PostgreSQL 9.1.24 手册

The PostgreSQL Global Development Group

PostgreSQL 9.1.24 手册

The PostgreSQL Global Development Group
版权 © 1996-2016 PostgreSQL 全球开发组

法律声明

PostgreSQL 版权所有 © 1996-2016 PostgreSQL 全球开发组，并按照下文所述加利福尼亚大学的许可条款分发。

Postgres95 版权所有 © 1994-5 加利福尼亚大学董事会。

特此授权，可为任何目的使用、复制、修改和分发本软件及其文档，无需付费，也无需书面协议，
但条件是所有副本中都必须包含上述版权声明、本段以及后续两段内容。

在任何情况下，加利福尼亚大学都不对任何一方承担因使用本软件及其文档而产生的任何直接、间接、特殊、
附带或后果性损害赔偿责任，包括利润损失，即使加利福尼亚大学已被告知此类损害发生的可能性。

加利福尼亚大学明确不作任何担保，包括但不限于适销性和特定用途适用性的默示担保。本协议项下提供的软件按“原样”提供，
加利福尼亚大学没有义务提供维护、支持、更新、增强或修改。

目录

前言 xxxii
1. 什么是PostgreSQL? xxxii
2. PostgreSQL简史 xxxiii
2.1. 伯克利 POSTGRES 项目 xxxiii
2.2. Postgres95 xxxiii
2.3. PostgreSQL xxxiv
3. 约定 xxxiv
4. 更多信息 xxxiv
5. 缺陷报告指南 xxxv
5.1. 识别缺陷 xxxv
5.2. 报告哪些内容 xxxvi
5.3. 到哪里报告缺陷 xxxvii
I. 教程 1
1. 从头开始 3
1.1. 安装 3
1.2. 架构基础 3
1.3. 创建数据库 3
1.4. 访问数据库 5
2. SQL语言 7
2.1. 简介 7
2.2. 概念 7
2.3. 创建一个新表 7
2.4. 向表中填充行 8
2.5. 查询表 9
2.6. 表之间的连接 10
2.7. 聚合函数 12
2.8. 更新 14
2.9. 删除 14
3. 高级特性 16
3.1. 简介 16
3.2. 视图 16
3.3. 外键 16
3.4. 事务 17
3.5. 窗口函数 19
3.6. 继承 21
3.7. 小结 23
II. SQL 语言 24
4. SQL语法 31
4.1. 词法结构 31
4.2. 值表达式 39
4.3. 调用函数 50
5. 数据定义 52
5.1. 表基础 52
5.2. 默认值 53
5.3. 约束 54
5.4. 系统列 60
5.5. 修改表 61
5.6. 权限 64
5.7. 模式 65
5.8. 继承 68
5.9. 分区 71

5.10. 外部数据 78
5.11. 其他数据库对象 79
5.12. 依赖跟踪 79
6. 数据操纵 81
6.1. 插入数据 81
6.2. 更新数据 82
6.3. 删除数据 82
7. 查询 84
7.1. 概述 84
7.2. 表表达式 84
7.3. 选择列表 94
7.4. 组合查询 96
7.5. 行排序 96
7.6. LIMIT和OFFSET 97
7.7. VALUES列表 98
7.8. WITH查询（公共表表达式） 99
8. 数据类型 104
8.1. 数字类型 105
8.2. 货币类型 109
8.3. 字符类型 110
8.4. 二进制数据类型 111
8.5. 日期/时间类型 113
8.6. 布尔类型 122
8.7. 枚举类型 123
8.8. 几何类型 125
8.9. 网络地址类型 127
8.10. 位串类型 129
8.11. 文本搜索类型 129
8.12. UUID类型 132
8.13. XML类型 132
8.14. 数组 134
8.15. 复合类型 142
8.16. 对象标识符类型 145
8.17. 伪类型 147
9. 函数和操作符 149
9.1. 逻辑操作符 149
9.2. 比较操作符 149
9.3. 数学函数和操作符 151
9.4. 字符串函数和操作符 156
9.5. 二进制串函数和操作符 168
9.6. 位串函数和操作符 171
9.7. 模式匹配 172
9.8. 数据类型格式化函数 185
9.9. 日期/时间函数和操作符 191
9.10. 枚举支持函数 203
9.11. 几何函数和操作符 204
9.12. 网络地址函数和操作符 207
9.13. 文本搜索函数和操作符 210
9.14. XML 函数 215
9.15. 序列操作函数 225
9.16. 条件表达式 227
9.17. 数组函数和操作符 230
9.18. 聚合函数 233

9.19. 窗口函数 ... 236
9.20. 子查询表达式 ... 237
9.21. 行和数组比较 ... 240
9.22. 集合返回函数 ... 242
9.23. 系统信息函数 ... 244
9.24. 系统管理函数 ... 253
9.25. 触发器函数 ... 260
10. 类型转换 ... 262
10.1. 概述 ... 262
10.2. 操作符 ... 263
10.3. 函数 ... 266
10.4. 值存储 ... 269
10.5. UNION、CASE及相关结构 ... 269
11. 索引 ... 272
11.1. 简介 ... 272
11.2. 索引类型 ... 273
11.3. 多列索引 ... 274
11.4. 索引和ORDER BY ... 275
11.5. 组合多个索引 ... 275
11.6. 唯一索引 ... 276
11.7. 表达式索引 ... 277
11.8. 部分索引 ... 277
11.9. 操作符类和操作符族 ... 279
11.10. 索引和排序规则 ... 280
11.11. 检查索引使用情况 ... 281
12. 全文搜索 ... 283
12.1. 介绍 ... 283
12.2. 表和索引 ... 286
12.3. 控制文本搜索 ... 288
12.4. 附加特性 ... 293
12.5. 解析器 ... 298
12.6. 词典 ... 300
12.7. 配置示例 ... 308
12.8. 测试和调试文本搜索 ... 309
12.9. GiST 和 GIN 索引类型 ... 313
12.10. psql 支持 ... 314
12.11. 限制 ... 317
12.12. 从 8.3 之前的文本搜索迁移 ... 318
13. 并发控制 ... 319
13.1. 介绍 ... 319
13.2. 事务隔离 ... 319
13.3. 显式锁定 ... 323
13.4. 应用级别的数据一致性检查 ... 328
13.5. 注意事项 ... 329
13.6. 锁定和索引 ... 329
14. 性能提示 ... 331
14.1. 使用EXPLAIN ... 331
14.2. 规划器使用的统计信息 ... 336
14.3. 用显式JOIN子句控制规划器 ... 337
14.4. 填充一个数据库 ... 339
14.5. 非持久设置 ... 341
III. 服务器管理 ... 343
15. 从源代码安装 ... 349

15.1. 简要说明 ... 349
15.2. 需求 ... 349
15.3. 获取源代码 ... 351
15.4. 安装过程 ... 351
15.5. 安装后设置 ... 361
15.6. 受支持的平台 ... 362
15.7. 平台相关说明 ... 363
16. 在 Windows 上从源代码安装 ... 372
16.1. 使用 Visual C++ 或 Platform SDK 构建 ... 372
16.2. 使用Visual C++或Borland C++构建libpq ... 376
17. 服务器设置和操作 ... 377
17.1. PostgreSQL用户账户 ... 377
17.2. 创建一个数据库集簇 ... 377
17.3. 启动数据库服务器 ... 378
17.4. 管理内核资源 ... 381
17.5. 关闭服务器 ... 389
17.6. 升级 PostgreSQL 集簇 ... 389
17.7. 防止服务器欺骗 ... 392
17.8. 加密选项 ... 392
17.9. 使用 SSL 的安全 TCP/IP 连接 ... 393
17.10. 通过 SSH 隧道建立安全 TCP/IP 连接 ... 395
18. 服务器配置 ... 397
18.1. 设置参数 ... 397
18.2. 文件位置 ... 398
18.3. 连接和认证 ... 399
18.4. 资源消耗 ... 403
18.5. 预写式日志 ... 407
18.6. 复制 ... 411
18.7. 查询规划 ... 413
18.8. 错误报告和日志 ... 418
18.9. 运行时统计数据 ... 426
18.10. 自动清理 ... 427
18.11. 客户端连接默认值 ... 429
18.12. 锁管理 ... 434
18.13. 版本和平台兼容性 ... 435
18.14. 错误处理 ... 437
18.15. 预置选项 ... 437
18.16. 自定义选项 ... 438
18.17. 开发者选项 ... 439
18.18. 短选项 ... 441
19. 客户端认证 ... 443
19.1. `pg_hba.conf` 文件 ... 443
19.2. 用户名映射 ... 449
19.3. 认证方法 ... 450
19.4. 认证问题 ... 457
20. 数据库角色 ... 458
20.1. 数据库角色 ... 458
20.2. 角色属性 ... 459
20.3. 角色成员资格 ... 460
20.4. 删除角色 ... 461
20.5. 函数和触发器安全性 ... 462
21. 管理数据库 ... 463
21.1. 概述 ... 463

21.2. 创建数据库 ... 463
21.3. 模板数据库 ... 464
21.4. 数据库配置 ... 465
21.5. 删除数据库 ... 465
21.6. 表空间 ... 466
22. 本地化 ... 468
22.1. 区域设置支持 ... 468
22.2. 排序规则支持 ... 470
22.3. 字符集支持 ... 472
23. 日常数据库维护任务 ... 479
23.1. 日常清理 ... 479
23.2. 日常重建索引 ... 484
23.3. 日志文件维护 ... 485
24. 备份和恢复 ... 486
24.1. SQL转储 ... 486
24.2. 文件系统级备份 ... 489
24.3. 持续归档和时间点恢复（PITR） ... 489
25. 高可用、负载均衡和复制 ... 499
25.1. 不同方案的比较 ... 499
25.2. 日志传送备库 ... 502
25.3. 故障切换 ... 507
25.4. 日志传送的另一种方法 ... 508
25.5. 热备 ... 509
26. 恢复配置 ... 516
26.1. 归档恢复设置 ... 516
26.2. 恢复目标设置 ... 517
26.3. 备库设置 ... 517
27. 监控数据库活动 ... 519
27.1. 标准 Unix 工具 ... 519
27.2. 统计收集器 ... 520
27.3. 查看锁 ... 528
27.4. 动态追踪 ... 528
28. 监控磁盘使用情况 ... 536
28.1. 确定磁盘使用情况 ... 536
28.2. 磁盘已满故障 ... 537
29. 可靠性与预写式日志 ... 538
29.1. 可靠性 ... 538
29.2. 预写式日志（WAL） ... 539
29.3. 异步提交 ... 540
29.4. WAL 配置 ... 541
29.5. WAL 内部机制 ... 542
30. 回归测试 ... 544
30.1. 运行测试 ... 544
30.2. 测试结果评估 ... 547
30.3. 变体比较文件 ... 549
30.4. 测试覆盖率检查 ... 549
IV. 客户端接口 ... 551
31. libpq - C 库 ... 556
31.1. 数据库连接控制函数 ... 556
31.2. 连接状态函数 ... 565
31.3. 命令执行函数 ... 569
31.4. 异步命令处理 ... 582
31.5. 取消进行中的查询 ... 586
31.6. 快速路径接口 ... 587

31.7. 异步通知 ... 588
31.8. 与COPY命令相关的函数 ... 589
31.9. 控制函数 ... 592
31.10. 杂项函数 ... 594
31.11. 通知处理 ... 596
31.12. 事件系统 ... 597
31.13. 环境变量 ... 603
31.14. 密码文件 ... 605
31.15. 连接服务文件 ... 605
31.16. 连接参数的 LDAP 查找 ... 605
31.17. SSL 支持 ... 606
31.18. 在线程化程序中的行为 ... 610
31.19. 构建libpq程序 ... 610
31.20. 示例程序 ... 611
32. 大对象 ... 622
32.1. 简介 ... 622
32.2. 实现特性 ... 622
32.3. 客户端接口 ... 622
32.4. 服务器端函数 ... 625
32.5. 示例程序 ... 626
33. ECPG - C 中的嵌入式 SQL ... 632
33.1. 概念 ... 632
33.2. 管理数据库连接 ... 632
33.3. 运行 SQL 命令 ... 635
33.4. 使用主变量 ... 637
33.5. 动态 SQL ... 650
33.6. pgtypes 库 ... 652
33.7. 使用描述符区域 ... 665
33.8. 错误处理 ... 678
33.9. 预处理器指令 ... 685
33.10. 处理嵌入式 SQL 程序 ... 687
33.11. 库函数 ... 687
33.12. 大对象 ... 688
33.13. C++ 应用 ... 690
33.14. 嵌入式 SQL 命令 ... 693
33.15. Informix兼容模式 ... 717
33.16. 内部 ... 732
34. 信息模式 ... 734
34.1. 模式 ... 734
34.2. 数据类型 ... 734
34.3. information_schema_catalog_name ... 735
34.4. administrable_role_authorizations ... 735
34.5. applicable_roles ... 735
34.6. attributes ... 736
34.7. character_sets ... 737
34.8. check_constraint_routine_usage ... 738
34.9. check_constraints ... 739
34.10. collations ... 739
34.11. collation_character_set_applicability ... 739
34.12. column_domain_usage ... 740
34.13. column_privileges ... 740
34.14. column_udt_usage ... 741
34.15. columns ... 741

34.16. constraint_column_usage 744
34.17. constraint_table_usage 744
34.18. data_type_privileges 745
34.19. domain_constraints 745
34.20. domain_udt_usage 746
34.21. domains 746
34.22. element_types 748
34.23. enabled_roles 750
34.24. foreign_data_wrapper_options 750
34.25. foreign_data_wrappers 750
34.26. foreign_server_options 751
34.27. foreign_servers 751
34.28. foreign_table_options 751
34.29. foreign_tables 752
34.30. key_column_usage 752
34.31. parameters 752
34.32. referential_constraints 754
34.33. role_column_grants 755
34.34. role_routine_grants 755
34.35. role_table_grants 756
34.36. role_usage_grants 756
34.37. routine_privileges 757
34.38. routines 757
34.39. schemata 761
34.40. sequences 761
34.41. sql_features 762
34.42. sql_implementation_info 762
34.43. sql_languages 763
34.44. sql_packages 763
34.45. sql_parts 764
34.46. sql_sizing 764
34.47. sql_sizing_profiles 764
34.48. table_constraints 765
34.49. table_privileges 765
34.50. tables 766
34.51. triggered_update_columns 767
34.52. triggers 767
34.53. usage_privileges 768
34.54. user_mapping_options 769
34.55. user_mappings 769
34.56. view_column_usage 770
34.57. view_routine_usage 770
34.58. view_table_usage 771
34.59. views 771
V. 服务器编程 773
35. 扩展 SQL 778
35.1. 可扩展性如何运作 778
35.2. PostgreSQL类型系统 778
35.3. 用户定义的函数 779
35.4. 查询语言（SQL）函数 780
35.5. 函数重载 792
35.6. 函数易变性分类 793

35.7. 过程语言函数 ... 794
35.8. 内部函数 ... 794
35.9. C 语言函数 ... 795
35.10. 用户定义的聚合 ... 817
35.11. 用户定义的类型 ... 819
35.12. 用户定义的操作符 ... 822
35.13. 操作符优化信息 ... 823
35.14. 索引扩展接口 ... 827
35.15. 将相关对象打包成扩展 ... 837
35.16. 扩展构建基础设施 ... 842
36. 触发器 ... 846
36.1. 触发器行为概述 ... 846
36.2. 数据更改的可见性 ... 848
36.3. 用 C 编写触发器函数 ... 848
36.4. 一个完整的触发器示例 ... 851
37. 规则系统 ... 855
37.1. 查询树 ... 855
37.2. 视图和规则系统 ... 856
37.3. INSERT、UPDATE和DELETE规则 ... 863
37.4. 规则和权限 ... 873
37.5. 规则和命令状态 ... 874
37.6. 规则与触发器 ... 875
38. 过程语言 ... 878
38.1. 安装过程语言 ... 878
39. PL/pgSQL – SQL 过程语言 ... 881
39.1. 概述 ... 881
39.2. PL/pgSQL的结构 ... 882
39.3. 声明 ... 883
39.4. 表达式 ... 888
39.5. 基本语句 ... 889
39.6. 控制结构 ... 895
39.7. 游标 ... 906
39.8. 错误和消息 ... 911
39.9. 触发器函数 ... 913
39.10. PL/pgSQL 内部机制 ... 920
39.11. PL/pgSQL开发提示 ... 923
39.12. 从Oracle PL/SQL 移植 ... 925
40. PL/Tcl - Tcl 过程语言 ... 935
40.1. 概述 ... 935
40.2. PL/Tcl 函数和参数 ... 935
40.3. PL/Tcl 中的数据值 ... 936
40.4. PL/Tcl 中的全局数据 ... 936
40.5. 从 PL/Tcl 访问数据库 ... 937
40.6. PL/Tcl 中的触发器函数 ... 939
40.7. 模块与unknown命令 ... 941
40.8. Tcl 过程名 ... 941
41. PL/Perl - Perl 过程语言 ... 942
41.1. PL/Perl 函数和参数 ... 942
41.2. PL/Perl 中的数据值 ... 946
41.3. 内置函数 ... 946
41.4. PL/Perl 中的全局值 ... 951
41.5. 受信任与不受信任的 PL/Perl ... 952
41.6. PL/Perl 触发器 ... 953

41.7. PL/Perl 内部机制 ... 954
42. PL/Python - Python 过程语言 ... 956
42.1. Python 2 与 Python 3 ... 956
42.2. PL/Python 函数 ... 957
42.3. 数据值 ... 958
42.4. 共享数据 ... 963
42.5. 匿名代码块 ... 963
42.6. 触发器函数 ... 963
42.7. 数据库访问 ... 964
42.8. 显式子事务 ... 966
42.9. 辅助函数 ... 967
42.10. 环境变量 ... 968
43. 服务器编程接口 ... 969
43.1. 接口函数 ... 969
43.2. 接口支持函数 ... 999
43.3. 内存管理 ... 1007
43.4. 数据更改的可见性 ... 1016
43.5. 示例 ... 1016
VI. 参考 ... 1020
I. SQL 命令 ... 1025
ABORT ... 1028
ALTER AGGREGATE ... 1029
ALTER COLLATION ... 1031
ALTER CONVERSION ... 1032
ALTER DATABASE ... 1033
ALTER DEFAULT PRIVILEGES ... 1035
ALTER DOMAIN ... 1038
ALTER EXTENSION ... 1041
ALTER FOREIGN DATA WRAPPER ... 1044
ALTER FOREIGN TABLE ... 1046
ALTER FUNCTION ... 1049
ALTER GROUP ... 1052
ALTER INDEX ... 1053
ALTER LANGUAGE ... 1055
ALTER LARGE OBJECT ... 1056
ALTER OPERATOR ... 1057
ALTER OPERATOR CLASS ... 1059
ALTER OPERATOR FAMILY ... 1060
ALTER ROLE ... 1064
ALTER SCHEMA ... 1067
ALTER SEQUENCE ... 1068
ALTER SERVER ... 1071
ALTER TABLE ... 1072
ALTER TABLESPACE ... 1082
ALTER TEXT SEARCH CONFIGURATION ... 1084
ALTER TEXT SEARCH DICTIONARY ... 1086
ALTER TEXT SEARCH PARSER ... 1088
ALTER TEXT SEARCH TEMPLATE ... 1089
ALTER TRIGGER ... 1090
ALTER TYPE ... 1091
ALTER USER ... 1094
ALTER USER MAPPING ... 1095
ALTER VIEW ... 1096
ANALYZE ... 1098

BEGIN 1100
CHECKPOINT 1102
CLOSE 1103
CLUSTER 1104
COMMENT 1106
COMMIT 1110
COMMIT PREPARED 1111
COPY 1112
CREATE AGGREGATE 1121
CREATE CAST 1124
CREATE COLLATION 1128
CREATE CONVERSION 1130
CREATE DATABASE 1132
CREATE DOMAIN 1135
CREATE EXTENSION 1137
CREATE FOREIGN DATA WRAPPER 1139
CREATE FOREIGN TABLE 1141
CREATE FUNCTION 1143
CREATE GROUP 1150
CREATE INDEX 1151
CREATE LANGUAGE 1157
CREATE OPERATOR 1160
CREATE OPERATOR CLASS 1163
CREATE OPERATOR FAMILY 1166
CREATE ROLE 1167
CREATE RULE 1172
CREATE SCHEMA 1175
CREATE SEQUENCE 1177
CREATE SERVER 1180
CREATE TABLE 1182
CREATE TABLE AS 1195
CREATE TABLESPACE 1198
CREATE TEXT SEARCH CONFIGURATION 1200
CREATE TEXT SEARCH DICTIONARY 1201
CREATE TEXT SEARCH PARSER 1203
CREATE TEXT SEARCH TEMPLATE 1205
CREATE TRIGGER 1206
CREATE TYPE 1211
CREATE USER 1218
CREATE USER MAPPING 1219
CREATE VIEW 1220
DEALLOCATE 1223
DECLARE 1224
DELETE 1227
DISCARD 1230
DO 1231
DROP AGGREGATE 1233
DROP CAST 1234
DROP COLLATION 1235
DROP CONVERSION 1236
DROP DATABASE 1237
DROP DOMAIN 1238
DROP EXTENSION 1239
DROP FOREIGN DATA WRAPPER 1240

DROP FOREIGN TABLE 1241
DROP FUNCTION 1242
DROP GROUP 1244
DROP INDEX 1245
DROP LANGUAGE 1246
DROP OPERATOR 1247
DROP OPERATOR CLASS 1249
DROP OPERATOR FAMILY 1251
DROP OWNED 1253
DROP ROLE 1254
DROP RULE 1255
DROP SCHEMA 1256
DROP SEQUENCE 1257
DROP SERVER 1258
DROP TABLE 1259
DROP TABLESPACE 1260
DROP TEXT SEARCH CONFIGURATION 1261
DROP TEXT SEARCH DICTIONARY 1262
DROP TEXT SEARCH PARSER 1263
DROP TEXT SEARCH TEMPLATE 1264
DROP TRIGGER 1265
DROP TYPE 1266
DROP USER 1267
DROP USER MAPPING 1268
DROP VIEW 1269
END 1270
EXECUTE 1271
EXPLAIN 1272
FETCH 1277
GRANT 1281
INSERT 1287
LISTEN 1290
LOAD 1292
LOCK 1293
MOVE 1296
NOTIFY 1298
PREPARE 1300
PREPARE TRANSACTION 1302
REASSIGN OWNED 1304
REINDEX 1305
RELEASE SAVEPOINT 1308
RESET 1309
REVOKE 1310
ROLLBACK 1314
ROLLBACK PREPARED 1315
ROLLBACK TO SAVEPOINT 1316
SAVEPOINT 1318
SECURITY LABEL 1320
SELECT 1322
SELECT INTO 1338
SET 1340
SET CONSTRAINTS 1343
SET ROLE 1344
SET SESSION AUTHORIZATION 1346

SET TRANSACTION 1348
SHOW 1350
START TRANSACTION 1352
TRUNCATE 1353
UNLISTEN 1355
UPDATE 1357
VACUUM 1361
VALUES 1364
II. PostgreSQL 客户端应用 1367
clusterdb 1368
createdb 1371
createlang 1374
createuser 1377
dropdb 1381
droplang 1383
dropuser 1386
ecpg 1389
pg_basebackup 1391
pg_config 1395
pg_dump 1398
pg_dumpall 1407
pg_restore 1412
psql 1419
reindexdb 1448
vacuumdb 1451
III. PostgreSQL 服务器端应用 1454
initdb 1455
pg_controldata 1458
pg_ctl 1459
pg_resetxlog 1464
postgres 1466
postmaster 1473
VII. 内部 1474
44. PostgreSQL 内部概述 1479
44.1. 查询的路径 1479
44.2. 连接是如何建立的 1479
44.3. 解析器阶段 1480
44.4. PostgreSQL 规则系统 1481
44.5. 规划器/优化器 1481
44.6. 执行器 1482
45. 系统目录 1484
45.1. 概述 1484
45.2. `pg_aggregate` 1485
45.3. `pg_am` 1486
45.4. `pg_amop` 1488
45.5. `pg_amproc` 1489
45.6. `pg_attrdef` 1489
45.7. `pg_attribute` 1490
45.8. `pg_authid` 1492
45.9. `pg_auth_members` 1493
45.10. `pg_cast` 1494
45.11. `pg_class` 1495
45.12. `pg_constraint` 1498

45.13. pg_collation 1500
45.14. pg_conversion 1501
45.15. pg_database 1501
45.16. pg_db_role_setting 1503
45.17. pg_default_acl 1503
45.18. pg_depend 1504
45.19. pg_description 1505
45.20. pg_enum 1506
45.21. pg_extension 1506
45.22. pg_foreign_data_wrapper 1507
45.23. pg_foreign_server 1508
45.24. pg_foreign_table 1508
45.25. pg_index 1508
45.26. pg_inherits 1511
45.27. pg_language 1511
45.28. pg_largeobject 1512
45.29. pg_largeobject_metadata 1513
45.30. pg_namespace 1513
45.31. pg_opclass 1513
45.32. pg_operator 1514
45.33. pg_opfamily 1515
45.34. pg_pltemplate 1515
45.35. pg_proc 1516
45.36. pg_rewrite 1519
45.37. pg_seclabel 1520
45.38. pg_shdepend 1521
45.39. pg_shdescription 1522
45.40. pg_statistic 1522
45.41. pg_tablespace 1524
45.42. pg_trigger 1525
45.43. pg_ts_config 1526
45.44. pg_ts_config_map 1527
45.45. pg_ts_dict 1527
45.46. pg_ts_parser 1528
45.47. pg_ts_template 1528
45.48. pg_type 1528
45.49. pg_user_mapping 1534
45.50. 系统视图 1534
45.51. pg_available_extensions 1535
45.52. pg_available_extension_versions 1535
45.53. pg_cursors 1536
45.54. pg_group 1537
45.55. pg_indexes 1537
45.56. pg_locks 1537
45.57. pg_prepared_statements 1540
45.58. pg_prepared_xacts 1541
45.59. pg_roles 1541
45.60. pg_rules 1542
45.61. pg_seclabels 1543
45.62. pg_settings 1543
45.63. pg_shadow 1545
45.64. pg_stats 1546

45.65. pg_tables ... 1548
45.66. pg_timezone_abbrevs ... 1548
45.67. pg_timezone_names ... 1549
45.68. pg_user ... 1549
45.69. pg_user_mappings ... 1550
45.70. pg_views ... 1550
46. 前端/后端协议 ... 1552
46.1. 概述 ... 1552
46.2. 消息流 ... 1553
46.3. 流复制协议 ... 1563
46.4. 消息数据类型 ... 1566
46.5. 消息格式 ... 1566
46.6. 错误和通知消息字段 ... 1581
46.7. 自协议 2.0 以来的变更总结 ... 1582
47. PostgreSQL 编码约定 ... 1584
47.1. 格式 ... 1584
47.2. 在服务器内部报告错误 ... 1584
47.3. 错误消息风格指南 ... 1586
48. 本地语言支持 ... 1591
48.1. 给翻译者 ... 1591
48.2. 给程序员 ... 1593
49. 编写过程语言调用处理器 ... 1596
50. 编写外部数据包装器 ... 1599
50.1. 外部数据包装器函数 ... 1599
50.2. 外部数据包装器回调例程 ... 1599
51. 遗传查询优化器 ... 1602
51.1. 将查询处理看成是一个复杂的优化问题 ... 1602
51.2. 遗传算法 ... 1602
51.3. PostgreSQL 中的遗传查询优化（GEQO） ... 1603
51.4. 进一步阅读 ... 1604
52. 索引访问方法接口定义 ... 1605
52.1. 索引的目录条目 ... 1605
52.2. 索引访问方法函数 ... 1606
52.3. 索引扫描 ... 1609
52.4. 索引锁定注意事项 ... 1610
52.5. 索引唯一性检查 ... 1611
52.6. 索引代价估算函数 ... 1612
53. GiST 索引 ... 1616
53.1. 简介 ... 1616
53.2. 可扩展性 ... 1616
53.3. 实现 ... 1616
53.4. 示例 ... 1624
54. GIN 索引 ... 1625
54.1. 简介 ... 1625
54.2. 可扩展性 ... 1625
54.3. 实现 ... 1627
54.4. GIN 提示和技巧 ... 1628
54.5. 限制 ... 1628
54.6. 示例 ... 1629
55. 数据库物理存储 ... 1630
55.1. 数据库文件布局 ... 1630
55.2. TOAST ... 1632
55.3. 空闲空间映射 ... 1633

55.4. 可见性映射 1633
55.5. 初始化分支 1634
55.6. 数据库页面布局 1634
56. BKI 后端接口 1637
56.1. BKI 文件格式 1637
56.2. BKI 命令 1637
56.3. 引导 BKI 文件的结构 1638
56.4. 示例 1639
57. 规划器如何使用统计信息 1640
57.1. 行估计示例 1640
VIII. 附录 1646
A. PostgreSQL错误代码 1667
B. Date/Time Support 1675
B.1. Date/Time Input Interpretation 1675
B.2. Date/Time Key Words 1676
B.3. Date/Time Configuration Files 1677
B.4. History of Units 1678
C. SQL 关键字 1680
D. SQL 符合性 1705
D.1. 支持的特性 1706
D.2. 不支持的特性 1720
E. 发布说明 1733
E.1. 版本 9.1.24 1733
E.2. 版本 9.1.23 1734
E.3. 版本 9.1.22 1737
E.4. 版本 9.1.21 1738
E.5. 版本 9.1.20 1739
E.6. 版本 9.1.19 1743
E.7. 版本 9.1.18 1746
E.8. 版本 9.1.17 1747
E.9. 版本 9.1.16 1747
E.10. 版本 9.1.15 1751
E.11. 版本 9.1.14 1756
E.12. 版本 9.1.13 1759
E.13. 版本 9.1.12 1760
E.14. 版本 9.1.11 1764
E.15. 版本 9.1.10 1765
E.16. 版本 9.1.9 1767
E.17. 版本 9.1.8 1770
E.18. 版本 9.1.7 1772
E.19. 版本 9.1.6 1775
E.20. 版本 9.1.5 1776
E.21. 版本 9.1.4 1779
E.22. 版本 9.1.3 1782
E.23. 版本 9.1.2 1786
E.24. 版本 9.1.1 1790
E.25. 版本 9.1 1791
E.26. 版本 9.0.23 1808
E.27. 版本 9.0.22 1811
E.28. 版本 9.0.21 1811
E.29. 版本 9.0.20 1812
E.30. 版本 9.0.19 1815
E.31. 版本 9.0.18 1820
E.32. 版本 9.0.17 1822

E.33. 版本 9.0.16 1823
E.34. 版本 9.0.15 1827
E.35. 版本 9.0.14 1828
E.36. 版本 9.0.13 1830
E.37. 版本 9.0.12 1832
E.38. 版本 9.0.11 1834
E.39. 版本 9.0.10 1836
E.40. 版本 9.0.9 1837
E.41. 版本 9.0.8 1839
E.42. 版本 9.0.7 1841
E.43. 版本 9.0.6 1845
E.44. 版本 9.0.5 1847
E.45. 版本 9.0.4 1851
E.46. 版本 9.0.3 1853
E.47. 版本 9.0.2 1855
E.48. 版本 9.0.1 1857
E.49. 版本 9.0 1858
E.50. 版本 8.4.22 1877
E.51. 版本 8.4.21 1879
E.52. 版本 8.4.20 1880
E.53. 版本 8.4.19 1883
E.54. 版本 8.4.18 1885
E.55. 版本 8.4.17 1886
E.56. 版本 8.4.16 1888
E.57. 版本 8.4.15 1889
E.58. 版本 8.4.14 1891
E.59. 版本 8.4.13 1892
E.60. 版本 8.4.12 1893
E.61. 版本 8.4.11 1895
E.62. 版本 8.4.10 1898
E.63. 版本 8.4.9 1900
E.64. 版本 8.4.8 1903
E.65. 版本 8.4.7 1905
E.66. 版本 8.4.6 1906
E.67. 版本 8.4.5 1908
E.68. 版本 8.4.4 1911
E.69. 版本 8.4.3 1913
E.70. 版本 8.4.2 1916
E.71. 版本 8.4.1 1919
E.72. 版本 8.4 1921
E.73. 版本 8.3.23 1941
E.74. 版本 8.3.22 1942
E.75. 版本 8.3.21 1944
E.76. 版本 8.3.20 1945
E.77. 版本 8.3.19 1946
E.78. 版本 8.3.18 1948
E.79. 版本 8.3.17 1950
E.80. 版本 8.3.16 1952
E.81. 版本 8.3.15 1954
E.82. 版本 8.3.14 1955
E.83. 版本 8.3.13 1956
E.84. 版本 8.3.12 1958
E.85. 版本 8.3.11 1961
E.86. 版本 8.3.10 1962

E.87. 版本 8.3.9 1964
E.88. 版本 8.3.8 1966
E.89. 版本 8.3.7 1968
E.90. 版本 8.3.6 1970
E.91. 版本 8.3.5 1972
E.92. 版本 8.3.4 1973
E.93. 版本 8.3.3 1975
E.94. 版本 8.3.2 1976
E.95. 版本 8.3.1 1978
E.96. 版本 8.3 1980
E.97. 版本 8.2.23 1996
E.98. 版本 8.2.22 1997
E.99. 版本 8.2.21 1999
E.100. 版本 8.2.20 2000
E.101. 版本 8.2.19 2001
E.102. 版本 8.2.18 2003
E.103. 版本 8.2.17 2005
E.104. 版本 8.2.16 2006
E.105. 版本 8.2.15 2008
E.106. 版本 8.2.14 2010
E.107. 版本 8.2.13 2011
E.108. 版本 8.2.12 2012
E.109. 版本 8.2.11 2014
E.110. 版本 8.2.10 2015
E.111. 版本 8.2.9 2016
E.112. 版本 8.2.8 2017
E.113. 版本 8.2.7 2018
E.114. 版本 8.2.6 2020
E.115. 版本 8.2.5 2022
E.116. 版本 8.2.4 2023
E.117. 版本 8.2.3 2024
E.118. 版本 8.2.2 2024
E.119. 版本 8.2.1 2025
E.120. 版本 8.2 2026
E.121. 版本 8.1.23 2041
E.122. 版本 8.1.22 2042
E.123. 版本 8.1.21 2044
E.124. 版本 8.1.20 2045
E.125. 版本 8.1.19 2046
E.126. 版本 8.1.18 2047
E.127. 版本 8.1.17 2048
E.128. 版本 8.1.16 2049
E.129. 版本 8.1.15 2050
E.130. 版本 8.1.14 2051
E.131. 版本 8.1.13 2052
E.132. 版本 8.1.12 2053
E.133. 版本 8.1.11 2055
E.134. 版本 8.1.10 2056
E.135. 版本 8.1.9 2057
E.136. 版本 8.1.8 2058
E.137. 版本 8.1.7 2058
E.138. 版本 8.1.6 2059
E.139. 版本 8.1.5 2060
E.140. 版本 8.1.4 2061

E.141. 版本 8.1.3 2063
E.142. 版本 8.1.2 2064
E.143. 版本 8.1.1 2066
E.144. 版本 8.1 2067
E.145. 版本 8.0.26 2080
E.146. 版本 8.0.25 2082
E.147. 版本 8.0.24 2083
E.148. 版本 8.0.23 2084
E.149. 版本 8.0.22 2085
E.150. 版本 8.0.21 2086
E.151. 版本 8.0.20 2087
E.152. 版本 8.0.19 2087
E.153. 版本 8.0.18 2088
E.154. 版本 8.0.17 2089
E.155. 版本 8.0.16 2090
E.156. 版本 8.0.15 2091
E.157. 版本 8.0.14 2093
E.158. 版本 8.0.13 2093
E.159. 版本 8.0.12 2094
E.160. 版本 8.0.11 2094
E.161. 版本 8.0.10 2095
E.162. 版本 8.0.9 2096
E.163. 版本 8.0.8 2097
E.164. 版本 8.0.7 2098
E.165. 版本 8.0.6 2099
E.166. 版本 8.0.5 2100
E.167. 版本 8.0.4 2101
E.168. 版本 8.0.3 2102
E.169. 版本 8.0.2 2104
E.170. 版本 8.0.1 2106
E.171. 版本 8.0 2107
E.172. 版本 7.4.30 2121
E.173. 版本 7.4.29 2122
E.174. 版本 7.4.28 2123
E.175. 版本 7.4.27 2124
E.176. 版本 7.4.26 2125
E.177. 版本 7.4.25 2126
E.178. 版本 7.4.24 2126
E.179. 版本 7.4.23 2127
E.180. 版本 7.4.22 2128
E.181. 版本 7.4.21 2128
E.182. 版本 7.4.20 2129
E.183. 版本 7.4.19 2130
E.184. 版本 7.4.18 2131
E.185. 版本 7.4.17 2131
E.186. 版本 7.4.16 2132
E.187. 版本 7.4.15 2133
E.188. 版本 7.4.14 2133
E.189. 版本 7.4.13 2134
E.190. 版本 7.4.12 2135
E.191. 版本 7.4.11 2136
E.192. 版本 7.4.10 2136
E.193. 版本 7.4.9 2137
E.194. 版本 7.4.8 2138

E.195. 版本 7.4.7 2140
E.196. 版本 7.4.6 2141
E.197. 版本 7.4.5 2142
E.198. 版本 7.4.4 2142
E.199. 版本 7.4.3 2143
E.200. 版本 7.4.2 2144
E.201. 版本 7.4.1 2146
E.202. 版本 7.4 2148
E.203. 版本 7.3.21 2163
E.204. 版本 7.3.20 2164
E.205. 版本 7.3.19 2164
E.206. 版本 7.3.18 2165
E.207. 版本 7.3.17 2165
E.208. 版本 7.3.16 2166
E.209. 版本 7.3.15 2166
E.210. 版本 7.3.14 2167
E.211. 版本 7.3.13 2168
E.212. 版本 7.3.12 2169
E.213. 版本 7.3.11 2169
E.214. 版本 7.3.10 2170
E.215. 版本 7.3.9 2171
E.216. 版本 7.3.8 2172
E.217. 版本 7.3.7 2173
E.218. 版本 7.3.6 2173
E.219. 版本 7.3.5 2174
E.220. 版本 7.3.4 2175
E.221. 版本 7.3.3 2176
E.222. 版本 7.3.2 2178
E.223. 版本 7.3.1 2180
E.224. 版本 7.3 2181
E.225. 版本 7.2.8 2193
E.226. 版本 7.2.7 2194
E.227. 版本 7.2.6 2194
E.228. 版本 7.2.5 2195
E.229. 版本 7.2.4 2196
E.230. 版本 7.2.3 2196
E.231. 版本 7.2.2 2197
E.232. 版本 7.2.1 2198
E.233. 版本 7.2 2198
E.234. 版本 7.1.3 2209
E.235. 版本 7.1.2 2209
E.236. 版本 7.1.1 2210
E.237. 版本 7.1 2210
E.238. 版本 7.0.3 2215
E.239. 版本 7.0.2 2216
E.240. 版本 7.0.1 2216
E.241. 版本 7.0 2217
E.242. 版本 6.5.3 2224
E.243. 版本 6.5.2 2224
E.244. 版本 6.5.1 2225
E.245. 版本 6.5 2226
E.246. 版本 6.4.2 2230
E.247. 版本 6.4.1 2231
E.248. 版本 6.4 2232

E.249. 版本 6.3.2 ... 2236
E.250. 版本 6.3.1 ... 2237
E.251. 版本 6.3 ... 2238
E.252. 版本 6.2.1 ... 2242
E.253. 版本 6.2 ... 2243
E.254. 版本 6.1.1 ... 2246
E.255. 版本 6.1 ... 2246
E.256. 版本 6.0 ... 2249
E.257. 版本 1.09 ... 2251
E.258. 版本 1.02 ... 2251
E.259. 版本 1.01 ... 2253
E.260. 版本 1.0 ... 2255
E.261. Postgres95 Release 0.03 ... 2256
E.262. Postgres95 Release 0.02 ... 2258
E.263. Postgres95 Release 0.01 ... 2258
F. 额外提供的模块与扩展 ... 2259
F.1. adminpack ... 2260
F.2. auth_delay ... 2262
F.3. auto_explain ... 2263
F.4. btree_gin ... 2265
F.5. btree_gist ... 2266
F.6. chkpass ... 2268
F.7. citext – 大小写不敏感的字符串类型 ... 2269
F.8. cube – 多维立方体数据类型 ... 2272
F.9. dblink ... 2276
F.10. dict_int – 用于整数的示例全文搜索词典 ... 2307
F.11. dict_xsyn – 示例同义词全文检索词典 ... 2308
F.12. dummy_seclabel ... 2310
F.13. earthdistance – 计算大圆距离 ... 2311
F.14. file_fdw ... 2313
F.15. fuzzystrmatch – 确定字符串的相似性和距离 ... 2314
F.16. hstore ... 2317
F.17. intagg – 整数聚合器和枚举器 ... 2323
F.18. intarray ... 2325
F.19. isn ... 2328
F.20. lo – 管理大对象 ... 2332
F.21. ltree ... 2334
F.22. oid2name ... 2341
F.23. pageinspect ... 2345
F.24. passwordcheck ... 2347
F.25. pg_archivecleanup ... 2348
F.26. pgbench ... 2350
F.27. pg_buffercache ... 2357
F.28. pgcrypto ... 2359
F.29. pg_freespacemap ... 2370
F.30. pgrowlocks – 显示表的行锁信息 ... 2372
F.31. pg_standby ... 2374
F.32. pg_stat_statements ... 2377
F.33. pgstattuple – 获取元组级统计信息 ... 2381
F.34. pg_test_fsync ... 2383
F.35. pg_trgm ... 2384
F.36. pg_upgrade ... 2387
F.37. seg – 用于线段或浮点区间的数据类型 ... 2393
F.38. sepgsql – 基于 SELinux 标签的强制访问控制（MAC）安全模块 ... 2396

F.39. spi 2402
F.40. sslinfo 2404
F.41. tablefunc 2406
F.42. test_parser 2416
F.43. tsearch2 2418
F.44. unaccent 2420
F.45. uuid-ossp 2422
F.46. vacuumlo 2424
F.47. xml2 2426
G. 外部项目 2430
G.1. 客户端接口 2430
G.2. 管理工具 2430
G.3. 过程语言 2430
G.4. 扩展 2431
H. 源代码仓库 2432
H.1. 通过 Git 获取源代码 2432
I. 文档 2433
I.1. DocBook 2433
I.2. 工具集 2433
I.3. 构建文档 2437
I.4. 文档编写 2440
I.5. 风格指南 2442
J. 缩略词 2444
参考书目 2450
索引 2452

插图清单

9.1. 转换 SQL/XML 输出到 HTML 的 XSLT 样式表 225
51.1. 遗传算法的结构图 1603

表格清单

4.1. 反斜线转义序列 33
4.2. 操作符优先级（递减） 38
8.1. 数据类型 104
8.2. 数字类型 105
8.3. 货币类型 109
8.4. 字符类型 110
8.5. 特殊字符类型 111
8.6. 二进制数据类型 111
8.7. `bytea`字面量中需要转义的字节 112
8.8. `bytea`输出转义字节 113
8.9. 日期/时间类型 113
8.10. 日期输入 115
8.11. 时间输入 116
8.12. 时区输入 116
8.13. 特殊日期/时间输入 118
8.14. 日期/时间输出风格 118
8.15. 日期顺序习惯 118
8.16. ISO 8601 间隔单位缩写 120
8.17. 间隔输入 121
8.18. 间隔输出风格示例 122
8.19. 布尔数据类型 122
8.20. 几何类型 125
8.21. 网络地址类型 127
8.22. `cidr`类型输入示例 128
8.23. 对象标识符类型 146
8.24. 伪类型 147
9.1. 比较操作符 149
9.2. 数学操作符 151
9.3. 数学函数 152
9.4. 随机函数 155
9.5. 三角函数 156
9.6. SQL字符串函数和操作符 156
9.7. 其他字符串函数 158
9.8. 内置转换 165
9.9. SQL二进制串函数和操作符 168
9.10. 其他二进制串函数 169
9.11. 位串操作符 171
9.12. 正则表达式匹配操作符 174
9.13. 正则表达式原子 178
9.14. 正则表达式量词 178
9.15. 正则表达式约束 179
9.16. 正则表达式字符输入转义 180
9.17. 正则表达式字符类简写转义 181
9.18. 正则表达式约束转义 181
9.19. 正则表达式反向引用 182
9.20. ARE 嵌入选项字母 182
9.21. 格式化函数 185
9.22. 用于日期/时间格式化的模板模式 186
9.23. 用于日期/时间格式化的模板模式修饰符 188
9.24. 用于数值格式化的模板模式 189
9.25. 用于数值格式化的模板模式修饰符 190

9.26. to_char示例 190
9.27. 日期/时间操作符 191
9.28. 日期/时间函数 192
9.29. AT TIME ZONE 变体 201
9.30. 枚举支持函数 203
9.31. 几何操作符 204
9.32. 几何函数 205
9.33. 几何类型转换函数 206
9.34. cidr 和 inet 操作符 207
9.35. cidr 和 inet 函数 208
9.36. macaddr 函数 210
9.37. 文本搜索操作符 210
9.38. 文本搜索函数 211
9.39. 文本搜索调试函数 214
9.40. 序列函数 226
9.41. 数组操作符 230
9.42. 数组函数 231
9.43. 通用聚合函数 233
9.44. 用于统计的聚合函数 235
9.45. 通用窗口函数 236
9.46. 序列生成函数 242
9.47. 下标生成函数 243
9.48. 会话信息函数 244
9.49. 访问权限查询函数 246
9.50. 模式可见性查询函数 248
9.51. 系统目录信息函数 249
9.52. 注释信息函数 252
9.53. 事务 ID 和快照 252
9.54. 快照组件 252
9.55. 配置设定函数 253
9.56. 服务器信号函数 253
9.57. 备份控制函数 254
9.58. 恢复信息函数 255
9.59. 恢复控制函数 256
9.60. 数据库对象尺寸函数 256
9.61. 数据库对象位置函数 258
9.62. 通用文件访问函数 258
9.63. 咨询锁函数 259
12.1. 默认解析器的词元类型 298
13.1. 标准 SQL 事务隔离级别 320
13.2. 冲突的锁模式 325
17.1. System V IPC参数 381
17.2. PostgreSQL 共享内存用量 387
17.3. SSL 服务器文件用法 394
18.1. 消息严重级别 421
18.2. 短选项对照 441
22.1. PostgreSQL 字符集 473
22.2. 客户端/服务器字符集转换 476
25.1. 高可用、负载均衡和复制特性矩阵 501
27.1. 标准统计视图 521
27.2. 统计访问函数 523
27.3. 内置 DTrace 探针 529
27.4. 探针参数中使用的已定义类型 533

31.1. SSL 模式描述 608
31.2. libpq/客户端 SSL 文件用法 609
33.1. 在 PostgreSQL 数据类型和 C 变量类型之间映射 639
33.2. PGTYPESdate_from_asc的合法输入格式 656
33.3. PGTYPESdate_fmt_asc的合法输入格式 658
33.4. rdefmtdate的合法输入格式 658
33.5. PGTYPEStimestamp_from_asc的合法输入格式 659
34.1. information_schema_catalog_name 列 735
34.2. administrable_role_authorizations 列 735
34.3. applicable_roles 列 735
34.4. attributes 列 736
34.5. character_sets 列 738
34.6. check_constraint_routine_usage 列 738
34.7. check_constraints 列 739
34.8. collations 列 739
34.9. collation_character_set_applicability 列 740
34.10. column_domain_usage 列 740
34.11. column_privileges 列 740
34.12. column_udt_usage 列 741
34.13. columns 列 741
34.14. constraint_column_usage 列 744
34.15. constraint_table_usage 列 744
34.16. data_type_privileges 列 745
34.17. domain_constraints 列 745
34.18. domain_udt_usage 列 746
34.19. domains 列 746
34.20. element_types 列 748
34.21. enabled_roles 列 750
34.22. foreign_data_wrapper_options 列 750
34.23. foreign_data_wrappers 列 750
34.24. foreign_server_options 列 751
34.25. foreign_servers 列 751
34.26. foreign_table_options 列 751
34.27. foreign_tables 列 752
34.28. key_column_usage 列 752
34.29. parameters 列 753
34.30. referential_constraints 列 754
34.31. role_column_grants 列 755
34.32. role_routine_grants 列 755
34.33. role_table_grants 列 756
34.34. role_usage_grants 列 756
34.35. routine_privileges 列 757
34.36. routines 列 757
34.37. schemata 列 761
34.38. sequences 列 761
34.39. sql_features 列 762
34.40. sql_implementation_info 列 762
34.41. sql_languages 列 763
34.42. sql_packages 列 763
34.43. sql_parts 列 764
34.44. sql_sizing 列 764
34.45. sql_sizing_profiles 列 765

34.46. table_constraints 列 765
34.47. table_privileges 列 765
34.48. tables 列 766
34.49. triggered_update_columns 列 767
34.50. triggers 列 767
34.51. usage_privileges 列 768
34.52. user_mapping_options 列 769
34.53. user_mappings 列 769
34.54. view_column_usage 列 770
34.55. view_routine_usage 列 770
34.56. view_table_usage 列 771
34.57. views 列 771
35.1. 内置 SQL 类型等效的 C 类型 798
35.2. B-树策略 828
35.3. 哈希策略 828
35.4. GiST 二维 “R 树” 策略 828
35.5. GIN 数组策略 829
35.6. B-树支持函数 829
35.7. 哈希支持函数 829
35.8. GiST 支持函数 830
35.9. GIN 支持函数 830
45.1. 系统目录 1484
45.2. pg_aggregate 列 1485
45.3. pg_am 列 1486
45.4. pg_amop 列 1488
45.5. pg_amproc 列 1489
45.6. pg_attrdef 列 1489
45.7. pg_attribute 列 1490
45.8. pg_authid 列 1492
45.9. pg_auth_members 列 1494
45.10. pg_cast 列 1494
45.11. pg_class 列 1495
45.12. pg_constraint 列 1498
45.13. pg_collation 列 1500
45.14. pg_conversion 列 1501
45.15. pg_database 列 1501
45.16. pg_db_role_setting 列 1503
45.17. pg_default_acl 列 1503
45.18. pg_depend 列 1504
45.19. pg_description 列 1505
45.20. pg_enum 列 1506
45.21. pg_extension 列 1506
45.22. pg_foreign_data_wrapper 列 1507
45.23. pg_foreign_server 列 1508
45.24. pg_foreign_table 列 1508
45.25. pg_index 列 1509
45.26. pg_inherits 列 1511
45.27. pg_language 列 1511
45.28. pg_largeobject 列 1512
45.29. pg_largeobject_metadata 列 1513
45.30. pg_namespace 列 1513
45.31. pg_opclass 列 1514

45.32. pg_operator 列 1514
45.33. pg_opfamily 列 1515
45.34. pg_pltemplate列 1516
45.35. pg_proc 列 1516
45.36. pg_rewrite 列 1519
45.37. pg_seclabel 列 1520
45.38. pg_shdepend 列 1521
45.39. pg_shdescription 列 1522
45.40. pg_statistic 列 1523
45.41. pg_tablespace 列 1524
45.42. pg_trigger 列 1525
45.43. pg_ts_config 列 1526
45.44. pg_ts_config_map 列 1527
45.45. pg_ts_dict 列 1527
45.46. pg_ts_parser 列 1528
45.47. pg_ts_template 列 1528
45.48. pg_type 列 1529
45.49. typcategory Codes 1533
45.50. pg_user_mapping 列 1534
45.51. 系统视图 1534
45.52. pg_available_extensions 列 1535
45.53. pg_available_extension_versions 列 1535
45.54. pg_cursors 列 1536
45.55. pg_group 列 1537
45.56. pg_indexes 列 1537
45.57. pg_locks 列 1538
45.58. pg_prepared_statements 列 1540
45.59. pg_prepared_xacts 列 1541
45.60. pg_roles 列 1541
45.61. pg_rules 列 1542
45.62. pg_seclabels 列 1543
45.63. pg_settings 列 1544
45.64. pg_shadow 列 1545
45.65. pg_stats 列 1546
45.66. pg_tables 列 1548
45.67. pg_timezone_abbrevs 列 1549
45.68. pg_timezone_names 列 1549
45.69. pg_user 列 1549
45.70. pg_user_mappings 列 1550
45.71. pg_views 列 1550
55.1. PGDATA 的内容 1630
55.2. Page Layout 1634
55.3. PageHeaderData Layout 1635
55.4. HeapTupleHeaderData Layout 1636
A.1. PostgreSQL错误代码 1667
B.1. Month Names 1676
B.2. 星期名称 1676
B.3. Date/Time Field Modifiers 1676
C.1. SQL 关键字 1680
F.1. adminpack 函数 2260
F.2. 已弃用的adminpack函数 2260
F.3. cube 外部表示 2272

F.4. cube GiST 操作符 2272
F.5. cube 函数 2273
F.6. 基于立方体的地球距离函数 2311
F.7. 基于点的地球距离操作符 2312
F.8. hstore 操作符 2317
F.9. hstore Functions 2319
F.10. intarray 函数 2325
F.11. intarray 操作符 2325
F.12. isn Data Types 2328
F.13. isn Functions 2329
F.14. ltree Operators 2335
F.15. ltree Functions 2336
F.16. pg_buffercache 列 2357
F.17. crypt()支持的算法 2360
F.18. crypt()的迭代计数 2360
F.19. Hash 算法速度 2361
F.20. 有无 OpenSSL 时的功能概要 2367
F.21. pgrowlocks 输出列 2372
F.22. pg_stat_statements 列 2377
F.23. pgstattuple 输出列 2381
F.24. pg_trgm函数 2384
F.25. pg_trgm操作符 2385
F.26. seg 外部表示 2393
F.27. 合法 seg 输入示例 2394
F.28. seg GiST 操作符 2394
F.29. tablefunc函数 2406
F.30. connectby参数 2413
F.31. 用于生成 UUID 的函数 2422
F.32. 返回 UUID 常量的函数 2422
F.33. 函数 2426
F.34. xpath_table 参数 2427
G.1. 由外部维护的客户端接口 2430
G.2. 由外部维护的过程语言 2431

范例清单

8.1. 使用字符类型 111
8.2. 使用boolean类型 123
8.3. 使用位串类型 129
10.1. 阶乘操作符类型解析 264
10.2. 字符串连接操作符类型解析 264
10.3. 绝对值与取反操作符类型解析 265
10.4. 域类型上的自定义操作符 266
10.5. round 函数参数类型解析 267
10.6. substr 函数类型解析 268
10.7. character 的存储类型转换 269
10.8. 未充分指定类型的 UNION 中的类型解析 270
10.9. 简单 UNION 中的类型解析 270
10.10. 次序对调的 UNION 中的类型解析 270
11.1. 建立一个部分索引以排除常见值 277
11.2. 建立一个部分索引以排除不感兴趣的值 278
11.3. 建立一个部分唯一索引 279
19.1. 示例 pg_hba.conf 项 447
19.2. 示例 pg_ident.conf 文件 450
31.1. libpq示例程序 1 611
31.2. libpq示例程序 2 614
31.3. libpq示例程序 3 616
32.1. 用libpq操作大对象的示例程序 626
33.1. 示例 SQLDA 程序 674
33.2. 访问大对象的 ECPG 程序 688
38.1. 手工安装 PL/Perl 879
39.1. 在动态查询中为值加引号 892
39.2. UPDATE/INSERT的异常 905
39.3. 一个 PL/pgSQL 触发器函数 914
39.4. 一个用于审计的 PL/pgSQL 触发器函数 915
39.5. 一个用于审计的 PL/pgSQL 视图触发器函数 916
39.6. 一个用于维护汇总表的 PL/pgSQL 触发器函数 917
39.7. 把一个简单函数从 PL/SQL 移植到 PL/pgSQL 926
39.8. 把一个创建其它函数的函数从 PL/SQL 移植到 PL/pgSQL 927
39.9. 把一个带字符串处理和 OUT 参数的过程从 PL/SQL 移植到 PL/pgSQL 928
39.10. 把一个过程从 PL/SQL 移植到 PL/pgSQL 930

前言

本书是PostgreSQL的官方文档。它由PostgreSQL开发人员及其他志愿者在 PostgreSQL软件开发过程中同步编写。它描述了当前版本的PostgreSQL正式支持的全部功能。

为了便于管理有关PostgreSQL的大量信息，本书被组织为若干部分。
每一部分都面向不同类型的用户，或者面向处于不同PostgreSQL使用阶段的用户：

- 第 I 部分是面向新用户的非正式介绍。
- 第 II 部分记述了SQL查询语言环境，包括数据类型、函数以及用户级性能调优。每个PostgreSQL用户都应阅读这一部分。
- 第 III 部分描述了服务器的安装和管理。无论是为自己使用还是为他人提供服务，凡是运行PostgreSQL服务器的人都应阅读这一部分。
- 第 IV 部分描述PostgreSQL客户端程序的编程接口。
- 第 V 部分包含面向高级用户的服务器可扩展能力相关信息，其中的主题包括用户定义数据类型和函数。
- 第 VI 部分包含有关 SQL 命令、客户端程序和服务器程序的参考信息。该部分以按命令或程序组织的结构化信息为其他部分提供支持。
- 第 VII 部分包含可能对PostgreSQL开发人员有用的各类信息。

1. 什么是PostgreSQL?

PostgreSQL是一种基于 POSTGRES, Version 4.2[1] 的对象关系数据库管理系统（ORDBMS），后者由加州大学伯克利分校计算机科学系开发。POSTGRES 首创了许多概念，而这些概念直到很久之后才在某些商业数据库系统中出现。

PostgreSQL是这套原始伯克利代码的开源后继版本。它支持 SQL 标准中的很大一部分，并提供了许多现代特性：

- 复杂查询
- 外键
- 触发器
- 视图
- 事务完整性
- 多版本并发控制

此外，PostgreSQL还允许用户以多种方式扩展，例如添加新的：

- 数据类型
- 函数
- 操作符
- 聚合函数
- 索引方法
- 过程语言

由于采用宽松的许可证，任何人都可以出于任何目的免费使用、修改和分发PostgreSQL，
无论是私人、商业还是学术用途。

[1] http://db.cs.berkeley.edu/postgres.html

2. PostgreSQL简史

如今名为PostgreSQL的对象关系数据库管理系统，源自加州大学伯克利分校编写的POSTGRES软件包。经过二十多年的发展，PostgreSQL如今已成为当今最先进的开源数据库。

2.1. 伯克利 POSTGRES 项目

由 Michael Stonebraker 教授领导的 POSTGRES 项目，得到了国防高级研究计划局（DARPA）、陆军研究办公室（ARO）、国家科学基金会（NSF）以及 ESL, Inc. 的资助。POSTGRES 的实现始于 1986 年。该系统的最初概念见于[ston86]，而最初数据模型的定义见于[rowe87]。当时规则系统的设计见于[ston87a]。存储管理器的设计动机和体系结构详见[ston87b]。

自那以后，POSTGRES 经历了几次重大版本发布。第一个“演示版”系统于 1987 年开始运行，并在 1988 年的 ACM-SIGMOD 大会上进行了展示。版本 1 在[ston90a]中有描述，并于 1989 年 6 月发布给少数外部用户。针对第一代规则系统所受到的批评（[ston89]），规则系统被重新设计（[ston90b]），版本 2 连同新的规则系统于 1990 年 6 月发布。版本 3 于 1991 年问世，增加了对多个存储管理器的支持、改进后的查询执行器以及重写后的规则系统。此后直到 Postgres95（见下文）之前的后续版本，主要聚焦于可移植性和可靠性。

POSTGRES 被用于实现许多不同的研究和生产应用。这些应用包括金融数据分析系统、喷气发动机性能监测软件包、小行星跟踪数据库、医疗信息数据库以及若干地理信息系统。POSTGRES 还在多所大学中被用作教学工具。最后，Illustra Information Technologies（后来并入 Informix[2]，该公司现归 IBM[3] 所有）接手了这套代码并将其商业化。在 1992 年末，POSTGRES 成为Sequoia 2000 科学计算项目[4]的主要数据管理器。

1993 年期间，外部用户社区的规模几乎翻了一番。人们越来越清楚地意识到，对原型代码的维护和支持工作占用了大量本应用于数据库研究的时间。为了减轻这一支持负担，伯克利 POSTGRES 项目在版本 4.2 时正式结束。

2.2. Postgres95

1994 年，Andrew Yu 和 Jolly Chen 为 POSTGRES 添加了一个 SQL 语言解释器。随后，它以 Postgres95 这一新名称发布到网络上，作为最初的 POSTGRES 伯克利代码的开源后继者自行发展。

Postgres95的代码完全采用 ANSI C 编写，并将体积缩减了 25%。许多内部更改提升了性能和可维护性。Postgres951.0.x 版本在 Wisconsin Benchmark 上快了大约 30-50%，比较对象是 POSTGRES4.2 版。除了错误修复之外，主要增强还包括：

- 查询语言 PostQUEL 被 SQL 所取代（在服务器中实现）。直到 PostgreSQL（见下文）才支持子查询，但在 Postgres95 中可以用用户定义的 SQL 函数来模拟。聚合函数也得到了重新实现，并增加了对 `GROUP BY` 查询子句的支持。
- 新增了一个用于交互式 SQL 查询的程序（psql），它使用了 GNU Readline。这在很大程度上取代了旧的 monitor 程序。
- 新增的前端库 `libpgtcl` 支持基于 Tcl 的客户端。示例 shell `pgtclsh` 提供了新的 Tcl 命令，用于让 Tcl 程序与 Postgres95 服务器交互。
- 大对象接口经过了全面改造。反转大对象成为存储大对象的唯一机制。（反转文件系统已被移除。）

[2] http://www.informix.com/
[3] http://www.ibm.com/
[4] http://meteora.ucsd.edu/s2k/s2k_home.html

- 实例级规则系统被移除。规则仍然可以作为重写规则使用。
- 源代码中附带了一份简短教程，介绍常规 SQL 特性以及 Postgres95 的特性。
- 构建时使用的是 GNU make（而不是 BSD make）。此外，Postgres95 可以用未经修改的 GCC 编译（双精度浮点数的数据对齐问题已修复）。

2.3. PostgreSQL

到了 1996 年，很明显“Postgres95”这个名字经不起时间的考验。我们选择了一个新名字 PostgreSQL，以体现最初的 POSTGRES 与后来具备 SQL 能力的版本之间的关系。同时，我们将版本号从 6.0 开始，使编号重新回到最初由伯克利 POSTGRES 项目开启的序列中。

许多人出于传统，或因为更容易发音，仍然将PostgreSQL称为“Postgres”（如今很少全部大写）。这一用法作为昵称或别名已被广泛接受。

在开发 Postgres95 期间，重点是识别并理解服务器代码中已有的问题。到了 PostgreSQL 阶段，重点则转向增强特性和能力，不过各个方面的工作仍在继续。

自那以后 PostgreSQL 发生了哪些变化，可见附录 E。

3. 约定

本书使用以下排版约定来标记某些文本：新术语、外来语和其他重要段落以斜体强调。所有表示计算机输入或输出的内容，特别是命令、程序代码和屏幕输出，均以等宽字体（示例）显示。在这类内容中，斜体（示例）表示占位符，你必须用实际值来替换占位符。有时，程序代码中相对前一个示例新增或修改的部分会以粗体（**示例**）强调。

以下约定用于命令概要：方括号（[和]）表示可选部分。（在 Tcl 命令的概要中，改用问号（?）表示，这是 Tcl 的惯例。）大括号（{和}）和竖线（|）表示必须选择其中一个替代项。点（...）表示前面的元素可以重复。

在有助于增强清晰度时，SQL 命令前会加上提示符=>，shell 命令前会加上提示符$。不过，通常并不显示提示符。

一般来说，管理员是负责安装和运行服务器的人。用户则可以是任何正在使用或希望使用PostgreSQL系统任一部分的人。不应把这些术语理解得过于狭隘；本书并不对系统管理流程作固定预设。

4. 更多信息

除了本文档，也就是本书之外，还有其他关于PostgreSQL的资源：

Wiki

PostgreSQL wiki[5] 收录了该项目的FAQ[6]（常见问题解答）列表、TODO[7] 列表，以及许多其他主题的详细信息。

网站

PostgreSQL 网站[8]提供最新版本的详细信息，以及其他能让你在工作中或日常使用PostgreSQL时更加高效的信息。

[5] http://wiki.postgresql.org
[6] http://wiki.postgresql.org/wiki/Frequently_Asked_Questions
[7] http://wiki.postgresql.org/wiki/Todo
[8] http://www.postgresql.org

邮件列表

邮件列表是一个让你的问题得到解答、与其他用户分享经验以及联系开发者的好地方。
详情请参阅PostgreSQL网站。

你自己！

PostgreSQL是一个开源项目。因此，它依赖用户社区提供持续支持。随着你开始使用PostgreSQL，无论是通过本文档还是通过邮件列表，你都会依赖他人的帮助。
请考虑把你的知识回馈给社区。阅读邮件列表并回答问题。
如果你学到了一些文档中尚未包含的内容，就把它写出来并贡献出去。
如果你为代码增加了功能，也请将其贡献出来。

5. 缺陷报告指南

当你在PostgreSQL中发现缺陷时，我们希望得知此事。你的缺陷报告对于提高 PostgreSQL的可靠性十分重要，因为即使再怎么小心，也无法保证 PostgreSQL的每一部分都能在每个平台、任何情况下正常工作。

以下建议旨在帮助你组织缺陷报告，以便能够高效处理。没有人必须遵循这些建议，
但这样做通常对各方都有好处。

我们无法承诺立刻修复每一个缺陷。如果该缺陷明显、严重，或者影响大量用户，
很可能会有人去调查它。也可能我们会让你升级到更新的版本，
看看在新版本中是否也会出现同样的缺陷。或者，我们可能会认定，
在某些计划中的重大重写完成之前，这个缺陷无法修复。再或者，它只是过于棘手，
而当前还有更重要的事情要处理。如果你需要立即获得帮助，请考虑购买商业支持合同。

5.1. 识别缺陷

在报告缺陷之前，请反复阅读文档，以确认你确实可以完成正在尝试做的事情。
如果从文档中无法明确看出某件事是否可行，也请报告；那就是文档中的缺陷。
如果程序实际做的事情与文档描述不同，那就是缺陷。这可能包括但不限于以下情况：

- 程序因致命信号而终止，或者给出一条表明程序本身存在问题的操作系统错误消息。（反例可能是“磁盘满”之类的消息，因为那得由你自己解决。）
- 程序对给定输入产生了错误的输出。
- 程序拒绝接受有效输入（按文档定义）。
- 程序接受了无效输入，却没有给出提示或错误消息。但请记住，你认为的无效输入，可能正是我们认为的扩展功能，或者是对传统做法的兼容。
- 在受支持的平台上，PostgreSQL无法按照说明完成编译、构建或安装。

这里的“程序”指的是任何可执行程序，不只是后端进程。

运行缓慢或消耗大量资源，并不一定就是缺陷。请阅读文档，
或在某个邮件列表中寻求调优应用程序的帮助。不符合 SQL标准，也不一定是缺陷，除非明确声称该特定特性符合标准。

在继续之前，请查看 TODO 列表和 FAQ，确认你的缺陷是否已经是已知问题。如果你看不懂TODO 列表中的信息，也请报告你的问题。至少我们可以把 TODO 列表写得更清楚。

5.2. 报告哪些内容

关于缺陷报告，最重要的一点是陈述全部事实，而且只陈述事实。
不要猜测你觉得哪里出了问题、“它看起来像是在做什么”，或者程序的哪一部分有错。
如果你不熟悉实现，十有八九会猜错，对我们毫无帮助。即使你熟悉实现，
有根据的解释也只能是事实的有益补充，而不能替代事实。如果我们要修复这个缺陷，
仍然必须先亲自重现它。陈述这些基本事实并不难做（你大概可以直接从屏幕上复制粘贴），
但太多时候，重要细节会被遗漏，因为有人觉得它们不重要，或者认为即使不说清楚，
报告也一样能被理解。

以下各项应包含在每一个缺陷报告中：

- 重现问题所需的精确步骤序列，而且必须是从程序启动开始。这些步骤应当是自包含的；
  如果输出依赖表中的数据，那么仅仅发来一条 `SELECT`语句，而不附带前面的`CREATE TABLE`和 `INSERT`语句，是不够的。我们没有时间去逆向推导你的数据库模式；如果要靠我们自己编造数据，多半会错过问题。

  对于 SQL 相关问题，测试用例的最佳形式是一个可由psql前端运行、并能展示问题的文件。（请务必确保你的`~/.psqlrc`启动文件中没有任何内容。）创建这个文件的一个简便方法，
  是用pg_dump导出搭建场景所需的表声明和数据，再附上触发问题的查询。
  我们鼓励你尽量缩小示例规模，但这并非绝对必要。如果缺陷可重现，
  我们无论如何都能找出来。

  如果你的应用使用的是别的客户端接口，例如PHP，那么请尽量隔离出触发问题的查询。
  我们大概不会为了重现你的问题去搭建一个 Web 服务器。无论如何，都请记得提供精确的输入文件；不要笼统地猜测问题发生在“大文件”或“中等大小的数据库”等情况下，
  因为这些信息不够精确，派不上用场。

- 你实际得到的输出。请不要只说它“没起作用”或“崩溃了”。如果有错误消息，
  请把它贴出来，即使你看不懂。如果程序因操作系统错误而终止，请说明是哪一种错误。
  如果什么都没发生，也请说清楚。即使你的测试用例导致程序崩溃，
  或出现其他显而易见的问题，这种情况在我们的平台上也未必会发生。如果可以，
  最简单的做法就是直接从终端复制输出。

  > **注意**
  >
  > 如果你要报告错误消息，请尽量获取其最详细的形式。在psql中，请预先执行`\set VERBOSITY verbose`。如果你是从服务器日志中提取消息，请将运行时参数log_error_verbosity设为`verbose`，以便记录全部细节。

  > **注意**
  >
  > 在致命错误的情况下，客户端报告的错误消息可能不包含全部可用信息。
  > 也请查看数据库服务器的日志输出。如果你平时没有保留服务器日志，
  > 现在正是开始这样做的好时机。

- 你期望得到的输出，也非常重要。如果你只是写“这个命令给了我那个输出。”
  或“这不是我期望的。”，我们可能会自己运行一遍，扫一眼输出，然后觉得它看起来没问题，
  也正是我们所期望的。我们不应该花时间去揣摩你的命令背后的确切语义。
  尤其不要只是说“这不是 SQL 规定的/Oracle 所做的。”从SQL标准中找出正确行为并不是一件

轻松的事，而且我们也不都了解其他所有关系数据库的行为。（如果你的问题是程序崩溃，显然可以省略这一项。）

- 任何命令行选项和其他启动选项，包括所有被你改动过、且与问题相关的环境变量或配置文件。同样，请提供精确的信息。如果你使用的是一种预打包发行版，并且它会在系统启动时自动启动数据库服务器，那么你应当尽量弄清楚它是怎么做到的。

- 任何偏离安装说明的做法。

- PostgreSQL的版本。你可以运行命令 `SELECT version();`来查明所连接服务器的版本。大多数可执行程序也支持 `--version`选项；至少`postgres --version`和`psql --version`应当可以工作。如果这个函数或这些选项都不存在，那就说明你的版本已经老到足以该升级了。如果你运行的是预打包版本，例如 RPM，请说明这一点，并附上该软件包可能带有的子版本号。如果你说的是一个 Git 快照，也请说明，并给出提交哈希值。

  如果你的版本早于 9.1.24，我们几乎肯定会建议你升级。每个新版本都会包含大量缺陷修复和改进，因此你在旧版 PostgreSQL中遇到的缺陷，很可能已经被修复。对于使用旧版PostgreSQL的站点，我们只能提供有限支持；如果你需要超出这一范围的帮助，请考虑购买商业支持合同。

- 平台信息。这包括内核名称和版本、C 库、处理器、内存信息等。在大多数情况下，报告供应商和版本就足够了，但不要想当然地认为所有人都知道“Debian”具体包含什么，或者所有人都运行在 i386 上。如果你遇到的是安装问题，那么你机器上的工具链信息（编译器、make等）也是必需的。

Do not be afraid if your bug report becomes rather lengthy. That is a fact of life. It is better to report everything the first time than us having to squeeze the facts out of you. On the other hand, if your input files are huge, it is fair to ask first whether somebody is interested in looking into it. Here is an article[9] that outlines some more tips on reporting bugs.

不要把全部时间都花在找出输入中的哪些变化会让问题消失上。这多半无助于解决问题。如果最后发现这个缺陷无法立刻修复，你仍然有时间去寻找并分享变通办法。还有，再强调一次，不要浪费时间去猜测这个缺陷为什么存在。我们迟早会查明原因。

在编写缺陷报告时，请避免使用含混的术语。整个软件包叫作“PostgreSQL”，有时简称“Postgres”。如果你特指后端进程，请明确说出来，不要只是说“PostgreSQL 崩溃了”。单个后端进程崩溃，与父“postgres”进程崩溃，是完全不同的事情；当你指的是某个单独的后端进程挂掉时，请不要说“服务器崩溃了”，反过来也一样。此外，客户端程序，例如交互式前端“psql”，与后端完全是分开的。请尽量明确问题是在客户端还是服务器端。

5.3. 到哪里报告缺陷

一般来说，请将缺陷报告发送到缺陷报告邮件列表 `<pgsql-bugs@postgresql.org>`。请为电子邮件使用描述性的主题，必要时可以摘取部分错误消息。

另一种方法是填写项目网站[10]上提供的缺陷报告网页表单。通过这种方式提交的缺陷报告，也会被发送到 `<pgsql-bugs@postgresql.org>`邮件列表。

如果你的缺陷报告涉及安全影响，而且你不希望它立刻出现在公开归档中，请不要发送到`pgsql-bugs`。安全问题可以私下报告给 `<security@postgresql.org>`。

[9] http://www.chiark.greenend.org.uk/~sgtatham/bugs.html
[10] http://www.postgresql.org/

不要将缺陷报告发送到任何用户邮件列表，例如 <pgsql-sql@postgresql.org>或 <pgsql-general@postgresql.org>。这些邮件列表用于回答用户问题，其订阅者通常并不希望收到缺陷报告。更重要的是，他们也不太可能去修复这些缺陷。

此外，请不要把报告发送到开发者邮件列表 <pgsql-hackers@postgresql.org>。这个列表是用来讨论 PostgreSQL开发的，最好把缺陷报告和这类讨论分开。如果这个问题需要更深入的审查，我们可能会选择在 pgsql-hackers上继续讨论你的缺陷报告。

如果你遇到的是文档问题，报告它的最佳地点是文档邮件列表 <pgsql-docs@postgresql.org>。请明确指出你对文档的哪一部分不满意。

如果你的缺陷是在某个不受支持平台上出现的可移植性问题，请发送邮件到 <pgsql-hackers@postgresql.org>，这样我们（以及你）就可以共同推进 PostgreSQL在该平台上的移植工作。

注意

由于垃圾邮件泛滥，上述所有电子邮件地址都是封闭邮件列表。也就是说，你必须订阅某个列表才能向它发帖。（不过，使用缺陷报告 Web 表单无需订阅。）如果你希望发邮件但不想接收列表邮件，可以订阅并把订阅选项设置为nomail。要了解更多信息，请向 <majordomo@postgresql.org>发送邮件，正文只写一个词 help。

部分 I. 教程

欢迎阅读PostgreSQL教程。本教程旨在介绍PostgreSQL、关系数据库概念以及 SQL 语言。我们假定读者具备基本的计算机使用常识，不要求具备特定的 Unix 或编程经验。本教程旨在让读者通过动手实践了解PostgreSQL系统的一些重要方面。它并不试图对所涵盖的主题作全面论述。

在成功完成本教程后，你可能会希望阅读第 II 部分部分，以更深入地理解 SQL 语言；或者阅读第 IV 部分，了解如何使用PostgreSQL开发应用程序。自行部署并管理 PostgreSQL 安装的读者也应阅读第 III 部分。

目录

1. 从头开始 3
1.1. 安装 3
1.2. 架构基础 3
1.3. 创建数据库 3
1.4. 访问数据库 5
2. SQL语言 7
2.1. 简介 7
2.2. 概念 7
2.3. 创建一个新表 7
2.4. 向表中填充行 8
2.5. 查询表 9
2.6. 表之间的连接 10
2.7. 聚合函数 12
2.8. 更新 14
2.9. 删除 14
3. 高级特性 16
3.1. 简介 16
3.2. 视图 16
3.3. 外键 16
3.4. 事务 17
3.5. 窗口函数 19
3.6. 继承 21
3.7. 小结 23

第 1 章 从头开始

1.1. 安装

当然，在你开始使用PostgreSQL之前，必须先安装它。
PostgreSQL可能已经安装在你的站点上，可能是因为它随操作系统发行版一起提供，
也可能是因为系统管理员已经安装了它。如果是这样，
你应该从操作系统文档或系统管理员那里了解如何访问PostgreSQL。

如果你不确定PostgreSQL是否已经可用，或者是否可以用它来做实验，那么你可以自行安装。
这并不难，而且也是一个不错的练习。PostgreSQL可由任何非特权用户安装，不需要超级用户（root）权限。

如果你要自己安装PostgreSQL，请参阅第 15 章中的安装说明，待安装完成后再回到本指南。务必仔细遵循关于设置适当环境变量的小节。

如果你的站点管理员没有按默认方式完成设置，你可能还需要做一些额外工作。例如，
如果数据库服务器所在的机器是远程机器，你就需要把`PGHOST`
环境变量设置为数据库服务器机器的名称。环境变量`PGPORT`也可能需要设置。
归根结底就是：如果你试着启动某个应用程序，而它抱怨无法连接到数据库，
那么你应该去咨询站点管理员；如果你自己就是管理员，
那就查阅文档以确保你的环境已经正确设置。如果你没有理解上一段的意思，请阅读下一节。

1.2. 架构基础

在继续之前，你应该先了解PostgreSQL的基本系统架构。了解PostgreSQL各部分如何交互，
会让本章更容易理解。

用数据库术语来说，PostgreSQL采用客户端/服务器模型。
一次PostgreSQL会话由以下协同工作的进程（程序）组成：

- 一个服务器进程，负责管理数据库文件、接受客户端应用到数据库的连接，
  并代表客户端执行数据库操作。数据库服务器程序叫做`postgres`。
- 用户用来执行数据库操作的客户端（前端）应用。
  客户端应用的形式可以非常多样：可以是一个文本工具，也可以是一个图形应用程序，
  或者是一个通过访问数据库来显示网页的 Web 服务器，还可以是专门的数据库维护工具。有些客户端应用随PostgreSQL发行版一起提供；大多数则由用户开发。

和典型的客户端/服务器应用一样，客户端和服务器可以位于不同的主机上。在这种情况下，
它们通过 TCP/IP 网络连接通信。你应该记住这一点，因为在客户端机器上可访问的文件，在数据库服务器机器上可能不可访问（或者只能使用不同的文件名访问）。

PostgreSQL服务器可以处理来自客户端的多个并发连接。为此，
它会为每个连接启动（“forks”）一个新进程。从那时起，客户端和新的服务器进程就无需原始`postgres`进程介入而直接通信。因此，这个主服务器进程始终在运行，等待客户端连接，
而客户端及其关联的服务器进程则不断创建和退出。（当然，这一切对用户都是不可见的。
这里只是为求完整而提及。）

1.3. 创建数据库

检验你能否访问数据库服务器的第一步，就是试着创建一个数据库。一个正在运行的
PostgreSQL服务器可以管理许多数据库。通常会为每个项目或每个用户使用一个独立的数据库。

你的站点管理员可能已经为你创建了一个可用的数据库。
他应该已经告诉你你的数据库名是什么。如果是这样，你就可以省略这一步，直接跳到下一节。

要创建一个新数据库（本例中命名为 mydb），请使用以下命令：

```
$ createdb mydb
```

如果这条命令没有任何输出，就说明这一步成功了，你可以跳过本节余下的内容。

如果你看到类似下面这样的信息：

```
createdb: command not found
```

那么说明PostgreSQL没有正确安装。要么根本没有安装，要么就是你的 shell 搜索路径没有把它包含进去。试着改用绝对路径调用该命令：

```
$ /usr/local/pgsql/bin/createdb mydb
```

你所在站点上的路径可能不同。请联系站点管理员，或者查看安装说明来修正这个问题。

另一种响应可能是这样：

```
createdb: could not connect to database postgres: could not connect to
 server: No such file or directory
        Is the server running locally and accepting
        connections on Unix domain socket "/tmp/.s.PGSQL.5432"?
```

这意味着服务器尚未启动，或者它没有在 createdb 期望的位置启动。同样，请查看安装说明或咨询管理员。

另一种响应可能是这样：

```
createdb: could not connect to database postgres: FATAL:  role "joe"
 does not exist
```

其中提到了你自己的登录名。如果管理员没有为你创建 PostgreSQL 用户帐号，就会发生这种情况。（PostgreSQL用户帐号与操作系统用户帐号是分开的。）如果你是管理员，请参阅第 20 章了解如何创建帐号。你需要切换成安装 PostgreSQL 时所用的操作系统用户（通常是 postgres），才能创建第一个用户帐号。也可能分配给你的 PostgreSQL 用户名与操作系统用户名不同；在这种情况下，你需要使用 -U 选项，或者设置 PGUSER 环境变量，来指定你的 PostgreSQL 用户名。

如果你有用户帐号，但它没有创建数据库所需的权限，那么你会看到下面的信息：

```
createdb: database creation failed: ERROR:  permission denied to
 create database
```

并非每个用户都被授权创建新数据库。如果 PostgreSQL 拒绝为你创建数据库，那么站点管理员就需要授予你创建数据库的权限。遇到这种情况时，请咨询站点管理员。如果 PostgreSQL 是你自己安装的，那么在本教程中，你应当以启动服务器时所用的用户帐号登录。[1]

你也可以创建其他名称的数据库。PostgreSQL允许你在一个给定站点上创建任意数量的数据库。数据库名的首字符必须是字母，并且长度限制为 63 字节。一种方便的做法是创建一个与当前用户

[1] 这之所以可行，是因为PostgreSQL用户名与操作系统用户帐号是分开的。当你连接到数据库时，可以选择以哪个PostgreSQL用户名连接；如果不指定，默认就会使用你当前操作系统帐号的名称。恰好总会有一个PostgreSQL用户帐号，其名称与启动服务器的操作系统用户相同，而该用户也总是有权创建数据库。除了直接以该用户身份登录之外，你还可以在各处使用 -U选项来选择连接时使用的PostgreSQL用户名。

名同名的数据库。许多工具都将该数据库名视为默认值，因此这样可以少打一些字。
要创建这个数据库，只需输入：

```
$ createdb
```

如果你不再想使用某个数据库，可以删除它。例如，如果你是数据库mydb的所有者（创建者），就可以使用下面的命令销毁它：

```
$ dropdb mydb
```

（对于这条命令，数据库名不会默认成用户帐号名，你始终需要显式指定它。）
这个操作会从物理上删除与该数据库关联的所有文件，而且无法撤销，因此务必三思而后行。

关于createdb和dropdb的更多信息，可分别参见createdb和dropdb。

1.4. 访问数据库

一旦你创建了数据库，你就可以通过以下方式访问它：

- 运行PostgreSQL的交互式终端程序 psql，它允许你以交互方式输入、编辑并执行SQL命令。
- 使用现有的图形前端工具，例如pgAdmin，或者使用具备ODBC或JDBC支持的办公套件来创建和操作数据库。这些方式不在本教程讨论范围内。
- 使用若干可用语言绑定之一编写自定义应用程序。关于这些方式的进一步讨论，请见第 IV 部分。

你可能需要启动psql来试验本教程中的示例。可以通过输入下面的命令连接到mydb数据库：

```
$ psql mydb
```

如果你不提供数据库名，那么它默认就是你的用户帐号名。在上一节使用createdb时，
你已经见识过这种机制了。

在psql中，你将看到下面的欢迎信息：

```
psql (9.1.24)
Type "help" for help.

mydb=>
```

最后一行也可能是：

```
mydb=#
```

这表示你是数据库超级用户；如果这个 PostgreSQL实例是你自己安装的，那很可能就是这种情况。超级用户意味着你不受访问控制限制。对本教程而言，这一点并不重要。

如果你在启动psql时遇到问题，请回到前一节。createdb和psql的诊断信息很相似，
如果前者能工作，后者也应该能工作。

由 psql 打印出的最后一行就是提示符，它表示 psql正在等待你的输入，你可以将 SQL 查询输入到由 psql 维护的工作区中。试试这些命令：

```
mydb=> SELECT version();
                                version
 ----------------------------------------------------------------------
---
```

```
 PostgreSQL 9.1.24 on i586-pc-linux-gnu, compiled by GCC 2.96, 32-bit
(1 row)

mydb=> SELECT current_date;
    date
------------
 2002-08-31
(1 row)

mydb=> SELECT 2 + 2;
 ?column?
----------
        4
(1 row)
```

psql程序有一些不属于 SQL 命令的内部命令。它们以反斜线“\”开头。例如，你可以输入下面的命令来查看各种PostgreSQL SQL命令的语法帮助：

```
mydb=> \h
```

要退出psql，输入：

```
mydb=> \q
```

psql将退出并返回到命令 shell。（要了解更多内部命令，可输入\?，在psql提示符下执行。）psql的完整功能见psql。在本教程中，我们不会明确使用这些功能，
但在有帮助时你可以自行使用它们。

第 2 章 SQL语言

2.1. 简介

本章概述如何使用SQL执行简单操作。本教程仅供你入门之用，绝不是一份完整的SQL教程。已有许多关于SQL的书籍，包括[melt93]和[date97]。你应当知道，PostgreSQL的某些语言特性是对标准的扩展。

在下面的示例中，我们假定你已经按照前一章所述创建了名为mydb的数据库，并且已经能够启动psql。

本手册中的示例也可以在PostgreSQL源代码发行包的src/tutorial/目录中找到。（PostgreSQL 的二进制发行包可能不包含这些源文件。）要使用这些源文件，请先切换到该目录并运行 make：

```
$ cd ..../src/tutorial
$ make
```

这会创建脚本并编译包含用户定义函数和类型的 C 文件。然后，要开始本教程，请执行以下操作：

```
$ cd ..../tutorial
$ psql -s mydb
...
mydb=> \i basics.sql
```

\i命令会从指定文件中读取命令。psql的 -s选项会让你进入单步模式，在将每条语句发送给服务器之前都会暂停。本节使用的命令都位于文件basics.sql中。

2.2. 概念

PostgreSQL是一种关系数据库管理系统（RDBMS）。这意味着它是一个用于管理存储在关系中的数据的系统。关系本质上是表的数学术语。如今，把数据存储在表中的观念已经十分常见，似乎显而易见，但组织数据库的方式还有很多。类 Unix 操作系统中的文件和目录就是层次数据库的一个例子。较新的发展则是面向对象数据库。

每个表都是一个有名字的行集合。给定表中的每一行都有相同的一组具名列，并且每一列都具有特定的数据类型。虽然列在每一行中的顺序是固定的，但一定要记住，SQL 并不以任何方式保证表内行的顺序（虽然可以为了显示而对它们显式排序）。

表被归入数据库，而由单个PostgreSQL服务器实例管理的数据库集合构成一个数据库集簇。

2.3. 创建一个新表

你可以通过指定表名以及各列的列名和类型来创建一个新表：

```
CREATE TABLE weather (
    city            varchar(80),
    temp_lo         int,           -- low temperature
    temp_hi         int,           -- high temperature
    prcp            real,          -- precipitation
```

```
    date            date
);
```

你可以在psql中按这种换行格式输入该命令。psql会识别出该命令在分号出现之前尚未结束。

在 SQL 命令中可以随意使用空白字符（即空格、制表符和换行符）。这意味着你可以按与上面不同的方式对齐该命令，甚至把它全部写在一行上。两个连字符（“--”）表示注释。
从它们开始直到行尾的内容都会被忽略。SQL 对关键字和标识符不区分大小写，除非用双引号括起标识符以保留大小写（上面的例子没有这么做）。

varchar(80)指定一种数据类型，它可以存储长度不超过 80 个字符的任意字符串。int是普通的整数类型。real是一种用于存储单精度浮点数的类型。date应当不言自明。（是的，类型为date的列也叫 date。这可能方便，也可能令人困惑 – 由你自己决定。）

PostgreSQL支持标准的SQL类型 int、smallint、real、double precision、char(*N*)、varchar(*N*)、date、time、timestamp和interval，
还支持其他一些通用类型以及丰富的几何类型。PostgreSQL 可以通过任意数量的用户定义数据类型进行定制。因此，类型名在语法中并不是关键字，只有在为了支持
SQL标准中的特殊情况而必须如此时才是例外。

第二个示例将存储城市及其关联的地理位置：

```
CREATE TABLE cities (
    name            varchar(80),
    location        point
);
```

point类型就是PostgreSQL特有数据类型的一个例子。

最后还要提一句，如果你不再需要某个表，或者想用不同的定义重新创建它，
可以使用下面的命令把它移除：

```
DROP TABLE tablename;
```

2.4. 向表中填充行

INSERT语句用于向表中填充行：

```
INSERT INTO weather VALUES ('San Francisco', 46, 50, 0.25, '1994-11
-27');
```

请注意，所有数据类型的输入格式都相当直观。那些并非简单数值的常量通常必须用单引号（'）括起，就像这个例子一样。date类型实际上可接受的格式相当灵活，不过在本教程中，
我们将坚持使用这里展示的这种无歧义格式。

point类型要求一个坐标对作为输入，如下所示：

```
INSERT INTO cities VALUES ('San Francisco', '(-194.0, 53.0)');
```

到目前为止所用的语法要求你记住列的顺序。另一种语法允许你显式列出列：

```
INSERT INTO weather (city, temp_lo, temp_hi, prcp, date)
    VALUES ('San Francisco', 43, 57, 0.0, '1994-11-29');
```

如果愿意，你可以按不同顺序列出列，甚至省略某些列，例如降水量未知时：

```
INSERT INTO weather (date, city, temp_hi, temp_lo)
    VALUES ('1994-11-29', 'Hayward', 54, 37);
```

许多开发者认为，与依赖隐含顺序相比，显式列出列是一种更好的风格。

请把上面展示的所有命令都输入一遍，这样你在下面各节中才有一些数据可供操作。

也可以使用COPY从纯文本文件装载大量数据。这通常更快，因为COPY命令针对这种用途进行了优化，不过灵活性不如INSERT。例如：

```
COPY weather FROM '/home/user/weather.txt';
```

其中，源文件必须可以在运行后端进程的机器上访问，而不是客户端机器，因为后端进程直接读取该文件。有关COPY命令的更多信息，请参见COPY。

2.5. 查询表

要从表中检索数据，就要查询该表。SQL的SELECT语句就是用来做这件事的。该语句分为选择列表（列出要返回的列）、表列表（列出从哪些表中检索数据）以及可选的限定条件（指定限制条件的部分）。例如，要检索表 weather的所有行，键入：

```
SELECT * FROM weather;
```

这里*是“所有列”的缩写。[1]因此，使用下面的语句也会得到相同结果：

```
SELECT city, temp_lo, temp_hi, prcp, date FROM weather;
```

输出应该是：

```
     city      | temp_lo | temp_hi | prcp |    date
---------------+---------+---------+------+------------
 San Francisco |      46 |      50 | 0.25 | 1994-11-27
 San Francisco |      43 |      57 |    0 | 1994-11-29
 Hayward       |      37 |      54 |      | 1994-11-29
(3 rows)
```

你可以在选择列表中编写表达式，而不仅仅是简单的列引用。例如，你可以这样做：

```
SELECT city, (temp_hi+temp_lo)/2 AS temp_avg, date FROM weather;
```

这样应该得到：

```
     city      | temp_avg |    date
---------------+----------+------------
 San Francisco |       48 | 1994-11-27
 San Francisco |       50 | 1994-11-29
 Hayward       |       45 | 1994-11-29
(3 rows)
```

注意这里AS子句如何用于重新命名输出列。（AS子句是可选的。）

查询可以被“限定”，方法是添加一个指定所需行的WHERE子句。WHERE子句包含一个布尔（真值）表达式，只有使该布尔表达式为真的行才会被返回。

[1] 虽然SELECT *对于即席查询很有用，但在生产代码中通常认为这是糟糕的风格，因为给表增加一个列就会改变结果。

在限定条件中可以使用常见的布尔操作符（AND、OR和NOT）。例如，下面的查询检索 San Francisco 在雨天的天气记录：

```
SELECT * FROM weather
    WHERE city = 'San Francisco' AND prcp > 0.0;
```

结果：

```
     city      | temp_lo | temp_hi | prcp |    date
---------------+---------+---------+------+------------
 San Francisco |      46 |      50 | 0.25 | 1994-11-27
(1 row)
```

你可以要求查询结果按排序后的顺序返回：

```
SELECT * FROM weather
    ORDER BY city;
```

```
     city      | temp_lo | temp_hi | prcp |    date
---------------+---------+---------+------+------------
 Hayward       |      37 |      54 |      | 1994-11-29
 San Francisco |      43 |      57 |    0 | 1994-11-29
 San Francisco |      46 |      50 | 0.25 | 1994-11-27
```

在这个例子中，排序顺序没有完全指定，因此两个 San Francisco 行可能以任意次序出现。但如果你使用下面的语句，就总会得到上面显示的结果：

```
SELECT * FROM weather
    ORDER BY city, temp_lo;
```

你可以要求从查询结果中去除重复行：

```
SELECT DISTINCT city
    FROM weather;
```

```
     city
---------------
 Hayward
 San Francisco
(2 rows)
```

这里同样要说明，结果行的顺序可能发生变化。你可以把DISTINCT 和ORDER BY一起使用，以确保结果一致：[2]

```
SELECT DISTINCT city
    FROM weather
    ORDER BY city;
```

2.6. 表之间的连接

到目前为止，我们的查询每次只访问一个表。查询可以同时访问多个表，也可以访问同一个表并同时处理其中的多行。一次访问同一个表或不同表中多行的查询称为连接查询。例如，假设要列出所有天气记录及其对应城市的位置。为此，需要比较city列在weather表各行中的值与name列在cities表所有行中的值，选出这些值匹配的行对。

[2] 在一些数据库系统中，包括较老版本的PostgreSQL，DISTINCT的实现会自动对行排序，因此 ORDER BY并非必需。但这并不是 SQL 标准的要求，而且当前的 PostgreSQL并不保证 DISTINCT会使行按顺序排列。

> **注意**
>
> 这只是一个概念模型。连接通常会以比真正比较每一个可能的行对更高效的方式执行，不过这对用户是不可见的。

下面的查询可以完成这一点：

```
SELECT *
    FROM weather, cities
    WHERE city = name;

     city      | temp_lo | temp_hi | prcp |    date    |     name
 | location
---------------+---------+---------+------+------------+--------------
-+-----------
 San Francisco |      46 |      50 | 0.25 | 1994-11-27 | San Francisco
 | (-194,53)
 San Francisco |      43 |      57 |    0 | 1994-11-29 | San Francisco
 | (-194,53)
(2 rows)
```

注意结果集的两点：

- 没有与城市 Hayward 对应的结果行。这是因为`cities`表中没有与 Hayward 匹配的项，所以连接会忽略`weather`表中的不匹配行。我们很快就会看到如何修正这一点。
- 有两列包含城市名称。这是正确的，因为来自`weather`和`cities`表的列列表被拼接在一起。不过，实际中通常不希望这样，因此可能更希望显式列出输出列，而不使用`*`：

  ```
  SELECT city, temp_lo, temp_hi, prcp, date, location
      FROM weather, cities
      WHERE city = name;
  ```

练习：. 试着确定省略`WHERE`子句时，这条查询的含义。

由于所有列的名称各不相同，解析器自动确定了它们所属的表。如果两个表中有重复的列名，就需要限定列名，以说明所指的是哪一列，例如：

```
SELECT weather.city, weather.temp_lo, weather.temp_hi,
       weather.prcp, weather.date, cities.location
    FROM weather, cities
    WHERE cities.name = weather.city;
```

通常认为，限定连接查询中的所有列名是一种良好风格，这样即使以后向某个表中添加了重复列名，查询也不会失败。

目前所见的连接查询，也可以采用下面这种形式编写：

```
SELECT *
    FROM weather INNER JOIN cities ON (weather.city = cities.name);
```

这种语法不如上面的语法常用，不过这里展示它是为了帮助理解后续主题。

现在来看看如何把 Hayward 的记录找回来。希望查询扫描`weather`表，并为其中每一行查找匹配的`cities`行。如果找不到匹配行，希望用一些“空值”代替`cities`表中的列。这种查询称为外连接。（此前所见的连接都是内连接。）命令如下：

```
SELECT *
    FROM weather LEFT OUTER JOIN cities ON (weather.city = cities.
name);

     city      | temp_lo | temp_hi | prcp |    date    |     name
 | location
---------------+---------+---------+------+------------+--------------
-+-----------
 Hayward       |      37 |      54 |      | 1994-11-29 |
  |
 San Francisco |      46 |      50 | 0.25 | 1994-11-27 | San Francisco
 | (-194,53)
 San Francisco |      43 |      57 |    0 | 1994-11-29 | San Francisco
 | (-194,53)
(3 rows)
```

这条查询称为左外连接，因为连接操作符左边的表，其每一行都至少在输出中出现一次，而右边的表只有与左表某行匹配的行才会输出。在输出某个没有右表匹配行的左表行时，右表的各列会用空值（null）代替。

练习：. 还有右外连接和全外连接。试着找出它们能做什么。

也可以把一个表与它自身连接，这称为自连接。例如，假设要找出处于其他天气记录温度范围内的所有天气记录。为此，需要比较`temp_lo`和`temp_hi`两列在每个`weather`行中的值与`temp_lo`和`temp_hi`两列在其他所有`weather`行中的值。可以使用下面的查询：

```
SELECT W1.city, W1.temp_lo AS low, W1.temp_hi AS high,
    W2.city, W2.temp_lo AS low, W2.temp_hi AS high
    FROM weather W1, weather W2
    WHERE W1.temp_lo < W2.temp_lo
    AND W1.temp_hi > W2.temp_hi;

     city      | low | high |     city      | low | high
---------------+-----+------+---------------+-----+------
 San Francisco |  43 |   57 | San Francisco |  46 |   50
 Hayward       |  37 |   54 | San Francisco |  46 |   50
(2 rows)
```

这里把 weather 表重新标记为`W1`和`W2`，以区分连接的左侧和右侧。在其他查询中也可以使用这种别名来减少输入，例如：

```
SELECT *
    FROM weather w, cities c
    WHERE w.city = c.name;
```

你会经常遇到这种缩写方式。

2.7. 聚合函数

与大多数其他关系数据库产品一样，PostgreSQL支持聚合函数。
聚合函数从多行输入中计算出单个结果。例如，有一些聚合函数可在一组行上计算`count`（计数）、`sum`（和）、`avg`（平均值）、`max`（最大值）和`min`（最小值）。

例如，我们可以用下面的语句找出所有记录中最低温度读数的最高值：

```
SELECT max(temp_lo) FROM weather;

 max
-----
  46
(1 row)
```

如果我们想知道该读数出现在哪个城市（或哪些城市），可能会尝试：

```
SELECT city FROM weather WHERE temp_lo = max(temp_lo);     错误
```

但这样行不通，因为聚合函数max不能用于 WHERE子句中（存在这个限制是因为WHERE 子句决定哪些行会被纳入聚合计算；因此显然它必须在聚合函数计算之前求值。）不过，这类情况下通常可以把查询改写成能够得到所需结果的形式，这里我们通过使用子查询来做到这一点：

```
SELECT city FROM weather
    WHERE temp_lo = (SELECT max(temp_lo) FROM weather);

     city
---------------
 San Francisco
(1 row)
```

这样做是可以的，因为子查询是一次独立的计算，它会与外层查询分开计算自己的聚合。

聚合与GROUP BY子句结合使用也很有用。例如，可以通过下面的查询，得到每个城市所观测到的最低温度的最大值：

```
SELECT city, max(temp_lo)
    FROM weather
    GROUP BY city;

     city      | max
---------------+-----
 Hayward       |  37
 San Francisco |  46
(2 rows)
```

这会为每个城市输出一行。每个聚合结果都是针对与该城市匹配的表行计算的。这些分组行可以用以下子句过滤：HAVING：

```
SELECT city, max(temp_lo)
    FROM weather
    GROUP BY city
    HAVING max(temp_lo) < 40;

  city   | max
---------+-----
 Hayward |  37
(1 row)
```

这样就只会为所有temp_lo值均小于 40 的城市返回相同的结果。最后，如果只关心名称以“S”开头的城市，可以这样做：

```
SELECT city, max(temp_lo)
    FROM weather
```

```
    WHERE city LIKE 'S%'❶
    GROUP BY city
    HAVING max(temp_lo) < 40;
```

❶ LIKE操作符执行模式匹配，详细说明见第 9.7 节。

理解聚合与SQL的WHERE和 HAVING子句之间的相互作用非常重要。WHERE和HAVING的根本区别在于：WHERE在分组和聚合计算之前选择输入行（因此它控制哪些行进入聚合计算），而HAVING在分组和聚合计算之后选择分组行。因此，WHERE子句中不能包含聚合函数；试图用聚合来决定哪些行应该成为聚合的输入是没有意义的。另一方面，HAVING子句总是包含聚合函数。（严格地说，也允许编写不使用聚合的HAVING子句，但那很少有用。同样的条件在WHERE阶段使用会更高效。）

在前面的例子中，我们可以在WHERE中应用城市名限制，因为它不需要聚合。这样比把限制放到HAVING中更高效，因为我们避免了对所有未通过WHERE检查的行进行分组和聚合计算。

2.8. 更新

你可以使用UPDATE命令更新已有的行。假设你发现 11 月 28 日之后的温度读数全部偏差了 2 度，那么就可以按下面的方式改正数据：

```
UPDATE weather
    SET temp_hi = temp_hi - 2,  temp_lo = temp_lo - 2
    WHERE date > '1994-11-28';
```

看看数据的新状态：

```
SELECT * FROM weather;

     city      | temp_lo | temp_hi | prcp |    date
---------------+---------+---------+------+------------
 San Francisco |      46 |      50 | 0.25 | 1994-11-27
 San Francisco |      41 |      55 |    0 | 1994-11-29
 Hayward       |      35 |      52 |      | 1994-11-29
(3 rows)
```

2.9. 删除

可以使用DELETE命令从表中移除行。假设你不再关心 Hayward 的天气，那么可以用下面的方法把这些行从表中删除：

```
DELETE FROM weather WHERE city = 'Hayward';
```

所有属于 Hayward 的天气记录都会被删除。

```
SELECT * FROM weather;

     city      | temp_lo | temp_hi | prcp |    date
---------------+---------+---------+------+------------
 San Francisco |      46 |      50 | 0.25 | 1994-11-27
 San Francisco |      41 |      55 |    0 | 1994-11-29
(2 rows)
```

对如下形式的语句必须保持警惕：

```
DELETE FROM tablename;
```

如果没有限定条件，`DELETE`将从指定表中删除所有行，使其变为空表。
系统在执行前不会请求你确认！

第 3 章 高级特性

3.1. 简介

在前一章中，我们已经介绍了如何使用SQL在PostgreSQL中存储和访问数据的基础知识。现在我们将讨论SQL的一些更高级特性，它们能够简化管理，并防止数据丢失或损坏。最后，我们还将看看一些PostgreSQL扩展。

本章有时会引用第 2 章中的示例，对其加以修改或改进，因此事先读过那一章会很有帮助。本章中的一些示例也可以在教程目录中的`advanced.sql`文件中找到。该文件还包含一些要装载的样例数据，这里不再重复。（关于如何使用该文件，参见第 2.1 节。）

3.2. 视图

回过头看第 2.6 节中的查询。假设天气记录和城市位置的组合清单对你的应用特别有用，但你又不想每次需要它时都键入这条查询。你可以在该查询之上创建一个视图，为该查询命名，以后就可以像引用普通表一样引用它：

```
CREATE VIEW myview AS
    SELECT city, temp_lo, temp_hi, prcp, date, location
        FROM weather, cities
        WHERE city = name;

SELECT * FROM myview;
```

大量使用视图是良好的 SQL 数据库设计的一个关键方面。视图允许你通过一致的接口封装表结构的细节，而这些细节可能会随着应用演进而变化。

几乎凡是真实表可以使用的地方，都可以使用视图。在其他视图之上再构建视图也很常见。

3.3. 外键

回想第 2 章中的`weather`和`cities`表。考虑这样一个问题：你希望确保没有人能向`weather`表中插入在`cities`表里没有匹配项的行。这称为维护数据的引用完整性。在较为简陋的数据库系统中，这通常是通过先查看`cities`表、检查是否存在匹配记录，然后再插入新的`weather`记录或拒绝插入来实现的（如果系统支持这种做法的话）。这种做法有很多问题，而且很不方便，所以PostgreSQL可以替你完成这件事。

新的表声明如下：

```
CREATE TABLE cities (
        city     varchar(80) primary key,
        location point
);

CREATE TABLE weather (
        city      varchar(80) references cities(city),
        temp_lo   int,
        temp_hi   int,
        prcp      real,
```

```
        date      date
);
```

现在试着插入一条无效记录：

```
INSERT INTO weather VALUES ('Berkeley', 45, 53, 0.0, '1994-11-28');

ERROR:  insert or update on table "weather" violates foreign key
 constraint "weather_city_fkey"
DETAIL:  Key (city)=(Berkeley) is not present in table "cities".
```

外键的行为可以针对应用进行精细调整。本教程不再超出这个简单例子继续展开，
更多信息请参见第 5 章。正确使用外键肯定能提高数据库应用的质量，因此强烈建议你了解它们。

3.4. 事务

事务是所有数据库系统中的一个基本概念。事务的要点在于，它把多个步骤打包成一个单一的、要么全部成功要么全部失败的操作。步骤之间的中间状态对其他并发事务不可见；
如果发生某种故障使事务无法完成，那么其中任何一步都不会对数据库产生影响。

例如，考虑一个银行数据库，其中保存着各个客户账户的余额，以及各营业网点的总存款余额。假设我们要记录一笔从Alice的账户向Bob的账户支付100.00美元的款项。大幅简化后，相应的SQL 命令可能是：

```
UPDATE accounts SET balance = balance - 100.00
    WHERE name = 'Alice';
UPDATE branches SET balance = balance - 100.00
    WHERE name = (SELECT branch_name FROM accounts WHERE name =
 'Alice');
UPDATE accounts SET balance = balance + 100.00
    WHERE name = 'Bob';
UPDATE branches SET balance = balance + 100.00
    WHERE name = (SELECT branch_name FROM accounts WHERE name = 'Bob')
;
```

这些命令的细节在这里并不重要；重要的是，为了完成这个相当简单的操作，
涉及了若干次彼此独立的更新。银行管理人员会希望确信这些更新要么全部发生，
要么一个也不发生。显然，不能因为系统故障导致Bob收到了100.00美元，
而Alice那边却没有被扣款。反过来，如果Alice被扣了款而Bob没有入账，她也不会满意。
我们需要保证：如果操作进行到一半出了问题，到目前为止已执行的步骤都不会生效。
把这些更新归为一个事务，就能提供这种保证。事务被称为原子的：从其他事务的角度看，
它要么完整发生，要么根本不发生。

我们还希望保证，一旦事务完成并得到数据库系统确认，它确实已经被永久记录下来，
即便紧接着发生崩溃也不会丢失。例如，如果我们正在记录Bob的一次现金提款，
我们当然不希望他刚走出银行大门，对他账户的扣款就在崩溃后消失。事务型数据库保证，
在把事务报告为完成之前，会先将该事务所做的全部更新记入永久存储（即磁盘）。

事务型数据库的另一个重要性质与原子更新的概念密切相关：当多个事务并发运行时，
每个事务都不应该看到其他事务尚未完成的更改。例如，
如果某个事务正在统计所有营业网点的余额，就不能让它只算进Alice所在网点的扣款，
却没有算进Bob所在网点的入账，反过来也不行。因此，
事务不仅在对数据库的持久影响上必须是全有或全无的，

在其发生过程中的可见性上也必须如此。一个未结束事务到目前为止所做的更新，对其他事务不可见；等到该事务完成时，所有更新会同时变得可见。

在PostgreSQL中，可以通过将事务中的 SQL 命令置于BEGIN和COMMIT命令之间来建立一个事务。因此，我们的银行事务实际上会是这样：

```
BEGIN;
UPDATE accounts SET balance = balance - 100.00
    WHERE name = 'Alice';
-- etc etc
COMMIT;
```

如果事务执行到一半时，我们决定不想提交了（例如刚刚发现Alice的余额变成了负数），就可以发出ROLLBACK而不是COMMIT，这样到目前为止的全部更新都会被取消。

PostgreSQL实际上把每条 SQL 语句都视为在一个事务中执行。如果你没有显式发出BEGIN，那么每条独立语句外围都会隐式包上一对BEGIN和（若成功）COMMIT。由BEGIN和COMMIT包围起来的一组语句，有时称为事务块。

> **注意**
>
> 某些客户端库会自动发出BEGIN和COMMIT命令，因此你可能在没有主动要求的情况下，也得到了事务块的效果。请查阅你所使用接口的文档。

也可以借助保存点，以更细粒度控制事务中的语句。保存点允许你有选择地丢弃事务的一部分，同时提交其余部分。使用SAVEPOINT定义保存点之后，如有需要，可以用ROLLBACK TO回滚到该保存点。事务中从定义保存点到回滚到它之间所做的数据库修改都会被丢弃，而早于该保存点的修改会被保留。

回滚到某个保存点之后，该保存点仍然保持定义状态，因此你可以多次回滚到它。反过来，如果你确定不再需要回滚到某个保存点，可以将其释放，以便系统回收一些资源。请记住，无论是释放某个保存点，还是回滚到某个保存点，都会自动释放在它之后定义的所有保存点。

所有这些都发生在事务块内部，因此其他数据库会话都看不到。等到你提交整个事务块时，被提交的动作才会作为一个整体对其他会话可见，而被回滚的动作则永远不会可见。

回到银行数据库，假设我们从Alice的账户扣除了100.00美元，并记入Bob的账户，后来却发现其实应该记入Wally的账户。我们可以像下面这样利用保存点来处理：

```
BEGIN;
UPDATE accounts SET balance = balance - 100.00
    WHERE name = 'Alice';
SAVEPOINT my_savepoint;
UPDATE accounts SET balance = balance + 100.00
    WHERE name = 'Bob';
-- oops ... forget that and use Wally's account
ROLLBACK TO my_savepoint;
UPDATE accounts SET balance = balance + 100.00
    WHERE name = 'Wally';
COMMIT;
```

当然，这个例子是过度简化的，但在事务块中借助保存点可以进行大量控制。此外，对于一个由于错误而被系统置为中止状态的事务块，要重新取得控制权，唯一的办法就是使用ROLLBACK TO；否则就只能把整个事务完全回滚并重新开始。

3.5. 窗口函数

窗口函数会在一组与当前行存在某种关联的表行上执行计算。
这与聚合函数能够完成的计算类型类似。但与普通聚合函数不同，
使用窗口函数不会把多行分组为一条输出行 – 各行仍然保留各自的独立身份。在幕后，窗口函数能够访问的不仅仅是查询结果中的当前行。

下面的例子说明如何将每个雇员的工资与其所在部门的平均工资进行比较：

```
SELECT depname, empno, salary, avg(salary) OVER (PARTITION BY depname)
 FROM empsalary;

  depname  | empno | salary |          avg
-----------+-------+--------+----------------------
 develop   |    11 |   5200 | 5020.0000000000000000
 develop   |     7 |   4200 | 5020.0000000000000000
 develop   |     9 |   4500 | 5020.0000000000000000
 develop   |     8 |   6000 | 5020.0000000000000000
 develop   |    10 |   5200 | 5020.0000000000000000
 personnel |     5 |   3500 | 3700.0000000000000000
 personnel |     2 |   3900 | 3700.0000000000000000
 sales     |     3 |   4800 | 4866.6666666666666667
 sales     |     1 |   5000 | 4866.6666666666666667
 sales     |     4 |   4800 | 4866.6666666666666667
(10 rows)
```

前三个输出列直接来自表empsalary，并且该表中的每一行对应一条输出行。第四列表示在所有depname值与当前行相同的表行上求得的平均值。（这其实与普通的avg
聚合函数是同一个函数，但OVER子句使它被当作窗口函数处理，并在一组适当的行上计算。）

窗口函数调用总是包含一个紧跟在窗口函数名和参数之后的OVER子句。
这正是它在语法上区别于普通函数或聚合函数的地方。OVER
子句精确决定查询中的哪些行会被拆分出来供窗口函数处理。OVER内部的PARTITION BY列表指定将具有相同PARTITION BY表达式值的行划分为组，也就是分区。对于每一行，窗口函数都是在与当前行处于同一分区的那些行上计算的。

你也可以用以下子句来控制窗口函数处理行的顺序：ORDER BY，将其放在OVER内部。（窗口中的ORDER BY甚至不必与行的输出顺序一致。）下面是一个例子：

```
SELECT depname, empno, salary, rank() OVER (PARTITION BY depname ORDER
 BY salary DESC) FROM empsalary;

  depname  | empno | salary | rank
-----------+-------+--------+------
 develop   |     8 |   6000 |    1
 develop   |    10 |   5200 |    2
 develop   |    11 |   5200 |    2
 develop   |     9 |   4500 |    4
 develop   |     7 |   4200 |    5
 personnel |     2 |   3900 |    1
 personnel |     5 |   3500 |    2
 sales     |     1 |   5000 |    1
 sales     |     4 |   4800 |    2
```

```
 sales     |     3 |   4800 |    2
(10 rows)
```

如上所示，rank函数会为当前行所在分区中每个不同的ORDER BY值，按ORDER BY子句定义的顺序产生一个数值排名。rank不需要显式参数，因为它的行为完全由OVER子句定义。

窗口函数所考虑的行，是查询的FROM子句产生并经WHERE、GROUP BY和HAVING子句（如果有）过滤后的那个“虚拟表”中的行。例如，由于不满足WHERE条件而被删除的行，不会被任何窗口函数看到。一个查询可以包含多个窗口函数，它们可以通过不同的OVER子句以不同方式划分数据，但它们都作用于这个虚拟表所定义的同一组行。

我们已经看到，如果行的顺序并不重要，就可以省略ORDER BY。PARTITION BY也可以省略，这时所有行构成一个分区。

与窗口函数相关的另一个重要概念是：对于每一行，在其所在分区内有一组行，称为它的窗口帧。许多（但不是全部）窗口函数只作用于窗口帧中的行，而不是整个分区中的所有行。默认情况下，如果提供了ORDER BY，则窗口帧包含从分区起始处直到当前行的所有行，再加上根据ORDER BY子句与当前行相等的所有后续行。如果省略ORDER BY，默认窗口帧则包含该分区中的所有行。[1]下面是一个使用sum的例子：

```
SELECT salary, sum(salary) OVER () FROM empsalary;

 salary |  sum
--------+-------
   5200 | 47100
   5000 | 47100
   3500 | 47100
   4800 | 47100
   3900 | 47100
   4200 | 47100
   4500 | 47100
   4800 | 47100
   6000 | 47100
   5200 | 47100
(10 rows)
```

上面由于OVER子句中没有ORDER BY，窗口帧与分区相同；而在没有PARTITION BY的情况下，这个分区就是整张表。换言之，每个求和都是在整张表上计算的，因此每条输出行得到的结果都相同。但是如果加上ORDER BY子句，结果就会大不一样：

```
SELECT salary, sum(salary) OVER (ORDER BY salary) FROM empsalary;

 salary |  sum
--------+-------
   3500 |  3500
   3900 |  7400
   4200 | 11600
   4500 | 16100
   4800 | 25700
   4800 | 25700
   5000 | 30700
   5200 | 41100
   5200 | 41100
```

[1] 还可以用其他方式定义窗口帧，但本教程不涉及这些选项。细节见第 4.2.8 节。

```
   6000 | 47100
(10 rows)
```

这里的求和是从第一条（最低）工资一直累加到当前行，
包括与当前行工资相同的所有重复值（注意那些重复工资对应的结果）。

窗口函数只允许出现在查询的SELECT列表和ORDER BY子句中。它们不允许出现在其他地方，例如GROUP BY、HAVING和WHERE子句中。这是因为从逻辑上讲，它们在这些子句处理完成之后才执行。另外，窗口函数在普通聚合函数之后执行。
这意味着在窗口函数的参数中包含聚合函数调用是合法的，反过来则不行。

如果需要在窗口计算执行之后再过滤或分组行，可以使用子查询。例如：

```
SELECT depname, empno, salary, enroll_date
FROM
  (SELECT depname, empno, salary, enroll_date,
     rank() OVER (PARTITION BY depname ORDER BY salary DESC, empno) AS
 pos
     FROM empsalary
  ) AS ss
WHERE pos < 3;
```

上面的查询只显示内部查询中rank小于3的那些行（也就是每个部门的前两行）。

当查询涉及多个窗口函数时，可以为每个函数分别写一个独立的OVER子句；
但如果多个函数都需要相同的窗口行为，这样写既重复，又容易出错。替代方法是，在WINDOW子句中给每一种窗口行为命名，然后在OVER中引用它。例如：

```
SELECT sum(salary) OVER w, avg(salary) OVER w
  FROM empsalary
  WINDOW w AS (PARTITION BY depname ORDER BY salary DESC);
```

关于窗口函数的更多细节可以在第 4.2.8 节、第 9.19 节、第 7.2.4 节以及SELECT参考页中找到。

3.6. 继承

继承是面向对象数据库中的一个概念。它为数据库设计打开了一些有趣的新可能性。

让我们创建两个表：表cities和表capitals。很自然，首都也是城市，
因此当你列出所有城市时，应该有某种办法能隐式地把首都也显示出来。如果你很有巧思，
也许会想出这样的方案：

```
CREATE TABLE capitals (
  name       text,
  population real,
  altitude  int,    -- (in ft)
  state      char(2)
);

CREATE TABLE non_capitals (
  name       text,
  population real,
  altitude  int     -- (in ft)
);
```

```
CREATE VIEW cities AS
  SELECT name, population, altitude FROM capitals
    UNION
  SELECT name, population, altitude FROM non_capitals;
```

就查询而言，这样做还行，但一旦需要更新多行，它就会变得很麻烦。

更好的解决办法是：

```
CREATE TABLE cities (
  name       text,
  population real,
  altitude  int     -- (in ft)
);

CREATE TABLE capitals (
  state      char(2)
) INHERITS (cities);
```

在这种情况下，capitals的一行会从它的父表cities继承全部列（name、population和altitude）。列name的类型是text，这是PostgreSQL内置的一种变长字符串类型。州首府有一个额外的列state，用来表示所在的州。在PostgreSQL中，一个表可以从零个或多个其他表继承。

例如，下面的查询会找出所有海拔超过500英尺的城市名称，其中也包括州首府：

```
SELECT name, altitude
  FROM cities
  WHERE altitude > 500;
```

返回：

```
   name    | altitude
-----------+-----------
 Las Vegas |      2174
 Mariposa  |      1953
 Madison   |       845
(3 rows)
```

另一方面，下面的查询会找出所有海拔500英尺或更高且不是州首府的城市：

```
SELECT name, altitude
    FROM ONLY cities
    WHERE altitude > 500;

   name    | altitude
-----------+-----------
 Las Vegas |      2174
 Mariposa  |      1953
(2 rows)
```

这里，放在cities前面的ONLY表示该查询只针对cities表执行，而不包括继承层次中位于cities之下的表。我们前面已经讨论过的许多命令 – SELECT、UPDATE和DELETE – 都支持这种ONLY记法。

注意

尽管继承经常很有用，但它尚未与唯一约束或外键集成，这限制了它的实用性。
更多细节见第 5.8 节。

3.7. 小结

PostgreSQL还有许多特性在这个面向SQL新用户的入门教程中没有涉及。
本书余下部分会更详细地讨论这些特性。

如果你觉得还需要更多入门材料，请访问 PostgreSQL 官方网站[2]，其中有指向更多资源的链接。

[2] http://www.postgresql.org

部分 II. SQL 语言

本部分描述SQL语言在PostgreSQL中的用法。我们首先介绍SQL的一般语法，然后说明如何创建表、如何填充数据库以及如何查询数据库。中间部分列出可在SQL命令中使用的数据类型和函数。最后，我们讨论数据库调优中的若干重要方面。

本部分的内容安排方式使初学者能够从头到尾依次阅读，并在不必过多向后查阅的情况下充分理解这些主题。各章力求自成体系，因此高级用户也可以按需单独阅读。相关内容按主题单元以叙述方式展开。希望获得某个命令完整描述的读者，建议查阅第 VI 部分。

读者应当知道如何连接到PostgreSQL数据库并发出SQL命令。不熟悉这些内容的读者，建议先阅读第 I 部分。SQL命令通常通过PostgreSQL的交互式终端程序psql输入，但也可以使用具有类似功能的其他程序。

目录

4. SQL语法 31
4.1. 词法结构 31
4.1.1. 标识符和关键词 31
4.1.2. 常量 32
4.1.3. 操作符 37
4.1.4. 特殊字符 37
4.1.5. 注释 37
4.1.6. 操作符优先级 38
4.2. 值表达式 39
4.2.1. 列引用 40
4.2.2. 位置参数 40
4.2.3. 下标 40
4.2.4. 字段选择 41
4.2.5. 操作符调用 41
4.2.6. 函数调用 41
4.2.7. 聚合表达式 42
4.2.8. 窗口函数调用 43
4.2.9. 类型转换 44
4.2.10. 排序规则表达式 45
4.2.11. 标量子查询 46
4.2.12. 数组构造器 46
4.2.13. 行构造器 47
4.2.14. 表达式计算规则 49
4.3. 调用函数 50
4.3.1. 使用位置记法 50
4.3.2. 使用命名记法 51
4.3.3. 使用混合记法 51
5. 数据定义 52
5.1. 表基础 52
5.2. 默认值 53
5.3. 约束 54
5.3.1. 检查约束 54
5.3.2. 非空约束 55
5.3.3. 唯一约束 56
5.3.4. 主键 57
5.3.5. 外键 58
5.3.6. 排他约束 60
5.4. 系统列 60
5.5. 修改表 61
5.5.1. 增加列 62
5.5.2. 移除列 62
5.5.3. 增加约束 62
5.5.4. 移除约束 63
5.5.5. 更改列的默认值 63
5.5.6. 修改列的数据类型 63
5.5.7. 重命名列 64
5.5.8. 重命名表 64
5.6. 权限 64
5.7. 模式 65
5.7.1. 创建模式 65
5.7.2. 公共模式 66

5.7.3. 模式搜索路径 ... 66
5.7.4. 模式和权限 ... 67
5.7.5. 系统目录模式 ... 68
5.7.6. 使用方式 ... 68
5.7.7. 可移植性 ... 68
5.8. 继承 ... 68
5.8.1. 注意事项 ... 71
5.9. 分区 ... 71
5.9.1. 概述 ... 72
5.9.2. 实现分区 ... 72
5.9.3. 分区管理 ... 75
5.9.4. 分区和约束排除 ... 76
5.9.5. 备选分区方法 ... 77
5.9.6. 注意事项 ... 78
5.10. 外部数据 ... 78
5.11. 其他数据库对象 ... 79
5.12. 依赖跟踪 ... 79
6. 数据操纵 ... 81
6.1. 插入数据 ... 81
6.2. 更新数据 ... 82
6.3. 删除数据 ... 82
7. 查询 ... 84
7.1. 概述 ... 84
7.2. 表表达式 ... 84
7.2.1. FROM子句 ... 84
7.2.2. WHERE子句 ... 91
7.2.3. GROUP BY和HAVING子句 ... 92
7.2.4. 窗口函数处理 ... 94
7.3. 选择列表 ... 94
7.3.1. 选择列表项 ... 94
7.3.2. 列标签 ... 95
7.3.3. DISTINCT ... 95
7.4. 组合查询 ... 96
7.5. 行排序 ... 96
7.6. LIMIT和OFFSET ... 97
7.7. VALUES列表 ... 98
7.8. WITH查询（公共表表达式） ... 99
7.8.1. WITH中的SELECT ... 99
7.8.2. WITH中的数据修改语句 ... 102
8. 数据类型 ... 104
8.1. 数字类型 ... 105
8.1.1. 整数类型 ... 106
8.1.2. 任意精度数值 ... 106
8.1.3. 浮点类型 ... 107
8.1.4. serial 类型 ... 108
8.2. 货币类型 ... 109
8.3. 字符类型 ... 110
8.4. 二进制数据类型 ... 111
8.4.1. bytea的十六进制格式 ... 112
8.4.2. bytea的转义格式 ... 112
8.5. 日期/时间类型 ... 113
8.5.1. 日期/时间输入 ... 115
8.5.2. 日期/时间输出 ... 118

8.5.3. 时区 118
8.5.4. 间隔输入 120
8.5.5. 间隔输出 121
8.5.6. 内部实现 122
8.6. 布尔类型 122
8.7. 枚举类型 123
8.7.1. 枚举类型的声明 123
8.7.2. 排序 124
8.7.3. 类型安全性 124
8.7.4. 实现细节 125
8.8. 几何类型 125
8.8.1. 点 125
8.8.2. 线段 125
8.8.3. 方框 126
8.8.4. 路径 126
8.8.5. 多边形 126
8.8.6. 圆 127
8.9. 网络地址类型 127
8.9.1. `inet` 127
8.9.2. `cidr` 127
8.9.3. `inet` 与 `cidr` 128
8.9.4. `macaddr` 128
8.10. 位串类型 129
8.11. 文本搜索类型 129
8.11.1. `tsvector` 130
8.11.2. `tsquery` 131
8.12. UUID类型 132
8.13. XML类型 132
8.13.1. 创建XML值 133
8.13.2. 编码处理 133
8.13.3. 访问XML值 134
8.14. 数组 134
8.14.1. 数组类型的声明 134
8.14.2. 数组值输入 135
8.14.3. 访问数组 136
8.14.4. 修改数组 138
8.14.5. 在数组中搜索 140
8.14.6. 数组输入和输出语法 141
8.15. 复合类型 142
8.15.1. 复合类型的声明 142
8.15.2. Composite Value Input 143
8.15.3. 访问复合类型 144
8.15.4. 修改组合值 144
8.15.5. 复合类型的输入和输出语法 145
8.16. 对象标识符类型 145
8.17. 伪类型 147
9. 函数和操作符 149
9.1. 逻辑操作符 149
9.2. 比较操作符 149
9.3. 数学函数和操作符 151
9.4. 字符串函数和操作符 156
9.5. 二进制串函数和操作符 168
9.6. 位串函数和操作符 171

9.7. 模式匹配 172
9.7.1. LIKE 172
9.7.2. SIMILAR TO正则表达式 173
9.7.3. POSIX正则表达式 174
9.8. 数据类型格式化函数 185
9.9. 日期/时间函数和操作符 191
9.9.1. EXTRACT, date_part 196
9.9.2. date_trunc 200
9.9.3. AT TIME ZONE 201
9.9.4. 当前日期/时间 201
9.9.5. 延时执行 203
9.10. 枚举支持函数 203
9.11. 几何函数和操作符 204
9.12. 网络地址函数和操作符 207
9.13. 文本搜索函数和操作符 210
9.14. XML 函数 215
9.14.1. 产生 XML 内容 215
9.14.2. XML 谓词 219
9.14.3. 处理 XML 221
9.14.4. 将表映射到 XML 222
9.15. 序列操作函数 225
9.16. 条件表达式 227
9.16.1. CASE 228
9.16.2. COALESCE 229
9.16.3. NULLIF 229
9.16.4. GREATEST和LEAST 230
9.17. 数组函数和操作符 230
9.18. 聚合函数 233
9.19. 窗口函数 236
9.20. 子查询表达式 237
9.20.1. EXISTS 237
9.20.2. IN 238
9.20.3. NOT IN 238
9.20.4. ANY/SOME 238
9.20.5. ALL 239
9.20.6. 单一行比较 239
9.21. 行和数组比较 240
9.21.1. IN 240
9.21.2. NOT IN 240
9.21.3. ANY/SOME（数组） 240
9.21.4. ALL（数组） 241
9.21.5. 按行比较 241
9.22. 集合返回函数 242
9.23. 系统信息函数 244
9.24. 系统管理函数 253
9.25. 触发器函数 260
10. 类型转换 262
10.1. 概述 262
10.2. 操作符 263
10.3. 函数 266
10.4. 值存储 269
10.5. UNION、CASE及相关结构 269
11. 索引 272

11.1. 简介 272
11.2. 索引类型 273
11.3. 多列索引 274
11.4. 索引和ORDER BY 275
11.5. 组合多个索引 275
11.6. 唯一索引 276
11.7. 表达式索引 277
11.8. 部分索引 277
11.9. 操作符类和操作符族 279
11.10. 索引和排序规则 280
11.11. 检查索引使用情况 281
12. 全文搜索 283
12.1. 介绍 283
12.1.1. 什么是文档? 283
12.1.2. 基本文本匹配 284
12.1.3. 配置 285
12.2. 表和索引 286
12.2.1. 搜索表 286
12.2.2. 创建索引 286
12.3. 控制文本搜索 288
12.3.1. 解析文档 288
12.3.2. 解析查询 289
12.3.3. 搜索结果排名 290
12.3.4. 高亮结果 292
12.4. 附加特性 293
12.4.1. 操纵文档 293
12.4.2. 操纵查询 294
12.4.3. 用于自动更新的触发器 296
12.4.4. 收集文档统计数据 297
12.5. 解析器 298
12.6. 词典 300
12.6.1. 停用词 301
12.6.2. 简单词典 301
12.6.3. 同义词词典 302
12.6.4. 分类词典 304
12.6.5. Ispell 词典 306
12.6.6. Snowball 词典 307
12.7. 配置示例 308
12.8. 测试和调试文本搜索 309
12.8.1. 配置测试 309
12.8.2. 解析器测试 311
12.8.3. 词典测试 313
12.9. GiST 和 GIN 索引类型 313
12.10. psql 支持 314
12.11. 限制 317
12.12. 从 8.3 之前的文本搜索迁移 318
13. 并发控制 319
13.1. 介绍 319
13.2. 事务隔离 319
13.2.1. 读已提交隔离级别 320
13.2.2. 可重复读隔离级别 321
13.2.3. 可串行化隔离级别 322
13.3. 显式锁定 323

13.3.1. 表级锁 324
13.3.2. 行级锁 326
13.3.3. 死锁 326
13.3.4. 咨询锁 327
13.4. 应用级别的数据一致性检查 328
13.4.1. 用可串行化事务强制一致性 328
13.4.2. 使用显式阻塞锁强制一致性 328
13.5. 注意事项 329
13.6. 锁定和索引 329
14. 性能提示 331
14.1. 使用EXPLAIN 331
14.2. 规划器使用的统计信息 336
14.3. 用显式JOIN子句控制规划器 337
14.4. 填充一个数据库 339
14.4.1. 禁用自动提交 339
14.4.2. 使用COPY 339
14.4.3. 移除索引 340
14.4.4. 移除外键约束 340
14.4.5. 增加maintenance_work_mem 340
14.4.6. 增加checkpoint_segments 340
14.4.7. 禁用 WAL 归档和流复制 340
14.4.8. 事后运行ANALYZE 341
14.4.9. 关于pg_dump的一些注记 341
14.5. 非持久设置 341

第 4 章 SQL语法

本章描述 SQL 的语法。它是理解后续各章的基础，后续各章将详细介绍如何使用 SQL 命令定义和修改数据。

我们也建议已经熟悉 SQL 的用户仔细阅读本章，因为其中包含一些在各 SQL 数据库之间实现并不一致的规则和概念，以及一些 PostgreSQL 特有的规则和概念。

4.1. 词法结构

SQL 输入由一系列命令组成。一个命令由一系列词元组成，并以分号（“;”）终止。输入流的结束也会终止一个命令。哪些词元是合法的，取决于具体命令的语法。

一个词元可以是关键词、标识符、带引号的标识符、字面量（或常量），或者特殊字符符号。
词元通常由空白（空格、制表符、换行）分隔，但如果不存在歧义，
则不必如此（一般只有特殊字符与其他类型的词元相邻时才会这样）。

例如，下面是一个（语法上）合法的SQL输入：

```
SELECT * FROM MY_TABLE;
UPDATE MY_TABLE SET A = 5;
INSERT INTO MY_TABLE VALUES (3, 'hi there');
```

这是一个由三个命令组成的序列，每行一个命令（虽然这并非必须；一行中可以包含多个命令，
命令也可以有意义地跨多行书写）。

此外，注释也可以出现在 SQL 输入中。它们不是词元，实际上等价于空白。

在哪些词元标识命令、哪些词元是操作数或参数这一点上，SQL 语法并不十分一致。最前面的几个词元通常是命令名，因此在上面的示例中，我们通常会说是“SELECT”、“UPDATE”
和“INSERT”命令。但是例如 `UPDATE` 命令总是要求在特定位置出现一个 `SET` 词元，而这种形式的 `INSERT` 也要求有一个 `VALUES` 才算完整。每个命令的精确语法规则见第 VI 部分。

4.1.1. 标识符和关键词

上例中的 `SELECT`、`UPDATE` 或 `VALUES` 这类词元都是关键词，也就是在 SQL 语言中具有固定含义的词。词元 `MY_TABLE` 和 `A` 则是标识符的例子。它们标识表、列或其他数据库对象的名称，这取决于它们所在的命令。因此，它们有时也简称为“名称”。
关键词和标识符具有相同的词法结构，这意味着如果不了解该语言，
就无法判断一个词元究竟是标识符还是关键词。完整的关键词列表可以在附录 C中找到。

SQL 标识符和关键词必须以字母（`a-z`，也包括带变音符的字母和非拉丁字母）或下划线（`_`）开头。标识符或关键词中的后续字符可以是字母、下划线、数字（`0-9`）或美元符号（`$`）。注意，按照 SQL 标准的字面规定，标识符中不允许使用美元符号，因此它可能会降低应用的可移植性。SQL 标准不会定义包含数字、以下划线开头或以下划线结尾的关键词，因此这种形式的标识符可以安全地避免与未来标准扩展发生冲突。

系统对标识符最多只使用 `NAMEDATALEN`-1 个字节；在命令中可以写更长的名称，但它们会被截断。默认情况下，`NAMEDATALEN` 为 64，因此标识符的最大长度是 63 字节。如果这个限制有问题，可以通过修改 `src/include/pg_config_manual.h` 中的 `NAMEDATALEN` 常量来提高。

关键词和未加引号的标识符不区分大小写。因此：

```
UPDATE MY_TABLE SET A = 5;
```

也可以等价地写成：

```
uPDaTE my_TabLE SeT a = 5;
```

一种常见约定是关键词使用大写、名称使用小写，例如：

```
UPDATE my_table SET a = 5;
```

还有第二类标识符：定界标识符或带引号的标识符。它通过把任意字符序列括在双引号中形成（"）。定界标识符始终是标识符，绝不会是关键词。因此，`"select"` 可以用来引用名为“select”的列或表，而不加引号的 `select` 会被视为关键词，因此在期望表名或列名的位置使用时会导致解析错误。这个示例可以用带引号的标识符写成：

```
UPDATE "my_table" SET "a" = 5;
```

带引号的标识符可以包含除编码为零的字符之外的任意字符。（如果要包含双引号，就写两个双引号。）这使得可以构造原本不可能的表名或列名，例如包含空格或和号的名称。长度限制仍然适用。

带引号标识符的一种变体允许包含按代码点标识的转义 Unicode 字符。这种变体在开头双引号之前紧接着写 `U&`（大写或小写的字母 U 后接和号），中间没有空格，例如 `U&"foo"`。（注意这会与操作符 `&` 产生歧义。请在该操作符两侧加空格以避免这个问题。）在引号内，可以通过写一个反斜线后跟 4 位十六进制代码点编号，或者一个反斜线后跟加号再跟 6 位十六进制代码点编号，来指定 Unicode 字符。例如，标识符 `"data"` 可以写成：

```
U&"d\0061t\+000061"
```

下面这个稍微复杂一些的示例，用西里尔字母写出了俄语单词“slon”（大象）：

```
U&"\0441\043B\043E\043D"
```

如果想使用反斜线以外的转义字符，可以在字符串后用 `UESCAPE` 子句指定，例如：

```
U&"d!0061t!+000061" UESCAPE '!'
```

转义字符可以是除十六进制数字、加号、单引号、双引号或空白字符之外的任何单个字符。注意，转义字符要写在单引号内，而不是双引号内。

如果要在标识符中按字面包含转义字符本身，把它写两次即可。

Unicode 转义语法只有在服务器编码为 `UTF8` 时才有效。使用其他服务器编码时，只能指定 ASCII 范围内的代码点（不超过 `\007F`）。4 位或 6 位转义形式都可以用来指定 UTF-16 代理对，以组合出代码点大于 U+FFFF 的字符，尽管从技术上说，由于有 6 位形式，这样做并无必要。（代理对不会被直接存储，而是会合并成一个代码点，然后以 UTF-8 编码。）

给标识符加上引号也会使其区分大小写，而不加引号的名称总是会被折叠成小写。例如，标识符 `FOO`、`foo` 和 `"foo"` 在 PostgreSQL 中被视为相同，但 `"Foo"` 和 `"FOO"` 与这三个都不同，彼此之间也不同。（PostgreSQL 将不加引号的名称折叠为小写，这与 SQL 标准不兼容；标准规定，不加引号的名称应折叠为大写。因此，按照标准，`foo` 应当等价于 `"FOO"`，而不是 `"foo"`。如果你想编写可移植的应用，建议对某个特定名称要么始终加引号，要么始终不加引号。）

4.1.2. 常量

在PostgreSQL中有三种隐式类型常量：字符串、位串和数字。也可以为常量指定显式类型，这样系统就能更准确地表示并更高效地处理它们。这些方式将在后续小节中讨论。

4.1.2.1. 字符串常量

SQL 中的字符串常量是由单引号（'）括起的任意字符序列，例如 `'This is a string'`。要在字符串常量中包含单引号字符，请写两个相邻的单引号，例如 `'Dianne''s horse'`。注意，这不同于双引号字符（"）。

两个只由空白及至少一个新行分隔的字符串常量会被连接在一起，
并且将作为一个写在一起的字符串常量来对待。例如：

```
SELECT 'foo'
'bar';
```

等同于：

```
SELECT 'foobar';
```

但是：

```
SELECT 'foo'      'bar';
```

则不是合法的语法（这种有些奇怪的行为是SQL指定的，PostgreSQL遵循了该标准）。

4.1.2.2. C风格转义的字符串常量

PostgreSQL也接受“转义”字符串常量，这也是SQL标准的一个扩展。
一个转义字符串常量可以通过在开单引号前面写一个字母`E`（大写或小写形式）来指定，例如 `E'foo'`（当一个转义字符串常量跨行时，只在第一个开引号之前写`E`）。
在一个转义字符串内部，一个反斜线字符（`\`）会开始一个 C 风格的反斜线转义序列，在其中反斜线和后续字符的组合表示一个特殊的字节值（如表 4.1中所示）。

表 4.1. 反斜线转义序列

反斜线转义序列	解释
`\b`	退格
`\f`	换页
`\n`	换行
`\r`	回车
`\t`	制表符
`\o`, `\oo`, `\ooo` (*o* = 0 - 7)	八进制字节值
`\xh`, `\xhh` (*h* = 0 - 9, A - F)	十六进制字节值
`\uxxxx`, `\Uxxxxxxxx` (*x* = 0 - 9, A - F)	16 或 32 位十六进制 Unicode 字符值

跟随在一个反斜线后面的任何其他字符被当做其字面意思。因此，要包括一个反斜线字符，
请写两个反斜线（`\\`）。在一个转义字符串中包括一个单引号除了普通方法`''`之外，还可以写成`\'`。

你有责任确保所创建的字节序列能够构成服务器字符集编码中的有效字符，
尤其是在使用八进制或十六进制转义时。当服务器编码为 UTF-8 时，应改用 Unicode 转义或者第 4.1.2.3 节中说明的另一种 Unicode 转义语法。（另一种做法是手工进行 UTF-8 编码并写出各字节，但那会非常繁琐。）

Unicode 转义语法只有在服务器编码为 `UTF8` 时才能完全有效。使用其他服务器编码时，只能指定 ASCII 范围内的代码点（不超过 `\u007F`）。4 位和 8 位形式都可以用来指定 UTF-16 代理对，

以组合出代码点大于 U+FFFF 的字符，尽管从技术上说，由于有 8 位形式，这样做并无必要。（在服务器编码为 `UTF8` 时使用代理对，它们会先合并成一个代码点，然后以 UTF-8 编码。）

> **小心**
>
> 如果配置参数standard_conforming_strings为`off`，那么PostgreSQL对常规字符串常量和转义字符串常量中的反斜线转义都识别。不过，从PostgreSQL 9.1 开始，该参数的默认值为`on`，意味着只在转义字符串常量中识别反斜线转义。这种行为更兼容标准，但是可能使依赖历史行为（始终识别反斜线转义）的应用无法正常工作。作为一种变通，你可以设置该参数为`off`，但最好修改应用，不再使用反斜线转义。如果你需要使用一个反斜线转义来表示一个特殊字符，为该字符串常量写上一个`E`。
>
> 在`standard_conforming_strings`之外，配置参数escape_string_warning和backslash_quote也决定了如何对待字符串常量中的反斜线。

编码为零的字符不能出现在字符串常量中。

4.1.2.3. 带有 Unicode 转义的字符串常量

PostgreSQL 还支持另一种字符串转义语法，它允许用代码点指定任意 Unicode 字符。Unicode 转义字符串常量在开引号前紧接着写 `U&`（大写或小写字母 U 后跟和号），中间没有任何空白，例如 `U&'foo'`。（注意这会与操作符 `&` 产生歧义。请在该操作符两侧加空格以避免这个问题。）在引号内，可以通过写一个反斜线后跟 4 位十六进制代码点编号，或者一个反斜线后跟加号再跟 6 位十六进制代码点编号，来指定 Unicode 字符。例如，字符串 `'data'` 可以写成

```
U&'d\0061t\+000061'
```

下面这个稍微复杂一些的示例，用西里尔字母写出了俄语单词“slon”（大象）：

```
U&'\0441\043B\043E\043D'
```

如果想要一个不是反斜线的转义字符，可以在字符串之后使用`UESCAPE`子句来指定，例如：

```
U&'d!0061t!+000061' UESCAPE '!'
```

转义字符可以是除十六进制数字、加号、单引号、双引号或空白字符之外的任意单个字符。

Unicode 转义语法只有在服务器编码为 `UTF8` 时才有效。使用其他服务器编码时，只能指定 ASCII 范围内的代码点（不超过 `\007F`）。4 位和 6 位形式都可以用来指定 UTF-16 代理对，以组合出代码点大于 U+FFFF 的字符，尽管从技术上说，由于有 6 位形式，这样做并无必要。（在服务器编码为 `UTF8` 时使用代理对，它们会先合并成一个代码点，然后以 UTF-8 编码。）

此外，字符串常量的 Unicode 转义语法只有在配置参数standard_conforming_strings开启时才有效。这是因为否则这种语法可能会混淆解析 SQL 语句的客户端，从而导致 SQL 注入以及类似的安全问题。如果该参数被设置为 off，这种语法将被拒绝并报错。

如果要在字符串中按字面包含转义字符本身，把它写两次即可。

4.1.2.4. 美元引用的字符串常量

指定字符串常量的标准语法通常很方便，但当所需字符串包含很多单引号或反斜线时，就可能难以理解，因为每个这样的字符都必须写成双份。为了在这种情况下让查询更易读，

PostgreSQL提供了另一种编写字符串常量的方法，称为“美元引用”。
美元引用的字符串常量由一个美元符号（`$`）、一个由零个或多个字符组成的可选“标签”、
另一个美元符号、构成字符串内容的任意字符序列、一个美元符号、
与该美元引用开头相同的标签，以及一个美元符号组成。例如，对于字符串“Dianne's horse”，
下面是使用美元引用指定它的两种不同方法：

```
$$Dianne's horse$$
$SomeTag$Dianne's horse$SomeTag$
```

注意，在美元引用的字符串中，单引号不需要转义就可以使用。事实上，
美元引用字符串内的任何字符都不会被转义：字符串内容始终按字面书写。
反斜线没有特殊含义，美元符号也没有，除非它是与开头标签相匹配的序列的一部分。

可以通过在每个嵌套层级选择不同的标签，来嵌套美元引用字符串常量。
这种做法最常见于编写函数定义时。例如：

```
$function$
BEGIN
    RETURN ($1 ~ $q$[\t\r\n\v\\]$q$);
END;
$function$
```

这里，序列 `$q$[\t\r\n\v\\]$q$` 表示一个美元引用的字面字符串 `[\t\r\n\v\\]`，当函数体被 PostgreSQL 执行时，它会被识别。但由于该序列与外层美元引用定界符 `$function$` 不匹配，就外层字符串而言，它只不过是常量中的一部分字符。

美元引用字符串的标签（如果有）遵循与未加引号标识符相同的规则，但不能包含美元符号。
标签区分大小写，因此 `$tag$String content$tag$` 是正确的，而 `$TAG$String content $tag$` 则不正确。

如果美元引用字符串位于某个关键词或标识符之后，必须用空白把它们分隔开，
否则美元引用定界符会被当作前一个标识符的一部分。

美元引用不是 SQL 标准的一部分，但在书写复杂字符串字面量时，它往往比符合标准的单引号语法更方便。当要表示的字符串常量位于其他常量内部时，它尤其有用，
这种情况在过程函数定义中很常见。如果使用单引号语法，
上一个示例中的每个反斜线都必须写成四个反斜线；
在解析原始字符串常量时它们会缩减为两个反斜线，
而在函数执行期间重新解析内层字符串常量时又会变成一个。

4.1.2.5. 位串常量

位串常量看起来像常规字符串常量在开引号之前（中间无空白）加了一个`B`（大写或小写形式），
例如`B'1001'`。位串常量中允许的字符只有`0`和`1`。

或者，也可以用十六进制记法指定位串常量，即在开头写一个 `X`（大写或小写形式），例如
`X'1FF'`。这种记法等价于把每个十六进制位替换成四个二进制位得到的位串常量。

两种形式的位串常量可以以常规字符串常量相同的方式跨行继续。
美元引用不能被用在位串常量中。

4.1.2.6. 数字常量

数字常量接受下列一般形式：

digits

```
digits.[digits][e[+-]digits]
[digits].digits[e[+-]digits]
digitse[+-]digits
```

其中 *digits* 是一个或多个十进制数字（0 到 9）。若使用小数点，则小数点前后至少有一侧必须有数字。若使用指数标记（e），其后至少必须有一位数字。常量中不能嵌入空格或其他字符。请注意，前导正号或负号实际上不属于常量本身，而是作用于常量的操作符。

这些是合法数字常量的示例：

```
42
3.5
4.
.001
5e2
1.925e-3
```

如果一个不包含小数点和指数的数字常量的值适合类型integer（32 位），它首先被假定为类型integer。否则如果它的值适合类型bigint（64 位），它被假定为类型bigint。再否则它会被取做类型numeric。包含小数点和/或指数的常量总是首先被假定为类型numeric。

为数字常量最初指定的数据类型，只是类型解析算法的起点。在大多数情况下，常量会根据上下文被自动强制转换成最合适的类型。必要时，你可以通过类型转换强制把一个数值解释为特定的数据类型。例如，可以通过下面的写法强制把一个数值当作类型 real（float4）处理：

```
REAL '1.23'  -- 字符串形式
1.23::REAL   -- PostgreSQL（历史）形式
```

这些实际上只是下面将要讨论的一般类型转换记法的特例。

4.1.2.7. 其他类型的常量

任意类型的常量都可以用下列记法之一输入：

```
type 'string'
'string'::type
CAST ( 'string' AS type )
```

字符串常量的文本会被传递给名为 *type* 的类型的输入转换例程。其结果是一个指定类型的常量。如果该常量必须具有的类型不存在歧义（例如，它被直接赋给某个表列时），就可以省略显式类型转换，在这种情况下它会被自动强制转换。

字符串常量可以使用常规 SQL 记法或美元引用书写。

也可以用类似函数调用的语法来指定类型强制转换：

```
typename ( 'string' )
```

但并非所有类型名都可以这样使用，详见第 4.2.9 节。

如第 4.2.9 节中所述，::、CAST() 以及函数调用语法也可以用来指定任意表达式的运行时类型转换。为避免语法歧义，*type* '*string*' 语法只能用来指定简单字面常量的类型。*type* '*string*' 语法的另一个限制是不能用于数组类型；要指定数组常量的类型，请使用 :: 或 CAST( )。

`CAST()`语法符合 SQL。*`type 'string'`*语法是该标准的一般化：SQL 指定这种语法只用于一些数据类型，但是PostgreSQL允许它用于所有类型。带有`::`的语法是PostgreSQL的历史用法，就像函数调用语法一样。

4.1.3. 操作符

操作符名称由下列列表中的字符组成，长度最多为 `NAMEDATALEN`-1 个字符（默认是 63）：

+ - * / < > = ~ ! @ # % ^ & | ` ?

不过，操作符名称也有一些限制：

- `--` 和 `/*` 不能出现在操作符名称中的任何位置，因为它们会被视为注释的开始。
- 多字符操作符名称不能以 `+` 或 `-` 结尾，除非该名称中还至少包含下列字符中的一个：

  ~ ! @ # % ^ & | ` ?

  例如，`@-` 是合法的操作符名称，但 `*-` 不是。这个限制使 PostgreSQL 能够在不要求词元之间必须有空格的情况下解析符合 SQL 的查询。

当使用非 SQL 标准的操作符名时，你通常需要用空格分隔相邻的操作符来避免歧义。例如，如果你定义了一个名为`@`的左一元操作符，你不能写`X*@Y`，你必须写`X* @Y`来确保PostgreSQL把它读作两个操作符名而不是一个。

4.1.4. 特殊字符

某些非字母数字字符具有不同于操作符的特殊含义。关于它们的详细用法，可以在描述相应语法元素的位置找到。本节只是为了提示它们的存在，并概述这些字符的用途。

- 美元符号（`$`）后跟数字时，用来表示函数定义体或预备语句中的位置参数。在其他上下文中，美元符号可以是标识符的一部分，也可以是美元引用字符串常量的一部分。
- 圆括号（`()`）具有它们通常的含义，用来对表达式分组并确定运算优先级。在某些情况中，圆括号被要求作为一个特定 SQL 命令的固定语法的一部分。
- 方括号（`[]`）被用来选择一个数组中的元素。更多关于数组的信息见第 8.14 节。
- 逗号（`,`）被用在某些语法结构中来分割一个列表的元素。
- 分号（`;`）结束一个 SQL 命令。它不能出现在一个命令中间的任何位置，除了在一个字符串常量中或者一个带引号的标识符中。
- 冒号（`:`）被用来从数组中选择“切片”（见第 8.14 节）。在某些 SQL 的“方言”（例如嵌入式 SQL）中，冒号被用来作为变量名的前缀。
- 星号（`*`）在某些上下文中用来表示表行或复合值的所有字段。当它被用作聚合函数的参数时，还有一种特殊含义，即该聚合不需要任何显式参数。
- 句点（`.`）被用在数字常量中，并且被用来分割模式、表和列名。

4.1.5. 注释

注释是一串以双连字符开始并延伸到行尾的字符，例如：

```
-- 这是一条标准 SQL 注释
```

另外，也可以使用 C 风格注释块：

```
/* 多行注释
 * 包含嵌套：/* 嵌套块注释 */
 */
```

这里该注释开始于/*并且延伸到匹配出现的*/。这些注释块可按照 SQL 标准中指定的方式嵌套，但和 C 中不同。这样我们可以注释掉一大段可能包含注释块的代码。

在进一步进行语法分析之前，注释会从输入流中移除，并实际被替换为空白。

4.1.6. 操作符优先级

表 4.2展示了 PostgreSQL 中操作符的优先级和结合性。大多数操作符具有相同的优先级，并且是左结合的。操作符的优先级和结合性是固定写在解析器中的。这可能导致不直观的行为；例如，布尔操作符 < 和 > 与布尔操作符 <= 和 >= 的优先级不同。此外，在组合使用二元和一元操作符时，有时需要添加圆括号。例如：

```
SELECT 5 ! - 6;
```

将被解析为：

```
SELECT 5 ! (- 6);
```

因为解析器并不知道 - 等到知道时已经太迟了 - !被定义为后缀操作符，而不是中缀操作符。要在这种情况下得到所需的行为，必须写成：

```
SELECT (5 !) - 6;
```

这是为可扩展性付出的代价。

表 4.2. 操作符优先级（递减）

操作符/元素	结合性	描述
.	左	表/列名分隔符
::	左	PostgreSQL-风格的类型转换
[ ]	左	数组元素选择
+ -	右	一元正号、一元负号
^	左	求幂
* / %	左	乘、除、取模
+ -	左	加、减
IS		IS TRUE、IS FALSE、IS NULL等
ISNULL		空值测试
NOTNULL		非空测试
（任意其他）	左	所有其他操作符
IN		集合成员关系
BETWEEN		范围包含
OVERLAPS		时间区间重叠

操作符/元素	结合性	描述
LIKE ILIKE SIMILAR		字符串模式匹配
< >		小于、大于
=	右	相等、赋值
NOT	右	逻辑否定
AND	左	逻辑合取
OR	左	逻辑析取

注意，这些操作符优先级规则也适用于名称与上述内置操作符相同的用户定义操作符。例如，如果你为某种自定义数据类型定义了一个“+”操作符，那么无论你的操作符做什么，它都将具有与内置“+”操作符相同的优先级。

当一个模式限定的操作符名被用在OPERATOR语法中时，如下面的示例：

```
SELECT 3 OPERATOR(pg_catalog.+) 4;
```

OPERATOR 结构在优先级上会被视为表 4.2中为“任意其他”操作符所示的优先级。无论OPERATOR() 中指定的是哪个具体操作符，这一点都成立。

4.2. 值表达式

值表达式被用于各种各样的环境中，例如在SELECT命令的目标列表中，作为INSERT或UPDATE中的新列值，或者作为若干命令中的搜索条件。为了与表表达式（其结果是一张表）区分开来，值表达式的结果有时被称为标量。因此，值表达式也称为标量表达式（甚至简称为表达式）。表达式语法允许使用算术、逻辑、集合以及其他操作，从基本部分计算出值。

一个值表达式是下列之一：

- 一个常量或字面值
- 一个列引用
- 在一个函数定义体或预备语句中的一个位置参数引用
- 一个下标表达式
- 一个字段选择表达式
- 一个操作符调用
- 一个函数调用
- 一个聚合表达式
- 一个窗口函数调用
- 一个类型转换
- 一个排序规则表达式
- 一个标量子查询
- 一个数组构造器

- 一个行构造器
- 另一个圆括号中的值表达式（用于对子表达式分组并覆盖优先级）

在这个列表之外，还有一些结构可以被分类为一个表达式，但是它们不遵循任何一般语法规则。这些通常具有一个函数或操作符的语义并且在第 9 章中的合适位置解释。一个示例是`IS NULL`子句。

我们已经在第 4.1.2 节中讨论过常量。下面的小节会讨论剩下的选项。

4.2.1. 列引用

一个列可以以下面的形式被引用：

correlation.*columnname*

*correlation*是一个表的名字（可能带有模式名限定），或者是在`FROM`子句中为某个表定义的别名。如果该列名在当前查询所用的所有表中都是唯一的，
则关联名称和分隔用的句点都可以省略（另见第 7 章）。

4.2.2. 位置参数

位置参数引用用来表示一个由 SQL 语句外部提供的值。参数可用于 SQL 函数定义和预备语句中。某些客户端库还支持把数据值与 SQL 命令字符串分开指定，在这种情况下，参数就用来引用这些命令字符串之外的数据值。参数引用的形式是：

$*number*

例如，考虑一个函数`dept`的定义：

```
CREATE FUNCTION dept(text) RETURNS dept
    AS $$ SELECT * FROM dept WHERE name = $1 $$
    LANGUAGE SQL;
```

这里`$1`引用函数被调用时第一个函数参数的值。

4.2.3. 下标

如果一个表达式得到了一个数组类型的值，那么可以抽取出该数组值的一个特定元素：

expression[*subscript*]

或者抽取出多个相邻元素（一个“数组切片”）：

expression[*lower_subscript*:*upper_subscript*]

（这里，方括号`[ ]`表示其字面意思）。每一个下标自身是一个表达式，它必须得到一个整数值。

通常，数组表达式必须加上圆括号，但如果要加下标的表达式只是列引用或位置参数，
则圆括号可以省略。另外，当原始数组是多维数组时，多个下标可以连写。例如：

```
mytable.arraycolumn[4]
mytable.two_d_column[17][34]
$1[10:42]
(arrayfunction(a,b))[42]
```

最后一个示例中的圆括号是必需的。详见第 8.14 节。

4.2.4. 字段选择

如果一个表达式得到一个复合类型（行类型）的值，那么可以抽取该行的指定字段：

expression.*fieldname*

通常，行表达式必须加上圆括号，但如果要取字段的表达式只是表引用或位置参数，则圆括号可以省略。例如：

```
mytable.mycolumn
$1.somecolumn
(rowfunction(a,b)).col3
```

（因此，限定列引用实际上只是字段选择语法的一种特例。）
一种重要的特例是从复合类型的表列中抽取字段：

```
(compositecol).somefield
(mytable.compositecol).somefield
```

这里需要圆括号来显示compositecol是一个列名而不是一个表名，在第二种情况中则是显示mytable是一个表名而不是一个模式名。

在选择列表中（参见第 7.3 节），你可以通过书写.*来请求一个组合值的所有字段：

```
(compositecol).*
```

4.2.5. 操作符调用

对于一次操作符调用，有三种可能的语法：

expression operator expression（二元中缀操作符）
operator expression（一元前缀操作符）
expression operator（一元后缀操作符）

其中 *operator* 词元遵循第 4.1.3 节的语法规则，或者是关键词 AND，OR 和 NOT 之一，或者是一个如下形式的模式限定操作符名：

OPERATOR(*schema*.*operatorname*)

具体存在哪些操作符以及它们是一元还是二元，取决于系统或用户定义了哪些操作符。第 9 章描述了内置操作符。

4.2.6. 函数调用

函数调用的语法是：函数名（可能带有模式名限定）后跟一组放在圆括号内的参数列表：

function_name ([*expression* [, *expression* ...]])

例如，下面会计算 2 的平方根：

```
sqrt(2)
```

内置函数的列表在第 9 章中。其他函数可以由用户增加。

参数也可以选择附带名称。详见第 4.3 节。

> **注意**
>
> 接受单个复合类型参数的函数，可以选择使用字段选择语法来调用；反过来，字段选择也可以写成函数风格。也就是说，`col(table)` 与 `table.col` 这两种记法可以互换。这种行为不是 SQL 标准的一部分，但 PostgreSQL 提供了它，因为这允许使用函数来模拟“计算字段”。详见第 35.4.2 节。

4.2.7. 聚合表达式

一个聚合表达式表示将聚合函数应用于查询选中的各行。
聚合函数将多个输入归约为单个输出值，例如输入的和或平均值。
聚合表达式的语法可以是以下形式之一：

```
aggregate_name (expression [ , ... ] [ order_by_clause ] )
aggregate_name (ALL expression [ , ... ] [ order_by_clause ] )
aggregate_name (DISTINCT expression [ , ... ] [ order_by_clause ] )
aggregate_name ( * )
```

其中，*aggregate_name* 是先前定义的聚合（可以带模式名限定），*expression* 是任何本身不包含聚合表达式或窗口函数调用的值表达式，而 *order_by_clause* 是一个可选的 `ORDER BY` 子句，在下文说明。

第一种形式的聚合表达式为每一个输入行调用一次聚合。第二种形式和第一种相同，因为`ALL`是默认选项。
第三种形式为输入行中表达式的每一个可区分值（或者对于多个表达式是值的可区分集合）调用一次聚合。最后一种形式为每一个输入行调用一次聚合，因为没有特定的输入值被指定，它通常只对于`count(*)`聚合函数有用。

大部分聚合函数忽略空输入，这样其中一个或多个表达式得到空值的行将被丢弃。除非另有说明，对于所有内置聚合都是这样。

例如，`count(*)`得到输入行的总数。`count(f1)`得到输入行中`f1`为非空的数量，因为`count`忽略空值。而`count(distinct f1)`得到`f1`的非空可区分值的数量。

通常，输入行以未指定的顺序传递给聚合函数。在许多情况下这无关紧要；例如，`min`无论以什么顺序接收输入都会产生相同的结果。但是，有些聚合函数（例如 `array_agg` 和 `string_agg`）的结果取决于输入行的顺序。使用这类聚合时，可选的 *order_by_clause* 可以用来指定所需的顺序。*order_by_clause*的语法与查询级 `ORDER BY` 子句相同，详见第 7.5 节，但其中的表达式只能是表达式，不能是输出列的名称或编号。例如：

```
SELECT array_agg(a ORDER BY b DESC) FROM table;
```

使用多参数聚合函数时，请注意 `ORDER BY` 子句位于所有聚合参数之后。例如，应写成：

```
SELECT string_agg(a, ',' ORDER BY a) FROM table;
```

而不是：

```
SELECT string_agg(a ORDER BY a, ',') FROM table;  -- incorrect
```

后者在语法上是有效的，但它表示调用一个带有两个 `ORDER BY` 排序键的单参数聚合函数（第二个排序键是常量，因此没什么用处）。

如果在 DISTINCT 之外还指定了 *order_by_clause*，那么所有 ORDER BY 表达式都必须与聚合的常规参数相匹配；也就是说，不能根据未包含在 DISTINCT 列表中的表达式排序。

> **注意**
>
> 在聚合函数中同时指定 DISTINCT 和 ORDER BY 的能力是 PostgreSQL 的扩展。

预定义的聚合函数在第 9.18 节中描述。其他聚合函数可以由用户增加。

一个聚合表达式只能出现在SELECT命令的结果列表或是HAVING子句中。
在其他子句（如WHERE）中禁止使用它，
因为那些子句的计算在逻辑上是在聚合的结果被形成之前。

当一个聚合表达式出现在一个子查询中（见第 4.2.11 节和第 9.20 节），聚合通常在该子查询的行上被计算。
但是如果该聚合的参数只包含外层变量则会产生一个异常：该聚合则属于最近的那个外层，
并且会在那个查询的行上被计算。
该聚合表达式从整体上则是对其所出现于的子查询的一种外层引用，
并且在那个子查询的任意一次计算中都作为一个常量。只出现在结果列表或HAVING
子句的限制适用于该聚合所属的查询层次。

4.2.8. 窗口函数调用

一次窗口函数调用表示在查询选出的部分行上应用一个类似聚合的函数。
与普通聚合函数调用不同，这并不意味着把选出的行分组为单个输出行 – 每一行在查询输出中仍然独立存在。不过，窗口函数能够扫描根据该窗口函数调用的分组说明（PARTITION BY 列表）属于当前行所在组的所有行。窗口函数调用的语法是下列之一：

```
function_name ([expression [, expression ... ]]) OVER window_name
function_name ([expression [, expression ... ]]) OVER ( window_
definition )
function_name ( * ) OVER window_name
function_name ( * ) OVER ( window_definition )
```

其中 *window_definition* 的语法为：

```
[ existing_window_name ]
[ PARTITION BY expression [, ...] ]
[ ORDER BY expression [ ASC | DESC | USING operator ] [ NULLS { FIRST
 | LAST } ] [, ...] ]
[ frame_clause ]
```

而可选的 *frame_clause* 可以是以下形式之一：

```
{ RANGE | ROWS } frame_start
{ RANGE | ROWS } BETWEEN frame_start AND frame_end
```

其中 *frame_start* 和 *frame_end* 可以是以下形式之一：

```
UNBOUNDED PRECEDING
value PRECEDING
CURRENT ROW
value FOLLOWING
```

UNBOUNDED FOLLOWING

这里，*expression*表示任何自身不含有窗口函数调用的值表达式。PARTITION BY和ORDER BY列表与整个查询的GROUP BY和ORDER BY子句具有本质上相同的语法和语义，唯一例外是它们的表达式始终只能是表达式，不能是输出列的名称或编号。*window_name*是对定义在查询的WINDOW子句中的一个命名窗口声明的引用。还可以使用在WINDOW子句中定义命名窗口的相同语法在圆括号内给定一个完整的*window_definition*，详见SELECT参考页。值得指出的是，OVER wname并不严格地等价于OVER (wname)，后者表示复制并修改窗口定义，并且在被引用窗口声明包括一个帧子句时会被拒绝。

*frame_clause*指定构成窗口帧的行集合，供那些作用于帧而不是整个分区的窗口函数使用。如果*frame_end*被省略，则默认为CURRENT ROW。限制是*frame_start*不能为UNBOUNDED FOLLOWING，*frame_end*不能为UNBOUNDED PRECEDING，并且在上述列表中，*frame_end*选项不能出现在*frame_start*选项之前 — 例如，不允许RANGE BETWEEN CURRENT ROW AND *value* PRECEDING。默认的帧选项是RANGE UNBOUNDED PRECEDING，它和RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW相同；这会把该帧设置为从分区开始一直到当前行的最后一个ORDER BY同等行的所有行（如果没有ORDER BY，则是所有行）。一般来说，UNBOUNDED PRECEDING表示该帧开始于分区的第一行，类似地UNBOUNDED FOLLOWING表示该帧结束于分区的最后一行（不管是RANGE还是ROWS模式）。在ROWS模式中，CURRENT ROW表示该帧开始或结束于当前行；而在RANGE模式中，它表示该帧开始或结束于当前行在ORDER BY次序中的第一个或最后一个同等行。*value* PRECEDING和*value* FOLLOWING目前只允许用于ROWS模式。它们表示帧开始或结束于当前行之前或之后那么多行的位置。*value*必须是一个不包含任何变量、聚合函数或窗口函数的整数表达式。其值不能为空值或负数；但可以为零，这会只选择当前行本身。

内置的窗口函数在表 9.45中介绍。用户可以加入其他窗口函数。此外，任何内置的或者用户定义的聚合函数都可以被用作窗口函数。

使用*的语法被用来把参数较少的聚合函数当作窗口函数调用，例如count(*) OVER (PARTITION BY x ORDER BY y)。*通常不被用于非聚合的窗口函数。与普通聚合函数不同，聚合窗口函数不允许在函数参数列表中使用DISTINCT或ORDER BY。

只有在SELECT列表和查询的ORDER BY子句中才允许窗口函数调用。

更多关于窗口函数的信息可以在第 3.5 节、第 9.19 节和第 7.2.4 节中找到。

4.2.9. 类型转换

一个类型转换指定从一种数据类型到另一种数据类型的转换。PostgreSQL接受两种等价的类型转换语法：

CAST (*expression* AS *type*)
expression::*type*

CAST语法遵从 SQL，而用::的语法是PostgreSQL的历史用法。

当类型转换被应用于一种已知类型的值表达式时，它表示一次运行时类型转换。只有在已经定义了合适的类型转换操作时，该类型转换才会成功。注意，这与常量上的类型转换（如第 4.1.2.7 节中所示）略有不同。应用于未修饰字符串字面量的类型转换，表示为一个字面常量值赋予初始类型，因此它对任意类型都能成功（前提是该字符串字面量的内容符合该数据类型的输入语法）。

如果一个值表达式必须产生的类型没有歧义（例如当它被赋给一个表列时），通常可以省略显式类型转换，在这种情况下系统会自动应用类型转换。但是，

自动类型转换只会对那些在系统目录中被标记为“可以隐式应用”的类型转换执行。
其他类型转换必须使用显式类型转换语法来调用。
施加这一限制的目的是防止出人意料的转换被悄无声息地应用。

还可以用类似函数调用的语法来指定类型转换：

typename (*expression*)

不过，这只对那些名称本身也可作为函数名使用的类型有效。例如，double precision 不能用这种方式，但等效的 float8 可以。另外，由于语法冲突，名称 interval、time 和 timestamp 只有在使用双引号引用时才能采用这种写法。因此，函数风格的类型转换语法会带来不一致性，通常应尽量避免。

> **注意**
>
> 函数风格的语法事实上只是一次函数调用。
> 当两种标准类型转换语法之一被用来做一次运行时转换时，
> 它将在内部调用一个已注册的函数来执行该转换。简而言之，
> 这些转换函数具有和它们的输出类型相同的名字，并且因此“函数风格的语法”
> 无非是对底层转换函数的一次直接调用。显然，一个可移植的应用不应当依赖于它。
> 详见CREATE CAST。

4.2.10. 排序规则表达式

COLLATE 子句会覆盖表达式的排序规则。它附加在所作用的表达式之后：

expr COLLATE *collation*

这里 *collation* 可以是一个带模式名限定的标识符。COLLATE 子句比操作符绑定得更紧，必要时可以使用圆括号。

如果没有显式指定排序规则，数据库系统会从表达式所涉及的列中得到一个排序规则，
如果该表达式没有涉及列，则会默认采用数据库的默认排序规则。

COLLATE 子句的两种常见用途是覆盖 ORDER BY 子句中的排序顺序，例如：

```
SELECT a, b, c FROM tbl WHERE ... ORDER BY a COLLATE "C";
```

以及覆盖具有区域敏感结果的函数或操作符调用的排序规则，例如：

```
SELECT * FROM tbl WHERE a > 'foo' COLLATE "C";
```

注意，在后一种情况中，COLLATE 子句是附加到我们希望影响的那个操作符的某个输入参数上的。把 COLLATE 子句附加到该操作符或函数调用的哪个参数上并不重要，因为操作符或函数实际采用的排序规则是综合所有参数得出的，而显式的 COLLATE 子句会覆盖其他所有参数的排序规则。（不过，若把不匹配的 COLLATE 子句附加到多个参数上，则会报错。详见第 22.2 节。）因此，这会得到与前一个示例相同的结果：

```
SELECT * FROM tbl WHERE a COLLATE "C" > 'foo';
```

但是这是一个错误：

```
SELECT * FROM tbl WHERE (a > 'foo') COLLATE "C";
```

因为它尝试把一个排序规则应用到>操作符的结果，而它的数据类型是不支持排序规则的数据类型 boolean。

4.2.11. 标量子查询

标量子查询是一个放在圆括号中的普通 SELECT 查询，它恰好返回一行一列（关于如何编写查询，见第 7 章）。该 SELECT 查询会被执行，其返回的单个值会用在外围值表达式中。把一个返回多于一行或多于一列的查询当作标量子查询来用是错误的。
（但如果在某次具体执行中该子查询没有返回任何行，则不算错误；
其标量结果会被视为空值。）该子查询可以引用外围查询中的变量，
这些变量在该子查询的每一次求值期间都会作为常量。有关其他涉及子查询的表达式，另见第 9.20 节。

例如，下列语句会寻找每个州中最大的城市人口：

```
SELECT name, (SELECT max(pop) FROM cities WHERE cities.state = states.
name)
    FROM states;
```

4.2.12. 数组构造器

数组构造器是一个表达式，它使用其成员元素的值来构造一个数组值。
简单的数组构造器由关键字 ARRAY、左方括号 [、一个用于指定数组元素值的表达式列表（以逗号分隔）以及最后的右方括号] 组成。例如：

```
SELECT ARRAY[1,2,3+4];
  array
---------
 {1,2,7}
(1 row)
```

默认情况下，数组元素类型是成员表达式的公共类型，其确定规则与 UNION 或 CASE 结构相同（见第 10.5 节）。你可以通过把数组构造器显式转换为所需类型来覆盖这一点，例如：

```
SELECT ARRAY[1,2,22.7]::integer[];
  array
----------
 {1,2,23}
(1 row)
```

这与把每个表达式分别转换为数组元素类型具有相同的效果。关于类型转换的更多信息，见第 4.2.9 节。

多维数组值可以通过嵌套数组构造器来构造。在内部构造器中，关键字 ARRAY 可以省略。例如，下面两种写法产生相同的结果：

```
SELECT ARRAY[ARRAY[1,2], ARRAY[3,4]];
     array
---------------
 {{1,2},{3,4}}
(1 row)

SELECT ARRAY[[1,2],[3,4]];
     array
```

```
---------------
 {{1,2},{3,4}}
(1 row)
```

由于多维数组必须是长方形的，因此同一级别的内部构造器必须生成维度完全相同的子数组。施加到外层 ARRAY 构造器上的任何类型转换，都会自动传播到所有内层构造器。

多维数组构造器中的元素可以是任何能够生成适当类型数组的表达式，而不仅仅是子 ARRAY 构造。例如：

```
CREATE TABLE arr(f1 int[], f2 int[]);

INSERT INTO arr VALUES (ARRAY[[1,2],[3,4]], ARRAY[[5,6],[7,8]]);

SELECT ARRAY[f1, f2, '{{9,10},{11,12}}'::int[]] FROM arr;
                     array
------------------------------------------------
 {{{1,2},{3,4}},{{5,6},{7,8}},{{9,10},{11,12}}}
(1 row)
```

你也可以构造空数组，但由于不可能存在没有类型的数组，因此必须把空数组显式转换为所需类型。例如：

```
SELECT ARRAY[]::integer[];
 array
-------
 {}
(1 row)
```

也可以根据子查询的结果构造数组。在这种形式中，数组构造器写为关键字 ARRAY 后跟一个用圆括号（而不是方括号）括起的子查询。例如：

```
SELECT ARRAY(SELECT oid FROM pg_proc WHERE proname LIKE 'bytea%');
                          ?column?
---------------------------------------------------------------
 {2011,1954,1948,1952,1951,1244,1950,2005,1949,1953,2006,31}
(1 row)
```

子查询必须返回单列。生成的数组为子查询结果中的每一行包含一个元素，元素类型与子查询输出列的类型相匹配。

用 ARRAY 构造出来的数组值，其下标总是从 1 开始。关于数组的更多信息，见第 8.14 节。

4.2.13. 行构造器

行构造器是一种表达式，它使用成员字段的值来构造一个行值（也称为组合值）。行构造器由关键字 ROW、左圆括号、零个或多个用于指定行各字段值的表达式（以逗号分隔），以及最后的右圆括号组成。例如：

```
SELECT ROW(1,2.5,'this is a test');
```

当列表中有多个表达式时，关键字 ROW 是可选的。

一个行构造器可以包括语法*rowvalue*.*，它将被扩展为该行值的元素的一个列表，就像在一个顶层SELECT列表中使用.*时发生的事情一样。例如，如果表t有列f1和f2，那么这些是相同的：

```
SELECT ROW(t.*, 42) FROM t;
SELECT ROW(t.f1, t.f2, 42) FROM t;
```

> **注意**
>
> 在PostgreSQL 8.2 以前，`.*` 语法不会被展开，因此写成 `ROW(t.*, 42)` 会创建一个具有两个字段的行，其中第一个字段是另一个行值。新的行为通常更有用。如果你需要旧版本那种嵌套行值的行为，写内层行值时不要使用 `.*`，例如 `ROW(t, 42)`。

默认情况下，`ROW` 表达式创建的值具有匿名记录类型。如有需要，可以把它转换为具名复合类型，也就是某个表的行类型，或者使用 `CREATE TYPE AS` 创建的复合类型。为避免歧义，可能需要显式类型转换。例如：

```
CREATE TABLE mytable(f1 int, f2 float, f3 text);

CREATE FUNCTION getf1(mytable) RETURNS int AS 'SELECT $1.f1' LANGUAGE
 SQL;

-- No cast needed since only one getf1() exists
SELECT getf1(ROW(1,2.5,'this is a test'));
 getf1
-------
     1
(1 row)

CREATE TYPE myrowtype AS (f1 int, f2 text, f3 numeric);

CREATE FUNCTION getf1(myrowtype) RETURNS int AS 'SELECT $1.f1'
 LANGUAGE SQL;

-- Now we need a cast to indicate which function to call:
SELECT getf1(ROW(1,2.5,'this is a test'));
ERROR:  function getf1(record) is not unique

SELECT getf1(ROW(1,2.5,'this is a test')::mytable);
 getf1
-------
     1
(1 row)

SELECT getf1(CAST(ROW(11,'this is a test',2.5) AS myrowtype));
 getf1
-------
    11
(1 row)
```

行构造器可以用来构造复合值，以存储在复合类型的表列中，或者传递给接受复合参数的函数。此外，还可以比较两个行值，或者用 `IS NULL` 或 `IS NOT NULL` 来测试一个行，例如：

```
SELECT ROW(1,2.5,'this is a test') = ROW(1, 3, 'not the same');

SELECT ROW(table.*) IS NULL FROM table;  -- detect all-null rows
```

更多详情见第 9.21 节。行构造器也可以与子查询结合使用，详见第 9.20 节。

4.2.14. 表达式计算规则

子表达式的计算顺序是未定义的。特别是，
操作符或函数的输入不一定会按照从左到右或任何其他固定顺序计算。

此外，如果一个表达式的结果可以通过只计算其一部分来决定，
那么其他子表达式可能完全不需要被计算。例如，如果我们写：

```
SELECT true OR somefunc();
```

那么`somefunc()`将（可能）完全不被调用。如果我们写成下面这样也是一样：

```
SELECT somefunc() OR true;
```

注意这和一些编程语言中布尔操作符从左至右的“短路”不同。

因此，在复杂表达式中使用带有副作用的函数是不明智的。在`WHERE`和`HAVING`
子句中依赖副作用或计算顺序尤其危险，
因为在建立一个执行计划时这些子句会被广泛地重新处理。这些子句中布尔表达式（`AND`/`OR`
/`NOT`的组合）可能会以布尔代数定律所允许的任何方式被重组。

当有必要强制计算顺序时，可以使用一个`CASE`结构（见第 9.16 节）。例如，在一个`WHERE`子句
中使用下面的方法尝试避免除零是不可靠的：

```
SELECT ... WHERE x > 0 AND y/x > 1.5;
```

但是这是安全的：

```
SELECT ... WHERE CASE WHEN x > 0 THEN y/x > 1.5 ELSE false END;
```

一个以这种风格使用的`CASE`结构将使得优化尝试失败，
因此只有必要时才这样做（在这个特别的示例中，最好通过写`y > 1.5*x`来回避这个问题）。

不过，`CASE` 并不是这类问题的万灵药。上述技术的一个限制是，它无法阻止常量子表达式被提前
计算。如第 35.6 节中所述，当查询被规划而不是被执行时，被标记成 `IMMUTABLE`的函数和操作
符可以被计算。因此

```
SELECT CASE WHEN x > 0 THEN x ELSE 1/0 END FROM tab;
```

很可能会导致一次除零失败，因为规划器尝试简化常量子表达式。即便是表中的每一行都有`x > 0`
（这样运行时永远不会进入到 `ELSE`分支）也是这样。

虽然这个例子看起来有些可笑，但在函数内部执行的查询中，
也可能出现一些看上去并未明显涉及常量的相关情况，
因为函数参数值和局部变量值为了规划目的，可能会作为常量插入到查询中。例如，在 PL/
pgSQL 函数中，使用 `IF-THEN-ELSE` 语句来保护一项有风险的计算，要比仅仅把它嵌套进一个
`CASE` 表达式安全得多。

另一个同类型的限制是，一个`CASE`无法阻止其所包含的聚合表达式的计算，因为在考虑`SELECT`
列表或`HAVING`子句中的其他表达式之前，会先计算聚合表达式。例如，
下面的查询可能会导致除零错误，尽管看起来似乎已经对这种情况做了防护：

```
SELECT CASE WHEN min(employees) > 0
            THEN avg(expenses / employees)
```

```
        END
    FROM departments;
```

min()和avg()聚合会在所有输入行上并行地计算，因此如果任何行有employees等于零，在有机会测试 min()的结果之前，就会发生除零错误。取而代之的是，可以使用一个WHERE子句来首先阻止有问题的输入行到达一个聚合函数。

4.3. 调用函数

PostgreSQL 允许对带命名参数的函数使用位置记法或命名记法来调用。命名记法对于拥有大量参数的函数尤其有用，因为它能让形参与实参之间的对应关系更明确、更可靠。在位置记法中，函数调用中的参数值按照它们在函数声明中定义的顺序书写；在命名记法中，实参与函数参数按名称匹配，因此可以按任意顺序书写。

在任意一种记法中，在函数声明中给出了默认值的参数都可以在调用中省略。不过这在命名记法中特别有用，因为任意组合的参数都可以被省略；而在位置记法中，参数只能从右向左省略。

PostgreSQL 也支持混合记法，它结合了位置记法和命名记法。在这种情况下，位置参数先写，命名参数写在后面。

下面的示例将展示这三种记法的用法，所用函数定义如下：

```
CREATE FUNCTION concat_lower_or_upper(a text, b text, uppercase
 boolean DEFAULT false)
RETURNS text
AS
$$
 SELECT CASE
        WHEN $3 THEN UPPER($1 || ' ' || $2)
        ELSE LOWER($1 || ' ' || $2)
        END;
$$
LANGUAGE SQL IMMUTABLE STRICT;
```

函数 concat_lower_or_upper 有两个必需参数 a 和 b。此外还有一个可选参数 uppercase，其默认值为 false。a 和 b 的输入将被连接起来，并根据 uppercase 参数的值转换成大写或小写。该函数定义的其余细节在这里并不重要（详见第 35 章）。

4.3.1. 使用位置记法

位置记法是在 PostgreSQL 中向函数传递参数的传统机制。例如：

```
SELECT concat_lower_or_upper('Hello', 'World', true);
 concat_lower_or_upper
-----------------------
 HELLO WORLD
(1 row)
```

所有参数都按顺序给出。由于 uppercase 被指定为 true，因此结果是大写。另一个例子是：

```
SELECT concat_lower_or_upper('Hello', 'World');
 concat_lower_or_upper
-----------------------
```

```
 hello world
(1 row)
```

这里省略了 uppercase 参数，因此它接收默认值 false，结果是小写输出。在位置记法中，只要参数具有默认值，就可以从右向左依次省略。

4.3.2. 使用命名记法

在命名记法中，每个参数名都使用 := 与其参数表达式分隔。例如：

```
SELECT concat_lower_or_upper(a := 'Hello', b := 'World');
 concat_lower_or_upper
-----------------------
 hello world
(1 row)
```

同样，uppercase 参数被省略，因此它被隐式设置为 false。使用命名记法的一个优点是，参数可以按任意顺序指定，例如：

```
SELECT concat_lower_or_upper(a := 'Hello', b := 'World', uppercase :=
 true);
 concat_lower_or_upper
-----------------------
 HELLO WORLD
(1 row)

SELECT concat_lower_or_upper(a := 'Hello', uppercase := true, b :=
 'World');
 concat_lower_or_upper
-----------------------
 HELLO WORLD
(1 row)
```

4.3.3. 使用混合记法

混合记法结合了位置记法和命名记法。不过，如前所述，命名参数不能出现在位置参数之前。例如：

```
SELECT concat_lower_or_upper('Hello', 'World', uppercase := true);
 concat_lower_or_upper
-----------------------
 HELLO WORLD
(1 row)
```

在上面的查询中，参数 a 和 b 采用位置方式指定，而 uppercase 采用命名方式指定。在这个示例中，这样做除了起到一定说明作用外并没有太多额外价值。
但对于拥有大量带默认值参数的更复杂函数，命名记法或混合记法可以大幅减少书写量，并降低出错机会。

注意

命名调用记法和混合调用记法当前只能用于常规函数，不能用于聚合函数或窗口函数。

第 5 章 数据定义

本章介绍如何创建用于保存数据的数据库结构。在关系数据库中，原始数据存储在表中，因此本章的大部分内容都在说明如何创建和修改表，以及有哪些特性可用于控制表中存储的数据。随后，我们将讨论如何把表组织到模式中，以及如何向表分配权限。最后，我们会简要介绍一些同样会影响数据存储的特性，例如继承、视图、函数和触发器。

5.1. 表基础

关系数据库中的表很像纸面上的表格：它由行和列组成。列的数量和顺序是固定的，每一列都有名称。行的数量则是可变的，它反映了某一时刻存储了多少数据。SQL 不对表中行的顺序作任何保证。读取表时，除非显式要求排序，否则行会以未指定的顺序出现，参见第 7 章。此外，SQL 不会为行分配唯一标识符，因此一个表中可能出现多行完全相同的记录。这是 SQL 所依据的数学模型带来的结果，但通常并不理想。本章稍后将看到如何处理这个问题。

每一列都有一种数据类型。数据类型限制了可以赋给该列的可能值集合，并为列中存储的数据赋予语义，从而使其能够用于计算。例如，声明为数值类型的列不能接受任意文本字符串，存储在这类列中的数据可以参与数学运算。相比之下，声明为字符串类型的列几乎能接受任何种类的数据，但并不适合用于数学计算，尽管仍可进行字符串拼接等其他操作。

PostgreSQL提供了相当丰富的一组内置数据类型，适用于许多应用场景。用户也可以定义自己的数据类型。大多数内置数据类型的名称和语义都很直观，因此我们把详细说明留到第 8 章。一些常用的数据类型包括：表示整数的`integer`、表示可能带小数部分数值的`numeric`、表示字符串的`text`、表示日期的`date`、表示一天中时间值的`time`，以及同时包含日期和时间的`timestamp`。

要创建一个表，可使用名副其实的CREATE TABLE命令。在该命令中，至少要指定新表的名称、各列的名称以及每列的数据类型。例如：

```
CREATE TABLE my_first_table (
    first_column text,
    second_column integer
);
```

这会创建一个名为`my_first_table`的表，它有两列。第一列名为`first_column`，数据类型为`text`；第二列名为`second_column`，类型为`integer`。表名和列名遵循第 4.1.1 节中解释的标识符语法。类型名通常也是标识符，但也有少数例外。注意，列列表以逗号分隔，并用圆括号括起来。

当然，前面的例子相当刻意。通常，你会为表和列起能体现其所存储数据含义的名称。下面看一个更实际的例子：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric
);
```

（`numeric`类型可以存储小数部分，这对金额很常见。）

> **提示**
>
> 在创建许多相互关联的表时，最好为表和列选择一致的命名模式。例如，表名可以使用单数名词或复数名词，两种做法都各有支持者。

一个表能包含的列数是有限的。根据列类型不同，这个上限在 250 到 1600 之间。不过，定义一个接近这个数量的表极不常见，而且其设计往往值得商榷。

如果不再需要某个表，可以使用DROP TABLE命令删除它。例如：

```
DROP TABLE my_first_table;
DROP TABLE products;
```

尝试删除不存在的表会报错。不过，在 SQL 脚本文件中，人们常常在创建每个表之前无条件地尝试删除它，并忽略可能出现的错误消息，这样脚本无论表是否存在都能工作。（如果愿意，也可以使用`DROP TABLE IF EXISTS`变体来避免错误消息，但这不是标准 SQL。）

如果我们需要修改一个已经存在的表，请参考本章稍后的第 5.5 节。

利用到目前为止所讨论的工具，我们已经可以创建功能完备的表。本章其余内容关注如何向表定义添加特性，以确保数据完整性、安全性或使用便利。如果你现在就急着向表中填充数据，可以先跳到第 6 章，稍后再回来阅读本章其余内容。

5.2. 默认值

列可以被赋予默认值。当创建新行时，如果没有为某些列指定值，这些列就会填入各自的默认值。数据操作命令也可以显式要求把某列设为其默认值，而不必知道默认值究竟是什么。（数据操作命令详见第 6 章。）

如果没有显式指定默认值，则默认值是空值。这是合理的，因为空值表示未知数据。

在表定义中，默认值写在列的数据类型之后。例如：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric DEFAULT 9.99
);
```

默认值也可以是表达式。每次需要插入默认值时都会对其求值（不是在创建表时）。一个常见例子是让`timestamp`列的默认值为`CURRENT_TIMESTAMP`，这样它会被设为插入该行时的时间。另一个常见例子是为每一行生成一个“序列号”。在PostgreSQL中，这通常可以这样实现：

```
CREATE TABLE products (
    product_no integer DEFAULT nextval('products_product_no_seq'),
    ...
);
```

其中`nextval()`函数从序列对象（见第 9.15 节）中依次取值。这种写法非常常见，因此还有一种专门的简写形式：

```
CREATE TABLE products (
```

```
    product_no SERIAL,
    ...
);
```

SERIAL简写会在第 8.1.4 节进一步讨论。

5.3. 约束

数据类型是一种限制能够存储在表中的数据类别的方法。但是对于很多应用来说，它们提供的约束太粗糙。例如，一个包含产品价格的列应该只接受正值。但是没有任何一种标准数据类型只接受正值。另一个问题是我们可能需要根据其他列或行来约束一个列中的数据。例如，在一个包含产品信息的表中，对于每个产品编号应该只有一行。

为此，SQL 允许我们在列和表上定义约束。约束让我们能够按照需要控制表中的数据。如果用户试图在列中存储违反约束的数据，就会报错。即使该值来自默认值定义，这条规则也同样适用。

5.3.1. 检查约束

检查约束是最通用的约束类型。它允许我们指定某一列中的值必须满足一个布尔（真值）表达式。例如，要要求产品价格为正值，可以使用：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric CHECK (price > 0)
);
```

如你所见，约束定义和默认值定义一样都写在数据类型之后。默认值和约束的先后顺序没有影响。检查约束由关键字CHECK以及其后放在圆括号中的表达式组成。检查约束表达式应当涉及被约束的列，否则这个约束就没有太大意义。

我们也可以为约束单独指定一个名称。这样可以让错误消息更清晰，也便于在需要修改约束时引用它。语法如下：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric CONSTRAINT positive_price CHECK (price > 0)
);
```

要指定一个命名约束，在约束名标识符前写关键字CONSTRAINT，再在其后写约束定义即可。（如果没有用这种方式指定约束名，系统会为你选择一个。）

一个检查约束也可以引用多个列。例如我们存储一个普通价格和一个打折后的价格，而我们希望保证打折后的价格低于普通价格：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric CHECK (price > 0),
    discounted_price numeric CHECK (discounted_price > 0),
    CHECK (price > discounted_price)
```

```
);
```

前两个约束看起来很相似。第三个则使用了一种新语法。它并没有依附在一个特定的列，而是作为一个独立的项出现在逗号分隔的列列表中。
列定义和这种约束定义可以以混合的顺序出现在列表中。

我们将前两个约束称为列约束，而第三个约束为表约束，因为它独立于任何一个列定义。列约束也可以写成表约束，但反过来不行，因为一个列约束只能引用它所依附的那一个列（PostgreSQL并不强制要求这个规则，但是如果我们希望表定义能够在其他数据库系统中工作，那就应该遵循它）。上述示例也可以写成：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric,
    CHECK (price > 0),
    discounted_price numeric,
    CHECK (discounted_price > 0),
    CHECK (price > discounted_price)
);
```

甚至是：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric CHECK (price > 0),
    discounted_price numeric,
    CHECK (discounted_price > 0 AND price > discounted_price)
);
```

这只是口味的问题。

表约束也可以用与列约束相同的方法来指定名称：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric,
    CHECK (price > 0),
    discounted_price numeric,
    CHECK (discounted_price > 0),
    CONSTRAINT valid_discount CHECK (price > discounted_price)
);
```

需要注意的是，当检查表达式的值为真或空值时，检查约束就被视为满足。由于当任一操作数为空时，大多数表达式都会计算为空值，所以检查约束不会阻止受约束列中出现空值。要确保某列不包含空值，可以使用下一节介绍的非空约束。

5.3.2. 非空约束

非空约束只是指定某列不能取空值。语法示例如下：

```
CREATE TABLE products (
```

```
    product_no integer NOT NULL,
    name text NOT NULL,
    price numeric
);
```

非空约束总是写成列约束。从功能上看，非空约束等价于创建检查约束CHECK (*column_name* IS NOT NULL)，但在PostgreSQL中创建显式的非空约束效率更高。缺点是无法为以这种方式创建的非空约束显式指定名称。

当然，一列可以有多个约束。只需将约束逐个写出：

```
CREATE TABLE products (
    product_no integer NOT NULL,
    name text NOT NULL,
    price numeric NOT NULL CHECK (price > 0)
);
```

顺序并不重要，也不一定决定检查约束的顺序。

该NOT NULL约束有一个反面形式：NULL约束。这并不表示该列必须为空值，那显然毫无用处。它只是显式选择列可以为空的默认行为。NULL约束不属于 SQL 标准，因此不应在需要可移植性的应用中使用。（它之所以被加入PostgreSQL，只是为了兼容某些其他数据库系统。）不过，有些用户喜欢它，因为它让在脚本文件中切换该约束变得比较容易。例如，可以从下面的定义开始：

```
CREATE TABLE products (
    product_no integer NULL,
    name text NULL,
    price numeric NULL
);
```

然后在需要的地方插入NOT关键字。

> **提示**
>
> 在大多数数据库设计中，多数列都应标记为非空。

5.3.3. 唯一约束

唯一约束保证某一列或某一组列中保存的数据在整个表的所有行之间都是唯一的。写成列约束时的语法是：

```
CREATE TABLE products (
    product_no integer UNIQUE,
    name text,
    price numeric
);
```

写成表约束时则是：

```
CREATE TABLE products (
    product_no integer,
    name text,
```

```
    price numeric,
    UNIQUE (product_no)
);
```

要为一组列定义一个唯一约束，把它写作一个表级约束，列名用逗号分隔：

```
CREATE TABLE example (
    a integer,
    b integer,
    c integer,
    UNIQUE (a, c)
);
```

这指定这些列的组合值在整个表的范围内是唯一的，
但其中任意一列的值并不需要是（一般也不是）唯一的。

我们可以按通常的方式为唯一约束指定名称：

```
CREATE TABLE products (
    product_no integer CONSTRAINT must_be_different UNIQUE,
    name text,
    price numeric
);
```

增加一个唯一约束会在约束中列出的列或列组上自动创建一个唯一 B-树索引。只覆盖某些行的唯一性限制不能写成唯一约束，但可以通过创建唯一的部分索引来强制这种限制。

通常情况下，如果表中有多行，且约束所包含的所有列的值都相等，就违反了唯一约束。不过，在这种比较中，两个空值从不被视为相等。这意味着即使存在唯一约束，也可以存储重复行，只要其中至少一个受约束列含有空值。此行为符合 SQL 标准，但我们听说其他 SQL 数据库可能不遵循此规则。因此，开发需要可移植性的应用时要小心。

5.3.4. 主键

一个主键约束表示可以用作表中行的唯一标识符的一个列或者一组列。
这要求那些值都是唯一的并且非空。因此，下面的两个表定义接受相同的数据：

```
CREATE TABLE products (
    product_no integer UNIQUE NOT NULL,
    name text,
    price numeric
);
```

```
CREATE TABLE products (
    product_no integer PRIMARY KEY,
    name text,
    price numeric
);
```

主键也可以包含多于一个列，其语法和唯一约束相似：

```
CREATE TABLE example (
    a integer,
    b integer,
    c integer,
```

```
    PRIMARY KEY (a, c)
);
```

增加一个主键会自动在主键中列出的列或列组上创建一个唯一 B-树索引，并强制这些列被标记为 NOT NULL。

一个表最多只能有一个主键（可以有任意数量的唯一和非空约束，
它们可以达到和主键几乎一样的功能，但只能有一个被标识为主键）。
关系数据库理论要求每一个表都要有一个主键。PostgreSQL 并未强制要求这一点，但通常最好遵循它。

主键对于文档和客户端应用都是有用的。例如，一个允许修改行值的 GUI 应用可能需要知道一个表的主键，以便能唯一地标识行。如果定义了主键，数据库系统也有多种方法来利用主键。
例如，主键定义了外键要引用的默认目标列。

5.3.5. 外键

一个外键约束指定一列（或一组列）中的值必须匹配出现在另一个表中某些行的值。
我们说这维持了两个关联表之间的引用完整性。

例如我们有一个使用过多次的产品表：

```
CREATE TABLE products (
    product_no integer PRIMARY KEY,
    name text,
    price numeric
);
```

让我们假设我们还有一个存储这些产品订单的表。
我们希望保证订单表中只包含真正存在的产品的订单。
因此我们在订单表中定义一个引用产品表的外键约束：

```
CREATE TABLE orders (
    order_id integer PRIMARY KEY,
    product_no integer REFERENCES products (product_no),
    quantity integer
);
```

现在就不可能创建包含不存在于产品表中的product_no值的订单。

我们说在这种情况下，订单表是引用表而产品表是被引用表。相应地，
也有引用和被引用列的说法。

我们也可以把上述命令简写为：

```
CREATE TABLE orders (
    order_id integer PRIMARY KEY,
    product_no integer REFERENCES products,
    quantity integer
);
```

因为如果缺少列的列表，则被引用表的主键将被用作被引用列。

一个外键也可以约束和引用一组列。照例，它需要被写成表约束的形式。下面是一个示例：

```
CREATE TABLE t1 (
```

```
  a integer PRIMARY KEY,
  b integer,
  c integer,
  FOREIGN KEY (b, c) REFERENCES other_table (c1, c2)
);
```

当然，被约束列的数量和类型应该匹配被引用列的数量和类型。

你可以按常规方式为外键约束指定自己的名称。

一个表可以含有多个外键约束。这可用于实现表之间的多对多关系。假设有产品表和订单表，现在希望一个订单可以包含多个产品（上面的结构不允许这样做）。可以使用如下表结构：

```
CREATE TABLE products (
    product_no integer PRIMARY KEY,
    name text,
    price numeric
);

CREATE TABLE orders (
    order_id integer PRIMARY KEY,
    shipping_address text,
    ...
);

CREATE TABLE order_items (
    product_no integer REFERENCES products,
    order_id integer REFERENCES orders,
    quantity integer,
    PRIMARY KEY (product_no, order_id)
);
```

注意，最后一个表中的主键与外键有重叠。

我们知道，外键不允许创建与任何产品都无关的订单。但如果先创建了引用某产品的订单，随后又删除该产品，会怎样呢？SQL 也允许处理这种情况。直观来看，有以下几种选择：

- 不允许删除被引用的产品
- 同时删除订单
- 其他处理方式？

为说明这一点，在上面的多对多关系示例中实行如下策略：如果有人要删除仍被订单引用（通过 `order_items`）的产品，则不允许删除。如果有人删除订单，则同时删除订单项：

```
CREATE TABLE products (
    product_no integer PRIMARY KEY,
    name text,
    price numeric
);

CREATE TABLE orders (
    order_id integer PRIMARY KEY,
    shipping_address text,
    ...
);
```

```
CREATE TABLE order_items (
    product_no integer REFERENCES products ON DELETE RESTRICT,
    order_id integer REFERENCES orders ON DELETE CASCADE,
    quantity integer,
    PRIMARY KEY (product_no, order_id)
);
```

限制删除和级联删除是最常见的两种选项。RESTRICT阻止删除被引用的行。NO ACTION表示检查约束时若仍存在引用行，就会报错；未指定任何选项时，这是默认行为。（这两种选项的根本区别在于，NO ACTION允许将检查延迟到事务稍后进行，而RESTRICT不允许。）CASCADE指定当被引用行被删除时，引用它的行也应自动删除。另有两个选项：SET NULL和SET DEFAULT。它们会在被引用行删除时，分别把引用列设为空值或默认值。注意，这并不能让你免于遵守其他约束。例如，如果某个动作指定了SET DEFAULT，但默认值本身不满足外键，则该操作仍会失败。

与ON DELETE类似，还有ON UPDATE，当被引用列发生变化（更新）时就会触发。可用动作相同。

由于从被引用表中DELETE行或UPDATE被引用列时，需要扫描引用表以查找匹配旧值的行，因此通常建议为引用列建立索引。由于这并不总是必需，而且索引方式也有很多种，因此声明外键约束时不会自动在引用列上创建索引。

更多关于更新和删除数据的信息请见第 6 章。

最后，我们要提到外键必须引用作为主键或组成唯一约束的列。如果外键引用的是一个唯一约束，那么在如何匹配空值方面还有一些额外的可能性。这些在CREATE TABLE的参考文档中有解释。

5.3.6. 排他约束

排他约束保证：对任意两行，若对指定列或表达式使用指定操作符进行比较，则这些操作符比较中至少有一个会返回假或空值。语法如下：

```
CREATE TABLE circles (
    c circle,
    EXCLUDE USING gist (c WITH &&)
);
```

详见CREATE TABLE ... CONSTRAINT ... EXCLUDE。

增加一个排他约束会自动创建约束声明中指定类型的索引。

5.4. 系统列

每个表都有几个由系统隐式定义的系统列。因此，这些名称不能用作用户定义列的名称。（注意，这些限制与名称是否为关键字无关；给名称加引号也不能绕过这些限制。）你其实不必关心这些列的细节，只需要知道它们存在即可。

oid

行的对象标识符（对象 ID）。只有使用WITH OIDS创建表，或创建时设置了default_with_oids配置变量，才存在此列。该列的类型是oid（与列同名）；关于该类型的更多信息，参见第 8.16 节。

`tableoid`

包含该行的表的 OID。对于从继承层次（见第 5.8 节）中查询数据的场景，这一列尤其方便，因为如果没有它，就很难判断一行究竟来自哪个具体表。`tableoid`可以与`pg_class`的`oid`列连接，以取得表名。

`xmin`

插入该行版本的事务 ID。（行版本是某一行的一个具体状态；对同一逻辑行的每次更新都会创建一个新的行版本。）

`cmin`

插入事务中的命令标识符（从0开始）。

`xmax`

删除事务的标识（事务 ID）；对于未删除的行版本则为 0。对于一个可见的行版本，该列也可能是非零值。这通常表示删除事务尚未提交，或者一次删除尝试被回滚了。

`cmax`

删除事务中的命令标识符，或者为0。

`ctid`

行版本在其表中的物理位置。注意尽管`ctid`可以被用来非常快速地定位行版本，但是一个行的`ctid`会在被更新或者被`VACUUM FULL`移动时改变。因此，`ctid`不适合用作长期行标识符。应使用 OID，或者更好地使用用户定义的序列号，来标识逻辑行。

OID 是 32 位的量，由整个集簇共用的单个计数器分配。在大型或长期运行的数据库中，该计数器可能发生回卷。因此，除非采取了确保唯一性的措施，否则不应假定 OID 唯一。如果需要标识表中的行，强烈建议使用序列生成器。不过，只要采取一些额外的预防措施，也可以使用 OID：

- 对于使用 OID 标识行的每个表，都应在其 OID 列上创建唯一约束。有这样的唯一约束（或唯一索引）时，系统会确保不生成与已有行相同的 OID。（当然，这只有在表的行数小于 2^{32}（40 亿）时才有可能；实际上，表的大小最好远小于此值，否则性能可能受影响。）
- 绝不能假定 OID 在不同表之间也唯一；如果需要数据库范围的标识符，应将`tableoid`与行 OID 组合使用。
- 当然，相关表必须以`WITH OIDS`创建。从PostgreSQL 8.1 开始，默认值是`WITHOUT OIDS`。

事务 ID也是 32 位的量。在一个长期运行的数据库中，事务 ID 可能会回卷。只要采取适当的维护措施，这并不是致命问题，详见第 23 章。不过，从长期来看（超过十亿个事务）依赖事务 ID 的唯一性是不明智的。

命令标识符也是 32 位的量。这为单个事务中的SQL命令数设置了一个硬上限：2^{32}（40 亿）。在实践中，这个限制并不是问题 — 注意，这里限制的是SQL命令的数量，而不是处理的行数。另外，从PostgreSQL 8.3 开始，只有真正修改数据库内容的命令才会消耗命令标识符。

5.5. 修改表

创建表之后，如果你意识到自己犯了错误，或者应用需求发生了变化，可以删掉表再重新创建。但如果表中已经有数据，或者该表已被其他数据库对象引用（例如被外键约束引用），这样做就不方便了。因此，PostgreSQL提供了一组命令来修改现有表。注意，这在概念上不同于修改表中保存的数据：这里关注的是修改表的定义，也就是表的结构。

利用这些命令，我们可以：

- 增加列
- 移除列
- 增加约束
- 移除约束
- 修改默认值
- 修改列数据类型
- 重命名列
- 重命名表

所有这些动作都由ALTER TABLE命令执行，其参考页面中包含更详细的信息。

5.5.1. 增加列

要添加一列，可以使用这样的命令：

```
ALTER TABLE products ADD COLUMN description text;
```

新列最初会填入给定的默认值（如果没有指定DEFAULT子句，则填入空值）。

也可以同时为该列定义约束，使用常规语法：

```
ALTER TABLE products ADD COLUMN description text CHECK (description <>
 '');
```

实际上，凡是在CREATE TABLE中可用于列描述的选项，在这里都可以使用。不过要记住，默认值必须满足给定约束，否则ADD就会失败。另一种做法是先把新列正确填好，
再在之后添加约束（见下文）。

> **提示**
>
> 添加带默认值的列需要更新表中的每一行（以存储新列值）。不过，
> 如果没有指定默认值，PostgreSQL就可以避免物理更新。因此，
> 如果打算为该列填入的大多数值都不是默认值，最好先添加不带默认值的列，用UPDATE填入正确的值，再按下面所述添加所需的默认值。

5.5.2. 移除列

要删除一列，使用如下命令：

```
ALTER TABLE products DROP COLUMN description;
```

该列中的数据会消失，涉及该列的表约束也会被删除。不过，
如果该列被其他表的外键约束引用，PostgreSQL不会静默删除该约束。
你可以通过添加CASCADE来授权删除所有依赖该列的对象：

```
ALTER TABLE products DROP COLUMN description CASCADE;
```

关于这个操作背后的一般性机制请见第 5.12 节。

5.5.3. 增加约束

添加约束时使用表约束语法。例如：

```
ALTER TABLE products ADD CHECK (name <> '');
ALTER TABLE products ADD CONSTRAINT some_name UNIQUE (product_no);
ALTER TABLE products ADD FOREIGN KEY (product_group_id) REFERENCES
 product_groups;
```

非空约束不能写成表约束，添加时应使用以下语法：

```
ALTER TABLE products ALTER COLUMN product_no SET NOT NULL;
```

系统会立即检查约束，因此只有表中的数据满足约束，才能将其添加。

5.5.4. 移除约束

要删除约束，需要知道它的名称。如果曾为它指定名称，这很容易；
否则系统会分配一个自动生成的名称，需要将其查出来。psql的命令`\d tablename`对此很有帮助；其他接口也可能提供检查表细节的方式。然后执行以下命令：

```
ALTER TABLE products DROP CONSTRAINT some_name;
```

（如果处理的是自动生成的约束名，例如`$2`，别忘了加上双引号，使其成为有效标识符。）

和删除列一样，如果要删除某些其他对象所依赖的约束，也需要加上`CASCADE`。例如，
外键约束就依赖于被引用列上的唯一约束或主键约束。

除非空约束之外，所有类型的约束都可以用相同方式删除。要删除非空约束，使用：

```
ALTER TABLE products ALTER COLUMN product_no DROP NOT NULL;
```

（请记住，非空约束没有名称。）

5.5.5. 更改列的默认值

要为一个列设置一个新默认值，使用命令：

```
ALTER TABLE products ALTER COLUMN price SET DEFAULT 7.77;
```

注意这不会影响表中任何已经存在的行，它只是为未来的`INSERT`命令改变了默认值。

要移除任何默认值，使用：

```
ALTER TABLE products ALTER COLUMN price DROP DEFAULT;
```

这等同于将默认值设置为空值。相应地，试图删除一个未被定义的默认值并不会引发错误，
因为默认值已经被隐式地设置为空值。

5.5.6. 修改列的数据类型

为了将一个列转换为一种不同的数据类型，使用如下命令：

```
ALTER TABLE products ALTER COLUMN price TYPE numeric(10,2);
```

只有当列中的每一个项都能通过一个隐式类型转换为新的类型时该操作才能成功。
如果需要一种更复杂的转换，应该加上一个`USING`子句来指定应该如何把旧值转换为新值。

PostgreSQL将尝试把列的默认值转换为新类型，其他涉及到该列的任何约束也是一样。
但是这些转换可能失败或者产生奇特的结果。
因此最好在修改类型之前先删除该列上所有的约束，
然后在修改完类型后重新加上相应修改过的约束。

5.5.7. 重命名列

要重命名一个列：

```
ALTER TABLE products RENAME COLUMN product_no TO product_number;
```

5.5.8. 重命名表

要重命名一个表：

```
ALTER TABLE products RENAME TO items;
```

5.6. 权限

一旦一个对象被创建，它会被分配一个所有者。所有者通常是执行创建语句的角色。
对于大部分类型的对象，初始状态下只有所有者（或者超级用户）能够对该对象做任何事情。
为了允许其他角色使用它，必须分配权限。

有不同种类的权限：`SELECT`、`INSERT`、`UPDATE`、`DELETE`、`TRUNCATE`、`REFERENCES`、`TRIGGER`、`CREATE`、`CONNECT`、`TEMPORARY`、`EXECUTE`和`USAGE`。
适用于特定对象的权限取决于对象的类型（表、函数等）。关于
PostgreSQL支持的不同权限类型的完整信息，请参阅GRANT参考页。
后续的各节和各章还将展示这些权限是如何使用的。

修改或销毁对象的权利始终只属于该对象的拥有者。

可以使用适合对象类型的`ALTER`命令，将对象分配给新的拥有者，例如ALTER TABLE。超级用户始终可以这样做；普通角色只有同时是对象的当前拥有者（或拥有者角色的成员），
并且是新拥有者角色的成员，才能这样做。

分配权限使用`GRANT`命令。例如，如果`joe`是一个已有角色，并且`accounts`是一个已有表，
则可以用以下命令授予更新该表的权限：

```
GRANT UPDATE ON accounts TO joe;
```

用`ALL`代替某个具体权限，会授予与该对象类型相关的全部权限。

特殊的“用户”名`PUBLIC`可用于向系统中的每个用户授予权限。此外，当数据库用户很多时，
可以设置“组”角色来协助管理权限 – 详见第 20 章。

要撤销权限，使用名称贴切的`REVOKE`命令：

```
REVOKE ALL ON accounts FROM PUBLIC;
```

对象拥有者的特殊权限（即执行`DROP`，`GRANT`，`REVOKE`等操作的权利）
始终隐含在拥有者身份中，不能授予或撤销。不过，对象拥有者可以选择撤销自己的普通权限，
例如将表设为对自己和其他人都只读。

通常，只有对象的拥有者（或超级用户）可以授予或撤销对象上的权限。不过，
也可以在授予权限时“附带授权选项”，使接收者有权进一步将该权限授予其他人。

如果之后撤销了授权选项，所有从该接收者直接或经过授权链取得此权限的用户都会失去权限。详见GRANT和REVOKE参考页。

5.7. 模式

一个PostgreSQL数据库集簇中包含一个或更多命名的数据库。用户和用户组被整个集簇共享，但数据库之间不共享其他任何数据。
任何一条到服务器的客户端连接都只能访问单个数据库中的数据，
也就是连接请求中指定的那个数据库。

> **注意**
>
> 一个集簇的用户并不必拥有访问集簇中每一个数据库的权限。
> 用户名的共享意味着不可能在同一个集簇中出现重名的不同用户，
> 例如两个数据库中都有叫`joe`的用户。但系统可以被配置为只允许`joe`访问某些数据库。

一个数据库包含一个或多个命名的模式，模式中又包含表。模式还包含其他类型的命名对象，包括数据类型、函数和操作符。在同一个模式中，同一类型的两个对象不能有相同的名称。
此外，表、序列、索引、视图、物化视图和外部表共享同一个名字空间，
因此例如当它们位于同一模式中时，索引和表就必须具有不同的名称。
相同的对象名可以在不同模式中重复使用而不发生冲突；例如，`schema1`和`myschema`
都可以包含名为`mytable`的表。与数据库不同，模式并不是被严格隔离的：只要拥有相应权限，用户就可以访问其所连接数据库中任意模式里的对象。

下面是一些使用模式的原因：

- 允许多个用户使用一个数据库并且不会互相干扰。
- 将数据库对象组织成逻辑组以便更容易管理。
- 第三方应用的对象可以放在独立的模式中，这样它们就不会与其他对象的名称发生冲突。

模式类似于操作系统层的目录，但是模式不能嵌套。

5.7.1. 创建模式

要创建一个模式，可使用CREATE SCHEMA命令，并且给出选择的模式名称。例如：

```
CREATE SCHEMA myschema;
```

在一个模式中创建或访问对象，需要使用由模式名和表名构成的限定名，
模式名和表名之间以点号分隔：

```
schema.table
```

在任何需要一个表名的地方都可以这样用，
包括表修改命令和后续章节要讨论的数据访问命令（为了简洁我们在这里只谈到表，
但是这种方式对其他类型的命名对象同样有效，例如类型和函数）。

实际上，也可以使用更通用的语法

```
database.schema.table
```

，但目前这只是为了形式上符合 SQL 标准。如果写出数据库名称，它必须与当前连接的数据库相同。

因此，如果要在一个新模式中创建一个表，可用：

```
CREATE TABLE myschema.mytable (
 ...
);
```

要删除一个为空的模式（其中的所有对象已经被删除），可用：

```
DROP SCHEMA myschema;
```

要删除一个模式以及其中包含的所有对象，可用：

```
DROP SCHEMA myschema CASCADE;
```

有关于此的更一般的机制请参见第 5.12 节。

我们常常希望创建一个由其他人所拥有的模式（因为这是将用户动作限制在良定义的名字空间中的方法之一）。其语法是：

```
CREATE SCHEMA schemaname AUTHORIZATION username;
```

我们甚至可以省略模式名称，在此种情况下模式名称将会使用用户名，参见第 5.7.6 节。

以`pg_`开头的模式名被保留用于系统目的，所以不能被用户所创建。

5.7.2. 公共模式

在前面的小节中，我们创建表时都没有指定模式名。默认情况下，这些表（以及其他对象）会自动放入名为“public”的模式中。每个新数据库都包含这样一个模式。因此，下面两条命令是等效的：

```
CREATE TABLE products ( ... );
```

以及：

```
CREATE TABLE public.products ( ... );
```

5.7.3. 模式搜索路径

限定名写起来很繁琐，而且通常最好不要把某个特定模式名硬编码到应用中。因此，表通常通过非限定名来引用，也就是只写表名。系统会沿着一条搜索路径来决定该名称指的是哪个表；搜索路径就是要查找的一组模式列表。搜索路径中第一个匹配的表会被视为目标表。如果搜索路径中没有任何匹配，就会报错，即使数据库的其他模式中存在同名表也是如此。

搜索路径中的第一个模式被称为当前模式。除了是第一个被搜索的模式外，如果`CREATE TABLE`命令没有指定模式名，它将是新创建表所在的模式。

要显示当前搜索路径，可使用下面的命令：

```
SHOW search_path;
```

在默认设置下这将返回：

```
 search_path
--------------
 "$user",public
```

第一个元素说明一个和当前用户同名的模式会被搜索。如果不存在这个模式，该项将被忽略。
第二个元素指向我们已经见过的公共模式。

搜索路径中第一个存在的模式，是创建新对象时的默认位置。
这就是默认情况下对象会被创建在公共模式中的原因。
当对象在其他任何未限定模式的上下文中被引用时（表修改、数据修改或查询命令），
系统会沿搜索路径查找，直到找到匹配对象为止。因此，在默认配置中，
任何非限定访问仍然只能指向公共模式。

要把新模式放在搜索路径中，我们可以使用：

```
SET search_path TO myschema,public;
```

（我们在这里省略了`$user`，因为我们并不立即需要它）。
然后我们可以删除该表而无需使用模式进行限定：

```
DROP TABLE mytable;
```

同样，由于`myschema`是路径中的第一个元素，新对象默认也会创建在其中。

我们也可以这样写：

```
SET search_path TO myschema;
```

这样一来，在没有显式限定时，我们就不再能访问公共模式了。公共模式本身并无特殊之处，
只是默认存在而已；它同样可以被删除。

其他操作模式搜索路径的方法请见第 9.23 节。

搜索路径对于数据类型名称、函数名称和操作符名称的作用与表名一样。
数据类型和函数名称可以使用和表名完全相同的限定方式。
如果我们需要在一个表达式中写一个限定的操作符名称，我们必须写成一种特殊的形式：

```
OPERATOR(schema.operator)
```

这是为了避免句法歧义。例如：

```
SELECT 3 OPERATOR(pg_catalog.+) 4;
```

实际上我们通常都会依赖于搜索路径来查找操作符，因此没有必要去写如此“丑陋”的东西。

5.7.4. 模式和权限

默认情况下，用户无法访问他们不拥有的模式中的任何对象。要允许这样做，
模式的所有者必须授予该模式上的`USAGE`权限。要允许用户使用模式中的对象，
可能需要授予其他权限，适用于该对象。

也可以允许用户在他人的模式中创建对象。为此，需要授予该模式上的`CREATE`权限。注意，
默认情况下，每个人都具有`CREATE`和`USAGE`权限，它们作用于模式`public`。
这使所有能够连接到某个数据库的用户都可以在它的`public`模式中创建对象。
如果你不希望允许这样做，可以撤销该权限：

```
REVOKE CREATE ON SCHEMA public FROM PUBLIC;
```

（第一个“public”是模式名，第二个“public”表示“每个用户”。前一种用法是标识符，后一种是关键字，因此大小写不同；请回顾第 4.1.1 节。）

5.7.5. 系统目录模式

除public和用户创建的模式之外，每一个数据库还包括一个pg_catalog模式，它包含了系统表和所有内置的数据类型、函数以及操作符。pg_catalog总是搜索路径的一个有效部分。如果没有在路径中显式地包括该模式，它将在路径中的模式之前被搜索。这保证了内置的名称总是能被找到。然而，如果我们希望用用户定义的名称重载内置的名称，可以显式的将pg_catalog放在搜索路径的末尾。

在PostgreSQL 7.3 之前的版本中，以pg_开头的表名是被保留的。现在不再如此：如果你愿意，可以在任何非系统模式中创建这样名字的表。不过，最好还是继续避免使用这样的名称，以确保将来某个版本定义出与你的表同名的系统表时，你不会遇到冲突。系统表会继续遵循以pg_开头的约定，因此只要用户避免使用pg_前缀，它们就不会与未经限定的用户表名发生冲突。

5.7.6. 使用方式

模式可以用多种方式组织数据。下面几种使用方式是推荐的做法，并且很容易得到默认配置的支持：

- 如果你不创建任何模式，那么所有用户都将隐式地访问 public 模式。这模拟了根本没有模式可用的情况。这种设置主要推荐用于数据库中只有单个用户或少数几个相互合作的用户的场景。这种设置也允许从不支持模式的环境平滑过渡。

- 你可以为每个用户创建一个与其同名的模式。请记住，默认搜索路径以$user开头，它会解析为用户名。因此，如果每个用户都有单独的模式，他们默认访问的就是自己的模式。

  如果采用这种设置，你可能还想撤销对 public 模式的访问（或者干脆删除它），使用户真正被限制在他们自己的模式中。

- 如果要安装共享应用（所有人都要使用的表、第三方提供的附加函数等），可以把它们放进单独的模式中。记得授予适当的权限，以便其他用户能够访问它们。这样，用户既可以通过带模式名的限定名来引用这些附加对象，也可以按自己的需要把这些附加模式加入搜索路径。

5.7.7. 可移植性

在 SQL 标准中，不存在同一模式中的对象由不同用户拥有这一概念。此外，有些实现不允许创建与其所有者名称不同的模式。事实上，在那些只实现了标准中基本模式支持的数据库系统中，模式和用户这两个概念几乎是等价的。因此，很多用户认为限定名实际上就是 *username*.*tablename*。如果你为每个用户都创建一个独立模式，PostgreSQL的行为实际上也会是这样。

同样，SQL 标准中也没有public模式这一概念。为了尽可能符合标准，你不应使用（甚至可以移除）public模式。

当然，某些SQL数据库系统可能根本没有实现模式，或者提供（很可能是有限制地）允许跨数据库访问的命名空间。如果需要使用这样的系统，为了获得最好的可移植性，最好不要使用模式。

5.8. 继承

PostgreSQL实现了表继承，这对数据库设计者来说是一种有用的工具（SQL: 1999及其后的版本定义了一种类型继承特性，但和这里介绍的继承有很大的不同）。

让我们从一个示例开始：假设我们要为城市建立一个数据模型。每个州有很多城市，但只有一个首府。我们希望能够快速检索任意特定州的首府城市。这可以通过创建两个表来实现：一个用于州首府，另一个用于非首府城市。然而，当我们想要查询某个城市的数据，而不关心它是不是首府时，会发生什么？继承特性将有助于解决这个问题。我们可以将capitals表定义为继承自cities表：

```
CREATE TABLE cities (
    name            text,
    population      float,
    altitude       int     -- in feet
);

CREATE TABLE capitals (
    state           char(2)
) INHERITS (cities);
```

在这种情况下，capitals表继承了它的父表cities的所有列。州首府还有一个额外的列state用来表示它所属的州。

在PostgreSQL中，一个表可以从0个或者多个其他表继承，而对一个表的查询则可以引用一个表的所有行或者该表的所有行加上它所有的后代表。默认情况是后一种行为。例如，下面的查询将查找所有高度高于500尺的城市的名称，包括州首府：

```
SELECT name, altitude
    FROM cities
    WHERE altitude > 500;
```

对于来自PostgreSQL教程（见第 2.1 节）的示例数据，它将返回：

```
   name    | altitude
-----------+-----------
 Las Vegas |      2174
 Mariposa  |      1953
 Madison   |       845
```

另一方面，下面的查询将找到所有高度超过 500 尺且不是州首府的城市：

```
SELECT name, altitude
    FROM ONLY cities
    WHERE altitude > 500;

   name    | altitude
-----------+-----------
 Las Vegas |      2174
 Mariposa  |      1953
```

这里的ONLY关键词指示查询只被应用于cities上，而其他在继承层次中位于cities之下的其他表都不会被该查询涉及。很多我们已经讨论过的命令（如SELECT、UPDATE和DELETE）都支持ONLY关键词。

我们也可以在表名后写上一个*来显式地将后代表包括在查询范围内：

```
SELECT name, altitude
    FROM cities*
    WHERE altitude > 500;
```

写*并非必需，因为这种行为就是默认的（除非你更改了sql_inheritance配置选项的设置）。不过，写*或许有助于强调将会搜索额外的表。

在某些情况下，我们可能希望知道一个特定行来自于哪个表。每个表中的系统列tableoid可以告诉我们行来自于哪个表：

```
SELECT c.tableoid, c.name, c.altitude
FROM cities c
WHERE c.altitude > 500;
```

将会返回：

```
 tableoid |   name    | altitude
----------+-----------+-----------
   139793 | Las Vegas |      2174
   139793 | Mariposa  |      1953
   139798 | Madison   |       845
```

（如果重新生成这个结果，可能会得到不同的OID数字。）通过与pg_class进行连接可以看到实际的表名：

```
SELECT p.relname, c.name, c.altitude
FROM cities c, pg_class p
WHERE c.altitude > 500 AND c.tableoid = p.oid;
```

将会返回：

```
 relname  |   name    | altitude
----------+-----------+-----------
 cities   | Las Vegas |      2174
 cities   | Mariposa  |      1953
 capitals | Madison   |       845
```

继承不会自动地将来自INSERT或COPY命令的数据传播到继承层次中的其他表中。在我们的示例中，下面的INSERT语句将会失败：

```
INSERT INTO cities (name, population, altitude, state)
VALUES ('New York', NULL, NULL, 'NY');
```

我们也许会希望数据能以某种方式被路由到capitals表中，但这不会发生：INSERT总是向指定的表中插入。在某些情况下，可以通过使用一个规则（见第 37 章）将插入动作重定向。但是这对上面的情况并没有帮助，因为cities表根本就不包含state列，因而这个命令会在触发规则之前就被拒绝。

父表上的所有检查约束和非空约束都将自动被它的后代所继承。其他类型的约束（唯一、主键和外键约束）则不会被继承。

一个表可以从超过一个的父表继承，在这种情况下它拥有父表们所定义的列的并集。任何定义在子表上的列也会被加入到其中。如果在这个集合中出现重名列，那么这些列将被“合并”，这样在子表中只会有一个这样的列。重名列能被合并的前提是这些列必须具有相同的数据类型，否则会导致错误。合并后的列将带有它所来自的任何一个列定义中的全部检查约束的拷贝，并且只要其中任何一个被标记为非空，它就会被标记为非空。

表继承通常在创建子表时建立，即通过CREATE TABLE语句中的INHERITS子句。已经创建好的表也可以通过ALTER TABLE的INHERIT变体再增加一个新的父表关系。要这么做，新子表必须已经包含与父表同名且数据类型相同的列。子表还必须包含与父表相同的检查约束和检查表达式。类似地，也可以使用ALTER TABLE的NO INHERIT变体，从子表中移除一条继承链接。动态添加和移除继承链接可用于实现表分区（见第 5.9 节）。

一种创建将来要用作子表的新表的方法，是在CREATE TABLE中使用LIKE子句。这样会创建一个与源表具有相同列的新表。如果源表上定义了任何CHECK约束，可以使用LIKE的INCLUDING CONSTRAINTS选项，让新表也包含与父表相同的约束。

当有任何一个子表存在时，父表不能被删除。当子表的列或者检查约束继承于父表时，它们也不能被删除或修改。如果希望移除一个表和它的所有后代，一种简单的方法是使用CASCADE选项删除父表。

ALTER TABLE会把列的数据定义或检查约束上的任何变化沿着继承层次向下传播。同样，删除被其他表依赖的列只能使用CASCADE选项。对于同名列的合并与拒绝，ALTER TABLE遵循与CREATE TABLE相同的规则。

注意表访问权限是如何处理的。查询父表可以自动访问子表中的数据，而无需进一步的访问权限检查。这保持了数据（也）在父表中的外观。但是，直接访问子表并不会被自动允许，需要授予进一步的权限。

5.8.1. 注意事项

注意，并非所有 SQL 命令都能作用于继承层次。用于数据查询、数据修改或模式修改的命令（例如SELECT、UPDATE、DELETE、大多数ALTER TABLE变体，但不包括INSERT或ALTER TABLE ... RENAME）通常默认包含子表，并支持使用ONLY记法将其排除。用于数据库维护和调优的命令（例如REINDEX、VACUUM）通常只作用于独立的物理表，不支持沿继承层次递归。各条命令的具体行为都记录在相应参考页中（SQL 命令）。

继承特性的一个严重限制是，索引（包括唯一约束）和外键约束只作用于单个表，而不作用于其继承子表。对于外键约束的引用端和被引用端，这一点都成立。因此，沿用上面的示例：

- 如果我们把cities.name声明为UNIQUE或PRIMARY KEY，这并不能阻止capitals表中出现与cities中城市同名的行。而且这些重复行默认还会出现在针对cities的查询结果中。事实上，默认情况下capitals根本没有唯一约束，因此它可以包含多行同名记录。你当然可以给capitals添加唯一约束，但这仍无法阻止相对于cities的重复。

- 相似地，如果我们指定cities.name REFERENCES某个其他表，该约束不会自动地传播到capitals。在此种情况下，我们可以变通地在capitals上手工创建一个相同的REFERENCES约束。

- 如果让另一个表的某列REFERENCES cities(name)，那么该表可以包含城市名称，但不能包含首府名称。对于这种情况，并没有什么好的变通办法。

这些不足未来可能会在某个发行版中修复，但在此期间，在决定继承是否适合你的应用时，仍需要非常小心。

5.9. 分区

PostgreSQL支持基本的表分区。
本节介绍为什么以及如何把分区作为数据库设计的一部分来实现。

5.9.1. 概述

分区是指将逻辑上的一个大表拆分成较小的物理部分。分区可以带来以下好处：

- 在某些情况下查询性能能够显著提升，
  特别是当那些访问压力大的行在一个分区或者少数几个分区时。分区替代了索引的前导列，
  从而减小索引大小，使索引中被大量使用的部分更有可能容纳在内存中。
- 当查询或更新访问单个分区的很大一部分时，可以通过使用该分区的顺序扫描来提高性能，
  而不是使用索引，这将需要分散在整个表中的随机访问读取。
- 如果在分区设计中考虑到了这种使用模式，
  就可以通过添加或移除分区来完成批量加载和删除。`ALTER TABLE NO INHERIT`和`DROP TABLE`都比批量操作快得多。这些命令还完全避免了批量`DELETE`所导致的`VACUUM`开销。
- 很少使用的数据可以被迁移到便宜且较慢的存储介质上。

通常只有表本身非常大时，这些好处才值得考虑。表从多大开始受益于分区，取决于具体应用；
不过，一条经验法则是表的大小应超过数据库服务器的物理内存。

目前，PostgreSQL通过表继承来支持分区。每个分区都必须作为单个父表的子表创建。
父表本身通常为空；它的存在只是为了表示整个数据集。在尝试设置分区之前，
你应当先熟悉继承（见第 5.8 节）。

在PostgreSQL中可以实现以下几种分区形式：

范围分区

按照某个键列或一组列定义的“范围”对表进行分区，不同分区所分配的值范围互不重叠。
例如，可以按日期范围，或特定业务对象的标识符范围分区。

列表分区

通过明确列出每个分区包含哪些键值来对表进行分区。

5.9.2. 实现分区

要设置一个分区表，可以按以下步骤操作：

1. 创建所有分区都将继承的“主”表。

   这个表将不包含数据。不要在这个表上定义任何检查约束，
   除非想让它们等同地应用到所有分区上。在这个表上定义索引或者唯一约束也没有意义。
2. 创建若干“子”表，每个都从主表继承。通常，
   这些表不会在从主表继承的列集合之外增加任何列。

   我们将这些子表称为分区，尽管它们在各方面都是普通的PostgreSQL表。
3. 为各个分区表添加表约束，定义每个分区中允许的键值。

   典型示例如下：

   ```
   CHECK ( x = 1 )
   CHECK ( county IN ( 'Oxfordshire', 'Buckinghamshire',
    'Warwickshire' ))
   CHECK ( outletID >= 100 AND outletID < 200 )
   ```

   确保约束保证不同分区所允许的键值互不重叠。常见错误是设置如下范围约束：

```
CHECK ( outletID BETWEEN 100 AND 200 )
CHECK ( outletID BETWEEN 200 AND 300 )
```

这是错误的，因为无法明确键值 200 属于哪个分区。

注意，范围分区和列表分区在语法上没有区别；这些术语只是描述性的。

4. 对于每个分区，在键列上创建索引，以及其他所需的索引。（键索引不是严格必需的，但在大多数场景中是有帮助的。如果希望键值唯一，则应始终为每个分区创建唯一或主键约束。）
5. 可以选择定义一个触发器或规则，把插入到主表的数据重定向到适当的分区。
6. 确认constraint_exclusion配置参数在`postgresql.conf`中没有被禁用，否则查询将无法按预期得到优化。

例如，假设正在为一家大型冰淇淋公司构建数据库。该公司每天测量最高温度，并统计各个地区的冰淇淋销量。从概念上看，我们需要如下表：

```
CREATE TABLE measurement (
    city_id         int not null,
    logdate         date not null,
    peaktemp        int,
    unitsales       int
);
```

我们知道，大多数查询只会访问最近一周、一个月或一个季度的数据，因为该表主要用于为管理层生成在线报表。为减少需要存储的旧数据量，决定只保留最近 3 年的数据，并在每月月初删除最早一个月的数据。

在这种情况下，可以利用分区来帮助满足测量表的各种不同需求。按照上面概述的步骤，可以这样设置分区：

1. 主表就是`measurement`表，完全按上面的方式声明。
2. 然后，为每个活动月份创建一个分区：

```
CREATE TABLE measurement_y2006m02 ( ) INHERITS (measurement);
CREATE TABLE measurement_y2006m03 ( ) INHERITS (measurement);
...
CREATE TABLE measurement_y2007m11 ( ) INHERITS (measurement);
CREATE TABLE measurement_y2007m12 ( ) INHERITS (measurement);
CREATE TABLE measurement_y2008m01 ( ) INHERITS (measurement);
```

每个分区本身都是完整的表，但它们从`measurement`表继承其定义。

这解决了我们的一个问题：删除旧数据。每个月，我们只需对最旧的子表执行`DROP TABLE`，并为新月份的数据创建一个新的子表。

3. 我们必须提供互不重叠的表约束。与其像上面那样只创建分区表，表创建脚本其实应该是：

```
 CREATE TABLE measurement_y2006m02 (
     CHECK ( logdate >= DATE '2006-02-01' AND logdate < DATE '2006
-03-01' )
 ) INHERITS (measurement);
 CREATE TABLE measurement_y2006m03 (
     CHECK ( logdate >= DATE '2006-03-01' AND logdate < DATE '2006
-04-01' )
 ) INHERITS (measurement);
 ...
```

```
 CREATE TABLE measurement_y2007m11 (
     CHECK ( logdate >= DATE '2007-11-01' AND logdate < DATE '2007
-12-01' )
 ) INHERITS (measurement);
 CREATE TABLE measurement_y2007m12 (
     CHECK ( logdate >= DATE '2007-12-01' AND logdate < DATE '2008
-01-01' )
 ) INHERITS (measurement);
 CREATE TABLE measurement_y2008m01 (
     CHECK ( logdate >= DATE '2008-01-01' AND logdate < DATE '2008
-02-01' )
 ) INHERITS (measurement);
```

4. 我们可能还需要在键列上创建索引：

```
 CREATE INDEX measurement_y2006m02_logdate ON measurement_y2006m02
 (logdate);
 CREATE INDEX measurement_y2006m03_logdate ON measurement_y2006m03
 (logdate);
...
 CREATE INDEX measurement_y2007m11_logdate ON measurement_y2007m11
 (logdate);
 CREATE INDEX measurement_y2007m12_logdate ON measurement_y2007m12
 (logdate);
 CREATE INDEX measurement_y2008m01_logdate ON measurement_y2008m01
 (logdate);
```

   我们此时选择不添加更多索引。
5. 我们希望应用程序能够执行`INSERT INTO measurement ...`，并使数据被重定向到适当的分区表。可以通过在主表上附加合适的触发器函数来实现这一点。
   如果数据只添加到最新的分区，可以使用非常简单的触发器函数：

```
 CREATE OR REPLACE FUNCTION measurement_insert_trigger()
 RETURNS TRIGGER AS $$
 BEGIN
     INSERT INTO measurement_y2008m01 VALUES (NEW.*);
     RETURN NULL;
 END;
 $$
 LANGUAGE plpgsql;
```

   创建函数后，创建一个调用该触发器函数的触发器：

```
CREATE TRIGGER insert_measurement_trigger
    BEFORE INSERT ON measurement
    FOR EACH ROW EXECUTE PROCEDURE measurement_insert_trigger();
```

   必须每月重新定义触发器函数，使其始终指向当前分区。不过，触发器定义无需更新。

   我们可能希望插入数据时，由服务器自动定位应添加该行的分区。
   这可以通过更复杂的触发器函数实现，例如：

```
CREATE OR REPLACE FUNCTION measurement_insert_trigger()
RETURNS TRIGGER AS $$
BEGIN
    IF ( NEW.logdate >= DATE '2006-02-01' AND
```

```
         NEW.logdate < DATE '2006-03-01' ) THEN
        INSERT INTO measurement_y2006m02 VALUES (NEW.*);
    ELSIF ( NEW.logdate >= DATE '2006-03-01' AND
            NEW.logdate < DATE '2006-04-01' ) THEN
        INSERT INTO measurement_y2006m03 VALUES (NEW.*);
    ...
    ELSIF ( NEW.logdate >= DATE '2008-01-01' AND
            NEW.logdate < DATE '2008-02-01' ) THEN
        INSERT INTO measurement_y2008m01 VALUES (NEW.*);
    ELSE
        RAISE EXCEPTION 'Date out of range.  Fix the measurement_
insert_trigger() function!';
    END IF;
    RETURN NULL;
END;
$$
LANGUAGE plpgsql;
```

触发器定义与之前相同。注意，每个IF测试必须与其子表的CHECK约束完全一致。

当该函数比单月形式更加复杂时，并不需要频繁地更新它，
因为可以在需要的时候提前加入分支。

> **注意**
>
> 在实践中，如果大部分插入都会进入最新的分区，最好先检查它。为了简洁，
> 我们为触发器的检查采用了和本例中其他部分一致的顺序。

如我们所见，一个复杂的分区方案可能需要大量的DDL。在上面的示例中，
我们可能每个月创建一个新分区，因此编写一个脚本来自动生成所需要的DDL可能会更好。

5.9.3. 分区管理

通常，最初定义表时建立的分区集合并不会保持不变。常见的需求是删除旧数据分区，
并定期为新数据添加新分区。分区最重要的优势之一，恰恰在于它允许通过操纵分区结构，
而不是物理地移动大量数据，来近乎瞬时地完成这种原本非常痛苦的任务。

移除旧数据最简单的选择，是直接删除不再需要的分区：

```
DROP TABLE measurement_y2006m02;
```

由于不必逐条删除每条记录，这可以非常快速地删除数百万条记录。

另一种通常更可取的选择，是将该分区从分区表中移除，但保留它作为独立表的可访问性：

```
ALTER TABLE measurement_y2006m02 NO INHERIT measurement;
```

这样在删除数据之前，还可以对其执行进一步的操作。例如，这常常是使用COPY、pg_
dump或类似工具备份数据的好时机。也可能是把数据聚合成更小格式、
执行其他数据处理或运行报告的好时机。

类似地，我们可以添加一个新分区来处理新数据。可以像上面创建初始分区那样，
在分区表中创建一个空分区：

```
CREATE TABLE measurement_y2008m02 (
    CHECK ( logdate >= DATE '2008-02-01' AND logdate < DATE '2008-03
-01' )
) INHERITS (measurement);
```

作为替代，有时更方便的做法是在分区结构之外创建新表，之后再把它变成正式的分区。这允许数据在出现在分区表中之前先被载入、检查和转换：

```
CREATE TABLE measurement_y2008m02
  (LIKE measurement INCLUDING DEFAULTS INCLUDING CONSTRAINTS);
ALTER TABLE measurement_y2008m02 ADD CONSTRAINT y2008m02
   CHECK ( logdate >= DATE '2008-02-01' AND logdate < DATE '2008-03
-01' );
\copy measurement_y2008m02 from 'measurement_y2008m02'
-- possibly some other data preparation work
ALTER TABLE measurement_y2008m02 INHERIT measurement;
```

5.9.4. 分区和约束排除

约束排除是一种查询优化技术，可提高按上述方式定义的分区表的性能。例如：

```
SET constraint_exclusion = on;
SELECT count(*) FROM measurement WHERE logdate >= DATE '2008-01-01';
```

如果没有约束排除，上述查询会扫描`measurement`表的每个分区。启用约束排除后，规划器会检查每个分区的约束，并尝试证明该分区不需要扫描，因为它不可能包含满足查询`WHERE`子句的行。如果规划器能够证明这一点，就会将该分区排除在查询计划之外。

可以使用`EXPLAIN`命令，显示启用`constraint_exclusion`时的计划与禁用时的计划之间的差异。对于这种表结构，典型的未优化计划如下：

```
SET constraint_exclusion = off;
EXPLAIN SELECT count(*) FROM measurement WHERE logdate >= DATE '2008
-01-01';

                                           QUERY PLAN
-----------------------------------------------------------------------
-------------------------
 Aggregate  (cost=158.66..158.68 rows=1 width=0)
   ->  Append  (cost=0.00..151.88 rows=2715 width=0)
         ->  Seq Scan on measurement  (cost=0.00..30.38 rows=543
 width=0)
               Filter: (logdate >= '2008-01-01'::date)
         ->  Seq Scan on measurement_y2006m02 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2008-01-01'::date)
         ->  Seq Scan on measurement_y2006m03 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2008-01-01'::date)
...
         ->  Seq Scan on measurement_y2007m12 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2008-01-01'::date)
         ->  Seq Scan on measurement_y2008m01 measurement  (cost=0.00.
.30.38 rows=543 width=0)
```

```
               Filter: (logdate >= '2008-01-01'::date)
```

某些或全部分区可能使用索引扫描，而不是全表顺序扫描，但这里的重点是，为回答该查询，根本不需要扫描较旧的分区。启用约束排除后，可以得到代价明显更低、结果却相同的计划：

```
SET constraint_exclusion = on;
EXPLAIN SELECT count(*) FROM measurement WHERE logdate >= DATE '2008
-01-01';
                                           QUERY PLAN
-----------------------------------------------------------------------
-------------------------
 Aggregate  (cost=63.47..63.48 rows=1 width=0)
   ->  Append  (cost=0.00..60.75 rows=1086 width=0)
         ->  Seq Scan on measurement  (cost=0.00..30.38 rows=543
 width=0)
               Filter: (logdate >= '2008-01-01'::date)
         ->  Seq Scan on measurement_y2008m01 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2008-01-01'::date)
```

注意，约束排除只由`CHECK`约束驱动，与是否存在索引无关。因此，不必在键列上定义索引。是否需要为某个分区创建索引，取决于预期查询扫描该分区时通常会扫描其中的大部分，还是仅扫描一小部分。后一种情况下索引会有帮助，前一种情况下则不会。

constraint_exclusion的默认（也是推荐的）设置不是`on`也不是`off`，而是一种被称为`partition`的中间设置，这会导致该技术仅被应用于可能访问分区表的查询。`on`设置导致规划器检查所有查询中的`CHECK`约束，甚至是那些不太可能受益的简单查询。

5.9.5. 备选分区方法

将插入重定向到适当分区表的另一种方法，是在根表上设置规则来代替触发器。例如：

```
CREATE RULE measurement_insert_y2006m02 AS
ON INSERT TO measurement WHERE
    ( logdate >= DATE '2006-02-01' AND logdate < DATE '2006-03-01' )
DO INSTEAD
    INSERT INTO measurement_y2006m02 VALUES (NEW.*);
...
CREATE RULE measurement_insert_y2008m01 AS
ON INSERT TO measurement WHERE
    ( logdate >= DATE '2008-01-01' AND logdate < DATE '2008-02-01' )
DO INSTEAD
    INSERT INTO measurement_y2008m01 VALUES (NEW.*);
```

规则的开销明显高于触发器，但这种开销每个查询只支付一次，而不是每行一次，因此在批量插入时，这种方法可能有优势。不过，在大多数情况下，触发器方法的性能更好。

注意`COPY`会忽略规则。如果想要使用`COPY`插入数据，则需要拷贝到正确的分区表而不是根表中。`COPY`会引发触发器，因此在使用触发器方法时可以正常使用它。

规则方法的另一个缺点是，如果规则集合无法覆盖插入日期，则没有简单的方法能够强制产生错误，数据将会无声无息地进入到根表中。

也可以使用`UNION ALL`视图来安排分区，而不是使用表继承。例如：

```
CREATE VIEW measurement AS
          SELECT * FROM measurement_y2006m02
UNION ALL SELECT * FROM measurement_y2006m03
...
UNION ALL SELECT * FROM measurement_y2007m11
UNION ALL SELECT * FROM measurement_y2007m12
UNION ALL SELECT * FROM measurement_y2008m01;
```

但是，重新创建视图的需要给数据集分区的添加和删除增加了一个额外步骤。实践中，与使用继承相比，这种方法没有什么可取之处。

5.9.6. 注意事项

分区表有以下注意事项：

- 没有自动方式验证所有CHECK约束是否互斥。与手工逐个编写相比，通过代码生成分区并创建和/或修改相关对象更安全。
- 这里展示的方案假定行的分区键列值永不改变，或者至少不会改变到必须把该行移入另一个分区的程度。由于CHECK约束的存在，试图那样做的UPDATE将会失败。如果需要处理这种情况，可以在分区表上放置合适的更新触发器，但这会让整个结构的管理复杂得多。
- 如果手动执行VACUUM或ANALYZE命令，不要忘记需要在每个分区上分别运行。例如，以下命令：

  ```
  ANALYZE measurement;
  ```

  只会处理根表。

约束排除有以下注意事项：

- 只有查询的WHERE子句包含常量时，约束排除才能有效果。参数化查询不会被优化，因为规划器无法知道参数值在运行时会选择哪些分区。出于同样的原因，必须避免诸如CURRENT_DATE这样的“稳定”函数。
- 保持分区约束尽量简单，否则规划器可能无法证明哪些分区不需要访问。如前面的示例所示，对列表分区使用简单的等值条件，对范围分区使用简单的范围测试。一条很好的经验法则是：分区约束应只包含分区列与常量之间使用 B-树可索引操作符的比较。
- 约束排除期间会检查根表的所有分区上的全部约束，因此大量分区很可能显著增加查询规划时间。使用这些技术进行分区，大概在不超过一百个分区时效果较好；不要尝试使用成千上万个分区。

5.10. 外部数据

PostgreSQL实现了部分的SQL/MED规定，允许我们使用普通SQL查询来访问位于PostgreSQL之外的数据。这种数据被称为外部数据（注意这种用法不要和外键混淆，后者是数据库中的一种约束）。

外部数据可以借助外部数据包装器来访问。外部数据包装器是一个库，它可以与外部数据源通信，并隐藏连接数据源以及从中获取数据的细节。在contrib模块中有一个外部数据包装器，它可以读取服务器上存储的普通数据文件。其他类型的外部数据包装器可以在第三方产品中找到。如果这些现有的外部数据包装器都不能满足需要，也可以自行编写，参见第 50 章。

要访问外部数据，我们需要创建一个外部服务器对象，
它根据某个特定外部数据包装器所使用的一组选项，定义如何连接到某个特定的外部数据源。
然后还需要创建一个或多个外部表，用于定义远程数据的结构。
外部表在查询中可以像普通表一样使用，但它在 PostgreSQL 服务器中并不存储数据。每次使用外部表时，PostgreSQL都会要求外部数据包装器从外部数据源获取数据。

访问远程数据可能需要在外部数据源进行认证。这些信息可以通过一个用户映射提供，
它会基于当前的PostgreSQL角色提供附加选项。

目前，外部表是只读的。这一限制可能会在将来的版本中解决。

5.11. 其他数据库对象

表是关系数据库结构中的核心对象，因为它们承载着数据。
但它们并不是数据库中唯一存在的对象。还可以创建许多其他种类的对象，
以便让数据的使用和管理更加高效或方便。本章不会讨论这些对象，但这里给出一个列表，
让你了解有哪些可能性：

- 视图
- 函数和操作符
- 数据类型和域
- 触发器和重写规则

这些主题的详细信息请见第 V 部分。

5.12. 依赖跟踪

当我们创建一个涉及到很多具有外键约束、视图、触发器、函数等的表的复杂数据库结构时，
我们隐式地创建了一张对象之间的依赖关系网。例如，
具有一个外键约束的表依赖于它所引用的表。

为了保证整个数据库结构的完整性，PostgreSQL确保我们无法删除仍然被其他对象依赖的对象。
例如，尝试删除第 5.3.5 节中的产品表会导致一个如下的错误消息，因为有订单表依赖于产品表：

```
DROP TABLE products;

ERROR:  cannot drop table products because other objects depend on it
DETAIL:  constraint orders_product_no_fkey on table orders depends on
 table products
HINT:  Use DROP ... CASCADE to drop the dependent objects too.
```

该错误消息包含了一个有用的提示：如果我们不想一个一个去删除所有的依赖对象，
我们可以执行：

```
DROP TABLE products CASCADE;
```

这样所有的依赖对象将被移除。在这种情况下，订单表不会被移除，
但是它的外键约束会被移除。（如果希望检查`DROP ... CASCADE`会干什么，运行不带`CASCADE`的`DROP`并阅读`DETAIL`输出。）

PostgreSQL中所有DROP命令都支持指定CASCADE。当然，
可能出现的依赖关系形态会随着对象类型不同而变化。你也可以写RESTRICT来代替CASCADE，
从而得到默认行为，也就是阻止删除任何被其他对象依赖的对象。

> **注意**
>
> 根据 SQL 标准，DROP命令必须指定RESTRICT或CASCADE。实际上，没有数据库系统强制执行这条规则，但不同系统的默认行为是RESTRICT还是CASCADE，各有不同。

对于用户定义函数，PostgreSQL会跟踪与函数外部可见属性有关的依赖，例如参数和结果类型，
但不会跟踪那些只有检查函数体才能得知的依赖。例如，考虑以下情况：

```
CREATE TYPE rainbow AS ENUM ('red', 'orange', 'yellow',
                             'green', 'blue', 'purple');

CREATE TABLE my_colors (color rainbow, note text);

CREATE FUNCTION get_color_note (rainbow) RETURNS text AS
  'SELECT note FROM my_colors WHERE color = $1'
  LANGUAGE SQL;
```

（关于 SQL 语言函数的说明，请参阅第 35.4 节。）PostgreSQL会知道get_color_note函数依赖于rainbow类型：删除该类型将迫使系统删除函数，因为它的参数类型将不再有定义。但是，
PostgreSQL不会认为get_color_note依赖于my_colors表，
因此不会在删除该表时删除函数。这种做法有缺点，也有好处。即使表不存在，
函数在某种意义上仍然有效，尽管执行它会报错；创建一个同名新表后，函数就可以重新工作。

第 6 章 数据操纵

前一章讨论了如何创建用于保存数据的表和其他结构。现在该把数据填入表中了。
本章介绍如何插入、更新和删除表数据。
下一章则终于会讲到如何从数据库中取回那些久寻不得的数据。

6.1. 插入数据

创建表时，其中不包含任何数据。在数据库真正派上用场之前，首先要做的就是插入数据。
从概念上说，数据一次插入一行。当然也可以插入多行，但不能插入少于一行的数据。
即使只知道部分列值，也必须创建完整的一行。

要创建新行，可使用INSERT命令。该命令需要提供表名和列值。例如，考虑第 5 章中的 products 表：

```
CREATE TABLE products (
    product_no integer,
    name text,
    price numeric
);
```

插入一行的示例命令如下：

```
INSERT INTO products VALUES (1, 'Cheese', 9.99);
```

数据值按列在表中出现的顺序列出，并以逗号分隔。通常，数据值会是字面量（常量），
但也允许使用标量表达式。

上述语法的缺点在于你必须知道表中列的顺序。为了避免这一点，你也可以显式列出列。例如，
下面两条命令的效果都与上面的命令相同：

```
INSERT INTO products (product_no, name, price) VALUES (1, 'Cheese', 9.
99);
INSERT INTO products (name, price, product_no) VALUES ('Cheese', 9.99,
 1);
```

许多用户认为，总是列出列名是一种好习惯。

如果你没有所有列的值，可以省略其中一些列。在这种情况下，这些列将填入各自的默认值。
例如：

```
INSERT INTO products (product_no, name) VALUES (1, 'Cheese');
INSERT INTO products VALUES (1, 'Cheese');
```

第二种形式是PostgreSQL 的一个扩展。它从左到右用给出的值填充各列，其余列则取默认值。

为求清晰，你也可以显式请求默认值，既可以针对单个列，也可以针对整行：

```
INSERT INTO products (product_no, name, price) VALUES (1, 'Cheese',
 DEFAULT);
INSERT INTO products DEFAULT VALUES;
```

你可以在一条命令中插入多行：

```
INSERT INTO products (product_no, name, price) VALUES
```

```
(1, 'Cheese', 9.99),
(2, 'Bread', 1.99),
(3, 'Milk', 2.99);
```

提示

当需要一次插入大量数据时，可以考虑使用COPY命令。它不像INSERT命令那样灵活，但效率更高。关于如何改善批量装载性能的更多信息，请参阅第 14.4 节。

6.2. 更新数据

修改数据库中已有的数据称为更新。你可以更新单独的行、表中的所有行，或者所有行中的一个子集。每一列都可以单独更新；其他列不受影响。

要更新现有的行，可使用UPDATE命令。这需要三部分信息：

1. 要更新的表名和列名
2. 列的新值
3. 要更新哪一行（或哪些行）

回想一下第 5 章中的内容，SQL 通常并不为行提供唯一标识符。因此，并不总能直接指定要更新哪一行。取而代之的是，你要指定一行必须满足哪些条件才会被更新。只有当表中存在主键时（不论你是否声明了它），才能通过选择一个匹配主键的条件来可靠地定位单独的行。图形化数据库访问工具正是依赖这一事实来允许你逐行更新。

例如，这条命令把所有价格为 5 的产品的价格更新为 10：

```
UPDATE products SET price = 10 WHERE price = 5;
```

这可能更新零行、一行或多行。尝试执行一个匹配不到任何行的更新并不算错误。

让我们仔细看看这条命令。首先是关键字 `UPDATE`，后面跟着表名。和平常一样，表名可以用模式限定；否则就在搜索路径中查找。接着是关键字 `SET`，后面跟着列名、一个等号以及新的列值。新列值可以是任意标量表达式，而不仅仅是常量。例如，如果你想把所有产品的价格提高 10%，可以使用：

```
UPDATE products SET price = price * 1.10;
```

如你所见，用于新值的表达式也可以引用该行中的现有值。我们还省略了`WHERE`子句。如果省略它，就意味着更新表中的所有行。如果给出了它，则只有匹配 `WHERE` 条件的那些行才会被更新。请注意，`SET` 子句中的等号表示赋值，而 `WHERE` 子句中的等号表示比较，但这不会造成歧义。当然，`WHERE` 条件不必是等值测试。还有许多其他操作符可用（见第 9 章）。不过，该表达式必须计算出一个布尔结果。

你可以在一条`UPDATE`命令中更新多个列，只需在 `SET`子句中列出多个赋值。例如：

```
UPDATE mytable SET a = 5, b = 3, c = 1 WHERE a > 0;
```

6.3. 删除数据

到目前为止，我们已经解释了如何向表中添加数据以及如何修改数据。接下来要讨论的是如何删除不再需要的数据。正如添加数据只能按整行进行一样，

从表中删除数据也只能按整行进行。前一节已经说明，SQL 不提供直接寻址单独行的方法。因此，只能通过指定待删除行必须满足的条件来删除它们。如果表中有主键，
那么你可以指定精确的一行；但你也可以删除匹配某个条件的一组行，
或者一次删除表中的所有行。

可以使用DELETE命令删除行；其语法与UPDATE命令非常相似。例如，要删除 products 表中所有价格为 10 的行，可以使用：

```
DELETE FROM products WHERE price = 10;
```

如果你只是写：

```
DELETE FROM products;
```

那么表中的所有行都会被删除！程序员须当心。

第 7 章 查询

前面的章节解释了如何创建表、如何用数据填充它们，以及如何操纵这些数据。
现在我们终于可以讨论如何从数据库中检索数据了。

7.1. 概述

从数据库检索数据的过程或用于检索数据的命令称为查询。在 SQL 中，SELECT命令用于指定查询。SELECT命令的一般语法为：

```
[WITH with_queries] SELECT select_list FROM table_expression [sort_specification]
```

以下各节将详细描述选择列表、表表达式和排序说明。WITH查询是一项高级特性，
因此放在最后介绍。

一种简单查询的形式如下：

```
SELECT * FROM table1;
```

假设存在一个名为table1的表，则这条命令会从table1中检索所有行以及所有列。
（具体的检索方式取决于客户端应用。例如，psql程序会在屏幕上显示一个 ASCII 形式的表格，
而客户端库则会提供从查询结果中提取单个值的函数。）选择列表说明*
表示表表达式所提供的全部列。选择列表也可以只选择部分可用列，或者利用这些列进行计算。
例如，如果table1有名为a、b和c的列（还可能有其他列），那么可以写出下面的查询：

```
SELECT a, b + c FROM table1;
```

（假设b和c都是数值数据类型）。有关更多细节请参见第 7.3 节。

FROM table1是一种简单的表表达式：它只读取一个表。通常，表表达式可以是由基本表、连接
和子查询组成的复杂结构。不过，你也可以完全省略表表达式，把SELECT
命令当成计算器来使用：

```
SELECT 3 * 4;
```

如果选择列表中的表达式会返回变化的结果，这种用法就更有用了。例如，
你可以这样调用函数：

```
SELECT random();
```

7.2. 表表达式

表表达式用于计算出一个表。表表达式包含一个FROM子句，后面可以根据需要跟上WHERE、
GROUP BY和HAVING子句。最简单的表表达式只是引用磁盘上的一个表，即所谓的基本表；但也
可以使用更复杂的表达式，以多种方式修改或组合基本表。

表表达式中可选的WHERE、GROUP BY和HAVING子句指定了一个连续转换的流水线，这些转换作
用于由FROM子句派生出的表。所有这些转换都会生成一个虚拟表，
该表提供的各行会传递给选择列表，以计算查询的输出行。

7.2.1. FROM子句

此FROM Clause从逗号分隔的表引用列表所给出的一个或多个其他表推导出一个表。

```
FROM table_reference [, table_reference [, ...]]
```

表引用可以是表名（可能带模式限定），也可以是子查询、JOIN结构等推导表，
或它们的复杂组合。如果FROM子句中列出了多个表引用，
这些表会进行交叉连接（即形成各表行的笛卡尔积，见下文）。FROM
列表的结果是一个中间虚拟表，随后可以用WHERE，GROUP BY和HAVING子句对它进行变换，最终得到整个表表达式的结果。

当表引用命名的是表继承层次中的父表时，除非在表名前加上关键字ONLY，
否则该表引用不仅会产生该表中的行，还会产生其所有后代表中的行。不过，
这种引用只会产生该命名表中出现的列 – 子表中新增的列会被忽略。

也可以不在表名前写ONLY，而是在表名后面写上*，显式指定要包含后代表。
由于这种行为是默认行为（除非你更改了sql_inheritance配置选项的设置），写*并无必要。
不过，写*或许有助于强调将会搜索额外的表。

7.2.1.1. 连接表

一个连接表是根据特定的连接类型的规则从两个其它表（真实表或生成表）中派生的表。
目前支持内连接、外连接和交叉连接。一个连接表的一般语法是：

```
T1 join_type T2 [ join_condition ]
```

所有类型的连接都可以被链在一起或者嵌套：*T1*和*T2*都可以是连接表。在JOIN
子句周围可以使用圆括号来控制连接顺序。如果不使用圆括号，JOIN子句会从左至右嵌套。

连接类型

交叉连接

```
T1 CROSS JOIN T2
```

对来自于*T1*和*T2*的行的每一种可能的组合（即笛卡尔积），
连接表将包含这样一行：它由所有*T1*里面的列后面跟着所有*T2*里面的列构成。
如果两个表分别有 N 和 M 行，连接表将有 N * M 行。

FROM *T1* CROSS JOIN *T2*等效于FROM *T1* INNER JOIN *T2* ON TRUE（见下文）。
它也等效于FROM *T1*,*T2*。

> **注意**
>
> 当出现两个以上的表时，后一种等价关系并不严格成立，因为JOIN的绑定强于逗号。
> 例如FROM *T1* CROSS JOIN *T2* INNER JOIN *T3* ON *condition*和FROM *T1*,*T2* INNER JOIN *T3* ON *condition*并不完全相同，因为第一种情况中的*condition*
> 可以引用*T1*，而第二种情况中则不行。

限定连接

```
T1 { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN T2
 ON boolean_expression
T1 { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN T2 USING (
 join column list )
```

T1 NATURAL { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN *T2*

INNER和OUTER在所有形式中都是可选的。INNER是默认值；LEFT、RIGHT和FULL表示外连接。

连接条件在ON或USING子句中指定，或者用关键字NATURAL隐含地指定。
连接条件决定来自两个源表中的哪些行是“匹配”的，这些我们将在后文详细解释。

限定连接有以下几种类型：

INNER JOIN

对于 T1 的每一行 R1，生成的连接表都有一行对应 T2 中的每一个满足和 R1 的连接条件的行。

LEFT OUTER JOIN

首先，执行一次内连接。然后，为 T1 中每一个无法在连接条件上匹配 T2 里任何一行的行返回一个连接行，该连接行中 T2 的列用空值补齐。因此，生成的连接表里为来自 T1 的每一行都至少包含一行。

RIGHT OUTER JOIN

首先，执行一次内连接。然后，为 T2 中每一个无法在连接条件上匹配 T1 里任何一行的行返回一个连接行，该连接行中 T1 的列用空值补齐。因此，生成的连接表里为来自 T2 的每一行都至少包含一行。

FULL OUTER JOIN

首先，执行一次内连接。然后，为 T1 中每一个无法在连接条件上匹配 T2 里任何一行的行返回一个连接行，该连接行中 T2 的列用空值补齐。同样，为 T2 中每一个无法在连接条件上匹配 T1 里任何一行的行返回一个连接行，该连接行中 T1 的列用空值补齐。

ON子句是最常见的连接条件的形式：它接收一个和WHERE子句里用的一样的布尔值表达式。
如果两个分别来自*T1*和*T2*的行在ON表达式上运算的结果为真，那么它们就算是匹配的行。

USING是个缩写符号，它允许你利用特殊的情况：连接的两端都具有相同的连接列名。
它接受共享列名的一个逗号分隔列表，
并且为其中每一个共享列构造一个包含等值比较的连接条件。例如用USING (a, b)连接*T1*和*T2*会产生连接条件ON *T1*.a = *T2*.a AND *T1*.b = *T2*.b。

更进一步，JOIN USING的输出会废除冗余列：不需要把匹配上的列都打印出来，因为它们必须具有相等的值。不过JOIN ON会先产生来自*T1*的所有列，后面跟上所有来自*T2*的列；而JOIN USING会先为列出的每一个列对产生一个输出列，然后先跟上来自*T1*的剩余列，最后跟上来自*T2*的剩余列。

最后，NATURAL是USING的一种缩写形式：它会形成一个USING列表，其中包含两个输入表中共同出现的所有列名。和USING一样，这些列在输出表中只出现一次。
如果没有公共列名，NATURAL的行为就与CROSS JOIN相同。

注意

对于被连接的关系发生列变化的情况，USING相当安全，
因为只有列出的列才会被合并。NATURAL的风险则大得多，

> 因为只要任一关系的模式变更导致出现新的同名列，
> 连接就会把这个新列也纳入合并。

综合起来看，假设我们有表t1：

```
 num | name
-----+------
   1 | a
   2 | b
   3 | c
```

和t2：

```
 num | value
-----+-------
   1 | xxx
   3 | yyy
   5 | zzz
```

然后我们用不同的连接方式可以获得各种结果：

```
=> SELECT * FROM t1 CROSS JOIN t2;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   1 | a    |   3 | yyy
   1 | a    |   5 | zzz
   2 | b    |   1 | xxx
   2 | b    |   3 | yyy
   2 | b    |   5 | zzz
   3 | c    |   1 | xxx
   3 | c    |   3 | yyy
   3 | c    |   5 | zzz
(9 rows)

=> SELECT * FROM t1 INNER JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   3 | c    |   3 | yyy
(2 rows)

=> SELECT * FROM t1 INNER JOIN t2 USING (num);
 num | name | value
-----+------+-------
   1 | a    | xxx
   3 | c    | yyy
(2 rows)

=> SELECT * FROM t1 NATURAL INNER JOIN t2;
 num | name | value
-----+------+-------
   1 | a    | xxx
```

```
   3 | c    | yyy
(2 rows)

=> SELECT * FROM t1 LEFT JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |   3 | yyy
(3 rows)

=> SELECT * FROM t1 LEFT JOIN t2 USING (num);
 num | name | value
-----+------+-------
   1 | a    | xxx
   2 | b    |
   3 | c    | yyy
(3 rows)

=> SELECT * FROM t1 RIGHT JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   3 | c    |   3 | yyy
     |      |   5 | zzz
(3 rows)

=> SELECT * FROM t1 FULL JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |   3 | yyy
     |      |   5 | zzz
(4 rows)
```

用ON指定的连接条件也可以包含与连接不直接相关的条件。这种功能可能对某些查询很有用，但是需要我们仔细想清楚。例如：

```
=> SELECT * FROM t1 LEFT JOIN t2 ON t1.num = t2.num AND t2.value =
 'xxx';
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |     |
(3 rows)
```

注意把限制放在WHERE子句中会产生不同的结果：

```
=> SELECT * FROM t1 LEFT JOIN t2 ON t1.num = t2.num WHERE t2.value =
 'xxx';
 num | name | num | value
-----+------+-----+-------
```

```
   1 | a    |   1 | xxx
(1 row)
```

这是因为放在ON子句中的约束会在连接之前处理，而放在WHERE
子句中的约束则会在连接之后处理。这对内连接无关紧要，但对外连接影响很大。

7.2.1.2. 表和列别名

你可以给表以及复杂的表引用指定一个临时名字，以便在查询的其余部分引用该派生表。
这被称为表别名。

要创建一个表别名，我们可以写：

```
FROM table_reference AS alias
```

或者

```
FROM table_reference alias
```

AS关键字只是可选的语法噪音。*alias*可以是任意标识符。

表别名的典型应用是给长表名赋予比较短的标识符，好让连接子句更易读。例如：

```
SELECT * FROM some_very_long_table_name s JOIN another_fairly_long_
name a ON s.id = a.num;
```

对于当前查询而言，
别名成为该表引用的新名称–不允许在查询的其他地方通过原名称引用该表。因此，
以下写法无效：

```
SELECT * FROM my_table AS m WHERE my_table.a > 5;    -- wrong
```

表别名主要用于简化符号，但是当把一个表连接到它自身时必须使用别名，例如：

```
SELECT * FROM people AS mother JOIN people AS child ON mother.id =
 child.mother_id;
```

此外，当表引用是子查询时也必须使用别名（参见第 7.2.1.3 节）。

圆括弧用于解决歧义。在下面的示例中，第一个语句将把别名b赋给my_table的第二个实例，
但是第二个语句把别名赋给连接的结果：

```
SELECT * FROM my_table AS a CROSS JOIN my_table AS b ...
SELECT * FROM (my_table AS a CROSS JOIN my_table) AS b ...
```

另外一种给表指定别名的形式是给表的列赋予临时名字，就像给表本身指定别名一样：

```
FROM table_reference [AS] alias ( column1 [, column2 [, ...]] )
```

如果指定的列别名比表里实际的列少，那么剩下的列就没有被重命名。
这种语法对于自连接或子查询特别有用。

如果用这些形式中的任何一种给一个JOIN子句的输出附加了一个别名，那么该别名就在JOIN
的作用下隐去了其原始的名字。例如：

```
SELECT a.* FROM my_table AS a JOIN your_table AS b ON ...
```

是合法 SQL，但是：

```
SELECT a.* FROM (my_table AS a JOIN your_table AS b ON ...) AS c
```

是不合法的：表别名a在别名c外面是看不到的。

7.2.1.3. 子查询

指定派生表的子查询必须用圆括号括起来，并且必须赋予一个表别名。（参见第 7.2.1.2 节。）例如：

```
FROM (SELECT * FROM table1) AS alias_name
```

这个示例等效于FROM table1 AS alias_name。更有趣的情况是在子查询里面有分组或聚合的时候，子查询不能被简化为一个简单的连接。

一个子查询也可以是一个VALUES列表：

```
FROM (VALUES ('anne', 'smith'), ('bob', 'jones'), ('joe', 'blow'))
     AS names(first, last)
```

同样，表别名是必须的。为VALUES列表中的列分配别名是可选的，但这是一个好习惯。
更多信息可参见第 7.7 节。

7.2.1.4. 表函数

表函数是那些生成行集合的函数，这些行可以由基础类型（标量类型）组成，
也可以由复合数据类型（表行）组成。它们在查询的FROM子句中用法类似于表、视图或子查询。
表函数返回的列可以像表、视图或子查询的列一样，出现在SELECT、JOIN或WHERE子句中。

如果一个表函数返回基础数据类型，唯一的结果列名与该函数名相同。
如果该函数返回一个复合类型，结果列会得到与该类型的各个属性相同的名称。

表函数可以在FROM子句中被赋予别名，但也可以不使用别名。如果一个函数在FROM
子句中使用时没有别名，该函数名将被用作结果表的表名。

示例：

```
CREATE TABLE foo (fooid int, foosubid int, fooname text);

CREATE FUNCTION getfoo(int) RETURNS SETOF foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT * FROM getfoo(1) AS t1;

SELECT * FROM foo
    WHERE foosubid IN (
                        SELECT foosubid
                        FROM getfoo(foo.fooid) z
                        WHERE z.fooid = foo.fooid
                      );

CREATE VIEW vw_getfoo AS SELECT * FROM getfoo(1);

SELECT * FROM vw_getfoo;
```

有时候，定义一个能够根据调用方式返回不同列集合的表函数很有用。为支持这一点，表函数可以被声明为返回伪类型record。当此类函数在查询中使用时，必须在查询本身中指定预期的行结构，这样系统才能知道如何分析和规划该查询。考虑下面的示例：

```
SELECT *
    FROM dblink('dbname=mydb', 'SELECT proname, prosrc FROM pg_proc')
      AS t1(proname name, prosrc text)
    WHERE proname LIKE 'bytea%';
```

dblink函数（dblink模块的一部分）执行远程查询。它被声明为返回record，因为它可能用于任意类型的查询。实际的列集必须在调用它的查询中指定，这样分析器才知道像*这样的写法应当扩展成什么。

7.2.2. WHERE子句

此WHERE Clause的语法是：

```
WHERE search_condition
```

其中，*search_condition*可以是任意值表达式（见第 4.2 节），只要其返回值的类型为boolean。

在完成FROM子句的处理之后，生成的虚拟表中的每一行都会根据搜索条件进行检查。如果条件结果为真，该行就会保留在输出表中；否则（也就是说，如果结果为假或空）就会被丢弃。搜索条件通常至少会引用FROM子句生成的表中的某一列；这并非必须，但若不引用这些列，WHERE子句通常就没什么用了。

> **注意**
>
> 内连接的连接条件既可以写在WHERE子句也可以写在JOIN子句里。例如，这些表表达式是等效的：
>
> ```
> FROM a, b WHERE a.id = b.id AND b.val > 5
> ```
>
> 和：
>
> ```
> FROM a INNER JOIN b ON (a.id = b.id) WHERE b.val > 5
> ```
>
> 或者可能还有：
>
> ```
> FROM a NATURAL JOIN b WHERE b.val > 5
> ```
>
> 选择其中哪一种主要只是风格问题。FROM子句中的JOIN语法虽然属于 SQL 标准，但在其他 SQL 数据库管理系统中可能没有那么容易移植。对于外连接则没有选择：它们必须放在FROM子句中完成。外连接的ON或USING子句与WHERE条件并不等价，因为它除了会在最终结果中移除行之外，还会为不匹配的输入行增加结果行。

下面是一些WHERE子句的例子：

```
SELECT ... FROM fdt WHERE c1 > 5

SELECT ... FROM fdt WHERE c1 IN (1, 2, 3)
```

```
SELECT ... FROM fdt WHERE c1 IN (SELECT c1 FROM t2)

SELECT ... FROM fdt WHERE c1 IN (SELECT c3 FROM t2 WHERE c2 = fdt.c1 +
 10)

SELECT ... FROM fdt WHERE c1 BETWEEN (SELECT c3 FROM t2 WHERE c2 =
 fdt.c1 + 10) AND 100

SELECT ... FROM fdt WHERE EXISTS (SELECT c1 FROM t2 WHERE c2 > fdt.c1)
```

fdt是由FROM子句推导得到的表。不满足WHERE子句搜索条件的行会从fdt中消除。注意，这里将标量子查询用作值表达式。与其他查询一样，子查询也可以使用复杂的表表达式。还要注意子查询中对fdt的引用方式。将c1限定为fdt.c1仅在c1也是子查询推导输入表中的列名时才有必要。不过，即使不必限定，限定列名仍能使含义更清晰。本例说明了外层查询的列命名作用域如何扩展到其内部查询。

7.2.3. GROUP BY和HAVING子句

通过WHERE过滤之后，派生出来的输入表还可以使用GROUP BY子句进行分组，并通过HAVING子句排除某些分组行。

```
SELECT select_list
    FROM ...
    [WHERE ...]
    GROUP BY grouping_column_reference [, grouping_column_reference].
..
```

此GROUP BY Clause用于把表中所有列出列的值均相同的行分到同一组。列的列出顺序没有影响。其效果是把具有相同值的每一组行合并为一个分组行，由它代表组内的所有行。这样可以消除输出中的冗余，或者计算应用于各组的聚合。例如：

```
=> SELECT * FROM test1;
 x | y
---+---
 a | 3
 c | 2
 b | 5
 a | 1
(4 rows)

=> SELECT x FROM test1 GROUP BY x;
 x
---
 a
 b
 c
(3 rows)
```

在第二个查询里，我们不能写成SELECT * FROM test1 GROUP BY x，因为列y没有一个单一值可以与每个分组对应。被分组的列可以在选择列表中引用是因为它们在每个组都有单一的值。

通常，如果一个表被分了组，那么没有在GROUP BY中列出的列都不能被引用，除非在聚合表达式中被引用。一个用聚合表达式的示例是：

```
=> SELECT x, sum(y) FROM test1 GROUP BY x;
 x | sum
---+-----
 a |   4
 b |   5
 c |   2
(3 rows)
```

这里的sum是一个聚合函数，它在整个组上计算出一个单一值。
有关可用的聚合函数的更多信息可以在第 9.18 节中找到。

> **提示**
>
> 不带聚合表达式的分组实际上是在计算某一列的非重复值集合。这也可以用DISTINCT子句实现（参阅第 7.3.3 节）。

这里是另外一个示例：它计算每种产品的总销售额（而不是所有产品的总销售额）：

```
SELECT product_id, p.name, (sum(s.units) * p.price) AS sales
    FROM products p LEFT JOIN sales s USING (product_id)
    GROUP BY product_id, p.name, p.price;
```

在这个示例里，列product_id、p.name和p.price必须在GROUP BY子句里，因为它们都在查询的选择列表里被引用到（但见下文）。列s.units不必在GROUP BY列表里，因为它只是在一个聚合表达式（sum(...)）里使用，它代表一种产品的销售额。对于每种产品，
这个查询都返回一行关于该产品所有销售额的汇总。

如果 products 表的设计使得product_id是主键，那么在上面的示例中只按product_id分组就足够了，因为名称和价格都函数依赖于产品 ID，因此对于每个产品 ID 分组该返回哪个名称和价格值不会有歧义。

在严格的 SQL 里，GROUP BY只能对源表的列进行分组，但PostgreSQL把这个扩展为也允许GROUP BY去根据选择列表中的列分组。也允许对值表达式进行分组，而不仅是简单的列名。

如果一个表已经用GROUP BY子句分了组，而你只对其中某些组感兴趣，那么就可以使用HAVING子句。它很像WHERE子句，用于从结果中删除某些组。其语法是：

```
SELECT select_list FROM ... [WHERE ...] GROUP BY ... HAVING boolean_
expression
```

在HAVING
子句中的表达式可以引用分组的表达式和未分组的表达式（后者必须涉及一个聚合函数）。

示例：

```
=> SELECT x, sum(y) FROM test1 GROUP BY x HAVING sum(y) > 3;
 x | sum
---+-----
 a |   4
 b |   5
(2 rows)
```

```
=> SELECT x, sum(y) FROM test1 GROUP BY x HAVING x < 'c';
 x | sum
---+-----
 a |   4
 b |   5
(2 rows)
```

再次，一个更现实的示例：

```
SELECT product_id, p.name, (sum(s.units) * (p.price - p.cost)) AS
 profit
    FROM products p LEFT JOIN sales s USING (product_id)
    WHERE s.date > CURRENT_DATE - INTERVAL '4 weeks'
    GROUP BY product_id, p.name, p.price, p.cost
    HAVING sum(p.price * s.units) > 5000;
```

在上面的示例里，WHERE
子句通过一个未分组的列来选择数据行（该表达式仅对最近四周内发生的销售为真），而HAVING子句则把输出限制为总销售收入超过 5000 的组。请注意，聚合表达式不必在查询的所有部分都完全相同。

如果一个查询包含聚合函数调用，但没有GROUP BY子句，分组仍然会发生：结果是一个单一组行（如果该单一组行随后又被HAVING消除，则可能根本没有结果行）。如果查询包含HAVING
子句，即使没有任何聚合函数调用或GROUP BY子句，情况也是如此。

7.2.4. 窗口函数处理

如果查询包含任何窗口函数（见第 3.5 节、第 9.19 节和第 4.2.8 节），这些函数将在任何分组、聚合和HAVING过滤被执行之后被计算。也就是说如果查询使用了任何聚合、GROUP BY或HAVING，则窗口函数看到的行是分组行而不是来自于FROM/WHERE的原始表行。

当使用多个窗口函数时，窗口定义中具有语法等价的PARTITION BY和ORDER BY
子句的所有窗口函数，都保证在数据的一次遍历中求值。因此，即使ORDER BY不能唯一确定排序顺序，它们看到的排序顺序也相同。不过，对于具有不同PARTITION BY
或ORDER BY定义的函数，其求值不作保证。（在这种情况下，窗口函数的各次求值遍历之间通常需要排序步骤，而排序不保证保留那些在其ORDER BY看来等价的行的顺序。）

目前，窗口函数总是要求输入数据预先排好序，因此查询输出总会按照某个窗口函数的
PARTITION BY/ORDER BY子句排序。不过，不建议依赖这一点。如果你希望确保结果按特定方式排序，请显式使用顶层的ORDER BY子句。

7.3. 选择列表

如前面几节所示，SELECT命令中的表表达式会构造出一个中间虚拟表，
其中可能涉及组合表和视图、消除行、分组等操作。这个表随后交由选择列表处理。
选择列表决定中间表中的哪些列会实际输出。

7.3.1. 选择列表项

最简单的选择列表是*，它会输出表表达式生成的所有列。否则，
选择列表就是一个逗号分隔的值表达式列表（定义见第 4.2 节）。例如，它可以是一个列名列表：

```
SELECT a, b, c FROM ...
```

列名a、b和c要么是FROM子句中所引用表的实际列名，要么是像第 7.2.1.2 节所述为它们指定的别名。选择列表中可用的名字空间与WHERE子句相同；如果使用了分组，则与HAVING子句相同。

如果超过一个表有同样的列名，那么你还必须给出表名字，如：

```
SELECT tbl1.a, tbl2.a, tbl1.b FROM ...
```

在使用多个表时，要求一个特定表的所有列也是有用的：

```
SELECT tbl1.*, tbl2.a FROM ...
```

（另请参阅第 7.2.2 节。）

如果在选择列表中使用任意值表达式，那么从概念上说，它会在返回表中增加一个新的虚拟列。该值表达式会对每个结果行计算一次，并用该行中的值替换其中的列引用。不过，选择列表中的表达式并不一定要引用FROM子句表表达式中的列；例如，它也可以是任意常量算术表达式。

7.3.2. 列标签

选择列表中的项可以被赋予名字，以供后续处理使用，例如在ORDER BY子句中引用，或者供客户端应用显示。例如：

```
SELECT a AS value, b + c AS sum FROM ...
```

如果没有使用AS指定输出列名，系统就会分配一个默认列名。对于简单的列引用，它就是被引用列的名字；对于函数调用，它就是函数的名字；对于复杂表达式，系统则会生成一个通用名称。

此AS关键字是可选的，但前提是新列名不能与任何PostgreSQL关键字相同（见附录 C）。为了避免意外使用关键字，可以用双引号括住列名。例如，VALUE是关键字，因此下面的写法无效：

```
SELECT a value, b + c AS sum FROM ...
```

但下面的写法有效：

```
SELECT a "value", b + c AS sum FROM ...
```

为防范将来可能新增的关键字，建议始终写出AS，或者用双引号括住输出列名。

> **注意**
>
> 输出列的命名和在FROM子句里的命名是不一样的（参阅第 7.2.1.2 节）。它实际上允许你对同一个列命名两次，但是在选择列表中分配的名字是要传递下去的名字。

7.3.3. DISTINCT

在处理完选择列表之后，结果表还可以选择性地删除重复行。我们可以直接在SELECT后面写上DISTINCT关键字来指定：

```
SELECT DISTINCT select_list ...
```

（如果不用DISTINCT，则可以用ALL关键字来显式指定保留所有行这一默认行为。）

显然，如果两行至少在一个列值上不同，就认为它们是不同的。
空值在这种比较中被认为是相等的。

另外，我们还可以用任意表达式来判断什么行可以被认为是可区分的：

SELECT DISTINCT ON (*expression* [, *expression* ...]) *select_list* ...

这里的*expression*是一个对所有行求值的任意值表达式。
如果某一组行在这些表达式上的值都相等，那么它们就会被视为重复行，
因此只保留该组中的第一行。请注意，某一组里的“第一行”是不可预测的，
除非查询在足够多的列上进行了排序，以保证到达DISTINCT过滤器的行顺序是唯一的。
（DISTINCT ON的处理发生在ORDER BY排序之后。）

DISTINCT ON子句不是 SQL 标准的一部分，有时会被认为风格不佳，因为它的结果可能具有不确定性。通过审慎使用GROUP BY以及FROM中的子查询，可以避免使用这种结构，但它往往是最方便的替代方案。

7.4. 组合查询

两个查询的结果可以用集合操作并、交、差进行组合。语法是

query1 UNION [ALL] *query2*
query1 INTERSECT [ALL] *query2*
query1 EXCEPT [ALL] *query2*

*query1*和*query2*都是可以使用以上所有特性的查询。集合操作也可以嵌套和链式使用，例如

query1 UNION *query2* UNION *query3*

它会被当作：

(*query1* UNION *query2*) UNION *query3*

来执行。

UNION有效地把*query2*的结果附加到*query1*
的结果上（不过我们不能保证这就是这些行实际被返回的顺序）。此外，
它将删除结果中所有重复的行，就像DISTINCT做的那样，除非你使用了UNION ALL。

INTERSECT返回那些同时存在于*query1*和*query2*的结果中的行，除非声明了INTERSECT ALL，否则所有重复行都被消除。

EXCEPT返回所有在*query1*结果中但不在*query2*结果中的行。
（这有时被称为两个查询的差。）同样，除非使用了EXCEPT ALL，否则重复行都会被消除。

为了计算两个查询的并、交、差，这两个查询必须是“并操作兼容的”，
也就意味着它们都返回同样数量的列，并且对应的列有兼容的数据类型，如第 10.5 节中描述的那样。

7.5. 行排序

当一个查询生成输出表之后（也就是处理完选择列表之后），还可以选择对其排序。
如果没有选择排序，行将按未指定的顺序返回。

这时的实际顺序取决于扫描和连接计划类型以及磁盘上的物理顺序，但绝不能依赖这些因素。只有显式选择排序步骤，才能保证特定的输出顺序。

ORDER BY子句指定了排序顺序：

```
SELECT select_list
    FROM table_expression
    ORDER BY sort_expression1 [ASC | DESC] [NULLS { FIRST | LAST }]
             [, sort_expression2 [ASC | DESC] [NULLS { FIRST | LAST }
] ...]
```

排序表达式可以是任何在查询选择列表中合法的表达式。例如：

```
SELECT a, b FROM table1 ORDER BY a + b, c;
```

当指定了多个表达式时，后面的值将用于对那些在前面值上相等的行继续排序。每个表达式后面都可以选择性地写上ASC或DESC关键字，以指定升序或降序排序方向。ASC是默认值。升序会把较小的值排在前面，而“较小”是由<操作符定义的。类似地，降序则由>操作符定义。[1]

NULLS FIRST和NULLS LAST选项可用于决定空值在排序顺序中是位于非空值之前还是之后。默认情况下，空值排序时被认为大于任何非空值；也就是说，在DESC顺序下默认是NULLS FIRST，否则默认是NULLS LAST。

注意顺序选项是对每一个排序列独立考虑的。例如ORDER BY x, y DESC表示ORDER BY x ASC, y DESC，而和ORDER BY x DESC, y DESC不同。

一个*sort_expression*也可以是输出列的列标签或编号，例如：

```
SELECT a + b AS sum, c FROM table1 ORDER BY sum;
SELECT a, max(b) FROM table1 GROUP BY a ORDER BY 1;
```

这两条查询都按第一个输出列排序。注意，输出列名必须单独出现，不能用于表达式中—例如，下面的写法是不正确的：

```
SELECT a + b AS sum, c FROM table1 ORDER BY sum + c;          -- wrong
```

这一限制是为了减少歧义。如果ORDER BY项是一个简单名称，既可能匹配输出列名，也可能匹配表表达式中的列，那么仍会有歧义。这种情况下使用输出列。只有在使用AS将输出列重命名为其他表列的名称时，才会造成这种混淆。

ORDER BY可以被应用于UNION、INTERSECT或EXCEPT组合的结果，但是在这种情况中它只被允许根据输出列名或编号排序，而不能根据表达式排序。

7.6. LIMIT和OFFSET

LIMIT和OFFSET允许仅检索查询其余部分所生成行的一部分：

```
SELECT select_list
    FROM table_expression
    [ ORDER BY ... ]
    [ LIMIT { number | ALL } ] [ OFFSET number ]
```

[1] 事实上，PostgreSQL会使用该表达式数据类型的默认 B-树操作符类来决定ASC和DESC的排序顺序。按惯例，数据类型会被设置成由<和>操作符来对应这种排序顺序，不过用户定义数据类型的设计者也可以选择不同的做法。

如果给出了一个限制计数，那么会返回数量不超过该限制的行（但可能更少些，因为查询本身可能生成的行数就比较少）。LIMIT ALL的效果和省略LIMIT子句一样。

OFFSET说明在开始返回结果行之前要跳过多少行。OFFSET 0的效果与省略OFFSET子句相同，而LIMIT NULL与省略LIMIT子句相同。如果OFFSET和LIMIT都出现，那么在开始返回LIMIT行之前，会先跳过OFFSET行。

如果使用LIMIT，那么用一个ORDER BY子句把结果行约束成一个唯一的顺序是很重要的。否则你就会拿到一个不可预料的该查询的行的子集。你要的可能是第十到第二十行，但以什么顺序的第十到第二十？除非你指定了ORDER BY，否则顺序是不知道的。

查询优化器在生成查询计划时会考虑LIMIT，因此如果你给定LIMIT和OFFSET，那么你很可能收到不同的规划（产生不同的行顺序）。因此，使用不同的LIMIT/OFFSET值选择查询结果的不同子集将生成不一致的结果，除非你用ORDER BY强制一个可预测的顺序。这并非bug，这是一个很自然的结果，因为 SQL 没有许诺把查询的结果按照任何特定的顺序发出，除非用了ORDER BY来约束顺序。

被OFFSET子句跳过的行仍然需要在服务器内部计算；因此，较大的OFFSET可能效率不高。

7.7. VALUES列表

VALUES提供了一种生成“常量表”的方法，可以在查询中使用，而不必在磁盘上实际创建并填充表。语法为：

```
VALUES ( expression [, ...] ) [, ...]
```

每个括号括住的表达式列表都会生成表中的一行。所有列表必须具有相同数量的元素（即表中的列数），并且各列表中对应项的数据类型必须兼容。为结果中各列分配的实际数据类型，使用与UNION相同的规则确定（见第 10.5 节）。

一个示例：

```
VALUES (1, 'one'), (2, 'two'), (3, 'three');
```

将会返回一个有两列三行的表。它实际上等效于：

```
SELECT 1 AS column1, 'one' AS column2
UNION ALL
SELECT 2, 'two'
UNION ALL
SELECT 3, 'three';
```

在默认情况下，PostgreSQL将column1、column2等名字分配给一个VALUES表的列。这些列名不是由SQL标准指定的，并且不同的数据库系统的做法也不同，因此通常最好使用表别名列表来重写这些默认的名字。

从语法上说，后面跟有表达式列表的VALUES列表被视为等同于

```
SELECT select_list FROM table_expression
```

，并且可以出现在SELECT能出现的任何地方。例如，你可以把它用作UNION的一部分，或者附加一个*sort_specification*（ORDER BY、LIMIT和/或OFFSET）给它。VALUES最常见的用途是作为一个INSERT命令的数据源，其次常见的用途是作为子查询。

更多信息请见VALUES。

7.8. WITH查询（公共表表达式）

WITH提供了一种为更大查询编写辅助语句的方法。这些语句通常被称为公共表表达式或CTE，可以视为仅在单个查询期间存在的临时表定义。WITH子句中的每个辅助语句都可以是SELECT、INSERT、UPDATE或DELETE；而WITH子句本身则附加到一个主语句上，该主语句同样可以是SELECT、INSERT、UPDATE或DELETE。

7.8.1. WITH中的SELECT

将SELECT用于WITH的基本价值是把复杂查询分解成更简单的部分。例如：

```
WITH regional_sales AS (
    SELECT region, SUM(amount) AS total_sales
    FROM orders
    GROUP BY region
), top_regions AS (
    SELECT region
    FROM regional_sales
    WHERE total_sales > (SELECT SUM(total_sales)/10 FROM regional_
sales)
)
SELECT region,
       product,
       SUM(quantity) AS product_units,
       SUM(amount) AS product_sales
FROM orders
WHERE region IN (SELECT region FROM top_regions)
GROUP BY region, product;
```

这条查询只显示销售额最高的几个地区中各产品的销售总额。WITH子句定义了两个辅助语句，名为regional_sales和top_regions，其中，regional_sales的输出用于top_regions，而top_regions的输出用于主SELECT查询。本例也可以不用WITH来编写，但那样就需要两层嵌套的子SELECT。采用这种写法更容易理解一些。

可选的RECURSIVE修饰符把WITH从单纯的语法便利变成一项能够完成标准 SQL 中原本无法完成之事的特性。使用RECURSIVE后，WITH查询可以引用自己的输出。
一个非常简单的例子是以下查询，它计算 1 到 100 的整数之和：

```
WITH RECURSIVE t(n) AS (
    VALUES (1)
  UNION ALL
    SELECT n+1 FROM t WHERE n < 100
)
SELECT sum(n) FROM t;
```

递归WITH查询的一般形式始终是一个非递归项，接着是UNION（或UNION ALL），然后是一个递归项，其中只有递归项才能包含对查询自身输出的引用。这样的查询按以下方式执行：

递归查询求值

1. 计算非递归项。对UNION（但不对UNION ALL），抛弃重复行。把所有剩余的行包括在递归查询的结果中，并且也把它们放在一个临时的工作表中。

2. 只要工作表不为空，重复下列步骤：

a. 计算递归项，用当前工作表的内容替换递归自引用。对UNION（不是UNION ALL），抛弃重复行以及那些与之前结果行重复的行。将剩下的所有行包括在递归查询的结果中，并且也把它们放在一个临时的中间表中。

b. 用中间表的内容替换工作表的内容，然后清空中间表。

> **注意**
>
> 严格地说，这个过程是迭代而非递归，但RECURSIVE是 SQL 标准委员会选定的术语。

在上面的示例中，工作表在每一步都只有一行，并且在连续步骤中依次取值 1 到 100。到了第 100 步，由于WHERE子句不再产生输出，因此查询终止。

递归查询通常用于处理层次结构或树形结构数据。
以下是一个有用的例子：只给定一个表示直接包含关系的表，
查询找出产品的所有直接和间接子部件：

```
WITH RECURSIVE included_parts(sub_part, part, quantity) AS (
    SELECT sub_part, part, quantity FROM parts WHERE part = 'our_
product'
  UNION ALL
    SELECT p.sub_part, p.part, p.quantity
    FROM included_parts pr, parts p
    WHERE p.part = pr.sub_part
)
SELECT sub_part, SUM(quantity) as total_quantity
FROM included_parts
GROUP BY sub_part
```

使用递归查询时，必须确保查询的递归部分最终会不返回任何元组，否则查询会无限循环。
有时，可以用UNION代替UNION ALL，通过丢弃与先前输出行重复的行来做到这一点。不过，循环往往并不涉及完全重复的输出行：可能需要只检查一个或几个字段，
以判断是否曾到达同一点。处理这种情况的标准方法是计算一个包含已访问值的数组。例如，
考虑以下查询，它搜索表graph，使用其中的link字段：

```
WITH RECURSIVE search_graph(id, link, data, depth) AS (
    SELECT g.id, g.link, g.data, 1
    FROM graph g
  UNION ALL
    SELECT g.id, g.link, g.data, sg.depth + 1
    FROM graph g, search_graph sg
    WHERE g.id = sg.link
)
SELECT * FROM search_graph;
```

如果link关系包含循环，此查询就会循环。由于需要输出“depth”，仅把UNION ALL改为UNION并不能消除循环。我们需要判断在沿特定链接路径前进时，是否再次到达了同一行。
为这个容易循环的查询添加path和cycle两列：

```
WITH RECURSIVE search_graph(id, link, data, depth, path, cycle) AS (
        SELECT g.id, g.link, g.data, 1,
          ARRAY[g.id],
```

```
          false
        FROM graph g
      UNION ALL
        SELECT g.id, g.link, g.data, sg.depth + 1,
          path || g.id,
          g.id = ANY(path)
        FROM graph g, search_graph sg
        WHERE g.id = sg.link AND NOT cycle
)
SELECT * FROM search_graph;
```

除了防止循环，数组值本身也常常很有用，因为它表示到达某一行所经过的“路径”。

在需要检查多个字段以识别循环的一般情况下，使用行数组。例如，如果需要比较字段`f1`和`f2`：

```
WITH RECURSIVE search_graph(id, link, data, depth, path, cycle) AS (
        SELECT g.id, g.link, g.data, 1,
          ARRAY[ROW(g.f1, g.f2)],
          false
        FROM graph g
      UNION ALL
        SELECT g.id, g.link, g.data, sg.depth + 1,
          path || ROW(g.f1, g.f2),
          ROW(g.f1, g.f2) = ANY(path)
        FROM graph g, search_graph sg
        WHERE g.id = sg.link AND NOT cycle
)
SELECT * FROM search_graph;
```

提示

在通常只有一个字段需要检查即可识别环的情况下，可以省略`ROW()`语法。
这样就能使用简单数组而不是复合类型数组，从而提高效率。

提示

递归查询求值算法按广度优先搜索顺序产生输出。可以让外层查询按这样构造的“路径”列进行`ORDER BY`排序，以深度优先搜索顺序显示结果。

当不确定查询是否会循环时，一个有用的测试技巧是在父查询中加入`LIMIT`。例如，
下面的查询将会无限循环，除非加上`LIMIT`：

```
WITH RECURSIVE t(n) AS (
    SELECT 1
  UNION ALL
    SELECT n+1 FROM t
)
SELECT n FROM t LIMIT 100;
```

这一技巧之所以有效，是因为PostgreSQL的实现只会对`WITH`
查询中父查询实际取出的那些行求值。不推荐在生产中使用这一技巧，

因为其他系统可能有不同的行为。另外，
如果让外层查询对递归查询结果排序或将其与其他表连接，这一技巧通常就不起作用，
因为在这种情况下，外层查询通常还是会尝试取出WITH查询的全部输出。

WITH查询的一个有用特性是，在父查询每次执行时，它只求值一次，即使父查询或同级WITH
查询对它有多次引用也是如此。因此，在多处需要的昂贵计算可以放在一个WITH查询中，
以避免重复工作。另一个用途是防止带副作用的函数被意外求值多次。不过，另一方面，
与普通子查询相比，优化器更难把父查询中的限制下推到WITH查询中。WITH
查询通常会按书写的样子求值，不会抑制父查询稍后可能丢弃的行。（不过，如上所述，
如果对该查询的引用只需要有限数量的行，求值可能会提前停止。）

上面的示例仅显示了WITH与SELECT一起使用，但它可以以相同的方式附加到INSERT、UPDATE
或DELETE。在每种情况下，它都等效于提供了可在主命令中引用的临时表。

7.8.2. WITH中的数据修改语句

你可以在WITH中使用数据修改语句（INSERT、UPDATE或DELETE）。
这样就可以在同一查询中执行多个不同的操作。例如：

```
WITH moved_rows AS (
    DELETE FROM products
    WHERE
        "date" >= '2010-10-01' AND
        "date" < '2010-11-01'
    RETURNING *
)
INSERT INTO products_log
SELECT * FROM moved_rows;
```

这个查询实际上将行从products移动到 products_log。在WITH中的DELETE 从products中
删除指定的行，通过其RETURNING子句返回它们的内容；然后主查询读取该输出并将其插入到
products_log中。

上述示例中有一个细节值得注意：WITH子句附加在INSERT上，而不是附加在INSERT
内部的子SELECT上。这是必要的，因为数据修改语句只允许出现在附着于顶层语句的WITH
子句中。不过，普通WITH的可见性规则仍然适用，因此仍然可以在该子SELECT中引用WITH
语句的输出。

如上面的例子所示，WITH中的数据修改语句通常带有RETURNING子句。
构成其他查询可引用临时表的，是RETURNING子句的输出，而不是数据修改语句的目标表。
如果WITH中的数据修改语句缺少RETURNING子句，那么它就不会形成临时表，
也无法被查询的其他部分引用。尽管如此，这样的语句仍然会被执行。
下面是一个并不特别有用的示例：

```
WITH t AS (
    DELETE FROM foo
)
DELETE FROM bar;
```

这个示例将从表foo和bar中移除所有行。被报告给客户端的受影响行的数目可能只包括从bar
中移除的行。

数据修改语句不允许递归自引用。有时，可以通过引用递归WITH的输出来绕过这一限制，例如：

```
WITH RECURSIVE included_parts(sub_part, part) AS (
```

```
    SELECT sub_part, part FROM parts WHERE part = 'our_product'
  UNION ALL
    SELECT p.sub_part, p.part
    FROM included_parts pr, parts p
    WHERE p.part = pr.sub_part
)
DELETE FROM parts
  WHERE part IN (SELECT part FROM included_parts);
```

此查询会删除某产品的所有直接和间接子部件。

WITH中的数据修改语句只会执行一次，并且总会执行到完成，
而不管主查询是否读取了它们的全部输出，甚至是否读取了任何输出。注意这与WITH中SELECT的规则不同：正如前一小节所述，SELECT只会执行到主查询实际需要其输出为止。

WITH中的子语句彼此之间以及与主查询是并发执行的。因此，在WITH中使用数据修改语句时，
实际更新发生的顺序是不可预知的。所有语句都使用同一个快照执行（参见第 13 章），因此它们无法“看见”彼此对目标表产生的影响。这减轻了实际行更新顺序不可预知所带来的影响，
也意味着RETURNING数据是在不同WITH子语句与主查询之间传递更改的唯一方式。例如，在

```
WITH t AS (
    UPDATE products SET price = price * 1.05
    RETURNING *
)
SELECT * FROM products;
```

外层SELECT返回的将是UPDATE操作发生之前的原始价格，而在

```
WITH t AS (
    UPDATE products SET price = price * 1.05
    RETURNING *
)
SELECT * FROM t;
```

外层SELECT将返回更新后的数据。

在单个语句中试图更新同一行两次是不受支持的。只会发生一次修改，但很难（有时根本无法）可靠地预测究竟是哪一次。
这同样适用于删除在同一语句中已经被更新过的行：只有更新会生效。因此，
通常应避免在一个语句中两次修改同一行。
特别要避免编写那些可能影响主语句或同级子语句所修改行的WITH子语句，
因为这类语句的效果将不可预测。

当前，用作WITH中数据修改语句目标的任何表不能有条件规则、ALSO
规则或扩展到多个语句的INSTEAD规则。

第 8 章 数据类型

PostgreSQL 提供了丰富的原生数据类型。用户可以使用CREATE TYPE命令向 PostgreSQL添加新类型。

表 8.1展示了所有内置的通用数据类型。“别名”列中列出的多数可选名称，都是 PostgreSQL出于历史原因在内部使用的名称。此外，还有一些内部使用或已废弃的类型也可用，
但未在此列出。

表 8.1. 数据类型

名字	别名	描述
bigint	int8	有符号的8字节整数
bigserial	serial8	自动递增的8字节整数
bit [(*n*)]		定长位串
bit varying [(*n*)]	varbit	变长位串
boolean	bool	逻辑布尔值（真/假）
box		平面上的矩形框
bytea		二进制数据（“字节数组”）
character varying [(*n*)]	varchar [(*n*)]	变长字符串
character [(*n*)]	char [(*n*)]	定长字符串
cidr		IPv4或IPv6网络地址
circle		平面上的圆
date		日历日期（年、月、日）
double precision	float8	双精度浮点数（8 字节）
inet		IPv4或IPv6主机地址
integer	int, int4	有符号4字节整数
interval [*fields*] [(*p*)]		时间段
line		平面上的无限长的线
lseg		平面上的线段
macaddr		MAC（Media Access Control）地址
money		货币数量
numeric [(*p*, *s*)]	decimal [(*p*, *s*)]	可选择精度的精确数字
path		平面上的几何路径
point		平面上的几何点
polygon		平面上的封闭几何路径
real	float4	单精度浮点数（4字节）
smallint	int2	有符号2字节整数
serial	serial4	自动递增的4字节整数
text		变长字符串
time [(*p*)] [without time zone]		一天中的时间（无时区）

名字	别名	描述
time [(*p*)] with time zone	timetz	一天中的时间，包括时区
timestamp [(*p*)] [without time zone]		日期和时间（无时区）
timestamp [(*p*)] with time zone	timestamptz	日期和时间，包括时区
tsquery		文本搜索查询
tsvector		文本搜索文档
txid_snapshot		用户级事务 ID 快照
uuid		通用唯一标识符
xml		XML 数据

兼容性

下列类型（或它们的这些拼写形式）由SQL规定：bigint、bit、bit varying、boolean、char、character varying、character、varchar、date、double precision、integer、interval、numeric、decimal、real、smallint、time（有时区或无时区）、timestamp（有时区或无时区）、xml。

每种数据类型都有一种由其输入和输出函数决定的外部表示。
许多内置类型的外部格式都很直观。不过，也有一些类型是 PostgreSQL所特有的，例如几何路径；还有一些类型可能存在多种可能的格式，例如日期/时间类型。
有些输入和输出函数并不可逆，也就是说，输出函数的结果与原始输入相比可能会丢失精度。

8.1. 数字类型

数字类型包括 2、4、8 字节的整数，4、8 字节的浮点数，以及可选精度的小数。表 8.2列出了所有可用类型。

表 8.2. 数字类型

名字	存储尺寸	描述	范围
smallint	2字节	小范围整数	-32768 到 +32767
integer	4字节	整数的典型选择	-2147483648 到 +2147483647
bigint	8字节	大范围整数	-9223372036854775808 到 9223372036854775807
decimal	可变	用户指定精度，精确	最高小数点前131072位，以及小数点后16383位
numeric	可变	用户指定精度，精确	最高小数点前131072位，以及小数点后16383位
real	4字节	可变精度，不精确	6位十进制精度
double precision	8字节	可变精度，不精确	15位十进制精度
serial	4字节	自动递增的整数	1到2147483647
bigserial	8字节	自动递增的大整数	1到9223372036854775807

数字类型常量的语法在第 4.1.2 节里描述。数字类型有一整套对应的算术操作符和函数。相关信息请参考第 9 章。下面的几节详细描述这些类型。

8.1.1. 整数类型

类型 smallint、integer 和 bigint 用于存储不同范围的整数，也就是没有小数部分的数。试图存储超出允许范围的值会导致错误。

常用的类型是integer，因为它提供了在范围、存储空间和性能之间的最佳平衡。
一般只有在磁盘空间紧张的时候才使用 smallint类型。只有当integer类型的范围不够时才应使用bigint类型，因为integer类型的速度绝对更快。

在非常精简的操作系统上，bigint类型可能无法正确工作，
因为它依赖编译器对八字节整数的支持。在这样的机器上，bigint的行为与integer相同，
但仍然占用八字节的存储空间。（我们不知道任何现代平台存在这种情况。）

SQL只规定了整数类型integer（或int）、smallint和bigint。类型int2、int4和int8都是扩展，也在某些其他SQL数据库系统中使用。

8.1.2. 任意精度数值

类型numeric可以存储非常多位的数字并精确地执行计算。
我们特别建议将它用于货币金额和其它要求精确的数量。不过，numeric
类型上的算术运算比整数类型或者下一节描述的浮点数类型要慢很多。

下面使用如下术语：numeric的精度是整个数中有效数字的总数，
即小数点两侧的数字位数之和。numeric的小数位数是小数部分中十进制数字的数量，
即小数点右侧的位数。因此，数值 23.5141 的精度为 6，小数位数为 4。整数可以视为小数位数为零。

可以配置numeric列的最大精度和最大小数位数。要声明numeric类型的列，使用以下语法：

NUMERIC(*precision*, *scale*)

精度必须为正数，小数位数必须为零或正数。也可以使用：

NUMERIC(*precision*)

这会选择小数位数为 0。指定：

NUMERIC

而不指定任何精度或小数位数，会创建一个可存储任意精度和小数位数数值的列，
直到达到实现对精度的限制。这种列不会把输入值强制转换为某个特定的小数位数，
而声明了小数位数的numeric列会将输入值强制转换为该小数位数。
（SQL标准要求默认小数位数为 0，即强制转换为整数精度。我们认为这没有多少用处。如果你关心可移植性，请始终显式指定精度和小数位数。）

> **注意**
>
> 在类型声明中显式指定精度时，所允许的最大值为 1000；未指定精度的NUMERIC受表 8.2中所述限制的约束。

如果要存储的值的小数位数大于该列声明的小数位数，系统会把该值舍入到指定的小数位数。然后，如果小数点左侧的位数超过了声明的精度减去声明的小数位数，就会报错。

数值在物理上存储时不会保留多余的前导零或尾随零。因此，列上声明的精度和小数位数只是最大值，而不是固定分配的空间（从这个意义上说，`numeric` 更像 `varchar(n)`，而不像 `char(n)`）。实际存储需求是每四个十进制数字组占两个字节，再加上 3 到 8 字节的开销。

除了普通数值外，`numeric`类型还允许特殊值`NaN`，表示“非数”。对`NaN`的任何运算都会产生另一个`NaN`。在 SQL 命令中将该值写为常量时，必须加上引号，例如`UPDATE table SET x = 'NaN'`。输入时，字符串`NaN`的识别不区分大小写。

> **注意**
>
> 在大多数“非数字”概念的实现中，`NaN` 都被认为不等于任何其他数值（包括 `NaN` 本身）。为了让 `numeric` 值能够排序并用于基于树的索引，PostgreSQL 将 `NaN` 值视为彼此相等，并且大于所有非 `NaN` 值。

类型`decimal`和`numeric`是等效的。两种类型都是SQL标准的一部分。

8.1.3. 浮点类型

数据类型`real`和`double precision`是不精确的、变精度的数字类型。实际上，在底层处理器、操作系统和编译器支持的范围内，这两种类型通常实现了IEEE 754 二进制浮点算术标准（分别对应单精度和双精度）。

所谓不精确，是指某些值无法被精确转换为内部格式，只能以近似值存储，因此存储并再取出一个值时，可能会看到轻微差异。如何处理这类误差以及它们在计算中如何传播，是数学和计算机科学中的一个独立领域，这里不再展开，只强调以下几点：

- 如果你要求准确的存储和计算（例如计算货币金额），应使用`numeric`类型。
- 如果你想用这些类型进行任何重要的复杂计算，尤其是依赖边界情形（无穷大、下溢）特定行为的计算，那么应该仔细评估其实现。
- 比较两个浮点值是否相等，未必总能得到符合预期的结果。

在大多数平台上，`real`类型的范围至少为 1E-37 到 1E+37，精度至少为 6 位十进制数字。`double precision`类型的范围通常约为 1E-307 到 1E+308，精度至少为 15 位数字。过大或过小的值会导致错误。如果输入数字的精度过高，可能会发生舍入。过于接近零且无法表示为与零不同的数值，会导致下溢错误。

> **注意**
>
> extra_float_digits设置控制浮点值转换为文本输出时所包含的额外有效数字位数。使用默认值`0`时，输出在 PostgreSQL 支持的每个平台上都相同。增大该值会使输出更精确地表示所存储的值，但可能不具有可移植性。

除了普通数值外，浮点类型还有几个特殊值：

```
Infinity
-Infinity
NaN
```

它们分别表示 IEEE 754 的特殊值“无穷大”，“负无穷大”和“非数字”。（在浮点运算不遵循 IEEE 754 的机器上，这些值可能无法按预期工作。）在 SQL 命令中将这些值写为常量时，必须加上引号，例如`UPDATE table SET x = 'Infinity'`。输入时，这些字符串的识别不区分大小写。

> **注意**
>
> IEEE 754 规定，`NaN` 不应与任何其他浮点值（包括`NaN`）相等。为了允许浮点值被排序，并可用于基于树的索引，PostgreSQL 将`NaN`视为彼此相等，并且大于所有非 `NaN`值。

PostgreSQL 也支持 SQL 标准记法 `float` 和 `float(p)` 来指定不精确数值类型。这里，*p* 指定可接受的最小二进制精度位数。PostgreSQL 将 `float(1)` 到 `float(24)` 视为选择 `real` 类型，而 `float(25)` 到 `float(53)` 则选择 `double precision`。*p* 超出允许范围会报错。未指定精度的 `float` 视为 `double precision`。

> **注意**
>
> 在PostgreSQL 7.4 之前，`float(p)` 中的精度被理解为这么多位十进制数字。这一点已被纠正为符合 SQL 标准，标准规定精度以二进制位衡量。假定`real`和`double precision`的尾数分别恰好有 24 位和 53 位，对于 IEEE 标准浮点实现而言是正确的。在非 IEEE 平台上可能会有少许偏差，但为简单起见，所有平台都使用相同的*p*取值范围。

8.1.4. serial 类型

`serial` 和 `bigserial` 并不是真正的数据类型，它们只是为了创建唯一标识符列而提供的记法便利（类似于其他一些数据库支持的 `AUTO_INCREMENT` 属性）。在当前实现中，下列语句：

```
CREATE TABLE tablename (
    colname SERIAL
);
```

等价于以下语句：

```
CREATE SEQUENCE tablename_colname_seq;
CREATE TABLE tablename (
    colname integer NOT NULL DEFAULT nextval('tablename_colname_seq')
);
ALTER SEQUENCE tablename_colname_seq OWNED BY tablename.colname;
```

这样就创建了一个整数列，并把它的默认值设置为从序列生成器中取值。同时还会加上`NOT NULL`约束，以确保不能插入空值。（在大多数情况下，你可能还会希望再加上`UNIQUE` 或`PRIMARY KEY`约束，以防意外插入重复值，但这不会自动发生。）最后，该序列会被标记为“属于”该列，这样当列或表被删除时，序列也会随之删除。

> **注意**
>
> 在PostgreSQL 7.3 之前，serial 隐含UNIQUE。现在不再自动如此。如果你希望 serial 列具有唯一约束或作为主键，现在必须像对待任何其他数据类型一样显式指定。

> **注意**
>
> 因为 serial 和 bigserial 是用序列实现的，所以即使没有删除任何行，列中出现的值序列也可能存在"空洞"或缺口。即使包含该值的行从未成功插入表中，
> 从序列分配出去的值仍会被视为"已使用"。例如，如果插入事务回滚，
> 就会发生这种情况。详细信息见第 9.15 节中的 nextval()。

要向 serial 列插入序列中的下一个值，应指定让serial列使用其默认值。这既可以通过在 INSERT 语句的列表中省略该列来实现，也可以通过使用 DEFAULT 关键字来实现。

类型名 serial 和 serial4 是等价的：二者都会创建 integer 列。类型名 bigserial 和 serial8 的工作方式相同，只是它们创建的是 bigint 列。如果预计表在其生命周期内会使用超过 2^{31} 个标识符，就应使用 bigserial。

为 serial 列创建的序列会在其所属列被删除时自动删除。你也可以在不删除该列的情况下删除该序列，但这会强制移除该列的默认值表达式。

8.2. 货币类型

money 类型以固定的小数精度存储货币金额；参见表 8.3。小数精度由数据库的lc_monetary设置决定。表中显示的范围假定有两位小数。输入支持多种格式，
包括整数字面量和浮点数字面量，以及典型的货币格式，例如 '$1,000.00'。输出通常也采用后一种形式，但会受到区域设置影响。

表 8.3. 货币类型

名字	存储尺寸	描述	范围
money	8 字节	货币额	-92233720368547758.08 到 +92233720368547758.07

由于该数据类型的输出依赖区域设置，因此在 lc_monetary 设置不同的数据库之间装入 money 数据时，可能无法正常工作。为了避免这类问题，在将转储恢复到新数据库之前，应确保新数据库的 lc_monetary 设置与原数据库相同或等效。

数据类型numeric、int和bigint的值可以转换为money。从数据类型real和double precision的转换可以通过先转换为numeric来实现，例如：

```
SELECT '12.34'::float8::numeric::money;
```

但是，我们不推荐这样做。浮点数不应该被用来处理货币，因为浮点数可能会有舍入误差。

一个money值可以在不损失精度的情况下转换为numeric。转换到其他类型可能会丢失精度，
并且必须采用两个阶段完成：

```
SELECT '52093.89'::money::numeric::float8;
```

当一个 money 值被另一个 money 值除时，结果是 double precision（即一个纯数字，而不是金额），因为在除法中货币单位被约掉了。

8.3. 字符类型

表 8.4. 字符类型

名字	描述
character varying(*n*),varchar(*n*)	有长度限制的变长
character(*n*),char(*n*)	定长，空白填充
text	无限长度的变长

表 8.4显示了在PostgreSQL里可用的一般用途的字符类型。

SQL 定义了两种主要字符类型：character varying(*n*) 和 character(*n*)，其中 *n* 是正整数。这两种类型都可以存储长度最多为 *n* 个字符（而不是字节）的字符串。试图把更长的字符串存入这些类型的列时会报错，除非超出的字符全部都是空格，
在这种情况下字符串会被截断到最大长度。（这一多少有些奇怪的例外是 SQL 标准要求的。）如果要存储的字符串短于声明长度，character 类型的值会用空格填充；character varying 类型则只是简单地存储较短的字符串。

如果显式把一个值转换为 character varying(*n*) 或 character(*n*)，那么超长的值会被截断为 *n* 个字符而不会报错。（这同样是 SQL 标准要求的。）

varchar(*n*) 和 char(*n*) 分别是 character varying(*n*) 和 character(*n*) 的别名。character 若不带长度说明则等同于 character(1)。如果 character varying 不带长度说明，则该类型接受任意长度的字符串。这是 PostgreSQL 的扩展。

此外，PostgreSQL 还提供 text 类型，用于存储任意长度的字符串。虽然 text 类型不在 SQL 标准中，但其他若干 SQL 数据库管理系统也提供了它。

character 类型的值在物理存储时会在右侧用空格填充到指定宽度 *n*，并且也会按这种形式存储和显示。但是，这些填充空格在语义上被视为不重要。在比较两个 character 值时会忽略尾随空格，把 character 值转换为其他字符串类型之一时，尾随空格也会被移除。请注意，在 character varying 和 text 值中，以及进行模式匹配（例如 LIKE 和正则表达式）时，尾随空格在语义上是有意义的。

短字符串（最长 126 字节）的存储需求是 1 个字节，再加上实际字符串本身；对 character 而言，这其中还包括填充的空格。更长的字符串则需要 4 个字节的额外开销，而不是 1 个字节。长字符串会被系统自动压缩，因此在磁盘上的实际占用可能更小。非常长的值还会被存储在后台表中，以免影响对较短列值的快速访问。无论如何，可存储的最长字符串大约为 1 GB。（数据类型声明中 *n* 允许的最大值比这还小。更改这个限制并没有意义，因为在多字节字符编码下，字符数和字节数可能差异很大。如果你想存储没有明确上限的长字符串，应使用 text 或未指定长度的 character varying，而不是随意给出一个长度上限。）

> **提示**
>
> 这三种类型之间没有性能差别，除了使用空格填充类型时会占用更多存储空间，
> 以及在写入带长度约束的列时需要少量额外 CPU 周期来检查长度。虽然在某些其他数据库系统中，character(*n*) 可能有一定性能优势，但在 PostgreSQL 中并不存在这种优势。

> 事实上，由于额外的存储开销，character(*n*) 通常是三种类型中最不常用的一种。在大多数情况下，应优先使用 text 或 character varying。

关于字符串常量的语法，请参见第 4.1.2.1 节；关于可用的操作符和函数，请参见第 9 章。数据库字符集决定了用于存储文本值的字符集；有关字符集支持的更多信息，请参阅第 22.3 节。

例 8.1. 使用字符类型

```
CREATE TABLE test1 (a character(4));
INSERT INTO test1 VALUES ('ok');
SELECT a, char_length(a) FROM test1; -- 1
  a   | char_length
------+-------------
 ok   |           2

CREATE TABLE test2 (b varchar(5));
INSERT INTO test2 VALUES ('ok');
INSERT INTO test2 VALUES ('good      ');
INSERT INTO test2 VALUES ('too long');
ERROR:  value too long for type character varying(5)
INSERT INTO test2 VALUES ('too long'::varchar(5)); -- explicit
 truncation
SELECT b, char_length(b) FROM test2;
   b   | char_length
-------+-------------
 ok    |           2
 good  |           5
 too l |           5
```

1 函数char_length在第 9.4 节中讨论。

PostgreSQL 中还有两种固定长度字符类型，如表 8.5所示。name 类型仅用于在内部系统目录中存储标识符，并非供一般用户使用。它的长度目前定义为 64 字节（63 个可用字符加结束符），但在 C 源代码中应使用常量 NAMEDATALEN 来引用。这个长度是在编译时设定的（因此可以针对特殊用途调整）；默认最大长度在未来版本中可能会变化。类型 "char"（注意带引号）不同于 char(1)，因为它只使用 1 个字节存储。它在系统目录中被用作一种简单的枚举类型。

表 8.5. 特殊字符类型

名字	存储尺寸	描述
"char"	1字节	单字节内部类型
name	64字节	用于对象名的内部类型

8.4. 二进制数据类型

bytea数据类型允许存储二进制串，参见表 8.6。

表 8.6. 二进制数据类型

名字	存储尺寸	描述
bytea	1或4字节外加真正的二进制串	变长二进制串

二进制串是八位组（或字节）的序列。二进制串与字符串有两点区别。首先，
二进制串明确允许存储值为零的字节以及其他“不可打印”的字节（通常指 32 到 126 范围之外的字节）。而字符串不允许零字节，
也不允许那些按数据库所选字符集编码看属于非法的其他字节值或字节序列。其次，
对二进制串的操作处理的是实际字节，而字符串的处理则取决于区域设置。简单地说，
二进制串适合存储程序员视为“裸字节”的数据，而字符串适合存储文本。

`bytea`类型支持两种用于输入和输出的外部格式：PostgreSQL历史的“转义”
格式和“十六进制”格式。在输入时这两种格式总是会被接受。
输出格式则取决于配置参数bytea_output，其默认值为十六进制（注意十六进制格式是在PostgreSQL 9.0中被引入的，早期的版本和某些工具无法理解它）。

SQL标准定义了一种不同的二进制串类型，叫做`BLOB`或者`BINARY LARGE OBJECT`。其输入格式和`bytea`不同，但是提供的函数和操作符大多一样。

8.4.1. bytea的十六进制格式

“十六进制”格式把二进制数据编码为每字节两个十六进制数字，最高有效半字节在前。
整个串以前缀 `\x` 开头（以便与转义格式区分）。在某些上下文中，这个开头的反斜线可能需要通过双写进行转义，具体场合与转义格式中的反斜线必须双写的场合相同；细节见下文。
十六进制数字可以使用大写或小写，并且每两个数字组成一组，
组与组之间允许出现空白（但组内以及起始的 `\x` 序列中不能有空白）。十六进制格式与大量外部应用和协议兼容，并且通常比转义格式转换得更快，因此更推荐使用。

例如：

```
SELECT E'\\xDEADBEEF';
```

8.4.2. bytea的转义格式

“转义”格式是 `bytea` 类型在 PostgreSQL 中的传统格式。它采用把二进制串表示为 ASCII 字符序列的方式，同时把那些不能表示为 ASCII 字符的字节转换为特殊的转义序列。如果从应用角度看，把字节当作字符表示是合理的，那么这种表示法会比较方便。但在实际中它往往让人困惑，
因为它模糊了二进制串和字符串之间的区别，而且所选用的转义机制也比较笨拙。
所以对大多数新应用来说，最好避免使用这种格式。

在转义格式中输入 `bytea` 值时，某些字节值必须转义，而所有字节值都可以转义。通常，转义一个字节的方法是把它写成三位八进制值，
并在前面加一个反斜线（如果使用转义字符串语法把该值写为字面量，则加两个反斜线）。
反斜线本身（字节值 92）也可以写成双反斜线。表 8.7展示了必须转义的字符，并给出了可用的替代转义序列。

表 8.7. bytea字面量中需要转义的字节

十进制字节值	描述	转义输入表示	示例	输出表示
0	0字节	`E'\\000'`	`SELECT E'\\000'::bytea;`	`\000`
39	单引号	`''''`或`E'\\047'`	`SELECT E'\''::bytea;`	`'`
92	反斜线	`E'\\\\'`或`E'\\134'`	`SELECT E'\\\\'::bytea;`	`\\`
0到31和127到255	“不可打印的”字节	`E'\\xxx'`（八进制值）	`SELECT E'\\001'::bytea;`	`\001`

是否必须转义这些不可打印字节，会因区域设置不同而有所差异。在某些情况下，
你可以不转义它们。请注意，表 8.7中每个示例的结果都恰好是一个字节长，尽管其输出表示有时多于一个字符。

如表 8.7所示之所以需要多个反斜线，原因在于以字符串字面量形式书写的输入串必须经过 PostgreSQL 服务器的两个解析阶段。每对反斜线中的第一个会被字符串常量解析器解释为转义字符（假定使用转义字符串语法）而被消耗，留下该对中的第二个反斜线。
（可以使用美元引用的字符串来避免这一层转义。）剩下的反斜线随后被 `bytea` 输入函数识别为某个三位八进制值的开始，或者是对另一个反斜线的转义。例如，
以字符串字面量形式传给服务器的 `E'\\001'` 在经过转义字符串解析器之后变为 `\001`。然后 `\001` 被送到 `bytea` 输入函数，在那里被转换为十进制值为 1 的单个八位组。注意，单引号字符不会被 `bytea` 特殊对待，因此它遵循字符串常量的常规规则。（另见第 4.1.2.1 节。）

`bytea` 的字节在输出时有时会被转义。一般而言，每个“不可打印”字节会被转换成等价的三位八进制值，并在前面加一个反斜线。大多数“可打印”
字节则会按其在客户端字符集中的标准表示输出。十进制值 92（反斜线）的字节在输出中会被双写。详情见表 8.8。

表 8.8. bytea输出转义字节

十进制字节值	描述	转义的输出表示	示例	输出结果
92	反斜线	`\\`	`SELECT E'\\134'::bytea;`	`\\`
0到31和127到255	“不可打印的”字节	`\`*xxx*（八进制值）	`SELECT E'\\001'::bytea;`	`\001`
32到126	“可打印的”字节	客户端字符集表示	`SELECT E'\\176'::bytea;`	`~`

取决于你所使用的 PostgreSQL 前端，在转义和反转义 `bytea` 串时可能还需要做额外工作。例如，如果你的接口会自动转换换行和回车，那么你可能还需要对它们进行转义。

8.5. 日期/时间类型

PostgreSQL 支持完整的一组 SQL 日期和时间类型，如表 8.9所示。这些数据类型上可用的操作见第 9.9 节。

表 8.9. 日期/时间类型

名字	存储尺寸	描述	最小值	最大值	解析度
`timestamp [ (p) ] [ without time zone ]`	8字节	包括日期和时间（无时区）	4713 BC	294276 AD	1微秒 / 14位
`timestamp [ (p) ] with time zone`	8字节	包括日期和时间，有时区	4713 BC	294276 AD	1微秒 / 14位
`date`	4字节	日期（没有一天中的时间）	4713 BC	5874897 AD	1日
`time [ (p) ] [ without time zone ]`	8字节	一天中的时间（无日期）	00:00:00	24:00:00	1微秒 / 14位

名字	存储尺寸	描述	最小值	最大值	解析度
time [(*p*)] with time zone	12字节	仅一天中的时间，带时区	00:00:00+1459	24:00:00-1459	1微秒 / 14位
interval [*fields*] [(*p*)]	16字节	时间间隔	-178000000年	178000000年	1微秒 / 14位

> **注意**
>
> SQL 标准要求仅写 timestamp 时，应等效于 timestamp without time zone，而 PostgreSQL 也遵循这种行为。（7.3 之前的版本将其视为 timestamp with time zone。）timestamptz 被接受为 timestamp with time zone 的简写，这是 PostgreSQL 的扩展。

time、timestamp 和 interval 都接受一个可选精度值 *p*，用来指定在秒字段中保留多少位小数。默认情况下，对精度没有显式上界。对于 timestamp 和 interval 类型，*p* 的允许范围是 0 到 6。

> **注意**
>
> 当 timestamp 值存储为八字节整数（目前的默认方式）时，在整个取值范围内都可以得到微秒精度。而当 timestamp 值改为存储为双精度浮点数（一个已废弃的编译期选项）时，有效精度上界可能小于 6。timestamp 值存储为相对于 2000-01-01 午夜之前或之后的秒数。当 timestamp 值采用浮点数实现时，对于距 2000-01-01 几年之内的日期可以达到微秒精度，但对更远的日期精度会下降。注意，使用浮点日期时间可以表示比上表所列更大的 timestamp 取值范围：从公元前 4713 年直到公元 5874897 年。
>
> 同一个编译期选项还决定 time 和 interval 值存储为浮点数还是八字节整数。在浮点存储的情况下，大的 interval 值会随着间隔长度的增大而降低精度。

对于 time 类型，采用八字节整数存储时，*p* 的允许范围是 0 到 6；采用浮点存储时，允许范围是 0 到 10。

interval 类型还有一个附加选项，可以通过写出下面这些短语之一来限制所存储字段的集合：

```
YEAR
MONTH
DAY
HOUR
MINUTE
SECOND
YEAR TO MONTH
DAY TO HOUR
DAY TO MINUTE
DAY TO SECOND
HOUR TO MINUTE
HOUR TO SECOND
```

MINUTE TO SECOND

注意，如果同时指定了 *fields* 和 *p*，那么 *fields* 必须包含 SECOND，因为精度只作用于秒。

time with time zone 类型由 SQL 标准定义，但其定义具有一些会让人怀疑其实用性的特性。在大多数情况下，date、time、timestamp without time zone 和 timestamp with time zone 的组合，就足以提供任何应用所需的完整日期/时间功能。

类型abstime和reltime是内部使用的低精度类型。不推荐在应用中使用这些类型；这些内部类型可能在将来的版本中消失。

8.5.1. 日期/时间输入

日期和时间输入几乎接受任何合理的格式，包括 ISO 8601、与 SQL 兼容的格式、传统 POSTGRES 格式等。对于某些格式，日期输入中日、月、年的顺序可能存在歧义，因此支持指定这些字段的预期顺序。将DateStyle参数设置为 MDY，表示采用月-日-年解释；设置为 DMY 表示采用日-月-年解释；而 YMD 表示采用年-月-日解释。

PostgreSQL 在处理日期/时间输入方面，比 SQL 标准要求的更灵活。有关日期/时间输入的精确解析规则，以及可识别的文本字段（包括月份、星期几和时区），请参见附录 B。

请记住，任何日期或时间字面值输入都必须像文本字符串一样用单引号括起来。更多信息请参见第 4.1.2.7 节。SQL 要求使用下列语法：

type [(*p*)] '*value*'

其中 *p* 是可选的精度说明，给出秒字段中保留的小数位数。精度可用于 time、timestamp 和 interval 类型，允许的取值范围见上文。如果在常量声明中没有指定精度，则默认采用该字面值本身的精度。

8.5.1.1. 日期

表 8.10显示了date类型可能的输入方式。

表 8.10. 日期输入

示例	描述
1999-01-08	ISO 8601；任何模式下的1月8日（推荐格式）
January 8, 1999	在任何datestyle输入模式下都无歧义
1/8/1999	MDY模式中的1月8日；DMY模式中的8月1日
1/18/1999	MDY模式中的1月18日；在其他模式中被拒绝
01/02/03	MDY模式中的2003年1月2日；DMY模式中的2003年2月1日；YMD模式中的2001年2月3日
1999-Jan-08	任何模式下的1月8日
Jan-08-1999	任何模式下的1月8日
08-Jan-1999	任何模式下的1月8日
99-Jan-08	YMD模式中的1月8日，否则错误
08-Jan-99	1月8日，除了在YMD模式中错误
Jan-08-99	1月8日，除了在YMD模式中错误
19990108	ISO 8601；任何模式中的1999年1月8日

示例	描述
990108	ISO 8601；任何模式中的1999年1月8日
1999.008	年和一年中的日子
J2451187	儒略日
January 8, 99 BC	公元前99年

8.5.1.2. 时间

一天中的时间类型包括 `time [ (p) ] without time zone` 和 `time [ (p) ] with time zone`。单独写 `time` 等效于 `time without time zone`。

这些类型的有效输入由一个一天中的时间，加上一个可选时区组成（见表 8.11和表 8.12）。如果在 `time without time zone` 的输入中指定了时区，它会被静默忽略。你也可以指定一个日期，但它同样会被忽略，除非你使用了涉及夏令时规则的时区名称，例如 `America/New_York`。在这种情况下，必须指定日期，以便确定应适用标准时间还是夏令时。
相应的时区偏移会被记录到 `time with time zone` 值中。

表 8.11. 时间输入

示例	描述
`04:05:06.789`	ISO 8601
`04:05:06`	ISO 8601
`04:05`	ISO 8601
`040506`	ISO 8601
`04:05 AM`	和04:05一样，AM并不影响值
`04:05 PM`	和16:05一样，输入的小时必须为 <= 12
`04:05:06.789-8`	ISO 8601
`04:05:06-08:00`	ISO 8601
`04:05-08:00`	ISO 8601
`040506-08`	ISO 8601
`04:05:06 PST`	缩写指定的时区
`2003-04-12 04:05:06 America/New_York`	全名指定的时区

表 8.12. 时区输入

示例	描述
`PST`	缩写（太平洋标准时间）
`America/New_York`	完整时区名
`PST8PDT`	POSIX风格的时区声明
`-8:00`	PST的ISO-8601偏移
`-800`	PST的ISO-8601偏移
`-8`	PST的ISO-8601偏移
`zulu`	UTC的军方缩写
`z`	`zulu`的缩写形式

关于如何指定时区，参见第 8.5.3 节。

8.5.1.3. 时间戳

时间戳类型的有效输入由一个日期和时间的串接组成，后面跟着一个可选时区，以及一个可选的 AD 或 BC（另外，AD/BC 也可以出现在时区前面，但这种顺序并不推荐）。因此：

```
1999-01-08 04:05:06
```

和：

```
1999-01-08 04:05:06 -8:00
```

都是遵循 ISO 8601 标准的有效值。另外，广泛使用的下列格式：

```
January 8 04:05:06 1999 PST
```

也被支持。

按照SQL标准，timestamp without time zone和timestamp with time zone字面量的区别在于，时间后是否有“+”或“-”符号及其后的时区偏移。因此，按照该标准，

```
TIMESTAMP '2004-10-19 10:23:54'
```

是timestamp without time zone，而

```
TIMESTAMP '2004-10-19 10:23:54+02'
```

是timestamp with time zone。PostgreSQL在确定字符串字面量的类型之前，从不检查其内容，因此会把上述两者都视为timestamp without time zone。为确保字面量被视为timestamp with time zone，应为它显式指定正确类型：

```
TIMESTAMP WITH TIME ZONE '2004-10-19 10:23:54+02'
```

若字面量已经被确定为timestamp without time zone，PostgreSQL会静默忽略任何时区标记。也就是说，所得值来自输入值中的日期/时间字段，不会根据时区调整。

对于timestamp with time zone，内部存储的值始终使用 UTC（协调世界时，传统上称为格林尼治标准时间，GMT）。对于显式指定时区的输入值，会使用该时区适当的偏移将它转换为 UTC。如果输入字符串中没有说明时区，则假定它属于系统timezone参数指定的时区，并使用timezone时区的偏移将它转换为 UTC。

当输出一个 timestamp with time zone 值时，它总会从 UTC 转换到当前 timezone 时区，并显示为该时区的本地时间。若要查看其他时区的时间，可以修改 timezone，或者使用 AT TIME ZONE 构造（见第 9.9.3 节）。

在 timestamp without time zone 和 timestamp with time zone 之间转换时，通常假定 timestamp without time zone 值应被解释为，或输出为，timezone 本地时间。要为该转换指定不同的时区，可以使用 AT TIME ZONE。

8.5.1.4. 特殊值

为了方便起见，PostgreSQL 支持若干特殊的日期/时间输入值，如表 8.13所示。infinity 和 -infinity 在系统内部有特殊表示，并且输出时会保持不变；其余值则只是记法上的简写，在读取时会被转换成普通日期/时间值。（特别是，now 及相关字符串在被读取后会立刻转换成某个特定的时间值。）这些值在 SQL 命令中作为常量使用时，都必须用单引号括起来。

表 8.13. 特殊日期/时间输入

输入串	合法类型	描述
epoch	date, timestamp	1970-01-01 00:00:00+00（Unix系统时间0）
infinity	date, timestamp	晚于所有其他时间戳
-infinity	date, timestamp	早于所有其他时间戳
now	date, time, timestamp	当前事务的开始时间
today	date, timestamp	今天午夜
tomorrow	date, timestamp	明天午夜
yesterday	date, timestamp	昨天午夜
allballs	time	00:00:00.00 UTC

以下与 SQL 兼容的函数也可用于获取相应数据类型的当前时间值：CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP、LOCALTIME 和 LOCALTIMESTAMP。后四个函数接受可选的亚秒级精度说明。（参见第 9.9.4 节。）请注意，这些是 SQL 函数，不会在日期/时间输入字符串中被识别。

8.5.2. 日期/时间输出

日期/时间类型的输出格式可以设为四种样式之一：ISO 8601、SQL（Ingres）、传统的 POSTGRES（Unix date 格式）或 German。默认是 ISO 格式。（SQL 标准要求使用 ISO 8601 格式。之所以存在名为“SQL”的输出格式，只是历史原因。）表 8.14展示了各种输出样式的示例。date 和 time 类型的输出当然只包含示例中的日期部分或时间部分。

表 8.14. 日期/时间输出风格

风格声明	描述	示例
ISO	ISO 8601/SQL标准	1997-12-17 07:37:16-08
SQL	传统样式	12/17/1997 07:37:16.00 PST
POSTGRES	原始样式	Wed Dec 17 07:37:16 1997 PST
German	地区样式	17.12.1997 07:37:16.00 PST

SQL和POSTGRES风格中，如果指定了 DMY 字段顺序，“日”将出现在“月”之前，否则“月”出现在“日”之前（有关该设置如何影响输入值的解释，请参考第 8.5.1 节）。表 8.15给出了一个示例。

表 8.15. 日期顺序习惯

datestyle设置	输入顺序	示例输出
SQL, DMY	日/月/年	17/12/1997 15:37:16.00 CET
SQL, MDY	月/日/年	12/17/1997 07:37:16.00 PST
Postgres, DMY	日/月/年	Wed 17 Dec 07:37:16 1997 PST

用户可以通过 SET datestyle 命令、postgresql.conf 配置文件中的DateStyle参数，或者服务器端或客户端上的 PGDATESTYLE 环境变量来选择日期/时间样式。格式化函数to_char（见第 9.8 节）也可以作为一个更灵活的方式来格式化日期/时间输出。

8.5.3. 时区

时区及其约定不仅受地球几何形状影响，也受政治决定影响。世界各地的时区在 20 世纪逐渐趋于标准化，但仍然容易发生任意变化，尤其是夏令时规则方面。PostgreSQL 使用广泛采用的 IANA（Olson）时区数据库来获取历史时区规则信息。对于未来时间，则假定某个时区最新已知的规则会无限期地持续下去。

PostgreSQL 努力在典型用法上与 SQL 标准定义保持兼容。不过，SQL 标准在日期和时间类型及其能力方面存在一些奇怪的混搭。两个显而易见的问题是：

- 尽管 `date` 类型不能有关联的时区，`time` 类型却可以。但现实世界中的时区如果不同时关联日期和时间，几乎没有意义，因为偏移量可能会随着夏令时切换而在一年中发生变化。

- 默认时区被指定为相对于 UTC 的一个固定数值偏移。因此，在跨越 DST 边界做日期/时间算术时，根本无法适应夏令时变化。

为了克服这些困难，我们建议在使用时区时采用同时包含日期和时间的日期/时间类型。我们不建议使用 `time with time zone` 类型（尽管 PostgreSQL 出于兼容旧应用以及遵循 SQL 标准的考虑而支持它）。PostgreSQL 对于任何只包含日期或时间的类型，都会假定其使用本地时区。

在系统内部，所有带时区的日期和时间都以 UTC 存储。显示给客户端之前，它们会被转换为由timezone配置参数指定的本地时间。

PostgreSQL 允许使用三种不同形式来指定时区：

- 完整时区名称，例如 `America/New_York`。识别到的时区名称列在 `pg_timezone_names` 视图中（见第 45.67 节）。PostgreSQL 为此使用广泛采用的 IANA 时区数据，因此同样的时区名称通常也会被其他软件识别。

- 时区缩写，例如 `PST`。与完整时区名称不同，这种指定方式只定义了一个特定的 UTC 偏移，而完整时区名称还可能隐含一套夏令时切换规则。识别到的缩写列在 `pg_timezone_abbrevs` 视图中（见第 45.66 节）。你不能把timezone或log_timezone配置参数设置为时区缩写，但可以在日期/时间输入值中以及与 `AT TIME ZONE` 操作符一起使用缩写。

- 除了时区名称和缩写之外，PostgreSQL 还接受形如 *`STDoffset`* 或 *`STDoffsetDST`* 的 POSIX 风格时区说明，其中 *`STD`* 是时区缩写，*`offset`* 是以小时计、相对 UTC 向西的数字偏移，而 *`DST`* 是可选的夏令时缩写，假定它表示比给定偏移早一小时。例如，如果 `EST5EDT` 还不是一个已被识别的时区名称，它也会被接受，并且在功能上等同于美国东海岸时间。当给出夏令时时区名称时，假定它按照 IANA 时区数据库 `posixrules` 条目中所用的相同夏令时转换日期规则使用。在标准的 PostgreSQL 安装中，`posixrules` 与 `US/Eastern` 相同，因此 POSIX 风格的时区说明遵循美国的夏令时规则。如果需要，可以通过替换 `posixrules` 文件来调整这一行为。

简而言之，缩写和完整名称的区别是：缩写表示特定的 UTC 偏移，而许多完整名称隐含本地夏令时规则，因此有两个可能的 UTC 偏移。例如，`2014-06-04 12:00 America/New_York`表示纽约当地时间的中午，在该日期使用的是东部夏令时间（UTC-4）。因此，`2014-06-04 12:00 EDT`指定相同的时刻。但是，`2014-06-04 12:00 EST`指定东部标准时间（UTC-5）的中午，而不管该日期是否名义上实行夏令时。

应当当心，POSIX 风格的时区特性可能导致静默接受伪劣输入，因为对时区缩写的合理性没有任何检查。例如，`SET TIMEZONE TO FOOBAR0` 也能工作，使系统实际上使用一个相当古怪的 UTC 缩写。另一个需要记住的问题是，在 POSIX 时区名中，正偏移用于格林尼治以西的地点；而在其他所有地方，PostgreSQL 都遵循 ISO-8601 约定，正的时区偏移位于格林尼治以东。

更复杂的是，一些司法辖区在不同时间使用同一时区缩写来表示不同的 UTC 偏移；例如在莫斯科，`MSK` 在某些年份表示 UTC+3，在另一些年份则表示 UTC+4。PostgreSQL 会按照该缩写在所

给日期上的含义（或最近一次的含义）来解释这类缩写；但与上面的 `EST` 例子一样，这并不一定等同于该日期的当地民用时间。

无论哪种形式，时区名称及其缩写都不区分大小写。（这是对 PostgreSQL 8.2 之前版本的一项改动；在那些版本中，时区名在某些环境下区分大小写，而在另一些环境下则不区分。）

时区名称和缩写并不是硬编码在服务器中的；它们来自安装目录下 `.../share/timezone/` 和 `.../share/timezonesets/` 子目录中的配置文件（见第 B.3 节）。

timezone配置参数可以在 `postgresql.conf` 文件中设置，也可以通过第 18 章中说明的其他标准方式设置。另外，还有几种特殊的设置方法：

- 如果没有在 `postgresql.conf` 中或以服务器命令行选项的方式指定 `timezone`，服务器会尝试使用 `TZ` 环境变量的值作为默认时区。如果 `TZ` 没有定义，或者不是 PostgreSQL 已知的任何时区名称，服务器会通过检查 C 库函数 `localtime()` 的行为来确定操作系统的默认时区。默认时区选取的是与 PostgreSQL 已知时区中最接近的匹配。（如果未指定，这些规则也用于选择log_timezone的默认值。）
- SQL 命令 `SET TIME ZONE` 用于设置会话的时区。它是 `SET TIMEZONE TO` 的另一种写法，语法上更符合 SQL 规范。
- `PGTZ` 环境变量会被 libpq 客户端用于在连接到服务器时发送一条 `SET TIME ZONE` 命令。

8.5.4. 间隔输入

`interval`值可以使用下列详细语法书写：

`[@] quantity unit [quantity unit...] [direction]`

其中*quantity*是一个数字（可以带有符号）；*unit*是`microsecond`、`millisecond`、`second`、`minute`、`hour`、`day`、`week`、`month`、`year`、`decade`、`century`、`millennium` 或它们的缩写或复数形式；*direction* 可以是 `ago` 或为空。at 符号（`@`）只是可选的噪声。不同单位的数量会按适当的符号规则隐式相加。`ago` 会将所有字段取反。如果IntervalStyle被设置为 `postgres_verbose`，该语法也会用于间隔输出。

日、小时、分钟和秒的数量也可以不写显式单位标记。例如，`'1 12:59:10'` 会被读作 `'1 day 12 hours 59 min 10 sec'`。同样，年和月的组合也可以用一个连字符表示，例如 `'200-10'` 会被读作 `'200 years 10 months'`。（事实上，这些较短形式正是 SQL 标准唯一允许的形式，并且在 `IntervalStyle` 被设置为 `sql_standard` 时也用于输出。）

间隔值也可以写成 ISO 8601 时间间隔，使用标准第 4.4.3.2 节的“带标志符的格式”，或第 4.4.3.3 节的“替代格式”。带标志符的格式如下：

`P quantity unit [ quantity unit ...] [ T [ quantity unit ...]]`

字符串必须以 `P` 开头，并且可以包含一个 `T` 来引出一天中的时间单位。可用的单位缩写见表 8.16。单位可以省略，也可以按任意顺序出现，但小于一天的单位必须出现在 `T` 之后。特别是，`M` 的含义取决于它是在 `T` 之前还是之后。

表 8.16. ISO 8601 间隔单位缩写

缩写	含义
Y	年
M	月（在日期部分中）

缩写	含义
W	周
D	日
H	小时
M	分钟 (在时间部分中)
S	秒

如果使用替代格式：

```
P [ years-months-days ] [ T hours:minutes:seconds ]
```

串必须以`P`开始，并且一个`T`分隔间隔的日期和时间部分。其值按照类似于 ISO 8601日期的数字给出。

当编写带有 *fields* 说明的间隔常量，或者把字符串赋给一个定义时带有 *fields* 说明的间隔列时，未标记数量的解释方式取决于 *fields*。例如，`INTERVAL '1' YEAR` 会被读作 1 年，而 `INTERVAL '1'` 表示 1 秒。此外，位于 *fields* 说明所允许的最小字段“右侧”的字段值会被静默丢弃。例如，写 `INTERVAL '1 day 2:03:04' HOUR TO MINUTE` 会导致秒字段被丢弃，而不是日字段。

根据 SQL 标准，一个间隔值的所有字段都必须带有相同符号，因此一个前导负号会作用于所有字段；例如间隔字面值 `'-1 2:03:04'` 中的负号会同时作用于日、小时、分钟和秒。PostgreSQL 允许各字段具有不同符号，并且传统上认为文本表示中的每个字段都带有各自独立的符号，因而在这个例子里，小时、分钟和秒部分会被视为正值。如果 `IntervalStyle` 被设置为 `sql_standard`，则会把前导符号视为作用于所有字段（但前提是没有出现额外符号）；否则将采用传统的 PostgreSQL 解释。为了避免混淆，我们建议只要有任何字段为负值，就为每个字段都显式写出符号。

在内部，`interval` 值以月、日和秒存储。这是因为一个月的天数各不相同，而涉及夏令时调整时，一天可能有 23 或 25 小时。月和日字段为整数，秒字段则可以存储小数秒。由于时间间隔通常由常量字符串或`timestamp`相减创建，这种存储方式在大多数情况下都很有效。函数 `justify_days`和 `justify_hours`可用于调整超出正常范围的日和小时。

在冗长输入格式中，以及在一些更紧凑输入格式的某些字段中，字段值可以带小数部分；例如 `'1.5 week'`或`'01:02:03.45'`。这样的输入会被转换为适当数量的月、日和秒来存储。当转换结果出现小数的月数或天数时，小数部分会按 1 月 = 30 天、1 天 = 24 小时的换算系数计入更低阶的字段。例如，`'1.5 month'`会变成 1 个月零 15 天。输出时，只有秒会显示小数部分。

表 8.17展示了一些有效`interval`输入的示例。

表 8.17. 间隔输入

示例	描述
1-2	SQL标准格式：1年2个月
3 4:05:06	SQL标准格式：3日4小时5分钟6秒
1 year 2 months 3 days 4 hours 5 minutes 6 seconds	传统Postgres格式：1年2个月3日4小时5分钟6秒钟
P1Y2M3DT4H5M6S	“带标志符的” ISO 8601 格式：含义同上
P0001-02-03T04:05:06	ISO 8601 的“替代格式”：含义同上

8.5.5. 间隔输出

间隔类型的输出格式可以通过 `SET intervalstyle` 命令设置为以下四种风格之一：`sql_standard`、`postgres`、`postgres_verbose` 或 `iso_8601`。默认值是 `postgres` 格式。表 8.18展示了每种输出风格的示例。

如果间隔值满足 SQL 标准的限制条件（仅有年-月或仅有日-时间，且不混合正负分量），`sql_standard` 风格会生成符合 SQL 标准的间隔字面值输出。否则，输出将表现为一个标准的年-月字面值串后跟一个日-时间字面值串，并显式加上符号，以消除正负混合间隔的歧义。

`postgres` 样式的输出与 PostgreSQL 8.4 之前版本在DateStyle参数设为 `ISO` 时的输出一致。

`postgres_verbose` 样式的输出与 PostgreSQL 8.4 之前版本在 `DateStyle` 参数设为非 `ISO` 输出时的结果一致。

`iso_8601` 风格的输出符合 ISO 8601 标准 4.4.3.2 节描述的“带标志符的格式”。

表 8.18. 间隔输出风格示例

风格声明	年-月间隔	日-时间间隔	混合间隔
`sql_standard`	1-2	3 4:05:06	-1-2 +3 -4:05:06
`postgres`	1 year 2 mons	3 days 04:05:06	-1 year -2 mons +3 days -04:05:06
`postgres_verbose`	@ 1 year 2 mons	@ 3 days 4 hours 5 mins 6 secs	@ 1 year 2 mons -3 days 4 hours 5 mins 6 secs ago
`iso_8601`	P1Y2M	P3DT4H5M6S	P-1Y-2M3DT-4H-5M-6S

8.5.6. 内部实现

PostgreSQL 在所有日期/时间计算中使用儒略日。这样具有一个有用的特性：在假定一年长度为 365.2425 天的前提下，能够正确计算从公元前 4713 年到遥远未来的日期。

19 世纪之前的日期惯例读起来很有意思，但不够一致，不值得写入日期/时间处理器中。

8.6. 布尔类型

PostgreSQL提供标准的SQL类型`boolean`，参见表 8.19。`boolean`可以有多个状态：“true（真）”、“false（假）”和第三种状态“unknown（未知）”，未知状态由SQL空值表示。

表 8.19. 布尔数据类型

名字	存储尺寸	描述
`boolean`	1字节	状态为真或假

“真”状态的有效字面值有：

```
TRUE
't'
'true'
'y'
'yes'
'on'
'1'
```

对于“假”状态，可以使用下列值：

```
FALSE
'f'
'false'
'n'
'no'
'off'
'0'
```

前导或尾随空白会被忽略，并且大小写不敏感。关键字 `TRUE` 和 `FALSE` 是首选的（符合 SQL）用法。

例 8.2表明，`boolean` 值使用字母 `t` 和 `f` 输出。

例 8.2. 使用boolean类型

```
CREATE TABLE test1 (a boolean, b text);
INSERT INTO test1 VALUES (TRUE, 'sic est');
INSERT INTO test1 VALUES (FALSE, 'non est');
SELECT * FROM test1;
 a |    b
---+---------
 t | sic est
 f | non est

SELECT * FROM test1 WHERE a;
 a |    b
---+---------
 t | sic est
```

8.7. 枚举类型

枚举（enum）类型是由一个静态的有序值集合构成的数据类型。
它们等效于很多编程语言所支持的`enum`类型。枚举类型的一个示例可以是一周中的各天，
或者一个数据的状态值集合。

8.7.1. 枚举类型的声明

可以使用CREATE TYPE命令创建枚举类型，例如：

```
CREATE TYPE mood AS ENUM ('sad', 'ok', 'happy');
```

创建之后，该 enum 类型就可以像任何其他类型一样在表定义和函数定义中使用：

```
CREATE TYPE mood AS ENUM ('sad', 'ok', 'happy');
CREATE TABLE person (
    name text,
    current_mood mood
);
INSERT INTO person VALUES ('Moe', 'happy');
SELECT * FROM person WHERE current_mood = 'happy';
 name | current_mood
```

```
------+--------------
 Moe  | happy
(1 row)
```

8.7.2. 排序

enum 类型中各值的顺序，就是创建该类型时列出的顺序。枚举支持所有标准比较操作符以及相关聚合函数。例如：

```
INSERT INTO person VALUES ('Larry', 'sad');
INSERT INTO person VALUES ('Curly', 'ok');
SELECT * FROM person WHERE current_mood > 'sad';
 name  | current_mood
-------+--------------
 Moe   | happy
 Curly | ok
(2 rows)

SELECT * FROM person WHERE current_mood > 'sad' ORDER BY current_mood;
 name  | current_mood
-------+--------------
 Curly | ok
 Moe   | happy
(2 rows)

SELECT name
FROM person
WHERE current_mood = (SELECT MIN(current_mood) FROM person);
 name
-------
 Larry
(1 row)
```

8.7.3. 类型安全性

每一种枚举数据类型都是独立的并且不能和其他枚举类型相比较。看这样一个示例：

```
CREATE TYPE happiness AS ENUM ('happy', 'very happy', 'ecstatic');
CREATE TABLE holidays (
    num_weeks integer,
    happiness happiness
);
INSERT INTO holidays(num_weeks,happiness) VALUES (4, 'happy');
INSERT INTO holidays(num_weeks,happiness) VALUES (6, 'very happy');
INSERT INTO holidays(num_weeks,happiness) VALUES (8, 'ecstatic');
INSERT INTO holidays(num_weeks,happiness) VALUES (2, 'sad');
ERROR:  invalid input value for enum happiness: "sad"
SELECT person.name, holidays.num_weeks FROM person, holidays
  WHERE person.current_mood = holidays.happiness;
ERROR:  operator does not exist: mood = happiness
```

如果你确实需要做这类事情，可以编写自定义操作符，或者在查询中添加显式类型转换：

```
SELECT person.name, holidays.num_weeks FROM person, holidays
```

```
  WHERE person.current_mood::text = holidays.happiness::text;
 name | num_weeks
------+-----------
 Moe  |         4
(1 row)
```

8.7.4. 实现细节

一个枚举值在磁盘上占据4个字节。一个枚举值的文本标签的长度受限于NAMEDATALEN设置，该设置被编译在PostgreSQL中，在标准编译下它表示最多63字节。

枚举标签是大小写敏感的，因此'happy'与'HAPPY'是不同的。标签中的空白也是有意义的。

从内部枚举值到文本标签的转换关系被保存在系统目录pg_enum中。直接查询该目录可能会有所帮助。

8.8. 几何类型

几何数据类型表示二维的空间物体。表 8.20展示了PostgreSQL中可以用的几何类型。最基础的类型是点，它构成了所有其他类型的基础。

表 8.20. 几何类型

名字	存储尺寸	表示	描述
point	16字节	(x,y)	平面上的点
line	32字节	((x1,y1),(x2,y2))	无限直线（未完全实现）
lseg	32字节	有限线段	((x1,y1),(x2,y2))
box	32字节	矩形框	((x1,y1),(x2,y2))
path	16+16n字节	封闭路径（类似于多边形）	((x1,y1),...)
path	16+16n字节	开放路径	[(x1,y1),...]
polygon	40+16n字节	多边形（类似于封闭路径）	((x1,y1),...)
circle	24字节	圆	<(x,y),r>（中心点和半径）

我们提供了丰富的函数和操作符来进行各种几何操作，例如缩放、平移、旋转以及求交等，详见第 9.11 节。

8.8.1. 点

点是几何类型的基本二维构造块。用下面任一种语法描述point类型的值：

```
( x , y )
  x , y
```

其中x和y分别是坐标，都是浮点数。

点使用第一种语法输出。

8.8.2. 线段

线段（lseg）用成对的点来表示。lseg类型的值用下面任一种语法指定：

```
[ ( x1 , y1 ) , ( x2 , y2 ) ]
( ( x1 , y1 ) , ( x2 , y2 ) )
  ( x1 , y1 ) , ( x2 , y2 )
    x1 , y1   ,   x2 , y2
```

其中(*x1*,*y1*) 和 (*x2*,*y2*) 是线段的端点。

线段使用第一种语法输出。

8.8.3. 方框

方框用其对角的点对表示。box类型的值使用下面任一种语法指定：

```
( ( x1 , y1 ) , ( x2 , y2 ) )
  ( x1 , y1 ) , ( x2 , y2 )
    x1 , y1   ,   x2 , y2
```

其中(*x1*,*y1*) 和 (*x2*,*y2*) 是方框的对角点。

方框使用第二种语法输出。

在输入时可以提供任意两个对角，但是值将根据需要被重新排列，
并依次存储为右上角和左下角。

8.8.4. 路径

路径由一系列连接的点组成。路径可能是开放的，
也就是认为列表中第一个点和最后一个点没有被连接起来；也可能是封闭的，
这时认为第一个和最后一个点被连接起来。

path类型的值用下面任一种语法指定：

```
[ ( x1 , y1 ) , ... , ( xn , yn ) ]
( ( x1 , y1 ) , ... , ( xn , yn ) )
  ( x1 , y1 ) , ... , ( xn , yn )
  ( x1 , y1   , ... ,   xn , yn )
    x1 , y1   , ... ,   xn , yn
```

其中的点是组成路径的线段的端点。方括弧（[]）表示一个开放的路径，圆括弧（()）
表示一个封闭的路径。如第三种到第五种语法所示，当最外面的圆括号被省略时，
路径将被假定为封闭。

路径的输出视情况使用第一种或第二种语法。

8.8.5. 多边形

多边形由点的列表（多边形的顶点）表示。多边形与封闭路径非常相似，但存储方式不同，
并且有自己的支持例程。

polygon类型的值用下列任一种语法指定：

```
( ( x1 , y1 ) , ... , ( xn , yn ) )
```

```
( x1 , y1 ) , ... , ( xn , yn )
( x1 , y1   , ... ,   xn , yn )
  x1 , y1   , ... ,   xn , yn
```

其中的点是组成多边形边界的线段的端点。

多边形的输出使用第一种语法。

8.8.6. 圆

圆由一个圆心和一个半径代表。`circle`类型的值用下面任一种语法指定：

```
< ( x , y ) , r >
( ( x , y ) , r )
  ( x , y ) , r
    x , y   , r
```

其中`(x,y)`是圆心，而`r`是圆的半径。

圆的输出用第一种语法。

8.9. 网络地址类型

PostgreSQL 提供了用于存储 IPv4、IPv6 和 MAC 地址的数据类型，如表 8.21所示。使用这些数据类型来存储网络地址，比使用纯文本类型更好，因为它们提供了输入错误检查以及专门的操作符和函数（见第 9.12 节）。

表 8.21. 网络地址类型

名字	存储尺寸	描述
`cidr`	7或19字节	IPv4和IPv6网络
`inet`	7或19字节	IPv4和IPv6主机以及网络
`macaddr`	6字节	MAC地址

在对 `inet` 或 `cidr` 数据类型排序时，IPv4 地址总是排在 IPv6 地址之前，即使后者是封装或映射 IPv4 地址所得的 IPv6 地址，例如 ::10.2.3.4 或 ::ffff:10.4.3.2。

8.9.1. `inet`

`inet` 在一个字段中保存 IPv4 或 IPv6 主机地址，以及可选的子网信息。子网由主机地址中包含的网络地址位数（即“网络掩码”）表示。如果网络掩码为 32 且地址是 IPv4，那么该值表示的不是子网，而只是单个主机。在 IPv6 中，地址长度为 128 位，因此 128 位表示一个唯一的主机地址。请注意，如果你只想接受网络地址，应使用 `cidr` 类型，而不是 `inet`。

该类型的输入格式为 *address/y*，其中 *address* 是 IPv4 或 IPv6 地址，而 *y* 是网络掩码中的位数。如果省略 */y* 部分，则 IPv4 的网络掩码取 32，IPv6 的网络掩码取 128，因此该值表示单个主机。显示时，如果 */y* 部分表示的是单个主机，则它不会被显示出来。

8.9.2. `cidr`

`cidr` 类型保存 IPv4 或 IPv6 网络说明。输入和输出格式遵循无类别域间路由（CIDR）约定。指定网络的格式为 *address/y*，其中 *address* 是网络的 IPv4 或 IPv6 表示，而 *y* 是网络掩码中的位

数。如果省略 *y*，则会按照旧式的分类网络编号系统假设来计算，但至少会大到足以覆盖输入中写出的所有八位组。若指定的网络地址在所给网络掩码右侧仍有置位比特，则会报错。

表 8.22展示了一些示例。

表 8.22. cidr类型输入示例

cidr输入	cidr输出	abbrev(cidr)
192.168.100.128/25	192.168.100.128/25	192.168.100.128/25
192.168/24	192.168.0.0/24	192.168.0/24
192.168/25	192.168.0.0/25	192.168.0.0/25
192.168.1	192.168.1.0/24	192.168.1/24
192.168	192.168.0.0/24	192.168.0/24
128.1	128.1.0.0/16	128.1/16
128	128.0.0.0/16	128.0/16
128.1.2	128.1.2.0/24	128.1.2/24
10.1.2	10.1.2.0/24	10.1.2/24
10.1	10.1.0.0/16	10.1/16
10	10.0.0.0/8	10/8
10.1.2.3/32	10.1.2.3/32	10.1.2.3/32
2001:4f8:3:ba::/64	2001:4f8:3:ba::/64	2001:4f8:3:ba::/64
2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1
::ffff:1.2.3.0/120	::ffff:1.2.3.0/120	::ffff:1.2.3/120
::ffff:1.2.3.0/128	::ffff:1.2.3.0/128	::ffff:1.2.3.0/128

8.9.3. inet 与 cidr

`inet` 和 `cidr` 两种数据类型的本质区别在于：`inet` 接受网络掩码右侧仍含有非零位的值，而 `cidr` 不接受。

> **提示**
>
> 如果你不喜欢 `inet` 或 `cidr` 值的输出格式，可以试试 `host`、`text` 和 `abbrev` 函数。

8.9.4. macaddr

`macaddr`类型存储 MAC 地址，例如以太网卡硬件地址（尽管 MAC 地址还用于其它用途）。可以接受下列格式的输入：

```
'08:00:2b:01:02:03'
'08-00-2b-01-02-03'
'08002b:010203'
'08002b-010203'
'0800.2b01.0203'
'08002b010203'
```

这些示例指定的都是同一个地址。十六进制数字 a 到 f 可使用大小写任意形式。输出总是采用上面展示的第一种形式。

IEEE标准802-2001指定所示的第二种形式（带连字符）作为MAC地址的规范形式，并指定第一种形式（带冒号）作为位反转表示法，因此 08-00-2b-01-02-03 = 01:00:4D:08:04:0C。这种约定现在被广泛忽略，仅适用于过时的网络协议（如Token Ring）。PostgreSQL不提供位反转的功能，所有接受的格式都使用规范的LSB顺序。

剩下的四种输入格式不属于任何标准。

8.10. 位串类型

位串是由 1 和 0 组成的字符串。它们可用于存储或展示位掩码。SQL 中有两种位类型：bit(*n*) 和 bit varying(*n*)，其中 *n* 是正整数。

bit 类型的数据长度必须与 *n* 完全一致；试图存储更短或更长的位串都会报错。bit varying 数据则是长度上限为 *n* 的变长类型，更长的串会被拒绝。不带长度的 bit 等效于 bit(1)，而不带长度说明的 bit varying 表示长度不受限。

> **注意**
>
> 如果显式地将一个位串值转换为 bit(*n*)，它会在右侧被截断或补零，直到长度恰好为 *n* 位，且不会报错。类似地，如果显式地将一个位串值转换为 bit varying(*n*)，而其长度超过 *n* 位，则会在右侧被截断。

关于位串常量的语法信息，请参见第 4.1.2.5 节。位逻辑操作符和字符串操作函数也可用；参见第 9.6 节。

例 8.3. 使用位串类型

```
CREATE TABLE test (a BIT(3), b BIT VARYING(5));
INSERT INTO test VALUES (B'101', B'00');
INSERT INTO test VALUES (B'10', B'101');
ERROR:  bit string length 2 does not match type bit(3)
INSERT INTO test VALUES (B'10'::bit(3), B'101');
SELECT * FROM test;
  a  |  b
-----+-----
 101 | 00
 100 | 101
```

一个位串值每 8 位需要 1 个字节，再加上 5 或 8 个字节的额外开销，具体取决于串的长度。（不过，长值可能会被压缩或移到行外存储，与第 8.3 节中对字符串的说明相同。）

8.11. 文本搜索类型

PostgreSQL 提供了两种专为支持全文检索而设计的数据类型。所谓全文检索，是指在一组自然语言文档中查找最匹配某个查询的文档。tsvector 类型以适合文本检索的优化形式表示文档，tsquery 类型则表示文本查询。关于这一功能的详细解释见第 12 章；相关函数和操作符的概览见第 9.13 节。

8.11.1. tsvector

tsvector 值是一个排好序且互不重复的词位（lexeme）列表，这些词已经过正规化，以便把同一单词的不同变体合并起来（详见第 12 章）。排序和去重会在输入时自动完成，如下例所示：

```
SELECT 'a fat cat sat on a mat and ate a fat rat'::tsvector;
                      tsvector
----------------------------------------------------
 'a' 'and' 'ate' 'cat' 'fat' 'mat' 'on' 'rat' 'sat'
```

若要表示包含空白或标点的词位，请用引号将它们括起来：

```
SELECT $$the lexeme '    ' contains spaces$$::tsvector;
                 tsvector
--------------------------------------------
 '    ' 'contains' 'lexeme' 'spaces' 'the'
```

（本例及下例使用美元引用的字符串常量，以避免在字符串内部必须双写引号所造成的混淆。）
嵌入的引号和反斜线必须双写：

```
SELECT $$the lexeme 'Joe''s' contains a quote$$::tsvector;
                    tsvector
--------------------------------------------------
 'Joe''s' 'a' 'contains' 'lexeme' 'quote' 'the'
```

还可以为词位附加整数形式的位置：

```
SELECT 'a:1 fat:2 cat:3 sat:4 on:5 a:6 mat:7 and:8 ate:9 a:10 fat:11
 rat:12'::tsvector;
                                  tsvector
---------------------------------------------------------------------------
---------
 'a':1,6,10 'and':8 'ate':9 'cat':3 'fat':2,11 'mat':7 'on':5 'rat':12
 'sat':4
```

一个位置通常表示源词在文档中的位置。位置信息可用于邻近度排序。位置值可以位于 1 到 16383 之间；更大的数字会被静默设为 16383。同一词位的重复位置会被丢弃。

带有位置的词位还可以进一步附加一个权重标签，其值可以是 A、B、C 或 D。D 是默认值，因此在输出中不会显示：

```
SELECT 'a:1A fat:2B,4C cat:5D'::tsvector;
          tsvector
----------------------------
 'a':1A 'cat':5 'fat':2B,4C
```

权重通常用于反映文档结构，例如把标题中的词和正文中的词区分开来。
文本检索排序函数可以为不同的权重标记分配不同优先级。

必须认识到，tsvector 类型本身并不会执行任何词语正规化；它假定输入的词已经按照应用需求完成正规化。例如：

```
select 'The Fat Rats'::tsvector;
      tsvector
--------------------
 'Fat' 'Rats' 'The'
```

对于大多数英文全文检索应用来说，上述词会被视为尚未正规化，但 tsvector 并不在意。原始文档文本通常应先经过 to_tsvector，以按检索需要对词语进行正规化：

```
SELECT to_tsvector('english', 'The Fat Rats');
   to_tsvector
-----------------
 'fat':2 'rat':3
```

更多细节仍请参见第 12 章。

8.11.2. tsquery

tsquery 值存储要搜索的词位，并按照布尔操作符 &（AND）、|（OR）和 !（NOT）把它们组合起来。可以使用圆括号强制指定这些操作符的分组方式：

```
SELECT 'fat & rat'::tsquery;
    tsquery
---------------
 'fat' & 'rat'

SELECT 'fat & (rat | cat)'::tsquery;
          tsquery
----------------------------
 'fat' & ( 'rat' | 'cat' )

SELECT 'fat & rat & ! cat'::tsquery;
        tsquery
------------------------
 'fat' & 'rat' & !'cat'
```

若没有圆括号，!（NOT）的绑定最紧，而 &（AND）比 |（OR）绑定得更紧。

可选地，tsquery 中的词位可以用一个或多个权重字母标注，这会限制它们只匹配 tsvector 中带有相应权重的词位：

```
SELECT 'fat:ab & cat'::tsquery;
    tsquery
------------------
 'fat':AB & 'cat'
```

此外，tsquery 中的词位还可以带上 * 标签来指定前缀匹配：

```
SELECT 'super:*'::tsquery;
  tsquery
-----------
 'super':*
```

这个查询将匹配 tsvector 中任何以 “super” 开头的词位。请注意，前缀会先由文本搜索配置处理，这意味着下面的比较会返回真：

```
SELECT to_tsvector( 'postgraduate' ) @@ to_tsquery( 'postgres:*' );
 ?column?
----------
 t
(1 row)
```

因为 postgres 会被词干化为 postgr：

```
SELECT to_tsquery('postgres:*');
 to_tsquery
------------
 'postgr':*
(1 row)
```

它随后就能匹配 postgraduate。

引号的使用规则与前面介绍 tsvector 时相同；同样，与 tsvector 一样，任何需要的词语正规化都必须在转换为 tsquery 类型之前完成。to_tsquery 函数很适合用来实现这种正规化：

```
SELECT to_tsquery('Fat:ab & Cats');
    to_tsquery
------------------
 'fat':AB & 'cat'
```

8.12. UUID类型

uuid 数据类型用于存储由 RFC 4122、ISO/IEC 9834-8:2005 及相关标准定义的通用唯一标识符（UUID）。（有些系统把这种数据类型称为全局唯一标识符，或 GUID。）这种标识符是一个 128 位的量，由某种算法生成，该算法的设计目标是：使已知宇宙中其他人用同一算法生成出相同标识符的概率极低。因此，对于分布式系统而言，
这类标识符能比序列生成器提供更好的唯一性保证，因为序列生成器仅在单个数据库内唯一。

UUID 写作一串小写十六进制数字，并用连字符分隔成若干组：先是 8 位一组，接着是三个 4 位组，最后是一个 12 位组。总共 32 个十六进制数字表示 128 个二进制位。标准形式的 UUID 例如：

```
a0eebc99-9c0b-4ef8-bb6d-6bb9bd380a11
```

PostgreSQL 也接受其他输入形式：可以使用大写字母、用花括号包围标准格式、省略部分或全部连字符，或者在任意 4 位分组后额外加上连字符。例如：

```
A0EEBC99-9C0B-4EF8-BB6D-6BB9BD380A11
{a0eebc99-9c0b-4ef8-bb6d-6bb9bd380a11}
a0eebc999c0b4ef8bb6d6bb9bd380a11
a0ee-bc99-9c0b-4ef8-bb6d-6bb9-bd38-0a11
{a0eebc99-9c0b4ef8-bb6d6bb9-bd380a11}
```

输出总是采用标准形式。

PostgreSQL为 UUID 提供存储和比较函数，但核心数据库不包含生成 UUID 的函数，因为没有一种算法适合所有应用。uuid-ossp模块提供实现多种标准算法的函数。或者，
可以由客户端应用或通过服务器端函数调用的其他库来生成 UUID。

8.13. XML类型

xml 数据类型可用于存储 XML 数据。与把 XML 数据存储在 text 字段中相比，它的优势在于会检查输入值的良构性，并且提供了可执行类型安全操作的支持函数；见第 9.14 节。使用该数据类型要求构建时使用 configure --with-libxml。

xml 类型既可以存储 XML 标准所定义的良构“文档”，也可以存储“内容”片段；后者由 XML 标准中的产生式 XMLDecl? content 定义。粗略地说，这意味着内容片段可以拥有多个顶层元素

或字符节点。表达式 *xmlvalue* IS DOCUMENT 可用于判断某个特定的 xml 值是完整文档，还是仅仅是一个内容片段。

8.13.1. 创建XML值

要从字符数据生成 xml 类型的值，可以使用函数 xmlparse：

```
XMLPARSE ( { DOCUMENT | CONTENT } value)
```

示例：

```
XMLPARSE (DOCUMENT '<?xml version="1.0"?><book><title>Manual</
title><chapter>...</chapter></book>')
XMLPARSE (CONTENT 'abc<foo>bar</foo><bar>foo</bar>')
```

按照 SQL 标准，这是把字符串转换为 XML 值的唯一方法；但 PostgreSQL 特有的下列语法：

```
xml '<foo>bar</foo>'
'<foo>bar</foo>'::xml
```

也可以使用。

即使输入值指定了文档类型声明（DTD），xml 类型也不会根据 DTD 来验证输入值。目前也没有内置支持可依据其他 XML 模式语言（如 XML Schema）来执行验证。

反向操作，也就是从 xml 生成字符串值，则使用函数 xmlserialize：

```
XMLSERIALIZE ( { DOCUMENT | CONTENT } value AS type )
```

type 可以是 character、character varying 或 text（或它们的别名）。同样，根据 SQL 标准，这是在 xml 与字符类型之间进行转换的唯一方式，但 PostgreSQL 也允许你直接进行类型转换。

当字符串值在不经过 XMLPARSE 或 XMLSERIALIZE 的情况下与 xml 类型互相转换时，选择 DOCUMENT 还是 CONTENT 由会话配置参数“XML 选项”决定，可以使用标准命令设置：

```
SET XML OPTION { DOCUMENT | CONTENT };
```

或使用更接近 PostgreSQL 风格的语法

```
SET xmloption TO { DOCUMENT | CONTENT };
```

默认值是 CONTENT，因此允许所有形式的 XML 数据。

> **注意**
>
> 在默认的 XML 选项设置下，如果字符串包含文档类型声明，则不能直接转换为 xml 类型，因为 XML 内容片段的定义不接受它们。如果需要这样做，可以使用 XMLPARSE，或者更改 XML 选项。

8.13.2. 编码处理

在客户端、服务器以及其间传输的 XML 数据上处理多种字符编码时，必须格外小心。使用文本模式向服务器发送查询并把查询结果返回给客户端时（这是通常使用的模式），PostgreSQL 会将

客户端与服务器之间传输的所有字符数据转换为目标端的字符编码，参见第 22.3 节。这也包括表示 XML 值的字符串，如上例所示。这通常意味着，由于字符数据在客户端和服务器之间传输时可能被转换为其他编码，XML 数据中包含的编码声明可能会失效，因为内嵌的编码声明本身并不会被修改。为处理这种情况，表示 `xml` 类型输入值的字符串中所包含的编码声明会被忽略，其内容被假定为当前服务器编码。因此，为了正确处理，客户端发出的 XML 数据字符串必须采用当前客户端编码。客户端负责在将文档发送给服务器之前把它们转换为当前客户端编码，或适当调整客户端编码。输出时，`xml` 类型值不会带有编码声明，而客户端应假定所有数据都采用当前客户端编码。

若使用二进制模式把查询参数传给服务器并把查询结果返回客户端，则不会执行字符集转换，因此情况有所不同。在这种情况下，XML 数据中的编码声明会被识别；如果缺少编码声明，则该数据会被假定为 UTF-8（这是 XML 标准的要求；另请注意 PostgreSQL 不支持 UTF-16）。输出时，除非客户端编码是 UTF-8（此时编码声明会被省略），否则数据会带有一个指明客户端编码的编码声明。

不言而喻，在 PostgreSQL 中处理 XML 数据时，如果 XML 数据编码、客户端编码和服务器编码三者相同，不仅更高效，也更不容易出错。由于 XML 数据在内部以 UTF-8 处理，如果服务器编码也是 UTF-8，则处理效率最高。

> **小心**
>
> 当服务器编码不是 UTF-8 时，某些与 XML 相关的函数在非 ASCII 数据上可能完全无法工作。尤其是 `xpath()`，这是一个已知问题。

8.13.3. 访问XML值

`xml` 数据类型有些特殊，因为它不提供任何比较操作符。这是因为对 XML 数据并不存在良定义且通用的比较算法。其结果是，你无法通过把某个 `xml` 列与搜索值比较来检索行。因此，XML 值通常应伴随一个独立的键字段，例如 ID。另一种比较 XML 值的办法，是先把它们转换成字符串；但请注意，字符串比较对 XML 的比较需求通常帮助不大。

由于 `xml` 数据类型没有可用的比较操作符，因此无法直接在这种类型的列上创建索引。如果需要在 XML 中快速搜索，可行方案包括：将表达式转换为字符串类型后为其建立索引，或者为某个 XPath 表达式建立索引。当然，实际查询也必须相应调整为使用该被索引的表达式。

PostgreSQL 的文本检索功能也可用于加速 XML 数据的全文检索。不过，目前 PostgreSQL 发行版中仍缺少所需的预处理支持。

8.14. 数组

PostgreSQL 允许将表列定义为变长多维数组。可以创建任何内置或用户定义的基础类型、枚举类型或复合类型的数组。尚不支持域的数组。

8.14.1. 数组类型的声明

为了说明数组类型的用法，我们创建下面这个表：

```
CREATE TABLE sal_emp (
    name            text,
    pay_by_quarter  integer[],
    schedule        text[][]
```

```
);
```

如上所示，数组数据类型通过在数组元素的数据类型名后附加方括号（[]）来命名。
上述命令将创建一个名为 sal_emp 的表，其中有一个 text 类型的列（name）、一个表示员工各季度薪资的一维 integer 数组（pay_by_quarter），以及一个表示员工每周日程安排的二维 text 数组（schedule）。

CREATE TABLE 的语法允许指定数组的确切大小，例如：

```
CREATE TABLE tictactoe (
    squares   integer[3][3]
);
```

不过，当前实现会忽略给出的任何数组大小限制，也就是说，其行为与未指定长度的数组相同。

当前实现也不会强制执行所声明的维度数。对于某一特定元素类型的数组，
无论大小或维度数如何，都会被视为同一种类型。因此，在 CREATE TABLE 中声明数组大小或维度数仅仅是文档说明；它不会影响运行时行为。

另一种使用关键字 ARRAY、并且符合 SQL 标准的语法可用于一维数组。pay_by_quarter 也可以定义为：

```
    pay_by_quarter  integer ARRAY[4],
```

或者，如果不指定数组大小：

```
    pay_by_quarter  integer ARRAY,
```

不过，与前面一样，PostgreSQL 在任何情况下都不会强制执行大小限制。

8.14.2. 数组值输入

要把数组值写成字面常量，请将元素值放在花括号内，并用逗号分隔。（如果你了解 C，这与 C 中初始化结构体的语法有些类似。）你可以给任意元素值加上双引号；如果它包含逗号或花括号，则必须这样做。（更多细节见下文。）因此，数组常量的一般格式如下：

```
'{ val1 delim val2 delim ... }'
```

其中 *delim* 是该类型的分隔符字符，它记录在其 pg_type 条目中。在 PostgreSQL 发行版提供的标准数据类型里，除类型 box 使用分号（;）外，其余都使用逗号（,）。每个 *val* 要么是数组元素类型的常量，要么是一个子数组。数组常量的一个示例是：

```
'{{1,2,3},{4,5,6},{7,8,9}}'
```

这个常量是一个 3 x 3 的二维数组，由三个整数子数组构成。

要把数组常量中的某个元素设为 NULL，请把该元素值写成 NULL。（NULL 的任意大小写变体都可以。）如果你想要实际的字符串值“NULL”，就必须为它加上双引号。

（这类数组常量实际上只是第 4.1.2.7 节中讨论的通用类型常量的一种特例。该常量最初会被当作字符串处理，然后传递给数组输入转换例程。可能需要显式指定类型。）

现在我们来看几个 INSERT 语句：

```
INSERT INTO sal_emp
    VALUES ('Bill',
```

```
    '{10000, 10000, 10000, 10000}',
    '{{"meeting", "lunch"}, {"training", "presentation"}}');

INSERT INTO sal_emp
    VALUES ('Carol',
    '{20000, 25000, 25000, 25000}',
    '{{"breakfast", "consulting"}, {"meeting", "lunch"}}');
```

前两条插入语句的结果如下：

```
SELECT * FROM sal_emp;
 name  |      pay_by_quarter       |                  schedule
-------+---------------------------+-----------------------------------
---------
 Bill  | {10000,10000,10000,10000} | {{meeting,lunch},{training,
presentation}}
 Carol | {20000,25000,25000,25000} | {{breakfast,consulting},{meeting,
lunch}}
(2 rows)
```

多维数组的各个子数组在每个维度上的长度必须匹配。不匹配会导致错误，例如：

```
INSERT INTO sal_emp
    VALUES ('Bill',
    '{10000, 10000, 10000, 10000}',
    '{{"meeting", "lunch"}, {"meeting"}}');
ERROR:  multidimensional arrays must have array expressions with
 matching dimensions
```

也可以使用 ARRAY 构造器语法：

```
INSERT INTO sal_emp
    VALUES ('Bill',
    ARRAY[10000, 10000, 10000, 10000],
    ARRAY[['meeting', 'lunch'], ['training', 'presentation']]);

INSERT INTO sal_emp
    VALUES ('Carol',
    ARRAY[20000, 25000, 25000, 25000],
    ARRAY[['breakfast', 'consulting'], ['meeting', 'lunch']]);
```

注意，数组元素是普通的 SQL 常量或表达式；例如，字符串字面值用单引号而不是双引号包围，而在数组字面量中则会用双引号。关于 ARRAY 构造器语法的更多细节见第 4.2.12 节。

8.14.3. 访问数组

现在，我们可以对该表执行一些查询。首先，演示如何访问数组中的单个元素。
下面的查询取回第二季度薪资发生变化的员工姓名：

```
SELECT name FROM sal_emp WHERE pay_by_quarter[1] <> pay_by_quarter[2];

 name
-------
 Carol
(1 row)
```

数组下标写在方括号内。默认情况下，PostgreSQL 对数组采用从 1 开始的编号约定，也就是说，一个有 *n* 个元素的数组从 `array[1]` 开始，到 `array[n]` 结束。

下面这个查询取回所有员工第三季度的薪资：

```
SELECT pay_by_quarter[3] FROM sal_emp;

 pay_by_quarter
----------------
          10000
          25000
(2 rows)
```

我们还可以访问数组的任意矩形切片，即子数组。数组切片通过在一个或多个数组维度上写成 *lower-bound*:*upper-bound* 的形式来表示。例如，下面这个查询取回 Bill 在一周中前两天里每天的首项日程：

```
SELECT schedule[1:2][1:1] FROM sal_emp WHERE name = 'Bill';

        schedule
------------------------
 {{meeting},{training}}
(1 row)
```

如果任何一个维度写成切片形式，也就是包含冒号，那么所有维度都会被当作切片处理。任何只有单个数字（没有冒号）的维度都会被视为从 1 到该数字指定的范围。例如，`[2]` 会被当作 `[1:2]`，如下例所示：

```
SELECT schedule[1:2][2] FROM sal_emp WHERE name = 'Bill';

                 schedule
-------------------------------------------
 {{meeting,lunch},{training,presentation}}
(1 row)
```

为了避免与非切片情况混淆，最好对所有维度都使用切片语法，例如写成 `[1:2][1:1]`，而不是 `[2][1:1]`。

如果数组本身或任一下标表达式为 NULL，则数组下标表达式将返回 NULL。此外，如果下标超出数组边界，也会返回 NULL（这种情况不会报错）。例如，如果 `schedule` 当前的维度是 `[1:3][1:2]`，那么引用 `schedule[3][3]` 会得到 NULL。类似地，使用错误数量的下标访问数组，得到的也是 NULL 而不是错误。

同样地，如果数组本身或任一下标表达式为 NULL，数组切片表达式也会返回 NULL。不过，在其他情况下，例如选择一个完全位于当前数组边界之外的数组切片时，切片表达式返回的是空（零维）数组而不是 NULL。（这与非切片行为不一致，是出于历史原因。）如果所请求的切片与数组边界仅部分重叠，那么它会被静默缩减为重叠区域，而不是返回 NULL。

任意数组值的当前维度都可以用 `array_dims` 函数取回：

```
SELECT array_dims(schedule) FROM sal_emp WHERE name = 'Carol';

 array_dims
------------
```

```
 [1:2][1:2]
(1 row)
```

array_dims 产生的是一个 text 结果，这对人工阅读比较方便，但对程序来说可能不太方便。也可以用 array_upper 和 array_lower 取回维度信息，它们分别返回指定数组维度的上界和下界：

```
SELECT array_upper(schedule, 1) FROM sal_emp WHERE name = 'Carol';

 array_upper
-------------
           2
(1 row)
```

array_length 会返回指定数组维度的长度：

```
SELECT array_length(schedule, 1) FROM sal_emp WHERE name = 'Carol';

 array_length
--------------
            2
(1 row)
```

8.14.4. 修改数组

数组值可以被整体替换：

```
UPDATE sal_emp SET pay_by_quarter = '{25000,25000,27000,27000}'
    WHERE name = 'Carol';
```

或者使用 ARRAY 表达式语法：

```
UPDATE sal_emp SET pay_by_quarter = ARRAY[25000,25000,27000,27000]
    WHERE name = 'Carol';
```

也可以更新数组中的单个元素：

```
UPDATE sal_emp SET pay_by_quarter[4] = 15000
    WHERE name = 'Bill';
```

或者更新其中一个切片：

```
UPDATE sal_emp SET pay_by_quarter[1:2] = '{27000,27000}'
    WHERE name = 'Carol';
```

已存储的数组值可以通过给尚不存在的元素赋值来扩展。
原有元素与新赋值元素之间的任何位置都将用 NULL 填充。例如，如果数组 myarray 当前有 4 个元素，那么在一次更新把值赋给 myarray[6] 之后，它将有 6 个元素；myarray[5] 将包含 NULL。目前，以这种方式扩展只允许用于一维数组，不允许用于多维数组。

带下标的赋值也允许创建不使用从 1 开始下标的数组。例如，可以给 myarray[-2:7] 赋值，从而创建一个下标值范围为 -2 到 7 的数组。

新的数组值也可以使用连接操作符 || 构造：

```
SELECT ARRAY[1,2] || ARRAY[3,4];
```

```
 ?column?
-----------
 {1,2,3,4}
(1 row)

SELECT ARRAY[5,6] || ARRAY[[1,2],[3,4]];
      ?column?
---------------------
 {{5,6},{1,2},{3,4}}
(1 row)
```

连接操作符允许把单个元素添加到一维数组的开头或末尾。它也接受两个 *N* 维数组，或者一个 *N* 维数组与一个 *N+1* 维数组。

当把单个元素添加到一维数组的开头或末尾时，结果数组会保留该数组操作数的下界下标。例如：

```
SELECT array_dims(1 || '[0:1]={2,3}'::int[]);
 array_dims
------------
 [0:2]
(1 row)

SELECT array_dims(ARRAY[1,2] || 3);
 array_dims
------------
 [1:3]
(1 row)
```

当连接两个维度数相同的数组时，结果会保留左侧操作数外层维度的下界下标。结果数组由左侧操作数的所有元素后跟右侧操作数的所有元素组成。例如：

```
SELECT array_dims(ARRAY[1,2] || ARRAY[3,4,5]);
 array_dims
------------
 [1:5]
(1 row)

SELECT array_dims(ARRAY[[1,2],[3,4]] || ARRAY[[5,6],[7,8],[9,0]]);
 array_dims
------------
 [1:5][1:2]
(1 row)
```

当把一个 *N* 维数组添加到一个 *N+1* 维数组的开头或末尾时，其结果与前面“单个元素与数组相连”的情况类似。每个 *N* 维子数组本质上都是该 *N+1* 维数组外层维度中的一个元素。例如：

```
SELECT array_dims(ARRAY[1,2] || ARRAY[[3,4],[5,6]]);
 array_dims
------------
 [1:3][1:2]
(1 row)
```

也可以使用 `array_prepend`、`array_append` 或 `array_cat` 函数来构造数组。前两个只支持一维数组，但 `array_cat` 支持多维数组。注意，与直接使用这些函数相比，更推荐使用上面讨

论的连接操作符。实际上，这些函数的存在主要是为了在实现连接操作符时使用。不过，在创建用户定义聚合时，它们可能也有直接的用处。示例如下：

```
SELECT array_prepend(1, ARRAY[2,3]);
 array_prepend
---------------
 {1,2,3}
(1 row)

SELECT array_append(ARRAY[1,2], 3);
 array_append
--------------
 {1,2,3}
(1 row)

SELECT array_cat(ARRAY[1,2], ARRAY[3,4]);
 array_cat
-----------
 {1,2,3,4}
(1 row)

SELECT array_cat(ARRAY[[1,2],[3,4]], ARRAY[5,6]);
      array_cat
---------------------
 {{1,2},{3,4},{5,6}}
(1 row)

SELECT array_cat(ARRAY[5,6], ARRAY[[1,2],[3,4]]);
      array_cat
---------------------
 {{5,6},{1,2},{3,4}}
```

8.14.5. 在数组中搜索

要在数组中搜索某个值，就必须检查每一个值。如果你知道数组的大小，这可以手工完成。例如：

```
SELECT * FROM sal_emp WHERE pay_by_quarter[1] = 10000 OR
                            pay_by_quarter[2] = 10000 OR
                            pay_by_quarter[3] = 10000 OR
                            pay_by_quarter[4] = 10000;
```

不过，对于大型数组，这样很快就会变得繁琐；而在数组大小未知时，这也没有帮助。另一种方法见第 9.21 节。上面的查询可以改写为：

```
SELECT * FROM sal_emp WHERE 10000 = ANY (pay_by_quarter);
```

此外，如果要查找数组中所有值都等于 10000 的行，可以使用：

```
SELECT * FROM sal_emp WHERE 10000 = ALL (pay_by_quarter);
```

另外，也可以使用 `generate_subscripts` 函数。例如：

```
SELECT * FROM
   (SELECT pay_by_quarter,
```

```
          generate_subscripts(pay_by_quarter, 1) AS s
     FROM sal_emp) AS foo
 WHERE pay_by_quarter[s] = 10000;
```

关于该函数的说明见表 9.47。

> **提示**
>
> 数组不是集合；搜索特定数组元素可能是数据库设计不当的信号。
> 可以考虑使用一个独立的表，让原本会成为数组元素的每个项各占一行。
> 这样会更容易搜索，而且在元素数量很多时通常也有更好的可伸缩性。

8.14.6. 数组输入和输出语法

数组值的外部文本表示由若干项构成，这些项会按照数组元素类型的 I/O 转换规则进行解释，再加上一些表示数组结构的修饰。修饰部分包括数组值外层的花括号（`{` 和 `}`），以及相邻项之间的分隔符字符。分隔符字符通常是逗号（`,`），但也可能是别的字符：它由数组元素类型的 `typdelim` 设置决定。在 PostgreSQL 发行版提供的标准数据类型中，除类型 `box` 使用分号（`;`）外，其余都使用逗号。在多维数组中，每个维度（行、平面、立方体等）都有自己的一层花括号，并且同一级别相邻的、由花括号括起的各项之间必须写出分隔符。

如果元素值是空字符串、包含花括号、分隔符字符、双引号、反斜线或空白，或者与单词 `NULL` 匹配，数组输出例程就会在元素值外面加上双引号。元素值中嵌入的双引号和反斜线会用反斜线转义。对于数字数据类型，可以安全地假定不会出现双引号；但对于文本数据类型，
则应准备好处理带引号和不带引号两种情况。

默认情况下，数组各维度的下界下标值都设为 1。要表示具有其他下界的数组，可以在写出数组内容之前显式指定数组下标范围。这种修饰由包围每个数组维度上下界的方括号（`[]`）构成，
中间以冒号（`:`）作为分隔符字符。数组维度修饰后面再跟一个等号（`=`）。例如：

```
SELECT f1[1][-2][3] AS e1, f1[1][-1][5] AS e2
 FROM (SELECT '[1:1][-2:-1][3:5]={{{1,2,3},{4,5,6}}}'::int[] AS f1) AS
 ss;

 e1 | e2
----+----
  1 |  6
(1 row)
```

只有当一个或多个下界不等于 1 时，数组输出例程才会在结果中包含显式维度信息。

如果某个元素写入的值是 `NULL`（任何大小写变体都可以），该元素就会被视为 NULL。若存在任何引号或反斜线，则会禁用这种行为，从而可以输入字面字符串值“NULL”。另外，为了与 PostgreSQL 8.2 之前的版本保持向后兼容，可以将array_nulls配置参数设为 `off`，从而禁止把 `NULL` 识别为 NULL。

如前所述，在写数组值时，可以给任意单个数组元素加上双引号。
如果元素值不加引号就会让数组值解析器产生混淆，那么你必须这样做。例如，包含花括号、
逗号（或者该数据类型的分隔符字符）、双引号、反斜线，或者前后带有空白的元素值，
必须使用双引号。空字符串以及与单词 `NULL` 匹配的字符串也必须加引号。要在一个带双引号的数组元素值中放入双引号或反斜线，需要使用转义字符串语法并在它前面加反斜线。或者，
你也可以避免使用引号，而改用反斜线转义来保护所有原本会被当作数组语法的数据字符。

你可以在左花括号之前或右花括号之后添加空白。
也可以在任意单个项字符串之前或之后添加空白。在所有这些情况下，这些空白都会被忽略。
不过，双引号元素内部的空白，或者元素中夹在两个非空白字符之间的空白，不会被忽略。

注意

请记住，你在 SQL 命令中写下的内容会先被解释为字符串字面值，然后才被解释为数组。这会使你所需的反斜线数量加倍。例如，要插入一个包含一个反斜线和一个双引号的 `text` 数组值，你需要写成：

```
INSERT ... VALUES (E'{"\\\\","\\""}');
```

转义字符串处理器会去掉一层反斜线，因此到达数组值解析器的内容形如 `{"\\","\""}`。而传给 `text` 数据类型输入例程的字符串则分别变成 `\` 和 `"`。（如果我们使用的是输入例程同样对反斜线做特殊处理的数据类型，例如 `bytea`，那么要在命令中写入八个反斜线才能让存储的数组元素中出现一个反斜线。）可以使用美元引用（见第 4.1.2.4 节）来避免双写反斜线。

提示

在 SQL 命令中写数组值时，`ARRAY` 构造器语法（见第 4.2.12 节）通常比数组字面量语法更容易使用。在 `ARRAY` 中，各个元素值的写法与它们不作为数组成员时完全相同。

8.15. 复合类型

复合类型表示一行或一条记录的结构；本质上它就是字段名及其数据类型的列表。
PostgreSQL允许像使用简单类型那样，在许多相同的场合使用复合类型。例如，
表中的一列可以声明为复合类型。

8.15.1. 复合类型的声明

下面是两个定义复合类型的简单示例：

```
CREATE TYPE complex AS (
    r       double precision,
    i       double precision
);

CREATE TYPE inventory_item AS (
    name            text,
    supplier_id     integer,
    price           numeric
);
```

这种语法与`CREATE TABLE`类似，但只能指定字段名和类型；目前还不能包含约束（例如`NOT NULL`）。注意，`AS`关键字必不可少；如果没有它，系统会认为你想使用另一种`CREATE TYPE`命令，并产生令人费解的语法错误。

定义了这些类型之后，我们可以用它们来创建表：

```
CREATE TABLE on_hand (
    item      inventory_item,
    count     integer
);

INSERT INTO on_hand VALUES (ROW('fuzzy dice', 42, 1.99), 1000);
```

或者创建函数：

```
CREATE FUNCTION price_extension(inventory_item, integer) RETURNS
 numeric
AS 'SELECT $1.price * $2' LANGUAGE SQL;

SELECT price_extension(item, 10) FROM on_hand;
```

每当创建一张表时，也会自动创建一个与该表同名的复合类型，用来表示表的行类型。例如，假如我们执行了：

```
CREATE TABLE inventory_item (
    name            text,
    supplier_id     integer REFERENCES suppliers,
    price           numeric CHECK (price > 0)
);
```

那么，与上文相同的`inventory_item`复合类型就会随之产生，而且可以像上文那样使用。不过，请注意当前实现的一个重要限制：由于复合类型本身不关联任何约束，表定义中的约束并不适用于表之外的复合类型值。（一种部分解决办法是使用域类型作为复合类型的成员。）

8.15.2. Composite Value Input

要把组合值写成字面常量，请将各字段值放在圆括号内，并用逗号分隔。你可以给任意字段值加双引号；如果它包含逗号或圆括号，则必须这样做。（更多细节见下文。）因此，组合常量的一般格式如下：

```
'( val1 , val2 , ... )'
```

An example is:

```
'("fuzzy dice",42,1.99)'
```

它是上面定义的`inventory_item`类型的一个合法值。要让某个字段为 NULL，在该字段在列表中的位置上什么也不写即可。例如，下面的常量指定第三个字段为 NULL：

```
'("fuzzy dice",42,)'
```

如果你想要的是空串而不是 NULL，请写双引号：

```
'("",42,)'
```

这里第一个字段是非 NULL 的空串，第三个字段是 NULL。

（这些常量实际上只是第 4.1.2.7 节中讨论的通用类型常量的一个特例。该常量最初被当作字符串处理，然后传递给组合类型的输入转换例程。有时可能需要显式指定类型。）

`ROW`表达式语法也可以用来构造组合值。在多数情况下，它比字符串字面量语法简单得多，因为你不必担心多层引号的问题。我们在上面已经用过这种方法：

```
ROW('fuzzy dice', 42, 1.99)
ROW('', 42, NULL)
```

只要表达式中的字段多于一个，ROW 关键字实际上可以省略，因此上面的例子可以简化为：

```
('fuzzy dice', 42, 1.99)
('', 42, NULL)
```

ROW表达式语法的更多细节见第 4.2.13 节。

8.15.3. 访问复合类型

要访问组合列中的某个字段，可以写一个点号再加字段名，这很像通过表名选取字段。实际上，它与通过表名选取字段太像了，以至于你通常必须使用圆括号，以免让解析器混淆。例如，你可能尝试从示例表on_hand中选取一些子字段：

```
SELECT item.name FROM on_hand WHERE item.price > 9.99;
```

这不会起作用，因为根据 SQL 语法规则，名称item会被当成表名，而不是on_hand的列名。你必须写成这样：

```
SELECT (item).name FROM on_hand WHERE (item).price > 9.99;
```

或者，如果你还需要使用表名（例如在多表查询中），可以这样写：

```
SELECT (on_hand.item).name FROM on_hand WHERE (on_hand.item).price >
 9.99;
```

现在，加上括号的对象就会被正确解释为对item列的引用，然后就可以从中选出子字段。

无论何时从组合值中选择字段，都会遇到类似的语法问题。例如，要从一个返回组合值的函数结果中只选取一个字段，你需要这样写：

```
SELECT (my_func(...)).field FROM ...
```

如果没有额外的圆括号，这将生成一个语法错误。

8.15.4. 修改组合值

下面是一些插入和更新组合列时正确语法的示例。先看插入或更新整个列值的情况：

```
INSERT INTO mytab (complex_col) VALUES((1.1,2.2));

UPDATE mytab SET complex_col = ROW(1.1,2.2) WHERE ...;
```

第一个示例省略了ROW，第二个示例使用了它；两种写法都可以。

我们也可以更新组合列中的单个子字段：

```
UPDATE mytab SET complex_col.r = (complex_col).r + 1 WHERE ...;
```

注意，这里不需要（事实上也不能）给紧跟在SET后面的列名加圆括号，但在等号右侧的表达式中引用同一列时，则需要加圆括号。

我们也可以把子字段指定为INSERT的目标：

```
INSERT INTO mytab (complex_col.r, complex_col.i) VALUES(1.1, 2.2);
```

如果我们没有为该列的所有子字段提供值，其余子字段就会填充为 NULL 值。

8.15.5. 复合类型的输入和输出语法

组合值的外部文本表示由两部分组成：一部分是按照各字段类型的 I/O 转换规则解释的项，另一部分是表明组合结构的附加符号。这些附加符号包括包围整个值的圆括号（(和)），以及相邻项之间的逗号（,）。圆括号外部的空白会被忽略；但在圆括号内部，空白会被视为字段值的一部分，其是否有意义取决于该字段数据类型的输入转换规则。例如，在

```
'(  42)'
```

中，如果字段类型是 integer，则空白会被忽略；如果是 text，则不会被忽略。

如前所示，在写组合值时，你可以给任意单个字段值加双引号。如果字段值本身可能让组合值解析器混淆，则必须这样做。特别是，包含圆括号、逗号、双引号或反斜杠的字段必须用双引号括起来。要在带引号的组合字段值中写入双引号或反斜杠，需要在其前面加一个反斜杠。（另外，带双引号的字段值内部成对出现的双引号会被视为一个双引号字符，这与 SQL 字面字符串中单引号的规则类似。）或者，你也可以完全不使用引号，而改用反斜杠转义，保护所有原本会被当作组合语法的数据字符。

完全空的字段值（即逗号或圆括号之间一个字符也没有）表示 NULL。要写一个空字符串值而不是 NULL，可以写成""。

如果字段值是空字符串，或者包含圆括号、逗号、双引号、反斜杠或空白字符，复合类型输出例程就会在其周围加上双引号。（对空白字符这样做并非必需，但有助于提高可读性。）嵌入字段值中的双引号和反斜杠会被加倍。

> **注意**
>
> 请记住，你在 SQL 命令中写的内容会先被解释为字符串字面量，然后再被解释为组合值。这使你需要的反斜线数量翻倍（假定使用转义字符串语法）。例如，要在组合值中插入一个包含双引号和反斜线的`text`字段，你需要写：
>
> ```
> INSERT ... VALUES (E'("\\\"\\\\")');
> ```
>
> 字符串字面量处理器会去掉一层反斜线，因此到达组合值解析器的内容看起来像`("\"\\")`。相应地，传给`text`数据类型输入例程的字符串变成`"\`。（如果我们使用的是输入例程也特殊处理反斜线的类型，例如`bytea`，则要在命令中写多达八个反斜线才能在存储的组合字段中得到一个反斜线。）可以使用美元引用（见第 4.1.2.4 节）来避免双写反斜线。

> **提示**
>
> 在 SQL 命令中编写组合值时，`ROW`构造器语法通常比组合字面量语法更容易使用。在`ROW`中，各个字段值的写法与它们不是组合成员时完全相同。

8.16. 对象标识符类型

对象标识符（OID）在PostgreSQL内部被用作若干系统表的主键。除非建表时指定WITH OIDS，或启用配置变量default_with_oids，否则不会向用户创建的表添加 OID。类型oid表示一个对象标识符。还有若干oid的别名类型：regproc、regprocedure、regoper、regoperator、regclass、regtype、regconfig和regdictionary。表 8.23给出了概要说明。

oid 类型目前实现为无符号 4 字节整数。因此，在大型数据库中，它不足以提供数据库范围内的唯一性，甚至在大型的单个表中也无法保证唯一性。因此，不推荐把用户创建表的 OID 列用作主键。OID 最好仅用于引用系统表。

oid 类型本身除比较之外几乎没有其他操作。不过，它可以转换为整数，然后再用标准整数操作符进行处理。（这样做时要注意有符号与无符号可能造成的混淆。）

OID 别名类型除了专门的输入和输出例程外，没有自己的操作。这些例程能够接受并显示系统对象的符号名称，而不是 oid 类型所使用的原始数值。别名类型简化了对象 OID 值的查找。例如，要查看与表 mytable 相关的 pg_attribute 行，可以写成：

```
SELECT * FROM pg_attribute WHERE attrelid = 'mytable'::regclass;
```

而不是：

```
SELECT * FROM pg_attribute
  WHERE attrelid = (SELECT oid FROM pg_class WHERE relname =
 'mytable');
```

虽然这样看起来也不算太糟，但其实仍然过于简化。如果不同模式中有多个名为 mytable 的表，就必须写一个复杂得多的子查询来选出正确的 OID。regclass 的输入转换器会按照模式路径设置来处理表查找，因此能够自动完成“正确的事情”。同样，把表的 OID 转换成 regclass，也很适合用来以符号形式显示数值 OID。

表 8.23. 对象标识符类型

名字	引用	描述	值示例
oid	任意	数字形式的对象标识符	564182
regproc	pg_proc	函数名称	sum
regprocedure	pg_proc	函数与参数类型	sum(int4)
regoper	pg_operator	操作符名称	+
regoperator	pg_operator	带参数类型的操作符	*(integer,integer) 或-(NONE,integer)
regclass	pg_class	关系名称	pg_type
regtype	pg_type	数据类型名称	integer
regconfig	pg_ts_config	文本搜索配置	english
regdictionary	pg_ts_dict	文本搜索字典	simple

所有 OID 别名类型都接受模式限定名称；如果不加限定就无法在当前搜索路径中找到对象，输出时也会显示模式限定名称。regproc和regoper别名类型只接受唯一的（未重载的）输入名称，因此用途有限；对于大多数用途，regprocedure或regoperator更合适。
对于regoperator，通过将未使用的操作数写为NONE来标识一元操作符。

大多数 OID 别名类型还有一个附加特性，就是会创建依赖关系。如果这些类型的常量出现在存储的表达式中（例如列默认表达式或视图），它就会对被引用对象建立依赖。例如，
如果某列的默认表达式是 nextval('my_seq'::regclass)，PostgreSQL 就会知道该默认表达式依赖于序列 my_seq，因此系统不允许在尚未移除该默认表达式时删除序列。

系统使用的另一种标识符类型是xid，即事务（缩写为xact）标识符。这是系统列xmin和xmax的数据类型。事务标识符是 32 位的量。

系统使用的第三种标识符类型是 cid，也就是命令标识符。这是系统列 cmin 和 cmax 的数据类型。命令标识符同样是 32 位量。

系统使用的最后一种标识符类型是 tid，也就是元组标识符（行标识符）。这是系统列 ctid 的数据类型。一个元组 ID 是一对值（块号、块内元组索引），用来标识该行在其所属表中的物理位置。

（系统列的更多说明见第 5.4 节。）

8.17. 伪类型

PostgreSQL 的类型系统中包含一些特殊用途的条目，统称为伪类型。伪类型不能作为列数据类型使用，但可以用于声明函数的参数类型或结果类型。
每一种伪类型都适用于这样的场景：函数的行为并不对应于简单地接受或返回某个特定 SQL 数据类型的值。表 8.24列出了现有的伪类型。

表 8.24. 伪类型

名字	描述
any	表示一个函数可以接受任意输入数据类型。
anyarray	表示一个函数可以接受任意数组数据类型（参见第 35.2.5 节）
anyelement	表示一个函数可以接受任意数据类型（参见第 35.2.5 节）。
anyenum	表示一个函数可以接受任意枚举数据类型（参见第 35.2.5 节和第 8.7 节）。
anynonarray	表示一个函数可以接受任意非数组数据类型（参见第 35.2.5 节）。
cstring	表示一个函数接受或返回一个以空字符结尾的 C 字符串。
internal	表示一个函数接受或返回一个服务器内部数据类型。
language_handler	表示过程语言调用处理器被声明为返回 language_handler。
fdw_handler	表示外部数据包装器处理器被声明为返回 fdw_handler。
record	标识一个返回未指定行类型的函数。
trigger	触发器函数被声明为返回trigger.
void	表示一个函数不返回值。
opaque	一个已废弃的类型名称，以前用于上述所有用途。

用 C 编写的函数（无论是内置的还是动态装载的）都可以声明为接受或返回这些伪类型中的任意一种。函数作者必须自行确保，当伪类型被用作参数类型时，该函数的行为仍然安全。

用过程语言编写的函数，只有在其实现语言允许的情况下才能使用伪类型。目前，
各过程语言都禁止把伪类型用作参数类型，而只允许把void和record
用作结果类型（如果函数被用作触发器，则也允许trigger）。
有些语言还支持使用anyarray、anyelement、anyenum和anynonarray实现多态函数。

`internal` 伪类型用于声明那些只应由数据库系统内部调用、而不应被 SQL 查询直接调用的函数。如果一个函数至少有一个 `internal` 类型参数，那么它就不能从 SQL 中调用。为了保持这一限制的类型安全性，必须遵守这样一条编码规则：除非函数至少有一个 `internal` 类型参数，否则不要声明任何返回 `internal` 的函数。

第 9 章 函数和操作符

PostgreSQL为内置数据类型提供了大量函数和操作符。用户也可以按第 V 部分中的说明定义自己的函数和操作符。psql命令`\df`和`\do`可分别用于列出所有可用的函数和操作符。

如果你关心可移植性，请注意，本章描述的大多数函数和操作符，
除了最简单的算术和比较操作符以及少数明确标记的函数之外，都不是SQL标准规定的。不过，
其中一些扩展功能也存在于其他SQL数据库管理系统中，而且在很多情况下，
各实现之间的这类功能是兼容且一致的。本章也并非详尽无遗；
手册的相关章节中还有其他函数。

9.1. 逻辑操作符

常用的逻辑操作符有：

```
AND
OR
NOT
```

SQL使用三值的逻辑系统，包括真、假和`null`，它表示“未知”。观察下面的真值表：

a	*b*	*a* AND *b*	*a* OR *b*
TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	FALSE	TRUE
TRUE	NULL	NULL	TRUE
FALSE	FALSE	FALSE	FALSE
FALSE	NULL	FALSE	NULL
NULL	NULL	NULL	NULL

a	NOT *a*
TRUE	FALSE
FALSE	TRUE
NULL	NULL

操作符`AND`和`OR`是可交换的，也就是说，你可以交换左右操作数而不影响结果。但请参见第 4.2.14 节获取有关子表达式计算顺序的更多信息。

9.2. 比较操作符

常见的比较操作符都可用，如表 9.1所示。

表 9.1. 比较操作符

操作符	描述
<	小于
>	大于
<=	小于或等于
>=	大于或等于

操作符	描述
=	等于
<> 或 !=	不等于

注意

!= 操作符在解析器阶段会转换为 <>。不可能实现行为不同的 != 和 <> 操作符。

比较操作符适用于所有相关的数据类型。所有比较操作符都是返回 `boolean` 类型值的二元操作符；类似 `1 < 2 < 3` 的表达式是无效的（因为没有用于比较布尔值和 `3` 的 `<` 操作符）。

除比较操作符之外，还可以使用特殊的`BETWEEN`结构：

```
a BETWEEN x AND y
```

等价于

```
a >= x AND a <= y
```

注意，`BETWEEN`将两个端点值都视为包含在范围内。`NOT BETWEEN`执行相反的比较：

```
a NOT BETWEEN x AND y
```

等价于

```
a < x OR a > y
```

`BETWEEN SYMMETRIC`与`BETWEEN`相同，但不要求`AND`左侧的参数小于或等于右侧的参数。如果不是这样，这两个参数会自动交换，以确保表示的范围始终非空。

当任一输入为 null 时，普通的比较操作符会得到 null（表示“未知”），而不是真或假。例如，`7 = NULL`得到 null，`7 <> NULL`也一样。如果这种行为不合适，可以使用`IS [ NOT ] DISTINCT FROM`结构：

```
a IS DISTINCT FROM b
a IS NOT DISTINCT FROM b
```

对于非 null 输入，`IS DISTINCT FROM`和`<>`操作符一样。不过，如果两个输入都为 null，它会返回假。而如果只有一个输入为 null，它会返回真。类似地，`IS NOT DISTINCT FROM`对于非 null 输入的行为与`=`相同，但是当两个输入都为 null 时它返回真，并且当只有一个输入为 null 时返回假。因此，这些结构实际上把 null 当作一种普通数据值，而不是“未知”。

要检查一个值是否为 null，使用下面的结构：

```
expression IS NULL
expression IS NOT NULL
```

或者等效但非标准的结构：

```
expression ISNULL
expression NOTNULL
```

不要写*expression* `= NULL`，因为`NULL`并不“等于”`NULL`。（null 值表示未知值，而我们并不知道两个未知值是否相等。）

> **提示**
>
> 有些应用可能期望表达式*expression* = NULL在*expression*求值为 null 时返回真。我们强烈建议此类应用修改为遵循 SQL 标准。但是，如果无法这样修改，那么可以使用配置变量transform_null_equals。如果将其打开，PostgreSQL会把x = NULL子句转换成x IS NULL。

如果*expression*是行值，那么当行表达式本身为 null 或其所有字段都为 null 时，IS NULL 为真；而当行表达式本身非 null 且其所有字段都非 null 时，IS NOT NULL 为真。由于这种行为，IS NULL和IS NOT NULL并不总是对行值表达式返回相反的结果；特别是，一个同时包含 null 和非 null 字段的行值表达式会对这两种测试都返回假。在某些情况下，写成*row* IS DISTINCT FROM NULL或者*row* IS NOT DISTINCT FROM NULL可能更合适，因为它们只会检查整个行值是否为 null，而不会再对行字段做额外测试。

布尔值也可以使用下列结构进行测试：

```
expression IS TRUE
expression IS NOT TRUE
expression IS FALSE
expression IS NOT FALSE
expression IS UNKNOWN
expression IS NOT UNKNOWN
```

这些谓词总是返回真或假，从不返回空值，即使操作数为 null 也是如此。null 输入被当作逻辑值“未知”。请注意，IS UNKNOWN和IS NOT UNKNOWN实际上分别等同于IS NULL和IS NOT NULL，只是输入表达式必须是布尔类型。

9.3. 数学函数和操作符

PostgreSQL为很多类型提供了数学操作符。对于那些没有标准数学惯例的类型（如日期/时间类型），我们将在后续小节中描述实际的行为。

表 9.2列出了可用的数学操作符。

表 9.2. 数学操作符

操作符	描述	示例	结果
+	加法	2 + 3	5
-	减法	2 - 3	-1
*	乘法	2 * 3	6
/	除法（整数除法会截断结果）	4 / 2	2
%	取模（余数）	5 % 4	1
^	求幂（从左向右结合）	2.0 ^ 3.0	8
|/	平方根	|/ 25.0	5
||/	立方根	||/ 27.0	3
!	阶乘	5 !	120
!!	阶乘（前缀操作符）	!! 5	120
@	绝对值	@ -5.0	5

操作符	描述	示例	结果
&	按位与	91 & 15	11
|	按位或	32 | 3	35
#	按位异或	17 # 5	20
~	按位非	~1	-2
<<	按位左移	1 << 4	16
>>	按位右移	8 >> 2	2

按位操作符仅适用于整数数据类型，也可用于位串类型 bit 和 bit varying，如表 9.11所示。

表 9.3列出了可用的数学函数。表中的 dp 表示 double precision。许多函数提供了参数类型不同的多种形式。除非另有说明，函数的每种形式都返回与其参数相同的数据类型。处理 double precision 数据的函数大多基于主机系统的 C 库实现；因此，其精度和边界情况下的行为可能因主机系统而异。

表 9.3. 数学函数

函数	abs(*x*)
返回类型	（与输入相同）
描述	绝对值
示例	abs(-17.4)
结果	17.4
函数	cbrt(dp)
返回类型	dp
描述	立方根
示例	cbrt(27.0)
结果	3
函数	ceil(dp or numeric)
返回类型	（与输入相同）
描述	大于或等于参数的最小整数
示例	ceil(-42.8)
结果	-42
函数	ceiling(dp or numeric)
返回类型	（与输入相同）
描述	大于或等于参数的最小整数（与 ceil 相同）
示例	ceiling(-95.3)
结果	-95
函数	degrees(dp)
返回类型	dp
描述	将弧度转换为角度

示例	degrees(0.5)
结果	28.6478897565412
函数	div(*y* numeric, *x* numeric)
返回类型	numeric
描述	*y*/*x* 的整数商
示例	div(9,4)
结果	2
函数	exp(dp or numeric)
返回类型	（与输入相同）
描述	指数函数
示例	exp(1.0)
结果	2.71828182845905
函数	floor(dp or numeric)
返回类型	（与输入相同）
描述	小于或等于参数的最大整数
示例	floor(-42.8)
结果	-43
函数	ln(dp or numeric)
返回类型	（与输入相同）
描述	自然对数
示例	ln(2.0)
结果	0.693147180559945
函数	log(dp or numeric)
返回类型	（与输入相同）
描述	以 10 为底的对数
示例	log(100.0)
结果	2
函数	log(*b* numeric, *x* numeric)
返回类型	numeric
描述	以 *b* 为底的对数
示例	log(2.0, 64.0)
结果	6.0000000000
函数	mod(*y*, *x*)
返回类型	（与参数类型相同）
描述	*y*/*x* 的余数

示例	mod(9,4)
结果	1
函数	pi()
返回类型	dp
描述	常数 “π”
示例	pi()
结果	3.14159265358979
函数	power(*a* dp, *b* dp)
返回类型	dp
描述	*a*的*b*次幂
示例	power(9.0, 3.0)
结果	729
函数	power(*a* numeric, *b* numeric)
返回类型	numeric
描述	*a*的*b*次幂
示例	power(9.0, 3.0)
结果	729
函数	radians(dp)
返回类型	dp
描述	将角度转换为弧度
示例	radians(45.0)
结果	0.785398163397448
函数	round(dp or numeric)
返回类型	（与输入相同）
描述	舍入到最接近的整数
示例	round(42.4)
结果	42
函数	round(*v* numeric, *s* int)
返回类型	numeric
描述	舍入到 *s* 位小数
示例	round(42.4382, 2)
结果	42.44
函数	sign(dp or numeric)
返回类型	（与输入相同）

描述	参数的符号（-1、0、+1）
示例	`sign(-8.4)`
结果	`-1`
函数	`sqrt(dp or numeric)`
返回类型	（与输入相同）
描述	平方根
示例	`sqrt(2.0)`
结果	`1.4142135623731`
函数	`trunc(dp or numeric)`
返回类型	（与输入相同）
描述	向零截断
示例	`trunc(42.8)`
结果	`42`
函数	`trunc(v numeric, s int)`
返回类型	`numeric`
描述	截断到 *s* 位小数
示例	`trunc(42.4382, 2)`
结果	`42.43`
函数	`width_bucket(op numeric, b1 numeric, b2 numeric, count int)`
返回类型	`int`
描述	返回 *operand* 在等宽直方图（在 *b1* 到 *b2* 的范围内有 *count* 个桶）中所属的桶编号
示例	`width_bucket(5.35, 0.024, 10.06, 5)`
结果	`3`
函数	`width_bucket(op dp, b1 dp, b2 dp, count int)`
返回类型	`int`
描述	返回 *operand* 在等宽直方图（在 *b1* 到 *b2* 的范围内有 *count* 个桶）中所属的桶编号
示例	`width_bucket(5.35, 0.024, 10.06, 5)`
结果	`3`

表 9.4展示了用于产生随机数的函数。

表 9.4. 随机函数

函数	返回类型	描述
`random()`	`dp`	范围 0.0 <= x < 1.0 内的随机值
`setseed(dp)`	`void`	为后续 `random()` 调用设置种子（值在 -1.0 与 1.0 之间，包含端点）

random() 返回值的特性取决于系统实现。它不适用于加密应用；替代方案见pgcrypto模块。

最后，表 9.5列出了可用的三角函数。所有三角函数的参数和返回值都是 double precision 类型。三角函数的参数以弧度表示。反三角函数的返回值以弧度表示。可使用上面的单位换算函数radians()和degrees()。

表 9.5. 三角函数

函数	描述
acos(*x*)	反余弦
asin(*x*)	反正弦
atan(*x*)	反正切
atan2(*y*, *x*)	*y*/*x* 的反正切
cos(*x*)	余弦
cot(*x*)	余切
sin(*x*)	正弦
tan(*x*)	正切

9.4. 字符串函数和操作符

本节描述用于检查和操作字符串值的函数和操作符。这里的字符串包括 character、character varying 和 text 类型的值。除非另有说明，下面列出的所有函数都适用于这些类型，但使用 character 类型时，应注意自动填充空格可能带来的影响。一些函数也为位串类型提供了原生实现。

SQL定义了一些字符串函数，它们使用关键字，而不是逗号来分隔参数。详情请见表 9.6，PostgreSQL也提供了这些函数使用正常函数调用语法的版本（见表 9.7）。

> **注意**
>
> 在 PostgreSQL 8.3 之前，由于某些非字符串数据类型到 text 的隐式强制转换，这些函数也会悄然接受那些类型的值。由于这些强制转换经常产生出人意料的行为，它们已被移除。不过，字符串串接操作符（||）仍接受非字符串输入，只要至少一个输入属于字符串类型，如表 9.6所示。对于其他情况，如果需要重现以前的行为，可添加到 text 的显式强制转换。

表 9.6. SQL字符串函数和操作符

函数	*string* || *string*
返回类型	text
描述	字符串串接
示例	'Post' || 'greSQL'
结果	PostgreSQL
函数	*string* || *non-string* 或 *non-string* || *string*
返回类型	text
描述	一个输入为非字符串的字符串串接

示例	`'Value: ' || 42`
结果	`Value: 42`
函数	`bit_length(string)`
返回类型	`int`
描述	字符串中的位数
示例	`bit_length('jose')`
结果	`32`
函数	`char_length(string)` 或 `character_length(string)`
返回类型	`int`
描述	字符串中的字符数
示例	`char_length('jose')`
结果	`4`
函数	`lower(string)`
返回类型	`text`
描述	将字符串转换为小写
示例	`lower('TOM')`
结果	`tom`
函数	`octet_length(string)`
返回类型	`int`
描述	字符串中的字节数
示例	`octet_length('jose')`
结果	`4`
函数	`overlay(string placing string from int [for int])`
返回类型	`text`
描述	替换子字符串
示例	`overlay('Txxxxas' placing 'hom' from 2 for 4)`
结果	`Thomas`
函数	`position(substring in string)`
返回类型	`int`
描述	指定子字符串的位置
示例	`position('om' in 'Thomas')`
结果	`3`
函数	`substring(string [from int] [for int])`
返回类型	`text`

描述	提取子字符串
示例	`substring('Thomas' from 2 for 3)`
结果	`hom`
函数	`substring(string from pattern)`
返回类型	`text`
描述	提取匹配 POSIX 正则表达式的子字符串。有关模式匹配的更多信息，参见第 9.7 节。
示例	`substring('Thomas' from '...$')`
结果	`mas`
函数	`substring(string from pattern for escape)`
返回类型	`text`
描述	提取匹配 SQL 正则表达式的子字符串。有关模式匹配的更多信息，参见第 9.7 节。
示例	`substring('Thomas' from '%#"o_a#"_' for '#')`
结果	`oma`
函数	`trim([leading | trailing | both] [characters] from string)`
返回类型	`text`
描述	从 *string* 的开头、结尾或两端（默认为两端）移除仅由 *characters* 中字符（默认为空格）组成的最长字符串
示例	`trim(both 'x' from 'xTomxx')`
结果	`Tom`
函数	`upper(string)`
返回类型	`text`
描述	将字符串转换为大写
示例	`upper('tom')`
结果	`TOM`

还有其他字符串操作函数，列在表 9.7中。其中一些用于内部实现表 9.6中列出的 SQL 标准字符串函数。

表 9.7. 其他字符串函数

函数	`ascii(string)`
返回类型	`int`
描述	参数首字符的 ASCII 码。对于 UTF8，返回该字符的 Unicode 码点。对于其他多字节编码，参数必须是 ASCII 字符。
示例	`ascii('x')`
结果	`120`
函数	`btrim(string text [, characters text])`
返回类型	`text`

描述	从 *string* 的开头和结尾移除仅由 *characters* 中字符（默认为空格）组成的最长字符串
示例	`btrim('xyxtrimyyx', 'xy')`
结果	`trim`
函数	`chr(int)`
返回类型	`text`
描述	指定编码值对应的字符。对于 UTF8，参数被视为 Unicode 码点。对于其他多字节编码，参数必须指定一个 ASCII 字符。不允许 NULL（0）字符，因为文本数据类型无法存储这样的字节。
示例	`chr(65)`
结果	`A`
函数	`concat(str "any" [, str "any" [, ...] ])`
返回类型	`text`
描述	串接所有参数的文本表示。忽略 NULL 参数。
示例	`concat('abcde', 2, NULL, 22)`
结果	`abcde222`
函数	`concat_ws(sep text, str "any" [, str "any" [, ...] ])`
返回类型	`text`
描述	使用分隔符串接除第一个参数以外的所有参数。第一个参数用作分隔字符串。忽略 NULL 参数。
示例	`concat_ws(',', 'abcde', 2, NULL, 22)`
结果	`abcde,2,22`
函数	`convert(string bytea, src_encoding name, dest_encoding name)`
返回类型	`bytea`
描述	将字符串转换为 *dest_encoding*。原编码由 *src_encoding* 指定。*string* 必须是该编码下的有效字符串。可以用 `CREATE CONVERSION` 定义转换。此外，也有一些预定义的转换。可用的转换见表 9.8。
示例	`convert('text_in_utf8', 'UTF8', 'LATIN1')`
结果	以 Latin-1 编码（ISO 8859-1）表示的 `text_in_utf8`
函数	`convert_from(string bytea, src_encoding name)`
返回类型	`text`
描述	将字符串转换为数据库编码。原编码由 *src_encoding* 指定。*string* 必须是该编码下的有效字符串。
示例	`convert_from('text_in_utf8', 'UTF8')`
结果	以当前数据库编码表示的 `text_in_utf8`
函数	`convert_to(string text, dest_encoding name)`
返回类型	`bytea`

描述	将字符串转换为 *dest_encoding*。
示例	convert_to('some text', 'UTF8')
结果	以 UTF8 编码表示的 some text
函数	decode(*string* text, *format* text)
返回类型	bytea
描述	从 *string* 中的文本表示解码二进制数据。*format* 的选项与 encode 相同。
示例	decode('MTIzAAE=', 'base64')
结果	\x3132330001
函数	encode(*data* bytea, *format* text)
返回类型	text
描述	将二进制数据编码为文本表示。支持的格式为：base64、hex、escape。escape 将零字节和最高位为 1 的字节转换为八进制序列（*nnn*），并将反斜杠双写。
示例	encode(E'123\\000\\001', 'base64')
结果	MTIzAAE=
函数	format(*formatstr* text [, *formatarg* "any" [, ...]])
返回类型	text
描述	格式化字符串。此函数类似于 C 函数 sprintf，但只识别下列转换说明：%s将对应参数作为字符串插入；%I将其参数转义为 SQL 标识符；%L将其参数转义为 SQL 字面量；%%输出一个字面 %。转换可以通过在转换说明符前加上*n*$来引用显式的参数位置，其中*n*是参数位置。另见例 39.1。
示例	format('Hello %s, %1$s', 'World')
结果	Hello World, World
函数	initcap(*string*)
返回类型	text
描述	将每个单词的首字母转换为大写，其余字母转换为小写。单词是由非字母数字字符分隔的字母数字字符序列。
示例	initcap('hi THOMAS')
结果	Hi Thomas
函数	left(*str* text, *n* int)
返回类型	text
描述	返回字符串中的前 *n* 个字符。当 *n* 为负数时，返回除最后 |*n*| 个字符以外的所有字符。
示例	left('abcde', 2)
结果	ab
函数	length(*string*)
返回类型	int
描述	*string* 中的字符数
示例	length('jose')

结果	4
函数	length(*string* bytea, *encoding* name)
返回类型	int
描述	采用给定 *encoding* 的 *string* 中的字符数。*string* 必须是该编码下的有效字符串。
示例	length('jose', 'UTF8')
结果	4
函数	lpad(*string* text, *length* int [, *fill* text])
返回类型	text
描述	在 *string* 前面添加字符 *fill*（默认为空格），将其填充到长度 *length*。如果 *string* 已经长于 *length*，则从右侧截断。
示例	lpad('hi', 5, 'xy')
结果	xyxhi
函数	ltrim(*string* text [, *characters* text])
返回类型	text
描述	从 *string* 的开头移除仅由 *characters* 中字符（默认为空格）组成的最长字符串
示例	ltrim('zzzytrim', 'xyz')
结果	test
函数	md5(*string*)
返回类型	text
描述	计算 *string* 的 MD5 hash，并以十六进制返回结果
示例	md5('abc')
结果	900150983cd24fb0 d6963f7d28e17f72
函数	pg_client_encoding()
返回类型	name
描述	当前客户端编码名称
示例	pg_client_encoding()
结果	SQL_ASCII
函数	quote_ident(*string* text)
返回类型	text
描述	为给定字符串添加适当的引号并返回，使其可用作 SQL 语句字符串中的标识符。仅在必要时添加引号（即字符串包含不能用于标识符的字符，或会被大小写折叠时）。内嵌引号会适当地双写。另见例 39.1。
示例	quote_ident('Foo bar')
结果	"Foo bar"
函数	quote_literal(*string* text)
返回类型	text

描述	为给定字符串添加适当的引号并返回，使其可用作 SQL 语句字符串中的字符串字面量。内嵌单引号和反斜杠会被适当地双写。注意，`quote_literal` 在输入为 null 时返回 null；如果参数可能为 null，`quote_nullable` 通常更合适。另见例 39.1。
示例	`quote_literal(E'O\'Reilly')`
结果	`'O''Reilly'`
函数	`quote_literal(value anyelement)`
返回类型	`text`
描述	将给定值强制转换为文本，然后作为字面量加引号。内嵌单引号和反斜杠会被适当地双写。
示例	`quote_literal(42.5)`
结果	`'42.5'`
函数	`quote_nullable(string text)`
返回类型	`text`
描述	为给定字符串添加适当的引号并返回，使其可用作 SQL 语句字符串中的字符串字面量；如果参数为 null，则返回 `NULL`。内嵌单引号和反斜杠会被适当地双写。另见例 39.1。
示例	`quote_nullable(NULL)`
结果	`NULL`
函数	`quote_nullable(value anyelement)`
返回类型	`text`
描述	将给定值强制转换为文本，然后作为字面量加引号；如果参数为 null，则返回 `NULL`。内嵌单引号和反斜杠会被适当地双写。
示例	`quote_nullable(42.5)`
结果	`'42.5'`
函数	`regexp_matches(string text, pattern text [, flags text])`
返回类型	`setof text[]`
描述	返回 POSIX 正则表达式与 *string* 匹配所得到的所有捕获子串。更多信息参见第 9.7.3 节。
示例	`regexp_matches('foobarbequebaz', '(bar)(beque)')`
结果	`{bar,beque}`
函数	`regexp_replace(string text, pattern text, replacement text [, flags text])`
返回类型	`text`
描述	替换匹配 POSIX 正则表达式的子字符串。更多信息参见第 9.7.3 节。
示例	`regexp_replace('Thomas', '.[mN]a.', 'M')`
结果	`ThM`
函数	`regexp_split_to_array(string text, pattern text [, flags text ])`
返回类型	`text[]`

描述	使用 POSIX 正则表达式作为分隔符拆分 *string*。更多信息参见第 9.7.3 节。
示例	`regexp_split_to_array('hello world', '\s+')`
结果	`{hello,world}`
函数	`regexp_split_to_table(string text, pattern text [, flags text])`
返回类型	`setof text`
描述	使用 POSIX 正则表达式作为分隔符拆分 *string*。更多信息参见第 9.7.3 节。
示例	`regexp_split_to_table('hello world', '\s+')`

结果	`hello`
	`world`
	（2 行）

函数	`repeat(string text, number int)`
返回类型	`text`
描述	将 *string* 重复指定的 *number* 次
示例	`repeat('Pg', 4)`
结果	`PgPgPgPg`
函数	`replace(string text, from text, to text)`
返回类型	`text`
描述	将 *string* 中每次出现的子字符串 *from* 替换为子字符串 *to*
示例	`replace('abcdefabcdef', 'cd', 'XX')`
结果	`abXXefabXXef`
函数	`reverse(str)`
返回类型	`text`
描述	返回反转后的字符串。
示例	`reverse('abcde')`
结果	`edcba`
函数	`right(str text, n int)`
返回类型	`text`
描述	返回字符串中的最后 *n* 个字符。当 *n* 为负数时，返回除前 |*n*| 个字符以外的所有字符。
示例	`right('abcde', 2)`
结果	`de`
函数	`rpad(string text, length int [, fill text])`
返回类型	`text`
描述	在 *string* 后面追加字符 *fill*（默认为空格），将其填充到长度 *length*。如果 *string* 已经长于 *length*，则截断它。
示例	`rpad('hi', 5, 'xy')`

结果	hixyx
函数	rtrim(*string* text [, *characters* text])
返回类型	text
描述	从 *string* 的结尾移除仅由 *characters* 中字符（默认为空格）组成的最长字符串
示例	rtrim('testim', 'xyz')
结果	test
函数	split_part(*string* text, *delimiter* text, *field* int)
返回类型	text
描述	按 *delimiter* 拆分 *string*，并返回指定字段（从一开始计数）
示例	split_part('abc~@~def~@~ghi', '~@~', 2)
结果	def
函数	strpos(*string*, *substring*)
返回类型	int
描述	指定子字符串的位置（与 position(*substring* in *string*) 相同，但注意参数顺序相反）
示例	strpos('high', 'ig')
结果	2
函数	substr(*string*, *from* [, *count*])
返回类型	text
描述	提取子字符串（与 substring(*string* from *from* for *count*) 相同）
示例	substr('alphabet', 3, 2)
结果	ph
函数	to_ascii(*string* text [, *encoding* text])
返回类型	text
描述	将 *string* 从其他编码转换为 ASCII（只支持从 LATIN1、LATIN2、LATIN9 和 WIN1250 编码转换）
示例	to_ascii('Karel')
结果	Karel
函数	to_hex(*number* int or bigint)
返回类型	text
描述	将 *number* 转换为等价的十六进制表示
示例	to_hex(2147483647)
结果	7fffffff
函数	translate(*string* text, *from* text, *to* text)
返回类型	text

描述	将 *string* 中与 *from* 集合中某字符匹配的每个字符替换为 *to* 集合中的对应字符。如果 *from* 比 *to* 长，则移除输入中出现的 *from* 中的多余字符。
示例	`translate('12345', '143', 'ax')`
结果	`a2x5`

另见第 9.18 节中的聚合函数 `string_agg`。

表 9.8. 内置转换

转换名称 [a]	源编码	目标编码
ascii_to_mic	SQL_ASCII	MULE_INTERNAL
ascii_to_utf8	SQL_ASCII	UTF8
big5_to_euc_tw	BIG5	EUC_TW
big5_to_mic	BIG5	MULE_INTERNAL
big5_to_utf8	BIG5	UTF8
euc_cn_to_mic	EUC_CN	MULE_INTERNAL
euc_cn_to_utf8	EUC_CN	UTF8
euc_jp_to_mic	EUC_JP	MULE_INTERNAL
euc_jp_to_sjis	EUC_JP	SJIS
euc_jp_to_utf8	EUC_JP	UTF8
euc_kr_to_mic	EUC_KR	MULE_INTERNAL
euc_kr_to_utf8	EUC_KR	UTF8
euc_tw_to_big5	EUC_TW	BIG5
euc_tw_to_mic	EUC_TW	MULE_INTERNAL
euc_tw_to_utf8	EUC_TW	UTF8
gb18030_to_utf8	GB18030	UTF8
gbk_to_utf8	GBK	UTF8
iso_8859_10_to_utf8	LATIN6	UTF8
iso_8859_13_to_utf8	LATIN7	UTF8
iso_8859_14_to_utf8	LATIN8	UTF8
iso_8859_15_to_utf8	LATIN9	UTF8
iso_8859_16_to_utf8	LATIN10	UTF8
iso_8859_1_to_mic	LATIN1	MULE_INTERNAL
iso_8859_1_to_utf8	LATIN1	UTF8
iso_8859_2_to_mic	LATIN2	MULE_INTERNAL
iso_8859_2_to_utf8	LATIN2	UTF8
iso_8859_2_to_windows_1250	LATIN2	WIN1250
iso_8859_3_to_mic	LATIN3	MULE_INTERNAL
iso_8859_3_to_utf8	LATIN3	UTF8
iso_8859_4_to_mic	LATIN4	MULE_INTERNAL
iso_8859_4_to_utf8	LATIN4	UTF8
iso_8859_5_to_koi8_r	ISO_8859_5	KOI8R

转换名称 [a]	源编码	目标编码
iso_8859_5_to_mic	ISO_8859_5	MULE_INTERNAL
iso_8859_5_to_utf8	ISO_8859_5	UTF8
iso_8859_5_to_windows_1251	ISO_8859_5	WIN1251
iso_8859_5_to_windows_866	ISO_8859_5	WIN866
iso_8859_6_to_utf8	ISO_8859_6	UTF8
iso_8859_7_to_utf8	ISO_8859_7	UTF8
iso_8859_8_to_utf8	ISO_8859_8	UTF8
iso_8859_9_to_utf8	LATIN5	UTF8
johab_to_utf8	JOHAB	UTF8
koi8_r_to_iso_8859_5	KOI8R	ISO_8859_5
koi8_r_to_mic	KOI8R	MULE_INTERNAL
koi8_r_to_utf8	KOI8R	UTF8
koi8_r_to_windows_1251	KOI8R	WIN1251
koi8_r_to_windows_866	KOI8R	WIN866
koi8_u_to_utf8	KOI8U	UTF8
mic_to_ascii	MULE_INTERNAL	SQL_ASCII
mic_to_big5	MULE_INTERNAL	BIG5
mic_to_euc_cn	MULE_INTERNAL	EUC_CN
mic_to_euc_jp	MULE_INTERNAL	EUC_JP
mic_to_euc_kr	MULE_INTERNAL	EUC_KR
mic_to_euc_tw	MULE_INTERNAL	EUC_TW
mic_to_iso_8859_1	MULE_INTERNAL	LATIN1
mic_to_iso_8859_2	MULE_INTERNAL	LATIN2
mic_to_iso_8859_3	MULE_INTERNAL	LATIN3
mic_to_iso_8859_4	MULE_INTERNAL	LATIN4
mic_to_iso_8859_5	MULE_INTERNAL	ISO_8859_5
mic_to_koi8_r	MULE_INTERNAL	KOI8R
mic_to_sjis	MULE_INTERNAL	SJIS
mic_to_windows_1250	MULE_INTERNAL	WIN1250
mic_to_windows_1251	MULE_INTERNAL	WIN1251
mic_to_windows_866	MULE_INTERNAL	WIN866
sjis_to_euc_jp	SJIS	EUC_JP
sjis_to_mic	SJIS	MULE_INTERNAL
sjis_to_utf8	SJIS	UTF8
windows_1258_to_utf8	WIN1258	UTF8
uhc_to_utf8	UHC	UTF8
utf8_to_ascii	UTF8	SQL_ASCII
utf8_to_big5	UTF8	BIG5
utf8_to_euc_cn	UTF8	EUC_CN

转换名称 [a]	源编码	目标编码
utf8_to_euc_jp	UTF8	EUC_JP
utf8_to_euc_kr	UTF8	EUC_KR
utf8_to_euc_tw	UTF8	EUC_TW
utf8_to_gb18030	UTF8	GB18030
utf8_to_gbk	UTF8	GBK
utf8_to_iso_8859_1	UTF8	LATIN1
utf8_to_iso_8859_10	UTF8	LATIN6
utf8_to_iso_8859_13	UTF8	LATIN7
utf8_to_iso_8859_14	UTF8	LATIN8
utf8_to_iso_8859_15	UTF8	LATIN9
utf8_to_iso_8859_16	UTF8	LATIN10
utf8_to_iso_8859_2	UTF8	LATIN2
utf8_to_iso_8859_3	UTF8	LATIN3
utf8_to_iso_8859_4	UTF8	LATIN4
utf8_to_iso_8859_5	UTF8	ISO_8859_5
utf8_to_iso_8859_6	UTF8	ISO_8859_6
utf8_to_iso_8859_7	UTF8	ISO_8859_7
utf8_to_iso_8859_8	UTF8	ISO_8859_8
utf8_to_iso_8859_9	UTF8	LATIN5
utf8_to_johab	UTF8	JOHAB
utf8_to_koi8_r	UTF8	KOI8R
utf8_to_koi8_u	UTF8	KOI8U
utf8_to_sjis	UTF8	SJIS
utf8_to_windows_1258	UTF8	WIN1258
utf8_to_uhc	UTF8	UHC
utf8_to_windows_1250	UTF8	WIN1250
utf8_to_windows_1251	UTF8	WIN1251
utf8_to_windows_1252	UTF8	WIN1252
utf8_to_windows_1253	UTF8	WIN1253
utf8_to_windows_1254	UTF8	WIN1254
utf8_to_windows_1255	UTF8	WIN1255
utf8_to_windows_1256	UTF8	WIN1256
utf8_to_windows_1257	UTF8	WIN1257
utf8_to_windows_866	UTF8	WIN866
utf8_to_windows_874	UTF8	WIN874
windows_1250_to_iso_8859_2	WIN1250	LATIN2
windows_1250_to_mic	WIN1250	MULE_INTERNAL
windows_1250_to_utf8	WIN1250	UTF8
windows_1251_to_iso_8859_5	WIN1251	ISO_8859_5

转换名称 [a]	源编码	目标编码
windows_1251_to_koi8_r	WIN1251	KOI8R
windows_1251_to_mic	WIN1251	MULE_INTERNAL
windows_1251_to_utf8	WIN1251	UTF8
windows_1251_to_windows_866	WIN1251	WIN866
windows_1252_to_utf8	WIN1252	UTF8
windows_1256_to_utf8	WIN1256	UTF8
windows_866_to_iso_8859_5	WIN866	ISO_8859_5
windows_866_to_koi8_r	WIN866	KOI8R
windows_866_to_mic	WIN866	MULE_INTERNAL
windows_866_to_utf8	WIN866	UTF8
windows_866_to_windows_1251	WIN866	WIN
windows_874_to_utf8	WIN874	UTF8
euc_jis_2004_to_utf8	EUC_JIS_2004	UTF8
utf8_to_euc_jis_2004	UTF8	EUC_JIS_2004
shift_jis_2004_to_utf8	SHIFT_JIS_2004	UTF8
utf8_to_shift_jis_2004	UTF8	SHIFT_JIS_2004
euc_jis_2004_to_shift_jis_2004	EUC_JIS_2004	SHIFT_JIS_2004
shift_jis_2004_to_euc_jis_2004	SHIFT_JIS_2004	EUC_JIS_2004

[a] 转换名称遵循标准命名规则：将源编码正式名称中的所有非字母数字字符替换为下划线，再接上 _to_，然后接上按相同方式处理的目标编码名称。因此，这些名称可能与惯用的编码名称不同。

9.5. 二进制串函数和操作符

本节描述用于检查和操作 bytea 类型值的函数和操作符。

SQL定义了一些使用关键字而不是逗号来分隔参数的字符串函数。详情请见表 9.9。PostgreSQL也提供了这些函数使用常规函数调用语法的版本（参阅表 9.10）。

> **注意**
>
> 本页显示的示例结果假定服务器参数 bytea_output 设为 escape（传统的 PostgreSQL格式）。

表 9.9. SQL二进制串函数和操作符

函数	*string* || *string*
返回类型	bytea
描述	字符串串接
示例	E'\\\\Post'::bytea || E'\\047gres\\000'::bytea
结果	\\Post'gres\000

函数	octet_length(*string*)
返回类型	int
描述	二进制串中的字节数
示例	octet_length(E'jo\\000se'::bytea)
结果	5
函数	overlay(*string* placing *string* from int [for int])
返回类型	bytea
描述	替换子字符串
示例	overlay(E'Th\\000omas'::bytea placing E'\\002\\003'::bytea from 2 for 3)
结果	T\\002\\003mas
函数	position(*substring* in *string*)
返回类型	int
描述	指定子字符串的位置
示例	position(E'\\000om'::bytea in E'Th\\000omas'::bytea)
结果	3
函数	substring(*string* [from int] [for int])
返回类型	bytea
描述	提取子字符串
示例	substring(E'Th\\000omas'::bytea from 2 for 3)
结果	h\000o
函数	trim([both] *bytes* from *string*)
返回类型	bytea
描述	从 *string* 的开头和结尾移除仅由 *bytes* 中字节组成的最长字符串
示例	trim(E'\\000'::bytea from E'\\000Tom\\000'::bytea)
结果	Tom

还有一些二进制串处理函数可以使用，在表 9.10列出。其中有一些是在内部使用，用于实现表 9.9列出的 SQL 标准串函数。

表 9.10. 其他二进制串函数

函数	btrim(*string* bytea, *bytes* bytea)
返回类型	bytea
描述	从 *string* 的开头和结尾移除仅由 *bytes* 中字节组成的最长字符串
示例	btrim(E'\\000trim\\000'::bytea, E'\\000'::bytea)
结果	trim

函数	decode(*string* text, *format* text)
返回类型	bytea
描述	从 *string* 中的文本表示解码二进制数据。*format* 的选项与 encode 相同。
示例	decode(E'123\\000456', 'escape')
结果	123\000456
函数	encode(*data* bytea, *format* text)
返回类型	text
描述	将二进制数据编码为文本表示。支持的格式为：base64、hex、escape。escape 将零字节和最高位为 1 的字节转换为八进制序列（*nnn*），并将反斜杠双写。
示例	encode(E'123\\000456'::bytea, 'escape')
结果	123\000456
函数	get_bit(*string*, *offset*)
返回类型	int
描述	从字符串中提取位
示例	get_bit(E'Th\\000omas'::bytea, 45)
结果	1
函数	get_byte(*string*, *offset*)
返回类型	int
描述	从字符串中提取字节
示例	get_byte(E'Th\\000omas'::bytea, 4)
结果	109
函数	length(*string*)
返回类型	int
描述	二进制串的长度
示例	length('jo\000se'::bytea)
结果	5
函数	md5(*string*)
返回类型	text
描述	计算 *string* 的 MD5 hash，并以十六进制返回结果
示例	md5('Th\000omas'::bytea)
结果	8ab2d3c9689aaf18 b4958c334c82d8b1
函数	set_bit(*string*, *offset*, *newvalue*)
返回类型	bytea
描述	设置字符串中的位
示例	set_bit('Th\000omas'::bytea, 45, 0)

结果	Th\000omAs
函数	set_byte(*string*, *offset*, *newvalue*)
返回类型	bytea
描述	设置字符串中的字节
示例	set_byte('Th\000omas'::bytea, 4, 64)
结果	Th\000o@as

函数get_byte和set_byte把二进制串中的第一个字节编号为字节 0。函数get_bit和set_bit在每一个字节中从右边起计数位；例如位 0 是第一个字节的最低有效位，而位 15 是第二个字节的最高有效位。

9.6. 位串函数和操作符

本节描述用于检查和操作位串的函数和操作符，也就是 bit 和 bit varying 类型的值。除常用比较操作符外，还可使用表 9.11中的操作符。&、| 和 # 的位串操作数长度必须相同。移位时会保留字符串的原始长度，如示例所示。

表 9.11. 位串操作符

操作符	描述	示例	结果
||	串接	B'10001' || B'011'	10001011
&	按位与	B'10001' & B'01101'	00001
|	按位或	B'10001' | B'01101'	11101
#	按位异或	B'10001' # B'01101'	11100
~	按位非	~ B'10001'	01110
<<	按位左移	B'10001' << 3	01000
>>	按位右移	B'10001' >> 2	00100

以下 SQL 标准函数既适用于位串，也适用于字符串：length、bit_length、octet_length、position、substring、overlay。

以下函数既适用于位串，也适用于二进制串：get_bit、set_bit。处理位串时，这些函数将字符串的第一位（最左侧的位）编号为位 0。

此外，还可以在整数值与bit类型之间进行类型转换。例如：

```
44::bit(10)                    0000101100
44::bit(3)                     100
cast(-44 as bit(12))           111111010100
'1110'::bit(4)::integer        14
```

注意，仅转换为 “bit” 意味着转换为bit(1)，因此只会得到该整数的最低有效位。

注意

在 PostgreSQL 8.0 之前，将整数转换为 `bit(n)` 会复制整数最左侧的 `n` 位，而现在复制的是最右侧的 `n` 位。此外，将整数转换为宽度大于该整数自身的位串时，会在左侧进行符号扩展。

9.7. 模式匹配

PostgreSQL提供了三种独立的模式匹配方法：传统的SQL `LIKE`操作符、较新的`SIMILAR TO`操作符（在 SQL:1999 中加入），以及POSIX风格的正则表达式。除了用于判断“这个字符串是否匹配这个模式？”的基本操作符外，还提供了提取或替换匹配子字符串、
在匹配位置分割字符串的函数。

提示

如果你的模式匹配的要求超出了这些，请考虑用 Perl 或 Tcl 写一个用户定义的函数。

小心

虽然大多数正则表达式搜索都能很快完成，
但特意构造的正则表达式可能需要任意长的处理时间和任意多的内存。
接受来自恶意来源的正则表达式搜索模式时应当谨慎。如果必须这样做，
建议设置语句超时。

使用`SIMILAR TO`模式进行搜索具有同样的安全风险，因为`SIMILAR TO`提供了许多与POSIX风格正则表达式相同的能力。

`LIKE`搜索比另外两种方法简单得多，因此，当模式可能来自恶意来源时，使用它更安全。

9.7.1. LIKE

```
string LIKE pattern [ESCAPE escape-character]
string NOT LIKE pattern [ESCAPE escape-character]
```

如果该*string*匹配了提供的*pattern*，那么`LIKE`表达式返回真（和预期的一样，如果`LIKE`返回真，那么`NOT LIKE`表达式返回假，反之亦然。一个等效的表达式是`NOT (string LIKE pattern)`）。

如果*pattern*不包含百分号或下划线，那么该模式只表示它本身的字符串；在这种情况下，`LIKE`的行为就像等号操作符。*pattern*中的下划线（`_`）代表（匹配）任意单个字符；
百分号（`%`）匹配任意由零个或多个字符组成的序列。

一些示例：

```
'abc' LIKE 'abc'    true
'abc' LIKE 'a%'     true
'abc' LIKE '_b_'    true
'abc' LIKE 'c'      false
```

LIKE模式匹配总是覆盖整个字符串。因此，如果想要匹配字符串内任意位置上的一个序列，该模式就必须以百分号开头并以百分号结尾。

要匹配字面量下划线或百分号而不是把它们当作通配符，*pattern*中相应的字符前面必须带有转义字符。默认的转义字符是反斜线，但也可以使用ESCAPE子句选择其他转义字符。要匹配转义字符本身，请写两个转义字符。

> **注意**
>
> 如果你关掉了standard_conforming_strings，你在字符串常量中写的任何反斜线都需要被双写。详见第 4.1.2.1 节。

也可以通过写 ESCAPE '' 来选择不使用转义字符。这会禁用转义机制，从而无法关闭模式中下划线和百分号的特殊含义。

可以用关键字 ILIKE 代替 LIKE，使匹配根据当前区域设置忽略大小写。这不属于 SQL 标准，而是 PostgreSQL 的扩展。

操作符~~等效于LIKE，而~~*对应ILIKE。还有 !~~和!~~*操作符分别代表NOT LIKE和NOT ILIKE。所有这些操作符都是PostgreSQL特有的。

9.7.2. SIMILAR TO正则表达式

string SIMILAR TO *pattern* [ESCAPE *escape-character*]
string NOT SIMILAR TO *pattern* [ESCAPE *escape-character*]

SIMILAR TO 操作符根据其模式是否匹配给定字符串返回真或假。它与 LIKE 类似，但按照 SQL 标准定义的正则表达式来解释模式。SQL 正则表达式是 LIKE 表示法和常见正则表达式表示法的一种奇特结合。

与LIKE类似，SIMILAR TO操作符只有在其模式匹配整个字符串时才算成功；这一点不同于普通正则表达式，后者可以匹配字符串的任意部分。与LIKE相同，SIMILAR TO也使用_和%作为通配符，分别匹配任意单个字符和任意字符串（分别类似于 POSIX 正则表达式中的.和.*）。

除了这些从LIKE借用的功能之外，SIMILAR TO支持下面这些从 POSIX 正则表达式借用的模式匹配元字符：

- |表示选择（两个候选之一）。
- *表示重复前面的项零次或更多次。
- +表示重复前面的项一次或更多次。
- ?表示重复前面的项零次或一次。
- {*m*}表示重复前面的项刚好*m*次。
- {*m*,}表示重复前面的项*m*次或更多次。
- {*m*,*n*}表示重复前面的项至少*m*次并且不超过*n*次。
- 可以使用圆括号()把多个项组合成一个逻辑项。
- 一个方括号表达式[...]声明一个字符类，就像 POSIX 正则表达式一样。

注意点号（.）不是SIMILAR TO的一个元字符。

与LIKE一样，反斜杠会禁用这些元字符的特殊含义；也可以用 ESCAPE 指定其他转义字符。

下面是一些示例：

```
'abc' SIMILAR TO 'abc'      true
'abc' SIMILAR TO 'a'        false
'abc' SIMILAR TO '%(b|d)%'  true
'abc' SIMILAR TO '(b|c)%'   false
```

带三个参数的 substring 函数，即 substring(*string* from *pattern* for *escape-character*)，用于提取匹配 SQL 正则表达式模式的子串。与 SIMILAR TO 一样，指定模式必须匹配整个数据字符串，否则函数失败并返回空值。要指出成功时应返回的模式部分，模式中必须出现两次后跟双引号（"）的转义字符。
函数返回匹配这两个标记之间的模式部分的文本。

下面是一些示例，其中使用#"界定返回字符串：

```
substring('foobar' from '%#"o_b#"%' for '#')   oob
substring('foobar' from '#"o_b#"%' for '#')    NULL
```

9.7.3. POSIX正则表达式

表 9.12列出了所有可用于 POSIX 正则表达式模式匹配的操作符。

表 9.12. 正则表达式匹配操作符

操作符	描述	示例
~	匹配正则表达式，区分大小写	'thomas' ~ '.*thomas.*'
~*	匹配正则表达式，不区分大小写	'thomas' ~* '.*Thomas.*'
!~	不匹配正则表达式，区分大小写	'thomas' !~ '.*Thomas.*'
!~*	不匹配正则表达式，不区分大小写	'thomas' !~* '.*vadim.*'

POSIX正则表达式提供了比LIKE和SIMILAR TO操作符更强大的模式匹配方式。许多 Unix 工具，例如egrep、sed或awk，都使用与这里描述的模式匹配语言相似的语言。

正则表达式是一个字符序列，是定义一组字符串（一个正则集）的简写。
如果字符串属于正则表达式描述的正则集，就称该字符串匹配此正则表达式。与LIKE一样，
模式中的字符精确匹配字符串中的字符，除非该模式字符在正则表达式语言中有特殊含义 – 但正则表达式使用的特殊字符与LIKE不同。与LIKE模式不同，
正则表达式可以匹配字符串中的任意位置，除非显式将其锚定到字符串开头或末尾。

下面是一些示例：

```
'abc' ~ 'abc'    true
'abc' ~ '^a'     true
'abc' ~ '(b|d)'  true
'abc' ~ '^(b|c)' false
```

POSIX模式语言的详细描述见下文。

带两个参数的 substring 函数，即 substring(*string* from *pattern*)，用于提取匹配 POSIX 正则表达式模式的子串。如果没有匹配，它返回空值；否则返回匹配模式的那部分文本。

但是，如果模式包含圆括号，则返回匹配第一个括号子表达式（左圆括号最先出现的那个）的文本。如果要在表达式内部使用圆括号而不触发这个例外，可以在整个表达式外再加一对圆括号。如果需要在想要提取的子表达式之前使用圆括号，请参见下文介绍的非捕获圆括号。

下面是一些示例：

```
substring('foobar' from 'o.b')      oob
substring('foobar' from 'o(.)b')    o
```

`regexp_replace`函数用新文本替换匹配 POSIX 正则表达式模式的子字符串。它的语法为 `regexp_replace(`*source*`,` *pattern*`,` *replacement* `[,` *flags* `])`。如果没有与 *pattern* 匹配的内容，则原样返回 *source* 字符串。如果存在匹配，则返回将匹配子字符串替换为 *replacement* 字符串后的 *source* 字符串。*replacement* 字符串可以包含 `\`*n*，其中 *n* 为 1 至 9，表示应插入与模式中第 *n* 个圆括号子表达式匹配的源子字符串；它也可以包含 `\&`，表示应插入与整个模式匹配的子字符串。如果需要在替换文本中放置字面的反斜杠，应写成 `\\`。*flags* 参数是可选的文本字符串，其中包含零个或多个改变函数行为的单字母标志。标志 `i` 指定不区分大小写的匹配，而标志 `g` 指定替换每个匹配的子字符串，而不仅是第一个。其他支持的标志在表 9.20中描述。

下面是一些示例：

```
regexp_replace('foobarbaz', 'b..', 'X')
                                    fooXbaz
regexp_replace('foobarbaz', 'b..', 'X', 'g')
                                    fooXX
regexp_replace('foobarbaz', 'b(..)', E'X\\1Y', 'g')
                                    fooXarYXazY
```

`regexp_matches`函数返回一个文本数组，其中包含与 POSIX 正则表达式模式匹配所得到的所有捕获子串。它的语法为 `regexp_matches(`*string*`,` *pattern* `[,` *flags* `])`。该函数可以不返回行、返回一行或多行（见下面的`g`标志）。如果*pattern*不匹配，该函数不返回任何行。如果模式不包含圆括号子表达式，则返回的每一行都是包含与整个模式匹配的子字符串的单元素文本数组。如果模式包含圆括号子表达式，该函数返回一个文本数组，其中第*n*个元素是匹配模式的第*n*个圆括号子表达式的子字符串（不包括“非捕获”括号；详情见下文）。*flags* 参数是一个可选的文本字符串，其中包含零个或多个单个字母标志，用于更改函数的行为。标志`g`使该函数查找字符串中的每个匹配项，而不仅仅是第一个，并为每个这样的匹配返回一行。其他支持的标志在表 9.20中描述。

下面是一些示例：

```
SELECT regexp_matches('foobarbequebaz', '(bar)(beque)');
 regexp_matches
----------------
 {bar,beque}
(1 row)

SELECT regexp_matches('foobarbequebazilbarfbonk', '(b[^b]+)(b[^b]+)',
 'g');
 regexp_matches
----------------
 {bar,beque}
 {bazil,barf}
(2 rows)
```

```
SELECT regexp_matches('foobarbequebaz', 'barbeque');
 regexp_matches
----------------
 {barbeque}
(1 row)
```

可以通过使用子查询强制regexp_matches()总是返回一行；
当希望返回所有行（包括不匹配的行）时，这在SELECT目标列表中特别有用：

```
SELECT col1, (SELECT regexp_matches(col2, '(bar)(beque)')) FROM tab;
```

regexp_split_to_table把一个 POSIX 正则表达式模式当作一个定界符来分割字符串。它的语法形式是regexp_split_to_table(*string*, *pattern* [, *flags*])。如果没有与*pattern*的匹配，该函数返回*string*。如果至少有一个匹配，
对每一个匹配它都返回从上一个匹配的末尾（或者串的开头）到这次匹配开头之间的文本。
当没有更多匹配时，它返回从上一次匹配的末尾到串末尾之间的文本。*flags*
参数是一个可选的文本串，它包含零个或更多单字母标志，这些标志可以改变该函数的行为。
regexp_split_to_table能支持的标志在表 9.20中描述。

regexp_split_to_array函数的行为和regexp_split_to_table相同，不过regexp_split_to_array会把它的结果以一个text数组的形式返回。它的语法是regexp_split_to_array(*string*, *pattern* [, *flags*])。这些参数和regexp_split_to_table的相同。

下面是一些示例：

```
SELECT foo FROM regexp_split_to_table('the quick brown fox jumped over
 the lazy dog', E'\\s+') AS foo;
  foo
--------
 the
 quick
 brown
 fox
 jumped
 over
 the
 lazy
 dog
(9 rows)

SELECT regexp_split_to_array('the quick brown fox jumped over the lazy
 dog', E'\\s+');
              regexp_split_to_array
-----------------------------------------------
 {the,quick,brown,fox,jumped,over,the,lazy,dog}
(1 row)

SELECT foo FROM regexp_split_to_table('the quick brown fox', E'\\s*')
 AS foo;
 foo
-----
 t
 h
```

```
 e
 q
 u
 i
 c
 k
 b
 r
 o
 w
 n
 f
 o
 x
(16 rows)
```

正如最后一个示例所示，
正则表达式分割函数会忽略出现在字符串开头或结尾或紧跟在前一个匹配项之后的零长度匹配。
这与`regexp_matches`实现的严格的正则表达式匹配定义相矛盾，
但在实践中通常是最方便的行为。其他软件系统如Perl使用类似的定义。

9.7.3.1. 正则表达式细节

PostgreSQL的正则表达式是使用 Henry Spencer 写的一个包来实现的。下面的正则表达式的大部分描述都是从他的手册页中逐字拷贝过来的。

正则表达式（RE），在POSIX 1003.2 中定义，它有两种形式：扩展的RE或者是ERE（大概地说就是那些在`egrep`里的），基本的RE或者是BRE（大概地说就是那些在`ed`里的）。
PostgreSQL支持两种形式，并且还实现了一些POSIX标准中没有但是在类似 Perl 或者 Tcl 这样的语言中得到广泛应用的一些扩展。使用了那些非POSIX扩展的RE叫高级RE，或者本文档里说的ARE。ARE 几乎完全是 ERE 的超集，但是 BRE 有几个符号上的不兼容（以及更多的限制）。我们首先描述 ARE 和 ERE 形式，描述那些只适用于 ARE 的特性，然后描述 BRE 的区别是什么。

> **注意**
>
> PostgreSQL最初总是假定正则表达式遵循 ARE 规则。不过，也可以像第 9.7.3.4 节中所述那样，在 RE 模式前加上一个嵌入选项，从而改用限制更多的 ERE 或 BRE 规则。这对于需要精确遵循POSIX 1003.2 规则的应用有助于保持兼容性。

一个正则表达式被定义为一个或更多分支，它们之间被 `|` 分隔。
只要能匹配其中一个分支的东西都能匹配正则表达式。

一个分支是零个或多个量化原子或约束，连接在一起。它匹配第一个的匹配项，
然后是第二个的匹配项，依此类推；一个空分支匹配空字符串。

一个量化原子是一个原子，后面可以跟一个量词。没有量词时，匹配一次原子所匹配的内容；
有量词时，按量词指定的次数匹配原子所匹配的内容。原子可以是表 9.13列出的任何一种形式。
可用量词及其含义见表 9.14。

一个约束匹配一个空串，但只是在满足特定条件下才匹配。
约束可以在能够使用原子的地方使用，只是它不能跟着量词。简单的约束在表 9.15里显示；更多的约束稍后描述。

表 9.13. 正则表达式原子

原子	描述
(*re*)	（其中*re*是任意正则表达式）匹配*re*所匹配的内容，并记录该匹配，以备输出结果
(?:*re*)	同上，但不记录匹配结果（“非捕获”圆括号；仅适用于 ARE）
.	匹配任意单个字符
[*chars*]	一个方括号表达式，匹配*chars*中的任意一个（详见第 9.7.3.2 节）
k	（其中*k*既不是字母也不是数字）把该字符视为普通字符并匹配它，例如，\\匹配反斜线字符
c	其中*c*是字母或数字（后面可能还有其他字符），这是一个转义，参见第 9.7.3.3 节（仅适用于 ARE；在 ERE 和 BRE 中，它匹配*c*）
{	如果后面跟着非数字字符，则匹配左花括号{；如果后面跟着数字，则是*bound*的开头（见下文）
x	其中*x*是一个没有其它意义的单个字符，则匹配该字符

RE 不能以反斜线（\）结尾。

> **注意**
>
> 如果你关掉了standard_conforming_strings，
> 你在字符串常量中写的任何反斜线都需要被双写。详见第 4.1.2.1 节。

表 9.14. 正则表达式量词

量词	匹配
*	一个由原子的 0 次或更多次匹配组成的序列
+	一个由原子的 1 次或更多次匹配组成的序列
?	一个由原子的 0 次或 1 次匹配组成的序列
{*m*}	一个由原子的正好*m*次匹配组成的序列
{*m*,}	一个由原子的*m*次或更多次匹配组成的序列
{*m*,*n*}	一个由原子的从*m*次到*n*次（包括）匹配组成的序列；*m*不能超过*n*
*?	*的非贪婪版本
+?	+的非贪婪版本
??	?的非贪婪版本
{*m*}?	{*m*}的非贪婪版本
{*m*,}?	{*m*,}的非贪婪版本
{*m*,*n*}?	{*m*,*n*}的非贪婪版本

使用{...}的形式被称作范围。一个范围内的数字*m*和*n*都是无符号十进制整数，允许的数值从 0 到 255（包含）。

非贪婪量词（仅适用于 ARE）与对应的普通（贪婪）量词匹配相同的可能内容，但优先选择最少的匹配次数，而不是最多的匹配次数。详见第 9.7.3.5 节。

> **注意**
>
> 一个量词不能紧跟在另外一个量词后面，例如`**`是非法的。
> 量词不能作为表达式或者子表达式的开头，也不能跟在`^`或者`|`后面。

表 9.15. 正则表达式约束

约束	描述
`^`	在字符串开头匹配
`$`	在字符串末尾匹配
`(?=`*re*`)`	正向先行断言在这样的位置匹配：存在从该位置开始且匹配*re*的子字符串（仅适用于 ARE）
`(?!`*re*`)`	负向先行断言在这样的位置匹配：不存在任何从该位置开始且匹配*re*的子字符串（仅适用于 ARE）

先行约束不能包含反向引用（参见第 9.7.3.3 节），其中的所有圆括号都视为非捕获圆括号。

9.7.3.2. 方括号表达式

方括号表达式是用`[]`括起来的字符列表。通常，
它匹配列表中的任意单个字符（但请参见下文）。如果列表以`^`开头，则匹配任意
不在列表剩余部分中的单个字符。如果列表中的两个字符用`-`分隔，
则表示排序序列中这两个字符之间的完整字符范围（包含两个端点）；例如，ASCII中的`[0-9]`
匹配任意十进制数字。两个范围共享一个端点是不合法的，例如`a-c-e`。
范围高度依赖排序序列，因此可移植程序应避免依赖它们。

想在列表中包含字面字符`]`，可以让它做列表的首字符（如果使用了`^`，需要放在其后）。
想在列表中包含字面字符`-`，可以让它做列表的首字符或者尾字符，或者一个范围的第二个端点。
想在列表中把字面字符`-`当做范围的起点，把它用`[.`和`.]`包围起来，
这样它就成为一个排序元素（见下文）。除了这些字符本身、一些用`[`的组合（见下段）
以及转义（只在 ARE 中有效）以外，所有其它特殊字符在方括号表达式里都失去它们的特殊含义。特别是，在 ERE 和 BRE 规则下`\`不是特殊的，但在 ARE 里，它是特殊的（引入一个转义）。

在一个方括号表达式里，一个排序元素（一个字符、
一个被当做一个单一字符排序的多字符序列或者一个表示上面两种情况的排序序列名称）包含在`[.`和`.]`里面的时候表示该排序元素的字符序列。该序列被当做该方括号列表的一个单一元素。
这允许一个包含多字符排序元素的方括号表达式去匹配多于一个字符，例如，
如果排序序列包含一个`ch`排序元素，那么 RE `[[.ch.]]*c`匹配`chchcc`的头五个字符。

> **注意**
>
> PostgreSQL当前不支持多字符排序元素。这些信息描述了将来可能有的行为。

在方括号表达式里，包围在`[=`和`=]`里的排序元素是一个等价类，
代表等效于那一个的所有排序元素的字符序列，包括它本身（如果没有其它等效排序元素，
那么就好像封装定界符是`[.`和 `.]`）。例如，如果`o`和`^`是一个等价类的成员，那么`[[=o=]]`、`[[=^=]]`和`[o^]`都是同义的。一个等价类不能是一个范围的端点。

在方括号表达式中，用 `[:` 和 `:]` 括起的字符类名表示属于该类的所有字符的列表。标准字符类名包括：`alnum`、`alpha`、`blank`、`cntrl`、`digit`、`graph`、`lower`、`print`、`punct`、`space`、`upper`、`xdigit`。它们表示 ctype 中定义的字符类。区域设置可以提供其他字符类。字符类不能用作范围的端点。

方括号表达式有两个特例：`[[:<:]]` 和 `[[:>:]]` 是约束，分别匹配单词开头和结尾的空字符串。单词定义为一个前后都没有单词字符的单词字符序列。单词字符是 `alnum` 字符（由 ctype 定义）或下划线。这是与 POSIX 1003.2 兼容但未由其规定的扩展，在打算移植到其他系统的软件中应谨慎使用。通常更适合使用下文介绍的约束转义；它们并不更标准，但更容易键入。

9.7.3.3. 正则表达式转义

转义是以`\`开头，后面跟着一个字母或数字字符的特殊序列。转义有好几种变体：字符输入转义、字符类简写转义、约束转义以及反向引用。在 ARE 里，如果一个`\`后面跟着一个字母或数字，但是并未组成一个合法的转义，那么它是非法的。在 ERE 中没有转义：在方括号表达式之外，一个后面跟着字母或数字字符的`\`只是表示该字符是一个普通的字符，而且在一个方括号表达式里，`\`是一个普通的字符（后者是 ERE 与 ARE 之间唯一的实际不兼容之处）。

字符输入转义便于在 RE 中指定不可打印或其他不便输入的字符，见表 9.16。

字符类简写转义用来提供一些常用的字符类简写。它们显示在表 9.17中。

约束转义是以转义形式书写的约束，在满足特定条件时匹配空字符串，见表 9.18。

反向引用（`\`*n*）匹配数字*n*指定的被前面的圆括号子表达式匹配的同一个串（参阅表 9.19）。例如，`([bc])\1`匹配`bb`或者`cc`，但是不匹配`bc`或者`cb`。RE 中子表达式必须完全在反向引用前面。子表达式以它们的左圆括号的顺序编号。非捕获圆括号并不定义子表达式。

> **注意**
>
> 请记住，把模式作为 SQL 字符串常量输入时，转义的开头`\`需要双写。例如：
>
> `'123' ~ E'^\\d{3}'` *true*

表 9.16. 正则表达式字符输入转义

转义	描述
`\a`	警告（响铃）字符，和 C 中一样
`\b`	退格，和 C 中一样
`\B`	反斜线（`\`）的同义词，用来减少双写反斜线
`\c`*X*	（其中*X*是任意字符）低序5位和*X*相同的字符，它的其他位都是零
`\e`	排序序列名称为`ESC`的字符；若不存在这样的字符，则使用八进制值为`033`的字符
`\f`	换页，和 C 中一样
`\n`	换行符，与 C 中相同
`\r`	回车，和 C 中一样
`\t`	水平制表符，和 C 中一样
`\u`*wxyz*	（其中*wxyz*正好是四个十六进制位）十六进制值为`0x`*wxyz*的字符

转义	描述
\U*stuvwxyz*	（其中*stuvwxyz*正好是八个十六进制位）十六进制值为0x*stuvwxyz*的字符
\v	垂直制表符，和 C 中一样
\x*hhh*	（其中*hhh*是十六进制位的任意序列）十六进制值为0x*hhh*的字符（一个单一字符，不管用了多少个十六进制位）
\0	值为0（空字节）的字符
xy	（其中*xy*正好是两个八进制位，并且不是一个反向引用）八进制值为0*xy*的字符
xyz	（其中*xyz*正好是三个八进制位，并且不是一个反向引用）八进制值为0*xyz*的字符

十六进制位是0-9、a-f和A-F。八进制位是0-7。

指定 ASCII 范围（0–127）之外的值的数字字符输入转义的含义取决于数据库编码。当编码是 UTF-8 时，转义值等价于 Unicode 代码点，例如 \u1234表示字符U+1234。对于其他多字节编码，字符输入转义通常只是指定该字符的字节值的串接。
如果该转义值不对应数据库编码中的任何合法字符，将不会发生错误，
但是它不会匹配任何数据。

字符输入转义总是被当作普通字符。例如，\135是 ASCII 中的]，但\135并不终止一个方括号表达式。

表 9.17. 正则表达式字符类简写转义

转义	描述
\d	[[:digit:]]
\s	[[:space:]]
\w	[[:alnum:]_]（注意包括下划线）
\D	[^[:digit:]]
\S	[^[:space:]]
\W	[^[:alnum:]_]（注意包括下划线）

在方括号表达式中，\d、\s 和 \w 会失去其外层方括号，而 \D、\S 和 \W 则是非法的。（因此，例如 [a-c\d] 等同于 [a-c[:digit:]]。此外，等同于 [a-c^[:digit:]] 的 [a-c\D] 是非法的。）

表 9.18. 正则表达式约束转义

转义	描述
\A	只在串开头匹配（与^的不同请参见第 9.7.3.5 节）
\m	只在一个词的开头匹配
\M	只在一个词的末尾匹配
\y	只在一个词的开头或末尾匹配
\Y	只在不属于单词开头或末尾的位置匹配
\Z	只在串的末尾匹配（与$的不同请参见第 9.7.3.5 节）

单词的定义与上文[[:<:]]和[[:>:]]的说明相同。方括号表达式中不允许使用约束转义。

表 9.19. 正则表达式反向引用

转义	描述
m	（其中*m*是一个非零数字）一个到第*m*个子表达式的反向引用
mnn	（其中*m*是一个非零数字，并且*nn*是后续的若干数字，并且十进制值*mnn*不大于此前已出现的捕获右圆括号数）一个到第*mnn*个子表达式的反向引用

> **注意**
>
> 八进制字符输入转义与反向引用之间存在固有歧义，
> 按上文提到的启发式规则解决：前导零始终表示八进制转义。单个非零数字，
> 如果后面没有其他数字，始终视为反向引用。不以零开头的多位数字序列，
> 如果前面已有相应的子表达式（即该数字在反向引用的合法范围内），则视为反向引用，
> 否则视为八进制转义。

9.7.3.4. 正则表达式元语法

除了上面描述的主要语法之外，还有几种特殊形式和杂项语法。

RE 可以以两种特殊的引导前缀之一开头。如果 RE 以`***:`开头，余下部分就被视为 ARE。（这在PostgreSQL中通常没有影响，因为 RE 默认被视为 ARE；但如果通过正则表达式函数的*flags*参数指定了 ERE 或 BRE 模式，它就会产生影响。）如果 RE 以`***=`开头，余下部分就作为按字面解释的字符串处理，所有字符都视为普通字符。

一个 ARE 可以以嵌入选项开头：一个序列`(?`*xyz*`)`（这里的*xyz*是一个或多个字母字符）声明影响剩余 RE 的选项。这些选项覆盖任何先前确定的选项 – 特别地，它们可以覆盖一个正则表达式操作符隐含的大小写敏感的行为，或者覆盖正则表达式函数的*flags*参数所指定的选项。
可用的选项字母在表 9.20中显示。注意这些同样的选项字母也被用在正则表达式函数的*flags*参数中。

表 9.20. ARE 嵌入选项字母

选项	描述
`b`	RE的剩余部分是一个BRE
`c`	大小写敏感的匹配（覆盖操作符类型）
`e`	RE的剩余部分是一个ERE
`i`	大小写不敏感的匹配（见第 9.7.3.5 节）（覆盖操作符类型）
`m`	`n`的历史原因的同义词
`n`	换行敏感的匹配（见第 9.7.3.5 节）
`p`	部分换行敏感的匹配（见第 9.7.3.5 节）
`q`	RE 的剩余部分按字面（“加引号”）解释，所有字符都视为普通字符
`s`	非换行敏感的匹配（默认）
`t`	紧凑语法（默认；见下文）
`w`	逆部分换行敏感（“怪异”）的匹配（见第 9.7.3.5 节）

选项	描述
x	扩展语法（见下文）

嵌入选项从结束该序列的)处开始生效。它们只能出现在 ARE 的开头（如果存在***:引导前缀，则位于该前缀之后）。

除了通常的紧凑 RE 语法（其中所有字符都有意义）之外，还有一种扩展语法，可以通过指定嵌入的x选项来使用。在扩展语法中，RE 中的空白字符会被忽略，同样被忽略的还有#与其后的换行符（或 RE 末尾）之间的所有字符。这使得复杂的 RE 可以分段并添加注释。此基本规则有三个例外：

- 空白字符或 # 的前面若有 \，该字符就会被保留。
- 方括号表达式里的空白或者#将被保留
- 在多字符符号里面不能出现空白和注释，例如(?:

在这里，空白字符包括空格、制表符、换行符，以及属于*space*字符类的任何字符。

最后，在 ARE 里，方括号表达式外面，序列(?#*ttt*)（其中*ttt*是任意不包含一个)的文本）是一个注释，它被完全忽略。同样，这样的东西是不允许出现在多字符符号的字符中间的，例如 (?:。这种注释更像是一种历史产物而不是一种有用的设施，并且它们的使用已经被废弃；请使用扩展语法来替代。

如果指定了开头的***=引导前缀，那么这些元语法扩展都不能使用，因为这表示把用户输入作为按字面解释的字符串，而非 RE 处理。

9.7.3.5. 正则表达式匹配规则

在 RE 可以在给定串中匹配多于一个子串的情况下，RE 匹配串中最靠前的那个子串。如果 RE 可以匹配在那个位置开始的多个子串，要么是取最长的子串，要么是最短的，具体哪种，取决于 RE 是贪婪的还是非贪婪的。

一个 RE 是否贪婪取决于下面规则：

- 大多数原子以及所有约束，都没有贪婪属性（因为它们毕竟无法匹配长度不定的文本）。
- 在一个 RE 周围加上圆括号并不会改变其贪婪性。
- 带一个固定重复次数量词（{*m*}或者{*m*}?）的量化原子和原子自身具有同样的贪婪性（可能是没有）。
- 一个带其他普通的量词（包括{*m*,*n*}中*m*等于*n*的情况）的量化原子是贪婪的（首选最长匹配）。
- 一个带非贪婪量词（包括{*m*,*n*}?中*m*等于 *n*的情况）的量化原子是非贪婪的（首选最短匹配）。
- 一个分支 – 也就是说，一个没有顶级|操作符的 RE – 和它里面的第一个有贪婪属性的量化原子有着同样的贪婪性。
- 一个由|操作符连接起来的两个或者更多分支组成的 RE 总是贪婪的。

上面的规则所描述的贪婪属性不仅仅适用于独立的量化原子，而且也适用于包含量化原子的分支和整个 RE。这里的意思是，匹配是按照分支或者整个 RE 作为一个整体匹配最长或者最短的可能子串。一旦整个匹配的长度确定，

那么匹配任意特定子表达式的部分就基于该子表达式的贪婪属性进行判断，在 RE 里面靠前的子表达式的优先级高于靠后的子表达式。

一个相应的示例：

```
SELECT SUBSTRING('XY1234Z', 'Y*([0-9]{1,3})');
结果：123
SELECT SUBSTRING('XY1234Z', 'Y*?([0-9]{1,3})');
结果：1
```

在第一个示例里，RE 作为整体是贪婪的，因为`Y*`是贪婪的。它可以匹配从`Y`开始的东西，并且它匹配从这个位置开始的最长的串，也就是，`Y123`。输出是这里的圆括号包围的部分，或者说是`123`。在第二个示例里，RE 总体上是一个非贪婪的 RE，因为`Y*?`是非贪婪的。它可以匹配从`Y`开始的最短的子串，也就是说`Y1`。子表达式`[0-9]{1,3}`是贪婪的，
但是它不能修改总体匹配长度的决定；因此它被迫只匹配`1`。

简而言之，如果一个 RE 同时包含贪婪和非贪婪的子表达式，那么总的匹配长度要么是尽可能长，要么是尽可能短，这取决于给整个 RE 赋予的属性。给子表达式赋予的属性只影响在这个匹配里，各个子表达式相对于其他子表达式能“吃掉”多少内容。

量词`{1,1}`和`{1,1}?`可以分别用于在一个子表达式或者整个 RE 上强制贪婪或者非贪婪。当需要整个 RE 具有不同于从其元素中推导出的贪婪属性时，这很有用。例如，假设我们尝试将一个包含一些数字的字符串分隔成数字以及在它们之前和之后的部分，我们可能会尝试这样做：

```
SELECT regexp_matches('abc01234xyz', '(.*)(\d+)(.*)');
结果：{abc0123,4,xyz}
```

这不会有用：第一个`.*`是贪婪的，因此它会“吃掉”尽可能多的字符而留下`\d+`
去匹配在最后一个可能位置上的最后一个数字。我们可能会通过让它变成非贪婪来修复：

```
SELECT regexp_matches('abc01234xyz', '(.*?)(\d+)(.*)');
结果：{abc,0,""}
```

这也不会有用：因为现在 RE 作为整体来说是非贪婪的，因此它会尽快结束全部的匹配。我们可以通过强制 RE 整体是贪婪的来得到我们想要的：

```
SELECT regexp_matches('abc01234xyz', '(?:(.*?)(\d+)(.*)){1,1}');
结果：{abc,01234,xyz}
```

将 RE 的整体贪婪性与各组件自身的贪婪性分开控制，为处理变长模式提供了很大的灵活性。

在决定更长或者更短的匹配时，匹配长度是以字符衡量的，而不是排序元素。
一个空串会被认为比什么都不匹配长。例如：`bb*`匹配`abbbc`的中间三个字符；`(week|wee)(night|knights)`匹配`weeknights`的所有十个字符；而`(.*).*`匹配 `abc`的时候，圆括号包围的子表达式匹配所有三个字符；当`(a*)*`被拿来匹配`bc`时，整个 RE 和圆括号子表达式都匹配一个空串。

如果指定不区分大小写的匹配，其效果近似于字母表中的所有大小写差别都消失了。
当存在大小写形式的字母作为普通字符出现在方括号表达式之外时，
实际上会转换为包含其大小写形式的方括号表达式，例如`x`变成`[xX]`。
当它出现在方括号表达式内部时，其所有大小写形式都会加入该表达式，例如`[x]`变成`[xX]`，`[^x]`变成`[^xX]`。

如果指定了换行敏感的匹配，`.`和使用`^`的方括号表达式将永远不会匹配换行字符（这样，
匹配就不会跨越行，除非 RE 显式安排了跨行匹配）并且`^`和`$`除了分别匹配串开头和结尾之外，还将分别匹配换行后面和前面的空串。但是 ARE 转义`\A`和`\Z`仍然只匹配串的开头和结尾。

如果指定了部分换行敏感的匹配，那么它影响 . 和方括号表达式，
这个时候和换行敏感的匹配一样，但是不影响 ^ 和 $。

如果指定了逆部分换行敏感匹配，那么它影响 ^ 和 $，其作用和在换行敏感的匹配里一样，
但是不影响 . 和方括号表达式。这个并不是很有用，只是为了满足对称性而提供的。

9.7.3.6. 限制和兼容性

在这个实现里，对 RE 的长度没有特别的限制。但是，那些希望高移植性的程序应该避免使用长度超过 256 字节的 RE，因为 POSIX 兼容的实现可以拒绝接受这样的 RE。

ARE 与 POSIX ERE 实际不兼容的唯一特性是：\在方括号表达式中不会失去特殊含义。其他所有 ARE 特性所使用的语法，在 POSIX ERE 中都是非法的，或其效果未定义或未指定；引导前缀的 *** 语法同样不属于 POSIX 的 BRE 或 ERE 语法。

许多 ARE 扩展借鉴自 Perl，但其中一些经过了整理和修改，也未实现少数 Perl 扩展。需要注意的不兼容之处包括 \b、\B、不对末尾换行符作特殊处理、
取反的方括号表达式也受换行敏感匹配影响、先行约束中对圆括号及反向引用的限制，
以及采用最长或最短匹配而非首次匹配的语义。

ARE 与旧版 ERE 语法存在两点重要的不兼容，这里比较的是 7.4 之前的PostgreSQL：

- 在 ARE 中，\ 后跟字母或数字字符时，要么表示转义，要么是错误；而在以前的版本中，它只是书写该字母或数字字符的另一种方式。这应该不是什么大问题，
  因为以前的版本没有理由使用这样的序列。
- 在 ARE 中，\ 在 [] 内仍是特殊字符，因此方括号表达式内的字面 \ 必须写成 \\。

9.7.3.7. 基本正则表达式

BRE 在几个方面和 ERE 不太一样。在 BRE 中，|、+ 和 ? 都是普通字符并且没有与它们功能等价的东西。范围的定界符是 \{ 和 \}，而 { 和 } 本身是普通字符。嵌套的子表达式的圆括号是 \(和 \)，而 (和) 自身是普通字符。除非在 RE 开头或者是圆括号子表达式开头，^ 都是一个普通字符。除非在 RE 结尾或者是圆括号子表达式的结尾，$ 是一个普通字符。如果 * 出现在 RE 开头或者是圆括号封装的子表达式开头（前面可能有 ^），那么它是个普通字符。最后，可以用单数字的反向引用，\< 和 \> 分别是 [[:<:]] 和 [[:>:]] 的同义词；在 BRE 中没有其它可用的转义。

9.8. 数据类型格式化函数

PostgreSQL格式化函数提供一套强大的工具用于把各种数据类型（日期/时间、整数、浮点数、数值）转换成格式化的字符串以及反过来从格式化的字符串转换成指定的数据类型。表 9.21列出了这些函数。这些函数都遵循一个公共的调用规范：第一个参数是待格式化的值，
而第二个是一个定义输出或输入格式的模板。

还有一个单参数的 to_timestamp 函数；它接受一个 double precision 参数，并把 Unix 纪元（自 1970-01-01 00:00:00+00 起的秒数）转换为 timestamp with time zone。（Integer Unix 纪元会被隐式转换为 double precision。）

表 9.21. 格式化函数

函数	返回类型	描述	示例
to_char(timestamp, text)	text	将时间戳转换为字符串	to_char(current_timestamp, 'HH12:MI:SS')

函数	返回类型	描述	示例
to_char(interval, text)	text	将时间间隔转换为字符串	to_char(interval '15h 2m 12s', 'HH24:MI:SS')
to_char(int, text)	text	将整数转换为字符串	to_char(125, '999')
to_char(double precision, text)	text	将 real/double precision 转换为字符串	to_char(125.8::real, '999D9')
to_char(numeric, text)	text	将 numeric 转换为字符串	to_char(-125.8, '999D99S')
to_date(text, text)	date	将字符串转换为日期	to_date('05 Dec 2000', 'DD Mon YYYY')
to_number(text, text)	numeric	将字符串转换为 numeric	to_number('12,454.8-', '99G999D9S')
to_timestamp(text, text)	timestamp with time zone	将字符串转换为时间戳	to_timestamp('05 Dec 2000', 'DD Mon YYYY')
to_timestamp(double precision)	timestamp with time zone	把 Unix 纪元转换为时间戳	to_timestamp(1284352323)

在一个to_char输出模板串中，
一些特定的模式可以被识别并且被替换成基于给定值的被恰当地格式化的数据。
任何不属于模板模式的文本都简单地照字面拷贝。同样，在一个输入模板串里（对其他函数），
模板模式标识由输入数据串提供的值。

表 9.22展示了可以用于格式化日期和时间值的模板模式。

表 9.22. 用于日期/时间格式化的模板模式

模式	描述
HH	一天中的小时（01-12）
HH12	一天中的小时（01-12）
HH24	一天中的小时（00-23）
MI	分钟（00-59）
SS	秒（00-59）
MS	毫秒（000-999）
US	微秒（000000-999999）
SSSS	自午夜起的秒数（0-86399）
AM, am, PM 或 pm	上午/下午标记（不带句点）
A.M., a.m., P.M. 或 p.m.	上午/下午标记（带句点）
Y,YYY	带逗号的年（4 位或者更多位）
YYYY	年（4 位或者更多位）
YYY	年的最后 3 位数字
YY	年的最后 2 位数字
Y	年的最后 1 位数字
IYYY	ISO 8601 周编号方式的年（4 位或更多位）

模式	描述
IYY	ISO 8601 周编号方式的年的最后 3 位数字
IY	ISO 8601 周编号方式的年的最后 2 位数字
I	ISO 8601 周编号方式的年的最后 1 位数字
BC, bc, AD 或 ad	纪元指示器（不带句号）
B.C., b.c., A.D. 或 a.d.	纪元指示器（带句号）
MONTH	大写的月份全称（空格补齐到 9 字符）
Month	首字母大写的月份全称（空格补齐到 9 字符）
month	小写的月份全称（空格补齐到 9 字符）
MON	简写的大写形式的月名（英文 3 字符，本地化长度可变）
Mon	简写的首字母大写形式的月名（英文 3 字符，本地化长度可变）
mon	简写的小写形式的月名（英文 3 字符，本地化长度可变）
MM	月份编号（01-12）
DAY	大写的星期全称（空格补齐到 9 字符）
Day	首字母大写的星期全称（空格补齐到 9 字符）
day	小写的星期全称（空格补齐到 9 字符）
DY	大写的星期简称（英语 3 字符，本地化长度可变）
Dy	首字母大写的星期简称（英语 3 字符，本地化长度可变）
dy	小写的星期简称（英语 3 字符，本地化长度可变）
DDD	一年中的第几天（001-366）
IDDD	ISO 8601 周编号年中的第几天（001-371；一年的第 1 天是第一个 ISO 周的星期一）
DD	月内日序数（01-31）
D	星期几，周日 (1) 到周六 (7)
ID	ISO 8601 星期几，周一 (1) 到周日 (7)
W	月内周序数（1-5）（第一周从该月第一天开始）
WW	一年中的第几周（1-53）（第一周从该年第一天开始）
IW	ISO 8601 周编号年中的第几周（01-53；该年的第一个星期四位于第 1 周）
CC	世纪（2 位数）（21 世纪开始于 2001-01-01）
J	儒略日（从公元前 4714 年 11 月 24 日 UTC 午夜开始的整数日数）
Q	季度（被to_date和to_timestamp忽略）
RM	以大写罗马数字表示的月份（I-XII；I=一月）
rm	以小写罗马数字表示的月份（i-xii；i=一月）
TZ	大写时区名称
tz	小写时区名称

修饰符可以被应用于模板模式来修改它们的行为。例如，FMMonth就是带着FM修饰符的Month模式。表 9.23展示了可用于日期/时间格式化的修饰符模式。

表 9.23. 用于日期/时间格式化的模板模式修饰符

修饰符	描述	示例
FM 前缀	填充模式（抑制前导零和填充的空格）	FMMonth
TH 后缀	大写形式的序数后缀	DDTH，例如，12TH
th 后缀	小写形式的序数后缀	DDth，例如，12th
FX 前缀	固定格式全局选项（见使用须知）	FX Month DD Day
TM 前缀	翻译模式（根据lc_time输出本地化的星期名和月份名）	TMMonth
SP 后缀	拼写模式（未实现）	DDSP

日期/时间格式化的使用注意事项：

- FM抑制了在模式输出中添加前导零和尾随空格的行为，这些前导零和尾随空格本来会被添加以使输出成为固定宽度。在PostgreSQL中，FM仅修改下一个格式说明，而在Oracle中FM影响所有后续格式说明，并且重复的FM修饰符切换填充模式的开启和关闭。

- TM不包含尾随空格。

- to_timestamp 和 to_date 会跳过输入字符串中的多个空格，除非使用 FX 选项。例如，to_timestamp('2000 JUN', 'YYYY MON') 可以工作，但 to_timestamp('2000 JUN', 'FXYYYY MON') 会报错，因为 to_timestamp 只接受一个空格。FX 必须指定为模板中的第一项。

- to_char 模板中允许普通文本，并会按字面输出。可以用双引号括起子串，使其即使包含模式关键字也强制按字面文本解释。例如，在 '"Hello Year "YYYY' 中，YYYY 会被年份数据替换，但 Year 中单独的 Y 不会被替换。在 to_date、to_number 和 to_timestamp 中，双引号字符串会跳过与该字符串所含字符数相同数量的输入字符，例如 "XX" 跳过两个输入字符。

- 如果要在输出中包含双引号，必须在它前面加上反斜杠，例如 '\"YYYY Month\"'。

- 在从字符串到timestamp或date的转换中，YYYY转换在处理超过4位数字的年份时有限制。您必须在YYYY后使用一些非数字字符或模板，否则年份总是被解释为4位数字。例如（使用年份20000）：to_date('200001131', 'YYYYMMDD')将被解释为4位年份；而应该在年份后使用非数字分隔符，如 to_date('20000-1131', 'YYYY-MMDD')或 to_date('20000Nov31', 'YYYYMonDD')。

- 在从字符串到timestamp或date的转换中，如果存在YYY、YYYY或Y,YYY字段，则忽略CC（世纪）字段。如果CC与 YY或Y一起使用，则年份按(CC-1)*100+YY计算。如果指定了世纪但未指定年份，则假定为该世纪的第一年。

- 可以向to_timestamp和to_date以下列两种方式之一指定 ISO 8601 周编号日期（与公历日期不同）：

  - 年份、周编号和星期几：例如to_date('2006-42-4', 'IYYY-IW-ID')返回日期2006-10-19。如果省略星期几，则假定它为 1（星期一）。

  - 年份和年内日序数：例如to_date('2006-291', 'IYYY-IDDD')也返回2006-10-19。

  尝试使用ISO 8601周编号字段和公历日期字段的混合输入日期是荒谬的，并将导致错误。在ISO 8601周编号年的背景下，“月份”或“月内日序数”的概念没有意义。在公历年的背景下，ISO周没有意义。

> **小心**
>
> 虽然to_date会拒绝混合使用公历和ISO周编号日期字段，但to_char不会，因为输出格式规范如YYYY-MM-DD (IYYY-IDDD)可能很有用。但要避免编写类似IYYY-MM-DD的内容；那会在年初附近产生令人惊讶的结果。（有关更多信息，请参见第 9.9.1 节。）

- 在从字符串到timestamp的转换中，毫秒（MS）或微秒（US）值被用作小数点后的秒数位。例如to_timestamp('12:3', 'SS:MS')不是3毫秒，而是300，因为转换将其计为12 + 0.3秒。这意味着对于格式SS:MS，输入值12:3、12:30和12:300指定相同数量的毫秒。要获得三毫秒，必须写成12:003，转换将其计为12 + 0.003 = 12.003秒。

  这是一个更复杂的示例：to_timestamp('15:12:02.020.001230', 'HH:MI:SS.MS.US'）为15小时12分钟，秒数为2秒 + 20毫秒 + 1230微秒 = 2.021230秒。

- to_char(..., 'ID')的星期几编号与extract(isodow from ...)函数匹配，但 to_char(..., 'D') 的星期几编号与 extract(dow from ...) 的不一致。

- to_char(interval) 按照如12小时制时钟上显示的方式格式化HH和HH12，即零小时和36小时都输出为12，而HH24输出完整的小时值，对于时间间隔它可以超过23。

表 9.24展示了可以用于格式化数值的模板模式。

表 9.24. 用于数值格式化的模板模式

模式	描述
9	数位（非有效位可以被省略）
0	数位（即便是非有效位也不会被省略）
.（句点）	小数点
,（逗号）	分组（千）分隔符
PR	尖括号内的负值
S	紧贴数值的正负号（使用区域设置）
L	货币符号（使用区域设置）
D	小数点（使用区域设置）
G	分组分隔符（使用区域设置）
MI	在指定位置的负号（如果数字 < 0）
PL	在指定位置的正号（如果数字 > 0）
SG	在指定位置的正/负号
RN	罗马数字（输入在 1 和 3999 之间）
TH 或 th	序数后缀
V	移动指定位数（参阅注解）
EEEE	科学记数的指数

数值格式化的使用注意事项：

- 9产生一个位数与9的个数相同的值。如果没有可用的数字，则输出一个空格。

- 使用SG、PL或MI格式化的正负号不紧贴数值；例如，to_char(-12, 'MI9999')会产生'-12'，但to_char(-12, 'S9999')会产生' -12'。（Oracle实现不允许在9之前使用MI，而是要求9在MI之前。）
- TH不会转换小于零的值，也不会转换小数。
- PL，SG和 TH是PostgreSQL 的扩展。
- V实际上将输入值乘以10^*n*，其中*n*是跟在V后面的数字位数。to_char 不支持与小数点结合使用的V（例如，不允许使用99.9V99）。
- EEEE（科学计数法）不能与任何其他格式模式或修饰符结合使用，除了数字和小数点模式之外，必须位于格式字符串的末尾（例如，9.99EEEE 是一个有效模式）。

某些修饰符可以被应用到任何模板来改变其行为。例如，FM9999是带有FM修饰符的9999模式。表 9.25中展示了用于数值格式化的模式修饰符。

表 9.25. 用于数值格式化的模板模式修饰符

修饰符	描述	示例
FM 前缀	填充模式（抑制前导零和填充的空白）	FM9999
TH 后缀	大写序数后缀	999TH
th 后缀	小写序数后缀	999th

表 9.26展示了一些使用to_char函数的示例。

表 9.26. to_char示例

表达式	结果
to_char(current_timestamp, 'Day, DD HH12:MI:SS')	'Tuesday , 06 05:39:18'
to_char(current_timestamp, 'FMDay, FMDD HH12:MI:SS')	'Tuesday, 6 05:39:18'
to_char(-0.1, '99.99')	' -.10'
to_char(-0.1, 'FM9.99')	'-.1'
to_char(0.1, '0.9')	' 0.1'
to_char(12, '9990999.9')	' 0012.0'
to_char(12, 'FM9990999.9')	'0012.'
to_char(485, '999')	' 485'
to_char(-485, '999')	'-485'
to_char(485, '9 9 9')	' 4 8 5'
to_char(1485, '9,999')	' 1,485'
to_char(1485, '9G999')	' 1 485'
to_char(148.5, '999.999')	' 148.500'
to_char(148.5, 'FM999.999')	'148.5'
to_char(148.5, 'FM999.990')	'148.500'
to_char(148.5, '999D999')	' 148,500'
to_char(3148.5, '9G999D999')	' 3 148,500'

表达式	结果
to_char(-485, '999S')	'485-'
to_char(-485, '999MI')	'485-'
to_char(485, '999MI')	'485 '
to_char(485, 'FM999MI')	'485'
to_char(485, 'PL999')	'+485'
to_char(485, 'SG999')	'+485'
to_char(-485, 'SG999')	'-485'
to_char(-485, '9SG99')	'4-85'
to_char(-485, '999PR')	'<485>'
to_char(485, 'L999')	'DM 485'
to_char(485, 'RN')	' CDLXXXV'
to_char(485, 'FMRN')	'CDLXXXV'
to_char(5.2, 'FMRN')	'V'
to_char(482, '999th')	' 482nd'
to_char(485, '"Good number:"999')	'Good number: 485'
to_char(485.8, '"Pre:"999" Post:" .999')	'Pre: 485 Post: .800'
to_char(12, '99V999')	' 12000'
to_char(12.4, '99V999')	' 12400'
to_char(12.45, '99V9')	' 125'
to_char(0.0004859, '9.99EEEE')	' 4.86e-04'

9.9. 日期/时间函数和操作符

表 9.28展示了可用于处理日期/时间值的函数，其细节在随后的小节中描述。表 9.27演示了基本算术操作符（+、*等）的行为。而与格式化相关的函数，可以参考第 9.8 节。你应当熟悉第 8.5 节中的日期/时间数据类型的背景知识。

所有下文描述的接受time或timestamp
输入的函数和操作符实际上都有两种变体：一种接收time with time zone或timestamp with time zone，另外一种接受time without time zone或者 timestamp without time zone。为了简化，这些变种没有被独立地展示。此外，+和*操作符都是可交换的操作符对（例如，date + integer 和 integer + date）；我们只显示每一对中的一个。

表 9.27. 日期/时间操作符

操作符	示例	结果
+	date '2001-09-28' + integer '7'	date '2001-10-05'
+	date '2001-09-28' + interval '1 hour'	timestamp '2001-09-28 01:00:00'
+	date '2001-09-28' + time '03:00'	timestamp '2001-09-28 03:00:00'
+	interval '1 day' + interval '1 hour'	interval '1 day 01:00:00'

操作符	示例	结果
+	timestamp '2001-09-28 01:00' + interval '23 hours'	timestamp '2001-09-29 00:00:00'
+	time '01:00' + interval '3 hours'	time '04:00:00'
-	- interval '23 hours'	interval '-23:00:00'
-	date '2001-10-01' - date '2001-09-28'	integer '3'（天）
-	date '2001-10-01' - integer '7'	date '2001-09-24'
-	date '2001-09-28' - interval '1 hour'	timestamp '2001-09-27 23:00:00'
-	time '05:00' - time '03:00'	interval '02:00:00'
-	time '05:00' - interval '2 hours'	time '03:00:00'
-	timestamp '2001-09-28 23:00' - interval '23 hours'	timestamp '2001-09-28 00:00:00'
-	interval '1 day' - interval '1 hour'	interval '1 day -01:00:00'
-	timestamp '2001-09-29 03:00' - timestamp '2001-09-27 12:00'	interval '1 day 15:00:00'
*	900 * interval '1 second'	interval '00:15:00'
*	21 * interval '1 day'	interval '21 days'
*	double precision '3.5' * interval '1 hour'	interval '03:30:00'
/	interval '1 hour' / double precision '1.5'	interval '00:40:00'

表 9.28. 日期/时间函数

函数	age(timestamp, timestamp)
返回类型	interval
描述	将两个参数相减，生成一个使用年和月，而不是只用日的“符号化”的结果
示例	age(timestamp '2001-04-10', timestamp '1957-06-13')
结果	43 years 9 mons 27 days
函数	age(timestamp)
返回类型	interval
描述	从 current_date（午夜）减去
示例	age(timestamp '1957-06-13')
结果	43 years 8 mons 3 days
函数	clock_timestamp()
返回类型	timestamp with time zone

描述	当前日期和时间（在语句执行期间变化）；参见第 9.9.4 节
示例	
结果	
函数	`current_date`
返回类型	`date`
描述	当前日期；参见第 9.9.4 节
示例	
结果	
函数	`current_time`
返回类型	`time with time zone`
描述	一天中的当前时刻；参见第 9.9.4 节
示例	
结果	
函数	`current_timestamp`
返回类型	`timestamp with time zone`
描述	当前日期和时间（当前事务开始时）；参见第 9.9.4 节
示例	
结果	
函数	`date_part(text, timestamp)`
返回类型	`double precision`
描述	获取子字段（等价于 `extract`）；参见第 9.9.1 节
示例	`date_part('hour', timestamp '2001-02-16 20:38:40')`
结果	`20`
函数	`date_part(text, interval)`
返回类型	`double precision`
描述	获取子字段（等价于 `extract`）；参见第 9.9.1 节
示例	`date_part('month', interval '2 years 3 months')`
结果	`3`
函数	`date_trunc(text, timestamp)`
返回类型	`timestamp`
描述	截断到指定精度；另见第 9.9.2 节
示例	`date_trunc('hour', timestamp '2001-02-16 20:38:40')`
结果	`2001-02-16 20:00:00`
函数	`extract(`*`field`*` from timestamp)`
返回类型	`double precision`

描述	获取子字段；参见第 9.9.1 节
示例	`extract(hour from timestamp '2001-02-16 20:38:40')`
结果	`20`
函数	`extract(field from interval)`
返回类型	`double precision`
描述	获取子字段；参见第 9.9.1 节
示例	`extract(month from interval '2 years 3 months')`
结果	`3`
函数	`isfinite(date)`
返回类型	`boolean`
描述	测试日期是否有限（不是正负无穷）
示例	`isfinite(date '2001-02-16')`
结果	`true`
函数	`isfinite(timestamp)`
返回类型	`boolean`
描述	测试时间戳是否有限（不是正负无穷）
示例	`isfinite(timestamp '2001-02-16 21:28:30')`
结果	`true`
函数	`isfinite(interval)`
返回类型	`boolean`
描述	测试时间间隔是否有限
示例	`isfinite(interval '4 hours')`
结果	`true`
函数	`justify_days(interval)`
返回类型	`interval`
描述	调整时间间隔，使 30 天的时段表示为月
示例	`justify_days(interval '35 days')`
结果	`1 mon 5 days`
函数	`justify_hours(interval)`
返回类型	`interval`
描述	调整时间间隔，使 24 小时的时段表示为天
示例	`justify_hours(interval '27 hours')`
结果	`1 day 03:00:00`

函数	justify_interval(interval)
返回类型	interval
描述	使用 justify_days 和 justify_hours调整时间间隔，并额外调整符号
示例	justify_interval(interval '1 mon -1 hour')
结果	29 days 23:00:00
函数	localtime
返回类型	time
描述	一天中的当前时刻；参见第 9.9.4 节
示例	
结果	
函数	localtimestamp
返回类型	timestamp
描述	当前日期和时间（当前事务开始时）；参见第 9.9.4 节
示例	
结果	
函数	now()
返回类型	timestamp with time zone
描述	当前日期和时间（当前事务开始时）；参见第 9.9.4 节
示例	
结果	
函数	statement_timestamp()
返回类型	timestamp with time zone
描述	当前日期和时间（当前语句开始时）；参见第 9.9.4 节
示例	
结果	
函数	timeofday()
返回类型	text
描述	当前的日期和时间（类似 clock_timestamp，但是采用 text 字符串）；参见第 9.9.4 节
示例	
结果	
函数	transaction_timestamp()
返回类型	timestamp with time zone
描述	当前日期和时间（当前事务开始时）；参见第 9.9.4 节
示例	

结果

除了这些函数以外，还支持 SQL 操作符OVERLAPS：

```
(start1, end1) OVERLAPS (start2, end2)
(start1, length1) OVERLAPS (start2, length2)
```

这个表达式在两个时间段（用它们的端点定义）重叠的时候得到真，当它们不重叠时得到假。端点可以用一对日期、时间或者时间戳来指定；或者是用一个后面跟着一个间隔的日期、时间或时间戳来指定。当一对值被提供时，起点或终点都可以被写在前面，OVERLAPS会自动地把较早的值作为起点。每一个时间段被认为是表示半开区间*start* <= *time* < *end*，除非*start*和*end*相等，这种情况下它表示单个时刻。例如这表示两个只有一个共同端点的时间段不重叠。

```
SELECT (DATE '2001-02-16', DATE '2001-12-21') OVERLAPS
       (DATE '2001-10-30', DATE '2002-10-30');
结果: true
SELECT (DATE '2001-02-16', INTERVAL '100 days') OVERLAPS
       (DATE '2001-10-30', DATE '2002-10-30');
结果: false
SELECT (DATE '2001-10-29', DATE '2001-10-30') OVERLAPS
       (DATE '2001-10-30', DATE '2001-10-31');
结果: false
SELECT (DATE '2001-10-30', DATE '2001-10-30') OVERLAPS
       (DATE '2001-10-30', DATE '2001-10-31');
结果: true
```

当把一个interval值加到某个timestamp with time zone值上（或从中减去一个interval值）时，天数部分会按指定的天数推进或退回timestamp with time zone的日期。当跨越夏令时变化时（会话时区设置为识别夏令时的时区），这意味着interval '1 day'不一定等于interval '24 hours'。例如，将会话时区设置为CST7CDT时，timestamp with time zone '2005-04-02 12:00-07' + interval '1 day'会得到timestamp with time zone '2005-04-03 12:00-06'，而向同一个初始timestamp with time zone加上interval '24 hours'会得到timestamp with time zone '2005-04-03 13:00-06'，这是因为在时区CST7CDT中，夏令时变化发生在2005-04-03 02:00。

注意，age返回的months字段可能存在歧义，因为不同月份的天数不同。PostgreSQL在计算不足整月的部分时，会采用两个日期中较早的那个日期所在的月份。例如：age('2004-06-01', '2004-04-30')使用 4 月得到1 mon 1 day，而如果使用 5 月则会得到1 mon 2 days，因为 5 月有 31 天，而 4 月只有 30 天。

9.9.1. EXTRACT, date_part

```
EXTRACT(field FROM source)
```

该extract函数从日期/时间值中提取年份、小时等子字段。*source*必须是以下类型的值表达式：timestamp，time，或interval。（date类型的表达式会转换为timestamp，因此也可以使用。）*field*是一个标识符或字符串，用于选择从源值中提取的字段。extract函数返回的值的类型为double precision。以下是有效的字段名称：

century

世纪

```
SELECT EXTRACT(CENTURY FROM TIMESTAMP '2000-12-16 12:21:13');
结果: 20
SELECT EXTRACT(CENTURY FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 21
```

第一个世纪始于公元 0001-01-01 00:00:00，尽管当时的人们并不知道。这一定义适用于所有采用格里高利历的国家。世纪编号没有 0，而是从 -1 世纪直接跳到 1 世纪。如果你对此有异议，请向梵蒂冈罗马圣彼得大教堂的教皇投诉。

PostgreSQL 8.0 之前的版本不遵循传统的世纪编号方式，而是直接返回年份字段除以 100 的结果。

`day`

对于 `timestamp` 值，表示月份中的日期字段（1 - 31）；对于 `interval` 值，表示天数

```
SELECT EXTRACT(DAY FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 16
SELECT EXTRACT(DAY FROM INTERVAL '40 days 1 minute');
结果: 40
```

`decade`

年份字段除以10

```
SELECT EXTRACT(DECADE FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 200
```

`dow`

一周的日子从星期天（`0`）到星期六（`6`）

```
SELECT EXTRACT(DOW FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 5
```

请注意`extract`函数的星期几编号与`to_char(..., 'D')`函数不同。

`doy`

一年中的第几天（1 - 365/366）

```
SELECT EXTRACT(DOY FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 47
```

`epoch`

对于`date`和`timestamp`值，自1970-01-01 00:00:00 UTC 以来的秒数（可以为负）；对于`interval`值，是间隔的总秒数

```
SELECT EXTRACT(EPOCH FROM TIMESTAMP WITH TIME ZONE '2001-02-16 20:
38:40.12-08');
结果: 982384720.12

SELECT EXTRACT(EPOCH FROM INTERVAL '5 days 3 hours');
结果: 442800
```

下面介绍如何把一个 epoch 值转换回时间戳：

```
SELECT TIMESTAMP WITH TIME ZONE 'epoch' + 982384720.12 * INTERVAL
 '1 second';
```

（to_timestamp函数封装了上述转换。）

hour

小时字段（0 - 23）

```
SELECT EXTRACT(HOUR FROM TIMESTAMP '2001-02-16 20:38:40');
```
结果：20

isodow

一周的日子从星期一（1）到星期日（7）

```
SELECT EXTRACT(ISODOW FROM TIMESTAMP '2001-02-18 20:38:40');
```
结果：7

这与dow相同，除了星期天。这匹配ISO 8601的星期几编号。

isoyear

日期所属的 ISO 8601 周编号年份（不适用于间隔）

```
SELECT EXTRACT(ISOYEAR FROM DATE '2006-01-01');
```
结果：2005
```
SELECT EXTRACT(ISOYEAR FROM DATE '2006-01-02');
```
结果：2006

每个 ISO 8601 周编号年都从包含 1 月 4 日的那一周的星期一开始，因此在 1 月初或 12 月末，ISO 年可能与格里高利年不同。更多信息请参见 week 字段。

PostgreSQL 8.3 之前的版本不支持此字段。

microseconds

秒字段，包括小数部分，乘以1 000 000；注意这包括完整的秒数

```
SELECT EXTRACT(MICROSECONDS FROM TIME '17:12:28.5');
```
结果：28500000

millennium

千年

```
SELECT EXTRACT(MILLENNIUM FROM TIMESTAMP '2001-02-16 20:38:40');
```
结果：3

1900 年代的年份属于第二个千年。第三个千年始于 2001 年 1 月 1 日。

PostgreSQL 8.0 之前的版本不遵循传统的千年编号方式，而是直接返回年份字段除以 1000 的结果。

milliseconds

秒字段，包括小数部分，乘以1000。请注意，这包括完整的秒数。

```
SELECT EXTRACT(MILLISECONDS FROM TIME '17:12:28.5');
```
结果：28500

minute

分钟字段（0 - 59）

```
SELECT EXTRACT(MINUTE FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 38
```

month

对于 timestamp 值，表示一年中的月份编号（1 - 12）；对于 interval 值，表示月数对 12 取模（0 - 11）

```
SELECT EXTRACT(MONTH FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 2
SELECT EXTRACT(MONTH FROM INTERVAL '2 years 3 months');
结果: 3
SELECT EXTRACT(MONTH FROM INTERVAL '2 years 13 months');
结果: 1
```

quarter

日期所在的一年中的季度（1 - 4）

```
SELECT EXTRACT(QUARTER FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 1
```

second

秒字段，包括小数部分（0 - 59[1]）

```
SELECT EXTRACT(SECOND FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 40
SELECT EXTRACT(SECOND FROM TIME '17:12:28.5');
结果: 28.5
```

timezone

与UTC的时区偏移量，以秒为单位。正值对应于UTC东部的时区，
负值对应于UTC西部的时区。（从技术上讲，PostgreSQL使用 UT1 而不是 UTC，因为不处理闰秒。）

timezone_hour

时区偏移的小时部分

timezone_minute

时区偏移的分钟部分

week

一年中按ISO 8601 周编号体系计算的周序号。根据定义，ISO 周从周一开始，一年的第一周包含该年的 1 月 4 日。换句话说，一年的第一个星期四在该年的第 1 周。

在 ISO 周编号体系中，1 月初的日期可能属于前一年的第 52 周或第 53 周，而 12 月末的日期可能属于下一年的第一周。例如，2005-01-01属于 2004 年的第 53 周，2006-01-01属于 2005 年的第 52 周，而2012-12-31属于 2013 年的第一周。建议将isoyear字段与week一起使用，以获得一致的结果。

```
SELECT EXTRACT(WEEK FROM TIMESTAMP '2001-02-16 20:38:40');
结果: 7
```

[1]如果操作系统实现了闰秒，则为 60

year

年份字段。请记住，没有0 AD，所以把BC年份从AD年份中减去时需要小心。

SELECT EXTRACT(YEAR FROM TIMESTAMP '2001-02-16 20:38:40');
结果：2001

extract函数主要的用途是做计算性处理。对于用于显示的日期/时间值格式化，参阅第 9.8 节。

date_part函数仿照传统的Ingres实现，后者对应SQL标准的extract函数：

date_part('*field*', *source*)

注意，此处的*field*参数必须是字符串值，而不能是名称。date_part
的有效字段名与extract相同。

SELECT date_part('day', TIMESTAMP '2001-02-16 20:38:40');
结果：16
SELECT date_part('hour', INTERVAL '4 hours 3 minutes');
结果：4

9.9.2. date_trunc

date_trunc函数在概念上和用于数字的trunc函数类似。

date_trunc('*field*', *source*)

*source*是timestamp或interval类型的值表达式。（类型为date和time
的值会自动转换为timestamp或interval，分别对应这两种输入类型。）*field*
选择输入值的截断精度。返回值的类型是timestamp或interval，
所有小于所选精度的字段都设为零（日和月则设为一）。

*field*的有效值是：

microseconds
milliseconds
second
minute
hour
day
week
month
quarter
year
decade
century
millennium

示例：

SELECT date_trunc('hour', TIMESTAMP '2001-02-16 20:38:40');
结果：2001-02-16 20:00:00

SELECT date_trunc('year', TIMESTAMP '2001-02-16 20:38:40');
结果：2001-01-01 00:00:00

9.9.3. AT TIME ZONE

AT TIME ZONE 结构允许把时间戳转换到不同的时区。表 9.29展示了它的各种变体。

表 9.29. AT TIME ZONE 变体

表达式	返回类型	描述
timestamp without time zone AT TIME ZONE *zone*	timestamp with time zone	将给定的不带时区时间戳视为指定时区中的时间
timestamp with time zone AT TIME ZONE *zone*	timestamp without time zone	将给定的带时区时间戳转换为新时区的时间，结果不带时区标识
time with time zone AT TIME ZONE *zone*	time with time zone	将给定的带时区时间转换为新时区的时间

在这些表达式里，所需的时区 *zone* 可以指定为文本值（例如 'PST'），也可以指定为一个间隔值（例如 INTERVAL '-08:00'）。在文本情况下，时区名称可以按第 8.5.3 节中描述的任意方式指定。

示例（假设本地时区为PST8PDT）：

```
SELECT TIMESTAMP '2001-02-16 20:38:40' AT TIME ZONE 'MST';
结果: 2001-02-16 19:38:40-08

SELECT TIMESTAMP WITH TIME ZONE '2001-02-16 20:38:40-05' AT TIME ZONE
 'MST';
结果: 2001-02-16 18:38:40
```

第一个示例取一个不带时区的时间戳，把它解释为 MST 时间（UTC-7），然后转换为 PST（UTC-8）用于显示。第二个示例取一个以 EST（UTC-5）指定的时间戳，把它转换为 MST（UTC-7）的本地时间。

函数timezone(*zone*, *timestamp*)等效于符合 SQL 标准的结构*timestamp* AT TIME ZONE *zone*。

9.9.4. 当前日期/时间

PostgreSQL提供了许多返回当前日期和时间的函数。这些 SQL 标准的函数全部都按照当前事务的开始时刻返回值：

```
CURRENT_DATE
CURRENT_TIME
CURRENT_TIMESTAMP
CURRENT_TIME(precision)
CURRENT_TIMESTAMP(precision)
LOCALTIME
LOCALTIMESTAMP
LOCALTIME(precision)
LOCALTIMESTAMP(precision)
```

CURRENT_TIME和CURRENT_TIMESTAMP返回带时区的值；LOCALTIME和LOCALTIMESTAMP返回不带时区的值。

CURRENT_TIME、CURRENT_TIMESTAMP、LOCALTIME和 LOCALTIMESTAMP可以有选择地接受一个精度参数，该精度会使结果的秒字段舍入到指定的小数位数。如果没有精度参数，结果将给出可用的全部精度。

下面是一些示例：

```
SELECT CURRENT_TIME;
结果： 14:39:53.662522-05
SELECT CURRENT_DATE;
结果： 2001-12-23
SELECT CURRENT_TIMESTAMP;
结果： 2001-12-23 14:39:53.662522-05
SELECT CURRENT_TIMESTAMP(2);
结果： 2001-12-23 14:39:53.66-05
SELECT LOCALTIMESTAMP;
结果： 2001-12-23 14:39:53.662522
```

因为这些函数全部都按照当前事务的开始时刻返回结果，
所以它们的值在事务运行的整个期间内都不改变。
我们认为这是一个特性：目的是为了允许一个事务在“当前”时间上有一致的概念，
这样在同一个事务里的多个修改可以保持同样的时间戳。

> **注意**
>
> 其他数据库系统可能会更频繁地推进这些值。

PostgreSQL还提供了返回当前语句的开始时间以及调用该函数时的实际当前时间的函数。这些非SQL 标准的函数列表如下：

```
transaction_timestamp()
statement_timestamp()
clock_timestamp()
timeofday()
now()
```

`transaction_timestamp()`等价于`CURRENT_TIMESTAMP`，
但是其命名清楚地反映了它的返回值。`statement_timestamp()`
返回当前语句的开始时刻（更准确地说，是接收到客户端最近一条命令消息的时间）。
`statement_timestamp()`和`transaction_timestamp()`
在一个事务的第一条命令期间返回值相同，但是在随后的命令中却不一定相同。`clock_timestamp()`返回真正的当前时间，因此它的值甚至在同一条 SQL 命令中都会变化。
`timeofday()`是一个有历史原因的PostgreSQL函数。和`clock_timestamp()`相似，
它也返回真实的当前时间，但是它的结果是一个格式化的`text`串，而不是`timestamp with time zone`值。`now()`是PostgreSQL中与`transaction_timestamp()`等价的传统函数。

所有日期/时间数据类型也都接受特殊字面值`now`
来指定当前日期和时间（同样解释为事务开始时间）。因此，下面三种写法都返回相同的结果：

```
SELECT CURRENT_TIMESTAMP;
SELECT now();
SELECT TIMESTAMP 'now';  -- incorrect for use with DEFAULT
```

> **提示**
>
> 在创建表时指定`DEFAULT`子句的情况下，你不希望使用第三种形式。
> 系统将在分析这个常量的时候把`now`转换为一个`timestamp`，

> 这样需要默认值时就会得到创建表的时间！而前两种形式要到实际使用默认值的时候才被计算，因为它们是函数调用。因此它们可以给出每次插入行的时刻。

9.9.5. 延时执行

以下函数可用于延迟服务器进程的执行：

`pg_sleep(seconds)`

`pg_sleep`使当前会话的进程休眠*seconds*秒。*seconds*的类型为`double precision`，因此可以指定带小数部分的秒数作为延迟时间。例如：

```
SELECT pg_sleep(1.5);
```

> **注意**
>
> 有效的休眠时间间隔精度与平台有关，0.01 秒是常见值。休眠延迟将至少持续指定的时长，也有可能由于服务器负荷而比指定的时间长。特别地，
> 休眠并不保证能刚好在指定的时刻结束，但它不会在比指定时刻早的时候醒来。

> **警告**
>
> 请确保在调用`pg_sleep`或者其变体时，你的会话没有持有不必要的锁。
> 否则其它会话可能必须等待你的休眠会话，因而减慢整个系统速度。

9.10. 枚举支持函数

对于枚举类型（见第 8.7 节），有些函数可以避免硬编码枚举类型中的特定值，使程序更简洁。这些函数列在表 9.30中。以下示例假定枚举类型按如下方式创建：

```
CREATE TYPE rainbow AS ENUM ('red', 'orange', 'yellow', 'green',
 'blue', 'purple');
```

表 9.30. 枚举支持函数

函数	描述	示例	示例结果
`enum_first(anyenum)`	返回输入枚举类型的第一个值。	`enum_first(null::rainbow)`	`red`
`enum_last(anyenum)`	返回输入枚举类型的最后一个值。	`enum_last(null::rainbow)`	`purple`
`enum_range(anyenum)`	将输入枚举类型的所有值作为一个有序的数组返回。	`enum_range(null::rainbow)`	`{red,orange,yellow,green,blue,purple}`
`enum_range(anyenum, anyenum)`	以有序数组返回两个给定枚举值之间的范围。两个值必须来自同一枚举类型。如果第一个参数为null，结果从该枚举类型	`enum_range('orange'::rainbow, 'green'::rainbow)`	`{orange,yellow,green}`
		`enum_range(NULL, 'green'::rainbow)`	`{red,orange,yellow,green}`

函数	描述	示例	示例结果
	的第一个值开始；如果第二个参数为 null，结果以该枚举类型的最后一个值结束。	enum_range('orange'::rainbow, NULL)	{orange,yellow, green,blue,purple}

请注意，除了enum_range的双参数形式外，这些函数都忽略传入的具体值，只关心其声明的数据类型。传入 null 或该类型的某个具体值，结果都相同。通常会将这些函数用于表列或函数参数，而不是像示例那样使用硬编码的类型名。

9.11. 几何函数和操作符

几何类型point、box、lseg、line、path、polygon和circle有大量内置支持函数和操作符，如表 9.31、表 9.32和表 9.33中所示。

> **小心**
>
> 注意，“相同”操作符 ~= 表示 point、box、polygon 和 circle 类型通常意义上的相等。其中一些类型也有 = 操作符，但 = 仅比较面积是否相等。这些类型的其他标量比较操作符（<= 等）同样比较面积。

表 9.31. 几何操作符

操作符	描述	示例
+	平移	box '((0,0),(1,1))' + point '(2.0,0)'
-	平移	box '((0,0),(1,1))' - point '(2.0,0)'
*	缩放/旋转	box '((0,0),(1,1))' * point '(2.0,0)'
/	缩放/旋转	box '((0,0),(2,2))' / point '(2.0,0)'
#	相交的点或矩形框	box '((1,-1),(-1,1))' # box '((1,1),(-2,-2))'
#	路径或多边形中的点数	# path '((1,0),(0,1),(-1,0))'
@-@	长度或周长	@-@ path '((0,0),(1,0))'
@@	中心	@@ circle '((0,0),10)'
##	第二个操作数上距离第一个操作数最近的点	point '(0,0)' ## lseg '((2,0),(0,2))'
<->	两者之间的距离	circle '((0,0),1)' <-> circle '((5,0),1)'
&&	是否重叠？（有一个公共点即为真。）	box '((0,0),(1,1))' && box '((0,0),(2,2))'
<<	是否严格位于左侧？	circle '((0,0),1)' << circle '((5,0),1)'
>>	是否严格位于右侧？	circle '((5,0),1)' >> circle '((0,0),1)'

操作符	描述	示例
&<	是否未超出对方的右边界？	box '((0,0),(1,1))' &< box '((0,0),(2,2))'
&>	是否未超出对方的左边界？	box '((0,0),(3,3))' &> box '((0,0),(2,2))'
<<|	是否严格位于下方？	box '((0,0),(3,3))' <<| box '((3,4),(5,5))'
|>>	是否严格位于上方？	box '((3,4),(5,5))' |>> box '((0,0),(3,3))'
&<|	是否未超出对方的上边界？	box '((0,0),(1,1))' &<| box '((0,0),(2,2))'
|&>	是否未超出对方的下边界？	box '((0,0),(3,3))' |&> box '((0,0),(2,2))'
<^	是否位于下方（允许接触）？	circle '((0,0),1)' <^ circle '((0,5),1)'
>^	是否位于上方（允许接触）？	circle '((0,5),1)' >^ circle '((0,0),1)'
?#	是否相交？	lseg '((-1,0),(1,0))' ?# box '((-2,-2),(2,2))'
?-	是否水平？	?- lseg '((-1,0),(1,0))'
?-	是否水平对齐？	point '(1,0)' ?- point '(0,0)'
?|	是否竖直？	?| lseg '((-1,0),(1,0))'
?|	是否竖直对齐？	point '(0,1)' ?| point '(0,0)'
?-|	是否互相垂直？	lseg '((0,0),(0,1))' ?-| lseg '((0,0),(1,0))'
?||	是否平行？	lseg '((-1,0),(1,0))' ?|| lseg '((-1,2),(1,2))'
@>	是否包含？	circle '((0,0),2)' @> point '(1,1)'
<@	是否位于内部或边界上？	point '(1,1)' <@ circle '((0,0),2)'
~=	是否相同？	polygon '((0,0),(1,1))' ~= polygon '((1,1),(0,0))'

注意

在 PostgreSQL 8.2 之前，包含操作符 @> 和 <@ 分别称为 ~ 和 @。这些名称仍然可用，但已被弃用，最终将被移除。

表 9.32. 几何函数

函数	返回类型	描述	示例
area(*object*)	double precision	面积	area(box '((0,0),(1,1))')

函数	返回类型	描述	示例
center(*object*)	point	中心	center(box '((0,0),(1,2))')
diameter(circle)	double precision	圆的直径	diameter(circle '((0,0),2.0)')
height(box)	double precision	矩形框的竖直尺寸	height(box '((0,0),(1,1))')
isclosed(path)	boolean	是否为闭合路径?	isclosed(path '((0,0),(1,1),(2,0))')
isopen(path)	boolean	是否为开放路径?	isopen(path '[(0,0),(1,1),(2,0)]')
length(*object*)	double precision	长度	length(path '((-1,0),(1,0))')
npoints(path)	int	点数	npoints(path '[(0,0),(1,1),(2,0)]')
npoints(polygon)	int	点数	npoints(polygon '((1,1),(0,0))')
pclose(path)	path	将路径转换为闭合路径	pclose(path '[(0,0),(1,1),(2,0)]')
popen(path)	path	将路径转换为开放路径	popen(path '((0,0),(1,1),(2,0))')
radius(circle)	double precision	圆的半径	radius(circle '((0,0),2.0)')
width(box)	double precision	矩形框的水平尺寸	width(box '((0,0),(1,1))')

表 9.33. 几何类型转换函数

函数	返回类型	描述	示例
box(circle)	box	将圆转换为矩形框	box(circle '((0,0),2.0)')
box(point, point)	box	将点转换为矩形框	box(point '(0,0)', point '(1,1)')
box(polygon)	box	将多边形转换为矩形框	box(polygon '((0,0),(1,1),(2,0))')
circle(box)	circle	将矩形框转换为圆	circle(box '((0,0),(1,1))')
circle(point, double precision)	circle	由圆心和半径构造圆	circle(point '(0,0)', 2.0)
circle(polygon)	circle	将多边形转换为圆	circle(polygon '((0,0),(1,1),(2,0))')
lseg(box)	lseg	将矩形框的对角线转换为线段	lseg(box '((-1,0),(1,0))')
lseg(point, point)	lseg	由点构造线段	lseg(point '(-1,0)', point '(1,0)')
path(polygon)	path	将多边形转换为路径	path(polygon '((0,0),(1,1),(2,0))')

函数	返回类型	描述	示例
point(double precision, double precision)	point	构造点	point(23.4, -44.5)
point(box)	point	矩形框的中心	point(box '((-1,0),(1,0))')
point(circle)	point	圆的中心	point(circle '((0,0),2.0)')
point(lseg)	point	线段的中心	point(lseg '((-1,0),(1,0))')
point(polygon)	point	多边形的中心	point(polygon '((0,0),(1,1),(2,0))')
polygon(box)	polygon	将矩形框转换为 4 个顶点的多边形	polygon(box '((0,0),(1,1))')
polygon(circle)	polygon	将圆转换为 12 个顶点的多边形	polygon(circle '((0,0),2.0)')
polygon(*npts*, circle)	polygon	将圆转换为 *npts* 个顶点的多边形	polygon(12, circle '((0,0),2.0)')
polygon(path)	polygon	将路径转换为多边形	polygon(path '((0,0),(1,1),(2,0))')

可以把一个point当作下标为 0 和 1 的数组，访问它的两个数值分量。例如，如果t.p是一个point列，那么SELECT p[0] FROM t检索 X 坐标而 UPDATE t SET p[1] = ...改变 Y 坐标。同样，box或者lseg类型的值可以当作两个point值组成的数组看待。

area 函数适用于 box、circle 和 path 类型。对于 path 数据类型，只有当 path 中的点构成的路径不自相交时，area 函数才能工作。例如，path '((0,0),(0,1),(2,1),(2,2),(1,2),(1,0),(0,0))'::PATH 无法使用；但下面这个视觉上相同的 path '((0,0),(0,1),(1,1),(1,2),(2,2),(2,1),(1,1),(1,0),(0,0))'::PATH 可以使用。如果难以理解自相交与非自相交 path 的区别，可以把上述两条 path 并排画在方格纸上。

9.12. 网络地址函数和操作符

表 9.34列出了可用于 cidr 和 inet 类型的操作符。操作符 <<、<<=、>> 和 >>= 测试子网包含关系。它们只考虑两个地址的网络部分（忽略主机部分），并判断一个网络是否与另一个网络相同，或是另一个网络的子网。

表 9.34. cidr 和 inet 操作符

操作符	描述	示例
<	小于	inet '192.168.1.5' < inet '192.168.1.6'
<=	小于或等于	inet '192.168.1.5' <= inet '192.168.1.5'
=	等于	inet '192.168.1.5' = inet '192.168.1.5'
>=	大于或等于	inet '192.168.1.5' >= inet '192.168.1.5'

操作符	描述	示例
>	大于	inet '192.168.1.5' > inet '192.168.1.4'
<>	不等于	inet '192.168.1.5' <> inet '192.168.1.4'
<<	被包含	inet '192.168.1.5' << inet '192.168.1/24'
<<=	被包含在内或等于	inet '192.168.1/24' <<= inet '192.168.1/24'
>>	包含	inet '192.168.1/24' >> inet '192.168.1.5'
>>=	包含或等于	inet '192.168.1/24' >>= inet '192.168.1/24'
~	按位非	~ inet '192.168.1.6'
&	按位与	inet '192.168.1.6' & inet '0.0.0.255'
|	按位或	inet '192.168.1.6' | inet '0.0.0.255'
+	加法	inet '192.168.1.6' + 25
-	减法	inet '192.168.1.43' - 36
-	减法	inet '192.168.1.43' - inet '192.168.1.19'

表 9.35列出了可用于 cidr 和 inet 类型的函数。abbrev、host 和 text 函数主要用于提供其他显示格式。

表 9.35. cidr 和 inet 函数

函数	abbrev(inet)
返回类型	text
描述	以文本表示的缩略显示格式
示例	abbrev(inet '10.1.0.0/16')
结果	10.1.0.0/16
函数	abbrev(cidr)
返回类型	text
描述	以文本表示的缩略显示格式
示例	abbrev(cidr '10.1.0.0/16')
结果	10.1/16
函数	broadcast(inet)
返回类型	inet
描述	网络的广播地址
示例	broadcast('192.168.1.5/24')
结果	192.168.1.255/24

函数	family(inet)
返回类型	int
描述	提取地址族；IPv4 为 4，IPv6 为 6
示例	family('::1')
结果	6
函数	host(inet)
返回类型	text
描述	以文本形式提取 IP 地址
示例	host('192.168.1.5/24')
结果	192.168.1.5
函数	hostmask(inet)
返回类型	inet
描述	构造网络的主机掩码
示例	hostmask('192.168.23.20/30')
结果	0.0.0.3
函数	masklen(inet)
返回类型	int
描述	提取网络掩码长度
示例	masklen('192.168.1.5/24')
结果	24
函数	netmask(inet)
返回类型	inet
描述	构造网络的网络掩码
示例	netmask('192.168.1.5/24')
结果	255.255.255.0
函数	network(inet)
返回类型	cidr
描述	提取地址的网络部分
示例	network('192.168.1.5/24')
结果	192.168.1.0/24
函数	set_masklen(inet, int)
返回类型	inet
描述	设置 inet 值的网络掩码长度
示例	set_masklen('192.168.1.5/24', 16)
结果	192.168.1.5/16

函数	`set_masklen(cidr, int)`
返回类型	`cidr`
描述	设置 `cidr` 值的网络掩码长度
示例	`set_masklen('192.168.1.0/24'::cidr, 16)`
结果	`192.168.0.0/16`
函数	`text(inet)`
返回类型	`text`
描述	以文本形式提取 IP 地址和网络掩码长度
示例	`text(inet '192.168.1.5')`
结果	`192.168.1.5/32`

任何 `cidr` 值都可以隐式或显式地转换为 `inet`；因此，上述处理 `inet` 的函数也适用于 `cidr` 值。（如果 `inet` 和 `cidr` 各有单独的函数，是因为这两种情况下的行为应当不同。）也允许将 `inet` 值转换为 `cidr`。这样做时，网络掩码右侧的所有位都会被直接置零，以创建有效的 `cidr` 值。此外，还可以用普通的类型转换语法将文本值转换为 `inet` 或 `cidr`：例如 `inet(`*`expression`*`)` 或 *`colname`*`::cidr`。

表 9.36列出了可用于 `macaddr` 类型的函数。函数 `trunc(macaddr)` 返回将最后 3 个字节置零的 MAC 地址。可以用余下的前缀确定制造商。

表 9.36. macaddr 函数

函数	`trunc(macaddr)`
返回类型	`macaddr`
描述	将最后 3 个字节设为零
示例	`trunc(macaddr '12:34:56:78:90:ab')`
结果	`12:34:56:00:00:00`

`macaddr` 类型还支持按字典顺序比较的标准关系操作符（>、<= 等），。

9.13. 文本搜索函数和操作符

表 9.37、表 9.38以及表 9.39总结了为全文检索提供的函数和操作符。PostgreSQL的文本检索功能的详细解释可参考第 12 章。

表 9.37. 文本搜索操作符

操作符	描述	示例	结果
`@@`	`tsvector` 是否匹配 `tsquery`?	`to_tsvector('fat cats ate rats') @@ to_tsquery('cat & rat')`	`t`
`@@@`	`@@` 的已弃用同义写法	`to_tsvector('fat cats ate rats') @@@ to_tsquery( 'cat & rat')`	`t`

操作符	描述	示例	结果
`||`	串接 `tsvector`	`'a:1 b:2'::tsvector || 'c:1 d:2 b:3'::tsvector`	`'a':1 'b':2,5 'c':3 'd':4`
`&&`	将 `tsquery` 进行 AND 组合	`'fat | rat'::tsquery && 'cat'::tsquery`	`( 'fat' | 'rat' ) & 'cat'`
`||`	将 `tsquery` 进行 OR 组合	`'fat | rat'::tsquery || 'cat'::tsquery`	`( 'fat' | 'rat' ) | 'cat'`
`!!`	对 `tsquery` 取反	`!! 'cat'::tsquery`	`!'cat'`
`@>`	`tsquery` 是否包含另一个查询？	`'cat'::tsquery @> 'cat & rat'::tsquery`	`f`
`<@`	`tsquery` 是否被包含于另一个查询？	`'cat'::tsquery <@ 'cat & rat'::tsquery`	`t`

注意

`tsquery` 包含操作符只考虑两个查询中列出的词位，忽略组合操作符。

除了表中列出的操作符，`tsvector` 和 `tsquery` 类型还定义了普通的 B-树比较操作符（`=`、`<` 等）。这些操作符对文本检索用处不大，但可以用于其他用途，例如在这些类型的列上建立唯一索引。

表 9.38. 文本搜索函数

函数	`get_current_ts_config()`
返回类型	`regconfig`
描述	获取默认的文本搜索配置
示例	`get_current_ts_config()`
结果	`english`
函数	`length(tsvector)`
返回类型	`integer`
描述	`tsvector` 中的词位数
示例	`length('fat:2,4 cat:3 rat:5A'::tsvector)`
结果	`3`
函数	`numnode(tsquery)`
返回类型	`integer`
描述	`tsquery` 中的词位数与操作符数之和
示例	`numnode('(fat & rat) | cat'::tsquery)`
结果	`5`

函数	plainto_tsquery([*config* regconfig ,] *query* text)
返回类型	tsquery
描述	生成 tsquery，忽略标点符号
示例	plainto_tsquery('english', 'The Fat Rats')
结果	'fat' & 'rat'
函数	querytree(*query* tsquery)
返回类型	text
描述	获取 tsquery 中可索引的部分
示例	querytree('foo & ! bar'::tsquery)
结果	'foo'
函数	setweight(tsvector, "char")
返回类型	tsvector
描述	为tsvector中的每个元素赋予权重
示例	setweight('fat:2,4 cat:3 rat:5B'::tsvector, 'A')
结果	'cat':3A 'fat':2A,4A 'rat':5A
函数	strip(tsvector)
返回类型	tsvector
描述	从 tsvector 中移除位置和权重
示例	strip('fat:2,4 cat:3 rat:5A'::tsvector)
结果	'cat' 'fat' 'rat'
函数	to_tsquery([*config* regconfig ,] *query* text)
返回类型	tsquery
描述	正规化单词并转换为 tsquery
示例	to_tsquery('english', 'The & Fat & Rats')
结果	'fat' & 'rat'
函数	to_tsvector([*config* regconfig ,] *document* text)
返回类型	tsvector
描述	将文档文本转换为 tsvector
示例	to_tsvector('english', 'The Fat Rats')
结果	'fat':2 'rat':3
函数	ts_headline([*config* regconfig,] *document* text, *query* tsquery [, *options* text])
返回类型	text
描述	显示查询匹配内容
示例	ts_headline('x y z', 'z'::tsquery)

结果	x y z
函数	ts_rank([*weights* float4[],] *vector* tsvector, *query* tsquery [, *normalization* integer])
返回类型	float4
描述	计算文档相对于查询的排名分值
示例	ts_rank(textsearch, query)
结果	0.818
函数	ts_rank_cd([*weights* float4[],] *vector* tsvector, *query* tsquery [, *normalization* integer])
返回类型	float4
描述	使用覆盖密度计算文档相对于查询的排名分值
示例	ts_rank_cd('{0.1, 0.2, 0.4, 1.0}', textsearch, query)
结果	2.01317
函数	ts_rewrite(*query* tsquery, *target* tsquery, *substitute* tsquery)
返回类型	tsquery
描述	将查询中的target替换为substitute
示例	ts_rewrite('a & b'::tsquery, 'a'::tsquery, 'foo|bar'::tsquery)
结果	'b' & ('foo' | 'bar')
函数	ts_rewrite(*query* tsquery, *select* text)
返回类型	tsquery
描述	使用 SELECT 命令提供的目标和替代内容进行替换
示例	SELECT ts_rewrite('a & b'::tsquery, 'SELECT t,s FROM aliases')
结果	'b' & ('foo' | 'bar')
函数	tsvector_update_trigger()
返回类型	trigger
描述	自动更新 tsvector 列的触发器函数
示例	CREATE TRIGGER ... tsvector_update_trigger(tsvcol, 'pg_catalog.swedish', title, body)
结果	
函数	tsvector_update_trigger_column()
返回类型	trigger
描述	自动更新 tsvector 列的触发器函数
示例	CREATE TRIGGER ... tsvector_update_trigger_column(tsvcol, configcol, title, body)
结果	

> **注意**
>
> 所有接受一个可选的regconfig参数的文本检索函数在省略该参数时，会使用由default_text_search_config指定的配置。

表 9.39中的函数被单独列出，因为它们通常不被用于日常的文本检索操作。它们有助于开发和调试新的文本检索配置。

表 9.39. 文本搜索调试函数

函数	ts_debug([*config* regconfig,] *document* text, OUT *alias* text, OUT *description* text, OUT *token* text, OUT *dictionaries* regdictionary[], OUT *dictionary* regdictionary, OUT *lexemes* text[])
返回类型	setof record
描述	测试配置
示例	ts_debug('english', 'The Brightest supernovaes')
结果	(asciiword,"Word, all ASCII",The,{english_stem},english_stem,{}) ...
函数	ts_lexize(*dict* regdictionary, *token* text)
返回类型	text[]
描述	测试词典
示例	ts_lexize('english_stem', 'stars')
结果	{star}
函数	ts_parse(*parser_name* text, *document* text, OUT *tokid* integer, OUT *token* text)
返回类型	setof record
描述	测试解析器
示例	ts_parse('default', 'foo - bar')
结果	(1,foo) ...
函数	ts_parse(*parser_oid* oid, *document* text, OUT *tokid* integer, OUT *token* text)
返回类型	setof record
描述	测试解析器
示例	ts_parse(3722, 'foo - bar')
结果	(1,foo) ...
函数	ts_token_type(*parser_name* text, OUT *tokid* integer, OUT *alias* text, OUT *description* text)
返回类型	setof record
描述	获取解析器定义的词元类型
示例	ts_token_type('default')

结果	`(1,asciiword,"Word, all ASCII") ...`
函数	`ts_token_type(parser_oid oid, OUT tokid integer, OUT alias text, OUT description text)`
返回类型	`setof record`
描述	获取解析器定义的词元类型
示例	`ts_token_type(3722)`
结果	`(1,asciiword,"Word, all ASCII") ...`
函数	`ts_stat(sqlquery text, [ weights text, ] OUT word text, OUT ndoc integer, OUT nentry integer)`
返回类型	`setof record`
描述	获取 `tsvector` 列的统计信息
示例	`ts_stat('SELECT vector from apod')`
结果	`(foo,10,15) ...`

9.14. XML 函数

本节中描述的函数以及类函数表达式都作用于类型 `xml` 的值。关于 `xml` 类型的详细信息请查阅第 8.13 节。用于在值与类型 `xml` 之间转换的类函数表达式 `xmlparse` 和 `xmlserialize` 记录在该节中，这里不再重复。使用这些函数中的大多数要求安装时使用了`configure --with-libxml`编译。

9.14.1. 产生 XML 内容

有一组函数和类函数的表达式可以用来从 SQL 数据产生 XML 内容。它们特别适合于将查询结果格式化成 XML 文档以便于在客户端应用中处理。

9.14.1.1. xmlcomment

```
xmlcomment(text)
```

函数 `xmlcomment` 创建一个 XML 值，其中包含以指定文本为内容的 XML 注释。该文本不能包含“`--`”，也不能以“`-`”结尾，以确保构造出的 XML 注释有效。如果参数为空值，结果也为空值。

示例：

```
SELECT xmlcomment('hello');

  xmlcomment
--------------
 <!--hello-->
```

9.14.1.2. xmlconcat

```
xmlconcat(xml[, ...])
```

函数xmlconcat将由各个独立 XML 值组成的列表串接成一个单独的值，这个值包含一个 XML 内容片断。空值会被忽略，只有当没有参数为非空时结果才为空。

示例：

```
SELECT xmlconcat('<abc/>', '<bar>foo</bar>');

      xmlconcat
----------------------
 <abc/><bar>foo</bar>
```

如果 XML 声明存在，它们会按照下面的方式被组合。如果所有的参数值都有相同的 XML 版本声明，该版本将被用在结果中，否则将不使用版本。如果所有参数值有独立声明值“yes”，那么该值将被用在结果中。如果所有参数值都有一个独立声明值并且至少有一个为“no”，则该值被用在结果中。否则结果中将没有独立声明。
如果结果被决定要要求一个独立声明但是没有版本声明，将会使用一个版本 1.0 的版本声明，因为 XML 要求一个 XML 声明要包含一个版本声明。编码声明会被忽略并且在所有情况中都会被移除。

示例：

```
SELECT xmlconcat('<?xml version="1.1"?><foo/>', '<?xml version="1.1"
 standalone="no"?><bar/>');

             xmlconcat
-----------------------------------
 <?xml version="1.1"?><foo/><bar/>
```

9.14.1.3. xmlelement

```
xmlelement(name name [, xmlattributes(value [AS attname] [, ... ])] [,
 content, ...])
```

表达式 xmlelement 使用给定的名称、属性和内容生成一个 XML 元素。

示例：

```
SELECT xmlelement(name foo);

 xmlelement
------------
 <foo/>

SELECT xmlelement(name foo, xmlattributes('xyz' as bar));

    xmlelement
------------------
 <foo bar="xyz"/>

SELECT xmlelement(name foo, xmlattributes(current_date as bar),
 'cont', 'ent');

             xmlelement
-------------------------------------
```

```
 <foo bar="2007-01-26">content</foo>
```

不是合法 XML 名字的元素名和属性名将被转义，转义的方法是将不合法的字符用序列_x*HHHH*_替换，其中*HHHH*是被替换字符的 Unicode 代码点的十六进制表示。例如：

```
SELECT xmlelement(name "foo$bar", xmlattributes('xyz' as "a&b"));

            xmlelement
-----------------------------------
 <foo_x0024_bar a_x0026_b="xyz"/>
```

如果属性值是一个列引用，则不需要指定一个显式的属性名，
在这种情况下列的名字将被默认用于属性的名字。在其他情况下，
属性必须被给定一个显式名称。因此这个示例是合法的：

```
CREATE TABLE test (a xml, b xml);
SELECT xmlelement(name test, xmlattributes(a, b)) FROM test;
```

但是下面这些不合法：

```
SELECT xmlelement(name test, xmlattributes('constant'), a, b) FROM
 test;
SELECT xmlelement(name test, xmlattributes(func(a, b))) FROM test;
```

如果指定了元素内容，它们将被根据其数据类型格式化。如果内容本身也是类型`xml`，
就可以构建复杂的 XML 文档。例如：

```
SELECT xmlelement(name foo, xmlattributes('xyz' as bar),
                            xmlelement(name abc),
                            xmlcomment('test'),
                            xmlelement(name xyz));

                  xmlelement
----------------------------------------------
 <foo bar="xyz"><abc/><!--test--><xyz/></foo>
```

其他类型的内容将被格式化为合法的 XML 字符数据。这意味着字符 <、> 和 & 将被转换为实体。二进制数据（数据类型`bytea`）将被表示成 base64 或十六进制编码，具体取决于配置参数xmlbinary的设置。个别数据类型的特殊行为将不断发展，以使SQL和PostgreSQL数据类型与XML Schema 规范对齐，届时将给出更精确的描述。

9.14.1.4. xmlforest

```
xmlforest(content [AS name] [, ...])
```

表达式 `xmlforest` 使用给定的名称和内容生成由元素构成的 XML 森林（序列）。

示例：

```
SELECT xmlforest('abc' AS foo, 123 AS bar);

          xmlforest
------------------------------
 <foo>abc</foo><bar>123</bar>
```

```
SELECT xmlforest(table_name, column_name)
FROM information_schema.columns
WHERE table_schema = 'pg_catalog';

                                          xmlforest
-----------------------------------------------------------------------
---------------------
 <table_name>pg_authid</table_name><column_name>rolname</column_name>
 <table_name>pg_authid</table_name><column_name>rolsuper</column_name>
 ...
```

如第二个示例所示，如果内容值是一个列引用，可以省略元素名称，此时默认使用列名。否则，必须指定名称。

如上文`xmlelement`所示，非法 XML 名字的元素名会被转义。相似地，内容数据也会被转义来产生合法的 XML 内容，除非它已经是一个`xml`类型。

注意如果 XML 森林由多于一个元素组成，那么它不是合法的 XML 文档，因此在`xmlelement`中包装`xmlforest`表达式会有用处。

9.14.1.5. xmlpi

```
xmlpi(name target [, content])
```

表达式 `xmlpi` 创建一个 XML 处理指令。如果提供了内容，其中不得包含字符序列 `?>`。

示例：

```
SELECT xmlpi(name php, 'echo "hello world";');

            xmlpi
-----------------------------
 <?php echo "hello world";?>
```

9.14.1.6. xmlroot

```
xmlroot(xml, version text | no value [, standalone yes|no|no value])
```

表达式`xmlroot`修改一个 XML 值的根结点的属性。如果指定了一个版本，它会替换根节点的版本声明中的值；如果指定了一个独立设置，它会替换根节点的独立声明中的值。

```
SELECT xmlroot(xmlparse(document '<?xml version="1.1"?><content>abc</
content>'),
               version '1.0', standalone yes);

                xmlroot
----------------------------------------
 <?xml version="1.0" standalone="yes"?>
 <content>abc</content>
```

9.14.1.7. xmlagg

```
xmlagg(xml)
```

和这里描述的其他函数不同，函数xmlagg是一个聚合函数。
它将聚合函数调用的输入值串接起来，非常像xmlconcat所做的事情，
除了串接是跨行发生的而不是在单一行的多个表达式上发生。聚合表达式的更多信息请见第 9.18 节。

示例：

```
CREATE TABLE test (y int, x xml);
INSERT INTO test VALUES (1, '<foo>abc</foo>');
INSERT INTO test VALUES (2, '<bar/>');
SELECT xmlagg(x) FROM test;
        xmlagg
----------------------
 <foo>abc</foo><bar/>
```

为了决定串接的顺序，可以为聚合调用增加一个ORDER BY子句，如第 4.2.7 节中所述。例如：

```
SELECT xmlagg(x ORDER BY y DESC) FROM test;
        xmlagg
----------------------
 <bar/><foo>abc</foo>
```

我们推荐在以前的版本中使用下列非标准方法，并且它们在特定情况下仍然有用：

```
SELECT xmlagg(x) FROM (SELECT * FROM test ORDER BY y DESC) AS tab;
        xmlagg
----------------------
 <bar/><foo>abc</foo>
```

9.14.2. XML 谓词

这一节描述的表达式检查xml值的属性。

9.14.2.1. IS DOCUMENT

```
xml IS DOCUMENT
```

如果参数 XML 值是一个正确的 XML 文档，则IS DOCUMENT返回真，如果不是则返回假（即它是一个内容片断），或者是参数为空时返回空。文档和内容片断之间的区别请见第 8.13 节。

9.14.2.2. XMLEXISTS

```
XMLEXISTS(text PASSING [BY REF] xml [BY REF])
```

如果第一个参数中的XPath表达式返回任何节点，函数xmlexists返回true，否则返回false。（如果任一参数为空，则结果为空。）

示例：

```
SELECT xmlexists('//town[text() = ''Toronto'']' PASSING BY REF
 '<towns><town>Toronto</town><town>Ottawa</town></towns>');

 xmlexists
```

```
------------
 t
(1 row)
```

BY REF 子句在 PostgreSQL 中没有效果，但为了 SQL 一致性以及与其他实现的兼容性，允许使用它们。按照 SQL 标准，第一个 BY REF 是必需的，第二个是可选的。另请注意，SQL 标准规定 xmlexists 结构接受 XQuery 表达式作为第一个参数，但 PostgreSQL 目前只支持 XPath，它是 XQuery 的一个子集。

9.14.2.3. xml_is_well_formed

```
xml_is_well_formed(text)
xml_is_well_formed_document(text)
xml_is_well_formed_content(text)
```

这些函数检查一个text串是不是一个良构的 XML，返回一个布尔结果。xml_is_well_formed_document检查一个良构的文档，而xml_is_well_formed_content检查良构的内容。如果xmloption配置参数被设置为DOCUMENT，xml_is_well_formed会做第一个函数的工作；如果配置参数被设置为CONTENT，它会做第二个函数的工作。这意味着xml_is_well_formed对于检查一个到类型xml的简单类型转换是否会成功非常有用，而其他两个函数对于检查XMLPARSE的对应变体是否会成功有用。

示例：

```
SET xmloption TO DOCUMENT;
SELECT xml_is_well_formed('<>');
 xml_is_well_formed
--------------------
 f
(1 row)

SELECT xml_is_well_formed('<abc/>');
 xml_is_well_formed
--------------------
 t
(1 row)

SET xmloption TO CONTENT;
SELECT xml_is_well_formed('abc');
 xml_is_well_formed
--------------------
 t
(1 row)

SELECT xml_is_well_formed_document('<pg:foo xmlns:pg="http://
postgresql.org/stuff">bar</pg:foo>');
 xml_is_well_formed_document
-----------------------------
 t
(1 row)

SELECT xml_is_well_formed_document('<pg:foo xmlns:pg="http://
postgresql.org/stuff">bar</my:foo>');
 xml_is_well_formed_document
```

```
-----------------------------
 f
(1 row)
```

最后一个示例显示了检查是否正确匹配命名空间。

9.14.3. 处理 XML

为了处理`xml`数据类型的值，PostgreSQL 提供了`xpath`和`xpath_exists`函数，它们计算 XPath 1.0 表达式。

```
xpath(xpath, xml [, nsarray])
```

函数`xpath`针对 XML 值*`xml`*计算 XPath 表达式*`xpath`*（一个`text`值）。它返回一个由 XPath 表达式产生的节点集所对应的 XML 值组成的数组。

第二个参数必须是一个良构的 XML 文档。特殊地，它必须有一个单一根结点元素。

该函数可选的第三个参数是一个名字空间映射的数组。这个数组应该是一个二维`text`数组，其第二轴长度等于2（即它应该是一个数组的数组，其中每一个都由刚好 2 个元素组成）。每个数组项的第一个元素是名字空间的名称（别名），第二个元素是名字空间的 URI。并不要求在这个数组中提供的别名和在 XML 文档本身中使用的那些名字空间相同（换句话说，在 XML 文档中和在`xpath`函数环境中，别名都是本地的）。

示例：

```
SELECT xpath('/my:a/text()', '<my:a xmlns:my="http://example.
com">test</my:a>',
             ARRAY[ARRAY['my', 'http://example.com']]);

 xpath
--------
 {test}
(1 row)
```

要处理默认（匿名）命名空间，做这样的事情：

```
SELECT xpath('//mydefns:b/text()', '<a xmlns="http://example.
com"><b>test</b></a>',
             ARRAY[ARRAY['mydefns', 'http://example.com']]);

 xpath
--------
 {test}
(1 row)
```

```
xpath_exists(xpath, xml [, nsarray])
```

函数`xpath_exists`是`xpath`函数的一个特殊形式。它不返回满足 XPath 的各个 XML 值，而是返回一个布尔值，指示查询是否被满足。这个函数等价于标准的`XMLEXISTS`谓词，只是它还支持名称空间映射参数。

示例：

```
SELECT xpath_exists('/my:a/text()', '<my:a xmlns:my="http://example.
com">test</my:a>',
                    ARRAY[ARRAY['my', 'http://example.com']]);
 xpath_exists
--------------
 t
(1 row)
```

9.14.4. 将表映射到 XML

以下函数将关系表的内容映射为 XML 值，可以将它们视为 XML 导出功能：

```
table_to_xml(tbl regclass, nulls boolean, tableforest boolean,
 targetns text)
query_to_xml(query text, nulls boolean, tableforest boolean, targetns
 text)
cursor_to_xml(cursor refcursor, count int, nulls boolean,
              tableforest boolean, targetns text)
```

每个函数的返回类型都是`xml`。

`table_to_xml`映射由参数*`tbl`*传递的命名表的内容。`regclass`类型接受使用常见标记标识表的字符串，包括可选的模式限定和双引号。`query_to_xml`执行由参数*`query`*传递的查询并且映射结果集。`cursor_to_xml`从*`cursor`*指定的游标中取出指定数量的行。如果需要映射一个大型的表，我们推荐这种变体，因为每一个函数都是在内存中构建结果值的。

如果*`tableforest`*为假，则结果的 XML 文档看起来像这样：

```
<tablename>
  <row>
    <columnname1>data</columnname1>
    <columnname2>data</columnname2>
  </row>

  <row>
    ...
  </row>

  ...
</tablename>
```

如果*`tableforest`*为真，结果是一个看起来像这样的 XML 内容片断：

```
<tablename>
  <columnname1>data</columnname1>
  <columnname2>data</columnname2>
</tablename>

<tablename>
  ...
</tablename>

...
```

如果没有表名可用，在映射一个查询或一个游标时，在第一种格式中使用串`table`，在第二种格式中使用`row`。

这几种格式的选择由用户决定。第一种格式是一个正确的 XML 文档，它在很多应用中都很重要。如果结果值要被重组为一个文档，第二种格式在`cursor_to_xml`函数中更有用。前文讨论的产生 XML 内容的函数（特别是`xmlelement`）可以被用来把结果修改成符合用户的要求。

数据值会被以前文的函数`xmlelement`中描述的相同方法映射。

参数*`nulls`*决定空值是否会被包含在输出中。如果为真，列中的空值被表示为：

```
<columnname xsi:nil="true"/>
```

其中`xsi`是 XML 模式实例的 XML 名字空间前缀。一个合适的名字空间声明将被加入到结果值中。如果为假，包含空值的列将被从输出中忽略掉。

参数*`targetns`*指定想要的结果的 XML 名字空间。如果没有想要的特定名字空间，将会传递一个空串。

以下函数返回 XML Schema 文档，描述上面相应函数所执行的映射：

```
table_to_xmlschema(tbl regclass, nulls boolean, tableforest boolean,
 targetns text)
query_to_xmlschema(query text, nulls boolean, tableforest boolean,
 targetns text)
cursor_to_xmlschema(cursor refcursor, nulls boolean, tableforest
 boolean, targetns text)
```

必须传入相同的参数，才能得到彼此匹配的 XML 数据映射和 XML Schema 文档。

以下函数在同一个文档（或森林）中生成 XML 数据映射和对应的 XML Schema，并将两者链接在一起。当需要自包含且自描述的结果时，这些函数很有用：

```
table_to_xml_and_xmlschema(tbl regclass, nulls boolean, tableforest
 boolean, targetns text)
query_to_xml_and_xmlschema(query text, nulls boolean, tableforest
 boolean, targetns text)
```

此外，还可以使用以下函数，为整个模式或整个当前数据库生成类似的映射：

```
schema_to_xml(schema name, nulls boolean, tableforest boolean,
 targetns text)
schema_to_xmlschema(schema name, nulls boolean, tableforest boolean,
 targetns text)
schema_to_xml_and_xmlschema(schema name, nulls boolean, tableforest
 boolean, targetns text)

database_to_xml(nulls boolean, tableforest boolean, targetns text)
database_to_xmlschema(nulls boolean, tableforest boolean, targetns
 text)
database_to_xml_and_xmlschema(nulls boolean, tableforest boolean,
 targetns text)
```

注意，这些函数可能生成大量数据，需要在内存中构建。请求大型模式或数据库的内容映射时，可以考虑分别映射各个表，甚至通过游标来完成。

一个模式内容映射的结果看起来像这样：

```
<schemaname>

table1-mapping

table2-mapping

...

</schemaname>
```

其中一个表映射的格式取决于上文解释的*tableforest*参数。

一个数据库内容映射的结果看起来像这样：

```
<dbname>

<schema1name>
  ...
</schema1name>

<schema2name>
  ...
</schema2name>

...

</dbname>
```

其中的模式映射如上所述。

作为一个使用这些函数产生的输出的示例，图 9.1展示了一个 XSLT 样式表，它将`table_to_xml_and_xmlschema`的输出转换为一个包含表数据的扁平转印的 HTML 文档。以一种相似的方式，这些函数的结果可以被转换成其他基于 XML 的格式。

图 9.1. 转换 SQL/XML 输出到 HTML 的 XSLT 样式表

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0"
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema"
    xmlns="http://www.w3.org/1999/xhtml"
>

  <xsl:output method="xml"
      doctype-system="http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.
dtd"
      doctype-public="-//W3C/DTD XHTML 1.0 Strict//EN"
      indent="yes"/>

  <xsl:template match="/*">
    <xsl:variable name="schema" select="//xsd:schema"/>
    <xsl:variable name="tabletypename"
                  select="$schema/xsd:element[@name=name(current())]/
@type"/>
    <xsl:variable name="rowtypename"
                  select="$schema/xsd:complexType[@name=
$tabletypename]/xsd:sequence/xsd:element[@name='row']/@type"/>

    <html>
      <head>
        <title><xsl:value-of select="name(current())"/></title>
      </head>
      <body>
        <table>
          <tr>
            <xsl:for-each select="$schema/xsd:complexType[@name=
$rowtypename]/xsd:sequence/xsd:element/@name">
              <th><xsl:value-of select="."/></th>
            </xsl:for-each>
          </tr>

          <xsl:for-each select="row">
            <tr>
              <xsl:for-each select="*">
                <td><xsl:value-of select="."/></td>
              </xsl:for-each>
            </tr>
          </xsl:for-each>
        </table>
      </body>
    </html>
  </xsl:template>

</xsl:stylesheet>
```

9.15. 序列操作函数

本节描述用于操作序列对象（也称为序列生成器，或简称序列）的函数。序列对象是使用CREATE SEQUENCE创建的特殊单行表。序列对象通常用于为表中的行生成惟一标识符。在表 9.40中列出的序列函数，提供了简单的、多用户安全方法，用于从序列对象中获取连续的序列值。

表 9.40. 序列函数

函数	返回类型	描述
`currval(regclass)`	`bigint`	返回最近对指定序列调用 `nextval` 得到的值
`lastval()`	`bigint`	返回最近对任意序列调用 `nextval` 得到的值
`nextval(regclass)`	`bigint`	推进序列并返回新值
`setval(regclass, bigint)`	`bigint`	设置序列的当前值
`setval(regclass, bigint, boolean)`	`bigint`	设置序列的当前值和 `is_called` 标志

序列函数要操作的序列由一个`regclass`参数指定，该参数就是序列在`pg_class`系统目录中的OID。不过，不必手动查找 OID，因为`regclass`数据类型的输入转换函数会替你完成这项工作。只需将序列名用单引号括起来，使其看上去像一个字面常量。为兼容普通SQL名称的处理方式，除非序列名被双引号括起，否则字符串会转换为小写。因此：

`nextval('foo')` 操作序列 *foo*
`nextval('FOO')` 操作序列 *foo*
`nextval('"Foo"')` 操作序列 *Foo*

必要时，序列名可以带模式限定：

`nextval('myschema.foo')` 操作 *myschema.foo*
`nextval('"myschema".foo')` 同上
`nextval('foo')` 在搜索路径中查找 *foo*

参见第 8.16 节以获取更多有关以下类型的信息：`regclass`。

注意

在 PostgreSQL 8.1 之前，序列函数的参数类型是 `text`，而不是 `regclass`；上述从文本字符串到 OID 值的转换会在每次调用时于运行期间执行。为保持向后兼容，这种能力仍然存在，但内部现在会在调用函数前，通过从 `text` 到 `regclass` 的隐式强制转换来处理。

将序列函数的参数写成不加修饰的字符串字面量时，它会成为以下类型的常量：`regclass`。由于它实际上只是一个 OID，即使后来发生重命名、模式变更等情况，它仍会指向最初标识的序列。这种“早绑定”行为通常适用于列默认值和视图中的序列引用。但有时你可能希望采用“后绑定”，在运行时解析序列引用。要获得后绑定行为，应强制将常量存储为`text`常量，而不是`regclass`：

`nextval('foo'::text)` *foo*运行时被查找

请注意，后绑定是下列旧版本唯一支持的行为：PostgreSQL8.1 之前的版本，因此可能需要这样做来保留旧应用的语义。

当然，序列函数的参数既可以是常量，也可以是表达式。如果它是文本表达式，隐式强制转换就会导致运行时查找。

可用的序列函数如下：

`nextval`

将序列对象推进到下一个值并返回该值。这个操作是原子的：即使多个会话并发执行 `nextval`，每个会话也会安全地获得一个不同的序列值。

如果使用默认参数创建序列对象，连续调用 `nextval` 会返回从 1 开始的连续值。可以在CREATE SEQUENCE命令中使用特殊参数获得其他行为；更多信息请参见该命令的参考页。

> **重要**
>
> 为了避免阻塞从同一序列获取数值的并发事务，`nextval`操作永远不会回滚；也就是说，一旦一个值被取出，它就被视为已使用，即使执行`nextval`的事务随后中止也是如此。因此，中止的事务可能在序列的已分配值中留下未使用的“空洞”。

`currval`

返回当前会话中最近一次针对该序列调用 `nextval` 所获得的值。（如果当前会话从未针对该序列调用过 `nextval`，则会报错。）由于返回的是会话局部值，无论其他会话是否在当前会话调用之后执行过 `nextval`，结果都是可预测的。

`lastval`

返回当前会话中最近一次调用 `nextval` 所返回的值。此函数与 `currval` 相同，但不接受序列名参数，而是引用当前会话中最近一次调用 `nextval` 时所操作的序列。如果当前会话尚未调用过 `nextval`，调用 `lastval` 会报错。

`setval`

重置序列对象的计数器值。双参数形式将序列的`last_value`字段设为指定值，并将其`is_called`字段设为`true`，这意味着下次调用`nextval`会在返回值之前推进序列。`currval`报告的值也会设为指定值。在三参数形式中，`is_called`可以设为`true`或`false`。`true`与双参数形式具有相同的效果。如果将它设为`false`，下次调用`nextval`将恰好返回指定值，并从再下一次调用`nextval`时开始推进序列。此外，在这种情况下，`currval`报告的值不会改变。例如：

```
SELECT setval('foo', 42);           下一次 nextval 将返回 43
SELECT setval('foo', 42, true);     同上
SELECT setval('foo', 42, false);    下一次 nextval 将返回 42
```

函数`setval`返回的结果就是其第二个参数的值。

> **重要**
>
> 由于序列是非事务性的，如果事务回滚，`setval`所做的更改不会被撤销。

9.16. 条件表达式

本节描述在PostgreSQL中可用的SQL兼容的条件表达式。

> **提示**
>
> 如果你的需求超过这些条件表达式的能力，
> 你可能会希望用一种更富表现力的编程语言写一个存储过程。

9.16.1. CASE

SQL CASE表达式是一种通用的条件表达式，类似于其它编程语言中的 if/else 语句：

```
CASE WHEN condition THEN result
     [WHEN ...]
     [ELSE result]
END
```

CASE子句可以用于任何表达式可以出现的地方。每一个*condition*是一个返回boolean结果的表达式。如果结果为真，那么CASE表达式的结果就是紧随该条件后的*result*，并且剩下的CASE表达式不会被处理。如果条件的结果不为真，那么以相同方式搜寻任何随后的WHEN子句。如果没有WHEN *condition*为真，那么CASE表达式的值就是在ELSE子句里的*result*。如果省略了ELSE子句而且没有条件为真，结果为 null。

示例：

```
SELECT * FROM test;

 a
---
 1
 2
 3

SELECT a,
       CASE WHEN a=1 THEN 'one'
            WHEN a=2 THEN 'two'
            ELSE 'other'
       END
    FROM test;

 a | case
---+-------
 1 | one
 2 | two
 3 | other
```

所有*result*表达式的数据类型都必须可以转换成单一的输出类型。参阅第 10.5 节获取细节。

下面这个“简单”形式的CASE表达式是上述通用形式的一个变种：

```
CASE expression
    WHEN value THEN result
    [WHEN ...]
    [ELSE result]
END
```

第一个*expression*会被计算，然后与所有在WHEN子句中的每一个*value*对比，直到找到一个相等的。如果没有找到匹配的，则返回在ELSE子句中的*result*（或者 null 值）。这类似于 C 里的switch语句。

上面的示例可以用简单CASE语法来写：

```
SELECT a,
       CASE a WHEN 1 THEN 'one'
              WHEN 2 THEN 'two'
              ELSE 'other'
       END
    FROM test;

 a | case
---+-------
 1 | one
 2 | two
 3 | other
```

CASE表达式并不计算任何无助于判断结果的子表达式。例如，下面是一个可以避免被零除错误的方法：

```
SELECT ... WHERE CASE WHEN x <> 0 THEN y/x > 1.5 ELSE false END;
```

> **注意**
>
> 如第 4.2.14 节所述，在多种情况下，表达式中的子表达式会在不同阶段求值，因此“CASE只计算必要的子表达式”这一原则并非绝对成立。例如，常量子表达式1/0通常会在规划时导致除零错误，即便它位于一个运行时永远不会进入的CASE分支中也是如此。

9.16.2. COALESCE

```
COALESCE(value [, ...])
```

COALESCE函数返回其参数中第一个非空的值。只有当所有参数都为空时才返回空。它常用于在为显示而检索数据时为空值替换一个默认值，例如：

```
SELECT COALESCE(description, short_description, '(none)') ...
```

和CASE表达式一样，COALESCE只计算确定结果所需的参数；也就是说，在第一个不为 null 的参数右边的参数不会被计算。这个 SQL 标准函数提供了类似于NVL和IFNULL的能力，它们被用在某些其他数据库系统中。

9.16.3. NULLIF

```
NULLIF(value1, value2)
```

当*value1*和*value2*相等时，NULLIF返回 null。否则它返回*value1*。这些可以用于执行前文给出的COALESCE示例的逆操作：

```
SELECT NULLIF(value, '(none)') ...
```

在这个示例中，如果`value`是`(none)`，将返回 null，否则返回`value`的值。

9.16.4. GREATEST和LEAST

`GREATEST(value [, ...])`

`LEAST(value [, ...])`

`GREATEST`和`LEAST`函数从一个任意数量的表达式列表中选出最大或最小的值。
这些表达式必须都能转换为一个公共数据类型，该类型也就是结果的类型（细节见第 10.5 节）。
列表中的空值会被忽略。只有当所有表达式都计算为空时，结果才为空。

注意，`GREATEST`和`LEAST`不在 SQL 标准中，但它们是一种常见的扩展。有些其他数据库让它们在任一参数为空时就返回空，而不是只在全部参数都为空时才返回空。

9.17. 数组函数和操作符

表 9.41列出了可用于数组类型的操作符。

表 9.41. 数组操作符

操作符	描述	示例	结果
=	等于	ARRAY[1.1,2.1,3.1]::int[] = ARRAY[1,2,3]	t
<>	不等于	ARRAY[1,2,3] <> ARRAY[1,2,4]	t
<	小于	ARRAY[1,2,3] < ARRAY[1,2,4]	t
>	大于	ARRAY[1,4,3] > ARRAY[1,2,4]	t
<=	小于等于	ARRAY[1,2,3] <= ARRAY[1,2,3]	t
>=	大于等于	ARRAY[1,4,3] >= ARRAY[1,4,3]	t
@>	包含	ARRAY[1,4,3] @> ARRAY[3,1,3]	t
<@	被包含	ARRAY[2,2,7] <@ ARRAY[1,7,4,2,6]	t
&&	重叠（有公共元素）	ARRAY[1,4,3] && ARRAY[2,1]	t
||	数组与数组串接	ARRAY[1,2,3] || ARRAY[4,5,6]	{1,2,3,4,5,6}
||	数组与数组串接	ARRAY[1,2,3] || ARRAY[[4,5,6],[7,8,9]]	{{1,2,3},{4,5,6},{7,8,9}}
||	元素与数组串接	3 || ARRAY[4,5,6]	{3,4,5,6}
||	数组与元素串接	ARRAY[4,5,6] || 7	{4,5,6,7}

数组比较操作符逐个元素地比较数组内容，使用元素数据类型的默认 B-树比较函数。多维数组的元素按行优先顺序访问（最后一个下标变化最快）。如果两个数组的内容相同，

但维度信息不同，则由维度信息中的第一个差异决定排序顺序。（这与 PostgreSQL 8.2 之前的版本不同：旧版本会将内容相同的两个数组视为相等，即使它们的维数或下标范围不同。）

参阅第 8.14 节获取有关数组操作符行为的更多细节。

表 9.42展示了可以用于数组类型的函数。参阅第 8.14 节获取更多信息以及使用这些函数的示例。

表 9.42. 数组函数

函数	`array_append(anyarray, anyelement)`
返回类型	`anyarray`
描述	向数组末尾追加一个元素
示例	`array_append(ARRAY[1,2], 3)`
结果	`{1,2,3}`
函数	`array_cat(anyarray, anyarray)`
返回类型	`anyarray`
描述	串接两个数组
示例	`array_cat(ARRAY[1,2,3], ARRAY[4,5])`
结果	`{1,2,3,4,5}`
函数	`array_ndims(anyarray)`
返回类型	`int`
描述	返回数组的维数
示例	`array_ndims(ARRAY[[1,2,3], [4,5,6]])`
结果	`2`
函数	`array_dims(anyarray)`
返回类型	`text`
描述	返回数组维度的文本表示
示例	`array_dims(ARRAY[[1,2,3], [4,5,6]])`
结果	`[1:2][1:3]`
函数	`array_fill(anyelement, int[] [, int[]])`
返回类型	`anyarray`
描述	返回按给定值和维度初始化的数组，可选地使用不为 1 的下界
示例	`array_fill(7, ARRAY[3], ARRAY[2])`
结果	`[2:4]={7,7,7}`
函数	`array_length(anyarray, int)`
返回类型	`int`
描述	返回数组指定维度的长度
示例	`array_length(array[1,2,3], 1)`

结果	3
函数	array_lower(anyarray, int)
返回类型	int
描述	返回数组指定维度的下界
示例	array_lower('[0:2]={1,2,3}'::int[], 1)
结果	0
函数	array_prepend(anyelement, anyarray)
返回类型	anyarray
描述	向数组开头添加一个元素
示例	array_prepend(1, ARRAY[2,3])
结果	{1,2,3}
函数	array_to_string(anyarray, text [, text])
返回类型	text
描述	使用给定分隔符和可选的 null 替代字符串串接数组元素
示例	array_to_string(ARRAY[1, 2, 3, NULL, 5], ',', '*')
结果	1,2,3,*,5
函数	array_upper(anyarray, int)
返回类型	int
描述	返回数组指定维度的上界
示例	array_upper(ARRAY[1,8,3,7], 1)
结果	4
函数	string_to_array(text, text [, text])
返回类型	text[]
描述	使用给定分隔符和可选的 null 替代字符串将字符串拆分为数组元素
示例	string_to_array('xx~^~yy~^~zz', '~^~', 'yy')
结果	{xx,NULL,zz}
函数	unnest(anyarray)
返回类型	setof anyelement
描述	将数组展开为一组行
示例	unnest(ARRAY[1,2])
结果	1 2 (2 rows)

在 `string_to_array` 中，如果分隔符参数为 NULL，输入字符串中的每个字符都会成为结果数组中的独立元素。如果分隔符为空字符串，则将整个输入字符串作为单元素数组返回。否则，在分隔符字符串的每次出现处拆分输入字符串。

在 `string_to_array` 中，如果省略 null 替代字符串参数，或该参数为 NULL，则不会将任何输入子串替换为 NULL。在 `array_to_string` 中，如果省略 null 替代字符串参数，或该参数为 NULL，则直接跳过数组中的所有 null 元素，不在输出字符串中表示它们。

> **注意**
>
> `string_to_array` 的行为与 PostgreSQL 9.1 之前的版本有两点不同。首先，当输入字符串长度为零时，它返回空数组（零元素），而不是 NULL。其次，如果分隔符字符串为 NULL，该函数会将输入拆分为单独的字符，而不是像以前那样返回 NULL。

也可参见第 9.18 节了解用于数组的聚合函数`array_agg`。

9.18. 聚合函数

聚合函数根据一组输入值计算出单个结果。内置聚合函数列于表 9.43和表 9.44。聚合函数的特殊语法说明见第 4.2.7 节。更多入门信息请参见第 2.7 节。

表 9.43. 通用聚合函数

函数	参数类型	返回类型	描述
`array_agg(`*expression*`)`	任意非数组类型	参数类型的数组	将输入值（包括 null）串接为数组
`avg(`*expression*`)`	`smallint`、`int`、`bigint`、`real`、`double precision`、`numeric` 或 `interval`	整数类型参数返回 `numeric`，浮点参数返回 `double precision`，其他情况与参数数据类型相同	所有输入值的平均值（算术平均值）
`bit_and(`*expression*`)`	`smallint`、`int`、`bigint` 或 `bit`	与参数数据类型相同	所有非空输入值的按位与；没有非空输入时为 null
`bit_or(`*expression*`)`	`smallint`、`int`、`bigint` 或 `bit`	与参数数据类型相同	所有非空输入值的按位或；没有非空输入时为 null
`bool_and(`*expression*`)`	`bool`	`bool`	所有输入值都为真时返回真，否则返回假
`bool_or(`*expression*`)`	`bool`	`bool`	至少一个输入值为真时返回真，否则返回假
`count(*)`		`bigint`	输入行数
`count(`*expression*`)`	任意	`bigint`	*expression* 的值不为 null 的输入行数
`every(`*expression*`)`	`bool`	`bool`	等价于 `bool_and`
`max(`*expression*`)`	任意数值、字符串、日期/时间、网络或枚举类型，或这些类型的数组	与参数类型相同	所有输入值中 *expression* 的最大值
`min(`*expression*`)`	任意数值、字符串、日期/时间、网络或枚举类型，或这些类型的数组	与参数类型相同	所有输入值中 *expression* 的最小值

函数	参数类型	返回类型	描述
string_agg(*expression*, *delimiter*)	text, text	text	将输入值串接为字符串，以分隔符分隔
sum(*expression*)	smallint、int、bigint、real、double precision、numeric、interval 或 money	smallint 或 int 参数返回 bigint，bigint 参数返回 numeric，其他情况与参数数据类型相同	对所有输入值的 *expression* 求和
xmlagg(*expression*)	xml	xml	串接 XML 值（另见第 9.14.1.7 节）

需要注意，除了count之外，这些函数在没有选中任何行时都会返回空值。特别地，sum在没有输入行时返回空值，而不是预期中的零；array_agg在没有输入行时返回空值，而不是空数组。必要时，可以用coalesce函数把空值替换成零或空数组。

> **注意**
>
> 布尔聚合bool_and和bool_or对应于标准 SQL 聚合every和any或some。至于any和some，标准语法似乎存在歧义：
>
> ```
> SELECT b1 = ANY((SELECT b2 FROM t2 ...)) FROM t1 ...;
> ```
>
> 此处的ANY既可以被视为引入一个子查询，也可以在该子查询返回一行布尔值时被视为聚合函数。因此，不能将标准名称用于这些聚合。

> **注意**
>
> 习惯于其他 SQL 数据库管理系统的用户，可能会对count聚合用于整个表时的性能感到失望。如下查询：
>
> ```
> SELECT count(*) FROM sometable;
> ```
>
> 将由PostgreSQL通过顺序扫描整个表来执行。

聚合函数array_agg、string_agg和xmlagg，以及类似的用户定义聚合函数，其结果值会随输入值的顺序发生实质性变化。默认情况下，输入顺序未指定，但可以在聚合调用中写入ORDER BY子句来控制，如第 4.2.7 节所示。也可以用已排序的子查询提供输入值，这通常也能奏效。例如：

```
SELECT xmlagg(x) FROM (SELECT x FROM test ORDER BY y DESC) AS tab;
```

但这种语法在 SQL 标准中是不允许的，并且不可移植到其他数据库系统。

表 9.44列出了统计分析常用的聚合函数。（将它们单独列出只是为了避免让更常用的聚合函数列表过于拥挤。）说明中提到的 *N* 表示所有输入表达式均非空的输入行数。在所有情况下，如果计算没有意义，例如 *N* 为零，都会返回空值。

表 9.44. 用于统计的聚合函数

函数	参数类型	返回类型	描述
corr(*Y*, *X*)	double precision	double precision	相关系数
covar_pop(*Y*, *X*)	double precision	double precision	总体协方差
covar_samp(*Y*, *X*)	double precision	double precision	样本协方差
regr_avgx(*Y*, *X*)	double precision	double precision	自变量的平均值（sum(*X*)/*N*）
regr_avgy(*Y*, *X*)	double precision	double precision	因变量的平均值（sum(*Y*)/*N*）
regr_count(*Y*, *X*)	double precision	bigint	两个表达式都非空的输入行数
regr_intercept(*Y*, *X*)	double precision	double precision	由（*X*, *Y*）数值对确定的最小二乘拟合线性方程的 y 轴截距
regr_r2(*Y*, *X*)	double precision	double precision	相关系数的平方
regr_slope(*Y*, *X*)	double precision	double precision	由（*X*, *Y*）数值对确定的最小二乘拟合线性方程的斜率
regr_sxx(*Y*, *X*)	double precision	double precision	sum(*X*^2) - sum(*X*)^2/*N*（自变量的“平方和”）
regr_sxy(*Y*, *X*)	double precision	double precision	sum(*X**Y) - sum(*X*) * sum(*Y*)/*N*（自变量与因变量的“乘积和”）
regr_syy(*Y*, *X*)	double precision	double precision	sum(*Y*^2) - sum(*Y*)^2/*N*（因变量的“平方和”）
stddev(*expression*)	smallint、int、bigint、real、double precision 或 numeric	浮点参数返回 double precision，其他情况返回 numeric	stddev_samp 的历史别名
stddev_pop(*expression*)	smallint、int、bigint、real、double precision 或 numeric	浮点参数返回 double precision，其他情况返回 numeric	输入值的总体标准差
stddev_samp(*expression*)	smallint、int、bigint、real、double precision 或 numeric	浮点参数返回 double precision，其他情况返回 numeric	输入值的样本标准差
variance(*expression*)	smallint、int、bigint、real、double precision 或 numeric	浮点参数返回 double precision，其他情况返回 numeric	var_samp 的历史别名
var_pop(*expression*)	smallint、int、bigint、real、double precision 或 numeric	浮点参数返回 double precision，其他情况返回 numeric	输入值的总体方差（总体标准差的平方）
var_samp(*expression*)	smallint、int、bigint、real、	浮点参数返回 double precision，其他情况返回 numeric	输入值的样本方差（样本标准差的平方）

函数	参数类型	返回类型	描述
	double precision 或 numeric		

9.19. 窗口函数

窗口函数提供了对与当前查询行相关的一组行执行计算的能力。该特性的介绍请参见第 3.5 节。

内置的窗口函数列在表 9.45中。注意，这些函数必须使用窗口函数语法调用，也就是说，需要一个OVER子句。

除了这些函数外，
任何内置或用户定义的聚合函数都可以用作窗口函数（内置聚合函数列表见第 9.18 节）。聚合函数只有在调用后跟随 OVER 子句时才作为窗口函数；否则作为常规聚合。

表 9.45. 通用窗口函数

函数	返回类型	描述
row_number()	bigint	当前行在其分区内的编号，从 1 开始计数
rank()	bigint	当前行的排名，允许空缺；等于其第一个同等行的 row_number
dense_rank()	bigint	当前行的排名，没有空缺；此函数对同等行组计数
percent_rank()	double precision	当前行的相对排名：（rank - 1）/（总行数 - 1）
cume_dist()	double precision	当前行的相对排名：（当前行之前或与当前行同等的行数）/（总行数）
ntile(*num_buckets* integer)	integer	范围为 1 到参数值的整数，将分区尽可能均等地划分
lag(*value* anyelement [, *offset* integer [, *default* anyelement]])	与 *value* 相同的类型	返回在分区内当前行之前 *offset* 行处计算的 *value*；如果没有这样的行，则返回 *default*（其类型必须与 *value* 相同）。*offset* 和 *default* 都针对当前行求值。如果省略，*offset* 默认为 1，*default* 默认为 null
lead(*value* anyelement [, *offset* integer [, *default* anyelement]])	与 *value* 相同的类型	返回在分区内当前行之后 *offset* 行处计算的 *value*；如果没有这样的行，则返回 *default*（其类型必须与 *value* 相同）。*offset* 和 *default* 都针对当前行求值。如果省略，*offset* 默认为 1，*default* 默认为 null
first_value(*value* any)	与 *value* 相同的类型	返回在窗口帧第一行处计算的 *value*
last_value(*value* any)	与 *value* 相同的类型	返回在窗口帧最后一行处计算的 *value*
nth_value(*value* any, *nth* integer)	与 *value* 相同的类型	返回在窗口帧第 *nth* 行处计算的 *value*（从 1 开始计数）；没有这样的行时返回 null

在表 9.45中列出的所有函数都依赖于相关窗口定义的ORDER BY子句指定的排序顺序。在ORDER BY排序中不能区分的行被称为是同等行；这四个排名函数的定义使它们对任何两个同等行都给出相同的答案。

注意first_value、last_value和nth_value只考虑“窗口帧”内的行，它默认情况下包含从分区的开始行直到当前行的最后一个同等行。这对last_value可能不会给出有用的结果，有时对nth_value也一样。你可以通过向OVER子句增加一个合适的帧声明（RANGE或ROWS）来重定义帧。关于帧声明的更多信息请参考第 4.2.8 节。

当一个聚合函数被用作窗口函数时，它将在当前行的窗口帧内的行上聚合。一个使用ORDER BY和默认窗口帧定义的聚合产生一种“累计求和”类型的行为，这可能是或者不是想要的结果。为了获取在整个分区上的聚合，省略ORDER BY或者使用ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING。其它窗口帧声明可以用来获得其它的效果。

注意

SQL 标准为lead、lag、first_value、last_value和nth_value定义了一个RESPECT NULLS或IGNORE NULLS选项。这在PostgreSQL中没有实现：行为总是与标准的默认相同，即RESPECT NULLS。同样，标准中用于nth_value的FROM FIRST或FROM LAST选项没有实现：只支持默认的FROM FIRST行为（你可以通过反转ORDER BY的排序达到FROM LAST的结果）。

9.20. 子查询表达式

本节描述PostgreSQL中可用的SQL兼容的子查询表达式。本节介绍的所有表达式形式都返回布尔值（真/假）结果。

9.20.1. EXISTS

EXISTS (*subquery*)

EXISTS的参数是一个任意的SELECT语句，或者说子查询。系统对子查询进行运算以判断它是否返回行。如果它至少返回一行，那么EXISTS的结果就为“真”；如果子查询没有返回行，那么EXISTS的结果是“假”。

子查询可以引用外层查询的变量，这些变量在该子查询的任何一次计算中都起常量的作用。

这个子查询通常只是运行到能判断它是否可以返回至少一行为止，而不是等到全部结束。在这里写任何有副作用的子查询都是不明智的（例如调用序列函数）；这些副作用是否发生是很难判断的。

因为结果只取决于是否会返回行，而不取决于这些行的内容，所以这个子查询的输出列表通常是无关紧要的。一个常用的编码习惯是用EXISTS(SELECT 1 WHERE ...)的形式写所有的EXISTS测试。不过这条规则有例外，例如那些使用INTERSECT的子查询。

下面这个简单的示例类似在col2上的一次内连接，但是它为每个 tab1的行最多生成一行输出，即使存在多个匹配tab2的行也如此：

```
SELECT col1
FROM tab1
WHERE EXISTS (SELECT 1 FROM tab2 WHERE col2 = tab1.col2);
```

9.20.2. IN

expression IN (*subquery*)

右手边是一个圆括号括起来的子查询，它必须恰好返回一列。左手边表达式会被求值，并与子查询结果逐行比较。如果找到任意相等的子查询行，那么IN的结果就是“真”。如果没有找到相等行，那么结果就是“假”（包括子查询不返回任何行的情况）。

请注意，如果左侧表达式得到空值，或者右侧没有相等的值且至少有一行得到空值，则IN结构的结果将是空值，而不是假。这符合 SQL 对空值布尔组合的一般规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor IN (*subquery*)

这种形式的IN左手边是一个行构造器，如第 4.2.13 节中所述。右手边是一个圆括号括起来的子查询，它必须返回与左手边行中表达式数量完全相同的列数。左手边表达式会被求值，并与子查询结果逐行比较。如果找到任意相等的子查询行，则IN的结果为“真”。如果没有找到相等行，那么结果为“假”（包括子查询不返回任何行的情况）。

通常，表达式或者子查询行里的空值是按照 SQL 布尔表达式的一般规则进行组合的。如果两个行对应的成员都非空并且相等，那么认为这两行相等；如果任意对应成员为非空且不等，那么这两行不等；否则这样的行比较的结果是未知（空值）。如果所有行的结果要么是不等，要么是空值，并且至少有一个空值，那么IN的结果是空值。

9.20.3. NOT IN

expression NOT IN (*subquery*)

右手边是一个圆括号括起来的子查询，它必须恰好返回一列。左手边表达式会被求值，并与子查询结果逐行比较。如果只找到不相等的子查询行（包括子查询不返回任何行的情况），那么NOT IN的结果是“真”。如果找到任何相等行，则结果为“假”。

请注意，如果左侧表达式得到空值，或者右侧没有相等的值且至少有一行得到空值，则NOT IN结构的结果将是空值，而不是真。这符合 SQL 对空值布尔组合的一般规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor NOT IN (*subquery*)

这种形式的NOT IN左侧是一个行构造器，如第 4.2.13 节中所述。右侧是一个用圆括号括起来的子查询，必须返回与左侧行中表达式数量完全相同的列数。对左侧表达式求值后，将其按行与子查询结果的每一行比较。
如果只找到不相等的子查询行（包括子查询不返回任何行的情况），则NOT IN的结果为“真”。如果找到任何相等行，则结果为“假”。

通常，表达式或者子查询行里的空值是按照 SQL 布尔表达式的一般规则进行组合的。如果两个行对应的成员都非空并且相等，那么认为这两行相等；如果任意对应成员为非空且不等，那么这两行不等；否则这样的行比较的结果是未知（空值）。如果所有行的结果要么是不等，要么是空值，并且至少有一个空值，那么NOT IN的结果是空值。

9.20.4. ANY/SOME

expression operator ANY (*subquery*)
expression operator SOME (*subquery*)

这种形式的右侧是一个用括号括起来的子查询，它必须恰好返回一列。左侧表达式会被求值，并使用给定的 *operator* 与子查询结果的每一行进行比较。该操作符必须产生布尔结果。如果得到任何真值结果，那么ANY的结果就是“真”。如果没有找到真值结果，那么结果是“假”（包括子查询没有返回任何行的情况）。

SOME是ANY的同义词。IN等价于= ANY。

请注意，如果没有任何比较返回真，并且至少有一个右侧行的操作符结果为空值，则ANY结构的结果将是空值，而不是假。这符合 SQL 对空值布尔组合的一般规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor operator ANY (*subquery*)
row_constructor operator SOME (*subquery*)

这种形式的ANY左侧是一个行构造器，如第 4.2.13 节所述。右侧是一个用括号括起来的子查询，它必须返回与左侧行中表达式数量完全相同的列数。左侧表达式会被求值，并使用给定的 *operator*与子查询结果的每一行逐行比较。如果比较对任何子查询行返回真，则ANY的结果为“真”。如果比较对每一个子查询行都返回假，则结果为“假”（包括子查询不返回行的情况）。如果比较对任何行都不返回真，并且至少有一次比较返回NULL，则结果为 NULL。

关于按行比较的详细含义请见第 9.21.5 节。

9.20.5. ALL

expression operator ALL (*subquery*)

右侧是一个用圆括号括起来的子查询，必须恰好返回一列。左侧表达式会被求值，并使用给定的 *operator*与子查询结果的每一行进行比较。该操作符必须产生布尔结果。
如果所有行都得到真（包括子查询不返回任何行的情况），则ALL的结果为“真”。
如果得到任何假值结果，则结果为“假”。如果与任何子查询行的比较都不返回假，并且至少有一次比较返回 NULL，则结果为 NULL。

NOT IN等价于<> ALL。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor operator ALL (*subquery*)

这种形式的 ALL 左侧是一个行构造器，如第 4.2.13 节所述。右侧是一个用括号括起来的子查询，它必须返回与左侧行中表达式一样多的列。左侧表达式会被求值，并使用给定的 *operator* 与子查询结果逐行比较。如果该比较对所有子查询行都返回真，那么ALL的结果就是“真”（包括子查询没有返回任何行的情况）。如果对任何子查询行的比较返回假，则结果为“假”。
如果比较对任何子查询行都不返回假，并且至少有一次比较返回 NULL，则结果为 NULL。

关于按行比较的详细含义请见第 9.21.5 节。

9.20.6. 单一行比较

row_constructor operator (*subquery*)

左侧是一个行构造器，如第 4.2.13 节所述。右侧是一个用圆括号括起来的子查询，必须返回与左侧行中表达式数量完全相同的列数。此外，该子查询不能返回超过一行；如果它返回零行，则结果为空值。对左侧求值后，将所得的行与子查询返回的唯一一行进行比较。

关于按行比较的详细含义请见第 9.21.5 节。

9.21. 行和数组比较

本节描述几个特殊的结构，用于在值的组之间进行多重比较。
这些形式语法上和前面一节的子查询形式相关，但是不涉及子查询。涉及数组子表达式的形式是PostgreSQL的扩展；其余形式是SQL兼容的。
本节介绍的所有表达式形式都返回布尔（Boolean）结果（真/假）。

9.21.1. IN

```
expression IN (value [, ...])
```

右侧是一个用圆括号括起来的标量表达式列表。
如果左侧表达式的结果等于右侧任一表达式的结果，则结果为“真”。这等价于以下写法：

```
expression = value1
OR
expression = value2
OR
...
```

请注意如果左手边表达式得到空值，
或者没有相等的右手边值并且至少有一个右手边的表达式得到空值，那么
`IN`结构的结果将为空值，而不是假。这符合 SQL 处理空值的布尔组合的一般规则。

9.21.2. NOT IN

```
expression NOT IN (value [, ...])
```

右侧是一个用圆括号括起来的标量表达式列表。
如果左侧表达式的结果与右侧所有表达式的结果都不相等，则结果为“真”。这等价于以下写法：

```
expression <> value1
AND
expression <> value2
AND
...
```

请注意如果左手边表达式得到空值，
或者没有相等的右手边值并且至少有一个右手边的表达式得到空值，那么`NOT IN`结构的结果将为空值，而不是我们可能天真地认为的真值。这符合 SQL 处理空值的布尔组合的一般规则。

> **提示**
>
> `x NOT IN y`在所有情况下都等效于`NOT (x IN y)`。但是，在处理空值的时候，用`NOT IN`比用`IN`更可能迷惑新手。最好尽可能用正逻辑来表达你的条件。

9.21.3. ANY/SOME（数组）

```
expression operator ANY (array expression)
expression operator SOME (array expression)
```

右侧是一个用括号括起来的表达式，它必须产生一个数组值。左侧表达式会被求值，
并使用给定的*operator*与数组的每个元素进行比较，该操作符必须产生布尔结果。

如果得到了任何真值结果，那么ANY的结果是“真”。
如果没有找到真值结果（包括数组有零个元素的情况），那么结果是“假”。

如果数组表达式得到的是 null 数组，那么ANY的结果将为 null。如果左手边的表达式得到 null，ANY通常也为 null（尽管非严格比较操作符可能得到不同结果）。另外，如果右手边数组包含任何 null 元素，并且没有得到真值比较结果，那么ANY的结果将为 null 而不是假（同样假设这里使用的是严格比较操作符）。这符合 SQL 处理 null 值布尔组合的一般规则。

SOME是ANY的同义词。

9.21.4. ALL（数组）

expression operator ALL (*array expression*)

右侧是一个用括号括起来的表达式，它必须产生一个数组值。左侧表达式会被求值，
并使用给定的*operator*与数组的每个元素进行比较，该操作符必须产生布尔结果。
如果所有比较都得到真值结果，那么ALL的结果是“真”（包括数组有零个元素的情况）。
如果有任何假值结果，那么结果是“假”。

如果数组表达式得到的是 null 数组，那么ALL的结果将为 null。如果左手边的表达式得到 null，ALL通常也为 null（尽管非严格比较操作符可能得到不同结果）。另外，如果右手边数组包含任何 null 元素，并且没有得到假值比较结果，那么ALL的结果将为 null 而不是真（同样假设这里使用的是严格比较操作符）。这符合 SQL 处理 null 值布尔组合的一般规则。

9.21.5. 按行比较

row_constructor operator row_constructor

两侧都是行构造器，如第 4.2.13 节所述。两个行值必须具有相同数量的字段。对两侧分别求值后，按行进行比较。当 *operator* 为 =、<>、<、<=、>、>=，或具有与其中之一类似语义的操作符时，允许进行行比较。（具体而言，如果某个操作符是某个 B-树操作符类的成员，或者是某个 B-树操作符类中 = 成员的否定子，它就可以作为行比较操作符。）

=和<>情况略有不同。如果两行的所有对应成员都是非空且相等则这两行被认为相等；
如果任何对应成员是非空但是不相等则这两行不相等；否则行比较的结果为未知（空值）。

对于<、<=、>和>=这几种情况，会从左到右比较各行元素，一旦找到一对不相等或含有 null 的元素就立即停止。如果这对元素中的任意一个为 null，那么行比较的结果就是未知（null）；否则，这对元素的比较结果决定整个行比较的结果。例如，ROW(1,2,NULL) < ROW(1,3,0)的结果为真，而不是 null，因为第三对元素并不会被考虑。

> **注意**
>
> 在PostgreSQL 8.2 之前，<、<=、>和>=这几种情况并不是按照 SQL 规范处理的。像ROW(a,b) < ROW(c,d)这样的比较会被实现为a < c AND b < d，而正确行为应当等价于a < c OR (a = c AND b < d)。

row_constructor IS DISTINCT FROM *row_constructor*

这个结构与<>行比较相似，但是它对于空值输入不会得到空值。
任何空值被认为和任何非空值不相等（有区别），并且任意两个空值被认为相等（无区别）。
因此结果将总是为真或为假，永远不会是空值。

row_constructor IS NOT DISTINCT FROM *row_constructor*

这个结构与=行比较相似，但是它对于空值输入不会得到空值。
任何空值被认为和任何非空值不相等（有区别），并且任意两个空值被认为相等（无区别）。
因此结果将总是为真或为假，永远不会是空值。

> **注意**
>
> SQL 规范要求在结果依赖于比较两个 NULL 值或者一个 NULL 与一个非 NULL 时行比较返回 NULL。PostgreSQL只有在比较两个行构造器的结果或者比较一个行构造器与一个子查询的输出时才这样做（如第 9.20 节中所述）。在其他比较两个复合类型值的环境中，两个 NULL 字段值被认为相等，并且一个 NULL 被认为大于一个非 NULL。为了得到复合类型的一致的排序和索引行为，这样做是必要的。

9.22. 集合返回函数

本节描述那些可能返回多于一行的函数。目前这个类中仅有的函数是序列生成函数，如表 9.46和表 9.47所述。

表 9.46. 序列生成函数

函数	参数类型	返回类型	描述
`generate_series(`*`start`*`, `*`stop`*`)`	`int`或`bigint`	`setof int`或`setof bigint`（与参数类型相同）	以 1 为步长，从 *`start`* 到 *`stop`* 生成一系列值
`generate_series(`*`start`*`, `*`stop`*`, `*`step`*`)`	`int`或`bigint`	`setof int`或`setof bigint`（与参数类型相同）	以 *`step`* 为步长，从 *`start`* 到 *`stop`* 生成一系列值
`generate_series(`*`start`*`, `*`stop`*`, `*`step`*` interval)`	`timestamp` 或 `timestamp with time zone`	`setof timestamp` 或 `setof timestamp with time zone`（与参数类型相同）	以 *`step`* 为步长，从 *`start`* 到 *`stop`* 生成一系列值

当*`step`*为正数时，如果*`start`*大于*`stop`*，则返回零行。反之，当*`step`*为负数时，如果*`start`*小于*`stop`*则返回零行。如果输入为`NULL`，也返回零行。如果*`step`*为零，则会报错。
下面是一些示例：

```
SELECT * FROM generate_series(2,4);
 generate_series
-----------------
               2
               3
               4
(3 rows)

SELECT * FROM generate_series(5,1,-2);
 generate_series
-----------------
               5
               3
               1
(3 rows)

SELECT * FROM generate_series(4,3);
```

```
 generate_series
-----------------
(0 rows)

-- 此示例使用日期加整数的操作符
SELECT current_date + s.a AS dates FROM generate_series(0,14,7) AS s(
a);
   dates
------------
 2004-02-05
 2004-02-12
 2004-02-19
(3 rows)

SELECT * FROM generate_series('2008-03-01 00:00'::timestamp,
                              '2008-03-04 12:00', '10 hours');
   generate_series
---------------------
 2008-03-01 00:00:00
 2008-03-01 10:00:00
 2008-03-01 20:00:00
 2008-03-02 06:00:00
 2008-03-02 16:00:00
 2008-03-03 02:00:00
 2008-03-03 12:00:00
 2008-03-03 22:00:00
 2008-03-04 08:00:00
(9 rows)
```

表 9.47. 下标生成函数

函数	返回类型	描述
`generate_subscripts(array anyarray, dim int)`	`setof int`	生成由给定数组下标组成的序列。
`generate_subscripts(array anyarray, dim int, reverse boolean)`	`setof int`	生成由给定数组下标组成的序列。当 *reverse* 为真时，按逆序返回该序列。

`generate_subscripts` 是一个便利函数，它为给定数组的指定维度生成有效下标的集合。对于没有所请求维度的数组或空值数组，返回零行（但对空值数组元素会返回有效下标）。
下面是一些示例：

```
-- 基本用法
SELECT generate_subscripts('{NULL,1,NULL,2}'::int[], 1) AS s;
 s
---
 1
 2
 3
 4
(4 rows)

-- 展示一个数组、其下标以及下标对应的值需要用到子查询
```

```
SELECT * FROM arrays;
         a
--------------------
 {-1,-2}
 {100,200,300}
(2 rows)

SELECT a AS array, s AS subscript, a[s] AS value
FROM (SELECT generate_subscripts(a, 1) AS s, a FROM arrays) foo;
     array     | subscript | value
---------------+-----------+-------
 {-1,-2}       |         1 |    -1
 {-1,-2}       |         2 |    -2
 {100,200,300} |         1 |   100
 {100,200,300} |         2 |   200
 {100,200,300} |         3 |   300
(5 rows)

-- 展开一个二维数组
CREATE OR REPLACE FUNCTION unnest2(anyarray)
RETURNS SETOF anyelement AS $$
select $1[i][j]
   from generate_subscripts($1,1) g1(i),
        generate_subscripts($1,2) g2(j);
$$ LANGUAGE sql IMMUTABLE;
CREATE FUNCTION
postgres=# SELECT * FROM unnest2(ARRAY[[1,2],[3,4]]);
 unnest2
---------
       1
       2
       3
       4
(4 rows)
```

9.23. 系统信息函数

表 9.48列出了多个用于提取会话和系统信息的函数。

除了本节列出的函数外，还有许多与统计系统相关的函数，这些函数也提供系统信息。有关更多信息，请参见第 27.2.2 节。

表 9.48. 会话信息函数

名称	返回类型	描述
current_catalog	name	当前数据库名称（在 SQL 标准中称为“目录”）
current_database()	name	当前数据库名称
current_query()	text	客户端提交的当前执行查询文本（可能包含多条语句）
current_schema[()]	name	当前模式名称
current_schemas(boolean)	name[]	搜索路径中的模式名称，可选地包含隐式模式

名称	返回类型	描述
current_user	name	当前执行上下文的用户名
inet_client_addr()	inet	远端连接地址
inet_client_port()	int	远端连接端口
inet_server_addr()	inet	本地连接地址
inet_server_port()	int	本地连接端口
pg_backend_pid()	int	服务当前会话的服务器进程的进程ID
pg_conf_load_time()	timestamp with time zone	配置加载时间
pg_is_other_temp_schema(oid)	boolean	该模式是否为另一个会话的临时模式?
pg_listening_channels()	setof text	会话当前监听的通道名称
pg_my_temp_schema()	oid	会话临时模式的 OID；不存在时为0
pg_postmaster_start_time()	timestamp with time zone	服务器启动时间
session_user	name	会话用户名
user	name	等价于 current_user
version()	text	PostgreSQL版本信息

注意

current_catalog、current_schema、current_user、session_user和user在SQL中具有特殊语法：调用时不得在后面加圆括号。在 PostgreSQL 中，current_schema可以选择加圆括号，其他函数则不可以。

session_user 通常是发起当前数据库连接的用户；但超级用户可以用SET SESSION AUTHORIZATION更改这个设置。current_user 是适用于权限检查的用户标识符。通常它等于会话用户，但可以用SET ROLE更改。在执行带有SECURITY DEFINER属性的函数期间，它也会发生改变。用 Unix 术语来说，会话用户是“真实用户”，而当前用户是“有效用户”。

current_schema 返回搜索路径中第一个模式的名称（如果搜索路径为空，则返回空值）。在创建表或其他命名对象时，如果未指定目标模式，就会使用该模式。current_schemas(boolean) 返回当前搜索路径中所有模式名称的数组。布尔选项决定是否在返回的搜索路径中包含 pg_catalog 等隐式包含的系统模式。

注意

可以在运行时修改搜索路径，命令如下：

SET search_path TO *schema* [, *schema*, ...]

pg_listening_channels返回当前会话正在侦听的信道名称集合。更多信息见LISTEN。

inet_client_addr 返回当前客户端的 IP 地址，inet_client_port 返回端口号。inet_server_addr 返回服务器接受当前连接所用的 IP 地址，inet_server_port 返回端口号。如果当前连接通过 Unix 域套接字建立，这些函数都返回 NULL。

`pg_conf_load_time` 返回服务器配置文件最近一次加载的时间，类型为 `timestamp with time zone`。（如果当时当前会话已经存在，则返回该会话自身重新读取配置文件的时间，因此不同会话中的时间会略有不同。否则，返回 postmaster 进程重新读取配置文件的时间。）

`pg_my_temp_schema` 返回当前会话的临时模式的 OID；如果没有临时模式（因为尚未创建任何临时表），则返回零。如果给定 OID 是另一个会话的临时模式的 OID，`pg_is_other_temp_schema` 返回真。（这可以用于从系统目录显示结果中排除其他会话的临时表等场景。）

`pg_postmaster_start_time` 返回服务器启动的时间，类型为 `timestamp with time zone`。

`version`返回一个描述PostgreSQL服务器版本的字符串。

表 9.49列出了允许用户以编程方式查询对象访问权限的函数。关于权限的更多信息，请参见第 5.6 节。

表 9.49. 访问权限查询函数

名称	返回类型	描述
`has_any_column_privilege(`*`user`*`, `*`table`*`, `*`privilege`*`)`	`boolean`	用户是否对表的至少一列具有权限
`has_any_column_privilege(`*`table`*`, `*`privilege`*`)`	`boolean`	当前用户是否对表的至少一列具有权限
`has_column_privilege(`*`user`*`, `*`table`*`, `*`column`*`, `*`privilege`*`)`	`boolean`	用户是否具有列权限
`has_column_privilege(`*`table`*`, `*`column`*`, `*`privilege`*`)`	`boolean`	当前用户是否具有列权限
`has_database_privilege(`*`user`*`, `*`database`*`, `*`privilege`*`)`	`boolean`	用户是否具有数据库权限
`has_database_privilege(`*`database`*`, `*`privilege`*`)`	`boolean`	当前用户是否具有数据库权限
`has_foreign_data_wrapper_privilege(`*`user`*`, `*`fdw`*`, `*`privilege`*`)`	`boolean`	用户是否具有外部数据包装器权限
`has_foreign_data_wrapper_privilege(`*`fdw`*`, `*`privilege`*`)`	`boolean`	当前用户是否具有外部数据包装器权限
`has_function_privilege(`*`user`*`, `*`function`*`, `*`privilege`*`)`	`boolean`	用户是否具有函数权限
`has_function_privilege(`*`function`*`, `*`privilege`*`)`	`boolean`	当前用户是否具有函数权限
`has_language_privilege(`*`user`*`, `*`language`*`, `*`privilege`*`)`	`boolean`	用户是否具有语言权限
`has_language_privilege(`*`language`*`, `*`privilege`*`)`	`boolean`	当前用户是否具有语言权限

名称	返回类型	描述
has_schema_privilege(*user*, *schema*, *privilege*)	boolean	用户是否具有模式权限
has_schema_privilege(*schema*, *privilege*)	boolean	当前用户是否具有模式权限
has_sequence_privilege(*user*, *sequence*, *privilege*)	boolean	用户是否具有序列权限
has_sequence_privilege(*sequence*, *privilege*)	boolean	当前用户是否具有序列权限
has_server_privilege(*user*, *server*, *privilege*)	boolean	用户是否具有外部服务器权限
has_server_privilege(*server*, *privilege*)	boolean	当前用户是否具有外部服务器权限
has_table_privilege(*user*, *table*, *privilege*)	boolean	用户是否具有表权限
has_table_privilege(*table*, *privilege*)	boolean	当前用户是否具有表权限
has_tablespace_privilege(*user*, *tablespace*, *privilege*)	boolean	用户是否具有表空间权限
has_tablespace_privilege(*tablespace*, *privilege*)	boolean	当前用户是否具有表空间权限
pg_has_role(*user*, *role*, *privilege*)	boolean	用户是否具有角色权限
pg_has_role(*role*, *privilege*)	boolean	当前用户是否具有角色权限

has_table_privilege检查用户是否可以以某种方式访问表。用户参数可以是名称、OID（pg_authid.oid），public（表示 PUBLIC 伪角色）；如果省略该参数，则使用current_user。表可以通过名称或 OID 指定。（因此，has_table_privilege实际上有六种变体，可以通过参数的数量和类型来区分。）通过名称指定时，如有需要，可以用模式限定名称。所需访问权限类型由文本字符串指定，其值必须为以下值之一：SELECT，INSERT，UPDATE，DELETE，TRUNCATE，REFERENCES，或TRIGGER。还可以添加WITH GRANT OPTION到权限类型后，以检查该权限是否附带授予选项。也可以用逗号分隔列出多个权限类型；只要拥有列出的任一权限，结果就为true。（权限字符串不区分大小写，权限名之间允许有额外的空白，但权限名内部不允许。）下面是一些示例：

```
SELECT has_table_privilege('myschema.mytable', 'select');
SELECT has_table_privilege('joe', 'mytable', 'INSERT, SELECT WITH
 GRANT OPTION');
```

has_sequence_privilege 检查用户是否可以以某种方式访问序列。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 USAGE、SELECT 或 UPDATE 之一。

has_any_column_privilege 检查用户是否可以以某种方式访问表的至少一列。其参数形式与 has_table_privilege 类似，但所需访问权限类型必须为 SELECT、INSERT、UPDATE 或 REFERENCES 的某种组合。注意，在表级别拥有这些权限中的任意一种，就隐式地对表的每一列拥有该权限，因此对于相同参数，如果 has_table_privilege 返回 true，has_any_column_privilege 也总是返回真。但是，只要至少有一列获得该权限的列级授权，has_any_column_privilege 也会成功。

has_column_privilege 检查用户是否可以以某种方式访问列。其参数形式与 has_table_privilege 类似，但还可以通过列名或属性编号指定列。所需访问权限类型必须为 SELECT、INSERT、UPDATE 或 REFERENCES 的某种组合。注意，在表级别拥有这些权限中的任意一种，就隐式地对表的每一列拥有该权限。

has_database_privilege 检查用户是否可以以某种方式访问数据库。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE、CONNECT、TEMPORARY 或 TEMP（等同于 TEMPORARY）的某种组合。

has_function_privilege检查用户是否可以以某种方式访问函数。其参数形式类似于has_table_privilege。通过文本字符串而不是 OID 指定函数时，允许的输入与regprocedure数据类型相同（参见第 8.16 节）。所需访问权限类型必须为EXECUTE。例如：

```
SELECT has_function_privilege('joeuser', 'myfunc(int, text)',
 'execute');
```

has_foreign_data_wrapper_privilege 检查用户是否可以以某种方式访问外部数据包装器。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 USAGE。

has_language_privilege 检查用户是否可以以某种方式访问过程语言。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 USAGE。

has_schema_privilege 检查用户是否可以以某种方式访问模式。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE 或 USAGE 的某种组合。

has_server_privilege 检查用户是否可以以某种方式访问外部服务器。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 USAGE。

has_tablespace_privilege 检查用户是否可以以某种方式访问表空间。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE。

pg_has_role检查用户是否可以以某种方式访问角色。它的参数可能性与has_table_privilege类似，但不允许使用public作为用户名。所需访问权限类型必须为MEMBER或USAGE的某种组合。MEMBER表示直接或间接地属于该角色（即有权执行SET ROLE），而USAGE表示无需执行SET ROLE就能立即使用该角色的权限。

表 9.50列出了判断某个特定对象是否可见的函数，其判断依据是当前模式搜索路径。例如，如果表所在的模式位于搜索路径中，并且在搜索路径的更前面没有同名表，就称该表可见。这等价于说，可以只通过表名引用该表，而不必显式地用模式限定。要列出所有可见表的名称：

```
SELECT relname FROM pg_class WHERE pg_table_is_visible(oid);
```

表 9.50. 模式可见性查询函数

名称	返回类型	描述
pg_collation_is_visible(*collation_oid*)	boolean	排序规则是否在搜索路径中可见
pg_conversion_is_visible(*conversion_oid*)	boolean	转换是否在搜索路径中可见
pg_function_is_visible(*function_oid*)	boolean	函数是否在搜索路径中可见
pg_opclass_is_visible(*opclass_oid*)	boolean	操作符类是否在搜索路径中可见
pg_operator_is_visible(*operator_oid*)	boolean	操作符是否在搜索路径中可见

名称	返回类型	描述
pg_table_is_visible(*table_oid*)	boolean	表是否在搜索路径中可见
pg_ts_config_is_visible(*config_oid*)	boolean	全文检索配置是否在搜索路径中可见
pg_ts_dict_is_visible(*dict_oid*)	boolean	全文检索词典是否在搜索路径中可见
pg_ts_parser_is_visible(*parser_oid*)	boolean	全文检索解析器是否在搜索路径中可见
pg_ts_template_is_visible(*template_oid*)	boolean	全文检索模板是否在搜索路径中可见
pg_type_is_visible(*type_oid*)	boolean	类型（或域）是否在搜索路径中可见

每个函数检查一种数据库对象的可见性。注意，pg_table_is_visible 也可用于视图、物化视图、索引、序列和外部表；pg_type_is_visible 也可用于域。对于函数和操作符，如果搜索路径更前面没有名称和参数数据类型都相同的对象，那么路径中的该对象就是可见的。对于操作符类，会同时考虑名称和关联的索引访问方法。

所有这些函数都需要用对象 OID 标识要检查的对象。如果想按名称测试对象，使用 OID 别名类型会很方便（regclass，regtype，regprocedure，regoperator，regconfig，或 regdictionary），例如：

```
SELECT pg_type_is_visible('myschema.widget'::regtype);
```

注意，用这种方式测试不带模式限定的类型名并没有太大意义：只要该名称能够被识别，它就必然可见。

表 9.51列出从系统目录中提取信息的函数。

pg_describe_object返回由目录 OID、对象 OID 和子对象 ID（可为零）指定的数据库对象的文本描述。该描述供人阅读，并且可能会根据服务器配置被翻译。这有助于确定存储在pg_depend目录中的对象标识。

表 9.51. 系统目录信息函数

名称	返回类型	描述
format_type(*type_oid*, *typemod*)	text	获取数据类型的 SQL 名称
pg_describe_object(*catalog_id*, *object_id*, *object_sub_id*)	text	获取数据库对象的描述
pg_get_constraintdef(*constraint_oid*)	text	获取约束定义
pg_get_constraintdef(*constraint_oid*, *pretty_bool*)	text	获取约束定义
pg_get_expr(*pg_node_tree*, *relation_oid*)	text	反编译表达式的内部形式，假定其中所有 Var 节点都引用第二个参数指定的关系

名称	返回类型	描述
pg_get_expr(*pg_node_tree*, *relation_oid*, *pretty_bool*)	text	反编译表达式的内部形式，假定其中所有 Var 节点都引用第二个参数指定的关系
pg_get_functiondef(*func_oid*)	text	获取函数定义
pg_get_function_arguments(*func_oid*)	text	获取函数定义的参数列表（包含默认值）
pg_get_function_identity_arguments(*func_oid*)	text	获取用于标识函数的参数列表（不含默认值）
pg_get_function_result(*func_oid*)	text	获取函数的 RETURNS 子句
pg_get_indexdef(*index_oid*)	text	获取索引的 CREATE INDEX 命令
pg_get_indexdef(*index_oid*, *column_no*, *pretty_bool*)	text	获取索引的 CREATE INDEX 命令；当 *column_no* 非零时，仅获取一个索引列的定义
pg_get_keywords()	setof record	获取 SQL 关键字及其类别的列表
pg_get_ruledef(*rule_oid*)	text	获取规则的 CREATE RULE 命令
pg_get_ruledef(*rule_oid*, *pretty_bool*)	text	获取规则的 CREATE RULE 命令
pg_get_serial_sequence(*table_name*, *column_name*)	text	获取 serial 或 bigserial 列使用的序列名称
pg_get_triggerdef(*trigger_oid*)	text	获取触发器的 CREATE [CONSTRAINT] TRIGGER 命令
pg_get_triggerdef(*trigger_oid*,*pretty_bool*)	text	获取触发器的 CREATE [CONSTRAINT] TRIGGER 命令
pg_get_userbyid(*role_oid*)	name	获取具有给定 OID 的角色名称
pg_get_viewdef(*view_name*)	text	获取视图或物化视图的底层 SELECT 命令（已弃用）
pg_get_viewdef(*view_name*, *pretty_bool*)	text	获取视图或物化视图的底层 SELECT 命令（已弃用）
pg_get_viewdef(*view_oid*)	text	获取视图或物化视图的底层 SELECT 命令
pg_get_viewdef(*view_oid*, *pretty_bool*)	text	获取视图或物化视图的底层 SELECT 命令
pg_options_to_table(*reloptions*)	setof record	获取存储选项的名称/值对集合
pg_tablespace_databases(*tablespace_oid*)	setof oid	获取在该表空间中具有对象的数据库 OID 集合
pg_typeof(*any*)	regtype	获取任意值的数据类型

format_type 返回数据类型的 SQL 名称，该类型由其类型 OID 和可能存在的类型修饰符标识。如果不知道具体的类型修饰符，请为类型修饰符参数传入 NULL。

pg_get_keywords 返回一组记录，描述服务器识别的 SQL 关键字。word 列包含关键字。catcode 列包含类别代码：U 表示非保留关键字，C 表示列名，T 表示类型名或函数名，R 表示保留关键字。catdesc 列包含描述该类别的字符串，该字符串可能已经过本地化。

pg_get_constraintdef、pg_get_indexdef、pg_get_ruledef和pg_get_triggerdef分别重建约束、索引、规则或触发器的创建命令。（注意，这是通过反编译重建的，并非命令的原始文本。）pg_get_expr 反编译单个表达式的内部形式，例如列的默认值。这在检查系统目录内容时可能很有用。如果表达式可能包含 Var 节点，请将它们所引用的关系的 OID 指定为第二个参数；如果预计不含 Var 节点，传入零即可。pg_get_viewdef 重建定义视图的 SELECT 查询。这些函数中的大多数有两种变体，其中一种可以选择“美化输出”结果。美化格式更易读，但默认格式更可能被未来版本的 PostgreSQL 以相同方式解释；转储时应避免使用美化输出。向美化输出参数传入 false，所得结果与完全没有该参数的变体相同。

pg_get_functiondef 返回一个函数的完整 CREATE OR REPLACE FUNCTION 语句。pg_get_function_arguments 以 CREATE FUNCTION 中所需的形式返回函数参数列表。pg_get_function_result 类似地返回该函数相应的 RETURNS 子句。pg_get_function_identity_arguments 返回标识函数所需的参数列表，例如以 ALTER FUNCTION 中所需的形式返回。此形式省略默认值。

pg_get_serial_sequence返回与某一列关联的序列名称；如果该列没有关联的序列，则返回NULL。第一个输入参数是可以带有模式名的表名，第二个参数是列名。由于第一个参数可能包含模式和表，它不会被当作双引号括起的标识符处理，因此默认转换为小写；第二个参数只包含列名，会被当作带双引号的标识符处理并保留大小写。函数返回的值采用适于传递给序列函数的格式（参见第 9.15 节）。这一关联可以使用ALTER SEQUENCE OWNED BY修改或移除。（该函数也许应该叫作pg_get_owned_sequence；它当前的名称反映了它通常用于serial或bigserial列这一事实。）

pg_get_userbyid 根据给定 OID 获取角色名称。

向 pg_options_to_table 传入 pg_class.reloptions 或 pg_attribute.attoptions 时，它返回存储选项名称/值对（option_name/option_value）的集合。

pg_tablespace_databases 用于检查表空间。它返回在该表空间中存储了对象的数据库 OID 集合。如果此函数返回任何行，则说明该表空间不为空，不能删除。要显示存放在该表空间中的具体对象，需要连接到 pg_tablespace_databases 标识的数据库，并查询它们的 pg_class 系统目录。

pg_typeof返回所传入值的数据类型的 OID。这有助于排查问题或动态构造 SQL 查询。函数声明的返回类型是regtype，它是一种 OID 别名类型（参见第 8.16 节）；这意味着它在比较时与 OID 相同，但显示为类型名。例如：

```
SELECT pg_typeof(33);

 pg_typeof
-----------
 integer
(1 row)

SELECT typlen FROM pg_type WHERE oid = pg_typeof(33);
 typlen
--------
      4
(1 row)
```

表 9.52中的函数用于提取此前通过COMMENT命令存储的注释。如果找不到与指定参数对应的注释，则返回空值。

表 9.52. 注释信息函数

名称	返回类型	描述
col_description(*table_oid*, *column_number*)	text	获取表列的注释
obj_description(*object_oid*, *catalog_name*)	text	获取数据库对象的注释
obj_description(*object_oid*)	text	获取数据库对象的注释（已弃用）
shobj_description(*object_oid*, *catalog_name*)	text	获取共享数据库对象的注释

col_description 返回表列的注释，列由所属表的 OID 和列号指定。（obj_description 不能用于表列，因为列没有自身的 OID。）

obj_description 的双参数形式返回数据库对象的注释，对象由其 OID 和所在系统目录的名称指定。例如，obj_description(123456,'pg_class') 会获取 OID 为 123456 的表的注释。obj_description 的单参数形式只需要对象 OID。该形式已被弃用，因为无法保证 OID 在不同系统目录之间唯一，因而可能返回错误的注释。

shobj_description 的用法与 obj_description 相同，但用于获取共享对象的注释。有些系统目录由数据库集簇中的所有数据库全局共享，其中对象的描述也全局存储。

表 9.53中的函数以可导出的形式提供服务器事务信息。这些函数主要用于确定两个快照之间有哪些事务提交。

表 9.53. 事务 ID 和快照

名称	返回类型	描述
txid_current()	bigint	获取当前事务 ID；如果当前事务尚无 ID，则分配一个新的 ID
txid_current_snapshot()	txid_snapshot	获取当前快照
txid_snapshot_xip(*txid_snapshot*)	setof bigint	获取快照中正在进行的事务 ID
txid_snapshot_xmax(*txid_snapshot*)	bigint	获取快照的 xmax
txid_snapshot_xmin(*txid_snapshot*)	bigint	获取快照的 xmin
txid_visible_in_snapshot(*bigint*, *txid_snapshot*)	boolean	该事务 ID 是否在快照中可见？（不要用于子事务 ID）

内部事务 ID 类型（xid）为 32 位，每经过约 40 亿个事务就会回卷一次。不过，这些函数导出的是用“纪元”计数器扩展的 64 位格式，在一次安装的整个使用期间都不会回卷。这些函数使用的 txid_snapshot 数据类型存储某一时刻的事务 ID 可见性信息。其组成部分见表 9.54。

表 9.54. 快照组件

名称	描述
xmin	仍然活动的最早事务 ID（txid）。所有更早的事务要么已经提交且可见，要么已经回滚而失效。

名称	描述
xmax	第一个尚未分配的 txid。所有大于或等于此值的 txid 在快照时尚未开始，因此不可见。
xip_list	快照时活动的 txid。该列表仅包含 xmin 与 xmax 之间的活动 txid；可能存在大于 xmax 的活动 txid。满足 xmin <= txid < xmax 且不在此列表中的 txid 在快照时已经完成，因此根据其提交状态，要么可见，要么失效。该列表不包含子事务的 txid。

txid_snapshot 的文本表示为 *xmin*:*xmax*:*xip_list*。例如，10:20:10,14,15 表示 xmin=10, xmax=20, xip_list=10, 14, 15。

9.24. 系统管理函数

表 9.55展示了那些可以用于查询以及修改运行时配置参数的函数。

表 9.55. 配置设定函数

名称	返回类型	描述
current_setting(*setting_name*)	text	获取配置项的当前值
set_config(*setting_name*, *new_value*, *is_local*)	text	设置参数并返回新值

函数current_setting返回设置*setting_name*的当前值。它对应于SQL命令SHOW。例如：

```
SELECT current_setting('datestyle');

 current_setting
-----------------
 ISO, MDY
(1 row)
```

set_config将参数*setting_name*设置为*new_value*。如果*is_local*为true，新值仅适用于当前事务。如果要让新值适用于当前会话，请改用false。该函数对应于 SQL 命令SET。例如：

```
SELECT set_config('log_statement_stats', 'off', false);

 set_config
------------
 off
(1 row)
```

在表 9.56中展示的函数向其它服务器进程发送控制信号。这些函数的使用仅限于超级用户。

表 9.56. 服务器信号函数

名称	返回类型	描述
pg_cancel_backend(*pid* int)	boolean	取消后端的当前查询

名称	返回类型	描述
pg_reload_conf()	boolean	使服务器进程重新加载配置文件
pg_rotate_logfile()	boolean	轮转服务器日志文件
pg_terminate_backend(*pid* int)	boolean	终止后端

这些函数在成功时都返回 true，否则返回 false。

pg_cancel_backend和pg_terminate_backend向由进程 ID 标识的后端进程发送信号（分别是SIGINT或SIGTERM）。一个活动后端的进程 ID可以从pg_stat_activity视图的procpid列中找到，或者通过在服务器上列出postgres进程（在 Unix 上使用ps或者在Windows上使用任务管理器）得到。

pg_reload_conf 向服务器发送 SIGHUP 信号，使所有服务器进程重新加载配置文件。

pg_rotate_logfile 向日志文件管理器发送信号，使其立即切换到新的输出文件。只有内置日志收集器正在运行时才有效，否则不存在日志文件管理器子进程。

表 9.57中列出的函数有助于进行在线备份。这些函数不能在恢复期间执行。

表 9.57. 备份控制函数

名称	返回类型	描述
pg_create_restore_point(*name* text)	text	创建用于恢复的命名点（默认仅限超级用户，但可以向其他用户授予EXECUTE权限以运行此函数）
pg_current_xlog_insert_location()	text	获取当前预写式日志插入位置
pg_current_xlog_location()	text	获取当前预写式日志写入位置
pg_start_backup(*label* text [, *fast* boolean])	text	准备执行在线备份（默认仅限超级用户，但可以向其他用户授予EXECUTE权限以运行此函数）
pg_stop_backup()	text	结束排他在线备份（默认仅限超级用户，但可以向其他用户授予EXECUTE权限以运行此函数）
pg_switch_xlog()	text	强制切换到新的预写式日志文件（默认仅限超级用户，但可以向其他用户授予EXECUTE权限以运行此函数）
pg_xlogfile_name(*location* text)	text	将事务日志位置字符串转换为文件名
pg_xlogfile_name_offset(*location* text)	text, integer	将事务日志位置字符串转换为文件名和文件内的十进制字节偏移量

pg_start_backup接受任意用户定义的备份标签。（通常是备份转储文件将要保存的名称。）该函数将备份标签文件（backup_label）写入数据库集簇的数据目录，执行检查点，然后以文本形式返回备份的起始事务日志位置。用户可以忽略该返回值，提供它只是因为它可能有用。

```
postgres=# select pg_start_backup('label_goes_here');
```

```
 pg_start_backup
-----------------
 0/D4445B8
(1 row)
```

还可以提供第二个可选参数，其类型为boolean。如果该参数为true，则尽快执行pg_start_backup。这会强制立即执行检查点，使 I/O 操作量陡增，并降低并发执行的查询的速度。

pg_stop_backup 删除 pg_start_backup 创建的标签文件，并在事务日志归档区域创建一个备份历史文件。历史文件包括传给 pg_start_backup 的标签、备份的起止事务日志位置以及备份的起止时间。返回值是备份的结束事务日志位置（同样可以忽略）。记录结束位置后，
当前事务日志插入点会自动推进到下一个事务日志文件的起始处，
使包含结束位置的事务日志文件可以立即归档，从而完成备份。

pg_switch_xlog 切换到下一个预写式日志文件，使当前文件可以归档（假设正在使用连续归档）。返回值是刚完成的预写式日志文件中的结束预写式日志位置加 1。如果自上次切换预写式日志以来没有发生任何预写式日志活动，pg_switch_xlog 不执行任何操作，并返回当前使用的预写式日志文件的起始位置。

pg_create_restore_point 创建一条可用作恢复目标的命名预写式日志记录，并返回相应的预写式日志位置。随后可以在recovery_target_name中使用该名称，指定恢复到哪一点。
应避免创建多个同名恢复点，因为恢复会在第一个名称匹配恢复目标的恢复点停止。

pg_current_xlog_location 显示当前预写式日志写入位置，格式与上述函数相同。类似地，pg_current_xlog_insert_location 显示当前预写式日志插入点。插入位置是预写式日志在任意时刻的“逻辑”末尾，而写入位置是实际从服务器内部缓冲区写出的内容的末尾。
写入位置是能从服务器外部检查到的内容的末尾，如果要归档尚未写满的预写式日志文件，
通常需要的是这个位置。提供插入位置主要是为了服务器调试。这些都是只读操作，
不需要超级用户权限。

可以使用pg_xlogfile_name_offset
从上述任一函数的结果中提取相应的预写式日志文件名和字节偏移。例如：

```
postgres=# SELECT * FROM pg_xlogfile_name_offset(pg_stop_backup());
        file_name         | file_offset
--------------------------+-------------
 00000001000000000000000D |     4039624
(1 row)
```

类似地，pg_xlogfile_name仅提取预写式日志文件名。
当指定的预写式日志位置恰好位于预写式日志文件边界时，
这两个函数都返回前一个预写式日志文件的名称。
这通常正是管理预写式日志归档时所需的行为，
因为前一个文件是当前需要归档的最后一个文件。

有关正确使用这些函数的详细信息，参见第 24.3 节。

表 9.58中展示的函数提供有关备库当前状态的信息。这些函数既可以在恢复过程中也可以在正常运行期间被执行。

表 9.58. 恢复信息函数

名称	返回类型	描述
pg_is_in_recovery()	bool	如果恢复仍在进行，则为真。

名称	返回类型	描述
pg_last_xlog_receive_location()	text	获取流复制最近接收并同步到磁盘的预写式日志位置。在流复制进行期间，该值单调增加。如果恢复已完成，该值将保持为恢复期间最后接收并同步到磁盘的 WAL 记录的位置。如果禁用了流复制，或者流复制尚未开始，此函数返回 NULL。
pg_last_xlog_replay_location()	text	获取恢复期间最近重放的预写式日志位置。如果恢复仍在进行，该值单调增加。如果恢复已完成，该值将保持为该次恢复期间最后应用的 WAL 记录的位置。如果服务器未经恢复而正常启动，此函数返回 NULL。
pg_last_xact_replay_timestamp()	timestamp with time zone	获取恢复期间最近重放事务的时间戳，即该事务的提交或中止 WAL 记录在主库上生成的时间。如果恢复期间尚未重放任何事务，此函数返回 NULL。否则，如果恢复仍在进行，该值单调增加。如果恢复已完成，该值将保持为该次恢复期间最后应用的事务的时间戳。如果服务器未经恢复而正常启动，此函数返回 NULL。

表 9.59列出的函数用于控制恢复进度。这些函数只能在恢复期间执行。

表 9.59. 恢复控制函数

名称	返回类型	描述
pg_is_xlog_replay_paused()	bool	如果恢复已暂停，则为真。
pg_xlog_replay_pause()	void	立即暂停恢复。
pg_xlog_replay_resume()	void	如果恢复已暂停，则重新开始恢复。

恢复暂停时，不再应用数据库更改。如果处于热备状态，
所有新查询都会看到同一个一致的数据库快照，并且在恢复继续之前不会再产生查询冲突。

如果禁用了流复制，则暂停状态可能会无限期地持续下去，不会出现问题。
如果正在进行流复制，那么将继续接收WAL记录，这将最终填满可用磁盘空间，
这取决于暂停持续时间、WAL生成速度和可用磁盘空间。

表 9.60中的函数计算数据库对象的磁盘空间用量。

表 9.60. 数据库对象尺寸函数

名称	返回类型	描述
pg_column_size(any)	int	存储某个值所使用的字节数（可能已压缩）
pg_database_size(oid)	bigint	指定 OID 的数据库占用的磁盘空间

名称	返回类型	描述
pg_database_size(name)	bigint	指定名称的数据库占用的磁盘空间
pg_indexes_size(regclass)	bigint	指定表的索引占用的磁盘空间总量
pg_relation_size(*relation* regclass, *fork* text)	bigint	指定表或索引的指定分支（'main'、'fsm'、'vm' 或 'init'）占用的磁盘空间
pg_relation_size(*relation* regclass)	bigint	pg_relation_size(..., 'main') 的简写
pg_size_pretty(bigint)	text	将字节大小转换为带有大小单位的人类可读格式
pg_table_size(regclass)	bigint	指定表占用的磁盘空间，不包括索引（但包括 TOAST、空闲空间映射和可见性映射）
pg_tablespace_size(oid)	bigint	指定 OID 的表空间占用的磁盘空间
pg_tablespace_size(name)	bigint	指定名称的表空间占用的磁盘空间
pg_total_relation_size(regclass)	bigint	指定表占用的磁盘空间总量，包括所有索引和 TOAST 数据

pg_column_size 显示存储任意单个数据值所用的空间。

pg_total_relation_size 接受表或 TOAST 表的 OID 或名称，返回该表占用的磁盘总空间，包括所有关联索引。此函数等同于 pg_table_size + pg_indexes_size。

pg_table_size 接受表的 OID 或名称，返回该表所需的磁盘空间，不包括索引。（包括 TOAST 空间、空闲空间映射和可见性映射。）

pg_indexes_size 接受表的 OID 或名称，返回该表所有附属索引占用的磁盘总空间。

pg_database_size 和 pg_tablespace_size 接受数据库或表空间的 OID 或名称，返回其中占用的磁盘总空间。使用 pg_database_size 必须对指定数据库拥有 CONNECT 权限（默认授予）。使用 pg_tablespace_size 必须对指定表空间拥有 CREATE 权限；但如果该表空间是当前数据库的默认表空间，则不受此限制。

pg_relation_size接受表、索引或 TOAST 表的 OID 或名称，并以字节为单位返回该关系某个分支的磁盘大小。（注意，大多数情况下使用更高层的函数会更方便，例如pg_total_relation_size或pg_table_size，它们会将所有分支的大小相加。）只带一个参数时，返回关系的主数据分支大小。可以提供第二个参数来指定要检查的分支：

- 'main' 返回关系的主数据分支大小。
- 'fsm' 返回关系所关联的空闲空间映射的大小（参见第 55.3 节）。
- 'vm' 返回关系所关联的可见性映射的大小（参见第 55.4 节）。
- 'init' 返回关系所关联的初始化分支的大小（如果存在，见第 55.5 节）。

pg_size_pretty 可以将其他函数的结果格式化为便于阅读的形式，适当地使用 kB、MB、GB 或 TB 单位。

上述操作表和索引的函数接受一个regclass参数，它是该表或索引在pg_class系统目录中的 OID。你不必手工去查找该 OID，因为regclass数据类型的输入转换器会为你代劳。只写包围在单引号内的表名，这样它看起来像一个字面量。为了与普通SQL名称的处理相兼容，该字符串将被转换为小写形式，除非其中在表名周围包含双引号。

表 9.61中展示的函数帮助标识数据库对象相关的磁盘文件。

表 9.61. 数据库对象位置函数

名称	返回类型	描述
pg_relation_filenode(*relation* regclass)	oid	指定关系的文件结点编号
pg_relation_filepath(*relation* regclass)	text	指定关系的文件路径名

pg_relation_filenode 接受表、索引、序列或 TOAST 表的 OID 或名称，返回当前分配给它的“文件结点”编号。文件结点是关系文件名的基本组成部分（更多信息见第 55.1 节）。对于大多数表，结果与 pg_class.relfilenode 相同；但某些系统目录的 relfilenode 为零，必须使用此函数获取正确值。如果传入的关系没有存储空间（例如视图），函数返回 NULL。

pg_relation_filepath 与 pg_relation_filenode 类似，但返回关系的完整文件路径名（相对于数据库集簇的数据目录 PGDATA）。

表 9.62中的函数提供对数据库服务器所在机器上的文件的本地访问。只能访问数据库集簇目录和 log_directory 中的文件。集簇目录中的文件使用相对路径，日志文件使用与 log_directory 配置设置相匹配的路径。除非另有说明，这些函数仅限超级用户使用。

表 9.62. 通用文件访问函数

名称	返回类型	描述
pg_ls_dir(*dirname* text)	setof text	列出目录内容。
pg_read_file(*filename* text [, *offset* bigint, *length* bigint])	text	返回文本文件的内容。
pg_read_binary_file(*filename* text [, *offset* bigint, *length* bigint])	bytea	返回文件内容。
pg_stat_file(*filename* text)	record	返回文件信息。

pg_ls_dir 返回指定目录中的所有名称，但不包括特殊项“.”和“..”。

pg_read_file 从给定的 *offset* 开始返回文本文件的一部分，最多返回 *length* 字节（如果先到达文件末尾，则返回更少的字节）。如果 *offset* 为负，则它相对于文件末尾。如果省略 *offset* 和 *length*，则返回整个文件。从文件读取的字节按服务器编码解释为字符串；如果它们在该编码下无效，则会报错。

pg_read_binary_file类似于pg_read_file，但结果是bytea值，因此不执行编码检查。与convert_from函数配合使用，可以按指定编码读取文件：

```
SELECT convert_from(pg_read_binary_file('file_in_utf8.txt'), 'UTF8');
```

pg_stat_file返回一条记录，包含文件大小、最后访问时间戳、最后修改时间戳、最后文件状态更改时间戳（仅限 Unix 平台）、文件创建时间戳（仅限 Windows），以及一个boolean值，表示它是否是目录。典型用法包括：

```
SELECT * FROM pg_stat_file('filename');
```

```
SELECT (pg_stat_file('filename')).modification;
```

表 9.63中展示的函数管理咨询锁。有关正确使用这些函数的细节请参考第 13.3.4 节。

表 9.63. 咨询锁函数

名称	返回类型	描述
pg_advisory_lock(*key* bigint)	void	获取排他会话级咨询锁
pg_advisory_lock(*key1* int, *key2* int)	void	获取排他会话级咨询锁
pg_advisory_lock_shared(*key* bigint)	void	获取共享会话级咨询锁
pg_advisory_lock_shared(*key1* int, *key2* int)	void	获取共享会话级咨询锁
pg_advisory_unlock(*key* bigint)	boolean	释放排他会话级咨询锁
pg_advisory_unlock(*key1* int, *key2* int)	boolean	释放排他会话级咨询锁
pg_advisory_unlock_all()	void	释放当前会话持有的所有会话级咨询锁
pg_advisory_unlock_shared(*key* bigint)	boolean	释放共享会话级咨询锁
pg_advisory_unlock_shared(*key1* int, *key2* int)	boolean	释放共享会话级咨询锁
pg_advisory_xact_lock(*key* bigint)	void	获取排他事务级咨询锁
pg_advisory_xact_lock(*key1* int, *key2* int)	void	获取排他事务级咨询锁
pg_advisory_xact_lock_shared(*key* bigint)	void	获取共享事务级咨询锁
pg_advisory_xact_lock_shared(*key1* int, *key2* int)	void	获取共享事务级咨询锁
pg_try_advisory_lock(*key* bigint)	boolean	如果可用，则获取排他会话级咨询锁
pg_try_advisory_lock(*key1* int, *key2* int)	boolean	如果可用，则获取排他会话级咨询锁
pg_try_advisory_lock_shared(*key* bigint)	boolean	如果可用，则获取共享会话级咨询锁
pg_try_advisory_lock_shared(*key1* int, *key2* int)	boolean	如果可用，则获取共享会话级咨询锁
pg_try_advisory_xact_lock(*key* bigint)	boolean	如果可用，则获取排他事务级咨询锁
pg_try_advisory_xact_lock(*key1* int, *key2* int)	boolean	如果可用，则获取排他事务级咨询锁
pg_try_advisory_xact_lock_shared(*key* bigint)	boolean	如果可用，则获取共享事务级咨询锁

名称	返回类型	描述
pg_try_advisory_xact_lock_shared(*key1* int, *key2* int)	boolean	如果可用，则获取共享事务级咨询锁

pg_advisory_lock 锁定应用定义的资源，可以用一个 64 位键值或两个 32 位键值标识资源（注意，这两个键空间不重叠）。如果另一个会话已持有相同资源标识符上的锁，此函数会等待，直到该资源可用。该锁是排他的。多个锁请求会叠加，因此如果同一资源被锁定了三次，就必须解锁三次，才能释放给其他会话使用。

pg_advisory_lock_shared 的行为与 pg_advisory_lock 相同，但该锁可以与其他请求共享锁的会话共享。只有试图获取排他锁的会话会被阻挡。

pg_try_advisory_lock 与 pg_advisory_lock 类似，但不会等待锁变得可用。它要么立即获得锁并返回 true，要么在无法立即获得锁时返回 false。

pg_try_advisory_lock_shared 的行为与 pg_try_advisory_lock 相同，但尝试获取的是共享锁，而不是排他锁。

pg_advisory_unlock 释放先前获得的会话级排他咨询锁。如果成功释放锁，则返回 true。如果未持有该锁，则返回 false，服务器还会报告一条 SQL 警告。

pg_advisory_unlock_shared 的行为与 pg_advisory_unlock 相同，但释放的是会话级共享咨询锁。

pg_advisory_unlock_all 释放当前会话持有的所有会话级咨询锁。（会话结束时会隐式调用此函数，即使客户端非正常断开连接也是如此。）

pg_advisory_xact_lock 的行为与 pg_advisory_lock 相同，但锁会在当前事务结束时自动释放，不能显式释放。

pg_advisory_xact_lock_shared 的行为与 pg_advisory_lock_shared 相同，但锁会在当前事务结束时自动释放，不能显式释放。

pg_try_advisory_xact_lock 的行为与 pg_try_advisory_lock 相同，但如果获得了锁，该锁会在当前事务结束时自动释放，不能显式释放。

pg_try_advisory_xact_lock_shared 的行为与 pg_try_advisory_lock_shared 相同，但如果获得了锁，该锁会在当前事务结束时自动释放，不能显式释放。

9.25. 触发器函数

目前 PostgreSQL 提供一个内置触发器函数 suppress_redundant_updates_trigger，它会阻止那些实际上不会改变行中数据的更新。通常的行为则是无论数据是否改变都执行更新。（这种通常行为不需要检查，因此更新运行得更快，并且在某些情况下也有用。）

理想情况下，应尽量避免执行实际上不会改变记录中数据的更新。
冗余更新会耗费大量不必要的时间，尤其是在需要改动许多索引时；它们还会使死行占用空间，
最终需要通过清理回收。但是，在客户端代码中检测这种情况并不总是容易，
甚至可能无法做到，而编写用于检测的表达式也容易出错。另一种办法是使用 `suppress_redundant_updates_trigger`，跳过不改变数据的更新。但使用时应当谨慎。
此触发器处理每条记录的耗时虽小，却不可忽略，
因此如果一次更新涉及的大多数记录确实会改变，使用此触发器反而会使更新更慢。

该`suppress_redundant_updates_trigger`函数可以按如下方式添加到表中：

```
CREATE TRIGGER z_min_update
BEFORE UPDATE ON tablename
FOR EACH ROW EXECUTE PROCEDURE suppress_redundant_updates_trigger();
```

在大多数情况下，应让该触发器对每一行最后触发。由于触发器按名称顺序触发，
应选择一个排序位于表上其他所有触发器名称之后的名称。

关于创建触发器的更多信息，请参见CREATE TRIGGER。

第 10 章 类型转换

SQL语句可能有意或无意地要求在同一表达式中混用不同的数据类型。
PostgreSQL提供了丰富的机制来求值这类混合类型表达式。

在很多情况下，用户并不需要理解类型转换机制的细节。不过，
PostgreSQL执行的隐式类型转换会影响查询结果。必要时，可以使用
显式类型转换来控制这些结果。

本章介绍PostgreSQL的类型转换机制和约定。
有关特定数据类型以及允许使用的函数和操作符的更多信息，请参阅第 8 章和第 9 章中的相关章节。

10.1. 概述

SQL是一种强类型语言。也就是说，
每个数据项都关联着一个决定其行为和允许用法的数据类型。
PostgreSQL拥有可扩展的类型系统，比其他SQL实现更通用也更灵活。因此，
PostgreSQL中的大多数类型转换行为都由通用规则控制，而不是依赖临时性的启发式规则。
这使得即使表达式中包含用户定义类型，也能够使用混合类型表达式。

PostgreSQL扫描器/解析器将词法元素分成五个基本类别：整数、非整数数字、字符串、
标识符和关键字。大多数非数值类型的常量会先被归类为字符串。
SQL语言定义允许给字符串指定类型名，而
PostgreSQL可以利用这一机制让解析器从一开始就沿着正确的路径处理。例如，查询：

```
SELECT text 'Origin' AS "label", point '(0,0)' AS "value";

 label  | value
--------+-------
 Origin | (0,0)
(1 row)
```

包含两个字面量，它们的类型分别是`text`和`point`。如果没有为某个字符串字面量指定类型，
那么它最初会被赋予占位类型`unknown`，并在后文所述的后续阶段解析。

在PostgreSQL解析器中，有四种基本的SQL结构需要采用不同的类型转换规则：

函数调用

PostgreSQL类型系统的很大一部分建立在一套丰富的函数之上。
函数可以有一个或多个参数。由于PostgreSQL支持函数重载，
单凭函数名无法唯一确定要调用哪个函数；
解析器必须根据所提供参数的数据类型选择正确的函数。

操作符

PostgreSQL允许表达式使用前缀和后缀一元（单参数）操作符，以及二元（双参数）
操作符。与函数一样，操作符也可以重载，因此同样存在选择正确操作符的问题。

值存储

SQL的`INSERT`和`UPDATE`语句会把表达式的结果放入表中。
语句中的表达式必须与目标列的类型匹配，并且在必要时转换为目标列类型。

UNION、CASE和相关结构

由于经UNION合并的SELECT语句的所有查询结果都必须出现在同一组列中，因此每个SELECT子句的结果类型必须彼此匹配，并转换为统一的一组类型。类似地，CASE结构中的结果表达式必须转换为某种公共类型，这样整个CASE表达式才有确定的输出类型。对于ARRAY构造以及GREATEST和LEAST函数，情况也是如此。

系统目录保存了哪些数据类型之间存在哪些类型转换，以及如何执行这些转换的信息。用户还可以使用CREATE CAST命令添加额外的类型转换。（这通常与定义新数据类型一并完成。内置类型之间的类型转换集合经过精心设计，最好不要改动。）

解析器还提供了一种额外的启发式规则，可在存在隐式类型转换的一组类型之间更准确地判定适当的转换行为。数据类型被划分为若干基本的类型分类，包括boolean、numeric、string、bitstring、datetime、timespan、geometric、network以及用户定义类型。（列表见表 45.49；不过也可以创建自定义类型分类。）每个分类中可以有一个或多个首选类型，当存在多种可能类型可选时，会优先选择它们。通过仔细选择首选类型和可用的隐式类型转换，可以确保有歧义的表达式（即存在多个候选解析方案的表达式）能够以有用的方式被解析。

所有类型转换规则都以以下几个原则为基础：

- 隐式转换绝不应产生令人意外或无法预测的结果。
- 如果查询不需要隐式类型转换，解析器和执行器就不应承担额外开销。也就是说，如果查询本身书写正确且类型已经匹配，那么执行时不应让解析器额外耗时，也不应在查询中引入不必要的隐式类型转换调用。
- 此外，如果某个查询通常需要先对参数做隐式类型转换才能调用某个函数，而此后用户又定义了一个参数类型正确的新函数，那么解析器应改为使用这个新函数，而不再通过隐式转换去调用旧函数。

10.2. 操作符

操作符表达式所引用的具体操作符按照下述过程确定。注意，该过程还会间接受到相关操作符优先级的影响，因为优先级决定了哪些子表达式会被视为哪些操作符的输入。详见第 4.1.6 节。

操作符类型解析

1. 从系统目录pg_operator中选取要考虑的操作符。如果使用的是未带模式限定的操作符名（通常如此），则考虑当前搜索路径中可见且名称和参数个数匹配的操作符（见第 5.7.3 节）。如果给出了限定名，则只考虑指定模式中的操作符。
   - (可选的) 如果搜索路径中找到多个参数类型完全相同的操作符，只考虑路径中最早出现的那一个。参数类型不同的操作符则不受搜索路径位置影响，处于同等地位。
2. 检查是否存在一个恰好接受输入参数类型的操作符。如果存在（在被考虑的操作符集合中至多只有一个精确匹配），就使用它。
   a. (可选的) 如果二元操作符调用的一个参数是unknown类型，则在本次检查中假定它与另一个参数类型相同。涉及两个unknown输入的调用，或带有unknown输入的一元操作符，在这一步都不可能找到匹配。

b. (可选的) 如果二元操作符调用的一个参数是unknown类型，而另一个是域类型，则接着检查是否存在两边都恰好接受该域基础类型的操作符；如果有，就使用它。

3. 寻找最佳匹配。

a. 丢弃那些输入类型不匹配，且也无法通过隐式类型转换实现匹配的候选操作符。为此，假定unknown字面量可以转换为任何类型。如果只剩一个候选操作符，就使用它；否则继续下一步。

b. 如果任何输入参数是域类型，则在后续所有步骤中都把它当作该域的基础类型。这样可以确保在解决操作符歧义时，域的行为与其基础类型一致。

c. 遍历所有候选操作符，保留那些在输入类型上精确匹配数最多的候选项。如果没有任何候选项存在精确匹配，则保留全部候选项。如果只剩一个候选操作符，就使用它；否则继续下一步。

d. 遍历所有候选操作符，在需要类型转换的位置上，保留接受首选类型（属于输入数据类型的类型分类）的位置数最多的候选项。如果没有候选项接受首选类型，则保留全部候选项。如果只剩一个候选操作符，就使用它；否则继续下一步。

e. 如果有任何输入参数是unknown，就检查剩余候选操作符在这些参数位置上接受哪些类型分类。在每个位置上，如果有任何候选项接受string分类，就选择该分类。（这种偏向字符串的做法是合适的，因为unknown类型的字面量看起来像字符串。）否则，如果所有剩余候选项都接受同一种类型分类，就选择该分类；否则失败，因为在缺少更多线索的情况下无法推断出正确选择。然后丢弃不接受所选类型分类的候选项。此外，如果有任何候选项接受该分类中的首选类型，就丢弃在该参数上接受非首选类型的候选项。

f. 如果只剩一个候选操作符，就使用它。如果没有候选操作符或者不止一个候选操作符剩余，则失败。

下面是一些示例。

例 10.1. 阶乘操作符类型解析

标准目录中只定义了一个阶乘操作符（后缀 !），它接受一个类型为 bigint 的参数。扫描器为以下查询表达式中的参数赋予初始类型 integer：

```
SELECT 40 ! AS "40 factorial";

                    40 factorial
--------------------------------------------------
 815915283247897734345611269596115894272000000000
(1 row)
```

因此，解析器会对操作数进行类型转换，该查询等价于：

```
SELECT CAST(40 AS bigint) ! AS "40 factorial";
```

例 10.2. 字符串连接操作符类型解析

处理字符串类型以及复杂的扩展类型时，常常会使用类似字符串的语法。未指定类型的字符串会与可能的操作符候选项进行匹配。

一个参数未指定类型的例子：

```
SELECT text 'abc' || 'def' AS "text and unknown";

 text and unknown
------------------
 abcdef
(1 row)
```

在这种情况下，解析器会查看是否存在一个两边参数都接受`text`的操作符。既然存在，它就会假定第二个参数应解释为`text`类型。

下面是两个未指定类型值的连接：

```
SELECT 'abc' || 'def' AS "unspecified";

 unspecified
-------------
 abcdef
(1 row)
```

这里对使用哪种类型没有初始提示，因为查询中没有指定任何类型。因此，解析器会查找所有候选操作符，并发现既有接受字符串分类输入的候选项，也有接受位串分类输入的候选项。由于在可用时优先选择字符串分类，所以会选中该分类，再使用字符串的首选类型`text`作为解析这些unknown类型字面量的具体类型。

例 10.3. 绝对值与取反操作符类型解析

PostgreSQL操作符目录中为前缀操作符`@`提供了多个条目，它们分别实现各种数值数据类型的绝对值操作。其中一个条目对应`float8`，它是数值分类中的首选类型。因此，当遇到一个`unknown`输入时，PostgreSQL会使用该条目：

```
SELECT @ '-4.5' AS "abs";
 abs
-----
 4.5
(1 row)
```

这里，系统在应用所选操作符之前，已经把unknown类型字面量隐式解析为`float8`。我们可以验证使用的确实是`float8`，而不是其他类型：

```
SELECT @ '-4.5e500' AS "abs";

ERROR:  "-4.5e500" is out of range for type double precision
```

另一方面，前缀操作符`~`（按位取反）只对整数数据类型定义，并没有为`float8`定义。因此，如果对`~`试一个类似的例子，会得到：

```
SELECT ~ '20' AS "negation";

ERROR:  operator is not unique: ~ "unknown"
HINT:  Could not choose a best candidate operator. You might need to
 add
explicit type casts.
```

这是因为系统无法决定几个可能的`~`操作符中应优先选择哪一个。我们可以通过显式类型转换来帮助它：

```
SELECT ~ CAST('20' AS int8) AS "negation";

 negation
----------
      -21
(1 row)
```

例 10.4. 域类型上的自定义操作符

用户有时会尝试声明只适用于某个域类型的操作符。这是可行的，但远没有看上去那样有用，因为操作符解析规则被设计成优先选择作用于域基础类型的操作符。考虑下面的例子：

```
CREATE DOMAIN mytext AS text CHECK(...);
CREATE FUNCTION mytext_eq_text (mytext, text) RETURNS boolean AS ...;
CREATE OPERATOR = (procedure=mytext_eq_text, leftarg=mytext, rightarg=
text);
CREATE TABLE mytable (val mytext);

SELECT * FROM mytable WHERE val = 'foo';
```

这个查询不会使用自定义操作符。解析器首先会检查是否存在`mytext = mytext`操作符（步骤 2.a），但并不存在；然后它会考虑该域的基础类型`text`，并检查是否存在`text = text`操作符（步骤 2.b），而这是存在的；于是它会把`unknown`类型的字面量解析为`text`，并使用`text = text`操作符。要让自定义操作符被使用，唯一的方法是显式地对该字面量做类型转换：

```
SELECT * FROM mytable WHERE val = text 'foo';
```

这样就会根据精确匹配规则立即找到`mytext = text`操作符。如果解析过程进入最佳匹配规则，它们会主动排斥作用于域类型的操作符。如果不是这样，这类操作符会导致过多的操作符歧义失败，因为类型转换规则总是把域视为可以转换到其基础类型或从其基础类型转换而来，因此域操作符会在所有与基础类型上同名操作符相同的场景中都被视为可用。

10.3. 函数

函数调用所引用的具体函数按照下述过程确定。

函数类型解析

1. 从系统目录`pg_proc`中选取要考虑的函数。如果使用的是未带模式限定的函数名，则考虑当前搜索路径中可见且名称和参数个数匹配的函数（见第 5.7.3 节）。如果给出了限定名，则只考虑指定模式中的函数。

   a. (可选的) 如果搜索路径中找到多个参数类型完全相同的函数，只考虑路径中最早出现的那一个。参数类型不同的函数则不受搜索路径位置影响，处于同等地位。

   b. (可选的) 如果函数声明了一个`VARIADIC`数组参数，而调用时没有使用`VARIADIC`关键字，则把数组参数视为一个或多个其元素类型的参数，个数按匹配该调用的需要确定。扩展之后，该函数的有效参数类型可能与某个非variadic函数完全相同。在这种情况下，使用搜索路径中较早出现的函数；如果两个函数位于同一模式，则优先选择非variadic的那个。

   c. (可选的) 带有参数默认值的函数，被认为可以匹配任何省略了零个或多个可使用默认值的参数位置的调用。如果不止一个这样的函数能匹配某个调用，

则使用搜索路径中最早出现的那一个。如果同一模式中存在两个或更多这样的函数，并且它们在未使用默认值的参数位置上的参数类型相同（如果它们具有不同的带有默认值的参数集合，这种情况是可能的），系统就无法判断该偏好哪一个；如果找不到更好的匹配，就会报“有歧义的函数调用”错误。

2. 检查是否存在一个恰好接受输入参数类型的函数。
   如果存在（在被考虑的函数集合中至多只有一个精确匹配），就使用它。（涉及`unknown`的情况在这一步永远找不到匹配。）

3. 如果没有找到精确匹配，就检查该函数调用是否看起来像一种特殊的类型转换请求。
   这种情况发生在：函数调用只有一个参数，且函数名与某个数据类型的（内部）名称相同。此外，函数参数必须是unknown类型字面量，或者是一种可以二进制强制转换为该命名数据类型的类型，或者是一种可以通过应用其 I/O 函数转换为该命名数据类型的类型（也就是说，该转换要么转到某种标准字符串类型，要么从某种标准字符串类型转来）。满足这些条件时，该函数调用会被当作一种`CAST`形式。[1]

4. 寻找最佳匹配。

   a. 丢弃那些输入类型不匹配，且也无法通过隐式类型转换实现匹配的候选函数。为此，假定`unknown`字面量可以转换为任何类型。如果只剩一个候选函数，就使用它；否则继续下一步。

   b. 如果任何输入参数是域类型，则在后续所有步骤中都把它当作该域的基础类型。这样可以确保在解决函数歧义时，域的行为与其基础类型一致。

   c. 遍历所有候选函数，保留那些在输入类型上精确匹配数最多的候选项。
      如果没有任何候选项存在精确匹配，则保留全部候选项。如果只剩一个候选函数，就使用它；否则继续下一步。

   d. 遍历所有候选函数，在需要类型转换的位置上，保留接受首选类型（属于输入数据类型的类型分类）的位置数最多的候选项。
      如果没有候选项接受首选类型，则保留全部候选项。如果只剩一个候选函数，就使用它；否则继续下一步。

   e. 如果有任何输入参数是`unknown`，就检查剩余候选函数在这些参数位置上接受哪些类型分类。在每个位置上，如果有任何候选项接受`string`分类，就选择该分类。
      （这种偏向字符串的做法是合适的，因为unknown类型的字面量看起来像字符串。）
      否则，如果所有剩余候选项都接受同一种类型分类，就选择该分类；否则失败，因为在缺少更多线索的情况下无法推断出正确选择。
      然后丢弃不接受所选类型分类的候选项。此外，如果有任何候选项接受该分类中的首选类型，就丢弃在该参数上接受非首选类型的候选项。

   f. 如果只剩一个候选函数，就使用它。如果没有候选函数或者不止一个候选函数剩余，则失败。

注意，操作符类型解析和函数类型解析的“最佳匹配”规则是完全相同的。下面是一些示例。

例 10.5. round 函数参数类型解析

只有一个接受两个参数的`round`函数；它的第一个参数类型是`numeric`，第二个参数类型是`integer`。因此下面的查询会自动把类型为`integer`的第一个参数转换为`numeric`：

[1] 之所以有这一步，是为了在不存在实际类型转换函数的情况下仍支持函数风格的类型转换写法。如果存在类型转换函数，按惯例它会以其输出类型命名，因此无需为此设置特殊情况。更多说明见CREATE CAST。

```
SELECT round(4, 4);

 round
--------
 4.0000
(1 row)
```

该查询实际上会被解析器改写成：

```
SELECT round(CAST (4 AS numeric), 4);
```

由于带小数点的数字常量最初会被赋予numeric类型，下面的查询不需要类型转换，因此可能稍微更高效一些：

```
SELECT round(4.0, 4);
```

例 10.6. substr 函数类型解析

存在多个substr函数，其中一个接受text和integer。如果以未指定类型的字符串常量调用，系统会选择接受首选类型分类string参数的候选函数（也就是text类型的那个）。

```
SELECT substr('1234', 3);

 substr
--------
     34
(1 row)
```

如果字符串被声明为varchar类型，例如它来自表中的某一列，那么解析器会尝试把它转换为text：

```
SELECT substr(varchar '1234', 3);

 substr
--------
     34
(1 row)
```

解析器会把它改写成如下形式：

```
SELECT substr(CAST (varchar '1234' AS text), 3);
```

> **注意**
>
> 解析器从pg_cast目录得知text与varchar二进制兼容，这意味着把其中一种类型传给接受另一种类型的函数时，无需做任何物理转换。因此，这种情况下实际上不会插入类型转换调用。

如果用integer类型参数调用该函数，解析器会尝试把它转换为text：

```
SELECT substr(1234, 3);
ERROR:  function substr(integer, integer) does not exist
HINT:  No function matches the given name and argument types. You
 might need
```

```
to add explicit type casts.
```

这行不通，因为integer并没有到text的隐式类型转换。不过，显式类型转换可以：

```
SELECT substr(CAST (1234 AS text), 3);

 substr
--------
     34
(1 row)
```

10.4. 值存储

要插入表中的值会按照以下步骤转换为目标列的数据类型。

值存储的类型转换

1. 检查是否与目标类型精确匹配。

2. 否则，尝试把表达式转换为目标类型。如果这两种类型之间存在已登记的类型转换，这就会成功。如果表达式是unknown类型字面量，则会把字符串字面量的内容送入目标类型的输入转换例程。

3. 检查目标类型是否有尺寸调整类型转换。尺寸调整类型转换是该类型到其自身的类型转换。如果在pg_cast目录中找到了这种转换，就在把表达式存入目标列之前先应用它。此类类型转换的实现函数总是额外接受一个integer参数，用来接收目标列的atttypmod值（通常是其声明长度，尽管不同数据类型对atttypmod的解释各不相同）；它还可能接受第三个boolean参数，用来说明该类型转换是显式还是隐式的。类型转换函数负责应用所有与长度相关的语义，例如长度检查或截断。

例 10.7. character 的存储类型转换

对于声明为character(20)的目标列，下面的语句表明存储值会被调整到正确的长度：

```
CREATE TABLE vv (v character(20));
INSERT INTO vv SELECT 'abc' || 'def';
SELECT v, octet_length(v) FROM vv;

          v           | octet_length
----------------------+--------------
 abcdef               |           20
(1 row)
```

这里实际发生的是，两个unknown类型字面量默认都被解析为text，从而使||操作符被解析为text的连接操作。然后，操作符得到的text结果被转换为bpchar（“用空格填充的字符”，blank-padded char，即character数据类型的内部名称），以匹配目标列类型。（由于从text到bpchar的转换是二进制可强制转换的，这一步不会插入任何真实的函数调用。）最后，系统从系统目录中找到尺寸调整函数bpchar(bpchar, integer, boolean)，并把它应用到操作符结果以及存储列长度上。这个特定于类型的函数会执行所需的长度检查并补齐填充空格。

10.5. UNION、CASE及相关结构

SQL 的UNION结构必须让可能不同的类型彼此匹配，以形成单一结果集。该解析算法会分别应用到联合查询的每个输出列。INTERSECT和EXCEPT结构以与UNION相同的方式解析不同类型。CASE、ARRAY、VALUES以及GREATEST和LEAST
构造使用完全相同的算法来匹配其组成表达式并选择结果数据类型。

UNION、CASE及相关结构的类型解析

1. 如果所有输入都是同一类型，且该类型不是unknown，就解析为该类型。
2. 如果任何输入是域类型，则在后续所有步骤中都把它当作该域的基础类型。[2]
3. 如果所有输入都是unknown类型，则解析为text类型（字符串分类的首选类型）。否则，忽略unknown输入。
4. 如果非unknown输入并非全都属于同一类型分类，则失败。
5. 选择该分类中的首选类型（如果第一个非unknown输入类型恰好是首选类型，则选它）。
6. 否则，
   选择最后一个能让所有在它之前的非unknown输入都隐式转换为它的非unknown输入类型。（这样的类型总是存在的，因为列表中至少第一个类型就满足这一条件。）
7. 把所有输入都转换为所选类型。如果某个输入到所选类型不存在转换，则失败。

下面是一些示例。

例 10.8. 未充分指定类型的 UNION 中的类型解析

```
SELECT text 'a' AS "text" UNION SELECT 'b';

 text
------
 a
 b
(2 rows)
```

这里，unknown类型字面量'b'会被解析为text类型。

例 10.9. 简单 UNION 中的类型解析

```
SELECT 1.2 AS "numeric" UNION SELECT 1;

 numeric
---------
       1
     1.2
(2 rows)
```

字面量1.2的类型是numeric，而integer值1可以隐式转换为numeric，因此使用该类型。

例 10.10. 次序对调的 UNION 中的类型解析

```
SELECT 1 AS "real" UNION SELECT CAST('2.2' AS REAL);
```

[2] 这在某种程度上类似于操作符和函数对域输入的处理方式；只要用户确保所有输入都隐式或显式地正好属于该域类型，这种行为就能让域类型在UNION或类似结构中得以保留。否则，将优先使用该域的基础类型。

```
 real
------
    1
  2.2
(2 rows)
```

这里，由于real类型不能隐式转换为integer，而integer可以隐式转换为real，因此UNION结果类型被解析为real。

第 11 章 索引

索引是增强数据库性能的一种常见方式。
索引允许数据库服务器比没有索引时更快地找到并检索特定行。
但索引也会给整个数据库系统带来额外开销，因此应当合理使用。

11.1. 简介

假设我们有一个类似如下的表：

```
CREATE TABLE test1 (
    id integer,
    content varchar
);
```

并且应用会发出很多如下形式的查询：

```
SELECT content FROM test1 WHERE id = constant;
```

如果没有预先准备，系统就必须逐行扫描整个test1表，以找出所有匹配项。如果test1中有很多行，而这种查询只会返回很少几行（甚至可能一行也没有，或者只有一行），这显然是低效的办法。但是如果系统被要求维护一个基于id列的索引，它就可以用更高效的方法定位匹配行。例如，它可能只需要在一棵搜索树中向下走几层。

大多数非虚构类书籍也采用了类似的方法：读者经常查找的术语和概念会被收集到书末按字母顺序排列的索引中。感兴趣的读者可以相对快速地浏览索引并翻到相应页面，而不必通读整本书才能找到自己感兴趣的内容。正如作者需要预判读者可能查找哪些条目一样，数据库程序员也需要预见哪些索引会有用。

如前所述，下列命令可用于在id列上创建索引：

```
CREATE INDEX test1_id_index ON test1 (id);
```

test1_id_index这个名字可以自由选择，但最好选一个以后仍能让你记住该索引用途的名字。

要删除索引，使用DROP INDEX命令。索引可以在任何时候添加到表上，也可以在任何时候从表上删除。

一旦创建了索引，就无需再进行额外干预：系统会在表被修改时更新索引，并且会在它认为这样做比顺序扫描表更高效时，在查询中使用该索引。不过，你可能仍需要定期运行ANALYZE命令来更新统计信息，以便查询规划器作出更有根据的决策。关于如何判断某个索引是否被使用，以及规划器何时和为何可能选择不使用索引，可参见第 14 章。

索引还可以让带有搜索条件的UPDATE和DELETE命令受益。索引也可以用于连接搜索。因此，定义在连接条件组成列上的索引也能显著加速带连接的查询。

在大表上创建索引可能需要很长时间。默认情况下，PostgreSQL允许在创建索引期间并行执行读取（SELECT语句），但写入（INSERT、UPDATE、DELETE）会被阻塞，直到索引构建完成。在生产环境中，这往往不可接受。可以允许在创建索引时并行写入，但有若干注意事项需要了解，详见并发构建索引。

索引创建之后，系统还必须使其与表保持同步。这会给数据操作增加开销。因此，那些在查询中很少使用或从不使用的索引应当被移除。

11.2. 索引类型

PostgreSQL提供了多种索引类型：B-树、Hash、GiST 和 GIN。每种索引类型都采用不同的算法，分别最适合不同类型的查询。默认情况下，`CREATE INDEX`命令创建的是 B-树索引，它适用于最常见的场景。

B-树能够处理可以按某种顺序排序的数据上的等值查询和范围查询。特别是，只要已索引列参与了下列任一操作符的比较，PostgreSQL查询规划器就会考虑使用 B-树索引：

```
<
<=
=
>=
>
```

与这些操作符组合等价的构造，例如`BETWEEN`和`IN`，也可以通过 B-树索引搜索来实现。此外，索引列上的`IS NULL`或`IS NOT NULL`条件也可以配合 B-树索引使用。

如果模式是常量，并且锚定在字符串起始位置，优化器也可以对涉及模式匹配操作符`LIKE`和`~`的查询使用 B-树索引，例如`col LIKE 'foo%'`或`col ~ '^foo'`，但不能用于`col LIKE '%bar'`。不过，如果你的数据库没有使用 C 区域设置，就需要以一个特殊操作符类来创建该索引，才能支持模式匹配查询的索引化，详见下文第 11.9 节。B-树索引也可以用于`ILIKE`和`~*`，但前提是模式以非字母字符开头，也就是不会受到大小写转换影响的字符。

B-树索引还可以用于按排序顺序取回数据。这并不总是比简单扫描再排序更快，但通常会很有帮助。

Hash 索引只能处理简单的等值比较。只要已索引列参与的是使用=操作符的比较，查询规划器就会考虑使用 Hash 索引。使用以下命令创建 Hash 索引：

```
CREATE INDEX name ON table USING hash (column);
```

> **小心**
>
> 哈希索引操作目前不会写入 WAL 日志，因此如果发生过未写入的更改，数据库崩溃后可能需要用`REINDEX`重建哈希索引。此外，在初始基础备份之后，哈希索引的更改不会通过流复制或基于文件的复制进行复制，因此它们会对随后使用它们的查询给出错误的答案。由于这些原因，目前不建议使用哈希索引。

GiST 索引并不是某一种单独的索引，而是一种基础设施，可在其中实现许多不同的索引策略。因此，GiST 索引可使用哪些具体操作符，取决于所采用的索引策略（即操作符类）。例如，PostgreSQL标准发布版中包含了若干二维几何数据类型的 GiST 操作符类，它们支持使用下列操作符的索引化查询：

```
<<
&<
&>
>>
<<|
&<|
|&>
|>>
```

```
@>
<@
~=
&&
```

（这些操作符的含义见第 9.11 节。）还有许多其他 GiST 操作符类可在`contrib`集合或独立项目中获得。更多信息见第 53 章。

GiST 索引还能够优化“最近邻”搜索，例如：

```
SELECT * FROM places ORDER BY location <-> point '(101,456)' LIMIT 10;
```

这会找出距离给定目标点最近的十个地点。是否能做到这一点，
同样取决于所使用的具体操作符类。

GIN 索引是倒排索引，可以处理包含多个键的值，例如数组。与 GiST 一样，GIN 也能支持多种不同的用户定义索引策略，GIN 索引可使用哪些具体操作符同样取决于索引策略。例如，PostgreSQL标准发布版中包含用于一维数组的 GIN 操作符类，它们支持使用下列操作符的索引化查询：

```
<@
@>
=
&&
```

（这些操作符的含义见第 9.17 节。）还有许多其他 GIN 操作符类可在`contrib`集合或独立项目中获得。更多信息见第 54 章。

11.3. 多列索引

一个索引可以定义在表的多个列上。例如，如果你有一个如下形式的表：

```
CREATE TABLE test2 (
  major int,
  minor int,
  name varchar
);
```

（假设你把`/dev`目录保存在数据库里.....）并且经常发出如下查询：

```
SELECT name FROM test2 WHERE major = constant AND minor = constant;
```

那么将`major`和`minor`两列一起建一个索引可能是合适的，例如：

```
CREATE INDEX test2_mm_idx ON test2 (major, minor);
```

当前，只有 B-树、GiST 和 GIN 索引类型支持多列索引。最多可以指定 32 列。（这个限制可以在构建 PostgreSQL 时修改；参见文件 `pg_config_manual.h`。）

多列 B-树索引可以用于涉及索引任意列子集的查询条件，但当对前导（最左）列存在约束时，索引效率最高。精确的规则是：对前导列的等值约束，
再加上第一个没有等值约束的列上的任何不等约束，会被用来限制索引扫描的范围。
这些列右侧列上的约束会在索引中进行检查，因此可以减少访问表本体的次数，
但不会缩小必须扫描的索引范围。例如，给定一个 `(a, b, c)` 上的索引和查询条件 `WHERE a = 5 AND b >= 42 AND c < 77`，索引必须从第一个 `a` = 5 且 `b` = 42 的条目扫描到最后一个 `a` = 5 的条目。`c` >= 77 的索引项会被排除，但仍然必须扫描这些项。原则上，这个索引也可以用于

只对 b 和/或 c 有约束、而对 a 没有约束的查询，但必须扫描整个索引，因此在大多数情况下，规划器会倾向于顺序扫描表，而不是使用该索引。

多列 GiST 索引可以用于涉及索引任意列子集的查询条件。附加列上的条件会限制索引返回的项，但决定索引需要扫描多少内容的，最重要的仍是第一列上的条件。
如果第一列只有很少几个非重复值，即使其他列有很多非重复值，GiST 索引也会相对低效。

多列 GIN 索引可以用于涉及索引任意列子集的查询条件。与 B-树或 GiST 不同，无论查询条件使用的是哪些索引列，GIN 的索引搜索效果都一样。

当然，每一列都必须配合适合该索引类型的操作符使用；涉及其他操作符的子句不会被考虑。

多列索引应谨慎使用。在大多数情况下，单列索引已经足够，而且更省空间、也更省时间。
除非表的使用方式极其程式化，否则超过三列的索引通常不会有帮助。
关于不同索引配置优缺点的讨论，还可参见第 11.5 节。

11.4. 索引和ORDER BY

除了简单地找到查询要返回的行之外，索引还可能能够以某种特定排序顺序交付这些行。
这使得查询中的ORDER BY要求无需额外排序步骤即可满足。在PostgreSQL当前支持的索引类型中，只有 B-树能产生有序输出，其他索引类型返回匹配行的顺序则未指定，并且依赖具体实现。

规划器在满足ORDER BY要求时，会考虑两种方案：扫描一个与该要求匹配的可用索引，或者按物理顺序扫描表再显式排序。对于需要扫描表中很大一部分内容的查询，
显式排序通常会比使用索引更快，因为它遵循顺序访问模式，需要的磁盘 I/O 更少。只需取出少量行时，索引更有价值。一个重要的特例是ORDER BY与LIMIT *n*组合使用：显式排序必须处理所有数据才能找出最前面的*n*行，而如果有一个与ORDER BY匹配的索引，就可以直接取回前*n*行，而完全不必扫描剩余数据。

默认情况下，B-树索引按升序存储其项，并将空值放在最后。这意味着，对列x
上的索引进行前向扫描，会产生满足ORDER BY x的输出（更完整地说，是ORDER BY x ASC NULLS LAST）。索引也可以反向扫描，从而产生满足ORDER BY x DESC的输出（更完整地说，是ORDER BY x DESC NULLS FIRST，因为NULLS FIRST是ORDER BY DESC的默认行为）。

你可以在创建 B-树索引时通过指定ASC、DESC、NULLS FIRST和/或NULLS LAST选项来调整索引的排序方式，例如：

```
CREATE INDEX test2_info_nulls_low ON test2 (info NULLS FIRST);
CREATE INDEX test3_desc_index ON test3 (id DESC NULLS LAST);
```

一个按升序且空值在前存储的索引，根据扫描方向不同，可以满足ORDER BY x ASC NULLS FIRST或ORDER BY x DESC NULLS LAST。

你可能会疑惑，既然借助两个选项再加上反向扫描的可能性就能覆盖所有ORDER BY变体，为什么还要提供四个选项。对于单列索引，这些选项确实是冗余的，
但对于多列索引它们就可能很有用。考虑一个(x, y)上的两列索引：前向扫描时它能满足ORDER BY x, y，反向扫描时能满足ORDER BY x DESC, y DESC。但应用也可能经常需要ORDER BY x ASC, y DESC。普通索引无法产生这种顺序，而如果把索引定义成(x ASC, y DESC)或(x DESC, y ASC)，就可以做到。

显然，具有非默认排序方式的索引是一项相当专门的特性，
但在某些查询上它们有时能带来巨大的加速。是否值得维护这样的索引，
取决于你需要这种特殊排序的查询出现得有多频繁。

11.5. 组合多个索引

单次索引扫描只能利用那些在索引列上使用了该索引操作符类中操作符、并且彼此以AND连接的查询子句。例如，给定一个(a, b)上的索引，查询条件WHERE a = 5 AND b = 6可以使用该索引，而WHERE a = 5 OR b = 6这样的查询就不能直接使用该索引。

幸运的是，PostgreSQL能够组合多个索引（包括同一个索引的多次使用），以处理单次索引扫描无法实现的情况。系统可以在多个索引扫描之间形成AND和OR条件。例如，WHERE x = 42 OR x = 47 OR x = 53 OR x = 99这样的查询，可以拆分成对x上的一个索引进行四次独立扫描，每次扫描使用一个查询子句。然后将这些扫描的结果做 OR 运算以得到最终结果。另一个例子是，如果我们在x和y上分别有索引，那么WHERE x = 5 AND y = 6这样的查询，其一种可能实现方式就是分别用两个索引处理相应条件，然后将索引结果做 AND 运算，以确定结果行。

为了组合多个索引，系统会扫描每个所需索引，并在内存中准备一个位图，给出满足该索引条件的表行位置。然后按查询需要对这些位图进行 AND 和 OR 运算。最后，再访问并返回实际的表行。表行是按物理顺序访问的，因为位图就是这样布局的；这意味着原始索引中的任何排序都会丢失，所以如果查询带有ORDER BY子句，就需要单独的排序步骤。也正因为如此，再加上每多一次索引扫描都会增加额外时间，规划器有时会选择使用简单的单一索引扫描，即使还有其他本来也可以利用的索引。

除了最简单的应用之外，通常会有多种可能有用的索引组合，数据库开发者必须权衡决定提供哪些索引。有时多列索引最好，但有时创建独立索引并依赖索引组合功能会更合适。例如，如果你的工作负载包含一组查询：有时只涉及列x，有时只涉及列y，有时同时涉及两列，那么你可以选择分别在x和y上创建两个独立索引，并依赖索引组合来处理同时使用两列的查询。你也可以创建一个(x, y)上的多列索引。对于同时涉及两列的查询，这个索引通常会比索引组合更高效，但正如第 11.3 节中所讨论的，它对于只涉及y的查询几乎没有用处，因此不应只创建这一个索引。将这个多列索引与一个单独的y索引组合起来，能够取得相当好的效果。对于只涉及x的查询，多列索引也可以使用，但它会比单独的x索引更大，因此更慢。最后一种选择是同时创建这三个索引，但这大概只有在该表被搜索的频率远高于被更新的频率，并且三类查询都很常见时才合理。如果其中一种查询远没有另外两种常见，那么你大概只需创建最适合常见查询类型的两个索引即可。

11.6. 唯一索引

索引还可以用于强制列值唯一，或者强制多个列组合值唯一。

```
CREATE UNIQUE INDEX name ON table (column [, ...]);
```

当前，只有 B-树索引可以声明为唯一。

当一个索引被声明为唯一时，就不允许存在多个具有相同索引值的表行。空值不被视为相等。多列唯一索引只会拒绝那种在多行中所有索引列都相等的情况。

当为表定义唯一约束或主键时，PostgreSQL会自动创建一个唯一索引。该索引覆盖构成主键或唯一约束的列（必要时是多列索引），并且正是它在强制执行该约束。

> **注意**
>
> 向表添加唯一约束的首选方式是ALTER TABLE ... ADD CONSTRAINT。用索引来强制唯一约束可以被视为一种不应直接访问的实现细节。不过应当知道，无需手工在唯一列上创建索引；那样做只会重复自动创建的索引。

11.7. 表达式索引

索引列不一定非得是底层表中的某一列，
也可以是一个由表中一列或多列计算得到的函数或标量表达式。
这个特性有助于基于计算结果来快速访问表。

例如，进行大小写不敏感比较的一种常见方式，是使用`lower`函数：

```
SELECT * FROM test1 WHERE lower(col1) = 'value';
```

如果为`lower(col1)`函数的结果定义了索引，这个查询就可以使用该索引：

```
CREATE INDEX test1_lower_col1_idx ON test1 (lower(col1));
```

如果我们把这个索引声明为`UNIQUE`，那么它不仅会阻止创建那些`col1`
值仅在大小写上不同的行，也会阻止创建`col1`值完全相同的行。因此，
表达式索引可用于强制那些无法定义为简单唯一约束的约束。

再例如，如果你经常执行如下查询：

```
SELECT * FROM people WHERE (first_name || ' ' || last_name) = 'John
 Smith';
```

那么创建一个如下索引可能是值得的：

```
CREATE INDEX people_names ON people ((first_name || ' ' || last_name)
);
```

正如第二个示例所示，`CREATE INDEX`命令的语法通常要求对索引表达式写上圆括号。当表达式只是一个函数调用时，则可以像第一个示例那样省略这些圆括号。

维护索引表达式的代价相对较高，因为在插入每一行时以及每次更新时都必须计算派生表达式。
不过，在使用索引搜索时，索引表达式不会被重新计算，因为它们已经存储在索引中。
在上面的两个示例里，系统看到的查询只是`WHERE indexedcolumn = 'constant'`，因此搜索速度与其他简单索引查询相同。因此，当检索速度比插入和更新速度更重要时，表达式索引会很有用。

11.8. 部分索引

部分索引是建立在表的一个子集上的索引；
这个子集由一个条件表达式定义（称为部分索引的谓词）。
索引中只包含满足该谓词的表行的项。部分索引是一种专门特性，但有几种场景下它会很有用。

使用部分索引的一个主要原因是避免索引常见值。
由于搜索常见值（即占全部表行百分之几以上的值）的查询反正也不会使用索引，
因此完全没有必要把这些行保留在索引里。这会减小索引尺寸，
从而加快那些确实会使用该索引的查询。它也会加快很多表更新操作，
因为索引并不需要在所有情况下都更新。例 11.1展示了这种思路的一种可能应用。

例 11.1. 建立一个部分索引以排除常见值

假设你把 Web 服务器访问日志存储在数据库中。大多数访问来自你所在组织的 IP 地址范围，但也有一些来自其他地方（例如使用拨号连接的员工）。如果你按 IP 搜索时主要关心外部访问，那么你大概没有必要为对应于组织内子网的 IP 范围建索引。

假设有一个如下表：

```
CREATE TABLE access_log (
    url varchar,
    client_ip inet,
    ...
);
```

要创建一个适合这个例子的部分索引，可使用如下命令：

```
CREATE INDEX access_log_client_ip_ix ON access_log (client_ip)
WHERE NOT (client_ip > inet '192.168.100.0' AND
           client_ip < inet '192.168.100.255');
```

一个能够使用此索引的典型查询如下：

```
SELECT *
FROM access_log
WHERE url = '/index.html' AND client_ip = inet '212.78.10.32';
```

以下查询无法使用此索引：

```
SELECT *
FROM access_log
WHERE client_ip = inet '192.168.100.23';
```

请注意，这类部分索引要求常见值事先可知，因此最适合用于数据分布不会变化的场景。
也可以偶尔重建这类索引以适应新的数据分布，但那会增加维护工作量。

部分索引的另一种可能用途，是把典型查询工作负载不感兴趣的值排除在索引之外，如例 11.2所示。这样会得到与上面相同的好处，但也意味着这些“不感兴趣”的值无法通过该索引访问，
即使在那种情况下索引扫描可能是有利的。显然，
为这种场景建立部分索引需要大量的谨慎和实验。

例 11.2. 建立一个部分索引以排除不感兴趣的值

如果你有一张表，其中同时包含已开账单和未开账单的订单，
而未开账单订单只占整张表的一小部分，却是最常被访问的那些行，
那么你可以通过只为未开账单的行创建索引来提升性能。创建该索引的命令如下：

```
CREATE INDEX orders_unbilled_index ON orders (order_nr)
    WHERE billed is not true;
```

一个可能会使用该索引的查询是：

```
SELECT * FROM orders WHERE billed is not true AND order_nr < 10000;
```

不过，这个索引也可以用于那些完全不涉及`order_nr`的查询，例如：

```
SELECT * FROM orders WHERE billed is not true AND amount > 5000.00;
```

这不像在`amount`列上建立部分索引那样高效，因为系统必须扫描整个索引。不过，
如果未开账单的订单相对较少，那么即使只是用这个部分索引来找出未开账单订单，
也可能是值得的。

请注意，下面这个查询不能使用该索引：

```
SELECT * FROM orders WHERE order_nr = 3501;
```

因为订单 3501 可能属于已开账单订单，也可能属于未开账单订单。

例 11.2也说明了：索引列和谓词中使用的列不必一致。PostgreSQL支持带任意谓词的部分索引，只要其中只涉及正在建立索引的那张表的列。不过要记住，谓词必须与那些希望从该索引受益的查询中使用的条件相匹配。更准确地说，只有当系统能够识别出查询的`WHERE`条件在数学上蕴含该索引的谓词时，部分索引才能用于该查询。PostgreSQL并没有一个复杂的定理证明器，来识别那些写法不同但数学上等价的表达式。（不仅构建这样一个通用定理证明器极其困难，而且它很可能也会慢到失去实际用途。）系统可以识别简单的不等式蕴含，例如“x < 1”蕴含“x < 2”；否则，谓词条件必须与查询`WHERE`条件的某一部分完全匹配，否则索引不会被识别为可用。匹配发生在查询规划阶段，而不是运行时。因此，参数化查询子句无法与部分索引配合工作。例如，一个带参数的预备查询可能写成“x < ?”，而它无法保证在参数的所有可能取值下都蕴含“x < 2”。

部分索引的第三种可能用途，甚至不要求索引被查询使用。这里的思路是像例 11.3那样，在表的一个子集上创建唯一索引。这样就能在满足索引谓词的那些行之间强制唯一性，而不会约束不满足谓词的行。

例 11.3. 建立一个部分唯一索引

假设我们有一张描述测试结果的表。我们希望确保对于给定的测试对象和目标组合，只有一条“成功”记录，但可以有任意多条“失败”记录。实现方法之一如下：

```
CREATE TABLE tests (
    subject text,
    target text,
    success boolean,
    ...
);

CREATE UNIQUE INDEX tests_success_constraint ON tests (subject,
 target)
    WHERE success;
```

当成功测试很少而失败测试很多时，这是一种特别高效的方法。

最后，部分索引还可以用来影响系统的查询计划选择。同样，分布异常的数据集可能导致系统在实际上不应该使用索引的时候选择使用它。在这种情况下，可以把索引建成对那个有问题的查询不可用。通常，PostgreSQL会对索引使用作出合理选择（例如，它会在检索常见值时避免使用索引，因此前面的例子实际上只是节省索引空间，而不是为了避免使用索引所必需的），如果出现明显错误的计划选择，那就应当提交 bug 报告。

请记住，建立部分索引意味着你至少与查询规划器一样了解情况，尤其是你知道索引何时可能是有利的。形成这种认识需要经验，以及对PostgreSQL中索引工作方式的理解。在大多数情况下，部分索引相比普通索引的优势都很小。

关于部分索引的更多信息可参见[ston89b]、[olson93]和[seshadri95]。

11.9. 操作符类和操作符族

索引定义可以为索引的每一列指定一个操作符类。

```
CREATE INDEX name ON table (column opclass [sort options] [, ...]);
```

操作符类确定索引针对该列使用哪些操作符。例如，建立在类型int4上的 B-树索引会使用int4_ops类；该操作符类包含针对以下类型的值的比较函数：int4。实际中，列数据类型的默认操作符类通常就足够了。之所以需要操作符类，主要是因为对于某些数据类型，可能存在不止一种有意义的索引行为。例如，可能希望按绝对值或实部对复数数据类型排序。可以为该数据类型定义两个操作符类，并在创建索引时选择合适的类。操作符类决定基本排序顺序（随后可以通过添加排序选项进行调整：COLLATE，ASC/DESC和/或NULLS FIRST/NULLS LAST）。

除了默认操作符类之外，还有一些内置操作符类：

- 操作符类text_pattern_ops、varchar_pattern_ops和bpchar_pattern_ops分别支持类型text、varchar和char上的 B-树索引。它们与默认操作符类的区别在于，值是严格按字符逐个比较，而不是按照区域设置相关的排序规则比较。因此，当数据库未使用标准“C”区域设置时，这些操作符类适用于涉及模式匹配表达式（LIKE或POSIX 正则表达式）的查询。例如，你可以像这样为一个varchar列建立索引：

  ```
  CREATE INDEX test_index ON test_table (col varchar_pattern_ops);
  ```

  注意，如果你希望涉及普通<、<=、>或>=比较的查询也能使用索引，那么还应该再创建一个使用默认操作符类的索引。这类查询不能使用*xxx*_pattern_ops操作符类。（不过普通等值比较是可以使用这些操作符类的。）同一列上可以创建多个使用不同操作符类的索引。如果你使用的是 C 区域设置，则不需要*xxx*_pattern_ops操作符类，因为带默认操作符类的索引在 C 区域设置下就可以用于模式匹配查询。

下面的查询会显示所有已定义的操作符类：

```
SELECT am.amname AS index_method,
       opc.opcname AS opclass_name
    FROM pg_am am, pg_opclass opc
    WHERE opc.opcmethod = am.oid
    ORDER BY index_method, opclass_name;
```

操作符类其实只是一个更大结构的子集，这个结构称为操作符族。当若干数据类型具有相似行为时，定义跨数据类型操作符并让索引支持它们，往往会很有用。为此，每种类型对应的操作符类都必须归入同一个操作符族。跨类型操作符属于该族，但不与该族中任何单独一个类关联。

下面这个查询会显示所有已定义的操作符族，以及每个族中包含的全部操作符：

```
SELECT am.amname AS index_method,
       opf.opfname AS opfamily_name,
       amop.amopopr::regoperator AS opfamily_operator
    FROM pg_am am, pg_opfamily opf, pg_amop amop
    WHERE opf.opfmethod = am.oid AND
          amop.amopfamily = opf.oid
    ORDER BY index_method, opfamily_name, opfamily_operator;
```

11.10. 索引和排序规则

每个索引列只能支持一种排序规则。如果你需要关心多种排序规则，就可能需要多个索引。

考虑以下语句：

```
CREATE TABLE test1c (
    id integer,
    content varchar COLLATE "x"
);

CREATE INDEX test1c_content_index ON test1c (content);
```

该索引会自动使用底层列的排序规则。因此，下面这种形式的查询：

```
SELECT * FROM test1c WHERE content > constant;
```

可以使用该索引，因为比较默认会使用该列的排序规则。不过，
这个索引不能加速涉及其他排序规则的查询。因此，如果你也关心如下形式的查询：

```
SELECT * FROM test1c WHERE content > constant COLLATE "y";
```

那么可以额外创建一个支持"y"排序规则的索引，例如：

```
CREATE INDEX test1c_content_y_index ON test1c (content COLLATE "y");
```

11.11. 检查索引使用情况

尽管PostgreSQL中的索引不需要维护或调优，
但检查真实查询工作负载到底实际使用了哪些索引，仍然很重要。
检查某个单独查询的索引使用情况，可以使用EXPLAIN命令；它在这方面的应用见第 14.1 节。也可以像第 27.2 节中描述的那样，在运行中的服务器上收集索引使用的总体统计信息。

很难给出一个确定应该创建哪些索引的通用流程。前面各节的示例已经展示了若干典型情形。
通常需要做大量实验。下面这部分会给出一些相关提示：

- 一定要先运行ANALYZE。这个命令会收集表中值分布的统计信息。估计查询会返回多少行，需要这些信息；而规划器要为每种可能的查询计划分配现实的代价，也需要这些信息。
  如果没有真实统计信息，系统就会假定一些默认值，而这些默认值几乎肯定并不准确。因此，在没有运行ANALYZE的情况下去检查应用的索引使用情况，基本上是徒劳的。更多信息见第 23.1.3 节和第 23.1.5 节。

- 用真实数据做实验。使用测试数据来设置索引，只会告诉你测试数据需要什么索引，仅此而已。

  特别要避免使用非常小的测试数据集。虽然在 100000 行里选 1000 行可能适合用索引，但在 100 行里选 1 行几乎不适合，因为这 100 行大概都能放进一个磁盘页，而没有任何计划能比顺序取一个磁盘页更快。

  另外，在编造测试数据时也要小心；如果应用尚未投入生产，这往往无法避免。非常相似、完全随机、或者按排序顺序插入的值，都会让统计信息偏离真实数据应有的分布。

- 当索引没有被使用时，强制使用它们对测试会很有帮助。
  有一些运行时参数可以关闭不同的计划类型（见第 18.7.1 节）。例如，关闭顺序扫描（enable_seqscan）和嵌套循环连接（enable_nestloop）这两种最基本的计划，
  就会迫使系统使用不同的计划。如果系统仍然选择顺序扫描或嵌套循环连接，
  那么索引未被使用很可能有更根本的原因，例如查询条件并不匹配该索引。
  （前面各节已经解释了什么样的查询可以使用什么样的索引。）

- 如果强制使用索引后确实使用了索引，那么就有两种可能：要么系统是对的，
  使用索引确实不合适；要么查询计划的代价估计并未反映现实。因此，

你应该分别在有索引和无索引的情况下对查询计时。`EXPLAIN ANALYZE`命令在这里会很有帮助。

- 如果发现代价估计是错的，那么又有两种可能。
  总代价是根据每个计划节点的逐行代价乘以该计划节点的选择率估计计算出来的。
  计划节点的估计代价可以通过运行时参数来调整（见第 18.7.2 节）。而选择率估计不准确，则是因为统计信息不足。可以尝试通过调整统计信息收集参数来改进这一点（见ALTER TABLE）。

  如果你无法把这些代价调得更合适，那么可能就不得不退而求其次，显式强制使用索引。
  你也可能希望联系PostgreSQL开发者来研究这个问题。

第 12 章 全文搜索

12.1. 介绍

全文检索（或简称文本检索）提供了识别满足查询条件的自然语言文档的能力，并且可按它们与查询的相关度进行排序。最常见的搜索类型，是找出所有包含给定查询词的文档，并按它们与查询的相似性排序返回。查询和相似性的概念都很灵活，取决于具体应用。最简单的搜索把查询视为一组词，把相似性视为查询词在文档中的出现频率。

文本检索操作符在数据库中已经存在很多年了。PostgreSQL为文本数据类型提供了 ~、~*、LIKE 和 ILIKE 操作符，但它们缺少现代信息系统所要求的许多关键特性：

- 缺少语言学支持，即便对英语也是如此。正则表达式并不足够，因为它们难以轻松处理词形变化，例如 satisfies 和 satisfy。搜索 satisfy 时，你大概也希望找到包含 satisfies 的文档，但实际上可能会漏掉它们。虽然可以用 OR 搜索多个派生形式，但这样既繁琐又容易出错（有些词甚至可能有数千种派生形式）。
- 它们不会对搜索结果排序（排名），因此当找到成千上万条匹配文档时就难以实用。
- 它们往往很慢，因为没有索引支持，所以每次搜索都必须处理全部文档。

全文索引允许先对文档进行预处理，并保存索引以供后续快速搜索。预处理包括：

将文档解析成词元。识别出不同类别的词元很有帮助，例如数字、单词、复合词、电子邮件地址等，这样就可以分别处理。原则上，词元类别取决于具体应用，但在大多数场景中，一套预定义的类别已经足够。PostgreSQL使用解析器执行这一步。系统提供了标准解析器，也可以根据特定需求创建自定义解析器。

将词元转换成词位。词位和词元一样都是字符串，但它已经过正规化，使同一个词的不同形式归为一致。例如，正规化几乎总是包含把大写字母折叠为小写，也常常会去掉后缀（如英语中的 s 或 es）。这样，搜索时无需繁琐地输入所有可能的变体，就能匹配同一个词的不同形式。此外，这一步通常还会去除停用词，也就是那些过于常见、对搜索没有帮助的词。（简言之，词元是文档文本的原始片段，而词位则是被认为适合用于索引和搜索的词。）PostgreSQL使用词典执行这一步。系统提供了多种标准词典，也可以按需创建自定义词典。

以适合搜索的形式存储预处理后的文档。例如，每个文档都可以表示为由正规化词位组成的有序数组。除词位之外，通常还希望存储位置信息，以便用于邻近排名；这样，查询词分布更“密集”的文档会比查询词分散出现的文档得到更高的排名。

词典允许对词元的正规化方式进行细粒度控制。借助合适的词典，你可以：

- 定义不应该被索引的停用词。
- 使用 Ispell 把同义词映射为同一个词。
- 使用分类词典把短语映射为同一个词。
- 使用 Ispell 词典把一个词的不同变体映射到一种规范形式。
- 使用 Snowball 词干分析规则把一个词的不同变体映射到一种规范形式。

系统提供了数据类型 tsvector 用于存储预处理后的文档，也提供了数据类型 tsquery 用于表示处理过的查询（第 8.11 节）。围绕这两种数据类型还有许多函数和操作符（第 9.13 节），其中最重要的是匹配操作符 @@，我们将在第 12.1.2 节中介绍。全文检索还可以借助索引加速（第 12.9 节）。

12.1.1. 什么是文档？

文档是全文检索系统中的搜索单位，例如一篇杂志文章或一封电子邮件。
文本检索引擎必须能够解析文档，并保存词位（关键字）与其所属文档之间的关联。随后，
就可以利用这些关联来搜索包含查询词的文档。

在 PostgreSQL 中进行搜索时，文档通常是数据库表某一行中的一个文本字段，或者是这类字段的组合（串接），这些字段可能分布在多个表里，也可能是动态取得的。换句话说，
文档可以由不同部分拼接而成以供索引，而且未必会以一个整体存储在任何地方。例如：

```
SELECT title || ' ' ||  author || ' ' ||  abstract || ' ' || body AS
 document
FROM messages
WHERE mid = 12;

SELECT m.title || ' ' || m.author || ' ' || m.abstract || ' ' || d.
body AS document
FROM messages m, docs d
WHERE mid = did AND mid = 12;
```

注意

实际上，这些示例查询中应使用 `coalesce`，以避免某个单独的 `NULL` 属性导致整个文档结果变成 `NULL`。

另一种做法是把文档存储为文件系统中的普通文本文件。在这种情况下，
数据库可以用于保存全文索引并执行搜索，
而文档本身则借助某个唯一标识符从文件系统中取回。不过，
从数据库外部检索文件需要超级用户权限或特殊函数支持，
因此这种方法通常不如把所有数据都保存在 PostgreSQL 内部方便。此外，把所有内容都放在数据库中，也便于访问文档元数据来辅助索引和展示。

为了进行文本检索，每个文档都必须被化简为预处理后的 `tsvector` 格式。搜索和排名完全基于文档的 `tsvector` 表示来执行 – 只有当文档被选中并准备展示给用户时，才需要取回原始文本。因此，我们常常把 `tsvector` 直接称作文档，但它当然只是完整文档的一种紧凑表示。

12.1.2. 基本文本匹配

在PostgreSQL中，全文检索基于匹配操作符`@@`。如果一个`tsvector`（文档）
匹配一个`tsquery`（查询），它就返回`true`。哪一种数据类型写在前面并不重要：

```
SELECT 'a fat cat sat on a mat and ate a fat rat'::tsvector @@ 'cat &
 rat'::tsquery;
 ?column?
----------
 t

SELECT 'fat & cow'::tsquery @@ 'a fat cat sat on a mat and ate a fat
 rat'::tsvector;
 ?column?
----------
 f
```

正如上例所示，`tsquery`并不只是原始文本，`tsvector`也不是。`tsquery`包含搜索词，
这些搜索词必须已经是正规化后的词位，并且可以用 AND、OR 和 NOT 操作符把多个搜索词组合

起来。（详情见第 8.11 节。）to_tsquery和plainto_tsquery有助于把用户输入的文本转换为合适的 tsquery，例如对文本中的词做正规化。类似地，to_tsvector 用于解析并正规化文档字符串。因此在实践中，文本检索匹配更像是这样：

```
SELECT to_tsvector('fat cats ate fat rats') @@ to_tsquery('fat & rat')
;
 ?column?
----------
 t
```

请注意，如果写成

```
SELECT 'fat cats ate fat rats'::tsvector @@ to_tsquery('fat & rat');
 ?column?
----------
 f
```

就不会匹配成功，因为这里不会对单词 rats 做正规化。tsvector 的元素是词位，被假定为已经正规化，因此 rats 不会匹配 rat。

@@操作符也支持text输入，因此在简单场景下可以跳过把文本字符串显式转换为tsvector或tsquery。可用的形式有：

```
tsvector @@ tsquery
tsquery  @@ tsvector
text @@ tsquery
text @@ text
```

前两种形式我们已经见过。text @@ tsquery 等价于 to_tsvector(x) @@ y。text @@ text 等价于 to_tsvector(x) @@ plainto_tsquery(y)。

12.1.3. 配置

前述的都是简单的文本检索示例。正如前面所提到的，
全文检索功能包括做更多事情的能力：跳过索引特定词（停用词）、
处理同义词并使用更高级的解析，例如不局限于按空白划分的解析。这个功能由
文本检索配置控制。PostgreSQL中有多种语言的预定义配置，
并且你可以很容易地创建你自己的配置（psql的\dF命令显示所有可用的配置）。

在安装期间会选择一个合适的配置，并据此在postgresql.conf中设置default_text_search_config。如果整个集簇都使用同一种文本检索配置，你可以直接使用postgresql.conf中的这个值。若要在整个集簇中使用不同配置，但保证每个数据库内部使用同一种配置，可以使用ALTER DATABASE ... SET。否则，你也可以在每个会话中设置default_text_search_config。

依赖一个配置的每一个文本检索函数都有一个可选的regconfig参数，
因此要使用的配置可以被显式指定。只有当省略这个参数时，default_text_search_config才被使用。

为了便于构建自定义文本检索配置，配置由更简单的数据库对象组合而成。PostgreSQL 的文本检索功能提供了四类与配置相关的数据库对象：

- 文本检索解析器将文档拆分成词元并对每个词元分类（例如，作为词或者数字）。
- 文本检索词典将词元转变成正规化的形式并拒绝停用词。

- 文本检索模板提供位于词典底层的函数（词典只需指定一个模板及其一组参数）。
- 文本检索配置选择一个解析器和一组用于将解析器产生的词元正规化的词典。

文本检索解析器和模板由底层 C 函数构成，因此开发新的解析器和模板需要 C 编程能力，将它们安装到数据库中则需要超级用户权限。（PostgreSQL 发行版的 contrib/ 目录中有附加解析器和模板的示例。）词典和配置只是为底层解析器和模板设置参数并将它们连接起来，因此创建新词典或配置不需要特殊权限。
本章后面会给出创建自定义词典和配置的示例。

12.2. 表和索引

在前一节中的示例演示了使用简单常量字符串进行全文匹配。本节展示如何搜索表数据，
以及可选择地使用索引。

12.2.1. 搜索表

即使没有索引，也可以执行全文检索。下面这个简单查询会输出每一行的title，其中对应的body字段包含单词friend：

```
SELECT title
FROM pgweb
WHERE to_tsvector('english', body) @@ to_tsquery('english', 'friend');
```

这还会找到相关词，例如friends和friendly，因为这些词都会被约简为同一个正规化词位。

上述查询指定使用 english 配置来解析并正规化字符串。我们也可以省略配置参数：

```
SELECT title
FROM pgweb
WHERE to_tsvector(body) @@ to_tsquery('friend');
```

该查询将使用由default_text_search_config设置的配置。

更复杂一点的例子，是选出最近的 10 个文档，它们的 title 和 body 合在一起包含 create 和 table：

```
SELECT title
FROM pgweb
WHERE to_tsvector(title || ' ' || body) @@ to_tsquery('create &
 table')
ORDER BY last_mod_date DESC
LIMIT 10;
```

为简洁起见，这里省略了 coalesce 调用；如果希望在这两个字段之一为 NULL 时，另一字段仍能参与搜索，就需要加上它。

虽然这些查询在没有索引的情况下也能工作，但除偶尔的临时搜索外，
大多数应用都会觉得这种方式太慢。文本检索在实际使用中通常都需要建立索引。

12.2.2. 创建索引

我们可以创建一个GIN索引（第 12.9 节）来加速文本检索：

```
CREATE INDEX pgweb_idx ON pgweb USING gin(to_tsvector('english', body)
);
```

注意这里使用的是to_tsvector的双参数版本。只有显式指定配置名称的文本检索函数，才能用于表达式索引（第 11.7 节）。这是因为索引内容必须不受default_text_search_config的影响。否则，索引内容就可能不一致，因为不同的索引项可能包含使用不同文本检索配置创建的tsvector，而且无法判断各自使用了哪一种配置。这样的索引也不可能被正确地转储和恢复。

由于上面的索引使用了 to_tsvector 的双参数版本，因此只有同样使用相同配置名的双参数版 to_tsvector 查询，才能使用该索引。也就是说，WHERE to_tsvector('english', body) @@ 'a & b' 可以使用该索引，而 WHERE to_tsvector(body) @@ 'a & b' 则不能。这样可以保证索引只会和创建索引项时所用的同一配置配合使用。

还可以建立更复杂的表达式索引，其中配置名由另一个列指定，例如：

```
CREATE INDEX pgweb_idx ON pgweb USING gin(to_tsvector(config_name,
 body));
```

这里config_name是pgweb表中的一个列。这样就允许在同一个索引中混合使用不同配置，同时记录每个索引项使用的是哪一种配置。例如，如果文档集合中包含不同语言的文档，这就会很有用。同样，打算使用该索引的查询也必须写成对应的形式，例如WHERE to_tsvector(config_name, body) @@ 'a & b'。

索引甚至可以串接多个列：

```
CREATE INDEX pgweb_idx ON pgweb USING gin(to_tsvector('english', title
 || ' ' || body));
```

另一种方法是创建一个单独的tsvector列来保存to_tsvector的输出。下面的示例把title和body串接起来，并用coalesce保证一个字段仍然可以被建立索引，即使另一个字段为NULL：

```
ALTER TABLE pgweb ADD COLUMN textsearchable_index_col tsvector;
UPDATE pgweb SET textsearchable_index_col =
     to_tsvector('english', coalesce(title,'') || ' ' || coalesce(
body,''));
```

然后我们创建一个GIN索引来加速搜索：

```
CREATE INDEX textsearch_idx ON pgweb USING gin(textsearchable_index_
col);
```

现在可以执行快速全文检索了：

```
SELECT title
FROM pgweb
WHERE textsearchable_index_col @@ to_tsquery('create & table')
ORDER BY last_mod_date DESC
LIMIT 10;
```

当使用一个单独的列来存储 tsvector 表示时，需要创建一个触发器来使 tsvector 列保持最新，以应对 title 或 body 的任何更改。第 12.4.3 节说明了如何做到这一点。

与表达式索引相比，单独列方法的一个优点是，为了利用索引，查询中不必显式指定文本检索配置。正如上面的例子所示，查询可以依赖default_text_

search_config。另一个优点是搜索会更快，因为它不必重新执行to_tsvector调用来验证索引匹配（使用 GiST 索引时这一点比使用 GIN 索引时更重要；见第 12.9 节）。不过，表达式索引方法更容易设置，而且占用更少磁盘空间，因为tsvector表示并没有被显式存储。

12.3. 控制文本搜索

要实现全文检索，必须有函数能够从文档创建 tsvector，并从用户查询创建 tsquery。此外，我们还希望结果能按有意义的顺序返回，因此还需要函数根据文档与查询的相关性进行比较。同样重要的，是把结果良好地展示出来。PostgreSQL为这些能力都提供了支持。

12.3.1. 解析文档

PostgreSQL 提供了 to_tsvector 函数，用于把文档转换成 tsvector 数据类型。

```
to_tsvector([ config regconfig, ] document text) returns tsvector
```

to_tsvector 会把文本文档解析为词元，将词元归约为词位，并返回一个 tsvector，其中列出各词位及其在文档中的位置。文档会按指定的或默认的文本检索配置进行处理。
下面是一个简单示例：

```
SELECT to_tsvector('english', 'a fat  cat sat on a mat - it ate a fat
 rats');
                   to_tsvector
-----------------------------------------------------
 'ate':9 'cat':3 'fat':2,11 'mat':7 'rat':12 'sat':4
```

在上面的示例中可以看到，结果 tsvector 不包含 a、on 和 it，单词 rats 变成了 rat，标点符号 - 也被忽略了。

to_tsvector 在内部会调用解析器，把文档文本拆分成词元并为每个词元分配类型。对于每个词元，系统都会查询一个词典列表（第 12.6 节），而这个列表会随词元类型而变化。第一个能够识别该词元的词典，会输出一个或多个正规化后的词位来表示它。例如，rats 之所以变成 rat，是因为某个词典识别出 rats 是 rat 的复数形式。某些词会被识别为停用词（第 12.6.1 节），于是被忽略，因为它们出现得过于频繁，对搜索没有帮助。在这个示例中，a、on 和 it 就是停用词。如果列表中的词典都无法识别某个词元，它也会被忽略。示例里的标点符号 - 就属于这种情况，因为其词元类型（Space symbols）实际上没有分配任何词典，也就是说，空白类词元永远不会被索引。解析器、词典以及需要索引哪些词元类型，都由所选的文本检索配置（第 12.7 节）决定。一个数据库里可以同时存在多种不同配置，并且系统已经为多种语言提供了预定义配置。在本例中，我们使用的是英语的默认配置 english。

setweight 函数可用于给 tsvector 中的项打上指定的权重标签，权重可以是四个字母之一：A、B、C 或 D。这通常用于标记来自文档不同部分的项，例如标题和正文。稍后，这些信息可以用于搜索结果排名。

由于 to_tsvector(NULL) 会返回 NULL，因此只要字段可能为 NULL，就建议使用 coalesce。下面是从结构化文档创建 tsvector 的推荐方法：

```
UPDATE tt SET ti =
    setweight(to_tsvector(coalesce(title,'')), 'A')    ||
    setweight(to_tsvector(coalesce(keyword,'')), 'B')  ||
    setweight(to_tsvector(coalesce(abstract,'')), 'C') ||
```

```
    setweight(to_tsvector(coalesce(body,'')), 'D');
```

这里我们使用 `setweight` 为最终 `tsvector` 中每个词位标注来源，然后再用 `tsvector` 连接操作符 `||` 把这些已标注的 `tsvector` 值合并起来（关于这些操作的细节，见第 12.4.1 节）。

12.3.2. 解析查询

PostgreSQL 提供了 `to_tsquery`和 `plainto_tsquery`，用于把查询转换成 `tsquery` 数据类型。`to_tsquery` 提供的特性比 `plainto_tsquery` 更丰富，但对输入也更严格。

```
to_tsquery([ config regconfig, ] querytext text) returns tsquery
```

`to_tsquery` 根据 *querytext* 创建一个 `tsquery` 值。输入必须由单个词元组成，这些词元之间使用布尔操作符 `&`（AND）、`|`（OR）和 `!`（NOT）分隔。这些操作符可使用圆括号分组。换句话说，传给 `to_tsquery` 的输入，必须已经遵循第 8.11 节中描述的 `tsquery` 一般输入规则。不同之处在于，基础 `tsquery` 输入是按字面接受词元，而 `to_tsquery` 会使用指定或默认配置，把每个词元正规化为词位，并丢弃那些按该配置被判定为停用词的词元。例如：

```
SELECT to_tsquery('english', 'The & Fat & Rats');
  to_tsquery
---------------
 'fat' & 'rat'
```

与基础`tsquery`输入一样，可以为每个词位附加权重，以限制它只匹配具有这些权重的`tsvector`词位。例如：

```
SELECT to_tsquery('english', 'Fat | Rats:AB');
    to_tsquery
------------------
 'fat' | 'rat':AB
```

此外，还可以在词位后附加`*`来指定前缀匹配：

```
SELECT to_tsquery('supern:*A & star:A*B');
        to_tsquery
--------------------------
 'supern':*A & 'star':*AB
```

这样的词位将匹配`tsvector`中任何以给定字符串开头的单词。

`to_tsquery`也可以接受单引号括起来的短语。
当配置中包含可能在这类短语上触发的分类词典时，这一点尤其有用。在下面的例子中，一个分类词典包含规则 `supernovae stars : sn`：

```
SELECT to_tsquery('''supernovae stars'' & !crab');
  to_tsquery
---------------
 'sn' & !'crab'
```

如果不加引号，对于那些未被 AND 或 OR 操作符分隔的词元，`to_tsquery`会产生语法错误。

```
plainto_tsquery([ config regconfig, ] querytext text) returns tsquery
```

plainto_tsquery把未格式化的文本*querytext*转换成tsquery。该文本会像to_tsvector那样被解析并正规化，然后在保留下来的词之间插入&（AND）布尔操作符。

示例：

```
SELECT plainto_tsquery('english', 'The Fat Rats');
 plainto_tsquery
-----------------
 'fat' & 'rat'
```

注意，plainto_tsquery不能识别其输入中的布尔操作符、权重标签或前缀匹配标签：

```
SELECT plainto_tsquery('english', 'The Fat & Rats:C');
   plainto_tsquery
---------------------
 'fat' & 'rat' & 'c'
```

这里，输入中的所有标点符号都被当作空白符丢弃。

12.3.3. 搜索结果排名

排名旨在衡量文档与特定查询的相关程度，以便在匹配很多时优先显示最相关的结果。PostgreSQL提供了两种预定义的排名函数，它们会综合考虑词法信息、邻近关系和结构信息；也就是说，会考虑查询词在文档中出现的频率、这些词彼此之间的距离，以及这些词所处文档部分的重要程度。不过，相关性这一概念本身就比较模糊，而且高度依赖具体应用。不同应用可能还需要额外信息参与排名，例如文档修改时间。内置排名函数仅仅是示例。你可以编写自己的排名函数，或者把它们的结果与其他因素结合起来，以满足特定需求。

目前可用的两种排名函数是：

```
ts_rank([ weights float4[], ] vector tsvector, query tsquery
 [, normalization integer ]) returns float4
```

根据向量中匹配词位的频率对向量进行排名。

```
ts_rank_cd([ weights float4[], ] vector tsvector,
 query tsquery [, normalization integer ]) returns float4
```

该函数为给定的文档向量和查询计算覆盖密度排名。该方法见 Clarke、Cormack 和 Tudhope 于 1999 年发表于期刊 "Information Processing and Management" 的文章 "Relevance Ranking for One to Three Term Queries"。

该函数需要输入中包含位置信息。因此，它对“被剥离的”tsvector值不起作用 — 结果总是为零。

对这两个函数来说，可选的 *weights* 参数允许根据词实例的标注情况赋予它们不同权重。权重数组按如下顺序指定各类词的权重：

{D-权重，C-权重，B-权重，A-权重}

如果没有提供 *weights*，则使用如下默认值：

{0.1, 0.2, 0.4, 1.0}

通常，权重用于标记来自文档特定区域的词，例如标题或开头摘要中的词，
从而使它们相较于正文中的词被赋予更高或更低的重要性。

由于较长的文档更有机会包含查询词，因此把文档大小纳入考量是合理的。例如，
一个一百词的文档里某个搜索词出现五次，
通常会比一个一千词的文档里同一搜索词也只出现五次更相关。两种排名函数都接受一个整数 *normalization* 选项，用于指定文档长度是否影响排名，以及具体如何影响。该整数选项控制多种行为，因此它是一个位掩码：你可以使用 `|` 指定一种或多种行为（例如 `2|4`）。

- 0（默认值）忽略文档长度
- 1 用 1 + 文档长度的对数除排名
- 2 用文档长度除排名
- 4 用匹配范围之间的平均调和距离除排名（仅由 `ts_rank_cd` 实现）
- 8 用文档中唯一词的数量除排名
- 16 用 1 + 文档中不同词数量的对数除排名
- 32 用排名 + 1 除排名

如果指定了多个标志位，这些变换将按上面列出的顺序依次应用。

需要注意的是，排名函数不会使用任何全局信息，因此不可能像有时人们希望的那样，
给出一种能够公平归一化到 1% 或 100% 的结果。归一化选项 32（`rank/(rank+1)`）可以把所有排名缩放到 0 到 1 的范围内，但这当然只是表面上的变化，不会影响搜索结果的顺序。

下面是只选择排名最高的十个匹配的示例：

```
SELECT title, ts_rank_cd(textsearch, query) AS rank
FROM apod, to_tsquery('neutrino|(dark & matter)') query
WHERE query @@ textsearch
ORDER BY rank DESC
LIMIT 10;
                     title                     |   rank
-----------------------------------------------+----------
 Neutrinos in the Sun                          |      3.1
 The Sudbury Neutrino Detector                 |      2.4
 A MACHO View of Galactic Dark Matter          |  2.01317
 Hot Gas and Dark Matter                       |  1.91171
 The Virgo Cluster: Hot Plasma and Dark Matter |  1.90953
 Rafting for Solar Neutrinos                   |      1.9
 NGC 4650A: Strange Galaxy and Dark Matter     |  1.85774
 Hot Gas and Dark Matter                       |   1.6123
 Ice Fishing for Cosmic Neutrinos              |      1.6
 Weak Lensing Distorts the Universe            | 0.818218
```

下面是使用归一化排名的同一示例：

```
SELECT title, ts_rank_cd(textsearch, query, 32 /* rank/(rank+1) */ )
 AS rank
FROM apod, to_tsquery('neutrino|(dark & matter)') query
WHERE  query @@ textsearch
ORDER BY rank DESC
LIMIT 10;
                     title                     |        rank
-----------------------------------------------+--------------------
 Neutrinos in the Sun                          | 0.756097569485493
 The Sudbury Neutrino Detector                 | 0.705882361190954
```

```
 A MACHO View of Galactic Dark Matter          | 0.668123210574724
 Hot Gas and Dark Matter                       |  0.65655958650282
 The Virgo Cluster: Hot Plasma and Dark Matter | 0.656301290640973
 Rafting for Solar Neutrinos                   | 0.655172410958162
 NGC 4650A: Strange Galaxy and Dark Matter     | 0.650072921219637
 Hot Gas and Dark Matter                       | 0.617195790024749
 Ice Fishing for Cosmic Neutrinos              | 0.615384618911517
 Weak Lensing Distorts the Universe            | 0.450010798361481
```

排名计算可能非常昂贵，因为它需要访问每个匹配文档的 `tsvector`，这往往会受 I/O 限制而变慢。不幸的是，这几乎无法避免，因为实际查询常常会产生大量匹配。

12.3.4. 高亮结果

在呈现搜索结果时，理想的方式是展示每个文档的一段摘录，并说明它与查询的关系。通常，搜索引擎会在文档片段中标出搜索词。PostgreSQL 提供了 `ts_headline` 函数来实现这一功能。

```
ts_headline([ config regconfig, ] document text, query tsquery [,
 options text ]) returns text
```

`ts_headline` 接收文档和查询，并返回文档中一段高亮查询词的摘录。用于解析文档的配置可通过 *`config`* 指定；若省略 *`config`*，则使用 `default_text_search_config` 配置。

如果指定了*`options`*字符串，它必须由一个或多个用逗号分隔的*`option=value`*对组成。可用选项有：

- `StartSel`、`StopSel`：用于界定文档中查询词的字符串，以便与摘录中的其他词区分。如果这些字符串包含空格或逗号，则必须用双引号括起来。
- `MaxWords`、`MinWords`：这两个数值决定输出摘要的最长和最短长度。
- `ShortWord`：长度不超过该值的词就会从摘要的开头和结尾处被丢弃。默认值 3 会去掉常见的英语冠词。
- `HighlightAll`：布尔标志；如果为 `true`，则整个文档都会被用作摘要，而忽略前面三个参数。
- `MaxFragments`：要显示的最大文本摘录或片段数。默认值 0 选择一种非面向片段的摘要生成方法。大于 0 的值选择基于片段的摘要生成方法。这种方法会寻找包含尽可能多查询词的文本片段，并把这些片段围绕查询词扩展。这样，查询词就接近每个片段的中间，两侧都有词语。每个片段至多包含 `MaxWords` 个词，并且长度不超过 `ShortWord` 的词会在每个片段的开头和结尾处被丢弃。如果并非所有查询词都在文档中找到，则显示由文档前 `MinWords` 个词组成的单个片段。
- `FragmentDelimiter`：当显示多个片段时，片段之间将以该字符串分隔。

未指定的选项将采用以下默认值：

```
StartSel=<b>, StopSel=</b>,
MaxWords=35, MinWords=15, ShortWord=3, HighlightAll=FALSE,
MaxFragments=0, FragmentDelimiter=" ... "
```

例如：

```
SELECT ts_headline('english',
  'The most common type of search
is to find all documents containing given query terms
```

```
and return them in order of their similarity to the
query.',
  to_tsquery('query & similarity'));
                        ts_headline
------------------------------------------------------------
 containing given <b>query</b> terms
 and return them in order of their <b>similarity</b> to the
 <b>query</b>.

SELECT ts_headline('english',
  'The most common type of search
is to find all documents containing given query terms
and return them in order of their similarity to the
query.',
  to_tsquery('query & similarity'),
  'StartSel = <, StopSel = >');
                      ts_headline
-------------------------------------------------------
 containing given <query> terms
 and return them in order of their <similarity> to the
 <query>.
```

`ts_headline` 使用的是原始文档，而不是 `tsvector` 摘要，因此它可能较慢，使用时应当谨慎。一个典型的错误是，当只需要显示十个文档时，却对每一个匹配文档都调用 `ts_headline`。SQL 子查询可以提供帮助，下面是一个例子：

```
SELECT id, ts_headline(body, q), rank
FROM (SELECT id, body, q, ts_rank_cd(ti, q) AS rank
      FROM apod, to_tsquery('stars') q
      WHERE ti @@ q
      ORDER BY rank DESC
      LIMIT 10) AS foo;
```

12.4. 附加特性

本节介绍一些在文本检索中很有用的附加函数和操作符。

12.4.1. 操纵文档

第 12.3.1 节展示了原始文本文档如何被转换成 `tsvector` 值。PostgreSQL 也提供了用于操纵已经处于 `tsvector` 形式文档的函数和操作符。

```
tsvector || tsvector
```

`tsvector` 连接操作符返回一个向量，它结合了两个参数向量中的词位和位置信息。位置和权重标签在连接过程中会被保留。
右侧向量中的位置会按左侧向量中出现的最大位置做偏移，
因此结果几乎等同于对两个原始文档字符串连接后的结果执行 `to_tsvector`。（这种等价并不完全成立，因为从左侧参数末尾移除的停用词不会影响结果；而如果直接做文本连接，它们本来会影响右侧参数中词位的位置。）

相比先连接文本再应用 `to_tsvector`，直接连接向量的一个优点是，你可以用不同配置来解析文档的不同部分。此外，由于 `setweight` 会以相同方式标记给定向量中的全部词位，

如果希望给文档不同部分打上不同权重，就必须先解析文本并调用 `setweight`，然后再做连接。

`setweight(vector tsvector, weight "char") returns tsvector`

`setweight` 返回输入向量的一个副本，其中每个位置都被标注为给定的 *weight*：A、B、C 或 D（D 是新向量的默认值，因此不会在输出中显示）。这些标签在向量连接时会被保留，从而允许排名函数对来自文档不同部分的词赋予不同权重。

注意权重标签是应用到位置而不是词位。如果输入向量已经被剥离了位置，则`setweight`什么也不会做。

`length(vector tsvector) returns integer`

返回存储在向量中的词位数。

`strip(vector tsvector) returns tsvector`

返回一个向量，其中列出与给定向量相同的词位，但不包含任何位置或权重信息。虽然返回的向量在相关度排名方面远不如未剥离的向量有用，但通常会比它小得多。

12.4.2. 操纵查询

第 12.3.2 节展示了原始文本查询如何被转换成 `tsquery` 值。PostgreSQL 也提供了用于操纵已经处于 `tsquery` 形式查询的函数和操作符。

`tsquery && tsquery`

返回用 AND 结合的两个给定查询。

`tsquery || tsquery`

返回用 OR 结合的两个给定查询。

`!! tsquery`

返回给定查询的否定（NOT）。

`numnode(query tsquery) returns integer`

返回`tsquery`中的节点数（词位加操作符）。这个函数可用于判断*query*是否有意义（返回值 > 0），或者是否只包含停用词（返回 0）。例如：

```
SELECT numnode(plainto_tsquery('the any'));
NOTICE:  query contains only stopword(s) or doesn't contain lexeme(
s), ignored
 numnode
---------
       0

SELECT numnode('foo & bar'::tsquery);
 numnode
---------
       3
```

querytree(*query* tsquery) returns text

返回一个tsquery中可用于搜索索引的部分。此函数可用于检测无法使用索引的查询，例如只包含停用词或只包含否定词项的查询。例如：

```
SELECT querytree(to_tsquery('!defined'));
 querytree
-----------
```

12.4.2.1. 查询重写

ts_rewrite 函数族会在给定的 tsquery 中搜索目标子查询的出现，并把每一次出现都替换为替换子查询。本质上，这就是面向 tsquery 的一种子串替换。目标与替换的组合可以看作一条查询重写规则。一组这样的重写规则可以成为非常强大的搜索辅助工具。例如，你可以利用同义词扩展搜索（如 new york、big apple、nyc、gotham），或者收窄搜索，把用户引向某个热门主题。该特性与分类词典（第 12.6.4 节）在功能上有一定重叠。不过，重写规则可以随时修改而无需重建索引，而更新分类词典则必须重建索引后才能生效。

ts_rewrite (*query* tsquery, *target* tsquery, *substitute* tsquery) returns tsquery

这种形式的 ts_rewrite 只应用一条重写规则：无论 *target* 在 *query* 中的何处出现，都用 *substitute* 替换它。例如：

```
SELECT ts_rewrite('a & b'::tsquery, 'a'::tsquery, 'c'::tsquery);
 ts_rewrite
------------
 'b' & 'c'
```

ts_rewrite (*query* tsquery, *select* text) returns tsquery

这种形式的 ts_rewrite 接受一个起始 *query* 和一个以文本字符串给出的 SQL *select* 命令。该 *select* 必须产生两列 tsquery 类型的值。对于 *select* 结果中的每一行，第一列值（目标）在当前 *query* 值中的所有出现都会被第二列值（替换）取代。例如：

```
CREATE TABLE aliases (t tsquery PRIMARY KEY, s tsquery);
INSERT INTO aliases VALUES('a', 'c');

SELECT ts_rewrite('a & b'::tsquery, 'SELECT t,s FROM aliases');
 ts_rewrite
------------
 'b' & 'c'
```

注意，当以这种方式应用多个重写规则时，应用顺序可能很重要；因此在实际使用中，你通常会要求源查询按某个排序键 ORDER BY。

我们来看一个现实中的天文示例。我们将使用表驱动的重写规则来扩展查询 supernovae：

```
CREATE TABLE aliases (t tsquery primary key, s tsquery);
INSERT INTO aliases VALUES(to_tsquery('supernovae'), to_tsquery(
'supernovae|sn'));

SELECT ts_rewrite(to_tsquery('supernovae & crab'), 'SELECT * FROM
 aliases');
           ts_rewrite
```

```
----------------------------------
 'crab' & ( 'supernova' | 'sn' )
```

我们只需更新表，就可以修改这些重写规则：

```
UPDATE aliases
SET s = to_tsquery('supernovae|sn & !nebulae')
WHERE t = to_tsquery('supernovae');

SELECT ts_rewrite(to_tsquery('supernovae & crab'), 'SELECT * FROM
 aliases');
                 ts_rewrite
---------------------------------------------
 'crab' & ( 'supernova' | 'sn' & !'nebula' )
```

当重写规则很多时，重写过程可能会很慢，因为它需要检查每一条规则是否可能匹配。
为了过滤掉显然不可能命中的规则，可以使用 `tsquery` 类型的包含操作符。在下面的例子中，我们只选择那些可能匹配原始查询的规则：

```
SELECT ts_rewrite('a & b'::tsquery,
                  'SELECT t,s FROM aliases WHERE ''a & b''::tsquery @>
 t');
 ts_rewrite
------------
 'b' & 'c'
```

12.4.3. 用于自动更新的触发器

当使用单独一列来存储文档的 `tsvector` 表示时，就需要创建触发器，以便在文档内容列发生变化时更新该 `tsvector` 列。系统为此提供了两个内置触发器函数，你也可以自行编写触发器函数。

```
tsvector_update_trigger(tsvector_column_name, config_name, text_
column_name [, ... ])
tsvector_update_trigger_column(tsvector_column_name, config_column_
name, text_column_name [, ... ])
```

这些触发器函数会在 `CREATE TRIGGER` 命令中给定参数的控制下，从一个或多个文本列自动计算出一个 `tsvector` 列。下面是它们的一个使用示例：

```
CREATE TABLE messages (
    title       text,
    body        text,
    tsv         tsvector
);

CREATE TRIGGER tsvectorupdate BEFORE INSERT OR UPDATE
ON messages FOR EACH ROW EXECUTE PROCEDURE
tsvector_update_trigger(tsv, 'pg_catalog.english', title, body);

INSERT INTO messages VALUES('title here', 'the body text is here');

SELECT * FROM messages;
   title    |         body          |            tsv
```

```
------------+-----------------------+----------------------------
 title here | the body text is here | 'bodi':4 'text':5 'titl':1

SELECT title, body FROM messages WHERE tsv @@ to_tsquery('title &
 body');
   title    |         body
------------+-----------------------
 title here | the body text is here
```

创建该触发器之后，对title或body的任何修改都会自动反映到tsv中，
而应用程序无需关心这一点。

第一个触发器参数必须是要更新的 tsvector 列名。第二个参数指定执行转换时要使用的文本检索配置。对于 tsvector_update_trigger，配置名直接作为第二个触发器参数给出。如上所示，它必须带模式限定，这样触发器行为就不会随着 search_path 的变化而变化。对于 tsvector_update_trigger_column，第二个触发器参数则是另一个表列的名称，该列必须是 regconfig 类型。这样就可以按行选择配置。其余参数是文本列的名称（类型为 text、varchar 或 char），它们会按给定顺序并入文档。NULL 值会被跳过（但其他列仍会被索引）。

这些内置触发器的一个限制是，它们会以相同方式处理所有输入列。若要对列做不同处理 — 例如让标题和正文具有不同权重 — 就需要编写自定义触发器。下面是一个使用 PL/pgSQL 作为触发器语言的示例：

```
CREATE FUNCTION messages_trigger() RETURNS trigger AS $$
begin
  new.tsv :=
     setweight(to_tsvector('pg_catalog.english', coalesce(new.title,
'')), 'A') ||
     setweight(to_tsvector('pg_catalog.english', coalesce(new.body,'')
), 'D');
  return new;
end
$$ LANGUAGE plpgsql;

CREATE TRIGGER tsvectorupdate BEFORE INSERT OR UPDATE
    ON messages FOR EACH ROW EXECUTE PROCEDURE messages_trigger();
```

请记住，在触发器中创建 tsvector 值时，明确指定配置名称非常重要，这样列的内容才不会受到 default_text_search_config 变化的影响。否则，很可能会导致问题，例如在转储并恢复之后搜索结果发生变化。

12.4.4. 收集文档统计数据

ts_stat 可用于检查配置，并寻找候选停用词。

```
ts_stat(sqlquery text, [ weights text, ]
        OUT word text, OUT ndoc integer,
        OUT nentry integer) returns setof record
```

sqlquery 是一个文本值，其中包含一条必须返回单一 tsvector 列的 SQL 查询。ts_stat 会执行该查询，并返回该 tsvector 数据中每个不同词位（单词）的统计信息。返回的列如下：

- *word* text — 一个词位的值

- *ndoc* integer – 词出现过的文档（tsvector）的数量
- *nentry* integer – 词出现的总次数

如果提供了 *weights*，则只统计具有这些权重之一的出现。

例如，要在一个文档集合中查找十个最频繁的词：

```
SELECT * FROM ts_stat('SELECT vector FROM apod')
ORDER BY nentry DESC, ndoc DESC, word
LIMIT 10;
```

同样的查询，但只统计权重为 A 或 B 的出现次数：

```
SELECT * FROM ts_stat('SELECT vector FROM apod', 'ab')
ORDER BY nentry DESC, ndoc DESC, word
LIMIT 10;
```

12.5. 解析器

文本检索解析器负责把未处理的文档文本划分成词元并标识每个词元的类型，
而可能的类型集合由解析器本身定义。注意，解析器完全不会修改文本 – 它只是识别看似合理的词边界。由于作用范围有限，相比自定义词典，
对应用相关的自定义解析器的需求没有那么强烈。目前PostgreSQL只提供一种内置解析器，
而它已经被证明对广泛的应用都很有用。

内置解析器被称为pg_catalog.default。它识别 23 种词元类型，如表 12.1所示。

表 12.1. 默认解析器的词元类型

别名	描述	示例
asciiword	单词，所有 ASCII 字母	elephant
word	单词，所有字母	mañana
numword	单词，字母和数字	beta1
asciihword	带连字符的单词，所有 ASCII	up-to-date
hword	带连字符的单词，所有字母	lógico-matemática
numhword	带连字符的单词，字母和数字	postgresql-beta1
hword_asciipart	带连字符的单词部分，所有 ASCII	在 postgresql-beta1 上下文中的 postgresql
hword_part	带连字符的单词部分，所有字母	在 lógico-matemática 上下文中的 lógico 或 matemática
hword_numpart	带连字符的单词部分，字母和数字	在 postgresql-beta1 上下文中的 beta1
email	电子邮件地址	foo@example.com
protocol	协议头部	http://
url	URL	example.com/stuff/index.html
host	主机	example.com
url_path	URL 路径	/stuff/index.html，在 URL 上下文中

别名	描述	示例
file	文件或路径名	/usr/local/foo.txt，如果不在 URL 中
sfloat	科学记数法	-1.234e56
float	十进制记数法	-1.234
int	有符号整数	-1234
uint	无符号整数	1234
version	版本号	8.3.0
tag	XML 标签	<a href="dictionaries.html">
entity	XML 实体	&
blank	空格符号	（其他不识别的任意空白或标点符号）

注意

解析器对“字母”的界定由数据库的区域设置决定，具体是lc_ctype。只包含基本 ASCII 字母的词被报告为一个单独的词元类型，因为有时区分它们会很有用。
在大部分欧洲语言中，词元类型word和asciiword应该被同样对待。

email 不支持 RFC 5322 定义的所有有效电子邮件字符。具体来说，电子邮件用户名中支持的非字母数字字符只有句点、短横线和下划线。

解析器有可能从同一段文本中产生重叠的词元。举例来说，
一个带连字符的词既会被报告为整个单词，也会被报告为各个组成部分：

```
SELECT alias, description, token FROM ts_debug('foo-bar-beta1');
      alias      |               description                |
 token
-----------------+------------------------------------------+--------
-------
 numhword        | Hyphenated word, letters and digits      | foo-bar-
beta1
 hword_asciipart | Hyphenated word part, all ASCII          | foo
 blank           | Space symbols                            | -
 hword_asciipart | Hyphenated word part, all ASCII          | bar
 blank           | Space symbols                            | -
 hword_numpart   | Hyphenated word part, letters and digits | beta1
```

这种行为是可取的，因为它既允许针对整个复合词搜索，也允许针对各个组成部分搜索。
下面是另一个有启发性的示例：

```
SELECT alias, description, token FROM ts_debug('http://example.com/
stuff/index.html');
  alias   |  description  |            token
----------+---------------+------------------------------
 protocol | Protocol head | http://
 url      | URL           | example.com/stuff/index.html
 host     | Host          | example.com
```

```
 url_path | URL path      | /stuff/index.html
```

12.6. 词典

词典用于消除不应参与搜索的词（停用词），并用于对词进行正规化，
以便同一个词的不同派生形式能够匹配。成功完成正规化的词被称为词位。除了改善搜索质量，
正规化和移除停用词还会减小文档的`tsvector`表示，从而提高性能。
正规化并不总是具有语言学意义，而且通常依赖于应用的语义。

一些正规化的示例：

- 语言学上的例子 – Ispell 词典尝试把输入词约简为一种正规化形式；词干分析器词典则去掉词尾
- URL 地址可以规范化，以便让等价的 URL 匹配：
  - http://www.pgsql.ru/db/mw/index.html
  - http://www.pgsql.ru/db/mw/
  - http://www.pgsql.ru/db/../db/mw/index.html
- 颜色名可以被它们的十六进制值替换，例如`red, green, blue, magenta -> FF0000, 00FF00, 0000FF, FF00FF`
- 如果要索引数字，我们可以去掉某些小数位，以缩小可能数值的范围。因此，
  如果只保留小数点后两位，那么例如 3.14159265359、3.1415926 和 3.14 在正规化后就会变成相同的值。

一个词典是一个程序，它接受一个词元作为输入，并返回：

- 如果输入的词元对词典是已知的，
  则返回一个词位数组（注意一个词元可能产生多于一个词位）
- 一个带有 `TSL_FILTER` 标志的单个词位，它会用一个新词元替换原始词元，并把它传递给后续词典（执行这种工作的词典称为过滤词典）
- 如果词典知道该词元但它是一个停用词，则返回一个空数组
- 如果词典不识别该输入词元，则返回`NULL`

PostgreSQL为许多语言提供了预定义的词典。
也有多种预定义模板可以被用于创建带自定义参数的新词典。
每一种预定义词典模板在下面描述。如果没有合适的现有模板，可以创建新的；示例见
PostgreSQL发行版的`contrib/`目录。

文本检索配置把一个解析器与一组用于处理解析器输出词元的词典绑定在一起。
对于解析器可能返回的每一种词元类型，配置都会指定一个单独的词典列表。
当解析器找到该类型的词元时，会按顺序依次查询列表中的每个词典，
直到有某个词典把它识别为已知词。如果它被识别为停用词，或者没有任何词典识别它，
那么该词元就会被丢弃，既不会建立索引，也不会参与搜索。通常，第一个返回非 `NULL` 输出的词典就决定结果，后续词典不会再被查询；
但过滤词典可以把给定单词替换为一个修改后的单词，再传递给后续词典。

配置词典列表的一般规则是，把适用范围最窄、最专门的词典放在最前面，
然后是更通用的词典，最后以一个非常通用的词典收尾，例如 Snowball 词干分析器，或能识别所有内容的 `simple`。例如，对于一个天文学相关搜索（配置名为 `astro_en`），可以把词元类型 `asciiword`（ASCII 词）绑定到一个天文学术语同义词词典、一个通用英语词典，以及一个Snowball 英语词干分析器：

```
ALTER TEXT SEARCH CONFIGURATION astro_en
    ADD MAPPING FOR asciiword WITH astrosyn, english_ispell, english_
stem;
```

过滤词典可以放在列表中的任何位置，但放在末尾没有用处。
过滤词典可用于先对词做部分正规化，以简化后续词典的工作。例如，
可以用过滤词典去掉带重音字母中的重音符号，就像unaccent模块所做的那样。

12.6.1. 停用词

停用词是非常常见、几乎出现在每个文档中的词，没有区分价值。因此，
在全文检索中可以忽略它们。例如，每篇英文文本都包含像 a 和 the 这样的词，所以把它们存储在索引中没有用处。不过，停用词确实会影响 tsvector 中的位置，而这又会影响排名：

```
SELECT to_tsvector('english','in the list of stop words');
        to_tsvector
----------------------------
 'list':3 'stop':5 'word':6
```

缺失的位置 1、2、4 就是由停用词造成的。对包含停用词和不包含停用词的文档，计算出来的排名会明显不同：

```
SELECT ts_rank_cd (to_tsvector('english','in the list of stop words'),
 to_tsquery('list & stop'));
 ts_rank_cd
------------
       0.05

SELECT ts_rank_cd (to_tsvector('english','list stop words'), to_
tsquery('list & stop'));
 ts_rank_cd
------------
        0.1
```

如何处理停用词取决于具体词典。例如，ispell词典会先正规化单词，再查看停用词列表，而Snowball词干分析器则先检查停用词列表。这种不同行为是为了尽量减少噪声。

12.6.2. 简单词典

simple 词典模板的工作方式是先把输入词元转换为小写，再根据停用词文件检查它。如果在文件中找到该词元，就返回一个空数组，从而丢弃该词元；否则，
返回其小写形式作为正规化后的词位。或者，也可以把该词典配置为把非停用词报告为未识别，
从而允许它们传递给列表中的下一个词典。

下面是一个使用 simple 模板的词典定义示例：

```
CREATE TEXT SEARCH DICTIONARY public.simple_dict (
    TEMPLATE = pg_catalog.simple,
    STOPWORDS = english
);
```

其中，english 是停用词文件的基名。文件完整名称将是 $SHAREDIR/tsearch_data/english.stop，其中 $SHAREDIR 表示 PostgreSQL 安装的共享数据目录，通常是 /usr/local/share/postgresql（如果不确定，可用 pg_config --sharedir 来确定）。文件格式只是一个单词列表，每行一个单词。空行和行尾空格会被忽略，大写会折叠为小写，但不会对文件内容做其他处理。

现在我们可以测试这个词典：

```
SELECT ts_lexize('public.simple_dict','YeS');
 ts_lexize
-----------
 {yes}

SELECT ts_lexize('public.simple_dict','The');
 ts_lexize
-----------
 {}
```

如果在停用词文件中找不到该词，我们也可以选择返回 `NULL`，而不是返回它的小写形式。这种行为可以通过把词典的 `Accept` 参数设置为 `false` 来启用。继续上面的例子：

```
ALTER TEXT SEARCH DICTIONARY public.simple_dict ( Accept = false );

SELECT ts_lexize('public.simple_dict','YeS');
 ts_lexize
-----------


SELECT ts_lexize('public.simple_dict','The');
 ts_lexize
-----------
 {}
```

在默认设置 `Accept` = `true` 下，只有把 `simple` 词典放在词典列表末尾才有意义，因为它不会把任何词元传递给后续词典。相反，`Accept` = `false` 只有在后面至少还有一个词典时才有用。

> **小心**
>
> 大多数词典类型都依赖配置文件，例如停用词文件。这些文件必须采用 UTF-8 编码保存。当它们被读入服务器时，如果数据库编码不同，就会被转换为实际的数据库编码。

> **小心**
>
> 通常，词典配置文件在某个数据库会话中第一次被使用时，只会被读取一次。
> 如果你修改了配置文件，并希望强制现有会话加载新内容，可以对该词典执行一条 `ALTER TEXT SEARCH DICTIONARY` 命令。这可以是一条“假”更新，也就是并不实际修改任何参数值的更新。

12.6.3. 同义词词典

这个词典模板用于创建把一个单词替换为同义词的词典。
不支持短语（对此请使用分类词典模板第 12.6.4 节）。同义词词典可以用来克服语言学上的问题，例如防止英语词干分析词典把 'Paris' 约简成 'pari'。只要在同义词词典中加入一行 `Paris paris`，并把该词典放在 `english_stem` 词典之前即可。例如：

```
SELECT * FROM ts_debug('english', 'Paris');
   alias   |   description   | token |  dictionaries  |  dictionary  |
 lexemes
```

```
-----------+-----------------+-------+----------------+--------------+
---------
 asciiword | Word, all ASCII | Paris | {english_stem} | english_stem |
 {pari}

CREATE TEXT SEARCH DICTIONARY my_synonym (
    TEMPLATE = synonym,
    SYNONYMS = my_synonyms
);

ALTER TEXT SEARCH CONFIGURATION english
    ALTER MAPPING FOR asciiword
    WITH my_synonym, english_stem;

SELECT * FROM ts_debug('english', 'Paris');
   alias   |   description   | token |       dictionaries        |
 dictionary | lexemes
-----------+-----------------+-------+---------------------------+----
--------+---------
 asciiword | Word, all ASCII | Paris | {my_synonym,english_stem} | my_
synonym | {paris}
```

synonym模板要求的唯一参数是SYNONYMS，它是其配置文件的基名 – 上例中的my_synonyms。该文件的完整名称将是$SHAREDIR/tsearch_data/my_synonyms.syn（其中$SHAREDIR表示PostgreSQL安装的共享数据目录）。
该文件格式是每行一个要被替换的词，后面跟着它的同义词，用空白分隔。
空行和结尾的空格会被忽略。

synonym模板还有一个可选参数CaseSensitive，默认值为false。当CaseSensitive为false时，同义词文件中的词会像输入词元一样折叠为小写。当它为true时，词和词元都不会被折叠为小写，而是按原样比较。

可以在配置文件中把星号（*）放在同义词末尾，表示该同义词是一个前缀。当该条目用于 to_tsvector() 时，星号会被忽略；但当它用于 to_tsquery() 时，结果将是一个带有前缀匹配标记的查询项（见第 12.3.2 节）。例如，假设在 $SHAREDIR/tsearch_data/synonym_sample.syn 中有以下条目：

```
postgres        pgsql
postgresql      pgsql
postgre pgsql
gogle   googl
indices index*
```

那么将得到如下结果：

```
mydb=# CREATE TEXT SEARCH DICTIONARY syn (template=synonym, synonyms=
'synonym_sample');
mydb=# SELECT ts_lexize('syn','indices');
 ts_lexize
-----------
 {index}
(1 row)

mydb=# CREATE TEXT SEARCH CONFIGURATION tst (copy=simple);
```

```
mydb=# ALTER TEXT SEARCH CONFIGURATION tst ALTER MAPPING FOR asciiword
 WITH syn;
mydb=# SELECT to_tsvector('tst','indices');
 to_tsvector
-------------
 'index':1
(1 row)

mydb=# SELECT to_tsquery('tst','indices');
 to_tsquery
------------
 'index':*
(1 row)

mydb=# SELECT 'indexes are very useful'::tsvector;
            tsvector
---------------------------------
 'are' 'indexes' 'useful' 'very'
(1 row)

mydb=# SELECT 'indexes are very useful'::tsvector @@ to_tsquery('tst',
'indices');
 ?column?
----------
 t
(1 row)
```

12.6.4. 分类词典

一个分类词典（有时被简写成TZ）是一个词的集合，其中包括了词与短语之间的联系，即广义词（BT）、狭义词（NT）、首选词、非首选词、相关词等。

分类词典基本上会用一个首选词替换所有非首选词，并且也可以选择保留原始词用于索引。PostgreSQL的分类词典的当前实现是同义词词典的一个扩展，并增加了短语支持。一个分类词典需要一个采用以下格式的配置文件：

```
# 这是一条注释
示例词（一个或多个） : 索引词（一个或多个）
更多示例词（一个或多个） : 更多索引词（一个或多个）
...
```

其中冒号（`:`）符号用作短语及其替换词之间的分隔符。

分类词典会使用一个子词典（在词典配置中指定）在检查短语匹配之前正规化输入文本。只能选择一个子词典。如果子词典无法识别某个词，就会报错。在这种情况下，你应当避免使用该词，或者让子词典学会它。你可以在某个被索引词的开头放置一个星号（`*`），以跳过将子词典应用于该词，但所有样例词都必须能被子词典识别。

如果有多个短语匹配输入，分类词典会选择最长的匹配；长度相同时使用最后的定义。

由子词典识别的特定停用词不能够被指定；改用`?`标记可出现任意停用词的位置。例如，假定根据子词典`a`和`the`是停用词：

```
? one ? two : swsw
```

匹配a one the two和the one a two；两者都将被swsw替换。

由于分类词典具备识别短语的能力，因此它必须记住自身状态并与解析器交互。
分类词典会利用这些映射关系来判断自己是应当处理下一个词，还是停止累积。
分类词典必须经过仔细配置。例如，如果分类词典只被映射到 asciiword 词元，那么像 one 7 这样的分类词典定义就无法工作，因为词元类型 uint 并没有映射给该分类词典。

> **小心**
>
> 分类词典会在建立索引时使用，因此其参数发生任何变化都要求重新建立索引。
> 对于大多数其他词典类型，
> 像增加或移除停用词这样的细小改动则不会强制重新建立索引。

12.6.4.1. 分类词典配置

要定义一个新的分类词典，可使用thesaurus模板。例如：

```
CREATE TEXT SEARCH DICTIONARY thesaurus_simple (
    TEMPLATE = thesaurus,
    DictFile = mythesaurus,
    Dictionary = pg_catalog.english_stem
);
```

这里：

- thesaurus_simple是新词典的名称
- mythesaurus是分类词典配置文件的基名（它的全名将是$SHAREDIR/tsearch_data/mythesaurus.ths，其中$SHAREDIR表示安装的共享数据目录）。
- pg_catalog.english_stem是要用于分类词典正规化的子词典（这里是一个 Snowball 英语词干分析器）。注意子词典将拥有它自己的配置（例如停用词），但这里没有展示。

现在可以在配置中把分类词典thesaurus_simple绑定到想要的词元类型上，例如：

```
ALTER TEXT SEARCH CONFIGURATION russian
    ALTER MAPPING FOR asciiword, asciihword, hword_asciipart
    WITH thesaurus_simple;
```

12.6.4.2. 分类词典示例

考虑一个简单的天文分类词典 thesaurus_astro，其中包含一些天文单词组合：

```
supernovae stars : sn
crab nebulae : crab
```

下面我们创建一个词典，并把一些词元类型绑定到天文分类词典和英语词干分析器上：

```
CREATE TEXT SEARCH DICTIONARY thesaurus_astro (
    TEMPLATE = thesaurus,
    DictFile = thesaurus_astro,
    Dictionary = english_stem
);
```

```
ALTER TEXT SEARCH CONFIGURATION russian
    ALTER MAPPING FOR asciiword, asciihword, hword_asciipart
    WITH thesaurus_astro, english_stem;
```

现在我们可以看看它是如何工作的。ts_lexize 对测试分类词典并不十分有用，因为它把输入当作一个单独的词元。相反，我们可以使用 plainto_tsquery 和 to_tsvector，它们会把输入字符串拆分成多个词元：

```
SELECT plainto_tsquery('supernova star');
 plainto_tsquery
-----------------
 'sn'

SELECT to_tsvector('supernova star');
 to_tsvector
-------------
 'sn':1
```

原则上，如果把参数用引号括起来，也可以使用 to_tsquery：

```
SELECT to_tsquery('''supernova star''');
 to_tsquery
------------
 'sn'
```

请注意，supernova star 能匹配 thesaurus_astro 中的 supernovae stars，因为我们在分类词典定义中指定了 english_stem 词干分析器。它去掉了词尾的 e 和 s。

如果既要为替换词建立索引，也要为原始短语建立索引，只需把原始短语也写到定义右侧即可：

```
supernovae stars : sn supernovae stars

SELECT plainto_tsquery('supernova star');
       plainto_tsquery
-----------------------------
 'sn' & 'supernova' & 'star'
```

12.6.5. Ispell 词典

Ispell词典模板支持形态词典，它可以把一个词的许多不同语言学形式正规化为同一个词位。例如，一个英语 Ispell 词典可以匹配搜索词 bank 的所有变格和变位形式，例如 banking、banked、banks、banks' 和 bank's。

标准PostgreSQL发行版不包含任何Ispell配置文件。大量语言的词典可从Ispell[1]获取。此外，还支持一些更现代的词典文件格式 – MySpell[2] (OO < 2.0.1) 和Hunspell[3](OO >= 2.0.2)。在OpenOffice Wiki[4]上有大量词典可用。

要创建一个Ispell词典，可使用内置的ispell模板并指定几个参数：

```
CREATE TEXT SEARCH DICTIONARY english_ispell (
```

[1] http://ficus-www.cs.ucla.edu/geoff/ispell.html
[2] http://en.wikipedia.org/wiki/MySpell
[3] http://sourceforge.net/projects/hunspell/
[4] http://wiki.services.openoffice.org/wiki/Dictionaries

```
    TEMPLATE = ispell,
    DictFile = english,
    AffFile = english,
    StopWords = english
);
```

这里，DictFile、AffFile和StopWords指定词典、词缀和停用词文件的基名。
停用词文件的格式和前面解释的simple词典类型相同。其他文件的格式在这里没有指定，
但是也可以从上面提到的网站获得。

Ispell 词典通常只能识别有限的词，因此其后应跟着另一个覆盖范围更广的词典；例如，一个 Snowball 词典，它可以识别所有输入。

Ispell 词典支持划分复合词，这是一个有用的特性。注意词缀文件应该用compoundwords controlled语句指定一个特殊标志，用于标记词典中可以参与构成复合词的词：

```
compoundwords  controlled z
```

下面是挪威语的一些示例：

```
SELECT ts_lexize('norwegian_ispell',
 'overbuljongterningpakkmesterassistent');
   {over,buljong,terning,pakk,mester,assistent}
SELECT ts_lexize('norwegian_ispell', 'sjokoladefabrikk');
   {sjokoladefabrikk,sjokolade,fabrikk}
```

> **注意**
>
> MySpell 不支持复合词。Hunspell则对复合词有更好的支持。当前，PostgreSQL只实现了 Hunspell 中基本的复合词操作。

12.6.6. Snowball 词典

Snowball词典模板基于 Martin Porter 的一个项目，他是流行的英语 Porter 词干分析算法的发明者。Snowball 现在对许多语言提供词干分析算法（详见Snowball 站点[5]）。每种算法都知道如何把其语言中词的常见变体约简为基本拼写形式，也就是词干。一个 Snowball 词典要求一个language参数来标识要用哪种词干分析器，并且可以选择指定一个stopword
文件名来给出一个要被消除的词列表（PostgreSQL的标准停用词列表也是由 Snowball 项目提供的）。例如，有一个内置的定义等效于

```
CREATE TEXT SEARCH DICTIONARY english_stem (
    TEMPLATE = snowball,
    Language = english,
    StopWords = english
);
```

停用词文件格式和已经解释的一样。

一个Snowball词典识别所有输入，不管它能不能简化该词，
因此它应当被放置在词典列表的最后。把它放在任何其他词典前面是没有用处的，
因为一个词元永远不会穿过它而进入到下一个词典。

[5] http://snowballstem.org/

12.7. 配置示例

一个文本检索配置指定了将一个文档转换成一个tsvector
所需的所有选项：用于把文本分解成词元的解析器，以及用于将每一个词元转换成词位的词典。
每一次to_tsvector或to_tsquery的调用都需要一个文本检索配置来执行其处理。
配置参数default_text_search_config指定了默认配置的名称，如果省略了显式的配置参数，
文本检索函数将会使用它。它可以在postgresql.conf中设置，或者使用SET
命令为一个单独的会话设置。

有一些预定义的文本检索配置可用，并且你可以容易地创建自定义的配置。
为了便于管理文本检索对象，可以使用一组SQL命令，并且有多个
psql命令可以显示有关文本检索对象（第 12.10 节）的信息。

作为一个示例，我们将创建一个配置pg，从复制内置的english配置开始：

```
CREATE TEXT SEARCH CONFIGURATION public.pg ( COPY = pg_catalog.
english );
```

我们将使用一个 PostgreSQL 相关的同义词列表，并将它存储在$SHAREDIR/tsearch_data/pg_dict.syn中。文件内容看起来像：

```
postgres    pg
pgsql       pg
postgresql  pg
```

我们定义同义词词典如下：

```
CREATE TEXT SEARCH DICTIONARY pg_dict (
    TEMPLATE = synonym,
    SYNONYMS = pg_dict
);
```

接下来我们注册Ispell词典english_ispell，它有其自己的配置文件：

```
CREATE TEXT SEARCH DICTIONARY english_ispell (
    TEMPLATE = ispell,
    DictFile = english,
    AffFile = english,
    StopWords = english
);
```

现在我们可以在配置pg中建立词的映射：

```
ALTER TEXT SEARCH CONFIGURATION pg
    ALTER MAPPING FOR asciiword, asciihword, hword_asciipart,
                      word, hword, hword_part
    WITH pg_dict, english_ispell, english_stem;
```

我们选择不索引或搜索某些内置配置确实处理的词元类型：

```
ALTER TEXT SEARCH CONFIGURATION pg
    DROP MAPPING FOR email, url, url_path, sfloat, float;
```

现在我们可以测试我们的配置：

```
SELECT * FROM ts_debug('public.pg', '
PostgreSQL, the highly scalable, SQL compliant, open source object-
relational
database management system, is now undergoing beta testing of the next
version of our software.
');
```

下一步是把会话设置为使用这个新配置，该配置是在 public 模式中创建的：

```
=> \dF
   List of text search configurations
 Schema  | Name | Description
---------+------+-------------
 public  | pg   |

SET default_text_search_config = 'public.pg';
SET

SHOW default_text_search_config;
 default_text_search_config
----------------------------
 public.pg
```

12.8. 测试和调试文本搜索

一个自定义文本检索配置的行为很容易让人困惑。
本节中描述的函数对于测试文本检索对象有用。你可以测试一个完整的配置，
或者独立测试解析器和词典。

12.8.1. 配置测试

函数ts_debug可方便地测试一个文本检索配置。

```
ts_debug([ config regconfig, ] document text,
         OUT alias text,
         OUT description text,
         OUT token text,
         OUT dictionaries regdictionary[],
         OUT dictionary regdictionary,
         OUT lexemes text[])
         returns setof record
```

ts_debug显示*document*的每一个词元的信息，词元由解析器产生并由配置的词典处理过。
该函数使用由*config*指定的配置，如果省略该参数则使用default_text_search_config
指定的配置。

ts_debug为解析器在文本中标识的每一个词元返回一行。被返回的列是：

- *alias* text — 词元类型的短名称
- *description* text — 词元类型的描述
- *token* text — 词元的文本
- *dictionaries* regdictionary[] — 配置为这种词元类型选择的词典

- *dictionary* regdictionary – 识别该词元的词典，如果没有词典能识别则为NULL
- *lexemes* text[] – 识别该词元的词典产生的词位，如果没有词典能识别则为NULL；一个空数组（{}）表示该词元被识别为一个停用词

以下是一个简单的示例：

```
SELECT * FROM ts_debug('english','a fat  cat sat on a mat - it ate a
 fat rats');
   alias   |   description   | token |  dictionaries  |  dictionary  |
 lexemes
-----------+-----------------+-------+----------------+--------------+
---------
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem |
 {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | fat   | {english_stem} | english_stem |
 {fat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | cat   | {english_stem} | english_stem |
 {cat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | sat   | {english_stem} | english_stem |
 {sat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | on    | {english_stem} | english_stem |
 {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem |
 {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | mat   | {english_stem} | english_stem |
 {mat}
 blank     | Space symbols   |       | {}             |              |
 blank     | Space symbols   | -     | {}             |              |
 asciiword | Word, all ASCII | it    | {english_stem} | english_stem |
 {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | ate   | {english_stem} | english_stem |
 {ate}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem |
 {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | fat   | {english_stem} | english_stem |
 {fat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | rats  | {english_stem} | english_stem |
 {rat}
```

为了做更完整的演示，我们先为英语创建一个public.english配置和 Ispell 词典：

```
CREATE TEXT SEARCH CONFIGURATION public.english ( COPY = pg_catalog.
english );
```

```
CREATE TEXT SEARCH DICTIONARY english_ispell (
    TEMPLATE = ispell,
    DictFile = english,
    AffFile = english,
    StopWords = english
);

ALTER TEXT SEARCH CONFIGURATION public.english
   ALTER MAPPING FOR asciiword WITH english_ispell, english_stem;

SELECT * FROM ts_debug('public.english','The Brightest supernovaes');
   alias   |   description   |    token    |         dictionaries
     |   dictionary   |   lexemes
-----------+-----------------+-------------+-------------------------
-----+----------------+-------------
 asciiword | Word, all ASCII | The         | {english_ispell,english_
stem} | english_ispell | {}
 blank     | Space symbols   |             | {}
     |                |
 asciiword | Word, all ASCII | Brightest   | {english_ispell,english_
stem} | english_ispell | {bright}
 blank     | Space symbols   |             | {}
     |                |
 asciiword | Word, all ASCII | supernovaes | {english_ispell,english_
stem} | english_stem   | {supernova}
```

在这个示例中，词Brightest被解析器识别为一个ASCII 词（别名asciiword）。对于这种词元类型，词典列表是english_ispell和english_stem。该词被english_ispell识别，并被这个词典归约为词位bright。词supernovaes对english_ispell词典来说是未知的，因此它会被传递给下一个词典；幸运的是，它随后被识别了。（实际上，english_stem是一个Snowball 词典，它能够识别所有输入；这也是为什么它被放在词典列表末尾。）

词The被english_ispell词典识别为一个停用词（第 12.6.1 节）并且将不会被索引。空格也被丢弃，因为该配置没有为它们提供词典。

你可以通过明确指定要查看的列来减少输出宽度：

```
SELECT alias, token, dictionary, lexemes
FROM ts_debug('public.english','The Brightest supernovaes');
   alias   |    token    |   dictionary   |   lexemes
-----------+-------------+----------------+-------------
 asciiword | The         | english_ispell | {}
 blank     |             |                |
 asciiword | Brightest   | english_ispell | {bright}
 blank     |             |                |
 asciiword | supernovaes | english_stem   | {supernova}
```

12.8.2. 解析器测试

下列函数允许直接测试一个文本检索解析器。

```
ts_parse(parser_name text, document text,
         OUT tokid integer, OUT token text) returns setof record
```

```
ts_parse(parser_oid oid, document text,
         OUT tokid integer, OUT token text) returns setof record
```

ts_parse解析给定的*document*，并返回一组记录，
每个由解析过程产生的词元对应一条记录。每条记录都包含一个 tokid，显示所分配的词元类型，以及一个 token，即该词元的文本。例如：

```
SELECT * FROM ts_parse('default', '123 - a number');
 tokid | token
-------+--------
    22 | 123
    12 |
    12 | -
     1 | a
    12 |
     1 | number
```

```
ts_token_type(parser_name text, OUT tokid integer,
              OUT alias text, OUT description text) returns setof
 record
ts_token_type(parser_oid oid, OUT tokid integer,
              OUT alias text, OUT description text) returns setof
 record
```

ts_token_type返回一个表，描述指定解析器能够识别的每一种词元类型。
对于每种词元类型，该表给出解析器用来标记该类词元的整数 tokid、在配置命令中命名该词元类型的 alias，以及简短的 description。例如：

```
SELECT * FROM ts_token_type('default');
 tokid |      alias      |               description
-------+-----------------+-------------------------------------------
     1 | asciiword       | Word, all ASCII
     2 | word            | Word, all letters
     3 | numword         | Word, letters and digits
     4 | email           | Email address
     5 | url             | URL
     6 | host            | Host
     7 | sfloat          | Scientific notation
     8 | version         | Version number
     9 | hword_numpart   | Hyphenated word part, letters and digits
    10 | hword_part      | Hyphenated word part, all letters
    11 | hword_asciipart | Hyphenated word part, all ASCII
    12 | blank           | Space symbols
    13 | tag             | XML tag
    14 | protocol        | Protocol head
    15 | numhword        | Hyphenated word, letters and digits
    16 | asciihword      | Hyphenated word, all ASCII
    17 | hword           | Hyphenated word, all letters
    18 | url_path        | URL path
    19 | file            | File or path name
    20 | float           | Decimal notation
    21 | int             | Signed integer
    22 | uint            | Unsigned integer
    23 | entity          | XML entity
```

12.8.3. 词典测试

`ts_lexize`函数便于测试词典。

```
ts_lexize(dict regdictionary, token text) returns text[]
```

如果输入的*token*是该词典已知的，则`ts_lexize`返回一个词位数组；
如果词元是词典已知的但是它是一个停用词，则返回一个空数组；或者如果它对词典是未知词，则返回`NULL`。

示例：

```
SELECT ts_lexize('english_stem', 'stars');
 ts_lexize
-----------
 {star}

SELECT ts_lexize('english_stem', 'a');
 ts_lexize
-----------
 {}
```

注意

`ts_lexize`函数期望的是单个词元，而不是文本。下面这种情况可能会让人困惑：

```
SELECT ts_lexize('thesaurus_astro','supernovae stars') is null;
 ?column?
----------
 t
```

分类词典 `thesaurus_astro` 确实知道短语 `supernovae  stars`，但 `ts_lexize` 会失败，因为它不会解析输入文本，而是把它当作一个单独的词元。测试分类词典时应使用 `plainto_tsquery` 或 `to_tsvector`，例如：

```
SELECT plainto_tsquery('supernovae stars');
 plainto_tsquery
-----------------
 'sn'
```

12.9. GiST 和 GIN 索引类型

有两种索引可以用来加速全文搜索。请注意，索引对于全文搜索并非强制要求，
但在定期搜索某一列的情况下，通常是可取的。

```
CREATE INDEX name ON table USING gist(column);
```

创建基于 GiST（广义搜索树）的索引。*column*可以是`tsvector`或`tsquery`类型。

```
CREATE INDEX name ON table USING gin(column);
```

创建基于 GIN（广义倒排索引）的索引。*column*必须是`tsvector`类型。

这两种索引类型之间存在显著的性能差异，因此理解它们的特性很重要。

一个 GiST 索引是有损的，这表示索引可能产生假匹配，并且有必要检查真实的表行来消除这种假匹配（PostgreSQL在需要时会自动做这一步）。GiST 索引之所以是有损的，是因为每一个文档在索引中被表示为一个定长的签名。该签名通过将每个词 hash 到一个 n 位串中的一位，再将所有这些位进行 OR 运算来生成，结果是一个 n 位的文档签名。当两个词 hash 到同一个位位置时就会产生假匹配。如果查询中所有词都有匹配（真或假），则必须检索表行查看匹配是否正确。

有损性导致的性能下降归因于不必要的表记录（即被证实为假匹配的记录）获取。
因为表记录的随机访问是较慢的，这限制了 GiST 索引的实用性。假匹配的可能性取决于几个因素，特别是不同词的数量，因此推荐使用词典来缩减这个数量。

GIN 索引对标准查询不是有损的，但其性能以对数方式依赖于不同词的数量。（不过，GIN 索引只存储`tsvector`值的词（词位），而不存储它们的权重标签。因此，
在使用涉及权重的查询时需要一次表行重新检查。）

在选择使用哪种索引类型（GiST 还是 GIN）时，请考虑这些性能差异：

- GIN 索引的查找大约比 GiST 快三倍
- GIN 索引的构建大约比 GiST 慢三倍
- GIN 索引的更新比 GiST 索引略慢，但如果禁用了快速更新支持则要慢大约 10 倍（详见第 54.3.1 节）
- GIN 索引比 GiST 索引大两到三倍

根据经验法则，GIN 索引最适合静态数据，因为查找更快。对于动态数据，GiST 索引的更新更快。具体来说，当不同词（词位）的数量在 100,000 以下时，GiST 索引对动态数据非常好而且快；而 GIN 索引能更好地处理 100,000 以上的词位，但更新较慢。

注意GIN索引的构建时间常常可以通过增加maintenance_work_mem来缩短，而
GiST索引的构建时间则对该参数不敏感。

对大集合分区并正确使用 GiST 和 GIN 索引可以实现带在线更新的极快搜索。分区可以在数据库层面上使用表继承来完成，或者通过将文档分布在服务器上，并使用
dblink模块收集搜索结果来完成。后者是可能的，因为排名函数只使用本地信息。

12.10. psql 支持

关于文本检索配置对象的信息可以在psql中使用一组命令获得：

```
\dF{d,p,t}[+] [PATTERN]
```

可选的+能产生更多细节。

可选参数 `PATTERN` 可以是文本检索对象的名称，也可以可选地带上模式限定。如果省略`PATTERN`，则会显示所有可见对象的信息。`PATTERN` 还可以是正则表达式，并且可以为模式名和对象名分别提供独立的匹配模式。下面的示例说明了这一点：

```
=> \dF *fulltext*
       List of text search configurations
 Schema |  Name        | Description
--------+--------------+-------------
 public | fulltext_cfg |

=> \dF *.fulltext*
```

```
      List of text search configurations
 Schema   |  Name        | Description
----------+----------------------------
 fulltext | fulltext_cfg |
 public   | fulltext_cfg |
```

可用命令如下：

```
\dF[+] [PATTERN]
```

列出文本检索配置（加上 + 可显示更多细节）。

```
=> \dF russian
            List of text search configurations
   Schema   |  Name   |            Description
------------+---------+-----------------------------------
 pg_catalog | russian | configuration for russian language

=> \dF+ russian
Text search configuration "pg_catalog.russian"
Parser: "pg_catalog.default"
      Token      | Dictionaries
-----------------+--------------
 asciihword      | english_stem
 asciiword       | english_stem
 email           | simple
 file            | simple
 float           | simple
 host            | simple
 hword           | russian_stem
 hword_asciipart | english_stem
 hword_numpart   | simple
 hword_part      | russian_stem
 int             | simple
 numhword        | simple
 numword         | simple
 sfloat          | simple
 uint            | simple
 url             | simple
 url_path        | simple
 version         | simple
 word            | russian_stem
```

```
\dFd[+] [PATTERN]
```

列出文本检索词典（加上 + 可显示更多细节）。

```
=> \dFd
                            List of text search dictionaries
   Schema   |      Name       |                        Description

------------+-----------------+-----------------------------------
-----------------------
 pg_catalog | danish_stem     | snowball stemmer for danish
 language
```

```
 pg_catalog | dutch_stem      | snowball stemmer for dutch language
 pg_catalog | english_stem    | snowball stemmer for english
 language
 pg_catalog | finnish_stem    | snowball stemmer for finnish
 language
 pg_catalog | french_stem     | snowball stemmer for french
 language
 pg_catalog | german_stem     | snowball stemmer for german
 language
 pg_catalog | hungarian_stem  | snowball stemmer for hungarian
 language
 pg_catalog | italian_stem    | snowball stemmer for italian
 language
 pg_catalog | norwegian_stem  | snowball stemmer for norwegian
 language
 pg_catalog | portuguese_stem | snowball stemmer for portuguese
 language
 pg_catalog | romanian_stem   | snowball stemmer for romanian
 language
 pg_catalog | russian_stem    | snowball stemmer for russian
 language
 pg_catalog | simple          | simple dictionary: just lower case
 and check for stopword
 pg_catalog | spanish_stem    | snowball stemmer for spanish
 language
 pg_catalog | swedish_stem    | snowball stemmer for swedish
 language
 pg_catalog | turkish_stem    | snowball stemmer for turkish
 language
```

```
    \dFp[+] [PATTERN]
```

列出文本检索解析器（加上 + 可显示更多细节）。

```
=> \dFp
        List of text search parsers
   Schema   |  Name   |     Description
------------+---------+---------------------
 pg_catalog | default | default word parser
=> \dFp+
    Text search parser "pg_catalog.default"
     Method      |    Function    | Description
-----------------+----------------+-------------
 Start parse     | prsd_start     |
 Get next token  | prsd_nexttoken |
 End parse       | prsd_end       |
 Get headline    | prsd_headline  |
 Get token types | prsd_lextype   |

        Token types for parser "pg_catalog.default"
   Token name    |               Description
-----------------+------------------------------------------
 asciihword      | Hyphenated word, all ASCII
 asciiword       | Word, all ASCII
```

```
 blank           | Space symbols
 email           | Email address
 entity          | XML entity
 file            | File or path name
 float           | Decimal notation
 host            | Host
 hword           | Hyphenated word, all letters
 hword_asciipart | Hyphenated word part, all ASCII
 hword_numpart   | Hyphenated word part, letters and digits
 hword_part      | Hyphenated word part, all letters
 int             | Signed integer
 numhword        | Hyphenated word, letters and digits
 numword         | Word, letters and digits
 protocol        | Protocol head
 sfloat          | Scientific notation
 tag             | XML tag
 uint            | Unsigned integer
 url             | URL
 url_path        | URL path
 version         | Version number
 word            | Word, all letters
(23 rows)
```

```
     \dFt[+] [PATTERN]
```

列出文本检索模板（加上 + 可显示更多细节）。

```
=> \dFt
                           List of text search templates
   Schema   |   Name    |                        Description
------------+-----------+-------------------------------------------
-----------------
 pg_catalog | ispell    | ispell dictionary
 pg_catalog | simple    | simple dictionary: just lower case and
 check for stopword
 pg_catalog | snowball  | snowball stemmer
 pg_catalog | synonym   | synonym dictionary: replace word by its
 synonym
 pg_catalog | thesaurus | thesaurus dictionary: phrase by phrase
 substitution
```

12.11. 限制

目前，PostgreSQL的文本检索功能存在以下限制：

- 每个词位的长度必须小于 2K 字节
- `tsvector` 的长度（词位加位置）必须小于 1 兆字节
- 词位数量必须小于 2^{64}
- `tsvector` 中的位置值必须大于 0 且不超过 16,383
- 每个词位的位置数不能超过 256 个
- `tsquery` 中的节点数（词位加操作符）必须小于 32,768

为了对比，PostgreSQL 8.1 的文档包含 10,441 个不同词，总数 335,420 个词，并且最频繁的词 “postgresql” 在 655 个文档中被提到 6,127 次。

另一个示例 – PostgreSQL的邮件列表归档在 461,020 条消息的 57,491,343 个词位中包含 910,989 个不同词。

12.12. 从 8.3 之前的文本搜索迁移

使用tsearch2模块进行文本搜索的应用需要进行一些调整，才能使用内置功能：

- 一些函数已被重命名，或在其参数列表上做了小调整，而且它们现在都位于 `pg_catalog`模式中，而在以前的安装中它们可能位于 `public`或其他非系统模式中。新版本的tsearch2 提供了一个兼容层，可解决这一领域的大多数问题。

- 从 8.3 之前的数据库加载pg_dump输出时，必须屏蔽旧的tsearch2函数和其他对象。虽然其中很多本来就无法加载，但少数可以加载并随后引发问题。
  处理这个问题的一种简单方法是在恢复转储之前先加载新的 tsearch2模块；这样它就会阻止旧对象被加载。

- 文本搜索配置的设置方式现在完全不同了。不再需要手动向配置表中插入行，
  而是通过本章前面介绍的专用 SQL 命令来配置搜索。没有将现有自定义配置转换为 8.3 形式的自动化支持；这一步需要你自己完成。

- 大多数类型的词典依赖于数据库之外的一些配置文件。这些文件与 8.3 之前的用法大体兼容，但请注意以下差异：
  - 配置文件现在必须放在一个指定的目录中（`$SHAREDIR/tsearch_data`），
    并且根据文件类型必须有特定的扩展名，如前面各种词典类型的描述中所述。
    添加这一限制是为了防止安全问题。
  - 无论使用什么数据库编码，配置文件都必须以 UTF-8 编码。
  - 在分类词典配置文件中，停用词必须用`?`标记。

第 13 章 并发控制

本章描述 PostgreSQL 数据库系统在两个或更多会话试图同时访问同一数据时的行为。在这种情况下，目标是在保持严格数据完整性的同时，为所有会话提供高效访问。
每一个数据库应用开发者都应该熟悉本章涉及的主题。

13.1. 介绍

PostgreSQL 为开发者提供了一组丰富的工具来管理对数据的并发访问。在内部，数据一致性是通过使用多版本模型（多版本并发控制，MVCC）来维护的。这意味着，在查询数据库时，
每个事务看到的都是某个较早时刻的数据快照（即一个数据库版本），
而不受底层数据当前状态的影响。这样可以保护事务不看到由（其他）
并发事务在同一数据行上执行更新所可能导致的不一致数据，从而为每个数据库会话提供
事务隔离。MVCC 摒弃了传统数据库系统的加锁方法，尽量减少锁争用，以便在多用户环境中获得合理的性能。

使用 MVCC 并发控制模型而不是加锁的主要优点在于，在 MVCC 中，为查询（读取）数据而获得的锁不会与为写入数据而获得的锁冲突，因此读取不会阻塞写入，写入也不会阻塞读取。
PostgreSQL 甚至在通过创新性的可串行化快照隔离（SSI）级别提供最严格的事务隔离时，也保持了这一保证。

PostgreSQL 也提供表级和行级锁设施，供那些通常不需要完整事务隔离、并且倾向于显式管理特定冲突点的应用使用。不过，正确使用 MVCC 通常会比加锁提供更好的性能。此外，由应用定义的咨询锁还提供了一种获取不依附于单个事务的锁的机制。

13.2. 事务隔离

SQL 标准定义了四个事务隔离级别。最严格的是可串行化，标准将它定义为：一组可串行化事务的任意并发执行，都保证与按照某种顺序一次只运行一个事务的效果相同。
其他三个级别则按现象定义，这些现象源自并发事务之间的相互作用，
而在各自级别上这些现象不得发生。标准指出，由于可串行化的定义，
这些现象在该级别下都不可能出现。（这并不奇怪，
如果事务的效果必须与一次只运行一个事务保持一致，
又怎么可能看到由相互作用引起的任何现象呢？）

在各个级别上被禁止的现象有：

脏读

一个事务读取了由另一个并发未提交事务写入的数据。

不可重复读

一个事务重新读取它之前读过的数据，
发现这些数据已经被另一个事务修改（且该事务已在初次读取之后提交）。

幻读

一个事务重新执行一个返回满足某个搜索条件的一组行的查询，
却发现由于另一个最近提交的事务，满足该条件的行集已经发生了变化。

四种事务隔离级别及其对应行为见表 13.1。

表 13.1. 标准 SQL 事务隔离级别

隔离级别	脏读	不可重复读	幻读
读未提交	可能	可能	可能
读已提交	不可能	可能	可能
可重复读	不可能	不可能	可能
可串行化	不可能	不可能	不可能

在 PostgreSQL 中，你可以请求四种标准事务隔离级别中的任意一种。但在内部，实际上只有三种不同的隔离级别，分别对应读已提交、可重复读和可串行化这几个级别。
当你选择读未提交级别时，实际得到的是读已提交；而在 PostgreSQL 的可重复读实现中，幻读是不可能发生的，因此实际的隔离级别可能比你选择的更严格。这是 SQL 标准所允许的：这四个隔离级别只定义了哪些现象不允许发生，并没有定义哪些现象必须发生。PostgreSQL 之所以只提供三种隔离级别，是因为要把标准隔离级别映射到多版本并发控制架构上，
这是唯一合理的做法。可用隔离级别的行为将在下面各小节中详细说明。

要设置事务的事务隔离级别，请使用命令SET TRANSACTION。

13.2.1. 读已提交隔离级别

读已提交是 PostgreSQL 中默认的隔离级别。当事务使用这一隔离级别时，`SELECT` 查询（不带 `FOR UPDATE/SHARE` 子句）只能看到查询开始之前已经提交的数据；它既看不到未提交的数据，也看不到查询执行期间并发事务提交的更改。实际上，`SELECT` 查询看到的是该查询开始运行瞬间的数据库快照。不过，`SELECT` 能看到其自身事务中先前执行的更新效果，即使这些更新尚未提交。还要注意，即使两个连续的 `SELECT` 命令位于同一个事务中，如果其他事务在第一个 `SELECT` 执行期间提交了更改，这两个命令也可能看到不同的数据。

`UPDATE`、`DELETE`、`SELECT FOR UPDATE` 和 `SELECT FOR SHARE` 命令在搜索目标行时的行为与 `SELECT` 相同：它们只会找到在命令开始时已经提交的目标行。不过，当找到这样的目标行时，它可能已经被其他并发事务更新、删除或者锁定。在这种情况下，
即将执行更新的事务会等待第一个更新事务提交或回滚（如果它仍在进行中）。
如果第一个更新事务回滚，那么它的作用将被撤销，
第二个更新事务就可以继续更新最初找到的行。如果第一个更新事务提交了，
那么若该行已被第一个更新者删除，第二个更新事务就会忽略该行；否则，
第二个更新者将尝试在该行的已更新版本上应用自己的操作。该命令的搜索条件（`WHERE` 子句）会被重新计算，以判断该行的已更新版本是否仍然符合搜索条件。如果符合，
则第二个更新者会基于该行的已更新版本继续执行操作。对于 `SELECT FOR UPDATE` 和 `SELECT FOR SHARE`，这意味着被锁住并返回给客户端的是该行的已更新版本。

由于上述规则，
更新命令可能看到一个不一致的快照：它能够看到并发更新命令在它试图更新的同一行上的效果，却看不到这些命令对数据库中其他行的影响。
这种行为使得读已提交模式不适合涉及复杂搜索条件的命令；不过，
对于较简单的场景它恰到好处。例如，考虑以如下事务更新银行余额：

```
BEGIN;
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 12345;
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 7534;
COMMIT;
```

如果两个这样的事务并发地尝试更改账户 12345 的余额，我们显然希望第二个事务从该账户行的已更新版本开始。因为每个命令只影响一个预先确定的行，
让它看到该行的已更新版本不会造成任何麻烦的不一致。

在读已提交模式下，更复杂的用法可能产生不理想的结果。例如，考虑一个 DELETE 命令，另一个命令正在修改数据，使某些行开始满足其筛选条件、另一些行不再满足。假设 website 是一个有两行的表，其中 website.hits 分别等于 9 和 10：

```
BEGIN;
UPDATE website SET hits = hits + 1;
-- 从另一个会话运行：  DELETE FROM website WHERE hits = 10;
COMMIT;
```

该 DELETE 将不会产生任何效果，即使在 UPDATE 之前和之后都存在一行 website.hits = 10。这是因为更新前值为 9 的那一行被跳过了，而当 UPDATE 完成并且 DELETE 获得锁时，新行值已经不再是 10 而是 11，因此不再匹配条件。

因为在读已提交模式中，每个命令都从一个新快照开始，
而该快照包含截至当时已提交的所有事务，
所以同一事务中的后续命令无论如何都会看到并发事务已提交的效果。上面真正的问题在于，
单个命令是否能看到数据库的绝对一致视图。

读已提交模式所提供的部分事务隔离对于许多应用来说已经足够，
而且这种模式既快速又容易使用。不过，它并不能满足所有场景。执行复杂查询和更新的应用，
可能需要比读已提交模式所提供的更严格一致的数据库视图。

13.2.2. 可重复读隔离级别

可重复读隔离级别只能看到事务开始前已提交的数据；它看不到未提交的数据，
也看不到事务执行期间并发事务提交的更改。（不过，
每个查询都能看到其自身事务中先前执行的更新效果，即使这些更新尚未提交。）这比 SQL 标准对该隔离级别的要求更强，能够防止表 13.1中描述的所有现象。正如前面提到的，这正是标准明确允许的，因为标准只描述每个隔离级别必须提供的最低保护。

这一点与读已提交不同：在可重复读事务中，查询看到的是该事务开始时的快照，
而不是事务中当前查询开始时的快照。因此，在单个事务中的连续 SELECT 命令会看到相同的数据，也就是说，它们不会看到在自身事务开始后由其他事务提交的更改。

使用这一隔离级别的应用必须准备好在发生串行化失败时重试事务。

UPDATE、DELETE、SELECT FOR UPDATE 和 SELECT FOR SHARE 命令在搜索目标行方面与SELECT相同：它们只会找到在事务开始时已提交的目标行。然而，这样的目标行可能在找到时已被另一个并发事务更新（或删除或锁定）。在这种情况下，
可重复读事务将等待第一个更新事务提交或回滚（如果仍在进行）。如果第一个更新者回滚，
则其效果将被取消，可重复读事务可以继续更新最初找到的行。
但如果第一个更新者提交了（并实际更新或删除了行，而不仅仅是锁定了它），
那么可重复读事务将被回滚，并显示以下消息：

```
ERROR:  could not serialize access due to concurrent update
```

因为可重复读事务不能修改或锁定在其开始之后已被其他事务更改的行。

当应用收到这条错误消息时，应当中止当前事务，并从头重试整个事务。第二次执行时，
该事务会把先前已提交的更改视为其初始数据库视图的一部分，
因此以该行的新版本作为新事务更新的起点时，就不存在逻辑冲突。

注意，只有更新事务才可能需要重试；只读事务永远不会发生串行化冲突。

可重复读模式提供了严格的保证，即每个事务都看到数据库的一个完全稳定的视图。不过，
这个视图并不一定总能与同一级别并发事务的某种串行（一次一个）执行保持一致。例如，

即使是该级别的只读事务，也可能看到一条表明某批次已完成的控制记录，却
看不到逻辑上属于该批次的某条明细记录，
因为创建该明细记录的事务读取了控制记录的较早版本。
如果不仔细使用显式锁来阻塞冲突事务，试图依靠运行在这一隔离级别的事务来强制业务规则，
往往无法正确工作。

> **注意**
>
> 在 PostgreSQL 9.1 之前，请求可串行化事务隔离级别会得到与这里描述完全相同的行为。
> 若要保留旧式的可串行化行为，现在应请求可重复读。

13.2.3. 可串行化隔离级别

可串行化隔离级别提供最严格的事务隔离。该级别模拟串行事务执行；也就是说，
效果就像这些事务不是并发执行，而是一个接一个地串行执行一样。不过，
与可重复读级别一样，使用该级别的应用必须准备好在发生串行化失败时重试事务。实际上，
这个隔离级别的工作方式与可重复读完全相同，
只是它还会监视那些可能使一组并发可串行化事务的执行结果，
与这些事务任何可能的串行（一次一个）执行都不一致的条件。
这种监视不会引入超出可重复读已有阻塞之外的任何阻塞，但会带来一定开销；
一旦检测到可能导致串行化异常的条件，就会触发串行化失败。

例如，考虑表`mytab`，它最初包含以下内容：

```
 class | value
-------+-------
     1 |    10
     1 |    20
     2 |   100
     2 |   200
```

假设可串行化事务 A 执行以下计算：

```
SELECT SUM(value) FROM mytab WHERE class = 1;
```

然后将结果（30）作为新行的`value`插入，且该行满足`class = 2`。与此同时，可串行化事务 B 执行以下计算：

```
SELECT SUM(value) FROM mytab WHERE class = 2;
```

得到结果 300，并将它插入一个新行，该行满足`class = 1`。然后两个事务都尝试提交。如果任一事务运行在可重复读隔离级别，两个事务都可以提交；
但由于不存在与该结果一致的串行执行顺序，使用可串行化事务时，只会允许一个事务提交，
另一个则会被回滚，并收到以下消息：

```
ERROR:  could not serialize access due to read/write dependencies
 among transactions
```

这是因为，如果 A 在 B 之前执行，B 算出的总和应该是 330，而不是 300；同样，按另一种顺序执行，A 算出的总和也会不同。

为了保证真正的可串行性，PostgreSQL 使用了谓词锁，也就是说，系统会保留一些锁，以便判断某个写操作如果先发生，是否会影响并发事务先前读取的结果。在 PostgreSQL 中，这些锁不会造成任何阻塞，因此不会参与形成死锁。

它们用于识别并标记并发可串行化事务之间的依赖关系，
而这些依赖在某些组合下可能导致串行化异常。相比之下，
想要保证数据一致性的读已提交或可重复读事务，可能需要在整张表上加锁，
这会阻塞其他试图使用该表的用户；或者它可能需要使用 SELECT FOR UPDATE 或 SELECT FOR SHARE，而这些做法不仅可能阻塞其他事务，还会造成磁盘访问。

与大多数其他数据库系统一样，PostgreSQL 中的谓词锁以事务实际访问的数据为基础。这些锁会出现在pg_locks系统视图中，其 mode 为 SIReadLock。查询执行期间获取的具体锁取决于查询所使用的计划，并且在事务过程中，多个更细粒度的锁（如元组锁）
可能会合并成较少的粗粒度锁（如页锁），以防耗尽用于跟踪锁的内存。如果某个 READ ONLY 事务检测到不再可能发生会导致串行化异常的冲突，它可以在完成前释放其 SIRead 锁。事实上，READ ONLY 事务通常在启动时就能确定这一点，从而避免获取任何谓词锁。如果你显式请求一个 SERIALIZABLE READ ONLY DEFERRABLE 事务，它会阻塞，直到能够确立这一事实为止。（这是唯一一种可串行化事务会阻塞而可重复读事务不会阻塞的情况。）另一方面，SIRead 锁往往需要保留到事务提交之后，直到与之重叠的读写事务完成。

统一采用可串行化事务可以简化开发。任何一组并发可串行化事务，
都保证与把它们一次一个运行的效果相同；这意味着，
如果你能够证明某个事务在单独运行时会做正确的事，那么即使不了解其他事务会做什么，
也可以相信它在任何可串行化事务的混合执行中也会做正确的事。重要的是，
采用这种技术的环境应当有一套通用机制来处理串行化失败（其 SQLSTATE 值总是 '40001'），因为很难准确预测究竟哪些事务会对读/写依赖关系有贡献，并因此需要被回滚来防止串行化异常。
监视读/写依赖关系会带来开销，被串行化失败中止的事务重新启动也会带来开销；
但在把这些开销与显式锁以及 SELECT FOR UPDATE 或 SELECT FOR SHARE 所涉及的成本和阻塞进行权衡之后，可串行化事务在某些环境中仍然是性能最佳的选择。

当依赖可串行化事务进行并发控制时，为了获得最佳性能，应考虑以下问题：

- 在可能时声明事务为READ ONLY。
- 控制活动连接的数量，并在需要时使用连接池。这始终是重要的性能考量，
  但在使用可串行化事务的繁忙系统中尤为重要。
- 单个事务中只包含为保证完整性所必需的内容。
- 不要让连接不必要地“闲置在事务中”。
- 在那些已经由可串行化事务自动提供保护的地方，消除不再需要的显式锁、SELECT FOR UPDATE 和 SELECT FOR SHARE。
- 当系统因谓词锁表内存不足而被迫把多个页级谓词锁合并为单个关系级谓词锁时，
  串行化失败的比例可能会上升。你可以通过增加max_pred_locks_per_transaction来避免这种情况。
- 顺序扫描总是需要关系级谓词锁。这可能导致串行化失败的比例上升。通过降低random_page_cost和/或提高cpu_tuple_cost来鼓励使用索引扫描，可能对此有所帮助。
  务必在事务回滚和重启次数减少所带来的收益，与查询执行时间整体变化之间进行权衡。

13.3. 显式锁定

PostgreSQL 提供了多种锁模式，用于控制对表中数据的并发访问。在 MVCC 无法给出期望行为的场景中，这些模式可用于由应用自行控制的加锁。此外，大多数 PostgreSQL 命令也会自动获取适当模式的锁，以确保在命令执行期间，被引用的表不会以不兼容的方式被删除或修改。
（例如，TRUNCATE 无法安全地与同一张表上的其他操作并发执行，因此它会在该表上获取独占锁来强制实现这一点。）

要查看数据库服务器中当前尚未释放的锁列表，可以使用 pg_locks 系统视图。有关监控锁管理器子系统状态的更多信息，请参见第 27 章。

13.3.1. 表级锁

下面的列表给出了可用的锁模式，以及它们在 PostgreSQL 中被自动使用的场景。你也可以通过命令LOCK显式获取其中任意一种锁。请记住，这些锁模式全部都是表级锁，
即使名称中包含“row”一词；这些名称只是历史遗留。在某种程度上，
这些名称反映了各锁模式的典型用途 – 但它们的语义是完全相同的。一个锁模式与另一个锁模式真正的区别，只在于它与哪些锁模式冲突（见表 13.2）。两个事务不能在同一时刻在同一张表上持有相互冲突模式的锁。（不过，事务永远不会与自己冲突。例如，它可能先获取 ACCESS EXCLUSIVE 锁，随后又在同一张表上获取 ACCESS SHARE 锁。）不冲突的锁模式可以被多个事务同时持有。特别要注意，有些锁模式与自身冲突（例如，一次只能有一个事务持有 ACCESS EXCLUSIVE 锁），而另一些锁模式并不与自身冲突（例如，可以有多个事务持有 ACCESS SHARE 锁）。

表级锁模式

ACCESS SHARE

只与 ACCESS EXCLUSIVE 锁模式冲突。

SELECT 命令会在被引用的表上获取这种模式的锁。通常，任何只读取表而不修改它的查询都会获取这种锁模式。

ROW SHARE

与 EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。

SELECT FOR UPDATE 和 SELECT FOR SHARE 命令会在目标表上获取这种模式的锁（此外，对于引用到但没有用 FOR UPDATE/FOR SHARE 选取的其他表，还会获取 ACCESS SHARE 锁）。

ROW EXCLUSIVE

与SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE和ACCESS EXCLUSIVE锁模式冲突。

UPDATE、DELETE 和 INSERT 命令会在目标表上获取这种锁模式（此外，对任何其他被引用的表还会获取 ACCESS SHARE 锁）。通常，任何修改表中数据的命令都会获取这种锁模式。

SHARE UPDATE EXCLUSIVE

与 SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式用于保护表不受并发模式更改和 VACUUM 运行的影响。

由 VACUUM（不带 FULL）、ANALYZE、CREATE INDEX CONCURRENTLY 以及某些形式的 ALTER TABLE 获取。

SHARE

与 ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式用于保护表不受并发数据更改的影响。

由 CREATE INDEX（不带 CONCURRENTLY）获取。

SHARE ROW EXCLUSIVE

与 ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式用于保护表不受并发数据更改的影响，并且是自排他的，因此同一时刻只能有一个会话持有它。

这种锁模式不会被任何 PostgreSQL 命令自动获取。

EXCLUSIVE

与 ROW SHARE、ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式只允许并发的 ACCESS SHARE 锁，也就是说，只有对该表的读操作可以与持有这种锁模式的事务并行执行。

没有任何 PostgreSQL 命令会自动在表上获取这种锁模式。

ACCESS EXCLUSIVE

与所有模式的锁冲突（ACCESS SHARE、ROW SHARE、ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE）。这种模式保证持有者是以任何方式访问该表的唯一事务。

由 ALTER TABLE、DROP TABLE、TRUNCATE、REINDEX、CLUSTER 和 VACUUM FULL 命令获取。这也是未显式指定模式的 LOCK TABLE 语句的默认锁模式。

> **提示**
>
> 只有 ACCESS EXCLUSIVE 锁才会阻塞 SELECT（不带 FOR UPDATE/SHARE）语句。

锁一旦被获取，通常会一直持有到事务结束。但是，如果锁是在建立保存点之后才获取的，那么在回滚到该保存点时，这个锁会立即释放。这与 ROLLBACK 会取消保存点之后所有命令效果的原则是一致的。相同的原则也适用于在 PL/pgSQL 异常块中获取的锁：从该块中因错误跳出时，会释放在块中获取的锁。

表 13.2. 冲突的锁模式

请求的锁模式	当前的锁模式							
	ACCESS SHARE	ROW SHARE	ROW EXCLUSIVE	SHARE UPDATE EXCLUSIVE	SHARE	SHARE ROW EXCLUSIVE	EXCLUSIVE	ACCESS EXCLUSIVE
ACCESS SHARE								X
ROW SHARE							X	X
ROW EXCLUSIVE					X	X	X	X
SHARE UPDATE EXCLUSIVE				X	X	X	X	X

请求的锁模式	当前的锁模式							
	ACCESS SHARE	ROW SHARE	ROW EXCLUSIVE	SHARE UPDATE EXCLUSIVE	SHARE	SHARE ROW EXCLUSIVE	EXCLUSIVE	ACCESS EXCLUSIVE
SHARE			X	X		X	X	X
SHARE ROW EXCLUSIVE			X	X	X	X	X	X
EXCLUSIVE		X	X	X	X	X	X	X
ACCESS EXCLUSIVE	X	X	X	X	X	X	X	X

13.3.2. 行级锁

除了表级锁之外，还有行级锁，它可以是排他锁或共享锁。当某一行被更新或删除时，会自动获取该行上的排他行级锁。与表级锁一样，该锁会被持有到事务提交或回滚为止。行级锁不会影响数据查询；它们只会阻塞对同一行的写入者。

要在不实际修改某一行的情况下获取该行上的排他行级锁，可以使用`SELECT FOR UPDATE`选择该行。请注意，一旦获取了行级锁，事务就可以多次更新该行而不必担心冲突。

要获取某一行上的共享行级锁，可以使用`SELECT FOR SHARE`选择该行。共享锁不会阻止其他事务获取相同的共享锁。但是，任何事务都不允许更新、删除或排他锁定其他事务持有共享锁的行。任何这样的尝试都会被阻塞，直到共享锁被释放。

PostgreSQL 不会在内存中保存任何关于已修改行的信息，因此一次加锁的行数没有限制。不过，锁定一行可能会导致一次磁盘写；例如，`SELECT FOR UPDATE` 会修改被选中的行以标记它们已被锁定，因此会产生磁盘写入。

除了表级锁和行级锁之外，还使用页级共享/排他锁来控制对共享缓冲池中表页面的读写访问。这些锁会在取出或更新某一行后立即释放。应用开发者通常不需要关心页级锁，这里提到它们只是为了完整性。

13.3.3. 死锁

使用显式锁可能会增加死锁的发生概率。所谓死锁，是指两个（或更多）事务各自持有对方想要的锁。例如，如果事务 1 在表 A 上获得了排他锁，然后试图再获取表 B 上的排他锁，而事务 2 已经持有表 B 上的排他锁，现在又想获取表 A 上的排他锁，那么两者都无法继续进行。PostgreSQL 会自动检测死锁，并通过中止其中一个事务来解决这个问题，使其他事务得以继续完成。（具体会中止哪个事务很难预测，也不应依赖这种预测。）

还要注意，死锁也可能由于行级锁而发生（因此，即使没有使用显式锁，它们也可能出现）。考虑下面这种情况：两个并发事务都在修改一个表。第一个事务执行：

```
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 11111;
```

这会在指定账号的那一行上获取一个行级锁。然后第二个事务执行：

```
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 22222;
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 11111;
```

第一条 `UPDATE` 语句成功地在指定行上获取了一个行级锁，因此它成功更新了那一行。然而，第二条 `UPDATE` 语句发现它试图更新的行已经被锁住了，于是它等待持有该锁的事务结束。此时，事务二正在等待事务一结束之后才能继续执行。现在，事务一执行：

```
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 22222;
```

事务一试图在指定行上获取一个行级锁，但它做不到，因为事务二已经持有了这个锁。
于是它必须等待事务二完成。这样一来，事务一被事务二阻塞，
而事务二又被事务一阻塞：死锁条件形成了。PostgreSQL 会检测到这种情况，并中止其中一个事务。

防止死锁的最好办法通常是避免它们出现，
也就是确保所有使用同一数据库的应用都以一致的顺序在多个对象上获取锁。在上面的例子中，
如果两个事务都是按相同顺序更新这些行，就不会发生死锁。
还应确保事务在某个对象上获取的第一个锁，就是该对象所需的最严格锁模式。
如果事先无法验证这一点，那么就应通过重试因死锁而中止的事务来动态处理死锁。

只要没有检测到死锁，寻求表级锁或行级锁的事务就会无限期等待冲突锁被释放。
这意味着让应用长时间保持事务打开并不是好主意（例如等待用户输入时）。

13.3.4. 咨询锁

PostgreSQL 提供了一种创建由应用自行定义含义的锁的方法。这类锁被称为咨询锁，因为系统并不强制要求使用它们 – 是否正确使用完全取决于应用。咨询锁对于那些与 MVCC 模型不太契合的加锁策略非常有用。例如，咨询锁的一个常见用途是模拟所谓“平面文件”
数据管理系统中典型的悲观锁策略。虽然也可以通过在表中存储一个标志位来达到相同目的，
但咨询锁速度更快，可以避免表膨胀，并且会在会话结束时由服务器自动清理。

在 PostgreSQL 中获取咨询锁有两种方式：会话级和事务级。会话级咨询锁一旦获取，就会一直保持到显式释放或会话结束。不同于标准锁请求，
会话级咨询锁请求不遵守事务语义：在随后回滚的事务中获取的锁，回滚后仍会保持；同样，
即使发出解锁请求的事务后来失败，解锁依然有效。一个进程可以多次获取同一个锁；
每成功一次加锁请求，都必须有对应的解锁请求，锁才会真正释放。另一方面，
事务级锁请求更像普通锁请求：它们会在事务结束时自动释放，而且没有显式解锁操作。
对于短期使用咨询锁，这种行为通常比会话级行为更方便。
针对同一咨询锁标识符的会话级和事务级锁请求，会按预期相互阻塞。
如果某个会话已经持有给定的咨询锁，那么它发出的附加请求总会成功，
即使其他会话正在等待该锁；无论现有锁和新请求属于会话级还是事务级，都是如此。

和 PostgreSQL 中的所有锁一样，任何会话当前持有的咨询锁完整列表，都可以在`pg_locks`系统视图中找到。

咨询锁和普通锁都存储在一个共享内存池中，其大小由配置变量max_locks_per_transaction和max_connections定义。必须小心不要耗尽这部分内存，
否则服务器将根本无法再授予任何锁。这也为服务器可授予的咨询锁数量设定了上限，
具体通常在数万到数十万个之间，取决于服务器的配置。

在某些使用咨询锁的方法中，特别是涉及显式排序和 `LIMIT` 子句的查询，必须注意控制由于 SQL 表达式求值顺序而获取的锁。例如：

```
SELECT pg_advisory_lock(id) FROM foo WHERE id = 12345; -- 可以
SELECT pg_advisory_lock(id) FROM foo WHERE id > 12345 LIMIT 100; -- 危险!
SELECT pg_advisory_lock(q.id) FROM
(
  SELECT id FROM foo WHERE id > 12345 LIMIT 100
```

) q; -- 可以

在上面的查询中，第二种形式是危险的，因为不能保证在执行加锁函数之前先应用 LIMIT。这可能导致获取了一些应用并未预期的锁，因此应用也不会去释放它们（直到会话结束）。从应用的角度看，这类锁会变成悬而未清的锁，尽管仍然可以在 pg_locks 中看到。

用于操作咨询锁的函数见表 9.63。

13.4. 应用级别的数据一致性检查

用读已提交事务强制执行有关数据完整性的业务规则非常困难，因为数据视图会随每条语句而变化，而且一旦发生写冲突，即使是单条语句也未必局限于该语句自己的快照。

虽然可重复读事务在整个执行期间都具有稳定的数据视图，但在使用 MVCC 快照进行数据一致性检查时，还存在一个微妙问题，即所谓的读/写冲突。如果一个事务写入数据，而并发事务试图读取相同的数据（无论是在写入之前还是之后），它都看不到另一个事务的工作。于是，读取者看起来就像是先执行的一方，而不管究竟是谁先启动，也不管谁先提交。如果事情只停留在这一步，就没有问题；但如果读取者还写入了数据，而这些数据又被另一个并发事务读取，那么现在就会有一个事务看起来像是在前面提到的任一事务之前执行。如果看起来最后执行的事务实际上最先提交，就很容易在事务执行顺序图中形成一个环。一旦出现这样的环，不借助额外机制，完整性检查就无法正确工作。

正如第 13.2.3 节中提到的，可串行化事务其实就是在可重复读事务的基础上，增加了对危险读/写冲突模式的非阻塞监控。当检测到某种模式可能在表面执行顺序中形成一个环时，其中一个相关事务就会被回滚，以打破这个环。

13.4.1. 用可串行化事务强制一致性

如果对所有写操作，以及所有需要一致数据视图的读操作都使用可串行化事务隔离级别，那么不需要额外工作就能确保一致性。从其他环境移植而来、按使用可串行化事务来保证一致性而编写的软件，在这方面应当能够在 PostgreSQL 中“正常工作”。

使用这种技术时，如果应用软件通过某个框架运行，并由该框架自动重试因串行化失败而回滚的事务，就可以避免给应用程序员带来不必要的负担。把 default_transaction_isolation 设置为 serializable 可能是个好主意。通过在触发器中检查事务隔离级别来采取某些措施，以确保不会因为疏忽或为了绕过完整性检查而使用其他事务隔离级别，也是明智的。

性能方面的建议见第 13.2.3 节。

> **警告**
>
> 利用可串行化事务提供的这一层完整性保护，尚未扩展到热备模式（第 25.5 节）。因此，使用热备的用户可能希望在主库上使用可重复读和显式锁定。

13.4.2. 使用显式阻塞锁强制一致性

当存在非可串行化写入时，要确保某一行当前仍然有效，并保护它不受并发更新影响，就必须使用 SELECT FOR UPDATE、SELECT FOR SHARE 或适当的 LOCK TABLE 语句。

（SELECT FOR UPDATE 和 SELECT FOR SHARE 只锁定返回的行以防止并发更新，而 LOCK TABLE 会锁住整张表。）从其他环境向 PostgreSQL 迁移应用时，应当考虑这一点。

对于从其他环境迁移而来的用户，还需要注意：SELECT FOR UPDATE 并不能保证并发事务不会更新或删除被选中的行。要在 PostgreSQL 中做到这一点，必须真正去更新该行，即使没有任何值需要改变。SELECT FOR UPDATE 只是临时阻塞其他事务，使它们不能获取同样的锁，也不能执行会影响被锁定行的 UPDATE 或 DELETE；但是一旦持有该锁的事务提交或回滚，被阻塞的事务就会继续执行冲突操作，除非在持锁期间已经对该行执行了实际的 UPDATE。

在非可串行化的 MVCC 环境下，全局有效性检查需要额外考虑。例如，一个银行应用可能希望检查一个表中的贷方金额总和等于另一个表中的借方金额总和，而这两个表都在被活跃更新。
在读已提交模式下，比较两个连续的 SELECT sum(...) 命令的结果并不可靠，因为第二个查询很可能会包含第一个查询没有统计到的事务提交结果。在单个可重复读事务中完成这两次求和，只能准确反映在该可重复读事务开始之前已提交事务的效果 – 但等到结果交付时，人们完全可能合理地怀疑这个答案是否仍然相关。
如果可重复读事务本身在尝试进行一致性检查之前已经应用了某些更改，
那么这种检查的意义就更加值得商榷，因为此时它包含了事务开始后的一部分而不是全部更改。
在这种情况下，谨慎的做法可能是锁定执行检查所需的所有表，
以获得当前真实状态的无可争议图景。SHARE 模式（或更高）的锁能够保证，在被锁定的表中，除了当前事务自身的更改之外，没有其他未提交更改。

注意，如果依赖显式锁定来防止并发更改，就应当使用读已提交模式，
或者在可重复读模式下小心地在执行查询之前先获取锁。
可重复读事务获得的锁能够保证没有其他修改该表的事务仍在运行，
但如果该事务看到的快照早于获取锁的时刻，
那么它看到的快照也可能早于表中某些现在已经提交的更改。
可重复读事务的快照实际上是在其第一条查询或数据修改命令（SELECT、INSERT、UPDATE、DELETE）开始时被冻结的，因此可以在快照冻结之前显式获取锁。

13.5. 注意事项

一些 DDL 命令（目前只有TRUNCATE和会重写表的ALTER TABLE形式）不是 MVCC 安全的。这意味着在截断或重写提交之后，如果并发事务使用的是在该 DDL 命令提交之前取得的快照，那么该表对它们来说会表现为空表。这只会对那些在 DDL 命令开始前没有访问过相关表的事务造成问题 – 任何在 DDL 命令开始前访问过该表的事务，都会持有至少一个 ACCESS SHARE 表锁，从而阻塞该 DDL 命令直到该事务完成。因此，这些命令不会在针对目标表的连续查询中造成明显的表内容不一致，但它们可能会导致目标表内容与数据库中其他表内容之间出现可见的不一致。

对可串行化事务隔离级别的支持尚未添加到热备复制目标（见第 25.5 节）。当前热备模式下支持的最严格隔离级别是可重复读。
虽然在主库上把所有永久数据库写操作都放在可串行化事务中执行，
能够确保所有备库最终达到一致状态，但在备库上运行的可重复读事务，
有时仍可能看到一种瞬态状态，而这种状态与主库上事务的任何串行执行都不一致。

13.6. 锁定和索引

尽管 PostgreSQL 为表数据提供了非阻塞的读写访问，但 PostgreSQL 当前实现的索引访问方法并不是每一种都能提供非阻塞的读写访问。各种索引类型的处理方式如下：

B-树和 GiST 索引

读写访问使用短期的页级共享/排他锁。每个索引行在被取出或插入之后，锁都会立即释放。这些索引类型提供最高的并发性，并且不会产生死锁。

Hash 索引

读写访问使用的是 Hash 桶级共享/排他锁。锁会在整个 Hash 桶处理完成后释放。桶级锁比索引级锁具有更好的并发性，但也可能产生死锁，
因为它们的持有时间比一次索引操作更长。

GIN 索引

读写访问使用短期的页级共享/排他锁。每个索引行在被取出或插入之后，锁都会立即释放。但要注意，插入一个使用 GIN 索引的值通常会导致每行产生多个索引键插入，因此 GIN 可能为了插入单个值而做大量工作。

目前，B-树索引为并发应用提供了最佳性能；由于它们还比 Hash 索引拥有更多特性，因此对于需要为标量数据建立索引的并发应用，推荐使用 B-树索引类型。处理非标量数据时，B-树就不再适用，此时应改用 GiST 或 GIN 索引。

第 14 章 性能提示

查询性能可能受许多因素影响。其中一些因素可以由用户控制，
另一些则属于系统底层设计的基本特性。本章提供一些帮助理解和调优PostgreSQL性能的提示。

14.1. 使用EXPLAIN

PostgreSQL会为收到的每个查询制定一个查询计划。
选择与查询结构和数据特性相匹配的正确计划，对获得良好的性能至关重要，
因此系统内置了一个复杂的规划器来尽量选出好的计划。
可以使用EXPLAIN命令查看规划器为任意查询生成的查询计划。
读懂计划是一门值得用长篇教程来阐述的技艺，本节并非那样的教程，
这里只给出一些基本信息。

查询计划的结构是一棵由计划节点组成的树。树的最底层是表扫描节点，
它们从表中返回原始行。不同的表访问方法对应不同类型的扫描节点，例如顺序扫描、
索引扫描和位图索引扫描。如果查询需要对原始行执行连接、聚合、排序或其他操作，
那么扫描节点之上还会出现额外的节点来完成这些操作。同样，
这些操作通常不止一种实现方式，因此这里也会出现不同的节点类型。EXPLAIN
会为计划树中的每个节点输出一行，
显示基本节点类型以及规划器对该计划节点执行代价的估计值。第一行（最顶层节点）
给出了整个计划的估计总执行代价；规划器力图最小化的正是这个数字。

下面是一个简单的示例，仅用于展示输出的样子：[1]

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..458.00 rows=10000 width=244)
```

EXPLAIN引用的数字从左到右依次表示：

- 估计启动开销（输出扫描开始前需要消耗的时间，例如排序节点执行排序所需的时间）
- 估计总开销（假定取回所有行，尽管可能并不会全部取回；例如，带LIMIT
  子句的查询将不会付清Limit计划节点的输入节点的全部代价）
- 该计划节点输出行数的估计值（同样，也是假定该节点会运行到结束）
- 预计该计划节点输出行的平均宽度，以字节计

这些开销使用由规划器代价参数决定的任意单位来衡量（见第 18.7.2 节）。传统上通常以磁盘页
读取作为代价单位；也就是说，惯例上将seq_page_cost设为1.0，
其他代价参数都相对它来设定。（本节中的示例都使用默认代价参数运行。）

需要注意的是，上层节点的开销包含了其所有子节点的开销。还要注意，
这个开销只反映规划器关心的内容。特别是，它没有考虑将结果行传输给客户端所消耗的时间，
而这些在实际耗时中可能是重要因素；但规划器会忽略这些代价，
因为它无法通过改变计划来影响它们。（我们相信每个正确的计划都会输出相同的行集。）

[1] 本节中的示例取自执行过VACUUM ANALYZE的回归测试数据库，使用的是 8.2 开发版本源码。如果自行尝试这些示例，通常应当能得到相近的结果，但估计代价和行计数可能略有差异，因为ANALYZE生成的统计信息来自随机采样而非精确计数。

rows值有些容易误解，因为它不是计划节点处理或扫描过的行数。它通常更小，反映了在该节点上应用的WHERE子句条件的估计选择度。理想情况下，顶层的行数估计应当接近查询实际返回、更新或删除的行数。

回到我们的示例：

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..458.00 rows=10000 width=244)
```

这已经是再直接不过的了。如果执行：

```
SELECT relpages, reltuples FROM pg_class WHERE relname = 'tenk1';
```

可以看到tenk1有 358 个磁盘页和 10000 行。估计代价按如下公式计算：（读取的磁盘页数 * seq_page_cost）+（扫描的行数 * cpu_tuple_cost）。默认情况下，seq_page_cost为 1.0，cpu_tuple_cost为 0.01，因此估计代价就是 (358 * 1.0) + (10000 * 0.01) = 458。

现在修改原始查询，添加一个WHERE条件：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 7000;

                         QUERY PLAN
------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..483.00 rows=7033 width=244)
   Filter: (unique1 < 7000)
```

注意，EXPLAIN的输出表明，WHERE子句被作为一个“过滤器”条件；这表示计划节点会对扫描到的每一行检查该条件，并且只输出满足条件的行。由于有了WHERE子句，估计输出行数减少了。但扫描仍需访问全部 10000 行，因此代价没有下降；实际上还略有上升（精确地说，增加了 10000 * cpu_operator_cost），以反映检查WHERE条件所消耗的额外 CPU 时间。

这条查询实际选出的行数是 7000，但估计的rows只是近似值。如果重复这个实验，很可能会得到略有不同的估计值。此外，由于ANALYZE生成的统计信息来自该表的随机采样，这个估计值可能在每次执行ANALYZE之后发生变化。

现在让条件更严格一些：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 100;

                                  QUERY PLAN
-----------------------------------------------------------------------
--------
 Bitmap Heap Scan on tenk1  (cost=2.37..232.35 rows=106 width=244)
   Recheck Cond: (unique1 < 100)
   ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37 rows=106
 width=0)
         Index Cond: (unique1 < 100)
```

这里规划器决定使用两步计划：底层计划节点访问索引，找出匹配索引条件的行的位置，然后上层计划节点再从表本身取出这些行。分别取出各行的代价远高于顺序读取，但由于不必访问表的所有页，这仍然比顺序扫描便宜。（使用两层计划的原因是，上层计划节点在读取行之前，会先把索引确定的行位置按物理顺序排列，以尽量降低逐行读取的代价。节点名称中的“位图”就是执行这种排序的机制。）

如果WHERE条件的选择性足够强，规划器可能改用一种“简单”索引扫描计划：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 3;

                                 QUERY PLAN
-----------------------------------------------------------------------
--------
 Index Scan using tenk1_unique1 on tenk1  (cost=0.00..10.00 rows=2
 width=244)
   Index Cond: (unique1 < 3)
```

在这种计划中，表行按索引顺序取出，因此读取代价更高，但行数很少，
不值得为排序行位置付出额外代价。对于仅取出一行的查询，以及带有与索引顺序匹配的ORDER BY条件的查询，最常见到这种计划。

再向WHERE子句添加另一个条件：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 3 AND stringu1 = 'xxx';

                                 QUERY PLAN
-----------------------------------------------------------------------
--------
 Index Scan using tenk1_unique1 on tenk1  (cost=0.00..10.01 rows=1
 width=244)
   Index Cond: (unique1 < 3)
   Filter: (stringu1 = 'xxx'::name)
```

新增的条件stringu1 = 'xxx'降低了估计输出行数，却没有降低代价，因为仍然必须访问同一组行。注意，stringu1子句不能用作索引条件（因为该索引只建立在unique1列上）。
它只能作为过滤条件，应用到通过索引取出的行上。因此，代价实际上略有增加，
以反映这项额外检查。

如果在WHERE引用的多个列上建有索引，规划器可能选择将这些索引做 AND 或 OR 组合：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 100 AND unique2 > 9000;

                                   QUERY PLAN
-----------------------------------------------------------------------
---------------
 Bitmap Heap Scan on tenk1  (cost=11.27..49.11 rows=11 width=244)
   Recheck Cond: ((unique1 < 100) AND (unique2 > 9000))
   ->  BitmapAnd  (cost=11.27..11.27 rows=11 width=0)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0)
               Index Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique2  (cost=0.00..8.65
 rows=1042 width=0)
               Index Cond: (unique2 > 9000)
```

但这需要访问两个索引，因此与只使用一个索引、把另一个条件作为过滤条件相比，
并不一定更划算。如果调整条件中的范围，就会看到计划相应改变。

使用之前讨论的列，尝试连接两个表：

```
EXPLAIN SELECT *
```

```
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;

                                       QUERY PLAN
-----------------------------------------------------------------------
----------------
 Nested Loop  (cost=2.37..553.11 rows=106 width=488)
   ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244)
         Recheck Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0)
               Index Cond: (unique1 < 100)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..3.01
 rows=1 width=244)
         Index Cond: (unique2 = t1.unique2)
```

在这个嵌套循环连接中，外侧（上层）扫描与我们前面看到的位图索引扫描相同，因此其代价和行计数也相同，因为`WHERE`子句`unique1 < 100`正是在该节点上应用的。`t1.unique2 = t2.unique2`子句此时还无关，因此不会影响外侧扫描的行计数。对于内侧（下层）扫描，当前外侧行的`unique2`值会被代入内侧索引扫描，产生一个形如`unique2 = constant`的索引条件。因此，得到的内侧扫描计划和代价与例如`EXPLAIN SELECT * FROM tenk2 WHERE unique2 = 42`的结果相同。随后，循环节点的代价建立在外侧扫描代价之上，再加上每个外侧行都要执行一次内侧扫描的代价（这里是 106 * 3.01），以及少量连接处理的 CPU 时间。

本例中，连接的输出行数等于两个扫描行数的乘积，但并非所有情况都如此，因为可能有同时涉及两个表的`WHERE`子句，因此只能在连接处应用，不能应用到任一输入扫描。例如，如果添加`WHERE ... AND t1.hundred < t2.hundred`，将减少连接节点的输出行数，但不改变任一输入扫描。

查看其他候选计划的一种方法，是使用第 18.7.1 节中介绍的启用/禁用标志，强制规划器忽略它认为代价最低的策略。（这是一种粗略但有用的工具。另请参见第 14.3 节。）

```
SET enable_nestloop = off;
EXPLAIN SELECT *
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;

                                         QUERY PLAN
-----------------------------------------------------------------------
--------------------
 Hash Join  (cost=232.61..741.67 rows=106 width=488)
   Hash Cond: (t2.unique2 = t1.unique2)
   ->  Seq Scan on tenk2 t2  (cost=0.00..458.00 rows=10000 width=244)
   ->  Hash  (cost=232.35..232.35 rows=106 width=244)
         ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244)
               Recheck Cond: (unique1 < 100)
               ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.
37 rows=106 width=0)
                     Index Cond: (unique1 < 100)
```

这个计划提议用同一个旧索引扫描取出`tenk1`的 100 个感兴趣的行，把它们存入一个内存哈希表，然后对`tenk2`做顺序扫描，并为每个`tenk2`行探查哈希表中`t1.unique2 =`

t2.unique2的可能匹配。读取tenk1并建立哈希表的代价是哈希连接的启动代价，因为在开始读取tenk2之前不会有任何输出。连接的总时间估计还包括一笔可观的 CPU 时间，用于探查哈希表 10000 次。但请注意，我们不是按 10000 次 232.35 计费；在这种计划类型中，哈希表的建立只进行一次。

可以通过EXPLAIN ANALYZE来检验规划器估计代价的准确性。这个命令会真正执行查询，然后显示在每个计划节点内累计的真实运行时间，以及普通EXPLAIN所显示的相同估计代价。例如，我们可能得到这样的结果：

```
EXPLAIN ANALYZE SELECT *
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;

                                                                  QUERY PLAN
-----------------------------------------------------------------------------
-------------------------------------------------------------
 Nested Loop  (cost=2.37..553.11 rows=106 width=488) (actual time=1.
392..12.700 rows=100 loops=1)
   ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244) (actual time=0.878..2.367 rows=100 loops=1)
         Recheck Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0) (actual time=0.546..0.546 rows=100 loops=1)
               Index Cond: (unique1 < 100)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..3.01
 rows=1 width=244) (actual time=0.067..0.078 rows=1 loops=100)
         Index Cond: (unique2 = t1.unique2)
 Total runtime: 14.452 ms
```

注意，“actual time”值是实际时间的毫秒数，而cost估计值则以任意单位表示，因此两者不太可能直接吻合。值得关注的是实际时间与估计代价之间的比率是否一致。

在某些查询计划中，一个子计划节点可能会执行多次。例如，
上面那个嵌套循环计划中的内侧索引扫描会对外侧的每一行执行一次。在这种情况下，loops值报告的是该节点的总执行次数，而 actual time 和 rows 显示的是每次执行的平均值。这样做是为了让这些数字更容易与代价估计的展示方式相比较。将它们乘以loops值，
就能得到该节点实际消耗的总时间。

EXPLAIN ANALYZE显示的Total runtime包含执行器的启动和关闭时间，但不包括解析、重写或计划时间。对于INSERT、UPDATE和DELETE命令，应用表更改所花费的时间被记入顶层的 Insert、Update 或 Delete 计划节点。（该节点之下的计划节点表示定位旧行和/或计算新行的工作。）如果有BEFORE触发器，执行它们所花费的时间被记入相关的 Insert、Update 或 Delete 节点，而执行AFTER触发器的时间则不会。每个触发器（BEFORE或AFTER）所花费的时间也会单独显示，并计入总运行时间。但要注意，延迟约束触发器直到事务结束才会执行，
因此不会被EXPLAIN ANALYZE显示。

EXPLAIN ANALYZE测得的运行时间，可能以两种重要方式偏离同一查询的正常执行时间。第一，由于不会向客户端传递任何输出行，因此网络传输开销和 I/O 格式化开销均不会被计入。第二，EXPLAIN ANALYZE增加的开销本身可能相当可观，尤其是在gettimeofday()内核调用较慢的机器上。

值得注意的是，不应将EXPLAIN结果外推到与实际测试不同的场景。例如，
在一个很小的表上得到的结果，不能假定也适用于大型表。规划器的代价估计不是线性的，
因此它可能会为更大或更小的表选择不同的计划。

一个极端示例是：对于只占用一个磁盘页的表，无论索引是否可用，
几乎总会得到顺序扫描计划。规划器认识到，无论如何处理该表都需要一次磁盘页读取，
因此再额外读取页面去查看索引并没有价值。

14.2. 规划器使用的统计信息

如上一节所见，查询规划器需要估计查询将检索多少行，才能对查询计划做出良好选择。
本节简要介绍系统用于这些估计的统计信息。

统计信息的一部分是各个表和索引的条目总数，以及各个表和索引占用的磁盘块数。
这些信息保存在表pg_class的reltuples和relpages列中。
可以使用类似下面的查询来查看：

```
SELECT relname, relkind, reltuples, relpages
FROM pg_class
WHERE relname LIKE 'tenk1%';

       relname        | relkind | reltuples | relpages
----------------------+---------+-----------+----------
 tenk1                | r       |     10000 |      358
 tenk1_hundred        | i       |     10000 |       30
 tenk1_thous_tenthous | i       |     10000 |       30
 tenk1_unique1        | i       |     10000 |       30
 tenk1_unique2        | i       |     10000 |       30
(5 rows)
```

这里可以看到，tenk1包含 10000 行，它的索引也一样，但索引比表小得多（这并不意外）。

出于效率考虑，reltuples和relpages不会实时更新，因此它们通常包含有些过时的值。
它们会在VACUUM、ANALYZE以及少数 DDL 命令（如CREATE INDEX）执行时被更新。独立执行的ANALYZE（即不属于VACUUM的那种）由于不会读取表中的每一行，生成的是reltuples的近似值。无论如何，规划器都会将它在pg_class中找到的值按当前物理表大小进行缩放，
从而得到更接近实际情况的近似值。

大多数查询只会检索表中一部分行，因为它们通过WHERE子句限制了需要检查的行。因此，
规划器需要估算WHERE子句的选择度，也就是满足WHERE子句中各个条件的行所占比例。
完成这项任务所需的信息存储在pg_statistic系统目录中。pg_statistic
中的条目由ANALYZE和VACUUM ANALYZE命令更新，而且即使刚更新完，也始终只是近似值。

手动检查统计信息时，与其直接查看pg_statistic，不如查看它的视图pg_stats。pg_stats旨在让统计信息更易读。此外，pg_stats所有用户都可以读取，而pg_statistic
只有超级用户可以读取。
（这可以防止没有相应权限的用户通过统计信息获知其他用户表中的内容。pg_stats
视图被限制为只显示当前用户有权读取的表的相关行。）例如，可以执行：

```
SELECT attname, inherited, n_distinct,
       array_to_string(most_common_vals, E'\n') as most_common_vals
FROM pg_stats
WHERE tablename = 'road';

 attname | inherited | n_distinct |          most_common_vals
```

```
---------+-----------+------------+----------------------------------
--
 name    | f         |  -0.363388 | I- 580                        Ramp
+
         |           |            | I- 880                        Ramp
+
         |           |            | Sp Railroad
 +
         |           |            | I- 580
 +
         |           |            | I- 680                        Ramp
 name    | t         |  -0.284859 | I- 880                        Ramp
+
         |           |            | I- 580                        Ramp
+
         |           |            | I- 680                        Ramp
+
         |           |            | I- 580
 +
         |           |            | State Hwy 13                  Ramp
(2 rows)
```

注意，同一列显示了两行，其中一行对应从road表开始的整个继承层次结构（inherited =t），另一行只包含road表本身（inherited=f）。

ANALYZE在pg_statistic中存储的信息量，特别是每列most_common_vals中的最大项数和histogram_bounds数组的大小，可以使用ALTER TABLE SET STATISTICS命令按列设置，也可以通过设置配置变量default_statistics_target进行全局设置。目前默认上限是 100 项。提高这一上限可能让规划器做出更准确的估计，尤其是对于数据分布不规则的列；代价则是pg_statistic占用更多空间，并且计算估计值所需时间也会略有增加。相反，对于数据分布较简单的列，较低的上限可能已经足够。

更多规划器对统计信息的使用可参阅第 57 章。

14.3. 用显式JOIN子句控制规划器

可以在一定程度上用显式JOIN语法控制查询规划器。要理解这一点为何重要，先需要一些背景知识。

在一个简单的连接查询中，例如：

```
SELECT * FROM a, b, c WHERE a.id = b.id AND b.ref = c.id;
```

规划器可以自由地按任意顺序连接这些表。例如，它可以生成一个查询计划，先利用WHERE条件a.id = b.id将 A 连接到 B，然后再利用另一个WHERE条件把 C 连接到这个结果上。也可以先连接 B 和 C，再把 A 连接到所得结果上。甚至还可以先连接 A 和 C，再与 B 连接，但这样效率会很低，因为必须先形成 A 和 C 的完整笛卡尔积，而WHERE子句中并没有可用于优化这一连接的条件。（PostgreSQL执行器中的所有连接都发生在两个输入表之间，因此结果必须以这些形式之一逐步构造出来。）关键在于，这些不同的连接可能性在语义上是等价的，但执行代价可能相差极大。因此，规划器会探索它们，力图找出最高效的查询计划。

当查询只涉及两个或三个表时，需要考虑的连接顺序并不多。但可能的连接顺序数量会随着表数增加而呈指数增长。输入表超过十个左右之后，

对所有可能性做穷举搜索实际上就不再可行，甚至六七个表也可能让规划耗时长得令人厌烦。输入表过多时，PostgreSQL规划器会从穷举搜索切换到一种遗传概率搜索，只考虑有限数量的可能性。（切换阈值由运行时参数geqo_threshold控制。）遗传搜索耗时更少，但并不一定能找到最优计划。

当查询涉及外连接时，规划器比处理普通（内）连接时拥有更小的自由度。例如，考虑：

```
SELECT * FROM a LEFT JOIN (b JOIN c ON (b.ref = c.id)) ON (a.id = b.
id);
```

尽管这个查询的约束表面上与前一个非常相似，但它们的语义不同，因为如果 A 中有某一行无法匹配 B 和 C 连接结果中的任何行，该行仍然必须被输出。因此这里规划器对连接顺序没有选择：它必须先连接 B 和 C，再把 A 连接到该结果上。相应地，这个查询比前一个查询需要更少的规划时间。在其他情况下，规划器可能会判断多种连接顺序都是安全的。例如：

```
SELECT * FROM a LEFT JOIN b ON (a.bid = b.id) LEFT JOIN c ON (a.cid =
 c.id);
```

将 A 首先连接到 B 或 C 都是有效的。当前，只有`FULL JOIN`完全约束连接顺序。大多数涉及`LEFT JOIN`或`RIGHT JOIN`的实际情况都在某种程度上可以被重新排列。

显式连接语法（`INNER JOIN`、`CROSS JOIN`或无修饰的`JOIN`）在语义上和`FROM`中列出输入关系是一样的，因此它不约束连接顺序。

即使大多数类型的`JOIN`并不会完全约束连接顺序，仍然可以指示PostgreSQL查询规划器将所有`JOIN`子句都当作带有连接顺序约束来处理。例如，下面三个查询在逻辑上是等价的：

```
SELECT * FROM a, b, c WHERE a.id = b.id AND b.ref = c.id;
SELECT * FROM a CROSS JOIN b CROSS JOIN c WHERE a.id = b.id AND b.ref
 = c.id;
SELECT * FROM a JOIN (b JOIN c ON (b.ref = c.id)) ON (a.id = b.id);
```

但如果告诉规划器遵循`JOIN`顺序，那么第二个和第三个查询在规划上会比第一个花费更少时间。对于只有三个表的连接来说，这种效果微不足道；但当表很多时，它可能非常关键。

要强制规划器遵循显式`JOIN`给出的连接顺序，可以将运行时参数join_collapse_limit设置为 1。（其他可能值见下文讨论。）

不必为了缩短搜索时间而完全约束连接顺序，因为可以在普通`FROM`列表的某一项中使用`JOIN`操作符。例如：

```
SELECT * FROM a CROSS JOIN b, c, d, e WHERE ...;
```

如果设置`join_collapse_limit = 1`，就会强制规划器先将 A 连接到 B，然后再与其他表连接，但不会进一步约束它的选择。在这个示例中，可能的连接顺序数量减少了 5 倍。

以这种方式约束规划器的搜索，是一种既可减少规划时间、又可引导规划器生成更好查询计划的实用技巧。如果规划器默认选择了糟糕的连接顺序，可以通过`JOIN`语法强制它采用更好的顺序，前提当然是确实知道哪个顺序更好。建议进行实验。

与此密切相关、同样会影响规划时间的另一个问题，是将子查询折叠进父查询。例如：

```
SELECT *
FROM x, y,
```

```
    (SELECT * FROM a, b, c WHERE something) AS ss
WHERE somethingelse;
```

这种情况可能出现在使用包含连接的视图时；该视图的SELECT规则会被插入到引用视图的位置，从而得到一个与上面非常相似的查询。通常，规划器会尝试把子查询折叠进父查询，得到：

```
SELECT * FROM x, y, a, b, c WHERE something AND somethingelse;
```

这通常会生成比单独规划子查询更好的计划。（例如，外层WHERE条件可能会先把 X 连接到 A，从而消除 A 中的大量行，也就避免了形成子查询完整逻辑输出的需要。）但与此同时，规划时间也增加了；这里把两个彼此独立的三路连接问题，替换成了一个五路连接问题。
由于可能性数量呈指数增长，这种差别可能非常大。为了避免陷入巨大的连接搜索问题，
如果折叠子查询会使父查询产生超过from_collapse_limit个FROM项，
规划器就会尝试通过停止提升子查询来规避这一点。可以通过调高或调低这个运行时参数，
在规划时间和计划质量之间做权衡。

from_collapse_limit和join_collapse_limit名称相似，
因为它们做的几乎是同一件事：一个控制规划器何时将子查询“平面化”，
另一个控制何时将显式连接平面化。通常，要么把join_collapse_limit设置为与from_collapse_limit相同，这样显式连接与子查询的行为类似；要么把join_collapse_limit设置为 1，如果想使用显式连接控制连接顺序。不过，也可以把它们设置为不同的值，以更细致地调节规划时间与运行时间之间的平衡。

14.4. 填充一个数据库

初次填充数据库时，可能需要插入大量数据。本节给出一些使这一过程尽可能高效的建议。

14.4.1. 禁用自动提交

使用多个INSERT时，应关闭自动提交，只在最后提交一次。（在普通 SQL 中，这意味着开始时发出BEGIN，结束时发出COMMIT。某些客户端库可能会替调用方完成这件事，
在这种情况下需要确认它们确实会在所需的时候这样做。）如果允许每次插入都单独提交，
PostgreSQL就必须为每一行的加入执行大量额外工作。
把所有插入都放在一个事务中的另一个好处是：如果其中某一行插入失败，
那么此前插入的所有行都会被回滚，这样就不会留下部分装载的数据。

14.4.2. 使用COPY

使用COPY在一条命令中装载所有记录，而不是使用一系列INSERT命令。COPY
命令针对装载大量行做了优化；它不如INSERT灵活，但在大规模数据装载时开销显著更小。由于COPY是一条单独的命令，因此采用这种方法填充表时无须关闭自动提交。

如果不能使用COPY，那么用PREPARE创建一个预备INSERT语句，再按需多次执行EXECUTE
也会有所帮助。这样可以避免重复解析和规划INSERT的开销。
不同接口以不同方式提供这一功能，可参阅接口文档中关于“预备语句”的说明。

请注意，在装载大量行时，使用COPY几乎总是比使用INSERT更快，即使已经使用了PREPARE，并把多次插入批量放入同一个事务中也是如此。

当COPY与更早的CREATE TABLE或TRUNCATE命令处于同一事务中时，速度最快。在这种情况下，不需要写 WAL，因为一旦出错，包含新装载数据的文件反正也会被移除。不过，这一点只适用于wal_level为minimal的非分区表；否则所有命令都必须写 WAL。

14.4.3. 移除索引

如果正在装载一个新创建的表，最快的方法是先创建表，用COPY批量装载数据，
然后再创建该表所需的索引。在已有数据的表上创建索引，
要比在每行装载时对索引做增量更新更快。

如果正在向现有表加入大量数据，那么删除索引、装载数据、再重建索引可能是更好的方案。
当然，在索引缺失期间，其他数据库用户的性能可能会下降。删除唯一索引之前也必须慎重，
因为唯一约束提供的错误检查会在索引缺失期间丧失。

14.4.4. 移除外键约束

与索引类似，“批量”检查外键约束比逐行检查更高效。因此，先删除外键约束、装载数据、
再重建约束可能很有用。同样，
这里也需要在装载速度与约束缺失期间失去错误检查之间做权衡。

更重要的是，当在已有外键约束的情况下向表中装载数据时，
每一行新数据都需要在服务器待处理的触发器事件列表中占一个条目，
因为外键约束检查是通过触发器触发完成的。
装载数百万行可能导致触发器事件队列溢出可用内存，造成无法接受的交换，
甚至让命令直接失败。因此在装载大量数据时，删除并重新应用外键可能是必须的，
而不仅仅是期望如此。如果不能临时移除约束，
唯一的替代办法可能就是把装载操作拆分成更小的事务。

14.4.5. 增加maintenance_work_mem

在装载大量数据时，临时增大maintenance_work_mem配置变量可以提升性能。
这个参数也有助于加速CREATE INDEX和ALTER TABLE ADD FOREIGN KEY命令。它对COPY
本身帮助不大，因此这个建议只有在采用前述一种或两种技巧时才有意义。

14.4.6. 增加checkpoint_segments

临时增大checkpoint_segments配置变量，也可以让大规模数据装载更快。这是因为向
PostgreSQL中装载大量数据，会导致检查点比平常更频繁地发生，而正常频率由checkpoint_
timeout配置变量指定。每次发生检查点时，所有脏页都必须刷写到磁盘。
通过在批量装载期间临时增大checkpoint_segments，可以减少所需的检查点次数。

14.4.7. 禁用 WAL 归档和流复制

当在使用 WAL 归档或流复制的安装中载入大量数据时，完成装载后重新做一次基础备份，可能比
处理大量增量 WAL 数据更快。为了避免在装载期间记录这些增量 WAL，可以通过将wal_level设
为minimal、将archive_mode设为off，并将max_wal_senders设为零，
来禁用归档和流复制。但请注意，修改这些设置需要重启服务器。

这样做除了省去归档进程或 WAL 发送进程处理 WAL 数据的时间外，还能使某些命令执行得更
快，因为这些命令被设计为在wal_level为minimal时完全不写 WAL。（它们可以在结束时执
行一次fsync，以比写 WAL 更低的代价保证崩溃安全性。）这适用于以下命令：

- CREATE TABLE AS SELECT
- CREATE INDEX（以及 ALTER TABLE ADD PRIMARY KEY 等变体）
- ALTER TABLE SET TABLESPACE

- CLUSTER
- COPY FROM，前提是目标表在同一事务中已被创建或截断

14.4.8. 事后运行ANALYZE

每当显著改变了表中数据的分布，都强烈建议运行ANALYZE。
这也包括向表中批量装载大量数据。运行ANALYZE（或VACUUM ANALYZE）可以确保规划器掌握该表的最新统计信息。如果没有统计信息，或者统计信息已经过时，
规划器在生成查询计划时就可能做出糟糕决定，从而导致相关表性能不佳。注意，
如果启用了自动清理守护进程，它可能会自动运行ANALYZE；详见第 23.1.3 节和第 23.1.5 节。

14.4.9. 关于pg_dump的一些注记

由pg_dump生成的转储脚本会自动应用上面若干条指导原则，但并非全部。
若要尽可能快速地还原pg_dump的转储，仍需手动做一些额外操作。（注意，这些要点适用于还原转储，而不是创建转储。无论是使用psql加载文本转储，还是使用pg_restore从pg_dump归档文件加载，相关要点都是一样的。）

默认情况下，pg_dump使用COPY；而当它生成完整的模式加数据转储时，
也会小心地先装载数据，再创建索引和外键。因此在这种情况下，
上述若干指导原则已经被自动处理。剩下需要做的是：

- 为maintenance_work_mem和checkpoint_segments设置适当的（即比正常值大的）值。
- 如果使用 WAL 归档或流复制，可以考虑在恢复期间禁用它们。为此，请在载入转储之前将archive_mode设为off、将wal_level设为minimal，并将max_wal_senders设为零。恢复完成后，再把这些设置改回正确的值，并重新做一次基础备份。
- 考虑是否应当把整个转储作为单个事务来恢复。要这样做，请把-1或--single-transaction命令行选项传给psql或pg_restore。使用这种模式时，
  即使是很小的错误也会回滚整个恢复过程，可能丢掉数小时的处理成果。
  视数据之间的关联程度而定，这种做法未必一定比手工清理更可取。如果使用单个事务并关闭 WAL 归档，COPY命令会运行得最快。
- 如果在数据库服务器上有多个 CPU 可用，可以考虑使用pg_restore的--jobs选项。这允许并行数据载入和索引创建。
- 之后运行ANALYZE。

仅包含数据的转储仍然会使用COPY，但它不会删除或重建索引，通常也不会处理外键。[2]因此，
在装载纯数据转储时，如果想采用这些技术，就需要自行负责删除并重建索引与外键。
装载数据期间增大checkpoint_segments仍然有益，但没有必要同时增大maintenance_work_mem；后者更适合留到之后手工重建索引和外键时再调大。
完成后也别忘了执行ANALYZE；详见第 23.1.3 节和第 23.1.5 节。

14.5. 非持久设置

持久性是一种数据库特性，保证即使服务器崩溃或断电，已提交的事务也会被记录下来。
但持久性会带来显著的数据库开销，因此如果你的应用场景不需要这种保证，可以将

[2] 可以通过使用--disable-triggers选项达到禁用外键的效果 — 但要注意，这样做是取消外键验证，而不仅仅是推迟它。因此如果使用该选项，就有可能插入坏数据。

PostgreSQL配置为以快得多的速度运行。以下配置更改可以在这些情况下提高性能。
除下文特别注明的情况外，数据库软件崩溃时仍然保证持久性；使用这些设置时，
只有操作系统突然停止运行才会造成数据丢失或损坏的风险。

- 将数据库集簇的数据目录放在一个内存支持的文件系统上（即RAM磁盘）。
  这消除了所有的数据库磁盘 I/O，但将数据存储限制到可用的内存量（可能有交换区）。
- 关闭fsync；不需要将数据刷入磁盘。
- 关闭full_page_writes；不需要警惕部分页面写入。
- 增加checkpoint_segments和checkpoint_timeout；这会降低检查点频率，但会增加`/pg_wal`的存储需求。
- 关闭synchronous_commit；这样可能无需在每次提交时将WAL写入磁盘。
  如果仅数据库本身崩溃，这种设置会带来事务丢失的风险（但不会造成数据损坏）。

部分 III. 服务器管理

本部分涵盖PostgreSQL管理员关心的主题，包括软件安装、服务器配置、用户和数据库管理以及维护任务。任何运行PostgreSQL服务器的人，即使仅供个人使用，尤其是在生产环境中，都应熟悉这些主题。

本部分的内容大体按照新用户应当阅读的顺序来安排。各章自成体系，也可按需单独阅读。
相关内容按主题单元以叙述方式展开。希望获得某个命令完整描述的读者，建议查阅第 VI 部分。

开头几章写得无需预备知识即可理解，因此需要自行搭建服务器的新用户可以从这里开始探索。
本部分其余内容涉及调优和管理；这些材料假定读者已经熟悉PostgreSQL数据库系统的一般用法。
建议读者再参阅第 I 部分和第 II 部分部分，以获取更多信息。

目录

15. 从源代码安装 349
15.1. 简要说明 349
15.2. 需求 349
15.3. 获取源代码 351
15.4. 安装过程 351
15.5. 安装后设置 361
15.5.1. 共享库 361
15.5.2. 环境变量 362
15.6. 受支持的平台 362
15.7. 平台相关说明 363
15.7.1. AIX 363
15.7.2. Cygwin 366
15.7.3. HP-UX 366
15.7.4. IRIX 367
15.7.5. MinGW/原生 Windows 368
15.7.6. SCO OpenServer 和 SCO UnixWare 368
15.7.7. Solaris 369
16. 在 Windows 上从源代码安装 372
16.1. 使用 Visual C++ 或 Platform SDK 构建 372
16.1.1. 需求 373
16.1.2. 64 位 Windows 的特殊注意事项 374
16.1.3. 构建 374
16.1.4. 清理和安装 375
16.1.5. 运行回归测试 375
16.1.6. 构建文档 375
16.2. 使用Visual C++或Borland C++构建libpq 376
16.2.1. 生成的文件 376
17. 服务器设置和操作 377
17.1. PostgreSQL用户账户 377
17.2. 创建一个数据库集簇 377
17.2.1. 二级文件系统的使用 378
17.2.2. 使用网络文件系统 378
17.3. 启动数据库服务器 378
17.3.1. 服务器启动失败 380
17.3.2. 客户端连接问题 381
17.4. 管理内核资源 381
17.4.1. 共享内存和信号量 381
17.4.2. 资源限制 387
17.4.3. Linux 内存过量分配 388
17.5. 关闭服务器 389
17.6. 升级 PostgreSQL 集簇 389
17.6.1. 通过pg_dump升级数据 390
17.6.2. 非转储升级方法 392
17.7. 防止服务器欺骗 392
17.8. 加密选项 392
17.9. 使用 SSL 的安全 TCP/IP 连接 393
17.9.1. 使用客户端证书 394
17.9.2. SSL 服务器文件用法 394
17.9.3. 创建自签名证书 395
17.10. 通过 SSH 隧道建立安全 TCP/IP 连接 395
18. 服务器配置 397

18.1. 设置参数 397
18.2. 文件位置 398
18.3. 连接和认证 399
18.3.1. 连接设置 399
18.3.2. 安全和认证 401
18.4. 资源消耗 403
18.4.1. 内存 403
18.4.2. 内核资源使用 404
18.4.3. 基于代价的清理延迟 405
18.4.4. 后台写入器 406
18.4.5. 异步行为 407
18.5. 预写式日志 407
18.5.1. 设置 407
18.5.2. 检查点 410
18.5.3. 归档 410
18.6. 复制 411
18.6.1. 主库 411
18.6.2. 备库 412
18.7. 查询规划 413
18.7.1. 规划器方法配置 413
18.7.2. 规划器代价常量 414
18.7.3. 遗传查询优化器 416
18.7.4. 其他规划器选项 416
18.8. 错误报告和日志 418
18.8.1. 日志记录到哪里 418
18.8.2. 什么时候记录日志 420
18.8.3. 记录哪些内容 422
18.8.4. 使用 CSV 格式的日志输出 425
18.9. 运行时统计数据 426
18.9.1. 查询和索引统计信息收集器 426
18.9.2. 统计监控 427
18.10. 自动清理 427
18.11. 客户端连接默认值 429
18.11.1. 语句行为 429
18.11.2. 区域设置和格式化 432
18.11.3. 其他默认值 433
18.12. 锁管理 434
18.13. 版本和平台兼容性 435
18.13.1. 以前的 PostgreSQL 版本 435
18.13.2. 平台和客户端兼容性 436
18.14. 错误处理 437
18.15. 预置选项 437
18.16. 自定义选项 438
18.17. 开发者选项 439
18.18. 短选项 441
19. 客户端认证 443
19.1. `pg_hba.conf` 文件 443
19.2. 用户名映射 449
19.3. 认证方法 450
19.3.1. 信任认证 450
19.3.2. 密码认证 451
19.3.3. GSSAPI 认证 451
19.3.4. SSPI 认证 452

19.3.5. Kerberos 认证 452
19.3.6. Ident 认证 453
19.3.7. Peer 认证 454
19.3.8. LDAP 认证 454
19.3.9. RADIUS 认证 456
19.3.10. 证书认证 456
19.3.11. PAM 认证 456
19.4. 认证问题 457
20. 数据库角色 458
20.1. 数据库角色 458
20.2. 角色属性 459
20.3. 角色成员资格 460
20.4. 删除角色 461
20.5. 函数和触发器安全性 462
21. 管理数据库 463
21.1. 概述 463
21.2. 创建数据库 463
21.3. 模板数据库 464
21.4. 数据库配置 465
21.5. 删除数据库 465
21.6. 表空间 466
22. 本地化 468
22.1. 区域设置支持 468
22.1.1. 概述 468
22.1.2. 行为 469
22.1.3. 问题 470
22.2. 排序规则支持 470
22.2.1. 概念 470
22.2.2. 管理排序规则 472
22.3. 字符集支持 472
22.3.1. 支持的字符集 473
22.3.2. 设置字符集 475
22.3.3. 服务器和客户端之间的自动字符集转换 476
22.3.4. 进一步阅读 478
23. 日常数据库维护任务 479
23.1. 日常清理 479
23.1.1. 清理基础 479
23.1.2. 回收磁盘空间 480
23.1.3. 更新规划器统计信息 481
23.1.4. 防止事务 ID 回卷失败 481
23.1.5. 自动清理守护进程 483
23.2. 日常重建索引 484
23.3. 日志文件维护 485
24. 备份和恢复 486
24.1. SQL转储 486
24.1.1. 恢复转储 487
24.1.2. 使用pg_dumpall 487
24.1.3. 处理大型数据库 488
24.2. 文件系统级备份 489
24.3. 持续归档和时间点恢复（PITR） 489
24.3.1. 设置 WAL 归档 490
24.3.2. 进行基础备份 492
24.3.3. 使用持续归档备份进行恢复 494
24.3.4. 时间线 495

24.3.5. 建议和示例 496
24.3.6. 注意事项 497
25. 高可用、负载均衡和复制 499
25.1. 不同方案的比较 499
25.2. 日志传送备库 502
25.2.1. 规划 502
25.2.2. 备库操作 503
25.2.3. 为备库准备主库 503
25.2.4. 设置备库 503
25.2.5. 流复制 504
25.2.6. 同步复制 505
25.3. 故障切换 507
25.4. 日志传送的另一种方法 508
25.4.1. 实现 508
25.4.2. 基于记录的日志传送 509
25.5. 热备 509
25.5.1. 用户概览 509
25.5.2. 处理查询冲突 511
25.5.3. 管理员概览 512
25.5.4. 热备参数参考 514
25.5.5. 注意事项 515
26. 恢复配置 516
26.1. 归档恢复设置 516
26.2. 恢复目标设置 517
26.3. 备库设置 517
27. 监控数据库活动 519
27.1. 标准 Unix 工具 519
27.2. 统计收集器 520
27.2.1. 统计信息收集配置 520
27.2.2. 查看收集到的统计信息 520
27.3. 查看锁 528
27.4. 动态追踪 528
27.4.1. 为动态追踪编译 528
27.4.2. 内置探针 528
27.4.3. 使用探针 533
27.4.4. 定义新探针 534
28. 监控磁盘使用情况 536
28.1. 确定磁盘使用情况 536
28.2. 磁盘已满故障 537
29. 可靠性与预写式日志 538
29.1. 可靠性 538
29.2. 预写式日志（WAL） 539
29.3. 异步提交 540
29.4. WAL 配置 541
29.5. WAL 内部机制 542
30. 回归测试 544
30.1. 运行测试 544
30.1.1. 针对临时安装运行测试 544
30.1.2. 针对现有安装运行测试 544
30.1.3. 附加测试套件 545
30.1.4. 区域设置和编码 545
30.1.5. 额外测试 546
30.1.6. 测试热备 546
30.2. 测试结果评估 547

30.2.1. 错误消息差异 547
30.2.2. 区域设置差异 547
30.2.3. 日期和时间差异 547
30.2.4. 浮点差异 548
30.2.5. 行顺序差异 548
30.2.6. 栈深度不足 548
30.2.7. “random” 测试 548
30.2.8. 配置参数 548
30.3. 变体比较文件 549
30.4. 测试覆盖率检查 549

第 15 章 从源代码安装

本章介绍如何使用源代码发布包安装 PostgreSQL。（如果你安装的是预打包的发行版，例如 RPM 或 Debian 软件包，请忽略本章，改为阅读打包者提供的说明。）

15.1. 简要说明

```
./configure
gmake
su
gmake install
adduser postgres
mkdir /usr/local/pgsql/data
chown postgres /usr/local/pgsql/data
su - postgres
/usr/local/pgsql/bin/initdb -D /usr/local/pgsql/data
/usr/local/pgsql/bin/postgres -D /usr/local/pgsql/data >logfile 2>&1 &
/usr/local/pgsql/bin/createdb test
/usr/local/pgsql/bin/psql test
```

详细说明见本章的其余部分。

15.2. 需求

一般来说，现代的 Unix 兼容平台应该能够运行 PostgreSQL。发布时经过具体测试的平台列于下方第 15.6 节。发行包的 `doc` 子目录中有若干平台特定的 FAQ 文档，遇到问题时可以查阅。

构建 PostgreSQL 需要下列软件包：

- 需要 GNU make 3.80 或更高版本；其他 make 程序或较旧版本的 GNU make 都无法工作。GNU make 通常以 `gmake` 这个名字安装；本文档将始终使用这个名字来称呼它。（在某些系统上，GNU make 是默认工具，名字为 `make`。）要检查是否为 GNU make，请输入：

  ```
  gmake --version
  ```

- 你需要一个 ISO/ANSI C 编译器（至少符合 C89）。推荐使用较新的 GCC 版本，不过已知 PostgreSQL 也可以使用来自不同厂商的多种编译器构建。

- 需要 tar 来解开源代码发布包，此外还需要 gzip 或 bzip2 之一。

- 默认使用 GNU Readline 库。它可以让 psql（PostgreSQL 命令行 SQL 解释器）记住你输入的每一条命令，并允许你使用方向键回忆和编辑此前的命令。这非常有用，强烈推荐。如果你不想使用它，则必须给 `configure` 指定 `--without-readline` 选项。作为替代方案，你通常也可以使用 BSD 许可的 `libedit` 库，它最初是在 NetBSD 上开发的。`libedit` 库与 GNU Readline 兼容；如果找不到 `libreadline`，或者在 `configure` 中使用了 `--with-libedit-preferred` 选项，就会使用它。如果你使用的是基于软件包的 Linux 发行版，请注意，如果你的发行版把它们拆成单独的软件包，那么需要同时安装 `readline` 和 `readline-devel`。

- 默认使用 zlib 压缩库。如果你不想使用它，则必须给 `configure` 指定 `--without-zlib` 选项。使用这个选项将禁用 pg_dump 和 pg_restore 对压缩归档的支持。

下列软件包是可选的。默认配置不需要它们，但在启用某些构建选项时会用到，如下所述：

- 要构建服务器端编程语言 PL/Perl，你需要一个完整的 Perl 安装，包括 `libperl` 库和头文件。由于 PL/Perl 是一个共享库，在大多数平台上，`libperl` 也必须是共享库。这在较新的 Perl 版本中似乎是默认行为，但在更早版本中并非如此；无论如何，这取决于你所在环境中安装 Perl 的人是如何选择的。如果你打算不仅仅是偶尔使用 PL/Perl，应确保 Perl 安装是在启用了 `usemultiplicity` 选项的情况下构建的（`perl -V` 会显示这一点）。

  如果你没有共享库但需要一个，PostgreSQL 构建期间会出现类似下面的消息来指出这一事实：

  ```
  *** Cannot build PL/Perl because libperl is not a shared library.
  *** You might have to rebuild your Perl installation.  Refer to
  *** the documentation for details.
  ```

  （如果你没有留意屏幕输出，你只会注意到 PL/Perl 库对象 `plperl.so` 或类似文件没有被安装。）如果看到这条消息，你就必须手动重新构建并安装 Perl，才能构建 PL/Perl。在配置 Perl 时，请要求生成共享库。

- 要构建 PL/Python 服务器端编程语言，你需要带有头文件以及 distutils 模块的 Python 安装。最低要求版本是 Python 2.2。支持 3.1 及更高版本的 Python 3；但使用 Python 3 时请参见第 42.1 节。

  由于 PL/Python 是一个共享库，在大多数平台上，`libpython` 也必须是共享库。默认的 Python 安装并非如此。如果在构建并安装 PostgreSQL 之后，你得到了一个名为 `plpython.so` 的文件（扩展名可能不同），那么一切正常。否则，你应该会看到类似下面的一闪而过的通知：

  ```
  *** Cannot build PL/Python because libpython is not a shared
   library.
  *** You might have to rebuild your Python installation.  Refer to
  *** the documentation for details.
  ```

  这意味着你必须重新构建（部分）Python 安装来创建这个共享库。

  如果遇到问题，请在 Python 2.3 或更高版本的 configure 中使用 `--enable-shared` 标志。在某些操作系统上，你不必构建共享库，但你必须让 PostgreSQL 构建系统相信这一点。详情请查阅 `src/pl/plpython` 目录中的 `Makefile`。

- 要构建 PL/Tcl 过程语言，你当然需要安装 Tcl。如果你使用的是 8.4 之前的 Tcl 版本，请确保它构建时没有启用多线程支持。

- 要启用本地语言支持（NLS），也就是以非英语语言显示程序消息的能力，你需要一个 Gettext API 的实现。某些操作系统已内置此功能（例如 Linux、NetBSD、Solaris）；对于其他系统，你可以从http://www.gnu.org/software/gettext/下载附加软件包。如果你使用的是 GNU C 库中的 Gettext 实现，那么还需要 GNU Gettext 软件包来提供某些实用程序。如果使用其他实现，则不需要它。

- 如果你希望支持使用这些服务进行认证或加密，就需要 Kerberos、OpenSSL、OpenLDAP 和/或 PAM。

如果你从 Git 树构建，而不是使用已发布的源码包，或者你要进行服务器开发，还需要以下软件包：

- 从 Git 检出构建，或者修改了扫描器和解析器的实际定义文件时，需要 GNU Flex 和 Bison。如果需要它们，请确保使用 Flex 2.5.31 或更高版本，以及 Bison 1.875 或更高版本。不能使用其他 lex 和 yacc 程序。

- 从 Git 检出构建，或者修改了使用 Perl 脚本的任何构建步骤的输入文件时，需要 Perl 5.8 或更高版本。在 Windows 上构建时，无论如何都需要 Perl。

如果需要获取 GNU 软件包，可以从本地的 GNU 镜像站点获取（站点列表见http://www.gnu.org/order/ftp.html），或者从ftp://ftp.gnu.org/gnu/获取。

还要检查磁盘空间是否充足。编译期间，源码树需要约 100 MB，安装目录需要约 20 MB。空数据库集簇约占 35 MB；数据库所需空间大约是存储相同数据的纯文本文件的五倍。如果要运行回归测试，还会临时额外需要最多 150 MB。使用 `df` 命令检查可用磁盘空间。

15.3. 获取源代码

要获取 PostgreSQL 9.1.24 的源码，可以通过匿名 FTP 从ftp://ftp.postgresql.org/pub/source/v9.1.24/postgresql-9.1.24.tar.gz下载。
其他下载选项可以在我们的网站上找到：http://www.postgresql.org/download/。
取得文件后，将其解包：

```
gunzip postgresql-9.1.24.tar.gz
tar xf postgresql-9.1.24.tar
```

这会在当前目录下创建目录 `postgresql-9.1.24`，其中包含 PostgreSQL 源码。切换到该目录，继续完成其余安装过程。

你也可以直接从版本控制仓库获取源码，参见附录 H。

15.4. 安装过程

1. **配置**

   安装过程的第一步是为你的系统配置源码树，并选择所需选项。这通过运行以下脚本完成：`configure`。对于默认安装，只需输入：

   ```
   ./configure
   ```

   此脚本会运行一系列测试，确定各个依赖系统的变量的值，并检测操作系统的特殊之处，最后在构建树中创建若干文件来记录结果。你也可以运行 `configure` 时使用源码树之外的目录，以将构建目录单独存放。这种过程也称为 VPATH 构建。方法如下：

   ```
   mkdir build_dir
   cd build_dir
   /path/to/source/tree/configure [options go here]
   gmake
   ```

   默认配置将构建服务器和实用程序，以及所有只需要 C 编译器的客户端应用程序和接口。默认情况下，所有文件都会安装到 `/usr/local/pgsql` 之下。

   你可以通过提供以下一个或多个命令行选项来定制构建和安装过程，这些选项传给 `configure`：

   `--prefix=`*`PREFIX`*

   把所有文件安装到目录 *`PREFIX`* 下，而不是 `/usr/local/pgsql` 下。实际文件会安装到各个子目录中；不会有任何文件直接安装到 *`PREFIX`* 目录中。

如果有特殊需求，还可以通过以下选项分别定制各个子目录。不过，
如果保留它们的默认值，安装就可以重定位，也就是说，安装完成后可以移动目录。（man 和 doc 的位置不受此影响。）

对于可重定位的安装，可以考虑使用 configure 的 --disable-rpath 选项。此外，还需要告知操作系统如何查找共享库。

--exec-prefix=*EXEC-PREFIX*

你可以把与体系结构相关的文件安装到与 *PREFIX* 不同的前缀 *EXEC-PREFIX* 下。这对于在多台主机之间共享与体系结构无关的文件很有用。如果省略该选项，那么 *EXEC-PREFIX* 会被设为与 *PREFIX* 相同，与体系结构相关和无关的文件都会安装到同一棵目录树下，这通常正是你想要的。

--bindir=*DIRECTORY*

指定可执行程序的目录。默认是 *EXEC-PREFIX*/bin，通常也就是 /usr/local/pgsql/bin。

--sysconfdir=*DIRECTORY*

设置各种配置文件的目录，默认是 *PREFIX*/etc。

--libdir=*DIRECTORY*

设置安装库和动态可加载模块的位置。默认是 *EXEC-PREFIX*/lib。

--includedir=*DIRECTORY*

设置安装 C 和 C++ 头文件的目录。默认是 *PREFIX*/include。

--datarootdir=*DIRECTORY*

设置多种只读数据文件的根目录。这只会为后续某些选项设置默认值。默认是 *PREFIX*/share。

--datadir=*DIRECTORY*

设置已安装程序使用的只读数据文件目录。默认是 *DATAROOTDIR*。注意，这与数据库文件实际存放的位置无关。

--localedir=*DIRECTORY*

设置安装区域设置数据的目录，尤其是消息翻译目录文件。默认是 *DATAROOTDIR*/locale。

--mandir=*DIRECTORY*

PostgreSQL 附带的手册页将安装到这个目录下各自对应的 man*x* 子目录中。默认是 *DATAROOTDIR*/man。

--docdir=*DIRECTORY*

设置安装文档文件（不包括“man”页）的根目录。这只会为后续选项设置默认值。
该选项的默认值是 *DATAROOTDIR*/doc/postgresql。

`--htmldir=DIRECTORY`

PostgreSQL 的 HTML 格式文档将安装到这个目录。默认是 *DATAROOTDIR*。

> **注意**
>
> 为了能把 PostgreSQL 安装到共享安装位置（例如 `/usr/local/include`），同时又不干扰系统其他部分的名字空间，我们做了特别处理。首先，除非完整展开后的目录名已经包含字符串“`postgres`”或“`pgsql`”，否则会自动把字符串“`/postgresql`”追加到 `datadir`、`sysconfdir` 和 `docdir` 上。例如，如果你选择 `/usr/local` 作为前缀，那么文档会安装到 `/usr/local/doc/postgresql`；但如果前缀是 `/opt/postgres`，那么它会安装到 `/opt/postgres/doc`。客户端接口的公共 C 头文件会安装到 `includedir` 中，并且不会污染名字空间。内部头文件和服务器头文件则会安装到 `includedir` 下的私有目录中。关于如何访问这些头文件，请参见各接口自己的文档。最后，如果有需要，也会在 `libdir` 下创建私有子目录，用于存放动态可加载模块。

`--with-includes=DIRECTORIES`

DIRECTORIES 是一个以冒号分隔的目录列表，这些目录会被加入编译器搜索头文件的路径中。如果你把可选软件包（如 GNU Readline）安装在非标准位置，就必须使用此选项，并且通常还要使用对应的 `--with-libraries` 选项。

例如：`--with-includes=/opt/gnu/include:/usr/sup/include`。

`--with-libraries=DIRECTORIES`

DIRECTORIES 是一个以冒号分隔的目录列表，用于搜索库文件。如果你把软件包安装在非标准位置，很可能需要使用这个选项（以及对应的 `--with-includes`）。

例如：`--with-libraries=/opt/gnu/lib:/usr/sup/lib`。

`--enable-nls[=LANGUAGES]`

启用本地语言支持（NLS），也就是以非英语语言显示程序消息的能力。*LANGUAGES* 是可选的、以空格分隔的语言代码列表，例如 `--enable-nls='de fr'`。（你的列表与实际提供的翻译集合之间的交集会自动计算。）如果你不指定列表，则会安装所有可用的翻译。

要使用此选项，需要一个 Gettext API 实现；参见上文。

`--with-pgport=NUMBER`

将 *NUMBER* 设为服务器和客户端的默认端口号。默认值是 5432。端口号以后始终都可以修改，但如果在这里指定，那么服务器和客户端都会编译进同一个默认值，这可能很方便。通常选择非默认值的唯一合理原因，是你打算在同一台机器上运行多个 PostgreSQL 服务器。

`--with-perl`

构建 PL/Perl 服务器端语言。

`--with-python`

构建 PL/Python 服务器端语言。

`--with-tcl`

构建 PL/Tcl 服务器端语言。

`--with-tclconfig=`*`DIRECTORY`*

Tcl 会安装文件 `tclConfig.sh`，其中包含构建与 Tcl 交互的模块所需的配置信息。通常会在已知位置自动找到此文件，但如果要使用其他版本的 Tcl，可以指定查找它的目录。

`--with-gssapi`

构建时支持 GSSAPI 认证。在许多系统上，GSSAPI 系统（通常是 Kerberos 安装的一部分）并不安装在默认搜索的位置（如 `/usr/include`、`/usr/lib`），因此除了此选项，还必须使用 `--with-includes` 和 `--with-libraries`。`configure` 会检查所需的头文件和库，确保 GSSAPI 安装满足要求，然后才会继续。

`--with-krb5`

构建时支持 Kerberos 5 认证。在许多系统上，Kerberos 系统并不安装在默认搜索的位置（如 `/usr/include`、`/usr/lib`），因此除了此选项，还必须使用 `--with-includes` 和 `--with-libraries`。`configure` 会检查所需的头文件和库，确保你的 Kerberos 安装满足要求，然后才会继续。

`--with-krb-srvnam=`*`NAME`*

Kerberos 服务主体的默认名称（也供 GSSAPI 使用）。默认值为 `postgres`。通常没有理由更改，除非使用 Windows 环境，此时必须将其设为大写的 `POSTGRES`。

`--with-openssl`

构建时支持 SSL（加密）连接。这要求已安装 OpenSSL 软件包。`configure` 会检查所需的头文件和库，确保 OpenSSL 安装满足要求，然后才会继续。

`--with-pam`

构建时支持 PAM（可插拔认证模块）。

`--with-ldap`

构建时支持 LDAP 认证和连接参数查找（更多信息参见第 31.16 节和第 19.3.8 节）。在 Unix 上，这要求已安装 OpenLDAP 软件包。在 Windows 上，使用默认的 WinLDAP 库。`configure` 会检查所需的头文件和库，确保 OpenLDAP 安装满足要求，然后才会继续。

`--without-readline`

禁止使用 Readline 库（也禁止 libedit）。此选项会禁用 psql 中的命令行编辑和历史记录，因此不建议使用。

`--with-libedit-preferred`

优先使用 BSD 许可的 libedit 库，而不是 GPL 许可的 Readline。只有在两个库都已安装时，该选项才有意义；在那种情况下，默认行为是使用 Readline。

`--with-bonjour`

构建时支持 Bonjour。这要求操作系统支持 Bonjour。建议在 Mac OS X 上启用。

`--with-ossp-uuid`

使用 OSSP UUID 库[1]构建组件。具体来说，构建 uuid-ossp模块，它提供生成 UUID 的函数。

`--with-libxml`

构建时支持 libxml（以启用 SQL/XML 支持）。这一特性要求 libxml 2.6.23 或更高版本。

libxml 会安装一个 `xml2-config` 程序，可用于检测所需的编译器和链接器选项。如果找到该程序，PostgreSQL 会自动使用它。要指定位于不常见位置的 libxml 安装，可以把环境变量 `XML2_CONFIG` 设为指向该安装对应的 `xml2-config` 程序，或使用 `--with-includes` 和 `--with-libraries` 选项。

`--with-libxslt`

构建 xml2模块时使用 libxslt。xml2 依赖此库对 XML 执行 XSL 转换。

`--disable-integer-datetimes`

禁用对时间戳和时间间隔的 64 位整数存储支持，改为以浮点数存储日期时间值。浮点日期时间存储是 8.4 之前PostgreSQL发行版的默认方式，但现在已被弃用，因为它无法在完整`timestamp`取值范围内支持微秒精度。不过，基于整数的日期时间存储需要 64 位整数类型。因此，在没有这种类型的平台上，或为了与为早期PostgreSQL版本编写的应用保持兼容，可以使用此选项。更多信息参见第 8.5 节。

`--disable-float4-byval`

禁用 float4 值的“按值”传递，改为“按引用”传递。此选项会降低性能，但为了兼容用 C 编写并使用“版本 0”调用约定的旧用户定义函数，可能需要使用它。更好的长期解决办法是更新这些函数，使其使用“版本 1”调用约定。

`--disable-float8-byval`

禁用 float8 值的“按值”传递，改为“按引用”传递。此选项会降低性能，但为了兼容用 C 编写并使用“版本 0”调用约定的旧用户定义函数，可能需要使用它。更好的长期解决办法是更新这些函数，使其使用“版本 1”调用约定。注意，此选项不仅影响 float8，还影响 int8 和 timestamp 等相关类型。在 32 位平台上，默认使用 `--disable-float8-byval`，且不允许选择 `--enable-float8-byval`。

[1] http://www.ossp.org/pkg/lib/uuid/

`--with-segsize=SEGSIZE`

设置段大小，单位为 GB。大型表会拆分为多个操作系统文件，每个文件的大小等于段大小。这避免了许多平台上的文件大小限制问题。默认段大小为 1 GB，在所有支持的平台上都是安全的。如果操作系统支持“大文件”（如今大多数系统都支持），
可以使用更大的段大小。这有助于减少处理超大表时消耗的文件描述符数量。
但应注意，不要选择超出平台和拟用文件系统支持范围的值。
你可能要使用的其他工具（如 tar）也可能限制可用的文件大小。建议将此值设为 2 的幂，但这并非硬性要求。注意，更改此值需要执行 initdb。

`--with-blocksize=BLOCKSIZE`

设置块大小，单位为 KB。这是表内存储和 I/O 的单位。默认值 8 KB 适用于大多数情况；在特殊情况下，其他值可能有用。该值必须是 1 到 32（KB）之间的 2 的幂。注意，更改此值需要执行 initdb。

`--with-wal-segsize=SEGSIZE`

设置 WAL 段大小，单位为 MB。这是 WAL 日志中每个文件的大小。调整此大小可能有助于控制 WAL 日志传送的粒度。默认大小为 16 MB。该值必须是 1 到 64（MB）之间的 2 的幂。注意，更改此值需要执行 initdb。

`--with-wal-blocksize=BLOCKSIZE`

设置 WAL 块大小，单位为 KB。这是 WAL 日志内存储和 I/O 的单位。默认值 8 KB 适用于大多数情况；在特殊情况下，其他值可能有用。该值必须是 1 到 64（KB）之间的 2 的幂。注意，更改此值需要执行 initdb。

`--disable-spinlocks`

即使 PostgreSQL 不支持该平台的 CPU 自旋锁，也允许构建成功。缺少自旋锁支持会导致性能不佳；因此，仅应在构建中止并告知平台缺少自旋锁支持时使用此选项。
如果在你的平台上构建 PostgreSQL 必须使用此选项，请向 PostgreSQL 开发者报告该问题。

`--disable-thread-safety`

禁用客户端库的线程安全性。这会使 libpq 和 ECPG 程序中的并发线程无法安全地控制各自私有的连接句柄。

`--with-system-tzdata=DIRECTORY`

PostgreSQL 自带了日期和时间操作所需的时区数据库。这个时区数据库实际上与很多操作系统（如 FreeBSD、Linux 和 Solaris）提供的 IANA 时区数据库兼容，因此再次安装它是多余的。使用此选项时，将使用位于 *DIRECTORY* 的、由系统提供的时区数据库，而不是 PostgreSQL 源码发布中自带的那一份。*DIRECTORY* 必须是绝对路径。在某些操作系统上，`/usr/share/zoneinfo` 是一个可能的目录。请注意，安装过程不会检测时区数据是否不匹配或有误。如果你使用该选项，建议运行回归测试，以验证你指定的时区数据能够与 PostgreSQL 正常配合工作。

这个选项主要面向那些非常了解其目标操作系统的二进制包发布者。
使用该选项的主要优点是，当本地众多夏令时规则中的任何一条发生变化时，Postgre

SQL 软件包都无需升级。另一个优点是，如果安装时不需要构建时区数据库文件，那么 PostgreSQL 的交叉编译也会更直接。

`--without-zlib`

禁止使用 Zlib 库。这会禁用 pg_dump 和 pg_restore 对压缩归档的支持。此选项仅用于没有该库的罕见系统。

`--enable-debug`

将所有程序和库编译为带调试符号的版本。这意味着你可以在调试器中运行程序，以分析问题。这会显著增大安装后的可执行文件大小，而且在非 GCC 编译器上，通常还会禁用编译器优化，从而导致变慢。不过，保留这些符号对于处理可能出现的各种问题极其有帮助。目前，只有在你使用 GCC 的情况下，才建议在生产安装中使用该选项。但如果你在做开发工作或运行测试版，就应始终启用它。

`--enable-coverage`

如果使用 GCC，所有程序和库都会在编译时插入代码覆盖率测试所需的检测代码。运行时，它们会在构建目录中生成包含代码覆盖率指标的文件。更多信息参见第 30.4 节。此选项仅用于使用 GCC 进行开发工作时。

`--enable-profiling`

如果使用 GCC，所有程序和库都会编译为可进行性能分析的形式。后端退出时，会创建一个子目录，其中包含用于性能分析的 `gmon.out` 文件。此选项仅用于使用 GCC 进行开发工作时。

`--enable-cassert`

在服务器中启用断言检查，用于测试许多“不可能发生”的条件。这对代码开发非常有价值，但这些测试可能会显著拖慢服务器速度。此外，启用这些测试并不一定会增强服务器稳定性！断言检查并未按严重程度分类，因此即使某个 bug 相对无害，只要触发了断言失败，仍会导致服务器重启。该选项不建议用于生产环境，但如果你在做开发工作或运行测试版，就应当启用它。

`--enable-depend`

启用自动依赖跟踪。启用后，makefile 会在任何头文件被修改时，重新构建所有受影响的目标文件。如果你在做开发工作，这很有用；但如果你只是打算编译一次并安装，这只是额外的开销。目前该选项只在 GCC 下有效。

`--enable-dtrace`

编译 PostgreSQL 时支持动态跟踪工具 DTrace。更多信息参见第 27.4 节。

可以设置环境变量 `DTRACE` 来指定 `dtrace` 程序。这通常是必要的，因为 `dtrace` 一般安装在 `/usr/sbin` 下，而该目录可能不在搜索路径中。

对于 `dtrace` 程序，可在以下环境变量中指定额外命令行选项：`DTRACEFLAGS`。在 Solaris 上，要在 64 位二进制文件中包含 DTrace 支持，必须指定 `DTRACEFLAGS="-64"` 给 configure。例如，使用 GCC 编译器时：

```
./configure CC='gcc -m64' --enable-dtrace DTRACEFLAGS='-64' ...
```

使用 Sun 的编译器时：

```
./configure CC='/opt/SUNWspro/bin/cc -xtarget=native64' --
enable-dtrace DTRACEFLAGS='-64' ...
```

如果你希望使用不同于 `configure` 所选的 C 编译器，可以将环境变量 `CC` 设为所选程序。默认情况下，`configure` 会在 `gcc` 可用时选择它，否则选择平台默认编译器（通常为 `cc`）。类似地，如有需要，可以使用 `CFLAGS` 变量覆盖默认编译器标志。

可以在 `configure` 命令行上指定环境变量，例如：

`./configure CC=/opt/bin/gcc CFLAGS='-O2 -pipe'`

下面列出可以按这种方式设置的重要变量：

`BISON`

Bison 程序

`CC`

C 编译器

`CFLAGS`

传给 C 编译器的选项

`CPP`

C 预处理器

`CPPFLAGS`

传给 C 预处理器的选项

`DTRACE`

`dtrace` 程序的位置

`DTRACEFLAGS`

传给 `dtrace` 程序的选项

`FLEX`

Flex 程序

`LDFLAGS`

用于链接可执行文件或共享库的选项

`LDFLAGS_EX`

仅用于链接可执行文件的附加选项

`LDFLAGS_SL`

仅用于链接共享库的附加选项

`MSGFMT`

用于本地语言支持的 `msgfmt` 程序

`PERL`

Perl 解释器的完整路径名。它将用于确定构建 PL/Perl 所需的依赖。

`PYTHON`

Python 解释器的完整路径名。它将用于确定构建 PL/Python 所需的依赖。此外，这里指定（或以其他方式隐式选择）Python 2 还是 3，决定了可用的 PL/Python 语言变体。更多信息参见第 42.1 节。

`TCLSH`

Tcl 解释器的完整路径名。它将用于确定构建 PL/Tcl 所需的依赖，并会被代入 Tcl 脚本中。

`XML2_CONFIG`

用于定位 libxml 安装的 `xml2-config` 程序。

2. **构建**

要开始构建，请输入：

`gmake`

（记得使用 GNU make。）构建需要几分钟，具体取决于硬件。显示的最后一行应该是：

```
All of PostgreSQL successfully made. Ready to install.
```

如果要构建所有可构建的内容，包括文档（HTML 和手册页）及附加模块（`contrib`），请改为输入：

`gmake world`

显示的最后一行应该是：

```
PostgreSQL, contrib, and documentation successfully made. Ready to
 install.
```

3. **回归测试**

如果希望在安装前测试刚构建的服务器，可以在此时运行回归测试。
回归测试是一套测试程序，用于验证 PostgreSQL 是否在你的机器上按开发者预期的方式运行。输入：

`gmake check`

（不能以 root 身份运行；请使用非特权用户。）第 30 章包含解读测试结果的详细信息。以后任何时候都可以通过执行同一命令来重复此测试。

4. **安装文件**

> **注意**
>
> 如果正在升级现有系统，请务必阅读第 17.6 节其中包含升级数据库集簇的说明。

要安装 PostgreSQL，请输入：

```
gmake install
```

这会把文件安装到步骤 1中指定的目录。请确保你有权限写入该区域。通常这一步需要以 root 身份执行。另一种做法是预先创建目标目录，并安排授予适当权限。

要安装文档（HTML 和手册页），请输入：

```
gmake install-docs
```

如果你上面执行的是 world 构建，则改为输入：

```
gmake install-world
```

这也会安装文档。

你可以用 `gmake install-strip` 代替 `gmake install`，在安装时剥离可执行文件和库中的符号。这会节省一些空间。如果你在构建时启用了调试支持，
那么剥离操作实际上会移除调试支持，因此只有在不再需要调试时才应该这样做。
`install-strip` 会尽量合理地节省空间，但它并不知道如何从每个可执行文件中去掉所有不需要的字节，所以如果你希望尽可能节省所有磁盘空间，仍需要手工处理。

标准安装会提供客户端应用开发以及服务器端程序开发（例如用 C 编写自定义函数或数据类型）所需的全部头文件。（在 PostgreSQL 8.0 之前，后者需要单独执行 `gmake install-all-headers` 命令，但此步骤现已并入标准安装。）

仅客户端安装：. 如果你只想安装客户端应用程序和接口库，可以使用以下命令：

```
gmake -C src/bin install
gmake -C src/include install
gmake -C src/interfaces install
gmake -C doc install
```

`src/bin` 中有少量仅供服务器使用的二进制文件，不过它们都很小。

在 Windows 上注册 eventlog：. 要在 Windows 操作系统中注册一个 eventlog 库，请在安装完成后执行以下命令：

```
regsvr32 pgsql_library_directory/pgevent.dll
```

这会创建事件查看器所使用的注册表项。

卸载：. 要撤销安装，请使用命令 `gmake uninstall`。不过，这不会删除任何已创建的目录。

清理：. 安装完成后，你可以使用命令 `gmake clean` 从源码树中删除构建产生的文件，以释放磁盘空间。这会保留 `configure` 程序生成的文件，以便

你之后可以用 `gmake` 重新构建全部内容。要把源码树重置到发布时的状态，请使用 `gmake distclean`。如果你打算在同一个源码树中为多个平台构建，就必须为每个平台执行这一步并重新配置。（另一种做法是为每个平台使用单独的构建树，这样源码树就不会被修改。）

如果你已经执行了一次构建，随后发现 `configure` 选项有误，或者修改了任何会被 `configure` 检查的内容（例如软件升级），那么在重新配置和重新构建之前先执行 `gmake distclean` 是一个好主意。否则，你对配置选项所做的更改可能不会传播到所有需要它们的地方。

15.5. 安装后设置

15.5.1. 共享库

在某些使用共享库的系统上，你需要告诉系统如何找到新安装的共享库。
不需要这样做的系统包括 BSD/OS、FreeBSD、HP-UX、IRIX、Linux、NetBSD、OpenBSD、Tru64 UNIX（前身为 Digital UNIX）和 Solaris。

设置共享库搜索路径的方法因平台而异，但最常见的方法是设置环境变量 `LD_LIBRARY_PATH`，方法如下。在 Bourne shell（`sh`，`ksh`，`bash`，`zsh`）：

```
LD_LIBRARY_PATH=/usr/local/pgsql/lib
export LD_LIBRARY_PATH
```

或者在 `csh` 或 `tcsh`：

```
setenv LD_LIBRARY_PATH /usr/local/pgsql/lib
```

请将 `/usr/local/pgsql/lib` 替换为你为 `--libdir` 指定的值，设置过程见步骤 1。应将这些命令放入 shell 启动文件，例如 `/etc/profile` 或 `~/.bash_profile`。关于此方法相关注意事项的一些有用信息，参见http://xahlee.org/UnixResource_dir/_/ldpath.html。

在某些系统上，更好的做法可能是在构建之前设置环境变量 `LD_RUN_PATH`。

在 Cygwin 上，请把库目录加入 `PATH`，或者把 `.dll` 文件移到 `bin` 目录中。

如果不确定，请参考你系统的手册页（可能是 `ld.so` 或 `rld`）。如果你随后收到类似下面这样的消息：

```
psql: error in loading shared libraries
libpq.so.2.1: cannot open shared object file: No such file or
 directory
```

那就说明这一步确实是必需的。到那时再处理即可。

如果你使用的是 BSD/OS、Linux 或 SunOS 4 且拥有 root 权限，可以在安装后运行：

```
/sbin/ldconfig /usr/local/pgsql/lib
```

（或者等效目录），这样可以让运行时链接器更快地找到共享库。更多信息请参见 `ldconfig` 的手册页。在 FreeBSD、NetBSD 和 OpenBSD 上，相应命令是：

```
/sbin/ldconfig -m /usr/local/pgsql/lib
```

其他系统尚不清楚是否有等效命令。

15.5.2. 环境变量

如果你安装到了 /usr/local/pgsql，或者安装到了其他默认不会搜索程序的目录，那么应该把 /usr/local/pgsql/bin（或者你在步骤 1中通过 --bindir 设置的值）加入你的 PATH。严格来说，这不是必需的，但它会让 PostgreSQL 的使用方便得多。

为此，请把下面内容加到你的 shell 启动文件中，例如 ~/.bash_profile（如果你希望影响所有用户，则用 /etc/profile）：

```
PATH=/usr/local/pgsql/bin:$PATH
export PATH
```

如果你使用的是 csh 或 tcsh，则使用这条命令：

```
set path = ( /usr/local/pgsql/bin $path )
```

要让系统能够找到 man 文档，除非你安装到了默认会搜索的位置，否则需要把类似下面的内容加到 shell 启动文件中：

```
MANPATH=/usr/local/pgsql/share/man:$MANPATH
export MANPATH
```

环境变量 PGHOST 和 PGPORT 会告诉客户端应用数据库服务器的主机和端口，从而覆盖编译时内置的默认值。如果你准备从远端运行客户端应用，那么让每个打算使用数据库的用户都设置 PGHOST 会比较方便。不过这不是必需的；对于大多数客户端程序，也可以通过命令行选项传递这些设置。

15.6. 受支持的平台

如果代码中包含了使某个平台（也就是一种 CPU 架构和操作系统的组合）工作的相关处理，并且最近已经在该平台上验证过能够成功构建并通过回归测试，那么 PostgreSQL 开发社区就认为该平台受支持。目前，大多数平台兼容性测试都是由 PostgreSQL 构建农场[2]中的测试机器自动完成的。如果你有兴趣在某个平台上使用 PostgreSQL，而该平台尚未出现在构建农场中，但代码已经能够工作或可以被修改得可工作，我们强烈鼓励你设置一个构建农场成员机器，以便持续确保兼容性。

一般来说，可以预期 PostgreSQL 能在以下 CPU 架构上工作：x86、x86_64、IA64、PowerPC、PowerPC 64、S/390、S/390x、Sparc、Sparc 64、Alpha、ARM、MIPS、MIPSEL、M68K 和 PA-RISC。代码中包含对 M32R、NS32K 和 VAX 的支持，但尚不清楚这些架构近期是否经过测试。对于不受支持的 CPU 类型，通常可以通过配置 --disable-spinlocks 来构建，但性能会很差。

可以预期 PostgreSQL 能在以下操作系统上工作：Linux（所有近期发行版）、Windows（Win2000 SP4 及更高版本）、FreeBSD、OpenBSD、NetBSD、Mac OS X、AIX、HP/UX、IRIX、Solaris、Tru64 Unix 和 UnixWare。其他类 Unix 系统也可能可用，但目前未在测试。大多数情况下，某个操作系统所支持的全部 CPU 架构也都可以工作。尤其是在使用较旧系统时，请查看下面的第 15.7 节，看看是否有针对你的操作系统的特别说明。

如果你在某个平台上遇到安装问题，而根据近期构建农场的结果该平台是受支持的，请把问题报告到 <pgsql-bugs@postgresql.org>。如果你有兴趣把 PostgreSQL 移植到一个新平台，那么 <pgsql-hackers@postgresql.org> 是合适的讨论地点。

[2] http://buildfarm.postgresql.org/

15.7. 平台相关说明

本节记录 PostgreSQL 安装和设置时的一些平台相关附加问题。请务必阅读安装说明，尤其是第 15.2 节。此外，也请参阅第 30 章中关于如何解释回归测试结果的说明。

这里未列出的平台，目前没有已知的平台特定安装问题。

15.7.1. AIX

PostgreSQL 可以在 AIX 上运行，但正确安装可能比较困难。AIX 4.3.3 至 6.1 被视为受支持版本。你可以使用 GCC 或原生 IBM 编译器 `xlc`。一般来说，使用较新的 AIX 和 PostgreSQL 版本会有所帮助。关于哪些 AIX 版本已知可用的最新信息，请查看构建农场。

对于受支持的 AIX 版本，建议至少达到以下修复级别：

AIX 4.3.3

维护级别 11，加上 ML11 后续补丁包

AIX 5.1

维护级别 9，加上 ML9 后续补丁包

AIX 5.2

技术级别 10，服务包 3

AIX 5.3

技术级别 7

AIX 6.1

基础级别

要检查当前修复级别，在 AIX 4.3.3 至 AIX 5.2 ML 7 上使用 `oslevel -r`，在更高版本上使用 `oslevel -s`。

如果将 Readline 或 libz 安装在 `/usr/local`，除了自己的选项外，还应使用以下 `configure` 标志：`--with-includes=/usr/local/include --with-libraries=/usr/local/lib`。

15.7.1.1. GCC 问题

在 AIX 5.3 上，使用 GCC 编译和运行 PostgreSQL 曾出现一些问题。

应使用晚于 3.3.2 的 GCC 版本，尤其是在使用预打包版本时。我们使用 4.0.1 的效果很好。较早版本的问题似乎更多与 IBM 打包 GCC 的方式有关，而非 GCC 本身的问题，因此如果自行编译 GCC，使用较早版本的 GCC 也很可能成功。

15.7.1.2. Unix 域套接字不可用

AIX 5.3 存在一个问题：`sockaddr_storage` 的大小定义不足。在 5.3 版中，IBM 增大了 Unix 域套接字地址结构 `sockaddr_un` 的大小，却未相应增大 `sockaddr_storage`。结果是，尝试

通过 Unix 域套接字使用 PostgreSQL 时，libpq 会写出该数据结构的边界。TCP/IP 连接正常，但 Unix 域套接字无法正常工作，导致回归测试无法运行。

此问题已报告给 IBM，缺陷报告编号为 PMR29657。升级到维护级别 5300-03 或更高版本即可包含此修复。一种快速解决办法是将 `/usr/include/sys/socket.h` 中的 `_SS_MAXSIZE` 改为 1025。无论采用哪种方法，在取得修正后的头文件后，都应重新编译 PostgreSQL。

15.7.1.3. 互联网地址问题

PostgreSQL 依赖系统的 `getaddrinfo` 函数解析 `listen_addresses`、`pg_hba.conf` 等位置中的 IP 地址。较旧的 AIX 版本在此函数中存在多种缺陷。如果遇到与这些设置有关的问题，更新到上面所列的适当 AIX 修复级别应该能够解决。

一位用户报告：

在 AIX 5.3 上部署 PostgreSQL 8.1 时，我们不时遇到统计信息收集器“莫名其妙地”无法成功启动的问题。这似乎是 IPv6 实现中意外行为所致。看起来在 AIX 5.3 上，PostgreSQL 与 IPv6 配合得不太好。

以下任意操作都能“修复”此问题。

- 删除 localhost 的 IPv6 地址：

  ```
  (as root)
  # ifconfig lo0 inet6 ::1/0 delete
  ```

- 从网络服务中移除 IPv6。在 AIX 上，文件 `/etc/netsvc.conf` 大致相当于 Solaris/Linux 上的 `/etc/nsswitch.conf`。AIX 上的默认设置如下：

  ```
  hosts=local,bind
  ```

  将其替换为：

  ```
  hosts=local4,bind4
  ```

  以停用对 IPv6 地址的搜索。

> **警告**
>
> 这实际上只是针对 IPv6 支持不成熟所引发问题的临时解决办法，而 IPv6 支持在 AIX 5.3 的各次发布中已有明显改进。此方法在 AIX 5.3 上有效，但并不是优雅的解决方案。据报告，在 IPv6 支持更成熟的 AIX 6.1 上，这种做法不仅没有必要，还会导致问题。

15.7.1.4. 内存管理

AIX 的内存管理方式有些特殊。服务器可能有数 GB 乃至更多的空闲 RAM，但运行应用程序时仍会出现内存不足或地址空间错误。例如，`createlang`可能因不寻常的错误而失败。以下是以 PostgreSQL 安装所有者身份运行的例子：

```
-bash-3.00$ createlang plperl template1
createlang: language installation failed: ERROR:  could not load
 library "/opt/dbs/pgsql748/lib/plperl.so": A memory address is not in
 the address space for the process.
```

以 PostgreSQL 安装所属组中非所有者的成员身份运行：

```
-bash-3.00$ createlang plperl template1
createlang: language installation failed: ERROR:  could not load
 library "/opt/dbs/pgsql748/lib/plperl.so": Bad address
```

另一个例子是 PostgreSQL 服务器日志中的内存不足错误，此时每次接近或超过 256 MB 的内存分配都会失败。

这些问题的总体原因是服务器进程使用的默认位数和内存模型。默认情况下，在 AIX 上构建的所有二进制文件都是 32 位的。这与硬件类型或所用内核无关。这些 32 位进程的内存上限为 4 GB，按几种模型之一以 256 MB 的段布局。默认情况下，堆与栈共享一个段，因此堆可用空间不足 256 MB。

对于上面的 `createlang` 示例，请检查你的 umask 和 PostgreSQL 安装中二进制文件的权限。该示例涉及的二进制文件是 32 位的，安装时使用模式 750 而非 755。由于权限以这种方式设置，只有所有者或所属组的成员可以加载该库。因为它不是所有用户可读，加载器会将该对象放入进程的堆中，而不是通常应放置的共享库段。

这个问题的“理想”解决办法是使用 64 位构建的 PostgreSQL，但这并非总是可行，因为采用 32 位处理器的系统可以构建 64 位二进制文件，却无法运行它们。

如果需要 32 位二进制文件，请在启动 PostgreSQL 服务器之前，将 `LDR_CNTRL` 设为 `MAXDATA=0x`*n*`0000000`，其中 1 <= n <= 8，并尝试不同的值及 `postgresql.conf` 设置，找到能令人满意地工作的配置。这样使用 `LDR_CNTRL` 会告知 AIX，为服务器的堆预留 `MAXDATA` 字节，按 256 MB 的段分配。找到可用配置后，可以使用 `ldedit` 修改二进制文件，使其默认使用所需的堆大小。也可以重新构建 PostgreSQL，传入 `configure LDFLAGS="-Wl,-bmaxdata:0x`*n*`0000000"` 来达到相同效果。

对于 64 位构建，将 `OBJECT_MODE` 设为 64，并向 `configure` 传入 `CC="gcc -maix64"` 和 `LDFLAGS="-Wl,-bbigtoc"`。（`xlc` 的选项可能不同。）如果未导出 `OBJECT_MODE`，构建可能因链接器错误而失败。设置 `OBJECT_MODE` 后，它会告知 AIX 的 `ar`、`as` 和 `ld` 等构建工具默认处理哪种对象。

默认情况下，可能发生分页空间过量分配。虽然我们尚未见过这种情况，
但当内存耗尽且访问了过量分配的空间时，AIX 会终止进程。我们遇到过最接近的情况是，系统认为没有足够内存容纳另一个进程，导致 fork 失败。与 AIX 的许多其他部分一样，如果这成为问题，可以在系统或进程级别配置分页空间分配方式和内存不足时终止进程的行为。

References and Resources

“Large Program Support[1]”. AIX Documentation: General Programming Concepts: Writing and Debugging Programs.

“Program Address Space Overview[2]”. AIX Documentation: General Programming Concepts: Writing and Debugging Programs.

“Performance Overview of the Virtual Memory Manager (VMM)[3]”. AIX Documentation: Performance Management Guide.

“Page Space Allocation[4]”. AIX Documentation: Performance Management Guide.

[1] http://publib.boulder.ibm.com/infocenter/pseries/topic/com.ibm.aix.doc/aixprggd/genprogc/lrg_prg_support.htm
[2] http://publib.boulder.ibm.com/infocenter/pseries/topic/com.ibm.aix.doc/aixprggd/genprogc/address_space.htm
[3] http://publib.boulder.ibm.com/infocenter/pseries/v5r3/topic/com.ibm.aix.doc/aixbman/prftungd/resmgmt2.htm
[4] http://publib.boulder.ibm.com/infocenter/pseries/v5r3/topic/com.ibm.aix.doc/aixbman/prftungd/memperf7.htm

"Paging-space thresholds tuning[5]" . AIX Documentation: Performance Management Guide.

Developing and Porting C and C++ Applications on AIX[6]. IBM Redbook.

15.7.2. Cygwin

可以使用 Cygwin 这个 Windows 上的类 Linux 环境来构建 PostgreSQL，但这种方式不如原生 Windows 构建（见第 16 章），且如今已不再推荐在 Cygwin 下运行服务器。

从源码构建时，请按照常规安装过程操作（即 `./configure; make` 等），同时注意以下 Cygwin 特有的差异：

- GNU make 命令名为 `make`，而不是 `gmake`。
- 请把路径设置成优先使用 Cygwin 的 bin 目录，而不是 Windows 工具目录。这有助于避免编译问题。
- 不支持 `adduser` 命令；请使用 Windows NT、2000 或 XP 中相应的用户管理应用。如果不是这些系统，则跳过此步骤。
- 不支持 `su` 命令；请在 Windows NT、2000 或 XP 上使用 ssh 模拟 su。如果不是这些系统，则跳过此步骤。
- 不支持 OpenSSL。
- 请启动 `cygserver` 以支持共享内存。为此，请输入命令 `/usr/sbin/cygserver &`。每次启动 PostgreSQL 服务器或初始化数据库集簇（`initdb`）时，该程序都必须在运行中。默认的 `cygserver` 配置可能需要修改（例如增大 `SEMMNS`），以防止 PostgreSQL 因系统资源不足而失败。
- 在某些使用非 C 区域设置的系统上，构建可能会失败。要修复这一点，请在构建前执行 `export LANG=C.utf8` 把区域设置改为 C，安装完 PostgreSQL 后再把它恢复为之前的设置。
- 并行回归测试（`make check`）可能因 `listen()` 的待处理连接队列溢出而误报回归测试失败；队列溢出会导致连接被拒绝的错误或挂起。可以使用 make 变量 `MAX_CONNECTIONS` 限制连接数，方法如下：

  ```
  make MAX_CONNECTIONS=5 check
  ```

  （在某些系统上，并发连接数最高可达约 10 个。）

可以把 `cygserver` 和 PostgreSQL 服务器安装为 Windows NT 服务。关于具体做法，请参阅 Cygwin 上 PostgreSQL 二进制包附带的 `README` 文档。它安装在 `/usr/share/doc/Cygwin` 目录中。

15.7.3. HP-UX

如果系统补丁级别和构建工具合适，PostgreSQL 7.3+ 应能在运行 HP-UX 10.X 或 11.X 的 Series 700/800 PA-RISC 机器上工作。至少有一位开发者经常在 HP-UX 10.20 上测试，我们也收到过在 HP-UX 11.00 和 11.11 上成功安装的报告。

除了 PostgreSQL 源码发行包，还需要 GNU make（HP 的 make 不可用），以及 GCC 或 HP 的完整 ANSI C 编译器。如果打算从 Git 源码而非发行 tar 包构建，还需要 Flex（GNU lex）和

[5] http://publib.boulder.ibm.com/infocenter/pseries/v5r3/topic/com.ibm.aix.doc/aixbman/prftungd/memperf6.htm
[6] http://www.redbooks.ibm.com/abstracts/sg245674.html?Open

Bison（GNU yacc）。我们也建议确保 HP 补丁相对较新。至少，在 HP-UX 11.11 上构建 64 位二进制文件时，可能需要 PHSS_30966（11.11）或其后续补丁，否则 `initdb` 可能挂起：

PHSS_30966 s700_800 ld(1) and linker tools cumulative patch

一般而言，应安装最新的 libc 和 ld/dld 补丁；如果使用 HP 的 C 编译器，也应安装最新的编译器补丁。请访问 HP 的支持站点，例如http://itrc.hp.com和ftp://us-ffs.external.hp.com/，免费获取最新补丁。

如果在 PA-RISC 2.0 机器上构建，并希望使用 GCC 生成 64 位二进制文件，则必须使用 64 位版本的 GCC。HP-UX PA-RISC 和 Itanium 的 GCC 二进制文件可从http://www.hp.com/go/gcc获取。别忘了同时获取并安装 binutils。

如果在 PA-RISC 2.0 机器上构建，并希望编译后的二进制文件能在 PA-RISC 1.1 机器上运行，需要在 `CFLAGS` 中指定 `+DAportable`。

如果在 HP-UX Itanium 机器上构建，需要最新的 HP ANSI C 编译器及其依赖补丁或后续补丁：

PHSS_30848 s700_800 HP C Compiler (A.05.57)
PHSS_30849 s700_800 u2comp/be/plugin library Patch

如果同时安装了 HP 的 C 编译器和 GCC，可以在运行 `configure` 时显式选择要使用的编译器：

```
./configure CC=cc
```

这会选择 HP 的 C 编译器；或者执行：

```
./configure CC=gcc
```

这会选择 GCC。如果省略此设置，configure 在可以选择时会选用 `gcc` 作为编译器。

默认安装目标位置为 `/usr/local/pgsql`，你可能希望将其改为 `/opt` 下的某个位置。如果是这样，请为 `configure` 使用 `--prefix` 选项。

在回归测试中，几何测试可能存在一些低位数字差异，具体取决于所用编译器和数学库的版本。任何其他错误都值得怀疑。

15.7.4. IRIX

据报告，PostgreSQL 可以在 MIPS r8000、r10000（ip25 和 ip27）以及 r12000（ip35）处理器上成功运行，操作系统为 IRIX 6.5.5m、6.5.12、6.5.13 和 6.5.26，使用的 MIPSPro 编译器版本为 7.30、7.3.1.2m、7.3 和 7.4.4m。

你需要 MIPSPro 的完整 ANSI C 编译器。使用 GCC 构建会存在问题。这是一个已知的 GCC 缺陷（在 3.0 版本中仍未修复），与使用返回某些结构体的函数有关。该缺陷影响的函数包括 `inet_ntoa`、`inet_lnaof`、`inet_netof`、`inet_makeaddr` 和 `semctl`。据说可以通过强制将这些函数与 libgcc 链接来修复，但这尚未经过测试。

已知 MIPSPro 编译器的 7.4.1m 版本会生成不正确的代码。症状是尝试启动数据库时出现“invalid primary checkpoint record”。7.4.4m 版本没有问题；中间版本的状态不确定。

可能会出现如下编译问题：

```
cc-1020 cc: ERROR File = pqcomm.c, Line = 427
  The identifier "TCP_NODELAY" is undefined.

                 if (setsockopt(port->sock, IPPROTO_TCP, TCP_NODELAY,
```

某些版本将 TCP 定义放在 `sys/xti.h` 中，因此需要在 `src/backend/libpq/pqcomm.c` 和 `src/interfaces/libpq/fe-connect.c` 中加入 `#include <sys/xti.h>`。如果遇到这种情况，请通知我们，以便我们开发适当的修复。

在回归测试中，几何测试的某些低位数字可能会因所使用的 FPU 不同而存在差异。任何其他错误都是值得怀疑的。

15.7.5. MinGW/原生 Windows

Windows 版 PostgreSQL 可以使用 MinGW（用于 Microsoft 操作系统的类 Unix 构建环境）构建，也可以使用 Microsoft 的 Visual C++ 编译器套件构建。MinGW 构建方式使用本章介绍的常规构建系统；Visual C++ 构建方式完全不同，详见第 16 章。后者是完全原生的构建，不使用 MinGW 等附加软件。PostgreSQL 主网站上提供现成的安装程序。

原生 Windows 移植版本要求 32 位或 64 位的 Windows 2000 或更高版本。更早的操作系统没有足够的基础设施（但可以在其上使用 Cygwin）。可以从http://www.mingw.org/下载 MinGW 这一类 Unix 构建工具，以及 MSYS 这一运行 `configure` 等 shell 脚本所需的 Unix 工具集。运行生成的二进制文件不需要它们；只有创建二进制文件时才需要。

要使用 MinGW 构建 64 位二进制文件，请从http://mingw-w64.sourceforge.net/安装 64 位工具集，将其 bin 目录放入 `PATH`，并使用 `--host=x86_64-w64-mingw32` 选项运行 `configure`。

安装好所有组件后，建议在 `CMD.EXE` 下运行 psql，因为 MSYS 控制台存在缓冲问题。

15.7.5.1. 在 Windows 上收集崩溃转储

如果 PostgreSQL 在 Windows 上崩溃，它能够生成 minidumps，可用于追踪崩溃原因，类似于 Unix 上的核心转储。这些转储可以使用 Windows Debugger Tools 或 Visual Studio 读取。要在 Windows 上启用转储生成，请在集簇数据目录中创建一个名为 `crashdumps` 的子目录。随后，转储会以唯一名称写入该目录，该名称基于崩溃进程的标识符以及崩溃发生时的当前时间。

15.7.6. SCO OpenServer 和 SCO UnixWare

PostgreSQL 可以在 SCO UnixWare 7 和 SCO OpenServer 5 上构建。在 OpenServer 上，你可以使用 OpenServer Development Kit，也可以使用 Universal Development Kit。不过，可能需要按下面的说明做一些调整。

15.7.6.1. Skunkware

你应当找到你的 SCO Skunkware CD 副本。Skunkware CD 随 UnixWare 7 和当前版本的 OpenServer 5 一起提供。Skunkware 包含 Internet 上许多流行程序的即装即用版本。例如，gzip、gunzip、GNU Make、Flex 和 Bison 都包含在内。对于 UnixWare 7.1，这张 CD 现在标为 "Open License Software Supplement"。如果你没有这张 CD，其上的软件可以从http://www.sco.com/skunkware/获取。

Skunkware 对 UnixWare 和 OpenServer 有不同的版本。请确保为你的操作系统安装正确的版本，下面注明的情况除外。

在 UnixWare 7.1.3 及更高版本中，UDK CD 上附带有 GCC 编译器，GNU Make 也是如此。

15.7.6.2. GNU Make

你需要使用 GNU Make 程序，它在 Skunkware CD 上。默认情况下，它会安装为 `/usr/local/bin/make`。为避免与 SCO 的 `make` 程序混淆，你可以将 GNU `make` 重命名为 `gmake`。

对于 UnixWare 7.1.3 及更高版本，GNU Make 程序位于 UDK CD 的 OSTK 部分，路径为 `/usr/gnu/bin/gmake`。

15.7.6.3. Readline

Readline 库在 Skunkware CD 上。但 UnixWare 7.1 的 Skunkware CD 并不包含它。如果你有 UnixWare 7.0.0 或 7.0.1 的 Skunkware CD，可以从那里安装。否则，请尝试http://www.sco.com/skunkware/。

默认情况下，Readline 会安装到 `/usr/local/lib` 和 `/usr/local/include`。但是，PostgreSQL 的 `configure` 程序在没有帮助的情况下无法在那里找到它。如果你安装了 Readline，请对 `configure` 使用以下选项：

```
./configure --with-libraries=/usr/local/lib --with-includes=/usr/
local/include
```

15.7.6.4. 在OpenServer上使用UDK

如果你在 OpenServer 上使用新的 Universal Development Kit（UDK）编译器，需要指定 UDK 库的位置：

```
./configure --with-libraries=/udk/usr/lib --with-includes=/udk/usr/
include
```

把这些与上面的 Readline 选项结合起来：

```
./configure --with-libraries="/udk/usr/lib /usr/local/lib" --with-
includes="/udk/usr/include /usr/local/include"
```

15.7.6.5. 阅读PostgreSQL手册页

默认情况下，PostgreSQL 的手册页会安装到 `/usr/local/pgsql/share/man`。默认情况下，UnixWare 不会到该目录查找手册页。要能阅读它们，你需要修改 `/etc/default/man` 中的 `MANPATH` 变量，例如：

```
MANPATH=/usr/lib/scohelp/%L/man:/usr/dt/man:/usr/man:/usr/share/man:
scohelp:/usr/local/man:/usr/local/pgsql/share/man
```

在 OpenServer 上，要让手册页可用还需要额外花些功夫研究，因为其手册系统与其他平台有些不同。目前，PostgreSQL 完全不会安装它们。

15.7.6.6. 7.1.1b特性补充中的 C99 问题

对于早于 OpenUNIX 8.0.0（UnixWare 7.1.2）随附编译器的编译器，包括 7.1.1b 特性补充，你可能需要在 `CFLAGS` 或 `CC` 环境变量中指定 `-Xb`。其表现是在编译 `tuplesort.c` 时出现引用内联函数的错误。显然，7.1.2(8.0.0) 及之后的编译器发生了变化。

15.7.6.7. UnixWare上的线程

对于线程，你必须在所有使用 libpq 的程序上使用 `-Kpthread`。libpq 使用 `pthread_*` 调用，而这些调用只有在使用 `-Kpthread`/`-Kthread` 标志时才可用。

15.7.7. Solaris

PostgreSQL 在 Solaris 上有良好支持。你的操作系统越新，遇到的问题越少；详情见下文。

15.7.7.1. 所需工具

你可以使用 GCC 或 Sun 编译器套件来构建。为了获得更好的代码优化，在 SPARC 架构上强烈推荐使用 Sun 编译器。我们收到过使用 GCC 2.95.1 时出现问题的报告；建议使用 GCC 2.95.3 或更高版本。如果使用 Sun 编译器，请注意不要选择 `/usr/ucb/cc`；应使用 `/opt/SUNWspro/bin/cc`。

你可以从http://developers.sun.com/sunstudio/downloads/下载 Sun Studio。许多 GNU 工具已经集成到 Solaris 10 中，或者包含在 Solaris 配套光盘中。如果你需要适用于较旧 Solaris 版本的软件包，可以到http://www.sunfreeware.com或http://www.blastwave.org查找这些工具。如果你更想要源码，请看http://www.gnu.org/order/ftp.html。

15.7.7.2. OpenSSL 的问题

在以 OpenSSL 支持构建 PostgreSQL 时，你可能会在以下文件中遇到编译错误：

- `src/backend/libpq/crypt.c`
- `src/backend/libpq/password.c`
- `src/interfaces/libpq/fe-auth.c`
- `src/interfaces/libpq/fe-connect.c`

这是因为标准的 `/usr/include/crypt.h` 头文件与 OpenSSL 提供的头文件之间存在命名空间冲突。

将你的 OpenSSL 安装升级到 0.9.6a 版可解决此问题。Solaris 9 及更高版本带有较新版本的 OpenSSL。

15.7.7.3. configure 报告测试程序失败

如果 `configure` 报告某个测试程序失败，这多半是因为运行时链接器找不到某些库，通常是 libz、libreadline，或其他非标准库如 libssl。要把它指向正确位置，请在 `configure` 命令行中设置环境变量 `LDFLAGS`，例如：

```
configure ... LDFLAGS="-R /usr/sfw/lib:/opt/sfw/lib:/usr/local/lib"
```

更多信息请参见 ld 手册页。

15.7.7.4. 64 位构建有时会崩溃

在 Solaris 7 及更早版本上，64 位 libc 的 `vsnprintf` 例程存在缺陷，导致 PostgreSQL 不定期发生核心转储。已知最简单的解决办法是强制 PostgreSQL 使用自带的 `vsnprintf`，而不是库中的版本。为此，在运行 `configure` 后，编辑由 `configure` 生成的文件：在 `src/Makefile.global` 中，将以下行：

```
LIBOBJS =
```

改为：

```
LIBOBJS = snprintf.o
```

（此变量中可能已经列有其他文件，顺序无关紧要。）然后照常构建。

15.7.7.5. 为获得最佳性能而编译

在 SPARC 架构上，强烈推荐使用 Sun Studio 进行编译。可以尝试使用 `-xO5` 优化标志，以生成显著更快的二进制文件。不要使用任何会改变浮点运算行为以及 `errno` 处理行为的标志（例如 `-fast`）。这些标志可能导致 PostgreSQL 出现不符合标准的行为，例如在日期和时间计算中。

如果你没有理由在 SPARC 上使用 64 位二进制文件，那么优先选择 32 位版本。64 位运算更慢，而且 64 位二进制文件也比 32 位变体慢。另一方面，在 AMD64 CPU 家族上，32 位代码并不是原生模式，因此 32 位代码在该 CPU 家族上会明显更慢。

关于 PostgreSQL 与 Solaris 性能调优的一些技巧，可以参见http://www.sun.com/servers/coolthreads/tnb/applications_postgresql.jsp。这篇文章主要针对 T2000 平台，但其中的许多建议对运行 Solaris 的其他硬件同样有用。

15.7.7.6. 使用 DTrace 跟踪 PostgreSQL

是的，可以使用 DTrace。更多信息见第 27.4 节。你还可以在这篇文章中找到更多信息：http://blogs.sun.com/robertlor/entry/user_level_dtrace_probes_in。

如果发现链接 `postgres` 可执行文件时中止，并出现类似以下错误消息：

```
Undefined                       first referenced
 symbol                             in file
AbortTransaction                    utils/probes.o
CommitTransaction                   utils/probes.o
ld: fatal: Symbol referencing errors. No output written to postgres
collect2: ld returned 1 exit status
gmake: *** [postgres] Error 1
```

则说明安装的 DTrace 太旧，无法处理静态函数中的探针。需要 Solaris 10u4 或更高版本。

第 16 章 在 Windows 上从源代码安装

建议大多数用户下载适用于 Windows 的二进制发行版，该发行版以一键安装程序包的形式提供，可从PostgreSQL网站获取。从源码构建仅面向开发PostgreSQL或扩展的人员。

在Windows上构建 PostgreSQL 有几种不同的方法。使用 Microsoft 工具进行构建的最简单方式，是安装受支持版本的Microsoft Platform SDK并使用其自带的编译器。也可以使用完整的Microsoft Visual C++ 2005 或 2008进行构建。在某些情况下，除编译器之外还需要安装Platform SDK。

也可以使用MinGW提供的 GNU 编译器工具构建 PostgreSQL，或者在较旧版本的Windows上使用Cygwin。

最后，客户端访问库（libpq）可以使用 Visual C++ 7.1或Borland C++构建，以便与使用这些工具静态链接构建的应用保持兼容。

使用MinGW或Cygwin构建时，采用的是常规构建系统，见第 15 章以及第 15.7.5 节和第 15.7.2 节中的专门说明。要在这些环境中生成原生 64 位二进制文件，请使用MinGW-w64工具。这些工具也可用于在其他主机（例如Linux和Darwin）上为 32 位和 64 位Windows目标进行交叉编译。不建议使用Cygwin运行生产服务器，它只应用于原生构建无法工作的较旧版本Windows，例如Windows 98。官方二进制文件是使用Visual Studio构建的。

原生构建的psql不支持命令行编辑。Cygwin构建支持命令行编辑，因此如果需要在Windows上交互式使用 psql，应当使用它。

16.1. 使用 Visual C++ 或 Platform SDK 构建

PostgreSQL 可以使用 Microsoft 的 Visual C++ 编译器套件构建。这些编译器既可以来自Visual Studio、Visual Studio Express，也可以来自某些版本的Platform SDK。如果你还没有配置好Visual Studio环境，最简单的办法是使用Platform SDK中的编译器，它可以从 Microsoft 免费下载。

PostgreSQL 支持Visual Studio 2005和 Visual Studio 2008的编译器。当只使用 Platform SDK，或为 64 位 Windows 构建时，只支持Visual Studio 2008。尚不支持Visual Studio 2010。

使用Platform SDK构建时，支持 SDK 的 6.0 到 7.0 版本。更旧或更新的版本都无法工作。特别是7.0a 及之后的版本不能使用，因为它们包含了来自Visual Studio 2010的编译器。

使用Visual C++构建所需的工具位于`src/tools/msvc`目录中。构建时，请确保系统 PATH 中没有来自MinGW或Cygwin的工具。同时，也要确保所需的 Visual C++ 工具都已在 PATH 中可用。在Visual Studio中，启动Visual Studio Command Prompt。在Platform SDK中，启动开始菜单中 SDK 项下列出的CMD shell。如果你想构建 64 位版本，就必须使用 64 位版本的命令提示符，反之亦然。所有命令都应在`src\tools\msvc`目录中运行。

在构建之前，你可能需要编辑`config.pl`文件，以反映你想更改的任何配置选项，或要使用的任何第三方库的路径。完整配置是先读取并解析`config_default.pl`文件，再应用来自`config.pl`的任何更改而得到的。例如，要指定Python安装位置，可在`config.pl`中写入以下内容：

```
$config->{python} = 'c:\python26';
```

你只需要指定那些与`config_default.pl`中不同的参数。

如果你需要设置其他环境变量，请创建一个名为`buildenv.pl`的文件，
并将所需命令放入其中。例如，如果 bison 不在 PATH 中，可以创建一个包含以下内容的文件，以将其路径加入 PATH：

```
$ENV{PATH}=$ENV{PATH} . ';c:\some\where\bison\bin';
```

16.1.1. 需求

构建PostgreSQL需要以下附加产品。请使用`config.pl`文件指定这些库所在的目录。

Microsoft Platform SDK

建议升级到Microsoft Platform SDK最新受支持的版本（目前为版本 7.0），可从http://www.microsoft.com/downloads/下载。

你必须始终包含 SDK 中的Windows Headers and Libraries部分。如果你安装的Platform SDK包含Visual C++ Compilers，则无需Visual Studio即可构建。

ActiveState Perl

运行构建生成脚本需要 ActiveState Perl。MinGW 或 Cygwin Perl 都无法工作。它还必须出现在 PATH 中。二进制文件可从http://www.activestate.com下载（注意：需要 5.8 或更高版本，免费的标准发行版即可）。

以下附加产品并非入门所必需，但若要构建完整的软件包则需要它们。请使用`config.pl`文件指定这些库所在的目录。

ActiveState TCL

构建 PL/TCL 需要它（注意：需要 8.4 版本，免费的标准发行版即可）。

Bison和Flex

从 Git 构建需要 Bison 和 Flex，但从发布文件构建则不需要。注意只有 Bison 1.875 或 2.2 及以上版本可以工作。此外，需要 Flex 2.5.31 或更高版本。Bison 可从http://gnuwin32.sourceforge.net下载。Flex 可从http://www.postgresql.org/ftp/misc/winflex/下载。

> **注意**
>
> 来自 GnuWin32 的 Bison 发行版似乎存在一个 bug：如果安装在名称中含有空格的目录中，Bison 就会工作异常，例如英文安装中的默认位置`C:\Program Files\GnuWin32`。请考虑安装到`C:\GnuWin32`。

Diff

运行回归测试需要 Diff，可从http://gnuwin32.sourceforge.net下载。

Gettext

构建带有 NLS 支持的版本需要 Gettext，可从http://gnuwin32.sourceforge.net下载。请注意，二进制文件、依赖项以及开发文件都是必需的。

MIT Kerberos

支持 Kerberos 认证需要 MIT Kerberos。MIT Kerberos 可从http://web.mit.edu/Kerberos/dist/index.html下载。

libxml2和libxslt

XML 支持需要它们。二进制文件可从http://zlatkovic.com/pub/libxml下载，源代码可从http://xmlsoft.org获取。请注意，libxml2 需要 iconv，而它可从相同的下载位置获得。

openssl

SSL 支持需要它。二进制文件可从http://www.slproweb.com/products/Win32OpenSSL.html下载，源码可从http://www.openssl.org下载。

ossp-uuid

支持 UUID-OSSP 需要它（仅 contrib）。源代码可从http://www.ossp.org/pkg/lib/uuid/下载。

Python

构建PL/Python需要它。二进制文件可从http://www.python.org下载。

zlib

pg_dump和pg_restore中的压缩支持需要它。二进制文件可从http://www.zlib.net下载。

16.1.2. 64 位 Windows 的特殊注意事项

PostgreSQL 在 64 位 Windows 上只会为 x64 架构构建，不支持 Itanium 处理器。

不支持在同一个构建树中混用 32 位和 64 位版本。构建系统会自动检测自己是在 32 位还是 64 位环境中运行，并据此构建 PostgreSQL。因此，在开始构建之前启动正确的命令提示符非常重要。

若要使用服务器端第三方库，例如python或openssl，该库也必须是 64 位的。不支持在 64 位服务器中加载 32 位库。PostgreSQL 支持的若干第三方库可能仅提供 32 位版本，在这种情况下，它们不能与 64 位 PostgreSQL 一起使用。

16.1.3. 构建

要以发布配置（默认值）构建 PostgreSQL 的全部内容，请运行以下命令：

```
build
```

要以调试配置构建 PostgreSQL 的全部内容，请运行以下命令：

```
build DEBUG
```

若只构建单个项目，例如 psql，请运行以下命令：

```
build psql
build DEBUG psql
```

若要将默认构建配置改为调试模式，请在`buildenv.pl`文件中加入以下内容：

```
$ENV{CONFIG}="Debug";
```

也可以在 Visual Studio 图形界面中进行构建。在这种情况下，你需要先从命令提示符运行：

```
perl mkvcbuild.pl
```

然后在 Visual Studio 中打开生成的`pgsql.sln`（位于源码树根目录）。

16.1.4. 清理和安装

大多数时候，Visual Studio 的自动依赖跟踪都会处理好变更过的文件。但如果变更较大，你可能需要清理安装。要执行此操作，只需运行`clean.bat`命令，它会自动清除所有生成的文件。
你也可以带上*`dist`*参数运行它，此时它的行为类似于**`make distclean`**，并且也会删除 flex/bison 的输出文件。

默认情况下，所有文件都会写入`debug`或`release`目录下的某个子目录。
若要按标准布局安装这些文件，并同时生成初始化和使用数据库所需的文件，请运行以下命令：

```
install c:\destination\directory
```

16.1.5. 运行回归测试

要运行回归测试，请先确保已完成所有必需部分的构建。此外，还要确保加载系统各部分所需的 DLL（例如过程语言所用的 Perl 和 Python DLL）都能在系统路径中找到。如果不能，请通过 `buildenv.pl` 文件设置路径。要运行测试，请在 `src\tools\msvc` 目录中执行以下命令之一：

```
vcregress check
vcregress installcheck
vcregress plcheck
vcregress contribcheck
```

要更改所用的测试调度方式（默认为 parallel），请将其追加到命令行，例如：

```
vcregress check serial
```

有关回归测试的更多信息，参见第 30 章。

16.1.6. 构建文档

将 PostgreSQL 文档构建为 HTML 格式，需要若干工具和文件。请为这些文件创建一个根目录，并将它们存放在下列子目录中。

OpenJade 1.3.1-2

从http://sourceforge.net/projects/openjade/files/openjade/1.3.1/openjade-1_3_1-2-bin.zip/download下载，并解压到子目录 `openjade-1.3.1` 中。

DocBook DTD 4.2

从http://www.oasis-open.org/docbook/sgml/4.2/docbook-4.2.zip下载，并解压到子目录 `docbook` 中。

DocBook DSSSL 1.79

从http://sourceforge.net/projects/docbook/files/docbook-dsssl/1.79/docbook-dsssl-1.79.zip/download下载，并解压到子目录 `docbook-dsssl-1.79` 中。

ISO 字符实体

从http://www.oasis-open.org/cover/ISOEnts.zip下载，并解压到子目录 `docbook` 中。

编辑 `buildenv.pl` 文件，并添加一个变量来指定根目录的位置，例如：

```
$ENV{DOCROOT}='c:\docbook';
```

要构建文档，请运行命令 `builddoc.bat`。注意，为了生成索引，实际上会构建两次。生成的 HTML 文件位于 `doc\src\sgml`。

16.2. 使用Visual C++或Borland C++构建libpq

只有当你需要带有不同调试/发布标志的版本，或需要静态库以链接到应用中时，才建议使用Visual C++ 7.1-9.0或Borland C++构建 libpq。一般用途推荐使用MinGW、Visual Studio或Platform SDK方法。

要使用Visual Studio 7.1 或更高版本构建libpq客户端库，请切换到`src`目录并输入命令：

```
nmake /f win32.mak
```

要使用Visual Studio 8.0 或更高版本构建 64 位版本的libpq客户端库，请切换到`src`目录并输入命令：

```
nmake /f win32.mak CPU=AMD64
```

关于所支持变量的更多细节，参见`win32.mak`文件。

要使用Borland C++构建libpq客户端库，请切换到`src`目录并输入命令：

```
make -N -DCFG=Release /f bcc32.mak
```

16.2.1. 生成的文件

将构建以下文件：

`interfaces\libpq\Release\libpq.dll`

动态可链接的前端库

`interfaces\libpq\Release\libpqdll.lib`

用于把你的程序链接到`libpq.dll`的导入库

`interfaces\libpq\Release\libpq.lib`

前端库的静态版本

通常不需要安装任何客户端文件。你应当把`libpq.dll`文件放在与应用程序可执行文件相同的目录中。除非绝对必要，不要把`libpq.dll`安装到你的`Windows`、`System`或`System32`目录。如果使用安装程序安装该文件，则应利用文件中包含的`VERSIONINFO`资源进行带版本检查的安装，以确保不会覆盖较新版本的库。

如果你计划在这台机器上使用libpq进行开发，则需要把源码树的`src\include`和`src\interfaces\libpq`子目录加入编译器设置中的包含路径。

要使用该库，你必须把`libpqdll.lib`文件加入你的项目。（在 Visual C++ 中，只需右键单击项目并选择添加它。）

第 17 章 服务器设置和操作

本章讨论如何设置和运行数据库服务器，以及它与操作系统的交互。

17.1. PostgreSQL用户账户

与任何可从外部访问的服务器守护进程一样，建议使用单独的用户账户运行 PostgreSQL。该用户账户应该只拥有服务器管理的数据，不应与其他守护进程共用。（例如，使用 nobody 用户就不是好主意。）不建议将可执行文件安装为由该用户所有，因为系统一旦被攻破，就可能修改自己的二进制文件。

要在系统中添加一个 Unix 用户账户，请查找useradd或adduser命令。通常使用的用户名是 postgres，本书也都按此假定，不过你也可以使用其他名称。

17.2. 创建一个数据库集簇

在做任何事情之前，必须先在磁盘上初始化一个数据库存储区域。我们把它称为数据库集簇。（SQL 标准使用术语目录集簇（catalog cluster）。）数据库集簇是由一个正在运行的数据库服务器实例管理的一组数据库。初始化之后，数据库集簇中会包含一个名为postgres的数据库，作为供工具、用户和第三方应用使用的默认数据库。数据库服务器本身并不要求该postgres数据库存在，但许多外部工具程序都假定它存在。初始化期间，集簇中还会创建另一个数据库，名为template1。顾名思义，它会作为后续新建数据库的模板；不应将它用于实际工作。（关于在集簇中创建新数据库的信息，见第 21 章。）

从文件系统角度看，数据库集簇就是一个目录，所有数据都存储在其下。我们将它称为数据目录或数据区域。数据存放位置完全由你决定，没有默认值；常用位置包括 /usr/local/pgsql/data或/var/lib/pgsql/data。要初始化数据库集簇，请使用命令initdb，该命令随以下产品安装：PostgreSQL。数据库集簇在文件系统中的目标位置通过 -D 选项指定，例如：

```
$ initdb -D /usr/local/pgsql/data
```

注意，必须先登录 PostgreSQL 用户账户，再执行此命令。该账户已在上一节介绍。

> **提示**
>
> 除了使用 -D 选项，也可以设置环境变量 PGDATA。

或者，也可以通过pg_ctl程序运行initdb，如下所示：

```
$ pg_ctl -D /usr/local/pgsql/data initdb
```

如果你打算用pg_ctl来启动和停止服务器（见第 17.3 节），这种方式可能更直观，因为这样pg_ctl就会成为你管理数据库服务器实例时使用的唯一命令。

如果指定的目录尚不存在，initdb会尝试创建它。当然，如果initdb对其父目录没有写权限，这会失败。通常建议让PostgreSQL用户不仅拥有数据目录，也拥有其父目录，这样就不会有这个问题。如果目标父目录也不存在，则需要先创建它；如果祖父目录不可写，就需要使用 root 权限。因此，过程可能如下所示：

```
root# mkdir /usr/local/pgsql
root# chown postgres /usr/local/pgsql
```

```
root# su postgres
postgres$ initdb -D /usr/local/pgsql/data
```

如果数据目录已存在且其中已有文件，initdb会拒绝运行；这是为了防止意外覆盖现有安装。

由于数据目录包含存储在数据库中的全部数据，必须保护它免遭未经授权的访问。因此，initdb 会撤销除 PostgreSQL 用户以外所有人的访问权限。

不过，尽管目录内容是安全的，默认的客户端认证设置却允许任何本地用户连接到数据库，甚至成为数据库超级用户。如果你不信任其他本地用户，我们建议在运行 initdb 时使用 -W、--pwprompt 或 --pwfile 之一，为数据库超级用户指定密码。另外，指定 -A md5 或 -A password，以避免使用默认的 trust 认证方式；或者在运行 initdb 之后、首次启动服务器之前修改生成的 pg_hba.conf 文件。（其他合理方法包括使用 peer 认证，或使用文件系统权限限制连接。更多信息见第 19 章。）

initdb 还会初始化数据库集簇的默认区域设置。通常，它会直接采用环境中的区域设置，并将其应用于初始化的数据库。可以为数据库指定不同的区域设置；更多信息见第 22.1 节。特定数据库集簇中使用的默认排序顺序由 initdb 设置，虽然你可以使用不同的排序顺序创建新数据库，但 initdb 创建的模板数据库所使用的顺序，只有删除并重建这些数据库才能更改。使用 C 或 POSIX 以外的区域设置还会影响性能。因此，一开始就作出正确选择很重要。

initdb还会为数据库集簇设置默认字符集编码。通常应选择与区域设置相匹配的字符集编码。详见第 22.3 节。

非C和非POSIX的区域设置依赖操作系统的排序规则库来决定字符集排序。这会控制索引中存储键值的顺序。因此，无论是通过快照恢复、二进制流复制、切换到不同操作系统，还是升级操作系统，都不能让一个集簇切换到不兼容版本的排序规则库。

17.2.1. 二级文件系统的使用

很多安装会把数据库集簇建在机器“根”卷以外的文件系统（卷）上。如果这样做，不建议把二级卷的最顶层目录（挂载点）直接用作数据目录。最佳实践是在挂载点目录下创建一个由PostgreSQL用户拥有的目录，再在其下创建数据目录。这样可以避免权限问题，尤其是在执行pg_upgrade之类的操作时；同时也能确保二级卷脱机时以干净的方式失败。

17.2.2. 使用网络文件系统

许多安装会在网络文件系统上创建数据库集簇。有时通过 NFS 完成，有时则使用内部采用 NFS 的网络附加存储（NAS）设备。PostgreSQL 不会对 NFS 文件系统做特殊处理，也就是说，它假定 NFS 的行为与本地连接的驱动器完全一致。如果客户端或服务器端的 NFS 实现不提供标准文件系统语义，就可能导致可靠性问题（参见http://www.time-travellers.org/shane/papers/NFS_considered_harmful.html）。具体而言，对 NFS 服务器的延迟（异步）写入可能导致数据损坏。如果可能，应以同步方式（不使用缓存）挂载 NFS 文件系统，以避免这一风险。此外，不建议对 NFS 文件系统使用软挂载。

存储区域网络（SAN）通常使用 NFS 以外的通信协议，可能存在也可能不存在此类风险。建议查阅供应商文档中有关数据一致性保证的说明。PostgreSQL 的可靠性不可能高于它所使用的文件系统。

17.3. 启动数据库服务器

在任何人访问数据库之前，必须先启动数据库服务器。数据库服务器程序名为 postgres。其中，postgres 程序必须知道应到哪里查找它要使用的数据。这通过 -D 选项指定。因此，启动服务器最简单的方法是：

```
$ postgres -D /usr/local/pgsql/data
```

这会让服务器在前台运行。必须先登录 PostgreSQL 用户账户再执行此操作。如果不使用 `-D`，服务器会尝试使用环境变量所指定的数据目录，该变量为 `PGDATA`。如果也未提供该变量，则会失败。

通常最好在后台启动`postgres`。做法是使用常见的 Unix shell 语法：

```
$ postgres -D /usr/local/pgsql/data >logfile 2>&1 &
```

如上所示，把服务器的stdout和stderr输出保存到某个地方非常重要。这有助于审计，也有助于诊断问题。（关于日志文件处理的更深入讨论，见第 23.3 节。）

`postgres`还接受许多其他命令行选项。更多信息请见postgres参考页以及下面的第 18 章。

这些 shell 语法很快就会让人觉得繁琐。因此提供了包装器程序pg_ctl来简化一些任务。例如：

```
pg_ctl start -l logfile
```

这会在后台启动服务器，并把输出写入指定的日志文件。这里的 `-D` 选项含义与在 `postgres` 中相同。`pg_ctl` 还可用于停止服务器。

通常，你会希望在计算机启动时就启动数据库服务器。自动启动脚本依赖于操作系统。PostgreSQL 在 `contrib/start-scripts` 目录中附带了一些示例脚本。安装这些脚本需要 root 权限。

不同系统在引导时启动守护进程的惯例各不相同。许多系统有`/etc/rc.local`或`/etc/rc.d/rc.local`文件，其他系统则使用`init.d`或`rc.d`目录。无论采用哪种方式，服务器都必须由PostgreSQL用户账户而不是 root或其他用户来启动。因此，你大概应该在命令中使用`su postgres -c '...'`这种形式。例如：

```
su postgres -c 'pg_ctl start -D /usr/local/pgsql/data -l serverlog'
```

下面是一些与操作系统相关的补充建议。（在每种情况下，请确保使用正确的安装目录和用户名；这里展示的是通用值。）

- 对于FreeBSD，请查看PostgreSQL源码发布包中的`contrib/start-scripts/freebsd`文件。
- 在OpenBSD上，把以下内容加入`/etc/rc.local`：

  ```
  if [ -x /usr/local/pgsql/bin/pg_ctl -a -x /usr/local/pgsql/bin/
  postgres ]; then
      su -l postgres -c '/usr/local/pgsql/bin/pg_ctl start -s -l /var/
  postgresql/log -D /usr/local/pgsql/data'
      echo -n ' postgresql'
  fi
  ```

- 在Linux系统上，可以把

  ```
  /usr/local/pgsql/bin/pg_ctl start -l logfile -D /usr/local/pgsql/
  data
  ```

  加入`/etc/rc.d/rc.local`或`/etc/rc.local`，或者查看PostgreSQL源码发布包中的`contrib/start-scripts/linux`文件。

- 在NetBSD上，可根据偏好使用FreeBSD或Linux的启动脚本。

- 在Solaris上，创建一个名为/etc/init.d/postgresql的文件，其中包含以下内容：

  ```
  su - postgres -c "/usr/local/pgsql/bin/pg_ctl start -l logfile -D /
  usr/local/pgsql/data"
  ```

  然后，在/etc/rc3.d中为它创建一个名为S99postgresql的符号链接。

当服务器运行时，其PID会保存在数据目录中的postmaster.pid文件里。
这可以防止多个服务器实例在同一个数据目录上运行，也可以用来关闭服务器。

17.3.1. 服务器启动失败

服务器启动失败有几种常见原因。请检查服务器日志文件，
或者手工启动服务器（不要重定向标准输出和标准错误），看看出现了什么错误消息。
下面将更详细地解释一些最常见的错误消息。

```
LOG:  could not bind IPv4 socket: Address already in use
HINT:  Is another postmaster already running on port 5432? If not,
 wait a few seconds and retry.
FATAL:  could not create TCP/IP listen socket
```

这通常就是字面上的意思：你试图在一个已有服务器运行的端口上再启动另一个服务器。不过，如果内核错误消息不是Address already in use或类似变体，也可能是别的问题。例如，试图在一个保留端口上启动服务器，可能会得到类似下面的消息：

```
$ postgres -p 666
LOG:  could not bind IPv4 socket: Permission denied
HINT:  Is another postmaster already running on port 666? If not, wait
 a few seconds and retry.
FATAL:  could not create TCP/IP listen socket
```

像下面这样的消息：

```
FATAL:  could not create shared memory segment: Invalid argument
DETAIL:  Failed system call was shmget(key=5440001, size=4011376640,
 03600).
```

可能表示内核对共享内存大小的限制小于 PostgreSQL 尝试创建的工作区（本例中为4011376640 字节）。也可能表示内核中根本没有配置 System V 风格的共享内存支持。作为临时解决办法，可以尝试以少于通常数量的缓冲区启动服务器（shared_buffers）。
最终还是需要重新配置内核，增大允许的共享内存大小。
如果尝试在同一台机器上启动多个服务器，而它们请求的总空间超出内核限制，
也可能看到这条消息。

像下面这样的错误：

```
FATAL:  could not create semaphores: No space left on device
DETAIL:  Failed system call was semget(5440126, 17, 03600).
```

并不意味着你已经用光了磁盘空间。它表示你的内核对System V信号量数量的限制，小于PostgreSQL想要创建的数量。和上面一样，你也许可以通过减少允许的连接数（max_connections）来暂时绕过这个问题，但最终还是应该提高内核限制。

如果收到“illegal system call”错误，很可能是内核根本不支持共享内存或信号量。在这种情况下，唯一的办法就是重新配置内核以启用这些功能。

关于配置System V IPC功能的细节请见第 17.4.1 节。

17.3.2. 客户端连接问题

客户端一侧可能出现的错误种类很多，并且依赖于具体应用，
但其中有一些可能直接与服务器的启动方式有关。除下面列出的几种情况外，
其他问题应查阅相应客户端应用的文档。

```
psql: could not connect to server: Connection refused
        Is the server running on host "server.joe.com" and accepting
        TCP/IP connections on port 5432?
```

这种常见的失败表示“我找不到可通信的服务器”。尝试 TCP/IP 通信时，它就会显示为上述形式。一个常见错误是忘记配置服务器以允许 TCP/IP 连接。

另一种情况是，尝试通过 Unix 域套接字与本地服务器通信时，会得到以下消息：

```
psql: could not connect to server: No such file or directory
        Is the server running locally and accepting
        connections on Unix domain socket "/tmp/.s.PGSQL.5432"?
```

最后一行有助于确认客户端是否正尝试连接到正确的位置。如果那里实际上没有运行服务器，
内核错误消息通常会像示例那样是 `Connection refused` 或 `No such file or directory`。（要注意，在这里 `Connection refused` 并不表示服务器收到了你的连接请求并拒绝了它；那种情况会产生另一条消息，如第 19.4 节所示。）其他错误消息，例如 `Connection timed out`，可能表示更底层的问题，例如网络不通。

17.4. 管理内核资源

一个大型PostgreSQL安装会很快耗尽各种操作系统资源限制。（在某些系统上，
出厂默认值非常低，你甚至不需要真正“大型”的安装就会遇到问题。）如果你遇到过这类问题，
请继续阅读。

17.4.1. 共享内存和信号量

共享内存和信号量统称为“System V IPC”（连同消息队列，但消息队列与PostgreSQL无关）。
几乎所有现代操作系统都提供这些特性，但其中许多系统默认没有开启它们或尺寸不足，
尤其是随着可用 RAM 和数据库应用的需求增长。（在 Windows 上，PostgreSQL 为这些设施提供了自己的替代实现，因此本节的大部分内容可以忽略。）

完全缺少这些功能时，通常会在服务器启动时出现 Illegal system call 错误。在这种情况下，除了重新配置内核别无选择。没有它们，PostgreSQL 就无法工作。不过，在现代操作系统中，这种情况很少见。

当PostgreSQL超出各种硬性IPC限制之一时，服务器会拒绝启动，
并应留下带有指导性的错误消息，说明问题所在以及应如何处理（另见第 17.3.1 节）。相关内核参数在不同系统上的命名基本一致，表 17.1给出了概览；不过，设置它们的方法却各不相同。下面给出一些平台上的建议。

表 17.1. System V IPC参数

名称	描述	合理的值
`SHMMAX`	共享内存段的最大尺寸（字节）	至少数 MB（见正文）

名称	描述	合理的值
SHMMIN	共享内存段的最小尺寸（字节）	1
SHMALL	可用共享内存的总量（字节或页面）	如果是字节，同SHMMAX；如果是页面，为ceil(SHMMAX/PAGE_SIZE)
SHMSEG	每个进程的最大共享内存段数目	只需要 1 段，但是默认值高很多
SHMMNI	系统范围内的最大共享内存段数目	像SHMSEG外加其他应用的空间
SEMMNI	信号量标识符（即，集合）的最大数目	至少 ceil((max_connections + autovacuum_max_workers + 4) / 16)
SEMMNS	系统范围内的最大信号量数目	ceil((max_connections + autovacuum_max_workers + 4) / 16) * 17，并为其他应用保留空间
SEMMSL	每个集合中信号量的最大数目	至少 17
SEMMAP	信号量映射中的项数	见文本
SEMVMX	信号量的最大值	至少 1000（默认值常常是 32767，如非必要不要更改）

最重要的共享内存参数是 SHMMAX，即共享内存段的最大字节数。如果你从 shmget 得到类似“Invalid argument”的错误消息，很可能是超出了这个限制。所需共享内存段的大小取决于若干 PostgreSQL 配置参数，如表 17.2所示。（你可能得到的任何错误消息都会包含失败分配请求的确切大小。）作为一种临时解决方案，你可以降低其中某些设置来避免失败。虽然 SHMMAX 小到 2 MB 时也能让 PostgreSQL 运行，但要获得可接受的性能需要大得多的值。理想的设置在几百 MB 到几 GB 之间。

有些系统还对系统中的共享内存总量有限制（SHMALL）。请确保它足以容纳 PostgreSQL 以及其他使用共享内存段的应用程序。注意，在许多系统上，SHMALL 以页而非字节为单位。

不太可能出问题的是共享内存段的最小尺寸（SHMMIN），对PostgreSQL来说应该最多大约是 500 kB。而系统范围（SHMMNI）或每个进程（SHMSEG）的最大共享内存段数目不太可能会导致问题，除非你的系统把它们设成零。

PostgreSQL 为每个允许的连接（max_connections）和自动清理工作进程（autovacuum_max_workers）各使用一个信号量，每 16 个组成一组。每组还包含第 17 个信号量，其中存放一个“魔数”，用来检测与其他应用程序使用的信号量集的冲突。系统中的信号量最大数量由 SEMMNS 设置，因此它必须至少等于 max_connections 加上 autovacuum_max_workers，再为每 16 个允许的连接和工作进程额外增加一个信号量（参见表 17.1中的公式）。参数 SEMMNI 限制系统中同时存在的信号量集的数量。因此，此参数必须至少为 ceil((max_connections + autovacuum_max_workers + 4) / 16)。降低允许的连接数，可以临时规避 semget 函数的失败；此类失败通常会报告令人困惑的“No space left on device”。

在某些情况下，可能还需要增大SEMMAP，使其至少与SEMMNS处于同一数量级。
此参数定义信号量资源映射的大小；映射中每个连续的可用信号量块都需要占用一项。
每当一个信号量集合被释放时，它要么会并入与该释放块相邻的现有项，
要么会登记为一个新的映射项。如果映射已满，被释放的信号量就会丢失（直到重启）。因此，
随着时间推移，信号量空间碎片化可能会导致可用信号量少于应有数量。

SEMMSL 参数决定一个信号量集中可以包含多少个信号量，对于PostgreSQL，它必须至少为 17。

与“信号量撤销”有关的其他各种设置，如SEMMNU和SEMUME，不会影响PostgreSQL。

AIX

至少从 5.1 版起，应该不需要对 SHMMAX 等参数进行任何特殊配置，因为它似乎已配置为允许将全部内存用作共享内存。这也是 DB/2 等其他数据库常用的配置。

不过，可能需要修改 /etc/security/limits 中的全局 ulimit 信息，因为文件大小（fsize）和文件数量（nofiles）的默认硬限制可能过低。

BSD/OS

共享内存. 默认只支持 4 MB 共享内存。请记住，共享内存不可换页；它被锁定在 RAM 中。要增加系统支持的共享内存量，可在内核配置文件中加入类似下面的内容：

```
options "SHMALL=8192"
options "SHMMAX=\(SHMALL*PAGE_SIZE\)"
```

SHMALL 以 4 kB 页为单位，因此值 1024 表示 4 MB 共享内存。上面的设置把最大共享内存区增加到 32 MB。对于 4.3 或更高版本，可能还需要把 KERNEL_VIRTUAL_MB 增大到默认值 248 以上。完成所有修改后，重新编译内核并重启。

信号量. 你可能还需要增加信号量的数量；系统默认总量 60 只能支持大约 50 个 PostgreSQL 连接。在内核配置文件中设置所需的值，例如：

```
options "SEMMNI=40"
options "SEMMNS=240"
```

FreeBSD

默认设置只适合小型安装（例如，默认 SHMMAX 为 32 MB）。可以通过 sysctl 或 loader 接口进行修改。下列参数可用sysctl设置：

```
# sysctl kern.ipc.shmall=32768
# sysctl kern.ipc.shmmax=134217728
```

要让这些设置在重启之后也保持，请修改/etc/sysctl.conf。

这些信号量相关设置对于 sysctl 是只读的；设置它们时，可修改 /boot/loader.conf：

```
kern.ipc.semmni=256
kern.ipc.semmns=512
kern.ipc.semmnu=256
```

修改这些值后，需要重启才能使新设置生效。（注意：FreeBSD 不使用 SEMMAP。旧版本会接受但忽略 kern.ipc.semmap 设置；新版本则干脆拒绝它。）

你可能还希望配置内核，把共享内存锁定在 RAM 中，防止其被换出到交换区。这可以通过 sysctl 设置 kern.ipc.shm_use_phys 来实现。

如果通过启用 sysctl 的 security.jail.sysvipc_allowed 在 FreeBSD jail 中运行，应让不同 jail 中的 postmaster 使用不同的操作系统用户。这能提高安全性，因为它可以防止非 root 用户干扰不同 jail 中的共享内存或信号量，也能让 PostgreSQL 的 IPC 清理代码正常工作。（在 FreeBSD 6.0 及更高版本中，IPC 清理代码无法正确检测其他 jail 中的进程，导致无法在不同 jail 中使用同一端口运行 postmaster。）

4.0 之前的 FreeBSD 版本与 OpenBSD 的行为相同（见下文）。

NetBSD

在 NetBSD 5.0 及更高版本中，可以通过以下命令调整 IPC 参数：`sysctl`。例如：

```
# sysctl -w kern.ipc.semmni=100
```

要让这些设置在重启后仍然生效，请修改 `/etc/sysctl.conf`。

你可能还希望配置内核，把共享内存锁定在 RAM 中，防止其被换出到交换区。这可以通过 `sysctl` 设置 `kern.ipc.shm_use_phys` 来实现。

5.0 之前的 NetBSD 版本与 OpenBSD 的行为相同（见下文），但设置参数时应使用关键字 `options`，而非 `option`。

OpenBSD

编译内核时需要启用 `SYSVSHM` 和 `SYSVSEM` 选项（它们默认启用）。共享内存的最大大小由选项 `SHMMAXPGS`（以页为单位）决定。以下示例展示了如何设置各项参数：

```
option        SYSVSHM
option        SHMMAXPGS=4096
option        SHMSEG=256

option        SYSVSEM
option        SEMMNI=256
option        SEMMNS=512
option        SEMMNU=256
option        SEMMAP=256
```

你可能还希望配置内核，把共享内存锁定在 RAM 中，防止其被换出到交换区。这可以通过 `sysctl` 设置 `kern.ipc.shm_use_phys` 来实现。

HP-UX

默认设置通常足以满足普通安装的需求。在 HP-UX 10 上，`SEMMNS` 的出厂默认值为 128，对较大的数据库站点来说可能过低。

可以在 System Administration Manager（SAM）的 Kernel Configuration→Configurable Parameters 中设置 IPC 参数。完成后选择 Create A New Kernel。

Linux

默认最大段大小为 32 MB，这只够非常小型的 PostgreSQL 安装使用。默认最大总大小为 2097152 页。除使用“大页”的特殊内核配置外，页大小几乎总是 4096 字节（使用 `getconf PAGE_SIZE` 验证）。这意味着默认上限为 8 GB，通常够用，但并非总是如此。

可以通过以下接口更改共享内存大小设置：`sysctl`。例如，要允许 16 GB：

```
$ sysctl -w kernel.shmmax=17179869184
$ sysctl -w kernel.shmall=4194304
```

此外，可以将这些设置保存在以下文件中，使其在重启后仍然生效：`/etc/sysctl.conf`。强烈建议这样做。

很旧的发行版可能没有 `sysctl` 程序，但可以通过操作 `/proc` 文件系统进行等效更改：

```
$ echo 17179869184 >/proc/sys/kernel/shmmax
$ echo 4194304 >/proc/sys/kernel/shmall
```

其他默认值都相当充裕，通常不需要更改。

Mac OS X

在 Mac OS X 中配置共享内存的推荐方法是创建一个名为`/etc/sysctl.conf`的文件，其中包含这样的变量赋值：

```
kern.sysv.shmmax=4194304
kern.sysv.shmmin=1
kern.sysv.shmmni=32
kern.sysv.shmseg=8
kern.sysv.shmall=1024
```

注意，在某些 OS X 版本中，全部五个共享内存参数都必须在`/etc/sysctl.conf`中设置，否则这些值会被忽略。

请注意，近期 OS X 版本会忽略将 `SHMMAX` 设为非 4096 整数倍的值的尝试。

在这个平台上，`SHMALL` 以 4 kB 的页为单位。

在较旧的 OS X 版本中，必须重启才能使共享内存参数的变更生效。从 10.5 起，除了 `SHMMNI`，都可以使用 sysctl 在线更改。但最好仍通过 `/etc/sysctl.conf` 设置所需的值，以便重启后保留这些值。

文件 `/etc/sysctl.conf` 只在 OS X 10.3.9 及更高版本中生效。如果运行的是更早的 10.3.x 版本，必须编辑文件 `/etc/rc` 并更改以下命令中的值：

```
sysctl -w kern.sysv.shmmax
sysctl -w kern.sysv.shmmin
sysctl -w kern.sysv.shmmni
sysctl -w kern.sysv.shmseg
sysctl -w kern.sysv.shmall
```

注意，`/etc/rc` 通常会被 OS X 系统更新覆盖，因此每次更新后可能都需要重新进行这些编辑。

在 OS X 10.2 及更早版本中，请改为编辑 `/System/Library/StartupItems/SystemTuning/SystemTuning` 文件中的这些命令。

SCO OpenServer

在默认配置中，每个共享内存段只允许 512 kB。要提高此设置，首先切换到目录`/etc/conf/cf.d`。要显示`SHMMAX`的当前值，运行：

```
./configure -y SHMMAX
```

要为`SHMMAX`设置新值，运行：

```
./configure SHMMAX=value
```

其中*`value`*是你要使用的新值（以字节为单位）。设置`SHMMAX`后，重新构建内核：

```
./link_unix
```

然后重启。

Solaris 2.6 至 2.9（Solaris 6 至 Solaris 9）

可以在以下文件中更改相关设置：`/etc/system`。例如：

```
set shmsys:shminfo_shmmax=0x2000000
set shmsys:shminfo_shmmin=1
set shmsys:shminfo_shmmni=256
set shmsys:shminfo_shmseg=256

set semsys:seminfo_semmap=256
set semsys:seminfo_semmni=512
set semsys:seminfo_semmns=512
set semsys:seminfo_semmsl=32
```

需要重启才能使变更生效。另请参见http://sunsite.uakom.sk/sunworldonline/swol-09-1997/swol-09-insidesolaris.html，了解旧版 Solaris 中共享内存的信息。

Solaris 2.10（Solaris 10）及更高版本
OpenSolaris

在 Solaris 10 及更高版本和 OpenSolaris 中，默认共享内存和信号量设置足以满足大多数 PostgreSQL 应用的需求。Solaris 现在默认将 `SHMMAX` 设为系统 RAM 的四分之一。要进一步调整此设置，请使用与以下用户关联的项目设置：`postgres`。例如，执行以下命令时使用的用户为 `root`：

```
projadd -c "PostgreSQL DB User" -K "project.max-shm-memory=(
privileged,8GB,deny)" -U postgres -G postgres user.postgres
```

该命令会新增`user.postgres`项目，并把`postgres`用户的共享内存上限设为 8GB；它会在该用户下次登录时或重启PostgreSQL时（不是重新加载）生效。
上述示例假定PostgreSQL由`postgres`组中的`postgres`用户运行。不需要重启操作系统。

对于会承载大量连接的数据库服务器，我们还建议修改以下内核设置：

```
project.max-shm-ids=(priv,32768,deny)
project.max-sem-ids=(priv,4096,deny)
project.max-msg-ids=(priv,4096,deny)
```

此外，如果你正在某个区（zone）中运行PostgreSQL，
可能也需要提高该区的资源使用限制。关于`projects`和`prctl`的更多信息，请参见System Administrator's Guide中的 "Chapter 2: Projects and Tasks"。

UnixWare

在 UnixWare 7 上，默认配置中共享内存段的最大尺寸为 512 kB。要显示`SHMMAX`的当前值，运行：

```
/etc/conf/bin/idtune -g SHMMAX
```

这会显示当前值、默认值、最小值和最大值。要为`SHMMAX`设置新值，运行：

```
/etc/conf/bin/idtune SHMMAX value
```

其中*value*是你要使用的新值（以字节为单位）。设置SHMMAX后，重新构建内核：

```
/etc/conf/bin/idbuild -B
```

然后重启。

表 17.2. PostgreSQL 共享内存用量

用途	所需共享内存字节数的近似值（8.3 版起）
连接	(1800 + 270 * max_locks_per_transaction) * max_connections
自动清理工作进程	(1800 + 270 * max_locks_per_transaction) * autovacuum_max_workers
预备事务	(770 + 270 * max_locks_per_transaction) * max_prepared_transactions
共享磁盘缓冲	(block_size + 208) * shared_buffers
WAL 缓冲	(wal_block_size + 8) * wal_buffers
固定空间需求	770 kB

17.4.2. 资源限制

类 Unix 操作系统会施加多种资源限制，这些限制可能干扰 PostgreSQL 服务器的运行。其中尤其重要的是：每个用户可用的进程数限制、每个进程可打开文件数限制，以及每个进程可用内存量限制。每一种限制都有“硬”限制和“软”限制。实际生效的是软限制，但用户可以自行把它调高到不超过硬限制；硬限制则只能由 root 用户修改。系统调用setrlimit负责设置这些参数。shell 的内置命令ulimit（Bourne shell）或limit（csh）可用于在命令行控制资源限制。在 BSD 派生系统上，/etc/login.conf文件控制登录时设置的各种资源限制。详细信息请参阅操作系统文档。相关参数有maxproc、openfiles和datasize。例如：

```
default:\
...
        :datasize-cur=256M:\
        :maxproc-cur=256:\
        :openfiles-cur=256:\
...
```

（-cur表示软限制；把它改为-max则表示设置硬限制。）

内核还可能对某些资源施加系统范围的限制。

- 在Linux上，内核参数 /proc/sys/fs/file-max确定内核支持的最大打开文件数。可以向该文件写入一个不同的数字来修改它，也可以在/etc/sysctl.conf中添加相应赋值。每个进程的文件数上限是在内核编译时固定的；请参阅 /usr/src/linux/Documentation/proc.txt获取更多信息。

PostgreSQL服务器为每个连接使用一个进程，因此你至少应提供与允许连接数相同数量的进程，再加上系统其他部分所需的进程数。通常这不是问题，但如果你在一台机器上运行多个服务器，资源可能就会变得紧张。

打开文件数的出厂默认限制通常被设为“对共享环境友好”的值，也就是允许许多用户在一台机器上共存，而不会占用不成比例的系统资源。如果你在一台机器上运行很多服务器，这也许正合适；但在专用服务器上，你可能会希望提高这个限制。

另一方面，有些系统允许单个进程打开非常多的文件；如果不止少数几个进程都这么做，就很容易超过系统范围的限制。如果你遇到这种情况，又不想修改系统范围的限制，可以设置PostgreSQL的max_files_per_process配置参数来限制其打开文件的消耗。

17.4.3. Linux 内存过量分配

在 Linux 2.4 及更高版本中，默认的虚拟内存行为对 PostgreSQL 并非最佳。由于内核实现内存“overcommit”（过量分配）的方式，如果 PostgreSQL 或其他进程的内存需求导致系统耗尽虚拟内存，内核可能会终止 PostgreSQL 的 postmaster（主管服务器进程）。

如果发生这种情况，你会看到类似下面的内核消息（至于应到哪里查看此类消息，请参阅你的系统文档和配置）：

```
Out of Memory: Killed process 12345 (postgres).
```

这表示 `postgres` 进程因内存压力而被终止。尽管现有数据库连接仍会继续正常运行，但新连接将不再被接受。要恢复服务，必须重启PostgreSQL。

避免该问题的一种办法，是让PostgreSQL运行在一台你能确定不会被其他进程耗尽内存的机器上。如果内存紧张，增加操作系统交换空间也有助于避免这个问题，因为内存不足（OOM）杀手只有在物理内存和交换空间都耗尽时才会被触发。

如果导致系统耗尽内存的正是 PostgreSQL 自身，那么你可以通过调整配置来避免该问题。在某些情况下，调低与内存相关的配置参数会有帮助，尤其是 `shared_buffers` 和 `work_mem`。在其他情况下，允许数据库服务器接受过多连接也可能导致这一问题。很多时候，更好的做法可能是减小 `max_connections`，并转而使用外部连接池软件。

在 Linux 2.6 及更高版本中，可以修改内核的行为，使其不会“过量分配”内存。虽然此设置无法完全阻止 OOM 杀手[1]被调用，但会显著降低发生概率，从而使系统行为更健壮。方法是选择严格的过量分配模式，使用以下命令：`sysctl`：

```
sysctl -w vm.overcommit_memory=2
```

或者在以下文件中加入等效条目：`/etc/sysctl.conf`。你可能还希望修改相关设置 `vm.overcommit_ratio`。详细信息参见内核文档文件 `Documentation/vm/overcommit-accounting`。

另一种方法可在修改或不修改 `vm.overcommit_memory` 的情况下使用：把 postmaster 进程专属的`oom_adj`值设为 `-17`，从而保证它不会成为 OOM 杀手的目标。最简单的做法是在postmaster 启动脚本中、调用 postmaster 之前执行：

```
echo -17 > /proc/self/oom_adj
```

请注意，这个操作必须以 root 身份完成，否则不会生效；因此，由 root 拥有的启动脚本是最容易执行该操作的位置。如果这样做，还可以在构建 PostgreSQL 时向 `CPPFLAGS` 添加 `-DLINUX_OOM_ADJ=0`。这会使 postmaster 子进程以常规的 `oom_adj`值零运行，以便 OOM 杀手在需要时仍可将它们作为目标。

> **注意**
>
> 据报告，某些厂商的 Linux 2.4 内核包含 2.6 版过量分配 `sysctl` 参数的早期实现。但是，在没有相关代码的 2.4 内核上将 `vm.overcommit_memory` 设为 2，会使情况更糟，而

[1] http://lwn.net/Articles/104179/

> 非更好。建议在 2.4 安装上尝试之前，检查实际内核源码（参见文件 mm/mmap.c 中的函数 vm_enough_memory），确认内核支持的功能。存在 overcommit-accounting 文档文件，不能作为该功能已存在的证据。如有任何疑问，请咨询内核专家或内核供应商。

17.5. 关闭服务器

有几种方法可以关闭数据库服务器。通过向主管进程发送不同的信号，可以控制关闭类型，该进程为 postgres。

SIGTERM

这是智能关闭模式。收到SIGTERM后，服务器禁止新连接，但允许现有会话正常结束工作。仅在所有会话终止后才会关闭。如果服务器处于在线备份模式，还会等待该模式结束。在线备份模式生效期间，仍允许新的超级用户连接（这一例外允许超级用户连接以结束在线备份模式），但不允许其他新连接。如果服务器在请求智能关闭时处于恢复状态，则只有在所有常规会话终止后，恢复和流复制才会停止。

SIGINT

这是快速关闭模式。服务器禁止新连接并向所有现有服务器进程发送SIGTERM，这将导致它们中止当前事务并迅速退出。然后等待所有服务器进程退出，最后关闭。如果服务器处于在线备份模式，该模式将被终止，导致备份不可用。

SIGQUIT

这是立即关闭模式。主 postgres 进程将向所有子进程发送 SIGQUIT 并立即退出，而不对自身进行正常的关闭。子进程在收到 SIGQUIT 时同样会立即退出。这会导致下次启动时通过重放 WAL 日志执行恢复。仅建议在紧急情况下使用。

pg_ctl程序提供了一个便于发送这些信号以关闭服务器的接口。另外，在非 Windows 系统上，你也可以用kill直接发送这些信号。可以用ps程序，或者从数据目录中的postmaster.pid文件找到postgres进程的PID。例如，要执行一次快速关闭：

```
$ kill -INT `head -1 /usr/local/pgsql/data/postmaster.pid`
```

> **重要**
>
> 最好不要使用 SIGKILL 关闭服务器。这样做会阻止服务器释放共享内存和信号量，因此在启动新服务器之前，可能必须手动释放它们。此外，SIGKILL 会终止 postgres 进程，使其没有机会将信号转发给子进程，因此还必须手动逐个终止子进程。

要终止单个会话并允许其他会话继续运行，可使用pg_terminate_backend()（参阅表 9.56），或者向与该会话相关的子进程发送SIGTERM信号。

17.6. 升级 PostgreSQL 集簇

本节讨论如何把数据库数据从一个PostgreSQL发行版升级到较新的发行版。

PostgreSQL的主版本由版本号的前两组数字表示，例如 8.4。PostgreSQL的次版本由版本号的第三组数字表示，例如 8.4.2 就是 8.4 的第二个次版本。次版本从不改变内部存储格式，并且始终与

同一主版本号中的更早和更晚次版本兼容，例如 8.4.2 与 8.4、8.4.1 和 8.4.6 兼容。要在兼容版本之间更新，只需在服务器关闭时替换可执行文件，然后重启服务器即可。数据目录保持不变 – 次版本升级就是这么简单。

对于 PostgreSQL 的major版本，内部数据存储格式可能会发生变化，因此升级会复杂得多。将数据迁移到新的主版本的传统方法是转储并重新载入数据库。此外还有其他方法可用，
如下所述。

新的主版本通常还会引入一些用户可见的不兼容变化，因此应用程序可能也需要做出修改。
所有用户可见的变化都列在发行说明（附录 E）中；请特别留意标为"迁移（Migration）"的小节。如果你跨越多个主版本进行升级，请务必阅读每个中间版本的发行说明。

谨慎的用户会希望在完全切换之前，先在新版本上测试客户端应用。因此，
同时安装旧版本和新版本通常是个好主意。测试PostgreSQL主版本升级时，
可以考虑以下几类可能的变化：

管理

管理员可用于监控和控制服务器的功能，在每个主版本中常常都会变化并有所增强。

SQL

通常这意味着 SQL 命令能力会增加，而行为不会变化，除非发行说明中特别提到。

库 API

通常像libpq这样的库只会增加新功能，除非发行说明中特别说明。

系统目录

系统目录的变化通常只影响数据库管理工具。

服务器端 C 语言 API

这涉及使用 C 语言编写的后端函数 API 的变更。这类变更会影响那些深入引用服务器内部后端函数的代码。

17.6.1. 通过pg_dump升级数据

一种升级方法是从某个主版本的 PostgreSQL 转储数据，再在另一个版本中重新载入。必须使用 pg_dump；文件系统级备份方法不起作用。（系统中有相应检查，会阻止你在不兼容版本的 PostgreSQL 上使用数据目录，因此尝试用错误版本的服务器启动某个数据目录，不会造成严重损害。）

建议使用来自新版 PostgreSQL 的 pg_dump 和 pg_dumpall 程序，以利用这些程序可能包含的改进。当前版本的转储程序可以读取从 7.0 起任意服务器版本的数据。

以下说明假定你现有的安装位于`/usr/local/pgsql`目录下，数据区域位于`/usr/local/pgsql/data`。请根据实际情况替换成你的路径。

1. 如果是在制作备份，请确认数据库此时没有正在进行更新。
   更新操作不会影响备份的完整性，但那些变更当然不会包含在备份中。必要时，可修改`/usr/local/pgsql/data/pg_hba.conf`（或等效文件）中的访问权限，
   禁止除你之外的其他人访问数据库。有关访问控制的更多信息见第 19 章。

要备份整个数据库安装，请输入：

pg_dumpall > *outputfile*

制作备份时，你可以使用当前正在运行版本中的 pg_dumpall 命令；如果需要保留 OID（例如把它们用作外键），可以在运行 pg_dumpall 时使用 -o 选项。不过，为了获得最佳结果，尽量使用PostgreSQL 9.1.24 自带的pg_dumpall命令，因为这一版本相较旧版本包含了缺陷修复和改进。这个建议看起来可能有些反常，因为你这时还没有安装新版本；但如果你计划让新旧版本并行安装，遵循这一建议是明智的。在那种情况下，你可以先正常完成新版本安装，稍后再迁移数据，这样也能减少停机时间。

2. 关闭旧服务器：

   pg_ctl stop

   在那些会在开机时自动启动PostgreSQL的系统上，可能已经有某个启动脚本能完成同样的工作。例如，在Red Hat Linux系统上，也许下面这条命令就可以：

   /etc/rc.d/init.d/postgresql stop

   有关服务器启动和停止的细节见第 17 章。

3. 如果是从备份恢复，请重命名或删除旧的安装目录，前提是它不是按版本区分的目录。与其删除，不如重命名，这样如果你遇到问题需要回退，它仍然还在。请记住，该目录可能会占用相当可观的磁盘空间。要重命名该目录，可以使用类似下面的命令：

   mv /usr/local/pgsql /usr/local/pgsql.old

   （请确保把该目录作为一个整体移动，这样相对路径才会保持不变。）

4. 安装新版本的 PostgreSQL，步骤见第 15.4 节。

5. 如有需要，创建一个新的数据库集簇。请记住，执行这些命令时必须登录到专用数据库用户账户（如果你正在升级，就已经拥有这个账户）。

   /usr/local/pgsql/bin/initdb -D /usr/local/pgsql/data

6. 恢复先前的`pg_hba.conf`，以及你对`postgresql.conf`所做的任何修改。

7. 启动数据库服务器，同样要使用该专门的数据库用户账户：

   /usr/local/pgsql/bin/postgres -D /usr/local/pgsql/data

8. 最后，使用以下命令从备份恢复数据：

   /usr/local/pgsql/bin/psql -d postgres -f *outputfile*

   这里应使用 psql。

若要把停机时间降到最低，可以把新服务器安装到不同目录中，并让新旧两个服务器在不同端口上并行运行。这样就可以使用类似下面的命令：

```
pg_dumpall -p 5432 | psql -d postgres -p 5433
```

来传输数据。

17.6.2. 非转储升级方法

pg_upgrade模块允许把一次安装从一个PostgreSQL主版本就地迁移到下一个主版本。
升级可在几分钟内完成。

也可以使用某些复制方法，例如 Slony，创建一个运行较新版本 PostgreSQL 的备库。之所以可行，是因为 Slony 支持不同主版本 PostgreSQL 之间进行复制。该备库可以与旧服务器部署在同一台机器上，也可以部署在不同机器上。一旦它与主库（运行旧版本 PostgreSQL）同步，就可以切换主备角色，让备库成为主库，并关闭旧的数据库实例。
这种计划内切换只会为升级带来几秒钟的停机时间。

17.7. 防止服务器欺骗

当服务器正在运行时，恶意用户不可能取代正常的数据库服务器。然而，当服务器关闭时，
本地用户却可能通过启动自己的服务器来冒充正常服务器。
这个伪造服务器可以读取客户端发出的密码和查询语句，但由于`PGDATA`
目录仍然受目录权限保护，它无法返回任何真实数据。之所以会出现这种冒充，
是因为任何用户都可以启动一个数据库服务器；而客户端除非经过专门配置，
否则无法识别服务器是否为伪造。

防止`local`连接被欺骗的最简单办法，是使用一个 Unix 域套接字目录（unix_socket_directory），并且只允许受信任的本地用户对该目录拥有写权限。
这样可以防止恶意用户在其中创建自己的套接字文件。如果你担心某些应用程序仍会引用`/tmp`
中的套接字文件，因此仍然容易受到欺骗，那么可以在操作系统启动时创建一个符号链接`/tmp/.s.PGSQL.5432`，使其指向迁移后的套接字文件。你可能还需要修改`/tmp`清理脚本，
防止该符号链接被删除。

要防止 TCP 连接上的服务器欺骗，最佳办法是使用 SSL 证书，并确保客户端检查服务器证书。为此，服务器必须配置为仅接受 `hostssl` 连接（第 19.1 节），并且具备 SSL `server.key`（密钥）和 `server.crt`（证书）文件（第 17.9 节）。TCP 客户端必须使用 `sslmode=verify-ca` 或 `verify-full` 连接，并安装适当的根证书文件（第 31.1 节）。

17.8. 加密选项

PostgreSQL提供了多个层次的加密，并在防止数据因数据库服务器被盗、
不诚实的管理员或不安全网络等原因而泄露方面提供了很高的灵活性。为了保护医疗记录、
金融交易等敏感数据，也可能需要使用加密。

密码存储加密

默认情况下，数据库用户密码以 MD5 哈希形式存储，因此管理员无法得知分配给用户的实际密码。

指定列加密

pgcrypto模块允许以加密形式存储特定字段。这在只有部分数据敏感时很有用。
客户端提供解密密钥，数据在服务器端解密后再发送给客户端。

当数据被解密并在服务器与客户端之间传输时，
解密后的数据和解密密钥会在服务器端短暂存在。
这就给那些可以完全访问数据库服务器的人（例如系统管理员）
提供了一个短暂的机会来截获密钥和数据。

在 Linux 上，可以使用“loopback device”（环回设备）在文件系统之上叠加加密。这允许把整个文件系统分区在磁盘上加密，并由操作系统解密。在 FreeBSD 上，等价的设施称为 GEOM Based Disk Encryption（gbde），许多其他操作系统也支持这类功能，包括 Windows。

数据分区加密

这种机制可以防止在整台计算机或磁盘被盗时，从驱动器中直接读取未加密数据。
但它无法防御文件系统已挂载时的攻击，因为在挂载之后，
操作系统会提供数据的解密视图。不过，要挂载该文件系统，
就必须以某种方式把加密密钥提供给操作系统，
而有时这个密钥就存放在挂载该磁盘的主机上的某处。

跨网络安全密码

`MD5` 认证方法在把密码发送给服务器之前先在客户端对其进行双重加密：先基于用户名做一次 MD5 加密，再基于建立数据库连接时服务器发送的随机盐做一次加密。经过双重加密的值才是实际通过网络发送到服务器的内容。双重加密不仅能防止密码被窃取，
还能防止他人日后用同一个加密后的密码连接到数据库服务器。

跨网络加密数据

SSL 连接会加密通过网络发送的全部数据：密码、查询和返回的数据。`pg_hba.conf` 文件允许管理员指定哪些主机可以使用非加密连接（`host`），哪些主机必须使用 SSL 加密连接（`hostssl`）。此外，客户端可以指定仅通过 SSL 连接到服务器。也可以使用 Stunnel 或 SSH 加密传输。

SSL 主机认证

客户端和服务器都可以向对方提供 SSL 证书。这需要双方做一些额外配置，但它提供的认证强度高于单纯依赖密码。它可以防止某台计算机短暂冒充服务器，
只为读取客户端发送的密码；同时也有助于防御“man in the middle”攻击，即某台位于客户端与服务器之间的计算机冒充服务器，读取并转发双方之间的全部数据。

客户端加密

如果服务器所在机器的系统管理员不可信，那么就有必要由客户端自行加密数据。
这样一来，未加密的数据就永远不会出现在数据库服务器上。
数据在发送给服务器之前就在客户端加密，而查询结果也必须在客户端解密后才能使用。

17.9. 使用 SSL 的安全 TCP/IP 连接

PostgreSQL原生支持使用SSL连接来加密客户端与服务器之间的通信，以提高安全性。
这要求客户端和服务器系统上都安装了OpenSSL，并且在构建
PostgreSQL时启用了该支持（见第 15 章）。

如果构建时启用了 SSL 支持，那么在 `postgresql.conf` 中将参数ssl设为 `on`，就可以在启动 PostgreSQL 服务器时启用 SSL。服务器会在同一个 TCP 端口上同时监听普通连接和 SSL 连接，并与每个连接进来的客户端协商是否使用 SSL。默认情况下，这由客户端决定；关于如何把服务器配置为要求某些或全部连接必须使用 SSL，请参阅第 19.1 节。

PostgreSQL会读取系统范围的OpenSSL配置文件。默认情况下，该文件名为`openssl.cnf`，
位于`openssl version -d`报告的目录中。可以通过把环境变量`OPENSSL_CONF`设为所需配置文件的名称，覆盖这一默认值。

OpenSSL支持种类繁多、强度不一的密码套件和认证算法。虽然可以在OpenSSL配置文件中指定密码套件列表，但你也可以通过修改postgresql.conf中的ssl_ciphers，专门为数据库服务器指定要使用的密码套件。

> **注意**
>
> 使用NULL-SHA或NULL-MD5密码套件，可以在没有加密开销的情况下完成认证。不过，中间人仍然能够读取并转发客户端与服务器之间的通信。此外，与认证的开销相比，加密开销本身很小。基于这些原因，不建议使用 NULL 密码套件。

要以SSL模式启动服务器，服务器的数据目录中必须存在 server.crt 和 server.key 文件。这两个文件应分别包含服务器证书和私钥。在 Unix 系统上，server.key的权限必须禁止组和其他用户访问；可以通过chmod 0600 server.key实现。如果私钥受口令保护，服务器会提示输入口令，并在输入之前一直等待，不会启动。

在某些情况下，服务器证书可能不是由客户端直接信任的证书颁发机构签名，而是由一个 “intermediate” 证书颁发机构签名。要使用这样的证书，把签名机构的证书追加到 server.crt 文件，然后是其上级机构的证书，如此逐级直到一个受客户端信任的 “root” 机构。只要 server.crt 包含多于一个证书，就应当把根证书包含在内。

17.9.1. 使用客户端证书

要要求客户端提供受信任的证书，请把你信任的证书颁发机构（CAs）的证书放入数据目录中的 root.crt 文件，并在 pg_hba.conf 中相应的 hostssl 行上把 clientcert 参数设为 1。这样，SSL 连接启动时就会向客户端请求证书。（关于如何在客户端设置证书，请参见第 31.17 节。）服务器会验证客户端证书是否由某个受信任的证书颁发机构签名。如果存在 root.crl 文件，还会检查证书吊销列表（CRL）条目。（展示 SSL 证书用法的示意图可参见http://h71000.www7.hp.com/doc/83final/ba554_90007/ch04s02.html。）

注意，root.crt 列出的是被认为可信任为客户端证书签名的顶层 CA。原则上它不必列出为服务器证书签名的 CA，不过在大多数情况下，那个 CA 对于客户端证书也会是受信任的。

pg_hba.conf 中的 clientcert 选项适用于所有认证方法，但只适用于指定为 hostssl 的行。未指定 clientcert 或将其设为 0 时，如果存在 root.crt 文件，服务器仍会依据它验证客户端提供的证书 – 但不会强制要求客户端提供证书。

如果正在设置客户端证书，可以考虑使用 cert 认证方法，让证书在提供连接安全性的同时也控制用户认证。详情参见第 19.3.10 节。

17.9.2. SSL 服务器文件用法

表 17.3概括了服务器端 SSL 设置相关的文件。

表 17.3. SSL 服务器文件用法

文件	内容	效果
$PGDATA/server.crt	服务器证书	发送给客户端，用于表明服务器身份
$PGDATA/server.key	服务器私钥	证明服务器证书是其所有者发送的，并不说明证书所有者是值得信任的

文件	内容	效果
`$PGDATA/root.crt`	受信任的证书颁发机构	检查客户端证书是由一个受信任的证书颁发机构签名的
`$PGDATA/root.crl`	被证书颁发机构吊销的证书	客户端证书不能出现在这个列表上

文件`server.key`、`server.crt`、`root.crt`和`root.crl`只在服务器启动期间检查；因此要让这些文件中的变更生效，必须重启服务器。

17.9.3. 创建自签名证书

要为服务器快速创建一个自签名证书，可以使用下面的OpenSSL命令：

```
openssl req -new -text -out server.req
```

按提示填写 openssl 询问的信息。确保把本地主机名作为“Common Name”输入；挑战口令可以留空。程序会生成一个受口令保护的密钥；它不接受少于四个字符的口令。
要去掉口令（如果希望服务器自动启动，就必须这样做），可以运行命令：

```
openssl rsa -in privkey.pem -out server.key
rm privkey.pem
```

输入旧口令以解锁现有密钥。然后执行：

```
openssl req -x509 -in server.req -text -key server.key -out server.crt
```

把证书转换为自签名证书，并把密钥和证书复制到服务器查找它们的位置。最后执行：

```
chmod og-rwx server.key
```

如果该文件权限比这更宽松，服务器会拒绝使用它。
关于如何创建服务器私钥和证书的更多细节，请参阅OpenSSL文档。

虽然自签名证书可用于测试，但在生产环境中应使用由证书颁发机构（CA）（既可以是全球性 CA 之一，也可以是本地机构）签名的证书，以便客户端能够验证服务器的身份。如果客户端全部位于组织内部，推荐使用本地 CA。

17.10. 通过 SSH 隧道建立安全 TCP/IP 连接

可以使用SSH来加密客户端与PostgreSQL服务器之间的网络连接。如果配置得当，
即使客户端本身不支持 SSL，这种方法也能提供足够安全的网络连接。

首先确认在PostgreSQL服务器所在的同一台机器上有一个正常运行的SSH服务器，
并且你可以使用`ssh`以某个用户身份登录；然后就可以建立通向远程服务器的安全隧道。
安全隧道会监听本地端口，并把所有流量转发到远程机器上的某个端口。
发送到该远程端口的流量可以到达其localhost地址，或者在需要时到达其他绑定地址；
对远端来说，这些流量看起来并不是来自你的本地机器。
下面这个命令会创建一条从客户端机器到远程机器`foo.com`的安全隧道：

```
ssh -L 63333:localhost:5432 joe@foo.com
```

`-L`参数中的第一个数字 63333 是隧道在本地的端口号；它可以是任何未使用的端口。（IANA 保留 49152 到 65535 端口供私用。）其后的名称或 IP 地址是你要连接的远程绑定地址，这里是默认值localhost。第二个数字 5432 是隧道远端的端口，例如数据库服务器所使用的端口。要通过这条隧道连接数据库服务器，只需连接本地机器上的 63333 端口：

```
psql -h localhost -p 63333 postgres
```

对数据库服务器来说，它看到的是来自主机foo.com上用户joe对localhost绑定地址的连接，并将应用为该用户到该绑定地址所配置的认证方式。请注意，服务器不会认为这是 SSL 加密连接，因为实际上SSH服务器与PostgreSQL服务器之间并没有加密。
不过这通常不会带来额外安全风险，因为两者位于同一台机器上。

要让这条隧道建立成功，你必须有权通过 ssh 以 joe@foo.com 身份连接，就像尝试使用 ssh 创建终端会话一样。

你也可以把端口转发设成这样：

```
ssh -L 63333:foo.com:5432 joe@foo.com
```

但这样一来，数据库服务器会把该连接看作是到达其foo.com绑定地址的连接，而默认设置 listen_addresses = 'localhost'并不会监听该地址。这通常不是你想要的结果。

如果你必须经由某个登录主机“跳转”到数据库服务器，那么一种可能的设置如下：

```
ssh -L 63333:db.foo.com:5432 joe@shell.foo.com
```

请注意，这种从shell.foo.com到db.foo.com的连接不会受到 SSH 隧道的加密保护。当网络受到各种限制时，SSH 还提供了许多其他配置方式。详情请参阅 SSH 文档。

提示

还有一些其他应用也能提供安全隧道，其思路与刚才描述的 SSH 方法类似。

第 18 章 服务器配置

有许多配置参数会影响数据库系统的行为。在本章的第一节中，我们描述如何设置配置参数。后续各节将详细讨论每个参数。

18.1. 设置参数

所有参数名都是大小写不敏感的。每个参数都可以接受五种类型之一的值：布尔、整数、浮点数、字符串或枚举。布尔值可以被写成on、off、true、false、yes、no、1、0（都是大小写不敏感的）或者这些值的任何无歧义前缀。

一些设置指定内存或时间值。这些设置中的每一个都有隐含单位，它可能是千字节、块（通常为 8 千字节）、毫秒、秒或分钟。默认单位可以通过参考 pg_settings.unit找到。为方便起见，也可以显式指定不同的单位。有效的内存单位是 kB（千字节）、MB（兆字节）和GB（吉字节）；有效的时间单位是 ms（毫秒）、s（秒）、min（分钟）、h（小时）和d（天）。注意，内存单位的乘数是 1024，而不是 1000。

“枚举”类型的参数以与字符串参数相同的方式指定，但被限制到一组有限的值。允许的值可以在 pg_settings.enumvals 中找到。枚举参数值是大小写无关的。

设置这些参数的一种方法是编辑文件 postgresql.conf，它通常保存在数据目录中。（在数据库集簇目录初始化时，会在那里安装该文件的一个默认副本。）该文件的内容可能如下所示：

```
# This is a comment
log_connections = yes
log_destination = 'syslog'
search_path = '"$user", public'
shared_buffers = 128MB
```

每行指定一个参数。名称和值之间的等号是可选的。空白不重要，空行会被忽略。井号（#）表示该行余下部分是注释。不是简单标识符或数字的参数值必须用单引号括起。要在参数值中嵌入单引号，可以写两个单引号（推荐）或使用反斜线转义单引号。

除了单个参数设置之外，postgresql.conf文件还可以包含 include 指令，用来指定另一个要读取和处理的文件，就像把该文件插入到配置文件的这个位置一样。include 指令的形式很简单：

```
include 'filename'
```

如果文件名不是绝对路径，则会被解释为相对于引用它的配置文件所在目录的路径。include 可以嵌套。

每当主服务器进程收到SIGHUP信号（最简单的方法是使用pg_ctl reload发送）时，配置文件都会被重新读取。主服务器进程还会把这个信号传播给所有正在运行的服务器进程，这样现有的会话也能采用新值。另外，你可以直接向单一的服务器进程发送该信号。有些参数只能在服务器启动时设置，在配置文件中对这些条目的任何更改都将被忽略，直到服务器重启。

设置这些配置参数的第二种方法是作为postgres命令的命令行选项给出它们，例如：

```
postgres -c log_connections=yes -c log_destination='syslog'
```

命令行选项会覆盖postgresql.conf中的任何冲突设置。注意，这意味着你将无法通过编辑postgresql.conf即时更改该值，因此虽然命令行方法可能很方便，但它之后可能会让你付出灵活性的代价。

偶尔，只给一个特定会话提供命令行选项是有用的。环境变量 PGOPTIONS可以在客户端用于此目的：

```
env PGOPTIONS='-c geqo=off' psql
```

（这适用于任何基于libpq的客户端应用，而不仅是 psql。）注意，这对于在服务器启动时固定或者必须在postgresql.conf中指定的参数无效。

此外，可以将一组参数设置分配给一个用户或一个数据库。每当会话启动时，会加载所涉及的用户和数据库的默认设置。ALTER ROLE和ALTER DATABASE命令分别被用来配置这些设置。每数据库设置会覆盖从 postgres命令行或配置文件得到的任何设置，转而又被每用户设置覆盖；两者都会被每会话设置覆盖。

有些参数可以在单独的SQL会话中用SET命令更改，例如：

```
SET ENABLE_SEQSCAN TO OFF;
```

如果允许SET，它会覆盖该参数所有其他的值来源。有些参数不能通过 SET更改：例如，如果它们控制的行为不重启整个 PostgreSQL服务器就无法改变。此外，某些SET或ALTER参数修改要求超级用户权限。

SHOW命令允许检查所有参数的当前值。

虚拟表pg_settings也允许显示和更新会话运行时参数；不同变量类型以及何时可以更改它们的详情和描述请参见第 45.62 节。pg_settings等效于SHOW 和SET，但使用起来可能更方便，因为它可以与其他表连接，或者用任何想要的选择条件从中选择。它还包含关于参数允许哪些值的更多信息。

18.2. 文件位置

除了已经提到过的postgresql.conf文件之外，PostgreSQL还使用另外两个手工编辑的配置文件，它们控制客户端认证（其使用在第 19 章中讨论）。默认情况下，所有三个配置文件都存放在数据库集簇的数据目录中。本节描述的参数允许配置文件放在别的地方（这么做可以简化管理，特别是如果配置文件被独立放置，可以很容易保证它得到恰当的备份）。

data_directory (string)

指定用于数据存储的目录。这个参数只能在服务器启动时设置。

config_file (string)

指定主服务器配置文件（通常叫postgresql.conf）。这个参数只能在postgres命令行上设置。

hba_file (string)

指定基于主机认证配置文件（通常叫pg_hba.conf）。这个参数只能在服务器启动的时候设置。

ident_file (string)

指定用于第 19.2 节用户名称映射的配置文件（通常叫pg_ident.conf）。这个参数只能在服务器启动的时候设置。

external_pid_file (string)

指定服务器应创建的额外进程 ID（PID）文件的名称，供服务器管理程序使用。这个参数只能在服务器启动的时候设置。

在默认安装中不会显式设置以上参数。相反，命令行参数-D或者环境变量PGDATA指定数据目录，并且上述配置文件都能在数据目录中找到。

如果希望把配置文件放在数据目录之外的地方，postgres的-D命令行选项或PGDATA环境变量必须指向包含配置文件的目录，并且必须在postgresql.conf（或在命令行上）设置data_directory参数，以指出数据目录实际所在的位置。注意，data_directory会覆盖-D和PGDATA对数据目录位置的指定，但不会覆盖配置文件的位置。

如果你愿意，可以使用选项config_file、hba_file和/或ident_file单独指定配置文件名称和位置。config_file只能在postgres命令行上指定，但是其他参数可以在主配置文件中设置。如果所有三个参数外加data_directory被显式地设置，则不必指定-D或PGDATA。

在设置任何这些参数时，相对路径将被解释为相对于启动 postgres 时所在目录的路径。

18.3. 连接和认证

18.3.1. 连接设置

listen_addresses (string)

指定服务器用于监听客户端应用连接的 TCP/IP 地址。此值采用以逗号分隔的主机名和/或数字 IP 地址列表的形式。特殊项*对应所有可用的 IP 接口。项0.0.0.0允许监听所有 IPv4 地址，而::允许监听所有 IPv6 地址。如果列表为空，服务器不会监听任何 IP 接口，此时只能通过 Unix 域套接字连接。默认值为localhost，只允许建立本地 TCP/IP “回环”连接。在客户端认证（第 19 章）允许对谁可以访问服务器进行细粒度控制的同时，listen_addresses 控制哪些接口接受连接尝试，这可以帮助防止在不安全的网络接口上重复恶意连接请求。此参数只能在服务器启动时设置。

port (integer)

服务器监听的 TCP 端口，默认是 5432。请注意，服务器监听的所有 IP 地址都使用同一个端口号。此参数只能在服务器启动时设置。

max_connections (integer)

决定数据库服务器允许的最大并发连接数。默认值通常是 100 个连接，但如果内核设置不支持（在 initdb 期间确定），则可能更少。这个参数只能在服务器启动时设置。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存或信号量。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

当运行一个备库时，你必须设置这个参数等于或大于主库上的参数。否则，备库上将不允许查询。

superuser_reserved_connections (integer)

决定为 PostgreSQL 超级用户连接保留多少个连接“槽”。同时活跃的连接数最多始终只能达到max_connections。当活跃并发连接数至少达到 max_connections 减去 superuser_reserved_connections 时，新连接将只接受超级用户，并且不再接受新的复制连接。

默认值是 3 个连接。该值必须小于 max_connections。这个参数只能在服务器启动时设置。

unix_socket_directory (string)

指定服务器用于监听来自客户端应用连接的 Unix 域套接字的目录。默认值通常是/tmp，但可以在构建时更改。这个参数只能在服务器启动时设置。

除了套接字文件本身（名为.s.PGSQL.*nnnn*，其中*nnnn*是服务器的端口号），还会在unix_socket_directory目录中创建一个名为.s.PGSQL.*nnnn*.lock的普通文件。这两个文件都不应该被手工移除。

Windows 没有 Unix 域套接字，因此此参数在 Windows 上没有意义。

unix_socket_group (string)

设置 Unix 域套接字的所属组（套接字的所属用户总是启动服务器的用户）。可以与选项unix_socket_permissions一起用于对 Unix域连接进行访问控制。默认是一个空字符串，表示服务器用户的默认组。这个参数只能在服务器启动时设置。

Windows 没有 Unix 域套接字，因此此参数在 Windows 上没有意义。

unix_socket_permissions (integer)

设置 Unix 域套接字的访问权限。Unix 域套接字使用通常的 Unix 文件系统权限集。参数值应是以 chmod 和 umask 系统调用所接受格式指定的数字权限模式。（要使用惯用的八进制格式，数字必须以 0（零）开头。）

默认权限是 0777，表示任何人都可以连接。合理的其他取值包括 0770（仅属主和所属组，另见 unix_socket_group）和 0700（仅属主）。（注意，对 Unix 域套接字而言，只有写权限起作用，因此设置或撤销读权限和执行权限没有意义。）

此访问控制机制独立于第 19 章中描述的机制。

此参数只能在服务器启动时设置。

此参数对完全忽略套接字权限的系统无效，尤其是 Solaris（截至 Solaris 10）。在这些系统上，可以将 unix_socket_directory 指向一个仅向目标用户授予搜索权限的目录，以达到类似效果。Windows 没有 Unix 域套接字，因此此参数在 Windows 上也没有意义。

bonjour (boolean)

启用通过 Bonjour 通告服务器存在的功能。默认值为关闭。此参数只能在服务器启动时设置。

bonjour_name (string)

指定 Bonjour 服务名。空字符串 ''（默认值）表示使用计算机名。如果编译时未启用 Bonjour 支持，则此参数会被忽略。此参数只能在服务器启动时设置。

tcp_keepalives_idle (integer)

指定在一个原本空闲的连接上发送 keepalive 包之前要等待的秒数。值 0 表示使用系统默认值。此参数仅在支持TCP_KEEPIDLE或TCP_KEEPALIVE符号的系统以及 Windows 上可用；在其他系统上，它必须为零。在通过 Unix 域套接字连接的会话中，此参数会被忽略，并始终读作零。

> **注意**
>
> 在 Windows 上，值 0 会将此参数设置为 2 小时，因为 Windows 不提供读取系统默认值的方法。

`tcp_keepalives_interval` (`integer`)

指定在一个原本空闲的连接上两次发送 keepalive 之间间隔的秒数。值 0 表示使用系统默认值。此参数仅在支持`TCP_KEEPINTVL`符号的系统以及 Windows 上可用；在其他系统上，它必须为零。在通过 Unix 域套接字连接的会话中，此参数会被忽略，并始终读作零。

> **注意**
>
> 在 Windows 上，值 0 会将此参数设置为 1 秒，因为 Windows 不提供读取系统默认值的方法。

`tcp_keepalives_count` (`integer`)

指定在一个原本空闲的连接上要发送的 keepalive 包数量。值 0 表示使用系统默认值。此参数仅在支持`TCP_KEEPCNT`符号的系统上可用；在其他系统上，它必须为零。在通过 Unix 域套接字连接的会话中，此参数会被忽略，并始终读作零。

> **注意**
>
> Windows 不支持此参数，它必须为零。

18.3.2. 安全和认证

`authentication_timeout` (`integer`)

允许完成客户端认证的最长时间，以秒为单位。
如果一个客户端没有在这段时间里完成认证协议，服务器将关闭连接。
这样就避免了出问题的客户端无限制地占有一个连接。默认值是 1分钟（`1m`）。这个参数只能在服务器命令行上或者在`postgresql.conf`文件中设置。

`ssl` (`boolean`)

启用SSL连接。使用前请阅读第 17.9 节。默认值是`off`。这个参数只能在服务器启动时设置。
SSL通信只能在 TCP/IP 连接上使用。

`ssl_renegotiation_limit` (`integer`)

指定在一个SSL加密连接上可以流过多少数据之后将进行会话密钥的重新协商。
当可以检查大量流量时，重新协商能降低攻击者进行密码分析的机会，
但它也会带来很大的性能损失。该限制以发送和接收流量的总和来检查。
如果这个参数被设置为 0，则禁用重新协商。默认值为`0`。

注意

2009 年 11 月之前的 SSL 库在使用 SSL 重新协商时不安全，原因是 SSL 协议中的一个漏洞。作为针对该漏洞的权宜修复，一些厂商发布的 SSL 库不具备重新协商能力。如果在客户端或服务器上正在使用这类库，应当禁用 SSL 重新协商。

警告

由于 OpenSSL 中的 bug，启用 SSL 重新协商（即把 `ssl_renegotiation_limit` 配置为非零）很可能导致诸如长连接中断之类的问题。

`ssl_ciphers` (`string`)

指定允许在安全连接上使用的SSL密码列表。支持的密码列表请参见openssl手册页。

`password_encryption` (`boolean`)

在CREATE USER或ALTER ROLE中指定密码，且既没有写 `ENCRYPTED` 也没有写 `UNENCRYPTED` 时，此参数决定密码是否被加密。默认值为 `on`（加密密码）。

`krb_server_keyfile` (`string`)

设置服务器的Kerberos密钥文件的位置。详情请参考第 19.3.5 节或第 19.3.3 节。这个参数只能在`postgresql.conf`文件中或者服务器命令行上设置。

`krb_srvname` (`string`)

设置Kerberos服务名。详情请参考第 19.3.5 节。这个参数只能在`postgresql.conf`文件中或者服务器命令行上设置。

`krb_caseins_users` (`boolean`)

设置是否应该以大小写不敏感的方式对待Kerberos和GSSAPI用户名。默认值是`off`（大小写敏感）。这个参数只能在`postgresql.conf`文件中或者服务器命令行上设置。

`db_user_namespace` (`boolean`)

此参数启用各数据库独立的用户名。默认关闭。此参数只能在`postgresql.conf`文件中或在服务器命令行上设置。

如果启用此参数，你应以`username@dbname`的形式创建用户。
连接的客户端传入`username`时，会在用户名后附加`@`和数据库名，
然后由服务器查找此数据库专属的用户名。请注意，在 SQL 环境中创建名称包含`@`的用户时，需要用引号括起用户名。

启用此参数后，仍然可以创建普通的全局用户。只需在客户端指定用户名时附加`@`，例如`joe@`。服务器查找用户名之前会去掉`@`。

`db_user_namespace`会使客户端和服务器的用户名表示形式不同。
认证检查始终使用服务器端的用户名，因此认证方法必须针对服务器端的用户名配置，
而不是客户端的用户名。由于`md5`在客户端和服务器上都使用用户名作为盐值，`md5`
不能与`db_user_namespace`一起使用。

> **注意**
>
> 此特性旨在作为找到完整解决方案之前的临时措施。届时将移除此选项。

18.4. 资源消耗

18.4.1. 内存

`shared_buffers` (`integer`)

设置数据库服务器用于共享内存缓冲区的内存量。默认值通常为 32 兆字节（`32MB`），但如果内核设置不支持，则可能更小（在 initdb 期间确定）。此设置必须至少为 128 千字节。（`BLCKSZ` 的非默认值会改变该最小值。）不过，要获得良好性能，通常需要远高于该最小值的设置。此参数只能在服务器启动时设置。

如果专用数据库服务器具有 1GB 或更多内存，`shared_buffers` 的合理初始值是系统内存的 25%。对于某些工作负载，将 `shared_buffers` 设得很大也有效，但由于 PostgreSQL 同时依赖操作系统缓存，将超过 40% 的内存分配给 `shared_buffers` 不太可能比更小的值效果更好。将 `shared_buffers` 设得更大时，通常还需要相应增加 `checkpoint_segments`，以便将大量新数据或已修改数据的写入过程分散到更长的时间内。

对于内存少于 1GB 的系统，适合使用更小的内存比例，以便为操作系统留出足够空间。另外，在 Windows 上，很大的 `shared_buffers` 设置并不那么有效。让该设置保持相对较低、更多地使用操作系统缓存，可能得到更好的结果。Windows 系统上 `shared_buffers` 的有效范围通常是从 64MB 到 512MB。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

`temp_buffers` (`integer`)

设置每个数据库会话使用的临时缓冲区的最大数量。这些是会话本地的缓冲区，仅用于访问临时表。默认值为 8 兆字节（`8MB`）。可以在单个会话内更改此设置，但必须在该会话首次使用临时表之前更改；此后尝试更改该值，对该会话不会产生影响。

会话会按需分配临时缓冲区，上限为 `temp_buffers`。对于实际不需要很多临时缓冲区的会话，将此参数设得较大时，开销仅为 `temp_buffers` 每增加一就多分配一个缓冲区描述符，约为 64 字节。不过，如果实际使用了某个缓冲区，还会为它额外消耗 8192 字节（一般而言为 `BLCKSZ` 字节）。

`max_prepared_transactions` (`integer`)

设置可同时处于“预备”状态的事务的最大数量（见PREPARE TRANSACTION）。将此参数设为零（默认值）会禁用预备事务功能。此参数只能在服务器启动时设置。

如果不打算使用预备事务，应将此参数设为零，以防意外创建预备事务。如果使用预备事务，通常应将 `max_prepared_transactions` 设为不小于max_connections的值，以便每个会话都能有一个待处理的预备事务。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

运行备库时，必须将此参数设为与主库相同或更大的值。否则，备库上将不允许执行查询。

work_mem (integer)

指定内部排序操作和哈希表在写入临时磁盘文件之前可使用的内存量。默认值为 1 兆字节（1MB）。请注意，复杂查询可能并行执行多个排序或哈希操作，
每个操作在开始向临时文件写入数据之前，都可以使用此值指定的内存量。此外，
多个正在运行的会话也可能并发执行此类操作。因此，使用的总内存量可能是 work_mem 值的数倍；选择此值时必须考虑这一点。排序操作用于 ORDER BY、DISTINCT 和归并连接。
哈希表用于哈希连接、基于哈希的聚合、以及基于哈希的 IN 子查询处理。

maintenance_work_mem (integer)

指定维护操作（如 VACUUM、CREATE INDEX 和 ALTER TABLE ADD FOREIGN KEY）可使用的最大内存量。默认值为 16 兆字节（16MB）。由于一个数据库会话一次只能执行一个此类操作，而一个数据库系统通常也不会并发运行很多此类操作，
因此可以安全地将该值设得远大于 work_mem。更大的设置可能改善清理和恢复数据库转储的性能。

注意，自动清理运行时，最多可能分配此内存量的autovacuum_max_workers倍，
因此不要将默认值设得过高。

max_stack_depth (integer)

指定服务器执行栈的最大安全深度。
此参数的理想设置是由内核强制执行的实际栈大小限制（如由ulimit -s或本地等效设置），减去大约一兆字节的安全余量。
需要安全余量是因为服务器中并非每个例程都检查栈深度，
而只在可能递归的关键例程（如表达式求值）中检查。默认设置为两兆字节（2MB），
这是保守且不太可能引起崩溃的小值。但是，这可能太小，无法执行复杂函数。
只有超级用户才能更改此设置。

把max_stack_depth
参数设置得高于实际的内核限制将意味着一个失控的递归函数可能会导致一个独立的后端进程崩溃。在PostgreSQL能够检测内核限制的平台上，
服务器将不允许把这个参数设置为一个不安全的值。不过，并非所有平台都能提供该信息，
所以我们还是建议你在选择值时要小心。

18.4.2. 内核资源使用

max_files_per_process (integer)

设置每个服务器子进程允许同时打开的最大文件数量。默认值为一千个文件。
如果内核强制实施了安全的每进程上限，就不必担心此设置。
但在某些平台上（尤其是大多数 BSD 系统），内核允许单个进程打开的文件数量很大，如果很多进程都尝试打开这么多文件，就会远超系统实际能够支持的总量。如果遇到“Too many open files”（打开的文件过多）错误，可尝试减小此设置。此参数只能在服务器启动时设置。

shared_preload_libraries (string)

这个变量指定一个或多个要在服务器启动时预载入的共享库。例如，'$libdir/mylib'
将导致从安装的标准库目录中预载入mylib.so（或在某些平台上是mylib.sl）。
除非用双引号括起，所有库名都会被转换为小写。如果要载入多个库，
用逗号分隔它们的名字。这个参数只能在服务器启动时设置。

可以用这个方法预装载PostgreSQL的过程语言库，通常是使用'$libdir/plXXX'语法，其中的XXX是pgsql、perl、tcl或python。

通过预载入一个共享库，当该库被第一次使用时就可以避免库的启动时间。不过，启动每个新服务器进程的时间可能会略有增加，即使该进程从不使用该库。因此，推荐只把这个参数用于那些要在大多数会话中使用的库上。

> **注意**
>
> 在 Windows 主机上，在服务器启动时预载入一个库并不会减少启动每个新服务器进程所需的时间；每一个服务器进程将会重新载入所有预载入的库。不过，shared_preload_libraries在 Windows 主机上仍然有用，因为某些共享库可能需要执行一些只在 postmaster 启动时才能进行的操作（例如，共享库可能需要预留轻量级锁或共享内存，而这些在 postmaster 启动之后就做不到了）。

如果指定的库没有找到，服务器将无法启动。

每一个 PostgreSQL 支持的库都有一个被检查以保证兼容性的“魔法块”。因此，非 PostgreSQL 的库不能用这种方式载入。

18.4.3. 基于代价的清理延迟

执行VACUUM和ANALYZE命令期间，系统维护一个内部计数器，记录已执行的各种 I/O 操作的估算代价。当累计代价达到上限（由 vacuum_cost_limit 指定）时，执行该操作的进程会休眠一小段时间，时长由 vacuum_cost_delay 指定。随后重置计数器并继续执行。

此功能让管理员能够降低这些命令对并发数据库活动的 I/O 影响。在许多情况下，VACUUM 和 ANALYZE 等维护命令是否快速完成并不重要，但避免它们显著干扰系统执行其他数据库操作的能力通常很重要。基于代价的清理延迟为管理员提供了实现这一点的方法。

对于手动执行的 VACUUM 命令，此功能默认禁用。要启用它，将 vacuum_cost_delay 变量设为非零值。

vacuum_cost_delay (integer)

超过代价上限后，进程将休眠的时长，单位为毫秒。默认值为零，表示禁用基于代价的清理延迟功能。正值会启用基于代价的清理。注意，在许多系统上，休眠延迟的有效分辨率为 10 毫秒；将 vacuum_cost_delay 设为不是 10 的倍数的值，可能与将它设为下一个更大的 10 的倍数效果相同。

使用基于代价的清理时，vacuum_cost_delay 的合适值通常较小，例如 10 或 20 毫秒。调整清理的资源消耗时，最好更改其他清理代价参数。

vacuum_cost_page_hit (integer)

清理在共享缓冲区缓存中找到的缓冲区时所计入的估算代价。它表示锁定缓冲池、查找共享哈希表和扫描页内容的代价。默认值为 1。

vacuum_cost_page_miss (integer)

清理必须从磁盘读取的缓冲区时所计入的估算代价。它表示锁定缓冲池、查找共享哈希表、从磁盘读取所需数据块并扫描其内容所需的工作量。默认值为 10。

`vacuum_cost_page_dirty` (`integer`)

清理操作修改原本干净的数据块时所计入的估算代价。它表示再次将脏块刷盘所需的额外 I/O。默认值为 20。

`vacuum_cost_limit` (`integer`)

会使清理进程休眠的累计代价。默认值为 200。

> **注意**
>
> 某些操作持有关键的锁，因此应尽快完成。这些操作期间不会发生基于代价的清理延迟，所以累计代价可能远超指定上限。为避免此时出现无益的长时间延迟，实际延迟按 `vacuum_cost_delay` * `accumulated_balance` / `vacuum_cost_limit` 计算，但最大不超过 `vacuum_cost_delay` * 4。

18.4.4. 后台写入器

有一个独立的服务器进程，称为后台写入器，负责写出“脏”（新的或修改过的）共享缓冲区。当干净的共享缓冲区数量似乎不足时，后台写入器会将一些脏缓冲区写入文件系统，并将其标记为干净。这可以降低处理用户查询的服务器进程找不到干净缓冲区、因而不得不自行写出脏缓冲区的可能性。不过，后台写入器确实会使总体 I/O 负载有所增加：反复变脏的页面原本可能在每个检查点间隔中只写出一次，而后台写入器可能在同一间隔内随着它变脏而多次写出。本节参数可用于根据实际需求调整此行为。

`bgwriter_delay` (`integer`)

指定后台写入器各轮活动之间的延迟。每一轮中，写入器会对一定数量的脏缓冲区发出写操作（由下面的参数控制），然后休眠 `bgwriter_delay` 毫秒，再重复此过程。默认值为 200 毫秒（`200ms`）。注意，在许多系统上，休眠延迟的有效分辨率为 10 毫秒；将 `bgwriter_delay` 设为不是 10 的倍数的值，可能与将它设为下一个更大的 10 的倍数效果相同。此参数只能在 `postgresql.conf` 文件中或服务器命令行上设置。

`bgwriter_lru_maxpages` (`integer`)

后台写入器每轮写出的缓冲区数量不会超过此值。设为零会禁用后台写入（检查点活动除外）。默认值为 100 个缓冲区。此参数只能在 `postgresql.conf` 文件中或服务器命令行上设置。

`bgwriter_lru_multiplier` (`floating point`)

每轮写出的脏缓冲区数量取决于最近几轮服务器进程所需的新缓冲区数量。将近期平均需求乘以 `bgwriter_lru_multiplier`，即可估算下一轮所需的缓冲区数量。写入器会写出脏缓冲区，直到可用的干净且可重用缓冲区达到这一数量。（不过，每轮写出的缓冲区数量不会超过 `bgwriter_lru_maxpages`。）因此，设为 1.0 表示采用“恰好及时”策略，写出的缓冲区数量恰好等于预测需求量。更大的值可为需求突增留出余量，而更小的值则有意将部分写操作留给服务器进程执行。默认值为 2.0。此参数只能在 `postgresql.conf` 文件中或服务器命令行上设置。

较小的 `bgwriter_lru_maxpages` 和 `bgwriter_lru_multiplier` 可以降低后台写入器造成的额外 I/O 负载，但也会增加服务器进程必须自行发出写操作的可能性，从而延迟交互式查询。

18.4.5. 异步行为

`effective_io_concurrency` (`integer`)

设置 PostgreSQL 预期可以同时执行的并发磁盘 I/O 操作数量。提高此值会增加单个 PostgreSQL 会话尝试并行发起的 I/O 操作数量。允许的范围为 1 至 1000，或设为零以禁用异步 I/O 请求。目前，此设置仅影响位图堆扫描。

对于磁盘，可以将为数据库提供存储的 RAID 0 条带或 RAID 1 镜像中的独立磁盘数量作为合理初始值。（对于 RAID 5，不应计入校验盘。）不过，如果数据库经常忙于执行并发会话发出的多个查询，较小的值可能就足以使磁盘阵列保持繁忙。
超过使磁盘保持繁忙所需的值只会增加 CPU 开销。

对于更特殊的系统，例如基于内存的存储或受总线带宽限制的 RAID 阵列，正确的值可能是可用 I/O 通路的数量。可能需要一些实验才能找到最佳值。

异步 I/O 依赖于有效的 `posix_fadvise` 函数，而某些操作系统缺少此函数。如果该函数不存在，将此参数设为任何非零值都会报错。在某些操作系统（如 Solaris）上，该函数虽然存在，却实际上不做任何事情。

18.5. 预写式日志

更多关于 WAL 和检查点调优的细节另见第 29.4 节。

18.5.1. 设置

`wal_level` (`enum`)

`wal_level`决定多少信息写入到 WAL 中。默认值是`minimal`，它只写入从崩溃或立即关机中进行恢复所需的信息。`archive`会增加 WAL 归档所需的日志，`hot_standby`会进一步增加在备库上运行只读查询所需的信息。这个参数只能在服务器启动时设置。

在`minimal`级别，可以安全地跳过一些批量操作的 WAL 日志记录，使这些操作快得多（参见第 14.4.7 节）。可以应用此优化的操作包括：

```
CREATE TABLE AS
CREATE INDEX
CLUSTER
```
向同一事务中创建或截断的表执行`COPY`

但 minimal 级别的 WAL 不包含足够的信息来从基础备份和 WAL 日志重建数据，因此必须使用`archive`或`hot_standby`级别来启用 WAL 归档（archive_mode）和流复制。

在`hot_standby`级别上，记录与`archive`相同的信息，以及从 WAL 重构正在运行事务的状态所需的信息。要在备库上启用只读查询，必须把主库上的`wal_level`设置为`hot_standby`，并且在备库上启用hot_standby。据认为，使用`hot_standby`和`archive`级别在性能上几乎没有可测量的差别，因此如果注意到任何生产环境影响，欢迎反馈。

`fsync` (`boolean`)

如果打开这个参数，PostgreSQL服务器将尝试确保更新被物理地写入到磁盘，做法是发出`fsync()`系统调用或者使用多种等价的方法（见wal_sync_method）。
这保证了数据库集簇在一次操作系统或者硬件崩溃后能恢复到一个一致的状态。

虽然关闭fsync常常可以得到性能上的收益，
但当发生断电或系统崩溃时可能造成不可恢复的数据损坏。因此，
只有在能很容易地从外部数据中重建整个数据库时才建议关闭fsync。

可以安全关闭fsync的情形包括：从备份文件初始装载一个新数据库集簇；
用数据库集簇处理一批数据，处理后就丢弃并重建该数据库；
或者使用经常重建且不用于故障切换的只读数据库克隆。仅有高质量硬件不足以成为关闭fsync的理由。

在很多情况下，为非关键事务关闭synchronous_commit，可以获得关闭fsync
所带来的大部分潜在性能收益，同时避免伴随的数据损坏风险。

fsync只能在postgresql.conf文件中或在服务器命令行上设置。如果你关闭这个参数，
请也考虑关闭full_page_writes。

synchronous_commit (enum)

指定事务提交是否等待 WAL 记录写入磁盘后，命令才向客户端返回“成功”指示。有效值为on、local和off。默认且安全的值是on。当为off时，向客户端报告成功后，
可能还要经过一段时间，事务才能真正保证不会因服务器崩溃而丢失。（最大延迟为wal_writer_delay的三倍。）与fsync不同，将此参数设为off
不会带来数据库不一致的风险：操作系统或数据库崩溃可能会使一些最近报告已提交的事务丢失，但数据库状态会与这些事务已正常中止时完全相同。因此，
当性能比完全确保事务持久性更重要时，关闭synchronous_commit
可以是一种有用的替代方案。更多讨论见第 29.3 节。

如果设置了synchronous_standby_names，此参数还控制事务提交是否等待该事务的 WAL 记录刷盘到磁盘并复制到备库。
提交等待将持续到当前的同步备库回复表明它已将该事务的提交记录写入持久存储为止。
如果使用同步复制，通常合理的做法是：要么等待 WAL 记录同时到达本地和远程磁盘，要么允许事务异步提交。不过，对于那些希望等待本地刷盘但不等待同步复制的事务，
可以使用特殊值local。

此参数可以随时更改；每个事务的行为由提交时生效的设置决定。因此，
让一些事务同步提交、另一些事务异步提交是可行且有用的。例如，
当默认设置要求同步提交时，可以在一个包含多条语句的事务中执行SET LOCAL synchronous_commit TO OFF，使该事务异步提交。

wal_sync_method (enum)

用于将 WAL 更新强制写入磁盘的方法。如果fsync关闭，此设置就没有作用，因为 WAL 文件更新根本不会被强制写入磁盘。可选值为：

- open_datasync（用open()选项O_DSYNC写 WAL 文件）
- fdatasync（在每次提交时调用fdatasync()）
- fsync（在每次提交时调用fsync()）
- fsync_writethrough（在每次提交时调用fsync()，强制穿透任何磁盘写缓存）
- open_sync（用open()选项O_SYNC写 WAL 文件）

open_* 选项还会使用O_DIRECT（如果可用）。不是在所有平台上都能使用所有这些选择。
默认值是列表中第一个被平台支持的那个，不过fdatasync是 Linux 中的默认值。默认值不

一定最合适；可能需要更改此设置或系统配置的其他方面，
以确保崩溃时的数据安全或达到最佳性能。这些方面在第 29.1 节中讨论。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

full_page_writes (boolean)

启用此参数时，PostgreSQL服务器会在检查点之后首次修改每个磁盘页面时，
将该页面的全部内容写入 WAL。这样做是因为，操作系统崩溃时正在进行的页面写入可能只完成了一部分，导致磁盘页面混有新旧数据。通常存储在 WAL 中的行级变更数据不足以在崩溃恢复时完整还原这样的页面。保存整页镜像能保证正确恢复页面，但会增加必须写入 WAL 的数据量。（由于 WAL 重放总是从检查点开始，只需在检查点之后首次修改每个页面时这样做。因此，减少整页写入开销的一种方法是增大检查点间隔参数。）

关闭此参数可以加快正常操作，
但系统故障后可能出现不可恢复的数据损坏或静默数据损坏。风险与关闭fsync类似，
虽然较小，但也只有在建议关闭fsync的相同情形下，才应关闭此参数。

关闭这个选项并不影响用于时间点恢复（PITR）的 WAL 归档使用（见第 24.3 节）。

这个参数只能在postgresql.conf文件中或在服务器命令行上设置。默认值是on。

wal_buffers (integer)

用于还未写入磁盘的 WAL 数据的共享内存量。默认值 -1 选择等于shared_buffers的 1/32 的尺寸（大约3%），但是不小于64kB也不大于 WAL 段的尺寸（通常为16MB）。如果自动的选择太大或太小可以手工设置该值，但是任何小于32kB的正值都将被当作32kB。
这个参数只能在服务器启动时设置。

在每次事务提交时，WAL 缓冲区的内容被写出到磁盘，因此极大的值不太可能带来显著收益。不过，
把这个值设置为至少几个兆字节可以在一个繁忙的服务器（其中很多客户端会在同一时间提交）上提高写性能。由默认设置 -1 选择的自动调节将在大部分情况下得到合理的结果。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

wal_writer_delay (integer)

指定 WAL 写入器两轮活动之间的延迟。在每一轮中，写入器会把 WAL 刷写到磁盘，然后休眠wal_writer_delay毫秒，再重复这一过程。默认值为 200 毫秒（200ms）。请注意，许多系统的休眠延迟有效分辨率为 10 毫秒；将wal_writer_delay设置为非 10 的倍数的值，可能与设置为下一个更大的 10 的倍数具有相同效果。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

commit_delay (integer)

当一个事务的提交数据被刷盘到磁盘时，当时已准备好的任何其他提交也会一起被刷出。
commit_delay在事务尝试把 WAL 缓冲区刷出到磁盘之前增加一个以微秒为单位的时间延迟。如果系统负载足够高，使得在给定的时间间隔内有额外的事务准备好提交，
那么一个非零的延迟可以让更多的事务只用一次刷盘操作提交。
但如果没有其他事务准备好提交，这个延迟就纯属浪费。因此，
只有当一个服务器进程写入其提交记录的那一刻，至少有commit_siblings
个其他事务处于活动状态时，才会执行延迟。默认的commit_delay为零（无延迟）。
因为无论此设置如何，每次刷盘都会写出所有挂起的提交数据，
所以增大此参数而增加延迟很少能真正提高性能。

commit_siblings (integer)

执行commit_delay延迟前要求的并发活动事务的最小数目。
大一些的值会导致在延迟间隔期间更可能有至少另外一个事务准备好提交。
默认值是五个事务。

18.5.2. 检查点

checkpoint_segments (integer)

自动 WAL 检查点之间日志文件段的最大数量（每个段通常为 16 兆字节）。默认值为三个段。增加这个参数会增加崩溃恢复所需的时间。此参数只能在postgresql.conf文件中或服务器命令行上设置。

checkpoint_timeout (integer)

自动 WAL 检查点之间的最长时间，以秒为单位。合理的范围在 30 秒到 1 小时之间。默认是 5 分钟（5min）。增加这个参数的值可能会增加崩溃恢复所需的时间。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

checkpoint_completion_target (floating point)

指定检查点完成的目标，作为检查点之间总时间的一部分。默认是 0.5。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

checkpoint_warning (integer)

如果由于填充检查点段文件导致的检查点之间的间隔低于这个参数表示的秒数，
那么就向服务器日志写一个消息（它建议增加checkpoint_segments的值）。默认值是 30 秒（30s）。零则关闭警告。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

18.5.3. 归档

archive_mode (boolean)

当启用archive_mode时，完成的WAL段会通过设置archive_command发送到归档存储。
archive_mode和archive_command是单独的变量，这样即使不离开归档模式也可以修改archive_command。这个参数只能在服务器启动时设置。当wal_level被设置为minimal时，不能启用archive_mode。

archive_command (string)

要归档一个已完成的 WAL 文件段而执行的 shell 命令。字符串中的任何%p都被替换成要归档文件的路径名，而%f只被文件名替换。（路径名是相对于服务器的工作目录，
即集簇的数据目录。）如果要在命令里嵌入一个真正的%字符，可以使用%%。
只有当命令成功时才返回零退出状态，这一点很重要。更多信息请见第 24.3.1 节。

这个参数只能在postgresql.conf文件或服务器命令行中设置。
除非在服务器启动时启用了archive_mode，否则将被忽略。如果archive_command是空字符串（默认值），而archive_mode已启用，WAL归档将暂时被禁用，
但服务器将继续积累WAL段文件，期望很快会提供命令。将archive_command设置为一个什么都不做但返回true的命令，例如/bin/true（Windows上为REM），

实际上禁用了归档，但也破坏了用于归档恢复所需的WAL文件链，
因此只应在不寻常的情况下使用。

`archive_timeout` (`integer`)

archive_command只针对已完成的 WAL 段调用。因此，如果您的服务器生成的WAL流量较少（或者在这样做时有间歇期），
在事务完成和安全记录到归档存储之间可能会有很长的延迟。为了限制未归档数据的年龄，
您可以将`archive_timeout`设置为强制服务器定期切换到新的WAL段文件。
当此参数大于零时，只要自上次段文件切换以来经过了指定的秒数，
并且存在任何数据库活动（包括单个检查点），服务器将切换到新的段文件。（增大 `checkpoint_timeout` 可以减少空闲系统上不必要的检查点。）请注意，由于强制切换而提前关闭的归档文件仍然与完全填满的文件长度相同。因此，使用非常短的`archive_timeout`是不明智的，它会使您的归档存储膨胀。通常，将`archive_timeout`
设置为一分钟左右是合理的。如果您希望数据比这更快地从主库复制出来，
您应该考虑使用流复制而不是归档。此参数只能在`postgresql.conf`
文件或服务器命令行中设置。

18.6. 复制

这些设置控制内建的流复制特性（参见第 25.2.5 节）的行为。其中一些参数必须设置在主库上，而另一些必须设置在将要接收复制数据的备库上。

18.6.1. 主库

这些参数可以设置在要把复制数据发送到一个或多个备库的主库上。注意除了这些参数之外，还必须在主库上恰当地设置wal_level，并且通常你还会想要启用 WAL 归档（参见第 18.5.3 节）。这些参数在备库上的值是无关紧要的，尽管你可以为备库将来可能成为主库而预先在那里设置它们。

`max_wal_senders` (`integer`)

指定来自备库或流式基础备份客户端的并发连接的最大数量（即同时运行的 WAL 发送进程的最大数）。默认值为 0，表示禁用复制。WAL 发送进程计入总连接数，因此此参数不能设得高于max_connections。这个参数只能在服务器启动时设置。要允许来自备库的连接，
`wal_level`必须设置为`archive`或`hot_standby`。

`wal_sender_delay` (`integer`)

指定 WAL 发送进程各轮活动之间的延迟。在每一轮中，WAL 发送进程会把自上一轮以来累积的所有 WAL 发送给备库，然后休眠`wal_sender_delay`毫秒，再重复此过程。休眠会被事务提交打断，因此已提交事务的效果会在提交发生时立即发送给备库，不受此设置的影响。
默认值为一秒（`1s`）。注意，在许多系统上，休眠延迟的有效分辨率为 10 毫秒；将`wal_sender_delay`设为不是 10 的倍数的值，可能与将它设为下一个更大的 10 的倍数效果相同。这个参数只能在`postgresql.conf`文件中或在服务器命令行上设置。

`wal_keep_segments` (`integer`)

指定在`pg_xlog`目录中保留的过去日志文件段的最小数量，
以便备库需要获取它们来进行流复制。每个段通常为 16 兆字节。如果连接到主库的备库落后超过`wal_keep_segments`个段，主库可能会移除备库仍需要的 WAL 段，在这种情况下，复制连接将被终止。（但是，如果使用了 WAL 归档，备库可以通过从归档中获取该段来恢复。）

此设置只指定在pg_xlog中保留的最小段数；系统可能需要为 WAL 归档或从检查点恢复而保留更多段。如果wal_keep_segments为零（默认值），系统不会为备库额外保留任何段，因此备库可用的旧 WAL 段数取决于前一个检查点的位置和 WAL 归档的状态。此参数对重启点没有影响。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

vacuum_defer_cleanup_age (integer)

指定VACUUM和HOT更新延迟清理死行版本的事务数。默认值为零个事务，意味着可以尽快移除死行版本，也就是在它们不再对任何打开的事务可见时立即移除。如第 25.5 节所述，在为热备服务器提供支持的主库上，你可能希望将此参数设置为非零值。这让备库上的查询有更多时间完成，而不会因过早清理行而发生冲突。但是，由于此值以主库上发生的写事务数计量，很难预测会给备库查询带来多少额外的宽限时间。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

你也应考虑设置hot_standby_feedback，作为使用此参数的替代方案。

replication_timeout (integer)

终止不活动时间超过指定毫秒数的复制连接。这有助于主库检测备库崩溃或网络中断。值为零将禁用超时机制。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。默认值为 60 秒。

为防止连接被过早终止，必须在备库上启用wal_receiver_status_interval，并且它的值必须小于replication_timeout的值。

synchronous_standby_names (string)

如第 25.2.6 节所述，这个参数指定一个支持同步复制的备库名称的逗号分隔列表。在任一时刻，最多只有一个活动的同步备库；在这个备库确认收到它们的数据之后，等待提交的事务将被允许继续下去。同步备库是列表中第一个被命名、并且当前已连接并且正在实时流式传输数据（如pg_stat_replication视图中streaming的状态所示）的服务器。在这个列表中出现得较晚的其他备库表示潜在的同步备库。如果当前的同步备库因为某种原因断开连接，它将立刻被下一个最高优先级的备库替代。指定多于一台备库名称可以得到非常高的可用性。

用于这一目的的备库名称是其 walreceiver 的primary_conninfo中设置的application_name。没有机制强制唯一性。在出现重复的情况下，匹配的备库之一将被选为同步备库，不过无法弄清到底是哪一个。特殊项*匹配任意application_name，包括默认的应用名walreceiver。

如果这里没有指定同步备库名称，那么不启用同步复制并且事务提交将不会等待复制。这是默认的配置。即便当同步复制被启用时，个体事务也可以被配置为不等待复制，做法是将synchronous_commit参数设置为local或off。

这个参数只能在postgresql.conf文件中或通过服务器命令行进行设置。

18.6.2. 备库

这些设置控制备库接收复制数据时的行为。它们在主库的值是无关的。

hot_standby (boolean)

指定在恢复期间，你是否能够连接并运行查询，如第 25.5 节中所述。默认值是off。这个参数只能在服务器启动时设置。它只在归档恢复期间或备库模式下才有效。

max_standby_archive_delay (integer)

当热备处于活动状态时，
此参数确定备库在取消与即将应用的WAL条目冲突的备库查询之前应等待多长时间，如第 25.5.2 节中所述。max_standby_archive_delay在从WAL归档中读取WAL数据时适用（因此不是当前的）。如果未指定单位，则将其视为毫秒。默认值为30秒。值为-1允许备库永远等待冲突查询完成。此参数只能在postgresql.conf文件或服务器命令行中设置。

注意，max_standby_archive_delay并不等同于查询在被取消前可以运行的最长时间；它表示应用任意一个 WAL 段的数据所允许的最长总时间。因此，如果某个查询先前在处理该 WAL 段时已造成显著延迟，后续冲突查询的宽限时间就会短得多。

max_standby_streaming_delay (integer)

当热备处于活动状态时，
此参数确定备库在取消与即将应用的WAL条目冲突的备库查询之前应等待多长时间，如第 25.5.2 节中所述。max_standby_streaming_delay在通过流复制接收WAL数据时应用。如果未指定单位，则将其视为毫秒。默认值为30秒。值为-1允许备库永远等待冲突查询完成。此参数只能在postgresql.conf文件或服务器命令行中设置。

注意，max_standby_streaming_delay
并不等同于查询在被取消前可以运行的最长时间；它表示从主库接收到 WAL 数据后，允许用于应用这些数据的最长总时间。因此，如果某个查询已造成显著延迟，
后续冲突查询的宽限时间就会短得多，直到备库再次赶上进度。

wal_receiver_status_interval (integer)

指定备库上的 WAL 接收进程向主库发送有关复制进度的信息的最小频度，它可以使用pg_stat_replication视图看到。备库将报告最后写入的事务日志位置、
最后刷盘的位置以及最后应用的位置。这个参数的值是报告之间的最大间隔，以秒为单位。每次写入或刷盘位置改变时会发送状态更新，或者至少按这个参数指定的频度发送。因此，应用位置可能比真实位置略微滞后。将此参数设置为零会完全禁用状态更新。
这个参数只能在postgresql.conf文件中或在服务器命令行上设置。默认值是 10 秒。

当在主库上启用了replication_timeout时，必须启用wal_receiver_status_interval，并且它的值必须小于replication_timeout的值。

hot_standby_feedback (boolean)

指定一个热备库是否将会向主库发送有关于备库上当前正被执行的查询的反馈。
这个参数可以被用来消除由清理记录引起的查询取消，
但是可能导致在主库上用于某些负载的数据库膨胀。反馈消息的发送频度不会高于每个wal_receiver_status_interval周期发送一次。默认值是off。
这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

18.7. 查询规划

18.7.1. 规划器方法配置

这些配置参数提供了一种较为粗糙的方法，用于影响查询优化器所选择的查询计划。
如果优化器为某个特定查询选择的默认计划并不理想，一种临时解决方案是使用这些配置参数，强制优化器选择另一种计划。

改善优化器所选计划质量的更好办法包括调整规划器代价常数（见第 18.7.2 节）、手工运行ANALYZE、增加default_statistics_target配置参数的值，以及使用 ALTER TABLE SET STATISTICS 增加为特定列收集的统计信息量。

enable_bitmapscan (boolean)

启用或禁用查询规划器对位图扫描计划类型的使用。默认值为 on。

enable_hashagg (boolean)

启用或禁用查询规划器对哈希聚合计划类型的使用。默认值为 on。

enable_hashjoin (boolean)

启用或禁用查询规划器对哈希连接计划类型的使用。默认值为 on。

enable_indexscan (boolean)

启用或禁用查询规划器对索引扫描计划类型的使用。默认值为 on。

enable_material (boolean)

允许或者禁止查询规划器使用物化。它不可能完全禁用物化，
但是关闭这个变量将阻止规划器插入物化节点，除非为了保证正确性。默认值是on。

enable_mergejoin (boolean)

启用或禁用查询规划器对归并连接计划类型的使用。默认值为 on。

enable_nestloop (boolean)

允许或禁止查询规划器使用嵌套循环连接计划。它不可能完全禁止嵌套循环连接，
但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是on。

enable_seqscan (boolean)

允许或禁止查询规划器使用顺序扫描计划类型。它不可能完全禁止顺序扫描，
但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是on。

enable_sort (boolean)

允许或禁止查询规划器使用显式排序步骤。它不可能完全禁止显式排序，
但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是on。

enable_tidscan (boolean)

允许或禁止查询规划器使用TID扫描计划类型。默认值是on。

18.7.2. 规划器代价常量

这一节中描述的代价变量可以按照任意尺度衡量。我们只关心它们的相对值，
将它们以相同的因子缩放不会影响规划器的选择。默认情况下，
这些代价变量是基于顺序页面获取的代价的，即seq_page_cost被设置为1.0
并且其他代价变量都参考它来设置。不过你可以使用你喜欢的不同尺度，
例如在一个特定机器上以毫秒为单位的实际执行时间。

> **注意**
>
> 没有明确的方法可以确定代价变量的理想值。
> 最好将它们视为某个数据库安装实例所接收的全部查询组合的平均值。因此，
> 仅凭少数几次试验就修改它们是非常冒险的。

`seq_page_cost`（`floating point`）

设置规划器对一系列顺序磁盘页面读取中单次读取的代价估计。默认值是 1.0。对于某个特定的表空间，可以通过设置该表空间的同名参数来覆盖此值（见ALTER TABLESPACE）。

`random_page_cost`（`floating point`）

设置规划器对一次非顺序磁盘页面读取的代价估计。默认值是 4.0。对于某个特定的表空间，可以通过设置该表空间的同名参数来覆盖此值（见ALTER TABLESPACE）。

减少这个值（相对于`seq_page_cost`）将导致系统更倾向于索引扫描；
提高它将让索引扫描看起来相对更昂贵。你可以一起提高或降低两个值来改变磁盘 I/O 代价相对于 CPU 代价的重要性，后者由下列参数描述。

> **提示**
>
> 尽管系统允许将`random_page_cost`设置得小于`seq_page_cost`，
> 但这不符合实际物理情况。不过，如果数据库完全缓存在 RAM 中，将它们设置为相等是合理的，因为此时非顺序访问页面不会产生额外代价。同样，
> 对于大部分数据已缓存的数据库，应相对于 CPU 参数降低这两个值，因为读取一个已在 RAM 中的页面的代价远小于通常的页面读取代价。

`cpu_tuple_cost`（`floating point`）

设置规划器对一次查询中处理每一行的代价估计。默认值是 0.01。

`cpu_index_tuple_cost`（`floating point`）

设置规划器对一次索引扫描中处理每一个索引项的代价估计。默认值是 0.005。

`cpu_operator_cost`（`floating point`）

设置规划器对于一次查询中处理每个操作符或函数的代价估计。默认值是 0.0025。

`effective_cache_size`（`integer`）

设置规划器对一个单一查询可用的有效磁盘缓存尺寸的假设。
这个参数会被考虑在使用一个索引的代价估计中，
更高的数值会使得索引扫描更可能被使用，更低的数值会使得顺序扫描更可能被使用。
在设置这个参数时，你还应该考虑
PostgreSQL的共享缓冲区以及将被用于PostgreSQL数据文件的内核磁盘缓存。另外，
还要考虑预计在不同表上的并发查询数目，因为它们必须共享可用的空间。这个参数对
PostgreSQL分配的共享内存尺寸没有影响，它也不会预留内核磁盘缓存，
它只用于估计的目的。系统也不会假设在查询之间数据会保留在磁盘缓存中。默认值是 128 兆字节（`128MB`）。

18.7.3. 遗传查询优化器

遗传查询优化器（GEQO）是一种使用启发式搜索进行查询规划的算法。
它可以缩短复杂查询（连接很多关系的查询）的规划时间，
代价是生成的计划有时不如常规穷举搜索算法找到的计划。更多信息见第 51 章。

`geqo` (`boolean`)

允许或禁止遗传查询优化。默认是启用。在生产环境中通常最好不要关闭它。`geqo_threshold`变量提供了对 GEQO 更细粒度的控制。

`geqo_threshold` (`integer`)

只有当涉及的`FROM`项数量至少有这么多个的时候，才使用遗传查询优化（注意一个`FULL OUTER JOIN`只被计为一个`FROM`项）。默认值是 12。对于更简单的查询，通常会使用普通的穷举搜索规划器，但是对于有很多表的查询穷举搜索会花很长时间，
通常比执行一个次优的计划带来的惩罚值还要长。因此，在查询尺寸上的一个阈值是管理 GEQO 使用的一种方便的方法。

`geqo_effort` (`integer`)

控制 GEQO 中规划时间和查询计划质量之间的权衡。此变量必须是 1 到 10 之间的整数。默认值是 5。较大的值会增加查询规划所用的时间，但也会提高选择高效查询计划的可能性。

`geqo_effort`实际并不直接做任何事情；它只是被用来计算其他影响 GEQO 行为的变量（如下所述）的默认值。如果你愿意，你可以手工设置其他参数。

`geqo_pool_size` (`integer`)

控制 GEQO 使用的池尺寸，它就是遗传种群中的个体数目。它必须至少为 2，且有用的值通常在 100 到 1000 之间。如果它被设置为零（默认设置）则会基于`geqo_effort`和查询中表的数量选择一个合适的值。

`geqo_generations` (`integer`)

控制 GEQO 使用的代数，即算法的迭代次数。它必须至少为 1，通常有用的值与池大小处于相同范围。如果设置为零（默认设置），则根据`geqo_pool_size`选择合适的值。

`geqo_selection_bias` (`floating point`)

控制 GEQO 使用的选择偏好。选择偏好是种群中的选择压力。值可以是 1.5 到 2.0 之间，后者是默认值。

`geqo_seed` (`floating point`)

控制 GEQO 使用的随机数生成器的初始值，随机数生成器用于在连接顺序搜索空间中选择随机路径。该值可以从 0（默认值）到 1。变化该值会改变被探索的连接路径集合，并且可能使找到的最优路径变得更好或更差。

18.7.4. 其他规划器选项

`default_statistics_target` (`integer`)

为没有通过`ALTER TABLE SET STATISTICS`设置列相关目标的表列设置默认统计目标。更大的值增加了需要做`ANALYZE`的时间，但是可能会改善规划器的估计质量。默认值是 100。有关PostgreSQL查询规划器使用的统计信息的更多内容，请参考第 14.2 节。

constraint_exclusion(enum)

控制查询规划器对表约束的使用，以优化查询。constraint_exclusion的允许值是on（对所有表检查约束）、off（从不检查约束）和partition（只对继承的子表和UNION ALL子查询检查约束）。partition是默认设置。
它通常与继承和分区表一起使用来提高性能。

当此参数允许对某个表使用约束排除时，规划器会比较查询条件与该表的CHECK约束，
并且忽略那些条件违反约束的表扫描。例如：

```
CREATE TABLE parent(key integer, ...);
CREATE TABLE child1000(check (key between 1000 and 1999)) INHERITS(
parent);
CREATE TABLE child2000(check (key between 2000 and 2999)) INHERITS(
parent);
...
SELECT * FROM parent WHERE key = 2400;
```

在启用约束排除时，这个SELECT将完全不会扫描child1000，从而提高性能。

目前，约束排除仅在通常用于实现表分区的情况下默认启用。
为所有表启用它会增加额外的规划开销，这在简单查询上相当明显，
而且通常不会为简单查询带来好处。如果没有分区表，你可能希望完全关闭它。

更多关于使用约束排除实现分区的信息请参阅第 5.9.4 节。

cursor_tuple_fraction(floating point)

设置规划器对将被检索的一个游标的行的比例的估计。默认值是 0.1。更小的值使得规划器偏向为游标使用“快速开始”计划，
它将很快地检索前几行但是可能需要很长时间来获取所有行。更大的值强调总的估计时间。
最大设置为 1.0，游标将和普通查询完全一样地被规划，只考虑总估计时间并且不考虑前几行会被多快地返回。

from_collapse_limit(integer)

如果生成的FROM列表不超过这么多项，规划器将把子查询融合到上层查询。
较小的值可以减少规划时间，但是可能会生成较差的查询计划。默认值是 8。详见第 14.3 节。

将这个值设置为geqo_threshold或更大，可能触发使用 GEQO 规划器，从而产生非最优计划。见第 18.7.3 节。

join_collapse_limit(integer)

如果得出的列表中不超过这么多项，那么规划器将把显式JOIN（除了FULL JOIN）结构重写到 FROM项列表中。较小的值可减少规划时间，但是可能会生成差些的查询计划。

默认情况下，这个变量被设置成和from_collapse_limit相同，这样适合大多数使用。
把它设置为 1 可避免任何显式JOIN的重排序。因此查询中指定的显式连接顺序就是关系被连接的实际顺序。因为查询规划器并不是总能选取最优的连接顺序，
高级用户可以选择暂时把这个变量设置为 1，然后显式地指定他们想要的连接顺序。更多信息请见第 14.3 节。

将这个值设置为geqo_threshold或更大，可能触发使用 GEQO 规划器，从而产生非最优计划。见第 18.7.3 节。

18.8. 错误报告和日志

18.8.1. 日志记录到哪里

`log_destination`(`string`)

PostgreSQL支持多种记录服务器消息的方法，包括stderr，csvlog和syslog。在 Windows 上，还支持eventlog。将此参数设为所需日志目的地的逗号分隔列表。
默认只将日志记录到stderr。此参数只能在`postgresql.conf`
文件中或在服务器命令行上设置。

如果csvlog被包括在`log_destination`中，日志项会以“逗号分隔值”（CSV）
格式被输出，这样可以很方便地把日志载入到程序中。详见第 18.8.4 节。要产生 CSV 格式的日志输出，必须启用logging_collector。

> **注意**
>
> 在大多数 Unix 系统上，你将需要修改系统的syslog守护进程的配置来使用`log_destination`的syslog选项。PostgreSQL可以在syslog设施`LOCAL0`到`LOCAL7`
> 中记录（见syslog_facility），但是大部分平台上的默认
> syslog配置会丢弃所有这种消息。你将需要增加这样的内容：
>
> ```
> local0.*    /var/log/postgresql
> ```
>
> 到syslog守护进程的配置文件来让它工作。

`logging_collector`(`boolean`)

这个参数启用日志收集器，它是一个捕捉被发送到stderr的日志消息的后台进程，
并且它会将这些消息重定向到日志文件中。这种方法比记录到syslog通常更有用，
因为某些类型的消息可能不会在
syslog输出中出现（一个常见的示例是动态链接器错误消息；另一个示例是由`archive_command`等脚本产生的错误消息）。这个参数只能在服务器启动时设置。

> **注意**
>
> 也可以不使用日志收集器而把日志记录到stderr，日志消息将只会去到服务器的
> stderr被定向到的位置。不过，那种方法只适合于低日志量，
> 因为它没有提供便捷的方法来轮转日志文件。还有，在某些平台上，
> 不使用日志收集器可能会导致日志输出丢失或混杂，
> 因为多个进程并发写入同一个日志文件时会覆盖彼此的输出。

> **注意**
>
> 日志收集器被设计成从来不会丢失消息。这意味着在极高的负载下，
> 如果服务器进程试图在收集器已经落后时发送更多的日志消息，
> 那么它可能会被阻塞。相反，syslog倾向于在无法写入消息时丢掉消息，
> 这意味着在这样的情况下它可能会无法记录某些消息，
> 但是它不会阻塞系统的其他部分。

log_directory (string)

当logging_collector被启用时，这个参数决定日志文件将被在哪个目录下创建。它可以被指定为一个绝对路径，也可以被指定为一个相对于集簇数据目录的相对路径。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。默认是pg_log。

log_filename (string)

当logging_collector被启用时，这个参数设置被创建的日志文件的文件名。该值被视为一种strftime模式，因此%转义可以被用来指定根据时间变化的文件名（注意如果有任何依赖时区的%转义，计算将在由log_timezone指定的时区中完成）。被支持的%转义和开放组织的strftime[1]说明中列举的类似。注意系统的strftime不会被直接使用，因此平台相关（非标准）的扩展无法工作。默认是postgresql-%Y-%m-%d_%H%M%S.log。

如果你不使用转义来指定一个文件名，你应该计划使用一个日志轮转工具来避免最终填满整个磁盘。在 8.4 发行之前，如果不存在%转义，PostgreSQL将追加新日志文件创建时间的纪元，但是现在已经不再这样做了。

如果在log_destination中启用了 CSV 格式输出，.csv将会被追加到时间戳日志文件名中来创建 CSV 格式输出（如果log_filename以.log结尾，该后缀会被替换）。

这个参数只能在postgresql.conf文件中或通过服务器命令行进行设置。

log_file_mode (integer)

在 Unix 系统上，当logging_collector被启用时，这个参数设置日志文件的权限（在微软 Windows 上这个参数将被忽略）。这个参数值应当是一个数字形式的模式，它可以被chmod和umask系统调用接受（要使用通常的八进制格式，该数字必须以一个0（零）开始）。

默认的权限是0600，表示只有服务器拥有者才能读取或写入日志文件。其他常用的设置是0640，它允许拥有者的组成员读取文件。不过要注意你需要修改log_directory为将文件存储在集簇数据目录之外的某个位置，才能利用这个设置。在任何情况下，让日志文件变成任何人都可读是不明智的，因为日志文件中可能包含敏感数据。

这个参数只能在postgresql.conf文件中或通过服务器命令行进行设置。

log_rotation_age (integer)

当logging_collector被启用时，这个参数决定单个日志文件的最长使用时间。经过指定的分钟数之后，将创建一个新的日志文件。将这个参数设置为零将禁用基于时间的新日志文件创建。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

log_rotation_size (integer)

当logging_collector被启用时，这个参数决定一个个体日志文件的最大尺寸。当指定千字节数的数据被写入一个日志文件后，将创建一个新的日志文件。设置为零时将禁用基于大小创建新的日志文件。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

[1] http://pubs.opengroup.org/onlinepubs/009695399/functions/strftime.html

log_truncate_on_rotation (boolean)

当logging_collector被启用时，这个参数将导致PostgreSQL截断（覆盖而不是追加）任何已有的同名日志文件。不过，截断只在一个新文件由于基于时间的轮转被打开时发生，在服务器启动或基于尺寸的轮转时不会发生。如果被关闭，
在所有情况下以前存在的文件将被追加。例如，使用这个设置和一个类似postgresql-%H.log的log_filename将导致产生 24 个每小时的日志文件，并且循环地覆盖它们。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

示例：保留7天的日志，每天一个日志文件，命名为server_log.Mon，server_log.Tue，等等，并自动用本周的日志覆盖上周的日志，将log_filename设置为server_log.%a，将log_truncate_on_rotation设置为on，将log_rotation_age设置为1440。

示例：要保留 24 小时的日志，每个小时一个日志文件，如果日志文件尺寸超过 1GB，也会提前轮转。可以这样做：将log_filename设置为server_log.%H%M、将log_truncate_on_rotation设置为on、将log_rotation_age设置为60并且将log_rotation_size设置为1000000。在log_filename中包含%M，
可让按文件大小触发的轮转选用与整点初始文件名不同的新文件名。

syslog_facility (enum)

当启用了向syslog记录时，这个参数决定要使用的syslog“设施”。你可以在LOCAL0、LOCAL1、LOCAL2、LOCAL3、LOCAL4、LOCAL5、LOCAL6、LOCAL7中选择，默认值是LOCAL0。还请参阅系统的syslog守护进程的文档。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

syslog_ident (string)

当启用了向syslog记录时，这个参数决定用来标识syslog中的PostgreSQL消息的程序名。默认值是postgres。这个参数只能在postgresql.conf
文件中或在服务器命令行上设置。

silent_mode (boolean)

静默地运行服务器。如果设置了这个参数，服务器将自动在后台运行并脱离控制终端。这个参数只能在服务器启动时设置。

> **小心**
>
> 当设置了这个参数时，服务器的标准输出和标准错误会被重定向到数据目录中的postmaster.log文件。没有任何轮转这个文件的机制，
> 所以除非服务器日志输出被其他设置重定向到别处，它将无限增长。
> 建议在使用这个选项时将log_destination设置为syslog或者启用logging_collector。即使采取这些措施，启动早期报告的错误也可能出现在postmaster.log中而不是正常的日志目标中。

18.8.2. 什么时候记录日志

client_min_messages (enum)

控制哪些消息级别被发送给客户端。有效值是DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、LOG、NOTICE、WARNING、ERROR、FATAL和PANIC。

每个级别都包括其后的所有级别。级别越靠后，被发送的消息越少。默认值是NOTICE。
注意LOG在这里的排序与log_min_messages中的不同。

log_min_messages (enum)

控制哪些消息级别被写入服务器日志。有效值为DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、INFO、NOTICE、WARNING、ERROR、LOG、FATAL和PANIC。
每个级别包括其后的所有级别。级别越靠后，发送到日志的消息越少。默认值为WARNING。
注意LOG在这里的排序与client_min_messages中的不同。
只有超级用户能更改这个设置。

log_min_error_statement (enum)

控制在服务器日志中记录哪些导致错误条件的SQL语句。
对于达到指定严重级别或更高级别的消息，其日志条目中会包含当前 SQL 语句。有效值为DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、INFO、NOTICE、WARNING、ERROR、LOG、FATAL和PANIC。默认值为ERROR，这意味着导致错误、日志消息、
致命错误或紧急情况的语句将被记录。要有效地关闭记录失败的语句，将此参数设置为PANIC。只有超级用户能更改这个设置。

log_min_duration_statement (integer)

如果一条已完成语句的运行时间至少达到指定毫秒数，就记录其持续时间。
将此值设置为零会打印所有语句的持续时间。负一（默认值）禁用语句持续时间记录。
例如，如果设置为250ms，则会记录所有运行 250ms 或更长时间的 SQL 语句。启用此参数有助于发现应用中未优化的查询。只有超级用户能更改这个设置。

对于使用扩展查询协议的客户端，Parse、Bind 和 Execute 步骤的持续时间将被独立记录。

注意

当把这个选项和log_statement一起使用时，已经被log_statement
记录的语句文本不会在持续时间日志消息中重复。如果你没有使用syslog，
我们推荐你使用log_line_prefix记录 PID 或会话 ID，这样你可以使用进程 ID 或会话 ID 把语句消息链接到后来的持续时间消息。

表 18.1解释了PostgreSQL所使用的消息严重级别。如果日志输出被发送到syslog或 Windows 的eventlog，严重级别会按照表中所示进行转换。

表 18.1. 消息严重级别

严重性	用法	syslog	eventlog
DEBUG1..DEBUG5	为开发者提供逐级更加详细的信息。	DEBUG	INFORMATION
INFO	提供用户隐式要求的信息，例如来自VACUUM VERBOSE的输出。	INFO	INFORMATION
NOTICE	提供可能对用户有用的信息，例如长标识符截断提示。	NOTICE	INFORMATION
WARNING	提供可能出现的问题的警告，例如在一个事务块外COMMIT。	NOTICE	WARNING

严重性	用法	syslog	eventlog
ERROR	报告一个导致当前命令中断的错误。	WARNING	ERROR
LOG	报告管理员可能感兴趣的信息，例如检查点活动。	INFO	INFORMATION
FATAL	报告一个导致当前会话中断的错误。	ERR	ERROR
PANIC	报告一个导致所有数据库会话中断的错误。	CRIT	ERROR

18.8.3. 记录哪些内容

application_name (string)

application_name可以是任意小于NAMEDATALEN个字符（标准编译中是 64 个字符）的字符串。应用通常在连接服务器时设置此值。该名称将被显示在pg_stat_activity视图中并被包括在 CSV 日志项中。也可以通过log_line_prefix将其包括在普通日志项中。只有可打印 ASCII 字符能被使用在application_name之中。其他字符将被替换为问号（?）。

debug_print_parse (boolean)
debug_print_rewritten (boolean)
debug_print_plan (boolean)

这些参数将会让多种调试输出被发出。当被设置时，它们为每一个被执行的查询打印结果分析树、查询重写器输出或执行计划。这些消息在LOG消息级别上被发出，因此默认情况下它们将出现在服务器日志中但不会被发送到客户端。你可以通过调整client_min_messages和/或log_min_messages来改变这种情况。这些参数默认是关闭的。

debug_pretty_print (boolean)

当被设置时，debug_pretty_print会缩进由debug_print_parse、debug_print_rewritten或 debug_print_plan产生的输出。这将导致比关闭参数时使用的“紧凑”模式可读性更强但是更长的输出。它默认是打开的。

log_checkpoints (boolean)

导致检查点和重启点在服务器日志中记录。日志消息中包括一些统计信息，包括写入的缓冲区数量和写入它们所花费的时间。此参数只能在 postgresql.conf文件或服务器命令行中设置。默认值为关闭。

log_connections (boolean)

记录每一次到服务器的连接尝试，以及客户端认证的成功完成。这个参数在会话开始后不能更改。默认为关闭。

> **注意**
>
> 某些客户端程序（例如psql）在判断是否需要密码时会尝试连接两次，因此重复的“收到连接”消息并不一定表示一个错误。

log_disconnections (boolean)

在会话终止时向服务器日志输出一行与log_connections类似的内容，并且包含会话的持续时间。默认情况下这是关闭的。这个参数在会话开始后不能更改。

log_duration (boolean)

记录每个已完成语句的持续时间。默认值为off。只有超级用户能更改这个设置。

对于使用扩展查询协议的客户端，Parse、Bind 和 Execute 步骤的持续时间将被独立记录。

> **注意**
>
> 启用这个选项和设置log_min_duration_statement为零之间的区别是，超过log_min_duration_statement指定的时长会强制记录查询文本，而这个选项不会。因此，如果log_duration为on并且log_min_duration_statement为正值，所有持续时间都将被记录，但是只有超过阈值的语句才会被记录查询文本。这种行为有助于在高负载安装中收集统计信息。

log_error_verbosity (enum)

控制在服务器日志中记录的每条消息的详细程度。有效值为TERSE，DEFAULT和VERBOSE，它们依次在显示的消息中增加更多字段。TERSE不包括DETAIL，HINT，QUERY和CONTEXT错误信息的记录。VERBOSE输出包括SQLSTATE错误代码（另请参见附录 A）以及生成错误的源代码文件名、函数名和行号。只有超级用户能更改这个设置。

log_hostname (boolean)

默认情况下，连接日志消息只显示连接主机的 IP 地址。打开这个参数将导致也记录主机名。注意根据你的主机名解析设置，这可能会导致不可忽视的性能开销。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

log_line_prefix (string)

这是一个printf风格的字符串，会输出在每个日志行的开头。%字符用于引入“转义序列”，它们会被替换为下表所述的状态信息。无法识别的转义会被忽略。其他字符会直接复制到日志行中。有些转义只被会话进程识别，后台进程（例如主服务器进程）会忽略它们。此参数只能在postgresql.conf文件中或在服务器命令行上设置。默认值是一个空字符串。

转义	效果	只限会话
%a	应用名	是
%u	用户名	是
%d	数据库名	是
%r	远程主机名或 IP 地址，以及远程端口	是
%h	远程主机名或 IP 地址	是
%p	进程 ID	否
%t	无毫秒的时间戳	否
%m	带毫秒的时间戳	否
%i	命令标签：会话当前命令的类型	是

转义	效果	只限会话
%e	SQLSTATE 错误代码	否
%c	会话 ID：见下文	否
%l	对每个会话或进程的日志行号，从 1 开始	否
%s	进程开始的时间戳	否
%v	虚拟事务 ID (backendID/local XID)	否
%x	事务 ID（如果未分配则为 0）	否
%q	不产生输出，但是告诉非会话进程在字符串的这一点停止；会话进程忽略	否
%%	字面字符 %	否

该%c转义会打印一个近乎唯一的会话标识符，由两个以点分隔的 4 字节十六进制数（不含前导零）组成。这两个数分别是进程启动时间和进程 ID，因此%c也可以用作节省空间的方式来打印这些信息。例如，要从pg_stat_activity生成会话标识符，可以使用以下查询：

```
SELECT to_hex(EXTRACT(EPOCH FROM backend_start)::integer) || '.' ||
       to_hex(procpid)
FROM pg_stat_activity;
```

提示

如果你为log_line_prefix设置了非空值，
你通常应该让它的最后一个字符为空格，
这样用以提供和日志行的剩余部分的视觉区别。也可以使用标点符号。

提示

Syslog产生自己的时间戳和进程 ID 信息，因此如果你记录到syslog你可能不希望包括那些转义。

log_lock_waits (boolean)

控制会话为获取锁而等待的时间超过deadlock_timeout时是否生成日志消息。
这对于确定锁等待是否导致性能不佳很有用。默认值为off。

log_statement (enum)

控制哪些 SQL 语句被记录。有效值是 none (off)、ddl、mod和 all（所有语句）。ddl记录所有数据定义语句，例如CREATE、ALTER和 DROP语句。mod记录所有ddl语句，外加INSERT、UPDATE、DELETE、TRUNCATE和COPY FROM等数据修改语句。如果PREPARE、EXECUTE和 EXPLAIN ANALYZE包含合适类型的命令，它们也会被记录。对于使用扩展查询协议的客户端，当收到一个 Execute 消息时会产生日志并且会包括 Bind 参数的值（任何内嵌的单引号会被双写）。

默认值为none。只有超级用户能更改这个设置。

> **注意**
>
> 即使使用log_statement = all设置，包含简单语法错误的语句也不会被记录。这是因为只有在完成基本语法解析并确定了语句类型之后才会发出日志消息。
> 在扩展查询协议的情况下，在 Execute 阶段之前（即在解析分析或规划期间）出错的语句也不会被记录。将log_min_error_statement设置为ERROR（或更低）来记录这种语句。

log_temp_files (integer)

控制临时文件名和大小的日志记录。临时文件可以用于排序、hash 和临时查询结果。每当删除临时文件时都会发出日志记录。值为零时记录所有临时文件信息，而正值仅记录大小大于或等于指定千字节数的文件。默认设置为-1，禁用此类日志记录。只有超级用户能更改这个设置。

log_timezone (string)

设置在服务器日志中写入的时间戳的时区。和timezone不同，这个值是集簇范围的，因此所有会话将报告一致的时间戳。如果没有显式设置，服务器会将这个变量初始化为其系统环境指定的时区。详见第 8.5.3 节。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

18.8.4. 使用 CSV 格式的日志输出

将csvlog加入log_destination列表中，可以方便地将日志文件导入数据库表。此选项以逗号分隔值（CSV）格式输出日志行，包含以下列：带毫秒的时间戳、用户名、数据库名、进程 ID、客户端主机:端口号、会话 ID、会话内行号、命令标签、会话开始时间、虚拟事务 ID、常规事务 ID、错误严重性、SQLSTATE 代码、错误消息、错误消息详情、提示、引发错误的内部查询（如果有）、该内部查询中错误位置的字符数、错误上下文、引发错误的用户查询（如果有且由log_min_error_statement启用）、该用户查询中错误位置的字符数、错误在 PostgreSQL 源代码中的位置（如果log_error_verbosity设置为verbose）和应用名称。以下是用于存储 CSV 格式日志输出的示例表定义：

```
CREATE TABLE postgres_log
(
  log_time timestamp(3) with time zone,
  user_name text,
  database_name text,
  process_id integer,
  connection_from text,
  session_id text,
  session_line_num bigint,
  command_tag text,
  session_start_time timestamp with time zone,
  virtual_transaction_id text,
  transaction_id bigint,
  error_severity text,
  sql_state_code text,
  message text,
  detail text,
  hint text,
  internal_query text,
```

```
  internal_query_pos integer,
  context text,
  query text,
  query_pos integer,
  location text,
  application_name text,
  PRIMARY KEY (session_id, session_line_num)
);
```

使用COPY FROM命令将一个日志文件导入到这个表中：

```
COPY postgres_log FROM '/full/path/to/logfile.csv' WITH csv;
```

你可以做一些事情来简化导入 CSV 日志文件：

1. 设置log_filename和log_rotation_age，为日志文件提供一致且可预测的命名方案。这样就能预测文件名，并知道单个日志文件何时已完成写入、可以导入。

2. 将log_rotation_size设置为 0 来禁用基于尺寸的日志轮转，因为它使得日志文件名难以预测。

3. 将log_truncate_on_rotation设置为on，这样在同一个文件中旧日志数据不会与新数据混杂。

4. 上述表定义包括一个主键声明。这有助于避免意外地两次导入相同的信息。COPY命令一次提交所有它导入的数据，因此任何错误将导致整个导入失败。如果你导入一个部分完成的日志文件并且稍后当它完全完成后再次导入，主键冲突将导致导入失败。请等到日志完成且被关闭之后再导入。这个过程也可以避免意外地导入部分完成的行，这种行也将导致COPY失败。

18.9. 运行时统计数据

18.9.1. 查询和索引统计信息收集器

这些参数控制服务器范围内的统计信息收集功能。启用统计信息收集后，产生的数据可以通过pg_stat和 pg_statio系列系统视图进行访问。更多信息请参阅第 27 章。

track_activities (boolean)

启用对每个会话当前执行命令的信息收集，包括命令开始执行的时间。此参数默认为开启状态。请注意，即使启用了此参数，该信息也只有超级用户和拥有被报告会话的用户才能看到，因此不应构成安全风险。只有超级用户能更改这个设置。

track_activity_query_size (integer)

为每个活动会话指定存储当前执行命令的文本所预留的字节数，它们被用于pg_stat_activity.current_query字段。默认值是 1024字节。这个参数只能在服务器启动时被设置。

track_counts (boolean)

启用对数据库活动的统计信息收集。此参数默认为开启，因为自动清理守护进程需要这些收集到的信息。只有超级用户能更改这个设置。

track_functions (enum)

启用函数调用次数和耗时的跟踪。指定pl以仅跟踪过程语言函数，all
以同时跟踪SQL和C语言函数。默认值为none，即禁用函数统计跟踪。
只有超级用户能更改这个设置。

> **注意**
>
> 简单到足以被“内联”到调用查询中的 SQL 语言函数不会被跟踪，而不管这个设置。

update_process_title (boolean)

启用后，每次服务器接收到新的 SQL 命令时都会更新进程标题。进程标题通常通过ps命令查看，在 Windows 上则可以使用Process Explorer。只有超级用户能更改这个设置。

stats_temp_directory (string)

设置存放临时统计信息数据的目录。它既可以是相对于数据目录的路径，
也可以是绝对路径。默认值是 pg_stat_tmp。将其指向基于 RAM 的文件系统可以减少物理 I/O 需求，并可能提升性能。这个参数只能在postgresql.conf文件中或服务器命令行上设置。

18.9.2. 统计监控

log_statement_stats (boolean)
log_parser_stats (boolean)
log_planner_stats (boolean)
log_executor_stats (boolean)

对于每个查询，将各自模块的性能统计输出到服务器日志中。
这是一个简单的性能分析工具，类似于Unix getrusage()操作系统功能。log_statement_stats报告整个语句的统计信息，而其他选项报告每个模块的统计信息。log_statement_stats不能与任何单独模块选项一起启用。
所有这些选项默认情况下都是禁用的。只有超级用户才能更改这些设置。

18.10. 自动清理

这些设置控制autovacuum特性的行为。更多信息见第 23.1.5 节。

autovacuum (boolean)

控制服务器是否运行自动清理启动器后台进程。默认为开启，不过要使自动清理正常工作，
还需要启用track_counts。该参数只能在postgresql.conf文件或服务器命令行中设置，
不过，通过更改表存储参数可以为表禁用自动清理。

注意即使该参数被禁用，系统也会在需要防止事务ID回卷时发起自动清理进程。
详情请见第 23.1.4 节。

log_autovacuum_min_duration (integer)

当自动清理执行的操作运行时间至少达到指定毫秒数时，就会记录该操作。
将此设置为零会记录所有自动清理操作。负一（默认值）会禁用记录自动清理操作。例如，

如果将其设置为250ms，则所有运行时间为250ms或更长的自动清理和分析都将被记录。此外，当此参数设置为任何非-1值时，如果由于冲突锁而跳过自动清理操作，则会记录消息。启用此参数可帮助跟踪自动清理活动。此参数只能在postgresql.conf文件或服务器命令行中设置；但可以通过更改表存储参数来覆盖对单个表的设置。

autovacuum_max_workers(integer)

指定任一时刻可能运行的自动清理进程（自动清理启动器除外）的最大数量。默认值为三个。这个参数只能在服务器启动时设置。

autovacuum_naptime(integer)

指定自动清理在任意给定数据库上各次运行之间的最小间隔。在每一轮中后台进程检查数据库并根据需要为数据库中的表发出VACUUM和ANALYZE命令。延迟以秒为单位。默认值为1分钟（1min）。该参数只能在postgresql.conf文件或在服务器命令行上设置。

autovacuum_vacuum_threshold(integer)

指定能在一个表上触发VACUUM的被更新或被删除元组的最小数量。默认值为50个元组。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

autovacuum_analyze_threshold(integer)

指定能在一个表上触发ANALYZE的被插入、被更新或被删除元组的最小数量。默认值为50个元组。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

autovacuum_vacuum_scale_factor(floating point)

指定一个表尺寸的分数，在决定是否触发VACUUM时将它加到autovacuum_vacuum_threshold上。默认值为0.2（表尺寸的20%）。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

autovacuum_analyze_scale_factor(floating point)

指定一个表尺寸的分数，在决定是否触发ANALYZE时将它加到autovacuum_analyze_threshold上。默认值为0.1（表尺寸的10%）。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

autovacuum_freeze_max_age(integer)

指定在一个VACUUM操作被强制执行来防止表中事务ID回卷之前，一个表的pg_class.relfrozenxid域能保持的最大年龄（事务的）。注意即便自动清理被禁用，系统也将发起自动清理进程来阻止回卷。

清理也允许从pg_clog子目录中移除旧文件，这也是为什么默认值被设置为较低的2亿事务。该参数只能在服务器启动时设置，但是对于个别表可以通过修改表存储参数来降低该设置。详见第 23.1.4 节。

autovacuum_vacuum_cost_delay(integer)

指定用于自动VACUUM操作中的代价延迟值。如果指定-1，则使用vacuum_cost_delay值。默认值为20毫秒。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

autovacuum_vacuum_cost_limit(integer)

指定用于自动VACUUM操作中的代价限制值。如果指定-1（默认值），则使用vacuum_cost_limit值。注意该值被按比例地分配到运行中的自动清理工作进程上（如果有多个），因此每一个工作进程的限制值之和不会超过这个变量中的值。该参数只能在postgresql.conf文件或在服务器命令行上设置。但是对个别表可以通过修改表存储参数来覆盖该设置。

18.11. 客户端连接默认值

18.11.1. 语句行为

search_path(string)

这个变量指定当一个对象（表、数据类型、函数等）被用一个无模式限定的简单名称引用时，搜索该对象时的模式顺序。当在不同模式中有同名对象时，将使用第一个在搜索路径中被找到的对象。一个不属于搜索路径中任何一个模式的对象只能通过用限定名（带点号）指定包含它的模式来引用。

search_path的值必须是一个逗号分隔的模式名列表。如果列表项之一是特殊值$user，则具有SESSION_USER返回的名字的模式将取代它（如果有这样一个模式；如果没有，$user会被忽略）。

系统目录模式pg_catalog总是被搜索，不管它是否在搜索路径中被提及。如果它在路径中被提及，那么它将被按照路径指定的顺序搜索。如果pg_catalog不在路径中，则它将在任何路径项之前被搜索。

同样，如果存在，当前会话的临时表模式pg_temp_*nnn*总是被搜索。可以通过使用别名pg_temp在路径中显式列出它。如果它没有列在路径中，则首先搜索它（甚至在pg_catalog之前）。然而，临时模式仅用于搜索关系（表、视图、序列等）和数据类型名称。它永远不会用于搜索函数或操作符名称。

当对象创建时没有指定一个特定目标模式，它们将被放置在搜索路径中列出的第一个模式中。如果搜索路径为空将报告一个错误。

这个参数的默认值是'"$user", public'（其中第二部分在没有名为public的模式时会被忽略）。这支持共享使用一个数据库（其中没有用户有私有模式，所有用户共用public）、每用户私有模式以及它们的组合。通过全局或者逐用户地改变默认搜索路径设置，可以得到其他效果。

搜索路径的当前有效值可以通过SQL函数current_schemas检查（见第 9.23 节）。它和检查search_path的值不太一样，因为current_schemas显示出现在search_path中的项如何被解析。

更多有关模式处理的信息，请参考第 5.7 节。

default_tablespace(string)

这个变量指定当一个CREATE命令没有显式指定一个表空间时，创建对象（表和索引）的默认表空间。

该值要么是一个表空间的名字，要么是一个指定使用当前数据库默认表空间的空字符串。如果该值和任何现有表空间的名字都不匹配，

PostgreSQL将自动使用当前数据库的默认表空间。如果指定了一个非默认的表空间，用户必须对它有CREATE权限，否则创建尝试将失败。

这个变量不被用于临时表，对临时表会使用temp_tablespaces。

创建数据库时也不会使用这个变量。默认情况下，一个新数据库会从它的模板数据库继承其表空间设置。

有关表空间的更多的信息，请见第 21.6 节。

temp_tablespaces (string)

这个变量指定当一个CREATE命令没有显式指定一个表空间时，创建临时对象（临时表和临时表上的索引）的默认表空间。用于排序大型数据集的临时文件也被创建在这些表空间中。

该值是一个表空间名字的列表。当列表中有多于一个名称时，每次一个临时对象被创建时PostgreSQL随机选择列表中的一个成员。例外是在一个事务中，连续创建的临时对象被依次放置在列表中的连续表空间中。如果列表中被选中的元素是一个空字符串，PostgreSQL将自动使用当前数据库的默认表空间。

当temp_tablespaces被交互式地设置时，指定一个不存在的表空间是一种错误，指定一个用户不具有CREATE权限的表空间也同样是错误。不过，当使用一个之前设置的值时，不存在的表空间会被忽略，就像用户缺少CREATE权限的表空间一样。特别是，使用一个在postgresql.conf中设置的值时，这条规则起效。

默认值是一个空字符串，它使得所有临时对象被创建在当前数据库的默认表空间中。

参阅default_tablespace。

check_function_bodies (boolean)

这个参数通常为打开。当设置为off时，它禁用CREATE FUNCTION期间对函数体字符串的验证。禁用验证避免了验证处理的副作用，也避免前向引用等问题导致的误报。在代表其他用户载入函数之前设置这个参数为off；pg_dump会自动这样做。

default_transaction_isolation (enum)

每个 SQL 事务都有一个隔离级别，可以是“读未提交”、“读已提交”、“可重复读”或者“可串行化”。这个参数控制每个新事务的默认隔离级别。默认是“读已提交”。

更多信息请参阅第 13 章和SET TRANSACTION。

default_transaction_read_only (boolean)

一个只读的 SQL 事务不能修改非临时表。这个参数控制每个新事务的默认只读状态。默认是off（读/写）。

更多信息请参考SET TRANSACTION。

default_transaction_deferrable (boolean)

当运行在serializable隔离级别时，一个可延迟只读 SQL 事务可能在获准继续之前被延迟一段时间。但是，一旦它开始执行就不会产生任何用来保证可串行性的开销；因此串行化代码将没有任何理由因为并发更新而强制它中止，使得这个选项适合于长时间运行的只读事务。

这个参数控制每个新事务的默认可延迟状态。
目前它对读写事务或者那些运行在低于serializable隔离级别上的事务无效。
默认值是off。

更多信息请参考SET TRANSACTION。

session_replication_role (enum)

控制当前会话中与复制相关的触发器和规则的触发。设置此变量需要超级用户权限，
并会丢弃任何先前缓存的查询计划。可用值为origin（默认值）、replica和local。
更多信息参见ALTER TABLE。

statement_timeout (integer)

中止任何使用了超过指定毫秒数的语句，从客户端命令到达服务器时开始计时。如果log_min_error_statement被设置为ERROR或更低，语句如果超时也会被记录。
一个零值（默认）将禁用超时。

我们不推荐在postgresql.conf中设置statement_timeout，因为它会影响所有会话。

vacuum_freeze_table_age (integer)

如果表的pg_class.relfrozenxid字段达到此设置指定的年龄，VACUUM
就会执行一次全表扫描。默认值为 1.5 亿个事务。尽管用户可以将该值设置在 0 到 10 亿之间，VACUUM仍会悄悄将其有效值限制为不超过autovacuum_freeze_max_age的 95%，以便在针对该表启动防回卷自动清理之前，周期性手工VACUUM仍有机会运行。详见第 23.1.4 节。

vacuum_freeze_min_age (integer)

指定VACUUM在扫描表时用来决定是否用FrozenXID替换事务 ID 的截止年龄（以事务计）。默认值是 5000 万个事务。尽管用户可以将该值设置为 0 到 10 亿之间的任意值，VACUUM会悄悄将有效值限制为不超过autovacuum_freeze_max_age的一半，这样强制 autovacuum 之间就不会间隔过短。更多信息请参见第 23.1.4 节。

bytea_output (enum)

设置bytea类型值的输出格式。有效值是hex（默认）和 escape（传统的 PostgreSQL 格式）。详见第 8.4 节。不管这个设置的值如何，bytea类型总是接受这两种格式的输入。

xmlbinary (enum)

设置二进制值如何被编码为 XML。例如，这适用于通过xmlelement函数或xmlforest函数将bytea值转换到 XML 值。可能的值有base64和hex，它们都是用 XML 模式标准定义的。
默认值是base64。更多关于 XML 相关函数的信息可参阅第 9.14 节。

这里的实际选择主要取决于偏好，只受客户端应用中可能存在的限制的约束。
两种方法都支持所有可能的值，尽管十六进制编码会比 base64 编码略大。

xmloption (enum)

设置在 XML 与字符串值之间进行转换时，隐含采用DOCUMENT还是CONTENT。有关说明参见第 8.13 节。有效值是DOCUMENT和CONTENT。默认值是CONTENT。

根据 SQL 标准，设置这个选项的命令是：

```
SET XML OPTION { DOCUMENT | CONTENT };
```

这种语法在 PostgreSQL 也可用。

18.11.2. 区域设置和格式化

`DateStyle` (`string`)

设置日期和时间值的显示格式，以及解释有歧义的日期输入值的规则。由于历史原因，这个变量包含两个独立的部分：输出格式声明（`ISO`、`Postgres`、`SQL`或`German`）、输入/输出的年/月/日顺序（`DMY`、`MDY`或`YMD`）。这些可以被独立设置或者一起设置。关键字`Euro`和`European`是`DMY`的同义词；关键字`US`、`NonEuro`和`NonEuropean`是`MDY`的同义词。详见第 8.5 节。内置默认值是`ISO`，`MDY`，但是initdb将用对应于选中的`lc_time`区域设置行为的设置初始化配置文件。

`IntervalStyle` (`enum`)

设置时间间隔值的显示格式。值`sql_standard`会生成符合SQL标准时间间隔字面量的输出。值`postgres`（默认值）的输出与PostgreSQL 8.4 之前版本中DateStyle设为`ISO`时的输出一致。值`postgres_verbose`的输出与PostgreSQL 8.4 之前版本中`DateStyle`设为非`ISO`输出时的输出一致。值`iso_8601`会生成符合 ISO 8601 第 4.4.3.2 节定义的时间间隔“带标志符格式”的输出。

`IntervalStyle`参数也会影响对有歧义的时间间隔输入的解释。详见第 8.5.4 节。

`timezone` (`string`)

设置用于显示和解释时间戳的时区。如果没有显式设置，服务器会将这个变量初始化为其系统环境指定的时区。详见第 8.5.3 节。

`timezone_abbreviations` (`string`)

设置服务器接受的日期时间输入中使用的时区缩写集合。默认值为`'Default'`，这个集合在全世界大多数地方都能工作。也还有`'Australia'`和`'India'`，以及可能为一种特定安装定义的其他集合。详见第 B.3 节。

`extra_float_digits` (`integer`)

此参数调整浮点值显示的位数，包括`float4`、`float8`和几何数据类型。参数值会加到标准位数（按情况使用`FLT_DIG`或`DBL_DIG`）上。此值最高可以设置为 3，以包含部分有效的数位；这对于转储需要精确还原的浮点数据特别有用。也可以设置为负数，以抑制不需要的数位。另见第 8.1.3 节。

`client_encoding` (`string`)

设置客户端编码（字符集）。默认使用数据库编码。PostgreSQL服务器所支持的字符集在第 22.3.1 节中描述。

`lc_messages` (`string`)

设置消息显示的语言。可接受的值是系统相关的；详见第 22.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

在一些系统上，这个区域设置类别并不存在。仍然可以设置这个变量，只是不会有任何效果。同样，所期望语言的翻译消息也可能不存在。在这种情况下，你将仍然继续看到英文消息。

只有超级用户能更改这个设置，因为它会同时影响发送到服务器日志和客户端的消息，设置不当可能降低服务器日志的可读性。

`lc_monetary`(`string`)

设置用于格式化货币量的区域设置，例如用`to_char`函数族。可接受的值是系统相关的；详见第 22.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

`lc_numeric`(`string`)

设置用于格式化数字的区域设置，例如用`to_char`函数族。可接受的值是系统相关的；详见第 22.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

`lc_time`(`string`)

设置用于格式化日期和时间的区域设置，例如用`to_char`函数族。可接受的值是系统相关的；详见第 22.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

`default_text_search_config`(`string`)

选择被那些没有显式参数指定配置的文本检索函数变体使用的文本检索配置。详见第 12 章。内置默认值是`pg_catalog.simple`，但是如果能够标识一个匹配该区域设置的配置，initdb将用对应于选中的`lc_ctype`区域设置的值初始化配置文件。

18.11.3. 其他默认值

`dynamic_library_path`(`string`)

如果需要打开一个可以动态装载的模块并且在`CREATE FUNCTION`或`LOAD`命令中指定的文件名没有目录部分（即名字中不包含斜线），那么系统将搜索这个路径以查找所需的文件。

参数`dynamic_library_path`的值必须是由冒号（Windows上为分号）分隔的绝对目录路径列表。如果某个列表元素以特殊字符串`$libdir`开头，则会使用编译时确定的PostgreSQL软件包的库目录来替换`$libdir`；该目录是标准PostgreSQL发行版所提供模块的安装位置。（使用`pg_config --pkglibdir`可以找出此目录的名称。）例如：

```
dynamic_library_path = '/usr/local/lib/postgresql:/home/my_project/
lib:$libdir'
```

或者，在 Windows 环境中：

```
dynamic_library_path = 'C:\tools\postgresql;H:\my_project\lib;
$libdir'
```

这个参数的默认值是`'$libdir'`。如果该值被设置为一个空字符串，则关闭自动路径搜索。

这个参数可以由超级用户在运行时更改，但以这种方式进行的设置只会持续到客户端连接结束，因此这种方法应该保留用于开发目的。推荐设置此参数的方法是在`postgresql.conf`配置文件中。

`gin_fuzzy_search_limit`(`integer`)

GIN 索引扫描返回的集合尺寸的软上限。详见第 54.4 节。

`local_preload_libraries`(`string`)

这个变量指定一个或多个要在连接开始时预载入的共享库。如果要载入多个库，用逗号分隔它们的名字。除非用双引号括起，所有库名都会被转换为小写。这个参数在一个特定会话开始之后不能被更改。

因为这不是一个仅限超级用户的选项，能被这样载入的库被严格限制为出现于安装的标准库目录中`plugins`子目录下的共享库（保证只有“安全的”库被安装到这里是数据库管理员的责任）。`local_preload_libraries`中的项可以显式指定这个目录，例如`$libdir/plugins/mylib`，或者只是指定库的名称 – `mylib` 和`$libdir/plugins/mylib`的效果是相同的。

与shared_preload_libraries不同，在会话开始时载入一个库而不是在第一次使用它时载入，并没有性能上的优势。这个特性的意图是允许调试或性能测量库被载入到特定会话中，而无需发出显式的`LOAD`命令。例如，可以通过`PGOPTIONS`环境变量为某个会话设置这个参数来启用调试。

如果一个指定的库没有找到，连接尝试将会失败。

每一个 PostgreSQL 支持的库都有一个被检查以保证兼容性的“魔法块”。因此，非 PostgreSQL 的库不能用这种方式载入。

18.12. 锁管理

`deadlock_timeout`(`integer`)

指定在检查是否发生死锁之前等待锁的毫秒数。检查死锁相对昂贵，因此服务器不会每次等待锁时都运行它。我们乐观地假设在生产应用程序中死锁并不常见，所以在检查死锁之前只是等待一段时间。增加此值会减少在不必要的死锁检查中浪费的时间，但会减慢实际死锁错误的报告速度。默认值为一秒（`1s`），这可能是你在实践中想要的最小值。在负载较重的服务器上，你可能希望提高它。理想情况下，设置应超过你的典型事务时间，以提高在等待者决定检查死锁之前释放锁的几率。

当log_lock_waits被设置时，这个参数也决定发出关于锁等待的日志之前等待的时间量。如果你想调查锁延迟，你可能希望设置一个比正常的`deadlock_timeout`小的值。

`max_locks_per_transaction`(`integer`)

共享锁表跟踪在`max_locks_per_transaction` * (max_connections + max_prepared_transactions) 个对象（如表）上的锁。因此，在任何一个时刻，只有不超过这么多个可区分对象能够被锁住。这个参数控制为每个事务分配的对象锁的平均数量。只要所有事务的锁都能放入锁表中，单个事务就可以锁住更多对象。这不是能被锁住的行数，那个值是没有限制的。默认值 64 已经被历史证明是足够的，但是如果你有在一个事务中访问很多不同表的客户端，你可能需要提高这个值。这个参数只能在服务器启动时设置。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

运行备库时，必须将此参数设置为大于或等于主库上的值。否则，备库上将不允许查询。

`max_pred_locks_per_transaction`(`integer`)

共享谓词锁表跟踪在`max_pred_locks_per_transaction` * (max_connections + max_prepared_transactions) 个对象（如表）上的锁。因此，在任何一个时刻，只有不超过这么

多个可区分对象能够被锁住。这个参数控制为每个事务分配的对象锁的平均数量。只要所有事务的锁都能放入锁表中，单个事务就可以锁住更多对象。这不是能被锁住的行数，那个值是没有限制的。默认值 64 已经在测试中被证明通常是足够的，但是如果你有会在单个可串行化事务中访问许多不同表的客户端，则可能需要提高这个值。这个参数只能在服务器启动时设置。

增大这个参数可能会导致PostgreSQL请求比操作系统默认配置所允许的更多的System V共享内存。如有必要，关于如何调整这些参数的信息请参见第 17.4.1 节。

18.13. 版本和平台兼容性

18.13.1. 以前的 PostgreSQL 版本

`array_nulls` (`boolean`)

这个参数控制数组输入解析器是否把未用引号的`NULL`识别为一个值为空值的数组元素。默认为`on`，允许输入包含空值的数组值。但是PostgreSQL 8.2 之前的版本不支持数组中的空值，并且因此将把`NULL`当作指定一个值为字符串 “NULL” 的正常数组元素。为了向后兼容那些要求旧行为的应用，这个变量可以被设置为`off`。

注意即使这个变量为`off`也能够创建包含空值的数组值。

`backslash_quote` (`enum`)

这个参数控制字符串字面量中的单引号是否能够用`\'`来表示。首选的 SQL 标准的方法是将其双写（`''`），但是PostgreSQL在历史上也接受`\'`。不过使用`\'`容易导致安全风险，因为在某些客户端字符集编码中，有多字节字符的最后一个字节在数值上等价于 ASCII 的`\`。如果客户端代码没有做到正确转义，那么就可能遭到 SQL 注入攻击。如果服务器拒绝看起来带有被反斜线转义的单引号的查询，那么就可以避免这种风险。`backslash_quote`的可用值是`on`（总是允许`\'`）、`off`（总是拒绝）以及`safe_encoding`（只有客户端编码不允许在多字节字符中存在 ASCII `\`时允许）。`safe_encoding`是默认设置。

注意在符合标准的字符串字面量中，`\`就表示`\`。这个参数只影响不符合标准的字面量的处理，包括转义字符串语法（`E'...'`）。

`default_with_oids` (`boolean`)

此参数控制在既未指定`WITH OIDS`也未指定`WITHOUT OIDS`时，`CREATE TABLE`和`CREATE TABLE AS`是否在新建表中包含 OID 列。它也决定`SELECT INTO`创建的表是否包含 OID。此参数默认为`off`；在PostgreSQL 8.0 及更早版本中，默认为打开。

在用户表中使用 OID 已被弃用，因此大多数安装应让此变量保持禁用状态。需要为特定表使用 OID 的应用应在创建表时指定`WITH OIDS`。可以启用此变量，以兼容不遵循这种做法的旧应用。

`escape_string_warning` (`boolean`)

打开时，如果在普通字符串字面量中（`'...'`语法）出现了一个反斜线（`\`）并且`standard_conforming_strings`为关闭，那么就会发出一个警告。默认值是`on`。

希望使用反斜线作为转义符的应用应该被修改来使用转义字符串语法（`E'...'`），因为按照 SQL 标准，普通字符串现在默认将反斜线视作一个普通字符。这个变量可以被启用来帮助定位需要被更改的代码。

`lo_compat_privileges` (`boolean`)

在PostgreSQL 9.0之前的版本中，大对象没有访问权限，实际上可以被所有用户读取和写入。将此变量设置为`on`会禁用新的权限检查，以保持与之前版本的兼容性。默认值为`off`。

设置此变量不会禁用所有与大对象相关的安全检查 – 只禁用那些在PostgreSQL 9.0 中默认行为发生变化的检查。例如，`lo_import()`和`lo_export()`需要超级用户权限，与此设置无关。

`quote_all_identifiers` (`boolean`)

当数据库产生 SQL 时，强制所有标识符被引号包围，即使它们（当前）不是关键字。这将影响`EXPLAIN`的输出以及`pg_get_viewdef`等函数的结果。另请参阅pg_dump和pg_dumpall的`--quote-all-identifiers`选项。

`sql_inheritance` (`boolean`)

这个设置控制不带修饰的表引用是否被认为包括继承子表。默认值为`on`，
即子表被包括在内（因此，默认假定有`*`后缀）。如果设为`off`，子表不被包括（因此，
假定有`ONLY`前缀）。SQL 标准要求包括子表，因此`off`设置不符合标准，但提供它是为了与7.1 之前版本的PostgreSQL兼容。更多信息见第 5.8 节。

将`sql_inheritance`关闭已被弃用，因为这种行为已被发现既容易出错又违背 SQL 标准。本手册其他地方对继承行为的讨论通常都假定它为`on`。

`standard_conforming_strings` (`boolean`)

控制普通字符串字面量（`'...'`）是否按照 SQL 标准把反斜线当普通文本。从PostgreSQL 9.1 开始，默认值为`on`（之前的发行中默认值为`off`）。应用可以检查这个参数来判断字符串字面量如何被处理。这个参数的存在也可以被当做转义字符串语法（`E'...'`）被支持的标志。如果一个应用希望反斜线被当做转义字符，应该使用转义字符串语法（第 4.1.2.2 节）。

`synchronize_seqscans` (`boolean`)

它允许对大型表的顺序扫描与其他扫描同步，
因此并发扫描可以在几乎相同的时刻读取相同的块，这样可以分担 I/O 负载。当启用这个参数时，一个扫描可能会从表的中间开始并且之后“绕回”到开头以覆盖所有的行，
这样可以与已在进行中的扫描活动同步。对于没有`ORDER BY`子句的查询，这样的扫描可能会在返回行的顺序中造成不可预料的改变。将这个参数设置为`off`以保证 8.3 之前的行为（顺序扫描总是从表的起始处开始）。默认值是`on`。

18.13.2. 平台和客户端兼容性

`transform_null_equals` (`boolean`)

当打开时，形如*`expr`* `= NULL`（或`NULL =` *`expr`*）的表达式将被当做*`expr`* `IS NULL`，也就是说，如果*`expr`*计算结果为空值则返回真，否则返回假。正确的 SQL 标准兼容的*`expr`* `= NULL`行为总是返回空值（未知）。因此这个参数默认为`off`。

不过，在Microsoft Access里的过滤表单生成的查询似乎使用*`expr`* `= NULL`来测试空值，因此，如果你使用这个接口访问数据库，你可能想把这个选项打开。因为*`expr`* `= NULL`形式的表达式总是返回空值（使用 SQL 标准解释），它们不是非常有用并且在普通应用中也不常见，因此这个选项实际上没有什么危害。
但是新用户常常对涉及空值的表达式语义感到困惑，因此这个选项默认为关闭。

请注意这个选项只影响= NULL形式，而不影响其它比较操作符或者其它与一些涉及等值操作符的表达式在计算上等效的其他表达式（例如IN）。因此，
这个选项不能普遍修复错误的程序写法。

相关信息请见第 9.2 节。

18.14. 错误处理

exit_on_error (boolean)

如果为true，任何错误将中止当前会话。默认情况下，这个值被设置为false，这样只有FATAL 错误（致命）将中止会话。

restart_after_crash (boolean)

当被设置为true（默认值），PostgreSQL将在一次后端崩溃后自动重新初始化。
让这个值设置为true通常是将数据库可用性最大化的最佳方法。但是在某些环境中，例如PostgreSQL被集群软件调用时，禁用重启可能很有用，
这样集群软件可以得到控制并且采取它认为适当的行动。

18.15. 预置选项

下列“参数”是只读的，在编译或安装 PostgreSQL 时确定。因此，它们未列入示例 postgresql.conf 文件。这些选项报告 PostgreSQL 行为的各个方面，某些应用（特别是管理前端）可能对此感兴趣。

block_size (integer)

报告一个磁盘块的大小。它由编译服务器时BLCKSZ的值确定。默认值是 8192 字节。有些配置变量的含义（例如shared_buffers）会被block_size影响。详见第 18.4 节。

integer_datetimes (boolean)

报告PostgreSQL是否在编译时启用了对 64 位整数日期和时间的支持。构建PostgreSQL时配置--disable-integer-datetimes可以禁用它。默认值是on。

lc_collate (string)

报告对文本数据进行排序所使用的区域设置。更多信息参见第 22.1 节。此值在创建数据库时确定。

lc_ctype (string)

报告决定字符分类的区域设置。更多信息参见第 22.1 节。此值在创建数据库时确定。通常它与lc_collate相同，但特殊应用可能将它设置为不同的值。

max_function_args (integer)

报告函数参数的最大数量。它由编译服务器时的FUNC_MAX_ARGS值决定。默认值是 100 个参数。

max_identifier_length (integer)

报告标识符的最大长度。它由编译服务器时的NAMEDATALEN值减一决定。NAMEDATALEN的默认值是 64；因此max_identifier_length的默认值是 63 字节，在使用多字节编码时，这可能不足 63 个字符。

max_index_keys (integer)

报告索引键的最大数目。它由编译服务器时的INDEX_MAX_KEYS值决定。默认值是 32 个键。

segment_size (integer)

报告一个文件段中可以存储的块（页）的数量。由编译服务器时的RELSEG_SIZE值决定。一个段文件的最大尺寸（以字节计）等于segment_size乘以block_size，默认是 1GB。

server_encoding (string)

报告数据库的编码（字符集）。这是在数据库被创建时决定的。通常，客户端只需要关心client_encoding的值。

server_version (string)

报告服务器的版本号。它是由编译服务器时的PG_VERSION值决定的。

server_version_num (integer)

以整数形式报告服务器的版本号。它是由编译服务器时的PG_VERSION_NUM值决定的。

wal_block_size (integer)

报告一个 WAL 磁盘块的尺寸。由编译服务器时的XLOG_BLCKSZ值决定。默认是 8192 字节。

wal_segment_size (integer)

报告 WAL 段文件中的块（页）数。WAL 段文件以字节计的总大小等于wal_segment_size乘以wal_block_size；默认总大小为 16MB。更多信息参见第 29.4 节。

18.16. 自定义选项

这个特性被设计用来允许PostgreSQL通常不知道的参数由附加模块（例如过程语言）添加。这允许附加模块以标准的方式配置。

custom_variable_classes (string)

这个变量指定一个或多个用于自定义变量的类名，形式为逗号分隔的列表。自定义变量是PostgreSQL本身通常不知道、但被某些附加模块使用的变量。这种变量的名字必须由类名、点和变量名组成。custom_variable_classes指定一个特定安装中使用的所有类名。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

在postgresql.conf中设置自定义变量的困难在于，该文件必须在附加模块载入之前被读取，因此自定义变量通常会被作为未知变量拒绝。当设置了custom_variable_classes时，服务器将接受每个指定类中任意变量的定义。这些变量会被当作占位符对待，并且在定义它们的模块被载入之前没有任何功能。当一个特定类的模块被载入时，它会为其类名添加正确的变量定义，按照那些定义转换所有占位符值，并对它所属类中剩余的任何无法识别的占位符发出警告。

下面的例子展示了使用自定义变量时postgresql.conf可能包含的内容：

```
custom_variable_classes = 'plpgsql,plperl'
plpgsql.variable_conflict = use_variable
plperl.use_strict = true
```

```
plruby.use_strict = true        # generates error: unknown class name
```

18.17. 开发者选项

下列参数用于PostgreSQL源代码的开发工作，在某些情况下也用于辅助恢复严重损坏的数据库。没有理由在生产数据库中使用它们。因此，它们被排除在示例postgresql.conf文件之外。请注意，许多参数需要特殊的源代码编译标志才能起作用。

allow_system_table_mods (boolean)

允许修改系统表的结构。此参数由initdb使用。此参数只能在服务器启动时设置。

debug_assertions (boolean)

打开各种断言检查。这是一种调试辅助手段。如果你遇到了奇怪的问题或崩溃，可以打开它，因为它可能暴露出编程错误。要使用这个参数，在构建PostgreSQL时必须定义宏USE_ASSERT_CHECKING（通过configure选项--enable-cassert完成）。注意，如果PostgreSQL构建时启用了断言，debug_assertions默认为on。

ignore_system_indexes (boolean)

读取系统表时忽略系统索引（但是修改系统表时依然同时更新索引）。这在从被破坏的系统索引中恢复数据时有用。这个参数在会话开始之后不能被更改。

post_auth_delay (integer)

如果非零，则在新服务器进程启动并执行认证过程之后，延迟指定的秒数。这旨在给开发者一个机会，用调试器附加到服务器进程上。此参数在会话开始之后不能更改。

pre_auth_delay (integer)

如果非零，则在新服务器进程刚刚派生之后、执行认证过程之前，延迟指定的秒数。这旨在给开发者一个机会，用调试器附加到服务器进程上，跟踪认证过程中的异常行为。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

trace_notify (boolean)

为LISTEN和NOTIFY命令生成大量调试输出。client_min_messages或log_min_messages必须是DEBUG1或者更低才能把这种输出分别发送到客户端或者服务器日志。

trace_recovery_messages (enum)

启用原本不会记录的恢复相关调试输出。此参数允许用户覆盖log_min_messages的常规设置，但仅针对特定消息。它旨在用于调试热备。有效值为DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1和LOG。默认值LOG完全不影响是否记录消息的决定。其他值会让该优先级或更高优先级的恢复相关调试消息按LOG优先级记录；对于log_min_messages的常见设置，这会无条件地将它们发送到服务器日志。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

trace_sort (boolean)

如果开启，输出排序操作中的资源使用信息。只有在编译PostgreSQL时定义了TRACE_SORT宏，这个参数才可用（不过，当前在默认情况下就定义了TRACE_SORT）。

`trace_locks` (`boolean`)

如果开启，发出锁使用情况的信息。被转储信息中包括锁操作的类型、
锁的类型和被锁或被解锁对象的唯一标识符。
同样包括的还有已经授予这个对象的锁类型的位掩码和等待这个对象的锁类型的位掩码。
还会转储每种锁类型已授予的锁数、等待的锁数，以及它们的总数。
一个日志文件输出的示例如下：

```
LOG:  LockAcquire: new: lock(0xb7acd844) id(24688,24696,0,0,0,1)
      grantMask(0) req(0,0,0,0,0,0,0)=0 grant(0,0,0,0,0,0,0)=0
      wait(0) type(AccessShareLock)
LOG:  GrantLock: lock(0xb7acd844) id(24688,24696,0,0,0,1)
      grantMask(2) req(1,0,0,0,0,0,0)=1 grant(1,0,0,0,0,0,0)=1
      wait(0) type(AccessShareLock)
LOG:  UnGrantLock: updated: lock(0xb7acd844) id(24688,24696,0,0,0,
1)
      grantMask(0) req(0,0,0,0,0,0,0)=0 grant(0,0,0,0,0,0,0)=0
      wait(0) type(AccessShareLock)
LOG:  CleanUpLock: deleting: lock(0xb7acd844) id(24688,24696,0,0,0,
1)
      grantMask(0) req(0,0,0,0,0,0,0)=0 grant(0,0,0,0,0,0,0)=0
      wait(0) type(INVALID)
```

被转储结构体的详细信息可以在`src/include/storage/lock.h`中找到。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`trace_lwlocks` (`boolean`)

如果开启，发出轻量级锁的使用信息。
轻量级锁主要是为了提供对共享内存数据结构的互斥访问。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`trace_userlocks` (`boolean`)

如果开启，发出关于用户锁使用的信息。与`trace_locks`的输出一样，但只用于咨询锁。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`trace_lock_oidmin` (`integer`)

如果设置，不会跟踪 OID 小于此值的表上的锁（用于避免在系统表上的输出）。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`trace_lock_table` (`integer`)

无条件地跟踪此表（OID）上的锁。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`debug_deadlocks` (`boolean`)

如果设置，当死锁超时发生时，转储所有当前锁的信息。

只有在编译PostgreSQL时定义了`LOCK_DEBUG`宏，这个参数才可用。

`log_btree_build_stats` (`boolean`)

如果设置，会记录 B-树操作上的系统资源使用情况统计（内存和 CPU）。

只有在编译PostgreSQL时定义了`BTREE_BUILD_STATS`宏，这个参数才可用。

`wal_debug` (`boolean`)

如果被打开，WAL 相关的调试输出将被发出。只有在编译PostgreSQL时定义了`WAL_DEBUG`宏的情况下，这个参数才可用。

`zero_damaged_pages` (`boolean`)

检测到损坏的页面头通常会导致PostgreSQL报告错误，中止当前事务。将`zero_damaged_pages`设置为on会导致系统报告警告，将内存中的损坏页面清零，并继续处理。这种行为会破坏数据，即损坏页面上的所有行。但是，它确实允许你跳过错误，
并从表中可能存在的未损坏页面中检索行。如果由于硬件或软件错误而发生损坏，
这对于恢复数据很有用。通常在放弃从表的损坏页面恢复数据的希望之前，
不应将其设置为on。清零的页面不会强制写入磁盘，
因此建议在再次关闭此参数之前重新创建表或索引。默认设置为`off`。
只有超级用户能更改这个设置。

18.18. 短选项

为了方便起见，系统中还为一些参数提供了单字母的命令行选项开关。它们在表 18.2中描述。其中一些选项是由于历史原因而存在，它们以单字母选项的形式存在，
并不一定表示鼓励频繁使用这些选项。

表 18.2. 短选项对照

Short Option	Equivalent
`-A` *x*	`debug_assertions =` *x*
`-B` *x*	`shared_buffers =` *x*
`-d` *x*	`log_min_messages = DEBUG`*x*
`-e`	`datestyle = euro`
`-fb`, `-fh`, `-fi`, `-fm`, `-fn`, `-fs`, `-ft`	`enable_bitmapscan = off`, `enable_hashjoin = off`, `enable_indexscan = off`, `enable_mergejoin = off`, `enable_nestloop = off`, `enable_seqscan = off`, `enable_tidscan = off`
`-F`	`fsync = off`
`-h` *x*	`listen_addresses =` *x*
`-i`	`listen_addresses = '*'`
`-k` *x*	`unix_socket_directory =` *x*
`-l`	`ssl = on`
`-N` *x*	`max_connections =` *x*
`-O`	`allow_system_table_mods = on`
`-p` *x*	`port =` *x*
`-P`	`ignore_system_indexes = on`
`-s`	`log_statement_stats = on`

Short Option	Equivalent
`-S x`	`work_mem = x`
`-tpa`, `-tpl`, `-te`	`log_parser_stats = on`, `log_planner_stats = on`, `log_executor_stats = on`
`-W x`	`post_auth_delay = x`

第 19 章 客户端认证

当客户端应用连接到数据库服务器时，它会指定要以哪个 PostgreSQL 数据库用户名连接，这很像以某个特定用户身份登录 Unix 计算机。在 SQL 环境中，当前活动的数据库用户名决定了对数据库对象的访问权限 – 详见第 20 章。因此，必须限制哪些数据库用户能够连接。

> **注意**
>
> 如第 20 章中所述，PostgreSQL 实际上是以“角色”为单位进行权限管理的。在本章中，我们统一使用数据库用户来表示“拥有 `LOGIN` 权限的角色”。

认证是数据库服务器确认客户端身份的过程，
并据此决定是否允许该客户端应用（或者运行该客户端应用的用户）
以请求的数据库用户名进行连接。

PostgreSQL 提供了多种不同的客户端认证方法。用于认证特定客户端连接的方法可以根据（客户端）主机地址、数据库和用户来选择。

PostgreSQL 数据库用户名在逻辑上独立于服务器所在操作系统中的用户名。如果某台服务器的所有用户在该机器上也都有账号，
那么为他们分配与操作系统用户名一致的数据库用户名是有意义的。不过，
接受远程连接的服务器可能有许多数据库用户并没有本地操作系统账号，在这种情况下，
数据库用户名与操作系统用户名之间就不必存在任何对应关系。

19.1. `pg_hba.conf` 文件

客户端认证由一个配置文件控制，该文件按惯例命名为 `pg_hba.conf`，并存放在数据库集簇的数据目录中。（HBA 代表 host-based authentication，即基于主机的认证。）当数据目录由 `initdb` 初始化时，会安装一个默认的 `pg_hba.conf` 文件。不过，也可以把认证配置文件放在别处；请参见配置参数hba_file。

`pg_hba.conf` 文件的一般格式是一组记录，每行一条。空行被忽略，# 注释字符之后的任何文本也被忽略。记录不能跨行。一条记录由若干字段组成，字段之间用空格和/或制表符分隔。
如果字段值被引号括起，字段中可以包含空白。把数据库、
用户或地址字段中的某个关键字（例如 `all` 或 `replication`）用引号括起会使该词失去特殊含义，只是匹配同名的数据库、用户或主机。

每条认证记录都指定一种连接类型、一个客户端 IP 地址范围（如果该连接类型需要）、一个数据库名、一个用户名，以及对匹配这些参数的连接要使用的认证方法。第一条同时匹配连接类型、客户端地址、请求数据库和用户名的记录会被用来执行认证。这里不存在“继续向后匹配”
或“后备”机制：如果选中某条记录而认证失败，就不会再考虑后续记录。
如果没有任何记录匹配，则拒绝访问。

记录可以采用以下七种格式之一

```
local      database  user  auth-method  [auth-options]
host       database  user  address  auth-method  [auth-options]
hostssl    database  user  address  auth-method  [auth-options]
hostnossl  database  user  address  auth-method  [auth-options]
```

```
host        database  user  IP-address  IP-mask  auth-method  [auth-
options]
hostssl     database  user  IP-address  IP-mask  auth-method  [auth-
options]
hostnossl   database  user  IP-address  IP-mask  auth-method  [auth-
options]
```

各字段的含义如下：

`local`

该记录匹配使用 Unix 域套接字的连接尝试。没有这种类型的记录时，Unix 域套接字连接将被禁止。

`host`

此记录匹配通过 TCP/IP 发起的连接尝试。`host` 记录既匹配 SSL 连接，也匹配非 SSL 连接。

> **注意**
>
> 除非服务器以适当的listen_addresses配置参数值启动，否则远程 TCP/IP 连接将不可用，因为默认行为是只在本地回环地址 `localhost` 上监听 TCP/IP 连接。

`hostssl`

该记录匹配使用 TCP/IP 发起的连接尝试，但仅当连接使用 SSL 加密时才匹配。

要使用此选项，服务器必须在构建时启用 SSL 支持。此外，还必须在服务器启动时通过设置ssl配置参数来启用 SSL（有关更多信息，请参见第 17.9 节）。

`hostnossl`

该记录类型与 `hostssl` 的行为相反；它只匹配通过 TCP/IP 发起且不使用 SSL 的连接尝试。

`database`

指定这条记录匹配哪些数据库名。`all` 表示匹配所有数据库。`sameuser` 表示如果请求的数据库名和请求的用户名相同，则匹配。`samerole` 表示请求的用户必须是与会话请求的数据库名相同的角色的成员。（`samegroup` 是 `samerole` 的一个已经废弃但仍然被接受的拼写。）The value `replication` specifies that the record matches if a replication connection is requested (note that replication connections do not specify any particular database). Otherwise, this is the name of a specific PostgreSQL database. Multiple database names can be supplied by separating them with commas. A separate file containing database names can be specified by preceding the file name with `@`.

`user`

指定这条记录匹配哪些数据库用户名。`all` 表示匹配所有用户。否则，它要么是一个具体数据库用户的名字，要么是一个前面带有 `+` 的组名。（回想一下，PostgreSQL 中用户和组之间没有真正的区别；`+` 标记的实际含义是“匹配直接或间接属于该角色成员的任何角色”，而不带 `+` 标记的名字只匹配那个具体的角色。）可以用逗号分隔多个用户名。也可以在文件名前加 `@`，指定一个单独存放用户名的文件。

address

指定此记录匹配的客户端机器地址。此字段可以包含主机名、IP地址范围或下面提到的特殊关键字之一。

IP地址范围使用标准的数字表示法来指定起始地址，然后是斜杠（/）和一个CIDR掩码长度。掩码长度表示客户端IP地址必须匹配的高位比特数。给定IP地址中右侧的比特应为零。IP地址、/和CIDR掩码长度之间不得有任何空白。

以这种方式指定的IPv4地址范围的典型示例包括172.20.143.89/32用于单个主机，或172.20.143.0/24用于小型网络，或10.6.0.0/16用于较大的网络。IPv6地址范围可能看起来像::1/128用于单个主机（在这种情况下是IPv6环回地址）或fe80::7a31:c1ff:0000:0000/96用于小型网络。0.0.0.0/0代表所有IPv4地址，::0/0代表所有IPv6地址。要指定单个主机，请对IPv4使用32的掩码长度，对IPv6使用128。在网络地址中，不要省略尾部的零。

以IPv4格式给出的条目将仅匹配IPv4连接，以IPv6格式给出的条目将仅匹配IPv6连接，即使所代表的地址在IPv4-in-IPv6范围内。请注意，如果系统的C库不支持IPv6地址，以IPv6格式给出的条目将被拒绝。

你也可以写all来匹配任何IP地址，samehost来匹配服务器自己的任何IP地址，或samenet来匹配服务器直接连接到的任何子网中的任何地址。

如果指定了主机名（任何不是IP地址范围或特殊关键字的内容都被视为主机名），则将该名称与客户端IP地址的反向名称解析结果进行比较（例如，如果使用DNS，则进行反向DNS查找）。主机名比较不区分大小写。如果匹配成功，则对主机名执行正向名称解析（例如，进行正向DNS查找），以检查其解析为的任何地址是否等于客户端IP地址。如果两个方向都匹配，则将条目视为匹配。（在pg_hba.conf中使用的主机名应该是客户端IP地址的地址到名称解析返回的名称，否则该行将不会匹配。一些主机名数据库允许将IP地址与多个主机名关联，但操作系统在要求解析IP地址时只会返回一个主机名。）

以点（.）开头的主机名规范匹配实际主机名的后缀。因此，.example.com将匹配foo.example.com（但不匹配单独的example.com）。

当在pg_hba.conf中指定主机名时，应确保名称解析相对快速。设置一个本地名称解析缓存可能会有帮助，如nscd。此外，还可能希望启用配置参数log_hostname，以便在日志中看到客户端的主机名而不是 IP 地址。

> 偶尔会有用户想知道，为什么要用这种看起来复杂的方式处理主机名——做两次名字解析并要求对 IP 地址做反向解析，而反向解析有时没有设置或者指向某个不合适的主机名。这主要是为了效率：一次连接尝试需要对当前客户端的地址做两次解析器查找。如果该地址存在解析器问题，那就只是那个客户端自己的问题。而假想的只做正向查找的替代实现，将不得不在每次连接尝试时解析 pg_hba.conf 中提到的每一个主机名。这本身就已经很慢了。而且如果其中某个主机名存在解析器问题，它就会成为所有人的问题。
>
> 此外，要实现后缀匹配特性也必须做反向解析，因为必须知道客户端的实际主机名才能将它与模式匹配。
>
> 请注意，这种行为与其他流行的基于主机名的访问控制实现一致，例如 Apache HTTP 服务器和 TCP Wrappers。

这个字段只适用于 `host`、`hostssl` 和 `hostnossl` 记录。

IP-address
IP-mask

这两个字段可以用作*IP-address/mask-length* 表示法的替代方案。不指定掩码长度，而是在单独一列中指定实际掩码。例如，`255.0.0.0`表示IPv4的CIDR掩码长度为8，而`255.255.255.255`表示CIDR掩码长度为32。

这些字段仅适用于 `host`、`hostssl` 和 `hostnossl` 记录。

auth-method

指定连接匹配此记录时要使用的认证方法。可选值在此处做了概述；详细说明见第 19.3 节。

`trust`

无条件允许连接。
这种方法允许任何能够连接到PostgreSQL数据库服务器的人以任意他们希望的PostgreSQL用户身份登录，无需密码或任何其他认证。详见第 19.3.1 节。

`reject`

无条件拒绝连接。这对于“过滤掉”某些主机很有用，例如一个`reject`行可以阻止特定主机连接，而后面的行允许特定网络中的其余主机连接。

`md5`

要求客户端提供一个经过双重 MD5 哈希的密码以完成认证。详见第 19.3.2 节。

`password`

要求客户端提供未加密的密码以完成认证。由于密码会以明文形式通过网络发送，因此不应在不受信任的网络上使用。详见第 19.3.2 节。

`gss`

使用 GSSAPI 认证用户。这仅适用于 TCP/IP 连接。详情参见第 19.3.3 节。

`sspi`

使用 SSPI 对用户进行认证。这仅适用于 Windows。详见第 19.3.4 节。

`krb5`

使用 Kerberos V5 对用户进行认证。这只适用于 TCP/IP 连接。详见第 19.3.5 节。

`ident`

通过联系客户端上的 ident 服务器获取客户端的操作系统用户名，并检查它是否与请求的数据库用户名匹配。Ident 认证只能用于 TCP/IP 连接。当为本地连接指定时，将改为使用 peer 认证。详见第 19.3.6 节。

`peer`

从操作系统获取客户端的操作系统用户名，并检查是否与请求的数据库用户名匹配。这仅适用于本地连接。有关详细信息，请参见第 19.3.7 节。

`ldap`

使用LDAP服务器进行认证。详见第 19.3.8 节。

`radius`

使用 RADIUS 服务器进行认证。详见第 19.3.9 节。

`cert`

使用 SSL 客户端证书进行认证。详见第 19.3.10 节。

`pam`

使用操作系统提供的可插拔认证模块（PAM）服务进行认证。详见第 19.3.11 节。

auth-options

在 *auth-method* 字段之后，可以有一个或多个形如 *name=value* 的字段，用于指定认证方法的选项。下文会详细说明各认证方法有哪些可用选项。

由 `@` 构造引用的文件会被读取为名称列表，其中的名称可以用空白或逗号分隔。注释仍然用 `#` 引入，与 `pg_hba.conf` 中相同，并且允许嵌套的 `@` 构造。除非 `@` 后面的文件名是绝对路径，否则它会被视为相对于引用它的文件所在目录。

由于每次连接尝试都会按顺序检查 `pg_hba.conf` 记录，因此记录的顺序很重要。通常，靠前的记录使用较严格的连接匹配条件和较弱的认证方法，
靠后的记录使用较宽松的匹配条件和较强的认证方法。例如，可能希望对本地 TCP/IP 连接使用 `trust` 认证，而要求远程 TCP/IP 连接提供密码。此时，为来自 127.0.0.1 的连接指定 `trust` 认证的记录，应放在为更大范围的允许客户端 IP 地址指定密码认证的记录之前。

在启动时以及主服务器进程收到 SIGHUP 信号时，`pg_hba.conf` 文件会被读取。如果你在运行中的系统上编辑了该文件，就需要通知 postmaster（使用 `pg_ctl reload` 或 `kill -HUP`）重新读取该文件。

> **提示**
>
> 要连接到一个特定数据库，一个用户必须不仅要通过`pg_hba.conf`检查，
> 还必须要有该数据库上的`CONNECT`权限。
> 如果你希望限制哪些用户能够连接到哪些数据库，授予/撤销`CONNECT`权限通常比在`pg_hba.conf`项中设置规则简单。

例 19.1展示了一些 `pg_hba.conf` 条目示例。有关不同认证方法的详细信息，参见下一节。

例 19.1. 示例 `pg_hba.conf` 项

```
# Allow any user on the local system to connect to any database with
# any database user name using Unix-domain sockets (the default for
 local
# connections).
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
local   all             all                                     trust
```

```
# The same using local loopback TCP/IP connections.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             127.0.0.1/32            trust

# The same as the previous line, but using a separate netmask column
#
# TYPE  DATABASE        USER            IP-ADDRESS      IP-MASK
      METHOD
host    all             all             127.0.0.1       255.255.255.
255     trust

# The same over IPv6.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             ::1/128                 trust

# The same using a host name (would typically cover both IPv4 and
 IPv6).
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             localhost               trust

# Allow any user from any host with IP address 192.168.93.x to connect
# to database "postgres" as the same user name that ident reports for
# the connection (typically the operating system user name).
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    postgres        all             192.168.93.0/24         ident

# Allow any user from host 192.168.12.10 to connect to database
# "postgres" if the user's password is correctly supplied.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    postgres        all             192.168.12.10/32        md5

# Allow any user from hosts in the example.com domain to connect to
# any database if the user's password is correctly supplied.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             .example.com            md5

# In the absence of preceding "host" lines, these two lines will
# reject all connections from 192.168.54.1 (since that entry will be
# matched first), but allow Kerberos 5 connections from anywhere else
# on the Internet.  The zero mask causes no bits of the host IP
# address to be considered, so it matches any host.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             192.168.54.1/32         reject
host    all             all             0.0.0.0/0               krb5

# Allow users from 192.168.x.x hosts to connect to any database, if
# they pass the ident check.  If, for example, ident says the user is
```

```
# "bryanh" and he requests to connect as PostgreSQL user "guest1", the
# connection is allowed if there is an entry in pg_ident.conf for map
# "omicron" that says "bryanh" is allowed to connect as "guest1".
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    all             all             192.168.0.0/16          ident
 map=omicron

# If these are the only three lines for local connections, they will
# allow local users to connect only to their own databases (databases
# with the same name as their database user name) except for
 administrators
# and members of role "support", who can connect to all databases.
 The file
# $PGDATA/admins contains a list of names of administrators.
 Passwords
# are required in all cases.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
local   sameuser        all                                     md5
local   all             @admins                                 md5
local   all             +support                                md5

# The last two lines above can be combined into a single line:
local   all             @admins,+support                        md5

# The database column can also use lists and file names:
local   db1,db2,@demodbs  all                                   md5
```

19.2. 用户名映射

当使用 Ident 或 GSSAPI 之类的外部认证系统时，发起连接的操作系统用户名可能不同于要使用的数据库用户。在这种情况下，可以通过用户名映射把操作系统用户名映射为数据库用户。
要使用用户名映射，需要在 `pg_hba.conf` 的选项字段中指定 `map=`*`map-name`*。此选项适用于所有会接收外部用户名的认证方法。由于不同连接可能需要不同映射，在 `pg_hba.conf` 中通过 *`map-name`* 参数指定要使用的映射，以表明每个连接应使用哪个映射。

用户名映射在 ident 映射文件中定义。该文件默认名为 `pg_ident.conf`，存储在集簇的数据目录中。（也可以将映射文件放在其他位置；参见ident_file配置参数。）ident 映射文件中各行的基本格式如下：

`map-name system-username database-username`

注释和空白的处理方式与 `pg_hba.conf` 相同。*`map-name`* 是任意名称，用于在 `pg_hba.conf` 中引用此映射。另两个字段指定操作系统用户名及与之匹配的数据库用户名。同一个 *`map-name`* 可以重复使用，以便在一项映射中指定多个用户对应关系。

一个操作系统用户可以对应多少个数据库用户，没有限制，反之亦然。因此，
映射中的条目应理解为“此操作系统用户可以作为此数据库用户连接”，而不是说二者等同。
只要存在一个映射条目，
将从外部认证系统取得的用户名与用户请求连接时使用的数据库用户名配对，就会允许该连接。

如果 *`system-username`* 字段以斜杠（`/`）开头，该字段的其余部分就会被视为正则表达式。（参见第 9.7.3.1 节，其中介绍了 PostgreSQL 的正则表达式语法。）正则表达式可以

包含一个捕获，即用圆括号括起来的子表达式，随后可以在 *database-username* 字段中使用 \1（反斜杠加数字一）来引用。这样就可以在一行中映射多个用户名，对简单的语法替换尤其有用。例如，以下条目

```
mymap   /^(.*)@mydomain\.com$      \1
mymap   /^(.*)@otherdomain\.com$   guest
```

会去掉系统用户名中的域名部分（如果用户名以 @mydomain.com 结尾），并允许系统用户名以 @otherdomain.com 结尾的任何用户登录为 guest。

> **提示**
>
> 记住在默认情况下，一个正则表达式可以只匹配字符串的一部分。如上例所示，使用^和$来强制匹配整个系统用户名通常是明智的。

在启动时以及主服务器进程收到 SIGHUP 信号时，pg_ident.conf 文件会被读取。如果你在运行中的系统上编辑了该文件，就需要通知 postmaster（使用 pg_ctl reload 或 kill -HUP）重新读取该文件。

例 19.2展示了一个可与例 19.1中的 pg_hba.conf 文件配合使用的 pg_ident.conf 文件。在这个示例中，任何登录到 192.168 网络中某台机器上的用户，如果其操作系统用户名不是 bryanh、ann 或 robert，都不会被授予访问权限。Unix 用户 robert 只有在尝试以 PostgreSQL 用户 bob 身份连接时才被允许访问，而不能以 robert 或其他身份连接。ann 只能以 ann 身份连接。用户 bryanh 则可以以 bryanh 或 guest1 身份连接。

例 19.2. 示例 pg_ident.conf 文件

```
# MAPNAME       SYSTEM-USERNAME         PG-USERNAME

omicron         bryanh                  bryanh
omicron         ann                     ann
# bob has user name robert on these machines
omicron         robert                  bob
# bryanh can also connect as guest1
omicron         bryanh                  guest1
```

19.3. 认证方法

下面几节会更详细地介绍这些认证方法。

19.3.1. 信任认证

当trust认证被指定时，PostgreSQL假设任何可以连接到服务器的人都被授权使用他们指定的任何数据库用户名（即使是超级用户）访问数据库。当然，在database和 user列中设置的限制仍然适用。只有当在操作系统层对进入服务器的连接有足够保护时，才应该使用这种方法。

trust 认证对单用户工作站上的本地连接来说是合适且非常方便的。在多用户机器上，通常单用它并不合适。不过，如果用文件系统权限限制对服务器 Unix 域套接字文件的访问，即使在多用户机器上也可以使用 trust。为此，可按第 18.3 节中的说明设置 unix_socket_permissions（可能还有 unix_socket_group）配置参数。或者，也可以设置 unix_socket_directory 配置参数，把套接字文件放在一个经过适当限制的目录中。

设置文件系统权限只能有助于 Unix 套接字连接。本地 TCP/IP 连接不会被文件系统权限限制。因此，如果你想利用文件系统权限来控制本地安全，那么从`pg_hba.conf`中移除`host ... 127.0.0.1 ...`行，或者把它改为一个非`trust`认证方法。

只有当你信任由 `pg_hba.conf` 中指定 `trust` 的行所允许连接的每台机器上的每个用户时，`trust` 认证才适合用于 TCP/IP 连接。对来自 localhost（127.0.0.1）以外的任何 TCP/IP 连接使用 `trust`，通常都不合理。

19.3.2. 密码认证

基于密码的认证方法有 `md5` 和 `password`。除了密码通过连接发送的方式——前者以 MD5 哈希发送、后者以明文发送——之外，这两种方法的操作是相似的。

如果你担心密码“嗅探”攻击，那么应优先使用 `md5`。如果可能，应始终避免使用明文 `password`。不过，`md5` 不能与db_user_namespace功能一起使用。如果连接被 SSL 加密保护着，那么可以安全地使用 `password`（不过如果依靠 SSL，SSL 证书认证可能是更好的选择）。

PostgreSQL 数据库密码独立于操作系统用户密码。每个数据库用户的密码存储在 `pg_authid` 系统目录中。可以使用 SQL 命令CREATE USER和ALTER ROLE管理密码，例如 **`CREATE USER foo WITH PASSWORD 'secret'`**，如果未为某个用户设置密码，存储的密码就是空值，该用户的密码认证始终会失败。

19.3.3. GSSAPI 认证

GSSAPI 是 RFC 2743 定义的安全认证行业标准协议。PostgreSQL 按照 RFC 1964 支持使用 Kerberos 认证的 GSSAPI。GSSAPI 为支持它的系统提供自动认证（单点登录）。认证过程本身是安全的，但除非使用 SSL，否则数据库连接上传输的数据不会加密。

当 GSSAPI 使用 Kerberos 时，它使用标准格式 *`servicename/hostname@realm`* 的 principal。有关 principal 的组成部分以及如何设置所需密钥的信息，见第 19.3.5 节。

当编译PostgreSQL时，GSSAPI 支持必须被启用，详见第 15 章。

以下配置选项适用于 GSSAPI：

`include_realm`

如果设为 1，已认证用户主体中的 realm 名称将包含在传递给用户名映射（第 19.2 节）的系统用户名中。这是推荐的配置，因为否则将无法区分来自不同 realm 的同名用户。该参数的默认值为 0（即系统用户名中不包含 realm），但在未来版本的 PostgreSQL 中可能会改为 1。用户可以显式设置它，以避免升级时出现任何问题。

`map`

允许在系统用户名与数据库用户名之间建立映射。详见第 19.2 节。对于 `username@EXAMPLE.COM`（或较少见的 `username/hostbased@EXAMPLE.COM`）这样的 GSSAPI/Kerberos 主体，映射默认使用的用户名是 `username`（或相应的 `username/hostbased`）；但如果按上文建议将 `include_realm` 设为了 1，那么映射时视为系统用户名的就是 `username@EXAMPLE.COM`（或 `username/hostbased@EXAMPLE.COM`）。

`krb_realm`

设置用于匹配用户主体名的 realm。如果设置了该参数，则只接受来自该 realm 的用户；如果未设置，则允许来自任意 realm 的用户连接，但仍受已执行的用户名映射约束。

19.3.4. SSPI 认证

SSPI 是一种提供安全认证和单点登录的 Windows 技术。PostgreSQL 会以 `negotiate` 模式使用 SSPI，尽可能使用 Kerberos，否则自动回退到 NTLM。只有服务器和客户端都运行 Windows，或者在非 Windows 平台上可用 GSSAPI 时，SSPI 认证才能工作。

当使用Kerberos认证时，SSPI和GSSAPI的工作方式相同，详见第 19.3.3 节。

SSPI 支持下列配置选项：

`include_realm`

如果设为 1，已认证用户主体中的 realm 名称将包含在传递给用户名映射（第 19.2 节）的系统用户名中。这是推荐的配置，因为否则将无法区分来自不同 realm 的同名用户。该参数的默认值为 0（即系统用户名中不包含 realm），但在未来版本的 PostgreSQL 中可能会改为 1。用户可以显式设置它，以避免升级时出现任何问题。

`map`

允许在系统用户名与数据库用户名之间建立映射。详见第 19.2 节。对于 SSPI/Kerberos 主体，例如 `username@EXAMPLE.COM`（或者较少见的 `username/hostbased@EXAMPLE.COM`），映射默认使用的用户名是 `username`（或相应的 `username/hostbased`）；但如果按上文建议将 `include_realm` 设为了 1，那么映射时视为系统用户名的就是 `username@EXAMPLE.COM`（或 `username/hostbased@EXAMPLE.COM`）。

`krb_realm`

设置用于匹配用户主体名的 realm。如果设置了该参数，则只接受来自该 realm 的用户；如果未设置，则允许来自任意 realm 的用户连接，但仍受已执行的用户名映射约束。

19.3.5. Kerberos 认证

> **注意**
>
> 原生 Kerberos 认证已被弃用，只应用于向后兼容。鼓励新建和升级的安装改用行业标准 GSSAPI 认证方法（见第 19.3.3 节）。

Kerberos 是一种适合在公共网络上进行分布式计算的行业标准安全认证系统。对 Kerberos 系统的描述超出了本文档的范围；在最一般的意义上它可能相当复杂（但也很强大）。Kerberos FAQ[1] 或 MIT Kerberos 页面[2]可以作为很好的探索起点。存在若干 Kerberos 发行版来源。Kerberos 提供安全认证，但不对网络上传递的查询或数据加密；如需加密请使用 SSL。

PostgreSQL 支持 Kerberos 版本 5。构建 PostgreSQL 时必须启用 Kerberos 支持；更多信息见第 15 章。

PostgreSQL 像一个普通的 Kerberos 服务那样运作。服务主体的名称为 *`servicename/hostname@realm`*。

`servicename` can be set on the server side using the krb_srvname configuration parameter, and on the client side using the `krbsrvname` connection parameter. (See also 第 31.1 节.) 安装时的默认值可以在构建时用 `./configure --with-krb-srvnam=`

[1] http://www.cmf.nrl.navy.mil/CCS/people/kenh/kerberos-faq.html
[2] http://web.mit.edu/kerberos/www/

whatever 从默认的 postgres 改变。在大多数环境中，这个参数永远不需要更改。但是，在同一台主机上支持多个 PostgreSQL 安装时需要更改它。Some Kerberos implementations might also require a different service name, such as Microsoft Active Directory which requires the service name to be in upper case (POSTGRES).

hostname 是服务器机器的完全限定主机名。服务主体的 realm 是服务器机器的首选 realm。

客户端主体的第一个组成部分必须是 PostgreSQL 数据库用户名，例如 pgusername@realm。或者，也可以使用用户名映射，将主体名的第一个组成部分映射为数据库用户名。默认情况下，PostgreSQL 不检查客户端的 realm。如果启用了跨 realm 认证并且需要验证 realm，请使用 krb_realm 参数，或者启用 include_realm 并使用用户名映射来检查 realm。

确保你的服务器密钥表文件可被 PostgreSQL 服务器账户读取（最好只有该账户可读）。（另见第 17.1 节。）密钥文件的位置由krb_server_keyfile配置参数指定。默认为 /usr/local/pgsql/etc/krb5.keytab（或构建时指定为 sysconfdir 的目录）。

keytab 文件由 Kerberos 软件生成；详情参见 Kerberos 文档。以下示例适用于兼容 MIT 的 Kerberos 5 实现：

```
kadmin% ank -randkey postgres/server.my.domain.org
kadmin% ktadd -k krb5.keytab postgres/server.my.domain.org
```

连接数据库时，请确保持有与请求的数据库用户名相匹配的主体票据。例如，数据库用户名为 fred 时，主体 fred@EXAMPLE.COM 可以连接。如果还要允许主体 fred/users.example.com@EXAMPLE.COM，请按第 19.2 节所述使用用户名映射。

如果你在 Apache Web 服务器上使用 mod_auth_kerb[3]和 mod_perl，可以在 mod_perl 脚本中使用 AuthType KerberosV5SaveCredentials。这可提供经过 Web 的安全数据库访问，而无需额外的口令。

Kerberos 支持以下配置选项：

map

允许在系统用户名和数据库用户名之间进行映射。详见第 19.2 节。

include_realm

如果设置为 1，已认证用户主体中的领域名将包含在传递给用户名映射（第 19.2 节）的系统用户名中。这有助于处理来自多个领域的用户。

krb_realm

设置用于匹配用户主体名的 realm。如果设置了该参数，则只接受来自该 realm 的用户；如果未设置，则允许来自任意 realm 的用户连接，但仍受已执行的用户名映射约束。

krb_server_hostname

设置服务主体的主机名部分。它与 krb_srvname 组合生成完整的服务主体，即 krb_srvname/krb_server_hostname@REALM。如果未设置，默认为服务器主机名。

19.3.6. Ident 认证

ident 认证方法通过从一个 ident 服务器获得客户端的操作系统用户名并且用它作为被允许的数据库用户名（和可选的用户名映射）来工作。它只在 TCP/IP 连接上支持。

[3] http://modauthkerb.sf.net

> **注意**
>
> 当为一个本地（非 TCP/IP）连接指定 ident 时，将实际使用 peer 认证（见第 19.3.7 节）。

以下配置选项适用于 ident：

`map`

允许在系统用户名和数据库用户名之间进行映射。详见第 19.2 节。

“标识协议”在 RFC 1413 中定义。几乎所有类 Unix 操作系统都自带 ident 服务器，默认监听 TCP 端口 113。ident 服务器的基本功能是回答这样的问题：“从你的端口 *X* 连到我的端口 *Y* 的连接，是哪个用户发起的？”由于建立物理连接时，PostgreSQL 已知 *X* 和 *Y*，因此可以询问连接客户端所在主机上的 ident 服务器，理论上能够确定任意给定连接的操作系统用户。

这个过程的缺点在于它依赖客户端本身的可信性：如果客户端机器不可信或者已被攻破，
攻击者几乎可以在 113 端口上运行任何程序，并返回任意他选择的用户名。因此，这种认证方法只适用于封闭网络，在这类网络中每台客户端机器都受到严格控制，
而且数据库管理员与系统管理员之间保持密切协作。换句话说，你必须信任运行 ident 服务器的那台机器。请注意下面的警告：

> 标识协议的本意不是作为一种授权或访问控制协议。
>
> —RFC 1413

有些 ident 服务器提供了一个非标准选项，会让返回的用户名被加密，而解密所需密钥只有发起连接机器的管理员才知道。将 ident 服务器与 PostgreSQL 配合使用时，绝不能启用这个选项，因为 PostgreSQL 无法解密返回的字符串，也就无法确定实际用户名。

19.3.7. Peer 认证

Peer 认证通过从内核获取客户端的操作系统用户名，并把它用作被允许的数据库用户名（可结合可选的用户名映射）来工作。这种方法只支持本地连接。

以下配置选项适用于 peer：

`map`

允许在系统用户名和数据库用户名之间进行映射。详见第 19.2 节。

Peer 认证只在提供 `getpeereid()` 函数、`SO_PEERCRED` 套接字参数或类似机制的操作系统上可用。目前这包括 Linux、大多数 BSD 变种（包括 OS X）以及 Solaris。

19.3.8. LDAP 认证

这种认证方法的工作方式与 `password` 类似，只不过它使用 LDAP 作为密码验证方法。LDAP 只用于验证用户名/密码对。因此，在使用 LDAP 进行认证之前，用户必须已经存在于数据库中。

LDAP 认证可以以两种模式运作。在第一种模式中，服务器将绑定到由 *prefix* *username* *suffix* 构造的可分辨名称上。通常，*prefix* 参数用于指定 `cn=`，或者在 Active Directory 环境中指定 *DOMAIN*\。*suffix* 用于在非 Active Directory 环境中指定 DN 的剩余部分。

在第二种模式中，服务器首先用一个固定的用户名和密码绑定到 LDAP 目录（由 *ldapbinddn* 和 *ldapbindpasswd* 指定），然后为尝试登录数据库的用户执行搜索。如果没有配置用户和密码，将尝试对目录做匿名绑定。搜索将在 *ldapbasedn* 的子树上进行，并尝试对 *ldapsearchattribute* 中指定的属性做精确匹配。如果没有指定属性，将使用 `uid` 属性。Once the user has been found in this search, the server disconnects and re-binds to the directory as this user, using the password specified by the client, to verify that the login is correct. This method allows for significantly more flexibility in where the user objects are located in the directory, but will cause two separate connections to the LDAP server to be made.

The following configuration options are supported for LDAP:

`ldapserver`

要连接的 LDAP 服务器的名字或 IP 地址。可以指定多个服务器，用空格分隔。

`ldapport`

要连接的LDAP服务器的端口号。如果未指定端口，则将使用LDAP库的默认端口设置。

`ldaptls`

设为 1 时，PostgreSQL 与 LDAP 服务器之间的连接会使用 TLS 加密。注意，这只加密与 LDAP 服务器之间的流量 – 除非使用 SSL，否则与客户端之间的连接仍不加密。

`ldapprefix`

在进行简单绑定认证时，附加到用户名前面以形成绑定 DN 的字符串。

`ldapsuffix`

在进行简单绑定认证时，附加到用户名后面以形成绑定 DN 的字符串。

`ldapbasedn`

在进行搜索+绑定认证时，用作用户搜索起点的根 DN。

`ldapbinddn`

在进行搜索+绑定认证时，用于绑定到目录并执行搜索的用户 DN。

`ldapbindpasswd`

在进行搜索+绑定认证时，用于绑定到目录并执行搜索的用户密码。

`ldapsearchattribute`

做 search+bind 认证时，在搜索中与用户名匹配的属性。

注意

由于 LDAP 通常使用逗号和空格来分隔 DN 的不同部分，配置 LDAP 选项时常常需要使用双引号括起的参数值，例如：

```
ldapserver=ldap.example.net ldapprefix="cn=" ldapsuffix=", dc=
example, dc=net"
```

19.3.9. RADIUS 认证

这种认证方法的工作方式与 `password` 类似，只不过它使用 RADIUS 作为密码验证方式。RADIUS 只用于验证用户名/密码对。因此，在使用 RADIUS 进行认证之前，用户必须已经存在于数据库中。

使用 RADIUS 认证时，会向配置好的 RADIUS 服务器发送一条 Access Request 消息。该请求的类型为 `Authenticate Only`，并包含 `user name`、`password`（加密的）以及 `NAS Identifier` 参数。该请求会使用与服务器共享的密钥进行加密。这台服务器会返回 `Access Accept` 或 `Access Reject` 作为响应。PostgreSQL 不支持 RADIUS 记账。

RADIUS 支持下列配置选项：

`radiusserver`

要连接的 RADIUS 服务器的名称或 IP 地址。此参数为必需项。

`radiussecret`

与 RADIUS 服务器进行安全通信时使用的共享密钥。它在 PostgreSQL 服务器和 RADIUS 服务器上必须完全相同。建议它至少是一个 16 个字符长的字符串。此参数为必需项。

> **注意**
>
> 只有当 PostgreSQL 在构建时启用了 OpenSSL 支持，所使用的加密向量才具有足够的密码学强度。在其他情况下，到 RADIUS 服务器的传输只能被视为经过混淆，而非受到安全保护；如有必要，应额外采取外部安全措施。

`radiusport`

要连接的 RADIUS 服务器端口号。如果未指定端口，则会使用默认端口 `1812`。

`radiusidentifier`

在 RADIUS 请求中用作 `NAS Identifier` 的字符串。这个参数可用作第二个参数，例如标识用户正尝试以哪个数据库用户进行认证，从而便于在 RADIUS 服务器上进行策略匹配。如果未指定标识符，则默认使用 `postgresql`。

19.3.10. 证书认证

这种认证方法使用 SSL 客户端证书进行认证，因此仅适用于 SSL 连接。使用此方法时，服务器要求客户端提供有效的证书，不会向客户端发送密码提示。服务器会将证书的 `cn`（通用名称）属性与请求的数据库用户名比较，匹配时才允许登录。可以使用用户名映射，允许 `cn` 与数据库用户名不同。

SSL 证书认证支持下列配置选项：

`map`

允许在系统用户名和数据库用户名之间进行映射。详见第 19.2 节。

19.3.11. PAM 认证

这种认证方法的工作方式与 `password` 类似，只是它使用 PAM（可插拔认证模块）作为认证机制。默认的 PAM 服务名是 `postgresql`。PAM 只用于验证用户名/密码对。因此，在能够使用 PAM 认证之前，用户必须已经存在于数据库中。For more information about PAM, please read the Linux-PAM Page[4] and the Solaris PAM Page[5].

PAM 支持下列配置选项：

`pamservice`

PAM 服务名称。

> **注意**
>
> 如果 PAM 被设置为读取 `/etc/shadow`，认证将会失败，因为 PostgreSQL 服务器是由非 root 用户启动的。不过，当 PAM 被配置为使用 LDAP 或其他认证方法时，这就不是问题。

19.4. 认证问题

认证失败及相关问题通常会表现为类似下面这样的错误消息：

```
FATAL:  no pg_hba.conf entry for host "123.123.123.123", user "andym",
 database "testdb"
```

这种情况通常意味着你已经成功联系到了服务器，但服务器不愿意接受你的连接。
正如消息所示，服务器拒绝了该连接请求，因为它没有在自己的 `pg_hba.conf` 配置文件中找到匹配项。

```
FATAL:  password authentication failed for user "andym"
```

这样的消息表示你已经联系到了服务器，而且服务器也愿意继续处理连接，
但前提是你必须先通过 `pg_hba.conf` 文件中指定的认证方法。请检查你提供的密码；如果错误消息提到了 Kerberos 或 ident 等认证类型，也请检查对应的软件配置。

```
FATAL:  user "andym" does not exist
```

指定的数据库用户不存在。

```
FATAL:  database "testdb" does not exist
```

你试图连接的数据库不存在。注意，如果你没有指定数据库名，
默认会使用数据库用户名作为数据库名，但这不一定是所需的数据库名。

> **提示**
>
> 服务器日志中可能包含比返回给客户端的更多认证失败信息。如果你不清楚失败原因，
> 请检查服务器日志。

[4] http://www.kernel.org/pub/linux/libs/pam/
[5] http://www.sun.com/software/solaris/pam/

第 20 章 数据库角色

PostgreSQL使用角色这一概念来管理数据库访问权限。角色可以被看作数据库用户，
也可以被看作一组数据库用户，这取决于该角色的设置方式。
角色可以拥有数据库对象（例如表和函数），并且可以把这些对象上的权限赋予其他角色，
以控制谁能访问哪些对象。此外，还可以把一个角色中的成员资格授予另一个角色，
从而允许成员角色使用授予该角色的权限。

角色这一概念涵盖了“用户”和“组”这两个概念。在 PostgreSQL 8.1 之前的版本中，用户和组是两类不同的实体，但现在只有角色。任意角色都可以充当用户、组，或者同时充当两者。

本章描述如何创建和管理角色。有关角色权限对各种数据库对象影响的更多信息，见第 5.6 节。

20.1. 数据库角色

数据库角色在概念上与操作系统用户完全分离。在实践中，
维持两者之间的对应关系可能比较方便，但这并非必需。
数据库角色在整个数据库集簇安装中是全局的（而非每个数据库各自独立）。要创建角色，
请使用CREATE ROLE SQL 命令：

```
CREATE ROLE name;
```

name 遵循 SQL 标识符规则：要么不加修饰且不含特殊字符，要么用双引号括起。（在实践中，通常还需要为该命令添加其他选项，例如 `LOGIN`。下文会给出更多细节。）要移除现有角色，请使用类似的DROP ROLE命令：

```
DROP ROLE name;
```

为了方便，提供了createuser和dropuser程序，作为这些 SQL 命令的包装器，可以从 shell 命令行调用：

```
createuser name
dropuser name
```

要确定现有角色的集合，请查看 `pg_roles` 系统目录，例如

```
SELECT rolname FROM pg_roles;
```

使用psql程序的 `\du` 元命令也可以列出现有角色。

为了完成数据库系统的初始引导，一个刚初始化的系统总会包含一个预定义角色。
该角色总是一个“超级用户”，并且除非另行指定了不同名称，否则它的名称与用`initdb`
初始化数据库集簇的操作系统用户相同。这个角色通常命名为`postgres`。要创建更多角色，
必须先以这个初始角色建立连接。

每个到数据库服务器的连接都是以某个特定角色名建立的，
而这个角色决定在该连接中发出命令时的初始访问权限。某个数据库连接要使用的角色名，
由发起连接请求的客户端以应用特定的方式指明。例如，`psql`
程序使用`-U`命令行选项来指明要以哪个角色连接。
很多应用默认假定使用当前操作系统用户的名字（包括`createuser`和`psql`）。因此，
在角色名和操作系统用户名之间维护对应关系通常很方便。

某个客户端连接能够以哪些数据库角色进行连接，由客户端认证设置决定，如第 19 章中所述。（因此，客户端并不限于只能以与其操作系统用户匹配的角色连接，
正如一个人的登录名不必与其真实姓名一致。）
由于角色身份决定了已连接客户端可用的权限集合，所以在设置多用户环境时，
仔细配置权限非常重要。

20.2. 角色属性

数据库角色可以具有若干定义其权限的属性，这些属性还会与客户端认证系统交互。

登录权限

只有具有LOGIN属性的角色才能用作数据库连接的初始角色名。具有LOGIN
属性的角色可以被视为与“数据库用户”等同。要创建具有登录权限的角色，可以使用：

```
CREATE ROLE name LOGIN;
CREATE USER name;
```

（CREATE USER等价于CREATE ROLE，只是CREATE USER默认假定LOGIN，而CREATE ROLE不是。）

超级用户状态

数据库超级用户会绕过除登录权和发起复制权限之外的所有权限检查。
这是一种危险的权限，不应草率使用；最好以非超级用户角色完成大部分工作。
要创建新的数据库超级用户，使用CREATE ROLE *name* SUPERUSER。这必须由已经是超级用户的角色来执行。默认情况下，创建超级用户还会授予发起流复制的权限。
为了提高安全性，可以使用CREATE ROLE *name* SUPERUSER NOREPLICATION 来禁止该权限。

创建数据库

必须显式授予角色创建数据库的权限（超级用户除外，因为超级用户绕过所有权限检查）。
要创建这样的角色，使用CREATE ROLE *name* CREATEDB。

创建角色

必须显式授予角色创建更多角色的权限（超级用户除外）。要创建这样的角色，使用CREATE ROLE *name* CREATEROLE。具有CREATEROLE权限的角色还可以修改和删除其他角色，以及授予或撤销其中的成员资格。但是，要创建、修改、
删除超级用户角色或更改其成员资格，需要超级用户身份；仅CREATEROLE是不够的。

发起复制

必须显式地给一个角色授予发起流复制的权限。用于流复制的角色还必须具有 LOGIN 权限。
要创建这样的角色，可使用 CREATE ROLE *name* REPLICATION LOGIN。

密码

只有当客户端认证方法要求用户在连接数据库时提供密码时，密码才有意义。
password和md5认证方法使用密码。数据库密码与操作系统密码是分开的。
在创建角色时用CREATE ROLE *name* PASSWORD '*string*'指定密码。

角色创建后，可以使用以下命令修改其属性：ALTER ROLE。详情请参见CREATE ROLE和ALTER ROLE命令的参考页。

> **提示**
>
> 一种良好做法是创建具有CREATEDB和CREATEROLE权限但不是超级用户的角色，然后使用此角色执行数据库和角色的所有日常管理工作。这种方法可以避免在实际上并不需要超级用户权限的任务中以超级用户身份操作所带来的风险。

角色还可以为第 18 章中描述的许多运行时配置设置指定角色特定默认值。例如，如果出于某种原因你希望每次连接时都禁用索引扫描（提示：这不是个好主意），你可以使用：

```
ALTER ROLE myname SET enable_indexscan TO off;
```

这会保存该设置（但不会立即生效）。在该角色后续建立的连接中，它看起来就像在会话开始之前执行了SET enable_indexscan TO off一样。你仍然可以在会话期间更改该设置；它只会作为默认值。要移除角色特定默认设置，使用ALTER ROLE *rolename* RESET *varname*。注意，附加到没有LOGIN权限的角色上的角色特定默认值几乎没有用，因为它们永远不会生效。

20.3. 角色成员资格

为了简化权限管理，把用户分组通常很方便：这样，权限就可以整体授予给一个组，或者从整个组中撤销。在PostgreSQL中，这通过创建一个表示该组的角色，然后把该组角色中的成员资格授予各个用户角色来实现。

要建立一个组角色，首先创建该角色：

```
CREATE ROLE name;
```

通常，作为组使用的角色不会有LOGIN属性，不过如果你愿意，也可以设置它。

组角色存在后，就可以使用GRANT和REVOKE命令添加和移除成员：

```
GRANT group_role TO role1, ... ;
REVOKE group_role FROM role1, ... ;
```

也可以把成员资格授予其他组角色（因为组角色与非组角色实际上没有任何区别）。数据库不允许建立循环成员关系。此外，也不允许将角色的成员资格授予 PUBLIC。

组角色的成员可以通过两种方式使用该角色的权限。首先，组的每个成员都可以显式执行SET ROLE，以暂时“成为”该组角色。在这种状态下，数据库会话可以使用组角色的权限，而不是原始登录角色的权限；创建的任何数据库对象也被视为由组角色而非登录角色拥有。其次，具有 INHERIT 属性的成员角色可以自动使用其所属角色的权限，包括这些角色继承的所有权限。例如，假定我们已经执行了：

```
CREATE ROLE joe LOGIN INHERIT;
CREATE ROLE admin NOINHERIT;
CREATE ROLE wheel NOINHERIT;
GRANT admin TO joe;
GRANT wheel TO admin;
```

以角色 joe 连接后，数据库会话立即可以使用直接授予 joe 的权限，以及授予 admin 的所有权限，因为 joe “继承” admin 的权限。不过，授予 wheel 的权限不可用，因为即使 joe 间接

属于 wheel，该成员资格也是通过 admin 获得的，而后者具有 NOINHERIT 属性。执行以下命令后：

```
SET ROLE admin;
```

会话将只能使用授予 admin 的权限，而不能使用授予 joe 的权限。执行以下命令后：

```
SET ROLE wheel;
```

会话将只能使用授予 wheel 的权限，而不能使用授予 joe 或 admin 的权限。使用以下任一命令可以恢复原始的权限状态：

```
SET ROLE joe;
SET ROLE NONE;
RESET ROLE;
```

注意

SET ROLE命令总是允许切换到原始登录角色直接或间接所属的任何角色。因此，在上例中，没有必要先成为admin再成为wheel。

注意

在 SQL 标准中，用户与角色之间有明确区别，并且用户不会自动继承权限，而角色会。这种行为可以在PostgreSQL中通过让用作 SQL 角色的角色具有INHERIT属性，而让用作 SQL 用户的角色具有NOINHERIT属性来获得。不过，出于与 8.1 之前版本向后兼容的考虑，PostgreSQL默认给所有角色都赋予INHERIT属性；在那些早期版本中，用户总能使用授予其所属组的权限。

角色属性LOGIN、SUPERUSER、CREATEDB和CREATEROLE可以被视为特殊权限，但它们从不会像数据库对象上的普通权限那样被继承。要使用其中之一，必须实际执行SET ROLE切换到拥有该属性的特定角色。继续上面的例子，我们可以选择把CREATEDB和CREATEROLE赋予admin角色。这样，一个以joe角色连接的会话并不会立即拥有这些权限，只有在执行SET ROLE admin之后才会拥有。

要删除组角色，请使用DROP ROLE：

```
DROP ROLE name;
```

该组角色中的所有成员资格都会自动被撤销（但成员角色不会受到其他影响）。

20.4. 删除角色

由于角色可以拥有数据库对象，并且可以拥有访问其他对象的权限，所以删除一个角色通常不只是简单执行一次DROP ROLE。必须先删除或重新分配该角色拥有的任何对象；并且必须撤销授予该角色的任何权限。

可以使用 ALTER 命令逐个转移对象的所有权，例如：

```
ALTER TABLE bobs_table OWNER TO alice;
```

另一种办法是使用REASSIGN OWNED命令，把待删除角色拥有的所有对象的所有权重新分配给另一个角色。由于 `REASSIGN OWNED` 无法访问其他数据库中的对象，因此必须在包含该角色所拥有对象的每个数据库中运行它。（请注意，首次执行这样的 `REASSIGN OWNED` 时，会更改待删除角色所拥有的所有跨数据库共享对象（即数据库或表空间）的所有权。）

一旦有价值的对象都已转移给新拥有者，待删除角色所拥有的其余对象即可使用DROP OWNED命令删除。同样，由于该命令不能访问其他数据库中的对象，
因此必须在包含该角色所拥有对象的每个数据库中运行它。另外，`DROP OWNED`不会删除整个数据库或表空间，因此如果该角色拥有任何尚未转移给新拥有者的数据库或表空间，
就必须手工删除它们。

`DROP OWNED`还会负责移除目标角色在其他角色所拥有的对象上获授的所有权限。由于`REASSIGN OWNED`不会处理这类对象，所以通常需要同时运行`REASSIGN OWNED`和`DROP OWNED`（按这个顺序！），才能完整移除待删除角色的依赖关系。

总之，移除曾用来拥有对象的角色，最通用的做法是：

```
REASSIGN OWNED BY doomed_role TO successor_role;
DROP OWNED BY doomed_role;
-- repeat the above commands in each database of the cluster
DROP ROLE doomed_role;
```

如果并非所有拥有的对象都要转移给同一个后继拥有者，最好先手工处理这些例外情况，
然后再执行上述步骤作最后清理。

如果在仍有依赖对象存在时尝试执行`DROP ROLE`，它会发出消息指出哪些对象需要被重新分配或删除。

20.5. 函数和触发器安全性

函数和触发器允许用户把代码插入到后端服务器中，而其他用户可能会无意间执行这些代码。
因此这两种机制都使得用户能比较容易地以“特洛伊木马”的方式坑害他人。
唯一真正的保护是严格控制谁可以定义函数。

函数在后端服务器进程中运行，并具有数据库服务器守护进程的操作系统权限。
如果函数所用的编程语言允许不受检查的内存访问，就有可能更改服务器的内部数据结构。
因此，这类函数可以做很多别的事情，其中就包括绕过任何系统访问控制。
允许这种访问的函数语言被认为是“不受信任的”，而
PostgreSQL只允许超级用户创建用这类语言编写的函数。

第 21 章 管理数据库

每个正在运行的PostgreSQL服务器实例都管理着一个或多个数据库。因此，数据库是组织SQL对象（“数据库对象”）的最高层级。本章描述数据库的属性，以及如何创建、管理和删除数据库。

21.1. 概述

数据库是一个命名的SQL对象（“数据库对象”）集合。通常，每个数据库对象（表、函数等）都属于且只属于一个数据库。（不过也有少量系统目录，例如`pg_database`，属于整个集簇，可以从集簇中的每个数据库访问。）更准确地说，数据库是模式的集合，而模式中包含表、函数等。因此完整的层次结构是：服务器、数据库、模式、表（或其他类型的对象，如函数）。

当连接到数据库服务器时，客户端必须在连接请求中指定要连接的数据库名。一个连接无法访问多个数据库。不过，一个应用可以不受限制地打开到同一数据库或其他数据库的多个连接。数据库在物理上是相互隔离的，访问控制在连接级别管理。因此，如果一个PostgreSQL服务器实例要容纳彼此应当分离、并且在大多数情况下互不知晓的项目或用户，建议把它们放到不同的数据库中。如果项目或用户是相互关联的，并且应该能够使用彼此的资源，则应当把它们放在同一个数据库中，但可能使用不同的模式。模式是纯逻辑结构，谁能访问什么由权限系统管理。更多关于管理模式的信息见第 5.7 节。

数据库使用`CREATE DATABASE`命令创建（见第 21.2 节），使用`DROP DATABASE`命令删除（见第 21.5 节）。要确定现有数据库的集合，可以检查系统目录`pg_database`，例如：

```
SELECT datname FROM pg_database;
```

psql程序的`\l`元命令和`-l`命令行选项也可用于列出现有数据库。

> **注意**
>
> SQL标准将数据库称为“目录”，但在实践中两者没有区别。

21.2. 创建数据库

要创建数据库，PostgreSQL服务器必须已经启动并正在运行（见第 17.3 节）。

数据库用 SQL 命令CREATE DATABASE创建：

```
CREATE DATABASE name;
```

其中*name*遵循SQL标识符的一般规则。当前角色会自动成为新数据库的拥有者。数据库拥有者有权在日后删除该数据库（这也会删除其中的全部对象，即使那些对象具有不同的拥有者）。

创建数据库是一项受限操作。关于如何授予该权限，见第 20.2 节。

由于执行`CREATE DATABASE`命令时必须先连接到数据库服务器，因此还会有一个问题：某个站点上的第一个数据库是如何创建的？第一个数据库总是由`initdb`命令在初始化数据存储区时创建的（见第 17.2 节）。这个数据库叫作`postgres`。因此，要创建第一个“普通”数据库，可以连接到`postgres`。

在初始化数据库集簇时，还会创建第二个数据库template1。
每当在集簇中创建一个新数据库时，新数据库本质上都是通过克隆template1得到的。
这意味着你在template1中所做的任何更改，都会传播到后来创建的所有数据库中。因此，
除非你希望这些对象出现在每一个新建数据库中，否则应避免在template1中创建对象。
更多细节见第 21.3 节。

为了方便，还可以从 shell 中执行一个程序来创建新数据库，即createdb。

```
createdb dbname
```

createdb并没有什么特殊之处。它连接到postgres数据库并发出CREATE DATABASE命令，
和上面描述的完全一样。关于调用细节见createdb参考页。注意，不带任何参数的createdb
会创建一个与当前用户名同名的数据库。

> **注意**
>
> 第 19 章包含关于如何限制谁能连接到给定数据库的信息。

有时你会想为其他人创建数据库，并让他们成为新数据库的拥有者，
以便他们自行配置和管理它。要实现这一点，可使用下列命令之一：

```
CREATE DATABASE dbname OWNER rolename;
```

这是在 SQL 环境中使用；或者：

```
createdb -O rolename dbname
```

这是在 shell 中使用。只有超级用户才允许为其他人（即为一个你不是其成员的角色）创建数据库。

21.3. 模板数据库

CREATE DATABASE实际上是通过复制一个现有数据库来创建新数据库。默认情况下，它复制名为template1的标准系统数据库。因此，该数据库就是创建新数据库时使用的“模板”。
如果你向template1中添加对象，这些对象也会被复制到以后创建的用户数据库中。
这种行为允许对数据库中的标准对象集合进行站点本地修改。例如，如果你将过程语言
PL/Perl安装到template1中，那么以后创建用户数据库时无需额外操作，它就会自动可用。

还有第二个标准系统数据库，名为template0。该数据库包含与template1
初始内容相同的数据，也就是说，只包含你的PostgreSQL版本预定义的标准对象。
在数据库集簇初始化之后，template0就不应再被修改。通过指示CREATE DATABASE复制template0而不是template1，你可以创建一个“纯净的”用户数据库，它不包含
template1中的任何站点本地附加对象。这在恢复pg_dump
转储时特别方便：转储脚本应当恢复到一个纯净的数据库中，
以确保重建出被转储数据库的正确内容，而不会与后来可能加入template1的对象发生冲突。

复制template0而不是template1的另一个常见原因是：复制template0
时可以指定新的编码和区域设置，而template1的副本则必须使用与它相同的设置。
这是因为template1可能包含依赖特定编码或区域设置的数据，而已知template0
不会包含这类数据。

要通过复制template0创建数据库，可在 SQL 环境中使用：

```
CREATE DATABASE dbname TEMPLATE template0;
```

或在 shell 中使用：

```
createdb -T template0 dbname
```

可以创建额外的模板数据库，实际上也可以通过将集簇中任意数据库的名称指定为CREATE DATABASE的模板来复制它。不过，必须理解的是，这（至少目前）并不是一个通用的“COPY DATABASE”功能。最主要的限制是：复制源数据库时，不能有其他会话连接到该数据库。如果CREATE DATABASE开始时存在任何其他连接，该命令就会失败；在复制操作期间，也会阻止到源数据库的新连接。

对每个数据库，pg_database中都有两个有用的标志列：datistemplate和datallowconn。datistemplate可设置为表明某个数据库将作为CREATE DATABASE的模板。如果设置了该标志，任何拥有CREATEDB权限的用户都可以克隆该数据库；如果未设置，则只有超级用户和数据库拥有者才能克隆它。如果datallowconn为假，则不允许建立到该数据库的新连接（但仅仅把该标志设为假并不会终止已有会话）。为了防止被修改，template0数据库通常被标记为datallowconn = false。template0和template1始终都应标记为datistemplate = true。

> **注意**
>
> template1和template0除了template1这个名字是CREATE DATABASE默认源数据库名这一点之外，并无任何特殊地位。例如，可以删除template1，然后再从template0重新创建它，而不会有任何不良后果。如果有人不小心在template1里加入了许多杂项对象，这样做也许是可取的。（要删除template1，必须先令其具有pg_database.datistemplate = false。）
>
> 初始化数据库集簇时也会创建postgres数据库。这个数据库被用作用户和应用连接时的默认数据库。它只是template1的一个副本，因此在必要时也可以删除并重建。

21.4. 数据库配置

正如第 18 章所述，PostgreSQL服务器提供了大量运行时配置变量。你可以为其中许多设置指定数据库特定的默认值。

例如，如果你因为某种原因想在某个数据库中禁用GEQO优化器，通常要么必须对所有数据库都禁用它，要么必须确保每个连入的客户端都会执行SET geqo TO off。若要让这个设置在某个特定数据库中成为默认值，可以执行如下命令：

```
ALTER DATABASE mydb SET geqo TO off;
```

这会保存该设置（但不会立即生效）。以后连接到该数据库时，它看起来就像是在会话开始前刚执行过SET geqo TO off;一样。注意，用户仍然可以在自己的会话中更改该设置；它只会作为默认值。要撤销这样的设置，可使用ALTER DATABASE *dbname* RESET *varname*。

21.5. 删除数据库

数据库使用DROP DATABASE命令删除：

```
DROP DATABASE name;
```

只有数据库拥有者或超级用户可以删除数据库。删除数据库会移除其中包含的全部对象。数据库的删除无法撤销。

当你连接到目标数据库时，不能执行DROP DATABASE命令。不过，你可以连接到任何其他数据库，包括template1数据库。若要删除某个集簇中的最后一个用户数据库，template1将是唯一的选择。

为了方便，还有一个 shell 程序可用于删除数据库，即dropdb。

```
dropdb dbname
```

（与createdb不同，默认操作并不是删除与当前用户名同名的数据库。）

21.6. 表空间

PostgreSQL中的表空间允许数据库管理员在文件系统中定义存放表示数据库对象的文件的位置。表空间一旦创建，在创建数据库对象时就可以通过名称引用它。

通过使用表空间，管理员可以控制PostgreSQL安装的磁盘布局。这至少有两方面的用途。首先，如果初始化集簇所在的分区或卷空间耗尽且无法扩展，可以在另一个分区上创建表空间并使用它，直到系统能够重新配置。

其次，表空间允许管理员利用对数据库对象使用模式的了解来优化性能。例如，一个使用非常频繁的索引可以放在速度非常快且高可用的磁盘上，例如昂贵的固态设备。同时，一个存放归档数据且很少使用或对性能并不关键的表，则可以放在成本更低、速度更慢的磁盘系统上。

要定义表空间，请使用CREATE TABLESPACE命令，例如：

```
CREATE TABLESPACE fastspace LOCATION '/mnt/sda1/postgresql/data';
```

该位置必须是一个已存在的、属于PostgreSQL操作系统用户的空目录。随后在该表空间中创建的所有对象都将存储在这个目录下的文件中。

> **注意**
>
> 通常来说，在每个逻辑文件系统上创建多个表空间意义不大，因为你无法控制单个文件在一个逻辑文件系统中的具体位置。不过，PostgreSQL并不强制这种限制，而且它实际上并不能直接感知你的系统中文件系统的边界。它只是在你指定的目录中存储文件。

表空间本身必须由数据库超级用户创建，但之后你可以允许普通数据库用户使用它。为此，需要向他们授予其上的CREATE权限。

表、索引以及整个数据库都可以被分配到特定的表空间。为此，对给定表空间拥有CREATE权限的用户必须把表空间名作为参数传给相应命令。例如，下列命令会在表空间space1中创建一个表：

```
CREATE TABLE foo(i int) TABLESPACE space1;
```

另一种方式是使用default_tablespace参数：

```
SET default_tablespace = space1;
CREATE TABLE foo(i int);
```

当default_tablespace被设置为非空字符串时，它会为未显式指定TABLESPACE子句的CREATE TABLE和CREATE INDEX命令隐含地补上该子句。

还有一个temp_tablespaces参数，用来决定临时表和索引以及为大数据集排序等用途所使用的临时文件应放置在哪里。它可以是一个表空间名列表，而不只是单个表空间，这样与临时对象相关的负载就可以分散到多个表空间上。每次创建临时对象时，都会从该列表中随机挑选一个成员。

与数据库关联的表空间用于存储该数据库的系统目录。此外，如果没有给出TABLESPACE子句，并且default_tablespace或temp_tablespaces（视情况而定）也没有指定其他选择，那么它还是在该数据库中创建的表、索引和临时文件所使用的默认表空间。如果创建数据库时没有为其指定表空间，它将使用其所复制模板数据库的同一个表空间。

初始化数据库集簇时会自动创建两个表空间。pg_global表空间只用于共享系统目录。pg_default表空间是template1和template0数据库的默认表空间（因此，除非被TABLESPACE子句覆盖，而该子句可在CREATE DATABASE中指定，否则它也会是其他数据库的默认表空间）。

表空间一旦创建，只要请求的用户具有足够权限，就可以从任何数据库中使用它。这意味着，只有当使用该表空间的所有数据库中的所有对象都已被移除之后，才能删除该表空间。

要移除一个空表空间，使用DROP TABLESPACE命令。

要确定现有表空间的集合，可以检查pg_tablespace系统目录，例如

```
SELECT spcname FROM pg_tablespace;
```

psql程序的\db元命令也可用于列出现有的表空间。

PostgreSQL使用符号链接来简化表空间的实现。这意味着表空间只能在支持符号链接的系统上使用。

目录$PGDATA/pg_tblspc包含指向集簇中定义的每个非内建表空间的符号链接。尽管不推荐，但可以通过重新定义这些链接来手工调整表空间布局。两个警告：不要在服务器运行期间这样做；并且在重启服务器之后，要用新位置更新pg_tablespace目录。（如果不这样做，pg_dump将继续输出旧的表空间位置。）

第 22 章 本地化

本章从管理员的角度介绍可用的本地化特性。PostgreSQL支持两类本地化机制：

- 利用操作系统的区域设置（locale）特性，提供与区域设置相关的排序顺序、数字格式、翻译后的消息以及其他方面的支持。详见第 22.1 节和第 22.2 节。
- 提供多种字符集，以支持存储各种语言的文本，并提供客户端与服务器之间的字符集转换。详见第 22.3 节。

22.1. 区域设置支持

区域设置支持是指应用程序在字母表、排序、数字格式等方面尊重文化偏好。PostgreSQL使用服务器操作系统提供的标准 ISO C 和POSIX区域设置机制。更多信息请参考系统文档。

22.1.1. 概述

在使用`initdb`创建数据库集簇时，区域设置支持会自动初始化。默认情况下，`initdb`会使用其执行环境中的区域设置来初始化数据库集簇，因此如果你的系统已经设置为数据库集簇想要使用的区域设置，就无需额外操作。如果想使用不同的区域设置（或者不确定系统当前设置了哪种区域设置），可以通过指定`--locale`选项，精确告诉`initdb`使用哪种区域设置。例如：

```
initdb --locale=sv_SE
```

这个 Unix 系统上的示例把区域设置为在瑞典（`SE`）使用的瑞典语（`sv`）。其他可能的值包括`en_US`（美国英语）和`fr_CA`（加拿大法语）。如果一个区域设置可以使用多于一种字符集，则说明可以采用 *language_territory.codeset*这样的形式。例如，`fr_BE.UTF-8`表示在比利时（BE）使用、采用 UTF-8字符集编码的法语（fr）。

系统上有哪些区域设置可用，以及它们以什么名称出现，取决于操作系统供应商提供了什么以及实际安装了什么。在大多数 Unix 系统上，命令`locale -a`会给出所有可用区域设置的列表。Windows 使用更冗长的区域设置名称，例如`German_Germany`或`Swedish_Sweden.1252`，但原则是相同的。

有时混合多个区域设置的规则会很有用，例如使用英语排序规则但使用西班牙语消息。为此，存在一组只控制本地化规则某些方面的区域设置子类别：

`LC_COLLATE`	字符串排序顺序
`LC_CTYPE`	字符分类（什么算字母？它的大写等价形式是什么？）
`LC_MESSAGES`	消息的语言
`LC_MONETARY`	货币数量使用的格式
`LC_NUMERIC`	数字的格式
`LC_TIME`	日期和时间的格式

这些类别名会转换成`initdb`选项名，用于覆盖特定类别的区域设置选择。例如，要将区域设置设为加拿大法语，但对货币格式采用美国规则，可以使用`initdb --locale=fr_CA --lc-monetary=en_US`。

如果希望系统表现得像没有区域设置支持一样，可使用特殊区域设置名`C`，或等效的`POSIX`。

某些区域设置类别的值必须在创建数据库时固定下来。你可以为不同数据库使用不同设置，但数据库一旦创建，这些类别的值就不能再更改。LC_COLLATE和LC_CTYPE就属于这类类别。它们会影响索引的排序顺序，因此必须保持固定，否则文本列上的索引会损坏。（不过，可以通过排序规则缓解这一限制，参见第 22.2 节。）这些类别的默认值在运行initdb 时确定，创建新数据库时会使用这些值，除非在CREATE DATABASE 命令中另行指定。

其他区域设置类别可以在任何时候通过设置与其同名的服务器配置参数来更改（详见第 18.11.2 节）。initdb 选择的值实际上只是写入配置文件postgresql.conf中，作为服务器启动时的默认值。如果从postgresql.conf中删除这些赋值，服务器就会从其执行环境继承相应设置。

请注意，服务器的区域设置行为由服务器看到的环境变量决定，而不是由任何客户端的环境决定。因此，务必在启动服务器前配置好正确的区域设置。由此带来的一个结果是，如果客户端和服务器使用不同的区域设置，消息可能会因其来源不同而显示为不同语言。

注意

这里所说的从执行环境继承区域设置，在大多数操作系统上是指：对于某个给定的区域设置类别，例如排序规则，会按以下顺序检查环境变量，直到找到一个已设置的变量为止：LC_ALL、LC_COLLATE（或相应类别对应的变量）、LANG。如果这些环境变量都没有设置，则区域设置默认为C。

某些消息本地化库还会查看环境变量LANGUAGE，它会覆盖所有其他用于设置消息的语言的区域设置。如果有疑问，请参考操作系统文档，尤其是关于gettext的文档。

若要让消息可以翻译为用户偏好的语言，构建时必须启用NLS（configure --enable-nls）。所有其他区域设置支持都会自动内置。

22.1.2. 行为

区域设置会影响下列 SQL 特性：

- 在对文本数据使用ORDER BY或标准比较操作符的查询中的排序顺序
- upper、lower和 initcap函数
- 模式匹配操作符（LIKE、SIMILAR TO 以及 POSIX 风格正则表达式）；区域设置既会影响大小写不敏感匹配，也会影响字符类正则表达式中的字符分类
- to_char函数族
- 对LIKE子句使用索引的能力

在PostgreSQL中使用C或 POSIX之外的区域设置，缺点是会影响性能。它会减慢字符处理速度，并阻止普通索引用于LIKE。因此，只有在确有需要时才应使用区域设置。

为了让PostgreSQL在非 C 区域设置下也能对 LIKE子句使用索引，存在若干自定义操作符类可供使用。它们允许创建执行严格逐字符比较、忽略区域设置比较规则的索引。详见第 11.9 节。另一种方法是创建使用C 排序规则的索引，如第 22.2 节所述。

22.1.3. 问题

如果区域设置支持没有像上面说明的那样工作，
请检查操作系统中的区域设置支持是否已经正确配置。要检查系统中安装了哪些区域设置，
可以使用 `locale -a`命令（如果操作系统提供该命令）。

请确认PostgreSQL实际使用的区域设置就是你预期的区域设置。`LC_COLLATE`和`LC_CTYPE`
是在创建数据库时确定的，除非重新创建数据库，否则不能更改。其他区域设置，包括 `LC_MESSAGES`和`LC_MONETARY`，最初由启动服务器时的环境决定，但可以在运行中动态修改。
你可以使用`SHOW` 命令检查当前生效的区域设置。

源码发布包中的`src/test/locale`目录包含 PostgreSQL区域设置支持的测试套件。

那些通过解析错误消息文本来处理服务器端错误的客户端应用，显然会遇到问题，
因为服务器消息可能使用不同语言。这类应用的作者应改用错误代码机制。

维护消息翻译目录需要许多志愿者持续投入，他们希望 PostgreSQL能够良好地支持他们偏好的语言。如果你所用语言的消息目前还不可用，或尚未完全翻译，我们将非常感谢你的协助。
如果你愿意帮忙，请参考第 48 章，或向开发者邮件列表写信。

22.2. 排序规则支持

排序规则特性允许为每一列甚至每一次操作指定数据的排序顺序和字符分类行为。
这缓解了数据库的`LC_COLLATE`和`LC_CTYPE` 设置在创建之后无法更改这一限制。

22.2.1. 概念

从概念上讲，每个支持排序规则的数据类型的表达式都有一个排序规则。
（内置的支持排序规则的数据类型包括`text`、`varchar`和`char`。
用户定义的基础类型也可以标记为支持排序规则，当然，
建立在支持排序规则的数据类型之上的域也支持排序规则。）如果表达式是列引用，
则该表达式的排序规则就是该列定义的排序规则。如果表达式是常量，
则其排序规则就是该常量数据类型的默认排序规则。更复杂表达式的排序规则则按下文所述，
从其输入表达式的排序规则推导出来。

表达式的排序规则可以是“默认”排序规则，这表示为数据库定义的区域设置。
表达式的排序规则也可能是不确定的。在这种情况下，
排序操作以及其他需要知道排序规则的操作都会失败。

当数据库系统必须执行排序或字符分类时，它会使用输入表达式的排序规则。这例如发生在
`ORDER BY`子句以及函数或操作符调用（如 `<`）中。应用于`ORDER BY`子句的排序规则就是排序键的排序规则。应用于函数或操作符调用的排序规则则按下文所述，从参数中推导出来。
除了比较操作符之外，在大小写之间进行转换的函数（例如`lower`、`upper`和 `initcap`）、模式匹配操作符，以及`to_char` 及相关函数，也都会考虑排序规则。

对于函数或操作符调用，通过检查参数排序规则推导出的排序规则，
会在运行时用于执行指定操作。如果该函数或操作符调用的结果属于支持排序规则的数据类型，
那么在解析时它也会被用作该函数或操作符表达式的已定义排序规则，
以便在外围表达式需要知道其排序规则时使用。

表达式的排序规则派生可以是显式的，也可以是隐式的。
这一区别会影响当一个表达式中出现多个不同排序规则时，系统如何把它们组合起来。使用
`COLLATE`子句时，会发生显式排序规则派生；其他所有排序规则派生都是隐式的。
当需要组合多个排序规则时，例如在函数调用中，将使用以下规则：

1. 如果任一输入表达式具有显式排序规则派生，
   那么输入表达式中所有显式派生的排序规则都必须相同，否则会报错。
   如果存在显式派生的排序规则，那么排序规则组合的结果就是该排序规则。

2. 否则，所有输入表达式都必须具有相同的隐式排序规则派生，或者为默认排序规则。
   如果存在任何非默认排序规则，那么排序规则组合的结果就是该排序规则；否则，
   结果就是默认排序规则。

3. 如果输入表达式之间存在相互冲突的非默认隐式排序规则，
   则该组合会被视为具有不确定排序规则。
   除非被调用的特定函数确实需要知道它应当使用哪个排序规则，否则这并不是错误。
   如果它确实需要，就会在运行时抛出错误。

例如，考虑如下表定义：

```
CREATE TABLE test1 (
    a text COLLATE "de_DE",
    b text COLLATE "es_ES",
    ...
);
```

那么在

```
SELECT a < 'foo' FROM test1;
```

中，<比较会按de_DE规则进行，因为该表达式组合了一个隐式派生的排序规则和默认排序规则。但在

```
SELECT a < ('foo' COLLATE "fr_FR") FROM test1;
```

中，比较会按fr_FR规则进行，因为显式排序规则派生覆盖了隐式派生。进一步，给定

```
SELECT a < b FROM test1;
```

解析器无法确定应当应用哪个排序规则，因为a列和b列具有冲突的隐式排序规则。由于 <操作符确实需要知道要使用哪个排序规则，因此这会导致一个错误。
该错误可以通过给任一输入表达式附加显式排序规则说明符来解决，例如：

```
SELECT a < b COLLATE "de_DE" FROM test1;
```

或者等价地：

```
SELECT a COLLATE "de_DE" < b FROM test1;
```

另一方面，结构相似的情况

```
SELECT a || b FROM test1;
```

不会导致错误，因为||操作符不关心排序规则：无论使用什么排序规则，其结果都相同。

如果函数或操作符返回的是支持排序规则的数据类型，
那么分配给该函数或操作符组合输入表达式的排序规则，也被认为适用于其结果。因此，在

```
SELECT * FROM test1 ORDER BY a || 'foo';
```

中，排序将按de_DE规则进行。但这个查询：

```
SELECT * FROM test1 ORDER BY a || b;
```

会报错，因为即使||操作符本身不需要知道排序规则，ORDER BY子句仍然需要。与之前一样，可以通过显式排序规则说明符来解决冲突：

```
SELECT * FROM test1 ORDER BY a || b COLLATE "fr_FR";
```

22.2.2. 管理排序规则

排序规则是一个 SQL 模式对象，它把某个 SQL 名称映射到操作系统的区域设置。具体来说，它对应于LC_COLLATE和LC_CTYPE 的一种组合。（顾名思义，排序规则的主要用途是设置控制排序顺序的 LC_COLLATE。但在实践中，很少有必要让 LC_CTYPE和LC_COLLATE不同，因此把二者归为一个概念，比再建立一套为每个表达式设置LC_CTYPE 的机制更方便。）此外，排序规则还与某种字符集编码绑定（见第 22.3 节）。同一个排序规则名称可能会在不同编码中出现。

所有平台都提供名为 default、C 和 POSIX 的排序规则。根据操作系统支持情况，还可能提供其他排序规则。default 排序规则选择创建数据库时指定的 LC_COLLATE 和 LC_CTYPE 值。C 和 POSIX 排序规则都采用“传统 C”行为，仅将 ASCII 字母“A”到“Z”视为字母，并严格按字符编码的字节值排序。

如果操作系统支持在单个程序中使用多个区域设置（newlocale 及相关函数），那么在初始化数据库集簇时，initdb会根据当时在操作系统中发现的所有区域设置，用排序规则填充系统目录pg_collation。例如，操作系统可能提供一个名为de_DE.utf8的区域设置。initdb随后会为UTF8编码创建一个名为de_DE.utf8的排序规则，其 LC_COLLATE和LC_CTYPE都设置为 de_DE.utf8。它还会再创建一个从名称中去掉 .utf8标签的排序规则。因此，你也可以使用 de_DE这个名称来使用该排序规则，这样写起来更方便，且名称与编码的耦合更小。不过请注意，初始排序规则名称集合仍然取决于平台。

如果需要一个LC_COLLATE和LC_CTYPE取值不同的排序规则，可以使用CREATE COLLATION命令创建新的排序规则。该命令也可以用来从一个已有排序规则创建新的排序规则，这有助于在应用中使用与操作系统无关的排序规则名称。

在任何特定数据库中，只有使用该数据库编码的排序规则才有意义。pg_collation中的其他条目会被忽略。因此，像 de_DE这样去掉编码后缀的排序规则名称，在某个给定数据库内可以视为唯一，即使它在全局范围内并不唯一。建议使用这种去掉后缀的排序规则名称，因为如果你决定改用另一种数据库编码，需要改动的地方会更少。不过要注意，default、C和POSIX 排序规则不受数据库编码影响，始终都可以使用。

PostgreSQL即使面对具有相同属性的不同排序规则对象，也会把它们视为不兼容。例如：

```
SELECT a COLLATE "C" < b COLLATE "POSIX" FROM test1;
```

即使C和POSIX排序规则的行为完全相同，这仍然会报错。因此，不建议混用去掉后缀和保留后缀的排序规则名称。

22.3. 字符集支持

PostgreSQL中的字符集支持允许你以多种字符集（也称为编码）存储文本，包括 ISO 8859 系列之类的单字节字符集，以及 EUC（扩展 Unix 编码）、UTF-8 和 Mule 内部编码等多字节字符集。所有受支持的字符集都可以被客户端透明地使用，但其中少数不支持在服务器内部使用（即不能作为服务器端编码）。默认字符集是在使用initdb初始化PostgreSQL 数据库集簇时选定的。创建数据库时可以覆盖该设置，因此你可以拥有多个数据库，并让每个数据库使用不同的字符集。

不过，一项重要限制是，每个数据库的字符集必须与数据库的 LC_CTYPE（字符分类）和 LC_COLLATE（字符串排序顺序）区域设置兼容。对于 C 或 POSIX 区域设置，允许使用任何字符集；

对于其他区域设置，只有一种字符集能够正确工作。（但在 Windows 上，UTF-8 编码可以与任何区域设置一起使用。）

22.3.1. 支持的字符集

表 22.1显示了PostgreSQL中可用的字符集。

表 22.1. PostgreSQL 字符集

名称	描述	语言	是否服务器端?	字节数/字符	别名
BIG5	Big Five	繁体中文	否	1-2	WIN950, Windows950
EUC_CN	扩展 Unix 编码-CN	简体中文	是	1-3	
EUC_JP	扩展 Unix 编码-JP	日语	是	1-3	
EUC_JIS_2004	扩展 Unix 编码-JP，JIS X 0213	日语	是	1-3	
EUC_KR	扩展 Unix 编码-KR	韩语	是	1-3	
EUC_TW	扩展 Unix 编码-TW	繁体中文、台湾语	是	1-3	
GB18030	国家标准	中文	否	1-4	
GBK	扩展国家标准	简体中文	否	1-2	WIN936, Windows936
ISO_8859_5	ISO 8859-5, ECMA 113	拉丁语/西里尔语	是	1	
ISO_8859_6	ISO 8859-6, ECMA 114	拉丁语/阿拉伯语	是	1	
ISO_8859_7	ISO 8859-7, ECMA 118	拉丁语/希腊语	是	1	
ISO_8859_8	ISO 8859-8, ECMA 121	拉丁语/希伯来语	是	1	
JOHAB	JOHAB	韩语	否	1-3	
KOI8R	KOI8-R	西里尔语（俄语）	是	1	KOI8
KOI8U	KOI8-U	西里尔语（乌克兰语）	是	1	
LATIN1	ISO 8859-1, ECMA 94	西欧	是	1	ISO88591
LATIN2	ISO 8859-2, ECMA 94	中欧	是	1	ISO88592
LATIN3	ISO 8859-3, ECMA 94	南欧	是	1	ISO88593
LATIN4	ISO 8859-4, ECMA 94	北欧	是	1	ISO88594
LATIN5	ISO 8859-9, ECMA 128	土耳其语	是	1	ISO88599
LATIN6	ISO 8859-10, ECMA 144	北欧	是	1	ISO885910

名称	描述	语言	是否服务器端?	字节数/字符	别名
LATIN7	ISO 8859-13	波罗的海	是	1	ISO885913
LATIN8	ISO 8859-14	凯尔特语	是	1	ISO885914
LATIN9	ISO 8859-15	带欧元符号和重音字符的LATIN1	是	1	ISO885915
LATIN10	ISO 8859-16, ASRO SR 14111	罗马尼亚语	是	1	ISO885916
MULE_INTERNAL	Mule 内部编码	多语种 Emacs	是	1-4	
SJIS	Shift JIS	日语	否	1-2	Mskanji, ShiftJIS ,WIN932, Windows932
SHIFT_JIS_2004	Shift JIS, JIS X 0213	日语	否	1-2	
SQL_ASCII	未指定（见正文）	任意	是	1	
UHC	统一韩语编码	韩语	否	1-2	WIN949, Windows949
UTF8	Unicode，8 位	所有	是	1-4	Unicode
WIN866	Windows CP866	西里尔语	是	1	ALT
WIN874	Windows CP874	泰语	是	1	
WIN1250	Windows CP1250	中欧	是	1	
WIN1251	Windows CP1251	西里尔语	是	1	WIN
WIN1252	Windows CP1252	西欧	是	1	
WIN1253	Windows CP1253	希腊语	是	1	
WIN1254	Windows CP1254	土耳其语	是	1	
WIN1255	Windows CP1255	希伯来语	是	1	
WIN1256	Windows CP1256	阿拉伯语	是	1	
WIN1257	Windows CP1257	波罗的海	是	1	
WIN1258	Windows CP1258	越南语	是	1	ABC, TCVN, TCVN5712, VSCII

并非所有客户端API都支持上表中的全部字符集。例如，PostgreSQL JDBC 驱动就不支持 MULE_INTERNAL、LATIN6、LATIN8和LATIN10。

SQL_ASCII 设置的行为与其他设置有很大不同。当服务器字符集为 SQL_ASCII 时，服务器按 ASCII 标准解释字节值 0-127，而将字节值 128-255 视为不作解释的字符。设置为 SQL_ASCII

时，不会进行任何编码转换。因此，这一设置与其说是声明使用某种特定编码，不如说是声明不关心编码。在大多数情况下，如果需要处理任何非 ASCII 数据，使用 `SQL_ASCII` 都是不明智的，因为 PostgreSQL 将无法帮助转换或验证非 ASCII 字符。

22.3.2. 设置字符集

`initdb`为PostgreSQL集簇定义默认字符集（编码）。例如：

```
initdb -E EUC_JP
```

这会把默认字符集设为`EUC_JP`（日语的扩展 Unix 编码）。如果你喜欢更长的选项形式，也可以用`--encoding`代替`-E`。如果既没有给出`-E`也没有给出 `--encoding`，`initdb`会根据指定的或默认的区域设置，尝试确定应使用的合适编码。

你可以在创建数据库时指定非默认编码，只要该编码与所选区域设置兼容：

```
createdb -E EUC_KR -T template0 --lc-collate=ko_KR.euckr --lc-ctype=
ko_KR.euckr korean
```

这将创建一个名为`korean`的数据库，它使用 `EUC_KR`字符集和`ko_KR`区域设置。另一种实现方式是使用以下 SQL 命令：

```
CREATE DATABASE korean WITH ENCODING 'EUC_KR' LC_COLLATE='ko_KR.euckr'
 LC_CTYPE='ko_KR.euckr' TEMPLATE=template0;
```

请注意，上述命令指定复制`template0`数据库。如果复制任何其他数据库，
就不能更改源数据库的编码和区域设置，因为这可能导致数据损坏。详见第 21.3 节。

数据库编码存储在系统目录`pg_database`中。可以通过 `psql`的`-l`选项或`\l` 命令查看它。

```
$ psql -l
                                      List of databases
   Name    |  Owner   | Encoding  |  Collation  |    Ctype    |
   Access Privileges
-----------+----------+-----------+-------------+-------------+-------
-----------------------------
 clocaledb | hlinnaka | SQL_ASCII | C           | C           |
 englishdb | hlinnaka | UTF8      | en_GB.UTF8  | en_GB.UTF8  |
 japanese  | hlinnaka | UTF8      | ja_JP.UTF8  | ja_JP.UTF8  |
 korean    | hlinnaka | EUC_KR    | ko_KR.euckr | ko_KR.euckr |
 postgres  | hlinnaka | UTF8      | fi_FI.UTF8  | fi_FI.UTF8  |
 template0 | hlinnaka | UTF8      | fi_FI.UTF8  | fi_FI.UTF8  | {=c/
hlinnaka,hlinnaka=CTc/hlinnaka}
 template1 | hlinnaka | UTF8      | fi_FI.UTF8  | fi_FI.UTF8  | {=c/
hlinnaka,hlinnaka=CTc/hlinnaka}
(7 rows)
```

重要

在大多数现代操作系统上，PostgreSQL能够判断`LC_CTYPE`设置所隐含的字符集，
并强制只允许使用与之匹配的数据库编码。在较老的系统上，
你需要自行确保所用编码正是所选区域设置期望的编码。
这里的错误很可能会导致区域设置相关操作（例如排序）出现奇怪的行为。

> 即使LC_CTYPE不是C或 POSIX，PostgreSQL 仍允许超级用户创建使用SQL_ASCII编码的数据库。正如前文所述，SQL_ASCII并不强制数据库中存储的数据具有任何特定编码，因此这种选择存在区域设置相关误行为的风险。这种设置组合已经被弃用，未来某一天可能会被完全禁止。

22.3.3. 服务器和客户端之间的自动字符集转换

PostgreSQL 支持在服务器和客户端之间，对某些字符集组合进行自动字符集转换。转换信息保存在系统目录 pg_conversion 中。PostgreSQL 提供了一些预定义转换，如表 22.2所示。可以使用 SQL 命令 CREATE CONVERSION 创建新的转换。

表 22.2. 客户端/服务器字符集转换

服务器字符集	可用的客户端字符集
BIG5	不支持作为服务器编码
EUC_CN	EUC_CN, MULE_INTERNAL, UTF8
EUC_JP	EUC_JP, MULE_INTERNAL, SJIS, UTF8
EUC_KR	EUC_KR, MULE_INTERNAL, UTF8
EUC_TW	EUC_TW, BIG5, MULE_INTERNAL, UTF8
GB18030	不支持作为服务器编码
GBK	不支持作为服务器编码
ISO_8859_5	ISO_8859_5, KOI8R, MULE_INTERNAL, UTF8, WIN866, WIN1251
ISO_8859_6	ISO_8859_6, UTF8
ISO_8859_7	ISO_8859_7, UTF8
ISO_8859_8	ISO_8859_8, UTF8
JOHAB	JOHAB, UTF8
KOI8R	KOI8R, ISO_8859_5, MULE_INTERNAL, UTF8, WIN866, WIN1251
KOI8U	KOI8U, UTF8
LATIN1	LATIN1, MULE_INTERNAL, UTF8
LATIN2	LATIN2, MULE_INTERNAL, UTF8, WIN1250
LATIN3	LATIN3, MULE_INTERNAL, UTF8
LATIN4	LATIN4, MULE_INTERNAL, UTF8
LATIN5	LATIN5, UTF8
LATIN6	LATIN6, UTF8
LATIN7	LATIN7, UTF8
LATIN8	LATIN8, UTF8
LATIN9	LATIN9, UTF8
LATIN10	LATIN10, UTF8
MULE_INTERNAL	MULE_INTERNAL, BIG5, EUC_CN, EUC_JP, EUC_KR, EUC_TW, ISO_8859_5, KOI8R, LATIN1 至 LATIN4, SJIS, WIN866, WIN1250, WIN1251
SJIS	不支持作为服务器编码

服务器字符集	可用的客户端字符集
SQL_ASCII	任何（不会执行转换）
UHC	不支持作为服务器编码
UTF8	所有受支持的编码
WIN866	WIN866, ISO_8859_5, KOI8R, MULE_INTERNAL, UTF8, WIN1251
WIN874	WIN874, UTF8
WIN1250	WIN1250, LATIN2, MULE_INTERNAL, UTF8
WIN1251	WIN1251, ISO_8859_5, KOI8R, MULE_INTERNAL, UTF8, WIN866
WIN1252	WIN1252, UTF8
WIN1253	WIN1253, UTF8
WIN1254	WIN1254, UTF8
WIN1255	WIN1255, UTF8
WIN1256	WIN1256, UTF8
WIN1257	WIN1257, UTF8
WIN1258	WIN1258, UTF8

若要启用自动字符集转换，必须告诉PostgreSQL 你希望客户端使用哪种字符集（编码）。有几种方法可以做到这一点：

- 在psql中使用\encoding 命令。\encoding允许动态更改客户端编码。例如，要把编码改成SJIS，可以输入：

  ```
  \encoding SJIS
  ```

- libpq（第 31.9 节）提供了控制客户端编码的函数。
- 使用SET client_encoding TO。可以使用以下 SQL 命令设置客户端编码：

  ```
  SET CLIENT_ENCODING TO 'value';
  ```

  也可以为此使用标准 SQL 语法SET NAMES：

  ```
  SET NAMES 'value';
  ```

  要查询当前客户端编码：

  ```
  SHOW client_encoding;
  ```

  要返回到默认编码：

  ```
  RESET client_encoding;
  ```

- 使用PGCLIENTENCODING。如果在客户端环境中定义了 PGCLIENTENCODING环境变量，那么连接服务器时会自动选择该客户端编码。
  （之后仍可以用上面提到的任何其他方法覆盖它。）
- 使用配置变量client_encoding。如果设置了 client_encoding变量，那么连接服务器时会自动选择该客户端编码。（之后仍可以用上面提到的任何其他方法覆盖它。）

如果某个特定字符无法完成转换，就会报错。例如，假设服务器使用 EUC_JP而客户端使用LATIN1，此时若返回了一些在LATIN1中没有表示形式的日语字符，就会出现错误。

如果客户端字符集定义为 `SQL_ASCII`，则无论服务器字符集是什么，都会禁用编码转换。与服务器一样，除非处理的全部是 ASCII 数据，否则使用 `SQL_ASCII` 是不明智的。

22.3.4. 进一步阅读

这些都是开始了解各种编码系统的良好资料。

CJKV Information Processing: Chinese, Japanese, Korean & Vietnamese Computing

其中对`EUC_JP`、`EUC_CN`、`EUC_KR`和`EUC_TW`做了详细说明。

http://www.unicode.org/

Unicode Consortium 网站。

RFC 3629

这里定义了UTF-8（8 位 UCS/Unicode 转换格式）。

第 23 章 日常数据库维护任务

和任何数据库软件一样，PostgreSQL为了获得最佳性能，需要定期执行某些任务。这里讨论的任务是必需的，但它们本质上是重复性的，因此可以很容易地用标准工具实现自动化，例如 cron 脚本或 Windows 的 Task Scheduler。建立合适的脚本并检查其是否成功执行，是数据库管理员的职责。

一个显而易见的维护任务，是按固定计划创建数据的备份副本。没有最近的备份，在灾难（磁盘故障、火灾、误删关键表等）发生后就没有恢复的可能。PostgreSQL提供的备份和恢复机制在第 24 章中有详细讨论。

另一大类维护任务是定期对数据库进行“清理”。这一活动在第 23.1 节中讨论。与之密切相关的是更新查询规划器将会使用的统计信息，这在第 23.1.3 节中讨论。

另一项可能需要定期关注的任务是日志文件管理。这在第 23.3 节中讨论。

check_postgres[1] 可用于监控数据库健康状况并报告异常情况。check_postgres能与 Nagios 和 MRTG 集成，但也可以单独运行。

和某些其他数据库管理系统相比，PostgreSQL的维护工作量较小。不过，适当地关注这些任务，将大大有助于确保你愉快而高效地使用该系统。

23.1. 日常清理

PostgreSQL数据库需要一种称为清理的周期性维护。对于许多安装，让第 23.1.5 节所述的自动清理守护进程执行清理就足够了。为了在你的场景中获得最佳效果，你可能需要调整其中描述的自动清理参数。有些数据库管理员希望用手工管理的 `VACUUM`命令来补充甚至取代该守护进程的工作，这类命令通常由 cron 或 Task Scheduler 脚本按计划执行。要正确设置手工管理的清理，理解下面几小节讨论的问题至关重要。依赖自动清理的管理员也不妨略读这一材料，以帮助理解和调整自动清理。

23.1.1. 清理基础

PostgreSQL 的VACUUM命令必须定期处理每个表，原因如下：

1. 回收或再利用被更新或删除的行所占用的磁盘空间。
2. 更新 PostgreSQL 查询规划器使用的数据统计信息。
3. 防止由于事务 ID 回卷而丢失非常旧的数据。

针对上述不同目的，需要以不同的频率和范围执行 `VACUUM` 操作，下面几节会详细说明。

`VACUUM`有两种变体：标准的 `VACUUM` 和 `VACUUM FULL`。`VACUUM FULL`可以回收更多磁盘空间，但执行速度要慢得多。此外，标准形式的 `VACUUM` 可以与生产数据库操作并行运行。（`SELECT`、`INSERT`、`UPDATE` 以及 `DELETE` 等命令会继续正常工作，不过在表被清理时，你将不能使用 `ALTER TABLE` 等命令修改该表的定义。）`VACUUM FULL` 需要对正在处理的表持有独占锁，因此不能与对该表的其他使用并行进行。一般来说，管理员应尽量使用标准 `VACUUM` 并避免 `VACUUM FULL`。

`VACUUM`会产生大量 I/O 流量，这可能导致其他活动会话性能变差。可以调整一些配置参数来降低后台清理对性能的影响，参见第 18.4.3 节。

[1] http://bucardo.org/wiki/Check_postgres

23.1.2. 回收磁盘空间

在 PostgreSQL 中，对某一行执行 UPDATE 或 DELETE 时，不会立即移除该行的旧版本。这种方法对于获得多版本并发控制（MVCC，见第 13 章）的好处是必需的：当行版本仍可能对其他事务可见时，就不能删除它。但最终，过时或已删除的行版本将不再是任何事务关心的对象。必须回收它占用的空间，以供新行重用，从而避免磁盘空间需求无限增长。这是通过运行 VACUUM 完成的。

标准形式的 VACUUM 会移除表和索引中的死行版本，并将空间标记为可供将来重用。不过，它不会把空间归还给操作系统，除非出现一种特殊情况：表尾的一个或多个页面完全空闲，并且能够轻松获得一个排他表锁。相比之下，VACUUM FULL 会主动压实表，它会写出一个不含死空间的全新表文件版本。这会将表的大小降到最小，但可能耗时很长。在操作完成之前，它还需要额外的磁盘空间来存放表的新副本。

例行清理的通常目标，是足够频繁地执行标准 VACUUM，从而避免需要 VACUUM FULL。自动清理守护进程就是按照这种方式工作的，事实上它永远不会发出 VACUUM FULL。这种方法的思路不是让表始终保持在最小尺寸，而是让磁盘空间的使用维持在稳态：每个表占用的空间相当于其最小尺寸，再加上两次清理之间又被用掉的空间。虽然 VACUUM FULL 可以把表重新缩小到最小尺寸，并把磁盘空间归还给操作系统，但如果该表之后还会再次增长，这样做意义并不大。因此，对于维护被频繁更新的表，与其不常执行 VACUUM FULL，不如适度频繁地执行标准 VACUUM。

有些管理员喜欢自己安排清理，例如在夜间负载较低时完成全部工作。按固定时间表执行清理的难点在于，如果某个表的更新活动出现意外高峰，它可能膨胀到确实需要 VACUUM FULL 才能回收空间的程度。使用自动清理守护进程可以缓解这个问题，因为守护进程会根据更新活动动态安排清理。除非工作负载极其可预测，否则完全禁用守护进程是不明智的。一种可能的折中办法是设置守护进程参数，使其只对异常繁重的更新活动作出反应，从而防止情况失控，而在负载正常时，则期望定期调度的 VACUUM 完成大部分工作。

对于不使用自动清理的人，一个典型的做法是在低使用时段每天安排一次面向整个数据库的 VACUUM，并在必要时更频繁地清理那些更新特别频繁的表。（某些更新率极高的安装，甚至会每隔几分钟就清理一次最繁忙的表。）如果你在一个集簇中有多个数据库，不要忘记对每一个数据库都执行 VACUUM；程序vacuumdb可能会有帮助。

> **提示**
>
> 当一个表由于大规模更新或删除活动而包含大量死行版本时，普通的 VACUUM 可能并不能令人满意。如果你有这样一个表，并且需要回收它所占用的多余磁盘空间，就需要使用 VACUUM FULL，或者改用CLUSTER，又或者使用ALTER TABLE的某一种重写表变体。这些命令会重写整个表的新副本，并为其构建新的索引。所有这些选项都需要独占锁。还要注意，它们会临时额外占用大约等于该表大小的磁盘空间，因为在新表和新索引完成之前，旧的表和索引副本都不能被释放。

> **提示**
>
> 如果你有一个表，其全部内容会被定期删除，考虑使用TRUNCATE，而不是先 DELETE 再 VACUUM。TRUNCATE 会立即移除表的全部内容，而不需要后续再执行 VACUUM 或 VACUUM FULL 来回收此时未使用的磁盘空间。缺点是它会破坏严格的 MVCC 语义。

23.1.3. 更新规划器统计信息

PostgreSQL 查询规划器依赖于有关表内容的统计信息，以便为查询生成良好的计划。这些统计信息由ANALYZE命令收集，它既可以单独调用，也可以作为 `VACUUM` 的一个可选步骤执行。拥有足够准确的统计信息很重要，否则糟糕的计划选择可能会降低数据库性能。

如果启用了自动清理守护进程，它会在表内容发生足够大变化时自动发出 `ANALYZE` 命令。不过，管理员也可能更愿意依靠手工调度的 `ANALYZE` 操作，尤其是在已知某个表上的更新活动不会影响“重要”列统计信息的情况下。守护进程严格依据插入或更新的行数来安排 `ANALYZE`；它并不知道这些变化是否会导致有意义的统计变化。

就像为了回收空间而进行的清理一样，
频繁更新统计信息对于更新频繁的表比对于很少更新的表更有用。但即便是更新频繁的表，
如果数据的统计分布变化不大，也未必需要更新统计信息。
一个简单的经验法则是考虑表中各列的最小值和最大值变化了多少。例如，
一个包含行更新时间的 `timestamp` 列，会随着行的插入和更新而持续增大其最大值；这样的列可能比例如存放网站页面 URL 的列更需要频繁更新统计信息。URL 列收到更改的频率可能一样高，但其值的统计分布变化大概相对缓慢。

可以对特定表，甚至仅对表中的特定列运行 `ANALYZE`，因此如果你的应用需要，确实可以比其他统计更频繁地更新某些统计信息。然而在实践中，通常最好直接分析整个数据库，
因为这是一项很快的操作。`ANALYZE` 使用对表行的统计随机抽样，而不是读取每一行。

> **提示**
>
> 虽然按列微调 `ANALYZE` 的频率未必很有成效，但按列调整 `ANALYZE` 收集的统计信息详细程度可能是值得的。在 `WHERE` 子句中被频繁使用且数据分布高度不规则的列，可能需要比其他列更细粒度的数据直方图。参见 `ALTER TABLE SET STATISTICS`，或者使用配置参数default_statistics_target更改数据库范围的默认值。
>
> 此外，默认情况下，关于函数选择率的信息很有限。不过，
> 如果创建了使用函数调用的表达式索引，系统就会收集关于该函数的有用统计信息，
> 这可以显著改善使用该表达式索引的查询计划。

23.1.4. 防止事务 ID 回卷失败

PostgreSQL 的 MVCC 事务语义依赖于比较事务 ID（XID）的大小：如果某个行版本的插入 XID 大于当前事务的 XID，它就处于“未来”，不应对当前事务可见。但是，事务 ID 的大小有限（32 位），长期运行的集簇（超过 40 亿个事务）会遇到事务 ID 回卷：XID 计数器回到零，原本处于过去的事务突然看起来处于未来 – 这意味着它们的结果变得不可见。也就是灾难性的数据丢失。（数据其实还在，但如果无法访问，也无济于事。）为避免这种情况，必须至少每经过 20 亿个事务，就清理一次每个数据库中的每个表。

定期清理之所以能解决这一问题，是因为 PostgreSQL 保留了一个称为 `FrozenXID` 的特殊 XID。该 XID 不遵循普通 XID 的比较规则，始终被视为比每个普通 XID 都老。普通 XID 使用模 2^{32} 算术进行比较。这意味着，对每个普通 XID，都有 20 亿个 XID 比它“更老”，另有 20 亿个 XID 比它“更新”；换句话说，普通 XID 空间是一个没有端点的环。因此，用某个普通 XID 创建行版本后，无论该 XID 是多少，在接下来的 20 亿个事务中，该行版本都会被视为处于“过去”。如果经过超过 20 亿个事务后它仍然存在，就会突然被视为处于未来。为避免这种情况，旧的行版本在到达 20 亿事务这一年限之前，必须被重新赋予 XID `FrozenXID`。一旦被赋予这个特殊 XID，无论是否发生回卷，对所有普通事务而言，它们都处于“过去”。所以，这类行版本在被删除之前始终有效，无论存在多久。这种旧 XID 的重新赋值由 `VACUUM` 完成。

vacuum_freeze_min_age控制某个 XID 值需要达到多老才会被替换为 FrozenXID。该设置取较大值可以更长时间地保留事务信息，
而取较小值则会增加在表必须再次被清理之前可以流逝的事务数量。

VACUUM 通常会跳过没有死行版本的页面，但这些页面中可能仍有 XID 值很旧的行版本。要确保所有旧 XID 都已被替换为 FrozenXID，就需要对整个表进行一次扫描。vacuum_freeze_table_age控制 VACUUM 何时执行这种操作：如果该表在 vacuum_freeze_table_age 减去 vacuum_freeze_min_age 个事务内未曾被完整扫描，就会强制进行一次全表扫描。将其设为 0 会强制 VACUUM 始终扫描所有页面，实际上是忽略可见性映射。

一个表可以不清理的最长时间，是 20 亿个事务减去 VACUUM 上次扫描全表时的 vacuum_freeze_min_age 值。超过这个期限仍不清理，就可能丢失数据。为确保不会发生这种情况，对于可能包含 XID 年龄大于配置参数autovacuum_freeze_max_age所指定年龄的任何表，都会调用自动清理。（即使禁用了自动清理，也会如此。）

这意味着，如果一个表本来不会因为其他原因被清理，大约每经过 autovacuum_freeze_max_age 减去 vacuum_freeze_min_age 个事务，就会在该表上触发一次自动清理。对于那些为了回收空间而经常清理的表，这一点并不重要。然而，对于静态表（包括只接收插入、
但没有更新或删除的表），并不需要为了回收空间而清理，
因此设法尽可能拉大这些非常大的静态表上强制自动清理之间的间隔，可能会很有用。显然，
这可以通过增大 autovacuum_freeze_max_age 或减小 vacuum_freeze_min_age 来做到。

vacuum_freeze_table_age 的实际最大值是 0.95 * autovacuum_freeze_max_age；高于该值的设置会被截断到该最大值。设置一个高于 autovacuum_freeze_max_age 的值没有意义，因为无论如何，防回卷自动清理都会在那时被触发，而 0.95 的乘数是为了在那之前留出一些余地以手工运行一次 VACUUM。经验上，vacuum_freeze_table_age 应设置为略低于 autovacuum_freeze_max_age 的值，并留出足够空档，使一次常规调度的 VACUUM 或由正常删除、更新活动触发的自动清理能在这个窗口内运行。若设置得过于接近，即使该表最近已经为了回收空间而清理过，也可能仍会触发防回卷自动清理；
而较低的值则会导致更频繁的全表扫描。

增大 autovacuum_freeze_max_age（以及相应的 vacuum_freeze_table_age）唯一的缺点，是数据库集簇的 pg_clog 子目录会占用更多空间，因为必须保存追溯到 autovacuum_freeze_max_age 视界的所有事务的提交状态。每个事务的提交状态占用两位，因此，如果将 autovacuum_freeze_max_age 设为允许的最大值 20 亿，pg_clog 预计会增长到约 0.5GB。如果这与数据库总大小相比微不足道，建议将 autovacuum_freeze_max_age 设为允许的最大值。否则，应根据愿意为 pg_clog 分配的存储空间来设置。（默认值为 2 亿个事务，对应约 50MB 的 pg_clog 存储。）

减小 vacuum_freeze_min_age 的一个缺点是，它可能导致 VACUUM 做无用功：如果某个行版本随后不久就被修改（因而获得新的 XID），把表行的 XID 改为 FrozenXID 就是浪费时间。另一个缺点是，关于究竟是哪个事务插入或修改了某一行的细节会更早丢失。这些信息有时很有用，特别是在数据库故障后试图分析哪里出了问题时。由于这两个原因，
除了完全静态的表之外，不建议减小该设置。

为了跟踪数据库中最老 XID 的年龄，VACUUM 会在系统表 pg_class 和 pg_database 中存储 XID 统计信息。具体来说，某个表在 pg_class 中对应行的 relfrozenxid 列，记录了 VACUUM 上次扫描全表时所使用的冻结截止 XID。所有比该截止 XID 更老的普通 XID 都保证已在表内被替换为 FrozenXID。类似地，某个数据库在 pg_database 中对应行的 datfrozenxid 列，是出现在该数据库中的普通 XID 的下界，它就是该数据库内各表 relfrozenxid 值的最小值。查看这些信息的一种方便方式是执行如下查询：

```
SELECT c.oid::regclass as table_name,
       greatest(age(c.relfrozenxid),age(t.relfrozenxid)) as age
```

```
FROM pg_class c
LEFT JOIN pg_class t ON c.reltoastrelid = t.oid
WHERE c.relkind = 'r';

SELECT datname, age(datfrozenxid) FROM pg_database;
```

age 列度量的是从该截止 XID 到当前事务 XID 之间的事务数。

VACUUM通常只扫描自上次清理以来被修改过的页面，但只有当扫描整个表时，relfrozenxid 才会被推进。当 relfrozenxid的年龄超过vacuum_freeze_table_age个事务、使用了 VACUUM 的 FREEZE 选项，或者所有页面碰巧都需要清理以移除死行版本时，就会扫描整个表。当 VACUUM 扫描整个表时，它应把 age(relfrozenxid) 设为略高于所用 vacuum_freeze_min_age 设置的值（更高的部分等于自 VACUUM 开始以来已启动的事务数）。如果直到达到 autovacuum_freeze_max_age 之前，该表都没有执行一次全表扫描的 VACUUM，那么很快就会被强制执行一次自动清理。

如果由于某种原因，自动清理未能清除表中的旧 XID，当数据库中最老的 XID 距离回卷点只剩 1000 万个事务时，系统就会开始发出类似以下的警告：

```
WARNING:  database "mydb" must be vacuumed within 177009986
 transactions
HINT:  To avoid a database shutdown, execute a database-wide VACUUM in
 "mydb".
```

（按照提示，手动执行 VACUUM 应能解决问题；但请注意，VACUUM 必须由超级用户执行，否则无法处理系统目录，也就无法推进数据库的 datfrozenxid。）如果忽略这些警告，一旦距离回卷只剩不到 100 万个事务，系统就会关闭并拒绝启动任何新事务：

```
ERROR:  database is not accepting commands to avoid wraparound data
 loss in database "mydb"
HINT:  Stop the postmaster and use a standalone backend to VACUUM in
 "mydb".
```

预留这 100 万个事务的安全余量，是为了让管理员可以手动执行必需的 VACUUM 命令，在不丢失数据的情况下恢复。不过，系统一旦进入安全关闭模式，就不会再执行命令，因此唯一的办法是停止服务器，并使用一个单用户后端执行 VACUUM。单用户后端不会强制进入关闭模式。有关使用单用户后端的详细信息，参见postgres参考页面。

23.1.5. 自动清理守护进程

PostgreSQL 具有一个可选但强烈推荐的特性，称为自动清理，其目的是自动执行 VACUUM 和 ANALYZE 命令。启用后，自动清理会检查那些已经积累了大量插入、更新或删除元组的表。这些检查依赖于统计收集功能；因此，除非track_counts被设置为 true，否则无法使用自动清理。在默认配置下，自动清理是启用的，并且相关配置参数也设置得比较合适。

所谓“自动清理守护进程”实际上由多个进程组成。其中有一个持久运行的守护进程，称为自动清理启动器，负责为所有数据库启动自动清理工作进程。启动器会把工作分散在时间上，力图在每个数据库中每隔autovacuum_naptime秒启动一个工作进程。（因此，如果安装中有 *N* 个数据库，就会每隔 autovacuum_naptime/*N* 秒启动一个新的工作进程。）同时最多允许运行autovacuum_max_workers个工作进程。如果需要处理的数据库数量多于 autovacuum_max_workers，那么一旦第一个工作进程结束，就会处理下一个数据库。每个工作进程会检查其数据库中的每个表，并根据需要执行 VACUUM 和/或 ANALYZE。

如果多个大型表在很短时间内都变得需要清理，那么所有自动清理工作进程都可能会长期忙于清理这些表。

这会导致其他表和数据库在某个工作进程空闲下来之前都得不到清理。
单个数据库中的工作进程数量没有上限，
但工作进程会尽量避免重复其他工作进程已经完成的工作。注意，
正在运行的工作进程数量不计入max_connections或superuser_reserved_connections的限制。

凡是 `relfrozenxid` 值的年龄超过autovacuum_freeze_max_age个事务的表，始终都会被清理（这也适用于通过存储参数修改了最大冻结年龄的表，见下文）。否则，如果自上次 `VACUUM` 以来失效的元组数超过“清理阈值”，就会清理该表。清理阈值定义为：

```
vacuum threshold = vacuum base threshold + vacuum scale factor *
 number of tuples
```

其中，清理基础阈值为autovacuum_vacuum_threshold，清理比例因子为autovacuum_vacuum_scale_factor，元组数为 `pg_class.reltuples`。失效元组数来自统计收集器；这是一个近似计数，由每次 `UPDATE` 和 `DELETE` 操作更新。（之所以只是近似值，是因为高负载下可能丢失部分信息。）如果表的 `relfrozenxid` 值的年龄超过 `vacuum_freeze_table_age` 个事务，就会扫描整个表以冻结旧元组并推进 `relfrozenxid`；否则，只扫描自上次清理以来被修改过的页面。

对于分析操作，也使用了一个类似的条件：其阈值定义如下：

```
analyze threshold = analyze base threshold + analyze scale factor *
 number of tuples
```

该阈值会与自上次 `ANALYZE` 以来插入、更新或删除的元组总数进行比较。

自动清理无法访问临时表。因此，应通过会话 SQL 命令执行适当的清理和分析操作。

默认阈值和尺度因子取自 `postgresql.conf`，但也可以按表覆盖它们；详情见存储参数。如果某个设置已经通过存储参数修改，则使用该值；否则使用全局设置。关于全局设置的更多细节，见第 18.10 节。

除了基础阈值和比例因子之外，还有六个可以通过存储参数为每张表设置的自动清理参数。
第一个参数 `autovacuum_enabled` 可以设为 `false`，以指示自动清理后台进程完全跳过该表。在这种情况下，只有为防止事务 ID 回卷所必需时，自动清理才会处理该表。另外两个参数 `autovacuum_vacuum_cost_delay` 和 `autovacuum_vacuum_cost_limit`，用于为基于代价的清理延迟特性设置表级值（见第 18.4.3 节）。`autovacuum_freeze_min_age`、`autovacuum_freeze_max_age` 和 `autovacuum_freeze_table_age` 分别用于设置vacuum_freeze_min_age、autovacuum_freeze_max_age和vacuum_freeze_table_age的值。

当多个工作进程同时运行时，代价延迟参数会在所有运行中的工作进程之间“平衡”，
这样无论实际运行了多少个工作进程，系统承受的总 I/O 影响都相同。不过，任何正在处理那些为每表存储参数 `autovacuum_vacuum_cost_delay` 或 `autovacuum_vacuum_cost_limit` 显式设置过值的表的工作进程，都不会被纳入这个均衡算法。

23.2. 日常重建索引

在某些情况下，值得周期性地使用REINDEX命令或一系列单独的重建步骤，来重建索引。

B-树索引中已经完全变空的索引页会被回收以供重用。不过，
仍然可能出现空间利用低效：如果某个页面上的索引键除了少数几个之外都被删除，
该页面仍会保留分配。因此，在一种使用模式下，

如果每个键范围中的大多数但不是全部键最终都被删除，就会出现糟糕的空间利用。
对于这种使用模式，建议定期重建索引。

对于非 B-树索引中的膨胀潜力，人们还没有做过充分研究。在使用任何非 B-树索引类型时，定期监控索引的物理大小是个好主意。

此外，对 B-树索引而言，新近构建的索引比已经多次更新过的索引在访问时稍快一些，因为在新建的索引中，逻辑上相邻的页面通常在物理上也彼此相邻。（这一考虑不适用于非 B-树索引。）即便只是为了提高访问速度，周期性重建索引也可能是值得的。

23.3. 日志文件维护

最好把数据库服务器的日志输出保存到某个地方，而不是简单地把它丢弃到 `/dev/null`。诊断问题时，日志输出极其宝贵。但是，日志输出往往十分庞大（尤其是在较高调试级别下），
因此你不会希望无限期保存它。你需要对日志文件进行轮转，以便周期性地启用新的日志文件，
并在合理时间后移除旧文件。

如果你只是把 `postgres` 的 stderr 重定向到一个文件，那么虽然也能得到日志输出，但截断该日志文件的唯一办法是停止并重新启动服务器。如果你是在开发环境中使用 PostgreSQL，这也许可以接受，但几乎没有生产服务器会认为这种行为可以接受。

更好的办法是把服务器的 stderr 输出发送给某种日志轮转程序。系统内置了日志轮转功能，你可以在 `postgresql.conf` 中将配置参数 `logging_collector` 设为 `true` 来使用它。该程序的控制参数见第 18.8.1 节。你还可以用这种方式以机器可读的 CSV（逗号分隔值）格式捕获日志数据。

也可以使用外部日志轮转程序，特别是已经在其他服务器软件上使用了某个此类程序时。例如，rotatelogs 工具包含在 Apache 发行版中，也可用于 PostgreSQL。为此，只需通过管道将服务器的 stderr 输出传递给所需程序。如果使用 `pg_ctl` 启动服务器，那么 stderr 已经被重定向到 stdout，因此只需一个管道命令，例如：

```
pg_ctl start | rotatelogs /var/log/pgsql_log 86400
```

管理日志输出的另一种生产级方法，是把它发送给 syslog，并让 syslog 负责文件轮转。为此，在 `postgresql.conf` 中把配置参数 `log_destination` 设为 `syslog`（如果只想记录到 syslog）。之后，每当你想强制它开始写入新的日志文件时，都可以向 syslog 守护进程发送一个 `SIGHUP` 信号。如果你想自动化日志轮转，也可以把 logrotate 配置为处理来自 syslog 的日志文件。

不过，在很多系统上，syslog 并不十分可靠，尤其是在日志消息很大时；它可能恰恰在你最需要消息的时候截断或丢弃它们。此外，在 Linux 上，syslog 会把每条消息都刷盘，导致性能较差。（你可以在 syslog 配置文件的文件名开头使用一个“-”来禁用同步。）

请注意，上面描述的所有方案都处理了按可配置间隔启动新日志文件的问题，
但它们并不处理删除那些已经没有用处的旧日志文件。你很可能还需要设置一个批处理任务，
定期删除旧日志文件。另一种可能是把轮转程序配置为循环覆盖旧日志文件。

pgFouine[2] 是一个进行复杂日志文件分析的外部项目。check_postgres[3] 可以在日志文件中出现重要消息时向 Nagios 发出警报，也能检测许多其他异常情况。

[2] http://pgfouine.projects.postgresql.org/
[3] http://bucardo.org/wiki/Check_postgres

第 24 章 备份和恢复

与任何保存重要数据的系统一样，PostgreSQL数据库也应定期备份。虽然其过程基本简单，但清楚理解其底层技术和前提假设非常重要。

备份PostgreSQL数据主要有三种根本不同的方法：

- SQL转储
- 文件系统级备份
- 持续归档

每种方法都有各自的优缺点，下面各节将依次讨论。

24.1. SQL转储

这种转储方法的思路是生成一个包含 SQL 命令的文本文件；将其重新送回服务器后，服务器就会把数据库重建为转储时的状态。PostgreSQL为此提供了工具程序pg_dump。该命令的基本用法是：

```
pg_dump dbname > outfile
```

如你所见，pg_dump把结果写到标准输出。稍后我们会看到这一点有何用处。

pg_dump是一个普通的PostgreSQL客户端应用程序（尽管它实现得相当巧妙）。这意味着你可以从任何能够访问该数据库的远程主机执行这种备份过程。但请记住，pg_dump并不会以特殊权限运行。特别是，它必须对你想要备份的所有表具有读权限，因此在实践中你几乎总是必须以数据库超级用户身份运行它。

要指定pg_dump应连接哪个数据库服务器，请使用命令行选项`-h host`和`-p port`。默认主机是本地主机，或者由环境变量`PGHOST`指定的主机。类似地，默认端口由环境变量`PGPORT`指定；如果未设置，则采用编译时内置的默认值。（方便的是，服务器通常也会使用同样的内置默认值。）

和其他PostgreSQL客户端应用程序一样，pg_dump默认会以与当前操作系统用户名相同的数据库用户名进行连接。若要覆盖这一点，可以指定`-U`选项，或者设置环境变量`PGUSER`。请记住，pg_dump连接同样受常规客户端认证机制约束（见第 19 章）。

与后文介绍的其他备份方法相比，pg_dump的一个重要优点在于，pg_dump的输出通常可以重新装入较新版本的PostgreSQL，而文件级备份和持续归档都高度依赖具体的服务器版本。pg_dump也是在不同机器体系结构之间迁移数据库时唯一可行的方法，例如从 32 位服务器迁移到 64 位服务器。

由pg_dump创建的转储在内部是一致的，也就是说，该转储表示的是pg_dump开始运行时的数据库快照。pg_dump在工作期间不会阻塞数据库上的其他操作。（例外是那些需要独占锁的操作，例如大多数形式的`ALTER TABLE`。）

> **重要**
>
> 如果你的数据库架构依赖于 OID（例如作为外键），你必须指示pg_dump也转储 OID。为此，可以使用`-o`命令行选项。

24.1.1. 恢复转储

由 pg_dump 创建的文本文件供 psql 程序读取。恢复转储的通用命令格式是

```
psql dbname < infile
```

其中，*infile* 是由 pg_dump 命令输出的文件。数据库 *dbname* 不会由此命令创建，因此必须自行基于 `template0` 创建它，再执行 psql（例如使用 `createdb -T template0 dbname`）。psql 支持与 pg_dump 类似的选项，用于指定要连接的数据库服务器及所用的用户名。更多信息参见psql参考页面。

在恢复 SQL 转储之前，所有拥有对象或者在被转储数据库中曾被授予对象权限的用户都必须已经存在。否则，恢复将无法以原有所有权和/或权限重新创建这些对象。（有时这正是你想要的，但通常不是。）

默认情况下，psql脚本在遇到 SQL 错误后仍会继续执行。你可能希望让psql在设置了`ON_ERROR_STOP`变量的情况下运行，以改变这一行为，并让psql在发生 SQL 错误时以退出状态码 3 退出：

```
psql --set ON_ERROR_STOP=on dbname < infile
```

无论采用上述哪种方式，你最终只会得到一个部分恢复的数据库。
另一种做法是指定将整个转储作为单个事务恢复，这样恢复要么全部完成，要么全部回滚。
可以通过把`-1`或`--single-transaction`命令行选项传给psql来启用这种模式。
使用这种模式时要注意，即便是一个很小的错误，
也可能回滚一个已经运行了许多小时的恢复过程。不过，
这仍可能比在部分恢复后手工清理一个复杂数据库更可取。

pg_dump和psql能够写入或读取管道，
这使得可以把数据库直接从一台服务器转储到另一台服务器，例如：

```
pg_dump -h host1 dbname | psql -h host2 dbname
```

> **重要**
>
> pg_dump生成的转储是以`template0`为基准的。这意味着通过`template1`增加的任何语言、过程等也都会被pg_dump转储。因此，
> 恢复时如果你使用的是定制过的`template1`，就必须像上面的例子那样，从`template0`创建空数据库。

恢复完备份后，最好对每个数据库运行ANALYZE，这样查询优化器就能拥有有用的统计信息；
更多信息见第 23.1.3 节和第 23.1.5 节。关于如何高效地将大量数据装入PostgreSQL，请参见第 14.4 节。

24.1.2. 使用pg_dumpall

pg_dump一次只转储一个数据库，
而且不会转储角色或表空间的信息（因为这些是整个集簇级别的，而不是每个数据库各自的）。
为了方便地转储整个数据库集簇的全部内容，提供了pg_dumpall程序。pg_dumpall会备份给定集簇中的每个数据库，并保留角色定义和表空间定义等集簇范围的数据。
该命令的基本用法是：

```
pg_dumpall > outfile
```

生成的转储可以用psql恢复：

```
psql -f infile postgres
```

（实际上，可以指定任意一个现有数据库名作为起点，但如果要装载到一个空集簇中，通常应使用`postgres`。）恢复pg_dumpall转储时始终需要数据库超级用户权限，因为恢复角色和表空间信息必须使用该权限。如果使用了表空间，请确保转储中的表空间路径适合新的安装。

pg_dumpall的工作方式是先发出重新创建角色、表空间和空数据库的命令，然后针对每个数据库调用pg_dump。这意味着每个数据库本身都是内部一致的，但不同数据库的快照可能并不完全同步。

24.1.3. 处理大型数据库

某些操作系统对文件大小有上限，这会在创建大型pg_dump输出文件时带来问题。幸运的是，pg_dump可以把输出写到标准输出，因此你可以利用标准 Unix 工具绕过这个潜在问题。有几种可行方法：

使用压缩转储。. 你可以使用自己喜欢的压缩程序，例如gzip：

```
pg_dump dbname | gzip > filename.gz
```

恢复时：

```
gunzip -c filename.gz | psql dbname
```

或者：

```
cat filename.gz | gunzip | psql dbname
```

使用`split`。. `split`命令允许你把输出分割成文件系统可接受大小的较小文件。例如，要切分为 1 MB 的块：

```
pg_dump dbname | split -b 1m - filename
```

恢复时：

```
cat filename* | psql dbname
```

使用pg_dump的自定义转储格式。. 如果构建PostgreSQL所用的系统安装了zlib压缩库，自定义转储格式会在向输出文件写入数据时对其进行压缩。这会产生与使用`gzip`大致相同大小的转储文件，但额外的优点是其中的表可以有选择地恢复。下面的命令使用自定义转储格式转储数据库：

```
pg_dump -Fc dbname > filename
```

自定义格式的转储不是供psql执行的脚本，而必须通过pg_restore恢复，例如：

```
pg_restore -d dbname filename
```

详细信息见pg_dump和pg_restore参考页。

对于非常大的数据库，你可能需要把`split`与另外两种方法之一结合使用。

24.2. 文件系统级备份

另一种备份策略是直接复制PostgreSQL用来存储数据库数据的文件；第 17.2 节解释了这些文件的位置。你可以使用自己喜欢的任何方法进行文件系统备份，例如：

```
tar -cf backup.tar /usr/local/pgsql/data
```

不过，这种方法有两个限制，使得它并不实用，至少也不如pg_dump方法：

1. 为了得到可用的备份，数据库服务器必须关闭。
   像禁止所有连接这样的折中办法不起作用（部分原因是`tar`
   和类似工具不会对文件系统状态执行原子快照，另一部分原因是服务器内部也存在缓冲）。
   有关停止服务器的信息见第 17.5 节。不言而喻，恢复数据前也必须关闭服务器。

2. 如果你已经深入了解数据库文件系统布局的细节，
   可能会想尝试仅通过某些表或数据库各自的文件或目录来备份或恢复它们。这是行不通的，
   因为这些文件中的信息离开记录所有事务提交状态的提交日志文件`pg_clog/*`就无法使用。
   一个表文件只有和这些信息一起才可用。当然，也不可能只恢复一个表以及相关的`pg_clog`
   数据，因为那会让数据库集簇中的其他所有表都无法使用。因此，
   文件系统备份只适用于整个数据库集簇的完整备份与恢复。

另一种文件系统备份方法是对数据目录制作一个“一致快照”，
前提是文件系统支持该功能（并且你愿意相信其实现是正确的）。
典型流程是：对包含数据库的卷制作一个“冻结快照”，
然后从该快照把整个数据目录（不是其中一部分，见上文）复制到备份设备，
再释放这个冻结快照。即使数据库服务器正在运行，这种方法也能工作。不过，
以这种方式创建的备份会把数据库文件保存为一种仿佛服务器没有被正确关闭的状态；因此，
当你基于备份数据启动数据库服务器时，它会认为前一个服务器实例发生了崩溃，并重放 WAL 日志。这不是问题；只要意识到这一点即可（并且务必把 WAL 文件也包含在备份中）。你可以在获取快照之前执行一次`CHECKPOINT`，以缩短恢复时间。

如果数据库分布在多个文件系统上，就可能根本无法对所有卷获得完全同步的冻结快照。例如，
如果数据文件和 WAL 日志位于不同磁盘，或者表空间位于不同文件系统上，就可能无法使用快照备份，因为这些快照必须同时获取。在这种情况下，在信任一致快照技术之前，
一定要非常仔细地阅读文件系统文档。

如果无法获得同时的快照，一种选择是将数据库服务器关闭足够长的时间，
以建立所有冻结快照。另一种选择是执行持续归档基础备份（第 24.3.2 节），因为这种备份不受备份期间文件系统变化的影响。这种方法只需在备份期间启用持续归档；
恢复则使用持续归档恢复（第 24.3.3 节）。

另一种选择是使用rsync进行文件系统备份。做法是先在数据库服务器运行时执行一次rsync，
然后将数据库服务器只关闭足够再做一次rsync的时间。第二次rsync会比第一次快得多，
因为它需要传输的数据相对较少，而最终结果会因为服务器已经关闭而保持一致。
这种方法可以在最小停机时间下完成文件系统备份。

注意，文件系统备份通常会比 SQL 转储更大。（例如，pg_dump不需要转储索引内容，只需转储重建索引的命令。）不过，进行文件系统备份可能更快。

24.3. 持续归档和时间点恢复（PITR）

在任何时候，PostgreSQL都会维护一个预写式日志（WAL），它位于集簇数据目录的`pg_xlog/`子目录中。该日志记录对数据库数据文件所做的每一次修改。
这个日志首先是为崩溃安全而存在：如果系统崩溃，可以通过“重放”

自上次检查点以来的日志记录，将数据库恢复到一致状态。不过，
日志的存在也使第三种数据库备份策略成为可能：我们可以把文件系统级备份与 WAL 文件备份结合起来。如果需要恢复，就先恢复文件系统备份，再重放已备份的 WAL 文件，使系统回到当前状态。与前两种方法相比，这种方法管理起来更复杂，但它有一些显著优点：

- 我们不需要一个完全一致的文件系统备份作为起点。
  备份中的任何内部不一致性都会通过日志重放纠正（这与崩溃恢复期间发生的事情并没有本质不同）。因此，我们不需要文件系统快照能力，只需要tar或类似的归档工具。

- 由于我们可以把任意长的一串 WAL 文件串联起来进行重放，只要持续归档 WAL 文件，就可以实现连续备份。这对于大型数据库尤其有价值，因为此时频繁进行完整备份可能并不方便。

- 不必把 WAL 记录一直重放到最后。我们可以在任意位置停止重放，并得到数据库在当时的一致快照。因此，这项技术支持
  时间点恢复：可以把数据库恢复到自基础备份之后任意时刻的状态。

- 如果我们持续把这一串 WAL 文件输送给另一台已经装载了同一基础备份文件的机器，就得到了一个温备系统：我们随时都可以启用第二台机器，而它将拥有一个几乎是最新的数据库副本。

> **注意**
>
> pg_dump和pg_dumpall不会生成文件系统级备份，
> 因此不能作为持续归档方案的一部分使用。这类转储是逻辑备份，不包含 WAL 重放所需的足够信息。

和普通文件系统备份技术一样，这种方法只能支持整个数据库集簇的恢复，
而不支持其中某个子集的恢复。此外，它需要大量归档存储：基础备份可能很庞大，
繁忙系统也会产生许多兆字节的、必须归档的 WAL 流量。尽管如此，在很多需要高可靠性的场景中，它仍是首选的备份技术。

要想利用持续归档（许多数据库厂商也称之为“在线备份”）成功恢复，你需要一串连续的已归档 WAL 文件，其时间至少要追溯到备份的开始时刻。因此，入门时应在进行第一次基础备份之前就建立并测试好归档 WAL 文件的流程。下面先讨论归档 WAL 文件的机制。

24.3.1. 设置 WAL 归档

从抽象角度看，运行中的 PostgreSQL 系统会产生一个无限延伸的 WAL 记录序列。在物理存储上，系统将该序列分为 WAL 段文件，通常每个为 16MB（但可以在构建 PostgreSQL 时修改段大小）。段文件使用数字名称，反映其在抽象 WAL 序列中的位置。不使用 WAL 归档时，系统通常只创建少量段文件，然后通过将不再需要的段文件重命名为更大的段号来“回收”它们。系统假定，内容早于倒数第二个检查点的段文件已不再需要，可以回收。

在归档 WAL 数据时，我们需要在每个段文件写满后捕获其内容，并在该段文件被回收重用之前把数据保存到某处。根据应用场景和可用硬件的不同，“把数据保存到某处”
可以有很多不同的方法：可以把段文件复制到另一台机器上的 NFS 挂载目录，把它们写到磁带机中（确保你有办法识别每个文件的原始文件名），把它们批量打包后刻录到 CD 上，或者采用完全不同的方式。为了给数据库管理员提供灵活性，PostgreSQL尽量不对归档方式作任何假设。
相反，PostgreSQL允许管理员指定一个 shell 命令，以便把已完成的段文件复制到它应去的地方。该命令可以只是一次简单的`cp`调用，也可以调用一个复杂的 shell 脚本，全由你决定。

要启用 WAL 归档，请将wal_level配置参数设为 `archive`（或 `hot_standby`），将archive_mode设为 `on`，并在archive_command配置参数中指定要执行的 shell 命令。实际使用中，这些设置始终放在 `postgresql.conf` 文件中。对于 `archive_command`，`%p` 会被替换为要归

档文件的路径名，而 %f 只会被替换为文件名。（路径名相对于当前工作目录，即集簇的数据目录。）使用 %% 可以在命令中嵌入实际的 % 字符。最简单的可用命令类似于：

```
archive_command = 'test ! -f /mnt/server/archivedir/%f && cp %p /mnt/
server/archivedir/%f'  # Unix
archive_command = 'copy "%p" "C:\\server\\archivedir\\%f"'  # Windows
```

这会将可归档的 WAL 段复制到目录 /mnt/server/archivedir。（这只是示例，并非推荐做法，而且不一定适用于所有平台。）在替换 %p 和 %f 参数后，实际执行的命令可能如下：

```
test ! -f /mnt/server/archivedir/00000001000000A900000065 &&
 cp pg_xlog/00000001000000A900000065 /mnt/server/archivedir/
00000001000000A900000065
```

每个需要归档的新文件都会生成一条类似的命令。

归档命令将以运行 PostgreSQL 服务器的同一用户身份执行。由于归档的一系列 WAL 文件实际上包含数据库中的全部内容，应确保归档数据不会被他人窥视；例如，
将其归档到不允许所属组或其他用户读取的目录。

归档命令必须当且仅当成功时才返回退出状态零。收到零状态后，PostgreSQL 会认为该文件已成功归档，并将其删除或回收。非零状态则告诉 PostgreSQL 该文件尚未归档；它会定期重试，直到成功。

通常应将归档命令设计为拒绝覆盖任何已存在的归档文件。这是一项重要的安全保护措施，
能在管理员操作失误时（例如将两台不同服务器的输出发送到同一归档目录）
维护归档的完整性。

建议测试所拟定的归档命令，确保它确实不会覆盖已有文件，并且在这种情况下返回非零状态。
上面给出的 Unix 示例命令通过单独加入一个test步骤，确保这两点。在某些 Unix 平台上，cp提供了诸如-i之类的开关，也可以更简洁地实现同样目的，
但在你确认它会返回正确的退出状态之前，不应依赖这些开关。（尤其是 GNU cp在使用-i且目标文件已存在时会返回状态零，这不是我们想要的行为。）

在设计归档方案时，请考虑如果归档命令因为某些环节需要操作员干预，
或者归档空间耗尽而反复失败，会发生什么。例如，
如果你在没有自动换带器的情况下向磁带写入数据，那么磁带满了以后，
在更换磁带之前将无法继续归档。
你应确保任何错误情况或对人工操作员的请求都能得到适当报告，以便问题能够比较快地解决。
在问题解决之前，pg_xlog/目录会继续堆积 WAL 段文件。（如果包含pg_xlog/的文件系统被写满，PostgreSQL将执行 PANIC 关闭。不会丢失已提交的事务，但在释放出一些空间之前，数据库会一直离线。）

归档命令的速度并不重要，只要它能跟上服务器生成 WAL 数据的平均速度即可。即使归档过程稍有滞后，正常操作也会继续。如果归档明显落后，灾难发生时可能丢失的数据量就会增加。
这还意味着pg_xlog/目录会包含大量尚未归档的段文件，最终可能耗尽可用磁盘空间。
建议监控归档过程，确保它按你的预期工作。

在编写归档命令时，应假定待归档文件名最长可达 64 个字符，并且可以包含 ASCII 字母、数字和点号的任意组合。无需保留原始相对路径（%p），但必须保留文件名（%f）。

注意，虽然 WAL 归档允许你恢复对PostgreSQL数据库中数据所做的任何修改，但它不会恢复对配置文件（即postgresql.conf、pg_hba.conf和pg_ident.conf）的修改，
因为这些文件是手工编辑的，而不是通过 SQL 操作修改的。你可能希望把配置文件放在常规文件系统备份过程能够覆盖的位置。关于如何重定位配置文件，见第 18.2 节。

归档命令只会针对已完成的 WAL 段调用。因此，如果服务器产生的 WAL 流量很小（或者存在低谷期），事务完成到其被安全写入归档存储之间可能会有很长延迟。
为了限制未归档数据可能有多旧，你可以设置archive_timeout，使服务器强制切换到新 WAL 段文件的间隔不超过这个值。注意，由强制切换而提前归档的文件长度仍与装满的文件相同。
因此，把archive_timeout设得很短并不明智，这会使归档存储膨胀。archive_timeout设为大约 1 分钟通常是合理的。

此外，如果你希望确保一个刚刚完成的事务尽快被归档，可以使用pg_switch_xlog手工强制一次段切换。其他与 WAL 管理相关的实用函数列在表 9.57中。

当wal_level为minimal时，某些 SQL 命令会像第 14.4.7 节中所述那样被优化为避免 WAL 记录。如果在执行这些语句期间启用了归档或流复制，WAL 将不包含归档恢复所需的足够信息。（崩溃恢复不受影响。）因此，wal_level只能在服务器启动时更改。然而，archive_command可以通过重新加载配置文件来更改。如果你希望暂时停止归档，一种办法是将archive_command设置为空字符串（''）。这会导致 WAL 文件在pg_xlog/中累积，直到重新建立可用的archive_command。

24.3.2. 进行基础备份

制作基础备份的过程相对简单：

1. 确保已启用 WAL 归档并且工作正常。

2. 以超级用户身份连接到数据库，并执行命令：

   ```
   SELECT pg_start_backup('label');
   ```

   其中，label 是用于唯一标识此次备份操作的任意字符串。（一种好做法是使用你打算存放备份转储文件的完整路径。）pg_start_backup 会创建一个备份标签文件，名为backup_label，位于集簇目录中，包含备份信息，例如开始时间和标签字符串。

   连接到集簇中的哪个数据库来执行该命令都无关紧要。可以忽略该函数返回的结果；但如果它报告错误，应先处理后再继续。

   默认情况下，pg_start_backup 可能需要很长时间才能完成。这是因为它会执行一次检查点，而检查点所需的 I/O 会分散在较长时间内，默认是检查点间隔的一半（参见配置参数checkpoint_completion_target）。这通常是理想的行为，
   因为它尽量减少了对查询处理的影响。如果希望尽快开始备份，请使用：

   ```
   SELECT pg_start_backup('label', true);
   ```

   这会强制尽快完成检查点。

3. 使用任何方便的文件系统备份工具执行备份，例如tar或cpio（不要使用pg_dump或pg_dumpall）。在此过程中既没有必要，也不希望停止数据库的正常运行。

4. 再次以超级用户身份连接到数据库，并执行命令：

   ```
   SELECT pg_stop_backup();
   ```

   这会终止备份模式，并自动切换到下一个 WAL 段。切换的目的是让备份期间写入的最后一个 WAL 段文件准备好归档。

5. 备份期间处于活动状态的 WAL 段文件全部归档后，备份就完成了。pg_stop_backup 返回结果所标识的文件，是构成完整备份文件集所需的最后一个段。如果启用了

archive_mode，pg_stop_backup 会等到最后一个段归档后才返回。由于已经配置了 archive_command，这些文件会自动归档。大多数情况下，归档很快就能完成，但建议监控归档系统，确保没有延迟。如果归档进程因归档命令失败而落后，它会不断重试，直到归档成功、备份完成。如果希望限制 pg_stop_backup 的执行时间，可以设置合适的 statement_timeout 值。

你也可以使用pg_basebackup工具来取得备份，而不是手工复制文件。这个工具会自动完成等价于pg_start_backup()、复制和pg_stop_backup()的步骤，并使用复制协议通过常规的PostgreSQL连接传输备份，而不要求文件系统级别的访问。pg_basebackup不会干扰使用pg_start_backup()/pg_stop_backup()进行的文件系统级备份。

某些文件系统备份工具在复制过程中，如果它们试图复制的文件发生变化，就会发出警告或错误。对活动数据库进行基础备份时，这种情况是正常的，并不表示出错。不过，你需要确保能够把这类提示与真正的错误区分开来。例如，某些版本的rsync会针对“源文件消失”返回单独的退出码，你可以编写一个驱动脚本，把这一退出码视为非错误情况。此外，某些版本的 GNU tar在文件被tar复制时如果发生截断，会返回与致命错误无法区分的错误码。幸运的是，GNU tar 1.16 及之后版本在备份过程中如果文件被更改会以 1 退出，而其他错误则以 2 退出。

不必担心pg_start_backup与实际备份开始之间、或者备份结束与pg_stop_backup之间经过的时间；几分钟的延迟不会造成任何损害。（但是，如果你通常在禁用full_page_writes的情况下运行服务器，你可能会注意到pg_start_backup和pg_stop_backup之间性能的下降，因为在备份模式期间full_page_writes实际上是被强制打开的。）你必须确保这些步骤按顺序执行，没有任何可能的重叠，否则会使备份失效。

务必确认你的备份包含数据库集簇目录（例如/usr/local/pgsql/data）下的全部文件。如果你使用的表空间不位于该目录之下，也要记得把它们包括进来（并确保备份把符号链接作为链接归档，否则恢复时会破坏表空间）。

不过，备份中应省略集簇pg_xlog/子目录内的文件。这个小调整很值得，因为它能降低恢复时出错的风险。如果pg_xlog/是一个指向集簇目录外某处的符号链接，就很容易做到这一点，而出于性能原因，这本来也是一种常见配置。你也可能想排除postmaster.pid和postmaster.opts，它们记录的是正在运行的postmaster的信息，而不是最终使用此备份的postmaster的信息。（这些文件可能会让pg_ctl感到困惑。）

要让该备份可用，你需要保留在文件系统备份期间以及之后生成的所有 WAL 段文件。为帮助完成这件事，pg_stop_backup函数会创建一个备份历史文件，并立即将其存入 WAL 归档区域。该文件以文件系统备份所需的第一个 WAL 段文件命名。例如，如果起始 WAL 文件是000000010000123400005 5CD，备份历史文件的名称将类似于000000010000123400005 5CD.007C9330.backup。（文件名第二部分表示该 WAL 文件中的一个精确位置，通常可以忽略。）一旦你已经安全归档了文件系统备份以及备份期间使用到的 WAL 段文件（如备份历史文件所指定），所有名称在数值上更小的已归档 WAL 段就不再是恢复该文件系统备份所必需的，可以删除。不过，你仍应考虑保留多个备份集，以绝对确保能够恢复数据。

备份历史文件只是一个很小的文本文件。它包含你提供给pg_start_backup的标签字符串，以及备份的起止时间和起止 WAL 段。如果你用该标签标识了关联的转储文件，那么已归档的历史文件就足以告诉你应恢复哪个转储文件。

由于你必须保留自上一次基础备份以来的所有已归档 WAL 文件，基础备份之间的间隔通常应根据你愿意为已归档 WAL 文件投入多少存储空间来决定。你还应考虑在确实需要恢复时，你愿意花多长时间进行恢复，因为系统必须重放所有这些 WAL 段；如果距离上次基础备份已经过去很久，这可能会花费一些时间。

另外值得一提的是，pg_start_backup函数会在数据库集簇目录中创建一个名为backup_label的文件，该文件会被pg_stop_backup删除。
该文件当然也会作为备份转储文件的一部分被归档。备份标签文件包含你提供给pg_start_backup的标签字符串、执行pg_start_backup的时间以及起始 WAL 文件的名称。因此，在发生混淆时，可以查看备份转储文件的内容，精确确定该文件来自哪一次备份会话。

服务器停止时也可以进行备份。在这种情况下，你显然无法使用pg_start_backup或pg_stop_backup，因此只能自己追踪各个备份的身份以及相关 WAL 文件最早需要追溯到哪里。通常最好还是遵循上面的持续归档过程。

24.3.3. 使用持续归档备份进行恢复

现在，最坏的情况发生了，你需要通过备份进行恢复。步骤如下：

1. 如果服务器仍在运行，就先停止它。

2. 如果你有足够空间，请把整个集簇数据目录以及所有表空间复制到临时位置，以备后续需要。注意，这一预防措施要求系统有足够空闲空间来保存现有数据库的两份副本。如果空间不够，至少也应保存集簇pg_xlog子目录中的内容，因为其中可能包含系统停机前尚未归档的 WAL 文件。

3. 删除集簇数据目录下以及所有正在使用的表空间根目录下的现有文件和子目录。

4. 从文件系统备份中恢复数据库文件。务必确保它们以正确的所有者（数据库系统用户，而不是root！）和正确的权限恢复。如果使用了表空间，还应验证pg_tblspc/中的符号链接是否已正确恢复。

5. 删除pg_xlog/中现有的所有文件；这些文件来自文件系统备份，因此很可能已经过时而不是最新的。如果你根本没有归档pg_xlog/，那么就以正确权限重新创建它，并注意如果它原先是符号链接，就要重新把它设置成符号链接。

6. 如果你手头还有第 2 步中保存下来的未归档 WAL 段文件，请把它们复制到pg_xlog/中。（最好复制，而不是移动，这样一旦出问题需要重来时，你手里仍然保留着未修改的原始文件。）

7. 在集簇数据目录中创建恢复命令文件 recovery.conf（参见第 26 章）。也可以临时修改pg_hba.conf，在确认恢复成功之前阻止普通用户连接。

8. 启动服务器。服务器会进入恢复模式，依次读取所需的已归档 WAL 文件。如果恢复因外部错误而终止，只需重启服务器，它就会继续恢复。恢复过程完成后，服务器会将 recovery.conf 重命名为 recovery.done（防止以后意外重新进入恢复模式），然后开始正常的数据库操作。

9. 检查数据库内容，确认你已经恢复到期望状态。如果不是，就回到第 1 步。如果一切正常，就把pg_hba.conf恢复为正常设置，让用户重新连接。

整个过程的关键，是设置一个恢复配置文件，说明希望如何恢复，以及恢复到什么位置。
可以使用 recovery.conf.sample（通常位于安装目录的 share/ 目录中）作为模板。必须在 recovery.conf 中指定 restore_command，它告诉 PostgreSQL 如何获取已归档的 WAL 段文件。与 archive_command 一样，它是一个 shell 命令字符串，可以包含 %f，该标记会被替换为所需日志文件的名称；还可以包含 %p，该标记会被替换为日志文件要复制到的路径名。（路径名相对于当前工作目录，即集簇的数据目录。）使用 %% 可以在命令中嵌入实际的 % 字符。最简单的可用命令类似于：

```
restore_command = 'cp /mnt/server/archivedir/%f %p'
```

这会复制先前归档的 WAL 段，来源目录为 `/mnt/server/archivedir`。当然，也可以使用复杂得多的命令，甚至可以使用要求操作人员挂载相应磁带的 shell 脚本。

重要的是，该命令在失败时必须返回非零退出状态。系统会调用该命令来请求归档中不存在的文件；遇到这种情况时，它必须返回非零值。这不是一种错误情况。并非所有被请求的文件都是 WAL 段文件；你还应预料到对后缀为`.backup`或`.history`的文件的请求。还要注意，`%p`路径的基本名称会与`%f`不同；不要认为它们可以互换。

在归档中找不到的 WAL 段会转而在`pg_xlog/`中查找；这使得可以使用最近尚未归档的段。不过，凡是归档中可用的段，都会优先于`pg_xlog/`中的文件使用。系统在取回归档文件时不会覆盖`pg_xlog/`中的现有内容。

通常，恢复会处理完所有可用的 WAL 段，从而把数据库恢复到当前时间点（或者在可用 WAL 段所允许的情况下尽可能接近当前时间点）。因此，一次正常恢复会以一条“file not found”消息结束，具体错误文本取决于你选择的`restore_command`。在恢复开始时，你也可能看到一条针对类似`00000001.history`文件的错误消息。这同样是正常的，在简单恢复场景中并不表示有问题；相关讨论见第 24.3.4 节。

如果希望恢复到过去的某个时间点（例如初级 DBA 删除主要事务表之前），只需在 `recovery.conf` 中指定所需的停止点。这个停止点称为“恢复目标”，可以通过日期/时间、命名恢复点，或某个特定事务 ID 的完成来指定。撰写本文时，只有日期/时间和命名恢复点这两种方式比较实用，因为还没有工具能帮助你准确确定应使用哪个事务 ID。

> **注意**
>
> 停止点必须晚于基础备份的结束时间，也就是`pg_stop_backup`的结束时间。你不能用某次基础备份恢复到该备份仍在进行中的时间点。（若要恢复到这样的时间点，必须回到更早的一次基础备份，再从那里向前重放日志。）

如果恢复过程中发现了损坏的 WAL 数据，恢复会在该点停止，服务器也不会启动。在这种情况下，可以从头重新执行恢复，并指定一个位于损坏点之前的“恢复目标”，使恢复能够正常完成。如果恢复因外部原因失败，例如系统崩溃或 WAL 归档变得不可访问，那么只需重新启动恢复，它几乎会从上次失败的位置继续。恢复重启的工作方式很像正常运行时的检查点：服务器会周期性地把自身状态强制写盘，然后更新`pg_control`文件，表明已处理过的 WAL 数据无需再次扫描。

24.3.4. 时间线

能够将数据库恢复到过去某个时间点，也会带来一些类似科幻故事中时间旅行和平行宇宙的复杂情况。例如，假设在数据库原来的历史中，你在周二下午 5:15 删除了一张重要的表，直到周三中午才发现错误。你从容地取出备份，将数据库恢复到周二下午 5:14，然后重新投入运行。在数据库宇宙的这段历史中，你从未删除过那张表。但假设你后来发现这样做不太合适，希望回到原来历史中的周三上午某个时刻。如果数据库恢复运行后覆盖了通往该时刻所需的某些 WAL 段文件，就无法回去了。因此，为了避免这种情况，需要区分时间点恢复后产生的一系列 WAL 记录与数据库原来历史中产生的记录。

为解决这一问题，PostgreSQL 引入了时间线的概念。每当归档恢复完成，系统都会创建一条新时间线，用于标识此次恢复之后产生的一系列 WAL 记录。时间线 ID 是 WAL 段文件名的一部分，

因此新时间线不会覆盖旧时间线产生的 WAL 数据。实际上，可以归档许多不同的时间线。这个功能看似没什么用，却常常能救急。比如，你不能确定应该恢复到哪个时间点，需要反复尝试时间点恢复，直到找到脱离旧历史的最佳分支点。没有时间线，这个过程很快就会乱得无法管理。有了时间线，就可以恢复到任何先前的状态，包括早先已放弃的时间线分支中的状态。

每当创建一条新的时间线时，PostgreSQL都会创建一个“时间线历史”文件，记录它是从哪条时间线、在何时分叉出来的。当从包含多条时间线的归档中恢复时，这些历史文件对于系统选取正确的 WAL 段文件是必需的。因此，它们会像 WAL 段文件一样被归档到 WAL 归档区域。历史文件只是很小的文本文件，因此长期保存它们既便宜也合适（而段文件通常很大）。如果你愿意，还可以在历史文件中加入注释，记录创建这条时间线的方式和原因。当你因实验而积累出一批错综复杂的时间线时，这类注释会特别有价值。

默认情况下，恢复会沿着制作基础备份时的当前时间线进行。如果希望恢复到某条子时间线（即返回到一次恢复尝试之后产生的某个状态），需要在 `recovery.conf` 中指定目标时间线 ID。不能恢复到在基础备份之前就已分支出去的时间线。

24.3.5. 建议和示例

这里给出一些配置持续归档的建议。

24.3.5.1. 独立热备份

可以利用PostgreSQL的备份设施生成独立热备份。这些备份不能用于时间点恢复，但它们的制作和恢复通常都比pg_dump转储快得多。（它们也比pg_dump转储大得多，因此在某些情况下速度优势可能会被抵消。）

要准备独立热备份，将 `wal_level` 设为 `archive`（或 `hot_standby`），将 `archive_mode` 设为 `on`，并配置一个 `archive_command`，使其仅在开关文件存在时执行归档。例如：

```
archive_command = 'test ! -f /var/lib/pgsql/backup_in_progress || (
test ! -f /var/lib/pgsql/archive/%f && cp %p /var/lib/pgsql/archive/
%f)'
```

当 `/var/lib/pgsql/backup_in_progress` 存在时，这条命令会执行归档；否则会静默地返回零退出状态（允许 PostgreSQL 回收不需要的 WAL 文件）。

做好上述准备后，就可以使用类似下面的脚本进行备份：

```
touch /var/lib/pgsql/backup_in_progress
psql -c "select pg_start_backup('hot_backup');"
tar -cf /var/lib/pgsql/backup.tar /var/lib/pgsql/data/
psql -c "select pg_stop_backup();"
rm /var/lib/pgsql/backup_in_progress
tar -rf /var/lib/pgsql/backup.tar /var/lib/pgsql/archive/
```

首先创建开关文件 `/var/lib/pgsql/backup_in_progress`，以启用已完成 WAL 文件的归档。备份完成后删除该开关文件。随后将已归档的 WAL 文件加入备份，使基础备份与所有必需的 WAL 文件都包含在同一个 tar 文件中。请记得在备份脚本中加入错误处理。

24.3.5.2. 压缩的归档日志

如果担心归档存储的大小，可以使用 pg_compresslog（http://pglesslog.projects.postgresql.org）去除 WAL 文件中不必要的full_page_writes内容和尾随空格。然后可以使用 gzip 进一步压缩 pg_compresslog 的输出：

```
archive_command = 'pg_compresslog %p - | gzip > /var/lib/pgsql/
archive/%f'
```

恢复时则需要使用 gunzip 和 pg_decompresslog：

```
restore_command = 'gunzip < /mnt/server/archivedir/%f | pg_
decompresslog - %p'
```

24.3.5.3. archive_command脚本

很多人选择使用脚本来定义自己的archive_command，这样postgresql.conf中的配置项就会显得非常简单：

```
archive_command = 'local_backup_script.sh "%p" "%f"'
```

只要你希望在归档过程中使用不止一条命令，就建议使用单独的脚本文件。这样一来，所有复杂性都可以在脚本内部管理，而脚本可以使用诸如bash或perl之类的常见脚本语言编写。

可以在脚本中处理的需求示例包括：

- 将数据复制到安全的异地数据存储
- 把 WAL 文件成批处理，使其每三个小时传输一次，而不是每次只传输一个
- 与其他备份和恢复软件对接
- 与监控软件对接以报告错误

> **提示**
>
> 使用archive_command脚本时，最好启用logging_collector。脚本写到stderr的任何消息都会出现在数据库服务器日志中，这样一来，当复杂配置失败时就更容易诊断。

24.3.6. 注意事项

在撰写本文时，持续归档技术还存在若干局限。这些问题很可能会在未来版本中得到修复：

- 哈希索引上的操作目前不会写入 WAL 日志，因此重放不会更新这些索引。这意味着任何新插入的行都会被索引忽略，被更新的行会看起来消失了，而被删除的行仍会保留指针。换句话说，如果你修改了一张带有哈希索引的表，那么在备库服务器上会得到不正确的查询结果。恢复完成后，建议在恢复操作结束后对每个这样的索引手工执行REINDEX。
- 如果在进行基础备份时执行了CREATE DATABASE命令，而该CREATE DATABASE所复制的模板数据库又在基础备份尚未结束时被修改，那么恢复时可能会把这些修改也传播到新建数据库中。这当然并不理想。为避免这种风险，最好在进行基础备份时不要修改任何模板数据库。
- CREATE TABLESPACE命令会以字面绝对路径写入 WAL，因此重放时会按相同的绝对路径创建表空间。如果 WAL 在另一台机器上重放，这可能并不理想。即使 WAL 在同一台机器上、但重放到新的数据目录中，也可能有危险：重放仍会覆盖原表空间的内容。为避免此类潜在陷阱，最佳做法是在创建或删除表空间后重新执行一次基础备份。

还应注意，默认的WAL格式相当臃肿，因为它包含许多磁盘页面快照。
这些页面快照是为支持崩溃恢复而设计的，因为我们可能需要修复部分写入的磁盘页。
根据你的系统硬件和软件情况，部分写入的风险可能小到可以忽略；在这种情况下，可以通过full_page_writes参数关闭页面快照，从而显著减少已归档 WAL 文件的总量。（在这样做之前，请先阅读第 29 章中的说明和警告。）关闭页面快照并不妨碍把 WAL 用于 PITR 操作。未来一个可能的开发方向，是在`full_page_writes`开启的情况下，通过去除不必要的页面副本来压缩归档 WAL 数据。在此之前，管理员可以考虑尽可能增大检查点间隔相关参数，以减少 WAL 中包含的页面快照数量。

第 25 章 高可用、负载均衡和复制

数据库服务器可以协同工作，从而在主库失效时让第二台服务器快速接管其任务（高可用性），
或者让多台计算机提供同一份数据（负载均衡）。理想情况下，
数据库服务器应当能够无缝协同工作。提供静态网页服务的 Web 服务器只需把 Web 请求分发到多台机器，就能很容易地组合起来。事实上，只读数据库服务器也相对容易组合起来。
不幸的是，大多数数据库服务器同时处理读写请求，而读/写服务器要更难组合。这是因为，
只读数据只需要在每台服务器上放置一次，
而对任意一台服务器的写入都必须传播到所有服务器，
以保证后续对这些服务器的读取请求返回一致的结果。

这种同步问题是服务器协同工作的根本难点。
由于不存在一种能够对所有使用场景都消除同步问题影响的单一方案，
因此出现了多种不同的解决方案。每一种方案都以不同方式处理这个问题，
并且针对特定负载将其影响降到最低。

某些方案通过只允许一台服务器修改数据来处理同步。能够修改数据的服务器称为读/写、主库（master）或主库（primary）服务器。跟踪主库变更的服务器称为备库（standby）或备库（slave）。只能在被提升为主库之后才能连接的备库称为温备，
而能够接受连接并提供只读查询的备库称为热备。

某些方案是同步的，
即一个修改数据的事务只有在所有服务器都提交该事务之后才被视为已提交。
这保证一次故障切换不会丢失任何数据，
并且所有负载均衡的服务器无论查询哪一台都将返回一致的结果。相反，
异步方案允许一次提交与其传播到其他服务器之间存在一定延迟，
这就带来了切换到备库时丢失某些事务的可能性，
也意味着负载均衡的服务器可能返回略微陈旧的结果。当同步通信过慢时，就会使用异步通信。

这些方案还可以按粒度分类。有些方案只能处理整个数据库服务器，
而另一些则允许在每个表或者每个数据库级别进行控制。

无论做出哪种选择，都必须考虑性能。功能与性能之间通常存在权衡。例如，
在低速网络上采用完全同步的方案，性能可能下降一半以上，
而异步方案对性能的影响则可能很小。

本节其余部分将概述多种故障切换、复制和负载均衡方案。另有一份术语表[1]可供参考。

25.1. 不同方案的比较

共享磁盘故障切换

共享磁盘故障切换通过只保留一份数据库拷贝来避免同步开销。
它使用一个由多个服务器共享的单一磁盘阵列。如果主库失效，
备库就可以挂载并启动数据库，就好像它正在从一次数据库崩溃中恢复一样。
这允许快速故障切换而不会丢失数据。

共享硬件功能常见于网络存储设备。使用网络文件系统也是可行的，
但必须注意文件系统是否具备完整的 POSIX 行为（见第 17.2.2 节）。这种方法的一大局限是，如果共享磁盘阵列失效或损坏，主库和备库都会无法工作。另一个问题是，
在主库运行期间，备库绝不应访问共享存储。

[1] http://www.postgres-r.org/documentation/terms

文件系统（块设备）复制

共享硬件功能的一种变体是文件系统复制，
其中对一个文件系统的所有更改都会镜像到位于另一台计算机上的文件系统。
唯一的限制是，这种镜像必须以能够保证备库拥有该文件系统一致拷贝的方式完成 – 特别是，对备库的写入必须与主库上的写入保持相同顺序。DRBD 是 Linux 上一种流行的文件系统复制方案。

使用时间点恢复（PITR）的温备和热备

温备和热备服务器能够通过读取预写式日志（WAL）记录流来保持最新状态。
如果主库失效，备库拥有主库的几乎全部数据，并且能够迅速成为新的主数据库服务器。
这是异步的，并且只能用于整个数据库服务器。

可以使用基于文件的日志传送（第 25.2 节）、流复制（见第 25.2.5 节）或两者结合的方式来实现 PITR 备库。有关热备的信息，见第 25.5 节。

基于触发器的主备复制

主备复制配置会将所有修改数据的查询发送到主库。主库将数据更改异步发送到备库。
主库运行期间，备库可以响应只读查询。备库很适合处理数据仓库查询。

Slony-I 是这种复制类型的一个示例。它按表粒度工作，并且支持多个备库。由于它会以批处理方式异步更新备库，因此在故障切换期间可能会发生数据丢失。

基于语句的复制中间件

使用基于语句的复制中间件时，一个程序会截获每条 SQL 查询，并将其发送到一台或所有服务器。每台服务器独立运行。读写查询必须发送到所有服务器，
确保每台服务器都收到所有更改。但只读查询可以只发送到一台服务器，
从而将读取负载分散到这些服务器上。

如果只是原样广播查询，`random()`、`CURRENT_TIMESTAMP` 等函数以及序列在不同服务器上可能产生不同的值。这是因为每台服务器独立运行，且广播的是 SQL 查询，而不是实际修改后的行。如果无法接受这种情况，中间件或应用就必须从同一台服务器查询这些值，
然后在写入查询中使用这些值。另一种选择是将这种复制方式与传统主备配置结合使用，
即只向主库发送修改数据的查询，再通过主备复制将其传播到备库，
而不是通过复制中间件传播。还必须确保所有事务在所有服务器上要么全部提交，
要么全部中止，为此可以使用两阶段提交（PREPARE TRANSACTION和COMMIT PREPARED）。Pgpool-II 和 Continuent Tungsten 就是这类复制的例子。

异步多主复制

对于那些并不经常连接的服务器（例如笔记本电脑或远程服务器），
保持服务器间的数据一致是一个挑战。使用异步多主复制时，每台服务器都独立工作，
并定期与其他服务器通信，以识别冲突事务。这些冲突可以由用户或冲突解决规则来解决。
Bucardo 是这种复制类型的一个示例。

同步多主复制

在同步多主复制中，每台服务器都能接受写请求，并且在每个事务提交之前，
修改过的数据都会从原始服务器传送给其他每台服务器。
繁重的写入活动可能导致过多锁定，进而带来较差的性能。事实上，
写性能往往比单台服务器更差。读请求可以发送给任意服务器。
某些实现使用共享磁盘来减少通信开销。同步多主复制尤其适合以读为主的负载，
尽管它的一大优势是任意服务器都能接受写请求——无需在主库和备库之间划分负载，

并且由于数据更改是从一台服务器传送到另一台服务器，因此不会出现 `random()` 这类非确定函数的问题。

PostgreSQL 本身不提供这种复制，不过可以在应用代码或中间件中利用 PostgreSQL 的两阶段提交（PREPARE TRANSACTION和COMMIT PREPARED）来实现这一功能。

商业解决方案

由于 PostgreSQL 是开源的，而且易于扩展，许多公司以 PostgreSQL 为基础，创建了具有独特故障切换、复制和负载均衡功能的商业闭源解决方案。

表 25.1总结了上述多种方案的能力。

表 25.1. 高可用、负载均衡和复制特性矩阵

Feature	Shared Disk Failover	File System Replication	Hot/Warm Standby Using PITR	Trigger-Based Master-Standby Replication	Statement-Based Replication Middleware	Asynchronous Multimaster Replication	Synchronous Multimaster Replication
Most Common Implementation	NAS	DRBD	PITR	Slony	pgpool-II	Bucardo	
Communication Method	shared disk	disk blocks	WAL	table rows	SQL	table rows	table rows and row locks
No special hardware required		•	•	•	•	•	•
Allows multiple master servers					•	•	•
No master server overhead	•		•		•		
No waiting for multiple servers	•		•	•		•	
Master failure will never lose data	•	•			•		•
Standby accept read-only queries			Hot only	•	•	•	•
Per-table granularity				•		•	•
No conflict resolution necessary	•	•	•	•			•

有一些方案不适合上述的类别：

数据分区

数据分区会把表拆分成多个数据集。每个数据集只能由一台服务器修改。例如，可以按办公室划分数据，如伦敦和巴黎，每个办公室各有一台服务器。如果有必要执行组合伦敦和巴黎数据的查询，应用可以同时查询两台服务器，或者可以使用主库/备库复制，在每台服务器上保留其他办公室数据的只读副本。

多服务器并行查询执行

上述许多方案都允许多台服务器处理多个查询，但没有一种允许单个查询同时使用多台服务器来更快完成。这种方案允许多台服务器在同一个查询上并发工作。通常的做法是把数据分散到各台服务器上，让每台服务器执行该查询中属于自己的部分，然后把结果返回给一台中心服务器，由它汇总结果并返回给用户。Pgpool-II就具备这种能力。此外，这种方案也可以使用PL/Proxy工具集来实现。

25.2. 日志传送备库

持续归档可用于创建一种高可用性（HA）集簇配置，其中有一个或多个备库随时准备在主库失效时接管操作。这种能力通常称为温备或日志传送。

主库和备库协同工作以提供这种能力，不过两者之间只是松耦合。主库运行在持续归档模式下，而每台备库都运行在持续恢复模式下，从主库读取 WAL 文件。启用这种能力不需要修改数据库表，因此与其他一些复制方案相比，它的管理开销较低。这种配置对主库的性能影响也相对较小。

将 WAL 记录直接从一台数据库服务器移动到另一台数据库服务器，通常称为日志传送。PostgreSQL 通过一次传送一个文件（WAL 段）中的 WAL 记录来实现基于文件的日志传送。WAL 文件（16MB）可以轻松且低成本地传送到任意距离，无论是邻近系统、同一站点的另一台系统，还是地球另一端的系统。该技术所需的带宽取决于主库的事务速率。基于记录的日志传送粒度更细，会通过网络连接增量地流式传输 WAL 变更（见第 25.2.5 节）。

需要注意的是，日志传送是异步的，也就是说 WAL 记录是在事务提交之后才被传送的。因此，如果主库发生灾难性故障，就会存在一个数据丢失窗口；尚未传送的事务将会丢失。对于基于文件的日志传送，可以通过使用 `archive_timeout` 参数来限制这个数据丢失窗口，它可以设置为低至数秒。不过，如此低的设置会显著增加文件传送所需的带宽。流复制（见第 25.2.5 节）允许把数据丢失窗口缩得更小。

恢复性能已经足够好，因此一旦备库被激活，通常只需片刻就能达到完全可用状态。因此，这种配置被称为温备配置，它提供了高可用性。从归档的基础备份中恢复服务器并向前重放则需要更长时间，因此这种技术只能用于灾难恢复，而不是高可用性。备库还可以用于只读查询，这种情况下它被称为热备服务器。更多信息见第 25.5 节。

25.2.1. 规划

通常，最好让主库和备库尽可能相似，至少从数据库服务器的角度看应如此。尤其是，与表空间相关的路径名会原样传递，因此如果使用该特性，主库和备库必须为表空间配置完全相同的挂载路径。请记住，如果在主库上执行CREATE TABLESPACE，则它所需的任何新挂载点都必须在执行该命令之前先在主库和所有备库上创建好。硬件不必完全相同，但经验表明，在应用和系统的整个生命周期内，维护两个相同的系统要比维护两个不同的系统更容易。无论如何，硬件架构必须相同 – 例如，从 32 位系统向 64 位系统传送日志是不可行的。

一般来说，不能在运行不同主版本 PostgreSQL 的服务器之间传送日志。PostgreSQL 全球开发组的策略是在次版本升级期间不改变磁盘格式，
因此主库和备库运行不同次版本通常也能正常工作。不过，这方面并没有正式支持，
因此仍建议主库和备库尽量保持在相同的发行级别。当升级到新的次版本时，
最安全的策略是先升级备库 – 新的次版本更有可能兼容读取前一个次版本生成的 WAL 文件，反过来则未必。

25.2.2. 备库操作

在备库模式下，服务器会持续应用从主库收到的 WAL。备库可以从 WAL 归档读取 WAL（参见restore_command），也可以通过 TCP 连接直接从主库读取（流复制）。备库还会尝试恢复其集簇 `pg_xlog` 目录中找到的所有 WAL。这通常发生在服务器重启后，此时备库会重新重放重启前从主库接收的 WAL；不过，也可以随时手动将文件复制到 `pg_xlog` 中，让备库重放它们。

启动时，备库首先调用 `restore_command`，恢复归档位置中所有可用的 WAL。当读到归档中可用 WAL 的末尾，且 `restore_command` 失败后，它会尝试恢复 `pg_xlog` 目录中的所有可用 WAL。如果这也失败，且已配置流复制，备库就会尝试连接主库，从归档或 `pg_xlog` 中找到的最后一条有效记录开始接收 WAL 流。如果连接失败、未配置流复制，或连接后来断开，备库就会返回第一步，再次尝试从归档恢复文件。这种依次从归档、`pg_xlog` 和流复制重试的循环，会一直持续到服务器停止，或触发文件触发故障切换。

运行 `pg_ctl promote` 或发现触发文件（`trigger_file`）时，服务器会退出备库模式，切换到正常运行。在故障切换前，会恢复归档或 `pg_xlog` 中能立即获取的所有 WAL，但不会尝试连接主库。

25.2.3. 为备库准备主库

如第 24.3 节所述，在主库上设置持续归档，将日志归档到备库可访问的归档目录。即使主库宕机，该归档位置也应该对备库可访问；也就是说，它应位于备库本身或另一台可信服务器上，
而不是位于主库上。

如果想使用流复制，就需要在主库上配置认证，以允许来自备库的复制连接；也就是说，
创建一个角色，并在 `pg_hba.conf` 中添加一个或多个把数据库字段设置为 `replication` 的合适项。还要确保主库配置文件中的 `max_wal_senders` 被设置为足够大的值。

如第 24.3.2 节所述，获取一个基础备份来引导备库。

25.2.4. 设置备库

要配置备库，请恢复从主库取得的基础备份（参见第 24.3.3 节）。在备库的集簇数据目录中创建恢复命令文件 `recovery.conf`，并启用 `standby_mode`。将 `restore_command` 设为一个从 WAL 归档复制文件的简单命令。如果计划使用多个备库提供高可用，请将 `recovery_target_timeline` 设为 `latest`，使备库能够跟随故障切换到另一个备库时发生的时间线变化。

> **注意**
>
> 不要将 pg_standby 或类似工具与这里介绍的内置备库模式一起使用。如果文件不存在，`restore_command` 应立即返回；服务器会在必要时重试该命令。有关 pg_standby 等工具的使用，参见第 25.4 节。

如果希望使用流复制，请在 `primary_conninfo` 中填写 libpq 连接字符串，包含连接主库所需的主机名（或 IP 地址）及其他信息。如果主库使用密码认证，还需要在 `primary_conninfo` 中指定密码。

如果你是为了高可用目的设置备库，那么也应像主库一样设置 WAL 归档、连接和认证，因为故障切换后该备库将作为主库工作。

如果使用 WAL 归档，可以借助archive_cleanup_command参数删除备库不再需要的文件，以尽量缩小归档大小。pg_archivecleanup 工具专门设计用于在典型的单备库配置中与 `archive_cleanup_command` 配合使用，见pg_archivecleanup。不过请注意，如果你还把归档用于备份目的，即使这些文件对备库已经不再需要，
也必须至少保留从最新基础备份恢复所需的那些文件。

一个简单的 `recovery.conf` 示例如下：

```
standby_mode = 'on'
primary_conninfo = 'host=192.168.1.50 port=5432 user=foo password=
foopass'
restore_command = 'cp /path/to/archive/%f %p'
archive_cleanup_command = 'pg_archivecleanup /path/to/archive %r'
```

备库的数量可以任意多，但如果使用流复制，请确保主库上的 `max_wal_senders` 设置得足够高，以允许它们同时连接。

25.2.5. 流复制

与基于文件的日志传送相比，流复制可以让备库保持得更接近最新状态。备库连接到主库，
主库在生成 WAL 记录时就会把它们流式发送给备库，而不必等待 WAL 文件被写满。

默认情况下流复制是异步的（见第 25.2.6 节），在这种情况下主库上提交一个事务与该变化在备库上变得可见之间存在短暂的延迟。不过这种延迟比基于文件的日志传送方式中要小得多，
在备库的能力足以跟得上负载的前提下，延迟通常低于一秒。在流复制中，不需要 `archive_timeout` 来缩减数据丢失窗口。

如果使用流复制而没有使用基于文件的连续归档，就必须在主库中把 `wal_keep_segments` 设置为足够大的值，以确保旧的 WAL 段不会在备库可能仍需要它们追赶时过早回收。如果备库落后太多，就需要用新的基础备份重新初始化。如果你设置了备库可以访问的 WAL 归档，则不需要 `wal_keep_segments`，因为备库总是可以使用归档来追赶。

要使用流复制，请先按第 25.2 节所述配置基于文件的日志传送备库。将这种备库转为流复制备库的关键步骤，是在 `recovery.conf` 文件中设置 `primary_conninfo`，使其指向主库。在主库上设置listen_addresses和认证选项（参见 `pg_hba.conf`），使备库能够连接主库的 `replication` 伪数据库（参见第 25.2.5.1 节）。

在支持 keepalive 套接字选项的系统上，设置tcp_keepalives_idle、tcp_keepalives_interval和tcp_keepalives_count有助于主库迅速注意到断开的连接。

设置来自备库的最大并发连接数（详见max_wal_senders）。

备库启动后，如果正确设置了 `primary_conninfo`，备库会在重放归档中所有可用的 WAL 文件后连接主库。如果连接成功，就会在备库上看到 walreceiver 进程，在主库上看到对应的 walsender 进程。

25.2.5.1. 认证

非常重要的一点是，复制的访问权限必须设置为只允许受信任的用户读取 WAL 流，因为从中很容易提取出特权信息。备库必须以具有 `REPLICATION` 权限的账户向主库进行认证。因此需要在主库上创建一个具有 `REPLICATION` 和 `LOGIN` 权限的角色。

> **注意**
>
> 建议为复制使用一个专用的用户账户。虽然 REPLICATION 权限默认授予超级用户账户，但不推荐使用超级用户账户进行复制。REPLICATION 权限赋予的权限已经很高，但它不允许用户修改主库系统上的任何数据，而 SUPERUSER 权限允许这样做。

复制的客户端认证由 pg_hba.conf 中的一条记录控制，该记录需要把 replication 指定在 *database* 字段中。例如，如果备库运行在主机 IP 192.168.1.100 上，并且用于复制的账户名为 foo，管理员可以在主库上的 pg_hba.conf 文件中加入下列行：

```
# Allow the user "foo" from host 192.168.1.100 to connect to the
 primary
# as a replication standby if the user's password is correctly
 supplied.
#
# TYPE  DATABASE        USER            ADDRESS                 METHOD
host    replication     foo             192.168.1.100/32        md5
```

主库的主机名、端口号、连接用户名和密码在 recovery.conf 文件中指定。也可以将密码设置在备库的 ~/.pgpass 文件中（将 replication 指定为 *database* 字段的值）。例如，如果主库所在主机的 IP 地址为 192.168.1.50，端口为 5432，用于复制的账户名为 foo，密码为 foopass，管理员可以将以下行添加到 recovery.conf 文件（位于备库）中：

```
# The standby connects to the primary that is running on host 192.168.
1.50
# and port 5432 as the user "foo" whose password is "foopass".
primary_conninfo = 'host=192.168.1.50 port=5432 user=foo password=
foopass'
```

25.2.5.2. 监控

流复制的一个重要健康指标，是主库上已经生成但尚未在备库上应用的 WAL 记录量。你可以通过比较主库上的当前 WAL 写入位置和备库收到的最后一个 WAL 位置来计算这种滞后。这些位置分别可以用主库上的 pg_current_xlog_location 和备库上的 pg_last_xlog_receive_location 取得（详见表 9.57和表 9.58）。备库上的最后一个 WAL 接收位置也会显示在 WAL 接收进程的进程状态中，即通过 ps 命令显示的状态（详见第 27.1 节）。

可以通过 pg_stat_replication 视图获取 WAL 发送进程列表。pg_current_xlog_location 与 sent_location 字段相差较大，可能表示主库负载很高；而 sent_location 与备库上 pg_last_xlog_receive_location 之间存在差异，则可能表示网络延迟或备库负载很高。

25.2.6. 同步复制

PostgreSQL 的流复制默认是异步的。如果主库崩溃，则某些已提交的事务可能尚未复制到备库，从而导致数据丢失。数据丢失量与故障切换时的复制延迟成正比。

同步复制能够确认一个事务所做的全部修改已经被传送到一台同步备库。
这扩展了事务提交所提供的标准持久性级别。在计算机科学理论中，这种保护级别被称为 2-safe 复制。

请求同步复制时，每个写事务的提交都会等待，直到收到确认，
表明该提交已被写入主库和备库磁盘上的事务日志。数据唯一可能丢失的情况，

是主库和备库同时崩溃。这可以提供更高的持久性级别，
不过前提是系统管理员必须谨慎地部署和管理这两台服务器。
等待确认会增强用户对服务器崩溃时更改不会丢失的信心，
但也必然会增加请求事务的响应时间。最短等待时间是主库与备库之间的往返时间。

只读事务和事务回滚不需要等待备库的回应。子事务提交也不需要等待备库响应，
只有顶层提交才需要等待。数据装载或索引构建等长时间运行的动作，
直到最终提交时才会等待。所有两阶段提交操作都需要等待提交，包括准备和提交两个阶段。

25.2.6.1. 基本配置

一旦流复制已经配置好，配置同步复制只需要额外一步配置：必须把synchronous_standby_names设置为非空值。`synchronous_commit` 也必须设置为 `on`，但由于这是默认值，通常无需更改。这样的配置会导致每次提交都等待确认备库已经把提交记录写入持久存储，
即使那需要很长时间。`synchronous_commit` 可以由单个用户设置，因此既可以在配置文件中配置，也可以针对特定用户或数据库配置，或者由应用动态配置，
从而在每事务级别控制持久性保证。

当提交记录已经在主库上写入磁盘之后，WAL 记录就会被发送到备库。每当新的一批 WAL 数据被写入磁盘时，备库就会发送回复消息，除非备库上的 `wal_receiver_status_interval` 被设置为零。如果根据主库上的 `synchronous_standby_names`，该备库是第一个匹配的备库，那么它发出的回复将被用于唤醒那些正在等待确认提交记录已被收到的用户。
这些参数允许管理员指定哪些备库应作为同步备库。注意，同步复制的配置主要在主库上进行。

如果请求快速关闭，用户将停止等待。不过，与使用异步复制时一样，在所有尚未传送的 WAL 记录传输到当前已连接的备库之前，服务器不会完全关闭。

25.2.6.2. 性能规划

同步复制通常要求对备库进行仔细规划和部署，才能保证应用具有可接受的性能。
等待本身不会占用系统资源，但事务锁会一直保持到传输得到确认。因此，
若不谨慎使用同步复制，数据库应用的性能会因为响应时间增加和争用加剧而下降。

PostgreSQL允许应用开发者通过复制来指定所需的持久性级别。这可以在整个系统级别指定，
也可以针对特定用户、特定连接，甚至单个事务指定。

例如，一个应用负载可能由如下部分组成：10% 的变更是重要的客户资料，而 90% 的变更是不太重要、即使丢失业务也较容易承受的数据，例如用户之间的聊天消息。

通过在应用级别（在主库上）指定同步复制选项，我们可以只对最重要的变更提供同步复制，
而不会拖慢大部分工作负载。
应用级别选项是让高性能应用获得同步复制收益的一种重要而实用的工具。

你应当确保网络带宽高于 WAL 数据的生成速率。

25.2.6.3. 高可用规划

当 `synchronous_commit` 设置为 `on` 时，事务提交将等待，直到同步备库做出响应。如果最后一台（或唯一一台）备库崩溃，这种响应可能永远不会到来。

对于避免数据丢失而言，最好的办法是确保你不会失去最后一台剩余的同步备库。这可以通过在 `synchronous_standby_names` 中命名多个潜在同步备库来实现。第一台被命名的备库将被用作同步备库。列在它之后的备库，会在第一台失效时接替同步备库的角色。

当一台备库第一次连接到主库时，它还没有正确同步。这种状态称为`catchup`模式。
一旦备库与主库之间的滞后第一次变为零，它就会进入实时的`streaming`状态。

备库刚创建之后，追赶阶段可能会持续较长时间。如果备库被关闭，
则追赶阶段会随着它停机时间的延长而变长。只有在到达streaming状态后，
备库才能成为同步备库。

如果主库在提交正等待确认时重启，这些等待中的事务会在主库恢复后被标记为已完全提交。
无法确定在主库崩溃时，所有备库是否已经收到全部待传送的 WAL 数据。因此，某些事务可能不会在备库上显示为已提交，即使它们在主库上显示为已提交。我们所提供的保证是：只有在确认 WAL 数据已经被该备库安全接收之后，应用才会收到事务成功提交的显式确认。

如果你确实失去了最后一台备库服务器，那么就应当禁用synchronous_standby_names，
然后在主库上重新加载配置文件。

如果主库与剩余的备库隔离开了，你应当故障切换到那些剩余备库中最佳的候选者。

如果需要在事务等待期间重新创建备库，请确保在 synchronous_commit = off 的会话中运行 pg_start_backup() 和 pg_stop_backup() 命令，否则这些请求会一直等待备库出现。

25.3. 故障切换

如果主库失效，备库就应该开始执行故障切换过程。

如果备库失效，则不需要发生故障切换。如果备库能够重新启动，即使是在稍后某个时间点，
恢复过程也可以立即重新开始，从而利用可重启恢复的优势。如果备库无法重新启动，
则应创建一个全新的备库实例。

如果主库失效，而备库成为新的主库，那么旧主库之后如果重新启动，
你必须有一种机制通知它，它已经不再是主库。这有时被称为STONITH（Shoot The Other Node In The Head），它对于避免两个系统都认为自己是主库的情况至关重要，因为那种情况会导致混乱，并最终造成数据丢失。

许多故障切换系统只使用两个系统，即主库和备库，并通过某种心跳机制连接它们，
以持续验证两者之间的连通性以及主库的可用性。也可以使用第三个系统（称为见证服务器）
来防止某些不恰当的故障切换，但除非设置得足够谨慎并经过严格测试，
否则额外增加的复杂性可能并不值得。

PostgreSQL并不提供用于识别主库故障并通知备库的系统软件。现在已经存在许多这样的工具，
并且它们通常能很好地与成功故障切换所需的操作系统设施整合在一起，例如 IP 地址迁移。

一旦故障切换到备库，系统中就只剩下一台服务器在运行。这被称为退化状态。
原来的备库现在成为主库，而原来的主库已经停机，并且可能持续停机。
要恢复到正常运行状态，就必须重新创建一台备库：要么在原主库恢复后在其上重建，
要么在第三台可能是全新的服务器上重建。一旦完成，就可以认为主库和备库已经交换了角色。
有些人会选择使用第三台服务器，在新的备库重建完成之前为新的主库提供后备支持，
但显然这会让系统配置和操作流程更加复杂。

因此，从主库切换到备库可以很快，但重新准备故障切换集簇仍然需要时间。
定期在主库与备库之间进行切换是有益的，因为它允许每个系统定期停机维护。
这也相当于对故障切换机制进行测试，以确保真正需要它时它能够正常工作。
建议编写书面的管理操作规程。

要触发日志传送备库的故障切换，可以运行 pg_ctl promote，或按照 recovery.conf 中 trigger_file 设置指定的文件名和路径创建触发文件。如果计划使用 pg_ctl promote 进行故障切换，就不需要设置 trigger_file。如果所配置的报表服务器只是用来分担主库的只读查询，而不用于高可用，则不需要提升它。

25.4. 日志传送的另一种方法

除了前几节介绍的内置备库模式，也可以使用一个轮询归档位置的 `restore_command`。在 8.4 及更早版本中，这是唯一可用的方法。采用这种配置时，应关闭 `standby_mode`，因为备库运行所需的轮询由你自己实现。其参考实现参见pg_standby模块。

注意，在这种模式下，服务器每次应用一整个 WAL 文件。因此，如果使用备库处理查询（参见热备），主库上的操作与备库上能看到该操作的结果之间，会有一段延迟，其长度相当于写满一个 WAL 文件所需的时间。可以用 `archive_timeout` 缩短这一延迟。还要注意，这种方法不能与流复制结合使用。

主库和备库上执行的操作都是普通的持续归档和恢复任务。两台数据库服务器之间唯一的联系，就是它们共享的 WAL 文件归档：主库写入归档，备库从归档读取。必须确保不同主库的 WAL 归档不会混在一起或弄错。如果归档仅用于备库运行，则无需保留很大的归档。

使这两台松耦合的服务器协同工作的诀窍，只是备库上的一个`restore_command`：当被要求提供下一个 WAL 文件时，它会等待该文件从主库那边变为可用。`restore_command`在备库服务器的`recovery.conf`文件中指定。正常的恢复处理会向 WAL 归档请求文件，如果文件不可用则报告失败。而对于备库处理来说，下一个 WAL 文件不可用是正常情况，因此备库必须等待它出现。对于以`.backup`或`.history`结尾的文件则无须等待，必须返回非零返回码。等待型的`restore_command`可以写成一个自定义脚本，在轮询下一个 WAL 文件是否存在之后循环等待。还必须有某种触发故障转移的方法，它应中断`restore_command`、打破循环并让备库退出恢复模式。

一个合适的 `restore_command` 的伪代码如下：

```
triggered = false;
while (!NextWALFileReady() && !triggered)
{
    sleep(100000L);         /* wait for ~0.1 sec */
    if (CheckForExternalTrigger())
        triggered = true;
}
if (!triggered)
        CopyWALFileForRecovery();
```

pg_standby模块提供了会等待的 `restore_command` 的可用示例。应参考它来正确实现上述逻辑，也可以按需扩展，以支持特定配置和环境。

触发故障切换的方法是规划和设计的重要部分。一个可能的选择是 `restore_command` 命令。每个 WAL 文件都会执行一次该命令，但运行 `restore_command` 的进程针对每个文件单独创建和结束，因此不存在守护进程或服务器进程，也无法使用信号或信号处理程序。所以，`restore_command` 不适合触发故障切换。可以使用简单的超时机制，尤其是在已知主库 `archive_timeout` 设置的情况下配合使用。不过，这容易出错，因为网络问题或主库繁忙就可能导致故障切换。如果能够安排，显式创建触发文件之类的通知机制是理想的方式。

25.4.1. 实现

使用这种替代方法配置备库的简要流程如下。各步骤的完整细节，参见所注明的前面章节。

1. 尽可能将主库和备库系统配置得相同，包括安装两个完全相同、发行版本一致的 PostgreSQL 副本。
2. 配置持续归档，将主库的 WAL 归档到备库上的一个目录。确保在主库上正确设置archive_mode、archive_command和archive_timeout（参见第 24.3.1 节）。

3. 制作主库的基础备份（参见第 24.3.2 节），并将这些数据装载到备库上。

4. 在备库上从本地 WAL 归档开始恢复，在 `recovery.conf` 中指定前面所述的会等待的 `restore_command`（参见第 24.3.3 节）。

恢复过程将 WAL 归档视为只读，因此，WAL 文件一旦复制到备库系统，就可以在备库读取它的同时，将其复制到磁带上。这样，既能运行备库提供高可用，
又能保存文件用于更长期的灾难恢复。

为了测试，可以在同一个系统上同时运行主库和备库。
这不会对服务器的稳健性带来有意义的提升，也不能称为高可用。

25.4.2. 基于记录的日志传送

也可以使用这种替代方法实现基于记录的日志传送，但需要自行开发，而且只有整个 WAL 文件传送完成后，更改才会对热备上的查询可见。

外部程序可以调用 `pg_xlogfile_name_offset()` 函数（参见第 9.24 节），取得当前 WAL 末尾所在的文件名及其在文件中的精确字节偏移。随后就可以直接访问 WAL 文件，将上次已知的 WAL 末尾到当前末尾之间的数据复制到备库。采用这种方法，可能丢失数据的时间窗口就是复制程序的轮询周期，可以很短，而且不会因为强制归档未写满的段文件而浪费带宽。注意，备库的 `restore_command` 脚本只能处理完整的 WAL 文件，因此这些增量复制的数据通常不会提供给备库使用。它们只有在主库失效时才有用 – 此时，在允许备库启动前，将最后一个不完整的 WAL 文件交给备库。要正确实现这一过程，`restore_command` 脚本必须与数据复制程序配合。

从 PostgreSQL 9.0 开始，可以使用流复制（参见第 25.2.5 节），更省力地获得同样的收益。

25.5. 热备

热备是用来描述服务器在归档恢复或备库模式下仍然可以接受连接并运行只读查询的能力的术语。这对于复制用途以及把备份以极高精度恢复到目标状态都很有用。
术语热备还指服务器能够在用户持续运行查询和/或保持连接不断开的同时，
从恢复状态切换到正常运行状态的能力。

在热备模式下运行查询与正常查询操作类似，不过正如下文所述，
在使用和管理上存在一些差异。

25.5.1. 用户概览

当备库上的hot_standby参数被设置为真时，一旦恢复把系统带到一致状态，
它就会开始接受连接。所有这类连接都严格是只读的，甚至不能写入临时表。

备库上的数据需要一些时间才能从主库到达，因此主库和备库之间会有可测量的延迟。因此，
在主库和备库上几乎同时运行同一查询，可能会返回不同的结果。我们说备库上的数据与主库是最终一致的。一旦某个事务的提交记录在备库上被重放，
该事务所做的修改就会对备库上之后取得的所有新快照可见。快照可以在每个查询开始时取得，
也可以在每个事务开始时取得，这取决于当前的事务隔离级别。详见第 13.2 节。

在热备期间启动的事务可以发出下列命令：

- 查询访问 -- `SELECT`、`COPY TO`

- 游标命令 -- `DECLARE`、`FETCH`、`CLOSE`

- 参数 -- SHOW、SET、RESET
- 事务管理命令
  - BEGIN, END, ABORT, START TRANSACTION
  - SAVEPOINT、RELEASE、ROLLBACK TO SAVEPOINT
  - EXCEPTION块及其他内部子事务
- LOCK TABLE，但仅当显式处于以下模式之一：ACCESS SHARE、ROW SHARE 或 ROW EXCLUSIVE。
- 计划和资源 -- PREPARE、EXECUTE、DEALLOCATE、DISCARD
- 插件和扩展 -- LOAD

热备期间启动的事务永远不会被分配事务 ID，也不能写系统预写式日志 to the system write-ahead log. Therefore, the following actions will produce error messages:

- Data Manipulation Language (DML) - INSERT, UPDATE, DELETE, COPY FROM, TRUNCATE . 注意，没有任何被允许的操作会导致触发器在恢复期间执行。这一限制甚至适用于临时表，因为不分配事务 ID 就无法读写表行，而热备环境目前无法分配事务 ID。
- Data Definition Language (DDL) - CREATE, DROP, ALTER, COMMENT. 这一限制甚至适用于临时表，因为执行这些操作需要更新系统目录表。
- SELECT ... FOR SHARE | UPDATE，因为不更新底层数据文件就无法取得行锁。
- 在 SELECT 语句上生成 DML 命令的规则。
- 显式请求高于 ROW EXCLUSIVE MODE 模式的 LOCK。
- 短默认形式的 LOCK，因为它请求的是 ACCESS EXCLUSIVE MODE。
- 显式设置非只读状态的事务管理命令：
  - BEGIN READ WRITE, START TRANSACTION READ WRITE
  - SET TRANSACTION READ WRITE、SET SESSION CHARACTERISTICS AS TRANSACTION READ WRITE
  - SET transaction_read_only = off
- 两阶段提交命令 -- PREPARE TRANSACTION、COMMIT PREPARED、ROLLBACK PREPARED，因为即使是只读事务，在准备阶段（两阶段提交的第一阶段）也需要写 WAL。
- 序列更新 -- nextval()、setval()
- LISTEN、UNLISTEN、NOTIFY

在正常运行中，“只读”事务允许更新序列，也允许使用LISTEN、UNLISTEN和NOTIFY，因此热备会话受到的限制比普通只读会话略严。其中某些限制有可能在未来的版本中放宽。

在热备期间，参数transaction_read_only始终为真，并且不能更改。不过，只要不尝试修改数据库，热备期间的连接行为与其他数据库连接大致相同。

如果发生故障切换或计划内切换，数据库将切换到正常处理模式。服务器切换模式时，会话仍会保持连接。一旦热备结束，就可以发起读写事务（即使该会话是在热备期间开始的）。

用户可以执行 `SHOW transaction_read_only`，判断自己的会话是否只读。此外，还有一组函数（表 9.58）可用于获取备库信息。借助它们，可以编写感知数据库当前状态的程序，用来监控恢复进度，或编写将数据库恢复到特定状态的复杂程序。

25.5.2. 处理查询冲突

主库和备库在许多方面都是松耦合的。主库上的动作会对备库产生影响，因此它们之间可能出现负面交互或冲突。最容易理解的冲突是性能：如果主库上正在执行一次大规模数据装载，那么备库上也会产生类似的 WAL 记录流，因此备库查询可能会争用系统资源，例如 I/O。

热备还可能发生其他类型的冲突。这些冲突属于硬冲突，因为查询可能需要被取消，并且在某些情况下还需要断开会话来解决冲突。系统为用户提供了若干种处理这些冲突的方法。冲突场景包括：

- 主库上获取的 Access Exclusive 锁，包括显式的`LOCK`命令和各种DDL操作，会与备库查询中的表访问发生冲突。
- 主库上删除表空间，会与备库查询把该表空间用于临时工作文件发生冲突。
- 主库上删除数据库，会与备库上连接到该数据库的会话发生冲突。
- 来自 WAL 的清理记录在应用时，会与那些快照仍然能够“看到”将被删除的行的备库事务发生冲突。
- 来自 WAL 的清理记录在应用时，会与备库上访问目标页面的查询发生冲突，而不管待删除的数据是否可见。

在主库上，这些情况只会导致等待；用户可以选择取消冲突动作中的任意一方。但是在备库上没有这种选择：已经被写入 WAL 的动作早已在主库上发生，因此备库在应用它时不能失败。此外，让 WAL 应用无限期等待通常是很不可取的，因为备库的状态会越来越落后于主库。因此，系统提供了一种机制，用来强制取消那些与即将应用的 WAL 记录发生冲突的备库查询。

一个典型例子是，主库上的管理员在某个表上执行`DROP TABLE`，而备库上正有查询访问该表。显然，如果`DROP TABLE`在备库上被应用，该备库查询就无法继续。如果这种情况发生在主库上，`DROP TABLE`会等待直到其他查询结束。但在主库上执行`DROP TABLE`时，主库并不知道备库上正在运行哪些查询，因此它不会等待这些备库查询。于是，WAL 变更记录会在备库查询仍在运行时到达备库，从而引发冲突。备库要么延迟应用这条 WAL 记录（以及之后的所有记录），要么取消冲突查询，以便应用`DROP TABLE`。

当冲突查询很短时，通常最好通过稍微延迟 WAL 应用来让它完成；但 WAL 应用长时间延迟通常并不可取。因此，取消机制提供了max_standby_archive_delay和max_standby_streaming_delay这两个参数，用于定义 WAL 应用允许的最大延迟。一旦应用新收到的 WAL 数据所花费的时间超过相应延迟设置，冲突查询就会被取消。之所以有两个参数，是为了分别针对从归档读取 WAL 数据的场景（即从基础备份进行初始恢复，或者“追赶”一台已经远远落后的备库）以及通过流复制读取 WAL 数据的场景指定不同的延迟值。

如果一台备库主要用于高可用性，那么最好把这些延迟参数设置得较短，这样服务器就不会因为备库查询导致的延迟而远远落后于主库。但是，如果该备库的用途是执行长时间运行的查询，那么较高甚至无限的延迟值可能更合适。不过要记住，如果某个长时间运行的查询延迟了 WAL 记录的应用，它也可能使备库上的其他会话看不到主库上的最新变更。

一旦max_standby_archive_delay或max_standby_streaming_delay指定的延迟被超越，冲突查询将被取消。这通常仅导致一个取消错误，尽管在重放一个DROP DATABASE的情况下整个冲突会话都将被终止。另外，如果冲突发生在一个被空闲事务持有的锁上，该冲突会话会被终止（这种行为可能在未来被改变）。

被取消的查询可以立即重试（当然，要先开始一个新事务）。由于查询取消取决于正在重放的 WAL 记录的性质，被取消的查询再次执行时完全可能成功。

请记住，延迟参数要与备库收到 WAL 数据之后经过的时间进行比较。因此，留给备库上任何一个查询的宽限期从不会超过延迟参数，并且如果备库已经由于等待之前的查询完成而落后或者因为过重的更新负载而无法跟上主库，宽限期可能会更少。

备库查询与 WAL 重放发生冲突的最常见原因是“过早清理”。正常情况下，当不再有事务需要看到旧行版本来保证符合 MVCC 规则的数据可见性时，PostgreSQL允许清理这些旧行版本。不过，这条规则只能应用于在主库上执行的事务。因此，主库上的清理有可能移除某个备库事务仍然可见的行版本。

有经验的用户应注意，行版本清理和行版本冻结都有可能与备库查询冲突。手动运行 VACUUM FREEZE 很可能导致冲突，即使表中没有更新过或删除过的行也是如此。

用户应当清楚，主库上经常并且大量更新的表，会很快导致备库上的长时间运行查询被取消。在这种情况下，把max_standby_archive_delay或max_standby_streaming_delay设置为有限值，可以视作类似于设置statement_timeout。

如果备库查询被取消的次数多得令人无法接受，也存在补救办法。第一种选择是设置hot_standby_feedback参数，它会阻止VACUUM移除最近死亡的行，因此不会发生清理冲突。如果这样做，你应当注意这会延迟主库上死行的清理，从而可能导致不希望出现的表膨胀。不过，清理情况不会比直接在主库上运行这些备库查询更糟，而且你仍然可以获得把执行卸载到备库上的好处。在这种情况下必须保持较大的 max_standby_archive_delay，因为延迟的 WAL 文件中可能已经包含与所需备库查询相冲突的条目。

另一种选择是增大主库上的vacuum_defer_cleanup_age，使死行不会像通常那样很快被清理。这样，无需将 max_standby_streaming_delay 设得很大，就能让备库查询在被取消前有更多执行时间。不过，这种方法难以保证具体的执行时间窗口，因为 vacuum_defer_cleanup_age 是按主库上执行的事务数来衡量的。

查询取消的数量及其原因可以通过备库上的pg_stat_database_conflicts系统视图查看。pg_stat_database系统视图也包含汇总信息。

25.5.3. 管理员概览

如果在 postgresql.conf 中把 hot_standby 打开为 on，并且存在 recovery.conf 文件，服务器将以热备模式运行。但是，可能需要一段时间才允许热备连接，因为在完成足够的恢复、提供一个可供查询运行的一致状态之前，服务器不会接受连接。在此期间，尝试连接的客户端将被拒绝并收到一条错误消息。要确认服务器已经启动，可以在应用中循环尝试连接，或者在服务器日志中查找这些消息：

```
LOG:  entering standby mode

... then some time later ...

LOG:  consistent recovery state reached
LOG:  database system is ready to accept read only connections
```

Consistency information is recorded once per checkpoint on the primary. It is not possible to enable hot standby when reading WAL written during a period when `wal_level` was not set to `hot_standby` on the primary. Reaching a consistent state can also be delayed in the presence of both of these conditions:

- A write transaction has more than 64 subtransactions
- 存活时间很长的写事务

If you are running file-based log shipping ("warm standby"), you might need to wait until the next WAL file arrives, which could be as long as the `archive_timeout` setting on the primary.

如果某些参数在主库上被更改过，备库上的相应设置就需要重新配置。对于这些参数，备库上的值必须等于或大于主库上的值。如果这些参数设置得不够高，备库将拒绝启动。此时可以提供更高的值并重启服务器以重新开始恢复。这些参数是：

- `max_connections`
- `max_prepared_transactions`
- `max_locks_per_transaction`

管理员为max_standby_archive_delay和max_standby_streaming_delay选择合适的设置非常重要。最佳选择取决于业务优先级。例如，如果服务器的主要任务是充当高可用服务器，那么你会希望延迟设置较低，甚至可能设为零，尽管这是一个非常激进的设置。如果备库承担的是决策支持查询的附加服务器角色，那么把最大延迟设置为数小时甚至 -1（意味着永远等待查询完成）也可能是可以接受的。

主库上写出的事务状态“提示位”不会被 WAL 记录，因此备库上的数据很可能会再次写出这些提示位。这样一来，即使所有用户都是只读的，备库仍然会执行磁盘写操作；不过数据值本身并不会发生变化。用户仍然会写出大型排序临时文件，并重新生成 relcache 信息文件，因此在热备模式下，数据库没有任何部分是真正只读的。还要注意，使用dblink模块写入远程数据库，以及借助 PL 函数执行其他数据库外部操作，依然是可能的，即使该事务在本地是只读的。

在恢复模式下，不接受下列类型的管理命令：

- 数据定义语言（DDL）：例如 `CREATE INDEX`
- 权限和所有权：`GRANT`、`REVOKE`、`REASSIGN`
- 维护命令：`ANALYZE`、`VACUUM`、`CLUSTER`、`REINDEX`

再次注意，这些命令中的某些在主库上的“只读”事务中实际上是被允许的。

因此，你不能创建只存在于备库上的额外索引，也不能创建只存在于备库上的统计信息。如果需要这些管理命令，应在主库上执行，最终这些变更会传播到备库。

`pg_cancel_backend()`对用户后端有效，但对执行恢复的 Startup 进程无效。`pg_stat_activity`不会显示 Startup 进程的条目，恢复中的事务也不会显示为活动。因此，恢复期间`pg_prepared_xacts`总是空的。如果想解决存疑的预备事务，请查看主库上的`pg_prepared_xacts`并在那里发出命令解决事务。

与正常情况一样，`pg_locks`会显示后端持有的锁。`pg_locks`还会显示一个由启动进程管理的虚拟事务，它拥有所有正在被恢复重放的事务所持有的`Access`

ExclusiveLocks。请注意，启动进程不会为了修改数据库而获取锁，因此除了Access ExclusiveLocks之外，其他锁不会在启动进程的pg_locks中显示；它们只是被假定存在。

Nagios的check_pgsql插件可以工作，因为它检查的简单信息是存在的。check_postgres监控脚本也可以工作，尽管其中某些被报告的值可能会给出不同或令人困惑的结果。例如，最近一次清理时间不会被维护，因为备库上不会发生清理。不过，主库上执行的清理仍然会把其变更发送到备库。

恢复期间，WAL 文件控制命令不可用，例如pg_start_backup、pg_switch_xlog等。

可动态载入的模块可以工作，包括pg_stat_statements。

咨询锁在恢复期间可以正常工作，包括死锁检测。注意，咨询锁从不会被 WAL 记录，因此主库或备库上的咨询锁都不可能与 WAL 重放发生冲突。同样，也不可能在主库上获取一个咨询锁，却在备库上触发一个类似的咨询锁。咨询锁只与获取它们的那台服务器相关。

基于触发器的复制系统，如Slony、Londiste和Bucardo，根本无法在备库上运行；不过只要它们的变更不会被发送到备库并在那里应用，它们在主库上就可以正常工作。WAL 重放并不是基于触发器的，因此你不能把备库作为任何需要额外数据库写操作或依赖触发器的系统的中继节点。

不能分配新的 OID，不过某些UUID生成器仍可工作，只要它们不依赖于向数据库写入新的状态。

目前，在只读事务期间不允许创建临时表，因此某些现有脚本在这种情况下将无法正常运行。这个限制可能会在未来版本中放宽。这既涉及 SQL 标准兼容性问题，也涉及技术问题。

只有在表空间为空时DROP TABLESPACE才能成功。某些备库用户可能正在通过他们的temp_tablespaces参数使用该表空间。如果该表空间中存在临时文件，所有活动查询都将被取消，以确保临时文件被移除，这样该表空间才能被移除并且 WAL 重放可以继续。

在主库上执行DROP DATABASE或ALTER DATABASE ... SET TABLESPACE会生成一条 WAL 记录，从而强制断开备库上所有连接到该数据库的用户。这个动作会立即发生，而不受max_standby_streaming_delay设置的影响。注意，ALTER DATABASE ... RENAME不会断开用户，这在大多数情况下不会被注意到，但如果程序依赖某种基于数据库名的机制，在某些情况下可能会导致混乱。

在普通（非恢复）模式下，如果你对一个具有登录能力的角色执行DROP USER或DROP ROLE，而该用户仍然处于连接状态，那么已连接用户不会发生任何变化 – 他们会继续保持连接，不过之后不能重新连接。这种行为在恢复期间同样适用，因此在主库上执行一次DROP USER并不会断开备库上的该用户连接。

统计收集器在恢复期间保持活动状态。所有扫描、读取、块、索引使用情况等，都会在备库上正常记录。重放操作不会重复记录其在主库上产生的统计影响，因此重放一次插入不会增加 pg_stat_user_tables 的 Inserts 列。统计文件会在恢复开始时被删除，因此主库和备库的统计信息不同；这是特性，不是缺陷。

恢复期间自动清理不会运行。它会在恢复结束时正常启动。

检查点进程和后台写入进程在恢复期间是活动的。检查点进程会执行重启点（类似于主库上的检查点），后台写入进程会执行正常的块清理活动。这可能包括更新存储在备库上的提示位信息。恢复期间接受CHECKPOINT命令，不过它执行的是重启点，而不是新的检查点。

25.5.4. 热备参数参考

多个参数已经在第 25.5.2 节和第 25.5.3 节中提到过。

在主库上，可以使用wal_level和vacuum_defer_cleanup_age参数。max_standby_archive_delay和max_standby_streaming_delay在主库上设置时没有效果。

在备库上，可以使用hot_standby、max_standby_archive_delay和max_standby_streaming_delay参数。只要服务器仍处于备库模式，vacuum_defer_cleanup_age就没有效果，不过备库成为主库后，它就会起作用。

25.5.5. 注意事项

热备有若干限制。这些限制在未来的版本中可以、也很可能会被修复：

- 哈希索引上的操作目前不会写入 WAL 日志，因此重放不会更新这些索引。
- 在能够取得快照之前，必须完整了解正在运行的事务。使用大量子事务（目前超过 64 个）的事务，会把只读连接的启动推迟到持续时间最长的写事务完成之后。如果发生这种情况，会向服务器日志发送解释性消息。
- 备库查询的有效起始点是在主库的每个检查点上生成的。如果主库处于关闭状态时备库被关闭，那么在主库再次启动并在 WAL 日志中生成更多起始点之前，备库可能无法重新进入热备状态。不过在最常见的场景中，这通常不是问题。一般来说，如果主库关闭且不再可用，很可能是出现了严重故障，此时无论如何都需要把备库提升为新的主库。而在主库被有意关闭的情况下，协调好备库顺利成为新主库，本来也就是标准流程。
- 在恢复结束时，由预备事务持有的`AccessExclusiveLocks`将需要正常数量两倍的锁表项。如果你计划运行大量并发的预备事务，而这些事务通常会持有`AccessExclusiveLocks`，或者你计划运行一个会持有许多`AccessExclusiveLocks`的大事务，那么建议选择更大的`max_locks_per_transaction`值，可能要达到主库上该参数值的两倍。如果你的`max_prepared_transactions`设置为 0，则完全不需要考虑这一点。
- 热备中尚不支持可串行化事务隔离级别。（详见第 13.2.3 节和第 13.4.1 节。）在热备模式下尝试把事务设置为可串行化隔离级别会产生错误。

第 26 章 恢复配置

本章介绍 recovery.conf 文件中可用的设置。它们只在恢复期间生效。如果之后还要进行恢复，必须重新设置这些参数。恢复开始后，就不能再更改它们。

recovery.conf 中的设置使用 name = 'value' 格式，每行指定一个参数。井号（#）表示该行余下内容为注释。要在参数值中嵌入单引号，请写两个单引号（''）。

安装目录的 share/ 目录中提供了示例文件 share/recovery.conf.sample。

26.1. 归档恢复设置

restore_command (string)

要执行的 shell 命令，用于取回 WAL 文件系列中的一个已归档段。这个参数是归档恢复所必需的，但对流复制是可选的。字符串中的任何 %f 都会被替换为要从归档中取回的文件名，而任何 %p 都会被替换为服务器上的副本目标路径名。（路径名相对于当前工作目录，即集簇的数据目录。）任何 %r 都会被替换为包含最后一个有效重启点的文件的名称。这是允许恢复可重启而必须保留的最早文件，
因此这一信息可以用来把归档截断到仅包含支持从当前恢复重启所需的最小内容。%r 通常只被温备配置使用（见第 25.2 节）。要嵌入一个实际的 % 字符，可以写 %%。

只有当命令成功时才返回零退出状态，这一点很重要。系统会请求归档中不存在的文件名；遇到这种情况时它必须返回非零值。例子：

```
restore_command = 'cp /mnt/server/archivedir/%f "%p"'
restore_command = 'copy "C:\\server\\archivedir\\%f" "%p"'  #
 Windows
```

archive_cleanup_command (string)

这个可选参数指定了一个 shell 命令，它将在每一个重启点被执行。archive_cleanup_command的目的是提供一种清除不再被备库需要的旧的已归档 WAL 文件的机制。任何%r会被替换为包含最后一个可用重启点的文件的名称。
那是使一次恢复变成可重启的所必须被保留的最早的文件，并且因此比%r
更早的所有文件可以被安全地移除。
这个信息可以被用来把归档截断为支持从当前恢复重启所需的最小值。对于单一备库配置，pg_archivecleanup模块常常被用在archive_cleanup_command中，例如：

```
archive_cleanup_command = 'pg_archivecleanup /mnt/server/archivedir
 %r'
```

但是注意，如果多个备库正在从同一个归档目录中恢复，
你将需要保证只有当所有服务器都不再需要这些 WAL 文件时才会删除它们。archive_cleanup_command通常被用于一种温备配置（见第 25.2 节）中。要在该命令中嵌入一个真正的%字符，需要写成%%。

如果命令返回非零退出状态，将写入一条 WARNING 日志消息。

recovery_end_command (string)

此参数指定一个 shell 命令，在恢复结束时仅执行一次。此参数可选。recovery_end_command 用于提供复制或恢复后的清理机制。与archive_cleanup_command一样，所有 %r 都会被替换为包含最后一个有效重启点的文件的名称。

如果命令返回非零退出状态，将写入一条 WARNING 日志消息，但数据库仍会继续启动。例外是，如果命令被信号终止，数据库将不会继续启动。

26.2. 恢复目标设置

`recovery_target_name` (`string`)

这个参数指定（`pg_create_restore_point()`所创建）的已命名的恢复点，恢复将进行到该恢复点。在 `recovery_target_name`、recovery_target_time或recovery_target_xid中，最多只能指定一个。默认行为是恢复到 WAL 日志的末尾。

`recovery_target_time` (`timestamp`)

此参数指定恢复要进行到的时间戳。在 `recovery_target_time`、recovery_target_name或recovery_target_xid中，最多只能指定一个。默认行为是恢复到 WAL 日志的末尾。精确的停止点还受recovery_target_inclusive影响。

`recovery_target_xid` (`string`)

此参数指定恢复要进行到的事务 ID。请注意，虽然事务 ID 在事务开始时按顺序分配，但事务完成的顺序可能与 ID 的数值顺序不同。恢复的事务是那些在指定事务之前提交的事务，并可选择包括指定事务本身。在 `recovery_target_xid`、recovery_target_name或recovery_target_time中，最多只能指定一个。默认行为是恢复到 WAL 日志的末尾。精确的停止点还受recovery_target_inclusive影响。

`recovery_target_inclusive` (`boolean`)

指定是在指定的恢复目标之后停止（`true`），还是在恢复目标之前停止（`false`）。适用于recovery_target_time和recovery_target_xid中为此恢复指定的那一个。它分别表示提交时间或 ID 恰好等于目标值的事务是否会被包含在恢复中。默认值是 `true`。

`recovery_target_timeline` (`string`)

指定恢复到某条特定时间线。默认沿着制作基础备份时的当前时间线恢复。设为 `latest` 时，会恢复到归档中找到的最新时间线，这对备库很有用。除此之外，只有在复杂的再次恢复场景中才需要设置此参数：需要返回的状态本身就是一次时间点恢复之后达到的状态。相关讨论参见第 24.3.4 节。

`pause_at_recovery_target` (`boolean`)

指定在达到恢复目标时恢复是否应暂停。默认值为 true。这是为了允许对数据库执行查询，以检查这个恢复目标是否为最合适的恢复位置。暂停的状态可以使用`pg_xlog_replay_resume()`（见表 9.59）继续，这会让恢复终结。如果这个恢复目标不是想要的停止点，那么关闭服务器，将恢复目标设置改为一个稍后的目标并且重启以继续恢复。

如果未启用hot_standby，或者未设置恢复目标，此设置不起作用。

26.3. 备库设置

`standby_mode` (`boolean`)

指定是否将 PostgreSQL 服务器作为备库启动。如果此参数为 `on`，服务器到达已归档 WAL 的末尾时不会停止恢复，而会继续尝试使用 `restore_command` 获取新的 WAL 段，和/或按照 `primary_conninfo` 设置连接主库，以继续恢复。

primary_conninfo (string)

指定备库用来连接主库的连接串。该串采用 libpq 的 PQconnectdb 函数所接受的格式，见第 31.1 节。如果串中未指定某个选项，则会检查相应的环境变量（见第 31.13 节）。如果环境变量也未设置，则使用默认值。

连接串应当指定主服务器的主机名（或地址），如果端口与备库的默认端口不同，还应指定端口号。还要指定一个用户名，它对应主库上一个具有 REPLICATION 和 LOGIN 权限的角色（见第 25.2.5.1 节）。如果主库要求密码认证，还需要提供密码。密码可以放在 primary_conninfo 串中，也可以放在备库服务器上一个单独的 ~/.pgpass 文件中（数据库名使用 replication）。不要在 primary_conninfo 串中指定数据库名。

如果 standby_mode 为 off，此设置不起作用。

trigger_file (string)

指定一个触发文件，其出现会使备库结束恢复。即使未设置此值，也仍可以使用 pg_ctl promote 提升备库。如果 standby_mode 为 off，此设置不起作用。

第 27 章 监控数据库活动

数据库管理员常常会问："系统现在正在做什么？"本章将讨论如何回答这个问题。

有几种工具可用于监控数据库活动和分析性能。
本章的大部分内容将用于描述PostgreSQL的统计收集器，但也不应忽略常见的 Unix 监控程序，例如ps、top、iostat和vmstat。此外，一旦确定了某个查询性能不佳，可能需要使用PostgreSQL的EXPLAIN命令进行进一步调查。第 14.1 节讨论了EXPLAIN以及理解单个查询行为的其他方法。

27.1. 标准 Unix 工具

在大多数 Unix 平台上，PostgreSQL会修改由ps报告的命令标题，以便轻松识别各个服务器进程。以下是一个显示示例：

```
$ ps auxww | grep ^postgres
postgres   960  0.0  1.1  6104 1480 pts/1    SN   13:17   0:00
 postgres -i
postgres   963  0.0  1.1  7084 1472 pts/1    SN   13:17   0:00
 postgres: writer process
postgres   965  0.0  1.1  6152 1512 pts/1    SN   13:17   0:00
 postgres: stats collector process
postgres   998  0.0  2.3  6532 2992 pts/1    SN   13:18   0:00
 postgres: tgl runbug 127.0.0.1 idle
postgres  1003  0.0  2.4  6532 3128 pts/1    SN   13:19   0:00
 postgres: tgl regression [local] SELECT waiting
postgres  1016  0.1  2.4  6532 3080 pts/1    SN   13:19   0:00
 postgres: tgl regression [local] idle in transaction
```

（ps的适当调用方式随平台而异，显示内容的细节也是如此。此示例来自较新的 Linux 系统。）
这里列出的第一个进程是主服务器进程。显示的命令参数与启动该进程时使用的参数相同。
接下来的两个进程是主进程自动启动的后台工作进程。（如果系统被设置为不启动统计收集器，则不会出现"stats collector"进程。）其余每个进程都是处理一个客户端连接的服务器进程。每个这样的进程都将其命令行显示设置为以下形式：

postgres: *user database host activity*

在客户端连接的整个生命周期内，用户、数据库和（客户端）主机这几项保持不变，
但活动指示符会改变。活动可以是idle（即等待客户端命令）、idle in transaction（在BEGIN块内等待客户端），或者命令类型名称，例如SELECT。此外，
服务器进程当前若正在等待另一会话持有的锁，还会附加waiting。在上面的示例中，
可以推断进程 1003 正在等待进程 1016 完成事务，从而释放某个锁。

如果您关闭了update_process_title，则活动指示器不会更新；
进程标题仅在新进程启动时设置一次。在某些平台上，这可以节省可观的每命令开销；
在其他平台上，则微不足道。

提示

Solaris需要特殊处理。您必须使用/usr/ucb/ps，而不是/bin/ps。
您还必须使用两个w标志，而不仅仅是一个。此外，您对postgres

命令的原始调用必须具有比每个服务器进程提供的更短的ps状态显示。
如果未满足这三个条件，每个服务器进程的ps输出都会显示原始的postgres命令行。

27.2. 统计收集器

PostgreSQL的统计收集器是一个支持收集和报告服务器活动信息的子系统。目前，
对表和索引的访问会按磁盘块和行两种口径计数。每个表中的总行数，
以及每个表的清理和分析操作的信息也被计数。
它还可以统计对用户定义函数的调用和每个函数中花费的总时间。

PostgreSQL还支持报告其他服务器进程当前正在执行的确切命令。此功能独立于收集器进程。

27.2.1. 统计信息收集配置

由于收集统计信息会增加查询执行的开销，因此可以配置系统是否收集信息。这由通常在postgresql.conf中设置的配置参数控制（有关设置配置参数的详细信息，请参阅第 18 章）。

参数track_counts控制是否收集关于表和索引访问的统计信息。

参数track_functions启用对用户定义函数使用的跟踪。

参数track_activities启用对任意服务器进程当前执行命令的监控。

通常，这些参数会设置在postgresql.conf中，以便它们适用于所有服务器进程，
但也可以在单个会话中使用SET命令打开或关闭它们。
（为防止普通用户隐藏其活动不被管理员发现，只有超级用户才能使用SET更改这些参数。）

统计收集器通过临时文件把收集到的信息传递给后端（包括自动清理）。这些文件存放在pg_stat_tmp子目录中。当 postmaster 关闭时，统计信息数据的永久副本会存储在global子目录中。为了获得更高的性能，可以把stats_temp_directory参数指向基于 RAM 的文件系统，以减少物理 I/O 需求。

27.2.2. 查看收集到的统计信息

有几个预定义视图列在表 27.1中，可用于显示统计信息收集的结果。也可以使用底层的统计函数构建自定义视图。

当使用统计信息来监视当前活动时，重要的是要意识到信息不会立即更新。
每个单独的服务器进程会在转为空闲之前把新的统计计数发送给收集器，
因此仍在进行中的查询或事务不会影响显示的总计。并且，收集器本身最多每隔PGSTAT_STAT_INTERVAL毫秒才会发出一份新报告一次（除非在构建服务器时做过修改，否则为 500）。因此，显示的信息会滞后于实际活动。然而，由track_activities
收集的当前查询信息始终是最新的。

另一个重要点是，当要求服务器进程显示这些统计信息中的任何一种时，
它会先取回收集器进程最近发出的报告，然后在当前事务结束之前，
对所有统计视图和函数一直使用这一快照。因此，只要你继续当前事务，
统计信息就会显示静态信息。类似地，在事务中第一次请求这类当前查询信息时，
就会收集所有会话当前查询的信息，并且在整个事务期间显示的都是同一份信息。
这是一个特性，而不是一个错误，因为它允许你对统计信息执行多个查询并对结果做关联，
而不必担心数字在你眼前变化。但如果希望每个查询都看到新的结果，
请确保在任何事务块之外执行这些查询。或者，也可以调用pg_stat_clear_snapshot()，
它会丢弃当前事务的统计快照（如果有）。下一次使用统计信息时，就会获取一个新的快照。

A transaction can also see its own statistics (as yet untransmitted to the collector) in the views pg_stat_xact_all_tables, pg_stat_xact_sys_tables, pg_stat_xact_user_tables, and pg_stat_xact_user_functions中，或者通过这些视图的底层函数看到自己的统计信息（这些统计信息尚未传送给收集器）。这些数字不像上面所述那样起作用；相反，它们在事务期间持续更新。

表 27.1. 标准统计视图

视图名称	Description
pg_stat_activity	每个服务器进程一行，显示数据库 OID、数据库名、进程 ID、用户 OID、用户名、应用名、客户端地址、主机名（如果可用）以及端口号、服务器进程、当前事务和当前查询开始执行的时间、进程的等待状态以及当前查询的文本。报告当前查询数据的列在参数track_activities被关闭时不可用。此外，只有查看视图的用户是超级用户或者与被报告进程的属主相同时，这些列才可见。客户端的主机名只有在设置了log_hostname或者在pg_hba.conf处理过程中需要查找用户的主机名时才可用。
pg_stat_bgwriter	仅一行，显示来自后台写入器的集簇范围统计信息：已调度的检查点数量、被请求的检查点数量、由检查点和清理扫描写出的缓冲区，以及后台写入器因已写出过多缓冲区而停止清理扫描的次数。还包括有关共享缓冲池的统计信息，包括由后端（即非后台写入器）写出的缓冲区、这些后端不得不执行它们自己的 fsync 调用的次数（通常即使后端自己做写操作，后台写入器也会处理这些调用）、分配的总缓冲区数以及最近一次统计信息重置的时间。
pg_stat_database	每个数据库一行，显示数据库 OID、数据库名、连接到该数据库的活动服务器进程数、在该数据库中提交和回滚的事务数、总磁盘块读取数、总缓冲区命中数（即因为块已在缓冲区缓存中被找到而避免的块读取请求数）、返回、抓取、插入、更新和删除的行数、因与恢复冲突（在备库上）而被取消的查询总数，以及最近一次统计信息重置的时间。
pg_stat_database_conflicts	每个数据库一行，显示数据库 OID、数据库名以及该数据库中由于被删除的表空间、锁超时、旧快照、被固定的缓冲区和死锁而被取消的查询数。只在备库上包含信息，因为主库上不会发生冲突。
pg_stat_replication	每个 WAL 发送进程一行，显示进程 ID、用户 OID、用户名、应用名、客户端地址、主机名（如果可用）和端口号、服务器进程开始执行的时间，以及当前 WAL 发送进程状态和事务日志位置。此外，备库会报告它收到并写入、刷出到磁盘以及重放的最近事务日志位置，这些信息也显示在这里。如果备库的应用名匹配synchronous_standby_names中的某一项设置，这里还会显示 sync_priority，即备库成为同步备库的顺序。详细说明连接具体在做什么的列只有当查看视图的用户是超级用户时才可见。客户端的主机名只有在设置了log_hostname或者在pg_hba.conf处理过程中需要查找用户的主机名时才可用。
pg_stat_all_tables	对当前数据库中的每个表（包括 TOAST 表）显示表 OID、模式和表名、发起的顺序扫描数、由顺序扫描抓取的活行数、发起的索引扫描数（在该表所有索引上）、由索引扫描抓取的活行数、插入、更新和

视图名称	Description
	删除的行数、HOT 更新（即没有单独索引更新）的行数、活行和死行数、该表最近一次被手工非FULL清理的时间、最近一次被自动清理守护进程清理的时间、最近一次被手工分析的时间、最近一次被自动清理守护进程分析的时间、被手工非FULL清理的次数、被自动清理守护进程清理的次数、被手工分析的次数以及被自动清理守护进程分析的次数。
pg_stat_sys_tables	和pg_stat_all_tables一样，但只显示系统表。
pg_stat_user_tables	和pg_stat_all_tables一样，但只显示用户表。
pg_stat_xact_all_tables	与 pg_stat_all_tables 相似，但只统计当前事务中截至目前执行的操作（这些操作尚未计入 pg_stat_all_tables 及相关视图）。此视图不包含存活行和死行数量以及清理、分析操作的列。
pg_stat_xact_sys_tables	和pg_stat_xact_all_tables一样，但只显示系统表。
pg_stat_xact_user_tables	和pg_stat_xact_all_tables一样，但只显示用户表。
pg_stat_all_indexes	当前数据库中的每个索引一行，显示表和索引 OID、模式、表和索引名、在该索引上发起的索引扫描次数、索引扫描返回的索引条目数，以及使用该索引的简单索引扫描取出的活表行数。
pg_stat_sys_indexes	和pg_stat_all_indexes一样，但只显示系统表上的索引。
pg_stat_user_indexes	和pg_stat_all_indexes一样，但只显示用户表上的索引。
pg_statio_all_tables	当前数据库中的每个表（包括 TOAST 表）一行，显示表 OID、模式和表名、从该表读取的磁盘块数、缓冲区命中数、该表所有索引的磁盘块读取数和缓冲区命中数、该表辅助 TOAST 表（若有）的磁盘块读取数和缓冲区命中数，以及 TOAST 表索引的磁盘块读取数和缓冲区命中数。
pg_statio_sys_tables	和pg_statio_all_tables一样，但只显示系统表。
pg_statio_user_tables	和pg_statio_all_tables一样，但只显示用户表。
pg_statio_all_indexes	当前数据库中的每个索引一行，显示表和索引 OID、模式、表和索引名，以及该索引中的磁盘块读取数和缓冲区命中数。
pg_statio_sys_indexes	和pg_statio_all_indexes一样，但只显示系统表上的索引。
pg_statio_user_indexes	和pg_statio_all_indexes一样，但只显示用户表上的索引。
pg_statio_all_sequences	当前数据库中的每个序列对象一行，显示序列 OID、模式和序列名，以及该序列中的磁盘块读取数和缓冲区命中数。
pg_statio_sys_sequences	和pg_statio_all_sequences一样，但只显示系统序列（目前没有定义系统序列，因此这个视图总是为空）。
pg_statio_user_sequences	和pg_statio_all_sequences一样，但只显示用户序列。

视图名称	Description
pg_stat_user_functions	对所有被跟踪的函数，显示函数 OID、模式、名称、调用次数、总时间和自身时间。自身时间是花在该函数本身的时间，总时间包括花在它调用的函数中的时间。时间值以毫秒计。
pg_stat_xact_user_functions	和pg_stat_user_functions相似，但是只统计在当前事务期间的调用（还没有被包括在pg_stat_user_functions中）。

每个索引的统计信息对于判断哪些索引正在被使用以及它们有多有效尤其有用。

索引既可以被直接使用，也可以通过“位图扫描”使用。在位图扫描中，多个索引的输出可以通过 AND 或 OR 规则组合；因此当使用位图扫描时，很难把单个堆行抓取与特定的索引关联起来。Therefore, a bitmap scan increments the pg_stat_all_indexes.idx_tup_read count(s) for the index(es) it uses, and it increments the pg_stat_all_tables.idx_tup_fetch count for the table, but it does not affect pg_stat_all_indexes.idx_tup_fetch.

> **注意**
>
> 在PostgreSQL 8.1 之前，idx_tup_read和idx_tup_fetch计数基本总是相等的。现在，即使不考虑位图扫描，它们也可能不同，因为idx_tup_read
> 统计从索引中取出的索引项，而idx_tup_fetch统计从表中取出的活行；
> 如果使用索引取出了任何死行或尚未提交的行，后者就会更小。

pg_statio_视图主要用于判断缓冲区缓存的有效性。
如果实际磁盘读取次数远少于缓冲区命中次数，缓存就能满足大多数读取请求，
而无需调用内核。不过，这些统计信息并不能反映全部情况：由于PostgreSQL处理磁盘 I/O 的方式，不在PostgreSQL缓冲区缓存中的数据仍可能驻留在内核的 I/O 缓存中，因而仍可在不执行物理读取的情况下取得。对于希望更详细地了解PostgreSQL I/O 行为的用户，建议将PostgreSQL统计收集器与能够深入查看内核 I/O 处理情况的操作系统工具结合使用。

其他查看统计信息的方法是编写查询，使用与这些标准视图相同的底层统计信息访问函数。
这些函数列在表 27.2中。针对每个数据库的访问函数把一个数据库 OID 作为参数来标识要报告哪个数据库。针对每个表和每个索引的函数要求表或索引 OID。针对函数调用统计的函数用一个函数 OID。（注意只有当前数据库中的表、索引和函数才能被这些函数看到。）针对每个服务器进程的访问函数要求一个服务器进程编号，范围从 1 到当前活动服务器进程的数目。

表 27.2. 统计访问函数

函数	返回类型	Description
pg_stat_get_db_numbackends(oid)	integer	数据库的活动服务器进程数
pg_stat_get_db_xact_commit(oid)	bigint	数据库中已提交的事务数
pg_stat_get_db_xact_rollback(oid)	bigint	数据库中已回滚的事务数
pg_stat_get_db_blocks_fetched(oid)	bigint	数据库的磁盘块读取请求数
pg_stat_get_db_blocks_hit(oid)	bigint	数据库在缓存中找到的磁盘块读取请求数
pg_stat_get_db_tuples_returned(oid)	bigint	数据库返回的元组数

函数	返回类型	Description
pg_stat_get_db_tuples_fetched(oid)	bigint	数据库获取的元组数
pg_stat_get_db_tuples_inserted(oid)	bigint	数据库中插入的元组数
pg_stat_get_db_tuples_updated(oid)	bigint	数据库中更新的元组数
pg_stat_get_db_tuples_deleted(oid)	bigint	数据库中删除的元组数
pg_stat_get_db_conflict_tablespace(oid)	bigint	由于与被删除表空间的恢复冲突而在数据库中被取消的查询数
pg_stat_get_db_conflict_lock(oid)	bigint	由于与锁的恢复冲突而在数据库中被取消的查询数
pg_stat_get_db_conflict_snapshot(oid)	bigint	由于与旧快照的恢复冲突而在数据库中被取消的查询数
pg_stat_get_db_conflict_bufferpin(oid)	bigint	由于与被固定缓冲区的恢复冲突而在数据库中被取消的查询数
pg_stat_get_db_conflict_startup_deadlock(oid)	bigint	由于与死锁的恢复冲突而在数据库中被取消的查询数
pg_stat_get_db_stat_reset_time(oid)	timestamptz	该数据库最近一次统计信息重置的时间。在每个数据库第一次连接时初始化为系统时间。当在该数据库上调用pg_stat_reset以及对其中任意表或索引执行pg_stat_reset_single_table_counters时，重置时间会被更新。
pg_stat_get_numscans(oid)	bigint	参数为表时是完成的顺序扫描数，参数为索引时是完成的索引扫描数
pg_stat_get_tuples_returned(oid)	bigint	参数为表时是顺序扫描读取的行数，参数为索引时是返回的索引条目数
pg_stat_get_tuples_fetched(oid)	bigint	参数为表时是位图扫描取出的表行数，参数为索引时是使用该索引的简单索引扫描取出的表行数
pg_stat_get_tuples_inserted(oid)	bigint	插入表中的行数
pg_stat_get_tuples_updated(oid)	bigint	表中更新的行数（包括 HOT 更新）
pg_stat_get_tuples_deleted(oid)	bigint	从表中删除的行数
pg_stat_get_tuples_hot_updated(oid)	bigint	表中被 HOT 更新的行数
pg_stat_get_live_tuples(oid)	bigint	表中的活行数
pg_stat_get_dead_tuples(oid)	bigint	表中的死行数
pg_stat_get_blocks_fetched(oid)	bigint	表或索引的磁盘块读取请求数

函数	返回类型	Description
pg_stat_get_blocks_hit(oid)	bigint	表或索引在缓存中找到的磁盘块请求数
pg_stat_get_last_vacuum_time(oid)	timestamptz	用户在此表上发起的最近一次非FULL清理的时间
pg_stat_get_last_autovacuum_time(oid)	timestamptz	该表最近一次由自动清理守护进程发起的清理的时间
pg_stat_get_last_analyze_time(oid)	timestamptz	该表最近一次由用户发起的分析的时间
pg_stat_get_last_autoanalyze_time(oid)	timestamptz	该表最近一次由自动清理守护进程发起的分析的时间
pg_stat_get_vacuum_count(oid)	bigint	此表被手工非FULL清理的次数
pg_stat_get_autovacuum_count(oid)	bigint	此表被自动清理守护进程清理的次数
pg_stat_get_analyze_count(oid)	bigint	此表被手工分析的次数
pg_stat_get_autoanalyze_count(oid)	bigint	此表被自动清理守护进程分析的次数
pg_stat_get_xact_numscans(oid)	bigint	在当前事务中，当参数是表时完成的顺序扫描数，或当参数是索引时完成的索引扫描数
pg_stat_get_xact_tuples_returned(oid)	bigint	在当前事务中，当参数是表时由顺序扫描读取的行数，或当参数是索引时返回的索引项数
pg_stat_get_xact_tuples_fetched(oid)	bigint	在当前事务中，当参数是表时由位图扫描抓取的表行数，或当参数是索引时由使用该索引的简单索引扫描抓取的表行数
pg_stat_get_xact_tuples_inserted(oid)	bigint	在当前事务中插入表的行数
pg_stat_get_xact_tuples_updated(oid)	bigint	在当前事务中更新的表行数（包括HOT 更新）
pg_stat_get_xact_tuples_deleted(oid)	bigint	在当前事务中从表删除的行数
pg_stat_get_xact_tuples_hot_updated(oid)	bigint	在当前事务中表中被 HOT 更新的行数
pg_stat_get_xact_blocks_fetched(oid)	bigint	在当前事务中为表或索引发出的磁盘块读取请求数
pg_stat_get_xact_blocks_hit(oid)	bigint	在当前事务中在缓存中找到的表或索引磁盘块请求数
pg_backend_pid()	integer	附属于当前会话的服务器进程的进程 ID
pg_stat_get_activity(integer)	setof record	返回一条记录，包含指定 PID 的后端的信息；如果指定NULL，则为系统中的每个活动后端返回一条记录。返回字段是pg_stat_activity视图字段的子集。

函数	返回类型	Description
pg_stat_get_function_calls(oid)	bigint	该函数被调用的次数
pg_stat_get_function_time(oid)	bigint	花在该函数上的总挂钟时间，以微秒计。包括花在它调用的函数中的时间。
pg_stat_get_function_self_time(oid)	bigint	仅花在该函数本身上的时间。不包括花在被调用函数中的时间。
pg_stat_get_xact_function_calls(oid)	bigint	在当前事务中该函数被调用的次数。
pg_stat_get_xact_function_time(oid)	bigint	在当前事务中在该函数中花费的总挂钟时间，以微秒计。包括它调用的函数所花的时间。
pg_stat_get_xact_function_self_time(oid)	bigint	在当前事务中仅在此函数中花费的时间。被调用函数所花的时间被排除在外。
pg_stat_get_backend_idset()	setof integer	当前活动服务器进程编号的集合（从 1 到活动服务器进程数）。用法示例见正文。
pg_stat_get_backend_pid(integer)	integer	给定服务器进程的进程 ID
pg_stat_get_backend_dbid(integer)	oid	给定服务器进程的数据库 ID
pg_stat_get_backend_userid(integer)	oid	给定服务器进程的用户 ID
pg_stat_get_backend_activity(integer)	text	给定服务器进程的活动命令，但仅当当前用户是超级用户或与被查询会话的用户相同（且track_activities已启用）时可见
pg_stat_get_backend_waiting(integer)	boolean	如果给定服务器进程正在等待锁则为真，但仅当当前用户是超级用户或与被查询会话的用户相同（且track_activities已启用）时可见
pg_stat_get_backend_activity_start(integer)	timestamp with time zone	给定服务器进程当前执行的查询的启动时间，但仅当当前用户是超级用户或与被查询会话的用户相同（且track_activities已启用）时可见
pg_stat_get_backend_xact_start(integer)	timestamp with time zone	给定服务器进程当前执行的事务的启动时间，但仅当当前用户是超级用户或与被查询会话的用户相同（且track_activities已启用）时可见
pg_stat_get_backend_start(integer)	timestamp with time zone	给定服务器进程的启动时间；如果当前用户既不是超级用户也不是被查询会话的用户，则为 null
pg_stat_get_backend_client_addr(integer)	inet	连接到给定服务器进程的客户端的 IP 地址；如果连接是通过 Unix 域套接字建立的，或者当前用户既不是超级用户也不是被查询会话的用户，则为 null

函数	返回类型	Description
pg_stat_get_backend_client_port(integer)	integer	连接到给定服务器进程的客户端的 TCP 端口号；如果连接是通过 Unix 域套接字建立的则为 -1；如果当前用户既不是超级用户也不是被查询会话的用户，则为 null
pg_stat_get_bgwriter_timed_checkpoints()	bigint	后台写进程启动的定时检查点次数 (because the checkpoint_timeout time has expired)
pg_stat_get_bgwriter_requested_checkpoints()	bigint	后台写进程因超出checkpoint_segments或因发出了CHECKPOINT命令而根据后端请求启动检查点的次数
pg_stat_get_bgwriter_buf_written_checkpoints()	bigint	检查点期间后台写进程写出的缓冲区数
pg_stat_get_bgwriter_buf_written_clean()	bigint	后台写进程为例行清理脏页而写出的缓冲区数
pg_stat_get_bgwriter_maxwritten_clean()	bigint	后台写进程因写出的缓冲区数超过bgwriter_lru_maxpages参数指定值而中止其清理扫描的次数
pg_stat_get_bgwriter_stat_reset_time()	timestamptz	后台写入器最近一次统计信息重置的时间，在数据库集簇上执行pg_stat_reset_shared('bgwriter')时更新。
pg_stat_get_buf_written_backend()	bigint	后端因需要分配新缓冲区而写出的缓冲区数
pg_stat_get_buf_alloc()	bigint	缓冲区分配总数
pg_stat_get_wal_senders()	setof record	每个活动 wal 发送进程一条记录。返回的字段是pg_stat_replication视图字段的一个子集。
pg_stat_clear_snapshot()	void	丢弃当前统计信息快照
pg_stat_reset()	void	将当前数据库的所有统计计数器重置为零（需要超级用户权限）
pg_stat_reset_shared(text)	void	将数据库集簇的部分共享统计计数器重置为零（需要超级用户权限）。调用pg_stat_reset_shared('bgwriter')将把pg_stat_bgwriter显示的所有值清零。
pg_stat_reset_single_table_counters(oid)	void	将当前数据库中单个表或索引的统计信息重置为零（需要超级用户权限）
pg_stat_reset_single_function_counters(oid)	void	将当前数据库中单个函数的统计信息重置为零（需要超级用户权限）

注意

pg_stat_get_blocks_fetched减去pg_stat_get_blocks_hit得到为该表、索引或数据库发出的内核read()调用次数；由于内核级的缓冲，

> 实际物理读取的次数通常更低。*_blks_read统计列使用这个减法，
> 即取出的减去命中的。

所有访问后端信息的函数都以 backend id 编号为索引，唯一例外是按 PID 索引的pg_stat_get_activity。函数pg_stat_get_backend_idset
提供了一种方便的方法为每个活动服务器进程产生一行。例如，要显示所有服务器进程的
PID和当前查询：

```
SELECT pg_stat_get_backend_pid(s.backendid) AS procpid,
       pg_stat_get_backend_activity(s.backendid) AS current_query
    FROM (SELECT pg_stat_get_backend_idset() AS backendid) AS s;
```

27.3. 查看锁

监控数据库活动的另外一个有用的工具是pg_locks系统表。
这样就允许数据库管理员查看锁管理器中当前存在的锁的信息。例如，这个功能可以被用于：

- 查看当前存在的所有锁、在一个特定数据库中的关系上所有的锁、
  在一个特定关系上所有的锁，或者由一个特定PostgreSQL会话持有的所有的锁。
- 判断当前数据库中带有最多未授予锁的关系（它可能是数据库客户端之间的竞争来源）。
- 判断锁竞争给数据库总体性能带来的影响，以及锁竞争随着整个数据库流量的变化范围。

pg_locks视图的细节在第 45.56 节中。更多有关PostgreSQL的锁和管理并发性的信息，请参考第 13 章。

27.4. 动态追踪

PostgreSQL提供了功能来支持数据库服务器的动态追踪。
这样就允许在代码中的特定点上调用外部工具来追踪执行过程。

一些探针或追踪点已经被插入在源代码中。这些探针的目的是被数据库开发者和管理员使用。
默认情况下，探针不被编译到PostgreSQL中；用户需要显式地告诉配置脚本使得探针可用。

当前，只支持DTrace[1]工具，它在 OpenSolaris、Solaris 10 和 Mac OS X Leopard 上可用。预期 DTrace 将来会出现在 FreeBSD 以及可能其他操作系统上。Linux 上的SystemTap[2]项目也提供了一个 DTrace 的等价物。通过修改src/include/utils/probes.h中宏的定义，理论上也可以支持其他动态追踪工具。

27.4.1. 为动态追踪编译

默认情况下，探针是不可用的，因此你将需要显式地告诉配置脚本让探针在PostgreSQL中可用。
要包括 DTrace 支持，在配置时指定--enable-dtrace。更多信息请见第 15.4 节。

27.4.2. 内置探针

如表 27.3所示，源代码中提供了一些标准探针。表 27.4显示了在探针中使用的类型。当然，可以增加更多探针来增强PostgreSQL的可观测性。

[1] http://opensolaris.org/os/community/dtrace/
[2] http://sourceware.org/systemtap/

表 27.3. 内置 DTrace 探针

名称	参数	Description
transaction-start	(LocalTransactionId)	在一个新事务开始时触发的探针。arg0 是事务 ID。
transaction-commit	(LocalTransactionId)	在一个事务成功完成时触发的探针。arg0 是事务 ID。
transaction-abort	(LocalTransactionId)	当一个事务失败结束时触发的探针。arg0 是事务 ID。
query-start	(const char *)	当一个查询的处理开始时触发的探针。arg0 是查询字符串。
query-done	(const char *)	当一个查询的处理完成时触发的探针。arg0 是查询字符串。
query-parse-start	(const char *)	当一个查询的解析开始时触发的探针。arg0 是查询字符串。
query-parse-done	(const char *)	当一个查询的解析完成时触发的探针。arg0 是查询字符串。
query-rewrite-start	(const char *)	当一个查询的重写开始时触发的探针。arg0 是查询字符串。
query-rewrite-done	(const char *)	当一个查询的重写完成时触发的探针。arg0 是查询字符串。
query-plan-start	()	当一个查询的规划开始时触发的探针。
query-plan-done	()	当一个查询的规划完成时触发的探针。
query-execute-start	()	当一个查询的执行开始时触发的探针。
query-execute-done	()	当一个查询的执行完成时触发的探针。
statement-status	(const char *)	服务器进程更新其`pg_stat_activity.current_query`状态时触发的探针。arg0 是新的状态字符串。
checkpoint-start	(int)	当一个检查点开始时触发的探针。arg0 传递位标志来区分不同的检查点类型，例如关闭（shutdown）、立即（immediate）或强制（force）。
checkpoint-done	(int, int, int, int, int)	当一个检查点完成时触发的探针（下面列出的探针会在检查点处理过程中依次触发）。arg0 是已写入的缓冲区数量。arg1 是缓冲区的总数。arg2、arg3 和 arg4 分别包含了增加、删除和循环回收的 WAL 文件的数量。
clog-checkpoint-start	(bool)	当一个检查点的 CLOG 部分开始时触发的探针。arg0 为真表示正常检查点，为假表示关闭检查点。
clog-checkpoint-done	(bool)	当一个检查点的 CLOG 部分完成时触发的探针。arg0 的含义与clog-checkpoint-start中相同。

名称	参数	Description
subtrans-checkpoint-start	(bool)	当一个检查点的 SUBTRANS 部分开始时触发的探针。arg0 为真表示正常检查点，为假表示关闭检查点。
subtrans-checkpoint-done	(bool)	当一个检查点的 SUBTRANS 部分完成时触发的探针。arg0 的含义与subtrans-checkpoint-start中相同。
multixact-checkpoint-start	(bool)	当一个检查点的 MultiXact 部分开始时触发的探针。arg0 为真表示正常检查点，为假表示关闭检查点。
multixact-checkpoint-done	(bool)	当一个检查点的 MultiXact 部分完成时触发的探针。arg0 的含义与multixact-checkpoint-start中相同。
buffer-checkpoint-start	(int)	当一个检查点的写缓冲区部分开始时触发的探针。arg0 传递位标志来区分不同的检查点类型，例如关闭（shutdown）、立即（immediate）或强制（force）。
buffer-sync-start	(int, int)	当我们在检查点期间开始写脏缓冲区时（在标识哪些缓冲区必须被写之后）触发的探针。arg0 是缓冲区总数，arg1 是当前为脏并且需要被写的缓冲区数量。
buffer-sync-written	(int)	在检查点期间当每个缓冲区被写完之后触发的探针。arg0 是缓冲区的 ID。
buffer-sync-done	(int, int, int)	当所有脏缓冲区被写之后触发的探针。arg0 是缓冲区总数。arg1 是检查点进程实际写的缓冲区数量。arg2 是期望写的数目（buffer-sync-start的 arg1）；arg1 和 arg2 的任何的不同反映在该检查点期间有其他进程刷写了缓冲区。
buffer-checkpoint-sync-start	()	在脏缓冲区被写入到内核之后并且在开始发出 fsync 请求之前触发的探针。
buffer-checkpoint-done	()	当同步缓冲区到磁盘完成时触发的探针。
twophase-checkpoint-start	()	当一个检查点的两阶段部分开始时触发的探针。
twophase-checkpoint-done	()	当一个检查点的两阶段部分完成时触发的探针。
buffer-read-start	(ForkNumber, BlockNumber, Oid, Oid, Oid, int, bool)	当一次缓冲区读开始时触发的探针。arg0 和 arg1 包含该页的分支号和块号（如果这是一次关系扩展请求，arg1 为 -1）。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。对一个本地缓冲区，arg5 是创建临时关

名称	参数	Description
		系的后端的 ID；对于一个共享缓冲区，arg5 是 InvalidBackendId（-1）。arg6 为真表示一次关系扩展请求，为假表示正常读。
buffer-read-done	(ForkNumber, BlockNumber, Oid, Oid, Oid, int, bool, bool)	当一次缓冲区读完成时触发的探针。arg0 和 arg1 包含该页的分支号和块号（如果这是一次关系扩展请求，arg1 现在包含新增加块的块号）。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。对一个本地缓冲区，arg5 是创建临时关系的后端的 ID；对于一个共享缓冲区，arg5 是 InvalidBackendId（-1）。arg6 为真表示一次关系扩展请求，为假表示正常读。arg7 为真表示在池中找到该缓冲区，为假表示没有找到。
buffer-flush-start	(ForkNumber, BlockNumber, Oid, Oid, Oid)	在发出对一个共享缓冲区的任意写请求之前触发的探针。arg0 和 arg1 包含该页的分支号和块号。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。
buffer-flush-done	(ForkNumber, BlockNumber, Oid, Oid, Oid)	当一个写请求完成时触发的探针（注意这只反映传递数据给内核的时间，它通常并没有实际地被写入到磁盘）。参数和buffer-flush-start的相同。
buffer-write-dirty-start	(ForkNumber, BlockNumber, Oid, Oid, Oid)	当一个服务器进程开始写一个脏缓冲区时触发的探针（如果这经常发生，表示shared_buffers太小，或需要调整后台写入器的控制参数）。arg0 和 arg1 包含该页的分支号和块号。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。
buffer-write-dirty-done	(ForkNumber, BlockNumber, Oid, Oid, Oid)	当一次脏缓冲区写完成时触发的探针。参数与buffer-write-dirty-start相同。
wal-buffer-write-dirty-start	()	服务器进程因 WAL 缓冲区空间耗尽而开始写出脏 WAL 缓冲区时触发的探针。（如果这种情况经常发生，说明wal_buffers太小。）
wal-buffer-write-dirty-done	()	当一次脏 WAL 缓冲区写入完成时触发的探针。
xlog-insert	(unsigned char, unsigned char)	当一个 WAL 记录被插入时触发的探针。arg0 是该记录的资源管理器（rmid）。arg1 包含 info 标志。
xlog-switch	()	当请求一次 WAL 段切换时触发的探针。
smgr-md-read-start	(ForkNumber, BlockNumber, Oid, Oid, Oid, int)	当开始从一个关系读取一块时触发的探针。arg0 和 arg1 包含该页

名称	参数	Description
		的分支号和块号。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。对一个本地缓冲区，arg5 是创建临时关系的后端的 ID；对于一个共享缓冲区，arg5 是InvalidBackendId（-1）。
smgr-md-read-done	(ForkNumber, BlockNumber, Oid, Oid, Oid, int, int, int)	当一次块读取完成时触发的探针。arg0 和 arg1 包含该页的分支号和块号。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。对一个本地缓冲区，arg5 是创建临时关系的后端的 ID；对于一个共享缓冲区，arg5 是InvalidBackendId（-1）。arg6 是实际读取的字节数，而 arg7 是请求读取的字节数（如果两者不同就意味着麻烦）。
smgr-md-write-start	(ForkNumber, BlockNumber, Oid, Oid, Oid, int)	当开始向一个关系中写入一个块时触发的探针。arg0 和 arg1 包含该页的分支号和块号。arg2、arg3 和 arg4 包含表空间、数据库和关系 OID 用以识别该关系。对一个本地缓冲区，arg5 是创建临时关系的后端的 ID；对于一个共享缓冲区，arg5 是InvalidBackendId（-1）。
smgr-md-write-done	(ForkNumber, BlockNumber, Oid, Oid, Oid, int, int, int)	当一个块写操作完成时触发的探针。arg0 和 arg1 包含该页的分支号和块号。arg2、arg3和arg4 包含表空间、数据库和关系 OID来标识该关系。对于一个本地缓冲区，arg5 是创建临时关系的后端 ID；对于一个共享缓冲区，arg5 是InvalidBackendId（-1）。arg6 是实际写的字节数，而 arg7 是要求写的字节数（如果这两者不同，则意味着麻烦）。
sort-start	(int, bool, int, int, bool)	当一次排序操作开始时触发的探针。arg0 指示是堆排序、索引排序或 datum 排序。arg1 为真表示强制唯一值。arg2 是键列数。arg3 是允许使用的工作内存大小（以千字节计）。如果要求随机访问排序结果，则 arg4 为真。
sort-done	(bool, long)	当一次排序完成时触发的探针。arg0 为真表示外部排序，为假表示内排序。arg1 是用于一次外部排序的磁盘块的数目，或用于一次内排序的以千字节计的内存。
lwlock-acquire	(LWLockId, LWLockMode)	获取到一个 LWLock 时触发的探针。arg0 是该 LWLock 的 ID。arg1 是请求的锁模式，即排他或共享。

名称	参数	Description
lwlock-release	(LWLockId)	释放一个 LWLock 时触发的探针（但注意被释放的等待者尚未被唤醒）。arg0 是该 LWLock 的 ID。
lwlock-wait-start	(LWLockId, LWLockMode)	LWLock 不能立即获取、服务器进程开始等待该锁可用时触发的探针。arg0 是该 LWLock 的 ID。arg1 是请求的锁模式，即排他或共享。
lwlock-wait-done	(LWLockId, LWLockMode)	服务器进程结束对 LWLock 的等待时触发的探针（它实际上尚未持有该锁）。arg0 是该 LWLock 的 ID。arg1 是请求的锁模式，即排他或共享。
lwlock-condacquire	(LWLockId, LWLockMode)	调用者指定不等待却成功获取到 LWLock 时触发的探针。arg0 是该 LWLock 的 ID。arg1 是请求的锁模式，即排他或共享。
lwlock-condacquire-fail	(LWLockId, LWLockMode)	调用者指定不等待且未能成功获取 LWLock 时触发的探针。arg0 是该 LWLock 的 ID。arg1 是请求的锁模式，即排他或共享。
lock-wait-start	(unsigned int, unsigned int, unsigned int, unsigned int, unsigned int, LOCKMODE)	当一个重量级锁（lmgr锁）的请求由于锁不可用开始等待时触发的探针。arg0 到 arg3 是标识被锁定对象的标签字段。arg4 指示被锁对象的类型。arg5 表示被请求的锁类型。
lock-wait-done	(unsigned int, unsigned int, unsigned int, unsigned int, unsigned int, LOCKMODE)	当一个重量级锁（lmgr 锁）的请求结束等待时（即已经得到锁）触发的探针。参数与lock-wait-start一样。
deadlock-found	()	当死锁检测器发现死锁时触发的探针。

表 27.4. 探针参数中使用的已定义类型

Type	定义
LocalTransactionId	unsigned int
LWLockId	int
LWLockMode	int
LOCKMODE	int
BlockNumber	unsigned int
Oid	unsigned int
ForkNumber	int
bool	char

27.4.3. 使用探针

下面的示例展示了一个分析系统中事务计数的 DTrace 脚本，可以用来代替一次性能测试之前和之后的`pg_stat_database`快照：

```
#!/usr/sbin/dtrace -qs

postgresql$1:::transaction-start
{
      @start["Start"] = count();
      self->ts  = timestamp;
}

postgresql$1:::transaction-abort
{
      @abort["Abort"] = count();
}

postgresql$1:::transaction-commit
/self->ts/
{
      @commit["Commit"] = count();
      @time["Total time (ns)"] = sum(timestamp - self->ts);
      self->ts=0;
}
```

当被执行时，该示例 D 脚本给出这样的输出：

```
# ./txn_count.d `pgrep -n postgres` or ./txn_count.d <PID>
^C

Start                                          71
Commit                                         70
Total time (ns)                        2312105013
```

> **注意**
>
> SystemTap 的追踪脚本记法与 DTrace 不同，但底层探针是兼容的。需要注意的一点是，截至本文编写时，SystemTap 脚本必须使用双下划线来代替连字符引用探针名。预计未来的 SystemTap 版本会修复这一点。

你应该记住，DTrace 脚本需要细心地编写和调试，否则被收集的追踪信息可能会毫无意义。在大多数发现问题的情况下，出错的是插桩，而不是底层系统。当讨论使用动态追踪发现的信息时，一定要附上使用的脚本以便其也被检查和讨论。

更多示例脚本可以在 PgFoundry 的dtrace 项目[3]中找到。

27.4.4. 定义新探针

开发者可以在代码中任意位置定义新的探针，当然这要重新编译之后才能生效。
下面是插入新探针的步骤：

1. 决定探针名称以及要通过探针提供的数据
2. 把该探针定义加入到`src/backend/utils/probes.d`

[3] http://pgfoundry.org/projects/dtrace/

3. 如果pg_trace.h尚未被包含该探针点的模块引用，则将它包含进来，并且在源代码中期望的位置插入TRACE_POSTGRESQL探针宏

4. 重新编译并验证新探针是可用的

示例：. 这里是一个如何增加一个探针来用事务 ID 追踪所有新事务的示例。

1. 决定探针将被命名为transaction-start并且需要一个LocalTransactionId类型的参数

2. 将该探针定义加入到src/backend/utils/probes.d：

   ```
   probe transaction__start(LocalTransactionId);
   ```

   注意探针名字中双下划线的使用。在一个使用探针的 DTrace 脚本中，双下划线需要被替换为一个连字符，因此，文档中应向用户说明的名称是transaction-start。

3. 在编译时，transaction__start被转换成一个名为TRACE_POSTGRESQL_TRANSACTION_START的宏（注意这里是单下划线），可以通过包含头文件pg_trace.h获得。将宏调用加入到源代码中的合适位置。在这种情况下，看起来类似：

   ```
   TRACE_POSTGRESQL_TRANSACTION_START(vxid.localTransactionId);
   ```

4. 在重新编译和运行新的二进制文件之后，通过运行下面的 DTrace 命令来检查新增的探针是否可用。你应该看到类似下面的输出：

   ```
   # dtrace -ln transaction-start
      ID    PROVIDER          MODULE           FUNCTION NAME
   18705 postgresql49878     postgres     StartTransactionCommand
    transaction-start
   18755 postgresql49877     postgres     StartTransactionCommand
    transaction-start
   18805 postgresql49876     postgres     StartTransactionCommand
    transaction-start
   18855 postgresql49875     postgres     StartTransactionCommand
    transaction-start
   18986 postgresql49873     postgres     StartTransactionCommand
    transaction-start
   ```

向C代码中添加追踪宏时，有一些事情需要注意：

- 需要小心的是，为探针参数指定的数据类型要匹配宏中使用的变量的数据类型，否则会发生编译错误。

- 在大多数平台上，如果用--enable-dtrace编译了PostgreSQL，无论何时当控制经过一个追踪宏时，都会对该宏的参数求值，即使没有进行追踪也会这样做。如果只是报告少数局部变量的值，通常无需担心这一点。但是要注意不要将开销大的函数调用放入参数中。如果你需要这样做，考虑通过检查追踪是否真的被启用来保护该宏：

  ```
  if (TRACE_POSTGRESQL_TRANSACTION_START_ENABLED())
      TRACE_POSTGRESQL_TRANSACTION_START(some_function(...));
  ```

  每个追踪宏有一个对应的ENABLED宏。

第 28 章 监控磁盘使用情况

本章讨论如何监控PostgreSQL数据库系统的磁盘使用情况。

28.1. 确定磁盘使用情况

每个表都有一个主堆磁盘文件，其中存储了大部分数据。如果该表有任何值可能较宽的列，还可能会有一个与该表关联的TOAST文件，用于存储宽到无法方便地放入主表中的值（见第 55.2 节）。如果存在，还会在TOAST表上有一个索引。也可能会有与基表关联的索引。每个表和索引都存储在单独的磁盘文件中 – 如果文件会超过 1 GB，则可能分成多个文件。这些文件的命名约定见第 55.1 节。

可以通过三种方式监控磁盘空间：使用表 9.60中列出的 SQL 函数，使用oid2name模块，或者手工检查系统目录。SQL 函数最容易使用，通常也是推荐的方法。本节其余部分展示如何通过检查系统目录来完成此事。

在最近进行过清理或分析的数据库上使用psql时，可以发出查询以查看任意表的磁盘使用情况：

```
SELECT pg_relation_filepath(oid), relpages FROM pg_class WHERE relname
 = 'customer';

 pg_relation_filepath | relpages
----------------------+----------
 base/16384/16806     |       60
(1 row)
```

每个页通常为 8 KB。（请记住，`relpages`只会由`VACUUM`、`ANALYZE`以及少数 DDL 命令（如`CREATE INDEX`）更新。）如果你想直接检查该表的磁盘文件，文件路径名就很有用。

要显示TOAST表使用的空间，可使用如下查询：

```
SELECT relname, relpages
FROM pg_class,
     (SELECT reltoastrelid
      FROM pg_class
      WHERE relname = 'customer') AS ss
WHERE oid = ss.reltoastrelid OR
      oid = (SELECT reltoastidxid
             FROM pg_class
             WHERE oid = ss.reltoastrelid)
ORDER BY relname;

       relname        | relpages
----------------------+----------
 pg_toast_16806       |        0
 pg_toast_16806_index |        1
```

也可以轻松显示索引大小：

```
SELECT c2.relname, c2.relpages
FROM pg_class c, pg_class c2, pg_index i
WHERE c.relname = 'customer' AND
      c.oid = i.indrelid AND
```

```
      c2.oid = i.indexrelid
ORDER BY c2.relname;

       relname        | relpages
----------------------+----------
 customer_id_indexdex |       26
```

利用这些信息，你可以很容易地找出最大的表和索引：

```
SELECT relname, relpages
FROM pg_class
ORDER BY relpages DESC;

       relname        | relpages
----------------------+----------
 bigtable             |     3290
 customer             |     3144
```

28.2. 磁盘已满故障

数据库管理员最重要的磁盘监控任务，是确保磁盘不会被写满。
数据磁盘写满不会导致数据损坏，但可能会阻止有用的活动继续进行。如果保存 WAL 文件的磁盘被写满，则数据库服务器可能会 panic 并随后关闭。

如果无法通过删除其他内容来释放磁盘上的额外空间，
可以利用表空间将部分数据库文件移到其他文件系统上。有关更多信息，请参见第 21.6 节。

> **提示**
>
> 有些文件系统在接近写满时性能会明显变差，因此不要等到磁盘完全写满才采取措施。

如果你的系统支持按用户设置磁盘配额，
那么数据库自然也会受到运行服务器的那个用户所拥有配额的限制。
超出配额会带来与磁盘空间完全耗尽相同的不良影响。

第 29 章 可靠性与预写式日志

本章说明如何使用预写式日志，实现高效、可靠的运行。

29.1. 可靠性

可靠性是任何严肃数据库系统的重要属性，而 PostgreSQL 会尽一切可能保证运行可靠。可靠运行的一个方面是，已提交事务所记录的所有数据都应存储在不受断电、操作系统故障和硬件故障影响的非易失性区域中（当然不包括该非易失性区域本身的故障）。通常，只要能成功将数据写入计算机的永久存储设备（磁盘驱动器或等效设备），就满足这一要求。实际上，即使一台计算机遭受致命损坏，只要磁盘驱动器得以幸存，就可以将它们移到另一台硬件相近的计算机上，所有已提交事务仍将完好无损。

虽然周期性地强制将数据写到磁盘盘片似乎是个简单操作，但事实并非如此。由于磁盘驱动器远慢于主存和 CPU，在计算机主存与磁盘盘片之间存在多层缓存。首先是操作系统的缓冲区缓存，它会缓存常用磁盘块并合并磁盘写入。幸运的是，所有操作系统都为应用提供了强制将数据从缓冲区缓存写入磁盘的方法，而 PostgreSQL 会使用这些机制。（关于如何调整这一过程，见参数wal_sync_method。）

其次，磁盘驱动器控制器中可能还有缓存；这在 RAID 控制卡上尤为常见。其中一些缓存是直写式，也就是说写请求一到达就立刻发送到驱动器。另一些是回写式，也就是说数据会在稍后的某个时间发送到驱动器。这类缓存可能构成可靠性隐患，因为磁盘控制器缓存中的内存是易失性的，断电时会丢失内容。较好的控制卡配有后备电池单元（BBU），也就是卡上带有电池，可在系统断电时维持缓存供电。电力恢复后，数据便会写入磁盘驱动器。

最后，大多数磁盘驱动器自身也带有缓存。有些是直写式，有些是回写式；对于回写式驱动器缓存，同样存在与磁盘控制器缓存相同的数据丢失隐患。消费级 IDE 和 SATA 驱动器尤其可能带有在断电后无法保留内容的回写式缓存。很多固态驱动器（SSD）也带有易失性的回写式缓存。

这些缓存通常都可以禁用；不过，具体方法会因操作系统和驱动器类型而异：

- 在 Linux 上，可以使用 `hdparm -I` 查询 IDE 驱动器；如果 `Write cache` 旁边有一个 `*`，就表示启用了写缓存。可以使用 `hdparm -W` 关闭写缓存。SCSI 驱动器可以使用sdparm[1]查询。使用 `sdparm --get=WCE` 检查是否启用了写缓存，使用 `sdparm --clear=WCE` 将其禁用。
- 在 FreeBSD 上，可以使用 `atacontrol` 查询 IDE 驱动器，并通过在 `/boot/loader.conf` 中设置 `hw.ata.wc=0` 来关闭写缓存；SCSI 驱动器也可以使用 `camcontrol identify` 查询，并在可用时借助 `sdparm` 查询和修改写缓存。
- 在 Solaris 上，磁盘写缓存由 `format -e`[2] 控制。（Solaris 的 ZFS 文件系统在启用磁盘写缓存时仍然是安全的，因为它会发出自己的磁盘缓存刷新命令。）
- 在 Windows 上，如果 `wal_sync_method` 是 `open_datasync`（默认值），可以通过取消勾选 `My Computer\Open\`*`disk drive`*`\Properties\Hardware\Properties\Policies\Enable write caching on the disk` 来禁用写缓存。另一种办法是将 `wal_sync_method` 设为 `fsync` 或 `fsync_writethrough`，这两者都会阻止写缓存。
- 在 Mac OS X 上，可以通过将 `wal_sync_method` 设为 `fsync_writethrough` 来防止写缓存。

[1] http://sg.danny.cz/sg/sdparm.html
[2] http://www.sun.com/bigadmin/content/submitted/format_utility.jsp

较新的 SATA 驱动器（遵循 ATAPI-6 或更高版本标准的那些）提供了驱动器缓存刷新命令（`FLUSH CACHE EXT`），而 SCSI 驱动器则早已支持类似的 `SYNCHRONIZE CACHE` 命令。这些命令对 PostgreSQL 并不直接可见，但某些文件系统（如 ZFS、ext4）可以利用它们将启用了回写缓存的驱动器上的数据刷新到盘片。遗憾的是，这类文件系统与带后备电池单元（BBU）的磁盘控制器配合时表现并不理想。在这种配置中，同步命令会强制把控制器缓存中的全部数据写到磁盘，削弱 BBU 的大部分好处。可以运行pg_test_fsync模块查看自己是否受此影响。如果确实受影响，而又能这样做，则可通过关闭文件系统中的写屏障或重新配置磁盘控制器，恢复 BBU 的性能优势。如果关闭了写屏障，一定要确保电池保持正常工作；电池故障有可能导致数据丢失。希望文件系统和磁盘控制器的设计者最终能解决这种次优行为。

当操作系统向存储硬件发出写请求后，它几乎做不了什么来确保数据已经到达真正的非易失性存储区域。相反，管理员有责任确保所有存储组件都能保证数据完整性。应避免使用写缓存没有电池后备的磁盘控制器。在驱动器层面，如果驱动器无法保证在关闭前把数据写出，就应禁用回写缓存。如果使用 SSD，要注意其中很多默认并不遵守缓存刷新命令。可以借助 `diskchecker.pl`[3] 测试 I/O 子系统行为是否可靠。

数据丢失的另一种风险来自磁盘盘片写操作本身。磁盘盘片被分为若干扇区，通常每个 512 字节。每一次物理读写都处理整个扇区。当写请求到达驱动器时，其数据量可能是 512 字节的某个倍数（PostgreSQL 通常一次写入 8192 字节，也就是 16 个扇区），而写入过程可能在任何时刻因断电失败，这就意味着某些 512 字节扇区写成功了，另一些却没有。为防范这类故障，PostgreSQL 会在修改磁盘上的实际页面之前，定期将整页镜像写入持久 WAL 存储。这样一来，在崩溃恢复时 PostgreSQL 就能从 WAL 恢复部分写入的页面。如果你使用了能阻止部分页面写入的文件系统软件（如 ZFS），可以通过关闭参数full_page_writes来禁用这种页面镜像。带后备电池单元（BBU）的磁盘控制器并不能阻止部分页面写入，除非它们保证写入 BBU 的数据总是完整的整页（8kB）。

29.2. 预写式日志（WAL）

预写式日志（WAL）是确保数据完整性的标准方法。大多数（即使不是全部）事务处理书籍中都能找到详细描述。简而言之，WAL 的核心思想是，对数据文件（即表和索引所在之处）的更改，必须在这些更改被记入日志之后才能写出；也就是说，必须在描述这些更改的 WAL 记录被刷入持久存储之后，再去写数据文件。如果遵循这个过程，就无需在每次事务提交时都把数据页刷盘，因为我们知道，一旦发生崩溃，就可以利用日志恢复数据库：任何尚未应用到数据页上的更改，都可以根据 WAL 记录重做。（这就是前滚恢复，也称 REDO。）

> **提示**
>
> 由于 WAL 会在崩溃后恢复数据库文件内容，因此要可靠地存储数据文件或 WAL 文件，并不需要日志化文件系统。实际上，日志开销反而可能降低性能，特别是在日志会导致文件系统数据刷盘时。幸运的是，通常可以通过文件系统挂载选项关闭日志期间的数据刷写，例如 Linux ext3 文件系统上的 `data=writeback`。不过，日志化文件系统确实能够提高崩溃后的启动速度。

使用 WAL 会显著减少磁盘写入次数，因为要保证事务已提交，通常只需把 WAL 文件刷入磁盘，而不必把该事务修改过的每个数据文件都刷盘。WAL 文件按顺序写入，因此同步 WAL 的代价远低于刷写数据页。对于处理大量小事务、且这些事务触及数据存储不同部分的服务

[3] http://brad.livejournal.com/2116715.html

器，这一点尤其明显。此外，当服务器正在处理大量并发的小事务时，对 WAL 文件执行一次 `fsync` 就可能足以提交多个事务。

WAL 还使在线备份和时间点恢复成为可能，如第 24.3 节所述。通过归档 WAL 数据，我们就能支持回退到可用 WAL 数据覆盖范围内的任意时刻：只需先安装数据库的一个较早物理备份，然后把 WAL 重放到目标时刻即可。更进一步说，这个物理备份不必是数据库状态的瞬时快照 – 即使备份是在一段时间内完成的，重放这段时间内的 WAL 也能修复任何内部不一致。

29.3. 异步提交

异步提交是一项可选功能，它允许事务更快完成，但代价是如果数据库发生崩溃，
最近的事务可能会丢失。在许多应用中，这是可以接受的权衡。

如前一节所述，事务提交通常是同步的：服务器要等到该事务的 WAL 记录被刷入持久存储后，才向客户端返回成功指示。因此，客户端可以确信那些已被报告为提交成功的事务一定会被保留，
即使服务器紧接着立刻崩溃也是如此。然而，对于短事务而言，
这种等待是总事务时间的主要组成部分。
选择异步提交模式意味着服务器会在事务从逻辑上完成后就立刻返回成功，而不必等到它生成的 WAL 记录真正到达磁盘。这会显著提升小事务的吞吐量。

异步提交引入了数据丢失风险。在向客户端报告事务完成与事务真正被提交（也就是说，
即使服务器崩溃也保证不会丢失）之间，存在一个很短的时间窗口。因此，
如果客户端会基于“系统一定会记住这个事务”这一假设采取外部动作，就不应使用异步提交。
例如，银行显然不会对记录 ATM 吐钞的事务使用异步提交。但在很多场景中，例如事件日志记录，并不需要这种强保证。

使用异步提交所承担的风险是数据丢失，而不是数据损坏。如果数据库发生崩溃，它会通过重放 WAL 直到最后一条已刷写记录来恢复。因此，数据库会被恢复到自洽状态，但任何尚未刷到磁盘的事务都不会反映在该状态中。其最终效果就是丢失最后几个事务。由于事务按提交顺序重放，
所以不会引入任何不一致性 – 例如，如果事务 B 做出的更改依赖于更早的事务 A 的效果，就不可能出现 A 的效果丢失而 B 的效果却被保留的情况。

用户可以为每个事务选择提交模式，因此同步提交事务和异步提交事务可以并发运行。
这使得在性能与事务持久性的确定性之间能够灵活权衡。提交模式由用户可设置的参数
synchronous_commit控制，而该参数可以用任何设置配置参数的方式来更改。
任一事务实际使用的模式取决于事务提交开始时 `synchronous_commit` 的值。

某些实用命令，例如 `DROP TABLE`，无论 `synchronous_commit` 如何设置，都会强制同步提交。这是为了保证服务器文件系统与数据库逻辑状态之间的一致性。支持两阶段提交的命令，
例如 `PREPARE TRANSACTION`，也始终是同步的。

如果数据库在异步提交与写出该事务 WAL 记录之间的风险窗口内崩溃，那么该事务所做的更改将丢失。该风险窗口的时长是受限的，因为一个后台进程（“WAL 写入器”）会将尚未写出的 WAL 记录每隔wal_writer_delay毫秒刷到磁盘。风险窗口的实际最大时长是 `wal_writer_delay` 的三倍，因为 WAL 写入器在系统繁忙时被设计为倾向于一次写出整页。

> **小心**
>
> 立即模式关闭等同于服务器崩溃，因此也会造成任何未刷写的异步提交丢失。

异步提交的行为不同于将fsync设为 off。`fsync` 是影响所有事务的全服务器设置。它会禁用 PostgreSQL 内部所有试图同步数据库不同部分写入的逻辑，因此系统崩溃（即硬件或操作系统

崩溃，而不是 PostgreSQL 本身故障）可能导致数据库状态遭受任意严重的损坏。在许多场景中，异步提交能带来关闭 fsync 所能获得的大部分性能提升，却不承担数据损坏的风险。

commit_delay听起来也很像异步提交，但它实际上是一种同步提交方法（事实上，在异步提交期间 commit_delay 会被忽略）。commit_delay 会在事务把 WAL 刷到磁盘之前引入一段延迟，希望某个事务执行的一次刷盘也能服务于其他差不多同时提交的事务。可以把这个设置看成是扩大一个时间窗口，让事务得以加入即将参与同一次刷盘的组，从而在多个事务之间摊销这次刷盘的代价。只有在存在许多并发事务时，设置 commit_delay 才可能有帮助。

29.4. WAL 配置

有若干与 WAL 相关的配置参数会影响数据库性能。本节解释它们的用法。有关设置服务器配置参数的一般信息，请参见第 18 章。

检查点是事务序列中的一些点，在这些点上可以保证堆和索引数据文件已经用检查点之前写入的全部信息进行了更新。执行检查点时，所有脏数据页都会刷盘，并向 WAL 文件写入一条特殊的检查点记录。（更改记录此前已写入 WAL 文件并刷盘。）如果发生崩溃，崩溃恢复过程会查看最新的检查点记录，以确定应该从 WAL 的哪个位置（称为重做记录）开始执行 REDO。该点之前对数据文件所做的任何更改都保证已经在磁盘上。因此，检查点之后，位于包含重做记录的段之前的 WAL 段不再需要，可以被回收或删除。（若正在执行 WAL 归档，则这些 WAL 段必须先归档，随后才能回收或删除。）

检查点要求将所有脏数据页刷盘，这可能造成可观的 I/O 负载。因此，检查点活动会被节流，使 I/O 从检查点开始时就启动，并在下一个检查点预计开始前完成；这样可将检查点期间的性能下降降到最低。

服务器的后台写进程会自动定期执行检查点。每写入checkpoint_segments个日志段，或者每经过checkpoint_timeout秒，就会开始一个检查点，以先到者为准。默认设置分别是 3 个段和 300 秒（5 分钟）。也可以使用 SQL 命令 CHECKPOINT 强制执行检查点。

减小 checkpoint_segments 和/或 checkpoint_timeout 会让检查点更频繁地发生。这样可以加快崩溃后恢复，因为需要重做的工作更少。不过，这必须与更频繁地将脏数据页刷盘所增加的成本权衡。如果设置了full_page_writes（默认就是如此），还要考虑另一个因素。为了保证数据页一致性，每个检查点之后对某个数据页的首次修改，都会导致把整页内容写入日志。在这种情况下，更短的检查点间隔会增加输出到 WAL 的数据量，部分抵消缩短间隔的目的，并且无论如何都会带来更多磁盘 I/O。

检查点的代价相当高，首先是因为它要求写出当前所有脏缓冲区，其次是因为它还会像上文所述那样带来额外的后续 WAL 流量。因此，明智的做法是将检查点参数设得足够高，避免检查点发生得过于频繁。作为对检查点参数的一种简单合理性检查，你可以设置参数checkpoint_warning。如果检查点之间的间隔小于 checkpoint_warning 秒，服务器日志就会输出一条消息，建议增大 checkpoint_segments。偶尔出现这样的消息无需惊慌，但如果它频繁出现，就应当增大检查点控制参数。大型 COPY 传输之类的批量操作，如果 checkpoint_segments 设得不够高，也可能导致出现多次此类警告。

为了避免大量页面写出在短时间内淹没 I/O 系统，检查点期间对脏缓冲区的写出会被分散到一段时间内。这段时间由checkpoint_completion_target控制，它表示为检查点间隔的一个分数。系统会调节 I/O 速率，使检查点在自检查点开始以来已消耗给定比例的 checkpoint_segments 个 WAL 段、或者已过去给定比例的 checkpoint_timeout 秒时完成，以较早者为准。使用默认值 0.5 时，可以预期 PostgreSQL 会在下一次检查点开始前的这段时间约过一半时完成每个检查点。在正常运行时已非常接近最大 I/O 吞吐量的系统上，你可能希望增大 checkpoint_completion_target，以降低检查点带来的 I/O 负载。这样做的缺点是，延长检查点会影响恢复时间，因为需要保留更多 WAL 段以备恢复时使用。虽然 checkpoint_completion_target

可以设到 1.0，但最好保持在该值以下（或许最多为 0.9），因为检查点除了写出脏缓冲区之外还包含其他活动。设置为 1.0 很可能导致检查点不能按时完成，从而因为所需 WAL 段数量的意外波动而损失性能。

WAL 段文件至少会有一个，通常不会多于 (2 + checkpoint_completion_target) * checkpoint_segments + 1 或 checkpoint_segments + wal_keep_segments + 1 个文件。每个段文件通常是 16 MB（不过在构建服务器时可以更改这个大小）。可以据此估算 WAL 的空间需求。通常，当旧的日志段文件不再需要时，它们会被回收（也就是重命名为编号序列中的未来段文件）。如果由于日志输出速率的短期突增，段文件数量超过了 3 * checkpoint_segments + 1，那么不再需要的段文件就会被删除而不是回收，直到系统重新回到该限制之下。

在归档恢复或备库模式下，服务器会周期性地执行重启点，其行为类似于正常运行时的检查点：服务器会强制将其所有状态刷盘，更新 pg_control 文件，以表明已经处理过的 WAL 数据无需再次扫描，然后回收 pg_xlog 目录中的旧 WAL 段文件。重启点的执行频率不会高于主库上的检查点，因为重启点只能在检查点记录处执行。重启点总是会等待至少 checkpoint_segments 个日志段或 checkpoint_timeout 秒（以较长者为准）。

有两个常用的内部 WAL 函数：LogInsert 和 XLogFlush。LogInsert 用于把新记录放入共享内存中的 WAL 缓冲区。如果新记录没有可用空间，LogInsert 就不得不把若干已满的 WAL 缓冲区写出（移入内核缓存）。这并不理想，因为 LogInsert 会在每一次数据库底层修改（例如插入行）时被调用，而在调用它时，受影响数据页上正持有排他锁，因此这一操作必须尽可能快。更糟的是，写 WAL 缓冲区还可能迫使系统创建新的 WAL 段，这会花费更多时间。正常情况下，WAL 缓冲区应由 XLogFlush 请求负责写出并刷盘；该请求大多发生在事务提交时，以确保事务记录已刷盘至持久存储。在 WAL 输出量很高的系统上，XLogFlush 请求可能不够频繁，无法避免 LogInsert 自己去执行写出。在这种系统上，应增加 WAL 缓冲区的数量，具体做法是修改参数wal_buffers。当设置了full_page_writes且系统非常繁忙时，把该值设置得更高有助于平滑每次检查点后紧接着那段时期的响应时间。

commit_delay参数定义了服务器进程在用 LogInsert 把一条提交记录写入日志之后、执行 XLogFlush 之前会睡眠多少微秒。这一延迟允许其他服务器进程把它们的提交记录追加到日志中，从而让所有这些记录通过一次日志同步一起刷盘。如果未启用fsync，或者当前处于活动事务中的其他会话少于commit_siblings个，则不会发生睡眠；这避免了在不太可能有其他会话很快提交时进行睡眠。注意，在大多数平台上，睡眠请求的分辨率是十毫秒，因此 1 到 10000 微秒之间的任何非零 commit_delay 设置效果都相同。这些参数的理想取值目前尚不明确；鼓励进行实验。

参数wal_sync_method决定 PostgreSQL 将如何请求内核把 WAL 更新强制刷盘。就可靠性而言，所有选项应当都是相同的；例外是 fsync_writethrough，它有时能够在其他选项做不到时强制将磁盘缓存中的数据刷盘。不过，具体哪个选项速度最快则高度依赖平台。可以使用pg_test_fsync模块测试不同选项的速度。请注意，如果 fsync 已被关闭，那么这个参数就没有意义。

启用配置参数wal_debug（前提是 PostgreSQL 编译时支持它）会导致每一次 LogInsert 和 XLogFlush 的 WAL 调用都被记录到服务器日志中。将来，这个选项可能会被更通用的机制所取代。

29.5. WAL 内部机制

WAL 会自动启用；管理员除了确保满足 WAL 文件的磁盘空间需求，并完成任何必要调优之外，不需要再做其他操作（见第 29.4 节）。

WAL 文件保存在数据目录下的 pg_xlog 目录中，由一组段文件组成，每个段通常为 16 MB（不过可在构建服务器时通过修改 --with-wal-segsize configure 选项来改变大小）。每个段又

分为若干页面，通常每页 8 kB（该大小可通过 configure 选项 `--with-wal-blocksize` 更改）。WAL 记录头定义在 `access/xlogrecord.h` 中；记录内容则取决于被记录的事件类型。段文件以不断递增的数字命名，从 `000000010000000000000000` 开始。这些编号不会回绕，但要把可用的编号耗尽还需要非常、非常长的时间。

如果能将 WAL 放在与主数据库文件不同的磁盘上，会更有利。这可以通过把 `pg_xlog` 目录移动到其他位置（当然要在服务器关闭时进行），然后在主数据目录中的原位置创建一个指向新位置的符号链接来实现。

WAL 的目标是确保数据库记录在被修改之前，日志已经先被写出；但如果磁盘驱动器实际上只是缓存了数据、尚未将其存储到磁盘上，却向内核谎报写入成功，这一目标就会被破坏。
在这种情况下，断电可能导致不可恢复的数据损坏。管理员应尽量确保保存 PostgreSQL WAL 文件的磁盘不会做出这种虚假报告。（见第 29.1 节。）

在完成检查点并刷写 WAL 之后，该检查点的位置会保存在文件 `pg_control` 中。因此，在恢复开始时，服务器会先读取 `pg_control`，再读取检查点记录；然后从检查点记录给出的 WAL 位置向前扫描并执行 REDO。由于检查点之后对数据页的第一次修改，会把整页内容保存在 WAL 中（假设没有禁用full_page_writes），因而自该检查点以来被修改的所有页面都会恢复到一致状态。

为处理 `pg_control` 损坏的情况，我们本应支持按相反顺序扫描现有 WAL 段 – 也就是从新到旧 –，以便找到最新检查点。这一功能尚未实现。`pg_control` 足够小（不足一个磁盘页），因此不会受部分写问题影响；到目前为止，也没有仅仅因为无法读取 `pg_control` 本身而导致数据库故障的报告。所以，尽管理论上它是个薄弱环节，`pg_control` 在实践中似乎不是问题。

第 30 章 回归测试

回归测试是针对PostgreSQL中 SQL 实现的一套全面测试。它们既测试标准 SQL 操作，也测试PostgreSQL的扩展能力。

30.1. 运行测试

回归测试既可以针对已经安装并正在运行的服务器执行，
也可以在构建树中使用临时安装来执行。此外，运行测试时还有“parallel”和“sequential”两种模式。顺序方式一次只运行一个测试脚本，而并行方式会启动多个服务器进程，
以并行方式运行一组测试。并行测试能进一步确认进程间通信和锁定机制是否正常工作。

30.1.1. 针对临时安装运行测试

要在构建完成后、安装之前运行并行回归测试，请输入：

```
gmake check
```

运行位置为顶层目录。（也可以切换到`src/test/regress`并在那里运行此命令。）结束时，应该会看到类似下面的结果：

```
=======================
 All 115 tests passed.
=======================
```

否则会显示哪些测试失败的说明。请参阅第 30.2 节中的下文说明，不要过早认定一次“失败”就意味着存在严重问题。

由于这种测试方法会运行一个临时服务器，如果你以 root 用户身份进行构建，它将无法工作，因为服务器不会以 root 身份启动。推荐做法是不要以 root 身份构建；否则应在安装完成后再执行测试。

如果你把PostgreSQL配置为安装到某个已经存在旧版 PostgreSQL安装的位置，并且在安装新版本之前执行 `gmake check`，可能会发现测试失败，因为新程序会尝试使用已安装的共享库。（典型症状是报出未定义符号。）如果你希望在覆盖旧安装之前运行测试，就需要使用`configure --disable-rpath` 进行构建。不过，不建议在最终安装时使用这个选项。

并行回归测试会在你的用户 ID 下启动相当多的进程。目前最大并发度为二十个并行测试脚本，也就是四十个进程：每个测试脚本都有一个服务器进程和一个 psql进程。因此，如果你的系统对每个用户可创建的进程数施加限制，请确保该限制至少有五十个左右，
否则并行测试中可能会出现看似随机的失败。如果无法提高该限制，可以通过设置`MAX_CONNECTIONS` 参数来降低并行度。例如：

```
gmake MAX_CONNECTIONS=10 check
```

这样一次最多只会并发运行十个测试。

30.1.2. 针对现有安装运行测试

要在安装之后运行测试（见第 15 章），请初始化一个数据区域并启动服务器，具体说明见第 17 章，然后输入：

```
gmake installcheck
```

或者，要并行测试：

```
gmake installcheck-parallel
```

测试默认通过本地主机和默认端口号连接服务器，除非通过PGHOST和PGPORT
环境变量另行指定。测试将在名为regression的数据库中运行；
任何现有的同名数据库都会被删除。这些测试还会临时创建名为 regressuser*N* 的用户身份。

30.1.3. 附加测试套件

gmake check 和 gmake installcheck 命令只运行“核心”回归测试，这些测试用于检验 PostgreSQL 服务器的内置功能。源码发行版还包含其他测试套件，其中大部分与附加功能有关，例如可选的过程语言。

要运行适用于已选定构建模块的全部测试套件（包括核心测试），
请在构建树顶层输入下列命令之一：

```
gmake check-world
gmake installcheck-world
```

这些命令分别使用临时服务器或已安装的服务器运行测试，与前面对gmake check和gmake installcheck的说明相同。其他注意事项也与前面对各个方法的说明相同。注意，gmake check-world会为每个被测试的模块构建一个独立的临时安装树，因此需要多得多的时间和磁盘空间，相比之下耗用较少的是gmake installcheck-world。

或者，也可以在构建树相应的子目录中执行gmake check 或gmake installcheck来运行单个测试套件。请记住，gmake installcheck假定你已经安装了相关模块，而不仅仅是核心服务器。

可以按这种方式调用的附加测试包括：

- 可选过程语言的回归测试（PL/pgSQL 除外，它由核心测试负责）。这些测试位于 src/pl 下。
- contrib模块的回归测试，位于 contrib下。并非所有contrib 模块都有测试。
- ECPG 接口库的回归测试，位于 src/interfaces/ecpg/test。
- 针对并发会话行为的压力测试，位于src/test/isolation。

使用 installcheck 模式时，这些测试会销毁任何名为 pl_regression、contrib_regression、isolationtest、regress1 或 connectdb 的现有数据库，以及名为 regression 的现有数据库。

30.1.4. 区域设置和编码

默认情况下，使用临时安装的测试会采用当前环境中定义的区域设置，以及由 initdb决定的相应数据库编码。通过设置适当的环境变量来测试不同的区域设置可能很有用，例如：

```
gmake check LANG=C
gmake check LC_COLLATE=en_US.utf8 LC_CTYPE=fr_CA.utf8
```

出于实现原因，设置LC_ALL不能用于此目的；其他所有与区域设置相关的环境变量都可以。

在针对现有安装测试时，区域设置由现有数据库集簇决定，不能为测试单独设置。

也可以通过设置变量ENCODING来显式选择数据库编码，例如：

```
gmake check LANG=C ENCODING=EUC_JP
```

只有在区域设置为 C 时，这样设置数据库编码通常才有意义；否则编码会从区域设置自动选择，而指定与区域设置不匹配的编码会导致错误。

无论是针对临时安装还是现有安装测试，都可以设置数据库编码；不过在后一种情况下，它必须与该安装的区域设置兼容。

30.1.5. 额外测试

核心回归测试套件包含几个默认不运行的测试文件，因为它们可能依赖于平台，或者运行时间很长。可以通过设置以下变量来运行这些或其他额外的测试文件：EXTRA_TESTS。例如，要运行numeric_big测试：

```
gmake check EXTRA_TESTS=numeric_big
```

要运行排序规则测试：

```
gmake check EXTRA_TESTS=collate.linux.utf8 LANG=en_US.utf8
```

其中，collate.linux.utf8测试仅适用于 Linux/glibc 平台，且只有在使用 UTF-8 编码的数据库中运行时才会成功。

30.1.6. 测试热备

源代码分发包中还包含针对热备静态行为的回归测试。这些测试需要一个正在运行的主库和一个正在运行的备库，备库正在从主库接收新的 WAL 更改（使用基于文件的日志传送或流复制均可）。这些服务器不会自动创建，复制的设置也不在此处文档说明。请查阅文档中有关所需命令和相关问题的各个章节。

要运行热备测试，首先在主库上创建一个名为regression 的数据库：

```
psql -h primary -c "CREATE DATABASE regression"
```

接下来，在主库上的 regression 数据库中运行准备脚本 src/test/regress/sql/hs_primary_setup.sql，例如：

```
psql -h primary -f src/test/regress/sql/hs_primary_setup.sql
 regression
```

等待这些更改传播到备库。

现在将默认数据库连接指向要测试的备库（例如，设置 PGHOST和PGPORT 环境变量）。最后，在回归测试目录中运行 gmake standbycheck：

```
cd src/test/regress
gmake standbycheck
```

还可以使用脚本 src/test/regress/sql/hs_primary_extremes.sql 在主库上生成一些极端行为，以测试备库的行为。

30.2. 测试结果评估

某些安装正确且功能完备的PostgreSQL实例，
可能会因为平台特有因素而在部分回归测试中“失败”，例如浮点表示或消息措辞不同。
目前这些测试是通过把输出与参考系统生成的输出做简单的diff比较来评估的，
因此结果会对细微的系统差异很敏感。报告某项测试“失败”时，
请务必检查预期结果与实际结果之间的差异；你可能会发现这些差异并不重要。尽管如此，
我们仍努力在所有受支持平台上维护准确的参考文件，因此原则上应期待所有测试都能通过。

回归测试的实际输出位于src/test/regress/results 目录中的文件里。测试脚本使用diff
将每个输出文件与存放在src/test/regress/expected目录中的参考输出进行比较。
所有差异都会保存在 src/test/regress/regression.diffs中供你检查。（运行核心测试
以外的测试套件时，这些文件当然会出现在相应的子目录中，而不是src/test/regress。）

如果由于某种原因，某个特定平台在某个测试上产生了“失败”，
但对输出的检查令你相信结果是有效的，那么可以新增一个比较文件，
让后续测试运行不再报告该失败。详情见第 30.3 节。

30.2.1. 错误消息差异

某些回归测试包含故意构造的非法输入值。错误消息可能来自 PostgreSQL代码，也可能来自宿
主平台的系统例程。在后一种情况下，消息会因平台而异，但应反映相近的信息。
这类消息差异会导致回归测试显示为“失败”，但可以通过检查确认其有效性。

30.2.2. 区域设置差异

如果你针对一个使用 C 以外排序规则顺序的区域设置初始化的服务器运行测试，就可能因为排序
顺序不同而出现差异，并导致后续失败。回归测试套件通过提供备用结果文件来处理这个问题；
这些文件已知可以覆盖大量区域设置。

使用临时安装方法时，要在不同区域设置下运行测试，可在 gmake命令行上传递合适的区域设置
相关环境变量，例如：

```
gmake check LANG=de_DE.utf8
```

（回归测试驱动器会取消设置LC_ALL，因此不能用该变量选择区域设置。）
如果不使用区域设置，要么取消所有与区域设置相关的环境变量（或将它们设为C），
要么使用下面这个特殊调用：

```
gmake check NO_LOCALE=1
```

针对现有安装运行测试时，区域设置由现有安装决定。要改变它，需要在初始化数据库集簇时向
initdb传入合适选项，使用另一种区域设置。

一般来说，建议在计划用于生产环境的区域设置下运行回归测试，
因为这样可以覆盖实际将在生产中使用的区域设置和编码相关代码部分。
根据操作系统环境不同，可能会出现失败，
但至少你会知道在真实应用运行时应当预期哪些与区域设置相关的行为。

30.2.3. 日期和时间差异

大多数日期和时间结果依赖于时区环境。参考文件是在时区 PST8PDT（加利福尼亚州伯克利）下
生成的，如果测试未在该时区设置下运行，就会出现表面上的失败。
回归测试驱动器会将环境变量 PGTZ设为PST8PDT，这通常能够确保获得正确结果。

30.2.4. 浮点差异

某些测试涉及根据表列计算 64 位浮点数（double precision）。我们已经观察到涉及double precision列数学函数的结果存在差异。float8和 geometry测试尤其容易在不同平台之间，甚至在不同编译器优化设置下出现细微差异。要判断这些通常出现在小数点右侧第 10 位附近的差异究竟是否重要，需要人工目测比较。

某些系统把负零显示为-0，而另一些只显示 0。

某些系统对pow()和exp() 发出错误的方式，与当前PostgreSQL代码所预期的机制不同。

30.2.5. 行顺序差异

你可能会看到这样的差异：同样的行在输出中的顺序与预期文件中的顺序不同。
在大多数情况下，严格说来这并不是缺陷。大多数回归测试脚本并没有细致到为每一个SELECT 都使用ORDER BY，因此按 SQL 规范，它们的结果行顺序并没有良好定义。实际上，由于我们看到的是同一软件在同一数据上执行相同查询，通常在所有平台上都会得到相同的结果顺序，
所以缺少ORDER BY并不是问题。不过，有些查询确实会表现出跨平台的顺序差异。针对已安装服务器测试时，非 C 区域设置或非默认参数设置，例如自定义的work_mem值或规划器代价参数，也可能导致顺序差异。

因此，如果你看到顺序差异，一般无需担心，除非查询确实包含 ORDER BY而你的结果违反了它。不过，仍请报告该问题，这样我们可以为那个特定查询加上ORDER BY，以在后续版本中消除这种虚假的“失败”。

你可能会好奇，为什么我们不显式地为所有回归测试查询排序，从而一劳永逸地解决这个问题。
原因在于，那样反而会降低回归测试的价值，
因为测试会倾向于覆盖能产生有序结果的查询计划类型，
而排除那些不能产生有序结果的计划类型。

30.2.6. 栈深度不足

如果errors测试在执行 select infinite_recurse()命令时导致服务器崩溃，这意味着平台对进程栈大小的限制比max_stack_depth参数所表明的更小。
可通过在更高的栈大小限制下运行服务器来修复（对于max_stack_depth的默认值，建议 4MB）。如果做不到，另一种办法是减小max_stack_depth 的值。

在支持getrlimit()的平台上，服务器应自动选择 max_stack_depth的安全值；因此，除非你手工覆盖了该设置，否则这类失败就是一个应报告的缺陷。

30.2.7. “random” 测试

random测试脚本本来就是要产生随机结果的。在极少数情况下，这会导致该回归测试失败。
输入：

```
diff results/random.out expected/random.out
```

通常只应产生一两行差异。除非 random 测试反复失败，否则不必担心。

30.2.8. 配置参数

在针对现有安装运行测试时，某些非默认参数设置可能导致测试失败。例如，改变enable_seqscan或 enable_indexscan等参数，可能导致计划发生变化，进而影响使用EXPLAIN的测试结果。

30.3. 变体比较文件

由于某些测试天生会产生依赖环境的结果，我们提供了指定备用“预期”结果文件的方法。
每个回归测试都可以有多个比较文件，用来展示不同平台上的可能结果。
对于每个测试应使用哪个比较文件，有两种相互独立的判定机制。

第一种机制允许针对特定平台选择比较文件。有一个映射文件 `src/test/regress/resultmap`，用于定义每个平台应使用哪个比较文件。
要为某个特定平台消除虚假的测试“失败”，你需要先选择或创建一个变体结果文件，然后在 `resultmap` 文件中增加一行。

映射文件中的每一行都采用如下形式：

```
testname:output:platformpattern=comparisonfilename
```

测试名就是相应回归测试模块的名称。输出值指明要检查哪个输出文件。对标准回归测试而言，这里始终是`out`。该值对应于输出文件的扩展名。平台模式是 Unix 工具`expr`风格的模式（也就是在开头隐含`^`锚点的正则表达式）。它会与`config.guess`打印出的平台名进行匹配。
比较文件名则是替代结果比较文件的基名。

例如：某些系统会把极小的浮点数值解释为零，而不是报告下溢错误。这会导致`float8`回归测试出现一些差异。因此，我们提供了一个变体比较文件`float8-small-is-zero.out`，其中包含这些系统上的预期结果。为了消除虚假的“失败”消息，在OpenBSD平台上，`resultmap`包含以下内容：

```
float8:out:i.86-.*-openbsd=float8-small-is-zero.out
```

如果一台机器上的`config.guess`输出匹配`i.86-.*-openbsd`，就会触发该规则。`resultmap`中的其他行会为其他适用的平台选择变体比较文件。

第二种用于选择变体比较文件的机制更加自动化：它只是在多个提供的比较文件中选取“最佳匹配”。回归测试驱动脚本既考虑某个测试的标准比较文件*testname*`.out`，也考虑名为*testname*`_`*digit*`.out` 的变体文件（其中*digit*可以是任一单个数字 `0-9`）。如果其中任何一个文件完全匹配，就认为测试通过；否则，会使用生成最短 diff 的那个文件来创建失败报告。（如果`resultmap`为该特定测试包含一项，那么基础*testname*就是 `resultmap`中给出的替代名称。）

例如，对于`char`测试，比较文件 `char.out`包含在`C`和 `POSIX`区域设置中应当出现的结果，而文件 `char_1.out`则包含在许多其他区域设置中出现的排序结果。

最佳匹配机制原本是为了解决依赖区域设置的结果问题，
但它也可以用于任何仅凭平台名无法轻易预测测试结果的场景。该机制的一个局限是，
测试驱动器无法判断哪一个变体对当前环境才真正“正确”；它只会选择看起来最合适的那个。
因此，
最安全的做法是只对那些你愿意在所有上下文中都视为同样有效的变体结果使用这种机制。

30.4. 测试覆盖率检查

可以用覆盖率测试插装编译 PostgreSQL 源代码，从而检查回归测试或任何其他与该代码一起运行的测试套件覆盖了代码的哪些部分。目前，这只在使用 GCC 编译时受支持，并且需要`gcov`和`lcov`程序。

一个典型的工作流程如下：

```
./configure --enable-coverage ... OTHER OPTIONS ...
gmake
gmake check # or other test suite
gmake coverage-html
```

然后将 HTML 浏览器指向`coverage/index.html`。这些`gmake`命令在子目录中也可以使用。

要在各次测试运行之间重置执行计数，请运行：

```
gmake coverage-clean
```

部分 IV. 客户端接口

本部分描述随PostgreSQL一起发布的客户端编程接口。这些章节彼此独立，可以分别阅读。
还有许多面向客户端程序的外部编程接口是单独发布的，它们各自也附带文档（附录 G列出了一些较常用的接口）。本部分的读者应熟悉如何使用SQL操作和查询数据库（见第 II 部分），当然还应熟悉自己所选的编程语言。

目录

31. libpq - C 库 556
31.1. 数据库连接控制函数 556
31.2. 连接状态函数 565
31.3. 命令执行函数 569
31.3.1. 主要函数 569
31.3.2. 检索查询结果信息 575
31.3.3. 检索其他结果信息 579
31.3.4. 用于在 SQL 命令中嵌入字符串的转义 580
31.4. 异步命令处理 582
31.5. 取消进行中的查询 586
31.6. 快速路径接口 587
31.7. 异步通知 588
31.8. 与COPY命令相关的函数 589
31.8.1. 发送COPY数据的函数 589
31.8.2. 接收COPY数据的函数 590
31.8.3. COPY的已废弃函数 591
31.9. 控制函数 592
31.10. 杂项函数 594
31.11. 通知处理 596
31.12. 事件系统 597
31.12.1. 事件类型 597
31.12.2. 事件回调函数 599
31.12.3. 事件支持函数 600
31.12.4. 事件实例 600
31.13. 环境变量 603
31.14. 密码文件 605
31.15. 连接服务文件 605
31.16. 连接参数的 LDAP 查找 605
31.17. SSL 支持 606
31.17.1. 服务器证书的客户端验证 607
31.17.2. 客户端证书 607
31.17.3. 不同模式中提供的保护 608
31.17.4. SSL 客户端文件使用 609
31.17.5. SSL 库初始化 609
31.18. 在线程化程序中的行为 610
31.19. 构建libpq程序 610
31.20. 示例程序 611
32. 大对象 622
32.1. 简介 622
32.2. 实现特性 622
32.3. 客户端接口 622
32.3.1. 创建一个大对象 622
32.3.2. 导入一个大对象 623
32.3.3. 导出一个大对象 623
32.3.4. 打开一个现有的大对象 623
32.3.5. 向大对象写入数据 624
32.3.6. 从大对象读取数据 624
32.3.7. 在大对象中定位 624
32.3.8. 获取大对象的当前位置 624
32.3.9. 截断一个大对象 625
32.3.10. 关闭一个大对象描述符 625

32.3.11. 移除一个大对象 625
32.4. 服务器端函数 625
32.5. 示例程序 626
33. ECPG - C 中的嵌入式 SQL 632
33.1. 概念 632
33.2. 管理数据库连接 632
33.2.1. 连接到数据库服务器 632
33.2.2. 选择一个连接 633
33.2.3. 关闭一个连接 634
33.3. 运行 SQL 命令 635
33.3.1. 执行 SQL 语句 635
33.3.2. 使用游标 635
33.3.3. 管理事务 636
33.3.4. 预备语句 636
33.4. 使用主变量 637
33.4.1. 概述 637
33.4.2. 声明节 637
33.4.3. 检索查询结果 638
33.4.4. 类型映射 639
33.4.5. 处理非原始 SQL 数据类型 645
33.4.6. 指示符 650
33.5. 动态 SQL 650
33.5.1. 执行没有结果集的语句 650
33.5.2. 执行一个有输入参数的语句 651
33.5.3. 执行一个有结果集的语句 651
33.6. pgtypes 库 652
33.6.1. numeric类型 652
33.6.2. 日期类型 655
33.6.3. 时间戳类型 659
33.6.4. interval 类型 663
33.6.5. decimal类型 663
33.6.6. pgtypeslib 的 errno 值 664
33.6.7. pgtypeslib 的特殊常量 665
33.7. 使用描述符区域 665
33.7.1. 命名 SQL 描述符区域 665
33.7.2. SQLDA 描述符区域 667
33.8. 错误处理 678
33.8.1. 设置回调 678
33.8.2. sqlca 680
33.8.3. SQLSTATE 与 SQLCODE 681
33.9. 预处理器指令 685
33.9.1. 包括文件 685
33.9.2. define 和 undef 指令 685
33.9.3. ifdef、ifndef、else、elif 和 endif 指令 686
33.10. 处理嵌入式 SQL 程序 687
33.11. 库函数 687
33.12. 大对象 688
33.13. C++ 应用 690
33.13.1. 主变量的可见范围 690
33.13.2. 使用外部 C 模块的 C++ 应用开发 691
33.14. 嵌入式 SQL 命令 693
33.15. Informix兼容模式 717
33.15.1. 附加类型 717

33.15.2. 附加的/缺少的嵌入式 SQL 语句 718
33.15.3. Informix-兼容的 SQLDA 描述符区域 718
33.15.4. 附加函数 721
33.15.5. 额外的常量 730
33.16. 内部 732
34. 信息模式 734
34.1. 模式 734
34.2. 数据类型 734
34.3. information_schema_catalog_name 735
34.4. administrable_role_authorizations 735
34.5. applicable_roles 735
34.6. attributes 736
34.7. character_sets 737
34.8. check_constraint_routine_usage 738
34.9. check_constraints 739
34.10. collations 739
34.11. collation_character_set_applicability 739
34.12. column_domain_usage 740
34.13. column_privileges 740
34.14. column_udt_usage 741
34.15. columns 741
34.16. constraint_column_usage 744
34.17. constraint_table_usage 744
34.18. data_type_privileges 745
34.19. domain_constraints 745
34.20. domain_udt_usage 746
34.21. domains 746
34.22. element_types 748
34.23. enabled_roles 750
34.24. foreign_data_wrapper_options 750
34.25. foreign_data_wrappers 750
34.26. foreign_server_options 751
34.27. foreign_servers 751
34.28. foreign_table_options 751
34.29. foreign_tables 752
34.30. key_column_usage 752
34.31. parameters 752
34.32. referential_constraints 754
34.33. role_column_grants 755
34.34. role_routine_grants 755
34.35. role_table_grants 756
34.36. role_usage_grants 756
34.37. routine_privileges 757
34.38. routines 757
34.39. schemata 761
34.40. sequences 761
34.41. sql_features 762
34.42. sql_implementation_info 762
34.43. sql_languages 763
34.44. sql_packages 763
34.45. sql_parts 764
34.46. sql_sizing 764

34.47. sql_sizing_profiles 764
34.48. table_constraints 765
34.49. table_privileges 765
34.50. tables 766
34.51. triggered_update_columns 767
34.52. triggers 767
34.53. usage_privileges 768
34.54. user_mapping_options 769
34.55. user_mappings 769
34.56. view_column_usage 770
34.57. view_routine_usage 770
34.58. view_table_usage 771
34.59. views 771

第 31 章 libpq - C 库

libpq是PostgreSQL的C应用程序编程接口。libpq是一组库函数，客户端程序可用它们向PostgreSQL后端服务器发送查询并接收查询结果。

libpq也是其他几个PostgreSQL应用程序接口的底层引擎，包括为 C++、Perl、Python、Tcl 和ECPG编写的接口。因此，对于这些包的用户，libpq行为的某些方面也很重要。特别是，第 31.13 节、第 31.14 节和第 31.17 节描述了任何使用libpq的应用程序的用户都能观察到的行为。

本章末尾（第 31.20 节）包含一些简短程序，展示如何编写使用libpq的程序。源代码发行包的`src/test/examples`目录中还提供了几个完整的libpq应用程序示例。

使用libpq的客户端程序必须包含头文件`libpq-fe.h`，并且必须与libpq库链接。

31.1. 数据库连接控制函数

以下函数用于建立到PostgreSQL后端服务器的连接。应用程序可以同时保持多个后端连接。（这样做的原因之一是访问多个数据库。）每个连接由一个`PGconn`对象表示，该对象通过函数`PQconnectdb`、`PQconnectdbParams`或`PQsetdbLogin`获取。注意，
除非内存太少以至于无法分配`PGconn`对象，这些函数总是返回非空的对象指针。
在通过连接对象发送查询之前，应调用`PQstatus`函数检查返回值以确认连接成功。

> **警告**
>
> 在 Unix 上，fork 一个带有打开的 libpq 连接的进程可能导致不可预测的结果，因为父进程和子进程共享相同的套接字和操作系统资源。因此不推荐这种用法，不过在子进程中执行`exec`来载入新的可执行文件是安全的。

> **注意**
>
> 在 Windows 上，如果单个数据库连接被反复地启动和关闭，有一种办法可以改进性能。在内部，libpq 分别调用 `WSAStartup()` 和 `WSACleanup()` 来启动和关闭连接。`WSAStartup()` 会增加一个内部 Windows 库引用计数，而 `WSACleanup()` 会递减该计数。当引用计数恰好为 1 时，调用 `WSACleanup()` 会释放所有资源并卸载所有 DLL。这是一项代价很高的操作。为避免这种情况，应用程序可以手工调用 `WSAStartup()`，这样在最后一个数据库连接关闭时资源就不会被释放。

`PQconnectdbParams`

开启一个到数据库服务器的新连接。

```
PGconn *PQconnectdbParams(const char **keywords, const char
 **values, int expand_dbname);
```

这个函数使用从两个以`NULL`结尾的数组中取得的参数打开一个新的数据库连接。
第一个数组`keywords`是一个字符串数组，其中每个元素都是一个关键词。第二个数组`values`给出每个关键词的值。和下面的`PQsetdbLogin`不同，
参数集合可以在不改变函数签名的情况下扩展，因此对于新应用，

最好使用这个函数（或者相应的非阻塞函数PQconnectStartParams和PQconnectPoll）。

当 expand_dbname 非零时，允许把*dbname*关键词的值识别为一个*conninfo*字符串。详见下文。

传入的数组可以为空以使用全部默认参数，也可以包含一个或多个参数设置。
两个数组的长度应当相同。处理会随着keywords数组的最后一个非NULL元素而停止。

当前识别的参数关键字如下：

host

要连接的主机名。如果主机名以斜杠开头，则指定的是 Unix 域通信，而非 TCP/IP 通信；此值是存放套接字文件的目录名。当未指定 host 或其值为空时，默认连接到 /tmp（或构建 PostgreSQL 时指定的套接字目录）中的 Unix 域套接字。在没有 Unix 域套接字的机器上，默认连接到 localhost。

hostaddr

要连接的主机的数字 IP 地址。应采用标准的 IPv4 地址格式，例如 172.28.40.9。如果机器支持 IPv6，也可以使用 IPv6 地址。只要此参数指定了非空字符串，就始终使用 TCP/IP 通信。

使用hostaddr代替host可以避免主机名查找，
这对于有时间限制的应用程序可能很重要。但是，采用 Kerberos、GSSAPI 或 SSPI 认证方式，以及进行verify-fullSSL 证书验证时，都需要主机名。遵循以下规则：

- If host is specified without hostaddr, a host name lookup occurs.
- 如果指定了hostaddr而没有指定host，则hostaddr的值给出服务器的网络地址。如果认证方法需要主机名，则连接尝试将失败。
- 如果同时指定了 host 和 hostaddr，则 hostaddr 的值给出服务器的网络地址。只有认证方法需要主机名时，才会将 host 的值用作主机名；否则忽略该值。

Note that authentication is likely to fail if host is not the name of the server at network address hostaddr. Also, note that host rather than hostaddr is used to identify the connection in ~/.pgpass (see 第 31.14 节).

如果既没有主机名也没有主机地址，libpq 会使用本地 Unix 域套接字连接；在没有 Unix 域套接字的机器上，则会尝试连接到 localhost。

port

连接到服务器主机的端口号，或者Unix域连接的套接字文件名扩展。

dbname

数据库名称。默认为与用户名相同。

user

建立连接所用的 PostgreSQL 用户名。默认与运行应用程序的操作系统用户名相同。

password

服务器要求密码认证时所使用的密码。

`connect_timeout`

连接时的最长等待时间，以秒为单位（写成十进制整数字符串）。
零或未指定表示无限等待。不建议使用小于2秒的超时时间。

`client_encoding`

这将为此连接设置`client_encoding`配置参数。除了对应服务器选项接受的值外，
您还可以使用`auto`来从客户端的当前区域设置（Unix系统上的`LC_CTYPE`环境变量）
确定正确的编码。

`options`

添加在运行时发送给服务器的命令行选项。例如，将其设置为`-c geqo=off`会把会话的`geqo`参数值设为`off`。有关可用选项的详细讨论，请参阅第 18 章。

`application_name`

指定application_name配置参数的值。

`fallback_application_name`

指定application_name配置参数的后备值。如果没有通过连接参数或`PGAPPNAME`
环境变量为`application_name`指定值，则将使用此值。
在通用实用程序中指定后备名称很有用，该程序希望设置默认应用程序名称，
但允许用户覆盖它。

`keepalives`

控制是否使用客户端 TCP keepalive。默认值为 1，表示开启；如果不需要 keepalive，可以将其设为 0，表示关闭。对于通过 Unix 域套接字建立的连接，此参数会被忽略。

`keepalives_idle`

控制在多久没有活动后，TCP 应向服务器发送 keepalive 消息，以秒为单位。值为零时使用系统默认值。对于通过 Unix 域套接字建立的连接，或禁用 keepalive 时，此参数会被忽略。此参数仅在支持 `TCP_KEEPIDLE` 或 `TCP_KEEPALIVE` 套接字选项的系统以及 Windows 上受支持；在其他系统上无效。

`keepalives_interval`

控制未被服务器确认收到的 TCP keepalive 消息在多少秒后应被重传。值为零时使用系统默认值。对于通过 Unix 域套接字建立的连接，或禁用 keepalive 时，此参数会被忽略。此参数仅在支持 `TCP_KEEPINTVL` 套接字选项的系统以及 Windows 上受支持；在其他系统上无效。

`keepalives_count`

控制在客户端与服务器之间的连接被视为中断之前，可以丢失多少个 TCP keepalive 消息。值为零时使用系统默认值。对于通过 Unix 域套接字建立的连接，或禁用 keepalive 时，此参数会被忽略。此参数仅在支持 `TCP_KEEPCNT` 套接字选项的系统上受支持；在其他系统上无效。

`tty`

忽略此参数（以前用于指定服务器调试输出的发送位置）。

`sslmode`

这个选项确定是否以及以何种优先级与服务器协商安全的SSL TCP/IP连接。有六种模式：

`disable`

仅尝试非SSL连接

`allow`

首先尝试非SSL连接；如果失败，则尝试SSL连接

`prefer` (default)

首先尝试SSL连接；如果失败，则尝试非SSL连接

`require`

仅尝试SSL连接。如果存在根CA文件，则验证证书的方式与指定了`verify-ca`时相同

`verify-ca`

仅尝试SSL连接，并验证服务器证书是否由受信任的证书颁发机构（CA）颁发

`verify-full`

仅尝试SSL连接，验证服务器证书是否由受信任的CA颁发，
并且请求的服务器主机名与证书中的匹配

See 第 31.17 节 for a detailed description of how these options work.

在 Unix 域套接字通信中，`sslmode` 会被忽略。如果PostgreSQL编译时未启用 SSL支持，使用选项`require`、`verify-ca`或 `verify-full`会导致错误，而选项`allow`和`prefer` 将被接受，但libpq实际上不会尝试建立SSL 连接。

`requiressl`

此选项已弃用，请改用 `sslmode` 设置。

如果设置为1，则需要与服务器建立SSL连接（这相当于`sslmode require`）。libpq将拒绝连接，如果服务器不接受 SSL连接。如果设置为0（默认值），libpq将与服务器协商连接类型（相当于`sslmode prefer`）。此选项仅在 PostgreSQL 编译时启用了 SSL 支持的情况下可用。

`sslcert`

这个参数指定客户端SSL证书的文件名，替换默认的 `~/.postgresql/postgresql.crt`。如果没有建立SSL连接，则此参数将被忽略。

`sslkey`

这个参数指定了用于客户端证书的密钥的位置。它可以指定一个文件名，
该文件名将被用来替代默认的 `~/.postgresql/postgresql.key`，或者它可以指定一个从外部“引擎”（引擎是OpenSSL可加载模块）获取的密钥。
外部引擎的指定形式应包含一个由冒号分隔的引擎名称和一个引擎特定的密钥标识符。
如果没有进行SSL连接，则此参数将被忽略。

`sslrootcert`

这个参数指定一个包含SSL证书颁发机构（CA）证书的文件名。如果文件存在，服务器的证书将被验证是否由这些机构之一签名。默认值是`~/.postgresql/root.crt`。

`sslcrl`

此参数指定 SSL 证书吊销列表（CRL）的文件名。如果该文件存在，在验证服务器证书时，会拒绝其中列出的证书。默认值为 `~/.postgresql/root.crl`。

`requirepeer`

这个参数指定了服务器的操作系统用户名，例如`requirepeer=postgres`。在建立Unix域套接字连接时，如果设置了这个参数，客户端会在连接开始时检查服务器进程是否在指定的用户下运行；如果不是，则连接会因错误而中止。这个参数可用于提供类似于在TCP/IP连接上使用SSL证书的服务器认证。（请注意，如果Unix域套接字位于`/tmp`或其他公共可写位置，任何用户都可以在那里启动一个服务器监听。使用这个参数来确保您连接到由受信任用户运行的服务器。）此选项仅在实现了`peer`认证方法的平台上受支持；请参见第 19.3.7 节。

`krbsrvname`

使用 Kerberos 5 或 GSSAPI 认证时所用的 Kerberos 服务名。这必须与服务器配置中指定的服务名匹配，Kerberos 认证才能成功。（另请参见第 19.3.5 节和第 19.3.3 节。）

`gsslib`

用于 GSSAPI 认证的 GSS 库。只在 Windows 上使用。设置为`gssapi`可强制 libpq 使用 GSSAPI 库而非默认的 SSPI 进行认证。

`service`

用于额外参数的服务名称。它指定了`pg_service.conf`中保存额外连接参数的服务名称。这允许应用程序只指定一个服务名称，以便可以集中维护连接参数。参见第 31.15 节。

If any parameter is unspecified, then the corresponding environment variable (see 第 31.13 节) is checked. If the environment variable is not set either, then the indicated built-in defaults are used.

如果 `expand_dbname` 非零并且*`dbname`*包含`=`符号，就会将其当作一个*`conninfo`*字符串处理，方式与将其传给`PQconnectdb`完全相同（见下文）。之前处理过的关键词会被*`conninfo`*字符串中的关键词覆盖。

通常，关键词按索引顺序从这些数组的开头开始处理。其效果是，当关键词重复出现时，保留最后处理的值。因此，通过仔细放置 *`dbname`* 关键词，可以决定哪些值可能被 *`conninfo`* 串覆盖，哪些不能。

`PQconnectdb`

开启一个到数据库服务器的新连接。

```
PGconn *PQconnectdb(const char *conninfo);
```

这个函数使用从字符串`conninfo`中得到的参数开启一个新的数据库连接。

传入的字符串可以为空以使用全部默认参数，
也可以包含一个或多个用空白分隔的参数设置。每一个参数设置的形式都是keyword = value。等号周围的空白是可选的。要写一个空值或一个包含空白的值，将它用单引号包围，例如keyword = 'a value'。值中的单引号和反斜线必须用一个反斜线转义，即\'和\\。

当前能被识别的参数关键词与上面相同。

PQsetdbLogin

开启一个到数据库服务器的新连接。

```
PGconn *PQsetdbLogin(const char *pghost,
                     const char *pgport,
                     const char *pgoptions,
                     const char *pgtty,
                     const char *dbName,
                     const char *login,
                     const char *pwd);
```

这是 PQconnectdb 的前身，使用固定的一组参数。除缺失参数始终采用默认值之外，功能相同。对于要使用默认值的任意固定参数，请传入 NULL 或空字符串。

如果 *dbName* 包含 = 符号，就会将其当作一个*conninfo*字符串处理，方式与将其传给PQconnectdb完全相同，然后按照上面的方式应用其余参数。

PQsetdb

开启一个到数据库服务器的新连接。

```
PGconn *PQsetdb(char *pghost,
                char *pgport,
                char *pgoptions,
                char *pgtty,
                char *dbName);
```

这是一个调用PQsetdbLogin的宏，其中为*login*和*pwd*参数使用空指针。
提供它是为了向后兼容非常老的程序。

PQconnectStartParams
PQconnectStart
PQconnectPoll

以非阻塞的方式建立一个到数据库服务器的连接。

```
PGconn *PQconnectStartParams(const char **keywords,
                             const char **values,
                             int expand_dbname);

PGconn *PQconnectStart(const char *conninfo);

PostgresPollingStatusType PQconnectPoll(PGconn *conn);
```

这三个函数被用来开启一个到数据库服务器的连接，
这样你的应用的执行线程不会因为远程的I/O而被阻塞。这种方法的要点在于等待 I/O 完成可能在应用的主循环中发生，而不是在PQconnectdbParams 或 PQconnectdb中，并且因此应用能够把这种操作和其他动作并行处理。

在PQconnectStartParams中，数据库连接使用从keywords和values数组中取得的参数创建，并且被expand_dbname控制，这和之前描述的PQconnectdbParams相同。

在PQconnectStart中，数据库连接使用从字符串conninfo中取得的参数创建，这和之前描述的PQconnectdb相同。

Neither PQconnectStartParams nor PQconnectStart nor PQconnectPoll will block, so long as a number of restrictions are met:

- 必须恰当地使用 hostaddr 和 host 参数，以确保不会进行名称和反向名称查询。详细信息请参见上文PQconnectdbParams中这些参数的说明。
- 如果你调用PQtrace，确保接收追踪输出的流对象不会阻塞。
- 如后文所述，你要确保在调用PQconnectPoll之前，套接字处于合适的状态。

注意：PQconnectStartParams的用法与下文展示的PQconnectStart类似。

要开始无阻塞的连接请求，可调用conn = PQconnectStart("*connection_info_string*")。如果conn为空，则libpq无法分配一个新的PGconn结构体。否则，一个有效的PGconn指针会被返回（不过还没有表示一个到数据库的有效连接）。从PQconnectStart返回后，调用status = PQstatus(conn)。如果status等于CONNECTION_BAD，则PQconnectStart失败。

如果PQconnectStart成功，下一个阶段是轮询libpq，这样它能够继续进行连接序列。使用PQsocket(conn)来获得该数据库连接底层的套接字描述符。这样循环：如果PQconnectPoll(conn)上一次返回PGRES_POLLING_READING，等到该套接字准备好读取（按照select()、poll()或类似的系统函数所指示的）。则再次调用PQconnectPoll(conn)。反之，如果PQconnectPoll(conn)上一次返回PGRES_POLLING_WRITING，等到该套接字准备好写入，则再次调用PQconnectPoll(conn)。如果你还没有调用过PQconnectPoll，即刚刚调用过PQconnectStart之后，行为就像是它上次返回了PGRES_POLLING_WRITING。持续这个循环直到PQconnectPoll(conn)返回PGRES_POLLING_FAILED指示连接过程已经失败，或者返回PGRES_POLLING_OK指示连接已经被成功地建立。

在连接过程中的任何时刻，都可以通过调用PQstatus来检查连接状态。如果该调用返回CONNECTION_BAD，则连接过程已经失败；如果返回CONNECTION_OK，则连接已就绪。通过以下函数的返回值也同样可以检测这两种状态：PQconnectPoll，该函数已在上文介绍。在异步连接过程中（也仅在此过程中）还可能出现其他状态。它们指明连接过程的当前阶段，例如可以用来向用户提供反馈。这些状态如下：

CONNECTION_STARTED

等待连接被建立。

CONNECTION_MADE

连接 OK，等待发送。

CONNECTION_AWAITING_RESPONSE

等待来自服务器的一个回应。

CONNECTION_AUTH_OK

收到认证，等待后端启动结束。

CONNECTION_SSL_STARTUP

协商 SSL 加密。

CONNECTION_SETENV

协商环境驱动的参数设置。

Note that, although these constants will remain (in order to maintain compatibility), an application should never rely upon these occurring in a particular order, or at all, or on the status always being one of these documented values. An application might do something like this:

```
switch(PQstatus(conn))
{
        case CONNECTION_STARTED:
            feedback = "Connecting...";
            break;

        case CONNECTION_MADE:
            feedback = "Connected to server...";
            break;
.
.
.
        default:
            feedback = "Connecting...";
}
```

在使用PQconnectPoll时，连接参数connect_timeout会被忽略：判断是否超时是应用的责任。除此之外，PQconnectStart后面跟着PQconnectPoll循环等效于PQconnectdb。

注意如果PQconnectStart返回一个非空的指针，你必须在用完它之后调用PQfinish来处理该结构体和任何相关的内存块。即使连接尝试失败或被放弃时也必须完成这些工作。

PQconndefaults

返回默认连接选项。

```
PQconninfoOption *PQconndefaults(void);

typedef struct
{
    char   *keyword;   /* 该选项的关键词 */
    char   *envvar;    /* 后备环境变量名 */
    char   *compiled;  /* 编译时设置的后备默认值 */
    char   *val;       /* 选项的当前值，或者 NULL */
    char   *label;     /* 连接对话框中字段的标签 */
    char   *dispchar;  /* 指示如何在连接对话框中显示此字段。可取值：
                          ""          显示输入的值
                          "*"         密码字段 – 隐藏值
                          "D"         调试选项 – 默认不显示 */
    int     dispsize;  /* 对话框中的字段宽度，以字符计 */
} PQconninfoOption;
```

返回一个连接选项数组。这可以用来确定所有可能的PQconnectdb
选项和它们的当前默认值。返回值指向一个PQconninfoOption结构体的数组，
该数组以一个包含空keyword指针的条目结束。如果无法分配内存，则返回空指针。
注意当前默认值（val字段）将依赖于环境变量和其他上下文。
调用者必须把连接选项数据当作只读对待。

在处理完选项数组后，把它交给PQconninfoFree释放。如果没有这么做，每次调用
PQconndefaults都会导致一小部分内存泄漏。

PQconninfoParse

返回从提供的连接字符串中解析到的连接选项。

```
PQconninfoOption *PQconninfoParse(const char *conninfo, char
 **errmsg);
```

解析一个连接字符串并且将结果选项作为一个数组返回，或者在连接字符串有问题时返回
NULL。这个函数可以用来抽取所提供的连接字符串中的PQconnectdb选项。
返回值指向一个PQconninfoOption结构体的数组，该数组以一个包含空keyword
指针的条目结束。

注意，只有在字符串中显式指定的选项才会在结果数组中被设置值；不会插入默认值。

如果 errmsg 不是 NULL，则成功时将 *errmsg 设为 NULL；失败时将其设为由 malloc
分配的、用于说明问题的错误字符串。（也可能出现 *errmsg 被设为 NULL，同时函数返回
NULL 的情况；这表示内存不足。）

在处理完选项数组后，把它交给PQconninfoFree释放。如果没有这么做，每次调用
PQconninfoParse都会导致一小部分内存泄漏。反过来，如果发生一个错误并且errmsg
不是NULL，确保使用PQfreemem释放错误字符串。

PQfinish

关闭与服务器的连接。同时释放PGconn对象使用的内存。

```
void PQfinish(PGconn *conn);
```

注意，即使与服务器的连接尝试失败（由PQstatus指示），应用也应当调用PQfinish
来释放PGconn对象使用的内存。不能在调用PQfinish之后再使用PGconn指针。

PQreset

重置与服务器的通信通道。

```
void PQreset(PGconn *conn);
```

此函数将关闭与服务器的连接，并尝试重新建立到同一服务器的新连接，
使用之前使用过的所有参数。这可能有助于在工作连接丢失后的错误恢复。

PQresetStart
PQresetPoll

以非阻塞方式重置与服务器的通信通道。

```
int PQresetStart(PGconn *conn);

PostgresPollingStatusType PQresetPoll(PGconn *conn);
```

这些函数会关闭与服务器的连接，并尝试重新建立到同一服务器的新连接，
使用之前使用过的所有参数。如果原本可用的连接丢失，这可以用于错误恢复。
它们与上文的 PQreset 不同之处在于采用非阻塞方式。它们受到与 PQconnectStartParams、PQconnectStart 和 PQconnectPoll 相同的限制。

要开始重置连接，请调用 PQresetStart。如果返回 0，表示重置失败。如果返回 1，则使用 PQresetPoll 轮询重置过程，方式与使用 PQconnectPoll 建立连接完全相同。

PQpingParams

PQpingParams报告服务器状态。它接受的连接参数与上文介绍的PQconnectdbParams相同。但取得服务器状态并不需要提供正确的用户名、密码或数据库名值。

```
PGPing PQpingParams(const char **keywords, const char **values, int
 expand_dbname);
```

该函数返回以下值之一：

PQPING_OK

The server is running and appears to be accepting connections.

PQPING_REJECT

服务器正在运行，但是处于一种不允许连接的状态（启动、关闭或崩溃恢复）。

PQPING_NO_RESPONSE

无法联系到服务器。这可能表示服务器没有运行，
或者给定的连接参数中有些错误（例如，错误的端口号），
或者有一个网络连接问题（例如，一个防火墙阻断了连接请求）。

PQPING_NO_ATTEMPT

没有尝试联系服务器，因为提供的参数显然不正确，或者有一些客户端问题（例如，内存用完）。

PQping

PQping报告服务器状态。它接受的连接参数与上文介绍的PQconnectdb相同。
但取得服务器状态并不需要提供正确的用户名、密码或数据库名值。

```
PGPing PQping(const char *conninfo);
```

返回值与 PQpingParams 相同。

31.2. 连接状态函数

这些函数可用于查询现有数据库连接对象的状态。

提示

编写 libpq 应用程序时，应注意维护 PGconn 的抽象。请使用下述访问函数获取 PGconn 的内容。不建议通过 libpq-int.h 引用 PGconn 的内部字段，因为这些字段将来可能改变。

以下函数返回建立连接时确定的参数值。这些值在PGconn对象的生命周期内保持不变。

PQdb

返回该连接的数据库名。

```
char *PQdb(const PGconn *conn);
```

PQuser

返回该连接的用户名。

```
char *PQuser(const PGconn *conn);
```

PQpass

返回该连接的密码。

```
char *PQpass(const PGconn *conn);
```

PQhost

返回连接的服务器主机名。可能是主机名、IP 地址或者一个目录路径（如果通过 Unix 套接字连接）。

```
char *PQhost(const PGconn *conn);
```

PQport

返回连接的端口。

```
char *PQport(const PGconn *conn);
```

PQtty

返回该连接用于调试的TTY。（此功能已过时，因为服务器不再使用TTY设置，但为保持向后兼容仍保留此函数。）

```
char *PQtty(const PGconn *conn);
```

PQoptions

返回连接请求中传递的命令行选项。

```
char *PQoptions(const PGconn *conn);
```

以下函数返回的状态数据，可能随着对 PGconn 对象执行操作而改变。

PQstatus

返回该连接的状态。

```
ConnStatusType PQstatus(const PGconn *conn);
```

该状态可以是一系列值之一。不过，其中只有两个在一个异步连接过程之外可见：CONNECTION_OK和CONNECTION_BAD。一个到数据库的完好连接的状态为CONNECTION_OK。一个失败的连接尝试则由状态CONNECTION_BAD表示。通常，一个 OK 状态将一直保持

到PQfinish，但是一次通信失败可能导致该状态过早地改变为CONNECTION_BAD。在那种情况下，该应用可以通过调用PQreset尝试恢复。

关于其他可能会被返回的状态代码，请见PQconnectStartParams、PQconnectStart和PQconnectPoll的条目。

PQtransactionStatus

返回服务器的当前事务内状态。

```
PGTransactionStatusType PQtransactionStatus(const PGconn *conn);
```

该状态可能是PQTRANS_IDLE（当前空闲）、PQTRANS_ACTIVE（一个命令运行中）、PQTRANS_INTRANS（空闲，处于一个有效的事务块中）或者PQTRANS_INERROR（空闲，处于一个失败的事务块中）。如果该连接异常，将会报告PQTRANS_UNKNOWN。只有当一个查询已经被发送给服务器并且还没有完成时，才会报告PQTRANS_ACTIVE。

> **小心**
>
> 在使用参数autocommit被设置为 off 的PostgreSQL 7.3 服务器时，PQtransactionStatus会给出不正确的结果。服务器端的 autocommit 特性已被废弃，在之后的服务器版本中不存在。

PQparameterStatus

查找服务器某个参数的当前设置。

```
const char *PQparameterStatus(const PGconn *conn, const char *paramName);
```

服务器会在连接启动时，以及某些参数值发生变化时，自动报告这些参数值。PQparameterStatus可用于查询这些设置。如果已知该参数，则返回其当前值；如果未知，则返回NULL。

当前版本报告的参数包括：server_version、server_encoding、client_encoding、application_name、is_superuser、session_authorization、DateStyle、IntervalStyle、TimeZone、integer_datetimes、standard_conforming_strings。（8.0 之前的版本不报告 server_encoding、TimeZone 和 integer_datetimes；8.1 之前的版本不报告 standard_conforming_strings；8.4 之前的版本不报告 IntervalStyle；9.0 之前的版本不报告 application_name。）注意，server_version、server_encoding 和 integer_datetimes 在启动后不能改变。

使用 3.0 之前协议的服务器不报告参数设置，但 libpq 仍包含获取 server_version 和 client_encoding 值的逻辑。建议应用程序使用 PQparameterStatus，而不是专门编写代码来确定这些值。（但要注意，在使用 3.0 之前协议的连接上，连接启动后通过 SET 改变 client_encoding，不会反映在 PQparameterStatus 的结果中。）对于 server_version，另请参见 PQserverVersion，它以数值形式返回此信息，更易于比较。

如果服务器未报告standard_conforming_strings的值，应用程序可以假定其为off，即反斜杠在字符串字面量中被视为转义字符。此外，服务器报告此参数也表明它接受转义字符串语法（E'...'）。

返回的指针虽然被声明为`const`，但实际上指向与`PGconn`结构体关联的可变存储。不能假定该指针在执行其他查询后仍然有效。

`PQprotocolVersion`

查询正在使用的前端/后端协议。

```
int PQprotocolVersion(const PGconn *conn);
```

应用程序可以使用此函数判断是否支持某些特性。目前，可能的值为 2（协议 2.0）、3（协议 3.0）或零（连接异常）。连接启动完成后，协议版本不会改变，但理论上可能在连接重置期间改变。通常，与 PostgreSQL 7.4 或更高版本的服务器通信时使用协议 3.0；7.4 之前的服务器仅支持协议 2.0。（协议 1.0 已过时，libpq 不支持它。）

`PQserverVersion`

返回一个表示后端版本的整数。

```
int PQserverVersion(const PGconn *conn);
```

应用可能会使用这个函数来判断它们连接到的数据库服务器的版本。这个数字是这样形成的：将主版本、次版本和修订版本号分别转换成两位十进制数，然后把它们拼接在一起。例如，版本8.1.5将被返回为80105，而版本8.2将被返回为80200（不显示前导零）。如果连接无效则返回零。

`PQerrorMessage`

返回连接上的一个操作最近产生的错误消息。

```
char *PQerrorMessage(const PGconn *conn);
```

几乎所有 libpq 函数在失败时都会设置一条供 `PQerrorMessage` 返回的消息。注意，按照 libpq 的约定，非空的 `PQerrorMessage` 结果可能包含多行，并以换行符结尾。调用者不应直接释放该结果；当关联的 `PGconn` 句柄被传给 `PQfinish` 时，结果会被释放。不能假定在对 `PGconn` 结构体执行其他操作后，结果字符串仍保持不变。

`PQsocket`

获取与服务器相连的套接字的文件描述符编号。有效描述符大于或等于 0。结果为 -1 表示当前没有打开服务器连接（在普通操作期间这将不会改变，但是在连接设置或重置期间可能改变）。

```
int PQsocket(const PGconn *conn);
```

`PQbackendPID`

返回处理这个连接的后端进程的进程ID（PID）。

```
int PQbackendPID(const PGconn *conn);
```

后端 PID 可用于调试，也可与 `NOTIFY` 消息进行比较（消息包含发出通知的后端进程的 PID）。注意，该 PID 属于在数据库服务器主机上运行的进程，而非本地主机上的进程！

`PQconnectionNeedsPassword`

如果连接认证方法要求一个密码但没有可用的密码，返回真（1）。否则返回假（0）。

```
int PQconnectionNeedsPassword(const PGconn *conn);
```

这个函数可以在连接尝试失败后被应用于决定是否向用户提示要求一个密码。

`PQconnectionUsedPassword`

如果连接认证方法使用一个密码，返回真（1）。否则返回假（0）。

```
int PQconnectionUsedPassword(const PGconn *conn);
```

这个函数能在一次连接尝试失败或成功后用于检测该服务器是否要求一个密码。

`PQgetssl`

返回在连接中使用的SSL结构体，如果未使用SSL，则返回NULL。

```
SSL *PQgetssl(const PGconn *conn);
```

该结构体可以用来验证加密级别、检查服务器证书等。有关此结构体的信息，请参阅OpenSSL文档。

必须定义 `USE_SSL` 才能得到这个函数正确的原型。这样做还会自动包含来自OpenSSL的`ssl.h`。

31.3. 命令执行函数

与数据库服务器的连接成功建立后，此处描述的函数用于执行 SQL 查询和命令。

31.3.1. 主要函数

`PQexec`

提交一个命令给服务器并且等待结果。

```
PGresult *PQexec(PGconn *conn, const char *command);
```

返回一个 `PGresult` 指针，也可能返回空指针。除内存不足或无法向服务器发送命令等严重错误外，通常都会返回非空指针。应调用 `PQresultStatus` 函数检查返回值是否表示错误（也应检查空指针，此时该函数返回 `PGRES_FATAL_ERROR`）。使用 `PQerrorMessage` 获取此类错误的更多信息。

命令字符串可以包含多个 SQL 命令（以分号分隔）。在一次`PQexec`调用中发送的多个查询会在同一个事务中处理，除非查询字符串中显式包含`BEGIN/COMMIT`命令将其划分为多个事务。
不过要注意，返回的`PGresult`结构体只描述该字符串中最后执行的命令的结果。
如果其中一条命令失败，就会在此处停止处理该字符串，返回的`PGresult`则描述该错误。

`PQexecParams`

提交一个命令给服务器并且等待结果，它可以在 SQL 命令文本之外独立地传递参数。

```
PGresult *PQexecParams(PGconn *conn,
                       const char *command,
                       int nParams,
                       const Oid *paramTypes,
                       const char * const *paramValues,
                       const int *paramLengths,
                       const int *paramFormats,
```

```
                                int resultFormat);
```

PQexecParams与PQexec相似，
但是提供了额外的功能：参数值可以与命令字符串分开指定，
并且可以以文本或二进制格式请求查询结果。PQexecParams 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

该函数的参数是：

conn

要在其中发送命令的连接对象。

command

要执行的 SQL 命令字符串。如果使用了参数，它们在该命令字符串中被引用为$1、$2等。

nParams

提供的参数数量。它是数组*paramTypes[]*、*paramValues[]*、*paramLengths[]*和*paramFormats[]*的长度（当*nParams*为零时，数组指针可以是NULL）。

paramTypes[]

通过 OID 指定要赋予给参数符号的数据类型。如果*paramTypes*为NULL或者该数组中任何特定元素为零，
服务器会用对待未指定类型的字符串字面量的方式为参数符号推测一种数据类型。

paramValues[]

指定参数的实际值。数组中的空指针表示对应参数为 null；否则，指针指向以零结尾的文本字符串（文本格式），或采用服务器所要求格式的二进制数据（二进制格式）。

paramLengths[]

指定二进制格式参数的实际数据长度。对于 null 参数和文本格式参数，该值会被忽略。如果没有二进制参数，数组指针可以为空指针。

paramFormats[]

指定参数采用文本格式（在对应数组元素中填入零）还是二进制格式（填入一）。
如果数组指针为空指针，则将所有参数视为文本字符串。

以二进制格式传递值时，需要了解后端所要求的内部表示形式。例如，
整数必须以网络字节序传递。传递 numeric 值时，需要了解服务器的存储格式，其实现见 src/backend/utils/adt/numeric.c::numeric_send() 和 src/backend/utils/adt/numeric.c::numeric_recv()。

resultFormat

指定零以获取文本格式的结果，指定一以获取二进制格式的结果。
（目前无法让不同结果列使用不同格式，尽管底层协议支持这样做。）

PQexecParams 相对于 PQexec 的主要优点是可以将参数值与命令字符串分开，从而避免繁琐且容易出错的加引号和转义工作。

和PQexec不同，PQexecParams至多允许在给定串中出现一个 SQL 命令（其中可以有分号，但是不能有超过一个非空命令）。这是底层协议的一个限制，但是有助于抵抗 SQL 注入攻击。

> **提示**
>
> 通过 OID 指定参数类型较为繁琐，尤其是在不希望将具体 OID 值写死在程序中时。不过，即使服务器无法自行确定参数类型，或者推断出的类型与你所需的不同，
> 也可以避免直接指定 OID。在 SQL 命令文本中，为参数符号添加显式类型转换，即可指定要发送的数据类型。例如：
>
> ```
> SELECT * FROM mytable WHERE x = $1::bigint;
> ```
>
> 这会强制将参数 $1 当作 bigint，而默认情况下会为它分配与 x 相同的类型。以二进制格式发送参数值时，强烈建议采用这种方式，或直接指定类型的数值 OID，来明确决定参数类型。因为二进制格式的冗余比文本格式更少，
> 服务器发现类型不匹配错误的机会也更少。

PQprepare

提交一个请求用给定参数创建一个预备语句并且等待完成。

```
PGresult *PQprepare(PGconn *conn,
                    const char *stmtName,
                    const char *query,
                    int nParams,
                    const Oid *paramTypes);
```

PQprepare 创建一个预备语句，供随后使用 PQexecPrepared 执行。此功能允许将被反复使用的命令只解析和规划一次，而不是每次执行时都重复进行。PQprepare 仅在使用协议 3.0 及更高版本的连接中受支持，使用协议 2.0 时会失败。

该函数从*query*串创建一个名为*stmtName*的预备语句，该串必须包含一个单一 SQL 命令。*stmtName*可以是""来创建一个未命名语句，
在这种情况下任何已存在的未命名语句将被自动替换。否则，
如果语句名称已经在当前会话中被定义，则是一种错误。如果使用了任何参数，
它们在查询中以$1、$2等引用。*nParams* 是在数组 *paramTypes[]* 中预先指定了类型的参数数量（当*nParams*为零时，该数组指针可以是NULL）。*paramTypes[]*通过 OID 指定要赋予给参数符号的数据类型。如果*paramTypes*是NULL或者该数组中任何特定元素为零，
服务器会用对待未指定类型的字符串字面量的方式为参数符号推测一种数据类型。还有，
查询能够使用编号高于*nParams*的参数符号，
它们的数据类型也会被自动推测（找出推测出的数据类型的方法见PQdescribePrepared）。

正如PQexec一样，结果通常是一个PGresult对象，其内容代表服务器端成功或失败。
返回空指针表示内存不足，或者根本无法发送命令。关于错误的更多信息请见PQerrorMessage。

用于PQexecPrepared的预备语句也可以通过执行 SQLPREPARE语句来创建。此外，虽然没有libpq函数可删除预备语句，但可以使用 SQLDEALLOCATE语句来完成。

PQexecPrepared

发送一个请求来执行一个带有给定参数的预备语句，并等待结果。

```
PGresult *PQexecPrepared(PGconn *conn,
                         const char *stmtName,
                         int nParams,
                         const char * const *paramValues,
                         const int *paramLengths,
                         const int *paramFormats,
                         int resultFormat);
```

PQexecPrepared类似于PQexecParams，但它通过已创建的预备语句的名称指定要执行的命令，而非提供查询字符串。此功能使重复使用的命令只需解析和规划一次，而不必在每次执行时都进行这些工作。该语句必须事先在当前会话中创建为预备语句。PQexecPrepared 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

参数与PQexecParams相同，只是给出了预备语句的名称而不是查询字符串，并且*paramTypes[]*参数不存在（因为在创建预备语句时已确定了参数类型）。

PQdescribePrepared

提交请求以获取有关指定预备语句的信息，并等待完成。

```
PGresult *PQdescribePrepared(PGconn *conn, const char *stmtName);
```

PQdescribePrepared允许应用程序获取关于先前创建的预备语句的信息。PQdescribePrepared 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

*stmtName*可以是""或NULL来引用未命名的语句，否则必须是现有预备语句的名称。成功时，返回一个状态为PGRES_COMMAND_OK的PGresult。函数PQnparams和PQparamtype可以应用于此 PGresult以获取有关预备语句参数的信息，函数PQnfields、PQfname、PQftype等提供有关语句的结果列（如果有）的信息。

PQdescribePortal

提交请求以获取有关指定 portal 的信息，并等待完成。

```
PGresult *PQdescribePortal(PGconn *conn, const char *portalName);
```

PQdescribePortal 允许应用程序获取先前创建的 portal 的信息。（libpq 不提供对 portal 的直接访问，但可以用此函数检查通过 DECLARE CURSOR SQL 命令创建的游标的属性。）PQdescribePortal 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

*portalName*可以是""或NULL来引用未命名的 portal，否则必须是现有 portal 的名称。成功时，将返回一个带有状态PGRES_COMMAND_OK的PGresult。函数PQnfields、PQfname、PQftype等可应用于 PGresult，以获取有关 portal 的结果列（如果有）的信息。

PGresult 结构体封装服务器返回的结果。编写 libpq 应用程序时，应注意维护 PGresult 的抽象。请使用下述访问函数获取 PGresult 的内容。避免直接引用 PGresult 结构体的字段，因为这些字段将来可能改变。

PQresultStatus

返回该命令的结果状态。

```
ExecStatusType PQresultStatus(const PGresult *res);
```

PQresultStatus可以返回以下值之一：

PGRES_EMPTY_QUERY

发送给服务器的字符串为空。

PGRES_COMMAND_OK

一个不返回数据的命令成功完成。

PGRES_TUPLES_OK

一个返回数据的命令（例如SELECT或者SHOW）成功完成。

PGRES_COPY_OUT

从服务器复制出数据的传输开始。

PGRES_COPY_IN

复制数据到服务器的传输开始。

PGRES_BAD_RESPONSE

无法理解服务器的响应。

PGRES_NONFATAL_ERROR

发生了一次非致命错误（一个提示或警告）。

PGRES_FATAL_ERROR

发生了一次致命错误。

PGRES_COPY_BOTH

向服务器复制数据/从服务器复制数据的传输开始。这当前只被用于流复制。

If the result status is PGRES_TUPLES_OK, then the functions described below can be used to retrieve the rows returned by the query. Note that a SELECT command that happens to retrieve zero rows still shows PGRES_TUPLES_OK. PGRES_COMMAND_OK is for commands that can never return rows (INSERT, UPDATE, etc.). A response of PGRES_EMPTY_QUERY might indicate a bug in the client software.

一个状态为PGRES_NONFATAL_ERROR的结果将不会被PQexec或者其他查询执行函数直接返回，这类结果将被传递给通知处理器（见第 31.11 节）。

PQresStatus

将 PQresultStatus 返回的枚举值转换为描述该状态码的字符串常量。调用者不应释放此结果。

```
char *PQresStatus(ExecStatusType status);
```

PQresultErrorMessage

返回与命令关联的错误消息；如果没有错误，则返回空字符串。

```
char *PQresultErrorMessage(const PGresult *res);
```

如果发生了错误，返回的字符串会以换行符结尾。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

紧跟着一个PQexec 或 PQgetResult调用，PQerrorMessage（在连接上）将返回与PQresultErrorMessage相同的字符串（在结果上）。不过，一个PGresult 将保持它的错误消息直到被销毁，而连接的错误消息将在后续操作被执行时被更改。当你想要知道与一个特定PGresult相关的状态，使用PQresultErrorMessage。而当你想要知道连接上最后一个操作的状态，使用PQerrorMessage。

PQresultErrorField

返回错误报告中的单个字段。

```
char *PQresultErrorField(const PGresult *res, int fieldcode);
```

fieldcode 是错误字段标识符，参见下文列出的符号。如果 PGresult 不是错误或警告结果，或者不包含指定字段，则返回 NULL。字段值通常不含末尾换行符。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

可以使用以下字段代码：

PG_DIAG_SEVERITY

严重性。字段的内容是ERROR、FATAL或PANIC（在一个错误消息中）。或者是WARNING、NOTICE、DEBUG、INFO或LOG（在一个提示消息中）。或者是其中之一的一个本地化翻译。总是存在。

PG_DIAG_SQLSTATE

用于错误的 SQLSTATE 代码。SQLSTATE 代码标识了已经发生的错误的类型，它可以被前端应用用来执行特定操作（例如错误处理）来响应一个特定数据库错误。一个可能的 SQLSTATE 代码列表可见附录 A。这个字段无法被本地化，并且总是存在。

PG_DIAG_MESSAGE_PRIMARY

主要的人类可读的错误消息（通常是一行）。总是存在。

PG_DIAG_MESSAGE_DETAIL

细节：一个可选的次级错误消息，它携带了关于问题的更多细节。可能有多行。

PG_DIAG_MESSAGE_HINT

提示：一个关于如何处理该问题的可选建议。它与细节的区别在于它提供了建议（可能不合适）而不是确切事实。可能有多行。

PG_DIAG_STATEMENT_POSITION

包含一个十进制整数的字符串，它表示一个错误游标位置，该位置是原始语句字符串的索引。第一个字符的索引是 1，位置以字符计算而不是以字节计算。

PG_DIAG_INTERNAL_POSITION

这被定义为与PG_DIAG_STATEMENT_POSITION字段相同，但是它被用在游标位置引用一个内部产生的命令而不是客户端提交的命令时。当这个字段出现时，PG_DIAG_INTERNAL_QUERY字段将总是出现。

PG_DIAG_INTERNAL_QUERY

一个失败的内部产生的命令的文本。例如，这可能是由一个 PL/pgSQL 函数发出的 SQL 查询。

PG_DIAG_CONTEXT

指示错误发生的上下文。
当前这包括活动过程语言函数的调用栈追踪以及内部生成的查询。追踪是每行一项，最近的排在最前面。

PG_DIAG_SOURCE_FILE

报告错误的源代码所在的文件名。

PG_DIAG_SOURCE_LINE

报告错误的源代码行号。

PG_DIAG_SOURCE_FUNCTION

报告错误的源代码函数的名字。

客户端负责按自身需要格式化所显示的信息，尤其应在需要时将长行折行。
错误消息字段中的换行符应当视为段落分隔，而非行分隔。

libpq 内部产生的错误包含严重性和主要消息，但通常没有其他字段。使用 3.0 之前协议的服务器返回的错误包含严重性和主要消息，有时还包含详细消息，但没有其他字段。

注意，错误字段只对PGresult对象有效，对PGconn对象无效。没有PQerrorField函数。

PQclear

释放与 PGresult 关联的存储空间。每个命令结果在不再需要时都应通过 PQclear 释放。

```
void PQclear(PGresult *res);
```

你可以在需要时一直保留PGresult对象；它不会在你发出新命令时消失，
甚至在关闭连接后也不会消失。要销毁它，你必须调用PQclear。
否则应用程序会发生内存泄漏。

31.3.2. 检索查询结果信息

这些函数用于从表示成功查询结果的PGresult对象（即状态为PGRES_TUPLES_OK的对象）中提取信息。它们也可用于提取成功 Describe 操作的结果信息：Describe 结果包含的列信息与实际执行查询时相同，但行数为零。对于其他状态值的对象，这些函数会将结果视为零行、零列。

PQntuples

返回查询结果中的行（元组）数。因为它返回一个整数结果，在 32 位操作系统上大的结果集可能会导致返回值溢出。

```
int PQntuples(const PGresult *res);
```

PQnfields

返回查询结果中每一行的列（字段）数。

```
int PQnfields(const PGresult *res);
```

PQfname

返回给定列号对应的列名。列号从 0 开始。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

```
char *PQfname(const PGresult *res,
              int column_number);
```

如果列号超出范围，将返回NULL。

PQfnumber

返回与给定列名相关联的列号。

```
int PQfnumber(const PGresult *res,
              const char *column_name);
```

如果给定的名字不匹配任何列，将返回 -1。

给定名称按 SQL 命令中的标识符处理，即除非用双引号引用，否则会转换为小写。例如，对于以下 SQL 命令生成的查询结果：

```
SELECT 1 AS FOO, 2 AS "BAR";
```

会得到以下结果：

```
PQfname(res, 0)              foo
PQfname(res, 1)              BAR
PQfnumber(res, "FOO")        0
PQfnumber(res, "foo")        0
PQfnumber(res, "BAR")        -1
PQfnumber(res, "\"BAR\"")    1
```

PQftable

返回给定列所取自的表的 OID。列号从 0 开始。

```
Oid PQftable(const PGresult *res,
             int column_number);
```

如果列号超出范围、指定的列不是对表列的简单引用，或者使用 3.0 之前的协议，则返回 InvalidOid。可以查询系统表 pg_class，确定所引用的具体表。

包含 libpq 头文件后，将定义类型 Oid 和常量 InvalidOid。它们都属于某种整数类型。

PQftablecol

返回指定查询结果列所对应的表列在表中的列号。查询结果的列号从 0 开始，而表列的编号非零。

```
int PQftablecol(const PGresult *res,
                int column_number);
```

如果列号超出范围、指定的列不是对表列的简单引用，或者使用 3.0 之前的协议，则返回零。

PQfformat

返回表示给定列格式的格式代码。列号从 0 开始。

```
int PQfformat(const PGresult *res,
              int column_number);
```

格式代码零表示文本数据，格式代码一表示二进制数据。（其他代码保留供将来定义。）

PQftype

返回与给定列号相关联的数据类型。被返回的整数是该类型的内部 OID 号。列号从 0 开始。

```
Oid PQftype(const PGresult *res,
            int column_number);
```

可以查询系统表 pg_type 来获取各种数据类型的名称和属性。内置数据类型的 OID 定义在源码树的 src/include/catalog/pg_type.h 文件中。

PQfmod

返回与给定列号相关联的列的类型修饰符。列号从 0 开始。

```
int PQfmod(const PGresult *res,
           int column_number);
```

修饰符值的含义由数据类型决定，通常表示精度或大小限制。值 -1 表示“没有可用信息”。大多数数据类型不使用修饰符，此时该值始终为 -1。

PQfsize

返回给定列号对应列的大小，以字节计。列号从 0 开始。

```
int PQfsize(const PGresult *res,
            int column_number);
```

PQfsize 返回数据库行中为该列分配的空间，即服务器内部表示该数据类型所需的大小。（因此，它对客户端用处不大。）负值表示该数据类型是变长类型。

PQbinaryTuples

如果PGresult包含二进制数据，返回 1。如果包含的是文本数据，返回 0。

```
int PQbinaryTuples(const PGresult *res);
```

此函数已弃用（与 COPY 配合使用的情况除外），因为单个 PGresult 可能在部分列中包含文本数据，而在其他列中包含二进制数据。推荐使用 PQfformat。只有结果中的所有列都采用二进制格式（格式 1）时，PQbinaryTuples 才返回 1。

PQgetvalue

返回 PGresult 中某一行的单个字段值。行号和列号从 0 开始。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

```
char *PQgetvalue(const PGresult *res,
                 int row_number,
                 int column_number);
```

对于文本格式的数据，PQgetvalue 返回字段值的字符串表示，以零字节结尾。对于二进制格式的数据，返回值采用该数据类型的 typsend 和 typreceive 函数所决定的二进制表示。（这种情况下，值后面实际上也有一个零字节，但通常没有用处，因为值本身很可能包含零字节。）

如果字段值为 null，则返回空字符串。关于如何区分 null 值与空字符串值，参见 PQgetisnull。

PQgetvalue 返回的指针指向属于 PGresult 结构体的存储空间。不应修改它所指向的数据；如果需要在 PGresult 结构体的生命周期结束后继续使用这些数据，就必须显式地将数据复制到其他存储空间。

PQgetisnull

检查字段是否为 null 值。行号和列号从 0 开始。

```
int PQgetisnull(const PGresult *res,
                int row_number,
                int column_number);
```

如果字段为 null，此函数返回 1；如果包含非 null 值，则返回 0。（注意，对于 null 字段，PQgetvalue 返回空字符串，而非空指针。）

PQgetlength

返回字段值的实际长度，以字节计。行号和列号从 0 开始。

```
int PQgetlength(const PGresult *res,
                int row_number,
                int column_number);
```

这是该数据值的实际数据长度，即 PQgetvalue 所指对象的大小。对于文本格式的数据，它与 strlen() 的结果相同。对于二进制格式，这是必不可少的信息。注意，不应依赖 PQfsize 来获取实际数据长度。

PQnparams

返回一个预备语句的参数数量。

```
int PQnparams(const PGresult *res);
```

只有在查看PQdescribePrepared的结果时，这个函数才有用。对于其他类型的查询，它将返回零。

PQparamtype

返回所指示的语句参数的数据类型。参数号从 0 开始。

```
Oid PQparamtype(const PGresult *res, int param_number);
```

只有在查看PQdescribePrepared的结果时，这个函数才有用。对于其他类型的查询，它将返回零。

PQprint

将所有行输出到指定的输出流，并可选择输出列名。

```
void PQprint(FILE *fout,      /* 输出流 */
             const PGresult *res,
             const PQprintOpt *po);
typedef struct
{
    pqbool  header;      /* 打印输出字段标题和行数 */
    pqbool  align;       /* 填充并对齐字段 */
    pqbool  standard;    /* 旧的简陋格式 */
    pqbool  html3;       /* 输出 HTML 表格 */
    pqbool  expanded;    /* 展开表格 */
    pqbool  pager;       /* 需要时使用分页器输出 */
    char    *fieldSep;   /* 字段分隔符 */
    char    *tableOpt;   /* 用于 HTML 表格元素的属性 */
    char    *caption;    /* HTML 表格标题 */
    char    **fieldName; /* 以空指针结尾的替换字段名数组 */
} PQprintOpt;
```

psql 以前用此函数打印查询结果，现在已不再使用。注意，此函数假定所有数据都是文本格式。

31.3.3. 检索其他结果信息

这些函数被用来从PGresult对象中抽取其他信息。

PQcmdStatus

返回来自于产生PGresult的 SQL 命令的命令状态标签。

```
char *PQcmdStatus(PGresult *res);
```

通常这就是该命令的名称，但是它可能包括额外数据，例如已被处理的行数。
调用者不应该直接释放该结果。它将在相关的PGresult句柄被传递给PQclear之后被释放。

PQcmdTuples

返回受该 SQL 命令影响的行数。

```
char *PQcmdTuples(PGresult *res);
```

此函数返回一个字符串，其中包含产生该 PGresult 的 SQL 语句所影响的行数。此函数只能在执行 SELECT、CREATE TABLE AS、INSERT、UPDATE、DELETE、MOVE、FETCH 或 COPY 语句之后使用，也可以在对包含 INSERT、UPDATE 或 DELETE 语句的预备查询执行 EXECUTE 之后使用。如果产生 PGresult 的是其他命令，PQcmdTuples 将返回空字符串。调用者不应直接释放返回值；当关联的 PGresult 句柄被传给 PQclear 时，返回值会被释放。

PQoidValue

如果该SQL命令是一个正好将一行插入到具有 OID 的表的INSERT，或者是一个包含合适INSERT语句的预备查询的EXECUTE，这个函数返回被插入行的 OID。否则，这个函数返回InvalidOid。如果被INSERT语句影响的表不包含 OID，这个函数也将返回InvalidOid。

```
Oid PQoidValue(const PGresult *res);
```

PQoidStatus

此函数不是线程安全的，现已弃用，推荐改用 PQoidValue。它返回包含插入行 OID 的字符串，而 PQoidValue 返回 OID 值。

```
char *PQoidStatus(const PGresult *res);
```

31.3.4. 用于在 SQL 命令中嵌入字符串的转义

PQescapeLiteral

```
char *PQescapeLiteral(PGconn *conn, const char *str, size_t length)
;
```

为了让一个字符串可用于 SQL 命令，PQescapeLiteral会对它进行转义。当在 SQL 命令中把数据值作为字符串字面量插入时，这个函数很有用。一些字符（例如引号和反斜杠）必须经过转义，才不会被 SQL 解析器解释成特殊含义。PQescapeLiteral执行这种操作。

PQescapeLiteral 返回 *str* 参数的转义版本，存放在通过 malloc() 分配的内存中。结果不再需要时，应使用 PQfreemem() 释放该内存。输入不必以零字节结尾，末尾零字节也不应计入 *length*。（如果在处理完 *length* 个字节之前遇到末尾零字节，PQescapeLiteral 会在该字节处停止；这一行为类似于 strncpy。）返回字符串中的所有特殊字符均已替换，以便 PostgreSQL 字符串字面量解析器正确处理。还会添加一个末尾零字节。包围 PostgreSQL 字符串字面量所需的单引号包含在结果字符串中。

发生错误时，PQescapeLiteral返回NULL并且一个合适的消息会被存储在*conn*对象中。

> **提示**
>
> 在处理来自不可信来源的字符串时，正确转义尤其重要。
> 否则就会有安全风险：你很容易受到“SQL 注入”攻击，不期望的 SQL 命令可能会被送入数据库。

注意，当一个数据值被作为PQexecParams或相关例程中的一个独立参数传递时，既没有必要做转义而且做转义也不正确。

PQescapeIdentifier

```
char *PQescapeIdentifier(PGconn *conn, const char *str, size_t
 length);
```

PQescapeIdentifier 对字符串进行转义，使其可用作 SQL 标识符，例如表名、列名或函数名。当用户提供的标识符可能包含未经转义就不会被 SQL 解析器视为标识符一部分的特殊字符，或包含需要保留大小写的大写字符时，此函数很有用。

PQescapeIdentifier 返回 *str* 参数作为 SQL 标识符的转义版本，存放在通过 malloc() 分配的内存中。结果不再需要时，必须使用 PQfreemem() 释放该内存。输入不必以零字节结尾，末尾零字节也不应计入 *length*。（如果在处理完 *length* 个字节之前遇到末尾零字节，PQescapeIdentifier 会在该字节处停止；这一行为类似于 strncpy。）返回字符串中的所有特殊字符均已替换，以便正确地作为 SQL 标识符处理。还会添加一个末尾零字节，并用双引号包围返回的字符串。

发生错误时，PQescapeIdentifier返回NULL并且一个合适的消息会被存储在*conn*对象中。

> **提示**
>
> 与字符串字面量一样，为防止 SQL 注入攻击，从不可信来源接收到的 SQL 标识符必须经过转义。

PQescapeStringConn

```
size_t PQescapeStringConn(PGconn *conn,
                          char *to, const char *from, size_t
 length,
                          int *error);
```

PQescapeStringConn 对字符串字面量进行转义，与 PQescapeLiteral 类似。与 PQescapeLiteral 不同，调用者需要提供大小合适的缓冲区。此外，PQescapeStringConn 不会生成包围 PostgreSQL 字符串字面量所需的单引号；应在包含转义结果的 SQL 命令中提供这些引号。*from* 参数指向待转义字符串的首字符，*length* 参数给出该字符串的字节数。输入不必以零字节结尾，末尾零字节也不应计入 *length*。（如果在处理完 *length* 个字节之前遇到末尾零字节，PQescapeStringConn 会在该字节处停止；这一行为类似于 strncpy。）*to* 必须指向一个缓冲区，其容量至少为 *length* 的两倍加一个字节，否则行为未定义。如果 *to* 与 *from* 字符串重叠，行为同样未定义。

如果 *error* 参数不是 NULL，则成功时将 *error 设为零，出错时设为非零。目前唯一可能的错误是源字符串中存在无效的多字节编码。出错时仍会生成输出字符串，但预计服务器会因其格式错误而拒绝它。发生错误时，无论 *error* 是否为 NULL，都会在 *conn* 对象中保存相应消息。

PQescapeStringConn返回写到*to*的字节数，不包括终止的零字节。

PQescapeString

PQescapeString 是 PQescapeStringConn 的旧版本，现已弃用。

```
size_t PQescapeString (char *to, const char *from, size_t length);
```

与 PQescapeStringConn 唯一的区别是，PQescapeString 不接受 PGconn 或 *error* 参数。因此，它无法根据连接属性（例如字符编码）调整行为，可能给出错误的结果。此外，它也无法报告错误情况。

PQescapeString 可以在同一时刻仅使用一个 PostgreSQL 连接的客户端程序中安全使用（在这种情况下，它能够“在内部”取得所需信息）。在其他情形下，它存在安全隐患，应改用 PQescapeStringConn。

PQescapeByteaConn

对二进制数据进行转义，使其能够在 SQL 命令中用作 bytea 类型的值。与 PQescapeStringConn 一样，这仅用于将数据直接插入 SQL 命令字符串的情况。

```
unsigned char *PQescapeByteaConn(PGconn *conn,
                                 const unsigned char *from,
                                 size_t from_length,
                                 size_t *to_length);
```

当某些字节值被用作一个SQL语句中的bytea字面量的一部分时，它们必须被转义。PQescapeByteaConn使用十六进制编码或反斜杠转义来转义这些字节。详见第 8.4 节。

from 参数指向待转义字符串的首字节，*from_length* 参数给出该二进制字符串的字节数。（末尾零字节既不需要，也不计入长度。）*to_length* 参数指向用于保存转义后字符串长度的变量。该结果字符串长度包含结果末尾的零字节。

PQescapeByteaConn 返回 *from* 参数所指二进制字符串的转义版本，存放在通过 malloc() 分配的内存中。结果不再需要时，应使用 PQfreemem() 释放该内存。返回字符串中的所有特殊字符都已替换，以便 PostgreSQL 字符串字面量解析器和 bytea 输入函数正确处理。还会添加一个末尾零字节。包围 PostgreSQL 字符串字面量所需的单引号不包含在结果字符串中。

在发生错误时，将返回一个空指针，并且一个合适的错误消息被存储在*conn*对象中。当前，唯一可能的错误是没有足够的内存用于结果串。

PQescapeBytea

PQescapeBytea 是 PQescapeByteaConn 的旧版本，现已弃用。

```
unsigned char *PQescapeBytea(const unsigned char *from,
                             size_t from_length,
                             size_t *to_length);
```

与 PQescapeByteaConn 唯一的区别是，PQescapeBytea 不接受 PGconn 参数。因此，PQescapeBytea 只能在同一时刻仅使用一个 PostgreSQL 连接的客户端程序中安全使用（在这种情况下，它能够“在内部”取得所需信息）。
如果在使用多个数据库连接的程序中使用，它可能给出错误的结果（此时应使用 PQescapeByteaConn）。

PQunescapeBytea

将二进制数据的字符串表示转换为二进制数据，这是PQescapeBytea的逆操作。
以文本格式取得bytea数据时需要此操作；以二进制格式取得时则不需要。

```
unsigned char *PQunescapeBytea(const unsigned char *from, size_t
 *to_length);
```

*from*参数指向一个字符串，例如对bytea列调用PQgetvalue时返回的字符串。
PQunescapeBytea将这个字符串表示转换为二进制表示。它返回指向通过malloc()
分配的缓冲区的指针，出错时返回NULL，并将缓冲区大小存入*to_length*。
不再需要结果时，必须使用PQfreemem释放它。

此转换并不完全是PQescapeBytea的逆操作，因为从PQgetvalue
收到的字符串并非经过“转义”的形式。具体而言，这意味着无需考虑字符串引号，
因此也不需要PGconn参数。

31.4. 异步命令处理

PQexec函数足以满足普通同步应用程序提交命令的需要。不过，它有两个缺点，
对某些用户可能很重要：

- PQexec waits for the command to be completed. The application might have other work to do (such as maintaining a user interface), in which case it won't want to block waiting for the response.
- 因为客户端应用的执行在它等待结果时会被挂起，
对于应用来说很难决定要不要尝试取消正在进行的命令（这可以在一个信号处理器中完成，
但别无他法）。

- PQexec只能返回一个PGresult结构体。如果提交的命令串包含多个SQL命令，除了最后一个PGresult之外都会被PQexec丢弃。

如果应用程序不希望受到这些限制，可以改用构成 PQexec 的底层函数：PQsendQuery 和 PQgetResult。此外，PQsendQueryParams、PQsendPrepare、PQsendQueryPrepared、PQsendDescribePrepared、PQsendDescribePortal 可以与 PQgetResult 配合使用，分别实现 PQexecParams、PQprepare、PQexecPrepared、PQdescribePrepared、PQdescribePortal 的功能。

PQsendQuery

向服务器提交命令，不等待结果。命令发送成功时返回 1，否则返回 0（此时可使用 PQerrorMessage 获取更多失败信息）。

```
int PQsendQuery(PGconn *conn, const char *command);
```

成功调用 PQsendQuery 后，应调用 PQgetResult 一次或多次来获取结果。在 PQgetResult 返回空指针、表明命令已完成之前，不得在同一连接上再次调用 PQsendQuery。

PQsendQueryParams

向服务器提交命令及独立指定的参数，不等待结果。

```
int PQsendQueryParams(PGconn *conn,
                      const char *command,
                      int nParams,
                      const Oid *paramTypes,
                      const char * const *paramValues,
                      const int *paramLengths,
                      const int *paramFormats,
                      int resultFormat);
```

该函数等价于 PQsendQuery，但查询参数可以与查询字符串分开指定。函数参数的处理方式与 PQexecParams 相同。与 PQexecParams 一样，它不能用于协议 2.0 的连接，并且查询字符串中只允许包含一条命令。

PQsendPrepare

发送按给定参数创建预备语句的请求，不等待完成。

```
int PQsendPrepare(PGconn *conn,
                  const char *stmtName,
                  const char *query,
                  int nParams,
                  const Oid *paramTypes);
```

这是 PQprepare 的异步版本：请求发送成功时返回 1，否则返回 0。调用成功后，再调用 PQgetResult，确定服务器是否成功创建了预备语句。函数参数的处理方式与 PQprepare 相同。与 PQprepare 一样，它不能用于协议 2.0 的连接。

PQsendQueryPrepared

发送使用给定参数执行预备语句的请求，不等待结果。

```
int PQsendQueryPrepared(PGconn *conn,
                        const char *stmtName,
```

```
                                int nParams,
                                const char * const *paramValues,
                                const int *paramLengths,
                                const int *paramFormats,
                                int resultFormat);
```

该函数类似于 PQsendQueryParams，但通过已创建的预备语句的名称指定要执行的命令，而非提供查询字符串。函数参数的处理方式与 PQexecPrepared 相同。与 PQexecPrepared 一样，它不能用于协议 2.0 的连接。

PQsendDescribePrepared

提交获取指定预备语句信息的请求，不等待完成。

```
int PQsendDescribePrepared(PGconn *conn, const char *stmtName);
```

这是PQdescribePrepared的异步版本：请求发送成功时返回 1，否则返回 0。调用成功后，再调用PQgetResult获取结果。函数参数的处理方式与PQdescribePrepared相同。与PQdescribePrepared一样，它不能用于协议 2.0 的连接。

PQsendDescribePortal

提交获取指定 portal 信息的请求，不等待完成。

```
int PQsendDescribePortal(PGconn *conn, const char *portalName);
```

这是PQdescribePortal的异步版本：请求发送成功时返回 1，否则返回 0。调用成功后，再调用PQgetResult获取结果。函数参数的处理方式与PQdescribePortal相同。与PQdescribePortal一样，它不能用于协议 2.0 的连接。

PQgetResult

等待先前的PQsendQuery、PQsendQueryParams、PQsendPrepare或PQsendQueryPrepared调用产生的下一个结果，并返回该结果。当命令执行完毕且不再有其他结果时，返回空指针。

```
PGresult *PQgetResult(PGconn *conn);
```

必须反复调用 PQgetResult，直到它返回空指针，表明命令已经完成。（如果当前没有正在执行的命令，调用 PQgetResult 会立即返回空指针。）对于 PQgetResult 返回的非空指针，应使用前文介绍的 PGresult 访问函数处理相应结果。使用完毕后，不要忘记调用 PQclear 释放每个结果对象。注意，只有存在正在执行的命令，且所需响应数据尚未被 PQconsumeInput 读取时，PQgetResult 才会阻塞。

> **注意**
>
> 即使PQresultStatus指示发生了致命错误，也应该调用PQgetResult直到它返回一个空指针，以便libpq完全处理错误信息。

使用PQsendQuery和PQgetResult解决了PQexec的一个问题：如果命令字符串包含多个SQL命令，就能分别获取这些命令的结果。（这也支持一种简单的重叠处理方式：客户端可以处理某条命令的结果，同时服务器继续处理同一命令字符串中后面的查询。）然而，调用PQgetResult

仍会导致客户端阻塞直到服务器完成下一条SQL命令。
通过恰当使用另外两个函数可以避免这一点：

PQconsumeInput

If input is available from the server, consume it.

```
int PQconsumeInput(PGconn *conn);
```

PQconsumeInput 通常返回 1，表示“没有错误”；发生问题时则返回 0（此时可查看 PQerrorMessage）。注意，返回值并不说明是否实际读取了输入数据。调用 PQconsumeInput 后，应用程序可以检查 PQisBusy 和/或 PQnotifies，以确定其状态是否发生变化。

即使应用程序尚未准备好处理结果或通知，也可以调用 PQconsumeInput。此函数会读取可用数据并将其保存在缓冲区中，从而清除 select() 的可读就绪指示。因此，应用程序可以用 PQconsumeInput 立即清除 select() 的就绪条件，随后在合适的时候检查结果。

PQisBusy

如果命令仍在忙碌，则返回 1，意味着PQgetResult会阻塞等待输入。返回 0 则表示可以调用PQgetResult，并保证不会阻塞。

```
int PQisBusy(PGconn *conn);
```

PQisBusy本身将不会尝试从服务器读取数据，因此必须先调用PQconsumeInput，否则繁忙状态将永远不会结束。

使用这些函数的典型应用程序会在主循环中通过 select() 或 poll() 等待需要响应的各种条件。其中一个条件是服务器有可读取的输入；对于 select()，这意味着 PQsocket 标识的文件描述符上有可读数据。主循环检测到输入就绪时，应调用 PQconsumeInput 读取输入，然后调用 PQisBusy。如果 PQisBusy 返回假（0），就可以接着调用 PQgetResult。还可以调用 PQnotifies 检测 NOTIFY 消息（见第 31.7 节）。

一个使用PQsendQuery/PQgetResult
的客户端也可以尝试取消一个正在被服务器处理的命令，见第 31.5 节。但是，不管PQcancel的返回值是什么，应用都必须继续使用PQgetResult进行正常的结果读取序列。
一次成功的取消只会导致命令比不取消时更快终止。

使用上述函数可以避免在等待数据库服务器输入时阻塞。不过，
应用程序仍可能在等待向服务器发送输出时阻塞。这种情况较少见，但发送很长的 SQL 命令或数据值时可能发生。（如果应用程序通过 COPY IN 发送数据，发生的可能性则大得多。）为了防止这种情况，实现完全非阻塞的数据库操作，可以使用以下附加函数。

PQsetnonblocking

设置连接的非阻塞状态。

```
int PQsetnonblocking(PGconn *conn, int arg);
```

如果*arg*为1，则将连接状态设置为非阻塞，如果 *arg*为0，则设置为阻塞。如果成功返回0，出错返回-1。

在非阻塞状态下，对 PQsendQuery、PQputline、PQputnbytes 和 PQendcopy 的调用不会阻塞；如果需要再次调用，它们会返回错误。

请注意，PQexec不遵守非阻塞模式；如果调用它，它将以阻塞方式执行。

PQisnonblocking

返回数据库连接的阻塞状态。

```
int PQisnonblocking(const PGconn *conn);
```

如果连接设置为非阻塞模式，则返回1，如果为阻塞，则返回0。

PQflush

尝试将发送队列中的输出数据发送到服务器。成功（或发送队列为空）时返回 0；因某种原因失败时返回 -1；如果尚未能发送队列中的全部数据，则返回 1（这种情况只可能发生在非阻塞连接上）。

```
int PQflush(PGconn *conn);
```

在非阻塞连接上发送命令或数据后，应调用 PQflush。如果返回 1，就等待套接字变为可读或可写。套接字可写时，再次调用 PQflush；可读时，先调用 PQconsumeInput，再调用 PQflush。重复上述步骤，直到 PQflush 返回 0。（必须检查套接字是否可读，并用 PQconsumeInput 读完输入，因为服务器可能在尝试向客户端发送数据时阻塞，例如发送 NOTICE 消息；在客户端读取这些数据之前，服务器不会读取客户端发送的数据。）当 PQflush 返回 0 后，等待套接字变为可读，再按前述方法读取响应。

31.5. 取消进行中的查询

客户端应用程序可以使用本节介绍的函数，请求取消服务器仍在处理的命令。

PQgetCancel

创建一个数据结构，其中包含取消通过特定数据库连接发出的命令所需的信息。

```
PGcancel *PQgetCancel(PGconn *conn);
```

PQgetCancel基于一个PGconn连接对象创建PGcancel对象。如果给定的*conn*为NULL或无效连接，则返回NULL。PGcancel是不透明结构体，不应由应用程序直接访问；它只能传给PQcancel或PQfreeCancel。

PQfreeCancel

释放由以下函数创建的数据结构：PQgetCancel。

```
void PQfreeCancel(PGcancel *cancel);
```

PQfreeCancel 释放之前由 PQgetCancel 创建的数据对象。

PQcancel

请求服务器停止处理当前命令。

```
int PQcancel(PGcancel *cancel, char *errbuf, int errbufsize);
```

返回值为 1 表示取消请求成功发送，为 0 表示未成功发送。如果未成功发送，*errbuf*将填充解释性错误消息。*errbuf*必须是大小为*errbufsize*的字符数组（推荐大小为 256 字节）。

不过，成功发送请求并不保证请求会产生效果。如果取消生效，
当前命令会提前终止并返回错误结果。如果取消失败（例如服务器已经处理完该命令），
则不会产生任何可见结果。

如果*errbuf*是信号处理程序中的局部变量，则PQcancel
可以安全地从信号处理程序中调用。对于PQcancel而言，PGcancel对象是只读的，
因此也可以从与操作PGconn对象的线程不同的线程中调用。

PQrequestCancel

PQrequestCancel是以下函数的已弃用变体：PQcancel。

```
int PQrequestCancel(PGconn *conn);
```

请求服务器放弃当前命令的处理。它直接作用于PGconn对象，失败时会把错误消息存储到
PGconn对象中（可通过PQerrorMessage获取）。虽然功能相同，
但这种方法在多线程程序或信号处理程序中并不安全，因为它可能覆盖PGconn
中的错误消息，从而破坏当前连接上正在进行的操作。

31.6. 快速路径接口

PostgreSQL提供一种快速路径接口来向服务器发送简单的函数调用。

> **提示**
>
> 此接口已显得有些过时，因为可以通过建立定义函数调用的预备语句，
> 获得相近的性能和更多功能。随后使用二进制格式传输参数和结果来执行该语句，
> 就可以替代快速路径函数调用。

函数PQfn请求通过快速路径接口执行服务器函数。

```
PGresult *PQfn(PGconn *conn,
               int fnid,
               int *result_buf,
               int *result_len,
               int result_is_int,
               const PQArgBlock *args,
               int nargs);

typedef struct
{
    int len;
    int isint;
    union
    {
        int *ptr;
        int integer;
    } u;
} PQArgBlock;
```

fnid 参数是要执行函数的 OID。*args* 和 *nargs* 指定传给函数的参数，必须与函数声明中的参数列表匹配。参数结构体的 *isint* 字段为真时，*u.integer* 值会以指定长度的整数发送到服务器，该长度必须是 2 或 4 字节，并会进行适当的字节序转换。*isint* 为假时，位于 **u.*

ptr 的指定数量字节会原样发送；数据必须符合服务器对该函数参数数据类型的二进制传输格式要求。（将 *u.ptr* 声明为 int * 是历史原因；将其视为 void * 更合适。）*result_buf* 指向用于存放函数返回值的缓冲区。调用者必须事先分配足够空间来保存返回值，这里不会检查！实际结果长度以字节为单位，返回到 *result_len* 指向的整数中。如果预期结果是 2 或 4 字节整数，将 *result_is_int* 设为 1，否则设为 0。将 *result_is_int* 设为 1 后，libpq 会按需转换字节序，使结果成为适合客户端机器的 int 值；注意，无论是哪种允许的结果大小，传入 **result_buf* 的都是 4 字节整数。*result_is_int* 为 0 时，服务器发送的二进制格式字节串会原样返回。（此时，将 *result_buf* 视为 void * 更合适。）

PQfn 总是返回有效的 PGresult 指针。使用结果前应检查其状态。不再需要结果时，调用者负责使用 PQclear 释放 PGresult。

注意，使用此接口时无法处理 NULL 参数、NULL 结果以及集合值结果。

31.7. 异步通知

PostgreSQL通过LISTEN和NOTIFY命令提供异步通知。客户端会话可使用LISTEN命令注册其感兴趣的特定通知通道（也可以用UNLISTEN命令停止监听）。当任何会话执行带有该通道名的NOTIFY命令时，所有监听该通道的会话都会异步收到通知。还可以传递一个“载荷”字符串，向监听者传递附加数据。

libpq应用把LISTEN、UNLISTEN和NOTIFY命令作为普通 SQL 命令提交。随后通过调用PQnotifies来检测NOTIFY消息的到达。

函数 PQnotifies 从已收到但尚未处理的服务器通知消息列表中返回下一条通知。没有待处理通知时返回空指针。一旦 PQnotifies 返回一条通知，该通知就被视为已处理，并从通知列表中移除。

```
PGnotify *PQnotifies(PGconn *conn);

typedef struct pgNotify
{
    char *relname;              /* 通知通道名称 */
    int  be_pid;                /* 发送通知的服务器进程 ID */
    char *extra;                /* 通知载荷字符串 */
} PGnotify;
```

处理完 PQnotifies 返回的 PGnotify 对象后，务必使用 PQfreemem 释放它。只需释放 PGnotify 指针；relname 和 extra 字段并非独立分配。（这些字段名称是历史遗留的；尤其是，通道名称与关系名称不必有任何关联。）

例 31.2给出了一个示例程序展示异步通知的使用。

PQnotifies实际上并不从服务器读取数据；它只是返回之前已被其他libpq函数读取的消息。在较早版本的libpq中，及时收到NOTIFY消息的唯一方法是不断提交命令，哪怕是空命令，然后在每次PQexec后检查PQnotifies。虽然这种方法仍然有效，但由于浪费处理能力，现已弃用。

当没有需要执行的实际命令时，一种更好的检查NOTIFY消息的方法是调用PQconsumeInput，然后检查PQnotifies。你可以使用select()等待服务器数据到达，这样在无事可做时就不会浪费CPU资源（参见PQsocket以获得可传给select()的文件描述符）。注意不管是用PQsendQuery/PQgetResult提交命令还是简单地使用PQexec，这种方法都能正常工作。不过，你应该记住在每次PQgetResult 或 PQexec之后检查PQnotifies，看看在命令的处理过程中是否有通知到达。

31.8. 与COPY命令相关的函数

PostgreSQL 的 COPY 命令提供了选项，可以通过 libpq 使用的网络连接读取或写入数据。本节介绍的函数允许应用程序通过提供或接收复制数据来使用这一能力。

整体流程如下：应用程序先通过 PQexec 或等效函数发出 SQL COPY 命令。如果命令没有错误，响应就是一个 PGresult 对象，其状态码为 PGRES_COPY_OUT 或 PGRES_COPY_IN，取决于指定的复制方向。应用程序随后应使用本节函数接收或发送数据行。数据传输完成后，会返回另一个 PGresult 对象，表示传输成功或失败：成功时状态为 PGRES_COMMAND_OK，出现问题时为 PGRES_FATAL_ERROR。此时可以通过 PQexec 继续发出 SQL 命令。（COPY 操作进行期间，不能在同一连接上执行其他 SQL 命令。）

如果一个COPY命令是通过PQexec在一个可能包含额外命令的字符串中发出的，那么应用在完成COPY序列之后必须继续用PQgetResult取得结果。只有在PQgetResult返回NULL时，我们才能确信PQexec的命令字符串已经处理完毕，并且可以安全地发出更多命令。

只有从 PQexec 或 PQgetResult 获得 PGRES_COPY_OUT 或 PGRES_COPY_IN 结果状态后，才应调用本节函数。

带有上述某个状态值的 PGresult 对象，还会携带关于即将开始的 COPY 操作的附加数据。这些数据可以通过下列函数获取，这些函数也用于查询结果：

PQnfields

返回要复制的列（字段）的数量。

PQbinaryTuples

0表示整体复制格式为文本（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式为二进制。有关更多信息，请参见COPY。

PQfformat

返回与复制操作的每列关联的格式代码（0表示文本，1表示二进制）。
当整体复制格式为文本时，每列的格式代码始终为零，
但二进制格式可以支持文本和二进制列。（但是，在当前的 COPY 实现中，只有二进制列出现在二进制复制中；因此，每列格式目前始终与整体格式匹配。）

> **注意**
>
> 这些附加数据值仅在使用协议 3.0 时可用。使用协议 2.0 时，这些函数都返回 0。

31.8.1. 发送COPY数据的函数

这些函数用于在 COPY FROM STDIN 期间发送数据。如果连接不处于 COPY_IN 状态，调用它们会失败。

PQputCopyData

在COPY_IN状态中向服务器发送数据。

```
int PQputCopyData(PGconn *conn,
```

```
                    const char *buffer,
                    int nbytes);
```

将指定 *buffer* 中长度为 *nbytes* 的 COPY 数据传输到服务器。数据发送成功时返回 1；因尝试会阻塞而未发送时返回零（仅可能发生在连接处于非阻塞模式时）；发生错误时返回 -1。（返回 -1 时，可用 PQerrorMessage 获取详细信息。返回零时，应等待可写就绪后重试。）

应用程序可以将 COPY 数据流分成任意方便大小的数据块，逐块装入缓冲区。发送时，这些数据块的边界没有语义含义。数据流内容必须符合 COPY 命令预期的数据格式；详见COPY。

PQputCopyEnd

在COPY_IN状态中向服务器发送数据结束的指示。

```
int PQputCopyEnd(PGconn *conn,
                 const char *errormsg);
```

如果 *errormsg* 为 NULL，则成功结束 COPY_IN 操作。如果 *errormsg* 不为 NULL，则强制 COPY 失败，并将 *errormsg* 指向的字符串用作错误消息。（但不应假定服务器一定会返回这条完全相同的错误消息，因为服务器可能已经因自身原因使 COPY 失败。还要注意，在使用 3.0 之前协议的连接上，强制失败选项不起作用。）

终止数据发送成功时返回 1；因尝试会阻塞而未发送时返回零（仅可能发生在连接处于非阻塞模式时）；发生错误时返回 -1。（返回 -1 时，可用 PQerrorMessage 获取详细信息。返回零时，应等待可写就绪后重试。）

成功调用 PQputCopyEnd 后，调用 PQgetResult 获取 COPY 命令的最终结果状态。可以按通常方式等待该结果就绪，然后恢复正常操作。

31.8.2. 接收COPY数据的函数

这些函数用于在COPY TO STDOUT的过程中接收数据。如果连接不在COPY_OUT状态，那么调用它们将会失败。

PQgetCopyData

在COPY_OUT状态下从服务器接收数据。

```
int PQgetCopyData(PGconn *conn,
                  char **buffer,
                  int async);
```

在 COPY 期间尝试从服务器获取下一行数据。每次总是返回一个完整数据行；如果只有部分行可用，则不返回。成功返回数据行时，会分配一块内存保存数据。*buffer* 参数必须为非 NULL。**buffer* 会被设置为指向所分配的内存；如果没有返回缓冲区，则设为 NULL。非 NULL 的结果缓冲区在不再需要时应使用 PQfreemem 释放。

成功返回一行时，返回值是该行的数据字节数，始终大于零。
返回的字符串总是以零字节结尾，不过这可能仅对文本 COPY 有用。返回零表示 COPY 仍在进行，但尚无可用行（仅在 *async* 为真时可能发生）。返回 -1 表示 COPY 已完成；返回 -2 表示发生了错误（可用 PQerrorMessage 查看原因）。

当 *async* 为真（非零）时，PQgetCopyData 不会阻塞等待输入；如果 COPY 仍在进行，但没有完整行可用，则返回零。（此时，应等待读就绪，先调用 PQconsumeInput，再调用

PQgetCopyData。）当 *async* 为假（零）时，PQgetCopyData 会阻塞，直到数据可用或操作完成。

在 PQgetCopyData 返回 -1 后，调用 PQgetResult 获取 COPY 命令的最终结果状态。可以按通常方式等待该结果就绪，然后恢复正常操作。

31.8.3. COPY的已废弃函数

这些函数使用较旧的方式处理 COPY。虽然仍然可用，但由于错误处理欠佳、检测数据结束的方式不便，而且缺少对二进制或非阻塞传输的支持，已被弃用。

PQgetline

将服务器传来的、以换行符结尾的一行字符读入大小为 *length* 的字符串缓冲区。

```
int PQgetline(PGconn *conn,
              char *buffer,
              int length);
```

此函数最多将 *length*-1 个字符复制到缓冲区，并将末尾的换行符转换为零字节。PQgetline 在输入结束时返回 EOF，读完一整行时返回 0，缓冲区已满但尚未读到末尾换行符时返回 1。

注意，应用程序必须检查新读入的一行是否仅由 \. 两个字符组成，这表示服务器已发送完 COPY 命令的结果。如果可能收到长度超过 *length*-1 个字符的行，必须确保正确识别 \. 行，例如不能把长数据行的末尾误当作终止行。

PQgetlineAsync

以非阻塞方式将服务器传来的一行 COPY 数据读入缓冲区。

```
int PQgetlineAsync(PGconn *conn,
                   char *buffer,
                   int bufsize);
```

此函数类似于 PQgetline，但可用于必须异步读取 COPY 数据的应用程序，即读取时不阻塞。发出 COPY 命令并收到 PGRES_COPY_OUT 响应后，应用程序应调用 PQconsumeInput 和 PQgetlineAsync，直到检测到数据结束信号。

与 PQgetline 不同，此函数会负责检测数据结束。

每次调用时，如果 libpq 的输入缓冲区中有完整数据行，PQgetlineAsync 就会返回数据；否则，要等该行剩余部分到达后才返回数据。识别到复制数据结束标记时返回 -1，没有可用数据时返回 0，否则返回正数，表示返回的数据字节数。返回 -1 后，调用者必须接着调用 PQendcopy，然后恢复正常处理。

返回的数据不会跨越数据行边界。只要可能，每次就返回一整行；
但如果调用者提供的缓冲区太小，容不下服务器发送的一行，则只返回部分行。
对于文本数据，可检查最后返回的字节是否为 \n，以判断是否返回了完整行。（对于二进制 COPY，则必须实际解析 COPY 数据格式才能作出相同判断。）返回的字符串不以零字节结尾。（如果要自行添加末尾的零字节，务必将传入的 *bufsize* 设置为比实际可用空间少一字节。）

PQputline

向服务器发送以零字节结尾的字符串。成功时返回 0，无法发送字符串时返回 EOF。

```
int PQputline(PGconn *conn,
              const char *string);
```

连续调用 PQputline 发送的 COPY 数据流，与 PQgetlineAsync 返回的数据格式相同。不过，应用程序不必在每次 PQputline 调用中恰好发送一个数据行；每次发送部分行或多行也可以。

> **注意**
>
> 在 PostgreSQL 协议 3.0 之前，应用程序必须显式发送由 \. 两个字符组成的最后一行，告知服务器应用程序已发送完 COPY 数据。虽然这种方式仍然有效，但已被弃用，\. 的特殊含义预计会在未来版本中移除。发送完实际数据后，调用 PQendcopy 即可。

PQputnbytes

向服务器发送不以零字节结尾的字符串。成功时返回 0，无法发送字符串时返回 EOF。

```
int PQputnbytes(PGconn *conn,
                const char *buffer,
                int nbytes);
```

此函数与 PQputline 完全相同，只是直接指定了要发送的字节数，因此数据缓冲区不必以零字节结尾。发送二进制数据时可使用此函数。

PQendcopy

与服务器同步。

```
int PQendcopy(PGconn *conn);
```

此函数会等待服务器完成复制。应在使用PQputline向服务器发送最后一个字符串后，或使用PGgetline从服务器接收最后一个字符串后调用它。必须调用此函数，否则服务器与客户端会“失去同步”。函数返回后，服务器便准备好接收下一条 SQL 命令。成功完成时返回 0，否则返回非零值。（返回非零值时，可用PQerrorMessage获取详细信息。）

使用 PQgetResult 时，收到 PGRES_COPY_OUT 结果后，应用程序应反复调用 PQgetline，并在看到终止行后调用 PQendcopy。随后应回到 PQgetResult 循环，直到 PQgetResult 返回空指针。类似地，收到 PGRES_COPY_IN 结果后，应连续调用 PQputline，再调用 PQendcopy，然后回到 PQgetResult 循环。这样可以保证嵌在一系列 SQL 命令中的 COPY 命令正确执行。

旧的应用很可能会通过PQexec提交一个COPY命令并且假定事务在PQendcopy之后完成。只有在COPY是命令字符串中唯一的SQL命令时才能正确工作。

31.9. 控制函数

这些函数控制 libpq 行为的各种细节。

PQclientEncoding

Returns the client encoding.

```
int PQclientEncoding(const PGconn *conn);
```

注意，返回的是编码 ID，而不是 `EUC_JP` 这样的符号字符串。要将编码 ID 转换为编码名称，可以使用：

```
char *pg_encoding_to_char(int encoding_id);
```

`PQsetClientEncoding`

设置客户端编码。

```
int PQsetClientEncoding(PGconn *conn, const char *encoding);
```

*conn*是到服务器的连接，而*encoding*是要使用的编码。如果成功设置编码，函数返回 0，否则返回 -1。此连接的当前编码可以通过以下函数确定：`PQclientEncoding`。

`PQsetErrorVerbosity`

设置以下函数所返回消息的详细程度：`PQerrorMessage`和`PQresultErrorMessage`。

```
typedef enum
{
    PQERRORS_TERSE,
    PQERRORS_DEFAULT,
    PQERRORS_VERBOSE
} PGVerbosity;

PGVerbosity PQsetErrorVerbosity(PGconn *conn, PGVerbosity
 verbosity);
```

`PQsetErrorVerbosity`设置详细程度模式，并返回该连接先前的设置。在TERSE模式下，返回的消息只包括严重性、主要文本和位置，通常一行就能容纳。
默认模式生成的消息除上述字段外，还包括详细信息、
提示或上下文字段（这些内容可能跨越多行）。VERBOSE模式包括所有可用字段。
改变详细程度不会影响已有`PGresult`对象中的消息，只影响随后创建的对象。

`PQtrace`

启用客户端/服务器通信跟踪，将跟踪信息写入调试文件流。

```
void PQtrace(PGconn *conn, FILE *stream);
```

> **注意**
>
> 在 Windows 上，如果 libpq 库与应用程序使用不同的标志编译，调用此函数会导致应用程序崩溃，因为 `FILE` 指针的内部表示不同。具体而言，库与所有使用它的应用程序都应使用相同的多线程/单线程、发布/调试和静态/动态标志。

`PQuntrace`

禁用由 `PQtrace` 启动的跟踪。

```
void PQuntrace(PGconn *conn);
```

31.10. 杂项函数

一如往常，总有一些函数不适合放在任何其他地方。

`PQfreemem`

释放libpq分配的内存。

```
void PQfreemem(void *ptr);
```

释放 libpq 分配的内存，特别是 `PQescapeByteaConn`、`PQescapeBytea`、`PQunescapeBytea` 和 `PQnotifies` 分配的内存。在 Microsoft Windows 上，务必使用此函数，而不是 `free()`。这是因为只有 DLL 与应用程序使用相同的多线程/单线程、发布/调试和静态/动态标志，才能在 DLL 中分配内存并在应用程序中释放它。在 Microsoft Windows 以外的平台上，此函数与标准库函数 `free()` 相同。

`PQconninfoFree`

释放`PQconndefaults` 或 `PQconninfoParse`分配的数据结构。

```
void PQconninfoFree(PQconninfoOption *connOptions);
```

仅调用 `PQfreemem` 不足以完成此项释放，因为数组还包含指向附属字符串的引用。

`PQencryptPassword`

准备一个PostgreSQL密码的加密形式。

```
char * PQencryptPassword(const char *passwd, const char *user);
```

这个函数旨在用于那些希望发送类似于`ALTER USER joe PASSWORD 'pwd'`命令的客户端应用。不在这样一个命令中发送原始的明文密码是一个好习惯，
因为它可能被暴露在命令日志、活动显示等等中。相反，
在发送之前使用这个函数可以将密码转换为加密的形式。
参数是明文密码和该密码所属用户的 SQL 名称。返回值是由`malloc`分配的字符串，如果内存不足则为`NULL`。调用者可以假定该字符串不包含任何需要转义的特殊字符。
用完之后用`PQfreemem`释放结果。

`PQmakeEmptyPGresult`

用给定的状态，构造一个空`PGresult`对象。

```
PGresult *PQmakeEmptyPGresult(PGconn *conn, ExecStatusType status);
```

这是libpq内部用于分配并初始化一个空`PGresult`对象的函数。如果无法分配内存，
此函数返回`NULL`。将它导出供外部调用，是因为一些应用需要自行生成结果对象，
特别是带有错误状态的对象。如果*conn* 不为 null，并且*status*表示错误，指定连接的当前错误消息会被复制到`PGresult`中。此外，如果*conn* 不为 null，连接中注册的所有事件过程也会被复制到`PGresult`中。（这些过程不会收到`PGEVT_RESULTCREATE`调用，但可参见`PQfireResultCreateEvents`。）注意，最终应对该对象调用`PQclear`，就像处理libpq自身返回的`PGresult`一样。

`PQfireResultCreateEvents`

为每一个在`PGresult`对象中注册的事件过程触发一个`PGEVT_RESULTCREATE`事件（见第 31.12 节）。成功时返回非 0，如果任何事件过程失败则返回 0。

```
int PQfireResultCreateEvents(PGconn *conn, PGresult *res);
```

conn 参数会传递给事件过程，但此函数不会直接使用它。如果事件过程不使用此参数，则可以传入 NULL。

已经接收到这个对象的PGEVT_RESULTCREATE或PGEVT_RESULTCOPY事件的事件过程不会被再次触发。

此函数与 PQmakeEmptyPGresult 分开的主要原因是，通常适合先创建 PGresult 并填充数据，然后再调用事件过程。

PQcopyResult

创建 PGresult 对象的副本。副本与源结果没有任何关联，不再需要副本时必须调用 PQclear。函数失败时返回 NULL。

```
PGresult *PQcopyResult(const PGresult *src, int flags);
```

这不是为了制作一个精确的副本。返回的结果总是放在PGRES_TUPLES_OK状态中，并且不复制源中的任何错误消息。（但是会复制命令状态字符串。）*flags* 参数确定要复制的其他内容。它是几个标志的按位或。PG_COPYRES_ATTRS 指定复制源结果的属性（列定义）。PG_COPYRES_TUPLES指定复制源结果的元组。（这也意味着复制属性。）PG_COPYRES_NOTICEHOOKS指定复制源结果的通知钩子。PG_COPYRES_EVENTS指定复制源结果的事件。（但不复制与源相关的任何实例数据。）

PQsetResultAttrs

设置PGresult对象的属性。

```
int PQsetResultAttrs(PGresult *res, int numAttributes, PGresAttDesc
 *attDescs);
```

提供的*attDescs*被复制到结果中。如果*attDescs*指针为NULL或*numAttributes*小于1，那么请求将被忽略并且函数成功。如果*res*已经包含属性，那么函数会失败。如果函数失败，返回值是 0。如果函数成功，返回值是非 0。

PQsetvalue

设置 PGresult 对象中某个元组的字段值。

```
int PQsetvalue(PGresult *res, int tup_num, int field_num, char
 *value, int len);
```

此函数会根据需要自动扩展结果内部的元组数组。不过，*tup_num* 参数必须小于或等于 PQntuples，也就是说，每次只能向元组数组增加一个元组。已有元组的任何字段都可以按任意顺序修改。如果 *field_num* 指定的位置已有值，该值会被覆盖。如果 *len* 为 -1 或 *value* 为 NULL，则将该字段设置为 SQL null 值。*value* 会被复制到结果的私有存储中，因此函数返回后就不再需要它。函数失败时返回零，成功时返回非零值。

PQresultAlloc

为一个PGresult对象分配附属存储。

```
void *PQresultAlloc(PGresult *res, size_t nBytes);
```

使用此函数分配的所有内存都会在清除 *res* 时释放。函数失败时返回 NULL。与 malloc 一样，返回的内存保证满足任意数据类型的对齐要求。

PQlibVersion

返回所使用的libpq版本。

```
int PQlibVersion(void);
```

在运行时，这个函数的结果可以被用来判断在当前已载入的 libpq 版本中特定的功能是否可用。例如，这个函数可以被用来判断PQconnectdb可以使用哪些连接选项，或者是否支持PostgreSQL 9.0 中加入的bytea的hex输出格式。

这个数字是这样形成的：将主版本、次版本和修订版本号分别转换成两位十进制数，然后把它们拼接在一起。例如，版本9.1将被返回为90100，而版本9.1.2将被返回为90102（不显示前导零）。

> **注意**
>
> 这个函数出现于PostgreSQL版本 9.1，因此它不能被用来在早期的版本中检测所需的功能，因为链接到它将会创建一个对版本9.1的链接依赖。

31.11. 通知处理

服务器生成的通知和警告消息不由查询执行函数返回，因为它们并不表示查询失败。这些消息会传递给通知处理函数，处理函数返回后，执行正常继续。默认通知处理函数将消息打印到 stderr，应用程序也可以提供自己的处理函数来替换这一行为。

由于历史原因，通知处理有两个级别，称为通知接收器和通知处理器。通知接收器的默认行为是格式化通知并且将一个字符串传递给通知处理器来打印。不过，如果一个应用选择提供自己的通知接收器，它通常会忽略通知处理器层并且在通知接收器中完成所有工作。

函数PQsetNoticeReceiver 为一个连接对象设置或者检查当前的通知接收器。相似地，PQsetNoticeProcessor 设置或检查当前的通知处理器。

```
typedef void (*PQnoticeReceiver) (void *arg, const PGresult *res);

PQnoticeReceiver
PQsetNoticeReceiver(PGconn *conn,
                    PQnoticeReceiver proc,
                    void *arg);

typedef void (*PQnoticeProcessor) (void *arg, const char *message);

PQnoticeProcessor
PQsetNoticeProcessor(PGconn *conn,
                     PQnoticeProcessor proc,
                     void *arg);
```

这些函数中的每一个会返回之前的通知接收器或处理器函数指针，并且设置新值。如果你提供了一个空函数指针，将不会采取任何动作，只会返回当前指针。

从服务器收到通知或警告消息，或者 libpq 内部生成这类消息时，便会调用通知接收器函数。消息以状态为 PGRES_NONFATAL_ERROR 的 PGresult 对象传给该函数。（接收器可使用 PQresultErrorField 提取单个字段，或使用 PQresultErrorMessage 获取完整的预格式化消息。）传给 PQsetNoticeReceiver 的同一个 void 指针也会传入。（必要时，可通过该指针访问应用程序特有的状态。）

默认的通知接收器会简单地抽取消息（使用PQresultErrorMessage）
并且将它传递给通知处理器。

通知处理器负责处理以文本形式提供的通知或警告消息。
传给它的参数包括消息字符串（含末尾的换行符），以及先前传给 PQsetNoticeProcessor 的同一个 void 指针。（必要时，可通过该指针访问应用程序特有的状态。）

默认的通知处理器很简单：

```
static void
defaultNoticeProcessor(void *arg, const char *message)
{
    fprintf(stderr, "%s", message);
}
```

设置通知接收器或处理器后，应当预期：只要 PGconn 对象或由它生成的 PGresult 对象仍存在，就可能调用该函数。创建 PGresult 时，会将 PGconn 当前的通知处理指针复制到 PGresult 中，供 PQgetvalue 等函数在需要时使用。

31.12. 事件系统

libpq的事件系统被设计为通知已注册的事件处理器它感兴趣的libpq事件，例如PGconn
以及PGresult对象的创建和销毁。一个主要用途是允许应用将自己的数据与一个PGconn
或者PGresult关联在一起，并且确保那些数据在适当的时候被释放。

每个注册的事件处理程序都与两项数据相关联，libpq仅将其视为不透明的void *指针。有一个透传指针，由应用程序在向 PGconn 注册事件处理程序时提供。透传指针在PGconn及其生成的所有PGresult的生命周期内永远不会更改；因此，如果使用，它必须指向长期存在的数据。
此外，还有一个实例数据指针，在每个PGconn和PGresult中一开始都是NULL。
可以使用PQinstanceData、PQsetInstanceData、PQresultInstanceData
和PQresultSetInstanceData函数来操作此指针。请注意，与透传指针不同，PGconn的实例数据不会自动继承到从中创建的PGresult。
libpq不知道透传和实例数据指针指向的内容（如果有的话），也永远不会尝试释放它们 – 这是事件处理程序的责任。

31.12.1. 事件类型

枚举PGEventId命名了事件系统处理的事件类型。它的所有值的名称都以PGEVT开始。
对于每一种事件类型，都有一个相应的事件信息结构体用来承载传递给事件处理器的参数。
事件类型是：

PGEVT_REGISTER

注册事件在PQregisterEventProc被调用时触发。
此时最适合初始化事件过程可能需要的instanceData。
每个连接中的每个事件过程只会触发一次注册事件。如果事件过程失败，则中止注册。

typedef struct

```
{
    PGconn *conn;
} PGEventRegister;
```

收到PGEVT_REGISTER事件时，应将*evtInfo*指针强制转换为PGEventRegister *。此结构体包含一个PGconn，它应处于CONNECTION_OK状态；如果调用PQregisterEventProc紧接在取得正常的PGconn之后，就能保证这一点。返回失败代码时，必须完成全部清理工作，因为不会发送PGEVT_CONNDESTROY事件。

PGEVT_CONNRESET

连接重置事件会在完成以下调用时触发：PQreset或PQresetPoll。在这两种情况下，只有重置成功才会触发该事件。如果事件过程失败，整个连接重置就会失败；PGconn会被置于CONNECTION_BAD状态，并且PQresetPoll将返回PGRES_POLLING_FAILED。

```
typedef struct
{
    PGconn *conn;
} PGEventConnReset;
```

收到PGEVT_CONNRESET事件时，应将*evtInfo*指针强制转换为PGEventConnReset *。虽然其中的PGconn刚刚被重置，但所有事件数据都保持不变。应利用此事件重置、重新加载或重新查询相关联的instanceData。注意，即使事件过程未能处理PGEVT_CONNRESET，它仍会在连接关闭时收到PGEVT_CONNDESTROY事件。

PGEVT_CONNDESTROY

连接销毁事件由以下调用触发：PQfinish。事件过程负责正确清理其事件数据，因为 libpq 无法管理这部分内存。如果不清理，就会造成内存泄漏。

```
typedef struct
{
    PGconn *conn;
} PGEventConnDestroy;
```

收到PGEVT_CONNDESTROY事件时，应将*evtInfo*指针强制转换为PGEventConnDestroy *。该事件触发于以下函数执行任何其他清理工作之前：PQfinish。事件过程的返回值会被忽略，因为无法通过以下函数报告失败：PQfinish。此外，事件过程失败不应中止清理不再使用的内存的过程。

PGEVT_RESULTCREATE

任何生成结果的查询执行函数都会触发结果创建事件，其中包括PQgetResult。只有成功创建结果后才会触发该事件。

```
typedef struct
{
    PGconn *conn;
    PGresult *result;
} PGEventResultCreate;
```

收到PGEVT_RESULTCREATE事件时，应将*evtInfo*指针转换为PGEventResultCreate *。其中，*conn*是用于生成结果的连接。这是初始化需要与结果关联的instanceData的理想位置。如果事件过程失败，结果会被清除，失败也会向外传递。事件过程不得自行调用PQclear来清除结果对象。返回失败代码时，必须完成所有清理工作，因为不会发送PGEVT_RESULTDESTROY事件。

PGEVT_RESULTCOPY

结果复制事件会在调用PQcopyResult时触发。只有复制完成后才会触发该事件。只有为源结果成功处理过PGEVT_RESULTCREATE或PGEVT_RESULTCOPY事件的事件过程，才会收到PGEVT_RESULTCOPY事件。

```
typedef struct
{
    const PGresult *src;
    PGresult *dest;
} PGEventResultCopy;
```

收到PGEVT_RESULTCOPY事件时，应将*evtInfo*指针强制转换为PGEventResultCopy *。其中，*src*结果是复制源，而*dest*结果是复制目标。可以利用此事件对instanceData进行深复制，因为PQcopyResult无法完成这项工作。如果事件过程失败，整个复制操作就会失败，并且*dest*结果将被清除。返回失败代码时，必须完成所有清理工作，因为不会为目标结果发送PGEVT_RESULTDESTROY事件。

PGEVT_RESULTDESTROY

结果销毁事件由以下调用触发：PQclear。事件过程负责正确清理其事件数据，因为 libpq 无法管理这部分内存。如果不清理，就会造成内存泄漏。

```
typedef struct
{
    PGresult *result;
} PGEventResultDestroy;
```

收到PGEVT_RESULTDESTROY事件时，应将*evtInfo*指针强制转换为PGEventResultDestroy *。该事件触发于以下函数执行任何其他清理工作之前：PQclear。事件过程的返回值会被忽略，因为无法通过以下函数报告失败：PQclear。此外，事件过程失败不应中止清理不再使用的内存的过程。

31.12.2. 事件回调函数

PGEventProc

PGEventProc 是通过 typedef 定义的事件过程指针类型，也就是接收 libpq 事件的用户回调函数的指针类型。事件过程的签名必须为：

```
int eventproc(PGEventId evtId, void *evtInfo, void *passThrough)
```

evtId 参数指示发生了哪一种 PGEVT 事件。必须将 *evtInfo* 指针强制转换为适当的结构体类型，以获取关于该事件的更多信息。*passThrough* 参数是在注册事件过程时传给 PQregisterEventProc 的指针。函数应在成功时返回非零值，在失败时返回零。

在任何一个PGconn中，一个特定事件过程只能被注册一次。这是因为该过程的地址被用作查找键来标识相关的实例数据。

> **小心**
>
> 在 Windows 上，函数可能有两个不同的地址：一个在 DLL 外部可见，另一个在 DLL 内部可见。使用 libpq 的事件过程函数时，务必始终使用其中同一个地址，否则会产

> 生混淆。确保代码正常工作的最简单做法，是将事件过程声明为 `static`。如果需要在过程所在的源文件之外取得其地址，应提供一个单独的函数来返回该地址。

31.12.3. 事件支持函数

`PQregisterEventProc`

为 libpq 注册一个事件回调过程。

```
int PQregisterEventProc(PGconn *conn, PGEventProc proc,
                        const char *name, void *passThrough);
```

对于希望接收其事件的每个 `PGconn`，都必须注册一次事件过程。一个连接可注册的事件过程数量只受内存限制。函数成功时返回非零值，失败时返回零。

当一个 libpq 事件被触发时，*proc*参数将被调用。它的内存地址也被用来查找`instanceData`。*name*参数被用来在错误消息中引用该事件过程。这个值不能是`NULL`或一个零长度串。名字串被复制到`PGconn`中，因此传递进来的东西不需要长期存在。当一个事件发生时，*passThrough*指针被传递给*proc*。这个参数可以是`NULL`。

`PQsetInstanceData`

设置连接*conn*的用于过程*proc*的`instanceData`为*data*。它在成功时返回非零值，失败时返回零（只有*proc*没有被正确地注册在*conn*中，才可能会失败）。

```
int PQsetInstanceData(PGconn *conn, PGEventProc proc, void *data);
```

`PQinstanceData`

返回连接*conn*的与过程*proc*相关的`instanceData`，如果没有则返回`NULL`。

```
void *PQinstanceData(const PGconn *conn, PGEventProc proc);
```

`PQresultSetInstanceData`

将结果中针对 *proc* 的 `instanceData` 设置为 *data*。成功时返回非零值，失败时返回零。（只有当 *proc* 未在结果中正确注册时，才可能失败。）

```
int PQresultSetInstanceData(PGresult *res, PGEventProc proc, void
 *data);
```

`PQresultInstanceData`

返回结果的与过程*proc*相关的`instanceData`，如果没有则返回`NULL`。

```
void *PQresultInstanceData(const PGresult *res, PGEventProc proc);
```

31.12.4. 事件实例

下面给出一个示例框架，用于管理与 libpq 连接和结果关联的私有数据。

```
/* required header for libpq events (note: includes libpq-fe.h) */
#include <libpq-events.h>
```

```
/* The instanceData */
typedef struct
{
    int n;
    char *str;
} mydata;

/* PGEventProc */
static int myEventProc(PGEventId evtId, void *evtInfo, void *pass
Through);

int
main(void)
{
    mydata *data;
    PGresult *res;
    PGconn *conn = PQconnectdb("dbname = postgres");

    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        PQfinish(conn);
        return 1;
    }

    /* called once on any connection that should receive events.
     * Sends a PGEVT_REGISTER to myEventProc.
     */
    if (!PQregisterEventProc(conn, myEventProc, "mydata_proc", NULL))
    {
        fprintf(stderr, "Cannot register PGEventProc\n");
        PQfinish(conn);
        return 1;
    }

    /* conn instanceData is available */
    data = PQinstanceData(conn, myEventProc);

    /* Sends a PGEVT_RESULTCREATE to myEventProc */
    res = PQexec(conn, "SELECT 1 + 1");

    /* result instanceData is available */
    data = PQresultInstanceData(res, myEventProc);

    /* If PG_COPYRES_EVENTS is used, sends a PGEVT_RESULTCOPY to my
EventProc */
    res_copy = PQcopyResult(res, PG_COPYRES_TUPLES | PG_COPYRES_
EVENTS);

    /* result instanceData is available if PG_COPYRES_EVENTS was
     * used during the PQcopyResult call.
     */
    data = PQresultInstanceData(res_copy, myEventProc);
```

```
    /* Both clears send a PGEVT_RESULTDESTROY to myEventProc */
    PQclear(res);
    PQclear(res_copy);

    /* Sends a PGEVT_CONNDESTROY to myEventProc */
    PQfinish(conn);

    return 0;
}

static int
myEventProc(PGEventId evtId, void *evtInfo, void *passThrough)
{
    switch (evtId)
    {
        case PGEVT_REGISTER:
        {
            PGEventRegister *e = (PGEventRegister *)evtInfo;
            mydata *data = get_mydata(e->conn);

            /* associate app specific data with connection */
            PQsetInstanceData(e->conn, myEventProc, data);
            break;
        }

        case PGEVT_CONNRESET:
        {
            PGEventConnReset *e = (PGEventConnReset *)evtInfo;
            mydata *data = PQinstanceData(e->conn, myEventProc);

            if (data)
              memset(data, 0, sizeof(mydata));
            break;
        }

        case PGEVT_CONNDESTROY:
        {
            PGEventConnDestroy *e = (PGEventConnDestroy *)evtInfo;
            mydata *data = PQinstanceData(e->conn, myEventProc);

            /* free instance data because the conn is being destroyed
 */
            if (data)
              free_mydata(data);
            break;
        }

        case PGEVT_RESULTCREATE:
        {
            PGEventResultCreate *e = (PGEventResultCreate *)evtInfo;
            mydata *conn_data = PQinstanceData(e->conn, myEventProc);
            mydata *res_data = dup_mydata(conn_data);
```

```
            /* associate app specific data with result (copy it from
 conn) */
            PQsetResultInstanceData(e->result, myEventProc, res_data);
            break;
        }

        case PGEVT_RESULTCOPY:
        {
            PGEventResultCopy *e = (PGEventResultCopy *)evtInfo;
            mydata *src_data = PQresultInstanceData(e->src, myEvent
Proc);
            mydata *dest_data = dup_mydata(src_data);

            /* associate app specific data with result (copy it from a
 result) */
            PQsetResultInstanceData(e->dest, myEventProc, dest_data);
            break;
        }

        case PGEVT_RESULTDESTROY:
        {
            PGEventResultDestroy *e = (PGEventResultDestroy *)evtInfo;
            mydata *data = PQresultInstanceData(e->result, myEvent
Proc);

            /* free instance data because the result is being
 destroyed */
            if (data)
              free_mydata(data);
            break;
        }

        /* unknown event ID, just return TRUE. */
        default:
            break;
    }

    return TRUE; /* event processing succeeded */
}
```

31.13. 环境变量

以下环境变量可用于选择连接参数的默认值，供 `PQconnectdb`、`PQsetdbLogin` 和 `PQsetdb` 在调用代码未直接指定参数值时使用。例如，这样可以避免在简单的客户端应用程序中硬编码数据库连接信息。

- `PGHOST`的行为与host连接参数相同。
- `PGHOSTADDR`的行为与hostaddr连接参数相同。这可以替代或者与`PGHOST`一起设置，以避免DNS查找开销。
- `PGPORT`的行为与port连接参数相同。
- `PGDATABASE`的行为与dbname连接参数相同。

- PGUSER的行为与user连接参数相同。
- PGPASSWORD的行为与password连接参数相同。出于安全原因，不建议使用此环境变量，因为一些操作系统允许非root用户通过ps查看进程环境变量；而应考虑使用~/.pgpass文件（参见第 31.14 节）。
- PGPASSFILE指定用于查找的密码文件名。如果未设置，默认为~/.pgpass（参见第 31.14 节）。
- PGSERVICE的行为与service连接参数相同。
- PGSERVICEFILE指定每个用户的连接服务文件的名称。如果未设置，默认为~/.pg_service.conf（参见第 31.15 节）。
- PGREALM设置与PostgreSQL一起使用的 Kerberos 域（如果它与本地域不同）。如果设置了PGREALM，libpq应用将尝试与该域中的服务器进行认证，并使用单独的票据文件以避免与本地票据文件冲突。只有当服务器选择了 Kerberos 认证时，才会使用这个环境变量。
- PGOPTIONS的行为与options连接参数相同。
- PGAPPNAME的行为与application_name连接参数相同。
- PGSSLMODE的行为与sslmode连接参数相同。
- PGREQUIRESSL的行为与requiressl连接参数相同。
- PGSSLCERT的行为与sslcert连接参数相同。
- PGSSLKEY的行为与sslkey连接参数相同。
- PGSSLROOTCERT的行为与sslrootcert连接参数相同。
- PGSSLCRL的行为与sslcrl连接参数相同。
- PGREQUIREPEER的行为与requirepeer连接参数相同。
- PGKRBSRVNAME的行为与krbsrvname连接参数相同。
- PGGSSLIB的行为与gsslib连接参数相同。
- PGCONNECT_TIMEOUT的行为与connect_timeout连接参数相同。
- PGCLIENTENCODING的行为与client_encoding连接参数相同。

以下环境变量可用于指定每个 PostgreSQL 会话的默认行为。（也可参见ALTER ROLE和ALTER DATABASE命令，了解按用户或按数据库设置默认行为的方法。）

- PGDATESTYLE设置日期/时间表示的默认风格（等同于SET datestyle TO ...）。
- PGTZ设置默认的时区（等同于SET timezone TO ...）。
- PGGEQO为遗传查询优化器设置默认模式（等同于SET geqo TO ...）。

有关这些环境变量的正确取值，请参见SQL命令SET。

下面的环境变量决定libpq的内部行为，它们会覆盖编译在程序中的默认值。

- PGSYSCONFDIR设置包含pg_service.conf文件以及未来版本中可能出现的其他系统范围配置文件的目录。
- PGLOCALEDIR设置包含用于消息国际化的locale文件的目录。

31.14. 密码文件

用户主目录中的.pgpass文件或PGPASSFILE引用的文件可保存密码，
供连接需要密码且尚未通过其他方式指定密码时使用。在 Microsoft Windows 上，文件名为 %APPDATA%\postgresql\pgpass.conf（其中%APPDATA%
指用户配置文件中的应用数据子目录）。

该文件中的行应采用以下格式：

hostname:*port*:*database*:*username*:*password*

（可以复制上面这一行，并在行首加上#，在文件中加入提示注释。）
前四个字段中的每一个都可以是字面值，或者是*，后者可以匹配任何内容。
将使用与当前连接参数匹配的第一行中的密码字段。（因此，使用通配符时，
应将更具体的条目放在前面。）如果条目需要包含:或\，请使用\转义该字符。
主机名localhost同时匹配来自本地机器的 TCP 连接（主机名localhost）和 Unix 域套接字连接（pghost为空或为默认套接字目录）。在备库中，数据库字段为replication时，
匹配连接到主库的流复制连接。除此之外，database字段的用途有限，
因为同一用户在同一数据库集簇的所有数据库中使用相同的密码。

在 Unix 系统上，.pgpass文件上的权限必须不允许所有人或组内访问，可以用chmod 0600 ~/.pgpass这样的命令实现。如果权限没有这么严格，该文件将被忽略。在微软 Windows 上，该文件被假定存储在一个安全的目录中，因此不会进行特别的权限检查。

31.15. 连接服务文件

连接服务文件允许 libpq 连接参数与一个单一服务名称关联。之后在一个 libpq 连接中就可以指定那个服务名称，与其相关的设置将被使用。这允许在不重新编译 libpq 应用的前提下修改连接参数。服务名称也可以使用PGSERVICE环境变量来指定。

连接服务文件可以是位于~/.pg_service.conf或由环境变量PGSERVICEFILE
指定的位置处的每用户服务文件，也可以是位于`pg_config --sysconfdir`/pg_service.conf或由环境变量 PGSYSCONFDIR指定的目录中的系统范围文件。如果在用户文件和系统文件中存在同名的服务定义，则用户文件优先。

该文件使用一种“INI 文件”格式，其中小节名是服务名并且参数是连接参数。列表见第 31.1 节。例如：

```
# comment
[mydb]
host=somehost
port=5433
user=admin
```

An example file is provided at share/pg_service.conf.sample.

31.16. 连接参数的 LDAP 查找

如果libpq已经在编译时打开了 LDAP 支持（configure的选项--with-ldap），就可以通过 LDAP 从一个中央服务器检索host或dbname之类的连接参数。这样做的好处是如果一个数据库的连接参数改变，不需要在所有的客户端机器上更新连接信息。

LDAP 连接参数查询使用连接服务文件pg_service.conf（参见第 31.15 节）。在pg_service.conf的配置段中，以ldap://开头的行会被识别为 LDAP URL，并执行 LDAP 查询。结果必须是一个keyword = value键值对列表，用于设置连接选项。URL 必须符合 RFC 1959，格式如下：

ldap://[*hostname*[:*port*]]/*search_base*?*attribute*?*search_scope*?*filter*

其中，*hostname*默认为localhost而*port*默认为 389。

LDAP 查找成功后就会停止处理 pg_service.conf；如果无法联系 LDAP 服务器，则会继续处理。这使后续指向其他 LDAP 服务器的 LDAP URL 行、常规的 keyword = value 对或默认连接选项能够作为后备。如果希望在这种情况下得到错误消息，可以在 LDAP URL 后添加一个语法不正确的行。

例如，使用以下 LDIF 文件创建的 LDAP 条目：

```
version:1
dn:cn=mydatabase,dc=mycompany,dc=com
changetype:add
objectclass:top
objectclass:groupOfUniqueNames
cn:mydatabase
uniqueMember:host=dbserver.mycompany.com
uniqueMember:port=5439
uniqueMember:dbname=mydb
uniqueMember:user=mydb_user
uniqueMember:sslmode=require
```

可以通过以下 LDAP URL 查询：

```
ldap://ldap.mycompany.com/dc=mycompany,dc=com?uniqueMember?one?(cn=
mydatabase)
```

你也可以将常规的服务文件条目和 LDAP 查找混合。pg_service.conf中一节的完整示例：

```
# 仅 host 和 port 存储在 LDAP 中；显式指定 dbname 和 user
[customerdb]
dbname=customer
user=appuser
ldap://ldap.acme.com/cn=dbserver,cn=hosts?pgconnectinfo?base?(
objectclass=*)
```

31.17. SSL 支持

PostgreSQL 原生支持使用 SSL 连接来加密客户端与服务器之间的通信，以提高安全性。有关服务器端 SSL 功能的详细信息，请参见第 17.9 节。

libpq读取系统范围的OpenSSL配置文件。默认情况下，这个文件被命名为openssl.cnf并且位于openssl version -d所报告的目录中。可以通过设置环境变量OPENSSL_CONF把这个默认值覆盖为想要的配置文件的名称。

31.17.1. 服务器证书的客户端验证

默认情况下，PostgreSQL不会对服务器证书执行任何验证。
这意味着可以在客户端不知情的情况下伪造服务器身份，例如修改 DNS 记录或接管服务器的 IP 地址。为了防止伪造，必须使用SSL证书验证。

如果参数sslmode设为verify-ca，libpq 会沿证书链检查到一个受信任的证书机构（CA），以验证服务器是否可信。如果sslmode设为verify-full，libpq 还会验证服务器主机名是否与其证书匹配。如果无法验证服务器证书，SSL 连接将失败。在大多数对安全敏感的环境中，建议使用verify-full。

在 verify-full 模式下，会将主机名与证书的 cn（通用名称）属性匹配。如果 cn 属性以星号（*）开头，该星号会被视为通配符，匹配除点（.）以外的所有字符。这意味着该证书不会匹配子域。如果使用 IP 地址而不是主机名建立连接，则会匹配该 IP 地址（不执行任何 DNS 查询）。

要允许服务器证书验证，必须将一个或多个受信任的CA的证书放置在用户主目录下的~/.postgresql/root.crt文件中（在 Microsoft Windows 上该文件名为%APPDATA%\postgresql\root.crt）。

如果文件~/.postgresql/root.crl存在（微软 Windows 上的%APPDATA%\postgresql\root.crl），证书撤销列表（CRL）项也会被检查。

可以通过设置连接参数 sslrootcert 和 sslcrl，或环境变量 PGSSLROOTCERT 和 PGSSLCRL，更改根证书文件与 CRL 的位置。

> **注意**
>
> 为与 PostgreSQL 的早期版本向后兼容，如果存在根 CA 文件，sslmode=require 的行为将与 verify-ca 相同，即根据 CA 验证服务器证书。不建议依赖这种行为；需要证书验证的应用程序应始终使用 verify-ca 或 verify-full。

31.17.2. 客户端证书

如果服务器请求一个受信任的客户端证书，libpq将发送存储在用户主目录下的文件~/.postgresql/postgresql.crt中的证书。该证书必须由服务器信任的证书机构（CA）之一签发。匹配的私钥文件~/.postgresql/postgresql.key也必须存在。
私钥文件的权限必须不允许任何对世界或组的访问；可以通过命令chmod 0600 ~/.postgresql/postgresql.key实现。在 Microsoft Windows 上，这些文件的名称分别为%APPDATA%\postgresql\postgresql.crt和%APPDATA%\postgresql\postgresql.key，并且不进行特别的权限检查，因为该目录被假定为安全的。
证书和密钥文件的位置可以通过连接参数sslcert和sslkey或环境变量PGSSLCERT和PGSSLKEY来覆盖。

在某些情况下，客户端证书可能由一个“中间”证书机构签发，
而不是由服务器直接信任的机构签发。要使用这样的证书，把签发机构的证书追加到postgresql.crt文件中，然后是其父机构的证书，依此类推直到一个服务器信任的“根”机构。在postgresql.crt包含多个证书的所有情况中，都应包含根证书。

注意，root.crt列出的是被认为可以信任其为服务器证书签名的顶层 CA。原则上它不必列出为客户端证书签名的 CA，尽管在大多数情况下该 CA 也会被信任用于服务器证书。

31.17.3. 不同模式中提供的保护

`sslmode` 参数的不同值提供不同级别的保护。SSL 可以防范三类攻击：

窃听

如果一个第三方能够检查客户端和服务器之间的网络流量，
它能读取连接信息（包括用户名和密码）以及被传递的数据。SSL使用加密来阻止这种攻击。

中间人（MITM）

如果第三方能修改客户端与服务器之间传输的数据，就可以冒充服务器，
进而查看和修改数据，即使数据已经加密。随后，
第三方可以将连接信息和数据转发给原来的服务器，使攻击无法被察觉。常见的手段包括 DNS 污染和地址劫持，从而将客户端引向预期之外的服务器。还有其他几种攻击手段可以达到同样的目的。SSL 使用证书验证，让客户端认证服务器身份，以防范这种攻击。

模仿

如果第三方能冒充获授权的客户端，就能直接访问其无权访问的数据。
这通常可能由不安全的密码管理导致。SSL 使用客户端证书，确保只有持有有效证书的客户端才能访问服务器，以防范这种攻击。

要确保连接安全，必须在建立连接之前，在客户端和服务器两端配置 SSL。如果仅在服务器上配置，客户端可能在得知服务器要求高安全性之前就已发送敏感信息（例如密码）。在 libpq 中，可以将 `sslmode` 参数设为 `verify-full` 或 `verify-ca`，并向系统提供用于验证的根证书，以确保连接安全。这类似于使用 `https` URL 进行加密的网页浏览。

服务器通过身份认证后，客户端便可以传送敏感数据。这意味着，在此之前，
客户端无需知道是否会使用证书进行认证，因此可以安全地仅在服务器配置中指定这一点。

所有 SSL 选项都会产生加密和密钥交换的开销，因此必须在性能与安全性之间作出权衡。表 31.1说明了不同 `sslmode` 值所能防范的风险，以及它们所表达的对安全性和开销的取舍。

表 31.1. SSL 模式描述

sslmode	**窃听保护**	**MITM protection**	**Statement**
`disable`	No	No	我不关心安全性，并且我不想为加密增加负荷。
`allow`	Maybe	No	我不关心安全性，但如果服务器坚持，我将承担加密带来的负荷。
`prefer`	Maybe	No	我不关心安全性，但如果服务器支持，我希望承担加密带来的负荷。
`require`	Yes	No	我想要对数据加密，并且我接受因此带来的负荷。我信任该网络会保证我总是连接到想要连接的服务器。
`verify-ca`	Yes	`Depends on CA`-policy	我想要对数据加密，并且我接受因此带来的负荷。我想要确保我连接到的是我信任的服务器。

sslmode	窃听保护	MITM protection	Statement
verify-full	Yes	Yes	我想要对数据加密，并且我接受因此带来的负荷。我想要确保我连接到的是我信任的服务器，并且就是我指定的那一个。

verify-ca和verify-full之间的区别取决于根CA的策略。如果使用了一个公共CA，verify-ca允许连接到那些可能已经被其他人注册到该CA的服务器。在这种情况下，总是应该使用verify-full。如果使用了一个本地CA或者甚至是一个自签名的证书，使用verify-ca常常就可以提供足够的保护。

sslmode 的默认值是 prefer。如表所示，从安全角度看，这一设置没有意义；它只会在可能时带来性能开销。将其作为默认值仅出于向后兼容的考虑，不建议在有安全要求的部署中使用。

31.17.4. SSL 客户端文件使用

表 31.2总结了与客户端 SSL 设置相关的文件。

表 31.2. libpq/客户端 SSL 文件用法

文件	内容	作用
~/.postgresql/postgresql.crt	客户端证书	被服务器请求
~/.postgresql/postgresql.key	客户端私钥	证明客户端证书由其所有者发送；并不表明证书所有者可信
~/.postgresql/root.crt	受信任的证书机构	检查服务器证书是否由受信任的证书机构签发
~/.postgresql/root.crl	被证书机构撤销的证书	服务器证书不能出现在这个列表中

31.17.5. SSL 库初始化

如果你的应用程序初始化libssl和/或libcrypto库，并且libpq 构建时带有SSL支持，你应该调用PQinitOpenSSL告诉libpq libssl和/或libcrypto库已被你的应用程序初始化，以便libpq不会再初始化这些库。SSL API 的详细信息见http://h71000.www7.hp.com/doc/83final/ba554_90007/ch04.html。

PQinitOpenSSL

允许应用程序选择要初始化的安全库。

```
void PQinitOpenSSL(int do_ssl, int do_crypto);
```

当*do_ssl*是非零时，libpq将在第一次打开数据库连接前初始化OpenSSL库。当*do_crypto*是非零时，libcrypto库将被初始化。默认情况下（如果没有调用PQinitOpenSSL），两个库都会被初始化。当 SSL 支持没有被编译时，这个函数也存在但是什么也不做。

如果你的应用使用并且初始化OpenSSL或者它的底层libcrypto库，你必须在第一次打开数据库连接前调用这个函数，并把相应参数设为零。同时要确保在打开一个数据库连接前已经完成了初始化。

PQinitSSL

允许应用程序选择要初始化的安全库。

```
void PQinitSSL(int do_ssl);
```

这个函数等效于PQinitOpenSSL(do_ssl, do_ssl)。这对于要么初始化OpenSSL以及libcrypto要么都不初始化的应用足够用了。

PQinitSSL从PostgreSQL 8.0 就存在了，而PQinitOpenSSL直到PostgreSQL 8.4 才被加入，因此PQinitSSL可能对那些需要与旧版本libpq一起工作的应用来说更合适。

31.18. 在线程化程序中的行为

libpq 默认是可重入且线程安全的。编译应用程序代码时，可能需要使用特殊的编译器命令行选项。有关如何构建支持线程的应用程序，请参阅系统文档，或查看 src/Makefile.global 中的 PTHREAD_CFLAGS 和 PTHREAD_LIBS。以下函数可用于查询 libpq 的线程安全状态：

PQisthreadsafe

返回libpq库的线程安全状态。

```
int PQisthreadsafe();
```

如果libpq是线程安全的则返回 1，否则返回 0。

线程使用的一项限制是：两个线程不能同时操作同一个 PGconn 对象。尤其不能通过同一个连接对象从不同线程并发发出命令。（如果需要并发执行命令，请使用多个连接。）

PGresult对象在创建后通常是只读的，因此可以在线程之间自由传递。不过，如果你使用第 31.10 节或第 31.12 节中描述的任何会修改PGresult的函数，则需要自行避免对同一个PGresult执行并发操作。

已弃用的函数 PQrequestCancel 和 PQoidStatus 不是线程安全的，不应在多线程程序中使用。PQrequestCancel 可以用 PQcancel 替代；PQoidStatus 可以用 PQoidValue 替代。

如果你在应用程序中使用 Kerberos（除了libpq内部之外），则需要在 Kerberos 调用周围加锁，因为 Kerberos 函数不是线程安全的。可参考libpq源代码中的PQregisterThreadLock函数，它提供了一种在libpq与应用程序之间协作加锁的方法。

如果你在多线程应用程序中遇到问题，请运行 src/tools/thread 目录中的程序，检查你的平台是否存在线程不安全的函数。configure 会运行这个程序，但对于二进制发行版，你自己的库可能与编译这些二进制文件时所用的库不一致。

31.19. 构建libpq程序

要构建（即编译和链接）一个使用 libpq 的程序，需要完成以下所有步骤：

- 包含 libpq-fe.h 头文件：

  ```
  #include <libpq-fe.h>
  ```

  如果没有这样做，编译器通常会给出类似以下内容的错误消息：

  ```
  foo.c: In function `main':
  foo.c:34: `PGconn' undeclared (first use in this function)
  foo.c:35: `PGresult' undeclared (first use in this function)
  foo.c:54: `CONNECTION_BAD' undeclared (first use in this function)
  foo.c:68: `PGRES_COMMAND_OK' undeclared (first use in this function)
  ```

```
foo.c:95: `PGRES_TUPLES_OK' undeclared (first use in this function)
```

- 向编译器提供 `-I`*`directory`* 选项，指定 PostgreSQL 头文件的安装目录。（某些情况下，编译器默认会搜索该目录，此时可以省略这个选项。）例如，编译命令行可以是：

  ```
  cc -c -I/usr/local/pgsql/include testprog.c
  ```

  如果使用 makefile，请将该选项添加到 `CPPFLAGS` 变量中：

  ```
  CPPFLAGS += -I/usr/local/pgsql/include
  ```

  如果你的程序可能由其他用户编译，那么你不应该像那样硬编码目录位置。你可以运行工具`pg_config`在本地系统上找出头文件在哪里：

  ```
  $ pg_config --includedir
  /usr/local/include
  ```

  未向编译器指定正确的选项将导致类似以下内容的错误消息：

  ```
  testlibpq.c:8:22: libpq-fe.h: No such file or directory
  ```

- 链接最终程序时，指定 `-lpq` 选项以链接 libpq 库，同时指定 `-L`*`directory`* 选项，告知编译器 libpq 库所在的目录。（同样，编译器默认会搜索某些目录。）为获得最大的可移植性，请将 `-L` 选项放在 `-lpq` 选项之前。例如：

  ```
  cc -o testprog testprog1.o testprog2.o -L/usr/local/pgsql/lib -lpq
  ```

  你也可以使用`pg_config`找出库目录：

  ```
  $ pg_config --libdir
  /usr/local/pgsql/lib
  ```

  指出这一部分问题的错误消息可能看起来像：

  ```
  testlibpq.o: In function `main':
  testlibpq.o(.text+0x60): undefined reference to `PQsetdbLogin'
  testlibpq.o(.text+0x71): undefined reference to `PQstatus'
  testlibpq.o(.text+0xa4): undefined reference to `PQerrorMessage'
  ```

  这意味着你忘记了 `-lpq`。

  ```
  /usr/bin/ld: cannot find -lpq
  ```

  这意味着你忘记了`-L`选项或者没有指定正确的目录。

31.20. 示例程序

这些示例和其他示例可在源代码发行包的`src/test/examples`目录中找到。

例 31.1. libpq示例程序 1

```
/*
 * testlibpq.c
 *
 *      Test the C version of libpq, the PostgreSQL frontend library.
 */
```

```
#include <stdio.h>
#include <stdlib.h>
#include <libpq-fe.h>

static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}

int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    int         nFields;
    int         i,
                j;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * Our test case here involves using a cursor, for which we must
 be inside
     * a transaction block.  We could do the whole thing with a single
     * PQexec() of "select * from pg_database", but that's too trivial
 to make
     * a good example.
     */
```

```
    /* Start a transaction block */
    res = PQexec(conn, "BEGIN");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "BEGIN command failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /*
     * Should PQclear PGresult whenever it is no longer needed to
 avoid memory
     * leaks
     */
    PQclear(res);

    /*
     * Fetch rows from pg_database, the system catalog of databases
     */
    res = PQexec(conn, "DECLARE myportal CURSOR FOR select * from pg_
database");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "DECLARE CURSOR failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }
    PQclear(res);

    res = PQexec(conn, "FETCH ALL in myportal");
    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
        fprintf(stderr, "FETCH ALL failed: %s", PQerrorMessage(conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /* first, print out the attribute names */
    nFields = PQnfields(res);
    for (i = 0; i < nFields; i++)
        printf("%-15s", PQfname(res, i));
    printf("\n\n");

    /* next, print out the rows */
    for (i = 0; i < PQntuples(res); i++)
    {
        for (j = 0; j < nFields; j++)
            printf("%-15s", PQgetvalue(res, i, j));
        printf("\n");
    }
```

```
    PQclear(res);

    /* close the portal ... we don't bother to check for errors ... */
    res = PQexec(conn, "CLOSE myportal");
    PQclear(res);

    /* end the transaction */
    res = PQexec(conn, "END");
    PQclear(res);

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

例 31.2. libpq示例程序 2

```
/*
 * testlibpq2.c
 *      Test of the asynchronous notification interface
 *
 * Start this program, then from psql in another window do
 *   NOTIFY TBL2;
 * Repeat four times to get this program to exit.
 *
 * Or, if you want to get fancy, try this:
 * populate a database with the following commands
 * (provided in src/test/examples/testlibpq2.sql):
 *
 *   CREATE TABLE TBL1 (i int4);
 *
 *   CREATE TABLE TBL2 (i int4);
 *
 *   CREATE RULE r1 AS ON INSERT TO TBL1 DO
 *     (INSERT INTO TBL2 VALUES (new.i); NOTIFY TBL2);
 *
 * and do this four times:
 *
 *   INSERT INTO TBL1 VALUES (10);
 */
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <errno.h>
#include <sys/time.h>
#include <libpq-fe.h>

static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}
```

```
int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    PGnotify   *notify;
    int         nnotifies;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * Issue LISTEN command to enable notifications from the rule's
 NOTIFY.
     */
    res = PQexec(conn, "LISTEN TBL2");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "LISTEN command failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /*
     * should PQclear PGresult whenever it is no longer needed to
 avoid memory
     * leaks
     */
    PQclear(res);
```

```
    /* Quit after four notifies are received. */
    nnotifies = 0;
    while (nnotifies < 4)
    {
        /*
         * Sleep until something happens on the connection.  We use
 select(2)
         * to wait for input, but you could also use poll() or similar
         * facilities.
         */
        int         sock;
        fd_set      input_mask;

        sock = PQsocket(conn);

        if (sock < 0)
            break;              /* shouldn't happen */

        FD_ZERO(&input_mask);
        FD_SET(sock, &input_mask);

        if (select(sock + 1, &input_mask, NULL, NULL, NULL) < 0)
        {
            fprintf(stderr, "select() failed: %s\n", strerror(errno));
            exit_nicely(conn);
        }

        /* Now check for input */
        PQconsumeInput(conn);
        while ((notify = PQnotifies(conn)) != NULL)
        {
            fprintf(stderr,
                    "ASYNC NOTIFY of '%s' received from backend PID %d
\n",
                    notify->relname, notify->be_pid);
            PQfreemem(notify);
            nnotifies++;
        }
    }

    fprintf(stderr, "Done.\n");

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

例 31.3. libpq示例程序 3

```
/*
 * testlibpq3.c
 *      Test out-of-line parameters and binary I/O.
```

```
 *
 * Before running this, populate a database with the following
 commands
 * (provided in src/test/examples/testlibpq3.sql):
 *
 * CREATE TABLE test1 (i int4, t text, b bytea);
 *
 * INSERT INTO test1 values (1, 'joe''s place', '\\000\\001\\002\\003\
\004');
 * INSERT INTO test1 values (2, 'ho there', '\\004\\003\\002\\001\
\000');
 *
 * The expected output is:
 *
 * tuple 0: got
 *  i = (4 bytes) 1
 *  t = (11 bytes) 'joe's place'
 *  b = (5 bytes) \000\001\002\003\004
 *
 * tuple 0: got
 *  i = (4 bytes) 2
 *  t = (8 bytes) 'ho there'
 *  b = (5 bytes) \004\003\002\001\000
 */
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <sys/types.h>
#include <libpq-fe.h>

/* for ntohl/htonl */
#include <netinet/in.h>
#include <arpa/inet.h>


static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}

/*
 * This function prints a query result that is a binary-format fetch
 from
 * a table defined as in the comment above.  We split it out because
 the
 * main() function uses it twice.
 */
static void
show_binary_results(PGresult *res)
{
    int         i,
                j;
```

```
   int         i_fnum,
               t_fnum,
               b_fnum;

   /* Use PQfnumber to avoid assumptions about field order in result
*/
   i_fnum = PQfnumber(res, "i");
   t_fnum = PQfnumber(res, "t");
   b_fnum = PQfnumber(res, "b");

   for (i = 0; i < PQntuples(res); i++)
   {
       char       *iptr;
       char       *tptr;
       char       *bptr;
       int         blen;
       int         ival;

       /* Get the field values (we ignore possibility they are null!)
*/
       iptr = PQgetvalue(res, i, i_fnum);
       tptr = PQgetvalue(res, i, t_fnum);
       bptr = PQgetvalue(res, i, b_fnum);

       /*
        * The binary representation of INT4 is in network byte order,
which
        * we'd better coerce to the local byte order.
        */
       ival = ntohl(*((uint32_t *) iptr));

       /*
        * The binary representation of TEXT is, well, text, and since
libpq
        * was nice enough to append a zero byte to it, it'll work
just fine
        * as a C string.
        *
        * The binary representation of BYTEA is a bunch of bytes,
which could
        * include embedded nulls so we have to pay attention to field
length.
        */
       blen = PQgetlength(res, i, b_fnum);

       printf("tuple %d: got\n", i);
       printf(" i = (%d bytes) %d\n",
              PQgetlength(res, i, i_fnum), ival);
       printf(" t = (%d bytes) '%s'\n",
              PQgetlength(res, i, t_fnum), tptr);
       printf(" b = (%d bytes) ", blen);
       for (j = 0; j < blen; j++)
           printf("\\%03o", bptr[j]);
       printf("\n\n");
```

```
    }
}

int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    const char *paramValues[1];
    int         paramLengths[1];
    int         paramFormats[1];
    uint32_t    binaryIntVal;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * The point of this program is to illustrate use of PQexecParams(
) with
     * out-of-line parameters, as well as binary transmission of data.
     *
     * This first example transmits the parameters as text, but
 receives the
     * results in binary format.  By using out-of-line parameters we
 can
     * avoid a lot of tedious mucking about with quoting and escaping,
 even
     * though the data is text.  Notice how we don't have to do
 anything
     * special with the quote mark in the parameter value.
     */
```

```
   /* Here is our out-of-line parameter value */
   paramValues[0] = "joe's place";

   res = PQexecParams(conn,
                      "SELECT * FROM test1 WHERE t = $1",
                      1,       /* one param */
                      NULL,    /* let the backend deduce param type
*/
                      paramValues,
                      NULL,    /* don't need param lengths since text
*/
                      NULL,    /* default to all text params */
                      1);      /* ask for binary results */

   if (PQresultStatus(res) != PGRES_TUPLES_OK)
   {
       fprintf(stderr, "SELECT failed: %s", PQerrorMessage(conn));
       PQclear(res);
       exit_nicely(conn);
   }

   show_binary_results(res);

   PQclear(res);

   /*
    * In this second example we transmit an integer parameter in
binary
    * form, and again retrieve the results in binary form.
    *
    * Although we tell PQexecParams we are letting the backend deduce
    * parameter type, we really force the decision by casting the
parameter
    * symbol in the query text.  This is a good safety measure when
sending
    * binary parameters.
    */

   /* Convert integer value "2" to network byte order */
   binaryIntVal = htonl((uint32_t) 2);

   /* Set up parameter arrays for PQexecParams */
   paramValues[0] = (char *) &binaryIntVal;
   paramLengths[0] = sizeof(binaryIntVal);
   paramFormats[0] = 1;        /* binary */

   res = PQexecParams(conn,
                      "SELECT * FROM test1 WHERE i = $1::int4",
                      1,       /* one param */
                      NULL,    /* let the backend deduce param type
*/
                      paramValues,
                      paramLengths,
```

```
                        paramFormats,
                        1);      /* ask for binary results */

    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
        fprintf(stderr, "SELECT failed: %s", PQerrorMessage(conn));
        PQclear(res);
        exit_nicely(conn);
    }

    show_binary_results(res);

    PQclear(res);

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

第 32 章 大对象

PostgreSQL提供一种大对象机制，允许以流式方式访问存储在专用大对象结构中的用户数据。在处理大到无法方便地整体操作的数据值时，这种流式访问非常有用。

本章介绍PostgreSQL大对象数据的实现，以及相应的编程接口和查询语言接口。本章示例使用libpq C 库，但大多数PostgreSQL原生编程接口都支持等效的功能。其他接口也可能在内部使用大对象接口来为大值提供通用支持，这里不作说明。

32.1. 简介

所有大对象都存放在一个名为`pg_largeobject`的系统表中。PostgreSQL还支持一种名为“TOAST”的存储系统，它会自动将大于单个数据库页的值存储到每个表各自的二级存储区域中。这使得大对象机制在一定程度上已经过时。大对象机制仍然保留的一个优势是，它允许的值大小可达 2 GB，而经过TOAST处理的字段最大只能到 1 GB。此外，大对象可以通过一种读写 API 进行随机修改，这比使用TOAST执行此类操作更高效。

32.2. 实现特性

大对象实现会把大对象拆分成若干“块”，并将这些块存储在数据库行中。一个 B-树索引保证在执行随机访问读写时，能够快速找到正确的块号。

为与以前的版本兼容，请参见lo_compat_privileges。

自PostgreSQL 9.0 起，大对象具有拥有者和一组访问权限，可用GRANT和REVOKE管理。读取大对象需要`SELECT`权限，写入或截断大对象需要`UPDATE`权限。只有大对象的拥有者（或者数据库超级用户）才能删除（unlink）、注释或更改大对象的拥有者。

32.3. 客户端接口

PostgreSQL客户端接口库为访问大对象提供了支持。使用这些函数对大对象进行的所有操作都必须发生在一个 SQL 事务块内。PostgreSQL的大对象接口是仿照Unix文件系统接口设计的，提供了与`open`、`read`、`write`、`lseek`等相对应的操作。

在libpq中使用大对象接口的客户端应用应包含头文件`libpq/libpq-fs.h`并与libpq库链接。

32.3.1. 创建一个大对象

函数

```
Oid lo_creat(PGconn *conn, int mode);
```

创建一个新的大对象。返回值是分配给新大对象的 OID，失败时为`InvalidOid`（零）。自PostgreSQL 8.1 起，*mode*未被使用且被忽略；不过，为了与更早版本的向后兼容，最好将其设置为`INV_READ`、`INV_WRITE`或`INV_READ | INV_WRITE`。（这些符号常量定义在头文件`libpq/libpq-fs.h`中。）

例如：

```
inv_oid = lo_creat(conn, INV_READ|INV_WRITE);
```

函数

```
Oid lo_create(PGconn *conn, Oid lobjId);
```

也会创建一个新的大对象。要分配的 OID 可以由*lobjId*指定；如果指定了它，而该 OID 已经被某个大对象使用，则会失败。如果*lobjId*是`InvalidOid`（零），则`lo_create`会分配一个未使用的 OID（其行为与`lo_creat`相同）。返回值是分配给新大对象的 OID，失败时为`InvalidOid`（零）。

`lo_create`是PostgreSQL 8.1 新增的；如果将该函数用于更早版本的服务器，它会失败并返回`InvalidOid`。

例如：

```
inv_oid = lo_create(conn, desired_oid);
```

32.3.2. 导入一个大对象

要把一个操作系统文件导入为大对象，调用

```
Oid lo_import(PGconn *conn, const char *filename);
```

*filename*指定要作为大对象导入的操作系统文件名。返回值是分配给新大对象的 OID，失败时为`InvalidOid`（零）。注意，该文件是由客户端接口库读取的，而不是由服务器读取的；因此它必须存在于客户端文件系统中，并且对客户端应用可读。

函数

```
Oid lo_import_with_oid(PGconn *conn, const char *filename, Oid lobjId)
;
```

也会导入一个新的大对象。要分配的 OID 可以由*lobjId*指定；如果指定了它，而该 OID 已经被某个大对象使用，则会失败。如果*lobjId*是`InvalidOid`（零），则`lo_import_with_oid`会分配一个未使用的 OID（其行为与`lo_import`相同）。返回值是分配给新大对象的 OID，失败时为`InvalidOid`（零）。

`lo_import_with_oid`是PostgreSQL 8.4 新增的，它在内部使用了 8.1 新增的`lo_create`；如果将该函数用于 8.0 或更早版本的服务器，它会失败并返回`InvalidOid`。

32.3.3. 导出一个大对象

要把一个大对象导出到操作系统文件中，调用

```
int lo_export(PGconn *conn, Oid lobjId, const char *filename);
```

*lobjId*参数指定要导出的大对象的 OID，*filename*参数指定操作系统文件名。注意，该文件是由客户端接口库写入的，而不是由服务器写入的。成功时返回 1，失败时返回 -1。

32.3.4. 打开一个现有的大对象

要打开一个现有的大对象以供读取或写入，调用

```
int lo_open(PGconn *conn, Oid lobjId, int mode);
```

*lobjId*参数指定要打开的大对象的 OID。*mode*位控制该对象是以读取（INV_READ）、写入（INV_WRITE）还是两者兼有的方式打开。（这些符号常量定义在头文件libpq/libpq-fs.h中。）lo_open返回一个（非负的）大对象描述符，供后续在lo_read、lo_write、lo_lseek、lo_tell和lo_close中使用。大对象在被创建之前不能被打开。
该描述符只在当前事务持续期间有效。失败时返回 -1。

服务器当前不区分INV_WRITE和INV_READ | INV_WRITE这两种模式：在这两种情况下都允许通过该描述符读取。不过，这两种模式与单独使用INV_READ存在一个重要区别：使用INV_READ时，不能通过该描述符写入，而且从该描述符读取到的数据会反映执行lo_open时活动事务快照中的大对象内容，而不受本事务或其他事务之后写入的影响。对于以INV_WRITE打开的描述符，读取返回的数据会反映其他已提交事务的所有写入以及当前事务的写入。
这类似于普通 SQL SELECT命令在REPEATABLE READ与READ COMMITTED事务模式下的行为差异。

例如：

```
inv_fd = lo_open(conn, inv_oid, INV_READ|INV_WRITE);
```

32.3.5. 向大对象写入数据

函数

```
int lo_write(PGconn *conn, int fd, const char *buf, size_t len);
```

将*buf*中的*len*字节写入大对象描述符*fd*。*fd*参数必须是先前由lo_open返回的大对象描述符。返回值是实际写入的字节数。发生错误时，返回值为负。

32.3.6. 从大对象读取数据

函数

```
int lo_read(PGconn *conn, int fd, char *buf, size_t len);
```

从大对象描述符*fd*中读取*len*字节到*buf*中。*fd*参数必须是先前由lo_open返回的大对象描述符。返回实际读取的字节数。发生错误时，返回值为负。

32.3.7. 在大对象中定位

要改变与大对象描述符关联的当前读或写位置，调用

```
int lo_lseek(PGconn *conn, int fd, int offset, int whence);
```

该函数将由*fd*标识的大对象描述符的当前位置指针移动到由*offset*指定的新位置。*whence*的有效值是SEEK_SET（从对象起始处定位）、SEEK_CUR（从当前位置定位）以及SEEK_END（从对象末尾定位）。返回值是新的位置指针，出错时为 -1。

32.3.8. 获取大对象的当前位置

要取得大对象描述符当前的读或写位置，调用

```
int lo_tell(PGconn *conn, int fd);
```

如果发生错误，返回值是 -1。

32.3.9. 截断一个大对象

要把一个大对象截断为给定长度，调用

```
int lo_truncate(PGcon *conn, int fd, size_t len);
```

该函数把大对象描述符*fd*对应的大对象截断为长度*len*。*fd*参数必须是先前由lo_open返回的大对象描述符。如果*len*大于大对象当前的长度，则会用零字节（'\0'）把该大对象扩展。成功时，lo_truncate返回零；出错时返回值为 -1。

文件偏移量不会改变。

lo_truncate是PostgreSQL 8.3 新增的；如果将该函数用于更早版本的服务器，它会失败并返回一个负值。

32.3.10. 关闭一个大对象描述符

可以通过调用

```
int lo_close(PGconn *conn, int fd);
```

来关闭一个大对象描述符，其中*fd*是由lo_open返回的大对象描述符。成功时，lo_close返回零；出错时返回值为 -1。

任何在事务结束时仍保持打开的大对象描述符都会被自动关闭。

32.3.11. 移除一个大对象

要从数据库中移除一个大对象，调用

```
int lo_unlink(PGconn *conn, Oid lobjId);
```

*lobjId*参数指定要移除的大对象的 OID。成功时返回 1，失败时返回 -1。

32.4. 服务器端函数

前面介绍的每个客户端函数都有对应的服务器端函数可以通过 SQL 调用；实际上，大多数客户端函数都只是等价服务器端函数的接口。通过 SQL 命令调用真正有用的是 lo_creat、lo_create、lo_unlink、lo_import 和 lo_export。下面是它们的使用示例：

```
CREATE TABLE image (
    name            text,
    raster          oid
);

SELECT lo_creat(-1);       -- returns OID of new, empty large object

SELECT lo_create(43213);   -- attempts to create large object with OID
 43213

SELECT lo_unlink(173454);  -- deletes large object with OID 173454

INSERT INTO image (name, raster)
```

```
    VALUES ('beautiful image', lo_import('/etc/motd'));

INSERT INTO image (name, raster)  -- same as above, but specify OID to
 use
    VALUES ('beautiful image', lo_import('/etc/motd', 68583));

SELECT lo_export(image.raster, '/tmp/motd') FROM image
    WHERE name = 'beautiful image';
```

服务器端 `lo_import` 和 `lo_export` 函数的行为与客户端对应函数有很大不同。这两个函数使用运行数据库服务器的用户的权限，读写服务器文件系统中的文件，因此它们的用途受限制。
相比之下，客户端导入和导出函数读写客户端文件系统中的文件，使用客户端程序的权限。
客户端函数不要求任何特殊的数据库权限。

`lo_read` 和 `lo_write` 的功能也可以通过服务器端调用获得，但服务器端函数名与客户端接口不同：不包含下划线。必须将这些函数调用为`loread` 和 `lowrite`。

32.5. 示例程序

例　32.1是一个示例程序，它展示了如何使用libpq中的大对象接口。程序中有些部分已经被注释掉，但仍保留在源代码中供读者参考。该程序也可以在源代码发行包的`src/test/examples/testlo.c`中找到。

例 32.1. 用libpq操作大对象的示例程序

```
/*--------------------------------------------------------------
 *
 * testlo.c--
 *    test using large objects with libpq
 *
 * Copyright (c) 1994, Regents of the University of California
 *
 *--------------------------------------------------------------
 */
#include <stdio.h>
#include "libpq-fe.h"
#include "libpq/libpq-fs.h"

#define BUFSIZE          1024

/*
 * importFile
 *    import file "in_filename" into database as large object "lobj
Oid"
 *
 */
Oid
importFile(PGconn *conn, char *filename)
{
    Oid         lobjId;
    int         lobj_fd;
    char        buf[BUFSIZE];
    int         nbytes,
```

```
                tmp;
    int         fd;

    /*
     * open the file to be read in
     */
    fd = open(filename, O_RDONLY, 0666);
    if (fd < 0)
    {                           /* error */
        fprintf(stderr, "cannot open unix file %s\n", filename);
    }

    /*
     * create the large object
     */
    lobjId = lo_creat(conn, INV_READ | INV_WRITE);
    if (lobjId == 0)
        fprintf(stderr, "cannot create large object\n");

    lobj_fd = lo_open(conn, lobjId, INV_WRITE);

    /*
     * read in from the Unix file and write to the inversion file
     */
    while ((nbytes = read(fd, buf, BUFSIZE)) > 0)
    {
        tmp = lo_write(conn, lobj_fd, buf, nbytes);
        if (tmp < nbytes)
            fprintf(stderr, "error while reading large object\n");
    }

    (void) close(fd);
    (void) lo_close(conn, lobj_fd);

    return lobjId;
}

void
pickout(PGconn *conn, Oid lobjId, int start, int len)
{
    int         lobj_fd;
    char       *buf;
    int         nbytes;
    int         nread;

    lobj_fd = lo_open(conn, lobjId, INV_READ);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    lo_lseek(conn, lobj_fd, start, SEEK_SET);
    buf = malloc(len + 1);
```

```
    nread = 0;
    while (len - nread > 0)
    {
        nbytes = lo_read(conn, lobj_fd, buf, len - nread);
        buf[nbytes] = ' ';
        fprintf(stderr, ">>> %s", buf);
        nread += nbytes;
    }
    free(buf);
    fprintf(stderr, "\n");
    lo_close(conn, lobj_fd);
}

void
overwrite(PGconn *conn, Oid lobjId, int start, int len)
{
    int         lobj_fd;
    char       *buf;
    int         nbytes;
    int         nwritten;
    int         i;

    lobj_fd = lo_open(conn, lobjId, INV_WRITE);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    lo_lseek(conn, lobj_fd, start, SEEK_SET);
    buf = malloc(len + 1);

    for (i = 0; i < len; i++)
        buf[i] = 'X';
    buf[i] = ' ';

    nwritten = 0;
    while (len - nwritten > 0)
    {
        nbytes = lo_write(conn, lobj_fd, buf + nwritten, len -
 nwritten);
        nwritten += nbytes;
    }
    free(buf);
    fprintf(stderr, "\n");
    lo_close(conn, lobj_fd);
}

/*
 * exportFile
 *    export large object "lobjOid" to file "out_filename"
 *
 */
```

```
void
exportFile(PGconn *conn, Oid lobjId, char *filename)
{
    int         lobj_fd;
    char        buf[BUFSIZE];
    int         nbytes,
                tmp;
    int         fd;

    /*
     * open the large object
     */
    lobj_fd = lo_open(conn, lobjId, INV_READ);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    /*
     * open the file to be written to
     */
    fd = open(filename, O_CREAT | O_WRONLY, 0666);
    if (fd < 0)
    {                           /* error */
        fprintf(stderr, "cannot open unix file %s\n",
                filename);
    }

    /*
     * read in from the inversion file and write to the Unix file
     */
    while ((nbytes = lo_read(conn, lobj_fd, buf, BUFSIZE)) > 0)
    {
        tmp = write(fd, buf, nbytes);
        if (tmp < nbytes)
        {
            fprintf(stderr, "error while writing %s\n",
                    filename);
        }
    }

    (void) lo_close(conn, lobj_fd);
    (void) close(fd);

    return;
}

void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}
```

```
int
main(int argc, char **argv)
{
    char       *in_filename,
               *out_filename;
    char       *database;
    Oid         lobjOid;
    PGconn     *conn;
    PGresult   *res;

    if (argc != 4)
    {
        fprintf(stderr, "Usage: %s database_name in_filename out_
filename\n",
                argv[0]);
        exit(1);
    }

    database = argv[1];
    in_filename = argv[2];
    out_filename = argv[3];

    /*
     * set up the connection
     */
    conn = PQsetdb(NULL, NULL, NULL, NULL, database);

    /* check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) == CONNECTION_BAD)
    {
        fprintf(stderr, "Connection to database '%s' failed.\n",
 database);
        fprintf(stderr, "%s", PQerrorMessage(conn));
        exit_nicely(conn);
    }

    res = PQexec(conn, "begin");
    PQclear(res);

    printf("importing file %s\n", in_filename);
/*  lobjOid = importFile(conn, in_filename); */
    lobjOid = lo_import(conn, in_filename);
/*
    printf("as large object %d.\n", lobjOid);

    printf("picking out bytes 1000-2000 of the large object\n");
    pickout(conn, lobjOid, 1000, 1000);

    printf("overwriting bytes 1000-2000 of the large object with X's
\n");
    overwrite(conn, lobjOid, 1000, 1000);
*/
```

```
    printf("exporting large object to file %s\n", out_filename);
/*    exportFile(conn, lobjOid, out_filename); */
    lo_export(conn, lobjOid, out_filename);

    res = PQexec(conn, "end");
    PQclear(res);
    PQfinish(conn);
    exit(0);
}
```

第 33 章 ECPG - C 中的嵌入式 SQL

这一章描述了用于PostgreSQL的嵌入式SQL包。它由 Linus Tolke（<linus@epact.se>）和 Michael Meskes（<meskes@postgresql.org>）编写。最初它是为了与C一起工作而编写的。它也能与C++配合，但是它还不识别所有的C++结构。

这份文档还很不完整。不过，由于这一接口是标准化的，因此可以在许多关于 SQL 的资料中找到补充信息。

33.1. 概念

嵌入式 SQL 程序由普通编程语言编写的代码（这里是 C）与位于特殊标记区段中的 SQL 命令混合组成。要构建该程序，首先要将源代码（*.pgc）交给嵌入式 SQL 预处理器，它会把源代码转换成普通 C 程序（*.c），之后就可以由 C 编译器处理。（关于编译和链接的细节见第 33.10 节。）转换后的 ECPG 应用会通过嵌入式 SQL 库（ecpglib）调用 libpq 库中的函数，并使用普通的前端/后端协议与 PostgreSQL 服务器通信。

与其他在 C 代码中处理SQL命令的方法相比，嵌入式SQL有其优势。首先，它能处理与你的C程序变量之间来回传递信息的繁琐工作。其次，程序中的 SQL 代码会在构建时接受语法正确性检查。第三，C 中的嵌入式SQL由SQL标准规定，并且得到许多其他SQL数据库系统的支持。PostgreSQL 的实现被设计为尽可能贴合该标准，因此通常可以比较容易地把为其他 SQL 数据库编写的嵌入式SQL程序移植到PostgreSQL。

如前所述，为嵌入式 SQL 接口编写的程序，就是插入了特殊代码的普通 C 程序，这些特殊代码用于执行数据库相关操作，其形式始终如下：

```
EXEC SQL ...;
```

这些语句在语法上占据一条 C 语句的位置。根据具体语句的不同，它们可以出现在全局层级，也可以出现在函数内部。嵌入式 SQL 语句遵循普通 SQL 代码的大小写规则，而不是 C 的规则。

下列小节解释了所有嵌入式 SQL 语句。

33.2. 管理数据库连接

这一节描述如何打开、关闭以及切换数据库连接。

33.2.1. 连接到数据库服务器

使用以下语句连接数据库：

```
EXEC SQL CONNECT TO target [AS connection-name] [USER user-name];
```

其中，*target* 可以按以下方式指定：

- *dbname*[@*hostname*][:*port*]
- tcp:postgresql://*hostname*[:*port*][/*dbname*][?*options*]
- unix:postgresql://*hostname*[:*port*][/*dbname*][?*options*]
- 包含上述某种形式的 SQL 字符串常量
- 对包含上述某种形式的字符变量的引用（见示例）

- DEFAULT

如果直接写出连接目标（即不通过变量引用），且未用引号括起该值，就会应用普通 SQL 的大小写不敏感规则。此时，也可以按需分别用双引号括起各个参数。实际使用中，
采用单引号括起的字符串字面值或变量引用，可能更不容易出错。连接目标 DEFAULT 会以默认用户名发起到默认数据库的连接，此时不能单独指定用户名或连接名。

也有不同的方法来指定用户名：

- *username*
- *username*/*password*
- *username* IDENTIFIED BY *password*
- *username* USING *password*

如上所述，参数*username*以及*password*可以是一个 SQL 标识符、一个 SQL 字符串常量或者一个对字符变量的引用。

*connection-name*被用来在一个程序中处理多个连接。如果一个程序只使用一个连接，
则可以省略它。最近被打开的连接将成为当前连接，当一个 SQL 语句要被执行时，将默认使用它（见这一章稍后的部分）。

以下是一些 CONNECT 语句的示例：

```
EXEC SQL CONNECT TO mydb@sql.mydomain.com;

EXEC SQL CONNECT TO unix:postgresql://sql.mydomain.com/mydb AS
 myconnection USER john;

EXEC SQL BEGIN DECLARE SECTION;
const char *target = "mydb@sql.mydomain.com";
const char *user = "john";
EXEC SQL END DECLARE SECTION;
 ...
EXEC SQL CONNECT TO :target USER :user;
```

最后一种形式使用了上面提到的字符变量引用。后续章节会说明，如何在 C 变量前加冒号，以便在 SQL 语句中使用它们。

请注意，连接目标的格式并未在 SQL 标准中规定。因此，如果你想开发可移植的应用，最好采用类似上面最后一个示例的方法，把连接目标字符串封装在某个地方。

33.2.2. 选择一个连接

嵌入式 SQL 程序中的 SQL 语句默认在当前连接上执行，即最近打开的连接。如果应用需要管理多个连接，可以采用两种方式。

第一种方式是显式地为每一个 SQL 语句选择一个连接，例如：

```
EXEC SQL AT connection-name SELECT ...;
```

如果应用需要以混合的顺序使用多个连接，这种方式特别合适。

如果你的应用使用多个线程执行，它们不能并发地共享一个连接。
你必须显式地控制对连接的访问（使用互斥量）或者为每个线程使用一个连接。
如果每个线程使用自己的连接，将需要使用 AT 子句来指定该线程将使用的连接。

第二种方式是执行一个语句来切换当前的连接。该语句是：

```
EXEC SQL SET CONNECTION connection-name;
```

如果有很多语句要在同一个连接上执行，这种方式尤其方便。它不是线程感知的。

这里有一个管理多个数据库连接的例子程序：

```
#include <stdio.h>

EXEC SQL BEGIN DECLARE SECTION;
    char dbname[1024];
EXEC SQL END DECLARE SECTION;

int
main()
{
    EXEC SQL CONNECT TO testdb1 AS con1 USER testuser;
    EXEC SQL CONNECT TO testdb2 AS con2 USER testuser;
    EXEC SQL CONNECT TO testdb3 AS con3 USER testuser;

    /* 这个查询将在最近打开的数据库 "testdb3" 中执行。 */
    EXEC SQL SELECT current_database() INTO :dbname;
    printf("current=%s (should be testdb3)\n", dbname);

    /* 使用 "AT" 在 "testdb2" 中运行查询 */
    EXEC SQL AT con2 SELECT current_database() INTO :dbname;
    printf("current=%s (should be testdb2)\n", dbname);

    /* 将当前连接切换到 "testdb1"。 */
    EXEC SQL SET CONNECTION con1;

    EXEC SQL SELECT current_database() INTO :dbname;
    printf("current=%s (should be testdb1)\n", dbname);

    EXEC SQL DISCONNECT ALL;
    return 0;
}
```

这个例子将产生这样的输出：

```
current=testdb3 (should be testdb3)
current=testdb2 (should be testdb2)
current=testdb1 (should be testdb1)
```

33.2.3. 关闭一个连接

要关闭连接，请使用以下语句：

```
EXEC SQL DISCONNECT [connection];
```

其中，*connection* 可以按以下方式指定：

- *connection-name*
- DEFAULT

- CURRENT
- ALL

如果未指定连接名，则关闭当前连接。

应用程序始终显式断开其打开的每个连接，是良好的编程习惯。

33.3. 运行 SQL 命令

任何 SQL 命令都可以在一个嵌入式 SQL 应用中被运行。下面是一些在嵌入式 SQL 应用中运行 SQL 命令的例子。

33.3.1. 执行 SQL 语句

创建一个表：

```
EXEC SQL CREATE TABLE foo (number integer, ascii char(16));
EXEC SQL CREATE UNIQUE INDEX num1 ON foo(number);
EXEC SQL COMMIT;
```

插入行：

```
EXEC SQL INSERT INTO foo (number, ascii) VALUES (9999, 'doodad');
EXEC SQL COMMIT;
```

删除行：

```
EXEC SQL DELETE FROM foo WHERE number = 9999;
EXEC SQL COMMIT;
```

更新：

```
EXEC SQL UPDATE foo
    SET ascii = 'foobar'
    WHERE number = 9999;
EXEC SQL COMMIT;
```

返回一个单一结果行的`SELECT`语句也可以直接使用`EXEC SQL`执行。要处理有多行的结果集，一个应用必须使用一个游标，可参考下面的第 33.3.2 节（作为一种特殊情况，一个应用可以一次取出多行到一个数组主变量中，参考第 33.4.4.3.1 节）。

单行选择：

```
EXEC SQL SELECT foo INTO :FooBar FROM table1 WHERE ascii = 'doodad';
```

还有，一个配置参数可以用`SHOW`命令检索：

```
EXEC SQL SHOW search_path INTO :var;
```

`:`*`something`*形式的词元是主变量，即它们指向 C 程序中的变量。其说明见第 33.4 节。

33.3.2. 使用游标

要读取包含多行的结果集，应用程序必须声明一个游标，并从中逐行取出数据。
使用游标的步骤如下：声明游标、打开游标、从中取出一行、重复取行，最后关闭游标。

使用游标选择：

```
EXEC SQL DECLARE foo_bar CURSOR FOR
    SELECT number, ascii FROM foo
    ORDER BY ascii;
EXEC SQL OPEN foo_bar;
EXEC SQL FETCH foo_bar INTO :FooBar, DooDad;
...
EXEC SQL CLOSE foo_bar;
EXEC SQL COMMIT;
```

有关游标声明的更多细节，参见DECLARE；有关 `FETCH` 命令的详细信息，参见FETCH。

> **注意**
>
> ECPG `DECLARE`命令实际上不会向 PostgreSQL 后端发送语句。在执行`OPEN`命令时，游标才会在后端被打开（使用后端的`DECLARE`命令）。

33.3.3. 管理事务

在默认模式下，只有发出`EXEC SQL COMMIT`时命令才会提交。嵌入式 SQL 接口也支持事务自动提交（类似于psql的默认行为），可以通过`ecpg`的`-t`命令行选项（见ecpg）或者`EXEC SQL SET AUTOCOMMIT TO ON`语句启用。在自动提交模式中，除非位于显式事务块内，每条命令都会自动提交。这种模式也可以通过`EXEC SQL SET AUTOCOMMIT TO OFF`显式关闭。

可以使用下列事务管理命令：

`EXEC SQL COMMIT`

提交一个进行中的事务。

`EXEC SQL ROLLBACK`

回滚一个进行中的事务。

`EXEC SQL SET AUTOCOMMIT TO ON`

启用自动提交模式。

`SET AUTOCOMMIT TO OFF`

禁用自动提交模式。这是默认值。

33.3.4. 预备语句

当要传给 SQL 语句的值在编译时未知，或者同一语句将使用多次时，预备语句就很有用。

语句使用命令`PREPARE`进行预备。对于还未知的值，使用占位符“`?`”：

```
EXEC SQL PREPARE stmt1 FROM "SELECT oid, datname FROM pg_database
 WHERE oid = ?";
```

如果一个语句返回一个单一行，应用可以在`PREPARE`之后调用`EXECUTE`来执行该语句，同时要用一个`USING`子句为占位符提供真实的值：

```
EXEC SQL EXECUTE stmt1 INTO :dboid, :dbname USING 1;
```

如果一个语句返回多行，应用可以使用一个基于该预备语句声明的游标。要绑定输入参数，该游标必须用一个USING子句打开：

```
EXEC SQL PREPARE stmt1 FROM "SELECT oid,datname FROM pg_database WHERE
 oid > ?";
EXEC SQL DECLARE foo_bar CURSOR FOR stmt1;

/* 到达结果集末尾时，跳出 while 循环 */
EXEC SQL WHENEVER NOT FOUND DO BREAK;

EXEC SQL OPEN foo_bar USING 100;
...
while (1)
{
    EXEC SQL FETCH NEXT FROM foo_bar INTO :dboid, :dbname;
    ...
}
EXEC SQL CLOSE foo_bar;
```

当你不再需要该预备语句时，你应该释放它：

```
EXEC SQL DEALLOCATE PREPARE name;
```

更多有关PREPARE的细节，可参考PREPARE。关于使用占位符和输入参数的细节，可参考第 33.5 节。

33.4. 使用主变量

在第 33.3 节中，你已经看到如何从嵌入式 SQL 程序中执行 SQL 语句。其中一些语句只使用固定值，既无法把用户提供的值插入语句中，也无法让程序处理查询返回的值。
这类语句在实际应用中并不真正有用。本节将详细解释如何使用一种称为主变量的简单机制，在 C 程序与嵌入式 SQL 语句之间传递数据。在嵌入式 SQL 程序中，我们把 SQL 语句视为 C 程序代码中的客体，而 C 是主语言。因此，C 程序中的变量被称为主变量。

另一种在 PostgreSQL 后端和 ECPG 应用之间交换值的方式是使用 SQL 描述符，它在第 33.7 节中介绍。

33.4.1. 概述

在嵌入式 SQL 中，在 C 程序和 SQL 语句之间传递数据特别简单。我们不必让程序把数据拼接进语句中去，那样会带来诸如正确为值加引号等各种复杂问题；只需在 SQL 语句中写出 C 变量的名称，并在前面加上一个冒号即可。例如：

```
EXEC SQL INSERT INTO sometable VALUES (:v1, 'foo', :v2);
```

该语句引用了两个名为v1和v2的 C 变量，同时还使用了一个普通的 SQL 字符串常量，以说明你并不局限于只使用某一种数据。

这种在 SQL 语句中插入 C 变量的方式可以用在 SQL 语句中每一个应该出现值表达式的地方。

33.4.2. 声明节

要把数据从程序传给数据库（例如作为查询参数），或者把数据从数据库传回程序，
用来承载这些数据的 C 变量必须声明在经过特殊标记的区段中，这样嵌入式 SQL 预处理器才能识别它们。

这个区段始于：

```
EXEC SQL BEGIN DECLARE SECTION;
```

并结束于：

```
EXEC SQL END DECLARE SECTION;
```

在这两行之间，必须是正常的 C 变量声明，例如：

```
int   x = 4;
char  foo[16], bar[16];
```

如你所见，可以选择为变量指定初始值。变量的作用域由其声明节在程序中的位置决定。
你也可以使用下面的语法声明变量，这种写法会隐式创建一个声明节：

```
EXEC SQL int i = 4;
```

一个程序中可以根据需要放置任意多个声明节。

这些声明也会作为普通 C 变量声明写入输出文件，因此无需再次声明它们。不准备在 SQL 命令中使用的变量可以照常在这些特殊声明节之外声明。

结构体或联合的定义也必须写在`DECLARE`声明节中。否则预处理器无法处理这些类型，
因为它不知道它们的定义。

33.4.3. 检索查询结果

现在你应该能够把程序产生的数据传递到一个 SQL 命令中了。但是怎么检索一个查询的结果呢？为此，嵌入式 SQL 提供了常规命令`SELECT`和`FETCH`的特殊变体。这些命令有一个特殊的`INTO`子句，它指定被检索到的值要被存储在哪些主变量中。`SELECT`被用于只返回单一行的查询，而`FETCH`被用于使用一个游标返回多行的查询。

这里是一个例子：

```
/*
 * 假定有下列表:
 * CREATE TABLE test1 (a int, b varchar(50));
 */

EXEC SQL BEGIN DECLARE SECTION;
int v1;
VARCHAR v2;
EXEC SQL END DECLARE SECTION;

 ...

EXEC SQL SELECT a, b INTO :v1, :v2 FROM test;
```

那么`INTO`子句出现在选择列表和`FROM`子句之间。选择列表中的元素数量必须和`INTO`
后面列表（也被称为目标列表）的元素数量相等。

这里有一个使用命令`FETCH`的例子：

```
EXEC SQL BEGIN DECLARE SECTION;
int v1;
VARCHAR v2;
```

```
EXEC SQL END DECLARE SECTION;

 ...

EXEC SQL DECLARE foo CURSOR FOR SELECT a, b FROM test;

 ...

do
{
    ...
    EXEC SQL FETCH NEXT FROM foo INTO :v1, :v2;
    ...
} while (...);
```

这里INTO子句出现在所有正常子句的后面。

33.4.4. 类型映射

当 ECPG 应用在 PostgreSQL 服务器和 C 应用之间交换值时（例如从服务器检索查询结果时或者用输入参数执行 SQL 语句时），值需要在 PostgreSQL 数据类型和主语言变量类型（具体来说是 C 语言数据类型）之间转换。ECPG 的要点之一就是它会在大多数情况下自动完成这种转换。

在这方面有两类数据类型：一些简单 PostgreSQL 数据类型（例如integer和text）可以被应用直接读取和写入。其他 PostgreSQL 数据类型（例如timestamp和numeric）只能通过特殊库函数访问，见第 33.4.4.2 节。

表 33.1展示了哪种 PostgreSQL 数据类型对应于哪一种 C 数据类型。当你希望发送或接收一种给定 PostgreSQL 数据类型的值时，你应该在声明节中声明一个具有相应 C 数据类型的 C 变量。

表 33.1. 在 PostgreSQL 数据类型和 C 变量类型之间映射

PostgreSQL 数据类型	主变量类型
smallint	short
integer	int
bigint	long long int
decimal	decimal[a]
numeric	numeric[a]
real	float
double precision	double
serial	int
bigserial	long long int
oid	unsigned int
character(*n*),varchar(*n*),text	char[*n*+1],VARCHAR[*n*+1][b]
name	char[NAMEDATALEN]
timestamp	timestamp[a]
interval	interval[a]
date	date[a]
boolean	bool[c]

[a]这种类型只能通过特殊的库函数访问，见第 33.4.4.2 节。

[b]在 ecpglib.h 中声明
[c]如果不是原生类型，则在ecpglib.h中声明

33.4.4.1. 处理字符串

要处理 SQL 字符串数据类型（例如varchar以及text），有两种可能的方式来声明主变量。

一种方式是使用char[]（一个char数组），这是在 C 中处理字符数据最常见的方式。

```
EXEC SQL BEGIN DECLARE SECTION;
    char str[50];
EXEC SQL END DECLARE SECTION;
```

注意你必须自行管理长度。如果你把这个主变量用作一个查询的目标变量并且该查询返回超过 49 个字符的字符串，那么将会发生缓冲区溢出。

另一种方式是使用VARCHAR类型，它是 ECPG 提供的一种特殊类型。VARCHAR数组的定义会为每个变量分别转换成一个具名struct。这样一个声明：

```
VARCHAR var[180];
```

会被转变成：

```
struct varchar_var { int len; char arr[180]; } var;
```

成员arr容纳包含一个终止零字节的字符串。因此，要在一个VARCHAR
主变量中存储一个字符串，该主变量必须被声明为具有包括零字节终止符的长度。成员len
保存存储在arr中的字符串的长度，不包括终止零字节。
当一个主变量被用作一个查询的输入时，如果strlen(arr)和len不同，将使用短的那一个。

VARCHAR可以被写成大写或小写形式，但是不能大小写混合。

char和VARCHAR主变量也可以保存其他 SQL 类型的值，它们将被存储为字符串形式。

33.4.4.2. 访问特殊数据类型

ECPG 包含一些特殊类型帮助你容易地与来自 PostgreSQL 服务器的一些特殊数据类型交互。特别地，它已经实现了对于numeric、decimal、date、timestamp以及interval类型的支持。这些数据类型无法有效地被映射到简单的主变量类型（例如int、long long int或者char[]），因为它们有一种复杂的内部结构。应用通过声明特殊类型的主变量以及使用 pgtypes 库中的函数来处理这些类型。pgtypes 库（在第 33.6 节中详细描述）包含了处理这些类型的基本函数，这样你不需要仅仅为了给一个时间戳加上一个时间间隔而发送一个查询给 SQL 服务器。

下面的小节描述了这些特殊数据类型。关于 pgtypes 库函数的更多细节，请参考第 33.6 节。

33.4.4.2.1. timestamp, date

这里有一种在 ECPG 主应用中处理timestamp变量的模式。

首先，程序必须包括用于timestamp类型的头文件：

```
#include <pgtypes_timestamp.h>
```

接着，在声明节中声明一个主变量为类型timestamp：

```
EXEC SQL BEGIN DECLARE SECTION;
timestamp ts;
```

```
EXEC SQL END DECLARE SECTION;
```

并且在读入一个值到该主变量中之后，使用 pgtypes 库函数处理它。在下面的例子中，timestamp值被PGTYPEStimestamp_to_asc()函数转变成文本（ASCII）形式：

```
EXEC SQL SELECT now()::timestamp INTO :ts;

printf("ts = %s\n", PGTYPEStimestamp_to_asc(ts));
```

这个例子将展示像下面形式的一些结果：

```
ts = 2010-06-27 18:03:56.949343
```

另外，DATE 类型可以用相同的方式处理。程序必须包括pgtypes_date.h，声明一个主变量为日期类型并且将一个 DATE 值使用PGTYPESdate_to_asc()函数转变成一种文本形式。关于 pgtypes 库函数的更多细节，请参考第 33.6 节。

33.4.4.2.2. interval

对interval类型的处理也类似于timestamp和date类型。不过，必须显式为一个interval类型分配内存。换句话说，该变量的内存空间必须在堆内存中分配，而不是在栈内存中分配。

这里是一个例子程序：

```
#include <stdio.h>
#include <stdlib.h>
#include <pgtypes_interval.h>

int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    interval *in;
EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO testdb;

    in = PGTYPESinterval_new();
    EXEC SQL SELECT '1 min'::interval INTO :in;
    printf("interval = %s\n", PGTYPESinterval_to_asc(in));
    PGTYPESinterval_free(in);

    EXEC SQL COMMIT;
    EXEC SQL DISCONNECT ALL;
    return 0;
}
```

33.4.4.2.3. numeric, decimal

numeric和decimal类型的处理类似于interval类型：需要定义一个指针、在堆上分配一些内存空间并且使用 pgtypes 库函数访问该变量。关于 pgtypes 库函数的更多细节，请参考第 33.6 节。

pgtypes 库没有特别为decimal类型提供函数。一个应用必须使用一个 pgtypes 库函数把它转变成一个numeric变量以便进一步处理。

这里是一个处理numeric和decimal类型变量的例子程序。

```
#include <stdio.h>
#include <stdlib.h>
#include <pgtypes_numeric.h>

EXEC SQL WHENEVER SQLERROR STOP;

int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    numeric *num;
    numeric *num2;
    decimal *dec;
EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO testdb;

    num = PGTYPESnumeric_new();
    dec = PGTYPESdecimal_new();

    EXEC SQL SELECT 12.345::numeric(4,2), 23.456::decimal(4,2) INTO :
num, :dec;

    printf("numeric = %s\n", PGTYPESnumeric_to_asc(num, 0));
    printf("numeric = %s\n", PGTYPESnumeric_to_asc(num, 1));
    printf("numeric = %s\n", PGTYPESnumeric_to_asc(num, 2));

    /* 将 decimal 转换为 numeric, 以显示 decimal 值。 */
    num2 = PGTYPESnumeric_new();
    PGTYPESnumeric_from_decimal(dec, num2);

    printf("decimal = %s\n", PGTYPESnumeric_to_asc(num2, 0));
    printf("decimal = %s\n", PGTYPESnumeric_to_asc(num2, 1));
    printf("decimal = %s\n", PGTYPESnumeric_to_asc(num2, 2));

    PGTYPESnumeric_free(num2);
    PGTYPESdecimal_free(dec);
    PGTYPESnumeric_free(num);

    EXEC SQL COMMIT;
    EXEC SQL DISCONNECT ALL;
    return 0;
}
```

33.4.4.3. 非简单类型的主变量

你也可以把数组、typedef、结构体和指针用作主变量。

33.4.4.3.1. 数组

将数组用作主变量有两种情况。第一种如第 33.4.4.1 节所述，是一种将一些文本字符串存储在char[]或VARCHAR[]中的方法。第二种是不用一个游标从一个查询结果中检索多行。

如果没有一个数组，要处理由多个行组成的查询结果，我们需要使用一个游标以及FETCH命令。但是使用数组主变量，多个行可以被一次收取。该数组的长度必须被定义成足以容纳所有的行，否则很可能会发生一次缓冲区溢出。

下面的例子扫描pg_database系统表并且显示所有可用数据库的 OID 和名称：

```
int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    int dbid[8];
    char dbname[8][16];
    int i;
EXEC SQL END DECLARE SECTION;

    memset(dbname, 0, sizeof(char)* 16 * 8);
    memset(dbid, 0, sizeof(int) * 8);

    EXEC SQL CONNECT TO testdb;

    /* 一次将多行取入数组。 */
    EXEC SQL SELECT oid,datname INTO :dbid, :dbname FROM pg_database;

    for (i = 0; i < 8; i++)
        printf("oid=%d, dbname=%s\n", dbid[i], dbname[i]);

    EXEC SQL COMMIT;
    EXEC SQL DISCONNECT ALL;
    return 0;
}
```

这个例子显示下面的结果（确切的值取决于本地环境）。

```
oid=1, dbname=template1
oid=11510, dbname=template0
oid=11511, dbname=postgres
oid=313780, dbname=testdb
oid=0, dbname=
oid=0, dbname=
oid=0, dbname=
```

33.4.4.3.2. 结构体

一个成员名称匹配查询结果列名的结构体可以被用来一次检索多列。
该结构体使得我们能够在一个单一主变量中处理多列值。

下面的例子通过 pg_database 系统表和 pg_database_size() 函数检索可用数据库的 OID、名称和尺寸。在这个例子中，一个成员名匹配SELECT结果的每一列的结构体变量dbinfo_t被用来检索结果行，而不需要把多个主变量放在FETCH语句中。

```
EXEC SQL BEGIN DECLARE SECTION;
    typedef struct
    {
       int oid;
       char datname[65];
```

```
        long long int size;
    } dbinfo_t;

    dbinfo_t dbval;
EXEC SQL END DECLARE SECTION;

    memset(&dbval, 0, sizeof(dbinfo_t));

    EXEC SQL DECLARE cur1 CURSOR FOR SELECT oid, datname, pg_database_
size(oid) AS size FROM pg_database;
    EXEC SQL OPEN cur1;

    /* 到达结果集末尾时，跳出 while 循环 */
    EXEC SQL WHENEVER NOT FOUND DO BREAK;

    while (1)
    {
        /* 将多列取入一个结构体。 */
        EXEC SQL FETCH FROM cur1 INTO :dbval;

        /* 打印结构体的成员。 */
        printf("oid=%d, datname=%s, size=%lld\n", dbval.oid, dbval.
datname, dbval.size);
    }

    EXEC SQL CLOSE cur1;
```

这个例子会显示下列结果（确切的值取决于本地环境）。

```
oid=1, datname=template1, size=4324580
oid=11510, datname=template0, size=4243460
oid=11511, datname=postgres, size=4324580
oid=313780, datname=testdb, size=8183012
```

结构体主变量“吸收”的列数与结构体的字段数相同。额外的列可以被分配给其他主变量。例如，上面的程序也可以使用结构体外部的`size`变量改写：

```
EXEC SQL BEGIN DECLARE SECTION;
    typedef struct
    {
        int oid;
        char datname[65];
    } dbinfo_t;

    dbinfo_t dbval;
    long long int size;
EXEC SQL END DECLARE SECTION;

    memset(&dbval, 0, sizeof(dbinfo_t));

    EXEC SQL DECLARE cur1 CURSOR FOR SELECT oid, datname, pg_database_
size(oid) AS size FROM pg_database;
    EXEC SQL OPEN cur1;

    /* 到达结果集末尾时，跳出 while 循环 */
```

```
    EXEC SQL WHENEVER NOT FOUND DO BREAK;

    while (1)
    {
        /* 将多列取入一个结构体。 */
        EXEC SQL FETCH FROM cur1 INTO :dbval, :size;

        /* 打印结构体的成员。 */
        printf("oid=%d, datname=%s, size=%lld\n", dbval.oid, dbval.
datname, size);
    }

    EXEC SQL CLOSE cur1;
```

33.4.4.3.3. typedef

使用 `typedef` 关键字将新类型映射到已有类型。

```
EXEC SQL BEGIN DECLARE SECTION;
    typedef char mychartype[40];
    typedef long serial_t;
EXEC SQL END DECLARE SECTION;
```

注意，也可以使用：

```
EXEC SQL TYPE serial_t IS long;
```

此声明不必放在声明区段内。

33.4.4.3.4. 指针

你可以声明指向最常见类型的指针。但要注意，只有使用自动内存分配时，
才能将指针用作查询的目标变量。关于自动内存分配的更多信息，参见第 33.7 节。

```
EXEC SQL BEGIN DECLARE SECTION;
    int   *intp;
    char **charp;
EXEC SQL END DECLARE SECTION;
```

33.4.5. 处理非原始 SQL 数据类型

本节介绍如何在 ECPG 应用中处理非标量以及用户定义的 SQL 层数据类型。注意，这与上一节介绍的非简单类型主变量的处理方式是不同的。

33.4.5.1. 数组

ECPG 不直接支持 SQL 层的多维数组。一维 SQL 数组可以映射为 C 数组主变量，反之亦然。不过，在创建语句时，ecpg 并不知道列的类型，因此无法检查某个 C 数组是否会作为对应 SQL 数组的输入。在处理 SQL 语句输出时，ecpg 则掌握所需信息，因此会检查双方是否都是数组。

如果查询是分别访问数组的元素，那就可以避开在 ECPG 中直接使用数组。此时应使用一种能映射到该元素类型的主变量。例如，如果某列的类型是`integer`数组，就可以使用`int`
类型的主变量；如果元素类型是`varchar`或`text`，则可以使用`char[]`或`VARCHAR[]`
类型的主变量。

下面是一个示例。假定有下列表：

```
CREATE TABLE t3 (
    ii integer[]
);

testdb=> SELECT * FROM t3;
     ii
-------------
 {1,2,3,4,5}
(1 row)
```

下面的示例程序检索该数组的第 4 个元素，并将其存入一个类型为`int`的主变量中：

```
EXEC SQL BEGIN DECLARE SECTION;
int ii;
EXEC SQL END DECLARE SECTION;

EXEC SQL DECLARE cur1 CURSOR FOR SELECT ii[4] FROM t3;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    EXEC SQL FETCH FROM cur1 INTO :ii ;
    printf("ii=%d\n", ii);
}

EXEC SQL CLOSE cur1;
```

该示例会显示下列结果：

```
ii=4
```

要把多个数组元素映射到一个数组类型主变量中的多个元素，
数组列的每一个元素以及主变量数组的每一个元素都必须被单独管理，例如：

```
EXEC SQL BEGIN DECLARE SECTION;
int ii_a[8];
EXEC SQL END DECLARE SECTION;

EXEC SQL DECLARE cur1 CURSOR FOR SELECT ii[1], ii[2], ii[3], ii[4]
 FROM t3;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    EXEC SQL FETCH FROM cur1 INTO :ii_a[0], :ii_a[1], :ii_a[2], :ii_a[
3];
    ...
}
```

注意

```
EXEC SQL BEGIN DECLARE SECTION;
```

```
int ii_a[8];
EXEC SQL END DECLARE SECTION;

EXEC SQL DECLARE cur1 CURSOR FOR SELECT ii FROM t3;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    /* 错误 */
    EXEC SQL FETCH FROM cur1 INTO :ii_a;
    ...
}
```

在这种情况中不会正确工作，因为你无法把一个数组类型列直接映射到一个数组主变量。

另一种变通办法是把数组的外部字符串表示存入`char[]`或`VARCHAR[]`类型的主变量中。关于这种表示的更多细节见第 8.14.2 节。注意，这意味着在宿主程序中无法自然地把它当作数组访问，除非再做一步解析文本表示的处理。

33.4.5.2. 复合类型

ECPG 不直接支持复合类型，但有一种简单的变通办法。可用方案与上文处理数组时类似：要么分别访问每个属性，要么使用外部字符串表示。

对于下列例子，假定有下面的类型和表：

```
CREATE TYPE comp_t AS (intval integer, textval varchar(32));
CREATE TABLE t4 (compval comp_t);
INSERT INTO t4 VALUES ( (256, 'PostgreSQL') );
```

最显而易见的解决方案是单独访问每一个属性。下面的程序通过单独选择类型`comp_t`的每一个属性从例子表中检索数据：

```
EXEC SQL BEGIN DECLARE SECTION;
int intval;
varchar textval[33];
EXEC SQL END DECLARE SECTION;

/* 将复合类型列的每个元素放入 SELECT 列表。 */
EXEC SQL DECLARE cur1 CURSOR FOR SELECT (compval).intval, (compval).
textval FROM t4;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    /* 将复合类型列的每个元素取入主变量。 */
    EXEC SQL FETCH FROM cur1 INTO :intval, :textval;

    printf("intval=%d, textval=%s\n", intval, textval.arr);
}
```

```
EXEC SQL CLOSE cur1;
```

这个示例还可以进一步改进：把在FETCH命令中接收值的主变量集中到一个结构体里。关于结构体形式主变量的更多细节，见第 33.4.4.3.2 节。改写成结构体形式后，两个主变量intval和textval就成为comp_t结构体的成员，并在FETCH命令中指定该结构体。

```
EXEC SQL BEGIN DECLARE SECTION;
typedef struct
{
    int intval;
    varchar textval[33];
} comp_t;

comp_t compval;
EXEC SQL END DECLARE SECTION;

/* 将复合类型列的每个元素放入 SELECT 列表。 */
EXEC SQL DECLARE cur1 CURSOR FOR SELECT (compval).intval, (compval).
textval FROM t4;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    /* 将 SELECT 列表中的所有值放入一个结构体。 */
    EXEC SQL FETCH FROM cur1 INTO :compval;

    printf("intval=%d, textval=%s\n", compval.intval, compval.textval.
arr);
}

EXEC SQL CLOSE cur1;
```

尽管在FETCH命令中使用了结构体，SELECT子句中的属性名仍要逐一列出。可以通过使用*来请求该复合类型值的所有属性，从而进一步改进：

```
...
EXEC SQL DECLARE cur1 CURSOR FOR SELECT (compval).* FROM t4;
EXEC SQL OPEN cur1;

EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    /* 将 SELECT 列表中的所有值放入一个结构体。 */
    EXEC SQL FETCH FROM cur1 INTO :compval;

    printf("intval=%d, textval=%s\n", compval.intval, compval.textval.
arr);
}
...
```

采用这种方法后，尽管 ECPG 本身并不理解复合类型，复合类型仍然几乎可以无缝映射到结构体上。

最后，也可以把复合类型值的外部字符串表示存入char[]或VARCHAR[]类型的主变量中。不过，如果采用这种方法，就很难在宿主程序中访问该值的各个字段。

33.4.5.3. 用户定义的基础类型

ECPG 并不直接支持新的用户定义的基础类型。你可以使用外部字符串表示以及类型char[]或VARCHAR[]的主变量，并且这种方案事实上对很多类型都是合适和足够的。

下面的示例使用数据类型 complex，它来自第 35.11 节中的示例。该类型的外部字符串表示为 (%f,%f)，由函数 complex_in() 和 complex_out() 定义，见第 35.11 节。以下示例将 complex 类型的值 (1,1) 和 (3,3) 分别插入列 a 和 b，然后从表中查询它们。

```
EXEC SQL BEGIN DECLARE SECTION;
    varchar a[64];
    varchar b[64];
EXEC SQL END DECLARE SECTION;

    EXEC SQL INSERT INTO test_complex VALUES ('(1,1)', '(3,3)');

    EXEC SQL DECLARE cur1 CURSOR FOR SELECT a, b FROM test_complex;
    EXEC SQL OPEN cur1;

    EXEC SQL WHENEVER NOT FOUND DO BREAK;

    while (1)
    {
        EXEC SQL FETCH FROM cur1 INTO :a, :b;
        printf("a=%s, b=%s\n", a.arr, b.arr);
    }

    EXEC SQL CLOSE cur1;
```

该示例的结果如下：

```
a=(1,1), b=(3,3)
```

另一种变通方案是避免在 ECPG 中直接使用用户定义的类型，而是创建一个在用户定义的类型和 ECPG 能处理的简单类型之间转换的函数或者类型转换。不过要注意，在类型系统中引入类型转换（特别是隐式类型转换）要非常小心。

例如，

```
CREATE FUNCTION create_complex(r double, i double) RETURNS complex
LANGUAGE SQL
IMMUTABLE
AS $$ SELECT $1 * complex '(1,0')' + $2 * complex '(0,1)' $$;
```

在这个定义之后，下面的语句

```
EXEC SQL BEGIN DECLARE SECTION;
double a, b, c, d;
EXEC SQL END DECLARE SECTION;

a = 1;
b = 2;
c = 3;
```

```
d = 4;

EXEC SQL INSERT INTO test_complex VALUES (create_complex(:a, :b),
 create_complex(:c, :d));
```

具有和

```
EXEC SQL INSERT INTO test_complex VALUES ('(1,2)', '(3,4)');
```

相同的效果。

33.4.6. 指示符

上面的示例没有处理空值。实际上，如果读取示例从数据库取得空值，就会抛出错误。
要向数据库传入空值或从数据库取得空值，需要在每个包含数据的主变量后，
再指定第二个主变量。第二个主变量称为指示符，其中包含一个标志，表示该数据是否为空值；
如果为空值，就会忽略实际主变量的值。下面的示例能正确处理空值的读取：

```
EXEC SQL BEGIN DECLARE SECTION;
VARCHAR val;
int val_ind;
EXEC SQL END DECLARE SECTION:

 ...

EXEC SQL SELECT b INTO :val :val_ind FROM test1;
```

指示符变量 `val_ind` 在数据非空时为零，在数据为空值时为负数。

指示符有另一种功能：如果指示符值为正，它表示值不为空，
但是当它被存储在主变量中时已被截断。

如果向预处理器`ecpg`传入参数`-r no_indicator`，它就会工作在“无指示符”模式下。在这种模式中，如果没有指定指示符变量，那么对于字符串类型，空值会在输入和输出时表现为空串；
对于整数类型，空值会表现为该类型可能的最小值（例如对于`int`就是`INT_MIN`）。

33.5. 动态 SQL

在很多情况中，一个应用必须要执行的特定 SQL 语句在编写该应用时就已知。不过在某些情况中，SQL 语句在运行时构造或者由一个外部来源提供。这样你就不能直接把 SQL 语句嵌入到 C 源代码，不过有一种功能允许你调用在一个字符串变量中提供的任意 SQL 语句。

33.5.1. 执行没有结果集的语句

执行一个任意 SQL 语句的最简单方法是使用命令`EXECUTE IMMEDIATE`。例如：

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "CREATE TABLE test1 (...);";
EXEC SQL END DECLARE SECTION;

EXEC SQL EXECUTE IMMEDIATE :stmt;
```

`EXECUTE IMMEDIATE`可以被用于不返回结果集的 SQL 语句（例如 DDL、`INSERT`、`UPDATE`、`DELETE`）。你不能用这种方法执行检索数据的语句（例如`SELECT`）。
下一节将描述如何执行这一种语句。

33.5.2. 执行一个有输入参数的语句

执行任意 SQL 语句的一种更强大的方法是将语句预备一次并且在每次需要时执行该预备语句。也可以预备一个一般化的语句，然后通过替换参数执行它的特定版本。在预备语句时，在你想要稍后替换参数的地方写上问号。例如：

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "INSERT INTO test1 VALUES(?, ?);";
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE mystmt FROM :stmt;
 ...
EXEC SQL EXECUTE mystmt USING 42, 'foobar';
```

当你不再需要该预备语句时，你应该释放它：

```
EXEC SQL DEALLOCATE PREPARE name;
```

33.5.3. 执行一个有结果集的语句

要执行一个只有单一结果行的 SQL 语句，可以使用`EXECUTE`。要保存结果，在其中增加一个`INTO`子句。

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "SELECT a, b, c FROM test1 WHERE a > ?";
int v1, v2;
VARCHAR v3[50];
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE mystmt FROM :stmt;
 ...
EXEC SQL EXECUTE mystmt INTO :v1, :v2, :v3 USING 37;
```

一个`EXECUTE`命令可以有一个`INTO`子句、一个`USING`子句，可以同时有这两个子句，也可以不带这两个子句。

如果一个查询被期望返回多于一个结果行，应该如下列例子所示使用一个游标（关于游标详见第 33.3.2 节）。

```
EXEC SQL BEGIN DECLARE SECTION;
char dbaname[128];
char datname[128];
char *stmt = "SELECT u.usename as dbaname, d.datname "
             "  FROM pg_database d, pg_user u "
             "  WHERE d.datdba = u.usesysid";
EXEC SQL END DECLARE SECTION;

EXEC SQL CONNECT TO testdb AS con1 USER testuser;

EXEC SQL PREPARE stmt1 FROM :stmt;

EXEC SQL DECLARE cursor1 CURSOR FOR stmt1;
EXEC SQL OPEN cursor1;
```

```
EXEC SQL WHENEVER NOT FOUND DO BREAK;

while (1)
{
    EXEC SQL FETCH cursor1 INTO :dbaname,:datname;
    printf("dbaname=%s, datname=%s\n", dbaname, datname);
}

EXEC SQL CLOSE cursor1;

EXEC SQL COMMIT;
EXEC SQL DISCONNECT ALL;
```

33.6. pgtypes 库

pgtypes 库将PostgreSQL数据库类型映射到 C 中等价的类型以便在 C 程序中使用。它还提供在 C 中对这些类型进行基本计算的函数，即不依赖PostgreSQL服务器进行计算。请看下面的例子：

```
EXEC SQL BEGIN DECLARE SECTION;
   date date1;
   timestamp ts1, tsout;
   interval iv1;
   char *out;
EXEC SQL END DECLARE SECTION;

PGTYPESdate_today(&date1);
EXEC SQL SELECT started, duration INTO :ts1, :iv1 FROM datetbl WHERE
 d=:date1;
PGTYPEStimestamp_add_interval(&ts1, &iv1, &tsout);
out = PGTYPEStimestamp_to_asc(&tsout);
printf("Started + duration: %s\n", out);
free(out);
```

33.6.1. numeric类型

numeric 类型支持任意精度计算。PostgreSQL服务器中的等效类型见第 8.1 节。为了支持任意精度，这类变量需要能够动态扩展和收缩。因此，numeric 变量只能在堆上创建，并通过`PGTYPESnumeric_new`和`PGTYPESnumeric_free`函数管理。decimal 类型与之相似，但精度有限，既可以在栈上创建，也可以在堆上创建。

下列函数可用于处理 numeric 类型：

`PGTYPESnumeric_new`

请求一个指向新分配的numeric变量的指针。

```
numeric *PGTYPESnumeric_new(void);
```

`PGTYPESnumeric_free`

释放一个 numeric 变量及其占用的全部内存。

```
void PGTYPESnumeric_free(numeric *var);
```

`PGTYPESnumeric_from_asc`

从字符串表示中解析一个 numeric 值。

```
numeric *PGTYPESnumeric_from_asc(char *str, char **endptr);
```

例如，可用的格式是：`-2`、`.794`、`+3.44`、`592.49E07`或者 `-32.84e-4`。如果值能被成功地解析，将返回一个有效的指针，否则返回 NULL 指针。目前 ECPG 总是解析整个字符串并且因此当前不支持把第一个非法字符的地址存储在`*endptr`中。你可以安全地把`endptr`设置为 NULL。

`PGTYPESnumeric_to_asc`

返回由`malloc`分配的字符串的指针，它包含numeric类型`num`的字符串表示。

```
char *PGTYPESnumeric_to_asc(numeric *num, int dscale);
```

numeric 值将以`dscale`位小数输出，必要时进行舍入。

`PGTYPESnumeric_add`

把两个numeric变量相加放到第三个numeric变量中。

```
int PGTYPESnumeric_add(numeric *var1, numeric *var2, numeric
 *result);
```

该函数把变量`var1`和`var2`相加放到结果变量`result`中。成功时该函数返回 0，出错时返回 -1。

`PGTYPESnumeric_sub`

把两个numeric变量相减并且把结果返回到第三个numeric变量。

```
int PGTYPESnumeric_sub(numeric *var1, numeric *var2, numeric
 *result);
```

该函数把变量`var2`从变量`var1`中减除。该操作的结果被存储在变量`result`中。成功时该函数返回 0，出错时返回 -1。

`PGTYPESnumeric_mul`

把两个numeric变量相乘并且把结果返回到第三个numeric变量。

```
int PGTYPESnumeric_mul(numeric *var1, numeric *var2, numeric
 *result);
```

该函数把变量`var1`和`var2`相乘。该操作的结果被存储在变量`result`中。成功时该函数返回 0，出错时返回 -1。

`PGTYPESnumeric_div`

把两个numeric变量相除并且把结果返回到第三个numeric变量。

```
int PGTYPESnumeric_div(numeric *var1, numeric *var2, numeric
 *result);
```

该函数用变量`var2`除变量`var1`。该操作的结果被存储在变量`result`中。成功时该函数返回 0，出错时返回 -1。

`PGTYPESnumeric_cmp`

比较两个numeric变量。

```
int PGTYPESnumeric_cmp(numeric *var1, numeric *var2)
```

这个函数比较两个numeric变量。错误时会返回`INT_MAX`。成功时，该函数返回三种可能结果之一：

- `var1`大于`var2`则返回 1
- 如果`var1`小于`var2`则返回 -1
- 如果`var1`和`var2`相等则返回 0

`PGTYPESnumeric_from_int`

将一个 int 变量转换为 numeric 变量。

```
int PGTYPESnumeric_from_int(signed int int_val, numeric *var);
```

该函数接收一个 signed int 类型的变量，并将其值存入 numeric 变量`var`中。成功时返回 0，失败时返回 -1。

`PGTYPESnumeric_from_long`

将一个 long int 变量转换为 numeric 变量。

```
int PGTYPESnumeric_from_long(signed long int long_val, numeric
 *var);
```

该函数接收一个 signed long int 类型的变量，并将其值存入 numeric 变量`var`中。成功时返回 0，失败时返回 -1。

`PGTYPESnumeric_copy`

把一个numeric变量复制到另一个中。

```
int PGTYPESnumeric_copy(numeric *src, numeric *dst);
```

这个函数把`src`指向的变量的值复制到`dst`指向的变量中。成功时该函数返回 0，出错时返回 -1。

`PGTYPESnumeric_from_double`

将一个 double 类型的变量转换为 numeric 变量。

```
int  PGTYPESnumeric_from_double(double d, numeric *dst);
```

该函数接收一个 double 类型的变量，并将结果存入`dst`指向的变量中。成功时返回 0，出错时返回 -1。

`PGTYPESnumeric_to_double`

将一个 numeric 类型的变量转换为 double。

```
int PGTYPESnumeric_to_double(numeric *nv, double *dp)
```

该函数将nv指向的变量中的 numeric 值转换为 double，并存入dp指向的变量中。成功时返回 0，发生错误（包括溢出）时返回 -1。发生溢出时，还会将全局变量errno设为PGTYPES_NUM_OVERFLOW。

PGTYPESnumeric_to_int

将一个 numeric 类型的变量转换为 int。

```
int PGTYPESnumeric_to_int(numeric *nv, int *ip);
```

该函数将nv指向的变量中的 numeric 值转换为 int，并存入ip指向的变量中。成功时返回 0，发生错误（包括溢出）时返回 -1。发生溢出时，还会将全局变量errno设为PGTYPES_NUM_OVERFLOW。

PGTYPESnumeric_to_long

将一个 numeric 类型的变量转换为 long。

```
int PGTYPESnumeric_to_long(numeric *nv, long *lp);
```

该函数将 nv 所指变量中的 numeric 值转换为 long，并存入 lp 所指的变量。成功时返回 0，发生错误（包括溢出）时返回 -1。发生溢出时，还会将全局变量 errno 设为 PGTYPES_NUM_OVERFLOW。

PGTYPESnumeric_to_decimal

将一个 numeric 类型的变量转换为 decimal。

```
int PGTYPESnumeric_to_decimal(numeric *src, decimal *dst);
```

该函数将src指向的变量中的 numeric 值转换成dst指向的 decimal 变量。成功时该函数返回 0，出错时返回 -1，包括溢出。发生溢出时，全局变量errno会被设置为PGTYPES_NUM_OVERFLOW。

PGTYPESnumeric_from_decimal

将一个 decimal 类型的变量转换为 numeric。

```
int PGTYPESnumeric_from_decimal(decimal *src, numeric *dst);
```

该函数将src指向的变量中的 decimal 值转换成dst指向的 numeric 变量。成功时该函数返回 0，出错时返回 -1。由于 decimal 类型被实现为 numeric 类型的一个受限版本，这种转换不会发生溢出。

33.6.2. 日期类型

C 中的 date 类型让程序能够处理 SQL date 类型的数据。PostgreSQL服务器中的等效类型见第 8.5 节。

下列函数可用于处理 date 类型：

PGTYPESdate_from_timestamp

从一个时间戳中抽取日期部分。

```
date PGTYPESdate_from_timestamp(timestamp dt);
```

该函数接收一个时间戳作为它的唯一参数并且从这个时间戳返回抽取的日期部分。

PGTYPESdate_from_asc

从日期的文本表示解析一个日期。

```
date PGTYPESdate_from_asc(char *str, char **endptr);
```

该函数接收一个 C char* 字符串str以及一个指向 C char* 字符串的指针endptr。目前 ECPG 始终解析整个字符串，因此尚不支持将第一个无效字符的地址存入*endptr中。可以放心地将endptr设为 NULL。

注意，该函数始终假定日期采用 MDY 格式，目前 ECPG 中没有变量可以改变这一点。

表 33.2展示了所有允许的输入格式。

表 33.2. PGTYPESdate_from_asc的合法输入格式

Input	Result
January 8, 1999	January 8, 1999
1999-01-08	January 8, 1999
1/8/1999	January 8, 1999
1/18/1999	January 18, 1999
01/02/03	February 1, 2003
1999-Jan-08	January 8, 1999
Jan-08-1999	January 8, 1999
08-Jan-1999	January 8, 1999
99-Jan-08	January 8, 1999
08-Jan-99	January 8, 1999
08-Jan-06	January 8, 2006
Jan-08-99	January 8, 1999
19990108	ISO 8601; January 8, 1999
990108	ISO 8601; January 8, 1999
1999.008	year and day of year
J2451187	Julian day
January 8, 99 BC	year 99 before the Common Era

PGTYPESdate_to_asc

返回一个日期变量的文本表示。

```
char *PGTYPESdate_to_asc(date dDate);
```

该函数接收日期dDate作为它的唯一参数。它将以形式1999-01-18输出该日期，即以YYYY-MM-DD格式输出。

PGTYPESdate_julmdy

从一个 date 类型的变量中提取日、月和年的值。

```
void PGTYPESdate_julmdy(date d, int *mdy);
```

该函数接收日期d以及一个指向有 3 个整数值的数组mdy的指针。变量名就表明了顺序：mdy[0]将被设置为包含月份，mdy[1]将被设置为日的值，而mdy[2]将包含年。

PGTYPESdate_mdyjul

根据一个包含 3 个整数的数组创建日期值，这些整数指定日期的日、月和年。

```
void PGTYPESdate_mdyjul(int *mdy, date *jdate);
```

该函数的第一个参数是包含 3 个整数的数组（mdy），第二个参数是一个指向 date 类型变量的指针，用来保存操作结果。

PGTYPESdate_dayofweek

为一个日期值返回表示它是星期几的数字。

```
int PGTYPESdate_dayofweek(date d);
```

这个函数接收日期变量d作为它唯一的参数并且返回一个整数说明这个日期是星期几。

- 0 - 星期日
- 1 - 星期一
- 2 - 星期二
- 3 - 星期三
- 4 - 星期四
- 5 - 星期五
- 6 - 星期六

PGTYPESdate_today

得到当前日期。

```
void PGTYPESdate_today(date *d);
```

该函数接收一个指向 date 变量的指针（d），并将其指向的变量设为当前日期。

PGTYPESdate_fmt_asc

使用格式掩码将一个 date 类型的变量转换为文本表示。

```
int PGTYPESdate_fmt_asc(date dDate, char *fmtstring, char *outbuf);
```

该函数接收要转换的日期（dDate）、格式掩码（fmtstring）
以及将要保存日期的文本表示的字符串（outbuf）。

成功时，返回 0；如果发生错误，则返回一个负值。

下列字面量是可用的字段说明符：

- dd - 一个月中的第几天。
- mm - 一年中的第几个月。
- yy - 两位数的年份。

- `yyyy` - 四位数的年份。
- `ddd` - 星期几的名称（简写）。
- `mmm` - 月份的名称（简写）。

All other characters are copied 1:1 to the output string.

表 33.3给出了一些可用格式，有助于理解此函数的用法。所有输出行都基于同一个日期：1959 年 11 月 23 日。

表 33.3. PGTYPESdate_fmt_asc的合法输入格式

Format	Result
mmddyy	112359
ddmmyy	231159
yymmdd	591123
yy/mm/dd	59/11/23
yy mm dd	59 11 23
yy.mm.dd	59.11.23
.mm.yyyy.dd.	.11.1959.23.
mmm. dd, yyyy	Nov. 23, 1959
mmm dd yyyy	Nov 23 1959
yyyy dd mm	1959 23 11
ddd, mmm. dd, yyyy	Mon, Nov. 23, 1959
(ddd) mmm. dd, yyyy	(Mon) Nov. 23, 1959

`PGTYPESdate_defmt_asc`

使用格式掩码把一个 C `char*` 字符串转换成 date 类型的值。

```
int PGTYPESdate_defmt_asc(date *d, char *fmt, char *str);
```

该函数接收一个指向用于保存操作结果的 date 值的指针（`d`）、用于解析日期的格式掩码（`fmt`），以及包含日期文本表示的 C char* 字符串（`str`）。文本表示应与格式掩码匹配，但字符串与格式掩码无需一一对应。该函数只分析先后顺序，并查找表示年份位置的`yy`或`yyyy`、表示月份位置的`mm`以及表示日位置的`dd`。

表 33.4给出了一些可用格式，有助于理解此函数的用法。

表 33.4. rdefmtdate的合法输入格式

Format	String	Result
ddmmyy	21-2-54	1954-02-21
ddmmyy	2-12-54	1954-12-02
ddmmyy	20111954	1954-11-20
ddmmyy	130464	1964-04-13
mmm.dd.yyyy	MAR-12-1967	1967-03-12
yy/mm/dd	1954, February 3rd	1954-02-03
mmm.dd.yyyy	041269	1969-04-12

Format	String	Result
yy/mm/dd	In the year 2525, in the month of July, mankind will be alive on the 28th day	2525-07-28
dd-mm-yy	I said on the 28th of July in the year 2525	2525-07-28
mmm.dd.yyyy	9/14/58	1958-09-14
yy/mm/dd	47/03/29	1947-03-29
mmm.dd.yyyy	oct 28 1975	1975-10-28
mmddyy	Nov 14th, 1985	1985-11-14

33.6.3. 时间戳类型

C 中的 timestamp 类型让程序能够处理 SQL timestamp 类型的数据。PostgreSQL服务器中的等效类型见第 8.5 节。

下列函数可用于处理 timestamp 类型：

`PGTYPEStimestamp_from_asc`

将时间戳的文本表示解析为 timestamp 变量。

```
timestamp PGTYPEStimestamp_from_asc(char *str, char **endptr);
```

该函数接收要解析的字符串（`str`）和一个指向 C char* 的指针（`endptr`）。目前 ECPG 始终解析整个字符串，因此尚不支持将第一个无效字符的地址存入`*endptr`中。可以放心地将`endptr`设为 NULL。

该函数在成功时返回解析后的时间戳。出错时返回`PGTYPESInvalidTimestamp`，并将`errno`设置为`PGTYPES_TS_BAD_TIMESTAMP`。关于该值的重要说明，见`PGTYPESInvalidTimestamp`。

通常，输入字符串可以包含允许的日期规范、空白字符和允许的时间规范的任意组合。请注意，ECPG 不支持时区。它可以解析时区，但不像PostgreSQL服务器那样应用任何计算。时区标识符会被静默丢弃。

表 33.5包含了一些输入字符串的示例。

表 33.5. `PGTYPEStimestamp_from_asc`的合法输入格式

Input	Result
1999-01-08 04:05:06	1999-01-08 04:05:06
January 8 04:05:06 1999 PST	1999-01-08 04:05:06
1999-Jan-08 04:05:06.789-8	1999-01-08 04:05:06.789 (time zone specifier ignored)
J2451187 04:05-08:00	1999-01-08 04:05:00 (time zone specifier ignored)

`PGTYPEStimestamp_to_asc`

将时间戳转换为 C char* 字符串。

```
char *PGTYPEStimestamp_to_asc(timestamp tstamp);
```

该函数接收时间戳tstamp作为其唯一参数，
并返回一个包含时间戳文本表示的已分配字符串。

PGTYPEStimestamp_current

检索当前时间戳。

```
void PGTYPEStimestamp_current(timestamp *ts);
```

该函数检索当前时间戳，并将其保存到ts指向的时间戳变量中。

PGTYPEStimestamp_fmt_asc

使用格式掩码将时间戳变量转换为 C char*。

```
int PGTYPEStimestamp_fmt_asc(timestamp *ts, char *output, int str_
len, char *fmtstr);
```

该函数接收一个指向要转换的时间戳的指针作为第一个参数（ts），
一个指向输出缓冲区的指针（output），已为输出缓冲区分配的最大长度（str_len），
以及用于转换的格式掩码（fmtstr）。

成功时返回 0，发生错误时返回负值。

格式掩码可以使用以下格式说明符。它们与libc中strftime函数使用的格式说明符相同。
不是格式说明符的内容都会被复制到输出缓冲区中。

- %A - 替换为完整星期名称的本地化表示。
- %a - 替换为缩写星期名称的本地化表示。
- %B - 替换为完整月份名称的本地化表示。
- %b - 替换为缩写月份名称的本地化表示。
- %C - 替换为年份除以 100 的结果，以十进制数表示；一位数前补零。
- %c - 替换为时间和日期的本地化表示。
- %D - 等同于 %m/%d/%y。
- %d - 替换为一个月中的日，以十进制数表示（01-31）。
- %E* %O* - POSIX 区域设置扩展。序列 %Ec %EC %Ex %EX %Ey %EY %Od %Oe %OH %OI %Om %OM %OS %Ou %OU %OV %Ow %OW %Oy 应该提供替代表示。

  此外，%OB提供月份名称的替代形式，用于单独显示月份、不提及具体日的场合。
- %e - 替换为一个月中的日，以十进制数表示（1-31）；一位数前补空格。
- %F - 等同于%Y-%m-%d。
- %G - 替换为包含世纪的年份，以十进制数表示。该年份是这一周中多数天所在的年份（以星期一为每周第一天）。

- `%g` - 替换为与 `%G` 相同的年份，但以不带世纪的十进制数表示（00-99）。
- `%H` - 替换为小时（24 小时制），以十进制数表示（00-23）。
- `%h` - 与`%b`相同。
- `%I` - 替换为小时（12 小时制），以十进制数表示（01-12）。
- `%j` - 替换为一年中的第几天，以十进制数表示（001-366）。
- `%k` - 替换为小时（24 小时制），以十进制数表示（0-23）；一位数前补空格。
- `%l` - 替换为小时（12 小时制），以十进制数表示（1-12）；一位数前补空格。
- `%M` - 替换为分钟，以十进制数表示（00-59）。
- `%m` - 替换为月份，以十进制数表示（01-12）。
- `%n` - 被换行符替换。
- `%O*` - 与`%E*`相同。
- `%p` - 按需要替换为“上午”或“下午”的本地化表示。
- `%R` - 等同于`%H:%M`。
- `%r` - 等同于`%I:%M:%S %p`。
- `%S` - 替换为秒，以十进制数表示（00-60）。
- `%s` - 替换为从 UTC 纪元起经过的秒数。
- `%T` - 等同于`%H:%M:%S`
- `%t` - 被替换为一个制表符。
- `%U` - 替换为一年中的周数，以星期日为每周第一天，以十进制数表示（00-53）。
- `%u` - 替换为星期几，以星期一为每周第一天，以十进制数表示（1-7）。
- `%V` - 替换为一年中的周数，以星期一为每周第一天，以十进制数表示（01-53）。如果包含 1 月 1 日的那一周有四天或更多天属于新年，该周就是第 1 周；否则，该周是上一年的最后一周，下一周才是第 1 周。
- `%v` - 等同于 `%e-%b-%Y`。
- `%W` - 替换为一年中的周数，以星期一为每周第一天，以十进制数表示（00-53）。
- `%w` - 替换为星期几，以星期日为每周第一天，以十进制数表示（0-6）。
- `%X` - 被时间的本地化表示替换。
- `%x` - 替换为日期的本地化表示。
- `%Y` - 替换为包含世纪的年份，以十进制数表示。
- `%y` - 替换为不带世纪的年份，以十进制数表示（00-99）。

- `%Z` - 被时区名称替换。
- `%z` - 替换为时区相对于 UTC 的偏移量；前导加号表示 UTC 以东，减号表示 UTC 以西，后接小时和分钟，各占两位数字，中间没有分隔符（RFC 822日期头的常见形式）。
- `%+` - 被日期和时间的本地化表示所替换。
- `%-*` - GNU libc扩展。在执行数字输出时不进行任何填充。
- $_* - GNU libc扩展。明确指定填充空间。
- `%0*` - GNU libc 扩展。显式指定用零填充。
- `%%` - 被`%`替换。

`PGTYPEStimestamp_sub`

从一个时间戳减去另一个时间戳，并将结果保存在 interval 类型变量中。

```
int PGTYPEStimestamp_sub(timestamp *ts1, timestamp *ts2, interval
 *iv);
```

该函数从`ts1`指向的时间戳变量中减去`ts2`指向的时间戳变量，并将结果存入`iv`指向的 interval 变量中。

成功时返回 0，发生错误时返回负值。

`PGTYPEStimestamp_defmt_asc`

使用格式掩码从文本表示中解析时间戳值。

```
int PGTYPEStimestamp_defmt_asc(char *str, char *fmt, timestamp *d);
```

该函数接收变量`str`中的时间戳文本表示，以及变量`fmt`中的格式掩码。结果存入`d`指向的变量中。

如果格式掩码`fmt`为 NULL，函数将退回到默认格式掩码，即`%Y-%m-%d %H:%M:%S`。

这是`PGTYPEStimestamp_fmt_asc`的逆函数。可用的格式掩码项见该函数的说明。

`PGTYPEStimestamp_add_interval`

向时间戳变量加上一个 interval 变量。

```
int PGTYPEStimestamp_add_interval(timestamp *tin, interval *span,
 timestamp *tout);
```

该函数接收一个指向 timestamp 变量`tin`的指针，以及一个指向 interval 变量`span`的指针。它将时间间隔加到时间戳上，并将得到的时间戳存入`tout`指向的变量中。

成功时返回 0，发生错误时返回负值。

`PGTYPEStimestamp_sub_interval`

从时间戳变量中减去一个 interval 变量。

```
int PGTYPEStimestamp_sub_interval(timestamp *tin, interval *span,
 timestamp *tout);
```

该函数从`tin`指向的时间戳变量中减去`span`指向的时间间隔变量，并将结果存入`tout`指向的变量中。

成功时返回 0，发生错误时返回负值。

33.6.4. interval 类型

C 中的 interval 类型允许程序处理 SQL interval 类型的数据。PostgreSQL服务器中的等效类型见第 8.5 节。

下列函数可用于处理 interval 类型：

`PGTYPESinterval_new`

返回一个指向新分配 interval 变量的指针。

```
interval *PGTYPESinterval_new(void);
```

`PGTYPESinterval_free`

释放先前分配的 interval 变量所占用的内存。

```
void PGTYPESinterval_new(interval *intvl);
```

`PGTYPESinterval_from_asc`

从文本表示中解析一个 interval 值。

```
interval *PGTYPESinterval_from_asc(char *str, char **endptr);
```

该函数解析输入字符串`str`，并返回一个已分配 interval 变量的指针。目前 ECPG 总是解析整个字符串，因此暂不支持把第一个非法字符的地址存储在`*endptr`中。
可以安全地把`endptr`设为 NULL。

`PGTYPESinterval_to_asc`

将 interval 类型变量转换成文本表示。

```
char *PGTYPESinterval_to_asc(interval *span);
```

该函数把`span`指向的 interval 变量转换成 C char*。输出形式类似于：`@ 1 day 12 hours 59 mins 10 secs`。

`PGTYPESinterval_copy`

复制一个 interval 类型变量。

```
int PGTYPESinterval_copy(interval *intvlsrc, interval *intvldest);
```

该函数把`intvlsrc`指向的 interval 变量复制到`intvldest`指向的 interval 变量中。注意，需要事先为目标变量分配好内存。

33.6.5. decimal类型

decimal类型和numeric类型相似。不过，它被限制为最大精度是 30 个有效位。与numeric类型只能在堆上创建相反，decimal类型既可以在栈上也可以在堆上创建（使用函数

PGTYPESdecimal_new 和PGTYPESdecimal_free）。在第 33.15 节中描述的Informix兼容模式中有很多其它函数可以处理decimal类型。

下列函数可用于处理 decimal 类型，并非仅在libcompat库中提供。

PGTYPESdecimal_new

要求一个指向新分配的decimal变量的指针。

```
decimal *PGTYPESdecimal_new(void);
```

PGTYPESdecimal_free

释放一个 decimal 变量及其占用的全部内存。

```
void PGTYPESdecimal_free(decimal *var);
```

33.6.6. pgtypeslib 的 errno 值

PGTYPES_NUM_BAD_NUMERIC

一个参数应该包含一个numeric变量（或者指向一个numeric变量），但是实际上它的内存表示非法。

PGTYPES_NUM_OVERFLOW

发生一次溢出。由于numeric类型可以处理几乎任何精度，将一个numeric变量转换成其他类型可能导致溢出。

PGTYPES_NUM_UNDERFLOW

发生一次下溢。由于numeric类型可以处理几乎任何精度，将一个numeric变量转换成其他类型可能导致下溢。

PGTYPES_NUM_DIVIDE_ZERO

尝试了一次除零。

PGTYPES_DATE_BAD_DATE

一个非法的日期字符串被传给了PGTYPESdate_from_asc函数。

PGTYPES_DATE_ERR_EARGS

非法参数被传给了PGTYPESdate_defmt_asc函数。

PGTYPES_DATE_ERR_ENOSHORTDATE

PGTYPESdate_defmt_asc函数在输入字符串中发现了一个非法词元。

PGTYPES_INTVL_BAD_INTERVAL

向PGTYPESinterval_from_asc函数传入了无效的 interval 字符串，或者向PGTYPESinterval_to_asc函数传入了无效的 interval 值。

PGTYPES_DATE_ERR_ENOTDMY

PGTYPESdate_defmt_asc函数中的日、月、年赋值不匹配。

`PGTYPES_DATE_BAD_DAY`

`PGTYPESdate_defmt_asc`函数发现了月中的一个非法日值。

`PGTYPES_DATE_BAD_MONTH`

`PGTYPESdate_defmt_asc`函数发现了一个非法的月值。

`PGTYPES_TS_BAD_TIMESTAMP`

一个非法的时间戳字符串被传给了`PGTYPEStimestamp_from_asc`函数，或者一个非法的时间戳值被传给了`PGTYPEStimestamp_to_asc`函数。

`PGTYPES_TS_ERR_EINFTIME`

在一个无法处理无限时间戳值的环境中遇到了这样一个值。

33.6.7. pgtypeslib 的特殊常量

`PGTYPESInvalidTimestamp`

一个表示无效时间戳的 timestamp 类型值。`PGTYPEStimestamp_from_asc`函数在解析出错时返回此值。注意，由于`timestamp`数据类型的内部表示，`PGTYPESInvalid Timestamp`同时也是一个有效的时间戳，其值为`1899-12-31 23:59:59`。为了检测错误，应用程序每次调用`PGTYPEStimestamp_from_asc`后，不仅要检查`PGTYPESInvalid Timestamp`，还要检查`errno != 0`。

33.7. 使用描述符区域

SQL 描述符区域是一种处理`SELECT`、`FETCH`或`DESCRIBE`语句结果的高级方法。SQL 描述符区域将一行数据及其元数据项组合到一个数据结构中。在执行动态 SQL 语句时，结果列的性质可能无法提前获知，此时元数据特别有用。PostgreSQL 提供两种使用描述符区域的方式：命名 SQL 描述符区域和 C 结构体 SQLDA。

33.7.1. 命名 SQL 描述符区域

一个命名 SQL 描述符区域由一个头部以及一个或多个条目描述符区域构成，头部包含与整个描述符相关的信息，而条目描述符区域则描述结果行中的每一列。

在使用 SQL 描述符区域之前，需要先分配一个：

```
EXEC SQL ALLOCATE DESCRIPTOR identifier;
```

identifier 会作为该描述符区域的“变量名”。当不再需要该描述符时，应当释放它：

```
EXEC SQL DEALLOCATE DESCRIPTOR identifier;
```

要使用一个描述符区域，把它指定为`INTO`子句的存储目标（而不是列出主变量）：

```
EXEC SQL FETCH NEXT FROM mycursor INTO SQL DESCRIPTOR mydesc;
```

如果结果集为空，该描述符区域仍然会包含查询的元数据，即字段的名称。

对于还没有执行的预备查询，`DESCRIBE`可以被用来得到其结果集的元数据：

```
EXEC SQL BEGIN DECLARE SECTION;
```

```
char *sql_stmt = "SELECT * FROM table1";
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE stmt1 FROM :sql_stmt;
EXEC SQL DESCRIBE stmt1 INTO SQL DESCRIPTOR mydesc;
```

在 PostgreSQL 9.0 之前，SQL关键词是可选的，因此使用DESCRIPTOR和SQL DESCRIPTOR都会产生命名 SQL 描述符区域。现在该关键词是强制性的，省略SQL关键词会产生 SQLDA 描述符区域（见第 33.7.2 节）。

在DESCRIBE和FETCH语句中，INTO和USING
关键词的用法相似：它们在描述符区域中生成结果集和元数据。

要从描述符区域取得数据，可以将它视为一个包含命名字段的结构。使用以下命令，
从头部取得某个字段的值，并存入主变量：

EXEC SQL GET DESCRIPTOR *name* :*hostvar* = *field*;

目前只定义了一个头部字段：*COUNT*，它表示有多少个条目描述符区域（即结果包含多少列）。
主变量必须是整数类型。要从条目描述符区域取得字段，请使用以下命令：

EXEC SQL GET DESCRIPTOR *name* VALUE *num* :*hostvar* = *field*;

*num*可以是整数字面值，也可以是包含整数的主变量。可用字段如下：

CARDINALITY（整数）

结果集中的行数

DATA

实际的数据项（因此，这个字段的数据类型取决于查询）

DATETIME_INTERVAL_CODE（整数）

当TYPE是9时，DATETIME_INTERVAL_CODE将具有以下值之一：1 表示 DATE，2 表示 TIME，3 表示 TIMESTAMP，4 表示 TIME WITH TIME ZONE，5 表示 TIMESTAMP WITH TIME ZONE。

DATETIME_INTERVAL_PRECISION（整数）

没有实现

INDICATOR（整数）

指示符（表示一个空值或者一个值截断）

KEY_MEMBER（整数）

没有实现

LENGTH（整数）

以字符计的数据长度

NAME（字符串）

列名

`NULLABLE`（整数）

没有实现

`OCTET_LENGTH`（整数）

以字节计的数据字符表示的长度

`PRECISION`（整数）

精度（用于类型`numeric`）

`RETURNED_LENGTH`（整数）

以字符计的数据长度

`RETURNED_OCTET_LENGTH`（整数）

以字节计的数据字符表示的长度

`SCALE`（整数）

小数位数（用于类型 `numeric`）

`TYPE`（整数）

列的数据类型的数字编码

在`EXECUTE`、`DECLARE`以及`OPEN`语句中，`INTO`和`USING`关键词的效果不同。
也可以手工建立一个描述符区域来为一个查询或者游标提供输入参数，并且`USING SQL DESCRIPTOR` *`name`*是用来传递输入参数给参数化查询的方法。建立一个命名 SQL 描述符区域的语句如下：

```
EXEC SQL SET DESCRIPTOR name VALUE num field = :hostvar;
```

PostgreSQL　支持在一条`FETCH`语句中检索多条记录。在这种情况下，要将数据存入主变量，该变量必须是数组。例如：

```
EXEC SQL BEGIN DECLARE SECTION;
int id[5];
EXEC SQL END DECLARE SECTION;

EXEC SQL FETCH 5 FROM mycursor INTO SQL DESCRIPTOR mydesc;

EXEC SQL GET DESCRIPTOR mydesc VALUE 1 :id = DATA;
```

33.7.2. SQLDA 描述符区域

SQLDA 描述符区域是一个 C 语言结构体，它也能被用来得到一个查询的结果集和元数据。一个结构体存储一个来自结果集的记录。

```
EXEC SQL include sqlda.h;
sqlda_t         *mysqlda;

EXEC SQL FETCH 3 FROM mycursor INTO DESCRIPTOR mysqlda;
```

注意，这里省略了SQL关键词。第 33.7.1 节中关于INTO和USING关键词用法的说明也适用于这里，并有一项补充规则：在DESCRIBE语句中，如果使用了INTO关键词，就可以完全省略DESCRIPTOR关键词：

```
EXEC SQL DESCRIBE prepared_statement INTO mysqlda;
```

使用 SQLDA 的程序的一般流程是：

1. 准备一个查询，并且为它声明一个游标。
2. 为结果行声明一个 SQLDA。
3. 为输入参数声明一个 SQLDA，并且初始化它们（内存分配、参数设置）。
4. 用输入 SQLDA 打开一个游标。
5. 从游标中取得行，并且把它们存储到一个输出 SQLDA。
6. 从输出 SQLDA 读取值到主变量中（必要时使用转换）。
7. 关闭游标。
8. 关闭为输入 SQLDA 分配的内存区域。

33.7.2.1. SQLDA 数据结构

SQLDA 使用三种数据结构类型：sqlda_t、sqlvar_t以及struct sqlname。

> **提示**
>
> PostgreSQL 的 SQLDA 数据结构与 IBM DB2 Universal Database 中的 SQLDA 相似，因此一些有关 DB2 SQLDA 的技术资料有助于更好地理解 PostgreSQL 的 SQLDA。

33.7.2.1.1. sqlda_t 结构体

结构体类型sqlda_t是实际 SQLDA 的类型。它保存一个记录。并且两个或者更多个sqlda_t结构体能够以desc_next字段中的指针连接成一个链表，这样可以表示一个有序的行集合。因此，当两个或多个行被取得时，应用可以通过沿着每一个sqlda_t节点中的desc_next指针读取它们。

sqlda_t的定义是：

```
struct sqlda_struct
{
    char            sqldaid[8];
    long            sqldabc;
    short           sqln;
    short           sqld;
    struct sqlda_struct *desc_next;
    struct sqlvar_struct sqlvar[1];
};

typedef struct sqlda_struct sqlda_t;
```

各字段的含义如下：

sqldaid

它包含一个字符串"SQLDA "。

sqldabc

它包含已分配空间的尺寸（以字节计）。

sqln

当它通过USING关键词传递给OPEN、DECLARE或EXECUTE语句时，它包含参数化查询的输入参数个数。如果它用作SELECT、EXECUTE或FETCH语句的输出，其值与sqld相同。

sqld

它包含一个结果集中的域的数量。

desc_next

如果查询返回不止一个记录，会返回多个链接在一起的 SQLDA 结构体，并且desc_next保存一个指向下一个项的指针。

sqlvar

这是结果集中列的数组。

33.7.2.1.2. sqlvar_t 结构体

结构体类型sqlvar_t保存一个列值和元数据（例如类型和长度）。该类型的定义是：

```
struct sqlvar_struct
{
    short          sqltype;
    short          sqllen;
    char          *sqldata;
    short         *sqlind;
    struct sqlname sqlname;
};

typedef struct sqlvar_struct sqlvar_t;
```

各字段的含义如下：

sqltype

包含该字段的类型标识符。值可以参考ecpgtype.h中的enum ECPGttype。

sqllen

包含字段的二进制长度，例如ECPGt_int是 4 字节。

sqldata

指向数据。数据的格式在第 33.4.4 节中描述。

`sqlind`

指向空值指示符。0 表示非空，-1 表示空。

`sqlname`

域的名称。

33.7.2.1.3. struct sqlname 结构体

一个`struct sqlname`结构体保存一个列名。它用作`sqlvar_t`结构体的成员。该结构体的定义如下：

```
#define NAMEDATALEN 64

struct sqlname
{
        short           length;
        char            data[NAMEDATALEN];
};
```

各字段的含义如下：

`length`

包含字段名称的长度。

`data`

包含实际的域名称。

33.7.2.2. 使用一个 SQLDA 检索一个结果集

通过一个 SQLDA 检索一个查询结果集的一般步骤是：

1. 声明一个`sqlda_t`结构体来接收结果集。
2. 执行 `FETCH/EXECUTE/DESCRIBE` 命令来处理一个指定已声明 SQLDA 的查询。
3. 通过查看`sqlda_t`结构体的成员`sqln`来检查结果集中记录的数量。
4. 从`sqlda_t`结构体的成员`sqlvar[0]`、`sqlvar[1]`等中得到每一列的值。
5. 沿着`sqlda_t`结构体的成员`desc_next`指针到达下一行（`sqlda_t`）。
6. 根据你的需要重复上述步骤。

这里是一个通过 SQLDA 检索结果集的例子。

首先，声明一个`sqlda_t`结构体来接收结果集。

```
sqlda_t *sqlda1;
```

接下来，在命令中指定 SQLDA。下面是一个`FETCH`命令的例子。

```
EXEC SQL FETCH NEXT FROM cur1 INTO DESCRIPTOR sqlda1;
```

运行一个循环顺着链表来检索行。

```
sqlda_t *cur_sqlda;

for (cur_sqlda = sqlda1;
     cur_sqlda != NULL;
     cur_sqlda = cur_sqlda->desc_next)
{
    ...
}
```

在循环内部，运行另一个循环来检索行中每一列的数据（`sqlvar_t`结构体）。

```
for (i = 0; i < cur_sqlda->sqld; i++)
{
    sqlvar_t v = cur_sqlda->sqlvar[i];
    char *sqldata = v.sqldata;
    short sqllen  = v.sqllen;
    ...
}
```

要得到一列的值，应检查`sqlvar_t`结构体的成员`sqltype`的值。然后，
根据列类型切换到一种合适的方法从`sqlvar`字段中复制数据到一个主变量。

```
char var_buf[1024];

switch (v.sqltype)
{
    case ECPGt_char:
        memset(&var_buf, 0, sizeof(var_buf));
        memcpy(&var_buf, sqldata, (sizeof(var_buf) <= sqllen ? sizeof(
var_buf) - 1 : sqllen));
        break;

    case ECPGt_int: /* integer */
        memcpy(&intval, sqldata, sqllen);
        snprintf(var_buf, sizeof(var_buf), "%d", intval);
        break;

    ...
}
```

33.7.2.3. 使用一个 SQLDA 传递查询参数

使用一个 SQLDA 传递输入参数给一个预备查询的一般步骤是：

1. 创建一个预备查询（预备语句）。
2. 声明一个 sqlda_t 结构，作为输入 SQLDA。
3. 为输入 SQLDA 分配内存区域（作为 sqlda_t 结构体）。
4. 在分配好的内存中设置（复制）输入值。
5. 打开一个说明了输入 SQLDA 的游标。

这里是一个例子。

首先，创建一个预备语句。

```
EXEC SQL BEGIN DECLARE SECTION;
char query[1024] = "SELECT d.oid, * FROM pg_database d, pg_stat_
database s WHERE d.oid = s.datid AND (d.datname = ? OR d.oid = ?)";
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE stmt1 FROM :query;
```

接下来为一个 SQLDA 分配内存，并且在`sqlda_t`结构体的`sqln`成员变量中设置输入参数的数量。当预备查询要求两个或多个输入参数时，应用必须分配额外的内存空间，空间的大小为 (参数数目 - 1) * sizeof(sqlvar_t)。这里的例子展示了为两个输入参数分配内存空间。

```
sqlda_t *sqlda2;

sqlda2 = (sqlda_t *) malloc(sizeof(sqlda_t) + sizeof(sqlvar_t));
memset(sqlda2, 0, sizeof(sqlda_t) + sizeof(sqlvar_t));

sqlda2->sqln = 2; /* 输入变量的数目 */
```

内存分配之后，把参数值存储到`sqlvar[]`数组（当 SQLDA 在接收结果集时，这也是用来检索列值的数组）。在这个例子中，输入参数是`"postgres"`（字符串类型）和`1`（整数类型）。

```
sqlda2->sqlvar[0].sqltype = ECPGt_char;
sqlda2->sqlvar[0].sqldata = "postgres";
sqlda2->sqlvar[0].sqllen  = 8;

int intval = 1;
sqlda2->sqlvar[1].sqltype = ECPGt_int;
sqlda2->sqlvar[1].sqldata = (char *) &intval;
sqlda2->sqlvar[1].sqllen  = sizeof(intval);
```

打开游标并指定事先设置好的 SQLDA，即可将输入参数传递给预备语句。

```
EXEC SQL OPEN cur1 USING DESCRIPTOR sqlda2;
```

最后，用完输入 SQLDA 后必须显式地释放已分配的内存空间，这与用于接收查询结果的 SQLDA 不同。

```
free(sqlda2);
```

33.7.2.4. 一个使用 SQLDA 的应用例子

这里是一个例子程序，它描述了如何按照输入参数的指定从系统目录中取得数据库的访问统计。

这个应用在数据库 OID 上连接两个系统表（pg_database 和 pg_stat_database），并且还取得和显示通过两个输入参数（一个数据库`postgres`和 OID `1`）检索到的数据库统计。

首先，为输入和输出分别声明一个 SQLDA。

```
EXEC SQL include sqlda.h;

sqlda_t *sqlda1; /* 一个输出描述符 */
```

```
sqlda_t *sqlda2; /* 一个输入描述符  */
```

接下来，连接到数据库，预备一个语句，并为该预备语句声明游标。

```
int
main(void)
{
    EXEC SQL BEGIN DECLARE SECTION;
    char query[1024] = "SELECT d.oid,* FROM pg_database d, pg_stat_
database s WHERE d.oid=s.datid AND ( d.datname=? OR d.oid=? )";
    EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO testdb AS con1 USER testuser;

    EXEC SQL PREPARE stmt1 FROM :query;
    EXEC SQL DECLARE cur1 CURSOR FOR stmt1;
```

然后，为输入参数在输入 SQLDA 中放入一些值。为输入 SQLDA 分配内存，并且在`sqln`中设置输入参数的数目。在`sqlvar`结构体的`sqltype`、`sqldata`和`sqllen`中存入类型、值和值长度。

```
    /* 为输入参数创建 SQLDA 结构。 */
    sqlda2 = (sqlda_t *) malloc(sizeof(sqlda_t) + sizeof(sqlvar_t));
    memset(sqlda2, 0, sizeof(sqlda_t) + sizeof(sqlvar_t));
    sqlda2->sqln = 2; /* 输入变量的数量 */

    sqlda2->sqlvar[0].sqltype = ECPGt_char;
    sqlda2->sqlvar[0].sqldata = "postgres";
    sqlda2->sqlvar[0].sqllen  = 8;

    intval = 1;
    sqlda2->sqlvar[1].sqltype = ECPGt_int;
    sqlda2->sqlvar[1].sqldata = (char *)&intval;
    sqlda2->sqlvar[1].sqllen  = sizeof(intval);
```

设置完输入 SQLDA 之后，用输入 SQLDA 打开一个游标。

```
    /* 用输入参数打开一个游标。 */
    EXEC SQL OPEN cur1 USING DESCRIPTOR sqlda2;
```

从打开的游标中取得行，并存入输出 SQLDA（通常，需要在循环中反复调用`FETCH`，才能取出结果集中的所有行）。

```
    while (1)
    {
        sqlda_t *cur_sqlda;

        /* 分配描述符给游标  */
        EXEC SQL FETCH NEXT FROM cur1 INTO DESCRIPTOR sqlda1;
```

接下来，沿着`sqlda_t`结构体的链表，从 SQLDA 中读取已取得的记录。

```
    for (cur_sqlda = sqlda1 ;
         cur_sqlda != NULL ;
         cur_sqlda = cur_sqlda->desc_next)
    {
        ...
```

读取第一个记录中的每一列。列的数量被存储在sqld中，第一列的实际数据被存储在sqlvar[0]中，两者都是sqlda_t结构体的成员。

```
        /* 打印一行中的每一列。 */
        for (i = 0; i < sqlda1->sqld; i++)
        {
            sqlvar_t v = sqlda1->sqlvar[i];
            char *sqldata = v.sqldata;
            short sqllen  = v.sqllen;

            strncpy(name_buf, v.sqlname.data, v.sqlname.length);
            name_buf[v.sqlname.length] = '\0';
```

现在，列数据已存储在变量v中。将每个数据项复制到主变量，并查看v.sqltype来确定列的类型。

```
            switch (v.sqltype) {
                int intval;
                double doubleval;
                unsigned long long int longlongval;

                case ECPGt_char:
                    memset(&var_buf, 0, sizeof(var_buf));
                    memcpy(&var_buf, sqldata, (sizeof(var_buf) <=
 sqllen ? sizeof(var_buf)-1 : sqllen));
                    break;

                case ECPGt_int: /* 整数 */
                    memcpy(&intval, sqldata, sqllen);
                    snprintf(var_buf, sizeof(var_buf), "%d", intval);
                    break;

                ...

                default:
                    ...
            }

            printf("%s = %s (type: %d)\n", name_buf, var_buf, v.
sqltype);
        }
```

处理所有记录后关闭游标，并且从数据库断开连接。

```
    EXEC SQL CLOSE cur1;
    EXEC SQL COMMIT;

    EXEC SQL DISCONNECT ALL;
```

整个程序显示在例 33.1中。

例 33.1. 示例 SQLDA 程序

```
#include <stdlib.h>
#include <string.h>
```

```
#include <stdlib.h>
#include <stdio.h>
#include <unistd.h>

EXEC SQL include sqlda.h;

sqlda_t *sqlda1; /* descriptor for output */
sqlda_t *sqlda2; /* descriptor for input */

EXEC SQL WHENEVER NOT FOUND DO BREAK;
EXEC SQL WHENEVER SQLERROR STOP;

int
main(void)
{
    EXEC SQL BEGIN DECLARE SECTION;
    char query[1024] = "SELECT d.oid,* FROM pg_database d, pg_stat_
database s WHERE d.oid=s.datid AND ( d.datname=? OR d.oid=? )";

    int intval;
    unsigned long long int longlongval;
    EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO uptimedb AS con1 USER uptime;

    EXEC SQL PREPARE stmt1 FROM :query;
    EXEC SQL DECLARE cur1 CURSOR FOR stmt1;

    /* Create a SQLDA structure for an input parameter */
    sqlda2 = (sqlda_t *)malloc(sizeof(sqlda_t) + sizeof(sqlvar_t));
    memset(sqlda2, 0, sizeof(sqlda_t) + sizeof(sqlvar_t));
    sqlda2->sqln = 2; /* a number of input variables */

    sqlda2->sqlvar[0].sqltype = ECPGt_char;
    sqlda2->sqlvar[0].sqldata = "postgres";
    sqlda2->sqlvar[0].sqllen  = 8;

    intval = 1;
    sqlda2->sqlvar[1].sqltype = ECPGt_int;
    sqlda2->sqlvar[1].sqldata = (char *) &intval;
    sqlda2->sqlvar[1].sqllen  = sizeof(intval);

    /* Open a cursor with input parameters. */
    EXEC SQL OPEN cur1 USING DESCRIPTOR sqlda2;

    while (1)
    {
        sqlda_t *cur_sqlda;

        /* Assign descriptor to the cursor  */
        EXEC SQL FETCH NEXT FROM cur1 INTO DESCRIPTOR sqlda1;

        for (cur_sqlda = sqlda1 ;
             cur_sqlda != NULL ;
```

```
             cur_sqlda = cur_sqlda->desc_next)
        {
            int i;
            char name_buf[1024];
            char var_buf[1024];

            /* Print every column in a row. */
            for (i=0 ; i<cur_sqlda->sqld ; i++)
            {
                sqlvar_t v = cur_sqlda->sqlvar[i];
                char *sqldata = v.sqldata;
                short sqllen  = v.sqllen;

                strncpy(name_buf, v.sqlname.data, v.sqlname.length);
                name_buf[v.sqlname.length] = '\0';

                switch (v.sqltype)
                {
                    case ECPGt_char:
                        memset(&var_buf, 0, sizeof(var_buf));
                        memcpy(&var_buf, sqldata, (sizeof(var_buf)<=
sqllen ? sizeof(var_buf)-1 : sqllen) );
                        break;

                    case ECPGt_int: /* integer */
                        memcpy(&intval, sqldata, sqllen);
                        snprintf(var_buf, sizeof(var_buf), "%d",
 intval);
                        break;

                    case ECPGt_long_long: /* bigint */
                        memcpy(&longlongval, sqldata, sqllen);
                        snprintf(var_buf, sizeof(var_buf), "%lld",
 longlongval);
                        break;

                    default:
                    {
                        int i;
                        memset(var_buf, 0, sizeof(var_buf));
                        for (i = 0; i < sqllen; i++)
                        {
                            char tmpbuf[16];
                            snprintf(tmpbuf, sizeof(tmpbuf), "%02x ",
 (unsigned char) sqldata[i]);
                            strncat(var_buf, tmpbuf, sizeof(var_buf));
                        }
                    }
                        break;
                }

                printf("%s = %s (type: %d)\n", name_buf, var_buf, v.
sqltype);
            }
```

```
            printf("\n");
        }
    }

    EXEC SQL CLOSE cur1;
    EXEC SQL COMMIT;

    EXEC SQL DISCONNECT ALL;

    return 0;
}
```

这个例子的输出应该看起来类似下面的结果（一些数字会变化）。

```
oid = 1 (type: 1)
datname = template1 (type: 1)
datdba = 10 (type: 1)
encoding = 0 (type: 5)
datistemplate = t (type: 1)
datallowconn = t (type: 1)
datconnlimit = -1 (type: 5)
datlastsysoid = 11510 (type: 1)
datfrozenxid = 379 (type: 1)
dattablespace = 1663 (type: 1)
datconfig =  (type: 1)
datacl = {=c/uptime,uptime=CTc/uptime} (type: 1)
datid = 1 (type: 1)
datname = template1 (type: 1)
numbackends = 0 (type: 5)
xact_commit = 113606 (type: 9)
xact_rollback = 0 (type: 9)
blks_read = 130 (type: 9)
blks_hit = 7341714 (type: 9)
tup_returned = 38262679 (type: 9)
tup_fetched = 1836281 (type: 9)
tup_inserted = 0 (type: 9)
tup_updated = 0 (type: 9)
tup_deleted = 0 (type: 9)

oid = 11511 (type: 1)
datname = postgres (type: 1)
datdba = 10 (type: 1)
encoding = 0 (type: 5)
datistemplate = f (type: 1)
datallowconn = t (type: 1)
datconnlimit = -1 (type: 5)
datlastsysoid = 11510 (type: 1)
datfrozenxid = 379 (type: 1)
dattablespace = 1663 (type: 1)
datconfig =  (type: 1)
datacl =  (type: 1)
datid = 11511 (type: 1)
datname = postgres (type: 1)
```

```
numbackends = 0 (type: 5)
xact_commit = 221069 (type: 9)
xact_rollback = 18 (type: 9)
blks_read = 1176 (type: 9)
blks_hit = 13943750 (type: 9)
tup_returned = 77410091 (type: 9)
tup_fetched = 3253694 (type: 9)
tup_inserted = 0 (type: 9)
tup_updated = 0 (type: 9)
tup_deleted = 0 (type: 9)
```

33.8. 错误处理

这一节描述如何在嵌入式 SQL 程序中处理异常情况和警告。为此提供了两种可以同时使用的机制。

- 可以使用`WHENEVER`命令配置回调来处理警告和错误情况。
- 可以从`sqlca`变量中获得错误或警告的详细信息。

33.8.1. 设置回调

捕获错误和警告的一种简单方法是指定一个动作，在特定条件发生时执行。一般形式如下：

EXEC SQL WHENEVER *condition action*;

*condition*可以是以下值之一：

`SQLERROR`

只要在 SQL 语句执行期间发生一个错误就调用指定的动作。

`SQLWARNING`

只要在 SQL 语句执行期间发生一个警告就调用指定的动作。

`NOT FOUND`

只要一个 SQL 语句检索或者影响零行就调用指定的动作（这种情况不是一个错误，但是你可能需要特别地处理它）。

*action*可以是以下值之一：

`CONTINUE`

这实际上表示该情况被忽略。这是默认值。

GOTO *label*
GO TO *label*

跳到指定的标签（使用一个 C `goto`语句）。

`SQLPRINT`

把一个消息打印到标准错误。这对简单程序或原型开发很有用。消息的细节无法配置。

`STOP`

调用`exit(1)`终止程序。

`DO BREAK`

执行 C 语句`break`。只应被用在循环或`switch`语句中。

`CALL` *`name`* `(`*`args`*`)`
`DO` *`name`* `(`*`args`*`)`

使用指定参数调用指定的 C 函数。

SQL 标准只规定了以下动作：`CONTINUE` 和 `GOTO`（以及 `GO TO`）。

这里有一个可能会用在简单程序中的例子。当一个警告发生时它打印一个简单消息，而发生一个错误时它会中止程序：

```
EXEC SQL WHENEVER SQLWARNING SQLPRINT;
EXEC SQL WHENEVER SQLERROR STOP;
```

语句`EXEC SQL WHENEVER`是 SQL 预处理器的指令，而不是 C 语句。它设置的错误或警告动作适用于源代码中出现在处理器设置位置之后的所有嵌入式 SQL 语句，除非在第一个`EXEC SQL WHENEVER`与触发该条件的 SQL 语句之间，为同一条件设置了其他动作。这与 C 程序的控制流无关。因此，下面两个 C 程序片段都不会产生预期效果：

```
/*
 * 错误
 */
int main(int argc, char *argv[])
{
    ...
    if (verbose) {
        EXEC SQL WHENEVER SQLWARNING SQLPRINT;
    }
    ...
    EXEC SQL SELECT ...;
    ...
}

/*
 * 错误
 */
int main(int argc, char *argv[])
{
    ...
    set_error_handler();
    ...
    EXEC SQL SELECT ...;
    ...
}

static void set_error_handler(void)
{
    EXEC SQL WHENEVER SQLERROR STOP;
}
```

33.8.2. sqlca

为了支持更强大的错误处理，嵌入式 SQL 接口提供了一个名为sqlca（SQL 通信区域）的全局变量，其结构如下：

```
struct
{
    char sqlcaid[8];
    long sqlabc;
    long sqlcode;
    struct
    {
        int sqlerrml;
        char sqlerrmc[SQLERRMC_LEN];
    } sqlerrm;
    char sqlerrp[8];
    long sqlerrd[6];
    char sqlwarn[8];
    char sqlstate[5];
} sqlca;
```

（在一个多线程程序中，每一个线程会自动得到它自己的sqlca副本。这和对于标准 C 全局变量errno的处理相似。）

sqlca覆盖了警告和错误。如果执行一个语句时发生了多个警告或错误，那么sqlca将只包含关于最后一个的信息。

如果在上一个SQL语句中没有产生错误，sqlca.sqlcode将为 0 并且sqlca.sqlstate将为"00000"。如果发生一个警告或错误，则sqlca.sqlcode将为负并且sqlca.sqlstate将不为"00000"。一个正的sqlca.sqlcode表示一种无害的情况，例如上一个查询返回零行。sqlcode和sqlstate是两种不同的错误编码方案，详见下文。

如果上一个 SQL 语句成功，则sqlca.sqlerrd[1]包含所处理行的 OID（如果适用），sqlca.sqlerrd[2]包含处理或返回的行数（如果适用于该命令）。

发生错误或警告时，sqlca.sqlerrm.sqlerrmc包含描述该错误的字符串。字段sqlca.sqlerrm.sqlerrml包含存储在sqlca.sqlerrm.sqlerrmc中的错误消息的长度（即strlen()的结果，对于 C 程序员通常没有多少用处）。注意，有些消息太长，无法存入定长的sqlerrmc数组，因此会被截断。

发生警告时，sqlca.sqlwarn[2]被设为W（其他情况下，它被设为非W的值）。如果sqlca.sqlwarn[1]被设为W，说明某个值在存入主变量时被截断了。如果其他任一元素被设置为指示警告，sqlca.sqlwarn[0]就会被设为W。

字段sqlcaid、sqlabc、sqlerrp，以及sqlerrd和sqlwarn的其余元素，目前不包含有用信息。

SQL 标准中没有定义sqlca结构体，但是在一些其他的 SQL 数据库系统中都有实现。在核心上这些定义都相似，但是如果你想要编写可移植的应用，那么你应该仔细研究不同的实现。

这里有一个整合使用WHENEVER和sqlca的例子，当一个错误发生时打印出sqlca的内容。在安装一个更“用户友好”的错误处理器之前，这可能对调试或开发原型应用有用。

```
EXEC SQL WHENEVER SQLERROR CALL print_sqlca();
```

```
void
print_sqlca()
{
    fprintf(stderr, "==== sqlca ====\n");
    fprintf(stderr, "sqlcode: %ld\n", sqlca.sqlcode);
    fprintf(stderr, "sqlerrm.sqlerrml: %d\n", sqlca.sqlerrm.sqlerrml);
    fprintf(stderr, "sqlerrm.sqlerrmc: %s\n", sqlca.sqlerrm.sqlerrmc);
    fprintf(stderr, "sqlerrd: %ld %ld %ld %ld %ld %ld\n", sqlca.
sqlerrd[0],sqlca.sqlerrd[1],sqlca.sqlerrd[2],
                                                          sqlca.
sqlerrd[3],sqlca.sqlerrd[4],sqlca.sqlerrd[5]);
    fprintf(stderr, "sqlwarn: %d %d %d %d %d %d %d %d\n", sqlca.
sqlwarn[0], sqlca.sqlwarn[1], sqlca.sqlwarn[2],
                                                          sqlca.
sqlwarn[3], sqlca.sqlwarn[4], sqlca.sqlwarn[5],
                                                          sqlca.
sqlwarn[6], sqlca.sqlwarn[7]);
    fprintf(stderr, "sqlstate: %5s\n", sqlca.sqlstate);
    fprintf(stderr, "===============\n");
}
```

结果看起来像（这里的错误是一个拼写错误的表名）：

```
==== sqlca ====
sqlcode: -400
sqlerrm.sqlerrml: 49
sqlerrm.sqlerrmc: relation "pg_databasep" does not exist on line 38
sqlerrd: 0 0 0 0 0 0
sqlwarn: 0 0 0 0 0 0 0 0
sqlstate: 42P01
===============
```

33.8.3. SQLSTATE 与 SQLCODE

字段`sqlca.sqlstate`和`sqlca.sqlcode`采用两种不同的错误编码方案。两者都源自 SQL 标准，但`SQLCODE`在 SQL-92 标准中已被标记为弃用，并在后续版本中被删除。因此，强烈建议新应用使用`SQLSTATE`。

`SQLSTATE`是一个包含五个字符的数组。这五个字符由数字或大写字母组成，表示各种错误和警告条件的代码。`SQLSTATE`采用分层编码方案：前两个字符表示条件所属的大类，后三个字符表示该大类中的子类。代码`00000`表示成功状态。大部分`SQLSTATE`代码由 SQL 标准定义。PostgreSQL服务器原生支持`SQLSTATE`错误码，因此在所有应用中统一使用这种错误编码方案，可以实现较高的一致性。更多信息见附录 A。

已弃用的错误编码方案`SQLCODE`使用一个简单的整数。0 表示成功，正值表示成功并有附加信息，负值表示错误。SQL 标准只定义了正值 +100，表示上一个命令返回或影响了零行，没有定义具体的负值。因此，这种方案的可移植性较差，也不采用分层编码。历史上，PostgreSQL的嵌入式 SQL 处理器为自身使用分配了一些特定的`SQLCODE`值，下文列出了它们的数值和符号名称。注意，这些值不能移植到其他 SQL 实现。为了便于将应用迁移到`SQLSTATE`方案，还列出了对应的`SQLSTATE`。不过，两种方案之间不存在一对一或一对多的映射关系（实际上是多对多），因此每次都应查阅附录 A中的完整`SQLSTATE`列表。

以下是已分配的SQLCODE值：

0 (ECPG_NO_ERROR)

表示没有错误（SQLSTATE 00000）。

100 (ECPG_NOT_FOUND)

这是一种无害情况，它表示上一个命令检索或者处理了零行，
或者你到达了游标的末尾（SQLSTATE 02000）。

在一个循环中处理一个游标时，你可以使用这个代码作为一种方法来检测何时中止该循环，
像这样：

```
while (1)
{
    EXEC SQL FETCH ... ;
    if (sqlca.sqlcode == ECPG_NOT_FOUND)
        break;
}
```

但是WHENEVER NOT FOUND DO BREAK实际上会在内部这样做，因此显式地把它写出来通常没有什么好处。

-12 (ECPG_OUT_OF_MEMORY)

表示你的虚拟内存已被耗尽。数字值被定义为-ENOMEM（SQLSTATE YE001）。

-200 (ECPG_UNSUPPORTED)

表示预处理器生成了该库无法识别的内容。
可能是所用的预处理器与库的版本不兼容（SQLSTATE YE002）。

-201 (ECPG_TOO_MANY_ARGUMENTS)

这表示命令指定了超过该命令预期数量的主变量（SQLSTATE 07001 或 07002）。

-202 (ECPG_TOO_FEW_ARGUMENTS)

这表示命令指定的主变量数量低于该命令的预期（SQLSTATE 07001 或 07002）。

-203 (ECPG_TOO_MANY_MATCHES)

这意味着一个查询已经返回了多个行，但是该语句只准备存储一个结果行（例如，
因为指定的变量不是数组）（SQLSTATE 21000）。

-204 (ECPG_INT_FORMAT)

主变量是类型int而数据库中的数据是一种不同的类型并且含有一个不能被解释为int的值。
该库使用strtol()进行这种转换（SQLSTATE 42804）。

-205 (ECPG_UINT_FORMAT)

主变量是类型unsigned int而数据库中的数据是一种不同的类型并且含有一个不能被解释为unsigned int的值。该库使用strtoul()进行这种转换（SQLSTATE 42804）。

-206 (ECPG_FLOAT_FORMAT)

主变量是类型float而数据库中的数据是另一种类型并且含有一个不能被解释为float的值。该库使用strtod()进行这种转换（SQLSTATE 42804）。

-207 (ECPG_NUMERIC_FORMAT)

主变量是类型numeric而数据库中的数据是另一种类型并且含有一个不能被解释为numeric的值（SQLSTATE 42804）。

-208 (ECPG_INTERVAL_FORMAT)

主变量是类型interval
而数据库中的数据是另一种类型并且含有一个不能被解释为interval的值（SQLSTATE 42804）。

-209 (ECPG_DATE_FORMAT)

主变量是类型date而数据库中的数据是另一种类型并且含有一个不能被解释为date的值（SQLSTATE 42804）。

-210 (ECPG_TIMESTAMP_FORMAT)

主变量是类型timestamp
而数据库中的数据是另一种类型并且含有一个不能被解释为timestamp的值（SQLSTATE 42804）。

-211 (ECPG_CONVERT_BOOL)

这表示主变量是类型bool而数据库中的数据既不是't'也不是'f'（SQLSTATE 42804）。

-212 (ECPG_EMPTY)

发送给PostgreSQL服务器的语句是空的（通常在一个嵌入式 SQL 程序中不会发生，因此它可能指向一个内部错误）（SQLSTATE YE002）。

-213 (ECPG_MISSING_INDICATOR)

返回了一个空值，但未提供空值指示符变量（SQLSTATE 22002）。

-214 (ECPG_NO_ARRAY)

在要求一个数组的地方使用了一个普通变量（SQLSTATE 42804）。

-215 (ECPG_DATA_NOT_ARRAY)

在一个要求数组值的地方数据库返回了一个普通变量（SQLSTATE 42804）。

-220 (ECPG_NO_CONN)

程序尝试访问一个不存在的连接（SQLSTATE 08003）。

-221 (ECPG_NOT_CONN)

程序尝试访问一个存在的连接但是它没有打开（这是一个内部错误）（SQLSTATE YE002）。

-230 (ECPG_INVALID_STMT)

尝试使用的语句尚未预备（SQLSTATE 26000）。

-239 (`ECPG_INFORMIX_DUPLICATE_KEY`)

重复键错误，违背唯一约束（Informix 兼容模式）（SQLSTATE 23505）。

-240 (`ECPG_UNKNOWN_DESCRIPTOR`)

未找到指定的描述符。尝试使用的语句尚未预备（SQLSTATE 33000）。

-241 (`ECPG_INVALID_DESCRIPTOR_INDEX`)

指定的描述符超出范围（SQLSTATE 07009）。

-242 (`ECPG_UNKNOWN_DESCRIPTOR_ITEM`)

请求了一个无效的描述符条目（这是一个内部错误）（SQLSTATE YE002）。

-243 (`ECPG_VAR_NOT_NUMERIC`)

执行动态语句时，数据库返回了数值，而主变量不是数值类型（SQLSTATE 07006）。

-244 (`ECPG_VAR_NOT_CHAR`)

执行动态语句时，数据库返回了非数值的值，而主变量是数值类型（SQLSTATE 07006）。

-284 (`ECPG_INFORMIX_SUBSELECT_NOT_ONE`)

子查询的结果不是单一行（Informix 兼容模式）（SQLSTATE 21000）。

-400 (`ECPG_PGSQL`)

PostgreSQL服务器导致了某个错误。该消息包含来自PostgreSQL服务器的错误消息。

-401 (`ECPG_TRANS`)

PostgreSQL服务器通知我们不能启动、提交或回滚事务（SQLSTATE 08007）。

-402 (`ECPG_CONNECT`)

到数据库的连接尝试没有成功（SQLSTATE 08001）。

-403 (`ECPG_DUPLICATE_KEY`)

重复键错误，违背唯一约束（SQLSTATE 23505）。

-404 (`ECPG_SUBSELECT_NOT_ONE`)

子查询的结果不是单一行（SQLSTATE 21000）。

-602 (`ECPG_WARNING_UNKNOWN_PORTAL`)

指定了一个非法的游标名（SQLSTATE 34000）。

-603 (`ECPG_WARNING_IN_TRANSACTION`)

事务正在进行（SQLSTATE 25001）。

-604 (`ECPG_WARNING_NO_TRANSACTION`)

没有活动（正在进行）的事务（SQLSTATE 25P01）。

-605 (`ECPG_WARNING_PORTAL_EXISTS`)

指定了一个现有的游标名（SQLSTATE 42P03）。

33.9. 预处理器指令

一些预处理器指令可以用来改变`ecpg`预处理器解析和处理一个文件的方式。

33.9.1. 包括文件

要在嵌入式 SQL 程序中包含外部文件，可使用：

```
EXEC SQL INCLUDE filename;
EXEC SQL INCLUDE <filename>;
EXEC SQL INCLUDE "filename";
```

嵌入式 SQL 预处理器会查找名为*filename*`.h`的文件，对它进行预处理，并将其包含在生成的 C 输出中。这样，被包含文件中的嵌入式 SQL 语句就能得到正确处理。

`ecpg`预处理器将以下列顺序在几个目录中搜索一个文件：

- 当前目录
- `/usr/local/include`
- PostgreSQL 的头文件目录，在编译时定义（例如`/usr/local/pgsql/include`）
- `/usr/include`

但是当使用`EXEC SQL INCLUDE "`*filename*`"`时，只有当前目录会被搜索。

在每一个目录中，预处理器将首先按给定的文件名搜索，如果没有找到将会追加`.h`到文件名并且重试（除非指定的文件名已经具有该后缀）。

注意`EXEC SQL INCLUDE`不同于：

```
#include <filename.h>
```

因为这样包含的文件不会经过 SQL 命令预处理。当然，仍可使用 C 的`#include`指令包含其他头文件。

> **注意**
>
> 被包含文件的名称区分大小写，尽管`EXEC SQL INCLUDE`命令的其他部分遵循通常的 SQL 大小写规则。

33.9.2. define 和 undef 指令

与 C 中我们熟知的指令`#define`相似，嵌入式 SQL 具有类似的概念：

```
EXEC SQL DEFINE name;
```

```
EXEC SQL DEFINE name value;
```

因此你可以定义一个名称：

```
EXEC SQL DEFINE HAVE_FEATURE;
```

并且你也可以定义常量：

```
EXEC SQL DEFINE MYNUMBER 12;
EXEC SQL DEFINE MYSTRING 'abc';
```

使用`undef`来移除一个之前的定义：

```
EXEC SQL UNDEF MYNUMBER;
```

当然在你的嵌入式 SQL 程序中你可以继续使用 C 版本的`#define`和`#undef`。区别在于你定义的值会在哪里被计算。如果你使用`EXEC SQL DEFINE`，那么`ecpg`预处理器会计算这些定义并且替换值。例如，如果你写：

```
EXEC SQL DEFINE MYNUMBER 12;
...
EXEC SQL UPDATE Tbl SET col = MYNUMBER;
```

那么`ecpg`将已经做过替换并且你的 C 编译器将永远不会看见名为`MYNUMBER`的任何名称或标识符。注意你不能把`#define`用于一个将要在一个嵌入式 SQL 查询中使用的常量，因为在这种情况下嵌入式 SQL 预处理器不能看到这个声明。

33.9.3. ifdef、ifndef、else、elif 和 endif 指令

可以使用以下指令对代码区段进行条件编译：

```
EXEC SQL ifdef name;
```

检查 *name*，如果已通过 `EXEC SQL define name` 创建了 *name*，就处理后续各行。

```
EXEC SQL ifndef name;
```

检查 *name*，如果尚未通过 `EXEC SQL define name` 创建 *name*，就处理后续各行。

```
EXEC SQL else;
```

开始处理一个备选区段，对应由 `EXEC SQL ifdef name` 或 `EXEC SQL ifndef name` 引入的区段。

```
EXEC SQL elif name;
```

检查 *name*，如果已通过 `EXEC SQL define name` 创建了 *name*，就开始一个备选区段。

```
EXEC SQL endif;
```

结束备选区段。

示例：

```
EXEC SQL ifndef TZVAR;
EXEC SQL SET TIMEZONE TO 'GMT';
EXEC SQL elif TZNAME;
```

```
EXEC SQL SET TIMEZONE TO TZNAME;
EXEC SQL else;
EXEC SQL SET TIMEZONE TO TZVAR;
EXEC SQL endif;
```

33.10. 处理嵌入式 SQL 程序

现在你已经对如何构造嵌入式 SQL C 程序有所了解了，你可能希望知道如何编译它们。在编译之前，你需要让该文件通过嵌入式SQL C预处理器，它会把你用到的SQL转换成特殊的函数调用。在编译之后，你必须链接一个包含所需函数的特殊库。这些函数从参数中取得信息、使用libpq执行SQL命令并且把结果放在指定的参数中用来输出。

该预处理器程序被称作`ecpg`并且被包括在一个正常的PostgreSQL安装中。嵌入式 SQL 程序通常带有扩展名`.pgc`。如果你有一个程序文件`prog1.pgc`，
你可以调用下面的命令对它进行预处理：

```
ecpg prog1.pgc
```

这将创建一个文件`prog1.c`。如果你的输入文件不遵循建议的命名模式，你可以用`-o`选项显式地指定输出文件。

预处理过的文件可以被正常地编译，例如：

```
cc -c prog1.c
```

生成的 C 源文件会包含PostgreSQL安装中的头文件。因此，如果PostgreSQL安装在默认搜索路径之外，就必须在编译命令行中添加类似`-I/usr/local/pgsql/include`的选项。

要链接嵌入式 SQL 程序，需要链接`libecpg`库，例如：

```
cc -o myprog prog1.o prog2.o ... -lecpg
```

同样，可能需要在该命令行中添加类似`-L/usr/local/pgsql/lib`的选项。

如果你使用make来管理一个大工程的构建过程，把下面的隐式规则包括在你的 makefile 中将会很方便：

```
ECPG = ecpg

%.c: %.pgc
        $(ECPG) $<
```

`ecpg`命令的完整语法可见ecpg。

ecpg库默认是线程安全的。不过，你可能需要使用一些线程命令行选项来编译你的客户端代码。

33.11. 库函数

`libecpg`库主要包含用于实现嵌入式 SQL 命令功能的“隐藏”函数，也有一些适合直接调用的函数。注意，直接调用这些函数会使代码失去可移植性。

- 调用`ECPGdebug(int on, FILE *stream)`时，如果第一个参数非零，就会启用调试日志。日志写入*stream*，包含代入全部输入变量后的所有SQL语句，以及PostgreSQL服务器返回的结果。这对于查找SQL语句中的错误很有用。

> **注意**
>
> 在 Windows 上，如果ecpg库和应用是使用不同标志编译的，这个函数调用将会使应用崩溃，因为FILE指针的内部表示不同。特别地，
> 库和使用库的应用应该使用相同的多线程/单线程、发行/调试以及静态/动态标志。

- ECPGget_PGconn(const char *connection_name) 返回由给定名称标识的库数据库连接句柄。如果connection_name被设置为NULL，当前连接句柄将被返回。
  如果无法定位到连接句柄，该函数返回NULL。如果需要，
  返回的连接句柄可以被用来调用任何其他来自于libpq的函数。

> **注意**
>
> 直接使用libpq例程来操纵ecpg中建立的数据库连接句柄是一种糟糕的做法。

- ECPGtransactionStatus(const char *connection_name) 返回由 connection_name 标识的连接的当前事务状态。返回状态码的详细信息，参见第 31.2 节和 libpq 的 PQtransactionStatus()。
- 如果你连接到了一个数据库，ECPGstatus(int lineno, const char* connection_name)会返回真；否则返回假。如果使用的是一个单一连接，connection_name可以为NULL。

33.12. 大对象

ECPG 并不直接支持大对象，在调用ECPGget_PGconn()函数获得所需的PGconn对象后，ECPG 应用能通过 libpq 大对象函数操纵大对象（不过，对ECPGget_PGconn()函数的使用以及直接接触PGconn对象都必须非常小心，并且最好不要与其他 ECPG 数据库访问调用混合在一起）。

更多关于ECPGget_PGconn()的细节可见第 33.11 节。大对象函数接口的相关信息可见第 32 章。

大对象函数必须在一个事务块中被调用，因此当自动提交关闭时，必须显式地发出BEGIN命令。

例 33.2给出了一个例子程序，它展示了在一个 ECPG 应用中如何创建、写入和读取一个大对象。

例 33.2. 访问大对象的 ECPG 程序

```
#include <stdio.h>
#include <stdlib.h>
#include <libpq-fe.h>
#include <libpq/libpq-fs.h>

EXEC SQL WHENEVER SQLERROR STOP;

int
main(void)
{
    PGconn     *conn;
```

```
Oid         loid;
int         fd;
char        buf[256];
int         buflen = 256;
char        buf2[256];
int         rc;

memset(buf, 1, buflen);

EXEC SQL CONNECT TO testdb AS con1;

conn = ECPGget_PGconn("con1");
printf("conn = %p\n", conn);

/* create */
loid = lo_create(conn, 0);
if (loid &lt; 0)
    printf("lo_create() failed: %s", PQerrorMessage(conn));

printf("loid = %d\n", loid);

/* write test */
fd = lo_open(conn, loid, INV_READ|INV_WRITE);
if (fd &lt; 0)
    printf("lo_open() failed: %s", PQerrorMessage(conn));

printf("fd = %d\n", fd);

rc = lo_write(conn, fd, buf, buflen);
if (rc &lt; 0)
    printf("lo_write() failed\n");

rc = lo_close(conn, fd);
if (rc &lt; 0)
    printf("lo_close() failed: %s", PQerrorMessage(conn));

/* read test */
fd = lo_open(conn, loid, INV_READ);
if (fd &lt; 0)
    printf("lo_open() failed: %s", PQerrorMessage(conn));

printf("fd = %d\n", fd);

rc = lo_read(conn, fd, buf2, buflen);
if (rc &lt; 0)
    printf("lo_read() failed\n");

rc = lo_close(conn, fd);
if (rc &lt; 0)
    printf("lo_close() failed: %s", PQerrorMessage(conn));

/* check */
rc = memcmp(buf, buf2, buflen);
printf("memcmp() = %d\n", rc);
```

```
    /* cleanup */
    rc = lo_unlink(conn, loid);
    if (rc &lt; 0)
        printf("lo_unlink() failed: %s", PQerrorMessage(conn));

    EXEC SQL COMMIT;
    EXEC SQL DISCONNECT ALL;
    return 0;
}
```

33.13. C++ 应用

ECPG 对 C++ 应用提供有限支持。本节介绍一些注意事项。

`ecpg` 预处理器接受一个用 C（或类似 C 的语言）和嵌入式 SQL 命令编写的输入文件，将嵌入式 SQL 命令转换为 C 代码片段，最终生成 `.c` 文件。在 C++ 中使用时，`ecpg` 生成的 C 代码片段所调用的库函数，其头文件声明会包在 `extern "C" { ... }` 块中，因此应该能在 C++ 中无缝工作。

不过，通常`ecpg`预处理器只理解 C，它无法处理 C++ 语言的特殊语法和保留词。因此，一些写在 C++ 应用代码中的使用了 C++ 特定复杂特性的嵌入式 SQL 代码可能无法被正确地预处理或者无法按预期工作。

使用 C++ 应用中嵌入式 SQL 代码的安全方法是把 ECPG 调用隐藏在一个 C 模块中，C++ 应用代码会调用它来访问数据库，还要把它和剩余的 C++ 代码链接起来。详见第 33.13.2 节。

33.13.1. 主变量的可见范围

`ecpg`预处理器理解 C 中变量的作用域。在 C 语言中，这比较简单，因为变量的作用域由所在代码块决定。但在 C++ 中，类成员变量可能在与声明位置不同的代码块中被引用，因此`ecpg`预处理器无法理解类成员变量的作用域。

例如，在下面的情况中，`ecpg`预处理器无法为`test`方法中的变量`dbname`找到任何声明，因此将发生一个错误。

```
class TestCpp
{
    EXEC SQL BEGIN DECLARE SECTION;
    char dbname[1024];
    EXEC SQL END DECLARE SECTION;

  public:
    TestCpp();
    void test();
    ~TestCpp();
};

TestCpp::TestCpp()
{
    EXEC SQL CONNECT TO testdb1;
}

void Test::test()
```

```
{
    EXEC SQL SELECT current_database() INTO :dbname;
    printf("current_database = %s\n", dbname);
}

TestCpp::~TestCpp()
{
    EXEC SQL DISCONNECT ALL;
}
```

这段代码将导致一个这样的错误：

```
ecpg test_cpp.pgc
test_cpp.pgc:28: ERROR: variable "dbname" is not declared
```

为避免这个作用域问题，可以修改`test`方法，使用局部变量作为中间存储。不过，这只是一个较差的变通办法，因为它使代码变得难看，还会降低性能。

```
void TestCpp::test()
{
    EXEC SQL BEGIN DECLARE SECTION;
    char tmp[1024];
    EXEC SQL END DECLARE SECTION;

    EXEC SQL SELECT current_database() INTO :tmp;
    strlcpy(dbname, tmp, sizeof(tmp));

    printf("current_database = %s\n", dbname);
}
```

33.13.2. 使用外部 C 模块的 C++ 应用开发

了解`ecpg`预处理器在 C++ 中的这些技术限制后，可能会认为：在链接阶段将 C 和 C++ 的目标文件链接起来，使 C++ 应用能够使用 ECPG 功能，会比直接在 C++ 代码中编写嵌入式 SQL 命令更好。本节用一个简单示例介绍如何将嵌入式 SQL 命令与 C++ 应用代码分离。在这个示例中，应用用 C++ 实现，连接 PostgreSQL 服务器的部分则使用 C 和 ECPG。

需要创建三种文件：C 文件（`*.pgc`）、头文件和 C++ 文件：

`test_mod.pgc`

一个执行嵌入在 C 中的 SQL 命令的子例程模块。它将被预处理器转换成`test_mod.c`。

```
#include "test_mod.h"
#include <stdio.h>

void
db_connect()
{
    EXEC SQL CONNECT TO testdb1;
}

void
db_test()
{
```

```
    EXEC SQL BEGIN DECLARE SECTION;
    char dbname[1024];
    EXEC SQL END DECLARE SECTION;

    EXEC SQL SELECT current_database() INTO :dbname;
    printf("current_database = %s\n", dbname);
}

void
db_disconnect()
{
    EXEC SQL DISCONNECT ALL;
}
```

test_mod.h

包含 C 模块（test_mod.pgc）中函数声明的头文件，由test_cpp.cpp包含。由于需要从 C++ 模块链接这些函数，声明必须包在extern "C"块中。

```
#ifdef __cplusplus
extern "C" {
#endif

void db_connect();
void db_test();
void db_disconnect();

#ifdef __cplusplus
}
#endif
```

test_cpp.cpp

应用的主要代码，包括main例程，在本示例中还包括一个 C++ 类。

```
#include "test_mod.h"

class TestCpp
{
  public:
    TestCpp();
    void test();
    ~TestCpp();
};

TestCpp::TestCpp()
{
    db_connect();
}

void
TestCpp::test()
{
    db_test();
}
```

```
TestCpp::~TestCpp()
{
    db_disconnect();
}

int
main(void)
{
    TestCpp *t = new TestCpp();

    t->test();
    return 0;
}
```

要构建该应用，按以下步骤处理。通过运行ecpg将test_mod.pgc转换为test_mod.c，并且用 C 编译器将test_mod.c编译成test_mod.o：

```
ecpg -o test_mod.c test_mod.pgc
cc -c test_mod.c -o test_mod.o
```

接着，用 C++ 编译器把test_cpp.cpp编译成test_cpp.o：

```
c++ -c test_cpp.cpp -o test_cpp.o
```

最后，通过 C++ 编译器驱动程序，将test_cpp.o和test_mod.o这两个目标文件链接成一个可执行文件：

```
c++ test_cpp.o test_mod.o -lecpg -o test_cpp
```

33.14. 嵌入式 SQL 命令

本节介绍专用于嵌入式 SQL 的全部 SQL 命令。另请参见SQL 命令中列出的 SQL 命令，除非另有说明，它们也可用于嵌入式 SQL。

ALLOCATE DESCRIPTOR

ALLOCATE DESCRIPTOR — allocate an SQL descriptor area

大纲

`ALLOCATE DESCRIPTOR name`

描述

`ALLOCATE DESCRIPTOR`分配一个新的命名 SQL 描述符区域，可用于在 PostgreSQL 服务器与主程序之间交换数据。

描述符区域使用后应通过`DEALLOCATE DESCRIPTOR`命令释放。

参数

name

SQL 描述符的名称，大小写敏感。这可以是一个 SQL 标识符或者一个主变量。

例子

```
EXEC SQL ALLOCATE DESCRIPTOR mydesc;
```

兼容性

SQL 标准中说明了`ALLOCATE DESCRIPTOR`。

参见

DEALLOCATE DESCRIPTOR, GET DESCRIPTOR, SET DESCRIPTOR

CONNECT

CONNECT — establish a database connection

大纲

```
CONNECT TO connection_target [ AS connection_name ] [ USER connection_
user ]
CONNECT TO DEFAULT
CONNECT connection_user
DATABASE connection_target
```

描述

`CONNECT`命令在客户端和 PostgreSQL 服务器之间建立一个连接。

参数

connection_target

*connection_target*以下列形式之一指定连接的目标服务器。

[*database_name*][@*host*][:*port*]

通过 TCP/IP 连接

`unix:postgresql://`*host*[:*port*] / [*database_name*][?*connection_option*]

通过 Unix 域套接字

`tcp:postgresql://`*host*[:*port*] / [*database_name*][?*connection_option*]

通过 TCP/IP 连接

SQL 字符串常量

包含上述形式之一的一个值

主变量

类型`char[]`或`VARCHAR[]`的主变量，它包含上述形式之一的一个值

connection_object

用于该连接的一个可选标识符，这样可以在其他命令中引用它。这可以是一个 SQL 标识符或者一个主变量。

connection_user

用于数据库连接的用户名。

使用 *user_name*/*password*、*user_name* `IDENTIFIED BY` *password*或者 *user_name* `USING` *password*之一，这个参数也能指定用户名和密码。

用户名和密码可以是 SQL 标识符、字符串常量或者主变量。

DEFAULT

按 libpq 的定义使用所有默认连接参数。

例子

这里是一些指定连接参数的变体：

```
EXEC SQL CONNECT TO "connectdb" AS main;
EXEC SQL CONNECT TO "connectdb" AS second;
EXEC SQL CONNECT TO "unix:postgresql://200.46.204.71/connectdb" AS
 main USER connectuser;
EXEC SQL CONNECT TO "unix:postgresql://localhost/connectdb" AS main
 USER connectuser;
EXEC SQL CONNECT TO 'connectdb' AS main;
EXEC SQL CONNECT TO 'unix:postgresql://localhost/connectdb' AS main
 USER :user;
EXEC SQL CONNECT TO :db AS :id;
EXEC SQL CONNECT TO :db USER connectuser USING :pw;
EXEC SQL CONNECT TO @localhost AS main USER connectdb;
EXEC SQL CONNECT TO REGRESSDB1 as main;
EXEC SQL CONNECT TO AS main USER connectdb;
EXEC SQL CONNECT TO connectdb AS :id;
EXEC SQL CONNECT TO connectdb AS main USER connectuser/connectdb;
EXEC SQL CONNECT TO connectdb AS main;
EXEC SQL CONNECT TO connectdb@localhost AS main;
EXEC SQL CONNECT TO tcp:postgresql://localhost/ USER connectdb;
EXEC SQL CONNECT TO tcp:postgresql://localhost/connectdb USER
 connectuser IDENTIFIED BY connectpw;
EXEC SQL CONNECT TO tcp:postgresql://localhost:20/connectdb USER
 connectuser IDENTIFIED BY connectpw;
EXEC SQL CONNECT TO unix:postgresql://localhost/ AS main USER
 connectdb;
EXEC SQL CONNECT TO unix:postgresql://localhost/connectdb AS main USER
 connectuser;
EXEC SQL CONNECT TO unix:postgresql://localhost/connectdb USER
 connectuser IDENTIFIED BY "connectpw";
EXEC SQL CONNECT TO unix:postgresql://localhost/connectdb USER
 connectuser USING "connectpw";
EXEC SQL CONNECT TO unix:postgresql://localhost/connectdb?connect_
timeout=14 USER connectuser;
```

这里是一个展示使用主变量指定连接参数的例子程序：

```
int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    char *dbname     = "testdb";    /* 数据库名 */
    char *user       = "testuser";  /* 连接用户名 */
    char *connection = "tcp:postgresql://localhost:5432/testdb";
                                    /* 连接字符串 */
    char ver[256];                  /* 存储版本字符串的缓冲区 */
EXEC SQL END DECLARE SECTION;
```

```
    ECPGdebug(1, stderr);

    EXEC SQL CONNECT TO :dbname USER :user;
    EXEC SQL SELECT version() INTO :ver;
    EXEC SQL DISCONNECT;

    printf("version: %s\n", ver);

    EXEC SQL CONNECT TO :connection USER :user;
    EXEC SQL SELECT version() INTO :ver;
    EXEC SQL DISCONNECT;

    printf("version: %s\n", ver);

    return 0;
}
```

兼容性

SQL 标准中说明了`CONNECT`，但是连接参数的格式是与实现相关的。

参见

DISCONNECT, SET CONNECTION

DEALLOCATE DESCRIPTOR

DEALLOCATE DESCRIPTOR — deallocate an SQL descriptor area

大纲

`DEALLOCATE DESCRIPTOR` *`name`*

描述

`DEALLOCATE DESCRIPTOR`释放一个命名的 SQL 描述符区域。

参数

`name`

要被释放的描述符的名称。它是大小写敏感的。这可以是一个 SQL 标识符或者一个主变量。

例子

```
EXEC SQL DEALLOCATE DESCRIPTOR mydesc;
```

兼容性

SQL 标准说明了`DEALLOCATE DESCRIPTOR`。

参见

ALLOCATE DESCRIPTOR, GET DESCRIPTOR, SET DESCRIPTOR

DECLARE

DECLARE — define a cursor

大纲

```
DECLARE cursor_name [ BINARY ] [ INSENSITIVE ] [ [ NO ] SCROLL ]
 CURSOR [ { WITH | WITHOUT } HOLD ] FOR prepared_name
DECLARE cursor_name [ BINARY ] [ INSENSITIVE ] [ [ NO ] SCROLL ]
 CURSOR [ { WITH | WITHOUT } HOLD ] FOR query
```

描述

`DECLARE`声明一个游标用来在一个预备语句的结果集上迭代。这个命令与直接的 SQL 命令`DECLARE`在语义上有一点点区别：后者会执行一个查询并且准备结果集用于检索，而这个嵌入式 SQL 命令仅仅声明一个名称作为“循环变量”用于在一个查询的结果集上迭代，实际的执行在游标被`OPEN`命令打开时才发生。

参数

cursor_name

一个游标名称，大小写敏感。这可以是一个 SQL 标识符或者一个主变量。

prepared_name

一个预备查询的名称，可以是一个 SQL 标识符或者一个主变量。

query

一个提供游标要返回的行的SELECT或者VALUES命令。

游标选项的含义请见DECLARE。

例子

为一个查询声明一个游标的例子：

```
EXEC SQL DECLARE C CURSOR FOR SELECT * FROM My_Table;
EXEC SQL DECLARE C CURSOR FOR SELECT Item1 FROM T;
EXEC SQL DECLARE cur1 CURSOR FOR SELECT version();
```

为一个预备语句声明一个游标的例子：

```
EXEC SQL PREPARE stmt1 AS SELECT version();
EXEC SQL DECLARE cur1 CURSOR FOR stmt1;
```

兼容性

SQL 标准中说明了`DECLARE`。

参见

OPEN, CLOSE, DECLARE

DESCRIBE

DESCRIBE — obtain information about a prepared statement or result set

大纲

```
DESCRIBE [ OUTPUT ] prepared_name USING [ SQL ] DESCRIPTOR descriptor_
name
DESCRIBE [ OUTPUT ] prepared_name INTO [ SQL ] DESCRIPTOR descriptor_
name
DESCRIBE [ OUTPUT ] prepared_name INTO sqlda_name
```

描述

`DESCRIBE`获取预备语句所含结果列的元数据，而不实际读取任何行。

参数

prepared_name

一个预备语句的名称。这可以是一个 SQL 标识符或者一个主变量。

descriptor_name

一个描述符名称。它是大小写敏感的。它可以是一个 SQL 标识符或者一个主变量。

sqlda_name

一个 SQLDA 变量的名称。

例子

```
EXEC SQL ALLOCATE DESCRIPTOR mydesc;
EXEC SQL PREPARE stmt1 FROM :sql_stmt;
EXEC SQL DESCRIBE stmt1 INTO SQL DESCRIPTOR mydesc;
EXEC SQL GET DESCRIPTOR mydesc VALUE 1 :charvar = NAME;
EXEC SQL DEALLOCATE DESCRIPTOR mydesc;
```

兼容性

SQL 标准中说明了`DESCRIBE`。

参见

ALLOCATE DESCRIPTOR, GET DESCRIPTOR

DISCONNECT

DISCONNECT — terminate a database connection

大纲

```
DISCONNECT connection_name
DISCONNECT [ CURRENT ]
DISCONNECT DEFAULT
DISCONNECT ALL
```

描述

`DISCONNECT`关闭一个（或者所有）到数据库的连接。

参数

connection_name

一个由`CONNECT`命令建立的数据库连接名称。

`CURRENT`

关闭“当前的”连接，它可以是最近打开的连接或者是由`SET CONNECTION`命令设置的连接。如果没有参数被传给`DISCONNECT`命令，这将是默认值。

`DEFAULT`

关闭默认连接。

`ALL`

关闭所有打开的连接。

例子

```
int
main(void)
{
    EXEC SQL CONNECT TO testdb AS DEFAULT USER testuser;
    EXEC SQL CONNECT TO testdb AS con1 USER testuser;
    EXEC SQL CONNECT TO testdb AS con2 USER testuser;
    EXEC SQL CONNECT TO testdb AS con3 USER testuser;

    EXEC SQL DISCONNECT CURRENT;  /* close con3          */
    EXEC SQL DISCONNECT DEFAULT;  /* close DEFAULT       */
    EXEC SQL DISCONNECT ALL;      /* close con2 and con1 */

    return 0;
}
```

兼容性

SQL 标准中说明了`DISCONNECT`。

参见

CONNECT, SET CONNECTION

EXECUTE IMMEDIATE

EXECUTE IMMEDIATE — dynamically prepare and execute a statement

大纲

`EXECUTE IMMEDIATE` *`string`*

描述

`EXECUTE IMMEDIATE`立刻预备并且执行一个动态指定的 SQL 语句，不检索结果行。

参数

`string`

C 字符串字面值，或包含待执行 SQL 语句的主变量。

例子

这里是一个用`EXECUTE IMMEDIATE`和一个名为`command`的主变量执行`INSERT`语句的例子：

```
sprintf(command, "INSERT INTO test (name, amount, letter) VALUES ('db:
 ''r1''', 1, 'f')");
EXEC SQL EXECUTE IMMEDIATE :command;
```

兼容性

SQL 标准中说明了`EXECUTE IMMEDIATE`。

GET DESCRIPTOR

GET DESCRIPTOR — get information from an SQL descriptor area

大纲

```
GET DESCRIPTOR descriptor_name :cvariable = descriptor_header_item [,
 ... ]
GET DESCRIPTOR descriptor_name VALUE column_number :cvariable =
 descriptor_item [, ... ]
```

描述

`GET DESCRIPTOR`从一个 SQL 描述符区域检索关于一个查询结果集的信息并且把它存储在主变量中。在使用这个命令把信息传输到主语言变量之前，一个描述符区域通常是用`FETCH`或`SELECT`填充的。

此命令有两种形式。第一种取得描述符的“头部”项目，它们适用于整个结果集，例如行数。第二种需要额外提供列号参数，取得特定列的信息，例如列名和实际列值。

参数

descriptor_name

一个描述符名称。

descriptor_header_item

一个标识要检索哪一个头部信息项的词元。当前只支持用于得到结果集中列数的`COUNT`。

column_number

要检索其信息的列号。计数从 1 开始。

descriptor_item

一个标识要检索哪一个有关一列信息的项的词元。被支持的项可见第 33.7.1 节。

cvariable

接收从描述符区域检索到的数据的主变量。

例子

检索一个结果集中列数的例子：

```
EXEC SQL GET DESCRIPTOR d :d_count = COUNT;
```

检索第一列中数据长度的例子：

```
EXEC SQL GET DESCRIPTOR d VALUE 1 :d_returned_octet_length = RETURNED_
OCTET_LENGTH;
```

把第二列的数据体检索成一个字符串的例子：

```
EXEC SQL GET DESCRIPTOR d VALUE 2 :d_data = DATA;
```

这里是执行`SELECT current_database();`并且显示列数、列数据长度和列数据的完整过程的例子:

```
int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    int  d_count;
    char d_data[1024];
    int  d_returned_octet_length;
EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO testdb AS con1 USER testuser;
    EXEC SQL ALLOCATE DESCRIPTOR d;

    /* 声明、打开游标，并为游标指定描述符 */
    EXEC SQL DECLARE cur CURSOR FOR SELECT current_database();
    EXEC SQL OPEN cur;
    EXEC SQL FETCH NEXT FROM cur INTO SQL DESCRIPTOR d;

    /* 得到全部列的数量 */
    EXEC SQL GET DESCRIPTOR d :d_count = COUNT;
    printf("d_count                 = %d\n", d_count);

    /* 得到一个返回列的长度 */
    EXEC SQL GET DESCRIPTOR d VALUE 1 :d_returned_octet_length =
 RETURNED_OCTET_LENGTH;
    printf("d_returned_octet_length = %d\n", d_returned_octet_length);

    /* 将返回的列取出成一个字符串 */
    EXEC SQL GET DESCRIPTOR d VALUE 1 :d_data = DATA;
    printf("d_data                  = %s\n", d_data);

    /* 关闭 */
    EXEC SQL CLOSE cur;
    EXEC SQL COMMIT;

    EXEC SQL DEALLOCATE DESCRIPTOR d;
    EXEC SQL DISCONNECT ALL;

    return 0;
}
```

当该例子被执行时，结果看起来是:

```
d_count                 = 1
d_returned_octet_length = 6
d_data                  = testdb
```

兼容性

SQL 标准中说明了`GET DESCRIPTOR`。

参见

ALLOCATE DESCRIPTOR, SET DESCRIPTOR

OPEN

OPEN — open a dynamic cursor

大纲

```
OPEN cursor_name
OPEN cursor_name USING value [, ... ]
OPEN cursor_name USING SQL DESCRIPTOR descriptor_name
```

描述

`OPEN`打开一个游标并且可选地绑定实际值到游标声明中的占位符。该游标必须之前用`DECLARE`命令声明。`OPEN`的执行会导致查询开始在服务器上执行。

参数

cursor_name

要被打开的游标的名称。这可以是一个 SQL 标识符或者一个主变量。

value

要被绑定到游标中一个占位符的值。这可以是一个 SQL 常量、一个主变量或者一个带有指示符的主变量。

descriptor_name

包含要绑定到游标中占位符的值的描述符的名称。这可以是一个 SQL 标识符或者一个主变量。

例子

```
EXEC SQL OPEN a;
EXEC SQL OPEN d USING 1, 'test';
EXEC SQL OPEN c1 USING SQL DESCRIPTOR mydesc;
EXEC SQL OPEN :curname1;
```

兼容性

SQL 标准中说明了`OPEN`。

参见

DECLARE, CLOSE

PREPARE

PREPARE — prepare a statement for execution

大纲

```
PREPARE name FROM string
```

描述

`PREPARE`预备一个以字符串动态指定的语句，以供执行。这不同于直接的 SQL 语句PREPARE，后者也可用于嵌入式程序。EXECUTE命令可用于执行任一种预备语句。

参数

prepared_name

预备查询的一个标识符。

string

C 字符串字面值，或包含可预备语句的主变量，语句可以是 SELECT、INSERT、UPDATE 或 DELETE。

例子

```
char *stmt = "SELECT * FROM test1 WHERE a = ? AND b = ?";

EXEC SQL ALLOCATE DESCRIPTOR outdesc;
EXEC SQL PREPARE foo FROM :stmt;

EXEC SQL EXECUTE foo USING SQL DESCRIPTOR indesc INTO SQL DESCRIPTOR
 outdesc;
```

兼容性

SQL 标准中说明了`PREPARE`。

参见

EXECUTE

SET AUTOCOMMIT

SET AUTOCOMMIT — set the autocommit behavior of the current session

大纲

```
SET AUTOCOMMIT { = | TO } { ON | OFF }
```

描述

`SET AUTOCOMMIT`设置当前数据库会话的自动提交行为。默认情况下，嵌入式 SQL 程序不处于自动提交模式，需要提交时必须显式发出`COMMIT`。此命令可以将会话改为自动提交模式，使每条语句都被隐式提交。

兼容性

`SET AUTOCOMMIT`是 PostgreSQL ECPG 的扩展。

SET CONNECTION

SET CONNECTION — select a database connection

大纲

```
SET CONNECTION [ TO | = ] connection_name
```

描述

`SET CONNECTION`设置“当前的”数据库连接，除非被覆盖，所有命令都会使用这个连接。

参数

connection_name

一个由`CONNECT`命令建立的数据库连接名称。

`DEFAULT`

将连接设为默认连接。

例子

```
EXEC SQL SET CONNECTION TO con2;
EXEC SQL SET CONNECTION = con1;
```

兼容性

SQL 标准中说明了`SET CONNECTION`。

参见

CONNECT, DISCONNECT

SET DESCRIPTOR

SET DESCRIPTOR — set information in an SQL descriptor area

大纲

```
SET DESCRIPTOR descriptor_name descriptor_header_item = value [, ... ]
SET DESCRIPTOR descriptor_name VALUE number descriptor_item = value [
, ...]
```

描述

`SET DESCRIPTOR`用值填充一个 SQL 描述符区域。然后该描述符区域通常会被用来在一个预备查询执行中绑定参数。

此命令有两种形式：第一种适用于描述符的“头部”，与某个特定数据项无关。
第二种为以编号标识的特定数据项赋值。

参数

descriptor_name

一个描述符名称。

descriptor_header_item

一个标识要设置哪个头部信息项的词元。当前只有设置描述符项数量的`COUNT`被支持。

number

要设置的描述符项的编号。计数从 1 开始。

descriptor_item

一个标识在描述符中要设置哪个信息项的词元。受支持的项的列表可见第 33.7.1 节。

value

要存入描述符项的值，可以是 SQL 常量或主变量。

例子

```
EXEC SQL SET DESCRIPTOR indesc COUNT = 1;
EXEC SQL SET DESCRIPTOR indesc VALUE 1 DATA = 2;
EXEC SQL SET DESCRIPTOR indesc VALUE 1 DATA = :val1;
EXEC SQL SET DESCRIPTOR indesc VALUE 2 INDICATOR = :val1, DATA = 'some
 string';
EXEC SQL SET DESCRIPTOR indesc VALUE 2 INDICATOR = :val2null, DATA = :
val2;
```

兼容性

SQL 标准中说明了`SET DESCRIPTOR`。

参见

ALLOCATE DESCRIPTOR, GET DESCRIPTOR

TYPE

TYPE — define a new data type

大纲

```
TYPE type_name IS ctype
```

描述

`TYPE`命令定义一个新的 C 类型。它等效于把一个`typedef`放在声明节中。

只有使用选项`-c`运行`ecpg`时才能识别这个命令。

参数

type_name

新类型的名称。这必须是一个合法的 C 类型名。

ctype

一个 C 类型说明。

例子

```
EXEC SQL TYPE customer IS
    struct
    {
        varchar name[50];
        int     phone;
    };

EXEC SQL TYPE cust_ind IS
    struct ind
    {
        short   name_ind;
        short   phone_ind;
    };

EXEC SQL TYPE c IS char reference;
EXEC SQL TYPE ind IS union { int integer; short smallint; };
EXEC SQL TYPE intarray IS int[AMOUNT];
EXEC SQL TYPE str IS varchar[BUFFERSIZ];
EXEC SQL TYPE string IS char[11];
```

这里是一个使用`EXEC SQL TYPE`的例子程序：

```
EXEC SQL WHENEVER SQLERROR SQLPRINT;

EXEC SQL TYPE tt IS
    struct
    {
        varchar v[256];
```

```
        int     i;
    };

EXEC SQL TYPE tt_ind IS
    struct ind {
        short   v_ind;
        short   i_ind;
    };

int
main(void)
{
EXEC SQL BEGIN DECLARE SECTION;
    tt t;
    tt_ind t_ind;
EXEC SQL END DECLARE SECTION;

    EXEC SQL CONNECT TO testdb AS con1;

    EXEC SQL SELECT current_database(), 256 INTO :t:t_ind LIMIT 1;

    printf("t.v = %s\n", t.v.arr);
    printf("t.i = %d\n", t.i);

    printf("t_ind.v_ind = %d\n", t_ind.v_ind);
    printf("t_ind.i_ind = %d\n", t_ind.i_ind);

    EXEC SQL DISCONNECT con1;

    return 0;
}
```

这个程序的输出看起来像:

```
t.v = testdb
t.i = 256
t_ind.v_ind = 0
t_ind.i_ind = 0
```

兼容性

`TYPE`命令是一种 PostgreSQL 扩展。

VAR

VAR — define a variable

大纲

```
VAR varname IS ctype
```

描述

`VAR`命令为主变量指定一个新的 C 数据类型。该主变量必须已在声明节中声明。

参数

varname

一个 C 变量名。

ctype

一个 C 类型说明。

例子

```
Exec sql begin declare section;
short a;
exec sql end declare section;
EXEC SQL VAR a IS int;
```

兼容性

`VAR`命令是一个 PostgreSQL 扩展。

WHENEVER

WHENEVER — specify the action to be taken when an SQL statement causes a specific class condition to be raised

大纲

```
WHENEVER { NOT FOUND | SQLERROR | SQLWARNING } action
```

描述

定义 SQL 执行结果出现特殊情况（未找到行、SQL 警告或错误）时采取的行为。

参数

参数描述见第 33.8.1 节。

例子

```
EXEC SQL WHENEVER NOT FOUND CONTINUE;
EXEC SQL WHENEVER NOT FOUND DO BREAK;
EXEC SQL WHENEVER SQLWARNING SQLPRINT;
EXEC SQL WHENEVER SQLWARNING DO warn();
EXEC SQL WHENEVER SQLERROR sqlprint;
EXEC SQL WHENEVER SQLERROR CALL print2();
EXEC SQL WHENEVER SQLERROR DO handle_error("select");
EXEC SQL WHENEVER SQLERROR DO sqlnotice(NULL, NONO);
EXEC SQL WHENEVER SQLERROR DO sqlprint();
EXEC SQL WHENEVER SQLERROR GOTO error_label;
EXEC SQL WHENEVER SQLERROR STOP;
```

典型用法是通过`WHENEVER NOT FOUND BREAK`处理遍历结果集的循环：

```
int
main(void)
{
    EXEC SQL CONNECT TO testdb AS con1;
    EXEC SQL ALLOCATE DESCRIPTOR d;
    EXEC SQL DECLARE cur CURSOR FOR SELECT current_database(), 'hoge',
 256;
    EXEC SQL OPEN cur;

    /* 到达结果集末尾时，跳出 while 循环 */
    EXEC SQL WHENEVER NOT FOUND DO BREAK;

    while (1)
    {
        EXEC SQL FETCH NEXT FROM cur INTO SQL DESCRIPTOR d;
        ...
    }

    EXEC SQL CLOSE cur;
    EXEC SQL COMMIT;
```

```
    EXEC SQL DEALLOCATE DESCRIPTOR d;
    EXEC SQL DISCONNECT ALL;

    return 0;
}
```

兼容性

SQL 标准中说明了`WHENEVER`，但是大部分动作是 PostgreSQL 扩展。

33.15. Informix兼容模式

`ecpg`可以运行在一种所谓的Informix 兼容模式中。如果这种模式被激活，它的行为就好像它是一个用于Informix E/SQL 的Informix预处理器。一般而言，这将允许你使用美元符号替代`EXEC SQL`来引入嵌入式 SQL 命令：

```
$int j = 3;
$CONNECT TO :dbname;
$CREATE TABLE test(i INT PRIMARY KEY, j INT);
$INSERT INTO test(i, j) VALUES (7, :j);
$COMMIT;
```

> **注意**
>
> `$`与后面的预处理器指令（如`include`、`define`、`ifdef`等）之间不能有任何空白。否则，预处理器会将该词元解析为主变量。

有两种兼容性模式：`INFORMIX`、`INFORMIX_SE`

在链接使用这种兼容性模式的程序时，要记得链接上和 ECPG 一起发布的`libcompat`。

除了之前解释过的语法糖，Informix兼容模式从 E/SQL 中移植了一些用于输入、输出和数据转换的函数以及嵌入式 SQL 语句到 ECPG 中。

Informix兼容模式与 ECPG 的 pgtypeslib 库密切相关。pgtypeslib 将 SQL 数据类型映射到 C 主程序中的数据类型，Informix兼容模式中的大多数附加函数用于操作这些 C 主程序类型。但兼容范围有限。它并不试图复制Informix的行为，而是允许执行大致相同的操作，并提供名称和基本行为相同的函数；如果目前使用的是Informix，则不能用它直接替代。
此外，有些数据类型也有差异。例如，PostgreSQL的日期时间和 interval 类型不支持`YEAR TO MINUTE`这样的范围，ECPG 也不支持。

33.15.1. 附加类型

Informix 模式现在支持 Informix 特有的“string”伪类型，用于存储去除尾随空格的字符串数据，无需使用`typedef`。事实上，在 Informix 模式中，ECPG 拒绝处理包含`typedef sometype string;`的源文件。

```
EXEC SQL BEGIN DECLARE SECTION;
string userid; /* 此变量将包含去除尾随空格的数据 */
EXEC SQL END DECLARE SECTION;
```

```
EXEC SQL FETCH MYCUR INTO :userid;
```

33.15.2. 附加的/缺少的嵌入式 SQL 语句

```
CLOSE DATABASE
```

这个语句关闭当前连接。实际上，这是ECPG的`DISCONNECT CURRENT`的同义词：

```
$CLOSE DATABASE;                /* close the current connection */
EXEC SQL CLOSE DATABASE;
```

```
FREE cursor_name
```

由于ECPG的工作方式与Informix的ESQL/C有所不同（即，哪些步骤纯粹是语法转换，哪些步骤依赖于底层运行时库），在ECPG中没有`FREE cursor_name`语句。这是因为在ECPG中，`DECLARE CURSOR`不会转换为使用游标名称的运行时库的函数调用。这意味着在ECPG运行时库中没有SQL游标的运行时记录，只在PostgreSQL服务器中有。

```
FREE statement_name
```

`FREE statement_name`是`DEALLOCATE PREPARE statement_name`的同义词。

33.15.3. Informix-兼容的 SQLDA 描述符区域

Informix 兼容模式支持的结构体与第 33.7.2 节中所述的不同，如下所示：

```
struct sqlvar_compat
{
    short   sqltype;
    int     sqllen;
    char   *sqldata;
    short  *sqlind;
    char   *sqlname;
    char   *sqlformat;
    short   sqlitype;
    short   sqlilen;
    char   *sqlidata;
    int     sqlxid;
    char   *sqltypename;
    short   sqltypelen;
    short   sqlownerlen;
    short   sqlsourcetype;
    char   *sqlowner name;
    int     sqlsourceid;
    char   *sqlilongdata;
    int     sqlflags;
    void   *sqlreserved;
};

struct sqlda_compat
{
    short  sqld;
    struct sqlvar_compat *sqlvar;
```

```
    char   desc_name[19];
    short  desc_occ;
    struct sqlda_compat *desc_next;
    void  *reserved;
};

typedef struct sqlvar_compat    sqlvar_t;
typedef struct sqlda_compat     sqlda_t;
```

全局属性是：

`sqld`

SQLDA描述符中的字段数量。

`sqlvar`

每一个域属性的指针。

`desc_name`

未使用，用零字节填充。

`desc_occ`

已分配结构体的尺寸。

`desc_next`

如果结果集包含多于一个记录，这个字段是下一个 SQLDA 结构体的指针。

`reserved`

未使用的指针，包含 NULL。为 Informix-兼容性而保留。

每个字段的属性如下，它们存储在 `sqlvar` 数组中：

`sqltype`

域的类型。可以使用的常量定义在`sqltypes.h`中。

`sqllen`

域数据的长度。

`sqldata`

指向字段数据的指针，类型为`char *`，所指的数据采用二进制格式。例如：

```
int intval;

switch (sqldata->sqlvar[i].sqltype)
{
    case SQLINTEGER:
        intval = *(int *)sqldata->sqlvar[i].sqldata;
        break;
  ...
```

```
}
```

sqlind

指向 NULL 指示符的指针。如果由 DESCRIBE 或 FETCH 返回，它始终是有效指针。如果作为 `EXECUTE ... USING sqlda;`的输入，则 NULL 指针表示此字段的值不是 NULL。否则，必须正确设置有效指针和`sqlitype`。例如：

```
if (*(int2 *)sqldata->sqlvar[i].sqlind != 0)
    printf("value is NULL\n");
```

sqlname

域的名称。以 0 终止的字符串。

sqlformat

在 Informix 中保留；这里是对该字段调用 `PQfformat()` 所得的值。

sqlitype

NULL 指示符数据的类型。从服务器返回数据时，它始终是 SQLSMINT。将`SQLDA`用于参数化查询时，按所设置的类型处理数据。

sqlilen

NULL 指示符数据的长度。

sqlxid

字段的扩展类型，即 `PQftype()` 的结果。

sqltypename
sqltypelen
sqlownerlen
sqlsourcetype
sqlownername
sqlsourceid
sqlflags
sqlreserved

未使用。

sqlilongdata

如果`sqllen`大于 32KB，它等于`sqldata`。

示例：

```
EXEC SQL INCLUDE sqlda.h;

    sqlda_t        *sqlda; /* 这不需要在嵌入式 DECLARE SECTION 下 */

    EXEC SQL BEGIN DECLARE SECTION;
    char *prep_stmt = "select * from table1";
    int i;
```

```
EXEC SQL END DECLARE SECTION;

...

EXEC SQL PREPARE mystmt FROM :prep_stmt;

EXEC SQL DESCRIBE mystmt INTO sqlda;

printf("# of fields: %d\n", sqlda->sqld);
for (i = 0; i < sqlda->sqld; i++)
  printf("field %d: \"%s\"\n", sqlda->sqlvar[i]->sqlname);

EXEC SQL DECLARE mycursor CURSOR FOR mystmt;
EXEC SQL OPEN mycursor;
EXEC SQL WHENEVER NOT FOUND GOTO out;

while (1)
{
  EXEC SQL FETCH mycursor USING sqlda;
}

EXEC SQL CLOSE mycursor;

free(sqlda); /* 主结构完全被 free(), sqlda 和 sqlda->sqlvar 在一个已分配区域中 */
```

更多信息参见 `sqlda.h` 头文件和 `src/interfaces/ecpg/test/compat_informix/sqlda.pgc` 回归测试。

33.15.4. 附加函数

`decadd`

将两个 decimal 类型的值相加。

```
int decadd(decimal *arg1, decimal *arg2, decimal *sum);
```

该函数接收指向第一个 decimal 类型操作数的指针（`arg1`）、指向第二个 decimal 类型操作数的指针（`arg2`），以及指向用于保存和的 decimal 类型值的指针（`sum`）。成功时返回 0，发生溢出时返回`ECPG_INFORMIX_NUM_OVERFLOW`，发生下溢时返回`ECPG_INFORMIX_NUM_UNDERFLOW`。其他失败情况下返回 -1，并将`errno`设为 pgtypeslib 中对应的`errno`值。

`deccmp`

比较两个 decimal 类型的变量。

```
int deccmp(decimal *arg1, decimal *arg2);
```

该函数接收指向第一个 decimal 值的指针（`arg1`）和指向第二个 decimal 值的指针（`arg2`），并返回一个整数值，表示哪个值更大。

- 1，如果`arg1`指向的值大于`var2`指向的值
- -1，如果`arg1`指向的值小于`arg2`指向的值

- 0，如果arg1指向的值和arg2指向的值相等

deccopy

复制一个 decimal 值。

```
void deccopy(decimal *src, decimal *target);
```

该函数的第一个参数是指向待复制 decimal 值的指针（src），第二个参数是指向目标 decimal 类型结构体的指针（target）。

deccvasc

将一个值的 ASCII 表示转换为 decimal 类型。

```
int deccvasc(char *cp, int len, decimal *np);
```

该函数接收指向待转换数值的字符串表示的指针（cp）及其长度len。np 是指向用于保存操作结果的 decimal 值的指针。

例如，有效的格式包括：-2、.794、+3.44、592.49E07或 -32.84e-4。

该函数成功返回0。如果发生溢出或下溢，则返回ECPG_INFORMIX_NUM_OVERFLOW或ECPG_INFORMIX_NUM_UNDERFLOW。如果ASCII表示无法解析，则返回ECPG_INFORMIX_BAD_NUMERIC，或者如果在解析指数时出现问题，则返回 ECPG_INFORMIX_BAD_EXPONENT。

deccvdbl

将 double 类型的值转换为 decimal 类型的值。

```
int deccvdbl(double dbl, decimal *np);
```

该函数接收应该被转换的 double 类型变量作为其第一个参数（dbl）。作为第二个参数（np），该函数接收一个指向应该保存操作结果的 decimal 变量的指针。

该函数在成功时返回0，在转换失败时返回负值。

deccvint

将 int 类型的值转换为 decimal 类型的值。

```
int deccvint(int in, decimal *np);
```

该函数接收应该被转换的 int 类型变量作为其第一个参数（in）。作为第二个参数（np），该函数接收一个指向应该保存操作结果的 decimal 变量的指针。

该函数在成功时返回0，在转换失败时返回负值。

deccvlong

将类型为 long 的值转换为类型为 decimal 的值。

```
int deccvlong(long lng, decimal *np);
```

该函数接收应该被转换的类型为 long 的变量作为其第一个参数（lng）。作为第二个参数（np），该函数接收一个指向应该保存操作结果的 decimal 变量的指针。

该函数在成功时返回0，在转换失败时返回负值。

decdiv

将两个 decimal 类型的变量相除。

```
int decdiv(decimal *n1, decimal *n2, decimal *result);
```

该函数接收指向第一个（n1）和第二个（n2）操作数的变量的指针，并计算n1/n2。result是应该保存操作结果的变量的指针。

在成功时返回0，如果除法失败则返回负值。如果发生溢出或下溢，则函数分别返回 ECPG_INFORMIX_NUM_OVERFLOW或 ECPG_INFORMIX_NUM_UNDERFLOW。如果尝试除以零，则函数返回ECPG_INFORMIX_DIVIDE_ZERO。

decmul

将两个 decimal 值相乘。

```
int decmul(decimal *n1, decimal *n2, decimal *result);
```

该函数接收指向第一个（n1）和第二个（n2）操作数的变量的指针，并计算n1*n2。result是一个指向应该保存操作结果的变量的指针。

在成功时返回0，如果乘法失败则返回负值。如果发生溢出或下溢，函数分别返回 ECPG_INFORMIX_NUM_OVERFLOW或 ECPG_INFORMIX_NUM_UNDERFLOW。

decsub

从一个 decimal 值中减去另一个 decimal 值。

```
int decsub(decimal *n1, decimal *n2, decimal *result);
```

该函数接收指向第一个（n1）和第二个（n2）操作数的变量的指针，并计算n1-n2。result是应该保存操作结果的变量的指针。

在成功时返回0，如果减法失败则返回负值。如果发生溢出或下溢，函数分别返回 ECPG_INFORMIX_NUM_OVERFLOW或 ECPG_INFORMIX_NUM_UNDERFLOW。

dectoasc

将 decimal 类型变量转换为 C char* 字符串中的 ASCII 表示。

```
int dectoasc(decimal *np, char *cp, int len, int right)
```

该函数接收指向 decimal 类型变量的指针（np），并将该变量转换为文本表示。cp是用于保存操作结果的缓冲区。参数right指定输出中小数点右侧的位数，
结果将舍入到这个小数位数。将right设为 -1 表示输出所有可用的小数位。如果len指定的输出缓冲区长度不足以容纳文本表示及末尾的零字节，则只在结果中存入一个*字符，并返回-1。

如果缓冲区cp太小，函数返回 -1；如果内存耗尽，则返回ECPG_INFORMIX_OUT_OF_MEMORY。

dectodbl

将 decimal 类型的变量转换为 double。

```
int dectodbl(decimal *np, double *dblp);
```

该函数接受指向待转换 decimal 值的指针（`np`），以及指向用于保存操作结果的 double 变量的指针（`dblp`）。

当成功时，返回0；如果转换失败，则返回负值。

`dectoint`

将 decimal 类型的变量转换为整数。

```
int dectoint(decimal *np, int *ip);
```

该函数接受指向待转换 decimal 值的指针（`np`），以及指向用于保存操作结果的整数变量的指针（`ip`）。

当成功时，返回0；如果转换失败，则返回负值。如果发生溢出，将返回`ECPG_INFORMIX_NUM_OVERFLOW`。

请注意，ECPG实现与Informix实现不同。Informix将整数限制在-32767到32767的范围内，而ECPG实现中的限制取决于架构（`-INT_MAX .. INT_MAX`）。

`dectolong`

将 decimal 类型的变量转换为长整数。

```
int dectolong(decimal *np, long *lngp);
```

该函数接受指向待转换 decimal 值的指针（`np`），以及指向用于保存操作结果的 long 变量的指针（`lngp`）。

当成功时，返回0；如果转换失败，则返回负值。如果发生溢出，将返回`ECPG_INFORMIX_NUM_OVERFLOW`。

请注意，ECPG实现与Informix实现不同。Informix将长整型限制在-2,147,483,647到2,147,483,647的范围内，而ECPG实现中的限制取决于架构（`-LONG_MAX .. LONG_MAX`）。

`rdatestr`

将日期转换成 C char* 字符串。

```
int rdatestr(date d, char *str);
```

该函数接受两个参数，第一个是待转换的日期（`d`），第二个是指向目标字符串的指针。
输出格式始终为 `yyyy-mm-dd`，因此必须为字符串分配至少 11 字节（包括值为零的终止字节）。

该函数在成功时返回0，在错误时返回负值。

请注意，ECPG的实现与Informix的实现不同。在Informix中，
格式可以通过设置环境变量来影响。然而，在ECPG中，你无法更改输出格式。

`rstrdate`

解析日期的文本表示。

```
int rstrdate(char *str, date *d);
```

该函数接收要转换的日期的文本表示（`str`）和指向类型为 date 的变量的指针（`d`）。此函数不允许你指定格式掩码。它使用Informix的默认格式掩码，即`mm/dd/yyyy`。在内部，此函数通过`rdefmtdate`实现。因此，`rstrdate`不会更快，如果可以选择，应选择允许你显式指定格式掩码的 `rdefmtdate`。

这个函数返回与`rdefmtdate`相同的值。

`rtoday`

得到当前日期。

```
void rtoday(date *d);
```

该函数接收一个指向 date 变量的指针（`d`），并将其指向的变量设为当前日期。

在内部，此函数使用`PGTYPESdate_today`函数。

`rjulmdy`

从一个 date 类型的变量中提取日、月和年的值。

```
int rjulmdy(date d, short mdy[3]);
```

该函数接收日期`d`和一个指向包含 3 个 short 整数值的数组`mdy`的指针。变量名表明了顺序：`mdy[0]`设为月份数，`mdy[1]`设为日，`mdy[2]`包含年份。

这个函数目前总是返回0。

在内部，该函数使用`PGTYPESdate_julmdy`函数。

`rdefmtdate`

使用格式掩码将字符串转换为日期类型的值。

```
int rdefmtdate(date *d, char *fmt, char *str);
```

该函数接收一个指向用于保存操作结果的 date 值的指针（`d`）、用于解析日期的格式掩码（`fmt`），以及包含日期文本表示的 C char* 字符串（`str`）。文本表示应与格式掩码匹配，但字符串与格式掩码无需一一对应。该函数只分析先后顺序，并查找表示年份位置的`yy`或`yyyy`、表示月份位置的`mm`以及表示日位置的`dd`。

该函数返回以下值：

- 0 - 函数成功终止。
- `ECPG_INFORMIX_ENOSHORTDATE` - 日期不包含日、月和年之间的分隔符。在这种情况下，输入字符串必须恰好为6或8个字节长，但实际不是。
- `ECPG_INFORMIX_ENOTDMY` - 格式字符串未正确指示年、月和日的顺序。
- `ECPG_INFORMIX_BAD_DAY` - 输入字符串不包含有效的日。
- `ECPG_INFORMIX_BAD_MONTH` - 输入字符串不包含有效的月。
- `ECPG_INFORMIX_BAD_YEAR` - 输入字符串不包含有效的年。

在内部，此函数实现为使用`PGTYPESdate_defmt_asc`函数。请参阅那里的参考资料，了解示例输入表。

`rfmtdate`

使用格式掩码将 date 类型的变量转换为文本表示。

```
int rfmtdate(date d, char *fmt, char *str);
```

该函数接收要转换的日期（`d`）、格式掩码（`fmt`）
和将保存日期文本表示的字符串（`str`）。

成功时，返回 0；如果发生错误，则返回一个负值。

在内部，此函数使用`PGTYPESdate_fmt_asc`函数，有关示例，请参阅那里的参考资料。

`rmdyjul`

根据一个包含 3 个 short 整数的数组创建日期值，这些整数指定日期的日、月和年。

```
int rmdyjul(short mdy[3], date *d);
```

该函数接收包含 3 个 short 整数的数组（`mdy`）和一个指向 date 类型变量的指针，用来保存操作结果。

目前该函数始终返回0。

在内部，该函数实现为使用函数`PGTYPESdate_mdyjul`。

`rdayofweek`

为一个日期值返回表示它是星期几的数字。

```
int rdayofweek(date d);
```

这个函数接收日期变量`d`作为它唯一的参数并且返回一个整数说明这个日期是星期几。

- 0 - 星期日
- 1 - 星期一
- 2 - 星期二
- 3 - 星期三
- 4 - 星期四
- 5 - 星期五
- 6 - 星期六

在内部，该函数被实现为使用函数`PGTYPESdate_dayofweek`。

`dtcurrent`

检索当前时间戳。

```
void dtcurrent(timestamp *ts);
```

该函数检索当前时间戳，并将其保存到`ts`指向的时间戳变量中。

`dtcvasc`

将时间戳的文本表示解析为 timestamp 变量。

```
int dtcvasc(char *str, timestamp *ts);
```

该函数接收要解析的字符串（`str`），以及指向用于保存操作结果的 timestamp 变量的指针（`ts`）。

该函数在成功时返回0，在错误时返回负值。

在内部，此函数使用`PGTYPEStimestamp_from_asc`函数。请参阅那里的参考资料，了解包含示例输入的表格。

`dtcvfmtasc`

使用格式掩码将时间戳的文本表示解析为 timestamp 变量。

```
dtcvfmtasc(char *inbuf, char *fmtstr, timestamp *dtvalue)
```

该函数接收要解析的字符串（`inbuf`）、使用的格式掩码（`fmtstr`），以及指向用于保存操作结果的 timestamp 变量的指针（`dtvalue`）。

这个函数是通过`PGTYPEStimestamp_defmt_asc`函数实现的。请参阅那里的文档，了解可用的格式说明符列表。

该函数在成功时返回0，在错误时返回负值。

`dtsub`

从一个时间戳中减去另一个时间戳，并返回一个 interval 类型的变量。

```
int dtsub(timestamp *ts1, timestamp *ts2, interval *iv);
```

该函数从`ts1`指向的时间戳变量中减去`ts2`指向的时间戳变量，并将结果存入`iv`指向的 interval 变量中。

成功时返回 0，发生错误时返回负值。

`dttoasc`

将 timestamp 变量转换为 C char* 字符串。

```
int dttoasc(timestamp *ts, char *output);
```

该函数接收指向待转换 timestamp 变量的指针（`ts`），以及用于保存操作结果的字符串（`output`）。它将`ts`转换为符合 SQL 标准的文本表示，格式为`YYYY-MM-DD HH:MM:SS`。

成功时返回 0，发生错误时返回负值。

`dttofmtasc`

使用格式掩码将 timestamp 变量转换为 C char*。

```
int dttofmtasc(timestamp *ts, char *output, int str_len, char
 *fmtstr);
```

该函数的第一个参数是指向待转换时间戳的指针（ts），随后是指向输出缓冲区的指针（output）、为输出缓冲区分配的最大长度（str_len），以及转换所用的格式掩码（fmtstr）。

成功时返回 0，发生错误时返回负值。

在内部，此函数使用PGTYPEStimestamp_fmt_asc函数。请参阅那里的参考资料，了解可以使用哪些格式掩码说明符。

intoasc

将 interval 变量转换为 C char* 字符串。

```
int intoasc(interval *i, char *str);
```

该函数接受指向待转换 interval 变量的指针（i），以及用于保存操作结果的字符串（str）。它将 i 转换为符合 SQL 标准的文本表示，格式为 YYYY-MM-DD HH:MM:SS。

成功时返回 0，发生错误时返回负值。

rfmtlong

将长整型值使用格式掩码转换为其文本表示形式。

```
int rfmtlong(long lng_val, char *fmt, char *outbuf);
```

该函数接收长整型值lng_val，格式掩码fmt和指向输出缓冲区outbuf的指针。它根据格式掩码将长整型值转换为其文本表示形式。

格式掩码可以由以下格式指定字符组成：

- *（星号）- 如果此位置为空白，用星号填充。
- &（和号）- 如果此位置为空白，用零填充。
- # - 将前导零转换为空格。
- < - 将数字左对齐在字符串中。
- ,（逗号）- 将四位或更多位的数字每三位分为一组，用逗号分隔。
- .（句点）- 此字符将整数部分与小数部分分隔开。
- -（减号）- 如果数字是负值，则显示减号。
- +（加号）- 如果数字是正值，则显示加号。
- （ - 这个字符替换负数前面的减号。减号不会显示。
- ）- 这个字符替换减号，并打印在负值后面。
- $ - 货币符号。

rupshift

将字符串转换为大写。

```
void rupshift(char *str);
```

该函数接收一个指向字符串的指针，并将每个小写字符转换为大写。

`byleng`

返回字符串中字符的数量，不包括末尾的空格。

```
int byleng(char *str, int len);
```

该函数期望一个固定长度的字符串作为第一个参数 (`str`)，并将其长度作为第二个参数 (`len`)。它返回有效字符的数量，即不包括末尾空格的字符串长度。

`ldchar`

将固定长度的字符串复制到以空字符结尾的字符串中。

```
void ldchar(char *src, int len, char *dest);
```

该函数接收要复制的固定长度字符串（`src`）、其长度（`len`）
和指向目标内存的指针（`dest`）。请注意，你需要为`dest`指向的字符串保留至少`len+1`字节。该函数最多复制`len`字节到新位置（如果源字符串具有尾随空格，则会少一些），并添加空字符终止符。

`rgetmsg`

```
int rgetmsg(int msgnum, char *s, int maxsize);
```

这个函数存在，但目前尚未实现！

`rtypalign`

```
int rtypalign(int offset, int type);
```

这个函数存在，但目前尚未实现！

`rtypmsize`

```
int rtypmsize(int type, int len);
```

这个函数存在，但目前尚未实现！

`rtypwidth`

```
int rtypwidth(int sqltype, int sqllen);
```

这个函数存在，但目前尚未实现！

`rsetnull`

将一个变量设置为NULL。

```
int rsetnull(int t, char *ptr);
```

该函数接收一个整数，表示变量的类型，以及一个指向变量本身的指针，
该指针被转换为C中的char*指针。

下列类型存在：

- `CCHARTYPE` - 用于`char`或`char*`类型的变量

- CSHORTTYPE - 用于short int类型的变量
- CINTTYPE - 用于int类型的变量
- CBOOLTYPE - 用于boolean类型的变量
- CFLOATTYPE - 用于float类型的变量
- CLONGTYPE - 用于long类型的变量
- CDOUBLETYPE - 用于double类型的变量
- CDECIMALTYPE - 用于decimal类型的变量
- CDATETYPE - 用于date类型的变量
- CDTIMETYPE - 用于timestamp类型的变量

这是调用此函数的示例：

```
$char c[] = "abc       ";
$short s = 17;
$int i = -74874;

rsetnull(CCHARTYPE, (char *) c);
rsetnull(CSHORTTYPE, (char *) &s);
rsetnull(CINTTYPE, (char *) &i);
```

risnull

测试变量是否为NULL。

```
int risnull(int t, char *ptr);
```

该函数接收要测试的变量类型（t）以及指向该变量的指针（ptr）。请注意，后者需要转换为char*类型。查看函数rsetnull以获取可能的变量类型列表。

这是如何使用这个函数的示例：

```
$char c[] = "abc       ";
$short s = 17;
$int i = -74874;

risnull(CCHARTYPE, (char *) c);
risnull(CSHORTTYPE, (char *) &s);
risnull(CINTTYPE, (char *) &i);
```

33.15.5. 额外的常量

注意这里所有的常量都描述错误并且它们都被定义为表示负值。在每一种不同常量的描述中，你还可以找到在当前实现中该常量表示的值。不过你不应该依赖于这个数字。但是你可以相信所有的这些常量都是被定义为表示负值。

ECPG_INFORMIX_NUM_OVERFLOW

如果在一次计算中发生了溢出，函数会返回这个值。在内部它被定义为 -1200（Informix定义）。

`ECPG_INFORMIX_NUM_UNDERFLOW`

如果在一次计算中发生了下溢，函数会返回这个值。在内部它被定义为 -1201（Informix定义）。

`ECPG_INFORMIX_DIVIDE_ZERO`

如果发现尝试除零，函数会返回这个值。在内部它被定义为 -1202（Informix定义）。

`ECPG_INFORMIX_BAD_YEAR`

如果解析日期时发现无效的年份值，函数返回此值。其内部定义为 -1204（Informix的定义）。

`ECPG_INFORMIX_BAD_MONTH`

如果解析日期时发现无效的月份值，函数返回此值。其内部定义为 -1205（Informix的定义）。

`ECPG_INFORMIX_BAD_DAY`

如果解析日期时发现无效的日值，函数返回此值。其内部定义为 -1206（Informix的定义）。

`ECPG_INFORMIX_ENOSHORTDATE`

如果解析例程需要短日期表示，但未得到长度正确的日期字符串，函数返回此值。其内部定义为 -1209（Informix的定义）。

`ECPG_INFORMIX_DATE_CONVERT`

如果在日期格式化时产生了一个错误，函数会返回这个值。在内部它被定义为 -1210（Informix定义）。

`ECPG_INFORMIX_OUT_OF_MEMORY`

如果在操作时内存被耗尽，函数会返回这个值。在内部它被定义为 -1211（Informix定义）。

`ECPG_INFORMIX_ENOTDMY`

如果一个解析例程被假定为得到一个格式掩码（如`mmddyy`）但是列出的域并不是全部正确，函数会返回这个值。在内部它被定义为 -1212（Informix定义）。

`ECPG_INFORMIX_BAD_NUMERIC`

如果解析例程因 numeric 值的文本表示有误而无法解析，或者例程因至少一个 numeric 变量无效而无法完成涉及 numeric 变量的计算，函数返回此值。其内部定义为 -1213（Informix的定义）。

`ECPG_INFORMIX_BAD_EXPONENT`

如果一个解析例程不能解析一个指数，函数会返回这个值。在内部它被定义为 -1216（Informix定义）。

`ECPG_INFORMIX_BAD_DATE`

如果一个解析例程不能解析一个日期，函数会返回这个值。在内部它被定义为 -1218（Informix定义）。

`ECPG_INFORMIX_EXTRA_CHARS`

如果一个解析例程被传递了它不能解析的额外字符，函数会返回这个值。在内部它被定义为 -1264（Informix定义）。

33.16. 内部

这一节解释ECPG在内部如何工作。这些信息有时有助于用户理解如何使用ECPG。

`ecpg`写入输出的前四行是固定的：其中两行是注释，另外两行是与库接口所必需的包含行。随后预处理器会通读整个文件并写出输出，通常它只是把所有内容原样回显到输出中。

当它看到`EXEC SQL`语句时，就会介入并对其进行改写。命令以`EXEC SQL`开始，以`;`结束。其间的所有内容都会被视为一条SQL语句，并进行变量替换解析。

当某个符号以冒号（`:`）开头时，就会发生变量替换。系统会在此前于`EXEC SQL DECLARE`声明节中声明的变量里查找同名变量。

库中最重要的函数是`ECPGdo`，负责执行大多数命令。它接收可变数量的参数，参数数量很容易达到 50 个左右；我们希望这在任何平台上都不会造成问题。

参数是：

一个行号

这是原始行的行号，只用于错误消息。

一个字符串

这是将要发出的SQL命令。输入变量会对它进行修改，也就是那些在编译时尚未知晓、但要填入命令中的变量。字符串中的`?`表示变量应当放置的位置。

输入变量

每个输入变量都会生成十个参数（见下文）。

`ECPGt_EOIT`

一个表明后面没有更多输入变量的`enum`值。

输出变量

每个输出变量都会生成十个参数（见下文）。这些变量将由该函数填充。

`ECPGt_EORT`

一个表明后面没有更多变量的`enum`值。

对于每一个作为SQL命令一部分的变量，该函数得到十个参数：

1. 用特殊符号表示的类型。
2. 一个值的指针或者一个指针的指针。
3. 如果变量是一个`char`或者`varchar`，这是它的尺寸。
4. 数组中元素的数量（用于数组获取）。

5. 数组中下一个元素的偏移量（用于数组获取）。
6. 作为一个特殊符号的指示符变量的类型。
7. 一个指示符变量的指针。
8. 0
9. 指示符数组中的元素数量（用于数组取数）。
10. 到指示符数组中下一个元素的偏移量（用于数组取数）。

注意并非所有 SQL 命令都被以这种方式对待。例如，一个打开游标语句：

```
EXEC SQL OPEN cursor;
```

它不会被复制到输出中。相反，游标的`DECLARE`命令会被放在`OPEN`命令的位置上，因为真正打开游标的正是它。

这里有一个完整的例子，它描述了一个文件`foo.pgc`的预处理器输出（对预处理器的每一个特定版本细节可能不同）：

```
EXEC SQL BEGIN DECLARE SECTION;
int index;
int result;
EXEC SQL END DECLARE SECTION;
...
EXEC SQL SELECT res INTO :result FROM mytable WHERE index = :index;
```

会被转换成：

```
/* Processed by ecpg (2.6.0) */
/* These two include files are added by the preprocessor */
#include <ecpgtype.h>;
#include <ecpglib.h>;

/* 声明节开始 */

#line 1 "foo.pgc"

 int index;
 int result;
/* 声明节结束 */
...
ECPGdo(__LINE__, NULL, "SELECT res FROM mytable WHERE index = ?     ",
        ECPGt_int,&(index),1L,1L,sizeof(int),
        ECPGt_NO_INDICATOR, NULL , 0L, 0L, 0L, ECPGt_EOIT,
        ECPGt_int,&(result),1L,1L,sizeof(int),
        ECPGt_NO_INDICATOR, NULL , 0L, 0L, 0L, ECPGt_EORT);
#line 147 "foo.pgc"
```

（这里的缩进是为了可读性而添加的，并非是预处理器做的处理）。

第 34 章 信息模式

信息模式由一组视图组成，其中包含当前数据库中定义的对象的信息。信息模式由 SQL 标准定义，因此可以预期它具有可移植性并保持稳定–而系统目录则不同，它们是 PostgreSQL 特有的，并且是围绕实现方面的考虑建模的。不过，信息模式视图不包含 PostgreSQL 特有功能的信息；要查询这些内容，需要查询系统目录或其他 PostgreSQL 特有的视图。

> **注意**
>
> 当在数据库中查询约束信息时，一个期望返回一行的标准兼容的查询可能返回多行。这是因为 SQL 标准要求约束名在一个模式中唯一，但是PostgreSQL并不强制这种限制。PostgreSQL自动产生的约束名避免在相同的模式中重复，但是用户能够指定这种重复的名称。
>
> 这个问题可能在查询信息模式视图时出现，例如`check_constraint_routine_usage`、`check_constraints`、`domain_constraints`和 `referential_constraints`。一些其他视图也有相似的问题，但是它们包含了表名来帮助区分重复行，例如`constraint_column_usage`、`constraint_table_usage`、`table_constraints`。

34.1. 模式

信息模式本身是一个名为`information_schema`的模式。这个模式自动存在于所有数据库中。这个模式的拥有者是集簇中的初始数据库用户，并且该用户自然地拥有这个模式上的所有权限，包括删除它的能力（但是这样节省的空间是很小的）。

默认情况下，信息模式不在模式搜索路径中，因此你需要使用限定名访问其中的所有对象。因为信息模式中的某些对象的名称是可能出现在用户应用中的一般名称，如果你想把该信息模式放在路径中，你应该小心。

34.2. 数据类型

信息模式视图的列使用在信息模式中定义的特殊数据类型。这些类型被定义为普通内置类型之上的简单域。不应在信息模式之外使用这些类型，但如果应用要从信息模式中查询数据，就必须能够处理它们。

这些类型为：

`cardinal_number`

一种非负整数。

`character_data`

一种字符串（没有指定最大长度）。

`sql_identifier`

一种字符串。该类型用于 SQL 标识符，其他任何文本数据则使用类型`character_data`。

time_stamp

在类型timestamp with time zone之上的一个域。

yes_or_no

一种字符串域，只包含YES或NO。它用于在信息模式中表示布尔（真/假）数据。（信息模式是在类型boolean被纳入 SQL 标准之前设计出来的，因此需要采用这种约定以保持向后兼容。）

信息模式中的每一列都具有这五种类型之一。

34.3. information_schema_catalog_name

information_schema_catalog_name 是一个表，它始终只包含一行一列，用于存放当前数据库的名称（按 SQL 术语称为当前目录）。

表 34.1. information_schema_catalog_name 列

名称	数据类型	描述
catalog_name	sql_identifier	包含这个信息模式的数据库名称

34.4. administrable_role_authorizations

视图administrable_role_authorizations标识当前用户对其有管理选项的所有角色。

表 34.2. administrable_role_authorizations 列

名称	数据类型	描述
grantee	sql_identifier	被授予该角色成员资格的角色名称（可以是当前用户，也可以是在嵌套角色成员资格情况下的另一个角色）
role_name	sql_identifier	一个角色的名称
is_grantable	yes_or_no	总是 YES

34.5. applicable_roles

视图applicable_roles标识当前用户可以使用其权限的所有角色。
这意味着从当前用户到相关角色之间存在某条角色授予链。当前用户本身也是一个适用角色。
适用角色的集合通常用于权限检查。

表 34.3. applicable_roles 列

名称	数据类型	描述
grantee	sql_identifier	被授予该角色成员资格的角色名称（可以是当前用户，也可以是在嵌套角色成员资格情况下的另一个角色）
role_name	sql_identifier	一个角色的名称
is_grantable	yes_or_no	如果被授权人在角色上有管理选项则为YES，如果没有则为NO

34.6. attributes

视图attributes包含数据库中定义的组合数据类型的属性的信息。
（注意该视图不给出表列的信息，在 PostgreSQL 语境中表列有时也被称为属性。）

表 34.4. attributes 列

名称	数据类型	描述
udt_catalog	sql_identifier	包含该数据类型的数据库名称（总是当前数据库）
udt_schema	sql_identifier	包含该数据类型的模式名称
udt_name	sql_identifier	数据类型名称
attribute_name	sql_identifier	属性名称
ordinal_position	cardinal_number	属性在该数据类型内部的顺序位置（从 1 开始计算）
attribute_default	character_data	该属性的默认表达式
is_nullable	yes_or_no	如果该属性是可能为空的，值为YES，否则为NO。
data_type	character_data	如果该属性是一个内置类型，此列值为该属性的数据类型；如果该属性是某种数组，此列值为ARRAY（在这种情况下，见视图element_types）；其他情况，此列值为USER-DEFINED（在这种情况下，该类型在attribute_udt_name和相关列中标识）。
character_maximum_length	cardinal_number	如果data_type标识字符类型或位串类型，这里是声明的最大长度；对于所有其他数据类型或未声明最大长度时为空。
character_octet_length	cardinal_number	如果data_type标识字符类型，这里是一个数据以八位字节（字节）计的最大可能长度；对于所有其他数据类型为空。最大八位字节长度取决于声明的字符最大长度（见上文）和服务器编码。
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	Not yet implemented
collation_schema	sql_identifier	Not yet implemented
collation_name	sql_identifier	Not yet implemented
numeric_precision	cardinal_number	如果data_type标识一种数字类型，这列包含这个属性类型的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。

名称	数据类型	描述
numeric_precision_radix	cardinal_number	如果data_type标识一种数字类型，这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。对于所有其他数据类型，这一列为空。
numeric_scale	cardinal_number	如果data_type标识精确数字类型，这一列包含该属性类型的标度（声明的或隐式的）。标度表示小数点右侧的有效数字位数。它可以按列numeric_precision_radix的指定表示为十进制（基数 10）或二进制（基数 2）。对于所有其他数据类型，这一列为空。
datetime_precision	cardinal_number	如果data_type标识一种日期、时间、时间戳或时间间隔类型，这一列包含这个属性类型的（声明的或隐式的）分数秒的精度，也就是秒值的小数点后的十进制位数。对于所有其他数据类型，这一列为空。
interval_type	character_data	Not yet implemented
interval_precision	character_data	Not yet implemented
attribute_udt_catalog	sql_identifier	属性数据类型被定义的数据库名（总是当前数据库）
attribute_udt_schema	sql_identifier	属性数据类型被定义的模式名
attribute_udt_name	sql_identifier	属性数据类型的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该列的数据类型描述符的一个标识符，在从属于该表的数据类型标识符之中唯一。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。
is_derived_reference_attribute	yes_or_no	对应 PostgreSQL 不支持的特性

关于某些列的详情，参见第 34.15 节之下的一个相似结构的视图。

34.7. character_sets

视图character_sets标识当前数据库中可用的字符集。因为 PostgreSQL 不支持在同一个数据库中有多个字符集，这个视图只显示一个字符集，它就是数据库编码。

请注意 SQL 标准中以下术语的用法：

character repertoire

一个抽象的字符集合，例如UNICODE，UCS或 LATIN1。不作为SQL对象公开，但在此视图中可见。

character encoding form

一种字符集的编码形式。大多数旧的字符集只使用一种编码形式，
因此它们没有单独的名称（例如，LATIN1是适用于LATIN1字符集的编码形式）。
但是例如Unicode有编码形式UTF8，UTF16等（并非所有都受PostgreSQL支持）。
编码形式不作为SQL对象公开，但在此视图中可见。

character set

一个命名的SQL对象，用于标识字符集、字符编码和默认排序规则。
预定义的字符集通常与编码形式同名，但用户可以定义其他名称。例如，字符集UTF8
通常会标识字符集 UCS，编码形式UTF8和一些默认排序规则。

在 PostgreSQL 中，可以把“编码”视为字符集，也可以视为字符编码形式。它们具有相同的名称，并且一个数据库中只能有一个。

表 34.5. character_sets 列

名称	数据类型	描述
character_set_catalog	sql_identifier	当前字符集并未被实现为模式对象，因此这一列为空。
character_set_schema	sql_identifier	当前字符集并未被实现为模式对象，因此这一列为空。
character_set_name	sql_identifier	该字符集的名字，当前实现为显示该数据库编码的名字
character_repertoire	sql_identifier	字符集合；如果编码为UTF8则显示UCS，否则仅显示编码名称
form_of_use	sql_identifier	字符编码形式，与数据库编码相同
default_collate_catalog	sql_identifier	包含该默认排序规则的数据库名（如果任意排序规则被标识，总是当前数据库）
default_collate_schema	sql_identifier	包含该默认排序规则的模式名
default_collate_name	sql_identifier	默认排序规则的名字。该默认排序规则被标识为匹配当前数据库的COLLATE和CTYPE设置的排序规则。如果没有这样的排序规则，则这一列以及相关的模式列和目录列都为空。

34.8. check_constraint_routine_usage

视图check_constraint_routine_usage标识检查约束所使用的例程（函数和过程）。
只有当前启用角色拥有的那些例程才会显示出来。

表 34.6. check_constraint_routine_usage 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含约束的数据库名称(总是当前数据库)

名称	数据类型	描述
constraint_schema	sql_identifier	包含约束的模式的名称
constraint_name	sql_identifier	约束的名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 34.38 节。

34.9. check_constraints

视图check_constraints包含所有检查约束，无论它们是定义在表上还是定义在域上，只要它们归当前启用角色所有即可显示（表或域的拥有者就是约束的拥有者）。

表 34.7. check_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含约束的数据库名称(总是当前数据库)
constraint_schema	sql_identifier	包含约束的模式的名称
constraint_name	sql_identifier	约束的名称
check_clause	character_data	该检查约束的检查表达式

34.10. collations

视图collations包含在当前数据库中可用的排序规则。

表 34.8. collations 列

名称	数据类型	描述
collation_catalog	sql_identifier	包含该排序规则的数据库名称（总是当前数据库）
collation_schema	sql_identifier	包含该排序规则的模式名称
collation_name	sql_identifier	默认排序规则的名称
pad_attribute	character_data	总是NO PAD（另一种选择PAD SPACE没有被 PostgreSQL 支持）

34.11. collation_character_set_applicability

视图collation_character_set_applicability标识可用的排序规则适用于哪些字符集。在 PostgreSQL 中，每个数据库中只有一种字符集（解释见第 34.7 节），因此这个视图没有提供很有用的信息。

表 34.9. collation_character_set_applicability 列

名称	数据类型	描述
collation_catalog	sql_identifier	包含该排序规则的数据库名称（总是当前数据库）
collation_schema	sql_identifier	包含该排序规则的模式名称
collation_name	sql_identifier	默认排序规则的名称
character_set_catalog	sql_identifier	字符集当前还未被实现为模式对象，所以这一列为空
character_set_schema	sql_identifier	字符集当前还未被实现为模式对象，所以这一列为空
character_set_name	sql_identifier	字符集名称

34.12. column_domain_usage

视图column_domain_usage
标识所有使用定义在当前数据库中并且被一个当前启用的角色拥有的域的列（表列或视图列）。

表 34.10. column_domain_usage 列

名称	数据类型	描述
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称
column_name	sql_identifier	列名称

34.13. column_privileges

视图column_privileges
标识所有授予给一个当前启用的角色或者被一个当前启用的角色授予的权限。对每一个列、授予者、被授予者的组合只有一行。

如果一个权限被授予在一整个表上，它在这个视图中被显示为在每一列上授予，但是只有可用于列粒度的权限类型才会这样：SELECT、INSERT、UPDATE、REFERENCES。

表 34.11. column_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
table_catalog	sql_identifier	包含该列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该列的表所在的模式名称
table_name	sql_identifier	包含该列的表名称

名称	数据类型	描述
column_name	sql_identifier	列名称
privilege_type	character_data	权限类型：SELECT、INSERT、UPDATE或 REFERENCES
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.14. column_udt_usage

视图column_udt_usage标识所有使用被一个当前启用的角色拥有的数据类型的列。
注意在PostgreSQL中，内置数据类型的行为和用户定义的类型相似，因此它们也被包括在这里。
详见第 34.15 节。

表 34.12. column_udt_usage 列

名称	数据类型	描述
udt_catalog	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的模式名称
udt_name	sql_identifier	该列数据类型（如果适用，底层的域类型）的名称
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称
column_name	sql_identifier	列名称

34.15. columns

视图columns包含数据库中有关所有表列（或视图列）的信息。系统列（oid等）
不被包括在内。只有那些当前用户能够访问（作为拥有者或具有某些权限）的列才被显示。

表 34.13. columns 列

名称	数据类型	描述
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称
column_name	sql_identifier	列名称
ordinal_position	cardinal_number	该列在表内的顺序位置（从 1 开始计）
column_default	character_data	该列的默认表达式
is_nullable	yes_or_no	如果该列可以为空，则为YES，否则为NO。一个非空约束是让一列成为不能为空的方法，但还有其他方法。

名称	数据类型	描述
data_type	character_data	如果该列的数据类型是一种内置类型，则为该列的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况下该类型被标识在udt_name和相关列中）。如果该列基于一个域，这一列引用该域底层的类型（该列被标识在domain_name和相关列中）。
character_maximum_length	cardinal_number	如果data_type标识字符类型或位串类型，这里是声明的最大长度；对于所有其他数据类型或未声明最大长度时为空。
character_octet_length	cardinal_number	如果data_type标识字符类型，这里是一个数据以八位字节（字节）计的最大可能长度；对于所有其他数据类型为空。最大八位字节长度取决于声明的字符最大长度（见上文）和服务器编码。
numeric_precision	cardinal_number	如果data_type标识一种数字类型，这列包含这个属性类型的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
numeric_precision_radix	cardinal_number	如果data_type标识一种数字类型，这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。对于所有其他数据类型，这一列为空。
numeric_scale	cardinal_number	如果data_type标识一种准确数字类型，这列包含这个属性类型的（声明的或隐式的）比例。标度表示小数点右侧的有效数字位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
datetime_precision	cardinal_number	如果data_type标识一种日期、时间、时间戳或时间间隔类型，这一列包含这个属性类型的（声明的或隐式的）分数秒的精度，也就是秒值的小数点后的十进制位数。对于所有其他数据类型，这一列为空。
interval_type	character_data	Not yet implemented
interval_precision	character_data	Not yet implemented
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	Not yet implemented
collation_schema	sql_identifier	Not yet implemented
collation_name	sql_identifier	Not yet implemented
domain_catalog	sql_identifier	如果该列有一个域类型，这里是该域所在的数据库名（总是当前数据库），否则为空。
domain_schema	sql_identifier	如果该列有一个域类型，这里是该域所在的模式名，否则为空。
domain_name	sql_identifier	如果该列有一个域类型，这里是该域的名称，否则为空。
udt_catalog	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的模式名称
udt_name	sql_identifier	该列数据类型（如果适用，底层的域类型）的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该列的数据类型描述符的一个标识符，在从属于该表的数据类型标识符之中唯一。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。
is_self_referencing	yes_or_no	对应 PostgreSQL 不支持的特性
is_identity	yes_or_no	对应 PostgreSQL 不支持的特性
identity_generation	character_data	对应 PostgreSQL 不支持的特性
identity_start	character_data	对应 PostgreSQL 不支持的特性
identity_increment	character_data	对应 PostgreSQL 不支持的特性
identity_maximum	character_data	对应 PostgreSQL 不支持的特性
identity_minimum	character_data	对应 PostgreSQL 不支持的特性
identity_cycle	yes_or_no	对应 PostgreSQL 不支持的特性
is_generated	character_data	对应 PostgreSQL 不支持的特性
generation_expression	character_data	对应 PostgreSQL 不支持的特性
is_updatable	yes_or_no	如果该列是可更新的，则为`YES`，否则为`NO`（基表中的列总是可更新的，视图中的列则不一定）

因为在 SQL 中有多种方式定义数据类型，而PostgreSQL还包含额外的方式来定义数据类型，它们在信息模式中的表示可能有点困难。列`data_type`应该标识列的底层内置类型。

在PostgreSQL中，这表示定义在系统目录模式pg_catalog中的类型。
如果应用能够特别地（例如以不同方式格式化数字类型或使用精度列中的数据）
处理众所周知的内置类型，这列可能会有用。列udt_name、udt_schema和udt_catalog总是
标识列的底层数据类型，即使该列是基于一个域的（因为
PostgreSQL对待内置类型和用户定义类型的方式是一样的，内置类型也出现在这里。这是 SQL
标准的一种扩展）。如果一个应用想要根据该类型以不同的方式处理数据，就应该使用这些列，
因为在那种情况下即使该列真地基于一个域也没有关系。如果该列是基于一个域，
该域的标识被存储在列domain_name、domain_schema和domain_catalog。
如果你想要把列和它们相关的数据类型配对并且把域视作单独的类型，你可以写coalesce(
domain_name, udt_name)等等。

34.16. constraint_column_usage

视图constraint_column_usage标识当前数据库中被某种约束使用的所有列。
只显示那些位于当前启用角色所拥有表中的列。对于检查约束，
此视图标识检查表达式中使用的列。对于非空约束，此视图标识定义该约束的列。
对于外键约束，此视图标识外键所引用的列。对于唯一约束或主键约束，
此视图标识受该约束限制的列。

表 34.14. constraint_column_usage 列

名称	数据类型	描述
table_catalog	sql_identifier	包含被某个约束使用的列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被某个约束使用的列的表所在的模式名称
table_name	sql_identifier	包含被某个约束使用的列的表名称
column_name	sql_identifier	包含被某个约束使用的列名称
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称

34.17. constraint_table_usage

视图constraint_table_usage
标识在当前数据库中被某个约束使用的所有表（这与视图table_constraints不同，
它标识哪些表约束定义在哪些表上）。对于一个外键约束，这个视图标识该外键引用的表。
对于一个唯一或主键约束，这个视图仅标识该约束属于的表。
检查约束和非空约束不被包括在这个视图中。

表 34.15. constraint_table_usage 列

名称	数据类型	描述
table_catalog	sql_identifier	包含被某个约束使用的表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被某个约束使用的表的模式名称
table_name	sql_identifier	包含被某个约束使用的表名称

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称

34.18. data_type_privileges

视图data_type_privileges
标识当前用户能够访问（作为被描述对象的拥有者或者具有其上的某种权限）
的所有数据类型描述符。只要一个数据类型被用在一个表列、
一个域或一个函数（作为参数或返回类型）
就会生成一个数据类型描述符并且在那个实例中存储一些有关该数据类型如何被使用的信息（例如，声明的最大长度，如果适用）。每一个数据类型描述符被赋予一个任意的标识符，
它在被赋予给一个对象（表、域、函数）的数据类型描述符中唯一。
这个视图对于应用可能没什么用，但是它被用于定义信息模式中的一些其他视图。

表 34.16. data_type_privileges 列

名称	数据类型	描述
object_catalog	sql_identifier	包含该被描述对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	包含该被描述对象的模式名称
object_name	sql_identifier	被描述对象的名称
object_type	character_data	被描述对象的类型：TABLE（从属于表的一列的数据类型描述符）、DOMAIN（从属于域的数据类型描述符）、ROUTINE（从属于函数的一个参数或返回数据类型的数据类型描述符）。
dtd_identifier	sql_identifier	数据类型描述符的标识符，它在同一对象的数据类型描述符之间唯一。

34.19. domain_constraints

视图domain_constraints包含当前数据库中定义的域的所有约束。
只显示当前用户是拥有者或对其有某些权限（SELECT 除外）的约束。

表 34.17. domain_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称

名称	数据类型	描述
domain_name	sql_identifier	域名称
is_deferrable	yes_or_no	如果该约束是可延迟的，则为YES，否则为NO
initially_deferred	yes_or_no	如果该约束是可延迟的且初始就被延迟，则为YES，否则为NO

34.20. domain_udt_usage

视图domain_udt_usage标识所有基于被一个当前启用的角色拥有的数据类型的域。
注意在PostgreSQL中，内置数据类型的行为相似于用户定义的类型，因此它们也被包括在这里。

表 34.18. domain_udt_usage 列

名称	数据类型	描述
udt_catalog	sql_identifier	该域数据类型被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该域数据类型被定义的模式名称
udt_name	sql_identifier	该域数据类型的名称
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称

34.21. domains

视图domains包含当前数据库中定义的所有域。
仅显示当前用户有权限访问的域（通过拥有者或某些权限）。

表 34.19. domains 列

名称	数据类型	描述
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称
data_type	character_data	该域的数据类型如果是一种内置类型，这里是该域的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	如果该域有一个字符或位串类型，这里是声明的最大长度；如果没有声明最大长度，则对于所有其他数据类型为空。
character_octet_length	cardinal_number	如果该域有一个字符类型，这里是一个数据的最大可能长度（以字节计）；对其他所有数据类型为空。

名称	数据类型	描述
		最大字节长度取决于声明的字符最大长度（见上文）和服务器编码。
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	Not yet implemented
collation_schema	sql_identifier	Not yet implemented
collation_name	sql_identifier	Not yet implemented
numeric_precision	cardinal_number	如果该域有一种数字类型，这列包含这个域类型的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
numeric_precision_radix	cardinal_number	如果该域有一种数字类型，这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。对于所有其他数据类型，这一列为空。
numeric_scale	cardinal_number	如果该域有一种准确数字类型，这列包含这个域类型的（声明的或隐式的）比例。标度表示小数点右侧的有效数字位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
datetime_precision	cardinal_number	如果data_type标识一种日期、时间、时间戳或时间间隔类型，这一列包含这个域类型的（声明的或隐式的）分数秒的精度，也就是秒值的小数点后的十进制位数。对于所有其他数据类型，这一列为空。
interval_type	character_data	Not yet implemented
interval_precision	character_data	Not yet implemented
domain_default	character_data	该域的默认表达式
udt_catalog	sql_identifier	该域数据类型被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该域数据类型被定义的模式名称
udt_name	sql_identifier	该域数据类型的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制

名称	数据类型	描述
dtd_identifier	sql_identifier	该域的数据类型描述符的一个标识符，在从属于该域的数据类型标识符之中唯一（这不重要，因为一个域只包含一个数据类型描述符）。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。

34.22. element_types

视图element_types包含数组元素的数据类型描述符。当一个表列、域、函数参数或函数返回值被定义为数组类型时，相应的信息模式视图的data_type列只包含ARRAY。要获得数组元素类型的信息，可以把相应的视图与这个视图连接。例如，要显示一个表的列及其数据类型和数组元素类型（如果适用），可以这样做：

```
SELECT c.column_name, c.data_type, e.data_type AS element_type
FROM information_schema.columns c LEFT JOIN information_schema.
element_types e
     ON ((c.table_catalog, c.table_schema, c.table_name, 'TABLE', c.
dtd_identifier)
       = (e.object_catalog, e.object_schema, e.object_name, e.object_
type, e.collection_type_identifier))
WHERE c.table_schema = '...' AND c.table_name = '...'
ORDER BY c.ordinal_position;
```

这个视图只包含当前用户通过作为拥有者或拥有某些权限而可访问的对象。

表 34.20. element_types 列

名称	数据类型	描述
object_catalog	sql_identifier	包含使用被描述的数组的对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	包含使用被描述的数组的对象的模式名称
object_name	sql_identifier	使用被描述的模式的对象名称
object_type	character_data	The type of the object that uses the array being described: one of TABLE (the array is used by a column of that table), DOMAIN (the array is used by that domain), ROUTINE (the array is used by a parameter or the return data type of that function).
collection_type_identifier	sql_identifier	被描述的数组的数据类型描述符的标识符。使用这个去与其他信息模式视图的dtd_identifier列连接。
data_type	character_data	如果数组元素的数据类型是内置类型，这里是数组元素的数据类型，否则为USER-DEFINED（在那种

名称	数据类型	描述
		情况下，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	Not yet implemented
collation_schema	sql_identifier	Not yet implemented
collation_name	sql_identifier	Not yet implemented
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
domain_default	character_data	Not yet implemented
udt_catalog	sql_identifier	元素的数据类型所在的数据库名称（总是当前数据库）
udt_schema	sql_identifier	定义元素的数据类型的模式的名称
udt_name	sql_identifier	元素的数据类型名
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该元素的数据类型描述符的标识符。当前没有用。

34.23. enabled_roles

视图enabled_roles标识当前“已启用的角色”。已启用的角色递归地定义为：当前用户，以及所有通过自动继承方式授予给已启用角色的角色。换句话说，这些角色都是当前用户直接或间接自动继承其成员资格的角色。

在权限检查中，使用的是“可应用角色”集合，它可能比已启用角色的集合更大。因此，一般而言最好使用视图applicable_roles而不是这个视图；另见那里的说明。

表 34.21. enabled_roles 列

名称	数据类型	描述
role_name	sql_identifier	一个角色的名称

34.24. foreign_data_wrapper_options

视图foreign_data_wrapper_options包含为当前数据库中外部数据包装器定义的所有选项。只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部数据包装器被显示。

表 34.22. foreign_data_wrapper_options 列

名称	数据类型	描述
foreign_data_wrapper_catalog	sql_identifier	该外部数据包装器所在的数据库名称（总是当前数据库）
foreign_data_wrapper_name	sql_identifier	该外部数据包装器的名称
option_name	sql_identifier	一个选项名称
option_value	character_data	该选项的值

34.25. foreign_data_wrappers

视图foreign_data_wrappers包含定义在当前数据库中的所有外部数据包装器。只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部数据包装器才会被显示。

表 34.23. foreign_data_wrappers 列

名称	数据类型	描述
foreign_data_wrapper_catalog	sql_identifier	包含该外部数据包装器的数据库名称（总是当前数据库）
foreign_data_wrapper_name	sql_identifier	该外部数据包装器的名称
authorization_identifier	sql_identifier	外部服务器拥有者的名称
library_name	character_data	实现这个外部数据包装器的库文件名称
foreign_data_wrapper_language	character_data	用于实现这个外部数据包装器的语言

34.26. foreign_server_options

视图foreign_server_options包含为当前数据库中外部服务器定义的所有选项。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部服务器才会被显示。

表 34.24. foreign_server_options 列

名称	数据类型	描述
foreign_server_catalog	sql_identifier	该外部服务器所在的数据库名称（总是当前数据库）
foreign_server_name	sql_identifier	该外部服务器的名称
option_name	sql_identifier	一个选项名称
option_value	character_data	该选项的值

34.27. foreign_servers

视图foreign_servers包含当前数据库中定义的所有外部服务器。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部服务器才会被显示。

表 34.25. foreign_servers 列

名称	数据类型	描述
foreign_server_catalog	sql_identifier	该外部服务器所在的数据库名称（总是当前数据库）
foreign_server_name	sql_identifier	该外部服务器的名称
foreign_data_wrapper_catalog	sql_identifier	包含被该外部服务器使用的外部数据包装器的数据库名称（总是当前数据库）
foreign_data_wrapper_name	sql_identifier	被该外部服务器所使用的外部数据包装器的名称
foreign_server_type	character_data	外部服务器类型信息，如果在创建时指定过
foreign_server_version	character_data	外部服务器版本信息，如果在创建时指定过
authorization_identifier	sql_identifier	外部服务器拥有者的名称

34.28. foreign_table_options

视图foreign_table_options包含为当前数据库中外部表定义的所有选项。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部表才会被显示。

表 34.26. foreign_table_options 列

名称	数据类型	描述
foreign_table_catalog	sql_identifier	包含该外部表的数据库名称（总是当前数据库）
foreign_table_schema	sql_identifier	包含该外部表的模式名称
foreign_table_name	sql_identifier	外部表的名字
option_name	sql_identifier	一个选项名称

名称	数据类型	描述
option_value	character_data	该选项的值

34.29. foreign_tables

视图foreign_tables包含定义在当前数据库中的所有外部表。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的外部表才会被显示。

表 34.27. foreign_tables 列

名称	数据类型	描述
foreign_table_catalog	sql_identifier	该外部表所在的数据库名称（总是当前数据库）
foreign_table_schema	sql_identifier	包含该外部表的模式名称
foreign_table_name	sql_identifier	外部表的名字
foreign_server_catalog	sql_identifier	该外部服务器所在的数据库名称（总是当前数据库）
foreign_server_name	sql_identifier	该外部服务器的名称

34.30. key_column_usage

视图key_column_usage标识当前数据库中所有被某种唯一、主键或外键约束限制的列。
检查约束不被包括在这个视图中。
只有那些当前用户能够访问的列才会被显示（作为拥有者或具有某些权限）。

表 34.28. key_column_usage 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
table_catalog	sql_identifier	包含被这个约束限制的列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被这个约束限制的列的表所在的模式名称
table_name	sql_identifier	包含被这个约束限制的列的表的名称
column_name	sql_identifier	被这个约束限制的列的名称
ordinal_position	cardinal_number	该列在约束键中的顺序位置（从 1 开始计数）
position_in_unique_constraint	cardinal_number	对于一个外键约束，被引用行在其唯一约束中的顺序位置（从 1 开始计数）；对于其他约束为空

34.31. parameters

视图parameters包含当前数据库中所有函数的参数的有关信息。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的函数才会被显示。

表 34.29. parameters 列

名称	数据类型	描述
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 34.38 节。
ordinal_position	cardinal_number	该参数在函数参数列表中的顺序位置（从 1 开始计数）
parameter_mode	character_data	IN表示输入参数，OUT表示输出参数，INOUT表示输入/输出参数。
is_result	yes_or_no	对应 PostgreSQL 不支持的特性
as_locator	yes_or_no	对应 PostgreSQL 不支持的特性
parameter_name	sql_identifier	参数名称，如果参数没有名称则为空
data_type	character_data	该参数的数据类型如果是一种内置类型，这里是该参数的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
collation_schema	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
collation_name	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型

名称	数据类型	描述
udt_catalog	sql_identifier	该参数的数据类型所在的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该参数的数据类型所在的模式名称
udt_name	sql_identifier	该参数的数据类型的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该参数的数据类型描述符的一个标识符，在从属于该函数的数据类型标识符之中唯一（这不重要，因为一个域只包含一个数据类型描述符）。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。

34.32. referential_constraints

视图referential_constraints包含当前数据库中的所有引用（外键）约束。
只显示当前用户对引用表有写访问权限（通过作为拥有者或拥有 SELECT 之外的某些权限）的约束。

表 34.30. referential_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含约束的数据库名称(总是当前数据库)
constraint_schema	sql_identifier	包含约束的模式的名称
constraint_name	sql_identifier	约束的名称
unique_constraint_catalog	sql_identifier	包含该外键约束所引用的唯一或主键约束的数据库名称（总是当前数据库）
unique_constraint_schema	sql_identifier	包含该外键约束所引用的唯一或主键约束的模式名称
unique_constraint_name	sql_identifier	包含该外键约束所引用的唯一或主键约束的名称
match_option	character_data	外键约束的匹配选项：FULL、PARTIAL或NONE。
update_rule	character_data	外键约束的更新规则：CASCADE、SET NULL、SET DEFAULT、RESTRICT或 NO ACTION。
delete_rule	character_data	外键约束的删除规则：CASCADE、SET NULL、SET DEFAULT、RESTRICT或 NO ACTION。

34.33. role_column_grants

视图role_column_grants标识授予者或被授予者是当前启用角色的所有列上的权限。更多信息可以在column_privileges下找到。这个视图与column_privileges之间唯一的有效区别是，这个视图省略了通过授予 public 而使当前用户可访问的列。

表 34.31. role_column_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
table_catalog	sql_identifier	包含该列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该列的表所在的模式名称
table_name	sql_identifier	包含该列的表名称
column_name	sql_identifier	列名称
privilege_type	character_data	权限类型：SELECT、INSERT、UPDATE或 REFERENCES
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.34. role_routine_grants

视图role_routine_grants标识授予者或被授予者是当前启用角色的所有函数上的权限。更多信息可以在routine_privileges下找到。这个视图与routine_privileges之间唯一的有效区别是，这个视图省略了通过授予 public 而使当前用户可访问的函数。

表 34.32. role_routine_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 34.38 节。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）
privilege_type	character_data	总是为EXECUTE（函数唯一的权限类型）
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.35. role_table_grants

视图role_table_grants标识授予者或被授予者是当前启用角色的所有表或视图上的权限。更多信息可以在table_privileges下找到。这个视图与table_privileges之间唯一的有效区别是，这个视图省略了通过授予 public 而使当前用户可访问的表。

表 34.33. role_table_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
privilege_type	character_data	该权限的类型：SELECT、INSERT、UPDATE、DELETE、TRUNCATE、REFERENCES或TRIGGER
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO
with_hierarchy	yes_or_no	对应 PostgreSQL 不支持的特性

34.36. role_usage_grants

视图role_usage_grants标识授予者或被授予者是当前启用角色的各类对象上的USAGE权限。更多信息可以在usage_privileges下找到。这个视图与usage_privileges之间唯一的有效区别是，这个视图省略了通过授予 public 而使当前用户可访问的对象。

表 34.34. role_usage_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
object_catalog	sql_identifier	包含该对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	如果适用，则为包含该对象的模式名称，否则为一个空字符串
object_name	sql_identifier	对象的名称
object_type	character_data	COLLATION、DOMAIN、FOREIGN DATA WRAPPER或FOREIGN SERVER
privilege_type	character_data	总是USAGE
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.37. routine_privileges

视图routine_privileges标识所有在函数上授予的权限，
其授予者或被授予者是当前已启用的角色。对于每一种函数、授予者和被授予者的组合，
这里都有一行。

表 34.35. routine_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 34.38 节。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）
privilege_type	character_data	总是为EXECUTE（函数唯一的权限类型）
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.38. routines

视图routines包含当前数据库中所有的函数。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的函数才会被显示。

表 34.36. routines 列

名称	数据类型	描述
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	该函数的“专用名”。这是一个在模式中唯一标识该函数的名称，即使该函数真正的名称已经被重载。专用名的格式尚未被定义，它应当仅被用来与指定例程名称的其他实例进行比较。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）

名称	数据类型	描述
routine_type	character_data	总是FUNCTION（将来可能会有其他类型的例程。）
module_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
module_schema	sql_identifier	对应 PostgreSQL 不支持的特性
module_name	sql_identifier	对应 PostgreSQL 不支持的特性
udt_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
udt_schema	sql_identifier	对应 PostgreSQL 不支持的特性
udt_name	sql_identifier	对应 PostgreSQL 不支持的特性
data_type	character_data	该函数的返回数据类型如果是一种内置类型，这里是该数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在type_udt_name和相关列中）。
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
collation_schema	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
collation_name	sql_identifier	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
type_udt_catalog	sql_identifier	该函数的返回数据类型所在的数据库名（总是当前数据库）。
type_udt_schema	sql_identifier	该函数的返回数据类型所在的模式名。
type_udt_name	sql_identifier	该函数的返回数据类型的名字。
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该函数返回数据类型的数据类型描述符的一个标识符，在从属于该函数的数据类型标识符之中唯一。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。
routine_body	character_data	如果该函数是一个 SQL 函数，则为SQL，否则为EXTERNAL。
routine_definition	character_data	该函数的源文本（如果该函数不归当前已启用的角色所有，则为空）（根据 SQL 标准，只有routine_body为SQL时这一列才适用。但是在PostgreSQL中，它将会包含该函数被创建时所指定的任何源文本。）
external_name	character_data	如果这个函数是一个 C 函数，则为该函数的外部名称（链接符号），否则为空（这会产生和显示在routine_definition中相同的值）。
external_language	character_data	该函数所用的语言
parameter_style	character_data	总是GENERAL（SQL 标准定义了其他参数风格，但在PostgreSQL中不可用）
is_deterministic	yes_or_no	如果该函数被声明为不变（在 SQL 标准中被称为确定性的），则为YES，否则为NO（你不能通过该信息模式查询在PostgreSQL中可用的其他易变性级别）。
sql_data_access	character_data	总是MODIFIES，表示该函数可能修改 SQL 数据。这种信息对PostgreSQL没有用处。
is_null_call	yes_or_no	如果该函数在任一参数为空时自动返回空值，则为YES，否则为NO。
sql_path	character_data	对应 PostgreSQL 不支持的特性
schema_level_routine	yes_or_no	总是YES（反例是一个用户定义类型的方法，这是在PostgreSQL不可用的一种特性）。
max_dynamic_result_sets	cardinal_number	对应 PostgreSQL 不支持的特性
is_user_defined_cast	yes_or_no	对应 PostgreSQL 不支持的特性
is_implicitly_invocable	yes_or_no	对应 PostgreSQL 不支持的特性
security_type	character_data	如果该函数以当前用户的权限运行，则为INVOKER；如果该函数以定义它的用户的权限运行，则为DEFINER。

名称	数据类型	描述
to_sql_specific_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
to_sql_specific_schema	sql_identifier	对应 PostgreSQL 不支持的特性
to_sql_specific_name	sql_identifier	对应 PostgreSQL 不支持的特性
as_locator	yes_or_no	对应 PostgreSQL 不支持的特性
created	time_stamp	对应 PostgreSQL 不支持的特性
last_altered	time_stamp	对应 PostgreSQL 不支持的特性
new_savepoint_level	yes_or_no	对应 PostgreSQL 不支持的特性
is_udt_dependent	yes_or_no	对应 PostgreSQL 不支持的特性
result_cast_from_data_type	character_data	对应 PostgreSQL 不支持的特性
result_cast_as_locator	yes_or_no	对应 PostgreSQL 不支持的特性
result_cast_char_max_length	cardinal_number	对应 PostgreSQL 不支持的特性
result_cast_char_octet_length	character_data	对应 PostgreSQL 不支持的特性
result_cast_char_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_char_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_char_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_numeric_precision	cardinal_number	对应 PostgreSQL 不支持的特性
result_cast_numeric_precision_radix	cardinal_number	对应 PostgreSQL 不支持的特性
result_cast_numeric_scale	cardinal_number	对应 PostgreSQL 不支持的特性
result_cast_datetime_precision	character_data	对应 PostgreSQL 不支持的特性
result_cast_interval_type	character_data	对应 PostgreSQL 不支持的特性
result_cast_interval_precision	character_data	对应 PostgreSQL 不支持的特性
result_cast_type_udt_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_type_udt_schema	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_type_udt_name	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
result_cast_scope_name	sql_identifier	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
result_cast_maximum_cardinality	cardinal_number	对应 PostgreSQL 不支持的特性
result_cast_dtd_identifier	sql_identifier	对应 PostgreSQL 不支持的特性

34.39. schemata

视图schemata包含当前数据库中由当前启用角色拥有的所有模式。

表 34.37. schemata 列

名称	数据类型	描述
catalog_name	sql_identifier	该模式所在的数据库名称（总是当前数据库）
schema_name	sql_identifier	模式的名称
schema_owner	sql_identifier	模式拥有者的名称
default_character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
default_character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
default_character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
sql_path	character_data	对应 PostgreSQL 不支持的特性

34.40. sequences

视图sequences包含所有定义在当前数据库中的序列。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的序列才会被显示。

表 34.38. sequences 列

名称	数据类型	描述
sequence_catalog	sql_identifier	包含该序列的数据库名称（总是当前数据库）
sequence_schema	sql_identifier	包含该序列的模式名称
sequence_name	sql_identifier	序列的名称
data_type	character_data	序列的数据类型。在 PostgreSQL 中，它目前总是 bigint。
numeric_precision	cardinal_number	这列包含这个序列数据类型（见上文）的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。
numeric_precision_radix	cardinal_number	这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。

名称	数据类型	描述
numeric_scale	cardinal_number	这一列包含该序列数据类型（见上文）的标度（声明的或隐式的）。标度表示小数点右侧的有效数字位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。
start_value	character_data	该序列的开始值
minimum_value	character_data	该序列的最小值
maximum_value	character_data	该序列的最大值
increment	character_data	该序列的增量
cycle_option	yes_or_no	如果该序列会循环，则为YES，否则为NO

注意依照 SQL 标准，开始值、最小值、最大值和增量值被作为字符串返回。

34.41. sql_features

表sql_features包含的信息指示了哪些 SQL 标准中定义的正式特性被PostgreSQL所支持。这和附录 D中的信息一样。这里你也能找到一些额外的背景信息。

表 34.39. sql_features 列

名称	数据类型	描述
feature_id	character_data	特性的标识符字符串
feature_name	character_data	特性的描述性名称
sub_feature_id	character_data	该子特性的标识符字符串，或者如果不是一个子特性则为一个长度为零的字符串
sub_feature_name	character_data	该子特性的描述性名称，或者如果不是一个子特性则为一个长度为零的字符串
is_supported	yes_or_no	如果当前版本的 PostgreSQL 完全支持该特性，则为YES，否则为NO
is_verified_by	character_data	总是为空，因为PostgreSQL开发组不进行特性一致性的正式测试
comments	character_data	可能会是关于该特性被支持状态的一段注释

34.42. sql_implementation_info

表sql_implementation_info包含的信息指示剩下的由 SQL 标准实现定义的多个方面。这类信息主要用来在 ODBC 接口的情境中使用；其它接口的用户可能将发现这类信息用处不大。由于这个原因，个体实现信息项没有在这里描述，你将会在 ODBC 接口的描述中找到它们。

表 34.40. sql_implementation_info 列

名称	数据类型	描述
implementation_info_id	character_data	实现信息项的标识符字符串

名称	数据类型	描述
implementation_info_name	character_data	实现信息项的描述性名称
integer_value	cardinal_number	实现信息项的整数值；如果该值在列character_value中给出，则本列为空
character_value	character_data	实现信息项的字符值；如果该值在列integer_value中给出，则本列为空
comments	character_data	可能是从属于该实现信息项的一段注释

34.43. sql_languages

表sql_languages为PostgreSQL支持的每一种 SQL 语言绑定包含一行。PostgreSQL支持直接 SQL 和 C 中的嵌入式 SQL；从此表中只能了解到这些信息。

此表已在 SQL:2008 中从 SQL 标准移除，因此没有引用 SQL:2003 之后标准的条目。

表 34.41. sql_languages 列

名称	数据类型	描述
sql_language_source	character_data	语言定义来源的名称；总是ISO 9075，即 SQL 标准
sql_language_year	character_data	sql_language_source引用的标准获批的年份。
sql_language_conformance	character_data	该语言绑定的标准符合性级别。对于 ISO 9075:2003，此值总是CORE。
sql_language_integrity	character_data	总是为空（此值与 SQL 标准的较早版本有关。）
sql_language_implementation	character_data	总是为空
sql_language_binding_style	character_data	语言绑定方式，为DIRECT或EMBEDDED
sql_language_programming_language	character_data	如果绑定方式为EMBEDDED，则为所用编程语言，否则为空。PostgreSQL只支持 C 语言。

34.44. sql_packages

表sql_packages包含PostgreSQL支持 SQL 标准所定义的哪些功能包的信息。功能包的背景信息请参阅附录 D。

表 34.42. sql_packages 列

名称	数据类型	描述
feature_id	character_data	功能包的标识符字符串
feature_name	character_data	功能包的描述性名称

名称	数据类型	描述
is_supported	yes_or_no	如果当前版本的PostgreSQL完全支持此功能包，则为YES，否则为NO
is_verified_by	character_data	总是为空，因为PostgreSQL开发组不进行特性一致性的正式测试
comments	character_data	关于此功能包支持状态的可选注释

34.45. sql_parts

表sql_parts包含的信息指示哪些定义在 SQL 标准中的部分被PostgreSQL支持。

表 34.43. sql_parts 列

名称	数据类型	描述
feature_id	character_data	包含该部分编号的一个标识符字符串
feature_name	character_data	该部分的描述性名称
is_supported	yes_or_no	如果当前版本的 PostgreSQL 完全支持该部分，则为YES，否则为NO
is_verified_by	character_data	总是为空，因为PostgreSQL开发组不进行特性一致性的正式测试
comments	character_data	可能会是关于该部分被支持状态的一段注释

34.46. sql_sizing

表sql_sizing包含有关PostgreSQL中多种尺寸限制和最大值的信息。这类信息主要用来在 ODBC 接口的情境中使用；其它接口的用户可能将发现这类信息用处不大。由于这个原因，个体实现信息项没有在这里描述，你将会在 ODBC 接口的描述中找到它们。

表 34.44. sql_sizing 列

名称	数据类型	描述
sizing_id	cardinal_number	尺寸项的标识符
sizing_name	character_data	尺寸项的描述性名称
supported_value	cardinal_number	尺寸项的值，如果尺寸是不受限制或不能确定的则为 0，如果尺寸项适用的特性不受支持则为空
comments	character_data	可能是从属于尺寸项的一段注释

34.47. sql_sizing_profiles

表sql_sizing_profiles包含 SQL 标准的各类配置规范所要求的sql_sizing值的信息。PostgreSQL不跟踪任何 SQL 配置规范，因此此表为空。

表 34.45. sql_sizing_profiles 列

名称	数据类型	描述
sizing_id	cardinal_number	尺寸项的标识符
sizing_name	character_data	尺寸项的描述性名称
profile_id	character_data	配置规范的标识符字符串
required_value	cardinal_number	SQL 配置规范对该尺寸限制项要求的值；如果配置规范对此项不设限制，则为 0；如果配置规范不要求此项所适用的任何功能，则为空
comments	character_data	关于此配置规范中该尺寸限制项的可选注释

34.48. table_constraints

视图table_constraints包含当前用户拥有的或对其有某些非 SELECT 权限的表的所有约束。

表 34.46. table_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
constraint_type	character_data	约束类型：CHECK（包括非空约束）、FOREIGN KEY、PRIMARY KEY 或 UNIQUE
is_deferrable	yes_or_no	如果该约束是可延迟的，则为YES，否则为NO
initially_deferred	yes_or_no	如果该约束是可延迟的且初始就被延迟，则为YES，否则为NO

34.49. table_privileges

视图table_privileges标识在表或视图上所有被授予的权限，
这些权限必须是被当前已启用角色授出或者被授予给当前已启用角色。对每一个表、
授予者和被授予者的组合都有一行。

表 34.47. table_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	被授予该权限的角色名称

名称	数据类型	描述
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
privilege_type	character_data	该权限的类型：SELECT、INSERT、UPDATE、DELETE、TRUNCATE、REFERENCES或TRIGGER
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO
with_hierarchy	yes_or_no	对应 PostgreSQL 不支持的特性

34.50. tables

视图tables包含定义在当前数据库中的所有表和视图。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的表和视图才会被显示。

表 34.48. tables 列

名称	数据类型	描述
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
table_type	character_data	该表的类型：BASE TABLE表示一个持久的基本表（常见表类型），VIEW表示一个视图，FOREIGN TABLE表示一个外部表，LOCAL TEMPORARY表示一个临时表
self_referencing_column_name	sql_identifier	对应 PostgreSQL 不支持的特性
reference_generation	character_data	对应 PostgreSQL 不支持的特性
user_defined_type_catalog	sql_identifier	如果该表是一个有类型的表，则是包含其底层数据类型的数据库名称（总是当前数据库），否则为空。
user_defined_type_schema	sql_identifier	如果该表是一个有类型的表，则是包含其底层数据类型的模式名，否则为空。
user_defined_type_name	sql_identifier	如果该表是一个有类型的表，则是其底层数据类型的名称，否则为空。
is_insertable_into	yes_or_no	如果该表能够被插入，则为YES，否则为NO（基本表总是能被插入，而视图则不一定）。
is_typed	yes_or_no	如果该表是一个有类型的表，则为YES，否则为NO
commit_action	character_data	如果该表是一个临时表，则为PRESERVE，否则为空。（SQL 标

名称	数据类型	描述
		准为临时表定义了其他提交动作，但PostgreSQL不支持它们。）

34.51. triggered_update_columns

对于当前数据库中指定了列列表（如UPDATE OF column1, column2）的触发器，视图triggered_update_columns标识这些列。不指定列列表的触发器不包含在这个视图中。只显示当前用户拥有的或对其有某些非 SELECT 权限的列。

表 34.49. triggered_update_columns 列

名称	数据类型	描述
trigger_catalog	sql_identifier	包含该触发器的数据库名称（总是当前数据库）
trigger_schema	sql_identifier	包含该触发器的模式名称
trigger_name	sql_identifier	触发器的名称
event_object_catalog	sql_identifier	包含触发器所在的表的数据库名称（总是当前数据库）
event_object_schema	sql_identifier	包含触发器所在的表的模式名称
event_object_table	sql_identifier	触发器所在的表的名称
event_object_column	sql_identifier	触发器所在的列的名称

34.52. triggers

视图triggers包含当前数据库中定义在当前用户拥有的或对其有某些非 SELECT 权限的表和视图上的所有触发器。

表 34.50. triggers 列

名称	数据类型	描述
trigger_catalog	sql_identifier	包含该触发器的数据库名称（总是当前数据库）
trigger_schema	sql_identifier	包含该触发器的模式名称
trigger_name	sql_identifier	触发器的名称
event_manipulation	character_data	触发该触发器的事件（INSERT、UPDATE或DELETE）
event_object_catalog	sql_identifier	包含触发器所在的表的数据库名称（总是当前数据库）
event_object_schema	sql_identifier	包含触发器所在的表的模式名称
event_object_table	sql_identifier	触发器所在的表的名称
action_order	cardinal_number	Not yet implemented
action_condition	character_data	触发器的WHEN条件，如果没有则为空（如果该表不被当前已启用角色拥有也是为空）
action_statement	character_data	触发器执行的语句（当前总是EXECUTE PROCEDURE *function*(...)）

名称	数据类型	描述
action_orientation	character_data	标识触发器是对每个被处理的行触发一次还是为每个语句触发一次（ROW或STATEMENT）
action_timing	character_data	触发器在什么时候触发（BEFORE、AFTER或INSTEAD OF）
action_reference_old_table	sql_identifier	对应 PostgreSQL 不支持的特性
action_reference_new_table	sql_identifier	对应 PostgreSQL 不支持的特性
action_reference_old_row	sql_identifier	对应 PostgreSQL 不支持的特性
action_reference_new_row	sql_identifier	对应 PostgreSQL 不支持的特性
created	time_stamp	对应 PostgreSQL 不支持的特性

PostgreSQL中的触发器有两点与 SQL 标准不兼容，这会影响在该信息模式中的表示。第一，在PostgreSQL中触发器的名字是局限于每个表的，而不是独立于模式对象。因此可能在一个模式中会有重复的触发器名称，只要它们属于不同的表（trigger_catalog和trigger_schema才真正标识了触发器被定义在哪个表上）。第二，在PostgreSQL中触发器可以被定义为在多个事件上触发（例如ON INSERT OR UPDATE），而在SQL 标准中只允许一个。如果一个触发器被定义为在多个事件上触发，它在信息模式中被表示为多行，每一行对应于一类事件。作为这两个问题的结果，视图triggers的主键实际上是(trigger_catalog, trigger_schema, event_object_table, trigger_name, event_manipulation)，而不是(trigger_catalog, trigger_schema, trigger_name)（这是 SQL 标准指定的）。尽管如此，如果你以符合 SQL 标准（在模式中触发器名称唯一并且每个触发器只能有一种事件类型）的方式定义你的触发器，这将不会影响你。

注意

在PostgreSQL 9.1 之前，这个视图的列 action_timing、action_reference_old_table、action_reference_new_table、action_reference_old_row和action_reference_new_row 分别被命名为 condition_timing、condition_reference_old_table、condition_reference_new_table、condition_reference_old_row和 condition_reference_new_row。那也是它们在SQL:1999 标准中的命名。新的命名遵循 SQL:2003 及其后的版本。

34.53. usage_privileges

视图usage_privileges标识各类对象上授予当前启用角色或由当前启用角色授予的USAGE权限。在PostgreSQL中，这当前适用于排序规则、域、外部数据包装器和外部服务器。每种对象、授予者和被授予者的组合各有一行。

由于排序规则和域在PostgreSQL中没有真正的权限，这个视图显示的是拥有者为所有排序规则和域隐式授予PUBLIC的不可授予的USAGE权限。而其他对象类型显示的是真正的权限。

表 34.51. usage_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称

名称	数据类型	描述
grantee	sql_identifier	被授予该权限的角色名称
object_catalog	sql_identifier	包含该对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	如果适用，则为包含该对象的模式名称，否则为一个空字符串
object_name	sql_identifier	对象的名称
object_type	character_data	COLLATION、DOMAIN、FOREIGN DATA WRAPPER或FOREIGN SERVER
privilege_type	character_data	总是USAGE
is_grantable	yes_or_no	如果该权限是可授予的，则为YES，否则为NO

34.54. user_mapping_options

视图user_mapping_options包含在当前数据库中为用户映射定义的所有选项。只有那些当前用户能够访问其相应外部服务器（作为拥有者或具有某些权限）的用户映射才会被显示。

表 34.52. user_mapping_options 列

名称	数据类型	描述
authorization_identifier	sql_identifier	被映射的用户名称，如果映射是公共的则为PUBLIC
foreign_server_catalog	sql_identifier	这个映射所使用的外部服务器所在的数据库名称（总是当前数据库）
foreign_server_name	sql_identifier	这个映射所使用的外部服务器的名称
option_name	sql_identifier	一个选项名称
option_value	character_data	选项的值。除非当前用户是被映射的用户或者映射是PUBLIC的并且当前用户是服务器拥有者或者超级用户，这一列将显示为空。这样做的目的是保护作为用户映射选项存储的密码信息。

34.55. user_mappings

视图user_mappings包含定义在当前数据库中的所有用户映射。只有当前用户能够访问其对应外部服务器（作为拥有者或具有某些权限）的用户映射才会被显示。

表 34.53. user_mappings 列

名称	数据类型	描述
authorization_identifier	sql_identifier	被映射的用户名称，如果映射是公共的则为PUBLIC

名称	数据类型	描述
foreign_server_catalog	sql_identifier	这个映射所使用的外部服务器所在的数据库名称（总是当前数据库）
foreign_server_name	sql_identifier	这个映射所使用的外部服务器的名称

34.56. view_column_usage

视图view_column_usage标识视图查询表达式（即定义该视图的 SELECT 语句）中使用的所有列。只有当包含该列的表归当前已启用角色所有时，该列才会出现在这个视图中。

> **注意**
>
> 系统表列不被包括。在某个时候这应该会被修复。

表 34.54. view_column_usage 列

名称	数据类型	描述
view_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
view_schema	sql_identifier	包含该视图的模式名称
view_name	sql_identifier	视图的名称
table_catalog	sql_identifier	被视图所使用的列所属表的数据库名称（总是当前数据库）
table_schema	sql_identifier	被视图所使用的列所属表的模式名称
table_name	sql_identifier	被视图所使用的列所属表的名称
column_name	sql_identifier	被该视图所使用的列名称

34.57. view_routine_usage

视图view_routine_usage标识视图查询表达式（即定义该视图的 SELECT 语句）中使用的所有例程（函数和过程）。只有归当前已启用角色所有的例程才会出现在这个视图中。

表 34.55. view_routine_usage 列

名称	数据类型	描述
table_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该视图的模式名称
table_name	sql_identifier	视图的名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 34.38 节。

34.58. view_table_usage

视图view_table_usage标识视图查询表达式（即定义该视图的 SELECT 语句）中使用的所有表。只有归当前已启用角色所有的表才会出现在这个视图中。

> **注意**
>
> 系统表没有被包括。这应当会在某个时候被修复。

表 34.56. view_table_usage 列

名称	数据类型	描述
view_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
view_schema	sql_identifier	包含该视图的模式名称
view_name	sql_identifier	视图的名称
table_catalog	sql_identifier	包含被该视图所使用的表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被该视图所使用的表的模式名称
table_name	sql_identifier	包含被该视图所使用的表的名称

34.59. views

视图views包含定义在当前数据库中的所有视图。
只有当前用户能够访问（作为拥有者或具有某些权限）的视图才会被显示。

表 34.57. views 列

名称	数据类型	描述
table_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该视图的模式名称
table_name	sql_identifier	视图的名称
view_definition	character_data	定义视图的查询表达式（如果该视图不被当前已启用角色拥有则为空）
check_option	character_data	对应 PostgreSQL 不支持的特性
is_updatable	yes_or_no	如果视图可更新（允许UPDATE和DELETE），则为YES，否则为NO
is_insertable_into	yes_or_no	如果视图可插入（允许INSERT），则为YES，否则为NO
is_trigger_updatable	yes_or_no	如果该视图上定义了 INSTEAD OF UPDATE 触发器，则为YES，否则为NO

名称	数据类型	描述
is_trigger_deletable	yes_or_no	如果该视图上定义了 INSTEAD OF DELETE 触发器，则为YES，否则为NO
is_trigger_insertable_into	yes_or_no	如果该视图上定义了 INSTEAD OF INSERT 触发器，则为YES，否则为NO

部分 V. 服务器编程

本部分介绍如何使用用户定义函数、数据类型、触发器等扩展服务器功能。这些属于高级主题，
通常应在读者已经理解其他有关PostgreSQL的用户文档之后再来学习。
本部分后面的章节介绍PostgreSQL发行版中提供的服务器端编程语言，
以及与服务器端编程有关的一般问题。在深入学习服务器端编程的内容之前，至少应先阅读第 35 章的前几节（涵盖函数）。

目录

35. 扩展 SQL 778
35.1. 可扩展性如何运作 778
35.2. PostgreSQL类型系统 778
35.2.1. 基础类型 778
35.2.2. 复合类型 778
35.2.3. 域 778
35.2.4. 伪类型 779
35.2.5. 多态类型 779
35.3. 用户定义的函数 779
35.4. 查询语言（SQL）函数 780
35.4.1. 基类型上的SQL函数 781
35.4.2. 复合类型上的SQL函数 782
35.4.3. 带参数名的 SQL 函数 785
35.4.4. 带输出参数的 SQL 函数 785
35.4.5. 带有可变数量参数的SQL函数 786
35.4.6. 带有参数默认值的SQL函数 787
35.4.7. SQL 函数作为表来源 788
35.4.8. 返回集合的SQL函数 788
35.4.9. 返回TABLE的SQL函数 790
35.4.10. 多态SQL函数 790
35.4.11. 带有排序规则的SQL函数 792
35.5. 函数重载 792
35.6. 函数易变性分类 793
35.7. 过程语言函数 794
35.8. 内部函数 794
35.9. C 语言函数 795
35.9.1. 动态载入 795
35.9.2. C 语言函数中的基础类型 796
35.9.3. 版本 0 调用约定 799
35.9.4. 版本 1 的调用约定 801
35.9.5. 编写代码 803
35.9.6. 编译和链接动态装载函数 804
35.9.7. 复合类型参数 807
35.9.8. 返回行（复合类型） 808
35.9.9. 返回集合 810
35.9.10. 多态参数和返回类型 815
35.9.11. 共享内存与 LWLocks 816
35.9.12. 把 C++ 用于可扩展性 817
35.10. 用户定义的聚合 817
35.11. 用户定义的类型 819
35.12. 用户定义的操作符 822
35.13. 操作符优化信息 823
35.13.1. COMMUTATOR 823
35.13.2. NEGATOR 824
35.13.3. RESTRICT 824
35.13.4. JOIN 825
35.13.5. HASHES 825
35.13.6. MERGES 826
35.14. 索引扩展接口 827
35.14.1. 索引方法和操作符类 827
35.14.2. 索引方法策略 828

35.14.3. 索引方法支持例程 ... 829
35.14.4. 示例 ... 831
35.14.5. 操作符类和操作符族 ... 833
35.14.6. 系统对操作符类的依赖 ... 835
35.14.7. 排序操作符 ... 836
35.14.8. 操作符类的特殊特性 ... 836
35.15. 将相关对象打包成扩展 ... 837
35.15.1. 扩展文件 ... 838
35.15.2. 扩展的可重定位性 ... 839
35.15.3. 扩展配置表 ... 840
35.15.4. 扩展更新 ... 841
35.15.5. 扩展示例 ... 841
35.16. 扩展构建基础设施 ... 842
36. 触发器 ... 846
36.1. 触发器行为概述 ... 846
36.2. 数据更改的可见性 ... 848
36.3. 用 C 编写触发器函数 ... 848
36.4. 一个完整的触发器示例 ... 851
37. 规则系统 ... 855
37.1. 查询树 ... 855
37.2. 视图和规则系统 ... 856
37.2.1. SELECT 规则如何工作 ... 857
37.2.2. 非 SELECT 语句中的视图规则 ... 861
37.2.3. PostgreSQL 中视图的威力 ... 862
37.2.4. 更新视图 ... 862
37.3. INSERT、UPDATE和DELETE规则 ... 863
37.3.1. 更新规则如何工作 ... 863
37.3.2. 与视图的协作 ... 867
37.4. 规则和权限 ... 873
37.5. 规则和命令状态 ... 874
37.6. 规则与触发器 ... 875
38. 过程语言 ... 878
38.1. 安装过程语言 ... 878
39. PL/pgSQL − SQL 过程语言 ... 881
39.1. 概述 ... 881
39.1.1. 使用PL/pgSQL的优点 ... 881
39.1.2. 支持的参数和结果数据类型 ... 881
39.2. PL/pgSQL的结构 ... 882
39.3. 声明 ... 883
39.3.1. 声明函数参数 ... 884
39.3.2. ALIAS ... 886
39.3.3. 复制类型 ... 886
39.3.4. 行类型 ... 887
39.3.5. 记录类型 ... 887
39.3.6. PL/pgSQL变量的排序规则 ... 887
39.4. 表达式 ... 888
39.5. 基本语句 ... 889
39.5.1. 赋值 ... 889
39.5.2. 执行没有结果的命令 ... 889
39.5.3. 执行返回单行结果的查询 ... 890
39.5.4. 执行动态命令 ... 891
39.5.5. 获取结果状态 ... 894
39.5.6. 什么也不做 ... 895

39.6. 控制结构 895
39.6.1. 从函数返回 895
39.6.2. 条件语句 897
39.6.3. 简单循环 900
39.6.4. 遍历查询结果 902
39.6.5. 遍历数组 903
39.6.6. 捕获错误 904
39.7. 游标 906
39.7.1. 声明游标变量 906
39.7.2. 打开游标 907
39.7.3. 使用游标 908
39.7.4. 遍历游标结果 911
39.8. 错误和消息 911
39.9. 触发器函数 913
39.10. PL/pgSQL 内部机制 920
39.10.1. 变量替换 920
39.10.2. 计划缓存 922
39.11. PL/pgSQL开发提示 923
39.11.1. 引号的处理 924
39.12. 从Oracle PL/SQL 移植 925
39.12.1. 移植示例 926
39.12.2. 其他要关注的事项 931
39.12.3. 附录 932
40. PL/Tcl - Tcl 过程语言 935
40.1. 概述 935
40.2. PL/Tcl 函数和参数 935
40.3. PL/Tcl 中的数据值 936
40.4. PL/Tcl 中的全局数据 936
40.5. 从 PL/Tcl 访问数据库 937
40.6. PL/Tcl 中的触发器函数 939
40.7. 模块与`unknown`命令 941
40.8. Tcl 过程名 941
41. PL/Perl - Perl 过程语言 942
41.1. PL/Perl 函数和参数 942
41.2. PL/Perl 中的数据值 946
41.3. 内置函数 946
41.3.1. 从 PL/Perl 访问数据库 946
41.3.2. PL/Perl 中的辅助函数 949
41.4. PL/Perl 中的全局值 951
41.5. 受信任与不受信任的 PL/Perl 952
41.6. PL/Perl 触发器 953
41.7. PL/Perl 内部机制 954
41.7.1. 配置 954
41.7.2. 限制与缺失特性 955
42. PL/Python - Python 过程语言 956
42.1. Python 2 与 Python 3 956
42.2. PL/Python 函数 957
42.3. 数据值 958
42.3.1. 数据类型映射 958
42.3.2. 空值、None 959
42.3.3. 数组、列表 959
42.3.4. 复合类型 960
42.3.5. 集合返回函数 961

42.4. 共享数据 963
42.5. 匿名代码块 963
42.6. 触发器函数 963
42.7. 数据库访问 964
42.7.1. 数据库访问函数 964
42.7.2. 捕获错误 965
42.8. 显式子事务 966
42.8.1. 子事务上下文管理器 966
42.8.2. 较早的 Python 版本 967
42.9. 辅助函数 967
42.10. 环境变量 968
43. 服务器编程接口 969
43.1. 接口函数 969
43.2. 接口支持函数 999
43.3. 内存管理 1007
43.4. 数据更改的可见性 1016
43.5. 示例 1016

第 35 章 扩展 SQL

在接下来的各节中，我们将讨论如何通过增加以下内容来扩展 PostgreSQL SQL 查询语言：

- 函数（从第 35.3 节开始）
- 聚合（从第 35.10 节开始）
- 数据类型（从第 35.11 节开始）
- 操作符（从第 35.12 节开始）
- 索引的操作符类（从第 35.14 节开始）
- 相关对象的包（从第 35.15 节开始）

35.1. 可扩展性如何运作

PostgreSQL 之所以具有可扩展性，是因为其运作由系统目录驱动。如果你熟悉标准的关系数据库系统，就会知道它们把有关数据库、表、
列等的信息存储在通常所说的系统目录中（有些系统把这称为数据字典）。
这些目录对用户而言看起来就像普通表一样，但 DBMS 会在其中保存自己的内部管理信息。PostgreSQL 与标准关系数据库系统的一个关键差别是，PostgreSQL 在目录中存储的信息要多得多：不仅有关于表和列的信息，还有关于数据类型、函数、访问方法等的信息。
这些表可以由用户修改，而 PostgreSQL 又是基于这些表来运行的，这意味着 PostgreSQL 可以由用户扩展。相比之下，传统数据库系统只能通过修改源代码中的硬编码过程，或加载由 DBMS 供应商专门编写的模块来扩展。

此外，PostgreSQL 服务器还能通过动态加载把用户编写的代码纳入自身。也就是说，用户可以指定一个实现了新类型或新函数的目标代码文件（例如共享库），而 PostgreSQL 会在需要时加载它。把用 SQL 编写的代码加入服务器就更为简单了。这种能够“即时”修改自身行为的能力，使 PostgreSQL 特别适合用于新应用和新存储结构的快速原型设计。

35.2. PostgreSQL类型系统

PostgreSQL 数据类型可以分为基础类型、复合类型、域和伪类型。

35.2.1. 基础类型

基础类型是在 SQL 语言层之下实现的类型（通常使用 C 等低级语言），例如 `int4`。它们通常对应于所谓的抽象数据类型。PostgreSQL 只能通过用户提供的函数操作这些类型，对其行为的理解也仅限于用户所描述的内容。基础类型进一步分为标量类型和数组类型。对于每个标量类型，都会自动创建相应的数组类型，以存储由该标量类型组成的可变大小数组。

35.2.2. 复合类型

用户每次创建表时，都会创建复合类型（也称行类型）。还可以使用CREATE TYPE定义不关联任何表的“独立”复合类型。复合类型就是一个带有对应字段名的类型列表。
复合类型的值是由字段值组成的一行或一条记录。用户可以在 SQL 查询中访问各个组成字段。关于复合类型的更多信息，参见第 8.15 节。

35.2.3. 域

域基于某一种特定的基础类型，并且在很多场合下都可以与其基础类型互换。不过，
域可以带有约束，把其合法值限制在底层基础类型所允许值的一个子集内。

可以使用 SQL 命令CREATE DOMAIN创建域。本章不讨论论域的创建和使用。

35.2.4. 伪类型

有少数几种“伪类型”用于特殊目的。伪类型不能作为表的列或复合类型的属性出现，
但它们可以用来声明函数的参数类型和结果类型。
这为在类型系统内部识别特殊类别的函数提供了一种机制。表 8.24列出了现有的伪类型。

35.2.5. 多态类型

四种特别值得关注的伪类型是 `anyelement`、`anyarray`、`anynonarray` 和 `anyenum`，它们统称为多态类型。使用这些类型声明的函数称为多态函数。
多态函数可以操作多种不同的数据类型，
具体的数据类型由特定调用中实际传递给它的数据类型决定。

多态参数和结果相互关联，在解析调用多态函数的查询时，会被解析为某种具体的数据类型 a query calling a polymorphic function is parsed. Each position (either argument or return value) declared as `anyelement` is allowed to have any specific actual data type, but in any given call they must all be the same actual type. 每个声明为 `anyarray` 的位置可以具有任意数组数据类型，但同样必须全部具有相同的类型。如果同时有声明为 `anyarray` 的位置和声明为 `anyelement` 的位置，`anyarray` 位置上的实际数组类型必须是一个元素类型与 `anyelement` 位置上出现的类型相同的数组。`anynonarray` 的处理方式与 `anyelement` 完全相同，但额外要求实际类型不能是数组类型。`anyenum` 的处理方式也与 `anyelement` 完全相同，但额外要求实际类型必须是枚举类型。

因此，当多个参数位置被声明为多态类型时，其总体效果就是只允许某些实际参数类型组合。
例如，声明为 `equal(anyelement, anyelement)` 的函数可以接受任意两个输入值，只要它们属于同一数据类型。

如果函数的返回值声明为多态类型，那么至少有一个参数位置也必须是多态的，
并且实际传入的参数数据类型决定该次调用的实际结果类型。例如，
如果尚不存在数组下标机制，可以将实现下标访问的函数定义为 `subscript(anyarray, integer) returns anyelement`。这一声明将实际的第一个参数约束为数组类型，并允许解析器根据第一个参数的实际类型推导正确的结果类型。另一个例子是，声明为 `f(anyarray) returns anyenum` 的函数只接受枚举类型的数组。

注意，`anynonarray` 和 `anyenum` 并不表示独立的类型变量；它们与 `anyelement` 是同一个类型，只是附加了额外约束。例如，把函数声明为 `f(anyelement, anyenum)`，等价于把它声明为 `f(anyenum, anyenum)`：两个实际参数都必须是同一种枚举类型。

可变参数函数（接受可变数量参数的函数，见第 35.4.5 节）可以是多态的：将其最后一个参数声明为 `VARIADIC anyarray` 即可。就参数匹配和确定实际结果类型而言，这样的函数与写出相应数量的 `anynonarray` 参数时行为相同。

35.3. 用户定义的函数

PostgreSQL提供四种函数：

- 查询语言函数（用SQL编写的函数）（第 35.4 节）
- 过程语言函数（例如，用PL/pgSQL或PL/Tcl编写的函数）（第 35.7 节）
- 内部函数（第 35.8 节）
- C 语言函数（第 35.9 节）

每一类函数可以采用基础类型、复合类型或者它们的组合作为参数。此外，
每一类函数可以返回一个基础类型或一个复合类型。
函数也能被定义成返回基础类型或复合类型值的集合。

很多类函数可以接受或者返回特定的伪类型（例如，多态类型），
但是可用的功能因函数类别而异。详情可以参考每一种函数的描述。

定义SQL函数最容易，因此我们将从讨论它们开始。大部分
SQL函数的概念也能用到其他类型的函数上。

在这一章中，查看CREATE FUNCTION命令的参考页有助于更好地理解示例。这章中的一些示例可以在PostgreSQL源代码发布的`src/tutorial`目录中的`funcs.sql`和`funcs.c`中找到。

35.4. 查询语言（SQL）函数

SQL 函数执行由任意 SQL 语句组成的一个列表，并返回列表中最后一个查询的结果。在简单（非集合）情况下，将返回最后一个查询结果的第一行。（请记住，除非使用 `ORDER BY`，否则多行结果中的“第一行”并没有明确定义。）如果最后一个查询完全没有返回任何行，则返回空值。

或者，一个 SQL 函数可以通过指定函数的返回类型为 `SETOF` *`sometype`* 被声明为返回一个集合（也就是多个行），或者等效地声明它为 `RETURNS TABLE(`*`columns`*`)`。在这种情况下，最后一个查询的结果的所有行会被返回。下文将给出进一步的细节。

SQL 函数的函数体必须是一个由分号分隔的 SQL 语句列表。最后一条语句后的分号是可选的。除非该函数被声明为返回 `void`，否则最后一条语句必须是 `SELECT`，或者是 `INSERT`、`UPDATE` 或 `DELETE` 并带有 `RETURNING` 子句。

任意一组 SQL 语言命令都可以打包并定义成函数。除 `SELECT` 查询外，这些命令还可以包含数据修改查询（`INSERT`，`UPDATE`、`DELETE`），以及其他 SQL 命令。（不能使用事务控制命令，例如 `COMMIT`，`SAVEPOINT`，以及某些工具命令，例如 `VACUUM`，来编写 SQL 函数。）不过，最后一条命令必须是 `SELECT`，或者带有 `RETURNING` 子句，并返回与函数声明返回类型相符的结果。或者，如果你想定义一个执行动作但没有有用返回值的 SQL 函数，也可以把它定义为返回 `void`。例如，下面这个函数会删除 `emp` 表中薪资为负的行：

```
CREATE FUNCTION clean_emp() RETURNS void AS '
    DELETE FROM emp
        WHERE salary < 0;
' LANGUAGE SQL;

SELECT clean_emp();

 clean_emp
-----------


(1 row)
```

`CREATE FUNCTION` 命令的语法要求把函数体写成一个字符串常量。对于字符串常量，通常使用美元引用最方便（见第 4.1.2.4 节）。如果你选择使用常规的单引号字符串常量语法，那么必须在函数体中将单引号（`'`）和反斜线（`\`）写成双份（假定使用转义字符串语法，见第 4.1.2.1 节）。

SQL 函数的参数在函数体中使用语法 `$`*`n`* 引用：`$1` 引用第一个参数，`$2` 引用第二个参数，依此类推。如果参数是组合类型，则可以用点号记法（例如 `$1.name`）访问该参数的属性。这些参数只能用作数据值，而不能用作标识符。因此，举例来说，这样写是合理的：

```
INSERT INTO mytable VALUES ($1);
```

但这样写行不通：

```
INSERT INTO $1 VALUES (42);
```

35.4.1. 基类型上的SQL函数

最简单的 SQL 函数没有参数，只是返回一个基础类型，例如 `integer`：

```
CREATE FUNCTION one() RETURNS integer AS $$
    SELECT 1 AS result;
$$ LANGUAGE SQL;

-- Alternative syntax for string literal:
CREATE FUNCTION one() RETURNS integer AS '
    SELECT 1 AS result;
' LANGUAGE SQL;

SELECT one();

 one
-----
   1
```

注意我们为该函数的结果在函数体内定义了一个列别名（名为`result`），
但是这个列别名在函数以外是不可见的。因此，结果被标记为`one`而不是`result`。

定义接受基本类型作为参数的 SQL 函数几乎同样容易。在下面的示例中，注意我们如何在函数中把参数引用为 `$1` 和 `$2`。

```
CREATE FUNCTION add_em(integer, integer) RETURNS integer AS $$
    SELECT $1 + $2;
$$ LANGUAGE SQL;

SELECT add_em(1, 2) AS answer;

 answer
--------
      3
```

下面是一个更有用的函数，它可以用来从银行账户中扣款：

```
CREATE FUNCTION tf1 (integer, numeric) RETURNS integer AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1;
    SELECT 1;
$$ LANGUAGE SQL;
```

用户可以这样执行该函数，从 17 号账户中扣除 $100.00：

```
SELECT tf1(17, 100.0);
```

实际上我们可能希望从该函数得到比常数 1 更有用的结果，因此一个更可能的定义是：

```
CREATE FUNCTION tf1 (integer, numeric) RETURNS numeric AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1;
    SELECT balance FROM bank WHERE accountno = $1;
$$ LANGUAGE SQL;
```

它会调整余额并返回新余额。同样的事情也可以用 RETURNING 在一条命令中完成：

```
CREATE FUNCTION tf1 (integer, numeric) RETURNS numeric AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1
    RETURNING balance;
$$ LANGUAGE SQL;
```

35.4.2. 复合类型上的SQL函数

在编写参数为组合类型的函数时，我们必须不仅指定想要哪个参数（如上面用 $1 和 $2 那样），还要指定该参数的所需属性（字段）。例如，假设 emp 是一个包含员工数据的表，因而也是该表每一行的组合类型的名称。下面的函数 double_salary 计算某人薪资翻倍后会是多少：

```
CREATE TABLE emp (
    name        text,
    salary      numeric,
    age         integer,
    cubicle     point
);

INSERT INTO emp VALUES ('Bill', 4200, 45, '(2,1)');

CREATE FUNCTION double_salary(emp) RETURNS numeric AS $$
    SELECT $1.salary * 2 AS salary;
$$ LANGUAGE SQL;

SELECT name, double_salary(emp.*) AS dream
    FROM emp
    WHERE emp.cubicle ~= point '(2,1)';

 name | dream
------+-------
 Bill |  8400
```

注意这里用 $1.salary 语法来选取参数行值中的一个字段。还要注意，调用时的 SELECT 命令使用 * 将表的当前整行取作一个复合值。该表行也可以仅用表名来引用：

```
SELECT name, double_salary(emp) AS dream
    FROM emp
    WHERE emp.cubicle ~= point '(2,1)';
```

但这种用法已被弃用，因为它很容易让人混淆。

有时候即时构造一个复合参数值会很方便。这可以用ROW构造器完成。例如，
我们可以调整被传递给函数的数据：

```
SELECT name, double_salary(ROW(name, salary*1.1, age, cubicle)) AS
 dream
    FROM emp;
```

也可以构建一个返回复合类型的函数。这是一个返回单一emp行的函数示例：

```
CREATE FUNCTION new_emp() RETURNS emp AS $$
    SELECT text 'None' AS name,
        1000.0 AS salary,
        25 AS age,
        point '(2,2)' AS cubicle;
$$ LANGUAGE SQL;
```

在这个示例中，我们为每一个属性指定了一个常量值，但是可以用任何计算来替换这些常量。

定义该函数时有两点重要注意事项：

- 查询中的选择列表顺序必须与列在该复合类型所关联的表中出现的顺序完全相同。（系统不考虑像上面这样指定的列名。）
- 你必须对表达式进行类型转换，使之与复合类型的定义匹配，否则会得到这样的错误：

  ```
  ERROR:  function declared to return emp returns varchar instead of
   text at column 1
  ```

定义同一个函数的另一种方法是：

```
CREATE FUNCTION new_emp() RETURNS emp AS $$
    SELECT ROW('None', 1000.0, 25, '(2,2)')::emp;
$$ LANGUAGE SQL;
```

这里我们编写了一个 SELECT，它只返回具有正确复合类型的单列。在这种情况下，这样做并没有更好，但在某些情况下它是一种方便的替代方法 – 例如，如果我们需要通过调用另一个返回所需复合值的函数来计算结果。

我们可以用两种方式之一直接调用这个函数：

```
SELECT new_emp();

         new_emp
--------------------------
 (None,1000.0,25,"(2,2)")

SELECT * FROM new_emp();

 name | salary | age | cubicle
------+--------+-----+---------
 None | 1000.0 |  25 | (2,2)
```

第二种方式在第 35.4.7 节中有更完整的描述。

当你使用返回复合类型的函数时，可能只需要取其结果中的一个字段（属性）。
可以使用下面这样的语法：

```
SELECT (new_emp()).name;
```

```
 name
------
 None
```

这里需要额外的括号，以免解析器产生歧义。如果不加括号，结果会是这样：

```
SELECT new_emp().name;
ERROR:  syntax error at or near "."
LINE 1: SELECT new_emp().name;
                        ^
```

另一种选择是使用函数记法来提取属性。简单的解释是：记法 `attribute(table)` 和 `table.attribute` 可以互换使用。

```
SELECT name(new_emp());

 name
------
 None

-- This is the same as:
-- SELECT emp.name AS youngster FROM emp WHERE emp.age < 30;

SELECT name(emp) AS youngster FROM emp WHERE age(emp) < 30;

 youngster
-----------
 Sam
 Andy
```

> **提示**
>
> 函数记法与属性记法之间的等价性，使得可以用组合类型上的函数来模拟“计算字段”。For example, using the previous definition for `double_salary(emp)`, we can write
>
> ```
> SELECT emp.name, emp.double_salary FROM emp;
> ```
>
> 使用这种写法的应用不需要直接了解 `double_salary` 并不是表的真实列。（也可以用视图来模拟计算字段。）
>
> 由于这一行为，
> 给一个接受单一组合类型参数的函数取与该组合类型任一字段相同的名称是不明智的。

另一种使用函数返回复合类型的方法是将结果传递给另一个接受正确行类型作为输入的函数：

```
CREATE FUNCTION getname(emp) RETURNS text AS $$
    SELECT $1.name;
$$ LANGUAGE SQL;

SELECT getname(new_emp());
 getname
---------
```

```
 None
(1 row)
```

使用返回复合类型的函数的又一种方式，是把它作为表函数调用，如第 35.4.7 节所述。

35.4.3. 带参数名的 SQL 函数

可以为函数的参数附上名字，例如

```
CREATE FUNCTION tf1 (acct_no integer, debit numeric) RETURNS numeric
 AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1
    RETURNING balance;
$$ LANGUAGE SQL;
```

这里第一个参数被命名为 `acct_no`，第二个参数被命名为 `debit`。就 SQL 函数本身而言，这些名字只是装饰；在函数体中你仍必须把参数引用为 `$1`、`$2` 等。（有些过程语言允许改用参数名。）不过，为参数附上名字对文档目的很有用。当函数有很多参数时，
在调用函数时使用这些名字也很有用，见第 4.3 节中的描述。

35.4.4. 带输出参数的 SQL 函数

描述函数结果的另一种方式是用输出参数来定义它，如下例所示：

```
CREATE FUNCTION add_em (IN x int, IN y int, OUT sum int)
AS 'SELECT $1 + $2'
LANGUAGE SQL;

SELECT add_em(3,7);
 add_em
--------
     10
(1 row)
```

这与第 35.4.1 节中所示的 `add_em` 版本没有本质区别。输出参数的真正价值在于，它们为定义返回若干列的函数提供了一种便捷方式。例如，

```
CREATE FUNCTION sum_n_product (x int, y int, OUT sum int, OUT product
 int)
AS 'SELECT $1 + $2, $1 * $2'
LANGUAGE SQL;

 SELECT * FROM sum_n_product(11,42);
 sum | product
-----+---------
  53 |     462
(1 row)
```

这里本质上发生的事情是：我们为该函数的结果创建了一个匿名复合类型。
上面的示例与下面这种写法的最终效果相同：

```
CREATE TYPE sum_prod AS (sum int, product int);
```

```
CREATE FUNCTION sum_n_product (int, int) RETURNS sum_prod
AS 'SELECT $1 + $2, $1 * $2'
LANGUAGE SQL;
```

但通常不必额外定义一个独立的复合类型会更方便。注意，
附在输出参数上的名称并非只是装饰，它们决定了这个匿名复合类型的列名。
（如果省略输出参数名称，系统会自行选择一个名称。）

在从 SQL 调用这样一个函数时，输出参数不会被包括在调用参数列表中。这是因为PostgreSQL
只考虑输入参数来定义函数的调用签名。
这也意味着在为诸如删除函数等目的引用该函数时只有输入参数有关系。
我们可以用下面的命令之一删除上述函数

```
DROP FUNCTION sum_n_product (x int, y int, OUT sum int, OUT product
 int);
DROP FUNCTION sum_n_product (int, int);
```

参数可以被标记为`IN`（默认）、`OUT`、`INOUT`或者`VARIADIC`。一个`INOUT`
参数既作为一个输入参数（调用参数列表的一部分）
又作为一个输出参数（结果记录类型的一部分）。`VARIADIC`参数是输入参数，
但被按照下文所述特殊对待。

35.4.5. 带有可变数量参数的SQL函数

SQL 函数可以被声明为接受可变数量的参数，只要所有“可选”参数都属于同一种数据类型。可选参数会以数组形式传递给函数。定义这类函数时，需要把最后一个参数标记为 `VARIADIC`；该参数必须被声明为数组类型。例如：

```
CREATE FUNCTION mleast(VARIADIC arr numeric[]) RETURNS numeric AS $$
    SELECT min($1[i]) FROM generate_subscripts($1, 1) g(i);
$$ LANGUAGE SQL;

SELECT mleast(10, -1, 5, 4.4);
 mleast
--------
     -1
(1 row)
```

实际上，位于 `VARIADIC` 位置及之后的所有实参都会被收集成一个一维数组，就像你写成了：

```
SELECT mleast(ARRAY[10, -1, 5, 4.4]);    -- doesn't work
```

不过你实际上不能这样写，至少它不会匹配这个函数定义。被标记为 `VARIADIC` 的参数匹配的是其元素类型出现一次或多次，而不是它自身的数组类型。

有时候，能够把一个已经构造好的数组传给可变参数函数会很有用，
尤其是当一个可变参数函数想把它的数组参数再传给另一个函数时。你可以在调用中指定
`VARIADIC` 来做到这一点：

```
SELECT mleast(VARIADIC ARRAY[10, -1, 5, 4.4]);
```

这样会阻止函数的可变参数按其元素类型展开，从而让数组实参能够按常规方式匹配。
`VARIADIC` 只能附加在函数调用的最后一个实参上。

在调用中指定VARIADIC也是向可变参数函数传递空数组的唯一方式，例如：

```
SELECT mleast(VARIADIC ARRAY[]::numeric[]);
```

仅仅写成SELECT mleast()是行不通的，因为可变参数必须匹配至少一个实参。（如果你希望允许这种调用，可以再定义一个同名且不带参数的函数mleast。）

从可变参数派生出的数组元素参数会被视为没有自己的名字。这意味着除非你指定了VARIADIC，否则不能使用命名参数来调用可变参数函数（第 4.3 节）。例如，下面的调用是可行的：

```
SELECT mleast(VARIADIC arr := ARRAY[10, -1, 5, 4.4]);
```

但这些就不行：

```
SELECT mleast(arr := 10);
SELECT mleast(arr := ARRAY[10, -1, 5, 4.4]);
```

35.4.6. 带有参数默认值的SQL函数

可以为部分或所有输入参数声明带默认值的函数。每当函数被调用时实参不够多，
就会插入默认值。由于实参只能从实参列表的末尾开始被省略，
一个有默认值的参数之后的所有参数也必须都有默认值。
（虽然使用命名实参记法可以让这一限制放松，但它仍被强制执行，
以便位置实参记法能合理工作。）

例如：

```
CREATE FUNCTION foo(a int, b int DEFAULT 2, c int DEFAULT 3)
RETURNS int
LANGUAGE SQL
AS $$
    SELECT $1 + $2 + $3;
$$;

SELECT foo(10, 20, 30);
 foo
-----
  60
(1 row)

SELECT foo(10, 20);
 foo
-----
  33
(1 row)

SELECT foo(10);
 foo
-----
  15
(1 row)

SELECT foo();  -- fails since there is no default for the first
 argument
```

```
ERROR:  function foo() does not exist
```

也可以用 = 符号代替关键字 DEFAULT。

35.4.7. SQL 函数作为表来源

所有的 SQL 函数都可以被用在查询的FROM子句中，但是对于返回复合类型的函数特别有用。如果函数被定义为返回一种基础类型，该表函数会产生一个单列表。
如果该函数被定义为返回一种复合类型，该表函数会为该复合类型的每一个属性产生一列。

以下是一个示例：

```
CREATE TABLE foo (fooid int, foosubid int, fooname text);
INSERT INTO foo VALUES (1, 1, 'Joe');
INSERT INTO foo VALUES (1, 2, 'Ed');
INSERT INTO foo VALUES (2, 1, 'Mary');

CREATE FUNCTION getfoo(int) RETURNS foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT *, upper(fooname) FROM getfoo(1) AS t1;

 fooid | foosubid | fooname | upper
-------+----------+---------+-------
     1 |        1 | Joe     | JOE
(1 row)
```

如示例所示，我们可以像操作普通表的列一样操作函数结果中的列。

注意我们只从函数得到了一行。这是因为我们没有使用SETOF。这会在下一节中介绍。

35.4.8. 返回集合的SQL函数

当一个 SQL 函数被声明为返回SETOF *sometype*时，该函数的最后一个查询会被执行完，并且它输出的每一行都会被作为结果集的一个元素返回。

这一特性通常是在查询的 FROM 子句中调用函数时使用的。在这种情况下，函数返回的每一行都会成为查询所见表中的一行。例如，假设表 foo 与上文相同，我们写：

```
CREATE FUNCTION getfoo(int) RETURNS SETOF foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT * FROM getfoo(1) AS t1;
```

那么会得到：

```
 fooid | foosubid | fooname
-------+----------+---------
     1 |        1 | Joe
     1 |        2 | Ed
(2 rows)
```

也可以通过输出参数定义的列来返回多行，例如：

```
CREATE TABLE tab (y int, z int);
INSERT INTO tab VALUES (1, 2), (3, 4), (5, 6), (7, 8);

CREATE FUNCTION sum_n_product_with_tab (x int, OUT sum int, OUT
 product int)
RETURNS SETOF record
AS $$
    SELECT $1 + tab.y, $1 * tab.y FROM tab;
$$ LANGUAGE SQL;

SELECT * FROM sum_n_product_with_tab(10);
 sum | product
-----+---------
  11 |      10
  13 |      30
  15 |      50
  17 |      70
(4 rows)
```

这里的关键点是：你必须写成 `RETURNS SETOF record`，以表明该函数返回的是多行而不是单行。如果只有一个输出参数，则写该参数的类型，而不是 `record`。

当前，返回集合的函数也可以在查询的 select 列表中被调用。对于查询本身生成的每一行，该返回集合的函数都会被调用，并为其结果集中的每一个元素生成一行输出。但请注意，
这一能力已被弃用，未来的发行版中可能会将其移除。下面是一个从 select 列表返回集合的函数示例：

```
CREATE FUNCTION listchildren(text) RETURNS SETOF text AS $$
    SELECT name FROM nodes WHERE parent = $1
$$ LANGUAGE SQL;

SELECT * FROM nodes;
   name    | parent
-----------+--------
 Top       |
 Child1    | Top
 Child2    | Top
 Child3    | Top
 SubChild1 | Child1
 SubChild2 | Child1
(6 rows)

SELECT listchildren('Top');
 listchildren
--------------
 Child1
 Child2
 Child3
(3 rows)

SELECT name, listchildren(name) FROM nodes;
  name  | listchildren
--------+--------------
```

```
 Top    | Child1
 Top    | Child2
 Top    | Child3
 Child1 | SubChild1
 Child1 | SubChild2
(5 rows)
```

在最后一个 SELECT 中，注意 Child2、Child3 等没有对应的输出行。这是因为 listchildren 对这些参数返回空集，因此不生成任何结果行。

> **注意**
>
> 如果一个函数的最后一条命令是 INSERT、UPDATE 或 DELETE 并带有 RETURNING，那么即使该函数没有声明为 SETOF，或者调用它的查询并未取走全部结果行，该命令也总会执行到完成。RETURNING 子句产生的额外结果行会被静默丢弃，但相应的表修改仍然会发生，并且会在函数返回前全部完成。

35.4.9. 返回TABLE的SQL函数

还有另一种方法可以把函数声明为返回一个集合，即使用 RETURNS TABLE(*columns*)语法。这等效于使用一个或者多个OUT参数外加把函数标记为返回 SETOF record（或者是SETOF单个输出参数的类型）。这种写法是在最近的 SQL 标准中指定的，因此可能比使用 SETOF的移植性更好。

例如，前面的求和并且相乘的示例也可以这样来做：

```
CREATE FUNCTION sum_n_product_with_tab (x int)
RETURNS TABLE(sum int, product int) AS $$
    SELECT $1 + tab.y, $1 * tab.y FROM tab;
$$ LANGUAGE SQL;
```

不允许把显式的OUT或者INOUT参数用于 RETURNS TABLE记法 – 必须把所有输出列放在 TABLE 列表中。

35.4.10. 多态SQL函数

SQL 函数可以声明为接受和返回多态类型 anyelement、anyarray、anynonarray 和 anyenum。关于多态函数的更详细解释，见第 35.2.5 节。下面是一个多态函数 make_array，它由两个任意数据类型的元素构建一个数组：

```
CREATE FUNCTION make_array(anyelement, anyelement) RETURNS anyarray AS
 $$
    SELECT ARRAY[$1, $2];
$$ LANGUAGE SQL;

SELECT make_array(1, 2) AS intarray, make_array('a'::text, 'b') AS
 textarray;
 intarray | textarray
----------+-----------
 {1,2}    | {a,b}
(1 row)
```

注意类型转换 `'a'::text` 的使用是为了指定该参数的类型是 `text`。如果该参数只是一个字符串字面量，就必须这样做，否则它会被当作 `unknown` 类型，并且 `unknown` 的数组也不是一种合法的类型。如果没有该类型转换，将得到这样的错误：

```
ERROR:  could not determine polymorphic type because input has type
 "unknown"
```

允许存在具有固定返回类型的多态参数，但反过来则不允许。例如：

```
CREATE FUNCTION is_greater(anyelement, anyelement) RETURNS boolean AS
 $$
    SELECT $1 > $2;
$$ LANGUAGE SQL;

SELECT is_greater(1, 2);
 is_greater
------------
 f
(1 row)

CREATE FUNCTION invalid_func() RETURNS anyelement AS $$
    SELECT 1;
$$ LANGUAGE SQL;
ERROR:  cannot determine result data type
DETAIL:  A function returning a polymorphic type must have at least
 one polymorphic argument.
```

多态也可以和带输出参数的函数一起使用。例如：

```
CREATE FUNCTION dup (f1 anyelement, OUT f2 anyelement, OUT f3
 anyarray)
AS 'select $1, array[$1,$1]' LANGUAGE SQL;

SELECT * FROM dup(22);
 f2 |   f3
----+---------
 22 | {22,22}
(1 row)
```

多态也可以用于可变参数函数。例如：

```
CREATE FUNCTION anyleast (VARIADIC anyarray) RETURNS anyelement AS $$
    SELECT min($1[i]) FROM generate_subscripts($1, 1) g(i);
$$ LANGUAGE SQL;

SELECT anyleast(10, -1, 5, 4);
 anyleast
----------
       -1
(1 row)

SELECT anyleast('abc'::text, 'def');
 anyleast
----------
 abc
```

```
(1 row)

CREATE FUNCTION concat_values(text, VARIADIC anyarray) RETURNS text AS
 $$
    SELECT array_to_string($2, $1);
$$ LANGUAGE SQL;

SELECT concat_values('|', 1, 4, 2);
 concat_values
---------------
 1|4|2
(1 row)
```

35.4.11. 带有排序规则的SQL函数

当一个 SQL 函数有一个或多个支持排序规则的数据类型参数时，会根据实参所带的排序规则，为每次函数调用确定一个排序规则，详见第 22.2 节。如果能够成功确定（也就是说，参数之间不存在隐式排序规则冲突），那么所有支持排序规则的参数都会被视为隐式带有该排序规则。
这会影响函数中对排序规则敏感的操作的行为。例如，使用上文的 `anyleast` 函数时，

```
SELECT anyleast('abc'::text, 'ABC');
```

的结果将取决于数据库的默认排序规则。在 `C` 区域设置下，结果会是 `ABC`，但在许多其他区域设置下则会是 `abc`。可以在任意参数上添加 `COLLATE` 子句来强制指定要使用的排序规则，例如：

```
SELECT anyleast('abc'::text, 'ABC' COLLATE "C");
```

另外，如果你希望某个函数无论以什么排序规则调用，都始终按特定排序规则工作，
可以在函数定义中按需插入 `COLLATE` 子句。下面这个版本的 `anyleast` 将始终使用 `en_US` 区域设置来比较字符串：

```
CREATE FUNCTION anyleast (VARIADIC anyarray) RETURNS anyelement AS $$
    SELECT min($1[i] COLLATE "en_US") FROM generate_subscripts($1, 1)
 g(i);
$$ LANGUAGE SQL;
```

但请注意，如果把它用于不支持排序规则的数据类型，就会抛出错误。

如果无法在实参之间确定共同的排序规则，那么 SQL 函数会把参数视为带有其数据类型的默认排序规则（通常是数据库的默认排序规则，但对域类型参数来说也可能不同）。

这种支持排序规则的参数的行为，可以看作是仅适用于文本数据类型的一种受限多态形式。

35.5. 函数重载

可以用同一个 SQL 名称定义多个函数，只要它们接受的参数不同即可。换句话说，函数名可以被重载。当一个查询被执行时，
服务器将根据所提供参数的数据类型和数量来决定要调用哪个函数。
重载也可用来模拟具有可变参数个数（最大个数有限）的函数。

在创建一个重载函数家族时，应该小心不要创建歧义。例如，给定函数：

```
CREATE FUNCTION test(int, real) RETURNS ...
CREATE FUNCTION test(smallint, double precision) RETURNS ...
```

对于`test(1, 1.5)`这样的输入就无法立刻清楚地知道应该调用哪个函数。当前实现的解决规则在第 10 章中有描述，但是设计一个依赖于这种行为的系统是不明智的。

具有单个复合类型参数的函数，通常不应与该类型的任何属性（字段）同名。回想一下，`attribute(table)`被视为等价于 `table.attribute`。如果在复合类型上的函数与复合类型的属性之间出现歧义，系统总会优先选择属性。可以通过给函数名加上模式限定（即 `schema.func(table)`）来覆盖这一选择，但更好的做法仍然是避免使用会冲突的名字。

另一类可能的冲突发生在可变参数函数和非可变参数函数之间。例如，可以同时创建 `foo(numeric)`和`foo(VARIADIC numeric[])`。这时，对于只提供一个 numeric 参数的调用（例如`foo(10.1)`），就不清楚该匹配哪个函数。规则是：优先使用在搜索路径中出现较早的那个函数；如果两者位于同一模式中，则优先使用非可变参数函数。

在重载 C 语言函数时有一个额外的约束：重载函数家族中的每一个函数的 C 名称必须与其他所有函数的 C 名称不同，不管是内部的还是动态载入的。如果这条规则被违背，该行为将不可移植。你可能会得到一个运行时链接器错误，或者这些函数之一将被调用（通常是内部的那一个）。SQL `CREATE FUNCTION`命令的`AS`子句的另一种形式可以把 SQL 函数名和 C 源代码中的函数名分离。例如：

```
CREATE FUNCTION test(int) RETURNS int
    AS 'filename', 'test_1arg'
    LANGUAGE C;
CREATE FUNCTION test(int, int) RETURNS int
    AS 'filename', 'test_2arg'
    LANGUAGE C;
```

这里的 C 函数名称反映了很多种可能的习惯之一。

35.6. 函数易变性分类

每一个函数都有一个易变性分类，可能的类别是 `VOLATILE`，`STABLE` 或 `IMMUTABLE`。`VOLATILE`会作为默认类别，条件是CREATE FUNCTION命令没有指定类别。易变性分类是给优化器的关于该函数行为的一种承诺：

- 一个`VOLATILE`函数可以做任何事情，包括修改数据库。在使用相同的参数连续调用时，它能返回不同的结果。优化器不会对这类函数的行为做任何假定。使用易变函数的查询会在需要该函数值的每一行上重新对它求值。

- 一个`STABLE`函数不能修改数据库并且被确保对一个语句中的所有行用给定的相同参数返回相同的结果。这种分类允许优化器把该函数的多个调用优化成一个调用。特别是，在一个索引扫描条件中使用包含这样一个函数的表达式是安全的（因为一次索引扫描只会计算一次比较值，而不是为每一行都计算一次，在一个索引扫描条件中不能使用 `VOLATILE`函数）。

- 一个`IMMUTABLE`函数不能修改数据库并且被确保用相同的参数永远返回相同的结果。这种分类允许优化器在一个查询用常量参数调用该函数时提前计算该函数。例如，一个`SELECT ... WHERE x = 2 + 2`这样的查询可以被简化为`SELECT ... WHERE x = 4`，因为整数加法操作符底层的函数被标记为`IMMUTABLE`。

为了最好的优化结果，你应该把函数标记为对它们合法的易变性分类中最严格的那种。

任何带有副作用的函数必须被标记为`VOLATILE`，这样对它的调用就不能被优化掉。甚至如果一个函数的值在一个查询中会变化，即使它没有副作用也需要被标记为`VOLATILE`。这样的示例有`random()`、`currval()`、`timeofday()`等。

另一种重要的示例是current_timestamp家族的函数有资格被标记为STABLE，因为它们的值在一个事务中不会改变。

在考虑先规划然后立即执行的简单交互式查询时，在STABLE和 IMMUTABLE分类间的区别相对较小：一个函数是在规划时只执行一次还是在查询执行开始期间只执行一次没有太大关系。但是如果计划被保存下来然后在后面被重用，区别就大了。如果把一个实际上并非不可变的函数标记为IMMUTABLE，就可能在规划期间过早将它折叠成常量，导致在后续重用该计划时继续使用陈旧的值。在使用预备语句，或使用会缓存执行计划的函数语言（如 PL/pgSQL）时，这会带来严重问题。

对于用 SQL 或者其他任何标准过程语言编写的函数，还有第二种由易变性分类决定的特性，即由调用该函数的 SQL 命令所作的数据修改的可见性。VOLATILE函数将看到这些更改，STABLE 或者IMMUTABLE函数则看不到。这种行为使用 MVCC 的快照行为（见第 13 章）实现：STABLE和IMMUTABLE函数使用一个在调用查询开始时建立的快照，而 VOLATILE函数在它们执行的每一个查询的开始都获得一个新鲜的快照。

> **注意**
>
> 用 C 编写的函数按照它们自己需要的方式管理快照，但是通常最好让 C 函数也按照上面的方式来。

由于这种快照行为，只包含 SELECT 命令的函数可以安全地标记为 STABLE，即使它查询的表可能正被并发查询修改。PostgreSQL 会使用为调用查询建立的快照来执行 STABLE 函数中的所有命令，因而该函数在整个查询期间看到的都是数据库的固定视图。

对 IMMUTABLE 函数中的 SELECT 也采用同样的快照行为。一般来说，在 IMMUTABLE 函数里查询数据库表并不明智，因为一旦表内容发生变化，就会破坏其不变性。不过，PostgreSQL 并不会强制禁止这样做。

一种常见的错误是当一个函数的结果依赖于一个配置参数时把它标记为 IMMUTABLE。例如，一个操纵时间戳的函数有可能结果依赖于timezone设置。为了安全起见，这类函数应该被标记为STABLE。

> **注意**
>
> 在 PostgreSQL 8.0 版之前，STABLE 和 IMMUTABLE 函数不能修改数据库这项要求并未被系统强制执行。8.0 及之后的版本会强制执行它，方式是要求这些类别的 SQL 函数和过程语言函数不包含 SELECT 之外的 SQL 命令。（这并不是一个完全万无一失的测试，因为这类函数仍可能调用修改数据库的 VOLATILE 函数。如果你这样做，会发现 STABLE 或 IMMUTABLE 函数不会注意到被调用函数所做的数据库更改，因为这些更改对它的快照是隐藏的。）

35.7. 过程语言函数

PostgreSQL允许用除 SQL 和 C 之外的语言编写用户定义的函数。这些语言通常被称为过程语言（PL）。过程语言并不内置在PostgreSQL服务器中，它们通过可装载模块提供。更多信息请见第 38 章以及接下来的章节。

35.8. 内部函数

内部函数由 C 编写并且已经被静态链接到PostgreSQL 服务器中。该函数定义的“主体”指定该函数的 C 语言名称，它不必与为 SQL 使用而声明的名称相同。（出于向后兼容的原因，也接受空主体，此时会认为 C 语言函数名与 SQL 函数名相同。）

通常，所有存在于服务器中的内部函数都在数据库集簇的初始化（见第 17.2 节）期间被声明，但是用户可以使用 `CREATE FUNCTION`为一个内部函数创建额外的别名。在`CREATE FUNCTION`中用语言名`internal`来声明内部函数。例如，要为 `sqrt`函数创建一个别名：

```
CREATE FUNCTION square_root(double precision) RETURNS double precision
    AS 'dsqrt'
    LANGUAGE internal
    STRICT;
```

（大部分内部函数应该被声明为“严格”）。

> **注意**
>
> 并非所有“预定义”函数在上述意义上都是“内部”函数。有些预定义的函数由 SQL 编写。

35.9. C 语言函数

用户定义的函数可以用 C 编写（或者可以与 C 兼容的语言，例如 C++）。这类函数被编译成动态载入对象（也被称为共享库）并且由服务器在需要时载入。动态载入是把“C语言”函数和“内部”函数区分开的特性 – 两者真正的编码习惯实际上是一样的（因此，标准内部函数库为用户定义的 C 函数提供了丰富的编码示例）。

目前 C 函数使用两种不同的调用约定。较新的“版本 1”调用约定通过为函数编写一个`PG_FUNCTION_INFO_V1()`宏来指示，如下文所示。没有这样的宏则表示这是一个旧式（“版本 0”）函数。无论哪种情况，`CREATE FUNCTION`中指定的语言名称都是`C`。旧式函数由于可移植性问题和功能缺乏，现在已被弃用，但出于兼容性原因仍然支持。

35.9.1. 动态载入

在一个会话中第一次调用一个特定可载入目标文件中的用户定义函数时，动态载入器会把那个目标文件载入到内存以便该函数被调用。因此用户定义的 C 函数的`CREATE FUNCTION`必须为该函数指定两块信息：可载入目标文件的名称，以及要在该目标文件中调用的特定函数的 C 名称（链接符号）。如果没有显式指定 C 名称，则它被假定为和 SQL 函数名相同。

下面的算法被用来基于`CREATE FUNCTION` 命令中给定的名称来定位共享目标文件：

1. 如果名称是一个绝对路径，则载入给定的文件。
2. 如果该名称以字符串`$libdir`开始，那么这一部分会被 PostgreSQL包的库目录名（在编译时确定）替换。
3. 如果该名称不包含目录部分，会在配置变量dynamic_library_path指定的路径中搜索该文件。
4. 否则（在该路径中没找到该文件，或者它包含一个非绝对目录），动态载入器将尝试接受给定的名称，这很可能会失败（依赖当前工作目录是不可靠的）。

如果这个序列不起作用，会把平台相关的共享库文件名扩展（通常是 `.so`）追加到给定的名称并且再次尝试上述的过程。如果还是失败，则载入失败。

建议通过相对于 `$libdir` 的路径，或者通过动态库路径来定位共享库。这样一来，如果新安装位于不同的位置，版本升级会更简单。`$libdir` 实际代表的目录可以通过命令 `pg_config --pkglibdir` 查出。

用于运行PostgreSQL服务器的用户 ID 必须能够通过要载入文件的路径。常见的错误是把文件或更高层的目录变得对postgres用户不可读或者不可执行。

在任何情况下，`CREATE FUNCTION`命令中给定的文件名会被原封不动地记录在系统目录中，这样如果需要再次载入该文件则会应用同样的过程。

> **注意**
>
> PostgreSQL不会自动编译 C 函数。在从`CREATE FUNCTION`命令中引用目标文件之前，它必须先被编译好。更多信息请见第 35.9.6 节。

为确保动态装载的目标文件不会被加载到不兼容的服务器中，
PostgreSQL会检查该文件是否包含内容适当的“魔数块”。
这样服务器就能检测出明显的不兼容情况，例如代码所针对的大版本不同于当前的
PostgreSQL。魔数块自 PostgreSQL 8.2 起就是必需的。要加入魔数块，请在模块某一个（且只能一个）源文件中写入下列内容，并事先包含头文件 `fmgr.h`：

```
#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif
```

可以省略 `#ifdef` 检查，如果代码不需要针对 8.2 之前的 PostgreSQL 版本编译的话。

在第一次使用之后，动态载入目标文件会保留在内存中。在同一个会话中，
后续对该文件中函数的调用只需付出一次很小的符号表查找开销。
如果需要强制重新载入一个目标文件（例如在重新编译之后），就需要开启一个新的会话。

动态装载的文件可以选择包含初始化和终结函数。如果文件包含一个名为 `_PG_init` 的函数，该函数会在文件装载后立即被调用。该函数不接收参数，并且应返回 void。如果文件包含一个名为 `_PG_fini` 的函数，该函数会在文件卸载前立即被调用。同样，该函数不接收参数，并且应返回 void。注意，`_PG_fini` 只会在卸载文件时被调用，而不会在进程终止时被调用。（目前，卸载已被禁用，永远不会发生，但将来可能会改变。）

35.9.2. C 语言函数中的基础类型

要了解如何编写 C 语言函数，你需要了解 PostgreSQL如何在内部表达基础类型以及如何与函数传递它们。在内部，PostgreSQL把基础类型视为一块“内存数据块”。
你为该类型定义的用户自定义函数，决定了 PostgreSQL 如何操作它。也就是说，PostgreSQL 只负责把数据存到磁盘、再从磁盘取回，而数据的输入、处理和输出则依赖你定义的这些函数。

基础类型可以有三种内部格式之一：

- 传值，定长
- 传引用，定长

- 传引用，变长

传值类型在长度上只能是 1、2 或 4 字节（如果你的机器上 sizeof(Datum) 是 8，则还有 8 字节）。你应当小心地定义你的类型，使其在所有架构上都具有相同的尺寸（字节）。例如，long 类型很危险，因为它在某些机器上是 4 字节但在另一些机器上是 8 字节；而 int 类型在大多数 Unix 机器上都是 4 字节。在 Unix 机器上 int4 类型的一种合理实现可以是：

```
/* 4-byte integer, passed by value */
typedef int int4;
```

另一方面，任意尺寸的定长类型都可以通过引用传递。例如，这里有一种 PostgreSQL类型的实现示例：

```
/* 16 字节结构，传引用 */
typedef struct
{
    double  x, y;
} Point;
```

在PostgreSQL函数中传进或传出这种类型时，只能使用指向这种类型的指针。
要返回这样一种类型的值，用 palloc分配正确的内存量，然后填充分配好的内存，并且返回一个指向该内存的指针（还有，如果只想返回与具有相同数据类型的一个输入参数相同的值，
可以跳过额外的palloc并且返回指向该输入值的指针）。

最后，所有变长类型也必须按引用传递。所有变长类型都必须以一个正好 4 字节的长度字段开始，而要存储在该类型中的所有数据都必须位于紧随该长度字段之后的内存中。
长度字段包含该结构的总长度，也就是说，它包含长度字段本身的尺寸。

另一个重点是要避免在数据类型值中留下未被初始化的位。例如，
要注意把可能存在于结构体中的任何对齐填充字节置零。如果不这样做，
你的数据类型的逻辑等价常量可能会被规划器认为是不等的，
进而导致低效的（不过还是正确的）计划。

> **警告**
>
> 绝不要修改通过引用传递的输入值的内容。如果这样做很可能会破坏磁盘上的数据，
> 因为给出的指针可能直接指向一个磁盘缓冲区。这条规则唯一的例外在第 35.10 节中有解释。

作为一个例子，我们可以把类型 text 定义为如下形式：

```
typedef struct {
    int4 length;
    char data[1];
} text;
```

显然，这里声明的 data 字段不足以容纳所有可能的字符串。由于在 C 中无法声明变长结构体，我们依赖于这样的知识：C 编译器不会对数组下标做范围检查。我们只需分配必要数量的空间，然后就像数组声明为正确长度那样去访问它。（这是一个常见的技巧，你可以在许多关于 C 的教科书中读到。）

在操作变长类型时，我们必须小心分配正确数量的内存，并正确设置长度字段。例如，
如果我们想在一个 text 结构体中存储 40 字节，我们可以使用这样的代码片段：

```
#include "postgres.h"
...
char buffer[40]; /* our source data */
...
text *destination = (text *) palloc(VARHDRSZ + 40);
SET_VARSIZE(destination, VARHDRSZ + 40);
memcpy(destination->data, buffer, 40);
...
```

VARHDRSZ 和 sizeof(int4) 一样，但是用宏 VARHDRSZ 来引用变长类型的额外开销的尺寸被认为是比较好的风格。还有，长度字段必须使用 SET_VARSIZE 宏来设置，而不是用简单的赋值来设置。

表 35.1说明了在编写使用 PostgreSQL 内置类型的 C 语言函数时，哪种 C 类型对应哪种 SQL 类型。“定义文件”列给出了需要包含的头文件，以获取类型定义。（实际的定义可能位于所列文件包含的其他文件中。建议用户坚持使用已定义的接口。）请注意，
在任何源文件中都应该始终首先包含postgres.h，因为它声明了很多你反正都会用到的内容。

表 35.1. 内置 SQL 类型等效的 C 类型

SQL Type	C Type	Defined In
abstime	AbsoluteTime	utils/nabstime.h
boolean	bool	postgres.h (maybe compiler built-in)
box	BOX*	utils/geo_decls.h
bytea	bytea*	postgres.h
"char"	char	(compiler built-in)
character	BpChar*	postgres.h
cid	CommandId	postgres.h
date	DateADT	utils/date.h
smallint (int2)	int2 or int16	postgres.h
int2vector	int2vector*	postgres.h
integer (int4)	int4 or int32	postgres.h
real (float4)	float4*	postgres.h
double precision (float8)	float8*	postgres.h
interval	Interval*	utils/timestamp.h
lseg	LSEG*	utils/geo_decls.h
name	Name	postgres.h
oid	Oid	postgres.h
oidvector	oidvector*	postgres.h
path	PATH*	utils/geo_decls.h
point	POINT*	utils/geo_decls.h
regproc	regproc	postgres.h
reltime	RelativeTime	utils/nabstime.h
text	text*	postgres.h
tid	ItemPointer	storage/itemptr.h
time	TimeADT	utils/date.h

SQL Type	C Type	Defined In
time with time zone	TimeTzADT	utils/date.h
timestamp	Timestamp*	utils/timestamp.h
tinterval	TimeInterval	utils/nabstime.h
varchar	VarChar*	postgres.h
xid	TransactionId	postgres.h

了解了基础类型所有可能的结构后，就可以看一些实际函数示例。

35.9.3. 版本 0 调用约定

我们先介绍“旧式”调用约定——尽管这种方法现在已被弃用，但入门时更容易上手。在版本 0 方法中，C 函数的参数和结果就按普通的 C 风格声明，只是要注意使用上文所示的每种 SQL 数据类型的 C 表示。

下面是一些示例：

```
#include "postgres.h"
#include <string.h>
#include "utils/geo_decls.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

/* by value */

int
add_one(int arg)
{
    return arg + 1;
}

/* by reference, fixed length */

float8 *
add_one_float8(float8 *arg)
{
    float8    *result = (float8 *) palloc(sizeof(float8));

    *result = *arg + 1.0;

    return result;
}

Point *
makepoint(Point *pointx, Point *pointy)
{
    Point     *new_point = (Point *) palloc(sizeof(Point));

    new_point->x = pointx->x;
    new_point->y = pointy->y;
```

```
    return new_point;
}

/* by reference, variable length */

text *
copytext(text *t)
{
    /*
     * VARSIZE is the total size of the struct in bytes.
     */
    text *new_t = (text *) palloc(VARSIZE(t));
    SET_VARSIZE(new_t, VARSIZE(t));
    /*
     * VARDATA is a pointer to the data region of the struct.
     */
    memcpy((void *) VARDATA(new_t), /* destination */
           (void *) VARDATA(t),     /* source */
           VARSIZE(t) - VARHDRSZ);  /* how many bytes */
    return new_t;
}

text *
concat_text(text *arg1, text *arg2)
{
    int32 new_text_size = VARSIZE(arg1) + VARSIZE(arg2) - VARHDRSZ;
    text *new_text = (text *) palloc(new_text_size);

    SET_VARSIZE(new_text, new_text_size);
    memcpy(VARDATA(new_text), VARDATA(arg1), VARSIZE(arg1) - VARHDRSZ)
;
    memcpy(VARDATA(new_text) + (VARSIZE(arg1) - VARHDRSZ),
           VARDATA(arg2), VARSIZE(arg2) - VARHDRSZ);
    return new_text;
}
```

假定上述代码已经写在文件`funcs.c`中并编译为共享对象，我们可以用类似下面的命令向PostgreSQL定义这些函数：

```
CREATE FUNCTION add_one(integer) RETURNS integer
     AS 'DIRECTORY/funcs', 'add_one'
     LANGUAGE C STRICT;

-- note overloading of SQL function name "add_one"
CREATE FUNCTION add_one(double precision) RETURNS double precision
     AS 'DIRECTORY/funcs', 'add_one_float8'
     LANGUAGE C STRICT;

CREATE FUNCTION makepoint(point, point) RETURNS point
     AS 'DIRECTORY/funcs', 'makepoint'
     LANGUAGE C STRICT;

CREATE FUNCTION copytext(text) RETURNS text
```

```
     AS 'DIRECTORY/funcs', 'copytext'
     LANGUAGE C STRICT;

CREATE FUNCTION concat_text(text, text) RETURNS text
     AS 'DIRECTORY/funcs', 'concat_text'
     LANGUAGE C STRICT;
```

这里*DIRECTORY*代表共享库文件所在的目录（例如 PostgreSQL的教程目录，其中包含本节示例所用的代码）。（更好的风格是在AS子句中只写'funcs'，前提是已把 *DIRECTORY*加入搜索路径。无论哪种情况，都可以省略共享库的系统特定扩展名，通常是 .so或.sl。）

注意我们把函数声明为“strict”（严格），意思是如果任何输入值为空，系统会自动假定结果为空。这样做可以避免在函数代码中检查空输入。否则，我们就必须显式检查空值，即检查每个传引用参数是否为空指针。（对于传值参数，我们甚至没有办法检查！）

尽管这种调用约定使用简单，但可移植性不好；在某些体系结构上，以这种方式传递小于int的数据类型会有问题。另外，也没有返回空结果的简单方法，除了把函数声明为严格函数之外，也没有其他处理空参数的简单方法。下面介绍的版本 1 约定克服了这些缺点。

35.9.4. 版本 1 的调用约定

版本-1 的调用约定依赖于宏来屏蔽传递参数和结果的大部分复杂性。版本-1 函数的 C 声明总是：

```
Datum funcname(PG_FUNCTION_ARGS)
```

此外，宏调用：

```
PG_FUNCTION_INFO_V1(funcname);
```

必须出现在同一个源文件中（按惯例会正好写在该函数本身之前）。这种宏调用不是internal语言函数所需要的，因为 PostgreSQL会假定所有内部函数都使用版本-1 调用约定。不过，对于动态载入函数是必需的。

在版本-1 函数中，每一个实参都使用对应于该参数数据类型的PG_GETARG_*xxx*()宏取得，结果要用对应于返回类型的PG_RETURN_*xxx*()宏返回。PG_GETARG_*xxx*()的参数是要取得的函数参数的编号，从零开始计。PG_RETURN_*xxx*()的参数是实际要返回的值。

下面我们用版本-1 风格编写与上面相同的函数：

```
#include "postgres.h"
#include <string.h>
#include "fmgr.h"
#include "utils/geo_decls.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

/* by value */

PG_FUNCTION_INFO_V1(add_one);

Datum
add_one(PG_FUNCTION_ARGS)
```

```
{
    int32   arg = PG_GETARG_INT32(0);

    PG_RETURN_INT32(arg + 1);
}

/* by reference, fixed length */

PG_FUNCTION_INFO_V1(add_one_float8);

Datum
add_one_float8(PG_FUNCTION_ARGS)
{
    /* The macros for FLOAT8 hide its pass-by-reference nature. */
    float8   arg = PG_GETARG_FLOAT8(0);

    PG_RETURN_FLOAT8(arg + 1.0);
}

PG_FUNCTION_INFO_V1(makepoint);

Datum
makepoint(PG_FUNCTION_ARGS)
{
    /* Here, the pass-by-reference nature of Point is not hidden. */
    Point     *pointx = PG_GETARG_POINT_P(0);
    Point     *pointy = PG_GETARG_POINT_P(1);
    Point     *new_point = (Point *) palloc(sizeof(Point));

    new_point->x = pointx->x;
    new_point->y = pointy->y;

    PG_RETURN_POINT_P(new_point);
}

/* by reference, variable length */

PG_FUNCTION_INFO_V1(copytext);

Datum
copytext(PG_FUNCTION_ARGS)
{
    text     *t = PG_GETARG_TEXT_P(0);
    /*
     * VARSIZE is the total size of the struct in bytes.
     */
    text     *new_t = (text *) palloc(VARSIZE(t));
    SET_VARSIZE(new_t, VARSIZE(t));
    /*
     * VARDATA is a pointer to the data region of the struct.
     */
    memcpy((void *) VARDATA(new_t), /* destination */
           (void *) VARDATA(t),     /* source */
           VARSIZE(t) - VARHDRSZ);  /* how many bytes */
```

```
    PG_RETURN_TEXT_P(new_t);
}

PG_FUNCTION_INFO_V1(concat_text);

Datum
concat_text(PG_FUNCTION_ARGS)
{
    text  *arg1 = PG_GETARG_TEXT_P(0);
    text  *arg2 = PG_GETARG_TEXT_P(1);
    int32 new_text_size = VARSIZE(arg1) + VARSIZE(arg2) - VARHDRSZ;
    text *new_text = (text *) palloc(new_text_size);

    SET_VARSIZE(new_text, new_text_size);
    memcpy(VARDATA(new_text), VARDATA(arg1), VARSIZE(arg1) - VARHDRSZ)
;
    memcpy(VARDATA(new_text) + (VARSIZE(arg1) - VARHDRSZ),
           VARDATA(arg2), VARSIZE(arg2) - VARHDRSZ);
    PG_RETURN_TEXT_P(new_text);
}
```

这些函数的`CREATE FUNCTION`命令与版本 0 的等价形式相同。

乍看之下，版本 1 的编码约定似乎只是无谓的故弄玄虚。不过，它们确实提供了许多改进，因为宏可以隐藏不必要的细节。例如，在编码`add_one_float8`时，我们不再需要知道`float8`是传引用类型。另一个例子是，变长类型的`GETARG`宏允许更有效地取得 “toasted”（压缩或外置）值。

版本-1 函数的一大改进是对空输入和结果的更好处理。宏`PG_ARGISNULL(n)`允许一个函数测试是否每一个输入为空值（当然，只需要在没有声明为 “strict” 的函数中这样做）。和`PG_GETARG_xxx()`宏一样，输入参数也是从零开始计数。
注意应该在验证了一个参数不是空值之后才执行`PG_GETARG_xxx()`。要返回空值结果，应执行`PG_RETURN_NULL()`，它对严格的以及非严格的函数都有用。

新式接口中提供的其他选项是`PG_GETARG_xxx()`宏的两个变种。其中的第一种是`PG_GETARG_xxx_COPY()`，
它确保返回的指定参数的拷贝可以被安全地写入（通常的宏有时会返回一个指向表中物理存储值的指针，不能写入该值。使用`PG_GETARG_xxx_COPY()`宏可以保证得到一个可写的结果）。
第二种变种`PG_GETARG_xxx_SLICE()`宏有三个参数。第一个是函数参数的编号（如上文）。
第二个和第三个是要被返回的段的偏移量和长度。偏移量从零开始计算，
而负值的长度则表示要求返回该值的剩余部分。当大型值的存储类型为 “external” 时，
这些宏提供了访问这些大型值的部分内容的更有效方法（列的存储类型可以使用`ALTER TABLE tablename ALTER COLUMN colname SET STORAGE storagetype`来指定。`storagetype`取`plain`、`external`、`extended`或者`main`）。

最后，版本-1 的函数调用约定可以返回集合结果（第 35.9.9 节）、实现触发器函数（第 36 章）和过程语言调用处理器（第 49 章）。版本-1 代码也比版本-0 更可移植，因为它不违反 C 标准中关于函数调用协议的限制。更多细节可见源代码发布中的`src/backend/utils/fmgr/README`。

35.9.5. 编写代码

在开始更高级的话题之前，我们应该讨论一下用于 PostgreSQL C 语言函数的编码规则。虽然有可能把不是 C 编写的函数载入到 PostgreSQL中，但即使能够做到，通常也很困难，因为其他语言

（例如 C++、FORTRAN 或者 Pascal）通常不会遵循和 C 相同的调用约定。也就是说，其他语言不会以同样的方式在函数之间传递参数以及返回值。由于这个原因，我们会假定你的 C 语言函数确实是用 C 编写的。

编写和构建 C 语言函数的基本规则如下：

- 使用 `pg_config --includedir-server` 查出 PostgreSQL 服务器头文件在你的系统上（或你的用户将要运行的系统上）安装于何处。
- 为了让你的代码能够被 PostgreSQL 动态装入，编译和链接时总是需要特殊的选项。关于如何在特定操作系统上完成这件事，详见第 35.9.6 节。
- 记得按照第 35.9.1 节中的说明，为你的共享库定义一个“魔数块”。
- 分配内存时，使用 PostgreSQL 提供的 `palloc` 和 `pfree`，而不是相应的 C 库函数 `malloc` 和 `free`。用 `palloc` 分配的内存会在每个事务结束时自动释放，从而避免内存泄漏。
- 总是使用 `memset` 将结构体的所有字节清零（或者一开始就用 `palloc0` 来分配它们）。即使你给结构体的每个字段都赋了值，
  结构体中仍可能存在包含垃圾值的对齐填充字节（也就是结构体中的空洞）。如果不这么做，
  就很难支持哈希索引或哈希连接，
  因为那时你必须只挑出数据结构中真正有意义的位来计算哈希值。
  规划器有时也依赖按位相等来比较常量，因此如果逻辑上等价的值在按位上不相等，
  就可能得到不理想的规划结果。
- PostgreSQL 的大多数内部类型都在 `postgres.h` 中声明，而函数管理器接口（`PG_FUNCTION_ARGS` 等）位于 `fmgr.h` 中，因此至少需要包含这两个文件。出于可移植性考虑，最好把 `postgres.h` 放在最前面，先于任何其他系统或用户头文件。包含 `postgres.h` 时，也会顺带为你包含 `elog.h` 和 `palloc.h`。
- 目标文件中定义的符号名不能彼此冲突，也不能与 PostgreSQL 服务器可执行文件中定义的符号冲突。如果你收到这类错误消息，就必须重命名相关函数或变量。

35.9.6. 编译和链接动态装载函数

在你能够使用以 C 编写的 PostgreSQL 扩展函数之前，必须以特殊方式对它们进行编译和链接，以生成一个可由服务器动态装载的文件。更准确地说，需要创建一个共享库。

若想了解本节未涵盖的信息，你应阅读操作系统的文档，特别是 C 编译器 `cc` 和链接编辑器 `ld` 的手册页。此外，PostgreSQL 源代码在 `contrib` 目录中包含若干可用的示例。不过，如果你依赖这些示例，就会使你的模块依赖于 PostgreSQL 源代码是否可用。

创建共享库通常与链接可执行文件类似：先把源文件编译为目标文件，
再把目标文件链接在一起。目标文件需要以位置无关代码（PIC）形式生成。从概念上讲，
这意味着当它们被可执行文件装载时，可以放在内存中的任意位置。
（面向可执行文件的目标文件通常不会这样编译。）链接共享库的命令中也包含一些特殊标志，
用来把它与链接可执行文件的命令区分开来（至少理论上如此，
某些系统上的实际做法要丑陋得多）。

在下面的示例中，我们假定你的源代码位于文件 `foo.c` 中，并将创建共享库 `foo.so`。除非另有说明，中间目标文件名为 `foo.o`。共享库可以包含多个目标文件，但这里我们只使用一个。

BSD/OS

生成 PIC 的编译器选项是 `-fpic`。创建共享库时使用的链接器选项是 `-shared`。

```
gcc -fpic -c foo.c
ld -shared -o foo.so foo.o
```

这从 BSD/OS 4.0 版起适用。

FreeBSD

生成 PIC 的编译器选项是 `-fpic`。创建共享库时使用的编译器选项是 `-shared`。

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

这从 FreeBSD 3.0 版起适用。

HP-UX

生成 PIC 的系统编译器选项是 `+z`。使用 GCC 时则是 `-fpic`。用于共享库的链接器选项是 `-b`。因此：

```
cc +z -c foo.c
```

or:

```
gcc -fpic -c foo.c
```

and then:

```
ld -b -o foo.sl foo.o
```

与大多数其他系统不同，HP-UX 使用 `.sl` 作为共享库扩展名。

IRIX

PIC 是默认行为，无须特殊的编译器选项。生成共享库时使用的链接器选项是 `-shared`。

```
cc -c foo.c
ld -shared -o foo.so foo.o
```

Linux

生成 PIC 的编译器选项是 `-fpic`。在某些平台的某些情况下，如果 `-fpic` 不起作用，则必须使用 `-fPIC`。更多信息请参阅 GCC 手册。创建共享库的编译器选项是 `-shared`。完整示例如下：

```
cc -fpic -c foo.c
cc -shared -o foo.so foo.o
```

Mac OS X

下面是一个示例。它假定开发者工具已经安装。

```
cc -c foo.c
cc -bundle -flat_namespace -undefined suppress -o foo.so foo.o
```

NetBSD

生成 PIC 的编译器选项是 `-fpic`。对于 ELF 系统，使用带 `-shared` 选项的编译器来链接共享库。在较旧的非 ELF 系统上，则使用 `ld -Bshareable`。

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

OpenBSD

生成 PIC 的编译器选项是 `-fpic`。链接共享库使用 `ld -Bshareable`。

```
gcc -fpic -c foo.c
ld -Bshareable -o foo.so foo.o
```

Solaris

使用 Sun 编译器时，生成 PIC 的编译器选项是 `-KPIC`；使用 GCC 时则是 `-fpic`。要链接共享库，两种编译器都使用编译器选项 `-G`，或者在使用 GCC 时改用 `-shared`。

```
cc -KPIC -c foo.c
cc -G -o foo.so foo.o
```

or

```
gcc -fpic -c foo.c
gcc -G -o foo.so foo.o
```

Tru64 UNIX

PIC 是默认值，因此编译命令就是常规的那条。链接时需要使用带特殊选项的`ld`。

```
cc -c foo.c
ld -shared -expect_unresolved '*' -o foo.so foo.o
```

使用 GCC 代替系统编译器时过程相同；不需要特殊选项。

UnixWare

生成 PIC 的编译器选项，对于 SCO 编译器是 `-K PIC`，而 `-fpic` 则用于 GCC。链接共享库时，编译器选项对于 SCO 编译器是 `-G`，对于 GCC 则是 `-shared`。

```
cc -K PIC -c foo.c
cc -G -o foo.so foo.o
```

或者

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

提示

如果这些内容对你来说过于复杂，可以考虑使用 GNU Libtool[1]，它通过统一接口隐藏了平台差异。

生成的共享库文件随后就可以装载到 PostgreSQL 中。在向 `CREATE FUNCTION` 命令指定文件名时，必须给出共享库文件名，而不是中间目标文件名。请注意，系统标准的共享

[1] http://www.gnu.org/software/libtool/

库扩展名（通常是 .so 或 .sl）可以在 CREATE FUNCTION 命令中省略，并且通常也应省略，以获得最佳可移植性。

关于服务器期望在何处找到共享库文件，请回头参见第 35.9.1 节。

35.9.7. 复合类型参数

复合类型没有像 C 结构体那样的固定布局。复合类型的实例可能包含空值字段。此外，继承层次中的复合类型可能具有和同一继承层次中其他成员不同的字段。因此，PostgreSQL提供了函数接口以便从 C 访问复合类型的字段。

假设我们想编写一个函数来回答如下查询：

```
SELECT name, c_overpaid(emp, 1500) AS overpaid
    FROM emp
    WHERE name = 'Bill' OR name = 'Sam';
```

使用版本 0 调用约定，我们可以将 c_overpaid 定义为：

```
#include "postgres.h"
#include "executor/executor.h"  /* for GetAttributeByName() */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

bool
c_overpaid(HeapTupleHeader t, /* the current row of emp */
           int32 limit)
{
    bool isnull;
    int32 salary;

    salary = DatumGetInt32(GetAttributeByName(t, "salary", &isnull));
    if (isnull)
        return false;
    return salary > limit;
}
```

按照版本-1 编码，上面的函数写成这样：

```
#include "postgres.h"
#include "executor/executor.h"  /* for GetAttributeByName() */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

PG_FUNCTION_INFO_V1(c_overpaid);

Datum
c_overpaid(PG_FUNCTION_ARGS)
{
    HeapTupleHeader  t = PG_GETARG_HEAPTUPLEHEADER(0);
```

```
    int32            limit = PG_GETARG_INT32(1);
    bool isnull;
    Datum salary;

    salary = GetAttributeByName(t, "salary", &isnull);
    if (isnull)
        PG_RETURN_BOOL(false);
    /* Alternatively, we might prefer to do PG_RETURN_NULL() for null
 salary. */

    PG_RETURN_BOOL(DatumGetInt32(salary) > limit);
}
```

`GetAttributeByName` 是 PostgreSQL 的一个系统函数，用于从指定行中取出属性。它有三个参数：传入函数的 `HeapTupleHeader` 类型参数、所需属性的名称，以及一个用于指示该属性是否为 null 的返回参数。`GetAttributeByName` 返回一个 `Datum` 值，你可以用适当的 `Datum Get*XXX*()` 宏把它转换为正确的数据类型。注意，如果 null 标志被设置，那么返回值本身没有意义；在尝试对结果做任何处理之前，务必先检查这个 null 标志。

也有`GetAttributeByNum`函数，它可以用目标属性的列号而不是属性名来选择目标属性。

下面的命令声明 SQL 中的`c_overpaid`：

```
CREATE FUNCTION c_overpaid(emp, integer) RETURNS boolean
    AS 'DIRECTORY/funcs', 'c_overpaid'
    LANGUAGE C STRICT;
```

注意我们用了`STRICT`，这样我们不需要检查输入参数是否为 NULL。

35.9.8. 返回行（复合类型）

要从 C 语言函数中返回一行或一个复合类型值，可以使用一套特殊的 API，它通过一组宏和函数隐藏了构造复合数据类型时的大部分复杂性。要使用这套 API，源文件中必须包含：

```
#include "funcapi.h"
```

构造复合数据值（下文简称“元组”）有两种方式：一种是从 Datum 值数组构造，另一种是从 C 字符串数组构造，这些字符串会传给该元组各列数据类型的输入转换函数。无论采用哪种方式，首先都需要获取或构造描述该元组结构的 `TupleDesc`。处理 Datum 时，需要把 `TupleDesc` 传给 `BlessTupleDesc`，然后为每一行调用 `heap_form_tuple`。处理 C 字符串时，则要把 `TupleDesc` 传给 `TupleDescGetAttInMetadata`，然后为每一行调用 `BuildTupleFromCStrings`。对于返回元组集合的函数，这些准备步骤可以在第一次调用函数时一次性完成。

有一些辅助函数可以用来设置所需的 `TupleDesc`。在大多数返回复合值的函数中，推荐的做法是调用：

```
TypeFuncClass get_call_result_type(FunctionCallInfo fcinfo,
                                   Oid *resultTypeId,
                                   TupleDesc *resultTupleDesc)
```

传入调用函数本身收到的同一个 `fcinfo` 结构体（这当然要求使用版本 1 调用约定）。`resultTypeId` 可以指定为 `NULL`，也可以指定为一个本地变量的地址，用于接收函数结果类型的 OID。`resultTupleDesc` 应当是一个本地 `TupleDesc` 变量的地址。检查返回结果是否为 `TYPEFUNC_COMPOSITE`；如果是，`resultTupleDesc` 就会被填入所需的 `TupleDesc`。（如

果不是，可以报告一个类似 “function returning record called in context that cannot accept type record” 的错误。）

提示

`get_call_result_type`能够解析一个多态函数结果的实际类型，
因此不仅在返回复合类型的函数中，在返回标量多态结果的函数中它也是非常有用的。
`resultTypeId`输出主要用于返回多态标量的函数。

注意

`get_call_result_type`有一个兄弟 `get_expr_result_type`，它被用来解析被表示为一棵表达式树的函数调用的输出类型。
在尝试从函数自身外部确定结果类型时可以用它。也有一个`get_func_result_type`，当只有函数的 OID 可用时可以用它。不过这些函数无法处理被声明为返回`record`的函数，并且`get_func_result_type`无法解析多态类型，因此你应该优先使用`get_call_result_type`。

更早、现在已被弃用的获取`TupleDesc`的函数有：

```
TupleDesc RelationNameGetTupleDesc(const char *relname)
```

它可以为一个指定名称的关系的行类型得到`TupleDesc`，还有：

```
TupleDesc TypeGetTupleDesc(Oid typeoid, List *colaliases)
```

可以基于一个类型 OID 得到`TupleDesc`。这可以被用来为一种基本或者复合类型获得`TupleDesc`。不过，对于返回`record`的函数它不起作用，并且它无法解析多态类型。

一旦有了一个`TupleDesc`，如果计划处理 Datum，可以调用：

```
TupleDesc BlessTupleDesc(TupleDesc tupdesc)
```

如果计划处理 C 字符串，可调用：

```
AttInMetadata *TupleDescGetAttInMetadata(TupleDesc tupdesc)
```

如果正在编写一个返回集合的函数，你可以把这些函数的结果保存在 `FuncCallContext`结构体中 – 分别使用 `tuple_desc`或者`attinmeta`字段。

在处理 Datum 时，使用

```
HeapTuple heap_form_tuple(TupleDesc tupdesc, Datum *values, bool
 *isnull)
```

来用 Datum 形式的用户数据构建一个`HeapTuple`。

在处理 C 字符串时，使用

```
HeapTuple BuildTupleFromCStrings(AttInMetadata *attinmeta, char
 **values)
```

来用 C 字符串形式的用户数据构建一个HeapTuple。values是一个 C 字符串数组，每一个元素是返回行的一个属性。每一个 C 字符串应该是该属性数据类型的输入函数所期望的格式。为了对一个属性返回空值，*values*数组中对应的指针应该被设置为NULL。
对于你返回的每一行都需要再次调用这个函数。

一旦已经构建了一个要从函数中返回的元组，它必须被转换成一个 Datum。使用

```
HeapTupleGetDatum(HeapTuple tuple)
```

可把一个HeapTuple转换成合法的 Datum。如果你只想返回一行，那么这个Datum可以被直接返回，在一个集合返回函数中它也可以被当做当前的返回值。

下一节中会有一个示例。

35.9.9. 返回集合

还有一种特殊的 API，为从 C 语言函数返回集合（多行）提供支持。集合返回函数必须遵循版本 1 调用约定。此外，与前述一样，源文件必须包含funcapi.h。

集合返回函数（SRF）为它返回的每一项被调用一次。因此，SRF 必须保存足够的状态，以记住它先前在做什么，并在每次调用时返回下一项。结构体 FuncCallContext 就是用来帮助控制这一过程的。在函数内部，fcinfo->flinfo->fn_extra 用于在调用之间保存指向 FuncCallContext 的指针。

```
typedef struct
{
    /*
     * 本次调用以前已经被调用过多少次
     *
     * SRF_FIRSTCALL_INIT() 会为你把 call_cntr 初始化为 0，
     * 并且在每次调用 SRF_RETURN_NEXT() 时增加。
     */
    uint32 call_cntr;

    /*
     * 可选：最大调用次数
     *
     * 这里的 max_calls 只是为了方便，设置它是可选的。
     * 如果没有设置，你必须提供替代的方法来了解函数什么时候做完。
     */
    uint32 max_calls;

    /*
     * 可选：指向结果槽的指针
     *
     * 此字段已过时，仅为向后兼容而保留，即兼容使用已弃用的
     * TupleDescGetSlot() 的用户定义 SRF。
     */
    TupleTableSlot *slot;

    /*
     * 可选：指向用户提供的上下文信息的指针
     *
     * user_fctx 是一个指向你自己的数据的指针，它可用来在函数的多次
```

```
     * 调用之间保存任意的上下文信息。
     */
    void *user_fctx;

    /*
     * 可选：指向包含属性类型输入元数据的结构的指针
     *
     * attinmeta 被用在返回元组（即复合数据类型）时，在返回基础类型
     * 时不会使用。只有想用BuildTupleFromCStrings()创建返回元组时才需要它。
     */
    AttInMetadata *attinmeta;

    /*
     * 用于保存必须在多次调用间都存在的结构的内存上下文
     *
     * SRF_FIRSTCALL_INIT() 会为你设置 multi_call_memory_ctx，并且由
     * SRF_RETURN_DONE() 来清理。对于任何需要在 SRF 的多次调用间都
     * 存在的内存来说，它是最合适的内存上下文。
     */
    MemoryContext multi_call_memory_ctx;

    /*
     * 可选：指向包含元组描述的结构的指针
     *
     * tuple_desc 被用在返回元组（即复合数据类型）时，并且只有在用
     * heap_form_tuple() 而不是 BuildTupleFromCStrings() 构建元组时才需要
它。
     * 注意这里存储的 TupleDesc 指针通常已经被先运行过 BlessTupleDesc()。
     */
    TupleDesc tuple_desc;

} FuncCallContext;
```

SRF 会使用若干自动操纵 `FuncCallContext` 结构体（并期望通过 `fn_extra` 找到它）的函数和宏。使用：

```
SRF_IS_FIRSTCALL()
```

来判断你的函数是第一次被调用还是后续调用。在第一次调用时（只在第一次调用时）使用：

```
SRF_FIRSTCALL_INIT()
```

来初始化 `FuncCallContext`。在每次函数调用时，包括第一次，使用：

```
SRF_PERCALL_SETUP()
```

来正确设置对 `FuncCallContext` 的使用，并清除上一次处理遗留的任何已返回数据。

如果你的函数有数据要返回，使用：

```
SRF_RETURN_NEXT(funcctx, result)
```

把它返回给调用者（`result`必须是类型`Datum`，
可以是一个单一值或者按上文所述准备好的元组）。最后，当函数完成了数据返回后，可使用：

```
SRF_RETURN_DONE(funcctx)
```

来清理并且结束SRF。

调用 SRF 时当前所处的内存上下文是一个瞬时上下文，它会在两次调用之间被清空。这意味着，你不必对用 palloc 分配的所有东西都调用 pfree，因为它们反正会自动释放。不过，如果你需要分配在多次调用之间持续存在的数据结构，就必须把它们放到别处。对于任何需要一直存活到 SRF 运行结束的数据，multi_call_memory_ctx 所指向的内存上下文就是合适的位置。在大多数情况下，这意味着你应当在做首次调用初始化时切换到 multi_call_memory_ctx。

一个完整的伪代码示例：

```
Datum
my_set_returning_function(PG_FUNCTION_ARGS)
{
    FuncCallContext  *funcctx;
    Datum             result;
    further declarations as needed

    if (SRF_IS_FIRSTCALL())
    {
        MemoryContext oldcontext;

        funcctx = SRF_FIRSTCALL_INIT();
        oldcontext = MemoryContextSwitchTo(funcctx->multi_call_memory_
ctx);
        /* 这里是一次性设置代码： */
        user code
        if returning composite
            build TupleDesc, and perhaps AttInMetadata
        endif returning composite
        user code
        MemoryContextSwitchTo(oldcontext);
    }

    /* 这里是每一次都要做的设置代码： */
    user code
    funcctx = SRF_PERCALL_SETUP();
    user code

    /* 这里只是一种测试是否执行完的方法： */
    if (funcctx->call_cntr < funcctx->max_calls)
    {
        /* 这里返回另一个项： */
        user code
        obtain result Datum
        SRF_RETURN_NEXT(funcctx, result);
    }
    else
    {
        /* 这里已经完成了项的返回，只需要做清理： */
        user code
        SRF_RETURN_DONE(funcctx);
    }
}
```

一个返回复合类型的简单 SRF 的完整示例如下：

```
PG_FUNCTION_INFO_V1(retcomposite);

Datum
retcomposite(PG_FUNCTION_ARGS)
{
    FuncCallContext     *funcctx;
    int                  call_cntr;
    int                  max_calls;
    TupleDesc            tupdesc;
    AttInMetadata       *attinmeta;

    /* stuff done only on the first call of the function */
    if (SRF_IS_FIRSTCALL())
    {
        MemoryContext   oldcontext;

        /* create a function context for cross-call persistence */
        funcctx = SRF_FIRSTCALL_INIT();

        /* switch to memory context appropriate for multiple function
 calls */
        oldcontext = MemoryContextSwitchTo(funcctx->multi_call_memory_
ctx);

        /* total number of tuples to be returned */
        funcctx->max_calls = PG_GETARG_UINT32(0);

        /* Build a tuple descriptor for our result type */
        if (get_call_result_type(fcinfo, NULL, &tupdesc) != TYPEFUNC_
COMPOSITE)
            ereport(ERROR,
                    (errcode(ERRCODE_FEATURE_NOT_SUPPORTED),
                     errmsg("function returning record called in
 context "
                            "that cannot accept type record")));

        /*
         * generate attribute metadata needed later to produce tuples
 from raw
         * C strings
         */
        attinmeta = TupleDescGetAttInMetadata(tupdesc);
        funcctx->attinmeta = attinmeta;

        MemoryContextSwitchTo(oldcontext);
    }

    /* stuff done on every call of the function */
    funcctx = SRF_PERCALL_SETUP();

    call_cntr = funcctx->call_cntr;
    max_calls = funcctx->max_calls;
    attinmeta = funcctx->attinmeta;
```

```
    if (call_cntr < max_calls)    /* do when there is more left to
 send */
    {
        char       **values;
        HeapTuple    tuple;
        Datum        result;

        /*
         * Prepare a values array for building the returned tuple.
         * This should be an array of C strings which will
         * be processed later by the type input functions.
         */
        values = (char **) palloc(3 * sizeof(char *));
        values[0] = (char *) palloc(16 * sizeof(char));
        values[1] = (char *) palloc(16 * sizeof(char));
        values[2] = (char *) palloc(16 * sizeof(char));

        snprintf(values[0], 16, "%d", 1 * PG_GETARG_INT32(1));
        snprintf(values[1], 16, "%d", 2 * PG_GETARG_INT32(1));
        snprintf(values[2], 16, "%d", 3 * PG_GETARG_INT32(1));

        /* build a tuple */
        tuple = BuildTupleFromCStrings(attinmeta, values);

        /* make the tuple into a datum */
        result = HeapTupleGetDatum(tuple);

        /* clean up (this is not really necessary) */
        pfree(values[0]);
        pfree(values[1]);
        pfree(values[2]);
        pfree(values);

        SRF_RETURN_NEXT(funcctx, result);
    }
    else    /* do when there is no more left */
    {
        SRF_RETURN_DONE(funcctx);
    }
}
```

在 SQL 中声明此函数的一种方法是：

```
CREATE TYPE __retcomposite AS (f1 integer, f2 integer, f3 integer);

CREATE OR REPLACE FUNCTION retcomposite(integer, integer)
    RETURNS SETOF __retcomposite
    AS 'filename', 'retcomposite'
    LANGUAGE C IMMUTABLE STRICT;
```

另一种方式是使用 OUT 参数：

```
CREATE OR REPLACE FUNCTION retcomposite(IN integer, IN integer,
    OUT f1 integer, OUT f2 integer, OUT f3 integer)
    RETURNS SETOF record
```

```
    AS 'filename', 'retcomposite'
    LANGUAGE C IMMUTABLE STRICT;
```

请注意，采用这种方法时，函数的输出类型在形式上是一个匿名的 record 类型。

源码发行版中的contrib/tablefunc模块包含更多集合返回函数的示例。

35.9.10. 多态参数和返回类型

C 语言函数可以声明为接受和返回多态类型 anyelement、anyarray、anynonarray 和 anyenum。关于多态函数的更详细解释，见第 35.2.5 节。当函数参数或者返回类型被定义为多态类型时，
函数的编写者无法提前知道会用什么数据类型调用该函数或者该函数需要返回什么数据类型。fmgr.h 中提供了两种例程，允许版本-1 的 C 函数发现其参数的实际数据类型以及它要返回的类型。这些例程被称为 get_fn_expr_rettype(FmgrInfo *flinfo) 和 get_fn_expr_argtype(FmgrInfo *flinfo, int argnum)。它们返回结果或者参数的类型的 OID，或者当该信息不可用时返回 InvalidOid。结构体 flinfo 通常被当做 fcinfo->flinfo 访问。参数 argnum 则是从零开始计。get_call_result_type 也可被用作 get_fn_expr_rettype 的替代。

例如，假设我们想要写一个接收一个任意类型元素并且返回一个该类型的一维数组的函数：

```
PG_FUNCTION_INFO_V1(make_array);
Datum
make_array(PG_FUNCTION_ARGS)
{
    ArrayType  *result;
    Oid         element_type = get_fn_expr_argtype(fcinfo->flinfo, 0);
    Datum       element;
    bool        isnull;
    int16       typlen;
    bool        typbyval;
    char        typalign;
    int         ndims;
    int         dims[MAXDIM];
    int         lbs[MAXDIM];

    if (!OidIsValid(element_type))
        elog(ERROR, "could not determine data type of input");

    /* 得到提供的元素，小心它为 NULL 的情况 */
    isnull = PG_ARGISNULL(0);
    if (isnull)
        element = (Datum) 0;
    else
        element = PG_GETARG_DATUM(0);

    /* 只有一个维度 */
    ndims = 1;
    /* 和一个元素 */
    dims[0] = 1;
    /* 且下界是 1 */
    lbs[0] = 1;
```

```
    /* 得到该元素类型所需的信息 */
    get_typlenbyvalalign(element_type, &typlen, &typbyval, &typalign);

    /* 现在构建数组 */
    result = construct_md_array(&element, &isnull, ndims, dims, lbs,
                                element_type, typlen, typbyval,
 typalign);

    PG_RETURN_ARRAYTYPE_P(result);
}
```

下面的命令在 SQL 中声明了 `make_array` 函数：

```
CREATE FUNCTION make_array(anyelement) RETURNS anyarray
    AS 'DIRECTORY/funcs', 'make_array'
    LANGUAGE C IMMUTABLE;
```

还有一种只对 C 语言函数可用的多态变体：它们可以被声明为接受 `"any"` 类型的参数。（注意，这个类型名必须用双引号括起来，因为它同时也是 SQL 保留字。）它与 `anyelement` 类似，但不会要求不同的 `"any"` 参数必须是同一种类型，也不会帮助确定函数的结果类型。C 语言函数还可以把最后一个参数声明为 `VARIADIC "any"`。这可以匹配一个或多个任意类型的实参（不必是同一种类型）。这些参数不会像普通可变参数函数那样被收集成一个数组，而是会单独传给函数。使用这种特性时，必须结合 `PG_NARGS()` 宏以及前面介绍的方法来确定实参的个数和类型。

35.9.11. 共享内存与 LWLocks

附加模块可以在服务器启动时预留 LWLock 和一块共享内存。附加模块的共享库必须通过在以下参数中指定它来预装载：shared_preload_libraries。共享内存可以通过下面的调用来预留：

```
void RequestAddinShmemSpace(int size)
```

该调用应放在你的 `_PG_init` 函数中。

LWLock 可以通过下面的调用来预留：

```
void RequestAddinLWLocks(int n)
```

该调用应放在 `_PG_init` 中。

为避免可能的竞争条件，每个后端都应使用 LWLock `AddinShmemInitLock` 来连接和初始化分配给它的共享内存，如下所示：

```
static mystruct *ptr = NULL;

if (!ptr)
{
        bool    found;

        LWLockAcquire(AddinShmemInitLock, LW_EXCLUSIVE);
        ptr = ShmemInitStruct("my struct name", size, &found);
        if (!found)
        {
                initialize contents of shmem area;
                acquire any requested LWLocks using:
                ptr->mylockid = LWLockAssign();
```

```
        }
        LWLockRelease(AddinShmemInitLock);
}
```

35.9.12. 把 C++ 用于可扩展性

尽管PostgreSQL后端是用 C 编写的，只要遵循下面的指导方针也可以用 C++ 编写扩展：

- 所有被后端访问的函数必须对后端呈现一种 C 接口，然后这些 C 函数调用 C++ 函数。例如，被后端访问的函数要求`extern C` 链接。对需要在后端和 C++ 代码之间作为指针传递的任何函数也要这样做。
- 使用合适的释放方法释放内存。例如，大部分后端内存是通过 `palloc()`分配的，所以应使用`pfree()` 来释放。在这种情况中使用 C++ 的`delete`会失败。
- 防止异常传播到 C 代码中（在所有`extern C`函数的顶层使用一个捕捉全部异常的块）。即使 C++ 代码不会显式地抛出任何异常也需要这样做，因为类似内存不足等事件仍会抛出异常。任何异常都必须被捕捉并且用适当的错误传回给 C 接口。如果可能，用 `-fno-exceptions` 来编译 C++ 以完全消灭异常。在这种情况下，你必须在 C++ 代码中检查失败，例如检查`new()` 返回的 NULL。
- 如果从 C++ 代码调用后端函数，确保 C++ 调用栈只包含传统 C 风格的数据结构（POD）。这是必要的，因为后端错误会产生远距离的`longjmp()`，它无法正确地展开具有非 POD 对象的 C++ 调用栈。

总之，最好把 C++ 代码放在与后端交互的`extern C`函数之后，并且避免异常、内存和调用栈泄漏。

35.10. 用户定义的聚合

PostgreSQL 中的聚合函数是依据状态值和状态转移函数来定义的。也就是说，聚合通过一个状态值来工作，并在处理每个后续输入行时更新该状态值。要定义一个新的聚合函数，
需要为状态值选定一种数据类型、为状态选定一个初始值，并指定一个状态转移函数。
状态转移函数只是普通函数，也可以在聚合的上下文之外使用。
如果聚合的期望结果不同于运行状态值中需要保存的数据，还可以指定一个最终函数。

因此，除了聚合用户所看到的参数和结果数据类型之外，还存在一种内部状态值数据类型，
它可能既不同于参数类型，也不同于结果类型。

如果我们定义一个不使用最终函数的聚合，
那么得到的就是一个对每一行的列值进行逐步计算的聚合。`sum` 就是这种聚合的一个示例。`sum` 从零开始，并始终把当前行的值加到运行总和中。例如，如果我们希望让 `sum` 聚合能用于复数数据类型，那么只需要该数据类型的加法函数。聚合定义如下：

```
CREATE AGGREGATE sum (complex)
(
    sfunc = complex_add,
    stype = complex,
    initcond = '(0,0)'
);

SELECT sum(a) FROM test_complex;
```

```
   sum
-----------
 (34,53.9)
```

（注意，这里我们依赖的是函数重载：名为 sum 的聚合不止一个，但 PostgreSQL 能判断出哪种 sum 适用于 complex 类型的列。）

如果没有非空输入值，上面这个 sum 的定义会返回零（初始状态条件）。也许我们希望在这种情况下返回 null —— SQL 标准就期望 sum 这样表现。只需省略 initcond 短语，使初始状态条件为 null，就可以做到这一点。通常这意味着 sfunc 需要检查空状态条件输入。但对于 sum 以及其他一些简单聚合（如 max 和 min），只需把第一个非空输入值插入状态变量，然后从第二个非空输入值开始应用转移函数即可。如果初始条件为 null 且转移函数被标记为“strict”（即不为空输入调用），PostgreSQL 会自动完成这项工作。

“strict” 转移函数的另一项默认行为是，只要遇到空值输入，就会保留先前的状态值不变。
因此，空值会被忽略。如果你需要对空值输入采用其他行为，就不要把转移函数声明为 strict，而应在函数代码中自行检测空值输入并完成所需处理。

avg（平均值）是更复杂的聚合示例。它需要两个运行状态：输入的总和与输入的个数。
最终结果由二者相除得到。平均值通常用数组作为状态值来实现。例如，avg(float8) 的内置实现如下所示：

```
CREATE AGGREGATE avg (float8)
(
    sfunc = float8_accum,
    stype = float8[],
    finalfunc = float8_avg,
    initcond = '{0,0,0}'
);
```

（float8_accum 需要一个三元素数组，而不只是两个元素，因为它累积的不仅是输入值的和与计数，还有平方和。这样它除了可用于 avg 之外，也可用于其他一些聚合。）

聚合函数可以使用多态的状态转移函数或最终函数，这样同一组函数就能用于实现多个聚合。
关于多态函数的解释，请参见第 35.2.5 节。再进一步，聚合函数自身也可以指定为具有多态输入类型和状态类型，从而让同一个聚合定义服务于多种输入数据类型。
下面是一个多态聚合的示例：

```
CREATE AGGREGATE array_accum (anyelement)
(
    sfunc = array_append,
    stype = anyarray,
    initcond = '{}'
);
```

对于任何给定的聚合调用，其实际状态类型都是一种数组类型，
该数组以实际输入类型为元素类型。这个聚合的行为是把所有输入串接成这种类型的一个数组。
（注意：内置聚合 array_agg 提供了类似的功能，而且性能比这个定义更好。）

下面是以两种不同的实际数据类型作为参数时的输出：

```
SELECT attrelid::regclass, array_accum(attname)
    FROM pg_attribute
    WHERE attnum > 0 AND attrelid = 'pg_tablespace'::regclass
    GROUP BY attrelid;
```

```
   attrelid    |              array_accum
---------------+---------------------------------------
 pg_tablespace | {spcname,spcowner,spclocation,spcacl}
(1 row)

SELECT attrelid::regclass, array_accum(atttypid::regtype)
    FROM pg_attribute
    WHERE attnum > 0 AND attrelid = 'pg_tablespace'::regclass
    GROUP BY attrelid;

   attrelid    |        array_accum
---------------+---------------------------
 pg_tablespace | {name,oid,text,aclitem[]}
(1 row)
```

用 C 编写的函数可以通过调用 `AggCheckCallContext` 来检测自己是否作为聚合转移函数或最终函数被调用，例如：

```
if (AggCheckCallContext(fcinfo, NULL))
```

检查这一点的一个原因是，如果对状态转移函数来说结果为真，
那么第一个输入一定是临时转移值，因此可以安全地原地修改，而不必分配一个新副本。示例见 `int8inc()`。（这是函数可以安全地修改按引用传递的输入的唯一情况。特别是，
聚合的最终函数在任何情况下都不得修改其输入，因为在某些情况下，
它们会在同一个最终转移值上被再次执行。）

更多细节请参见CREATE AGGREGATE命令。

35.11. 用户定义的类型

如第 35.2 节中所述，PostgreSQL可以扩展以支持新的数据类型。本节描述如何定义新的基础类型，也就是在SQL语言层之下定义的数据类型。创建新的基础类型需要用底层语言（通常是 C）实现操作该类型的函数。

本节中的示例位于源码分发包的`src/tutorial`目录中的 `complex.sql`和`complex.c`。关于如何运行这些示例，请参见该目录中的`README`文件。

用户定义类型必须始终具有输入函数和输出函数。这些函数决定该类型如何以字符串形式出现（供用户输入和向用户输出），以及该类型在内存中如何组织。
输入函数接受一个以空字符结尾的字符串作为参数，并返回该类型的内部（内存中）表示。
输出函数接受该类型的内部表示作为参数，并返回一个以空字符结尾的字符串。
如果我们希望该类型除了存储之外还能做别的事情，
就必须提供额外的函数来实现我们希望该类型支持的各种操作。

假设我们要定义一种表示复数的类型`complex`。在内存中表示复数的一种自然方式是下面这个 C 结构体：

```
typedef struct Complex {
    double      x;
    double      y;
} Complex;
```

我们需要把它做成按引用传递的类型，因为它太大了，无法放进单个 `Datum`值中。

作为该类型的外部字符串表示，我们选择形如`(x,y)`的字符串。

输入和输出函数通常不难编写，尤其是输出函数。但在定义该类型的外部字符串表示时，请记住，你最终必须为该表示编写一个完整而健壮的解析器作为输入函数。例如：

```
PG_FUNCTION_INFO_V1(complex_in);

Datum
complex_in(PG_FUNCTION_ARGS)
{
    char       *str = PG_GETARG_CSTRING(0);
    double      x,
                y;
    Complex    *result;

    if (sscanf(str, " ( %lf , %lf )", &x, &y) != 2)
        ereport(ERROR,
                (errcode(ERRCODE_INVALID_TEXT_REPRESENTATION),
                 errmsg("invalid input syntax for complex: \"%s\"",
                        str)));

    result = (Complex *) palloc(sizeof(Complex));
    result->x = x;
    result->y = y;
    PG_RETURN_POINTER(result);
}
```

输出函数可以简单地写成：

```
PG_FUNCTION_INFO_V1(complex_out);

Datum
complex_out(PG_FUNCTION_ARGS)
{
    Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
    char       *result;

    result = (char *) palloc(100);
    snprintf(result, 100, "(%g,%g)", complex->x, complex->y);
    PG_RETURN_CSTRING(result);
}
```

应当注意让输入函数和输出函数互为逆运算。如果不是这样，当你需要把数据转储到文件中再读回时，就会遇到严重问题。这在涉及浮点数时尤为常见。

可选地，用户定义类型还可以提供二进制输入和输出例程。二进制 I/O 通常比文本 I/O 更快，但可移植性较差。与文本 I/O 一样，外部二进制表示的精确定义完全由你决定。大多数内置数据类型都尽量提供与机器无关的二进制表示。对于`complex`，我们将借助类型`float8`的二进制 I/O 转换器：

```
PG_FUNCTION_INFO_V1(complex_recv);

Datum
complex_recv(PG_FUNCTION_ARGS)
{
```

```
    StringInfo  buf = (StringInfo) PG_GETARG_POINTER(0);
    Complex    *result;

    result = (Complex *) palloc(sizeof(Complex));
    result->x = pq_getmsgfloat8(buf);
    result->y = pq_getmsgfloat8(buf);
    PG_RETURN_POINTER(result);
}

PG_FUNCTION_INFO_V1(complex_send);

Datum
complex_send(PG_FUNCTION_ARGS)
{
    Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
    StringInfoData buf;

    pq_begintypsend(&buf);
    pq_sendfloat8(&buf, complex->x);
    pq_sendfloat8(&buf, complex->y);
    PG_RETURN_BYTEA_P(pq_endtypsend(&buf));
}
```

一旦我们写好了 I/O 函数并将它们编译进共享库，就可以在 SQL 中定义 `complex`类型。首先将其声明为一种 shell 类型：

```
CREATE TYPE complex;
```

这会建立一个占位符，使我们可以在定义其 I/O 函数时引用该类型。现在我们可以定义这些 I/O 函数：

```
CREATE FUNCTION complex_in(cstring)
    RETURNS complex
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_out(complex)
    RETURNS cstring
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_recv(internal)
   RETURNS complex
   AS 'filename'
   LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_send(complex)
   RETURNS bytea
   AS 'filename'
   LANGUAGE C IMMUTABLE STRICT;
```

最后，我们可以给出该数据类型的完整定义：

```
CREATE TYPE complex (
   internallength = 16,
```

```
    input = complex_in,
    output = complex_out,
    receive = complex_recv,
    send = complex_send,
    alignment = double
);
```

当你定义一种新的基础类型时，PostgreSQL会自动提供对该类型数组的支持。该数组类型通常与基础类型同名，只是在前面加一个下划线字符（_）。

一旦该数据类型存在，我们就可以声明额外的函数，为该数据类型提供有用的操作。随后可以在这些函数之上定义操作符；如果需要，还可以创建操作符类以支持该数据类型的索引。这些附加层会在后续各节中讨论。

如果你的数据类型的值在内部形式上的大小可变，就应该让该数据类型支持 TOAST（见第 55.2 节）。即使这些值总是小到不需要压缩或外部存储，也应该这样做，因为 TOAST 还可以通过减少头部 overhead.

为此，内部表示必须遵循可变长度数据的标准布局：前四个字节必须是一个从不直接访问的 `char[4]`字段（惯例上命名为 `vl_len_`）。必须使用 `SET_VARSIZE()`在该字段中存储该 datum 的大小，并使用 `VARSIZE()`取回它。操作该数据类型的 C 级函数必须始终使用`PG_DETOAST_DATUM`对传给它们的任何 TOAST 化值执行去 TOAST 化处理。（这一细节通常通过定义特定于该类型的 `GETARG_DATATYPE_P`宏来隐藏。）然后，在执行 `CREATE TYPE`命令时，把内部长度指定为 `variable`并选择合适的存储选项。

如果数据对齐并不重要（无论只是对某个特定函数而言，还是因为该数据类型本来就指定了字节对齐），那么就有可能避免 `PG_DETOAST_DATUM`的一部分开销。你可以改用 `PG_DETOAST_DATUM_PACKED`（通常通过定义 `GETARG_DATATYPE_PP`宏来隐藏），并使用 `VARSIZE_ANY_EXHDR`和`VARDATA_ANY` 宏访问一个可能采用打包形式的 datum。再次注意，即使数据类型定义指定了对齐方式，这些宏返回的数据也不是对齐的。如果对齐很重要，就必须使用常规的`PG_DETOAST_DATUM`接口。

> **注意**
>
> 较旧的代码常把`vl_len_`声明为 `int32`字段，而不是`char[4]`字段。只要结构体定义中还有其他至少按`int32`对齐的字段，这样做是可以的。但在处理可能未对齐的 datum 时使用这种结构体定义就很危险；编译器可能据此假定该 datum 实际上是对齐的，从而在对齐要求严格的体系结构上导致核心转储。

更多细节见CREATE TYPE命令的说明。

35.12. 用户定义的操作符

每个操作符都是对执行实际工作的底层函数调用的一种“语法糖”；因此，你必须先创建底层函数，才能创建操作符。不过，操作符并不仅仅是语法糖，因为它还携带一些额外信息，可帮助查询规划器优化使用该操作符的查询。下一节将专门解释这些附加信息。

PostgreSQL支持左一元、右一元和二元操作符。操作符可以被重载；也就是说，同一个操作符名可以用于不同的操作符，而这些操作符具有不同数量和类型的操作数。执行查询时，系统会根据提供的操作数数量和类型确定应调用哪个操作符。

下面是创建一个用于将两个复数相加的操作符的示例。假定我们已经创建了类型complex的定义（见第 35.11 节）。首先需要一个完成实际工作的函数，然后就可以定义操作符：

```
CREATE FUNCTION complex_add(complex, complex)
    RETURNS complex
    AS 'filename', 'complex_add'
    LANGUAGE C IMMUTABLE STRICT;

CREATE OPERATOR + (
    leftarg = complex,
    rightarg = complex,
    procedure = complex_add,
    commutator = +
);
```

现在我们就可以执行下面这样的查询：

```
SELECT (a + b) AS c FROM test_complex;

        c
-----------------
 (5.2,6.05)
 (133.42,144.95)
```

上面展示了如何创建一个二元操作符。要创建一元操作符，只需省略 leftarg（左一元操作符）或 rightarg（右一元操作符）之一。procedure 子句和参数子句是 CREATE OPERATOR 中唯一必需的项。示例中出现的 commutator 子句，则是给查询优化器的一个可选提示。关于 commutator 以及其他优化器提示的更多细节，会在下一节说明。

35.13. 操作符优化信息

PostgreSQL 的操作符定义可以包含若干可选子句，用来告诉系统该操作符行为的一些有用信息。只要适用，就应当提供这些子句，因为它们可以显著加快使用该操作符的查询执行速度。但如果你提供了这些信息，就必须确保它们是正确的！错误地使用优化子句可能导致查询变慢、输出出现隐蔽错误，或者引发其他糟糕后果。如果你拿不准，完全可以省略某个优化子句；唯一的后果，只是查询可能会比本来需要的更慢。

未来版本的 PostgreSQL 可能会加入更多优化子句。这里描述的是 9.1.24 版本所理解的全部子句。

35.13.1. COMMUTATOR

如果给出 COMMUTATOR 子句，它指定一个与正在定义的操作符互为交换子的操作符。若对所有可能的输入值 x、y，都有 (x A y) 等于 (y B x)，则称操作符 A 是操作符 B 的交换子。注意，B 也同样是 A 的交换子。例如，对某种特定数据类型而言，< 和 > 通常互为交换子，而操作符 + 通常与其自身可交换。但操作符 - 通常并不与任何操作符可交换。

一个可交换操作符的左操作数类型，与其交换子的右操作数类型相同，反之亦然。因此，PostgreSQL 只需知道交换子操作符的名称，就可以查找出该交换子，而这也正是 COMMUTATOR 子句中所需提供的全部内容。

对于将用于索引和连接子句的操作符，提供交换子信息至关重要，因为这样查询优化器才能把这类子句“翻转”成不同计划类型所需的形式。例如，

考虑这样一个查询，其 WHERE 子句形如 `tab1.x = tab2.y`，其中 `tab1.x` 和 `tab2.y` 都是某种用户定义类型，并假设 `tab2.y` 上建有索引。除非优化器能够确定如何把该子句翻转成 `tab2.y = tab1.x`，否则它就无法生成索引扫描，因为索引扫描机制要求传给它的操作符左侧必须是已建立索引的列。PostgreSQL 不会仅凭假设就认为这种转换有效；= 操作符的创建者必须通过为该操作符标记交换子信息，明确声明这种转换是有效的。

定义一个与自身可交换的操作符时，直接定义即可。但定义一对可交换操作符时，情况就稍微复杂一些：第一个要定义的操作符如何引用另一个尚未定义的操作符呢？这个问题有两种解决办法：

- 一种办法是在定义第一个操作符时省略 `COMMUTATOR` 子句，然后在第二个操作符的定义中提供该子句。由于 PostgreSQL 知道可交换操作符是成对出现的，因此它在看到第二个定义时，会自动回头补全第一个定义中缺少的 `COMMUTATOR` 子句。

- 另一种更直接的办法是在两个定义中都包含 `COMMUTATOR` 子句。当 PostgreSQL 处理第一个定义并发现 `COMMUTATOR` 引用了一个不存在的操作符时，系统会在系统目录中为该操作符建立一个占位项。这个占位项只有操作符名称、左右操作数类型和结果类型包含有效数据，因为这就是 PostgreSQL 此时能够推断出的全部信息。第一个操作符的目录项将链接到这个占位项。之后，当你定义第二个操作符时，系统会用第二个定义中的附加信息更新该占位项。如果在占位操作符补全之前尝试使用它，就只会得到一条错误消息。

35.13.2. NEGATOR

如果给出 `NEGATOR` 子句，它指定一个与正在定义的操作符互为求反器的操作符。若操作符 A 和 B 都返回布尔结果，并且对所有可能的输入 x、y，都有 (x A y) 等于 NOT (x B y)，则称 A 是 B 的求反器。注意，B 也同样是 A 的求反器。例如，对大多数数据类型而言，< 和 >= 构成一对求反器。一个操作符永远不可能合法地成为其自身的求反器。

与交换子不同，一对一元操作符完全可能合法地被标记为彼此的求反器；这意味着对所有 x，都有 (A x) 等于 NOT (B x)，或者对右一元操作符有等价的关系。

一个操作符的求反器必须与待定义操作符具有相同的左操作数类型和/或右操作数类型，因此与 `COMMUTATOR` 一样，在 `NEGATOR` 子句中只需给出操作符名称即可。

提供求反器对查询优化器很有帮助，因为它允许把 `NOT (x = y)` 这样的表达式简化为 `x <> y`。这种情况比你想象得更常见，因为 `NOT` 操作可能会作为其他重排的结果被插入进来。

可以使用上面解释的定义交换子对的相同方法，来定义成对的求反器操作符。

35.13.3. RESTRICT

如果给出 `RESTRICT` 子句，它指定该操作符的限制选择率估算函数。（注意，这里是函数名，而不是操作符名。）`RESTRICT` 子句只对返回 `boolean` 的二元操作符有意义。限制选择率估算器的作用，是针对当前操作符和某个特定常量值，猜测一张表中有多少比例的行会满足如下形式的 `WHERE` 子句条件：

```
column OP constant
```

这会帮助优化器大致了解具有这种形式的 `WHERE` 子句将淘汰多少行。（你可能会问：如果常量在左边会怎样？嗯，这正是 `COMMUTATOR` 的作用之一.....）

编写新的限制选择率估算函数远远超出了本章的范围，不过幸运的是，对于你自己的很多操作符，通常都可以直接使用系统提供的某个标准估算器。标准的限制选择率估算器如下：

eqsel 用于 =
neqsel 用于 <>
scalarltsel 用于 < 或 <=
scalargtsel 用于 > 或 >=

这些分类可能看起来有点奇怪，但仔细想想就会发现它们是有道理的。= 通常只会接受表中很小一部分行；<> 通常只会拒绝很小一部分行。< 接受的比例取决于给定常量落在该表列的值范围中的什么位置（恰好，这正是 ANALYZE 收集并提供给选择率估算器的信息）。对于相同的比较常量，<= 接受的比例会比 < 略大，但二者足够接近，不值得区分，尤其是无论如何我们很可能也只能做出粗略猜测。类似的说法也适用于 > 和 >=。

对于选择率非常高或非常低的操作符，即使它们实际上并不是真正的相等或不等比较，
你也常常可以勉强使用 eqsel 或 neqsel。例如，几何类型中的近似相等操作符就使用 eqsel，其依据是它们通常只会匹配表中很小一部分项。

对于那些能够以某种合理方式转换为数值标量、从而可进行范围比较的数据类型，你可以使用 scalarltsel 和 scalargtsel。如果可能的话，请把该数据类型加入函数 convert_to_scalar()（位于 src/backend/utils/adt/selfuncs.c）所理解的范围中。（最终，这个函数应被通过 pg_type 系统目录某一列标识的、按数据类型划分的函数所取代；但这件事目前还没有发生。）如果不这样做，系统仍然可以工作，
但优化器的估算效果就不会像本可达到的那样好。

在 src/backend/utils/adt/geo_selfuncs.c 中，还为几何操作符提供了其他选择率估算函数：areasel、positionsel 和 contsel。截至目前，这些函数都还只是桩实现，但你也许仍会想使用它们（或者更好的是，改进它们）。

35.13.4. JOIN

如果给出 JOIN 子句，它指定该操作符的连接选择率估算函数。（注意，这里是函数名，而不是操作符名。）JOIN 子句只对返回 boolean 的二元操作符有意义。连接选择率估算器的作用，是针对当前操作符，猜测两张表中有多少比例的行对会满足如下形式的 WHERE 子句条件：

```
table1.column1 OP table2.column2
```

与 RESTRICT 子句一样，这会极大地帮助优化器判断在若干可能的连接顺序中，哪一种预计工作量最小。

与前面一样，本章不会尝试解释如何编写连接选择率估算函数，
而只是建议你在适用时使用某个标准估算器：

eqjoinsel 用于 =
neqjoinsel 用于 <>
scalarltjoinsel 用于 < 或 <=
scalargtjoinsel 用于 > 或 >=
areajoinsel 用于基于二维面积的比较
positionjoinsel 用于基于二维位置的比较
contjoinsel 用于基于二维包含关系的比较

35.13.5. HASHES

如果存在 HASHES 子句，它会告知系统：在基于该操作符的连接中，可以使用哈希连接方法。HASHES 只对返回 boolean 的二元操作符有意义；在实践中，该操作符还必须表示某种数据类型或某对数据类型上的相等关系。

哈希连接背后的假设是：只有当左值和右值被哈希到同一个哈希码时，连接操作符才有可能返回 true。如果两个值被放进不同的哈希桶，连接过程就根本不会去比较它们，这实际上隐含假定连接操作符的结果必定为 false。因此，对于那些并不表示某种相等关系的操作符，指定 `HASHES` 永远没有意义。在大多数情况下，只有对两侧都接受同一数据类型的操作符支持哈希才是现实可行的。不过，有时也可以为两种或更多数据类型设计兼容的哈希函数；也就是说，尽管这些值的表示不同，但对于“相等”的值，函数仍会生成相同的哈希码。例如，对不同位宽的整数做到这一点就相当简单。

若要被标记为 `HASHES`，该连接操作符必须出现在某个哈希索引操作符族中。创建操作符时并不会强制这一点，因为那时引用它的操作符族当然还不存在。但如果没有这样的操作符族存在，运行时尝试在哈希连接中使用该操作符就会失败。系统需要通过该操作符族来查找与该操作符输入数据类型相对应的哈希函数。当然，在创建该操作符族之前，你还必须先创建合适的哈希函数。

准备哈希函数时应格外小心，因为它有一些依赖机器的方式可能会导致结果不正确。例如，如果你的数据类型是某种结构体，其中可能包含无意义的填充位，那就不能简单地把整个结构体传给 `hash_any`。（除非你编写了其他操作符和函数，以确保这些未使用位始终为零，而这正是推荐的策略。）再比如，在符合 IEEE 浮点标准的机器上，负零和正零是不同的值（位模式不同），但它们被定义为比较相等。如果某个浮点值可能包含负零，就需要采取额外步骤，确保它生成与正零相同的哈希值。

一个可参与哈希连接的操作符，必须拥有一个交换子（如果两侧操作数数据类型相同，则为它自身；如果不同，则应是相关的相等操作符），并且该交换子也必须出现在同一个操作符族中。否则，在使用该操作符时可能会发生规划器错误。另外，对于支持多种数据类型的哈希操作符族，最好（虽然并非严格必需）为每一种数据类型组合都提供相等操作符；这样可以获得更好的优化效果。

> **注意**
>
> 作为可哈希连接操作符底层实现的函数，必须标记为 immutable 或 stable。如果它是 volatile，系统就永远不会尝试把该操作符用于哈希连接。

> **注意**
>
> 如果某个可哈希连接操作符的底层函数被标记为 strict，那么该函数还必须是完备的：也就是说，对于任意两个非空输入，它都应返回 true 或 false，而绝不能返回 null。如果不遵循这条规则，对 `IN` 操作进行哈希优化时可能会产生错误结果。（具体来说，`IN` 可能会在按标准本应返回 null 的地方返回 false；或者抛出一条错误，抱怨它没有为 null 结果做好准备。）

35.13.6. MERGES

如果存在 `MERGES` 子句，它会告知系统：在基于该操作符的连接中，可以使用归并连接方法。`MERGES` 只对返回 `boolean` 的二元操作符有意义；在实践中，该操作符还必须表示某种数据类型或某对数据类型上的相等关系。

归并连接的基本思想，是先把左表和右表分别排序，然后同步扫描它们。因此，两种数据类型都必须能够被完全排序，

而连接操作符必须只能在那对值位于排序次序中“同一位置”时才成功。实际效果上，这意味着连接操作符必须表现得像相等比较一样。不过，只要两种不同的数据类型在逻辑上兼容，也完全可以对它们进行归并连接。例如，smallint 与 integer 之间的相等操作符就是可归并连接的。我们只需要能把两种数据类型都带入逻辑兼容排序序列的排序操作符即可。

若要被标记为 MERGES，该连接操作符必须作为相等成员出现在某个 btree 索引操作符族中。创建操作符时并不会强制这一点，因为那时引用它的操作符族当然还不存在。但除非能找到匹配的操作符族，否则该操作符实际上不会被用于归并连接。因此，MERGES 标记的作用，是向规划器提供一个提示，告诉它值得去查找匹配的操作符族。

一个可参与归并连接的操作符，必须拥有一个交换子（如果两侧操作数数据类型相同，则为它自身；如果不同，则应是相关的相等操作符），并且该交换子也必须出现在同一个操作符族中。否则，在使用该操作符时可能会发生规划器错误。另外，对于支持多种数据类型的 btree 操作符族，最好（虽然并非严格必需）为每一种数据类型组合都提供相等操作符；这样可以获得更好的优化效果。

> **注意**
>
> 作为可归并连接操作符底层实现的函数，必须标记为 immutable 或 stable。如果它是 volatile，系统就永远不会尝试把该操作符用于归并连接。

35.14. 索引扩展接口

到目前为止所描述的过程使我们能够定义新类型、新函数以及新操作符。然而，我们还不能在一种新数据类型的列上定义索引。为此，必须为该新数据类型定义一个操作符类。本节稍后将用一个示例说明这一概念：为 B-树索引方法定义一个新的操作符类，以便按绝对值升序存储和排序复数。

操作符类可以分组成操作符族，以展示语义兼容的类之间的关系。只涉及单一数据类型时，一个操作符类就已足够，因此我们先关注这种情况，然后再回到操作符族。

35.14.1. 索引方法和操作符类

pg_am 表为每个索引方法（内部称为访问方法）保存一行。对表进行常规访问的支持内置于 PostgreSQL 中，但所有索引方法都在 pg_am 中描述。可以定义所需的接口例程，然后在 pg_am 中创建一行，从而添加新的索引方法 – 但这超出了本章的范围（参见第 52 章）。

索引方法的例程并不直接知道它将处理哪些数据类型。相反，一个操作符类标识了索引方法在处理特定数据类型时需要使用的那组操作。之所以称为操作符类，是因为它指定的一项内容就是可与索引一起使用的 WHERE 子句操作符集合（也就是能被转换成索引扫描条件的操作符）。操作符类还可以指定索引方法内部操作所需的一些支持函数，但这些函数并不直接对应任何可与索引一起使用的 WHERE 子句操作符。

可以为同一种数据类型和索引方法定义多个操作符类。这样就能为一种数据类型定义多套索引语义。例如，一个 B-树索引要求为其处理的每一种数据类型定义一种排序顺序。对于复数数据类型，也许既需要一个按复数绝对值排序的 B-树操作符类，也需要另一个按实部排序的操作符类，等等。通常，其中一个操作符类会被视为最常用，并标记为该数据类型在该索引方法上的默认操作符类。

同一个操作符类名可以用于多个不同的索引方法（例如，B-树和哈希索引方法都有名为 `int4_ops` 的操作符类），但每一个这样的类都是独立实体，必须分别定义。

35.14.2. 索引方法策略

与操作符类关联的操作符通过“策略号”来标识，用以表示每个操作符在其操作符类上下文中的语义。例如，B-树对键施加了严格的从小到大的顺序，因此像“小于”和“大于等于”这样的操作符，对 B-树来说就很重要。由于 PostgreSQL 允许用户定义操作符，PostgreSQL 不能仅凭操作符名称（例如 `<` 或 `>=`）就判断它是哪一类比较。取而代之的是，索引方法定义了一组“策略”，可以把它们看成是广义的操作符。每个操作符类都会说明，对于某种特定数据类型和某种索引语义解释，每一种策略分别对应哪个实际操作符。

B-树索引方法定义了五种策略，如表 35.2所示。

表 35.2. B-树策略

Operation	Strategy Number
less than	1
less than or equal	2
equal	3
greater than or equal	4
greater than	5

哈希索引只支持等值比较，因此它们只使用一种策略，如表 35.3所示。

表 35.3. 哈希策略

Operation	Strategy Number
equal	1

GiST 索引更加灵活：它们根本没有固定的策略集合。相反，每个特定 GiST 操作符类中负责“一致性”检查的支持例程会按自己的方式解释策略号。举例来说，一些内置的 GiST 索引操作符类会为二维几何对象建立索引，并提供“R 树”策略，如表 35.4所示。其中四个是真正的二维测试（重叠、相同、包含、被包含），四个只考虑 X 方向，另外四个则在 Y 方向上提供相同测试。

表 35.4. GiST 二维“R 树”策略

Operation	Strategy Number
strictly left of	1
does not extend to right of	2
overlaps	3
does not extend to left of	4
strictly right of	5
same	6
contains	7
contained by	8
does not extend above	9
strictly below	10

Operation	Strategy Number
strictly above	11
does not extend below	12

GIN 索引在灵活性方面与 GiST 索引类似：它们没有固定的策略集合。相反，每个操作符类的支持例程会根据该操作符类的定义解释策略号。举例来说，内置数组操作符类所使用的策略号如表 35.5所示。

表 35.5. GIN 数组策略

Operation	Strategy Number
overlap	1
contains	2
is contained by	3
equal	4

注意，上面列出的所有操作符都返回布尔值。实际上，
所有被定义为索引方法搜索操作符的操作符都必须返回 `boolean`，因为要与索引配合使用，它们必须出现在 `WHERE` 子句的顶层。（某些索引访问方法还支持排序操作符，这类操作符通常不返回布尔值；该特性见第 35.14.7 节。）

35.14.3. 索引方法支持例程

仅靠策略信息通常不足以让系统知道如何使用索引。实际上，
索引方法还需要额外的支持例程才能工作。例如，B-树索引方法必须能够比较两个键，
并判断其中一个是大于、等于还是小于另一个。类似地，
哈希索引方法必须能够为键值计算哈希码。这些操作并不对应 SQL 命令条件中使用的操作符；它们是索引方法内部使用的管理例程。

与策略一样，操作符类会标识对于给定的数据类型和语义解释，
应由哪些具体函数承担这些角色。索引方法定义它需要的函数集合，
而操作符类则会通过为函数分配由索引方法规定的“支持函数号”来标识正确的函数。

B-tree 要求一个支持函数，如表 35.6所示。

表 35.6. B-树支持函数

Function	Support Number
Compare two keys and return an integer less than zero, zero, or greater than zero, indicating whether the first key is less than, equal to, or greater than the second	1

哈希索引同样只需要一个支持函数，如表 35.7所示。

表 35.7. 哈希支持函数

Function	Support Number
Compute the hash value for a key	1

GiST 索引需要七个支持函数，外加一个可选的第八个，如表 35.8所示。

表 35.8. GiST 支持函数

Function	Description	Support Number
`consistent`	determine whether key satisfies the query qualifier	1
`union`	compute union of a set of keys	2
`compress`	compute a compressed repres entation of a key or value to be indexed	3
`decompress`	compute a decompressed representation of a compre ssed key	4
`penalty`	compute penalty for inserting new key into subtree with given subtree's key	5
`picksplit`	determine which entries of a page are to be moved to the new page and compute the union keys for resulting pages	6
`equal`	compare two keys and return true if they are equal	7
`distance`	(optional method) determine distance from key to query value	8

GIN 索引需要四个支持函数，外加一个可选的第五个，如表 35.9所示。

表 35.9. GIN 支持函数

Function	Description	Support Number
`compare`	compare two keys and return an integer less than zero, zero, or greater than zero, indica ting whether the first key is less than, equal to, or greater than the second	1
`extractValue`	extract keys from a value to be indexed	2
`extractQuery`	extract keys from a query condition	3
`consistent`	determine whether value matches query condition	4
`comparePartial`	(optional method) compar e partial key from query and key from index, and return an integer less than zero, zero, or greater than zero, indicating whether GIN should ignore this index entry, treat the entry as a match, or stop the index scan	5

与搜索操作符不同，支持函数返回特定索引方法所期望的数据类型; for example in the case of the comparison function for B-trees, a signed integer. The number and types of the

arguments to each support function are likewise dependent on the index method. For B-树和哈希索引，支持函数接受与操作符类所含操作符相同的输入数据类型 in the operator class, but this is not the case for most GIN and GiST support functions.

35.14.4. 示例

现在我们已经了解了这些基本思想，下面给出先前承诺的创建新操作符类示例。（这个可运行示例位于源码发布包中的`src/tutorial/complex.c`和`src/tutorial/complex.sql`。）该操作符类封装了一组按绝对值顺序对复数排序的操作符，因此我们把它命名为`complex_abs_ops`。首先，我们需要一组操作符。定义操作符的过程已经在第 35.12 节中讨论过。对于 B-树上的操作符类，我们需要如下操作符：

- 绝对值小于（策略 1）
- 绝对值小于等于（策略 2）
- 绝对值等于（策略 3）
- 绝对值大于等于（策略 4）
- 绝对值大于（策略 5）

定义一组相关比较操作符时，最不容易出错的方式是先编写 B-树比较支持函数，再把其他函数写成围绕该支持函数的一行包装器函数。这样可以降低在边界情况下得到不一致结果的概率。按照这种方法，我们首先编写：

```
#define Mag(c)  ((c)->x*(c)->x + (c)->y*(c)->y)

static int
complex_abs_cmp_internal(Complex *a, Complex *b)
{
    double      amag = Mag(a),
                bmag = Mag(b);

    if (amag < bmag)
        return -1;
    if (amag > bmag)
        return 1;
    return 0;
}
```

现在，小于函数如下所示：

```
PG_FUNCTION_INFO_V1(complex_abs_lt);

Datum
complex_abs_lt(PG_FUNCTION_ARGS)
{
    Complex    *a = (Complex *) PG_GETARG_POINTER(0);
    Complex    *b = (Complex *) PG_GETARG_POINTER(1);

    PG_RETURN_BOOL(complex_abs_cmp_internal(a, b) < 0);
}
```

其他四个函数的区别只在于它们如何比较内部函数的结果与 0。

接下来，在 SQL 中声明这些函数，以及基于这些函数的操作符：

```
CREATE FUNCTION complex_abs_lt(complex, complex) RETURNS bool
```

```
    AS 'filename', 'complex_abs_lt'
    LANGUAGE C IMMUTABLE STRICT;

CREATE OPERATOR < (
   leftarg = complex, rightarg = complex, procedure = complex_abs_lt,
   commutator = > , negator = >= ,
   restrict = scalarltsel, join = scalarltjoinsel
);
```

必须指定正确的交换子和求反器操作符，以及合适的限制选择率与连接选择率函数，否则优化器无法有效使用索引。注意，小于、等于和大于这几种情况应使用不同的选择率函数。

这里还有几点值得注意：

- 只能有一个名为 = 且两个操作数都为 `complex` 类型的操作符。在这个例子里，我们并没有任何其他 = 操作符可用于 `complex`；但如果我们是在构造一种实际使用的数据类型，可能会希望 = 表示复数的普通相等，而不是绝对值相等。在那种情况下，我们就需要为 `complex_abs_eq` 选用其他操作符名。
- 尽管 PostgreSQL 能处理 SQL 名称相同但参数数据类型不同的函数，C 却只能处理给定名称的一个全局函数。因此，我们不应该把 C 函数简单命名成 `abs_eq` 之类。通常，在 C 函数名中包含数据类型名称是个好习惯，这样就不会与其他数据类型的函数发生冲突。
- 我们原本也可以把该函数的 SQL 名称取为 `abs_eq`，并依靠 PostgreSQL 通过参数数据类型把它与其他同名 SQL 函数区分开。为了让示例保持简单，这里我们让 C 层和 SQL 层的函数使用相同的名称。

下一步是注册 B-树要求的支持例程。实现该例程的 C 示例代码与操作符函数位于同一个文件中。该函数的声明如下：

```
CREATE FUNCTION complex_abs_cmp(complex, complex)
    RETURNS integer
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;
```

现在我们已经有了所需的操作符和支持例程，就可以最终创建操作符类：

```
CREATE OPERATOR CLASS complex_abs_ops
    DEFAULT FOR TYPE complex USING btree AS
        OPERATOR        1       < ,
        OPERATOR        2       <= ,
        OPERATOR        3       = ,
        OPERATOR        4       >= ,
        OPERATOR        5       > ,
        FUNCTION        1       complex_abs_cmp(complex, complex);
```

这样就完成了！现在应该可以在 `complex` 列上创建并使用 B-树索引。

我们本来也可以把操作符项写得更详细一些，例如：

```
        OPERATOR        1       < (complex, complex) ,
```

但是当操作符接受的数据类型与该操作符类所服务的数据类型相同时，就没有必要这样写。

上述示例假定你希望把这个新操作符类设为 `complex` 数据类型的默认 B-树操作符类。如果不是这样，只需省去 `DEFAULT` 这个词。

35.14.5. 操作符类和操作符族

到目前为止，我们一直隐含地假定一个操作符类只处理一种数据类型。
虽然某个特定的索引列当然只能有一种数据类型，
但对把被索引列与另一种数据类型的值进行比较的操作建立索引往往也很有用。此外，
如果某个与操作符类相关的跨数据类型操作符有用，
通常另一种数据类型本身也会有一个相关的操作符类。
把相关类之间的联系显式表示出来会很有帮助，因为这有助于规划器优化 SQL 查询（尤其是对 B-树操作符类而言，因为规划器中包含大量有关如何使用它们的知识）。

为了满足这些需求，PostgreSQL使用操作符族这一概念。
一个操作符族包含一个或多个操作符类，还可以包含属于整个族、
但不属于族中任何单一类的可索引操作符及其相应的支持函数。
我们称这样的操作符和函数在该族中是“松散”的，而不是绑定在某个特定类中。通常，
每个操作符类只包含单一数据类型的操作符，
而跨数据类型操作符则作为操作符族中的松散成员存在。

一个操作符族中的所有操作符和函数都必须具有兼容的语义，而兼容性的要求由索引方法设定。
因此，你也许会疑惑，为什么还要把该族的某些子集单独划成操作符类；事实上，
对很多用途而言，类的划分并不重要，真正有意义的分组只有操作符族。之所以定义操作符类，
是因为它们规定了支持特定索引所需的操作符族内容。如果某个索引使用了某个操作符类，
那么在不删除该索引的情况下就不能删除该操作符类 — 但操作符族中的其他部分，也就是其他操作符类和松散操作符，则可以被删除。因此，
一个操作符类应当只包含在特定数据类型上支持索引所合理需要的最小操作符和函数集合，
而那些相关但非必需的操作符，则可以作为操作符族的松散成员加入。

As an example, PostgreSQL has a built-in B-tree operator family `integer_ops`, which includes operator classes `int8_ops`, `int4_ops`, and `int2_ops` for indexes on `bigint` (`int8`), `integer` (`int4`), and `smallint` (`int2`) 列上的索引。该族还包含跨数据类型比较操作符，允许这些类型中的任意两个相互比较，这样其中一种类型上的索引就可以用另一种类型的比较值来搜索。该族可以用这些定义复制出来：

```
CREATE OPERATOR FAMILY integer_ops USING btree;

CREATE OPERATOR CLASS int8_ops
DEFAULT FOR TYPE int8 USING btree FAMILY integer_ops AS
  -- standard int8 comparisons
  OPERATOR 1 < ,
  OPERATOR 2 <= ,
  OPERATOR 3 = ,
  OPERATOR 4 >= ,
  OPERATOR 5 > ,
  FUNCTION 1 btint8cmp(int8, int8) ;

CREATE OPERATOR CLASS int4_ops
DEFAULT FOR TYPE int4 USING btree FAMILY integer_ops AS
  -- standard int4 comparisons
  OPERATOR 1 < ,
  OPERATOR 2 <= ,
  OPERATOR 3 = ,
  OPERATOR 4 >= ,
  OPERATOR 5 > ,
  FUNCTION 1 btint4cmp(int4, int4) ;
```

```
CREATE OPERATOR CLASS int2_ops
DEFAULT FOR TYPE int2 USING btree FAMILY integer_ops AS
  -- standard int2 comparisons
  OPERATOR 1 < ,
  OPERATOR 2 <= ,
  OPERATOR 3 = ,
  OPERATOR 4 >= ,
  OPERATOR 5 > ,
  FUNCTION 1 btint2cmp(int2, int2) ;

ALTER OPERATOR FAMILY integer_ops USING btree ADD
  -- cross-type comparisons int8 vs int2
  OPERATOR 1 < (int8, int2) ,
  OPERATOR 2 <= (int8, int2) ,
  OPERATOR 3 = (int8, int2) ,
  OPERATOR 4 >= (int8, int2) ,
  OPERATOR 5 > (int8, int2) ,
  FUNCTION 1 btint82cmp(int8, int2) ,

  -- cross-type comparisons int8 vs int4
  OPERATOR 1 < (int8, int4) ,
  OPERATOR 2 <= (int8, int4) ,
  OPERATOR 3 = (int8, int4) ,
  OPERATOR 4 >= (int8, int4) ,
  OPERATOR 5 > (int8, int4) ,
  FUNCTION 1 btint84cmp(int8, int4) ,

  -- cross-type comparisons int4 vs int2
  OPERATOR 1 < (int4, int2) ,
  OPERATOR 2 <= (int4, int2) ,
  OPERATOR 3 = (int4, int2) ,
  OPERATOR 4 >= (int4, int2) ,
  OPERATOR 5 > (int4, int2) ,
  FUNCTION 1 btint42cmp(int4, int2) ,

  -- cross-type comparisons int4 vs int8
  OPERATOR 1 < (int4, int8) ,
  OPERATOR 2 <= (int4, int8) ,
  OPERATOR 3 = (int4, int8) ,
  OPERATOR 4 >= (int4, int8) ,
  OPERATOR 5 > (int4, int8) ,
  FUNCTION 1 btint48cmp(int4, int8) ,

  -- cross-type comparisons int2 vs int8
  OPERATOR 1 < (int2, int8) ,
  OPERATOR 2 <= (int2, int8) ,
  OPERATOR 3 = (int2, int8) ,
  OPERATOR 4 >= (int2, int8) ,
  OPERATOR 5 > (int2, int8) ,
  FUNCTION 1 btint28cmp(int2, int8) ,

  -- cross-type comparisons int2 vs int4
  OPERATOR 1 < (int2, int4) ,
  OPERATOR 2 <= (int2, int4) ,
```

```
OPERATOR 3 = (int2, int4) ,
OPERATOR 4 >= (int2, int4) ,
OPERATOR 5 > (int2, int4) ,
FUNCTION 1 btint24cmp(int2, int4) ;
```

注意，此定义“重载”了操作符策略号和支持函数号：每个编号在操作符族中出现多次。只要同一编号的每个实例具有不同的输入数据类型，这就是允许的。两个输入类型都等于某个操作符类输入类型的实例，是该操作符类的主要操作符和支持函数，通常应声明为操作符类的一部分，而不是操作符族的松散成员。

在一个 B-tree 操作符族中，该族中的所有操作符都必须以兼容的方式排序，也就是说，传递律必须适用于该族支持的所有数据类型：“如果 A = B 且 B = C，则 A = C”，以及“如果 A < B 且 B < C，则 A < C”。对族中的每一个操作符，都必须有一个具有相同两个输入数据类型的支持函数。建议让操作符族保持完整，也就是说，对每一种数据类型组合都应包含全部操作符。每个操作符类只应包含其数据类型对应的非跨数据类型操作符和支持函数。

要构建一个多数据类型的哈希操作符族，必须为该族支持的每一种数据类型创建相互兼容的哈希支持函数。这里的兼容性是指：对任意两个被该族中的等值操作符视为相等的值，这些函数都保证返回相同的哈希码，即使这两个值属于不同类型也是如此。当这些类型具有不同的物理表示时，这通常难以实现，但在某些情况下可以做到。注意，每种数据类型只有一个支持函数，而不是每个等值操作符一个。建议让操作符族保持完整，也就是说，对每一种数据类型组合都提供一个等值操作符。每个操作符类只应包含其数据类型对应的非跨数据类型等值操作符和支持函数。

GIN 和 GiST 索引没有任何显式的跨数据类型操作概念。它们所支持的操作符集合，就是给定操作符类的主要支持函数所能处理的那些操作符。

> **注意**
>
> 在 PostgreSQL 8.3 之前，并没有操作符族这一概念，因此任何打算与索引一起使用的跨数据类型操作符都必须直接绑定到该索引的操作符类中。虽然这种做法仍然有效，但已被弃用，因为它会使索引的依赖关系过于宽泛，而且当两种数据类型都在同一操作符族中拥有操作符时，规划器能更有效地处理跨数据类型比较。

35.14.6. 系统对操作符类的依赖

PostgreSQL利用操作符类来从多方面推断操作符的属性，而不仅仅是判断它们能否用于索引。因此，即便你并不打算为自己的数据类型列建立索引，也可能会想创建操作符类。

特别地，`ORDER BY`和`DISTINCT`等 SQL 特性要求对值的比较和排序。为了在用户定义的数据类型上实现这些特性，PostgreSQL会为数据类型查找默认 B-树操作符类。这个操作符类的“相等”成员定义了用于`GROUP BY`和`DISTINCT`的值的等值概念，而该操作符类施加的排序顺序定义了默认的`ORDER BY`顺序。

用户定义类型的数组比较也依赖于该类型默认 B-树操作符类所定义的语义。

如果一种数据类型没有默认的 B-树操作符类，系统就会查找默认的哈希操作符类。但由于这类操作符类只提供等值语义，因此在实践中它只足以支持数组相等比较。

如果某种数据类型没有默认操作符类，而你又试图将这些 SQL 特性用于该数据类型，就会得到类似“无法识别排序操作符”这样的错误。

> **注意**
>
> 在版本 7.4 以前的PostgreSQL中，排序和分组操作将隐式地使用名为=、<以及>的操作符。新的依赖于默认操作符类的行为避免了对具有特定名字的操作符行为作出任何假设。

另一个重要点在于，出现在哈希操作符族中的操作符，都是哈希连接、哈希聚合以及相关优化的候选对象。这里哈希操作符族至关重要，因为它标识了应当使用的哈希函数。

35.14.7. 排序操作符

有些索引访问方法（目前只有 GiST）支持排序操作符的概念。我们到目前为止讨论的是搜索操作符。对于搜索操作符，可以搜索索引以找出满足以下条件的所有行：`WHERE indexed_column operator constant`。注意，不保证匹配行的返回顺序。排序操作符则不限制可以返回的行集合，而是确定这些行的顺序。对于排序操作符，可以扫描索引，按以下表达式表示的顺序返回行：`ORDER BY indexed_column operator constant`。这样定义排序操作符，是因为当操作符用于度量距离时，它可以支持最近邻搜索。例如，如下查询：

```
SELECT * FROM places ORDER BY location <-> point '(101,456)' LIMIT 10;
```

可以找到距离指定目标点最近的十个地点。location 列上的 GiST 索引能够高效完成此操作，因为 `<->`是排序操作符。

搜索操作符必须返回布尔结果，而排序操作符通常返回其他类型的结果，例如用于表示距离的 float 或 numeric。这种类型通常不同于被索引的数据类型。为了避免对不同数据类型行为作硬编码假设，在定义排序操作符时，必须指定一个 B-树操作符族，用来说明结果数据类型的排序顺序。正如上一节所述，B-树操作符族定义了 PostgreSQL 的顺序概念，因此这是一种自然的表示方式。由于点的 `<->` 操作符返回 `float8`，因此可以在创建操作符类时这样指定它：

```
OPERATOR 15    <-> (point, point) FOR ORDER BY float_ops
```

其中 `float_ops` 是包含针对 `float8` 的操作的内置操作符族。这个声明表明，该索引能够按 `<->` 操作符值递增的顺序返回行。

35.14.8. 操作符类的特殊特性

还有两个操作符类的特殊特性我们尚未讨论，主要是因为它们对最常用的索引方法没有用处。

通常，把一个操作符声明为操作符类（或操作符族）的成员，意味着索引方法能够正好检索出满足使用该操作符的 `WHERE` 条件的行集。例如：

```
SELECT * FROM table WHERE integer_column < 4;
```

可以由整数列上的 B-tree 索引精确满足。但也存在这样的情况：索引只是查找匹配行的不精确向导。例如，如果一个 GiST 索引只为几何对象存储包围盒，那么它就无法精确满足测试多边形等非矩形对象之间重叠的 `WHERE` 条件。不过，我们可以用索引找出包围盒与目标对象的包围盒重叠的对象，然后只对索引找到的对象做精确的重叠测试。如果适用这种情形，就称该索引对该操作符是“有损”的。有损索引搜索的实现方式是：当某行可能真正满足、也可能不满足查询条件时，由索引方法返回一个 recheck（重新检查）标志。随后，核心系统会在检索到的行上测试原始查询条件，看它是否应作为有效匹配返回。只要索引保证返回所有需要的行（可能还外加一些额外的行，它们可以通过执行原始操作符调用来排除），这种方法就是可行的。

支持有损搜索的索引方法（目前是 GiST 和 GIN）允许各个操作符类的支持函数设置 recheck 标志，因此这实质上是操作符类层面的特性。

再次考虑只在索引中存储多边形等复杂对象的包围盒的情况。此时，
在索引条目中存储整个多边形没有多少价值，不如只存储一个更简单的对象，其类型为box。
这种情况由STORAGE选项表达，该选项位于CREATE OPERATOR CLASS：可以写成如下形式：

```
CREATE OPERATOR CLASS polygon_ops
    DEFAULT FOR TYPE polygon USING gist AS
        ...
        STORAGE box;
```

目前，只有 GiST 和 GIN 索引方法支持与列数据类型不同的STORAGE类型。GiST 的compress和decompress支持函数在使用STORAGE时必须处理数据类型转换。在 GIN 中，STORAGE类型标识“键”值的类型，通常与被索引列的类型不同。例如，
整数数组列的操作符类可以只使用整数作为键。GIN 的extractValue和extractQuery支持函数负责从被索引值中提取键。

35.15. 将相关对象打包成扩展

一个有用的 PostgreSQL 扩展通常包含多个 SQL 对象；例如，一种新的数据类型将需要新的函数、新的操作符，并且很可能还需要新的索引操作符类。
把所有这些对象收集到一个单独的包中，有助于简化数据库管理。PostgreSQL 将这样的包称为扩展。要定义一个扩展，至少需要一个脚本文件，其中包含用于创建扩展对象的 SQL 命令，以及一个控制文件，用来指定扩展本身的一些基本属性。如果扩展包含 C 代码，通常还会有一个共享库文件，C 代码已被构建到其中。一旦准备好这些文件，只需执行一个简单的CREATE EXTENSION命令，就能把这些对象装入数据库。

使用扩展而不是仅仅运行 SQL 脚本把一堆“松散”对象装入数据库，最主要的优点在于 PostgreSQL 能够理解这些对象是属于同一个扩展的。你可以用一条DROP EXTENSION命令删除全部对象（无需维护单独的“卸载”脚本）。更有用的是，pg_dump 知道自己不应转储扩展的各个成员对象 – 它只会在转储中包含一条 CREATE EXTENSION 命令。这极大地简化了迁移到扩展新版本的过程，即使新版本包含比旧版本更多或不同的对象也是如此。不过要注意，在把这样的转储装载到新数据库时，必须能够访问该扩展的控制文件、
脚本文件以及其他相关文件。

PostgreSQL 不允许你删除扩展中包含的单个对象，除非删除整个扩展。另外，虽然你可以修改扩展成员对象的定义（例如对函数使用 CREATE OR REPLACE FUNCTION），但要记住，被修改后的定义不会被 pg_dump 转储。这样的修改通常只有在你同时在扩展脚本文件中做出同样修改时才有意义。（不过，对于包含配置数据的表有特殊规定；见见下文。）

扩展机制还提供了用于打包修改脚本的支持，以便调整扩展中所含 SQL 对象的定义。例如，如果扩展 1.1 版相比 1.0 版增加了一个函数，并修改了另一个函数的函数体，那么扩展作者可以提供一个更新脚本来只完成这两项修改。随后就可以使用 ALTER EXTENSION UPDATE 命令应用这些修改，并跟踪在某个数据库中实际安装的是该扩展的哪个版本。

哪些 SQL 对象种类可以成为扩展成员，见ALTER EXTENSION的说明。特别是，数据库集簇范围内的对象，如数据库、角色和表空间，不能成为扩展成员，因为扩展只在单个数据库内可见。
（尽管并不禁止扩展脚本创建这类对象，但如果这样做，
它们不会作为扩展的一部分受到跟踪。）还要注意，虽然表可以成为扩展成员，
但其附属对象（如索引）并不直接被视为扩展成员。另一个重要点是，模式可以属于扩展，
但反过来不成立：扩展本身只有一个非限定名，并不“位于”任何模式中。不过，
扩展的成员对象会在其对象类型适用的情况下属于某个模式。
扩展是否应当拥有其成员对象所在的模式，则要视具体情况而定。

35.15.1. 扩展文件

CREATE EXTENSION命令依赖于每个扩展都有一个控制文件，其名称必须是扩展名加上 `.control` 后缀，并放在安装目录的 `SHAREDIR/extension` 目录中。还必须至少有一个 SQL 脚本文件，其名称遵循 *`extension--version.sql`* 的模式（例如，扩展 `foo` 的 `1.0` 版本使用 `foo--1.0.sql`）。默认情况下，脚本文件也放在 `SHAREDIR/extension` 目录中，但控制文件可以为脚本文件指定其他目录。

扩展控制文件的格式与 `postgresql.conf` 文件相同，即由一组 *`parameter_name = value`* 赋值组成，每行一条。允许空行和以 `#` 引入的注释。任何不是单个单词或数字的值都要记得加引号。

控制文件可以设置下列参数：

`directory` (`string`)

包含扩展 SQL 脚本文件的目录。除非给出的是绝对路径，否则该名称相对于安装的 `SHAREDIR` 目录。默认行为等价于指定 `directory = 'extension'`。

`default_version` (`string`)

扩展的默认版本（即在 `CREATE EXTENSION` 中未指定版本时将安装的版本）。虽然这个参数可以省略，但那样一来，如果没有给出 `VERSION` 选项，`CREATE EXTENSION` 就会失败，因此一般不希望这样做。

`comment` (`string`)

关于扩展的注释（任意字符串）。另外，也可以通过脚本文件中的COMMENT命令来设置该注释。

`encoding` (`string`)

脚本文件所使用的字符集编码。如果脚本文件包含任何非 ASCII 字符，就应指定这个参数。否则将假定这些文件使用数据库编码。

`module_pathname` (`string`)

该参数的值会替换脚本文件中每次出现的 `MODULE_PATHNAME`。如果未设置该参数，则不会进行替换。通常会把它设置为 `$libdir/`*`shared_library_name`*，然后在 C 语言函数的 `CREATE FUNCTION` 命令中使用 `MODULE_PATHNAME`，这样脚本文件就无需把共享库的名字硬编码进去。

`requires` (`string`)

本扩展所依赖的其他扩展名称列表，例如 `requires = 'foo, bar'`。这些被依赖的扩展必须先安装好，本扩展才能安装。

`superuser` (`boolean`)

如果此参数为 `true`（默认值），则只有超级用户可以创建扩展或将其更新到新版本。如果设为 `false`，则只需具有执行安装或更新脚本中命令所需的权限。

`relocatable` (`boolean`)

如果一个扩展在初次创建之后仍然可以把其包含的对象移动到不同的模式中，那么它就是可重定位的。默认值为 `false`，也就是该扩展不可重定位。详见下文。

schema (string)

该参数只能为不可重定位扩展设置。它强制扩展装载到指定名称的模式中，而不能装载到其他模式。详见下文。

除主控制文件 *extension*.control 外，扩展还可以有按如下样式命名的次级控制文件：*extension*--*version*.control。如果提供了这些文件，它们必须位于脚本文件目录中。次级控制文件遵循与主控制文件相同的格式。在安装或更新到该扩展的相应版本时，次级控制文件中设置的任何参数都会覆盖主控制文件中的设置。不过，directory 和 default_version 这两个参数不能在次级控制文件中设置。

扩展的 SQL 脚本文件可以包含任何 SQL 命令，但事务控制命令（BEGIN、COMMIT 等）和无法在事务块内执行的命令（例如 VACUUM）除外。这是因为脚本文件会被隐式地放在事务块中执行。

扩展的 SQL 脚本文件也可以包含以 \echo 开头的行，扩展机制会忽略这些行（将其视为注释）。这一约定通常用于在脚本文件被直接交给 psql 而不是通过 CREATE EXTENSION 装载时抛出错误（见下文中的示例脚本）。如果没有这种机制，用户可能会意外地把扩展内容作为“松散”对象装载，而不是作为一个扩展来装载，这种状态恢复起来会有些麻烦。

虽然脚本文件可以包含指定编码允许的任意字符，但控制文件应只包含纯 ASCII 字符，因为 PostgreSQL 无法知道控制文件采用的是什么编码。在实践中，只有当你想在扩展注释中使用非 ASCII 字符时这才会成为问题。对此推荐的做法是不要使用控制文件中的 comment 参数，而是在脚本文件中使用 COMMENT ON EXTENSION 来设置注释。

35.15.2. 扩展的可重定位性

用户经常希望把扩展中的对象装载到与扩展作者原先设想不同的模式中。对于这种可重定位性，支持三个级别：

- 完全可重定位的扩展可以在任何时候移动到另一个模式中，即使它已经被装载到数据库之后也是如此。这通过 ALTER EXTENSION SET SCHEMA 命令完成，该命令会自动把所有成员对象重命名到新模式中。通常，只有当扩展对其任何对象所在模式都没有内部假设时，才有可能做到这一点。此外，扩展的对象一开始必须全部位于同一个模式中（不属于任何模式的对象，如过程语言，不算在内）。要把一个扩展标记为完全可重定位，只需在其控制文件中设置 relocatable = true。

- 扩展可能在安装期间可重定位，但安装之后不可重定位。如果扩展脚本文件需要显式引用目标模式，例如为 SQL 函数设置 search_path 属性时，通常就是这种情况。对于这样的扩展，应在控制文件中设置 relocatable = false，并在脚本文件中使用 @extschema@ 来引用目标模式。在脚本执行前，该字符串的每次出现都会被替换为实际目标模式的名称。用户可以通过 CREATE EXTENSION 的 SCHEMA 选项设置目标模式。

- 如果扩展完全不支持重定位，应在控制文件中设置 relocatable = false，并把 schema 设置为预定目标模式的名称。这样将阻止使用 CREATE EXTENSION 的 SCHEMA 选项，除非它指定的正是控制文件中命名的那个模式。如果扩展对模式名称有无法通过 @extschema@ 替换解决的内部假设，通常就需要采用这种方式。在这种情况下，@extschema@ 替换机制仍然可用，只是由于模式名由控制文件决定，其用途比较有限。

在所有情况下，脚本文件执行时，其初始search_path都会指向目标模式；也就是说，CREATE EXTENSION 做的事情等价于：

```
SET LOCAL search_path TO @extschema@;
```

这使得脚本文件创建的对象能够进入目标模式。脚本文件当然也可以修改 search_path，但通常不建议这样做。CREATE EXTENSION 完成后，search_path 会恢复为先前的设置。

目标模式由控制文件中的 schema 参数决定（如果给出了该参数）；否则由 CREATE EXTENSION 的 SCHEMA 选项决定（如果给出了该选项）；否则由当前默认的对象创建模式决定（也就是调用者的 search_path 中第一个模式）。当使用控制文件中的 schema 参数时，如果目标模式尚不存在，则会自动创建；而在另外两种情况下，目标模式必须已经存在。

如果控制文件的 requires 中列出了任何前置扩展，它们的目标模式会追加到 search_path 的初始设置中。这样新扩展的脚本文件就可以看到这些扩展的对象。

虽然不可重定位扩展可以包含分布在多个模式中的对象，
但通常仍然希望把所有供外部使用的对象放在单个模式中，
这个模式会被视为该扩展的目标模式。这种安排与创建依赖扩展时 search_path 的默认设置配合起来会比较方便。

35.15.3. 扩展配置表

有些扩展包含配置表，其中存放的数据可能会在扩展安装后被用户新增或修改。通常，
如果表属于某个扩展，那么 pg_dump 既不会转储该表的定义，也不会转储其内容。但这种行为对于配置表来说并不理想；用户所做的数据更改必须包含在转储中，否则在转储并恢复之后，
扩展的行为就会发生变化。

为了解决这个问题，扩展脚本文件可以把它创建的某个表标记为配置表，这样 pg_dump 就会把该表的内容（而不是定义）包含到转储中。做法是在创建表之后调用函数 pg_extension_config_dump(regclass, text)，例如：

```
CREATE TABLE my_config (key text, value text);
SELECT pg_catalog.pg_extension_config_dump('my_config', '');
```

可以用这种方式标记任意数量的表。

当 pg_extension_config_dump 的第二个参数是空字符串时，pg_dump 会转储该表的全部内容。通常只有当该表在扩展脚本创建时最初为空时，这样做才是正确的。
如果表中混有初始数据和用户提供的数据，那么 pg_extension_config_dump 的第二个参数就提供了一个用于选择要转储数据的 WHERE 条件。例如，你可以这样做：

```
CREATE TABLE my_config (key text, value text, standard_entry boolean);

SELECT pg_catalog.pg_extension_config_dump('my_config', 'WHERE NOT
 standard_entry');
```

然后确保只有扩展脚本创建的那些行，其 standard_entry 才为真。

更复杂的情况，例如用户可能会修改最初提供的数据行，
可以通过在配置表上创建触发器来处理，以确保被修改的行被正确标记。

你可以再次调用 pg_extension_config_dump 来修改与配置表关联的过滤条件。（这通常在扩展更新脚本中很有用。）要把某个表标记为不再是配置表，唯一的方法是通过 ALTER EXTENSION ... DROP TABLE 将它与扩展解除关联。

注意，这些表之间的外键关系会决定 pg_dump 转储它们的顺序。具体来说，pg_dump 会尝试先转储被引用表，再转储引用表。由于外键关系是在 CREATE EXTENSION 时建立的（那时数据尚未装入这些表中），因此不支持环状依赖。若存在环状依赖，数据仍会被转储出来，
但该转储将无法直接恢复，需要用户介入处理。

35.15.4. 扩展更新

扩展机制的一个优点是，它为管理定义扩展对象的 SQL 命令的更新提供了便利方式。这是通过为扩展安装脚本的每个已发布版本关联一个版本名或版本号来实现的。此外，
如果你希望用户能够把数据库从一个版本动态更新到下一个版本，就应提供更新脚本，
以完成从一个版本切换到下一版本所需的修改。更新脚本的名称遵循如下模式：*extension--oldversion--newversion*.sql（例如，foo--1.0--1.1.sql包含把扩展foo的版本 1.0 修改为版本 1.1的命令）。

在有合适更新脚本可用的前提下，ALTER EXTENSION UPDATE 命令会把已安装的扩展更新到指定的新版本。更新脚本运行在 CREATE EXTENSION 为安装脚本提供的同一环境中：尤其是，search_path 的设置方式完全相同，而且脚本创建的任何新对象都会自动加入扩展中。

如果扩展有次级控制文件，那么用于更新脚本的控制参数就是与该脚本目标（新）
版本相关联的那些参数。

更新机制可以用于解决一个重要的特殊情况：将“松散”对象集合转换为扩展。在 PostgreSQL 于 9.1 加入扩展机制之前，许多人编写的扩展模块只是创建各种未打包的对象。对于包含这些对象的现有数据库，如何将它们转换为正确打包的扩展？删除它们再执行普通的 CREATE EXTENSION 是一种办法，但如果对象具有依赖关系（例如，某些表列使用扩展创建的数据类型），就不适合这样做。解决方法是先创建一个空扩展，再使用 ALTER EXTENSION ADD 将每个现有对象附加到扩展中，最后创建当前扩展版本中存在、而未打包版本中没有的新对象。CREATE EXTENSION 通过 FROM *old_version* 选项支持这种情况：它不运行目标版本的普通安装脚本，而是运行名为 *extension--old_version--target_version*.sql 的更新脚本。用作 *old_version* 的虚拟版本名由扩展作者选择，不过通常约定使用 unpackaged。如果需要将多个先前版本更新为扩展形式，应使用不同的虚拟版本名来标识它们。

ALTER EXTENSION 能够执行一系列更新脚本文件来完成请求的更新。例如，如果只有 foo--1.0--1.1.sql 和 foo--1.1--2.0.sql 可用，那么当当前安装的是 1.0，而请求更新到 2.0 时，ALTER EXTENSION 就会按顺序应用这两个脚本。

PostgreSQL 并不对版本名称的属性作任何假设：例如，它并不知道 1.1 是否跟在 1.0 之后。它只是匹配可用的版本名，并选择需要应用更新脚本最少的那条路径。（实际上，版本名可以是任何不包含 --，且不以 - 开头或结尾的字符串。）

有时提供“降级”脚本也是有用的，例如 foo--1.1--1.0.sql 可用于回退与 1.1 版相关的修改。如果你这样做，要注意某个降级脚本可能会因为产生更短的路径而意外被采用。
危险情况是，存在一个跨越多个版本的“快速路径”更新脚本，
同时又有一个可降级到该快速路径起点的脚本。这时，先降级再走快速路径，
可能比按版本逐步前进所需的步数更少。如果降级脚本删除了任何不可替代的对象，
就会产生不希望看到的结果。

要检查是否存在意外的更新路径，可使用以下命令：

```
SELECT * FROM pg_extension_update_paths('extension_name');
```

它会显示指定扩展的每一对不同已知版本名，以及从源版本到目标版本将采用的更新路径序列；如果没有可用的更新路径，则显示 NULL。路径会以文本形式显示，并使用 -- 作为分隔符。如果你更喜欢数组形式，可以使用 regexp_split_to_array(path,'--')。

35.15.5. 扩展示例

下面给出一个纯 SQL 扩展的完整示例：一个双元素复合类型，它可以在两个槽位中存储任意类型的值，这两个槽位名为“k”和“v”。非文本值会自动强制转换为文本后再存储。

脚本文件`pair--1.0.sql`如下所示：

```
-- complain if script is sourced in psql, rather than via CREATE
 EXTENSION
\echo Use "CREATE EXTENSION pair" to load this file. \quit

CREATE TYPE pair AS ( k text, v text );

CREATE OR REPLACE FUNCTION pair(anyelement, text)
RETURNS pair LANGUAGE SQL AS 'SELECT ROW($1, $2)::pair';

CREATE OR REPLACE FUNCTION pair(text, anyelement)
RETURNS pair LANGUAGE SQL AS 'SELECT ROW($1, $2)::pair';

CREATE OR REPLACE FUNCTION pair(anyelement, anyelement)
RETURNS pair LANGUAGE SQL AS 'SELECT ROW($1, $2)::pair';

CREATE OR REPLACE FUNCTION pair(text, text)
RETURNS pair LANGUAGE SQL AS 'SELECT ROW($1, $2)::pair;';

CREATE OPERATOR ~> (LEFTARG = text, RIGHTARG = anyelement, PROCEDURE =
 pair);
CREATE OPERATOR ~> (LEFTARG = anyelement, RIGHTARG = text, PROCEDURE =
 pair);
CREATE OPERATOR ~> (LEFTARG = anyelement, RIGHTARG = anyelement,
 PROCEDURE = pair);
CREATE OPERATOR ~> (LEFTARG = text, RIGHTARG = text, PROCEDURE = pair)
;
```

控制文件`pair.control`如下所示：

```
# pair extension
comment = 'A key/value pair data type'
default_version = '1.0'
relocatable = true
```

虽然你几乎不需要专门写一个 makefile 来把这两个文件安装到正确的目录中，但你也可以使用如下内容的 `Makefile`：

```
EXTENSION = pair
DATA = pair--1.0.sql

PG_CONFIG = pg_config
PGXS := $(shell $(PG_CONFIG) --pgxs)
include $(PGXS)
```

这个 makefile 依赖于 PGXS，其说明见第 35.16 节。执行 `make install` 命令会把控制文件和脚本文件安装到 pg_config 报告的正确目录中。

安装文件后，使用CREATE EXTENSION命令将这些对象装入某个具体数据库。

35.16. 扩展构建基础设施

如果你打算分发自己的 PostgreSQL 扩展模块，那么为它们搭建一个可移植的构建系统可能相当困难。因此，PostgreSQL 安装提供了一套称为 PGXS 的扩展构建基础设施，使

简单的扩展模块可以针对已安装好的服务器直接构建。PGXS 主要面向包含 C 代码的扩展，不过它也能用于纯 SQL 扩展。注意，PGXS 并不打算成为一个可以用来构建任意与 PostgreSQL 交互软件的通用构建系统框架；它只是把简单服务器扩展模块的常见构建规则自动化。对于更复杂的软件包，你可能还是需要自己编写构建系统。

要为你的扩展使用PGXS基础设施，必须编写一个简单的 makefile。在 makefile 中，你需要设置一些变量，并在最后包含全局的PGXSmakefile。下面的示例构建一个扩展模块，名为`isbn_issn`，由包含一些 C 代码的共享库、扩展控制文件、SQL 脚本以及文档文本文件组成：

```
MODULES = isbn_issn
EXTENSION = isbn_issn
DATA = isbn_issn--1.0.sql
DOCS = README.isbn_issn

PG_CONFIG = pg_config
PGXS := $(shell $(PG_CONFIG) --pgxs)
include $(PGXS)
```

最后三行应该总是相同的。在文件更前面的部分，你可以为变量赋值或者添加自定义的make规则。

设定下列三个变量中的一个，以指定要构建的内容：

`MODULES`

要从同名源文件构建的共享库对象列表（列表中不要包含库后缀）

`MODULE_big`

要从多个源文件构建的共享库（在 `OBJS` 中列出目标文件）

`PROGRAM`

要构建的可执行程序（在 `OBJS` 中列出目标文件）

The following variables can also be set:

`EXTENSION`

扩展名称；对于每个名称，你都必须提供一个 *extension*`.control` 文件，它将安装到 *prefix*`/share/extension`

`MODULEDIR`

prefix`/share` 下的子目录，用于安装 DATA 和 DOCS 文件（若未设置，则在设置了 `EXTENSION` 时默认为 `extension`，否则默认为 `contrib`）

`DATA`

要安装到 *prefix*`/share/$MODULEDIR` 的任意文件

`DATA_built`

要安装到 *prefix*`/share/$MODULEDIR` 的任意文件，但它们需要先被构建

`DATA_TSEARCH`

要安装到 *prefix*`/share/tsearch_data` 下的任意文件

DOCS

要安装到 *prefix*/doc/$MODULEDIR 下的任意文件

SCRIPTS

要安装到 *prefix*/bin 的脚本文件（不是二进制文件）

SCRIPTS_built

要安装到 *prefix*/bin 的脚本文件（不是二进制文件），但它们需要先被构建

REGRESS

回归测试用例列表（不带后缀），详见下文

REGRESS_OPTS

传递给 pg_regress 的额外开关

EXTRA_CLEAN

在 make clean 中要额外删除的文件

PG_CPPFLAGS

将被添加到 CPPFLAGS

PG_LIBS

将被加入 PROGRAM 的链接命令行

SHLIB_LINK

将被加入 MODULE_big 的链接命令行

PG_CONFIG

要针对其进行构建的 PostgreSQL 安装所对应的 pg_config 程序路径（通常只写 pg_config，表示使用你 PATH 中找到的第一个）

把这个 makefile 命名为 Makefile，并放在保存扩展的目录中。然后你就可以执行 make 进行编译，再执行 make install 安装你的模块。默认情况下，该扩展会针对你 PATH 中找到的第一个 pg_config 所对应的 PostgreSQL 安装进行编译和安装。你也可以使用不同的安装，只需让 PG_CONFIG 指向它的 pg_config 程序，无论是在 makefile 中设置，还是在 make 命令行上设置都可以。

> **小心**
>
> 更改PG_CONFIG只在针对PostgreSQL 8.3 或更新版本构建时有效。对于更老的发行版本，把它设置为pg_config之外的任何值都不起作用；你必须修改你的PATH
> 来选择要针对其构建的安装。

在 REGRESS 变量中列出的脚本用于对模块做回归测试，可在执行完 make install 之后，通过 make installcheck 来运行。要让它工作，你必须有一个正在运行的 PostgreSQL 服务器。

列在 `REGRESS` 中的脚本文件必须位于扩展目录下名为 `sql/` 的子目录中。这些文件必须具有 `.sql` 扩展名，而该扩展名不能出现在 makefile 的 `REGRESS` 列表中。对于每个测试，还应在名为 `expected/` 的子目录中有一个包含期望输出的文件，其主干名相同，扩展名为 `.out`。`make installcheck` 会用 psql 执行每个测试脚本，并把得到的输出与对应的期望文件比较。任何差异都会以 `diff -c` 格式写入 `regression.diffs` 文件。注意，如果尝试运行一个缺少期望文件的测试，将被报告为“trouble”，所以请确保所有期望文件都已准备好。

提示

创建期望文件最简单的方法是先建立空文件，然后运行一次测试（当然这会报告差异）。检查 `results/` 目录中的实际结果文件；如果它们与你对测试的预期一致，就把它们复制到 `expected/` 中。

第 36 章 触发器

本章提供有关编写触发器函数的一般信息。触发器函数可以用大多数可用的过程语言编写，包括 PL/pgSQL（第 39 章）、PL/Tcl（第 40 章）、PL/Perl（第 41 章）和 PL/Python（第 42 章）。读完本章后，你还应查阅所用过程语言对应的章节，了解用该语言编写触发器的语言相关细节。

也可以用 C 编写触发器函数，不过大多数人会觉得使用某种过程语言更容易。目前还不能用普通 SQL 函数语言编写触发器函数。

36.1. 触发器行为概述

触发器是一种规定：每当执行某种类型的操作时，数据库就自动执行某个特定函数。
触发器可以附加到表和视图上。

在表上，触发器可以定义为在任何 `INSERT`、`UPDATE` 或 `DELETE` 操作之前或之后执行；既可以对每个被修改的行执行一次，也可以对每个 SQL 语句执行一次。`UPDATE` 触发器还可以设置成只有在 `UPDATE` 语句的 `SET` 子句中提及某些列时才触发。Triggers can also fire for `TRUNCATE` statements. If a trigger event occurs, the trigger's function is called at the appropriate time to handle the event.

在视图上，触发器可以定义为代替 `INSERT`、`UPDATE` 或 `DELETE` 操作执行。`INSTEAD OF` 触发器会对视图中每个需要修改的行触发一次。触发器函数负责对视图底层的基表执行必要的修改，并在适当时返回该修改后的行，使其在视图中呈现为应有的样子。
视图上的触发器也可以定义为对每个 SQL 语句执行一次，在 `INSERT`、`UPDATE` 或 `DELETE` 操作之前或之后触发。

必须先定义触发器函数，之后才能创建触发器本身。
触发器函数必须声明为不接受参数并返回类型 `trigger` 的函数。（触发器函数通过特殊传入的 `TriggerData` 结构体接收输入，而不是通过普通函数参数。）

创建出合适的触发器函数后，就可以使用CREATE TRIGGER创建触发器。同一个触发器函数可以供多个触发器使用。

PostgreSQL 同时提供每行触发器和每语句触发器。对于每行触发器，触发它的语句所影响的每一行都会调用一次触发器函数。相反，每语句触发器只会在执行相应语句时调用一次，
不管该语句影响了多少行。特别是，即使某个语句一行也没有影响，
仍会执行所有适用的每语句触发器。这两类触发器有时也分别称为行级触发器和语句级触发器。`TRUNCATE` 的触发器只能定义为语句级。在视图上，之前或之后触发的触发器只能定义为语句级，而代替 `INSERT`、`UPDATE` 或 `DELETE` 触发的触发器只能定义为行级。

触发器还可以按照它们是在操作之前、之后，还是取代该操作而触发来分类。
它们分别被称为`BEFORE` 触发器、`AFTER` 触发器和`INSTEAD OF` 触发器。Statement-level `BEFORE` triggers naturally fire before the statement starts to do anything, while statement-level `AFTER` triggers fire at the very end of the statement. These types of triggers may be defined on tables or views. Row-level `BEFORE` triggers fire immediately before a particular row is operated on, while row-level `AFTER` triggers fire at the end of the statement (but before any statement-level `AFTER` triggers). These types of triggers may only be defined on tables. Row-level `INSTEAD OF` triggers may only be defined on views, and fire immediately as each row in the view is identified as needing to be operated on.

由每语句触发器调用的触发器函数应始终返回 `NULL`。由每行触发器调用的触发器函数则可以根据需要向调用它的执行器返回一个表行（即类型 `HeapTuple` 的值）。在操作之前触发的行级触发器可以作出如下选择：

- 它可以返回 NULL，以跳过对当前行的操作。这会指示执行器不要执行引发该触发器的行级操作（即对某个具体表行的插入、修改或删除）。
- 仅对于行级 INSERT 和 UPDATE 触发器，返回的行会成为将要插入的行，或替换正在更新的行。这使得触发器函数可以修改即将插入或更新的行。

不打算产生上述任何一种行为的行级 BEFORE 触发器，必须小心地返回传入的同一行作为结果（也就是说，INSERT 和 UPDATE 触发器返回 NEW 行，DELETE 触发器返回 OLD 行）。

行级 INSTEAD OF 触发器应当要么返回 NULL，表示它没有修改视图底层基表中的任何数据；要么返回传入的视图行（INSERT 和 UPDATE 操作返回 NEW 行，DELETE 操作返回 OLD 行）。非空返回值用于表明触发器完成了视图中必要的数据修改。这会使该命令所影响行数的计数递增。对于 INSERT 和 UPDATE 操作，触发器可以在返回前修改 NEW 行。这会改变 INSERT RETURNING or UPDATE RETURNING, and is useful when the view will not show exactly the same data that was provided.

对于在操作之后触发的行级触发器，其返回值会被忽略，因此它们可以返回 NULL。

如果为同一关系上的同一事件定义了多个触发器，它们会按触发器名称的字母顺序触发。对于 BEFORE 和 INSTEAD OF 触发器，每个触发器返回的、可能已经被修改过的行都会成为下一个触发器的输入。如果任何一个 BEFORE 或 INSTEAD OF 触发器返回 NULL，则会放弃对该行执行该操作，并且后续触发器（针对该行）都不会再触发。

触发器定义还可以指定一个布尔 WHEN 条件，用来测试是否应当触发该触发器。对于行级触发器，WHEN 条件可以检查该行各列的旧值和/或新值。（语句级触发器也可以有 WHEN 条件，不过这个特性对它们用处不大。）在 BEFORE 触发器中，WHEN 条件就在函数即将执行或本会执行之前求值，因此使用 WHEN 与在触发器函数开头测试同样的条件并无实质差别。不过，在 AFTER 触发器中，WHEN 条件会在行更改发生后立即求值，并决定是否要将一个事件排入队列，以便在语句末尾触发该触发器。因此，当 AFTER 触发器的 WHEN 条件不返回真时，就不必排队该事件，也不必在语句末尾重新取出该行。如果触发器只需要针对少数几行触发，这会让修改大量行的语句显著加快。INSTEAD OF 触发器不支持 WHEN 条件。

通常，行级 BEFORE 触发器用于检查或修改将要插入或更新的数据。例如，BEFORE 触发器可以用来把当前时间写入某个 timestamp 列，或者检查该行的两个元素是否一致。行级 AFTER 触发器最适合用于把更新传播到其他表，或对其他表执行一致性检查。之所以这样分工，是因为 AFTER 触发器可以确定自己看到的是该行的最终值，而 BEFORE 触发器不能；它之后可能还会有其他 BEFORE 触发器触发。如果没有特定理由把触发器做成 BEFORE 或 AFTER，那么 BEFORE 形式更高效，因为关于该操作的信息无需一直保存到语句末尾。

如果触发器函数执行 SQL 命令，那么这些命令可能会再次触发触发器。这就是所谓的级联触发器。对级联层数没有直接限制。级联还有可能导致同一个触发器被递归调用；例如，一个 INSERT 触发器可能执行一条向同一个表再插入一行的命令，从而导致该 INSERT 触发器再次触发。因此，在这种场景下，避免无限递归是触发器编写者自己的责任。

在定义触发器时，可以为它指定参数。在触发器定义中包含参数，
是为了让需求相似的不同触发器能够调用同一个函数。举例来说，可以有一个通用触发器函数，
它接受两个列名作为参数，把当前用户写入其中一个列，把当前时间戳写入另一个列。
只要编写得当，这个触发器函数就应独立于它所作用的具体表。因此，
同一个函数可用于任何具有适当列的表上的 INSERT 事件，例如自动跟踪某个事务表中记录的创建。如果把它定义成 UPDATE 触发器，还可以用来跟踪最近更新事件。

每种支持触发器的编程语言都有自己的方法，使触发器输入数据可供触发器函数使用。
这些输入数据包括触发器事件的类型（例如 INSERT 或 UPDATE），以及 CREATE TRIGGER 中列出的任何参数。对于行级触发器，输入数据还包括 INSERT 和 UPDATE 触发器的 NEW 行，以及/或者 UPDATE 和 DELETE 触发器的 OLD 行。语句级触发器目前无法检查该语句修改的各个行。

36.2. 数据更改的可见性

如果你在触发器函数中执行 SQL 命令，而这些命令又访问该触发器所属的表，就需要了解数据可见性规则，因为这些规则决定了这些 SQL 命令是否能看到引发触发器的数据更改。简而言之：

- 语句级触发器遵循简单的可见性规则：语句所做的任何更改对在语句之前调用的语句级触发器都不可见，而所有修改对语句级 `AFTER` 触发器都可见。
- 导致触发器触发的数据更改（插入、更新或删除）对于在行级 `BEFORE` 触发器中执行的 SQL 命令自然是不可见的，因为它尚未发生。
- 然而，在行级 `BEFORE` 触发器中执行的 SQL 命令会看到同一外层命令中先前已处理各行上的数据更改结果。这里必须小心，因为这些更改事件的顺序通常不可预测；一条影响多行的 SQL 命令可能按任意顺序访问这些行。
- 类似地，行级 `INSTEAD OF` 触发器会看到同一外层命令中此前触发的 `INSTEAD OF` 触发器所造成的数据更改结果。
- 当行级 `AFTER` 触发器触发时，外层命令所做的所有数据更改都已经完成，并且对被调用的触发器函数可见。

如果你的触发器函数是用任何一种标准过程语言编写的，那么只有在该函数被声明为 `VOLATILE` 时，上述说法才成立。被声明为 `STABLE` 或 `IMMUTABLE` 的函数在任何情况下都看不到调用命令所做的更改。

有关数据可见性规则的更多信息可见第 43.4 节。第 36.4 节中的示例展示了这些规则。

36.3. 用 C 编写触发器函数

本节说明触发器函数接口的底层细节。这些信息仅在用 C 编写触发器函数时才需要。如果你使用更高层语言，这些细节会由系统代为处理。在多数情况下，在用 C 编写触发器之前，应先考虑使用过程语言。每种过程语言的文档都会说明如何用该语言编写触发器。

触发器函数必须使用“版本 1”函数管理器接口。

当函数由触发器管理器调用时，不会向它传递任何普通参数，但会传递一个指向 `TriggerData` 结构体的“上下文”指针。C 函数可以通过执行下列宏来检查自己是否由触发器管理器调用：

```
CALLED_AS_TRIGGER(fcinfo)
```

它会展开为：

```
((fcinfo)->context != NULL && IsA((fcinfo)->context, TriggerData))
```

如果该宏返回真，就可以安全地把 `fcinfo->context` 转换为 `TriggerData *` 类型，并使用其所指向的 `TriggerData` 结构体。函数绝不能修改 `TriggerData` 结构体本身，也不能修改它所指向的任何数据。

`struct TriggerData`定义在`commands/trigger.h`：

```
typedef struct TriggerData
{
    NodeTag       type;
    TriggerEvent  tg_event;
```

```
    Relation      tg_relation;
    HeapTuple     tg_trigtuple;
    HeapTuple     tg_newtuple;
    Trigger      *tg_trigger;
    Buffer        tg_trigtuplebuf;
    Buffer        tg_newtuplebuf;
} TriggerData;
```

其中各成员的定义如下：

type

始终是 T_TriggerData。

tg_event

描述调用函数的事件。可以使用以下宏检查 tg_event：

TRIGGER_FIRED_BEFORE(tg_event)

如果触发器在操作之前触发，则返回真。

TRIGGER_FIRED_AFTER(tg_event)

如果触发器在操作之后触发，则返回真。

TRIGGER_FIRED_INSTEAD(tg_event)

如果触发器是取代该操作而触发，则返回真。

TRIGGER_FIRED_FOR_ROW(tg_event)

如果触发器是因行级事件触发，则返回真。

TRIGGER_FIRED_FOR_STATEMENT(tg_event)

如果触发器是因语句级事件触发，则返回真。

TRIGGER_FIRED_BY_INSERT(tg_event)

如果触发器是由 INSERT 命令触发，则返回真。

TRIGGER_FIRED_BY_UPDATE(tg_event)

如果触发器是由 UPDATE 命令触发，则返回真。

TRIGGER_FIRED_BY_DELETE(tg_event)

如果触发器是由 DELETE 命令触发，则返回真。

TRIGGER_FIRED_BY_TRUNCATE(tg_event)

如果触发器是由 TRUNCATE 命令触发，则返回真。

tg_relation

指向一个描述该触发器所针对关系的结构体。有关此结构体的细节见 utils/rel.h。其中最值得关注的是 tg_relation->rd_att（关系元组的描述符）和 tg_relation->rd_

rel->relname（关系名；其类型不是 char*，而是 NameData；如果需要名称的副本，请使用 SPI_getrelname(tg_relation) 取得一个 char*）。

tg_trigtuple

指向触发该触发器的那一行。这是正在被插入、更新或删除的行。如果该触发器因 INSERT 或 DELETE 而触发，那么如果你不想用另一行替换该行（在 INSERT 的情形下）或跳过该操作，就应从函数中返回它。

tg_newtuple

如果触发器因 UPDATE 而触发，则指向该行的新版本；如果因 INSERT 或 DELETE 而触发，则为 NULL。如果事件是 UPDATE，而你不想用另一行替换该行或跳过该操作，就必须从函数中返回它。

tg_trigger

一个指向Trigger类型结构体的指针，该结构体定义在utils/rel.h：

```
typedef struct Trigger
{
    Oid         tgoid;
    char       *tgname;
    Oid         tgfoid;
    int16       tgtype;
    char        tgenabled;
    bool        tgisinternal;
    Oid         tgconstrrelid;
    Oid         tgconstrindid;
    Oid         tgconstraint;
    bool        tgdeferrable;
    bool        tginitdeferred;
    int16       tgnargs;
    int16       tgnattr;
    int16      *tgattr;
    char      **tgargs;
    char       *tgqual;
} Trigger;
```

where tgname is the trigger's name, tgnargs is the number of arguments in tgargs, and tgargs is an array of pointers to the arguments specified in the CREATE TRIGGER statement. The other members are for internal use only.

tg_trigtuplebuf

包含 tg_trigtuple 的缓冲区；如果没有这样的元组，或它没有存储在磁盘缓冲区中，则为 InvalidBuffer。

tg_newtuplebuf

包含 tg_newtuple 的缓冲区；如果没有这样的元组，或它没有存储在磁盘缓冲区中，则为 InvalidBuffer。

触发器函数必须返回一个 HeapTuple 指针或一个 NULL 指针（不是 SQL 空值，也就是说不要把 *isNull* 设为真）。如果你不想修改正在处理的行，就要小心地根据情况返回适当的 tg_trigtuple 或 tg_newtuple。

36.4. 一个完整的触发器示例

这里给出一个非常简单的、用 C 编写的触发器函数示例。（用过程语言编写的触发器示例可见各过程语言的文档。）

函数 `trigf` 会报告表 `ttest` 中的行数，并在命令试图向列 `x` 中插入空值时跳过实际操作。（这样，这个触发器就起到了非空约束的作用，但不会中止事务。）

首先，表定义：

```
CREATE TABLE ttest (
    x integer
);
```

下面是触发器函数的源代码：

```
#include "postgres.h"
#include "executor/spi.h"       /* this is what you need to work with
 SPI */
#include "commands/trigger.h"   /* ... and triggers */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

extern Datum trigf(PG_FUNCTION_ARGS);

PG_FUNCTION_INFO_V1(trigf);

Datum
trigf(PG_FUNCTION_ARGS)
{
    TriggerData *trigdata = (TriggerData *) fcinfo->context;
    TupleDesc   tupdesc;
    HeapTuple   rettuple;
    char       *when;
    bool        checknull = false;
    bool        isnull;
    int         ret, i;

    /* make sure it's called as a trigger at all */
    if (!CALLED_AS_TRIGGER(fcinfo))
        elog(ERROR, "trigf: not called by trigger manager");

    /* tuple to return to executor */
    if (TRIGGER_FIRED_BY_UPDATE(trigdata->tg_event))
        rettuple = trigdata->tg_newtuple;
    else
        rettuple = trigdata->tg_trigtuple;

    /* check for null values */
    if (!TRIGGER_FIRED_BY_DELETE(trigdata->tg_event)
        && TRIGGER_FIRED_BEFORE(trigdata->tg_event))
        checknull = true;
```

```
    if (TRIGGER_FIRED_BEFORE(trigdata->tg_event))
        when = "before";
    else
        when = "after ";

    tupdesc = trigdata->tg_relation->rd_att;

    /* connect to SPI manager */
    if ((ret = SPI_connect()) < 0)
        elog(ERROR, "trigf (fired %s): SPI_connect returned %d", when,
 ret);

    /* get number of rows in table */
    ret = SPI_exec("SELECT count(*) FROM ttest", 0);

    if (ret < 0)
        elog(ERROR, "trigf (fired %s): SPI_exec returned %d", when,
 ret);

    /* count(*) returns int8, so be careful to convert */
    i = DatumGetInt64(SPI_getbinval(SPI_tuptable->vals[0],
                                    SPI_tuptable->tupdesc,
                                    1,
                                    &isnull));

    elog (INFO, "trigf (fired %s): there are %d rows in ttest", when,
 i);

    SPI_finish();

    if (checknull)
    {
        SPI_getbinval(rettuple, tupdesc, 1, &isnull);
        if (isnull)
            rettuple = NULL;
    }

    return PointerGetDatum(rettuple);
}
```

编译完源代码后（参见第 35.9.6 节），声明函数和触发器：

```
CREATE FUNCTION trigf() RETURNS trigger
    AS 'filename'
    LANGUAGE C;

CREATE TRIGGER tbefore BEFORE INSERT OR UPDATE OR DELETE ON ttest
    FOR EACH ROW EXECUTE PROCEDURE trigf();

CREATE TRIGGER tafter AFTER INSERT OR UPDATE OR DELETE ON ttest
    FOR EACH ROW EXECUTE PROCEDURE trigf();
```

现在可以测试该触发器的行为：

```
=> INSERT INTO ttest VALUES (NULL);
INFO:  trigf (fired before): there are 0 rows in ttest
INSERT 0 0

-- Insertion skipped and AFTER trigger is not fired

=> SELECT * FROM ttest;
 x
---
(0 rows)

=> INSERT INTO ttest VALUES (1);
INFO:  trigf (fired before): there are 0 rows in ttest
INFO:  trigf (fired after ): there are 1 rows in ttest
                                       ^^^^^^^^
                             remember what we said about visibility.
INSERT 167793 1
vac=> SELECT * FROM ttest;
 x
---
 1
(1 row)

=> INSERT INTO ttest SELECT x * 2 FROM ttest;
INFO:  trigf (fired before): there are 1 rows in ttest
INFO:  trigf (fired after ): there are 2 rows in ttest
                                       ^^^^^^
                             remember what we said about visibility.
INSERT 167794 1
=> SELECT * FROM ttest;
 x
---
 1
 2
(2 rows)

=> UPDATE ttest SET x = NULL WHERE x = 2;
INFO:  trigf (fired before): there are 2 rows in ttest
UPDATE 0
=> UPDATE ttest SET x = 4 WHERE x = 2;
INFO:  trigf (fired before): there are 2 rows in ttest
INFO:  trigf (fired after ): there are 2 rows in ttest
UPDATE 1
vac=> SELECT * FROM ttest;
 x
---
 1
 4
(2 rows)

=> DELETE FROM ttest;
INFO:  trigf (fired before): there are 2 rows in ttest
INFO:  trigf (fired before): there are 1 rows in ttest
INFO:  trigf (fired after ): there are 0 rows in ttest
```

```
INFO:  trigf (fired after ): there are 0 rows in ttest
                                       ^^^^^^
                             remember what we said about visibility.
DELETE 2
=> SELECT * FROM ttest;
 x
---
(0 rows)
```

更复杂的示例可见 `src/test/regress/regress.c` 和spi。

第 37 章 规则系统

本章讨论PostgreSQL中的规则系统。产生式规则系统在概念上很简单，但在实际使用时涉及许多微妙之处。

有些其他数据库系统定义了主动型数据库规则，它们通常表现为存储过程和触发器。在PostgreSQL中，这类功能同样可以用函数和触发器实现。

规则系统（更准确地说，查询重写规则系统）与存储过程和触发器完全不同。它会修改查询，使其将规则纳入考虑，然后把修改后的查询交给查询规划器进行规划和执行。它非常强大，可用于查询语言过程、视图和版本等许多用途。关于这一规则系统的理论基础及其能力，另见[ston90b]和[ong90]。

37.1. 查询树

要理解规则系统如何工作，必须先知道它在何时被调用，以及它的输入和输出是什么。

规则系统位于解析器和规划器之间。它接收解析器的输出，即一棵查询树，以及用户定义的重写规则；这些规则本身也是带有一些附加信息的查询树。它会生成零棵或多棵查询树作为结果。因此，它的输入和输出始终都是解析器本身也可能产生的东西，所以它看到的任何内容基本上都可以表示为一个SQL语句。

那么什么是查询树？它是SQL语句的一种内部表示，其中构成语句的各个部分被分别存储。如果你设置了配置参数`debug_print_parse`、`debug_print_rewritten`或`debug_print_plan`，这些查询树就可以显示在服务器日志中。规则动作也以查询树的形式存储在系统目录`pg_rewrite`中。它们的格式不像日志输出，但包含的却是完全相同的信息。

阅读原始查询树需要一些经验。但由于查询树的SQL表示已足以理解规则系统，本章不会讲解如何阅读它们。

阅读本章查询树的 SQL 表示时，必须能够识别语句以查询树结构表示时被拆分成的各个部分。查询树的组成部分如下：

命令类型

这是一个简单的值，用来说明是哪一种命令（`SELECT`、`INSERT`、`UPDATE`、`DELETE`）产生了该查询树。

范围表

范围表是查询中使用的关系列表。在`SELECT`语句中，它们就是关键字`FROM`后面给出的关系。

每个范围表项标识一个表或视图，并说明在查询的其他部分以哪个名称来称呼它。在查询树中，范围表项是按编号而不是按名称引用的，因此即使像SQL语句中那样出现重名，这里也无关紧要。这种情况可能出现在规则的范围表被合并之后。本章中的示例不会涉及这种情形。

结果关系

这是范围表中的一个索引，用来标识查询结果应写入哪个关系。

`SELECT`查询没有结果关系。（`SELECT INTO`这一特殊情况与`CREATE TABLE`后接`INSERT ... SELECT`几乎相同，这里不再单独讨论。）

对于INSERT、UPDATE和DELETE命令，结果关系就是要让修改生效的表（或视图！）。

目标列表

目标列表是定义查询结果的表达式列表。对于SELECT，这些表达式构成查询的最终输出。它们对应于关键字SELECT和FROM之间的表达式。（*只是某个关系全部列名的缩写。解析器会把它展开为各个独立列，因此规则系统永远不会看到它。）

DELETE命令不需要普通的目标列表，因为它们不产生任何结果。相反，规划器会向空目标列表加入一个特殊的CTID项，以便执行器找到要删除的行。（当结果关系是普通表时，会加入CTID；如果结果关系是视图，则规则系统会改为加入一个整行变量，如第 37.2.4 节所述。）

对于INSERT命令，目标列表描述的是将要进入结果关系的新行。它由VALUES子句中的表达式，或INSERT ... SELECT里SELECT子句中的表达式组成。重写过程的第一步会为原始命令未赋值但带有默认值的列补上目标列表项。其余列（既没有给定值也没有默认值）将由规划器填入常量空值表达式。

对于UPDATE命令，目标列表描述要替换旧行的新行。在规则系统中，它只包含命令中SET column = expression部分的表达式。规划器会通过插入把旧行值复制到新行的表达式来处理缺失列。与DELETE一样，还会加入一个CTID或整行变量，以便执行器标识要更新的旧行。

目标列表中的每一项都包含一个表达式，它可以是常量值、指向范围表中某个关系列的变量、一个参数，或者由函数调用、常量、变量、操作符等构成的表达式树。

条件

查询的条件是一个表达式，与目标列表项中的表达式很相似。该表达式的结果值是布尔值，用来说明是否应对最终结果行执行相应操作（INSERT、UPDATE、DELETE或SELECT）。它对应于SQL语句的WHERE子句。

连接树

查询的连接树显示了FROM子句的结构。对于SELECT ... FROM a, b, c这样的简单查询，连接树只是FROM项的一个列表，因为允许按任意顺序连接它们。但当使用JOIN表达式，特别是外连接时，就必须按连接所显示的顺序进行连接。在这种情况下，连接树展示的是JOIN表达式的结构。与特定JOIN子句相关的限制（来自ON或USING表达式）会作为条件表达式附着在相应的连接树节点上。把顶层WHERE表达式也存储为附着在顶层连接树项上的一个条件，会很方便。因此，连接树实际上同时表示SELECT的FROM和WHERE子句。

其他

查询树的其他部分，如ORDER BY子句，这里不作关注。规则系统在应用规则时会替换其中某些项，但这与规则系统的基本原理关系不大。

37.2. 视图和规则系统

在 PostgreSQL 中，视图通过规则系统实现。实际上，以下命令：

```
CREATE VIEW myview AS SELECT * FROM mytab;
```

与下面这两条命令基本没有区别：

```
CREATE TABLE myview (same column list as mytab);
CREATE RULE "_RETURN" AS ON SELECT TO myview DO INSTEAD
    SELECT * FROM mytab;
```

因为这正是 CREATE VIEW 命令在内部所做的事情。这会带来一些副作用。其中之一是，在 PostgreSQL 系统目录中，视图的信息与表的信息完全相同。因此，对解析器而言，表和视图完全没有区别。它们是同一种东西：关系。

37.2.1. SELECT 规则如何工作

ON SELECT规则会在所有查询上作为最后一步应用，即使给出的命令是INSERT、UPDATE或DELETE也一样。它们与其他命令类型上的规则在语义上不同，因为它们是就地修改查询树，而不是创建新的查询树。因此我们先讨论SELECT规则。

目前，一个ON SELECT规则中只能有一个动作，而且它必须是一个无条件、带有INSTEAD的SELECT动作。之所以有这个限制，是为了让规则足够安全，从而能够向普通用户开放；它也把ON SELECT规则限制为像视图那样工作。

本章的示例是两个进行一些计算的连接视图，以及另外一些依次使用它们的视图。最初两个视图中的一个会在后面通过为INSERT、UPDATE和DELETE操作添加规则来定制，从而最终得到一个在行为上像真正的表、但又带有某些特殊功能的视图。作为入门示例，这并不算简单，因此会让理解变得更难一些。但与其使用许多可能令人混淆的不同示例，不如用一个示例逐步覆盖这里讨论的全部要点。

这个例子需要一个返回两个整数中较小者的 min 函数。我们这样创建它：

```
CREATE FUNCTION min(integer, integer) RETURNS integer AS $$
    SELECT CASE WHEN $1 < $2 THEN $1 ELSE $2 END
$$ LANGUAGE SQL STRICT;
```

在前两节对规则系统的描述中，我们需要用到如下真实表：

```
CREATE TABLE shoe_data (
    shoename   text,          -- primary key
    sh_avail   integer,       -- available number of pairs
    slcolor    text,          -- preferred shoelace color
    slminlen   real,          -- minimum shoelace length
    slmaxlen   real,          -- maximum shoelace length
    slunit     text           -- length unit
);

CREATE TABLE shoelace_data (
    sl_name    text,          -- primary key
    sl_avail   integer,       -- available number of pairs
    sl_color   text,          -- shoelace color
    sl_len     real,          -- shoelace length
    sl_unit    text           -- length unit
);

CREATE TABLE unit (
    un_name    text,          -- primary key
    un_fact    real           -- factor to transform to cm
);
```

如你所见，它们表示的是鞋店数据。

这些视图是这样创建的：

```
CREATE VIEW shoe AS
    SELECT sh.shoename,
           sh.sh_avail,
           sh.slcolor,
           sh.slminlen,
           sh.slminlen * un.un_fact AS slminlen_cm,
           sh.slmaxlen,
           sh.slmaxlen * un.un_fact AS slmaxlen_cm,
           sh.slunit
      FROM shoe_data sh, unit un
     WHERE sh.slunit = un.un_name;

CREATE VIEW shoelace AS
    SELECT s.sl_name,
           s.sl_avail,
           s.sl_color,
           s.sl_len,
           s.sl_unit,
           s.sl_len * u.un_fact AS sl_len_cm
      FROM shoelace_data s, unit u
     WHERE s.sl_unit = u.un_name;

CREATE VIEW shoe_ready AS
    SELECT rsh.shoename,
           rsh.sh_avail,
           rsl.sl_name,
           rsl.sl_avail,
           least(rsh.sh_avail, rsl.sl_avail) AS total_avail
      FROM shoe rsh, shoelace rsl
     WHERE rsl.sl_color = rsh.slcolor
       AND rsl.sl_len_cm >= rsh.slminlen_cm
       AND rsl.sl_len_cm <= rsh.slmaxlen_cm;
```

创建shoelace视图的CREATE VIEW命令（这是我们这里最简单的一个例子）会创建一个关系shoelace，并在pg_rewrite中创建一项，说明只要查询的范围表中引用了关系shoelace，就必须应用一条重写规则。该规则没有规则条件（稍后在讨论非SELECT规则时再谈，因为目前SELECT规则不能有规则条件），并且它是INSTEAD规则。注意，规则条件与查询条件不是一回事。这里我们的规则动作带有一个查询条件。规则动作本身是一棵查询树，它是视图创建命令中SELECT语句的一个副本。

> **注意**
>
> 你在pg_rewrite项中看到的、用于NEW和OLD的两个额外范围表项，与SELECT规则无关。

现在我们填充 unit，shoe_data 和 shoelace_data，然后对视图运行一个简单查询：

```
INSERT INTO unit VALUES ('cm', 1.0);
INSERT INTO unit VALUES ('m', 100.0);
INSERT INTO unit VALUES ('inch', 2.54);
```

```
INSERT INTO shoe_data VALUES ('sh1', 2, 'black', 70.0, 90.0, 'cm');
INSERT INTO shoe_data VALUES ('sh2', 0, 'black', 30.0, 40.0, 'inch');
INSERT INTO shoe_data VALUES ('sh3', 4, 'brown', 50.0, 65.0, 'cm');
INSERT INTO shoe_data VALUES ('sh4', 3, 'brown', 40.0, 50.0, 'inch');

INSERT INTO shoelace_data VALUES ('sl1', 5, 'black', 80.0, 'cm');
INSERT INTO shoelace_data VALUES ('sl2', 6, 'black', 100.0, 'cm');
INSERT INTO shoelace_data VALUES ('sl3', 0, 'black', 35.0 , 'inch');
INSERT INTO shoelace_data VALUES ('sl4', 8, 'black', 40.0 , 'inch');
INSERT INTO shoelace_data VALUES ('sl5', 4, 'brown', 1.0 , 'm');
INSERT INTO shoelace_data VALUES ('sl6', 0, 'brown', 0.9 , 'm');
INSERT INTO shoelace_data VALUES ('sl7', 7, 'brown', 60 , 'cm');
INSERT INTO shoelace_data VALUES ('sl8', 1, 'brown', 40 , 'inch');

SELECT * FROM shoelace;

 sl_name   | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
-----------+----------+----------+--------+---------+-----------
 sl1       |        5 | black    |     80 | cm      |        80
 sl2       |        6 | black    |    100 | cm      |       100
 sl7       |        7 | brown    |     60 | cm      |        60
 sl3       |        0 | black    |     35 | inch    |      88.9
 sl4       |        8 | black    |     40 | inch    |     101.6
 sl8       |        1 | brown    |     40 | inch    |     101.6
 sl5       |        4 | brown    |      1 | m       |       100
 sl6       |        0 | brown    |    0.9 | m       |        90
(8 rows)
```

这是你能在这些视图上执行的最简单的SELECT，因此我们借此机会说明视图规则的基础。SELECT * FROM shoelace由解析器解释后，会生成如下查询树：

```
SELECT shoelace.sl_name, shoelace.sl_avail,
       shoelace.sl_color, shoelace.sl_len,
       shoelace.sl_unit, shoelace.sl_len_cm
  FROM shoelace shoelace;
```

然后它会被交给规则系统。规则系统遍历范围表，检查其中的关系是否有相应规则。在处理shoelace的范围表项时（到目前为止只有这一个），它会找到带有如下查询树的_RETURN规则：

```
SELECT s.sl_name, s.sl_avail,
       s.sl_color, s.sl_len, s.sl_unit,
       s.sl_len * u.un_fact AS sl_len_cm
  FROM shoelace old, shoelace new,
       shoelace_data s, unit u
 WHERE s.sl_unit = u.un_name;
```

为了展开视图，重写器只需创建一个子查询范围表条目，其中包含规则的动作查询树，然后用这个范围表条目替换原先引用视图的条目。
所得的重写后查询树几乎等同于输入以下语句的结果：

```
SELECT shoelace.sl_name, shoelace.sl_avail,
       shoelace.sl_color, shoelace.sl_len,
       shoelace.sl_unit, shoelace.sl_len_cm
```

```
  FROM (SELECT s.sl_name,
               s.sl_avail,
               s.sl_color,
               s.sl_len,
               s.sl_unit,
               s.sl_len * u.un_fact AS sl_len_cm
          FROM shoelace_data s, unit u
         WHERE s.sl_unit = u.un_name) shoelace;
```

不过有一点不同：子查询的范围表中有两个额外条目，shoelace old 和 shoelace new。这些条目不直接参与查询，因为子查询的连接树或目标列表并未引用它们。
重写器用它们保存原先引用视图的范围表条目中的访问权限检查信息。这样，
即使重写后的查询没有直接使用视图，执行器仍会检查用户是否具有访问该视图所需的权限。

这就是应用的第一条规则。
规则系统接着会检查顶层查询中剩余的范围表项（本例中已经没有了），
并递归检查新增子查询中的范围表项，看它们是否引用了视图。（但它不会展开old或new，
否则就会出现无限递归！）在这个例子中，shoelace_data和unit都没有重写规则，
因此重写到此结束，上面的结果就是交给规划器的最终结果。

现在我们想写一个查询，找出商店里目前哪些鞋子有颜色和长度都匹配的鞋带，
并且完全匹配的总双数大于等于二。

```
SELECT * FROM shoe_ready WHERE total_avail >= 2;

 shoename | sh_avail | sl_name | sl_avail | total_avail
----------+----------+---------+----------+-------------
 sh1      |        2 | sl1     |        5 |           2
 sh3      |        4 | sl7     |        7 |           4
(2 rows)
```

这一次解析器的输出是如下查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM shoe_ready shoe_ready
 WHERE shoe_ready.total_avail >= 2;
```

首先应用的是用于 shoe_ready 视图的规则，它得到如下查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM (SELECT rsh.shoename,
               rsh.sh_avail,
               rsl.sl_name,
               rsl.sl_avail,
               least(rsh.sh_avail, rsl.sl_avail) AS total_avail
          FROM shoe rsh, shoelace rsl
         WHERE rsl.sl_color = rsh.slcolor
           AND rsl.sl_len_cm >= rsh.slminlen_cm
           AND rsl.sl_len_cm <= rsh.slmaxlen_cm) shoe_ready
 WHERE shoe_ready.total_avail >= 2;
```

类似地，shoe 和 shoelace 的规则会被替换进子查询的范围表中，最终得到一棵三层查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM (SELECT rsh.shoename,
               rsh.sh_avail,
               rsl.sl_name,
               rsl.sl_avail,
               least(rsh.sh_avail, rsl.sl_avail) AS total_avail
          FROM (SELECT sh.shoename,
                       sh.sh_avail,
                       sh.slcolor,
                       sh.slminlen,
                       sh.slminlen * un.un_fact AS slminlen_cm,
                       sh.slmaxlen,
                       sh.slmaxlen * un.un_fact AS slmaxlen_cm,
                       sh.slunit
                  FROM shoe_data sh, unit un
                 WHERE sh.slunit = un.un_name) rsh,
               (SELECT s.sl_name,
                       s.sl_avail,
                       s.sl_color,
                       s.sl_len,
                       s.sl_unit,
                       s.sl_len * u.un_fact AS sl_len_cm
                  FROM shoelace_data s, unit u
                 WHERE s.sl_unit = u.un_name) rsl
         WHERE rsl.sl_color = rsh.slcolor
           AND rsl.sl_len_cm >= rsh.slminlen_cm
           AND rsl.sl_len_cm <= rsh.slmaxlen_cm) shoe_ready
 WHERE shoe_ready.total_avail > 2;
```

事实证明，规划器会把这棵树折叠成一个两层的查询树：最底层的 SELECT 命令会被“上拉”到中间的 SELECT 中，因为没有必要单独处理它们。但中间的 SELECT 会保持与顶层分离，因为它包含聚合函数。如果把它们也上拉，就会改变最顶层 SELECT 的行为，而这不是我们想要的。不过，折叠查询树是一种优化，重写系统不必关心它。

37.2.2. 非 SELECT 语句中的视图规则

上文对视图规则的说明没有涉及查询树中的两个细节：命令类型和结果关系。实际上，视图规则并不需要命令类型，但结果关系可能会影响查询重写器的工作方式，因为当结果关系是视图时，必须进行特殊处理。

SELECT的查询树与其他任何命令的查询树之间只有少数差别。显然，它们的命令类型不同；对于SELECT之外的命令，结果关系会指向结果应写入的那个范围表项。除此之外，其余部分完全相同。因此，假设有两个表t1和t2，都具有列a和b，那么下面两条语句的查询树：

```
SELECT t2.b FROM t1, t2 WHERE t1.a = t2.a;

UPDATE t1 SET b = t2.b FROM t2 WHERE t1.a = t2.a;
```

几乎是一样的。特别是：

- 范围表中包含表t1和t2的项。
- 目标列表中都包含一个变量，该变量指向表t2的范围表项中的列b。

- 条件表达式会比较两个范围表项中的列a是否相等。
- 连接树都表示t1与t2之间的一次简单连接。

结果是，这两个查询树都会产生相似的执行计划：它们都是对这两个表的连接。对于UPDATE，规划器会把t1中缺失的列补入目标列表，最终查询树会变成：

```
UPDATE t1 SET a = t1.a, b = t2.b FROM t2 WHERE t1.a = t2.a;
```

因而，执行器在这个连接上运行时会产生与下面语句完全相同的结果集：

```
SELECT t1.a, t2.b FROM t1, t2 WHERE t1.a = t2.a;
```

但在UPDATE中有个小问题：执行器计划中负责连接的那一部分并不关心连接结果将被用于什么。它只是生成一个行结果集。一个是SELECT命令，另一个是UPDATE命令，这一差别是在执行器更高层处理的；在那里，系统知道这是一个UPDATE，也知道结果应写入表t1。但问题在于：其中哪一行应当被新行替换？

为了解决这个问题，UPDATE（以及DELETE）语句的目标列表中会额外加入一项：当前元组ID（CTID）。这是一个系统列，包含该行所在的文件块号以及在块中的位置。已知表之后，就可以利用CTID取回要更新的t1原始行。把CTID加入目标列表后，查询实际上会变成：

```
SELECT t1.a, t2.b, t1.ctid FROM t1, t2 WHERE t1.a = t2.a;
```

现在还要引入PostgreSQL的另一个细节。旧表行不会被覆盖，这也是ROLLBACK之所以很快的原因。在UPDATE中，新结果行会被插入表中（去掉CTID之后），而CTID所指向的旧行，其行头中的cmax和xmax项会被设置为当前命令计数器和当前事务 ID。于是旧行被隐藏起来，事务提交后，清理器最终就可以移除这条死行。

知道了这些以后，我们就可以用完全相同的方式把视图规则应用到任何命令上。没有区别。

37.2.3. PostgreSQL 中视图的威力

上文演示了规则系统如何把视图定义整合进原始查询树。在第二个示例中，从一个视图发出的简单SELECT最终生成了一棵四表连接的查询树（unit以不同名称使用了两次）。

用规则系统实现视图的好处在于：规划器能够在单棵查询树中同时看到哪些表必须被扫描、这些表之间的关系、来自视图的限制条件，以及原始查询自身的条件。即使原始查询本身已经是对若干视图的连接，情况也仍然如此。规划器必须决定执行查询的最佳路径，而它掌握的信息越多，这个决定通常就越好。PostgreSQL实现的规则系统能够保证，到那个阶段为止，关于该查询的全部可用信息都已经集中在这里。

37.2.4. 更新视图

如果把一个视图命名为 INSERT、UPDATE 或 DELETE 的目标关系会发生什么？简单地做上述替换会得到一个结果关系指向子查询范围表项的查询树，这是行不通的。相反，重写器假定该操作将由视图上的一个 INSTEAD OF 触发器处理。（如果没有这样的触发器，执行器会在开始执行时抛出一个错误。）在这种情形下，重写的工作方式略有不同。对于 INSERT，重写器完全不处理该视图，而是让它继续作为查询的结果关系。对于 UPDATE 和 DELETE，仍然需要展开视图查询以产生 the “old” rows that the command will attempt to update or delete. So the view is expanded as normal, but another unexpanded range-table entry is added to the query to represent the view in its capacity as the result relation.

此时出现的问题是，如何标识视图中需要更新的行。回想一下，当结果关系是表时，目标列表中会加入一个特殊的CTID项，用来标识待更新行的物理位置。如果结果关系是视图，这就行不通了，因为视图没有CTID，它的行并没有实际物理位置。取而代之的是，对于UPDATE或DELETE操作，目标列表中会加入一个特殊的wholerow项，它会展开为该视图的全部列。执行器利用这个值把“旧”行传给INSTEAD OF触发器。之后由触发器根据旧行值和新行值决定应当更新什么。

如果没有更新视图的 INSTEAD OF 触发器，执行器将抛出一个错误，因为它自己无法自动更新视图。要改变这一点，我们可以定义规则来修改视图上 INSERT、UPDATE 和 DELETE 命令的行为。这些规则会重写命令，通常是把它改写成更新一个或多个表而不是视图的命令。这正是下一节的主题。

注意，规则总是先被处理，也就是在原始查询被规划和执行之前先完成重写。因此，如果一个视图上同时有INSTEAD OF触发器以及INSERT、UPDATE或DELETE规则，那么规则会先执行；根据其结果，触发器甚至可能完全不会被用到。

37.3. INSERT、UPDATE和DELETE规则

定义在INSERT、UPDATE和DELETE上的规则与前几节描述的视图规则有明显的不同。首先，它们的CREATE RULE命令允许更多：

- 它们可以没有动作。
- 它们可以有多个动作。
- 它们可以是INSTEAD或ALSO（缺省）。
- 伪关系NEW和OLD可以派上用场。
- 它们可以有规则条件。

第二，它们不是就地修改查询树，而是创建零棵或多棵新查询树，并且可能丢弃原始查询树。

37.3.1. 更新规则如何工作

记住以下语法：

```
CREATE [ OR REPLACE ] RULE name AS ON event
    TO table [ WHERE condition ]
    DO [ ALSO | INSTEAD ] { NOTHING | command | ( command ; command .
.. ) }
```

在后文中，更新规则是指定义在INSERT、UPDATE或DELETE上的规则。

当查询树的结果关系和命令类型分别等于CREATE RULE命令中给出的对象和事件时，规则系统就会应用更新规则。对于更新规则，规则系统会创建一个查询树列表，初始时该列表为空。动作可以有零个（关键字NOTHING）、一个或多个。为简化说明，我们先看只有一个动作的规则。这个规则可以有条件，也可以没有条件；它还可以是INSTEAD或ALSO（默认值）。

什么是规则条件？它是一种限制，用来说明何时执行规则动作、何时不执行。这个条件只能引用伪关系NEW和/或OLD，它们基本上表示作为对象给出的那个关系，只是带有特殊含义。

因此，对于只有一个动作的规则，有以下三种情况会产生相应的查询树。

没有条件，有ALSO或INSTEAD

规则动作的查询树，再附加原始查询树的条件

给出了条件，有ALSO

规则动作的查询树，再附加规则条件和原始查询树的条件

给出了条件，有INSTEAD

规则动作的查询树，再附加规则条件和原始查询树的条件；
以及附加了规则条件取反后的原始查询树

最后，如果规则是 ALSO，就把未经更改的原始查询树加入列表。由于只有带条件的 INSTEAD 规则已经加入了原始查询树，因此，对于只有一个动作的规则，最终会得到一棵或两棵输出查询树。

对于ON INSERT规则，原始查询（如果未被INSTEAD抑制）会先于规则添加的任何动作执行。这样一来，这些动作就能看到被插入的行。但对ON UPDATE和ON DELETE规则，原始查询会在规则添加的动作之后执行。这就保证了这些动作能够看到将被更新或删除的行；否则，
动作可能什么也做不了，因为它们找不到符合条件的行。

由规则动作生成的查询树会再次送入重写系统，随后可能还会有更多规则被应用，
从而产生更多或更少的查询树。因此，一个规则的动作必须具有不同的命令类型，
或者具有与该规则自身不同的结果关系。否则，这种递归过程就会陷入无限循环。
（规则的递归展开会被检测出来，并作为错误报告。）

pg_rewrite系统目录中动作里的查询树只是模板。由于它们可以引用NEW和OLD的范围表项，
因此在使用前必须进行一些替换。对任何NEW的引用，
都会先在原始查询的目标列表中查找相应项。如果找到了，就用该项的表达式替换该引用。
否则，NEW就与OLD含义相同（对于UPDATE），或者被替换为一个空值（对于INSERT）。
任何对OLD的引用，都会被替换为对结果关系对应范围表项的引用。

在系统完成应用更新规则后，它再应用视图规则到生成的查询树上。
视图无法插入新的更新动作，所以没有必要向视图重写的输出应用更新规则。

37.3.1.1. 第一个规则：逐步分析

假设我们想跟踪shoelace_data关系中sl_avail列的变化。因此，
我们建立一个日志表和一条规则，使其在shoelace_data上执行UPDATE时，
有条件地写入一条日志记录。

```
CREATE TABLE shoelace_log (
    sl_name    text,          -- shoelace changed
    sl_avail   integer,       -- new available value
    log_who    text,          -- who did it
    log_when   timestamp      -- when
);

CREATE RULE log_shoelace AS ON UPDATE TO shoelace_data
    WHERE NEW.sl_avail <> OLD.sl_avail
    DO INSERT INTO shoelace_log VALUES (
                                    NEW.sl_name,
                                    NEW.sl_avail,
                                    current_user,
```

```
                                    current_timestamp
                                );
```

现在有人执行了：

```
UPDATE shoelace_data SET sl_avail = 6 WHERE sl_name = 'sl7';
```

然后我们查看日志表：

```
SELECT * FROM shoelace_log;

 sl_name | sl_avail | log_who | log_when
---------+----------+---------+----------------------------------
 sl7     |        6 | Al      | Tue Oct 20 16:14:45 1998 MET DST
(1 row)
```

这正是我们预期的结果。后台发生的事情如下。解析器创建了查询树：

```
UPDATE shoelace_data SET sl_avail = 6
  FROM shoelace_data shoelace_data
 WHERE shoelace_data.sl_name = 'sl7';
```

此时存在一条ON UPDATE规则log_shoelace，它带有如下规则条件表达式：

```
NEW.sl_avail <> OLD.sl_avail
```

其动作是：

```
INSERT INTO shoelace_log VALUES (
       new.sl_name, new.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old;
```

（这看起来有点奇怪，因为通常你不能写INSERT ... VALUES ... FROM。这里的FROM子句只是为了表明查询树中存在用于new和old的范围表项。之所以需要这些项，是为了让INSERT命令查询树中的变量能够引用它们。）

该规则是一条带条件的ALSO规则，因此规则系统必须返回两棵查询树：修改后的规则动作，以及原始查询树。第 1 步中，原始查询的范围表会并入规则动作的查询树，得到：

```
INSERT INTO shoelace_log VALUES (
       new.sl_name, new.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old,
       shoelace_data shoelace_data;
```

第 2 步将规则条件加进去，因此结果集被限制为sl_avail发生变化的行：

```
INSERT INTO shoelace_log VALUES (
       new.sl_name, new.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old,
       shoelace_data shoelace_data
 WHERE new.sl_avail <> old.sl_avail;
```

（这看起来更奇怪，因为INSERT ... VALUES同样没有WHERE子句，但规划器和执行器处理它并无困难。反正它们本来也需要为INSERT ... SELECT支持相同功能。）

第 3 步把原始查询树的条件加进去，把结果集进一步限制为只有那些原始查询会触及的行：

```
INSERT INTO shoelace_log VALUES (
       new.sl_name, new.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old,
       shoelace_data shoelace_data
 WHERE new.sl_avail <> old.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

第 4 步把NEW引用替换为原始查询树中的目标列表项，或者替换为结果关系中相应的变量引用：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old,
       shoelace_data shoelace_data
 WHERE 6 <> old.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

第 5 步把OLD引用替换为结果关系引用：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data new, shoelace_data old,
       shoelace_data shoelace_data
 WHERE 6 <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

至此就完成了。由于规则是ALSO，我们还要输出原始查询树。简而言之，
规则系统输出的是一个包含两棵查询树的列表，它们对应于以下语句：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data
 WHERE 6 <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';

UPDATE shoelace_data SET sl_avail = 6
 WHERE sl_name = 'sl7';
```

它们会按这个顺序执行，而这正是该规则想要达到的效果。

上述替换以及附加的条件能够保证：如果原始查询是下面这样，就不会写入任何日志记录：

```
UPDATE shoelace_data SET sl_color = 'green'
 WHERE sl_name = 'sl7';
```

在这种情况下，原始查询树不包含sl_avail的目标列表项，因此NEW.sl_avail
会被shoelace_data.sl_avail替换。于是，规则生成的额外命令是：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, shoelace_data.sl_avail,
       current_user, current_timestamp )
```

```
  FROM shoelace_data
 WHERE shoelace_data.sl_avail <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

而该条件永远不可能为真。

如果原始查询修改多行，这种机制同样能够正常工作。例如，假设有人发出如下命令：

```
UPDATE shoelace_data SET sl_avail = 0
 WHERE sl_color = 'black';
```

实际上会更新四行（sl1、sl2、sl3和sl4）。但sl3本来就已经是sl_avail = 0。在这种情况下，原始查询树的条件不同，因此规则会产生额外的查询树：

```
INSERT INTO shoelace_log
SELECT shoelace_data.sl_name, 0,
       current_user, current_timestamp
  FROM shoelace_data
 WHERE 0 <> shoelace_data.sl_avail
   AND shoelace_data.sl_color = 'black';
```

这棵查询树必然会插入三条新的日志记录。这完全正确。

到这里就能看出，为什么原始查询树最后执行至关重要。如果先执行UPDATE，那么所有行都已经被设为零，记日志的INSERT就找不到任何满足0 <> shoelace_data.sl_avail的行了。

37.3.2. 与视图的协作

为了防止有人像前面提到的那样对视图关系执行INSERT、UPDATE或DELETE，一种简单的办法是让那些查询树直接被丢弃。因此我们可以创建如下规则：

```
CREATE RULE shoe_ins_protect AS ON INSERT TO shoe
    DO INSTEAD NOTHING;
CREATE RULE shoe_upd_protect AS ON UPDATE TO shoe
    DO INSTEAD NOTHING;
CREATE RULE shoe_del_protect AS ON DELETE TO shoe
    DO INSTEAD NOTHING;
```

如果现在某人尝试对视图关系shoe执行这些操作，规则系统就会应用这些规则。由于这些规则没有动作而且是INSTEAD，生成的查询树列表将为空；整个查询也就什么都不会做，因为规则系统处理完后已经没有任何内容可供优化或执行。

另一种更完善的做法，是创建一些规则，把查询树重写成在真实表上执行正确操作的查询树。要在视图shoelace上做到这一点，我们创建下列规则：

```
CREATE RULE shoelace_ins AS ON INSERT TO shoelace
    DO INSTEAD
    INSERT INTO shoelace_data VALUES (
           NEW.sl_name,
           NEW.sl_avail,
           NEW.sl_color,
           NEW.sl_len,
           NEW.sl_unit
    );
```

```
CREATE RULE shoelace_upd AS ON UPDATE TO shoelace
    DO INSTEAD
    UPDATE shoelace_data
       SET sl_name = NEW.sl_name,
           sl_avail = NEW.sl_avail,
           sl_color = NEW.sl_color,
           sl_len = NEW.sl_len,
           sl_unit = NEW.sl_unit
     WHERE sl_name = OLD.sl_name;

CREATE RULE shoelace_del AS ON DELETE TO shoelace
    DO INSTEAD
    DELETE FROM shoelace_data
     WHERE sl_name = OLD.sl_name;
```

如果你想在视图上支持RETURNING查询，就需要让规则包含计算视图行的RETURNING子句。对于单表视图，这通常很简单；但对于像shoelace这样的连接视图，就会有点烦琐。插入场景的一个示例如下：

```
CREATE RULE shoelace_ins AS ON INSERT TO shoelace
    DO INSTEAD
    INSERT INTO shoelace_data VALUES (
           NEW.sl_name,
           NEW.sl_avail,
           NEW.sl_color,
           NEW.sl_len,
           NEW.sl_unit
    )
    RETURNING
           shoelace_data.*,
           (SELECT shoelace_data.sl_len * u.un_fact
            FROM unit u WHERE shoelace_data.sl_unit = u.un_name);
```

请注意，这一条规则同时支持该视图上的INSERT和INSERT RETURNING查询；对于普通INSERT，RETURNING子句会被简单忽略。

现在假定商店每隔一段时间就会收到一包鞋带，随附一份很长的部件清单。但你不想每次都手工更新 shoelace 视图。因此，我们建立两个小表：一个用于插入部件清单中的项目，另一个则采用一个特殊技巧。创建命令如下：

```
CREATE TABLE shoelace_arrive (
    arr_name    text,
    arr_quant   integer
);

CREATE TABLE shoelace_ok (
    ok_name     text,
    ok_quant    integer
);

CREATE RULE shoelace_ok_ins AS ON INSERT TO shoelace_ok
    DO INSTEAD
    UPDATE shoelace
       SET sl_avail = sl_avail + NEW.ok_quant
     WHERE sl_name = NEW.ok_name;
```

现在可以将部件清单中的数据填入 shoelace_arrive 表：

```
SELECT * FROM shoelace_arrive;

 arr_name | arr_quant
----------+-----------
 sl3      |        10
 sl6      |        20
 sl8      |        20
(3 rows)
```

快速查看一下当前数据：

```
SELECT * FROM shoelace;

 sl_name  | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
----------+----------+----------+--------+---------+-----------
 sl1      |        5 | black    |     80 | cm      |        80
 sl2      |        6 | black    |    100 | cm      |       100
 sl7      |        6 | brown    |     60 | cm      |        60
 sl3      |        0 | black    |     35 | inch    |      88.9
 sl4      |        8 | black    |     40 | inch    |     101.6
 sl8      |        1 | brown    |     40 | inch    |     101.6
 sl5      |        4 | brown    |      1 | m       |       100
 sl6      |        0 | brown    |    0.9 | m       |        90
(8 rows)
```

现在把到货的鞋带入库：

```
INSERT INTO shoelace_ok SELECT * FROM shoelace_arrive;
```

然后检查结果：

```
SELECT * FROM shoelace ORDER BY sl_name;

 sl_name  | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
----------+----------+----------+--------+---------+-----------
 sl1      |        5 | black    |     80 | cm      |        80
 sl2      |        6 | black    |    100 | cm      |       100
 sl7      |        6 | brown    |     60 | cm      |        60
 sl4      |        8 | black    |     40 | inch    |     101.6
 sl3      |       10 | black    |     35 | inch    |      88.9
 sl8      |       21 | brown    |     40 | inch    |     101.6
 sl5      |        4 | brown    |      1 | m       |       100
 sl6      |       20 | brown    |    0.9 | m       |        90
(8 rows)

SELECT * FROM shoelace_log;

 sl_name | sl_avail | log_who| log_when
---------+----------+--------+----------------------------------
 sl7     |        6 | Al     | Tue Oct 20 19:14:45 1998 MET DST
 sl3     |       10 | Al     | Tue Oct 20 19:25:16 1998 MET DST
 sl6     |       20 | Al     | Tue Oct 20 19:25:16 1998 MET DST
 sl8     |       21 | Al     | Tue Oct 20 19:25:16 1998 MET DST
```

```
(4 rows)
```

从一个INSERT ... SELECT到这些结果，中间要经历相当长的一段过程。本章最后再来描述这一查询树转换的细节。首先，看一下解析器的输出：

```
INSERT INTO shoelace_ok
SELECT shoelace_arrive.arr_name, shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok;
```

现在应用第一条规则shoelace_ok_ins，它会把这一输出转换成：

```
UPDATE shoelace
   SET sl_avail = shoelace.sl_avail + shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace
 WHERE shoelace.sl_name = shoelace_arrive.arr_name;
```

同时会丢弃针对shoelace_ok的原始INSERT。这个重写后的查询会再次交给规则系统，而接下来应用的规则shoelace_upd会生成：

```
UPDATE shoelace_data
   SET sl_name = shoelace.sl_name,
       sl_avail = shoelace.sl_avail + shoelace_arrive.arr_quant,
       sl_color = shoelace.sl_color,
       sl_len = shoelace.sl_len,
       sl_unit = shoelace.sl_unit
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace old,
       shoelace new, shoelace_data shoelace_data
 WHERE shoelace.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = shoelace.sl_name;
```

这同样是一条INSTEAD规则，因此前一个查询树会被丢弃。注意，这个查询仍然使用视图shoelace。但规则系统尚未完成，因此会继续在其上应用_RETURN规则，于是得到：

```
UPDATE shoelace_data
   SET sl_name = s.sl_name,
       sl_avail = s.sl_avail + shoelace_arrive.arr_quant,
       sl_color = s.sl_color,
       sl_len = s.sl_len,
       sl_unit = s.sl_unit
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace old,
       shoelace new, shoelace_data shoelace_data,
       shoelace old, shoelace new,
       shoelace_data s, unit u
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name;
```

最后，规则log_shoelace被应用，生成额外的查询树：

```
INSERT INTO shoelace_log
```

```
SELECT s.sl_name,
       s.sl_avail + shoelace_arrive.arr_quant,
       current_user,
       current_timestamp
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace old,
       shoelace new, shoelace_data shoelace_data,
       shoelace old, shoelace new,
       shoelace_data s, unit u,
       shoelace_data old, shoelace_data new
       shoelace_log shoelace_log
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name
   AND (s.sl_avail + shoelace_arrive.arr_quant) <> s.sl_avail;
```

到这里，规则系统已经没有更多规则可用，并返回生成的查询树。

因此，我们最终得到两棵查询树，它们等效于以下SQL语句：

```
INSERT INTO shoelace_log
SELECT s.sl_name,
       s.sl_avail + shoelace_arrive.arr_quant,
       current_user,
       current_timestamp
  FROM shoelace_arrive shoelace_arrive, shoelace_data shoelace_data,
       shoelace_data s
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name
   AND s.sl_avail + shoelace_arrive.arr_quant <> s.sl_avail;

UPDATE shoelace_data
   SET sl_avail = shoelace_data.sl_avail + shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive,
       shoelace_data shoelace_data,
       shoelace_data s
 WHERE s.sl_name = shoelace_arrive.sl_name
   AND shoelace_data.sl_name = s.sl_name;
```

结果是：来自一个关系的数据被插入到另一个关系中，这个插入被改写为对第三个关系的更新，再被改写为对第四个关系的更新外加在第五个关系中记录该更新，
最后整个过程被化简为两个查询。

这里有个稍显难看的小细节。观察这两个查询会发现，`shoelace_data`
关系在范围表中出现了两次，而实际上完全可以缩减成一次。规划器不会处理这一点，
因此规则系统为`INSERT`输出的执行计划会是

```
Nested Loop
  ->  Merge Join
        ->  Seq Scan
              ->  Sort
                    ->  Seq Scan on s
        ->  Seq Scan
              ->  Sort
                    ->  Seq Scan on shoelace_arrive
```

```
  ->  Seq Scan on shoelace_data
```

而省略那个额外的范围表项则会得到

```
Merge Join
  ->  Seq Scan
        ->  Sort
              ->  Seq Scan on s
  ->  Seq Scan
        ->  Sort
              ->  Seq Scan on shoelace_arrive
```

这会在日志表中生成完全相同的项。因此，规则系统导致了对shoelace_data表的一次完全不必要的额外扫描。而在UPDATE中，同样的冗余扫描还会再发生一次。不过，能让这一切总体上工作起来，已经是一项相当艰巨的工作。

现在我们来做一个关于PostgreSQL规则系统及其威力的最后演示。
假设你向数据库中加入了一些颜色特别的鞋带：

```
INSERT INTO shoelace VALUES ('sl9', 0, 'pink', 35.0, 'inch', 0.0);
INSERT INTO shoelace VALUES ('sl10', 1000, 'magenta', 40.0, 'inch', 0.
0);
```

我们想建立一个视图，用来检查哪些shoelace项在颜色上与任何鞋子都不匹配。这个视图是：

```
CREATE VIEW shoelace_mismatch AS
    SELECT * FROM shoelace WHERE NOT EXISTS
        (SELECT shoename FROM shoe WHERE slcolor = sl_color);
```

它的输出是：

```
SELECT * FROM shoelace_mismatch;

 sl_name | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
---------+----------+----------+--------+---------+-----------
 sl9     |        0 | pink     |     35 | inch    |      88.9
 sl10    |     1000 | magenta  |     40 | inch    |     101.6
```

现在希望将没有库存的不匹配鞋带从数据库中删除。为了给 PostgreSQL 增加一点难度，我们不直接删除，而是再创建一个视图：

```
CREATE VIEW shoelace_can_delete AS
    SELECT * FROM shoelace_mismatch WHERE sl_avail = 0;
```

然后这样执行：

```
DELETE FROM shoelace WHERE EXISTS
    (SELECT * FROM shoelace_can_delete
             WHERE sl_name = shoelace.sl_name);
```

Voilà：

```
SELECT * FROM shoelace;

 sl_name | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
---------+----------+----------+--------+---------+-----------
 sl1     |        5 | black    |     80 | cm      |        80
```

```
 sl2     |        6 | black    |    100 | cm      |       100
 sl7     |        6 | brown    |     60 | cm      |        60
 sl4     |        8 | black    |     40 | inch    |     101.6
 sl3     |       10 | black    |     35 | inch    |      88.9
 sl8     |       21 | brown    |     40 | inch    |     101.6
 sl10    |     1000 | magenta  |     40 | inch    |     101.6
 sl5     |        4 | brown    |      1 | m       |       100
 sl6     |       20 | brown    |    0.9 | m       |        90
(9 rows)
```

作用于一个视图上的DELETE，其子查询条件总共使用了四个嵌套/连接的视图，
其中一个视图自身又带有一个包含视图的子查询条件，并且还用到了计算得到的视图列，
最终仍会被重写成单棵查询树，从真正的表中删除所请求的数据。

在现实世界中，大概只有很少的场景会需要这样的构造。但知道它确实能工作，总归让人安心。

37.4. 规则和权限

由于PostgreSQL规则系统会重写查询，因此可能访问原始查询中并未使用的其他表或视图。
使用更新规则时，这甚至可能包括对表的写访问。

重写规则没有单独的所有者。关系（表或视图）
的所有者会自动成为为其定义的重写规则的所有者。
PostgreSQL规则系统改变了默认访问控制系统的行为。
由于规则而使用的关系都会根据规则所有者的权限进行检查，而不是调用规则的用户。
这意味着用户只需要对其查询中明确命名的表/视图具有所需的权限。

例如，某个用户有一份电话号码列表，其中一部分是私人的，另一部分对办公室秘书有用。
该用户可以这样构造：

```
CREATE TABLE phone_data (person text, phone text, private boolean);
CREATE VIEW phone_number AS
    SELECT person, CASE WHEN NOT private THEN phone END AS phone
    FROM phone_data;
GRANT SELECT ON phone_number TO assistant;
```

除了该用户本人（以及数据库超级用户）之外，没有人可以访问phone_data表。但由于GRANT
的存在，秘书可以对phone_number视图执行SELECT。规则系统会把对phone_number
的SELECT重写成对phone_data的SELECT。由于该用户是phone_number的所有者，
因此也是规则的所有者，对phone_data的读访问会按照该用户的权限进行检查，
于是查询被允许。同时，对phone_number本身的访问检查仍会执行，
但这是针对调用用户进行的，因此除了用户本人和秘书之外，没有其他人能使用它。

权限是逐条规则检查的。因此，目前秘书是唯一能看到公开电话号码的人。
但秘书还可以再建立一个视图，并把该视图授予公众访问。这样，
任何人都可以通过秘书的视图看到phone_number中的数据。
秘书做不到的是创建一个直接访问phone_data的视图。（其实秘书可以创建，
但它不会起作用，因为每次访问都会在权限检查时被拒绝。）而且，一旦用户发现秘书开放了其
phone_number视图，用户就可以撤销秘书的访问权限。那样一来，
对秘书视图的任何访问都会立即失败。

这种逐条规则的检查看起来似乎像是一个安全漏洞，但实际上并不是。因为即使不这样工作，
秘书也可以建立一个与phone_number具有相同列的表，每天把数据复制进去。那样一来，
这些就是秘书自己的数据，秘书同样可以把访问权限授予任何人。GRANT

的含义本来就是“我信任你”。如果某个你信任的人做了上面的事，那么该重新考虑这份信任，并使用REVOKE了。

需要注意的是，虽然视图可以用上面展示的技术来隐藏某些列的内容，但它们不能被用来可靠地隐藏不可见行中的数据。例如，下面这个视图是不安全的：

```
CREATE VIEW phone_number AS
    SELECT person, phone FROM phone_data WHERE phone NOT LIKE '412%';
```

This view might seem secure, since the rule system will rewrite any SELECT from phone_number into a SELECT from phone_data and add the qualification that only entries where phone does not begin with 412 are wanted. But if the user can create his or her own functions, it is not difficult to convince the planner to execute the user-defined function prior to the NOT LIKE expression. For example:

```
CREATE FUNCTION tricky(text, text) RETURNS bool AS $$
BEGIN
    RAISE NOTICE '% => %', $1, $2;
    RETURN true;
END
$$ LANGUAGE plpgsql COST 0.0000000000000000000001;

SELECT * FROM phone_number WHERE tricky(person, phone);
```

Every person and phone number in the phone_data table will be printed as a NOTICE, because the planner will choose to execute the inexpensive tricky function before the more expensive NOT LIKE. Even if the user is prevented from defining new functions, built-in functions can be used in similar attacks. (For example, most casting functions include their input values in the error messages they produce.)

类似的考虑也适用于更新规则。在上一节的示例中，示例数据库中那些表的所有者可以把shoelace视图上的SELECT、INSERT、UPDATE和DELETE权限授予其他用户，但对shoelace_log只授予SELECT权限。写日志记录的规则动作仍会成功执行，因此其他用户可以看到日志记录。但他们不能伪造记录，也不能操纵或删除已有记录。在这个例子里，不存在通过说服规划器改变操作顺序来破坏规则的可能性，因为唯一引用shoelace_log的规则是一条无条件的INSERT。在更复杂的场景中，情况未必如此。

37.5. 规则和命令状态

PostgreSQL服务器会为收到的每条命令返回一个命令状态字符串，例如INSERT 149592 1。在没有规则参与时，这很简单；但当查询被规则重写时会发生什么呢？

规则对命令状态的影响如下：

- 如果该查询没有无条件的INSTEAD规则，那么原始给出的查询就会照常执行，并返回它自身的命令状态。（但请注意，如果存在任何带条件的INSTEAD规则，那么这些规则条件的取反会被加到原始查询上。这可能会减少它处理的行数，从而影响报告的状态。）
- 如果该查询存在任意无条件的INSTEAD规则，那么原始查询将完全不会执行。在这种情况下，服务器会返回由INSTEAD规则（带条件或不带条件）插入的、最后一条且与原始查询具有相同命令类型（INSERT、UPDATE或DELETE）的查询的命令状态。如果没有任何规则添加出满足这些条件的查询，那么返回的命令状态会显示原始查询类型，并且把行计数和 OID 字段都置零。

(This system was established in PostgreSQL 7.3. In versions before that, the command status might show different results when rules exist.)

在第二种情况下，程序员可以通过为所需的INSTEAD
规则指定一个在活动规则中按字母顺序排在最后的规则名，来确保由它设置命令状态，
因为它会最后被应用。

37.6. 规则与触发器

许多能够用触发器完成的事情，同样也可以用PostgreSQL规则系统实现。
不能用规则实现的内容之一是某些约束，尤其是外键。
你可以放置一条带条件的规则：当某列的值没有出现在另一张表中时，把命令重写成NOTHING。
但那样做会悄无声息地丢弃数据，这并不是好主意。如果需要检查值的有效性，
并在值无效时生成错误消息，就必须使用触发器。

在本章中，我们重点讨论了用规则来更新视图。本章中所有更新规则的示例，
也都可以用视图上的INSTEAD OF触发器来实现。编写这类触发器通常比编写规则更容易，尤其是在需要用复杂逻辑执行更新时。

对于两者都能实现的事情，哪一种更好取决于数据库的使用方式。
触发器会对每个受影响的行触发一次，而规则会修改查询，或生成额外的查询。因此，
如果一条语句会影响很多行，那么发出一条额外命令的规则往往会比触发器更快，
因为触发器必须为每一行调用一次，并反复判断该做什么。不过，
触发器方法在概念上要比规则方法简单得多，也更容易让新手写对。

下面通过一个例子说明，在一种具体情形中选择规则或触发器会有什么不同。这里有两个表：

```
CREATE TABLE computer (
    hostname        text,    -- indexed
    manufacturer    text     -- indexed
);

CREATE TABLE software (
    software        text,    -- indexed
    hostname        text     -- indexed
);
```

两个表都有数千行，并且 hostname 上的索引都是唯一索引。规则或触发器应实现这样一个约束：从 software 中删除引用了已删除计算机的行。触发器会使用以下命令：

```
DELETE FROM software WHERE hostname = $1;
```

由于从 computer 中删除的每一行都会调用触发器，因此触发器可以为这条命令准备并保存执行计划，并通过参数传入 hostname 值。规则则会写成：

```
CREATE RULE computer_del AS ON DELETE TO computer
    DO DELETE FROM software WHERE hostname = OLD.hostname;
```

现在我们来看不同类型的删除。对于下面这种情况：

```
DELETE FROM computer WHERE hostname = 'mypc.local.net';
```

computer表会通过索引扫描（很快），由触发器发出的命令也会使用索引扫描（同样很快）。
规则产生的额外查询是：

```
DELETE FROM software WHERE computer.hostname = 'mypc.local.net'
```

```
                        AND software.hostname = computer.hostname;
```

由于已经建立了合适的索引，规划器将创建一个规划

```
Nestloop
  ->  Index Scan using comp_hostidx on computer
  ->  Index Scan using soft_hostidx on software
```

因此，触发器实现和规则实现之间的速度差异不会太大。

在下一个删除场景中，我们想去掉全部 2000 台hostname以old开头的计算机。有两种命令可以做到这件事。其中一种是：

```
DELETE FROM computer WHERE hostname >= 'old'
                        AND hostname <  'ole'
```

规则添加出来的命令将是：

```
DELETE FROM software WHERE computer.hostname >= 'old' AND computer.
hostname < 'ole'
                        AND software.hostname = computer.hostname;
```

其计划为：

```
Hash Join
  ->  Seq Scan on software
  ->  Hash
    ->  Index Scan using comp_hostidx on computer
```

另一个可能的命令是：

```
DELETE FROM computer WHERE hostname ~ '^old';
```

它会让规则所添加的命令得到下面这个执行计划：

```
Nestloop
  ->  Index Scan using comp_hostidx on computer
  ->  Index Scan using soft_hostidx on software
```

这说明，当多个条件表达式通过AND组合在一起时，规划器并没有意识到，对computer中hostname的限制同样也可以用于software上的索引扫描，
而在该命令的正则表达式版本里它却能做到。触发器会对需要删除的 2000 台旧计算机中的每一台各调用一次，这意味着对computer做一次索引扫描，并对software做 2000 次索引扫描。规则实现则用两条使用索引的命令完成这件事。不过，在顺序扫描这种情况下，
规则是否仍然更快，还取决于software表的总体大小。即使所有索引块很快都会进入缓存，
通过 SPI 管理器执行来自触发器的 2000 条命令仍然要耗费一些时间。

我们要看的最后一个命令是：

```
DELETE FROM computer WHERE manufacturer = 'bim';
```

同样，这也可能导致从computer中删除很多行。因此，
触发器同样会通过执行器运行很多条命令。规则生成的命令则是：

```
DELETE FROM software WHERE computer.manufacturer = 'bim'
                        AND software.hostname = computer.hostname;
```

这个命令的计划又将是在两个索引扫描上的嵌套循环，只不过使用了computer
上的另一个索引：

```
Nestloop
  ->  Index Scan using comp_manufidx on computer
  ->  Index Scan using soft_hostidx on software
```

在上述任何一种情况下，规则系统产生的额外命令都或多或少与该命令影响的行数无关。

总结来说，只有当规则动作导致了规模很大且条件很差的连接，而规划器又无能为力时，
规则才会明显慢于触发器。

第 38 章 过程语言

PostgreSQL允许使用 SQL 和 C 之外的其他语言编写用户定义的函数。这些语言统称为过程语言（PL）。对于用过程语言编写的函数，
数据库服务器并不内置关于如何解释函数源文本的知识。相反，
这项任务会交给一个了解该语言细节的专门的调用处理器。该调用处理器既可以自行完成解析、语法分析、执行等全部工作，也可以在
PostgreSQL与某种现有编程语言实现之间充当“粘合剂”。与其他任何 C 函数一样，调用处理器本身也是一个被编译进共享对象并按需装载的 C 语言函数。

标准PostgreSQL发行版当前提供四种过程语言：PL/pgSQL（第 39 章）、PL/Tcl（第 40 章）、PL/Perl（第 41 章）以及 PL/Python（第 42 章）。另有一些可用的过程语言并未包含在核心发行版中。附录 G提供了查找它们的信息。此外，用户还可以自行定义其他语言；开发新过程语言的基础知识见第 49 章。

38.1. 安装过程语言

过程语言必须在每个要使用它的数据库中“安装”。不过，安装在数据库`template1`
中的过程语言会自动在随后创建的所有数据库中可用，因为它们在`template1`中的条目会由`CREATE DATABASE`复制。因此，数据库管理员可以决定哪些数据库提供哪些语言，并且如果愿意，也可以让某些语言默认可用。

对于标准发行版附带的语言，只需执行 `CREATE EXTENSION` *`language_name`*，即可将该语言安装到当前数据库中。此外，也可以从 shell 命令行使用程序createlang来完成这项工作。例如，要把语言 PL/Perl 安装到数据库 `template1`，可以使用：

```
createlang plperl template1
```

下文所述的手工过程只建议用于安装那些尚未打包为扩展的语言。

手工安装过程语言

在数据库中安装过程语言需要五个步骤，且必须由数据库超级用户执行。多数情况下，所需的
SQL 命令都应打包成某个“扩展”的安装脚本，这样就可以用`CREATE EXTENSION`来执行它们。

1. 该语言调用处理器的共享对象必须先编译好并安装到合适的库目录中。
   这与构建和安装含有普通用户定义 C 函数的模块的方法相同；见第 35.9.6 节。该语言调用处理器往往还会依赖一个提供实际编程语言引擎的外部库；如果是这样，该库也必须安装。

2. 必须用如下命令声明该调用处理器：

   ```
   CREATE FUNCTION handler_function_name()
       RETURNS language_handler
       AS 'path-to-shared-object'
       LANGUAGE C;
   ```

   特殊返回类型`language_handler`会告诉数据库系统，该函数返回的不是某种已定义的
   SQL数据类型，并且不能在 SQL语句中直接使用。

3. (可选的) 可选地，语言调用处理器可以提供一个“内联”处理器函数，用于执行用该语言编写的匿名代码块（DO命令）。如果该语言提供了内联处理器函数，
   可用类似下面的命令声明它：

   ```
   CREATE FUNCTION inline_function_name(internal)
   ```

```
    RETURNS void
    AS 'path-to-shared-object'
    LANGUAGE C;
```

4. (可选的) 可选地，语言调用处理器可以提供一个“验证器”函数，用于在不实际执行的情况下检查函数定义是否正确。如果存在验证器函数，`CREATE FUNCTION`就会调用它。如果该语言提供了验证器函数，可用类似下面的命令声明它：

```
CREATE FUNCTION validator_function_name(oid)
    RETURNS void
    AS 'path-to-shared-object'
    LANGUAGE C STRICT;
```

5. 最后，必须用下面的命令声明该 PL：

```
CREATE [TRUSTED] [PROCEDURAL] LANGUAGE language-name
    HANDLER handler_function_name
    [INLINE inline_function_name]
    [VALIDATOR validator_function_name] ;
```

可选的关键字 `TRUSTED` 指明该语言不会授予用户本来没有的数据访问权限。可信语言是为普通数据库用户（没有超级用户权限的用户）设计的，
允许他们安全地创建函数和触发器过程。由于 PL 函数在数据库服务器内部执行，`TRUSTED` 标志只应授予那些不允许访问数据库服务器内部或文件系统的语言。语言 PL/pgSQL、PL/Tcl 和 PL/Perl 被认为是可信的；而语言 PL/TclU、PL/PerlU 和 PL/PythonU 被设计用来提供不受限制的功能，不应被标记为可信。

例 38.1展示了以 PL/Perl为例时，手工安装过程是如何进行的。

例 38.1. 手工安装 PL/Perl

下面的命令告诉数据库服务器到哪里查找 PL/Perl语言调用处理器函数的共享对象：

```
CREATE FUNCTION plperl_call_handler() RETURNS language_handler AS
    '$libdir/plperl' LANGUAGE C;
```

PL/Perl具有内联处理器函数和验证器函数，因此我们也声明它们：

```
CREATE FUNCTION plperl_inline_handler(internal) RETURNS void AS
    '$libdir/plperl' LANGUAGE C;

CREATE FUNCTION plperl_validator(oid) RETURNS void AS
    '$libdir/plperl' LANGUAGE C STRICT;
```

下面的命令：

```
CREATE TRUSTED PROCEDURAL LANGUAGE plperl
    HANDLER plperl_call_handler
    INLINE plperl_inline_handler
    VALIDATOR plperl_validator;
```

随后指定应为以下函数和触发器过程调用前面声明的函数：其语言属性为`plperl`。

在默认的PostgreSQL安装中，PL/pgSQL语言的调用处理器会被构建并安装到“库”目录中；
此外，PL/pgSQL语言本身也安装在所有数据库中。如果在构建时配置了Tcl支持，那么 PL/Tcl和 PL/TclU 的调用处理器会被构建并安装到该库目录中，但这两种语言本身默认并不会安装到任何

数据库中。同样，如果在构建时配置了 Perl 支持，就会构建并安装 PL/Perl和PL/PerlU 的调用处理器；如果在构建时配置了 Python 支持，则会安装 PL/PythonU的调用处理器，但这些语言默认也不会安装到任何数据库中。

第 39 章 PL/pgSQL – SQL 过程语言

39.1. 概述

PL/pgSQL 是一种用于 PostgreSQL 数据库系统的可载入过程语言。PL/pgSQL 的设计目标，是创建一种具备以下特性的可载入过程语言：

- 可用于创建函数和触发器函数，
- 为SQL语言增加控制结构，
- 可以执行复杂计算，
- 继承所有用户定义的类型、函数和操作符，
- 可定义为受服务器信任的语言，
- 便于使用。

用 PL/pgSQL 创建的函数，可以用于任何可以使用内置函数的地方。例如，可以创建带有复杂条件逻辑的计算函数，随后用它们定义操作符，或者把它们用于索引表达式。

在 PostgreSQL 9.0 及更高版本中，PL/pgSQL 默认安装。不过它仍然是一个可载入模块，因此对安全性特别敏感的管理员也可以选择移除它。

39.1.1. 使用PL/pgSQL的优点

SQL 是 PostgreSQL 以及大多数其他关系数据库使用的查询语言。它具有可移植性，也容易学习。但每条 SQL 语句都必须由数据库服务器单独执行。

这意味着客户端应用必须把每个查询发送给数据库服务器，等待其处理完成，接收并处理结果，做一些计算，然后再向服务器发送更多查询。这一切都会产生进程间通信开销；
如果客户端与数据库服务器不在同一台机器上，还会额外产生网络开销。

借助 PL/pgSQL，你可以把一段计算逻辑和一系列查询放在数据库服务器内部执行。这样既具备了过程语言的能力，也保留了 SQL 的易用性，同时还能显著减少客户端/服务器通信开销。

- 消除了客户端和服务器之间额外的往返通信
- 客户端不需要的中间结果不必在服务器与客户端之间编组或传输
- 可以避免多轮查询解析

与不使用存储函数的应用相比，这可能带来相当可观的性能提升。

此外，通过 PL/pgSQL 你可以使用 SQL 的全部数据类型、操作符和函数。

39.1.2. 支持的参数和结果数据类型

用 PL/pgSQL 编写的函数，可以接受服务器支持的任意标量或数组数据类型作为参数，也可以返回这些类型中的任意一种。它们还可以接受或返回任何按名称指定的复合类型（行类型）。
也可以把 PL/pgSQL 函数声明为返回 `record`，这表示结果是一种行类型，其列由调用查询中的指定内容决定，如第 7.2.1.4 节所述。

PL/pgSQL 函数还可以通过使用 `VARIADIC` 标记声明为接受可变数量的参数。这与 SQL 函数中的行为完全相同，详见第 35.4.5 节。

PL/pgSQL 函数也可以声明为接受和返回多态类型 anyelement、anyarray、anynonarray 和 anyenum。多态函数处理的实际数据类型可以在不同调用之间变化，如第 35.2.5 节所述。示例见第 39.3.1 节。

PL/pgSQL 函数还可以声明为返回任意可单独返回的数据类型的“集合”（或表）。这类函数可以通过对结果集中的每个目标元素执行 RETURN NEXT 生成输出，也可以通过 RETURN QUERY 输出某个查询的结果。

最后，如果某个 PL/pgSQL 函数没有有意义的返回值，可以将其声明为返回 void。

PL/pgSQL 函数也可以声明为使用输出参数，而不是显式指定返回类型。这并没有为该语言增加新的基本能力，但通常会更方便，尤其是在需要返回多个值时。RETURNS TABLE 记法也可以用来代替 RETURNS SETOF。

具体示例见第 39.3.1 节和第 39.6.1 节。

39.2. PL/pgSQL的结构

PL/pgSQL 是一种块结构语言。函数定义的完整文本必须是一个块。块的定义如下：

```
[ <<label>> ]
[ DECLARE
    declarations ]
BEGIN
    statements
END [ label ];
```

块中的每个声明和每条语句都以分号结束。如上所示，嵌套在其他块中的块，必须在 END 之后写分号；但是结束整个函数体的最后一个 END 不需要分号。

> **提示**
>
> 一个常见错误是在 BEGIN 后面立刻写分号。这是不正确的，并会导致语法错误。

只有在你希望通过 EXIT 语句标识某个块，或者希望用块名限定块内声明的变量名时，才需要 *label*。如果在 END 之后写了标签，它必须与块开始处的标签一致。

所有关键字都不区分大小写。标识符除非用双引号括起，否则会像普通 SQL 命令中那样被隐式转换为小写。

PL/pgSQL 代码中的注释与普通 SQL 相同。双连字符（--）开始一段持续到行尾的注释。/* 开始一段块注释，它会持续到匹配的 */。块注释可以嵌套。

块的语句部分中的任意语句都可以是一个子块。子块可用于逻辑分组，
或者把变量局部化到一小组语句中。在子块存续期间，
子块中声明的变量会遮蔽外层块中的同名变量。不过，如果使用块标签限定外层变量名，
仍然可以访问它们。例如：

```
CREATE FUNCTION somefunc() RETURNS integer AS $$
<< outerblock >>
DECLARE
    quantity integer := 30;
BEGIN
```

```
    RAISE NOTICE 'Quantity here is %', quantity;  -- Prints 30
    quantity := 50;
    --
    -- 创建一个子块
    --
    DECLARE
        quantity integer := 80;
    BEGIN
        RAISE NOTICE 'Quantity here is %', quantity;  -- Prints 80
        RAISE NOTICE 'Outer quantity here is %', outerblock.quantity;
 -- Prints 50
    END;

    RAISE NOTICE 'Quantity here is %', quantity;  -- Prints 50

    RETURN quantity;
END;
$$ LANGUAGE plpgsql;
```

> **注意**
>
> 实际上，任何 PL/pgSQL 函数体外面都包着一个隐藏的“外层块”。该块提供函数参数（如果有）的声明，以及一些诸如 FOUND 之类的特殊变量（见第 39.5.5 节）。这个外层块带有函数名标签，这意味着参数和特殊变量可以用函数名来限定。

需要特别注意，不要把BEGIN/END在PL/pgSQL中用于分组语句的用法，与用于事务控制的同名 SQL 命令混淆。PL/pgSQL中的BEGIN/END只用于分组，不会开始或结束事务。函数和触发器函数总是在外层查询建立的事务中执行 – 它们不能开始或提交该事务，否则就没有可供其执行的上下文。但是，包含EXCEPTION子句的块实际上会形成一个可回滚而不影响外层事务的子事务。详见第 39.6.6 节。

39.3. 声明

在一个块中使用的所有变量，都必须在该块的声明部分声明。
（唯一的例外是：在整数范围上迭代的 FOR 循环变量会被自动声明为整数变量；同样，在游标结果上迭代的 FOR 循环变量会被自动声明为记录变量。）

PL/pgSQL 变量可以是任意 SQL 数据类型，例如 integer、varchar 和 char。

这里是变量声明的一些示例：

```
user_id integer;
quantity numeric(5);
url varchar;
myrow tablename%ROWTYPE;
myfield tablename.columnname%TYPE;
arow RECORD;
```

一个变量声明的一般语法是：

```
name [ CONSTANT ] type [ COLLATE collation_name ] [ NOT NULL ] [ {
 DEFAULT | := } expression ];
```

如果给定 DEFAULT 子句，它会指定进入该块时赋给该变量的初始值。如果没有给出 DEFAULT 子句，则变量会被初始化为 SQL 空值。CONSTANT 选项会阻止该变量在初始化之后再次被赋值，因此其值在整个块的持续期间保持不变。COLLATE 选项指定该变量使用的排序规则（见第 39.3.6 节）。如果指定了 NOT NULL，给该变量赋空值将导致运行时错误。所有声明为 NOT NULL 的变量都必须指定非空默认值。

变量的默认值会在每次进入该块时重新计算并赋给该变量，而不是每次函数调用只计算一次。因此，例如把 now() 赋给一个 timestamp 类型变量，会使该变量得到当前函数调用时的时间，而不是函数预编译时的时间。

例如：

```
quantity integer DEFAULT 32;
url varchar := 'http://mysite.com';
user_id CONSTANT integer := 10;
```

39.3.1. 声明函数参数

传递给函数的参数使用标识符 $1、$2 等命名。也可以为这些 $*n* 形式的参数名声明别名，以提高可读性。之后既可以使用别名，也可以使用数字标识符来引用参数值。

有两种方式可以创建别名。推荐的方式是在 CREATE FUNCTION 命令中直接为参数命名。例如：

```
CREATE FUNCTION sales_tax(subtotal real) RETURNS real AS $$
BEGIN
    RETURN subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

另一种方式——这也是 PostgreSQL 8.0 之前唯一可用的方式——是使用声明语法显式声明别名。

```
name ALIAS FOR $n;
```

同样的示例，按这种写法如下：

```
CREATE FUNCTION sales_tax(real) RETURNS real AS $$
DECLARE
    subtotal ALIAS FOR $1;
BEGIN
    RETURN subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

注意

这两个示例并不完全等价。在第一种情况下，subtotal 可以写成 sales_tax.subtotal 来引用；但在第二种情况下则不行。（如果我们给内层块附上一个标签，那么 subtotal 可以用那个标签来限定。）

更多一些示例：

```
CREATE FUNCTION instr(varchar, integer) RETURNS integer AS $$
DECLARE
    v_string ALIAS FOR $1;
    index ALIAS FOR $2;
```

```
BEGIN
    -- 这里是一些使用 v_string 和 index 的计算
END;
$$ LANGUAGE plpgsql;

CREATE FUNCTION concat_selected_fields(in_t sometablename) RETURNS
 text AS $$
BEGIN
    RETURN in_t.f1 || in_t.f3 || in_t.f5 || in_t.f7;
END;
$$ LANGUAGE plpgsql;
```

当 PL/pgSQL 函数声明了输出参数时，输出参数也会像普通输入参数一样获得 `$n` 名称和可选别名。输出参数本质上是一个初始值为 NULL 的变量，应在函数执行期间给它赋值。该参数的最终值就是返回值。例如，销售税的示例也可以这样写：

```
CREATE FUNCTION sales_tax(subtotal real, OUT tax real) AS $$
BEGIN
    tax := subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

注意这里省略了 `RETURNS real` – 当然也可以写上，但那只是冗余。

当需要返回多个值时，输出参数尤其有用。下面是一个简单示例：

```
CREATE FUNCTION sum_n_product(x int, y int, OUT sum int, OUT prod int)
 AS $$
BEGIN
    sum := x + y;
    prod := x * y;
END;
$$ LANGUAGE plpgsql;
```

如第 35.4.4 节所述，这实际上会为函数结果创建一个匿名记录类型。如果写了 `RETURNS` 子句，它必须是 `RETURNS record`。

声明 PL/pgSQL 函数的另一种方式是使用 `RETURNS TABLE`，例如：

```
CREATE FUNCTION extended_sales(p_itemno int)
RETURNS TABLE(quantity int, total numeric) AS $$
BEGIN
    RETURN QUERY SELECT s.quantity, s.quantity * s.price FROM sales AS
 s
                 WHERE s.itemno = p_itemno;
END;
$$ LANGUAGE plpgsql;
```

这与声明一个或多个 `OUT` 参数并指定 `RETURNS SETOF` *sometype* 完全等效。

当 PL/pgSQL 函数的返回类型被声明为多态类型（`anyelement`、`anyarray`、`anynonarray` 或 `anyenum`）时，会创建一个特殊参数 `$0`。它的数据类型就是该函数的实际返回类型，由实际输入类型推导而来（见第 35.2.5 节）。这样，函数就能像第 39.3.3 节所示的那样访问自己的实际返回类型。`$0` 被初始化为空值，并且可以由函数修改，因此如果需要，可以用它来保存返回值，不过这并不是必须的。`$0` 也可以被赋予别名。例如，下面的函数适用于任何带有 + 操作符的数据类型：

```
CREATE FUNCTION add_three_values(v1 anyelement, v2 anyelement, v3
 anyelement)
RETURNS anyelement AS $$
DECLARE
    result ALIAS FOR $0;
BEGIN
    result := v1 + v2 + v3;
    RETURN result;
END;
$$ LANGUAGE plpgsql;
```

把一个或多个输出参数声明为多态类型，也可以达到同样的效果。在这种情况下，不使用特殊参数 `$0`，输出参数本身就承担相同作用。例如：

```
CREATE FUNCTION add_three_values(v1 anyelement, v2 anyelement, v3
 anyelement,
                                 OUT sum anyelement)
AS $$
BEGIN
    sum := v1 + v2 + v3;
END;
$$ LANGUAGE plpgsql;
```

39.3.2. ALIAS

newname ALIAS FOR *oldname*;

`ALIAS` 语法比上一节所展示的更一般化：你可以为任何变量声明别名，而不只是函数参数。它最主要的实际用途，是为那些名称预先固定的变量指定另一个名字，例如触发器函数中的 `NEW` 或 `OLD`。

示例：

```
DECLARE
  prior ALIAS FOR old;
  updated ALIAS FOR new;
```

由于 `ALIAS` 为同一个对象提供了两种命名方式，滥用它会让代码变得混乱。最好只把它用于改写那些预先固定的名称。

39.3.3. 复制类型

variable%TYPE

`%TYPE`提供变量或表列的数据类型。可以用它声明将保存数据库值的变量。例如，假设有一个名为`user_id`的列，位于`users`表中。要声明一个数据类型与`users.user_id`相同的变量，可以写：

```
user_id users.user_id%TYPE;
```

使用 `%TYPE` 的好处是，你不必知道所引用结构的实际数据类型；更重要的是，如果被引用项的数据类型将来发生变化（例如把 `user_id` 的类型从 `integer` 改成 `real`），你可能就不需要修改函数定义。

`%TYPE` 在多态函数中特别有价值，因为内部变量所需的数据类型可能在不同调用之间变化。可以把 `%TYPE` 应用到函数参数或结果占位符上，以创建合适的变量。

39.3.4. 行类型

name table_name%ROWTYPE;
name composite_type_name;

复合类型的变量称为行变量（或行类型变量）。只要查询的列集合与该变量声明的类型相匹配，这种变量就可以保存 SELECT 或 FOR 查询结果中的整行。行值的各个字段可以使用通常的点号记法访问，例如 rowvar.field。

行变量既可以通过 *table_name*%ROWTYPE 记法声明为与现有表或视图的行具有相同类型，也可以通过给出某个复合类型的名称来声明。（由于每个表都有一个同名的关联复合类型，所以在 PostgreSQL 中实际上写不写 %ROWTYPE 并无区别；不过带 %ROWTYPE 的形式可移植性更好。）

函数参数也可以是复合类型（完整的表行）。在这种情况下，相应的标识符 $*n* 就是一个行变量，并且可以从中选取字段，例如 $1.user_id。

在行类型变量中，只能访问表行中用户定义的列，不能访问 OID 或其他系统列（因为该行可能来自视图）。对于char(*n*)这样的数据类型，行类型中的字段会继承表字段的大小或精度。

下面是一个使用复合类型的示例。table1 和 table2 是已经存在的表，它们至少包含下面提到的字段：

```
CREATE FUNCTION merge_fields(t_row table1) RETURNS text AS $$
DECLARE
    t2_row table2%ROWTYPE;
BEGIN
    SELECT * INTO t2_row FROM table2 WHERE ... ;
    RETURN t_row.f1 || t2_row.f3 || t_row.f5 || t2_row.f7;
END;
$$ LANGUAGE plpgsql;

SELECT merge_fields(t.*) FROM table1 t WHERE ... ;
```

39.3.5. 记录类型

name RECORD;

记录变量与行类型变量类似，但没有预定义结构。它会在 SELECT 或 FOR 命令为其赋值时采用相应行的实际结构。记录变量的内部结构在每次被赋值时都可能变化。
其后果是：在记录变量第一次被赋值之前，它没有任何子结构，
任何试图访问其中字段的行为都会引发运行时错误。

注意，RECORD 并不是真正的数据类型，它只是一个占位符。还需要认识到，PL/pgSQL 函数被声明为返回 record，与记录变量并不是完全相同的概念，尽管这样的函数可能会用记录变量保存结果。这两种情况下，在编写函数时都不知道实际的行结构；但对于返回 record 的函数，实际结构会在解析调用查询时确定，而记录变量的行结构则可以在运行过程中随时变化。

39.3.6. PL/pgSQL变量的排序规则

当 PL/pgSQL 函数具有一个或多个支持排序规则的数据类型参数时，每次函数调用都会根据分配给实际参数的排序规则确定出一个排序规则，如第 22.2 节所述。如果该排序规则能成功确定出来（即参数之间的隐式排序规则没有冲突），
那么所有支持排序规则的参数都会被视为隐式带有该排序规则。
这会影响函数中那些受排序规则影响的操作。例如，考虑

```
CREATE FUNCTION less_than(a text, b text) RETURNS boolean AS $$
BEGIN
    RETURN a < b;
END;
$$ LANGUAGE plpgsql;

SELECT less_than(text_field_1, text_field_2) FROM table1;
SELECT less_than(text_field_1, text_field_2 COLLATE "C") FROM table1;
```

第一次调用 `less_than` 时，比较会使用 `text_field_1` 与 `text_field_2` 的共同排序规则；第二次则会使用 `C` 排序规则。

此外，确定出的排序规则也会被视为任何支持排序规则的数据类型局部变量的排序规则。因此，即使把这个函数写成下面这样，其行为也不会有任何不同：

```
CREATE FUNCTION less_than(a text, b text) RETURNS boolean AS $$
DECLARE
    local_a text := a;
    local_b text := b;
BEGIN
    RETURN local_a < local_b;
END;
$$ LANGUAGE plpgsql;
```

如果函数没有支持排序规则的数据类型的参数，或者无法为它们确定共同排序规则，那么参数和局部变量将使用其数据类型的默认排序规则（通常是数据库默认排序规则，但对于域类型变量也可能不同）。

通过在支持排序规则的数据类型局部变量的声明中加入 `COLLATE` 选项，可以为其指定不同的排序规则，例如

```
DECLARE
    local_a text COLLATE "en_US";
```

这个选项会覆盖按上述规则原本应赋给该变量的排序规则。

当然，如果某个函数希望在特定操作中强制使用特定排序规则，也可以在函数内部显式写出 `COLLATE` 子句。例如：

```
CREATE FUNCTION less_than_c(a text, b text) RETURNS boolean AS $$
BEGIN
    RETURN a < b COLLATE "C";
END;
$$ LANGUAGE plpgsql;
```

这会覆盖表达式中表列、参数或局部变量所关联的排序规则，就像在普通 SQL 命令中一样。

39.4. 表达式

PL/pgSQL 语句中使用的所有表达式都由服务器的主 SQL 执行器处理。例如，当你写出这样一条 PL/pgSQL 语句

```
IF expression THEN ...
```

PL/pgSQL 会通过把如下查询交给主 SQL 引擎来对表达式求值：

```
SELECT expression
```

在形成 SELECT 命令时，所有出现的 PL/pgSQL 变量名都会被替换为参数，详见第 39.10.1 节。这样，SELECT 的查询计划只需准备一次，随后就可以在变量值不同时重复使用。因此，首次使用表达式时，实际发生的事情本质上是执行一个 PREPARE 命令。例如，如果已经声明了两个整数变量 x 和 y，并写下

```
IF x < y THEN ...
```

背后发生的事情就等同于：

```
PREPARE statement_name(integer, integer) AS SELECT $1 < $2;
```

随后，对该预备语句调用 EXECUTE，以完成每次 IF 语句的执行，并把 PL/pgSQL 变量的当前值作为参数值提供。以这种方式准备的查询计划会在数据库连接的整个生命周期内保存，如第 39.10.2 节所述。通常，这些细节对 PL/pgSQL 用户并不重要，但在诊断问题时了解它们很有帮助。

39.5. 基本语句

在本节和后续各节中，我们描述PL/pgSQL能够明确理解的所有语句类型。
凡是不被识别为这些类型之一的内容，都被视为 SQL 命令并发送给主数据库引擎执行，详见第 39.5.2 节和第 39.5.3 节。

39.5.1. 赋值

向PL/pgSQL变量赋值的写法为：

```
variable := expression;
```

如前所述，这种语句中的表达式通过发送给主数据库引擎的 SQLSELECT命令求值。表达式必须产生单个值（如果变量是行变量或记录变量，也可以是行值）。
目标变量可以是简单变量（可以用块名限定）、行变量或记录变量的字段，
或者由简单变量或字段表示的数组中的元素。

如果该表达式的结果数据类型不匹配变量的数据类型，或者变量具有特定的大小/精度（例如 char(20)），结果值将由 PL/pgSQL解释器使用结果类型的输出函数和变量类型的输入函数隐式转换。注意如果结果值的字符串形式无法被输入函数所接受，
这可能会导致由输入函数产生的运行时错误。

例如：

```
tax := subtotal * 0.06;
my_record.user_id := 20;
```

39.5.2. 执行没有结果的命令

对于不返回行的 SQL 命令，例如没有RETURNING子句的INSERT，只需写出命令，就可以在PL/pgSQL函数中执行它。

命令文本中出现的任何PL/pgSQL变量名都被视为参数，
然后在运行时将变量的当前值作为参数值提供。这与前面描述的表达式处理完全相同；
详情见第 39.10.1 节。

以这种方式执行 SQL 命令时，PL/pgSQL 只对命令做一次规划，并在数据库连接的整个生命周期内的后续执行中重用该计划。这样做的影响在第 39.10.2 节中有详细讨论。

有时，对表达式或SELECT查询求值但丢弃结果会很有用，
例如调用有副作用但没有有用结果值的函数时。要在PL/pgSQL中这样做，可以使用PERFORM
语句：

```
PERFORM query;
```

这会执行*query*并丢弃结果。编写*query*的方式与编写 SQLSELECT命令相同，只需把开头的关键字SELECT替换为PERFORM。对于WITH查询，使用PERFORM，然后用圆括号括住查询。
（此时查询只能返回一行。）PL/pgSQL变量会被替换到查询中，
方式与不返回结果的命令相同，计划也会以相同方式缓存。另外，特殊变量FOUND
在查询至少产生一行时设为真，没有产生行时设为假（见第 39.5.5 节）。

> **注意**
>
> 我们可能期望直接写SELECT能实现这个结果，但是当前唯一被接受的方式是PERFORM。一个能返回行的 SQL 命令（例如SELECT）将被当成一个错误拒绝，除非它像下一节中讨论的有一个INTO子句。

一个示例：

```
PERFORM create_mv('cs_session_page_requests_mv', my_query);
```

39.5.3. 执行返回单行结果的查询

产生单行（可能有多列）结果的 SQL 命令，其结果可以赋给记录变量、行类型变量或标量变量列表。方法是在基本 SQL 命令中添加INTO子句。例如：

```
SELECT select_expressions INTO [STRICT] target FROM ...;
INSERT ... RETURNING expressions INTO [STRICT] target;
UPDATE ... RETURNING expressions INTO [STRICT] target;
DELETE ... RETURNING expressions INTO [STRICT] target;
```

其中，*target*可以是记录变量、行变量，或逗号分隔的简单变量及记录/行字段列表。
PL/pgSQL变量会被替换到查询的其余部分，计划也会被缓存，
正如上面描述的不返回行的命令一样。这适用于SELECT，INSERT/UPDATE/DELETE
带有RETURNING的情况，以及返回行集结果的工具命令（例如EXPLAIN）。除了INTO
子句以外，SQL 命令的写法与在PL/pgSQL之外的写法相同。

> **提示**
>
> 注意带INTO的SELECT的这种解释和PostgreSQL常规的SELECT INTO命令有很大的不同，后者的INTO目标是一个新创建的表。
> 如果你想要在一个PL/pgSQL函数中从一个SELECT的结果创建一个表，请使用语法
> CREATE TABLE ... AS SELECT。

如果使用行变量或变量列表作为目标，查询结果列的数量和数据类型必须与目标结构完全匹配，
否则会发生运行时错误。如果目标是记录变量，它会自动将自身配置为查询结果列的行类型。

INTO子句几乎可以出现在 SQL 命令中的任何位置。通常它被写成刚好在SELECT命令中的*select_expressions*列表之前或之后，或者在其他命令类型的命令最后。
我们推荐你遵循这种惯例，以防PL/pgSQL的解析器在未来的版本中变得更严格。

如果STRICT没有在INTO子句中指定，那么*target*会被设为查询返回的第一行；
如果查询没有返回行，则设为空值。（注意，“第一行”并没有明确定义，除非使用了ORDER BY。）第一行之后的所有结果行都会被丢弃。可以检查特殊的FOUND变量（见第 39.5.5 节），以确定是否返回了行：

```
SELECT * INTO myrec FROM emp WHERE empname = myname;
IF NOT FOUND THEN
    RAISE EXCEPTION 'employee % not found', myname;
END IF;
```

如果指定STRICT选项，查询必须恰好返回一行，否则会报告运行时错误：NO_DATA_FOUND（没有行）或TOO_MANY_ROWS（多于一行）。如果希望捕获错误，可以使用异常块，例如：

```
BEGIN
    SELECT * INTO STRICT myrec FROM emp WHERE empname = myname;
    EXCEPTION
        WHEN NO_DATA_FOUND THEN
            RAISE EXCEPTION 'employee % not found', myname;
        WHEN TOO_MANY_ROWS THEN
            RAISE EXCEPTION 'employee % not unique', myname;
END;
```

成功执行带有STRICT的命令总会将FOUND设为真。

对于带有RETURNING的INSERT/UPDATE/DELETE，即使没有指定STRICT，
PL/pgSQL也会针对多于一个返回行的情况报告一个错误。这是因为没有类似于ORDER BY的选项可以用来决定应该返回哪个被影响的行。

> **注意**
>
> STRICT选项匹配 Oracle PL/SQL 的SELECT INTO和相关语句的行为。

对于需要处理 SQL 查询的多个结果行的情况，请参见第 39.6.4 节。

39.5.4. 执行动态命令

很多时候你将想要在PL/pgSQL函数中产生动态命令，
也就是每次执行中会涉及到不同表或不同数据类型的命令。
PL/pgSQL通常对于命令所做的缓存计划尝试（如第 39.10.2 节中讨论）在这种情境下无法工作。
要处理这一类问题，提供了EXECUTE语句：

```
EXECUTE command-string [ INTO [STRICT] target ] [ USING expression [,
 ... ] ];
```

其中*command-string*是一个表达式，其求值结果（类型为 text）是包含待执行命令的字符串。可选的*target*是一个记录变量、一个行变量或者一个逗号分隔的简单变量以及记录/行字段的列表，该命令的结果将存储在其中。可选的USING
表达式提供要被插入到该命令中的值。

在计算得到的命令字符串中，不会做PL/pgSQL变量的替换。
任何所需的变量值必须在命令字符串被构造时被插入其中，或者你可以使用下面描述的参数。

另外，通过 EXECUTE 执行的命令不会缓存计划。相反，每次运行该语句时都会重新准备命令。因此，可以在函数内动态创建命令字符串，以对不同的表和列执行操作。

INTO子句指定一个返回行的 SQL 命令的结果应该被赋值到哪里。如果提供了一个行变量或变量列表，它必须完全匹配查询命令结果的结构（当一个记录变量被提供时，它会自动把它自己配置为匹配结果结构）。如果返回多个行，只有第一个行会被赋值给INTO变量。如果没有返回行，NULL 会被赋值给INTO变量。如果没有指定INTO子句，该查询结果会被抛弃。

如果给出了STRICT选项，除非该查询刚好产生一行，否则将会报告一个错误。

命令字符串可以使用参数值，它们在命令中用$1、$2等引用。这些符号引用在USING子句中提供的值。这种方法通常比把数据值作为文本插入命令字符串更可取：它避免了将该值转换为文本以及转换回来的运行时负荷，并且它更不容易被 SQL 注入攻击，因为不需要引用或转义。一个示例是：

```
EXECUTE 'SELECT count(*) FROM mytable WHERE inserted_by = $1 AND
 inserted <= $2'
   INTO c
   USING checked_user, checked_date;
```

注意，参数符号只能用于数据值 – 如果要使用动态确定的表名或列名，必须将其作为文本插入命令字符串。例如，如果前面的查询需要针对动态选择的表执行，可以这样写：

```
EXECUTE 'SELECT count(*) FROM '
    || tabname::regclass
    || ' WHERE inserted_by = $1 AND inserted <= $2'
   INTO c
   USING checked_user, checked_date;
```

参数符号的另一个限制是，它们只能用于SELECT，INSERT，UPDATE以及DELETE命令。在其他语句类型（统称为工具语句）中，即使只是数据值，也必须以文本形式插入。

带简单常量命令字符串和一些 USING 参数的 EXECUTE（如上面第一个示例那样），在功能上等价于直接在 PL/pgSQL 中写出该命令，并让 PL/pgSQL 变量的替换自动发生。重要的区别在于：EXECUTE 会在每次执行时重新规划命令，生成针对当前参数值的特定计划；而 PL/pgSQL 通常创建的是通用计划并将其缓存以供重用。在最佳计划强烈依赖参数值的情况下，EXECUTE 可能明显更快；而当计划对参数值不敏感时，重新规划就是浪费。

EXECUTE目前不支持SELECT INTO。但是可以执行一个纯的SELECT命令并且指定INTO作为EXECUTE本身的一部分。

注意

PL/pgSQL中的EXECUTE语句与EXECUTE这一由PostgreSQL服务器支持的 SQL 语句无关。服务器的EXECUTE语句不能直接在PL/pgSQL函数中使用（并且也没有必要）。

例 39.1. 在动态查询中为值加引号

在使用动态命令时，经常需要处理单引号的转义。我们推荐在函数体中使用美元引用来引用固定文本。（如果你有未使用美元引用的旧代码，请参阅第 39.11.1 节中的概述；在把这类代码转换成更合理的写法时，它会帮你省下一些工夫。）

将被插入到构造出的查询中的动态值需要被小心地处理，因为它们自身可能包含引号字符。一个示例（这假设你对函数整体使用了美元引用，因此引号标记不需要被双写）：

```
EXECUTE 'UPDATE tbl SET '
        || quote_ident(colname)
        || ' = '
        || quote_literal(newvalue)
        || ' WHERE key = '
        || quote_literal(keyvalue);
```

这个示例展示了quote_ident和quote_literal函数的用法（见第 9.4 节）。为了安全起见，在插入动态查询之前，包含列名或表名标识符的表达式应先传给quote_ident。在构造出的命令中应作为字符串字面量出现的值，则应传给quote_literal。这两个函数都会采取适当措施，分别返回用双引号或单引号括起来的输入文本，并正确转义其中嵌入的特殊字符。

由于 quote_literal 被标记为 STRICT，因此用 null 参数调用时它总会返回 null。在上面的示例中，如果 newvalue 或 keyvalue 为 null，整个动态查询字符串都会变成 null，进而导致 EXECUTE 报错。可以通过使用 quote_nullable 函数来避免这个问题；它与 quote_literal 的工作方式相同，只是在用 null 参数调用时会返回字符串 NULL。例如：

```
EXECUTE 'UPDATE tbl SET '
        || quote_ident(colname)
        || ' = '
        || quote_nullable(newvalue)
        || ' WHERE key = '
        || quote_nullable(keyvalue);
```

如果正在处理的参数值可能为空值，那么通常应该用 quote_nullable 来代替 quote_literal。

通常，必须小心地确保查询中的空值不会产生意料之外的结果。例如如果keyvalue为空值，下面的WHERE子句

```
'WHERE key = ' || quote_nullable(keyvalue)
```

永远不会成功，因为在=操作符中使用空值操作数得到的结果总是空值。如果想让空值像普通键值一样工作，你应该将上面的命令重写成

```
'WHERE key IS NOT DISTINCT FROM ' || quote_nullable(keyvalue)
```

（目前，IS NOT DISTINCT FROM的处理效率不如=，因此只有在非常必要时才这样做。关于空值和IS DISTINCT的详细信息请见第 9.2 节）。

请注意美元符号引用只对引用固定文本有用。尝试写出下面这个示例是一个非常糟糕的主意：

```
EXECUTE 'UPDATE tbl SET '
        || quote_ident(colname)
        || ' = $$'
        || newvalue
        || '$$ WHERE key = '
        || quote_literal(keyvalue);
```

因为如果newvalue的内容碰巧含有$$，那么这段代码就会出问题。同样的问题也适用于你选择的任何其他美元符号引用定界符。因此，要想安全地引用事先不知道的文本，必须恰当地使用quote_literal、quote_nullable或quote_ident。

动态 SQL 语句也可以使用format（见第 9.4 节）函数来安全地构造。例如：

```
EXECUTE format('UPDATE tbl SET %I = %L WHERE key = %L', colname,
 newvalue, keyvalue);
```

format函数可以和 USING子句一起使用：

```
EXECUTE format('UPDATE tbl SET %I = $1 WHERE key = $2', colname)
   USING newvalue, keyvalue;
```

这种形式更高效，因为参数newvalue和keyvalue 不会被转换为文本。

动态命令和EXECUTE的一个更大的示例可以在例 39.8中找到，它会构建并且执行一个CREATE FUNCTION命令来定义一个新的函数。

39.5.5. 获取结果状态

有好几种方法可以判断一条命令的效果。第一种方法是使用 GET DIAGNOSTICS 命令，其形式如下：

```
GET DIAGNOSTICS variable = item [ , ... ];
```

这条命令允许检索系统状态指示符。每个 *item* 是一个关键字，它标识一个要赋给指定变量（该变量应具有正确的数据类型来接收它）的状态值。当前可用的状态项是 ROW_COUNT（发送到 SQL 引擎的最后一条 SQL 命令所处理的行数）和 RESULT_OID（最近一条 SQL 命令插入的最后一行的 OID）。注意，RESULT_OID 只有在向包含 OID 的表执行 INSERT 命令之后才有用。

例如：

```
GET DIAGNOSTICS integer_var = ROW_COUNT;
```

第二种确定命令效果的方法是检查名为FOUND的特殊变量，类型为boolean。
在每次PL/pgSQL函数调用中，FOUND的初始值都是 false。它由以下类型的语句设置：

- SELECT INTO语句在为目标赋上一行值时将FOUND设置为true，如果没有返回行则设置为false。
- PERFORM语句在生成（和丢弃）一个或多个行时将FOUND设置为true，如果没有生成行则设置为false。
- UPDATE、INSERT和DELETE语句在至少影响一行时将FOUND设置为真，如果没有影响行则设置为假。
- FETCH语句在返回行时将FOUND设置为true，如果没有返回行则设置为false。
- MOVE语句在成功重新定位游标时将FOUND设置为true，否则设置为false。
- FOR或FOREACH语句在迭代一次或多次时将 FOUND设置为true，否则设置为false。当循环退出时，FOUND会按上述方式设置；在循环执行过程中，FOUND不会被循环语句修改，尽管它可能会被循环体内的其他语句执行修改。
- RETURN QUERY和RETURN QUERY EXECUTE语句在查询返回至少一行时将FOUND设置为true，如果没有返回行则设置为false。

Other PL/pgSQL statements do not change the state of FOUND. Note in particular that EXECUTE changes the output of GET DIAGNOSTICS, but does not change FOUND.

FOUND是每个PL/pgSQL函数的局部变量；任何对它的修改只影响当前的函数。

39.5.6. 什么也不做

有时一个什么也不做的占位语句也很有用。例如，它能够指示 if/then/else 链中故意留出的空分支。可以使用NULL语句达到这个目的：

```
NULL;
```

例如，下面的两段代码是等价的：

```
BEGIN
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN
        NULL;  -- 忽略错误
END;

BEGIN
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN  -- 忽略错误
END;
```

究竟使用哪一种取决于各人的喜好。

> **注意**
>
> 在 Oracle 的 PL/SQL 中，不允许出现空语句列表，并且因此在这种情况下必须使用NULL语句。而PL/pgSQL允许你什么也不写。

39.6. 控制结构

控制结构可能是PL/pgSQL中最有用的（以及最重要）的部分了。利用PL/pgSQL的控制结构，你可以以非常灵活而且强大的方法操纵PostgreSQL的数据。

39.6.1. 从函数返回

有两个命令让我们能够从函数中返回数据：RETURN和RETURN NEXT。

39.6.1.1. RETURN

```
RETURN expression;
```

带有一个表达式的RETURN用于终止函数并把*expression*的值返回给调用者。
这种形式被用于不返回集合的PL/pgSQL函数。

如果函数返回的是标量类型，任何表达式都可以使用。
表达式结果会按照赋值部分的说明自动转换为函数的返回类型。但要返回一个复合（行）值，你必须把记录变量或行变量写为 *expression*。

如果你声明带输出参数的函数，那么就只需要写不带表达式的RETURN。
输出参数变量的当前值将被返回。

如果你声明函数返回void，一个RETURN语句可以被用来提前退出函数；但是不要在RETURN后面写一个表达式。

一个函数的返回值不能是未定义。如果控制到达了函数最顶层块的末尾而没有碰到一个RETURN语句，那么会发生一个运行时错误。不过，这个限制不适用于带输出参数的函数以及返回void的函数。在这些情况中，如果顶层的块结束，将自动执行一个RETURN语句。

39.6.1.2. RETURN NEXT 和 RETURN QUERY

```
RETURN NEXT expression;
RETURN QUERY query;
RETURN QUERY EXECUTE command-string [ USING expression [, ... ] ];
```

当 PL/pgSQL 函数被声明为返回 SETOF *sometype* 时，返回过程会略有不同。在这种情况下，要返回的各个项通过一系列 RETURN NEXT 或 RETURN QUERY 命令指定，最后再用一个不带参数的 RETURN 命令表明函数已经执行完毕。RETURN NEXT 可用于标量和复合数据类型；对于复合结果类型，会返回完整的结果“表”。RETURN QUERY 会把查询执行结果追加到函数的结果集中。在同一个集合返回函数中，RETURN NEXT 和 RETURN QUERY 可以自由混用，此时它们的结果会被串接起来。

RETURN NEXT和RETURN QUERY实际上不会从函数中返回 – 它们简单地向函数的结果集中追加零或多行。然后会继续执行PL/pgSQL函数中的下一条语句。随着后继的RETURN NEXT和RETURN QUERY命令的执行，结果集就建立起来了。最后一个RETURN（应该没有参数）会导致控制退出该函数（或者你可以让控制到达函数的结尾）。

RETURN QUERY有一种变体RETURN QUERY EXECUTE，它可以动态指定要被执行的查询。可以通过USING向计算出的查询字符串插入参数表达式，这和在EXECUTE命令中的方式相同。

如果你声明函数带有输出参数，只需要写不带表达式的RETURN NEXT。在每一次执行时，输出参数变量的当前值将被保存下来用于最终返回为结果的一行。
注意为了创建一个带有输出参数的集合返回函数，在有多个输出参数时，你必须声明函数为返回SETOF record；或者如果只有一个类型为*sometype*的输出参数时，声明函数为SETOF *sometype*。

下面是一个使用 RETURN NEXT 的函数示例：

```
CREATE TABLE foo (fooid INT, foosubid INT, fooname TEXT);
INSERT INTO foo VALUES (1, 2, 'three');
INSERT INTO foo VALUES (4, 5, 'six');

CREATE OR REPLACE FUNCTION getAllFoo() RETURNS SETOF foo AS
$BODY$
DECLARE
    r foo%rowtype;
BEGIN
    FOR r IN SELECT * FROM foo
    WHERE fooid > 0
    LOOP
        -- can do some processing here
        RETURN NEXT r; -- return current row of SELECT
    END LOOP;
    RETURN;
END
$BODY$
LANGUAGE 'plpgsql' ;
```

```
SELECT * FROM getallfoo();
```

> **注意**
>
> 如上所述，目前RETURN NEXT和RETURN QUERY的实现在从函数返回之前会把整个结果集都保存起来。这意味着如果一个PL/pgSQL函数生成一个非常大的结果集，性能可能会很差：数据将被写到磁盘上以避免内存耗尽，但是函数本身在整个结果集都生成之前不会退出。将来的PL/pgSQL版本可能会允许用户定义没有这种限制的集合返回函数。目前，数据开始被写入到磁盘的时机由配置变量work_mem控制。拥有足够内存来存储大型结果集的管理员可以考虑增大这个参数。

39.6.2. 条件语句

IF 和 CASE 语句让你可以根据条件执行不同的命令。PL/pgSQL 有三种形式的 IF：

- IF ... THEN
- IF ... THEN ... ELSE
- IF ... THEN ... ELSIF ... THEN ... ELSE

and two forms of CASE:

- CASE ... WHEN ... THEN ... ELSE ... END CASE
- CASE WHEN ... THEN ... ELSE ... END CASE

39.6.2.1. IF-THEN

```
IF boolean-expression THEN
    statements
END IF;
```

IF-THEN 语句是 IF 的最简单形式。如果条件为真，在 THEN 和 END IF 之间的语句将被执行。否则，将忽略它们。

示例：

```
IF v_user_id <> 0 THEN
    UPDATE users SET email = v_email WHERE user_id = v_user_id;
END IF;
```

39.6.2.2. IF-THEN-ELSE

```
IF boolean-expression THEN
    statements
ELSE
    statements
END IF;
```

IF-THEN-ELSE 语句对 IF-THEN 进行了增加，它让你能够指定一组在条件不为真时应该被执行的语句（注意这也包括条件求值为 NULL 的情况）。

示例：

```
IF parentid IS NULL OR parentid = ''
THEN
    RETURN fullname;
ELSE
    RETURN hp_true_filename(parentid) || '/' || fullname;
END IF;

IF v_count > 0 THEN
    INSERT INTO users_count (count) VALUES (v_count);
    RETURN 't';
ELSE
    RETURN 'f';
END IF;
```

39.6.2.3. IF-THEN-ELSIF

```
IF boolean-expression THEN
    statements
[ ELSIF boolean-expression THEN
    statements
[ ELSIF boolean-expression THEN
    statements
    ...]]
[ ELSE
    statements ]
END IF;
```

有时会有多于两种选择。IF-THEN-ELSIF 则提供了一个简便的方法来检查多个条件。IF 条件会被一个接一个测试，直到找到第一个为真的。然后执行相关语句，之后控制会被交给 END IF 之后的下一个语句（后续的任何 IF 条件不会被测试）。如果没有一个 IF 条件为真，那么 ELSE 块（如果有）将被执行。

示例：

```
IF v_user_id <> 0 THEN
    UPDATE users SET email = v_email WHERE user_id = v_user_id;
END IF;
```

IF-THEN-ELSE语句对IF-THEN进行了增加，
它让你能够指定一组在条件不为真时应该被执行的语句（注意这也包括条件为 NULL 的情况）。

另一个可以完成相同任务的方法是嵌套 IF-THEN-ELSE 语句，如下例：

```
IF demo_row.sex = 'm' THEN
    pretty_sex := 'man';
ELSE
    IF demo_row.sex = 'f' THEN
        pretty_sex := 'woman';
    END IF;
END IF;
```

However, 这种方法要求为每个 IF 编写匹配的 END IF，因此当有许多备选方案时，它比使用 ELSIF 笨拙得多。

39.6.2.4. 简单 CASE

```
CASE search-expression
    WHEN expression [, expression [ ... ]] THEN
      statements
  [ WHEN expression [, expression [ ... ]] THEN
      statements
    ... ]
  [ ELSE
      statements ]
END CASE;
```

CASE 的简单形式提供基于操作数相等的条件执行。*search-expression* 被求值（一次），并依次与 WHEN 子句中的每个 *expression* 比较。如果找到匹配，则执行相应的 *statements*，然后控制传递到 END CASE 之后的下一个语句。（后续的 WHEN 表达式不会被求值。）如果没有找到匹配，则执行 ELSE 的 *statements*；但如果没有 ELSE，则会引发 CASE_NOT_FOUND 异常。

这里有一个简单示例：

```
CASE x
    WHEN 1, 2 THEN
        msg := 'one or two';
    ELSE
        msg := 'other value than one or two';
END CASE;
```

39.6.2.5. 搜索式 CASE

```
CASE
    WHEN boolean-expression THEN
      statements
  [ WHEN boolean-expression THEN
      statements
    ... ]
  [ ELSE
      statements ]
END CASE;
```

CASE 的搜索形式提供基于布尔表达式真实性的条件执行。每个 WHEN 子句的 *boolean-expression* 依次被求值，直到找到产生 true 的那一个。然后执行相应的 *statements*，之后控制传递到 END CASE 之后的下一个语句。（后续的 WHEN 表达式不会被求值。）如果没有找到为真的结果，则执行 ELSE 的 *statements*；但如果没有 ELSE，则会引发 CASE_NOT_FOUND 异常。

下面是一个例子：

```
CASE
    WHEN x BETWEEN 0 AND 10 THEN
        msg := 'value is between zero and ten';
    WHEN x BETWEEN 11 AND 20 THEN
        msg := 'value is between eleven and twenty';
END CASE;
```

这种形式的 CASE 完全等价于 IF-THEN-ELSIF，唯一的规则差异是：到达被省略的 ELSE 子句时会报错，而不是什么也不做。

39.6.3. 简单循环

使用LOOP、EXIT、CONTINUE、WHILE、FOR和FOREACH语句，可以让PL/pgSQL函数重复执行一系列命令。

39.6.3.1. LOOP

```
[ <<label>> ]
LOOP
    statements
END LOOP [ label ];
```

LOOP 定义了一个无条件循环，它会无限重复，直到被 EXIT 或 RETURN 语句终止。可选的 *label* 可以供嵌套循环中的 EXIT 和 CONTINUE 语句使用，以指定这些语句作用的是哪一个循环。

39.6.3.2. EXIT

```
EXIT [ label ] [ WHEN boolean-expression ];
```

如果没有给出 *label*，那么最内层的循环会被终止，然后跟在 END LOOP 后面的语句会被执行。如果给出了 *label*，那么它必须是当前或者更高层的嵌套循环或者语句块的标签。然后该命名循环或块就会被终止，并且控制会转移到该循环/块相应的 END 之后的语句上。

如果指定了 WHEN，则只有在 *boolean-expression* 为真时才会退出循环。否则，控制传递到 EXIT 之后的语句。

EXIT 可以与所有类型的循环一起使用；它并不限于在无条件循环中使用。

当与 BEGIN 块一起使用时，EXIT 会把控制转交给该块结束后的下一条语句。需要注意的是，为此必须使用标签；未加标签的 EXIT 永远不会被视为匹配某个 BEGIN 块。这与 PostgreSQL 8.4 之前的版本不同，旧版本允许未加标签的 EXIT 匹配 BEGIN 块。）

例子：

```
LOOP
    -- some computations
    IF count > 0 THEN
        EXIT;  -- exit loop
    END IF;
END LOOP;

LOOP
    -- some computations
    EXIT WHEN count > 0;  -- same result as previous example
END LOOP;

<<ablock>>
BEGIN
    -- some computations
    IF stocks > 100000 THEN
        EXIT ablock;  -- causes exit from the BEGIN block
    END IF;
    -- computations here will be skipped when stocks > 100000
END;
```

39.6.3.3. CONTINUE

```
CONTINUE [ label ] [ WHEN boolean-expression ];
```

如果没有给出 *label*，最内层循环的下一次迭代会开始。也就是，循环体中剩余的所有语句将被跳过，并且控制会返回到循环控制表达式（如果有）来决定是否需要另一次循环迭代。如果 *label* 存在，它指定应该继续执行的循环的标签。

如果指定了 WHEN，则只有在 *boolean-expression* 为真时才会开始循环的下一次迭代。否则，控制传递到 CONTINUE 之后的语句。

CONTINUE 可以与所有类型的循环一起使用；它并不限于在无条件循环中使用。

示例：

```
LOOP
    -- 一些计算
    IF count > 0 THEN
        EXIT;  -- 退出循环
    END IF;
END LOOP;

LOOP
    -- 一些计算
    EXIT WHEN count > 0;  -- 和前一个示例相同的结果
END LOOP;

<<ablock>>
BEGIN
    -- 一些计算
    IF stocks > 100000 THEN
        EXIT ablock;  -- 导致从 BEGIN 块中退出
    END IF;
    -- 当stocks > 100000时，这里的计算将被跳过
END;
```

39.6.3.4. WHILE

```
[ <<label>> ]
WHILE boolean-expression LOOP
    statements
END LOOP [ label ];
```

只要 *boolean-expression* 被计算为真，WHILE 语句就会重复一个语句序列。在每次进入到循环体之前都会检查该表达式。

例如：

```
WHILE amount_owed > 0 AND gift_certificate_balance > 0 LOOP
    -- some computations here
END LOOP;

WHILE NOT done LOOP
    -- some computations here
END LOOP;
```

39.6.3.5. FOR（整型变体）

```
[ <<label>> ]
FOR name IN [ REVERSE ] expression .. expression [ BY expression ]
 LOOP
    statements
END LOOP [ label ];
```

这种形式的 FOR 会创建一个在一个整数范围上迭代的循环。变量 *name* 会自动定义为类型 integer 并且只在循环内存在（任何该变量名的现有定义在此循环内都将被忽略）。给出范围上下界的两个表达式在进入循环的时候计算一次。如果没有指定 BY 子句，迭代步长为 1，否则步长是 BY 中指定的值，该值也只在循环进入时计算一次。如果指定了 REVERSE，那么在每次迭代后会减去步长，而不是加上步长。

整数 FOR 循环的一些示例：

```
FOR i IN 1..10 LOOP
    -- i will take on the values 1,2,3,4,5,6,7,8,9,10 within the loop
END LOOP;

FOR i IN REVERSE 10..1 LOOP
    -- i will take on the values 10,9,8,7,6,5,4,3,2,1 within the loop
END LOOP;

FOR i IN REVERSE 10..1 BY 2 LOOP
    -- i will take on the values 10,8,6,4,2 within the loop
END LOOP;
```

如果下界大于上界（或者在 REVERSE 情况下是小于），循环体根本不会被执行。而且不会抛出任何错误。

如果 FOR 循环附加了 *label*，则整数循环变量可以用该 *label* 限定的名称来引用。

39.6.4. 遍历查询结果

使用另一种形式的FOR循环，可以迭代一个查询的结果并相应地处理这些数据。语法是：

```
[ <<label>> ]
FOR target IN query LOOP
    statements
END LOOP [ label ];
```

The *target* is a record variable, row variable, or comma-separated list of scalar variables. The *target* is successively assigned each row resulting from the *query* and the loop body is executed for each row. Here is an example:

```
CREATE FUNCTION cs_refresh_mviews() RETURNS integer AS $$
DECLARE
    mviews RECORD;
BEGIN
    RAISE NOTICE 'Refreshing materialized views...';

    FOR mviews IN SELECT * FROM cs_materialized_views ORDER BY sort_
key LOOP
```

```
        -- Now "mviews" has one record from cs_materialized_views

        RAISE NOTICE 'Refreshing materialized view %s ...', quote_
ident(mviews.mv_name);
        EXECUTE 'TRUNCATE TABLE ' || quote_ident(mviews.mv_name);
        EXECUTE 'INSERT INTO '
                   || quote_ident(mviews.mv_name) || ' '
                   || mviews.mv_query;
    END LOOP;

    RAISE NOTICE 'Done refreshing materialized views.';
    RETURN 1;
END;
$$ LANGUAGE plpgsql;
```

If the loop is terminated by an `EXIT` statement, the last assigned row value is still accessible after the loop.

这种形式的 `FOR` 语句所使用的 *query* 可以是任何向调用者返回行的 SQL 命令：`SELECT` 是最常见的情况，但也可以使用带 `RETURNING` 子句的 `INSERT`、`UPDATE` 或 `DELETE`。某些工具命令（如 `EXPLAIN`）也可以。

PL/pgSQL 变量会被替换到查询文本中，并且查询计划会被缓存以备可能的重用，详见第 39.10.1 节和第 39.10.2 节。

`FOR-IN-EXECUTE` 语句是另一种遍历行的方式：

```
[ <<label>> ]
FOR target IN EXECUTE text_expression [ USING expression [, ... ] ]
 LOOP
    statements
END LOOP [ label ];
```

它与前一种形式类似，只是源查询被指定为一个字符串表达式，在每次进入 `FOR` 循环时都会重新求值和重新规划。这让程序员可以在预先规划的查询的速度与动态查询的灵活性之间做出选择，就像使用普通的 `EXECUTE` 语句一样。与 `EXECUTE` 一样，可以通过 `USING` 向动态命令中插入参数值。

指定要遍历其结果的查询的另一种方式，是把它声明为游标。这在第 39.7.4 节中描述。

39.6.5. 遍历数组

`FOREACH`循环与`FOR`循环很相似，但它不是迭代一个 SQL 查询返回的行，而是迭代一个数组值的元素。（一般而言，`FOREACH`旨在循环遍历组合值表达式的组成部分；
将来可能会增加遍历数组之外的其他组合的变体。）用于遍历数组的`FOREACH`语句是：

```
[ <<label>> ]
FOREACH target [ SLICE number ] IN ARRAY expression LOOP
    statements
END LOOP [ label ];
```

未指定`SLICE`或指定了`SLICE 0`时，循环会遍历计算*expression*得到的数组的各个元素。*target*变量被依次赋予每个元素的值，循环体为每个元素执行一次。
下面是遍历一个整数数组元素的例子：

```
CREATE FUNCTION sum(int[]) RETURNS int8 AS $$
```

```
DECLARE
  s int8 := 0;
  x int;
BEGIN
  FOREACH x IN ARRAY $1
  LOOP
    s := s + x;
  END LOOP;
  RETURN s;
END;
$$ LANGUAGE plpgsql;
```

The elements are visited in storage order, regardless of the number of array dimensions. Although the *target* is usually just a single variable, it can be a list of variables when looping through an array of composite values (records). In that case, for each array element, the variables are assigned from successive columns of the composite value.

当SLICE值为正数时，FOREACH迭代的是数组的切片而非单个元素。SLICE值必须是一个不大于数组维数的整数常量。*target*变量必须是一个数组，它会依次收到数组值的各个切片，每个切片的维数由SLICE指定。下面是遍历一维切片的例子：

```
CREATE FUNCTION scan_rows(int[]) RETURNS void AS $$
DECLARE
  x int[];
BEGIN
  FOREACH x SLICE 1 IN ARRAY $1
  LOOP
    RAISE NOTICE 'row = %', x;
  END LOOP;
END;
$$ LANGUAGE plpgsql;

SELECT scan_rows(ARRAY[[1,2,3],[4,5,6],[7,8,9],[10,11,12]]);

NOTICE:  row = {1,2,3}
NOTICE:  row = {4,5,6}
NOTICE:  row = {7,8,9}
NOTICE:  row = {10,11,12}
```

39.6.6. 捕获错误

默认情况下，PL/pgSQL 函数中发生的任何错误都会中止函数及其外围事务的执行。你可以使用带有 EXCEPTION 子句的 BEGIN 块来捕获错误并从中恢复。其语法是在普通 BEGIN 块语法上的扩展：

```
[ <<label>> ]
[ DECLARE
    declarations ]
BEGIN
    statements
EXCEPTION
    WHEN condition [ OR condition ... ] THEN
        handler_statements
    [ WHEN condition [ OR condition ... ] THEN
```

```
          handler_statements
      ... ]
END;
```

如果没有发生错误，这种形式的块只是简单地执行所有 *statements*，并且接着控制转到 END 之后的下一个语句。但是如果在该 *statements* 内发生了一个错误，则会放弃对 *statements* 的进一步处理，然后控制会转到 EXCEPTION 列表。系统会在列表中寻找匹配所发生错误的第一个 *condition*。如果找到一个匹配，则执行对应的 *handler_statements*，并且接着把控制转到 END 之后的下一个语句。如果没有找到匹配，该错误就会传播出去，就好像根本没有 EXCEPTION 一样：错误可以被一个带有 EXCEPTION 的外围块捕捉，如果没有这样的块则中止该函数的处理。

condition 的名字可以是附录 A中显示的任何名字。一个分类名匹配其中所有的错误。特殊的条件名 OTHERS 匹配除了 QUERY_CANCELED 之外的所有错误类型（虽然可能但通常并不明智，还是可以用名字捕获 QUERY_CANCELED）。条件名是大小写无关的。一个错误条件也可以通过 SQLSTATE 代码指定，例如以下是等价的：

```
WHEN division_by_zero THEN ...
WHEN SQLSTATE '22012' THEN ...
```

如果在选中的 *handler_statements* 内发生了新的错误，那么它不能被这个 EXCEPTION 子句捕获，而是被传播出去。一个外层的 EXCEPTION 子句可以捕获它。

当一个错误被 EXCEPTION 子句捕获时，PL/pgSQL 函数的局部变量会保持错误发生时的值，但是该块中所有对持久数据库状态的改变都会被回滚。例如，考虑这个片段：

```
INSERT INTO mytab(firstname, lastname) VALUES('Tom', 'Jones');
BEGIN
    UPDATE mytab SET firstname = 'Joe' WHERE lastname = 'Jones';
    x := x + 1;
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN
        RAISE NOTICE 'caught division_by_zero';
        RETURN x;
END;
```

当控制到达对 y 的赋值时，它会以 division_by_zero 错误失败。该错误会被 EXCEPTION 子句捕获。RETURN 语句返回的值将是 x 递增后的值，但 UPDATE 命令的效果已被回滚。不过，块之前的 INSERT 命令没有被回滚，因此最终结果是数据库中包含的是 Tom Jones 而不是 Joe Jones。

> **提示**
>
> 包含 EXCEPTION 子句的块，其进入和退出的开销明显高于没有该子句的块。因此，不要在没有必要的情况下使用 EXCEPTION。

在异常处理器中，SQLSTATE 变量包含与所引发异常对应的错误代码（可能错误代码的列表参见表 A.1）。SQLERRM 变量包含与该异常关联的错误消息。这些变量在异常处理器之外是未定义的。

例 39.2. UPDATE/INSERT的异常

这个例子使用异常处理来恰当地执行UPDATE或INSERT：

```
CREATE TABLE db (a INT PRIMARY KEY, b TEXT);

CREATE FUNCTION merge_db(key INT, data TEXT) RETURNS VOID AS
$$
BEGIN
    LOOP
        -- first try to update the key
        UPDATE db SET b = data WHERE a = key;
        IF found THEN
            RETURN;
        END IF;
        -- not there, so try to insert the key
        -- if someone else inserts the same key concurrently,
        -- we could get a unique-key failure
        BEGIN
            INSERT INTO db(a,b) VALUES (key, data);
            RETURN;
        EXCEPTION WHEN unique_violation THEN
            -- Do nothing, and loop to try the UPDATE again.
        END;
    END LOOP;
END;
$$
LANGUAGE plpgsql;

SELECT merge_db(1, 'david');
SELECT merge_db(1, 'dennis');
```

This example assumes the `unique_violation` error is caused by the `INSERT`, and not by an `INSERT` trigger function on the table.

39.7. 游标

和一次执行整个查询不同，可以建立一个游标来封装该查询，
并且接着一次读取该查询结果的一些行。
这样做的原因之一是在结果中包含大量行时避免内存不足（不过，
PL/pgSQL用户通常不需要担心这些，因为`FOR`
循环在内部会自动使用一个游标来避免内存问题）。
一种更有趣的用法是返回一个函数已经创建的游标的引用，允许调用者读取行。
这提供了一种有效的方法从函数中返回大型行集。

39.7.1. 声明游标变量

所有在PL/pgSQL中对游标的访问都会通过游标变量，它总是特殊的数据类型`refcursor`。
创建游标变量的一种方法是把它声明为一个类型为`refcursor`的变量。
另外一种方法是使用游标声明语法，通常是：

```
name [ [ NO ] SCROLL ] CURSOR [ ( arguments ) ] FOR query;
```

（为兼容 Oracle，可以用 `IS` 代替 `FOR`。）如果指定了 `SCROLL`，游标就支持向后滚动；如果指定了 `NO SCROLL`，向后提取会被拒绝；如果两者都未指定，是否允许向后提取则取决于查询本身。如果指定了 *`arguments`*，它就是一个由 *`name datatype`* 对组成的逗号分隔列表，这些名字会在给定查询中被参数值替换。实际替换这些名字的值会在打开游标时提供。

Some examples:

```
DECLARE
    curs1 refcursor;
    curs2 CURSOR FOR SELECT * FROM tenk1;
    curs3 CURSOR (key integer) IS SELECT * FROM tenk1 WHERE unique1 =
 key;
```

所有这三个变量都是 `refcursor` 类型，但是第一个可以用于任何查询，而第二个已经被绑定了一个完全指定的查询，并且最后一个被绑定了一个参数化查询。（游标被打开时，`key` 将被一个整数参数值替换。）变量 `curs1` 被称为未绑定，因为它没有被绑定到任何特定查询。

39.7.2. 打开游标

在能够使用游标检索行之前，必须先将其打开（这等效于 SQL 命令`DECLARE CURSOR`）。PL/pgSQL有三种形式的`OPEN`命令，其中两种用于未绑定游标变量，另一种用于已绑定游标变量。

> **注意**
>
> 可以通过第 39.7.4 节中描述的`FOR`语句在不显式打开游标的情况下使用已绑定的游标变量。

39.7.2.1. OPEN FOR *query*

```
OPEN unbound_cursorvar [ [ NO ] SCROLL ] FOR query;
```

该游标变量会被打开，并被赋予要执行的指定查询。该游标不能已经处于打开状态，
并且它必须已被声明为未绑定游标变量（即，一个简单的`refcursor`变量）。
该查询必须是`SELECT`，或其他会返回行的命令（例如`EXPLAIN`）。
该查询会以与PL/pgSQL中其他 SQL 命令相同的方式处理：替换PL/pgSQL变量名，并缓存查询计划以备后续重用。当一个PL/pgSQL变量被替换到游标查询中时，被替换的是它在`OPEN`
时刻所具有的值；之后对该变量的更改不会影响游标的行为。`SCROLL`和`NO SCROLL`选项与已绑定游标中的含义相同。

一个示例：

```
OPEN curs1 FOR SELECT * FROM foo WHERE key = mykey;
```

39.7.2.2. OPEN FOR EXECUTE

```
OPEN unbound_cursorvar [ [ NO ] SCROLL ] FOR EXECUTE query_string
                                      [ USING expression [, ... ] ];
```

该游标变量会被打开，并被赋予要执行的指定查询。该游标不能已经处于打开状态，
并且必须已被声明为未绑定游标变量（即，一个简单的`refcursor`变量）。
该查询以字符串表达式的形式给出，这一点与`EXECUTE`命令相同。像往常一样，
这提供了灵活性，因此查询计划可以在不同执行之间变化（见第 39.10.2 节），同时也意味着不会在该命令字符串上执行变量替换。与`EXECUTE`一样，可以通过`USING`
把参数值插入动态命令中。`SCROLL`和`NO SCROLL`选项与已绑定游标中的含义相同。

一个示例：

```
OPEN curs1 FOR EXECUTE 'SELECT * FROM ' || quote_ident(tabname)
                                        || ' WHERE col1 = $1' USING
 keyvalue;
```

在这个示例中，表名以文本方式插入到查询中，因此建议使用 `quote_ident()`来防范 SQL 注入。`col1` 的比较值通过`USING`参数插入，因此无需加引号。

39.7.2.3. 打开已绑定游标

```
OPEN bound_cursorvar [ ( argument_values ) ];
```

这种形式的`OPEN`用于打开一个查询在声明时已绑定到其上的游标变量。该游标不能已被打开。当且仅当游标被声明为接受参数时，才必须出现实参值表达式列表。这些值将被代入查询中。绑定游标的查询计划总是被视为可缓存的；这种情况没有等价的`EXECUTE`。注意，不能指定`SCROLL`和`NO SCROLL`，因为游标的滚动行为已经确定。

注意，因为会在已绑定游标的查询上执行变量替换，所以有两种向游标传值的方式：要么给 `OPEN` 一个显式参数，要么在查询中隐式地引用 PL/pgSQL 变量。但是，只有在已绑定游标声明之前声明的变量才会被替换进去。无论哪种方式，要传递的值都是在 `OPEN` 时刻确定的。

示例：

```
OPEN curs2;
OPEN curs3(42);
```

39.7.3. 使用游标

一旦一个游标已经被打开，那么就可以用这里描述的语句操作它。

这些操作不必发生在最初打开该游标的同一个函数中。你可以从函数中返回一个`refcursor`值，让调用者来操作该游标。（在内部，`refcursor`值只是一个所谓 portal 的字符串名称，该 portal 包含了该游标活动查询的状态。这个名称可以被传递、赋给其他`refcursor`变量等等，而不会干扰该 portal。）

所有 portal 都会在事务结束时被隐式关闭。因此，`refcursor`值只能在事务结束之前用于引用一个打开的游标。

39.7.3.1. FETCH

```
FETCH [ direction { FROM | IN } ] cursor INTO target;
```

`FETCH`从游标中检索下一行到目标中，目标可以是一个行变量、记录变量或者逗号分隔的简单变量列表，就像`SELECT INTO`一样。如果没有下一行，目标会被设置为 NULL。与`SELECT INTO`一样，可以检查特殊变量`FOUND`来看是否获得了一行。

direction 子句可以是 SQL FETCH命令中允许的任何变体，除了那些能够取得多于一行的。即它可以是 `NEXT`、`PRIOR`、`FIRST`、`LAST`、`ABSOLUTE` *count*、`RELATIVE` *count*、`FORWARD` 或者 `BACKWARD`。省略 *direction* 和指定 `NEXT` 是一样的。要求反向移动的 *direction* 值很可能会失败，除非游标被使用 `SCROLL` 选项声明或打开。

*cursor*必须是一个引用已打开游标 portal 的`refcursor`变量名。

示例：

```
FETCH curs1 INTO rowvar;
FETCH curs2 INTO foo, bar, baz;
FETCH LAST FROM curs3 INTO x, y;
FETCH RELATIVE -2 FROM curs4 INTO x;
```

39.7.3.2. MOVE

```
MOVE [ direction { FROM | IN } ] cursor;
```

MOVE重新定位一个游标而不检索任何数据。MOVE的工作方式与FETCH完全一样，只是仅重新定位游标，而不返回移动到的行。与SELECT INTO一样，可以检查特殊变量FOUND来看是否存在可移动到的下一行。

direction 子句可以是 SQL FETCH命令中允许的任何变体，即 NEXT、PRIOR、FIRST、LAST、ABSOLUTE *count*、RELATIVE *count*、ALL、FORWARD [*count* | ALL] 或 BACKWARD [*count* | ALL]。省略 *direction* 和指定 NEXT 是一样的。要求反向移动的 *direction* 值很可能会失败，除非游标被使用 SCROLL 选项声明或打开。

示例：

```
MOVE curs1;
MOVE LAST FROM curs3;
MOVE RELATIVE -2 FROM curs4;
MOVE FORWARD 2 FROM curs4;
```

39.7.3.3. UPDATE/DELETE WHERE CURRENT OF

```
UPDATE table SET ... WHERE CURRENT OF cursor;
DELETE FROM table WHERE CURRENT OF cursor;
```

当游标定位在某个表行上时，可以使用该游标来标识该行，并对其执行更新或删除。游标查询的形式有一些限制（尤其不能包含分组），并且在这类场景中最好对游标使用 FOR UPDATE。详见DECLARE参考页。

示例：

```
FETCH curs1 INTO rowvar;
FETCH curs2 INTO foo, bar, baz;
FETCH LAST FROM curs3 INTO x, y;
FETCH RELATIVE -2 FROM curs4 INTO x;
```

39.7.3.4. CLOSE

```
CLOSE cursor;
```

CLOSE关闭打开游标所基于的 portal。这样就可以在事务结束之前提前释放资源，或者释放该游标变量以便再次打开。

示例：

```
MOVE curs1;
MOVE LAST FROM curs3;
MOVE RELATIVE -2 FROM curs4;
MOVE FORWARD 2 FROM curs4;
```

39.7.3.5. 返回游标

PL/pgSQL函数可以向调用者返回游标。这对于返回多行或多列，
特别是非常大的结果集时很有用。要做到这一点，函数需要打开游标，
并把游标名返回给调用者（或者直接使用调用者指定或已知的 portal 名称来打开游标）。随后调用者就可以从该游标中提取行。游标既可以由调用者关闭，也会在事务结束时自动关闭。

游标使用的 portal 名称既可以由程序员指定，也可以自动生成。要指定 portal 名称，只需在打开`refcursor`变量之前给它赋一个字符串值。`OPEN`会把该`refcursor`变量的字符串值用作底层 portal 的名称。不过，如果`refcursor`变量为 null，`OPEN`就会自动生成一个与任何现有 portal 都不冲突的名称，并把它赋回给`refcursor`变量。

> **注意**
>
> 已绑定游标变量会被初始化为表示其名称的字符串值，
> 因此除非程序员在打开游标之前通过赋值覆盖它，否则 portal 名称与游标变量名相同。而未绑定游标变量在初始时默认为空值，因此除非被覆盖，
> 否则它会得到一个自动生成的唯一名称。

一个示例：

```
CLOSE curs1;
```

下面的示例使用了自动游标名生成：

```
CREATE FUNCTION reffunc2() RETURNS refcursor AS '
DECLARE
    ref refcursor;
BEGIN
    OPEN ref FOR SELECT col FROM test;
    RETURN ref;
END;
' LANGUAGE plpgsql;

-- 需要在一个事务中使用游标。
BEGIN;
SELECT reffunc2();

      reffunc2
--------------------
 <unnamed cursor 1>
(1 row)

FETCH ALL IN "<unnamed cursor 1>";
COMMIT;
```

下面的示例展示了从一个函数中返回多个游标的一种方法：

```
CREATE FUNCTION myfunc(refcursor, refcursor) RETURNS SETOF refcursor
 AS $$
BEGIN
    OPEN $1 FOR SELECT * FROM table_1;
    RETURN NEXT $1;
```

```
    OPEN $2 FOR SELECT * FROM table_2;
    RETURN NEXT $2;
END;
$$ LANGUAGE plpgsql;

-- 需要在一个事务中使用游标。
BEGIN;

SELECT * FROM myfunc('a', 'b');

FETCH ALL FROM a;
FETCH ALL FROM b;
COMMIT;
```

39.7.4. 遍历游标结果

FOR 语句还有一种变体，允许循环通过游标返回的行。语法是：

```
[ <<label>> ]
FOR recordvar IN bound_cursorvar [ ( argument_values ) ] LOOP
    statements
END LOOP [ label ];
```

该游标变量必须在声明时已经被绑定到某个查询，并且它不能已经处于打开状态。FOR 语句会自动打开游标，并且在退出循环时自动关闭游标。当且仅当游标被声明要使用参数时，
才必须出现一个实际参数值表达式的列表。这些值会被替换到查询中，采用 OPEN 期间的方式。
变量 *recordvar* 会被自动定义为 record 类型，并且只存在于循环内部（循环中该变量名任何已有定义都会被忽略）。
每一个由游标返回的行都会被陆续地赋值给这个记录变量并且执行循环体。

39.8. 错误和消息

使用RAISE语句报告消息和抛出错误。

```
RAISE [ level ] 'format' [, expression [, ... ]] [ USING option =
 expression [, ... ] ];
RAISE [ level ] condition_name [ USING option = expression [, ... ] ];
RAISE [ level ] SQLSTATE 'sqlstate' [ USING option = expression [, ..
. ] ];
RAISE [ level ] USING option = expression [, ... ];
RAISE ;
```

其中，*level*选项指定错误的严重程度。允许的级别为DEBUG，LOG，INFO，NOTICE，WARNING以及EXCEPTION，其中EXCEPTION是默认值。EXCEPTION
会抛出错误（通常会中止当前事务）；其他级别只会生成不同优先级的消息。
特定优先级的消息是报告给客户端、写入服务器日志，还是两者都做，由log_min_messages和client_min_messages配置变量控制。更多信息见第 18 章中的说明。

在*level*（如果有）之后，可以写一个*format*（必须是简单的字符串字面量，
不能是表达式）。格式字符串指定要报告的错误消息文本。
格式字符串之后可以跟上可选的参数表达式，其值将被插入消息中。在格式字符串内，%
会被替换为下一个可选参数值的字符串表示。写成%%可以输出一个字面的%。

在这个示例中，v_job_id的值将替换字符串中的%：

```
RAISE NOTICE 'Calling cs_create_job(%)', v_job_id;
```

可以通过写出 USING，后面跟上 *option* = *expression* 项目，为错误报告附加额外信息。允许的 *option* 关键字是 MESSAGE、DETAIL、HINT 和 ERRCODE，每个 *expression* 都可以是任意字符串值表达式。MESSAGE 设置错误消息文本（该选项不能用于在 USING 之前包含格式字符串的 RAISE 形式）。DETAIL 提供错误的详细信息，HINT 提供一个提示消息。ERRCODE 指定要报告的错误代码（SQLSTATE），可以用附录 A中所示的条件名，或者直接作为一个五字符 SQLSTATE 代码。

这个例子会中止事务，并给出指定的错误消息和提示：

```
RAISE EXCEPTION 'Nonexistent ID --> %', user_id
      USING HINT = 'Please check your user ID';
```

这两个示例展示了设置 SQLSTATE 的两种等价的方法：

```
RAISE 'Duplicate user ID: %', user_id USING ERRCODE = 'unique_
violation';
RAISE 'Duplicate user ID: %', user_id USING ERRCODE = '23505';
```

还有第二种RAISE语法，其中主参数是要报告的条件名或 SQLSTATE，例如：

```
RAISE division_by_zero;
RAISE SQLSTATE '22012';
```

在这种语法中，USING可以用来提供自定义的错误消息、细节或提示。
另一种达到前面示例同样效果的方式是：

```
RAISE unique_violation USING MESSAGE = 'Duplicate user ID: ' || user_
id;
```

还有另一种变体是写RAISE USING或RAISE *level* USING，并把其余内容都放在USING列表里。

RAISE的最后一种变体根本没有参数。这种形式只能被用在一个BEGIN块的EXCEPTION子句中，它导致当前正在被处理的错误被重新抛出。

> **注意**
>
> 在PostgreSQL 9.1 之前，没有参数的RAISE被解释为重新抛出来自包含活动异常处理器的块的错误。因此一个嵌套在那个处理器中的EXCEPTION子句无法捕捉它，即使RAISE 位于嵌套EXCEPTION子句的块中也是这样。这种行为很奇怪，也并不兼容 Oracle 的 PL/SQL。

如果在RAISE EXCEPTION命令中没有指定条件名称或SQLSTATE，则默认使用RAISE_EXCEPTION (P0001)。如果没有指定消息文本，则默认使用条件名称或SQLSTATE作为消息文本。

> **注意**
>
> 当用 SQLSTATE 代码指定错误代码时，你并不受限于预定义错误代码，而是可以选择任何由五位数字和/或大写 ASCII 字母构成的错误代码，唯一不能使用的是 00000。我们建议尽

> 量避免抛出以三个零结尾的错误代码，因为这些是类别代码，
> 只能通过捕获整个类别来捕获这类错误。

39.9. 触发器函数

PL/pgSQL可以被用来定义触发器函数。触发器函数用CREATE FUNCTION命令创建，它被声明为一个没有参数并且返回类型为trigger的函数。注意，即便该函数准备接收一些在CREATE TRIGGER中指定的参数 – 这类参数通过TG_ARGV传递（如下所述），也必须把它声明为没有参数。

当一个 PL/pgSQL 函数作为触发器被调用时，会在顶层块中自动创建一些特殊变量。它们是：

NEW

数据类型为 RECORD；该变量保存行级触发器中用于 INSERT/UPDATE 操作的新数据行。在语句级触发器和 DELETE 操作中，该变量为 NULL。

OLD

数据类型为 RECORD；该变量保存行级触发器中用于 UPDATE/DELETE 操作的旧数据行。在语句级触发器和 INSERT 操作中，该变量为 NULL。

TG_NAME

数据类型为 name；包含实际触发的触发器名称的变量。

TG_WHEN

数据类型为 text；根据触发器定义，其值为字符串 BEFORE、AFTER 或 INSTEAD OF。

TG_LEVEL

数据类型为 text；根据触发器定义，其值为字符串 ROW 或 STATEMENT。

TG_OP

数据类型为 text；表示触发器对应操作的字符串：INSERT、UPDATE、DELETE 或 TRUNCATE。

TG_RELID

数据类型为 name；包含实际触发的触发器名称的变量。

TG_RELNAME

数据类型为 name；导致触发器调用的表名。该变量已弃用，未来版本可能移除；请改用 TG_TABLE_NAME。

TG_TABLE_NAME

数据类型为 name；导致触发器调用的表名。

TG_TABLE_SCHEMA

数据类型为 name；导致触发器调用的表所在模式名。

TG_NARGS

数据类型为integer；CREATE TRIGGER语句中传给触发器函数的参数个数。

TG_ARGV[]

数据类型为text数组；来自CREATE TRIGGER语句的参数。索引从 0 开始；非法索引（小于 0 或大于等于tg_nargs）返回 null。

一个触发器函数必须返回NULL或者是一个与触发器为之引发的表结构完全相同的记录/行值。

BEFORE 行级触发器可以返回 null，以通知触发器管理器跳过该行后续的操作（也就是说，不再触发后续触发器，并且不会对该行执行INSERT/UPDATE/DELETE）。如果返回非 null 值，则操作会继续进行，并使用该行值。返回一个不同于原始NEW的行值会改变即将插入或更新的行。因此，如果触发器函数希望触发动作正常成功而不修改行值，就必须返回NEW（或与之相等的值）。若要修改将被存储的行，可以直接替换NEW中的单个值并返回修改后的NEW，或者构造一个完整的新记录/行来返回。对于作用于DELETE的 before 触发器，返回值本身没有直接效果，但必须为非 null 才能让触发器动作继续。注意，在DELETE触发器中NEW为 null，因此通常没有理由返回它。在DELETE触发器中，常见写法是返回OLD。

INSTEAD OF触发器（总是行级触发器，并且只能用于视图）能够返回 null 来表示它们没有执行任何更新，并且对该行剩余的操作应被跳过（即后续的触发器不会被引发，并且该行不会被计入外围INSERT/UPDATE/DELETE的行影响状态中）。否则应该返回一个非 null 值用以表示该触发器执行了所请求的操作。对于INSERT 和UPDATE操作，返回值应该是NEW，触发器函数可能对它进行了修改来支持INSERT RETURNING和UPDATE RETURNING（这也将影响被传递给任何后续触发器的行值）。对于DELETE操作，返回值应该是OLD。

一个AFTER行级触发器或一个BEFORE或AFTER语句级触发器的返回值总是会被忽略，因此也可以返回 null。不过，任何这些类型的触发器可能仍会通过抛出一个错误来中止整个操作。

例 39.3展示了PL/pgSQL中一个触发器函数的示例。

例 39.3. 一个 PL/pgSQL 触发器函数

这个示例触发器保证：任何时候一个行在表中被插入或更新时，当前用户名和时间也会被标记在该行中。并且它会检查给出了一个雇员的姓名以及薪水是一个正值。

```
CREATE TABLE emp (
    empname text,
    salary integer,
    last_date timestamp,
    last_user text
);

CREATE FUNCTION emp_stamp() RETURNS trigger AS $emp_stamp$
    BEGIN
        -- Check that empname and salary are given
        IF NEW.empname IS NULL THEN
            RAISE EXCEPTION 'empname cannot be null';
        END IF;
        IF NEW.salary IS NULL THEN
            RAISE EXCEPTION '% cannot have null salary', NEW.empname;
```

```
        END IF;

        -- Who works for us when she must pay for it?
        IF NEW.salary < 0 THEN
            RAISE EXCEPTION '% cannot have a negative salary', NEW.
empname;
        END IF;

        -- Remember who changed the payroll when
        NEW.last_date := current_timestamp;
        NEW.last_user := current_user;
        RETURN NEW;
    END;
$emp_stamp$ LANGUAGE plpgsql;

CREATE TRIGGER emp_stamp BEFORE INSERT OR UPDATE ON emp
    FOR EACH ROW EXECUTE PROCEDURE emp_stamp();
```

另一种记录对表的改变的方法涉及到创建一个新表来为每一个发生的插入、更新或删除保持一行。这种方法可以被认为是对一个表的改变的审计。例 39.4展示了PL/pgSQL中一个审计触发器函数的示例。

例 39.4. 一个用于审计的 PL/pgSQL 触发器函数

这个示例触发器保证了在emp表上的任何插入、更新或删除一行的动作都被记录（即审计）在emp_audit表中。当前时间和用户名会被记录到行中，还有在其上执行的操作类型。

```
CREATE TABLE emp (
    empname           text NOT NULL,
    salary            integer
);

CREATE TABLE emp_audit(
    operation         char(1)   NOT NULL,
    stamp             timestamp NOT NULL,
    userid            text      NOT NULL,
    empname           text      NOT NULL,
    salary integer
);

CREATE OR REPLACE FUNCTION process_emp_audit() RETURNS TRIGGER AS
 $emp_audit$
    BEGIN
        --
        -- Create a row in emp_audit to reflect the operation
 performed on emp,
        -- make use of the special variable TG_OP to work out the
 operation.
        --
        IF (TG_OP = 'DELETE') THEN
            INSERT INTO emp_audit SELECT 'D', now(), user, OLD.*;
            RETURN OLD;
        ELSIF (TG_OP = 'UPDATE') THEN
            INSERT INTO emp_audit SELECT 'U', now(), user, NEW.*;
```

```
            RETURN NEW;
        ELSIF (TG_OP = 'INSERT') THEN
            INSERT INTO emp_audit SELECT 'I', now(), user, NEW.*;
            RETURN NEW;
        END IF;
        RETURN NULL; -- result is ignored since this is an AFTER
 trigger
    END;
$emp_audit$ LANGUAGE plpgsql;

CREATE TRIGGER emp_audit
AFTER INSERT OR UPDATE OR DELETE ON emp
    FOR EACH ROW EXECUTE PROCEDURE process_emp_audit();
```

前一个示例的一个变体使用视图把主表和审计表连接起来，
以显示每个条目最后一次被修改的时间。这种方法仍然记录了表更改的完整审计轨迹，
同时也提供了一个简化视图，只显示从审计轨迹中导出的每个条目的最后修改时间戳。例 39.5展示了PL/pgSQL中一个定义在视图上的审计触发器示例。

例 39.5. 一个用于审计的 PL/pgSQL 视图触发器函数

这个示例在视图上使用了一个触发器让它变得可更新，并且确保视图中一行的任何插入、
更新或删除被记录（即审计）在`emp_audit`表中。
当前时间和用户名会被与执行的操作类型一起记录，并且该视图会显示每一行的最后修改时间。

```
CREATE TABLE emp (
    empname           text PRIMARY KEY,
    salary            integer
);

CREATE TABLE emp_audit(
    operation         char(1)   NOT NULL,
    userid            text      NOT NULL,
    empname           text      NOT NULL,
    salary            integer,
    stamp             timestamp NOT NULL
);

CREATE VIEW emp_view AS
    SELECT e.empname,
           e.salary,
           max(ea.stamp) AS last_updated
      FROM emp e
      LEFT JOIN emp_audit ea ON ea.empname = e.empname
     GROUP BY 1, 2;

CREATE OR REPLACE FUNCTION update_emp_view() RETURNS TRIGGER AS $$
    BEGIN
        --
        -- Perform the required operation on emp, and create a row in
 emp_audit
        -- to reflect the change made to emp.
        --
        IF (TG_OP = 'DELETE') THEN
```

```
            DELETE FROM emp WHERE empname = OLD.empname;
            IF NOT FOUND THEN RETURN NULL; END IF;

            OLD.last_updated = now();
            INSERT INTO emp_audit VALUES('D', user, OLD.*);
            RETURN OLD;
        ELSIF (TG_OP = 'UPDATE') THEN
            UPDATE emp SET salary = NEW.salary WHERE empname = OLD.
empname;
            IF NOT FOUND THEN RETURN NULL; END IF;

            NEW.last_updated = now();
            INSERT INTO emp_audit VALUES('U', user, NEW.*);
            RETURN NEW;
        ELSIF (TG_OP = 'INSERT') THEN
            INSERT INTO emp VALUES(NEW.empname, NEW.salary);

            NEW.last_updated = now();
            INSERT INTO emp_audit VALUES('I', user, NEW.*);
            RETURN NEW;
        END IF;
    END;
$$ LANGUAGE plpgsql;

CREATE TRIGGER emp_audit
INSTEAD OF INSERT OR UPDATE OR DELETE ON emp_view
    FOR EACH ROW EXECUTE PROCEDURE update_emp_view();
```

触发器的一种用途是维护另一个表的汇总表。
得到的汇总结果可以在某些查询中替代原始表使用，而且通常能大幅缩短运行时间。
这种技术在数据仓库中很常见，其中用于度量或观察数据的表（称为事实表）可能会非常大。
例 39.6展示了PL/pgSQL中一个为数据仓库事实表维护汇总表的触发器函数示例。

例 39.6. 一个用于维护汇总表的 PL/pgSQL 触发器函数

这里详述的模式有一部分是基于 Ralph Kimball 所作的数据仓库工具包中的Grocery Store示例。

```
--
-- Main tables - time dimension and sales fact.
--
CREATE TABLE time_dimension (
    time_key                    integer NOT NULL,
    day_of_week                 integer NOT NULL,
    day_of_month                integer NOT NULL,
    month                       integer NOT NULL,
    quarter                     integer NOT NULL,
    year                        integer NOT NULL
);
CREATE UNIQUE INDEX time_dimension_key ON time_dimension(time_key);

CREATE TABLE sales_fact (
    time_key                    integer NOT NULL,
    product_key                 integer NOT NULL,
    store_key                   integer NOT NULL,
```

```
    amount_sold                 numeric(12,2) NOT NULL,
    units_sold                  integer NOT NULL,
    amount_cost                 numeric(12,2) NOT NULL
);
CREATE INDEX sales_fact_time ON sales_fact(time_key);

--
-- Summary table - sales by time.
--
CREATE TABLE sales_summary_bytime (
    time_key                    integer NOT NULL,
    amount_sold                 numeric(15,2) NOT NULL,
    units_sold                  numeric(12) NOT NULL,
    amount_cost                 numeric(15,2) NOT NULL
);
CREATE UNIQUE INDEX sales_summary_bytime_key ON sales_summary_bytime(
time_key);

--
-- Function and trigger to amend summarized column(s) on UPDATE,
 INSERT, DELETE.
--
CREATE OR REPLACE FUNCTION maint_sales_summary_bytime() RETURNS
 TRIGGER
AS $maint_sales_summary_bytime$
    DECLARE
        delta_time_key          integer;
        delta_amount_sold       numeric(15,2);
        delta_units_sold        numeric(12);
        delta_amount_cost       numeric(15,2);
    BEGIN

        -- Work out the increment/decrement amount(s).
        IF (TG_OP = 'DELETE') THEN

            delta_time_key = OLD.time_key;
            delta_amount_sold = -1 * OLD.amount_sold;
            delta_units_sold = -1 * OLD.units_sold;
            delta_amount_cost = -1 * OLD.amount_cost;

        ELSIF (TG_OP = 'UPDATE') THEN

            -- forbid updates that change the time_key -
            -- (probably not too onerous, as DELETE + INSERT is how
 most
            -- changes will be made).
            IF ( OLD.time_key != NEW.time_key) THEN
                RAISE EXCEPTION 'Update of time_key : % -> % not
 allowed',
                                                        OLD.time_key,
 NEW.time_key;
            END IF;

            delta_time_key = OLD.time_key;
```

```
            delta_amount_sold = NEW.amount_sold - OLD.amount_sold;
            delta_units_sold = NEW.units_sold - OLD.units_sold;
            delta_amount_cost = NEW.amount_cost - OLD.amount_cost;

        ELSIF (TG_OP = 'INSERT') THEN

            delta_time_key = NEW.time_key;
            delta_amount_sold = NEW.amount_sold;
            delta_units_sold = NEW.units_sold;
            delta_amount_cost = NEW.amount_cost;

        END IF;


        -- Insert or update the summary row with the new values.
        <<insert_update>>
        LOOP
            UPDATE sales_summary_bytime
                SET amount_sold = amount_sold + delta_amount_sold,
                    units_sold = units_sold + delta_units_sold,
                    amount_cost = amount_cost + delta_amount_cost
                WHERE time_key = delta_time_key;

            EXIT insert_update WHEN found;

            BEGIN
                INSERT INTO sales_summary_bytime (
                            time_key,
                            amount_sold,
                            units_sold,
                            amount_cost)
                    VALUES (
                            delta_time_key,
                            delta_amount_sold,
                            delta_units_sold,
                            delta_amount_cost
                           );

                EXIT insert_update;

            EXCEPTION
                WHEN UNIQUE_VIOLATION THEN
                    -- do nothing
            END;
        END LOOP insert_update;

        RETURN NULL;

    END;
$maint_sales_summary_bytime$ LANGUAGE plpgsql;

CREATE TRIGGER maint_sales_summary_bytime
AFTER INSERT OR UPDATE OR DELETE ON sales_fact
    FOR EACH ROW EXECUTE PROCEDURE maint_sales_summary_bytime();
```

```
INSERT INTO sales_fact VALUES(1,1,1,10,3,15);
INSERT INTO sales_fact VALUES(1,2,1,20,5,35);
INSERT INTO sales_fact VALUES(2,2,1,40,15,135);
INSERT INTO sales_fact VALUES(2,3,1,10,1,13);
SELECT * FROM sales_summary_bytime;
DELETE FROM sales_fact WHERE product_key = 1;
SELECT * FROM sales_summary_bytime;
UPDATE sales_fact SET units_sold = units_sold * 2;
SELECT * FROM sales_summary_bytime;
```

39.10. PL/pgSQL 内部机制

本节讨论一些实现细节，了解这些细节对 PL/pgSQL 用户通常很重要。

39.10.1. 变量替换

在PL/pgSQL函数中，SQL 语句和表达式可以引用函数的变量和参数。在内部，PL/pgSQL会用查询参数替换这些引用。只有在语法上允许参数或列引用的位置才会进行参数替换。
作为一个极端例子，来看下面这种不良编程风格：

```
INSERT INTO foo (foo) VALUES (foo);
```

第一次出现的`foo`在语法上必须是表名，所以不会被替换，即使函数有一个变量也叫`foo`。
第二次出现的位置必须是该表的列名，所以也不会被替换。
只有第三次出现的位置才可能是对函数变量的引用。

> **注意**
>
> PostgreSQL 9.0 以前的版本会尝试在这三个位置都进行变量替换，从而导致语法错误。

由于变量名在语法上与表列名没有区别，
所以在同时引用表的语句中就可能产生歧义：某个给定名称到底是指表列，
还是变量？把前面的示例改成下面这样：

```
INSERT INTO dest (col) SELECT foo + bar FROM src;
```

这里，`dest` 和 `src` 必须是表名，`col` 也必须是 `dest` 的一列，但 `foo` 和 `bar` 既可能是该函数的变量，也可能是 `src` 的列。

默认情况下，如果 SQL 语句中的名称既可能指变量，也可能指表列，PL/pgSQL就会报告错误。
可以通过重命名变量或列、限定有歧义的引用，或者告诉
PL/pgSQL优先采用哪种解释来解决这类问题。

最简单的解决方案是重命名变量或列。
一种常用的编码规则是为PL/pgSQL变量使用一种不同于列名的命名习惯。例如，
如果你将函数变量统一地命名为`v_`*`something`*，而你的列名不会开始于`v_`，就不会发生冲突。

另外你可以限定有歧义的引用让它们变清晰。在上面的示例中，`src.foo`
将是对表列的一种无歧义的引用。要创建对一个变量的无歧义引用，
在一个被标记的块中声明它并且使用块的标签（见第 39.2 节）。例如

```
<<block>>
DECLARE
    foo int;
BEGIN
    foo := ...;
    INSERT INTO dest (col) SELECT block.foo + bar FROM src;
```

这里block.foo表示变量，即使在src中有一个列foo。函数参数以及诸如FOUND的特殊变量，都能通过函数的名称被限定，因为它们被隐式地声明在一个带有该函数名称的外层块中。

有时候在一个大型的PL/pgSQL代码体中修复所有的有歧义引用是不现实的。在这种情况下，你可以指定PL/pgSQL应该将有歧义的引用解析为变量（这与PL/pgSQL在 PostgreSQL 9.0 之前的行为兼容）或表列（这与某些其他系统兼容，例如Oracle）。

要在系统范围内改变这种行为，将配置参数 plpgsql.variable_conflict 设置为 error、use_variable 或者 use_column（这里 error 是出厂设置）之一。这个参数会影响 PL/pgSQL 函数中语句的后续编译，但是不会影响在当前会话中已经编译过的语句。要在 PL/pgSQL 装载之前设置该参数，必须先把“plpgsql”加入 postgresql.conf 中的custom_variable_classes列表。因为改变这个设置能够导致 PL/pgSQL 函数中行为的意想不到的改变，所以只能由一个超级用户来更改它。

你也可以按函数单独设置这一行为，方法是在函数文本开头插入以下特殊命令之一：

```
#variable_conflict error
#variable_conflict use_variable
#variable_conflict use_column
```

这些命令只影响它们所属的函数，并且会覆盖plpgsql.variable_conflict的设置。一个示例是：

```
CREATE FUNCTION stamp_user(id int, comment text) RETURNS void AS $$
    #variable_conflict use_variable
    DECLARE
        curtime timestamp := now();
    BEGIN
        UPDATE users SET last_modified = curtime, comment = comment
          WHERE users.id = id;
    END;
$$ LANGUAGE plpgsql;
```

在UPDATE命令中，curtime、comment以及id将引用该函数的变量和参数，不管users有没有这些名称的列。注意，我们不得不在WHERE子句中对users.id的引用加以限定，以便让它引用表列。但是我们不需要对UPDATE列表中作为目标的comment引用加以限定，因为语法上那必须是users的一列。我们可以用下面的方式写一个相同的不依赖于variable_conflict设置的函数：

```
CREATE FUNCTION stamp_user(id int, comment text) RETURNS void AS $$
    <<fn>>
    DECLARE
        curtime timestamp := now();
    BEGIN
        UPDATE users SET last_modified = fn.curtime, comment = stamp_
user.comment
          WHERE users.id = stamp_user.id;
```

```
    END;
$$ LANGUAGE plpgsql;
```

传递给 EXECUTE 及其变体的命令字符串中，不会发生变量替换。如果你需要向这种命令中插入变化的值，应在构造字符串值时完成，或者像第 39.5.4 节所说明的那样使用 USING。

目前，变量替换只在SELECT、INSERT、UPDATE和DELETE命令中生效，因为主 SQL 引擎只允许在这些命令中使用查询参数。若要在其他语句类型（统称为工具语句）中使用非常量名称或值，就必须把该工具语句构造为字符串，再用EXECUTE执行。

39.10.2. 计划缓存

在函数第一次被调用时（每个会话中都会如此），PL/pgSQL解释器会解析函数源文本，并生成一棵内部的二进制指令树。该指令树完整表示了PL/pgSQL语句结构，但函数中使用的各个SQL表达式和SQL命令并不会立即被分析。

函数中的每个表达式和SQL命令首次执行时，PL/pgSQL解释器都会创建一个预备好的执行计划（使用SPI管理器的SPI_prepare和SPI_saveplan函数）。之后对该表达式或命令的访问会复用该预备计划。因此，一个包含大量可能需要执行计划的语句的条件代码函数，在数据库连接的生存期内只会准备和保存真正用到的那些计划。这可以显著减少为PL/pgSQL函数中的语句解析和生成执行计划所需的总时间。一个不利之处是，特定表达式或命令中的错误要到函数执行到相应部分时才会被发现。（琐碎的语法错误会在初始解析阶段被发现，但更深层的错误要到执行时才会被发现。）

如果查询所用任何表的架构发生改变，或者查询中用到的任何用户定义函数被重新定义，已保存的计划将被自动重新规划。这使预备计划的重用在大多数情况下是透明的，但也存在可能重用过期计划的边角情况。例如，删除并重新创建一个用户定义操作符不会影响已缓存的计划；如果该操作符的底层函数没有被改变，这些计划会继续调用原来的底层函数。必要时，可以通过启动一个新的数据库会话来刷新缓存。

由于 PL/pgSQL 以这种方式保存执行计划，直接出现在 PL/pgSQL 函数中的 SQL 命令在每次执行时必须引用相同的表和列；也就是说，不能在 SQL 命令中把参数用作表名或列名。要绕过这一限制，可以使用 PL/pgSQL 的 EXECUTE 语句构造动态命令 – 代价是每次执行都要构造一个新的执行计划。

另一个要点是，预备计划是参数化的，允许 PL/pgSQL 变量的值在一次使用与下一次使用之间变化，如前文详细讨论的那样。有时这意味着计划不如按特定变量值生成的计划高效。举例来说，考虑

```
SELECT * INTO myrec FROM dictionary WHERE word LIKE search_term;
```

其中 search_term 是一个 PL/pgSQL 变量。该查询的缓存计划永远不会使用 word 上的索引，因为规划器不能假定 LIKE 模式在运行时是左锚定的。要使用索引，必须在规划查询时提供具体的常量 LIKE 模式。这是另一种情况，可以用 EXECUTE 强制为每次执行生成新计划。

记录变量的可变特性在这里还会带来另一个问题。当记录变量的字段被用于表达式或语句中时，这些字段的数据类型不能在函数的不同调用之间发生变化，因为每个表达式都会按照第一次执行到它时所看到的数据类型来规划。必要时，可以用 EXECUTE 绕过这个问题。

如果同一个函数被用作多个表的触发器，PL/pgSQL 会针对每个这样的表独立地准备并缓存计划 – 也就是说，缓存是按触发器函数 + 表的组合建立的，而不是每个函数只有一个缓存。这缓解了

数据类型变化带来的部分问题；例如，即使不同表中名为 key 的列类型不同，一个触发器函数也仍然能够成功使用它。

同样，
具有多态参数类型的函数也会为它们已经被调用的每一种实参类型组合都保留一个独立的缓存，这样数据类型差异不会导致意想不到的失败。

计划缓存有时可能在对时间敏感的值的解释上产生令人惊讶的效果。
例如这两个函数做的事情就有区别：

```
CREATE FUNCTION logfunc1(logtxt text) RETURNS void AS $$
    BEGIN
        INSERT INTO logtable VALUES (logtxt, 'now');
    END;
$$ LANGUAGE plpgsql;
```

and:

```
CREATE FUNCTION logfunc2(logtxt text) RETURNS void AS $$
    DECLARE
        curtime timestamp;
    BEGIN
        curtime := 'now';
        INSERT INTO logtable VALUES (logtxt, curtime);
    END;
$$ LANGUAGE plpgsql;
```

在 logfunc1 中，PostgreSQL 的主解析器在为 INSERT 准备计划时就知道字符串 'now' 应该被解释为 timestamp，因为 logtable 的目标列是这种类型。因此，'now' 将在 INSERT 被规划时转换为一个常量，然后在该会话的生命周期内被用于所有对 logfunc1 的调用。不用说，这不是程序员想要的。

在 logfunc2 中，PostgreSQL 的主解析器不知道 'now' 应该变成什么类型，因此返回一个 text 类型的数据值，其中包含字符串 now。在随后给局部变量 curtime 赋值时，PL/pgSQL 解释器通过调用用于该转换的 text_out 和 timestamp_in 函数，把这个字符串转换为 timestamp 类型。因此，计算得到的时间戳会按程序员预期在每次执行时更新。

39.11. PL/pgSQL开发提示

在PL/pgSQL中进行开发的一种好方法，是用你喜欢的文本编辑器编写函数，
并在另一个窗口里使用psql载入和测试这些函数。如果你采用这种方式，最好使用CREATE OR REPLACE FUNCTION来编写函数。这样只需重新载入该文件，就可以更新函数定义。例如：

```
CREATE OR REPLACE FUNCTION testfunc(integer) RETURNS integer AS $$
          ....
$$ LANGUAGE plpgsql;
```

在运行psql期间，你可以用下面的命令载入或者重载这样一个函数定义文件：

```
\i filename.sql
```

并且接着立即发出 SQL 命令来测试该函数。

另一种使用PL/pgSQL进行开发的好方法，是使用有助于过程语言开发的 GUI 数据库访问工具。
这类工具的一个例子是pgAdmin，当然也有其他工具。它们通常提供一些方便的特性，
例如转义单引号，以及让重新创建和调试函数变得更容易。

39.11.1. 引号的处理

一个PL/pgSQL函数的代码在一个CREATE FUNCTION中被指定为一个字符串字面量。如果你用通常的方式把该字符串写在单引号中间，那么该函数体中的任何单引号都必须被双写；同样任何反斜线也必须被双写（假定使用了转义字符串语法）。双写引号本身就很繁琐，并且在更复杂的情况中代码会变得完全无法理解，因为你很容易发现你需要半打或者更多相邻的引号。我们推荐你转而把函数体写成一个“美元引用”的字符串（见第 4.1.2.4 节）。在美元引用方法中，你从不需要双写任何引号。但是要注意为你需要的每一层嵌套选择一个不同的美元引用定界符。例如，你可能把CREATE FUNCTION命令写成：

```
CREATE OR REPLACE FUNCTION testfunc(integer) RETURNS integer AS $PROC$
          ....
$PROC$ LANGUAGE plpgsql;
```

在这里面，你可以在 SQL 命令中为简单字符串使用引号并且用$$来界定被你组装成字符串的 SQL 命令片段。如果你需要引用包括$$的文本，你可以使用Q等等。

下表展示了在不使用美元引用时，编写引号必须采取的做法。在把旧式的非美元引用代码改写成更易理解的形式时，它可能会有所帮助。

1 个引号

用来开始和结束函数体，例如：

```
CREATE FUNCTION foo() RETURNS integer AS '
          ....
' LANGUAGE plpgsql;
```

在一个单引号引用的函数体中的任何位置，引号必须成对出现。

2 个引号

用于函数体内的字符串字面量，例如：

```
a_output := ''Blah'';
SELECT * FROM users WHERE f_name=''foobar'';
```

在美元引用方法中，你只需要写：

```
a_output := 'Blah';
SELECT * FROM users WHERE f_name='foobar';
```

这恰好就是PL/pgSQL解析器在两种情况中会看到的。

4 个引号

当你在函数内的一个字符串常量中需要一个单引号时，例如：

```
a_output := a_output || '' AND name LIKE ''''foobar'''' AND xyz''
```

实际会被追加到a_output的值将是：AND name LIKE 'foobar' AND xyz。

在美元引用方法中，你可以写：

```
a_output := a_output || $$ AND name LIKE 'foobar' AND xyz$$
```

要小心在这周围的任何美元引用定界符都不能是`$$`。

6 个引号

当在函数体内的一个字符串中的一个单引号与该字符串常量末尾相邻，例如：

```
a_output := a_output || '' AND name LIKE ''''foobar''''''
```

被追加到`a_output`的值则将是：`AND name LIKE 'foobar'`。

在美元引用方法中，这会变成：

```
a_output := a_output || $$ AND name LIKE 'foobar'$$
```

10 个引号

当字符串常量中需要两个单引号（这需要 8 个引号），而且它们紧邻该字符串常量的末尾（还需 2 个引号）时。通常只有在编写生成其他函数的函数时（如例 39.8所示），才会需要这种写法。例如：

```
a_output := a_output || '' if v_'' ||
    referrer_keys.kind || '' like '''''''''
    || referrer_keys.key_string || '''''''''
    then return ''''''  || referrer_keys.referrer_type
    || ''''''; end if;'';
```

`a_output`的值将是：

```
if v_... like ''...'' then return ''...''; end if;
```

在美元引用方法中，这会变成：

```
a_output := a_output || $$ if v_$$ || referrer_keys.kind || $$ like
 '$$
    || referrer_keys.key_string || $$'
    then return '$$  || referrer_keys.referrer_type
    || $$'; end if;$$;
```

这里假定只需在`a_output`中放入单引号，因为在使用前会对它再次加引号。

39.12. 从Oracle PL/SQL 移植

这一节解释了PostgreSQL的PL/pgSQL语言和 Oracle 的PL/SQL语言之间的差别，用以帮助那些从Oracle®向PostgreSQL移植应用的人。

PL/pgSQL在许多方面都与 PL/SQL 类似。它是一种具有块结构的命令式语言，所有变量都必须声明。赋值、循环和条件语句也都很相似。在从PL/SQL移植到PL/pgSQL时，
应当记住以下主要差异：

- 如果 SQL 命令中使用的名字既可能是表的列名，也可能是对函数中变量的引用，PL/SQL会将其当作列名。这对应于PL/pgSQL的`plpgsql.variable_conflict = use_column`行为，这不是默认行为，如第 39.10.1 节所述。通常最好一开始就避免这种歧义，但如果必须移植大量依赖这种行为的代码，设置`variable_conflict`可能是最好的解决办法。

- 在PostgreSQL中，函数体必须写成字符串字面量。
  因此你需要使用美元符引用或者转义函数体中的单引号（见第 39.11.1 节）。

- 应该用模式把函数组织成不同的分组，而不是用包。
- 因为没有包，所以也没有包级变量。这一点有时会带来不便。
  你可以改为在临时表中保存会话级状态。
- 带有REVERSE的整数FOR循环的工作方式不同：PL/SQL中是从第二个数向第一个数倒数，而PL/pgSQL是从第一个数向第二个数倒数，因此在移植时需要交换循环边界。
  不幸的是这种不兼容性是不太可能改变的（见第 39.6.3.5 节）。
- 查询上的FOR循环（不是游标）的工作方式同样不同：目标变量必须已经被声明，而PL/SQL总是会隐式地声明它们。但是这样做的优点是在退出循环后，变量值仍然可以访问。
- 应该用模式把函数组织成不同的分组，而不是用包。

39.12.1. 移植示例

例 39.7展示了如何从PL/SQL移植一个简单的函数到PL/pgSQL中。

例 39.7. 把一个简单函数从 PL/SQL 移植到 PL/pgSQL

这里有一个Oracle PL/SQL函数：

```
CREATE OR REPLACE FUNCTION cs_fmt_browser_version(v_name varchar,
                                                  v_version varchar)
RETURN varchar IS
BEGIN
    IF v_version IS NULL THEN
        RETURN v_name;
    END IF;
    RETURN v_name || '/' || v_version;
END;
/
show errors;
```

让我们过一遍这个函数并且看看与PL/pgSQL相比有什么样的不同：

- 在函数原型中（不是函数体中）的RETURN关键字在PostgreSQL中变成了RETURNS。还有，IS变成了AS，并且你还需要增加一个LANGUAGE子句，因为PL/pgSQL并非唯一可用的函数语言。
- 在PostgreSQL中，函数体被认为是一个字符串字面量，所以你需要使用引号或者美元引用定界符包围它。这代替了Oracle 方法中的用于终止的/。
- 在PostgreSQL中没有show errors命令，并且也不需要这个命令，因为错误是自动报告的。

这个函数被移植到PostgreSQL后看起来会是这样：

```
CREATE OR REPLACE FUNCTION cs_fmt_browser_version(v_name varchar,
                                                  v_version varchar)
RETURNS varchar AS $$
BEGIN
    IF v_version IS NULL THEN
        RETURN v_name;
    END IF;
    RETURN v_name || '/' || v_version;
```

```
END;
$$ LANGUAGE plpgsql;
```

例 39.8展示了如何移植一个会创建另一个函数的函数，以及如何处理引号问题。

例 39.8. 把一个创建其它函数的函数从 PL/SQL 移植到 PL/pgSQL

下面的过程从 `SELECT` 语句读取行，并将结果写入 `IF` 语句，从而构造一个大型函数，以提高效率。

这是 Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_update_referrer_type_proc IS
    CURSOR referrer_keys IS
        SELECT * FROM cs_referrer_keys
        ORDER BY try_order;
    func_cmd VARCHAR(4000);
BEGIN
    func_cmd := 'CREATE OR REPLACE FUNCTION cs_find_referrer_type(v_
host IN VARCHAR,
                 v_domain IN VARCHAR, v_url IN VARCHAR) RETURN VARCHAR
 IS BEGIN';

    FOR referrer_key IN referrer_keys LOOP
        func_cmd := func_cmd ||
          ' IF v_' || referrer_key.kind
          || ' LIKE ''' || referrer_key.key_string
          || ''' THEN RETURN ''' || referrer_key.referrer_type
          || '''; END IF;';
    END LOOP;

    func_cmd := func_cmd || ' RETURN NULL; END;';

    EXECUTE IMMEDIATE func_cmd;
END;
/
show errors;
```

下面是这个函数的最终移植结果，目标数据库为PostgreSQL：

```
CREATE OR REPLACE FUNCTION cs_update_referrer_type_proc() RETURNS void
 AS $func$
DECLARE
    referrer_keys CURSOR IS
        SELECT * FROM cs_referrer_keys
        ORDER BY try_order;
    func_body text;
    func_cmd text;
BEGIN
    func_body := 'BEGIN';

    FOR referrer_key IN referrer_keys LOOP
        func_body := func_body ||
          ' IF v_' || referrer_key.kind
          || ' LIKE ' || quote_literal(referrer_key.key_string)
```

```
          || ' THEN RETURN ' || quote_literal(referrer_key.referrer_
type)
          || '; END IF;' ;
    END LOOP;

    func_body := func_body || ' RETURN NULL; END;';

    func_cmd :=
      'CREATE OR REPLACE FUNCTION cs_find_referrer_type(v_host
 varchar,
                                                        v_domain
 varchar,
                                                        v_url varchar)
        RETURNS varchar AS '
      || quote_literal(func_body)
      || ' LANGUAGE plpgsql;' ;

    EXECUTE func_cmd;
END;
$func$ LANGUAGE plpgsql;
```

注意，这里单独构造函数体，然后将其传给quote_literal，使其中的每个引号都变成两个。这种技术是必需的，
因为不能安全地使用美元引用来定义新函数：我们无法确定会插入什么字符串，其来源是referrer_key.key_string字段。（这里假定referrer_key.kind可信，其值总是host，domain或url，但是referrer_key.key_string可能是任何内容，
尤其可能包含美元符号。）这个函数实际上改进了 Oracle 原版：当referrer_key.key_string或referrer_key.referrer_type中包含引号时，它也不会生成有问题的代码。

例 39.9展示了如何移植一个带有OUT参数和字符串处理的函数。PostgreSQL没有内置的instr函数，但是你可以用其它函数的组合来创建一个。在第 39.12.3 节中有一个instr的PL/pgSQL实现，你可以用它让你的移植变得更容易。

例 39.9. 把一个带字符串处理和 OUT 参数的过程从 PL/SQL 移植到 PL/pgSQL

下面的Oracle PL/SQL 过程被用来解析一个 URL 并且返回一些元素（主机、路径和查询）。

这是 Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_update_referrer_type_proc IS
    CURSOR referrer_keys IS
        SELECT * FROM cs_referrer_keys
        ORDER BY try_order;
    func_cmd VARCHAR(4000);
BEGIN
    func_cmd := 'CREATE OR REPLACE FUNCTION cs_find_referrer_type(v_
host IN VARCHAR,
                 v_domain IN VARCHAR, v_url IN VARCHAR) RETURN VARCHAR
 IS BEGIN';

    FOR referrer_key IN referrer_keys LOOP
        func_cmd := func_cmd ||
          ' IF v_' || referrer_key.kind
          || ' LIKE ''' || referrer_key.key_string
          || ''' THEN RETURN ''' || referrer_key.referrer_type
```

```
            || '''; END IF;';
    END LOOP;

    func_cmd := func_cmd || ' RETURN NULL; END;';

    EXECUTE IMMEDIATE func_cmd;
END;
/
show errors;
```

这里给出一种可能的 PL/pgSQL 写法：

```
CREATE OR REPLACE FUNCTION cs_parse_url(
    v_url IN VARCHAR,
    v_host OUT VARCHAR,  -- 这个值将被返回
    v_path OUT VARCHAR,  -- 这个也是
    v_query OUT VARCHAR) -- 还有这个
AS $$
DECLARE
    a_pos1 INTEGER;
    a_pos2 INTEGER;
BEGIN
    v_host := NULL;
    v_path := NULL;
    v_query := NULL;
    a_pos1 := instr(v_url, '//');

    IF a_pos1 = 0 THEN
        RETURN;
    END IF;
    a_pos2 := instr(v_url, '/', a_pos1 + 2);
    IF a_pos2 = 0 THEN
        v_host := substr(v_url, a_pos1 + 2);
        v_path := '/';
        RETURN;
    END IF;

    v_host := substr(v_url, a_pos1 + 2, a_pos2 - a_pos1 - 2);
    a_pos1 := instr(v_url, '?', a_pos2 + 1);

    IF a_pos1 = 0 THEN
        v_path := substr(v_url, a_pos2);
        RETURN;
    END IF;

    v_path := substr(v_url, a_pos2, a_pos1 - a_pos2);
    v_query := substr(v_url, a_pos1 + 1);
END;
$$ LANGUAGE plpgsql;
```

这个函数可以这样使用：

```
SELECT * FROM cs_parse_url('http://foobar.com/query.cgi?baz');
```

例 39.10展示了如何移植一个使用了多种 Oracle 专属特性的过程。

例 39.10. 把一个过程从 PL/SQL 移植到 PL/pgSQL

Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_create_job(v_job_id IN INTEGER) IS
    a_running_job_count INTEGER;
    PRAGMA AUTONOMOUS_TRANSACTION;❶
BEGIN
    LOCK TABLE cs_jobs IN EXCLUSIVE MODE;❷

    SELECT count(*) INTO a_running_job_count FROM cs_jobs WHERE end_
stamp IS NULL;

    IF a_running_job_count > 0 THEN
        COMMIT; -- free lock❸
        raise_application_error(-20000,
                 'Unable to create a new job: a job is currently
 running.');
    END IF;

    DELETE FROM cs_active_job;
    INSERT INTO cs_active_job(job_id) VALUES (v_job_id);

    BEGIN
        INSERT INTO cs_jobs (job_id, start_stamp) VALUES (v_job_id,
 now());
    EXCEPTION
        WHEN dup_val_on_index THEN NULL; -- don't worry if it already
 exists
    END;
    COMMIT;
END;
/
show errors
```

这样的过程可以很容易地转换为PostgreSQL中返回`void`的函数。这个过程尤其值得关注，因为它能说明以下几点：

❶ PostgreSQL中没有`PRAGMA`语句。
❷ 如果在PL/pgSQL中执行`LOCK TABLE`，锁要等到调用事务结束时才会释放。
❸ 不能在PL/pgSQL函数中发出`COMMIT`。函数运行在某个外层事务之中，因此`COMMIT`意味着终止函数的执行。不过，在这个特定情形中本来也不需要提交，因为抛出错误时，`LOCK TABLE`获得的锁就会被释放。

下面展示了如何移植这个过程，目标语言为PL/pgSQL：

```
CREATE OR REPLACE FUNCTION cs_create_job(v_job_id integer) RETURNS
 void AS $$
DECLARE
    a_running_job_count integer;
BEGIN
    LOCK TABLE cs_jobs IN EXCLUSIVE MODE;

    SELECT count(*) INTO a_running_job_count FROM cs_jobs WHERE end_
stamp IS NULL;
```

```
    IF a_running_job_count > 0 THEN
        RAISE EXCEPTION 'Unable to create a new job: a job is
 currently running';❶
    END IF;

    DELETE FROM cs_active_job;
    INSERT INTO cs_active_job(job_id) VALUES (v_job_id);

    BEGIN
        INSERT INTO cs_jobs (job_id, start_stamp) VALUES (v_job_id,
 now());
    EXCEPTION
        WHEN unique_violation THEN ❷
            -- don't worry if it already exists
    END;
END;
$$ LANGUAGE plpgsql;
```

❶ RAISE的语法与 Oracle 的语句相当不同，尽管基本的形式RAISE *exception_name*工作起来是相似的。

❷ PL/pgSQL所支持的异常名称不同于 Oracle。内置的异常名称集合要更大（见附录 A）。目前没有办法声明用户定义的异常名称，尽管你能够抛出用户选择的 SQLSTATE 值。

The main functional difference between this procedure and the Oracle equivalent is that the exclusive lock on the `cs_jobs` table will be held until the calling transaction completes. Also, if the caller later aborts (for example due to an error), the effects of this procedure will be rolled back.

39.12.2. 其他要关注的事项

这一节解释了在移植 Oracle PL/SQL函数到PostgreSQL中时要关注的一些其他问题。

39.12.2.1. 异常后隐式回滚

在PL/pgSQL中，当异常被EXCEPTION子句捕获后，自该块的BEGIN
以来所做的所有数据库更改都会被自动回滚。也就是说，这种行为等效于你在 Oracle 中使用下面的代码所得到的效果：

```
BEGIN
    SAVEPOINT s1;
    ... code here ...
EXCEPTION
    WHEN ... THEN
        ROLLBACK TO s1;
        ... code here ...
    WHEN ... THEN
        ROLLBACK TO s1;
        ... code here ...
END;
```

如果你正在移植一个以这种方式使用SAVEPOINT和ROLLBACK TO的 Oracle 过程，工作会比较简单：只要省略SAVEPOINT和ROLLBACK TO即可。如果你的 Oracle 过程以不同方式使用SAVEPOINT和ROLLBACK TO，那就需要认真思考了。

39.12.2.2. EXECUTE

PL/pgSQL中的EXECUTE与PL/SQL中的版本工作方式相似，但必须记得按照第 39.5.4 节中的说明使用quote_literal和quote_ident。像 EXECUTE 'SELECT * FROM $1'; 这样的写法如果不借助这些函数，将无法可靠地工作。

39.12.2.3. 优化 PL/pgSQL 函数

PostgreSQL提供了两种函数创建修饰符来优化执行：“volatility”（对于给定的相同参数，函数是否总是返回相同的结果）以及“strictness”（如果任何参数为空值，函数是否返回空值）。详见CREATE FUNCTION参考页。

在利用这些优化属性时，你的CREATE FUNCTION语句可能像这样：

```
CREATE FUNCTION foo(...) RETURNS integer AS $$
...
$$ LANGUAGE plpgsql STRICT IMMUTABLE;
```

39.12.3. 附录

这一节包含了一组 Oracle 兼容的instr函数代码，你可以用它来简化你的移植工作。

```
--
-- instr functions that mimic Oracle's counterpart
-- Syntax: instr(string1, string2, [n], [m]) where [] denotes optional
 parameters.
--
-- Searches string1 beginning at the nth character for the mth
 occurrence
-- of string2.  If n is negative, search backwards.  If m is not
 passed,
-- assume 1 (search starts at first character).
--

CREATE FUNCTION instr(varchar, varchar) RETURNS integer AS $$
DECLARE
    pos integer;
BEGIN
    pos:= instr($1, $2, 1);
    RETURN pos;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;


CREATE FUNCTION instr(string varchar, string_to_search varchar, beg_
index integer)
RETURNS integer AS $$
DECLARE
    pos integer NOT NULL DEFAULT 0;
    temp_str varchar;
    beg integer;
    length integer;
    ss_length integer;
BEGIN
```

```
    IF beg_index > 0 THEN
        temp_str := substring(string FROM beg_index);
        pos := position(string_to_search IN temp_str);

        IF pos = 0 THEN
            RETURN 0;
        ELSE
            RETURN pos + beg_index - 1;
        END IF;
    ELSE
        ss_length := char_length(string_to_search);
        length := char_length(string);
        beg := length + beg_index - ss_length + 2;

        WHILE beg > 0 LOOP
            temp_str := substring(string FROM beg FOR ss_length);
            pos := position(string_to_search IN temp_str);

            IF pos > 0 THEN
                RETURN beg;
            END IF;

            beg := beg - 1;
        END LOOP;

        RETURN 0;
    END IF;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;


CREATE FUNCTION instr(string varchar, string_to_search varchar,
                      beg_index integer, occur_index integer)
RETURNS integer AS $$
DECLARE
    pos integer NOT NULL DEFAULT 0;
    occur_number integer NOT NULL DEFAULT 0;
    temp_str varchar;
    beg integer;
    i integer;
    length integer;
    ss_length integer;
BEGIN
    IF beg_index > 0 THEN
        beg := beg_index;
        temp_str := substring(string FROM beg_index);

        FOR i IN 1..occur_index LOOP
            pos := position(string_to_search IN temp_str);

            IF i = 1 THEN
                beg := beg + pos - 1;
            ELSE
                beg := beg + pos;
```

```
            END IF;

            temp_str := substring(string FROM beg + 1);
        END LOOP;

        IF pos = 0 THEN
            RETURN 0;
        ELSE
            RETURN beg;
        END IF;
    ELSE
        ss_length := char_length(string_to_search);
        length := char_length(string);
        beg := length + beg_index - ss_length + 2;

        WHILE beg > 0 LOOP
            temp_str := substring(string FROM beg FOR ss_length);
            pos := position(string_to_search IN temp_str);

            IF pos > 0 THEN
                occur_number := occur_number + 1;

                IF occur_number = occur_index THEN
                    RETURN beg;
                END IF;
            END IF;

            beg := beg - 1;
        END LOOP;

        RETURN 0;
    END IF;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;
```

第 40 章 PL/Tcl - Tcl 过程语言

PL/Tcl 是 PostgreSQL 数据库系统的一种可加载过程语言，可以使用 Tcl 语言[1]编写函数和触发器函数。

40.1. 概述

PL/Tcl提供了函数编写者在C语言中所具备的大部分能力，但会施加少量限制，并额外提供Tcl可用的强大字符串处理库。

其中一个颇具吸引力的良性限制是，所有内容都在 Tcl 解释器的安全上下文中执行。除了安全 Tcl 受限的命令集之外，只提供了少数通过 SPI 访问数据库以及通过`elog()`发出消息的命令。与 C 函数不同，PL/Tcl 不提供访问数据库服务器内部机制的方法，也不能在PostgreSQL服务器进程的权限下获得操作系统级访问。因此，可以信任非特权数据库用户使用这种语言；
它不会授予他们无限制的权限。

另一项值得注意的实现限制是，Tcl 函数不能用来为新数据类型创建输入/输出函数。

有时需要编写不受安全 Tcl 限制的 Tcl 函数。例如，可能希望某个 Tcl 函数能够发送电子邮件。为处理这类情况，PL/Tcl 提供了一个名为`PL/TclU`的变体（其中 U 表示不受信任的 Tcl）。除了使用完整的 Tcl 解释器之外，它与前者是完全相同的语言。如果使用PL/TclU，就必须将其安装为一种不受信任的过程语言，这样就只有数据库超级用户才能在其中创建函数。
PL/TclU函数的编写者必须确保该函数不会被用来做任何不希望发生的事情，
因为它能够执行任何以数据库管理员身份登录的用户所能做的事情。

如果在安装过程的配置步骤中指定了 Tcl 支持，则PL/Tcl和PL/TclU调用处理器的共享对象代码会自动构建并安装到PostgreSQL的库目录中。要在某个特定数据库中安装PL/Tcl和/或PL/TclU，请使用`CREATE EXTENSION`命令或`createlang`程序，例如`createlang pltcl *dbname*`或`createlang pltclu *dbname*`。

40.2. PL/Tcl 函数和参数

要用PL/Tcl语言创建函数，可使用标准的CREATE FUNCTION语法：

```
CREATE FUNCTION funcname (argument-types) RETURNS return-type AS $$
    # PL/Tcl 函数体
$$ LANGUAGE pltcl;
```

PL/TclU的写法相同，只是语言必须指定为`pltclu`。

函数体就是一段 Tcl 脚本。调用函数时，参数值以变量 `$1` ... `$n` 的形式传入 Tcl 脚本。Tcl 代码按通常方式使用 `return` 语句返回结果。

例如，一个返回两个整数值中较大值的函数可以定义为：

```
CREATE FUNCTION tcl_max(integer, integer) RETURNS integer AS $$
    if {$1 > $2} {return $1}
    return $2
$$ LANGUAGE pltcl STRICT;
```

注意`STRICT`子句，它让我们不必考虑空值输入：如果传入的是空值，函数根本不会被调用，
而是会自动返回空值结果。

[1] http://www.tcl.tk/

在非严格函数中，如果某个参数的实际值为空值，对应的$*n*变量会被设置为空串。
要检测某个特定参数是否为空值，可使用函数argisnull。例如，假设我们希望在tcl_max的两个参数中一个为空值、一个非空值时返回非空值参数，而不是返回空值：

```
CREATE FUNCTION tcl_max(integer, integer) RETURNS integer AS $$
    if {[argisnull 1]} {
        if {[argisnull 2]} { return_null }
        return $2
    }
    if {[argisnull 2]} { return $1 }
    if {$1 > $2} {return $1}
    return $2
$$ LANGUAGE pltcl;
```

如上所示，要从 PL/Tcl 函数返回空值，请执行return_null。无论函数是否为严格函数，都可以这样做。

复合类型参数会作为 Tcl 数组传递给函数。数组元素名就是该复合类型的属性名。如果传入行中的某个属性为空值，它就不会出现在数组中。下面是一个示例：

```
CREATE TABLE employee (
    name text,
    salary integer,
    age integer
);

CREATE FUNCTION overpaid(employee) RETURNS boolean AS $$
    if {200000.0 < $1(salary)} {
        return "t"
    }
    if {$1(age) < 30 && 100000.0 < $1(salary)} {
        return "t"
    }
    return "f"
$$ LANGUAGE pltcl;
```

目前尚不支持返回复合类型的结果值，也不支持返回集合。

PL/Tcl目前对域类型还没有完整支持：它把域当作底层标量类型同样对待。
这意味着与该域关联的约束不会被强制执行。对于函数参数来说这不是问题，但如果把PL/Tcl 函数声明为返回域类型，就会带来隐患。

40.3. PL/Tcl 中的数据值

提供给 PL/Tcl 函数代码的参数值，只是将输入参数转换成文本形式（就像用SELECT语句将它们显示出来一样）。反过来，return命令会接受任何字符串，
只要它是该函数声明返回类型的可接受输入格式。因此，在 PL/Tcl 函数内部，所有值都只是文本字符串。

40.4. PL/Tcl 中的全局数据

有时需要在一次函数调用结束后保留某些全局数据，供下一次调用继续使用，
或者在不同函数之间共享全局数据。在 PL/Tcl 中这很容易做到，但必须理解其中的一些限制。

出于安全原因，PL/Tcl 会针对每个 SQL 角色，在该角色各自独立的 Tcl 解释器中执行其调用的函数。这样可以防止一个用户无意或恶意地干扰另一用户的 PL/Tcl 函数行为。每个这样的解释器都为任何 “global” Tcl 变量维护各自的值。因此，当且仅当两个 PL/Tcl 函数由同一个 SQL 角色执行时，它们才会共享同一组全局变量。在某些应用中，一个会话会在多个 SQL 角色下执行代码（例如通过`SECURITY DEFINER`函数、使用`SET ROLE`等），这时你可能需要显式采取一些步骤，确保 PL/Tcl 函数能够共享数据。为此，应确保需要互相通信的函数由同一用户拥有，并将它们标记为`SECURITY DEFINER`。当然，你必须小心，确保这类函数不会被用来执行任何非预期操作。

一个会话中使用的所有 PL/TclU 函数都在同一个 Tcl 解释器中执行，而这个解释器当然不同于 PL/Tcl 函数所使用的解释器。因此，PL/TclU 函数之间会自动共享全局数据。这不被视为安全风险，因为所有 PL/TclU 函数都在相同的信任级别上执行，也就是数据库超级用户的级别。

为帮助防止 PL/Tcl 函数无意间彼此干扰，每个函数都可通过`upvar`命令访问一个全局数组。该变量的全局名称是函数的内部名称，局部名称是`GD`。建议将`GD`用于保存函数的持久私有数据。
只有当你明确希望某些值在多个函数之间共享时，才应使用常规的 Tcl 全局变量。（注意`GD`数组只在某个特定解释器内部是全局的，因此它们不会绕过上文提到的安全限制。）

下面`spi_execp`的示例展示了`GD`的用法。

40.5. 从 PL/Tcl 访问数据库

在 PL/Tcl 函数体中，可以使用下列命令访问数据库：

`spi_exec ?-count n? ?-array name? command ?loop-body?`

执行以字符串形式给出的 SQL 命令。命令出错时会引发错误。否则，`spi_exec`的返回值是该命令处理的行数（选出、插入、更新或删除的行），如果命令是工具语句则返回零。
此外，如果命令是`SELECT`语句，则所选列的值会按下文所述放入 Tcl 变量中。

可选的 `-count` 值告诉 `spi_exec` 在该命令中最多处理多少行。其效果等同于把查询设置为一个游标，然后执行 `FETCH n`。

如果命令是 `SELECT` 语句，则结果列的值会放入以列名命名的 Tcl 变量中。如果给定了 `-array` 选项，则列值会存储在指定的关联数组元素中，列名用作数组索引。

如果命令是 `SELECT` 语句且未给出 *loop-body* 脚本，则只会把结果的第一行存入 Tcl 变量中；其余行如果存在，会被忽略。如果查询没有返回任何行，则不会进行任何存储。（这种情况可以通过检查 `spi_exec` 的结果来发现。）例如：

```
spi_exec "SELECT count(*) AS cnt FROM pg_proc"
```

会把 Tcl 变量 `$cnt` 设置为 `pg_proc` 系统目录中的行数。

如果给出了可选的 *loop-body* 参数，它就是一段 Tcl 脚本，对查询结果中的每一行执行一次。（如果给定的命令不是 `SELECT`，则忽略 *loop-body*。）在每次迭代之前，当前行各列的值会被存入 Tcl 变量。例如：

```
spi_exec -array C "SELECT * FROM pg_class" {
    elog DEBUG "have table $C(relname)"
}
```

会为 `pg_class` 的每一行打印一条日志消息。这一特性的工作方式类似于其他 Tcl 循环构造；特别是 `continue` 和 `break` 在循环体中按通常方式工作。

如果查询结果的某一列为空值，对应的目标变量将被“取消设置”而不是被设置。

`spi_prepare` *`query typelist`*

为以后的执行准备并保存一个查询计划。保存的计划将在当前会话的生存期内保留。

查询可以使用参数，也就是在计划实际执行时才提供值的占位符。在查询字符串中，用符号 `$1` ... `$n` 引用参数。If the query uses parameters, the names of the parameter types must be given as a Tcl list. (Write an empty list for *`typelist`* if no parameters are used.)

`spi_prepare` 的返回值是一个查询 ID，供后续调用 `spi_execp` 时使用。示例见 `spi_execp`。

`spi_execp ?-count` *`n`*`? ?-array` *`name`*`? ?-nulls` *`string`*`?` *`queryid ?value-list ? ?loop-body`*`?`

执行一个之前由 `spi_prepare` 准备的查询。*`queryid`* 是 `spi_prepare` 返回的 ID。如果查询引用了参数，必须提供一个 *`value-list`*。这是一个由参数实际值组成的 Tcl 列表。该列表的长度必须与之前提供给 `spi_prepare` 的参数类型列表相同。如果查询没有参数，则省略 *`value-list`*。

可选的 `-nulls` 值是一个由空格和 `'n'` 字符组成的字符串，告诉 `spi_execp` 哪些参数是空值。如果给出，它的长度必须与 *`value-list`* 完全相同。If it is not given, all the parameter values are nonnull.

除了指定查询及其参数的方式之外，`spi_execp` 的工作方式与 `spi_exec` 完全相同。`-count`、`-array` 和 *`loop-body`* 选项都相同，返回值也相同。

下面是一个使用已准备好的计划的 PL/Tcl 函数例子：

```
CREATE FUNCTION t1_count(integer, integer) RETURNS integer AS $$
    if {![ info exists GD(plan) ]} {
        # prepare the saved plan on the first call
        set GD(plan) [ spi_prepare \
                "SELECT count(*) AS cnt FROM t1 WHERE num >= \$1
 AND num <= \$2" \
                [ list int4 int4 ] ]
    }
    spi_execp -count 1 $GD(plan) [ list $1 $2 ]
    return $cnt
$$ LANGUAGE pltcl;
```

We need backslashes inside the query string given to `spi_prepare` to ensure that the `$n` markers will be passed through to `spi_prepare` as-is, and not replaced by Tcl variable substitution.

`spi_lastoid`

返回最近一次 `spi_exec` 或 `spi_execp` 插入的行的 OID，前提是该命令是单行 `INSERT` 且被修改的表包含 OID。（否则返回零。）

`quote` *`string`*

把给定字符串中所有单引号和反斜线字符加倍。这可以用来安全地引用将被插入到交给 `spi_exec` 或 `spi_prepare` 的 SQL 命令中的字符串。例如，考虑下面这样的 SQL 命令字符串：

```
"SELECT '$val' AS ret"
```

where the Tcl variable `val` actually contains `doesn't`. This would result in the final command string:

```
SELECT 'doesn't' AS ret
```

which would cause a parse error during `spi_exec` or `spi_prepare`. To work properly, the submitted command should contain:

```
SELECT 'doesn''t' AS ret
```

which can be formed in PL/Tcl using:

```
"SELECT '[ quote $val ]' AS ret"
```

One advantage of `spi_execp` is that you don't have to quote parameter values like this, since the parameters are never parsed as part of an SQL command string.

`elog` *`level msg`*

发出一条日志或错误消息。可用的级别有 `DEBUG`、`LOG`、`INFO`、`NOTICE`、`WARNING`、`ERROR` 和 `FATAL`。`ERROR` 会引发一个错误条件；如果周围的 Tcl 代码没有捕获它，该错误会传播到调用查询，导致当前事务或子事务被中止。This is effectively the same as the Tcl `error` command. `FATAL` aborts the transaction and causes the current session to shut down. (There is probably no good reason to use this error level in PL/Tcl functions, but it's provided for completeness.) The other levels only generate messages of different priority levels. Whether messages of a particular priority are reported to the client, written to the server log, or both is controlled by the log_min_messages and client_min_messages configuration variables. See 第 18 章 for more information.

40.6. PL/Tcl 中的触发器函数

可以用 PL/Tcl 编写触发器函数。PostgreSQL 要求作为触发器调用的函数必须声明为无参数、返回类型为 `trigger` 的函数。

触发器管理器的信息通过以下变量传入函数体：

`$TG_name`

`CREATE TRIGGER`语句中触发器的名称。

`$TG_relid`

导致触发器函数被调用的表的对象 ID。

`$TG_table_name`

导致触发器函数被调用的表的名称。

`$TG_table_schema`

导致触发器函数被调用的表所属的模式。

`$TG_relatts`

表列名构成的 Tcl 列表，其前面带有一个空列表元素。因此，使用Tcl的`lsearch`命令在该列表中查找列名时，返回的元素编号会从 1 开始表示第一列，这与PostgreSQL中列的惯常编号

方式一致。（空列表元素也会出现在已删除列的位置上，
这样其右侧列的属性编号仍然正确。）

`$TG_when`

其值为字符串`BEFORE`、`AFTER`或`INSTEAD OF`，具体取决于触发器事件的类型。

`$TG_level`

其值为字符串`ROW`或`STATEMENT`，取决于触发器事件的类型。

`$TG_op`

其值为字符串`INSERT`、`UPDATE`、`DELETE`或`TRUNCATE`，取决于触发器事件的类型。

`$NEW`

对于`INSERT`或`UPDATE`动作，这是一个包含新表行值的关联数组；对于`DELETE`则为空。
该数组以列名为索引。值为空值的列不会出现在数组中。对于语句级触发器，
不会设置这个变量。

`$OLD`

对于`UPDATE`或`DELETE`动作，这是一个包含旧表行值的关联数组；对于`INSERT`则为空。
该数组以列名为索引。值为空值的列不会出现在数组中。对于语句级触发器，
不会设置这个变量。

`$args`

一个 Tcl 列表，包含 `CREATE TRIGGER` 语句中给出的触发器函数参数。在函数体中，也可以通过 `$1` ... `$n` 访问这些参数。

触发器过程的返回值可以是字符串 `OK` 或 `SKIP`，也可以是由 Tcl 命令 `array get` 返回的列表。如果返回值是 `OK`，引发触发器的操作（`INSERT`/`UPDATE`/`DELETE`）将正常进行。`SKIP` 告诉触发器管理器静默地抑制这一行的操作。如果返回值是一个列表，它告诉 PL/Tcl 向触发器管理器返回一个修改后的行。这只对行级 `BEFORE INSERT` 或 `UPDATE` 触发器有意义，此时被插入的将是修改后的行，而不是 `$NEW` 中给出的行；或者对行级 `INSTEAD OF INSERT` 或 `UPDATE` 触发器有意义，此时返回的行用于支持 `INSERT RETURNING` 和 `UPDATE RETURNING` 命令。对于其他类型的触发器，返回值会被忽略。

下面是一个简单的触发器函数示例，强制用表中的一个整数值记录该行的更新次数。
新插入的行将此值初始化为 0，以后每次更新操作都将其递增。

```
CREATE FUNCTION trigfunc_modcount() RETURNS trigger AS $$
    switch $TG_op {
        INSERT {
            set NEW($1) 0
        }
        UPDATE {
            set NEW($1) $OLD($1)
            incr NEW($1)
        }
        default {
            return OK
        }
    }
```

```
    return [array get NEW]
$$ LANGUAGE pltcl;

CREATE TABLE mytab (num integer, description text, modcnt integer);

CREATE TRIGGER trig_mytab_modcount BEFORE INSERT OR UPDATE ON mytab
    FOR EACH ROW EXECUTE PROCEDURE trigfunc_modcount('modcnt');
```

注意，触发器函数本身不知道列名，列名通过触发器参数提供。这样，同一个触发器函数就可以在不同表上复用。

40.7. 模块与unknown命令

PL/Tcl 支持在使用时自动装载 Tcl 代码。它会识别一张特殊的表 pltcl_modules，该表被假定包含若干 Tcl 代码模块。如果这张表存在，模块unknown会从该表取出，并在数据库会话中第一次执行 PL/Tcl 函数之前立即装载到 Tcl 解释器中。（如果一个会话中使用了多个 Tcl 解释器，则对每个解释器分别执行这一过程；参见第 40.4 节。）

虽然unknown模块实际上可以包含你需要的任何初始化脚本，但它通常会定义一个 Tcl 的unknown过程，每当 Tcl 无法识别被调用的过程名时，就会调用该过程。PL/Tcl的这一过程的标准版本会尝试在pltcl_modules中找到一个能够定义所需过程的模块。如果找到了，就把它装载到解释器中，然后允许继续执行最初尝试的过程调用。辅助表pltcl_modfuncs提供了哪个模块定义了哪些函数的索引，使查找相当快速。

PostgreSQL发行版中包含用于维护这些表的支持脚本：pltcl_loadmod、pltcl_listmod、pltcl_delmod，以及标准unknown模块的源代码，位于share/unknown.pltcl。要支持自动装载机制，最初必须把这个模块装载到每个数据库中。

表pltcl_modules和pltcl_modfuncs 必须对所有用户可读，但最好只让数据库管理员拥有并可以写入它们。作为一种安全预防措施，除非pltcl_modules由超级用户拥有，否则 PL/Tcl 会忽略它（因而不尝试装载unknown模块）。不过，如果你足够信任其他用户，可以把这张表的 UPDATE 权限授予他们。

40.8. Tcl 过程名

在PostgreSQL中，只要参数个数或参数类型不同，就可以复用同一个函数名。不过，Tcl 要求所有过程名都必须不同。PL/Tcl 处理这一问题的方式是：在内部 Tcl 过程名中包含系统表 pg_proc 中的函数对象 ID 作为其名称的一部分。因此，名称相同但参数类型不同的PostgreSQL函数，也会对应不同的 Tcl 过程。这通常不是 PL/Tcl 程序员需要关心的事情，但在调试时可能会看见。

第 41 章 PL/Perl - Perl 过程语言

PL/Perl 是一种可加载的过程语言，可以用 Perl 编程语言[1]编写 PostgreSQL 函数。

使用 PL/Perl 的主要好处，是可以在存储函数中使用 Perl 提供的多种“字符串处理”操作符和函数。与使用 PL/pgSQL 提供的字符串函数和控制结构相比，用 Perl 解析复杂字符串可能更容易。

要在特定数据库中安装 PL/Perl，可使用 `CREATE EXTENSION plperl`，或者从 shell 命令行使用 `createlang plperl dbname`。

> **提示**
>
> 如果把某种语言安装到`template1`中，之后创建的所有数据库都会自动安装该语言。

> **注意**
>
> 使用源码包的用户必须在安装过程中专门启用 PL/Perl 的构建（更多信息见第 15 章）。使用二进制包的用户可能会在单独的子包中找到 PL/Perl。

41.1. PL/Perl 函数和参数

要用 PL/Perl 语言创建一个函数，可使用标准的CREATE FUNCTION语法：

```
CREATE FUNCTION funcname (argument-types) RETURNS return-type AS $$
    # PL/Perl 函数体
$$ LANGUAGE plperl;
```

函数的主体就是普通的 Perl 代码。事实上，PL/Perl 的粘合代码会把它包裹在一个 Perl 子程序中。PL/Perl 函数在标量上下文中被调用，因此不能返回列表。如下文所述，可以通过返回引用来返回非标量值（数组、记录和集合）。

PL/Perl 也支持用DO语句调用的匿名代码块：

```
DO $$
    # PL/Perl 代码
$$ LANGUAGE plperl;
```

匿名代码块不接收参数，它可能返回的任何值都会被丢弃。除此之外，
它的行为与函数完全相同。

> **注意**
>
> 在 Perl 中使用命名嵌套子程序是有危险的，特别是当它们引用外层作用域中的词法变量时。因为 PL/Perl 函数被包装成一个子程序，任何放在其中的命名子程序都会成为嵌套子程序。一般来说，创建通过代码引用（coderef）调用的匿名子程序会安全得多。

[1] http://www.perl.org

> 更多信息可见 perldiag手册页中的`Variable "%s" will not stay shared`以及`Variable "%s" is not available`，或者在互联网上搜索“perl nested named subroutine”。

`CREATE FUNCTION` 命令的语法要求函数体写成一个字符串常量。通常对该字符串常量使用美元引用（见第 4.1.2.4 节）最方便。如果选择使用转义字符串语法 `E''`，则必须将函数体中用到的单引号（`'`）和反斜线（`\`）都写成双份（见第 4.1.2.1 节）。

参数和结果的处理与其他 Perl 子程序一样：参数通过 `@_` 传入，结果值通过 `return` 返回，或者作为函数中最后一个被求值的表达式返回。

例如，一个返回两个整数值中较大值的函数可以定义为：

```
CREATE FUNCTION perl_max (integer, integer) RETURNS integer AS $$
    if ($_[0] > $_[1]) { return $_[0]; }
    return $_[1];
$$ LANGUAGE plperl;
```

> **注意**
>
> 参数会从数据库编码转换为 PL/Perl 中使用的 UTF-8，返回时再从 UTF-8 转回数据库编码。

如果一个 SQL 空值被传给一个函数，在 Perl 中该参数值将呈现为“未定义”。上述函数定义对于空值输入的行为并不理想（实际上，它会把它们当作零）。我们可以在函数定义中添加 `STRICT`，让 PostgreSQL 采取更合理的做法：如果传入空值，函数将根本不会被调用，而是自动返回空值结果。另一种方式是在函数体中检查未定义输入。例如，假设我们希望在 `perl_max` 的两个参数中一个为空值、另一个非空值时，返回非空值参数而不是空值：

```
CREATE FUNCTION perl_max (integer, integer) RETURNS integer AS $$
    my ($x, $y) = @_;
    if (not defined $x) {
        return undef if not defined $y;
        return $y;
    }
    return $x if not defined $y;
    return $x if $x > $y;
    return $y;
$$ LANGUAGE plperl;
```

如上所示，要从 PL/Perl 函数返回 SQL 空值，就返回一个未定义值。无论函数是否为严格函数，都可以这样做。

函数参数中任何不是引用的值都是字符串，其格式是相应数据类型的标准 PostgreSQL 外部文本表示。对于普通的数值或文本类型，Perl 通常会正确处理，程序员一般不必操心。不过在其他情况下，参数可能需要转换成 Perl 中更易用的形式。例如，可以用 `decode_bytea` 函数把 `bytea` 类型的参数转换成未转义的二进制数据。

同样，返回给 PostgreSQL 的值也必须采用外部文本表示格式。例如，可以用 `encode_bytea` 函数把二进制数据转义成适合作为 `bytea` 类型返回值的形式。

Perl 可以把 PostgreSQL 数组返回为对 Perl 数组的引用。这里有一个示例：

```
CREATE OR REPLACE function returns_array()
RETURNS text[][] AS $$
    return [['a"b','c,d'],['e\\f','g']];
$$ LANGUAGE plperl;

select returns_array();
```

Perl 把 PostgreSQL 数组作为一个 blessed 的 PostgreSQL::InServer::ARRAY 对象传递。这个对象可以被当作数组引用或字符串对待，允许为 PostgreSQL 9.1 之前版本编写的 Perl 代码继续运行。例如：

```
CREATE OR REPLACE FUNCTION concat_array_elements(text[]) RETURNS TEXT
 AS $$
    my $arg = shift;
    my $result = "";
    return undef if (!defined $arg);

    # as an array reference
    for (@$arg) {
        $result .= $_;
    }

    # also works as a string
    $result .= $arg;

    return $result;
$$ LANGUAGE plperl;

SELECT concat_array_elements(ARRAY['PL','/','Perl']);
```

> **注意**
>
> Multi-dimensional arrays are represented as references to lower-dimensional arrays of references in a way common to every Perl programmer.

复合类型参数被作为哈希的引用传递给函数。哈希的键是复合类型的属性名。这里是一个示例：

```
CREATE TABLE employee (
    name text,
    basesalary integer,
    bonus integer
);

CREATE FUNCTION empcomp(employee) RETURNS integer AS $$
    my ($emp) = @_;
    return $emp->{basesalary} + $emp->{bonus};
$$ LANGUAGE plperl;

SELECT name, empcomp(employee.*) FROM employee;
```

PL/Perl 函数也可以用相同的方法返回复合类型：返回一个具有所需属性的哈希的引用。例如：

```
CREATE TYPE testrowperl AS (f1 integer, f2 text, f3 text);

CREATE OR REPLACE FUNCTION perl_row() RETURNS testrowperl AS $$
    return {f2 => 'hello', f1 => 1, f3 => 'world'};
$$ LANGUAGE plperl;

SELECT * FROM perl_row();
```

在声明的结果数据类型中，凡是哈希中未出现的列都将作为空值返回。

PL/Perl 函数还可以返回标量类型或复合类型的集合。通常会希望逐行返回，这既能缩短启动时间，也能避免把整个结果集排入内存。如下所示，可以用 `return_next` 实现这一点。注意，在最后一次 `return_next` 之后，必须写上 `return` 或者（更好）`return undef`。

```
CREATE OR REPLACE FUNCTION perl_set_int(int)
RETURNS SETOF INTEGER AS $$
    foreach (0..$_[0]) {
        return_next($_);
    }
    return undef;
$$ LANGUAGE plperl;

SELECT * FROM perl_set_int(5);

CREATE OR REPLACE FUNCTION perl_set()
RETURNS SETOF testrowperl AS $$
    return_next({ f1 => 1, f2 => 'Hello', f3 => 'World' });
    return_next({ f1 => 2, f2 => 'Hello', f3 => 'PostgreSQL' });
    return_next({ f1 => 3, f2 => 'Hello', f3 => 'PL/Perl' });
    return undef;
$$ LANGUAGE plperl;
```

对于较小的结果集，可以返回一个数组引用，其中对简单类型、
数组类型和复合类型分别包含标量、数组引用或哈希引用。
下面是把整个结果集作为数组引用返回的几个简单示例：

```
CREATE OR REPLACE FUNCTION perl_set_int(int) RETURNS SETOF INTEGER AS
 $$
    return [0..$_[0]];
$$ LANGUAGE plperl;

SELECT * FROM perl_set_int(5);

CREATE OR REPLACE FUNCTION perl_set() RETURNS SETOF testrowperl AS $$
    return [
        { f1 => 1, f2 => 'Hello', f3 => 'World' },
        { f1 => 2, f2 => 'Hello', f3 => 'PostgreSQL' },
        { f1 => 3, f2 => 'Hello', f3 => 'PL/Perl' }
    ];
$$ LANGUAGE plperl;

SELECT * FROM perl_set();
```

如果希望在代码中使用 `strict` 编译指示，可以有几种选择。对于临时的全局用法，可以 `SET plperl.use_strict` 为真。这会影响后续编译的 PL/Perl 函数，但不会影响当前会话中已经编

译的函数。对于永久的全局用法，可以在 postgresql.conf 文件中将 plperl.use_strict 设为真。

如果要在特定函数中永久启用它，可以简单地把

```
use strict;
```

放在函数体顶部。

如果 Perl 版本为 5.10.0 或更高，也可以通过 use 使用 feature 编译指示。

41.2. PL/Perl 中的数据值

提供给 PL/Perl 函数代码的参数值，就是转换成文本形式的输入参数（就像它们由 SELECT 语句显示出来一样）。相反，return 和 return_next 命令接受任何可作为该函数声明返回类型输入格式的字符串。

41.3. 内置函数

41.3.1. 从 PL/Perl 访问数据库

可以通过下列函数在 Perl 函数中访问数据库本身：

spi_exec_query(*query* [, *max-rows*])

spi_exec_query 执行 SQL 命令，并返回整个行集；返回值是对一个数组的引用，数组元素为哈希引用。只有在确定结果集会比较小时，才应使用此命令。下面是一个查询（SELECT 命令）的示例，指定了可选的最大行数：

```
$rv = spi_exec_query('SELECT * FROM my_table', 5);
```

这会返回最多 5 行，来自表 my_table。如果 my_table 包含列 my_column，就可以取得结果中第 $i 行的该列值，方法如下：

```
$foo = $rv->{rows}[$i]->{my_column};
```

要取得 SELECT 查询返回的总行数，可以这样做：

```
$nrows = $rv->{processed}
```

下面是使用另一种命令类型的示例：

```
$query = "INSERT INTO my_table VALUES (1, 'test')";
$rv = spi_exec_query($query);
```

可以这样访问命令状态（例如 SPI_OK_INSERT）：

```
$res = $rv->{status};
```

要获取受影响的行数，可使用：

```
$nrows = $rv->{processed};
```

这里是一个完整的示例：

```
CREATE TABLE test (
    i int,
    v varchar
);

INSERT INTO test (i, v) VALUES (1, 'first line');
INSERT INTO test (i, v) VALUES (2, 'second line');
INSERT INTO test (i, v) VALUES (3, 'third line');
INSERT INTO test (i, v) VALUES (4, 'immortal');

CREATE OR REPLACE FUNCTION test_munge() RETURNS SETOF test AS $$
    my $rv = spi_exec_query('select i, v from test;');
    my $status = $rv->{status};
    my $nrows = $rv->{processed};
    foreach my $rn (0 .. $nrows - 1) {
        my $row = $rv->{rows}[$rn];
        $row->{i} += 200 if defined($row->{i});
        $row->{v} =~ tr/A-Za-z/a-zA-Z/ if (defined($row->{v}));
        return_next($row);
    }
    return undef;
$$ LANGUAGE plperl;

SELECT * FROM test_munge();
```

spi_query(*command*)
spi_fetchrow(*cursor*)
spi_cursor_close(*cursor*)

spi_query 和 spi_fetchrow 需要配合使用，适用于结果集可能很大，或者希望在行到达时立即返回的情况。spi_fetchrow 只能与 spi_query 一起使用。下面的示例演示了它们的配合方式：

```
CREATE TYPE foo_type AS (the_num INTEGER, the_text TEXT);

CREATE OR REPLACE FUNCTION lotsa_md5 (INTEGER) RETURNS SETOF foo_
type AS $$
    use Digest::MD5 qw(md5_hex);
    my $file = '/usr/share/dict/words';
    my $t = localtime;
    elog(NOTICE, "opening file $file at $t" );
    open my $fh, '<', $file # 注意，这里访问了文件!
        or elog(ERROR, "cannot open $file for reading: $!");
    my @words = <$fh>;
    close $fh;
    $t = localtime;
    elog(NOTICE, "closed file $file at $t");
    chomp(@words);
    my $row;
    my $sth = spi_query("SELECT * FROM generate_series(1,$_[0]) AS
 b(a)");
    while (defined ($row = spi_fetchrow($sth))) {
        return_next({
            the_num => $row->{a},
```

```
                the_text => md5_hex($words[rand @words])
            });
        }
        return;
$$ LANGUAGE plperlu;

SELECT * from lotsa_md5(500);
```

通常，应重复调用 spi_fetchrow，直到它返回 undef，这表示已经没有更多行可读。当 spi_fetchrow 返回 undef 时，spi_query 返回的游标会被自动释放。如果不打算读取所有行，则应调用 spi_cursor_close 释放游标。否则会导致内存泄漏。

spi_prepare(*command*, *argument types*)
spi_query_prepared(*plan*, *arguments*)
spi_exec_prepared(*plan* [, *attributes*], *arguments*)
spi_freeplan(*plan*)

spi_prepare, spi_query_prepared, spi_exec_prepared, 以及 spi_freeplan 实现相同的功能，但用于预备查询。spi_prepare 接受包含编号参数占位符（$1、$2 等）的查询字符串，以及由参数类型字符串组成的列表：

```
$plan = spi_prepare('SELECT * FROM test WHERE id > $1 AND name =
 $2',
                                                        'INTEGER',
 'TEXT');
```

调用 spi_prepare 预备查询计划后，就可以使用该计划代替查询字符串。可以将其用于 spi_exec_prepared，其结果与以下函数返回的结果相同：spi_exec_query；也可以用于 spi_query_prepared，它返回游标，行为与 spi_query 完全相同。随后可以将该游标传给 spi_fetchrow。spi_exec_prepared 可选的第二个参数是属性哈希引用；目前唯一支持的属性是 limit，用于设置查询返回的最大行数。

预备查询的优点在于，一个准备好的计划可以用于多次查询执行。当计划不再需要时，可以用 spi_freeplan 将其释放：

```
CREATE OR REPLACE FUNCTION init() RETURNS VOID AS $$
        $_SHARED{my_plan} = spi_prepare('SELECT (now() + $1)::date
 AS now',
                                        'INTERVAL');
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION add_time( INTERVAL ) RETURNS TEXT AS $$
        return spi_exec_prepared(
                $_SHARED{my_plan},
                $_[0]
        )->{rows}->[0]->{now};
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION done() RETURNS VOID AS $$
        spi_freeplan( $_SHARED{my_plan});
        undef $_SHARED{my_plan};
$$ LANGUAGE plperl;

SELECT init();
```

```
SELECT add_time('1 day'), add_time('2 days'), add_time('3 days');
SELECT done();

  add_time  |  add_time  |  add_time
------------+------------+------------
 2005-12-10 | 2005-12-11 | 2005-12-12
```

请注意，spi_prepare 中的参数下标由 $1、$2、$3 等表示，因此应避免用双引号声明查询字符串，以免轻易引入难以察觉的错误。

下面的另一个示例展示了如何在 spi_exec_prepared 中使用可选参数：

```
CREATE TABLE hosts AS SELECT id, ('192.168.1.'||id)::inet AS
 address
                      FROM generate_series(1,3) AS id;

CREATE OR REPLACE FUNCTION init_hosts_query() RETURNS VOID AS $$
        $_SHARED{plan} = spi_prepare('SELECT * FROM hosts
                                      WHERE address << $1', 'inet')
;
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION query_hosts(inet) RETURNS SETOF hosts AS
 $$
        return spi_exec_prepared(
                $_SHARED{plan},
                {limit => 2},
                $_[0]
        )->{rows};
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION release_hosts_query() RETURNS VOID AS $$
        spi_freeplan($_SHARED{plan});
        undef $_SHARED{plan};
$$ LANGUAGE plperl;

SELECT init_hosts_query();
SELECT query_hosts('192.168.1.0/30');
SELECT release_hosts_query();

    query_hosts
-----------------
 (1,192.168.1.1)
 (2,192.168.1.2)
(2 rows)
```

41.3.2. PL/Perl 中的辅助函数

elog(*level*, *msg*)

发出日志消息或错误信息。可用级别包括 DEBUG、LOG、INFO、NOTICE、WARNING以及ERROR。ERROR 会引发错误条件；如果周围的 Perl 代码没有捕获它，错误就会传播到调用查询，导致当前事务或子事务被中止。这实质上等同于 Perl 的 die 命令。其他级别只会生成不同优先级的消息。某一优先级的消息是报告给客户端、写入服务

器日志，还是两者兼有，由配置变量log_min_messages和client_min_messages控制。详见第 18 章。

`quote_literal(string)`

返回给定字符串适当加引号后的形式，以便把它用作 SQL 语句字符串中的字符串字面量。嵌入的单引号和反斜线会被正确地双写。注意，对于 undef 输入，`quote_literal` 会返回 undef；如果参数可能为 undef，`quote_nullable` 往往更合适。

`quote_nullable(string)`

返回给定字符串适当加引号后的形式，以便把它用作 SQL 语句字符串中的字符串字面量；如果参数为 undef，则返回未加引号的字符串“NULL”。嵌入的单引号和反斜线会被正确地双写。

`quote_ident(string)`

返回给定字符串适当加引号后的形式，以便把它用作 SQL 语句字符串中的标识符。只有在必要时才会添加引号，也就是当字符串包含非标识符字符，或者会发生大小写折叠时。嵌入的引号会被正确地双写。

`decode_bytea(string)`

返回由给定字符串的内容表示的未转义二进制数据，该字符串应为 `bytea` 编码形式。

`encode_bytea(string)`

返回给定字符串中二进制数据内容的 `bytea` 编码形式。

`encode_array_literal(array)`
`encode_array_literal(array, delimiter)`

将引用数组的内容以数组字面量格式（见第 8.14.2 节）返回为字符串。若参数不是数组引用，则原样返回该参数值。如果未指定分隔符，或者分隔符为 undef，则数组字面量元素之间默认使用“, ”作为分隔符。

`encode_typed_literal(value, typename)`

将一个 Perl 变量转换为第二个参数指定的数据类型的值，并返回该值的字符串表示。它能正确处理嵌套数组和复合类型的值。

`encode_array_constructor(array)`

将引用数组的内容以数组构造器格式（第 4.2.12 节）返回为字符串。其中每个值都使用 `quote_nullable` 加引号。如果参数不是数组引用，则返回使用 `quote_nullable` 加引号后的参数值。

`looks_like_number(string)`

如果按照 Perl 的规则看，给定字符串的内容像数字，则返回真值，否则返回假值。如果参数为 undef，则返回 undef。前导和尾随空格会被忽略。`Inf` 和 `Infinity` 被视为数字。

`is_array_ref(argument)`

如果给定参数可被视为数组引用，则返回真值；也就是该参数的 ref 值为 `ARRAY` 或 `PostgreSQL::InServer::ARRAY`。否则返回假值。

41.4. PL/Perl 中的全局值

可以使用全局哈希 `%_SHARED` 在当前会话的整个生命周期内跨函数调用存储数据，包括代码引用。

这是共享数据的一个简单示例：

```
CREATE OR REPLACE FUNCTION set_var(name text, val text) RETURNS text
 AS $$
    if ($_SHARED{$_[0]} = $_[1]) {
        return 'ok';
    } else {
        return "cannot set shared variable $_[0] to $_[1]";
    }
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION get_var(name text) RETURNS text AS $$
    return $_SHARED{$_[0]};
$$ LANGUAGE plperl;

SELECT set_var('sample', 'Hello, PL/Perl!  How''s tricks?');
SELECT get_var('sample');
```

下面是一个稍复杂一些的、使用代码引用的示例：

```
CREATE OR REPLACE FUNCTION myfuncs() RETURNS void AS $$
    $_SHARED{myquote} = sub {
        my $arg = shift;
        $arg =~ s/(['\\])/\\$1/g;
        return "'$arg'";
    };
$$ LANGUAGE plperl;

SELECT myfuncs(); /* 初始化函数 */

/* 创建一个使用加引号函数的函数 */

CREATE OR REPLACE FUNCTION use_quote(TEXT) RETURNS text AS $$
    my $text_to_quote = shift;
    my $qfunc = $_SHARED{myquote};
    return &$qfunc($text_to_quote);
$$ LANGUAGE plperl;
```

（上面的代码也可以替换为单行 `return $_SHARED{myquote}->($_[0]);`，但代价是可读性会变差。）

出于安全原因，PL/Perl 会在每个 SQL 角色各自独立的 Perl 解释器中执行该角色调用的函数。这可以防止一个用户意外或恶意地干扰另一个用户的 PL/Perl 函数行为。每个这样的解释器都有自己的 `%_SHARED` 变量值和其他全局状态。因此，当且仅当两个 PL/Perl 函数都由同一 SQL 角色执行时，它们才会共享 `%_SHARED` 的值。在某些应用中，一个会话可能会在多个 SQL 角色下执行代码（通过 `SECURITY DEFINER` 函数、使用 `SET ROLE` 等），这时可能需要显式采取措施，确保 PL/Perl 函数能够通过 `%_SHARED` 共享数据。为此，要确保需要相互通信的函数由同一用户拥有，并将它们标记为 `SECURITY DEFINER`。当然，必须注意不要让这些函数被用于任何非预期用途。

41.5. 受信任与不受信任的 PL/Perl

通常，PL/Perl 会被安装为一种名为 `plperl` 的“受信任的”编程语言。在这种设置下，为了保持安全性，某些 Perl 操作会被禁用。一般来说，受限制的是那些与环境交互的操作，包括文件句柄操作、`require` 和 `use`（针对外部模块）。它无法像 C 函数那样访问数据库服务器进程的内部，也无法以服务器进程的权限获取操作系统级访问。因此，
可以允许任何无特权的数据库用户使用这种语言。

下面示例中的函数将无法工作，因为出于安全原因不允许它做文件操作：

```
CREATE FUNCTION badfunc() RETURNS integer AS $$
    my $tmpfile = "/tmp/badfile";
    open my $fh, '>', $tmpfile
        or elog(ERROR, qq{could not open the file "$tmpfile": $!});
    print $fh "Testing writing to a file\n";
    close $fh or elog(ERROR, qq{could not close the file "$tmpfile":
 $!});
    return 1;
$$ LANGUAGE plperl;
```

这个函数的创建会失败，因为验证器会捕捉到它使用了禁用的操作。

有时需要编写不受这些限制的 Perl 函数。例如，可能需要一个能发送邮件的 Perl 函数。为处理这类情况，也可以把 PL/Perl 安装成一种“不受信任的”语言（通常称为 PL/PerlU）。在这种情况下，完整的 Perl 语言都可用。安装该语言时，使用语言名 `plperlu` 就会选择不受信任的 PL/Perl 变体。

PL/PerlU 函数的编写者必须注意，函数不能被用于任何非预期用途，因为它能够执行以数据库管理员身份登录的用户所能做的任何事情。请注意，
数据库系统只允许数据库超级用户用不受信任的语言创建函数。

如果上述函数是由超级用户使用 `plperlu` 语言创建的，执行就会成功。

同样地，如果把语言指定为 `plperlu` 而不是 `plperl`，用 Perl 编写的匿名代码块也可以使用原本受限的操作，但调用者必须是超级用户。

> **注意**
>
> 虽然 PL/Perl 函数会为每个 SQL 角色在独立的 Perl 解释器中运行，但给定会话中执行的所有 PL/PerlU 函数都运行在同一个 Perl 解释器中（而且这个解释器并不是任何一个用于 PL/Perl 函数的解释器）。这允许 PL/PerlU 函数自由共享数据，但 PL/Perl 与 PL/PerlU 函数之间不能进行通信。

> **注意**
>
> Perl 无法在单个进程内支持多个解释器，除非构建时使用了适当的标志，即 `usemultiplicity` 或 `useithreads` 之一。（除非确实需要使用线程，否则更推荐 `usemultiplicity`。更多细节见 perlembed 手册页。）如果 PL/Perl 使用的是未按这种方式构建的 Perl 副本，那么每个会话中就只能有一个 Perl 解释器，因此任一会话只能执行 PL/PerlU 函数，或者执行全部由同一 SQL 角色调用的 PL/Perl 函数。

41.6. PL/Perl 触发器

PL/Perl 可用于编写触发器函数。在触发器函数中，哈希引用 `$_TD` 包含有关当前触发器事件的信息。`$_TD` 是一个全局变量，对触发器的每一次调用都会得到一个单独的局部值。`$_TD` 哈希引用包含以下字段：

`$_TD->{new}{foo}`

列 `foo` 的 `NEW` 值

`$_TD->{old}{foo}`

列 `foo` 的 `OLD` 值

`$_TD->{name}`

被调用的触发器名称

`$_TD->{event}`

触发器事件：`INSERT`、`UPDATE`、`DELETE`、`TRUNCATE`或者`UNKNOWN`

`$_TD->{when}`

触发器被调用的时机：`BEFORE`、`AFTER`、`INSTEAD OF` 或 `UNKNOWN`

`$_TD->{level}`

触发器级别：`ROW`、`STATEMENT`或者`UNKNOWN`

`$_TD->{relid}`

触发该触发器的表的 OID

`$_TD->{table_name}`

触发该触发器的表名

`$_TD->{relname}`

触发该触发器的表名。该字段已弃用，并且可能会在未来版本中移除。请改用 $_TD->{table_name}。

`$_TD->{table_schema}`

触发该触发器的表所在模式的名称

`$_TD->{argc}`

触发器函数的参数数目

`@{$_TD->{args}}`

触发器函数的参数。如果 `$_TD->{argc}` 为 0，则该字段不存在

行级触发器可以返回下列之一：

`return;`

执行操作

`"SKIP"`

不执行操作

`"MODIFY"`

指示触发器函数修改了 `NEW` 行

下面的触发器函数示例演示了上述部分内容：

```
CREATE TABLE test (
    i int,
    v varchar
);

CREATE OR REPLACE FUNCTION valid_id() RETURNS trigger AS $$
    if (($_TD->{new}{i} >= 100) || ($_TD->{new}{i} <= 0)) {
        return "SKIP";    # 跳过 INSERT/UPDATE 命令
    } elsif ($_TD->{new}{v} ne "immortal") {
        $_TD->{new}{v} .= "(modified by trigger)";
        return "MODIFY";  # 修改行并执行 INSERT/UPDATE 命令
    } else {
        return;           # 执行 INSERT/UPDATE 命令
    }
$$ LANGUAGE plperl;

CREATE TRIGGER test_valid_id_trig
    BEFORE INSERT OR UPDATE ON test
    FOR EACH ROW EXECUTE PROCEDURE valid_id();
```

41.7. PL/Perl 内部机制

41.7.1. 配置

本节列出影响 PL/Perl 的配置参数。要在 PL/Perl 装载之前设置这些参数，必须先把 “`plperl`” 加入 `postgresql.conf` 中的custom_variable_classes列表。

`plperl.on_init` (`string`)

指定在 Perl 解释器首次初始化时、在它被专用于 `plperl` 或 `plperlu` 之前，要执行的 Perl 代码。执行这段代码时 SPI 函数不可用。如果代码因错误而失败，就会中止解释器初始化，并把错误传播到调用查询，导致当前事务或子事务被中止。

Perl 代码仅限一个字符串。更长的代码可以放在模块中，并由 `on_init` 字符串加载。示例：

```
plperl.on_init = 'require "plperlinit.pl"'
plperl.on_init = 'use lib "/my/app"; use MyApp::PgInit;'
```

任何通过 `plperl.on_init` 直接或间接加载的模块，都可供 `plperl` 使用。这可能带来安全风险。要查看已经加载了哪些模块，可以使用：

```
DO 'elog(WARNING, join ", ", sort keys %INC)' language plperl;
```

如果 plperl 库被包含在shared_preload_libraries中，初始化就会在 postmaster 中发生；在这种情况下，需要额外考虑使 postmaster 不稳定的风险。使用此特性的主要原因是，通过 `plperl.on_init` 加载的 Perl 模块只需在 postmaster 启动时加载一次，之后在各个数据库会话中都可立即使用，而无需承担加载开销。不过要记住，
这种开销只会对数据库会话使用的第一个 Perl 解释器被省掉，也就是 PL/PerlU，或者第一个调用 PL/Perl 函数的 SQL 角色所对应的 PL/Perl。数据库会话中后来创建的任何额外 Perl 解释器，都必须重新执行 `plperl.on_init`。此外，在 Windows 上预加载完全不会带来节省，因为在 postmaster 进程中创建的 Perl 解释器不会传播到子进程中。

这个参数只能在 `postgresql.conf` 文件中，或在服务器命令行上设置。

`plperl.on_plperl_init` (`string`)
`plperl.on_plperlu_init` (`string`)

这些参数分别指定在 Perl 解释器被专用于 `plperl` 或 `plperlu` 时要执行的 Perl 代码。这会在数据库会话中首次执行 PL/Perl 或 PL/PerlU 函数时发生；如果因为调用另一种语言，或某个新的 SQL 角色调用了 PL/Perl 函数而需要创建额外解释器时，也会发生。这是在 `plperl.on_init` 完成的任何初始化之后进行的。执行这段代码时 SPI 函数不可用。`plperl.on_plperl_init` 中的 Perl 代码是在对解释器进行“锁定”之后执行的，因此只能执行受信任的操作。

如果代码因错误而失败，就会中止初始化，并把错误传播到调用查询，
导致当前事务或子事务被中止。在 Perl 中已经完成的任何动作都不会被撤销；不过，该解释器将不再被使用。如果再次使用该语言，就会在一个新的 Perl 解释器中再次尝试初始化。

只有超级用户可以更改这些设置。尽管这些设置可以在会话中修改，
但这类更改不会影响已经用于执行函数的 Perl 解释器。

`plperl.use_strict` (`boolean`)

如果将其设为真，后续编译的 PL/Perl 函数都会启用 `strict` 编译指示。该参数不影响当前会话中已经编译的函数。

41.7.2. 限制与缺失特性

PL/Perl 目前仍缺少下列特性，但欢迎为此作出贡献。

- PL/Perl 函数不能直接调用彼此。
- SPI 尚未完全实现。
- 如果使用 `spi_exec_query` 取回非常大的数据集，应注意这些数据都会进入内存。可以像前面所示那样，通过使用 `spi_query`/`spi_fetchrow` 来避免这种情况。

  如果集合返回函数通过 `return` 把一个大型行集返回给 PostgreSQL，也会出现类似问题。前面已经说明过，可以改为对每一行使用 `return_next`，从而避免这个问题。
- 当会话正常结束（而不是由于致命错误结束）时，任何已经定义的 `END` 块都会被执行。目前不会执行其他操作。具体来说，文件句柄不会自动刷写，对象也不会自动销毁。

第 42 章 PL/Python - Python 过程语言

PL/Python 过程语言允许用 Python 语言[1]编写 PostgreSQL 函数。

要在特定数据库中安装 PL/Python，可使用`CREATE EXTENSION plpythonu`，或者从 shell 命令行使用`createlang plpythonu` *dbname*（另见第 42.1 节）。

> **提示**
>
> 如果把某种语言安装到`template1`中，之后创建的所有数据库都会自动安装该语言。

从 PostgreSQL 7.4 开始，PL/Python 只以一种“非可信”语言的形式提供，这意味着它不提供任何限制用户在其中做什么的手段。因此它已被重命名为 `plpythonu`。如果 Python 中开发出新的安全执行机制，可信变体 `plpython` 将来可能会重新可用。非可信 PL/Python 函数的编写者必须注意，该函数不能被用来做任何不该做的事，因为它能做任何以数据库管理员身份登录的用户所能做的事情。只有超级用户才能用 `plpythonu`.

> **注意**
>
> 源码包用户必须在安装过程中专门启用 PL/Python 的构建（更多信息请参阅安装说明）。二进制包用户则可能会在单独的子包中找到 PL/Python。

42.1. Python 2 与 Python 3

PL/Python 同时支持 Python 2 和 Python 3 两种语言变体。（PostgreSQL 安装说明中可能包含关于所支持的确切 Python 次版本的更精确信息。）由于 Python 2 和 Python 3 语言变体在一些重要方面不兼容，PL/Python 使用下面的命名和过渡方案来避免混用它们：

- The PostgreSQL language named `plpython2u` implements PL/Python based on the Python 2 language variant.
- PostgreSQL 中名为 `plpython3u` 的语言实现基于 Python 3 语言变体的 PL/Python。
- 名为 `plpythonu` 的语言实现基于默认 Python 语言变体（目前是 Python 2）的 PL/Python。（这个默认与任何本地 Python 安装自己认为的“默认”无关，例如 `/usr/bin/python` 是什么。）取决于 Python 社区向 Python 3 迁移的进度，这个默认在 PostgreSQL 较远的将来版本中可能会被改为 Python 3。

It depends on the build configuration or the installed packages whether PL/Python for Python 2 or Python 3 or both are available.

> **提示**
>
> 所构建的变体取决于安装时找到的 Python 版本，或使用`PYTHON`环境变量显式设置的版本；请参见第 15.4 节。要在同一个安装中同时提供两种 PL/Python 变体，需要对源代码树进行两次配置和构建。

[1] http://www.python.org

由此可以采用以下使用和迁移策略：

- 现有用户和目前不打算使用 Python 3 的用户，可以使用语言名 `plpythonu`，在可预见的将来无需更改任何内容。建议通过迁移到 Python 2.6/2.7，逐步让代码“适应未来”，以简化最终迁移到 Python 3 的工作。

  实际上，许多 PL/Python 函数只需很少或不做改动即可迁移到 Python 3。

- 如果用户清楚自己的代码严重依赖 Python 2，且不打算更改，可以使用语言名 `plpython2u`。它在很久以后仍会继续工作，直到 PostgreSQL 可能完全停止支持 Python 2。

- 希望开始使用 Python 3 的用户可以使用语言名 `plpython3u`，按照目前的标准，它会一直可用。在遥远的未来，当 Python 3 可能成为默认版本时，用户也许会出于美观而去掉名称中的“3”。

- 希望构建仅包含 Python 3 的操作系统环境的大胆用户，可以修改 `pg_pltemplate` 的内容，使 `plpythonu` 等同于 `plpython3u`，但要记住，这会使该安装环境与绝大多数其他环境不兼容。

有关迁移到 Python 3 的更多信息，也可参见文档 Python 3.0 的新变化[2]。

不允许在同一个会话中同时使用基于 Python 2 和基于 Python 3 的 PL/Python，因为动态模块中的符号会冲突，可能导致 PostgreSQL 服务器进程崩溃。系统会检查同一会话是否混用了不同的 Python 大版本，如果发现不匹配，就会中止会话。不过，可以在同一数据库的不同会话中分别使用这两种 PL/Python。

42.2. PL/Python 函数

PL/Python 中的函数通过标准的CREATE FUNCTION语法声明：

```
CREATE FUNCTION funcname (argument-list)
  RETURNS return-type
AS $$
  # PL/Python 函数体
$$ LANGUAGE plpythonu;
```

函数体就是一段 Python 脚本。调用函数时，参数作为列表 `args` 的元素传入；命名参数还会作为普通变量传给 Python 脚本。使用命名参数通常更易读。Python 代码按通常方式用 `return` 或 `yield`（返回结果集时）返回结果。如果没有提供返回值，Python 默认返回 `None`。PL/Python 会将 Python 的 `None` 转换为 SQL 空值。

例如，一个返回两个整数中较大值的函数可以定义为：

```
CREATE FUNCTION pymax (a integer, b integer)
  RETURNS integer
AS $$
  if a > b:
    return a
  return b
$$ LANGUAGE plpythonu;
```

函数定义体中给出的 Python 代码会被转换成一个 Python 函数。例如，上面的定义会变成：

```
def __plpython_procedure_pymax_23456():
```

[2] http://docs.python.org/py3k/whatsnew/3.0.html

```
  if a > b:
    return a
  return b
```

这里假定 23456 是由PostgreSQL分配给该函数的 OID。

参数会被设置为全局变量。由于 Python 的作用域规则，这会带来一个微妙的结果：除非在代码块中再次将该变量声明为全局变量，
否则不能在函数内部把参数变量重新赋值为一个引用该变量名自身的表达式的值。例如，
下面这样是行不通的：

```
CREATE FUNCTION pystrip(x text)
  RETURNS text
AS $$
  x = x.strip()  # 错误
  return x
$$ LANGUAGE plpythonu;
```

因为对`x`赋值会使`x`在整个代码块中都成为局部变量，因此赋值语句右侧的`x`
指向的是尚未赋值的局部变量`x`，而不是 PL/Python 函数参数。使用`global`语句后，这样写就能工作：

```
CREATE FUNCTION pystrip(x text)
  RETURNS text
AS $$
  global x
  x = x.strip()  # 现在可以正常工作
  return x
$$ LANGUAGE plpythonu;
```

但是，建议不要依赖 PL/Python 的这一实现细节。最好将函数参数视为只读。

42.3. 数据值

一般来说，PL/Python 的目标是在 PostgreSQL 世界和 Python 世界之间提供一种“自然”的映射。下面描述的数据映射规则就是基于这一目标。

42.3.1. 数据类型映射

函数参数会被从其 PostgreSQL 类型转换为对应的 Python 类型：

- PostgreSQL 的 `boolean` 被转换为 Python 的 `bool`。
- PostgreSQL 的 `smallint` 和 `int` 被转换为 Python 的 `int`。PostgreSQL 的 `bigint` 在 Python 2 中被转换为 `long`，在 Python 3 中被转换为 `int`。
- PostgreSQL 的 `real`、`double` 和 `numeric` 被转换为 Python 的 `float`。注意对于 `numeric` 这会丢失信息并可能导致不正确的结果。这一点可能会在将来的版本中修复。
- PostgreSQL 的 `bytea` 在 Python 2 中被转换为 `str`，在 Python 3 中被转换为 `bytes`。在 Python 2 中，该字符串应当被当作没有任何字符编码的字节序列。
- 所有其他数据类型，包括 PostgreSQL 的字符串类型，都被转换为 Python 的 `str`。在 Python 2 中，该字符串将采用 PostgreSQL 服务器编码；在 Python 3 中，它将和所有字符串一样是 Unicode 字符串。

- 对于非标量数据类型，请参见下文。

函数返回值会按如下方式转换为声明的 PostgreSQL 返回数据类型：

- 当 PostgreSQL 返回类型为`boolean`时，返回值会按照Python规则进行真值判定。也就是说，0 和空字符串为假，但值得注意的是，`'f'` 为真。
- 如果 PostgreSQL 返回类型是 `bytea`，会先使用相应的 Python 内置函数，将返回值转换为字符串（Python 2）或 bytes（Python 3），再将结果转换为 `bytea`。
- 对于所有其他 PostgreSQL 返回类型，返回值会使用 Python 内置函数`str`转换为字符串，然后将结果传给 PostgreSQL 数据类型的输入函数。

  Python 2 中的字符串传给 PostgreSQL 时，必须采用 PostgreSQL 服务器编码。在当前服务器编码中无效的字符串会引发错误，但并非所有编码不匹配都能被检测到，
  因此处理不当仍可能产生乱码数据。Unicode 字符串会自动转换为正确的编码，因而使用它们可能更安全、更方便。在 Python 3 中，所有字符串都是 Unicode 字符串。

- 对于非标量数据类型，请参见下文。

请注意，即使声明的 PostgreSQL 返回类型与实际返回对象的 Python 数据类型在逻辑上并不匹配，也不会报错；无论如何该值都会被转换。

42.3.2. 空值、None

如果将 SQL 空值传递给函数，该参数值在 Python 中会表现为 `None`。例如，函数 `pymax` 的定义（见第 42.2 节）对空值输入会返回错误结果。可以在函数定义中添加 `STRICT`，让 PostgreSQL 采取更合理的行为：如果传入空值，就完全不调用函数，而是自动返回空值结果。也可以在函数体中检查空值输入：

```
CREATE FUNCTION pymax (a integer, b integer)
  RETURNS integer
AS $$
  if (a is None) or (b is None):
    return None
  if a > b:
    return a
  return b
$$ LANGUAGE plpythonu;
```

如上所示，要从 PL/Python 函数返回 SQL 空值，只需返回 `None`。无论函数是否严格，都可以这样做。

42.3.3. 数组、列表

SQL 数组值会作为 Python 列表传入 PL/Python。要从 PL/Python 函数返回 SQL 数组值，请返回一个 Python 序列，例如列表或元组：

```
CREATE FUNCTION return_arr()
  RETURNS int[]
AS $$
return (1, 2, 3, 4, 5)
$$ LANGUAGE plpythonu;
```

```
SELECT return_arr();
 return_arr
-------------
 {1,2,3,4,5}
(1 row)
```

请注意，在 Python 中，字符串也是序列，这可能会带来一些 Python 程序员熟悉但并不理想的效果：

```
CREATE FUNCTION return_str_arr()
  RETURNS varchar[]
AS $$
return "hello"
$$ LANGUAGE plpythonu;

SELECT return_str_arr();
 return_str_arr
----------------
 {h,e,l,l,o}
(1 row)
```

42.3.4. 复合类型

复合类型参数会以 Python 映射的形式传给函数。映射中的元素名就是复合类型的属性名。如果传入行中的某个属性为空值，那么它在映射中的值就是`None`。例如：

```
CREATE TABLE employee (
  name text,
  salary integer,
  age integer
);

CREATE FUNCTION overpaid (e employee)
  RETURNS boolean
AS $$
  if e["salary"] > 200000:
    return True
  if (e["age"] < 30) and (e["salary"] > 100000):
    return True
  return False
$$ LANGUAGE plpythonu;
```

有多种方法可以从 Python 函数返回行类型或复合类型。以下示例假定我们有：

```
CREATE TYPE named_value AS (
  name   text,
  value  integer
);
```

复合结果可以按以下形式返回：

序列类型（元组或列表，但不能是集合，因为集合不可通过索引访问）

返回的序列对象，其项目数必须与复合结果类型的字段数相同。索引 0 的项目赋给复合类型的第一个字段，索引 1 的项目赋给第二个字段，依此类推。例如：

```
CREATE FUNCTION make_pair (name text, value integer)
  RETURNS named_value
AS $$
  return [ name, value ]
  # 也可以使用元组： return ( name, value )
$$ LANGUAGE plpythonu;
```

要为某一列返回 SQL 空值，请将 `None` 放在对应位置。

映射（字典）

结果类型中每一列的值，都使用列名作为键从映射中取得。例如：

```
CREATE FUNCTION make_pair (name text, value integer)
  RETURNS named_value
AS $$
  return { "name": name, "value": value }
$$ LANGUAGE plpythonu;
```

字典中多余的键值对会被忽略，缺少键则会被视为错误。要为某一列返回 SQL 空值，请插入 `None`，并以对应列名为键。

对象（提供方法`__getattr__`的任何对象）

其工作方式与映射相同。例如：

```
CREATE FUNCTION make_pair (name text, value integer)
  RETURNS named_value
AS $$
  class named_value:
    def __init__ (self, n, v):
      self.name = n
      self.value = v
  return named_value(name, value)

  # 或者简写为
  class nv: pass
  nv.name = name
  nv.value = value
  return nv
$$ LANGUAGE plpythonu;
```

也支持带`OUT`参数的函数。例如：

```
CREATE FUNCTION multiout_simple(OUT i integer, OUT j integer) AS $$
return (1, 2)
$$ LANGUAGE plpythonu;

SELECT * FROM multiout_simple();
```

42.3.5. 集合返回函数

PL/Python函数也可以返回标量类型或复合类型的集合。实现方式有多种，
因为返回的对象在内部会被转换成一个迭代器。以下示例假定我们有如下复合类型：

```
CREATE TYPE greeting AS (
```

```
  how text,
  who text
);
```

集合结果可以通过以下对象返回：

序列类型（元组、列表、集合）

```
CREATE FUNCTION greet (how text)
  RETURNS SETOF greeting
AS $$
  # 返回包含列表的元组，以列表表示复合类型
  # 其他组合方式也都可用
  return ( [ how, "World" ], [ how, "PostgreSQL" ], [ how, "PL/
Python" ] )
$$ LANGUAGE plpythonu;
```

迭代器（任何提供 `__iter__` 和 `next` 方法的对象）

```
CREATE FUNCTION greet (how text)
  RETURNS SETOF greeting
AS $$
  class producer:
    def __init__ (self, how, who):
      self.how = how
      self.who = who
      self.ndx = -1

    def __iter__ (self):
      return self

    def next (self):
      self.ndx += 1
      if self.ndx == len(self.who):
        raise StopIteration
      return ( self.how, self.who[self.ndx] )

  return producer(how, [ "World", "PostgreSQL", "PL/Python" ])
$$ LANGUAGE plpythonu;
```

生成器（`yield`）

```
CREATE FUNCTION greet (how text)
  RETURNS SETOF greeting
AS $$
  for who in [ "World", "PostgreSQL", "PL/Python" ]:
    yield ( how, who )
$$ LANGUAGE plpythonu;
```

警告

由于 Python 的缺陷 #1483133[3]，某些调试版本的 Python 2.4（以`--with-pydebug`选项配置并编译）

[3] http://bugs.python.org/issue1483133

> 在使用迭代器返回集合结果时会导致PostgreSQL服务器崩溃。未打补丁的 Fedora 4 就包含此缺陷。在 Python 的正式版本或已打补丁的 Fedora 4 中不会发生这种情况。

也支持带OUT参数的集合返回函数（使用RETURNS SETOF record）。例如：

```
CREATE FUNCTION multiout_simple_setof(n integer, OUT integer, OUT
 integer) RETURNS SETOF record AS $$
return [(1, 2)] * n
$$ LANGUAGE plpythonu;

SELECT * FROM multiout_simple_setof(3);
```

42.4. 共享数据

全局字典 SD 可用于在函数调用之间保存数据。该变量是私有静态数据。全局字典 GD 则是公共数据，可供一个会话中的所有 Python 函数使用。使用时要小心。

每个函数在 Python 解释器中都有自己的执行环境，因此myfunc中的全局数据和函数参数对myfunc2不可见。例外是前面提到的GD字典中的数据。

42.5. 匿名代码块

PL/Python 也支持通过DO语句调用的匿名代码块：

```
DO $$
    # PL/Python 代码
$$ LANGUAGE plpythonu;
```

匿名代码块不接受任何参数，而且无论返回什么值都会被丢弃。除此之外，
它的行为与函数完全相同。

42.6. 触发器函数

当函数被用作触发器时，字典TD包含与触发器相关的值：

TD["event"]

以字符串形式包含事件：INSERT、UPDATE、DELETE或者TRUNCATE。

TD["when"]

包含BEFORE、AFTER或者INSTEAD OF之一。

TD["level"]

包含ROW或者STATEMENT。

TD["new"]
TD["old"]

对于行级触发器，这两个字段中的一个或两个会根据触发器事件包含相应的触发行。

`TD["name"]`

包含触发器名称。

`TD["table_name"]`

包含触发器所在表的名称。

`TD["table_schema"]`

包含触发器所在表的模式。

`TD["relid"]`

包含触发器所在表的 OID。

`TD["args"]`

如果`CREATE TRIGGER`命令包含参数，这些参数可在`TD["args"][0]`到`TD["args"][n-1]`中取得。

如果`TD["when"]`是`BEFORE`或`INSTEAD OF`，且`TD["level"]`是`ROW`，那么可以从 Python 函数返回`None`或`"OK"`来表示该行未被修改，返回`"SKIP"`来中止该事件；如果`TD["event"]`是`INSERT`或`UPDATE`，还可以返回`"MODIFY"`来表示你已经修改了新行。否则返回值会被忽略。

42.7. 数据库访问

PL/Python 语言模块会自动导入一个名为`plpy`的 Python 模块。该模块中的函数和常量在 Python 代码中可以通过`plpy.`*`foo`*这样的形式访问。

42.7.1. 数据库访问函数

`plpy` 模块提供了两个函数：`execute` 和 `prepare`。使用查询字符串和一个可选的限制参数调用`plpy.execute`会执行该查询，并把结果在一个结果对象中返回。结果对象模拟了列表或字典对象。可以通过行号和列名访问结果对象。它还有这些附加方法：`nrows` 返回查询返回的行数，`status` 是 `SPI_execute()` 的返回值。结果对象可以被修改。

例如：

```
rv = plpy.execute("SELECT * FROM my_table", 5)
```

会从 `my_table` 中返回最多 5 行。如果 `my_table` 有一列名为 `my_column`，则可以这样访问它：

```
foo = rv[i]["my_column"]
```

第二个函数 `plpy.prepare` 为查询准备执行计划。若查询中有参数引用，则它以查询字符串和参数类型列表作为参数。例如：

```
plan = plpy.prepare("SELECT last_name FROM my_users WHERE first_name =
 $1", [ "text" ])
```

`text` 是你要传给 `$1` 的变量的类型。准备好语句之后，使用函数 `plpy.execute` 运行它：

```
rv = plpy.execute(plan, [ "name" ], 5)
```

第三个参数是限制，它是可选的。

查询参数和结果行字段会按照第 42.3 节中所述，在 PostgreSQL 与 Python 数据类型之间进行转换。

当你使用 PL/Python 模块准备一个计划时，它会被自动保存。关于这意味着什么，请参阅 SPI 文档（第 43 章）。为了在函数调用之间有效利用这一点，需要使用持久存储字典 `SD` 或 `GD` 之一（见第 42.4 节）。例如：

```
CREATE FUNCTION usesavedplan() RETURNS trigger AS $$
    if SD.has_key("plan"):
        plan = SD["plan"]
    else:
        plan = plpy.prepare("SELECT 1")
        SD["plan"] = plan
    # rest of function
$$ LANGUAGE plpythonu;
```

42.7.2. 捕获错误

访问数据库的函数可能会遇到错误，这会导致它们中止并抛出异常。`plpy.execute`和`plpy.prepare`都可能抛出`plpy.SPIError`某个子类的实例，默认情况下这会终止函数。这个错误可以像其他 Python 异常一样，通过`try/except`结构来处理。例如：

```
CREATE FUNCTION try_adding_joe() RETURNS text AS $$
    try:
        plpy.execute("INSERT INTO users(username) VALUES ('joe')")
    except plpy.SPIError:
        return "something went wrong"
    else:
        return "Joe added"
$$ LANGUAGE plpythonu;
```

所抛出异常的实际类，对应引发错误的具体条件。参见表 A.1中列出的可能条件。模块 `plpy.spiexceptions` 为每种 PostgreSQL 条件定义了一个异常类，类名由条件名派生。例如，`division_by_zero` 变为 `DivisionByZero`，`unique_violation` 变为 `UniqueViolation`，`fdw_error` 变为 `FdwError`，依此类推。所有这些异常类都继承自 `SPIError`。这样区分之后，更容易处理特定错误，例如：

```
CREATE FUNCTION insert_fraction(numerator int, denominator int)
 RETURNS text AS $$
from plpy import spiexceptions
try:
    plan = plpy.prepare("INSERT INTO fractions (frac) VALUES ($1 / $2)
", ["int", "int"])
    plpy.execute(plan, [numerator, denominator])
except spiexceptions.DivisionByZero:
    return "denominator cannot equal zero"
except spiexceptions.UniqueViolation:
    return "already have that fraction"
except plpy.SPIError, e:
    return "other error, SQLSTATE %s" % e.sqlstate
else:
    return "fraction inserted"
$$ LANGUAGE plpythonu;
```

注意，因为 `plpy.spiexceptions` 模块中的所有异常都继承自 `SPIError`，所以处理它的 `except` 子句会捕获任何数据库访问错误。

作为处理不同错误情况的另一种方法，你可以捕获`SPIError`异常，并在`except`块中通过查看异常对象的`sqlstate`属性来判断具体的错误条件。该属性是一个包含“SQLSTATE”错误代码的字符串值。这种方法大体上提供了相同的功能。

42.8. 显式子事务

如第 42.7.2 节中所述，从数据库访问引发的错误中恢复，可能会造成一种不理想的情况：在某个操作失败之前，其他一些操作已经成功，而在从该错误恢复后，数据却处于不一致状态。PL/Python 以显式子事务的形式为这个问题提供了解决方案。

42.8.1. 子事务上下文管理器

考虑以下实现两个账户之间转账的函数：

```
CREATE FUNCTION transfer_funds() RETURNS void AS $$
try:
    plpy.execute("UPDATE accounts SET balance = balance - 100 WHERE
 account_name = 'joe'")
    plpy.execute("UPDATE accounts SET balance = balance + 100 WHERE
 account_name = 'mary'")
except plpy.SPIError, e:
    result = "error transferring funds: %s" % e.args
else:
    result = "funds transferred correctly"
plan = plpy.prepare("INSERT INTO operations (result) VALUES ($1)", [
"text"])
plpy.execute(plan, [result])
$$ LANGUAGE plpythonu;
```

如果第二条 `UPDATE` 语句引发异常，此函数会报告错误，但第一条 `UPDATE` 的结果仍会提交。换句话说，资金会从 Joe 的账户中扣除，却不会转入 Mary 的账户。

为避免此类问题，可以将 `plpy.execute` 调用放在显式子事务中。`plpy` 模块提供了用于管理显式子事务的辅助对象，可通过 `plpy.subtransaction()` 函数创建。此函数创建的对象实现了上下文管理器接口[4]。使用显式子事务后，可以将函数改写为：

```
CREATE FUNCTION transfer_funds2() RETURNS void AS $$
try:
    with plpy.subtransaction():
        plpy.execute("UPDATE accounts SET balance = balance - 100
 WHERE account_name = 'joe'")
        plpy.execute("UPDATE accounts SET balance = balance + 100
 WHERE account_name = 'mary'")
except plpy.SPIError, e:
    result = "error transferring funds: %s" % e.args
else:
    result = "funds transferred correctly"
plan = plpy.prepare("INSERT INTO operations (result) VALUES ($1)", [
"text"])
```

[4] http://docs.python.org/library/stdtypes.html#context-manager-types

```
plpy.execute(plan, [result])
$$ LANGUAGE plpythonu;
```

注意，仍需要使用 `try/catch`。否则，异常会传播到 Python 调用栈顶层，使整个函数因 PostgreSQL 错误而中止，从而不会向 `operations` 表插入任何行。子事务上下文管理器不会捕获错误，只保证在其作用域内执行的所有数据库操作以原子方式提交或回滚。
任何异常退出都会使子事务块回滚，并不限于数据库访问错误。显式子事务块中抛出的普通 Python 异常，也会导致该子事务回滚。

42.8.2. 较早的 Python 版本

使用 `with` 关键字的上下文管理器语法，从 Python 2.6 起默认可用。如果 PL/Python 使用更早的 Python 版本，仍然可以使用显式子事务，只是没那么方便。可以通过便捷别名 `enter` 和 `exit`，调用子事务管理器的 `__enter__` 和 `__exit__` 函数。转账示例函数可以写成：

```
CREATE FUNCTION transfer_funds_old() RETURNS void AS $$
try:
    subxact = plpy.subtransaction()
    subxact.enter()
    try:
        plpy.execute("UPDATE accounts SET balance = balance - 100
 WHERE account_name = 'joe'")
        plpy.execute("UPDATE accounts SET balance = balance + 100
 WHERE account_name = 'mary'")
    except:
        import sys
        subxact.exit(*sys.exc_info())
        raise
    else:
        subxact.exit(None, None, None)
except plpy.SPIError, e:
    result = "error transferring funds: %s" % e.args
else:
    result = "funds transferred correctly"

plan = plpy.prepare("INSERT INTO operations (result) VALUES ($1)", [
"text"])
plpy.execute(plan, [result])
$$ LANGUAGE plpythonu;
```

> **注意**
>
> 虽然 Python 2.5 已实现上下文管理器，但在该版本中使用 `with` 语法，需要使用 future 语句[5]。不过，由于实现细节的限制，PL/Python 函数中不能使用 future 语句。

42.9. 辅助函数

`plpy`模块还提供以下函数：`plpy.debug(msg)`、`plpy.log(msg)`、`plpy.info(msg)`、`plpy.notice(msg)`、`plpy.warning(msg)`、`plpy.error(msg)` 和 `plpy.fatal(`

[5] http://docs.python.org/release/2.5/ref/future.html

msg)。plpy.error和plpy.fatal实际上会引发 Python 异常；如果这些异常未被捕获，就会传播到调用查询，导致当前事务或子事务中止。raise plpy.Error(*msg*)和raise plpy.Fatal(*msg*)分别等价于调用plpy.error和plpy.fatal。其他函数只会生成不同优先级的消息。某一优先级的消息是报告给客户端、写入服务器日志还是两者兼有，由配置变量log_min_messages和client_min_messages控制。更多信息见第 18 章。

另一组辅助函数是plpy.quote_literal(*string*)、plpy.quote_nullable(*string*)以及plpy.quote_ident(*string*)。它们等价于第 9.4 节中描述的内置加引号函数。在构造临时查询时，这些函数很有用。例 39.1中动态 SQL 的一个 PL/Python 等价写法如下：

```
plpy.execute("UPDATE tbl SET %s = %s WHERE key = %s" % (
    plpy.quote_ident(colname),
    plpy.quote_nullable(newvalue),
    plpy.quote_literal(keyvalue)))
```

42.10. 环境变量

Python 解释器接受的一些环境变量也可以用来影响 PL/Python 的行为。它们需要在主 PostgreSQL 服务器进程的环境中设置，例如在启动脚本中。可用的环境变量取决于 Python 的版本；详情见 Python 文档。撰写本文时，假设 Python 版本足够新，下列环境变量会影响 PL/Python：

- PYTHONHOME
- PYTHONPATH
- PYTHONY2K
- PYTHONOPTIMIZE
- PYTHONDEBUG
- PYTHONVERBOSE
- PYTHONCASEOK
- PYTHONDONTWRITEBYTECODE
- PYTHONIOENCODING
- PYTHONUSERBASE

(It appears to be a Python implementation detail beyond the control of PL/Python that some of the environment variables listed on the python man page are only effective in a command-line interpreter and not an embedded Python interpreter.)

第 43 章 服务器编程接口

服务器编程接口（SPI）使用户定义 C 函数的编写者能够在其函数中运行 SQL 命令。SPI 是一组接口函数，用于简化对解析器、规划器和执行器的访问。SPI 还负责一部分内存管理工作。

> **注意**
>
> 可用的过程语言提供了多种从过程中执行 SQL 命令的方法。这些设施大多基于 SPI，因此本文档对这些语言的用户也会有帮助。

为避免误解，本文用“函数”指SPI接口函数，而用“过程”指使用SPI的用户定义 C 函数。

注意，如果通过 SPI 调用的某条命令失败，控制不会返回到你的过程。相反，执行该过程的事务或子事务会被回滚。（考虑到 SPI 函数大多记录了错误返回约定，这一点可能看起来有些意外。不过，这些约定只适用于在 SPI 函数自身内部检测到的错误。）如果在可能失败的 SPI 调用外围建立自己的子事务，则可以在出错后重新取得控制权。

SPI 函数在成功时返回非负结果（要么直接作为整数返回值，要么如后文所述存放在全局变量 `SPI_result` 中）。发生错误时，则返回负值或 `NULL`。

使用 SPI 的源代码文件必须包含头文件 `executor/spi.h`。

43.1. 接口函数

SPI_connect

SPI_connect — connect a procedure to the SPI manager

大纲

```
int SPI_connect(void)
```

描述

`SPI_connect` 会为某次过程调用打开到 SPI 管理器的连接。如果要通过 SPI 执行命令，就必须调用此函数。不过，有些 SPI 辅助函数可以在未连接的过程中调用。

如果你的过程已经处于连接状态，`SPI_connect`将返回错误代码→ `SPI_ERROR_CONNECT`。当某个直接调用了`SPI_connect`的过程又直接调用另一个调用`SPI_connect`的过程时，就会发生这种情况。虽然当通过 SPI 调用的 SQL 命令调用另一个使用 SPI 的函数时，允许对 SPI 管理器的递归调用，但直接嵌套调用`SPI_connect`和`SPI_finish`是被禁止的。（但请参见`SPI_push`和`SPI_pop`。）

返回值

`SPI_OK_CONNECT`

成功时

`SPI_ERROR_CONNECT`

发生错误时

SPI_finish

SPI_finish — disconnect a procedure from the SPI manager

大纲

```
int SPI_finish(void)
```

描述

`SPI_finish` 关闭到 SPI 管理器的现有连接。在完成当前这次过程调用所需的 SPI 操作后，必须调用此函数。不过，如果通过 `elog(ERROR)` 中止了事务，就不必担心是否显式调用它；这种情况下 SPI 会自行清理。

如果在没有有效连接的情况下调用`SPI_finish`，它将返回`SPI_ERROR_UNCONNECTED`。这并不会带来根本性问题；它只表示 SPI 管理器无事可做。

返回值

`SPI_OK_FINISH`

如果正确地断开连接

`SPI_ERROR_UNCONNECTED`

如果从未连接的过程中调用

SPI_push

SPI_push — push SPI stack to allow recursive SPI usage

大纲

```
void SPI_push(void)
```

描述

SPI_push应当在执行另一个自身可能希望使用 SPI 的过程之前调用。在SPI_push之后，SPI 不再处于“连接”状态，除非重新执行SPI_connect，否则 SPI 函数调用将被拒绝。这确保了你的过程的 SPI 状态与你所调用的另一个过程的状态之间的清晰分离。在另一个过程返回之后，调用SPI_pop恢复对你自己 SPI 状态的访问。

注意，SPI_execute及相关函数在把控制交还给 SQL 执行引擎之前，会自动完成相当于SPI_push的操作，因此在使用这些函数时无需为此操心。只有当你直接调用可能包含SPI_connect调用的任意代码时，才需要发出SPI_push和SPI_pop。

SPI_pop

SPI_pop — pop SPI stack to return from recursive SPI usage

大纲

```
void SPI_pop(void)
```

描述

`SPI_pop`从 SPI 调用栈中弹出先前的环境。参见`SPI_push`。

SPI_execute

SPI_execute — execute a command

大纲

```
int SPI_execute(const char * command, bool read_only, long count)
```

描述

SPI_execute 执行指定的 SQL 命令，并最多检索 *count* 行。如果 *read_only* 为 true，该命令必须是只读的，且执行开销会略有降低。

此函数只能从已连接的过程中调用。

如果 *count* 为零，则该命令会针对其适用的所有行执行。如果 *count* 大于零，则最多检索 *count* 行；达到该计数时就会停止执行，这很像给查询增加了一个 LIMIT 子句。例如：

```
SPI_execute("SELECT * FROM foo", true, 5);
```

最多会从该表中检索 5 行。注意，这种限制只有在命令实际返回行时才有效。例如：

```
SPI_execute("INSERT INTO foo SELECT * FROM bar", false, 5);
```

会忽略 *count* 参数，把 bar 中的所有行都插入进去。不过：

```
SPI_execute("INSERT INTO foo SELECT * FROM bar RETURNING *", false, 5)
;
```

最多只会插入 5 行，因为取到第 5 行 RETURNING 结果后就会停止执行。

你可以在一个字符串中传递多条命令，
但后面的命令不能依赖于该字符串中先前命令所创建的对象，
因为整个字符串会在执行开始之前被完整解析和规划。SPI_execute 返回最后执行的那条命令的结果。*count* 限制会分别作用于每条命令（尽管实际返回的只有最后一条命令的结果）。该限制不适用于规则生成的任何隐藏命令。

当 *read_only* 为 false 时，SPI_execute 会递增命令计数器，并在执行字符串中的每条命令前计算新的快照。如果当前事务隔离级别是 SERIALIZABLE 或 REPEATABLE READ，这个快照实际上不会变化；但在 READ COMMITTED 模式下，更新快照会让每条命令都能看到其他会话中新近提交事务的结果。这对于修改数据库的命令获得一致行为至关重要。

当 *read_only* 为 true 时，SPI_execute 不会更新快照和命令计数器，并且只允许命令字符串中出现普通的 SELECT 命令。这些命令会使用外围查询先前建立的快照来执行。由于消除了每条命令的额外开销，这种执行模式比读写模式略快。它还允许构造真正
稳定的函数：由于连续执行都会使用同一个快照，结果也就不会发生变化。

在同一个使用 SPI 的函数中混合只读命令和读写命令通常并不明智，因为只读查询看不到读写查询所做的数据库更新，这可能导致非常令人困惑的行为。

（最后一条）命令实际执行所处理的行数，会通过全局变量 SPI_processed 返回。如果函数返回值是 SPI_OK_SELECT、SPI_OK_INSERT_RETURNING、SPI_OK_DELETE_RETURNING 或 SPI_OK_UPDATE_RETURNING，则可以通过全局指针 SPITupleTable *SPI_tuptable 访问结果行。有些工具命令（如 EXPLAIN）也会返回结果行集，此时 SPI_tuptable 同样会保存结果。

结构 SPITupleTable 定义如下：

```
typedef struct
{
    MemoryContext tuptabcxt;    /* 结果表的内存上下文 */
    uint32      alloced;        /* 已分配的 vals 数量 */
    uint32      free;           /* 空闲的 vals 数量 */
    TupleDesc   tupdesc;        /* 行描述符 */
    HeapTuple  *vals;           /* 行 */
} SPITupleTable;
```

vals 是一个指向行的指针数组。（有效项数为 SPI_processed。）tupdesc 是一个行描述符，可以传给处理行的 SPI 函数。tuptabcxt，alloced 和 free 是内部字段，不供 SPI 调用者使用。

SPI_finish 会释放当前过程调用期间分配的全部 SPITupleTable。如果某个结果表已经不再需要，也可以提前调用 SPI_freetuptable 释放它。

参数

const char * *command*

包含待执行命令的字符串

bool *read_only*

true 表示只读执行

long *count*

要返回的最大行数，或者用 0 表示不限制

返回值

如果命令执行成功，则返回下列（非负）值之一：

SPI_OK_SELECT

执行了 SELECT（但不是 SELECT INTO）

SPI_OK_SELINTO

执行了 SELECT INTO

SPI_OK_INSERT

执行了 INSERT

SPI_OK_DELETE

执行了 DELETE

SPI_OK_UPDATE

执行了 UPDATE

`SPI_OK_INSERT_RETURNING`

执行了 `INSERT RETURNING`

`SPI_OK_DELETE_RETURNING`

执行了 `DELETE RETURNING`

`SPI_OK_UPDATE_RETURNING`

执行了 `UPDATE RETURNING`

`SPI_OK_UTILITY`

执行了工具命令（例如 `CREATE TABLE`）

`SPI_OK_REWRITTEN`

命令被规则重写成了另一类命令（例如 `UPDATE` 变成了 `INSERT`）

出错时，返回以下负值之一：

`SPI_ERROR_ARGUMENT`

command 为 `NULL`，或者 *count* 小于 0

`SPI_ERROR_COPY`

尝试执行了 `COPY TO stdout` 或 `COPY FROM stdin`

`SPI_ERROR_TRANSACTION`

尝试执行了事务控制命令（`BEGIN`、`COMMIT`、`ROLLBACK`、`SAVEPOINT`、`PREPARE TRANSACTION`、`COMMIT PREPARED`、`ROLLBACK PREPARED` 及其各种变体）

`SPI_ERROR_OPUNKNOWN`

命令类型未知（理论上不应发生）

`SPI_ERROR_UNCONNECTED`

如果从未连接的过程中调用

注解

所有 SPI 查询执行函数都会设置 `SPI_processed` 和 `SPI_tuptable`（只设置指针，而不更改结构体内容）。如果需要在后续调用之后继续访问 `SPI_execute` 或其他查询执行函数的结果表，请把这两个全局变量保存到过程局部变量中。

SPI_exec

SPI_exec — execute a read/write command

大纲

```
int SPI_exec(const char * command, long count)
```

描述

`SPI_exec`和 `SPI_execute`相同，但后者的 *`read_only`*参数的值总是取 `false`。

参数

`const char * command`

包含待执行命令的字符串

`long count`

要返回的最大行数，或者用 `0` 表示不限制

返回值

见`SPI_execute`。

SPI_execute_with_args

SPI_execute_with_args — execute a command with out-of-line parameters

大纲

```
int SPI_execute_with_args(const char *command,
                          int nargs, Oid *argtypes,
                          Datum *values, const char *nulls,
                          bool read_only, long count)
```

描述

`SPI_execute_with_args` 执行一条可能包含外部提供参数引用的命令。命令文本中的参数引用写作 `$n`，而该调用会为每个这类符号指定数据类型和值。*read_only* 和 *count* 的含义与 `SPI_execute` 中相同。

相比 `SPI_execute`，这个例程的主要优点在于可将数据值插入命令而无需繁琐的引用和转义，因此能够显著降低 SQL 注入攻击的风险。

也可以通过先调用 `SPI_prepare`，再调用 `SPI_execute_plan` 达到类似效果。不过，使用本函数时，查询计划会针对所提供的具体参数值进行定制。对于一次性查询执行，
应优先选择本函数。如果同一条命令要用许多不同参数重复执行，
则两种方式孰快取决于重新规划的代价与定制计划收益之间的权衡。

参数

`const char *` *command*

命令字符串

`int` *nargs*

输入参数的数量（`$1`、`$2` 等）

`Oid *` *argtypes*

一个数组，包含各参数数据类型的 OID

`Datum *` *values*

一个实际参数值的数组

`const char *` *nulls*

一个描述哪些参数为空值的数组

如果 *nulls* 为 `NULL`，则 `SPI_execute_with_args` 假定没有参数为空值。

`bool` *read_only*

`true` 表示只读执行

`long` *count*

要返回的最大行数，或者用 `0` 表示不限制

返回值

返回值与 `SPI_execute` 相同。

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute` 相同。

SPI_prepare

SPI_prepare — prepare a plan for a command, without executing it yet

大纲

```
SPIPlanPtr SPI_prepare(const char * command, int nargs, Oid
 * argtypes)
```

描述

SPI_prepare 为指定命令创建并返回一个执行计划，但不执行该命令。该函数只能从已连接的过程调用。

当同一条或相似命令需要反复执行时，只做一次规划可能是有利的。SPI_prepare 把命令字符串转换成一个执行计划，之后可以使用 SPI_execute_plan 反复执行。

预备命令可以通过在普通命令原本应写常量的位置写入参数（$1、$2 等）而得到泛化。这些参数的实际值会在调用 SPI_execute_plan 时指定。这样，预备语句就能适用于比无参形式更广泛的场景。

SPI_prepare 返回的计划只能在当前这次过程调用中使用，因为 SPI_finish 会释放为计划分配的内存。但可以使用函数 SPI_saveplan 把计划保存得更久。

参数

const char * *command*

命令字符串

int *nargs*

输入参数的数量（$1、$2 等）

Oid * *argtypes*

一个数组指针，它指向的数组包含参数的数据类型的 OID

返回值

SPI_prepare 返回一个指向执行计划的非空指针。发生错误时会返回 NULL，并将 SPI_result 设为 SPI_execute 所使用的那些错误码之一；但如果 *command* 为 NULL，或者 *nargs* 小于 0，或者 *nargs* 大于 0 且 *argtypes* 为 NULL，则会将其设置为 SPI_ERROR_ARGUMENT。

注解

SPIPlanPtr 在 spi.h 中被声明为指向不透明结构体类型的指针。尝试直接访问其内容并不明智，因为这会让你的代码在 PostgreSQL 后续版本中更容易失效。

使用参数有一个不利之处：由于规划器不知道将为参数提供什么值，
它做出的规划选择可能比在所有常量都可见的普通命令下更差。

SPI_prepare_cursor

SPI_prepare_cursor — prepare a plan for a command, without executing it yet

大纲

```
SPIPlanPtr SPI_prepare_cursor(const char * command, int nargs,
                              Oid * argtypes, int cursorOptions)
```

描述

SPI_prepare_cursor 与 SPI_prepare 相同，但额外允许指定规划器的“游标选项”参数。该参数是一个位掩码，其可取值对应于 nodes/parsenodes.h 中 DeclareCursorStmt 的 options 字段。SPI_prepare 总是把游标选项设为零。

参数

const char * *command*

命令字符串

int *nargs*

输入参数的数量（$1、$2 等）

Oid * *argtypes*

一个数组指针，它指向的数组包含参数的数据类型的 OID

int *cursorOptions*

整数形式的游标选项位掩码，零会导致默认行为

返回值

SPI_prepare_cursor 的返回约定与 SPI_prepare 相同。

注解

在 *cursorOptions* 中可设置的有用位包括 CURSOR_OPT_SCROLL、CURSOR_OPT_NO_SCROLL 和 CURSOR_OPT_FAST_PLAN。特别要注意，CURSOR_OPT_HOLD 会被忽略。

SPI_prepare_params

SPI_prepare_params — prepare a plan for a command, without executing it yet

大纲

```
SPIPlanPtr SPI_prepare_params(const char * command,
                              ParserSetupHook parserSetup,
                              void * parserSetupArg,
                              int cursorOptions)
```

描述

`SPI_prepare_params` 为指定命令创建并返回一个执行计划，但不执行该命令。该函数等价于 `SPI_prepare_cursor`，只是额外允许调用者指定解析器钩子函数，以控制外部参数引用的解析。

参数

`const char * command`

命令字符串

`ParserSetupHook parserSetup`

解析器钩子设置函数

`void * parserSetupArg`

传给 *parserSetup* 的透传参数

`int cursorOptions`

整数形式的游标选项位掩码，零会导致默认行为

返回值

`SPI_prepare_params` 的返回约定与 `SPI_prepare` 相同。

SPI_getargcount

SPI_getargcount — return the number of arguments needed by a plan prepared by `SPI_prepare`

大纲

```
int SPI_getargcount(SPIPlanPtr plan)
```

描述

`SPI_getargcount` 返回执行由 `SPI_prepare` 准备的计划所需的参数个数。

参数

`SPIPlanPtr` *`plan`*

执行计划（由 `SPI_prepare` 返回）

返回值

`plan` 所需的参数个数。如果 *`plan`* 为 `NULL` 或无效，则 `SPI_result` 会被设置为 `SPI_ERROR_ARGUMENT`，并返回 -1。

SPI_getargtypeid

SPI_getargtypeid — return the data type OID for an argument of a plan prepared by `SPI_prepare`

大纲

`Oid SPI_getargtypeid(SPIPlanPtr plan, int argIndex)`

描述

`SPI_getargtypeid` 返回由 `SPI_prepare` 准备的计划中第 *argIndex* 个参数的类型的 OID。第一个参数的下标为零。

参数

`SPIPlanPtr plan`

执行计划（由 `SPI_prepare` 返回）

`int argIndex`

参数的索引，从零开始

返回值

给定索引处参数的类型 OID。如果 *plan* 为 `NULL` 或无效，或者 *argIndex* 小于 0，或者不小于为 *plan* 声明的参数个数，则 `SPI_result` 会被设置为 `SPI_ERROR_ARGUMENT`，并返回 `InvalidOid`。

SPI_is_cursor_plan

SPI_is_cursor_plan — return `true` if a plan prepared by `SPI_prepare` can be used with `SPI_cursor_open`

大纲

```
bool SPI_is_cursor_plan(SPIPlanPtr plan)
```

描述

如果由 `SPI_prepare` 准备的计划可以作为参数传给 `SPI_cursor_open`，则 `SPI_is_cursor_plan` 返回 `true`；否则返回 `false`。判定标准是 *plan* 必须表示单条命令，并且该命令会向调用者返回元组。例如，不包含 `INTO` 子句的 `SELECT` 是允许的，而 `UPDATE` 只包含 `RETURNING` 子句时才允许。

参数

`SPIPlanPtr` *plan*

执行计划（由 `SPI_prepare` 返回）

返回值

返回 `true` 或 `false`，表示 *plan* 能否产生游标，同时将 `SPI_result` 设为零。如果无法确定答案（例如 *plan* 为 `NULL` 或无效，或者在未连接到 SPI 时调用），则会将 `SPI_result` 设为合适的错误码，并返回 `false`。

SPI_execute_plan

SPI_execute_plan — execute a plan prepared by `SPI_prepare`

大纲

```
int SPI_execute_plan(SPIPlanPtr plan, Datum * values, const char
 * nulls,
                     bool read_only, long count)
```

描述

`SPI_execute_plan` 执行由 `SPI_prepare` 准备的计划。*`read_only`* 和 *`count`* 的含义与 `SPI_execute` 中相同。

参数

`SPIPlanPtr` *`plan`*

执行计划（由 `SPI_prepare` 返回）

`Datum *` *`values`*

一个实际参数值的数组。必须与该计划的参数个数等长。

`const char *` *`nulls`*

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。`n` 表示空值（*`values`* 中的对应项会被忽略）；空格表示非空值（*`values`* 中的对应项有效）。

如果 *`nulls`* 为 `NULL`，则 `SPI_execute_plan` 假定没有参数为空值。

`bool` *`read_only`*

`true` 表示只读执行

`long` *`count`*

要返回的最大行数，或者用 `0` 表示不限制

返回值

返回值和`SPI_execute`相同，还有下列额外可能的错误（负值）结果：

`SPI_ERROR_ARGUMENT`

如果*`plan`*为`NULL` 或者非法，或者*`count`*小于 0

`SPI_ERROR_PARAM`

如果*`values`*为`NULL`，但 *`plan`*在准备时带有参数

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute` 相同。

SPI_execute_plan_with_paramlist

SPI_execute_plan_with_paramlist — execute a plan prepared by `SPI_prepare`

大纲

```
int SPI_execute_plan_with_paramlist(SPIPlanPtr plan,
                                    ParamListInfo params,
                                    bool read_only,
                                    long count)
```

描述

`SPI_execute_plan_with_paramlist` 执行由 `SPI_prepare` 准备好的计划。该函数等价于 `SPI_execute_plan`，只是向查询传递参数值信息的形式不同。`ParamListInfo` 表示形式便于传递已经按这种格式存在的值，也支持通过 `ParamListInfo` 中指定的钩子函数使用动态参数集。

参数

`SPIPlanPtr plan`

执行计划（由 `SPI_prepare` 返回）

`ParamListInfo params`

包含参数类型和值的数据结构；没有参数时为 NULL

`bool read_only`

`true` 表示只读执行

`long count`

要返回的最大行数，或者用 `0` 表示不限制

返回值

返回值与 `SPI_execute_plan` 相同。

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute_plan` 相同。

SPI_execp

SPI_execp — execute a plan in read/write mode

大纲

```
int SPI_execp(SPIPlanPtr plan, Datum * values, const char * nulls,
 long count)
```

描述

SPI_execp 与 SPI_execute_plan 相同，只不过后者的 *read_only* 参数固定为 false。

参数

SPIPlanPtr *plan*

执行计划（由 SPI_prepare 返回）

Datum * *values*

一个实际参数值的数组。必须与该计划的参数个数等长。

const char * *nulls*

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。n 表示空值（*values* 中的对应项会被忽略）；空格表示非空值（*values* 中的对应项有效）。

如果 *nulls* 为 NULL，则 SPI_execp 假定没有参数为空值。

long *count*

要返回的最大行数，或者用 0 表示不限制

返回值

见 SPI_execute_plan。

成功时，SPI_processed 和 SPI_tuptable 的设置方式与 SPI_execute 相同。

SPI_cursor_open

SPI_cursor_open — set up a cursor using a plan created with `SPI_prepare`

大纲

```
Portal SPI_cursor_open(const char * name, SPIPlanPtr plan,
                       Datum * values, const char * nulls,
                       bool read_only)
```

描述

`SPI_cursor_open` 建立一个游标（在内部是一个 portal），该游标将执行由 `SPI_prepare` 准备好的一个计划。参数具有和 `SPI_execute_plan` 的相应参数相同的含义。

使用游标而不是直接执行该计划有两个好处。首先，可以一次只取出少量结果行，从而避免对返回大量行的查询耗尽内存。其次，一个 portal 可以比当前的过程存活更久（实际上，它可以一直存活到当前事务结束）。把 portal 的名称返回给该过程的调用者，就提供了一种把行集作为结果返回的方法。

传入的参数数据会被复制到该游标的 portal 中，因此即使该游标仍然存在，也可以释放这些参数数据。

参数

`const char * name`

portal 的名称，或为`NULL`以让系统选择名称

`SPIPlanPtr plan`

执行计划（由 `SPI_prepare` 返回）

`Datum * values`

一个实际参数值的数组。必须与该计划的参数个数等长。

`const char * nulls`

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。`n` 表示空值（*values* 中的对应项会被忽略）；空格表示非空值（*values* 中的对应项有效）。

如果 *nulls* 为 `NULL`，则 `SPI_cursor_open` 假定没有参数为空值。

`bool read_only`

`true` 表示只读执行

返回值

指向包含该游标的 portal 的指针。注意这里没有错误返回约定，任何错误都将通过`elog`报告。

SPI_cursor_open_with_args

SPI_cursor_open_with_args — set up a cursor using a query and parameters

大纲

```
Portal SPI_cursor_open_with_args(const char *name,
                                 const char *command,
                                 int nargs, Oid *argtypes,
                                 Datum *values, const char *nulls,
                                 bool read_only, int cursorOptions)
```

描述

`SPI_cursor_open_with_args`建立一个将执行指定查询的游标（在内部是一个 portal）。大部分参数具有和 `SPI_prepare_cursor` 和`SPI_cursor_open`中相应参数相同的含义。

对于一次性查询执行，应优先使用此函数，而不是先调用 `SPI_prepare_cursor` 再调用 `SPI_cursor_open`。如果同一条命令要用许多不同参数执行，哪种方法更快取决于重新规划的代价与定制计划收益之间的权衡。

传入的参数数据会被复制到该游标的 portal 中，因此即使该游标仍然存在，也可以释放这些参数数据。

参数

`const char * name`

portal 的名称，或为`NULL`以让系统选择名称

`const char * command`

命令字符串

`int nargs`

输入参数的数量（`$1`、`$2` 等）

`Oid * argtypes`

一个数组，包含各参数数据类型的 OID

`Datum * values`

一个实际参数值的数组

`const char * nulls`

一个描述哪些参数为空值的数组

如果 *nulls* 为 `NULL`，则 `SPI_cursor_open_with_args` 假定没有参数为空值。

`bool read_only`

`true` 表示只读执行

`int` *`cursorOptions`*

整数形式的游标选项位掩码，零会导致默认行为

返回值

指向包含该游标的 portal 的指针。注意这里没有错误返回约定，任何错误都将通过`elog`报告。

SPI_cursor_open_with_paramlist

SPI_cursor_open_with_paramlist — set up a cursor using parameters

大纲

```
Portal SPI_cursor_open_with_paramlist(const char *name,
                                      SPIPlanPtr plan,
                                      ParamListInfo params,
                                      bool read_only)
```

描述

`SPI_cursor_open_with_paramlist` 建立一个游标（在内部是一个 portal），该游标将执行由 `SPI_prepare` 准备好的计划。该函数等价于 `SPI_cursor_open`，只是向查询传递参数值信息的形式不同。`ParamListInfo` 表示形式便于传递已经按这种格式存在的值，也支持通过 `ParamListInfo` 中指定的钩子函数使用动态参数集。

传入的参数数据会被复制到该游标的 portal 中，因此即使该游标仍然存在，也可以释放这些参数数据。

参数

`const char * name`

portal 的名称，或为`NULL`以让系统选择名称

`SPIPlanPtr plan`

执行计划（由 `SPI_prepare` 返回）

`ParamListInfo params`

包含参数类型和值的数据结构；没有参数时为 NULL

`bool read_only`

`true` 表示只读执行

返回值

指向包含该游标的 portal 的指针。注意这里没有错误返回约定，任何错误都将通过`elog`报告。

SPI_cursor_find

SPI_cursor_find — find an existing cursor by name

大纲

```
Portal SPI_cursor_find(const char * name)
```

描述

`SPI_cursor_find` 按名称查找现有的 portal。这主要用于解析由其他函数以文本形式返回的游标名。

参数

`const char * name`

该 portal 的名称

返回值

带有指定名称的 portal 的指针，如果没有找到就是 `NULL`

SPI_cursor_fetch

SPI_cursor_fetch — fetch some rows from a cursor

大纲

`void SPI_cursor_fetch(Portal portal, bool forward, long count)`

描述

`SPI_cursor_fetch`从一个游标取得一些行。这等效于 SQL 命令`FETCH`的一个子集（更多功能见`SPI_scroll_cursor_fetch`）。

参数

`Portal portal`

包含该游标的 portal

`bool forward`

为真表示向前获取，为假表示向后获取

`long count`

要取得的最大行数

返回值

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute` 相同。

注解

如果该游标的计划不是用`CURSOR_OPT_SCROLL` 选项创建的，向后获取可能失败。

SPI_cursor_move

SPI_cursor_move — move a cursor

大纲

```
void SPI_cursor_move(Portal portal, bool forward, long count)
```

描述

SPI_cursor_move跳过一个游标中的一些行。这等效于 SQL 命令MOVE的一个子集（更多的功能请见SPI_scroll_cursor_move）。

参数

Portal *portal*

包含该游标的 portal

bool *forward*

为真表示前移，为假表示后移

long *count*

要移动的最大行数

注解

如果该游标的计划不是用CURSOR_OPT_SCROLL 选项创建的，向后移动可能失败。

SPI_scroll_cursor_fetch

SPI_scroll_cursor_fetch — fetch some rows from a cursor

大纲

```
void SPI_scroll_cursor_fetch(Portal portal, FetchDirection direction,
                             long count)
```

描述

`SPI_scroll_cursor_fetch`从一个游标中取出一些行。这等效于 SQL 命令`FETCH`。

参数

`Portal` *`portal`*

包含该游标的 portal

`FetchDirection` *`direction`*

`FETCH_FORWARD`、`FETCH_BACKWARD`、`FETCH_ABSOLUTE`或者 `FETCH_RELATIVE`之一

`long` *`count`*

`FETCH_FORWARD`或 `FETCH_BACKWARD`方式中要取出的行数；`FETCH_ABSOLUTE`方式中要取出的绝对行号；`FETCH_RELATIVE`方式中要取出的相对行号

返回值

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute` 相同。

注解

参数*`direction`*和 *`count`*的详细解释请见 SQL FETCH命令。

如果该游标的计划不是用`CURSOR_OPT_SCROLL` 选项创建的，除`FETCH_FORWARD`之外的方向值可能失败。

SPI_scroll_cursor_move

SPI_scroll_cursor_move — move a cursor

大纲

```
void SPI_scroll_cursor_move(Portal portal, FetchDirection direction,
                            long count)
```

描述

`SPI_scroll_cursor_move`在一个游标中跳过一定数量的行。这等效于 SQL 命令`MOVE`。

参数

`Portal portal`

包含该游标的 portal

`FetchDirection direction`

`FETCH_FORWARD`、`FETCH_BACKWARD`、`FETCH_ABSOLUTE`或者 `FETCH_RELATIVE`之一

`long count`

`FETCH_FORWARD`或者 `FETCH_BACKWARD`方式中要移动的行数；`FETCH_ABSOLUTE`方式中要移动到的绝对行号；`FETCH_RELATIVE`方式中要移动到的相对行号

返回值

成功时，`SPI_processed` 的设置方式与 `SPI_execute` 相同。`SPI_tuptable` 会被设置为`NULL`，因为本函数不会返回任何行。

注解

参数*direction*和 *count*的详细解释请见 SQL FETCH命令。

如果该游标的计划不是用`CURSOR_OPT_SCROLL` 选项创建的，除`FETCH_FORWARD`之外的方向值可能失败。

SPI_cursor_close

SPI_cursor_close — close a cursor

大纲

`void SPI_cursor_close(Portal portal)`

描述

`SPI_cursor_close`关闭一个之前创建的游标并且释放它的 portal 存储。

所有打开的游标会在事务结束时自动被关闭。只有在希望尽快释放资源时，才需要调用 `SPI_cursor_close`。

参数

`Portal portal`

包含该游标的 portal

SPI_saveplan

SPI_saveplan — save a plan

大纲

`SPIPlanPtr SPI_saveplan(SPIPlanPtr plan)`

描述

`SPI_saveplan` 会把传入的计划（由 `SPI_prepare` 准备）保存到不会被 `SPI_finish` 或事务管理器释放的内存中，并返回指向所保存计划的指针。这让你能够在当前会话后续的过程调用中重用预备计划。

参数

`SPIPlanPtr plan`

要保存的计划

返回值

指向所保存计划的指针；如果不成功则为 `NULL`。出错时，`SPI_result` 会被设置为：

`SPI_ERROR_ARGUMENT`

如果 *plan* 为 `NULL` 或无效

`SPI_ERROR_UNCONNECTED`

如果从未连接的过程中调用

注解

原始传入的计划不会被释放，因此你可能需要对其调用 `SPI_freeplan`，以避免在 `SPI_finish` 之前发生内存泄露。

如果预备计划所引用的对象（表、函数等）之一被删除或重新定义，那么 `SPI_execute_plan` 未来的执行可能会失败，或者返回与该计划最初指示不同的结果。

43.2. 接口支持函数

这里介绍的函数提供了一个接口，用于从 `SPI_execute` 及其他 SPI 函数返回的结果集中提取信息。

本节介绍的所有函数都可以在已连接和未连接的过程中使用。

SPI_fname

SPI_fname — determine the column name for the specified column number

大纲

```
char * SPI_fname(TupleDesc rowdesc, int colnumber)
```

描述

`SPI_fname` 返回指定列列名的一个副本。（不再需要时，可以用 `pfree` 释放它。）

参数

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

返回值

列名；如果 *`colnumber`* 超出范围则返回 `NULL`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_fnumber

SPI_fnumber — determine the column number for the specified column name

大纲

```
int SPI_fnumber(TupleDesc rowdesc, const char * colname)
```

描述

`SPI_fnumber` 返回指定列名对应的列号。

如果 *colname* 指向的是系统列（例如 `oid`），则会返回相应的负列号。调用者应当通过检查返回值是否恰好等于 `SPI_ERROR_NOATTRIBUTE` 来判断错误；除非本就要拒绝系统列，否则测试结果是否小于等于 0 并不正确。

参数

`TupleDesc` *rowdesc*

输入行描述

`const char *` *colname*

列名

返回值

列号（从 1 开始计），如果没有找到所提到的列名则返回 `SPI_ERROR_NOATTRIBUTE`。

SPI_getvalue

SPI_getvalue — return the string value of the specified column

大纲

```
char * SPI_getvalue(HeapTuple row, TupleDesc rowdesc, int colnumber)
```

描述

SPI_getvalue返回指定列的值的字符串表示。

结果保存在由 palloc 分配的内存中返回。（不再需要时，可以用 pfree 释放该内存。）

参数

HeapTuple *row*

要检查的输入行

TupleDesc *rowdesc*

输入行描述

int *colnumber*

列号（从 1 开始计）

返回值

列值；如果该列为空值、*colnumber* 超出范围（SPI_result 被设置为 SPI_ERROR_NOATTRIBUTE）或者没有输出函数可用（SPI_result 被设置为 SPI_ERROR_NOOUTFUNC）则返回 NULL。

SPI_getbinval

SPI_getbinval — return the binary value of the specified column

大纲

```
Datum SPI_getbinval(HeapTuple row, TupleDesc rowdesc, int colnumber,
                    bool * isnull)
```

描述

`SPI_getbinval` 以内部形式（即 `Datum` 类型）返回指定列的值。

此函数不会为该 datum 分配新空间。对于传引用的数据类型，返回值将是指向传入行内部数据的指针。

参数

`HeapTuple` *`row`*

要检查的输入行

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

`bool *` *`isnull`*

列中是否为空值的标志

返回值

返回该列的二进制值。如果该列为空值，则 *`isnull`* 指向的变量会被设为真，否则设为假。

出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_gettype

SPI_gettype — return the data type name of the specified column

大纲

```
char * SPI_gettype(TupleDesc rowdesc, int colnumber)
```

描述

`SPI_gettype`返回该指定列的数据类型名称的副本（当你不再需要该副本后，可以使用`pfree`释放它）。

参数

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

返回值

指定列的数据类型名称；出错时返回 `NULL`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_gettypeid

SPI_gettypeid — return the data type OID of the specified column

大纲

`Oid SPI_gettypeid(TupleDesc rowdesc, int colnumber)`

描述

`SPI_gettypeid`返回该指定列的数据类型的 OID。

参数

`TupleDesc rowdesc`

输入行描述

`int colnumber`

列号（从 1 开始计）

返回值

指定列数据类型的 OID；出错时返回 `InvalidOid`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_getrelname

SPI_getrelname — return the name of the specified relation

大纲

```
char * SPI_getrelname(Relation rel)
```

描述

`SPI_getrelname`返回该指定关系的名称的副本（当你不再需要该副本后，可以使用`pfree` 释放它）。

参数

`Relation rel`

输入关系

返回值

指定关系的名称。

SPI_getnspname

SPI_getnspname — return the namespace of the specified relation

大纲

```
char * SPI_getnspname(Relation rel)
```

描述

SPI_getnspname 返回指定 Relation 所属名字空间名称的一个副本。这等价于该关系的模式名。用完该返回值后，应调用 pfree 释放它。

参数

Relation *rel*

输入关系

返回值

指定关系的名字空间的名称。

43.3. 内存管理

PostgreSQL 在内存上下文中分配内存。内存上下文为管理那些在许多不同位置创建、且生命周期各不相同的内存分配提供了便捷手段。销毁某个上下文时，会释放其中分配的全部内存。因此，不必为了避免内存泄漏而跟踪每个单独对象；只需管理数量相对较少的上下文即可。palloc 及相关函数都从“当前”上下文中分配内存。

SPI_connect 创建新的内存上下文并将其设为当前上下文。SPI_finish 则恢复先前的当前上下文，并销毁由 SPI_connect 创建的上下文。这些操作可以确保在过程内部所做的临时内存分配会在过程退出时被回收，从而避免内存泄漏。

不过，如果过程需要返回位于已分配内存中的对象（例如传引用数据类型的值），就不能使用 palloc 来分配这块内存，至少在连接到 SPI 的期间不能这么做。否则，该对象会在 SPI_finish 时被释放，过程也就无法可靠工作。解决办法是用 SPI_palloc 为返回对象分配内存。SPI_palloc 在“上层执行器上下文”中分配内存，也就是调用 SPI_connect 时的当前内存上下文；这正是从过程返回值最合适的上下文。

如果在过程尚未连接到 SPI 时调用SPI_palloc，它的行为与普通的palloc相同。在过程连接到 SPI 管理器之前，当前内存上下文就是上层执行器上下文，因此过程通过palloc或 SPI 辅助函数所做的所有分配都位于这个上下文中。

调用 SPI_connect 时，会把该过程的私有上下文（由 SPI_connect 创建）设为当前上下文。所有通过 palloc、repalloc 或 SPI 辅助函数分配的内存（但SPI_copytuple、SPI_returntuple、SPI_modifytuple 和SPI_palloc除外）都位于这个上下文中。当过程通过 SPI_finish 与 SPI 管理器断开连接时，当前上下文会恢复为上层执行器上下文，而在该过程内存上下文中分配的所有内存都会被释放，之后就不能再使用。

本节介绍的所有函数既可由已连接的过程使用，也可由未连接的过程使用。在未连接的过程中，它们的行为与底层普通服务器函数（palloc等）相同。

SPI_palloc

SPI_palloc — allocate memory in the upper executor context

大纲

```
void * SPI_palloc(Size size)
```

描述

`SPI_palloc` 在上层执行器上下文中分配内存。

参数

`Size size`

要分配的存储空间大小（以字节计）

返回值

指向具有指定大小的新存储空间的指针

SPI_repalloc

SPI_repalloc — reallocate memory in the upper executor context

大纲

```
void * SPI_repalloc(void * pointer, Size size)
```

描述

`SPI_repalloc`改变之前用`SPI_palloc` 分配的内存段的大小。

这个函数如今与普通的`repalloc`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`void * pointer`

指向要改变的现有存储空间的指针

`Size size`

要分配的存储空间大小（以字节计）

返回值

指向具有指定大小的新存储空间的指针，现有区域的内容会被复制到其中

SPI_pfree

SPI_pfree — free memory in the upper executor context

大纲

```
void SPI_pfree(void * pointer)
```

描述

`SPI_pfree`释放之前使用 `SPI_palloc`或者 `SPI_repalloc`分配的内存。

这个函数如今与普通的`pfree`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`void * pointer`

指向要释放的现有存储空间的指针

SPI_copytuple

SPI_copytuple — make a copy of a row in the upper executor context

大纲

`HeapTuple SPI_copytuple(HeapTuple row)`

描述

`SPI_copytuple` 在上层执行器上下文中创建一行的副本。它通常用于从触发器中返回修改后的行。对于声明为返回复合类型的函数，应改用 `SPI_returntuple`。

参数

`HeapTuple row`

要复制的行

返回值

复制后的行；仅当*tuple*为 `NULL` 时才返回 `NULL`

SPI_returntuple

SPI_returntuple — prepare to return a tuple as a Datum

大纲

`HeapTupleHeader SPI_returntuple(HeapTuple row, TupleDesc rowdesc)`

描述

`SPI_returntuple` 在上层执行器上下文中创建一行的副本，并以行类型 `Datum` 的形式返回。返回的指针只需在返回前用 `PointerGetDatum` 转换为 `Datum` 即可。

请注意，它应用于声明为返回复合类型的函数，而不用于触发器；
触发器中返回修改后的行应使用 `SPI_copytuple`。

参数

`HeapTuple row`

要复制的行

`TupleDesc rowdesc`

行描述符（若要获得最佳缓存效果，每次都传入同一个描述符）

返回值

指向复制后行的 `HeapTupleHeader`；仅当*row* 或 *rowdesc*为 `NULL` 时才返回 `NULL`

SPI_modifytuple

SPI_modifytuple — create a row by replacing selected fields of a given row

大纲

```
HeapTuple SPI_modifytuple(Relation rel, HeapTuple row, int ncols,
                          int * colnum, Datum * values, const char
 * nulls)
```

描述

`SPI_modifytuple` 通过为选定列替换新值、并在其他位置复制原行的列来创建一行新记录。输入行本身不会被修改。新行会在上层执行器上下文中返回。

参数

`Relation rel`

仅用作该行的行描述符来源。（传递关系而不是行描述符是个设计缺陷。）

`HeapTuple row`

要被修改的行

`int ncols`

数组 *colnum* 中列号的数量

`int * colnum`

一个数组，包含要被修改的列的编号（列号从 1 开始）

`Datum * values`

指定列的新值

`const char * Nulls`

哪些新值为空值（如果有的话；格式参见 `SPI_execute_plan`）

返回值

修改后的新行，分配在上层执行器上下文中；仅当*row* 为 `NULL` 时才返回 `NULL`

出错时，`SPI_result`被设置如下：

`SPI_ERROR_ARGUMENT`

如果*rel*为`NULL`，或者 *row*为`NULL`，或者*ncols* 小于等于 0，或者*colnum*为`NULL`，或者*values*为`NULL`。

`SPI_ERROR_NOATTRIBUTE`

如果*colnum*包含一个无效的列号（小于等于 0 或者大于 *row*中的列数）。

SPI_freetuple

SPI_freetuple — free a row allocated in the upper executor context

大纲

```
void SPI_freetuple(HeapTuple row)
```

描述

`SPI_freetuple`释放之前在上层执行器上下文中分配的一行。

这个函数如今与普通的`heap_freetuple`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`HeapTuple` *`row`*

要释放的行

SPI_freetuptable

SPI_freetuptable — free a row set created by `SPI_execute` or a similar function

大纲

```
void SPI_freetuptable(SPITupleTable * tuptable)
```

描述

`SPI_freetuptable` 释放由先前的 SPI 命令执行函数（例如 `SPI_execute`）创建的结果行集。因此，调用它时经常会直接把全局变量 `SPI_tupletable` 作为参数。

如果某个使用 SPI 的过程需要执行多条命令，并且不想将前面命令的结果一直保留到过程结束，那么这个函数就很有用。注意，任何尚未释放的结果行集都会在 `SPI_finish` 时被释放。

参数

`SPITupleTable * tuptable`

要释放的行集的指针

SPI_freeplan

SPI_freeplan — free a previously saved plan

大纲

`int SPI_freeplan(SPIPlanPtr plan)`

描述

`SPI_freeplan` 释放先前由 `SPI_prepare` 返回或由 `SPI_saveplan` 保存的命令执行计划。

参数

`SPIPlanPtr plan`

要释放的计划的指针

返回值

如果 *plan* 为 `NULL` 或无效，则返回 `SPI_ERROR_ARGUMENT`

43.4. 数据更改的可见性

下列规则决定了使用 SPI 的函数（以及其他任何 C 函数）中数据更改的可见性：

- 在某条 SQL 命令执行期间，该命令所做的任何数据更改对该命令本身都是不可见的。例如，在

  ```
  INSERT INTO a SELECT * FROM a;
  ```

  中，被插入的行对`SELECT`部分不可见。

- 某条命令 C 所做的更改，对所有在 C 之后启动的命令都可见，无论这些命令是在 C 内部（即 C 执行期间）启动，还是在 C 完成之后启动。

- 在由 SQL 命令调用的函数中（无论是普通函数还是触发器）通过 SPI 执行的命令，究竟遵循上述哪条规则，取决于传给 SPI 的读写标志。只读模式下执行的命令遵循第一条规则：它们看不到调用它们的命令所做的更改。
  读写模式下执行的命令遵循第二条规则：它们能够看到截至当前为止的所有更改。

- 所有标准过程语言都会根据函数的易变性属性设置 SPI 的读写模式。`STABLE` 和 `IMMUTABLE` 函数中的命令会以只读模式执行，而 `VOLATILE` 函数中的命令则以读写模式执行。虽然 C 函数作者能够违背这一约定，但通常并不是好主意。

下一节给出了一个展示这些规则如何应用的示例。

43.5. 示例

本节给出一个非常简单的 SPI 用法示例。过程 `execq` 以一条 SQL 命令作为第一个参数、以行数计数作为第二个参数，使用 `SPI_exec` 执行该命令，并返回该命令实际处理的行数。更复杂的 SPI 示例可以在源码树的 `src/test/regress/regress.c` 以及spi模块中找到。

```
#include "postgres.h"

#include "executor/spi.h"
#include "utils/builtins.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

int execq(text *sql, int cnt);

int
execq(text *sql, int cnt)
{
    char *command;
    int ret;
    int proc;

    /* Convert given text object to a C string */
    command = text_to_cstring(sql);

    SPI_connect();

    ret = SPI_exec(command, cnt);

    proc = SPI_processed;
    /*
     * If some rows were fetched, print them via elog(INFO).
     */
    if (ret > 0 && SPI_tuptable != NULL)
    {
        TupleDesc tupdesc = SPI_tuptable->tupdesc;
        SPITupleTable *tuptable = SPI_tuptable;
        char buf[8192];
        int i, j;

        for (j = 0; j < proc; j++)
        {
            HeapTuple tuple = tuptable->vals[j];

            for (i = 1, buf[0] = 0; i <= tupdesc->natts; i++)
                snprintf(buf + strlen (buf), sizeof(buf) - strlen(buf)
, " %s%s",
                         SPI_getvalue(tuple, tupdesc, i),
                         (i == tupdesc->natts) ? " " : " |");
            elog(INFO, "EXECQ: %s", buf);
        }
    }

    SPI_finish();
    pfree(command);

    return (proc);
}
```

（此函数使用版本 0 调用约定，以使示例更易于理解。在实际应用中，你应该使用新的版本 1 接口。）

将该函数编译成共享库之后（详见第 35.9.6 节），可以这样声明它：

```
CREATE FUNCTION execq(text, integer) RETURNS integer
    AS 'filename'
    LANGUAGE C;
```

以下是一个示例会话：

```
=> SELECT execq('CREATE TABLE a (x integer)', 0);
 execq
-------
     0
(1 row)

=> INSERT INTO a VALUES (execq('INSERT INTO a VALUES (0)', 0));
INSERT 0 1
=> SELECT execq('SELECT * FROM a', 0);
INFO:  EXECQ:  0    -- 由 execq 插入
INFO:  EXECQ:  1    -- 由 execq 返回并被外层 INSERT 插入

 execq
-------
     2
(1 row)

=> SELECT execq('INSERT INTO a SELECT x + 2 FROM a', 1);
 execq
-------
     1
(1 row)

=> SELECT execq('SELECT * FROM a', 10);
INFO:  EXECQ:  0
INFO:  EXECQ:  1
INFO:  EXECQ:  2    -- 0 + 2，按指定只插入一行

 execq
-------
     3              -- 10 只是最大值，3 才是真正的行数
(1 row)

=> DELETE FROM a;
DELETE 3
=> INSERT INTO a VALUES (execq('SELECT * FROM a', 0) + 1);
INSERT 0 1
=> SELECT * FROM a;
 x
---
 1                  -- a 中没有行 (0) + 1
(1 row)

=> INSERT INTO a VALUES (execq('SELECT * FROM a', 0) + 1);
```

```
INFO:  EXECQ:  1
INSERT 0 1
=> SELECT * FROM a;
 x
---
 1
 2                  -- a 中原有一行 + 1
(2 rows)

-- 下面演示数据更改可见性规则:

=> INSERT INTO a SELECT execq('SELECT * FROM a', 0) * x FROM a;
INFO:  EXECQ:  1
INFO:  EXECQ:  2
INFO:  EXECQ:  1
INFO:  EXECQ:  2
INFO:  EXECQ:  2
INSERT 0 2
=> SELECT * FROM a;
 x
---
 1
 2
 2                  -- 2 行 * 1 (第一行中的 x)
 6                  -- 3 行 (前面的 2 行加上刚插入的 1 行) * 2 (第二行中的
 x)
(4 rows)               ^^^^^^
                       不同调用中 execq() 可见的行
```

部分 VI. 参考

本参考中的各条目旨在以适当的篇幅，对各自的主题给出权威、完整且正式的总结。关于如何使用PostgreSQL的更多信息，例如叙述性说明、教程或示例，可在本书的其他部分找到。
另见各参考页列出的交叉引用。

这些参考条目也提供传统的“man”手册页形式。

目录

I. SQL 命令 1025
ABORT 1028
ALTER AGGREGATE 1029
ALTER COLLATION 1031
ALTER CONVERSION 1032
ALTER DATABASE 1033
ALTER DEFAULT PRIVILEGES 1035
ALTER DOMAIN 1038
ALTER EXTENSION 1041
ALTER FOREIGN DATA WRAPPER 1044
ALTER FOREIGN TABLE 1046
ALTER FUNCTION 1049
ALTER GROUP 1052
ALTER INDEX 1053
ALTER LANGUAGE 1055
ALTER LARGE OBJECT 1056
ALTER OPERATOR 1057
ALTER OPERATOR CLASS 1059
ALTER OPERATOR FAMILY 1060
ALTER ROLE 1064
ALTER SCHEMA 1067
ALTER SEQUENCE 1068
ALTER SERVER 1071
ALTER TABLE 1072
ALTER TABLESPACE 1082
ALTER TEXT SEARCH CONFIGURATION 1084
ALTER TEXT SEARCH DICTIONARY 1086
ALTER TEXT SEARCH PARSER 1088
ALTER TEXT SEARCH TEMPLATE 1089
ALTER TRIGGER 1090
ALTER TYPE 1091
ALTER USER 1094
ALTER USER MAPPING 1095
ALTER VIEW 1096
ANALYZE 1098
BEGIN 1100
CHECKPOINT 1102
CLOSE 1103
CLUSTER 1104
COMMENT 1106
COMMIT 1110
COMMIT PREPARED 1111
COPY 1112
CREATE AGGREGATE 1121
CREATE CAST 1124
CREATE COLLATION 1128
CREATE CONVERSION 1130
CREATE DATABASE 1132
CREATE DOMAIN 1135
CREATE EXTENSION 1137
CREATE FOREIGN DATA WRAPPER 1139

CREATE FOREIGN TABLE 1141
CREATE FUNCTION 1143
CREATE GROUP 1150
CREATE INDEX 1151
CREATE LANGUAGE 1157
CREATE OPERATOR 1160
CREATE OPERATOR CLASS 1163
CREATE OPERATOR FAMILY 1166
CREATE ROLE 1167
CREATE RULE 1172
CREATE SCHEMA 1175
CREATE SEQUENCE 1177
CREATE SERVER 1180
CREATE TABLE 1182
CREATE TABLE AS 1195
CREATE TABLESPACE 1198
CREATE TEXT SEARCH CONFIGURATION 1200
CREATE TEXT SEARCH DICTIONARY 1201
CREATE TEXT SEARCH PARSER 1203
CREATE TEXT SEARCH TEMPLATE 1205
CREATE TRIGGER 1206
CREATE TYPE 1211
CREATE USER 1218
CREATE USER MAPPING 1219
CREATE VIEW 1220
DEALLOCATE 1223
DECLARE 1224
DELETE 1227
DISCARD 1230
DO 1231
DROP AGGREGATE 1233
DROP CAST 1234
DROP COLLATION 1235
DROP CONVERSION 1236
DROP DATABASE 1237
DROP DOMAIN 1238
DROP EXTENSION 1239
DROP FOREIGN DATA WRAPPER 1240
DROP FOREIGN TABLE 1241
DROP FUNCTION 1242
DROP GROUP 1244
DROP INDEX 1245
DROP LANGUAGE 1246
DROP OPERATOR 1247
DROP OPERATOR CLASS 1249
DROP OPERATOR FAMILY 1251
DROP OWNED 1253
DROP ROLE 1254
DROP RULE 1255
DROP SCHEMA 1256
DROP SEQUENCE 1257
DROP SERVER 1258
DROP TABLE 1259
DROP TABLESPACE 1260

DROP TEXT SEARCH CONFIGURATION 1261
DROP TEXT SEARCH DICTIONARY 1262
DROP TEXT SEARCH PARSER 1263
DROP TEXT SEARCH TEMPLATE 1264
DROP TRIGGER 1265
DROP TYPE 1266
DROP USER 1267
DROP USER MAPPING 1268
DROP VIEW 1269
END 1270
EXECUTE 1271
EXPLAIN 1272
FETCH 1277
GRANT 1281
INSERT 1287
LISTEN 1290
LOAD 1292
LOCK 1293
MOVE 1296
NOTIFY 1298
PREPARE 1300
PREPARE TRANSACTION 1302
REASSIGN OWNED 1304
REINDEX 1305
RELEASE SAVEPOINT 1308
RESET 1309
REVOKE 1310
ROLLBACK 1314
ROLLBACK PREPARED 1315
ROLLBACK TO SAVEPOINT 1316
SAVEPOINT 1318
SECURITY LABEL 1320
SELECT 1322
SELECT INTO 1338
SET 1340
SET CONSTRAINTS 1343
SET ROLE 1344
SET SESSION AUTHORIZATION 1346
SET TRANSACTION 1348
SHOW 1350
START TRANSACTION 1352
TRUNCATE 1353
UNLISTEN 1355
UPDATE 1357
VACUUM 1361
VALUES 1364
II. PostgreSQL 客户端应用 1367
clusterdb 1368
createdb 1371
createlang 1374
createuser 1377
dropdb 1381
droplang 1383
dropuser 1386

ecpg 1389
pg_basebackup 1391
pg_config 1395
pg_dump 1398
pg_dumpall 1407
pg_restore 1412
psql 1419
reindexdb 1448
vacuumdb 1451
III. PostgreSQL 服务器端应用 1454
initdb 1455
pg_controldata 1458
pg_ctl 1459
pg_resetxlog 1464
postgres 1466
postmaster 1473

SQL 命令

本部分包含PostgreSQL支持的SQL命令的参考信息。这里的 “SQL” 指的是一般意义上的该语言；每条命令的标准符合性和兼容性信息可在相应的参考页中找到。

目录

ABORT 1028
ALTER AGGREGATE 1029
ALTER COLLATION 1031
ALTER CONVERSION 1032
ALTER DATABASE 1033
ALTER DEFAULT PRIVILEGES 1035
ALTER DOMAIN 1038
ALTER EXTENSION 1041
ALTER FOREIGN DATA WRAPPER 1044
ALTER FOREIGN TABLE 1046
ALTER FUNCTION 1049
ALTER GROUP 1052
ALTER INDEX 1053
ALTER LANGUAGE 1055
ALTER LARGE OBJECT 1056
ALTER OPERATOR 1057
ALTER OPERATOR CLASS 1059
ALTER OPERATOR FAMILY 1060
ALTER ROLE 1064
ALTER SCHEMA 1067
ALTER SEQUENCE 1068
ALTER SERVER 1071
ALTER TABLE 1072
ALTER TABLESPACE 1082
ALTER TEXT SEARCH CONFIGURATION 1084
ALTER TEXT SEARCH DICTIONARY 1086
ALTER TEXT SEARCH PARSER 1088
ALTER TEXT SEARCH TEMPLATE 1089
ALTER TRIGGER 1090
ALTER TYPE 1091
ALTER USER 1094
ALTER USER MAPPING 1095
ALTER VIEW 1096
ANALYZE 1098
BEGIN 1100
CHECKPOINT 1102
CLOSE 1103
CLUSTER 1104
COMMENT 1106
COMMIT 1110
COMMIT PREPARED 1111
COPY 1112
CREATE AGGREGATE 1121
CREATE CAST 1124

CREATE COLLATION 1128
CREATE CONVERSION 1130
CREATE DATABASE 1132
CREATE DOMAIN 1135
CREATE EXTENSION 1137
CREATE FOREIGN DATA WRAPPER 1139
CREATE FOREIGN TABLE 1141
CREATE FUNCTION 1143
CREATE GROUP 1150
CREATE INDEX 1151
CREATE LANGUAGE 1157
CREATE OPERATOR 1160
CREATE OPERATOR CLASS 1163
CREATE OPERATOR FAMILY 1166
CREATE ROLE 1167
CREATE RULE 1172
CREATE SCHEMA 1175
CREATE SEQUENCE 1177
CREATE SERVER 1180
CREATE TABLE 1182
CREATE TABLE AS 1195
CREATE TABLESPACE 1198
CREATE TEXT SEARCH CONFIGURATION 1200
CREATE TEXT SEARCH DICTIONARY 1201
CREATE TEXT SEARCH PARSER 1203
CREATE TEXT SEARCH TEMPLATE 1205
CREATE TRIGGER 1206
CREATE TYPE 1211
CREATE USER 1218
CREATE USER MAPPING 1219
CREATE VIEW 1220
DEALLOCATE 1223
DECLARE 1224
DELETE 1227
DISCARD 1230
DO 1231
DROP AGGREGATE 1233
DROP CAST 1234
DROP COLLATION 1235
DROP CONVERSION 1236
DROP DATABASE 1237
DROP DOMAIN 1238
DROP EXTENSION 1239
DROP FOREIGN DATA WRAPPER 1240
DROP FOREIGN TABLE 1241
DROP FUNCTION 1242
DROP GROUP 1244
DROP INDEX 1245
DROP LANGUAGE 1246
DROP OPERATOR 1247
DROP OPERATOR CLASS 1249
DROP OPERATOR FAMILY 1251
DROP OWNED 1253
DROP ROLE 1254

DROP RULE 1255
DROP SCHEMA 1256
DROP SEQUENCE 1257
DROP SERVER 1258
DROP TABLE 1259
DROP TABLESPACE 1260
DROP TEXT SEARCH CONFIGURATION 1261
DROP TEXT SEARCH DICTIONARY 1262
DROP TEXT SEARCH PARSER 1263
DROP TEXT SEARCH TEMPLATE 1264
DROP TRIGGER 1265
DROP TYPE 1266
DROP USER 1267
DROP USER MAPPING 1268
DROP VIEW 1269
END 1270
EXECUTE 1271
EXPLAIN 1272
FETCH 1277
GRANT 1281
INSERT 1287
LISTEN 1290
LOAD 1292
LOCK 1293
MOVE 1296
NOTIFY 1298
PREPARE 1300
PREPARE TRANSACTION 1302
REASSIGN OWNED 1304
REINDEX 1305
RELEASE SAVEPOINT 1308
RESET 1309
REVOKE 1310
ROLLBACK 1314
ROLLBACK PREPARED 1315
ROLLBACK TO SAVEPOINT 1316
SAVEPOINT 1318
SECURITY LABEL 1320
SELECT 1322
SELECT INTO 1338
SET 1340
SET CONSTRAINTS 1343
SET ROLE 1344
SET SESSION AUTHORIZATION 1346
SET TRANSACTION 1348
SHOW 1350
START TRANSACTION 1352
TRUNCATE 1353
UNLISTEN 1355
UPDATE 1357
VACUUM 1361
VALUES 1364

ABORT

ABORT — abort the current transaction

大纲

```
ABORT [ WORK | TRANSACTION ]
```

描述

`ABORT`回滚当前事务，并丢弃该事务所做的全部更新。这个命令在行为上与标准 SQL 命令ROLLBACK完全相同，只是出于历史原因而保留。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用COMMIT可成功地终止一个事务。

在事务块之外发出`ABORT`会产生一条警告，除此之外没有效果。

示例

要中止所有更改：

```
ABORT;
```

兼容性

这个命令是PostgreSQL扩展，只是出于历史原因而保留。`ROLLBACK`是与之等效的标准 SQL 命令。

另见

BEGIN, COMMIT, ROLLBACK

ALTER AGGREGATE

ALTER AGGREGATE — change the definition of an aggregate function

大纲

```
ALTER AGGREGATE name ( type [ , ... ] ) RENAME TO new_name
ALTER AGGREGATE name ( type [ , ... ] ) OWNER TO new_owner
ALTER AGGREGATE name ( type [ , ... ] ) SET SCHEMA new_schema
```

描述

`ALTER AGGREGATE`更改聚合函数的定义。

要使用`ALTER AGGREGATE`，必须拥有该聚合函数。要更改聚合函数的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在聚合函数所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建聚合函数完成的操作。不过，超级用户无论如何都可以更改任何聚合函数的所有权。）

参数

name

现有聚合函数的名称（可以用模式限定）。

type

聚合函数作用于其上的输入数据类型。要引用零参数聚合函数，
请在参数说明列表的位置写上`*`。

new_name

聚合函数的新名称。

new_owner

聚合函数的新拥有者。

new_schema

聚合函数的新模式。

示例

要把用于类型`integer`的聚合函数 `myavg`重命名为`my_average`：

```
ALTER AGGREGATE myavg(integer) RENAME TO my_average;
```

要把用于类型`integer`的聚合函数 `myavg`的拥有者改为`joe`：

```
ALTER AGGREGATE myavg(integer) OWNER TO joe;
```

要把用于`float8`的聚合`myavg`移动到模式`myschema`中：

```
ALTER AGGREGATE myavg(float8) SET SCHEMA myschema;
```

兼容性

在 SQL 标准中没有`ALTER AGGREGATE`语句。

另见

CREATE AGGREGATE, DROP AGGREGATE

ALTER COLLATION

ALTER COLLATION — change the definition of a collation

大纲

```
ALTER COLLATION name RENAME TO new_name
ALTER COLLATION name OWNER TO new_owner
ALTER COLLATION name SET SCHEMA new_schema
```

描述

`ALTER COLLATION`更改排序规则的定义。

要使用`ALTER COLLATION`，必须拥有该排序规则。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在排序规则所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建排序规则完成的操作。不过，超级用户无论如何都可以更改任何排序规则的所有权。）

参数

name

一个现有排序规则的名称（可以是模式限定的）。

new_name

排序规则的新名称。

new_owner

排序规则的新拥有者。

new_schema

排序规则的新模式。

示例

要将排序规则`de_DE`重命名为`german`：

```
ALTER COLLATION "de_DE" RENAME TO german;
```

要将排序规则`en_US`的拥有者改为`joe`：

```
ALTER COLLATION "en_US" OWNER TO joe;
```

兼容性

SQL 标准中没有`ALTER COLLATION`语句。

参见

CREATE COLLATION, DROP COLLATION

ALTER CONVERSION

ALTER CONVERSION — change the definition of a conversion

大纲

```
ALTER CONVERSION name RENAME TO new_name
ALTER CONVERSION name OWNER TO new_owner
ALTER CONVERSION name SET SCHEMA new_schema
```

描述

`ALTER CONVERSION`更改一个转换的定义。

要使用`ALTER CONVERSION`，必须拥有该转换。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在转换所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建转换完成的操作。不过，超级用户无论如何都可以更改任何转换的所有权。）

参数

name

一个现有转换的名称（可以是模式限定的）。

new_name

转换的新名称。

new_owner

转换的新拥有者。

new_schema

转换的新模式。

示例

要把转换`iso_8859_1_to_utf8`重命名为`latin1_to_unicode`：

```
ALTER CONVERSION iso_8859_1_to_utf8 RENAME TO latin1_to_unicode;
```

要把转换`iso_8859_1_to_utf8`的拥有者改成`joe`：

```
ALTER CONVERSION iso_8859_1_to_utf8 OWNER TO joe;
```

兼容性

在 SQL 标准中没有`ALTER CONVERSION`语句。

另见

CREATE CONVERSION, DROP CONVERSION

ALTER DATABASE

ALTER DATABASE — change a database

大纲

```
ALTER DATABASE name [ [ WITH ] option [ ... ] ]

where option can be:

    CONNECTION LIMIT connlimit

ALTER DATABASE name RENAME TO new_name

ALTER DATABASE name OWNER TO new_owner

ALTER DATABASE name SET TABLESPACE new_tablespace

ALTER DATABASE name SET configuration_parameter { TO | = } { value |
 DEFAULT }
ALTER DATABASE name SET configuration_parameter FROM CURRENT
ALTER DATABASE name RESET configuration_parameter
ALTER DATABASE name RESET ALL
```

描述

`ALTER DATABASE`更改一个数据库的属性。

第一种形式更改某些数据库级设置（详见下文）。
只有数据库拥有者或者超级用户才能更改这些设置。

第二种形式更改数据库的名称。只有数据库拥有者或者超级用户可以重命名数据库；
非超级用户拥有者还必须具有`CREATEDB`权限。
当前连接的数据库不能被重命名（如果需要这样做，请连接到另一个数据库）。

第三种形式更改数据库的所有者。要更改所有者，必须拥有该数据库，
并且还必须是新所有者角色的直接或间接成员，并且必须拥有`CREATEDB`权限。（请注意，
超级用户会自动拥有所有这些权限。）

第四种形式更改数据库的默认表空间。只有数据库拥有者或者超级用户可以这样做；
你还必须对新表空间具有创建权限。
该命令会将数据库旧默认表空间中的所有表和索引在物理上移动到新表空间中。
对于该数据库而言，新默认表空间必须为空，并且不能有人连接到该数据库。
位于非默认表空间中的表和索引不受影响。

其余形式会更改某个PostgreSQL数据库的运行时配置变量的会话默认值。
此后每当在该数据库中启动一个新会话时，指定的值就会成为会话默认值。
数据库特定的默认值会覆盖`postgresql.conf`中的设置，或者从`postgres`
命令行接收到的设置。只有数据库拥有者或者超级用户才能更改该数据库的会话默认值。
某些变量不能以这种方式设置，或者只能由超级用户设置。

参数

name

要修改其属性的数据库名称。

connlimit

这个数据库允许多少并发连接。-1 表示没有限制。

new_name

数据库的新名称。

new_owner

数据库的新拥有者。

new_tablespace

数据库的新默认表空间。

configuration_parameter
value

将此数据库在指定配置参数上的会话默认值设为给定值。如果 *value*是`DEFAULT`，或者等效地使用了 `RESET`，则数据库特定设置会被移除，因此新会话将继承系统范围的默认设置。使用 `RESET ALL`可清除所有数据库特定设置。`SET FROM CURRENT`会把该参数在当前会话中的值保存为数据库特定值。

关于允许的参数名和值的更多信息，见SET和第 18 章。

注解

也可以把会话默认值绑定到特定角色，而不是数据库；见ALTER ROLE。如果发生冲突，角色特定设置会覆盖数据库特定设置。

示例

要在数据库`test`中默认禁用索引扫描：

```
ALTER DATABASE test SET enable_indexscan TO off;
```

兼容性

`ALTER DATABASE`语句是一个PostgreSQL扩展。

参见

CREATE DATABASE, DROP DATABASE, SET, CREATE TABLESPACE

ALTER DEFAULT PRIVILEGES

ALTER DEFAULT PRIVILEGES — define default access privileges

大纲

```
ALTER DEFAULT PRIVILEGES
    [ FOR { ROLE | USER } target_role [, ...] ]
    [ IN SCHEMA schema_name [, ...] ]
    abbreviated_grant_or_revoke

where abbreviated_grant_or_revoke is one of:

GRANT { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES |
 TRIGGER }
    [, ...] | ALL [ PRIVILEGES ] }
    ON TABLES
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { USAGE | SELECT | UPDATE }
    [, ...] | ALL [ PRIVILEGES ] }
    ON SEQUENCES
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON FUNCTIONS
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

REVOKE [ GRANT OPTION FOR ]
    { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES |
 TRIGGER }
    [, ...] | ALL [ PRIVILEGES ] }
    ON TABLES
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { USAGE | SELECT | UPDATE }
    [, ...] | ALL [ PRIVILEGES ] }
    ON SEQUENCES
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { EXECUTE | ALL [ PRIVILEGES ] }
    ON FUNCTIONS
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]
```

描述

`ALTER DEFAULT PRIVILEGES` 允许设置将应用于将来创建的对象的权限。（它不影响已存在对象的权限。）目前，只能更改表（包括视图和外部表）、序列和函数的权限。

虽然你可以修改自己的默认权限，也可以修改你所属角色的默认权限，但在创建对象时，新对象的权限只受当前角色的默认权限影响，不会从当前角色所属的任何角色继承。

如GRANT下所述，任何对象类型的默认权限通常会向对象所有者授予所有可授予的权限，也可能向`PUBLIC`授予一些权限。不过，可以通过使用`ALTER DEFAULT PRIVILEGES`更改全局默认权限来改变这种行为。

参数

target_role

当前角色所属的现有角色的名称。如果省略`FOR ROLE`，则假定为当前角色。

schema_name

一个现有模式的名称。如果指定，将修改以后在该模式中创建的对象的默认权限。如果省略`IN SCHEMA`，则修改全局默认权限。

role_name

要为其授予或撤销权限的现有角色的名称。此参数以及 *abbreviated_grant_or_revoke* 中的所有其他参数，均按GRANT或REVOKE中的描述工作，只不过这里设置的是整类对象的权限，而不是特定的具名对象。

注解

使用psql的`\ddp`命令获取现有默认权限分配的信息。权限值的含义与GRANT下对`\dp`的说明相同。

如果你希望删除一个其默认权限已被修改的角色，则必须先撤销对其默认权限所做的更改，或者使用`DROP OWNED BY` 去除该角色的默认权限条目。

示例

为你后续在模式`myschema`中创建的所有表（以及视图）向所有人授予 SELECT 权限，并且也允许角色`webuser`对它们执行 INSERT：

```
ALTER DEFAULT PRIVILEGES IN SCHEMA myschema GRANT SELECT ON TABLES TO
 PUBLIC;
ALTER DEFAULT PRIVILEGES IN SCHEMA myschema GRANT INSERT ON TABLES TO
 webuser;
```

撤销上述操作，使后续创建的表不再拥有任何超出正常情况的权限：

```
ALTER DEFAULT PRIVILEGES IN SCHEMA myschema REVOKE SELECT ON TABLES
 FROM PUBLIC;
ALTER DEFAULT PRIVILEGES IN SCHEMA myschema REVOKE INSERT ON TABLES
 FROM webuser;
```

移除通常会在函数上授予的公共 EXECUTE 权限，使后续由角色`admin`创建的所有函数不再自动具有该权限：

```
ALTER DEFAULT PRIVILEGES FOR ROLE admin REVOKE EXECUTE ON FUNCTIONS
 FROM PUBLIC;
```

兼容性

在 SQL 标准中没有`ALTER DEFAULT PRIVILEGES`语句。

另见

GRANT, REVOKE

ALTER DOMAIN

ALTER DOMAIN — change the definition of a domain

大纲

```
ALTER DOMAIN name
    { SET DEFAULT expression | DROP DEFAULT }
ALTER DOMAIN name
    { SET | DROP } NOT NULL
ALTER DOMAIN name
    ADD domain_constraint
ALTER DOMAIN name
    DROP CONSTRAINT constraint_name [ RESTRICT | CASCADE ]
ALTER DOMAIN name
    OWNER TO new_owner
ALTER DOMAIN name
    SET SCHEMA new_schema
```

描述

`ALTER DOMAIN`更改一个现有域的定义。有以下几种子形式：

SET/DROP DEFAULT

这些形式为域设置或移除默认值。注意，默认值只适用于后续的`INSERT`命令；它们不会影响已存在于使用该域的表中的行。

SET/DROP NOT NULL

这些形式更改域是被标记为允许 NULL 值还是拒绝 NULL 值。只有当使用该域的列中不包含空值时，才能执行`SET NOT NULL`。

ADD *domain_constraint*

这种形式使用与CREATE DOMAIN相同的语法向域添加一个新约束。只有当所有使用该域的列都满足新约束时，这一操作才会成功。

DROP CONSTRAINT

这种形式删除域上的约束。

OWNER

这种形式将域的拥有者更改为指定用户。

SET SCHEMA

这种形式更改域所在的模式。与该域关联的任何约束也会一并移动到新模式中。

要使用`ALTER DOMAIN`，必须拥有该域。要更改域的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在域所在模式上拥有`CREATE`权限。

（这些限制确保更改所有者不会执行任何无法通过删除并重新创建域完成的操作。不过，超级用户无论如何都可以更改任何域的所有权。）

参数

name

要修改的现有域的名称（可以是模式限定的）。

domain_constraint

该域的新约束。

constraint_name

要删除的现有约束的名称。

CASCADE

自动删除依赖于该约束的对象。

RESTRICT

如果存在任何依赖对象，则拒绝删除该约束。这是默认行为。

new_owner

域的新拥有者的用户名。

new_schema

域的新模式。

注解

目前，如果命名域或任何派生域被数据库中任意表的复合类型列使用，ALTER DOMAIN ADD CONSTRAINT和ALTER DOMAIN SET NOT NULL将失败。最终应改进这些命令，使其能够验证此类嵌套列上的新约束。

示例

向域添加一个NOT NULL约束：

```
ALTER DOMAIN zipcode SET NOT NULL;
```

从域中移除一个NOT NULL约束：

```
ALTER DOMAIN zipcode DROP NOT NULL;
```

向域添加一个检查约束：

```
ALTER DOMAIN zipcode ADD CONSTRAINT zipchk CHECK (char_length(VALUE) =
 5);
```

从域中移除一个检查约束：

```
ALTER DOMAIN zipcode DROP CONSTRAINT zipchk;
```

将域移动到另一个模式中：

```
ALTER DOMAIN zipcode SET SCHEMA customers;
```

兼容性

`ALTER DOMAIN` 符合 SQL 标准，但 `OWNER` 和 `SET SCHEMA` 变体是 PostgreSQL 扩展。

另见

CREATE DOMAIN, DROP DOMAIN

ALTER EXTENSION

ALTER EXTENSION — change the definition of an extension

大纲

```
ALTER EXTENSION extension_name UPDATE [ TO new_version ]
ALTER EXTENSION extension_name SET SCHEMA new_schema
ALTER EXTENSION extension_name ADD member_object
ALTER EXTENSION extension_name DROP member_object

where member_object is:

  AGGREGATE agg_name (agg_type [, ...] ) |
  CAST (source_type AS target_type) |
  COLLATION object_name |
  CONVERSION object_name |
  DOMAIN object_name |
  FOREIGN DATA WRAPPER object_name |
  FOREIGN TABLE object_name |
  FUNCTION function_name ( [ [ argmode ] [ argname ] argtype [, ...] ]
 ) |
  OPERATOR operator_name (left_type, right_type) |
  OPERATOR CLASS object_name USING index_method |
  OPERATOR FAMILY object_name USING index_method |
  [ PROCEDURAL ] LANGUAGE object_name |
  SCHEMA object_name |
  SEQUENCE object_name |
  SERVER object_name |
  TABLE object_name |
  TEXT SEARCH CONFIGURATION object_name |
  TEXT SEARCH DICTIONARY object_name |
  TEXT SEARCH PARSER object_name |
  TEXT SEARCH TEMPLATE object_name |
  TYPE object_name |
  VIEW object_name
```

描述

`ALTER EXTENSION`更改已安装扩展的定义。其子形式如下：

`UPDATE`

这种形式将扩展更新到较新的版本。该扩展必须提供适当的更新脚本（或一系列脚本），能够将当前已安装版本修改为所请求的版本。

`SET SCHEMA`

这种形式将扩展的对象移入另一个模式。要使该命令成功，扩展必须是可重定位的。

`ADD` *`member_object`*

这种形式将一个现有对象加入扩展。这主要在扩展更新脚本中有用。
该对象随后将被视为扩展的成员；特别是，只能通过删除该扩展来删除它。

DROP *member_object*

这种形式从扩展中移除一个成员对象。这主要在扩展更新脚本中有用。该对象不会被删除，而只是与扩展解除关联。

有关这些操作的更多信息，见第 35.15 节。

要使用ALTER EXTENSION，你必须拥有该扩展。ADD/DROP形式还要求你拥有被添加/移除对象的所有权。

参数

extension_name

一个已安装扩展的名称。

new_version

扩展要更新到的新版本。它可以写成标识符或字符串常量。如果未指定，ALTER EXTENSION UPDATE会尝试更新到该扩展控制文件中标明的默认版本。

new_schema

扩展的新模式。

object_name
agg_name
function_name
operator_name

要被添加到扩展或从扩展移除的对象的名称。表、聚合、域、外部表、函数、操作符、操作符类、操作符族、序列、文本搜索对象、类型和视图的名称可以带模式限定。

agg_type

聚合函数作用于其上的输入数据类型。要引用一个零参数聚合函数，请在输入数据类型列表的位置写 *。

source_type

类型转换的源数据类型的名称。

target_type

类型转换的目标数据类型的名称。

argmode

函数或聚合函数参数的模式：IN、OUT、INOUT或VARIADIC。如果省略，默认值为IN。请注意，ALTER EXTENSION实际上不会关注OUT参数，因为确定函数身份只需要输入参数。因此，列出IN、INOUT和VARIADIC参数就足够了。

argname

函数或聚合函数参数的名称。请注意，ALTER EXTENSION实际上不会关注参数名称，因为确定函数身份只需要参数数据类型。

argtype

函数或聚合函数参数的数据类型。

left_type
right_type

操作符参数的数据类型（可带模式限定）。对于前缀或后缀操作符缺少的参数，写作`NONE`。

`PROCEDURAL`

这是一个噪声词。

示例

将`hstore`扩展更新到版本 2.0：

```
ALTER EXTENSION hstore UPDATE TO '2.0';
```

将`hstore`扩展的模式更改为`utils`：

```
ALTER EXTENSION hstore SET SCHEMA utils;
```

将现有函数添加到`hstore`扩展中：

```
ALTER EXTENSION hstore ADD FUNCTION populate_record(anyelement,
 hstore);
```

兼容性

`ALTER EXTENSION`是一个PostgreSQL 扩展。

另见

CREATE EXTENSION, DROP EXTENSION

ALTER FOREIGN DATA WRAPPER

ALTER FOREIGN DATA WRAPPER — change the definition of a foreign-data wrapper

大纲

```
ALTER FOREIGN DATA WRAPPER name
    [ HANDLER handler_function | NO HANDLER ]
    [ VALIDATOR validator_function | NO VALIDATOR ]
    [ OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ]) ]
ALTER FOREIGN DATA WRAPPER name OWNER TO new_owner
```

描述

`ALTER FOREIGN DATA WRAPPER`更改外部数据包装器的定义。该命令的第一种形式更改外部数据包装器的支持函数或通用选项（至少需要一个子句）。
第二种形式更改外部数据包装器的拥有者。

只有超级用户可以更改外部数据包装器。此外，
只有超级用户才能成为外部数据包装器的拥有者。

参数

name

一个现有外部数据包装器的名称。

`HANDLER` *handler_function*

为外部数据包装器指定一个新的处理器函数。

`NO HANDLER`

用于指定该外部数据包装器不再具有一个处理器函数。

注意，使用没有处理器函数的外部数据包装器的外部表将无法访问。

`VALIDATOR` *validator_function*

为外部数据包装器指定一个新的验证器函数。

注意，更改验证器之后，外部数据包装器、服务器和用户映射的选项有可能已经变为无效。在使用该外部数据包装器之前，用户有责任确保这些选项是正确的。

`NO VALIDATOR`

用于指定该外部数据包装器不再具有验证器函数。

`OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] )`

更改外部数据包装器的选项。`ADD`、`SET` 和`DROP`指定要执行的动作。如果未显式指定操作，则假定为`ADD`。选项名称必须唯一；如果存在验证器函数，
选项名称和值也会使用该外部数据包装器的验证器函数进行验证。

示例

更改一个外部数据包装器dbi，添加选项foo，删除bar：

```
ALTER FOREIGN DATA WRAPPER dbi OPTIONS (ADD foo '1', DROP 'bar');
```

将外部数据包装器dbi的验证器更改为 bob.myvalidator：

```
ALTER FOREIGN DATA WRAPPER dbi VALIDATOR bob.myvalidator;
```

兼容性

ALTER FOREIGN DATA WRAPPER 符合 ISO/IEC 9075-9（SQL/MED），但 HANDLER、VALIDATOR 和 OWNER TO 子句属于扩展。

另见

CREATE FOREIGN DATA WRAPPER, DROP FOREIGN DATA WRAPPER

ALTER FOREIGN TABLE

ALTER FOREIGN TABLE — change the definition of a foreign table

大纲

```
ALTER FOREIGN TABLE name
    action [, ... ]
ALTER FOREIGN TABLE name
    RENAME [ COLUMN ] column TO new_column
ALTER FOREIGN TABLE name
    RENAME TO new_name
ALTER FOREIGN TABLE name
    SET SCHEMA new_schema

where action is one of:

    ADD [ COLUMN ] column type [ NULL | NOT NULL ]
    DROP [ COLUMN ] [ IF EXISTS ] column [ RESTRICT | CASCADE ]
    ALTER [ COLUMN ] column [ SET DATA ] TYPE type
    ALTER [ COLUMN ] column { SET | DROP } NOT NULL
    OWNER TO new_owner
    OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ])
```

描述

`ALTER FOREIGN TABLE` 更改现有外部表的定义。有几种子形式：

`ADD COLUMN`

这种形式使用与CREATE FOREIGN TABLE相同的语法向外部表添加一个新列。

`DROP COLUMN [ IF EXISTS ]`

该形式从外部表中删除一列。如果该表之外的任何对象依赖于该列，例如视图，则需要指定 `CASCADE`。如果指定了`IF EXISTS` 而该列不存在，则不会抛出错误，而是改为发出一条提示。

`SET DATA TYPE`

该形式更改外部表中一列的类型。

`SET/DROP NOT NULL`

把一列标记为允许或不允许空值。

`OWNER`

该形式把外部表的拥有者更改为指定用户。

`RENAME`

`RENAME` 形式更改外部表的名称，或者更改外部表中某一列的名称。

`SET SCHEMA`

这种形式把外部表移动到另一个模式中。

`OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] )`

更改外部表的选项。ADD、SET 和 DROP 指定要执行的动作。如果没有显式指定操作，则默认为 ADD。选项名必须唯一；名称和值也会通过外部数据包装器库进行验证。

除了RENAME和SET SCHEMA之外，所有动作都可以组合在一个包含多项修改的列表中一并应用。例如，可以在一条命令中添加多个列和/或更改多个列的类型。

要使用 ALTER FOREIGN TABLE，必须拥有该表。要更改外部表的模式，还必须在新模式上拥有 CREATE 权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在表所在模式上拥有 CREATE 权限。（这些限制确保更改所有者不会执行任何无法通过删除并重新创建该表完成的操作。不过，超级用户无论如何都可以更改任何表的所有权。）

参数

name

要修改的现有外部表的名称（可以带模式限定）。

column

新列或现有列的名称。

new_column

现有列的新名称。

new_name

表的新名称。

type

新列的数据类型，或现有列的新数据类型。

CASCADE

自动删除依赖于该列的对象（例如引用该列的视图）。

RESTRICT

如果存在任何依赖对象，则拒绝删除该列。这是默认行为。

new_owner

该表的新拥有者的用户名。

new_schema

该表将被移动到的模式名称。

注解

关键字COLUMN只是噪声，可以省略。

使用`ADD COLUMN`或`DROP COLUMN`添加或移除列，添加`NOT NULL`约束，或使用`SET DATA TYPE`更改列类型时，不会检查与外部服务器的一致性。用户有责任确保表定义与远端一致。

关于有效参数的进一步说明，请参见CREATE FOREIGN TABLE。

示例

要把一列标记为非空：

```
ALTER FOREIGN TABLE distributors ALTER COLUMN street SET NOT NULL;
```

要更改外部表的选项：

```
ALTER FOREIGN TABLE myschema.distributors OPTIONS (ADD opt1 'value',
 SET opt2, 'value2', DROP opt3 'value3');
```

兼容性

`ADD`、`DROP`和`SET DATA TYPE` 形式符合 SQL 标准。其他形式是 PostgreSQL对 SQL 标准的扩展。另外，在单条`ALTER FOREIGN TABLE`命令中指定多个修改也是一种扩展。

`ALTER FOREIGN TABLE DROP COLUMN`可用于删除外部表的唯一一列，从而留下一个零列的表。这是 SQL 的一种扩展，因为 SQL 不允许零列外部表。

ALTER FUNCTION

ALTER FUNCTION — change the definition of a function

大纲

```
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    action [ ... ] [ RESTRICT ]
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    RENAME TO new_name
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    OWNER TO new_owner
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    SET SCHEMA new_schema

where action is one of:

    CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
    IMMUTABLE | STABLE | VOLATILE
    [ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER
    COST execution_cost
    ROWS result_rows
    SET configuration_parameter { TO | = } { value | DEFAULT }
    SET configuration_parameter FROM CURRENT
    RESET configuration_parameter
    RESET ALL
```

描述

`ALTER FUNCTION`更改函数的定义。

要使用`ALTER FUNCTION`，必须拥有该函数。要更改函数的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在函数所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建函数完成的操作。不过，
超级用户无论如何都可以更改任何函数的所有权。）

参数

name

现有函数的名称（可以用模式限定）。

argmode

参数的模式：`IN`、`OUT`、`INOUT`或`VARIADIC`。如果省略，默认为`IN`。注意`ALTER FUNCTION` 实际上并不关心任何`OUT`参数，因为确定函数身份只需要输入参数。因此，只列出`IN`、`INOUT`和`VARIADIC`参数就足够了。

argname

参数的名称。注意`ALTER FUNCTION`实际上并不关心参数名，因为确定函数身份只需要参数数据类型。

argtype

函数参数的数据类型（如果有，可以用模式限定）。

new_name

该函数的新名称。

new_owner

该函数的新拥有者。注意，如果该函数被标记为 SECURITY DEFINER，随后它将以新拥有者的身份执行。

new_schema

该函数的新模式。

CALLED ON NULL INPUT
RETURNS NULL ON NULL INPUT
STRICT

CALLED ON NULL INPUT将该函数改为在其部分或全部参数为 null 时仍会被调用。RETURNS NULL ON NULL INPUT或 STRICT将该函数改为只要任一参数为 null 就不调用，而是自动假定结果为 null。详见CREATE FUNCTION。

IMMUTABLE
STABLE
VOLATILE

将该函数的易变性更改为指定的设置。详见CREATE FUNCTION。

[EXTERNAL] SECURITY INVOKER
[EXTERNAL] SECURITY DEFINER

更改该函数是否为安全性定义者。关键字EXTERNAL 为了符合 SQL 标准而被忽略。关于这一能力的更多信息，参见CREATE FUNCTION。

COST *execution_cost*

更改该函数的估计执行代价。详见CREATE FUNCTION。

ROWS *result_rows*

更改集合返回函数估计返回的行数。详见CREATE FUNCTION。

configuration_parameter
value

添加或更改在调用该函数时要赋给某个配置参数的值。如果 *value*是DEFAULT，或者等效地使用了RESET，则会移除函数局部设置，这样该函数就会使用其环境中现有的值执行。使用 RESET ALL 可以清除所有函数局部设置。SET FROM CURRENT 会把执行ALTER FUNCTION 时该参数的当前值保存为进入函数时要应用的值。

关于允许的参数名和值的更多信息，参见SET和第 18 章。

RESTRICT

为符合 SQL 标准而被忽略。

示例

要将类型integer的函数sqrt 重命名为square_root：

```
ALTER FUNCTION sqrt(integer) RENAME TO square_root;
```

要将类型integer的函数sqrt 的拥有者改为joe：

```
ALTER FUNCTION sqrt(integer) OWNER TO joe;
```

要将类型integer的函数sqrt 的模式改为maths：

```
ALTER FUNCTION sqrt(integer) SET SCHEMA maths;
```

要调整自动为函数设置的搜索路径：

```
ALTER FUNCTION check_password(text) SET search_path = admin, pg_temp;
```

要禁用自动为函数设置的search_path：

```
ALTER FUNCTION check_password(text) RESET search_path;
```

现在该函数将使用其调用者所使用的搜索路径执行。

兼容性

该语句与 SQL 标准中的ALTER FUNCTION语句部分兼容。标准允许修改函数的更多属性，但不提供重命名函数、将函数设为安全性定义者、为函数附加配置参数值，或者更改函数的拥有者、模式或易变性的能力。标准还要求使用RESTRICT关键字，而它在 PostgreSQL中是可选的。

另见

CREATE FUNCTION, DROP FUNCTION

ALTER GROUP

ALTER GROUP — change role name or membership

大纲

```
ALTER GROUP group_name ADD USER user_name [, ... ]
ALTER GROUP group_name DROP USER user_name [, ... ]

ALTER GROUP group_name RENAME TO new_name
```

描述

`ALTER GROUP`更改用户组的属性。这是一个已废弃的命令，但出于向后兼容的考虑仍然接受它，因为组（以及用户）已经被更一般的角色概念所取代。

前两个变体将用户加入一个组，或将其从组中移除。（为此目的，任何角色都可以充当“用户”或“组”。）这些变体实际上等效于对名为“组”的角色授予或撤销成员资格，因此更推荐使用GRANT或REVOKE。

第三个变体更改组的名称。这完全等同于使用ALTER ROLE重命名角色。

参数

group_name

要修改的组（角色）的名称。

user_name

要加入该组或从该组中移除的用户（角色）。这些用户必须已经存在；`ALTER GROUP`不会创建或删除用户。

new_name

该组的新名称。

示例

向组中添加用户：

```
ALTER GROUP staff ADD USER karl, john;
```

从组中移除用户：

```
ALTER GROUP workers DROP USER beth;
```

兼容性

SQL 标准中没有`ALTER GROUP`语句。

另见

GRANT, REVOKE, ALTER ROLE

ALTER INDEX

ALTER INDEX — change the definition of an index

大纲

```
ALTER INDEX name RENAME TO new_name
ALTER INDEX name SET TABLESPACE tablespace_name
ALTER INDEX name SET ( storage_parameter = value [, ... ] )
ALTER INDEX name RESET ( storage_parameter [, ... ] )
```

描述

`ALTER INDEX`更改现有索引的定义。有几种子形式：

`RENAME`

`RENAME`形式更改索引的名称。存储的数据不受影响。

`SET TABLESPACE`

这种形式把索引的表空间更改为指定的表空间，
并把与该索引关联的数据文件移动到新表空间。另见CREATE TABLESPACE。

`SET ( storage_parameter = value [, ... ] )`

该形式更改索引的一个或多个索引方法特定存储参数。有关可用参数的详情，请参见CREATE INDEX。请注意，该命令不会立即修改索引内容；根据参数的不同，可能需要使用REINDEX重建索引才能达到预期效果。

`RESET ( storage_parameter [, ... ] )`

这种形式将一个或多个索引方法相关的存储参数重置为默认值。与 `SET`一样，可能仍需要执行一次`REINDEX` 才能让该索引被完全更新。

参数

name

要修改的现有索引名称（可以是模式限定的）。

new_name

该索引的新名称。

tablespace_name

该索引将被移动到的表空间。

storage_parameter

一个索引方法相关的存储参数的名称。

value

一个索引方法相关的存储参数的新值。根据具体参数，它可能是数字或一个词。

注解

这些操作也可以使用ALTER TABLE完成。实际上，`ALTER INDEX`只是`ALTER TABLE`中适用于索引的那些形式的别名。

以前曾有一种`ALTER INDEX OWNER`变体，但现在会被忽略（同时发出警告）。索引的拥有者不能不同于其所属表的拥有者。更改表的拥有者时，索引的拥有者也会自动更改。

不允许更改系统目录索引的任何部分。

示例

要重命名现有索引：

```
ALTER INDEX distributors RENAME TO suppliers;
```

要将索引移到另一个表空间：

```
ALTER INDEX distributors SET TABLESPACE fasttablespace;
```

要更改索引的填充因子（假设该索引方法支持fillfactor参数）：

```
ALTER INDEX distributors SET (fillfactor = 75);
REINDEX INDEX distributors;
```

兼容性

`ALTER INDEX`是一种 PostgreSQL扩展。

另见

CREATE INDEX, REINDEX

ALTER LANGUAGE

ALTER LANGUAGE — change the definition of a procedural language

大纲

```
ALTER [ PROCEDURAL ] LANGUAGE name RENAME TO new_name
ALTER [ PROCEDURAL ] LANGUAGE name OWNER TO new_owner
```

描述

`ALTER LANGUAGE`更改一种过程语言的定义。唯一的功能是重命名该语言或为其指定新的所有者。要使用`ALTER LANGUAGE`，你必须是超级用户或者该语言的所有者。

参数

name

语言的名称

new_name

该语言的新名称

new_owner

该语言的新所有者

兼容性

SQL 标准中没有`ALTER LANGUAGE`语句。

另见

CREATE LANGUAGE, DROP LANGUAGE

ALTER LARGE OBJECT

ALTER LARGE OBJECT — change the definition of a large object

大纲

`ALTER LARGE OBJECT` *`large_object_oid`* `OWNER TO` *`new_owner`*

描述

`ALTER LARGE OBJECT`更改一个大对象的定义。其唯一功能是分配新的所有者。必须以超级用户或大对象所有者的身份使用 `ALTER LARGE OBJECT`。

参数

`large_object_oid`

要修改的大对象的 OID

`new_owner`

该大对象的新拥有者

兼容性

SQL 标准中没有`ALTER LARGE OBJECT`语句。

另见

第 32 章

ALTER OPERATOR

ALTER OPERATOR — change the definition of an operator

大纲

```
ALTER OPERATOR name ( { left_type | NONE } , { right_type | NONE } )
 OWNER TO new_owner
ALTER OPERATOR name ( { left_type | NONE } , { right_type | NONE } )
 SET SCHEMA new_schema
```

描述

`ALTER OPERATOR`更改一个操作符的定义。目前唯一可用的功能是更改操作符的拥有者。

要使用`ALTER OPERATOR`，必须拥有该操作符。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在操作符所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建操作符完成的操作。不过，超级用户无论如何都可以更改任何操作符的所有权。）

参数

name

现有操作符的名称（可带模式限定）。

left_type

操作符左操作数的数据类型；如果操作符没有左操作数，写作`NONE`。

right_type

操作符右操作数的数据类型；如果操作符没有右操作数，写作`NONE`。

new_owner

操作符的新所有者。

new_schema

操作符的新模式。

示例

更改自定义操作符的限制和连接选择率估计函数，操作符为`a && b`，类型为`int[]`：

```
ALTER OPERATOR && (_int4, _int4) SET (RESTRICT = _int_contsel, JOIN =
 _int_contjoinsel);
```

兼容性

在 SQL 标准中没有`ALTER OPERATOR`语句。

另见

CREATE OPERATOR, DROP OPERATOR

ALTER OPERATOR CLASS

ALTER OPERATOR CLASS — change the definition of an operator class

大纲

```
ALTER OPERATOR CLASS name USING index_method RENAME TO new_name
ALTER OPERATOR CLASS name USING index_method OWNER TO new_owner
ALTER OPERATOR CLASS name USING index_method SET SCHEMA new_schema
```

描述

`ALTER OPERATOR CLASS`更改一个操作符类的定义。

要使用`ALTER OPERATOR CLASS`，必须拥有该操作符类。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在操作符类所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建操作符类完成的操作。不过，超级用户无论如何都可以更改任何操作符类的所有权。）

参数

name

一个现有操作符类的名称（可以是模式限定的）。

index_method

该操作符类所对应的索引方法的名称。

new_name

该操作符类的新名称。

new_owner

该操作符类的新拥有者。

new_schema

该操作符类的新模式。

兼容性

SQL 标准中没有`ALTER OPERATOR CLASS`语句。

另见

CREATE OPERATOR CLASS, DROP OPERATOR CLASS, ALTER OPERATOR FAMILY

ALTER OPERATOR FAMILY

ALTER OPERATOR FAMILY — change the definition of an operator family

大纲

```
ALTER OPERATOR FAMILY name USING index_method ADD
  {  OPERATOR strategy_number operator_name ( op_type, op_type ) [ FOR
 SEARCH | FOR ORDER BY sort_family_name ]
   | FUNCTION support_number [ ( op_type [ , op_type ] ) ] function_
name ( argument_type [, ...] )
  } [, ... ]
ALTER OPERATOR FAMILY name USING index_method DROP
  {  OPERATOR strategy_number ( op_type [ , op_type ] )
   | FUNCTION support_number ( op_type [ , op_type ] )
  } [, ... ]
ALTER OPERATOR FAMILY name USING index_method RENAME TO new_name
ALTER OPERATOR FAMILY name USING index_method OWNER TO new_owner
ALTER OPERATOR FAMILY name USING index_method SET SCHEMA new_schema
```

描述

`ALTER OPERATOR FAMILY`更改一个操作符族的定义。你可以向该族添加操作符和支持函数、从该族中移除它们，或者更改该族的名称或拥有者。

使用`ALTER OPERATOR FAMILY`向一个族中添加操作符和支持函数时，它们并不属于该族内任何特定的操作符类，而只是作为该族中的“松散”成员存在。
这表明这些操作符和函数与该族的语义兼容，但并不是任何特定索引正确工作所必需的。
（那些确实必需的操作符和函数应当声明为某个操作符类的一部分；参见CREATE OPERATOR CLASS。）PostgreSQL允许随时从一个族中删除松散成员，
但如果不删除整个类以及依赖它的任何索引，就不能删除操作符类的成员。通常，
单一数据类型的操作符和函数属于操作符类，因为支持该特定数据类型上的索引需要它们，
而跨数据类型的操作符和函数则作为该族中的松散成员。

要使用`ALTER OPERATOR FAMILY`，你必须是超级用户。（之所以这样限制，是因为错误的操作符族定义可能会使服务器混乱，甚至导致崩溃。）

`ALTER OPERATOR FAMILY`目前不会检查操作符族定义是否包括该索引方法所要求的全部操作符和函数，也不会检查这些操作符和函数是否构成一个自洽的集合。
定义一个合法的操作符族是用户的责任。

更多信息请参见第 35.14 节。

参数

name

一个现有操作符族的名称（可以是模式限定的）。

index_method

该操作符族所对应的索引方法名称。

strategy_number

与该操作符族关联的操作符的索引方法策略号。

operator_name

与该操作符族关联的操作符名称（可以是模式限定的）。

op_type

在OPERATOR子句中，操作符的操作数数据类型，或使用NONE表示左一元或右一元操作符。与CREATE OPERATOR CLASS中的对应语法不同，必须始终指定操作数数据类型。

在 ADD FUNCTION 子句中，如果函数要支持的操作数数据类型不同于函数的输入数据类型，则指定这些数据类型。对于 B-树和哈希索引，不需要指定 *op_type*，因为函数的输入数据类型始终是应使用的正确类型。对于 GIN 和 GiST 索引，则必须指定该函数要配合使用的输入数据类型。

在DROP FUNCTION子句中，必须指定该函数意图支持的操作数数据类型。

sort_family_name

一个现有btree操作符族的名称（可以是模式限定的），
它描述与某个排序操作符关联的排序顺序。

如果既没有指定FOR SEARCH，也没有指定 FOR ORDER BY，则默认值是FOR SEARCH。

support_number

与操作符族关联的函数的索引方法支持函数编号。

function_name

作为操作符族索引方法支持函数的函数名称（可带模式限定）。

argument_type

该函数的参数数据类型。

new_name

该操作符族的新名称。

new_owner

该操作符族的新拥有者。

new_schema

该操作符族所在的新模式。

OPERATOR和FUNCTION子句可以以任何顺序出现。

注解

注意，DROP语法只通过策略号或支持函数编号以及输入数据类型来指定操作符族中的“槽位”，不会提及占用该槽位的操作符或函数名称。另外，对于DROP FUNCTION，要指定的类型是该函

数意图支持的输入数据类型；对于 GiST、SP-GiST 和 GIN 索引，这些类型可能与该函数实际的输入参数类型毫无关系。

因为索引机制在使用函数之前不会检查函数的访问权限，
将一个函数或者操作符包括在一个操作符族中相当于在其上授予公共执行权限。
这对操作符族中很有用的这类函数来说通常不成问题。

不应由 SQL 函数定义这些操作符。SQL 函数很可能会被内联到调用查询中，从而阻止优化器识别该查询与索引匹配。

在PostgreSQL 8.4 之前，`OPERATOR`子句可以包含`RECHECK`选项。现在不再支持该选项，因为索引操作符是否“有损”会在运行时动态确定。
这样可以高效处理操作符可能有损也可能无损的情况。

示例

下列示例命令向一个已经包含数据类型`int4`和`int2` 的 B-树操作符类的操作符族中添加跨数据类型的操作符和支持函数。

```
ALTER OPERATOR FAMILY integer_ops USING btree ADD

  -- int4 vs int2
  OPERATOR 1 < (int4, int2) ,
  OPERATOR 2 <= (int4, int2) ,
  OPERATOR 3 = (int4, int2) ,
  OPERATOR 4 >= (int4, int2) ,
  OPERATOR 5 > (int4, int2) ,
  FUNCTION 1 btint42cmp(int4, int2) ,

  -- int2 vs int4
  OPERATOR 1 < (int2, int4) ,
  OPERATOR 2 <= (int2, int4) ,
  OPERATOR 3 = (int2, int4) ,
  OPERATOR 4 >= (int2, int4) ,
  OPERATOR 5 > (int2, int4) ,
  FUNCTION 1 btint24cmp(int2, int4) ;
```

要再次移除这些条目：

```
ALTER OPERATOR FAMILY integer_ops USING btree DROP

  -- int4 vs int2
  OPERATOR 1 (int4, int2) ,
  OPERATOR 2 (int4, int2) ,
  OPERATOR 3 (int4, int2) ,
  OPERATOR 4 (int4, int2) ,
  OPERATOR 5 (int4, int2) ,
  FUNCTION 1 (int4, int2) ,

  -- int2 vs int4
  OPERATOR 1 (int2, int4) ,
  OPERATOR 2 (int2, int4) ,
  OPERATOR 3 (int2, int4) ,
  OPERATOR 4 (int2, int4) ,
  OPERATOR 5 (int2, int4) ,
```

```
FUNCTION 1 (int2, int4) ;
```

兼容性

在 SQL 标准中没有 `ALTER OPERATOR FAMILY`语句。

另见

CREATE OPERATOR FAMILY, DROP OPERATOR FAMILY, CREATE OPERATOR CLASS, ALTER OPERATOR CLASS, DROP OPERATOR CLASS

ALTER ROLE

ALTER ROLE — change a database role

大纲

```
ALTER ROLE name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | REPLICATION | NOREPLICATION
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'

ALTER ROLE name RENAME TO new_name

ALTER ROLE name [ IN DATABASE database_name ] SET configuration_
parameter { TO | = } { value | DEFAULT }
ALTER ROLE name [ IN DATABASE database_name ] SET configuration_
parameter FROM CURRENT
ALTER ROLE name [ IN DATABASE database_name ] RESET configuration_
parameter
ALTER ROLE name [ IN DATABASE database_name ] RESET ALL
```

描述

`ALTER ROLE`更改 PostgreSQL角色的属性。

概要中列出的此命令的第一种形式可以更改CREATE ROLE中可指定的许多角色属性。（涵盖了所有可能的属性，但没有用于添加或移除成员资格的选项；请使用GRANT和REVOKE完成此操作。）命令中未提及的属性保留其以前的设置。数据库超级用户可以更改任何角色的任何这些设置。拥有`CREATEROLE`权限的角色可以更改任何这些设置，但只能针对非超级用户且非复制角色。普通角色只能更改自己的密码。

第二种形式更改角色的名称。数据库超级用户可以重命名任何角色。拥有`CREATEROLE`权限的角色可以重命名非超级用户角色。当前会话用户不能被重命名。（如果需要这样做，请以其他用户连接。）由于`MD5`加密的密码使用角色名作为密码学盐，重命名角色会清除其密码（如果密码使用`MD5`加密）。

其余的变体更改角色的某个配置变量的会话默认值，可以是针对所有数据库，也可以在指定了 `IN DATABASE` 子句时只针对指定数据库中的会话。每当该角色随后启动新会话时，指定的值就成为会话默认值，覆盖 `postgresql.conf` 中的设置或从 `postgres` 命令行接收的设置。这只发生在登录时；执行SET ROLE或SET SESSION AUTHORIZATION不会导致设置新的配置值。为所有数

据库设置的值会被附加到角色的数据库特定设置覆盖。超级用户可以更改任何人的会话默认值。具有 CREATEROLE 权限的角色可以更改非超级用户角色的默认值。普通角色只能为自己设置默认值。某些配置变量不能用这种方式设置，或者只有超级用户发出命令时才能设置。

参数

name

将要修改其属性的角色名称。

SUPERUSER
NOSUPERUSER
CREATEDB
NOCREATEDB
CREATEROLE
NOCREATEROLE
CREATEUSER
NOCREATEUSER
INHERIT
NOINHERIT
LOGIN
NOLOGIN
REPLICATION
NOREPLICATION
CONNECTION LIMIT *connlimit*
PASSWORD *password*
ENCRYPTED
UNENCRYPTED
VALID UNTIL '*timestamp*'

这些子句更改最初由CREATE ROLE设置的属性。更多信息请参见CREATE ROLE参考页。

new_name

角色的新名称。

database_name

应在其中设置该配置变量的数据库名称。

configuration_parameter
value

将该角色在指定配置参数上的会话默认值设为给定值。如果 *value*是DEFAULT，或者等效地使用了RESET，则会移除角色特定变量设置，
因此该角色在新会话中将继承系统范围的默认设置。使用 RESET ALL可清除所有角色特定设置。SET FROM CURRENT会将该参数在当前会话中的值保存为角色特定值。如果指定了IN DATABASE，则只会为给定角色和数据库设置或移除该配置参数。

角色特定的变量设置只在登录时生效；SET ROLE和SET SESSION AUTHORIZATION不会处理角色特定的变量设置。

关于允许的参数名称和值详见SET和第 18 章。

注解

使用CREATE ROLE添加新角色，使用DROP ROLE移除角色。

`ALTER ROLE`不能更改角色的成员资格。请使用GRANT和REVOKE完成此操作。

用这个命令指定未加密密码时必须加以注意。密码将以明文形式传输到服务器，
并且也可能被记录在客户端命令历史或服务器日志中。psql包含一个命令`\password`，
可用于更改角色的密码而不暴露明文密码。

也可以把会话默认值绑定到特定数据库，而不是角色；详见ALTER DATABASE。如果发生冲突，
同时针对数据库和角色的设置会覆盖角色特定设置，而角色特定设置又会覆盖数据库特定设置。

示例

更改角色的密码：

```
ALTER ROLE davide WITH PASSWORD 'hu8jmn3';
```

移除角色的密码：

```
ALTER ROLE davide WITH PASSWORD NULL;
```

更改密码的失效日期，并指定该密码应在比 UTC快 1 小时的时区中，于 2015 年 5 月 4 日中午失效：

```
ALTER ROLE chris VALID UNTIL 'May 4 12:00:00 2015 +1';
```

使密码永久有效：

```
ALTER ROLE fred VALID UNTIL 'infinity';
```

授予角色创建其他角色和新数据库的能力：

```
ALTER ROLE miriam CREATEROLE CREATEDB;
```

为角色指定maintenance_work_mem参数的非默认设置：

```
ALTER ROLE worker_bee SET maintenance_work_mem = 100000;
```

为角色指定client_min_messages参数的数据库特定的非默认设置：

```
ALTER ROLE fred IN DATABASE devel SET client_min_messages = DEBUG;
```

兼容性

`ALTER ROLE`语句是 PostgreSQL的扩展。

另见

CREATE ROLE, DROP ROLE, SET

ALTER SCHEMA

ALTER SCHEMA — change the definition of a schema

大纲

```
ALTER SCHEMA name RENAME TO new_name
ALTER SCHEMA name OWNER TO new_owner
```

描述

`ALTER SCHEMA`更改一个模式的定义。

要使用`ALTER SCHEMA`，必须拥有该模式。要重命名模式，还必须在数据库上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且必须在数据库上拥有`CREATE`权限。（请注意，超级用户会自动拥有所有这些权限。）

参数

name

现有模式的名称。

new_name

该模式的新名称。新名称不能以`pg_`开头，因为这类名称是为系统模式保留的。

new_owner

该模式的新拥有者。

兼容性

SQL 标准中没有`ALTER SCHEMA`语句。

另见

CREATE SCHEMA, DROP SCHEMA

ALTER SEQUENCE

ALTER SEQUENCE — change the definition of a sequence generator

大纲

```
ALTER SEQUENCE name [ INCREMENT [ BY ] increment ]
    [ MINVALUE minvalue | NO MINVALUE ] [ MAXVALUE maxvalue | NO
 MAXVALUE ]
    [ START [ WITH ] start ]
    [ RESTART [ [ WITH ] restart ] ]
    [ CACHE cache ] [ [ NO ] CYCLE ]
    [ OWNED BY { table.column | NONE } ]
ALTER SEQUENCE name OWNER TO new_owner
ALTER SEQUENCE name RENAME TO new_name
ALTER SEQUENCE name SET SCHEMA new_schema
```

描述

`ALTER SEQUENCE`更改现有序列发生器的参数。在 `ALTER SEQUENCE`命令中未明确设置的参数，会保留其原有设置。

要使用`ALTER SEQUENCE`，必须拥有该序列。要更改序列的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在序列所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建序列完成的操作。不过，
超级用户无论如何都可以更改任何序列的所有权。）

参数

name

要更改的序列名称（可以是模式限定的）。

increment

子句`INCREMENT BY` *increment*是可选的。正值会创建升序序列，负值会创建降序序列。如果未指定，则保留原有的增量值。

minvalue
`NO MINVALUE`

可选子句`MINVALUE` *minvalue*确定序列可生成的最小值。如果指定了`NO MINVALUE`，则升序序列和降序序列将分别使用默认值 1 和 $-2^{63}-1$。如果两种选项都未指定，则保留当前最小值。

maxvalue
`NO MAXVALUE`

可选子句`MAXVALUE` *maxvalue*确定序列的最大值。如果指定了`NO MAXVALUE`，则升序序列和降序序列将分别使用默认值 $2^{63}-1$ 和 -1。如果两种选项都未指定，则保留当前最大值。

start

可选子句START WITH *start*更改序列记录的起始值。这不会影响当前序列值；它只是设置未来ALTER SEQUENCE RESTART命令将使用的值。

restart

可选子句RESTART [WITH *restart*]更改序列的当前值。这等价于以is_called = false调用setval函数：指定的值会由对nextval的下一次调用返回。写成不带*restart*值的RESTART，等价于提供由CREATE SEQUENCE记录、或最近由ALTER SEQUENCE START WITH设置的起始值。

cache

CACHE *cache*子句启用预分配序列号并将其存储在内存中，以便更快访问。最小值为 1（一次只能生成一个值，即不缓存）。如果未指定，则保留旧的缓存值。

CYCLE

可选的CYCLE关键字可用于启用序列回绕：当递增或递减序列分别达到*maxvalue*或*minvalue*时回绕。如果达到限制，下一个生成的数将分别为*minvalue*或*maxvalue*。

NO CYCLE

如果指定了可选关键字NO CYCLE，则当序列达到其最大值后，任何对 nextval的调用都会返回错误。如果既未指定CYCLE也未指定 NO CYCLE，则保留原有的循环行为。

OWNED BY *table.column*
OWNED BY NONE

OWNED BY选项使序列与特定表列关联；如果该列（或其整个表）被删除，序列也会自动删除。如果指定此选项，该关联会替换序列此前指定的任何关联。
指定的表必须与序列拥有相同的所有者，并且位于同一模式中。指定OWNED BY NONE会移除现有的关联，使序列成为“独立的”对象。

new_owner

序列的新所有者的用户名。

new_name

该序列的新拥有者的用户名。

new_schema

该序列的新名称。

注解

为了避免阻塞从同一序列获取数字的并发事务，ALTER SEQUENCE对序列生成参数的更改永远不会被回滚；这些更改会立即生效且不可逆转。不过，OWNED BY、OWNER TO、RENAME TO和SET SCHEMA子句引起的是普通的目录更新，可以回滚。

ALTER SEQUENCE不会立即影响除当前后端之外、已经预分配（缓存）了序列值的其他后端中的 nextval结果。它们会先用完所有缓存值，然后才会注意到序列生成参数已经改变。当前后端则会立即受到影响。

`ALTER SEQUENCE`不会影响该序列的`currval`状态。（在PostgreSQL 8.3 之前，它有时会这样做。）

由于历史原因，序列也可以使用`ALTER TABLE`；但允许用于序列的 `ALTER TABLE`变体仅限于与上述形式等价的那些。

示例

将一个名为`serial`的序列重启到 105：

```
ALTER SEQUENCE serial RESTART WITH 105;
```

兼容性

`ALTER SEQUENCE`符合SQL标准，但 `START WITH`、`OWNED BY`、`OWNER TO`、`RENAME TO`和`SET SCHEMA`子句是PostgreSQL扩展。

另见

CREATE SEQUENCE, DROP SEQUENCE

ALTER SERVER

ALTER SERVER — change the definition of a foreign server

大纲

```
ALTER SERVER server_name [ VERSION 'new_version' ]
    [ OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] ) ]
ALTER SERVER server_name OWNER TO new_owner
```

描述

`ALTER SERVER`更改外部服务器的定义。第一种形式更改服务器的版本字符串或其通用选项（至少需要一个子句）。第二种形式更改服务器的拥有者。

要更改服务器，必须是服务器的所有者。此外，要更改所有者，还必须拥有该服务器，是新所有者角色的直接或间接成员，并且必须在服务器的外部数据包装器上拥有`USAGE`权限。（请注意，超级用户会自动满足所有这些条件。）

参数

server_name

一个现有服务器的名称。

new_version

新服务器版本。

`OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] )`

更改服务器的选项。`ADD`、`SET`和 `DROP`指定要执行的动作。如果未显式指定操作，则假定为`ADD`。选项名称必须唯一；名称和值也会使用该服务器的外部数据包装器库进行验证。

示例

更改服务器`foo`，添加连接选项：

```
ALTER SERVER foo OPTIONS (host 'foo', dbname 'foodb');
```

更改服务器`foo`，更改版本并修改 `host`选项：

```
ALTER SERVER foo VERSION '8.4' OPTIONS (SET host 'baz');
```

兼容性

`ALTER SERVER` 符合 ISO/IEC 9075-9（SQL/MED）。

另见

CREATE SERVER, DROP SERVER

ALTER TABLE

ALTER TABLE — change the definition of a table

大纲

```
ALTER TABLE [ ONLY ] name [ * ]
    action [, ... ]
ALTER TABLE [ ONLY ] name [ * ]
    RENAME [ COLUMN ] column TO new_column
ALTER TABLE name
    RENAME TO new_name
ALTER TABLE name
    SET SCHEMA new_schema

where action is one of:

    ADD [ COLUMN ] column data_type [ COLLATE collation ] [ column_
constraint [ ... ] ]
    DROP [ COLUMN ] [ IF EXISTS ] column [ RESTRICT | CASCADE ]
    ALTER [ COLUMN ] column [ SET DATA ] TYPE data_type [
 COLLATE collation ] [ USING expression ]
    ALTER [ COLUMN ] column SET DEFAULT expression
    ALTER [ COLUMN ] column DROP DEFAULT
    ALTER [ COLUMN ] column { SET | DROP } NOT NULL
    ALTER [ COLUMN ] column SET STATISTICS integer
    ALTER [ COLUMN ] column SET ( attribute_option = value [, ... ] )
    ALTER [ COLUMN ] column RESET ( attribute_option [, ... ] )
    ALTER [ COLUMN ] column SET STORAGE { PLAIN | EXTERNAL | EXTENDED
 | MAIN }
    ADD table_constraint [ NOT VALID ]
    ADD table_constraint_using_index
    VALIDATE CONSTRAINT constraint_name
    DROP CONSTRAINT [ IF EXISTS ]  constraint_name [ RESTRICT |
 CASCADE ]
    DISABLE TRIGGER [ trigger_name | ALL | USER ]
    ENABLE TRIGGER [ trigger_name | ALL | USER ]
    ENABLE REPLICA TRIGGER trigger_name
    ENABLE ALWAYS TRIGGER trigger_name
    DISABLE RULE rewrite_rule_name
    ENABLE RULE rewrite_rule_name
    ENABLE REPLICA RULE rewrite_rule_name
    ENABLE ALWAYS RULE rewrite_rule_name
    CLUSTER ON index_name
    SET WITHOUT CLUSTER
    SET WITH OIDS
    SET WITHOUT OIDS
    SET ( storage_parameter = value [, ... ] )
    RESET ( storage_parameter [, ... ] )
    INHERIT parent_table
    NO INHERIT parent_table
    OF type_name
```

```
    NOT OF
    OWNER TO new_owner
    SET TABLESPACE new_tablespace
```

and *table_constraint_using_index* is:

```
    [ CONSTRAINT constraint_name ]
    { UNIQUE | PRIMARY KEY } USING INDEX index_name
    [ DEFERRABLE | NOT DEFERRABLE ] [ INITIALLY DEFERRED | INITIALLY
 IMMEDIATE ]
```

描述

ALTER TABLE 更改现有表的定义。下面介绍几种子形式：

ADD COLUMN

这种形式使用与CREATE TABLE相同的语法向表添加一个新列。

DROP COLUMN [IF EXISTS]

这种形式从表中删除一个列。涉及该列的索引和表约束也将被自动删除。
如果表外的任何对象依赖于该列（例如外键引用或视图），则需要指定 CASCADE。如果指定了 IF EXISTS 而该列不存在，则不会抛出错误，而是发出一个提示。

SET DATA TYPE

该形式更改表中某列的类型。
涉及该列的索引和简单表约束会通过重新解析最初提供的表达式，
自动转换为使用新列类型。可选的COLLATE子句为新列指定排序规则；如果省略，
排序规则是新列类型的默认排序规则。可选的USING子句指定如何根据旧值计算新列值；
如果省略，默认转换与从旧数据类型到新数据类型的赋值转换相同。
如果旧类型到新类型之间不存在隐式转换或赋值转换，则必须提供USING子句。

SET/DROP DEFAULT

这些形式为列设置或移除默认值。默认值只会应用于后续的 INSERT 命令；它们不会导致表中已有的行发生变化。也可以为视图创建默认值，这种情况下，在应用视图的 ON INSERT 规则之前，默认值会被插入到视图上的 INSERT 语句中。

SET/DROP NOT NULL

这些形式更改列是否被标记为允许空值，或拒绝空值。只有当列不包含空值时，才能使用SET NOT NULL。

SET STATISTICS

该形式为后续ANALYZE操作设置每列的统计信息收集目标。目标可以设置在 0 到 10000 范围内；也可以将其设置为 -1，以恢复使用系统默认的统计目标（default_statistics_target）。
有关PostgreSQL查询规划器使用统计信息的更多信息，请参见第 14.2 节。

SET (*attribute_option* = *value* [, ...])
RESET (*attribute_option* [, ...])

该形式设置或重置每属性选项。目前定义的每属性选项只有n_distinct和n_distinct_inherited，它们会覆盖后续ANALYZE操作所作的不同值数量估计。n_distinct
影响表本身的统计信息，而n_distinct_inherited

影响为表及其继承子表收集的统计信息。当设置为正值时，ANALYZE
会假定该列恰好包含指定数量的不同非空值。当设置为负值时（该值必须大于或等于 -1），ANALYZE会假定列中不同非空值的数量与表大小成线性关系；
具体数量通过将估计的表大小乘以给定数值的绝对值计算。例如，-1 表示列中的所有值都不同，而 -0.5 表示平均每个值出现两次。当表大小随时间变化时，这可能很有用，因为只有在查询规划时才会执行与表中行数相乘的操作。指定 0 可恢复正常估计不同值数量。有关PostgreSQL查询规划器使用统计信息的更多信息，请参见第 14.2 节。

SET STORAGE

该形式为列设置存储模式。这控制该列是内联保存还是保存在辅助TOAST表中，
以及是否压缩数据。对于integer等定长值，必须使用PLAIN，
数据以内联且未压缩的形式保存。MAIN用于内联的可压缩数据。EXTERNAL
用于外部保存的未压缩数据，EXTENDED用于外部保存的压缩数据。对于支持非PLAIN
存储的大多数数据类型，EXTENDED是默认值。使用EXTERNAL会使非常大的text和bytea
值上的子字符串操作运行得更快，但会增加存储空间。请注意，SET STORAGE本身不会更改表中的任何内容，它只会设置今后更新表时采用的策略。更多信息请参见第 55.2 节。

ADD *table_constraint* [NOT VALID]

这种形式使用与CREATE TABLE相同的语法向表添加一个新约束，另外还提供 NOT VALID 选项，目前该选项只允许用于外键约束。如果约束被标记为 NOT VALID，则会跳过验证表中所有行都满足该约束的这一可能耗时很长的初始检查。
后续的插入或更新仍会强制执行该约束（也就是说，除非被引用表中存在匹配的行，
否则操作会失败）。但是，在使用 VALIDATE CONSTRAINT 选项验证该约束之前，数据库不会假定它对表中的所有行都成立。

ADD *table_constraint_using_index*

该形式根据现有唯一索引向表添加新的PRIMARY KEY或UNIQUE约束。索引的所有列都会包含在约束中。

索引不能包含表达式列，也不能是部分索引。此外，它必须是使用默认排序方式的 B-树索引。这些限制确保该索引等效于使用常规ADD PRIMARY KEY或ADD UNIQUE命令构建的索引。

如果指定了PRIMARY KEY，且索引的列尚未标记为NOT NULL，则此命令会尝试对每个此类列执行ALTER COLUMN SET NOT NULL。这需要完整扫描表，以验证这些列不包含空值。在其他所有情况下，这都是快速操作。

如果提供了约束名称，索引将被重命名以匹配约束名称。否则，约束将与索引同名。

执行此命令后，索引将由约束“拥有”，其方式与常规ADD PRIMARY KEY或ADD UNIQUE命令构建索引时相同。特别是，删除约束也会使索引消失。

> **注意**
>
> 在需要添加新约束且不希望长时间阻塞表更新的情况下，
> 使用现有索引添加约束可能很有用。为此，请使用CREATE INDEX CONCURRENTLY创建索引，然后使用此语法将其安装为正式约束。请参见下面的示例。

VALIDATE CONSTRAINT

这种形式通过扫描表来验证先前以 NOT VALID 方式创建的外键约束，确保不存在不匹配的行。如果该约束已经被标记为有效，则不会发生任何事情。

在较大的表上，验证可能是一个耗时的过程，并且当前要求持有`ACCESS EXCLUSIVE`锁。把验证与初始创建分开的意义在于：可以把验证推迟到系统较空闲的时候进行，也可以利用这段时间来修正已有的错误，同时防止产生新的错误。

`DROP CONSTRAINT [ IF EXISTS ]`

该形式从表中删除指定的约束。如果指定`IF EXISTS`且约束不存在，则不会抛出错误；在这种情况下会发出通知。

`DISABLE/ENABLE [ REPLICA | ALWAYS ] TRIGGER`

这些形式配置属于表的触发器的触发。已禁用的触发器仍为系统所知，但在触发事件发生时不会执行。对于延迟触发器，会在事件发生时检查启用状态，而不是在实际执行触发器函数时检查。可以按名称指定单个触发器，或指定表上的所有触发器，或仅指定用户触发器，从而禁用或启用触发器（最后一种选项不包括内部生成的约束触发器，例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器）。禁用或启用内部生成的约束触发器需要超级用户权限；应谨慎操作，因为如果不执行触发器，当然无法保证约束的完整性。触发器触发机制还会受到配置变量session_replication_role的影响。简单启用的触发器会在复制角色为“origin”（默认值）或“local”时触发。配置为`ENABLE REPLICA`的触发器只会在会话处于“replica”模式时触发，配置为`ENABLE ALWAYS`的触发器则无论当前复制模式为何都会触发。

`DISABLE/ENABLE [ REPLICA | ALWAYS ] RULE`

这些形式配置属于表的重写规则的触发。已禁用的规则仍为系统所知，但在查询重写期间不会应用。其语义与禁用或启用触发器时相同。对于`ON SELECT`规则会忽略此配置；为了使视图即使在当前会话处于非默认复制角色时也能正常工作，这类规则始终会被应用。

`CLUSTER`

该形式为将来的CLUSTER操作选择默认索引。它实际上不会对表重新聚簇。

`SET WITHOUT CLUSTER`

该形式从表中移除最近使用的CLUSTER索引规范。这会影响将来的聚簇操作（这些操作未指定索引）。

`SET WITH OIDS`

该形式向表添加一个`oid`系统列（参见第 5.4 节）。如果表已有 OID，则不执行任何操作。

请注意，这并不等同于`ADD COLUMN oid oid`；后者会添加一个碰巧命名为`oid`的普通列，而不是系统列。

`SET WITHOUT OIDS`

该形式从表中移除`oid`系统列。这完全等同于`DROP COLUMN oid RESTRICT`，但如果表中已经没有`oid`列，则不会报错。

`SET ( storage_parameter = value [, ... ] )`

该形式更改表的一个或多个存储参数。有关可用参数的详情，请参见存储参数。请注意，该命令不会立即修改表内容；根据参数的不同，可能需要重写表才能达到预期效果。可以使用VACUUM FULL、CLUSTER，或`ALTER TABLE`中会强制重写表的某种形式来完成重写。

> **注意**
>
> 虽然CREATE TABLE允许在WITH (*storage_parameter*)语法中指定OIDS，但ALTER TABLE不会将OIDS视为存储参数。应改用SET WITH OIDS和SET WITHOUT OIDS形式来更改 OID 状态。

RESET (*storage_parameter* [, ...])

该形式把一个或多个存储参数重置为默认值。与SET一样，可能仍需要进行表重写才能让整张表完全更新。

INHERIT *parent_table*

该形式将目标表添加为指定父表的新子表。之后，对父表执行的查询将包含目标表的记录。要作为子表添加，目标表必须已经包含与父表相同的所有列（也可以有额外的列）。列必须具有匹配的数据类型；如果父表中的列具有NOT NULL约束，则子表中的相应列也必须具有NOT NULL约束。

父表的所有 CHECK 约束还必须在子表中有匹配的约束。目前不考虑 UNIQUE、PRIMARY KEY 和 FOREIGN KEY 约束，但将来可能会改变。

NO INHERIT *parent_table*

该形式把目标表从指定父表的子表列表中移除。对父表的查询将不再包含来自目标表的记录。

OF *type_name*

该形式将表关联到一个复合类型，就像由CREATE TABLE OF创建该表一样。表的列名和类型列表必须与复合类型的完全匹配；是否存在oid系统列可以不同。该表不能继承任何其他表。这些限制确保CREATE TABLE OF允许等效的表定义。

NOT OF

该形式会解除类型化表与其类型之间的关联。

OWNER

该形式把表、序列或视图的所有者更改为指定用户。

SET TABLESPACE

这种形式把表的表空间更改为指定的表空间，并把与该表关联的数据文件移动到新表空间。表上的索引（如果有）不会被移动；但可以用额外的 SET TABLESPACE 命令单独移动。另见CREATE TABLESPACE。

RENAME

RENAME 形式更改表（或索引、序列或视图）的名称，或者更改表中某一列的名称。存储的数据不受影响。

SET SCHEMA

该形式把表移动到另一个模式。相关索引、约束以及由表列拥有的序列也会一并移动。

对单个表执行操作的 ALTER TABLE 所有形式，除了RENAME和SET SCHEMA之外，都可以组合成一个列表，一起应用多项更改。例如，可以在一条命令中添加多列和/或更改多列的类型。这对大表尤其有用，因为只需对表执行一次遍历。

要使用 ALTER TABLE，必须拥有该表。要更改表的模式，还必须在新模式上拥有 CREATE 权限。要将表作为父表的新子表添加，还必须拥有父表。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在表所在模式上拥有 CREATE 权限。（这些限制确保更改所有者不会执行任何无法通过删除并重新创建该表完成的操作。不过，超级用户无论如何都可以更改任何表的所有权。）

参数

name

要更改的现有表的名称（可带模式限定）。如果在表名之前指定ONLY，则只更改该表。如果未指定ONLY，则更改该表及其所有后代表（如果有）。也可以在表名后指定*，以明确表示包含后代表。

column

新列或现有列的名称。

new_column

现有列的新名称。

new_name

表的新名称。

type

新列的数据类型，或现有列的新数据类型。

table_constraint

表的新约束。

constraint_name

新约束或现有约束的名称。

CASCADE

自动删除依赖于被删除列或约束的对象（例如引用该列的视图）。

RESTRICT

如果存在任何依赖对象，则拒绝删除列或约束。这是默认行为。

trigger_name

要禁用或启用的单个触发器的名称。

ALL

禁用或启用属于表的所有触发器。（如果触发器中有内部生成的约束触发器，例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器，则需要超级用户权限。）

`USER`

禁用或启用属于表的所有触发器，但内部生成的约束触发器除外，例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器。

index_name

现有索引的名称。

storage_parameter

表存储参数的名称。

value

表存储参数的新值。根据参数的不同，这可以是数字或单词。

parent_table

要与该表关联或解除关联的父表。

new_owner

表的新所有者的用户名。

new_tablespace

表将被移动到的表空间名称。

new_schema

表将被移动到的模式名称。

注解

关键字`COLUMN`只是噪声，可以省略。

使用`ADD COLUMN`添加列时，表中的所有现有行都会使用该列的默认值初始化（如果未指定`DEFAULT`子句，则使用 NULL）。如果没有 DEFAULT 子句，这只是一次元数据更改，不需要立即更新表数据；添加的 NULL 值会在读取时提供。

使用 DEFAULT 子句添加列，或更改现有列的类型，都需要重写整个表及其索引。作为更改现有列类型时的例外，如果`USING`子句不改变列内容，并且旧类型要么可以二进制强制转换为新类型，要么是新类型之上的无约束域，则不需要重写表；但受影响列上的任何索引仍必须重建。添加或移除系统`oid`列也需要重写整个表。对于大型表，重建表和/或索引可能需要相当长的时间，并且会暂时需要最多两倍的磁盘空间。

添加`CHECK`或`NOT NULL`约束需要扫描表，以验证现有行满足约束，但不需要重写表。

允许在单条`ALTER TABLE`命令中指定多项更改，主要是因为这样可以将多次表扫描或重写合并为一次遍历。

`DROP COLUMN`形式不会从物理上删除列，而只是使其对 SQL 操作不可见。表后续的插入和更新操作会为该列存储空值。因此，删除列的速度很快，但不会立即减少表在磁盘上的大小，因为被删除列占用的空间不会被回收。随着现有行被更新，空间会逐渐回收。（删除系统`oid`列时不适用这些说明；删除该列会立即重写表。）

若要强制立即回收已删除列所占的空间，可以执行任何一种会导致整表重写的ALTER TABLE形式。这样会重建每一行，并用空值替换被删除的列。

会重写表的ALTER TABLE形式对于 MVCC 来说并不安全。表重写完成后，如果并发事务使用的是在重写发生之前取得的快照，那么该表在这些并发事务看来会像一张空表。详见第 13.5 节。

SET DATA TYPE的USING选项实际上可以指定任何涉及该行旧值的表达式；也就是说，它既可以引用正在转换的列，也可以引用其他列。这使得使用SET DATA TYPE语法完成非常通用的转换成为可能。正因为这种灵活性，USING表达式不会应用到列的默认值（如果有）上，
因为其结果可能不是默认值所要求的常量表达式。这意味着，
当从旧类型到新类型不存在隐式或赋值类型转换时，即便提供了USING子句，SET DATA TYPE也可能仍然无法转换默认值。在这种情况下，可以先用DROP DEFAULT删除默认值，执行ALTER TYPE，然后再用SET DEFAULT添加一个合适的新默认值。类似的考虑也适用于涉及该列的索引和约束。

如果表有任何后代表，则不允许只在父表中添加、重命名或更改列的类型，
或重命名继承的约束，而不对后代表执行相同操作。也就是说，ALTER TABLE ONLY会被拒绝。这确保后代表始终拥有与父表匹配的列。

只有当某个后代表中的列既不是从其他父表继承而来，也从未有过该列的独立定义时，递归的DROP COLUMN操作才会移除该后代表中的此列。非递归的DROP COLUMN（即ALTER TABLE ONLY ... DROP COLUMN）永远不会移除任何后代列，而只会把它们标记为独立定义，而非继承得到。

TRIGGER、CLUSTER、OWNER 和 TABLESPACE 操作永远不会递归到后代表；也就是说，它们的行为始终如同指定了 ONLY。添加约束时，只有 CHECK 约束可以递归，而且对这类约束必须递归。

不允许更改系统目录表的任何部分。

有关有效参数的进一步说明，请参见CREATE TABLE。第 5 章中还有关于继承的更多信息。

示例

要添加类型为varchar的列到表中：

```
ALTER TABLE distributors ADD COLUMN address varchar(30);
```

要从表中删除一列：

```
ALTER TABLE distributors DROP COLUMN address RESTRICT;
```

要在一个操作中更改两个现有列的类型：

```
ALTER TABLE distributors
    ALTER COLUMN address TYPE varchar(80),
    ALTER COLUMN name TYPE varchar(100);
```

要把一个包含 Unix 时间戳的整数列改为 timestamp with time zone，并通过USING子句完成转换：

```
ALTER TABLE foo
    ALTER COLUMN foo_timestamp SET DATA TYPE timestamp with time zone
    USING
        timestamp with time zone 'epoch' + foo_timestamp * interval '1
 second';
```

如果该列带有一个不能自动转换为新数据类型的默认值表达式，也是同样的做法：

```
ALTER TABLE foo
    ALTER COLUMN foo_timestamp DROP DEFAULT,
    ALTER COLUMN foo_timestamp TYPE timestamp with time zone
    USING
        timestamp with time zone 'epoch' + foo_timestamp * interval '1
 second',
    ALTER COLUMN foo_timestamp SET DEFAULT now();
```

要重命名一个现有列：

```
ALTER TABLE distributors RENAME COLUMN address TO city;
```

重命名一个现有的表：

```
ALTER TABLE distributors RENAME TO suppliers;
```

为一列增加一个非空约束：

```
ALTER TABLE distributors ALTER COLUMN street SET NOT NULL;
```

从一列移除一个非空约束：

```
ALTER TABLE distributors ALTER COLUMN street DROP NOT NULL;
```

要向一个表及其所有后代添加一个检查约束：

```
ALTER TABLE distributors ADD CONSTRAINT zipchk CHECK (char_length(
zipcode) = 5);
```

要从一个表及其所有后代移除一个检查约束：

```
ALTER TABLE distributors DROP CONSTRAINT zipchk;
```

只从一个表移除一个检查约束：

```
ALTER TABLE ONLY distributors DROP CONSTRAINT zipchk;
```

（该检查约束在所有子表上仍然保留。）

为一个表增加一个外键约束：

```
ALTER TABLE distributors ADD CONSTRAINT distfk FOREIGN KEY (address)
 REFERENCES addresses (address);
```

为一个表增加一个（多列）唯一约束：

```
ALTER TABLE distributors ADD CONSTRAINT dist_id_zipcode_key UNIQUE (
dist_id, zipcode);
```

为一个表增加一个自动命名的主键约束，注意一个表只能拥有一个主键：

```
ALTER TABLE distributors ADD PRIMARY KEY (dist_id);
```

把一个表移动到一个不同的表空间：

```
ALTER TABLE distributors SET TABLESPACE fasttablespace;
```

把一个表移动到一个不同的模式：

```
ALTER TABLE myschema.distributors SET SCHEMA yourschema;
```

重建一个主键约束，并且在重建索引期间不阻塞更新：

```
CREATE UNIQUE INDEX CONCURRENTLY dist_id_temp_idx ON distributors (
dist_id);
ALTER TABLE distributors DROP CONSTRAINT distributors_pkey,
    ADD CONSTRAINT distributors_pkey PRIMARY KEY USING INDEX dist_id_
temp_idx;
```

兼容性

以下形式符合 SQL 标准：`ADD`（不带`USING INDEX`）、`DROP`、`SET DEFAULT`以及`SET DATA TYPE`（不带`USING`）。其他形式是PostgreSQL对 SQL 标准的扩展。此外，在单条`ALTER TABLE`命令中指定多个操作的能力也是一种扩展。

`ALTER TABLE DROP COLUMN`可以被用来删除一个表的唯一的列，从而留下一个零列的表。这是一种 SQL 的扩展，SQL 中不允许零列的表。

另见

CREATE TABLE

ALTER TABLESPACE

ALTER TABLESPACE — change the definition of a tablespace

大纲

```
ALTER TABLESPACE name RENAME TO new_name
ALTER TABLESPACE name OWNER TO new_owner
ALTER TABLESPACE name SET ( tablespace_option = value [, ... ] )
ALTER TABLESPACE name RESET ( tablespace_option [, ... ] )
```

描述

`ALTER TABLESPACE`可以用于更改表空间的定义。

要使用`ALTER TABLESPACE`，必须拥有该表空间。要更改所有者，还必须是新所有者角色的直接或间接成员。（请注意，超级用户会自动拥有这些权限。）

参数

name

现有表空间的名称。

new_name

该表空间的新名称。新名称不能以`pg_`开头，因为这类名称保留给系统表空间使用。

new_owner

该表空间的新拥有者。

tablespace_parameter

要设置或重置的表空间参数。目前可用的参数只有`seq_page_cost`和`random_page_cost`。为特定表空间设置其中任一值，
会覆盖规划器对读取该表空间中表的数据页成本的通常估计；
该估计由同名配置参数确定（参见seq_page_cost、random_page_cost）。
如果一个表空间位于比 I/O 子系统其余部分更快或更慢的磁盘上，这可能很有用。

示例

将表空间`index_space`重命名为`fast_raid`：

```
ALTER TABLESPACE index_space RENAME TO fast_raid;
```

更改表空间`index_space`的拥有者：

```
ALTER TABLESPACE index_space OWNER TO mary;
```

兼容性

在 SQL 标准中没有 `ALTER TABLESPACE`语句。

另见

CREATE TABLESPACE, DROP TABLESPACE

ALTER TEXT SEARCH CONFIGURATION

ALTER TEXT SEARCH CONFIGURATION — change the definition of a text search configuration

大纲

```
ALTER TEXT SEARCH CONFIGURATION name
    ADD MAPPING FOR token_type [, ... ] WITH dictionary_name [, ... ]
ALTER TEXT SEARCH CONFIGURATION name
    ALTER MAPPING FOR token_type [, ... ] WITH dictionary_name [, ...
 ]
ALTER TEXT SEARCH CONFIGURATION name
    ALTER MAPPING REPLACE old_dictionary WITH new_dictionary
ALTER TEXT SEARCH CONFIGURATION name
    ALTER MAPPING FOR token_type [, ... ] REPLACE old_dictionary
 WITH new_dictionary
ALTER TEXT SEARCH CONFIGURATION name
    DROP MAPPING [ IF EXISTS ] FOR token_type [, ... ]
ALTER TEXT SEARCH CONFIGURATION name RENAME TO new_name
ALTER TEXT SEARCH CONFIGURATION name OWNER TO new_owner
ALTER TEXT SEARCH CONFIGURATION name SET SCHEMA new_schema
```

描述

`ALTER TEXT SEARCH CONFIGURATION` 更改一个文本搜索配置的定义。你可以修改其从记号类型到字典的映射，或者更改该配置的名称或拥有者。

要使用`ALTER TEXT SEARCH CONFIGURATION`，你必须是该配置的拥有者。

参数

name

一个现有文本搜索配置的名称（可以是模式限定的）。

token_type

由该配置的解析器发出的记号类型的名称。

dictionary_name

用于指定记号类型的文本搜索字典名称。如果列出了多个字典，
将按指定顺序依次查阅它们。

old_dictionary

在映射中要替换的文本搜索字典的名称。

new_dictionary

被用来替代*old_dictionary* 的文本搜索字典的名称。

new_name

该文本搜索配置的新名称。

new_owner

该文本搜索配置的新拥有者。

new_schema

该文本搜索配置的新模式。

`ADD MAPPING FOR`形式为指定的词元类型安装要查询的字典列表；如果任何词元类型已经存在映射，则会报错。`ALTER MAPPING FOR`形式执行相同操作，但会先移除这些词元类型的现有映射。`ALTER MAPPING REPLACE`形式在任何出现旧字典的地方用*new_dictionary*替换*old_dictionary*。当出现`FOR`时，只对指定的词元类型执行此操作；不出现时，
则对该配置的所有映射执行。`DROP MAPPING`形式移除指定词元类型的所有字典，使这些类型的词元被文本搜索配置忽略。如果词元类型没有映射，则会报错，除非出现`IF EXISTS`。

示例

以下示例在`my_config`中使用`english`的任何地方，都将`english`字典替换为`swedish`字典。

```
ALTER TEXT SEARCH CONFIGURATION my_config
  ALTER MAPPING REPLACE english WITH swedish;
```

兼容性

在 SQL 标准中没有 `ALTER TEXT SEARCH CONFIGURATION` 语句。

另见

CREATE TEXT SEARCH CONFIGURATION, DROP TEXT SEARCH CONFIGURATION

ALTER TEXT SEARCH DICTIONARY

ALTER TEXT SEARCH DICTIONARY — change the definition of a text search dictionary

大纲

```
ALTER TEXT SEARCH DICTIONARY name (
    option [ = value ] [, ... ]
)
ALTER TEXT SEARCH DICTIONARY name RENAME TO new_name
ALTER TEXT SEARCH DICTIONARY name OWNER TO new_owner
ALTER TEXT SEARCH DICTIONARY name SET SCHEMA new_schema
```

描述

`ALTER TEXT SEARCH DICTIONARY`更改文本搜索字典的定义。你可以更改该字典的模板相关选项，也可以更改该字典的名称或拥有者。

要使用`ALTER TEXT SEARCH DICTIONARY`，你必须是该字典的拥有者。

参数

name

一个现有文本搜索字典的名称（可以是模式限定的）。

option

要为此字典设置的模板相关选项的名称。

value

模板相关选项要使用的新值。如果省略等号和值，则会从该字典中移除该选项之前的设置，从而允许使用默认值。

new_name

该文本搜索字典的新名称。

new_owner

该文本搜索字典的新拥有者。

new_schema

该文本搜索字典的新模式。

模板相关选项可以以任意顺序出现。

示例

下面的示例命令更改了一个基于 Snowball 的字典的停用词列表。其他参数保持不变。

```
ALTER TEXT SEARCH DICTIONARY my_dict ( StopWords = newrussian );
```

下面的示例命令将语言选项更改为`dutch`，并完全移除了停用词选项。

```
ALTER TEXT SEARCH DICTIONARY my_dict ( language = dutch, StopWords );
```

下面的示例命令“更新”了该字典的定义，但实际上并没有做任何更改。

```
ALTER TEXT SEARCH DICTIONARY my_dict ( dummy );
```

（之所以可行，是因为选项移除代码在不存在该选项时也不会报错。）
这种技巧在修改该字典的配置文件时很有用：`ALTER` 会强制现有数据库会话重新读取配置文件，而如果它们先前已经读取过这些文件，本来是不会再次读取的。

兼容性

在 SQL 标准中没有 `ALTER TEXT SEARCH DICTIONARY`语句。

另见

CREATE TEXT SEARCH DICTIONARY, DROP TEXT SEARCH DICTIONARY

ALTER TEXT SEARCH PARSER

ALTER TEXT SEARCH PARSER — change the definition of a text search parser

大纲

```
ALTER TEXT SEARCH PARSER name RENAME TO new_name
ALTER TEXT SEARCH PARSER name SET SCHEMA new_schema
```

描述

`ALTER TEXT SEARCH PARSER`更改全文检索解析器的定义。当前唯一支持的功能是更改该解析器的名称。

要使用`ALTER TEXT SEARCH PARSER`，你必须是超级用户。

参数

name

一个现有全文检索解析器的名称（可以是模式限定的）。

new_name

该全文检索解析器的新名称。

new_schema

该全文检索解析器的新模式。

兼容性

SQL 标准中没有`ALTER TEXT SEARCH PARSER`语句。

另见

CREATE TEXT SEARCH PARSER, DROP TEXT SEARCH PARSER

ALTER TEXT SEARCH TEMPLATE

ALTER TEXT SEARCH TEMPLATE — change the definition of a text search template

大纲

```
ALTER TEXT SEARCH TEMPLATE name RENAME TO new_name
ALTER TEXT SEARCH TEMPLATE name SET SCHEMA new_schema
```

描述

`ALTER TEXT SEARCH TEMPLATE`更改文本搜索模板的定义。当前唯一支持的功能是更改该模板的名称。

要使用 `ALTER TEXT SEARCH TEMPLATE`，你必须是超级用户。

参数

name

一个现有文本搜索模板的名称（可以是模式限定的）。

new_name

该文本搜索模板的新名称。

new_schema

该文本搜索模板的新模式。

兼容性

SQL 标准中没有 `ALTER TEXT SEARCH TEMPLATE` 语句。

另见

CREATE TEXT SEARCH TEMPLATE, DROP TEXT SEARCH TEMPLATE

ALTER TRIGGER

ALTER TRIGGER — change the definition of a trigger

大纲

```
ALTER TRIGGER name ON table RENAME TO new_name
```

描述

`ALTER TRIGGER`更改现有触发器的属性。`RENAME`子句更改给定触发器的名称，但不以其他方式更改触发器定义。

若要更改触发器的属性，你必须拥有该触发器所作用的表。

参数

name

要修改的现有触发器的名称。

table

该触发器所作用的表的名称。

new_name

该触发器的新名称。

注解

临时启用或禁用触发器的能力由ALTER TABLE提供，而不是由`ALTER TRIGGER`提供，因为`ALTER TRIGGER`没有便捷的方式来表达一次启用或禁用表中所有触发器的选项。

示例

要重命名一个现有触发器：

```
ALTER TRIGGER emp_stamp ON emp RENAME TO emp_track_chgs;
```

兼容性

`ALTER TRIGGER`是 PostgreSQL 对 SQL 标准的扩展。

参见

ALTER TABLE

ALTER TYPE

ALTER TYPE — change the definition of a type

大纲

```
ALTER TYPE name action [, ... ]
ALTER TYPE name OWNER TO new_owner
ALTER TYPE name RENAME ATTRIBUTE attribute_name TO new_attribute_name
 [ CASCADE | RESTRICT ]
ALTER TYPE name RENAME TO new_name
ALTER TYPE name SET SCHEMA new_schema
ALTER TYPE name ADD VALUE new_enum_value [ { BEFORE | AFTER }
 existing_enum_value ]

where action is one of:

    ADD ATTRIBUTE attribute_name data_type [ COLLATE collation ] [
 CASCADE | RESTRICT ]
    DROP ATTRIBUTE [ IF EXISTS ] attribute_name [ CASCADE | RESTRICT ]
    ALTER ATTRIBUTE attribute_name [ SET DATA ] TYPE data_type [
 COLLATE collation ] [ CASCADE | RESTRICT ]
```

描述

`ALTER TYPE`更改现有类型的定义。有几种子形式：

`ADD ATTRIBUTE`

该形式使用与CREATE TYPE相同的语法向复合类型添加新属性。

`DROP ATTRIBUTE [ IF EXISTS ]`

这种形式从复合类型中删除一个属性。如果指定了`IF EXISTS`而该属性不存在，则不会抛出错误，而是发出一条提示。

`SET DATA TYPE`

这种形式更改复合类型中某个属性的数据类型。

`OWNER`

这种形式更改类型的拥有者。

`RENAME`

该形式更改类型的名称或复合类型中单个属性的名称。

`SET SCHEMA`

这种形式将类型移动到另一个模式中。

`ADD VALUE [ BEFORE | AFTER ]`

这种形式向枚举类型中添加一个新值。如果没有使用 `BEFORE` 或 `AFTER` 指定新值在枚举排序中的位置，则新项会被放在值列表的末尾。

`CASCADE`

自动将该操作传播到正在修改的类型的类型化表及其后代。

`RESTRICT`

如果正在修改的类型是某个类型化表的类型，则拒绝该操作。这是默认行为。

`ADD ATTRIBUTE`、`DROP ATTRIBUTE` 和 `ALTER ATTRIBUTE` 操作可以组合成一个包含多项修改的列表，以便并行应用。例如，可以在一条命令中添加多个属性和/或更改多个属性的数据类型。

要使用 `ALTER TYPE`，必须拥有该类型。要更改类型的模式，还必须在新模式上拥有 `CREATE` 权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在类型所在模式上拥有 `CREATE` 权限。（这些限制确保更改所有者不会执行任何无法通过删除并重新创建该类型完成的操作。不过，超级用户无论如何都可以更改任何类型的所有权。）

参数

name

要修改的现有类型的名称（可以是模式限定的）。

new_name

该类型的新名称。

new_owner

该类型新的拥有者的用户名。

new_schema

该类型的新模式。

attribute_name

要添加、修改或删除的属性名称。

new_attribute_name

要重命名的属性的新名称。

data_type

要添加的属性的数据类型，或者要修改的属性的新类型。

new_enum_value

要添加到枚举类型值列表中的新值。与所有枚举字面量一样，它必须加引号。

existing_enum_value

一个现有枚举值；在枚举类型的排序顺序中，新值会紧邻该值，被添加到它之前或之后。与所有枚举字面量一样，它必须加引号。

注解

`ALTER TYPE ... ADD VALUE`（向枚举类型添加新值的形式）不能在事务块内执行。

涉及新增枚举值的比较有时会比只涉及枚举类型原有成员的比较慢。通常只有在使用BEFORE或AFTER将新值的排序位置设置在列表末尾以外时，才会发生这种情况。但是，即使新值添加在末尾，有时也会发生这种情况（如果自枚举类型最初创建以来 OID 计数器发生了“回绕”）。速度下降通常并不明显；但如果这确实重要，可以通过删除并重新创建枚举类型，或转储并重新加载数据库，恢复最佳性能。

示例

要重命名一个数据类型：

```
ALTER TYPE electronic_mail RENAME TO email;
```

要将类型email的拥有者改为 joe：

```
ALTER TYPE email OWNER TO joe;
```

要将类型email所在的模式改为 customers：

```
ALTER TYPE email SET SCHEMA customers;
```

向类型添加新属性：

```
ALTER TYPE compfoo ADD ATTRIBUTE f3 int;
```

要在枚举类型的特定排序位置添加一个新值：

```
ALTER TYPE colors ADD VALUE 'orange' AFTER 'red';
```

兼容性

添加和删除属性的这些变体属于 SQL 标准；其他变体则是 PostgreSQL 扩展。

另见

CREATE TYPE, DROP TYPE

ALTER USER

ALTER USER — change a database role

大纲

```
ALTER USER name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | REPLICATION | NOREPLICATION
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'

ALTER USER name RENAME TO new_name

ALTER USER name SET configuration_parameter { TO | = } { value |
 DEFAULT }
ALTER USER name SET configuration_parameter FROM CURRENT
ALTER USER name RESET configuration_parameter
ALTER USER name RESET ALL
```

描述

`ALTER USER`现为ALTER ROLE的别名。

兼容性

`ALTER USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义交由具体实现决定。

另见

ALTER ROLE

ALTER USER MAPPING

ALTER USER MAPPING — change the definition of a user mapping

大纲

```
ALTER USER MAPPING FOR { user_name | USER | CURRENT_USER | PUBLIC }
    SERVER server_name
    OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] )
```

描述

`ALTER USER MAPPING`更改用户映射的定义。

外部服务器的拥有者可以为该服务器上的任何用户修改用户映射。此外，
如果某个用户已被授予该服务器上的`USAGE`权限，
那么该用户也可以修改其自己用户名对应的用户映射。

参数

user_name

映射的用户名。`CURRENT_USER`和`USER`与当前用户名称匹配。`PUBLIC`
用于匹配系统中当前及未来的所有用户名。

server_name

该用户映射所属服务器的名称。

`OPTIONS ( [ ADD | SET | DROP ] option ['value'] [, ... ] )`

更改该用户映射的选项。新选项会覆盖先前指定的任何选项。`ADD`、`SET`和`DROP` 指定要执行的动作。如果未显式指定操作，则假定为`ADD`。选项名必须唯一；
这些选项还会由该服务器的外部数据包装器进行验证。

示例

更改用户映射`bob`、服务器`foo`的密码：

```
ALTER USER MAPPING FOR bob SERVER foo OPTIONS (SET password 'public');
```

兼容性

`ALTER USER MAPPING`符合 ISO/IEC 9075-9 (SQL/MED)。这里有一个细微的语法问题：该标准省略了`FOR` 关键字。由于`CREATE USER MAPPING`和 `DROP USER MAPPING`都在类似位置使用`FOR`，而 IBM DB2（另一个主要的 SQL/MED 实现）也要求在 `ALTER USER MAPPING`中使用它，因此为了保持一致性和互操作性，PostgreSQL 在这里偏离了标准。

另见

CREATE USER MAPPING, DROP USER MAPPING

ALTER VIEW

ALTER VIEW — change the definition of a view

大纲

```
ALTER VIEW name ALTER [ COLUMN ] column SET DEFAULT expression
ALTER VIEW name ALTER [ COLUMN ] column DROP DEFAULT
ALTER VIEW name OWNER TO new_owner
ALTER VIEW name RENAME TO new_name
ALTER VIEW name SET SCHEMA new_schema
```

描述

`ALTER VIEW`更改视图的各种辅助属性。（如果要修改视图的定义查询，请使用 `CREATE OR REPLACE VIEW`。）

要使用`ALTER VIEW`，必须拥有该视图。要更改视图的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在视图所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建视图完成的操作。不过，
超级用户无论如何都可以更改任何视图的所有权。）

参数

name

现有视图的名称（可以带模式限定）。

`SET/DROP DEFAULT`

这些形式为列设置或移除默认值。如果 `INSERT` 语句没有为某一列指定值，那么在应用该视图的 `ON INSERT` 规则之前，与该视图列关联的默认值会被插入到视图上的 `INSERT` 语句中。

new_owner

视图新拥有者的用户名。

new_name

该视图的新名称。

new_schema

该视图的新模式。

注解

出于历史原因，视图也可以使用 `ALTER TABLE`；但允许用于视图的 `ALTER TABLE` 变体，仅限于与上面所示形式等价的那些。

示例

将视图 `foo` 重命名为 `bar`：

```
ALTER VIEW foo RENAME TO bar;
```

兼容性

`ALTER VIEW` 是 PostgreSQL 对 SQL 标准的扩展。

另见

CREATE VIEW, DROP VIEW

ANALYZE

ANALYZE — collect statistics about a database

大纲

```
ANALYZE [ VERBOSE ] [ table [ ( column [, ...] ) ] ]
```

描述

ANALYZE收集数据库中各表内容的统计信息，并将结果存储到pg_statistic 系统目录中。随后，查询规划器会使用这些统计信息来帮助确定查询的最高效执行计划。

如果不带参数，ANALYZE 会检查当前数据库中的每个表。带有参数时，ANALYZE 只检查该表。还可以提供列名列表，这时只会收集这些列的统计信息。

参数

VERBOSE

启用进度消息显示。

table

要分析的特定表的名称（可带模式限定）。默认为当前数据库中的所有表。

column

要分析的特定列的名称。默认为所有列。

输出

指定VERBOSE时，ANALYZE会输出进度消息，指示当前正在处理哪个表，
同时还会打印这些表的各种统计信息。

注解

在默认的 PostgreSQL 配置中，自动清理守护进程（见第 23.1.5 节）会在表首次装载数据时，以及在常规运行过程中数据发生变化时，自动分析这些表。当自动清理被禁用时，最好定期运行ANALYZE，或者在对表内容做出重大更改之后立即运行。
准确的统计信息有助于规划器选择最合适的查询计划，从而提高查询处理速度。
一个常见的策略是在每天使用率较低的时段运行一次VACUUM和 ANALYZE。

ANALYZE只需获取目标表上的读锁，因此可以与该表上的其他活动并行运行。

ANALYZE收集的统计信息通常包括每列中某些高频值的列表，
以及显示每列大致数据分布的直方图。如果ANALYZE认为它们没有意义（例如，
唯一键列中没有高频值），或者列的数据类型不支持相应的操作符，
则可能省略其中一项或两项。更多统计信息见第 23 章。

对于大型表，ANALYZE会对表内容进行随机采样，而不是检查每一行。这使得即使是很大的表，也能在较短时间内完成分析。不过要注意，这些统计信息只是近似值，

并且即使实际表内容没有变化，每次运行ANALYZE时统计信息也会略有变化。
这可能导致EXPLAIN中显示的规划器估算代价略有变化。在少数情况下，
这种非确定性会导致规划器在运行ANALYZE之后改用不同的查询计划。要避免这种情况，
可以按下文所述提高ANALYZE收集的统计信息量。

可以通过调整配置变量default_statistics_target来控制分析程度，也可以针对单个列使用ALTER TABLE ... ALTER COLUMN ... SET STATISTICS设置每列的统计信息目标（参见ALTER TABLE）。目标值会设置高频值列表中的最大条目数，以及直方图中的最大桶数。默认目标值是 100，但可以把它调高或调低，以在规划器估算精度、ANALYZE耗费的时间以及pg_statistic占用的空间之间作出权衡。特别地，
将统计信息目标设置为零会禁用该列的统计信息收集。对于那些从不出现在查询 WHERE、GROUP BY或ORDER BY子句中的列，这样做可能很有用，因为规划器不会用到这些列的统计信息。

被分析列中最大的统计信息目标决定了为了生成统计信息而需要采样的表行数。
增加该目标会导致执行ANALYZE所需的时间和空间按比例增加。

ANALYZE估算的值之一是每一列中出现的非重复值数量。由于只检查了部分行，
即使使用可能的最大统计信息目标，这种估计有时也可能相当不精确。
如果这种不精确导致查询计划不佳，就可以手工确定一个更精确的值，然后用 ALTER TABLE ... ALTER COLUMN ... SET (n_distinct = ...) 把该值设置进去（参见ALTER TABLE）。

如果被分析的表有一个或多个子表，ANALYZE 会两次收集统计信息：一次只针对父表中的行，另一次针对父表及其所有子表中的行。不过，
自动清理守护进程在决定是否为该表触发自动分析时，只会考虑对父表本身的插入或更新。
如果该表很少被插入或更新，那么除非手工运行 ANALYZE，否则继承统计信息就不会保持最新。

兼容性

SQL 标准中没有ANALYZE语句。

另见

VACUUM, vacuumdb, 第 18.4.3 节, 第 23.1.5 节

BEGIN

BEGIN — start a transaction block

大纲

```
BEGIN [ WORK | TRANSACTION ] [ transaction_mode [, ...] ]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
    [ NOT ] DEFERRABLE
```

描述

`BEGIN`启动一个事务块，也就是说，在 `BEGIN`命令之后的所有语句都会在同一个事务中执行，直到显式给出COMMIT或ROLLBACK为止。默认情况下（不使用`BEGIN`），PostgreSQL以“自动提交”模式执行事务，也就是说，每条语句都在自己的事务中执行，并在语句结束时隐式提交（如果执行成功，否则回滚）。

在事务块中执行语句通常会更快，因为事务的启动和提交需要大量的 CPU 和磁盘活动。在进行多个相关更改时，把多条语句放在一个事务中执行也有助于保证一致性：其他会话将无法看到那些相关更新尚未全部完成时的中间状态。

如果指定了隔离级别、读写模式或可延迟模式，新事务就会具有这些特征，就像执行了SET TRANSACTION一样。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

关于本语句其他参数含义的信息，参见SET TRANSACTION。

注解

START TRANSACTION与`BEGIN`具有相同功能。

使用COMMIT或ROLLBACK来结束事务块。

如果在已经处于事务块中时发出`BEGIN`，将会产生一条警告消息。事务状态不会受影响。若要在事务块中实现事务嵌套，请使用保存点（参见SAVEPOINT）。

出于向后兼容的考虑，连续多个*transaction_modes*之间的逗号可以省略。

示例

开始一个事务块：

```
BEGIN;
```

兼容性

`BEGIN`是PostgreSQL的一种语言扩展。它等价于 SQL 标准命令START TRANSACTION，后者的参考页中包含更多兼容性信息。

`DEFERRABLE` *`transaction_mode`* 是PostgreSQL的一种语言扩展。

另外，在嵌入式 SQL 中，关键字`BEGIN`用于不同的目的。在移植数据库应用程序时，应当谨慎处理事务语义。

另见

COMMIT, ROLLBACK, START TRANSACTION, SAVEPOINT

CHECKPOINT

CHECKPOINT — force a transaction log checkpoint

大纲

```
CHECKPOINT
```

描述

预写式日志（WAL）会每隔一段时间在事务日志中设置一个检查点。（要调整自动检查点间隔，请参见运行时配置选项checkpoint_segments和checkpoint_timeout。）`CHECKPOINT` 命令会在发出该命令时强制立即执行检查点，而不等待预定的检查点。

检查点是事务日志序列中的一个位置，到这一位置时，所有数据文件都已更新，以反映日志中的信息。所有数据文件都会被刷出到磁盘。有关 WAL 系统的更多信息，参见第 29 章。

如果在恢复期间执行 `CHECKPOINT` 命令，它将强制执行一个重启点，而不是写入一个新的检查点。

只有超级用户才能调用 `CHECKPOINT`。该命令并非设计用于正常运行期间。

兼容性

`CHECKPOINT`命令是 PostgreSQL的一种语言扩展。

CLOSE

CLOSE — close a cursor

大纲

```
CLOSE { name | ALL }
```

描述

CLOSE释放与一个打开的游标相关联的资源。游标关闭后，不允许再对其执行任何后续操作。当不再需要游标时，应将其关闭。

每个不可保持的打开游标都会在事务通过COMMIT或 ROLLBACK结束时被隐式关闭。可保持游标会在创建它的事务通过ROLLBACK中止时被隐式关闭。如果创建它的事务成功提交，则该可保持游标会一直保持打开状态，直到显式执行 CLOSE，或者客户端断开连接。

参数

name

要关闭的打开游标的名称。

ALL

关闭所有打开的游标。

注解

PostgreSQL没有显式的OPEN游标语句；游标在声明时即被视为打开。请使用DECLARE语句来声明游标。

可以通过查询pg_cursors 系统视图查看所有可用游标。

如果在某个保存点之后关闭了游标，而该保存点后来又被回滚，则CLOSE不会被回滚；也就是说，该游标仍然保持关闭状态。

示例

关闭游标liahona：

```
CLOSE liahona;
```

兼容性

CLOSE完全符合 SQL 标准。CLOSE ALL是PostgreSQL的一种扩展。

另见

DECLARE, FETCH, MOVE

CLUSTER

CLUSTER — cluster a table according to an index

大纲

```
CLUSTER [VERBOSE] table_name [ USING index_name ]
CLUSTER [VERBOSE]
```

描述

`CLUSTER` 指示 PostgreSQL 按照 *`index_name`* 指定的索引，对 *`table_name`* 指定的表进行聚簇。该索引必须已经定义在 *`table_name`* 上。

当一个表被聚簇时，它会根据索引信息在物理上重新排序。
聚簇是一次性操作：之后如果表再被更新，这些更改不会再次被聚簇。也就是说，
系统不会试图按照索引顺序存储新行或更新后的行。（如果需要，
可以定期再次执行该命令来重新聚簇。此外，将表的 `FILLFACTOR` 存储参数设置为小于 100%，有助于在更新期间保持聚簇顺序，因为如果页面上有足够空间，更新后的行会保留在同一页面中。）

当一个表被聚簇时，PostgreSQL 会记住该表是按哪个索引聚簇的。形式 `CLUSTER` *`table_name`* 会使用与之前相同的索引对表重新聚簇。你也可以使用ALTER TABLE的 `CLUSTER` 或 `SET WITHOUT CLUSTER` 形式，设置未来聚簇操作要使用的索引，或者清除任何先前的设置。

不带任何参数的 `CLUSTER` 会对当前数据库中所有此前已聚簇且属于调用用户的表重新执行聚簇；如果由超级用户调用，则会对所有此前已聚簇的表重新执行聚簇。这种形式的 `CLUSTER` 不能在事务块内执行。

当对一个表执行聚簇时，会在该表上获取 `ACCESS EXCLUSIVE` 锁。这会阻止任何其他数据库操作（包括读和写）在 `CLUSTER` 完成前访问该表。

参数

`table_name`

表的名称（可能是模式限定的）。

`index_name`

索引的名称。

`VERBOSE`

在每个表被聚簇时打印进度报告。

注解

当你在表中随机访问单行时，数据在表中的实际顺序并不重要。但是，
如果访问某些数据比其他数据更频繁，并且存在将这些数据放在一起的索引，使用`CLUSTER`
就会有益。如果要从表中获取一段索引值范围，或者获取有多行匹配的单个索引值，`CLUSTER`

会有帮助，因为一旦索引确定第一个匹配行所在的表页，其他匹配行很可能已在同一个表页中，从而减少磁盘访问并加快查询。

`CLUSTER` 可以通过在指定索引上进行索引扫描，或者（如果该索引是 B-树）先执行顺序扫描再排序，来对表重新排序。它会根据规划器代价参数和可用的统计信息，尝试选择速度更快的方法。

使用索引扫描时，会创建一个表的临时副本，其中表数据按索引顺序排列。还会为表上的每个索引创建临时副本。因此，所需的磁盘空闲空间至少应等于表大小与索引大小之和。

使用顺序扫描加排序时，还会创建临时排序文件，因此临时空间需求峰值可达表大小的两倍再加上索引大小。这种方法通常比索引扫描更快，但如果无法接受其磁盘空间需求，可以临时将enable_sort设置为`off`来禁用这种选择。

建议在聚簇之前将maintenance_work_mem设置为一个合理的大值（但不超过可专门用于`CLUSTER`操作的内存量）。

因为规划器会记录有关表中数据顺序的统计信息，建议在新近聚簇过的表上运行ANALYZE。否则，规划器可能会选择很差的查询计划。

因为 `CLUSTER` 会记住哪些索引已被设为聚簇索引，你可以第一次先手工聚簇需要聚簇的表，然后设置一个定期运行的维护脚本，执行不带任何参数的 `CLUSTER`，这样这些表就会被周期性地重新聚簇。

示例

按照索引 `employees_ind` 对表 `employees` 进行聚簇：

```
CLUSTER employees USING employees_ind;
```

使用之前用过的同一个索引对 `employees` 表进行聚簇：

```
CLUSTER employees;
```

对数据库中此前已聚簇过的所有表执行聚簇：

```
CLUSTER;
```

兼容性

SQL 标准中没有 `CLUSTER` 语句。

语法

```
CLUSTER index_name ON table_name
```

也受支持，以兼容 8.3 之前的 PostgreSQL 版本。

另见

clusterdb

COMMENT

COMMENT — define or change the comment of an object

大纲

```
COMMENT ON
{
  AGGREGATE agg_name (agg_type [, ...] ) |
  CAST (source_type AS target_type) |
  COLLATION object_name |
  COLUMN relation_name.column_name |
  CONSTRAINT constraint_name ON table_name |
  CONVERSION object_name |
  DATABASE object_name |
  DOMAIN object_name |
  EXTENSION object_name |
  FOREIGN DATA WRAPPER object_name |
  FOREIGN TABLE object_name |
  FUNCTION function_name ( [ [ argmode ] [ argname ] argtype [, ...] ]
 ) |
  INDEX object_name |
  LARGE OBJECT large_object_oid |
  OPERATOR operator_name (left_type, right_type) |
  OPERATOR CLASS object_name USING index_method |
  OPERATOR FAMILY object_name USING index_method |
  [ PROCEDURAL ] LANGUAGE object_name |
  ROLE object_name |
  RULE rule_name ON table_name |
  SCHEMA object_name |
  SEQUENCE object_name |
  SERVER object_name |
  TABLE object_name |
  TABLESPACE object_name |
  TEXT SEARCH CONFIGURATION object_name |
  TEXT SEARCH DICTIONARY object_name |
  TEXT SEARCH PARSER object_name |
  TEXT SEARCH TEMPLATE object_name |
  TRIGGER trigger_name ON table_name |
  TYPE object_name |
  VIEW object_name
} IS 'text'
```

描述

COMMENT存储数据库对象的注释。

每个对象只存储一个注释字符串，因此要修改注释，请对同一对象发出新的COMMENT命令。要移除注释，请用NULL代替文本字符串。对象删除时，其注释也会自动删除。

对于大多数对象类型，只有对象的拥有者才能设置注释。角色没有拥有者，因此COMMENT ON ROLE的规则是：要为超级用户角色添加注释，必须是超级用户；要为非超级用户角色添加注释，必须拥有CREATEROLE权限。

可以使用psql的\d 系列命令查看这些注释。其他用于检索注释的用户界面，也可以建立在psql使用的同一组内置函数之上，即 obj_description、col_description 和shobj_description（见表 9.52）。

参数

object_name
relation_name.column_name
agg_name
constraint_name
function_name
operator_name
rule_name
trigger_name

要添加注释的对象名称。表、聚合、排序规则、转换、域、外部表、函数、索引、操作符、操作符类、操作符族、序列、文本检索对象、类型和视图的名称可以带模式限定。
为列添加注释时，*relation_name*必须引用表、视图、复合类型或外部表。

agg_type

聚合函数作用于其上的输入数据类型。要引用一个零参数聚合函数，
请在输入数据类型列表的位置写 *。

source_type

类型转换的源数据类型的名称。

target_type

类型转换的目标数据类型的名称。

argmode

函数参数的模式：IN、OUT、INOUT 或 VARIADIC。如果省略，默认为 IN。注意 COMMENT ON FUNCTION 实际上并不关心 OUT 参数，因为确定函数的标识只需要输入参数。因此只列出 IN、INOUT 和 VARIADIC 参数就足够了。

argname

函数参数的名称。注意 COMMENT ON FUNCTION 实际上并不关心参数名称，因为确定函数的标识只需要参数的数据类型。

argtype

大对象的 OID。

large_object_oid

大对象的 OID。

left_type
right_type

操作符参数的数据类型（可带模式限定）。对于前缀或后缀操作符缺失的参数，请写NONE。

PROCEDURAL

这是一个噪声词。

text

新的注释，以字符串字面量形式书写；也可以写 NULL 来移除该注释。

注解

目前没有用于查看注释的安全机制：任何连接到某个数据库的用户都可以看到该数据库中所有对象的注释。对于数据库、角色和表空间这类共享对象，注释是全局存储的，
因此连接到集簇中任一数据库的任何用户都可以看到共享对象的全部注释。因此，
不要在注释中放入与安全密切相关的信息。

示例

为表mytable附加一条注释：

```
COMMENT ON TABLE mytable IS 'This is my table.';
```

再将其移除：

```
COMMENT ON TABLE mytable IS NULL;
```

更多示例：

```
COMMENT ON AGGREGATE my_aggregate (double precision) IS 'Computes
 sample variance';
COMMENT ON CAST (text AS int4) IS 'Allow casts from text to int4';
COMMENT ON COLLATION "fr_CA" IS 'Canadian French';
COMMENT ON COLUMN my_table.my_column IS 'Employee ID number';
COMMENT ON CONVERSION my_conv IS 'Conversion to UTF8';
COMMENT ON CONSTRAINT bar_col_cons ON bar IS 'Constrains column col';
COMMENT ON DATABASE my_database IS 'Development Database';
COMMENT ON DOMAIN my_domain IS 'Email Address Domain';
COMMENT ON EXTENSION hstore IS 'implements the hstore data type';
COMMENT ON FOREIGN DATA WRAPPER mywrapper IS 'my foreign data
 wrapper';
COMMENT ON FOREIGN TABLE my_foreign_table IS 'Employee Information in
 other database';
COMMENT ON FUNCTION my_function (timestamp) IS 'Returns Roman
 Numeral';
COMMENT ON INDEX my_index IS 'Enforces uniqueness on employee ID';
COMMENT ON LANGUAGE plpython IS 'Python support for stored
 procedures';
COMMENT ON LARGE OBJECT 346344 IS 'Planning document';
COMMENT ON MATERIALIZED VIEW my_matview IS 'Summary of order history';
COMMENT ON OPERATOR ^ (text, text) IS 'Performs intersection of two
 texts';
```

```
COMMENT ON OPERATOR - (NONE, integer) IS 'Unary minus';
COMMENT ON OPERATOR CLASS int4ops USING btree IS '4 byte integer
 operators for btrees';
COMMENT ON OPERATOR FAMILY integer_ops USING btree IS 'all integer
 operators for btrees';
COMMENT ON ROLE my_role IS 'Administration group for finance tables';
COMMENT ON RULE my_rule ON my_table IS 'Logs updates of employee
 records';
COMMENT ON SCHEMA my_schema IS 'Departmental data';
COMMENT ON SEQUENCE my_sequence IS 'Used to generate primary keys';
COMMENT ON SERVER myserver IS 'my foreign server';
COMMENT ON TABLE my_schema.my_table IS 'Employee Information';
COMMENT ON TABLESPACE my_tablespace IS 'Tablespace for indexes';
COMMENT ON TEXT SEARCH CONFIGURATION my_config IS 'Special word
 filtering';
COMMENT ON TEXT SEARCH DICTIONARY swedish IS 'Snowball stemmer for
 Swedish language';
COMMENT ON TEXT SEARCH PARSER my_parser IS 'Splits text into words';
COMMENT ON TEXT SEARCH TEMPLATE snowball IS 'Snowball stemmer';
COMMENT ON TRIGGER my_trigger ON my_table IS 'Used for RI';
COMMENT ON TYPE complex IS 'Complex number data type';
COMMENT ON VIEW my_view IS 'View of departmental costs';
```

兼容性

SQL 标准中没有COMMENT命令。

COMMIT

COMMIT — commit the current transaction

大纲

```
COMMIT [ WORK | TRANSACTION ]
```

描述

`COMMIT`提交当前事务。该事务所做的所有更改都会对其他会话可见，并且如果发生崩溃，也能保证其持久性。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用ROLLBACK中止事务。

在事务之外发出 `COMMIT` 不会造成危害，但会产生一条警告消息。

示例

提交当前事务，并使所有更改永久生效：

```
COMMIT;
```

兼容性

SQL 标准只规定了 `COMMIT` 和 `COMMIT WORK` 这两种形式。在其他方面，此命令完全符合标准。

另见

BEGIN, ROLLBACK

COMMIT PREPARED

COMMIT PREPARED — commit a transaction that was earlier prepared for two-phase commit

大纲

```
COMMIT PREPARED transaction_id
```

描述

`COMMIT PREPARED`提交一个处于预备状态的事务。

参数

transaction_id

要提交的事务的事务标识符。

注解

要提交预备事务，执行者必须是最初执行该事务的同一用户，或者是超级用户；但不必处在执行该事务的同一会话中。

这个命令不能在事务块内执行。该预备事务会被立即提交。

当前所有处于预备状态的事务都列在 `pg_prepared_xacts` 系统视图中。

示例

提交事务标识符为`foobar`的事务：

```
COMMIT PREPARED 'foobar';
```

另见

PREPARE TRANSACTION, ROLLBACK PREPARED

COPY

COPY — copy data between a file and a table

大纲

```
COPY table_name [ ( column [, ...] ) ]
    FROM { 'filename' | STDIN }
    [ [ WITH ] ( option [, ...] ) ]

COPY { table_name [ ( column [, ...] ) ] | ( query ) }
    TO { 'filename' | STDOUT }
    [ [ WITH ] ( option [, ...] ) ]

where option can be one of:

    FORMAT format_name
    OIDS [ boolean ]
    DELIMITER 'delimiter_character'
    NULL 'null_string'
    HEADER [ boolean ]
    QUOTE 'quote_character'
    ESCAPE 'escape_character'
    FORCE_QUOTE { ( column [, ...] ) | * }
    FORCE_NOT_NULL ( column [, ...] )
    ENCODING 'encoding_name'
```

描述

`COPY`在 PostgreSQL表与标准文件系统文件之间传输数据。`COPY TO`将表的内容复制到文件，而`COPY FROM` 则将数据从文件复制到表中（追加到表中已有的数据之后）。`COPY TO`也可以复制`SELECT`查询的结果。

如果指定了列列表，`COPY`将只在指定列与文件之间复制数据。如果表中存在未列入列列表的列，`COPY FROM`将为这些列插入默认值。

带文件名的 `COPY` 会指示 PostgreSQL 服务器直接从文件读取或向文件写入。该文件必须可由服务器访问，并且其名称必须从服务器的视角指定。当指定 `STDIN` 或 `STDOUT` 时，数据通过客户端与服务器之间的连接传输。

参数

table_name

一个现有表的名称（可以是模式限定的）。

column

要复制的可选列列表。如果没有指定列列表，则会复制该表所有列。

query

其结果将被复制的SELECT或VALUES命令。注意，查询外层必须带圆括号。

filename

输入或输出文件的绝对路径名。Windows 用户可能需要使用 `E''` 字符串，并把路径名中用到的反斜线都写成双份。

`STDIN`

指定输入来自客户端应用程序。

`STDOUT`

指定输出发送到客户端应用程序。

boolean

指定所选选项是否开启。可以写`TRUE`、`ON`或`1`来启用选项，写`FALSE`、`OFF`或`0`来禁用它。也可以省略*boolean*值，此时假定为`TRUE`。

`FORMAT`

选择要读取或写入的数据格式：`text`、`csv`（逗号分隔值）或`binary`。默认为`text`。

`OIDS`

指定复制每一行的 OID。（如果为没有 OID 的表指定了 `OIDS`，或者复制的是一个 *query*，就会报错。）

`DELIMITER`

指定文件中每行内分隔列的字符。文本格式默认为制表符，`CSV`格式默认为逗号。它必须是单个单字节字符。使用`binary`格式时不允许此选项。

`NULL`

指定表示一个空值的字符串。文本格式中默认是 `\N`（反斜线-N），`CSV`格式中默认是一个未加引用的空串。在你不想区分空值和空串的情况下，即使在文本格式中你也可能更喜欢空串。使用`binary`格式时不允许这个选项。

> **注意**
>
> 在使用`COPY FROM`时，任何匹配此字符串的数据项都会被存储为空值，因此应确保这里使用的字符串与 `COPY TO`时使用的相同。

`HEADER`

指定文件包含一个标题行，其中列出文件中每一列的名称。输出时，第一行包含表中的列名；输入时，第一行会被忽略。只有使用 `CSV` 格式时才允许此选项。

`QUOTE`

指定在对数据值加引号时使用的引用字符。默认是双引号。这必须是一个单一的单字节字符。只有使用 `CSV`格式时才允许这个选项。

ESCAPE

指定在与QUOTE值匹配的数据字符之前应出现的字符。默认值与QUOTE值相同（这样当引用字符出现在数据中时，就会被双写）。这必须是一个单一的单字节字符。只有使用CSV格式时才允许这个选项。

FORCE_QUOTE

强制对每个指定列中的所有非NULL值使用引号。NULL输出永远不会加引号。如果指定了*，则所有列中的非NULL值都会加引号。此选项仅允许用于 COPY TO，且只能在使用CSV格式时使用。

FORCE_NOT_NULL

不要将指定列的值与空值串进行匹配。在空值串就是空串的默认情况下，这意味着空串将被读作长度为零的字符串而不是空值（即使它们没有被引用）。只有在COPY FROM中使用 CSV格式时才允许这个选项。

ENCODING

指定文件采用*encoding_name*编码。如果省略此选项，将使用当前客户端编码。详见下文注解。

输出

成功完成时，COPY命令会返回形如

COPY *count*

的命令标签。*count*为复制的行数。

注解

COPY只能用于普通表，不能用于视图。但可以写 COPY (SELECT * FROM *viewname*) TO ...。

COPY只处理指定的表本身；它不会从子表复制数据或向子表复制数据。因此，例如COPY *table* TO显示的数据与 SELECT * FROM ONLY *table*相同。但COPY (SELECT * FROM *table*) TO ...可用于导出继承层次结构中的所有数据。

你必须对COPY TO读取其值的表具有 SELECT权限，并对COPY FROM 插入其值的表具有INSERT权限。对于命令中列出的列，具有列级权限即可。

COPY 命令中命名的文件由服务器直接读取或写入，而不是由客户端应用进行。因此，这些文件必须位于数据库服务器机器上或可由其访问，而不是客户端上。它们必须可由 PostgreSQL 用户（服务器运行时使用的用户 ID）访问并可读或可写，而不是由客户端访问。带有文件名的 COPY 只允许数据库超级用户使用，因为它允许读取或写入服务器有权访问的任何文件。

不要把 COPY 与 psql 的指令 \copy 混淆。\copy 会调用 COPY FROM STDIN 或 COPY TO STDOUT，然后在 psql 客户端可访问的文件中获取/存储数据。因此，使用 \copy 时，文件的可访问性和访问权限取决于客户端而不是服务器。

建议在COPY中使用的文件名始终指定为绝对路径。对于COPY TO，服务器会强制这一点；但对于COPY FROM，你仍可选择从使用相对路径指定的文件中读取。该路径将相对于服务器进程的工作目录（通常是集簇的数据目录）而非客户端的工作目录进行解释。

COPY FROM将调用目标表上的任何触发器和检查约束。但是它不会调用规则。

COPY的输入和输出会受 DateStyle影响。为确保数据能移植到其他可能使用非默认 DateStyle设置的PostgreSQL 安装中，使用COPY TO前应将 DateStyle设置为ISO。同样也建议避免在 IntervalStyle设置为sql_standard时转储数据，因为负的 interval 值可能会被采用不同 IntervalStyle设置的服务器误解。

即使数据会被服务器直接从一个文件读取或者写入一个文件而不通过客户端，
输入数据也会被根据ENCODING选项或者当前客户端编码解释，并且输出数据会被根据
ENCODING或者当前客户端编码进行编码。

COPY 会在遇到第一个错误时停止操作。对于 COPY TO，这应当不会导致问题；但对于 COPY FROM，目标表此时已经接收了前面的行。这些行不可见也不可访问，但仍占用磁盘空间。如果在一次大型复制操作已经进行很久后才失败，可能会浪费大量磁盘空间。可以调用 VACUUM 回收这些浪费的空间。

文件格式

文本格式

在使用text格式时，读取或写入的是一个文本文件，其中表中的每一行对应文件中的一行。
每行中的列由分隔符字符隔开。
列值本身是由各属性数据类型的输出函数生成或可被其输入函数接受的字符串。
对于为空值的列，会使用指定的空值串代替。
如果输入文件中的任何一行包含的列数多于或少于预期，COPY FROM就会报错。如果指定了OIDS，OID 会作为第一列读取或写入，位于用户数据列之前。

数据结束可以表示为只包含反斜线加点号（\.）的一行。从文件读取时，不需要数据结束标记，因为文件结束已足够；只有在使用 3.0 之前版本的客户端协议、向客户端应用程序复制数据或从其复制数据时，才需要该标记。

在COPY数据中，可以使用反斜线字符（\）
来转义那些原本可能被当作行或列分隔符的数据字符。特别是，
如果下列字符作为列值的一部分出现，那么它们前面必须加一个反斜线：反斜线本身、换行、
回车以及当前分隔符字符。

COPY TO输出指定的空值串时不会添加任何反斜线；相反，COPY FROM会在去除反斜线之前先将输入与空值串进行匹配。因此，像\N这样的空值串不会与实际的数据值\N混淆，
因为后者会表示为\\N。

COPY FROM识别下列特殊的反斜线序列：

序列	表示
\b	退格 (ASCII 8)
\f	换页 (ASCII 12)
\n	新行 (ASCII 10)
\r	回车 (ASCII 13)
\t	制表 (ASCII 9)
\v	纵向制表 (ASCII 11)
digits	反斜线后跟一到三个八进制数字表示该数字代码对应的字节

序列	表示
\xdigits	反斜线加x后跟一到两个十六进制数字表示该数字代码对应的字节

目前，COPY TO从不会输出八进制或十六进制数字反斜线序列，但对这些控制字符确实会使用上表列出的其他序列。

任何上表中未提到的其他反斜线字符都将表示其自身。不过，要注意不要不必要地添加反斜线，因为那可能意外地产生与数据结束标记（\.）或空值串（默认是\N）匹配的字符串。这些字符串会在进行任何其他反斜线处理之前先被识别出来。

强烈建议生成COPY数据的应用将数据中的换行和回车分别转换为\n和\r序列。目前，仍然可以用反斜线加回车表示数据回车，用反斜线加换行表示数据换行。不过，未来版本可能不再接受这些表示方式。如果COPY文件在不同机器之间传输（例如从 Unix 到 Windows，或反之），这些表示方式也非常容易被破坏。

COPY TO会用 Unix 风格的换行（“\n”）结束每一行。运行在 Microsoft Windows 上的服务器则会输出回车/换行（“\r\n”），但这只适用于复制到服务器文件的COPY；为保证跨平台一致性，COPY TO STDOUT总是发送“\n”，与服务器平台无关。COPY FROM能够处理以换行、回车或回车/换行结束的行。为减少本应是数据的未加反斜线新行或回车带来的风险，如果输入中的行结束符并不一致，COPY FROM将会报错。

CSV 格式

这种格式选项用于导入和导出许多其他程序（如电子表格）使用的逗号分隔值（CSV）文件格式。不同于 PostgreSQL标准文本格式使用的转义规则，它会生成并识别通用的 CSV转义机制。

每条记录中的值由DELIMITER字符分隔。如果某个值包含分隔符字符、QUOTE字符、NULL字符串、回车或换行字符，那么整个值都会以前后各一个QUOTE字符包围，并且该值内每次出现QUOTE字符或ESCAPE 字符之前都会加上转义字符。对于指定列中的非NULL值输出，还可以使用FORCE_QUOTE来强制加引号。

CSV格式没有标准方式区分NULL值和空字符串。PostgreSQL的COPY通过引号来处理这一区别。NULL会按照NULL参数字符串输出，且不会被加引号；而与NULL参数字符串匹配的非NULL 值会被加引号。例如，在默认设置下，NULL会写成一个未加引号的空字符串，而空字符串数据值会写成双引号包围的形式（""）。读取值时遵循类似规则。你可以使用FORCE_NOT_NULL来阻止对指定列进行NULL输入比较。

因为反斜线在 CSV 格式中不是特殊字符，数据结束标记 \. 也可能作为数据值出现。为避免误解，当 \. 数据值作为一行中的唯一字段出现时，输出时会自动给它加上引号；输入时，如果它带有引号，就不会被解释为数据结束标记。如果正在装载由其他应用程序创建的文件，其中只有一个不带引号的列，而且该列可能包含值 \.，那么可能需要在输入文件中给该值加上引号。

> **注意**
>
> 在CSV格式中，所有字符都有意义。被空白字符或 DELIMITER之外其他字符包围的带引号值，会把这些字符也包含进值中。如果你导入的数据来自某个会用空白把CSV 行填充到固定宽度的系统，这可能导致错误。出现这种情况时，你可能需要在将数据导入PostgreSQL之前，先预处理CSV文件以移除尾随空白。

> **注意**
>
> CSV 格式既能识别也能生成这样的 CSV 文件：其中带引号的值包含内嵌的回车和换行。因此，这类文件不像文本格式文件那样严格地一行对应表中的一行。

> **注意**
>
> 很多程序会生成奇怪、甚至近乎反常的CSV文件，
> 因此这种文件格式更像一种约定而非标准。
> 因而你可能会遇到无法用这种机制导入的文件，而`COPY`
> 也可能生成其他程序无法处理的文件。

二进制格式

`binary`格式选项会使所有数据以二进制格式而不是文本格式存储或读取。它比文本和`CSV`格式稍快一些，但二进制格式文件在不同的机器架构和PostgreSQL版本之间的可移植性较差。此外，二进制格式与数据类型高度相关。例如，不能从`smallint`列输出二进制数据再读入到`integer`列中，尽管这种做法在文本格式下是可行的。

`binary`文件格式由文件头、零个或多个包含行数据的元组以及一个文件尾构成。头部和数据都以网络字节序表示。

> **注意**
>
> 7.4 之前的PostgreSQL版本使用一种不同的二进制文件格式。

文件头

文件头由 15 字节的固定字段组成，后面跟着一个可变长度的头部扩展区域。固定字段如下：

签名

11 字节序列`PGCOPY\n\377\r\n\0` — 注意，零字节是签名中必不可少的一部分。（该签名的设计目的是便于识别那些在不具备 8 位透明性的传输过程中遭到破坏的文件。行尾转换过滤器、零字节丢失、高位丢失或奇偶校验变化等情况都会改变该签名。）

标志域

这是一个 32 位整数位掩码，用于表示文件格式的重要属性。位编号从 0（LSB）到 31（MSB）。注意，此字段与文件格式中的所有整数字段一样，采用网络字节序存储（最高有效字节在前）。位 16-31 保留用于表示文件格式的关键问题；如果读取程序发现这个范围内有非预期的位被置位，就应中止。位 0-15 保留用于表示向后兼容的格式问题；读取程序应直接忽略这个范围内非预期的置位。目前仅定义了一个标志位，其余位必须为零：

位 16

如果为 1，则数据中包含 OID；如果为 0，则不包含。

头部扩展区长度

32 位整数，表示头部剩余部分的长度（以字节计），不包括该字段本身。当前该值为零，因此其后紧接着第一个元组。未来对这种格式的更改可能允许在头部中包含额外数据。
如果读取程序不知道如何处理头部扩展区数据，应静默跳过它。

头部扩展区被设想为包含一系列可自我标识的块。
标志域并不用于告诉读取程序扩展区中包含哪些内容。
头部扩展内容的具体设计留待后续版本决定。

这种设计既允许向后兼容的头部新增（增加头部扩展块，或设置低位标志位），
也允许不向后兼容的更改（设置高位标志位来表明这类更改，
并在需要时向扩展区增加支持数据）。

元组

每个元组都以一个 16 位整数计数开头，用于表示该元组中的字段数。（目前，一个表中的所有元组都应有相同的计数，但这未必永远如此。）随后，对元组中的每个字段，都会有一个 32 位长度字，后跟该字段数据的相应字节数。（长度字不包括其本身，且可以为零。）特殊情况下，-1 表示一个 NULL 字段值；在 NULL 情况下，后面不会跟随任何值字节。

字段之间没有对齐填充或任何其他额外数据。

当前，二进制格式文件中的所有数据值都假定为二进制格式（格式代码一）。可以预见，
未来的扩展可能会增加一个允许为各列分别指定格式代码的头部字段。

要确定实际元组数据应采用的二进制格式，你应该参考 PostgreSQL源码，特别是各列数据类型对应的`*send`和`*recv`函数（这些函数通常可以在源码分发包的`src/backend/utils/adt/`目录中找到）。

如果文件中包含 OID，则 OID 字段紧跟在字段计数值之后。它是一个普通字段，只是不计入字段数。特别是，它带有一个长度字段 – 这样就能较容易地处理 4 字节或 8 字节的 OID，也允许在将来确有需要时把 OID 表示为空值。

文件尾

文件尾由一个值为 -1 的 16 位整数构成。这很容易与元组的字段计数字区分开来。

如果字段计数字既不是 -1 也不是预期的列数，读取程序应报告错误。这提供了一项额外检查，以防与数据失去同步。

示例

下面的示例使用竖线（`|`）作为字段分隔符将一个表复制到客户端：

```
COPY country TO STDOUT (DELIMITER '|');
```

要将文件中的数据复制到`country`表中：

```
COPY country FROM '/usr1/proj/bray/sql/country_data';
```

只把名称以 'A' 开头的国家复制到一个文件中：

```
COPY (SELECT * FROM country WHERE country_name LIKE 'A%') TO '/usr1/
proj/bray/sql/a_list_countries.copy';
```

下面给出适合从`STDIN`复制到表中的示例数据：

```
AF      AFGHANISTAN
AL      ALBANIA
DZ      ALGERIA
ZM      ZAMBIA
ZW      ZIMBABWE
```

注意每一行中的空白实际上是一个制表符。

下面是用二进制格式输出的相同数据。该数据是用 Unix 工具 `od -c`过滤后显示的。该表具有三列，第一列类型是`char(2)`，第二列类型是`text`，第三列类型是`integer`。
所有行在第三列都是空值。

```
0000000   P   G   C   O   P   Y  \n 377  \r  \n  \0  \0  \0  \0  \0
 \0
0000020  \0  \0  \0  \0 003  \0  \0  \0 002   A   F  \0  \0  \0 013
 A
0000040   F   G   H   A   N   I   S   T   A   N 377 377 377 377  \0
 003
0000060  \0  \0  \0 002   A   L  \0  \0  \0 007   A   L   B   A   N
 I
0000100   A 377 377 377 377  \0 003  \0  \0  \0 002   D   Z  \0  \0
 \0
0000120 007   A   L   G   E   R   I   A 377 377 377 377  \0 003  \0
 \0
0000140  \0 002   Z   M  \0  \0  \0 006   Z   A   M   B   I   A 377
 377
0000160 377 377  \0 003  \0  \0  \0 002   Z   W  \0  \0  \0  \b   Z
 I
0000200   M   B   A   B   W   E 377 377 377 377 377 377
```

兼容性

SQL 标准中没有`COPY`语句。

以下语法曾在PostgreSQL9.0 之前的版本中使用，目前仍受支持：

```
COPY table_name [ ( column_name [, ...] ) ]
    FROM { 'filename' | STDIN }
    [ [ WITH ]
          [ BINARY ]
          [ OIDS ]
          [ DELIMITER [ AS ] 'delimiter_character' ]
          [ NULL [ AS ] 'null string' ]
          [ CSV [ HEADER ]
                [ QUOTE [ AS ] 'quote_character' ]
                [ ESCAPE [ AS ] 'escape_character' ]
                [ FORCE NOT NULL column_name [, ...] ] ] ]

COPY { table_name [ ( column_name [, ...] ) ] | ( query ) }
    TO { 'filename' | STDOUT }
    [ [ WITH ]
          [ BINARY ]
          [ OIDS ]
          [ DELIMITER [ AS ] 'delimiter_character' ]
```

```
          [ NULL [ AS ] 'null string' ]
          [ CSV [ HEADER ]
                [ QUOTE [ AS ] 'quote_character' ]
                [ ESCAPE [ AS ] 'escape_character' ]
                [ FORCE QUOTE { column_name [, ...] | * } ] ] ]
```

注意，在这种语法中，BINARY和CSV被当作独立的关键字，而不是FORMAT选项的参数。

以下语法曾在PostgreSQL7.3 之前的版本中使用，目前仍受支持：

```
COPY [ BINARY ] table_name [ WITH OIDS ]
    FROM { 'filename' | STDIN }
    [ [USING] DELIMITERS 'delimiter_character' ]
    [ WITH NULL AS 'null_string' ]

COPY [ BINARY ] table_name [ WITH OIDS ]
    TO { 'filename' | STDOUT }
    [ [USING] DELIMITERS 'delimiter_character' ]
    [ WITH NULL AS 'null_string' ]
```

CREATE AGGREGATE

CREATE AGGREGATE — define a new aggregate function

大纲

```
CREATE AGGREGATE name ( input_data_type [ , ... ] ) (
    SFUNC = sfunc,
    STYPE = state_data_type
    [ , FINALFUNC = ffunc ]
    [ , INITCOND = initial_condition ]
    [ , SORTOP = sort_operator ]
)

or the old syntax

CREATE AGGREGATE name (
    BASETYPE = base_type,
    SFUNC = sfunc,
    STYPE = state_data_type
    [ , FINALFUNC = ffunc ]
    [ , INITCOND = initial_condition ]
    [ , SORTOP = sort_operator ]
)
```

描述

`CREATE AGGREGATE` 定义一个新的聚合函数。系统自带了一些基本且常用的聚合函数；它们在第 9.18 节中有文档说明。如果需要定义新类型或需要一个尚未提供的聚合函数，则可以使用 `CREATE AGGREGATE` 来实现所需功能。

如果给出了一个模式名（例如`CREATE AGGREGATE myschema.myagg ...`），则该聚合函数会在指定模式中创建。否则会在当前模式中创建。

聚合函数由其名称和输入数据类型标识。同一模式中的两个聚合如果操作于不同的输入类型，则可以有相同的名称。
聚合的名称和输入数据类型还必须与同一模式中每个普通函数的名称和输入数据类型都不同。

一个聚合函数由一个或两个普通函数构成：一个状态转移函数*sfunc*，
以及一个可选的最终计算函数*ffunc*。它们的用法如下：

```
sfunc( internal-state, next-data-values ) ---> next-internal-state
ffunc( internal-state ) ---> aggregate-value
```

PostgreSQL会创建一个数据类型为 *stype*的临时变量，用来保存聚合的当前内部状态。对于每个输入行，都会先计算聚合参数值，然后以当前状态值和新的参数值调用状态转移函数，
从而计算新的内部状态值。等到所有行都处理完毕后，
再调用一次最终函数来计算聚合的返回值。如果没有最终函数，则原样返回结束时的状态值。

聚合函数可以提供一个初始条件，即内部状态值的初始值。它在数据库中以 `text`类型的值指定和存储，但它必须是该状态值数据类型常量的合法外部表示。如果未提供，则状态值初始为空值。

如果状态转移函数被声明为“strict”的，那么它不能以空值输入被调用。使用这种转移函数时，聚合的执行行为如下：具有任一空值输入的行会被忽略（不调用函数并保留先前的状态值）。如果初始状态值为空，那么在第一个所有输入均非空的行上，第一个参数值会取代状态值，而转移函数会在后续所有输入均非空的行上被调用。这对于实现`max`这类聚合很方便。注意，只有当*`state_data_type`*与第一个*`input_data_type`*相同时，此行为才可用。当这两种类型不同时，必须提供非空的初始条件或者使用非严格的转移函数。

如果状态转移函数不是严格的，那么它会在每个输入行被无条件调用，并且必须自行处理空值输入和空值转移值。这使聚合作者能完全控制聚合对空值的处理。

如果最终函数被声明为“strict”，那么当结束状态值为空时不会调用它；而是自动返回空结果。（这当然只是 strict 函数的正常行为。）无论如何，最终函数都可以选择返回空值。例如，`avg`的最终函数在发现输入行数为零时会返回空值。

行为类似于`MIN`或`MAX`的聚合，有时可以通过查阅索引而不是扫描每个输入行来优化。如果该聚合可以这样优化，请通过指定一个排序操作符来表明。基本要求是，该聚合必须返回该操作符所诱导的排序顺序中的第一个元素；换句话说：

```
SELECT agg(col) FROM tab;
```

必须等价于：

```
SELECT col FROM tab ORDER BY col USING sortop LIMIT 1;
```

进一步的假设是，该聚合忽略空输入，并且当且仅当不存在非空输入时返回空结果。通常，某种数据类型的<操作符是 `MIN`的合适排序操作符，而>是 `MAX`的合适排序操作符。注意，除非指定的操作符是 B-树索引操作符类中“小于”或“大于”策略成员，否则这种优化实际上永远不会生效。

参数

`name`

要创建的聚合函数的名称（可带模式限定）。

`input_data_type`

此聚合函数所操作的一个输入数据类型。要创建一个零参数的聚合函数，可以在输入数据类型列表的位置写`*`。（这种聚合的一个例子是`count(*)`。）

`base_type`

在`CREATE AGGREGATE`的旧语法中，输入数据类型由`basetype`参数指定，而不是写在聚合名称旁边。注意这种语法只允许一个输入参数。要定义零参数的聚合函数，需将`basetype`指定为`"ANY"`（而不是`*`）。

`sfunc`

要为每个输入行调用的状态转移函数的名称。对于一个*`N`*参数的聚合函数，*`sfunc`*必须接受*`N`*+1个参数，第一个参数的类型为*`state_data_type`*，其余参数与聚合声明的输入数据类型相匹配。该函数必须返回*`state_data_type`*类型的值。此函数接受当前状态值和当前输入数据值，并返回下一个状态值。

`state_data_type`

聚合状态值的数据类型。

ffunc

在遍历完所有输入行之后调用的、用于计算聚合结果的最终函数的名称。
该函数必须接受单个类型为*state_data_type*的参数。
聚合的返回数据类型被定义为该函数的返回类型。如果没有指定*ffunc*，
则结束状态值被用作聚合的结果，返回类型为*state_data_type*。

initial_condition

状态值的初始设置。这必须是 *state_data_type* 数据类型所接受形式的字符串常量。如果未指定，状态值将从 null 开始。

sort_operator

类MIN或MAX聚合的关联排序操作符。这只是一个操作符名称（可以用模式限定）。
假定该操作符具有与聚合相同的输入数据类型（聚合必须是单参数聚合）。

CREATE AGGREGATE的参数可以按任意顺序书写，而不只是按上面展示的顺序。

示例

见第 35.10 节。

兼容性

CREATE AGGREGATE是 PostgreSQL的语言扩展。SQL 标准不提供用户定义聚合函数。

另见

ALTER AGGREGATE, DROP AGGREGATE

CREATE CAST

CREATE CAST — define a new cast

大纲

```
CREATE CAST (source_type AS target_type)
    WITH FUNCTION function_name (argument_type [, ...])
    [ AS ASSIGNMENT | AS IMPLICIT ]

CREATE CAST (source_type AS target_type)
    WITHOUT FUNCTION
    [ AS ASSIGNMENT | AS IMPLICIT ]

CREATE CAST (source_type AS target_type)
    WITH INOUT
    [ AS ASSIGNMENT | AS IMPLICIT ]
```

描述

`CREATE CAST`定义一种新的类型转换。类型转换规定如何在两种数据类型之间执行转换。例如，

```
SELECT CAST(42 AS float8);
```

通过调用一个预先指定的函数（此处是 `float8(int4)`）把整型常量 42 转换成 `float8`类型。（如果没有定义合适的类型转换，转换就会失败。）

两种类型可以是二进制可强制转换的，这意味着无需调用任何函数，就可以“免费”执行转换。这要求对应的值使用相同的内部表示。例如，`text`和`varchar`这两种类型在两个方向上都是二进制可强制转换的。二进制可强制转换性不一定是对称关系。例如，在当前实现中，从`xml`到`text`的类型转换可以免费执行，但反方向则需要一个至少执行语法检查的函数。（双向都二进制可强制转换的两种类型也称为二进制兼容。）

使用`WITH INOUT`语法，你可以把一种类型转换定义为基于 I/O 的类型转换。基于 I/O 的类型转换通过调用源数据类型的输出函数，并将得到的字符串传给目标数据类型的输入函数来执行。在许多常见情况下，这项特性避免了为转换单独编写类型转换函数的必要。基于 I/O 的类型转换与常规的基于函数的类型转换行为相同，只是实现方式不同。

默认情况下，只有显式请求类型转换时才会调用一种类型转换，也就是显式使用`CAST(`*`x`* `AS` *`typename`*`)`或 *`x`*`::`*`typename`* 这种构造。

如果一种类型转换被标记为`AS ASSIGNMENT`，那么在把值赋给目标数据类型的列时就可以隐式调用它。例如，假设`foo.f1` 是一个`text`类型的列，那么如果从`integer`到 `text`的类型转换被标记为`AS ASSIGNMENT`，则：

```
INSERT INTO foo (f1) VALUES (42);
```

就会被允许，否则不会。（我们通常用术语赋值类型转换来描述这种类型转换。）

如果一种类型转换被标记为`AS IMPLICIT`，那么无论是在赋值上下文中还是在表达式内部，都可以在任何上下文中隐式调用它。（我们通常用术语隐式类型转换来描述这种类型转换。）例如，考虑这个查询：

```
SELECT 2 + 4.0;
```

解析器最初分别把这两个常量标记为integer和 numeric类型。系统目录中没有integer + numeric操作符，但有一个 numeric + numeric操作符。因此，如果存在一种从integer到numeric的可用类型转换，并且被标记为AS IMPLICIT – 实际上确实如此 – 该查询就会成功。解析器将应用该隐式类型转换，并把该查询解析为如同它被写成了：

```
SELECT CAST ( 2 AS numeric ) + 4.0;
```

现在，系统目录还提供了一种从numeric到integer的类型转换。如果该类型转换被标记为AS IMPLICIT – 实际上并没有 – 那么解析器就必须在上面的解释方式和另一种方案之间作出选择：把numeric常量转换成integer，然后应用 integer + integer操作符。由于它不知道该偏向哪一种选择，就会放弃并将该查询判定为有歧义。两种类型转换中只有一种是隐式的，正是借此我们让解析器倾向于把混合了 numeric和integer的表达式解析为 numeric；系统对此并没有内置知识。

把类型转换标记为隐式时应当保持保守。过多的隐式类型转换路径可能导致 PostgreSQL对命令作出令人意外的解释，或者因为存在多种可能的解释而根本无法解析命令。一个好的经验法则是，只有对同一一般类型分类中且能保留信息的类型间转换，才让它可以被隐式调用。例如，从int2到int4的类型转换可以合理地设为隐式，但从float8到int4的类型转换大概应仅限赋值使用。跨类型分类的类型转换，例如从text到int4，最好只允许显式调用。

> **注意**
>
> 有时出于可用性或标准兼容性的原因，有必要在一组类型之间提供多种隐式类型转换，这会带来像上面那样无法避免的歧义。解析器有一种基于类型分类和首选类型的后备启发式规则，在这种情况下有助于提供期望的行为。详见CREATE TYPE。

要能够创建一个类型转换，你必须拥有源数据类型或目标数据类型之一。要创建一个二进制可强制转换的转换，你必须是超级用户。（之所以有这一限制，是因为错误的二进制可强制转换很容易使服务器崩溃。）

参数

source_type

该类型转换的源数据类型的名称。

target_type

该类型转换的目标数据类型的名称。

function_name(*argument_type* [, ...])

用于执行该类型转换的函数。函数名可以用模式限定；如果没有，则会在模式搜索路径中查找该函数。该函数的结果数据类型必须与类型转换的目标类型一致。其参数见下文。

WITHOUT FUNCTION

表示源类型对目标类型是二进制可强制转换的，因此执行该类型转换不需要函数。

`WITH INOUT`

表示该类型转换是一种基于 I/O 的类型转换，其执行方式是调用源数据类型的输出函数，并将得到的字符串传给目标数据类型的输入函数。

`AS ASSIGNMENT`

表示该类型转换可以在赋值上下文中隐式调用。

`AS IMPLICIT`

表示该类型转换可以在任何上下文中隐式调用。

类型转换实现函数可以有一到三个参数。第一个参数类型必须与源类型相同，
或者可以由源类型进行二进制强制转换得到。第二个参数（如果有）必须是 `integer`类型；它接收与目标类型关联的类型修饰符，如果没有则为`-1`。第三个参数（如果有）必须是`boolean` 类型；如果该类型转换是显式类型转换，它接收`true`，否则接收`false`。（奇怪的是，SQL 标准在某些情况下要求显式类型转换和隐式类型转换具有不同的行为。
这个参数是为必须实现这类类型转换的函数提供的。
不建议你把自己的数据类型设计成需要关心这一点。）

类型转换函数的返回类型必须与类型转换的目标类型相同，
或者对该目标类型是二进制可强制转换的。

通常，类型转换的源数据类型和目标数据类型必须不同。不过，
如果它具有一个接受多个参数的类型转换实现函数，
则允许声明源类型和目标类型相同的类型转换。
这用于在系统目录中表示特定类型的长度强制函数。
所命名的函数用于将该类型的值强制为其第二个参数给定的类型修饰符值。

当一种类型转换的源类型和目标类型不同，且其函数接受多个参数时，
它支持在一个步骤中同时完成从一种类型到另一种类型的转换并应用长度强制。
如果没有这样的条目，
对使用类型修饰符的类型进行强制就需要两个类型转换步骤：先在数据类型之间进行转换，
再应用该修饰符。

注解

使用DROP CAST删除用户定义的转换。

请记住，如果你希望能够双向转换类型，就需要在两个方向上分别显式声明类型转换。

通常没有必要在用户定义类型与标准字符串类型（`text`、`varchar`和`char(n)`，
以及被定义为属于字符串分类的用户定义类型）之间创建类型转换。
PostgreSQL会自动为此提供基于 I/O 的类型转换。转换到字符串类型的自动类型转换被视为赋值类型转换，而从字符串类型出发的自动类型转换则只允许显式调用。
你可以声明自己的类型转换来替代自动类型转换，从而覆盖这种行为，
但通常这样做的唯一原因，是希望该转换比标准的仅赋值或仅显式设置更容易调用。
另一种可能的原因是，你希望该转换的行为不同于该类型的 I/O 函数；但这已经足够反常，你应该三思这是不是一个好主意。（确实有少数内置类型在转换行为上有所不同，大多是由于 SQL 标准的要求。）

在 PostgreSQL 7.3 之前，每个与某种数据类型同名、返回该数据类型且接受一个不同类型参数的函数都会自动成为类型转换函数。随着模式的引入，

以及为了能在系统目录中表示二进制可强制转换，这一惯例已被放弃。
内建的类型转换函数仍然遵循这一命名方案，但它们也必须显示为系统目录 `pg_cast` 中的转换。

虽然这不是强制要求，但仍建议你继续遵循这种老惯例，
即按目标数据类型为类型转换实现函数命名。
许多用户已经习惯于用函数风格的记法进行类型转换，也就是*typename*(*x*)。
这种记法实际上无非就是调用类型转换实现函数；它不会被特别当作类型转换处理。
如果你的转换函数没有按这种惯例命名，那么用户会感到意外。由于 PostgreSQL允许同一函数名按不同参数类型重载，
因此让来自不同类型的多个转换函数都使用目标类型的名称并不存在困难。

注意

实际上，前一段说得过于简单了：有两种情况下，
即使一个函数调用形式没有匹配到实际存在的函数，也会被当作类型转换请求。
如果函数调用 *name*(*x*)不能与任何现有函数精确匹配，但*name*是一个数据类型名，并且 `pg_cast`为从*x*的类型到该类型提供了二进制强制转换，那么该调用会被解释为二进制强制转换。作出这一例外，是为了让二进制强制转换即使没有任何函数，
也能使用函数语法调用。同样，如果没有`pg_cast`项，
但该类型转换的目标或源是字符串类型，则该调用会被解释为基于 I/O 的类型转换。这一例外允许基于 I/O 的类型转换使用函数语法调用。

注意

还有一个例外中的例外：从复合类型到字符串类型的基于 I/O 的类型转换不能使用函数语法调用，而必须写成显式类型转换语法（`CAST` 或`::`记法）。增加这一例外，是因为在引入自动提供的基于 I/O 的类型转换之后，如果原意是函数调用或列引用，就太容易意外地触发这种类型转换了。

示例

要使用函数`int4(bigint)`创建一种从类型 `bigint`到类型`int4`的赋值类型转换：

```
CREATE CAST (bigint AS int4) WITH FUNCTION int4(bigint) AS ASSIGNMENT;
```

（在系统中这种类型转换已经被预定义。）

兼容性

`CREATE CAST`命令符合SQL标准，不过 SQL 没有对二进制可强制转换的类型或实现函数的额外参数作出规定。`AS IMPLICIT`也是 PostgreSQL的扩展。

另见

CREATE FUNCTION、CREATE TYPE、DROP CAST

CREATE COLLATION

CREATE COLLATION — define a new collation

大纲

```
CREATE COLLATION name (
    [ LOCALE = locale, ]
    [ LC_COLLATE = lc_collate, ]
    [ LC_CTYPE = lc_ctype ]
)
CREATE COLLATION name FROM existing_collation
```

描述

`CREATE COLLATION`使用指定的操作系统区域设置，或者通过复制现有排序规则，定义一个新的排序规则。

要创建排序规则，你必须拥有目标模式上的 `CREATE`权限。

参数

name

排序规则的名称。排序规则名称可以带有模式限定。如果没有，
排序规则将在当前模式中定义。排序规则名称在该模式中必须唯一。
（系统目录中可能包含其他编码的同名排序规则，但如果数据库编码不匹配，
则这些排序规则将被忽略。）

locale

这是同时设置 `LC_COLLATE` 和 `LC_CTYPE` 的快捷方式。如果指定了此选项，则不能指定这两个参数中的任何一个。

lc_collate

为 `LC_COLLATE` 区域类别使用指定的操作系统区域设置。该区域设置必须适用于当前数据库编码。（有关精确规则，请参见CREATE DATABASE。）

lc_ctype

为 `LC_CTYPE` 区域类别使用指定的操作系统区域设置。该区域设置必须适用于当前数据库编码。（有关精确规则，请参见CREATE DATABASE。）

existing_collation

要复制的现有排序规则的名称。新的排序规则将具有与现有排序规则相同的属性，
但它是一个独立对象。

注解

使用`DROP COLLATION`可移除用户定义的排序规则。

关于 PostgreSQL 中排序规则支持的更多信息，见第 22.2 节。

示例

创建排序规则，使用操作系统区域设置`fr_FR.utf8`（假设当前数据库编码为`UTF8`）：

```
CREATE COLLATION french (LOCALE = 'fr_FR.utf8');
```

从现有排序规则创建排序规则：

```
CREATE COLLATION german FROM "de_DE";
```

这在应用程序中使用与操作系统无关的排序规则名称时非常方便。

兼容性

在 SQL 标准中有一个`CREATE COLLATION`语句，但它仅限于复制现有排序规则。创建新排序规则的语法是 PostgreSQL扩展。

另见

ALTER COLLATION, DROP COLLATION

CREATE CONVERSION

CREATE CONVERSION — define a new encoding conversion

大纲

```
CREATE [ DEFAULT ] CONVERSION name
    FOR source_encoding TO dest_encoding FROM function_name
```

描述

CREATE CONVERSION 定义字符集编码之间的新转换。此外，被标记为 DEFAULT 的转换可用于客户端与服务器之间的自动编码转换。为此目的，必须定义两个转换：从编码 A 到 B 以及从编码 B 到 A。

要创建一个转换，你必须对该函数具有EXECUTE权限，并在目标模式上具有CREATE权限。

参数

DEFAULT

DEFAULT子句表示该转换是这一特定源编码到目标编码的默认转换。对于某个编码对，在一个模式中应该只有一个默认转换。

name

转换的名称。转换名可以是模式限定的；如果不是，则该转换定义在当前模式中。转换名在一个模式中必须唯一。

source_encoding

源编码名称。

dest_encoding

目标编码名称。

function_name

用于执行该转换的函数。函数名可以是模式限定的；如果不是，则会在搜索路径中查找该函数。

该函数必须具有以下签名：

```
conv_proc(
    integer,  -- source encoding ID
    integer,  -- destination encoding ID
    cstring,  -- source string (null terminated C string)
    internal, -- destination (fill with a null terminated C string)
    integer   -- source string length
) RETURNS void;
```

注解

使用`DROP CONVERSION`可以删除用户定义的转换。

创建转换所要求的权限在未来版本中可能会更改。

示例

使用`myfunc`创建从编码`UTF8`到 `LATIN1`的转换：

```
CREATE CONVERSION myconv FOR 'UTF8' TO 'LATIN1' FROM myfunc;
```

兼容性

`CREATE CONVERSION`是 PostgreSQL的一个扩展。SQL 标准中没有`CREATE CONVERSION`语句，但有一个在目的和语法上都非常相似的 `CREATE TRANSLATION`语句。

另见

ALTER CONVERSION, CREATE FUNCTION, DROP CONVERSION

CREATE DATABASE

CREATE DATABASE — create a new database

大纲

```
CREATE DATABASE name
    [ [ WITH ] [ OWNER [=] user_name ]
           [ TEMPLATE [=] template ]
           [ ENCODING [=] encoding ]
           [ LC_COLLATE [=] lc_collate ]
           [ LC_CTYPE [=] lc_ctype ]
           [ TABLESPACE [=] tablespace ]
           [ CONNECTION LIMIT [=] connlimit ] ]
```

描述

`CREATE DATABASE` 创建一个新的 PostgreSQL 数据库。

要创建数据库，你必须是超级用户，或者拥有特殊的 `CREATEDB` 权限。见CREATE USER。

默认情况下，新数据库通过克隆标准系统数据库 `template1` 来创建。可以通过写成 `TEMPLATE` *`name`* 指定其他模板。特别是，写成 `TEMPLATE template0` 时，可以创建一个全新的数据库，它只包含你的 PostgreSQL 版本预定义的标准对象。如果你希望避免复制任何可能已添加到 `template1` 中的站点本地附加对象，这会很有用。

参数

`name`

要创建的数据库名称。

`user_name`

新数据库所有者的角色名称，或使用 `DEFAULT` 以使用默认值（即执行命令的用户）。要创建由其他角色拥有的数据库，你必须是该角色的直接或间接成员，或为超级用户。

`template`

用于创建新数据库的模板名称，或者指定 `DEFAULT` 以使用默认模板（`template1`）。

`encoding`

新数据库要使用的字符集编码。可指定字符串常量（例如 `'SQL_ASCII'`）、整数编码编号，或者指定 `DEFAULT` 以使用默认编码（即模板数据库的编码）。PostgreSQL 服务器支持的字符集见第 22.3.1 节。其他限制见下文。

`lc_collate`

新数据库中使用的排序规则（`LC_COLLATE`）。这会影响字符串的排序顺序，例如在带有 ORDER BY 的查询中，以及文本列索引中使用的顺序。默认值是使用模板数据库的排序规则。请参见下文的其他限制。

lc_ctype

新数据库中使用的字符分类（LC_CTYPE）。这会影响字符的分类，例如小写、大写和数字。默认值是使用模板数据库的字符分类。请参见下文的其他限制。

tablespace

将与新数据库关联的表空间名称，或者指定 DEFAULT 以使用模板数据库的表空间。该表空间将成为在此数据库中创建对象时使用的默认表空间。详见CREATE TABLESPACE。

connlimit

可对该数据库建立的并发连接数。-1（默认值）表示不受限制。

可选参数可以按任意顺序书写，不一定要按照上面的顺序。

注解

CREATE DATABASE 不能在事务块内执行。

形如 “could not initialize database directory” 的错误，多半与数据目录权限不足、磁盘已满或其他文件系统问题有关。

使用DROP DATABASE删除数据库。

程序createdb是这个命令的一个包装器程序，为方便使用而提供。

尽管可以通过把某个数据库名指定为模板，从而复制它而不是复制 template1，但这（至少目前）并不打算作为一种通用的“COPY DATABASE”功能。主要限制是，在复制模板数据库期间，不能有任何其他会话连接到该数据库。CREATE DATABASE 在启动时如果发现存在任何其他连接，就会失败；否则，在 CREATE DATABASE 完成之前，将阻止对模板数据库建立新的连接。详见第 21.3 节。

为新数据库指定的字符集编码必须与所选区域设置（LC_COLLATE 和 LC_CTYPE）兼容。如果区域设置为 C（或等价的 POSIX），则允许所有编码；但对于其他区域设置，只有一种编码能够正常工作。（不过，在 Windows 上，UTF-8 编码可与任何区域设置一起使用。）CREATE DATABASE 允许超级用户不考虑区域设置而指定 SQL_ASCII 编码，但这种选择已弃用；如果数据库中存储了与该区域设置不兼容编码的数据，字符串函数的行为可能会出错。

编码和区域设置必须与模板数据库的设置一致，除非使用 template0 作为模板。这是因为其他数据库可能包含与指定编码不匹配的数据，或者包含排序顺序会受 LC_COLLATE 和 LC_CTYPE 影响的索引。复制这样的数据会导致数据库在新设置下被视为损坏。不过，已知 template0 不包含任何会受此影响的数据或索引。

CONNECTION LIMIT 选项只是近似地被强制执行；如果两个新会话几乎同时启动，而该数据库只剩下一个连接“槽”，则两者都可能失败。此外，该限制对超级用户无效。

示例

要创建一个新数据库：

```
CREATE DATABASE lusiadas;
```

要创建一个由用户 salesapp 拥有、默认表空间为 salesspace 的数据库 sales：

```
CREATE DATABASE sales OWNER salesapp TABLESPACE salesspace;
```

要创建支持 ISO-8859-1 字符集的数据库`music`：

```
CREATE DATABASE music ENCODING 'LATIN1' TEMPLATE template0;
```

在这个示例中，只有当`template1`的编码不是 ISO-8859-1 时，才需要`TEMPLATE template0`子句。注意，更改编码可能还需要选择新的`LC_COLLATE`和`LC_CTYPE`设置。

兼容性

SQL 标准中没有 `CREATE DATABASE` 语句。数据库相当于目录，而目录的创建由实现定义。

参见

ALTER DATABASE, DROP DATABASE

CREATE DOMAIN

CREATE DOMAIN — define a new domain

大纲

```
CREATE DOMAIN name [ AS ] data_type
    [ COLLATE collation ]
    [ DEFAULT expression ]
    [ constraint [ ... ] ]

where constraint is:

[ CONSTRAINT constraint_name ]
{ NOT NULL | NULL | CHECK (expression) }
```

描述

`CREATE DOMAIN`创建一个新域。域本质上是一种带有可选约束（即对允许值集合的限制）的数据类型。定义域的用户将成为其拥有者。

如果给定了模式名（例如`CREATE DOMAIN myschema.mydomain ...`），则该域会在指定模式中创建。否则它会在当前模式中创建。域名在其所在模式中的现有类型和域之间必须唯一。

域适合把字段上的常见约束抽象到单一位置进行维护。例如，
若有多个表都包含电子邮件地址列，并且都需要同一个 CHECK 约束来验证地址语法，那么定义一个域会比在每个表上分别设置该约束更合适。

参数

name

要创建的域名（可选地带模式限定）。

data_type

该域的底层数据类型。它可以包含数组说明符。

collation

该域的可选排序规则。如果未指定排序规则，则使用底层数据类型的默认排序规则。
如果指定了`COLLATE`，则底层类型必须是一种支持排序规则的数据类型。

`DEFAULT` *expression*

`DEFAULT`子句为该域数据类型的列指定默认值。
该值可以是任意不含变量的表达式（但不允许子查询）。
默认表达式的数据类型必须与该域的数据类型匹配。如果未指定默认值，则默认值为空值。

默认表达式会在任何未为该列指定值的插入操作中使用。如果为某个特定列定义了默认值，
它就会覆盖与该域关联的任何默认值。反过来，
域默认值又会覆盖与底层数据类型关联的任何默认值。

`CONSTRAINT constraint_name`

约束的可选名称。如果未指定，系统会生成一个名称。

`NOT NULL`

该域的值通常不允许为空值。不过，
如果给该域赋一个已成为空值的匹配域类型值（例如通过左外连接，或 `INSERT INTO tab (domcol) VALUES ((SELECT domcol FROM tab WHERE false))`），带此约束的域仍可能取到空值。

`NULL`

该域的值允许为空值。这是默认行为。

该子句仅用于与非标准 SQL 数据库兼容。不鼓励在新应用中使用它。

`CHECK (expression)`

`CHECK`子句指定该域的值必须满足的完整性约束或测试。
每个约束都必须是一个产生布尔结果的表达式。它应使用关键字`VALUE`来引用被测试的值。

当前，`CHECK`表达式不能包含子查询，也不能引用 `VALUE`之外的其他变量。

示例

这个示例创建`us_postal_code`数据类型，然后在一个表定义中使用该类型。
这里使用正则表达式测试来验证该值看起来是否为一个合法的美国邮政编码：

```
CREATE DOMAIN us_postal_code AS TEXT
CHECK(
   VALUE ~ '^\d{5}$'
OR VALUE ~ '^\d{5}-\d{4}$'
);

CREATE TABLE us_snail_addy (
  address_id SERIAL PRIMARY KEY,
  street1 TEXT NOT NULL,
  street2 TEXT,
  street3 TEXT,
  city TEXT NOT NULL,
  postal us_postal_code NOT NULL
);
```

兼容性

命令`CREATE DOMAIN`符合 SQL 标准。

另见

ALTER DOMAIN, DROP DOMAIN

CREATE EXTENSION

CREATE EXTENSION — install an extension

大纲

```
CREATE EXTENSION [ IF NOT EXISTS ] extension_name
    [ WITH ] [ SCHEMA schema_name ]
             [ VERSION version ]
             [ FROM old_version ]
```

描述

`CREATE EXTENSION`将一个新扩展装载到当前数据库中。同名扩展不能已经被装载。

装载一个扩展本质上就是运行该扩展的脚本文件。该脚本通常会创建新的 SQL 对象，例如函数、数据类型、操作符以及索引支持方法。`CREATE EXTENSION`还会记录所有已创建对象的标识，这样在发出`DROP EXTENSION`时就可以将它们一并删除。

加载扩展需要与创建其组件对象所需的权限相同。对于大多数扩展，
这意味着需要超级用户或数据库所有者权限。运行 `CREATE EXTENSION` 的用户将成为扩展的所有者，用于后续的权限检查，以及扩展脚本创建的任何对象的所有者。

参数

`IF NOT EXISTS`

如果已存在同名扩展，则不要抛出错误。这种情况下会发出一条提示。注意，
这并不保证现有扩展与根据当前可用脚本文件本应创建出的扩展有任何相似之处。

`extension_name`

要安装的扩展名称。PostgreSQL将根据 `SHAREDIR/extension/`*`extension_name`*`.control` 文件中的详细信息创建该扩展。

`schema_name`

如果该扩展允许其内容被重定位，则这是安装该扩展所含对象的模式名称。
指定的模式必须已经存在。如果未指定，并且扩展控制文件中也未指定模式，
则使用当前默认的对象创建模式。

请记住，扩展本身不被视为属于任何模式：扩展的名称是非限定名，
并且必须在整个数据库范围内唯一。但是，属于该扩展的对象可以位于模式中。

`version`

要安装的扩展版本。它可以写成标识符，也可以写成字符串常量。
默认版本由扩展控制文件指定。

`old_version`

`FROM` *`old_version`* 必须在且仅在尝试安装一个替换“旧式”模块的扩展时指定，该模块仅是一组未打包为扩展的对象集合。此选项会导致 `CREATE EXTENSION` 运行一个替代安装脚

本，将现有对象吸收进扩展中，而不是创建新对象。请确保 `SCHEMA` 指定的是包含这些现有对象的模式。

*`old_version`*的值由扩展作者决定，如果存在多个可升级为扩展的旧式模块版本，该值可能会有所不同。对于PostgreSQL 9.1 之前版本提供的标准附加模块，将模块更新为扩展形式时应使用`unpackaged`作为*`old_version`*的值。

注解

在使用 `CREATE EXTENSION` 将扩展装载到数据库之前，必须先安装好该扩展的支持文件。关于安装 PostgreSQL 随附扩展的信息，可见额外提供的模块。

当前可供装载的扩展可从系统视图 `pg_available_extensions` 或 `pg_available_extension_versions` 中查出。

关于编写新扩展的信息，见第 35.15 节。

示例

将 hstore 扩展安装到当前数据库中：

```
CREATE EXTENSION hstore;
```

将 PostgreSQL 9.1 之前版本中安装的`hstore`更新为扩展形式：

```
CREATE EXTENSION hstore SCHEMA public FROM unpackaged;
```

请注意指定安装现有`hstore`对象时所用的模式。

兼容性

`CREATE EXTENSION` 是 PostgreSQL 的扩展。

另见

ALTER EXTENSION, DROP EXTENSION

CREATE FOREIGN DATA WRAPPER

CREATE FOREIGN DATA WRAPPER — define a new foreign-data wrapper

大纲

```
CREATE FOREIGN DATA WRAPPER name
    [ HANDLER handler_function | NO HANDLER ]
    [ VALIDATOR validator_function | NO VALIDATOR ]
    [ OPTIONS ( option 'value' [, ... ] ) ]
```

描述

`CREATE FOREIGN DATA WRAPPER`创建一个新的外部数据包装器。定义该外部数据包装器的用户将成为它的拥有者。

外部数据包装器名称在数据库内必须唯一。

只有超级用户能够创建外部数据包装器。

参数

name

要创建的外部数据包装器的名称。

`HANDLER` *handler_function*

handler_function 是一个预先注册的函数的名称，该函数将被调用以取得外部表所需的执行函数。该处理器函数不能接受任何参数，并且其返回类型必须是 `fdw_handler`。

可以创建没有处理器函数的外部数据包装器，但使用这种包装器的外部表只能被声明，不能被访问。

`VALIDATOR` *validator_function*

validator_function 是一个预先注册的函数的名称，该函数将被调用以检查提供给该外部数据包装器的通用选项，以及用于外部服务器、用户映射和使用该外部数据包装器的外部表的选项。如果未指定验证器函数或者指定了 `NO VALIDATOR`，则在创建时不会检查这些选项。（外部数据包装器可能会在运行时忽略无效的选项声明，或者拒绝它们，这取决于其实现。）
验证器函数必须接受两个参数：一个参数的类型为 `text[]`，其中包含系统目录中存储的选项数组；另一个参数的类型为`oid`，其值是包含这些选项的系统目录的 OID。返回类型会被忽略；该函数应当使用 `ereport(ERROR)`报告无效选项。

`OPTIONS (` *option* `'`*value*`' [, ... ] )`

这个子句为新的外部数据包装器指定选项。
允许的选项名和值都依赖于各个外部数据包装器，并通过该外部数据包装器的验证器函数进行验证。选项名必须唯一。

注解

目前，外部数据包装器功能还处于初级阶段。不支持更新外部表，
查询优化也很原始（而且大部分留给包装器自己处理）。

提供了一个内建的外部数据包装器验证器函数：`postgresql_fdw_validator`，它接受与 libpq 连接参数对应的选项。

示例

创建一个无用的外部数据包装器`dummy`：

```
CREATE FOREIGN DATA WRAPPER dummy;
```

用处理器函数`file_fdw_handler`创建一个外部数据包装器 `file`：

```
CREATE FOREIGN DATA WRAPPER file HANDLER file_fdw_handler;
```

用一些选项创建一个外部数据包装器`mywrapper`：

```
CREATE FOREIGN DATA WRAPPER mywrapper
    OPTIONS (debug 'true');
```

兼容性

`CREATE FOREIGN DATA WRAPPER` 符合 ISO/IEC 9075-9（SQL/MED），但 `HANDLER` 和 `VALIDATOR` 子句属于扩展，而标准中的 `LIBRARY` 和 `LANGUAGE` 子句在 PostgreSQL 中尚未实现。

不过需要注意，SQL/MED 功能整体上尚未完全符合该标准。

另见

ALTER FOREIGN DATA WRAPPER, DROP FOREIGN DATA WRAPPER, CREATE SERVER, CREATE USER MAPPING

CREATE FOREIGN TABLE

CREATE FOREIGN TABLE — define a new foreign table

大纲

```
CREATE FOREIGN TABLE [ IF NOT EXISTS ] table_name ( [
  { column_name data_type [ NULL | NOT NULL ] }
    [, ... ]
] )
  SERVER server_name
[ OPTIONS ( option 'value' [, ... ] ) ]
```

描述

`CREATE FOREIGN TABLE`将在当前数据库中创建一个新的外部表。该表归发出该命令的用户所有。

如果给出了模式名（例如 `CREATE FOREIGN TABLE myschema.mytable ...`），则表将在指定模式中创建。否则，它将在当前模式中创建。外部表名必须与同一模式中任何其他关系（表、序列、索引、视图、物化视图或外部表）的名称不同。

`CREATE FOREIGN TABLE`还会自动创建一种数据类型，用以表示与该外部表一行对应的复合类型。因此，外部表名不能与同一模式中任何已有数据类型同名。

参数

`IF NOT EXISTS`

已经存在同名关系时不要抛出错误。这种情况下会发出一个提示。注意，已存在的关系不保证与原本将要创建的关系有任何相似之处。

`table_name`

要创建的表名（可选地带有模式限定）。

`column_name`

要在新表中创建的列名。

`data_type`

该列的数据类型，可以包含数组说明符。有关PostgreSQL支持的数据类型的更多信息，参见第 8 章。

`NOT NULL`

该列不允许包含空值。

`NULL`

该列允许包含空值。这是默认行为。

此子句仅用于与非标准 SQL 数据库兼容。在新应用程序中不建议使用。

server_name

用于该外部表的现有服务器的名称。

OPTIONS (*option* '*value*' [, ...])

与新外部表关联的选项。允许的选项名称和值取决于每个外部数据包装器，并通过外部数据包装器的验证函数进行验证。选项名必须唯一。

示例

创建带有服务器 film_server 的外部表 films：

```
CREATE FOREIGN TABLE films (
    code        char(5) NOT NULL,
    title       varchar(40) NOT NULL,
    did         integer NOT NULL,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute
)
SERVER film_server;
```

兼容性

CREATE FOREIGN TABLE 命令在很大程度上符合 SQL 标准；不过，与 CREATE TABLE 类似，允许使用 NULL 约束和零列外部表。

另见

ALTER FOREIGN TABLE, DROP FOREIGN TABLE, CREATE TABLE, CREATE SERVER

CREATE FUNCTION

CREATE FUNCTION — define a new function

大纲

```
CREATE [ OR REPLACE ] FUNCTION
    name ( [ [ argmode ] [ argname ] argtype [ { DEFAULT | = }
 default_expr ] [, ...] ] )
    [ RETURNS rettype
      | RETURNS TABLE ( column_name column_type [, ...] ) ]
  { LANGUAGE lang_name
    | WINDOW
    | IMMUTABLE | STABLE | VOLATILE
    | CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
    | [ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER
    | COST execution_cost
    | ROWS result_rows
    | SET configuration_parameter { TO value | = value | FROM
 CURRENT }
    | AS 'definition'
    | AS 'obj_file', 'link_symbol'
  } ...
    [ WITH ( attribute [, ...] ) ]
```

描述

CREATE FUNCTION定义一个新函数。CREATE OR REPLACE FUNCTION将创建一个新函数，或者替换现有定义。要定义函数，用户必须具有该语言上的USAGE权限。

如果包含模式名，则函数将在指定的模式中创建。否则将在当前模式中创建。
新函数的名称不能与同一模式中具有相同输入参数类型的任何现有函数匹配。然而，
不同参数类型的函数可以共享一个名称（这称为重载）。

要替换一个现有函数的当前定义，可以使用CREATE OR REPLACE FUNCTION。但不能用这种方式更改函数的名称或者参数类型（如果尝试这样做，实际上就会创建一个新的不同函数）。
此外，CREATE OR REPLACE FUNCTION也不允许更改现有函数的返回类型。要做到这一点，必须删除该函数并重新创建。（使用OUT参数时，这意味着除非删除该函数，否则不能更改任何OUT参数的类型。）

当CREATE OR REPLACE FUNCTION被用来替换一个现有函数时，该函数的拥有权和权限不会改变。所有其他的函数属性会按照该命令中指定的或者隐含的值赋值。
必须拥有（包括成为拥有角色的成员）该函数才能替换它。

如果删除函数后再重新创建，新函数就不再是旧函数的同一实体；
你将必须删除引用旧函数的现有规则、视图、触发器等。使用CREATE OR REPLACE FUNCTION可以在不破坏引用该函数的对象的情况下更改函数定义。此外，ALTER FUNCTION还可用于更改现有函数的大多数辅助属性。

创建该函数的用户将成为该函数的拥有者。

参数

name

要创建的函数名称（可以被模式限定）。

argmode

参数的模式可以是：IN、OUT、INOUT或者VARIADIC。如果省略，则默认为IN。只有OUT参数可以跟在VARIADIC参数之后。此外，OUT和INOUT参数不能与RETURNS TABLE记法一起使用。

argname

参数的名称。某些语言（包括 SQL 和 PL/pgSQL）允许在函数体中使用该名称。对于其他语言，就函数本身而言，输入参数的名称只是额外文档；
但你可以在调用函数时使用输入参数名来提高可读性（见第 4.3 节）。无论如何，输出参数的名称很重要，因为它定义了结果行类型中的列名。（如果省略输出参数的名称，
系统将选择一个默认列名。）

argtype

该函数参数（如果有）的数据类型（可以是模式限定的）。参数类型可以是基础类型、
复合类型或者域类型，也可以引用一个表列的类型。

根据实现语言的不同，也可能允许指定诸如cstring这样的“伪类型”。
伪类型表示实际参数类型要么没有被完整指定，要么不属于普通 SQL 数据类型集合。

写成*table_name*.*column_name*%TYPE即可引用一个列的类型。
使用这种特性有时有助于让函数独立于表定义的变化。

default_expr

如果未指定该参数，则用作默认值的表达式。该表达式必须能被强制转换为该参数的类型。
只有输入参数（包括INOUT）才能有默认值。
所有跟在具有默认值参数之后的输入参数也都必须有默认值。

rettype

该函数的返回数据类型（可以是模式限定的）。返回类型可以是基础类型、
复合类型或者域类型，也可以引用一个表列的类型。根据实现语言的不同，
也可能允许指定诸如cstring这样的“伪类型”。如果函数不应该返回值，
请把返回类型指定为void。

当存在OUT或INOUT参数时，可以省略RETURNS子句。如果写出该子句，
它必须与输出参数所隐含的结果类型一致：如果有多个输出参数，则为RECORD；
如果只有一个输出参数，则为该输出参数的类型。

SETOF修饰符表示该函数将返回一组项，而不是单个项。

写成*table_name*.*column_name*%TYPE即可引用一个列的类型。

column_name

RETURNS TABLE语法中输出列的名称。这实际上是声明一个具名OUT参数的另一种方式，只不过RETURNS TABLE还隐含了RETURNS SETOF。

column_type

RETURNS TABLE语法中的输出列的数据类型。

lang_name

实现该函数所用语言的名称。它可以是SQL、C、internal，
也可以是用户定义的过程语言的名称。为了向后兼容，名称可以用单引号括起。

WINDOW

WINDOW表示该函数是窗口函数而不是普通函数。目前这只对用 C 编写的函数有用。在替换现有函数定义时，不能更改WINDOW属性。

IMMUTABLE
STABLE
VOLATILE

这些属性会告诉查询优化器该函数的行为。最多只能指定其中一个。
如果这些属性都没有出现，则默认假定为VOLATILE。

IMMUTABLE表示该函数不能修改数据库，并且在给定相同参数值时总会返回相同结果；
也就是说，它不会执行数据库查找，
也不会以其他方式使用未直接出现在其参数列表中的信息。如果给出此选项，
任何使用全常量参数对该函数的调用都可以立即替换为该函数值。

STABLE表示该函数不能修改数据库，并且在一次表扫描内，
对于相同参数值会一致地返回相同结果，但其结果可能在不同 SQL 语句之间发生变化。这适用于结果依赖于数据库查找、参数变量（例如当前时区）等的函数。
（对于希望查询由当前命令修改过的行的AFTER触发器，这样做并不合适。）另请注意，
current_timestamp函数族也属于稳定函数，因为它们的值在一个事务内不会变化。

VOLATILE表示该函数的值即使在一次表扫描内也可能发生变化，因此无法进行任何优化。
从这个意义上说，真正不稳定的数据库函数相对较少；一些例子是random()、currval()、timeofday()。但请注意，任何有副作用的函数都必须归类为不稳定，
即使其结果相当可预测，也必须如此，以防其调用被优化掉；例如setval()。

更多细节见第 35.6 节。

CALLED ON NULL INPUT
RETURNS NULL ON NULL INPUT
STRICT

CALLED ON NULL INPUT（默认）表示当某些参数为空值时，仍会正常调用该函数。如果有需要，则由函数作者负责检查空值并作出适当响应。

RETURNS NULL ON NULL INPUT或STRICT表示只要任一参数为空值，该函数总是返回空值。如果指定了这个选项，那么在参数中出现空值时不会执行该函数，
而是自动假定结果为空值。

[EXTERNAL] SECURITY INVOKER
[EXTERNAL] SECURITY DEFINER

SECURITY INVOKER表示该函数将以调用它的用户的权限执行。这是默认设置。SECURITY DEFINER指定该函数将以创建它的用户的权限执行。

为了符合 SQL，允许使用关键字EXTERNAL。但它是可选的，因为与 SQL 不同，这个特性适用于所有函数，而不仅仅是外部函数。

execution_cost

一个正数，给出该函数的估计执行代价，单位为cpu_operator_cost。
如果该函数返回一个集合，则这是每个返回行的代价。如果未指定代价，则对 C 语言和内部函数假定为 1 个单位，对其他所有语言的函数假定为 100 个单位。较大的值会让规划器尽量避免对该函数进行不必要的频繁求值。

result_rows

一个正数，给出规划器应预计该函数返回的行数。
只有函数被声明为返回集合时才允许使用此项。默认假定为 1000 行。

configuration_parameter
value

SET选项会在进入函数时将指定的配置参数设为给定值，并在函数退出时恢复为先前的值。
SET FROM CURRENT会把执行CREATE FUNCTION时该参数的当前值保存下来，作为进入函数时要应用的值。

如果函数附带了SET选项，那么在函数内针对同一变量执行的SET LOCAL命令，其效果会被限制在该函数内部：函数退出时，配置参数先前的值仍会被恢复。不过，普通的SET
命令（不带LOCAL）会覆盖SET选项，就像它会覆盖先前的SET LOCAL命令一样：这类命令的效果会在函数退出后继续保持，除非当前事务被回滚。

关于允许的参数名和值的更多信息，见SET和第 18 章。

definition

一个定义该函数的字符串常量，其含义取决于所用语言。它可以是一个内部函数名称、
一个对象文件的路径、一个 SQL 命令，或者用一种过程语言编写的文本。

使用美元引用（见第 4.1.2.4 节）来书写函数定义字符串通常会更有帮助，而不是使用普通的单引号语法。如果没有美元引用，
函数定义中的任何单引号或者反斜线都必须用双写来转义。

obj_file, *link_symbol*

当 C 语言源代码中的函数名称与 SQL 函数名称不同时，这种形式的 AS 子句用于可动态加载的 C 语言函数。字符串 *obj_file* 是包含该动态可加载对象的文件名，而 *link_symbol* 是函数的链接符号，即 C 语言源代码中函数的名称。如果省略链接符号，则假定它与正在定义的 SQL 函数名称相同。

当重复的CREATE FUNCTION调用引用同一个对象文件时，该文件在每个会话中只装载一次。要卸载并重新装载该文件（例如在开发期间），请启动一个新会话。

attribute

用于指定函数可选信息的历史方式。以下属性可以出现在此处：

isStrict

等同于STRICT或RETURNS NULL ON NULL INPUT。

isCachable

isCachable是IMMUTABLE的已废弃等价写法；出于向后兼容的原因，目前仍接受它。

属性名称不区分大小写。

有关编写函数的详细信息，请参阅第 35.3 节。

重载

PostgreSQL 允许函数重载；也就是说，只要输入参数类型不同，同一个名称就可以用于多个不同的函数。但是，所有函数的 C 名称都必须不同，因此必须为重载的 C 函数指定不同的 C 名称（例如，把参数类型用作 C 名称的一部分）。

如果两个函数具有相同的名称和输入参数类型，它们被认为相同（不考虑任何OUT参数）。
因此这些声明会冲突：

```
CREATE FUNCTION foo(int) ...
CREATE FUNCTION foo(int, out text) ...
```

参数类型列表不同的函数在创建时不会被视为冲突，但如果提供了默认值，
则在使用时可能发生冲突。例如，考虑下面这些声明：

```
CREATE FUNCTION foo(int) ...
CREATE FUNCTION foo(int, int default 42) ...
```

调用foo(10)会失败，因为系统无法确定应该调用哪个函数。

注解

允许使用完整的SQL类型语法来声明函数参数和返回值。不过，CREATE FUNCTION会丢弃带圆括号的类型修饰符（例如numeric类型的精度字段）。因此，CREATE FUNCTION foo (varchar(10)) ...与CREATE FUNCTION foo (varchar) ...完全等同。

在用CREATE OR REPLACE FUNCTION替换现有函数时，更改参数名会受到限制。不能更改已经分配给任何输入参数的名称（但可以给先前没有名称的参数补上名称）。
如果输出参数多于一个，也不能更改输出参数的名称，
因为那会改变描述函数结果的匿名复合类型的列名。这些限制是为了确保函数被替换时，
已有的函数调用不会停止工作。

如果一个函数被声明为带有VARIADIC参数的STRICT函数，则严格性检查测试的是可变参数数组作为一个整体是否非空。如果该数组包含空值元素，仍会调用该函数。

示例

下面给出一些简单示例，帮助你开始使用。有关更多信息和示例，请参见第 35.3 节。

```
CREATE FUNCTION add(integer, integer) RETURNS integer
    AS 'select $1 + $2;'
    LANGUAGE SQL
    IMMUTABLE
    RETURNS NULL ON NULL INPUT;
```

在PL/pgSQL中，使用参数名把一个整数加 1：

```
CREATE OR REPLACE FUNCTION increment(i integer) RETURNS integer AS $$
        BEGIN
                RETURN i + 1;
        END;
$$ LANGUAGE plpgsql;
```

返回一个包含多个输出参数的记录：

```
CREATE FUNCTION dup(in int, out f1 int, out f2 text)
    AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
    LANGUAGE SQL;

SELECT * FROM dup(42);
```

你也可以用一个显式命名的复合类型，更详细地表达同样的意思：

```
CREATE TYPE dup_result AS (f1 int, f2 text);

CREATE FUNCTION dup(int) RETURNS dup_result
    AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
    LANGUAGE SQL;

SELECT * FROM dup(42);
```

另一种返回多列的方法是使用`TABLE`函数：

```
CREATE FUNCTION dup(int) RETURNS TABLE(f1 int, f2 text)
    AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
    LANGUAGE SQL;

SELECT * FROM dup(42);
```

不过，`TABLE`函数与前面的示例不同，因为它实际返回的是一组记录，而不只是单条记录。

安全地编写 SECURITY DEFINER函数

因为 `SECURITY DEFINER` 函数要以创建它的用户的权限执行，所以必须小心确保该函数不会被滥用。出于安全考虑，应将search_path设置为排除任何可被不受信任用户写入的模式。
这可以防止恶意用户创建对象来遮蔽该函数使用的对象。在这方面尤其重要的是临时表模式；默认情况下它最先被搜索，而且通常任何人都可写。
一个安全的安排是强制把临时模式放到搜索顺序的最后。要做到这一点，应把 `pg_temp` 写成 `search_path` 中的最后一项。下面这个函数展示了安全用法：

```
CREATE FUNCTION check_password(uname TEXT, pass TEXT)
RETURNS BOOLEAN AS $$
DECLARE passed BOOLEAN;
BEGIN
        SELECT  (pwd = $2) INTO passed
        FROM    pwds
        WHERE   username = $1;

        RETURN passed;
END;
$$  LANGUAGE plpgsql
    SECURITY DEFINER
```

```
-- Set a secure search_path: trusted schema(s), then 'pg_temp'.
SET search_path = admin, pg_temp;
```

在PostgreSQL 8.3 版本之前，`SET`选项尚不可用，因此较旧的函数可能包含相当复杂的逻辑来保存、设置和恢复`search_path`。使用`SET`选项更容易实现此目的。

还需注意，默认情况下，新创建的函数会将执行权限授予`PUBLIC`（有关更多信息，请参见GRANT）。通常，你会希望将安全定义器函数的使用限制为某些用户。为此，必须撤销默认的`PUBLIC`权限，然后有选择地授予执行权限。
为避免新函数在一段时间内对所有人可访问，应在同一个事务中创建函数并设置权限。例如：

```
BEGIN;
CREATE FUNCTION check_password(uname TEXT, pass TEXT) ... SECURITY
 DEFINER;
REVOKE ALL ON FUNCTION check_password(uname TEXT, pass TEXT) FROM
 PUBLIC;
GRANT EXECUTE ON FUNCTION check_password(uname TEXT, pass TEXT) TO
 admins;
COMMIT;
```

兼容性

`CREATE FUNCTION`命令在 SQL:1999 及更高版本中定义。PostgreSQL版本与其相似，但并不完全兼容。这些属性和可用的不同语言都不具备可移植性。

为与某些其他数据库系统兼容，*`argmode`*可以写在*`argname`*之前或之后，但只有前一种写法符合标准。

SQL 标准没有规定参数默认值。带 `DEFAULT` 关键字的语法来自 Oracle，而且它在某种程度上与标准的精神一致：SQL/PSM 将它用于变量默认值。带 = 的语法用于 T-SQL 和 Firebird。

另见

ALTER FUNCTION, DROP FUNCTION, GRANT, LOAD, REVOKE, createlang

CREATE GROUP

CREATE GROUP — define a new database role

大纲

```
CREATE GROUP name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE role_name [, ...]
    | IN GROUP role_name [, ...]
    | ROLE role_name [, ...]
    | ADMIN role_name [, ...]
    | USER role_name [, ...]
    | SYSID uid
```

描述

`CREATE GROUP` 现在是CREATE ROLE的别名。

兼容性

SQL 标准中没有 `CREATE GROUP` 语句。

另见

CREATE ROLE

CREATE INDEX

CREATE INDEX — define a new index

大纲

```
CREATE [ UNIQUE ] INDEX [ CONCURRENTLY ] [ name ] ON table [
 USING method ]
    ( { column | ( expression ) } [ COLLATE collation ] [ opclass ] [
 ASC | DESC ] [ NULLS { FIRST | LAST } ] [, ...] )
    [ WITH ( storage_parameter = value [, ... ] ) ]
    [ TABLESPACE tablespace ]
    [ WHERE predicate ]
```

描述

`CREATE INDEX` 在指定表的指定列上构建一个索引。索引主要用于提升数据库性能（但使用不当也可能导致性能下降）。

索引的键字段指定为列名，或者指定为写在圆括号中的表达式。如果索引方法支持多列索引，则可以指定多个字段。

索引字段可以是根据表行中一个或多个列值计算得到的表达式。
该特性可用于根据基础数据的某种变换来快速访问数据。例如，在`upper(col)`上计算的索引可让子句 `WHERE upper(col) = 'JIM'`使用索引。

PostgreSQL提供了索引方法 B-树、hash、GiST、SP-GiST 以及 GIN。用户也可以定义自己的索引方法，但这相当复杂。

当`WHERE`子句存在时，会创建一个部分索引。部分索引只包含表中一部分行的索引项，通常这一部分比表的其余部分更适合建立索引。例如，
如果一个表同时包含已开票和未开票订单，而未开票订单只占整个表的一小部分，
但这一部分又经常被访问，就可以只对这部分创建索引来提升性能。另一种可能的应用是将`WHERE`与`UNIQUE`结合使用，以便在表的一个子集上强制唯一性。更多讨论请见第 11.8 节。

`WHERE`子句中使用的表达式只能引用底层表的列，但它可以使用所有列，
而不仅仅是被索引的列。当前，`WHERE`中也禁止使用子查询和聚合表达式。
同样的限制也适用于作为表达式的索引字段。

所有在索引定义中使用的函数和操作符都必须是“不可变的”，也就是说，
它们的结果只能依赖其参数，而不能受任何外部因素影响（例如另一个表的内容或当前时间）。
这种限制确保索引的行为定义明确。要在索引表达式或`WHERE`子句中使用用户定义的函数，
记得在创建该函数时将其标记为不可变。

参数

`UNIQUE`

使系统在创建索引时（如果数据已经存在）以及每次添加数据时，检查表中的重复值。
任何会导致重复项的插入或更新操作都会报错。

`CONCURRENTLY`

使用此选项时，PostgreSQL将在构建索引时不获取任何会阻止对表进行并发插入、更新或删除的锁；而标准索引构建会阻塞对表的写入（但不会阻塞读取），直到构建完成。使用此选项时有几个注意事项需要了解–请参见并发构建索引。

name

要创建的索引名称。此处不能包含模式名称；索引始终创建在其父表所在的同一模式中。如果省略名称，PostgreSQL会根据父表名称和被索引的列名选择合适的名称。

table

要建立索引的表名（可以是模式限定名）。

method

要使用的索引方法的名称。选择包括 `btree`、`hash`、`gist` 和 `gin`。默认方法是 `btree`。

column

一个表列的名称。

expression

一个基于表中一个或多个列的表达式。通常必须像语法中所示那样写在外围圆括号中。不过，如果该表达式是函数调用形式，则可以省略圆括号。

collation

将用于该索引的排序规则名称。默认情况下，索引使用被索引列声明的排序规则，或者被索引表达式的结果排序规则。对于涉及使用非默认排序规则表达式的查询，使用非默认排序规则的索引可能会很有用。

opclass

一个操作符类的名称。详见下文。

`ASC`

指定升序排序（默认）。

`DESC`

指定降序排序。

`NULLS FIRST`

指定把空值排序在非空值前面。在指定`DESC`时，这是默认行为。

`NULLS LAST`

指定把空值排序在非空值后面。在没有指定`DESC`时，这是默认行为。

storage_parameter

索引方法专用存储参数的名称。有关详细信息，请参见索引存储参数。

tablespace

在其中创建索引的表空间。如果未指定，将查阅default_tablespace；
对于临时表上的索引，则查阅temp_tablespaces。

predicate

部分索引的约束表达式。

索引存储参数

可选的 `WITH` 子句为索引指定存储参数。每一种索引方法都有其各自允许的存储参数集合。B-树、哈希和 GiST 索引方法都接受一个参数：

`FILLFACTOR`

索引的填充因子是一个百分比，用于确定索引方法将尝试把索引页填充到多满。对于 B-树，在初始构建索引期间，以及向右扩展索引（添加新的最大键值）时，
叶页都会填充到该百分比。如果之后页面变得完全填满，就会进行拆分，
导致索引效率逐渐下降。B-树使用默认填充因子 90，但可以选择 10 到 100 之间的任意整数值。如果表是静态的，填充因子 100 最适合将索引的物理大小降至最低；但对于频繁更新的表，较小的填充因子更适合减少页面拆分的需要。
其他索引方法以不同但大致类似的方式使用填充因子；默认填充因子因方法而异。

GIN 索引接受不同的参数：

`FASTUPDATE`

此设置控制第 54.3.1 节中描述的快速更新技术的使用。这是一个布尔参数：`ON`启用快速更新，`OFF`禁用快速更新。（如第 18.1 节中所述，允许使用`ON`和`OFF`的其他拼写形式。）默认值为`ON`。

> **注意**
>
> 通过`ALTER INDEX`关闭`fastupdate` 会阻止后续插入进入待处理索引项列表，但这本身不会刷新现有条目。之后可能需要对该表执行`VACUUM`，
> 以确保待处理列表被清空。

并发构建索引

创建索引可能会干扰数据库的正常运行。通常 PostgreSQL会锁住要建立索引的表，阻止其写入，并通过一次扫描完成整个索引构建。其他事务仍可读取该表，但如果它们试图在表中插入、更新或删除行，就会阻塞直到索引构建完成。如果系统是在线生产数据库，
这可能产生严重影响。对非常大的表建立索引可能需要很多小时，即便是较小的表，
索引构建也可能在一段对生产系统而言不可接受的时间内阻止写入者操作。

PostgreSQL支持在不阻止写入的情况下构建索引。这种方法通过在`CREATE INDEX`中指定 `CONCURRENTLY`选项来启用。使用该选项时，PostgreSQL必须对该表执行两次扫描，此外还必须等待所有现有、可能修改或使用该索引的事务结束。因此，
这种方法比标准索引构建需要更多总工作量，完成时间也明显更长。不过，
由于它允许在构建索引期间继续进行正常操作，所以这种方法适合在生产环境中新增索引。
当然，创建索引带来的额外 CPU 和 I/O 负载也可能拖慢其他操作。

在并发索引构建中，索引实际上会在一个事务中录入系统目录，
然后在另外两个事务中执行两次表扫描。每次表扫描之前，
索引构建都必须等待已修改该表的现有事务结束。第二次扫描之后，
索引构建必须等待所有持有早于第二次扫描的快照（参见第 13 章）的事务结束，其中包括其他表上并发索引构建任一阶段使用的事务。最后，索引才可以标记为可用，`CREATE INDEX`命令随之结束。不过即便如此，索引也可能无法立即用于查询：在最坏情况下，
只要还存在早于索引构建开始的事务，就不能使用它。

如果在扫描表时出现问题，例如死锁或唯一索引中的唯一性冲突，`CREATE INDEX`命令将失败，但会留下一个“无效”索引。由于该索引可能不完整，查询时会忽略它；但是，
它仍会带来更新开销。该psql `\d`命令会将此类索引报告为`INVALID`：

```
postgres=# \d tab
       Table "public.tab"
 Column |  Type   | Modifiers
--------+---------+-----------
 col    | integer |
Indexes:
    "idx" btree (col) INVALID
```

在这种情况下，建议的恢复方法是删除索引，然后重新尝试执行`CREATE INDEX CONCURRENTLY`。（另一种可能性是重建索引，使用`REINDEX`。但是，由于`REINDEX`
不支持并发构建，此选项不太有吸引力。）

并发构建唯一索引时的另一项注意事项是，在第二次表扫描开始时，
唯一性约束就已经开始对其他事务生效了。这意味着在该索引可供使用之前，
其他查询就可能报告约束违规，甚至在索引构建最终失败的情况下也是如此。另外，
如果第二次扫描确实失败了，那个“无效”索引之后仍会继续强制执行其唯一性约束。

也支持并发构建表达式索引和部分索引。计算这些表达式时发生的错误，
可能导致与上文所述唯一性约束违规类似的行为。

普通索引构建允许同一表上的其他普通索引构建并行进行，
但一张表上一次只能进行一个并发索引构建。在这两种情况下，
期间都不允许对该表进行其他类型的模式修改。另一个区别是，普通`CREATE INDEX`命令可以在事务块中执行，而`CREATE INDEX CONCURRENTLY`不能。

注解

关于索引何时能被使用、何时不被使用以及什么情况下它们有用的信息请见第 11 章。

> **小心**
>
> 哈希索引操作目前不会写入 WAL 日志，因此如果数据库崩溃时还有未写入的更改，哈希索引可能需要用`REINDEX`重建。此外，在初始基础备份之后，
> 对哈希索引的更改不会通过流复制或基于文件的复制进行复制，
> 因此之后使用这些索引的查询会得到错误的答案。
> 哈希索引在时间点恢复期间也不能被正确恢复。由于这些原因，
> 目前不鼓励使用哈希索引。

目前，只有 B-树、GiST 和 GIN 索引方法支持多列索引。默认最多可指定 32 个字段。（构建PostgreSQL时可以更改此限制。）目前只有 B-树支持唯一索引。

索引的每一列都可以指定一个操作符类。操作符类标识索引用于该列的操作符。例如，四字节整数上的 B-树索引会使用`int4_ops`类；此操作符类包含四字节整数的比较函数。实际上，列数据类型的默认操作符类通常已足够。操作符类存在的主要原因是，某些数据类型可能有多个有意义的排序方式。例如，我们可能希望按绝对值或实部对复数数据类型排序。我们可以为该数据类型定义两个操作符类，并在创建索引时选择合适的类。有关操作符类的更多信息，请参见第 11.9 节和第 35.14 节。

对于支持有序扫描的索引方法（当前只有 B-树），可以指定可选子句 `ASC`、`DESC`、`NULLS FIRST` 和/或`NULLS LAST`来修改索引的排序顺序。由于有序索引可以向前或向后扫描，因此创建单列`DESC`索引通常并无用处 — 常规索引已经提供了这种排序顺序。这些选项的价值在于可以创建与混合排序查询所要求顺序相匹配的多列索引，例如 `SELECT ... ORDER BY x ASC, y DESC`。如果需要在依靠索引避免排序步骤的查询中支持“空值排在低位”，而不是默认的“空值排在高位”行为，那么`NULLS`选项就很有用。

对于大多数索引方法，索引的创建速度取决于maintenance_work_mem的设置。较大的值将会减少索引创建所需的时间，当然不要把它设置得超过实际可用的内存量（那会迫使机器进行交换）。

使用DROP INDEX删除索引。

早期版本的PostgreSQL还提供过一种 R-tree 索引方法。该方法已经被移除，因为它相对于 GiST 方法并无明显优势。如果指定了`USING rtree`，`CREATE INDEX` 会将其解释为`USING gist`，以简化旧数据库向 GiST 的转换。

示例

创建 B-树索引，列为`title`，所在表为`films`：

```
CREATE UNIQUE INDEX title_idx ON films (title);
```

要在表达式`lower(title)`上创建一个索引，以便高效执行不区分大小写的搜索：

```
CREATE INDEX ON films ((lower(title)));
```

（在这个示例中，索引名称被省略，因此系统会选择一个名字，通常为`films_lower_idx`。）

要创建一个具有非默认排序规则的索引：

```
CREATE INDEX title_idx_german ON films (title COLLATE "de_DE");
```

要创建一个具有非默认空值排序顺序的索引：

```
CREATE INDEX title_idx_nulls_low ON films (title NULLS FIRST);
```

要创建一个具有非默认填充因子的索引：

```
CREATE UNIQUE INDEX title_idx ON films (title) WITH (fillfactor = 70);
```

要创建一个禁用快速更新的GIN索引：

```
CREATE INDEX gin_idx ON documents_table USING GIN (locations) WITH (
fastupdate = off);
```

要在表`films`的列`code`上创建一个索引，并让该索引驻留在表空间`indexspace`中：

```
CREATE INDEX code_idx ON films (code) TABLESPACE indexspace;
```

要在点属性上创建一个 GiST 索引，以便能够在转换函数的结果上高效地使用 box 操作符：

```
CREATE INDEX pointloc
    ON points USING gist (box(location,location));
SELECT * FROM points
    WHERE box(location,location) && '(0,0),(1,1)'::box;
```

要在不阻止对表执行写操作的情况下创建索引：

```
CREATE INDEX CONCURRENTLY sales_quantity_index ON sales_table (
quantity);
```

兼容性

`CREATE INDEX`是 PostgreSQL的语言扩展。SQL 标准中没有关于索引的规定。

另见

ALTER INDEX, DROP INDEX

CREATE LANGUAGE

CREATE LANGUAGE — define a new procedural language

大纲

```
CREATE [ OR REPLACE ] [ PROCEDURAL ] LANGUAGE name
CREATE [ OR REPLACE ] [ TRUSTED ] [ PROCEDURAL ] LANGUAGE name
    HANDLER call_handler [ INLINE inline_handler ] [
 VALIDATOR valfunction ]
```

描述

CREATE LANGUAGE在PostgreSQL数据库中注册一种新的过程语言。随后，可以用这种新语言定义函数和触发器过程。

> **注意**
>
> 从PostgreSQL 9.1 开始，大多数过程语言都已变成“扩展”，因此应使用CREATE EXTENSION而不是CREATE LANGUAGE来安装。现在直接使用CREATE LANGUAGE应仅限于扩展安装脚本。如果数据库中有一个“裸”语言（例如升级的结果），可以使用CREATE EXTENSION *langname* FROM unpackaged将其转换为扩展。

CREATE LANGUAGE实际上是将该语言名称与负责执行用这种语言编写的函数的处理器函数关联起来。关于语言处理器的更多信息，参见第 49 章。

CREATE LANGUAGE命令有两种形式。第一种形式中，用户只提供所需语言的名称，PostgreSQL服务器会查询pg_pltemplate系统目录，以确定正确的参数。第二种形式中，用户在提供语言名称的同时也提供语言参数。第二种形式可用来创建未在pg_pltemplate中定义的语言，但这种做法被认为已经过时。

当服务器在pg_pltemplate目录中找到给定语言名称的条目时，即使命令包含语言参数，也会使用目录中的数据。这种行为简化了旧转储文件的装载，因为它们很可能包含关于语言支持函数的过时信息。

通常，用户必须具有PostgreSQL超级用户权限才能注册新语言。不过，如果语言列在pg_pltemplate目录中，并且标记为允许数据库所有者创建（tmpldbacreate为真），数据库所有者就可以在该数据库中注册这种新语言。默认情况下，数据库所有者可以创建受信任的语言，但超级用户可以通过修改pg_pltemplate的内容来调整这一点。语言的创建者成为其所有者，以后可以删除它、重命名它或将它分配给新的所有者。

CREATE OR REPLACE LANGUAGE会创建新语言，或替换现有定义。如果语言已经存在，则会根据指定值或从pg_pltemplate获取的值更新其参数，但不会改变语言的所有权和权限设置，并且假定用该语言编写的任何现有函数仍然有效。除了满足创建语言的常规权限要求外，用户还必须是超级用户或现有语言的所有者。REPLACE主要用于确保语言存在。如果语言具有pg_pltemplate条目，则REPLACE实际上不会改变现有定义的任何内容，除非出现少见的情况：该pg_pltemplate条目在语言创建后被修改过。

参数

TRUSTED

TRUSTED指定该语言不会让用户获得其原本无权访问的数据访问能力。
如果在注册该语言时省略这个关键字，只有拥有
PostgreSQL超级用户权限的用户才能使用该语言创建新函数。

PROCEDURAL

这是一个噪声词。

name

新过程语言的名称。语言名称不区分大小写。
该名称必须在与数据库中的其他语言名称不同。

为了向后兼容，名称可以用单引号括起来。

HANDLER *call_handler*

call_handler 是一个先前注册的函数名称，它会被调用来执行该过程语言中的函数。过程语言的调用处理器必须使用某种编译型语言（例如 C）编写，并采用版本 1 调用约定。它必须在PostgreSQL 中注册为一个不带参数且返回language_handler类型的函数。该返回类型是一种占位符类型，仅用于把该函数标识为调用处理器。

INLINE *inline_handler*

inline_handler 是一个先前注册的函数名称，它会被调用来执行该语言中的匿名代码块（DO命令）。如果没有指定*inline_handler* 函数，则该语言不支持匿名代码块。该处理器函数必须接受一个 internal类型的参数，该参数将是 DO命令的内部表示，并且它通常返回void。处理器的返回值会被忽略。

VALIDATOR *valfunction*

valfunction 是一个先前注册的函数名称，在创建该语言中的新函数时会调用它来验证该新函数。如果没有指定验证器函数，那么新函数在创建时不会被检查。
验证器函数必须接受一个oid类型的参数，它将是待创建函数的 OID，并且它通常返回void。

验证器函数通常会检查函数体的语法正确性，但它也可以检查函数的其他属性，
例如该语言是否支持某些参数类型。要报告错误，验证器函数应使用ereport()函数。
函数的返回值会被忽略。

TRUSTED选项和支持函数名称会被忽略，前提是服务器在pg_pltemplate
中有指定语言名称的条目。

注解

createlang程序是对CREATE LANGUAGE命令的一个简单包装。它使从 shell 命令行安装过程语言变得更加容易。

使用DROP LANGUAGE删除过程语言，或者更好的方法是使用droplang程序。

系统目录pg_language（见第 45.27 节）记录当前已安装语言的信息。此外，createlang有一个可以列出已安装语言的选项。

要用某种过程语言创建函数，用户必须拥有该语言的 USAGE权限。默认情况下，对于受信任的语言，USAGE会授予PUBLIC（即所有人）。如果需要，也可以撤销这项授权。

过程语言是各个数据库本地的。不过，可以将某种语言安装到 template1数据库中，这样它就会在之后创建的所有数据库中自动可用。

如果服务器在pg_pltemplate中没有该语言的条目，则调用处理器函数、
内联处理器函数（如果有）和验证器函数（如果有）必须已经存在。但如果有条目，
则这些函数不必事先存在；如果数据库中不存在它们，就会自动定义它们。
（如果安装中没有实现该语言的共享库，这可能导致CREATE LANGUAGE失败。）

在 7.3 之前的PostgreSQL版本中，必须将处理器函数声明为返回占位符类型opaque，而不是language_handler。为了支持装载旧转储文件，CREATE LANGUAGE会接受声明为返回opaque的函数，但会发出提示，并将函数声明的返回类型改为language_handler。

示例

创建任何标准过程语言的首选方式很简单：

```
CREATE LANGUAGE plperl;
```

对于 pg_pltemplate 目录中未知的语言，需要类似以下的命令序列：

```
CREATE FUNCTION plsample_call_handler() RETURNS language_handler
    AS '$libdir/plsample'
    LANGUAGE C;
CREATE LANGUAGE plsample
    HANDLER plsample_call_handler;
```

兼容性

CREATE LANGUAGE是一种 PostgreSQL扩展。

另见

ALTER LANGUAGE, CREATE FUNCTION, DROP LANGUAGE, GRANT, REVOKE, createlang, droplang

CREATE OPERATOR

CREATE OPERATOR — define a new operator

大纲

```
CREATE OPERATOR name (
    PROCEDURE = function_name
    [, LEFTARG = left_type ] [, RIGHTARG = right_type ]
    [, COMMUTATOR = com_op ] [, NEGATOR = neg_op ]
    [, RESTRICT = res_proc ] [, JOIN = join_proc ]
    [, HASHES ] [, MERGES ]
)
```

描述

`CREATE OPERATOR`定义一个新的操作符 *name*。定义该操作符的用户将成为其拥有者。如果给出了模式名，该操作符将在指定的模式中创建；否则将在当前模式中创建。

操作符名称是一个字符序列，最多包含 `NAMEDATALEN`-1 个字符（默认为 63 个），字符取自以下列表：

+ - * / < > = ~ ! @ # % ^ & | ` ?

名称的选择有以下几项限制：

- `--`和`/*`不能出现在操作符名称的任何位置，因为它们会被视为注释的开始。
- 多字符操作符名称不能以`+`或`-` 结尾，除非该名称中还至少包含下列字符之一：

  ~ ! @ # % ^ & | ` ?

  例如，`@-`是允许的操作符名称，而 `*-`则不允许。这一限制使 PostgreSQL能够在不要求记号之间必须有空格的情况下解析符合 SQL 规范的命令。
- 使用`=>`作为操作符名称的做法已被弃用。未来版本可能完全禁止这种用法。

输入`!=`时会被映射为`<>`，因此这两个名称始终等效。

必须至少定义`LEFTARG`和`RIGHTARG`中的一个。对于二元操作符，必须同时定义两者。
对于右一元操作符，只应定义`LEFTARG`；对于左一元操作符，只应定义`RIGHTARG`。

*function_name*过程必须事先已用`CREATE FUNCTION`定义，并且其定义必须接受正确数量（一个或两个）的指定类型参数。

其他子句用于指定可选的操作符优化属性。其含义详见第 35.13 节。

参数

name

要定义的操作符名称。允许使用的字符请见上文。该名称可以带模式限定，例如`CREATE OPERATOR myschema.+ (...)`。如果不带模式限定，该操作符将在当前模式中创建。同一

模式中的两个操作符如果针对不同的数据类型进行操作，可以具有相同的名称。这被称为重载。

function_name

用于实现该操作符的函数。

left_type

该操作符左操作数的数据类型（如果有）。对于左一元操作符，应省略此选项。

right_type

该操作符右操作数的数据类型（如果有）。对于右一元操作符，应省略此选项。

com_op

该操作符的交换子。

neg_op

该操作符的求反器。

res_proc

该操作符的限制选择度估计函数。

join_proc

该操作符的连接选择度估算函数。

HASHES

表示该操作符可以支持哈希连接。

MERGES

表示该操作符可以支持归并连接。

要在*com_op*或其他可选参数中给出带模式限定的操作符名称，请使用OPERATOR()语法，例如：

```
COMMUTATOR = OPERATOR(myschema.===) ,
```

注解

更多信息请参见第 35.12 节。

无法在CREATE OPERATOR中指定一个操作符的词法优先级，因为解析器的优先级行为是硬写在代码中的。详见第 4.1.6 节。

废弃的选项SORT1、SORT2、LTCMP以及GTCMP
以前被用来指定与可归并连接的操作符相关联的排序操作符名称。现在已不再需要这样做，因为相关操作符的信息会通过查找 B-树操作符族来获得。如果给出这些选项中的任意一个，除了会隐式将MERGES设为真之外，其余都会被忽略。

使用DROP OPERATOR从数据库中删除用户定义的操作符。使用ALTER OPERATOR修改数据库中的操作符。

示例

以下命令定义一个新的面积相等操作符，适用的数据类型为 `box`：

```
CREATE OPERATOR === (
    LEFTARG = box,
    RIGHTARG = box,
    PROCEDURE = area_equal_procedure,
    COMMUTATOR = ===,
    NEGATOR = !==,
    RESTRICT = area_restriction_procedure,
    JOIN = area_join_procedure,
    HASHES, MERGES
);
```

兼容性

`CREATE OPERATOR`是 PostgreSQL扩展。在 SQL 标准中没有用户定义操作符的规定。

另见

ALTER OPERATOR, CREATE OPERATOR CLASS, DROP OPERATOR

CREATE OPERATOR CLASS

CREATE OPERATOR CLASS — define a new operator class

大纲

```
CREATE OPERATOR CLASS name [ DEFAULT ] FOR TYPE data_type
  USING index_method [ FAMILY family_name ] AS
  {  OPERATOR strategy_number operator_name [ ( op_type, op_type ) ] [
 FOR SEARCH | FOR ORDER BY sort_family_name ]
   | FUNCTION support_number [ ( op_type [ , op_type ] ) ] function_
name ( argument_type [, ...] )
   | STORAGE storage_type
  } [, ... ]
```

描述

CREATE OPERATOR CLASS创建一个新的操作符类。操作符类定义特定数据类型如何用于索引。它指定某些操作符为该数据类型和索引方法承担特定角色或“策略”。
它还指定为索引列选择该操作符类时，索引方法要使用的支持函数。在创建操作符类之前，必须先定义它使用的所有操作符和函数。

如果给出了一个模式名称，那么该操作符类会被创建在指定模式中。否则，它会被创建在当前模式中。
同一模式中的两个操作符类只有在被用于不同的索引方法时才可以具有相同的名称。

定义操作符类的用户将成为其拥有者。目前，创建者必须是超级用户。（之所以有此限制，是因为错误的操作符类定义可能会使服务器混乱，甚至崩溃。）

CREATE OPERATOR CLASS当前不会检查操作符类定义是否包括该索引方法所要求的所有操作符和函数，也不会检查这些操作符和函数是否构成一个自洽的集合。
定义一个合法的操作符类是用户的责任。

相关的操作符类可以分组为操作符族。要把一个新的操作符类加入现有操作符族，可以在CREATE OPERATOR CLASS中指定FAMILY选项。如果没有这个选项，新类会被放入一个与其同名的操作符族中（如果该操作符族尚不存在，就会创建它）。

进一步的信息可参考第 35.14 节。

参数

name

要创建的操作符类的名称。该名称可以被模式限定。

DEFAULT

如果出现，该操作符类将成为其数据类型的默认操作符类。
对于某个特定的数据类型和索引方法，最多只能有一个默认操作符类。

data_type

该操作符类对应的列数据类型。

index_method

该操作符类对应的索引方法名称。

family_name

要把该操作符类加入其中的现有操作符族名称。如果没有指定，
则使用一个与该操作符类同名的操作符族（如果它还不存在则创建）。

strategy_number

与该操作符类关联的操作符在此索引方法中的策略号。

operator_name

与该操作符类关联的操作符名称（可以是模式限定的）。

op_type

在OPERATOR子句中，表示该操作符的操作数数据类型，或者用NONE
表示一个左一元或右一元操作符。在通常情况下，
如果操作数数据类型与操作符类的数据类型相同，则可以省略操作数数据类型。

在 FUNCTION 子句中，表示该函数意图支持的操作数数据类型；如果它不同于函数的输入数据类型（对于 B-树和哈希索引），或者不同于该类的数据类型（对于 GIN 和 GiST 索引），则需要在这里指定。这些默认值总是正确的，因此在 CREATE OPERATOR CLASS 的 FUNCTION 子句中指定 *op_type* 没有意义，但提供这一选项是为了与 ALTER OPERATOR FAMILY 中的类似语法保持一致。

sort_family_name

一个现有btree操作符族的名称（可以是模式限定的），
它描述与一种排序操作符相关联的排序顺序。

如果既没有指定FOR SEARCH，也没有指定 FOR ORDER BY，则默认值是FOR SEARCH。

support_number

与操作符类关联的函数在索引方法中的支持函数编号。

function_name

作为该操作符类的索引方法支持函数的函数名称（可带模式限定）。

argument_type

该函数的参数数据类型。

storage_type

实际存储在索引中的数据类型。通常它与列数据类型相同，但是有些索引方法（目前是 GiST 和 GIN）允许它们不同。除非索引方法允许使用不同类型，否则必须省略STORAGE子句。

OPERATOR、FUNCTION和STORAGE 子句可以以任何顺序出现。

注解

因为索引机制在使用函数之前不会检查函数的访问权限，
将一个函数或者操作符包括在一个操作符类中，相当于授予其公共执行权限。
对适合放入操作符类的那类函数而言，这通常不是问题。

操作符不应该由 SQL 函数定义。SQL 函数很有可能会被内联到调用查询中，这会妨碍优化器识别该查询匹配一个索引。

在PostgreSQL 8.4 之前，`OPERATOR`子句可以包含`RECHECK`选项。现在不再支持该选项，因为索引操作符是否“有损”会在运行时动态确定。
这样可以高效处理操作符可能有损也可能无损的情况。

示例

下面的示例为数据类型`_int4`（`int4`数组）定义了一个 GiST 索引操作符类。完整示例见intarray模块。

```
CREATE OPERATOR CLASS gist__int_ops
    DEFAULT FOR TYPE _int4 USING gist AS
        OPERATOR        3       &&,
        OPERATOR        6       = (anyarray, anyarray),
        OPERATOR        7       @>,
        OPERATOR        8       <@,
        OPERATOR        20      @@ (_int4, query_int),
        FUNCTION        1       g_int_consistent (internal, _int4,
 int, oid, internal),
        FUNCTION        2       g_int_union (internal, internal),
        FUNCTION        3       g_int_compress (internal),
        FUNCTION        4       g_int_decompress (internal),
        FUNCTION        5       g_int_penalty (internal, internal,
 internal),
        FUNCTION        6       g_int_picksplit (internal, internal),
        FUNCTION        7       g_int_same (_int4, _int4, internal);
```

兼容性

`CREATE OPERATOR CLASS`是一种 PostgreSQL扩展。在 SQL 标准中没有`CREATE OPERATOR CLASS`语句。

另见

ALTER OPERATOR CLASS, DROP OPERATOR CLASS, CREATE OPERATOR FAMILY, ALTER OPERATOR FAMILY

CREATE OPERATOR FAMILY

CREATE OPERATOR FAMILY — define a new operator family

大纲

```
CREATE OPERATOR FAMILY name USING index_method
```

描述

`CREATE OPERATOR FAMILY`创建一个新的操作符族。操作符族定义了一组相关的操作符类，以及可能还有一些与这些操作符类兼容、
但对任何单个索引的正常工作都不是必需的额外操作符和支持函数。
（那些对索引必不可少的操作符和函数应归入相应的操作符类，而不是作为操作符族中的“松散”成员。通常，单一数据类型的操作符属于操作符类，
而跨数据类型的操作符则可以作为操作符族中的松散成员，
该操作符族包含这两种数据类型对应的操作符类。）

新的操作符族最初是空的。应当通过随后发出的`CREATE OPERATOR CLASS`命令向其中添加所包含的操作符类，并可选择通过`ALTER OPERATOR FAMILY`命令添加“松散”操作符及其对应的支持函数。

如果给出了模式名称，该操作符族会被创建在指定的模式中。否则，它会被创建在当前模式中。
只有当同一模式中的两个操作符族用于不同的索引方法时，它们才能具有相同的名称。

定义操作符族的用户将成为其拥有者。当前，创建该对象的用户必须是超级用户。
（之所以这样限制，是因为错误的操作符族定义可能会使服务器混乱，甚至导致崩溃。）

更多信息请参见第 35.14 节。

参数

name

要创建的操作符族名称。该名称可以是模式限定的。

index_method

该操作符族所对应的索引方法名称。

兼容性

`CREATE OPERATOR FAMILY`是一种 PostgreSQL扩展。SQL 标准中没有`CREATE OPERATOR FAMILY`语句。

另见

ALTER OPERATOR FAMILY, DROP OPERATOR FAMILY, CREATE OPERATOR CLASS, ALTER OPERATOR CLASS, DROP OPERATOR CLASS

CREATE ROLE

CREATE ROLE — define a new database role

大纲

```
CREATE ROLE name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | REPLICATION | NOREPLICATION
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE role_name [, ...]
    | IN GROUP role_name [, ...]
    | ROLE role_name [, ...]
    | ADMIN role_name [, ...]
    | USER role_name [, ...]
    | SYSID uid
```

描述

`CREATE ROLE`向 PostgreSQL数据库集簇中新增一个角色。角色是一种可以拥有数据库对象并拥有数据库权限的实体；根据其使用方式，它既可以视为“用户”，也可以视为“组”，或者兼具这两种身份。有关用户管理和认证的信息，请参见第 20 章和第 19 章。要使用该命令，你必须拥有 `CREATEROLE`权限，或者是数据库超级用户。

注意，角色是在数据库集簇级别定义的，因此在该集簇中的所有数据库里都有效。

参数

name

新角色的名称。

`SUPERUSER`
`NOSUPERUSER`

这些子句决定新角色是否为“超级用户”；超级用户可以绕过数据库中的所有访问限制。超级用户身份具有危险性，只有在确有需要时才应使用。要创建新的超级用户，你自己必须先是超级用户。若未指定，默认值是 `NOSUPERUSER`。

CREATEDB
NOCREATEDB

这些子句定义角色创建数据库的能力。如果指定CREATEDB，
则允许正在定义的角色创建新数据库。指定NOCREATEDB会禁止角色创建数据库。
如果未指定，默认值为NOCREATEDB。

CREATEROLE
NOCREATEROLE

这些子句决定是否允许角色创建新角色（即执行CREATE ROLE）。具有CREATEROLE权限的角色还可以更改和删除其他角色。如果未指定，默认值为NOCREATEROLE。

CREATEUSER
NOCREATEUSER

这些子句是 SUPERUSER 和 NOSUPERUSER 的一种已过时但仍被接受的写法。注意，它们不等价于 CREATEROLE，尽管可能有人会天真地这么认为！

INHERIT
NOINHERIT

这些子句决定角色是否“继承”其所属角色的权限。具有INHERIT
属性的角色可以自动使用直接或间接所属的所有角色获得的任何数据库权限。没有INHERIT
时，成为另一个角色的成员只授予对该角色执行SET ROLE的能力；另一个角色的权限只有在执行该操作后才可用。如果未指定，默认值为INHERIT。

LOGIN
NOLOGIN

这些子句决定是否允许角色登录；也就是说，在客户端连接期间，
是否可以将该角色作为初始会话授权名称。具有 LOGIN 属性的角色可以视为用户。没有该属性的角色可用于管理数据库权限，但通常意义上不属于用户。如果未指定，默认值为NOLOGIN，但通过其另一种拼写 CREATE USER 调用 CREATE ROLE 时除外。

REPLICATION
NOREPLICATION

这些子句决定是否允许角色发起流复制，以及让系统进入和退出备份模式。具有
REPLICATION属性的角色权限非常高，因此该属性只应赋予实际用于复制的角色。
如果未指定，默认值为NOREPLICATION。

CONNECTION LIMIT *connlimit*

如果该角色可以登录，此项指定该角色可建立多少并发连接。-1（默认值）表示不限制。
注意，只有普通连接会计入此限制。

PASSWORD *password*

设置角色的密码。（密码只对具有LOGIN属性的角色有用，但即使没有该属性，
你仍然可以为角色定义密码。）如果你不打算使用密码认证，可以省略此选项。
如果未指定密码，密码将被设为 null，并且该用户的密码认证将始终失败。也可以显式写为PASSWORD NULL。

ENCRYPTED
UNENCRYPTED

这些关键字控制密码是否以加密形式存储在系统目录中。（如果两者都未指定，默认行为由配置参数password_encryption决定。）如果给出的密码字符串已经是 MD5 加密格式，那么无论指定的是ENCRYPTED还是 UNENCRYPTED，它都会按原样以加密形式存储（因为系统无法解密指定的已加密密码字符串）。这样便可在转储/恢复期间重新装载已加密的密码。

注意，较老的客户端可能缺少对 MD5 认证机制的支持，而处理加密存储的密码需要这种机制。

VALID UNTIL '*timestamp*'

VALID UNTIL子句设置一个日期和时间，在此之后角色的密码将不再有效。如果省略此子句，密码将始终有效。

IN ROLE *role_name*

IN ROLE子句列出一个或多个现有角色，新角色会立即作为新成员加入这些角色。（请注意，无法通过此选项将新角色添加为管理员；要实现这一点，请使用单独的GRANT命令。）

IN GROUP *role_name*

IN GROUP是IN ROLE的已废弃写法。

ROLE *role_name*

ROLE子句列出一个或多个现有角色，这些角色会自动作为新角色的成员加入。（实际上，这会使新角色成为一个“组”。）

ADMIN *role_name*

ADMIN子句类似于ROLE，但指定的角色会以WITH ADMIN OPTION加入新角色，从而获得向其他人授予该角色成员资格的权利。

USER *role_name*

USER子句是ROLE子句的已废弃写法。

SYSID *uid*

SYSID子句会被忽略，但为向后兼容而接受。

注解

使用ALTER ROLE更改角色属性，并使用DROP ROLE删除角色。CREATE ROLE指定的所有属性都可以由后续的ALTER ROLE命令修改。

向作为组使用的角色添加和移除成员，推荐使用GRANT和REVOKE。

VALID UNTIL子句只为密码定义过期时间，而不是为角色本身定义过期时间。特别是，使用非基于密码的认证方式登录时，不会强制执行该过期时间。

INHERIT属性控制可授予权限的继承（即数据库对象的访问权限和角色成员资格）。它不适用于CREATE ROLE和ALTER ROLE设置的特殊角色属性。例如，即使设置了INHERIT，成

为拥有CREATEDB权限的角色成员也不会立即获得创建数据库的能力；在创建数据库之前，必须先通过SET ROLE成为该角色。

INHERIT属性出于向后兼容的原因而成为默认值：在PostgreSQL的早期版本中，用户总能访问其所属组的所有权限。不过，NOINHERIT更接近 SQL 标准规定的语义。

请谨慎对待CREATEROLE权限。CREATEROLE角色的权限不存在继承这一概念。这意味着，即使某个角色没有特定权限但被允许创建其他角色，它仍可轻易创建一个权限不同于自身的角色（创建具有超级用户权限的角色除外）。例如，如果角色“user”具有CREATEROLE权限但没有CREATEDB权限，它仍然可以创建一个具有CREATEDB权限的新角色。因此，应将拥有CREATEROLE权限的角色视为几乎等同超级用户的角色。

PostgreSQL包含一个程序createuser，其功能与CREATE ROLE相同（实际上它就是调用此命令），但它可以从命令 shell 中运行。

CONNECTION LIMIT选项只是近似地被强制执行；如果两个新会话在几乎同一时刻启动，而此时该角色只剩下一个连接“槽”，则两者都可能失败。此外，该限制从不对超级用户强制执行。

用此命令指定未加密密码时必须谨慎。密码会以明文形式传输到服务器，也可能被记录在客户端命令历史或服务器日志中。而命令createuser传输的是加密密码。此外，psql包含 \password命令，可用于稍后安全地更改密码。

示例

创建一个可登录但没有密码的角色：

```
CREATE ROLE jonathan LOGIN;
```

创建一个带密码的角色：

```
CREATE USER davide WITH PASSWORD 'jw8s0F4';
```

（CREATE USER与CREATE ROLE相同，只是它隐含指定了LOGIN。）

创建一个密码有效期截止到 2004 年底的角色。当 2005 年的第一秒过去后，该密码就不再有效。

```
CREATE ROLE miriam WITH LOGIN PASSWORD 'jw8s0F4' VALID UNTIL '2005-01
-01';
```

创建一个可以创建数据库并管理角色的角色：

```
CREATE ROLE admin WITH CREATEDB CREATEROLE;
```

兼容性

该CREATE ROLE语句在 SQL 标准中有定义，但该标准只要求以下语法

```
CREATE ROLE name [ WITH ADMIN role_name ]
```

多个初始管理员，以及CREATE ROLE的所有其他选项，都是PostgreSQL扩展。

SQL 标准定义了用户和角色的概念，但将它们视为不同的概念，并将定义用户的所有命令留给各数据库实现自行规定。在PostgreSQL中，我们选择将用户和角色统一为一种实体。因此，与标准相比，角色具有更多可选属性。

将用户赋予 `NOINHERIT` 属性，而将角色赋予 `INHERIT` 属性，最接近 SQL 标准规定的行为。

参见

SET ROLE, ALTER ROLE, DROP ROLE, GRANT, REVOKE, createuser

CREATE RULE

CREATE RULE — define a new rewrite rule

大纲

```
CREATE [ OR REPLACE ] RULE name AS ON event
    TO table [ WHERE condition ]
    DO [ ALSO | INSTEAD ] { NOTHING | command | ( command ; command .
.. ) }
```

描述

CREATE RULE定义一条应用于指定表或视图的新规则。CREATE OR REPLACE RULE将创建一条新规则，或者替换同一个表上同名的现有规则。

PostgreSQL规则系统允许定义在数据库表上执行插入、更新或删除时的替代动作。粗略地说，当在某个给定表上执行某个命令时，一条规则会导致额外的命令被执行。另一种情况是，INSTEAD 规则可以用另一个命令替换给定命令，或者让某个命令完全不执行。规则也用于实现SQL 视图。必须认识到，规则实际上是一种命令转换机制，或者说是命令宏。这种转换发生在命令开始执行之前。如果你真正需要的是对每个物理行独立触发的操作，
那么你很可能应该使用触发器，而不是规则。有关规则系统的更多信息见第 37 章。

目前，ON SELECT规则必须是无条件的INSTEAD规则，其动作必须由单个SELECT命令组成。因此，ON SELECT规则实际上会把表变为视图，其可见内容是该规则的SELECT命令返回的行，而不是表中原来存储的内容（如果有）。直接编写CREATE VIEW命令，被认为比创建真实表并为其定义ON SELECT规则的风格更好。

可以通过定义 ON INSERT、ON UPDATE 和 ON DELETE 规则（或其中足以满足你需求的任意子集），把视图上的更新动作替换为对其他表的适当更新，从而营造出可更新视图的效果。如果想支持 INSERT RETURNING 等功能，那么务必在这些规则中的每一条里都放入合适的 RETURNING 子句。此外，也可以使用 INSTEAD OF 触发器实现可更新视图（见CREATE TRIGGER）。

如果你尝试对复杂视图更新使用有条件规则，
这里有一个陷阱：对于你希望允许在该视图上执行的每一种动作，必须都有一条无条件的INSTEAD规则。如果该规则是有条件的，或者不是 INSTEAD规则，那么系统仍会拒绝执行该更新动作，因为它认为在某些情况下最终可能还是会尝试在该视图的虚拟表上执行该动作。
如果你想用有条件规则处理所有有用的情况，可以加上一条无条件的 DO INSTEAD NOTHING 规则，以确保系统明白它永远不会被要求去更新这个虚拟表。然后，把这些有条件规则改成非INSTEAD规则；在它们适用的情况下，它们会附加在默认的 INSTEAD NOTHING动作之上。（不过，这种方法当前仍无法支持 RETURNING查询。）

参数

name

要创建的规则的名称。它必须与同一个表上任何其他规则的名称相区分。
同一个表上同一种事件类型的多条规则会按名称的字母顺序应用。

event

事件是SELECT、INSERT、UPDATE或者 DELETE之一。

table

规则适用的表或者视图的名称（可以是模式限定的）。

condition

任意的SQL条件表达式（返回 boolean）。条件表达式不能引用NEW和 OLD之外的任何表，并且不能包含聚合函数。

INSTEAD

INSTEAD表示这些命令应当替代原始命令执行。

ALSO

ALSO表示这些命令应当在原始命令之外执行。

如果既未指定ALSO，也未指定 INSTEAD，则默认值为ALSO。

command

组成规则动作的命令。可用的命令有SELECT、INSERT、UPDATE、DELETE或者NOTIFY。

在*condition*和 *command*中，可以使用特殊表名 NEW和OLD来引用目标表中的值。在ON INSERT和ON UPDATE规则中，NEW可用于引用待插入或待更新的新行。在 ON UPDATE和ON DELETE规则中，OLD可用于引用将被更新或删除的现有行。

注解

要在表上创建或者修改规则，必须是表的拥有者。

在视图上的INSERT、UPDATE或 DELETE规则中，可以添加一个输出视图各列的 RETURNING子句。如果该规则分别由 INSERT RETURNING、UPDATE RETURNING 或DELETE RETURNING命令触发，那么这个子句将用于计算输出。当规则由不带RETURNING的命令触发时，该规则的 RETURNING子句会被忽略。当前实现只允许无条件的 INSTEAD规则包含RETURNING；此外，对于同一事件的所有规则，总共最多只能有一个RETURNING子句。（这可以确保只有一个候选RETURNING子句可用于计算结果。）
如果任何可用规则中都没有 RETURNING子句，那么视图上的RETURNING查询将被拒绝。

避免循环规则非常重要。例如，尽管下面的两条规则定义都被 PostgreSQL所接受，SELECT命令将导致 PostgreSQL报告错误，因为规则会发生递归展开：

```
CREATE RULE "_RETURN" AS
    ON SELECT TO t1
    DO INSTEAD
        SELECT * FROM t2;

CREATE RULE "_RETURN" AS
    ON SELECT TO t2
    DO INSTEAD
        SELECT * FROM t1;

SELECT * FROM t1;
```

当前，如果一个规则动作包含一个NOTIFY命令，该NOTIFY命令将被无条件执行，也就是说，即使没有任何应当应用该规则的行，也会发出NOTIFY。例如：

```
CREATE RULE notify_me AS ON UPDATE TO mytable DO ALSO NOTIFY mytable;

UPDATE mytable SET name = 'foo' WHERE id = 42;
```

在`UPDATE`期间会发送一个`NOTIFY`事件，无论是否存在匹配条件`id = 42`的行。这是一个实现限制，未来版本中可能会修复。

兼容性

`CREATE RULE`是一种 PostgreSQL语言扩展，整个查询重写系统也是如此。

CREATE SCHEMA

CREATE SCHEMA — define a new schema

大纲

```
CREATE SCHEMA schema_name [ AUTHORIZATION user_name ] [ schema_element
 [ ... ] ]
CREATE SCHEMA AUTHORIZATION user_name [ schema_element [ ... ] ]
```

描述

CREATE SCHEMA在当前数据库中创建一个新模式。该模式名必须与当前数据库中任何现有模式的名称不同。

模式本质上是一个命名空间：它包含具名对象（表、数据类型、函数和操作符），这些对象的名称可以与其他模式中的对象重名。访问具名对象时，可以在其名称前加上模式名作为前缀来“限定”名称，或者设置一个包含所需模式的搜索路径。指定非限定对象名的CREATE命令会在当前模式中创建该对象（即搜索路径最前面的模式，可通过函数current_schema确定）。

CREATE SCHEMA还可以选择包含子命令，以便在新模式中创建对象。这些子命令基本上会被当作在创建模式之后单独发出的命令来处理。不过，如果使用了AUTHORIZATION子句，则所有创建的对象都将归该用户所有。

参数

schema_name

要创建的模式名称。如果省略，*user_name*将被用作模式名。该名称不能以pg_开头，因为这类名称保留给系统模式。

user_name

将拥有新模式的用户的角色名。如果省略，则默认为执行该命令的用户。要创建由另一个角色拥有的模式，你必须是该角色的直接或间接成员，或者是超级用户。

schema_element

定义要在该模式中创建的对象的 SQL 语句。当前，只有CREATE TABLE、CREATE VIEW、CREATE INDEX、CREATE SEQUENCE、CREATE TRIGGER和GRANT可作为 CREATE SCHEMA中的子句使用。其他类型的对象可以在模式创建之后通过单独的命令创建。

注解

要创建一个模式，执行该命令的用户必须拥有当前数据库的CREATE 权限（当然，超级用户可以绕过这项检查）。

示例

创建一个模式：

```
CREATE SCHEMA myschema;
```

为用户joe创建一个模式，该模式也将被命名为 joe：

```
CREATE SCHEMA AUTHORIZATION joe;
```

创建一个模式并且在其中创建一个表和视图：

```
CREATE SCHEMA hollywood
    CREATE TABLE films (title text, release date, awards text[])
    CREATE VIEW winners AS
        SELECT title, release FROM films WHERE awards IS NOT NULL;
```

请注意，各个子命令末尾都不带分号。

下面是实现相同结果的一种等效写法：

```
CREATE SCHEMA hollywood;
CREATE TABLE hollywood.films (title text, release date, awards text[]
);
CREATE VIEW hollywood.winners AS
    SELECT title, release FROM hollywood.films WHERE awards IS NOT
 NULL;
```

兼容性

SQL 标准允许在CREATE SCHEMA中使用 DEFAULT CHARACTER SET子句，并允许使用比PostgreSQL当前所接受的更多子命令类型。

SQL 标准规定，CREATE SCHEMA中的子命令可以按任意顺序出现。当前的PostgreSQL实现并不能处理子命令中所有前向引用的情况；有时可能需要重新排列子命令的顺序，以避免前向引用。

根据 SQL 标准，模式的拥有者总是拥有其中的所有对象。PostgreSQL允许模式包含由模式拥有者之外的用户所拥有的对象。只有当模式拥有者把其模式上的CREATE权限授予其他人，或者超级用户选择在该模式中创建对象时，才会发生这种情况。

另见

ALTER SCHEMA, DROP SCHEMA

CREATE SEQUENCE

CREATE SEQUENCE — define a new sequence generator

大纲

```
CREATE [ TEMPORARY | TEMP ] SEQUENCE name [ INCREMENT [ BY ] increment
 ]
    [ MINVALUE minvalue | NO MINVALUE ] [ MAXVALUE maxvalue | NO
 MAXVALUE ]
    [ START [ WITH ] start ] [ CACHE cache ] [ [ NO ] CYCLE ]
    [ OWNED BY { table.column | NONE } ]
```

描述

`CREATE SEQUENCE`创建一个新的序列号发生器。这会创建并初始化一个名为*name*的特殊单行表。该发生器归发出该命令的用户所有。

如果给定了模式名，则序列会在指定模式中创建。否则，它会在当前模式中创建。
临时序列存在于一个特殊模式中，因此创建临时序列时不能指定模式名。
序列名必须不同于同一模式中任何其他关系（表、序列、索引、视图、物化视图或外部表）的名称。

创建序列后，可以使用函数 `nextval`、`currval`以及 `setval`来操作该序列。这些函数的文档见第 9.15 节。

虽然不能直接更新序列，但可以使用如下查询：

```
SELECT * FROM name;
```

来查看序列的参数和当前状态。特别是，序列的 `last_value`字段会显示任一会话最近分配的值。（当然，如果其他会话正在调用`nextval`，那么到该值被打印出来时，它可能已经过时。）

参数

`TEMPORARY` or `TEMP`

如果指定该项，则该序列对象只为当前会话创建，并会在会话退出时自动删除。
临时序列存在期间，已有的同名永久序列在本会话中不可见，
除非使用模式限定名来引用它们。

name

要创建的序列的名称（可以是模式限定的）。

increment

可选子句`INCREMENT BY` *increment*指定加到当前序列值上以生成新值的增量。正值生成升序序列，负值生成降序序列。默认值为 1。

minvalue
`NO MINVALUE`

可选子句`MINVALUE` *minvalue*确定序列可生成的最小值。如果未提供此子句，或者指定了`NO MINVALUE`，则使用默认值。升序序列和降序序列的默认值分别为 1 和 $-2^{63}-1$。

maxvalue
NO MAXVALUE

可选子句MAXVALUE *maxvalue*确定序列的最大值。如果未提供此子句，或者指定了 NO MAXVALUE，则使用默认值。升序序列和降序序列的默认值分别为 2^{63}-1 和 -1。

start

可选子句START WITH *start* 允许序列从任意值开始。升序序列的默认起始值为*minvalue*，降序序列则为*maxvalue*。

cache

可选子句CACHE *cache*指定要预分配并存储在内存中以加快访问的序列号数量。最小值为 1（一次只能生成一个值，即不缓存），这也是默认值。

CYCLE
NO CYCLE

CYCLE选项允许序列在升序序列达到*maxvalue*、或降序序列达到*minvalue*时分别回绕。如果达到该限制，下一个生成的数字将分别是*minvalue*或*maxvalue*。

如果指定NO CYCLE，则当序列达到其限值后，任何对 nextval的调用都会返回错误。如果既未指定 CYCLE也未指定NO CYCLE，则默认使用NO CYCLE。

OWNED BY *table.column*
OWNED BY NONE

OWNED BY选项使序列与特定表列关联，从而在删除该列（或整个表）时，也自动删除该序列。指定表必须与序列具有相同的所有者，并且位于同一模式中。默认值OWNED BY NONE表示不存在这种关联。

注解

使用DROP SEQUENCE删除序列。

序列基于bigint算术，因此其范围不能超过八字节整数的范围（-9223372036854775808 到 9223372036854775807）。

如果对一个将由多个会话并发使用的序列对象设置了大于 1 的*cache*，则可能得到意想不到的结果。每个会话会在一次访问该序列对象时分配并缓存连续的序列值，并相应增加该序列对象的 last_value。随后，该会话中接下来的 *cache*-1 次nextval 调用会直接返回预分配的值，而不接触序列对象。因此，任何在会话中已分配但未使用的数字都会在会话结束时丢失，从而在序列中造成“空洞”。

此外，虽然多个会话保证会分配到互不相同的序列值，但从所有会话整体来看，这些值的生成顺序可能会乱序。例如，当*cache*设置为 10 时，会话 A 可能保留值 1..10 并返回nextval=1；然后会话 B 可能保留值 11..20 并返回nextval=11，而此时会话 A 尚未生成 nextval=2。因此，当*cache*设置为 1 时，可以安全地假定 nextval值按顺序生成；当*cache*设置大于 1 时，只应假定 nextval值彼此不同，而不应假定它们严格按顺序生成。此外，last_value会反映任一会话最近保留的值，不管该值是否已经由nextval返回。

另一个需要注意的问题是，在这种序列上执行的setval 不会立刻被其他会话察觉，直到它们用完自己缓存的所有预分配值为止。

示例

创建一个名为`serial`的升序序列，从 101 开始：

```
CREATE SEQUENCE serial START 101;
```

从该序列中取出下一个数字：

```
SELECT nextval('serial');

 nextval
---------
     101
```

从该序列中取出下一个数字：

```
SELECT nextval('serial');

 nextval
---------
     102
```

在`INSERT`命令中使用该序列：

```
INSERT INTO distributors VALUES (nextval('serial'), 'nothing');
```

在一次`COPY FROM`之后更新序列值：

```
BEGIN;
COPY distributors FROM 'input_file';
SELECT setval('serial', max(id)) FROM distributors;
END;
```

兼容性

`CREATE SEQUENCE`符合SQL 标准，但下列情况除外：

- 不支持标准的`AS <data type>`表达式。
- 获取下一个值是使用`nextval()`函数，而不是标准的 `NEXT VALUE FOR`表达式完成的。
- `OWNED BY`子句是PostgreSQL扩展。

另见

ALTER SEQUENCE, DROP SEQUENCE

CREATE SERVER

CREATE SERVER — define a new foreign server

大纲

```
CREATE SERVER server_name [ TYPE 'server_type' ] [ VERSION 'server_
version' ]
    FOREIGN DATA WRAPPER fdw_name
    [ OPTIONS ( option 'value' [, ... ] ) ]
```

描述

`CREATE SERVER`定义一个新的外部服务器。定义该服务器的用户将成为其拥有者。

外部服务器通常封装外部数据包装器用来访问外部数据资源的连接信息。
还可以通过用户映射指定额外的、特定于用户的连接信息。

服务器名称在数据库内必须唯一。

要创建服务器，必须在所使用的外部数据包装器上具有`USAGE`权限。

参数

server_name

要创建的外部服务器的名称。

server_type

可选的服务器类型。

server_version

可选的服务器版本。

fdw_name

管理该服务器的外部数据包装器的名称。

`OPTIONS ( option 'value' [, ... ] )`

这个子句指定服务器的选项。这些选项通常定义该服务器的连接细节，
但实际的名称和值取决于该服务器的外部数据包装器。

注解

在使用 dblink 模块（见dblink）时，可以把外部服务器的名称用作dblink_connect函数的一个参数，以指示连接参数。更多示例也请参见那里。要以这种方式使用它，
必须在该外部服务器上具有 `USAGE` 权限。

示例

创建一个使用内建外部数据包装器 `default` 的服务器 `foo`：

```
CREATE SERVER foo FOREIGN DATA WRAPPER "default";
```

创建一个使用外部数据包装器 `pgsql` 的服务器 `myserver`：

```
CREATE SERVER myserver FOREIGN DATA WRAPPER pgsql OPTIONS (host 'foo',
 dbname 'foodb', port '5432');
```

兼容性

`CREATE SERVER`符合 ISO/IEC 9075-9 (SQL/MED)。

另见

ALTER SERVER, DROP SERVER, CREATE FOREIGN DATA WRAPPER, CREATE USER MAPPING

CREATE TABLE

CREATE TABLE — define a new table

大纲

```
CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } | UNLOGGED ] TABLE [
 IF NOT EXISTS ] table_name ( [
  { column_name data_type [ COLLATE collation ] [ column_constraint [
 ... ] ]
    | table_constraint
    | LIKE parent_table [ like_option ... ] }
    [, ... ]
] )
[ INHERITS ( parent_table [, ... ] ) ]
[ WITH ( storage_parameter [= value] [, ... ] ) | WITH OIDS | WITHOUT
 OIDS ]
[ ON COMMIT { PRESERVE ROWS | DELETE ROWS | DROP } ]
[ TABLESPACE tablespace ]

CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } | UNLOGGED ] TABLE [
 IF NOT EXISTS ] table_name
    OF type_name [ (
  { column_name WITH OPTIONS [ column_constraint [ ... ] ]
    | table_constraint }
    [, ... ]
) ]
[ WITH ( storage_parameter [= value] [, ... ] ) | WITH OIDS | WITHOUT
 OIDS ]
[ ON COMMIT { PRESERVE ROWS | DELETE ROWS | DROP } ]
[ TABLESPACE tablespace ]

where column_constraint is:

[ CONSTRAINT constraint_name ]
{ NOT NULL |
  NULL |
  CHECK ( expression ) |
  DEFAULT default_expr |
  UNIQUE index_parameters |
  PRIMARY KEY index_parameters |
  REFERENCES reftable [ ( refcolumn ) ] [ MATCH FULL | MATCH PARTIAL |
 MATCH SIMPLE ]
    [ ON DELETE action ] [ ON UPDATE action ] }
[ DEFERRABLE | NOT DEFERRABLE ] [ INITIALLY DEFERRED | INITIALLY
 IMMEDIATE ]

and table_constraint is:

[ CONSTRAINT constraint_name ]
{ CHECK ( expression ) |
  UNIQUE ( column_name [, ... ] ) index_parameters |
```

```
  PRIMARY KEY ( column_name [, ... ] ) index_parameters |
  EXCLUDE [ USING index_method ] ( exclude_element WITH operator [, ..
. ] ) index_parameters [ WHERE ( predicate ) ] |
  FOREIGN KEY ( column_name [, ... ] ) REFERENCES reftable [ (
 refcolumn [, ... ] ) ]
    [ MATCH FULL | MATCH PARTIAL | MATCH SIMPLE ] [ ON DELETE action ]
 [ ON UPDATE action ] }
[ DEFERRABLE | NOT DEFERRABLE ] [ INITIALLY DEFERRED | INITIALLY
 IMMEDIATE ]

and like_option is:

{ INCLUDING | EXCLUDING } { DEFAULTS | CONSTRAINTS | INDEXES | STORAGE
 | COMMENTS | ALL }

index_parameters in UNIQUE, PRIMARY KEY, and EXCLUDE constraints are:

[ WITH ( storage_parameter [= value] [, ... ] ) ]
[ USING INDEX TABLESPACE tablespace ]

exclude_element in an EXCLUDE constraint is:

{ column | ( expression ) } [ opclass ] [ ASC | DESC ] [ NULLS { FIRST
 | LAST } ]
```

描述

`CREATE TABLE` 将在当前数据库中创建一个新的、初始为空的表。该表归发出该命令的用户所有。

如果给出了模式名（例如 `CREATE TABLE myschema.mytable ...`），则表将在指定模式中创建。否则，它将在当前模式中创建。临时表存在于一个特殊模式中，
因此创建临时表时不能给出模式名。表名必须与同一模式中任何其他表、序列、索引、
视图或外部表的名称不同。

`CREATE TABLE` 还会自动创建一种数据类型，用以表示与该表一行对应的复合类型。因此，表名不能与同一模式中任何已有数据类型同名。

可选的约束子句指定插入或更新要成功时，新行或更新后的行必须满足的约束（测试）。
约束是一种 SQL 对象，可用多种方式帮助定义表中允许的值集合。

定义约束有两种方式：表约束和列约束。列约束作为列定义的一部分定义。
表约束则不绑定到特定列，并且可以涵盖多个列。每个列约束也都可以写成表约束；
当约束只影响一列时，列约束只是一种书写上的方便。

参数

`TEMPORARY` or `TEMP`

如果指定该选项，表将创建为临时表。临时表会在会话结束时自动删除，
或者也可在当前事务结束时删除（见下文 `ON COMMIT`）。在临时表存在期间，同名的现有永久表对当前会话不可见，除非使用带模式限定的名称引用它们。
在临时表上创建的任何索引也都会自动成为临时索引。

自动清理守护进程不能访问并且因此也不能清理或分析临时表。由于这个原因，应该通过会话的 SQL 命令执行合适的清理和分析操作。例如，如果一个临时表将要被用于复杂的查询，最好在把它填充完毕后在其上运行ANALYZE。

可以选择在 TEMPORARY 或 TEMP 前写上 GLOBAL 或 LOCAL。这在 PostgreSQL 中没有任何区别，但参见兼容性。

UNLOGGED

如果指定了此选项，表将被创建为不记录 WAL 的表。写入不记录 WAL 的表的数据不会写入预写式日志（见第 29 章），这使得它们比普通表快得多。但是，它们不具备崩溃安全性：在崩溃或非正常关闭之后，不记录 WAL 的表会被自动截断。不记录 WAL 的表的内容也不会复制到备库。在不记录 WAL 的表上创建的任何索引也都自动是不记录 WAL 的；但是，目前不支持不记录 WAL 的 GiST 索引，不能在不记录 WAL 的表上创建它们。

IF NOT EXISTS

如果已存在同名关系，则不抛出错误，而是发出一条提示。注意，这并不保证现有关系与本应创建出的关系有任何相似之处。

table_name

要创建的表名（可选地带模式限定）。

OF *type_name*

创建一个类型化表，其结构取自指定的复合类型（名称可以带模式限定）。类型化表与其类型绑定；例如，如果删除该类型（使用DROP TYPE ... CASCADE），该表也会被删除。

创建类型化表时，列的数据类型由底层复合类型决定，不由CREATE TABLE命令指定。不过，CREATE TABLE命令可以为表添加默认值和约束，并指定存储参数。

column_name

要在新表中创建的列名。

data_type

列的数据类型。这可以包括数组说明符。有关 PostgreSQL 支持的数据类型的更多信息，请参见第 8 章。

COLLATE *collation*

COLLATE子句为列指定排序规则（该列必须属于支持排序规则的数据类型）。如果未指定，则使用列数据类型的默认排序规则。

INHERITS (*parent_table* [, ...])

可选的 INHERITS 子句指定一组表，新表将自动从中继承所有列。

使用 INHERITS 会在新子表与其父表之间建立持久关系。对父表的模式修改通常也会传播到子表，且默认情况下，对父表的扫描会包含子表的数据。

如果同一列名出现在多个父表中，除非这些父表中该列的数据类型全部匹配，否则会报错。如果没有冲突，这些重复列会合并为新表中的单个列。
如果新表的列名列表中包含一个同样来自继承的列名，其数据类型也必须与继承列匹配，

并且列定义会合并为一个。如果新表显式为该列指定了默认值，
该默认值会覆盖继承声明中的任何默认值。否则，
任何为该列指定默认值的父表都必须指定相同的默认值，否则会报错。

CHECK 约束基本上按与列相同的方式合并：如果多个父表和/或新表定义中包含同名的 CHECK 约束，则这些约束必须拥有相同的检查表达式，否则会报错。同名且表达式相同的约束将被合并为一份。注意，新表中未命名的 CHECK 约束永远不会被合并，因为系统总会为它选择一个唯一的名称。

列的 STORAGE 设置也会从父表复制过来。

LIKE *parent_table* [*like_option* ...]

LIKE 子句指定一个表，新表会自动从中复制所有列名、数据类型及其非空约束。

与 INHERITS 不同，新表和原表在创建完成后就完全脱钩了。对原表的修改不会应用到新表，也不可能在扫描原表时包含新表的数据。

只有指定INCLUDING DEFAULTS时，才会复制所复制列定义的默认表达式。默认行为是不包含默认表达式，因此新表中的所复制列将具有空默认值。请注意，
复制调用数据库修改函数（例如nextval）的默认值，
可能会在原表和新表之间创建功能性关联。

非空约束始终会复制到新表。只有指定INCLUDING CONSTRAINTS时，才会复制CHECK约束。列约束和表约束之间不作区分。

只有指定INCLUDING INDEXES时，才会在新表上创建原表的索引、PRIMARY KEY、UNIQUE和EXCLUDE约束。新索引和约束的名称按照默认规则选择，与原名称无关。
（此行为可以避免新索引可能发生名称重复错误。）

只有指定INCLUDING STORAGE时，才会复制所复制列定义的STORAGE设置。默认行为是不包含STORAGE设置，因此新表中复制的列使用其类型特定的默认设置。有关STORAGE
设置的更多信息，请参见第 55.2 节。

只有指定INCLUDING COMMENTS时，才会复制所复制列、约束和索引的注释。默认行为是不包含注释，因此新表中复制的列和约束没有注释。

INCLUDING ALL是INCLUDING DEFAULTS INCLUDING CONSTRAINTS INCLUDING INDEXES INCLUDING STORAGE INCLUDING COMMENTS的简写形式。

请注意，与INHERITS不同，LIKE复制的列和约束不会与同名的列和约束合并。
如果显式指定了相同的名称，或在另一个LIKE子句中指定了相同的名称，则会报错。

CONSTRAINT *constraint_name*

列约束或表约束的可选名称。如果约束被违反，错误消息中会包含该约束名，因此诸如 col must be positive 这样的约束名可以向客户端应用传达有用的约束信息。（若约束名中包含空格，则需要用双引号指定。）如果未指定约束名，系统会生成一个。

NOT NULL

该列不允许包含空值。

NULL

该列允许包含空值。这是默认情况。

该子句仅为兼容非标准 SQL 数据库而提供，不建议在新应用中使用。

`CHECK ( expression )`

`CHECK` 子句指定一个产生布尔结果的表达式。要使插入或更新成功，新行或更新后的行必须满足该表达式。计算结果为 TRUE 或 UNKNOWN 的表达式视为成功。如果插入或更新操作中的任何一行得到 FALSE 结果，就会抛出错误异常，并且插入或更新不会修改数据库。作为列约束指定的检查约束只应引用该列的值，而出现在表约束中的表达式可以引用多个列。

当前，`CHECK` 表达式不能包含子查询，也不能引用当前行的列之外的变量。

`DEFAULT default_expr`

`DEFAULT`子句为其所在列定义的列指定默认数据值。
该值可以是任何不含变量的表达式（不允许子查询，
也不允许交叉引用当前表中的其他列）。
默认表达式的数据类型必须与该列的数据类型匹配。

默认值表达式会用于任何未为该列指定值的插入操作。如果一列没有默认值，则默认值为 null。

`UNIQUE` (column constraint)
`UNIQUE ( column_name [, ... ] )` (table constraint)

`UNIQUE` 约束指定表中一列或多列组成的一组只能包含唯一值。表级唯一约束的行为与列级唯一约束相同，只是它还能跨越多列。因此，
该约束要求任意两行在这些列中至少有一列不同。

对于唯一约束，空值不被视为相等。

每个唯一表约束必须命名一组列，这组列不能与该表主键约束所命名的列集合相同。（否则它就只是把同一个约束列出了两次。）

`PRIMARY KEY` (column constraint)
`PRIMARY KEY ( column_name [, ... ] )` (table constraint)

`PRIMARY KEY` 约束指定表的一列或多列只能包含唯一（不重复）且非空的值。无论作为列约束还是表约束，一个表都只能指定一个主键。

主键约束所引用的列集合应不同于同一表上定义的任何唯一约束所引用的列集合。（否则，该唯一约束是冗余的，会被丢弃。）

`PRIMARY KEY` 强制的数据约束与 `UNIQUE` 和 `NOT NULL` 的组合相同。不过，将一组列标识为主键还会为模式设计提供元数据，因为主键意味着其他表可以将这组列作为行的唯一标识符来依赖。

`EXCLUDE [ USING index_method ] ( exclude_element WITH operator [, ... ] ) index_parameters [ WHERE ( predicate ) ]`

`EXCLUDE` 子句定义一个排他约束。它保证如果任意两行在指定列或表达式上使用指定操作符进行比较，这些比较不会全部返回 `TRUE`。如果所有指定操作符都测试相等，这就等价于 `UNIQUE` 约束，尽管普通唯一约束会更快。不过，排他约束可以指定比简单相等更一般的约束。例如，你可以通过使用 `&&` 操作符来指定一个约束，使表中不存在两个包含重叠圆的行（见第 8.8 节）。

排他约束通过索引实现，因此每个指定的操作符都必须与索引访问方法*index_method*的适当操作符类关联（见第 11.9 节）。这些操作符必须满足交换律。每个*exclude_element*都可以选择指定操作符类和/或排序选项；详见CREATE INDEX。

访问方法必须支持 amgettuple（见第 52 章）；目前这意味着不能使用 GIN。虽然允许在排他约束上使用 B-树或哈希索引，但这没有什么意义，因为普通唯一约束能把它做得更好。因此在实践中，访问方法总是 GiST。

predicate 允许你只在表的一个子集上指定排他约束；在内部，这会创建一个部分索引。注意，predicate 周围的圆括号是必需的。

REFERENCES *reftable* [(*refcolumn*)] [MATCH *matchtype*] [ON DELETE *action*] [ON UPDATE *action*] (column constraint)
FOREIGN KEY (*column* [, ...]) REFERENCES *reftable* [(*refcolumn* [, ...])] [MATCH *matchtype*] [ON DELETE *action*] [ON UPDATE *action*] (table constraint)

这些子句指定外键约束，
要求新表的一个或多个列组成的列组只能包含与被引用表某一行的被引用列值匹配的值。
如果省略*refcolumn*列表，则使用*reftable*的主键。
被引用列必须是被引用表中不可延迟的唯一约束或主键约束的列。请注意，
不能在临时表和永久表之间定义外键约束。

插入到引用列中的值会按照给定的匹配类型，与被引用表及其被引用列中的值进行匹配。
共有三种匹配类型：MATCH FULL、MATCH PARTIAL 和 MATCH SIMPLE（默认值）。
MATCH FULL 不允许多列外键中的某一列为空，除非所有外键列都为空。MATCH SIMPLE 允许某些外键列为空而外键的其他部分不为空。MATCH PARTIAL 目前尚未实现。

此外，当被引用列中的数据发生变化时，会对本表列中的数据执行某些操作。ON DELETE子句指定删除被引用表中的被引用行时要执行的操作。同样，ON UPDATE子句指定将被引用表中的被引用列更新为新值时要执行的操作。如果行被更新，但被引用列实际上没有变化，
则不执行任何操作。除NO ACTION检查以外的引用操作都不能延迟，即使该约束声明为可延迟也是如此。每个子句可以指定以下操作：

NO ACTION

产生错误，指出删除或更新会违反外键约束。如果该约束被延迟，
则会在约束检查时仍存在引用行的情况下产生这个错误。这是默认操作。

RESTRICT

产生错误，指出删除或更新会违反外键约束。这与NO ACTION相同，但检查不能延迟。

CASCADE

分别删除任何引用已删除行的行，或将引用列的值更新为被引用列的新值。

SET NULL

将引用列设置为空值。

SET DEFAULT

将引用列设置为它们的默认值。

如果被引用列经常变化，可以考虑在外键列上添加索引，
使与该外键列关联的引用动作能够更高效地执行。

DEFERRABLE
NOT DEFERRABLE

这控制约束是否可以延迟。不可延迟的约束会在每条命令之后立即检查。
可延迟约束的检查可以推迟到事务结束（使用SET CONSTRAINTS命令）。NOT DEFERRABLE是默认值。目前，只有UNIQUE、PRIMARY KEY、EXCLUDE和REFERENCES（外键）约束接受此子句。NOT NULL和CHECK约束不可延迟。

INITIALLY IMMEDIATE
INITIALLY DEFERRED

如果约束可延迟，则此子句指定检查约束的默认时间。如果约束为INITIALLY IMMEDIATE，则在每条语句之后检查。这是默认值。如果约束为INITIALLY DEFERRED，则仅在事务结束时检查。可以使用SET CONSTRAINTS命令更改约束检查时间。

WITH (*storage_parameter* [= *value*] [, ...])

该子句为表或索引指定可选的存储参数；详情见存储参数。表的WITH子句还可以包含OIDS=TRUE（或仅写OIDS），以指定为新表的行分配 OID（对象标识符）；也可以包含OIDS=FALSE，以指定行不应具有 OID。如果未指定OIDS，默认设置取决于default_with_oids配置参数。（如果新表继承自任何具有 OID 的表，则会强制使用OIDS=TRUE，即使命令指定了OIDS=FALSE也是如此。）

如果显式或隐式指定了OIDS=FALSE，新表将不存储 OID，也不会为插入其中的行分配 OID。通常认为这样做是值得的，因为它会减少 OID 的消耗，从而推迟 32 位 OID 计数器回卷。一旦计数器回卷，就不能再假定 OID 是唯一的，这会大大降低它们的用途。此外，不在表中包含 OID 可以减少在磁盘上存储该表所需的空间，在大多数机器上每行可减少 4 字节，从而略微提高性能。

要在表创建后移除其 OID，请使用ALTER TABLE。

WITH OIDS
WITHOUT OIDS

这些是过时的语法，分别等价于WITH (OIDS)和WITH (OIDS=FALSE)。如果要同时指定OIDS设置和存储参数，必须使用WITH (...)语法；见上文。

ON COMMIT

可以使用 ON COMMIT 控制临时表在事务块结束时的行为。三种选项如下：

PRESERVE ROWS

在事务结束时不执行任何特殊操作。这是默认行为。

DELETE ROWS

临时表中的所有行都会在每个事务块结束时删除。实际上，
每次提交时都会自动执行一次TRUNCATE。

DROP

在当前事务块结束时删除临时表。

TABLESPACE *tablespace*

*tablespace_name*是要创建新表的表空间名称。如果未指定，则查询default_tablespace；如果表是临时表，则查询temp_tablespaces。

USING INDEX TABLESPACE *tablespace*

该子句允许选择与 UNIQUE、PRIMARY KEY 或 EXCLUDE 约束相关联的索引要创建在哪个表空间中。若未指定，则参考default_tablespace；如果该表是临时表，则参考temp_tablespaces。

存储参数

WITH子句可以为表以及与UNIQUE、PRIMARY KEY或EXCLUDE约束关联的索引指定存储参数。索引的存储参数记载于CREATE INDEX。当前可用于表的存储参数列在下面。除非特别注明，对每个参数而言，都有一个同名的附加参数，其前缀为toast.，可用于控制该表的二级TOAST表（如果有）的行为（有关 TOAST 的更多信息请参见第 55.2 节）。注意，如果没有设置toast.autovacuum_*设置，TOAST 表会从其父表继承autovacuum_*的值。

fillfactor(integer)

表的填充因子是 10 到 100 之间的百分比。100（完全填充）是默认值。指定较小的填充因子时，INSERT操作只将表页填充到指定百分比；每页的剩余空间保留用于更新该页上的行。这样，UPDATE就有机会将行的更新副本放在与原行相同的页面上，这比放在不同页面上更高效。对于从不更新其条目的表，完全填充是最佳选择；但对于频繁更新的表，适合使用较小的填充因子。不能为 TOAST 表设置此参数。

autovacuum_enabled, toast.autovacuum_enabled(boolean)

在特定表上启用或禁用自动清理守护进程。如果为真，当已更新或已删除的元组数超过 autovacuum_vacuum_threshold 加上 autovacuum_vacuum_scale_factor 乘以当前估计存在于该关系中的活元组数时，自动清理守护进程将对该表发起一次 VACUUM 操作。类似地，当已插入、已更新或已删除的元组数超过 autovacuum_analyze_threshold 加上 autovacuum_analyze_scale_factor 乘以当前估计存在于该关系中的活元组数时，它将发起一次 ANALYZE 操作。如果为假，则该表不会被自动清理，但为了防止事务 ID 回卷，仍可能对其执行自动清理。有关回卷防护的更多信息，见第 23.1.4 节。请注意，这个变量继承autovacuum设置的值。

autovacuum_vacuum_threshold, toast.autovacuum_vacuum_threshold(integer)

在特定表上发起一次 VACUUM 操作之前，已更新或已删除元组的最小数量。

autovacuum_vacuum_scale_factor, toast.autovacuum_vacuum_scale_factor(float4)

要加到 autovacuum_vacuum_threshold 上的 reltuples 的倍数。

autovacuum_analyze_threshold(integer)

在特定表上发起一次 ANALYZE 操作之前，已插入、已更新或已删除元组的最小数量。

autovacuum_analyze_scale_factor(float4)

要加到 autovacuum_analyze_threshold 上的 reltuples 的倍数。

autovacuum_vacuum_cost_delay, toast.autovacuum_vacuum_cost_delay(integer)

autovacuum_vacuum_cost_delay参数的每表取值。

`autovacuum_vacuum_cost_limit, toast.autovacuum_vacuum_cost_limit (integer)`

autovacuum_vacuum_cost_limit参数的每表取值。

`autovacuum_freeze_min_age, toast.autovacuum_freeze_min_age (integer)`

vacuum_freeze_min_age参数的每表取值。注意，自动清理会忽略大于系统范围autovacuum_freeze_max_age设置一半的每表 `autovacuum_freeze_min_age` 参数。

`autovacuum_freeze_max_age, toast.autovacuum_freeze_max_age (integer)`

autovacuum_freeze_max_age参数的每表取值。注意，自动清理会忽略大于系统范围设置的每表 `autovacuum_freeze_max_age` 参数（它只能设置得更小）。

`autovacuum_freeze_table_age, toast.autovacuum_freeze_table_age (integer)`

vacuum_freeze_table_age参数的每表取值。

注解

不建议在新应用中使用 OID：在可能的情况下，优先使用`SERIAL`或其他序列生成器作为表的主键。不过，如果应用确实使用 OID 来标识表中的特定行，建议在该表的`oid`列上创建唯一约束，以确保即使计数器回卷，表中的 OID 也确实能唯一标识行。不要假定 OID 在不同表之间唯一；如果需要数据库范围的唯一标识符，请组合使用`tableoid`和行 OID。

> **提示**
>
> 对于没有主键的表，不建议使用`OIDS=FALSE`，因为既没有 OID，也没有唯一数据键时，很难标识特定行。

PostgreSQL为每一个唯一约束和主键约束自动创建一个索引来强制唯一性。因此，没有必要显式地为主键列创建一个索引（详见CREATE INDEX）。

在当前的实现中，唯一约束和主键不会被继承。这使得继承与唯一约束的组合相当不实用。

一个表不能有超过 1600 列（实际上，由于元组长度限制，有效的限制通常更低）。

示例

创建表`films`和表`distributors`：

```
CREATE TABLE films (
    code        char(5) CONSTRAINT firstkey PRIMARY KEY,
    title       varchar(40) NOT NULL,
    did         integer NOT NULL,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute
);
```

```
CREATE TABLE distributors (
     did    integer PRIMARY KEY DEFAULT nextval('serial'),
     name   varchar(40) NOT NULL CHECK (name <> '')
);
```

创建一个带二维数组列的表：

```
CREATE TABLE array_int (
    vector  int[][]
);
```

为表films定义一个唯一表约束。唯一表约束可以定义在表的一列或多列上：

```
CREATE TABLE films (
    code        char(5),
    title       varchar(40),
    did         integer,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute,
    CONSTRAINT production UNIQUE(date_prod)
);
```

定义一个列检查约束：

```
CREATE TABLE distributors (
    did     integer CHECK (did > 100),
    name    varchar(40)
);
```

定义一个表检查约束：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    CONSTRAINT con1 CHECK (did > 100 AND name <> '')
);
```

为表films定义一个主键表约束：

```
CREATE TABLE films (
    code        char(5),
    title       varchar(40),
    did         integer,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute,
    CONSTRAINT code_title PRIMARY KEY(code,title)
);
```

为表distributors定义一个主键约束。下面的两个示例是等价的，第一个使用表约束语法，第二个使用列约束语法：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
```

```
    PRIMARY KEY(did)
);

CREATE TABLE distributors (
    did     integer PRIMARY KEY,
    name    varchar(40)
);
```

为列name指定一个字面常量默认值，将列did的默认值设为从某个序列对象中取下一个值，并让modtime的默认值为插入该行的时间：

```
CREATE TABLE distributors (
    name      varchar(40) DEFAULT 'Luso Films',
    did       integer DEFAULT nextval('distributors_serial'),
    modtime   timestamp DEFAULT current_timestamp
);
```

在表distributors上定义两个NOT NULL列约束，其中一个显式指定了名称：

```
CREATE TABLE distributors (
    did     integer CONSTRAINT no_null NOT NULL,
    name    varchar(40) NOT NULL
);
```

为name列定义一个唯一约束：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40) UNIQUE
);
```

同样的唯一约束用表约束指定：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    UNIQUE(name)
);
```

创建同样的表，并为该表及其唯一索引都指定 70% 的填充因子：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    UNIQUE(name) WITH (fillfactor=70)
)
WITH (fillfactor=70);
```

创建表circles，并添加一个排他约束以防任意两个圆重叠：

```
CREATE TABLE circles (
    c circle,
    EXCLUDE USING gist (c WITH &&)
);
```

在表空间diskvol1中创建表cinemas：

```
CREATE TABLE cinemas (
        id serial,
        name text,
        location text
) TABLESPACE diskvol1;
```

创建一个复合类型和一个类型化表：

```
CREATE TYPE employee_type AS (name text, salary numeric);

CREATE TABLE employees OF employee_type (
    PRIMARY KEY (name),
    salary WITH OPTIONS DEFAULT 1000
);
```

兼容性

CREATE TABLE 命令符合 SQL 标准，但有下列例外。

临时表

尽管 CREATE TEMPORARY TABLE 的语法看起来类似于 SQL 标准，但其效果并不相同。按标准，临时表只需定义一次，并会自动存在于每个需要它的会话中（内容初始为空）。而 PostgreSQL 要求每个会话都为每个要使用的临时表发出自己的 CREATE TEMPORARY TABLE 命令。这使不同会话可以出于不同目的使用相同的临时表名；而标准做法则要求给定临时表名的所有实例都必须具有相同的表结构。

标准对临时表行为的定义在实践中被广泛忽略。PostgreSQL 在这一点上的行为与多种其他 SQL 数据库相似。

标准中全局临时表和局部临时表的区分在 PostgreSQL 中并不存在，因为这一区分依赖于模块的概念，而 PostgreSQL 没有这一概念。出于兼容性考虑，PostgreSQL 会接受在临时表声明中使用 GLOBAL 和 LOCAL 关键字，但它们没有效果。

临时表的 ON COMMIT 子句也与 SQL 标准相似，但存在一些差异。如果省略 ON COMMIT 子句，SQL 规定默认行为是 ON COMMIT DELETE ROWS。然而，PostgreSQL 中的默认行为是 ON COMMIT PRESERVE ROWS。SQL 中不存在 ON COMMIT DROP 选项。

非延迟唯一性约束

当 UNIQUE 或 PRIMARY KEY 约束不可延迟时，只要有行被插入或修改，PostgreSQL 就会立刻检查唯一性。SQL 标准规定应只在语句结束时强制唯一性；例如，当单个命令会更新多个键值时，这两者就会产生差异。若要获得符合标准的行为，应将约束声明为 DEFERRABLE 但不延迟（即 INITIALLY IMMEDIATE）。注意，这可能明显慢于立即检查唯一性。

列检查约束

SQL 标准规定，CHECK 列约束只能引用其所作用的列；只有 CHECK 表约束才能引用多列。PostgreSQL 并不强制这一限制；它对列检查约束和表检查约束一视同仁。

EXCLUDE 约束

EXCLUDE 约束类型是 PostgreSQL 的扩展。

NULL “约束”

NULL “约束”（实际上并不是约束）是 PostgreSQL 对 SQL 标准的扩展；提供它是为了与其他一些数据库系统兼容（以及与 NOT NULL 约束保持对称）。由于它本来就是任意列的默认情况，所以它的存在只是噪声。

继承

通过 INHERITS 子句实现的多重继承是 PostgreSQL 的语言扩展。SQL:1999 及后续标准使用不同的语法和语义定义了单继承。PostgreSQL 尚不支持 SQL:1999 风格的继承。

零列的表

PostgreSQL 允许创建没有列的表（例如 CREATE TABLE foo();）。这是对 SQL 标准的扩展，标准不允许零列的表。零列的表本身并不十分有用，但若禁止它们，就会让 ALTER TABLE DROP COLUMN 出现奇怪的特殊情况，因此忽略这一规范限制看起来更整洁。

LIKE 子句

虽然 SQL 标准中存在 LIKE 子句，但 PostgreSQL 接受的许多 LIKE 选项并不在标准中，而标准中的某些选项又没有被 PostgreSQL 实现。

WITH 子句

WITH 子句是 PostgreSQL 的扩展；存储参数和 OID 都不属于标准内容。

表空间

PostgreSQL 的表空间概念不是标准的一部分。因此，TABLESPACE 和 USING INDEX TABLESPACE 子句都是扩展。

类型化表

类型化表实现了 SQL 标准的一个子集。按照标准，类型化表除了具有与底层复合类型相对应的列之外，还应有一个额外的“自引用列”。PostgreSQL 不显式支持自引用列，但使用 OID 功能可以达到相同的效果。

另见

ALTER TABLE, DROP TABLE, CREATE TABLESPACE, CREATE TYPE

CREATE TABLE AS

CREATE TABLE AS — define a new table from the results of a query

大纲

```
CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } | UNLOGGED ]
 TABLE table_name
    [ (column_name [, ...] ) ]
    [ WITH ( storage_parameter [= value] [, ... ] ) | WITH OIDS |
 WITHOUT OIDS ]
    [ ON COMMIT { PRESERVE ROWS | DELETE ROWS | DROP } ]
    [ TABLESPACE tablespace ]
    AS query
    [ WITH [ NO ] DATA ]
```

描述

`CREATE TABLE AS`创建一个表，并用 `SELECT`命令计算得到的数据填充该表。该表的列具有与`SELECT`输出列对应的名称和数据类型（但可以通过显式给出新的列名列表来覆盖列名）。

`CREATE TABLE AS`与创建视图有些相似，但实际上差异很大：它会创建一个新表，并且只执行一次该查询来完成新表的初始填充。这个新表不会跟踪该查询所引用源表的后续变化。
相比之下，视图每次被查询时都会重新执行其定义中的 `SELECT`语句。

参数

`GLOBAL` or `LOCAL`

出于兼容性考虑会被忽略。详见CREATE TABLE。

`TEMPORARY` or `TEMP`

如果指定，将把该表创建为临时表。详见CREATE TABLE。

`UNLOGGED`

如果指定，该表将创建为不记录 WAL 的表。详见CREATE TABLE。

`table_name`

要创建的表名（可选地带模式限定）。

`column_name`

新表中某一列的名称。如果未提供列名，则采用查询输出列的列名。如果表是从 `EXECUTE` 命令创建的，则不能指定列名列表。

`WITH ( storage_parameter [= value] [, ... ] )`

此子句为新表指定可选存储参数；有关更多信息，请参见存储参数。`WITH`
子句还可以包含`OIDS=TRUE`（或仅包含`OIDS`），以指定应为新表的行分配 OID（对象标识符），或者包含`OIDS=FALSE`，以指定这些行不应有 OID。更多信息请参见CREATE TABLE。

WITH OIDS
WITHOUT OIDS

这些是分别等同于WITH (OIDS)和WITH (OIDS=FALSE)的过时语法。如果希望同时指定OIDS设置和存储参数，则必须使用WITH (...)语法；见上文。

ON COMMIT

临时表在事务块结束时的行为可以使用以下选项控制：ON COMMIT。三个选项是：

PRESERVE ROWS

在事务结束时不采取特殊动作。这是默认行为。

DELETE ROWS

临时表中的所有行将在每个事务块结束时删除。本质上，
每次提交时都会执行一次自动的TRUNCATE。

DROP

临时表会在当前事务块结束时被删除。

TABLESPACE *tablespace*

*tablespace_name*是要创建新表的表空间名称。如果未指定，则查询default_tablespace；如果表是临时表，则查询temp_tablespaces。

query

可以是SELECT、TABLE或VALUES命令，也可以是运行已预备的EXECUTESELECT、TABLE或VALUES查询的命令。

WITH [NO] DATA

这个子句指定是否将查询产生的数据复制到新表中。如果不复制，则只复制表结构。
默认会复制数据。

注解

这个命令在功能上类似于SELECT INTO，但更推荐使用它，因为它更不容易与SELECT INTO语法的其他用法相混淆。此外，CREATE TABLE AS提供的功能是SELECT INTO所提供功能的超集。

在 PostgreSQL 8.0 之前，CREATE TABLE AS创建的表总是包含 OID。从 PostgreSQL 8.0 起，CREATE TABLE AS命令允许用户显式指定是否包含 OID。如果未显式指定是否存在 OID，则使用default_with_oids配置变量。而从 PostgreSQL 8.1 起，该变量默认为 false，因此默认行为与 8.0 之前的发行版不再相同。需要在CREATE TABLE AS所创建表中包含 OID 的应用，应显式指定WITH (OIDS)以确保所需行为。

示例

创建一个新表films_recent，其中只包含表 films中的近期条目：

```
CREATE TABLE films_recent AS
  SELECT * FROM films WHERE date_prod >= '2002-01-01';
```

要完整复制一个表，也可以使用TABLE命令的简写形式：

```
CREATE TABLE films2 AS
  TABLE films;
```

创建一个新的临时表films_recent，其中只包含来自以下表的近期条目：films，使用预备语句。新表具有 OID，并将在提交时删除：

```
PREPARE recentfilms(date) AS
  SELECT * FROM films WHERE date_prod > $1;
CREATE TEMP TABLE films_recent WITH (OIDS) ON COMMIT DROP AS
  EXECUTE recentfilms('2002-01-01');
```

兼容性

CREATE TABLE AS符合SQL标准。以下是非标准扩展：

- 标准要求在子查询子句外围加括号；在 PostgreSQL中，这些括号是可选的。
- 在标准中，WITH [NO] DATA子句是必需的；在 PostgreSQL 中它是可选的。
- PostgreSQL处理临时表的方式与标准有相当大的不同；详见CREATE TABLE。
- WITH子句是PostgreSQL扩展；标准中既没有存储参数，也没有 OID。
- PostgreSQL的表空间概念不是标准的一部分。因此，TABLESPACE子句是一种扩展。

另见

CREATE TABLE, EXECUTE, SELECT, SELECT INTO, VALUES

CREATE TABLESPACE

CREATE TABLESPACE — define a new tablespace

大纲

```
CREATE TABLESPACE tablespace_name [ OWNER user_name ] LOCATION
 'directory'
```

描述

`CREATE TABLESPACE`注册一个新的集簇级别的表空间。该表空间的名称必须不同于数据库集簇中任何现有表空间的名称。

表空间允许超级用户在文件系统中定义一个替代位置，使数据库对象（如表和索引）的数据文件可以存放在那里。

具有适当权限的用户可以将 *tablespace_name*传递给 `CREATE DATABASE`、`CREATE TABLE`、`CREATE INDEX`或`ADD CONSTRAINT`，从而将这些对象的数据文件存储在指定的表空间中。

参数

tablespace_name

要创建的表空间名称。名称不能以`pg_`开头，因为这类名称保留给系统表空间使用。

user_name

将成为该表空间所有者的用户名。若省略，则默认为执行该命令的用户。
只有超级用户可以创建表空间，但他们可以将表空间的所有权赋予非超级用户。

directory

将用于表空间的目录。该目录应为空，并且必须归PostgreSQL系统用户所有。
必须使用绝对路径名指定该目录。

注解

表空间仅在支持符号链接的系统上受支持。

`CREATE TABLESPACE`不能在事务块内执行。

示例

创建一个表空间`dbspace`，位置为`/data/dbs`：

```
CREATE TABLESPACE dbspace LOCATION '/data/dbs';
```

创建一个表空间`indexspace`，位置为`/data/indexes`，所有者为用户`genevieve`：

```
CREATE TABLESPACE indexspace OWNER genevieve LOCATION '/data/indexes';
```

兼容性

`CREATE TABLESPACE`是一种PostgreSQL扩展。

另见

CREATE DATABASE, CREATE TABLE, CREATE INDEX, DROP TABLESPACE, ALTER TABLESPACE

CREATE TEXT SEARCH CONFIGURATION

CREATE TEXT SEARCH CONFIGURATION — define a new text search configuration

大纲

```
CREATE TEXT SEARCH CONFIGURATION name (
    PARSER = parser_name |
    COPY = source_config
)
```

描述

`CREATE TEXT SEARCH CONFIGURATION` 创建一个新的文本搜索配置。文本搜索配置指定一个能把字符串拆分为记号的文本搜索解析器，以及一些可用于判断哪些记号对搜索有意义的词典。

如果只指定了解析器，那么新文本搜索配置最初没有从记号类型到词典的映射，因而会忽略所有词。要让该配置真正可用，必须使用后续的`ALTER TEXT SEARCH CONFIGURATION`命令创建映射。另一种方式是复制一个现有的文本搜索配置。

如果给出了一个模式名称，则文本搜索配置会被创建在指定的模式中。否则它将会被创建在当前模式中。

定义该文本搜索配置的用户会成为其拥有者。

更多信息请参见第 12 章。

参数

name

要创建的文本搜索配置的名称。该名称可以是模式限定的。

parser_name

此配置要使用的文本搜索解析器的名称。

source_config

要复制的现有文本搜索配置的名称。

注解

`PARSER`和`COPY`选项是互斥的，因为当一个已有的配置被复制时，它的解析器选择也会一并被复制。

兼容性

在 SQL 标准中没有 `CREATE TEXT SEARCH CONFIGURATION`语句。

另见

ALTER TEXT SEARCH CONFIGURATION, DROP TEXT SEARCH CONFIGURATION

CREATE TEXT SEARCH DICTIONARY

CREATE TEXT SEARCH DICTIONARY — define a new text search dictionary

大纲

```
CREATE TEXT SEARCH DICTIONARY name (
    TEMPLATE = template
    [, option = value [, ... ]]
)
```

描述

`CREATE TEXT SEARCH DICTIONARY`创建一个新的文本搜索字典。文本搜索字典指定一种在搜索时识别哪些词值得关注、哪些词不值得关注的方式。字典依赖于文本搜索模板，后者规定了实际执行这项工作的函数。通常，字典会提供一些选项，用来控制模板函数的具体行为。

如果给出了一个模式名称，那么该文本搜索字典会被创建在指定的模式中。否则它会被创建在当前模式中。

定义文本搜索字典的用户将成为其拥有者。

更多信息请参见第 12 章。

参数

name

要创建的文本搜索字典的名称。该名称可以被模式限定。

template

用于定义该字典基本行为的文本搜索模板名称。

option

要为此字典设置的模板相关选项的名称。

value

模板相关选项要使用的值。如果该值不是简单的标识符或数字，则必须用引号括起（当然，如果你愿意，也始终可以加引号）。

这些选项可以按任意顺序出现。

示例

下面的示例命令创建了一个基于 Snowball 且使用非标准停用词列表的字典。

```
CREATE TEXT SEARCH DICTIONARY my_russian (
    template = snowball,
    language = russian,
```

```
    stopwords = myrussian
);
```

兼容性

在 SQL 标准中没有 `CREATE TEXT SEARCH DICTIONARY`语句。

另见

ALTER TEXT SEARCH DICTIONARY, DROP TEXT SEARCH DICTIONARY

CREATE TEXT SEARCH PARSER

CREATE TEXT SEARCH PARSER — define a new text search parser

大纲

```
CREATE TEXT SEARCH PARSER name (
    START = start_function ,
    GETTOKEN = gettoken_function ,
    END = end_function ,
    LEXTYPES = lextypes_function
    [, HEADLINE = headline_function ]
)
```

描述

`CREATE TEXT SEARCH PARSER`创建一个新的全文检索解析器。全文检索解析器定义了一种方法，用于将文本字符串拆分成记号并为这些记号指定类型（类别）。
解析器本身并没有太大用处，必须与一些全文检索字典一起绑定到一个全文检索配置中，
才能用于搜索。

如果给出了一个模式名称，那么全文检索解析器将被创建在指定的模式中。
否则它会被创建在当前模式中。

要使用`CREATE TEXT SEARCH PARSER`，你必须是超级用户。之所以有此限制，是因为错误的全文检索解析器定义可能会让服务器陷入混乱，甚至崩溃。

更多信息请参见第 12 章。

参数

name

要创建的全文检索解析器的名称。该名称可以是模式限定的。

start_function

该解析器的启动函数的名称。

gettoken_function

该解析器的获取下一个记号的函数名称。

end_function

该解析器的结束函数的名称。

lextypes_function

该解析器的 lextypes 函数的名称（该函数返回它所产生的记号类型集合的信息）。

headline_function

该解析器的 headline 函数的名称（该函数对一组记号生成摘要）。

如有必要，函数名称可以是模式限定的。这里未给出参数类型，
因为每类函数的参数列表都是预先确定的。除 headline 函数外，其余函数都是必需的。

这些参数可以按任意顺序出现，不必与上面展示的顺序一致。

兼容性

在 SQL 标准中没有 `CREATE TEXT SEARCH PARSER`语句。

另见

ALTER TEXT SEARCH PARSER, DROP TEXT SEARCH PARSER

CREATE TEXT SEARCH TEMPLATE

CREATE TEXT SEARCH TEMPLATE — define a new text search template

大纲

```
CREATE TEXT SEARCH TEMPLATE name (
    [ INIT = init_function , ]
    LEXIZE = lexize_function
)
```

描述

`CREATE TEXT SEARCH TEMPLATE`创建一个新的全文检索模板。全文检索模板定义实现全文检索字典的函数。模板本身并无直接用途，必须先实例化为字典后才能使用。
字典通常会指定要传递给模板函数的参数。

如果给出了模式名称，则全文检索模板会被创建在指定模式中。否则它会被创建在当前模式中。

要使用`CREATE TEXT SEARCH TEMPLATE`，你必须是超级用户。之所以有此限制，是因为错误的全文检索模板定义可能使服务器陷入混乱，甚至崩溃。将模板与字典分离的原因在于，
模板封装了定义字典时那些“不安全”的方面。而在定义字典时可设置的参数，
对非特权用户来说是安全的，因此创建字典不必是特权操作。

更多信息请参见第 12 章。

参数

name

要创建的全文检索模板名称。该名称可以是模式限定的。

init_function

该模板的 init 函数的名称。

lexize_function

该模板的 lexize 函数名称。

如有必要，函数名称可以是模式限定的。这里没有给出参数类型，
因为每一类函数的参数列表都是预先确定的。lexize 函数是必需的，但 init 函数是可选的。

这些参数可以按任意顺序出现，不必局限于上面显示的顺序。

兼容性

在 SQL 标准中没有 `CREATE TEXT SEARCH TEMPLATE`语句。

另见

ALTER TEXT SEARCH TEMPLATE, DROP TEXT SEARCH TEMPLATE

CREATE TRIGGER

CREATE TRIGGER — define a new trigger

大纲

```
CREATE [ CONSTRAINT ] TRIGGER name { BEFORE | AFTER | INSTEAD OF } {
 event [ OR ... ] }
    ON table
    [ FROM referenced_table_name ]
    [ NOT DEFERRABLE | [ DEFERRABLE ] { INITIALLY IMMEDIATE |
 INITIALLY DEFERRED } ]
    [ FOR [ EACH ] { ROW | STATEMENT } ]
    [ WHEN ( condition ) ]
    EXECUTE PROCEDURE function_name ( arguments )

where event can be one of:

    INSERT
    UPDATE [ OF column_name [, ... ] ]
    DELETE
    TRUNCATE
```

描述

`CREATE TRIGGER`创建一个新触发器。触发器会关联到指定的表、视图或外部表，并在发生某些事件时执行指定的函数*function_name*。

可以指定触发器在尝试对某一行执行该操作之前引发（即在检查约束以及尝试执行`INSERT`、`UPDATE`或`DELETE`之前）；也可以在该操作完成之后引发（即在检查约束以及完成 `INSERT`、`UPDATE`或 `DELETE`之后）；或者取代该操作执行（用于视图上的插入、更新或删除）。如果触发器在事件之前引发，或者改为取代该事件执行，则它可以跳过对当前行的操作，
或者修改待插入的行（仅适用于`INSERT`和`UPDATE`操作）。如果触发器在事件之后引发，
则所有更改（包括其他触发器的影响）对该触发器都是“可见”的。

被标记为`FOR EACH ROW`的触发器会针对该操作修改的每一行调用一次。例如，一个影响 10 行的`DELETE`会导致目标关系上的 `ON DELETE`触发器分别被调用 10 次，即每个被删除的行调用一次。相比之下，被标记为`FOR EACH STATEMENT`的触发器对于任意给定操作只执行一次，不管该操作修改了多少行（特别是，修改零行的操作仍会导致所有适用的 `FOR EACH STATEMENT`触发器执行）。

被指定为取代该触发事件执行的 `INSTEAD OF` 触发器必须标记为 `FOR EACH ROW`，并且只能定义在视图上。视图上的 `BEFORE` 和 `AFTER` 触发器必须标记为 `FOR EACH STATEMENT`。

此外，也可以把触发器定义为在 `TRUNCATE` 时引发，但只能是 `FOR EACH STATEMENT`。

下表汇总了哪些类型的触发器可用于表、视图和外部表：

When	Event	Row-level	Statement-level
`BEFORE`	`INSERT/UPDATE/DELETE`	Tables	Tables and views

When	Event	Row-level	Statement-level
	TRUNCATE	–	Tables
AFTER	INSERT/UPDATE/DELETE	Tables	Tables and views
	TRUNCATE	–	Tables
INSTEAD OF	INSERT/UPDATE/DELETE	Views	–
	TRUNCATE	–	–

此外，触发器定义还可以指定一个布尔型 WHEN 条件，用于测试是否应当引发该触发器。在行级触发器中，WHEN 条件可以检查该行各列的新值和/或旧值。语句级触发器也可以带有 WHEN 条件，不过这一特性对它们的用处不大，因为该条件无法引用表中的任何值。

如果针对同一事件定义了多个同类触发器，它们将按名称的字母顺序引发。

指定CONSTRAINT选项时，此命令会创建约束触发器。这与普通触发器相同，只是可以使用SET CONSTRAINTS调整触发器触发的时间。约束触发器必须是表上的AFTER ROW触发器。它们可以在导致触发事件的语句结束时触发，也可以在包含该语句的事务结束时触发；后一种情况下称为延迟。也可以使用SET CONSTRAINTS强制尚待执行的延迟触发器立即触发。约束触发器应在其实现的约束被违反时引发异常。

SELECT 不会修改任何行，因此无法创建 SELECT 触发器。对于这类情形，规则和视图更为合适。

参见第 36 章以了解有关触发器的更多信息。

参数

name

新触发器的名称。它必须不同于同一表上的任何其他触发器名称。该名称不能是模式限定名 – 触发器会继承其所属表的模式。对于约束触发器，这也是使用SET CONSTRAINTS 修改触发器行为时所使用的名称。

BEFORE
AFTER
INSTEAD OF

决定该函数是在事件之前、之后调用，还是取代该事件执行。约束触发器只能指定为 AFTER。

event

INSERT、UPDATE、DELETE或 TRUNCATE 之一；它指定了哪个事件将引发该触发器。可以使用OR指定多个事件。

对于 UPDATE 事件，可以使用以下语法指定列列表：

UPDATE OF *column_name1* [, *column_name2* ...]

只有当列出的列中至少有一列被列为 UPDATE 命令的目标时，触发器才会触发。

INSTEAD OF UPDATE 事件不支持列列表。

table

该触发器所作用的表、视图或外部表的名称（可以是模式限定的）。

referenced_table_name

约束所引用的另一张表的名称（可以是模式限定的）。此选项用于外键约束，不建议用于一般用途。它只能为约束触发器指定。

DEFERRABLE
NOT DEFERRABLE
INITIALLY IMMEDIATE
INITIALLY DEFERRED

触发器的默认时机。有关这些约束选项的细节，请参见CREATE TABLE。它们只能为约束触发器指定。

FOR EACH ROW
FOR EACH STATEMENT

这指定触发器过程应针对触发事件影响的每一行触发一次，还是仅针对每条 SQL 语句触发一次。如果两者均未指定，则默认值为FOR EACH STATEMENT。约束触发器只能指定FOR EACH ROW。

condition

一个布尔表达式，用于决定触发器函数是否会实际执行。如果指定了WHEN，则仅当 *condition* 返回 true 时才会调用该函数。在 FOR EACH ROW 触发器中，WHEN 条件可以分别写成 OLD.*column_name* 或 NEW.*column_name* 来引用旧行值和/或新行值中的列。当然，INSERT 触发器不能引用 OLD，而 DELETE 触发器不能引用 NEW。

INSTEAD OF触发器不支持WHEN条件。

当前，WHEN 表达式不能包含子查询。

请注意，对于约束触发器，WHEN 条件的求值不会被延迟，而是在行操作执行后立即发生。如果该条件求值结果不为真，则该触发器不会被加入延迟执行队列。

function_name

一个由用户提供的函数，声明为不接受参数并返回 trigger 类型；当触发器被触发时，就会执行该函数。

arguments

一个可选的、以逗号分隔的参数列表，在执行触发器时会传给该函数。这些参数是字符串字面常量。这里也可以写简单名称和数字常量，但它们都会被转换成字符串。请查阅该触发器函数所用实现语言的说明，以了解如何在函数内部访问这些参数；这可能与普通函数参数不同。

注解

要在表上创建触发器，用户必须拥有该表的TRIGGER权限。
用户还必须拥有触发器函数的EXECUTE权限。

使用DROP TRIGGER删除触发器。

列限定触发器（即使用 UPDATE OF *column_name* 语法定义的触发器）会在其任一列被列为 UPDATE 命令的 SET 列表目标时引发。即便触发器没有引发，列值仍有可能发生变化，因为 BEFORE UPDATE 触发器对行内容所作的更改不会被考虑在内。反过来，诸如 UPDATE ... SET x = x ... 这样的命令会引发列 x 上的触发器，即使该列的值实际上并未变化。

在BEFORE触发器中，WHEN条件会在函数执行前（或者本应执行前）立即求值，因此使用WHEN与在触发器函数开头测试相同条件并无实质区别。特别要注意的是，该条件看到的NEW行是当前值，它可能已经被更早触发的触发器修改过。此外，BEFORE触发器的WHEN条件不允许检查NEW行的系统列（例如oid），因为这些列此时尚未被设置。

在 AFTER 触发器中，WHEN 条件会在行变更发生后立即求值，并决定是否将一个事件排入队列，以便在语句结束时引发该触发器。因此，当 AFTER 触发器的 WHEN 条件不返回真时，就没有必要将事件排队，也不需要在语句结束时重新获取该行。如果触发器只需针对少数几行引发，这能显著提升那些会修改很多行的语句的速度。

在PostgreSQL 7.3 之前的版本中，必须将触发器函数声明为返回占位类型opaque，而不是trigger。为了支持加载旧转储文件，CREATE TRIGGER将接受声明为返回opaque的函数，但会发出通知，并将该函数声明的返回类型更改为trigger。

示例

执行该函数check_account_update，每当表accounts中的一行即将被更新时调用它：

```
CREATE TRIGGER check_update
    BEFORE UPDATE ON accounts
    FOR EACH ROW
    EXECUTE PROCEDURE check_account_update();
```

相同，但仅当列balance被指定为UPDATE命令的目标时才执行该函数：

```
CREATE TRIGGER check_update
    BEFORE UPDATE OF balance ON accounts
    FOR EACH ROW
    EXECUTE PROCEDURE check_account_update();
```

这种形式仅在列balance的值确实发生变化时才执行该函数：

```
CREATE TRIGGER check_update
    BEFORE UPDATE ON accounts
    FOR EACH ROW
    WHEN (OLD.balance IS DISTINCT FROM NEW.balance)
    EXECUTE PROCEDURE check_account_update();
```

调用一个函数记录 accounts 的更新，但仅在确有内容发生变化时才调用：

```
CREATE TRIGGER log_update
    AFTER UPDATE ON accounts
    FOR EACH ROW
    WHEN (OLD.* IS DISTINCT FROM NEW.*)
    EXECUTE PROCEDURE log_account_update();
```

执行该函数view_insert_row，针对每一行将行插入视图底层的表中：

```
CREATE TRIGGER view_insert
    INSTEAD OF INSERT ON my_view
```

```
    FOR EACH ROW
    EXECUTE PROCEDURE view_insert_row();
```

第 36.4 节包含一个用 C 编写的触发器函数的完整示例。

兼容性

PostgreSQL 中的 `CREATE TRIGGER` 语句实现了 SQL 标准的一个子集。当前仍缺少以下功能：

- SQL 允许为在触发动作定义中使用的“旧”行或“新”行定义别名（例如 `CREATE TRIGGER ... ON tablename REFERENCING OLD ROW AS somename NEW ROW AS othername ...`）。由于 PostgreSQL 允许用任意多种用户定义的语言编写触发器过程，对数据的访问是以语言相关的方式处理的。
- PostgreSQL 只允许通过执行用户定义函数来完成触发动作。标准则允许把许多其他 SQL 命令（例如 `CREATE TABLE`）作为触发动作执行。通过创建一个执行所需命令的用户定义函数，这一限制并不难绕过。

SQL 规定多个触发器应按创建时间顺序引发。PostgreSQL 则按名称顺序引发，这被认为更方便。

SQL 规定，级联删除上的 `BEFORE DELETE` 触发器应在级联 `DELETE` 完成之后引发。PostgreSQL 的行为则是 `BEFORE DELETE` 总会在删除动作之前引发，即使是级联删除也一样。这被认为更一致。如果 `BEFORE` 触发器在由引用动作引起的更新过程中修改行或阻止更新，也会出现非标准行为。这可能导致约束违反，或者使存储的数据不符合引用约束。

使用 `OR` 为单个触发器指定多个动作的能力，是 PostgreSQL 对 SQL 标准的扩展。

能够在 `TRUNCATE` 上引发触发器，是 PostgreSQL 对 SQL 标准的扩展；能够在视图上定义语句级触发器也是如此。

`CREATE CONSTRAINT TRIGGER`是PostgreSQL对SQL标准的扩展。

参见

CREATE FUNCTION, ALTER TRIGGER, DROP TRIGGER, SET CONSTRAINTS

CREATE TYPE

CREATE TYPE — define a new data type

大纲

```
CREATE TYPE name AS
    ( [ attribute_name data_type [ COLLATE collation ] [, ... ] ] )

CREATE TYPE name AS ENUM
    ( [ 'label' [, ... ] ] )

CREATE TYPE name (
    INPUT = input_function,
    OUTPUT = output_function
    [ , RECEIVE = receive_function ]
    [ , SEND = send_function ]
    [ , TYPMOD_IN = type_modifier_input_function ]
    [ , TYPMOD_OUT = type_modifier_output_function ]
    [ , ANALYZE = analyze_function ]
    [ , INTERNALLENGTH = { internallength | VARIABLE } ]
    [ , PASSEDBYVALUE ]
    [ , ALIGNMENT = alignment ]
    [ , STORAGE = storage ]
    [ , LIKE = like_type ]
    [ , CATEGORY = category ]
    [ , PREFERRED = preferred ]
    [ , DEFAULT = default ]
    [ , ELEMENT = element ]
    [ , DELIMITER = delimiter ]
    [ , COLLATABLE = collatable ]
)

CREATE TYPE name
```

描述

`CREATE TYPE`在当前数据库中注册一种新的数据类型。定义该类型的用户将成为其拥有者。

如果给出了模式名，则该类型会在指定模式中创建；否则会在当前模式中创建。
类型名必须不同于同一模式中任何现有类型或域的名称。（由于表也具有关联的数据类型，类型名还必须不同于同一模式中任何现有表的名称。）

复合类型

第一种形式的`CREATE TYPE`创建复合类型。复合类型由属性名和数据类型列表指定。如果某个属性的数据类型支持排序规则，还可以指定该属性的排序规则。
复合类型本质上与表的行类型相同，但如果目的只是定义一种类型，使用`CREATE TYPE`就不必实际创建表。例如，独立的复合类型可用作函数的参数类型或返回类型。

枚举类型

如第 8.7 节中所述，第二种形式的 `CREATE TYPE`创建枚举（enum）类型。枚举类型接受一个带引号标签的列表，其中每个标签的长度都必须小于 `NAMEDATALEN`字节（在标准 PostgreSQL构建中为 64 字节）。

基础类型

第四种形式的`CREATE TYPE`创建一种新的基础类型（标量类型）。要创建新的基础类型，你必须是超级用户。（这样限制是因为错误的类型定义可能使服务器陷入混乱，甚至导致其崩溃。）

这些参数可以按任意顺序出现，不必局限于上面展示的顺序，而且大多数都是可选的。
在定义该类型之前，必须先注册两个或更多函数（使用 `CREATE FUNCTION`）。支持函数 *`input_function`*和 *`output_function`*是必需的；*`receive_function`*、*`send_function`*、*`type_modifier_input_function`*、*`type_modifier_output_function`*、*`analyze_function`*则是可选的。通常这些函数必须用 C 或其他低级语言编写。

*`input_function`*将类型的外部文本表示转换为该类型的操作符和函数所使用的内部表示。*`output_function`* 执行相反的转换。输入函数可以声明为接受一个`cstring` 参数，或者接受三个参数，类型分别为`cstring`、`oid`、`integer`。第一个参数是以 C 字符串表示的输入文本，第二个参数是该类型自身的 OID（数组类型例外，此时传入的是其元素类型的 OID），第三个参数是在已知情况下目标列的 `typmod`（未知则传入 -1）。输入函数必须返回该数据类型本身的值。通常输入函数应声明为 STRICT；否则，在读取 NULL 输入值时会以 NULL 作为第一个参数调用它。除非函数抛出错误，否则在这种情况下仍必须返回 NULL。（这种情况主要是为了支持可能需要拒绝 NULL 输入的域输入函数。）输出函数必须声明为接受一个新数据类型参数，并且必须返回`cstring`类型。对于 NULL 值不会调用输出函数。

可选的*`receive_function`* 把类型的外部二进制表示转换为内部表示。如果未提供此函数，该类型就不能参与二进制输入。外部二进制表示应选择为既能低成本转换为内部形式，
又具有合理可移植性。（例如，标准整数数据类型把网络字节序用作外部二进制表示，
而内部表示则使用机器的本地字节序。）接收函数应进行充分检查以确保值有效。
它可以声明为接受一个 `internal`参数，或者接受三个参数，类型分别为 `internal`、`oid`、`integer`。第一个参数是指向保存已接收字节串的`StringInfo`缓冲区的指针；
其余可选参数与文本输入函数相同。接收函数必须返回该数据类型本身的值。通常，
接收函数应声明为 STRICT；否则，在读取 NULL 输入值时会以 NULL 作为第一个参数调用它。除非函数抛出错误，否则在这种情况下仍必须返回 NULL。（这种情况主要是为了支持可能需要拒绝 NULL 输入的域接收函数。）类似地，可选的 *`send_function`*把内部表示转换为外部二进制表示。如果未提供此函数，该类型就不能参与二进制输出。
发送函数必须声明为接受一个新数据类型参数，并且必须返回 `bytea`类型。对于 NULL 值不会调用发送函数。

读到这里，你可能会问：既然新类型本身还没创建，
输入和输出函数怎么能声明为返回或接受这个新类型呢？答案是，应先把该类型定义为一种 shell 类型，它除了名称和拥有者外没有任何属性，只是一个占位符类型。这可通过执行不带任何附加参数的命令 `CREATE TYPE` *`name`* 完成。然后就可以定义引用该 shell 类型的 I/O 函数。最后，再用带完整定义的`CREATE TYPE`替换该 shell 条目，生成一个完整且有效的类型定义，此后新类型就能正常使用。

如果该类型支持修饰符，也就是附加在类型声明上的可选约束，例如 `char(5)`或`numeric(30,2)`，就需要可选的*`type_modifier_input_function`*和 *`type_modifier_output_function`*。PostgreSQL允许用户定义类型接受一个或多个简单常量或标识符作为修饰符。
不过，这些信息必须能够打包成单个非负整数值，以便存储在系统目录中。声明的修饰符会以`cstring`数组的形式传递给*`type_modifier_input_function`*。

它必须检查这些值是否有效（若无效则抛出错误），若有效则返回一个非负integer值，该值将作为列的“typmod”存储。如果该类型没有*type_modifier_input_function*，就会拒绝类型修饰符。*type_modifier_output_function* 则把内部整数 typmod 值转换回适合用户显示的正确形式。它必须返回一个cstring值，即精确追加到类型名后的字符串；例如，numeric的该函数可能返回 (30,2)。允许省略*type_modifier_output_function*；在这种情况下，默认显示格式只是把存储的 typmod 整数值放在圆括号中。

可选的*analyze_function* 为该数据类型的列执行类型专用的统计信息收集。默认情况下，如果该类型有默认的 B-树操作符类，ANALYZE将尝试使用该类型的“equals”和“less-than”操作符收集统计信息。对于非标量类型，这种行为很可能不合适，因此可以通过指定自定义分析函数来覆盖。分析函数必须声明为接受一个 internal参数并返回boolean结果。分析函数的详细 API 见src/include/commands/vacuum.h。

虽然新类型内部表示的细节只有 I/O 函数以及你为该类型编写的其他函数才知道，但仍有若干内部表示属性必须向 PostgreSQL声明。其中最重要的是 *internallength*。基础类型可以是定长的，此时*internallength*为正整数；也可以是变长的，此时将*internallength*设为VARIABLE。（在内部，这通过把 typlen设为 -1 表示。）所有变长类型的内部表示都必须以一个 4 字节整数开头，用来给出该类型该值的总长度。

可选标志PASSEDBYVALUE表示该数据类型的值按值传递，而不是按引用传递。内部表示大于Datum类型大小（多数机器上为 4 字节，少数为 8 字节）的类型不能按值传递。

*alignment*参数指定该数据类型所需的存储对齐方式。允许的值分别对应按 1、2、4 或 8 字节边界对齐。注意，变长类型的对齐至少必须为 4，因为它们的第一个组成部分必然是一个int4。

*storage*参数允许为变长数据类型选择存储策略。（定长类型只允许 plain。）plain表示该类型数据始终内联存储且不压缩。extended表示系统会先尝试压缩较长的数据值，如果仍然过长，就把该值移出主表行。external允许把值移出主表，但系统不会尝试压缩它。main允许压缩，但不鼓励把值移出主表。（采用这种存储策略的数据项在没有其他办法让一行适配时仍可能被移出主表，但与extended和external 数据项相比，它们会被优先保留在主表中。）

like_type
参数提供了指定数据类型基本表示属性的另一种方法：从某个现有类型复制这些属性。*internallength*、*passedbyvalue*、*alignment*和 *storage*的值都从指定类型复制而来。（虽然可以通过同时给出LIKE子句和这些选项来覆盖其中某些值，但通常不建议这样做。）当新类型的底层实现以某种方式“借用”现有类型时，以这种方式指定表示属性尤其有用。

*category*和 *preferred*参数可用于在存在歧义时帮助控制应用哪一种隐式类型转换。每种数据类型都属于一个由单个 ASCII 字符命名的类别，并且在其类别内要么是“首选”的，要么不是。当这一规则有助于解析重载函数或操作符时，解析器会优先转换为首选类型（但只会从同一类别中的其他类型转换）。更多细节见第 10 章。对于与任何其他类型之间都没有隐式类型转换的类型，保持这些设置的默认值就足够了。不过，对于一组彼此存在隐式类型转换的相关类型，把它们都标记为属于同一类别，并选择一两个“最通用”的类型作为该类别的首选类型，通常会有帮助。*category*参数在将用户定义类型加入现有内置类别（例如数值类型或字符串类型）时尤其有用。不过，也可以创建全新的纯用户定义类型类别。为这种类别命名时，可选择任一非大写字母的 ASCII 字符。

如果用户希望该数据类型列的默认值不是空值，可以指定默认值。用 DEFAULT关键字指定默认值。（该默认值可以被附加到具体列上的显式DEFAULT子句覆盖。）

若要表明一种类型是数组类型，可用ELEMENT 关键字指定数组元素的类型。例如，要定义由 4 字节整数（int4）构成的数组，可指定 ELEMENT = int4。更多关于数组类型的细节见下文。

若要指定该类型数组在外部表示中用于分隔各值的分隔符，可把 *delimiter*设为特定字符。默认分隔符是逗号（,）。注意，这个分隔符关联的是数组元素类型，而不是数组类型本身。

如果可选的布尔参数 *collatable*为真，则该类型的列定义和表达式可以通过COLLATE子句携带排序规则信息。是否实际使用这些排序规则信息取决于操作该类型的函数实现；
仅仅把类型标记为支持排序规则并不会自动实现这一点。

数组类型

每当创建用户定义类型时，PostgreSQL 都会自动创建一个关联的数组类型，其名称由元素类型名前加一个下划线组成；必要时还会截断，以保持其长度小于 NAMEDATALEN字节。（如果这样生成的名称与现有类型名冲突，就会重复这一过程，直到找到不冲突的名称。）
这种隐式创建的数组类型是变长的，并使用内置输入/输出函数 array_in和array_out。该数组类型会跟踪其元素类型的拥有者或模式的任何变化，并在元素类型被删除时一并删除。

如果系统会自动创建正确的数组类型，你可能会合理地问，为什么还需要ELEMENT选项。
ELEMENT唯一有用的情况是：你正在创建一种定长类型，
而它在内部恰好是若干相同元素组成的数组，并且除了为整个类型提供的操作之外，
你还希望允许通过下标直接访问这些元素。例如，类型point在内部就表示为两个浮点数，
可以用point[0]和 point[1]访问。注意，这种机制只适用于内部形式恰好是一串相同定长字段的定长类型。可用下标访问的变长类型必须采用array_in和array_out
使用的通用内部表示。由于历史原因（也就是说这显然不对，但现在改已经太晚了），
定长数组类型的下标从零开始，而变长数组则从一开始。

参数

name

要创建的类型名称（可选地带有模式限定）。

attribute_name

复合类型的一个属性（列）的名称。

data_type

将成为复合类型一列的现有数据类型名称。

label

表示枚举类型某个值所关联文本标签的字符串字面量。

input_function

将数据从类型的外部文本形式转换为内部形式的函数名。

output_function

将数据从类型的内部形式转换为外部文本形式的函数名。

receive_function

将数据从类型的外部二进制形式转换成内部形式的函数名。

send_function

将数据从类型的内部形式转换为外部二进制形式的函数名。

type_modifier_input_function

将类型的修饰符数组转换为内部形式的函数名。

type_modifier_output_function

将类型修饰符的内部形式转换为外部文本形式的函数名。

analyze_function

为该数据类型执行统计分析的函数名。

internallength

一个数字常量，用于指定新类型内部表示的字节长度。默认假定它是变长的。

alignment

该数据类型的存储对齐需求。如果被指定，它必须是 `char`、`int2`、`int4`或者`double`。默认是 `int4`。

storage

该数据类型的存储策略。如果被指定，必须是 `plain`、`external`、`extended`或者`main`。默认是`plain`。

like_type

与新类型具有相同表示形式的现有数据类型名称。除非在本 `CREATE TYPE`命令的其他位置显式覆盖，否则 *internallength*、*passedbyvalue*、*alignment*和 *storage*的值都会从该类型复制。

category

该类型的类别码（单个 ASCII 字符）。默认值是表示“用户定义类型”的`'U'`。其他标准类别码见表 45.49。你也可以选择其他 ASCII 字符来创建自定义类别。

preferred

若该类型是其类型类别中的首选类型，则为真，否则为假。默认值为假。
在现有类型类别中创建新的首选类型时要格外小心，因为这可能导致出人意料的行为变化。

default

该数据类型的默认值。若省略，默认值为 null。

element

正在创建的类型是数组；该参数指定数组元素类型。

delimiter

在由该类型构成的数组中各值之间使用的分隔符字符。

collatable

如果该类型的操作可以使用排序规则信息，则为真。默认为假。

注解

由于数据类型一旦创建，其使用方式就不再受限制，因此创建基础类型或范围类型，
相当于对类型定义中提到的那些函数授予公共执行权限。
这对适合用于类型定义的那类函数来说通常不是问题。但如果设计一种类型时，
需要在把它转换为外部形式或从外部形式转换回来时使用“秘密”信息，就应当三思。

在PostgreSQL 8.3 之前，自动生成的数组类型名称总是恰好等于元素类型名称前加一个下划线字符（_）。（因此，类型名称的长度限制比其他名称少一个字符。）虽然现在通常仍是这样，
但在名称达到最大长度或与以下划线开头的用户类型名冲突时，数组类型名称可能与此不同。
因此，依赖这一约定编写代码的做法已经弃用。请改用 `pg_type.typarray` 来定位与给定类型关联的数组类型。

建议避免使用以下划线开头的类型名和表名。
虽然服务器会改变生成的数组类型名称以避免与用户给定的名称冲突，但仍然存在混淆风险，
特别是对旧客户端软件而言，它们可能会假定以下划线开头的类型名总是表示数组。

在PostgreSQL 8.2 之前，不存在 shell 类型创建语法`CREATE TYPE name`。创建新基础类型的做法是先创建它的输入函数。在这种做法下，
PostgreSQL会首先把新数据类型名视为输入函数的返回类型。此时 shell 类型会被隐式创建，然后就可以在其余 I/O 函数的定义中引用它。这种做法仍然有效，但已弃用，并且可能在未来某个版本中被禁止。另外，为避免由于函数定义中的简单拼写错误而意外使系统目录充满 shell 类型，只有在输入函数用 C 编写时，才会以这种方式创建 shell 类型。

在PostgreSQL 7.3 以前的版本中，通常完全不创建 shell 类型，而是将函数中对类型名的前向引用替换为占位伪类型 `opaque`。在 7.3 以前，`cstring`参数和结果也必须声明为`opaque`。为了支持载入旧转储文件，`CREATE TYPE`会接受使用`opaque`声明的 I/O 函数，但会发出通知，并修改函数声明以使用正确的类型。

示例

这个示例创建一种复合类型，并在函数定义中使用它：

```
CREATE TYPE compfoo AS (f1 int, f2 text);

CREATE FUNCTION getfoo() RETURNS SETOF compfoo AS $$
    SELECT fooid, fooname FROM foo
$$ LANGUAGE SQL;
```

这个示例创建一种枚举类型，并在表定义中使用它：

```
CREATE TYPE bug_status AS ENUM ('new', 'open', 'closed');

CREATE TABLE bug (
    id serial,
    description text,
    status bug_status
);
```

这个示例创建基础类型`box`，然后在表定义中使用它：

```
CREATE TYPE box;

CREATE FUNCTION my_box_in_function(cstring) RETURNS box AS ... ;
```

```
CREATE FUNCTION my_box_out_function(box) RETURNS cstring AS ... ;

CREATE TYPE box (
    INTERNALLENGTH = 16,
    INPUT = my_box_in_function,
    OUTPUT = my_box_out_function
);

CREATE TABLE myboxes (
    id integer,
    description box
);
```

如果box的内部结构是由四个 float4元素构成的数组，则也可以改为这样写：

```
CREATE TYPE box (
    INTERNALLENGTH = 16,
    INPUT = my_box_in_function,
    OUTPUT = my_box_out_function,
    ELEMENT = float4
);
```

这样就能通过下标访问 box 值的各个分量。除此之外，该类型的行为与前例相同。

这个示例创建一种大对象类型，并在表定义中使用它：

```
CREATE TYPE bigobj (
    INPUT = lo_filein, OUTPUT = lo_fileout,
    INTERNALLENGTH = VARIABLE
);
CREATE TABLE big_objs (
    id integer,
    obj bigobj
);
```

更多示例（包括配套的输入和输出函数）请见第 35.11 节。

兼容性

第一种CREATE TYPE形式，即创建复合类型的形式，符合SQL标准。其他形式都是 PostgreSQL 扩展。SQL标准中的CREATE TYPE 语句还定义了PostgreSQL尚未实现的其他形式。

支持创建零属性的复合类型，是 PostgreSQL对标准的一种特有背离（类似于CREATE TABLE中的同类情况）。

另见

ALTER TYPE, CREATE DOMAIN, CREATE FUNCTION, DROP TYPE

CREATE USER

CREATE USER — define a new database role

大纲

```
CREATE USER name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | REPLICATION | NOREPLICATION
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE role_name [, ...]
    | IN GROUP role_name [, ...]
    | ROLE role_name [, ...]
    | ADMIN role_name [, ...]
    | USER role_name [, ...]
    | SYSID uid
```

描述

`CREATE USER`现为CREATE ROLE的别名。唯一的区别在于：当命令写为`CREATE USER`时，默认假定为`LOGIN`；而当命令写为 `CREATE ROLE`时，默认假定为`NOLOGIN`。

兼容性

`CREATE USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义留给具体实现决定。

另见

CREATE ROLE

CREATE USER MAPPING

CREATE USER MAPPING — define a new mapping of a user to a foreign server

大纲

```
CREATE USER MAPPING FOR { user_name | USER | CURRENT_USER | PUBLIC }
    SERVER server_name
    [ OPTIONS ( option 'value' [ , ... ] ) ]
```

描述

`CREATE USER MAPPING`定义用户到外部服务器的映射。用户映射通常封装连接信息，外部数据包装器会将其与外部服务器所封装的信息结合起来，以访问外部数据资源。

外部服务器的拥有者可以为该服务器上的任何用户创建用户映射。此外，
如果某个用户已被授予该服务器上的`USAGE`权限，
那么该用户也可以为自己的用户名创建用户映射。

参数

user_name

被映射到外部服务器的现有用户的名称。`CURRENT_USER` 和`USER`都匹配当前用户的名称。当指定 `PUBLIC`时，会创建一个所谓的公共映射；当没有适用的特定用户映射时，就会使用它。

server_name

将为其创建用户映射的现有服务器的名称。

`OPTIONS ( option 'value' [, ... ] )`

该子句指定用户映射的选项。这些选项通常定义该映射实际的用户名和密码。
选项名必须唯一。允许的选项名和值取决于该服务器的外部数据包装器。

示例

为用户`bob`、服务器`foo`创建一个用户映射：

```
CREATE USER MAPPING FOR bob SERVER foo OPTIONS (user 'bob', password
 'secret');
```

兼容性

`CREATE USER MAPPING`符合 ISO/IEC 9075-9（SQL/MED）。

另见

ALTER USER MAPPING, DROP USER MAPPING, CREATE FOREIGN DATA WRAPPER, CREATE SERVER

CREATE VIEW

CREATE VIEW — define a new view

大纲

```
CREATE [ OR REPLACE ] [ TEMP | TEMPORARY ] VIEW name [ ( column_name [
, ...] ) ]
    AS query
```

描述

`CREATE VIEW`定义一个视图，其内容来自一个查询。该视图不会被实际物化。相反，每次在查询中引用该视图时，都会执行定义它的查询。

`CREATE OR REPLACE VIEW`与之类似，但如果已经存在同名视图，则会用新定义替换它。新查询必须生成与现有视图查询相同的列（即列名、列顺序和数据类型都相同），
但可以在列表末尾附加额外的列。生成输出列的计算方式则可以完全不同。

如果给出了模式名称（例如，`CREATE VIEW myschema.myview ...`），则视图将在指定的模式中创建。否则，它将在当前模式中创建。临时视图存在于一个特殊模式中，
因此在创建临时视图时不能给出模式名称。视图的名称必须与同一模式中的任何其他关系（表、序列、索引、视图、物化视图或外部表）的名称不同。

参数

`TEMPORARY` or `TEMP`

如果指定该选项，视图将被创建为临时视图。临时视图会在当前会话结束时自动删除。
当临时视图存在时，除非使用模式限定名称引用，
否则具有相同名称的现有永久关系对当前会话不可见。

如果视图引用的任何表是临时表，则该视图也会被创建为临时视图（无论是否指定了`TEMPORARY`）。

name

要创建的视图名称（可以带模式限定）。

column_name

视图各列使用的名称列表，可选。若未给出，则从查询中推断列名。

query

一个将为该视图提供列和行的SELECT或VALUES命令。

注解

目前，视图是只读的：系统不允许在视图上执行插入、更新或删除。可以通过在视图上创建`INSTEAD`触发器来获得可更新视图的效果，这些触发器必须把对视图的插入等尝试转换为对其他表的适当动作。更多信息见CREATE TRIGGER。另一种可能的做法是创建规则（见CREATE RULE），但在实践中触发器更容易理解并且更容易正确使用。

使用DROP VIEW语句删除视图。

应留意视图列的名称和类型是否按预期确定。例如：

```
CREATE VIEW vista AS SELECT 'Hello World';
```

这种写法在两方面都不好：列名默认是`?column?`，而且列数据类型默认为`unknown`。
如果想要在视图结果中使用字符串字面量，可使用类似下面的写法：

```
CREATE VIEW vista AS SELECT text 'Hello World' AS hello;
```

对视图中引用的表的访问由视图所有者的权限决定。在某些情况下，
这可以用于为底层表提供安全但受限的访问。不过，并非所有视图都能防止篡改；
详情请参见第 37.4 节。视图调用的函数与直接从使用该视图的查询中调用时的处理方式相同。因此，视图的用户必须拥有调用视图所用全部函数的权限。

对现有视图使用`CREATE OR REPLACE VIEW`时，只会更改视图的定义性 SELECT 规则。其他视图属性（包括所有权、权限和非 SELECT 规则）保持不变。必须拥有该视图才能替换它（包括作为所有者角色的成员）。

示例

创建一个包含所有喜剧电影的视图：

```
CREATE VIEW comedies AS
    SELECT *
    FROM films
    WHERE kind = 'Comedy';
```

该视图将包含视图创建时位于`film`表中的列。尽管使用了`*`创建视图，
但以后添加到表中的列不会成为视图的一部分。

兼容性

SQL 标准为`CREATE VIEW`语句指定了一些额外的能力：

```
CREATE VIEW name [ ( column_name [, ...] ) ] AS query [ WITH [
 CASCADED | LOCAL ] CHECK OPTION ]
```

The optional clauses for the full SQL command are:

`CHECK OPTION`

该选项与可更新视图有关。对视图的所有 `INSERT` 和 `UPDATE` 命令都会被检查，以确保数据满足视图定义条件（也就是说，新数据可以通过视图看到）。如果不满足，更新将被拒绝。

`LOCAL`

在该视图上检查完整性。

`CASCADED`

在该视图以及任何依赖视图上检查完整性。如果既未指定`CASCADED`也未指定`LOCAL`，
则假定使用`CASCADED`。

`CREATE OR REPLACE VIEW` 是 PostgreSQL 的语言扩展。临时视图的概念也是如此。

另见

ALTER VIEW, DROP VIEW

DEALLOCATE

DEALLOCATE — deallocate a prepared statement

大纲

```
DEALLOCATE [ PREPARE ] { name | ALL }
```

描述

`DEALLOCATE`用于释放一个先前准备好的 SQL 语句。如果不显式释放预备语句，则它会在会话结束时被释放。

更多关于预备语句的信息，请参见PREPARE。

参数

`PREPARE`

此关键字会被忽略。

`name`

要释放的预备语句的名称。

`ALL`

释放所有预备语句。

兼容性

SQL 标准中包含一个`DEALLOCATE`语句，但它只用于嵌入式 SQL。

另见

EXECUTE, PREPARE

DECLARE

DECLARE — define a cursor

大纲

```
DECLARE name [ BINARY ] [ INSENSITIVE ] [ [ NO ] SCROLL ]
    CURSOR [ { WITH | WITHOUT } HOLD ] FOR query
```

描述

DECLARE允许用户创建游标，游标可用于从较大的查询中一次取出少量行。游标创建后，可使用FETCH从中取出行。

> **注意**
>
> 本页面描述的是 SQL 命令层面的游标用法。如果想要在 PL/pgSQL函数中使用游标，规则会有所不同 – 见第 39.7 节。

参数

name

要创建的游标名称。它必须与该会话中任何其他活动游标的名称不同。

BINARY

使游标以二进制格式而不是文本格式返回数据。

INSENSITIVE

表示从游标取出的数据不应受到游标创建后对其底层表所做更新的影响。在PostgreSQL中，这是默认行为；因此该关键字没有效果，仅为兼容 SQL 标准而接受它。

SCROLL
NO SCROLL

SCROLL指定游标可用于以非顺序方式（例如向后）取出行。根据查询执行计划的复杂程度，指定SCROLL可能会给查询执行带来性能开销。NO SCROLL指定游标不能以非顺序方式取出行。默认情况下只在某些情形下允许滚动，这与显式指定 SCROLL并不相同。详见下文注解。

WITH HOLD
WITHOUT HOLD

WITH HOLD指定在创建该游标的事务成功提交后，仍可继续使用该游标。WITHOUT HOLD指定该游标不能在创建它的事务之外使用。如果既未指定WITHOUT HOLD也未指定WITH HOLD，默认值是WITHOUT HOLD。

query

一个提供游标要返回的行的SELECT或者VALUES命令。

关键字BINARY、INSENSITIVE和SCROLL可以按任意顺序出现。

注解

普通游标以文本格式返回数据，就像SELECT产生的结果一样。BINARY
选项指定游标应以二进制格式返回数据。这减少了服务器和客户端两端的转换工作量，
但代价是程序员需要付出更多精力来处理与平台相关的二进制数据格式。举例来说，
如果某个查询从一个整型列返回值 1，那么默认游标会返回字符串 1，而二进制游标则会返回一个 4 字节字段，其中包含该值的内部表示（采用大端字节序）。

应谨慎使用二进制游标。许多应用程序（包括 psql）都没有准备好处理二进制游标，而是期望返回的数据为文本格式。

> **注意**
>
> 当客户端应用使用“扩展查询”协议发出FETCH命令时，Bind 协议消息会指定数据应以文本格式还是二进制格式提取。这一选择会覆盖定义游标时所指定的方式。因此，
> 在使用扩展查询协议时，二进制游标这一概念实际上已经过时了 – 任何游标都可以按文本或二进制方式处理。

除非指定了WITH HOLD，否则该命令创建的游标只能在当前事务中使用。因此，DECLARE若不带WITH HOLD，在事务块之外就毫无用处：游标只能存活到该语句执行完成。所以，如果在事务块之外使用这种命令，PostgreSQL会报错。可使用BEGIN和COMMIT（或ROLLBACK）
来定义事务块。

如果指定了WITH HOLD，且创建游标的事务成功提交，那么在同一会话中的后续事务里仍可继续访问该游标。（但如果创建事务被中止，游标会被移除。）使用WITH HOLD创建的游标会在对其发出显式CLOSE命令时关闭，或在会话结束时关闭。在当前实现中，
这种游标所表示的行会被复制到临时文件或内存区域中，以便它们在后续事务中仍然可用。

当查询包括FOR UPDATE或FOR SHARE时，不能指定WITH HOLD。

在定义将用于向后取出行的游标时，应指定SCROLL选项。这是 SQL 标准所要求的。不过，为了兼容早期版本，如果游标的查询计划足够简单，以致支持向后取出不需要额外开销，
PostgreSQL也允许在未指定SCROLL的情况下向后取出行。不过，
建议应用开发者不要依赖于从未用SCROLL创建的游标中向后取出行。如果指定了NO SCROLL，那么无论如何都不允许向后取出行。

当查询包含FOR UPDATE或FOR SHARE时，同样不允许向后取出行。因此在这种情况下不能指定SCROLL。

> **小心**
>
> 可滚动游标和WITH HOLD游标如果调用了任何易变函数（参见第 35.6 节），可能产生意外结果。重新取出先前已取出的行时，函数可能会再次执行，导致结果与第一次不同。
> 对于这种情况，一个变通方法是将游标声明为WITH HOLD，并在读取其中任何行之前提交事务。这样会强制将游标的整个输出物化到临时存储中，
> 使易变函数对每行恰好执行一次。

如果游标的查询包含FOR UPDATE或FOR SHARE，则返回的行会在首次取出时被锁定，与带有这些选项的普通SELECT命令相同。此外，返回的行将是最新版本；因此，这些选项提供了相当于

SQL 标准所称的“敏感游标”的行为。（将INSENSITIVE与FOR UPDATE或FOR SHARE一起指定会报错。）

> **小心**
>
> 通常都建议使用FOR UPDATE，如果游标打算与 UPDATE ... WHERE CURRENT OF或DELETE ... WHERE CURRENT OF一起使用。使用FOR UPDATE可以防止其他会话在这些行被取出之后、被更新之前更改它们。如果不使用FOR UPDATE，而某一行在游标创建后已经被更改，那么后续的WHERE CURRENT OF命令将不会有任何效果。
>
> 使用 FOR UPDATE 的另一个原因是：如果没有它，当游标查询不满足 SQL 标准中“简单可更新”的规则（特别是，游标必须只引用一个表，并且不使用分组或 ORDER BY）时，后续的 WHERE CURRENT OF 可能会失败。Cursors that are not simply updatable might work, or might not, depending on plan choice details; so in the worst case, an application might work in testing and then fail in production.
>
> 不将FOR UPDATE与WHERE CURRENT OF一起使用的主要原因，是你需要游标可滚动，或者需要它对后续更新不敏感（也就是说，继续显示旧数据）。如果这是需求，
> 请务必仔细注意上面的警告。

SQL 标准只为嵌入式SQL中的游标作出规定。PostgreSQL服务器没有为游标实现OPEN语句；游标在声明时即被视为打开。不过，ECPG是PostgreSQL的嵌入式 SQL 预处理器，它支持标准 SQL 的游标约定，包括涉及DECLARE和OPEN语句的那些约定。

你可以通过查询pg_cursors 系统视图查看所有可用游标。

示例

声明一个游标：

```
DECLARE liahona CURSOR FOR SELECT * FROM films;
```

更多游标用法示例见FETCH。

兼容性

SQL 标准规定，游标默认是否对底层数据的并发更新敏感，由实现决定。在PostgreSQL中，游标默认不敏感，可以通过指定FOR UPDATE使其敏感。其他产品的行为可能不同。

SQL 标准只允许在嵌入式SQL和模块中使用游标。PostgreSQL允许以交互方式使用游标。

二进制游标是PostgreSQL的一种扩展。

另见

CLOSE, FETCH, MOVE

DELETE

DELETE — delete rows of a table

大纲

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
DELETE FROM [ ONLY ] table [ * ] [ [ AS ] alias ]
    [ USING using_list ]
    [ WHERE condition | WHERE CURRENT OF cursor_name ]
    [ RETURNING * | output_expression [ [ AS ] output_name ] [, ...] ]
```

描述

DELETE从指定表中删除满足 WHERE子句的行。如果省略WHERE 子句，效果就是删除表中的所有行。结果是一个合法但为空的表。

> **提示**
>
> TRUNCATE是一种PostgreSQL扩展，它提供了一种更快的机制来移除表中的所有行。

有两种方法可以利用数据库中其他表所包含的信息来删除一个表中的行：使用子选择，或者在USING子句中指定附加表。哪种技术更合适取决于具体情况。

可选的RETURNING子句使DELETE 基于每个实际被删除的行计算并返回一个或多个值。可以计算使用该表列和/或USING中提到的其他表列的任意表达式。RETURNING列表的语法与SELECT 的输出列表相同。

要从一个表中删除行，必须具有该表上的DELETE权限，以及USING
子句中出现的任何表或者其值会在 *condition*中读取的任何表上的SELECT权限。

参数

with_query

WITH子句允许你指定一个或多个子查询，这些子查询可在DELETE查询中按名称引用。
详见第 7.8 节和SELECT。

table

要从中删除行的表名（可以是模式限定的）。如果在表名前指定了 ONLY，只会从所提及表中删除匹配行。如果未指定 ONLY，还会删除继承自该表的任何表中的匹配行。可选地，可以在表名后指定*，以显式指示包含后代表。

alias

目标表的替代名称。提供别名时，它会完全隐藏该表的实际名称。例如，给定DELETE FROM foo AS f，DELETE语句的其余部分必须将该表称为 f，而不是foo。

using_list

一个表表达式列表，允许其他表的列出现在WHERE条件中。这类似于SELECT语句的FROM Clause中可以指定的表列表；例如，可以为表名指定别名。不要把目标表重复写在*using_list*中，除非打算进行自连接。

condition

一个返回boolean类型值的表达式。只有使这个表达式返回true的行才会被删除。

cursor_name

在WHERE CURRENT OF条件中使用的游标名称。要删除的行是最近从该游标取出的那一行。该游标必须是针对DELETE目标表的非分组查询。注意，WHERE CURRENT OF不能与布尔条件一起指定。有关游标使用的更多信息，请参阅DECLARE中关于WHERE CURRENT OF的说明。

output_expression

每行删除之后，由DELETE命令计算并返回的表达式。该表达式可以使用*table_name*所指定的表或USING中列出的表中的任何列名。写成*可返回所有列。

output_name

返回列所使用的名称。

输出

成功完成后，DELETE 命令返回如下形式的命令标签：

DELETE *count*

其中 *count* 是被删除的行数。如果 *count* 为 0，说明没有行匹配 *condition*（这不算错误）。

如果DELETE命令包含RETURNING 子句，其结果将类似于一个SELECT语句，其中包含RETURNING列表中定义的列和值，并在该命令删除的行上进行计算。

注解

PostgreSQL允许通过在 USING子句中指定其他表，在 WHERE条件中引用这些表的列。例如，要删除由给定制片人制作的所有电影，可以这样做：

```
DELETE FROM films USING producers
  WHERE producer_id = producers.id AND producers.name = 'foo';
```

本质上，这里发生的是在films和 producers之间进行连接，并将所有成功连接到的films行标记为删除。这种语法不是标准的。更标准的写法是：

```
DELETE FROM films
  WHERE producer_id IN (SELECT id FROM producers WHERE name = 'foo');
```

在某些情况下，连接形式比子选择形式更容易书写或者执行更快。

示例

删除所有电影，但音乐剧除外：

```
DELETE FROM films WHERE kind <> 'Musical';
```

清空表films：

```
DELETE FROM films;
```

删除已完成的任务，并返回被删除行的完整详情：

```
DELETE FROM tasks WHERE status = 'DONE' RETURNING *;
```

删除游标c_tasks当前所定位的 tasks行：

```
DELETE FROM tasks WHERE CURRENT OF c_tasks;
```

兼容性

这个命令符合SQL标准，不过 USING和RETURNING子句是 PostgreSQL扩展，在 DELETE中使用WITH也是扩展。

DISCARD

DISCARD — discard session state

大纲

```
DISCARD { ALL | PLANS | TEMPORARY | TEMP }
```

描述

`DISCARD`释放与数据库会话关联的内部资源。这些资源通常在会话结束时释放。

`DISCARD TEMP`删除当前会话中创建的所有临时表。`DISCARD PLANS`释放所有内部缓存的查询计划。`DISCARD ALL`将会话重置为新会话的状态。

参数

`TEMPORARY` or `TEMP`

删除当前会话中创建的所有临时表。

`PLANS`

删除当前会话中创建的所有临时表。

`ALL`

释放与当前会话关联的所有临时资源，并将会话重置为初始状态。目前，这与执行以下语句序列具有相同效果：

```
SET SESSION AUTHORIZATION DEFAULT;
RESET ALL;
DEALLOCATE ALL;
CLOSE ALL;
UNLISTEN *;
SELECT pg_advisory_unlock_all();
DISCARD PLANS;
DISCARD SEQUENCES;
DISCARD TEMP;
```

注解

`DISCARD ALL`不能在事务块内部执行。

兼容性

`DISCARD`是PostgreSQL扩展。

DO

DO — execute an anonymous code block

大纲

```
DO [ LANGUAGE lang_name ] code
```

描述

DO执行匿名代码块，换言之，就是执行一个以过程语言编写的瞬时匿名函数。

该代码块会被当作一个不带参数、返回void的函数体。它只会被解析并执行一次。

可选的LANGUAGE子句可以写在代码块之前，也可以写在之后。

参数

code

要执行的过程语言代码。与CREATE FUNCTION一样，它必须以字符串字面值形式指定。建议使用美元引用的字符串字面值。

lang_name

编写该代码所用的过程语言名称。如果省略，默认值为plpgsql。

注解

要使用的过程语言必须已经通过CREATE LANGUAGE安装到当前数据库中。plpgsql默认已安装，而其他语言则不会。

用户必须拥有该过程语言的USAGE权限；如果该语言是不受信任的，则必须是超级用户。这与用该语言创建函数时的权限要求相同。

示例

将模式public中所有视图上的全部权限授予角色webuser：

```
DO $$DECLARE r record;
BEGIN
    FOR r IN SELECT table_schema, table_name FROM information_schema.
tables
             WHERE table_type = 'VIEW' AND table_schema = 'public'
    LOOP
        EXECUTE 'GRANT ALL ON ' || quote_ident(r.table_schema) || '.'
 || quote_ident(r.table_name) || ' TO webuser';
    END LOOP;
END$$;
```

兼容性

SQL 标准中没有DO语句。

另见

CREATE LANGUAGE

DROP AGGREGATE

DROP AGGREGATE — remove an aggregate function

大纲

```
DROP AGGREGATE [ IF EXISTS ] name ( type [ , ... ] ) [ CASCADE |
 RESTRICT ]
```

描述

`DROP AGGREGATE`移除一个现有聚合函数。要执行此命令，当前用户必须是该聚合函数的拥有者。

参数

`IF EXISTS`

如果该聚合函数不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

现有聚合函数的名称（可以被模式限定）。

`type`

聚合函数作用于其上的输入数据类型。要引用一个零参数聚合函数，请在参数说明列表的位置写`*`。

`CASCADE`

自动删除依赖于该聚合函数的对象。

`RESTRICT`

如果有任何对象依赖于该聚合函数，则拒绝删除它。这是默认值。

示例

要移除用于类型`integer`的聚合函数`myavg`：

```
DROP AGGREGATE myavg(integer);
```

兼容性

SQL 标准中没有`DROP AGGREGATE`语句。

另见

ALTER AGGREGATE, CREATE AGGREGATE

DROP CAST

DROP CAST — remove a cast

大纲

```
DROP CAST [ IF EXISTS ] (source_type AS target_type) [ CASCADE |
 RESTRICT ]
```

描述

`DROP CAST`移除一个先前定义的类型转换。

要删除一种类型转换，你必须拥有其源数据类型或目标数据类型。
这与创建该类型转换所需的权限相同。

参数

`IF EXISTS`

如果该类型转换不存在，则不要抛出错误。这种情况下会发出一条提示。

`source_type`

该类型转换的源数据类型的名称。

`target_type`

该类型转换的目标数据类型的名称。

`CASCADE`
`RESTRICT`

这些关键字没有任何作用，因为没有对象依赖于类型转换。

示例

要删除从类型`text`到类型`int`的类型转换：

```
DROP CAST (text AS int);
```

兼容性

`DROP CAST`命令符合 SQL 标准。

另见

CREATE CAST

DROP COLLATION

DROP COLLATION — remove a collation

大纲

```
DROP COLLATION [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP COLLATION`删除一个先前定义的排序规则。要删除一个排序规则，你必须拥有该排序规则。

参数

`IF EXISTS`

如果该排序规则不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

排序规则的名称。该名称可以使用模式限定。

`CASCADE`

自动删除依赖于该排序规则的对象。

`RESTRICT`

如果有任何对象依赖于该排序规则，则拒绝删除该排序规则。这是默认值。

示例

删除名为`german`的排序规则：

```
DROP COLLATION german;
```

兼容性

`DROP COLLATION`命令符合SQL标准，但`IF EXISTS`选项是PostgreSQL扩展。

另见

ALTER COLLATION, CREATE COLLATION

DROP CONVERSION

DROP CONVERSION — remove a conversion

大纲

```
DROP CONVERSION [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP CONVERSION`移除一个先前定义的转换。要能够删除一个转换，你必须拥有该转换。

参数

`IF EXISTS`

如果该转换不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

转换的名称。转换名可以是模式限定的。

`CASCADE`
`RESTRICT`

这些关键字没有任何作用，因为没有对象依赖于转换。

示例

移除名为`myname`的转换：

```
DROP CONVERSION myname;
```

兼容性

SQL 标准中没有`DROP CONVERSION`语句，但有一个 `DROP TRANSLATION`语句，它与`CREATE TRANSLATION` 语句配套，而后者又类似于 PostgreSQL 中的`CREATE CONVERSION`语句。

另见

ALTER CONVERSION, CREATE CONVERSION

DROP DATABASE

DROP DATABASE — remove a database

大纲

```
DROP DATABASE [ IF EXISTS ] name
```

描述

`DROP DATABASE`删除一个数据库。它会移除数据库的系统目录条目，并删除包含数据的目录。只有数据库拥有者才能执行此命令。此外，当你或任何其他人连接到目标数据库时，都不能执行它。（请连接到`postgres`或任何其他数据库来发出此命令。）

`DROP DATABASE`无法撤销。请谨慎使用！

参数

`IF EXISTS`

如果该数据库不存在，则不要抛出错误。这种情况下会发出一个提示。

name

要移除的数据库名称。

注解

`DROP DATABASE`不能在事务块内执行。

由于该命令不能在连接到目标数据库时执行，因此改用程序dropdb可能更方便，它是此命令的一个包装器。

兼容性

SQL 标准中没有`DROP DATABASE`语句。

另见

CREATE DATABASE

DROP DOMAIN

DROP DOMAIN — remove a domain

大纲

```
DROP DOMAIN [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP DOMAIN`移除一个域。只有域的拥有者才能移除它。

参数

`IF EXISTS`

如果该域不存在，则不会抛出错误，而是发出一条提示。

`name`

要移除的现有域的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该域的对象（例如表列）。

`RESTRICT`

如果有任何对象依赖于该域，则拒绝删除它。这是默认值。

示例

要移除域`box`：

```
DROP DOMAIN box;
```

兼容性

此命令符合 SQL 标准，但`IF EXISTS`选项是 PostgreSQL扩展。

另见

CREATE DOMAIN, ALTER DOMAIN

DROP EXTENSION

DROP EXTENSION — remove an extension

大纲

```
DROP EXTENSION [ IF EXISTS ] extension_name [, ...] [ CASCADE |
 RESTRICT ]
```

描述

`DROP EXTENSION`从数据库中移除扩展。删除扩展也会删除它所包含的对象。

要使用`DROP EXTENSION`，必须拥有该扩展。

参数

`IF EXISTS`

如果该扩展不存在，则不要抛出错误。这种情况下会发出一个提示。

`extension_name`

一个已安装扩展的名称。

`CASCADE`

自动删除依赖于该扩展的对象。

`RESTRICT`

如果有任何对象依赖于扩展，则拒绝删除它（其自身成员对象以及同一`DROP`命令中列出的其他扩展除外）。这是默认行为。

示例

要从当前数据库中移除扩展`hstore`：

```
DROP EXTENSION hstore;
```

如果数据库中使用了`hstore`的任何对象，例如某些表具有`hstore`类型的列，则该命令会失败。加上`CASCADE`选项可以强制一并移除这些依赖对象。

兼容性

`DROP EXTENSION`是PostgreSQL扩展。

另见

CREATE EXTENSION, ALTER EXTENSION

DROP FOREIGN DATA WRAPPER

DROP FOREIGN DATA WRAPPER — remove a foreign-data wrapper

大纲

```
DROP FOREIGN DATA WRAPPER [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP FOREIGN DATA WRAPPER`移除一个现有的外部数据包装器。要执行此命令，当前用户必须是该外部数据包装器的拥有者。

参数

`IF EXISTS`

如果该外部数据包装器不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

现有外部数据包装器的名称。

`CASCADE`

自动删除依赖于该外部数据包装器的对象（例如服务器）。

`RESTRICT`

如果有任何对象依赖于该外部数据包装器，则拒绝删除它。这是默认值。

示例

删除外部数据包装器`dbi`：

```
DROP FOREIGN DATA WRAPPER dbi;
```

兼容性

`DROP FOREIGN DATA WRAPPER`符合 ISO/IEC 9075-9（SQL/MED）。`IF EXISTS`子句是PostgreSQL扩展。

另见

CREATE FOREIGN DATA WRAPPER, ALTER FOREIGN DATA WRAPPER

DROP FOREIGN TABLE

DROP FOREIGN TABLE — remove a foreign table

大纲

```
DROP FOREIGN TABLE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP FOREIGN TABLE`移除一个外部表。只有外部表的拥有者才能移除它。

参数

`IF EXISTS`

如果该外部表不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

要删除的外部表名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该外部表的对象（例如视图）。

`RESTRICT`

如果有任何对象依赖于该外部表，则拒绝删除它。这是默认值。

示例

要销毁两个外部表`films`和`distributors`：

```
DROP FOREIGN TABLE films, distributors;
```

兼容性

此命令符合 ISO/IEC 9075-9（SQL/MED），但标准每条命令只允许删除一个外部表，而`IF EXISTS`选项是PostgreSQL扩展。

另见

ALTER FOREIGN TABLE, CREATE FOREIGN TABLE

DROP FUNCTION

DROP FUNCTION — remove a function

大纲

```
DROP FUNCTION [ IF EXISTS ] name ( [ [ argmode ] [ argname ] argtype [
, ...] ] )
    [ CASCADE | RESTRICT ]
```

描述

DROP FUNCTION移除一个现有函数的定义。要执行此命令，用户必须是该函数的拥有者。必须指定该函数的参数类型，因为可能存在多个名称相同但参数列表不同的函数。

参数

IF EXISTS

如果该函数不存在，则不要抛出错误。这种情况下会发出一个提示。

name

现有函数的名称（可以被模式限定）。

argmode

参数的模式：IN、OUT、INOUT或VARIADIC。如果省略，默认值为IN。注意，DROP FUNCTION实际上并不关注OUT参数，因为确定函数标识只需要输入参数。因此，只列出IN、INOUT和 VARIADIC参数就足够了。

argname

一个参数的名称。注意，DROP FUNCTION 实际上并不关注参数名，因为确定函数标识只需要参数数据类型。

argtype

函数参数的数据类型（如果有，可以被模式限定）。

CASCADE

自动删除依赖于该函数的对象（例如操作符或触发器）。

RESTRICT

如果有任何对象依赖于该函数，则拒绝删除它。这是默认值。

示例

此命令删除平方根函数：

```
DROP FUNCTION sqrt(integer);
```

兼容性

`DROP FUNCTION`语句在 SQL 标准中有定义，但与该命令不兼容。

另见

CREATE FUNCTION, ALTER FUNCTION

DROP GROUP

DROP GROUP — remove a database role

大纲

```
DROP GROUP [ IF EXISTS ] name [, ...]
```

描述

`DROP GROUP` 现在是DROP ROLE的别名。

兼容性

SQL 标准中没有 `DROP GROUP` 语句。

另见

DROP ROLE

DROP INDEX

DROP INDEX — remove an index

大纲

```
DROP INDEX [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP INDEX`从数据库系统中删除一个现有索引。要执行此命令，你必须是该索引的拥有者。

参数

`IF EXISTS`

如果该索引不存在，则不要抛出错误。这种情况下会发出一条提示。

`name`

要移除的索引名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该索引的对象。

`RESTRICT`

如果有任何对象依赖于该索引，则拒绝删除它。这是默认值。

示例

此命令会删除索引`title_idx`：

```
DROP INDEX title_idx;
```

兼容性

`DROP INDEX`是PostgreSQL 的一种语言扩展。SQL 标准中没有关于索引的规定。

另见

CREATE INDEX

DROP LANGUAGE

DROP LANGUAGE — remove a procedural language

大纲

```
DROP [ PROCEDURAL ] LANGUAGE [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP LANGUAGE`移除一种先前注册的过程语言的定义。要使用`DROP LANGUAGE`，必须是超级用户或该语言的所有者。

> **注意**
>
> 从PostgreSQL 9.1 起，大多数过程语言都已被实现为“扩展”，因此应当使用DROP EXTENSION而不是`DROP LANGUAGE`来移除它们。

参数

`IF EXISTS`

如果该语言不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

现有过程语言的名称。为向后兼容，名称可以用单引号括起。

`CASCADE`

自动删除依赖于该语言的对象（例如以该语言编写的函数）。

`RESTRICT`

如果有任何对象依赖于该语言，则拒绝删除它。这是默认值。

示例

此命令移除过程语言`plsample`：

```
DROP LANGUAGE plsample;
```

兼容性

SQL 标准中没有`DROP LANGUAGE`语句。

另见

ALTER LANGUAGE, CREATE LANGUAGE, droplang

DROP OPERATOR

DROP OPERATOR — remove an operator

大纲

```
DROP OPERATOR [ IF EXISTS ] name ( { left_type | NONE } , { right_type
 | NONE } ) [ CASCADE | RESTRICT ]
```

描述

`DROP OPERATOR`从数据库系统中删除一个现有操作符。要执行此命令，你必须是该操作符的拥有者。

参数

`IF EXISTS`

如果该操作符不存在，则不要抛出错误。这种情况下会发出一条提示。

`name`

一个现有操作符的名称（可以是模式限定的）。

`left_type`

操作符左操作数的数据类型；如果该操作符没有左操作数，请写 `NONE`。

`right_type`

操作符右操作数的数据类型；如果操作符没有右操作数，请写`NONE`。

`CASCADE`

自动删除依赖于该操作符的对象。

`RESTRICT`

如果有任何对象依赖于该操作符，则拒绝删除它。这是默认值。

示例

删除幂操作符`a^b`，其类型为 `integer`：

```
DROP OPERATOR ^ (integer, integer);
```

删除前缀一元按位取反操作符 `~b`，其类型为 `bit`：

```
DROP OPERATOR ~ (none, bit);
```

删除后缀一元阶乘操作符 `x!`，其类型为 `bigint`：

```
DROP OPERATOR ! (bigint, none);
```

兼容性

SQL 标准中没有`DROP OPERATOR`语句。

另见

CREATE OPERATOR, ALTER OPERATOR

DROP OPERATOR CLASS

DROP OPERATOR CLASS — remove an operator class

大纲

```
DROP OPERATOR CLASS [ IF EXISTS ] name USING index_method [ CASCADE |
 RESTRICT ]
```

描述

`DROP OPERATOR CLASS`删除一个现有操作符类。要执行此命令，你必须是该操作符类的拥有者。

`DROP OPERATOR CLASS`不会删除该类所引用的任何操作符或函数。如果有索引依赖于该操作符类，你将需要指定`CASCADE` 才能完成删除。

参数

`IF EXISTS`

如果该操作符类不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

一个现有操作符类的名称（可以被模式限定）。

`index_method`

该操作符类所属索引访问方法的名称。

`CASCADE`

自动删除依赖于该操作符类的对象。

`RESTRICT`

如果有任何对象依赖于该操作符类，则拒绝删除它。这是默认值。

注解

`DROP OPERATOR CLASS`不会删除包含该类的操作符族，即使该族中已经没有任何其他成员（尤其是在该族由`CREATE OPERATOR CLASS`隐式创建的情况下）。空的操作符族并无害处，但为了整洁起见，你也许会希望用 `DROP OPERATOR FAMILY`删除该族；或者更好的做法是，一开始就直接使用`DROP OPERATOR FAMILY`。

示例

移除 B-树操作符类`widget_ops`：

```
DROP OPERATOR CLASS widget_ops USING btree;
```

如果仍有索引使用该操作符类，此命令将不会成功。加上`CASCADE`可以在删除操作符类的同时删除这类索引。

兼容性

SQL 标准中没有`DROP OPERATOR CLASS`语句。

另见

ALTER OPERATOR CLASS, CREATE OPERATOR CLASS, DROP OPERATOR FAMILY

DROP OPERATOR FAMILY

DROP OPERATOR FAMILY — remove an operator family

大纲

```
DROP OPERATOR FAMILY [ IF EXISTS ] name USING index_method [ CASCADE |
 RESTRICT ]
```

描述

`DROP OPERATOR FAMILY`删除一个现有的操作符族。要执行此命令，你必须是该操作符族的拥有者。

`DROP OPERATOR FAMILY`还会删除该族中包含的所有操作符类，但不会删除该族引用的任何操作符或函数。如果有任何索引依赖于该族内的操作符类，你需要指定`CASCADE`才能完成删除。

参数

`IF EXISTS`

如果该操作符族不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

一个现有操作符族的名称（可以是模式限定的）。

`index_method`

该操作符族所对应的索引访问方法名称。

`CASCADE`

自动删除依赖于该操作符族的对象。

`RESTRICT`

如果有任何对象依赖于该操作符族，则拒绝删除它。这是默认值。

示例

移除 B-tree 操作符族`float_ops`：

```
DROP OPERATOR FAMILY float_ops USING btree;
```

如果仍有索引使用该族内的操作符类，此命令将不会成功。加上`CASCADE`可以在删除操作符族的同时删除这类索引。

兼容性

SQL 标准中没有`DROP OPERATOR FAMILY`语句。

另见

ALTER OPERATOR FAMILY, CREATE OPERATOR FAMILY, ALTER OPERATOR CLASS, CREATE OPERATOR CLASS, DROP OPERATOR CLASS

DROP OWNED

DROP OWNED — remove database objects owned by a database role

大纲

```
DROP OWNED BY name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP OWNED`删除当前数据库中由指定角色之一拥有的所有对象。同时，授予给这些角色的、针对当前数据库中的对象或共享对象（数据库、表空间）的任何权限也会被撤销。

参数

name

一个角色的名称；该角色拥有的对象将被删除，并且授予给该角色的权限将被撤销。

`CASCADE`

自动删除依赖于该受影响对象的对象。

`RESTRICT`

如果有其他数据库对象依赖于某个受影响对象，则拒绝删除该角色拥有的对象。
这是默认值。

注解

`DROP OWNED`经常被用来为移除一个或多个角色做准备。因为`DROP OWNED`只影响当前数据库中的对象，所以通常需要在包含待移除角色所拥有对象的每个数据库中执行此命令。

使用`CASCADE`选项可能会使该命令递归到由其他用户拥有的对象。

REASSIGN OWNED命令提供了另一种选择，它会重新分配由一个或多个角色拥有的所有数据库对象的拥有关系。不过，`REASSIGN OWNED`不处理针对其他对象的权限。

由这些角色拥有的数据库和表空间不会被移除。

更多讨论请见第 20.4 节。

兼容性

`DROP OWNED`命令是PostgreSQL扩展。

另见

REASSIGN OWNED, DROP ROLE

DROP ROLE

DROP ROLE — remove a database role

大纲

```
DROP ROLE [ IF EXISTS ] name [, ...]
```

描述

`DROP ROLE`移除指定的角色。要删除超级用户角色，你自己也必须是超级用户；要删除非超级用户角色，必须拥有`CREATEROLE`权限。

如果某角色仍在该集簇的任何数据库中被引用，就不能移除它，否则会报错。在删除角色之前，必须删除它拥有的所有对象（或重新分配其拥有关系），
并撤销该角色在其他对象上被授予的任何权限。REASSIGN OWNED和DROP OWNED命令可用于此目的；更多讨论参见第 20.4 节。

不过，不必手工移除与该角色相关的角色成员资格；`DROP ROLE`会自动撤销目标角色在其他角色中的任何成员资格，以及其他角色在目标角色中的任何成员资格。其他角色不会被删除，
也不会受到影响。

参数

`IF EXISTS`

如果该角色不存在，则不要抛出错误。这种情况下会发出一个提示。

name

要移除的角色名称。

注解

PostgreSQL包含一个程序dropuser，它与此命令具有相同功能（实际上它就是调用此命令），但可以从命令 shell 中运行。

示例

删除一个角色：

```
DROP ROLE jonathan;
```

兼容性

SQL 标准定义了`DROP ROLE`，但它一次只允许删除一个角色，并且规定的权限要求也不同于PostgreSQL。

另见

CREATE ROLE, ALTER ROLE, SET ROLE

DROP RULE

DROP RULE — remove a rewrite rule

大纲

```
DROP RULE [ IF EXISTS ] name ON table [ CASCADE | RESTRICT ]
```

描述

`DROP RULE`移除一条重写规则。

参数

`IF EXISTS`

如果该规则不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

要移除的规则名称。

`table`

该规则适用的表或视图名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该规则的对象。

`RESTRICT`

如果有任何对象依赖于该规则，则拒绝删除它。这是默认值。

示例

要移除重写规则`newrule`：

```
DROP RULE newrule ON mytable;
```

兼容性

SQL 标准中没有 `DROP RULE` 语句。

另见

CREATE RULE

DROP SCHEMA

DROP SCHEMA — remove a schema

大纲

```
DROP SCHEMA [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP SCHEMA`从数据库中移除一个或多个模式。

一个模式只能由其拥有者或超级用户删除。注意，
即使该模式的拥有者并不拥有模式内的某些对象，
也仍然可以删除该模式（从而删除其中的全部对象）。

参数

`IF EXISTS`

如果该模式不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

模式名称。

`CASCADE`

自动删除模式中包含的对象（表、函数等）。

`RESTRICT`

如果该模式中含有任何对象，则拒绝删除它。这是默认值。

示例

要从数据库中移除模式`mystuff`及其包含的全部内容：

```
DROP SCHEMA mystuff CASCADE;
```

兼容性

`DROP SCHEMA`完全符合 SQL 标准，但标准每条命令只允许删除一个模式，另外`IF EXISTS`选项是PostgreSQL扩展。

另见

ALTER SCHEMA, CREATE SCHEMA

DROP SEQUENCE

DROP SEQUENCE — remove a sequence

大纲

```
DROP SEQUENCE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP SEQUENCE`移除序列。只有该序列的拥有者或超级用户才能删除它。

参数

`IF EXISTS`

如果该序列不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

要删除的序列名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该序列的对象。

`RESTRICT`

如果有任何对象依赖于该序列，则拒绝删除它。这是默认值。

示例

要移除序列`serial`：

```
DROP SEQUENCE serial;
```

兼容性

`DROP SEQUENCE`符合SQL标准，但标准每条命令只允许删除一个序列，而`IF EXISTS`选项是PostgreSQL扩展。

另见

CREATE SEQUENCE, ALTER SEQUENCE

DROP SERVER

DROP SERVER — remove a foreign server descriptor

大纲

```
DROP SERVER [ IF EXISTS ] server_name [ CASCADE | RESTRICT ]
```

描述

`DROP SERVER`移除一个现有外部服务器描述符。要执行此命令，当前用户必须是该服务器的拥有者。

参数

`IF EXISTS`

如果该服务器不存在，则不要抛出错误。这种情况下会发出一个提示。

`server_name`

现有服务器的名称。

`CASCADE`

自动删除依赖于该服务器的对象（例如用户映射）。

`RESTRICT`

如果有任何对象依赖于该服务器，则拒绝删除它。这是默认值。

示例

如果服务器`foo`存在，则删除它：

```
DROP SERVER IF EXISTS foo;
```

兼容性

`DROP SERVER`符合 ISO/IEC 9075-9（SQL/MED）。`IF EXISTS`子句是PostgreSQL扩展。

另见

CREATE SERVER, ALTER SERVER

DROP TABLE

DROP TABLE — remove a table

大纲

```
DROP TABLE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP TABLE` 从数据库中删除表。只有表的拥有者才能删除表。要在不破坏表的情况下清空表中的行，可使用DELETE或TRUNCATE。

`DROP TABLE`总会移除目标表上存在的所有索引、规则、触发器和约束。但是，要删除一个被视图或另一张表的外键约束引用的表，必须指定`CASCADE`。（`CASCADE`
会将依赖于该表的视图整个删除，但在外键场景下，它只会删除外键约束，
不会把另一张表整体删除。）

参数

`IF EXISTS`

如果该表不存在，则不要抛出错误。这种情况下会发出一个提示。

name

要删除的表名称（可选地使用模式限定）。

`CASCADE`

自动删除依赖于该表的对象（例如视图）。

`RESTRICT`

如果有任何对象依赖于该表，则拒绝删除它。这是默认值。

示例

要删除两个表`films`和`distributors`：

```
DROP TABLE films, distributors;
```

兼容性

此命令符合 SQL 标准，只是标准每条命令只允许删除一个表，且`IF EXISTS`选项是PostgreSQL扩展。

另见

ALTER TABLE, CREATE TABLE

DROP TABLESPACE

DROP TABLESPACE — remove a tablespace

大纲

```
DROP TABLESPACE [ IF EXISTS ] tablespace_name
```

描述

`DROP TABLESPACE`从系统中移除一个表空间。

一个表空间只能由其拥有者或超级用户删除。删除前，该表空间中必须不含任何数据库对象。即使当前数据库中没有对象正在使用该表空间，
其他数据库中的对象也可能仍然驻留在该表空间中。此外，如果任何活动会话的temp_tablespaces设置中列出了该表空间，那么由于该表空间中可能仍有临时文件，`DROP`也可能失败。

参数

`IF EXISTS`

如果该表空间不存在，则不要抛出错误。这种情况下会发出一个提示。

`tablespace_name`

表空间的名称。

注解

`DROP TABLESPACE`不能在事务块内执行。

示例

要从系统中移除表空间`mystuff`：

```
DROP TABLESPACE mystuff;
```

兼容性

`DROP TABLESPACE`是一种PostgreSQL扩展。

另见

CREATE TABLESPACE, ALTER TABLESPACE

DROP TEXT SEARCH CONFIGURATION

DROP TEXT SEARCH CONFIGURATION — remove a text search configuration

大纲

```
DROP TEXT SEARCH CONFIGURATION [ IF EXISTS ] name [ CASCADE |
 RESTRICT ]
```

描述

`DROP TEXT SEARCH CONFIGURATION`删除一个现有文本检索配置。要执行此命令，你必须是该配置的拥有者。

参数

`IF EXISTS`

如果该文本检索配置不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

现有文本检索配置的名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该文本检索配置的对象。

`RESTRICT`

如果有任何对象依赖于该文本检索配置，则拒绝删除该配置。这是默认值。

示例

移除文本检索配置`my_english`：

```
DROP TEXT SEARCH CONFIGURATION my_english;
```

如果已有索引在`to_tsvector`调用中引用了该配置，此命令将不会成功。加上`CASCADE`可以在删除该文本检索配置的同时删除这类索引。

兼容性

SQL 标准中没有`DROP TEXT SEARCH CONFIGURATION`语句。

另见

ALTER TEXT SEARCH CONFIGURATION, CREATE TEXT SEARCH CONFIGURATION

DROP TEXT SEARCH DICTIONARY

DROP TEXT SEARCH DICTIONARY — remove a text search dictionary

大纲

```
DROP TEXT SEARCH DICTIONARY [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP TEXT SEARCH DICTIONARY`删除一个现有文本检索词典。要执行此命令，你必须是该词典的拥有者。

参数

`IF EXISTS`

如果该文本检索词典不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

一个现有文本检索词典的名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该文本检索词典的对象。

`RESTRICT`

如果有任何对象依赖于该文本检索词典，则拒绝删除它。这是默认值。

示例

移除文本检索词典`english`：

```
DROP TEXT SEARCH DICTIONARY english;
```

如果已有文本检索配置使用该词典，此命令将不会成功。加上`CASCADE`可以在删除词典的同时删除这类配置。

兼容性

SQL 标准中没有`DROP TEXT SEARCH DICTIONARY`语句。

另见

ALTER TEXT SEARCH DICTIONARY, CREATE TEXT SEARCH DICTIONARY

DROP TEXT SEARCH PARSER

DROP TEXT SEARCH PARSER — remove a text search parser

大纲

```
DROP TEXT SEARCH PARSER [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP TEXT SEARCH PARSER`删除一个现有文本检索解析器。要使用此命令，你必须是超级用户。

参数

`IF EXISTS`

如果该文本检索解析器不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

一个现有文本检索解析器的名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该文本检索解析器的对象。

`RESTRICT`

如果有任何对象依赖于该文本检索解析器，则拒绝删除它。这是默认值。

示例

移除文本检索解析器`my_parser`：

```
DROP TEXT SEARCH PARSER my_parser;
```

如果已有文本检索配置使用该解析器，此命令将不会成功。加上`CASCADE`可以在删除解析器的同时删除这类配置。

兼容性

SQL 标准中没有`DROP TEXT SEARCH PARSER`语句。

另见

ALTER TEXT SEARCH PARSER, CREATE TEXT SEARCH PARSER

DROP TEXT SEARCH TEMPLATE

DROP TEXT SEARCH TEMPLATE — remove a text search template

大纲

```
DROP TEXT SEARCH TEMPLATE [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

描述

`DROP TEXT SEARCH TEMPLATE`删除一个现有文本检索模板。要使用此命令，你必须是超级用户。

参数

`IF EXISTS`

如果该文本检索模板不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

现有文本检索模板的名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该文本检索模板的对象。

`RESTRICT`

如果有任何对象依赖于该文本检索模板，则拒绝删除该模板。这是默认值。

示例

移除文本检索模板`thesaurus`：

```
DROP TEXT SEARCH TEMPLATE thesaurus;
```

如果已有文本检索词典使用该模板，此命令将不会成功。加上`CASCADE` 可以在删除该模板的同时删除这类词典。

兼容性

SQL 标准中没有`DROP TEXT SEARCH TEMPLATE`语句。

另见

ALTER TEXT SEARCH TEMPLATE, CREATE TEXT SEARCH TEMPLATE

DROP TRIGGER

DROP TRIGGER — remove a trigger

大纲

```
DROP TRIGGER [ IF EXISTS ] name ON table [ CASCADE | RESTRICT ]
```

描述

`DROP TRIGGER`移除一个现有触发器的定义。要执行此命令，当前用户必须是定义该触发器的表的所有者。

参数

`IF EXISTS`

如果该触发器不存在，则不要抛出错误。这种情况下会发出一个提示。

`name`

要移除的触发器的名称。

`table`

定义该触发器的表的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该触发器的对象。

`RESTRICT`

如果有任何对象依赖于该触发器，则拒绝删除它。这是默认值。

示例

删除表`films`上的触发器`if_dist_exists`：

```
DROP TRIGGER if_dist_exists ON films;
```

兼容性

PostgreSQL中的`DROP TRIGGER`语句与 SQL 标准不兼容。在 SQL 标准中，触发器名称不是表局部的，因此该命令仅为 `DROP TRIGGER name`。

参见

CREATE TRIGGER

DROP TYPE

DROP TYPE — remove a data type

大纲

```
DROP TYPE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP TYPE`移除一个用户定义的数据类型。只有该类型的拥有者才能移除它。

参数

`IF EXISTS`

如果该类型不存在，则不会抛出错误，而是发出一条提示。

`name`

要移除的数据类型的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该类型的对象（例如表列、函数和操作符）。

`RESTRICT`

如果有任何对象依赖于该类型，则拒绝删除它。这是默认值。

示例

要移除数据类型`box`：

```
DROP TYPE box;
```

兼容性

此命令与 SQL 标准中的对应命令相似，但`IF EXISTS`选项是 PostgreSQL扩展。不过请注意，`CREATE TYPE`命令的大部分内容，以及 PostgreSQL中的数据类型扩展机制，都不同于 SQL 标准。

另见

ALTER TYPE, CREATE TYPE

DROP USER

DROP USER — remove a database role

大纲

```
DROP USER [ IF EXISTS ] name [, ...]
```

描述

`DROP USER`只是DROP ROLE的另一种拼写。

兼容性

`DROP USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义留给具体实现决定。

另见

DROP ROLE

DROP USER MAPPING

DROP USER MAPPING — remove a user mapping for a foreign server

大纲

```
DROP USER MAPPING [ IF EXISTS ] FOR { user_name | USER | CURRENT_USER
 | PUBLIC } SERVER server_name
```

描述

`DROP USER MAPPING`删除外部服务器上的一个现有用户映射。

外部服务器的拥有者可以删除该服务器上任何用户的用户映射。此外，
如果某个用户已被授予该服务器上的`USAGE`权限，
那么该用户也可以删除其自己用户名对应的用户映射。

参数

`IF EXISTS`

如果该用户映射不存在，则不要报错。这种情况下会发出一个提示。

`user_name`

该映射的用户名。`CURRENT_USER`和`USER`都匹配当前用户的名称。`PUBLIC`
用于匹配系统中当前及未来的所有用户名。

`server_name`

该用户映射所属服务器的名称。

示例

如果用户`bob`在服务器`foo`上的用户映射存在，则将其删除：

```
DROP USER MAPPING IF EXISTS FOR bob SERVER foo;
```

兼容性

`DROP USER MAPPING`符合 ISO/IEC 9075-9（SQL/MED）。`IF EXISTS`子句是PostgreSQL扩展。

另见

CREATE USER MAPPING, ALTER USER MAPPING

DROP VIEW

DROP VIEW — remove a view

大纲

```
DROP VIEW [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP VIEW`删除一个现有视图。要执行此命令，你必须是该视图的拥有者。

参数

`IF EXISTS`

如果该视图不存在，则不要抛出错误。这种情况下会发出一个提示。

name

要移除的视图名称（可选地使用模式限定）。

`CASCADE`

自动删除依赖于该视图的对象（例如其他视图）。

`RESTRICT`

如果有任何对象依赖于该视图，则拒绝删除它。这是默认值。

示例

这个命令将移除名为`kinds`的视图：

```
DROP VIEW kinds;
```

兼容性

此命令符合 SQL 标准，只是标准每条命令只允许删除一个视图，且`IF EXISTS`选项是PostgreSQL扩展。

另见

ALTER VIEW, CREATE VIEW

END

END — commit the current transaction

大纲

```
END [ WORK | TRANSACTION ]
```

描述

`END`提交当前事务。该事务所做的所有更改都会对其他会话可见，并且如果发生崩溃，也能保证其持久性。该命令是PostgreSQL扩展，等效于COMMIT。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用ROLLBACK中止事务。

在事务块之外发出`END`不会造成任何影响，但会产生一条警告消息。

示例

提交当前事务，并使所有更改永久生效：

```
END;
```

兼容性

命令`END`是PostgreSQL扩展，提供了与 SQL 标准规定的COMMIT等效的功能。

另见

BEGIN, COMMIT, ROLLBACK

EXECUTE

EXECUTE — execute a prepared statement

大纲

```
EXECUTE name [ ( parameter [, ...] ) ]
```

Description

EXECUTE用于执行一个先前创建的预备语句。由于预备语句只在一个会话期间内存在，因此该预备语句必须由当前会话中较早执行的PREPARE语句创建。

如果创建该语句的PREPARE语句指定了某些参数，那么传递给EXECUTE语句的参数集合就必须与之兼容，否则将引发错误。注意，预备语句（不同于函数）不会根据其参数的类型或个数进行重载；预备语句的名称在一个数据库会话内必须唯一。

有关预备语句的创建和用法的更多信息，请参见PREPARE。

Parameters

name

要执行的预备语句名称。

parameter

预备语句中某个参数的实际值。它必须是一个表达式，并且其求值结果必须与该参数的数据类型兼容，该数据类型在创建预备语句时已确定。

Outputs

EXECUTE返回的命令标签是该预备语句的命令标签，而不是EXECUTE。

Examples

示例位于示例一节，详见PREPARE文档。

Compatibility

SQL 标准包含EXECUTE语句，但它只用于嵌入式 SQL。这里的EXECUTE语句还使用了稍有不同的语法。

See Also

DEALLOCATE, PREPARE

EXPLAIN

EXPLAIN — show the execution plan of a statement

大纲

```
EXPLAIN [ ( option [, ...] ) ] statement
EXPLAIN [ ANALYZE ] [ VERBOSE ] statement

where option can be one of:

    ANALYZE [ boolean ]
    VERBOSE [ boolean ]
    COSTS [ boolean ]
    BUFFERS [ boolean ]
    FORMAT { TEXT | XML | JSON | YAML }
```

描述

这个命令显示PostgreSQL规划器为给定语句生成的执行计划。
执行计划会显示将如何扫描该语句引用的表 – 例如普通顺序扫描、索引扫描等 –，如果引用了多个表，还会显示将使用哪些连接算法来汇集每个输入表中的所需行。

显示结果中最关键的部分是语句执行代价的估计值，
它是规划器对运行该语句需要多长时间的猜测（以任意代价单位衡量，
但按惯例表示磁盘页面抓取次数）。实际上会显示两个数字：返回第一行之前的启动代价，
以及返回全部行的总代价。对大多数查询来说，总代价才是关键；但在某些场景中，例如EXISTS中的子查询，
规划器会选择启动代价最小而不是总代价最小的计划（因为无论如何执行器都会在得到一行后停止）。此外，如果你用LIMIT子句限制返回的行数，规划器会在端点代价之间作适当插值，
以估计哪个计划实际上最便宜。

ANALYZE 选项导致语句被真正执行，而不仅仅是规划。每个计划节点内部耗费的总时间（以毫秒计）以及它实际返回的总行数会被加入显示。这有助于查看规划器的估计是否接近现实。

> **重要**
>
> 请记住，使用 ANALYZE 选项时，语句会被实际执行。虽然 EXPLAIN 会丢弃 SELECT 返回的任何输出，但语句的其他副作用仍会照常发生。如果希望使用 EXPLAIN ANALYZE 分析 INSERT, UPDATE, DELETE, CREATE TABLE AS 或 EXECUTE 语句而不让命令影响数据，可以采用以下方法：
>
> ```
> BEGIN;
> EXPLAIN ANALYZE ...;
> ROLLBACK;
> ```

不使用括号包围选项列表时，只能指定ANALYZE和VERBOSE选项，而且只能按这个顺序指定。在PostgreSQL 9.0 之前，只支持这种不带括号的语法。预计所有新选项都将仅在带括号的语法中得到支持。

参数

ANALYZE

执行命令并显示实际的运行时间。此参数默认为 FALSE。

VERBOSE

显示有关计划的附加信息。具体包括：计划树中每个节点的输出列列表、模式限定的表名和函数名、总是用其范围表别名标注表达式中的变量，以及总是打印显示统计信息的每个触发器名称。此参数默认为FALSE。

COSTS

包含每个计划节点的估计启动代价和总代价，以及估计行数和每行的估计宽度。此参数默认为TRUE。

BUFFERS

包含缓冲区使用情况的信息。具体来说，包含共享块命中、读和写的次数，本地块命中、读和写的次数，以及临时块读和写的次数。共享块、本地块和临时块分别包含表和索引、临时表和临时索引，以及排序和物化计划中使用的磁盘块。
上层节点显示的块数包括其所有子节点使用的块。在文本格式中，只打印非零值。
此参数只能与 ANALYZE 参数一起使用。默认为 FALSE。

FORMAT

指定输出格式，可以是 TEXT、XML、JSON 或 YAML。非文本输出包含与文本输出相同的信息，但更容易被程序解析。此参数默认为TEXT。

boolean

指定所选选项应开启还是关闭。可以写TRUE、ON或1来启用选项，写FALSE、OFF或0来禁用它。*boolean*值也可以省略，在这种情况下假定其值为TRUE。

statement

任何你希望查看其执行计划的SELECT、INSERT、UPDATE、DELETE、VALUES、EXECUTE、DECLARE或CREATE TABLE AS语句。

注解

关于 PostgreSQL 中优化器对代价信息的使用，文档很少。更多信息参见第 14.1 节。

为了让 PostgreSQL 查询规划器在优化查询时能做出合理知情的决定，应当运行ANALYZE语句来记录表中数据分布的统计信息。如果你没有这样做（或者自上次运行 ANALYZE 以来表中数据的统计分布发生了显著变化），估计的代价不太可能符合查询的真实属性，
因此可能选出一个较差的查询计划。

为了测量执行计划中每个节点的运行时代价，EXPLAIN ANALYZE 的当前实现可能给查询执行增加相当大的剖析开销。因此，对某个查询运行 EXPLAIN ANALYZE 有时可能比正常执行该查询花费显著更长的时间。开销的大小取决于查询的性质。

示例

要显示一个只有单个integer列且包含 10000 行的表上的简单查询计划：

```
EXPLAIN SELECT * FROM foo;

                       QUERY PLAN
-------------------------------------------------------
 Seq Scan on foo  (cost=0.00..155.00 rows=10000 width=4)
(1 row)
```

下面是同一查询，但使用 JSON 输出格式：

```
EXPLAIN (FORMAT JSON) SELECT * FROM foo;
           QUERY PLAN
--------------------------------
 [                              +
   {                            +
     "Plan": {                  +
       "Node Type": "Seq Scan",+
       "Relation Name": "foo", +
       "Alias": "foo",         +
       "Startup Cost": 0.00,   +
       "Total Cost": 155.00,   +
       "Plan Rows": 10000,     +
       "Plan Width": 4         +
     }                          +
   }                            +
 ]
(1 row)
```

如果存在索引，并且我们使用了带有可索引WHERE条件的查询，EXPLAIN
可能会显示不同的计划：

```
EXPLAIN SELECT * FROM foo WHERE i = 4;

                         QUERY PLAN
----------------------------------------------------------------
 Index Scan using fi on foo  (cost=0.00..5.98 rows=1 width=4)
   Index Cond: (i = 4)
(2 rows)
```

下面是同一查询，但采用 YAML 格式：

```
EXPLAIN (FORMAT YAML) SELECT * FROM foo WHERE i='4';
          QUERY PLAN
-------------------------------
 - Plan:                      +
     Node Type: "Index Scan"  +
     Scan Direction: "Forward"+
     Index Name: "fi"         +
     Relation Name: "foo"     +
     Alias: "foo"             +
     Startup Cost: 0.00       +
     Total Cost: 5.98         +
     Plan Rows: 1             +
     Plan Width: 4            +
     Index Cond: "(i = 4)"
```

```
(1 row)
```

XML 格式留给读者自行练习。

下面是同一计划，但隐藏了代价估计：

```
EXPLAIN (COSTS FALSE) SELECT * FROM foo WHERE i = 4;

        QUERY PLAN
----------------------------
 Index Scan using fi on foo
   Index Cond: (i = 4)
(2 rows)
```

下面是使用聚合函数的查询的查询计划示例：

```
EXPLAIN SELECT sum(i) FROM foo WHERE i < 10;

                            QUERY PLAN
-----------------------------------------------------------------------
 Aggregate  (cost=23.93..23.93 rows=1 width=4)
   ->  Index Scan using fi on foo  (cost=0.00..23.92 rows=6 width=4)
         Index Cond: (i < 10)
(3 rows)
```

下面是使用 `EXPLAIN EXECUTE` 显示预备查询执行计划的示例：

```
PREPARE query(int, int) AS SELECT sum(bar) FROM test
    WHERE id > $1 AND id < $2
    GROUP BY foo;

EXPLAIN ANALYZE EXECUTE query(100, 200);

                                                        QUERY PLAN

-----------------------------------------------------------------------
--------------------------------------------------
 HashAggregate  (cost=9.54..9.54 rows=1 width=8) (actual time=0.156..
0.161 rows=11 loops=1)
   Group Key: foo
   ->  Index Scan using test_pkey on test  (cost=0.29..9.29 rows=50
 width=8) (actual time=0.039..0.091 rows=99 loops=1)
         Index Cond: ((id > $1) AND (id < $2))
 Planning time: 0.197 ms
 Execution time: 0.225 ms
(6 rows)
```

当然，这里显示的具体数字取决于所涉及表的实际内容。还要注意，由于规划器的改进，这些数字甚至所选查询策略在PostgreSQL的不同版本之间都可能发生变化。此外，`ANALYZE` 命令使用随机采样来估计数据统计信息；因此，即使表中数据的实际分布没有变化，在重新运行一次`ANALYZE`之后，代价估计也可能发生变化。

兼容性

SQL 标准中没有定义`EXPLAIN`语句。

另见

ANALYZE

FETCH

FETCH — retrieve rows from a query using a cursor

大纲

```
FETCH [ direction [ FROM | IN ] ] cursor_name

where direction can be empty or one of:

    NEXT
    PRIOR
    FIRST
    LAST
    ABSOLUTE count
    RELATIVE count
    count
    ALL
    FORWARD
    FORWARD count
    FORWARD ALL
    BACKWARD
    BACKWARD count
    BACKWARD ALL
```

描述

FETCH使用先前创建的游标检索行。

游标带有一个关联位置，FETCH会使用该位置。游标位置可以位于查询结果的第一行之前、结果中的任意一行上，或者位于结果的最后一行之后。游标创建时位于第一行之前。提取若干行后，游标会定位在最近提取的那一行上。如果FETCH 越过了可用行的末尾，那么游标会留在最后一行之后；如果是向后提取，则会留在第一行之前。FETCH ALL或者 FETCH BACKWARD ALL总是会让游标位于最后一行之后或第一行之前。

NEXT、PRIOR、FIRST、LAST、ABSOLUTE、RELATIVE 这些形式会在适当地移动游标后提取一行。如果不存在这样的行，则返回空结果，并根据情况将游标定位在第一行之前或最后一行之后。

使用FORWARD和BACKWARD的形式会在向前或向后的方向上提取指定数量的行，并将游标定位在最后返回的那一行上（如果*count*超过可用行数，则定位在所有行之后或之前）。

RELATIVE 0、FORWARD 0以及 BACKWARD 0都会请求提取当前行而不移动游标，也就是重新提取最近一次提取的行。除非游标位于第一行之前或最后一行之后，否则该操作都会成功；在这两种情况下，不会返回任何行。

> **注意**
>
> 本页面描述的是 SQL 命令层面上的游标用法。如果想要在 PL/pgSQL函数中使用游标，规则会有所不同 – 请见第 39.7 节。

参数

direction

direction 定义抓取方向和要抓取的行数。它可以是以下值之一：

NEXT

提取下一行。如果省略*direction*，这将是默认值。

PRIOR

提取前一行。

FIRST

提取查询的第一行（与ABSOLUTE 1相同）。

LAST

提取查询的最后一行（与ABSOLUTE -1相同）。

ABSOLUTE *count*

提取查询的第*count*行；如果*count*为负，则提取从末尾算起的第abs(*count*)行。如果 *count*超出范围，则定位在第一行之前或最后一行之后。特别地，ABSOLUTE 0会定位在第一行之前。

RELATIVE *count*

抓取后面的第*count*行，或前面的第abs(*count*)行（当*count*为负时）。RELATIVE 0会重新抓取当前行（如果存在）。

count

提取接下来的*count*行（与FORWARD *count*相同）。

ALL

提取所有剩余行（与FORWARD ALL相同）。

FORWARD

提取下一行（与NEXT相同）。

FORWARD *count*

提取接下来的*count*行。FORWARD 0会重新提取当前行。

FORWARD ALL

提取所有剩余行。

BACKWARD

提取前一行（与PRIOR相同）。

BACKWARD *count*

提取前面的*count*行（反向扫描）。BACKWARD 0会重新提取当前行。

BACKWARD ALL

提取之前的所有行（反向扫描）。

count

*count*是一个可以带符号的整数常量，决定抓取位置或抓取行数。对于FORWARD和BACKWARD，指定负的*count*等同于交换FORWARD和BACKWARD的方向。

cursor_name

一个已打开游标的名称。

输出

成功完成时，FETCH命令会返回如下形式的命令标签：

FETCH *count*

其中*count*是提取到的行数（可能为零）。注意，在psql中，命令标签实际上不会显示，因为psql会改为显示提取到的行。

注解

如果打算使用除FETCH NEXT或带正数计数的 FETCH FORWARD之外的任何FETCH 变体，则应当在声明游标时使用SCROLL选项。对于简单查询，PostgreSQL允许从未使用SCROLL 声明的游标向后提取，但最好不要依赖这种行为。如果游标声明为 NO SCROLL，则不允许向后提取。

ABSOLUTE提取并不比通过相对移动到达目标行更快：无论如何，
底层实现都必须遍历所有中间行。
负值的绝对提取甚至更糟：必须先把查询读到末尾以找到最后一行，然后再从那里反向遍历。
不过，回卷到查询起始处（如FETCH ABSOLUTE 0）是很快的。

DECLARE用于定义游标。使用MOVE可以在不检索数据的情况下改变游标位置。

示例

下面的示例展示了如何使用游标遍历一个表：

```
BEGIN WORK;

-- 建立一个游标:
DECLARE liahona SCROLL CURSOR FOR SELECT * FROM films;

-- 从游标 liahona 中提取前 5 行:
FETCH FORWARD 5 FROM liahona;

 code  |          title          | did | date_prod  |   kind   |  len
-------+-------------------------+-----+------------+----------+-----
--
```

```
 BL101 | The Third Man           | 101 | 1949-12-23 | Drama    | 01:44
 BL102 | The African Queen       | 101 | 1951-08-11 | Romantic | 01:43
 JL201 | Une Femme est une Femme | 102 | 1961-03-12 | Romantic | 01:25
 P_301 | Vertigo                 | 103 | 1958-11-14 | Action   | 02:08
 P_302 | Becket                  | 103 | 1964-02-03 | Drama    | 02:28

-- 提取前一行:
FETCH PRIOR FROM liahona;

 code  |  title  | did | date_prod  |  kind  |  len
-------+---------+-----+------------+--------+-------
 P_301 | Vertigo | 103 | 1958-11-14 | Action | 02:08

-- 关闭游标并结束事务:
CLOSE liahona;
COMMIT WORK;
```

兼容性

SQL 标准只为嵌入式 SQL 定义了FETCH。这里描述的FETCH变体会像 SELECT结果那样返回数据，而不是把数据放入主变量中。除此之外，FETCH与 SQL 标准完全向上兼容。

涉及FORWARD和BACKWARD的 FETCH形式，以及形式FETCH *count*和FETCH ALL（其中隐含了FORWARD）都是 PostgreSQL扩展。

SQL 标准只允许在游标名之前使用FROM；允许使用 IN，或者将两者都完全省略，都是扩展。

另见

CLOSE, DECLARE, MOVE

GRANT

GRANT — define access privileges

大纲

```
GRANT { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES |
 TRIGGER }
    [, ...] | ALL [ PRIVILEGES ] }
    ON { [ TABLE ] table_name [, ...]
         | ALL TABLES IN SCHEMA schema_name [, ...] }
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { SELECT | INSERT | UPDATE | REFERENCES } ( column [, ...] )
    [, ...] | ALL [ PRIVILEGES ] ( column [, ...] ) }
    ON [ TABLE ] table_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { USAGE | SELECT | UPDATE }
    [, ...] | ALL [ PRIVILEGES ] }
    ON { SEQUENCE sequence_name [, ...]
         | ALL SEQUENCES IN SCHEMA schema_name [, ...] }
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { CREATE | CONNECT | TEMPORARY | TEMP } [, ...] | ALL [
 PRIVILEGES ] }
    ON DATABASE database_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { USAGE | ALL [ PRIVILEGES ] }
    ON FOREIGN DATA WRAPPER fdw_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { USAGE | ALL [ PRIVILEGES ] }
    ON FOREIGN SERVER server_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON { FUNCTION function_name ( [ [ argmode ] [ arg_name ] arg_type
 [, ...] ] ) [, ...]
         | ALL FUNCTIONS IN SCHEMA schema_name [, ...] }
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { USAGE | ALL [ PRIVILEGES ] }
    ON LANGUAGE lang_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { SELECT | UPDATE } [, ...] | ALL [ PRIVILEGES ] }
    ON LARGE OBJECT loid [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { { CREATE | USAGE } [, ...] | ALL [ PRIVILEGES ] }
```

```
    ON SCHEMA schema_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT { CREATE | ALL [ PRIVILEGES ] }
    ON TABLESPACE tablespace_name [, ...]
    TO { [ GROUP ] role_name | PUBLIC } [, ...] [ WITH GRANT OPTION ]

GRANT role_name [, ...] TO role_name [, ...] [ WITH ADMIN OPTION ]
```

描述

GRANT命令有两种基本变体：一种用于授予数据库对象（表、列、视图、外部表、序列、数据库、外部数据包装器、外部服务器、函数、过程语言、模式或表空间）上的权限，另一种用于授予角色的成员资格。这两种变体在许多方面很相似，但差别也足够大，因此分别介绍。

在数据库对象上 GRANT

这种 GRANT 命令变体将数据库对象上的特定权限授予一个或多个角色。如果此前已经授予过某些权限，新授予的权限会加到现有权限之上。

还可以选择在一个或多个模式中，对同一类型的所有对象授予权限。目前仅对表、序列和函数支持这一功能（但请注意，ALL TABLES被认为包括视图和外部表）。

关键字 PUBLIC 表示要把权限授予所有角色，包括以后可能创建的角色。PUBLIC 可以视为一个隐式定义的组，并且始终包含所有角色。任何特定角色实际拥有的权限，是直接授予给它的权限、授予给它当前所属任一角色的权限，以及授予给 PUBLIC 的权限之和。

如果指定了 WITH GRANT OPTION，权限接收者随后可以再把该权限授予其他人。没有授予选项时，接收者不能这样做。授予选项不能授予给 PUBLIC。

没有必要向对象拥有者（通常是创建它的用户）授予权限，因为拥有者默认拥有全部权限。（不过，出于安全考虑，拥有者也可以选择撤销自己的某些权限。）

删除对象或以任何方式更改其定义的权利，不被视为一种可授予的权限；它是拥有者固有的，不能被授予或撤销。（不过，可以通过授予或撤销拥有该对象的角色的成员资格，获得类似效果；见下文。）拥有者还隐式拥有该对象上的全部授予选项。

根据对象类型的不同，初始的默认权限可能包括向 PUBLIC 授予某些权限。默认情况下，表、列、模式和表空间没有公共访问权限；数据库有 CONNECT 权限和创建 TEMP 表的权限；函数有 EXECUTE 权限；语言有 USAGE 权限。当然，对象拥有者可以撤销这些权限。（为了最大程度的安全，应在创建对象的同一个事务中发出 REVOKE；这样其他用户就没有可以使用该对象的时间窗口。）此外，这些初始默认权限设置可以用ALTER DEFAULT PRIVILEGES命令更改。

可用权限如下：

SELECT

允许使用SELECT读取指定表、视图或序列的任何列，或列出的特定列。还允许使用COPY TO。在UPDATE或DELETE中引用现有列值时也需要此权限。对于序列，此权限还允许使用 currval函数。对于大对象，此权限允许读取该对象。

INSERT

允许使用INSERT向指定表插入新行。如果列出了特定列，则INSERT 命令只能给这些列赋值（其他列因此会获得默认值）。还允许使用COPY FROM。

UPDATE

允许使用UPDATE更新指定表的任何列，或列出的特定列。（实际上，任何非简单的`UPDATE`命令还需要`SELECT`权限，因为它必须引用表列，以确定要更新的行和/或计算列的新值。）`SELECT ... FOR UPDATE`和`SELECT ... FOR SHARE`除了需要`SELECT`权限外，也要求至少在一列上具有此权限。对于序列，此权限允许使用`nextval`和`setval`函数。对于大对象，此权限允许写入或截断该对象。

DELETE

允许使用DELETE从指定表删除一行。（实际上，任何非简单的`DELETE`命令还需要`SELECT`权限，因为它必须引用表列，以确定要删除的行。）

TRUNCATE

允许在指定表上使用TRUNCATE。

REFERENCES

要创建外键约束，必须在引用列和被引用列上都拥有此权限。此权限可以授予表的所有列，也可以只授予特定列。

TRIGGER

允许在指定表上创建触发器。（参见CREATE TRIGGER语句。）

CREATE

对于数据库，允许在数据库中创建新的模式。

对于模式，允许在模式中创建新对象。要重命名现有对象，必须拥有该对象，并且在包含它的模式上具有此权限。

对于表空间，允许在其中创建表、索引和临时文件，也允许创建以该表空间为默认表空间的数据库。（注意，撤销此权限不会改变现有对象的存放位置。）

CONNECT

允许用户连接到指定数据库。此权限在连接启动时检查（此外还会检查`pg_hba.conf`施加的任何限制）。

TEMPORARY
TEMP

允许在使用指定数据库时创建临时表。

EXECUTE

允许使用指定的函数，以及基于该函数实现的任何操作符。这是唯一适用于函数的权限类型。（此语法也适用于聚合函数。）

USAGE

对于过程语言，允许使用指定语言创建以该语言编写的函数。这是唯一适用于过程语言的权限类型。

对于模式，允许访问指定模式中包含的对象（假定也满足对象自身的权限要求）。本质上，这允许被授权者“查找”模式内的对象。没有此权限，仍然可能看到对象名称，例如通过查询系统表。此外，撤销此权限后，现有后端中可能仍有先前已执行过这种查找的语句，因此这并不是阻止对象访问的完全安全的方法。

对于序列，此权限允许使用`currval`和`nextval`函数。

对于外部数据包装器，此权限使被授权者能够使用该外部数据包装器创建新的服务器。

对于服务器，此权限允许被授予者创建、更改和删除与该服务器关联的自己的用户映射。此外，它还允许被授予者查询该服务器以及关联的用户映射的选项。

ALL PRIVILEGES

一次授予所有可用权限。`PRIVILEGES`关键字在PostgreSQL中是可选的，但在严格 SQL 中是必需的。

The privileges required by other commands are listed on the reference page of the respective command.

角色上的 GRANT

这种`GRANT`命令变体把一个角色的成员资格授予一个或多个其他角色。角色成员资格之所以重要，是因为它会把授予该角色的权限传递给其每个成员。

如果指定了`WITH ADMIN OPTION`，成员就可以将该角色的成员资格继续授予其他人，也可以撤销该角色的成员资格。没有管理选项时，普通用户不能这样做。一个角色不被认为在其自身上持有`WITH ADMIN OPTION`，但是在会话用户与该角色相符的数据库会话中，它可以把自身角色的成员资格授予其他角色，或撤销这种资格。数据库超级用户可以向任何人授予或撤销任何角色的成员资格。具有`CREATEROLE`权限的角色可以授予或撤销任何非超级用户角色的成员资格。

与权限不同，角色成员资格不能授予`PUBLIC`。还要注意，这种形式的命令不允许使用无实际作用的`GROUP`一词。

注解

REVOKE命令用于撤销访问权限。

从 PostgreSQL 8.1 起，用户和组的概念已统一为一种称为角色的单一实体。因此，不再需要使用关键字 `GROUP` 来标识被授权者是用户还是组。`GROUP` 仍可出现在命令中，但它只是一个噪声词。

如果用户对某一列本身，或者对其所在整张表拥有该权限，就可以在该列上执行 `SELECT`、`INSERT` 等操作。在表级授予某项权限后，再在单列上撤销该权限，并不会产生你可能期望的效果：表级授权不会受到列级操作的影响。

当对象的非拥有者试图在该对象上执行 `GRANT` 时，如果该用户在该对象上完全没有任何权限，命令会立即失败。只要有某项权限可用，命令就会继续执行，但只会授予那些该用户持有授予选项的权限。如果未持有任何授予选项，`GRANT ALL PRIVILEGES` 形式会发出警告；而其他形式如果命令中特别列出的任一权限未持有其授予选项，也会发出警告。（原则上，这些说明也适用于对象拥有者；但由于拥有者总是被视为持有全部授予选项，这种情况实际上不会发生。）

需要注意，数据库超级用户可以访问所有对象，而不受对象权限设置的影响。这可类比于 Unix 系统中的 `root` 权限。和 `root` 一样，除非绝对必要，否则不宜以超级用户身份操作。

如果超级用户选择执行 GRANT 或 REVOKE 命令，该命令会像由受影响对象的拥有者发出一样执行。特别是，通过这种命令授予的权限看起来会像是由对象拥有者授予的。
（对于角色成员资格，则看起来像是由被授予成员资格的角色本身授予的。）

GRANT 和 REVOKE 也可以由并非受影响对象拥有者的角色执行，只要该角色是拥有该对象之角色的成员，或者是持有该对象上 WITH GRANT OPTION 权限之角色的成员。在这种情况下，权限会记录为由实际拥有该对象的角色，或者由持有 WITH GRANT OPTION 权限的角色授予。例如，如果表 t1 由角色 g1 拥有，而角色 u1 是它的成员，那么 u1 可以把 t1 上的权限授予给 u2，但这些权限看起来会像是直接由 g1 授予的。角色 g1 的任何其他成员之后都可以撤销这些权限。

如果执行 GRANT 的角色通过多条角色成员资格路径间接持有所需权限，则系统不会指明会被记录为执行该授权的是哪一个上层角色。在这种情况下，最佳做法是使用 SET ROLE 切换成你希望作为其身份执行 GRANT 的那个具体角色。

在表上授予权限，并不会自动把权限扩展到该表使用的任何序列，包括绑定到 SERIAL 列的序列。序列上的权限必须单独设置。

使用psql的 \dp 命令可以获取表和列的现有权限信息。例如：

```
=> \dp mytable
                              Access privileges
 Schema |  Name   | Type  |   Access privileges   | Column access
 privileges
--------+---------+-------+-----------------------+-------------------
-------
 public | mytable | table | miriam=arwdDxt/miriam | col1:
                          : =r/miriam             :   miriam_rw=rw/
miriam
                          : admin=arw/miriam
(1 row)
```

由 \dp 显示的条目解释如下：

```
rolename=xxxx -- privileges granted to a role
        =xxxx -- privileges granted to PUBLIC

            r -- SELECT ("read")
            w -- UPDATE ("write")
            a -- INSERT ("append")
            d -- DELETE
            D -- TRUNCATE
            x -- REFERENCES
            t -- TRIGGER
            X -- EXECUTE
            U -- USAGE
            C -- CREATE
            c -- CONNECT
            T -- TEMPORARY
      arwdDxt -- ALL PRIVILEGES (for tables, varies for other objects)
            * -- grant option for preceding privilege

        /yyyy -- role that granted this privilege
```

上例中的显示结果会出现在用户 miriam 创建表 mytable 并执行以下命令之后：

```
GRANT SELECT ON mytable TO PUBLIC;
```

```
GRANT SELECT, UPDATE, INSERT ON mytable TO admin;
GRANT SELECT (col1), UPDATE (col1) ON mytable TO miriam_rw;
```

对于非表对象，还有其他`\d`命令可以显示其权限。

如果某个对象的“Access privileges”列为空，表示该对象具有默认权限（即其权限列为 null）。默认权限始终包括所有者的全部权限，并且可能根据对象类型包含授予`PUBLIC`的某些权限，如上所述。在对象上首次执行`GRANT`或`REVOKE`时，会先实例化默认权限（例如生成`{miriam=arwdDxt/miriam}`），然后根据指定请求修改它们。同样，“Column access privileges”中也只会显示具有非默认权限的列的条目。（注意：此处的“默认权限”始终指该对象类型的内置默认权限。权限受`ALTER DEFAULT PRIVILEGES`命令影响的对象，始终会显示显式权限条目，其中包含`ALTER`的效果。）

注意，访问权限显示中不会标记所有者隐含的授权选项。只有在显式向某人授予授权选项时，才会出现`*`。

示例

将表 `films` 上的插入权限授予所有用户：

```
GRANT INSERT ON films TO PUBLIC;
```

将视图 `kinds` 上的所有可用权限授予用户 `manuel`：

```
GRANT ALL PRIVILEGES ON kinds TO manuel;
```

请注意，如果上述命令由超级用户或 `kinds` 的拥有者执行，确实会授予所有权限；但如果由其他人执行，则只会授予该执行者持有授予选项的那些权限。

将角色 `admins` 的成员资格授予用户 `joe`：

```
GRANT admins TO joe;
```

兼容性

根据 SQL 标准，`ALL PRIVILEGES` 中的 `PRIVILEGES` 关键字是必需的。SQL 标准也不支持每条命令对多个对象设置权限。

PostgreSQL 允许对象拥有者撤销自己的普通权限：例如，表拥有者可以通过撤销自己的 `INSERT`、`UPDATE`、`DELETE` 和 `TRUNCATE` 权限，使该表对自己变成只读。这在 SQL 标准中是不可能的。原因是 PostgreSQL 把拥有者的权限视为拥有者授予给自己的；因此他们也可以撤销这些权限。在 SQL 标准中，拥有者的权限由一个假定实体“_SYSTEM”授予。由于拥有者并不是“_SYSTEM”，因此不能撤销这些权利。

SQL 标准还为其他种类的对象提供 `USAGE` 权限：字符集、排序规则、翻译。

数据库、表空间、模式和语言上的权限都是 PostgreSQL 扩展。

参见

REVOKE, ALTER DEFAULT PRIVILEGES

INSERT

INSERT — create new rows in a table

大纲

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
INSERT INTO table [ ( column [, ...] ) ]
    { DEFAULT VALUES | VALUES ( { expression | DEFAULT } [, ...] ) [,
 ...] | query }
    [ RETURNING * | output_expression [ [ AS ] output_name ] [, ...] ]
```

描述

INSERT将新行插入表中。可以插入由值表达式指定的一行或多行，
也可以插入由查询产生的零行或多行。

目标列名可以按任意顺序列出。如果根本未给出列名列表，
则默认使用按声明顺序排列的全部列；或者如果VALUES子句或 *query*只提供了*N*列，则默认使用按声明顺序排列的前*N*个列名。VALUES子句或*query* 提供的值，会按从左到右的顺序与显式或隐式列列表对应起来。

任何未出现在显式或隐式列列表中的列都会填入默认值；如果没有声明默认值，则填入空值。

如果任何列的表达式数据类型不正确，将尝试自动进行类型转换。

可选的RETURNING子句使INSERT 基于每个实际插入的行计算并返回一个或多个值。这主要用于获取由默认值提供的值，例如 serial 序列号。不过，也允许使用任何引用该表列的表达式。RETURNING列表的语法与 SELECT的输出列表相同。

要向表中插入行，必须具有该表上的INSERT权限。如果指定了列列表，则只需要对所列的列具有INSERT权限。使用RETURNING子句要求具有SELECT权限，涵盖RETURNING中提到的所有列。如果使用*query*子句从查询中插入行，当然需要在查询使用的任何表或列上具有SELECT权限。

参数

with_query

WITH子句允许指定一个或多个子查询，这些子查询可以在INSERT查询中按名称引用。
详见第 7.8 节和SELECT。

query（SELECT语句）本身也可以包含 WITH子句。在这种情况下，*query*中可以引用两组*with_query*，但由于第二组嵌套得更近，它具有更高的优先级。

table

现有表的名称（可选地使用模式限定）。

column

table 中某一列的名称。如有需要，列名可以用子字段名或数组下标限定。（只向组合列的某些字段插入时，其余字段保持为空。）

`DEFAULT VALUES`

所有列都会填入各自的默认值。

expression

要赋予对应 *column* 的表达式或值。

`DEFAULT`

对应的 *column* 将用它的默认值填充。

query

提供要插入行的查询（`SELECT`语句）。其语法说明请参见SELECT。

output_expression

`INSERT`命令在插入每行后要计算并返回的表达式。表达式可以使用*table_name*所指定表中的任何列名。写`*`可返回所插入行的所有列。

output_name

用于返回列的名称。

输出

成功完成时，`INSERT` 命令返回以下形式的命令标签：

`INSERT` *oid count*

其中，*count* 是插入的行数。如果 *count* 恰好为一，且目标表有 OID，则 *oid* 表示分配给插入行的 OID。该单行必须是插入的，而不是更新的。否则 *oid* 为零。

如果`INSERT`命令包含`RETURNING` 子句，则结果将类似于一个`SELECT`语句，其中包含`RETURNING`列表定义的列和值，它们是基于该命令插入或更新的行计算出来的。

示例

向表`films`插入一行：

```
INSERT INTO films VALUES
    ('UA502', 'Bananas', 105, '1971-07-13', 'Comedy', '82 minutes');
```

在这个示例中，省略了`len`列，因此它将取默认值：

```
INSERT INTO films (code, title, did, date_prod, kind)
    VALUES ('T_601', 'Yojimbo', 106, '1961-06-16', 'Drama');
```

这个示例对日期列使用`DEFAULT`子句，而不是显式指定值：

```
INSERT INTO films VALUES
    ('UA502', 'Bananas', 105, DEFAULT, 'Comedy', '82 minutes');
INSERT INTO films (code, title, did, date_prod, kind)
    VALUES ('T_601', 'Yojimbo', 106, DEFAULT, 'Drama');
```

要插入一行，其所有列都由默认值构成：

```
INSERT INTO films DEFAULT VALUES;
```

使用多行VALUES语法插入多行：

```
INSERT INTO films (code, title, did, date_prod, kind) VALUES
    ('B6717', 'Tampopo', 110, '1985-02-10', 'Comedy'),
    ('HG120', 'The Dinner Game', 140, DEFAULT, 'Comedy');
```

这个示例从与films列布局相同的 tmp_films表中选出一些行并插入到 films表：

```
INSERT INTO films SELECT * FROM tmp_films WHERE date_prod < '2004-05
-07';
```

这个示例向数组列插入值：

```
-- Create an empty 3x3 gameboard for noughts-and-crosses
INSERT INTO tictactoe (game, board[1:3][1:3])
    VALUES (1, '{{" "," "," "},{" "," "," "},{" "," "," "}}');
-- The subscripts in the above example aren't really needed
INSERT INTO tictactoe (game, board)
    VALUES (2, '{{X," "," "},{" ",O," "},{" ",X," "}}');
```

向表distributors插入一行，并返回由 DEFAULT子句生成的序号：

```
INSERT INTO distributors (did, dname) VALUES (DEFAULT, 'XYZ Widgets')
   RETURNING did;
```

为负责 Acme Corporation 账户的销售人员增加销量计数，并把整个更新后的行连同当前时间记录到日志表中：

```
WITH upd AS (
  UPDATE employees SET sales_count = sales_count + 1 WHERE id =
    (SELECT sales_person FROM accounts WHERE name = 'Acme
 Corporation')
    RETURNING *
)
INSERT INTO employees_log SELECT *, current_timestamp FROM upd;
```

兼容性

INSERT符合 SQL 标准，但 RETURNING子句是 PostgreSQL扩展，在 INSERT中使用WITH的能力也是扩展。此外，标准不允许省略列名列表却又不是所有列都由VALUES 子句或*query*填充的情况。

*query*子句可能存在的限制见SELECT。

LISTEN

LISTEN — listen for a notification

大纲

```
LISTEN channel
```

描述

LISTEN将当前会话注册为名为*channel*的通知通道的监听者。
如果当前会话已经在该通知通道上注册为监听者，则不会执行任何操作。

每当命令NOTIFY *channel*被执行时，无论是由当前会话还是由另一个连接到同一数据库的会话执行，所有当前正在监听该通知通道的会话都会收到通知，
而每个会话随后都会通知其所连接的客户端应用。

会话可以使用UNLISTEN命令取消在给定通知通道上的监听注册。会话结束时，
其监听注册会被自动清除。

客户端应用必须采用何种方法来检测通知事件，取决于它所使用的 PostgreSQL应用程序编程接口。使用 libpq库时，应用程序将LISTEN 作为普通 SQL 命令发出，然后必须定期调用 PQnotifies函数，以判断是否已经收到任何通知事件。其他接口，例如libpgtcl，则提供了处理通知事件的更高层方法；事实上，在libpgtcl中，应用程序员甚至不应直接发出LISTEN 或UNLISTEN。详见所使用接口的相关文档。

NOTIFY中对LISTEN和NOTIFY的使用有更详细的讨论。

参数

channel

通知通道的名称（任意标识符）。

注解

LISTEN会在事务提交时生效。如果在随后被回滚的事务中执行了LISTEN或 UNLISTEN，则被监听的通知通道集合不会发生变化。

执行过LISTEN的事务不能为两阶段提交做准备。

示例

在psql中配置并执行一组 listen/notify 操作：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
```

兼容性

SQL 标准中没有LISTEN语句。

另见

NOTIFY, UNLISTEN

LOAD

LOAD — load a shared library file

大纲

`LOAD 'filename'`

描述

该命令将一个共享库文件载入PostgreSQL服务器的地址空间。如果该文件已经载入过，则此命令不执行任何操作。包含 C 函数的共享库文件会在调用其中某个函数时自动载入。因此，显式执行`LOAD`通常只在要载入的是通过“钩子”修改服务器行为、而不是提供一组函数的库时才有需要。

文件名的指定方式与CREATE FUNCTION中共享库名的指定方式相同；特别地，可以依赖搜索路径以及系统标准共享库文件名扩展名的自动添加。关于此主题的更多信息，见第 35.9 节。

非超级用户只能将`LOAD`用于位于`$libdir/plugins/`的库文件 – 指定的*filename*必须以该字符串原样开头。（数据库管理员有责任确保只在该目录中安装“安全的”库。）

兼容性

`LOAD`是PostgreSQL的一种扩展。

另见

CREATE FUNCTION

LOCK

LOCK — lock a table

大纲

```
LOCK [ TABLE ] [ ONLY ] name [ * ] [, ...] [ IN lockmode MODE ] [
 NOWAIT ]

where lockmode is one of:

    ACCESS SHARE | ROW SHARE | ROW EXCLUSIVE | SHARE UPDATE EXCLUSIVE
    | SHARE | SHARE ROW EXCLUSIVE | EXCLUSIVE | ACCESS EXCLUSIVE
```

描述

LOCK TABLE获取一个表级锁；如有必要，会等待任何冲突锁被释放。如果指定了NOWAIT，LOCK TABLE就不会等待获取所需的锁：如果无法立即获得，命令将被中止并报错。锁一旦获得，就会一直持有到当前事务结束。（没有UNLOCK TABLE命令；锁总是在事务结束时释放。）

为引用表的命令自动获取锁时，PostgreSQL始终使用限制最少的可用锁模式。LOCK TABLE用于可能需要更严格锁定的情况。例如，假设一个应用在READ COMMITTED隔离级别运行事务，并且需要确保表中的数据在事务期间保持稳定。为此，可以在查询之前对表获取SHARE锁。这会阻止并发数据更改，确保后续对表的读取能看到已提交数据的稳定视图，因为SHARE锁模式与写入者获取的ROW EXCLUSIVE锁冲突，而你的LOCK TABLE *name* IN SHARE MODE语句会等待，直到所有并发持有ROW EXCLUSIVE模式锁的事务提交或回滚。因此，一旦获得该锁，就不存在尚未提交的写入；而且在释放该锁之前，也不会开始任何写入。

若要在REPEATABLE READ或SERIALIZABLE 隔离级别的事务中达到类似效果，你必须在执行任何SELECT 或数据修改语句之前执行LOCK TABLE语句。REPEATABLE READ或SERIALIZABLE事务的数据视图，会在其第一条SELECT或数据修改语句开始时冻结。在事务稍后再执行LOCK TABLE仍然可以阻止并发写入 – 但它不能保证该事务读取到的是最新已提交的值。

如果这类事务还要修改表中的数据，那么它应使用SHARE ROW EXCLUSIVE锁模式，而不是SHARE模式。这样可以确保同一时间只有一个这类事务在运行。否则就可能发生死锁：两个事务都可能先获得SHARE模式，然后都无法再获得实际执行更新所需的ROW EXCLUSIVE模式。（注意，事务自己的锁永远不会互相冲突，因此事务在持有SHARE模式时仍可获得 ROW EXCLUSIVE模式，但前提是没有其他人持有SHARE模式。）为避免死锁，要确保所有事务都按相同顺序对相同对象获取锁；如果同一对象需要多种锁模式，则事务应始终先获取限制最严格的模式。

关于锁模式和锁策略的更多信息，请参见第 13.3 节。

参数

name

要锁定的现有表的名称（可选模式限定）。如果在表名前指定了 ONLY，则只有该表会被锁定。如果未指定ONLY，则该表及其所有后代表（如果有）都会被锁定。也可以在表名后指定 *，以显式表明包含后代表。

命令LOCK TABLE a, b;等效于 LOCK TABLE a; LOCK TABLE b;。这些表会按 LOCK TABLE命令中指定的顺序逐个锁定。

lockmode

锁模式指定该锁会与哪些锁冲突。锁模式见第 13.3 节。

如果未指定锁模式，则使用限制最严格的ACCESS EXCLUSIVE模式。

NOWAIT

指定LOCK TABLE不等待任何冲突锁被释放：如果指定的锁无法在不等待的情况下立即获得，事务就会中止。

注解

LOCK TABLE ... IN ACCESS SHARE MODE要求对目标表具有SELECT权限。所有其他形式的LOCK 都要求具有表级UPDATE、DELETE或TRUNCATE权限。

LOCK TABLE在事务块外毫无用处：锁只会一直持有到该语句结束。因此，如果在事务块外使用LOCK，PostgreSQL会报告错误。请使用BEGIN和COMMIT（或ROLLBACK）来定义事务块。

LOCK TABLE只处理表级锁，因此名称中带有ROW的模式其实都不准确。这些模式名称通常应理解为：用户打算在被锁定的表中获取行级锁。此外，ROW EXCLUSIVE模式本身也是一种可共享的表锁。请记住，就LOCK TABLE而言，所有锁模式的语义完全相同，差别只在于哪些模式彼此冲突。关于如何获取真正的行级锁，请参阅第 13.3.2 节以及FOR UPDATE/FOR SHARE Clause一节（位于SELECT参考文档中）。

示例

在准备向外键表执行插入时，在主键表上获取一个SHARE锁：

```
BEGIN WORK;
LOCK TABLE films IN SHARE MODE;
SELECT id FROM films
    WHERE name = 'Star Wars: Episode I - The Phantom Menace';
-- 如果未返回记录则执行 ROLLBACK
INSERT INTO films_user_comments VALUES
    (_id_, 'GREAT! I was waiting for it for so long!');
COMMIT WORK;
```

在准备执行删除操作时，在主键表上获取一个SHARE ROW EXCLUSIVE锁：

```
BEGIN WORK;
LOCK TABLE films IN SHARE ROW EXCLUSIVE MODE;
DELETE FROM films_user_comments WHERE id IN
    (SELECT id FROM films WHERE rating < 5);
DELETE FROM films WHERE rating < 5;
COMMIT WORK;
```

兼容性

SQL 标准中没有LOCK TABLE，而是使用SET TRANSACTION 来指定事务的并发级别。PostgreSQL也支持这一点；详见SET TRANSACTION。

除ACCESS SHARE、ACCESS EXCLUSIVE和 SHARE UPDATE EXCLUSIVE锁模式外，PostgreSQL的锁模式和LOCK TABLE语法与Oracle中的对应语法兼容。

MOVE

MOVE — position a cursor

大纲

```
MOVE [ direction [ FROM | IN ] ] cursor_name

where direction can be empty or one of:

    NEXT
    PRIOR
    FIRST
    LAST
    ABSOLUTE count
    RELATIVE count
    count
    ALL
    FORWARD
    FORWARD count
    FORWARD ALL
    BACKWARD
    BACKWARD count
    BACKWARD ALL
```

描述

`MOVE`在不检索任何数据的情况下重新定位游标。`MOVE`与`FETCH`命令的工作方式完全相同，只是它只定位游标而不返回行。

`MOVE`命令的参数与`FETCH`命令完全相同；有关语法和用法的细节，请参阅FETCH。

输出

成功完成时，`MOVE`命令会返回如下形式的命令标签：

```
MOVE count
```

其中*count*是用相同参数执行 `FETCH`命令时本应返回的行数（可能为零）。

示例

```
BEGIN WORK;
DECLARE liahona CURSOR FOR SELECT * FROM films;

-- Skip the first 5 rows:
MOVE FORWARD 5 IN liahona;
MOVE 5

-- Fetch the 6th row from the cursor liahona:
FETCH 1 FROM liahona;
```

```
 code  | title  | did | date_prod  |  kind  |  len
-------+--------+-----+------------+--------+-------
 P_303 | 48 Hrs | 103 | 1982-10-22 | Action | 01:37
(1 row)

-- Close the cursor liahona and end the transaction:
CLOSE liahona;
COMMIT WORK;
```

兼容性

SQL 标准中没有`MOVE`语句。

另见

CLOSE, DECLARE, FETCH

NOTIFY

NOTIFY — generate a notification

大纲

NOTIFY *channel* [, *payload*]

描述

NOTIFY 命令向当前数据库中曾为指定通道名执行过 LISTEN *channel* 的每个客户端应用发送一个通知事件，以及一个可选的“载荷”字符串。

NOTIFY为访问同一个PostgreSQL数据库的一组进程提供了一种简单的进程间通信机制。通知中可以附带载荷字符串；如果需要传递结构化数据，也可以借助数据库中的表，将附加数据从通知发送者传递给一个或多个监听者，从而构建用于传递结构化数据的更高层机制。

传递给客户端的通知事件信息包括：通知通道名、发出通知的会话对应的服务器进程PID，以及载荷字符串。如果未指定载荷字符串，则该字符串为空串。

在某个数据库中使用哪些通道名以及各自含义，由数据库设计者自行决定。常见的做法是让通道名与数据库中的某个表同名，而通知事件基本上意味着：“我改动了这张表，去看看有什么新变化”。不过，NOTIFY和LISTEN命令并不会强制这种关联。例如，数据库设计者可以使用多个不同的通道名，来标识同一张表上的不同类型变更；或者也可以利用载荷字符串区分不同情况。

当NOTIFY用于表明某张特定表发生变化时，一种很有用的编程技巧是把NOTIFY放进由表更新触发的规则中。这样一来，每当表发生变化时就会自动发出通知，应用程序员也不容易忘记这么做。

NOTIFY会以几种重要方式与 SQL 事务交互。首先，如果在事务内部执行NOTIFY，那么只有在事务提交之后，通知事件才会被递送；如果事务被中止，则其中所有命令都不会生效，NOTIFY也不例外。这种行为是合理的，但如果你期望通知立即送达，可能会感到意外。其次，如果某个正在监听的会话在事务内部收到了通知信号，那么在该事务结束（提交或中止）之前，通知事件都不会递送给它所连接的客户端。原因同样在于：如果通知在事务内部就已递送，而该事务后来又被中止，我们会希望通知也能被撤销，但服务器一旦把通知发送给客户端，就无法再把它“收回”。因此，通知事件只会在事务之间递送。由此得出的结论是，使用NOTIFY做实时信号的应用，应尽量让事务保持短小。

如果在同一个事务中，针对同一通道名多次发送完全相同载荷字符串的通知，那么数据库服务器可以决定只递送一个通知。另一方面，载荷字符串不同的通知始终会作为不同通知递送。类似地，来自不同事务的通知也绝不会被折叠为一个通知。除了会丢弃重复通知中较后的那些实例之外，NOTIFY还保证来自同一事务的通知会按发送顺序递送；来自不同事务的消息，则会按事务提交顺序递送。

执行NOTIFY的客户端自己同时也在监听同一通知通道，这是很常见的情形。在这种情况下，它会像其他监听会话一样收到一个通知事件。根据应用逻辑，这可能导致无用功，例如再次读取自己刚刚写入更新的数据库表。要避免这种额外工作，可以检查通知事件消息中提供的发出通知的服务器进程PID，是否与当前会话自身的PID（可通过libpq获得）相同。如果二者相同，就说明这是当前会话自己发出的通知回送给自己，可以直接忽略。

参数

channel

要发信号的通知通道名称（任意标识符）。

payload

随通知一起传递的“载荷”字符串。它必须指定为简单的字符串字面值。在默认配置下，它必须短于 8000 字节。（如果需要传递二进制数据或大量信息，最好把它们放入数据库表中，并发送该记录的键。）

注解

系统中有一个队列，用来保存那些已经发送但尚未被所有监听会话处理的通知。
如果这个队列被塞满，那么调用NOTIFY的事务会在提交时失败。该队列非常大（标准安装中为8GB），几乎足以满足所有用例。不过，如果某个会话执行了LISTEN
后又长时间停留在一个事务里，就无法进行清理。一旦队列占用达到一半，
你就会在日志文件中看到警告，指出究竟是哪个会话阻碍了清理。在这种情况下，
应确保该会话结束其当前事务，以便清理能够继续进行。

执行过NOTIFY的事务不能为两阶段提交做准备。

pg_notify

要发送通知，也可以使用函数pg_notify(text, text)。该函数的第一个参数是通道名，第二个参数是载荷。如果需要处理非常量的通道名和载荷，它会比NOTIFY命令更容易使用。

示例

在psql中配置并执行一组 listen/notify 操作：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
NOTIFY virtual, 'This is the payload';
Asynchronous notification "virtual" with payload "This is the payload"
 received from server process with PID 8448.

LISTEN foo;
SELECT pg_notify('fo' || 'o', 'pay' || 'load');
Asynchronous notification "foo" with payload "payload" received from
 server process with PID 14728.
```

兼容性

SQL 标准中没有NOTIFY语句。

另见

LISTEN, UNLISTEN

PREPARE

PREPARE — prepare a statement for execution

大纲

```
PREPARE name [ ( data_type [, ...] ) ] AS statement
```

描述

`PREPARE` 创建一个预备语句。预备语句是一种可用于优化性能的服务器端对象。执行 `PREPARE` 语句时，指定的语句会被解析、重写和规划。随后对预备语句发出 `EXECUTE` 命令时，只需执行该预备语句即可。这样，解析、重写和规划阶段只执行一次，而不是每次执行语句时都进行。

预备语句可以带参数，也就是在执行时会代入语句中的值。创建预备语句时，
可以按位置引用参数，例如 `$1`、`$2` 等。也可以选择指定相应的参数数据类型列表。如果某个参数的数据类型未指定，或被声明为 `unknown`，则其类型会从该参数第一次被引用时所在的上下文中推断出来（如果可能）。执行该语句时，需要在 `EXECUTE` 语句中为这些参数提供实际值。有关详情，参见EXECUTE。

预备语句只在当前数据库会话期间存在。会话结束时，预备语句就会被遗忘，
因此再次使用前必须重新创建。这也意味着单个预备语句不能由多个并发的数据库客户端共用；
不过，每个客户端都可以创建自己要使用的预备语句。预备语句也可以用
DEALLOCATE命令手工释放。

当单个会话要执行大量相似语句时，预备语句能带来最大的性能优势。
如果语句在规划或重写时比较复杂，例如查询涉及许多表的连接，或者需要应用多条规则，
那么这种性能差异会特别明显。如果语句相对容易规划和重写，但执行本身代价较高，
则预备语句带来的性能优势就不那么明显。

参数

name

为这个特定预备语句指定的任意名称。它在单个会话内必须唯一，
随后用于执行或释放先前已准备的语句。

data_type

预备语句中某个参数的数据类型。如果某个参数的数据类型未指定，或被指定为 `unknown`，
则其类型会从该参数第一次被引用时所在的上下文中推断出来。
要在预备语句本身中引用参数，可使用 `$1`、`$2` 等。

statement

任何`SELECT`、`INSERT`、`UPDATE`、`DELETE`或`VALUES`语句。

注解

在某些情况下，
为预备语句生成的查询计划可能不如语句被正常提交并执行时所选出的查询计划。This is because when the statement is planned and the planner attempts to determine the

optimal query plan, the actual values of any parameters specified in the statement are unavailable. PostgreSQL collects statistics on the distribution of data in the table, 并可以利用语句中的常量值推测执行该语句可能得到的结果. Since this data is unavailable when planning prepared statements with parameters, the chosen plan might be suboptimal. To examine the query plan PostgreSQL has chosen for a prepared statement, use EXPLAIN.

有关查询规划以及 PostgreSQL 为此收集的统计信息的更多内容，请参见ANALYZE文档。

可以通过查询 `pg_prepared_statements` 系统视图查看一个会话的所有可用预备语句。

示例

为一个`INSERT` 语句创建预备语句，然后执行它：

```
PREPARE fooplan (int, text, bool, numeric) AS
    INSERT INTO foo VALUES($1, $2, $3, $4);
EXECUTE fooplan(1, 'Hunter Valley', 't', 200.00);
```

为一个 `SELECT` 语句创建预备语句，然后执行它：

```
PREPARE usrrptplan (int) AS
    SELECT * FROM users u, logs l WHERE u.usrid=$1 AND u.usrid=l.usrid
    AND l.date = $2;
EXECUTE usrrptplan(1, current_date);
```

注意，第二个参数的数据类型未指定，因此会从 `$2` 所在的使用上下文中推断出来。

兼容性

SQL 标准包含`PREPARE`语句，但它只用于嵌入式 SQL。这里的`PREPARE`语句在语法上也略有不同。

另见

DEALLOCATE, EXECUTE

PREPARE TRANSACTION

PREPARE TRANSACTION — prepare the current transaction for two-phase commit

大纲

PREPARE TRANSACTION *transaction_id*

描述

PREPARE TRANSACTION为两阶段提交准备当前事务。执行此命令后，该事务将不再与当前会话相关联；相反，它的状态会被完整地存储到磁盘上，因此即使在请求提交之前数据库发生崩溃，它也极有可能成功提交。

事务进入预备状态后，稍后可以分别使用COMMIT PREPARED或ROLLBACK PREPARED来提交或回滚。发出这些命令的不必是执行原始事务的那个会话，任何会话都可以这样做。

从发出该命令的会话来看，PREPARE TRANSACTION颇似ROLLBACK：执行之后，将不再有活动的当前事务，并且该预备事务的效果也不再可见。（如果该事务随后被提交，这些效果会再次可见。）

如果PREPARE TRANSACTION因任何原因失败，它就等同于执行了一次ROLLBACK：当前事务会被取消。

参数

transaction_id

一个任意标识符，后续可通过它在COMMIT PREPARED或ROLLBACK PREPARED中标识该事务。该标识符必须写成字符串字面值，长度必须小于 200 字节，并且不能与当前任何已处于预备状态的事务标识符相同。

注解

PREPARE TRANSACTION并非供应用程序或交互式会话使用。它的目的是让外部事务管理器能够跨多个数据库或其他事务性资源执行原子性的全局事务。除非是在编写事务管理器，否则通常不应使用PREPARE TRANSACTION。

该命令必须在事务块内部使用。请用BEGIN启动事务块。

目前不允许 PREPARE 一个执行过任何涉及临时表操作、创建过 WITH HOLD 游标，或执行过 LISTEN 或 UNLISTEN 的事务。这些特性与当前会话绑定得太紧，无法在要预备的事务中使用.

如果该事务曾用SET修改过任何运行时参数（且未使用LOCAL选项），那么这些效果在执行PREPARE TRANSACTION之后仍会保留，并且不会受到后续任何COMMIT PREPARED或ROLLBACK PREPARED的影响。因此，仅就这一点而言，PREPARE TRANSACTION的行为更像COMMIT而不是ROLLBACK。

当前所有处于预备状态的事务都列在pg_prepared_xacts系统视图中。

> **小心**
>
> 让事务长时间停留在预备状态并不明智。这会妨碍VACUUM回收存储空间，在极端情况下甚至可能导致数据库为防止事务 ID 回卷而关闭（见第 23.1.4 节）。还要记住，该事务会继续持有它原本持有的所有锁。此功能的预期用法是：一旦外部事务管理器确认其他数据库也已准备好提交，就尽快提交或回滚该预备事务。
>
> 如果没有设置外部事务管理器来跟踪预备事务并确保它们能被及时结束，最好将max_prepared_transactions设为零，从而禁用预备事务功能。这样可以防止意外创建预备事务，而这些事务随后可能被遗忘并最终引发问题。

示例

为两阶段提交准备当前事务，并使用foobar作为事务标识符：

```
PREPARE TRANSACTION 'foobar';
```

另见

COMMIT PREPARED, ROLLBACK PREPARED

REASSIGN OWNED

REASSIGN OWNED — change the ownership of database objects owned by a database role

大纲

`REASSIGN OWNED BY old_role [, ...] TO new_role`

描述

`REASSIGN OWNED`指示系统将任何 *old_roles*所拥有的数据库对象的所有权更改为 *new_role*。

参数

old_role

角色的名称。该角色在当前数据库中拥有的全部对象的所有权将被重新赋予 *new_role*。

new_role

将成为受影响对象新拥有者的角色名称。

注解

`REASSIGN OWNED` 常用于为移除一个或多个角色做准备。由于 `REASSIGN OWNED` 只影响当前数据库中的对象，通常需要在每个包含待移除角色所拥有对象的数据库中执行此命令。

`REASSIGN OWNED` 要求对源角色和目标角色都具有权限。

DROP OWNED命令提供了另一种选择，它会直接删除一个或多个角色所拥有的全部数据库对象。

`REASSIGN OWNED`命令不会影响授予 *old_roles*的、在非其所拥有的对象上的任何权限。同样，它也不会影响使用如需撤销这类权限，请使用`DROP OWNED`。

更多讨论请见第 20.4 节。

`REASSIGN OWNED` 命令不影响该角色所拥有的任何数据库的所有权。要重新赋予那种所有权，请使用ALTER DATABASE。

兼容性

`REASSIGN OWNED`命令是 PostgreSQL扩展。

另见

DROP OWNED, DROP ROLE, ALTER DATABASE

REINDEX

REINDEX — rebuild indexes

大纲

```
REINDEX { INDEX | TABLE | DATABASE | SYSTEM } name [ FORCE ]
```

描述

REINDEX使用索引所属表中存储的数据重建索引，并替换索引的旧副本。以下几种场景适合使用REINDEX：

- 一个索引已经损坏，不再包含有效数据。尽管理论上这不该发生，但在实践中索引可能因软件缺陷或硬件故障而损坏。REINDEX提供了一种恢复方法。
- 一个索引已经“膨胀”，也就是说其中包含许多空页或几乎为空的页。在 PostgreSQL 中，B-树索引在某些不常见的访问模式下可能出现这种情况。REINDEX可通过写入一个不含死页的新版本索引来减少索引的空间消耗。详见第 23.2 节。
- 你修改了某个索引的存储参数（例如 fillfactor），并希望确保该更改已经完全生效。
- 使用 CONCURRENTLY 选项构建索引失败后，会留下一个“无效”索引。这类索引没有用处，但用 REINDEX 重建它们可能很方便。注意，REINDEX 不会进行并发构建。要在不干扰生产运行的情况下构建索引，应删除该索引并重新执行 CREATE INDEX CONCURRENTLY 命令。

参数

INDEX

重新创建指定的索引。

TABLE

重新创建指定表的所有索引。如果该表有一个辅助“TOAST”表，也会对其重新索引。

DATABASE

重新创建当前数据库中的所有索引。共享系统目录上的索引也会被处理。这种形式的REINDEX不能在事务块内执行。

SYSTEM

重新创建当前数据库内系统目录上的所有索引。共享系统目录上的索引也包含在内。用户表上的索引不会被处理。这种形式的 REINDEX不能在事务块内执行。

name

要重新索引的特定索引、表或数据库的名称。索引名和表名可以带模式限定。目前，REINDEX DATABASE和 REINDEX SYSTEM只能对当前数据库重新索引，所以其参数必须与当前数据库名匹配。

FORCE

在每个索引被重建时打印进度报告。

注解

如果怀疑某个用户表上的索引已经损坏，可以使用 REINDEX INDEX或REINDEX TABLE 直接重建该索引，或者重建该表上的所有索引。

如果需要从系统表上的索引损坏中恢复，情况就更复杂了。在这种情况下，重要的是系统本身没有使用任何可疑索引。（事实上，在这种场景下，你可能会发现服务器进程在启动时立即崩溃，因为它依赖损坏的索引。）要安全恢复，必须用-P选项启动服务器，该选项会阻止服务器在查找系统目录时使用索引。

一种做法是关闭服务器，并在命令行中包含-P选项来启动单用户 PostgreSQL 服务器。然后可以根据希望重建的范围，执行REINDEX DATABASE、REINDEX SYSTEM、REINDEX TABLE 或REINDEX INDEX。如果拿不准，就使用REINDEX SYSTEM来重建该数据库中的所有系统索引。然后退出单用户服务器会话并重新启动常规服务器。关于如何与单用户服务器接口交互的更多信息，参见postgres参考页。

另一种方法是启动一个常规服务器会话，并在其命令行选项中包含 -P。具体做法因客户端而异，但对于所有基于 libpq的客户端，都可以在启动客户端之前将环境变量PGOPTIONS设置为-P。注意，尽管这种方法不需要阻止其他客户端，但在修复完成之前，阻止其他用户连接到受损数据库可能仍然更稳妥。

REINDEX类似于删除并重新创建索引，因为索引内容都是从头重建的。不过，两者在锁方面的考量相当不同。REINDEX会阻止该索引所属表上的写入，但不阻止读取。它还会对正在处理的特定索引获取独占锁，从而阻塞试图使用该索引的读取。相比之下，DROP INDEX会短暂地对父表获取独占锁，同时阻塞写入和读取。随后的 CREATE INDEX会阻止写入但不阻止读取；由于索引不存在，读取不会尝试使用它，因此不会发生阻塞，但读取可能被迫使用代价高昂的顺序扫描。

对单个索引或表重新索引，要求用户是该索引或表的拥有者。对数据库重新索引，则要求用户是该数据库的拥有者（因此，拥有者可以重建其他用户所拥有表的索引）。当然，超级用户始终可以重新索引任何对象。

在 PostgreSQL 8.1 之前，REINDEX DATABASE 只处理系统索引，而不是像其名字所暗示的那样处理所有索引。This has been changed to reduce the surprise factor. The old behavior is available as REINDEX SYSTEM.

重建表my_table上的所有索引：REINDEX TABLE my_table;

示例

重建单个索引：

```
REINDEX INDEX my_index;
```

重建表my_table上的所有索引：

```
REINDEX TABLE my_table;
```

在不假定系统索引已经有效的情况下，重建某个数据库中的所有索引：

```
$ export PGOPTIONS="-P"
$ psql broken_db
...
broken_db=> REINDEX DATABASE broken_db;
```

```
broken_db=> \q
```

兼容性

在 SQL 标准中没有`REINDEX`命令。

RELEASE SAVEPOINT

RELEASE SAVEPOINT — destroy a previously defined savepoint

大纲

```
RELEASE [ SAVEPOINT ] savepoint_name
```

描述

`RELEASE SAVEPOINT` 销毁先前在当前事务中定义的保存点。

销毁保存点后，就不能再将它用作回滚点，但除此之外没有其他用户可见的行为。它不会撤销保存点建立之后执行的命令的效果。（要这样做，请参见ROLLBACK TO SAVEPOINT。）在不再需要保存点时将其销毁，可以让系统在事务结束之前回收一些资源。

`RELEASE SAVEPOINT` 还会销毁在指定保存点建立之后建立的所有保存点。

参数

savepoint_name

要销毁的保存点名称。

注解

指定一个此前未定义的保存点名称是一种错误。

当事务处于已中止状态时，不能释放保存点。

如果有多个保存点使用相同名称，则只会释放最近定义的那个。

示例

要建立保存点，并在稍后销毁它：

```
BEGIN;
    INSERT INTO table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

上述事务将同时插入 3 和 4。

兼容性

该命令符合SQL标准。标准规定关键字`SAVEPOINT`是强制的，但PostgreSQL允许省略它。

另见

BEGIN, COMMIT, ROLLBACK, ROLLBACK TO SAVEPOINT, SAVEPOINT

RESET

RESET — restore the value of a run-time parameter to the default value

大纲

```
RESET configuration_parameter
RESET ALL
```

描述

`RESET`将运行时参数恢复为其默认值。`RESET`是

```
SET configuration_parameter TO DEFAULT
```

的另一种拼写。详见SET。

默认值指的是：在当前会话中，如果从未对该参数执行过`SET`，它本应具有的值。
该值的实际来源可能是编译时内置的默认值、配置文件、命令行选项，
或者针对特定数据库或特定用户的默认设置。这与将其定义为“该参数在会话开始时的值”
略有不同，因为如果该值来自配置文件，那么它会被重置为配置文件当前指定的值。
详见第 18 章。

`RESET`的事务行为和`SET`相同：它的效果会在事务回滚时被撤销。

参数

configuration_parameter

可设置的运行时参数名称。可用参数见第 18 章以及SET参考页。

`ALL`

将所有可设置的运行时参数重置为默认值。

示例

将`timezone`配置参数设置为其默认值：

```
RESET timezone;
```

兼容性

`RESET`是一种 PostgreSQL扩展。

另见

SET, SHOW

REVOKE

REVOKE — remove access privileges

大纲

```
REVOKE [ GRANT OPTION FOR ]
    { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES |
 TRIGGER }
    [, ...] | ALL [ PRIVILEGES ] }
    ON { [ TABLE ] table_name [, ...]
         | ALL TABLES IN SCHEMA schema_name [, ...] }
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { SELECT | INSERT | UPDATE | REFERENCES } ( column [, ...] )
    [, ...] | ALL [ PRIVILEGES ] ( column [, ...] ) }
    ON [ TABLE ] table_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { USAGE | SELECT | UPDATE }
    [, ...] | ALL [ PRIVILEGES ] }
    ON { SEQUENCE sequence_name [, ...]
         | ALL SEQUENCES IN SCHEMA schema_name [, ...] }
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { CREATE | CONNECT | TEMPORARY | TEMP } [, ...] | ALL [
 PRIVILEGES ] }
    ON DATABASE database_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { USAGE | ALL [ PRIVILEGES ] }
    ON FOREIGN DATA WRAPPER fdw_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { USAGE | ALL [ PRIVILEGES ] }
    ON FOREIGN SERVER server_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { EXECUTE | ALL [ PRIVILEGES ] }
```

```
    ON { FUNCTION function_name ( [ [ argmode ] [ arg_name ] arg_type
 [, ...] ] ) [, ...]
         | ALL FUNCTIONS IN SCHEMA schema_name [, ...] }
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { USAGE | ALL [ PRIVILEGES ] }
    ON LANGUAGE lang_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { SELECT | UPDATE } [, ...] | ALL [ PRIVILEGES ] }
    ON LARGE OBJECT loid [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { CREATE | USAGE } [, ...] | ALL [ PRIVILEGES ] }
    ON SCHEMA schema_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { CREATE | ALL [ PRIVILEGES ] }
    ON TABLESPACE tablespace_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ ADMIN OPTION FOR ]
    role_name [, ...] FROM role_name [, ...]
    [ CASCADE | RESTRICT ]
```

描述

REVOKE 命令从一个或多个角色那里撤销先前授予的权限。关键字 PUBLIC 指的是由所有角色组成的隐式定义组。

有关各种权限类型的含义，请参见GRANT命令的描述。

请注意，任何特定角色实际拥有的权限，是直接授予给它的权限、授予给它当前所属任一角色的权限，以及授予给 PUBLIC 的权限之和。因此，例如，从 PUBLIC 撤销 SELECT 权限，并不一定意味着所有角色都失去了该对象上的 SELECT 权限：那些被直接授予该权限或通过其他角色获得该权限的角色仍然拥有它。类似地，从某个用户撤销 SELECT 权限，如果 PUBLIC 或其所属的其他角色仍然拥有 SELECT 权限，该用户仍可能使用 SELECT。

如果指定了 GRANT OPTION FOR，则只撤销该权限的授予选项，而不撤销权限本身。否则，权限和授予选项都会被撤销。

如果某个用户持有带授予选项的权限，并且已经将其授予其他用户，那么其他用户持有的这些权限称为依赖权限。若正在撤销第一个用户持有的该权限或其授予选项，并且存在依赖权限，则在指定 CASCADE 时这些依赖权限也会被一并撤销；否则，该撤销操作会失败。此递归撤销只影响那些通过某条可

追溯到本 REVOKE 命令目标用户的用户链授予的权限。因此，如果受影响用户还通过其他用户获得了同一权限，那么他们实际上可能仍保有该权限。

撤销表上的权限时，对应的列权限（如果有）也会自动在每个列上被撤销。

在撤销角色成员资格时，GRANT OPTION 改称为 ADMIN OPTION，但行为类似。另请注意，这种形式的命令不允许使用噪声词 GROUP。

注解

使用psql的 \dp 命令可以显示现有表和列上授予的权限。有关格式的信息，参见GRANT。对于非表对象，也有其他 \d 命令可以显示它们的权限。

用户只能撤销由自己直接授予的权限。例如，如果用户 A 已将某项带授予选项的权限授予用户 B，而用户 B 又将其授予用户 C，则用户 A 不能直接从 C 撤销该权限。相反，用户 A 可以从 B 撤销授予选项，并使用 CASCADE 选项，这样该权限就会进一步从 C 撤销。再例如，如果 A 和 B 都将同一权限授予了 C，A 可以撤销他自己授予的那份，但不能撤销 B 授予的那份，因此 C 实际上仍会拥有该权限。

当对象的非拥有者尝试对该对象执行 REVOKE 时，如果该用户在该对象上完全没有任何权限，命令将立即失败。只要该用户至少拥有某项权限，命令就会继续执行，
但只会撤销那些该用户持有授予选项的权限。如果未持有任何授予选项，REVOKE ALL PRIVILEGES 形式会发出警告；而其他形式如果命令中特别列出的任一权限未持有其授予选项，也会发出警告。（原则上，这些说明也适用于对象拥有者；
但由于拥有者始终被视为持有全部授予选项，这种情况实际上不会发生。）

如果超级用户选择执行 GRANT 或 REVOKE 命令，则该命令会像由受影响对象的拥有者发出那样执行。由于所有权限最终都来自对象拥有者（可能经由授予选项链间接传递），
超级用户可以撤销所有权限，但如上所述，这可能需要使用 CASCADE。

REVOKE 也可以由并非受影响对象拥有者的角色执行，只要该角色是拥有该对象之角色的成员，或者是持有该对象上 WITH GRANT OPTION 权限之角色的成员。在这种情况下，该命令会视同由实际拥有该对象的角色，或持有该对象上 WITH GRANT OPTION 权限的角色发出。例如，如果表 t1 由角色 g1 拥有，而角色 u1 是其成员，那么 u1 可以撤销 t1 上那些记录为由 g1 授出的权限。这包括由 u1 以及角色 g1 的其他成员作出的授权。

如果执行 REVOKE 的角色通过多条角色成员资格路径间接持有权限，则系统不会指明将使用哪一个上层角色来执行该命令。在这种情况下，最佳做法是使用 SET ROLE 切换成你希望作为其身份执行 REVOKE 的那个具体角色。否则，可能会撤销掉并非你本意要撤销的权限，或者根本没有撤销任何权限。

示例

撤销表 films 上授予所有用户的插入权限：

```
REVOKE INSERT ON films FROM PUBLIC;
```

撤销用户 manuel 在视图 kinds 上的所有权限：

```
REVOKE ALL PRIVILEGES ON kinds FROM manuel;
```

请注意，这实际上意味着“撤销所有由我授予的权限”。

撤销用户 joe 在角色 admins 中的成员资格：

```
REVOKE admins FROM joe;
```

兼容性

GRANT命令的兼容性注解同样适用于 `REVOKE`。按照标准，关键词 `RESTRICT` 或 `CASCADE` 是必需的，但 PostgreSQL 默认假定为 `RESTRICT`。

另见

GRANT

ROLLBACK

ROLLBACK — abort the current transaction

大纲

```
ROLLBACK [ WORK | TRANSACTION ]
```

描述

`ROLLBACK`回滚当前事务，并丢弃该事务所做的全部更新。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用COMMIT成功结束事务。

在事务块外执行 `ROLLBACK` 会发出警告，除此之外没有任何效果。

示例

要中止所有更改：

```
ROLLBACK;
```

兼容性

SQL 标准只规定了 `ROLLBACK` 和 `ROLLBACK WORK` 两种形式。除此之外，该命令完全符合标准。

另见

BEGIN, COMMIT, ROLLBACK TO SAVEPOINT

ROLLBACK PREPARED

ROLLBACK PREPARED — cancel a transaction that was earlier prepared for two-phase commit

大纲

```
ROLLBACK PREPARED transaction_id
```

描述

`ROLLBACK PREPARED`回滚一个处于预备状态的事务。

参数

`transaction_id`

要回滚的事务的事务标识符。

注解

要回滚预备事务，执行者必须是最初执行该事务的同一用户，或者是超级用户；但不必处在执行该事务的同一会话中。

这个命令不能在事务块内执行。该预备事务会被立即回滚。

当前所有处于预备状态的事务都列在 `pg_prepared_xacts` 系统视图中。

示例

回滚事务标识符为`foobar`的事务：

```
ROLLBACK PREPARED 'foobar';
```

另见

PREPARE TRANSACTION, COMMIT PREPARED

ROLLBACK TO SAVEPOINT

ROLLBACK TO SAVEPOINT — roll back to a savepoint

大纲

```
ROLLBACK [ WORK | TRANSACTION ] TO [ SAVEPOINT ] savepoint_name
```

描述

回滚该保存点建立后执行的所有命令。该保存点仍然有效，如有需要，
之后还可以再次回滚到它。

`ROLLBACK TO SAVEPOINT`会隐式销毁在指定保存点之后建立的所有保存点。

参数

`savepoint_name`

要回滚到的保存点名称。

注解

使用RELEASE SAVEPOINT可以销毁一个保存点，而不丢弃在它建立之后执行的命令所产生的效果。

指定一个尚未建立的保存点名称是一种错误。

就保存点而言，游标带有一些非事务性的行为。凡是在某个保存点内打开的游标，
在回滚该保存点时都会被关闭。如果先前打开的游标在某个随后又被回滚的保存点内受到了`FETCH`或`MOVE`命令的影响，那么该游标会保留在`FETCH`使其指向的位置上（也就是说，由`FETCH`引起的游标移动不会被回滚）。关闭游标同样不会因回滚而撤销。不过，
如果游标查询导致了其他副作用（例如该查询调用的易变函数带来的副作用），
且这些副作用发生在后来被回滚的保存点期间，那么它们会被回滚。
如果某个游标的执行导致事务中止，该游标会进入不可执行状态，因此即使事务可以通过`ROLLBACK TO SAVEPOINT`恢复，该游标也不能再使用。

示例

要撤销在`my_savepoint`建立后执行的命令的效果：

```
ROLLBACK TO SAVEPOINT my_savepoint;
```

游标位置不受保存点回滚的影响：

```
BEGIN;

DECLARE foo CURSOR FOR SELECT 1 UNION SELECT 2;

SAVEPOINT foo;
```

```
FETCH 1 FROM foo;
 ?column?
----------
        1

ROLLBACK TO SAVEPOINT foo;

FETCH 1 FROM foo;
 ?column?
----------
        2

COMMIT;
```

兼容性

SQL标准规定关键字`SAVEPOINT`是强制的，但 PostgreSQL和Oracle允许省略它。SQL 只允许在`ROLLBACK`之后使用`WORK`作为噪声词，不允许使用`TRANSACTION`。此外，SQL 还有一个可选子句 `AND [ NO ] CHAIN`，而PostgreSQL当前尚不支持。除此之外，该命令符合 SQL 标准。

另见

BEGIN, COMMIT, RELEASE SAVEPOINT, ROLLBACK, SAVEPOINT

SAVEPOINT

SAVEPOINT — define a new savepoint within the current transaction

大纲

```
SAVEPOINT savepoint_name
```

描述

`SAVEPOINT`在当前事务中建立一个新的保存点。

保存点是事务中的一种特殊标记，它允许回滚在其建立之后执行的所有命令，
从而将事务状态恢复到建立保存点时的状态。

参数

savepoint_name

赋予新保存点的名称。

注解

使用ROLLBACK TO SAVEPOINT回滚到保存点。使用RELEASE SAVEPOINT销毁保存点，同时保留其建立后执行的命令所产生的效果。

保存点只能在事务块内部建立。一个事务中可以定义多个保存点。

示例

要建立一个保存点，并在稍后撤销其建立后执行的所有命令的效果：

```
BEGIN;
    INSERT INTO table1 VALUES (1);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (2);
    ROLLBACK TO SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (3);
COMMIT;
```

上述事务将插入值 1 和 3，而不会插入 2。

要建立保存点，并在稍后销毁它：

```
BEGIN;
    INSERT INTO table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

上述事务将同时插入 3 和 4。

兼容性

SQL 要求在建立另一个同名保存点时自动销毁原有保存点。在 PostgreSQL中，旧保存点会被保留，不过在回滚或释放时只会使用最近建立的那个。（如果使用 `RELEASE SAVEPOINT`释放较新的保存点，较旧的保存点将再次可供 `ROLLBACK TO SAVEPOINT` 和 `RELEASE SAVEPOINT` 使用。）除此之外，`SAVEPOINT` 完全符合 SQL。

另见

BEGIN, COMMIT, RELEASE SAVEPOINT, ROLLBACK, ROLLBACK TO SAVEPOINT

SECURITY LABEL

SECURITY LABEL — define or change a security label applied to an object

大纲

```
SECURITY LABEL [ FOR provider ] ON
{
  TABLE object_name |
  COLUMN table_name.column_name |
  AGGREGATE agg_name (agg_type [, ...] ) |
  DOMAIN object_name |
  FOREIGN TABLE object_name
  FUNCTION function_name ( [ [ argmode ] [ argname ] argtype [, ...] ]
 ) |
  LARGE OBJECT large_object_oid |
  [ PROCEDURAL ] LANGUAGE object_name |
  SCHEMA object_name |
  SEQUENCE object_name |
  TYPE object_name |
  VIEW object_name
} IS 'label'
```

描述

`SECURITY LABEL`为数据库对象设置安全标签。一个给定的数据库对象可以关联任意数量的安全标签，每个标签提供者对应一个。标签提供者是使用函数`register_label_provider`注册自身的可加载模块。

> **注意**
>
> `register_label_provider`不是一个 SQL 函数；它只能从加载到后端的 C 代码中调用。

标签提供者决定给定标签是否有效，以及是否允许将该标签赋给给定对象。
给定标签的含义同样由标签提供者自行决定。PostgreSQL不限制标签提供者是否解释安全标签，也不限制其如何解释；它仅提供一种存储安全标签的机制。实际上，
此功能旨在支持与基于标签的强制访问控制（MAC）系统（例如SELinux）集成。
这类系统基于对象标签，而不是基于用户和组等传统的自主访问控制（DAC）概念，
做出所有访问控制决策。

参数

object_name
table_name.column_name
agg_name
function_name

要加上安全标签的对象名称。表、聚合、域、外部表、函数、序列、
类型和视图的名称可以带模式限定。

provider

要与该标签关联的标签提供者名称。指定的提供者必须已加载，并且必须同意所提议的标签设置操作。若只加载了一个提供者，则为简洁起见可以省略其名称。

arg_type

聚合函数作用于其上的输入数据类型。要引用一个零参数聚合函数，请在输入数据类型列表的位置写 `*`。

argmode

函数参数的模式：`IN`、`OUT`、`INOUT` 或 `VARIADIC`。如果省略，默认为 `IN`。注意 `SECURITY LABEL ON FUNCTION` 实际上并不关心 `OUT` 参数, since only the input arguments are needed to determine the function's identity. So it is sufficient to list the `IN`, `INOUT`, and `VARIADIC` arguments.

argname

函数参数的名称。注意 `SECURITY LABEL ON FUNCTION` 实际上并不关心参数名称，因为确定函数的标识只需要参数的数据类型。

argtype

大对象的 OID。

large_object_oid

大对象的 OID。

`PROCEDURAL`

这是一个噪声词。

label

新的安全标签，写成字符串字面量；也可以使用 `NULL` 来删除安全标签。

示例

下面的示例展示了如何更改表的安全标签。

```
SECURITY LABEL FOR selinux ON TABLE mytable IS 'system_u:object_r:
sepgsql_table_t:s0';
```

兼容性

在 SQL 标准中没有`SECURITY LABEL`命令。

另见

sepgsql, dummy_seclabel

SELECT

SELECT, TABLE, WITH — retrieve rows from a table or view

大纲

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
    * | expression [ [ AS ] output_name ] [, ...]
    [ FROM from_item [, ...] ]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING condition [, ...] ]
    [ WINDOW window_name AS ( window_definition ) [, ...] ]
    [ { UNION | INTERSECT | EXCEPT } [ ALL | DISTINCT ] select ]
    [ ORDER BY expression [ ASC | DESC | USING operator ] [ NULLS {
 FIRST | LAST } ] [, ...] ]
    [ LIMIT { count | ALL } ]
    [ OFFSET start [ ROW | ROWS ] ]
    [ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ]
    [ FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ] [..
.] ]

where from_item can be one of:

    [ ONLY ] table_name [ * ] [ [ AS ] alias [ ( column_alias [, ...]
 ) ] ]
    ( select ) [ AS ] alias [ ( column_alias [, ...] ) ]
    with_query_name [ [ AS ] alias [ ( column_alias [, ...] ) ] ]
    function_name ( [ argument [, ...] ] ) [ AS ] alias [ ( column_
alias [, ...] | column_definition [, ...] ) ]
    function_name ( [ argument [, ...] ] ) AS ( column_definition [, .
..] )
    from_item [ NATURAL ] join_type from_item [ ON join_condition |
 USING ( join_column [, ...] ) ]

and with_query is:

    with_query_name [ ( column_name [, ...] ) ] AS ( select | insert |
 update | delete )

TABLE [ ONLY ] table_name [ * ]
```

描述

SELECT 从零个或多个表中检索行。SELECT 的一般处理流程如下：

1. WITH列表中的所有查询都会被计算。这些实际上充当临时表，可以在FROM列表中引用。在FROM列表中多次引用的WITH查询只会计算一次。（参见下面的WITH Clause。）

2. 所有FROM列表中的元素都会被计算。（FROM列表中的每个元素都是一个真实或虚拟表。）如果在FROM列表中指定了多个元素，则它们会被交叉连接在一起。（参见下面的FROM Clause。）

3. 如果指定了WHERE子句，则不满足条件的所有行将从输出中删除。（请参见下面的WHERE Clause。）

4. 如果指定了GROUP BY子句，或者存在聚合函数调用，输出将被组合成在一个或多个值上匹配的行组，并计算聚合函数的结果。如果存在HAVING子句，它将消除不满足给定条件的组。（参见GROUP BY Clause和HAVING Clause。）

5. 实际输出行是使用每个选定行或行组的SELECT输出表达式计算的。（参见下面的SELECT List。）

6. SELECT DISTINCT消除结果中的重复行。SELECT DISTINCT ON会消除在所有指定表达式上匹配的行，只保留每组中的第一行。SELECT ALL（默认）将返回所有候选行，包括重复行。（参见下面的DISTINCT Clause。）

7. 使用操作符UNION、INTERSECT和EXCEPT，可以将多个SELECT语句的输出合并成一个结果集。UNION操作符返回在一个或两个结果集中的所有行。INTERSECT操作符返回同时出现在两个结果集中的所有行。EXCEPT操作符返回在第一个结果集中但不在第二个结果集中的行。在这三种情况下，除非指定ALL，否则将消除重复行。还可以添加噪声词DISTINCT，以明确指定去重。请注意，这里的默认行为是DISTINCT，即使SELECT本身的默认行为是ALL。（请参见下面的UNION Clause、INTERSECT Clause和EXCEPT Clause。）

8. 如果指定了ORDER BY子句，则返回的行按指定顺序排序。如果没有给出ORDER BY，则按系统认为最快的顺序返回行。（参见下面的ORDER BY Clause。）

9. 如果指定了LIMIT（或FETCH FIRST）或OFFSET子句，SELECT语句只返回结果行的子集。（参见下面的LIMIT Clause。）

10. 如果指定了 FOR UPDATE 或 FOR SHARE，SELECT 语句将选定的行锁定，防止并发更新（参见下面的FOR UPDATE/FOR SHARE Clause）。

你必须拥有 SELECT 命令中使用到的每一列上的 SELECT 权限。使用 FOR UPDATE 或 FOR SHARE 还要求 UPDATE 权限（对于每个这样选中的表，至少要有一列具有该权限）。

参数

WITH Clause

WITH 子句允许你指定一个或多个可在主查询中按名称引用的子查询。这些子查询在主查询执行期间实际上充当临时表或视图。每个子查询都可以是一条 SELECT、INSERT、UPDATE 或 DELETE 语句。在 WITH 中写数据修改语句（INSERT、UPDATE 或 DELETE）时，通常会带上 RETURNING 子句。构成主查询所读取临时表的，是 RETURNING 的输出，而不是语句所修改的底层表。如果省略 RETURNING，语句仍会执行，但不产生任何输出，因此主查询无法把它当作表来引用。

对于每个WITH查询，都必须指定一个名称（不带模式限定）。还可以指定一个列名列表；如果省略，则列名将从子查询中推导出来。

如果指定了RECURSIVE，则允许一个 SELECT子查询使用名称引用自身。这样一个子查询的形式必须是

non_recursive_term UNION [ALL | DISTINCT] *recursive_term*

其中递归自引用必须出现在UNION的右手边。每个查询中只允许一个递归自引用。
不支持递归数据修改语句，但是可以在一个数据查询语句中使用一个递归 SELECT查询的结果。
示例可见第 7.8 节。

RECURSIVE的另一个效果是 WITH查询不需要被排序：一个查询可以引用另一个在列表中比它靠后的查询（不过，循环引用或者互递归没有实现）。如果没有RECURSIVE，WITH 查询只能引用在WITH列表中位置更前面的兄弟 WITH查询。

WITH查询的一个关键特性是，每次执行主查询时，它们都只会求值一次，
即使主查询多次引用它们。特别是，无论主查询是否读取了它们的全部输出或任何输出，
数据修改语句都保证执行一次且仅执行一次。

主查询和WITH查询（概念上）都在同一时间执行。这意味着，除了读取其RETURNING
输出之外，查询的其他部分都看不到WITH中数据修改语句的效果。
如果两个这样的数据修改语句试图修改同一行，结果未指定。

更多信息请见第 7.8 节。

FROM Clause

FROM 子句为 SELECT 指定一个或更多源表。如果指定了多个源表，结果将是所有源表的笛卡尔积（交叉连接）。但通常会增加限定条件，把返回的行限制为该笛卡尔积的一个小子集。

该 FROM 子句可以包含以下元素：

table_name

要扫描的现有表或视图的名称（可选模式限定符）。如果在表名之前指定ONLY，
则仅扫描该表。如果未指定ONLY，则扫描该表及其所有后代表（如果有）。可选地，
可以在表名后指定*，以明确指示包括后代表。

alias

包含该别名的FROM项的替代名称。别名可用于简写，
或者消除自连接（同一张表被扫描多次）中的歧义。提供别名后，
它会完全隐藏表或函数的实际名称；例如给定FROM foo AS f，SELECT的其余部分必须把
这个FROM 项写成f而不是foo。如果写了别名，还可以写列别名列表，为该表的一个或多个
列提供替代名称。

select

子SELECT可以出现在FROM子句中，它的作用就像在这个SELECT
命令的执行期间创建了一个临时表。注意，子SELECT必须用圆括号括起来，并且
必须为其提供别名。这里也可以使用VALUES命令。

with_query_name

WITH查询通过写出它的名称来引用，就像该查询名是表名一样。（事实上，
对于主查询而言，WITH查询会遮蔽任何同名的真实表；如有必要，
可以通过模式限定表名来引用该同名真实表。）也可以像对待表一样为它提供别名。

function_name

函数调用可以出现在FROM子句中。（这对于返回结果集的函数尤其有用，
但任何函数都可以使用。）其效果就像在这条SELECT命令执行期间，

将函数的输出创建成了一张临时表。也可以使用别名。如果写了别名，
还可以写一个列别名列表，为函数复合返回类型中的一个或多个属性提供替代名称。
如果函数被定义为返回record数据类型，则必须给出别名或关键字AS，
后面跟一个如下形式的列定义列表：(*column_name data_type* [, ...])。列定义列表必须与函数实际返回的列数和类型相匹配。

join_type

以下之一：

- [INNER] JOIN
- LEFT [OUTER] JOIN
- RIGHT [OUTER] JOIN
- FULL [OUTER] JOIN
- CROSS JOIN

For the INNER and OUTER join types, a join condition must be specified, namely exactly one of NATURAL, ON *join_condition*, or USING (*join_column* [, ...]). See below for the meaning. For CROSS JOIN, none of these clauses can appear.

JOIN 子句组合两个 FROM 项。如有必要可用圆括号确定嵌套的顺序。在没有圆括号时，JOIN 从左到右嵌套。在任何情况下，JOIN 的结合都比分隔 FROM 项的逗号更紧。

CROSS JOIN和INNER JOIN产生简单的笛卡尔积，与在FROM顶层列出这两个表得到的结果相同，但会受到连接条件（如果有）的限制。CROSS JOIN等价于INNER JOIN ON (TRUE)，也就是说，没有行会被条件过滤掉。这些连接类型只是提供了一种方便的记法，因为它们所做的一切都可以用普通的FROM和WHERE完成。

LEFT OUTER JOIN返回已限定笛卡尔积中的所有行（即，通过其连接条件的所有组合行），外加左侧表中每一行的一个副本；对于这些行，不存在通过连接条件的右侧行。
这样的左侧行会通过在右侧列中插入空值而扩展到连接表的完整宽度。注意，
在决定哪些行有匹配项时，只考虑JOIN子句自身的条件；外层条件是在之后应用的。

相反，RIGHT OUTER JOIN返回所有连接后的行，再加上每个未匹配右侧行对应的一行（左侧用空值扩展）。这只是一种记法上的便利，因为你可以通过交换左右表将其改写成LEFT OUTER JOIN。

FULL OUTER JOIN返回所有连接后的行，再加上每个未匹配的左侧行（右侧用空值扩展），以及每个未匹配的右侧行（左侧用空值扩展）。

ON *join_condition*

join_condition 是一个表达式，其结果类型为 boolean（类似于 WHERE 子句），用于指定连接中哪些行被视为匹配。

USING (*join_column* [, ...])

一个形如USING (a, b, ...)的子句是ON left_table.a = right_table.a AND left_table.b = right_table.b ...的简写。此外，USING意味着只有每对等价列中的一个会包含在连接输出中，而不是两者都包含。

NATURAL

NATURAL 是一个简写，表示一个列出两个表中所有同名列的 USING 列表。

WHERE Clause

可选的WHERE子句的形式

WHERE *condition*

其中*condition* 是任一计算得到boolean类型结果的表达式。任何不满足这个条件的行都会从输出中被消除。如果用一行的实际值替换其中的变量引用后，该表达式返回真，则该行符合条件。

GROUP BY Clause

可选的 GROUP BY 子句的一般形式为

GROUP BY *expression* [, ...]

GROUP BY会把所有在分组表达式上具有相同值的已选中行压缩成单独一行。*expression*可以是输入列名、输出列（SELECT列表项）的名称或序号或者由输入列值构成的任意表达式。在出现歧义时，GROUP BY名称将被解释为输入列名而不是输出列名。

（相反，如果没有GROUP BY，聚集会对所有被选中的行计算产生单个值。）当存在GROUP BY子句时，SELECT列表表达式不能引用非分组列，除非是在聚合函数内部，或者该非分组列函数依赖于分组列，因为否则对于一个非分组列可能会有多个可返回的值。如果分组列（或其子集）是包含该非分组列的表的主键，则存在函数依赖。

HAVING Clause

可选的HAVING子句的形式

HAVING *condition*

其中*condition*与 WHERE子句中指定的条件相同。

HAVING消除不满足该条件的分组行。HAVING与WHERE不同：WHERE会在应用GROUP BY之前过滤个体行，而HAVING过滤由 GROUP BY创建的分组行。*condition*中引用的每一列都必须无歧义地引用某个分组列，除非该引用出现在聚合函数中，或者该非分组列函数依赖于分组列。

即使没有GROUP BY子句，HAVING 的存在也会把一个查询转变成一个分组查询。这和查询中包含聚合函数但没有 GROUP BY子句时的情况相同。所有被选择的行都被认为是一个单一分组，并且SELECT列表和 HAVING子句只能从聚合函数内部引用表列。如果该 HAVING条件为真，这样的查询将输出单独一行；否则不返回行。

WINDOW Clause

可选的 WINDOW 子句的一般形式为：

WINDOW *window_name* AS (*window_definition*) [, ...]

其中，*window_name* 是一个名称，可从 OVER 子句或后续窗口定义中引用，而 *window_definition* 的定义为：

[*existing_window_name*]
[PARTITION BY *expression* [, ...]]
[ORDER BY *expression* [ASC | DESC | USING *operator*] [NULLS { FIRST | LAST }] [, ...]]

[*frame_clause*]

如果指定了一个*existing_window_name*，它必须引用WINDOW列表中一个更早出现的项。新窗口将从该项中复制它的分区子句，以及排序子句（如果有）。在这种情况下，新窗口不能指定它自己的PARTITION BY子句，并且只有在被复制的窗口没有排序子句时，才可以指定 ORDER BY。新窗口总是使用自己的帧子句，而被复制的窗口不得指定帧子句。

PARTITION BY列表元素的解释以GROUP BY Clause元素的方式进行，不过它们总是简单表达式并且绝不能是输出列的名称或编号。另一个区别是这些表达式可以包含聚合函数调用，而这在常规GROUP BY 子句中是不被允许的。它们被允许的原因是窗口是出现在分组和聚合之后的。

类似地，ORDER BY列表元素的解释也以语句级ORDER BY Clause元素的方式进行，不过该表达式总是被当做简单表达式并且绝不会是输出列的名称或编号。

可选的 *frame_clause* 定义窗口帧，供依赖于帧的窗口函数使用（并非所有窗口函数都依赖帧）。窗口帧是与查询中每一行相关联的一组行；查询中的这一行称为当前行。*frame_clause* 可以是以下之一：

{ RANGE | ROWS } *frame_start*
{ RANGE | ROWS } BETWEEN *frame_start* AND *frame_end*

其中，*frame_start* 和 *frame_end* 可以是以下之一：

UNBOUNDED PRECEDING
value PRECEDING
CURRENT ROW
value FOLLOWING
UNBOUNDED FOLLOWING

如果省略 *frame_end*，则默认为 CURRENT ROW。其限制是：*frame_start* 不能是 UNBOUNDED FOLLOWING，*frame_end* 不能是 UNBOUNDED PRECEDING，并且 *frame_end* 在上面列表中的位置不能早于 *frame_start* 的位置 – 例如，RANGE BETWEEN CURRENT ROW AND *value* PRECEDING 是不允许的。

默认的帧选项是RANGE UNBOUNDED PRECEDING，它等同于RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW；它将帧设置为从分区起点到当前行最后一个同等行的所有行（同等行是ORDER BY认为与当前行等价的行；如果没有ORDER BY，则所有行都属于同等行）。一般而言，UNBOUNDED PRECEDING表示帧从分区的第一行开始，而UNBOUNDED FOLLOWING同样表示帧在分区的最后一行结束（无论采用RANGE还是ROWS模式）。在ROWS模式下，CURRENT ROW表示帧从当前行开始或在当前行结束；但在RANGE模式下，它表示帧从ORDER BY排序中当前行的第一个同等行开始，或在最后一个同等行结束。目前，*value* PRECEDING和*value* FOLLOWING仅允许用于ROWS模式。它们表示帧从当前行之前或之后相应行数的那一行开始或结束。*value*必须是不包含任何变量、聚合函数或窗口函数的整数表达式。该值不能为空值或负数，但可以为零，此时选择当前行本身。

请注意，如果ORDER BY排序没有将行排成唯一顺序，ROWS选项可能产生不可预测的结果。RANGE选项旨在确保ORDER BY排序中的同等行得到相同处理；所有同等行都会位于同一帧中。

WINDOW子句的目的是指定出现在查询的SELECT List或ORDER BY Clause中的窗口函数的行为。这些函数可以在它们的 OVER子句中用名称引用WINDOW 子句项。不过，WINDOW子句项不是必须被引用。如果在查询中没有用到它，它会被简单地忽略。可以使用根本没有任何 WINDOW子句的窗口函数，因为窗口函数调用可以直接在其OVER子句中指定它的窗口定义。不过，当多个窗口函数都需要相同的窗口定义时，WINDOW子句能够减少输入量。

窗口函数的详细描述在第 3.5 节、第 4.2.8 节以及第 7.2.4 节中。

SELECT List

SELECT列表（位于关键词 SELECT和FROM之间）指定构成 SELECT语句输出行的表达式。这些表达式可以（并且通常确实会）引用FROM子句中计算得到的列。

正如在表中一样，SELECT 的每一个输出列都有一个名称。在一个简单的 SELECT 中，这个名称只是被用来标记要显示的列，但是当 SELECT 是一个大型查询的子查询时，大型查询会把该名称看做子查询产生的虚表的列名。要指定输出列使用的名称，可在列的表达式后面写上 AS *output_name*。（可以省略 AS，但仅当期望的输出名不与任何 PostgreSQL 关键字（见附录 C）冲突时。为了防止将来可能新增关键字，建议始终写上 AS，或者给输出名加上双引号。）如果没有指定列名，名称将由 PostgreSQL 自动选择：如果列的表达式是一个简单的列引用，所选名称就是该列的名称；在更复杂的情况下，通常会使用形如 ?column*N*? 的生成名称。

一个输出列的名称可以被用来在ORDER BY以及 GROUP BY子句中引用该列的值，但是不能用于WHERE和HAVING子句（在其中必须写出表达式）。

可以在输出列表中写*来取代表达式，它是被选中行的所有列的一种简写方式。还可以写 *table_name*.*，它是只来自那个表的所有列的简写形式。在这些情况中无法用 AS指定新的名称，输出行的名称将和表列的名称相同。

DISTINCT Clause

如果指定了SELECT DISTINCT，所有重复的行会被从结果集中移除（为每一组重复的行保留一行）。SELECT ALL则指定相反的行为：所有行都会被保留，这也是默认情况。

SELECT DISTINCT ON (*expression* [, ...]) 只保留在给定表达式上计算相等的行集合中的第一行。DISTINCT ON表达式使用和 ORDER BY相同的规则（见上文）解释。注意，除非用 ORDER BY来确保所期望的行出现在第一位，每一个集合的“第一行”是不可预测的。例如：

```
SELECT DISTINCT ON (location) location, time, report
    FROM weather_reports
    ORDER BY location, time DESC;
```

为每个地点检索最近的天气报告。但是如果我们不使用 ORDER BY来强制对每个地点的时间值进行降序排序，我们为每个地点得到的报告的时间可能是无法预测的。

DISTINCT ON表达式必须匹配最左边的 ORDER BY表达式。ORDER BY子句通常将包含额外的表达式，这些额外的表达式用于决定在每一个 DISTINCT ON分组内行的优先级。

UNION Clause

UNION 子句的一般形式如下：

select_statement UNION [ALL | DISTINCT] *select_statement*

select_statement 是任何不带 ORDER BY、LIMIT、FOR UPDATE 或 FOR SHARE 子句的 SELECT 语句。（如果用圆括号括起，ORDER BY 和 LIMIT 可以附着到子表达式上。如果没有圆括号，这些子句将被认为作用于 UNION 的结果，而不是其右侧的输入表达式。）

UNION操作符计算相关 SELECT语句所返回的行的并集。如果一行至少出现在两个结果集中的一个内，它就会在并集中。作为 UNION两个操作数的 SELECT语句必须产生相同数量的列并且对应位置上的列必须具有兼容的数据类型。

UNION的结果不会包含重复行，除非指定了 ALL选项。ALL会阻止消除重复（因此，UNION ALL通常显著地快于UNION，尽量使用ALL）。也可以写上DISTINCT，
以显式指定默认的去重行为。

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 UNION操作符会从左至右计算。

当前，FOR UPDATE 和 FOR SHARE 不能为 UNION 的结果或 UNION 的任何输入指定。

INTERSECT Clause

INTERSECT 子句的一般形式如下：

select_statement INTERSECT [ALL | DISTINCT] *select_statement*

select_statement is any SELECT statement without an ORDER BY, LIMIT, FOR UPDATE, or FOR SHARE clause.

INTERSECT操作符计算相关 SELECT语句返回的行的交集。如果一行同时出现在两个结果集中，它就在交集中。

INTERSECT的结果不会包含重复行，除非指定了 ALL选项。如果有ALL，一个在左表中有 *m*次重复并且在右表中有*n* 次重复的行将会在结果中出现 min(*m*,*n*) 次。也可以写上DISTINCT，以显式指定默认的去重行为。

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 INTERSECT操作符会从左至右计算。INTERSECT的优先级比 UNION更高。也就是说，A UNION B INTERSECT C将被读成A UNION (B INTERSECT C)。

当前，FOR UPDATE 和 FOR SHARE 不能为 INTERSECT 的结果或 INTERSECT 的任何输入指定。

EXCEPT Clause

EXCEPT 子句的一般形式如下：

select_statement EXCEPT [ALL | DISTINCT] *select_statement*

select_statement is any SELECT statement without an ORDER BY, LIMIT, FOR UPDATE, or FOR SHARE clause.

EXCEPT操作符计算位于左侧 SELECT语句的结果中但不在右侧语句结果中的行集合。

EXCEPT的结果不会包含重复行，除非指定了 ALL选项。如果有ALL，一个在左表中有 *m*次重复并且在右表中有 *n*次重复的行将会在结果集中出现 max(*m*-*n*,0) 次。也可以写上DISTINCT，以显式指定默认的去重行为。

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 EXCEPT操作符会从左至右计算。EXCEPT的优先级与 UNION相同。

当前，FOR UPDATE 和 FOR SHARE 不能为 EXCEPT 的结果或 EXCEPT 的任何输入指定。

ORDER BY Clause

可选的ORDER BY子句的形式如下：

ORDER BY *expression* [ASC | DESC | USING *operator*] [NULLS { FIRST | LAST }] [, ...]

ORDER BY子句导致结果行被按照指定的表达式排序。如果两行按照最左边的表达式是相等的，则会根据下一个表达式比较它们，依次类推。如果按照所有指定的表达式它们都是相等的，则它们被返回的顺序取决于实现。

每一个*expression* 可以是输出列（SELECT列表项）的名称或者序号，它也可以是由输入列值构成的任意表达式。

序号指的是输出列的顺序（从左至右）位置。这种特性可以为不具有唯一名称的列定义一个顺序。这不是绝对必要的，因为总是可以使用 AS子句为输出列赋予一个名称。

也可以在ORDER BY子句中使用任意表达式，包括没有出现在SELECT输出列表中的列。因此，下面的语句是合法的：

```
SELECT name FROM distributors ORDER BY code;
```

这种特性的一个限制是一个应用在UNION、INTERSECT或EXCEPT子句结果上的 ORDER BY只能指定输出列名称或序号，但不能指定表达式。

如果一个ORDER BY表达式是一个既匹配输出列名称又匹配输入列名称的简单名称，ORDER BY将把它解读成输出列名称。这与在同样情况下GROUP BY会做出的选择相反。这种不一致是为了与SQL 标准兼容。

可以在ORDER BY子句中任一表达式之后附加关键字 ASC（升序）或DESC（降序）。如果没有指定，ASC被假定为默认值。或者，可以在USING 子句中指定一个特定的排序操作符名称。一个排序操作符必须是某个 B-树操作符族的小于或者大于成员。ASC通常等价于 USING <而DESC通常等价于 USING >（但是一种用户定义数据类型的创建者可以准确地定义默认排序顺序是什么，并且它可能会对应于其他名称的操作符）。

如果指定NULLS LAST，空值会排在非空值之后；如果指定 NULLS FIRST，空值会排在非空值之前。如果都没有指定，在指定或者隐含ASC时的默认行为是NULLS LAST，而指定或者隐含DESC时的默认行为是 NULLS FIRST（因此，默认行为是空值大于非空值）。当指定USING时，默认的空值顺序取决于该操作符是否为小于或者大于操作符。

注意顺序选项只应用到它们所跟随的表达式上。例如 ORDER BY x, y DESC和 ORDER BY x DESC, y DESC是不同的。

字符串数据会按适用于被排序列的排序规则来排序。根据需要可以通过在 *expression*中包括一个 COLLATE子句来覆盖，例如 ORDER BY mycolumn COLLATE "en_US"。更多信息请见第 4.2.10 节和第 22.2 节。

LIMIT Clause

该 LIMIT 子句由两个独立的子句组成：

```
LIMIT { count | ALL }
OFFSET start
```

count 指定最多返回多少行，而 *start* 指定开始返回行之前要跳过的行数。如果两者都指定，则先跳过 *start* 行，然后开始计数并返回 *count* 行。

如果 *count* 表达式计算结果为 NULL，则视为 LIMIT ALL，即不限制。如果 *start* 计算结果为 NULL，则等同于 OFFSET 0。

SQL:2008 引入了另一种实现相同结果的语法，PostgreSQL 也支持它。其形式如下：

```
OFFSET start { ROW | ROWS }
FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY
```

在这种语法中，要为 *start* 或 *count* 写除简单整数常量以外的任何内容，必须给它加上圆括号。如果在 FETCH 子句中省略 *count*，它默认为 1。ROW、ROWS 以及 FIRST、NEXT 都是不影响这些子句效果的赘词。按照标准，如果两个子句同时出现，OFFSET 子句必须在 FETCH 子句之前；但 PostgreSQL 更宽松，允许任意顺序。

在使用LIMIT时，用一个ORDER BY子句把结果行约束到一个唯一顺序是个好办法。否则你将得到该查询结果行的一个不可预测的子集 – 你可能要求从第 10 到第 20 行，但是在什么顺序下的第 10 到第 20 呢？除非指定ORDER BY，你是不知道顺序的。

查询规划器在生成一个查询计划时会考虑LIMIT，因此根据你使用的LIMIT和OFFSET，你很可能得到不同的计划（得到不同的行序）。所以，使用不同的 LIMIT/OFFSET值来选择一个查询结果的不同子集将会给出不一致的结果，除非你用ORDER BY强制一种可预测的结果顺序。这不是一个缺陷，它是 SQL 不承诺以任何特定顺序（除非使用 ORDER BY来约束顺序）给出一个查询结果这一事实造成的必然后果。

如果没有一个ORDER BY来强制选择一个确定的子集，重复执行同样的LIMIT查询甚至可能会返回一个表中行的不同子集。同样，这也不是一种缺陷，在这种情况下也无法保证结果的确定性。

FOR UPDATE/FOR SHARE Clause

FOR UPDATE 子句的形式如下：

```
FOR UPDATE [ OF table_name [, ...] ] [ NOWAIT ]
```

与它密切相关的 FOR SHARE 子句的形式如下：

```
FOR SHARE [ OF table_name [, ...] ] [ NOWAIT ]
```

FOR UPDATE 导致由 SELECT 语句检索到的行被锁定，就像要更新它们一样。这可以防止它们被其他事务修改或删除，直到当前事务结束。也就是说，其他事务中试图对这些行执行 UPDATE、DELETE 或 SELECT FOR UPDATE 的操作将被阻塞，直到当前事务结束。此外，如果来自另一个事务的 UPDATE、DELETE 或 SELECT FOR UPDATE 已经锁定了某个被选中的行，SELECT FOR UPDATE 将等待该其他事务完成，然后锁定并返回更新后的行（如果该行已被删除，则不返回任何行）。而在 REPEATABLE READ 或 SERIALIZABLE 事务中，如果要锁定的行自事务开始以来已被更改，则会抛出一个错误。更多讨论见第 13 章。

FOR SHARE 的行为类似，但它在每个被检索的行上获取共享锁而非排他锁。共享锁阻塞其他事务对这些行执行 UPDATE、DELETE 或 SELECT FOR UPDATE，但不会阻止它们执行 SELECT FOR SHARE。

为了防止该操作等待其他事务提交，可使用NOWAIT 选项。使用NOWAIT时，如果选中的行不能被立即锁定，该语句会直接报错而不是等待。注意，NOWAIT只适用于行级锁；所需的ROW SHARE 表级锁仍会按常规方式取得（见第 13 章）。如果想要不等待的表级锁，你可以先使用带NOWAIT 的LOCK。

如果在 FOR UPDATE 或 FOR SHARE 中提到了特定的表，则只有来自那些表的行会被锁定；SELECT 中用到的任何其他表还是被简单地照常读取。一个没有表列表的 FOR UPDATE 或 FOR SHARE 子句会影响该语句中用到的所有表。如果把 FOR UPDATE 或 FOR SHARE 应用到一个视图或子查询，它会影响该视图或子查询中用到的所有表。However, FOR UPDATE/FOR SHARE do not apply to WITH queries referenced by the primary query. If you want row locking to occur within a WITH query, specify FOR UPDATE or FOR SHARE within the WITH query.

如果有必要对不同的表指定不同的锁定行为，可以写多个 FOR UPDATE 和 FOR SHARE 子句。如果同一个表同时被 FOR UPDATE 和 FOR SHARE 子句提到（或被隐式地影响到），那么它会按 FOR UPDATE 处理。类似地，如果在任何影响一个表的子句中指定了 NOWAIT，就会按此行为处理该表。

在返回的行无法清楚地与各个表行对应起来的上下文中，不能使用 FOR UPDATE 和 FOR SHARE；例如它们不能与聚合一起使用。

当 FOR UPDATE 或 FOR SHARE 出现在一个 SELECT 查询的顶层时，被锁定的行正好就是该查询返回的行；在连接查询的情况下，被锁定的行是那些对返回的连接行有贡献的行。此外，自该查询的快照起满足查询条件的行将被锁定，即使它们在该快照之后被更新并因此不再满足查询条件，也不会被返回。如果使用了 LIMIT，一旦返回的行数已满足限制，锁定就会停止（但注意被 OFFSET 跳过的行也会被锁定）。类似地，如果在游标的查询中使用 FOR UPDATE 或 FOR SHARE，只有游标实际取出或越过的行才会被锁定。

当 FOR UPDATE 或 FOR SHARE 出现在一个子 SELECT 中时，被锁定的行是子查询返回给外层查询的行。这可能比只检查子查询本身所显示的要少，因为来自外层查询的条件可能被用来优化子查询的执行。例如，

```
SELECT * FROM (SELECT * FROM mytable FOR UPDATE) ss WHERE col1 = 5;
```

将只锁定 col1 = 5 的行，即使该条件在文本上并不在子查询之内。

小心

避免锁定一行然后又在后面的保存点或 PL/pgSQL 异常块中修改它。后续的回滚会导致锁丢失。例如：

```
BEGIN;
SELECT * FROM mytable WHERE key = 1 FOR UPDATE;
SAVEPOINT s;
UPDATE mytable SET ... WHERE key = 1;
ROLLBACK TO s;
```

在执行 ROLLBACK 之后，该行实际上已被解锁，而不是回到保存点之前已锁定但未修改的状态。如果当前事务中锁定的行被更新或删除，或者共享锁被升级为排他锁，就会出现这种风险：在所有这些情况下，之前的锁状态都会被遗忘。如果事务随后回滚到原始锁定命令和后续更改之间的某个状态，该行将看起来完全没有被锁定。这是一个实现上的缺陷，将在 PostgreSQL 的未来版本中解决。

小心

一个运行在 READ COMMITTED 事务隔离级别并且使用 ORDER BY 和 FOR UPDATE/SHARE 的 SELECT 命令有可能返回无序的行。这是因为 ORDER BY 会被首先应用。该命令对结果排序，但可能接着在尝试获得一行或多行上的锁时阻塞。一旦 SELECT 不再阻塞，某些排序列的值可能已经被修改，导致这些行看起来乱序（尽管按照原始列值来看它们是有序的）。如有需要，可以通过把 FOR UPDATE/SHARE 子句放在子查询中来绕过这个问题，例如

```
SELECT * FROM (SELECT * FROM mytable FOR UPDATE) ss ORDER BY
 column1;
```

> 注意，这样会锁定 mytable 的所有行，而顶层的 FOR UPDATE 只会锁定实际返回的行。如果 ORDER BY 还与 LIMIT 或其他限制条件结合使用，这可能带来显著的性能差异。因此，只有在预期排序列会并发更新且严格要求结果有序时，才建议使用这种技术。
>
> 在REPEATABLE READ或SERIALIZABLE事务隔离级别下，这将导致串行化失败（SQLSTATE为'40001'），因此在这些隔离级别下不可能接收到无序的行。

TABLE 命令

命令

TABLE *name*

等价于

SELECT * FROM *name*

它可以作为顶层命令使用，也可以作为复杂查询某些部分中一种节省空间的语法变体。

示例

连接表 films 和表 distributors：

```
SELECT f.title, f.did, d.name, f.date_prod, f.kind
    FROM distributors d, films f
    WHERE f.did = d.did

       title       | did |     name     | date_prod  |   kind
-------------------+-----+--------------+------------+----------
 The Third Man     | 101 | British Lion | 1949-12-23 | Drama
 The African Queen | 101 | British Lion | 1951-08-11 | Romantic
 ...
```

要对所有电影的len列求和并且用 kind对结果分组：

```
SELECT kind, sum(len) AS total FROM films GROUP BY kind;

   kind   | total
----------+-------
 Action   | 07:34
 Comedy   | 02:58
 Drama    | 14:28
 Musical  | 06:42
 Romantic | 04:38
```

要对所有电影的len列求和、对结果按照 kind分组并且显示总长小于 5 小时的分组：

```
SELECT kind, sum(len) AS total
    FROM films
    GROUP BY kind
    HAVING sum(len) < interval '5 hours';
```

```
   kind   | total
----------+-------
 Comedy   | 02:58
 Romantic | 04:38
```

下面两个示例都是根据第二列（name）的内容来排序结果：

```
SELECT * FROM distributors ORDER BY name;
SELECT * FROM distributors ORDER BY 2;

 did |       name
-----+------------------
 109 | 20th Century Fox
 110 | Bavaria Atelier
 101 | British Lion
 107 | Columbia
 102 | Jean Luc Godard
 113 | Luso films
 104 | Mosfilm
 103 | Paramount
 106 | Toho
 105 | United Artists
 111 | Walt Disney
 112 | Warner Bros.
 108 | Westward
```

接下来的示例展示了如何得到表distributors和 actors的并集，把结果限制为那些在每个表中以字母 W 开始的行。这里只需要不重复的行，因此省略了关键词 ALL。

```
distributors:               actors:
 did |     name              id |     name
-----+--------------        ----+----------------
 108 | Westward                1 | Woody Allen
 111 | Walt Disney             2 | Warren Beatty
 112 | Warner Bros.            3 | Walter Matthau
 ...                          ...

SELECT distributors.name
    FROM distributors
    WHERE distributors.name LIKE 'W%'
UNION
SELECT actors.name
    FROM actors
    WHERE actors.name LIKE 'W%';

      name
----------------
 Walt Disney
 Walter Matthau
 Warner Bros.
 Warren Beatty
 Westward
 Woody Allen
```

这个示例展示了如何在FROM子句中使用函数，分别使用和不使用列定义列表：

```
CREATE FUNCTION distributors(int) RETURNS SETOF distributors AS $$
    SELECT * FROM distributors WHERE did = $1;
$$ LANGUAGE SQL;

SELECT * FROM distributors(111);
 did |    name
-----+-------------
 111 | Walt Disney

CREATE FUNCTION distributors_2(int) RETURNS SETOF record AS $$
    SELECT * FROM distributors WHERE did = $1;
$$ LANGUAGE SQL;

SELECT * FROM distributors_2(111) AS (f1 int, f2 text);
 f1  |     f2
-----+-------------
 111 | Walt Disney
```

这个例子展示了如何使用一个简单的 `WITH` 子句：

```
WITH t AS (
    SELECT random() as x FROM generate_series(1, 3)
  )
SELECT * FROM t
UNION ALL
SELECT * FROM t

         x
--------------------
  0.534150459803641
  0.520092216785997
 0.0735620250925422
  0.534150459803641
  0.520092216785997
 0.0735620250925422
```

请注意，`WITH` 查询只被求值了一次，因此我们得到了两组相同的三个随机值。

这个示例使用`WITH RECURSIVE`从一个只显示直接下属的表中寻找雇员 Mary 的所有下属（直接的或者间接的）以及他们的间接层数：

```
WITH RECURSIVE employee_recursive(distance, employee_name, manager_
name) AS (
    SELECT 1, employee_name, manager_name
    FROM employee
    WHERE manager_name = 'Mary'
  UNION ALL
    SELECT er.distance + 1, e.employee_name, e.manager_name
    FROM employee_recursive er, employee e
    WHERE er.employee_name = e.manager_name
  )
SELECT distance, employee_name FROM employee_recursive;
```

注意这种递归查询的典型形式：一个初始条件，后面跟着 `UNION`，然后是查询的递归部分。要确保查询的递归部分最终将不返回任何行，否则该查询将无限循环（更多示例见第 7.8 节）。

兼容性

当然，SELECT语句与 SQL 标准兼容。但它也有一些扩展和缺失的特性。

省略的FROM子句

PostgreSQL 允许省略 FROM 子句。它可以直接用于计算简单表达式的结果：

```
SELECT 2+2;

 ?column?
----------
        4
```

其他一些 SQL 数据库必须引入一个只有一行的虚拟表，才能从该表执行 SELECT。

请注意，如果未指定 FROM 子句，查询就不能引用任何数据库表。例如，以下查询是无效的：

```
SELECT distributors.* WHERE distributors.name = 'Westward';
```

PostgreSQL 在 8.1 之前的版本会接受这种形式的查询，并为查询引用的每个表在该查询的 FROM 子句中添加一个隐式项。现在不再允许这种做法。

省略AS关键词

在 SQL 标准中，只要新列名是一个合法的列名（就是说与任何保留关键词不同），就可以省略输出列名之前的可选关键词AS。
PostgreSQL要稍微严格些：只要新列名匹配任何关键词（保留或者非保留）就需要AS。
推荐的习惯是使用 AS或者带双引号的输出列名来防止与未来增加的关键词可能的冲突。

在FROM项中，标准和 PostgreSQL都允许在非保留关键字别名前省略AS。但是由于语法歧义，这种写法无法用于输出列名。

ONLY与继承

SQL 标准要求在使用ONLY时用圆括号括起表名，例如SELECT * FROM ONLY (tab1), ONLY (tab2) WHERE ...。PostgreSQL认为这些圆括号是可选的。

PostgreSQL允许在末尾写上*，显式指定包含子表的非ONLY行为。标准不允许这种写法。

（这些要点同样适用于所有支持ONLY选项的 SQL 命令。）

GROUP BY和ORDER BY可用的名字空间

在 SQL-92 标准中，一个ORDER BY子句只能使用输出列名或者序号，而一个GROUP BY子句只能使用基于输入列名的表达式。PostgreSQL扩展了这两种子句以允许它们使用其他的选择（但如果有歧义时还是使用标准的解释）。
PostgreSQL也允许两种子句指定任意表达式。
注意出现在一个表达式中的名称将总是被当做输入列名而不是输出列名。

SQL:1999 及其后的标准使用了一种略微不同的定义，它并不完全向后兼容 SQL-92。不过，在大部分的情况下，PostgreSQL会以与 SQL:1999 相同的方式解释ORDER BY或GROUP BY表达式。

函数依赖

只有当一个表的主键被包括在GROUP BY列表中时，PostgreSQL才识别函数依赖（允许从GROUP BY中省略列）。SQL 标准指定了应该要识别的额外情况。

WINDOW子句的限制

SQL 标准为窗口*frame_clause*提供了更多选项。PostgreSQL目前仅支持上面列出的选项。

LIMIT和OFFSET

LIMIT和OFFSET子句是PostgreSQL特有的语法，MySQL也使用这种语法。SQL:2008 标准引入了OFFSET ... FETCH {FIRST|NEXT} ...子句来实现相同功能，如上面的LIMIT Clause所示。IBM DB2也使用这种语法。（为Oracle编写的应用经常采用一种变通办法，通过自动生成的rownum列实现这些子句的效果，而 PostgreSQL 中没有这一列。）

FOR UPDATE and FOR SHARE

尽管 FOR UPDATE 出现在 SQL 标准中，但标准只允许它作为 DECLARE CURSOR 的一个选项。PostgreSQL 允许它出现在任何 SELECT 查询以及子 SELECT 中，但这是一个扩展。FOR SHARE 变体以及 NOWAIT 选项未出现在标准中。

WITH中的数据修改语句

PostgreSQL允许将INSERT、UPDATE和DELETE用作WITH查询。SQL 标准中没有这种用法。

非标准子句

DISTINCT ON子句未在 SQL 标准中定义。

SELECT INTO

SELECT INTO — define a new table from the results of a query

大纲

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
    * | expression [ [ AS ] output_name ] [, ...]
    INTO [ TEMPORARY | TEMP | UNLOGGED ] [ TABLE ] new_table
    [ FROM from_item [, ...] ]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING condition [, ...] ]
    [ WINDOW window_name AS ( window_definition ) [, ...] ]
    [ { UNION | INTERSECT | EXCEPT } [ ALL | DISTINCT ] select ]
    [ ORDER BY expression [ ASC | DESC | USING operator ] [ NULLS {
 FIRST | LAST } ] [, ...] ]
    [ LIMIT { count | ALL } ]
    [ OFFSET start [ ROW | ROWS ] ]
    [ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ]
    [ FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ] [..
.] ]
```

描述

`SELECT INTO`创建一个新表，并用查询计算得到的数据填充该表。与普通的`SELECT`不同，这些数据不会返回给客户端。新表的列具有与`SELECT`输出列对应的名称和数据类型。

参数

`TEMPORARY` or `TEMP`

如果指定，该表将创建为临时表。详见CREATE TABLE。

`UNLOGGED`

如果指定，该表将创建为不记录 WAL 的表。详见CREATE TABLE。

`new_table`

要创建的表的名称（可选地带模式限定）。

所有其他参数都在SELECT中有详细说明。

注解

CREATE TABLE AS在功能上与 `SELECT INTO`类似。`CREATE TABLE AS` 是推荐使用的语法，因为这种形式的`SELECT INTO`在ECPG 或PL/pgSQL中不可用，因为它们对 `INTO`子句有不同的解释。此外，`CREATE TABLE AS`提供的功能是 `SELECT INTO`所提供功能的超集。

在 PostgreSQL 8.1 之前，`SELECT INTO` 创建的表默认包含 OID。而在 PostgreSQL 8.1 中不再如此——要在新表中包含 OID，必须启用default_with_oids配置变量。或者，也可以使用 `CREATE TABLE AS` 加上 `WITH OIDS` 子句。

示例

创建一个新表`films_recent`，它只包含表 `films`中的最近条目：

```
SELECT * INTO films_recent FROM films WHERE date_prod >= '2002-01-01';
```

兼容性

SQL 标准使用`SELECT INTO`表示将值选入宿主程序的标量变量，而不是创建一个新表。这确实是 ECPG（见第 33 章）和 PL/pgSQL（见第 39 章）中的用法。PostgreSQL使用`SELECT INTO`表示创建表则是历史遗留用法。对于新代码，最好为此目的使用`CREATE TABLE AS`。

另见

CREATE TABLE AS

SET

SET — change a run-time parameter

大纲

```
SET [ SESSION | LOCAL ] configuration_parameter { TO | = } { value |
 'value' | DEFAULT }
SET [ SESSION | LOCAL ] TIME ZONE { timezone | LOCAL | DEFAULT }
```

描述

SET命令用于更改运行时配置参数。许多在第 18 章中列出的运行时参数都可以通过 SET即时更改。（有些参数需要超级用户权限才能更改，
还有一些参数在服务器启动或会话开始之后便无法再更改。）SET只影响当前会话所使用的值。

如果在一个随后被中止的事务中发出SET（或者等效的SET SESSION），那么在事务回滚时，SET命令的效果就会消失。一旦该事务提交，这些效果就会持续到会话结束，除非被另一个SET覆盖。

SET LOCAL的效果只持续到当前事务结束，无论事务是否提交。一种特殊情况是：在同一个事务中先执行 SET，再执行SET LOCAL。这样SET LOCAL设置的值会一直生效到事务结束，但在此之后（如果该事务被提交），SET设置的值将生效。

SET或SET LOCAL的效果，也会因为回滚到早于该命令的保存点而被取消。

如果在一个针对同一变量带有SET选项的函数内使用 SET LOCAL（见CREATE FUNCTION），那么SET LOCAL命令的效果会在函数退出时消失；也就是说，无论如何都会恢复函数被调用时生效的值。这使得SET LOCAL 可以在函数内部用于动态或重复地更改某个参数，同时仍能方便地利用SET选项保存并恢复调用者的值。不过，普通的 SET命令会覆盖任何外围函数的SET选项；除非回滚，否则它的效果会一直保留。

> **注意**
>
> 在PostgreSQL 8.0 到 8.2 版本中，SET LOCAL的效果会因为释放一个更早的保存点，或者成功退出某个PL/pgSQL异常块而被取消。由于这种行为被认为不够直观，
> 因此后来对其进行了更改。

参数

SESSION

指定该命令对当前会话生效。（如果既没有出现SESSION，也没有出现LOCAL，
这就是默认情况。）

LOCAL

指定该命令只对当前事务生效。在COMMIT或 ROLLBACK之后，会话级设置会再次生效。注意，如果在 BEGIN 块之外执行，SET LOCAL 看起来不会有任何效果。

configuration_parameter

可设置的运行时参数名称。可用参数见第 18 章以及下文。

value

参数的新值。根据具体参数的不同，值可以指定为字符串常量、标识符、数字，
或由这些构成的逗号分隔列表。也可以写成DEFAULT，
表示把该参数重置为其默认值（即当前会话中如果从未执行过 SET，它本应具有的值）。

除了第 18 章中记载的配置参数外，还有几个参数只能通过SET命令调整，或者具有特殊语法：

SCHEMA

SET SCHEMA '*value*'是 SET search_path TO *value*的别名。使用这种语法时只能指定一个模式。

NAMES

SET NAMES *value*是 SET client_encoding TO *value*的别名。

SEED

设置随机数生成器（即函数 random）的内部种子。允许的值是 -1 到 1 之间的浮点数，这些值随后会乘以 2^{31}-1。

种子也可以通过调用函数setseed来设置：

SELECT setseed(*value*);

TIME ZONE

SET TIME ZONE *value*是以下写法的别名：SET timezone TO *value*。语法SET TIME ZONE允许使用特殊语法指定时区。下面是一些有效值的示例：

'PST8PDT'

加利福尼亚州伯克利所使用的时区。

'Europe/Rome'

意大利所使用的时区。

-7

比 UTC 向西 7 小时的时区（等同于 PDT）。正值表示位于 UTC 以东。

INTERVAL '-08:00' HOUR TO MINUTE

比 UTC 向西 8 小时的时区（等同于 PST）。

LOCAL
DEFAULT

把时区设置为你的本地时区（即服务器 timezone 的默认值；如果没有在任何地方显式设置过它，将是服务器操作系统默认的时区）。

See 第 8.5.3 节 for more information about time zones.

注解

函数`set_config`提供了等效功能；见第 9.24 节。此外，也可以对 `pg_settings` 系统视图执行 UPDATE，以实现与`SET`等效的操作。

示例

设置模式搜索路径：

```
SET search_path TO my_schema, public;
```

将日期风格设置为传统POSTGRES风格，并采用“日在月之前”的输入惯例：

```
SET datestyle TO postgres, dmy;
```

设置加利福尼亚州伯克利的时区：

```
SET TIME ZONE 'PST8PDT';
```

将时区设置为意大利所使用的时区：

```
SET TIME ZONE 'Europe/Rome';
```

兼容性

`SET TIME ZONE`扩展了 SQL 标准定义的语法。标准只允许使用数值形式的时区偏移量，而 PostgreSQL 允许更灵活的时区指定方式。其他所有`SET`特性都是 PostgreSQL扩展。

另见

RESET, SHOW

SET CONSTRAINTS

SET CONSTRAINTS — set constraint check timing for the current transaction

大纲

```
SET CONSTRAINTS { ALL | name [, ...] } { DEFERRED | IMMEDIATE }
```

描述

SET CONSTRAINTS设置当前事务中约束检查的行为。IMMEDIATE约束会在每条语句结束时检查。DEFERRED约束要等到事务提交时才检查。每个约束都有各自的IMMEDIATE或DEFERRED模式。

创建约束时，它会被赋予以下三种特性之一：DEFERRABLE INITIALLY DEFERRED、DEFERRABLE INITIALLY IMMEDIATE或者 NOT DEFERRABLE。第三类始终是 IMMEDIATE，不会受到 SET CONSTRAINTS命令的影响。前两类在每个事务开始时都处于其指定的模式，但其行为可以在事务中通过 SET CONSTRAINTS修改。

带约束名称列表的SET CONSTRAINTS只会修改这些约束的模式（它们都必须是可延迟的）。每个约束名称都可以带模式限定。如果未指定模式名称，
就会使用当前模式搜索路径查找第一个匹配的名称。SET CONSTRAINTS ALL会修改所有可延迟约束的模式。

当SET CONSTRAINTS将某个约束的模式从 DEFERRED改为IMMEDIATE时，新模式具有追溯效力：任何原本会在事务结束时才检查的未决数据修改，都会改为在执行SET CONSTRAINTS命令期间检查。如果违反了任何此类约束，SET CONSTRAINTS就会失败（并且不会改变该约束的模式）。因此，可以利用SET CONSTRAINTS强制在事务中的特定点执行约束检查。

当前，只有UNIQUE、PRIMARY KEY、REFERENCES（外键）以及EXCLUDE 约束受到这个设置的影响。NOT NULL和CHECK约束总是在一行被插入或修改时立即检查（不是在语句结束时）。未声明为DEFERRABLE的唯一约束和排他约束也会立即检查。

被声明为“约束触发器”的触发器，其引发也受此设置控制 – 它们会在相关约束应当被检查的同时引发。

注解

因为PostgreSQL并不要求约束名称在同一模式内唯一（只要求在每个表内唯一），
所以指定的约束名称有可能匹配到多个约束。在这种情况下，SET CONSTRAINTS会作用于所有匹配项。对于未带模式限定的名称，一旦在搜索路径中的某个模式里找到一个或多个匹配项，
就不会再搜索路径中更靠后的模式。

这个命令只会改变当前事务中约束的行为。因此，如果你在事务块（BEGIN/COMMIT对）
之外执行此命令，它将不会产生任何效果。

兼容性

这个命令符合 SQL 标准定义的行为，但有一个限制：在 PostgreSQL中，它不会应用在NOT NULL和CHECK约束上。此外，PostgreSQL会立即检查不可延迟的唯一约束，而不是像标准所暗示的那样在语句结束时检查。

SET ROLE

SET ROLE — set the current user identifier of the current session

大纲

```
SET [ SESSION | LOCAL ] ROLE role_name
SET [ SESSION | LOCAL ] ROLE NONE
RESET ROLE
```

描述

该命令把当前 SQL 会话的当前用户标识符设置为 *role_name*。角色名可以写成标识符或字符串字面量。在 SET ROLE 之后，SQL 命令的权限检查会视同指定角色是最初登录的角色。

当前会话用户必须是指定的 *role_name* 角色的成员。（如果会话用户是超级用户，则可以选择任意角色。）

SESSION 和 LOCAL 修饰符的作用与常规的SET命令相同。

NONE 和 RESET 形式会把当前用户标识符重置为当前会话用户标识符。这些形式可以由任何用户执行。

注解

使用这个命令，既可以增加权限，也可以限制自己的权限。如果会话用户角色具有 INHERIT 属性，它会自动拥有每个可通过 SET ROLE 切换到的角色的全部权限；在这种情况下，SET ROLE 实际上会去掉直接分配给会话用户及其所属其他角色的全部权限，只保留指定角色可用的权限。另一方面，如果会话用户角色具有 NOINHERIT 属性，SET ROLE 会放弃直接分配给会话用户的权限，转而获得指定角色可用的权限。

特别是，当超级用户选择通过 SET ROLE 切换到非超级用户角色时，会失去其超级用户权限。

SET ROLE的效果与SET SESSION AUTHORIZATION相近，但其中涉及的权限检查完全不同。此外，SET SESSION AUTHORIZATION会决定后续 SET ROLE命令允许使用哪些角色，而通过 SET ROLE更改角色并不会改变后续 SET ROLE允许使用的角色集合。

SET ROLE不会处理角色的ALTER ROLE设置所指定的会话变量；这只会在登录期间发生。

SET ROLE不能在 SECURITY DEFINER函数中使用。

示例

```
SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 peter        | peter

SET ROLE 'paul';
```

```
SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 peter        | paul
```

兼容性

PostgreSQL允许使用标识符语法（`"rolename"`），而 SQL 标准要求将角色名写成字符串字面量。SQL 不允许在事务中执行这个命令；PostgreSQL不施加这一限制，因为没有理由这样做。`SESSION`和`LOCAL`修饰符，以及 `RESET`语法，都是PostgreSQL扩展。

另见

SET SESSION AUTHORIZATION

SET SESSION AUTHORIZATION

SET SESSION AUTHORIZATION — set the session user identifier and the current user identifier of the current session

大纲

```
SET [ SESSION | LOCAL ] SESSION AUTHORIZATION user_name
SET [ SESSION | LOCAL ] SESSION AUTHORIZATION DEFAULT
RESET SESSION AUTHORIZATION
```

描述

这个命令将当前 SQL 会话的会话用户标识符和当前用户标识符设置为 *user_name*。用户名可以写成一个标识符或者一个字符串字面量。例如，可以使用这个命令临时变为一个非特权用户，之后再切换回超级用户。

会话用户标识符初始时被设置为客户端提供的（可能已认证的）用户名。
当前用户标识符通常与会话用户标识符相同，但可能在 SECURITY DEFINER函数和类似机制的上下文中临时更改；它也可以通过SET ROLE更改。当前用户标识符用于权限检查。

只有当初始会话用户（即已认证用户）最初具有超级用户权限时，才能更改会话用户标识符。
否则，只有当该命令指定的是已认证用户名时，才会被接受。

SESSION 和 LOCAL 修饰符的作用与常规的SET命令相同。

DEFAULT和RESET形式会将会话用户标识符和当前用户标识符重置为最初已认证的用户名。
这些形式可以由任何用户执行。

注解

SET SESSION AUTHORIZATION不能在一个 SECURITY DEFINER函数中使用。

示例

```
SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 peter        | peter

SET SESSION AUTHORIZATION 'paul';

SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 paul         | paul
```

兼容性

SQL 标准允许在字面值*user_name*的位置上使用某些其他表达式，但这些选项在实践中并不重要。PostgreSQL允许使用标识符语法（"username"），而 SQL 标准不允许。SQL 不允许在事

务中使用这个命令；PostgreSQL并不做此限制，因为没有理由这样做。`SESSION`和`LOCAL`修饰符以及`RESET`语法都是PostgreSQL扩展。

标准把执行这个命令所需的权限留给实现定义。

另见

SET ROLE

SET TRANSACTION

SET TRANSACTION — set the characteristics of the current transaction

大纲

```
SET TRANSACTION transaction_mode [, ...]
SET SESSION CHARACTERISTICS AS TRANSACTION transaction_mode [, ...]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
    [ NOT ] DEFERRABLE
```

描述

SET TRANSACTION命令设置当前事务的特性。它对任何后续事务都没有影响。SET SESSION CHARACTERISTICS设置一个会话中后续事务的默认事务特性。对于单个事务，这些默认值可以用 SET TRANSACTION覆盖。

可用的事务特性包括事务隔离级别、事务访问模式（读/写或只读）以及可延迟模式。

一个事务的隔离级别决定当其他事务并发运行时该事务能看见什么数据：

READ COMMITTED

一条语句只能看到在它开始之前已提交的行。这是默认值。

REPEATABLE READ

当前事务中的所有语句都只能看到在该事务中执行第一条查询或数据修改语句之前已提交的行。

SERIALIZABLE

当前事务中的所有语句只能看到在该事务中执行第一条查询或数据修改语句之前已提交的行。如果一组并发可串行化事务之间的读写模式会造成任何串行（一次一个）执行都不可能出现的情况，其中一个事务将带着 serialization_failure SQLSTATE 被回滚。

The SQL standard defines one additional level, READ UNCOMMITTED. In PostgreSQL READ UNCOMMITTED is treated as READ COMMITTED.

在事务执行第一条查询或数据修改语句（SELECT, INSERT, DELETE, UPDATE, FETCH, 或 COPY）之后，事务隔离级别就不能再更改。有关事务隔离和并发控制的更多信息，请参见第 13 章。

事务访问模式决定事务是读/写还是只读。读/写是默认值。当事务为只读时，下列 SQL 命令会被禁止：INSERT、UPDATE、DELETE 以及 COPY FROM，前提是它们要写入的表不是临时表；所有 CREATE、ALTER 和 DROP 命令；COMMENT、GRANT、REVOKE、TRUNCATE；以及

`EXPLAIN ANALYZE` 和 `EXECUTE`，前提是它们要执行的命令属于上述列表。这是一个较高层面的只读概念，并不会阻止所有写入磁盘的行为。

除非事务同时是 `SERIALIZABLE` 且 `READ ONLY` 的，否则 `DEFERRABLE` 事务属性没有任何效果。When all of these properties are set on a transaction, the transaction may block when first acquiring its snapshot, 之后它就可以在不承担 `SERIALIZABLE` 事务正常开销、也不会导致或因序列化失败被取消的任何风险的情况下运行。这种模式非常适合长时间运行的报告或备份。

注解

如果在没有先执行 `START TRANSACTION` 或 `BEGIN` 的情况下执行 `SET TRANSACTION`，它将看起来没有任何效果，因为事务会立即结束。

也可以不使用 `SET TRANSACTION`，而是在 `BEGIN` 或 `START TRANSACTION` 中直接指定所需的 *`transaction_modes`*。

会话默认的事务模式也可以通过配置参数default_transaction_isolation、default_transaction_read_only和default_transaction_deferrable来设置（实际上，`SET SESSION CHARACTERISTICS`只是用 `SET`设置这些变量的一种更冗长的等价写法）。这意味着可以通过配置文件、`ALTER DATABASE`等方式设置默认值。详见第 18 章。

兼容性

这两个命令都在 SQL 标准中有定义。标准中的默认事务隔离级别是 `SERIALIZABLE`。在 PostgreSQL 中默认通常是 `READ COMMITTED`，但可以按上文所述更改。

在 SQL 标准中，还可以用这些命令设置另一项事务特性：诊断区域的大小。这个概念特定于嵌入式 SQL，因此没有在 PostgreSQL服务器中实现。

`DEFERRABLE` *`transaction_mode`* 是 PostgreSQL 的语言扩展。

SQL 标准要求在连续的*`transaction_modes`*之间有逗号，但出于历史原因，PostgreSQL允许省略逗号。

SHOW

SHOW — show the value of a run-time parameter

大纲

```
SHOW name
SHOW ALL
```

描述

SHOW将显示运行时参数的当前设置。这些变量可以使用SET语句来设置，也可以通过编辑postgresql.conf配置文件、使用 PGOPTIONS环境变量（在使用 libpq或基于libpq的应用时），或者在启动postgres服务器时通过命令行标志来设置。详见第 18 章。

参数

name

运行时参数的名称。可用参数的说明见第 18 章以及SET参考页。此外，还有几个可以显示但不能设置的参数：

SERVER_VERSION

显示服务器的版本号。

SERVER_ENCODING

显示服务器端字符集编码。目前，该参数可以显示但不能设置，因为编码在数据库创建时就已确定。

LC_COLLATE

显示数据库用于排序规则（文本排序）的区域设置。目前，该参数可以显示但不能设置，因为该设置在数据库创建时就已确定。

LC_CTYPE

显示数据库用于字符分类的区域设置。目前，该参数可以显示但不能设置，因为该设置在数据库创建时就已确定。

IS_SUPERUSER

如果当前角色具有超级用户权限，则为真。

ALL

显示所有配置参数的值及其描述。

注解

函数current_setting也会产生等效的输出，见第 9.24 节。另外，pg_settings 系统视图也会给出相同的信息。

示例

显示参数DateStyle的当前设置：

```
SHOW DateStyle;
 DateStyle
-----------
 ISO, MDY
(1 row)
```

显示参数geqo的当前设置：

```
SHOW geqo;
 geqo
------
 on
(1 row)
```

显示所有设置：

```
SHOW ALL;
            name          | setting |                 description
--------------------------+---------+-----------------------------------
---------------
 allow_system_table_mods | off     | Allows modifications of the
 structure of ...
    .
    .
    .
 xmloption               | content | Sets whether XML data in implicit
 parsing ...
 zero_damaged_pages      | off     | Continues processing past damaged
 page headers.
(196 rows)
```

兼容性

SHOW命令是一种 PostgreSQL扩展。

另见

SET, RESET

START TRANSACTION

START TRANSACTION — start a transaction block

大纲

```
START TRANSACTION [ transaction_mode [, ...] ]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
    [ NOT ] DEFERRABLE
```

描述

该命令开始一个新的事务块。如果指定了隔离级别、读写模式或可延迟模式，
新事务就会具有这些特性，就像执行了SET TRANSACTION一样。它与BEGIN命令具有相同功能。

参数

关于本语句参数含义的信息，参见SET TRANSACTION。

兼容性

在 SQL 标准中，并不需要发出`START TRANSACTION`来开始一个事务块：任何 SQL 命令都会隐式开始一个事务块。PostgreSQL的行为可以看作是：对每条不跟在`START TRANSACTION`（或`BEGIN`）之后的命令，都会在其后隐式发出一个`COMMIT`，因此这种行为通常被称为“自动提交”。其他关系型数据库系统也可能为了方便而提供自动提交特性。

`DEFERRABLE` *`transaction_mode`* 是PostgreSQL的一种语言扩展。

SQL 标准要求在连续的*`transaction_modes`*之间有逗号，但出于历史原因，PostgreSQL允许省略逗号。

另见SET TRANSACTION的兼容性小节。

另见

BEGIN, COMMIT, ROLLBACK, SAVEPOINT, SET TRANSACTION

TRUNCATE

TRUNCATE — empty a table or set of tables

大纲

```
TRUNCATE [ TABLE ] [ ONLY ] name [ * ] [, ... ]
    [ RESTART IDENTITY | CONTINUE IDENTITY ] [ CASCADE | RESTRICT ]
```

描述

TRUNCATE可以快速移除一组表中的所有行。它的效果与对每个表执行不带条件的DELETE相同，但由于它并不实际扫描这些表，因此速度更快。此外，它会立即回收磁盘空间，而不是要求后续再执行一次VACUUM操作。这一点对于大表尤其有用。

参数

name

要截断的表名（可以是模式限定的）。如果在表名前指定了 ONLY，则只截断该表。如果未指定ONLY，则该表及其所有后代表（如果有）都会被截断。也可以在表名后指定 *，以显式表明包含后代表。

RESTART IDENTITY

自动重新启动由被截断表列拥有的 sequence。

CONTINUE IDENTITY

不更改 sequence 的值。这是默认值。

CASCADE

自动截断所有以外键引用任一已命名表的表，以及任何由于 CASCADE而被加入该组的表。

RESTRICT

如果任一表具有来自命令中未列出表的外键引用，则拒绝截断。这是默认值。

注解

要截断一个表，你必须拥有该表上的TRUNCATE权限。

TRUNCATE会在其操作的每个表上获取 ACCESS EXCLUSIVE锁，这会阻塞该表上的所有其他并发操作。当指定RESTART IDENTITY时，任何需要重新启动的 sequence 也会同样被排他锁定。如果需要对某个表进行并发访问，则应改用 DELETE命令。

TRUNCATE不能用于被其他表通过外键引用的表，除非所有这类表也在同一条命令中被截断。因为在这种情况下检查有效性将需要扫描表，而该命令的意义恰恰在于避免这样做。CASCADE 选项可用于自动包含所有依赖表 – 但使用该选项时一定要非常小心，否则你可能会丢失并非有意删除的数据！

TRUNCATE不会引发这些表上可能存在的任何 ON DELETE触发器，但会引发 ON TRUNCATE触发器。如果这些表中的任何一个定义了ON TRUNCATE触发器，那么所有 BEFORE TRUNCATE触发器都会在任何截断发生之前引发，而所有AFTER TRUNCATE触发器都会在最后一次截断完成且所有 sequence 被重置之后引发。触发器会按照表的处理顺序引发（先是命令中列出的表，然后是由于级联而加入的表）。

TRUNCATE不是多版本并发控制（MVCC）安全的。截断之后，
如果并发事务使用的是在截断发生前取得的快照，该表对这些并发事务而言将表现为空。
详见第 13.5 节。

就表中数据而言，TRUNCATE是事务安全的：如果外围事务没有提交，截断将被安全地回滚。

当指定RESTART IDENTITY时，隐含的 ALTER SEQUENCE RESTART操作也会以事务方式执行；也就是说，如果外围事务没有提交，它们也会被回滚。这与 ALTER SEQUENCE RESTART通常的行为不同。请注意，如果在事务回滚前又对这些重启后的 sequence 执行了额外操作，这些操作对 sequence 本身的影响会被回滚，但对currval()的影响不会被回滚。也就是说，事务结束后，currval()仍将反映失败事务内取得的最后一个 sequence 值，即使 sequence 本身可能已经不再与之保持一致。这与失败事务之后currval()的通常行为类似。

示例

截断表bigtable和 fattable：

```
TRUNCATE bigtable, fattable;
```

同样的操作，并重置所有相关联的 sequence 生成器：

```
TRUNCATE bigtable, fattable RESTART IDENTITY;
```

截断表othertable，并级联到任何通过外键约束引用othertable的表：

```
TRUNCATE othertable CASCADE;
```

兼容性

SQL:2008 标准包括了一个TRUNCATE命令，语法是TRUNCATE TABLE *tablename*。子句 CONTINUE IDENTITY/RESTART IDENTITY 在该标准中也有出现，但其含义虽相关却略有不同。该命令的某些并发行为在标准中被留作实现定义，因此必要时应结合上述注解并与其他实现进行比较。

UNLISTEN

UNLISTEN — stop listening for a notification

大纲

```
UNLISTEN { channel | * }
```

描述

UNLISTEN用于移除现有的 NOTIFY事件注册。UNLISTEN会取消当前 PostgreSQL会话作为名为*channel*的通知通道监听者的任何现有注册。特殊通配符*会取消当前会话的所有监听注册。

NOTIFY中对LISTEN和 NOTIFY的使用有更详细的讨论。

参数

channel

通知通道的名称（任意标识符）。

*

该会话当前的所有监听注册都会被清除。

注解

可以取消监听自己并未监听的通道，不会出现警告或错误。

每个会话结束时，都会自动执行UNLISTEN *。

执行过UNLISTEN的事务不能为两阶段提交做准备。

示例

建立一个监听注册：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
```

执行UNLISTEN之后，后续的NOTIFY 消息将被忽略：

```
UNLISTEN virtual;
NOTIFY virtual;
-- no NOTIFY event is received
```

兼容性

SQL 标准中没有UNLISTEN语句。

另见

LISTEN, NOTIFY

UPDATE

UPDATE — update rows of a table

大纲

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
UPDATE [ ONLY ] table [ * ] [ [ AS ] alias ]
    SET { column = { expression | DEFAULT } |
          ( column [, ...] ) = ( { expression | DEFAULT } [, ...] ) }
 [, ...]
    [ FROM from_list ]
    [ WHERE condition | WHERE CURRENT OF cursor_name ]
    [ RETURNING * | output_expression [ [ AS ] output_name ] [, ...] ]
```

描述

UPDATE会更改所有满足条件的行中指定列的值。只需在SET子句中提及需要修改的列；未被显式修改的列会保留其原有值。

有两种方法可以利用数据库中其他表所包含的信息来修改一个表：使用子选择，或者在FROM子句中指定附加表。哪种技术更合适取决于具体情况。

可选的RETURNING子句使UPDATE 基于每个实际更新的行计算并返回一个或多个值。可以计算任何使用该表列和/或 FROM中提到的其他表列的表达式。默认使用该表列的新值（更新后值），但也可以请求旧值（更新前值）。RETURNING列表的语法与SELECT 的输出列表相同。

必须拥有该表上的UPDATE权限，或者至少拥有所列待更新列的该权限。还必须拥有SELECT权限，覆盖在*expressions*或*condition*中读取其值的所有列。

参数

with_query

WITH子句允许你指定一个或多个子查询，这些子查询可在UPDATE查询中按名称引用。详见第 7.8 节和SELECT。

table

要更新的表名（可以是模式限定的）。如果在表名前指定了 ONLY，只会更新所提及表中的匹配行。如果未指定 ONLY，还会更新继承自该表的任何表中的匹配行。可选地，可以在表名后指定*，以显式指示包含后代表。

alias

目标表的替代名称。提供别名时，它会完全隐藏该表的实际名称。例如，给定UPDATE foo AS f，UPDATE语句的其余部分必须将该表称为 f，而不是foo。

column

由*table_name*命名的表中的列名。如有需要，列名可以使用子字段名或数组下标进行限定。指定目标列时不要包含表名 — 例如，UPDATE tab SET tab.col = 1是无效的。

expression

要赋给该列的表达式。该表达式可以使用该表中这一列和其他列的旧值。

DEFAULT

把该列设置为其默认值（如果没有为它指定特定的默认表达式，则该值为 NULL）。

from_list

一个表表达式列表，允许其他表的列出现在WHERE条件以及更新表达式中。这类似于SELECT语句的FROM Clause中可以指定的表列表。注意，除非打算进行自连接（此时它必须在*from_list*中带有别名出现），否则目标表不能出现在*from_list*中。

condition

一个返回boolean类型值的表达式。只有使这个表达式返回true的行才会被更新。

cursor_name

要在WHERE CURRENT OF条件中使用的游标名。要被更新的行是最近一次从该游标中取出的那一行。该游标必须是针对UPDATE目标表的非分组查询。注意，WHERE CURRENT OF不能与布尔条件同时指定。参见DECLARE，了解将游标用于WHERE CURRENT OF的更多信息。

output_expression

在每一行被更新后，由UPDATE命令计算并返回的表达式。该表达式可以使用由*table_name*命名的表或FROM中列出的表中的任何列名。写成 *可返回所有列。

output_name

用于返回列的名称。

输出

成功完成后，UPDATE 命令返回如下形式的命令标签：

UPDATE *count*

The *count* is the number of rows updated. If *count* is 0, no rows matched the *condition* (this is not considered an error).

如果UPDATE命令包含RETURNING 子句，其结果将类似于一个SELECT语句，其中包含RETURNING列表中定义的列和值，并在该命令更新的行上进行计算。

注解

当存在FROM子句时，本质上会将目标表与 *from_list*中提到的表连接起来，而连接的每一条输出行都代表对目标表的一次更新操作。使用 FROM时，应确保对每一个要修改的目标行，连接至多生成一条输出行。换言之，一条目标行不应与其他表中的多于一行成功连接。
如果发生这种情况，则只会使用其中某一条连接行来更新目标行，
但具体使用哪一条并不容易预测。

因为存在这种不确定性，所以仅在子选择中引用其他表会更安全，
尽管这种写法通常比使用连接更难阅读、也更慢。

示例

把表`films`的`kind`列中的单词`Drama`改为`Dramatic`：

```
UPDATE films SET kind = 'Dramatic' WHERE kind = 'Drama';
```

在表`weather`的一行中调整温度项，并将降水量重置为其默认值：

```
UPDATE weather SET temp_lo = temp_lo+1, temp_hi = temp_lo+15, prcp =
 DEFAULT
  WHERE city = 'San Francisco' AND date = '2003-07-03';
```

执行同样的操作，并返回更新后的各项：

```
UPDATE weather SET temp_lo = temp_lo+1, temp_hi = temp_lo+15, prcp =
 DEFAULT
  WHERE city = 'San Francisco' AND date = '2003-07-03'
  RETURNING temp_lo, temp_hi, prcp;
```

使用另一种列列表语法完成同样的更新：

```
UPDATE weather SET (temp_lo, temp_hi, prcp) = (temp_lo+1, temp_lo+15,
 DEFAULT)
  WHERE city = 'San Francisco' AND date = '2003-07-03';
```

将负责 Acme Corporation 账户的销售人员的销量计数加一，使用 `FROM` 子句语法：

```
UPDATE employees SET sales_count = sales_count + 1 FROM accounts
  WHERE accounts.name = 'Acme Corporation'
  AND employees.id = accounts.sales_person;
```

执行同样的操作，在 `WHERE` 子句中使用子选择：

```
UPDATE employees SET sales_count = sales_count + 1 WHERE id =
  (SELECT sales_person FROM accounts WHERE name = 'Acme Corporation');
```

尝试插入一个新库存项及其库存量。如果该项已存在，则改为更新现有项的库存量。
要在不致使整个事务失败的情况下做到这一点，请使用保存点：

```
BEGIN;
-- 其他操作
SAVEPOINT sp1;
INSERT INTO wines VALUES('Chateau Lafite 2003', '24');
-- 假设上面的语句因唯一键冲突而失败，
-- 那么现在发出这些命令：
ROLLBACK TO sp1;
UPDATE wines SET stock = stock + 24 WHERE winename = 'Chateau Lafite
 2003';
-- 继续执行其他操作，最后
COMMIT;
```

更改 `kind` 列，该列位于表 `films` 中，所更改的行是游标 `c_films` 当前定位的行：

```
UPDATE films SET kind = 'Dramatic' WHERE CURRENT OF c_films;
```

兼容性

这个命令符合SQL标准，不过 FROM和RETURNING子句是 PostgreSQL扩展，把WITH 与UPDATE一起使用的能力也是扩展。

按照标准，列列表语法应允许用一个单一的行值表达式（例如子查询）为一组列赋值：

```
UPDATE accounts SET (contact_last_name, contact_first_name) =
    (SELECT last_name, first_name FROM salesmen
     WHERE salesmen.id = accounts.sales_id);
```

这一点目前尚未实现–来源必须是独立表达式的列表。

某些其他数据库系统提供一种 FROM 选项，其中目标表被认为要在 FROM 中再次列出。PostgreSQL 不是这样解释 FROM 的。移植使用这一扩展的应用时要小心。

VACUUM

VACUUM — garbage-collect and optionally analyze a database

大纲

```
VACUUM [ ( { FULL | FREEZE | VERBOSE | ANALYZE } [, ...] ) ] [ table [
 (column [, ...] ) ] ]
VACUUM [ FULL ] [ FREEZE ] [ VERBOSE ] [ table ]
VACUUM [ FULL ] [ FREEZE ] [ VERBOSE ] ANALYZE [ table [ (column [, ..
.] ) ] ]
```

描述

VACUUM回收死元组占用的存储空间。在正常的 PostgreSQL运行中，被删除或因更新而过时的元组并不会从其表中物理移除；它们会一直保留，直到执行 VACUUM。因此有必要定期执行VACUUM，尤其是对频繁更新的表。

不带参数时，VACUUM会处理当前数据库中当前用户有权清理的每个表。带参数时，VACUUM则只处理该表。

VACUUM ANALYZE会对每个选定表先执行 VACUUM，再执行ANALYZE。这种便捷的组合形式很适合例行维护脚本使用。关于其处理细节，参见ANALYZE。

普通的VACUUM（不带FULL）只是回收空间并使其可被重用。
这种形式的命令可以与表的正常读写并行运行，因为它不会获得独占锁。不过，
在大多数情况下，额外空间不会返还给操作系统；它只是保留在同一张表内供再次使用。VACUUM FULL会把表的全部内容重写到一个没有额外空闲空间的新磁盘文件中，
从而让未使用的空间能够返还给操作系统。这种形式要慢得多，
并且在处理每个表时都需要独占锁。

当选项列表用圆括号括起来时，选项可以按任意顺序书写。如果不加圆括号，
则必须严格按照上面所示的顺序指定选项。带圆括号的语法是在PostgreSQL 9.0 中加入的；不带圆括号的语法已被弃用。

参数

FULL

选择“完全”清理，它可以回收更多空间，但耗时更长，并且会独占锁定该表。
这种方法还需要额外的磁盘空间，因为它会写出该表的一个新副本，
并且在操作完成之前不会释放旧副本。通常，
只有当需要从表内回收大量空间时才应使用这种方法。

FREEZE

选择激进的元组“冻结”。指定FREEZE等价于在vacuum_freeze_min_age参数设置为零的情况下执行VACUUM。

VERBOSE

为每个表输出详细的清理活动报告。

ANALYZE

更新规划器用来确定查询最高效执行方式的统计信息。

table

要清理的特定表的名称（可选地带模式限定）。默认为当前数据库中的所有表。

column

要分析的特定列名。默认会分析所有列。如果指定了列列表，则隐含指定ANALYZE。

输出

指定VERBOSE时，VACUUM会输出进度消息，指示当前正在处理哪个表，
同时还会打印这些表的各种统计信息。

注解

要清理一个表，通常调用者必须是该表的拥有者或超级用户。不过，
数据库拥有者可以清理其数据库中的所有表，但共享系统目录除外。
（对共享系统目录的这一限制意味着，真正意义上的全数据库 VACUUM 只能由超级用户执行。）
VACUUM 会跳过调用用户无权清理的任何表。

VACUUM不能在一个事务块内被执行。

对于带有GIN索引的表，VACUUM（任何形式）还会通过把待处理的索引项移动到主 GIN索引结构中的适当位置，来完成所有挂起的索引插入。详见第 54.3.1 节。

我们建议对活跃的生产数据库频繁执行清理（至少每晚一次），以移除死行。
在添加或删除大量行之后，对受影响的表执行VACUUM ANALYZE命令可能是个好主意。这会用所有近期变更的结果更新系统目录，使PostgreSQL查询规划器在规划查询时能够做出更好的选择。

FULL选项不建议在日常场景中使用，但在某些特殊情况下可能很有用。例如，
当你删除或更新了表中的绝大多数行，并希望该表在物理上收缩以占用更少磁盘空间、
让表扫描更快时，这个选项就比较合适。VACUUM FULL通常会比普通 VACUUM更大幅度地收缩表。

VACUUM会显著增加 I/O 流量，这可能导致其他活动会话性能变差。因此，有时建议使用基于代价的清理延迟特性。详见第 18.4.3 节。

PostgreSQL提供了一个“自动清理（autovacuum）”机制，可以自动执行常规清理维护。
有关自动与手动清理的更多信息，参见第 23.1 节。

示例

下面是对回归数据库中的一个表运行 VACUUM 的例子：

```
regression=# VACUUM (VERBOSE, ANALYZE) onek;
INFO:  vacuuming "public.onek"
INFO:  index "onek_unique1" now contains 1000 tuples in 14 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.08u sec elapsed 0.18 sec.
```

```
INFO:  index "onek_unique2" now contains 1000 tuples in 16 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.00s/0.07u sec elapsed 0.23 sec.
INFO:  index "onek_hundred" now contains 1000 tuples in 13 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.08u sec elapsed 0.17 sec.
INFO:  index "onek_stringu1" now contains 1000 tuples in 48 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.09u sec elapsed 0.59 sec.
INFO:  "onek": removed 3000 tuples in 108 pages
DETAIL:  CPU 0.01s/0.06u sec elapsed 0.07 sec.
INFO:  "onek": found 3000 removable, 1000 nonremovable tuples in 143
 pages
DETAIL:  0 dead tuples cannot be removed yet.
There were 0 unused item pointers.
0 pages are entirely empty.
CPU 0.07s/0.39u sec elapsed 1.56 sec.
INFO:  analyzing "public.onek"
INFO:  "onek": 36 pages, 1000 rows sampled, 1000 estimated total rows
VACUUM
```

兼容性

在SQL标准中没有VACUUM语句。

另见

vacuumdb, 第 18.4.3 节, 第 23.1.5 节

VALUES

VALUES — compute a set of rows

大纲

```
VALUES ( expression [, ...] ) [, ...]
    [ ORDER BY sort_expression [ ASC | DESC | USING operator ] [, ...
] ]
    [ LIMIT { count | ALL } ]
    [ OFFSET start [ ROW | ROWS ] ]
    [ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ]
```

描述

VALUES计算由值表达式指定的一个行值或一组行值。
它最常用于在更大的命令中生成一个“常量表”，但也可以单独使用。

当指定多于一行时，所有行都必须具有相同数量的元素。
结果表各列的数据类型由出现在该列中的表达式的显式类型或推断类型组合决定，所用规则与UNION相同（见第 10.5 节）。

在更大的命令中，从语法上说，凡是允许出现SELECT的地方，也都允许出现VALUES。
因为在语法中它被当作 SELECT处理，所以可以对VALUES 命令使用ORDER BY、LIMIT（或者等效的FETCH FIRST）以及OFFSET子句。

参数

expression

要在结果表（行集合）中相应位置计算并插入的常量或表达式。在出现在INSERT顶层的VALUES列表中，*expression*可以用 DEFAULT替代，以表示应插入目标列的默认值。当VALUES出现在其他上下文中时，不能使用 DEFAULT。

sort_expression

一个指示结果行如何排序的表达式或整数常量。该表达式可以将VALUES
结果的列引用为column1、column2等。详细信息参见ORDER BY Clause。

operator

一个排序操作符。详细信息参见ORDER BY Clause。

count

返回的最大行数。详细信息参见LIMIT Clause。

start

开始返回行之前要跳过的行数。详细信息参见LIMIT Clause。

注解

应避免使用行数非常多的VALUES列表，因为这可能导致内存不足错误或性能不佳。出现在INSERT中的 VALUES属于一种特殊情况（因为所需的列类型可以从INSERT的目标表得知，并且无须通过扫描 VALUES列表来推断），因此它能处理比其他上下文中实际可行的更大的列表。

示例

一个独立的VALUES命令：

```
VALUES (1, 'one'), (2, 'two'), (3, 'three');
```

这将返回一个两列三行的表。它实际上等效于：

```
SELECT 1 AS column1, 'one' AS column2
UNION ALL
SELECT 2, 'two'
UNION ALL
SELECT 3, 'three';
```

更常见的是，VALUES会用在更大的 SQL 命令中。最常见的用法是在INSERT中：

```
INSERT INTO films (code, title, did, date_prod, kind)
    VALUES ('T_601', 'Yojimbo', 106, '1961-06-16', 'Drama');
```

在INSERT的上下文中，VALUES列表中的项可以写成DEFAULT，表示此处应使用该列的默认值，而不是显式指定一个值：

```
INSERT INTO films VALUES
    ('UA502', 'Bananas', 105, DEFAULT, 'Comedy', '82 minutes'),
    ('T_601', 'Yojimbo', 106, DEFAULT, 'Drama', DEFAULT);
```

VALUES也可以用在原本可以编写子SELECT的地方，例如在FROM子句中：

```
SELECT f.*
  FROM films f, (VALUES('MGM', 'Horror'), ('UA', 'Sci-Fi')) AS t (
studio, kind)
  WHERE f.studio = t.studio AND f.kind = t.kind;

UPDATE employees SET salary = salary * v.increase
  FROM (VALUES(1, 200000, 1.2), (2, 400000, 1.4)) AS v (depno, target,
 increase)
  WHERE employees.depno = v.depno AND employees.sales >= v.target;
```

注意，在FROM子句中使用VALUES时，需要有一个AS子句，这一点与SELECT相同。虽然AS子句不要求为所有列都指定名称，但这样做是一个好习惯。（在PostgreSQL中，VALUES的默认列名是 column1、column2等，但在其他数据库系统中，这些名称可能不同。）

当在INSERT中使用VALUES时，所有值都会自动转换为相应目标列的数据类型。当它用于其他上下文时，可能有必要指定正确的数据类型。如果各项都是带引号的字符串常量，只需转换第一项，就足以为所有项确定假定的数据类型：

```
SELECT * FROM machines
WHERE ip_address IN (VALUES('192.168.0.1'::inet), ('192.168.0.10'), (
'192.168.1.43'));
```

> **提示**
>
> 对于简单的`IN`测试，最好使用`IN`的标量列表形式，而不是写一个像上面那样的`VALUES`查询。标量列表方法写法更简洁，而且通常效率更高。

兼容性

`VALUES`符合 SQL 标准。`LIMIT`和`OFFSET`是 PostgreSQL扩展，另见SELECT。

另见

INSERT, SELECT

PostgreSQL 客户端应用

本部分包含PostgreSQL客户端应用和工具的参考信息。并非所有这些命令都适用于一般用途；有些可能需要特殊权限。这些应用的共同点是，它们都可以在任意主机上运行，而不受数据库服务器所在位置的限制。

目录

clusterdb 1368
createdb 1371
createlang 1374
createuser 1377
dropdb 1381
droplang 1383
dropuser 1386
ecpg 1389
pg_basebackup 1391
pg_config 1395
pg_dump 1398
pg_dumpall 1407
pg_restore 1412
psql 1419
reindexdb 1448
vacuumdb 1451

clusterdb

clusterdb — cluster a PostgreSQL database

大纲

`clusterdb` [*connection-option*...] [[--verbose] | [-v]] [--table | -t *table*] [*dbname*]

`clusterdb` [*connection-option*...] [[--verbose] | [-v]] [[--all] | [-a]]

描述

clusterdb是一个用于重新聚簇PostgreSQL 数据库中各表的工具。它会找出此前已经聚簇过的表，并按照上次使用的同一索引再次对其进行聚簇。从未聚簇过的表不会受到影响。

clusterdb是 SQL 命令CLUSTER的一个包装器。使用该工具聚簇数据库，与使用其他访问服务器的方法聚簇数据库，并没有实质区别。

选项

clusterdb接受以下命令行参数：

`-a`
`--all`

聚簇所有数据库。

`[-d] dbname`
`[--dbname=]dbname`

指定要聚簇的数据库名。如果未指定并且没有使用 `-a`（或`--all`），则从环境变量 `PGDATABASE` 读取数据库名。如果该环境变量也未设置，则使用为该连接指定的用户名作为数据库名。

`-e`
`--echo`

回显clusterdb生成并发送到服务器的命令。

`-q`
`--quiet`

不显示进度消息。

`-t table`
`--table=table`

只对 *table* 执行聚簇。

`-v`
`--verbose`

不显示进度消息。

`-V`
`--version`

打印clusterdb的版本并退出。

`-?`
`--help`

显示有关clusterdb命令行参数的帮助信息，并退出。

clusterdb还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host=`*`host`*

指定运行服务器的机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

用于连接的用户名。

`-w`
`--no-password`

绝不发出密码提示。如果服务器要求密码认证，而又无法通过 `.pgpass` 文件等其他方式获得密码，则连接尝试将失败。这个选项在批处理作业和脚本中很有用，
因为这些场景下没有用户可以输入密码。

`-W`
`--password`

强制clusterdb在连接到数据库之前提示输入密码。

这个选项并非必不可少，因为如果服务器要求密码认证，clusterdb会自动提示输入密码。
不过，clusterdb会浪费一次连接尝试，才知道服务器需要密码。在某些情况下，
为了避免这次额外的连接尝试，提前指定`-W`是值得的。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

该工具和大多数其他PostgreSQL工具一样，也使用 libpq支持的环境变量（见第 31.13 节）。

诊断

如果遇到问题，请参见CLUSTER和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须在目标主机上运行。此外，
libpq前端库所使用的任何默认连接设置和环境变量都会生效。

示例

要聚簇数据库`test`：

```
$ clusterdb test
```

要聚簇名为`xyzzy`的数据库中的单个表`foo`：

```
$ clusterdb --table foo xyzzy
```

参见

CLUSTER

createdb

createdb — create a new PostgreSQL database

大纲

`createdb` [*connection-option*...] [*option*...] [*dbname*] [*description*]

描述

createdb创建一个新的PostgreSQL数据库。

通常，执行此命令的数据库用户会成为新数据库的所有者。不过，
如果执行用户具有相应的权限，也可以通过`-O`选项指定不同的所有者。

createdb是SQL命令CREATE DATABASE的包装器。通过此工具创建数据库与通过其他访问服务器的方法创建数据库，在效果上没有区别。

选项

createdb接受以下命令行参数：

dbname

指定要创建的数据库名称。该名称必须在此数据库集簇中的所有PostgreSQL数据库中唯一。默认情况下，会创建一个与当前系统用户同名的数据库。

description

指定要与新创建的数据库关联的注释。

`-D` *tablespace*
`--tablespace=`*tablespace*

指定数据库的默认表空间。

`-e`
`--echo`

回显createdb生成并发送到服务器的命令。

`-E` *encoding*
`--encoding=`*encoding*

指定该数据库使用的字符集编码。PostgreSQL服务器支持的字符集见第 22.3.1 节。

`-l` *locale*
`--locale=`*locale*

指定此数据库要使用的区域设置。这相当于同时指定`--lc-collate`和`--lc-ctype`。

`--lc-collate=`*locale*

指定该数据库要使用的 LC_COLLATE 设置。

`--lc-ctype=locale`

指定该数据库要使用的 LC_CTYPE 设置。

`-O owner`
`--owner=owner`

指定将拥有新数据库的数据库用户。

`-T template`
`--template=template`

指定用于构建该数据库的模板数据库。

`-V`
`--version`

打印createdb的版本并退出。

`-?`
`--help`

显示有关createdb命令行参数的帮助信息，并退出。

`-D`、`-l`、`-E`、`-O`和`-T`选项对应底层 SQL 命令CREATE DATABASE的选项；有关这些选项的更多信息请参见该命令。

createdb还接受以下用于连接参数的命令行参数：

`-h host`
`--host=host`

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p port`
`--port=port`

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U username`
`--username=username`

以该用户名进行连接。

`-w`
`--no-password`

从不发出密码提示。如果服务器要求密码认证，而密码又无法通过诸如`.pgpass`文件等其他方式获得，则连接尝试将失败。
该选项适合没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制createdb在连接到数据库之前提示输入密码。

这个选项并非必不可少，因为如果服务器要求密码认证，createdb会自动提示输入密码。不过，createdb会浪费一次连接尝试，才知道服务器需要密码。在某些情况下，为了避免这次额外的连接尝试，提前指定`-W`是值得的。

环境

`PGDATABASE`

如果设置了该变量，其值就是要创建的数据库名称，除非被命令行覆盖。

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数。如果命令行和`PGDATABASE`都没有指定要创建的数据库名称，则`PGUSER`也会决定该名称。

与大多数其他PostgreSQL工具一样，该工具也使用libpq支持的环境变量（见第 31.13 节）。

诊断

如果遇到问题，请参阅CREATE DATABASE和psql中关于潜在问题和错误消息的讨论。数据库服务器必须在目标主机上运行。此外，由libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要使用默认数据库服务器创建`demo`数据库：

```
$ createdb demo
```

要使用主机`eden`上、端口为 5000 的服务器，并以 `LATIN1`编码方案创建`demo`数据库，可以看看下面的底层命令：

```
$ createdb -p 5000 -h eden -E LATIN1 -e demo
CREATE DATABASE demo ENCODING 'LATIN1';
```

参见

dropdb, CREATE DATABASE

createlang

createlang — install a PostgreSQL procedural language

大纲

`createlang` [*connection-option*...] *langname* [*dbname*]

`createlang` [*connection-option*...] [--list] | [-l] *dbname*

描述

createlang 是一个向 PostgreSQL 数据库中添加过程语言的工具。

createlang 只是对CREATE EXTENSION SQL 命令的一个包装。

> **小心**
>
> createlang 已被弃用，在未来的 PostgreSQL 版本中可能会被移除。建议直接使用 `CREATE EXTENSION` 命令。

选项

createlang accepts the following command-line arguments:

langname

指定要安装的过程语言的名称。

`[-d]` *dbname*
`[--dbname=]`*dbname*

指定要向哪个数据库添加该语言。默认使用与当前系统用户同名的数据库。

`-e`
`--echo`

在执行 SQL 命令时把它们显示出来。

`-l`
`--list`

显示目标数据库中已安装语言的列表。

`-V`
`--version`

打印 createlang 的版本并退出。

`-?`
`--help`

显示有关 createlang 命令行参数的帮助并退出。

createlang 还接受下列用于连接参数的命令行参数：

`-h` *host*
`--host=`*host*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

`-p` *port*
`--port=`*port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *username*
`--username=`*username*

要用来连接的用户名。

`-w`
`--no-password`

绝不提示输入密码。如果服务器要求密码认证，而密码又无法通过诸如`.pgpass`文件等其他方式获得，则连接尝试将失败。
该选项可用于没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制 createlang 在连接数据库之前提示输入密码。

该选项绝非必需，因为如果服务器要求密码认证，createlang会自动提示输入密码。不过，createlang会浪费一次连接尝试来发现服务器需要密码。在某些情况下，输入`-W`可以避免这次额外的连接尝试。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

和大部分其他 PostgreSQL 工具一样，这个工具也使用 libpq 支持的环境变量（参见第 31.13 节）。

诊断

大多数错误消息都是不言自明的。如果不然，可以带 `--echo` 选项运行 createlang，然后查看相应的 SQL 命令了解详情。此外，libpq 前端库使用的任何默认连接设置和环境变量也都适用。

注解

使用droplang可移除一个语言。

示例

要将语言 `pltcl` 安装到数据库 `template1` 中：

```
$ createlang pltcl template1
```

注意，将该语言安装到 `template1` 中后，之后创建的数据库中也会自动安装它。

参见

droplang, CREATE EXTENSION, CREATE LANGUAGE

createuser

createuser — define a new PostgreSQL user account

大纲

`createuser` [*connection-option*...] [*option*...] [*username*]

描述

createuser创建一个新的PostgreSQL用户（更准确地说，是一个角色）。只有超级用户以及具有`CREATEROLE`权限的用户才能创建新用户，因此调用
createuser的用户必须能够以超级用户或具有`CREATEROLE`权限的用户身份连接。

如果希望创建新的超级用户，必须以超级用户身份连接，而不能仅凭`CREATEROLE`权限。
超级用户意味着能够绕过数据库内所有访问权限检查，因此不应轻易授予超级用户身份。

createuser是SQL命令CREATE ROLE的包装器。通过此工具创建用户与通过其他访问服务器的方法创建用户，在效果上没有区别。

选项

createuser接受以下命令行参数：

username

指定要创建的PostgreSQL用户的名称。
该名称必须不同于此PostgreSQL安装中所有现有角色的名称。

`-c` *number*
`--connection-limit=`*number*

为新用户设置最大连接数。默认不设限制。

`-d`
`--createdb`

允许新用户创建数据库。

`-D`
`--no-createdb`

允许新用户创建数据库。

`-e`
`--echo`

回显createuser生成并发送给服务器的命令。

`-E`
`--encrypted`

加密存储在数据库中的用户密码。如果未指定，则使用默认的密码行为。

`-i`
`--inherit`

新角色将自动继承其所属角色的权限。这是默认设置。

`-I`
`--no-inherit`

新角色将不会自动继承其所属角色的权限。

`-l`
`--login`

允许新用户登录（也就是说，该用户名可用作初始会话用户标识符）。这是默认设置。

`-L`
`--no-login`

不允许新用户登录。（不具有登录权限的角色仍可用于管理数据库权限。）

`-N`
`--unencrypted`

不加密存储在数据库中的用户密码。如果未指定，则使用默认的密码行为。

`-P`
`--pwprompt`

如果给出此选项，createuser会提示输入新用户的密码。如果不打算使用密码认证，则不需要这样做。

`-r`
`--createrole`

允许新用户创建新角色（即该用户将具有`CREATEROLE`权限）。

`-R`
`--no-createrole`

不允许新用户创建新角色。

`-s`
`--superuser`

新用户将成为超级用户。

`-S`
`--no-superuser`

新用户将成为超级用户。

`-V`
`--version`

打印createuser版本并退出。

`-?`
`--help`

显示有关createuser命令行参数的帮助并退出。

如果命令行上没有指定名称和其他缺少的信息，将会提示你输入。

createuser还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host=`*`host`*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

指定连接时使用的用户名（不是要创建的用户名）。

`-w`
`--no-password`

绝不提示输入密码。如果服务器要求密码认证，而密码又无法通过诸如`.pgpass`文件等其他方式获得，则连接尝试将失败。
该选项可用于没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制createuser提示输入密码（用于连接到服务器，而不是新用户的密码）。

该选项绝非必需，因为如果服务器要求密码认证，createuser会自动提示输入密码。不过，createuser会浪费一次连接尝试来发现服务器需要密码。在某些情况下，输入`-W`可以避免这次额外的连接尝试。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他PostgreSQL工具一样，此工具也使用 libpq支持的环境变量（见第 31.13 节）。

诊断

若遇到困难，请参见CREATE ROLE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须在目标主机上运行。此外，libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要在默认数据库服务器上创建用户 `joe`：

```
$ createuser joe
Shall the new role be a superuser? (y/n) n
Shall the new role be allowed to create databases? (y/n) n
Shall the new role be allowed to create more new roles? (y/n) n
```

要使用主机 `eden`、端口 5000 上的服务器创建同一个用户 `joe`，避开提示并查看底层命令：

```
$ createuser -h eden -p 5000 -S -D -R -e joe
CREATE ROLE joe NOSUPERUSER NOCREATEDB NOCREATEROLE INHERIT LOGIN;
```

创建用户`joe`作为超级用户，并立即分配密码：

```
$ createuser -P -s -e joe
Enter password for new role: xyzzy
Enter it again: xyzzy
CREATE ROLE joe PASSWORD 'md5b5f5ba1a423792b526f799ae4eb3d59e'
 SUPERUSER CREATEDB CREATEROLE INHERIT LOGIN;
```

上面的示例中，输入密码时实际上不会回显新密码，但为清晰起见，我们显示了所输入的内容。如你所见，密码在发送给客户端之前会被加密。如果使用了`--unencrypted`选项，密码将出现在回显的命令中（也可能出现在服务器日志以及其他地方），因此在这种情况下，如果还有其他人能看到你的屏幕，就不要使用`-e`。

参见

dropuser, CREATE ROLE

dropdb

dropdb — remove a PostgreSQL database

大纲

`dropdb` [*connection-option*...] [*option*...] *dbname*

描述

dropdb删除一个现有的 PostgreSQL数据库。执行此命令的用户必须是数据库超级用户或该数据库的拥有者。

dropdb是SQL命令DROP DATABASE的包装器。通过此工具删除数据库与通过其他访问服务器的方法删除数据库，在效果上没有区别。

选项

dropdb接受以下命令行参数：

dbname

指定要移除的数据库名称。

`-e`
`--echo`

回显dropdb生成并发送给服务器的命令。

`-i`
`--interactive`

在执行任何破坏性操作之前发出确认提示。

`-V`
`--version`

打印dropdb版本并退出。

`-?`
`--help`

显示有关dropdb命令行参数的帮助并退出。

dropdb也接受下列用于连接参数的命令行参数：

`-h` *host*
`--host=`*host*

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p` *port*
`--port=`*port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

指定用于连接的用户名。

`-w`
`--no-password`

从不发出密码提示。如果服务器要求密码认证，而密码又无法通过诸如 `.pgpass`文件等其他方式获得，则连接尝试将失败。该选项适合没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制dropdb在连接数据库之前提示输入密码。

这个选项并非必不可少，因为如果服务器要求密码认证，dropdb会自动提示输入密码。不过，dropdb会浪费一次连接尝试，才知道服务器需要密码。在某些情况下，为了避免这次额外的连接尝试，提前指定`-W`是值得的。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他PostgreSQL工具一样，此工具也使用 libpq支持的环境变量（见第 31.13 节）。

诊断

若遇到困难，请参见DROP DATABASE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须运行在目标主机上。此外，libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要在默认数据库服务器上删除数据库`demo`：

```
$ dropdb demo
```

要使用主机`eden`、端口 5000 上的服务器删除数据库`demo`，并带确认提示以及查看底层命令：

```
$ dropdb -p 5000 -h eden -i -e demo
Database "demo" will be permanently deleted.
Are you sure? (y/n) y
DROP DATABASE demo;
```

另见

createdb, DROP DATABASE

droplang

droplang — remove a PostgreSQL procedural language

大纲

`droplang` [*`connection-option`*...] *`langname`* [*`dbname`*]

`droplang` [*`connection-option`*...] [--list] | [-l] *`dbname`*

描述

droplang 是一个从 PostgreSQL 数据库中移除现有过程语言的工具。

droplang 只是对DROP EXTENSION SQL 命令的一个包装。

> **小心**
>
> droplang 已被弃用，在未来的 PostgreSQL 版本中可能会被移除。建议直接使用 `DROP EXTENSION` 命令。

选项

droplang 接受下列命令行参数：

`langname`

指定要移除的过程语言的名称。

`[-d]` *`dbname`*
`[--dbname=]`*`dbname`*

指定要从哪个数据库中移除该语言。默认使用与当前系统用户同名的数据库。

`-e`
`--echo`

在执行 SQL 命令时把它们显示出来。

`-l`
`--list`

显示目标数据库中已安装语言的列表。

`-V`
`--version`

打印 droplang 的版本并退出。

`-?`
`--help`

显示有关 droplang 命令行参数的帮助并退出。

droplang 还接受下列用于连接参数的命令行参数：

`-h` *`host`*
`--host=`*`host`*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

要用来连接的用户名。

`-w`
`--no-password`

绝不提示输入密码。如果服务器要求密码认证，而密码又无法通过诸如`.pgpass`文件等其他方式获得，则连接尝试将失败。
该选项可用于没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制 droplang 在连接数据库之前提示输入密码。

该选项绝非必需，因为如果服务器要求密码认证，droplang会自动提示输入密码。不过，droplang会浪费一次连接尝试来发现服务器需要密码。在某些情况下，输入`-W`可以避免这次额外的连接尝试。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

和大部分其他 PostgreSQL 工具一样，这个工具也使用 libpq 支持的环境变量（参见第 31.13 节）。

诊断

大多数错误消息都是不言自明的。如果不然，可以带`--echo`选项运行 droplang，然后查看相应的 SQL 命令了解详情。此外，libpq 前端库使用的任何默认连接设置和环境变量也都适用。

注解

使用createlang可添加一个语言。

示例

要移除语言 `pltcl`：

```
$ droplang pltcl dbname
```

参见

createlang, DROP EXTENSION, DROP LANGUAGE

dropuser

dropuser — remove a PostgreSQL user account

大纲

`dropuser` [*connection-option*...] [*option*...] [*username*]

描述

dropuser移除现有的PostgreSQL用户。只有超级用户和拥有`CREATEROLE`权限的用户才能移除PostgreSQL用户。（要移除超级用户，你自己也必须是超级用户。）

dropuser是SQL命令DROP ROLE的包装器。通过此工具删除用户与通过其他访问服务器的方法删除用户，在效果上没有区别。

选项

dropuser 接受下列命令行参数：

username

指定要移除的 PostgreSQL 用户的名称。如果命令行上没有指定名称，将会提示你输入。

`-e`
`--echo`

回显dropuser生成并发送给服务器的命令。

`-i`
`--interactive`

在实际移除用户之前提示确认。

`-V`
`--version`

打印dropuser版本并退出。

`-?`
`--help`

显示有关dropuser命令行参数的帮助并退出。

dropuser也接受下列用于连接参数的命令行参数：

`-h` *host*
`--host=`*host*

指定服务器运行所在主机的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字所在目录。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

用于连接的用户名（不是要删除的用户名）。

`-w`
`--no-password`

从不发出密码提示。如果服务器要求密码认证，而密码又无法通过诸如 `.pgpass`文件之类的其他方式获得，则连接尝试将失败。
该选项适合没有用户在场输入密码的批处理作业和脚本。

`-W`
`--password`

强制dropuser在连接到数据库之前提示输入密码。

这个选项绝非必需，因为如果服务器要求密码认证，dropuser会自动提示输入密码。不过，dropuser会浪费一次连接尝试来发现服务器需要密码。在某些情况下，输入`-W`值得一试，以避免这次额外的连接尝试。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他PostgreSQL工具一样，此工具也使用 libpq支持的环境变量（见第 31.13 节）。

诊断

若遇到困难，请参见DROP ROLE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须在目标主机上运行。此外，libpq前端库使用的任何默认连接设置和环境变量也都适用。

示例

要从默认数据库服务器上移除用户`joe`：

```
$ dropuser joe
```

要使用主机`eden`、端口 5000 上的服务器移除用户`joe`，并带确认提示以及查看底层命令：

```
$ dropuser -p 5000 -h eden -i -e joe
Role "joe" will be permanently removed.
Are you sure? (y/n) y
DROP ROLE joe;
```

另见

createuser, DROP ROLE

ecpg

ecpg — embedded SQL C preprocessor

大纲

ecpg [*option*...] *file*...

描述

ecpg是供 C 程序使用的嵌入式 SQL 预处理器。它会把嵌入了 SQL 语句的 C 程序转换成普通 C 代码，方法是将 SQL 调用替换为特殊的函数调用。生成的输出文件随后可以由任何 C 编译器工具链处理。

ecpg 会把命令行上给出的每个输入文件转换为对应的 C 输出文件。输入文件最好带有扩展名 .pgc，这种情况下会用 .c 替换该扩展名来确定输出文件名。如果输入文件的扩展名不是 .pgc，则通过在整个文件名后追加 .c 来计算输出文件名。也可以使用 -o 选项覆盖输出文件名。

本参考页不描述嵌入式 SQL 语言。有关该主题的更多信息，见第 33 章。

选项

ecpg 接受以下命令行参数：

-c

自动从 SQL 代码生成某些 C 代码。目前，这适用于EXEC SQL TYPE。

-C *mode*

设置兼容模式。*mode*可以是INFORMIX或INFORMIX_SE。

-D *symbol*

定义一个 C 预处理器符号。

-i

也解析系统 include 文件。

-I *directory*

指定一个附加的 include 路径，用于查找通过EXEC SQL INCLUDE包含的文件。默认路径依次为 .（当前目录）、/usr/local/include、在编译时定义的PostgreSQL include 目录（默认：/usr/local/pgsql/include），以及 /usr/include。

-o *filename*

指定ecpg应将全部输出写入给定的*filename*。

-r *option*

选择运行时行为。*option* 可以是以下值之一：

`no_indicator`

不使用指示器，而改用特殊值表示空值。历史上曾有数据库采用这种方式。

`prepare`

在使用前预备所有语句。libecpg会维护一个已预备语句的缓存；
如果某条语句再次执行，就会复用它。若缓存已满，
libecpg会释放使用次数最少的语句。

`questionmarks`

出于兼容性原因，允许使用问号作为占位符。这在很久以前曾是默认行为。

`-t`

打开事务自动提交。在此模式下，除非 SQL 命令位于显式事务块中，否则每条 SQL 命令都会自动提交。在默认模式下，只有发出`EXEC SQL COMMIT`时才会提交命令。

`-v`

打印附加信息，包括版本和 include 路径。

`--version`

打印ecpg版本并退出。

`--help`

显示关于ecpg命令行参数的帮助信息，并退出。

注解

在编译预处理后的 C 代码文件时，编译器必须能够在PostgreSQL的 include 目录中找到ECPG头文件。因此，调用编译器时你可能需要使用`-I`选项（例如`-I/usr/local/pgsql/include`）。

使用嵌入式 SQL 的 C 程序必须与`libecpg`库链接，例如可使用链接器选项`-L/usr/local/pgsql/lib -lecpg`。

适合该安装的这两个目录的值可以通过pg_config查出。

示例

如果你有一个名为`prog1.pgc`的嵌入式 SQL C 源文件，就可以使用下列命令序列创建一个可执行程序：

```
ecpg prog1.pgc
cc -I/usr/local/pgsql/include -c prog1.c
cc -o prog1 prog1.o -L/usr/local/pgsql/lib -lecpg
```

pg_basebackup

pg_basebackup — take a base backup of a PostgreSQL cluster

大纲

`pg_basebackup` [*option*...]

描述

pg_basebackup用于获取正在运行的PostgreSQL数据库集簇的基础备份。
备份过程不会影响数据库的其他客户端，并且该备份既可用于时间点恢复（见第 24.3 节），也可用作日志传送或流复制备库的起点（见第 25.2 节）。

pg_basebackup创建数据库集簇文件的二进制副本，同时确保系统自动进入和退出备份模式。
备份始终针对整个数据库集簇；无法只备份单个数据库或数据库对象。要备份单个数据库，
必须使用pg_dump之类的工具。

备份通过使用复制协议的普通 PostgreSQL 连接进行。建立连接时必须使用具有 `REPLICATION` 权限的用户（参见第 20.2 节），并且必须在 `pg_hba.conf` 中授予该用户显式权限。服务器还必须将max_wal_senders设置得足够高，以便至少为备份保留一个可用会话。

可以同时运行多个`pg_basebackup`，但从性能角度来看，最好只执行一次备份，
然后复制其结果。

选项

以下命令行选项控制输出的位置和格式。

`-D` *directory*
`--pgdata=`*directory*

写入输出的目标目录。

当备份采用 tar 格式时，目标目录可以指定为`-`（短横线），从而将 tar 文件写到`stdout`。

此参数是必需的。

`-F` *format*
`--format=`*format*

选择输出格式。*format*可以是以下值之一：

`p`
`plain`

将输出写为普通文件，其布局与源服务器的数据目录和表空间相同。
当集簇没有额外表空间时，整个数据库都会放在目标目录中。如果集簇包含额外表空间，
则主数据目录会放在目标目录中，
而其他所有表空间都会放在与服务器上相同的绝对路径中。

这是默认格式。

`t`
`tar`

将输出写为目标目录中的 tar 文件。主数据目录的内容会写入名为`base.tar`的文件中，而每个其他表空间都会写入一个以该表空间 OID 命名的独立 tar 文件中。

如果目标目录指定为`-`（短横线），tar 内容将写入标准输出，适合通过管道传给例如 gzip。只有当集簇没有额外表空间时，才允许这样做。

`-x`
`--xlog`

在备份中包含所需的事务日志（WAL）文件。这将包括备份期间生成的所有事务日志。
如果指定了此选项，就可以在解包后的目录中直接启动 postmaster，而无需查阅日志归档，从而使输出成为一个完全独立的备份。

> **注意**
>
> 事务日志文件在备份结束时收集。因此，必须把wal_keep_segments参数设置得足够高，使日志在备份结束之前不会被移除。
> 如果在传输日志时它已经被轮转，备份将失败且不可用。

`-z`
`--gzip`

启用 tar 文件输出的 gzip 压缩，使用默认压缩级别。压缩仅在使用 tar 格式时可用。

`-Z` *`level`*
`--compress=`*`level`*

启用 tar 文件输出的 gzip 压缩，并指定压缩级别（0 到 9，其中 0 表示不压缩，9 表示最佳压缩）。压缩仅在使用 tar 格式时可用。

以下命令行选项控制备份的生成和程序的运行。

`-c` *`fast|spread`*
`--checkpoint=`*`fast|spread`*

将检查点模式设置为fast（立即）或spread（默认）。

`-l` *`label`*
`--label=`*`label`*

设置备份的标签。如果未指定，将使用默认值 `pg_basebackup base backup`。

`-P`
`--progress`

启用进度报告。打开此选项后，会在备份过程中给出一个近似的进度报告。
由于数据库在备份过程中可能发生变化，因此这只是近似值，最终未必恰好结束在`100%`。
特别是当备份中包含 WAL 时，总数据量无法预先估计；在这种情况下，一旦进度超过不含 WAL 时的总估计值，估计目标大小就会继续增加。

启用此选项后，备份会先遍历整个数据库以统计其大小，然后再返回发送实际内容。
这可能使备份稍微耗时更长，尤其会使首次发送数据之前的等待时间更长。

`-v`
`--verbose`

启用详细模式。它会在启动和关闭过程中输出一些额外步骤；如果同时启用了进度报告，还会显示当前正在处理的确切文件名。

以下命令行选项控制数据库连接参数。

`-h` *host*
`--host=`*host*

指定运行服务器的机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。默认值取自 `PGHOST` 环境变量（如果已设置）；否则会尝试使用 Unix 域套接字连接。

`-p` *port*
`--port=`*port*

指定服务器监听连接所使用的 TCP 端口，或本地 Unix 域套接字文件扩展名。默认使用 `PGPORT`环境变量中的值（如果已设置），否则使用编译时确定的默认值。

`-U` *username*
`--username=`*username*

指定连接时使用的用户名。

`-w`
`--no-password`

禁止发出密码提示。如果服务器要求密码认证，而又无法通过其他方式（例如`.pgpass`文件）获得密码，则连接尝试将失败。此选项对于批处理作业和脚本很有用，因为那种场景下通常没有用户在场输入密码。

`-W`
`--password`

强制pg_basebackup在连接数据库之前提示输入密码。

此选项绝非必需，因为如果服务器要求密码认证，pg_basebackup会自动提示输入密码。不过，pg_basebackup会浪费一次连接尝试来发现服务器需要密码。在某些情况下，输入`-W`以避免额外的连接尝试是值得的。

还可使用其他一些不太常用的参数：

`-V`
`--version`

打印 pg_basebackup 的版本并退出。

`-?`
`--help`

显示pg_basebackup命令行参数的帮助并退出。

环境

与大多数其他PostgreSQL工具一样，此工具也使用libpq支持的环境变量（见第 31.13 节）。

注解

备份将包含数据目录和表空间中的所有文件，
包括配置文件以及第三方放置在该目录中的任何附加文件。数据目录中只允许常规文件和目录，
不允许符号链接或特殊设备文件。

按照 PostgreSQL 管理表空间的方式，无论何时恢复备份，所有附加表空间的路径都必须相同。
不过，主数据目录可以重定位到任何位置。

示例

要为服务器`mydbserver`创建一个基础备份，并将其存储到本地目录`/usr/local/pgsql/data`中：

```
$ pg_basebackup -h mydbserver -D /usr/local/pgsql/data
```

要为本地服务器创建一个备份，为每个表空间各生成一个压缩的 tar 文件，并将其存储在目录`backup`中，同时在运行期间显示进度报告：

```
$ pg_basebackup -D backup -Ft -z -P
```

要为一个仅包含单个表空间的本地数据库创建备份，并使用bzip2进行压缩：

```
$ pg_basebackup -D - -Ft | bzip2 > backup.tar.bz2
```

（如果该数据库中有多个表空间，此命令将失败。）

另见

pg_dump

pg_config

pg_config — retrieve information about the installed version of PostgreSQL

大纲

`pg_config` [*option*...]

Description

pg_config工具打印当前已安装的PostgreSQL版本的配置参数。例如，希望与PostgreSQL进行接口的软件包可以使用它，以便找到所需的头文件和库。

Options

要使用 pg_config，请提供下列一个或多个选项：

`--bindir`

打印用户可执行文件的位置。例如，可以用它来寻找 `psql` 程序。通常这也是 `pg_config` 程序所在的位置。

`--docdir`

打印文档文件的位置。

`--htmldir`

打印文档文件所在的位置。

`--includedir`

打印 HTML 文档文件所在的位置。

`--pkgincludedir`

打印客户端接口的 C 头文件所在的位置。

`--includedir-server`

打印其他 C 头文件所在的位置。

`--libdir`

打印服务器编程所用 C 头文件所在的位置。

`--pkglibdir`

打印目标代码库所在的位置。

`--localedir`

打印区域支持文件的位置。（如果在构建 PostgreSQL 时没有配置区域支持，这将是一个空字符串。）

`--mandir`

打印手册页的位置。

`--sharedir`

打印手册页所在的位置。

`--sysconfdir`

打印与体系结构无关的支持文件所在的位置。

`--pgxs`

打印系统范围的配置文件所在的位置。

`--configure`

打印为构建 PostgreSQL 而配置时传递给 `configure` 脚本的选项。这可用于复现完全相同的配置，或者查明某个二进制包是用什么选项构建的。（不过注意，
二进制包通常包含厂商特有的自定义补丁。）另请参见下面的例子。

`--cc`

打印为构建`PostgreSQL`而运行configure脚本时给定的选项。
这可用于重现完全相同的配置，或者查明某个二进制软件包是用哪些选项构建的。
（但请注意，二进制软件包通常包含供应商特定的自定义补丁。）另请参见下面的示例。

`--cppflags`

打印构建 PostgreSQL 时使用的 `CPPFLAGS` 变量的值。它显示预处理时需要的 C 编译器开关（通常是 `-I` 开关）。

`--cflags`

打印构建 PostgreSQL 时使用的 `CFLAGS` 变量的值。它显示 C 编译器开关。

`--cflags_sl`

打印构建PostgreSQL时所用的`CFLAGS`变量值。这显示 C 编译器开关。

`--ldflags`

打印构建PostgreSQL时所用的`CFLAGS_SL`变量值。这显示构建共享库时使用的额外 C 编译器开关。

`--ldflags_ex`

打印构建PostgreSQL时所用的`LDFLAGS`变量值。这显示链接器开关。

`--ldflags_sl`

打印构建PostgreSQL时所用的`LDFLAGS_EX`变量值。
这显示仅用于构建可执行文件的链接器开关。

`--libs`

打印构建 PostgreSQL 时使用的 `LIBS` 变量的值。它通常包含链接进 PostgreSQL 的外部库的 `-l` 开关。

`--version`

打印 PostgreSQL 的版本。

If more than one option is given, the information is printed in that order, one item per line. If no options are given, all available information is printed, with labels.

Notes

`--includedir-server` 选项是在 PostgreSQL 7.2 中加入的。在之前的版本中，服务器头文件与客户端头文件安装在同一位置，可以用 `--includedir` 选项查询。要让你的软件包同时处理这两种情况，请先尝试较新的选项，并测试退出状态以确定它是否成功。

选项`--docdir`、`--pkgincludedir`、`--localedir`、`--mandir`、`--sharedir`、`--sysconfdir`、`--cc`、`--cppflags`、`--cflags`、`--cflags_sl`、`--ldflags`、`--ldflags_sl`以及`--libs`是在PostgreSQL 8.1 中加入的。选项`--htmldir`是在PostgreSQL 8.4 中加入的。选项`--ldflags_ex`是在PostgreSQL 9.0 中加入的。

在 PostgreSQL 7.1 之前的发行版中，还没有 `pg_config`，当时不存在获取等效配置信息的方法。

Example

要重现当前 PostgreSQL 安装的构建配置，请运行下列命令：

```
eval ./configure `pg_config --configure`
```

`pg_config --configure`的输出包含 shell 引号，因此带空格的参数能够被正确表示。因此，要得到正确的结果，需要使用`eval`。

pg_dump

pg_dump — extract a PostgreSQL database into a script file or other archive file

大纲

`pg_dump` [*connection-option*...] [*option*...] [*dbname*]

描述

pg_dump是一个用于备份PostgreSQL数据库的工具。即使数据库正在被并发使用，它也能生成一致的备份。pg_dump不会阻塞其他用户访问数据库（无论读还是写）。

转储可以输出为脚本格式或归档文件格式。脚本转储是纯文本文件，包含把数据库重建到保存时状态所需的 SQL 命令。要从这样的脚本恢复，只需将其交给psql。脚本文件甚至可以在其他机器和其他体系结构上用于重建数据库；经过一些修改后，甚至也可以用于其他 SQL 数据库产品。

另一类归档文件格式必须结合pg_restore来重建数据库。它们允许 pg_restore 有选择地恢复某些内容，甚至在恢复之前重新排列条目。归档文件格式被设计为可跨体系结构移植。

使用归档文件格式并配合 pg_restore 时，pg_dump 提供了一种灵活的归档和传输机制。pg_dump 可用于备份整个数据库，而 pg_restore 可用于检查归档和/或选择要恢复的数据库部分。最灵活的输出文件格式是“custom”格式（`-Fc`）。它允许选择和重新排序所有归档条目，并且默认会压缩。

运行 pg_dump 时，应检查输出中是否有任何警告（打印到标准错误），尤其要结合下面列出的限制来查看。

选项

以下命令行选项控制输出的内容和格式。

dbname

指定要转储的数据库名称。如果未指定，则使用环境变量`PGDATABASE`。如果未设置该变量，则使用连接指定的用户名。

`-a`
`--data-only`

只转储数据，不转储模式（数据定义）。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-b`
`--blobs`

在转储中包含大对象。除非指定了 `--schema`、`--table` 或 `--schema-only`，否则这是默认行为。因此，`-b` 开关只在需要把大对象加入选择性转储时才有用。

`-c`
`--clean`

在输出创建数据库对象的命令之前，先输出清理（删除）这些对象的命令。（恢复时可能会产生一些无害的错误。）

此选项只对纯文本格式有意义。对于归档格式，可以在调用 pg_restore 时指定该选项。

-C
--create

让输出以创建数据库本身并重新连接到新建数据库的命令开始。（使用这种形式的脚本时，在运行脚本之前连接到哪个数据库无关紧要。）

此选项只对纯文本格式有意义。对于归档格式，可以在调用 pg_restore 时指定该选项。

-E *encoding*
--encoding=*encoding*

在指定的字符集编码中创建转储。默认情况下，转储将以数据库编码创建。
（获得相同结果的另一种方法是将PGCLIENTENCODING环境变量设置为所需的转储编码。）

-f *file*
--file=*file*

将输出发送到指定文件。对于基于文件的输出格式，可以省略此参数，这时使用标准输出。
不过，对于目录输出格式，必须给出此参数，因为它指定的是目标目录而不是文件。
在这种情况下，该目录由 pg_dump 创建，并且在此之前必须不存在。

-F *format*
--format=*format*

选择输出格式。*format*可以是以下值之一：

p
plain

输出纯文本 SQL 脚本文件（默认）。

c
custom

输出适合供 pg_restore 使用的 custom 格式归档。与目录输出格式一起，这是最灵活的输出格式，因为它允许在恢复时手工选择并重新排序归档条目。这种格式默认也会压缩。

d
directory

输出适合供 pg_restore 使用的 directory 格式归档。这会创建一个目录，其中每个被转储的表和大对象各有一个文件，外加一个所谓的目录表（Table of Contents）文件，以机器可读格式描述被转储对象，pg_restore 可以读取它。directory 格式归档可以用标准 Unix 工具来操作；例如，未压缩归档中的文件可以使用 gzip 工具进行压缩。该格式默认压缩，并且支持并行转储。

t
tar

输出 tar 格式归档，适合供 pg_restore。tar 格式与 directory 格式兼容：解开一个 tar 格式归档就会得到一个有效的 directory 格式归档。不过，tar 格式不支持压缩。另外，使用 tar 格式时，在恢复过程中不能改变表数据项的相对顺序。

`-i`
`--ignore-version`

一个已弃用、现在会被忽略的选项。

`-n` *schema*
`--schema=`*schema*

只转储匹配 *schema* 的模式；这既会选择模式本身，也会选择其中包含的所有对象。未指定此选项时，将转储目标数据库中的所有非系统模式。可以通过写多个 `-n` 开关来选择多个模式。*schema* 参数按照 psql 的 `\d` 命令所使用的同样规则进行解释（见模式），因此也可以通过在模式中使用通配符来选择多个模式。使用通配符时，如有需要请小心为模式加引号，以防止 shell 展开通配符；见下文的Examples。

> **注意**
>
> 当指定 `-n` 时，pg_dump 不会尝试转储所选模式可能依赖的任何其他数据库对象。因此，不能保证特定模式转储的结果能够单独成功恢复到一个干净的数据库中。

> **注意**
>
> 指定 `-n` 时，不会转储大对象等非模式对象。可以使用 `--blobs` 开关把大对象加回转储中。

`-N` *schema*
`--exclude-schema=`*schema*

不转储任何名称匹配*schema*的模式。该匹配条件按照与`-n`相同的规则解释。可以多次指定`-N`，以排除匹配多个条件的模式。

当同时给出 `-n` 和 `-N` 时，其行为是只转储至少匹配一个 `-n` 开关但不匹配任何 `-N` 开关的模式。如果出现 `-N` 而没有 `-n`，那么匹配 `-N` 的模式会从原本的正常转储中排除。

`-o`
`--oids`

将对象标识符（OID）作为每个表的数据的一部分进行转储。如果应用程序以某种方式引用OID列（例如在外键约束中），请使用此选项。否则，不应使用此选项。

`-O`
`--no-owner`

不输出用于把对象所有权设置成与原始数据库一致的命令。默认情况下，pg_dump 会发出 `ALTER OWNER` 或 `SET SESSION AUTHORIZATION` 语句来设置新建数据库对象的所有权。除非脚本由超级用户（或拥有脚本中所有对象的同一用户）启动，否则这些语句会在运行脚本时报错。若要创建一个可由任意用户恢复、并让该用户拥有所有对象的脚本，请指定 `-O`。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-R`
`--no-reconnect`

这个选项已经过时，但仍然被接受以保持向后兼容性。

`-s`
`--schema-only`

只转储对象定义（模式），不转储数据或统计信息。

`-S` *`username`*
`--superuser=`*`username`*

指定在禁用触发器时要使用的超级用户名。这只在使用 `--disable-triggers` 时相关。（通常最好省略它，而是以超级用户身份运行生成的脚本。）

`-t` *`table`*
`--table=`*`table`*

只转储名称匹配 *`table`* 的表（或视图、序列或外部表）。可以通过写多个 `-t` 开关来选择多个表。*`table`* 参数按照 psql 的 `\d` 命令所使用的同样规则进行解释（见模式），因此也可以通过在模式中使用通配符来选择多个表。使用通配符时，如有需要请小心为模式加引号，以防止 shell 展开通配符；见下文的Examples。

使用 `-t` 时，`-n` 和 `-N` 开关没有作用，因为由 `-t` 选中的表无论这些开关如何设置都会被转储，而非表对象则不会被转储。

> **注意**
>
> 当指定 `-t` 时，pg_dump 不会尝试转储所选表可能依赖的任何其他数据库对象。因此，不能保证特定表转储的结果能够单独成功恢复到一个干净的数据库中。

> **注意**
>
> `-t`开关的行为与 8.2 之前的PostgreSQL版本并不完全向上兼容。以前，写成`-t tab`会转储所有名为`tab`的表，现在则只转储默认搜索路径中可见的那个表。要获得旧行为，可以写成`-t '*.tab'`。此外，要选择特定模式中的表，必须写成类似`-t sch.tab`的形式，而不是旧的`-n sch -t tab`写法。

`-T` *`table`*
`--exclude-table=`*`table`*

不转储任何匹配 *`table`* 的表。该模式按照与 `-t` 相同的规则解释。`-T` 可以给出多次，以排除匹配多个模式的表。

当同时给出 `-t` 和 `-T` 时，其行为是只转储至少匹配一个 `-t` 开关但不匹配任何 `-T` 开关的表。如果出现 `-T` 而没有 `-t`，那么匹配 `-T` 的表会从原本的正常转储中排除。

`-v`
`--verbose`

指定详细模式。这会使 pg_dump 将详细的对象注释、开始/停止时间写入转储文件，并把进度消息写到标准错误。

`-V`
`--version`

打印 pg_dump 的版本并退出。

`-x`
`--no-privileges`
`--no-acl`

不转储访问权限（grant/revoke 命令）。

`-Z 0..9`
`--compress=0..9`

指定所用的压缩级别。零表示不压缩。对于自定义格式归档，
该选项指定对各个表数据段的压缩，默认以适中的级别压缩。对于纯文本输出，
设置非零压缩级别会压缩整个输出文件，就像将其传给gzip处理一样；但默认不压缩。tar 归档格式目前完全不支持压缩。

`--binary-upgrade`

此选项供就地升级实用程序使用。不建议或支持将其用于其他目的。
该选项的行为可能在未来的版本中更改而不另行通知。

`--column-inserts`
`--attribute-inserts`

将数据转储为带有显式列名的`INSERT`命令（`INSERT INTO table (column, ...) VALUES ...`）。这会使恢复变得非常缓慢；它主要用于生成可装入非 PostgreSQL 数据库的转储文件。不过，由于此选项为每一行生成单独的命令，重新装载某一行时出错只会导致该行丢失，而不会导致整个表的内容丢失。

`--disable-dollar-quoting`

此选项禁用函数体中的 dollar quoting，并强制改用 SQL 标准字符串语法对它们进行引用。

`--disable-triggers`

此选项只在创建包含数据但不包含模式的转储时才相关。它指示 pg_dump 在输出中包含一些命令，以便在恢复数据时临时禁用目标表上的触发器。
如果这些表上存在不希望在数据恢复期间触发的引用完整性检查或其他触发器，
请使用此选项。

目前，为 `--disable-triggers` 输出的这些命令必须由超级用户执行。因此，还应通过 `-S` 指定一个超级用户名，或者更好的做法是谨慎地以超级用户身份运行生成的脚本。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`--inserts`

将数据转储为`INSERT`命令（而不是`COPY`）。这会使恢复变得非常缓慢；
它主要用于生成可装载到非PostgreSQL数据库的转储。不过，
由于此选项为每一行生成单独的命令，重新装载某一行时出错只会导致该行丢失，
而不会导致整个表的内容丢失。注意，如果重新排列了列顺序，恢复可能会完全失败。
`--column-inserts`选项不受列顺序变化影响，但会更慢。

`--lock-wait-timeout=timeout`

在转储开始时，不要无限等待获取共享表锁。如果无法在指定的 *timeout* 内锁定某个表，就让转储失败。超时可以用 `SET statement_timeout` 接受的任意格式指定。（允许的值因被转储源服务器的版本而异，但 7.3 以来的所有版本都接受以毫秒为单位的整数。从 7.3 之前的服务器转储时，此选项会被忽略。）

`--no-security-labels`

不要转储安全标签。

`--no-tablespaces`

不输出用于选择表空间的命令。使用此选项时，
所有对象在恢复时都会创建在当时默认的表空间中。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`--no-unlogged-table-data`

不要转储不记录 WAL 的表和序列的内容。此选项不会影响是否转储表和序列的定义（模式）；它只会抑制表和序列数据的转储。从备库转储时，不记录 WAL 的表和序列中的数据始终会被排除。

`--quote-all-identifiers`

强制为所有标识符加引号。当从某个服务器转储数据库，而该服务器的 PostgreSQL 主版本与 pg_dump 不同，或者输出打算装入另一主版本的服务器时，推荐使用此选项。默认情况下，pg_dump 只会为在其自身主版本中属于保留字的标识符加引号。这有时会在处理其他版本服务器时导致兼容性问题，因为这些版本的保留字集合可能略有不同。使用 `--quote-all-identifiers` 可以防止此类问题，但代价是转储脚本更难阅读。

`--serializable-deferrable`

使用 `serializable` 事务来执行转储，以确保所用快照与后续数据库状态一致；但这是通过等待事务流到达一个不会出现异常的时点来实现的，从而避免转储失败或导致其他事务因 `serialization_failure` 而回滚。有关事务隔离和并发控制的更多信息，请参见第 13 章。

这个选项对于仅用于灾难恢复的转储没有好处。但对于用来加载数据库副本，
以供报表或其他只读负载共享，而原始数据库继续更新的转储，它可能有用。
如果不使用它，转储可能反映出一种与最终提交事务的任何串行执行都不一致的状态。
例如，如果使用批处理技术，转储中可能显示某个批次已经关闭，
但批次中的全部条目却并未出现。

如果在启动 pg_dump 时没有活动的读写事务，此选项不会带来任何差别。如果存在活动的读写事务，转储开始时间可能会被延迟一个不确定的时长。一旦开始运行，
使用或不使用该开关的性能都是相同的。

`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是 `ALTER OWNER` 命令来确定对象所有权。这会让转储更符合标准，但根据转储中对象的历史，可能无法正确恢复。另外，使用 `SET SESSION AUTHORIZATION` 的转储肯定需要超级用户权限才能正确恢复，而 `ALTER OWNER` 只需较低权限。

`-?`
`--help`

显示关于 pg_dump 命令行参数的帮助信息，并退出。

下列命令行选项控制数据库连接参数。

`-h` *host*
`--host=`*host*

指定运行服务器的机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。默认值取自 `PGHOST` 环境变量（如果已设置）；否则会尝试使用 Unix 域套接字连接。

`-p` *port*
`--port=`*port*

指定服务器正在监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。默认值取自 `PGPORT` 环境变量（如果已设置），否则使用编译时默认值。

`-U` *username*
`--username=`*username*

用于连接的用户名。

`-w`
`--no-password`

绝不发出密码提示。如果服务器要求密码认证，而又无法通过 `.pgpass` 文件等其他方式获得密码，则连接尝试将失败。在没有用户在场输入密码的批处理作业和脚本中，此选项很有用。

`-W`
`--password`

强制 pg_dump 在连接数据库之前提示输入密码。

这个选项绝非必需，因为如果服务器要求密码认证，pg_dump 会自动提示输入密码。不过，pg_dump 会先浪费一次连接尝试来发现服务器需要密码。在某些情况下值得输入 `-W`，因为可以避免这次额外的连接尝试。

`--role=`*rolename*

指定用于创建转储的角色名称。此选项会让 pg_dump 在连接数据库后发出 `SET ROLE` *rolename* 命令。当经认证用户（由 `-U` 指定）缺少 pg_dump 所需权限，但可以切换到具有所需权限的角色时，此选项很有用。有些安装环境有禁止直接以超级用户登录的策略，使用此选项就可以在不违反该策略的情况下进行转储。

环境

`PGDATABASE`
`PGHOST`
`PGOPTIONS`
`PGPORT`
`PGUSER`

默认连接参数。

和大多数其他 PostgreSQL 工具一样，这个工具也使用 libpq 支持的环境变量（见第 31.13 节）。

诊断

pg_dump 在内部执行 `SELECT` 语句。如果运行 pg_dump 时遇到问题，请确认能够例如使用psql从数据库中查询信息。此外，libpq 前端库所使用的任何默认连接设置和环境变量也都会生效。

pg_dump 的数据库活动通常会被累积统计系统收集。如果不希望如此，可以将参数 `track_counts` 通过 `PGOPTIONS` 或 `ALTER USER` 命令设置为 false。

注意

如果数据库集簇在 `template1` 数据库中有任何本地添加的内容，要小心把 pg_dump 的输出恢复到一个真正空的数据库中；否则很可能因为这些新增对象的重复定义而报错。要创建一个不含任何本地添加的空数据库，应从 `template0` 而不是 `template1` 复制，例如：

```
CREATE DATABASE foo WITH TEMPLATE template0;
```

当选择不包含模式的转储并使用 `--disable-triggers` 选项时，pg_dump 会在插入数据前发出命令禁用用户表上的触发器，并在数据插入完成后发出命令重新启用它们。
如果恢复在中途停止，系统目录可能会保持在错误状态。

pg_dump生成的转储文件不包含优化器用于决定查询计划的统计信息。因此，
从转储文件恢复后，最好运行`ANALYZE`以确保最佳性能；更多信息请参阅第 23.1.3 节和第 23.1.5 节。转储文件也不包含任何`ALTER DATABASE ... SET`命令；这些设置由pg_dumpall与数据库用户及其他整个安装环境的设置一起转储。

由于 pg_dump 常被用来把数据迁移到更新版本的 PostgreSQL，因此通常可以期望 pg_dump 的输出能够装载到 PostgreSQL 服务器中，而这些服务器的版本比 pg_dump 更高。pg_dump 也可以从版本比它自身更旧的 PostgreSQL 服务器上转储数据。（目前支持回溯到 7.0 版的服务器。）然而，pg_dump 不能从主版本高于它自身的 PostgreSQL 服务器上转储；它甚至会拒绝尝试，以免冒生成无效转储的风险。另外，也不能保证 pg_dump 的输出能够装载到更旧主版本的服务器上 – 即使转储正是从该版本服务器上取得的。把转储文件装载到较旧服务器时，可能需要手工编辑转储文件，移除旧服务器无法理解的语法。在跨版本场景中，建议使用 `--quote-all-identifiers` 选项，因为它可以防止不同 PostgreSQL 版本保留字列表差异带来的问题。

Examples

要将名为 `mydb` 的数据库转储到 SQL 脚本文件中：

```
$ pg_dump mydb > db.sql
```

要将这样的脚本重新装载到一个新创建的数据库中，其名称为 `newdb`：

```
$ psql -d newdb -f db.sql
```

要把一个数据库转储为 custom 格式归档文件：

```
$ pg_dump -Fc mydb > db.dump
```

要把一个数据库转储为 directory 格式归档：

```
$ pg_dump -Fd mydb -f dumpdir
```

要把一个归档文件重新装入到一个（新创建的）名为 newdb 的数据库：

```
$ pg_restore -d newdb db.dump
```

要转储一个名为 mytab 的表：

```
$ pg_dump -t mytab mydb > db.sql
```

要转储 detroit 模式中名称以 emp 开头的所有表，但排除名为 employee_log 的表：

```
$ pg_dump -t 'detroit.emp*' -T detroit.employee_log mydb > db.sql
```

要转储名称以 east 或 west 开头、并且以 gsm 结尾的所有模式，同时排除名称中包含单词 test 的任何模式：

```
$ pg_dump -n 'east*gsm' -n 'west*gsm' -N '*test*' mydb > db.sql
```

同样，使用正则表达式记法来合并这些开关：

```
$ pg_dump -n '(east|west)*gsm' -N '*test*' mydb > db.sql
```

要转储除名称以 ts_ 开头的表之外的所有数据库对象：

```
$ pg_dump -T 'ts_*' mydb > db.sql
```

要在 -t 及相关开关中指定大写或大小写混合的名称，必须用双引号括起该名称；否则它会被折叠为小写（参见模式）。但双引号对 shell 有特殊意义，因此还必须用引号保护这些双引号。因此，要转储一个名称大小写混合的表，需要写成类似如下的形式：

```
$ pg_dump -t "\"MixedCaseName\"" mydb > mytab.sql
```

另见

pg_dumpall, pg_restore, psql

pg_dumpall

pg_dumpall — extract a PostgreSQL database cluster into a script file

大纲

`pg_dumpall` [*connection-option*...] [*option*...]

描述

pg_dumpall是一个实用程序，用于写出（“转储”）
一个集簇中的所有PostgreSQL数据库到一个脚本文件中。该脚本文件包含SQL命令，可作为psql的输入来恢复这些数据库。它通过对集簇中的每个数据库调用pg_dump来实现这一点。pg_dumpall还会转储所有数据库共有的全局对象。（pg_dump不会保存这些对象。）
目前包括数据库用户和组、表空间，以及适用于整个数据库的访问权限等属性的信息。

由于pg_dumpall会读取所有数据库中的表，因此通常必须以数据库超级用户身份连接，
才能生成完整的转储。另外，要执行保存的脚本，也需要具备超级用户权限，
这样才能添加用户和组并创建数据库。

SQL 脚本将写入标准输出。使用 [-f|file] 选项或 shell 操作符将其重定向到文件。

pg_dumpall需要多次连接到PostgreSQL服务器（每个数据库一次）。如果使用密码认证，
它每次都会提示输入密码。在这种情况下准备一个`~/.pgpass`文件会比较方便。详见第 31.14 节。

选项

以下命令行选项控制输出的内容和格式。

`-a`
`--data-only`

只转储数据，不转储模式（数据定义）。

`-c`
`--clean`

包含用于在重新创建数据库之前清理（删除）数据库的 SQL 命令。也会添加用于角色和表空间的`DROP`命令。

`-f` *filename*
`--file=`*filename*

将输出发送到指定文件。如果省略此选项，则使用标准输出。

`-g`
`--globals-only`

只转储全局对象（角色和表空间），不转储数据库。

`-i`
`--ignore-version`

一个已废弃的选项，现在会被忽略。

`-o`
`--oids`

将对象标识符（OID）作为每个表的数据的一部分进行转储。如果应用程序以某种方式引用 OID 列（例如在外键约束中），请使用此选项。否则，不应使用此选项。

`-O`
`--no-owner`

不要输出用于将对象所有权设置为与原始数据库一致的命令。默认情况下，pg_dumpall 会发出 `ALTER OWNER` 或 `SET SESSION AUTHORIZATION` 语句，以设置已创建模式元素的所有权。除非该脚本由超级用户（或拥有脚本中全部对象的同一用户）启动，否则这些语句在运行时会失败。若要创建一个可由任意用户恢复、并让该用户获得所有对象所有权的脚本，请指定 `-O`。

`-r`
`--roles-only`

只转储对象定义（模式），不转储数据。

`-s`
`--schema-only`

只转储对象定义（模式），而不转储数据。

`-S` *`username`*
`--superuser=`*`username`*

指定在禁用触发器时要使用的超级用户名。只有在使用 `--disable-triggers` 时才相关。（通常更好的做法是省略此选项，而以超级用户身份运行生成的脚本。）

`-t`
`--tablespaces-only`

只转储表空间，不转储数据库或角色。

`-v`
`--verbose`

指定详细模式。这会让 pg_dumpall 把开始/停止时间写入转储文件，并将进度消息输出到标准错误。它还会启用 pg_dump 的详细输出。

`-V`
`--version`

打印 pg_dumpall 的版本并退出。

`-x`
`--no-privileges`
`--no-acl`

此选项供就地升级工具使用。不建议也不支持将其用于其他用途。
该选项的行为在将来的发行版中可能会在不另行通知的情况下发生变化。

`--binary-upgrade`

此选项供原地升级工具使用。不建议将其用于其他目的，也不支持这样用。
该选项的行为可能在未来的版本中不经通知而改变。

`--column-inserts`
`--attribute-inserts`

将数据转储为带有显式列名的 `INSERT` 命令（`INSERT INTO table (column, ...) VALUES ...`）。这会使恢复变得非常缓慢；它主要用于生成可以装入非 PostgreSQL 数据库的转储。

`--disable-dollar-quoting`

此选项禁用函数体中的 dollar quoting，并强制改用 SQL 标准字符串语法对它们进行引用。

`--disable-triggers`

此选项只在创建只包含数据而不包含模式的转储时才相关。它指示 pg_dumpall 在输出中包含一些命令，以便在重新载入数据时临时禁用目标表上的触发器。
如果这些表上存在不希望在数据重新载入期间触发的引用完整性检查或其他触发器，
请使用此选项。

目前，为 `--disable-triggers` 输出的这些命令必须由超级用户执行。因此，还应通过 `-S` 指定一个超级用户名，或者更好的做法是确保以超级用户身份运行生成的脚本。

`--inserts`

将数据转储为 `INSERT` 命令（而不是 `COPY`）。这会使恢复非常缓慢；它主要用于生成可以装入非 PostgreSQL 数据库的转储。注意，如果重新安排了列顺序，恢复可能会彻底失败。`--column-inserts` 选项更安全，但速度更慢。

`--lock-wait-timeout=`*`timeout`*

在转储开始时，不要无限等待获取共享表锁。如果无法在指定的 *`timeout`* 内锁定某个表，就让转储失败。超时可以用 `SET statement_timeout` 接受的任意格式指定。允许的值因被转储的服务器版本而异，但从 7.3 起，所有版本都接受以毫秒为单位的整数。对 7.3 之前的服务器进行转储时，此选项会被忽略。

`--no-security-labels`

不输出创建表空间的命令，也不输出用于为对象选择表空间的命令。使用此选项时，
所有对象在恢复时都会创建在当时默认的表空间中。

`--no-tablespaces`

不要转储不记录 WAL 的表的内容。此选项不影响是否转储表定义（模式）；它只会抑制转储表数据。

`--no-unlogged-table-data`

不要转储不记录 WAL 的表的内容。此选项不影响是否转储表定义（模式）；它只会抑制转储表数据。

`--quote-all-identifiers`

强制为所有标识符加引号。当从其 PostgreSQL 主版本与 pg_dumpall 不同的服务器转储数据库时，或者当输出打算装入另一主版本的服务器中时，建议使用此选项。默认情况下，pg_dumpall 只会给在其自身主版本中属于保留字的标识符加引号。处理其他版本的服务器时，这有时会带来兼容性问题，因为它们的保留字集合可能略有不同。使用 `--quote-all-identifiers` 可以避免这类问题，但代价是转储脚本更难阅读。

`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是用 `ALTER OWNER` 命令来确定对象所有权。这会让转储更符合标准，但根据转储中对象的历史，可能无法正确恢复。

`-?`
`--help`

显示关于 pg_dumpall 命令行参数的帮助信息，并退出。

下列命令行选项控制数据库连接参数。

`-h` *`host`*
`--host=`*`host`*

指定运行数据库服务器的机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。默认值取自 `PGHOST` 环境变量（如果已设置）；否则会尝试使用 Unix 域套接字连接。

`-l` *`dbname`*
`--database=`*`dbname`*

指定连接来转储全局对象并发现还应转储哪些数据库的数据库名。如果未指定，则使用`postgres`数据库；如果该数据库不存在，则使用`template1`。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。默认值取自`PGPORT`环境变量（如果已设置），否则使用编译时默认值。

`-U` *`username`*
`--username=`*`username`*

用于连接的用户名。

`-w`
`--no-password`

绝不提示输入密码。如果服务器要求密码认证，而又无法通过`.pgpass`文件等其他方式获得密码，则连接尝试会失败。该选项适用于批处理作业和脚本，因为这些场景下没有用户可以输入密码。

`-W`
`--password`

强制pg_dumpall在连接数据库之前提示输入密码。

此选项从来都不是必需的，因为如果服务器要求密码认证，pg_dumpall会自动提示输入密码。不过，pg_dumpall会浪费一次连接尝试来发现服务器需要密码。在某些情况下，键入`-W`以避免这次额外的连接尝试是值得的。

注意，对每个要转储的数据库都会再次提示输入密码。通常，最好设置一个`~/.pgpass`文件，而不是依赖手工输入密码。

`--role=rolename`

指定一个角色名，用于创建转储。此选项会使pg_dumpall在连接数据库后发出`SET ROLE rolename`命令。当已认证用户（由`-U`指定）缺少pg_dumpall所需权限，但可以切换到具备所需权限的角色时，这很有用。有些安装环境不允许直接以超级用户身份登录，而使用此选项可以在不违反该策略的情况下完成转储。

环境

`PGHOST`
`PGOPTIONS`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他PostgreSQL工具一样，此实用程序也使用libpq支持的环境变量（见第 31.13 节）。

注解

由于pg_dumpall在内部调用pg_dump，因此某些诊断消息会提到pg_dump。

恢复完成后，最好对每个数据库运行`ANALYZE`，以便优化器获得有用的统计信息。也可以运行`vacuumdb -a -z`来分析所有数据库。

pg_dumpall要求在恢复之前，所有必需的表空间目录都已经存在；否则，位于非默认位置的数据库在创建时将会失败。

示例

要转储所有数据库：

```
$ pg_dumpall > db.out
```

要从此文件重新装载数据库，可以使用：

```
$ psql -f db.out postgres
```

（这里连接哪个数据库并不重要，因为 pg_dumpall 创建的脚本文件会包含适当的命令，用于创建并连接到已保存的数据库。）

另见

有关可能出现的错误情况，请参见pg_dump。

pg_restore

pg_restore — restore a PostgreSQL database from an archive file created by pg_dump

大纲

`pg_restore` [*connection-option*...] [*option*...] [*filename*]

描述

pg_restore 是一个用于从pg_dump以非纯文本格式创建的归档中恢复 PostgreSQL 数据库的工具。它会发出必要的命令，将数据库重建为保存时的状态。归档文件还允许 pg_restore 有选择地恢复其中的内容，甚至在恢复前重新排列各项的顺序。归档文件被设计为可跨体系结构移植。

pg_restore 可以以两种模式运行。如果指定了数据库名，pg_restore 就会连接到该数据库，并将归档内容直接恢复到数据库中。否则，它会创建一个脚本，其中包含重建数据库所需的 SQL 命令，并将其写入文件或标准输出。这种脚本输出等价于 pg_dump 的纯文本输出格式。因此，一些控制输出的选项与 pg_dump 的选项相对应。

显然，pg_restore 无法恢复归档文件中不存在的信息。例如，如果归档是使用“将数据转储为 `INSERT` 命令”选项生成的，pg_restore 就不能使用 `COPY` 语句装载数据。

选项

pg_restore接受以下命令行参数：

filename

指定要恢复的归档文件所在位置（对于目录格式归档则为目录）。如果未指定，则使用标准输入。

`-a`
`--data-only`

只恢复数据，不恢复模式（数据定义）。

`-c`
`--clean`

在重新创建数据库对象之前清理（删除）它们。

`-C`
`--create`

在恢复之前先创建数据库。（使用此选项时，由 `-d` 指定的数据库仅用于发出初始的 `CREATE DATABASE` 命令。所有数据都会恢复到归档中出现的数据库名中。）

`-d` *dbname*
`--dbname=`*dbname*

连接到数据库 *dbname*，并直接恢复到该数据库中。

`-e`
`--exit-on-error`

一个已弃用、现在会被忽略的选项。

`-f` *`filename`*
`--file=`*`filename`*

指定生成脚本的输出文件，或者在与 `-l` 一起使用时指定列表输出文件。默认为标准输出。

`-F` *`format`*
`--format=`*`format`*

Specify format of the archive. It is not necessary to specify the format, since pg_restore will determine the format automatically. If specified, it can be one of the following:

`c`
`custom`

归档采用 pg_dump 的自定义格式。

`d`
`directory`

归档是目录格式归档。

`t`
`tar`

归档是 `tar` 归档。

`-i`
`--ignore-version`

一个已废弃的选项，现在会被忽略。

`-I` *`index`*
`--index=`*`index`*

仅恢复指定名称的索引的定义。

`-j` *`number-of-jobs`*
`--jobs=`*`number-of-jobs`*

使用多个并发作业执行pg_restore中最耗时的步骤 — 装载数据、创建索引或创建约束 —。此选项可以显著缩短将大型数据库恢复到运行在多处理器机器上的服务器所需的时间。

每个作业都是一个进程或一个线程，具体取决于操作系统，
并使用一个到服务器的独立连接。

此选项的最佳取值取决于服务器、客户端和网络的硬件配置。影响因素包括 CPU 核心数和磁盘配置。一个较好的起点是服务器上的 CPU 核心数，但在很多情况下，更大的值也可能带来更快的恢复速度。当然，取值过高会因为频繁争用而导致性能下降。

只有自定义归档格式支持此选项。输入文件必须是常规文件（例如不能是管道）。
当生成脚本而不是直接连接到数据库服务器时，此选项会被忽略。另外，多个作业不能与 `--single-transaction` 选项一起使用。

-l
--list

列出归档的内容。此操作的输出可用作 -L 选项的输入。请注意，如果在 -l 中使用 -n 或 -t 等过滤开关，它们会限制被列出的项目。

-L *list-file*
--use-list=*list-file*

仅恢复在 *list-file* 中列出的归档元素，并按它们在文件中出现的顺序进行恢复。请注意，如果在 -L 中使用 -n 或 -t 等过滤开关，它们还会进一步限制被恢复的项目。

list-file 通常是通过编辑先前 -l 操作的输出来创建的。可以移动或删除行，也可以通过在行首放置分号（;）将其注释掉。示例见下文。

-n *namespace*
--schema=*schema*

仅恢复指定模式中的对象。这可以与 -t 选项结合使用，只恢复特定表。

-O
--no-owner

不输出用于设置对象所有权以匹配原始数据库的命令。默认情况下，pg_restore 会发出 ALTER OWNER 或 SET SESSION AUTHORIZATION 语句，为所创建的模式元素设置所有权。除非对数据库的初始连接由超级用户发起（或者由脚本中所有对象的同一所有者发起），否则这些语句会失败。使用 -O 时，初始连接可以使用任意用户名，并且该用户将拥有所有创建的对象。

-P *function-name(argtype [, ...])*
--function=*function-name(argtype [, ...])*

仅恢复指定函数。请务必按转储文件目录中显示的形式准确拼写函数名和参数。

-R
--no-reconnect

此选项已过时，但为了向后兼容仍然被接受。

-s
--schema-only

只恢复模式（数据定义），不恢复数据（表内容）。当前的序列值也不会被恢复。（不要将它与 --schema 选项混淆，那里“schema”一词的含义不同。）

-S *username*
--superuser=*username*

指定在禁用触发器时要使用的超级用户用户名。仅当使用 --disable-triggers 时才相关。

-t *table*
--table=*table*

只恢复指定表的定义和/或数据。此选项可以与 -n 选项结合使用，以指定模式。

`-T` *trigger*
`--trigger=`*trigger*

仅恢复指定名称的触发器。

`-v`
`--verbose`

指定详细模式。

`-V`
`--version`

打印 pg_restore 的版本并退出。

`-x`
`--no-privileges`
`--no-acl`

阻止恢复访问权限（grant/revoke 命令）。

`-1`
`--single-transaction`

将恢复作为单个事务执行（也就是把发出的命令包裹在 `BEGIN`/`COMMIT` 中）。这可确保要么所有命令都成功完成，要么不应用任何更改。此选项隐含 `--exit-on-error`。

`--disable-triggers`

此选项仅在执行仅数据恢复时才相关。它会指示 pg_restore 在恢复数据期间发出命令，临时禁用目标表上的触发器。
如果你不希望在恢复数据期间触发表上的引用完整性检查或其他触发器，请使用此选项。

目前，为 `--disable-triggers` 发出的命令必须以超级用户身份执行。因此，你还应通过 `-S` 指定超级用户用户名，或者更好的是，直接以 PostgreSQL 超级用户身份运行 pg_restore。

`--no-data-for-failed-tables`

默认情况下，即使表的创建命令失败（例如因为表已经存在），表数据仍然会被恢复。
使用此选项时，这类表的数据会被跳过。如果目标数据库已经包含所需的表内容，
这种行为会很有用。例如，PostgreSQL 扩展（如 PostGIS）的辅助表可能已经装载到目标数据库中；指定此选项可以防止向这些表装载重复或过时的数据。

此选项仅在直接恢复到数据库时有效，而在生成 SQL 脚本输出时无效。

`--no-security-labels`

不输出用于恢复安全标签的命令，即使归档中包含它们。

`--no-tablespaces`

不输出用于选择表空间的命令。使用此选项后，
所有对象都会在恢复期间默认的表空间中创建。

`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是使用 `ALTER OWNER` 命令来确定对象所有权。这样会使转储结果更符合标准，但根据转储中对象的历史，可能无法正确恢复。

`-?`
`--help`

显示关于 pg_restore 命令行参数的帮助信息，并退出。

pg_restore还接受以下用于连接参数的命令行参数：

`-h` *host*
`--host=`*host*

指定服务器运行所在机器的主机名。如果该值以斜杠开头，则它被用作 Unix 域套接字目录。默认值取自 `PGHOST` 环境变量（如果已设置），否则会尝试使用 Unix 域套接字连接。

`-p` *port*
`--port=`*port*

指定服务器监听连接所使用的 TCP 端口或本地 Unix 域套接字文件扩展名。默认值取自 `PGPORT` 环境变量（如果已设置），否则使用编译时的默认值。

`-U` *username*
`--username=`*username*

连接时使用的用户名。

`-w`
`--no-password`

从不发出密码提示。如果服务器要求使用密码认证，而又无法通过其他方式（如 `.pgpass` 文件）获得密码，则连接尝试会失败。此选项适用于批处理任务和脚本中没有用户在场输入密码的场景。

`-W`
`--password`

强制 pg_restore 在连接数据库之前提示输入密码。

此选项实际上并非必需，因为如果服务器要求密码认证，pg_restore 会自动提示输入密码。不过，pg_restore 会浪费一次连接尝试来发现服务器需要密码。在某些情况下，键入 `-W` 以避免这一次额外的连接尝试是值得的。

`--role=`*rolename*

指定执行恢复时要使用的角色名。此选项会使 pg_restore 在连接数据库后发出 `SET ROLE` *rolename* 命令。当已认证用户（由 `-U` 指定）缺少 pg_restore 所需的权限，但可以切换到具有所需权限的角色时，此选项会很有用。有些安装环境禁止直接以超级用户登录，而使用此选项可以在不违反该策略的情况下执行恢复。

环境

`PGHOST`
`PGOPTIONS`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他 PostgreSQL 工具一样，此工具也使用 libpq 支持的环境变量（见第 31.13 节）。

诊断

当使用 `-d` 选项指定直接数据库连接时，pg_restore 会在内部执行 SQL 语句。如果运行 pg_restore 时遇到问题，请确保你能够使用例如psql从该数据库中查询信息。此外，libpq 前端库使用的任何默认连接设置和环境变量也都会生效。

注解

如果你的安装在 `template1` 数据库中有任何本地添加内容，请务必将 pg_restore 的输出装载到一个真正空的数据库中；否则很可能会因为这些附加对象的重复定义而报错。
要创建一个不带任何本地添加内容的空数据库，应从 `template0` 而不是 `template1` 复制，例如：

```
CREATE DATABASE foo WITH TEMPLATE template0;
```

pg_restore 的局限性详述如下。

- 当把数据恢复到一个预先存在的表中，并且使用了 `--disable-triggers` 选项时，pg_restore 会在插入数据前发出命令，禁用用户表上的触发器，然后在数据插入后再发出命令重新启用它们。如果恢复在中途停止，系统目录可能会处于错误状态。
- pg_restore 不能有选择地恢复大对象；例如，不能只恢复属于某个特定表的大对象。如果归档中包含大对象，那么要么所有大对象都会被恢复，要么在通过 `-L`、`-t` 或其他选项排除它们时一个也不会恢复。

关于 pg_dump 的局限性的细节也可参见pg_dump文档。

恢复完成后，最好对每个恢复的表运行`ANALYZE`，以便优化器获得有用的统计信息；
更多信息请参阅第 23.1.3 节和第 23.1.5 节。

示例

假设我们已经把一个名为 `mydb` 的数据库转储到一个自定义格式转储文件中：

```
$ pg_dump -Fc mydb > db.dump
```

要删除该数据库并从转储中重新创建它：

```
$ dropdb mydb
$ pg_restore -C -d postgres db.dump
```

`-d` 选项中指定的数据库可以是集簇中任何一个已存在的数据库；pg_restore 只用它来为 `mydb` 发出 `CREATE DATABASE` 命令。使用 `-C` 时，数据总是恢复到转储文件中出现的那个数据库名中。

要将转储重新装载到一个新数据库中，其名称为 `newdb`：

```
$ createdb -T template0 newdb
$ pg_restore -d newdb db.dump
```

注意，我们没有使用 `-C`，而是直接连接到要恢复到的数据库。还要注意，我们从 `template0` 而不是 `template1` 克隆新数据库，以确保它最初为空。

要重新排列数据库项的顺序，首先需要转储归档的目录：

```
$ pg_restore -l db.dump > db.list
```

列表文件由一个头部和每个项各占一行的内容组成，例如：

```
;
; Archive created at Mon Sep 14 13:55:39 2009
;     dbname: DBDEMOS
;     TOC Entries: 81
;     Compression: 9
;     Dump Version: 1.10-0
;     Format: CUSTOM
;     Integer: 4 bytes
;     Offset: 8 bytes
;     Dumped from database version: 8.3.5
;     Dumped by pg_dump version: 8.3.8
;
;
; Selected TOC Entries:
;
3; 2615 2200 SCHEMA - public pasha
1861; 0 0 COMMENT - SCHEMA public pasha
1862; 0 0 ACL - public pasha
317; 1247 17715 TYPE public composite pasha
319; 1247 25899 DOMAIN public domain0 pasha
```

分号表示一条注释的开始，而行首的数字表示分配给每个项的内部归档 ID。

文件中的行可以被注释掉、删除并重新排序。例如：

```
10; 145433 TABLE map_resolutions postgres
;2; 145344 TABLE species postgres
;4; 145359 TABLE nt_header postgres
6; 145402 TABLE species_records postgres
;8; 145416 TABLE ss_old postgres
```

可以将这样的文件作为 pg_restore 的输入，这样它就只会按该顺序恢复项 10 和 6：

```
$ pg_restore -L db.list db.dump
```

参见

pg_dump, pg_dumpall, psql

psql

psql — PostgreSQL interactive terminal

大纲

psql [*option*...] [*dbname* [*username*]]

描述

psql是PostgreSQL的一个基于终端的前端。它使你能够交互式地输入查询，将其发送给PostgreSQL，并查看查询结果。也可以从文件提供输入。此外，
它还提供了若干元命令和多种类似 shell 的特性，以便于编写脚本和自动化执行各种任务。

选项

-a
--echo-all

在读入时将所有非空输入行打印到标准输出（不适用于交互式行读取）。
这等效于把变量ECHO设置为 all。

-A
--no-align

切换到非对齐输出模式（默认输出模式是对齐的）。

-c *command*
--command=*command*

指定psql执行一个命令字符串*command*，然后退出。这在 shell 脚本中很有用。使用此选项时，启动文件（psqlrc和~/.psqlrc）会被忽略。

*command*必须是一个服务器完全可解析的命令字符串（即不包含psql专有的特性）
或者单个反斜线命令。因此不能用这个选项混合SQL和psql元命令。要那样做，
可以把字符串用管道输送到psql中，例如echo '\x \\ SELECT * FROM foo;' | psql（\\是分隔符元命令）。

如果命令字符串包含多条 SQL 命令，它们会在单个事务中处理，除非字符串中包含显式的BEGIN/COMMIT命令将其分成多个事务。
这与把同一字符串送入psql标准输入时的行为不同。此外，只有最后一条 SQL 命令的结果会被返回。

由于这些历史行为，在 -c 字符串中放入多条命令常常会产生意外结果。最好把多条命令送入psql 的标准输入，可以像上面所示使用 echo，也可以使用 shell 的 here-document，例如：

```
psql <<EOF
\x
SELECT * FROM foo;
EOF
```

-d *dbname*
--dbname=*dbname*

指定要连接的数据库的名称。这等效于指定*dbname*为命令行上的第一个非选项参数。

如果这个参数包含一个 = 符号，它将被当作一个 *conninfo* 字符串处理。更多信息见第 31.1 节。

-e
--echo-queries

也把发送到服务器的所有 SQL 命令复制到标准输出。这等效于把变量ECHO设置为queries。

-E
--echo-hidden

回显\d以及其他反斜线命令生成的实际查询。可以用它来学习psql的内部操作。这等效于把变量ECHO_HIDDEN设置为on。

-f *filename*
--file=*filename*

使用文件*filename*作为命令来源，而不是交互式地读取命令。文件处理完毕后，psql终止。这在很多方面等价于元命令\i。

如果*filename*是-（连字符），则会读取标准输入，直到遇到 EOF 指示或\q元命令。不过请注意，这种情况下不会使用 Readline（很像指定了-n时的情况）。

使用这个选项与写成psql < *filename*有细微差别。通常两种形式都会得到你期望的结果，但使用-f可以启用一些有用的特性，例如带行号的错误消息。
使用这个选项也还有一点机会降低启动开销。另一方面，使用 shell 输入重定向的形式在理论上能保证得到与你手工逐行输入时完全相同的输出。

-F *separator*
--field-separator=*separator*

使用*separator*作为非对齐输出的字段分隔符。这等效于\pset fieldsep或者\f。

-h *hostname*
--host=*hostname*

指定运行服务器的机器的主机名。如果该值以斜线开头，则它会被用作 Unix 域套接字所在的目录。

-H
--html

切换到HTML表格输出模式。这等效于\pset format html或者\H命令。

-l
--list

列出所有可用的数据库，然后退出。其他非连接选项会被忽略。这与元命令\list类似。

`-L` *`filename`*
`--log-file=`*`filename`*

除了把所有查询输出写到普通输出目标之外，还写到文件*`filename`*中。

`-n`
`--no-readline`

不要使用Readline进行行编辑，也不要使用命令历史记录。
这有助于在剪切和粘贴时关闭TAB 补全。

`-o` *`filename`*
`--output=`*`filename`*

把所有查询输出放到文件*`filename`*中。这等效于命令`\o`。

`-p` *`port`*
`--port=`*`port`*

指定服务器用于监听连接的 TCP 端口或者本地 Unix 域套接字文件扩展名。默认是`PGPORT`环境变量的值，如果没有设置，则默认为编译时指定的端口号（通常是5432）。

`-P` *`assignment`*
`--pset=`*`assignment`*

以 `\pset` 的形式指定打印选项。注意，这里必须用一个等号而不是空格来分隔名称和值。例如，要把输出格式设置为 LaTeX，可以写 `-P format=latex`。

`-q`
`--quiet`

指定psql应该安静地工作。默认情况下，它会打印出欢迎消息和各种提示信息。
如果使用了这个选项，以上那些就都不会输出。在使用`-c`选项时，配合这个选项很有用。
这等效于设置变量`QUIET`为`on`。

`-R` *`separator`*
`--record-separator=`*`separator`*

把*`separator`*用作非对齐输出的记录分隔符。这等效于`\pset recordsep`命令。

`-s`
`--single-step`

运行在单步模式中。这意味着在每个命令被发送给服务器之前都会提示用户，
并允许取消执行。使用这个选项可以调试脚本。

`-S`
`--single-line`

运行在单行模式中，其中换行符会终止一个 SQL 命令，就像分号的作用一样。

> **注意**
>
> 这种模式是为坚持使用它的用户提供的，但并不一定值得推荐。特别是，
> 如果在一行中混合了SQL和元命令，对于没有经验的用户来说，
> 它们的执行顺序未必总是清楚的。

`-t`
`--tuples-only`

关闭打印列名和结果行计数页脚等。这等效于`\t`命令。

`-T` *`table_options`*
`--table-attr=`*`table_options`*

指定要放在HTML `table`标签内的选项。详见`\pset`。

`-U` *`username`*
`--username=`*`username`*

作为用户*`username`*而不是默认用户连接到数据库（当然，你必须具有这样做的权限）。

`-v` *`assignment`*
`--set=`*`assignment`*
`--variable=`*`assignment`*

执行一次变量赋值，和`\set`元命令相似。
注意你必须在命令行上用等号分隔名字和值（如果有）。要取消变量的设置，
去掉等号就行。要把一个变量设为空字符串，使用等号但是去掉值。
这些赋值在启动的非常早期阶段完成，因此为内部目的保留的变量可能会在稍后被覆盖。

`-V`
`--version`

打印psql版本并且退出。

`-w`
`--no-password`

永不发出密码提示。如果服务器要求密码认证，而密码又无法从其他来源获得，例如`.pgpass`文件，则连接尝试将失败。这个选项对于批处理任务和脚本很有用，
因为那时通常没有用户在场输入密码。

请注意，这个选项在整个会话期间都会保持生效，因此它不仅影响初始连接尝试，
也会影响元命令`\connect`。

`-W`
`--password`

强制psql在连接数据库之前提示输入密码。

这个选项并非必不可少，因为当服务器要求密码认证时，psql会自动提示输入密码。不过，
psql需要先浪费一次连接尝试，才能知道服务器要求密码。在某些情况下，显式指定
`-W`值得用来避免这次额外的连接尝试。

请注意，这个选项在整个会话期间都会保持生效，因此它不仅影响初始连接尝试，
也会影响元命令`\connect`。

`-x`
`--expanded`

打开扩展表格式模式。这等效于`\x`命令。

`-X`,
`--no-psqlrc`

不读取启动文件（既不读取系统范围的`psqlrc`文件，也不读取用户的`~/.psqlrc`文件）。

`-1`
`--single-transaction`

当 psql 使用 `-f` 选项执行一个脚本时，加上这个选项会在脚本前后包上 `BEGIN`/`COMMIT`，把它作为单个事务执行。这确保要么所有命令都成功完成，要么不应用任何更改。

如果脚本本身使用了`BEGIN`、`COMMIT`或`ROLLBACK`，这个选项就不会产生期望的效果。
另外，如果脚本中包含不能在事务块中执行的命令，
指定这个选项将导致该命令（进而整个事务）失败。

`-?`
`--help`

显示有关psql命令行参数的帮助并且退出。

退出状态

如果psql正常结束，它会向 shell 返回 0；如果它自身发生致命错误（例如内存耗尽、找不到文件），则返回 1；如果到服务器的连接发生故障且该会话不是交互式的，则返回 2；如果脚本中发生错误且变量`ON_ERROR_STOP`已设置，则返回 3。

用法

连接到数据库

psql 是常规的 PostgreSQL 客户端应用程序。要连接到数据库，你需要知道目标数据库名称、服务器的主机名和端口号，以及要以哪个数据库用户名连接。psql 可以通过命令行选项 `-d`、`-h`、`-p` 和 `-U` 分别指定这些参数。如果遇到一个不属于任何选项的参数，它将被解释为数据库名（如果数据库名已经给出，则解释为数据库用户名）。并非所有这些选项都是必需的；
它们都有有用的默认值。如果省略主机名，psql 将通过 Unix 域套接字连接到本地主机上的服务器，而在没有 Unix 域套接字的机器上则通过 TCP/IP 连接到 `localhost`。默认端口号在编译时确定。由于数据库服务器使用相同的默认值，因此在大多数情况下不必指定端口。
默认用户名是你的操作系统用户名，默认数据库名也是如此。请注意，
你不能随意以任意数据库用户名连接到任意数据库。
数据库管理员应当已经告知你拥有的访问权限。

当默认值不完全合适时，可以通过把环境变量`PGDATABASE`、`PGHOST`、`PGPORT`和`PGUSER`设置为适当的值来少敲一些键盘（额外的环境变量见第 31.13 节）。另外，准备一个`~/.pgpass`文件也很方便，这样就不必经常手工输入密码。详见第 31.14 节。

指定连接参数的另一种方法是使用一个 *conninfo* 字符串，它可用来代替数据库名。这种机制可以让你对连接进行非常细致的控制。例如：

```
$ psql "service=myservice sslmode=require"
```

用这种方式，你也可以把 LDAP 用于第 31.16 节中描述的连接参数查找。所有可用连接选项的更多信息见第 31.1 节。

如果由于任何原因（例如权限不足、服务器没有在目标主机上运行等）导致连接无法建立，
psql将返回一个错误并且终止。

如果标准输入或标准输出中至少有一个是终端，那么 psql 会把客户端编码设置为“auto”，从区域设置中检测合适的客户端编码（在 Unix 系统上是 `LC_CTYPE` 环境变量）。如果结果不符合预期，可以使用环境变量 `PGCLIENTENCODING` 覆盖客户端编码。

输入 SQL 命令

在正常操作时，psql会提供一个提示符，该提示符是
psql当前连接到的数据库名称后面跟上字符串`=>`。例如：

```
$ psql testdb
psql (9.1.24)
Type "help" for help.

testdb=>
```

在提示符下，用户可以输入SQL命令。通常，当遇到表示命令结束的分号时，
输入的内容就会被发送给服务器。行结束并不会终止一条命令，因此为了提高清晰度，
命令可以分布在多行上。如果命令被发送并成功执行，其结果就会显示在屏幕上。

每当执行命令时，psql也会轮询由LISTEN和NOTIFY生成的异步通知事件。

元命令

你输入到psql中的任何以未加引号的反斜线开始的东西都是一个psql元命令，它们由
psql自行处理。这些命令让psql对管理和编写脚本更有用。
元命令常常被称作斜线或者反斜线命令。

psql命令的格式是用反斜线后面直接跟上一个命令动词，然后是一些参数。
参数与命令动词和其他参数之间用任意多个空白字符分隔开。

要在参数中包含空白，可以用单引号将它括起来。要在这样的参数中包含一个单引号，
可以写两个单引号。单引号中的内容还会接受类似 C 语言的替换：`\n`（换行）、`\t`（制表符）、`\digits`（八进制）和 `\xdigits`（十六进制）。

如果未加引号的参数以冒号（`:`）开头，它会被当作一个 psql 变量，该变量的值将被用作参数。如果变量名被单引号包围（例如 `:'var'`），它会被作为一个 SQL 字面量转义，结果将被用作参数。如果变量名被双引号包围，它会被作为一个 SQL 标识符转义，结果将被用作参数。

用反引号（`` ` ``）包围的参数会被当作传给 shell 的命令行。该命令的输出（去掉末尾的换行符）将被用作参数值。上述转义序列在反引号中同样适用。

有些命令把SQL标识符（例如表名）作为参数。
这些参数遵循SQL的语法规则：未加引号的字母会被强制转换为小写，而双引号（`"`）
可以保护字母不发生大小写转换，并允许在标识符中包含空白。在双引号内，
成对的双引号会在结果名称中折叠成一个双引号。例如，`FOO"BAR"BAZ`会被解释为`foo BARbaz`，而`"A weird"" name"`会变成`A weird" name`。

参数解析会在行尾或遇到另一个未加引号的反斜线时停止。
未加引号的反斜线会被视为新元命令的开始。特殊序列`\\`（两个反斜线）表示参数结束，
并继续解析SQL命令（如果还有）。通过这种方式，
SQL命令和psql命令可以自由地混合在同一行中。但无论如何，
元命令的参数都不能延续到下一行。

定义了以下元命令：

`\a`

如果当前表格输出格式是非对齐，则切换为对齐；否则切换为非对齐。
保留此命令是为了向后兼容。更通用的解决方案请参见`\pset`。

`\c` or `\connect [ -reuse-previous=on|off ] [ dbname [ username ] [ host ] [ port ] | conninfo ]`

建立到PostgreSQL服务器的新连接。连接参数既可以使用按位置指定的语法，也可以使用中详述的conninfo第 31.1 节连接串。

当命令省略数据库名、用户、主机或端口时，新连接可以复用前一个连接的值。
默认情况下，除了在处理`conninfo`串时，其余情况都会复用前一个连接的值。传递 `-reuse-previous=on`或 `-reuse-previous=off`作为第一个参数可以覆盖该默认设置。
当命令既不指定也不复用特定参数时，将使用 libpq的默认值。将 *dbname*、*username*、*host*和*port*中的任何一个指定为`-`，都等价于省略该参数。

如果成功建立了新连接，则关闭先前的连接。如果连接尝试失败（用户名错误、
访问被拒绝等），当psql处于交互模式时，会保留先前的连接。但在执行非交互式脚本时，
会立即报错并停止处理。这一区别一方面使用户能够方便地应对输入错误，
另一方面也作为安全机制，防止脚本意外地作用于错误的数据库。

例如：

```
=> \c mydb myuser host.dom 6432
=> \c service=foo
=> \c "host=localhost port=5432 dbname=mydb connect_timeout=10
 sslmode=disable"
```

`\C [ title ]`

设置作为查询结果打印的表的标题，或取消此类标题。该命令等价于`\pset title title`。（此命令的名称源自“caption”，因为它过去只用于设置HTML表的标题。）

`\cd [ directory ]`

将当前工作目录更改为*directory*。如果没有参数，则切换到当前用户的主目录。

> **提示**
>
> 要打印当前工作目录，请使用`\! pwd`。

`\conninfo`

输出当前数据库连接的信息。

`\copy { table [(column_list)] | (query) } { from | to } { filename | stdin | stdout | pstdin | pstdout } [with] [binary] [oids] [delimiter [as] 'character'] [null [as] 'string'] [csv [header] [quote [as] 'character'] [escape [as] 'character'] [force quote column_list | *] [force not null column_list]]`

执行前端（客户端）复制。该操作会运行一个SQL COPY命令，但并不是由服务器读取或写入指定文件，而是由psql读取或写入文件，并在服务器与本地文件系统之间转送数据。
这意味着文件可访问性和权限取决于本地用户，而不是服务器，也不需要 SQL 超级用户权限。

该命令的语法与 SQL 的COPY命令类似。注意，正因为如此，`\copy` 命令适用特殊的解析规则。特别是，变量替换规则和反斜线转义在这里不适用。

`\copy ... from stdin | to stdout` 分别基于命令的输入和输出进行读取/写入。所有行都从发出该命令的同一输入源读取，直到读到仅包含 `\.` 的一行或流到达 EOF 为止。输出会发送到与命令输出相同的位置。若要读写psql的标准输入或输出，请使用 `pstdin` 或 `pstdout`。这个选项适合在 SQL 脚本文件中以内联方式填充表。

> **提示**
>
> 这个操作不像SQL的`COPY`命令那样高效，因为所有数据都必须经过客户端/服务器连接。对于大量数据，SQL命令可能更可取。

`\copyright`

显示PostgreSQL的版权和分发条款。

`\d[S+] [ pattern ]`

对于每个匹配*pattern*的关系（表、视图、物化视图、索引、序列或外部表）或复合类型，显示所有列及其类型、表空间（如果不是默认表空间），以及`NOT NULL`或默认值等特殊属性。还会显示关联的索引、约束、规则和触发器。对于外部表，还会显示关联的外部服务器。（“匹配模式”的定义见下面的模式。）

`\d+` 形式的命令与之相同，但会显示更多信息：表各列关联的注释、表中是否存在 OID、视图定义（如果关系是视图），以及通用选项（如果关系是外部表）。

默认只显示用户创建的对象；提供模式或`S`修饰符可包含系统对象。

> **注意**
>
> 如果使用`\d`而没有*pattern*参数，它等同于`\dtvmsE`，它将显示所有可见的表、视图、物化视图、序列和外部表的列表。这纯粹是一种便利措施。

`\da[S] [ pattern ]`

列出聚合函数及其返回类型和操作的数据类型。如果指定了*pattern*，则只显示名称匹配该模式的聚合函数。默认只显示用户创建的对象；提供模式或`S`修饰符可包含系统对象。

`\db[+] [ pattern ]`

列出表空间。如果指定了*pattern*，则只显示名称匹配该模式的表空间。如果在命令名后附加`+`，每个对象还会与它关联的权限一起列出。

`\dc[S] [ pattern ]`

列出字符集编码之间的转换。如果指定了 *pattern*，则只列出名称匹配该模式的转换。默认只显示用户创建的对象；提供模式或 `S` 修饰符可包含系统对象。

`\dC [ pattern ]`

列出类型转换。如果指定了 *pattern*，则只列出源类型或目标类型匹配该模式的类型转换。

\dd[S] [*pattern*]

显示匹配 *pattern* 的对象的描述，如果没有给出参数则显示所有可见对象的描述。但无论哪种情况，都只列出有描述的对象。默认只显示用户创建的对象；提供模式或 S 修饰符可包含系统对象。“对象”涵盖聚合、函数、操作符、类型、关系（表、视图、索引、序列）、大对象、规则和触发器。例如：

```
=> \dd version
                      Object descriptions
   Schema   |  Name   |  Object  |        Description
------------+---------+----------+---------------------------
 pg_catalog | version | function | PostgreSQL version string
(1 row)
```

可以用COMMENT SQL命令为对象创建描述。

\ddp [*pattern*]

列出默认访问权限设置。对于默认权限设置已偏离内置默认值的每个角色（以及模式，如果适用），都会显示一个条目。如果指定了*pattern*，则只列出角色名或模式名匹配该模式的条目。

ALTER DEFAULT PRIVILEGES命令用于设置默认访问权限。权限显示的含义在GRANT中有解释。

\dD[S] [*pattern*]

列出域。如果指定了 *pattern*，则只显示名称匹配该模式的域。默认只显示用户创建的对象；提供模式或 S 修饰符可包含系统对象。

\dE[S+] [*pattern*]
\di[S+] [*pattern*]
\ds[S+] [*pattern*]
\dt[S+] [*pattern*]
\dv[S+] [*pattern*]

在这组命令中，字母 E、i、s、t 和 v 分别代表外部表、索引、序列、表和视图。可以按任意顺序指定这些字母中的任意一个或全部，以获取相应类型的对象列表。例如，\dit 列出索引和表。如果在命令名后附加 +，还会列出每个对象在磁盘上的物理大小以及关联的描述（如果有）。如果指定了 *pattern*，则只列出名称匹配该模式的对象。默认只显示用户创建的对象；提供模式或 S 修饰符可包含系统对象。

\des[+] [*pattern*]

列出外部服务器（助记词：“external servers”）。如果指定了*pattern*，则只列出名称匹配该模式的服务器。如果使用\des+形式，则显示每个服务器的完整说明，包括服务器的ACL、类型、版本、选项和描述。

\det[+] [*pattern*]

列出外部表（助记词：“external tables”）。如果指定了 *pattern*，则只列出表名或模式名匹配该模式的条目。如果使用 \det+ 形式，还会显示通用选项。

\deu[+] [*pattern*]

列出用户映射（助记词：“external users”）。如果指定了*pattern*，则只列出用户名匹配该模式的映射。如果使用\deu+形式，还会显示每个映射的附加信息。

> **小心**
>
> \deu+还可能显示远程用户的用户名和密码，因此应注意不要泄露它们。

\dew[+] [*pattern*]

列出外部数据包装器（助记词：“external wrappers”）。如果指定了 *pattern*，则只列出名称匹配该模式的外部数据包装器。如果使用 \dew+ 形式，还会显示外部数据包装器的 ACL 和选项。

\df[antwS+] [*pattern*]

列出函数及其参数、返回类型和函数类型。函数类型分为“agg”（聚合）、“normal”、“trigger”或“window”。要仅显示特定类型的函数，可在命令中添加对应的字母a、n、t或w。如果指定了*pattern*，则只显示名称匹配该模式的函数。默认只显示用户创建的对象；提供模式或S修饰符可包含系统对象。如果使用\df+形式，还会显示每个函数的附加信息，包括安全分类、易变性、拥有者、语言、源代码和描述。

> **提示**
>
> 要查找接受特定类型参数或返回特定类型值的函数，请使用分页器的搜索功能浏览\df输出。

\dF[+] [*pattern*]

列出全文检索配置。如果指定了*pattern*，则只显示名称匹配该模式的配置。如果使用\dF+形式，则显示每个配置的完整说明，包括底层全文检索解析器和每种解析器词元类型的词典列表。

\dFd[+] [*pattern*]

列出全文检索词典。如果指定了*pattern*，则只显示名称匹配该模式的词典。如果使用\dFd+形式，还会显示每个选中词典的附加信息，包括底层全文检索模板和选项值。

\dFp[+] [*pattern*]

列出全文检索解析器。如果指定了*pattern*，则只显示名称匹配该模式的解析器。如果使用\dFp+形式，则显示每个解析器的完整说明，包括底层函数和可识别的词元类型列表。

\dFt[+] [*pattern*]

列出全文检索模板。如果指定了*pattern*，则只显示名称匹配该模式的模板。如果使用\dFt+形式，还会显示每个模板的附加信息，包括底层函数名。

\dg[+] [*pattern*]

列出数据库角色。如果指定了 *pattern*，则只列出名称匹配该模式的角色。（此命令现在实际上等价于 \du。）如果使用 \dg+ 形式，还会显示每个角色的附加信息，包括每个角色的注释。

\dl

这是\lo_list的别名，用于显示大对象列表。

`\dL[S+] [ pattern ]`

列出过程语言。如果指定了*pattern*，则只列出名称匹配该模式的语言。
默认只显示用户创建的语言；提供S修饰符可包含系统对象。如果在命令名后附加+，
还会列出每种语言的调用处理器、验证器、访问权限，以及它是否为系统对象。

`\dn[S+] [ pattern ]`

列出模式（命名空间）。如果指定了*pattern*，则只列出名称匹配该模式的模式。
默认只显示用户创建的对象；提供模式或S修饰符可包含系统对象。如果在命令名后附加+，
还会列出每个对象关联的权限和描述（如果有）。

`\do[S] [ pattern ]`

列出操作符及其操作数类型和结果类型。如果指定了 *pattern*，则只列出名称匹配该模式的操作符。默认只显示用户创建的对象；提供模式或 S 修饰符可包含系统对象。

`\dO[S+] [ pattern ]`

列出排序规则。如果指定了*pattern*，则只列出名称匹配该模式的排序规则。
默认只显示用户创建的对象；提供模式或S修饰符可包含系统对象。如果在命令名后附加+，
还会列出每个排序规则关联的描述（如果有）。请注意，
只会显示可用于当前数据库编码的排序规则，
因此同一安装中的不同数据库可能会得到不同结果。

`\dp [ pattern ]`

列出表、视图和序列及其关联的访问权限。如果指定了*pattern*，
则只列出名称匹配该模式的表、视图和序列。

GRANT和REVOKE命令用于设置访问权限。权限显示的含义在GRANT中有解释。

`\drds [ role-pattern [ database-pattern ] ]`

列出已定义的配置设置。这些设置可以特定于角色、特定于数据库，或同时特定于两者。
*role-pattern*和*database-pattern*分别用于选择要列出的角色和数据库。
省略某个模式参数或将其指定为*时，不会按该参数筛选，
还会分别包含不特定于角色或不特定于数据库的设置。

ALTER ROLE和ALTER DATABASE命令用于定义角色专属和数据库专属的配置设置。

`\dT[S+] [ pattern ]`

列出数据类型。如果指定了 *pattern*，则只列出名称匹配该模式的类型。如果在命令名后附加 +，每个类型都会连同其内部名称和大小一起列出，如果它是 `enum` 类型，还会列出其允许值。默认只显示用户创建的对象；提供模式或 S 修饰符可包含系统对象。

`\du[+] [ pattern ]`

列出数据库角色。如果指定了 *pattern*，则只列出名称匹配该模式的角色。如果使用 `\du+` 形式，还会显示每个角色的附加信息，包括每个角色的注释。

`\dx[+] [ pattern ]`

列出已安装的扩展。如果指定了*pattern*，则只列出名称匹配该模式的扩展。如果使用`\dx+`形式，则列出属于每个匹配扩展的所有对象。

`\e` or `\edit [ filename ] [ line_number ]`

如果指定了*filename*，则编辑该文件；编辑器退出后，将其内容复制回查询缓冲区。如果未给出*filename*，则将当前查询缓冲区复制到临时文件，再以相同方式编辑。

随后，按照psql的正常规则重新解析新的查询缓冲区，其中整个缓冲区被视为一行。（因此，不能用这种方式编写脚本。应使用`\i`来处理这类脚本。）这意味着，如果查询以分号结束（或包含分号），它会立即执行；否则它只会在查询缓冲区中等待；键入分号或`\g`发送它，或键入`\r`取消。

如果指定了行号，psql会将光标定位到文件或查询缓冲区中的指定行。请注意，如果只给出一个全由数字组成的参数，psql会假定它是行号，而不是文件名。

> **提示**
>
> 有关如何配置和定制编辑器，请参见下面的环境。

`\echo text [ ... ]`

将参数打印到标准输出，参数之间用一个空格分隔，末尾跟随换行符。这可用于在脚本输出中插入信息。例如：

```
=> \echo `date`
Tue Oct 26 21:40:57 CEST 1999
```

如果第一个参数是未加引号的`-n`，则不输出末尾的换行符。

> **提示**
>
> 如果使用`\o`命令重定向查询输出，可能会想用`\qecho`代替这个命令。

`\ef [ function_description [ line_number ] ]`

这个命令获取并编辑指定函数的定义，形式为`CREATE OR REPLACE FUNCTION`命令。编辑方式与`\edit`相同。编辑器退出后，更新后的命令会在查询缓冲区中等待；键入分号或`\g`发送它，或用`\r`取消。

目标函数可以只用名称指定，也可以同时给出名称和参数，例如`foo(integer, text)`。如果存在多个同名函数，就必须给出参数类型。

如果未指定函数，则会呈现一个空的`CREATE FUNCTION`模板供编辑。

如果指定了行号，psql会将光标定位到函数体中的指定行。（请注意，函数体通常并不从文件的第一行开始。）

> **提示**
>
> 有关如何配置和定制编辑器，请参见下面的环境。

`\encoding [ encoding ]`

设置客户端字符集编码。没有参数时，此命令显示当前编码。

`\f [ string ]`

设置非对齐查询输出的字段分隔符。默认值是竖线（`|`）。另请参见 `\pset`，那里介绍了设置输出选项的通用方法。

`\g [ filename ]`
`\g [ |command ]`

把当前查询输入缓冲区发送给服务器，并可选地把查询输出存储到 *filename*，或把输出通过管道传给 shell 命令 *command*。单独的 `\g` 基本上等同于一个分号。带参数的 `\g` 是 `\o` 命令的一种 “一次性” 替代。

`\h` or `\help [ command ]`

给出指定SQL命令的语法帮助。如果未指定*command*，则psql将列出所有可用语法帮助的命令。如果*command*是星号（`*`），则显示所有SQL命令的语法帮助。

> **注意**
>
> 为了简化输入，由多个单词组成的命令不需要加引号。因此，可以直接输入**`\help alter table`**。

`\H` or `\html`

打开HTML查询输出格式。如果HTML格式已经打开，则切换回默认的对齐文本格式。此命令是为兼容性和便利性而保留的；设置其他输出选项的方法见`\pset`。

`\i` or `\include filename`

从文件*filename*中读取输入，并像在键盘上输入一样执行它。

如果*filename*是`-`（连字符），则从标准输入读取，直到遇到 EOF 指示或`\q`元命令。这可用于将交互式输入与文件输入交错使用。请注意，只有在最外层启用了 Readline，此处才会使用 Readline 功能。

> **注意**
>
> 如果想在屏幕上看到被读入的各行，请将变量`ECHO`设置为`all`。

`\l` (or `\list`)
`\l+` (or `\list+`)

列出服务器中所有数据库的名称、拥有者、字符集编码和访问权限。如果在命令名后附加 `+`，还会显示数据库大小、默认表空间和描述。
（大小信息只对当前用户能够连接的数据库可用。）

`\lo_export` *loid filename*

从数据库中读取具有OID*loid*的大对象，并将其写入*filename*。请注意，这与服务器函数`lo_export`略有不同，后者使用运行数据库服务器的用户的权限，
并在服务器的文件系统上操作。

> **提示**
>
> 使用`\lo_list`命令来查找大对象的OID。

`\lo_import` *filename* [*comment*]

将文件存储到一个PostgreSQL大对象中。可选地，它将给定的注释与对象关联起来。例如：

```
foo=> \lo_import '/home/peter/pictures/photo.xcf' 'a picture of me'
lo_import 152801
```

响应表明大对象获得了对象 ID 152801，这个 ID 可以用来在将来访问新创建的大对象。为便于阅读，建议始终为每个对象关联一条便于人阅读的注释。查看 OID 和注释时，可以使用`\lo_list`命令。

请注意，此命令与服务器端的`lo_import`略有不同，
因为它作为本地用户在本地文件系统上操作，而不是服务器的用户和文件系统。

`\lo_list`

列出当前存储在数据库中的所有PostgreSQL大对象，以及为它们提供的注释。

`\lo_unlink` *loid*

从数据库中删除OID 为 *loid*的大对象。

> **提示**
>
> 使用`\lo_list`命令来查找大对象的OID。

`\o` or `\out` [*filename*]
`\o` or `\out` [|*command*]

将后续查询结果保存到文件*filename*，或通过管道传给 shell 命令*command*。如果没有指定参数，查询输出将恢复为标准输出。

“查询结果”包括从数据库服务器获取的所有表格、命令响应和通知，
以及查询数据库的各种反斜杠命令的输出（例如`\d`），但不包括错误消息。

> **提示**
>
> 要在查询结果之间插入文本输出，使用`\qecho`。

`\p` or `\print`

将当前查询缓冲区打印到标准输出。

`\password [ username ]`

更改指定用户（默认为当前用户）的密码。此命令提示输入新密码，对其进行加密，并将其作为`ALTER ROLE`命令发送到服务器。这样可以确保新密码不会以明文形式出现在命令历史记录、服务器日志或其他地方。

`\prompt [ text ] name`

提示用户提供文本，将其赋值给变量 *name*。可以指定一个可选的提示字符串 *text*。（对于多个单词的提示，用单引号括起文本。）

默认情况下，`\prompt` 使用终端进行输入和输出。然而，如果使用了 `-f` 命令行开关，`\prompt` 将使用标准输入和标准输出。

`\pset option [ value ]`

这个命令设置影响查询结果表输出的选项。*option*指定要设置哪个选项。*value*的含义取决于所选的选项。对于某些选项，省略*value*会切换或取消设置该选项，具体见各选项的说明。如果没有提及这类行为，那么省略*value*只会显示当前设置。

Adjustable printing options are:

`border`

value 必须是数字。一般来说，数字越大，表格的边框和分隔线就越多，但这取决于具体的格式。在 HTML 格式中，它会直接转换为 `border=...` 属性；在其他格式中，只有值 0（无边框）、1（内部分隔线）和 2（表格外框）有意义。

`columns`

为 `wrapped` 格式设置目标宽度，同时设置用于确定输出是否宽到需要分页器的宽度上限。零（默认值）表示目标宽度由环境变量 `COLUMNS` 控制，如果未设置 `COLUMNS`，则使用检测到的屏幕宽度。此外，如果 `columns` 为零，则 `wrapped` 格式只影响屏幕输出。如果 `columns` 非零，则文件和管道输出也会折行到该宽度。

`expanded` (or `x`)

如果指定了 *value*，它必须是 `on` 或 `off`，分别启用或禁用扩展模式。如果省略 *value*，该命令会在常规模式和扩展模式之间切换。启用扩展模式时，查询结果以两列显示，左侧为列名，右侧为数据。如果数据在通常的“横向”模式下无法在屏幕上完整显示，这种模式就很有用。

`fieldsep`

指定非对齐输出格式使用的字段分隔符。这样就可以生成例如制表符分隔或逗号分隔的输出，其他程序可能更喜欢这种格式。要把制表符设置为字段分隔符，可以键入 `\pset fieldsep '\t'`。默认的字段分隔符是 `'|'`（竖线）。

`footer`

如果指定了 *value*，它必须是 `on` 或 `off`，分别启用或禁用表格页脚（`(n rows)` 计数）的显示。如果省略 *value*，该命令会切换页脚显示的开关状态。

`format`

将输出格式设置为 `unaligned`、`aligned`、`wrapped`、`html`、`latex` 或 `troff-ms` 之一。允许使用无歧义的缩写。（这意味着一个字母就足够了。）

`unaligned` 格式把一行的所有列写在一行上，用当前生效的字段分隔符分隔。这对于生成可能要由其他程序读入的输出（例如制表符分隔或逗号分隔格式）很有用。

`unaligned`格式将一行的所有列写在同一行上，以当前生效的字段分隔符分隔。这适合创建供其他程序读取的输出（例如制表符分隔或逗号分隔格式）。

`wrapped` 格式与 `aligned` 相似，但会把较宽的数据值折成多行，使输出适应目标列宽。目标宽度的确定方式见 `columns` 选项的说明。注意，psql 不会尝试对列标题折行；因此，如果列标题所需的总宽度超过目标宽度，`wrapped` 格式的行为就与 `aligned` 相同。

`html`、`latex` 和 `troff-ms` 格式生成的表格旨在包含在使用相应标记语言的文档中。它们不是完整的文档！（这在 HTML 中可能不是必需的，但在 LaTeX 中，必须有一个完整文档的外层结构。）

`linestyle`

将边框线绘制样式设置为 `ascii`、`old-ascii` 或 `unicode` 之一。允许使用无歧义的缩写。（这意味着一个字母就足够了。）默认设置为 `ascii`。此选项只影响 `aligned` 和 `wrapped` 输出格式。

`ascii` 样式使用普通的 ASCII 字符。数据中的换行符以右边缘的 `+` 符号表示。当 `wrapped` 格式在没有换行符的位置把数据折到下一行时，会在第一行的右边缘显示一个点（`.`），并在下一行的左边缘再次显示。

`old-ascii` 样式使用普通的 ASCII 字符，采用 PostgreSQL 8.4 及更早版本的格式样式。数据中的换行符以替代左侧列分隔符的 `:` 符号表示。当数据在没有换行符的位置折到下一行时，则用 `;` 符号替代左侧列分隔符。

`unicode` 样式使用 Unicode 框线绘制字符。数据中的换行符以右边缘的回车符号表示。当数据在没有换行符的位置折到下一行时，会在第一行的右边缘显示省略号符号，并在下一行的左边缘再次显示。

当 `border` 设置大于零时，这个选项还决定用哪些字符绘制边框线。普通的 ASCII 字符在任何环境中都可用，但在支持 Unicode 的显示设备上，Unicode 字符更美观。

`null`

设置用于代替空值打印的字符串。默认不打印任何内容，这很容易被误认为空字符串。例如，你可能更喜欢使用 `\pset null '(null)'`。

`numericlocale`

如果指定了 *`value`*，它必须是 `on` 或 `off`，分别启用或禁用使用区域设置特定的字符来分隔小数点左侧的数字组。如果省略 *`value`*，该命令会在常规数字输出和区域设置特定的数字输出之间切换。

`pager`

控制查询和 psql 帮助输出是否使用分页器程序。如果设置了环境变量 `PAGER`，输出会通过管道传递给指定的程序。否则，使用与平台有关的默认程序（如 `more`）。

当 pager 选项为 off 时，不使用分页器程序。当 pager 选项为 on 时，会在适当的时候使用分页器，即输出目标为终端且内容无法在屏幕上完整显示时。pager 选项也可以设为 always，这样所有终端输出都会使用分页器，无论内容是否能在屏幕上完整显示。不带 *value* 的 \pset pager 会切换分页器的使用状态。

recordsep

指定非对齐输出格式使用的记录（行）分隔符。默认是换行符。

tableattr (or T)

指定要放在 html 输出格式的 HTML table 标签内部的属性。例如可以是 cellpadding 或 bgcolor。注意，你可能不想在这里指定 border，因为这已经由 \pset border 处理。如果没有给出 *value*，则取消表属性的设置。

title

设置随后打印的所有表格的标题。这可以为输出提供描述性标签。如果没有给出 *value*，则取消标题的设置。

tuples_only (or t)

如果指定了 *value*，它必须是 on 或 off，分别启用或禁用仅元组模式。如果省略 *value*，该命令会在常规输出和仅元组输出之间切换。常规输出包含列标题、表格标题和各种页脚等附加信息。在仅元组模式下，只显示实际的表格数据。

这些不同格式的外观示例可参见示例一节。

提示

有各种用于 \pset 的快捷命令。请参见 \a、\C、\H、\t、\T 和 \x。

注意

不带任何参数调用 \pset 是一个错误。在将来的版本中，这种情况可能会显示所有打印选项的当前状态。

\q or \quit

退出psql程序。在脚本文件中，只会终止该脚本的执行。

\qecho *text* [...]

这个命令与\echo相同，只是输出会写入由\o设置的查询输出通道。

\r or \reset

重置（清空）查询缓冲区。

\s [*filename*]

打印psql的命令行历史记录到*filename*。如果省略*filename*，历史记录将被写入标准输出（如果适用，将使用分页器）。如果psql在构建时没有使用Readline支持，则此命令不可用。

`\set [ name [ value [ ... ] ] ]`

把内部变量 *name* 设置为 *value*；如果给出多个值，则设置为所有值的串接。如果没有给出第二个参数，变量只是被设置为没有值。要取消变量的设置，使用 `\unset` 命令。

合法的变量名可以包含字母、数字和下划线。详情见下面的变量。变量名区分大小写。

尽管你可以随意将任何变量设置为任何值，但psql将其中若干变量视为特殊变量。这些变量在下面有关变量的章节中介绍。

> **注意**
>
> 这个命令与 SQL 命令SET完全无关。

`\sf[+] function_description`

这个命令获取并显示指定函数的定义，以`CREATE OR REPLACE FUNCTION`命令的形式呈现。定义将打印到当前查询输出通道，由`\o`设置。

目标函数可以只用名称指定，也可以同时给出名称和参数，例如`foo(integer, text)`。如果存在多个同名函数，就必须给出参数类型。

如果在命令名后附加`+`，则会为输出行编号，函数体的第一行编号为 1。

`\t`

切换输出中的列名标题和行数页脚的显示状态。这个命令等价于`\pset tuples_only`，提供它是为了使用方便。

`\T table_options`

指定在HTML输出格式中放在`table`标签内的属性。这个命令等价于`\pset tableattr table_options`。

`\timing [ on | off ]`

不带参数时，切换以毫秒为单位显示每条 SQL 语句执行耗时的开关状态。带参数时，设置为相同的值。

`\w` or `\write filename`
`\w` or `\write |command`

将当前查询缓冲区输出到文件*filename*，或通过管道传递给 shell 命令*command*。

`\x`

设置或切换扩展表格格式模式。它等价于`\pset expanded`。

`\z [ pattern ]`

列出表、视图和序列及其关联的访问权限。如果指定了*pattern*，则只列出名称匹配该模式的表、视图和序列。

这是`\dp`的别名（“显示权限”）。

`\! [ command ]`

进入一个单独的 shell，或执行 shell 命令 *command*。参数不会被进一步解释；shell 会原样看到它们。

`\?`

显示有关反斜线命令的帮助信息。

模式

很多`\d`命令都可以用一个*pattern*参数来指定要被显示的对象名称。在最简单的情况下，模式正好就是该对象的准确名称。在模式中的字符通常会被变成小写形式（就像在 SQL 名称中那样），例如`\dt FOO`将会显示名为`foo`的表。就像在 SQL 名称中那样，把模式放在双引号中可以阻止它被转换成小写形式。如果需要在一个模式中包括一个真正的双引号字符，则需要在双引号包围的文本内把它写成两个相邻的双引号，这同样是符合 SQL 加引号标识符的规则。例如，`\dt "FOO""BAR"`将显示名为`FOO"BAR`（不是`foo"bar`）的表。和普通的 SQL 名称规则不同，你可以只在模式的一部分周围放上双引号，例如`\dt FOO"FOO"BAR`将会显示名为`fooFOObar`的表。

只要完全省略*pattern*参数，`\d`命令就会显示当前模式搜索路径中可见的全部对象 – 这等价于用`*`作为模式（如果一个对象所在的模式位于搜索路径中，并且在搜索路径中该模式之前没有同类且同名的对象，则该对象就是可见的。这表示可以直接用名称引用该对象，而不需要用模式来限定）。要查看数据库中所有对象而不考虑其可见性，可以把`*.*`用作模式。

如果放在一个模式中，`*`将匹配任意字符序列（包括空序列），而`?`会匹配任意的单个字符（这种记号方法就像 Unix shell 的文件名模式一样）。例如，`\dt int*`会显示名称以`int`开始的表。但是如果被放在双引号内，`*`和`?`就会失去这些特殊含义而变成普通的字符。

包含点号（`.`）的模式会被解释为模式名称的匹配模式，后接对象名称的匹配模式。例如，`\dt foo*.*bar*`显示模式名以`foo`开头且表名包含`bar`的所有表。如果没有点号，则该匹配模式只匹配当前模式搜索路径中可见的对象。同样，双引号内的点号会失去其特殊含义，而按字面匹配。

高级用户可以使用字符类等正则表达式记法，如`[0-9]`可以匹配任意数字。所有的正则表达式特殊字符都按照第 9.7.3 节所说的工作，以下字符除外：`.`会按照上面所说的作为一种分隔符，`*`会被翻译成正则表达式记号`.*`，`?`会被翻译成`.`，而`$`则按字面意思匹配。根据需要，可以用`?`模拟`.`，用`(R+|)`模拟`R*`，或用`(R|)`模拟`R?`。`$`不需要作为一个正则表达式字符，因为模式必须匹配整个名称，而不是像正则表达式的常规用法那样解释（换句话说，`$`会被自动地追加到模式上）。如果不希望该模式的匹配位置被固定，可以在开头或者结尾写上`*`。注意在双引号内，所有的正则表达式特殊字符会失去其特殊含义并且按照其字面意思进行匹配。还有，在操作符名称模式中（即作为`\do`的参数），正则表达式特殊字符也按照字面意思进行匹配。

高级特性

变量

psql 提供了与常见 Unix 命令 shell 相似的变量替换特性。变量就是简单的名称/值对，其中值可以是任意长度的任意字符串。要设置变量，使用 psql 元命令 `\set`：

```
testdb=> \set foo bar
```

把变量 foo 设置为值 bar。要获取变量的内容，在名称前加上冒号，并把它用作任何斜线命令的参数：

```
testdb=> \echo :foo
bar
```

> **注意**
>
> \set的参数服从与其他命令相同的替换规则。因此可以构造有趣的引用，例如\set :foo 'something'以及分别得到Perl或者PHP的“软链接”或者“可变变量”。不幸的是（或者幸运的是？），这些构造出来的东西并没有什么用处。在另一方面，\set bar :foo是一种很好的拷贝变量的方法。

如果调用 \set 时没有第二个参数，该变量会被设置，其值为空字符串。要取消设置（即删除）一个变量，使用命令 \unset。

psql 的内部变量名可以由字母、数字和下划线按任意顺序、任意数量组成。其中一些变量会被 psql 特殊对待。它们表示特定的选项设置（运行时可以通过更改该变量的值来改变），或者表示应用的某种状态。尽管你可以把这些变量用于其他目的，但不建议这样做，
因为程序的行为可能会很快变得非常奇怪。按照惯例，
所有被特殊对待的变量都由全大写字母（以及可能的数字和下划线）组成。
为了确保将来最大的兼容性，请避免把这类变量名用于自己的目的。
下面是所有被特殊对待的变量的列表。

AUTOCOMMIT

在被设置为on（默认）时，每一个 SQL 命令在成功完成时会被自动提交。在这种模式中要推迟提交，必须输入一个BEGIN或者START TRANSACTION SQL 命令。当被设置为off或者被取消设置时，在显式发出COMMIT或者END之前，SQL 命令不会被提交。自动提交关闭模式会为你发出一个隐式的BEGIN，这会发生在任何不在一个事务块中且本身既不是BEGIN
及其他事务控制命令且不是无法在事务块中执行的命令（例如VACUUM）之前。

> **注意**
>
> 在自动提交关闭模式中，必须通过ABORT或者ROLLBACK
> 显式地放弃任何失败的事务。还要记住，如果退出会话时没有提交，
> 则所有的工作都会丢失。

> **注意**
>
> 自动提交打开模式是PostgreSQL的传统行为，但是自动提交关闭模式更接近于 SQL 的规范。如果更喜欢自动提交关闭模式，可以在系统级的psqlrc文件或者个人的~/.psqlrc文件中设置它。

DBNAME

当前已连接的数据库名称。每次连接到一个数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

ECHO

如果被设置为all，
所有非空输入行会在读入时打印到标准输出（不适用于交互式读取的行）。
要在程序开始时选择这种行为，可以使用开关-a。如果被设置为queries，
psql会在发送每个查询给服务器时将它们打印到标准输出。选择这种行为的开关是-e。
如果未设置，或被设置为上述值以外的其他值，则不会显示任何查询。

ECHO_HIDDEN

当这个变量被设置为on且一个反斜线命令查询数据库时，相应的查询会被先显示。
这种特性可以帮助我们学习
PostgreSQL的内部并且在自己的程序中提供类似的功能（要在程序开始时选择这种行为，
可以使用开关-E）。如果把这个变量设置为值noexec，
则对应的查询只会被显示而并不真正被发送给服务器执行。

ENCODING

当前的客户端字符集编码。

FETCH_COUNT

如果这个变量被设置为一个大于 0 的整数值，SELECT查询的结果会以一组一组的方式取出并
且显示（而不是像默认的那样把整个结果集拿到以后再显示），
每组包含的行数等于该整数值。因此，这种方式只会使用有限的内存量，
而不管整个结果集的大小。在启用这个特性时，通常会使用 100 到 1000 的设置。记住在使用
这种特性时，一个查询可能会在已经显示了一些行之后失败。

提示

尽管可以把这种特性用于任何的输出格式，但是默认的aligned
格式看起来会比较糟糕，因为每一组的FETCH_COUNT行将被单独格式化，
这就会导致不同的行组的列宽不同。其他的输出格式会更好。

HISTCONTROL

如果这个变量被设置为ignorespace，则以一个空格开始的行不会被放入到历史列表中。
如果被设置为值ignoredups，则与上一条历史记录相同的行不会被放入。值ignoreboth
组合了上述两种值。如果未设置，或被设置为上述值以外的其他值，
所有在交互模式中被读入的行都会保存在历史列表中。

注意

这个特性是可耻地从Bash抄袭过来的。

HISTFILE

用于存储历史记录列表的文件名。默认值是~/.psql_history。例如，将以下内容：

```
\set HISTFILE ~/.psql_history- :DBNAME
```

放入 ~/.psqlrc 会使 psql 为每个数据库维护单独的历史记录。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`HISTSIZE`

存储在命令历史中的命令数量。默认值是 500。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`HOST`

当前连接到的数据库服务器主机。
每次连接到数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

`IGNOREEOF`

如果未设置，向一个psql的交互式会话发送一个EOF字符（通常是**Control+D**）
将会终止应用。如果设置为一个数字值，则会有该数量的EOF字符被忽略，
然后应用才会终止。如果该变量被设置但没有数字值，则默认为 10。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`LASTOID`

最后被影响的 OID 的值，这可能会由`INSERT`或者`\lo_import`命令返回。这个变量只保证在下一个SQL命令的结果被显示完之前有效。

`ON_ERROR_ROLLBACK`

当被设置为`on`时，如果事务块中的一个语句产生一个错误，
该错误会被忽略并且该事务会继续。当被设置为`interactive`时，
只在交互式会话中忽略这类错误，而读取脚本文件时则不会忽略错误。当未设置或被设置为`off`时，事务块中产生错误的一个语句会中止整个事务。
错误回滚模式的工作原理是在事务块的每个命令之前都为你发出一个隐式的`SAVEPOINT`，
然后在该命令失败时回滚到该保存点。

`ON_ERROR_STOP`

默认情况下，发生错误后命令处理会继续。当这个变量被设置为`on`时，处理将立即停止。
在交互模式下，psql会返回到命令提示符；否则，psql会退出，并返回错误代码 3，以便与错误代码 1 所表示的致命错误区分开来。在两种情况下，任何当前正在运行的脚本（顶层脚本以及它可能调用的其他脚本）都会立即中止。如果顶层命令字符串包含多个 SQL 命令，则会在当前命令处停止处理。

PORT

当前连接到的数据库服务器端口。
每次连接到数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

PROMPT1
PROMPT2
PROMPT3

这些变量指定psql发出的提示符的模样。见下文的提示符。

QUIET

把这个变量设置为on等效于命令行选项-q。在交互模式下可能用处不大。

SINGLELINE

设置这个变量为on等效于命令行选项-S。

SINGLESTEP

设置这个变量为on等效于命令行选项-s。

USER

当前连接的数据库用户。每次连接到一个数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

VERBOSITY

这个变量可以被设置为值default、verbose或者terse来控制错误报告的详细程度。

SQL Interpolation

psql 变量的另一个有用特性是可以把它们替换（“插值”）到常规 SQL 语句中。psql 提供了专门的机制，确保被用作 SQL 字面量和标识符的值被正确转义。不加任何特殊转义地插值一个值的语法同样是在变量名前加上冒号（:）：

```
testdb=> \set foo 'my_table'
testdb=> SELECT * FROM :foo;
```

将查询表 my_table。注意这可能不安全：变量的值会被按字面拷贝，因此它甚至可能包含不平衡的引号或反斜线命令。必须确保把它放在那里是有意义的。

当一个值要用作 SQL 字面量或标识符时，最安全的做法是为它安排转义。要把变量值作为 SQL 字面量转义，写一个冒号，后跟用单引号括起的变量名。要把变量值作为 SQL 标识符转义，写一个冒号，后跟用双引号括起的变量名。前面的示例用下面这种写法更安全：

```
testdb=> \set foo 'my_table'
testdb=> SELECT * FROM :"foo";
```

不会对加了引号的 SQL 实体执行变量插值。

这种机制的一个可能用途是把一个文件的内容拷贝到一个表列中。
首先把该文件载入到一个变量，然后按上面的做法进行：

```
testdb=> \set content `cat my_file.txt`
```

```
testdb=> INSERT INTO my_table VALUES (:'content');
```

（注意如果 `my_file.txt` 包含 NUL 字节，这样仍然不行。psql 不支持在变量值中嵌入 NUL 字节。）

由于冒号可以合法地出现在 SQL 命令中，一次明显的插值尝试（如 `:name`、`:'name'` 或 `:"name"`）不会被改变，除非所指的变量当前已设置。无论哪种情况，都可以用反斜线对冒号转义，以避免它被替换。（变量的冒号语法是嵌入式查询语言（如 ECPG）的标准 SQL 语法。数组切片和类型转换的冒号语法是 PostgreSQL 的扩展，因此存在冲突。把变量值作为 SQL 字面量或标识符转义的冒号语法是 psql 的扩展。）

提示符

psql 发出的提示符可以按你的喜好进行定制。`PROMPT1`、`PROMPT2` 和 `PROMPT3` 这三个变量包含描述提示符外观的字符串和特殊转义序列。提示符 1 是 psql 请求新命令时发出的常规提示符。提示符 2 会在录入命令期间还需要更多输入时发出，例如命令尚未以分号结束，或者引号尚未闭合时。在执行 SQL `COPY FROM STDIN` 命令并需要在终端中输入一行值时，会发出提示符 3。

所选提示符变量的值会原样打印，除非遇到百分号（`%`）。
此时会根据下一个字符替换为其他文本。已定义的替换项如下：

`%M`

数据库服务器的完整主机名（含域名）；如果通过 Unix 域套接字连接，则为`[local]`；如果 Unix 域套接字不在编译时指定的默认位置，则为`[local:/dir/name]`。

`%m`

数据库服务器的主机名，在第一个点号处截断；如果通过 Unix 域套接字连接，则为`[local]`。

`%>`

数据库服务器监听的端口号。

`%n`

数据库会话用户名。（在数据库会话期间，`SET SESSION AUTHORIZATION`命令可能改变该值的扩展结果。）

`%/`

当前数据库的名称。

`%~`

类似 `%/`，但如果该数据库是你的默认数据库，则输出 `~`（波浪号）。

`%#`

如果会话用户是数据库超级用户，则为`#`，否则为`>`。（在数据库会话期间，`SET SESSION AUTHORIZATION`命令可能改变该值的扩展结果。）

`%R`

在提示符 1 中，通常为 `=`；但如果处于单行模式，则为 `^`；如果会话已与数据库断开连接（这可能发生在 `\connect` 失败时），则为 `!`。在提示符 2 中，`%R` 会被替换为一个字符，该字符

取决于 psql 为什么还期待更多输入：如果命令只是尚未终止，则为 `-`；如果存在未结束的 `/* ... */` 注释，则为 `*`；如果存在未结束的带引号字符串，则为单引号；如果存在未结束的带引号标识符，则为双引号；如果存在未结束的美元引用字符串，则为美元符号；
如果存在未匹配的左括号，则为 `(`。在提示符 3 中，`%R` 不会产生任何输出。

`%x`

事务状态：如果当前不在事务块中，则为空字符串；如果处于事务块中，则为 `*`；如果处于失败的事务块中，则为 `!`；如果事务状态不确定（例如因为当前没有连接），则为 `?`。

`%digits`

替换为指定八进制代码对应的字符。

`%:name:`

psql 变量 *name* 的值。详见变量。

`%`` `command` ``

command 的输出，类似普通的“反引号”替换。

`%[ ... %]`

提示符中可以包含终端控制字符，例如改变提示文本的颜色、背景或样式，
或者改变终端窗口标题。为了让 Readline 的行编辑功能正常工作，这些不可打印的控制字符必须用 `%[` 和 `%]` 包围起来，以标记为不可见字符。提示符中可以出现多组这样的标记。例如：

```
testdb=> \set PROMPT1 '%[%033[1;33;40m%]%n@%/%R%[%033[0m%]%# '
```

其效果是在兼容 VT100 且支持颜色的终端上，生成一个粗体（`1;`）的黑底黄字提示符（`33;40`）。

要在提示符中插入百分号，请写为`%%`。默认情况下，提示符 1 和 2 为`'%/%R%# '`，提示符 3 为`'>> '`。

> **注意**
>
> 这个特性是可耻地从tcsh抄袭过来的。

命令行编辑

psql 支持 Readline 库，便于编辑和检索输入行。命令历史记录会在 psql 退出时自动保存，并在 psql 启动时重新载入。也支持 Tab 补全，不过其补全逻辑并不声称自己是 SQL 解析器。如果出于某种原因你不喜欢 Tab 补全，可以将以下内容放入主目录下名为 `.inputrc` 的文件中，将其关闭：

```
$if psql
set disable-completion on
$endif
```

（这不是 psql 的功能，而是 Readline 的功能。更多细节请阅读其文档。）

环境

`COLUMNS`

如果 `\pset columns` 为零，这个环境变量控制用于 `wrapped` 格式的宽度，以及用来确定较宽的输出是否需要分页器的宽度。

`PAGER`

如果查询结果无法在屏幕上完整显示，则会通过管道传递给这个命令。典型值为 `more` 或 `less`。默认值取决于平台。可以使用 `\pset` 命令禁用分页器的使用。

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数（见第 31.13 节）。

`PSQL_EDITOR`
`EDITOR`
`VISUAL`

`\e`和`\ef`命令使用的编辑器。按列出的顺序检查这些变量，使用第一个已设置的变量。

内置的默认编辑器在 Unix 系统上为`vi`，在 Windows 系统上为`notepad.exe`。

`PSQL_EDITOR_LINENUMBER_ARG`

当 `\e` 或 `\ef` 带有行号参数使用时，此变量指定将起始行号传给用户编辑器时所用的命令行参数。对于 Emacs 或 vi 之类的编辑器，这个参数是一个加号。如果选项名和行号之间需要空格，请在变量值中包含末尾空格。例如：

```
PSQL_EDITOR_LINENUMBER_ARG='+'
PSQL_EDITOR_LINENUMBER_ARG='--line '
```

在 Unix 系统上默认是`+`（对应于默认编辑器`vi`，且对很多其他常见编辑器可用）。在 Windows 系统上没有默认值。

`SHELL`

被`\!`命令执行的命令。

`TMPDIR`

存储临时文件的目录。默认是`/tmp`。

和大部分其他PostgreSQL工具一样，这个工具也使用libpq所支持的环境变量（见第 31.13 节）。

文件

- 除非向它传递 `-X` 或 `-c` 选项，否则 psql 在启动之前会尝试从系统级的 `psqlrc` 文件和用户的 `~/.psqlrc` 文件中读取并执行命令。（在 Windows 上，用户的启动文件名为 `%APPDATA%\postgresql\psqlrc.conf`。）有关设置系统级文件的信息，参见 *PREFIX/*

share/psqlrc.sample。它可用来按个人喜好设置客户端或服务器（使用 \set 和 SET 命令）。

- 系统范围的 psqlrc 文件和用户的 ~/.psqlrc 文件都可以通过在文件名后附加连字符和 PostgreSQL 的版本号来使其与版本相关，例如 ~/.psqlrc-9.1.24。匹配的版本特定文件将优先于非版本特定的文件被读取。
- 命令行历史被存储在文件~/.psql_history中，或者是 Windows 的文件%APPDATA%\postgresql\psql_history中。

注解

- 在早期版本中，psql 允许一个单字母反斜线命令的第一个参数直接写在该命令后面，中间不需要空白。从 PostgreSQL 8.4 起不再允许这样做。
- 只保证 psql 能与相同版本的服务器顺利配合工作。这并不意味着其他组合会完全失败，但可能出现或微妙或不那么微妙的问题。如果服务器的版本比 psql 本身更新，反斜线命令特别容易失败。然而，\d 系列的反斜线命令应该可以与最低至 7.4 版本的服务器配合工作，但不一定适用于比 psql 本身更新的服务器。

给 Windows 用户的注解

psql是一个“控制台应用”。由于 Windows 的控制台窗口使用的是一种和系统中其他应用不同的编码，在psql中使用 8 位字符时要特别注意。如果psql检测到一个有问题的控制台代码页，它将会在启动时警告你。要更改控制台代码页，有两件事是必要的：

- 输入**cmd.exe /c chcp 1252**可以设置代码页（1252 是适用于德语的一个代码页，请在这里替换成你的值）。如果正在使用 Cygwin，可以把这个命令放在/etc/profile中。
- 把控制台字体设置为Lucida Console，因为栅格字体无法与 ANSI 代码页一起使用。

示例

第一个示例展示如何将一个命令分散在多行输入中。请注意提示符的变化：

```
testdb=> CREATE TABLE my_table (
testdb(>  first integer not null default 0,
testdb(>  second text)
testdb-> ;
CREATE TABLE
```

现在再看看表定义：

```
testdb=> \d my_table
             Table "my_table"
 Attribute |  Type   |      Modifier
-----------+---------+--------------------
 first     | integer | not null default 0
 second    | text    |
```

现在我们把提示符改得更有趣一些：

```
testdb=> \set PROMPT1 '%n@%m %~%R%# '
peter@localhost testdb=>
```

假设你已经在表中填入数据，并想查看一下：

```
peter@localhost testdb=> SELECT * FROM my_table;
 first | second
-------+--------
     1 | one
     2 | two
     3 | three
     4 | four
(4 rows)
```

要以不同方式显示表格，可以使用\pset命令：

```
peter@localhost testdb=> \pset border 2
Border style is 2.
peter@localhost testdb=> SELECT * FROM my_table;
+-------+--------+
| first | second |
+-------+--------+
|     1 | one    |
|     2 | two    |
|     3 | three  |
|     4 | four   |
+-------+--------+
(4 rows)

peter@localhost testdb=> \pset border 0
Border style is 0.
peter@localhost testdb=> SELECT * FROM my_table;
first second
----- ------
    1 one
    2 two
    3 three
    4 four
(4 rows)

peter@localhost testdb=> \pset border 1
Border style is 1.
peter@localhost testdb=> \pset format unaligned
Output format is unaligned.
peter@localhost testdb=> \pset fieldsep ","
Field separator is ",".
peter@localhost testdb=> \pset tuples_only
Showing only tuples.
peter@localhost testdb=> SELECT second, first FROM my_table;
one,1
two,2
three,3
four,4
```

也可以使用简短命令：

```
peter@localhost testdb=> \a \t \x
```

```
Output format is aligned.
Tuples only is off.
Expanded display is on.
peter@localhost testdb=> SELECT * FROM my_table;
-[ RECORD 1 ]-
first  | 1
second | one
-[ RECORD 2 ]-
first  | 2
second | two
-[ RECORD 3 ]-
first  | 3
second | three
-[ RECORD 4 ]-
first  | 4
second | four
```

reindexdb

reindexdb — reindex a PostgreSQL database

大纲

`reindexdb` [*connection-option*...] [--table | -t *table*] [--index | -i *index*] [*dbname*]

`reindexdb` [*connection-option*...] [--all | -a]

`reindexdb` [*connection-option*...] [--system | -s] [*dbname*]

描述

reindexdb是用于重建PostgreSQL数据库中索引的工具。

reindexdb 是 SQL 命令REINDEX的一个包装器。通过该工具重建数据库索引与通过其他方法访问服务器来重建索引，在效果上没有区别。

选项

reindexdb接受以下命令行参数：

`-a`
`--all`

重建所有数据库的索引。

`[-d]` *dbname*
`[--dbname=]`*dbname*

指定要重建索引的数据库名称。如果未指定并且没有使用 `-a`（或`--all`），则从环境变量`PGDATABASE`中读取数据库名称。如果该变量未设置，则使用为连接指定的用户名。

`-e`
`--echo`

回显reindexdb生成并发送给服务器的命令。

`-i` *index*
`--index=`*index*

只重建 *index*。

`-q`
`--quiet`

不显示进度消息。

`-s`
`--system`

重建数据库的系统目录上的索引。

`-t table`
`--table=table`

只对 *table* 重新建立索引。

`-V`
`--version`

打印reindexdb的版本并退出。

`-?`
`--help`

显示关于reindexdb命令行参数的帮助信息，并退出。

reindexdb还接受以下用于连接参数的命令行参数：

`-h host`
`--host=host`

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p port`
`--port=port`

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U username`
`--username=username`

要作为其身份连接的用户名。

`-w`
`--no-password`

永远不发出密码提示。如果服务器要求密码认证，而又无法通过其他方式获取密码，例如通过`.pgpass`文件，则连接尝试会失败。
此选项可用于没有用户可输入密码的批处理作业和脚本。

`-W`
`--password`

强制reindexdb在连接数据库前提示输入密码。

该选项并非必需，因为如果服务器要求密码认证，reindexdb会自动提示输入密码。不过，reindexdb会先浪费一次连接尝试来确认服务器需要密码。在某些场景下，值得使用 `-W`，以避免这次额外的连接尝试。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他PostgreSQL工具一样，该工具也使用libpq支持的环境变量（见第 31.13 节）。

诊断

如果遇到困难，请参阅REINDEX和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须运行在目标主机上。此外，
libpq前端库所使用的任何默认连接设置和环境变量也都会生效。

注解

reindexdb 可能需要多次连接到 PostgreSQL 服务器，每次都要求输入密码。在这种情况下，使用 `~/.pgpass` 文件会很方便。更多信息见第 31.14 节。

示例

要重建数据库`test`中的索引：

```
$ reindexdb test
```

要重建名为`abcd`的数据库中表`foo`的索引和索引`bar`：

```
$ reindexdb --table foo --index bar abcd
```

另见

REINDEX

vacuumdb

vacuumdb — garbage-collect and analyze a PostgreSQL database

大纲

`vacuumdb` [*connection-option*...] [[--full] | [-f]] [[--freeze] | [-F]] [[--verbose] | [-v]] [[--analyze] | [-z]] [[--analyze-only] | [-Z]] [--table | -t *table* [(*column* [,...])]] [*dbname*]

`vacuumdb` [*connection-option*...] [[--full] | [-f]] [[--freeze] | [-F]] [[--verbose] | [-v]] [[--analyze] | [-z]] [[--analyze-only] | [-Z]] [[--all] | [-a]]

描述

vacuumdb 是用于清理 PostgreSQL 数据库的工具。vacuumdb 还会生成供 PostgreSQL 查询优化器使用的内部统计信息。

vacuumdb是 SQL 命令VACUUM的包装器。通过该工具或通过访问服务器的其他方式来清理和分析数据库，在效果上没有区别。

选项

vacuumdb接受下列命令行参数：

`-a`
`--all`

清理所有数据库。

`[-d]` *dbname*
`[--dbname=]`*dbname*

指定要清理或分析的数据库名称。如果未指定并且没有使用 `-a`（或`--all`），则从环境变量 `PGDATABASE` 中读取数据库名。如果该变量未设置，则使用连接时指定的用户名。

`-e`
`--echo`

回显 vacuumdb 生成并发送给服务器的命令。

`-f`
`--full`

执行“完全”清理。

`-F`
`--freeze`

激进地“冻结”元组。

`-q`
`--quiet`

不显示进度消息。

`-t` *`table`* `[ (`*`column`* `[,...]) ]`
`--table=`*`table`* `[ (`*`column`* `[,...]) ]`

只清理或分析 *`table`*。列名只能与 `--analyze` 或 `--analyze-only` 选项一起指定。

> **提示**
>
> 如果指定了列，你很可能需要在 shell 中转义括号。（见下面的示例。）

`-v`
`--verbose`

在处理过程中输出详细信息。

`-V`
`--version`

输出 vacuumdb 的版本并退出。

`-z`
`--analyze`

同时计算供优化器使用的统计信息。

`-Z`
`--analyze-only`

仅计算供优化器使用的统计信息（不执行清理）。

`-?`
`--help`

显示 vacuumdb 命令行参数的帮助并退出。

vacuumdb还接受下列用于连接参数的命令行参数：

`-h` *`host`*
`--host=`*`host`*

指定服务器所在机器的主机名。如果该值以斜杠开头，则它会被用作 Unix 域套接字目录。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接所用的 TCP 端口，或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username=`*`username`*

连接时使用的用户名。

`-w`
`--no-password`

绝不提示输入密码。如果服务器要求密码认证，而又无法通过其他方式获得密码，例如 `pgpass` 文件，则连接尝试将失败。此选项可用于无人输入密码的批处理作业和脚本。

`-W`
`--password`

强制 vacuumdb 在连接数据库之前提示输入密码。

这个选项绝非必需，因为如果服务器要求密码认证，vacuumdb 会自动提示输入密码。不过，vacuumdb 会先浪费一次连接尝试，才得知服务器需要密码。在某些情况下，输入 `-W` 以避免这次额外的连接尝试是值得的。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

与大多数其他 PostgreSQL 工具一样，此工具也使用 libpq 支持的环境变量（见第 31.13 节）。

诊断

如遇困难，请参见VACUUM和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须在目标主机上运行。此外，libpq 前端库所使用的任何默认连接设置和环境变量也都会生效。

注解

vacuumdb可能需要多次连接到PostgreSQL服务器，每次都要询问密码。在这种情况下，使用`~/.pgpass`文件会比较方便。更多信息见第 31.14 节。

示例

清理数据库 `test`：

```
$ vacuumdb test
```

清理并为优化器分析数据库 `bigdb`：

```
$ vacuumdb --analyze bigdb
```

清理单个表 `foo`，它位于数据库 `xyzzy` 中，并为优化器分析该表的单列 `bar`：

```
$ vacuumdb --analyze --verbose --table 'foo(bar)' xyzzy
```

另见

VACUUM

PostgreSQL 服务器端应用

本部分包含PostgreSQL服务器端应用和支持工具的参考信息。
这些命令只有在数据库服务器所在的主机上运行时才有意义。其他工具程序列在PostgreSQL 客户端应用中。

目录

initdb 1455
pg_controldata 1458
pg_ctl 1459
pg_resetxlog 1464
postgres 1466
postmaster 1473

initdb

initdb — create a new PostgreSQL database cluster

大纲

initdb [*option*...] [--pgdata] | [-D]*directory*

描述

initdb创建一个新的PostgreSQL数据库集簇。
数据库集簇是由单个服务器实例管理的一组数据库。

创建数据库集簇包括创建用于存放数据库数据的目录，
生成共享系统目录表（属于整个集簇而不是某个特定数据库的表），以及创建template1
和postgres数据库。以后创建新数据库时，会复制template1数据库中的所有内容。（因此，
安装在template1中的任何东西都会自动复制到以后创建的每个数据库中。）postgres
数据库是一个默认数据库，供用户、工具程序和第三方应用程序使用。

虽然initdb会尝试创建指定的数据目录，但如果所需数据目录的父目录归 root 所有，它可能没有足够的权限。要在这种环境中初始化，可先由 root 创建一个空的数据目录，然后用chown将该目录的所有权赋予数据库用户账户，再用 su切换为该数据库用户来运行initdb。

initdb必须以将拥有服务器进程的用户身份运行，因为服务器需要访问 initdb创建的文件和目录。由于服务器不能以 root 身份运行，因此也绝不能以 root 身份运行initdb。（实际上它会拒绝这样做。）

initdb会初始化数据库集簇的默认区域设置和字符集编码。字符集编码、排序顺序（LC_COLLATE）和字符集分类（LC_CTYPE，例如大写、小写、数字）
可以在创建数据库时单独设置。initdb为template1数据库确定这些设置，
它们将作为所有其他数据库的默认值。

要更改默认排序顺序或字符集分类，请使用--lc-collate和--lc-ctype选项。使用C
或POSIX之外的排序顺序也会带来性能损失。因此，在运行initdb
时选择正确的区域设置很重要。

其余区域设置类别可以在以后启动服务器时更改。
还可以使用--locale设置所有区域设置类别的默认值，包括排序顺序和字符集分类。
所有服务器区域设置值（lc_*）都可以通过SHOW ALL显示。更多细节见第 22.1 节。

要修改默认编码，请使用--encoding。更多细节见第 22.3 节。

选项

-A *authmethod*
--auth=*authmethod*

此选项指定 pg_hba.conf 中用于本地用户的认证方法。除非你信任系统上的所有本地用户，否则不要使用 trust。为了便于安装，默认使用 trust。

`-D` *directory*
`--pgdata=`*directory*

该选项指定数据库集簇应存放的目录。这是`initdb`所需的唯一信息，但也可以通过设置`PGDATA`环境变量来省去显式写出它；这通常更方便，因为数据库服务器（`postgres`）之后也可以通过同一变量找到数据目录。

`-E` *encoding*
`--encoding=`*encoding*

选择模板数据库的编码。这也会成为以后创建的任何数据库的默认编码，除非在创建时覆盖它。默认值由区域设置推导而来，如果无法推导，则使用`SQL_ASCII`。PostgreSQL服务器支持的字符集在第 22.3.1 节中介绍。

`--locale=`*locale*

设置数据库集簇的默认区域设置。如果未指定该选项，区域设置将继承自`initdb`运行时所在的环境。区域设置支持见第 22.1 节。

`--lc-collate=`*locale*
`--lc-ctype=`*locale*
`--lc-messages=`*locale*
`--lc-monetary=`*locale*
`--lc-numeric=`*locale*
`--lc-time=`*locale*

类似于`--locale`，但只在指定的类别中设置区域设置。

`--no-locale`

等价于`--locale=C`。

`--pwfile=`*filename*

使`initdb`从文件中读取数据库超级用户的密码。文件的第一行会被当作密码。

`--text-search-config=`*CFG*

设置默认文本检索配置。更多信息见default_text_search_config。

`-U` *username*
`--username=`*username*

选择数据库超级用户的用户名。默认值是运行`initdb`的有效用户的名称。超级用户的名称本身并不重要，不过即使操作系统用户名称不同，也可以选择沿用惯常的名称 postgres。

`-W`
`--pwprompt`

使`initdb`提示输入要赋给数据库超级用户的密码。如果不打算使用密码认证，这一点并不重要。否则，在设置密码之前将无法使用密码认证。

`-X` *directory*
`--xlogdir=`*directory*

该选项指定事务日志应存放的目录。

还可以使用以下较少用到的选项：

`-d`
`--debug`

打印引导后端的调试输出，以及少量普通用户通常不感兴趣的其他消息。引导后端是 `initdb` 用来创建系统目录表的程序。该选项会产生大量极其乏味的输出。

`-L` *`directory`*

指定`initdb`初始化数据库集簇时应到哪里查找其输入文件。通常不需要这样做。若需要显式指定其位置，系统会提示。

`-n`
`--noclean`

默认情况下，如果`initdb`发现某个错误使其无法完整创建数据库集簇，就会删除它在发现无法完成任务之前可能已创建的所有文件。该选项会禁止这种清理，因此对调试有用。

`-V`
`--version`

打印 initdb 的版本并退出。

`-?`
`--help`

显示有关initdb命令行参数的帮助并退出。

环境

`PGDATA`

指定数据库集簇应存放的目录；可使用`-D`选项覆盖。

与大多数其他PostgreSQL工具程序一样，此工具也使用libpq支持的环境变量（参见第 31.13 节）。

注解

也可以通过`pg_ctl initdb`调用`initdb`。

参见

pg_ctl, postgres

pg_controldata

pg_controldata — display control information of a PostgreSQL database cluster

大纲

`pg_controldata` [*option*] [*datadir*]

描述

`pg_controldata`会打印在`initdb`期间初始化的信息，例如系统目录版本。
它还会显示有关预写式日志和检查点处理的信息。这些信息是整个集簇范围内的，
不针对任何单个数据库。

该工具只能由初始化该集簇的用户运行，因为它需要对数据目录具有读取权限。
可以在命令行中指定数据目录，也可以使用环境变量`PGDATA`。该工具支持`-V`和`--version`选项，用于输出 pg_controldata的版本并退出。它还支持`-?`和`--help`选项，用于输出可用参数。

环境

`PGDATA`

默认数据目录位置

pg_ctl

pg_ctl — initialize, start, stop, or control a PostgreSQL server

大纲

`pg_ctl` init[db] [-s] [-D *`datadir`*] [-o *`initdb-options`*]

`pg_ctl` start [-w] [-t *`seconds`*] [-s] [-D *`datadir`*] [-l *`filename`*] [-o *`options`*] [-p *`path`*] [-c]

`pg_ctl` stop [-W] [-t *`seconds`*] [-s] [-D *`datadir`*] [-m [s[mart]] | [f[ast]] | [i[mmediate]]]

`pg_ctl` restart [-w] [-t *`seconds`*] [-s] [-D *`datadir`*] [-c] [-m [s[mart]] | [f[ast]] | [i[mmediate]]] [-o *`options`*]

`pg_ctl` reload [-s] [-D *`datadir`*]

`pg_ctl` status [-D *`datadir`*]

`pg_ctl` promote [-s] [-D *`datadir`*]

`pg_ctl` kill *`signal_name process_id`*

`pg_ctl` register [-N *`servicename`*] [-U *`username`*] [-P *`password`*] [-D *`datadir`*] [-S [a[uto]] | [d[emand]]] [-w] [-t *`seconds`*] [-s] [-o *`options`*]

`pg_ctl` unregister [-N *`servicename`*]

描述

pg_ctl是一个实用工具，用于初始化PostgreSQL数据库集簇，启动、停止或重启PostgreSQL数据库服务器（postgres），或者显示正在运行服务器的状态。虽然服务器也可以手工启动，但pg_ctl将重定向日志输出、正确地与终端和进程组脱离等任务封装了起来。它还提供了便于实施受控关闭的选项。

`init`或`initdb`模式会创建一个新的PostgreSQL数据库集簇，也就是由单个服务器实例管理的一组数据库。该模式会调用`initdb`命令。详见initdb。

`start`模式会启动一个新服务器。服务器在后台启动，其标准输入连接到`/dev/null`（在Windows 上则是`nul`）。在类 Unix 系统上，默认情况下，服务器的标准输出和标准错误会被发送到pg_ctl的标准输出（而不是标准错误）。因此，pg_ctl的标准输出应被重定向到文件，或者通过管道传给另一个进程，例如rotatelogs这样的日志轮转程序；否则`postgres`会在后台将其输出写入控制终端，并且不会脱离 shell 的进程组。在 Windows 上，默认情况下，服务器的标准输出和标准错误会被发送到终端。使用`-l`将服务器输出追加到日志文件，可以改变这些默认行为。建议使用`-l`或输出重定向。

`stop`模式会关闭在指定数据目录中运行的服务器。可以使用`-m`选项选择三种不同的关闭方法。“Smart”模式（默认值）等待所有活动客户端断开，以及所有在线备份结束。如果服务器处于热备状态，那么在所有客户端断开后，恢复和流复制都会终止。“Fast”模式不等待客户端断开，并会终止正在进行的在线备份。所有活动事务都会回滚，客户端会被强制断开，

然后服务器关闭。“Immediate”模式会立即中止所有服务器进程，而不执行干净关闭。
这样会导致服务器在下次启动时进入一次崩溃恢复周期。

`restart` 模式实际上是先执行停止再执行启动。这允许更改 `postgres` 的命令行选项。

`reload`模式只是向`postgres`服务器进程发送一个SIGHUP信号，使其重新读取配置文件（`postgresql.conf`、`pg_hba.conf`等）。
这样就可以修改那些无需完全重启服务器即可生效的配置文件选项。

`status` 模式检查在指定数据目录中是否有一个服务器正在运行。如果有，将显示它的 PID 以及调用它时使用的命令行选项。

在`promote`模式下，会指示在指定数据目录中运行的备库退出恢复，并开始进行读写操作。

`kill`模式向指定进程发送信号。这在没有内置kill命令的Microsoft Windows上尤其有用。使用`--help`可查看受支持的信号名称列表。

`register`模式允许在Microsoft Windows上注册一个系统服务。`-S`选项允许选择服务启动类型，可以是“auto”（系统启动时自动启动服务）或“demand”（按需启动服务）。

`unregister`模式会在Microsoft Windows上注销一个系统服务。这会撤销`register`命令的效果。

选项

`-c`

在支持的平台上，通过解除对核心转储文件施加的任何软资源限制，
尝试允许服务器在崩溃时生成核心转储文件。
这样就可以从失败的服务器进程中获得栈跟踪，从而有助于调试或诊断问题。

`-D` *`datadir`*

指定数据库配置文件所在的文件系统位置。若省略此选项，则使用环境变量`PGDATA`。

`-l` *`filename`*

将服务器日志输出追加到*`filename`*。如果该文件不存在，就会创建它。umask被设置为077，因此默认情况下其他用户无法访问该日志文件。

`-m` *`mode`*

指定关闭模式。*`mode`*可以是`smart`、`fast`或`immediate`，也可以是这三者之一的首字母。
若省略此选项，则使用`smart`。

`-o` *`options`*

指定要直接传递给`postgres`命令的选项。

通常应将选项放在单引号或双引号中，以确保它们作为一个整体被传递。

`-o` *`initdb-options`*

指定要直接传递给`initdb`命令的选项。

通常应将选项放在单引号或双引号中，以确保它们作为一个整体被传递。

`-p` *path*

指定postgres可执行程序的位置。默认情况下，postgres可执行程序取自与pg_ctl相同的目录；如果那里没有，则取自硬编码的安装目录。除非采用了某些非常规方式，并收到找不到postgres可执行程序的错误，否则通常不需要使用此选项。

在init模式中，此选项同样指定initdb可执行程序的位置。

`-s`

只打印错误，不打印信息性消息。

`-t`

等待启动或关闭完成时最多等待多少秒。默认值为环境变量PGCTLTIMEOUT的值；如果未设置该环境变量，则默认为 60 秒。

`-w`

等待启动或关闭完成。等待是关闭操作的默认选项，但不是启动操作的。等待启动时，pg_ctl 会反复尝试连接服务器。等待关闭时，pg_ctl 会等待服务器移除其 PID 文件。pg_ctl 会根据启动或关闭是否成功返回相应的退出代码。

`-W`

不等待启动或关闭完成。对 start 和 restart 模式来说，这是默认行为。

用于 Windows 的选项

`-N` *servicename*

要注册的系统服务名称。该名称会同时用作服务名和显示名。

`-P` *password*

用于启动该服务的用户密码。

`-S` *start-type*

要注册的系统服务的启动类型。start-type 可以是auto、demand，或者两者之一的首字母。若省略此选项，则使用auto。

`-U` *username*

用于启动该服务的用户名。对于域用户，请使用DOMAIN\username格式。

环境

PGCTLTIMEOUT

等待启动或关闭完成时，默认的等待秒数上限。如果未设置，默认值为 60 秒。

PGDATA

默认数据目录位置。

pg_ctl与大多数其他PostgreSQL工具一样，也使用libpq所支持的环境变量（参见第 31. 13 节）。其他服务器相关变量见postgres。

文件

postmaster.pid

数据目录中该文件的存在与否，用于帮助pg_ctl确定服务器当前是否正在运行。

postmaster.opts

如果该文件存在于数据目录中，pg_ctl（在restart模式下）
会将该文件的内容作为选项传递给postgres，除非被-o选项覆盖。
该文件的内容也会在status模式下显示出来。

示例

启动服务器

要启动服务器：

```
$ pg_ctl start
```

要启动服务器并等待其开始接受连接：

```
$ pg_ctl -w start
```

要使用端口 5433 启动服务器，并在未启用fsync的情况下运行，可使用：

```
$ pg_ctl -o "-F -p 5433" start
```

停止服务器

要停止服务器，可使用：

```
$ pg_ctl stop
```

-m选项允许控制服务器如何关闭：

```
$ pg_ctl stop -m fast
```

重启服务器

重启服务器几乎等同于先停止服务器再重新启动，只不过pg_ctl
会保存并重用传给前一个运行实例的命令行选项。以最简单的形式重启服务器，可使用：

```
$ pg_ctl restart
```

要重启服务器并等待其关闭和重启完成：

```
$ pg_ctl -w restart
```

要使用端口 5433 重启服务器，并在重启时禁用fsync，可使用：

```
$ pg_ctl -o "-F -p 5433" restart
```

显示服务器状态

下面是pg_ctl状态输出的示例：

```
$ pg_ctl status
pg_ctl: server is running (PID: 13718)
/usr/local/pgsql/bin/postgres "-D" "/usr/local/pgsql/data" "-p" "5433"
 "-B" "128"
```

这是在重启模式下会调用的命令行。

参见

initdb, postgres

pg_resetxlog

pg_resetxlog — reset the write-ahead log and other control information of a PostgreSQL database cluster

大纲

pg_resetxlog [-f] [-n] [-o*oid*] [-x *xid*] [-e *xid_epoch*] [-m *mxid*] [-O *mxoff*] [-l *timelineid*,*fileid*,*seg*] *datadir*

描述

pg_resetxlog 清空事务日志（WAL），并有选择地重置 pg_control 文件中存储的一些其他控制信息。当这些文件损坏时，有时就需要这项功能。它应当只作为最后的手段使用，即当服务器因这种损坏而无法启动时。

运行这条命令之后，应当就能启动服务器了，但请记住，由于部分提交的事务，数据库中可能包含不一致的数据。你应当立即转储数据、运行 initdb 并重新装载。重新装载之后，检查不一致之处并按需修复。

这个工具只能由安装服务器的用户运行，因为它需要对数据目录的读写访问权限。出于安全原因，你必须在命令行上指定数据目录。pg_resetxlog 不使用环境变量 PGDATA。

如果 pg_resetxlog 抱怨说无法确定 pg_control 的有效数据，你可以通过指定 -f（强制）选项迫使它继续执行。在这种情况下，缺失的数据会用貌似合理的值代替。大多数字段都可以期望是匹配的，但下一个 OID、下一个事务 ID 及其纪元、下一个多事务 ID 及偏移量、WAL 起始地址这些字段可能需要人工协助。这些字段可以用下面讨论的选项来设置。如果你无法为所有这些字段确定正确的值，仍可以使用 -f，但恢复出的数据库必须受到比平常更多的怀疑：必须立即转储并重新装载。在转储之前不要在数据库中执行任何修改数据的操作，因为任何此类操作都很可能使损坏加剧。

-o、-x、-e、-m、-O 和 -l 选项允许手动设置下一个 OID、下一个事务 ID、下一个事务 ID 的纪元、下一个和最旧的多事务 ID、下一个多事务偏移量以及 WAL 起始地址。只有当 pg_resetxlog 无法通过读取 pg_control 确定合适的值时，才需要这些选项。安全值可以按如下方法确定：

- 下一个事务 ID（-x）的安全值可以通过在数据目录下的 pg_clog 目录中查找数值最大的文件名、加一、再乘以 1048576 来确定。注意文件名为十六进制。通常最简单的做法也是用十六进制指定选项值。例如，如果 0011 是 pg_clog 中最大的项，则 -x 0x1200000 可行（五个尾随零提供正确的乘数）。
- 下一个多事务 ID（-m）的安全值可以这样确定：在数据目录下的 pg_multixact/offsets 目录中找出数值最大的文件名，加一，再乘以 65536。如上所述，文件名是十六进制的，所以最简单的做法是用十六进制指定开关值并加上四个零。
- 下一个多事务偏移（-O）的安全值可以这样确定：在数据目录下的 pg_multixact/members 目录中找出数值最大的文件名，加一，再乘以 65536。如上所述，文件名是十六进制的，所以最简单的做法是用十六进制指定开关值并加上四个零。
- WAL 起始地址（-l）应当大于数据目录下 pg_xlog 目录中现存的任何 WAL 段文件名。这些名称也是十六进制的，并且由三部分组成。第一部分是“时间线 ID”，通常应保持不变。第三部分不要选择大于 255（0xFF）的值；应当递增第二部分并把第三部分重置为 0。例如，如果 000000010000000320000004A 是 pg_xlog 中最大的项，-l 0x1,0x32,0x4B 即可；但如果最大的项是 000000010000003A000000FF，请选择 -l 0x1,0x3B,0x0 或更大。

> **注意**
>
> `pg_resetxlog` 本身会查看 `pg_xlog` 中的文件，并选择一个超出最后一个现存文件名的默认 `-l` 设置。因此，只有当你知道存在当前不在 `pg_xlog` 中的 WAL 段文件（例如离线归档中的条目），或者 `pg_xlog` 的内容已完全丢失时，才需要手动调整 `-l`。

- 没有同样简单的方法来确定一个超出数据库中最大 OID 的下一个 OID，但幸运的是，下一个 OID 设置得是否正确并不关键。
- 事务 ID 纪元实际上并不存储在数据库中的任何地方（由 `pg_resetxlog` 设置的字段除外），因此就数据库本身而言，任何值都可以。你可能需要调整这个值以确保 Slony-I 等复制系统正常工作—— 如果需要，应该可以从下游复制数据库的状态中获得合适的值。

`-n`（无操作）开关指示 `pg_resetxlog` 打印从 `pg_control` 重构出的值，然后不做任何修改直接退出。This is mainly a debugging tool, but can be useful as a sanity check before allowing `pg_resetxlog` to proceed for real.

`-V` 和 `--version` 选项打印 pg_resetxlog 的版本并退出。`-?` 和 `--help` 选项显示支持的参数，并退出。

注解

服务器正在运行时不得使用这条命令。如果 `pg_resetxlog` 在数据目录中发现服务器锁文件，它将拒绝启动。如果服务器崩溃过，可能会留下锁文件；这种情况下你可以删除锁文件以允许 `pg_resetxlog` 运行。但在这样做之前，请务必再三确认没有仍然存活的服务器进程。

postgres

postgres — PostgreSQL database server

大纲

postgres [*option*...]

描述

postgres是 PostgreSQL数据库服务器。客户端应用要访问数据库时，会（通过网络或本地）连接到一个正在运行的postgres实例。该postgres
实例随后会启动一个独立的服务器进程来处理该连接。

一个postgres实例始终只管理一个数据库集簇的数据。
数据库集簇是一组存储在同一文件系统位置（“数据区域”）中的数据库。
只要使用不同的数据区域和不同的通信端口（见下文），一个系统上就可以同时运行多个postgres实例。当 postgres启动时，它需要知道数据区域的位置。该位置必须通过 -D选项或PGDATA环境变量指定；没有默认值。通常，-D或 PGDATA会直接指向由initdb创建的数据区域目录。其他可能的文件布局见第 18.2 节。

默认情况下，postgres在前台启动，并将日志消息打印到标准错误流。在实际应用中，postgres 应作为后台进程启动，例如在系统启动时启动。

postgres命令也可以在单用户模式下调用。
这种模式的主要用途是在initdb执行引导期间使用。有时也会用它来调试或进行灾难恢复；
但请注意，运行单用户服务器其实并不适合调试服务器，
因为不会发生真实的进程间通信和锁定。从 shell 以单用户模式调用时，用户可以输入查询，结果会打印到屏幕上，但其形式对开发者更有用，而非面向终端用户。在单用户模式下，
会话用户会被设置为 ID 为 1 的用户，并向该用户隐式授予超级用户权限。该用户实际上不必存在，因此单用户模式可用于手工恢复系统目录遭受某些意外损坏的情况。

选项

postgres接受以下命令行参数。有关这些选项的详细讨论，请参阅第 18 章。通过设置配置文件，你可以省去键入其中大多数选项的麻烦。某些（安全的）
选项还可以由连接的客户端以应用相关的方式设置，并且仅对该会话生效。例如，
如果设置了环境变量PGOPTIONS，那么基于libpq的客户端就会将该字符串传给服务器，
服务器会将其解释为 postgres命令行选项。

通用选项

-A 0|1

启用运行时断言检查，这是一种帮助检测编程错误的调试辅助手段。只有在编译 PostgreSQL 时启用了断言，此选项才可用；如果是这样，默认值为开。

-B *nbuffers*

设置供服务器进程使用的共享缓冲区数量。该参数的默认值由initdb自动选择。
指定此选项等效于设置shared_buffers配置参数。

`-c` *`name=value`*

设置一个命名的运行时参数。PostgreSQL支持的配置参数见第 18 章。大多数其他命令行选项实际上都是这种参数赋值的简写形式。`-c`可以出现多次以设置多个参数。

`-d` *`debug-level`*

设置调试级别。该值设得越高，写入服务器日志的调试输出就越多。取值范围是 1 到 5。还可以针对某个特定会话传入`-d 0`，从而阻止父`postgres`进程的服务器日志级别传播到该会话。

`-D` *`datadir`*

指定数据目录或配置文件的文件系统位置。详见第 18.2 节。

`-e`

将默认日期风格设置为“European”，即输入日期字段采用 `DMY`顺序。这也会在某些日期输出格式中使日期显示为日在月前。更多信息见第 8.5 节。

`-F`

禁用`fsync`调用以提高性能，但在系统崩溃时会有数据损坏的风险。指定此选项等效于禁用fsync配置参数。使用前请先阅读详细文档！

`-h` *`hostname`*

指定`postgres`用来侦听来自客户端应用的 TCP/IP 连接的 IP 主机名或地址。该值也可以是逗号分隔的地址列表，或者用`*`指定侦听所有可用接口。空值表示不侦听任何 IP 地址，在这种情况下只能使用 Unix 域套接字连接到服务器。默认只侦听 localhost。指定此选项等效于设置listen_addresses配置参数。

`-i`

允许远程客户端通过 TCP/IP（Internet 域）连接。没有此选项时，只接受本地连接。该选项等效于在 `postgresql.conf`中，或通过`-h`将 `listen_addresses`设置为`*`。

该选项已被弃用，因为它无法访问listen_addresses的全部功能。
通常最好直接设置`listen_addresses`。

`-k` *`directory`*

指定 `postgres` 监听来自客户端应用连接的 Unix 域套接字的目录。默认通常是 `/tmp`，但可以在构建时更改。

`-l`

启用使用SSL的安全连接。若要使用此选项，PostgreSQL必须在编译时启用 SSL支持。关于使用SSL的更多信息，请参阅第 17.9 节。

`-N` *`max-connections`*

设置该服务器将接受的最大客户端连接数。该参数的默认值由initdb自动选择。
指定此选项等效于设置max_connections配置参数。

`-o` *`extra-options`*

在*`extra-options`*中指定的命令行形式选项，会传给由这个`postgres`
进程启动的所有服务器进程。如果选项字符串中包含空格，整个字符串必须加引号。

此选项已过时；服务器进程的所有命令行选项都可以直接在postgres命令行上指定。

-p *port*

指定postgres 用来侦听客户端应用连接的 TCP/IP 端口或本地 Unix 域套接字文件扩展名。默认值取自PGPORT环境变量；如果PGPORT未设置，
则默认值就是在编译期间确定的值（通常为 5432）。如果你指定了非默认端口，那么所有客户端应用都必须通过命令行选项或PGPORT指定相同的端口。

-s

在每条命令结束时打印时间信息和其他统计信息。
这对基准测试或者调整缓冲区数量很有用。

-S *work-mem*

指定内部排序和散列操作在转而使用临时磁盘文件之前可使用的内存量。有关work_mem
配置参数的说明，请参阅第 18.4.1 节。

--*name=value*

设置一个命名的运行时参数；这是-c的较短形式。

--describe-config

该选项会以制表符分隔的COPY格式导出服务器内部配置变量、说明以及默认值。
它主要是为管理工具设计的。

半内部选项

这里描述的选项主要用于调试，并且在某些情况下可协助恢复严重损坏的数据库。
在生产数据库环境中没有理由使用它们。这里列出这些选项，仅供PostgreSQL 系统开发者使用。此外，这些选项在未来版本中可能会更改或被移除，且不另行通知。

-f { s | i | m | n | h }

禁止使用特定的扫描和连接方法：s 和 i 分别禁用顺序扫描和索引扫描，而 n、m 和 h 分别禁用嵌套循环、归并和哈希连接。

顺序扫描和嵌套循环连接都无法被完全禁用；-fs和 -fn选项只是在优化器有其他选择时，尽量不使用这些计划类型。

-n

此选项用于调试导致服务器进程异常终止的问题。在这种情况下，
通常的策略是通知所有其他服务器进程必须终止，然后重新初始化共享内存和信号量。
这是因为出错的服务器进程可能在终止前破坏了某些共享状态。此选项指定postgres
不重新初始化共享数据结构。
经验丰富的系统程序员随后可以使用调试器检查共享内存和信号量的状态。

-O

允许修改系统表的结构。该选项由 initdb使用。

-P

读取系统表时忽略系统索引，但在修改表时仍然更新这些索引。
这在从损坏的系统索引中恢复时很有用。

`-t pa[rser] | pl[anner] | e[xecutor]`

打印与各个主要系统模块相关的每个查询的时间统计信息。该选项不能与`-s`选项一起使用。

`-T`

此选项用于调试导致服务器进程异常终止的问题。在这种情况下，
通常的策略是通知所有其他服务器进程必须终止，然后重新初始化共享内存和信号量。
这是因为出错的服务器进程可能在终止前破坏了某些共享状态。此选项指定`postgres`
通过发送`SIGSTOP`信号暂停所有其他服务器进程，但不使它们终止。
这允许系统程序员手工收集所有服务器进程的核心转储。

`-v` *`protocol`*

指定特定会话要使用的前端/后端协议版本号。该选项仅供内部使用。

`-W` *`seconds`*

新服务器进程在完成认证过程后，会延迟这么多秒。
这是为了给调试器附着到该服务器进程提供机会。

用于单用户模式的选项

以下选项仅适用于单用户模式。

`--single`

选择单用户模式。这必须是命令行上的第一个参数。

`database`

指定要访问的数据库名称。这必须是命令行上的最后一个参数。如果省略，
则默认为用户名。

`-E`

回显所有命令。

`-j`

禁用换行符作为语句分隔符。

`-r` *`filename`*

将所有服务器日志输出发送到*`filename`*。在正常的多用户模式下，该选项会被忽略，
所有进程都使用stderr。

环境

`PGCLIENTENCODING`

客户端使用的默认字符编码。（客户端可以单独覆盖它。）该值也可以在配置文件中设置。

`PGDATA`

默认数据目录位置

PGDATESTYLE

DateStyle运行时参数的默认值。（该环境变量的用法已被弃用。）

PGPORT

默认端口号（最好在配置文件中设置）

TZ

Server time zone

诊断

提到semget或 shmget的失败消息，很可能表明你需要配置内核，以提供足够的共享内存和信号量。更多讨论见第 17.4 节。你也许可以通过降低shared_buffers来减少PostgreSQL的共享内存消耗，以及/或通过降低max_connections来减少信号量消耗，从而推迟重新配置内核。

表明另一个服务器已经在运行的失败消息应仔细检查，例如可根据你的系统使用下面的命令：

```
$ ps ax | grep postgres
```

或

```
$ ps -ef | grep postgres
```

如果你确信没有冲突的服务器在运行，可以删除消息中提到的锁文件，然后重试。

表明无法绑定端口的失败消息，可能意味着该端口已经被某个非PostgreSQL进程占用。
你在终止postgres 后如果立即使用同一端口重新启动它，也可能会得到这个错误；在这种情况下，你只需等待几秒钟，直到操作系统关闭该端口后再重试。最后，
如果你指定了一个操作系统认为是保留的端口号，也可能会收到这个错误。例如，许多版本的 Unix 认为低于 1024 的端口号是“受信任的”，并且只允许 Unix 超级用户访问它们。

注解

实用命令pg_ctl可用于安全、便捷地启动和关闭postgres服务器。

只要有可能，就不要使用 SIGKILL杀死主 postgres服务器。这样会阻止 postgres在终止前释放其持有的系统资源（例如共享内存和信号量）。这可能会导致新的postgres
实例在启动时出现问题。

要正常终止postgres服务器，可以使用 SIGTERM、SIGINT或 SIGQUIT信号。第一个会在退出前等待所有客户端终止，第二个会强制断开所有客户端，
而第三个会在不进行正确关闭的情况下立即退出，从而在重启期间导致一次恢复运行。

SIGHUP信号会重新加载服务器配置文件。也可以向单个服务器进程发送 SIGHUP，但这通常没有意义。

要取消一个正在运行的查询，可向运行该命令的进程发送 SIGINT 信号。

postgres 服务器使用 SIGTERM 通知下级服务器进程正常退出，使用 SIGQUIT 在不做正常清理的情况下终止。这些信号不应由用户使用。向服务器进程发送 SIGKILL 同样是不明智的—主 postgres 进程会将其解释为崩溃，并作为标准崩溃恢复过程的一部分强制所有兄弟进程退出。

缺陷

--选项在FreeBSD或OpenBSD上不起作用。请改用-c。这是受影响操作系统中的一个缺陷；
如果这个问题未被修复，未来版本的PostgreSQL 将提供变通方案。

用法

要启动单用户模式服务器，可以使用类似下面的命令：

postgres --single -D /usr/local/pgsql/data *other-options* my_database

使用-D提供正确的数据库目录路径，或者确保已经设置了环境变量PGDATA。
还要指定你想要操作的那个数据库的名称。

通常，单用户模式服务器将换行视为命令输入终止符；它不像psql那样会对分号进行智能处理。
要让一个命令跨越多行，必须在除最后一个换行之外的每个换行前输入反斜线。

但是，如果你使用-j命令行开关，换行并不会终止命令输入。在这种情况下，
服务器会读取标准输入直到文件末尾（EOF）标记，然后把输入作为单个命令字符串处理。
在这种情况下，反斜线-换行不会被特殊处理。

要退出会话，输入EOF（通常是**Control+D**）。如果你使用了-j，则需要连续输入两次
EOF才能退出。

请注意，单用户模式服务器不提供复杂的行编辑功能（例如没有命令历史）。
单用户模式也不会执行任何后台处理，例如自动检查点。

示例

要使用默认值在后台启动postgres，输入：

```
$ nohup postgres >logfile 2>&1 </dev/null &
```

要用指定端口启动postgres，例如 1234：

```
$ postgres -p 1234
```

要使用psql连接到该服务器，请用 -p 选项指定这个端口：

```
$ psql -p 1234
```

或者设置环境变量PGPORT：

```
$ export PGPORT=1234
$ psql
```

命名的运行时参数可以采用以下任一种形式设置：

```
$ postgres -c work_mem=1234
$ postgres --work-mem=1234
```

两种形式都会覆盖 postgresql.conf中可能存在的 work_mem设置。请注意，参数名中的下划
线在命令行中既可以写成下划线，也可以写成连字符。除了短期实验之外，
与其依赖命令行开关来设置参数，通常更好的做法是直接编辑 postgresql.conf中的设置。

另见

initdb, pg_ctl

postmaster

postmaster — PostgreSQL database server

大纲

`postmaster` [*`option`*...]

描述

`postmaster`是`postgres`的一个已弃用别名。

另见

postgres

部分 VII. 内部

本部分包含一些可能对PostgreSQL开发人员有用的各类信息。

目录

44. PostgreSQL 内部概述 1479
44.1. 查询的路径 1479
44.2. 连接是如何建立的 1479
44.3. 解析器阶段 1480
44.3.1. 解析器 1480
44.3.2. 转换过程 1480
44.4. PostgreSQL 规则系统 1481
44.5. 规划器/优化器 1481
44.5.1. 生成可能的计划 1481
44.6. 执行器 1482
45. 系统目录 1484
45.1. 概述 1484
45.2. pg_aggregate 1485
45.3. pg_am 1486
45.4. pg_amop 1488
45.5. pg_amproc 1489
45.6. pg_attrdef 1489
45.7. pg_attribute 1490
45.8. pg_authid 1492
45.9. pg_auth_members 1493
45.10. pg_cast 1494
45.11. pg_class 1495
45.12. pg_constraint 1498
45.13. pg_collation 1500
45.14. pg_conversion 1501
45.15. pg_database 1501
45.16. pg_db_role_setting 1503
45.17. pg_default_acl 1503
45.18. pg_depend 1504
45.19. pg_description 1505
45.20. pg_enum 1506
45.21. pg_extension 1506
45.22. pg_foreign_data_wrapper 1507
45.23. pg_foreign_server 1508
45.24. pg_foreign_table 1508
45.25. pg_index 1508
45.26. pg_inherits 1511
45.27. pg_language 1511
45.28. pg_largeobject 1512
45.29. pg_largeobject_metadata 1513
45.30. pg_namespace 1513
45.31. pg_opclass 1513
45.32. pg_operator 1514
45.33. pg_opfamily 1515
45.34. pg_pltemplate 1515
45.35. pg_proc 1516
45.36. pg_rewrite 1519
45.37. pg_seclabel 1520
45.38. pg_shdepend 1521
45.39. pg_shdescription 1522

45.40. pg_statistic 1522
45.41. pg_tablespace 1524
45.42. pg_trigger 1525
45.43. pg_ts_config 1526
45.44. pg_ts_config_map 1527
45.45. pg_ts_dict 1527
45.46. pg_ts_parser 1528
45.47. pg_ts_template 1528
45.48. pg_type 1528
45.49. pg_user_mapping 1534
45.50. 系统视图 1534
45.51. pg_available_extensions 1535
45.52. pg_available_extension_versions 1535
45.53. pg_cursors 1536
45.54. pg_group 1537
45.55. pg_indexes 1537
45.56. pg_locks 1537
45.57. pg_prepared_statements 1540
45.58. pg_prepared_xacts 1541
45.59. pg_roles 1541
45.60. pg_rules 1542
45.61. pg_seclabels 1543
45.62. pg_settings 1543
45.63. pg_shadow 1545
45.64. pg_stats 1546
45.65. pg_tables 1548
45.66. pg_timezone_abbrevs 1548
45.67. pg_timezone_names 1549
45.68. pg_user 1549
45.69. pg_user_mappings 1550
45.70. pg_views 1550
46. 前端/后端协议 1552
46.1. 概述 1552
46.1.1. 消息概述 1552
46.1.2. 扩展查询概述 1553
46.1.3. 格式和格式代码 1553
46.2. 消息流 1553
46.2.1. 启动 1553
46.2.2. 简单查询 1555
46.2.3. 扩展查询 1557
46.2.4. 函数调用 1559
46.2.5. COPY操作 1560
46.2.6. 异步操作 1561
46.2.7. 取消正在处理的请求 1561
46.2.8. 终止 1562
46.2.9. SSL会话加密 1562
46.3. 流复制协议 1563
46.4. 消息数据类型 1566
46.5. 消息格式 1566
46.6. 错误和通知消息字段 1581
46.7. 自协议 2.0 以来的变更总结 1582
47. PostgreSQL 编码约定 1584
47.1. 格式 1584

47.2. 在服务器内部报告错误 1584
47.3. 错误消息风格指南 1586
48. 本地语言支持 1591
48.1. 给翻译者 1591
48.1.1. 要求 1591
48.1.2. 概念 1591
48.1.3. 创建和维护消息目录 1592
48.1.4. 编辑 PO 文件 1593
48.2. 给程序员 1593
48.2.1. 实现机制 1593
48.2.2. 消息编写指南 1594
49. 编写过程语言调用处理器 1596
50. 编写外部数据包装器 1599
50.1. 外部数据包装器函数 1599
50.2. 外部数据包装器回调例程 1599
51. 遗传查询优化器 1602
51.1. 将查询处理看成是一个复杂的优化问题 1602
51.2. 遗传算法 1602
51.3. PostgreSQL 中的遗传查询优化（GEQO） 1603
51.3.1. 使用GEQO生成可能的计划 1604
51.3.2. PostgreSQL GEQO的未来实现任务 1604
51.4. 进一步阅读 1604
52. 索引访问方法接口定义 1605
52.1. 索引的目录条目 1605
52.2. 索引访问方法函数 1606
52.3. 索引扫描 1609
52.4. 索引锁定注意事项 1610
52.5. 索引唯一性检查 1611
52.6. 索引代价估算函数 1612
53. GiST 索引 1616
53.1. 简介 1616
53.2. 可扩展性 1616
53.3. 实现 1616
53.4. 示例 1624
54. GIN 索引 1625
54.1. 简介 1625
54.2. 可扩展性 1625
54.3. 实现 1627
54.3.1. GIN 快速更新技术 1627
54.3.2. 部分匹配算法 1628
54.4. GIN 提示和技巧 1628
54.5. 限制 1628
54.6. 示例 1629
55. 数据库物理存储 1630
55.1. 数据库文件布局 1630
55.2. TOAST 1632
55.3. 空闲空间映射 1633
55.4. 可见性映射 1633
55.5. 初始化分支 1634
55.6. 数据库页面布局 1634
56. BKI 后端接口 1637
56.1. BKI 文件格式 1637
56.2. BKI 命令 1637
56.3. 引导 BKI 文件的结构 1638

56.4. 示例 1639
57. 规划器如何使用统计信息 1640
57.1. 行估计示例 1640

第 44 章 PostgreSQL 内部概述

作者

本章最初源自[sim98] Stefan Simkovics 在维也纳技术大学完成的硕士论文，该论文由 O. Univ.Prof.Dr. Georg Gottlob 和 Univ.Ass. Mag. Katrin Seyr 指导。

本章概述 PostgreSQL 后端的内部结构。阅读以下各节之后，你应该能够大致了解一个查询是如何被处理的。本章并不试图详细描述 PostgreSQL 的内部运作，因为那样的文档会非常庞大。相反，本章旨在帮助读者理解，从后端收到一个查询开始，到将结果返回给客户端为止，内部通常有哪些操作步骤。

44.1. 查询的路径

这里将简要概述一个查询为得到结果必须经过的各个阶段。

1. 必须先建立从应用程序到 PostgreSQL 服务器的连接。应用程序将查询发送给服务器，并等待接收服务器返回的结果。

2. 解析器阶段检查应用程序传来的查询是否具有正确的语法，并创建一棵查询树。

3. 重写系统接收由解析器阶段创建的查询树，并查找任何可应用于该查询树的规则（存储在系统目录中）。它执行这些规则体中给出的转换。

   重写系统的一个应用是实现视图。每当对一个视图（即虚拟表）发出查询时，重写系统都会将用户的查询重写成一个改为访问视图定义中给出的基表的查询。

4. 规划器/优化器接收（重写后的）查询树，并创建一个查询计划，将作为执行器的输入。

   它首先生成所有能得到同一结果的可能路径。例如，如果待扫描的某个关系上有一个索引，那么该扫描就有两条路径：一种是简单的顺序扫描，另一种是使用该索引。接着会估算执行每条路径的代价，并选择代价最低的路径。代价最低的路径会被展开成一个完整的计划，供执行器使用。

5. 执行器会递归地遍历计划树，并以计划所表示的方式提取行。执行器在扫描关系时会使用存储系统，执行排序和连接，计算限定条件，最后返回得到的行。

在后续各节中，我们将更详细地介绍上述各项内容，以便更好地理解 PostgreSQL 的内部控制和数据结构。

44.2. 连接是如何建立的

PostgreSQL 使用一种简单的“每用户一个进程”客户端/服务器模型。在这种模型中，一个客户端进程恰好连接到一个服务器进程。由于事先不知道会建立多少个连接，因此必须使用一个主进程，在每次收到连接请求时派生一个新的服务器进程。这个主进程称为 `postgres`，在指定的 TCP/IP 端口上监听传入连接。每当检测到连接请求时，`postgres` 进程就会派生一个新的服务器进程。服务器任务之间使用信号量和共享内存通信，以确保并发数据访问期间的数据完整性。

客户端进程可以是任何理解 PostgreSQL 协议（见第 46 章）的程序。很多客户端基于 C 语言库 libpq，但该协议也有若干独立实现，例如 Java 的 JDBC 驱动。

一旦连接建立起来，客户端进程就可以向后端（服务器）发送查询。查询以纯文本方式传输，也就是说，前端（客户端）不进行解析。服务器解析查询，创建一个执行计划，执行该计划，并通过已建立的连接把检索到的行返回给客户端。

44.3. 解析器阶段

解析器阶段由两部分组成：

- 解析器定义在 `gram.y` 和 `scan.l` 中，它使用 Unix 工具 bison 和 flex 构建。
- 转换过程，它会对解析器返回的数据结构做修改和补充。

44.3.1. 解析器

解析器必须检查查询串（以普通 ASCII 文本形式到达）的语法是否有效。如果语法正确，就构建一棵解析树并返回；否则返回错误。解析器和词法分析器是用著名的 Unix 工具bison和flex实现的。

词法分析器定义在文件 `scan.l` 中，负责识别标识符、SQL 关键字等。每找到一个关键字或标识符，就会生成一个词元并交给解析器。

解析器定义在文件 `gram.y` 中，由一组语法规则和动作组成，每当某条规则被触发时，对应动作就会执行。动作中的代码（实际上是 C 代码）用于构造语法解析树。

文件 `scan.l` 会被转换成 C 源文件 `scan.c`，所用程序是 flex；而 `gram.y` 则会被转换成 `gram.c`，所用程序是 bison。这些转换完成之后，就可以使用普通的 C 编译器来构建解析器。绝不要修改这些生成出来的 C 文件，因为下一次调用 flex 或 bison 时，它们都会被覆盖。

> **注意**
>
> 上述转换和编译通常都是借助 makefiles 自动完成的，这些文件随 PostgreSQL 源代码发行版一同提供。

对 bison 的详细描述，或者对 `gram.y` 中给出的语法规则的说明，都超出了本文的范围。有许多书籍和文档专门讨论 flex 和 bison。在开始研究之前，你应该先熟悉 bison，然后再去看 `gram.y` 中给出的语法，否则你将无法理解其中发生了什么。

44.3.2. 转换过程

解析器阶段仅依据 SQL 语法结构的固定规则创建一棵语法解析树。它不会在系统目录中做任何查找，因此不可能理解所请求操作的详细语义。解析器完成之后，
转换过程会接收解析器返回的树，并进行必要的语义解释，以理解查询引用了哪些表、函数和操作符。为表示这些信息而构建的数据结构称为查询树。

之所以将原始解析与语义分析分开，是因为系统目录查找只能在事务内部进行，
而我们不希望在刚收到查询字符串时就立刻启动事务。
原始解析阶段已经足以识别事务控制命令（`BEGIN`、`ROLLBACK` 等），因此这些命令可以在无需进一步分析的情况下被正确执行。一旦我们知道当前处理的是实际查询（例如 `SELECT` 或 `UPDATE`），如果尚未处于事务中，就可以启动事务。只有在那之后才能调用转换过程。

转换过程创建的查询树在大多数地方的结构都与原始语法解析树相似，但在细节上有许多不同。例如，语法解析树中的一个 `FuncCall` 节点表示某个在语法上看起来像函数调用的东西。它可能

被转换成 `FuncExpr` 节点，也可能被转换成 `Aggref` 节点，这取决于被引用的名称最终被判定为普通函数还是聚合函数。此外，
关于列和表达式结果的实际数据类型的信息也会被加入到查询树中。

44.4. PostgreSQL 规则系统

PostgreSQL 提供了一个强大的规则系统，用于定义视图以及处理有歧义的视图更新。最初，PostgreSQL 的规则系统由两种实现组成：

- 第一种实现使用行级处理，并且深度实现于执行器内部。每当访问到单独一行时，
  规则系统就会被调用。这种实现在 1995 年被移除，当时 Berkeley Postgres 项目的最后一个官方发行版被转换成了 Postgres95。
- 规则系统的第二种实现是一种称为查询重写的技术。
  重写系统是位于解析器阶段与规划器/优化器之间的一个模块。这种技术至今仍在使用。

关于查询重写器，第 37 章中已有相当详细的讨论，因此这里没有必要再展开。我们只指出一点：重写器的输入和输出都是查询树，也就是说，树的表示形式以及语义细节层次都不会发生变化。重写可以被看作某种形式的宏展开。

44.5. 规划器/优化器

规划器/优化器的任务是创建一个最优的执行计划。给定的一个 SQL 查询（因此也就是一棵查询树）实际上可以用多种不同方式执行，而这些方式都会产生同样的结果集。如果在计算上可行，查询优化器将考察这些可能执行计划中的每一种，并最终选择预计运行最快的那个执行计划。

> **注意**
>
> 在某些情况下，考察查询的每一种可能执行方式会耗费过多时间和内存。
> 尤其是在执行涉及大量连接操作的查询时更是如此。
> 为了在合理时间内确定一个合理的（不一定最优的）查询计划，PostgreSQL 会使用遗传查询优化器（见第 51 章），前提是连接数量超过某个阈值（见geqo_threshold）。

规划器的搜索过程实际上使用一种称为路径的数据结构，它只是计划的简化表示，
仅包含规划器做出决策所需的信息。在确定出代价最低的路径之后，会构建一棵完整的
计划树传递给执行器。这棵树以足够细致的方式表示了期望的执行计划，使执行器能够运行它。
在本节余下部分，我们将忽略路径与计划之间的区别。

44.5.1. 生成可能的计划

规划器/优化器首先为查询中使用的每个单独关系（表）生成扫描计划。
可能的计划由每个关系上可用的索引决定。对一个关系总是可以执行顺序扫描，
因此一定会生成顺序扫描计划。假设某个关系上定义了一个索引（例如一个 B-树索引），并且查询中包含限制条件 `relation.attribute OPR constant`。如果 `relation.attribute` 恰好匹配该 B-树索引的键，而且 `OPR` 是该索引操作符类中列出的操作符之一，那么就会生成另一个使用该 B-树索引扫描该关系的计划。如果还存在其他索引，并且查询中的限制条件恰好匹配某个索引的键，则还会考虑更多计划。对于那些排序顺序能够匹配查询 `ORDER BY` 子句（如果存在）或者可能有助于进行归并连接（见下文）的索引，也会生成索引扫描计划。

如果查询需要连接两个或更多关系，那么会在找出所有可行的单关系扫描计划之后，
再考虑关系连接计划。有以下三种可用的连接策略：

- 嵌套循环连接：左关系中找到的每一行，都会使右关系被扫描一次。
  这种策略实现起来很容易，但可能非常耗时。（不过，如果右关系可以通过索引扫描，
  那么这也可能是一种不错的策略。可以把左关系当前行中的值作为右关系索引扫描的键。）

- 归并连接：在连接开始之前，每个关系都会先按照连接属性排序。然后两个关系并行扫描，
  将匹配的行组合成连接结果行。这种连接方式很有吸引力，因为每个关系只需扫描一次。
  所需的排序既可以通过显式排序步骤完成，
  也可以利用连接键上的索引按适当顺序扫描关系来完成。

- 哈希连接：首先扫描右关系，并使用其连接属性作为哈希键将其装入一个哈希表。
  接着扫描左关系，并将找到的每一行中的相应值作为哈希键，
  用来在该哈希表中定位匹配的行。

当查询涉及两个以上的关系时，最终结果必须由一棵连接步骤树构建出来，
每个连接步骤都有两个输入。规划器会考察不同的可能连接顺序，以找出代价最低的那个。

如果查询使用的关系少于geqo_threshold个，就会执行一次近乎穷举的搜索，
以找出最佳连接顺序。对于任意两个关系，只要在 `WHERE` 限定条件中存在相应的连接子句（即存在类似 `where rel1.attr1=rel2.attr2` 这样的限制），规划器就会优先考虑它们之间的连接。没有连接子句的连接对只有在别无选择时才会被考虑，也就是说，
某个关系与其他任何关系之间都没有可用的连接子句。对于规划器所考虑的每一对连接，
都会生成所有可能的计划，并选择其中估计代价最低的那个。

当超过 `geqo_threshold` 时，被考虑的连接顺序将由启发式方法决定，如第 51 章中所述。除此之外，处理过程与前面相同。

最终完成的计划树由基表的顺序扫描或索引扫描，再加上所需的嵌套循环、
归并或哈希连接节点，以及任何需要的辅助步骤（例如排序节点或聚合函数计算节点）组成。
这些计划节点类型中的大多数还具有执行选择（丢弃不满足指定布尔条件的行）
和投影（根据给定列值计算派生列集，也就是在需要时计算标量表达式）的额外能力。
规划器的职责之一，就是把来自 `WHERE` 子句的选择条件以及所需输出表达式的计算附加到计划树中最合适的节点上。

44.6. 执行器

执行器接收由规划器/优化器创建的计划，并递归地处理它，以提取所需的行集合。
这本质上是一种按需拉取的流水线机制。每当某个计划节点被调用时，它必须再交付一行，
或者报告自己已经完成了行的交付。

为了给出一个具体示例，假设顶层节点是一个 `MergeJoin` 节点。在执行归并之前，必须先取到两行（每个子计划各一行）。因此执行器会递归调用自身去处理这些子计划（从挂接在 `lefttree` 上的子计划开始）。新的顶层节点，也就是左子计划的顶层节点，假设是一个 `Sort` 节点，那么又需要通过递归来取得一个输入行。`Sort` 的子节点可能是一个 `SeqScan` 节点，它表示真正去读取一个表。执行该节点会使执行器从表中取出一行，并把它返回给调用者节点。`Sort` 节点会反复调用它的子节点，以取得所有待排序的行。当输入耗尽时（子节点返回的是 NULL 而不是一行），`Sort` 代码就执行排序，并最终能够返回它的第一条输出行，也就是排序后的第一行。它会把其余行保存起来，以便在后续请求中按排序后的顺序交付这些行。

`MergeJoin` 节点同样会向它的右子计划请求第一行。然后它比较这两行，看它们是否可以连接；
如果可以，就向调用者返回一条连接结果行。在下一次调用时，
或者在当前这一对输入无法连接时立即，
它会前进到某一个表的下一行或另一个表的下一行（取决于比较结果），并再次检查是否匹配。
最终，其中一个子计划会耗尽，`MergeJoin` 节点就返回 NULL，以表明无法再形成更多连接结果行。

复杂查询可能涉及许多层计划节点，但总体方法是相同的：每个节点在每次被调用时，
都会计算并返回它的下一条输出行。
每个节点还负责应用规划器分配给它的任何选择或投影表达式。

执行器机制用于求值全部四种基本 SQL 查询类型：SELECT、INSERT、UPDATE 和 DELETE。对于 SELECT，顶层执行器代码只需要把查询计划树返回的每一行发送给客户端。对于 INSERT，每个返回的行会被插入到为该 INSERT 指定的目标表中。这项工作是在一个名为 ModifyTable 的特殊顶层计划节点中完成的。（一个简单的 INSERT ... VALUES 命令会创建一棵极为简单的计划树，它只包含一个计算单条结果行的 Result 节点，以及位于其上执行插入的 ModifyTable。但 INSERT ... SELECT 可能需要执行器机制的全部能力。）对于 UPDATE，规划器会安排每个计算得到的行包含所有被更新列的值，再加上原始目标行的 TID（tuple ID，即元组 ID 或行 ID）；这些数据会被送入 ModifyTable 节点，该节点利用这些信息创建一个新的更新后行，并将旧行标记为已删除。对于 DELETE，计划实际返回的唯一列就是 TID，而 ModifyTable 节点只需利用该 TID 访问每个目标行并将其标记为已删除。

第 45 章 系统目录

系统目录是关系型数据库管理系统存放模式元数据的地方，
例如表和列的信息以及内部记账信息。PostgreSQL 的系统目录是普通表。你可以删除并重建这些表、增加列、插入和更新数值，并借此把系统严重搞坏。通常，不应手工修改系统目录，
总是有 SQL 命令可用于完成这些操作。（例如，`CREATE DATABASE` 会向 `pg_database` 目录中插入一行 – 并且实际上会在磁盘上创建该数据库。）对于某些特别深奥的操作存在一些例外，
例如增加索引访问方法。

45.1. 概述

表 45.1列出了系统目录。每个目录更详细的文档见后文。

大多数系统目录在创建数据库时都会从模板数据库复制过来，因此之后是数据库专有的。
少数目录在物理上由一个集簇中的所有数据库共享；这些目录会在各自的说明中注明。

表 45.1. 系统目录

Catalog Name	Purpose
pg_aggregate	aggregate functions
pg_am	index access methods
pg_amop	access method operators
pg_amproc	access method support procedures
pg_attrdef	column default values
pg_attribute	table columns (“attributes”)
pg_authid	authorization identifiers (roles)
pg_auth_members	authorization identifier membership relati onships
pg_cast	casts (data type conversions)
pg_class	tables, indexes, sequences, views (“relations”)
pg_constraint	check constraints, unique constraints, primary key constraints, foreign key constraints
pg_collation	collations (locale information)
pg_conversion	encoding conversion information
pg_database	databases within this database cluster
pg_db_role_setting	per-role and per-database settings
pg_default_acl	default privileges for object types
pg_depend	dependencies between database objects
pg_description	descriptions or comments on database objects
pg_enum	enum label and value definitions
pg_extension	installed extensions
pg_foreign_data_wrapper	foreign-data wrapper definitions
pg_foreign_server	foreign server definitions
pg_foreign_table	additional foreign table information
pg_index	additional index information

Catalog Name	Purpose
pg_inherits	table inheritance hierarchy
pg_language	languages for writing functions
pg_largeobject	data pages for large objects
pg_largeobject_metadata	metadata for large objects
pg_namespace	schemas
pg_opclass	access method operator classes
pg_operator	operators
pg_opfamily	access method operator families
pg_pltemplate	template data for procedural languages
pg_proc	functions and procedures
pg_rewrite	query rewrite rules
pg_seclabel	security labels on database objects
pg_shdepend	dependencies on shared objects
pg_shdescription	comments on shared objects
pg_statistic	planner statistics
pg_tablespace	tablespaces within this database cluster
pg_trigger	triggers
pg_ts_config	text search configurations
pg_ts_config_map	text search configurations' token mappings
pg_ts_dict	text search dictionaries
pg_ts_parser	text search parsers
pg_ts_template	text search templates
pg_type	data types
pg_user_mapping	mappings of users to foreign servers

45.2. pg_aggregate

目录pg_aggregate存储关于聚合函数的信息。
聚合函数是对一组值（典型的是每个匹配查询条件的行中的同一个列的值）进行操作的函数，它返回从这些值中计算出的单个值。典型的聚合函数是 sum、count和max。pg_aggregate里的每个项都是一个pg_proc项的扩展。pg_proc项记载该聚合的名字、输入和输出数据类型，以及其他一些和普通函数类似的信息。

表 45.2. pg_aggregate 列

Name	Type	References	Description
aggfnoid	regproc	pg_proc.oid	pg_proc OID of the aggregate function
aggtransfn	regproc	pg_proc.oid	Transition function
aggfinalfn	regproc	pg_proc.oid	Final function (zero if none)
aggsortop	oid	pg_operator.oid	Associated sort operat or (zero if none)

Name	Type	References	Description
aggtranstype	oid	pg_type.oid	Data type of the aggregate function's internal transition (state) data
agginitval	text		The initial value of the transition state. This is a text field containing the initial value in its external string repres entation. If this field is null, the transition state value starts out null.

新的聚合函数可通过CREATE AGGREGATE命令注册。更多关于编写聚合函数以及转移函数的含义等信息请参见第 35.10 节。

45.3. pg_am

目录pg_am存储关于索引访问方法的信息。
系统支持的每种索引访问方法在这个目录中都有一行。此目录的内容在第 52 章中详细讨论。

表 45.3. pg_am 列

Name	Type	References	Description
amname	name		Name of the access method
amstrategies	int2		Number of operat or strategies for this access method, or zero if access method does not have a fixed set of operator strategies
amsupport	int2		Number of support routines for this access method
amcanorder	bool		Does the access method support ordere d scans sorted by the indexed column's value?
amcanorderbyop	bool		Does the access method support ordere d scans sorted by the result of an operat or on the indexed column?
amcanbackward	bool		Does the access method support backward scanning?
amcanunique	bool		Does the access method support unique indexes?

Name	Type	References	Description
amcanmulticol	bool		Does the access method support multic olumn indexes?
amoptionalkey	bool		Does the access method support a scan without any constr aint for the first index column?
amsearchnulls	bool		Does the access method support IS NULL/NOT NULL search es?
amstorage	bool		Can index storage data type differ from column data type?
amclusterable	bool		Can an index of this type be clustered on?
ampredlocks	bool		Does an index of this type manage fine-grained predicate locks?
amkeytype	oid	pg_type.oid	Type of data stored in index, or zero if not a fixed type
aminsert	regproc	pg_proc.oid	“Insert this tuple” function
ambeginscan	regproc	pg_proc.oid	“Prepare for index scan” function
amgettuple	regproc	pg_proc.oid	“Next valid tuple” function, or zero if none
amgetbitmap	regproc	pg_proc.oid	“Fetch all valid tuples” function, or zero if none
amrescan	regproc	pg_proc.oid	“(Re)start index scan” function
amendscan	regproc	pg_proc.oid	“Clean up after index scan” function
ammarkpos	regproc	pg_proc.oid	“Mark current scan position” function
amrestrpos	regproc	pg_proc.oid	“Restore marked scan position” function
ambuild	regproc	pg_proc.oid	“Build new index” function
ambuildempty	regproc	pg_proc.oid	“Build empty index” function
ambulkdelete	regproc	pg_proc.oid	Bulk-delete function
amvacuumcleanup	regproc	pg_proc.oid	Post-VACUUM cleanup function

Name	Type	References	Description
amcostestimate	regproc	pg_proc.oid	Function to estimate cost of an index scan
amoptions	regproc	pg_proc.oid	Function to parse and validate reloptions for an index

45.4. pg_amop

目录pg_amop存储关于与访问方法操作符族相关的操作符信息。
对于操作符族中的每个成员操作符，此目录中都有一行。一个成员可以是搜索操作符，也可以是排序操作符。一个操作符可以出现在多个族中，但在同一个族中既不能出现在多个搜索位置，也不能出现在多个排序位置（虽然不太可能，但允许一个操作符同时用于搜索和排序）。

表 45.4. pg_amop 列

Name	Type	References	Description
amopfamily	oid	pg_opfamily.oid	The operator family this entry is for
amoplefttype	oid	pg_type.oid	Left-hand input data type of operator
amoprighttype	oid	pg_type.oid	Right-hand input data type of operator
amopstrategy	int2		Operator strategy number
amoppurpose	char		Operator purpose, either s for search or o for ordering
amopopr	oid	pg_operator.oid	OID of the operator
amopmethod	oid	pg_am.oid	Index access method operator family is for
amopsortfamily	oid	pg_opfamily.oid	The btree operator family this entry sorts according to, if an ordering operator; zero if a search operator

“搜索”操作符项表明可以搜索该操作符族的一个索引来找到所有满足 WHERE *indexed_column operator constant* 的行。显然，这样的操作符必须返回布尔值，并且其左输入类型必须匹配索引列的数据类型。

“排序”操作符项表明可以扫描该操作符族的一个索引，按照 ORDER BY *indexed_column operator constant* 所表示的顺序返回行。这样的操作符可以返回任何可排序的数据类型，但其左输入类型同样必须匹配索引列的数据类型。ORDER BY 的确切语义由 amopsortfamily 列指定，该列必须引用操作符结果类型的一个 btree 操作符族。

> **注意**
>
> 当前假定排序操作符的排序顺序是被引用操作符族的默认值，即 ASC NULLS LAST。将来可能会通过增加额外的列来显式指定排序选项，从而放宽这一假定。

一个项的amopmethod必须和其所属操作符族的opfmethod相匹配（此处加入amopmethod是出于性能原因而对目录结构故意做出的反规范化）。此外，amoplefttype和amoprighttype也必须匹配被引用的pg_operator项的oprleft和oprright域。

45.5. pg_amproc

目录pg_amproc存储关于访问方法操作符族相关的支持函数。
属于一个操作符族的每一个支持函数在这个目录中都有一行。

表 45.5. pg_amproc 列

Name	Type	References	Description
amprocfamily	oid	pg_opfamily.oid	The operator family this entry is for
amproclefttype	oid	pg_type.oid	Left-hand input data type of associated operator
amprocrighttype	oid	pg_type.oid	Right-hand input data type of associated operator
amprocnum	int2		Support procedure number
amproc	regproc	pg_proc.oid	OID of the procedure

amproclefttype 和 amprocrighttype 字段的通常解释是，它们标识某个特定支持过程所支持的操作符的左、右输入类型。For some access methods these match the input data type(s) of the support procedure itself, for others not. There is a notion of “default” support procedures for an index, which are those with amproclefttype and amprocrighttype both equal to the index opclass's opcintype.

45.6. pg_attrdef

目录pg_attrdef存储列默认值。列的主要信息存储在pg_attribute中（见下文）。
只有在创建表或添加列时显式指定默认值的列，才会在这里有一项。

表 45.6. pg_attrdef 列

Name	Type	References	Description
adrelid	oid	pg_class.oid	The table this column belongs to
adnum	int2	pg_attribute.attnum	The number of the column
adbin	pg_node_tree		The internal representation of the column default value
adsrc	text		A human-readable representation of the default value

adsrc字段是历史遗留，最好不要使用，因为它不会跟踪可能影响默认值表示的外部变化。
显示默认值时，更好的方法是反向编译adbin字段（例如使用pg_get_expr）。

45.7. pg_attribute

目录pg_attribute存储关于表列的信息。数据库中每个表的每一列，在pg_attribute中都恰好有一行。（其中也包括索引的属性项，实际上，凡是在pg_class中有项的对象，都有对应的属性项。）

术语属性等同于列，这里使用它只是出于历史原因。

表 45.7. pg_attribute 列

Name	Type	References	Description
attrelid	oid	pg_class.oid	The table this column belongs to
attname	name		The column name
atttypid	oid	pg_type.oid	The data type of this column
attstattarget	int4		attstattarget controls the level of detail of statistics accumulated for this column by ANALYZE. A zero value indicates that no statistics should be collected. A negative value says to use the system default statistics target. The exact meaning of positive values is data type-dependent. For scalar data types, attstattarget is both the target number of "most common values" to collect, and the target number of histogram bins to create.
attlen	int2		A copy of pg_type.typlen of this column's type
attnum	int2		The number of the column. Ordinary columns are numbered from 1 up. System columns, such as oid, have (arbitrary) negative numbers.
attndims	int4		Number of dimensions, if the column is an array type; otherwise 0. (Presently, the number of dimensions of an array is not enforced, so any

Name	Type	References	Description
			nonzero value effect ively means “it's an array” .)
attcacheoff	int4		Always -1 in storage, but when loaded into a row descriptor in memory this might be updated to cache the offset of the attribute within the row
atttypmod	int4		atttypmod record s type-specific data supplied at table creati on time (for example, the maximum length of a varchar column) . It is passed to type-specific input functions and length coercion functions. The value will generally be -1 for types that do not need atttypmod.
attbyval	bool		A copy of pg_type. typbyval of this column's type
attstorage	char		Normally a copy of pg_ type.typstorage of this column's type. For TOAST-able data types, this can be altered after column creati on to control storage policy.
attalign	char		A copy of pg_type. typalign of this column's type
attnotnull	bool		This represents a not-null constraint. It is possible to change this column to enable or disable the constraint.
atthasdef	bool		This column has a default value, in which case there will be a corresponding entry in the pg_attrdef catalo g that actually defines the value.
attisdropped	bool		This column has been dropped and is no longer valid. A droppe d column is still physic ally present in the

Name	Type	References	Description
			table, but is ignored by the parser and so cannot be accessed via SQL.
attislocal	bool		This column is define d locally in the relati on. Note that a column can be locally define d and inherited simult aneously.
attinhcount	int4		The number of direct ancestors this column has. A column with a nonzero number of ancestors cannot be dropped nor renamed.
attcollation	oid	pg_collation.oid	The defined collation of the column, or zero if the column is not of a collatable data type.
attacl	aclitem[]		Column-level access privileges, if any have been granted specif ically on this column
attoptions	text[]		Attribute-level options, as "keyword=value" strings

在被删除列的 pg_attribute 条目中，atttypid 被重置为零，但 attlen 以及其他从 pg_type 复制的字段仍然有效。这种安排用于应对被删除列的数据类型后来也被删除、因而不再有相应 pg_type 行的情况。attlen 和其他字段可用于解释表中一行的内容。

45.8. pg_authid

目录pg_authid包含关于数据库授权标识符（角色）的信息。角色涵盖了"用户"和"组"这两个概念。用户本质上只是设置了rolcanlogin标志的角色。
任何角色（无论是否设置了rolcanlogin）都可以拥有其他角色作为成员；参见pg_auth_members。

由于这个目录包含密码，它不应公开可读。pg_roles是建立在pg_authid之上的公开可读视图，它会隐藏密码字段。

第 20 章包含关于用户和权限管理的详细信息。

由于用户标识符是集簇范围的，pg_authid
在一个集簇的所有数据库之间共享：在一个集簇中只有一份pg_authid拷贝，而不是每个数据库一份。

表 45.8. pg_authid 列

Name	Type	Description
rolname	name	Role name

Name	Type	Description
rolsuper	bool	Role has superuser privileges
rolinherit	bool	Role automatically inheri ts privileges of roles it is a member of
rolcreaterole	bool	Role can create more roles
rolcreatedb	bool	Role can create databases
rolcatupdate	bool	Role can update system catalo gs directly. (Even a superu ser cannot do this unless this column is true)
rolcanlogin	bool	Role can log in. That is, this role can be given as the initial session authorization identifier
rolreplication	bool	Role is a replication role. That is, this role can initiate stream ing replication (see 第 25.2.5 节) and set/unset the system backup mode using pg_start_backup and pg_stop_backup
rolconnlimit	int4	For roles that can log in, this sets maximum number of concurrent connections this role can make. -1 means no limit.
rolpassword	text	Password (possibly encrypted) ; null if none. If the passwo rd is encrypted, this column will begin with the string md5 followed by a 32-character hexadecimal MD5 hash. The MD5 hash will be of the user's password concatenated to their user name. For exampl e, if user joe has password xyzzy, PostgreSQL will store the md5 hash of xyzzyjoe. A password that does not follow that format is assumed to be unencrypted.
rolvaliduntil	timestamptz	Password expiry time (only used for password authenticati on); null if no expiration

45.9. pg_auth_members

目录pg_auth_members展示了角色之间的成员关系。允许任何无环的关系集合。

由于用户标识符是集簇范围的，pg_auth_members
在一个集簇的所有数据库之间共享：在一个集簇中只有一份pg_auth_members拷贝，
而不是每个数据库一份。

表 45.9. pg_auth_members 列

Name	Type	References	Description
roleid	oid	pg_authid.oid	ID of a role that has a member
member	oid	pg_authid.oid	ID of a role that is a member of roleid
grantor	oid	pg_authid.oid	ID of the role that granted this member ship
admin_option	bool		True if member can grant membership in roleid to others

45.10. pg_cast

目录 pg_cast 存储数据类型转换路径，包括内建的路径和用CREATE CAST定义的路径。

需要注意的是，pg_cast并不表示系统知道如何执行的所有类型转换，
它只包括那些不能从某些通用规则推导出的转换。例如，
一个域及其基础类型之间的转换并未显式地在pg_cast中展示。
另一个重要的例外是“自动的基于 I/O 的类型转换”，它们通过数据类型自己的 I/O 函数来转换成（或者转换自）text或其他字符串类型，这些转换也没有显式地在pg_cast中表示。

表 45.10. pg_cast 列

Name	Type	References	Description
castsource	oid	pg_type.oid	OID of the source data type
casttarget	oid	pg_type.oid	OID of the target data type
castfunc	oid	pg_proc.oid	The OID of the functi on to use to perform this cast. Zero is stored if the cast method doesn't require a functi on.
castcontext	char		Indicates what contex ts the cast can be invoked in. e means only as an explicit cast (using CAST or :: syntax). a means implicitly in assign ment to a target column, as well as explicitly. i means implicitly in expres sions, as well as the other cases.
castmethod	char		Indicates how the cast is performed. f means that the function specif ied in the castfunc

Name	Type	References	Description
			field is used. i means that the input/output functions are used. b means that the types are binary-coercible, thus no conversion is required.

在pg_cast里列出的类型转换函数必须总是以转换的源类型作为它的第一个参数类型，并且返回转换的目标类型作为它的结果类型。一个类型转换函数最多有三个参数。如果出现了第二个参数，必须是integer类型；它接受与目标类型关联的修饰词，如果没有，就是 -1。如果出现了第三个参数，那么必须是boolean类型；如果该类型转换是一种明确的转换，那么它接受true，否则接受false。

在pg_cast里创建一条源类型和目标类型相同的记录是合理的，只要相关联的函数接受多过一个参数。这样的记录代表“长度转换函数”，它们把该类型的值转换为对特定的类型合法的值。

如果一个pg_cast的项有着不同的原类型和目标类型，并且有一个接收多于一个参数的函数，那么它会在一个步骤中完成从一种类型到另外一种类型的转换并应用一个长度转换。如果没有这样的项，使用一个类型修改器的转换涉及两个步骤，一个是在数据类型之间转换，另外一个是应用修改器。

45.11. pg_class

目录 pg_class 记录表和大多数其他具有列或者类似于表的对象。这包括索引（但另见 pg_index）、序列、视图、组合类型和 TOAST 表；见 relkind。Below, when we mean all of these kinds of objects we speak of “relations”. Not all columns are meaningful for all relation types.

表 45.11. pg_class 列

Name	Type	References	Description
relname	name		Name of the table, index, view, etc.
relnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this relation
reltype	oid	pg_type.oid	The OID of the data type that corresponds to this table's row type, if any (zero for indexes, which have no pg_type entry)
reloftype	oid	pg_type.oid	For typed tables, the OID of the underlying composite type, zero for all other relations
relowner	oid	pg_authid.oid	Owner of the relation
relam	oid	pg_am.oid	If this is an index, the access method used (B-tree, hash, etc.)

Name	Type	References	Description
relfilenode	oid		Name of the on-disk file of this relation; zero means this is a “mapped” relation whose disk file name is determined by low-level state
reltablespace	oid	pg_tablespace.oid	The tablespace in which this relation is stored. If zero, the database's default tablespace is implied. (Not meaningful if the relation has no on-disk file.)
relpages	int4		Size of the on-disk representation of this table in pages (of size BLCKSZ). This is only an estimate used by the planner. It is updated by VACUUM, ANALYZE, and a few DDL commands such as CREATE INDEX.
reltuples	float4		Number of rows in the table. This is only an estimate used by the planner. It is update d by VACUUM, ANALYZ E, and a few DDL commands such as CREATE INDEX.
reltoastrelid	oid	pg_class.oid	OID of the TOAST table associated with this table, 0 if none. The TOAST table stores large attributes “out of line” in a secondary table.
reltoastidxid	oid	pg_class.oid	For a TOAST table, the OID of its index. 0 if not a TOAST table.
relhasindex	bool		True if this is a table and it has (or recently had) any indexes
relisshared	bool		True if this table is shared across all databases in the cluste r. Only certain system catalogs (such as pg_database) are shared.

Name	Type	References	Description
relpersistence	char		p = permanent table, u = unlogged table, t = temporary table
relkind	char		r = ordinary table, i = index, S = sequence, v = view, c = composite type, t = TOAST table, f = foreign table
relnatts	int2		Number of user columns in the relation (system columns not counted). There must be this many corres ponding entries in pg_attribute. See also pg_attribute.attnum.
relchecks	int2		Number of CHECK constraints on the table; see pg_constraint catalog
relhasoids	bool		True if we generate an OID for each row of the relation
relhaspkey	bool		True if the table has (or once had) a primary key
relhasrules	bool		True if table has (or once had) rules; see pg_rewrite catalog
relhastriggers	bool		True if table has (or once had) triggers; see pg_trigger catalo g
relhassubclass	bool		True if table has (or once had) any inheri tance children
relfrozenxid	xid		All transaction IDs before this one have been replaced with a permanent (“frozen”) transaction ID in this table. This is used to track whether the table needs to be vacuum ed in order to prevent transaction ID wrapar ound or to allow pg_clog to be shrunk. Zero (InvalidTransactionId) if the relation is not a table.

Name	Type	References	Description
relacl	aclitem[]		Access privileges; see GRANT and REVOKE for details
reloptions	text[]		Access-method-specific options, as "keyword=value" strings

pg_class中的一些布尔标志采用延迟维护：当条件成立时，保证它们为真；但当条件不再成立时，可能不会立即将它们重置为假。例如，relhasindex由CREATE INDEX设置，但它从不会被DROP INDEX清除。作为替代，VACUUM会在找到无索引表后清除其relhasindex。这种安排避免了竞争条件并且提高了并发性。

45.12. pg_constraint

目录pg_constraint存储表上的检查约束、主键约束、唯一约束、外键约束和排他约束。（列约束不会被特殊对待。每一个列约束都等价于某种表约束。）非空约束在pg_attribute目录中表示，而不是在这里。

用户定义的约束触发器（使用CREATE CONSTRAINT TRIGGER创建）也会在这个表中产生一项。

域上的检查约束也存储在这里。

表 45.12. pg_constraint 列

Name	Type	References	Description
conname	name		Constraint name (not necessarily unique!)
connamespace	oid	pg_namespace.oid	The OID of the namespace that contains this constraint
contype	char		c = check constraint, f = foreign key constraint, p = primary key constraint, u = unique constraint, t = constraint trigger, x = exclusion constraint
condeferrable	bool		Is the constraint deferrable?
condeferred	bool		Is the constraint deferred by default?
convalidated	bool		Has the constraint been validated? Currently, can only be false for foreign keys
conrelid	oid	pg_class.oid	The table this constraint is on; 0 if not a table constraint
contypid	oid	pg_type.oid	The domain this constraint is on; 0 if not a domain constraint

Name	Type	References	Description
conindid	oid	pg_class.oid	The index supporting this constraint, if it's a unique, primary key, foreign key, or exclusion constraint; else 0
confrelid	oid	pg_class.oid	If a foreign key, the referenced table; else 0
confupdtype	char		Foreign key update action code: a = no action, r = restrict, c = cascade, n = set null, d = set default
confdeltype	char		Foreign key deletion action code: a = no action, r = restrict, c = cascade, n = set null, d = set default
confmatchtype	char		Foreign key match type: f = full, p = partial, u = simple (unspecified)
conislocal	bool		This constraint is defined locally for the relation. Note that a constraint can be locally defined and inherited simultaneously.
coninhcount	int4		The number of direct inheritance ancestors this constraint has. A constraint with a nonzero number of ancestors cannot be dropped nor renamed.
conkey	int2[]	pg_attribute.attnum	If a table constraint (including foreign keys, but not constraint triggers), list of the constrained columns
confkey	int2[]	pg_attribute.attnum	If a foreign key, list of the referenced columns
conpfeqop	oid[]	pg_operator.oid	If a foreign key, list of the equality operators for PK = FK comparisons
conppeqop	oid[]	pg_operator.oid	If a foreign key, list of the equality operators for PK = PK comparisons
conffeqop	oid[]	pg_operator.oid	If a foreign key, list of the equality operat

Name	Type	References	Description
			ors for FK = FK compar isons
conexclop	oid[]	pg_operator.oid	If an exclusion constr aint, list of the per-column exclusion operators
conbin	pg_node_tree		If a check constraint, an internal repres entation of the expres sion
consrc	text		If a check constraint, a human-readable representation of the expression

对于排他约束，conkey仅对作为简单列引用的约束元素有用。对于其他情况，conkey中会出现一个 0，必须查阅关联索引来确定被约束的表达式。（因此，conkey与该索引的 pg_index.indkey具有相同的内容。）

注意

被引用对象发生变化时，consrc不会更新；例如，它不会跟踪列重命名。最好使用pg_get_constraintdef()提取检查约束的定义，而不是依赖这个字段。

注意

pg_class.relchecks需要和每个关系在此目录中的检查约束数量保持一致。

45.13. pg_collation

目录pg_collation描述可用的排序规则，其本质上是从 SQL 名字到操作系统区域设置类别的映射。更多信息参见第 22.2 节。

表 45.13. pg_collation 列

Name	Type	References	Description
collname	name		Collation name (unique per namespace and encoding)
collnamespace	oid	pg_namespace.oid	The OID of the namesp ace that contains this collation
collowner	oid	pg_authid.oid	Owner of the collation
collencoding	int4		Encoding in which the collation is applicable, or -1 if it works for any encoding

Name	Type	References	Description
collcollate	name		LC_COLLATE for this collation object
collctype	name		LC_CTYPE for this collation object

注意在这个目录中的唯一键是（collname、collencoding、collnamespace），不仅仅是（collname、collnamespace）。PostgreSQL 通常会忽略所有 collencoding 既不等于当前数据库编码、也不等于 -1 的排序规则，并且禁止创建与 collencoding = -1 的项同名的新项。因此，使用限定的 SQL 名字（*schema.name*）来标识一个排序规则已经足够，即使按目录定义它并不唯一。之所以将该目录定义成这样，是因为 initdb 在集簇初始化时会用系统上所有可用的区域设置填充它，因此它必须能够容纳集簇中未来可能使用到的所有编码的项。

在template0数据库中，创建与数据库编码不匹配的排序规则可能很有用，因为它们可以匹配之后从template0克隆出的数据库的编码。目前这必须手工完成。

45.14. pg_conversion

目录pg_conversion描述编码转换函数。更多信息参见CREATE CONVERSION。

表 45.14. pg_conversion 列

Name	Type	References	Description
conname	name		Conversion name (unique within a namespace)
connamespace	oid	pg_namespace.oid	The OID of the namespace that contains this conversion
conowner	oid	pg_authid.oid	Owner of the conversion
conforencoding	int4		Source encoding ID
contoencoding	int4		Destination encoding ID
conproc	regproc	pg_proc.oid	Conversion procedure
condefault	bool		True if this is the default conversion

45.15. pg_database

目录pg_database存储有关可用数据库的信息。数据库通过CREATE DATABASE命令创建。更多关于其参数的信息请查阅第 21 章。

和大部分系统目录不同，pg_database是在集簇的所有数据库之间共享的：在一个集簇中只有一份pg_database拷贝，而不是每个数据库一份。

表 45.15. pg_database 列

Name	Type	References	Description
datname	name		Database name

Name	Type	References	Description
datdba	oid	pg_authid.oid	Owner of the database, usually the user who created it
encoding	int4		Character encoding for this database (pg_encoding_to_char() can translate this number to the encoding name)
datcollate	name		LC_COLLATE for this database
datctype	name		LC_CTYPE for this database
datistemplate	bool		If true then this database can be used in the TEMPLATE clause of CREATE DATABASE to create a new database as a clone of this one
datallowconn	bool		If false then no one can connect to this database. This is used to protect the template0 database from being altered.
datconnlimit	int4		Sets maximum number of concurrent connections that can be made to this database. -1 means no limit.
datlastsysoid	oid		Last system OID in the database; useful particularly to pg_dump
datfrozenxid	xid		All transaction IDs before this one have been replaced with a permanent (“frozen”) transaction ID in this database. This is used to track whether the database needs to be vacuumed in order to prevent transaction ID wraparound or to allow pg_clog to be shrunk. It is the minimum of the per-table pg_class.relfrozenxid values.
dattablespace	oid	pg_tablespace.oid	The default tablespace for the database. Within this database, all tables for which pg_class.reltablespac

Name	Type	References	Description
			e is zero will be stored in this tablespace; in particular, all the non-shared system catalogs will be there.
datacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

45.16. pg_db_role_setting

目录 pg_db_role_setting为每个角色与数据库组合记录运行时配置变量的默认值。

和大部分系统目录不同，pg_db_role_setting
是在集簇的所有数据库之间共享的：在一个集簇中只有一份pg_db_role_setting拷贝，
而不是每个数据库一份。

表 45.16. pg_db_role_setting 列

Name	Type	References	Description
setdatabase	oid	pg_database.oid	The OID of the database the setting is applicable to, or zero if not database-specific
setrole	oid	pg_authid.oid	The OID of the role the setting is applicable to, or zero if not role-specific
setconfig	text[]		Defaults for run-time configuration variables

45.17. pg_default_acl

目录pg_default_acl存储要被分配给新创建对象的初始权限。

表 45.17. pg_default_acl 列

Name	Type	References	Description
defaclrole	oid	pg_authid.oid	The OID of the role associated with this entry
defaclnamespace	oid	pg_namespace.oid	The OID of the namespace associated with this entry, or 0 if none
defaclobjtype	char		Type of object this entry is for: r = relation (table, view), S = sequence, f = function
defaclacl	aclitem[]		Access privileges that this type of object

Name	Type	References	Description
			should have on creati on

一个pg_default_acl项表示要赋给属于指定用户的对象的初始权限。
目前有两类项：defaclnamespace = 0 的“全局”项，以及引用特定模式的“按模式”项。如果存在全局项，则它会覆盖该对象类型通常写死的默认权限。如果存在按模式项，
则表示要将这些权限附加到全局默认权限或写死的默认权限之上。

注意，当另一个目录中的 ACL 项为 NULL 时，它表示该对象采用写死的默认权限，而不是当前 pg_default_acl 中可能存在的内容。pg_default_acl 只会在对象创建期间被查阅。

45.18. pg_depend

目录pg_depend记录数据库对象之间的依赖关系。这些信息允许DROP命令查找必须被DROP CASCADE删除的其他对象，或者在DROP RESTRICT情况下阻止删除。

另请参阅pg_shdepend，它对在一个数据库集簇中共享的对象之间的依赖提供了相似的功能。

表 45.18. pg_depend 列

Name	Type	References	Description
classid	oid	pg_class.oid	The OID of the system catalog the dependent object is in
objid	oid	any OID column	The OID of the specific dependent object
objsubid	int4		For a table column, this is the column number (the objid and classid refer to the table itself). For all other object types, this column is zero.
refclassid	oid	pg_class.oid	The OID of the system catalog the referenced object is in
refobjid	oid	any OID column	The OID of the specific referenced object
refobjsubid	int4		For a table column, this is the column number (the refobjid and refclassid refer to the table itself). For all other object types, this column is zero.
deptype	char		A code defining the specific semantics of this dependency relati onship; see text

在所有情况下，一条pg_depend记录都表示：如果不同时删除依赖对象，
就不能删除被引用对象。不过，还存在若干由deptype标识的子类型：

DEPENDENCY_NORMAL (n)

两个独立创建的对象之间的正常关系。依赖对象可以被删除而不影响被引用对象。
被引用对象只能通过指定CASCADE来删除，这样依赖对象也会被删除。
例如：表列对其数据类型有一个正常依赖关系。

DEPENDENCY_AUTO (a)

依赖对象可以与被引用对象分开删除，并且在被引用对象删除时应自动删除（无论使用RESTRICT还是CASCADE模式）。例如，表上的命名约束会自动依赖于该表，
因此删除表时该约束也会消失。

DEPENDENCY_INTERNAL (i)

依赖对象作为被引用对象创建过程的一部分而创建，实际上只是其内部实现的一部分。
对依赖对象执行DROP会被直接禁止（我们会告诉用户改为对被引用对象执行DROP）。
对被引用对象执行DROP会传播到依赖对象，使其被删除，无论是否指定CASCADE。例如，
为实施外键约束而创建的触发器会在内部依赖于该约束的pg_constraint项。

DEPENDENCY_EXTENSION (e)

依赖对象是作为被引用对象的扩展的一个成员（参见pg_extension）。
只能通过对被引用对象执行DROP EXTENSION来删除依赖对象。在功能上，这种依赖类型与内部依赖相同，但为了清晰和简化pg_dump而单独区分。

DEPENDENCY_PIN (p)

不存在依赖对象；这种项表示系统本身依赖于被引用对象，因此该对象绝不能删除。
这种项只由initdb创建。用于表示依赖对象的各列包含零。

未来可能需要其他依赖关系的变种。

45.19. pg_description

目录pg_description存储对每一个数据库对象可选的描述（注释）。
描述可以通过COMMENT操作，并可使用psql的\d命令查看。在pg_description
的初始内容中提供了很多内置系统对象的描述。

参见pg_shdescription，它对在一个数据库集簇中共享的对象的描述提供了相似的功能。

表 45.19. pg_description 列

Name	Type	References	Description
objoid	oid	any OID column	The OID of the object this description pertai ns to
classoid	oid	pg_class.oid	The OID of the system catalog this object appears in
objsubid	int4		For a comment on a table column, this is the column number (the objoid and classoid refer to the table itself). For all

Name	Type	References	Description
			other object types, this column is zero.
description	text		Arbitrary text that serves as the descri ption of this object

45.20. pg_enum

pg_enum目录包含每一个枚举类型的项，其中包括了值和标签。
一个给定枚举值的内部表示实际上是它在pg_enum中的相关行的OID。

表 45.20. pg_enum 列

Name	Type	References	Description
enumtypid	oid	pg_type.oid	The OID of the pg_type entry owning this enum value
enumsortorder	float4		The sort position of this enum value within its enum type
enumlabel	name		The textual label for this enum value

pg_enum
行的OID值遵循一种特殊的规则：即OID的数值被保证按照其枚举类型的排序顺序进行排序。
即如果两个偶数OID属于同一枚举类型，较小的OID必然具有较小的enumsortorder值。
奇数OID值不需要遵循排序顺序。
这种规则使得枚举比较例程在很多常见情况下可以避免系统目录查找。
创建和修改枚举类型的例程将尝试尽可能地为枚举值分配偶数OID。

当一个枚举类型被创建后，其成员会被分配排序位置 1..*n*。但后来增加的成员可能会被赋予负值或分数值的enumsortorder。对这些值的唯一要求是它们必须排序正确并且保持唯一。

45.21. pg_extension

目录pg_extension存储有关已安装扩展的信息。有关扩展的细节请参见第 35.15 节。

表 45.21. pg_extension 列

Name	Type	References	Description
extname	name		Name of the extension
extowner	oid	pg_authid.oid	Owner of the extension
extnamespace	oid	pg_namespace.oid	Schema containing the extension's exported objects
extrelocatable	bool		True if extension can be relocated to anothe r schema
extversion	text		Version name for the extension

Name	Type	References	Description
extconfig	oid[]	pg_class.oid	Array of regclass OIDs for the extension's configuration table(s), or NULL if none
extcondition	text[]		Array of WHERE-clause filter conditions for the extension's configuration table(s), or NULL if none

注意和大部分具有一个 “namespace” 列的模式不同，extnamespace 不是用来表示扩展属于该模式。扩展的名字从不用模式进行限定。extnamespace 表明该模式包含了该扩展的大部分或全部对象。如果extrelocatable为真，则该模式事实上必须包含属于此扩展的全部模式限定的对象。

45.22. pg_foreign_data_wrapper

目录pg_foreign_data_wrapper存储外部数据包装器定义。外部数据包装器是一种访问位于外部服务器上数据的机制。

表 45.22. pg_foreign_data_wrapper 列

Name	Type	References	Description
fdwname	name		Name of the foreign-data wrapper
fdwowner	oid	pg_authid.oid	Owner of the foreign-data wrapper
fdwhandler	oid	pg_proc.oid	References a handler function that is responsible for supplying execution routines for the foreign-data wrapper. Zero if no handler is provided
fdwvalidator	oid	pg_proc.oid	References a validator function that is responsible for checking the validity of the options given to the foreign-data wrapper, as well as options for foreign servers and user mappings using the foreign-data wrapper. Zero if no validator is provided
fdwacl	aclitem[]		Access privileges; see GRANT and REVOKE for details
fdwoptions	text[]		Foreign-data wrapper specific options, as

Name	Type	References	Description
			“keyword=value” strings

45.23. pg_foreign_server

目录pg_foreign_server存储外部服务器定义。外部服务器定义了外部数据的来源，例如一个远程服务器。外部服务器通过外部数据包装器来访问。

表 45.23. pg_foreign_server 列

Name	Type	References	Description
srvname	name		Name of the foreign server
srvowner	oid	pg_authid.oid	Owner of the foreign server
srvfdw	oid	pg_foreign_data_wrapper.oid	OID of the foreign-data wrapper of this foreign server
srvtype	text		Type of the server (optional)
srvversion	text		Version of the server (optional)
srvacl	aclitem[]		Access privileges; see GRANT and REVOKE for details
srvoptions	text[]		Foreign server specific options, as “keyword=value” strings

45.24. pg_foreign_table

目录pg_foreign_table包含关于外部表的辅助信息。一个外部表和普通表一样，主要由一个pg_class项表示。它的pg_foreign_table项包含外部表所特有的信息。

表 45.24. pg_foreign_table 列

Name	Type	References	Description
ftrelid	oid	pg_class.oid	OID of the pg_class entry for this foreign table
ftserver	oid	pg_foreign_server.oid	OID of the foreign server for this foreign table
ftoptions	text[]		Foreign table options, as “keyword=value” strings

45.25. pg_index

目录pg_index包含关于索引的部分信息。其他信息大部分在pg_class中。

表 45.25. pg_index 列

Name	Type	References	Description
indexrelid	oid	pg_class.oid	The OID of the pg_class entry for this index
indrelid	oid	pg_class.oid	The OID of the pg_class entry for the table this index is for
indnatts	int2		The number of columns in the index (duplicates pg_class.relnatts)
indisunique	bool		If true, this is a unique index
indisprimary	bool		If true, this index represents the primary key of the table (indisunique should always be true when this is true)
indisexclusion	bool		If true, this index supports an exclusion constraint
indimmediate	bool		If true, the uniqueness check is enforced immediately on insertion (irrelevant if indisunique is not true)
indisclustered	bool		If true, the table was last clustered on this index
indisvalid	bool		If true, the index is currently valid for queries. False means the index is possibly incomplete: it must still be modified by INSERT/UPDATE operations, but it cannot safely be used for queries. If it is unique, the uniqueness property is not true either.
indcheckxmin	bool		If true, queries must not use the index until the xmin of this pg_index row is below their TransactionXmin event horizon, because the table may contain broken HOT chains with incomp

Name	Type	References	Description
			atible rows that they can see
indisready	bool		If true, the index is currently ready for inserts. False means the index must be ignored by INSERT/UPDATE operations.
indkey	int2vector	pg_attribute.attnum	This is an array of indnatts values that indicate which table columns this index indexes. For example a value of 1 3 would mean that the first and the third table columns make up the index key. A zero in this array indicates that the corresponding index attribute is an expres sion over the table columns, rather than a simple column refere nce.
indcollation	oidvector	pg_collation.oid	For each column in the index key, this contains the OID of the collation to use for the index.
indclass	oidvector	pg_opclass.oid	For each column in the index key, this contains the OID of the operat or class to use. See pg_opclass for details.
indoption	int2vector		This is an array of indnatts values that store per-column flag bits. The meaning of the bits is defined by the index's access method.
indexprs	pg_node_tree		Expression trees (in nodeToString() representation) for index attributes that are not simple column references. This is a list with one elemen t for each zero entry in indkey. Null if all index attributes are simple references.
indpred	pg_node_tree		Expression tree (in nodeToString() representation) for

Name	Type	References	Description
			partial index predic ate. Null if not a partia l index.

45.26. pg_inherits

目录pg_inherits记录有关表继承层次的信息。数据库中每个直接子表在这里都有一项。（非直接继承可以通过顺着项构成的链来确定。）

表 45.26. pg_inherits 列

Name	Type	References	Description
inhrelid	oid	pg_class.oid	The OID of the child table
inhparent	oid	pg_class.oid	The OID of the parent table
inhseqno	int4		If there is more than one direct parent for a child table (multip le inheritance), this number tells the order in which the inheri ted columns are to be arranged. The count starts at 1.

45.27. pg_language

目录pg_language注册了可用于编写函数或存储过程的语言。
更多关于语言处理器的信息请参阅CREATE LANGUAGE和第 38 章。

表 45.27. pg_language 列

Name	Type	References	Description
lanname	name		Name of the language
lanowner	oid	pg_authid.oid	Owner of the language
lanispl	bool		This is false for intern al languages (such as SQL) and true for user-defined langua ges. Currently, pg_ dump still uses this to determine which languages need to be dumped, but this might be replaced by a differ ent mechanism in the future.
lanpltrusted	bool		True if this is a truste d language, which means that it is believ ed not to grant access

Name	Type	References	Description
			to anything outside the normal SQL execut ion environment. Only superusers can create functions in untrusted languages.
lanplcallfoid	oid	pg_proc.oid	For noninternal langua ges this references the language handle r, which is a special function that is respon sible for executing all functions that are written in the partic ular language
laninline	oid	pg_proc.oid	This references a functi on that is responsible for executing “inline” anonymous code blocks (DO blocks). Zero if inline blocks are not supported.
lanvalidator	oid	pg_proc.oid	This references a language validator function that is respon sible for checking the syntax and validity of new functions when they are created. Zero if no validator is provid ed.
lanacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

45.28. pg_largeobject

目录pg_largeobject保存构成“大对象”的数据。一个大对象在被创建时会被分配一个OID。每个大对象被分解成足够小的段或“页”，以便方便地作为行存储在pg_largeobject中。每页中的数据量被定义为LOBLKSIZE（目前是BLCKSZ/4或是2 kB）。

在PostgreSQL 9.0 之前，大对象没有相关的权限结构。因此，pg_largeobject曾是公开可读的，并且可以用来获得系统中所有大对象的 OID（和内容）。但现在已经不是这样了；可使用pg_largeobject_metadata来获得大对象 OID 的列表。

表 45.28. pg_largeobject 列

Name	Type	References	Description
loid	oid	pg_largeobject_metadata.oid	Identifier of the large object that includes this page
pageno	int4		Page number of this page within its large

Name	Type	References	Description
			object (counting from zero)
data	bytea		Actual data stored in the large object. This will never be more than LOBLKSIZE bytes and might be less.

pg_largeobject的每一行保存一个大对象的一个页的数据，从对象内部的字节偏移量（pageno * LOBLKSIZE）开始。现在的实现允许稀疏存储：页面可能丢失，并且可能比LOBLKSIZE字节短（即便不是最后一页）。一个大对象中丢失的区域会被读出为0。

45.29. pg_largeobject_metadata

目录pg_largeobject_metadata保持着与大对象有关的元数据。
真正的大对象数据被存储在pg_largeobject中。

表 45.29. pg_largeobject_metadata 列

Name	Type	References	Description
lomowner	oid	pg_authid.oid	Owner of the large object
lomacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

45.30. pg_namespace

目录pg_namespace存储名字空间。
名字空间是SQL模式之下的结构：每个名字空间拥有一个独立的表、类型等的集合，
且其中没有名字冲突。

表 45.30. pg_namespace 列

Name	Type	References	Description
nspname	name		Name of the namespace
nspowner	oid	pg_authid.oid	Owner of the namespace
nspacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

45.31. pg_opclass

目录 pg_opclass 定义索引访问方法操作符类。每个操作符类为一种特定数据类型和一种特定索引访问方法的索引列定义语义。
操作符类本质上指定一个特定的操作符族适用于一种特定的可索引列数据类型。

该操作符族中实际可用于被索引列的操作符集合，
是那些接受该列数据类型作为其左输入的操作符。

操作符类详见第 35.14 节。

表 45.31. pg_opclass 列

Name	Type	References	Description
opcmethod	oid	pg_am.oid	Index access method operator class is for
opcname	name		Name of this operator class
opcnamespace	oid	pg_namespace.oid	Namespace of this operator class
opcowner	oid	pg_authid.oid	Owner of the operator class
opcfamily	oid	pg_opfamily.oid	Operator family containing the operator class
opcintype	oid	pg_type.oid	Data type that the operator class indexes
opcdefault	bool		True if this operator class is the default for opcintype
opckeytype	oid	pg_type.oid	Type of data stored in index, or zero if same as opcintype

一个操作符类的opcmethod必须匹配包含它的操作符族的opfmethod。而且，对于任何给定的opcmethod和opcintype组合，只有不超过一个pg_opclass行的opcdefault值为真。

45.32. pg_operator

目录pg_operator存储关于操作符的信息。详见CREATE OPERATOR和第 35.12 节。

表 45.32. pg_operator 列

Name	Type	References	Description
oprname	name		Name of the operator
oprnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this operator
oprowner	oid	pg_authid.oid	Owner of the operator
oprkind	char		b = infix (“both”), l = prefix (“left”), r = postfix (“right”)
oprcanmerge	bool		This operator supports merge joins
oprcanhash	bool		This operator supports hash joins
oprleft	oid	pg_type.oid	Type of the left operand

Name	Type	References	Description
oprright	oid	pg_type.oid	Type of the right operand
oprresult	oid	pg_type.oid	Type of the result
oprcom	oid	pg_operator.oid	Commutator of this operator, if any
oprnegate	oid	pg_operator.oid	Negator of this operat or, if any
oprcode	regproc	pg_proc.oid	Function that implem ents this operator
oprrest	regproc	pg_proc.oid	Restriction selectivity estimation function for this operator
oprjoin	regproc	pg_proc.oid	Join selectivity estima tion function for this operator

未使用的列包含零。例如，前缀操作符的oprleft为零。

45.33. pg_opfamily

目录pg_opfamily定义了操作符族。每一个操作符族是操作符和相关支持例程的集合，支持例程用于实现一个特定索引访问方法的语义。此外，按照访问方法指定的某种方式，一个族内的操作符都是“兼容的”。操作符族概念允许在索引中使用跨数据类型操作符，并可以使用访问方法语义的知识推导出。

操作符族最终在第 35.14 节中描述。

表 45.33. pg_opfamily 列

Name	Type	References	Description
opfmethod	oid	pg_am.oid	Index access method operator family is for
opfname	name		Name of this operator family
opfnamespace	oid	pg_namespace.oid	Namespace of this operator family
opfowner	oid	pg_authid.oid	Owner of the operator family

定义操作符族的主要信息不在它的pg_opfamily行，而是在相关的pg_amop、pg_amproc和pg_opclass行中。

45.34. pg_pltemplate

目录pg_pltemplate存储过程语言的“模板”信息。有了语言模板，就可以用简单的CREATE LANGUAGE命令在特定数据库中创建该语言，无需指定实现细节。

与大多数系统目录不同，pg_pltemplate由一个集簇中的所有数据库共享：每个集簇只有一份pg_pltemplate，而不是每个数据库各有一份。这样，每个数据库都可以在需要时访问这些信息。

表 45.34. pg_pltemplate列

Name	Type	Description
tmplname	name	Name of the language this template is for
tmpltrusted	boolean	True if language is considered trusted
tmpldbacreate	boolean	True if language may be create d by a database owner
tmplhandler	text	Name of call handler function
tmplinline	text	Name of anonymous-block handler function, or null if none
tmplvalidator	text	Name of validator function, or null if none
tmpllibrary	text	Path of shared library that implements language
tmplacl	aclitem[]	Access privileges for template (not actually used)

目前没有用于操纵过程语言模板的命令；要更改内置信息，超级用户必须使用普通的INSERT、DELETE或UPDATE命令修改此表。

> **注意**
>
> pg_pltemplate很可能会在PostgreSQL的某个未来版本中移除，
> 改为将这些过程语言的信息保存在各自的扩展安装脚本中。

45.35. pg_proc

目录pg_proc存放有关函数（或过程）的信息。更多信息请参见CREATE FUNCTION和第 35. 3 节。

该表同时包含聚合函数和普通函数的数据。如果proisagg为真，在pg_aggregate
中应该有一个相匹配的行。

表 45.35. pg_proc 列

Name	Type	References	Description
proname	name		Name of the function
pronamespace	oid	pg_namespace.oid	The OID of the namesp ace that contains this function
proowner	oid	pg_authid.oid	Owner of the function
prolang	oid	pg_language.oid	Implementation language or call interf ace of this function
procost	float4		Estimated execution cost (in units of cpu_ operator_cost); if

Name	Type	References	Description
			proretset, this is cost per row returned
prorows	float4		Estimated number of result rows (zero if not proretset)
provariadic	oid	pg_type.oid	Data type of the variadic array parameter's elements, or zero if the function does not have a variadic parameter
proisagg	bool		Function is an aggregate function
proiswindow	bool		Function is a window function
prosecdef	bool		Function is a security definer (i.e., a “setuid” function)
proisstrict	bool		Function returns null if any call argument is null. In that case the function won't actually be called at all. Functions that are not “strict” must be prepared to handle null inputs.
proretset	bool		Function returns a set (i.e., multiple values of the specified data type)
provolatile	char		provolatile tells whether the function's result depends only on its input arguments, or is affected by outside factors. It is i for “immutable” functions, which always deliver the same result for the same inputs. It is s for “stable” functions, whose results (for fixed inputs) do not change within a scan. It is v for “volatile” functions, whose results might change at any time. (Use v also for functions with side-effects, so that calls to them cannot get optimized away.)
pronargs	int2		Number of input arguments

Name	Type	References	Description
pronargdefaults	int2		Number of arguments that have defaults
prorettype	oid	pg_type.oid	Data type of the return value
proargtypes	oidvector	pg_type.oid	An array with the data types of the function arguments. This includes only input arguments (including INOUT and VARIADIC arguments), and thus represents the call signature of the function.
proallargtypes	oid[]	pg_type.oid	An array with the data types of the function arguments. This includes all arguments (including OUT and INOUT arguments); however, if all the arguments are IN arguments, this field will be null. Note that subscripting is 1-based, whereas for historical reasons proargtypes is subscripted from 0.
proargmodes	char[]		An array with the modes of the function arguments, encoded as i for IN arguments, o for OUT arguments, b for INOUT arguments, v for VARIADIC arguments, t for TABLE arguments. If all the arguments are IN arguments, this field will be null. Note that subscripts correspond to positions of proallargtypes not proargtypes.
proargnames	text[]		An array with the names of the function arguments. Arguments without a name are set to empty strings in the array. If none of the arguments have a name, this field will be null. Note that subscripts correspond to positions of proall

Name	Type	References	Description
			argtypes not proargtypes.
proargdefaults	pg_node_tree		Expression trees (in nodeToString() representation) for default values. This is a list with pronargdefaults elements, corresponding to the last *N* input arguments (i.e., the last *N* proargtypes positions). If none of the arguments have defaults, this field will be null.
prosrc	text		This tells the function handler how to invoke the function. It might be the actual source code of the function for interpreted languages, a link symbol, a file name, or just about anything else, depending on the implementation language/call convention.
probin	text		Additional information about how to invoke the function. Again, the interpretation is language-specific.
proconfig	text[]		Function's local settings for run-time configuration variables
proacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

对于编译好的函数，包括内置函数和动态装载函数，prosrc包含函数的 C 语言名字（链接符号）。所有其他已知的语言类型，prosrc包含函数的源码文本。
除了对于动态载入的C函数之外，probin不使用，对于动态载入的C函数，它给定了包含该函数的共享库文件的名称。

45.36. pg_rewrite

目录pg_rewrite存储对于表和视图的重写规则。

表 45.36. pg_rewrite 列

Name	Type	References	Description
rulename	name		Rule name
ev_class	oid	pg_class.oid	The table this rule is for

Name	Type	References	Description
ev_attr	int2		The column this rule is for (currently, always -1 to indicate the whole table)
ev_type	char		Event type that the rule is for: 1 = SELECT, 2 = UPDATE, 3 = INSERT, 4 = DELETE
ev_enabled	char		Controls in which session_replication_role modes the rule fires. O = rule fires in "origin" and "local" modes, D = rule is disabled, R = rule fires in "replica" mode, A = rule fires always.
is_instead	bool		True if the rule is an INSTEAD rule
ev_qual	pg_node_tree		Expression tree (in the form of a nodeToString() representation) for the rule's qualifying condition
ev_action	pg_node_tree		Query tree (in the form of a nodeToString() representation) for the rule's action

注意

如果一个表在这个目录中有任何规则，pg_class.relhasrules必须为真。

45.37. pg_seclabel

目录pg_seclabel存储数据库对象上的安全标签。安全标签可以通过SECURITY LABEL命令操纵。查看安全标签的更简单方法请见第 45.61 节。

表 45.37. pg_seclabel 列

Name	Type	References	Description
objoid	oid	any OID column	The OID of the object this security label pertains to
classoid	oid	pg_class.oid	The OID of the system catalog this object appears in
objsubid	int4		For a security label on a table column,

Name	Type	References	Description
			this is the column number (the `objoid` and `classoid` refer to the table itself). For all other object types, this column is zero.
`provider`	`text`		The label provider associated with this label.
`label`	`text`		The security label applied to this object.

45.38. pg_shdepend

目录`pg_shdepend`记录数据库对象和共享对象之间的依赖关系，例如角色。这些信息使得PostgreSQL可以确保对象在被删除时没有被其他对象引用。

另请参阅`pg_depend`，它对单个数据库中对象之间的依赖提供了相似的功能。

与大部分其他系统目录不同，`pg_shdepend`
在整个集簇的所有数据库之间共享：在每一个集簇中只有一个`pg_shdepend`的拷贝，
而不是每个数据库一份。

表 45.38. pg_shdepend 列

Name	Type	References	Description
`dbid`	`oid`	`pg_database.oid`	The OID of the database the dependent object is in, or zero for a shared object
`classid`	`oid`	`pg_class.oid`	The OID of the system catalog the dependent object is in
`objid`	`oid`	any OID column	The OID of the specific dependent object
`objsubid`	`int4`		For a table column, this is the column number (the `objid` and `classid` refer to the table itself). For all other object types, this column is zero.
`refclassid`	`oid`	`pg_class.oid`	The OID of the system catalog the referenced object is in (must be a shared catalog)
`refobjid`	`oid`	any OID column	The OID of the specific referenced object
`deptype`	`char`		A code defining the specific semantics of this dependency relationship; see text

在所有情况下，一条pg_shdepend记录都表示：如果不同时删除依赖对象，
就不能删除被引用对象。不过，还存在若干由deptype标识的子类型：

SHARED_DEPENDENCY_OWNER (o)

被引用对象（必须是一个角色）是依赖对象的拥有者。

SHARED_DEPENDENCY_ACL (a)

被引用对象（必须是角色）出现在依赖对象的 ACL（访问控制列表，即权限列表）中。（不会为对象拥有者创建 SHARED_DEPENDENCY_ACL 项，因为对象拥有者无论如何都会有一个 SHARED_DEPENDENCY_OWNER 项。）

SHARED_DEPENDENCY_PIN (p)

不存在依赖对象；这种项表示系统本身依赖于被引用对象，因此该对象绝不能删除。
这种项只由initdb创建。用于表示依赖对象的各列包含零。

将来可能还需要其他依赖类型。特别要注意的是，当前定义只支持把角色作为被引用对象。

45.39. pg_shdescription

目录pg_shdescription存储共享数据库对象的可选描述（注释）。
这些描述可以通过COMMENT命令设置或修改，并且可以使用psql的\d命令查看。

另请参阅pg_description，它对单个数据库内对象的描述提供了相似的功能。

与大部分其他系统目录不同，pg_shdescription
在整个集簇的所有数据库之间共享：在每一个集簇中只有一个pg_shdescription的拷贝，
而不是每个数据库一份。

表 45.39. pg_shdescription 列

Name	Type	References	Description
objoid	oid	any OID column	The OID of the object this description pertai ns to
classoid	oid	pg_class.oid	The OID of the system catalog this object appears in
description	text		Arbitrary text that serves as the descri ption of this object

45.40. pg_statistic

目录pg_statistic存储有关数据库内容的统计数据。其中的项由ANALYZE创建，
查询规划器会使用这些数据来进行查询规划。注意所有的统计数据天然就是近似的，
即使它刚刚被更新。

通常，每个已分析的表列都有一个条目，其中stainherit = false。如果表具有继承子表，则还会创建第二个条目，其中stainherit = true。此行表示继承树上列的统计信息，即你可以通过SELECT *column* FROM *table**看到的数据的统计信息，而stainherit = false行表示SELECT *column* FROM ONLY *table*的结果。

pg_statistic也存储关于索引表达式值的统计数据，就好像它们是真正的数据列，但在这种情况中starelid指索引。对一个普通非表达式索引列不会创建项，因为它将是底层表列的项的冗余。当前，索引表达式的项都具有stainherit = false。

由于不同种类的数据可能适合不同种类的统计信息，pg_statistic 在设计上尽量不对所存储的统计信息种类作出假定。只有极为通用的统计信息（例如空值情况）才在 pg_statistic 中有专用的列。其余统计信息都存储在“槽位”中。每个槽位都是一组相关的列，其内容由其中一列的代码编号来标识。更多信息见 src/include/catalog/pg_statistic.h。

pg_statistic不应公开可读，因为即使是表内容的统计信息也可能被认为是敏感的（例如，一个工资列的最大值和最小值就可能非常引人关注）。pg_stats是建立在pg_statistic之上的公开可读视图，它只会显示当前用户可读取的表的信息。

表 45.40. pg_statistic 列

Name	Type	References	Description
starelid	oid	pg_class.oid	The table or index that the described column belongs to
staattnum	int2	pg_attribute.attnum	The number of the described column
stainherit	bool		If true, the stats include inheritance child columns, not just the values in the specified relation
stanullfrac	float4		The fraction of the column's entries that are null
stawidth	int4		The average stored width, in bytes, of nonnull entries
stadistinct	float4		The number of distinct nonnull data values in the column. A value greater than zero is the actual number of distinct values. A value less than zero is the negative of a multiplier for the number of rows in the table; for example, a column in which about 80% of the values are nonnull and each nonnull value appears about twice on average could be represented by stadistinct = -0.4. A zero value means the number of distinct values is unknown.
stakind*N*	int2		A code number indicating the kind of statis

Name	Type	References	Description
			tics stored in the *N* th “slot” of the pg_statistic row.
staop*N*	oid	pg_operator.oid	An operator used to derive the statis tics stored in the *N*th “slot”. For example, a histogram slot would show the < operator that defines the sort order of the data.
stanumbers*N*	float4[]		Numerical statistics of the appropriate kind for the *N*th “slot”, or null if the slot kind does not involve numerical values
stavalues*N*	anyarray		Column data values of the appropriate kind for the *N*th “slot”, or null if the slot kind does not store any data values. Each array's element values are actually of the specific column's data type, so there is no way to define these column s' type more specif ically than anyarray.

45.41. pg_tablespace

目录pg_tablespace存储关于可用表空间的信息。
表可以被放置在特定表空间中以实现磁盘布局的管理。

与大部分其他系统目录不同，pg_tablespace
在整个集簇的所有数据库之间共享：在每一个集簇中只有一个pg_tablespace的拷贝，
而不是每个数据库一份。

表 45.41. pg_tablespace 列

Name	Type	References	Description
spcname	name		Tablespace name
spcowner	oid	pg_authid.oid	Owner of the tables pace, usually the user who created it
spclocation	text		Location (directory path) of the tablespace
spcacl	aclitem[]		Access privileges; see GRANT and REVOKE for details

Name	Type	References	Description
spcoptions	text[]		Tablespace-level options, as “keyword=value” strings

45.42. pg_trigger

目录pg_trigger存储表和视图上的触发器。详见CREATE TRIGGER。

表 45.42. pg_trigger 列

Name	Type	References	Description
tgrelid	oid	pg_class.oid	The table this trigger is on
tgname	name		Trigger name (must be unique among triggers of same table)
tgfoid	oid	pg_proc.oid	The function to be called
tgtype	int2		Bit mask identifying trigger firing conditions
tgenabled	char		Controls in which session_replication_role modes the trigger fires. O = trigger fires in “origin” and “local” modes, D = trigger is disabled, R = trigger fires in “replica” mode, A = trigger fires always.
tgisinternal	bool		True if trigger is internally generated (usually, to enforce the constraint identified by tgconstraint)
tgconstrrelid	oid	pg_class.oid	The table referenced by a referential integrity constraint
tgconstrindid	oid	pg_class.oid	The index supporting a unique, primary key, or referential integrity constraint
tgconstraint	oid	pg_constraint.oid	The pg_constraint entry associated with the trigger, if any
tgdeferrable	bool		True if constraint trigger is deferrable
tginitdeferred	bool		True if constraint trigger is initially deferred

Name	Type	References	Description
tgnargs	int2		Number of argument strings passed to trigger function
tgattr	int2vector	pg_attribute.attnum	Column numbers, if trigger is column-specific; otherwise an empty array
tgargs	bytea		Argument strings to pass to trigger, each NULL-terminated
tgqual	pg_node_tree		Expression tree (in nodeToString() representation) for the trigger's WHEN condition, or null if none

当前，列限定触发器只支持UPDATE事件，因此tgattr只用于这种事件类型。tgtype也可以包含用于其他事件类型的位，但那些事件类型对应的是表级触发器，并且会忽略tgattr。

> **注意**
>
> 当tgconstraint非零时，tgconstrrelid、tgconstrindid、tgdeferrable和tginitdeferred与被引用的pg_constraint 项有很大的冗余。但是，存在将一个不可延迟触发器关联到一个可延迟约束的可能性：外键约束可以有一些可延迟和一些不可延迟触发器。

> **注意**
>
> 如果一个关系在本目录中拥有任何触发器，其pg_class.relhastriggers必须为真。

45.43. pg_ts_config

pg_ts_config系统目录包含表示文本检索配置的条目。一个配置指定某个特定的文本检索解析器，以及针对该解析器每种输出词元类型所定义的词典列表。解析器记录在pg_ts_config条目中，而词元到词典的映射则由pg_ts_config_map中的辅助项定义。

PostgreSQL的文本检索特性在第 12 章中有更详尽的描述。

表 45.43. pg_ts_config 列

Name	Type	References	Description
cfgname	name		Text search configuration name
cfgnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this configuration
cfgowner	oid	pg_authid.oid	Owner of the configuration

Name	Type	References	Description
cfgparser	oid	pg_ts_parser.oid	The OID of the text search parser for this configuration

45.44. pg_ts_config_map

pg_ts_config_map 系统目录中的条目说明了，对于每个文本检索配置所用解析器的每种输出词元类型，应当查询哪些文本检索词典以及查询的顺序。

PostgreSQL的文本检索特性在第 12 章中有更详尽的描述。

表 45.44. pg_ts_config_map 列

Name	Type	References	Description
mapcfg	oid	pg_ts_config.oid	The OID of the pg_ts_config entry owning this map entry
maptokentype	integer		A token type emitted by the configuration's parser
mapseqno	integer		Order in which to consult this entry (lower mapseqnos first)
mapdict	oid	pg_ts_dict.oid	The OID of the text search dictionary to consult

45.45. pg_ts_dict

pg_ts_dict系统目录包含定义文本检索词典的项。一个词典依赖于一个文本检索模板，它指定了所有需要的实现函数，词典本身则为模板支持的用户可设置参数提供值。这种分工允许普通用户创建词典。参数由一个文本串dictinitoption定义，其格式和意义随着模板而变化。

PostgreSQL的文本检索特性在第 12 章中有更详尽的描述。

表 45.45. pg_ts_dict 列

Name	Type	References	Description
dictname	name		Text search dictionary name
dictnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this dictionary
dictowner	oid	pg_authid.oid	Owner of the dictionary
dicttemplate	oid	pg_ts_template.oid	The OID of the text search template for this dictionary
dictinitoption	text		Initialization option string for the template

45.46. pg_ts_parser

pg_ts_parser系统目录包含定义文本检索解析器的项。
一个解析器负责将输入文本分割成词位并为每一个词位分配一个词元类型。
由于一个解析器必须用 C 语言级别的函数实现，创建新解析器的工作只限于数据库的超级用户。

PostgreSQL的文本检索特性在第 12 章中有更详尽的描述。

表 45.46. pg_ts_parser 列

Name	Type	References	Description
prsname	name		Text search parser name
prsnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this parser
prsstart	regproc	pg_proc.oid	OID of the parser's startup function
prstoken	regproc	pg_proc.oid	OID of the parser's next-token function
prsend	regproc	pg_proc.oid	OID of the parser's shutdown function
prsheadline	regproc	pg_proc.oid	OID of the parser's headline function
prslextype	regproc	pg_proc.oid	OID of the parser's lextype function

45.47. pg_ts_template

pg_ts_template系统目录包含定义文本检索模板的项。
一个模板是一类文本检索词典的实现骨架。由于一个模板必须用 C 语言级别的函数实现，新模板的创建只限于数据库超级用户。

PostgreSQL的文本检索特性在第 12 章中有更详尽的描述。

表 45.47. pg_ts_template 列

Name	Type	References	Description
tmplname	name		Text search template name
tmplnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this template
tmplinit	regproc	pg_proc.oid	OID of the template's initialization function
tmpllexize	regproc	pg_proc.oid	OID of the template's lexize function

45.48. pg_type

目录pg_type存储有关数据类型的信息。基础类型和枚举类型（标量类型）使用CREATE TYPE创建，而域使用CREATE DOMAIN创建。数据库中的每一个表都会有一个自动创建的复合类型，用于表示表的行结构。也可以使用CREATE TYPE AS创建复合类型。

表 45.48. pg_type 列

Name	Type	References	Description
typname	name		Data type name
typnamespace	oid	pg_namespace.oid	The OID of the namespace that contains this type
typowner	oid	pg_authid.oid	Owner of the type
typlen	int2		For a fixed-size type, typlen is the number of bytes in the internal representation of the type. But for a variable-length type, typlen is negative. -1 indicates a “varlena” type (one that has a length word), -2 indicates a null-terminated C string.
typbyval	bool		typbyval determines whether internal routines pass a value of this type by value or by reference. typbyval had better be false if typlen is not 1, 2, or 4 (or 8 on machines where Datum is 8 bytes). Variable-length types are always passed by reference. Note that typbyval can be false even if the length would allow pass-by-value.
typtype	char		typtype is b for a base type, c for a composite type (e.g., a table's row type), d for a domain, e for an enum type, or p for a pseudo-type. See also typrelid and typbasetype.
typcategory	char		typcategory is an arbitrary classification of data types that is used by the parser to determine which implicit casts should be “preferred”. See 表 45.49.

Name	Type	References	Description
typispreferred	bool		True if the type is a preferred cast target within its typcategory
typisdefined	bool		True if the type is defined, false if this is a placeholder entry for a not-yet-define d type. When typisd efined is false, nothin g except the type name, namespace, and OID can be relied on.
typdelim	char		Character that separa tes two values of this type when parsing array input. Note that the delimiter is associ ated with the array element data type, not the array data type.
typrelid	oid	pg_class.oid	If this is a composite type (see typtype), then this column points to the pg_class entry that defines the corres ponding table. (For a free-standing compos ite type, the pg_class entry doesn't really represent a table, but it is needed anyway for the type's pg_attrib ute entries to link to.) Zero for non-compos ite types.
typelem	oid	pg_type.oid	If typelem is not 0 then it identifies another row in pg_type. The current type can then be subscripted like an array yielding values of type typelem. A “true” array type is variable length (typlen = -1), but some fixed-length (typlen > 0) types also have nonzero typelem, for example name and point. If a fixed-length type has a typelem then its internal repres entation must be some number of values of the typelem data type

Name	Type	References	Description
			with no other data. Variable-length array types have a header defined by the array subroutines.
typarray	oid	pg_type.oid	If typarray is not 0 then it identifies another row in pg_type, which is the “true” array type having this type as element
typinput	regproc	pg_proc.oid	Input conversion function (text format)
typoutput	regproc	pg_proc.oid	Output conversion function (text format)
typreceive	regproc	pg_proc.oid	Input conversion function (binary format), or 0 if none
typsend	regproc	pg_proc.oid	Output conversion function (binary format), or 0 if none
typmodin	regproc	pg_proc.oid	Type modifier input function, or 0 if type does not support modifiers
typmodout	regproc	pg_proc.oid	Type modifier output function, or 0 to use the standard format
typanalyze	regproc	pg_proc.oid	Custom ANALYZE function, or 0 to use the standard function
typalign	char		typalign表示存储此类型值时所要求的对齐方式。它适用于磁盘存储，以及该值在 PostgreSQL 内部的大多数表示形式。当多个值连续存放时，例如在磁盘上完整行的表示中，这种类型的数据前会插入填充，使其从指定边界开始。对齐参考点是该序列中第一个数据的起始位置。 可能的值为： • c = char对齐，即不需要对齐。 • s = short对齐（在大部分机器上为2字节）。

Name	Type	References	Description
			• i = int对齐（在大部分机器上为4字节）。 • d = double对齐（在很多机器上为8字节，但绝不是全部）。 **注意** 对于系统表中使用的类型，关键的是在pg_type定义的大小和对齐与编译器在表示表行的结构体中布局列的方式一致。
typstorage	char		typstorage用于指示 varlena 类型（即typlen = -1）是否支持TOAST，以及这种类型的属性应采用什么默认策略。可能的值为： • p：值必须始终以普通方式存储。 • e：值可以存储在一个“二级”关系中（如果关系有这样的关系，参见pg_class.reltoastrelid）。 • m：值可以被压缩并内联存储。 • x：值可以被压缩并内联存储，或存储在“二级”存储中。 注意，m列也可以被移动到二级存储，但只能作为最后手段（e和x列会先被移动）。
typnotnull	bool		typnotnull表示类型上的一个非空约束。只用于域。
typbasetype	oid	pg_type.oid	如果这是一个域（见typtype），则typbasetype标识该域所基于的类型。如果此类型不是域，则为 0。
typtypmod	int4		域使用typtypmod来记录被应用于它们基础类型

Name	Type	References	Description
			的typmod（如果基础类型不使用typmod，则为-1）。如果此类型不是一个域则为-1。
typndims	int4		对于基于数组的域，typndims是数组维度数（即，typbasetype是一个数组类型）。对于不是基于数组的域的类型，此列为零。
typcollation	oid	pg_collation.oid	typcollation指定此类型的排序规则。如果类型不支持排序规则，此列为零。支持排序规则的基础类型在这里会有DEFAULT_COLLATION_OID。基于支持排序规则的类型的域，如果为其指定了排序规则 OID，则它的排序规则 OID 可以不同于其基础类型的排序规则 OID。
typdefaultbin	pg_node_tree		如果typdefaultbin非空，那么它就是该类型默认表达式的nodeToString()表示形式。此列只用于域。
typdefault	text		如果某类型没有相关默认值，那么typdefault为空。如果typdefaultbin不为空，那么typdefault必须包含由typdefaultbin表示的默认表达式的人类可读版本。如果typdefaultbin为空而typdefault不为空，则typdefault是该类型默认值的外部表示形式，它可以被交给该类型的输入转换器来产生一个常量。

表 45.49列出了typcategory的系统定义值。任何未来对此列表的增加都将是大写ASCII字母。所有其他ASCII字符都保留给用户定义的类别。

表 45.49. typcategory Codes

Code	Category
A	Array types
B	Boolean types
C	Composite types
D	Date/time types
E	Enum types
G	Geometric types

Code	Category
I	Network address types
N	Numeric types
P	Pseudo-types
S	String types
T	Timespan types
U	User-defined types
V	Bit-string types
X	unknown type

45.49. pg_user_mapping

目录pg_user_mapping存储从本地用户到远程的映射。对这个目录的访问对普通用户有限制，可使用视图pg_user_mappings替代。

表 45.50. pg_user_mapping 列

Name	Type	References	Description
umuser	oid	pg_authid.oid	OID of the local role being mapped, 0 if the user mapping is public
umserver	oid	pg_foreign_server.oid	The OID of the foreign server that contains this mapping
umoptions	text[]		User mapping specific options, as "keyword=value" strings

45.50. 系统视图

除了系统目录之外，PostgreSQL还提供了一些内置视图。
一些系统视图提供了对系统目录中一些常用查询的便捷访问。
其他视图则提供了对内部服务器状态的访问。

信息模式（第 34 章）提供了一组与系统视图功能重叠的替代视图。由于信息模式是SQL标准，而这里描述的视图是PostgreSQL特有的，所以如果信息模式提供了所需的全部信息，
通常最好使用信息模式。

表 45.51列出了这里描述的系统视图。每个视图的更详细文档见下文。还有一些额外的视图提供了对统计信息收集器结果的访问；它们在表 27.1中描述。

除非另有说明，否则这里描述的所有视图都是只读的。

表 45.51. 系统视图

View Name	Purpose
pg_available_extensions	available extensions
pg_available_extension_versions	available versions of extensions
pg_cursors	open cursors
pg_group	groups of database users

View Name	Purpose
pg_indexes	indexes
pg_locks	currently held locks
pg_prepared_statements	prepared statements
pg_prepared_xacts	prepared transactions
pg_roles	database roles
pg_rules	rules
pg_seclabels	security labels
pg_settings	parameter settings
pg_shadow	database users
pg_stats	planner statistics
pg_tables	tables
pg_timezone_abbrevs	time zone abbreviations
pg_timezone_names	time zone names
pg_user	database users
pg_user_mappings	user mappings
pg_views	views

45.51. pg_available_extensions

pg_available_extensions视图列出了可供安装的扩展。另请参阅pg_extension目录，显示当前已安装的扩展。

表 45.52. pg_available_extensions 列

Name	Type	Description
name	name	Extension name
default_version	text	Name of default version, or NULL if none is specified
installed_version	text	Currently installed version of the extension, or NULL if not installed
comment	text	Comment string from the extension's control file

pg_available_extensions视图是只读的。

45.52. pg_available_extension_versions

pg_available_extension_versions视图列出了可供安装的特定扩展版本。另请参阅pg_extension目录，显示当前已安装的扩展。

表 45.53. pg_available_extension_versions 列

Name	Type	Description
name	name	Extension name

Name	Type	Description
version	text	Version name
installed	bool	True if this version of this extension is currently installed
superuser	bool	True if only superusers are allowed to install this extension
relocatable	bool	True if extension can be reloca ted to another schema
schema	name	Name of the schema that the extension must be installed into, or NULL if partially or fully relocatable
requires	name[]	Names of prerequisite extens ions, or NULL if none
comment	text	Comment string from the extension's control file

pg_available_extension_versions视图是只读的。

45.53. pg_cursors

该pg_cursors视图列出当前可用的游标。游标可以通过几种方式定义：

- 通过SQL中的DECLARE语句
- 通过前端/后端协议中的Bind消息，如第 46.2.3 节中所述
- 通过服务器编程接口(SPI)，如第 43.1 节中所述

该pg_cursors视图会显示通过上述任意方式创建的游标。
游标仅在定义它的事务持续期间存在，除非将其声明为WITH HOLD。因此，不可保持的游标只会在此视图中保留到创建它的事务结束为止。

> **注意**
>
> 游标在内部用于实现PostgreSQL的某些组件，如过程语言。因此，pg_cursors视图可能包含用户未明确创建的游标。

表 45.54. pg_cursors 列

Name	Type	Description
name	text	The name of the cursor
statement	text	The verbatim query string submitted to declare this cursor
is_holdable	boolean	true if the cursor is holdable (that is, it can be accessed after the transaction that declared the cursor has committed); false otherwise
is_binary	boolean	true if the cursor was declar ed BINARY; false otherwise

Name	Type	Description
is_scrollable	boolean	true if the cursor is scrollable (that is, it allows rows to be retrieved in a nonsequential manner); false otherwise
creation_time	timestamptz	The time at which the cursor was declared

pg_cursors视图是只读的。

45.54. pg_group

视图pg_group存在是为了向后兼容性：它模拟了在PostgreSQL版本8.1之前存在的目录。
它显示了所有被标记为非rolcanlogin的角色的名称和成员，
这是对被用作组的角色集合的近似。

表 45.55. pg_group 列

Name	Type	References	Description
groname	name	pg_authid.rolname	Name of the group
grosysid	oid	pg_authid.oid	ID of this group
grolist	oid[]	pg_authid.oid	An array containing the IDs of the roles in this group

45.55. pg_indexes

视图pg_indexes提供了有关数据库中每个索引的有用信息。

表 45.56. pg_indexes 列

Name	Type	References	Description
schemaname	name	pg_namespace.nspname	Name of schema containing table and index
tablename	name	pg_class.relname	Name of table the index is for
indexname	name	pg_class.relname	Name of index
tablespace	name	pg_tablespace.spcname	Name of tablespace containing index (null if default for database)
indexdef	text		Index definition (a reconstructed CREATE INDEX command)

45.56. pg_locks

视图pg_locks提供了对数据库服务器中打开的事务持有的锁的信息的访问。
有关锁定的更多讨论，请参见第 13 章。

pg_locks包含每个活动的可锁定对象、请求的锁模式和相关事务的一行。因此，如果多个事务正在持有或等待锁定它，同一可锁定对象可能会出现多次。但是，当前没有任何锁定的对象将不会出现。

有几种不同类型的可锁定对象：整个关系（例如，表），关系的单个页面，关系的单个元组，事务ID（虚拟和永久ID均包括），以及一般的数据库对象（由类OID和对象OID标识，与pg_description或 pg_depend中的方式相同）。此外，还可以对具有用户定义含义的数字施加“咨询锁”。

表 45.57. pg_locks 列

Name	Type	References	Description
locktype	text		Type of the lockable object: relation, extend, page, tuple, transactionid, virtualxid, object, userlock, or advisory
database	oid	pg_database.oid	OID of the database in which the object exists, or zero if the object is a shared object, or null if the object is a transaction ID
relation	oid	pg_class.oid	OID of the relation, or null if the object is not a relation or part of a relation
page	integer		Page number within the relation, or null if the object is not a tuple or relation page
tuple	smallint		Tuple number within the page, or null if the object is not a tuple
virtualxid	text		Virtual ID of a transaction, or null if the object is not a virtual transaction ID
transactionid	xid		ID of a transaction, or null if the object is not a transaction ID
classid	oid	pg_class.oid	OID of the system catalog containing the object, or null if the object is not a general database object
objid	oid	any OID column	OID of the object within its system catalog, or null if the object is not a general database object
objsubid	smallint		Column number targeted by the lock (the

Name	Type	References	Description
			classid and objid refer to the table itself), or zero if the target is some other general database object, or null if the target is not a general database object
virtualtransaction	text		Virtual ID of the transaction that is holding or awaiting this lock
pid	integer		Process ID of the server process holding or awaiting this lock, or null if the lock is held by a prepared transaction
mode	text		Name of the lock mode held or desired by this process (see 第 13.3.1 节 and 第 13.2.3 节)
granted	boolean		True if lock is held, false if lock is awaited

在表示由指定事务持有的锁的行中，granted 为真。为假表示该事务当前正在等待获取此锁，这意味着某个其他事务正在同一可锁定对象上持有冲突的锁模式。等待的事务会休眠，直到另一个锁被释放（或检测到死锁情况）。单个事务一次最多只能等待获取一个锁。

每个事务在其整个持续期内都对其虚拟事务ID持有独占锁。
如果为事务分配了永久ID（通常仅在事务改变数据库状态时才会发生），
它还会对其永久事务ID持有独占锁直到事务结束。
当一个事务发现有必要专门等待另一个事务时，
它会尝试获取另一个事务ID（根据情况是虚拟ID还是永久ID）的共享锁。
只有当另一个事务终止并释放其锁时，这才会成功。

虽然元组是一种可锁定的对象类型，但关于行级锁的信息存储在磁盘上，而不是内存中，
因此行级锁通常不会出现在此视图中。如果一个事务正在等待行级锁，它通常会出现在视图中，
等待当前持有该行锁的永久事务ID。

咨询锁可以在由单个 bigint 值或两个整数值组成的键上获取。bigint 键显示时，其高位一半放在 classid 列，低位一半放在 objid 列，而 objsubid 等于 1。整数键显示时，第一个键放在 classid 列，第二个键放在 objid 列，而 objsubid 等于 2。The actual meaning of the keys is up to the user. Advisory locks are local to each database, so the database column is meaningful for an advisory lock.

pg_locks提供了集簇中所有锁的全局视图，不仅包括与当前数据库相关的锁。虽然它的relation列可以与pg_class.oid 连接来识别被锁定的关系，但这仅对当前数据库中的关系有效（即 database 列为当前数据库的 OID 或零的那些关系）。

pid 列可以连接到 pg_stat_activity 视图的 procpid 列，以获得关于持有或等待持有每个锁的会话的更多信息，例如

```
SELECT * FROM pg_locks pl LEFT JOIN pg_stat_activity psa
    ON pl.pid = psa.procpid;
```

Also, if you are using prepared transactions, the `virtualtransaction` column can be joined to the `transaction` column of the `pg_prepared_xacts` view to get more information on prepared transactions that hold locks. (A prepared transaction can never be waiting for a lock, but it continues to hold the locks it acquired while running.) For example:

```
SELECT * FROM pg_locks pl LEFT JOIN pg_prepared_xacts ppx
    ON pl.virtualtransaction = '-1/' || ppx.transaction;
```

`pg_locks` 视图显示来自常规锁管理器和谓词锁管理器这两个独立系统的数据。访问该视图时，每个锁管理器的内部数据结构会被短暂加锁，并制作副本供视图显示。
因此每个锁管理器都会产生一组一致的结果，但由于我们没有同时锁住两个锁管理器，
在查询常规锁管理器之后、查询谓词锁管理器之前，可能会有锁被取得或释放。
每个锁管理器只被加锁尽可能短的时间以减少查询该视图对性能的影响，但如果频繁访问它，
仍可能对数据库性能产生一些影响。

45.57. pg_prepared_statements

`pg_prepared_statements` 视图显示当前会话中所有可用的预备语句。关于预备语句的更多信息，参见PREPARE。

`pg_prepared_statements` 为每个预备语句包含一行。创建新的预备语句时，会向该视图添加一行；释放预备语句时（例如通过DEALLOCATE命令）则会移除该行。

表 45.58. pg_prepared_statements 列

Name	Type	Description
`name`	`text`	The identifier of the prepared statement
`statement`	`text`	The query string submitted by the client to create this prepared statement. For prepared statements created via SQL, this is the `PREPARE` statement submitted by the client. For prepared statements created via the frontend/backend protocol, this is the text of the prepared statement itself.
`prepare_time`	`timestamptz`	The time at which the prepared statement was created
`parameter_types`	`regtype[]`	The expected parameter types for the prepared statement in the form of an array of `regtype`. The OID corresponding to an element of this array can be obtained by casting the `regtype` value to `oid`.
`from_sql`	`boolean`	`true` if the prepared statement was created via the `PREPARE` SQL statement; `false` if the statement was prepared via the frontend/backend protocol

pg_prepared_statements视图是只读的。

45.58. pg_prepared_xacts

视图pg_prepared_xacts显示当前准备进行两阶段提交的事务的信息 (详见PREPARE TRANSACTION)。

pg_prepared_xacts 为每个预备事务包含一行。事务提交或回滚时，对应条目会被移除。

表 45.59. pg_prepared_xacts 列

Name	Type	References	Description
transaction	xid		Numeric transaction identifier of the prepared transaction
gid	text		Global transaction identifier that was assigned to the transaction
prepared	timestamp with time zone		Time at which the transaction was prepared for commit
owner	name	pg_authid.rolname	Name of the user that executed the transaction
database	name	pg_database.datname	Name of the database in which the transaction was executed

当访问pg_prepared_xacts视图时，内部事务管理器数据结构会被暂时锁定，
并为视图制作一份副本用以显示。这确保视图生成一致的结果集，
同时不会不必要地阻塞正常操作。尽管如此，如果频繁访问该视图，
可能会对数据库性能产生一些影响。

45.59. pg_roles

视图pg_roles提供数据库角色的信息。它本质上是 pg_authid 的公开可读视图，并隐藏密码字段。

此视图显式显示底层表的 OID 列，因为与其他目录连接时需要此列。

表 45.60. pg_roles 列

Name	Type	References	Description
rolname	name		Role name
rolsuper	bool		Role has superuser privileges
rolinherit	bool		Role automatically inherits privileges of roles it is a member of

Name	Type	References	Description
rolcreaterole	bool		Role can create more roles
rolcreatedb	bool		Role can create databa ses
rolcatupdate	bool		Role can update system catalogs direct ly. (Even a superuser cannot do this unless this column is true)
rolcanlogin	bool		Role can log in. That is, this role can be given as the initial session authorization identi fier
rolreplication	bool		Role is a replication role. That is, this role can initiate streaming replication (see 第 25.2. 5 节) and set/unset the system backup mode using pg_start_ backup and pg_stop_ backup
rolconnlimit	int4		For roles that can log in, this sets maximum number of concurrent connections this role can make. -1 means no limit.
rolpassword	text		Not the password (always reads as ********)
rolvaliduntil	timestamptz		Password expiry time (only used for password authentication); null if no expiration
rolconfig	text[]		Role-specific defaul ts for run-time config uration variables
oid	oid	pg_authid.oid	ID of role

45.60. pg_rules

视图pg_rules提供了有关查询重写规则的有用信息。

表 45.61. pg_rules 列

Name	Type	References	Description
schemaname	name	pg_namespace. nspname	Name of schema containing table
tablename	name	pg_class.relname	Name of table the rule is for

Name	Type	References	Description
rulename	name	pg_rewrite.rulename	Name of rule
definition	text		Rule definition (a reconstructed creation command)

pg_rules 视图不包含视图的 ON SELECT 规则；那些可以在 pg_views 中看到。

45.61. pg_seclabels

视图pg_seclabels提供有关安全标签的信息。它是 pg_seclabel目录的更易查询版本。

表 45.62. pg_seclabels 列

Name	Type	References	Description
objoid	oid	any OID column	The OID of the object this security label pertains to
classoid	oid	pg_class.oid	The OID of the system catalog this object appears in
objsubid	int4		For a security label on a table column, this is the column number (the objoid and classoid refer to the table itself). For all other object types, this column is zero.
objtype	text		The type of object to which this label applies, as text.
objnamespace	oid	pg_namespace.oid	The OID of the namespace for this object, if applicable; otherwise NULL.
objname	text		The name of the object to which this label applies, as text.
provider	text	pg_seclabel.provider	The label provider associated with this label.
label	text	pg_seclabel.label	The security label applied to this object.

45.62. pg_settings

视图pg_settings提供对服务器运行时参数的访问。它实质上是SHOW和SET命令的替代接口。它还提供了一些关于每个参数的事实，这些事实无法直接从SHOW中获取，例如最小值和最大值。

表 45.63. pg_settings 列

Name	Type	Description
name	text	Run-time configuration parameter name
setting	text	Current value of the parameter
unit	text	Implicit unit of the parameter
category	text	Logical group of the parameter
short_desc	text	A brief description of the parameter
extra_desc	text	Additional, more detailed, description of the parameter
context	text	Context required to set the parameter's value (see below)
vartype	text	Parameter type (bool, enum, integer, real, or string)
source	text	Source of the current parameter value
min_val	text	Minimum allowed value of the parameter (null for non-numeric values)
max_val	text	Maximum allowed value of the parameter (null for non-numeric values)
enumvals	text[]	Allowed values of an enum parameter (null for non-enum values)
boot_val	text	Parameter value assumed at server startup if the parameter is not otherwise set
reset_val	text	Value that RESET would reset the parameter to in the current session
sourcefile	text	Configuration file the current value was set in (null for values set from sources other than configuration files, or when examined by a non-superuser); helpful when using include directives in configuration files
sourceline	integer	Line number within the configuration file the current value was set at (null for values set from sources other than configuration files, or when examined by a non-superuser)

有几种可能的context值。按照更改设置的难度递减的顺序，它们是：

internal

这些设置不能直接更改；它们反映了内部确定的值。
其中一些可能可通过使用不同的配置选项重新构建服务器，或通过更改提供给initdb的选项来进行调整。

postmaster

这些设置只能在服务器启动时应用，因此任何更改都需要重启服务器。
这些设置的值通常存储在 postgresql.conf 文件中，或者在服务器启动时通过命令行传递。当然，任何较低 context 类型的设置也都可以在服务器启动时设置。

sighup

可以在 postgresql.conf 中更改这些设置，而无需重启服务器。向 postmaster 发送 SIGHUP 信号，会使其重新读取 postgresql.conf 并应用这些更改。postmaster 还会将 SIGHUP 信号转发给其子进程，以便它们都采用新值。

backend

可以在 postgresql.conf 中更改这些设置，而无需重启服务器。也可以在连接请求报文中为特定会话设置这些值（例如，通过 libpq 的 PGOPTIONS 环境变量）；任何用户都可以为其会话进行此类更改。但是，在会话启动后，这些设置永远不会更改。如果在 postgresql.conf 中更改它们，请向 postmaster 发送 SIGHUP 信号，使其重新读取 postgresql.conf。新值只会影响随后启动的会话。

superuser

这些设置可以从postgresql.conf中设置，或者通过SET命令在会话中设置；
但只有超级用户可以通过SET更改它们。如果没有使用SET建立会话本地值，那么在postgresql.conf中的更改也会影响到现有会话。

user

这些设置可以从postgresql.conf中设置，或通过SET命令在会话中设置。
任何用户都可以更改其会话本地值。如果没有使用SET建立会话本地值，那么在postgresql.conf中的更改也会影响到现有会话。

有关更改这些参数的各种方法，详见第 18.1 节。

pg_settings视图不能插入或删除，但可以更新。对 pg_settings 的某一行执行 UPDATE，等同于对该命名参数执行SET命令。更改只影响当前会话所用的值。
如果在之后被中止的事务中发出了 UPDATE，那么事务回滚时 UPDATE 命令的效果也会消失。一旦外围事务提交，这些效果会持续到会话结束，除非被另一个 UPDATE 或 SET 覆盖。

45.63. pg_shadow

视图pg_shadow存在是为了向后兼容性：它模拟了在PostgreSQL 8.1版本之前存在的目录。它显示了在pg_authid中标记为rolcanlogin的所有角色的属性。

这个名称源于该表不应向公众开放读取，因为它包含密码。pg_user 是 pg_shadow 的公开可读视图，并隐藏密码字段。

表 45.64. pg_shadow 列

Name	Type	References	Description
usename	name	pg_authid.rolname	User name

Name	Type	References	Description
usesysid	oid	pg_authid.oid	ID of this user
usecreatedb	bool		User can create databa ses
usesuper	bool		User is a superuser
usecatupd	bool		User can update system catalogs. (Even a superuser cannot do this unless this column is true.)
passwd	text		Password (possibly encrypted); null if none. See pg_authid for details of how encrypted passwords are stored.
valuntil	abstime		Password expiry time (only used for password authentication)
useconfig	text[]		Session defaults for run-time configuration variables

45.64. pg_stats

视图pg_stats提供对存储在pg_statistic目录中信息的访问。
此视图仅允许访问用户具有读取权限的表对应的pg_statistic行，
因此可以安全地允许对此视图进行公共读取访问。

pg_stats也旨在以比底层目录更易读的格式呈现信息— 但其模式必须在为pg_statistic定义新的槽类型时进行扩展。

表 45.65. pg_stats 列

Name	Type	References	Description
schemaname	name	pg_namespace. nspname	Name of schema containing table
tablename	name	pg_class.relname	Name of table
attname	name	pg_attribute. attname	Name of the column described by this row
inherited	bool		If true, this row includ es inheritance child columns, not just the values in the specified table
null_frac	real		Fraction of column entries that are null
avg_width	integer		Average width in bytes of column's entries
n_distinct	real		If greater than zero, the estimated number of distinct values in the

Name	Type	References	Description
			column. If less than zero, the negative of the number of distinct values divided by the number of rows. (The negated form is used when `ANALYZE` believes that the number of distinct values is likely to increase as the table grows; the positive form is used when the column seems to have a fixed number of possible values.) For example, -1 indicates a unique column in which the number of distinct values is the same as the number of rows.
`most_common_vals`	`anyarray`		A list of the most common values in the column. (Null if no values seem to be more common than any others.) For some data types such as `tsvector`, this is a list of the most common element values rather than values of the type itself.
`most_common_freqs`	`real[]`		A list of the frequencies of the most common values or elements, i.e., number of occurrences of each divided by total number of rows. (Null when `most_common_vals` is.) For some data types such as `tsvector`, it can also store some additional information, making it longer than the `most_common_vals` array.
`histogram_bounds`	`anyarray`		A list of values that divide the column's values into groups of approximately equal population. The values in `most_common_vals`, if present, are omitted from this histogram calculation. (This

Name	Type	References	Description
			column is null if the column data type does not have a < operator or if the most_common_vals list accounts for the entire population.)
correlation	real		Statistical correlation between physical row ordering and logical ordering of the column values. This ranges from -1 to +1. When the value is near -1 or +1, an index scan on the column will be estimated to be cheaper than when it is near zero, due to reduction of random access to the disk. (This column is null if the column data type does not have a < operator.)

most_common_vals 和 histogram_bounds 数组中的最大项数可以逐列地用 ALTER TABLE SET STATISTICS 命令设置，或者全局地通过设置default_statistics_target运行时参数来设置。

45.65. pg_tables

视图pg_tables提供了对数据库中每个表的有用信息的访问。

表 45.66. pg_tables 列

Name	Type	References	Description
schemaname	name	pg_namespace.nspname	Name of schema containing table
tablename	name	pg_class.relname	Name of table
tableowner	name	pg_authid.rolname	Name of table's owner
tablespace	name	pg_tablespace.spcname	Name of tablespace containing table (null if default for database)
hasindexes	boolean	pg_class.relhasindex	True if table has (or recently had) any indexes
hasrules	boolean	pg_class.relhasrules	True if table has (or once had) rules
hastriggers	boolean	pg_class.relhastriggers	True if table has (or once had) triggers

45.66. pg_timezone_abbrevs

视图pg_timezone_abbrevs提供当前被日期时间输入例程识别的时区缩写列表。当修改timezone_abbreviations运行时参数时，该视图的内容会发生变化。

表 45.67. pg_timezone_abbrevs 列

Name	Type	Description
abbrev	text	Time zone abbreviation
utc_offset	interval	Offset from UTC (positive means east of Greenwich)
is_dst	boolean	True if this is a daylight-savings abbreviation

虽然大多数时区缩写代表与UTC的固定偏移量，但也有一些在历史上变化过（有关更多信息，请参见第 B.3 节）。在这种情况下，此视图呈现它们当前的含义。

45.67. pg_timezone_names

视图pg_timezone_names提供了一个由SET TIMEZONE识别的时区名称列表，以及它们的相关缩写、UTC偏移和夏令时状态。（从技术上讲，PostgreSQL使用UT1而不是UTC，因为不处理闰秒。）与在pg_timezone_abbrevs中显示的缩写不同，这些名称中的许多暗示了一组夏令时转换日期规则。因此，相关信息在本地夏令时边界上发生变化。显示的信息是基于CURRENT_TIMESTAMP的当前值计算的。

表 45.68. pg_timezone_names 列

Name	Type	Description
name	text	Time zone name
abbrev	text	Time zone abbreviation
utc_offset	interval	Offset from UTC (positive means east of Greenwich)
is_dst	boolean	True if currently observing daylight savings

45.68. pg_user

视图pg_user提供数据库用户的信息。它本质上是 pg_shadow 的公开可读视图，并隐藏密码字段。

表 45.69. pg_user 列

Name	Type	Description
usename	name	User name
usesysid	int4	User ID (arbitrary number used to reference this user)
usecreatedb	bool	User can create databases
usesuper	bool	User is a superuser
usecatupd	bool	User can update system catalogs. (Even a superuser cannot

Name	Type	Description
		do this unless this column is true.)
passwd	text	Not the password (always reads as ********)
valuntil	abstime	Password expiry time (only used for password authentication)
useconfig	text[]	Session defaults for run-time configuration variables

45.69. pg_user_mappings

视图pg_user_mappings提供用户映射的信息。它本质上是 pg_user_mapping 的公开可读视图，但如果用户没有使用相应选项的权限，则会省略选项字段。

表 45.70. pg_user_mappings 列

Name	Type	References	Description
umid	oid	pg_user_mapping.oid	OID of the user mapping
srvid	oid	pg_foreign_server.oid	The OID of the foreign server that contains this mapping
srvname	text		Name of the foreign server
umuser	oid	pg_authid.oid	OID of the local role being mapped, 0 if the user mapping is public
usename	name		Name of the local user to be mapped
umoptions	text[]		User mapping specific options, as “keyword=value” strings, if the current user is the owner of the foreign server, else null

45.70. pg_views

视图pg_views提供了对数据库中每个视图的有用信息的访问。

表 45.71. pg_views 列

Name	Type	References	Description
schemaname	name	pg_namespace.nspname	Name of schema containing view
viewname	name	pg_class.relname	Name of view
viewowner	name	pg_authid.rolname	Name of view's owner

Name	Type	References	Description
`definition`	`text`		View definition (a reconstructed `SELECT` query)

第 46 章 前端/后端协议

PostgreSQL使用一种基于消息的协议在前端和后端（客户端与服务器）之间进行通信。
该协议既支持TCP/IP，也支持 Unix 域套接字。端口号 5432 已在 IANA 注册为支持该协议的服务器的惯用 TCP 端口号，但实际上任何非特权端口号都可以使用。

本文档描述协议 3.0 版，它在 PostgreSQL 7.4 及以后版本中实现。早期协议版本的描述参见 PostgreSQL 文档的以前发行版。单个服务器可以支持多个协议版本。最初的启动请求消息告诉服务器客户端试图使用哪个协议版本，然后服务器在有能力的情况下遵循该协议。

为了高效地为多个客户端提供服务，服务器会为每个客户端启动一个新的“后端”进程。
在当前实现中，一旦检测到传入连接，就会立刻创建新的子进程。不过，
这一点对协议而言是透明的。就协议而言，术语“后端”和“服务器”可以互换；同样，“前端”和“客户端”也可以互换。

46.1. 概述

协议分为启动和正常操作两个阶段。在启动阶段，前端打开到服务器的连接，
并完成服务器所要求的认证。（这可能只涉及一条消息，
也可能因所用认证方法不同而需要多条消息。）如果一切顺利，
服务器随后会向前端发送状态信息，并最终进入正常操作。除最初的启动请求消息外，
协议的这一部分由服务器驱动。

在正常操作中，前端向后端发送查询及其他命令，后端则返回查询结果和其他响应。
少数情况下（例如 `NOTIFY`），后端会发送未请求的消息，但会话中的绝大多数交互仍由前端请求驱动。

会话通常由前端选择终止，但在某些情况下也可能由后端强制终止。无论哪种情况，
后端关闭连接时，都会在退出前回滚所有打开的（未完成的）事务。

在正常操作中，SQL 命令可以通过两种子协议之一执行。在“简单查询”协议中，前端只需发送文本形式的查询字符串，后端会立即对其进行解析并执行。在“扩展查询”协议中，
查询处理被拆分为多个步骤：解析、参数值绑定以及执行。这带来了更高的灵活性和性能收益，
但代价是额外的复杂性。

正常操作还包含用于`COPY`等特殊操作的额外子协议。

46.1.1. 消息概述

所有通信都通过消息流进行。消息的第一个字节标识消息类型，
接下来的四个字节给出消息其余部分的长度（该长度计数包含自身，但不包括消息类型字节）。
消息剩余内容由消息类型决定。由于历史原因，客户端发送的第一条消息（启动消息）
没有开头的消息类型字节。

为了避免与消息流失去同步，服务器和客户端通常都会先根据字节计数把整条消息读入缓冲区，
然后再处理其内容。这样一来，如果在处理内容时检测到错误，就比较容易恢复。
在极端情况下（例如没有足够内存缓冲整条消息），
接收方也可以利用字节计数判断在恢复读取消息之前需要跳过多少输入。

反过来，服务器和客户端也必须注意绝不能发送不完整的消息。通常的做法是在开始发送之前，
先在缓冲区中整理好整条消息。如果在发送或接收消息的途中发生通信故障，
唯一合理的做法就是放弃连接，因为几乎不可能重新恢复消息边界的同步。

46.1.2. 扩展查询概述

在扩展查询协议中，SQL 命令的执行被划分为多个步骤。各步骤之间保留的状态由两类对象表示：预备语句和 portal。预备语句表示对文本查询字符串完成解析和语义分析（可选地还有规划）后的结果。预备语句本身还不一定能直接执行，因为它可能缺少特定参数的值。portal 表示一条已经可以执行、或已经部分执行过的语句，其中所有缺失的参数值都已补齐。（对于 `SELECT` 语句，portal 等价于一个打开的游标；但由于游标不能处理非 `SELECT` 语句，这里采用了不同的措辞。）

整个执行周期包括一个解析步骤，它从文本查询字符串创建预备语句；一个绑定步骤，
它根据预备语句和所需参数值创建 portal；以及一个执行步骤，用于执行 portal 中的查询。对于返回行的查询（`SELECT`、`SHOW`等），可以要求执行步骤只取回有限数量的行，
因此可能需要多次执行步骤才能完成整个操作。

后端可以跟踪多个预备语句和 portal（但请注意，它们只存在于单个会话内，绝不会在会话之间共享）。已有的预备语句和 portal 都通过创建时赋予的名称来引用。此外，还存在一个“未命名”的预备语句和 portal。虽然它们的行为与有名对象大体相同，但对未命名对象的操作是为“只执行一次然后丢弃”的场景优化的，
而对有名对象的操作则是基于会被多次使用的预期进行优化的。

46.1.3. 格式和格式代码

某一特定数据类型的数据可以使用多种不同的格式之一进行传输。自 PostgreSQL 7.4 起，当前只支持“文本”和“二进制”两种格式，但协议为未来扩展留出了空间。任意值所需的格式由
格式代码指定。客户端可以为每个传输的参数值以及查询结果的每一列指定格式代码。
文本格式的代码为零，二进制格式的代码为一，其他格式代码则保留供将来定义。

值的文本表示是相应数据类型的输入/输出转换函数生成和接受的字符串。在传输形式中，
值的末尾没有空字符；前端若要将收到的值作为 C 字符串处理，必须自行添加一个。（文本格式也不允许内嵌空字符。）

整数的二进制表示采用网络字节序（最高有效字节在前）。至于其他数据类型，
请查阅文档或源代码了解其二进制表示形式。要注意，
复杂数据类型的二进制表示可能会在不同服务器版本之间发生变化；
文本格式通常是可移植性更好的选择。

46.2. 消息流

本节描述消息流以及各种消息类型的语义（每种消息的精确格式见第 46.5 节）。根据连接所处的状态不同，存在若干不同的子协议：启动、查询、函数调用、`COPY`以及终止。
对于异步操作（包括通知响应和命令取消）还有专门规定，
它们可能在启动阶段结束后的任何时刻发生。

46.2.1. 启动

要开始一个会话，前端会打开到服务器的连接并发送一条启动消息。
该消息包含用户名以及用户希望连接的数据库名；它还指明要使用的协议版本。
（启动消息也可以选择性地包含运行时参数的附加设置。）
随后服务器会结合这些信息以及配置文件（例如 `pg_hba.conf`）的内容，初步判断是否接受该连接，以及需要何种额外认证（如果需要）。

接着服务器会发送适当的认证请求消息，
前端必须以适当的认证响应消息进行回应（例如密码）。对于除 GSSAPI 和 SSPI 之外的认证方

法，最多只会有一次请求和一次响应。在某些方法中，前端根本不需要发送响应，
因此也不会出现认证请求。对于 GSSAPI 和 SSPI，则可能需要多轮报文交换才能完成认证。

认证周期要么以服务器拒绝连接（ErrorResponse）结束，要么以 AuthenticationOk 结束。

该阶段服务器可能发送的消息有：

ErrorResponse

连接尝试被拒绝。随后服务器立即关闭连接。

AuthenticationOk

认证交换成功完成。

AuthenticationKerberosV5

前端现在必须参与与服务器之间的 Kerberos V5 认证对话（此处不再描述，它属于 Kerberos 规范的一部分）。如果对话成功，服务器会响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationCleartextPassword

前端现在必须发送一个以明文形式包含密码的 PasswordMessage。如果密码正确，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationMD5Password

前端现在必须发送一个 PasswordMessage，其中包含使用 AuthenticationMD5Password 消息中指定的 4 字符盐经过 MD5 加密的密码。如果密码正确，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationSCMCredential

此响应仅适用于支持 SCM 凭据消息的平台上的本地 Unix 域连接。前端必须发出一条 SCM 凭据消息，然后发送一个单字节数据。（该数据字节的内容无关紧要；它仅用于确保服务器等待足够长的时间来接收凭据消息。）如果凭据可接受，服务器响应 AuthenticationOk；否则响应 ErrorResponse。（该消息类型仅由 9.1 之前的服务器发出。它最终可能会从协议规范中移除。）

AuthenticationGSS

前端现在必须发起一次 GSSAPI 协商。前端将发送一条带有 GSSAPI 数据流第一部分的 PasswordMessage 消息作为响应。如果还需要进一步的消息，服务器会响应 AuthenticationGSSContinue。

AuthenticationSSPI

前端现在必须发起一次 SSPI 协商。前端将发送一条带有 SSPI 数据流第一部分的 PasswordMessage 作为响应。如果还需要进一步的消息，服务器会响应 AuthenticationGSSContinue。

AuthenticationGSSContinue

该消息包含前一步 GSSAPI 或 SSPI 协商（AuthenticationGSS、AuthenticationSSPI 或前一个 AuthenticationGSSContinue）的响应数据。如果该消息中的 GSSAPI 或 SSPI 数据表明完成认证还需要更多数据，前端必须把所需数据作为另一条 PasswordMessage 消息发送。如果该消息已完成 GSSAPI 或 SSPI 认证，服务器接下来会发送 AuthenticationOk 表示认证成功，或发送 ErrorResponse 表示认证失败。

如果前端不支持服务器要求的认证方式，那么它应该马上关闭连接。

在收到 AuthenticationOk 消息之后，前端必须继续等待来自服务器的后续消息。在这个阶段，一个后端进程正在启动，而前端只是旁观者。启动尝试仍可能失败（ErrorResponse），但通常情况下，后端会发送一些 ParameterStatus 消息、BackendKeyData，以及最后的 ReadyForQuery。

在这个阶段，后端会尝试应用启动消息中给出的任何额外运行时参数设置。如果成功，这些值就会成为会话的默认值。发生错误则会导致 ErrorResponse 并退出。

这个阶段来自后端的可能消息是：

BackendKeyData

该消息提供密钥数据。如果前端希望稍后发送取消请求，就必须保存这些数据。前端不应响应该消息，而应继续等待 ReadyForQuery 消息。

ParameterStatus

该消息告知前端后端参数的当前（初始）设置，例如client_encoding或DateStyle。前端可以忽略这些信息，也可以记录下来供后续使用；详见第 46.2.6 节。前端不应响应该消息，而应继续等待 ReadyForQuery 消息。

ReadyForQuery

启动成功，前端现在可以发出命令。

ErrorResponse

启动失败，在发送完这个消息之后连接被关闭。

NoticeResponse

已发出一条警告消息。前端应显示该消息，但仍应继续等待 ReadyForQuery 或 ErrorResponse。

ReadyForQuery 消息与后端在每个命令周期结束后发出的消息是同一个。前端可根据自身编码需要，将 ReadyForQuery 视为一个命令周期的开始，或视为启动阶段以及后续每个命令周期的结束。

46.2.2. 简单查询

一个简单查询周期由前端向后端发送一条 Query 消息来启动。该消息包含一条或多条以文本字符串表示的 SQL 命令。后端随后会根据查询命令串的内容向前端发送一条或多条响应消息，最后再发送一条 ReadyForQuery 响应消息。ReadyForQuery 通知前端它现在可以安全地发送新的命令。（实际上，前端并不一定非要等到收到 ReadyForQuery 才发送下一条命令，但这样一来，前端就必须自己处理“先前命令失败而后续已发出的命令成功”这种情况。）

后端可能返回的响应消息有：

CommandComplete

一条 SQL 命令正常完成。

CopyInResponse

后端已准备好把数据从前端复制到表中；参见第 46.2.5 节。

CopyOutResponse

后端已准备好把数据从表中复制到前端；参见第 46.2.5 节。

RowDescription

表示即将返回行作为对SELECT、FETCH等查询的响应。此消息的内容描述了行的列布局。该消息之后，每个返回给前端的行都对应一条 DataRow 消息。

DataRow

由SELECT、FETCH等查询返回的一组行中的一个。

EmptyQueryResponse

识别出了一条空查询字符串。

ErrorResponse

发生了一个错误。

ReadyForQuery

查询字符串的处理已完成。发送一个单独的消息来指示这一点，因为查询字符串可能包含多个SQL命令。（Command Complete标记了一个SQL命令的处理结束，而不是整个字符串的结束。）无论处理是成功还是出现错误，都将始终发送ReadyForQuery。

NoticeResponse

与查询相关的警告消息已发出。通知是其他响应的补充，即后端将继续处理命令。

SELECT 查询（或其他返回行集的查询，如 EXPLAIN 或 SHOW）的响应通常包含 Row Description、零条或多条 DataRow 消息，以及最后的 CommandComplete。在前端与服务器之间执行 COPY 输入或输出时，会使用第 46.2.5 节所述的特殊协议。所有其他类型的查询通常只产生一条 CommandComplete 消息。

由于查询字符串可能包含若干条查询（以分号分隔），因此在后端完成整个查询字符串的处理之前，可能会出现多个这样的响应序列。只有在整个字符串处理完毕且后端已准备好接受新的查询字符串时，才会发出 ReadyForQuery 消息。

如果收到的是一条完全为空的查询字符串（除了空白字符之外没有任何内容），响应就是一条 EmptyQueryResponse，后面跟着 ReadyForQuery。

一旦发生错误，就会发出一条 ErrorResponse 消息，后面跟着 ReadyForQuery。Error Response 会中止该查询字符串中后续所有处理（即使其中还包含其他查询）。请注意，这种情况可能发生在处理单条查询所生成的消息序列中途。

在简单 Query 模式下，取回值的格式总是文本，除非给定命令是从一个用 BINARY 选项声明的游标进行的 FETCH。那种情况下，取回的值为二进制格式。RowDescription 消息中给出的格式代码指明使用的是哪种格式。

当前端正在等待其他类型的消息时，也必须准备好接收 ErrorResponse 和 NoticeResponse 消息。参见第 46.2.6 节，了解后端因外部事件而可能生成的消息。

建议以状态机的方式编写前端，使其能够在任何合理的时机接收相应类型的消息，而不把消息确切顺序的假设写死在代码中。

46.2.3. 扩展查询

扩展查询协议把上文描述的简单查询协议拆分成多个步骤。准备步骤的结果可以重复使用，从而提高效率。此外，它还提供了额外特性，例如可以把数据值作为独立参数提供，而不必直接插入查询字符串中。

在扩展协议中，前端首先发送一条 Parse 消息，其中包含文本查询字符串、可选的参数占位符数据类型信息，以及目标预备语句对象的名称（空字符串表示未命名预备语句）。响应要么是 ParseComplete，要么是 ErrorResponse。参数数据类型可以用 OID 指定；如果未给出，解析器会像处理无类型字面字符串常量那样尝试推断其数据类型。

> **注意**
>
> 一个参数的数据类型可以通过设为零，或让参数类型 OID 数组短于查询字符串中参数符号（`$n`）的数量来保持未指定。另一个特例是，参数类型可以指定为 `void`（即伪类型 `void` 的 OID）。这样做是为了允许参数符号用于那些实际上是 OUT 参数的函数参数。通常并不存在可使用 `void` 参数的上下文，但如果这样的参数符号出现在函数参数列表中，它实际上会被忽略。例如，像 `foo($1,$2,$3,$4)` 这样的函数调用，如果 `$3` 和 `$4` 被指定为 `void` 类型，就可能匹配一个带有两个 IN 参数和两个 OUT 参数的函数。

> **注意**
>
> Parse 消息中的查询字符串不能包含多于一条 SQL 语句；否则会报告语法错误。这个限制在简单查询协议中并不存在，但在扩展协议中必须如此，因为若允许预备语句或 portal 包含多条命令，会使协议变得过于复杂。

If successfully created, a named prepared-statement object lasts till 当前会话结束，除非被显式销毁。未命名预备语句只会持续到下一条把未命名语句作为目标的 Parse 语句发出为止。（注意，简单 Query 消息也会销毁未命名语句。）有名预备语句在被另一条 Parse 消息重新定义之前必须显式关闭，但未命名语句则无此要求。有名预备语句也可以在 SQL command level, using `PREPARE` and `EXECUTE`.

一旦预备语句存在，就可以用 Bind 消息把它准备为可执行状态。Bind 消息给出源预备语句的名称（空字符串表示未命名预备语句）、目标 portal 的名称（空字符串表示未命名 portal），以及预备语句中所有参数占位符应使用的值。所提供的参数集必须与预备语句所需参数相匹配。（如果你在 Parse 消息中声明了任何 `void` 参数，那么在 Bind 消息中应为它们传递 NULL 值。）Bind 还会指定查询返回数据所使用的格式；格式既可以统一指定，也可以按列指定。响应要么是 BindComplete，要么是 ErrorResponse。

> **注意**
>
> 输出采用文本还是二进制格式，由 Bind 中给出的格式代码决定，而与涉及的 SQL 命令无关。在使用扩展查询协议时，游标声明中的 `BINARY` 属性并不起作用。

对命名预备语句对象的查询规划发生在处理 Parse 消息时。如果一条查询将用不同参数反复执行，那么发送一条包含参数化查询的 Parse 消息，随后再发送多条 Bind 和 Execute 消息可能是有益的。这样可以避免在每次执行时重新规划查询。

如果 Parse 消息未定义参数，未命名预备语句同样在 Parse 处理期间完成规划。但如果存在参数，查询规划会在每次提供 Bind 参数时进行。这使得规划器可以利用每条 Bind 消息提供的实际参数值，而不是使用一般性估计。

> **注意**
>
> 从参数化查询生成的查询计划，
> 可能不如用实际参数值替换后的等价查询生成的计划高效。
> 在为分配给命名预备语句对象的参数化查询做规划时，
> 查询规划器无法基于实际参数值（例如索引选择性）做决策。
> 使用未命名语句可以避免这种可能的性能损失，
> 因为直到有了实际参数值之后才会进行规划。代价是每次 Bind 都必须重新规划，即使查询保持不变。

如果成功创建，命名 portal 对象会一直持续到当前事务结束，除非被显式销毁。未命名 portal 会在事务结束时销毁，或者在下一条把未命名 portal 作为目标的 Bind 语句发出时立即销毁。（注意，简单 Query 消息也会销毁未命名 portal。）有名 portal 在被另一条 Bind 消息重新定义之前必须显式关闭，而未命名 portal 则无此要求。有名 portal 也可以在 SQL command level, using `DECLARE CURSOR` and `FETCH`.

一旦 portal 存在，就可以使用 Execute 消息执行它。Execute 消息指定 portal 的名称（空字符串表示未命名 portal）以及一个最大的结果行计数（零表示“取回全部行”）。结果行计数只对包含返回行集命令的 portal 有意义；在其他情况下，命令总会执行到完成，而行计数会被忽略。Execute 的可能响应与通过简单查询协议发出的查询相同，只是 Execute 不会导致后端发送 ReadyForQuery 或 RowDescription。

如果 Execute 在 portal 执行完成之前终止（因为达到了非零的结果行计数），它会发送一条 PortalSuspended 消息；该消息表明前端应当针对同一个 portal 再发出一条 Execute 消息，以完成此次操作。在 portal 执行完成之前，不会发送表示源 SQL 命令结束的 Command Complete 消息。因此，一个 Execute 阶段总会以下列消息中的恰好一条结束：Command Complete、EmptyQueryResponse（如果 portal 是从空查询字符串创建的）、ErrorResponse 或 PortalSuspended。

每一组扩展查询消息完成后，前端都应发送一条 Sync 消息。这条无参数消息会让后端关闭当前事务，如果当前事务并不处在 `BEGIN/COMMIT` 事务块内（这里的“关闭”指的是：无错误则提交，有错误则回滚）。随后会发送一条 ReadyForQuery 响应。Sync 的目的是为错误恢复提供一个重新同步点。如果在处理任何扩展查询消息时检测到错误，后端会发出 ErrorResponse，然后持续读取并丢弃消息，直到遇到 Sync，再发送 ReadyForQuery 并回到正常的消息处理流程。（但要注意，如果是在处理 Sync 期间检测到错误，则不会跳过任何消息，这样就能保证每条 Sync 恰好对应一条 ReadyForQuery。）

> **注意**
>
> Sync 不会关闭由`BEGIN`打开的事务块。之所以能够检测出这种情况，是因为 ReadyForQuery 消息中包含事务状态信息。

除了这些基本的、必需的操作之外，在扩展查询协议里还有几种可选的操作可以使用。

Describe 消息（portal 变体）指定一个现有 portal 的名称（或者用空字符串表示未命名 portal）。响应要么是一条 RowDescription 消息，用于描述执行该 portal 时将返回的行；要么是一条 NoData 消息，如果该 portal 不包含会返回行的查询；要么是 ErrorResponse，如果不存在这样的 portal。

Describe 消息（语句变体）指定一个现有预备语句的名称（或者用空字符串表示未命名预备语句）。响应是一条 ParameterDescription 消息，用于描述该语句所需的参数，随后是一条 RowDescription 消息，用于描述该语句最终执行时将返回的行（如果该语句不返回行，则为 NoData 消息）。如果不存在这样的预备语句，则返回 ErrorResponse。请注意，由于还没有发出 Bind，后端尚不知道返回列将使用什么格式；因此在这种情况下，RowDescription 消息中的格式代码字段将为零。

> **提示**
>
> 在大多数场景下，前端都应在发出 Execute 之前先发送某一种 Describe 变体，以确保它知道应如何解释接收到的结果。

Close 消息会关闭一个现有的预备语句或 portal，并释放相关资源。针对不存在的语句或 portal 发出 Close 并不算错误。响应通常是 CloseComplete，但如果在释放资源时遇到困难，也可能返回 ErrorResponse。请注意，关闭预备语句会隐式关闭所有基于该语句构造出的打开 portal。

Flush 消息本身不会产生任何特定输出，但会强制后端发送其输出缓冲区中所有尚待发送的数据。如果前端希望在发出更多命令之前检查某条扩展查询命令的结果，那么除了 Sync 之外，任何扩展查询命令之后都必须发送 Flush。如果没有 Flush，后端返回的消息会尽量合并成最少的数据包，以降低网络开销。

> **注意**
>
> 简单 Query 消息大致等价于一串 Parse、Bind、portal Describe、Execute、Close、Sync 操作，它们使用未命名预备语句和未命名 portal 对象，并且不带参数。不同之处在于，简单 Query 会接受包含多条 SQL 语句的查询字符串，并依次自动为每条语句执行绑定、描述和执行序列；另一个区别是它不会返回 ParseComplete、BindComplete、CloseComplete 或 NoData 消息。

46.2.4. 函数调用

函数调用子协议允许客户端请求直接调用数据库`pg_proc`系统目录中的任意函数。
客户端必须具有该函数的执行权限。

> **注意**
>
> 函数调用子协议是一个遗留的特性，在新代码里可能最好避免用它。
> 类似的结果可以通过设置一个执行`SELECT function($1, ...)`的预备语句得到。这样函数调用周期就可以用 Bind/Execute 代替。

函数调用周期由前端向后端发送一条 FunctionCall 消息来启动。后端随后根据函数调用的结果发送一条或多条响应消息，最后发送一条 ReadyForQuery 响应消息。ReadyForQuery 告知前端，可以安全地发送新的查询或函数调用。

来自后端的可能的响应消息是：

ErrorResponse

发生了一个错误。

FunctionCallResponse

函数调用完成并且在消息中返回一个结果（请注意函数调用协议只能处理单个标量结果，不能处理行类型或者结果集合）。

ReadyForQuery

函数调用处理完成。ReadyForQuery将总是被发送，不管是成功完成处理还是发生一个错误。

NoticeResponse

发出了一条有关该函数调用的警告信息。通知是附加在其他响应上的，也就是说，后端将继续处理该命令。

46.2.5. COPY操作

COPY命令允许在服务器和客户端之间进行高速大批量数据传输。拷贝入和拷贝出操作每个都把连接切换到一个独立的子协议中，并且持续到操作结束。

拷贝入模式（向服务器传输数据）在后端执行COPY FROM STDIN SQL 语句时启动。后端会向前端发送一条 CopyInResponse 消息。随后前端应发送零条或多条 CopyData 消息，构成一条输入数据流。（消息边界与行边界之间没有任何对应关系要求，尽管让它们对齐通常是合理的选择。）前端可以通过发送 CopyDone 消息来结束拷贝入模式（允许成功结束），也可以发送 CopyFail 消息（这会使COPY语句以错误失败）。然后后端会恢复到COPY开始之前的命令处理模式，也就是简单查询协议或扩展查询协议。接下来它会发送 CommandComplete（成功时）或 ErrorResponse（失败时）。

如果在拷贝入模式期间后端检测到错误（包括收到 CopyFail 消息），后端会发出一条 ErrorResponse 消息。如果COPY命令是通过扩展查询消息发出的，那么后端会从此开始丢弃前端消息，直到收到一条 Sync 消息，然后发出 ReadyForQuery 并恢复正常处理。如果COPY命令是在简单 Query 消息中发出的，那么该消息的剩余部分会被丢弃，并发送 ReadyForQuery。无论哪种情况，前端随后发出的任何 CopyData、CopyDone 或 CopyFail 消息都会被直接丢弃。

后端会忽略在拷贝入模式期间收到的 Flush 和 Sync 消息。收到任何其他非拷贝类型的消息都会构成错误，并按上述方式中止拷贝入状态。（Flush 和 Sync 的例外是为了方便那些总是在 Execute 消息之后发送 Flush 或 Sync、而不检查待执行命令是否为COPY FROM STDIN的客户端库。）

拷贝出模式（数据从服务器发出）是在后端执行一个COPY TO STDOUT语句时启动的。后端发出一个CopyOutResponse消息给前端，后面跟着零或者多个CopyData消息（总是每行一个），然后跟着CopyDone。然后后端回退到它在COPY开始之前的命令处理模式，然后发送CommandComplete。前端不能中止传输（除非是关闭连接或者发出一个Cancel请求），但是它可以抛弃不需要的CopyData和CopyDone消息。

在拷贝出模式中，如果后端检测到错误，那么它将发出一个ErrorResponse消息并且回到正常的处理。前端应该把收到ErrorResponse当作终止拷贝出模式的标志。

NoticeResponse 和 ParameterStatus 消息可能穿插在 CopyData 消息之间；前端必须处理这些情况，并应准备好处理其他异步消息类型（参见第 46.2.6 节）。除此之外，可以将任何除 CopyData 或 CopyDone 以外的消息类型视为拷贝出模式的终止标志。

还有一种与 COPY 相关的模式称为 Copy-both，它允许向服务器以及从服务器进行高速批量数据传输。当处于 WAL 发送进程模式的后端执行一条 START_REPLICATION 语句时，就会启动 Copy-both 模式。后端向前端发送一条 CopyBothResponse 消息，之后后端和前端都可以发送 CopyData 消息，直到连接终止。见第 46.3 节。

CopyInResponse、CopyOutResponse 和 CopyBothResponse 消息包含一些字段，用于告知前端每行的列数以及每列所使用的格式代码。（在当前实现中，同一次COPY操作的所有列都使用相同格式，但消息设计并不作此假设。）

46.2.6. 异步操作

在若干情况下，后端会发送并非由前端命令流直接触发的消息。前端必须随时准备处理这些消息，即使当前并未处于查询过程中。至少，在开始读取查询响应之前应检查这些情况。

NoticeResponse 消息可能因外部活动而产生；例如，如果数据库管理员发起一次“快速”数据库关闭，后端会在关闭连接之前发送一条 NoticeResponse 说明这一事实。因此，前端应始终准备好接收并显示 NoticeResponse 消息，即使连接表面上处于空闲状态。

只要后端认为前端应当知晓的某个参数的当前有效值发生变化，就会生成 ParameterStatus 消息。最常见的情况是响应前端执行的SET命令，这种情况实际上是同步的；但也可能是管理员修改了配置文件，然后向服务器发送SIGHUP信号，从而导致参数状态发生变化。同样，如果某条SET命令被回滚，也会生成适当的 ParameterStatus 消息，用于报告当前生效的值。

目前，会为一组固定的参数生成 ParameterStatus，参数如下：server_version、server_encoding、client_encoding、application_name、is_superuser、session_authorization、DateStyle、IntervalStyle、TimeZone、integer_datetimes、standard_conforming_strings。（8.0 之前的版本不报告 server_encoding、TimeZone 和 integer_datetimes；8.1 之前的版本不报告 standard_conforming_strings；8.4 之前的版本不报告 IntervalStyle；9.0 之前的版本不报告 application_name。）注意，server_version、server_encoding 和 integer_datetimes 是启动后不能改变的伪参数。这组参数将来可能变化，甚至可能变为可配置。因此，前端应直接忽略其不理解或不关心的参数的 ParameterStatus。

如果前端发出LISTEN命令，那么每当针对同一通道名执行NOTIFY命令时，后端都会发送一条 NotificationResponse 消息（不要与 NoticeResponse 混淆）。

> **注意**
>
> 目前，NotificationResponse只能在一个事务外面发送，因此它将不会在一个命令响应序列中间出现，但是它可能正好在ReadyForQuery之前出现。不过，在前端逻辑中做上述假设是不明智的。好的做法是在协议的任何点上都可以接受NotificationResponse。

46.2.7. 取消正在处理的请求

在一条查询正在处理的时候，前端可以请求取消该查询。这种取消请求不是直接通过打开的连接发送给后端的，这么做是因为实现的效率：我们不希望后端在处理查询的过程中不停地检查前端来的输入。取消请求应该相对而言比较少见，所以我们把取消做得稍微笨拙一些，以便不影响正常状况的性能。

要发出取消请求，前端会新建到服务器的连接，并发送 CancelRequest 消息，而不是新连接通常发送的 StartupMessage 消息。服务器处理该请求后便会关闭连接。出于安全原因，服务器不会直接回复取消请求消息。

除非CancelRequest消息包含在连接启动过程中传递给前端的相同的密钥数据（PID 和密钥），否则它将被忽略。如果该请求匹配当前运行着的后端的PID和密钥，
则中止当前查询的处理（目前的实现里采用的方法是向正在处理该查询的后端进程发送一个特殊的信号）。

取消信号可能有效，也可能无效；例如，如果它在后端已经处理完查询之后才到达，
就不会起作用。如果取消生效，当前命令就会以一条错误消息提前终止。

这么做是对安全性和效率通盘考虑的结果，前端没有直接的方法获知一个取消请求是否成功。
它必须继续等待后端对查询的响应。发出一个取消仅仅是增加了当前查询快些结束的可能性，
同时也增加了当前查询会伴随着一条错误消息失败而不是成功执行的可能性。

由于取消请求是通过一条新的连接发送给服务器，而不是通过常规的前端/后端通信链路发送，
因此发出取消请求的可以是任意进程，而不一定非要是要取消查询的那个前端。
这为构建多进程应用提供了额外的灵活性，同时也带来了安全风险，
因为未授权用户可能会尝试取消查询。通过要求在取消请求中提供动态生成的密钥，
可以缓解这一安全风险。

46.2.8. 终止

通常优雅的终止过程是前端发送一条Terminate消息并且立刻关闭连接。一旦收到消息，
后端马上关闭连接并且终止。

在少数情况下（比如一个管理员命令数据库关闭），
后端可能在没有任何前端请求的情况下断开连接。在这种情况下，
后端将在它断开连接之前尝试发送一个错误或者通知消息给出断开的原因。

其他终止场景来自各种故障，例如任一端发生 core dump、通信链路中断、消息边界同步丢失等。如果前端或后端看到连接意外关闭，就应清理并终止。若前端不想自行终止，
也可以重新联系服务器以启动一个新的后端。如果收到无法识别的消息类型，
同样建议关闭连接，因为这通常意味着消息边界同步已经丢失。

不管是正常还是不正常的终止，任何打开的事务都会回滚而不是提交。不过，
我们应该注意的是如果一个前端在一个非`SELECT`查询正在处理的时候断开，
那么后端很可能在发现断开之前先完成查询的处理。如果查询处于任何事务块之外（`BEGIN` ... `COMMIT`序列），那么其结果可能在发现连接断开之前被提交。

46.2.9. SSL会话加密

If PostgreSQL was built with SSL support, frontend/backend communications can be encrypted using SSL. This provides 在攻击者可能能够捕获会话流量的环境中提供通信安全。关于用 SSL 加密 PostgreSQL 会话的更多信息，参见第 17.9 节。

要发起 SSL 加密连接，前端首先发送 SSLRequest 消息，而不是 StartupMessage。服务器随后返回包含 `S` 或 `N` 的单个字节，分别表示愿意或不愿意进行 SSL 通信。如果前端不满意此响应，可以在此时关闭连接。要在收到 `S` 后继续，应与服务器执行 SSL 启动握手（这里不作说明，它属于 SSL 规范的一部分）。如果握手成功，则继续发送通常的 StartupMessage。此时，Startup Message 及其后的所有数据都将经过 SSL 加密。要在收到 `N` 后继续，则发送通常的 Startup Message，并以不加密的方式继续通信。

前端还应准备好处理服务器对 SSLRequest 的 ErrorMessage 响应。这种情况只会发生在服务器版本早于 PostgreSQL 加入 SSL 支持的时候。此时必须关闭连接，但前端可以选择打开一个新连接并继续，而不请求 SSL。

如果建立连接是为了发送 CancelRequest 消息，也可以先发送 SSLRequest。

虽然协议本身没有提供让服务器强制使用SSL加密的方法，但管理员可以配置服务器，使其在认证检查中拒绝未加密的会话。

46.3. 流复制协议

要发起流式复制，前端在启动消息中发送 `replication` 参数。这告知后端进入 WAL 发送进程模式，在该模式下可以发出一小组复制命令，而不是 SQL 语句。WAL 发送进程模式只能使用简单查询协议。WAL 发送进程模式接受以下命令：

IDENTIFY_SYSTEM

请求服务器标识自身。服务器返回一个只有一行的结果集，包含三个字段：

systemid

标识集簇的唯一系统标识符。
可用于检查初始化备库所用的基础备份是否来自同一个集簇。

timeline

当前 TimelineID。也可用于检查备库是否与主库一致。

xlogpos

当前 xlog 刷盘位置。可用于获取事务日志中一个已知的位置，以便从该处开始流式传输。

START_REPLICATION *XXX*/*XXX*

指示服务器开始流式传输 WAL，从 WAL 位置 *XXX*/*XXX* 开始。服务器可能回复错误，例如所请求的 WAL 段已被回收。成功时，服务器会回复一条 CopyBothResponse 消息，然后开始向前端流式传输 WAL。WAL 将持续流式传输，直到连接断开为止；不会再接受其他命令。

WAL 数据以一系列 CopyData 消息发送。（这样可以混合发送其他信息；尤其是服务器在开始流式传输后遇到故障时，可以发送 ErrorResponse 消息。）每条 CopyData 消息的有效载荷遵循以下格式：

XLogData (B)

Byte1('w')

标识此消息为 WAL 数据。

Byte8

本消息中 WAL 数据的起始点，以 XLogRecPtr 格式给出。

Byte8

服务器上 WAL 的当前末尾，以 XLogRecPtr 格式给出。

Byte8

传输时服务器的系统时钟，以 TimestampTz 格式给出。

Byte*n*

WAL 数据流的一段。

单条 WAL 记录绝不会拆分到两条 CopyData 消息中。当一条 WAL 记录跨越 WAL 页边界（因而已经用继续记录拆分）时，它可以在页边界处拆分。换句话说，第一条主 WAL 记录与它的继续记录可以在不同的 CopyData 消息中发送。

注意，WAL 数据内部以及上面描述的头部中的所有字段，都采用发送服务器的本机格式。字节序和时间戳的格式是不可预测的，除非接收方已验证发送方的系统标识符与自己 `pg_control` 中的内容一致。

如果 WAL 发送进程正常终止（postmaster 关闭期间），它会在退出前发送一条 Command Complete 消息。当然，异常关闭时可能不会发生这种情况。

接收进程可以随时使用以下消息格式之一回复发送端（同样放在 CopyData 消息的有效载荷中）：

备库状态更新 (F)

Byte1('r')

将该消息标识为接收端状态更新。

Byte8

备库上已接收并写入磁盘的最后一个 WAL 字节 + 1 的位置，以 XLogRecPtr 格式给出。

Byte8

备库上已刷出到磁盘的最后一个 WAL 字节 + 1 的位置，以 XLogRecPtr 格式给出。

Byte8

备库上已应用的最后一个 WAL 字节 + 1 的位置，以 XLogRecPtr 格式给出。

Byte8

传输时服务器的系统时钟，以 TimestampTz 格式给出。

热备反馈消息 (F)

Byte1('h')

将该消息标识为热备反馈消息。

Byte8

传输时服务器的系统时钟，以 TimestampTz 格式给出。

Byte4

备库当前的 xmin。如果备库发送通知说此连接将不再发送热备反馈，则该值可能为 0。之后的非零消息可以重新启动反馈机制。

Byte4

备库当前的纪元。

BASE_BACKUP [`LABEL '`*`label`*`'`] [`PROGRESS`] [`FAST`] [`WAL`] [`NOWAIT`]

指示服务器开始流式传输基础备份。在备份开始之前，系统将自动进入备份模式，并在备份完成后退出备份模式。接受以下选项：

`LABEL 'label'`

设置备份的标签。如果未指定，则将使用`base backup`作为备份标签。标签的引号使用规则与打开standard_conforming_strings的标准SQL字符串相同。

`PROGRESS`

请求生成进度报告所需的信息。这将在每个表空间的首部发送一个近似大小，可用于计算流式传输的进度。这是通过在传输开始之前先枚举所有文件大小来计算的，可能会对性能产生负面影响。特别是，在流式传输数据之前可能需要更长的时间。由于备份期间数据库文件可能会发生变化，因此大小仅为近似值，在估算与实际发送文件之间的这段时间里可能会增长或缩小。

`FAST`

请求快速检查点。

`WAL`

在备份中包含必要的 WAL 段。这会把开始备份到停止备份之间的所有文件放入基础目录 tar 文件内的`pg_xlog`目录中。

`NOWAIT`

默认情况下，备份会等待最后一个必需的 WAL 段完成归档；如果未启用日志归档，则发出警告。指定 `NOWAIT` 会同时禁用等待和警告，由客户端负责确保所需日志可用。

备份开始时，服务器首先发送两个普通结果集，然后发送一个或多个 CopyResponse 结果。

第一个普通结果集包含备份的起始位置，以 XLogRecPtr 格式作为单行单列给出。

第二个普通结果集中的每个表空间都有一行。这一行中的字段是：

spcoid

表空间的 OID，如果是基础目录则为 `NULL`。

spclocation

表空间目录的完整路径，如果是基础目录则为 `NULL`。

size

如果请求了进度报告，则为表空间的大致大小；否则为 `NULL`。

第二个普通结果集之后，将发送一个或多个 CopyResponse 结果，PGDATA 一个，除 `pg_default` 和 `pg_global` 之外的每个表空间各一个。CopyResponse 结果中的数据是表空间内容的 tar 格式（遵循 POSIX 1003.1-2008 标准规定的“ustar 交换格式”）转储，但省略了标准中规定的结尾两个全零块。tar 数据完成后，会发送最后一个普通结果集。

数据目录和每个表空间的tar归档将包含目录中的所有文件，无论它们是PostgreSQL文件还是添加到同一目录的其他文件。唯一排除的文件是：

- `postmaster.pid`
- `postmaster.opts`
- `pg_xlog`，包括子目录。如果备份时包含了 WAL 文件，则会包含一个合成的 pg_xlog 版本，但其中只包含使备份能够工作所必需的文件，而不包含其余内容。

如果服务器上的底层文件系统支持，则会设置属主、组和文件模式。

所有表空间发送完毕后，会发送最后一个常规结果集。该结果集包含备份的结束位置，以 XLogRecPtr 格式作为单行单列给出。

46.4. 消息数据类型

本节描述消息中使用的基础数据类型。

Int*n*(*i*)

以网络字节序（最高有效字节在前）表示的 *n* 位整数。如果指定了 *i*，则表示其精确值；否则表示该值可变。例如：Int16、Int32(42)。

Int*n*[*k*]

由 *k* 个 *n* 位整数组成的数组，每个整数都按网络字节序排列。数组长度 *k* 总是由消息中更早的某个字段确定。例如：Int16[M]。

String(*s*)

一个以空字符结尾的字符串（C 风格字符串）。字符串没有特定的长度限制。如果指定了 *s*，则表示其精确值；否则表示该值可变。例如：String、String("user")。

> **注意**
>
> 后端返回的字符串长度没有预定义的上限。前端的好做法是使用可扩展的缓冲区，这样凡是能装进内存的内容都能接受。如果做不到这一点，就读取整个字符串，然后丢弃固定大小缓冲区装不下的尾部字符。

Byte*n*(*c*)

精确 *n* 字节。如果字段宽度 *n* 不是常数，它总是可以由消息中更早的字段确定。如果指定了 *c*，则表示其精确值。例如：Byte2、Byte1('\n')。

46.5. 消息格式

本节描述每条消息的详细格式。每条消息都标记了可由前端（F）、后端（B）或双方（F&B）发送。请注意，虽然每条消息开头都带有字节计数，
但大多数消息格式都定义为无需参考该计数也能确定消息边界。
这一设计最初是出于历史原因（早期已废弃的协议 v2 没有显式长度字段），同时也有助于有效性校验。

AuthenticationOk (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(0)

表示认证成功。

AuthenticationKerberosV5 (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(2)

表示需要 Kerberos V5 认证。

AuthenticationCleartextPassword (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(3)

表示需要明文密码。

AuthenticationMD5Password (B)

Byte1('R')

将该消息标识为认证请求。

Int32(12)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(5)

表示需要经过 MD5 加密的密码。

Byte4

加密密码时使用的盐。

AuthenticationSCMCredential (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(6)

表示需要 SCM 凭证消息。

AuthenticationGSS (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(7)

表示需要 GSSAPI 认证。

AuthenticationSSPI (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(9)

表示需要 SSPI 认证。

AuthenticationGSSContinue (B)

Byte1('R')

将该消息标识为认证请求。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(8)

表示此消息包含 GSSAPI 或 SSPI 数据。

Byte$n$

GSSAPI 或 SSPI 认证数据。

BackendKeyData (B)

Byte1('K')

将此消息标识为取消请求密钥数据。如果前端希望以后能够发送 CancelRequest 消息，就必须保存这些值。

Int32(12)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

此后端的进程 ID。

Int32

此后端的密钥。

Bind (F)

Byte1('B')

将该消息标识为 Bind 命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

目标 portal 的名称（空字符串选择未命名的 portal）。

String

源预备语句的名称（空字符串选择未命名的预备语句）。

Int16

后续参数格式码的数量（下文以 *C* 表示）。可以为零，表示没有参数，或者所有参数都使用默认格式（文本）；也可以为一，此时指定的格式码应用于所有参数；
还可以等于实际参数数量。

Int16[*C*]

参数格式码。目前每个格式码必须为零（文本）或一（二进制）。

Int16

后续参数值的数量（可以为零）。必须与查询所需的参数数量一致。

接下来，每个参数都有以下一对字段：

Int32

参数值的长度，以字节为单位（不包括此长度字段本身）。可以为零。特殊值 -1 表示 NULL 参数值。为 NULL 时，后面不再有值的字节。

Byte*n*

参数值，格式由对应的格式码指明。*n* 为上述长度。

最后一个参数之后是以下字段：

Int16

后续结果列格式码的数量（下文以 *R* 表示）。可以为零，表示没有结果列，或者所有结果列都应使用默认格式（文本）；也可以为一，
此时指定的格式码应用于所有结果列（如果有）；还可以等于查询实际的结果列数量。

Int16[*R*]

结果列格式码。目前每个格式码必须为零（文本）或一（二进制）。

BindComplete (B)

Byte1('2')

将该消息标识为 Bind 完成指示。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CancelRequest (F)

Int32(16)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(80877102)

取消请求代码。此值的最高 16 位为 `1234`，最低 16 位为 `5678`。（为避免混淆，此代码不能与任何协议版本号相同。）

Int32

目标后端的进程 ID。

Int32

目标后端的密钥。

Close (F)

Byte1('C')

将该消息标识为 Close 命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

“`S`”表示关闭预备语句；“`P`”表示关闭 portal。

String

要关闭的预备语句或 portal 的名称（空字符串选择未命名的预备语句或 portal）。

CloseComplete (B)

Byte1('3')

将该消息标识为 Close 完成指示。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CommandComplete (B)

Byte1('C')

将该消息标识为命令完成响应。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

命令标签。通常是一个单词，用来标识已完成的 SQL 命令。

对于 `INSERT` 命令，标签是 `INSERT` *`oid rows`*，其中 *`rows`* 是插入的行数。如果 *`rows`* 为 1 且目标表具有 OID，则 *`oid`* 是插入行的对象 ID；否则 *`oid`* 为 0。

对于`DELETE`命令，标签是`DELETE` *`rows`*，其中*`rows`*表示删除的行数。

对于`UPDATE`命令，标签是`UPDATE` *`rows`*，其中*`rows`*是更新的行数。

对于`SELECT`或`CREATE TABLE AS`命令，标签是`SELECT` *`rows`*，其中*`rows`*是检索到的行数。

对于`MOVE`命令，标签是`MOVE` *`rows`*，其中*`rows`*表示游标位置改变的行数。

对于`FETCH`命令，标签是`FETCH` *`rows`*，其中*`rows`*是从游标中检索出的行数。

对于 `COPY` 命令，标签为 `COPY` *`rows`*，其中 *`rows`* 是复制的行数。（注意：行数仅出现在 PostgreSQL 8.2 及更高版本中。）

CopyData (F & B)

Byte1('d')

标识消息为`COPY`数据。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte*n*

数据是`COPY`数据流的一部分。来自后端的消息始终对应单个数据行，
但来自前端的消息可能会任意划分数据流。

CopyDone (F & B)

Byte1('c')

将消息标识为`COPY`完成指示符。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CopyFail (F)

Byte1('f')

将消息标识为COPY失败指示器。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

作为失败原因报告的错误消息。

CopyInResponse (B)

Byte1('G')

标识消息为开始复制输入的响应。前端此时必须发送复制输入数据（如果尚未准备好，应发送 CopyFail 消息）。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int8

0表示整体COPY格式是文本的（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式是二进制的（类似于DataRow格式）。更多信息请参见COPY。

Int16

要复制的数据中的列数（以下用*N*表示）。

Int16[*N*]

各列使用的格式代码。目前每个代码必须为零（文本）或一（二进制）。
如果整体复制格式为文本，则所有代码都必须为零。

CopyOutResponse (B)

Byte1('H')

标识消息为开始复制输出的响应。该消息之后会发送复制输出数据。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int8

0表示整体COPY格式是文本的（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式是二进制的（类似于DataRow格式）。更多信息请参见COPY。

Int16

要复制的数据中的列数（以下用*N*表示）。

Int16[*N*]

各列使用的格式代码。目前每个代码必须为零（文本）或一（二进制）。
如果整体复制格式为文本，则所有代码都必须为零。

CopyBothResponse (B)

Byte1('W')

标识消息为开始双向复制的响应。此消息仅用于流复制。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int8

0表示整体`COPY`格式是文本的（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式是二进制的（类似于DataRow格式）。更多信息请参见COPY。

Int16

要复制的数据中的列数（以下用*N*表示）。

Int16[*N*]

各列使用的格式代码。目前每个代码必须为零（文本）或一（二进制）。
如果整体复制格式为文本，则所有代码都必须为零。

DataRow (B)

Byte1('D')

标识消息为数据行。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

后面跟着的列值的数量（可能为零）。

接下来，每列都有以下两个字段：

Int32

列值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示列值为 NULL，此时后面没有值字节。

Byte*n*

列的值，格式由相关的格式代码指示。*n*是上述长度。

Describe (F)

Byte1('D')

标识消息为描述命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

'S' 表示描述一个预备语句；或者 'P' 表示描述一个 portal。

String

要描述的预备语句或 portal 的名称（空字符串选择未命名的预备语句或 portal）。

EmptyQueryResponse (B)

Byte1('I')

标识消息为对空查询字符串的响应。（此消息替代 CommandComplete。）

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

ErrorResponse (B)

Byte1('E')

将消息标识为错误。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

消息体由一个或多个带标识的字段组成，最后以一个零字节终止。
字段可以按任意顺序出现。每个字段包含以下内容：

Byte1

一个用于标识字段类型的代码；如果为零，则这是消息终止符，后面没有字符串。
目前定义的字段类型列在第 46.6 节中。由于将来可能会添加更多的字段类型，前端应该静默地忽略未识别类型的字段。

String

字段值。

Execute (F)

Byte1('E')

标识消息为一个执行命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

要执行的 portal 的名称（空字符串选择未命名的 portal）。

Int32

如果 portal 包含返回行的查询，则这是最多返回的行数（否则忽略此值）。零表示“无限制”。

Flush (F)

Byte1('H')

将消息标识为Flush命令。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

FunctionCall (F)

Byte1('F')

标识消息为函数调用。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

指定要调用的函数的对象 ID。

Int16

后续参数格式代码的数量（以下用 *C* 表示）。可以为零，表示没有参数，或所有参数都采用默认格式（文本）；也可以为一，表示将指定的格式代码用于所有参数；
还可以等于实际参数数量。

Int16[*C*]

参数格式代码。每个目前必须是零（文本）或一（二进制）。

Int16

指定传递给函数的参数数量。

接下来，每个参数都有以下两个字段：

Int32

参数值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示参数值为 NULL，此时后面没有值字节。

Byte*n*

参数的值，以相关格式代码指示的格式表示。*n*是上述长度。

最后一个参数之后还有以下字段：

Int16

函数结果的格式代码。目前必须为零（文本）或一（二进制）。

FunctionCallResponse (B)

Byte1('V')

标识消息为函数调用结果。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

函数结果值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示函数结果为 NULL，此时后面没有值字节。

Byte*n*

函数结果的值，格式由相关的格式代码指示。*n*是上述长度。

NoData (B)

Byte1('n')

将消息标识为无数据指示器。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

NoticeResponse (B)

Byte1('N')

将消息标识为通知。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

消息体由一个或多个带标识的字段组成，最后以一个零字节终止。
字段可以按任意顺序出现。每个字段包含以下内容：

Byte1

一个用于标识字段类型的代码；如果为零，则这是消息终止符，后面没有字符串。
目前定义的字段类型列在第 46.6 节中。由于将来可能会添加更多的字段类型，前端应该静默地忽略未识别类型的字段。

String

字段值。

NotificationResponse (B)

Byte1('A')

标识消息为通知响应。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

发出通知的后端进程的进程 ID。

String

发出该通知的通道名称。

String

通知进程传来的“载荷”字符串。

ParameterDescription (B)

Byte1('t')

标识消息为参数描述。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

语句使用的参数数量（可以为零）。

接下来，每个参数都有以下内容：

Int32

指定参数数据类型的对象 ID。

ParameterStatus (B)

Byte1('S')

标识消息为运行时参数状态报告。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

所报告的运行时参数的名称。

String

参数的当前值。

Parse (F)

Byte1('P')

将消息标识为解析命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

目标预备语句的名称（空字符串选择未命名的预备语句）。

String

要解析的查询字符串。

Int16

指定的参数数据类型的数量（可以为零）。请注意，
这不是查询字符串中可能出现的参数数量的指示，
而是前端希望为其预先指定类型的参数数量。

接下来，每个参数都有以下内容：

Int32

指定参数数据类型的对象 ID。此处填零等同于不指定类型。

ParseComplete (B)

Byte1('1')

将消息标识为解析完成指示器。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

PasswordMessage (F)

Byte1('p')

标识消息为密码响应。请注意，这也用于GSSAPI、SSPI和SASL响应消息。
可以从上下文中推断出确切的消息类型。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

密码（如果需要，已加密）。

PortalSuspended (B)

Byte1('s')

标识消息为 portal 挂起指示器。请注意，仅当执行消息的行数限制达到时才会出现此消息。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

Query (F)

Byte1('Q')

标识消息为简单查询。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

查询字符串本身。

ReadyForQuery (B)

Byte1('Z')

标识消息类型。ReadyForQuery在后端准备好进行新的查询周期时发送。

Int32(5)

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

当前后端事务状态指示器。可能的值为'I'，如果空闲（不在事务块中）；'T'，如果在事务块中；或'E'，如果在失败的事务块中（查询将被拒绝，直到块结束）。

RowDescription (B)

Byte1('T')

标识消息为行描述。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

指定一行中的字段数量（可以为零）。

接下来，每个字段都有以下内容：

String

字段名称。

Int32

如果能够确定该字段是某个特定表的列，则为该表的对象 ID；否则为零。

Int16

如果能够确定该字段是某个特定表的列，则为该列的属性编号；否则为零。

Int32

字段数据类型的对象 ID。

Int16

数据类型大小（参见`pg_type.typlen`）。注意，负值表示可变宽度类型。

Int32

类型修饰符（参见`pg_attribute.atttypmod`）。修饰符的含义是特定于类型的。

Int16

字段所使用的格式代码。目前为零（文本）或一（二进制）。对于 Describe 的语句变体所返回的 RowDescription，格式代码尚未确定，始终为零。

SSLRequest (F)

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(80877103)

SSL请求代码。该值被选择为在最高的16位中包含`1234`，在最低的16位中包含`5679`。（为避免混淆，此代码不得与任何协议版本号相同。）

StartupMessage (F)

Int32

消息内容以字节计的长度，包括自身。

Int32(196608)

协议版本号。最高 16 位是协议主版本号（本文档描述的协议为 3）。最低 16 位是协议次版本号（本文档描述的协议为 0）。

The protocol version number is followed by one or more pairs of parameter name and value strings. A zero byte is required as a terminator after the last name/value pair. Parameters can appear in any order. `user` is required, others are optional. Each parameter is specified as:

String

参数名。目前可识别的名称有：

`user`

要连接的数据库用户名。必需；没有默认值。

`database`

要连接的数据库。默认为用户名。

`options`

后端的命令行参数。（已弃用，建议设置单独的运行时参数。）

In addition to the above, any run-time parameter that can be set at backend start time might be listed. Such settings will be applied during backend start (after parsing the command-line options if any). The values will act as session defaults.

String

参数值。

Sync (F)

Byte1('S')

将消息标识为同步命令。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

Terminate (F)

Byte1('X')

标识消息为终止。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

46.6. 错误和通知消息字段

本节描述了ErrorResponse和NoticeResponse消息中可能出现的字段。
每个字段类型都有一个单字节的标识符。请注意，
任何给定的字段类型在消息中最多只能出现一次。

`S`

严重性：字段内容为 `ERROR`、`FATAL` 或 `PANIC`（在错误消息中），或 `WARNING`、`NOTICE`、`DEBUG`、`INFO` 或 `LOG`（在通知消息中），或这些内容的本地化翻译。始终存在。

`C`

代码：错误的 SQLSTATE 代码（见附录 A）。不可本地化。始终存在。

`M`

消息：主要的可读错误消息。应准确而简洁（通常为一行）。始终存在。

`D`

详细信息：可选的次要错误消息，提供有关该问题的更多细节。可能占多行。

`H`

提示：可选的建议，说明如何处理该问题。它与详细信息的区别在于，
它提供建议（可能并不合适），而非确凿事实。可能占多行。

`P`

位置：字段值是一个十进制ASCII整数，表示错误光标位置，作为原始查询字符串的索引。
第一个字符的索引为1，位置以字符而非字节计量。

`p`

内部位置：这与`P`字段的定义相同，
但在光标位置指向内部生成的命令而不是客户端提交的命令时使用。当此字段出现时，`q`
字段将始终出现。

`q`

内部查询：一个失败的内部生成命令的文本。例如，这可能是由PL/pg
SQL函数发出的SQL查询。

`W`

上下文：说明错误发生时的上下文。
目前包括活跃的过程语言函数和内部生成的查询的调用栈回溯。回溯每行一个条目，
最近的调用在前。

`F`

文件：报告错误的源代码位置所在的文件名。

`L`

行号：报告错误的源代码位置所在的行号。

`R`

例程：报告错误的源代码例程的名称。

客户端负责按自身需要格式化显示的信息；尤其应根据需要将长行断开。
错误消息字段中的换行符应视为段落分隔符，而非行分隔符。

46.7. 自协议 2.0 以来的变更总结

本节提供一份简要的变更清单，供准备将现有客户端库更新到协议 3.0 的开发者参考。

初始启动包采用灵活的字符串列表格式，取代了固定格式。注意，
运行时参数的会话默认值现在可以直接在启动包中指定。（实际上，以前也能通过 `options` 字段实现，但由于 `options` 的宽度有限，且无法用引号保护值中的空白字符，这种方法并不稳妥。）

现在，所有消息都在消息类型字节之后紧跟一个长度计数（启动包除外，它没有类型字节）。
另请注意，PasswordMessage 现在也有类型字节。

ErrorResponse 和 NoticeResponse（`'E'` 和 `'N'`）消息现在包含多个字段，客户端代码可以利用这些字段组合出所需详细程度的错误消息。注意，各字段通常不会以换行符结束，
而旧协议发送的单个字符串总是以换行符结束。

ReadyForQuery（`'Z'`）消息包含一个事务状态指示器。

BinaryRow 与 DataRow 消息类型不再有区别；单一的 DataRow 消息类型用于返回所有格式的数据。注意，DataRow 的布局已经改变，使其更容易解析。另外，二进制值的表示方式也已改变，不再直接取决于服务器内部的表示方式。

新增了“扩展查询”子协议，其中增加了前端消息类型 Parse、Bind、Execute、Describe、Close、Flush 和 Sync，以及后端消息类型 ParseComplete、BindComplete、Portal

Suspended、ParameterDescription、NoData 和 CloseComplete。现有客户端不必关注这个子协议，但使用它可能有助于改善性能或功能。

COPY 数据现在封装在 CopyData 和 CopyDone 消息中。COPY 期间的错误恢复已有明确定义的方法。不再需要特殊的最后一行“\.”，COPY OUT 期间也不再发送它。（在 COPY IN 期间仍会将它识别为终止符，但这种用法已弃用，最终将被移除。）现已支持二进制 COPY。CopyInResponse 和 CopyOutResponse 消息包含了表示列数及各列格式的字段。

FunctionCall 和 FunctionCallResponse 消息的布局已经改变。FunctionCall 现在支持向函数传递 NULL 参数，也可以使用文本或二进制格式传递参数和取得结果。由于它不再提供对服务器内部数据表示的直接访问，因此也不再有理由将 FunctionCall 视为潜在的安全漏洞。

在连接启动期间，后端会为其认为客户端库关注的所有参数发送 ParameterStatus（'S'）消息。之后，只要这些参数中任何一个的当前值发生变化，就会发送一条 ParameterStatus 消息。

RowDescription（'T'）消息针对所描述行中的每一列，增加了表 OID 和列编号字段，并显示各列的格式代码。

后端不再生成 CursorResponse（'P'）消息。

NotificationResponse（'A'）消息还有一个额外的字符串字段，它可以携带从 NOTIFY 事件发送方传递来的“载荷”字符串。

EmptyQueryResponse（'I'）消息以前包含一个空字符串参数，现在已将其移除。

第 47 章 PostgreSQL 编码约定

47.1. 格式

源代码格式使用 4 列制表宽度，并保留制表符（即，不会把制表符展开为空格）。每个逻辑缩进层级都对应再增加一个制表位。

布局规则（花括号位置等）遵循 BSD 约定。特别是，`if`、`while`、`switch` 等控制语句所对应代码块的花括号都单独占一行。

应限制行长度，使代码在 80 列窗口中可读。（这并不是说绝不能超过 80 列。例如，仅仅为了让代码保持在 80 列内而任意拆开一条很长的错误消息字符串，未必会提升可读性。）

不要使用 C++ 风格注释（`//` 注释）。严格的 ANSI C 编译器不接受这种注释。同样，也不要使用在块中途声明新变量这样的 C++ 扩展。

多行注释块的首选样式是

```
/*
 * comment text begins here
 * and continues here
 */
```

注意，从第 1 列开始的注释块会被 pgindent 原样保留，但它会把缩进的注释块当作普通文本重新排版。如果想保留缩进块中的换行，请像下面这样加上横线：

```
    /*----------
     * comment text begins here
     * and continues here
     *----------
     */
```

尽管提交的补丁并不绝对必须遵循这些格式规则，但这样做是个好主意。
你的代码会在下一个版本发布前经过 pgindent 处理，因此按另一套格式约定把它写得再漂亮也没有意义。补丁的一个经验法则是“让新代码看起来像周围现有的代码”。

`src/tools` 目录中包含可供 emacs、xemacs 或 vim 编辑器使用的示例设置文件，以帮助确保它们按这些约定格式化代码。

文本浏览工具 more 和 less 可以这样调用：

```
more -x4
less -x4
```

以便正确显示制表符。

47.2. 在服务器内部报告错误

服务器代码内生成的错误、警告和日志消息应使用 `ereport`，或其更老的近亲 `elog` 来创建。这一函数的用法相当复杂，因此需要一些解释。

每条消息都必须包含两个元素：严重性级别（范围从 `DEBUG` 到 `PANIC`）以及主消息文本。此外还可以有可选元素，其中最常见的是遵循 SQL 规范 SQLSTATE 约定的错误标识符代码。`ereport` 本身只是一个包装函数，主要为了语法上的便利，使消息生成在 C 源代码中看起来像一次函数调

用。ereport 唯一直接接受的参数是严重性级别。主消息文本以及任何可选消息元素都是通过在 ereport 调用中调用辅助函数（例如 errmsg）来生成的。

对ereport的一次典型调用可能如下：

```
ereport(ERROR,
        (errcode(ERRCODE_DIVISION_BY_ZERO),
         errmsg("division by zero")));
```

这指定了错误严重性级别ERROR（一种普通错误）。errcode调用使用定义在src/include/utils/errcodes.h中的一个宏指定 SQLSTATE 错误代码。errmsg调用提供主消息文本。注意辅助函数调用外面多了一层圆括号 – 它们虽然烦人，但在语法上是必需的。

这里有一个更复杂的示例：

```
ereport(ERROR,
        (errcode(ERRCODE_AMBIGUOUS_FUNCTION),
         errmsg("function %s is not unique",
                func_signature_string(funcname, nargs,
                                      NIL, actual_arg_types)),
         errhint("Unable to choose a best candidate function. "
                 "You might need to add explicit typecasts.")));
```

这展示了如何用格式代码把运行时值嵌入消息文本中。此外还提供了一条可选的“提示”消息。

适用于 ereport 的辅助例程有：

- errcode(sqlerrcode) 为该条件指定 SQLSTATE 错误标识符代码。如果不调用这个例程，则默认错误标识符在错误严重性级别为 ERROR 或更高时为 ERRCODE_INTERNAL_ERROR，在错误级别为 WARNING 时为 ERRCODE_WARNING，否则（对于 NOTICE 及以下）为 ERRCODE_SUCCESSFUL_COMPLETION。虽然这些默认值常常很方便，但在省略 errcode() 调用之前，始终要先想想它们是否合适。

- errmsg(const char *msg, ...) 指定主错误消息文本，以及可能要插入其中的运行时值。插入项通过 sprintf 风格的格式代码指定。除了 sprintf 接受的标准格式代码外，还可以使用格式代码 %m 插入 strerror 为 errno 当前值返回的错误消息。[1] %m does not require any corresponding entry in the parameter list for errmsg. Note that the message string will be run through gettext for possible localization before format codes are processed.

- errmsg_internal(const char *msg, ...) 与 errmsg 相同，只是消息字符串不会被翻译，也不会被收入国际化消息字典。这应当用于那些“不可能发生”、大概不值得为之投入翻译精力的情况。

- errmsg_plural(const char *fmt_singular, const char *fmt_plural, unsigned long n, ...) 类似于 errmsg，但支持消息的各种复数形式。*fmt_singular* 是英文单数格式，*fmt_plural* 是英文复数格式，*n* 是决定需要哪种复数形式的整数值，其余参数按所选格式字符串进行格式化。更多信息见第 48.2.2 节。

- errdetail(const char *msg, ...) 提供一条可选的“详情”消息；当有额外信息但似乎不适合放在主消息中时，可使用它。消息字符串的处理方式与 errmsg 完全相同。

- errdetail_internal(const char *msg, ...) 与 errdetail 相同，只是消息字符串不会被翻译，也不会被收入国际化消息字典。这应当用于那些不值得投入翻译精力的详情消息，例如它们对大多数用户来说技术性太强而无甚用处。

[1] 也就是说，是到达 ereport 调用点时的那个值；辅助报告例程内部对 errno 的更改不会影响它。如果显式写出 strerror(errno) 作为 errmsg 的参数列表内容，就不是这样了；因此不要这么做。

- `errdetail_plural(const char *fmt_singular, const char *fmt_plural, unsigned long n, ...)` 类似于 `errdetail`，但支持消息的各种复数形式。更多信息见第 48.2.2 节。

- `errdetail_log(const char *msg, ...)` 与 `errdetail` 相同，只是该字符串只会写入服务器日志，绝不会发送给客户端。如果同时使用 `errdetail`（或其上面的某个等价函数）和 `errdetail_log`，那么一条字符串会发往客户端，另一条会发往日志。对于那些因安全性过于敏感或内容过于庞大而不适合放入发给客户端的报告中的错误细节，这很有用。

- `errhint(const char *msg, ...)` 提供一条可选的“提示”消息；当要就如何修复问题给出建议（而非关于出了什么问题的事实性详情）时，可使用它。消息字符串的处理方式与 `errmsg` 完全相同。

- `errcontext(const char *msg, ...)` 通常不会直接在 `ereport` 消息处调用；它用于 `error_context_stack` 回调函数中，提供错误发生时的上下文信息，例如 PL 函数中的当前位置。消息字符串的处理方式与 `errmsg` 完全相同。与其他辅助函数不同，它在每次 `ereport` 调用中可以被调用多次；这样提供的 successive 字符串会用换行符分隔拼接起来。

- `errposition(int cursorpos)` 指定错误在查询字符串中的文本位置。目前，它只对在查询处理的词法和语法分析阶段检测到的错误有用。

- `errcode_for_file_access()` 是一个便捷函数，用于为与文件访问相关的系统调用失败选择合适的 SQLSTATE 错误标识符。它使用保存下来的 `errno` 来确定要生成哪种错误代码。通常应将它与主错误消息文本中的 `%m` 结合使用。

- `errcode_for_socket_access()` 是一个便捷函数，用于为与套接字相关的系统调用失败选择合适的 SQLSTATE 错误标识符。

- `errhidestmt(bool hide_stmt)` 可被调用以指定在 postmaster 日志中抑制消息的 `STATEMENT:` 部分。一般来说，如果消息文本已经包含当前语句，这样做就是合适的。

还有一个较旧的函数`elog`，至今仍被大量使用。一个`elog`调用：

```
elog(level, "format string", ...);
```

完全等价于：

```
ereport(level, (errmsg_internal("format string", ...)));
```

注意，SQLSTATE 错误代码总会取默认值，而且消息字符串不会被翻译。因此，`elog`只应用于内部错误和低层调试日志。凡是普通用户可能感兴趣的消息，都应通过`ereport`。尽管如此，系统中仍有足够多的内部“不可能发生”错误检查，因此`elog`依然被广泛使用；对这类消息来说，由于记法更简洁，它更受青睐。

关于如何编写良好的错误消息，可参见第 47.3 节。

47.3. 错误消息风格指南

提供这份风格指南，是希望在 PostgreSQL 生成的所有消息中保持一致、对用户友好的风格。

各部分写什么

主消息应简短、陈述事实，并避免提及具体函数名等实现细节。“简短”的意思是“在正常情况下应能放在一行里”。如果为了保持主消息简短，

或者你觉得有必要提及特定失败的系统调用之类的实现细节，可使用详情消息。
主消息和详情消息都应当陈述事实。提示消息则用于给出如何修复问题的建议，
尤其是在该建议未必总是适用时。

例如，不要写：

```
IpcMemoryCreate: shmget(key=%d, size=%u, 0%o) failed: %m
(plus a long addendum that is basically a hint)
```

而要写：

```
Primary:    could not create shared memory segment: %m
Detail:     Failed syscall was shmget(key=%d, size=%u, 0%o).
Hint:       the addendum
```

原理：保持主消息简短，有助于使其切中要点，也便于客户端假定一行足以容纳错误消息，
从而安排屏幕空间。详情消息和提示消息可以归入详细模式，或者放到弹出的错误详情窗口中。
此外，为节省空间，详情消息和提示消息通常会从服务器日志中省略。最好避免提及实现细节，
因为用户反正也不知道这些细节。

格式

不要在消息文本中写入任何特定的格式假设。应预期客户端和服务器日志会按照各自需要折行。
对于较长的消息，可以使用换行字符（\n）来表示建议的分段。不要让消息以换行结束。
不要使用制表符或其他格式控制字符。（在错误上下文显示中，
会自动添加换行以分隔函数调用等上下文层次。）

原理：消息不一定显示在终端类设备上。在 GUI 显示或浏览器中，这些格式指令充其量也只是被忽略。

引号

英文文本在适合引用时应使用双引号。其他语言的文本应一致地使用一种引号，
这种引号应符合该语言的出版习惯以及其他程序的计算机输出习惯。

原理：选择双引号而不是单引号多少有些武断，但通常更符合首选用法。有人建议按照 SQL 约定，根据对象类型来选择引号的种类（即字符串用单引号，标识符用双引号）。
但这是语言内部的技术问题，很多用户甚至并不熟悉；它也无法扩展到其他类型的被引术语，
更无法翻译到其他语言中，而且本身也没什么意义。

引号的用法

对文件名、用户提供的标识符以及其他可能包含单词的变量，始终使用引号界定。
对于不会包含单词的变量（例如操作符名），则不要加引号。

后端中有些函数会在需要时自行给其输出加上双引号（例如 `format_type_be()`）。不要再对这类函数的输出额外加引号。

原理：对象的名称在嵌入消息时可能造成歧义。对于插入的名称从哪里开始、到哪里结束，
要保持一致。但也不要用不必要或重复的引号把消息弄得杂乱无章。

语法和标点

主错误消息与详情/提示消息的规则不同：

主错误消息：首字母不要大写。消息末尾不要加句号。更不要想着在消息末尾加感叹号。

详情消息和提示消息：使用完整句子，并且每条都用句号结束。句子的第一个词要首字母大写。
如果后面还有一句，在句号后放两个空格（针对英文文本；对其他语言可能并不合适）。

错误上下文字符串：首字母不要大写，字符串末尾不要加句号。
上下文字符串通常不应是完整句子。

原理：避免使用标点，能让客户端应用更容易把消息嵌入各种语法上下文中。
主消息往往本来也不是语法完整的句子。（而如果它们长到不止一句，
就应当拆分成主消息和详情消息。）不过，详情消息和提示消息更长，
也可能需要包含多个句子。为了保持一致，即便只有一个句子，它们也应遵循完整句子的风格。

大写与小写

消息措辞应使用小写，包括主错误消息的首字母。如果消息中出现 SQL 命令或关键字，则使用大写形式。

原理：这样更容易让所有消息看起来一致，因为有些消息是完整句子，而有些则不是。

避免被动语态

使用主动语态。有施动者时，用完整句子（“A could not do B”）。如果施动者就是程序本身，则使用无主语的电报式写法，但不要把程序写成“I”。

原理：程序不是人。不要假装它是。

现在时与过去时

如果一次尝试做某事失败了，但下次仍可能成功（也许在修复某个问题之后），就使用过去时。
如果失败显然是永久性的，就使用现在时。

下面两种句式在语义上有明显差别：

```
could not open file "%s": %m
```

和：

```
cannot open file "%s"
```

第一种表示尝试打开文件失败了。消息应给出原因，例如“磁盘已满”或“文件不存在”。
过去时更合适，因为下次磁盘可能就不满了，或者所请求的文件可能已经存在。

第二种形式表示，程序中根本不存在打开该命名文件的功能，或者从概念上就不可能。
现在时更合适，因为这种情况会无限期持续下去。

原理：诚然，普通用户未必能仅凭消息的时态得出什么重要结论，但既然语言提供了语法，
我们就应当正确使用它。

对象类型

在引用对象名称时，要说明它是什么类型的对象。

原理：否则没人会知道“foo.bar.baz”指的是什么。

括号

方括号只应用于：（1）命令概要中表示可选参数，或（2）表示数组下标。

原理：其他任何用法都不符合广为人知的习惯用法，而且会让人困惑。

组装错误消息

当消息中包含其他地方生成的文本时，应按以下方式嵌入：

```
could not open file %s: %m
```

原理：要把这种文本拼进一个流畅的单句中，同时照顾所有可能的错误代码会很困难，
因此需要某种标点。也有人建议把嵌入的文本放在圆括号里，
但如果嵌入的文本很可能是消息中最重要的部分（而这往往确实如此），那样就显得不自然了。

错误原因

消息总应说明错误发生的原因。例如：

```
BAD:    could not open file %s
BETTER: could not open file %s (I/O failure)
```

如果不知道原因，最好去修复代码。

函数名

不要在错误文本中包含报告该错误的例程名。需要时，我们有其他机制可以找出它，
而且对大多数用户来说，这项信息没有帮助。如果去掉函数名后错误文本就不够清楚，
请重新措辞。

```
BAD:    pg_atoi: error in "z": cannot parse "z"
BETTER: invalid input syntax for integer: "z"
```

也要避免提及被调用函数的名称；应改为说明代码想要做什么：

```
BAD:    open() failed: %m
BETTER: could not open file %s: %m
```

如果确实有必要，可在详情消息中提及系统调用。（在某些情况下，
在详情消息中提供传给该系统调用的实际值也许是合适的信息。）

原理：用户并不知道那些函数都做了什么。

应避免的词语

Unable. “Unable” 几乎就是被动语态。更好的做法是视情况使用 “cannot” 或 “could not” 。

Bad. 像 “bad result” 这样的错误消息，实在很难被合理理解。最好写出结果为什么是 “bad” ，例如 “invalid format” 。

Illegal. “Illegal” 指的是违法，其余情况应使用 “invalid” 。更好的是，直接说明为什么无效。

Unknown. 尽量避免“unknown”。考虑“error: unknown response”。如果你都不知道响应是什么，又如何知道它有错？“Unrecognized”往往是更好的选择。此外，一定要包含被抱怨的值。

```
BAD:    unknown node type
BETTER: unrecognized node type: 42
```

Find vs. Exists. 如果程序为了定位某个资源使用了非平凡算法（例如路径搜索），而该算法失败了，那么说程序没能“find”该资源是公平的。另一方面，如果资源的预期位置已知，但程序无法在那里访问它，那么就应说该资源并不“exist”。在这种情况下使用“find”显得含糊，也会混淆问题。

May、Can 与 Might. “May”暗示许可（例如，"You may borrow my rake."），在文档或错误消息中用途不大。“Can”暗示能力（例如，"I can lift that log."），而“might”暗示可能性（例如，"It might rain today."）。使用恰当的词能澄清含义，也有助于翻译。

Contractions. 避免使用缩略形式，例如“can't”；应改用“cannot”。

完整拼写

单词应完整拼写。例如，避免使用：

- spec
- stats
- parens
- auth
- xact

原理：这会提升一致性。

本地化

请记住，错误消息文本需要被翻译成其他语言。请遵循第 48.2.2 节中的指导，避免给翻译者制造麻烦。

第 48 章 本地语言支持

Peter Eisentraut

48.1. 给翻译者

PostgreSQL 程序（服务器和客户端）可以用你所偏好的语言发出消息 — 前提是这些消息已经被翻译。创建和维护翻译后的消息集，需要那些精通自己语言并愿意为 PostgreSQL 项目作出贡献的人的帮助。做这件事完全不必是程序员。本节说明如何提供帮助。

48.1.1. 要求

我们不会评判你的语言能力 — 本节讨论的是软件工具。理论上，你只需要一个文本编辑器。但这只有在你不打算试用自己翻译的消息时才成立，而这种情况不太可能发生。
配置你的源代码树时，请务必使用 `--enable-nls` 选项。这还会检查 libintl 库和 `msgfmt` 程序，而终端用户无论如何也需要它们。要试用你的工作成果，请遵循安装说明中适用的部分。

如果你想开始一项新的翻译工作，或者想执行一次消息目录合并（后文会描述），则分别需要 GNU 兼容实现的 `xgettext` 和 `msgmerge` 程序。今后我们会尽量做到：如果你使用打包的源码发布版，就不需要 `xgettext` 了。（如果从 Git 工作，你仍然需要它。）目前推荐使用 GNU Gettext 0.10.36 或更高版本。

你本地的 gettext 实现应该自带文档。其中有些内容可能与下文重复，但要了解更多细节，你应当去查阅那些文档。

48.1.2. 概念

原始（英语）消息及其（可能存在的）译文对应项，以成对形式保存在消息目录中；
每个程序以及每种目标语言各有一份消息目录（尽管相关程序可以共享同一个消息目录）。
消息目录有两种文件格式：第一种是“PO”文件（Portable Object，可移植对象），它是带有特殊语法的纯文本文件，由翻译者编辑。第二种是“MO”文件（Machine Object，机器对象），它是根据相应的 PO 文件生成的二进制文件，在国际化后的程序运行时使用。翻译者不需要处理 MO 文件；事实上几乎没有人需要。

消息目录文件的扩展名，毫不意外地是 `.po` 或 `.mo`。基本名则视情况而定，要么是其所对应的程序名，要么是该文件所对应的语言。这多少有些令人困惑。例子包括 `psql.po`（psql 的 PO 文件）或 `fr.mo`（法语的 MO 文件）。

PO 文件的格式示例如下：

```
# comment

msgid "original string"
msgstr "translated string"

msgid "more original"
msgstr "another translated"
"string can be broken up like this"

...
```

msgid 行是从程序源代码中提取出来的。（并非一定如此，但这是最常见的做法。）msgstr 行起初为空，由翻译者填入有意义的字符串。字符串可以包含 C 风格的转义字符，也可以像上例那样跨多行延续。（下一行必须从行首开始。）

字符引入注释。如果 # 字符后面紧跟空白，那么这就是由翻译者维护的注释。也可能存在自动注释，即 # 后紧跟非空白字符的注释。这些注释由各种处理 PO 文件的工具维护，目的是帮助翻译者。

```
#. automatic comment
#: filename.c:1023
#, flags, flags
```

#. 风格的注释是从使用该消息的源文件中提取出来的。程序员可能已经为翻译者插入了信息，例如期望的对齐方式。#: 注释指出该消息在源代码中使用的精确位置。翻译者不必查看程序源代码，但如果对正确译法有疑问，也可以去看。#, 注释包含以某种方式描述该消息的标志。目前有两种标志：如果某条消息可能因程序源代码的更改而过期，则会设置 `fuzzy`。翻译者随后可以核实这一点，并在可能时移除 fuzzy 标志。注意，fuzzy 消息不会提供给终端用户。另一个标志是 `c-format`，它表示该消息是一个 `printf` 风格的格式模板。这意味着译文也应该是一个格式字符串，并具有相同数量和类型的占位符。有一些工具会据此验证格式字符串，它们就是根据 c-format 标志来检查的。

48.1.3. 创建和维护消息目录

好，那么该如何创建一个“空白”消息目录呢？首先，进入包含你想翻译其消息的程序的目录。如果那里有一个 `nls.mk` 文件，就说明这个程序已经为翻译做好准备了。

如果已经有一些 `.po` 文件，那么就说明已经有人做过部分翻译工作。这些文件命名为 *`language`*`.po`，其中 *`language`* 是 ISO 639-1 两字母语言代码（小写）[1]，例如法语对应 `fr.po`。如果某种语言确实需要进行多套翻译工作，这些文件也可以命名为 *`language_region`*`.po`，其中 *`region`* 是 ISO 3166-1 两字母国家代码（大写）[2]，例如巴西葡萄牙语对应 `pt_BR.po`。如果你找到了自己想要的语言，就可以直接开始编辑那个文件。

如果需要启动一项新的翻译工作，请先运行命令：

```
gmake init-po
```

这会创建文件*`progname`*`.pot`。（用`.pot`来与“正在使用中”的 PO 文件区分，`T`代表“模板”。）把这个文件复制为*`language`*`.po`并编辑它。为了让别人知道新语言已可用，还要编辑文件`nls.mk`，把语言（或语言和国家）代码加到形如下面的行中：

```
AVAIL_LANGUAGES := de fr
```

（当然也可以是其他语言。）

随着底层程序或库的变化，程序员可能会更改或添加消息。这种情况下不必从头开始，只需运行命令：

```
gmake update-po
```

该命令会创建一个新的空白消息目录文件（即你最初使用的 pot 文件），并把它与现有的 PO 文件合并。如果合并算法对某个消息没有把握，就会如上所述把它标记为“模糊”。新的 PO 文件以`.po.new`扩展名保存。

[1] http://www.loc.gov/standards/iso639-2/php/English_list.php
[2] http://www.iso.org/iso/country_names_and_code_elements

48.1.4. 编辑 PO 文件

PO 文件可以用普通文本编辑器编辑。翻译者只应修改 msgstr 指令后引号之间的内容、添加注释，以及调整 fuzzy 标志。Emacs 里（不出意外地）有一个 PO 模式，相当有用。

PO 文件不必完全填满。如果没有可用译文（或译文为空），软件会自动回退到原始字符串。把不完整的翻译提交到源码树中也没有问题；这样别人就有机会接手你的工作。不过，完成一次合并之后，建议优先清除 fuzzy 条目。记住，fuzzy 条目不会被安装；它们只用来作为可能正确译文的参考。

编辑译文时要记住以下几点：

- 务必确保如果原文以换行结束，译文也同样以换行结束。制表符等也是如此。
- 如果原文是一个 `printf` 格式字符串，译文也必须如此。译文还需要按相同顺序保留相同的格式说明符。有时语言自身的规则会让这一点变得不可能，或者至少很别扭。在这种情况下，你可以像下面这样修改格式说明符：

  ```
  msgstr "Die Datei %2$s hat %1$u Zeichen."
  ```

  这样一来，第一个占位符实际上会使用参数列表中的第二个参数。*digits*`$` 必须紧跟在 `%` 后面，位于其他任何格式修饰符之前。（这个特性确实存在于 `printf` 函数族中。你以前可能没听说过它，因为在消息国际化之外，它几乎没有用武之地。）
- 如果原始字符串包含语言错误，就报告它（或者你自己在程序源代码中修正它），然后照常翻译。等程序源代码更新之后，修正过的字符串就可以被合并进来。如果原始字符串包含事实性错误，就报告它（或者你自己修正它），并且不要翻译它。相反，你可以在 PO 文件里为该字符串加上一条注释。
- 保持原始字符串的风格和语气。特别是，那些不是完整句子的消息（`cannot open file %s`）通常不应以大写字母开头（如果你的语言区分大小写），也不应以句号结尾（如果你的语言使用标点符号）。阅读第 47.3 节可能会有帮助。
- 如果你不知道某条消息是什么意思，或者它有歧义，就到开发者邮件列表里提问。很可能说英语的终端用户也同样看不懂它，或者会觉得它有歧义，因此最好直接改进那条消息。

48.2. 给程序员

48.2.1. 实现机制

本节描述如何在 PostgreSQL 发行版中的程序或库里实现本地语言支持。目前它只适用于 C 程序。

为程序添加 NLS 支持

1. 把下面的代码插入到该程序的启动序列中：

   ```
   #ifdef ENABLE_NLS
   #include <locale.h>
   #endif

   ...
   ```

```
#ifdef ENABLE_NLS
setlocale(LC_ALL, "");
bindtextdomain("progname", LOCALEDIR);
textdomain("progname");
#endif
```

（其中 *progname* 实际上可以自由选择。）

2. 凡是遇到可能需要翻译的消息，都需要插入对 `gettext()` 的调用。例如：

   ```
   fprintf(stderr, "panic level %d\n", lvl);
   ```

   将改成：

   ```
   fprintf(stderr, gettext("panic level %d\n"), lvl);
   ```

   （如果没有配置 NLS 支持，`gettext` 会被定义成一个空操作。）

   这样往往会增加很多杂乱内容。一个常用的简写是：

   ```
   #define _(x) gettext(x)
   ```

   如果该程序的大部分通信都是通过一个或少数几个函数完成的，例如后端中的 `ereport()`，另一种可行的解决办法是让这个函数在内部对所有输入字符串调用 `gettext`。

3. 在程序源代码所在目录中添加一个`nls.mk`文件。这个文件会作为 makefile 读取。这里需要设置以下变量：

   `CATALOG_NAME`

   程序名，即 `textdomain()` 调用中提供的名称。

   `AVAIL_LANGUAGES`

   已提供的翻译列表 – 最初为空。

   `GETTEXT_FILES`

   包含可翻译字符串的文件列表，也就是那些使用 `gettext` 或其他替代方案标记过的文件。最终，这会包含该程序几乎所有的源文件。如果这个列表太长，
   可以让第一个“文件”是一个 `+`，第二个词则是一个文件名，该文件每行包含一个文件名。

   `GETTEXT_TRIGGERS`

   为翻译者生成消息目录的工具需要知道哪些函数调用包含可翻译字符串。默认只认识 `gettext()` 调用。如果你使用了 `_` 或其他标识符，就需要在这里列出它们。如果可翻译字符串不是第一个参数，条目就需要写成 `func:2` 这样的形式（表示第二个参数）。如果你有一个支持复数形式消息的函数，条目应写成 `func:1,2` 这样（分别标识单数和复数消息参数）。

构建系统会自动处理消息目录的构建和安装。

48.2.2. 消息编写指南

下面列出一些便于翻译的消息编写指南。

- 不要像下面这样在运行时拼接句子：

  ```
  printf("Files were %s.\n", flag ? "copied" : "removed");
  ```

  句子中的词序在其他语言里可能完全不同。另外，即使你记得对每个片段都调用 `gettext()`，这些片段分开来看也可能难以妥善翻译。最好稍微重复一点代码，让每条待翻译消息都成为一个语义完整的整体。只有数字、文件名之类的运行时变量才应该在运行时插入消息文本中。

- 出于类似的原因，下面这样也行不通：

  ```
  printf("copied %d file%s", n, n!=1 ? "s" : "");
  ```

  因为它假定了复数形式的构造方式。如果你以为可以这样解决：

  ```
  if (n==1)
      printf("copied 1 file");
  else
      printf("copied %d files", n):
  ```

  那就会失望了。有些语言的复数形式不止两种，而且规则相当特殊。通常最好从设计上彻底避开这个问题，例如这样：

  ```
  printf("number of copied files: %d", n);
  ```

  如果你确实想构造一条正确处理复数的消息，也有相应支持，只是有点别扭。在 `ereport()` 中生成主错误消息或详细错误消息时，可以写成这样：

  ```
  errmsg_plural("copied %d file",
                "copied %d files",
                n,
                n)
  ```

  第一个参数是适用于英语单数形式的格式字符串，第二个参数是适用于英语复数形式的格式字符串，第三个参数是决定使用哪种复数形式的整数控制值。后续参数会像往常一样按照格式字符串进行格式化。（通常，这个复数控制值本身也会是要格式化的值之一，所以必须写两次。）在英语中，只需要区分 *n* 是否为 1；而在其他语言中，可能存在许多不同的复数形式。翻译者会把这两个英语形式视为一组，并有机会提供多个替代字符串，再根据运行时的 *n* 值选择合适的那个。

  如果你需要对一条不是直接用于 `errmsg` 或 `errdetail` 报告的消息做复数处理，就必须使用底层函数 `ngettext`。参见 gettext 文档。

- 如果你想向翻译者传达某些信息，例如一条消息应如何与其他输出对齐，就在该字符串出现之前放置一个以 `translator` 开头的注释，例如：

  ```
  /* translator: This message is not what it seems to be. */
  ```

  这些注释会被复制到消息目录文件中，这样翻译者就能看到它们。

第 49 章 编写过程语言调用处理器

凡是不是使用当前针对编译型语言的“版本 1”接口编写的函数，在被调用时都会经过该语言专用的调用处理器函数。这包括用户定义过程语言中的函数、用 SQL 编写的函数，以及使用版本 0 编译语言接口的函数。调用处理器负责以恰当的方式执行该函数，例如解释所提供的源文本。本章概述如何编写新的过程语言调用处理器。

过程语言的调用处理器是一个“普通”函数，必须使用诸如 C 这样的编译型语言、按照版本 1 接口编写，并在 PostgreSQL 中注册为不接受参数且返回 `language_handler` 类型。这个特殊伪类型会将该函数标识为调用处理器，并阻止它在 SQL 命令中被直接调用。有关 C 语言调用约定和动态装载的更多细节，见第 35.9 节。

调用处理器与其他函数的调用方式相同：它接收一个指向 `FunctionCallInfoData struct` 的指针，其中包含参数值和被调用函数的信息，并应返回一个 `Datum` 结果（如果要返回 SQL 空值，还需要设置 `FunctionCallInfoData` 结构体的 `isnull` 字段）。调用处理器与普通被调用函数的区别在于，`FunctionCallInfoData` 结构体的 `flinfo->fn_oid` 字段包含的是实际要调用的函数的 OID，而不是调用处理器自身的 OID。调用处理器必须使用该字段确定应执行哪个函数。此外，传入的参数列表也是按照目标函数的声明设置的，而不是按照调用处理器的声明。

调用处理器必须从系统目录 `pg_proc` 中取出该函数的条目，并分析被调用函数的参数类型和返回类型。该函数的 `CREATE FUNCTION` 命令中的 `AS` 子句内容，保存在 `pg_proc` 对应行的 `prosrc` 列中。这里通常是一段过程语言源文本，但理论上也可以是其他内容，例如某个文件的路径名，或者任何能详细告诉调用处理器该做什么的信息。

同一个函数在执行一条 SQL 语句期间往往会被调用很多次。调用处理器可以利用 `flinfo->fn_extra` 字段，避免重复查找被调用函数的信息。该字段起初为 `NULL`，但调用处理器可以把它设置为指向与被调用函数有关的信息。在后续调用中，如果 `flinfo->fn_extra` 已经不是 `NULL`，就可以直接使用它并跳过信息查找步骤。调用处理器必须确保 `flinfo->fn_extra` 指向的内存至少能存活到当前查询结束，因为 `FmgrInfo` 数据结构可能会保留这么久。一种做法是在 `flinfo->fn_mcxt` 指定的内存上下文中分配这些额外数据；这类数据通常会与 `FmgrInfo` 本身具有相同的生命周期。不过，处理器也可以选择使用生命周期更长的内存上下文，以便跨查询缓存函数定义信息。

过程语言函数作为触发器调用时，不会按通常方式传入参数，但 `FunctionCallInfoData` 的 `context` 字段会指向一个 `TriggerData` 结构体，而不是像普通函数调用那样为 `NULL`。语言调用处理器应提供让过程语言函数获取触发器信息的机制。

下面是用 C 编写的过程语言处理器模板：

```
#include "postgres.h"
#include "executor/spi.h"
#include "commands/trigger.h"
#include "fmgr.h"
#include "access/heapam.h"
#include "utils/syscache.h"
#include "catalog/pg_proc.h"
#include "catalog/pg_type.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

PG_FUNCTION_INFO_V1(plsample_call_handler);
```

```
Datum
plsample_call_handler(PG_FUNCTION_ARGS)
{
    Datum          retval;

    if (CALLED_AS_TRIGGER(fcinfo))
    {
        /*
         * 作为触发器过程调用
         */
        TriggerData    *trigdata = (TriggerData *) fcinfo->context;

        retval = ...
    }
    else
    {
        /*
         * 作为函数调用
         */

        retval = ...
    }

    return retval;
}
```

只需在省略号处添加几千行代码，就能完成这个调用处理器。

将处理器函数编译为可加载模块后（参见第 35.9.6 节），可用以下命令注册示例过程语言：

```
CREATE FUNCTION plsample_call_handler() RETURNS language_handler
    AS 'filename'
    LANGUAGE C;
CREATE LANGUAGE plsample
    HANDLER plsample_call_handler;
```

虽然仅提供调用处理器就足以创建一种最小可用的过程语言，但还可以额外提供另外两个函数，以便让该语言更易于使用。它们是验证器和内联处理器。可以提供验证器，以便在CREATE FUNCTION期间执行与该语言相关的检查。也可以提供内联处理器，以便让该语言支持通过DO命令执行匿名代码块。

如果过程语言提供了验证器，它必须声明为一个接受单个 `oid` 参数的函数。验证器的返回值会被忽略，因此习惯上将其声明为返回 `void`。在 `CREATE FUNCTION` 命令创建或更新了一个用该过程语言编写的函数之后，验证器会在该命令结束时被调用。传入的 OID 是该函数在 `pg_proc` 中对应行的 OID。验证器必须按通常方式取出这一行，并执行适当的检查。首先，应调用 `CheckFunctionValidatorAccess()`，以诊断那些用户无法通过 `CREATE FUNCTION` 实现的对验证器的显式调用。典型检查包括确认该语言支持该函数的参数类型和结果类型，以及确认函数体在该语言中语法正确。如果验证器认为该函数没有问题，就应直接返回；如果发现错误，则应通过常规的 `ereport()` 错误报告机制加以报告。抛出错误会强制事务回滚，从而阻止错误的函数定义被提交。

验证器函数通常应遵守check_function_bodies参数：关闭该参数时，应跳过代价高昂或依赖上下文的检查。如果语言允许在编译时执行代码，验证器必须禁止会引发这种执行的检查。特别是，pg_dump 会关闭此参数，以便装载过程语言

函数，而无需担心副作用或函数体对其他数据库对象的依赖。（因此，
调用处理器不应假定验证器已经完整检查了函数。验证器的目的不是让调用处理器省略检查，
而是在 CREATE FUNCTION 命令存在明显错误时立即通知用户。）虽然具体检查哪些内容主要由验证器函数自行决定，但要注意，核心 CREATE FUNCTION 代码只有在 check_function_bodies 开启时，才执行函数附带的 SET 子句。因此，关闭 check_function_bodies 时，必须跳过那些结果可能受 GUC 参数影响的检查，以免重新装载转储时误报错误。

如果过程语言提供了内联处理器，它必须声明为一个接受单个 internal 参数的函数。内联处理器的返回值会被忽略，因此习惯上将其声明为返回 void。当执行指定了该过程语言的 DO 语句时，就会调用内联处理器。实际传入的参数是一个指向 InlineCodeBlock 结构体的指针，其中包含 DO 语句参数的相关信息，特别是待执行的匿名代码块文本。内联处理器应执行这段代码并返回。

建议你将以上所有函数声明以及 CREATE LANGUAGE 命令本身都封装进一个扩展中，这样只需执行一条简单的 CREATE EXTENSION 命令就足以安装该语言。关于如何编写扩展，见第 35.15 节。

标准发行版中附带的各过程语言，在尝试编写自己的语言调用处理器时都是很好的参考资料。
请查看源码树中的 src/pl 子目录。CREATE LANGUAGE参考页中也包含一些有用的细节。

第 50 章 编写外部数据包装器

对外部表的所有操作都通过它的外部数据包装器处理，
外部数据包装器由一组供规划器和执行器调用的函数组成。
外部数据包装器负责从远程数据源取回数据，并把数据返回给 PostgreSQL 执行器。本章概述如何编写一个新的外部数据包装器。

标准发行中包含的外部数据包装器是编写自己的包装器时的好参考。可以查看源代码树的 `contrib/file_fdw` 子目录。CREATE FOREIGN DATA WRAPPER参考页也提供了一些有用的细节。

> **注意**
>
> SQL 标准规定了一个用于编写外部数据包装器的接口。但是，PostgreSQL 没有实现该 API，因为要让 PostgreSQL 适配它需要大量工作，而且该标准 API 也没有得到广泛采用。

50.1. 外部数据包装器函数

FDW 作者需要实现一个处理器函数，并且可以选择实现一个验证器函数。这两个函数都必须使用 C 之类的编译型语言编写，并采用版本 1 接口。关于 C 语言调用约定和动态加载的细节，请参阅第 35.9 节。

处理器函数只是返回一个结构体，
其中包含指向将由规划器和执行器调用的回调函数的函数指针。编写 FDW 的大部分工作都在实现这些回调函数上。The handler function must be registered with PostgreSQL as taking no arguments and returning the special pseudo-type `fdw_handler`. The callback functions are plain C functions and are not visible or callable at the SQL level. The callback functions are described in 第 50.2 节.

验证器函数负责验证 `CREATE` 和 `ALTER` 命令中指定的选项，这些命令可以针对其外部数据包装器，也可以针对使用该包装器的外部服务器、用户映射和外部表。注册验证器函数时，
必须声明它接受两个参数：一个 text 数组，包含待验证的选项；以及一个 OID，表示选项所属对象的类型（使用存放该对象的系统目录的 OID，即 `ForeignDataWrapperRelationId`、`ForeignServerRelationId`、`UserMappingRelationId` 或 `ForeignTableRelationId`）。如果未提供验证器函数，创建或修改对象时就不会检查选项。

50.2. 外部数据包装器回调例程

FDW 处理器函数返回一个用 palloc 分配的 `FdwRoutine` 结构体，其中包含指向下列回调函数的指针：

```
FdwPlan *
PlanForeignScan (Oid foreigntableid,
                 PlannerInfo *root,
                 RelOptInfo *baserel);
```

规划一个外部表上的扫描。这在查询被规划时调用。`foreigntableid` 是外部表的 `pg_class` OID。`root` 是规划器关于该查询的全局信息，而 `baserel` 是规划器关于这个表的信息。该函数必须返回一个用 palloc 分配的结构体，其中包含代价估计以及以后执行外部扫描所需的任何 FDW 私有信息。（注意，私有信息必须以 `copyObject` 会复制的形式表示。）

root 和 baserel 中的信息可用来减少必须从外部表取回的信息量（从而降低代价估计）。baserel->baserestrictinfo 尤其值得注意，它包含可用于过滤要取回的行的限制条件（WHERE 子句）。（并不要求 FDW 强制执行这些条件，因为完成的计划无论如何都会重新检查它们。）baserel->reltargetlist 可用来确定需要取回哪些列。

除了返回代价估计之外，该函数还应在考虑限制条件所做的过滤之后，把 baserel->rows 更新为扫描预期返回的行数。baserel->rows 的初始值只是一个常量默认估计，应当尽可能替换掉它。如果该函数能计算出平均结果行宽度的更好估计，也可以选择更新 baserel->width。

```
void
ExplainForeignScan (ForeignScanState *node,
                    ExplainState *es);
```

为外部表扫描打印额外的 EXPLAIN 输出。如果不需要打印任何内容，直接返回即可。否则，它应调用 ExplainPropertyText 及相关函数向 EXPLAIN 输出中添加字段。es 中的标志字段可用来确定要打印什么，而在 EXPLAIN ANALYZE 的情形下，可以检查 ForeignScanState 节点的状态以提供运行时统计信息。

```
void
BeginForeignScan (ForeignScanState *node,
                  int eflags);
```

开始执行一次外部扫描。这在执行器启动期间被调用。它应执行扫描开始前所需的任何初始化，但不要开始执行实际的扫描（那应该在第一次调用 IterateForeignScan 时进行）。ForeignScanState 节点已经创建，但其 fdw_state 字段仍为 NULL。有关被扫描表的信息可以通过 ForeignScanState 节点访问（特别是通过底层的 ForeignScan 计划节点，它包含指向由 PlanForeignScan 返回的 FdwPlan 结构的指针）。

注意，当(eflags & EXEC_FLAG_EXPLAIN_ONLY)为真时，这个函数不应执行任何外部可见的动作；它只应做使节点状态对ExplainForeignScan和EndForeignScan有效所需的最少工作。

```
TupleTableSlot *
IterateForeignScan (ForeignScanState *node);
```

从外部数据源取得一行，并放入元组表槽中返回（应使用节点的 ScanTupleSlot）。如果没有更多行，则返回 NULL。元组表槽机制允许返回物理元组或虚拟元组；从性能角度看，大多数情况下后者更合适。注意，此函数在短期内存上下文中调用，该上下文会在各次调用之间重置。如果需要更长久的存储，请在 BeginForeignScan 中创建内存上下文，或者使用 es_query_cxt，它属于节点的 EState。

返回的行必须与被扫描外部表的列签名匹配。如果选择优化掉不需要取回的列，则应当在这些列的位置上填入空值。

注意，PostgreSQL 的执行器并不关心返回的行是否违反外部表列上定义的 NOT NULL 约束——但规划器关心这一点，如果一个声明为不包含 NULL 值的列中出现了 NULL 值，规划器可能会错误地优化查询。当用户已声明不应出现 NULL 值却遇到了 NULL 值时，引发一个错误可能是恰当的做法（正如在数据类型不匹配的情况下你需要做的那样）。

```
void
ReScanForeignScan (ForeignScanState *node);
```

从头重新扫描。注意，扫描所依赖的参数值可能已经改变，因此新扫描不一定返回完全相同的行。

```
void
```

```
EndForeignScan (ForeignScanState *node);
```

结束扫描并释放资源。通常无需专门释放 palloc 分配的内存，但应清理打开的文件、到远程服务器的连接等资源。

`FdwRoutine` 和 `FdwPlan` 结构类型声明在 `src/include/foreign/fdwapi.h` 中，更多细节请参见该文件。

第 51 章 遗传查询优化器

Martin Utesch, University of Mining and Technology

作者

由 Martin Utesch <utesch@aut.tu-freiberg.de> 为德国弗赖贝格矿业和技术大学自动控制研究所编写。

51.1. 将查询处理看成是一个复杂的优化问题

在所有关系操作符中，最难处理和优化的是连接。随着查询中连接数目的增加，可能的查询计划数量会呈指数增长。为了处理单个连接而支持多种连接方法（例如 PostgreSQL 中的嵌套循环、哈希连接和归并连接），以及作为关系访问路径的多种索引（例如 PostgreSQL 中的 B-树、哈希、GiST 和 GIN），也进一步增加了优化工作量。

常规的PostgreSQL查询优化器会在可选策略空间中执行近似穷举搜索。该算法最早出现在 IBM 的 System R 数据库中，能够产生近乎最优的连接顺序；但当查询中的连接数变得很大时，它可能耗费极大的时间和内存空间。因此，普通的PostgreSQL查询优化器并不适合需要连接大量表的查询。

德国弗赖贝格矿业和技术大学自动控制研究所在尝试将 PostgreSQL 用作一个用于电网维护的基于知识的决策支持系统后端时遇到了一些问题。该 DBMS 需要为该基于知识系统中的推理机处理大型连接查询。这些查询中的连接数量使得使用普通查询优化器变得不可行。

下文将介绍一种遗传算法的实现，它以适合涉及大量连接的查询的高效方式来解决连接顺序问题。

51.2. 遗传算法

遗传算法（GA）是一种通过随机化搜索进行工作的启发式优化方法。优化问题的可能解集合被视为由若干个体组成的种群。个体对其环境的适应程度由其适应度表示。

一个个体在搜索空间中的坐标由染色体表示，本质上是一组字符串。基因是染色体的一个片段，它编码某个待优化参数的值。基因的典型编码可以是二进制或整数。

通过模拟重组、变异和选择这些进化操作，可以找到平均适应度高于前代的新一代搜索点。

根据comp.ai.genetic FAQ 中的说法，再怎么强调也不过分：GA并不是为了求解问题而进行的纯粹随机搜索。GA会使用随机过程，但其结果显然并非随机的（优于随机）。

图 51.1. 遗传算法的结构图

P(t)	generation of ancestors at a time t
P''(t)	generation of descendants at a time t

```
+==========================================+
|>>>>>>>>>>>  Algorithm GA  <<<<<<<<<<<<<<|
+==========================================+
| INITIALIZE t := 0                        |
+==========================================+
| INITIALIZE P(t)                          |
+==========================================+
| evaluate FITNESS of P(t)                 |
+==========================================+
| while not STOPPING CRITERION do          |
|   +--------------------------------------+
|   | P'(t)  := RECOMBINATION{P(t)}        |
|   +--------------------------------------+
|   | P''(t) := MUTATION{P'(t)}            |
|   +--------------------------------------+
|   | P(t+1) := SELECTION{P''(t) + P(t)}   |
|   +--------------------------------------+
|   | evaluate FITNESS of P''(t)           |
|   +--------------------------------------+
|   | t := t + 1                           |
+===+======================================+
```

51.3. PostgreSQL 中的遗传查询优化（GEQO）

GEQO模块把查询优化问题视为著名的旅行商问题（TSP）来处理。
可能的查询计划被编码为整数串。每个串表示查询中从一个关系到下一个关系的连接顺序。
例如，连接树

```
   /\
  /\ 2
 /\ 3
4  1
```

会被编码为整数串 '4-1-3-2'，这意味着先连接关系 '4' 和 '1'，再连接 '3'，最后连接 '2'；其中 1、2、3、4 是PostgreSQL优化器内部的关系 ID。

PostgreSQL 中 GEQO 实现的具体特征包括：

- 采用稳态 GA（替换种群中适应度最低的个体，而不是整代替换），能够快速收敛到更优的查询计划。这对于在合理时间内处理查询至关重要；
- 采用边重组交叉，它特别适合在利用 GA 求解 TSP 时将边损失保持在较低水平；
- 不使用变异作为遗传操作符，因此无须借助修复机制来生成合法的 TSP 回路。

GEQO模块的部分内容改编自 D. Whitley 的 Genitor 算法。

GEQO模块使PostgreSQL查询优化器能够通过非穷举搜索有效支持大型连接查询。

51.3.1. 使用GEQO生成可能的计划

GEQO 规划过程使用标准规划器代码来生成各个关系扫描的计划，然后再用遗传方法生成连接计划。如上所示，每个候选连接计划都表示为一个连接基本关系的顺序。在初始阶段，GEQO 代码只是随机生成一些可能的连接序列。对于每一个待考察的连接序列，
都会调用标准规划器代码来估算按该序列执行查询的代价。（对于连接序列中的每一步，
都会考虑全部三种可能的连接策略；并且所有预先确定的关系扫描计划也都可用。
估计代价取这些可能性中最低者。）
估计代价较低的连接序列会被认为比代价较高的序列“更适合”。
遗传算法会丢弃最不适合的候选，然后通过组合更适合候选的基因来生成新的候选，
也就是从已知低代价连接序列中随机选取片段，构造新的序列以供考察。这个过程会重复进行，
直到已经考察的连接序列数量达到预设值；然后，
在搜索过程中任何时刻找到的最佳序列都会被用来生成最终计划。

由于初始种群选择以及后续对最佳候选的“变异”都包含随机选择，
因此这一过程天生就是非确定性的。为了避免所选计划发生令人意外的变化，每次运行 GEQO 算法时，都会用当前的geqo_seed参数设置重新初始化其随机数生成器。只要 `geqo_seed`、其他 GEQO 参数以及统计信息等其他规划器输入保持不变，对于给定查询就会生成相同的计划。若想试验不同的搜索路径，可以尝试修改 `geqo_seed`。

51.3.2. PostgreSQL GEQO的未来实现任务

为了改进遗传算法的参数设置，仍有一些工作要做。在文件`src/backend/optimizer/geqo/geqo_main.c`中的例程`gimme_pool_size`和`gimme_number_generations`里，
我们必须为参数设置找到一种折中，以满足两个相互竞争的需求：

- 查询计划的最优性
- 计算时间

在当前实现中，
每个候选连接序列的适应度都是通过从头运行标准规划器的连接选择和代价估算代码来估算的。
只要不同候选使用了相似的连接子序列，就会有大量工作被重复执行。
如果能保留子连接的代价估算，就可以显著加快这一过程。问题在于，
必须避免为了保存这种状态而消耗不合理数量的内存。

从更基础的层面看，用一个为 TSP 设计的 GA 算法来解决查询优化问题是否合适，也并不明确。在 TSP 情况下，与任何子串（部分巡回）相关的代价都独立于巡回的其余部分，但对于查询优化显然并非如此。因此，边重组交叉是否是最有效的变异过程，仍然值得怀疑。

51.4. 进一步阅读

下列资源包含关于遗传算法的更多信息：

- 《Hitch-Hiker 演化计算指南》[1]（comp.ai.genetic[2] FAQ）
- 《演化计算及其在艺术和设计中的应用》[3]，Craig Reynolds 著
- [elma04]
- [fong]

[1] http://www.aip.de/~ast/EvolCompFAQ/
[2] news://comp.ai.genetic
[3] http://www.red3d.com/cwr/evolve.html

第 52 章 索引访问方法接口定义

本章定义 PostgreSQL 核心系统与管理各个索引类型的索引访问方法之间的接口。除这里规定的内容之外，核心系统对索引一无所知，因此可以通过编写附加代码来开发全新的索引类型。

PostgreSQL 中所有索引在技术上都称为二级索引；也就是说，索引与它所描述的表文件在物理上是分离的。每个索引都存储为独立的物理关系，因此在 pg_class 目录中都有一个条目。索引的内容完全由其索引访问方法控制。实践中，
所有索引访问方法都会把索引划分为标准大小的页面，
以便使用常规的存储管理器和缓冲区管理器访问索引内容。（现有的所有索引访问方法还都使用第 55.6 节中描述的标准页面布局，而且它们都对索引元组头使用相同的格式；但这些决定并不是访问方法必须遵循的。）

索引本质上是从一些数据键值到行版本（元组）在索引父表中的元组标识符（TIDs）的映射。
一个 TID 由块号以及该块中的项号组成（见第 55.6 节）。这些信息足以从表中取出某个特定的行版本。索引并不直接知道在 MVCC 之下同一逻辑行可能会存在多个现存版本；对索引来说，每个元组都是一个独立对象，需要它自己的索引条目。因此，
对一行的更新总会为该行创建全新的索引条目，即使键值并未改变也是如此。（HOT 元组是这一说法的例外；但索引同样不直接处理它们。）当死元组自身被回收时（通过清理），
它们对应的索引条目也会被回收。

52.1. 索引的目录条目

每种索引访问方法都由系统目录 pg_am 中的一行描述（参见第 45.3 节）。pg_am 行的主要内容是一些指向 pg_proc 条目的引用，它们标识了该访问方法所提供的索引访问函数。这些函数的 API 将在本章后面定义。此外，pg_am 行还指定了访问方法的一些固定属性，例如它是否支持多列索引。目前没有任何专门用于创建或删除 pg_am 条目的支持；能够编写新访问方法的人，理应有能力自己插入合适的行。

要使一种索引访问方法真正可用，还必须在 pg_opfamily、pg_opclass、pg_amop 和 pg_amproc 中定义一个或多个操作符族和操作符类。这些条目使规划器能够确定该访问方法的索引可以使用哪些查询限定条件。关于操作符族和操作符类的说明见第 35.14 节；阅读本章前应先掌握这些内容。

单个索引由两个系统目录条目共同定义：一个 pg_class 条目把它描述为一个物理关系，另一个 pg_index 条目给出索引的逻辑内容，也就是它包含哪些索引列，以及由相关操作符类定义的这些列的语义。索引列（键值）既可以是底层表的简单列，也可以是基于表行的表达式。
索引访问方法通常并不关心索引键值来自何处（它拿到的总是预先计算好的键值），
但会非常关心 pg_index 中的操作符类信息。这两个目录条目都可以作为传递给索引上所有操作的 Relation 数据结构的一部分来访问。

pg_am 的某些标志列还有一些不那么直观的含义。amcanunique 的要求见第 52.5 节。amcanmulticol 标志表明该访问方法支持多列索引，而 amoptionalkey 表明它允许在第一索引列上没有给出可索引限制子句时进行扫描。当 amcanmulticol 为假时，amoptionalkey 实质上表示该访问方法是否支持没有任何限制子句的全索引扫描。支持多个索引列的访问方法必须支持在第一列之后省略任意一个或全部列限制条件的扫描；
但它们可以要求第一索引列必须出现某种限制条件，这通过把 amoptionalkey 设为假来表示。索引 AM 可能把 amoptionalkey 设为假的一个原因，是它不索引 NULL 值。由于大多数可索引操作符都是严格的（strict），因此对 NULL 输入不可能返回 TRUE，所以乍看之下不存储 NULL 值的索引条目似乎很有吸引力：无论如何，这些条目似乎都不可能被索引扫描返回。然而，
当某个索引扫描对给定索引列没有限制子句时，这个论证就不成立了。实践中这意味着，
若索引把 amoptionalkey 设为真，就必须索引 NULL 值，因为规划器可能决定在完全没有扫描键的情况下使用这种索引。与此相关的另一个限制是，支持多个索引列的索引访问方法

必须支持对第一列之后各列中的 NULL 值建立索引，因为规划器会假定该索引可用于不限制这些列的查询。例如，考虑一个在 (a,b) 上的索引以及查询 `WHERE a = 4`。系统会假定该索引可用于扫描满足 `a = 4` 的行；如果索引省略了 `b` 为 NULL 的行，这个假定就是错误的。不过，省略第一索引列为 NULL 的行是可以的。对 NULL 值建立索引的索引访问方法还可以设置 `amsearchnulls`，表示它支持把 `IS NULL` 和 `IS NOT NULL` 子句作为搜索条件。

52.2. 索引访问方法函数

索引访问方法必须提供如下索引构建和维护函数：

```
IndexBuildResult *
ambuild (Relation heapRelation,
         Relation indexRelation,
         IndexInfo *indexInfo);
```

构建新索引。索引关系已在物理上创建，但内容为空。必须填入访问方法所需的固定数据，以及表中所有已有元组对应的条目。通常，`ambuild`函数会调用`IndexBuildHeapScan()`扫描表中现有的元组，并计算需要插入索引的键。函数必须返回一个通过 palloc 分配的结构体，其中包含新索引的统计信息。

```
void
ambuildempty (Relation indexRelation);
```

构建一个空索引，并将其写入给定关系的初始化分支（INIT_FORKNUM）。只有不记录 WAL 的表才会调用此方法；写入初始化分支的空索引会在每次服务器重启时复制到主关系分支上。

```
bool
aminsert (Relation indexRelation,
          Datum *values,
          bool *isnull,
          ItemPointer heap_tid,
          Relation heapRelation,
          IndexUniqueCheck checkUnique);
```

将一个新元组插入已有索引。`values`和`isnull`数组给出要索引的键值，`heap_tid`则是要索引的 TID。如果访问方法支持唯一索引（其`pg_am.amcanunique`标志为 true），则`checkUnique`指明要执行的唯一性检查类型，具体取决于唯一约束是否可延迟；详情参见第 52.5 节。通常，访问方法只在执行唯一性检查时才需要`heapRelation`参数（因为此时必须查看堆以确认元组是否存活）。

只有当`checkUnique`为`UNIQUE_CHECK_PARTIAL`时，该函数的布尔返回值才有意义。在这种情况下，TRUE 表示已知新条目唯一，而 FALSE 表示它可能不唯一（并且必须安排一次延迟唯一性检查）。对于其他情况，建议始终返回 FALSE。

某些索引可能不会为所有元组建立索引。如果某个元组不应被索引，`aminsert` 应当直接返回而不做任何事。

```
IndexBulkDeleteResult *
ambulkdelete (IndexVacuumInfo *info,
              IndexBulkDeleteResult *stats,
              IndexBulkDeleteCallback callback,
              void *callback_state);
```

从索引中删除元组。这是一个“批量删除”操作，旨在通过扫描整个索引并检查每个条目是否应被删除来实现。必须调用传入的 `callback` 函数，其调用形式为 `callback(TID, callback_state) returns bool`，以确定由其引用的 TID

所标识的某个索引条目是否应删除。该函数必须返回 NULL，或者返回一个通过 palloc 分配的结构体，其中包含此次删除操作影响的统计信息。如果不需要向 amvacuumcleanup 传递信息，返回 NULL 也是可以的。

由于 maintenance_work_mem 有限，当待删除的元组很多时，ambulkdelete 可能需要被调用多次。参数 stats 是此前对该索引上一次调用的结果（在一次 VACUUM 操作中的第一次调用时它为 NULL）。这使得 AM 可以在整个操作过程中累积统计信息。通常，如果传入的 stats 非 NULL，ambulkdelete 会修改并返回同一个结构体。

```
IndexBulkDeleteResult *
amvacuumcleanup (IndexVacuumInfo *info,
                 IndexBulkDeleteResult *stats);
```

在一次 VACUUM 操作（零次或多次 ambulkdelete 调用）之后执行清理。它不一定要做返回索引统计信息之外的事情，但也可能执行批量清理，例如回收空索引页面。stats 是最后一次 ambulkdelete 调用返回的结果；如果因为没有需要删除的元组而没有调用 ambulkdelete，则为 NULL。如果结果不是 NULL，它必须是一个通过 palloc 分配的结构体。其中的统计信息将用于更新 pg_class，并在指定了 VERBOSE 时由 VACUUM 报告。如果索引在整个 VACUUM 操作期间完全没有变化，返回 NULL 也是可以的；否则应返回正确的统计信息。

从PostgreSQL 8.4 开始，amvacuumcleanup 也会在 ANALYZE 操作完成时被调用。在这种情况下，stats 总是 NULL，而且任何返回值都会被忽略。这种情况可以通过检查 info->analyze_only 来区分。我们建议访问方法在这种调用中除了做插入后的清理外什么都不做，而且只在自动清理工作进程中这样做。

```
void
amcostestimate (PlannerInfo *root,
                IndexOptInfo *index,
                List *indexQuals,
                List *indexOrderBys,
                RelOptInfo *outer_rel,
                Cost *indexStartupCost,
                Cost *indexTotalCost,
                Selectivity *indexSelectivity,
                double *indexCorrelation);
```

估计索引扫描的代价。该函数在下文第 52.6 节中有完整描述。

```
bytea *
amoptions (ArrayType *reloptions,
           bool validate);
```

解析并验证索引的 reloptions 数组。只有当该索引存在非 NULL 的 reloptions 数组时才会调用此函数。*reloptions* 是一个 text 数组，其中的条目形如 *name=value*。该函数应构造一个 bytea 值，并将其复制到索引 relcache 条目的 rd_options 字段中。这个 bytea 值中的数据内容由访问方法自行定义；大多数标准访问方法使用结构体 StdRdOptions。当 *validate* 为真时，若存在未识别选项或无效取值，函数应报告合适的错误消息；当 *validate* 为假时，无效条目应被静默忽略。（当装载已存储在 pg_catalog 中的选项时，*validate* 为假；此时只有在访问方法改变了选项规则时才可能发现无效条目，而忽略过时条目是合适的。）如果希望采用默认行为，返回 NULL 也是可以的。

索引的目的当然是支持扫描那些匹配可索引 WHERE 条件的元组，这种条件常被称为限定词或扫描键。关于索引扫描的语义，将在下面的第 52.3 节中更详细地说明。一种索引访问方法可以支持“普通”索引扫描、“位图”索引扫描，或者两者都支持。
索引访问方法必须或可以提供的扫描相关函数如下：

```
IndexScanDesc
ambeginscan (Relation indexRelation,
             int nkeys,
             int norderbys);
```

为一次索引扫描做准备。nkeys 和 norderbys 参数表示扫描中将使用的限定条件和排序操作符数量，这些信息可能有助于空间分配。请注意，此时还没有提供扫描键的实际值。结果必须是一个通过 palloc 分配的结构体。出于实现上的原因，索引访问方法必须通过调用 RelationGetIndexScan() 来创建这个结构体。大多数情况下，ambeginscan 除了做这次调用以及也许获取一些锁之外，不会做太多工作；索引扫描启动中真正有意思的部分在 amrescan 中。

```
void
amrescan (IndexScanDesc scan,
          ScanKey keys,
          int nkeys,
          ScanKey orderbys,
          int norderbys);
```

开始或重新开始一次索引扫描，并且可以使用新的扫描键。（若要用之前传入的键重新开始，则为 keys 和/或 orderbys 传递 NULL。）请注意，键的数量或排序操作符的数量都不允许超过传给 ambeginscan 的数量。实践中，这种重启特性通常用于嵌套循环连接选中了新的外层元组，因此需要新的键比较值，但扫描键结构体本身保持不变的场景。

```
boolean
amgettuple (IndexScanDesc scan,
            ScanDirection direction);
```

沿给定方向（在索引中向前或向后）取出给定扫描中的下一个元组。若取得元组则返回 TRUE，若没有剩余匹配元组则返回 FALSE。返回 TRUE 时，元组 TID 存入scan结构体中。注意，“成功”仅表示索引中包含与扫描键匹配的条目，并不表示该元组一定仍存在于堆中，或者能够通过调用方的快照测试。成功时，amgettuple还必须将scan->xs_recheck设为 TRUE 或 FALSE。FALSE 表示可以确定该索引条目匹配扫描键。TRUE 表示无法确定，因此在取出堆元组后必须重新检查扫描键所代表的条件。这一规定支持“有损”索引操作符。注意，重新检查仅针对扫描条件；部分索引谓词（如果有）永远不会由amgettuple的调用方重新检查。

只有当访问方法支持“普通”索引扫描时，才需要提供 amgettuple 函数。如果不支持，其 pg_am 行中的 amgettuple 字段必须设为零。

```
int64
amgetbitmap (IndexScanDesc scan,
             TIDBitmap *tbm);
```

取出给定扫描中的所有元组，并将其加入调用者提供的 TIDBitmap 中（也就是把这组元组 ID 与位图中已有的集合做 OR）。返回值是取得的元组数量（这可能只是近似计数，例如某些 AM 不会检测重复项）。在把元组 ID 插入位图时，amgetbitmap 可以指出某些具体的元组 ID 需要重新检查扫描条件。这类似于 amgettuple 的 xs_recheck 输出参数。注意：在当前实现中，对这一特性的支持与位图自身的有损存储支持混在一起，因此调用者会对需要重检的元组同时重新检查扫描条件和部分索引谓词（如果有）。不过，这并不一定永远如此。amgetbitmap 和 amgettuple 不能在同一次索引扫描中同时使用；使用 amgetbitmap 时还有其他限制，详见第 52.3 节。

只有当访问方法支持“位图”索引扫描时，才需要提供 amgetbitmap 函数。如果不支持，其 pg_am 行中的 amgetbitmap 字段必须设为零。

```
void
amendscan (IndexScanDesc scan);
```

结束扫描并释放资源。scan 结构体本身不应被释放，但访问方法内部获取的任何锁或引用（pin）都必须被释放。

```
void
ammarkpos (IndexScanDesc scan);
```

标记当前扫描位置。访问方法只需要为每次扫描支持一个被记住的扫描位置。

```
void
amrestrpos (IndexScanDesc scan);
```

将扫描恢复到最近一次标记的位置。

按照惯例，索引访问方法函数的 pg_proc 条目应显示正确的参数个数，但把所有参数都声明为 internal 类型（因为大多数参数的类型对 SQL 而言是未知的，而且我们也不希望用户直接调用这些函数）。返回类型根据实际情况声明为 void、internal 或 boolean。唯一的例外是 amoptions，它应当正确地声明为接受 text[] 和 bool 并返回 bytea。这一规定允许客户端代码执行 amoptions 来测试选项设置的有效性。

52.3. 索引扫描

在索引扫描中，索引访问方法负责返回它所知道的、匹配扫描键的所有元组的 TID。访问方法不负责从索引父表中实际取出这些元组，
也不负责判断它们是否能通过扫描的可见性测试或其他条件。

扫描键是如下形式的 WHERE 子句的内部表示：*index_key operator constant*。其中，索引键是该索引的一列，而操作符是与该索引列关联的操作符族成员之一。
一次索引扫描可以有零个或多个扫描键，这些扫描键会被隐式地以 AND 连接，也就是说，返回的元组预期应满足所有给定条件。

对于某个特定查询，访问方法可以报告该索引是有损的，或者要求重新检查。
这意味着索引扫描会返回所有通过扫描键的条目，并可能额外返回一些未通过扫描键的条目。
随后，核心系统的索引扫描机制会再次把索引条件应用到堆元组上，
以验证它是否真的应被选中。如果没有指定重检选项，
索引扫描就必须精确返回匹配条目的集合。

请注意，确保正确找出通过全部给定扫描键的所有且仅有的条目，
这项工作完全由访问方法负责。此外，核心系统只是简单地把所有与索引键和操作符族匹配的 WHERE 子句交给访问方法，而不会做任何语义分析来判断它们是否冗余或互相矛盾。例如，给定 WHERE x > 4 AND x > 14，其中 x 是一个 B-树索引列，那么由 B-树的 amrescan 函数负责识别第一个扫描键是冗余的并可被丢弃。amrescan 期间需要做多少预处理，取决于索引访问方法需要在多大程度上把扫描键归约成一种“正规化”形式。

某些访问方法会按明确定义的顺序返回索引条目，另一些则不会。实际上，
访问方法支持有序输出有两种不同方式：

- 总是按数据自然顺序返回条目的访问方法（例如 btree）应把 pg_am.amcanorder 设为真。目前，这类访问方法必须对其等值和排序操作符使用与 btree 兼容的策略号。
- 支持排序操作符的访问方法应把 pg_am.amcanorderbyop 设为真。这表示索引能够按满足 ORDER BY *index_key operator constant* 的顺序返回条目。前面已经提到，这种形式的扫描修饰符可以传给 amrescan。

amgettuple 函数有一个 direction 参数，它可以是 ForwardScanDirection（通常情况）或 BackwardScanDirection。如果 amrescan 之后的第一次调用指定了 BackwardScanDirection，那么匹配条件的索引条目集合就要按从后到前的方向扫描，而不是通常的从前到后，因此 amgettuple 必须返回索引中最后一个匹配元组，而不是通常情况下的第一个。（这只会发生在把 amcanorder 设为真的访问方法上。）第一次调用之后，amgettuple 必须准备好从最近一次返回条目的位置继续按任一方向推进扫描。（但如果 pg_am.amcanbackward 为假，则后续所有调用的方向都必须与第一次相同。）

支持有序扫描的访问方法必须支持在扫描中“标记”一个位置，并在之后返回到该标记位置。
同一个位置可能会被恢复多次。不过，每次扫描只需要记住一个位置；新的 ammarkpos 调用会覆盖先前标记的位置。不支持有序扫描的访问方法仍应在 pg_am 中提供标记和恢复函数，但让这些函数在被调用时报错也是可以的。

无论是扫描位置还是标记位置（如果有），在面对索引中的并发插入或删除时都必须保持一致。
如果某个新插入的条目没有被一次扫描返回，
而假如该条目在扫描开始前就已存在则本应被返回，这也是可以接受的；同样，
即使某个条目在第一次扫描通过时未被返回，
后续重新扫描或反向扫描时返回它也是可以接受的。类似地，
并发删除也可能会或者不会反映在扫描结果中。重要的是，插入或删除本身不能导致扫描漏掉，
或者重复返回那些本身并未被插入或删除的条目。

Instead of using amgettuple, an index scan can be done with amgetbitmap to fetch all tuples in one call. This can be noticeably more efficient than amgettuple because it allows avoiding lock/unlock cycles within the access method. In principle amgetbitmap should have the same effects as repeated amgettuple calls, but we impose several restrictions to simplify matters. First of all, amgetbitmap returns all tuples at once and marking or restoring scan positions isn't supported. 其次，元组是以位图形式返回的，没有任何特定顺序，这也正是 amgetbitmap 不接受 direction 参数的原因。（这种扫描也永远不会提供排序操作符。）Finally, amgetbitmap does not guarantee any locking of the returned tuples, with implications spelled out in 第 52.4 节.

请注意，如果访问方法的内部实现并不适合某一个 API，那么它可以只实现 amgetbitmap 而不实现 amgettuple，反之亦然。

52.4. 索引锁定注意事项

索引访问方法必须处理多个进程并发更新同一个索引的情况。PostgreSQL 核心系统会在索引扫描期间对索引获取 AccessShareLock，并在更新索引时（包括普通的 VACUUM）获取 RowExclusiveLock。由于这些锁类型彼此不冲突，所以访问方法必须自行处理它可能需要的任何细粒度锁定。对整个索引的排他锁只会在索引创建、销毁或 REINDEX 期间获取。

构建一种支持并发更新的索引类型，通常需要对所需行为进行大量而细致的分析。对于 B-树和 hash 索引类型，可以阅读 src/backend/access/nbtree/README 和 src/backend/access/hash/README 中涉及的设计决策。

除了索引自身内部一致性的要求外，并发更新还会带来父表（即堆）与索引之间一致性的问题。
由于 PostgreSQL 把对堆的访问和更新与对索引的访问和更新分离开来，因此存在一些窗口期，
在这些窗口期内索引可能与堆不一致。我们通过以下规则处理这个问题：

- 先创建新的堆条目，再创建它对应的索引条目。
  （因此并发索引扫描很可能看不到这个堆条目。这是可以接受的，
  因为索引读取者本来也不会关心未提交的行。但也请见第 52.5 节。）

- 当某个堆条目将要被删除（由 VACUUM 执行）时，必须先删除它的所有索引条目。

- 索引扫描必须在保存 amgettuple 最近一次返回条目的索引页面上保持钉住状态（pin），而 ambulkdelete 不能从被其他后端钉住的页面中删除条目。下面会解释为什么需要这条规则。

Without the third rule, it is possible for an index reader to see an index entry just before it is removed by VACUUM, and then to arrive at the corresponding heap entry after that was removed by VACUUM. This creates no serious problems if that item number is still unused when the reader reaches it, since an empty item slot will be ignored by heap_fetch(). But what if a third backend has already re-used the item slot for something else? When using an MVCC-compliant snapshot, there is no problem because the new occupant of the slot is certain to be too new to pass the snapshot test. However, with a non-MVCC-compliant snapshot (such as SnapshotNow), it would be possible to accept and return a row that does not in fact match the scan keys. We could defend against this scenario by requiring the scan keys to be rechecked against the heap row in all cases, but that is too expensive. Instead, we use a pin on an index page as a proxy to indicate that the reader might still be "in flight" from the index entry to the matching heap entry. Making ambulkdelete block on such a pin ensures that VACUUM cannot delete the heap entry before the reader is done with it. This solution costs little in run time, and adds blocking overhead only in the rare cases where there actually is a conflict.

这种解决方案要求索引扫描是"同步的"：我们必须在扫描到相应索引条目之后立刻取出对应的堆元组。由于多种原因，这会比较昂贵。相反，"异步"扫描可以先从索引中收集许多 TID，再在稍后某个时刻访问堆元组，这样索引锁定的开销要小得多，也能实现更高效的堆访问模式。
根据前面的分析，对于非 MVCC 兼容快照，我们必须采用同步方式；而对于使用 MVCC 快照的查询，异步扫描则是可行的。

在 amgetbitmap 索引扫描中，访问方法不会为任何返回的元组保持索引页面的钉住状态。因此，只有把这种扫描与 MVCC 兼容的快照一起使用才是安全的。

当未设置 ampredlocks 标志时，在可串行化事务中使用该索引访问方法的任何扫描，都会在整个索引上获取一个非阻塞谓词锁。
这会与并发可串行化事务向该索引插入任何元组形成读写冲突。
如果在一组并发可串行化事务之间检测到某些特定模式的读写冲突，为了保护数据完整性，
其中某个事务可能会被取消。设置该标志则表示该索引访问方法实现了更细粒度的谓词锁，
这通常会降低这类事务取消的频率。

52.5. 索引唯一性检查

PostgreSQL 使用唯一索引来强制执行 SQL 唯一性约束；所谓唯一索引，就是不允许存在多个具有相同键值条目的索引。支持这一特性的访问方法会把 pg_am.amcanunique 设为真。（目前只有 B-树支持这一点。）

由于 MVCC 的存在，索引中在物理上总是必须允许存在重复条目：这些条目可能指向同一个逻辑行的连续版本。我们真正想强制的行为是，任何 MVCC 快照都不能同时包含两行拥有相同索引键的记录。在向唯一索引插入一行新数据时，必须检查以下几种情况：

- 如果某一条冲突的有效行已被当前事务删除，那么这是允许的。（特别是，因为一次 UPDATE 总是在插入新版本前删除旧版本，这就允许对某一行执行不改变键值的 UPDATE。）

- 如果某条冲突的行是由一个尚未提交的事务插入的，那么当前准备插入的事务必须等待，
  看看那个事务是否提交。如果它回滚，就不存在冲突；
  如果它提交并且没有再次删除那条冲突行，就发生了唯一性违背。（实际上，
  我们只是等待另一个事务结束，然后把可见性检查整个重新做一遍。）

- 类似地，如果某条冲突的有效行是由一个尚未提交的事务删除的，
  那么当前准备插入的事务必须等待该事务提交或中止，然后重新执行测试。

此外，就在按照上述规则报告唯一性违背之前，
访问方法必须重新检查正在插入那一行的存活性。如果它已经是提交后死亡的状态，
就不应报告违背。（这种情况不会出现在插入由当前事务刚创建的行这一普通场景中，但在 `CREATE UNIQUE INDEX CONCURRENTLY` 期间却可能发生。）

我们要求索引访问方法自行应用这些测试，
这意味着它必须深入堆中检查那些根据索引内容显示为具有重复键的行的提交状态。毫无疑问，
这样做既丑陋又不够模块化，但它避免了重复工作：如果我们单独再做一次探测，
那么在寻找新行索引条目插入位置时，查找冲突行的索引搜索实际上就会被重复执行。更何况，
除非把冲突检查作为插入新索引条目动作的一个组成部分，
否则也没有明显的方法可以避免竞争条件。

如果唯一约束是可延迟的，情况会更复杂：我们需要能够为新行插入一个索引条目，
但把任何唯一性违背错误延迟到语句结束时甚至更晚才报告。
为了避免对索引进行不必要的重复搜索，
索引访问方法应在初始插入期间执行一次初步唯一性检查。
如果这表明确实不存在冲突的存活元组，那么事情就结束了。否则，
我们会安排在真正强制约束时再做一次重检。若在重检时，
插入的元组与另外某个具有相同键值的元组都仍然存活，就必须报告错误。（注意，
就此用途而言，“存活”实际上是指“索引条目 HOT 链中至少有一个元组是存活的”。）为实现这一点，传给 `aminsert` 函数的 `checkUnique` 参数会取以下值之一：

- `UNIQUE_CHECK_NO` 表示不应执行唯一性检查（这不是唯一索引）。

- `UNIQUE_CHECK_YES` 表示这是一个不可延迟的唯一索引，必须按前述方式立即执行唯一性检查。

- `UNIQUE_CHECK_PARTIAL` 表示该唯一约束是可延迟的。PostgreSQL 会使用此模式为每一行插入索引条目。访问方法必须允许索引中出现重复条目，并通过让 `aminsert` 返回 FALSE来报告任何潜在的重复。对于每一行返回 FALSE的情况，系统都会安排一次延迟重检。

  访问方法必须识别任何可能违反唯一约束的行，但报告误报并不算错误。
  这使得检查可以在不等待其他事务结束的情况下完成；此处报告的冲突不会被当作错误处理，
  而会在之后重新检查，到那时它们可能已不再冲突。

- `UNIQUE_CHECK_EXISTING` 表示这是对某一行的延迟重检，此前该行被报告为可能违反唯一性。虽然这同样是通过调用 `aminsert` 来实现，但访问方法在这种情况下不得插入新的索引条目。该索引条目已经存在。相反，访问方法必须检查是否存在另一个存活的索引条目；
  如果存在，并且目标行本身也仍然存活，就应报告错误。

  建议在一次 `UNIQUE_CHECK_EXISTING` 调用中，访问方法进一步验证目标行在索引中确实已有现存条目；若没有，就报告错误。这样做是个好主意，因为传给 `aminsert` 的索引元组值会被重新计算。如果索引定义涉及并非真正 immutable 的函数，我们就可能在索引的错误区域上做检查。确认重检时确实找到了目标行，
  可以验证我们扫描的仍是原始插入时使用的同一组元组值。

52.6. 索引代价估算函数

`amcostestimate` 函数会收到描述某种可能索引扫描方式的信息，其中包括已经确定可用于该索引的 `WHERE` 子句和 `ORDER BY` 子句列表。它必须返回访问该索引的代价估算，以及 `WHERE` 子句选择率的估计值（也就是在索引扫描期间将从父表中检索出的行所占比例）。对于简单情况，代价估算器几乎所有工作都可以通过调用优化器中的标准例程来完成；之所以提供 `amcostestimate` 函数，是为了让索引访问方法能够提供与索引类型有关的专门知识，以便在可能时改进标准估计。

每个 amcostestimate 函数都必须具有下面的签名：

```
void
amcostestimate (PlannerInfo *root,
                IndexOptInfo *index,
                List *indexQuals,
                List *indexOrderBys,
                RelOptInfo *outer_rel,
                Cost *indexStartupCost,
                Cost *indexTotalCost,
                Selectivity *indexSelectivity,
                double *indexCorrelation);
```

The first five parameters are inputs:

root

The planner's information about the query being processed.

index

正在被考虑的索引。

indexQuals

索引限制条件子句的列表（隐含地做了 AND）；NIL 列表表示没有可用的限制条件。注意该列表包含顶部带 RestrictInfo 节点的表达式树，而不是裸表达式。

indexOrderBys

可索引的 ORDER BY 操作符的列表，若无则为 NIL。注意该列表包含表达式树，而不是 Scan Key。

outer_rel

如果该索引正被考虑用于连接的内层索引扫描，则为规划器关于连接外侧的信息；否则为 NULL。当不为 NULL 时，部分限定子句将是与该关系的连接子句，而不是简单的限制子句。此外，代价估算器应预期该索引扫描会对外层关系的每一行重复一次。

最后四个参数是按引用传递的输出参数：

**indexStartupCost*

设为索引启动处理的代价。

**indexTotalCost*

设为索引处理的总代价。

**indexSelectivity*

设为索引选择率。

**indexCorrelation*

设为索引扫描顺序与底层表顺序之间的相关系数。

请注意，代价估算函数必须用 C 编写，而不能用 SQL 或任何可用的过程语言，因为它们必须访问规划器/优化器的内部数据结构。

索引访问代价应使用 `src/backend/optimizer/path/costsize.c` 所用的参数来计算：顺序磁盘块访问的代价为 `seq_page_cost`，非顺序访问的代价为 `random_page_cost`，而处理一个索引行的代价通常取为 `cpu_index_tuple_cost`。此外，对于索引处理（尤其是 `indexQuals` 本身的求值）期间调用的每个比较操作符，都应计入 `cpu_operator_cost` 的适当倍数。

访问代价应包括与扫描索引本身有关的全部磁盘和 CPU 代价，但不包括取出或处理由索引标识出的父表行的代价。

“启动代价”是整个扫描总代价中必须在开始取第一行之前先付出的那一部分。
对大多数索引来说，这可以视为零；但启动代价较高的索引类型可能希望把它设为非零。

`indexSelectivity` 应设为在索引扫描期间将从父表中检索出的行的估计比例。对于有损查询，这个值通常会高于实际通过给定限定条件的行比例。

`indexCorrelation` 应设为索引顺序与表顺序之间的相关性（范围从 -1.0 到 1.0）。该值用于调整从父表取行代价的估计。

在连接的情况下，返回的数值应是该索引任意一次扫描的预期平均值。

代价估算

一个典型的代价估算器会按如下步骤进行：

1. 根据给定的限定条件，估计并返回将被访问的父表行所占的比例。
   在没有索引类型相关知识的情况下，使用标准优化器函数 `clauselist_selectivity()`：

   ```
   *indexSelectivity = clauselist_selectivity(root, indexQuals,
                                              index->rel->relid,
                                              JOIN_INNER, NULL);
   ```

2. 估计扫描期间将被访问的索引行数。对于许多索引类型，它等于 *`indexSelectivity`* 乘以索引中的行数，但也可能更多。（注意，索引的大小（页数和行数）可以从 `IndexOptInfo` 结构体中获得。）

3. 估计扫描期间将读取的索引页面数。它可能仅仅是 *`indexSelectivity`* 乘以索引总页面数。

4. 计算索引访问代价。一个通用估计器可能会这样做：

   ```
   /*
    * Our generic assumption is that the index pages will be read
    * sequentially, so they cost seq_page_cost each, not random_page_
   cost.
    * Also, we charge for evaluation of the indexquals at each index
    row.
    * All the costs are assumed to be paid incrementally during the
    scan.
    */
   cost_qual_eval(&index_qual_cost, indexQuals, root);
   *indexStartupCost = index_qual_cost.startup;
   *indexTotalCost = seq_page_cost * numIndexPages +
   ```

```
    (cpu_index_tuple_cost + index_qual_cost.per_tuple) * numIndex
Tuples;
```

不过，上述做法没有考虑连接情况下重复索引扫描之间索引读取的摊销效果。

5. 估计索引的相关性。对于单列上的简单有序索引，这个值可以从 pg_statistic 中取得。如果相关性未知，保守估计应为零（即无相关性）。

代价估算器函数的示例可在 `src/backend/utils/adt/selfuncs.c` 中找到。

第 53 章 GiST 索引

53.1. 简介

GiST是 Generalized Search Tree（通用搜索树）的缩写。它是一种平衡的树形访问方法，可作为实现任意索引方案的基本模板。B-树、R 树以及许多其他索引方案都可以在GiST中实现。

GiST的一个优点是，它使数据类型所属领域的专家而不是数据库专家，
能够开发带有适当访问方法的自定义数据类型。

这里的一些内容来自加州大学伯克利分校的 GiST 索引项目网站[1]以及 Marcel Kornacker 的论文 Access Methods for Next-Generation Database Systems[2]。PostgreSQL中的GiST 实现主要由 Teodor Sigaev 和 Oleg Bartunov 维护，他们的网站[3]上还有更多信息。

53.2. 可扩展性

传统上，实现一种新的索引访问方法意味着大量艰难的工作。必须理解数据库的内部机制，
例如锁管理器和预写式日志。GiST接口具有很高的抽象层次，
只要求访问方法实现者实现被访问数据类型的语义。GiST层本身会处理并发、
日志记录以及树结构的搜索。

这种可扩展性不应与其他标准搜索树在可处理数据方面的可扩展性相混淆。例如，
PostgreSQL支持可扩展的 B-树和 hash 索引。这意味着你可以用PostgreSQL在任意数据类型上构建 B-树或 hash 索引。但 B-树只支持范围谓词（<、=、>），而 hash 索引只支持等值查询。

因此，如果你用PostgreSQL的 B-树为一个图像集合建立索引，你只能发出诸如“imagex 是否等于 imagey”、“imagex 是否小于 imagey”以及“imagex 是否大于 imagey”之类的查询。取决于你如何在这种上下文中定义“等于”、“小于”和“大于”，这可能仍然有用。不过，
使用基于GiST的索引，你就可以构造出能够提出特定领域问题的查询方式，
例如“找出所有马的图片”或者“找出所有曝光过度的图片”。

要让一个GiST访问方法运行起来，只需实现几个用户定义的方法，
这些方法定义了树中键的行为。当然，要支持复杂查询，这些方法本身也必须足够巧妙；
但对于所有标准查询（B-树、R 树等），它们都相对直接。简而言之，GiST把可扩展性与通用性、代码复用以及清晰的接口结合了起来。

53.3. 实现

一个GiST索引操作符类必须提供七个方法，另外还有一个可选方法。通过正确实现`same`、`consistent`和`union`方法可以保证索引的正确性，而索引的效率（大小与速度）则取决于`penalty`和`picksplit`方法。另外两个基本方法是`compress`和`decompress`，
它们允许索引的内部树数据使用与其所索引数据不同的类型。叶子必须是被索引数据类型，
而其他树节点可以是任意 C 结构体（但这里仍必须遵守PostgreSQL的数据类型规则，关于变长数据可参见`varlena`）。如果树的内部数据类型在 SQL 层存在，可以使用`CREATE OPERATOR CLASS`命令的`STORAGE`选项。可选的第八个方法是`distance`，若操作符类希望支持有序扫描（最近邻搜索），则需要它。

[1] http://gist.cs.berkeley.edu/
[2] http://www.sai.msu.su/~megera/postgres/gist/papers/concurrency/access-methods-for-next-generation.pdf.gz
[3] http://www.sai.msu.su/~megera/postgres/gist/

consistent

给定一个索引项p和一个查询值q，该函数判断该索引项是否与该查询“一致”；也就是说，该索引项所代表的某一行是否可能使谓词“*indexed_column indexable_operator* q”为真。对于叶子索引项，这等同于测试该可索引条件；而对于内部树节点，这决定是否有必要扫描该树节点所表示的索引子树。当结果为true时，还必须返回一个recheck标志。它表示该谓词是确定为真，还是仅可能为真。如果recheck = false，则该索引已经精确测试了谓词条件；如果recheck = true，则该行只是候选匹配。在这种情况下，系统会自动针对实际行值计算*indexable_operator*，以判断它是否真的匹配。这种约定使GiST能够同时支持无损和有损的索引结构。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_consistent(internal, data_type,
 smallint, oid, internal)
RETURNS bool
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_consistent(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_consistent);

Datum
my_consistent(PG_FUNCTION_ARGS)
{
    GISTENTRY  *entry = (GISTENTRY *) PG_GETARG_POINTER(0);
    data_type  *query = PG_GETARG_DATA_TYPE_P(1);
    StrategyNumber strategy = (StrategyNumber) PG_GETARG_UINT16(2);
    /* Oid subtype = PG_GETARG_OID(3); */
    bool       *recheck = (bool *) PG_GETARG_POINTER(4);
    data_type  *key = DatumGetDataType(entry->key);
    bool        retval;

    /*
     * 根据 strategy、key 和 query 确定返回值。
     *
     * 使用 GIST_LEAF(entry) 判断当前调用位于索引树的哪个位置。
     * 例如，支持 = 操作符时这很有用（可以在非叶节点检查
     * union() 是否非空，在叶节点检查是否相等）。
     */

    *recheck = true;        /* 如果检查是精确的，则为 false */

    PG_RETURN_BOOL(retval);
}
```

这里，key是索引中的一个元素，而query是在该索引中查找的值。StrategyNumber参数指示应用的是操作符类中的哪个操作符，它对应于CREATE OPERATOR CLASS命令中的某个操作符编号。取决于你在该类中包含了哪些操作符，query的数据类型可能会随操作符而变化，但上面的框架假设它不会变化。

union

该方法用于汇总树中的信息。给定一组项，该函数生成一个新的索引项，用来表示所有给定项。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_union(internal, internal)
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_union(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_union);

Datum
my_union(PG_FUNCTION_ARGS)
{
    GistEntryVector *entryvec = (GistEntryVector *) PG_GETARG_
POINTER(0);
    GISTENTRY  *ent = entryvec->vector;
    data_type  *out,
               *tmp,
               *old;
    int         numranges,
                i = 0;

    numranges = entryvec->n;
    tmp = DatumGetDataType(ent[0].key);
    out = tmp;

    if (numranges == 1)
    {
        out = data_type_deep_copy(tmp);

        PG_RETURN_DATA_TYPE_P(out);
    }

    for (i = 1; i < numranges; i++)
    {
        old = out;
        tmp = DatumGetDataType(ent[i].key);
        out = my_union_implementation(out, tmp);
    }

    PG_RETURN_DATA_TYPE_P(out);
}
```

如你所见，在这个框架里，我们处理的是一种满足`union(X, Y, Z) = union(union(X, Y), Z)`的数据类型。对于不满足这一性质的数据类型，只需在这个GiST支持方法中实现正确的 union 算法即可。

`union`实现函数应返回一个指向新近通过`palloc()`分配的内存的指针。不能原样返回输入值。

compress

将一个数据项转换成适合在索引页中物理存储的格式。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_compress(internal)
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_compress(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_compress);

Datum
my_compress(PG_FUNCTION_ARGS)
{
    GISTENTRY  *entry = (GISTENTRY *) PG_GETARG_POINTER(0);
    GISTENTRY  *retval;

    if (entry->leafkey)
    {
        /* 将 entry->key 替换为压缩后的形式 */
        compressed_data_type *compressed_data = palloc(sizeof(
compressed_data_type));

        /* 根据 entry->key 填充 *compressed_data ... */

        retval = palloc(sizeof(GISTENTRY));
        gistentryinit(*retval, PointerGetDatum(compressed_data),
                      entry->rel, entry->page, entry->offset,
 FALSE);
    }
    else
    {
        /* 通常无需对非叶项做任何处理 */
        retval = entry;
    }

    PG_RETURN_POINTER(retval);
}
```

当然，为了压缩叶子节点，你必须把*compressed_data_type*改成要转换成的具体类型。

根据你的需要，你可能还需要注意在其中压缩NULL值，例如像gist_circle_compress那样存储(Datum) 0。

decompress

compress方法的逆操作。将数据项的索引表示转换成数据库能够操作的格式。

该SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_decompress(internal)
```

```
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

C 模块中相应的代码可以采用以下框架：

```
Datum       my_decompress(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_decompress);

Datum
my_decompress(PG_FUNCTION_ARGS)
{
    PG_RETURN_POINTER(PG_GETARG_POINTER(0));
}
```

上述框架适用于不需要解压的情况。

`penalty`

返回一个值，指示把新项插入树中特定分支的“代价”。项会沿着树中`penalty`最小的路径插入。`penalty`返回的值应为非负；如果返回负值，它将被按零处理。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_penalty(internal, internal, internal)
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;  -- 某些情况下 penalty 函数不必是严格函数
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_penalty(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_penalty);

Datum
my_penalty(PG_FUNCTION_ARGS)
{
    GISTENTRY  *origentry = (GISTENTRY *) PG_GETARG_POINTER(0);
    GISTENTRY  *newentry = (GISTENTRY *) PG_GETARG_POINTER(1);
    float      *penalty = (float *) PG_GETARG_POINTER(2);
    data_type  *orig = DatumGetDataType(origentry->key);
    data_type  *new = DatumGetDataType(newentry->key);

    *penalty = my_penalty_implementation(orig, new);
    PG_RETURN_POINTER(penalty);
}
```

`penalty`函数对于索引的良好性能至关重要。
它会在插入时用于决定在树中应沿着哪个分支向下，以便选择把新项加到哪里。在查询时，索引越平衡，查找就越快。

`picksplit`

当索引页必须分裂时，该函数决定页面上的哪些项留在旧页中，哪些移到新页中。

该 SQL 声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_picksplit(internal, internal)
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

C 模块中相应的代码可以采用以下框架：

```
Datum       my_picksplit(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_picksplit);

Datum
my_picksplit(PG_FUNCTION_ARGS)
{
    GistEntryVector *entryvec = (GistEntryVector *) PG_GETARG_
POINTER(0);
    OffsetNumber maxoff = entryvec->n - 1;
    GISTENTRY  *ent = entryvec->vector;
    GIST_SPLITVEC *v = (GIST_SPLITVEC *) PG_GETARG_POINTER(1);
    int         i,
                nbytes;
    OffsetNumber *left,
                *right;
    data_type  *tmp_union;
    data_type  *unionL;
    data_type  *unionR;
    GISTENTRY **raw_entryvec;

    maxoff = entryvec->n - 1;
    nbytes = (maxoff + 1) * sizeof(OffsetNumber);

    v->spl_left = (OffsetNumber *) palloc(nbytes);
    left = v->spl_left;
    v->spl_nleft = 0;

    v->spl_right = (OffsetNumber *) palloc(nbytes);
    right = v->spl_right;
    v->spl_nright = 0;

    unionL = NULL;
    unionR = NULL;

    /* Initialize the raw entry vector. */
    raw_entryvec = (GISTENTRY **) malloc(entryvec->n * sizeof(void
 *));
    for (i = FirstOffsetNumber; i <= maxoff; i = OffsetNumberNext(
i))
        raw_entryvec[i] = &(entryvec->vector[i]);

    for (i = FirstOffsetNumber; i <= maxoff; i = OffsetNumberNext(
i))
    {
        int         real_index = raw_entryvec[i] - entryvec-
>vector;
```

```
        tmp_union = DatumGetDataType(entryvec->vector[real_index].
key);
        Assert(tmp_union != NULL);

        /*
         * Choose where to put the index entries and update unionL
 and unionR
         * accordingly. Append the entries to either v_spl_left or
         * v_spl_right, and care about the counters.
         */

        if (my_choice_is_left(unionL, curl, unionR, curr))
        {
            if (unionL == NULL)
                unionL = tmp_union;
            else
                unionL = my_union_implementation(unionL, tmp_union)
;

            *left = real_index;
            ++left;
            ++(v->spl_nleft);
        }
        else
        {
            /*
             * Same on the right
             */
        }
    }

    v->spl_ldatum = DataTypeGetDatum(unionL);
    v->spl_rdatum = DataTypeGetDatum(unionR);
    PG_RETURN_POINTER(v);
}
```

和penalty一样，picksplit函数对于索引的良好性能至关重要。设计合适的penalty和picksplit实现，正是实现高性能GiST索引的难点所在。

same

如果两个索引项相同则返回真，否则返回假。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_same(internal, internal, internal)
RETURNS internal
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_same(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_same);
```

```
Datum
my_same(PG_FUNCTION_ARGS)
{
    prefix_range *v1 = PG_GETARG_PREFIX_RANGE_P(0);
    prefix_range *v2 = PG_GETARG_PREFIX_RANGE_P(1);
    bool       *result = (bool *) PG_GETARG_POINTER(2);

    *result = my_eq(v1, v2);
    PG_RETURN_POINTER(result);
}
```

出于历史原因，`same`函数并不是直接返回一个布尔结果；相反，
它必须把该标志存储到第三个参数指示的位置。

`distance`

给定一个索引项`p`和一个查询值`q`，该函数确定索引项与查询值之间的“距离”。
如果操作符类包含任何排序操作符，就必须提供此函数。
使用排序操作符的查询会优先返回“距离”值最小的索引项，
因此结果必须与该操作符的语义一致。对于叶子索引项，结果仅表示到该索引项的距离；
对于内部树节点，结果必须是其任意子项可能具有的最小距离。

该函数的SQL声明必须如下所示：

```
CREATE OR REPLACE FUNCTION my_distance(internal, data_type,
 smallint, oid)
RETURNS float8
AS 'MODULE_PATHNAME'
LANGUAGE C STRICT;
```

而 C 模块中的对应代码则可以遵循如下框架：

```
Datum       my_distance(PG_FUNCTION_ARGS);
PG_FUNCTION_INFO_V1(my_distance);

Datum
my_distance(PG_FUNCTION_ARGS)
{
    GISTENTRY  *entry = (GISTENTRY *) PG_GETARG_POINTER(0);
    data_type  *query = PG_GETARG_DATA_TYPE_P(1);
    StrategyNumber strategy = (StrategyNumber) PG_GETARG_UINT16(2);
    /* Oid subtype = PG_GETARG_OID(3); */
    data_type  *key = DatumGetDataType(entry->key);
    double      retval;

    /*
     * 根据 strategy、key 和 query 确定返回值。
     */

    PG_RETURN_FLOAT8(retval);
}
```

`distance`函数的参数与`consistent`函数的参数相同，只是不使用 recheck 标志。到叶子索引项的距离必须始终精确确定，因为元组一旦返回就无法再重新排序。
在确定到内部树节点的距离时允许有一定近似，

只要结果永不大于任一子节点的实际距离即可。因此，例如在几何应用中，到包围盒的距离通常就足够了。结果值可以是任意有限的`float8`值。（无穷大和负无穷在内部用于处理空值等情况，因此不建议`distance`函数返回这些值。）

53.4. 示例

PostgreSQL源代码发行包包含了若干使用GiST实现的索引方法示例。核心系统目前提供了文本检索支持（为`tsvector`和`tsquery`建立索引），并为某些内置几何数据类型提供了与 R 树等价的功能（见`src/backend/access/gist/gistproc.c`）。下列`contrib`模块中也包含GiST操作符类：

`btree_gist`

为多种数据类型提供与 B-树等价的功能

`cube`

多维立方体的索引

`hstore`

用于存储（键，值）对的模块

`intarray`

一维 int4 值数组的 RD 树

`ltree`

树状结构的索引

`pg_trgm`

基于三字符组匹配的文本相似度

`seg`

“float 范围” 的索引

第 54 章 GIN 索引

54.1. 简介

GIN 是 Generalized Inverted Index（通用倒排索引）的缩写。GIN 被设计用于处理这样一类情况：要建立索引的项是组合值，
而索引需要处理的查询则要搜索出现在这些组合项内部的元素值。例如，这些项可以是文档，
而查询可以是搜索包含特定词语的文档。

我们用项来指代要被索引的组合值，用键来指代其中的元素值。GIN 总是存储并搜索键，而不是项值本身。

GIN 索引存储一组（键，倒排列表）对，其中一个倒排列表（posting list）是一个包含出现该键的行 ID 的集合。由于一个项可以包含多个键，同一个行 ID 可以出现在多个倒排列表中。每个键值只会被存储一次，因此在同一个键大量重复出现的场景下，GIN 索引会非常紧凑。

GIN 之所以是通用的，是因为 GIN 访问方法的代码不需要知道它所加速的具体操作。相反，它使用为特定数据类型定义的自定义策略。该策略定义了如何从被索引项和查询条件中提取键，
以及如何判断一个包含查询中某些键值的行是否真正满足该查询。

GIN 的一个优点在于，它允许由数据类型所属领域的专家来开发带有合适访问方法的自定义数据类型，而不必依赖数据库专家。这一点与使用 GiST 的优势非常相似。

PostgreSQL 中的 GIN 实现主要由 Teodor Sigaev 和 Oleg Bartunov 维护。关于 GIN 的更多信息，可在他们的网站[1]上找到。

54.2. 可扩展性

GIN 接口具有很高的抽象层次，访问方法实现者只需实现所访问数据类型的语义。GIN 层本身会处理并发、日志记录以及树结构的搜索。

要让一个 GIN 访问方法工作起来，只需实现四（或五）个用户定义的方法，它们定义了树中键的行为，以及键、被索引项和可索引查询之间的关系。简言之，GIN 将可扩展性与通用性、代码重用以及清晰的接口结合在一起。

GIN 操作符类必须提供以下四个方法：

`int compare(Datum a, Datum b)`

> 比较两个键（不是被索引项！），并返回一个整数：小于零、等于零或大于零，
> 分别表示第一个键小于、等于或大于第二个键。null 键绝不会被传递给这个函数。

`Datum *extractValue(Datum itemValue, int32 *nkeys, bool **nullFlags)`

> 给定一个要建立索引的项，返回一个用 palloc 分配的键数组。返回键的数量必须存入 `*nkeys`。如果任一键可以为 null，还要另外 palloc 一个包含 `*nkeys` 个布尔值的数组，将其地址存入 `*nullFlags`，并按需设置这些空值标志。如果所有键都不是 null，`*nullFlags` 可以保持为 NULL（其初始值）。如果该项不包含任何键，则返回值可以为 NULL。

[1] http://www.sai.msu.su/~megera/wiki/Gin

`Datum *extractQuery(Datum query, int32 *nkeys, StrategyNumber n, bool **pmatch, Pointer **extra_data, bool **nullFlags, int32 *searchMode)`

给定一个待查询的值，返回一个用 palloc 分配的键数组；也就是说，`query` 是一个可索引操作符右侧的值，而该操作符左侧是被索引列。`n` 是该操作符在操作符类中的策略号（见第 35.14.2 节）。通常，`extractQuery` 需要查看 `n`，以确定 `query` 的数据类型，以及应采用何种方法提取键值。返回键的数量必须存入 `*nkeys`。如果任一键可以为 null，还要另外 palloc 一个包含 `*nkeys` 个布尔值的数组，将其地址存入 `*nullFlags`，并按需设置这些空值标志。如果所有键都不是 null，`*nullFlags` 可以保持为 NULL（其初始值）。如果 `query` 不包含任何键，则返回值可以为 NULL。

`searchMode` 是一个输出参数，用于让 `extractQuery` 指定搜索如何执行。若 `*searchMode` 被设置为 `GIN_SEARCH_MODE_DEFAULT`（调用前它会被初始化为该值），则只有至少匹配一个返回键的项才会被视为候选匹配。若 `*searchMode` 被设置为 `GIN_SEARCH_MODE_INCLUDE_EMPTY`，则除至少包含一个匹配键的项之外，完全不含任何键的项也会被视为候选匹配。（例如，该模式对于实现是子集操作符很有用。）若 `*searchMode` 被设置为 `GIN_SEARCH_MODE_ALL`，则索引中所有非 null 项都会被视为候选匹配，无论它们是否匹配任一返回键。（该模式比前两种选择慢得多，
因为它基本上需要扫描整个索引；但为了正确处理某些边界情况，可能有此必要。
在大多数情况下都需要此模式的操作符，大概并不适合作为 GIN 操作符类的候选。）用于设置该模式的符号定义在 `access/gin.h` 中。

`pmatch` 是一个在支持部分匹配时使用的输出参数。要使用它，`extractQuery` 必须分配一个包含 `*nkeys` 个布尔值的数组，并将其地址存入 `*pmatch`。若相应键需要部分匹配，则数组对应元素应设置为 TRUE，否则设置为 FALSE。如果 `*pmatch` 被设置为 NULL，那么 GIN 认为不需要部分匹配。该变量在调用前会初始化为 NULL，因此不支持部分匹配的操作符类可以直接忽略此参数。

`extra_data` 是一个输出参数，用于让 `extractQuery` 向 `consistent` 和 `comparePartial` 方法传递额外数据。要使用它，`extractQuery` 必须分配一个包含 `*nkeys` 个指针的数组，并将其地址存入 `*extra_data`，然后把所需内容存入各个指针。该变量在调用前会初始化为 NULL，因此不需要额外数据的操作符类可以直接忽略此参数。如果设置了 `*extra_data`，则整个数组会传给 `consistent` 方法，而对应元素会传给 `comparePartial` 方法。

`bool consistent(bool check[], StrategyNumber n, Datum query, int32 nkeys, Pointer extra_data[], bool *recheck, Datum queryKeys[], bool nullFlags[])`

如果被索引项满足策略号为 `n` 的查询操作符，则返回 TRUE；如果同时返回需要复核的指示，那么 TRUE 也可以只表示它可能满足。该函数无法直接访问被索引项的值，因为 GIN 并不显式存储项。它所能利用的是这样一种信息：从查询中提取出的哪些键值出现在给定的被索引项中。`check` 数组长度为 `nkeys`，这与先前针对该 `query` datum 由 `extractQuery` 返回的键数量相同。如果被索引项包含相应查询键，则 `check` 数组中的对应元素为 TRUE；也就是说，如果 (check[i] == TRUE)，则 `extractQuery` 结果数组中的第 i 个键存在于该被索引项中。传入原始 `query` datum，是为了让 `consistent` 方法在需要时可以查看它；同样也会传入先前由 `extractQuery` 返回的 `queryKeys[]` 和 `nullFlags[]` 数组。`extra_data` 则是 `extractQuery` 返回的额外数据数组，如果没有则为 NULL。

当 `extractQuery` 在 `queryKeys[]` 中返回一个 null 键时，若被索引项包含 null 键，则对应的 `check[]` 元素为 TRUE；也就是说，`check[]` 的语义类似于 `IS NOT DISTINCT FROM`。如果 `consistent` 函数需要区分普通值匹配与 null 匹配，它可以检查对应的 `nullFlags[]` 元素。

成功时，如果需要根据查询操作符重新检查堆元组，则 `*recheck` 应设置为 TRUE；如果索引测试是精确的，则设置为 FALSE。也就是说，返回 FALSE 保证该堆元组不匹配查询；返回 TRUE 且 `*recheck` 为 FALSE 保证该堆元组匹配查询；返回 TRUE 且 `*recheck` 为 TRUE 则表示该堆元组可能匹配查询，因此需要取出该元组，并直接对最初被索引的项值重新计算查询操作符以完成复核。

Optionally, an operator class for GIN can supply a fifth method:

```
int comparePartial(Datum partial_key, Datum key, StrategyNumber n,
 Pointer extra_data)
```

比较部分匹配查询键与索引键。返回一个整数，
其符号表示结果：小于零表示索引键不匹配查询，但索引扫描应继续；
零表示索引键匹配查询；大于零表示索引扫描应停止，因为后续不可能再有匹配项。
产生该部分匹配查询的操作符的策略号 `n` 会被传入，以便在需要其语义来决定何时结束扫描时使用。另外，`extra_data` 是由 `extractQuery` 生成的额外数据数组中的对应元素，如果没有则为 NULL。null 键绝不会被传递给这个函数。

要支持“部分匹配”查询，操作符类必须提供 `comparePartial` 方法，并且其 `extractQuery` 方法在遇到部分匹配查询时必须设置 `pmatch` 参数。详见第 54.3.2 节。

上文提到的各种 `Datum` 值，其实际数据类型会因操作符类而异。传给 `extractValue` 的项值始终是该操作符类的输入类型，而所有键值都必须是该类的 `STORAGE` 类型。传给 `extractQuery`、`consistent` 的 `query` 参数类型，是由该策略号标识的类成员操作符右侧输入类型。只要能够从中提取出正确类型的键值，它就不必与项类型相同。

54.3. 实现

在内部，一个 GIN 索引包含一个基于键构建的 B-树索引，其中每个键都是一个或多个被索引项中的某个元素（例如数组成员），而叶子页中的每个元组要么包含一个指向堆指针 B-树的指针（“倒排树”），要么在列表足够小、能够连同键值一起放入单个索引元组时，
直接包含一个简单的堆指针列表（“倒排列表”）。

从 PostgreSQL 9.1 开始，索引中可以包含 NULL 键值。此外，对于那些自身为 NULL，或按照 `extractValue` 的定义不包含任何键的被索引项，索引中也会包含占位符 NULL 值。这样一来，应该找到空项的搜索就能够找到它们。

多列 GIN 索引的实现方式，是在组合值（列号，键值）之上构建一个单独的 B-树。不同列的键值可以具有不同类型。

54.3.1. GIN 快速更新技术

由于倒排索引的内在性质，更新 GIN 索引往往比较慢：插入或更新一个堆行，可能会导致向索引中执行多次插入（从被索引项中提取出的每个键都要插入一次）。从PostgreSQL 8.4 开始，GIN 可以通过把新元组插入一个临时的、未排序的待处理列表，来推迟其中的大部分工作。当表被清理时，或者待处理列表变得过大（大于work_mem）时，这些项就会被移入主要的 GIN 数据结构中，所用的是与初始索引创建时相同的批量插入技术。即便把额外的清理开销计算在内，这也会大幅提高 GIN 索引的更新速度。此外，这部分开销工作还可以由后台进程完成，而不是在前台查询处理过程中完成。

这种方法的主要缺点是，搜索除了要查找常规索引之外，还必须扫描待处理列表，
因此大型待处理列表会显著拖慢搜索。另一个缺点是，虽然大多数更新都很快，
但一旦某次更新使待处理列表变得“过大”，就会立刻触发一次清理周期，
因此该次更新会比其他更新慢得多。恰当地使用自动清理可以将这两个问题都尽量减轻。

如果稳定的响应时间比更新速度更重要，可以通过关闭 GIN 索引的 `FASTUPDATE` 存储参数来禁用待处理列表机制。详见CREATE INDEX。

54.3.2. 部分匹配算法

GIN 可以支持“部分匹配”查询。在这种查询中，查询无法确定一个或多个键的精确匹配，但可能的匹配会落在一个相对狭窄的键值范围内（该范围位于由 `compare` 支持方法确定的键排序顺序中）。`extractQuery` 方法不是返回一个要精确匹配的键值，而是返回待搜索范围的下界键值，并将 `pmatch` 标志设为 true。随后使用 `comparePartial` 方法扫描该键范围。对于匹配的索引键，`comparePartial` 必须返回零；对于虽然不匹配但仍处在待搜索范围内的索引键，返回小于零；如果索引键已经超出了可能匹配的范围，则返回大于零。

54.4. GIN 提示和技巧

创建与插入

向 GIN 索引插入数据可能较慢，因为每个项很可能需要插入许多键。因此，对于向表中执行的批量插入，建议先删除 GIN 索引，待批量插入完成后再重建它。

从PostgreSQL 8.4 开始，由于采用了延迟索引，这项建议已不那么必要（详见第 54.3.1 节）。但对于非常大的更新，删除并重建索引仍然可能是最佳选择。

maintenance_work_mem

GIN 索引的构建时间对 `maintenance_work_mem` 设置非常敏感；在创建索引时节省工作内存并不划算。

work_mem

在对现有 GIN 索引执行一系列插入且其 `FASTUPDATE` 已启用时，系统会在待处理项列表增长到超过 `work_mem` 时清理该列表。为避免观测到的响应时间波动，最好让待处理列表的清理在后台发生（即通过自动清理）。可以通过增大 `work_mem` 或让自动清理更积极，来避免前台清理操作。不过，增大 `work_mem` 也意味着，一旦确实发生前台清理，耗时会更长。

gin_fuzzy_search_limit

开发 GIN 索引的主要目标，是为 PostgreSQL 中高度可伸缩的全文检索提供支持，而全文检索返回一个非常大的结果集的情况也很常见。此外，当查询包含非常常见的词时，
通常也会出现这种情况，以至于这个大型结果集本身并无多大用处。
由于从磁盘读取大量元组并对其排序可能需要很长时间，这在生产环境中是不可接受的。
（注意，索引搜索本身是非常快的。）

为了便于受控地执行这类查询，GIN 对返回行数设置了一个可配置的软上限，即配置参数 `gin_fuzzy_search_limit`。默认值为 0（表示无限制）。如果设置了非零限制，那么返回集合将是整个结果集的一个随机子集。

“软”的意思是，实际返回结果的数量可能会与指定的限制略有不同，
这取决于查询以及系统随机数生成器的质量。

根据经验，数千量级的值（例如 5000 – 20000）效果不错。

54.5. 限制

GIN 假定可索引操作符是严格的。这意味着，当项值为 NULL 时，根本不会对其调用 `extractValue`（而是自动创建一个占位符索引项）；当查询值为 NULL 时，也不会调用 `extractQuery`

（而是认为该查询不可满足）。不过要注意，非 null 的组合项或查询值内部包含的 NULL 键值仍然受支持。

54.6. 示例

PostgreSQL 源码发行版包含 `tsvector` 以及所有内部类型的一维数组的 GIN 操作符类。`tsvector` 中的前缀搜索是使用 GIN 部分匹配特性实现的。下列 `contrib` 模块也包含 GIN 操作符类：

`btree_gin`

为若干数据类型提供等效于 B-树的功能

`hstore`

用于存储（键，值）对的模块

`intarray`

对 `int[]` 的增强支持

`pg_trgm`

基于三字符组匹配的文本相似度

第 55 章 数据库物理存储

本章概述 PostgreSQL 数据库所使用的物理存储格式。

55.1. 数据库文件布局

本节从文件和目录的层次描述存储格式。

数据库集簇所需的全部数据都存放在集簇的数据目录中，该目录通常称为 PGDATA（名字来自可用于定义它的环境变量）。PGDATA 的一个常见位置是 /var/lib/pgsql/data。同一台机器上可以存在多个由不同服务器实例管理的集簇。

PGDATA 目录包含若干子目录和控制文件，如表 55.1所示。除这些必需项之外，集簇配置文件 postgresql.conf、pg_hba.conf 和 pg_ident.conf 传统上也存放在 PGDATA 中（不过在 PostgreSQL 8.0 及以后版本中，可以把它们放在其他地方）。

表 55.1. PGDATA 的内容

Item	Description
PG_VERSION	A file containing the major version number of PostgreSQL
base	Subdirectory containing per-database subdirectories
global	Subdirectory containing cluster-wide tables, such as pg_database
pg_clog	Subdirectory containing transaction commit status data
pg_multixact	Subdirectory containing multitransaction status data (used for shared row locks)
pg_notify	Subdirectory containing LISTEN/NOTIFY status data
pg_serial	Subdirectory containing information about committed serializable transactions
pg_stat_tmp	Subdirectory containing temporary files for the statistics subsystem
pg_subtrans	Subdirectory containing subtransaction status data
pg_tblspc	Subdirectory containing symbolic links to tablespaces
pg_twophase	Subdirectory containing state files for prepared transactions
pg_xlog	Subdirectory containing WAL (Write Ahead Log) files
postmaster.opts	A file recording the command-line options the server was last started with
postmaster.pid	A lock file recording the current postmaster process ID (PID), cluster data directory path, postmaster start timestamp, port number, Unix-domain socket directory path (empty on Windows), first valid listen_address (IP address or *, or empty if not listening on TCP), and

Item	Description
	shared memory segment ID (this file is not present after server shutdown)

对于集簇中的每个数据库，PGDATA/base 中都有一个子目录，其名称是该数据库在 pg_database 中的 OID。这个子目录是该数据库文件的默认位置；特别是，它的系统目录就存放在这里。

每个表和索引都存储在单独的文件中。对于普通关系，这些文件以表或索引的 filenode 编号命名，该编号可在 pg_class.relfilenode 中找到。但对于临时关系，文件名的形式为 t*BBB*_*FFF*，其中 *BBB* 是创建该文件的后端进程号，*FFF* 是 filenode 编号。无论哪种情况，除了主文件（也称主分支）之外，每个表和索引还有一个空闲空间映射（见第 55.3 节），用于存储该关系可用空闲空间的信息。空闲空间映射存放在以 filenode 编号加上 _fsm 后缀命名的文件中。表还拥有可见性映射，存放在带有 _vm 后缀的分支中，用于跟踪哪些页面已知不含死元组。可见性映射在第 55.4 节中有进一步说明。不记录 WAL 的表和索引还有第三个分支，称为初始化分支，存放在带有 _init 后缀的分支中（见第 55.5 节）。

> **小心**
>
> 请注意，虽然表的 filenode 往往与其 OID 相同，但这并非必然如此；有些操作，例如 TRUNCATE、REINDEX、CLUSTER 以及某些形式的 ALTER TABLE，会在保留 OID 的同时改变 filenode。不要假定 filenode 和表 OID 一定相同。此外，对于某些系统目录（包括 pg_class 本身），pg_class.relfilenode 中存放的是零。这些目录的实际 filenode 编号保存在更底层的数据结构中，可以使用 pg_relation_filenode() 函数取得。

当表或索引超过 1 GB 时，会被划分成若干个 1 GB 大小的段。第一段的文件名与 filenode 相同，后续各段则命名为 filenode.1、filenode.2，依此类推。这样的安排避免了在文件大小受限的平台上出现问题。（实际上，1 GB 只是默认段大小。段大小可以通过配置选项 --with-segsize 在构建 PostgreSQL 时调整。）原则上，空闲空间映射和可见性映射分支也可能需要多个段，不过在实践中这不太可能发生。

如果一个表包含可能具有很大条目的列，它就会有关联的 TOAST 表，用于对那些大到无法直接保存在表行中的字段值进行线外存储。pg_class.reltoastrelid 会把该表关联到其 TOAST 表（如果存在的话）。更多信息见第 55.2 节。

表和索引的内容将在第 55.6 节中进一步讨论。

表空间使情况变得更复杂。每个用户定义的表空间都在 PGDATA/pg_tblspc 目录中有一个符号链接，指向其物理表空间目录（即该表空间的 CREATE TABLESPACE 命令中指定的位置）。该符号链接以表空间的 OID 命名。在物理表空间目录中，有一个名称依赖于 PostgreSQL 服务器版本的子目录，例如 PG_9.0_201008051。（使用这个子目录的原因是，为了让数据库的后续版本可以在不发生冲突的情况下复用同一个 CREATE TABLESPACE 位置值。）在这个版本相关的子目录中，对于每个在该表空间中拥有对象的数据库，都有一个以该数据库 OID 命名的子目录。表和索引就使用 filenode 命名方案存储在该目录中。pg_default 表空间不通过 pg_tblspc 访问，而是对应于 PGDATA/base。同样，pg_global 表空间也不通过 pg_tblspc 访问，而是对应于 PGDATA/global。

pg_relation_filepath() 函数会显示任意关系的完整路径（相对于 PGDATA）。它常可用作记住上述众多规则的替代方法。但请记住，该函数只给出该关系主分支第一段的名字，若要找出与该关系关联的所有文件，你可能还需要追加段号以及/或者 _fsm、_vm 或 _init。

临时文件（用于诸如排序的数据多于内存可容纳量之类的操作）会创建在 PGDATA/base/pgsql_tmp 中；或者创建在表空间目录下的 pgsql_tmp 子目录中，如果为其指定的是不同于

pg_default 的表空间。临时文件的名称形式为 pgsql_tmp*PPP*.*NNN*，其中 *PPP* 是所属后端的 PID，*NNN* 用来区分该后端的不同临时文件。

55.2. TOAST

本节概述TOAST（超长属性存储技术）。

PostgreSQL 使用固定页面大小（通常为 8 kB），并且不允许元组跨越多个页面。因此，无法直接存储非常大的字段值。为克服这一限制，大字段值会被压缩并且/或者拆分成多个物理行。这个过程对用户是透明的，而且对大部分后端代码的影响很小。这项技术被亲切地称为 TOAST（或者“自切片面包问世以来最棒的东西”）。

只有某些数据类型支持 TOAST，因为没有必要让那些不可能产生大字段值的数据类型承担这部分额外开销。要支持 TOAST，数据类型必须具有变长（varlena）表示形式，其中任何已存储值的第一个 32 位字都包含该值以字节计的总长度（包括其自身）。TOAST 不约束表示形式的其余部分。所有支持可 TOAST 数据类型的 C 级函数都必须小心处理已 TOAST 化的输入值。（通常是在对输入值做任何处理之前调用 PG_DETOAST_DATUM，但在某些情况下也可以采用更高效的方法。）

TOAST 会占用 varlena 长度字中的两个位（在大端机器上是最高位，在小端机器上是最低位），从而把任何可 TOAST 数据类型值的逻辑大小限制为 1 GB（2^{30} - 1 字节）。当这两个位都为零时，该值就是这种数据类型的普通、未经 TOAST 处理的值，长度字的其余位给出该 datum 的总大小（包括长度字），以字节计。当最高位或最低位被置位时，该值使用的不是通常的四字节头部，而是单字节头部；该字节的其余位给出 datum 的总大小（包括长度字节），以字节计。作为特殊情况，如果其余位全为零（对于自包含长度而言这是不可能的），则该值是指向存储在单独 TOAST 表中的线外数据的指针。（TOAST 指针的大小由 datum 的第二个字节给出。）带单字节头部的值同样不按任何特定边界对齐。最后，当最高位或最低位为零、但相邻的一位被置位时，该 datum 的内容已被压缩，使用前必须先解压。在这种情况下，长度字的其余位给出的是压缩后 datum 的总大小，而不是原始数据的大小。请注意，线外数据也可能被压缩，但 varlena 头部不会告诉你是否发生了压缩，这一点要由 TOAST 指针的内容来说明。

如果一个表的任意列是可 TOAST 的，该表就会有一个相关联的 TOAST 表，其 OID 存储在该表的 pg_class.reltoastrelid 项中。线外 TOAST 化值保存在 TOAST 表中，下文会更详细地描述。

所用的压缩技术是 LZ 压缩技术族中一种相当简单且非常快速的方法。详情见 src/backend/utils/adt/pg_lzcompress.c。

线外值会被划分成最多 TOAST_MAX_CHUNK_SIZE 字节的块（如果启用了压缩，则在压缩之后再分块；默认情况下该值的选取方式是让四个块行可以装入一页，因此大约是 2000 字节）。每个块都作为独立的一行，存储在所属表的 TOAST 表中。每个 TOAST 表都有 chunk_id 列（标识某个特定 TOAST 化值的 OID）、chunk_seq 列（该块在其值内的序号）以及 chunk_data 列（该块的实际数据）。在 chunk_id 和 chunk_seq 上建立的唯一索引可提供快速检索。因此，一个表示线外 TOAST 化值的指针 datum，需要存储要查找的 TOAST 表的 OID，以及该特定值的 OID（即其 chunk_id）。为方便起见，指针 datum 还会存储逻辑 datum 大小（原始未压缩数据长度）和物理存储大小（如果应用了压缩，这两者会不同）。再加上 varlena 头部字节，TOAST 指针 datum 的总大小因此恒为 18 字节，与所表示值的实际大小无关。

TOAST 代码只有在要存入表中的行值宽于 TOAST_TUPLE_THRESHOLD 字节（通常为 2 kB）时才会被触发。TOAST 代码会压缩并且/或者把字段值移到线外，直到该行值短于 TOAST_TUPLE_TARGET 字节（通常也为 2 kB），或者已经无法再获得更多收益。在 UPDATE 操作中，未更改字段的值通常会原样保留；因此，如果更新一行时其线外值都没有变化，便不会产生 TOAST 成本。

TOAST 代码为存储可 TOAST 的列识别出四种不同的策略：

- `PLAIN` 既不允许压缩，也不允许线外存储；此外，它还会禁止 varlena 类型使用单字节头部。这是不可 TOAST 数据类型列唯一可能的策略。

- `EXTENDED` 同时允许压缩和线外存储。这是大多数可 TOAST 数据类型的默认策略。系统会先尝试压缩，如果行仍然过大，再使用线外存储。

- `EXTERNAL` 允许线外存储，但不允许压缩。使用 `EXTERNAL` 会让宽 `text` 和 `bytea` 列上的子串操作更快（代价是占用更多存储空间），因为在值未压缩时，这些操作经过优化，只需提取线外值中所需的部分。

- `MAIN` 允许压缩，但不允许线外存储。（实际上，对于这类列，仍然可能进行线外存储，但只有在别无他法、必须这样做才能让行足够小以放入页面时，才会把它作为最后手段。）

每种可 TOAST 的数据类型都会为该数据类型的列指定默认策略，但可以通过以下命令更改某个表列的策略：`ALTER TABLE SET STORAGE`。

与允许行值跨页之类更直接的方法相比，这种方案有不少优点。
假定查询通常是通过与相对较小的键值比较来限定的，
执行器的大部分工作都会只使用主行项完成。TOAST 化属性的大值只有在结果集发送给客户端时才会被取出（前提是它们确实被选中了）。因此，主表会小得多，
其更多的行能够装入共享缓冲区缓存，这是没有线外存储时做不到的。排序集也会缩小，
因此排序更常能够完全在内存中完成。一个小测试表明，一个包含典型 HTML 页面及其 URL 的表，其总存储量（包括 TOAST 表）大约只有原始数据大小的一半，而主表只包含全部数据的大约 10%（URL 以及一些较小的 HTML 页面）。与未进行 TOAST 处理的对照表相比，运行时没有差异；在那个对照表中，所有 HTML 页面都被裁剪到 7 kB 以内以便放得下。

55.3. 空闲空间映射

每个堆关系和索引关系（哈希索引除外）都有一个空闲空间映射（FSM），
用于跟踪该关系中的可用空间。它与主关系数据并列存放在一个独立的关系分支中，
其名称由该关系的 filenode 编号加上 `_fsm` 后缀构成。例如，如果某关系的 filenode 是 12345，那么 FSM 就存储在名为 `12345_fsm` 的文件中，与主关系文件位于同一目录。

空闲空间映射被组织成一棵由 FSM 页面构成的树。最底层的 FSM 页面存储每个堆页面（或索引页面）上的可用空闲空间，每个这样的页面用一个字节表示。上层页面则聚合来自下层的信息。

每个 FSM 页面内部都有一棵二叉树，以数组形式存储，每个节点占一个字节。每个叶子节点表示一个堆页面或下层 FSM 页面。每个非叶子节点中保存其子节点值中较大的那个，因此叶子节点中的最大值会保存在根节点。

请参见 `src/backend/storage/freespace/README`，了解 FSM 的结构以及如何更新和搜索它的更多细节。pg_freespacemap模块可用于检查空闲空间映射中存储的信息。

55.4. 可见性映射

每个堆关系都有一个可见性映射（VM），
用于跟踪哪些页面只包含已知对所有活动事务都可见的元组。
它与主关系数据并列存放在一个独立的关系分支中，名称由关系的 filenode 编号加上 `_vm` 后缀构成。例如，如果某关系的 filenode 是 12345，那么 VM 会存储在名为 `12345_vm` 的文件中，与主关系文件位于同一目录。请注意，索引没有 VM。

可见性映射只是为每个堆页面存储一个位。
该位被置位表示页面上的所有元组都已知对所有事务可见。

这意味着该页面不包含任何需要被清理的元组；
将来它还可以用来避免为可见性检查而访问该页面。这个映射是保守的，也就是说，
只要某个位被置位，我们就能确定相应条件确实成立；但如果某个位没有置位，
该条件既可能为真，也可能不为真。

55.5. 初始化分支

每个不记录 WAL 的表以及该表上的每个索引，都有一个初始化分支。初始化分支是一个适当类型的空表或空索引。当不记录 WAL 的表因崩溃而必须重置为空时，初始化分支会被复制覆盖到主分支上，而其他分支都会被擦除（它们会在需要时自动重建）。

55.6. 数据库页面布局

本节概述 PostgreSQL 表和索引内部所使用的页面格式。[1]序列和 TOAST 表的格式与普通表相同。

在下面的说明中，假定一个字节包含 8 个位。另外，术语项指的是存储在页面上的单个数据值。在表中，项是一行；在索引中，项是一条索引条目。

每个表和索引都存储为固定大小的页面数组（通常为 8 kB，不过在编译服务器时可以选择不同的页面大小）。在表中，所有页面在逻辑上都是等价的，因此某个特定项（行）
可以存放在任意页面中。在索引中，第一页通常保留为元页，用于保存控制信息；
并且根据索引访问方法的不同，索引中还可能存在不同类型的页面。

表 55.2展示了页面的总体布局。每个页面包含五个部分。

表 55.2. 总体页面布局

Item	Description
PageHeaderData	24 bytes long. Contains general information about the page, including free space pointers.
ItemIdData	Array of (offset,length) pairs pointing to the actual items. 4 bytes per item.
Free space	The unallocated space. New item pointers are allocated from the start of this area, new items from the end.
Items	The actual items themselves.
Special space	Index access method specific data. Different methods store different data. Empty in ordinary tables.

每个页面的前 24 个字节由页头（PageHeaderData）组成。它的格式详见表 55.3。前两个字段跟踪与该页面有关的最新 WAL 记录。接下来是一个 2 字节字段，包含标志位。随后是三个 2 字节整数字段（`pd_lower`、`pd_upper` 和 `pd_special`）。它们分别保存从页面起始位置到未分配空间起始位置、未分配空间结束位置以及特殊空间起始位置的字节偏移量。页头接下来的 2 个字节 `pd_pagesize_version` 同时存储页面大小和版本指示器。从 PostgreSQL 8.3 起，版本号为 4；PostgreSQL 8.1 和 8.2 使用版本号 3；PostgreSQL 8.0 使用版本号 2；PostgreSQL 7.3 和 7.4 使用版本号 1；更早的版本使用版本号 0。（这些版本中的大多数，其基本页面布局和页头格式并未改变，但堆行头部的布局发生过

[1] 实际上，索引访问方法不要求必须使用这种页面格式。现有的所有索引方法都使用这种基本格式，但索引元页中保存的数据通常并不遵循项布局规则。

变化。）页面大小字段基本上只是用于交叉检查；一个安装中并不支持同时存在多种页面大小。
最后一个字段是一个提示，用于显示对页面进行剪枝是否可能有利：它跟踪页面上最老的、
尚未剪枝的 XMAX。

表 55.3. PageHeaderData 布局

Field	Type	Length	Description
pd_lsn	XLogRecPtr	8 bytes	LSN: next byte after last byte of xlog record for last change to this page
pd_tli	uint16	2 bytes	TimeLineID of last change (only its lowest 16 bits)
pd_flags	uint16	2 bytes	Flag bits
pd_lower	LocationIndex	2 bytes	Offset to start of free space
pd_upper	LocationIndex	2 bytes	Offset to end of free space
pd_special	LocationIndex	2 bytes	Offset to start of specia l space
pd_pagesize_version	uint16	2 bytes	Page size and layout version number inform ation
pd_prune_xid	TransactionId	4 bytes	Oldest unpruned XMAX on page, or zero if none

所有细节都可以在 `src/include/storage/bufpage.h` 中找到。

页头之后是项标识符（`ItemIdData`），每个需要四个字节。
一个项标识符包含项起始位置的字节偏移量、其字节长度，以及若干影响解释方式的属性位。
新的项标识符会根据需要从未分配空间的起始处分配。当前已有多少个项标识符，可以通过查看 `pd_lower` 得知；分配新标识符时它会增加。因为一个项标识符在被释放之前永远不会移动，所以即使项本身为了压缩空闲空间而在页面内移动，其索引仍然可以被长期用来引用该项。
事实上，每个指向项的指针（`ItemPointer`，也称 `CTID`）在 PostgreSQL 中都由页号和项标识符索引构成。

项本身存储在从未分配空间末尾开始、向后分配的区域中。
其确切结构取决于表要包含什么内容。表和序列都使用一种名为 `HeapTupleHeaderData` 的结构体，如下所述。

最后一部分是“特殊部分”，其中可以存放访问方法希望保存的任何内容。例如，B-
树索引会在这里保存指向页面左、右兄弟的链接，以及其他一些与索引结构相关的数据。
普通表完全不使用特殊部分（通过将 `pd_special` 设为页面大小来表示）。

所有表行的结构都相同。它们都有一个固定大小的头部（在大多数机器上占 23 字节），后面跟着可选的空值位图、可选的对象 ID 字段以及用户数据。头部的详细格式见表 55.4。实际用户数据（行的各列）从 `t_hoff` 指示的偏移位置开始，它必须始终是该平台 MAXALIGN 对齐单位的整数倍。只有当 HEAP_HASNULL 位在 `t_infomask` 中被置位时，空值位图才存在。若存在，它紧随固定头部之后，并占用足够多的字节，以便为每个数据列提供一位（也就是说，总共为 `t_natts` 位）。在这个位图中，1 表示非空，0 表示空值。当位图不存在时，假定所有列都非空。只有当 HEAP_HASOID 位在 `t_infomask` 中被置位时，对象 ID 才存在。若存在，它位于 `t_hoff` 边界

之前。为了让 `t_hoff` 成为 MAXALIGN 的整数倍所需的任何填充，都会出现在空值位图与对象 ID 之间。（这又反过来保证了对象 ID 的对齐是合适的。）

表 55.4. HeapTupleHeaderData 布局

Field	Type	Length	Description
t_xmin	TransactionId	4 bytes	insert XID stamp
t_xmax	TransactionId	4 bytes	delete XID stamp
t_cid	CommandId	4 bytes	insert and/or delete CID stamp (overlays with t_xvac)
t_xvac	TransactionId	4 bytes	XID for VACUUM operat ion moving a row version
t_ctid	ItemPointerData	6 bytes	current TID of this or newer row version
t_infomask2	int16	2 bytes	number of attributes, plus various flag bits
t_infomask	uint16	2 bytes	various flag bits
t_hoff	uint8	1 byte	offset to user data

所有细节都可以在 `src/include/access/htup.h` 中找到。

要解释实际数据，必须借助从其他表中获得的信息，其中大部分来自 `pg_attribute`。识别字段位置所需的关键值是 `attlen` 和 `attalign`。除非所有字段都是定宽且没有空值，否则没有办法直接取得某个特定属性。所有这些技巧都封装在 heap_getattr、fastgetattr 和 heap_getsysattr 这些函数中。

读取数据时，需要依次检查每个属性。首先根据空值位图判断该字段是否为 NULL。如果是，就继续下一个。然后确认对齐是否正确。如果字段是定宽字段，那么它的所有字节都是直接摆放的；如果它是变长字段（attlen = -1），情况就会稍复杂一些。所有变长数据类型都共享一个通用头部结构体 `struct varlena`，其中包含已存储值的总长度以及一些标志位。根据这些标志，数据可能是线内存储的，也可能位于 TOAST 表中；它也可能是经过压缩的（见第 55.2 节）。

第 56 章 BKI 后端接口

后端接口（BKI）文件是使用一种特殊语言编写的脚本，PostgreSQL 后端在“引导”模式下运行时能够理解这种语言。引导模式允许从零开始创建并填充系统目录，而普通 SQL 命令要求目录已经存在。因此，BKI 文件可以用于最初创建数据库系统。（它们可能也没有其他用途。）

initdb 在创建新的数据库集簇时，使用 BKI 文件完成部分工作。initdb 使用的输入文件作为构建和安装 PostgreSQL 的一部分，由名为 `genbki.pl` 的程序创建。该程序读取源码树 `src/include/catalog/` 目录中采用特殊格式的 C 头文件。创建的 BKI 文件名为 `postgres.bki`，通常安装在安装树的 `share` 子目录中。

相关信息可以在 initdb 的文档中找到。

56.1. BKI 文件格式

本节描述 PostgreSQL 后端如何解释 BKI 文件。如果手边有一份 `postgres.bki` 文件作为示例，这段说明会更容易理解。

BKI 输入由一系列命令组成。根据命令语法的不同，每条命令由若干词元组成。词元通常由空白分隔，但如果不会产生歧义，也可以不分隔。没有专门的命令分隔符；
语法上不可能属于前一条命令的下一个词元，就会开始一条新命令。（通常为了清晰，
你会把新命令写在新的一行。）词元可以是某些关键字、特殊字符（圆括号、逗号等）、数字，
或者双引号字符串。所有内容都区分大小写。

以 # 开头的行会被忽略。

56.2. BKI 命令

`create` *`tablename`* *`tableoid`* [`bootstrap`] [`shared_relation`] [`without_oids`] [`rowtype_oid` *`oid`*] (*`name1`* = *`type1`* [, *`name2`* = *`type2`*, ...])

创建一个名为 *`tablename`*、OID 为 *`tableoid`* 的表，其列定义在括号中给出。

`bootstrap.c` 直接支持以下列类型：`bool`、`bytea`、`char`（1 字节）、`name`、`int2`、`int4`、`regproc`、`regclass`、`regtype`、`text`、`oid`、`tid`、`xid`、`cid`、`int2vector`、`oidvector`、`_int4`（数组）、`_text`（数组）、`_oid`（数组）、`_char`（数组）以及 `_aclitem`（数组）。虽然也可以创建包含其他类型列的表，但这必须等到 `pg_type` 创建完成并填入合适条目之后才能做到。（实际上，这意味着引导目录中只能使用这些列类型，而非引导目录则可以包含任何内置类型。）

指定 `bootstrap` 时，表只会在磁盘上创建；不会在 `pg_class`、`pg_attribute` 等目录中为它写入任何信息。因此，在通过 `insert` 命令以较为原始的方式补入这些条目之前，普通 SQL 操作无法访问该表。这个选项用于创建 `pg_class` 等目录自身。

如果指定了 `shared_relation`，该表就会作为共享表创建。除非指定 `without_oids`，否则该表将具有 OID。该表的行类型 OID（即 `pg_type` 中对应的 OID）还可以通过 `rowtype_oid` 子句显式指定；如果不指定，就会自动为其生成一个 OID。（如果指定了 `bootstrap`，那么 `rowtype_oid` 子句其实没有用，但仍然可以写上，作为文档说明。）

`open` *`tablename`*

打开名为 *`tablename`* 的表，以便插入数据。当前任何已打开的表都会被关闭。

close [*tablename*]

关闭当前打开的表。可以给出表名进行交叉校验，但并非必须。

insert [OID = *oid_value*] (*value1* *value2* ...)

向当前打开的表插入一行新数据，使用 *value1*、*value2* 等作为各列的值，使用 *oid_value* 作为其 OID。如果 *oid_value* 为零（0）或省略该子句，并且该表具有 OID，则分配下一个可用 OID。

NULL 值可以用特殊关键字 _null_ 表示。包含空格的值必须用双引号括起来。

declare [unique] index *indexname* *indexoid* on *tablename* using *amname* (*opclass1* *name1* [, ...])

在名为 *tablename* 的表上，使用 *amname* 访问方法，创建一个名为 *indexname*、OID 为 *indexoid* 的索引。要建立索引的字段分别称为 *name1*、*name2* 等；使用的操作符类分别为 *opclass1*、*opclass2* 等。该命令会创建索引文件并写入相应的目录条目，但不会初始化索引内容。

declare toast *toasttableoid* *toastindexoid* on *tablename*

为名为 *tablename* 的表创建一个 TOAST 表。该 TOAST 表的 OID 设为 *toasttableoid*，其索引的 OID 设为 *toastindexoid*。与 declare index 一样，索引填充会被延后。

build indices

填充之前已声明的索引。

56.3. 引导 BKI 文件的结构

在 open 命令可以使用之前，它所依赖的那些表必须已经存在，并且其中已经有了将要被打开之表对应的条目。（这些最基本的表包括 pg_class、pg_attribute、pg_proc 和 pg_type。）为了让这些表本身也能够被填充，带 bootstrap 选项的 create 命令会隐式打开刚创建的表，以便插入数据。

同样，declare index 和 declare toast 命令也必须等到它们所需的系统目录已经创建并填充完毕之后才能使用。

因此，postgres.bki 文件的结构必须如下：

1. 对某个关键表执行 create bootstrap
2. 用 insert 插入至少足以描述这些关键表的数据
3. close
4. 对其他关键表重复上述步骤。
5. 对某个非关键表执行 create（不带 bootstrap）
6. open
7. 用 insert 插入所需数据
8. close

9. 对其他非关键表重复上述步骤。
10. 定义索引和 TOAST 表。
11. `build indices`

无疑还存在其他未文档化的顺序依赖。

56.4. 示例

以下命令序列会创建表`test_table`，其 OID 为 420，包含两列`cola`和`colb`，类型分别为`int4`和`text`，并向表中插入两行：

```
create test_table 420 (cola = int4, colb = text)
open test_table
insert OID=421 ( 1 "value1" )
insert OID=422 ( 2 _null_ )
close test_table
```

第 57 章 规划器如何使用统计信息

本章建立在第 14.1 节和第 14.2 节所介绍的内容之上，进一步说明规划器如何利用系统统计信息来估计查询各部分可能返回的行数。这是规划过程中的重要组成部分，
为代价计算提供了大量原始材料。

本章的目的不是详细说明代码，而是概述其工作方式。
这或许能让随后想要阅读代码的人更容易入门。

57.1. 行估计示例

下面的示例使用 PostgreSQL 回归测试数据库中的表。所示输出取自版本 8.3，更早（或更晚）的版本行为可能有所不同。另外，由于 `ANALYZE` 在生成统计信息时使用随机采样，每次重新执行 `ANALYZE` 后，结果都会略有变化。

让我们从一个非常简单的查询开始：

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..458.00 rows=10000 width=244)
```

How the planner determines the cardinality of `tenk1` is covered in 第 14.2 节, but is repeated here for completeness. The number of pages and rows is looked up in `pg_class`:

```
SELECT relpages, reltuples FROM pg_class WHERE relname = 'tenk1';

 relpages | reltuples
----------+-----------
      358 |     10000
```

These numbers are current as of the last `VACUUM` or `ANALYZE` on the table. The planner then fetches the actual current number of pages in the table (this is a cheap operation, not requiring a table scan). If that is different from `relpages` then `reltuples` is scaled accordingly to arrive at a current number-of-rows estimate. In the example above, the values are correct so the rows estimate is the same as `reltuples`.

接着看一个包含范围条件的示例，其条件位于`WHERE`子句中：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 1000;

                                  QUERY PLAN
-----------------------------------------------------------------------
----------
 Bitmap Heap Scan on tenk1  (cost=24.06..394.64 rows=1007 width=244)
   Recheck Cond: (unique1 < 1000)
   ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..23.80 rows=1007
 width=0)
         Index Cond: (unique1 < 1000)
```

规划器检查`WHERE`子句中的条件，并查找操作符`<`的选择率函数，所在的系统目录是`pg_operator`。此函数保存在`oprrest`列中，本例的条目为`scalarltsel`。`scalarltsel`

函数取得unique1的直方图，来源为pg_statistics。手动查询时，查看更简单的pg_stats视图更方便：

```
SELECT histogram_bounds FROM pg_stats
WHERE tablename='tenk1' AND attname='unique1';

                   histogram_bounds
-------------------------------------------------------
 {0,993,1997,3050,4040,5036,5957,7057,8029,9016,9995}
```

接下来，计算“< 1000”在直方图中所占的比例，这就是选择率。直方图把范围划分为等频率的桶，因此只需找到目标值所在的桶，计入该桶的部分和前面所有桶的全部。值 1000 显然在第二个桶（993-1997）中。假设每个桶内部的值呈线性分布，可以按如下方式计算选择率：

```
selectivity = (1 + (1000 - bucket[2].min)/(bucket[2].max - bucket[2].
min))/num_buckets
            = (1 + (1000 - 993)/(1997 - 993))/10
            = 0.100697
```

即一个完整的桶加上第二个桶中的线性比例，再除以桶数。用选择率乘以表的基数，即可计算估计行数。这里所用的表为tenk1：

```
rows = rel_cardinality * selectivity
     = 10000 * 0.100697
     = 1007  (rounding off)
```

接下来，考虑一个包含等值条件的示例，其条件位于WHERE子句中：

```
EXPLAIN SELECT * FROM tenk1 WHERE stringu1 = 'CRAAAA';

                        QUERY PLAN
----------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..483.00 rows=30 width=244)
   Filter: (stringu1 = 'CRAAAA'::name)
```

规划器同样会检查WHERE子句中的条件，查找操作符=的选择率函数，即eqsel。
对于等值估计，直方图没有用；应该使用高频值（MCV）列表来确定选择率。查看 MCV，并同时查看后面会用到的其他一些列：

```
SELECT null_frac, n_distinct, most_common_vals, most_common_freqs FROM
 pg_stats
WHERE tablename='tenk1' AND attname='stringu1';

null_frac         | 0
n_distinct        | 676
most_common_vals  | {EJAAAA,BBAAAA,CRAAAA,FCAAAA,FEAAAA,GSAAAA,JOAAAA,
MCAAAA,NAAAAA,WGAAAA}
most_common_freqs | {0.00333333,0.003,0.003,0.003,0.003,0.003,0.003,0.
003,0.003,0.003}
```

由于CRAAAA出现在 MCV 列表中，选择率就是最常见值频率（MCF）列表中的对应条目：

```
selectivity = mcf[3]
            = 0.003
```

与之前一样，估计行数就是这个值与表的基数的乘积。这里所用的表为tenk1：

```
rows = 10000 * 0.003
     = 30
```

现在考虑同一个查询，但使用一个不在MCV列表中的常量：

```
EXPLAIN SELECT * FROM tenk1 WHERE stringu1 = 'xxx';

                        QUERY PLAN
-----------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..483.00 rows=15 width=244)
   Filter: (stringu1 = 'xxx'::name)
```

这是一个完全不同的问题：当值不在MCV列表中时，如何估计选择率。
方法是利用该值不在列表中的事实，再结合所有MCV的频率信息：

```
selectivity = (1 - sum(mvf))/(num_distinct - num_mcv)
            = (1 - (0.00333333 + 0.003 + 0.003 + 0.003 + 0.003 + 0.003
 +
                    0.003 + 0.003 + 0.003 + 0.003))/(676 - 10)
            = 0.0014559
```

也就是说，把所有MCV的频率加起来，用一减去这个总和，再除以其他非重复值的数量。
这相当于假设该列中不属于任何 MCV 的那部分值，均匀分布在其他所有非重复值上。注意，这里没有空值，所以不必考虑它们（否则还应从分子中减去空值比例）。然后照常计算估计行数：

```
rows = 10000 * 0.0014559
     = 15  (rounding off)
```

前面条件为`unique1 < 1000`的示例，过度简化了`scalarltsel`的实际行为。现在已经看过 MCV 的使用示例，可以补充一些细节。前面的示例就其涉及的范围而言是正确的，因为`unique1`是唯一列，所以没有 MCV（显然，没有哪个值会比其他值更常见）。对于非唯一列，通常既有直方图，也有 MCV 列表，而且直方图不包括该列总体中由 MCV 表示的那一部分。这样处理可以使估计更准确。在这种情况下，`scalarltsel`直接将条件（例如“< 1000”）应用于 MCV 列表中的每个值，并将满足条件的 MCV 的频率相加。这可以精确估计表中 MCV 所代表部分的选择率。然后按前面所述的方法使用直方图，估计表中非 MCV 部分的选择率，再将两个值结合起来，估计整体选择率。例如，考虑以下查询：

```
EXPLAIN SELECT * FROM tenk1 WHERE stringu1 < 'IAAAAA';

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..483.00 rows=3077 width=244)
   Filter: (stringu1 < 'IAAAAA'::name)
```

我们已经看过`stringu1`的 MCV 信息，下面是它的直方图：

```
SELECT histogram_bounds FROM pg_stats
WHERE tablename='tenk1' AND attname='stringu1';

                                 histogram_bounds
-----------------------------------------------------------------------
----------
 {AAAAAA,CQAAAA,FRAAAA,IBAAAA,KRAAAA,NFAAAA,PSAAAA,SGAAAA,VAAAAA,
XLAAAA,ZZAAAA}
```

检查 MCV 列表可以发现，条件`stringu1 < 'IAAAAA'`被前六个条目满足，而后四个条目不满足，因此总体中 MCV 部分的选择率为：

```
selectivity = sum(relevant mvfs)
            = 0.00333333 + 0.003 + 0.003 + 0.003 + 0.003 + 0.003
            = 0.01833333
```

将所有 MCF 相加，还可以得知 MCV 所代表的部分占总体的比例为 0.03033333，因此直方图所代表的部分占比为 0.96966667（这里同样没有空值，否则还需要将其排除）。可以看到，值IAAAAA接近第三个直方图桶的末尾。规划器对不同字符的频率作出一些粗略假设，
估计直方图所代表的总体中有 0.298387 的部分小于IAAAAA。然后将 MCV 和非 MCV 两部分的估计结合起来：

```
selectivity = mcv_selectivity + histogram_selectivity * histogram_
fraction
            = 0.01833333 + 0.298387 * 0.96966667
            = 0.307669

rows        = 10000 * 0.307669
            = 3077  (rounding off)
```

在这个特定的示例中，MCV 列表带来的修正相当小，因为该列的分布实际上相当均匀（统计信息显示这些特定值比其他值更常见，主要是采样误差造成的）。更常见的情况是，
某些值明显比其他值更常见；这时，这个复杂过程能够有效提高准确性，
因为最常见值的选择率是精确求得的。

现在考虑另一种情形，其中多个条件出现在WHERE子句中：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 1000 AND stringu1 = 'xxx';

                                  QUERY PLAN
----------------------------------------------------------------------
----------
 Bitmap Heap Scan on tenk1  (cost=23.80..396.91 rows=1 width=244)
   Recheck Cond: (unique1 < 1000)
   Filter: (stringu1 = 'xxx'::name)
   ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..23.80 rows=1007
 width=0)
         Index Cond: (unique1 < 1000)
```

规划器假定这两个条件相互独立，因此可以将各个子句的选择率相乘：

```
selectivity = selectivity(unique1 < 1000) * selectivity(stringu1 =
 'xxx')
            = 0.100697 * 0.0014559
            = 0.0001466

rows        = 10000 * 0.0001466
            = 1  (rounding off)
```

注意，位图索引扫描的估计返回行数只反映用于索引的条件。这一点很重要，
因为它会影响后续堆访问的代价估计。

最后，来看一个涉及连接的查询：

```
EXPLAIN SELECT * FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 50 AND t1.unique2 = t2.unique2;

                                        QUERY PLAN
```

```
-----------------------------------------------------------------------
----------------
 Nested Loop  (cost=4.64..456.23 rows=50 width=488)
   ->  Bitmap Heap Scan on tenk1 t1  (cost=4.64..142.17 rows=50 width=
244)
         Recheck Cond: (unique1 < 50)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..4.63
 rows=50 width=0)
               Index Cond: (unique1 < 50)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..6.27
 rows=1 width=244)
         Index Cond: (unique2 = t1.unique2)
```

对于tenk1, unique1 < 50这一限制会在嵌套循环连接之前求值。处理方法与之前的范围查询示例类似。这次，值 50 落在unique1直方图的第一个桶中：

```
selectivity = (0 + (50 - bucket[1].min)/(bucket[1].max - bucket[1].
min))/num_buckets
            = (0 + (50 - 0)/(993 - 0))/10
            = 0.005035

rows        = 10000 * 0.005035
            = 50  (rounding off)
```

连接的限制条件是t2.unique2 = t1.unique2。操作符就是熟悉的=，不过选择率函数来自oprjoin列，该列属于pg_operator，本例中为eqjoinsel。eqjoinsel
查询两个表的统计信息，这两个表是tenk2和tenk1：

```
SELECT tablename, null_frac,n_distinct, most_common_vals FROM pg_stats
WHERE tablename IN ('tenk1', 'tenk2') AND attname='unique2';

tablename  | null_frac | n_distinct | most_common_vals
-----------+-----------+------------+------------------
 tenk1     |         0 |         -1 |
 tenk2     |         0 |         -1 |
```

本例中，没有MCV信息可用于unique2，因为所有值看起来都是唯一的。
因此使用的算法仅依赖于两个关系的非重复值数量及其空值比例：

```
selectivity = (1 - null_frac1) * (1 - null_frac2) * min(1/num_
distinct1, 1/num_distinct2)
            = (1 - 0) * (1 - 0) / max(10000, 10000)
            = 0.0001
```

也就是说，对两个关系分别用一减去空值比例，再除以两者非重复值数量中的较大者。
预计连接输出的行数，通过将两个输入的笛卡尔积基数乘以选择率来计算：

```
rows = (outer_cardinality * inner_cardinality) * selectivity
     = (50 * 10000) * 0.0001
     = 50
```

如果这两列存在 MCV 列表，eqjoinsel就会通过直接比较 MCV 列表，确定由 MCV 表示的那部分列值总体中的连接选择率。剩余部分总体的估计则沿用这里展示的相同方法。

注意，我们把 inner_cardinality 写成了 10000，也就是 tenk2 未加修改的大小。从 EXPLAIN 输出来看，连接行数的估计似乎来自 50 * 1，即外层行数乘以对 tenk2

的每次内层索引扫描估计得到的行数。但事实并非如此：连接关系的大小是在考虑任何具体连接计划之前估计的。如果一切正常，这两种估计连接大小的方式会得到大致相同的答案，但由于舍入误差和其他因素，它们有时会明显分歧。

如果想了解更多细节，表大小（在任何`WHERE`子句之前）的估计是在`src/backend/optimizer/util/plancat.c`中完成的。子句选择率的一般逻辑位于`src/backend/optimizer/path/clausesel.c`。按操作符区分的选择率函数大多位于`src/backend/utils/adt/selfuncs.c`中。

部分 VIII. 附录

目录

A. PostgreSQL错误代码 1667
B. Date/Time Support 1675
B.1. Date/Time Input Interpretation 1675
B.2. Date/Time Key Words 1676
B.3. Date/Time Configuration Files 1677
B.4. History of Units 1678
C. SQL 关键字 1680
D. SQL 符合性 1705
D.1. 支持的特性 1706
D.2. 不支持的特性 1720
E. 发布说明 1733
E.1. 版本 9.1.24 1733
E.1.1. 迁移到版本 9.1.24 1733
E.1.2. 变更 1733
E.2. 版本 9.1.23 1734
E.2.1. 迁移到版本 9.1.23 1735
E.2.2. 变更 1735
E.3. 版本 9.1.22 1737
E.3.1. 迁移到版本 9.1.22 1737
E.3.2. 变更 1737
E.4. 版本 9.1.21 1738
E.4.1. 迁移到版本 9.1.21 1738
E.4.2. 变更 1738
E.5. 版本 9.1.20 1739
E.5.1. 迁移到版本 9.1.20 1739
E.5.2. 变更 1739
E.6. 版本 9.1.19 1743
E.6.1. 迁移到版本 9.1.19 1743
E.6.2. 变更 1743
E.7. 版本 9.1.18 1746
E.7.1. 迁移到版本 9.1.18 1746
E.7.2. 变更 1746
E.8. 版本 9.1.17 1747
E.8.1. 迁移到版本 9.1.17 1747
E.8.2. 变更 1747
E.9. 版本 9.1.16 1747
E.9.1. 迁移到版本 9.1.16 1747
E.9.2. 变更 1748
E.10. 版本 9.1.15 1751
E.10.1. 迁移到版本 9.1.15 1751
E.10.2. 变更 1751
E.11. 版本 9.1.14 1756
E.11.1. 迁移到版本 9.1.14 1756
E.11.2. 变更 1756
E.12. 版本 9.1.13 1759
E.12.1. 迁移到版本 9.1.13 1759
E.12.2. 变更 1759
E.13. 版本 9.1.12 1760
E.13.1. 迁移到版本 9.1.12 1760
E.13.2. 变更 1760
E.14. 版本 9.1.11 1764

E.14.1. 迁移到版本 9.1.11 1764
E.14.2. 变更 1764
E.15. 版本 9.1.10 1765
E.15.1. 迁移到版本 9.1.10 1765
E.15.2. 变更 1766
E.16. 版本 9.1.9 1767
E.16.1. 迁移到版本 9.1.9 1767
E.16.2. 变更 1768
E.17. 版本 9.1.8 1770
E.17.1. 迁移到版本 9.1.8 1770
E.17.2. 变更 1770
E.18. 版本 9.1.7 1772
E.18.1. 迁移到版本 9.1.7 1772
E.18.2. 变更 1772
E.19. 版本 9.1.6 1775
E.19.1. 迁移到版本 9.1.6 1775
E.19.2. 变更 1775
E.20. 版本 9.1.5 1776
E.20.1. 迁移到版本 9.1.5 1776
E.20.2. 变更 1777
E.21. 版本 9.1.4 1779
E.21.1. 迁移到版本 9.1.4 1779
E.21.2. 变更 1779
E.22. 版本 9.1.3 1782
E.22.1. 迁移到版本 9.1.3 1782
E.22.2. 变更 1782
E.23. 版本 9.1.2 1786
E.23.1. 迁移到版本 9.1.2 1786
E.23.2. 变更 1787
E.24. 版本 9.1.1 1790
E.24.1. 迁移到版本 9.1.1 1791
E.24.2. 变更 1791
E.25. 版本 9.1 1791
E.25.1. 概述 1791
E.25.2. 迁移到版本 9.1 1792
E.25.3. 变更 1794
E.26. 版本 9.0.23 1808
E.26.1. 迁移到版本 9.0.23 1808
E.26.2. 变更 1808
E.27. 版本 9.0.22 1811
E.27.1. 迁移到版本 9.0.22 1811
E.27.2. 变更 1811
E.28. 版本 9.0.21 1811
E.28.1. 迁移到版本 9.0.21 1812
E.28.2. 变更 1812
E.29. 版本 9.0.20 1812
E.29.1. 迁移到版本 9.0.20 1812
E.29.2. 变更 1812
E.30. 版本 9.0.19 1815
E.30.1. 迁移到版本 9.0.19 1815
E.30.2. 变更 1815
E.31. 版本 9.0.18 1820
E.31.1. 迁移到版本 9.0.18 1820
E.31.2. 变更 1820

E.32. 版本 9.0.17 1822
E.32.1. 迁移到版本 9.0.17 1822
E.32.2. 变更 1822
E.33. 版本 9.0.16 1823
E.33.1. 迁移到版本 9.0.16 1823
E.33.2. 变更 1824
E.34. 版本 9.0.15 1827
E.34.1. 迁移到版本 9.0.15 1827
E.34.2. 变更 1827
E.35. 版本 9.0.14 1828
E.35.1. 迁移到版本 9.0.14 1828
E.35.2. 变更 1829
E.36. 版本 9.0.13 1830
E.36.1. 迁移到版本 9.0.13 1830
E.36.2. 变更 1830
E.37. 版本 9.0.12 1832
E.37.1. 迁移到版本 9.0.12 1832
E.37.2. 变更 1832
E.38. 版本 9.0.11 1834
E.38.1. 迁移到版本 9.0.11 1834
E.38.2. 变更 1834
E.39. 版本 9.0.10 1836
E.39.1. 迁移到版本 9.0.10 1836
E.39.2. 变更 1836
E.40. 版本 9.0.9 1837
E.40.1. 迁移到版本 9.0.9 1837
E.40.2. 变更 1837
E.41. 版本 9.0.8 1839
E.41.1. 迁移到版本 9.0.8 1839
E.41.2. 变更 1839
E.42. 版本 9.0.7 1841
E.42.1. 迁移到版本 9.0.7 1841
E.42.2. 变更 1842
E.43. 版本 9.0.6 1845
E.43.1. 迁移到版本 9.0.6 1845
E.43.2. 变更 1845
E.44. 版本 9.0.5 1847
E.44.1. 迁移到版本 9.0.5 1847
E.44.2. 变更 1848
E.45. 版本 9.0.4 1851
E.45.1. 迁移到版本 9.0.4 1851
E.45.2. 变更 1852
E.46. 版本 9.0.3 1853
E.46.1. 迁移到版本 9.0.3 1854
E.46.2. 变更 1854
E.47. 版本 9.0.2 1855
E.47.1. 迁移到版本 9.0.2 1855
E.47.2. 变更 1855
E.48. 版本 9.0.1 1857
E.48.1. 迁移到版本 9.0.1 1858
E.48.2. 变更 1858
E.49. 版本 9.0 1858
E.49.1. 概述 1859
E.49.2. 迁移到版本 9.0 1859

E.49.3. 变更 1861
E.50. 版本 8.4.22 1877
E.50.1. 迁移到版本 8.4.22 1877
E.50.2. 变更 1878
E.51. 版本 8.4.21 1879
E.51.1. 迁移到版本 8.4.21 1880
E.51.2. 变更 1880
E.52. 版本 8.4.20 1880
E.52.1. 迁移到版本 8.4.20 1881
E.52.2. 变更 1881
E.53. 版本 8.4.19 1883
E.53.1. 迁移到版本 8.4.19 1884
E.53.2. 变更 1884
E.54. 版本 8.4.18 1885
E.54.1. 迁移到版本 8.4.18 1885
E.54.2. 变更 1885
E.55. 版本 8.4.17 1886
E.55.1. 迁移到版本 8.4.17 1886
E.55.2. 变更 1886
E.56. 版本 8.4.16 1888
E.56.1. 迁移到版本 8.4.16 1888
E.56.2. 变更 1888
E.57. 版本 8.4.15 1889
E.57.1. 迁移到版本 8.4.15 1889
E.57.2. 变更 1889
E.58. 版本 8.4.14 1891
E.58.1. 迁移到版本 8.4.14 1891
E.58.2. 变更 1891
E.59. 版本 8.4.13 1892
E.59.1. 迁移到版本 8.4.13 1892
E.59.2. 变更 1892
E.60. 版本 8.4.12 1893
E.60.1. 迁移到版本 8.4.12 1894
E.60.2. 变更 1894
E.61. 版本 8.4.11 1895
E.61.1. 迁移到版本 8.4.11 1895
E.61.2. 变更 1896
E.62. 版本 8.4.10 1898
E.62.1. 迁移到版本 8.4.10 1898
E.62.2. 变更 1898
E.63. 版本 8.4.9 1900
E.63.1. 迁移到版本 8.4.9 1900
E.63.2. 变更 1900
E.64. 版本 8.4.8 1903
E.64.1. 迁移到版本 8.4.8 1903
E.64.2. 变更 1904
E.65. 版本 8.4.7 1905
E.65.1. 迁移到版本 8.4.7 1905
E.65.2. 变更 1905
E.66. 版本 8.4.6 1906
E.66.1. 迁移到版本 8.4.6 1906
E.66.2. 变更 1906
E.67. 版本 8.4.5 1908
E.67.1. 迁移到版本 8.4.5 1908

E.67.2. 变更 1908
E.68. 版本 8.4.4 1911
E.68.1. 迁移到版本 8.4.4 1911
E.68.2. 变更 1911
E.69. 版本 8.4.3 1913
E.69.1. 迁移到版本 8.4.3 1913
E.69.2. 变更 1913
E.70. 版本 8.4.2 1916
E.70.1. 迁移到版本 8.4.2 1916
E.70.2. 变更 1916
E.71. 版本 8.4.1 1919
E.71.1. 迁移到版本 8.4.1 1920
E.71.2. 变更 1920
E.72. 版本 8.4 1921
E.72.1. 概述 1921
E.72.2. 迁移到版本 8.4 1922
E.72.3. 变更 1924
E.73. 版本 8.3.23 1941
E.73.1. 迁移到版本 8.3.23 1941
E.73.2. 变更 1941
E.74. 版本 8.3.22 1942
E.74.1. 迁移到版本 8.3.22 1942
E.74.2. 变更 1942
E.75. 版本 8.3.21 1944
E.75.1. 迁移到版本 8.3.21 1944
E.75.2. 变更 1944
E.76. 版本 8.3.20 1945
E.76.1. 迁移到版本 8.3.20 1945
E.76.2. 变更 1945
E.77. 版本 8.3.19 1946
E.77.1. 迁移到版本 8.3.19 1946
E.77.2. 变更 1947
E.78. 版本 8.3.18 1948
E.78.1. 迁移到版本 8.3.18 1948
E.78.2. 变更 1948
E.79. 版本 8.3.17 1950
E.79.1. 迁移到版本 8.3.17 1950
E.79.2. 变更 1950
E.80. 版本 8.3.16 1952
E.80.1. 迁移到版本 8.3.16 1952
E.80.2. 变更 1952
E.81. 版本 8.3.15 1954
E.81.1. 迁移到版本 8.3.15 1954
E.81.2. 变更 1954
E.82. 版本 8.3.14 1955
E.82.1. 迁移到版本 8.3.14 1956
E.82.2. 变更 1956
E.83. 版本 8.3.13 1956
E.83.1. 迁移到版本 8.3.13 1957
E.83.2. 变更 1957
E.84. 版本 8.3.12 1958
E.84.1. 迁移到版本 8.3.12 1958
E.84.2. 变更 1958
E.85. 版本 8.3.11 1961

E.85.1. 迁移到版本 8.3.11 1961
E.85.2. 变更 1961
E.86. 版本 8.3.10 1962
E.86.1. 迁移到版本 8.3.10 1962
E.86.2. 变更 1962
E.87. 版本 8.3.9 1964
E.87.1. 迁移到版本 8.3.9 1964
E.87.2. 变更 1964
E.88. 版本 8.3.8 1966
E.88.1. 迁移到版本 8.3.8 1966
E.88.2. 变更 1967
E.89. 版本 8.3.7 1968
E.89.1. 迁移到版本 8.3.7 1968
E.89.2. 变更 1968
E.90. 版本 8.3.6 1970
E.90.1. 迁移到版本 8.3.6 1970
E.90.2. 变更 1970
E.91. 版本 8.3.5 1972
E.91.1. 迁移到版本 8.3.5 1972
E.91.2. 变更 1972
E.92. 版本 8.3.4 1973
E.92.1. 迁移到版本 8.3.4 1973
E.92.2. 变更 1973
E.93. 版本 8.3.3 1975
E.93.1. 迁移到版本 8.3.3 1976
E.93.2. 变更 1976
E.94. 版本 8.3.2 1976
E.94.1. 迁移到版本 8.3.2 1976
E.94.2. 变更 1976
E.95. 版本 8.3.1 1978
E.95.1. 迁移到版本 8.3.1 1979
E.95.2. 变更 1979
E.96. 版本 8.3 1980
E.96.1. 概述 1981
E.96.2. 迁移到版本 8.3 1981
E.96.3. 变更 1984
E.97. 版本 8.2.23 1996
E.97.1. 迁移到版本 8.2.23 1996
E.97.2. 变更 1996
E.98. 版本 8.2.22 1997
E.98.1. 迁移到版本 8.2.22 1998
E.98.2. 变更 1998
E.99. 版本 8.2.21 1999
E.99.1. 迁移到版本 8.2.21 2000
E.99.2. 变更 2000
E.100. 版本 8.2.20 2000
E.100.1. 迁移到版本 8.2.20 2001
E.100.2. 变更 2001
E.101. 版本 8.2.19 2001
E.101.1. 迁移到版本 8.2.19 2002
E.101.2. 变更 2002
E.102. 版本 8.2.18 2003
E.102.1. 迁移到版本 8.2.18 2003
E.102.2. 变更 2003

E.103. 版本 8.2.17 2005
E.103.1. 迁移到版本 8.2.17 2005
E.103.2. 变更 2005
E.104. 版本 8.2.16 2006
E.104.1. 迁移到版本 8.2.16 2007
E.104.2. 变更 2007
E.105. 版本 8.2.15 2008
E.105.1. 迁移到版本 8.2.15 2008
E.105.2. 变更 2008
E.106. 版本 8.2.14 2010
E.106.1. 迁移到版本 8.2.14 2010
E.106.2. 变更 2010
E.107. 版本 8.2.13 2011
E.107.1. 迁移到版本 8.2.13 2011
E.107.2. 变更 2011
E.108. 版本 8.2.12 2012
E.108.1. 迁移到版本 8.2.12 2013
E.108.2. 变更 2013
E.109. 版本 8.2.11 2014
E.109.1. 迁移到版本 8.2.11 2014
E.109.2. 变更 2014
E.110. 版本 8.2.10 2015
E.110.1. 迁移到版本 8.2.10 2015
E.110.2. 变更 2015
E.111. 版本 8.2.9 2016
E.111.1. 迁移到版本 8.2.9 2017
E.111.2. 变更 2017
E.112. 版本 8.2.8 2017
E.112.1. 迁移到版本 8.2.8 2017
E.112.2. 变更 2017
E.113. 版本 8.2.7 2018
E.113.1. 迁移到版本 8.2.7 2018
E.113.2. 变更 2019
E.114. 版本 8.2.6 2020
E.114.1. 迁移到版本 8.2.6 2020
E.114.2. 变更 2020
E.115. 版本 8.2.5 2022
E.115.1. 迁移到版本 8.2.5 2022
E.115.2. 变更 2022
E.116. 版本 8.2.4 2023
E.116.1. 迁移到版本 8.2.4 2023
E.116.2. 变更 2023
E.117. 版本 8.2.3 2024
E.117.1. 迁移到版本 8.2.3 2024
E.117.2. 变更 2024
E.118. 版本 8.2.2 2024
E.118.1. 迁移到版本 8.2.2 2024
E.118.2. 变更 2025
E.119. 版本 8.2.1 2025
E.119.1. 迁移到版本 8.2.1 2026
E.119.2. 变更 2026
E.120. 版本 8.2 2026
E.120.1. 概述 2026
E.120.2. 迁移到版本 8.2 2027

E.120.3. 变更 2029
E.121. 版本 8.1.23 2041
E.121.1. 迁移到版本 8.1.23 2041
E.121.2. 变更 2041
E.122. 版本 8.1.22 2042
E.122.1. 迁移到版本 8.1.22 2042
E.122.2. 变更 2042
E.123. 版本 8.1.21 2044
E.123.1. 迁移到版本 8.1.21 2044
E.123.2. 变更 2044
E.124. 版本 8.1.20 2045
E.124.1. 迁移到版本 8.1.20 2045
E.124.2. 变更 2045
E.125. 版本 8.1.19 2046
E.125.1. 迁移到版本 8.1.19 2046
E.125.2. 变更 2046
E.126. 版本 8.1.18 2047
E.126.1. 迁移到版本 8.1.18 2047
E.126.2. 变更 2048
E.127. 版本 8.1.17 2048
E.127.1. 迁移到版本 8.1.17 2049
E.127.2. 变更 2049
E.128. 版本 8.1.16 2049
E.128.1. 迁移到版本 8.1.16 2049
E.128.2. 变更 2050
E.129. 版本 8.1.15 2050
E.129.1. 迁移到版本 8.1.15 2050
E.129.2. 变更 2050
E.130. 版本 8.1.14 2051
E.130.1. 迁移到版本 8.1.14 2051
E.130.2. 变更 2051
E.131. 版本 8.1.13 2052
E.131.1. 迁移到版本 8.1.13 2053
E.131.2. 变更 2053
E.132. 版本 8.1.12 2053
E.132.1. 迁移到版本 8.1.12 2053
E.132.2. 变更 2053
E.133. 版本 8.1.11 2055
E.133.1. 迁移到版本 8.1.11 2055
E.133.2. 变更 2055
E.134. 版本 8.1.10 2056
E.134.1. 迁移到版本 8.1.10 2057
E.134.2. 变更 2057
E.135. 版本 8.1.9 2057
E.135.1. 迁移到版本 8.1.9 2057
E.135.2. 变更 2057
E.136. 版本 8.1.8 2058
E.136.1. 迁移到版本 8.1.8 2058
E.136.2. 变更 2058
E.137. 版本 8.1.7 2058
E.137.1. 迁移到版本 8.1.7 2058
E.137.2. 变更 2059
E.138. 版本 8.1.6 2059
E.138.1. 迁移到版本 8.1.6 2059

E.138.2. 变更 2059
E.139. 版本 8.1.5 2060
E.139.1. 迁移到版本 8.1.5 2060
E.139.2. 变更 2060
E.140. 版本 8.1.4 2061
E.140.1. 迁移到版本 8.1.4 2061
E.140.2. 变更 2062
E.141. 版本 8.1.3 2063
E.141.1. 迁移到版本 8.1.3 2063
E.141.2. 变更 2063
E.142. 版本 8.1.2 2064
E.142.1. 迁移到版本 8.1.2 2065
E.142.2. 变更 2065
E.143. 版本 8.1.1 2066
E.143.1. 迁移到版本 8.1.1 2066
E.143.2. 变更 2066
E.144. 版本 8.1 2067
E.144.1. 概述 2067
E.144.2. 迁移到版本 8.1 2068
E.144.3. 附加变更 2070
E.145. 版本 8.0.26 2080
E.145.1. 迁移到版本 8.0.26 2080
E.145.2. 变更 2080
E.146. 版本 8.0.25 2082
E.146.1. 迁移到版本 8.0.25 2082
E.146.2. 变更 2082
E.147. 版本 8.0.24 2083
E.147.1. 迁移到版本 8.0.24 2083
E.147.2. 变更 2083
E.148. 版本 8.0.23 2084
E.148.1. 迁移到版本 8.0.23 2084
E.148.2. 变更 2084
E.149. 版本 8.0.22 2085
E.149.1. 迁移到版本 8.0.22 2085
E.149.2. 变更 2085
E.150. 版本 8.0.21 2086
E.150.1. 迁移到版本 8.0.21 2086
E.150.2. 变更 2086
E.151. 版本 8.0.20 2087
E.151.1. 迁移到版本 8.0.20 2087
E.151.2. 变更 2087
E.152. 版本 8.0.19 2087
E.152.1. 迁移到版本 8.0.19 2088
E.152.2. 变更 2088
E.153. 版本 8.0.18 2088
E.153.1. 迁移到版本 8.0.18 2088
E.153.2. 变更 2088
E.154. 版本 8.0.17 2089
E.154.1. 迁移到版本 8.0.17 2089
E.154.2. 变更 2089
E.155. 版本 8.0.16 2090
E.155.1. 迁移到版本 8.0.16 2090
E.155.2. 变更 2090
E.156. 版本 8.0.15 2091

E.156.1. 迁移到版本 8.0.15 2091
E.156.2. 变更 2092
E.157. 版本 8.0.14 2093
E.157.1. 迁移到版本 8.0.14 2093
E.157.2. 变更 2093
E.158. 版本 8.0.13 2093
E.158.1. 迁移到版本 8.0.13 2094
E.158.2. 变更 2094
E.159. 版本 8.0.12 2094
E.159.1. 迁移到版本 8.0.12 2094
E.159.2. 变更 2094
E.160. 版本 8.0.11 2094
E.160.1. 迁移到版本 8.0.11 2095
E.160.2. 变更 2095
E.161. 版本 8.0.10 2095
E.161.1. 迁移到版本 8.0.10 2095
E.161.2. 变更 2095
E.162. 版本 8.0.9 2096
E.162.1. 迁移到版本 8.0.9 2096
E.162.2. 变更 2096
E.163. 版本 8.0.8 2097
E.163.1. 迁移到版本 8.0.8 2097
E.163.2. 变更 2097
E.164. 版本 8.0.7 2098
E.164.1. 迁移到版本 8.0.7 2098
E.164.2. 变更 2098
E.165. 版本 8.0.6 2099
E.165.1. 迁移到版本 8.0.6 2099
E.165.2. 变更 2099
E.166. 版本 8.0.5 2100
E.166.1. 迁移到版本 8.0.5 2100
E.166.2. 变更 2100
E.167. 版本 8.0.4 2101
E.167.1. 迁移到版本 8.0.4 2101
E.167.2. 变更 2101
E.168. 版本 8.0.3 2102
E.168.1. 迁移到版本 8.0.3 2103
E.168.2. 变更 2103
E.169. 版本 8.0.2 2104
E.169.1. 迁移到版本 8.0.2 2104
E.169.2. 变更 2104
E.170. 版本 8.0.1 2106
E.170.1. 迁移到版本 8.0.1 2106
E.170.2. 变更 2106
E.171. 版本 8.0 2107
E.171.1. 概述 2107
E.171.2. 迁移到版本 8.0 2108
E.171.3. 废弃特性 2109
E.171.4. 变更 2109
E.172. 版本 7.4.30 2121
E.172.1. 迁移到版本 7.4.30 2121
E.172.2. 变更 2121
E.173. 版本 7.4.29 2122
E.173.1. 迁移到版本 7.4.29 2122

E.173.2. 变更 2122
E.174. 版本 7.4.28 2123
E.174.1. 迁移到版本 7.4.28 2123
E.174.2. 变更 2123
E.175. 版本 7.4.27 2124
E.175.1. 迁移到版本 7.4.27 2124
E.175.2. 变更 2124
E.176. 版本 7.4.26 2125
E.176.1. 迁移到版本 7.4.26 2125
E.176.2. 变更 2125
E.177. 版本 7.4.25 2126
E.177.1. 迁移到版本 7.4.25 2126
E.177.2. 变更 2126
E.178. 版本 7.4.24 2126
E.178.1. 迁移到版本 7.4.24 2126
E.178.2. 变更 2127
E.179. 版本 7.4.23 2127
E.179.1. 迁移到版本 7.4.23 2127
E.179.2. 变更 2127
E.180. 版本 7.4.22 2128
E.180.1. 迁移到版本 7.4.22 2128
E.180.2. 变更 2128
E.181. 版本 7.4.21 2128
E.181.1. 迁移到版本 7.4.21 2128
E.181.2. 变更 2128
E.182. 版本 7.4.20 2129
E.182.1. 迁移到版本 7.4.20 2129
E.182.2. 变更 2129
E.183. 版本 7.4.19 2130
E.183.1. 迁移到版本 7.4.19 2130
E.183.2. 变更 2130
E.184. 版本 7.4.18 2131
E.184.1. 迁移到版本 7.4.18 2131
E.184.2. 变更 2131
E.185. 版本 7.4.17 2131
E.185.1. 迁移到版本 7.4.17 2132
E.185.2. 变更 2132
E.186. 版本 7.4.16 2132
E.186.1. 迁移到版本 7.4.16 2132
E.186.2. 变更 2132
E.187. 版本 7.4.15 2133
E.187.1. 迁移到版本 7.4.15 2133
E.187.2. 变更 2133
E.188. 版本 7.4.14 2133
E.188.1. 迁移到版本 7.4.14 2133
E.188.2. 变更 2133
E.189. 版本 7.4.13 2134
E.189.1. 迁移到版本 7.4.13 2134
E.189.2. 变更 2134
E.190. 版本 7.4.12 2135
E.190.1. 迁移到版本 7.4.12 2135
E.190.2. 变更 2135
E.191. 版本 7.4.11 2136
E.191.1. 迁移到版本 7.4.11 2136

E.191.2. 变更 2136
E.192. 版本 7.4.10 2136
E.192.1. 迁移到版本 7.4.10 2136
E.192.2. 变更 2137
E.193. 版本 7.4.9 2137
E.193.1. 迁移到版本 7.4.9 2137
E.193.2. 变更 2137
E.194. 版本 7.4.8 2138
E.194.1. 迁移到版本 7.4.8 2138
E.194.2. 变更 2139
E.195. 版本 7.4.7 2140
E.195.1. 迁移到版本 7.4.7 2140
E.195.2. 变更 2140
E.196. 版本 7.4.6 2141
E.196.1. 迁移到版本 7.4.6 2141
E.196.2. 变更 2141
E.197. 版本 7.4.5 2142
E.197.1. 迁移到版本 7.4.5 2142
E.197.2. 变更 2142
E.198. 版本 7.4.4 2142
E.198.1. 迁移到版本 7.4.4 2143
E.198.2. 变更 2143
E.199. 版本 7.4.3 2143
E.199.1. 迁移到版本 7.4.3 2143
E.199.2. 变更 2143
E.200. 版本 7.4.2 2144
E.200.1. 迁移到版本 7.4.2 2144
E.200.2. 变更 2145
E.201. 版本 7.4.1 2146
E.201.1. 迁移到版本 7.4.1 2146
E.201.2. 变更 2147
E.202. 版本 7.4 2148
E.202.1. 概述 2148
E.202.2. 迁移到版本 7.4 2149
E.202.3. 变更 2150
E.203. 版本 7.3.21 2163
E.203.1. 迁移到版本 7.3.21 2163
E.203.2. 变更 2163
E.204. 版本 7.3.20 2164
E.204.1. 迁移到版本 7.3.20 2164
E.204.2. 变更 2164
E.205. 版本 7.3.19 2164
E.205.1. 迁移到版本 7.3.19 2164
E.205.2. 变更 2165
E.206. 版本 7.3.18 2165
E.206.1. 迁移到版本 7.3.18 2165
E.206.2. 变更 2165
E.207. 版本 7.3.17 2165
E.207.1. 迁移到版本 7.3.17 2165
E.207.2. 变更 2166
E.208. 版本 7.3.16 2166
E.208.1. 迁移到版本 7.3.16 2166
E.208.2. 变更 2166
E.209. 版本 7.3.15 2166

E.209.1. 迁移到版本 7.3.15 2166
E.209.2. 变更 2167
E.210. 版本 7.3.14 2167
E.210.1. 迁移到版本 7.3.14 2167
E.210.2. 变更 2168
E.211. 版本 7.3.13 2168
E.211.1. 迁移到版本 7.3.13 2168
E.211.2. 变更 2168
E.212. 版本 7.3.12 2169
E.212.1. 迁移到版本 7.3.12 2169
E.212.2. 变更 2169
E.213. 版本 7.3.11 2169
E.213.1. 迁移到版本 7.3.11 2169
E.213.2. 变更 2169
E.214. 版本 7.3.10 2170
E.214.1. 迁移到版本 7.3.10 2170
E.214.2. 变更 2171
E.215. 版本 7.3.9 2171
E.215.1. 迁移到版本 7.3.9 2172
E.215.2. 变更 2172
E.216. 版本 7.3.8 2172
E.216.1. 迁移到版本 7.3.8 2172
E.216.2. 变更 2172
E.217. 版本 7.3.7 2173
E.217.1. 迁移到版本 7.3.7 2173
E.217.2. 变更 2173
E.218. 版本 7.3.6 2173
E.218.1. 迁移到版本 7.3.6 2173
E.218.2. 变更 2173
E.219. 版本 7.3.5 2174
E.219.1. 迁移到版本 7.3.5 2174
E.219.2. 变更 2174
E.220. 版本 7.3.4 2175
E.220.1. 迁移到版本 7.3.4 2175
E.220.2. 变更 2175
E.221. 版本 7.3.3 2176
E.221.1. 迁移到版本 7.3.3 2176
E.221.2. 变更 2176
E.222. 版本 7.3.2 2178
E.222.1. 迁移到版本 7.3.2 2179
E.222.2. 变更 2179
E.223. 版本 7.3.1 2180
E.223.1. 迁移到版本 7.3.1 2180
E.223.2. 变更 2180
E.224. 版本 7.3 2181
E.224.1. 概述 2181
E.224.2. 迁移到版本 7.3 2182
E.224.3. 变更 2182
E.225. 版本 7.2.8 2193
E.225.1. 迁移到版本 7.2.8 2193
E.225.2. 变更 2193
E.226. 版本 7.2.7 2194
E.226.1. 迁移到版本 7.2.7 2194
E.226.2. 变更 2194

E.227. 版本 7.2.6 2194
E.227.1. 迁移到版本 7.2.6 2195
E.227.2. 变更 2195
E.228. 版本 7.2.5 2195
E.228.1. 迁移到版本 7.2.5 2195
E.228.2. 变更 2195
E.229. 版本 7.2.4 2196
E.229.1. 迁移到版本 7.2.4 2196
E.229.2. 变更 2196
E.230. 版本 7.2.3 2196
E.230.1. 迁移到版本 7.2.3 2196
E.230.2. 变更 2197
E.231. 版本 7.2.2 2197
E.231.1. 迁移到版本 7.2.2 2197
E.231.2. 变更 2197
E.232. 版本 7.2.1 2198
E.232.1. 迁移到版本 7.2.1 2198
E.232.2. 变更 2198
E.233. 版本 7.2 2198
E.233.1. 概述 2199
E.233.2. 迁移到版本 7.2 2199
E.233.3. 变更 2200
E.234. 版本 7.1.3 2209
E.234.1. 迁移到版本 7.1.3 2209
E.234.2. 变更 2209
E.235. 版本 7.1.2 2209
E.235.1. 迁移到版本 7.1.2 2209
E.235.2. 变更 2210
E.236. 版本 7.1.1 2210
E.236.1. 迁移到版本 7.1.1 2210
E.236.2. 变更 2210
E.237. 版本 7.1 2210
E.237.1. 迁移到版本 7.1 2211
E.237.2. 变更 2211
E.238. 版本 7.0.3 2215
E.238.1. 迁移到版本 7.0.3 2215
E.238.2. 变更 2215
E.239. 版本 7.0.2 2216
E.239.1. 迁移到版本 7.0.2 2216
E.239.2. 变更 2216
E.240. 版本 7.0.1 2216
E.240.1. 迁移到版本 7.0.1 2216
E.240.2. 变更 2216
E.241. 版本 7.0 2217
E.241.1. 迁移到版本 7.0 2217
E.241.2. 变更 2218
E.242. 版本 6.5.3 2224
E.242.1. 迁移到版本 6.5.3 2224
E.242.2. 变更 2224
E.243. 版本 6.5.2 2224
E.243.1. 迁移到版本 6.5.2 2224
E.243.2. 变更 2224
E.244. 版本 6.5.1 2225
E.244.1. 迁移到版本 6.5.1 2225

E.244.2. 变更 2225
E.245. 版本 6.5 2226
E.245.1. 迁移到版本 6.5 2227
E.245.2. 变更 2227
E.246. 版本 6.4.2 2230
E.246.1. 迁移到版本 6.4.2 2231
E.246.2. 变更 2231
E.247. 版本 6.4.1 2231
E.247.1. 迁移到版本 6.4.1 2231
E.247.2. 变更 2231
E.248. 版本 6.4 2232
E.248.1. 迁移到版本 6.4 2232
E.248.2. 变更 2232
E.249. 版本 6.3.2 2236
E.249.1. 变更 2236
E.250. 版本 6.3.1 2237
E.250.1. 变更 2237
E.251. 版本 6.3 2238
E.251.1. 迁移到版本 6.3 2239
E.251.2. 变更 2239
E.252. 版本 6.2.1 2242
E.252.1. 从版本 6.2 迁移到版本 6.2.1 2243
E.252.2. 变更 2243
E.253. 版本 6.2 2243
E.253.1. 从版本 6.1 迁移到版本 6.2 2243
E.253.2. Migration from version 1.x to version 6.2 2244
E.253.3. 变更 2244
E.254. 版本 6.1.1 2246
E.254.1. 从版本 6.1 迁移到版本 6.1.1 2246
E.254.2. 变更 2246
E.255. 版本 6.1 2246
E.255.1. 迁移到版本 6.1 2247
E.255.2. 变更 2247
E.256. 版本 6.0 2249
E.256.1. 从版本 1.09 迁移到版本 6.0 2249
E.256.2. 从 1.09 之前版本迁移到版本 6.0 2249
E.256.3. 变更 2249
E.257. 版本 1.09 2251
E.258. 版本 1.02 2251
E.258.1. 从版本 1.02 迁移到版本 1.02.1 2251
E.258.2. 转储/重载程序 2252
E.258.3. 变更 2252
E.259. 版本 1.01 2253
E.259.1. 从版本 1.0 迁移到版本 1.01 2253
E.259.2. 变更 2255
E.260. 版本 1.0 2255
E.260.1. 变更 2255
E.261. Postgres95 Release 0.03 2256
E.261.1. 变更 2256
E.262. Postgres95 Release 0.02 2258
E.262.1. 变更 2258
E.263. Postgres95 Release 0.01 2258
F. 额外提供的模块与扩展 2259

F.1. adminpack 2260
F.2. auth_delay 2262
F.2.1. 配置参数 2262
F.2.2. 作者 2262
F.3. auto_explain 2263
F.3.1. 配置参数 2263
F.3.2. 示例 2264
F.3.3. 作者 2264
F.4. btree_gin 2265
F.4.1. 用法示例 2265
F.4.2. 作者 2265
F.5. btree_gist 2266
F.5.1. 用法示例 2266
F.5.2. 作者 2267
F.6. chkpass 2268
F.6.1. 作者 2268
F.7. citext – 大小写不敏感的字符串类型 2269
F.7.1. 原理 2269
F.7.2. 如何使用 2269
F.7.3. 字符串比较行为 2269
F.7.4. 限制 2270
F.7.5. 作者 2270
F.8. cube – 多维立方体数据类型 2272
F.8.1. 语法 2272
F.8.2. 精度 2272
F.8.3. 用法 2272
F.8.4. 默认规则 2274
F.8.5. 注意 2274
F.8.6. 致谢 2274
F.9. dblink 2276
F.10. dict_int – 用于整数的示例全文搜索词典 2307
F.10.1. 配置 2307
F.10.2. 用法 2307
F.11. dict_xsyn – 示例同义词全文检索词典 2308
F.11.1. 配置 2308
F.11.2. 用法 2308
F.12. dummy_seclabel 2310
F.12.1. 缘由 2310
F.12.2. 用法 2310
F.12.3. 作者 2310
F.13. earthdistance – 计算大圆距离 2311
F.13.1. 基于立方体的地球距离 2311
F.13.2. 基于点的地球距离 2312
F.14. file_fdw 2313
F.15. fuzzystrmatch – 确定字符串的相似性和距离 2314
F.15.1. Soundex 2314
F.15.2. Levenshtein 2314
F.15.3. Metaphone 2315
F.15.4. Double Metaphone 2316
F.16. hstore 2317
F.16.1. `hstore` 外部表示 2317
F.16.2. `hstore` 操作符和函数 2317
F.16.3. 索引 2320

F.16.4. 示例 2320
F.16.5. 统计信息 2321
F.16.6. 兼容性 2322
F.16.7. 作者 2322
F.17. intagg – 整数聚合器和枚举器 2323
F.17.1. 函数 2323
F.17.2. 使用示例 2323
F.18. intarray 2325
F.18.1. `intarray` 函数和操作符 2325
F.18.2. 索引支持 2326
F.18.3. 示例 2326
F.18.4. 基准测试 2327
F.18.5. 作者 2327
F.19. isn 2328
F.19.1. 数据类型 2328
F.19.2. 类型转换 2329
F.19.3. 函数和操作符 2329
F.19.4. 示例 2330
F.19.5. 参考文献 2330
F.19.6. 作者 2331
F.20. lo – 管理大对象 2332
F.20.1. 原理 2332
F.20.2. 如何使用 2332
F.20.3. 限制 2332
F.20.4. 作者 2333
F.21. ltree 2334
F.21.1. 定义 2334
F.21.2. 操作符和函数 2335
F.21.3. 索引 2337
F.21.4. 示例 2338
F.21.5. 作者 2340
F.22. oid2name 2341
F.22.1. Overview 2341
F.22.2. oid2name Options 2341
F.22.3. 示例 2342
F.22.4. Limitations 2344
F.22.5. 作者 2344
F.23. pageinspect 2345
F.23.1. 函数 2345
F.24. passwordcheck 2347
F.25. pg_archivecleanup 2348
F.25.1. Usage 2348
F.25.2. pg_archivecleanup 选项 2348
F.25.3. 示例 2348
F.25.4. Supported Server Versions 2349
F.25.5. 作者 2349
F.26. pgbench 2350
F.26.1. 概述 2350
F.26.2. pgbench 初始化选项 2351
F.26.3. pgbench Benchmarking Options 2351
F.26.4. pgbench Common Options 2352
F.26.5. 在 pgbench 中实际执行的“事务”是什么？ 2353
F.26.6. 自定义脚本 2353

F.26.7. 逐事务日志记录 2355
F.26.8. 逐语句延迟 2355
F.26.9. 良好实践 2356
F.27. pg_buffercache 2357
F.27.1. pg_buffercache 视图 2357
F.27.2. 示例输出 2357
F.27.3. 作者 2358
F.28. pgcrypto 2359
F.28.1. 通用哈希函数 2359
F.28.2. 密码哈希函数 2359
F.28.3. PGP 加密函数 2361
F.28.4. 原始加密函数 2366
F.28.5. 随机数据函数 2367
F.28.6. 注意事项 2367
F.28.7. 作者 2369
F.29. pg_freespacemap 2370
F.29.1. 函数 2370
F.29.2. 示例输出 2370
F.29.3. 作者 2371
F.30. pgrowlocks – 显示表的行锁信息 2372
F.30.1. 概述 2372
F.30.2. 示例输出 2372
F.30.3. 作者 2373
F.31. pg_standby 2374
F.31.1. Usage 2374
F.31.2. pg_standby Options 2375
F.31.3. 示例 2375
F.31.4. Supported Server Versions 2376
F.31.5. 作者 2376
F.32. pg_stat_statements 2377
F.32.1. pg_stat_statements 视图 2377
F.32.2. 函数 2378
F.32.3. 配置参数 2378
F.32.4. 示例输出 2379
F.32.5. Author 2380
F.33. pgstattuple – 获取元组级统计信息 2381
F.33.1. 函数 2381
F.33.2. 作者 2382
F.34. pg_test_fsync 2383
F.34.1. Usage 2383
F.34.2. 作者 2383
F.35. pg_trgm 2384
F.35.1. 三字符组（Trigram 或 Trigraph）概念 2384
F.35.2. 函数和操作符 2384
F.35.3. 索引支持 2385
F.35.4. 文本检索集成 2386
F.35.5. 参考 2386
F.35.6. 作者 2386
F.36. pg_upgrade 2387
F.36.1. 支持的版本 2387
F.36.2. pg_upgrade 选项 2387
F.36.3. 升级步骤 2388
F.36.4. 从 PostgreSQL 8.3 升级时的限制 2391

F.36.5. 注解 2391
F.37. seg – 用于线段或浮点区间的数据类型 2393
F.37.1. 原理 2393
F.37.2. 语法 2393
F.37.3. 精度 2394
F.37.4. 用法 2394
F.37.5. 注意 2395
F.37.6. 致谢 2395
F.38. sepgsql – 基于 SELinux 标签的强制访问控制（MAC）安全模块 2396
F.38.1. 概述 2396
F.38.2. 安装 2396
F.38.3. 回归测试 2397
F.38.4. GUC 参数 2398
F.38.5. 特性 2399
F.38.6. 限制 2401
F.38.7. 外部资源 2401
F.38.8. 作者 2401
F.39. spi 2402
F.39.1. refint – 用于实现引用完整性的函数 2402
F.39.2. timetravel – 实现时间旅行的函数 2402
F.39.3. autoinc – 用于字段自动递增的函数 2403
F.39.4. insert_username – 用于跟踪谁修改了表的函数 2403
F.39.5. moddatetime – 用于跟踪最后修改时间的函数 2403
F.40. sslinfo 2404
F.40.1. 提供的函数 2404
F.40.2. 作者 2405
F.41. tablefunc 2406
F.41.1. 提供的函数 2406
F.41.2. 作者 2415
F.42. test_parser 2416
F.42.1. 用法 2416
F.43. tsearch2 2418
F.43.1. 可移植性问题 2418
F.43.2. 转换 8.3 之前的安装 2418
F.43.3. 参考文献 2419
F.44. unaccent 2420
F.44.1. 配置 2420
F.44.2. 用法 2420
F.44.3. 函数 2421
F.45. uuid-ossp 2422
F.45.1. `uuid-ossp` 函数 2422
F.45.2. 作者 2423
F.46. vacuumlo 2424
F.46.1. 用法 2424
F.46.2. 方法 2425
F.46.3. 作者 2425
F.47. xml2 2426
F.47.1. 弃用说明 2426
F.47.2. 函数说明 2426
F.47.3. `xpath_table` 2427
F.47.4. XSLT 函数 2429
F.47.5. 作者 2429
G. 外部项目 2430

G.1. 客户端接口 2430
G.2. 管理工具 2430
G.3. 过程语言 2430
G.4. 扩展 2431
H. 源代码仓库 2432
H.1. 通过 Git 获取源代码 2432
I. 文档 2433
I.1. DocBook 2433
I.2. 工具集 2433
I.2.1. Linux RPM 安装 2434
I.2.2. FreeBSD 安装 2434
I.2.3. Debian 软件包 2435
I.2.4. 从源码手动安装 2435
I.2.5. 通过`configure`检测 2437
I.3. 构建文档 2437
I.3.1. HTML 2438
I.3.2. 手册页 2438
I.3.3. Print Output via JadeTeX 2438
I.3.4. 超宽文本 2439
I.3.5. 通过RTF生成打印输出 2439
I.3.6. 纯文本文件 2440
I.3.7. 语法检查 2440
I.4. 文档编写 2440
I.4.1. Emacs/PSGML 2441
I.4.2. 其他 Emacs 模式 2441
I.5. 风格指南 2442
I.5.1. 参考页 2442
J. 缩略词 2444

附录 A. PostgreSQL错误代码

PostgreSQL 服务器发出的所有消息都被分配了五字符错误代码，遵循 SQL 标准对“SQLSTATE”代码的约定。需要知道发生了哪一种错误条件的应用，通常应测试错误代码，而不是查看错误消息的文本内容。这些错误代码在不同 PostgreSQL 版本之间更不容易发生变化，也不会因错误消息的本地化而改变。请注意，PostgreSQL 产生的错误代码中，只有一部分由 SQL 标准定义；还有一些用于标准未定义情况的附加错误代码，是新设的或从其他数据库借用而来的。

根据标准，错误代码的前两个字符表示错误类别，后三个字符表示该类别中的特定条件。因此，即使应用无法识别某个具体错误代码，仍可能从错误类别推断出应采取的操作。

表 A.1列出了 PostgreSQL 9.1.24 中定义的所有错误代码。（其中有些当前实际上并未使用，不过它们由 SQL 标准定义。）其中也显示了错误类别。对于每个错误类别，都有一个后三个字符为 `000` 的“标准”错误代码。该代码只用于那些属于该类别但没有分配更具体代码的错误情况。

“Condition Name”列中显示的符号也是 PL/pgSQL 中使用的条件名。条件名可以写成大写或小写。（注意 PL/pgSQL 不识别警告（而非错误）类条件名，即类 00、01 和 02。）

表 A.1. PostgreSQL错误代码

Error Code	Condition Name
Class 00 — Successful Completion	
`00000`	`successful_completion`
Class 01 — Warning	
`01000`	`warning`
`0100C`	`dynamic_result_sets_returned`
`01008`	`implicit_zero_bit_padding`
`01003`	`null_value_eliminated_in_set_function`
`01007`	`privilege_not_granted`
`01006`	`privilege_not_revoked`
`01004`	`string_data_right_truncation`
`01P01`	`deprecated_feature`
Class 02 — No Data (this is also a warning class per the SQL standard)	
`02000`	`no_data`
`02001`	`no_additional_dynamic_result_sets_returned`
Class 03 — SQL Statement Not Yet Complete	
`03000`	`sql_statement_not_yet_complete`
Class 08 — Connection Exception	
`08000`	`connection_exception`
`08003`	`connection_does_not_exist`
`08006`	`connection_failure`
`08001`	`sqlclient_unable_to_establish_sqlconnection`
`08004`	`sqlserver_rejected_establishment_of_sqlconnection`

Error Code	Condition Name
08007	transaction_resolution_unknown
08P01	protocol_violation
Class 09 – Triggered Action Exception	
09000	triggered_action_exception
Class 0A – Feature Not Supported	
0A000	feature_not_supported
Class 0B – Invalid Transaction Initiation	
0B000	invalid_transaction_initiation
Class 0F – Locator Exception	
0F000	locator_exception
0F001	invalid_locator_specification
Class 0L – Invalid Grantor	
0L000	invalid_grantor
0LP01	invalid_grant_operation
Class 0P – Invalid Role Specification	
0P000	invalid_role_specification
Class 20 – Case Not Found	
20000	case_not_found
Class 21 – Cardinality Violation	
21000	cardinality_violation
Class 22 – Data Exception	
22000	data_exception
2202E	array_subscript_error
22021	character_not_in_repertoire
22008	datetime_field_overflow
22012	division_by_zero
22005	error_in_assignment
2200B	escape_character_conflict
22022	indicator_overflow
22015	interval_field_overflow
2201E	invalid_argument_for_logarithm
22014	invalid_argument_for_ntile_function
22016	invalid_argument_for_nth_value_function
2201F	invalid_argument_for_power_function
2201G	invalid_argument_for_width_bucket_function
22018	invalid_character_value_for_cast
22007	invalid_datetime_format
22019	invalid_escape_character
2200D	invalid_escape_octet

Error Code	Condition Name
22025	invalid_escape_sequence
22P06	nonstandard_use_of_escape_character
22010	invalid_indicator_parameter_value
22023	invalid_parameter_value
2201B	invalid_regular_expression
2201W	invalid_row_count_in_limit_clause
2201X	invalid_row_count_in_result_offset_clause
22009	invalid_time_zone_displacement_value
2200C	invalid_use_of_escape_character
2200G	most_specific_type_mismatch
22004	null_value_not_allowed
22002	null_value_no_indicator_parameter
22003	numeric_value_out_of_range
22026	string_data_length_mismatch
22001	string_data_right_truncation
22011	substring_error
22027	trim_error
22024	unterminated_c_string
2200F	zero_length_character_string
22P01	floating_point_exception
22P02	invalid_text_representation
22P03	invalid_binary_representation
22P04	bad_copy_file_format
22P05	untranslatable_character
2200L	not_an_xml_document
2200M	invalid_xml_document
2200N	invalid_xml_content
2200S	invalid_xml_comment
2200T	invalid_xml_processing_instruction
Class 23 – Integrity Constraint Violation	
23000	integrity_constraint_violation
23001	restrict_violation
23502	not_null_violation
23503	foreign_key_violation
23505	unique_violation
23514	check_violation
23P01	exclusion_violation
Class 24 – Invalid Cursor State	
24000	invalid_cursor_state

Error Code	Condition Name
Class 25 – Invalid Transaction State	
25000	invalid_transaction_state
25001	active_sql_transaction
25002	branch_transaction_already_active
25008	held_cursor_requires_same_isolation_level
25003	inappropriate_access_mode_for_branch_transaction
25004	inappropriate_isolation_level_for_branch_transaction
25005	no_active_sql_transaction_for_branch_transaction
25006	read_only_sql_transaction
25007	schema_and_data_statement_mixing_not_supported
25P01	no_active_sql_transaction
25P02	in_failed_sql_transaction
Class 26 – Invalid SQL Statement Name	
26000	invalid_sql_statement_name
Class 27 – Triggered Data Change Violation	
27000	triggered_data_change_violation
Class 28 – Invalid Authorization Specification	
28000	invalid_authorization_specification
28P01	invalid_password
Class 2B – Dependent Privilege Descriptors Still Exist	
2B000	dependent_privilege_descriptors_still_exist
2BP01	dependent_objects_still_exist
Class 2D – Invalid Transaction Termination	
2D000	invalid_transaction_termination
Class 2F – SQL Routine Exception	
2F000	sql_routine_exception
2F005	function_executed_no_return_statement
2F002	modifying_sql_data_not_permitted
2F003	prohibited_sql_statement_attempted
2F004	reading_sql_data_not_permitted
Class 34 – Invalid Cursor Name	
34000	invalid_cursor_name
Class 38 – External Routine Exception	
38000	external_routine_exception
38001	containing_sql_not_permitted
38002	modifying_sql_data_not_permitted

Error Code	Condition Name
38003	prohibited_sql_statement_attempted
38004	reading_sql_data_not_permitted
Class 39 – External Routine Invocation Exception	
39000	external_routine_invocation_exception
39001	invalid_sqlstate_returned
39004	null_value_not_allowed
39P01	trigger_protocol_violated
39P02	srf_protocol_violated
Class 3B – Savepoint Exception	
3B000	savepoint_exception
3B001	invalid_savepoint_specification
Class 3D – Invalid Catalog Name	
3D000	invalid_catalog_name
Class 3F – Invalid Schema Name	
3F000	invalid_schema_name
Class 40 – Transaction Rollback	
40000	transaction_rollback
40002	transaction_integrity_constraint_violation
40001	serialization_failure
40003	statement_completion_unknown
40P01	deadlock_detected
Class 42 – Syntax Error or Access Rule Violation	
42000	syntax_error_or_access_rule_violation
42601	syntax_error
42501	insufficient_privilege
42846	cannot_coerce
42803	grouping_error
42P20	windowing_error
42P19	invalid_recursion
42830	invalid_foreign_key
42602	invalid_name
42622	name_too_long
42939	reserved_name
42804	datatype_mismatch
42P18	indeterminate_datatype
42P21	collation_mismatch
42P22	indeterminate_collation
42809	wrong_object_type
42703	undefined_column

Error Code	Condition Name
42883	undefined_function
42P01	undefined_table
42P02	undefined_parameter
42704	undefined_object
42701	duplicate_column
42P03	duplicate_cursor
42P04	duplicate_database
42723	duplicate_function
42P05	duplicate_prepared_statement
42P06	duplicate_schema
42P07	duplicate_table
42712	duplicate_alias
42710	duplicate_object
42702	ambiguous_column
42725	ambiguous_function
42P08	ambiguous_parameter
42P09	ambiguous_alias
42P10	invalid_column_reference
42611	invalid_column_definition
42P11	invalid_cursor_definition
42P12	invalid_database_definition
42P13	invalid_function_definition
42P14	invalid_prepared_statement_definition
42P15	invalid_schema_definition
42P16	invalid_table_definition
42P17	invalid_object_definition
Class 44 – WITH CHECK OPTION Violation	
44000	with_check_option_violation
Class 53 – Insufficient Resources	
53000	insufficient_resources
53100	disk_full
53200	out_of_memory
53300	too_many_connections
Class 54 – Program Limit Exceeded	
54000	program_limit_exceeded
54001	statement_too_complex
54011	too_many_columns
54023	too_many_arguments
Class 55 – Object Not In Prerequisite State	
55000	object_not_in_prerequisite_state

Error Code	Condition Name
55006	object_in_use
55P02	cant_change_runtime_param
55P03	lock_not_available
Class 57 – Operator Intervention	
57000	operator_intervention
57014	query_canceled
57P01	admin_shutdown
57P02	crash_shutdown
57P03	cannot_connect_now
57P04	database_dropped
Class 58 – System Error (errors external to PostgreSQL itself)	
58030	io_error
58P01	undefined_file
58P02	duplicate_file
Class F0 – Configuration File Error	
F0000	config_file_error
F0001	lock_file_exists
Class HV – Foreign Data Wrapper Error (SQL/MED)	
HV000	fdw_error
HV005	fdw_column_name_not_found
HV002	fdw_dynamic_parameter_value_needed
HV010	fdw_function_sequence_error
HV021	fdw_inconsistent_descriptor_inform ation
HV024	fdw_invalid_attribute_value
HV007	fdw_invalid_column_name
HV008	fdw_invalid_column_number
HV004	fdw_invalid_data_type
HV006	fdw_invalid_data_type_descriptors
HV091	fdw_invalid_descriptor_field_identi fier
HV00B	fdw_invalid_handle
HV00C	fdw_invalid_option_index
HV00D	fdw_invalid_option_name
HV090	fdw_invalid_string_length_or_buffer_ length
HV00A	fdw_invalid_string_format
HV009	fdw_invalid_use_of_null_pointer
HV014	fdw_too_many_handles
HV001	fdw_out_of_memory
HV00P	fdw_no_schemas

Error Code	Condition Name
HV00J	fdw_option_name_not_found
HV00K	fdw_reply_handle
HV00Q	fdw_schema_not_found
HV00R	fdw_table_not_found
HV00L	fdw_unable_to_create_execution
HV00M	fdw_unable_to_create_reply
HV00N	fdw_unable_to_establish_connection
Class P0 – PL/pgSQL Error	
P0000	plpgsql_error
P0001	raise_exception
P0002	no_data_found
P0003	too_many_rows
Class XX – Internal Error	
XX000	internal_error
XX001	data_corrupted
XX002	index_corrupted

附录 B. Date/Time Support

PostgreSQL 对所有日期/时间输入都使用一个内部的启发式解析器。日期和时间以字符串形式输入，并被拆分为不同的字段，同时初步判断每个字段中可能包含哪类信息。
每个字段都会被解释，并被赋予数值、忽略或拒绝。
解析器内部为所有文本字段都维护了查找表，包括月份、星期几以及时区。

本附录包含关于这些查找表内容的信息，并描述了解析器解码日期和时间时所使用的步骤。

B.1. Date/Time Input Interpretation

日期/时间输入字符串按照下面的过程进行解码。

1. 将输入字符串拆分为词元，并把每个词元归类为字符串、时间、时区或数字。

   a. 如果数字词元包含冒号（`:`），则它是一个时间字符串。
      把其后的所有数字和冒号都包含进来。

   b. 如果数字词元包含连字符（`-`）、斜杠（`/`）或两个以上的点（`.`），
      则它是一个日期字符串，并且可能带有文本形式的月份。如果已经见过日期词元，
      则它会被解释为时区名称（例如 `America/New_York`）。

   c. 如果词元仅由数字组成，那么它要么是单个字段，要么是 ISO 8601 的拼接式日期（例如，`19990113` 表示 1999 年 1 月 13 日）或时间（例如，`141516` 表示 14:15:16）。

   d. 如果词元以加号（`+`）或减号（`-`）开头，那么它要么是数字时区，要么是特殊字段。

2. 如果词元是一个纯字母字符串，则将其与可能的字符串匹配：

   a. 先看该词元是否匹配任何已知的时区缩写。

   b. 如果没有找到，则搜索内部表，将该词元识别为特殊字符串（例如 `today`）、星期名（例如 `Thursday`）、月份名（例如 `January`）或噪声词（例如 `at`、`on`）。

   c. 如果仍然没有找到，则抛出错误。

3. 当词元是一个数字或数字字段时：

   a. 如果有八位或六位数字，并且之前还没有读取到其他日期字段，
      则将其解释为“拼接式日期”（例如 `19990118` 或 `990118`）。其解释方式分别为 `YYYYMMDD` 或 `YYMMDD`。

   b. 如果词元是三位数字，并且已经读取到了年份，则将其解释为一年中的第几天。

   c. 如果是四位或六位数字，并且已经读取到了年份，则将其解释为时间（`HHMM` 或 `HHMMSS`）。

   d. 如果有三位或更多位数字，并且尚未找到任何日期字段，
      则将其解释为年份（这会强制其余日期字段按 yy-mm-dd 的顺序解释）。

   e. 否则，假定日期字段顺序遵循 `DateStyle` 设置：mm-dd-yy、dd-mm-yy 或 yy-mm-dd。如果发现月份字段或日字段超出范围，则抛出错误。

4. 如果指定了 BC，则将年份取反并加一后再用于内部存储。（格里高利历中没有 0 年，因此从数值上说，公元前 1 年会变成年 0。）

5. 如果没有指定 BC，并且年份字段的长度为两位，则将年份调整为四位。如果该字段小于 70，则加上 2000，否则加上 1900。

> **提示**
>
> 格里高利历公元 1-99 年可以使用带前导零的四位数字输入（例如，`0099` 表示公元 99 年）。

B.2. Date/Time Key Words

表 B.1展示了哪些词元会被识别为月份名称。

表 B.1. Month Names

Month	Abbreviations
January	Jan
February	Feb
March	Mar
April	Apr
May	
June	Jun
July	Jul
August	Aug
September	Sep, Sept
October	Oct
November	Nov
December	Dec

表 B.2展示了哪些词元会被识别为星期名称。

表 B.2. 星期名称

Day	Abbreviations
Sunday	Sun
Monday	Mon
Tuesday	Tue, Tues
Wednesday	Wed, Weds
Thursday	Thu, Thur, Thurs
Friday	Fri
Saturday	Sat

表 B.3展示了用于各种修饰用途的词元。

表 B.3. Date/Time Field Modifiers

Identifier	Description
`AM`	Time is before 12:00

Identifier	Description
AT	Ignored
JULIAN, JD, J	Next field is Julian Day
ON	Ignored
PM	Time is on or after 12:00
T	Next field is time

B.3. Date/Time Configuration Files

由于时区缩写没有得到良好的标准化，PostgreSQL 提供了一种定制服务器所接受缩写集合的方法。运行时参数timezone_abbreviations决定当前使用的缩写集合。
任何数据库用户都可以修改此参数，但它的可选值由数据库管理员控制；
这些值实际上是安装目录中 `.../share/timezonesets/` 下的配置文件名。管理员可以通过在该目录中添加或修改文件，制定本地时区缩写策略。

如果文件名完全由字母组成，那么 `timezone_abbreviations` 可以被设置为 `.../share/timezonesets/` 中找到的任意文件名。（禁止在 `timezone_abbreviations` 中使用非字母字符，不仅可以防止读取目标目录之外的文件，也能防止读取编辑器备份文件和其他无关文件。）

时区缩写文件可以包含空行，以及以下字符开头的注释：`#`。非注释行必须采用以下格式之一：

```
zone_abbreviation offset
zone_abbreviation offset D
zone_abbreviation time_zone_name
@INCLUDE file_name
@OVERRIDE
```

zone_abbreviation 就是要定义的缩写。*offset* 是一个整数，表示相对于 UTC 的等效偏移秒数，正数表示格林尼治以东，负数表示以西。例如，-18000 表示格林尼治以西五小时，即北美东海岸标准时间。`D` 表示该时区名称代表当地夏令时间，而非标准时间。

也可以给出 *time_zone_name*，引用 IANA 时区数据库中定义的时区名。系统会查阅该时区的定义，判断其中是否正在或曾经使用该缩写；如果是，
则采用合适的含义：即正在确定其值的时间戳所处时刻使用的含义；
如果该时刻并未使用此缩写，则取此前最近使用的含义；如果只在该时刻之后使用过，
则取最早的含义。这对于处理含义曾经发生变化的缩写至关重要。
也允许用一个并未出现该缩写的时区名来定义缩写；此时，使用缩写就等价于写出该时区名。

> **提示**
>
> 如果某个缩写相对 UTC 的偏移从未改变过，那么定义它时最好使用简单的整数 *offset*，因为这类缩写的处理成本远低于那些需要查阅时区定义的缩写。

`@INCLUDE` 语法允许包含 `.../share/timezonesets/` 目录中的另一个文件。包含可以嵌套，但深度有限。

`@OVERRIDE` 语法表示文件中后续的条目可以覆盖之前的条目（通常是从被包含文件中取得的条目）。如果没有此指令，同一时区缩写的冲突定义会被视为错误。

在未经修改的安装中，`Default` 文件包含世界上大多数地区中所有互不冲突的时区缩写。另外还为澳大利亚和印度这两个地区提供了 `Australia` 和 `India` 文件：这些文件会先包含 `Default` 文件，然后按需添加或修改缩写。

作为参考，标准安装还包含 `Africa.txt`、`America.txt` 等文件，其中包含 IANA 时区数据库中已知正在使用的每个时区缩写的信息。
可以按需把这些文件中找到的时区名称定义复制并粘贴到自定义配置文件中。注意，
这些文件不能直接作为 `timezone_abbreviations` 设置来引用，因为它们的名称中带有点号。

注意

如果在读取时区缩写集时发生错误，则不会应用新值，而会保留旧集合。
如果该错误发生在数据库启动期间，则数据库启动会失败。

小心

配置文件中定义的时区缩写会覆盖 PostgreSQL 内置的非时区含义。例如，`Australia` 配置文件定义了 `SAT`（南澳大利亚标准时间）。当该文件处于活动状态时，`SAT` 将不会被识别为 Saturday 的缩写。

小心

如果你修改了 `.../share/timezonesets/` 中的文件，就需要自行做好备份 – 普通的数据库转储不会包含这个目录。

B.4. History of Units

儒略历由 Julius Caesar 于公元前 45 年引入。在 1582 年各国开始改用格里高利历之前，它在西方世界一直被广泛使用。在儒略历中，回归年近似为 365 1/4 天 = 365.25 天。这会在大约 128 年里累计 1 天的误差。

不断累积的历法误差促使教皇格里高利十三世按照特伦托会议的指示改革历法。
在格里高利历中，回归年近似为 365 + 97 / 400 天 = 365.2425 天。因此，大约要经过 3300 年，回归年与格里高利历之间才会累积出 1 天的偏差。

365+97/400 这一近似值是通过每 400 年设置 97 个闰年来实现的，具体规则如下：

每个能被 4 整除的年份都是闰年。
不过，每个能被 100 整除的年份都不是闰年。
但是，每个能被 400 整除的年份仍然是闰年。

因此，1700、1800、1900、2100 和 2200 都不是闰年，但 1600、2000 和 2400 是闰年。相比之下，较早的儒略历规定所有能被 4 整除的年份都是闰年。

1582 年 2 月发布的教皇诏书规定，应从 1582 年 10 月删去 10 天，使 10 月 15 日紧接在 10 月 4 日之后。这一改革在意大利、波兰、葡萄牙和西班牙得到执行。其他天主教国家不久后也跟进，但新教国家不愿改变，而希腊正教国家直到 20 世纪初才改用格里高利历。大不列颠及其属

地（包括今天的美国）在 1752 年实施了这项改革。因此，1752 年 9 月 2 日之后紧接着就是 1752 年 9 月 14 日。这就是为什么提供 `cal` 程序的 Unix 系统会产生如下输出：

```
$ cal 9 1752
   September 1752
 S  M Tu  W Th  F  S
       1  2 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
```

SQL 标准指出，“在‘日期时间字面量’的定义中，‘日期时间值’受格里高利历中日期和时间自然规则的约束”。1582-10-05 到 1582-10-14 之间的日期，虽然在某些国家因教皇敕令而被删除，但它们符合“自然规则”，因此是有效日期。PostgreSQL 遵循 SQL 标准的做法，完全按照格里高利历计算日期，即便对于该历法尚未使用之前的年份也是如此。

世界上不同地区发展出了不同的历法，其中许多都早于格里高利体系。例如，
中国历法的起源可以追溯到公元前 14 世纪。传说黄帝在公元前 2637 年发明了这种历法。中华人民共和国在民用方面使用格里高利历。中国历法则用于确定节日。

“儒略日”与“儒略历”并无关系。儒略日系统由法国学者 Joseph Justus Scaliger（1540-1609）发明，其名称大概取自他的父亲、意大利学者 Julius Caesar Scaliger（1484-1558）。在儒略日系统中，每一天都有一个连续编号，从 JD 0 开始（有时称为起始儒略日）。JD 0 对应儒略历公元前 4713 年 1 月 1 日，或格里高利历公元前 4714 年 11 月 24 日。儒略日计数最常用于天文学家标记夜间观测，因此一天从 UTC 正午开始，到下一个 UTC 正午结束，而不是从午夜到午夜：JD 0 表示从 UTC 公元前 4713 年 1 月 1 日正午到 UTC 公元前 4713 年 1 月 2 日正午的 24 小时。

虽然 PostgreSQL 支持在日期输入和输出中使用儒略日记法（并且也会在某些内部日期/时间计算中使用儒略日），但它并不遵循日期从正午算到正午这一约定。PostgreSQL 将儒略日视为从午夜到午夜。

附录 C. SQL 关键字

表 C.1列出了在 SQL 标准以及 PostgreSQL 9.1.24 中被视为关键字的所有词元。背景信息可参见第 4.1.1 节。

SQL 区分保留关键字和非保留关键字。根据标准，保留关键字才是真正的关键字，它们绝不允许作为标识符使用。非保留关键字只在特定上下文中具有特殊含义，
在其他上下文中则可以作为标识符使用。大多数非保留关键字实际上是 SQL 规定的内置表和内置函数名称。非保留关键字这一概念，
本质上只是为了声明某个词在某些上下文中附带了预定义含义。

在PostgreSQL解析器中，情况要更复杂一些。这里的词元分属若干不同类别，
从绝不能用作标识符的词元，到在解析器中完全没有特殊地位、
只是被当作普通标识符处理的词元都有。（后一类通常就是 SQL 规定的函数名。）即使是保留关键字，在PostgreSQL中也不是绝对保留的，它们仍然可以用作列标签（例如`SELECT 55 AS CHECK`，尽管`CHECK`是一个保留关键字）。

在表 C.1中针对PostgreSQL的那一列里，我们把解析器明确识别、但又允许作为列名或表名的那些关键字归类为“非保留”。某些原本属于非保留的关键字不能用作函数名或数据类型名，
因此会另外加以标记。（这类词大多表示具有特殊语法的内置函数或数据类型。
函数或类型本身仍然可用，但用户不能重新定义它们。）
不允许作为列名或表名的词元则标为“保留”。有些保留关键字仍允许作为函数名或数据类型名，这一点也会在表中标出。如果没有这样的标记，保留关键字就只允许作为“AS”列标签名使用。

一般来说，如果你在某条命令中把这里列出的关键字当作标识符使用，
并因此遇到意外的解析器错误，应尝试为该标识符加上引号，看看问题是否消失。

在研读表 C.1之前，有一点非常重要：某个关键字在PostgreSQL中不被保留，并不意味着与该词相关的特性尚未实现；反过来，某个关键字的存在也不意味着相应特性一定存在。

表 C.1. SQL 关键字

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
A		non-reserved	non-reserved		
ABORT	non-reserved				
ABS		reserved	reserved	non-reserved	
ABSENT		non-reserved			
ABSOLUTE	non-reserved	non-reserved	non-reserved	reserved	reserved
ACCESS	non-reserved				
ACCORDING		non-reserved			
ACTION	non-reserved	non-reserved	non-reserved	reserved	reserved
ADA		non-reserved	non-reserved	non-reserved	non-reserved
ADD	non-reserved	non-reserved	non-reserved	reserved	reserved
ADMIN	non-reserved	non-reserved	non-reserved	reserved	
AFTER	non-reserved	non-reserved	non-reserved	reserved	
AGGREGATE	non-reserved			reserved	
ALIAS				reserved	
ALL	reserved	reserved	reserved	reserved	reserved
ALLOCATE		reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
ALSO	non-reserved				
ALTER	non-reserved	reserved	reserved	reserved	reserved
ALWAYS	non-reserved	non-reserved	non-reserved		
ANALYSE	reserved				
ANALYZE	reserved				
AND	reserved	reserved	reserved	reserved	reserved
ANY	reserved	reserved	reserved	reserved	reserved
ARE		reserved	reserved	reserved	reserved
ARRAY	reserved	reserved	reserved	reserved	
ARRAY_AGG		reserved			
AS	reserved	reserved	reserved	reserved	reserved
ASC	reserved	non-reserved	non-reserved	reserved	reserved
ASENSITIVE		reserved	reserved	non-reserved	
ASSERTION	non-reserved	non-reserved	non-reserved	reserved	reserved
ASSIGNMENT	non-reserved	non-reserved	non-reserved	non-reserved	
ASYMMETRIC	reserved	reserved	reserved	non-reserved	
AT	non-reserved	reserved	reserved	reserved	reserved
ATOMIC		reserved	reserved	non-reserved	
ATTRIBUTE	non-reserved	non-reserved	non-reserved		
ATTRIBUTES		non-reserved	non-reserved		
AUTHORIZATION	reserved (can be function or type)	reserved	reserved	reserved	reserved
AVG		reserved	reserved	non-reserved	reserved
BACKWARD	non-reserved				
BASE64		non-reserved	non-reserved		
BEFORE	non-reserved	non-reserved	non-reserved	reserved	
BEGIN	non-reserved	reserved	reserved	reserved	reserved
BERNOULLI		non-reserved	non-reserved		
BETWEEN	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
BIGINT	non-reserved (cannot be function or type)	reserved	reserved		
BINARY	reserved (can be function or type)	reserved	reserved	reserved	
BIT	non-reserved (cannot be function or type)			reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
BITVAR				non-reserved	
BIT_LENGTH				non-reserved	reserved
BLOB		reserved	reserved	reserved	
BLOCKED		non-reserved	non-reserved		
BOM		non-reserved			
BOOLEAN	non-reserved (cannot be function or type)	reserved	reserved	reserved	
BOTH	reserved	reserved	reserved	reserved	reserved
BREADTH		non-reserved	non-reserved	reserved	
BY	non-reserved	reserved	reserved	reserved	reserved
C		non-reserved	non-reserved	non-reserved	non-reserved
CACHE	non-reserved				
CALL		reserved	reserved	reserved	
CALLED	non-reserved	reserved	reserved	non-reserved	
CARDINALITY		reserved	reserved	non-reserved	
CASCADE	non-reserved	non-reserved	non-reserved	reserved	reserved
CASCADED	non-reserved	reserved	reserved	reserved	reserved
CASE	reserved	reserved	reserved	reserved	reserved
CAST	reserved	reserved	reserved	reserved	reserved
CATALOG	non-reserved	non-reserved	non-reserved	reserved	reserved
CATALOG_NAME		non-reserved	non-reserved	non-reserved	non-reserved
CEIL		reserved	reserved		
CEILING		reserved	reserved		
CHAIN	non-reserved	non-reserved	non-reserved	non-reserved	
CHAR	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
CHARACTER	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
CHARACTERISTICS	non-reserved	non-reserved	non-reserved		
CHARACTERS		non-reserved	non-reserved		
CHARACTER_LENGTH		reserved	reserved	non-reserved	reserved
CHARACTER_SET_CATALOG		non-reserved	non-reserved	non-reserved	non-reserved
CHARACTER_SET_NAME		non-reserved	non-reserved	non-reserved	non-reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
CHARACTER_SET_SCHEMA		non-reserved	non-reserved	non-reserved	non-reserved
CHAR_LENGTH		reserved	reserved	non-reserved	reserved
CHECK	reserved	reserved	reserved	reserved	reserved
CHECKED				non-reserved	
CHECKPOINT	non-reserved				
CLASS	non-reserved			reserved	
CLASS_ORIGIN		non-reserved	non-reserved	non-reserved	non-reserved
CLOB		reserved	reserved	reserved	
CLOSE	non-reserved	reserved	reserved	reserved	reserved
CLUSTER	non-reserved				
COALESCE	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
COBOL		non-reserved	non-reserved	non-reserved	non-reserved
COLLATE	reserved	reserved	reserved	reserved	reserved
COLLATION	non-reserved	non-reserved	non-reserved	reserved	reserved
COLLATION_CATALOG		non-reserved	non-reserved	non-reserved	non-reserved
COLLATION_NAME		non-reserved	non-reserved	non-reserved	non-reserved
COLLATION_SCHEMA		non-reserved	non-reserved	non-reserved	non-reserved
COLLECT		reserved	reserved		
COLUMN	reserved	reserved	reserved	reserved	reserved
COLUMNS		non-reserved			
COLUMN_NAME		non-reserved	non-reserved	non-reserved	non-reserved
COMMAND_FUNCTION		non-reserved	non-reserved	non-reserved	non-reserved
COMMAND_FUNCTION_CODE		non-reserved	non-reserved	non-reserved	
COMMENT	non-reserved				
COMMENTS	non-reserved				
COMMIT	non-reserved	reserved	reserved	reserved	reserved
COMMITTED	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
COMPLETION				reserved	
CONCURRENTLY	reserved (can be function or type)				
CONDITION		reserved	reserved		
CONDITION_NUMBER		non-reserved	non-reserved	non-reserved	non-reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
CONFIGURATION	non-reserved				
CONNECT		reserved	reserved	reserved	reserved
CONNECTION	non-reserved	non-reserved	non-reserved	reserved	reserved
CONNECTION_NAME		non-reserved	non-reserved	non-reserved	non-reserved
CONSTRAINT	reserved	reserved	reserved	reserved	reserved
CONSTRAINTS	non-reserved	non-reserved	non-reserved	reserved	reserved
CONSTRAINT_CATALOG		non-reserved	non-reserved	non-reserved	non-reserved
CONSTRAINT_NAME		non-reserved	non-reserved	non-reserved	non-reserved
CONSTRAINT_SCHEMA		non-reserved	non-reserved	non-reserved	non-reserved
CONSTRUCTOR		non-reserved	non-reserved	reserved	
CONTAINS		non-reserved	non-reserved	non-reserved	
CONTENT	non-reserved	non-reserved	non-reserved		
CONTINUE	non-reserved	non-reserved	non-reserved	reserved	reserved
CONTROL		non-reserved	non-reserved		
CONVERSION	non-reserved				
CONVERT		reserved	reserved	non-reserved	reserved
COPY	non-reserved				
CORR		reserved	reserved		
CORRESPONDING		reserved	reserved	reserved	reserved
COST	non-reserved				
COUNT		reserved	reserved	non-reserved	reserved
COVAR_POP		reserved	reserved		
COVAR_SAMP		reserved	reserved		
CREATE	reserved	reserved	reserved	reserved	reserved
CROSS	reserved (can be function or type)	reserved	reserved	reserved	reserved
CSV	non-reserved				
CUBE		reserved	reserved	reserved	
CUME_DIST		reserved	reserved		
CURRENT	non-reserved	reserved	reserved	reserved	reserved
CURRENT_CATALOG	reserved	reserved			
CURRENT_DATE	reserved	reserved	reserved	reserved	reserved
CURRENT_DEFAULT_		reserved	reserved		

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
TRANSFORM_GROUP					
CURRENT_PATH		reserved	reserved	reserved	
CURRENT_ROLE	reserved	reserved	reserved	reserved	
CURRENT_SCHEMA	reserved (can be function or type)	reserved			
CURRENT_TIME	reserved	reserved	reserved	reserved	reserved
CURRENT_TIMESTAMP	reserved	reserved	reserved	reserved	reserved
CURRENT_TRANSFORM_GROUP_FOR_TYPE		reserved	reserved		
CURRENT_USER	reserved	reserved	reserved	reserved	reserved
CURSOR	non-reserved	reserved	reserved	reserved	reserved
CURSOR_NAME		non-reserved	non-reserved	non-reserved	non-reserved
CYCLE	non-reserved	reserved	reserved	reserved	
DATA	non-reserved	non-reserved	non-reserved	reserved	non-reserved
DATABASE	non-reserved				
DATALINK		reserved	reserved		
DATE		reserved	reserved	reserved	reserved
DATETIME_INTERVAL_CODE		non-reserved	non-reserved	non-reserved	non-reserved
DATETIME_INTERVAL_PRECISION		non-reserved	non-reserved	non-reserved	non-reserved
DAY	non-reserved	reserved	reserved	reserved	reserved
DB		non-reserved	non-reserved		
DEALLOCATE	non-reserved	reserved	reserved	reserved	reserved
DEC	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
DECIMAL	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
DECLARE	non-reserved	reserved	reserved	reserved	reserved
DEFAULT	reserved	reserved	reserved	reserved	reserved
DEFAULTS	non-reserved	non-reserved	non-reserved		
DEFERRABLE	reserved	non-reserved	non-reserved	reserved	reserved
DEFERRED	non-reserved	non-reserved	non-reserved	reserved	reserved
DEFINED		non-reserved	non-reserved	non-reserved	

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
DEFINER	non-reserved	non-reserved	non-reserved	non-reserved	
DEGREE		non-reserved	non-reserved		
DELETE	non-reserved	reserved	reserved	reserved	reserved
DELIMITER	non-reserved				
DELIMITERS	non-reserved				
DENSE_RANK		reserved	reserved		
DEPTH		non-reserved	non-reserved	reserved	
DEREF		reserved	reserved	reserved	
DERIVED		non-reserved	non-reserved		
DESC	reserved	non-reserved	non-reserved	reserved	reserved
DESCRIBE		reserved	reserved	reserved	reserved
DESCRIPTOR		non-reserved	non-reserved	reserved	reserved
DESTROY				reserved	
DESTRUCTOR				reserved	
DETERMINISTIC		reserved	reserved	reserved	
DIAGNOSTICS		non-reserved	non-reserved	reserved	reserved
DICTIONARY	non-reserved			reserved	
DISABLE	non-reserved				
DISCARD	non-reserved				
DISCONNECT		reserved	reserved	reserved	reserved
DISPATCH		non-reserved	non-reserved	non-reserved	
DISTINCT	reserved	reserved	reserved	reserved	reserved
DLNEWCOPY		reserved	reserved		
DLPREVIOUSCOPY		reserved	reserved		
DLURLCOMPLETE		reserved	reserved		
DLURLCOMPLETEONLY		reserved	reserved		
DLURLCOMPLETEWRITE		reserved	reserved		
DLURLPATH		reserved	reserved		
DLURLPATHONLY		reserved	reserved		
DLURLPATHWRITE		reserved	reserved		
DLURLSCHEME		reserved	reserved		
DLURLSERVER		reserved	reserved		
DLVALUE		reserved	reserved		
DO	reserved				
DOCUMENT	non-reserved	non-reserved	non-reserved		

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
DOMAIN	non-reserved	non-reserved	non-reserved	reserved	reserved
DOUBLE	non-reserved	reserved	reserved	reserved	reserved
DROP	non-reserved	reserved	reserved	reserved	reserved
DYNAMIC		reserved	reserved	reserved	
DYNAMIC_FUNCTION		non-reserved	non-reserved	non-reserved	non-reserved
DYNAMIC_FUNCTION_CODE		non-reserved	non-reserved	non-reserved	
EACH	non-reserved	reserved	reserved	reserved	
ELEMENT		reserved	reserved		
ELSE	reserved	reserved	reserved	reserved	reserved
EMPTY		non-reserved			
ENABLE	non-reserved				
ENCODING	non-reserved	non-reserved			
ENCRYPTED	non-reserved				
END	reserved	reserved	reserved	reserved	reserved
END-EXEC		reserved	reserved	reserved	reserved
ENUM	non-reserved				
EQUALS		non-reserved	non-reserved	reserved	
ESCAPE	non-reserved	reserved	reserved	reserved	reserved
EVERY		reserved	reserved	reserved	
EXCEPT	reserved	reserved	reserved	reserved	reserved
EXCEPTION			non-reserved	reserved	reserved
EXCLUDE	non-reserved	non-reserved	non-reserved		
EXCLUDING	non-reserved	non-reserved	non-reserved		
EXCLUSIVE	non-reserved				
EXEC		reserved	reserved	reserved	reserved
EXECUTE	non-reserved	reserved	reserved	reserved	reserved
EXISTING				non-reserved	
EXISTS	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
EXP		reserved	reserved		
EXPLAIN	non-reserved				
EXTENSION	non-reserved				
EXTERNAL	non-reserved	reserved	reserved	reserved	reserved
EXTRACT	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
FALSE	reserved	reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
FAMILY	non-reserved				
FETCH	reserved	reserved	reserved	reserved	reserved
FILE		non-reserved	non-reserved		
FILTER		reserved	reserved		
FINAL		non-reserved	non-reserved	non-reserved	
FIRST	non-reserved	non-reserved	non-reserved	reserved	reserved
FIRST_VALUE		reserved			
FLAG		non-reserved			
FLOAT	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
FLOOR		reserved	reserved		
FOLLOWING	non-reserved	non-reserved	non-reserved		
FOR	reserved	reserved	reserved	reserved	reserved
FORCE	non-reserved				
FOREIGN	reserved	reserved	reserved	reserved	reserved
FORTRAN		non-reserved	non-reserved	non-reserved	non-reserved
FORWARD	non-reserved				
FOUND		non-reserved	non-reserved	reserved	reserved
FREE		reserved	reserved	reserved	
FREEZE	reserved (can be function or type)				
FROM	reserved	reserved	reserved	reserved	reserved
FS		non-reserved	non-reserved		
FULL	reserved (can be function or type)	reserved	reserved	reserved	reserved
FUNCTION	non-reserved	reserved	reserved	reserved	
FUNCTIONS	non-reserved				
FUSION		reserved	reserved		
G		non-reserved	non-reserved	non-reserved	
GENERAL		non-reserved	non-reserved	reserved	
GENERATED		non-reserved	non-reserved	non-reserved	
GET		reserved	reserved	reserved	reserved
GLOBAL	non-reserved	reserved	reserved	reserved	reserved
GO		non-reserved	non-reserved	reserved	reserved
GOTO		non-reserved	non-reserved	reserved	reserved
GRANT	reserved	reserved	reserved	reserved	reserved
GRANTED	non-reserved	non-reserved	non-reserved	non-reserved	
GREATEST	non-reserved (cannot be				

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
	function or type)				
GROUP	reserved	reserved	reserved	reserved	reserved
GROUPING		reserved	reserved	reserved	
HANDLER	non-reserved				
HAVING	reserved	reserved	reserved	reserved	reserved
HEADER	non-reserved				
HEX		non-reserved	non-reserved		
HIERARCHY		non-reserved	non-reserved	non-reserved	
HOLD	non-reserved	reserved	reserved	non-reserved	
HOST				reserved	
HOUR	non-reserved	reserved	reserved	reserved	reserved
ID		non-reserved			
IDENTITY	non-reserved	reserved	reserved	reserved	reserved
IF	non-reserved				
IGNORE		non-reserved		reserved	
ILIKE	reserved (can be function or type)				
IMMEDIATE	non-reserved	non-reserved	non-reserved	reserved	reserved
IMMUTABLE	non-reserved				
IMPLEMENTATION		non-reserved	non-reserved	non-reserved	
IMPLICIT	non-reserved				
IMPORT		reserved	reserved		
IN	reserved	reserved	reserved	reserved	reserved
INCLUDING	non-reserved	non-reserved	non-reserved		
INCREMENT	non-reserved	non-reserved	non-reserved		
INDENT		non-reserved			
INDEX	non-reserved				
INDEXES	non-reserved				
INDICATOR		reserved	reserved	reserved	reserved
INFIX				non-reserved	
INHERIT	non-reserved				
INHERITS	non-reserved				
INITIALIZE				reserved	
INITIALLY	reserved	non-reserved	non-reserved	reserved	reserved
INLINE	non-reserved				
INNER	reserved (can be function or type)	reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
INOUT	non-reserved (cannot be function or type)	reserved	reserved	reserved	
INPUT	non-reserved	non-reserved	non-reserved	reserved	reserved
INSENSITIVE	non-reserved	reserved	reserved	non-reserved	reserved
INSERT	non-reserved	reserved	reserved	reserved	reserved
INSTANCE		non-reserved	non-reserved	non-reserved	
INSTANTIABLE		non-reserved	non-reserved	non-reserved	
INSTEAD	non-reserved	non-reserved			
INT	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
INTEGER	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
INTEGRITY		non-reserved	non-reserved		
INTERSECT	reserved	reserved	reserved	reserved	reserved
INTERSECTION		reserved	reserved		
INTERVAL	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
INTO	reserved	reserved	reserved	reserved	reserved
INVOKER	non-reserved	non-reserved	non-reserved	non-reserved	
IS	reserved (can be function or type)	reserved	reserved	reserved	reserved
ISNULL	reserved (can be function or type)				
ISOLATION	non-reserved	non-reserved	non-reserved	reserved	reserved
ITERATE				reserved	
JOIN	reserved (can be function or type)	reserved	reserved	reserved	reserved
K		non-reserved	non-reserved	non-reserved	
KEY	non-reserved	non-reserved	non-reserved	reserved	reserved
KEY_MEMBER		non-reserved	non-reserved	non-reserved	
KEY_TYPE		non-reserved	non-reserved	non-reserved	
LABEL	non-reserved				
LAG		reserved			
LANGUAGE	non-reserved	reserved	reserved	reserved	reserved
LARGE	non-reserved	reserved	reserved	reserved	

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
LAST	non-reserved	non-reserved	non-reserved	reserved	reserved
LAST_VALUE		reserved			
LATERAL		reserved	reserved	reserved	
LC_COLLATE	non-reserved				
LC_CTYPE	non-reserved				
LEAD		reserved			
LEADING	reserved	reserved	reserved	reserved	reserved
LEAST	non-reserved (cannot be function or type)				
LEFT	reserved (can be function or type)	reserved	reserved	reserved	reserved
LENGTH		non-reserved	non-reserved	non-reserved	non-reserved
LESS				reserved	
LEVEL	non-reserved	non-reserved	non-reserved	reserved	reserved
LIBRARY		non-reserved	non-reserved		
LIKE	reserved (can be function or type)	reserved	reserved	reserved	reserved
LIKE_REGEX		reserved			
LIMIT	reserved	non-reserved	non-reserved	reserved	
LINK		non-reserved	non-reserved		
LISTEN	non-reserved				
LN		reserved	reserved		
LOAD	non-reserved				
LOCAL	non-reserved	reserved	reserved	reserved	reserved
LOCALTIME	reserved	reserved	reserved	reserved	
LOCALTIMESTAMP	reserved	reserved	reserved	reserved	
LOCATION	non-reserved	non-reserved			
LOCATOR		non-reserved	non-reserved	reserved	
LOCK	non-reserved				
LOWER		reserved	reserved	non-reserved	reserved
M		non-reserved	non-reserved	non-reserved	
MAP		non-reserved	non-reserved	reserved	
MAPPING	non-reserved	non-reserved	non-reserved		
MATCH	non-reserved	reserved	reserved	reserved	reserved
MATCHED		non-reserved	non-reserved		
MAX		reserved	reserved	non-reserved	reserved
MAXVALUE	non-reserved	non-reserved	non-reserved		

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
MAX_CARDINALITY		reserved			
MEMBER		reserved	reserved		
MERGE		reserved	reserved		
MESSAGE_LENGTH		non-reserved	non-reserved	non-reserved	non-reserved
MESSAGE_OCTET_LENGTH		non-reserved	non-reserved	non-reserved	non-reserved
MESSAGE_TEXT		non-reserved	non-reserved	non-reserved	non-reserved
METHOD		reserved	reserved	non-reserved	
MIN		reserved	reserved	non-reserved	reserved
MINUTE	non-reserved	reserved	reserved	reserved	reserved
MINVALUE	non-reserved	non-reserved	non-reserved		
MOD		reserved	reserved	non-reserved	
MODE	non-reserved				
MODIFIES		reserved	reserved	reserved	
MODIFY				reserved	
MODULE		reserved	reserved	reserved	reserved
MONTH	non-reserved	reserved	reserved	reserved	reserved
MORE		non-reserved	non-reserved	non-reserved	non-reserved
MOVE	non-reserved				
MULTISET		reserved	reserved		
MUMPS		non-reserved	non-reserved	non-reserved	non-reserved
NAME	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
NAMES	non-reserved	non-reserved	non-reserved	reserved	reserved
NAMESPACE		non-reserved			
NATIONAL	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
NATURAL	reserved (can be function or type)	reserved	reserved	reserved	reserved
NCHAR	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
NCLOB		reserved	reserved	reserved	
NESTING		non-reserved	non-reserved		
NEW		reserved	reserved	reserved	
NEXT	non-reserved	non-reserved	non-reserved	reserved	reserved
NFC		non-reserved			
NFD		non-reserved			

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
NFKC		non-reserved			
NFKD		non-reserved			
NIL		non-reserved			
NO	non-reserved	reserved	reserved	reserved	reserved
NONE	non-reserved (cannot be function or type)	reserved	reserved	reserved	
NORMALIZE		reserved	reserved		
NORMALIZED		non-reserved	non-reserved		
NOT	reserved	reserved	reserved	reserved	reserved
NOTHING	non-reserved				
NOTIFY	non-reserved				
NOTNULL	reserved (can be function or type)				
NOWAIT	non-reserved				
NTH_VALUE		reserved			
NTILE		reserved			
NULL	reserved	reserved	reserved	reserved	reserved
NULLABLE		non-reserved	non-reserved	non-reserved	non-reserved
NULLIF	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
NULLS	non-reserved	non-reserved	non-reserved		
NUMBER		non-reserved	non-reserved	non-reserved	non-reserved
NUMERIC	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
OBJECT	non-reserved	non-reserved	non-reserved	reserved	
OCCURRENCES_REGEX		reserved			
OCTETS		non-reserved	non-reserved		
OCTET_LENGTH		reserved	reserved	non-reserved	reserved
OF	non-reserved	reserved	reserved	reserved	reserved
OFF	non-reserved	non-reserved	non-reserved	reserved	
OFFSET	reserved	reserved			
OIDS	non-reserved				
OLD		reserved	reserved	reserved	
ON	reserved	reserved	reserved	reserved	reserved
ONLY	reserved	reserved	reserved	reserved	reserved
OPEN		reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
OPERATION				reserved	
OPERATOR	non-reserved				
OPTION	non-reserved	non-reserved	non-reserved	reserved	reserved
OPTIONS	non-reserved	non-reserved	non-reserved	non-reserved	
OR	reserved	reserved	reserved	reserved	reserved
ORDER	reserved	reserved	reserved	reserved	reserved
ORDERING		non-reserved	non-reserved		
ORDINALITY		non-reserved	non-reserved	reserved	
OTHERS		non-reserved	non-reserved		
OUT	non-reserved (cannot be function or type)	reserved	reserved	reserved	
OUTER	reserved (can be function or type)	reserved	reserved	reserved	reserved
OUTPUT		non-reserved	non-reserved	reserved	reserved
OVER	reserved (can be function or type)	reserved	reserved		
OVERLAPS	reserved (can be function or type)	reserved	reserved	non-reserved	reserved
OVERLAY	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	
OVERRIDING		non-reserved	non-reserved	non-reserved	
OWNED	non-reserved				
OWNER	non-reserved				
P		non-reserved			
PAD		non-reserved	non-reserved	reserved	reserved
PARAMETER		reserved	reserved	reserved	
PARAMETERS				reserved	
PARAMETER_MODE		non-reserved	non-reserved	non-reserved	
PARAMETER_NAME		non-reserved	non-reserved	non-reserved	
PARAMETER_ORDINAL_POSITION		non-reserved	non-reserved	non-reserved	
PARAMETER_SPECIFIC_CATALOG		non-reserved	non-reserved	non-reserved	

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
PARAMETER_SPECIFIC_NAME		non-reserved	non-reserved	non-reserved	
PARAMETER_SPECIFIC_SCHEMA		non-reserved	non-reserved	non-reserved	
PARSER	non-reserved				
PARTIAL	non-reserved	non-reserved	non-reserved	reserved	reserved
PARTITION	non-reserved	reserved	reserved		
PASCAL		non-reserved	non-reserved	non-reserved	non-reserved
PASSING	non-reserved	non-reserved			
PASSTHROUGH		non-reserved	non-reserved		
PASSWORD	non-reserved				
PATH		non-reserved	non-reserved	reserved	
PERCENTILE_CONT		reserved	reserved		
PERCENTILE_DISC		reserved	reserved		
PERCENT_RANK		reserved	reserved		
PERMISSION		non-reserved	non-reserved		
PLACING	reserved	non-reserved	non-reserved		
PLANS	non-reserved				
PLI		non-reserved	non-reserved	non-reserved	non-reserved
POSITION	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
POSITION_REGEX		reserved			
POSTFIX				reserved	
POWER		reserved	reserved		
PRECEDING	non-reserved	non-reserved	non-reserved		
PRECISION	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
PREFIX				reserved	
PREORDER				reserved	
PREPARE	non-reserved	reserved	reserved	reserved	reserved
PREPARED	non-reserved				
PRESERVE	non-reserved	non-reserved	non-reserved	reserved	reserved
PRIMARY	reserved	reserved	reserved	reserved	reserved
PRIOR	non-reserved	non-reserved	non-reserved	reserved	reserved
PRIVILEGES	non-reserved	non-reserved	non-reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
PROCEDURAL	non-reserved				
PROCEDURE	non-reserved	reserved	reserved	reserved	reserved
PUBLIC		non-reserved	non-reserved	reserved	reserved
QUOTE	non-reserved				
RANGE	non-reserved	reserved	reserved		
RANK		reserved	reserved		
READ	non-reserved	non-reserved	non-reserved	reserved	reserved
READS		reserved	reserved	reserved	
REAL	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
REASSIGN	non-reserved				
RECHECK	non-reserved				
RECOVERY		non-reserved	non-reserved		
RECURSIVE	non-reserved	reserved	reserved	reserved	
REF	non-reserved	reserved	reserved	reserved	
REFERENCES	reserved	reserved	reserved	reserved	reserved
REFERENCING		reserved	reserved	reserved	
REGR_AVGX		reserved	reserved		
REGR_AVGY		reserved	reserved		
REGR_COUNT		reserved	reserved		
REGR_INTERCEPT		reserved	reserved		
REGR_R2		reserved	reserved		
REGR_SLOPE		reserved	reserved		
REGR_SXX		reserved	reserved		
REGR_SXY		reserved	reserved		
REGR_SYY		reserved	reserved		
REINDEX	non-reserved				
RELATIVE	non-reserved	non-reserved	non-reserved	reserved	reserved
RELEASE	non-reserved	reserved	reserved		
RENAME	non-reserved				
REPEATABLE	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
REPLACE	non-reserved				
REPLICA	non-reserved				
REQUIRING		non-reserved	non-reserved		
RESET	non-reserved				
RESPECT		non-reserved			
RESTART	non-reserved	non-reserved	non-reserved		
RESTORE		non-reserved	non-reserved		

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
RESTRICT	non-reserved	non-reserved	non-reserved	reserved	reserved
RESULT		reserved	reserved	reserved	
RETURN		reserved	reserved	reserved	
RETURNED_CARDINALITY		non-reserved	non-reserved		
RETURNED_LENGTH		non-reserved	non-reserved	non-reserved	non-reserved
RETURNED_OCTET_LENGTH		non-reserved	non-reserved	non-reserved	non-reserved
RETURNED_SQLSTATE		non-reserved	non-reserved	non-reserved	non-reserved
RETURNING	reserved	non-reserved			
RETURNS	non-reserved	reserved	reserved	reserved	
REVOKE	non-reserved	reserved	reserved	reserved	reserved
RIGHT	reserved (can be function or type)	reserved	reserved	reserved	reserved
ROLE	non-reserved	non-reserved	non-reserved	reserved	
ROLLBACK	non-reserved	reserved	reserved	reserved	reserved
ROLLUP		reserved	reserved	reserved	
ROUTINE		non-reserved	non-reserved	reserved	
ROUTINE_CATALOG		non-reserved	non-reserved	non-reserved	
ROUTINE_NAME		non-reserved	non-reserved	non-reserved	
ROUTINE_SCHEMA		non-reserved	non-reserved	non-reserved	
ROW	non-reserved (cannot be function or type)	reserved	reserved	reserved	
ROWS	non-reserved	reserved	reserved	reserved	reserved
ROW_COUNT		non-reserved	non-reserved	non-reserved	non-reserved
ROW_NUMBER		reserved	reserved		
RULE	non-reserved				
SAVEPOINT	non-reserved	reserved	reserved	reserved	
SCALE		non-reserved	non-reserved	non-reserved	non-reserved
SCHEMA	non-reserved	non-reserved	non-reserved	reserved	reserved
SCHEMA_NAME		non-reserved	non-reserved	non-reserved	non-reserved
SCOPE		reserved	reserved	reserved	
SCOPE_CATALOG		non-reserved	non-reserved		
SCOPE_NAME		non-reserved	non-reserved		
SCOPE_SCHEMA		non-reserved	non-reserved		
SCROLL	non-reserved	reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
SEARCH	non-reserved	reserved	reserved	reserved	
SECOND	non-reserved	reserved	reserved	reserved	reserved
SECTION		non-reserved	non-reserved	reserved	reserved
SECURITY	non-reserved	non-reserved	non-reserved	non-reserved	
SELECT	reserved	reserved	reserved	reserved	reserved
SELECTIVE		non-reserved	non-reserved		
SELF		non-reserved	non-reserved	non-reserved	
SENSITIVE		reserved	reserved	non-reserved	
SEQUENCE	non-reserved	non-reserved	non-reserved	reserved	
SEQUENCES	non-reserved				
SERIALIZABLE	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
SERVER	non-reserved	non-reserved	non-reserved		
SERVER_NAME		non-reserved	non-reserved	non-reserved	non-reserved
SESSION	non-reserved	non-reserved	non-reserved	reserved	reserved
SESSION_USER	reserved	reserved	reserved	reserved	reserved
SET	non-reserved	reserved	reserved	reserved	reserved
SETOF	non-reserved (cannot be function or type)				
SETS		non-reserved	non-reserved	reserved	
SHARE	non-reserved				
SHOW	non-reserved				
SIMILAR	reserved (can be function or type)	reserved	reserved	non-reserved	
SIMPLE	non-reserved	non-reserved	non-reserved	non-reserved	
SIZE		non-reserved	non-reserved	reserved	reserved
SMALLINT	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
SOME	reserved	reserved	reserved	reserved	reserved
SOURCE		non-reserved	non-reserved	non-reserved	
SPACE		non-reserved	non-reserved	reserved	reserved
SPECIFIC		reserved	reserved	reserved	
SPECIFICTYPE		reserved	reserved	reserved	
SPECIFIC_NAME		non-reserved	non-reserved	non-reserved	
SQL		reserved	reserved	reserved	reserved
SQLCODE					reserved
SQLERROR					reserved
SQLEXCEPTION		reserved	reserved	reserved	

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
SQLSTATE		reserved	reserved	reserved	reserved
SQLWARNING		reserved	reserved	reserved	
SQRT		reserved	reserved		
STABLE	non-reserved				
STANDALONE	non-reserved	non-reserved	non-reserved		
START	non-reserved	reserved	reserved	reserved	
STATE		non-reserved	non-reserved	reserved	
STATEMENT	non-reserved	non-reserved	non-reserved	reserved	
STATIC		reserved	reserved	reserved	
STATISTICS	non-reserved				
STDDEV_POP		reserved	reserved		
STDDEV_SAMP		reserved	reserved		
STDIN	non-reserved				
STDOUT	non-reserved				
STORAGE	non-reserved				
STRICT	non-reserved				
STRIP	non-reserved	non-reserved	non-reserved		
STRUCTURE		non-reserved	non-reserved	reserved	
STYLE		non-reserved	non-reserved	non-reserved	
SUBCLASS_ ORIGIN		non-reserved	non-reserved	non-reserved	non-reserved
SUBLIST				non-reserved	
SUBMULTISET		reserved	reserved		
SUBSTRING	non-reserv ed (cannot be function or type)	reserved	reserved	non-reserved	reserved
SUBSTRING_ REGEX		reserved			
SUM		reserved	reserved	non-reserved	reserved
SYMMETRIC	reserved	reserved	reserved	non-reserved	
SYSID	non-reserved				
SYSTEM	non-reserved	reserved	reserved	non-reserved	
SYSTEM_USER		reserved	reserved	reserved	reserved
T		non-reserved			
TABLE	reserved	reserved	reserved	reserved	reserved
TABLES	non-reserved				
TABLESAMPLE		reserved	reserved		
TABLESPACE	non-reserved				
TABLE_NAME		non-reserved	non-reserved	non-reserved	non-reserved
TEMP	non-reserved				

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
TEMPLATE	non-reserved				
TEMPORARY	non-reserved	non-reserved	non-reserved	reserved	reserved
TERMINATE				reserved	
TEXT	non-reserved				
THAN				reserved	
THEN	reserved	reserved	reserved	reserved	reserved
TIES		non-reserved	non-reserved		
TIME	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
TIMESTAMP	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
TIMEZONE_HOUR		reserved	reserved	reserved	reserved
TIMEZONE_MINUTE		reserved	reserved	reserved	reserved
TO	reserved	reserved	reserved	reserved	reserved
TOKEN		non-reserved	non-reserved		
TOP_LEVEL_COUNT		non-reserved	non-reserved		
TRAILING	reserved	reserved	reserved	reserved	reserved
TRANSACTION	non-reserved	non-reserved	non-reserved	reserved	reserved
TRANSACTIONS_COMMITTED		non-reserved	non-reserved	non-reserved	
TRANSACTIONS_ROLLED_BACK		non-reserved	non-reserved	non-reserved	
TRANSACTION_ACTIVE		non-reserved	non-reserved	non-reserved	
TRANSFORM		non-reserved	non-reserved	non-reserved	
TRANSFORMS		non-reserved	non-reserved	non-reserved	
TRANSLATE		reserved	reserved	non-reserved	reserved
TRANSLATE_REGEX		reserved			
TRANSLATION		reserved	reserved	reserved	reserved
TREAT	non-reserved (cannot be function or type)	reserved	reserved	reserved	
TRIGGER	non-reserved	reserved	reserved	reserved	
TRIGGER_CATALOG		non-reserved	non-reserved	non-reserved	

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
TRIGGER_NAME		non-reserved	non-reserved	non-reserved	
TRIGGER_SCHEMA		non-reserved	non-reserved	non-reserved	
TRIM	non-reserved (cannot be function or type)	reserved	reserved	non-reserved	reserved
TRIM_ARRAY		reserved			
TRUE	reserved	reserved	reserved	reserved	reserved
TRUNCATE	non-reserved	reserved			
TRUSTED	non-reserved				
TYPE	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
UESCAPE		reserved	reserved		
UNBOUNDED	non-reserved	non-reserved	non-reserved		
UNCOMMITTED	non-reserved	non-reserved	non-reserved	non-reserved	non-reserved
UNDER		non-reserved	non-reserved	reserved	
UNENCRYPTED	non-reserved				
UNION	reserved	reserved	reserved	reserved	reserved
UNIQUE	reserved	reserved	reserved	reserved	reserved
UNKNOWN	non-reserved	reserved	reserved	reserved	reserved
UNLINK		non-reserved	non-reserved		
UNLISTEN	non-reserved				
UNLOGGED	non-reserved				
UNNAMED		non-reserved	non-reserved	non-reserved	non-reserved
UNNEST		reserved	reserved	reserved	
UNTIL	non-reserved				
UNTYPED		non-reserved			
UPDATE	non-reserved	reserved	reserved	reserved	reserved
UPPER		reserved	reserved	non-reserved	reserved
URI		non-reserved			
USAGE		non-reserved	non-reserved	reserved	reserved
USER	reserved	reserved	reserved	reserved	reserved
USER_DEFINED_TYPE_CATALOG		non-reserved	non-reserved	non-reserved	
USER_DEFINED_TYPE_CODE		non-reserved	non-reserved		
USER_DEFINED_TYPE_NAME		non-reserved	non-reserved	non-reserved	
USER_DEFINED_TYPE_SCHEMA		non-reserved	non-reserved	non-reserved	
USING	reserved	reserved	reserved	reserved	reserved

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
VACUUM	non-reserved				
VALID	non-reserved	non-reserved			
VALIDATE	non-reserved				
VALIDATOR	non-reserved				
VALUE	non-reserved	reserved	reserved	reserved	reserved
VALUES	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
VARBINARY		reserved			
VARCHAR	non-reserved (cannot be function or type)	reserved	reserved	reserved	reserved
VARIABLE				reserved	
VARIADIC	reserved				
VARYING	non-reserved	reserved	reserved	reserved	reserved
VAR_POP		reserved	reserved		
VAR_SAMP		reserved	reserved		
VERBOSE	reserved (can be function or type)				
VERSION	non-reserved	non-reserved	non-reserved		
VIEW	non-reserved	non-reserved	non-reserved	reserved	reserved
VOLATILE	non-reserved				
WHEN	reserved	reserved	reserved	reserved	reserved
WHENEVER		reserved	reserved	reserved	reserved
WHERE	reserved	reserved	reserved	reserved	reserved
WHITESPACE	non-reserved	non-reserved	non-reserved		
WIDTH_BUCKET		reserved	reserved		
WINDOW	reserved	reserved	reserved		
WITH	reserved	reserved	reserved	reserved	reserved
WITHIN		reserved	reserved		
WITHOUT	non-reserved	reserved	reserved	reserved	
WORK	non-reserved	non-reserved	non-reserved	reserved	reserved
WRAPPER	non-reserved	non-reserved	non-reserved		
WRITE	non-reserved	non-reserved	non-reserved	reserved	reserved
XML	non-reserved	reserved	reserved		
XMLAGG		reserved	reserved		
XMLATTRIBUTES	non-reserved (cannot be function or type)	reserved	reserved		

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
XMLBINARY		reserved	reserved		
XMLCAST		reserved			
XMLCOMMENT		reserved	reserved		
XMLCONCAT	non-reserved (cannot be function or type)	reserved	reserved		
XMLDECLARATION		non-reserved			
XMLDOCUMENT		reserved			
XMLELEMENT	non-reserved (cannot be function or type)	reserved	reserved		
XMLEXISTS	non-reserved (cannot be function or type)	reserved			
XMLFOREST	non-reserved (cannot be function or type)	reserved	reserved		
XMLITERATE		reserved			
XMLNAMESPACES		reserved	reserved		
XMLPARSE	non-reserved (cannot be function or type)	reserved	reserved		
XMLPI	non-reserved (cannot be function or type)	reserved	reserved		
XMLQUERY		reserved			
XMLROOT	non-reserved (cannot be function or type)		reserved		
XMLSCHEMA		non-reserved			
XMLSERIALIZE	non-reserved (cannot be function or type)	reserved	reserved		
XMLTABLE		reserved			
XMLTEXT		reserved			
XMLVALIDATE		reserved			
YEAR	non-reserved	reserved	reserved	reserved	reserved
YES	non-reserved	non-reserved			

Key Word	PostgreSQL	SQL:2008	SQL:2003	SQL:1999	SQL-92
ZONE	non-reserved	non-reserved	non-reserved	reserved	reserved

附录 D. SQL 符合性

本节尝试概述PostgreSQL在多大程度上符合当前的 SQL 标准。下文并不是一份完整的符合性声明，但会在合理且对用户有用的范围内，尽可能详细地介绍主要主题。

SQL 标准的正式名称是 ISO/IEC 9075 “数据库语言 SQL”。该标准会不时发布修订版本；最近一次更新出现在 2008 年。2008 年的版本被称为 ISO/IEC 9075:2008，或简称 SQL:2008。在它之前的版本分别是 SQL:2003、SQL:1999 和 SQL-92。每个版本都会替代前一个版本，因此声称与更早版本一致的说法并没有官方效力。PostgreSQL 开发目标是在不与特性和历史兼容性冲突的前提下，尽量符合标准的最新官方版本 does not contradict traditional features or common sense. Many of the features required by the SQL standard are supported, though sometimes with slightly differing syntax or function. Further moves towards conformance can be expected over time.

SQL-92为符合性定义了三个特性集：入门（Entry）、中级（Intermediate）和完整（Full）。
多数宣称符合SQL标准的数据库管理系统，实际上只符合入门（Entry）级别，
因为中级（Intermediate）和完整（Full）级别中的全部特性要么过于庞杂，
要么与既有行为相冲突。

从SQL:1999开始，SQL 标准不再沿用SQL-92中那种效果不佳且过于宽泛的三级分类，而是定义了大量独立特性。其中很大一个子集构成了核心（“Core”）特性，也就是每个符合 SQL 标准的实现都必须提供的特性。其余特性则完全是可选的。某些可选特性被组合在一起，形成“包”，
SQL 实现可以声明符合这些包，从而表明符合特定的特性组。

SQL:2008和SQL:2003标准版本也分为若干部分，每部分都有一个简称。
注意这些部分的编号并不连续。

- ISO/IEC 9075-1 框架（SQL/Framework）
- ISO/IEC 9075-2 基础（SQL/Foundation）
- ISO/IEC 9075-3 调用级接口（SQL/CLI）
- ISO/IEC 9075-4 持久存储模块（SQL/PSM）
- ISO/IEC 9075-9 外部数据管理（SQL/MED）
- ISO/IEC 9075-10 对象语言绑定（SQL/OLB）
- ISO/IEC 9075-11 信息和定义模式（SQL/Schemata）
- ISO/IEC 9075-13 使用 Java 语言的例程和类型（SQL/JRT）
- ISO/IEC 9075-14 XML 相关规范（SQL/XML）

PostgreSQL核心部分涵盖第 1、2、9、11 和 14 部分。第 3 部分由 ODBC 驱动实现，第 13 部分由 PL/Java 插件实现，但目前尚未核验这些组件的确切符合性。PostgreSQL目前还没有第 4 和 10 部分的实现。

PostgreSQL 支持 SQL:2008 的大部分主要特性。在完全符合 Core 所要求的 179 个强制特性中，PostgreSQL 至少符合 160 个。此外，还有一长串受支持的可选特性。值得注意的是，截至撰写时，尚无任何数据库管理系统的当前版本声称完全符合 Core SQL:2008。

在下面两节中，我们先列出PostgreSQL支持的特性，然后列出SQL:2008中定义但PostgreSQL尚不支持的特性。这两个列表都是近似的：被列为受支持的特性可能存在个别不符合标准的细节，而不受支持的特性也可能实际上已实现了很大一部分。关于什么可用什么不可用的最准确信息，始终以文档正文为准。

注意

带连字符的特性代码表示子特性。因此，如果某个特定子特性不受支持，则即使该主特性的其他子特性受支持，该主特性也会被列为不受支持。

D.1. 支持的特性

标识符	包	描述	注释
B012		Embedded C	
B021		Direct SQL	
E011	Core	Numeric data types	
E011-01	Core	INTEGER and SMALLINT data types	
E011-02	Core	REAL, DOUBLE PRECISION, and FLOAT data types	
E011-03	Core	DECIMAL and NUMERIC data types	
E011-04	Core	Arithmetic operators	
E011-05	Core	Numeric comparison	
E011-06	Core	Implicit casting among the numeric data types	
E021	Core	Character data types	
E021-01	Core	CHARACTER data type	
E021-02	Core	CHARACTER VARYING data type	
E021-03	Core	Character literals	
E021-04	Core	CHARACTER_LENGTH function	trims trailing spaces from CHARACTER values before counting
E021-05	Core	OCTET_LENGTH function	
E021-06	Core	SUBSTRING function	
E021-07	Core	Character concatenation	
E021-08	Core	UPPER and LOWER functions	

标识符	包	描述	注释
E021-09	Core	TRIM function	
E021-10	Core	Implicit casting among the character string types	
E021-11	Core	POSITION function	
E021-12	Core	Character comparison	
E031	Core	Identifiers	
E031-01	Core	Delimited identifiers	
E031-02	Core	Lower case identifiers	
E031-03	Core	Trailing underscore	
E051	Core	Basic query specificatio n	
E051-01	Core	SELECT DISTINCT	
E051-02	Core	GROUP BY clause	
E051-04	Core	GROUP BY can contain columns not in <select list>	
E051-05	Core	Select list items can be renamed	
E051-06	Core	HAVING clause	
E051-07	Core	Qualified * in select list	
E051-08	Core	Correlation names in the FROM clause	
E051-09	Core	Rename columns in the FROM clause	
E061	Core	Basic predicates and search conditions	
E061-01	Core	Comparison predicate	
E061-02	Core	BETWEEN predicate	
E061-03	Core	IN predicate with list of values	
E061-04	Core	LIKE predicate	
E061-05	Core	LIKE predicate ESCAPE clause	
E061-06	Core	NULL predicate	
E061-07	Core	Quantified comparison predicate	
E061-08	Core	EXISTS predicate	
E061-09	Core	Subqueries in compar ison predicate	
E061-11	Core	Subqueries in IN predic ate	
E061-12	Core	Subqueries in quanti fied comparison predic ate	
E061-13	Core	Correlated subqueries	

标识符	包	描述	注释
E061-14	Core	Search condition	
E071	Core	Basic query expres sions	
E071-01	Core	UNION DISTINCT table operator	
E071-02	Core	UNION ALL table operator	
E071-03	Core	EXCEPT DISTINCT table operator	
E071-05	Core	Columns combined via table operators need not have exactly the same data type	
E071-06	Core	Table operators in subqueries	
E081-01	Core	SELECT privilege	
E081-02	Core	DELETE privilege	
E081-03	Core	INSERT privilege at the table level	
E081-04	Core	UPDATE privilege at the table level	
E081-05	Core	UPDATE privilege at the column level	
E081-06	Core	REFERENCES privilege at the table level	
E081-07	Core	REFERENCES privilege at the column level	
E081-08	Core	WITH GRANT OPTION	
E081-10	Core	EXECUTE privilege	
E091	Core	Set functions	
E091-01	Core	AVG	
E091-02	Core	COUNT	
E091-03	Core	MAX	
E091-04	Core	MIN	
E091-05	Core	SUM	
E091-06	Core	ALL quantifier	
E091-07	Core	DISTINCT quantifier	
E101	Core	Basic data manipu lation	
E101-01	Core	INSERT statement	
E101-03	Core	Searched UPDATE statement	
E101-04	Core	Searched DELETE statement	
E111	Core	Single row SELECT statement	

标识符	包	描述	注释
E121	Core	Basic cursor support	
E121-01	Core	DECLARE CURSOR	
E121-02	Core	ORDER BY columns need not be in select list	
E121-03	Core	Value expressions in ORDER BY clause	
E121-04	Core	OPEN statement	
E121-06	Core	Positioned UPDATE statement	
E121-07	Core	Positioned DELETE statement	
E121-08	Core	CLOSE statement	
E121-10	Core	FETCH statement implic it NEXT	
E121-17	Core	WITH HOLD cursors	
E131	Core	Null value support (nulls in lieu of values)	
E141	Core	Basic integrity constr aints	
E141-01	Core	NOT NULL constraints	
E141-02	Core	UNIQUE constraints of NOT NULL columns	
E141-03	Core	PRIMARY KEY constr aints	
E141-04	Core	Basic FOREIGN KEY constraint with the NO ACTION default for both referential delete action and referential update action	
E141-06	Core	CHECK constraints	
E141-07	Core	Column defaults	
E141-08	Core	NOT NULL inferred on PRIMARY KEY	
E141-10	Core	Names in a foreign key can be specified in any order	
E151	Core	Transaction support	
E151-01	Core	COMMIT statement	
E151-02	Core	ROLLBACK statement	
E152	Core	Basic SET TRANSA CTION statement	
E152-01	Core	SET TRANSACTION statement: ISOLATION LEVEL SERIALIZABLE clause	

标识符	包	描述	注释
E152-02	Core	SET TRANSACTION statement: READ ONLY and READ WRITE clauses	
E161	Core	SQL comments using leading double minus	
E171	Core	SQLSTATE support	
F021	Core	Basic information schema	
F021-01	Core	COLUMNS view	
F021-02	Core	TABLES view	
F021-03	Core	VIEWS view	
F021-04	Core	TABLE_CONSTRAINTS view	
F021-05	Core	REFERENTIAL_CONSTRAINTS view	
F021-06	Core	CHECK_CONSTRAINTS view	
F031	Core	Basic schema manipulation	
F031-01	Core	CREATE TABLE statement to create persistent base tables	
F031-02	Core	CREATE VIEW statement	
F031-03	Core	GRANT statement	
F031-04	Core	ALTER TABLE statement: ADD COLUMN clause	
F031-13	Core	DROP TABLE statement: RESTRICT clause	
F031-16	Core	DROP VIEW statement: RESTRICT clause	
F031-19	Core	REVOKE statement: RESTRICT clause	
F032		CASCADE drop behavior	
F033		ALTER TABLE statement: DROP COLUMN clause	
F034		Extended REVOKE statement	
F034-01		REVOKE statement performed by other than the owner of a schema object	
F034-02		REVOKE statement: GRANT OPTION FOR clause	

标识符	包	描述	注释
F034-03		REVOKE statement to revoke a privilege that the grantee has WITH GRANT OPTION	
F041	Core	Basic joined table	
F041-01	Core	Inner join (but not necessarily the INNER keyword)	
F041-02	Core	INNER keyword	
F041-03	Core	LEFT OUTER JOIN	
F041-04	Core	RIGHT OUTER JOIN	
F041-05	Core	Outer joins can be nested	
F041-07	Core	The inner table in a left or right outer join can also be used in an inner join	
F041-08	Core	All comparison operat ors are supported (rather than just =)	
F051	Core	Basic date and time	
F051-01	Core	DATE data type (includ ing support of DATE literal)	
F051-02	Core	TIME data type (includ ing support of TIME literal) with fractional seconds precision of at least 0	
F051-03	Core	TIMESTAMP data type (including support of TIMESTAMP literal) with fractional second s precision of at least 0 and 6	
F051-04	Core	Comparison predicate on DATE, TIME, and TIMESTAMP data types	
F051-05	Core	Explicit CAST betwee n datetime types and character string types	
F051-06	Core	CURRENT_DATE	
F051-07	Core	LOCALTIME	
F051-08	Core	LOCALTIMESTAMP	
F052	Enhanced datetime facilities	Intervals and datetime arithmetic	
F053		OVERLAPS predicate	
F081	Core	UNION and EXCEPT in views	

标识符	包	描述	注释
F111		Isolation levels other than SERIALIZABLE	
F111-01		READ UNCOMMITTED isolation level	
F111-02		READ COMMITTED isolation level	
F111-03		REPEATABLE READ isolation level	
F131	Core	Grouped operations	
F131-01	Core	WHERE, GROUP BY, and HAVING clauses supported in queries with grouped views	
F131-02	Core	Multiple tables suppor ted in queries with grouped views	
F131-03	Core	Set functions suppor ted in queries with grouped views	
F131-04	Core	Subqueries with GROUP BY and HAVING clauses and grouped views	
F131-05	Core	Single row SELECT with GROUP BY and HAVING clauses and grouped views	
F171		Multiple schemas per user	
F191	Enhanced integrity management	Referential delete actions	
F200		TRUNCATE TABLE statement	
F201	Core	CAST function	
F221	Core	Explicit defaults	
F222		INSERT statement: DEFAULT VALUES clause	
F231		Privilege tables	
F231-01		TABLE_PRIVILEGES view	
F231-02		COLUMN_PRIVILEGES view	
F231-03		USAGE_PRIVILEGES view	
F251		Domain support	
F261	Core	CASE expression	
F261-01	Core	Simple CASE	
F261-02	Core	Searched CASE	

标识符	包	描述	注释
F261-03	Core	NULLIF	
F261-04	Core	COALESCE	
F262		Extended CASE expres sion	
F271		Compound character literals	
F281		LIKE enhancements	
F302		INTERSECT table operat or	
F302-01		INTERSECT DISTINCT table operator	
F302-02		INTERSECT ALL table operator	
F304		EXCEPT ALL table operator	
F311-01	Core	CREATE SCHEMA	
F311-02	Core	CREATE TABLE for persis tent base tables	
F311-03	Core	CREATE VIEW	
F311-05	Core	GRANT statement	
F321		User authorization	
F361		Subprogram support	
F381		Extended schema manipulation	
F381-01		ALTER TABLE statem ent: ALTER COLUMN clause	
F381-02		ALTER TABLE statem ent: ADD CONSTRAINT clause	
F381-03		ALTER TABLE statem ent: DROP CONSTRAINT clause	
F382		Alter column data type	
F391		Long identifiers	
F392		Unicode escapes in identifiers	
F393		Unicode escapes in literals	
F401		Extended joined table	
F401-01		NATURAL JOIN	
F401-02		FULL OUTER JOIN	
F401-04		CROSS JOIN	
F402		Named column joins for LOBs, arrays, and multisets	

标识符	包	描述	注释
F411	Enhanced datetime facilities	Time zone specification	differences regarding literal interpretation
F421		National character	
F431		Read-only scrollable cursors	
F431-01		FETCH with explicit NEXT	
F431-02		FETCH FIRST	
F431-03		FETCH LAST	
F431-04		FETCH PRIOR	
F431-05		FETCH ABSOLUTE	
F431-06		FETCH RELATIVE	
F441		Extended set function support	
F442		Mixed column refere nces in set functions	
F471	Core	Scalar subquery values	
F481	Core	Expanded NULL predic ate	
F491	Enhanced integrity management	Constraint manage ment	
F501	Core	Features and confor mance views	
F501-01	Core	SQL_FEATURES view	
F501-02	Core	SQL_SIZING view	
F501-03	Core	SQL_LANGUAGES view	
F502		Enhanced docume ntation tables	
F502-01		SQL_SIZING_PROFILES view	
F502-02		SQL_IMPLEMENTATION_ INFO view	
F502-03		SQL_PACKAGES view	
F531		Temporary tables	
F555	Enhanced datetime facilities	Enhanced seconds precision	
F561		Full value expressions	
F571		Truth value tests	
F591		Derived tables	
F611		Indicator data types	
F641		Row and table constr uctors	
F651		Catalog name qualif iers	
F661		Simple tables	

标识符	包	描述	注释
F672		Retrospective check constraints	
F690		Collation support	but no character set support
F692		Extended collation support	
F701	Enhanced integrity management	Referential update actions	
F711		ALTER domain	
F731		INSERT column privil eges	
F761		Session management	
F762		CURRENT_CATALOG	
F763		CURRENT_SCHEMA	
F771		Connection manage ment	
F781		Self-referencing operat ions	
F791		Insensitive cursors	
F801		Full set function	
F850		Top-level <order by clause> in <query expression>	
F851		<order by clause> in subqueries	
F852		Top-level <order by clause> in views	
F855		Nested <order by clause> in <query expression>	
F856		Nested <fetch first clause> in <query expression>	
F857		Top-level <fetch first clause> in <query expression>	
F858		<fetch first clause> in subqueries	
F859		Top-level <fetch first clause> in views	
F860		<fetch first row count> in <fetch first clause>	
F861		Top-level <result offset clause> in <query expression>	
F862		<result offset clause> in subqueries	

标识符	包	描述	注释
F863		Nested <result offset clause> in <query expression>	
F864		Top-level <result offset clause> in views	
F865		<offset row count> in <result offset clause>	
S071	Enhanced object support	SQL paths in function and type name resolu tion	
S092		Arrays of user-defined types	
S095		Array constructors by query	
S096		Optional array bounds	
S098		ARRAY_AGG	
S111	Enhanced object support	ONLY in query expres sions	
S201		SQL-invoked routines on arrays	
S201-01		Array parameters	
S201-02		Array as result type of functions	
S211	Enhanced object support	User-defined cast functions	
T031		BOOLEAN data type	
T071		BIGINT data type	
T121		WITH (excluding RECURSIVE) in query expression	
T122		WITH (excluding RECURSIVE) in subque ry	
T131		Recursive query	
T132		Recursive query in subquery	
T141		SIMILAR predicate	
T151		DISTINCT predicate	
T152		DISTINCT predicate with negation	
T171		LIKE clause in table definition	
T172		AS subquery clause in table definition	
T173		Extended LIKE clause in table definition	

标识符	包	描述	注释
T191	Enhanced integrity management	Referential action RESTRICT	
T201	Enhanced integrity management	Comparable data types for referential constr aints	
T211-01	Active database, Enhanced integrity management	Triggers activated on UPDATE, INSERT, or DELETE of one base table	
T211-02	Active database, Enhanced integrity management	BEFORE triggers	
T211-03	Active database, Enhanced integrity management	AFTER triggers	
T211-04	Active database, Enhanced integrity management	FOR EACH ROW triggers	
T211-05	Active database, Enhanced integrity management	Ability to specify a search condition that must be true before the trigger is invoked	
T211-07	Active database, Enhanced integrity management	TRIGGER privilege	
T212	Enhanced integrity management	Enhanced trigger capability	
T213		INSTEAD OF triggers	
T231		Sensitive cursors	
T241		START TRANSACTION statement	
T271		Savepoints	
T281		SELECT privilege with column granularity	
T312		OVERLAY function	
T321-01	Core	User-defined functions with no overloading	
T321-03	Core	Function invocation	
T321-06	Core	ROUTINES view	
T321-07	Core	PARAMETERS view	
T322	PSM	Overloading of SQL-invoked functions and procedures	
T323		Explicit security for external routines	
T331		Basic roles	
T351		Bracketed SQL comments (/*...*/ comments)	

标识符	包	描述	注释
T441		ABS and MOD functions	
T461		Symmetric BETWEEN predicate	
T501		Enhanced EXISTS predicate	
T551		Optional key words for default syntax	
T581		Regular expression substring function	
T591		UNIQUE constraints of possibly null columns	
T614		NTILE function	
T615		LEAD and LAG functi ons	
T617		FIRST_VALUE and LAST_ VALUE function	
T621		Enhanced numeric functions	
T631	Core	IN predicate with one list element	
T651		SQL-schema statem ents in SQL routines	
T655		Cyclically dependent routines	
X010		XML type	
X011		Arrays of XML type	
X016		Persistent XML values	
X020		XMLConcat	
X031		XMLElement	
X032		XMLForest	
X034		XMLAgg	
X035		XMLAgg: ORDER BY option	
X036		XMLComment	
X037		XMLPI	
X040		Basic table mapping	
X041		Basic table mapping: nulls absent	
X042		Basic table mapping: null as nil	
X043		Basic table mapping: table as forest	
X044		Basic table mapping: table as element	
X045		Basic table mapping: with target namespace	

标识符	包	描述	注释
X046		Basic table mapping: data mapping	
X047		Basic table mapping: metadata mapping	
X048		Basic table mapping: base64 encoding of binary strings	
X049		Basic table mapping: hex encoding of binary strings	
X050		Advanced table mapping	
X051		Advanced table mapping: nulls absent	
X052		Advanced table mapping: null as nil	
X053		Advanced table mapping: table as forest	
X054		Advanced table mapping: table as element	
X055		Advanced table mapping: target namespace	
X056		Advanced table mapping: data mappin g	
X057		Advanced table mapping: metadata mapping	
X058		Advanced table mapping: base64 encoding of binary strings	
X059		Advanced table mapping: hex encodi ng of binary strings	
X060		XMLParse: Charac ter string input and CONTENT option	
X061		XMLParse: Charac ter string input and DOCUMENT option	
X070		XMLSerialize: Charac ter string serialization and CONTENT option	
X071		XMLSerialize: Charac ter string serialization and DOCUMENT option	

标识符	包	描述	注释
X072		XMLSerialize: Charac ter string serialization	
X090		XML document predic ate	
X120		XML parameters in SQL routines	
X121		XML parameters in external routines	
X400		Name and identifier mapping	

D.2. 不支持的特性

PostgreSQL的本版本未实现下列SQL:2008中定义的特性。在少数情况下，有等价的功能可用。

Identifier	Package	Description	Comment
B011		Embedded Ada	
B013		Embedded COBOL	
B014		Embedded Fortran	
B015		Embedded MUMPS	
B016		Embedded Pascal	
B017		Embedded PL/I	
B031		Basic dynamic SQL	
B032		Extended dynamic SQL	
B032-01		<describe input statem ent>	
B033		Untyped SQL-invoked function arguments	
B034		Dynamic specification of cursor attributes	
B035		Non-extended descri ptor names	
B041		Extensions to embedd ed SQL exception declarations	
B051		Enhanced execution rights	
B111		Module language Ada	
B112		Module language C	
B113		Module language COBOL	
B114		Module language Fortran	
B115		Module language MUMPS	
B116		Module language Pascal	

Identifier	Package	Description	Comment
B117		Module language PL/I	
B121		Routine language Ada	
B122		Routine language C	
B123		Routine language COBOL	
B124		Routine language Fortran	
B125		Routine language MUMPS	
B126		Routine language Pascal	
B127		Routine language PL/I	
B128		Routine language SQL	
E081	Core	Basic Privileges	
E081-09	Core	USAGE privilege	
E153	Core	Updatable queries with subqueries	
E182	Core	Module language	
F121		Basic diagnostics management	
F121-01		GET DIAGNOSTICS statement	
F121-02		SET TRANSACTION statement: DIAGNO STICS SIZE clause	
F122		Enhanced diagnostics management	
F123		All diagnostics	
F181	Core	Multiple module support	
F202		TRUNCATE TABLE: identity column restar t option	
F263		Comma-separated predicates in simple CASE expression	
F291		UNIQUE predicate	
F301		CORRESPONDING in query expressions	
F311	Core	Schema definition statement	
F311-04	Core	CREATE VIEW: WITH CHECK OPTION	
F312		MERGE statement	
F313		Enhanced MERGE statement	

Identifier	Package	Description	Comment
F341		Usage tables	no ROUTINE_*_USAGE tables
F394		Optional normal form specification	
F403		Partitioned joined tables	
F451		Character set definition	
F461		Named character sets	
F521	Enhanced integrity management	Assertions	
F671	Enhanced integrity management	Subqueries in CHECK	intentionally omitted
F693		SQL-session and client module collations	
F695		Translation support	
F696		Additional translation documentation	
F721		Deferrable constraints	foreign and unique keys only
F741		Referential MATCH types	no partial match yet
F751		View CHECK enhanc ements	
F812	Core	Basic flagging	
F813		Extended flagging	
F821		Local table references	
F831		Full cursor update	
F831-01		Updatable scrollable cursors	
F831-02		Updatable ordered cursors	
F841		LIKE_REGEX predicate	
F842		OCCURENCES_REGEX function	
F843		POSITION_REGEX functi on	
F844		SUBSTRING_REGEX function	
F845		TRANSLATE_REGEX function	
F846		Octet support in regula r expression operators	
F847		Nonconstant regular expressions	
S011	Core	Distinct data types	

Identifier	Package	Description	Comment
S011-01	Core	USER_DEFINED_TYPES view	
S023	Basic object support	Basic structured types	
S024	Enhanced object support	Enhanced structured types	
S025		Final structured types	
S026		Self-referencing structured types	
S027		Create method by specific method name	
S028		Permutable UDT options list	
S041	Basic object support	Basic reference types	
S043	Enhanced object support	Enhanced reference types	
S051	Basic object support	Create table of type	partially supported
S081	Enhanced object support	Subtables	
S091		Basic array support	partially supported
S091-01		Arrays of built-in data types	
S091-02		Arrays of distinct types	
S091-03		Array expressions	
S094		Arrays of reference types	
S097		Array element assignment	
S151	Basic object support	Type predicate	
S161	Enhanced object support	Subtype treatment	
S162		Subtype treatment for references	
S202		SQL-invoked routines on multisets	
S231	Enhanced object support	Structured type locators	
S232		Array locators	
S233		Multiset locators	
S241		Transform functions	
S242		Alter transform statement	
S251		User-defined orderings	
S261		Specific type method	
S271		Basic multiset support	
S272		Multisets of user-defined types	

Identifier	Package	Description	Comment
S274		Multisets of reference types	
S275		Advanced multiset support	
S281		Nested collection types	
S291		Unique constraint on entire row	
S301		Enhanced UNNEST	
S401		Distinct types based on array types	
S402		Distinct types based on distinct types	
S403		MAX_CARDINALITY	
S404		TRIM_ARRAY	
T011		Timestamp in Inform ation Schema	
T021		BINARY and VARBINARY data types	
T022		Advanced support for BINARY and VARBINARY data types	
T023		Compound binary literal	
T024		Spaces in binary litera ls	
T041	Basic object support	Basic LOB data type support	
T041-01	Basic object support	BLOB data type	
T041-02	Basic object support	CLOB data type	
T041-03	Basic object support	POSITION, LENGTH, LOWER, TRIM, UPPER, and SUBSTRING functi ons for LOB data types	
T041-04	Basic object support	Concatenation of LOB data types	
T041-05	Basic object support	LOB locator: non-holdable	
T042		Extended LOB data type support	
T043		Multiplier T	
T044		Multiplier P	
T051		Row types	
T052		MAX and MIN for row types	
T053		Explicit aliases for all-fields reference	
T061		UCS support	

Identifier	Package	Description	Comment
T101		Enhanced nullability determination	
T111		Updatable joins, unions, and columns	
T174		Identity columns	
T175		Generated columns	
T176		Sequence generator support	
T177		Sequence generator support: simple restart option	
T178		Identity columns: simple restart option	
T211	Active database, Enhanced integrity management	Basic trigger capability	
T211-06	Active database, Enhanced integrity management	Support for run-time rules for the interaction of triggers and constraints	
T211-08	Active database, Enhanced integrity management	Multiple triggers for the same event are executed in the order in which they were created in the catalog	intentionally omitted
T251		SET TRANSACTION statement: LOCAL option	
T261		Chained transactions	
T272		Enhanced savepoint management	
T285		Enhanced derived column names	
T301		Functional dependencies	partially supported
T321	Core	Basic SQL-invoked routines	
T321-02	Core	User-defined stored procedures with no overloading	
T321-04	Core	CALL statement	
T321-05	Core	RETURN statement	
T324		Explicit security for SQL routines	
T325		Qualified SQL parameter references	
T326		Table functions	
T332		Extended roles	mostly supported

Identifier	Package	Description	Comment
T431	OLAP	Extended grouping capabilities	
T432		Nested and concat enated GROUPING SETS	
T433		Multiargument GROUPI NG function	
T434		GROUP BY DISTINCT	
T471		Result sets return value	
T491		LATERAL derived table	
T511		Transaction counts	
T541		Updatable table refere nces	
T561		Holdable locators	
T571		Array-returning extern al SQL-invoked functi ons	
T572		Multiset-returning external SQL-invoked functions	
T601		Local cursor references	
T611	OLAP	Elementary OLAP operations	most forms supported
T612		Advanced OLAP operat ions	some forms supported
T613		Sampling	
T616		Null treatment option for LEAD and LAG functions	
T618		NTH_VALUE function	function exists, but some options missing
T641		Multiple column assign ment	only some syntax variants supported
T652		SQL-dynamic statem ents in SQL routines	
T653		SQL-schema statem ents in external routin es	
T654		SQL-dynamic statem ents in external routin es	
M001		Datalinks	
M002		Datalinks via SQL/CLI	
M003		Datalinks via Embedd ed SQL	
M004		Foreign data support	partially supported
M005		Foreign schema support	

Identifier	Package	Description	Comment
M006		GetSQLString routine	
M007		TransmitRequest	
M009		GetOpts and GetStatistics routines	
M010		Foreign data wrapper support	
M011		Datalinks via Ada	
M012		Datalinks via C	
M013		Datalinks via COBOL	
M014		Datalinks via Fortran	
M015		Datalinks via M	
M016		Datalinks via Pascal	
M017		Datalinks via PL/I	
M018		Foreign data wrapper interface routines in Ada	
M019		Foreign data wrapper interface routines in C	
M020		Foreign data wrapper interface routines in COBOL	
M021		Foreign data wrapper interface routines in Fortran	
M022		Foreign data wrapper interface routines in MUMPS	
M023		Foreign data wrapper interface routines in Pascal	
M024		Foreign data wrapper interface routines in PL/I	
M030		SQL-server foreign data support	
M031		Foreign data wrapper general routines	
X012		Multisets of XML type	
X013		Distinct types of XML type	
X014		Attributes of XML type	
X015		Fields of XML type	
X025		XMLCast	
X030		XMLDocument	
X038		XMLText	

Identifier	Package	Description	Comment
X065		XMLParse: BLOB input and CONTENT option	
X066		XMLParse: BLOB input and DOCUMENT option	
X068		XMLSerialize: BOM	
X069		XMLSerialize: INDENT	
X073		XMLSerialize: BLOB serialization and CONTENT option	
X074		XMLSerialize: BLOB serialization and DOCUMENT option	
X075		XMLSerialize: BLOB serialization	
X076		XMLSerialize: VERSION	
X077		XMLSerialize: explicit ENCODING option	
X078		XMLSerialize: explicit XML declaration	
X080		Namespaces in XML publishing	
X081		Query-level XML namespace declar ations	
X082		XML namespace declar ations in DML	
X083		XML namespace declar ations in DDL	
X084		XML namespace declar ations in compound statements	
X085		Predefined namespace prefixes	
X086		XML namespace declar ations in XMLTable	
X091		XML content predicate	
X096		XMLExists	
X100		Host language suppor t for XML: CONTENT option	
X101		Host language suppor t for XML: DOCUMENT option	
X110		Host language suppor t for XML: VARCHAR mapping	
X111		Host language suppor t for XML: CLOB mappin g	

Identifier	Package	Description	Comment
X112		Host language suppor t for XML: BLOB mappin g	
X113		Host language suppor t for XML: STRIP WHITES PACE option	
X114		Host language suppor t for XML: PRESERVE WHITESPACE option	
X131		Query-level XMLBIN ARY clause	
X132		XMLBINARY clause in DML	
X133		XMLBINARY clause in DDL	
X134		XMLBINARY clause in compound statements	
X135		XMLBINARY clause in subqueries	
X141		IS VALID predicate: data-driven case	
X142		IS VALID predicate: ACCORDING TO clause	
X143		IS VALID predicate: ELEMENT clause	
X144		IS VALID predicate: schema location	
X145		IS VALID predicate outside check constr aints	
X151		IS VALID predicate with DOCUMENT option	
X152		IS VALID predicate with CONTENT option	
X153		IS VALID predicate with SEQUENCE option	
X155		IS VALID predicate: NAMESPACE without ELEMENT clause	
X157		IS VALID predicate: NO NAMESPACE with ELEMENT clause	
X160		Basic Information Schema for registered XML Schemas	
X161		Advanced Information Schema for registered XML Schemas	

Identifier	Package	Description	Comment
X170		XML null handling options	
X171		NIL ON NO CONTENT option	
X181		XML(DOCUMENT(UNTYPED)) type	
X182		XML(DOCUMENT(ANY)) type	
X190		XML(SEQUENCE) type	
X191		XML(DOCUMENT(XMLSCHEMA)) type	
X192		XML(CONTENT(XMLSCH EMA)) type	
X200		XMLQuery	
X201		XMLQuery: RETURNING CONTENT	
X202		XMLQuery: RETURNING SEQUENCE	
X203		XMLQuery: passing a context item	
X204		XMLQuery: initializing an XQuery variable	
X205		XMLQuery: EMPTY ON EMPTY option	
X206		XMLQuery: NULL ON EMPTY option	
X211		XML 1.1 support	
X221		XML passing mechan ism BY VALUE	
X222		XML passing mechan ism BY REF	
X231		XML(CONTENT(UNTYPE D)) type	
X232		XML(CONTENT(ANY)) type	
X241		RETURNING CONTENT in XML publishing	
X242		RETURNING SEQUENCE in XML publishing	
X251		Persistent XML values of XML(DOCUMENT(UNTYPED)) type	
X252		Persistent XML values of XML(DOCUMENT(ANY)) type	
X253		Persistent XML values of XML(CONTENT(UNTYPED)) type	

Identifier	Package	Description	Comment
X254		Persistent XML values of XML(CONTENT(ANY)) type	
X255		Persistent XML values of XML(SEQUENCE) type	
X256		Persistent XML values of XML(DOCUMENT(XMLSCHEMA)) type	
X257		Persistent XML values of XML(CONTENT(XMLSCHEMA)) type	
X260		XML type: ELEMENT clause	
X261		XML type: NAMESP ACE without ELEMENT clause	
X263		XML type: NO NAMESP ACE with ELEMENT clause	
X264		XML type: schema location	
X271		XMLValidate: data-driven case	
X272		XMLValidate: ACCORD ING TO clause	
X273		XMLValidate: ELEMENT clause	
X274		XMLValidate: schema location	
X281		XMLValidate: with DOCUMENT option	
X282		XMLValidate with CONTENT option	
X283		XMLValidate with SEQUENCE option	
X284		XMLValidate NAMESP ACE without ELEMENT clause	
X286		XMLValidate: NO NAMESPACE with ELEMENT clause	
X300		XMLTable	
X301		XMLTable: derived column list option	
X302		XMLTable: ordinality column option	
X303		XMLTable: column default option	

Identifier	Package	Description	Comment
X304		XMLTable: passing a context item	
X305		XMLTable: initializing an XQuery variable	

附录 E. 发布说明

发布说明包含每个PostgreSQL发布中的重要变更，并将主要特性和迁移问题列在最前面。
发布说明不包含只影响少数用户的变更，也不包含仅限内部实现、因而对用户不可见的变更。
例如，优化器几乎在每次发布中都会得到改进，
但用户通常只会把这些改进感知为查询速度更快。

每个版本的一份完整变更列表可以通过查看该版本的 Git 日志获得。`pgsql-committers` 邮件列表[1]也记录了所有源代码变更。另有一个Web 界面[2]可显示特定文件的变更。

每个条目旁边标出的姓名表示该条目的主要开发者。当然，
所有变更都经历了社区讨论和补丁审查，因此每个条目实际上都是社区共同努力的成果。

E.1. 版本 9.1.24

发行日期
2016-10-27

本次发布包含对 9.1.23 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

这预计是 PostgreSQL 9.1.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.1.1. 迁移到版本 9.1.24

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.1.2. 变更

- 修复涉及 CTE 扫描的 EvalPlanQual 重新检查（Tom Lane）

  重新检查总是把 CTE 视为不返回任何行，通常导致无法更新最近已更新的行。

- 修复子查询中哈希聚合不当重复先前结果的问题（Andrew Gierth）

  判断能否复用先前计算的聚合状态值哈希表的测试，
  忽略了聚合参数表达式中出现外层查询引用的可能性。
  这种引用的值发生变化时本应重新计算哈希表，但实际没有。

- 修复 `VACUUM` 为截断表而等待排他表锁时的超时长度（Simon Riggs）

  该超时本应为 50 毫秒，但实际上只有 50 微秒，导致 `VACUUM` 比预期更容易放弃截断。将其设置为预期值。

- 移除对 `numeric_in()` 和 `numeric_recv()` 所接受值的人为限制（Tom Lane）

[1] http://archives.postgresql.org/pgsql-committers/
[2] http://git.postgresql.org/gitweb?p=postgresql.git;a=summary

我们允许数值达到存储格式的极限（超过 `1e100000`），因此 `numeric_in()` 拒绝科学计数法指数超过 1000 的做法相当无意义。同样，`numeric_recv()` 拒绝输入值超过 1000 位数字也很荒谬。

- 避免因未持有缓冲区锁测试元组可见性而导致的极低概率数据损坏（Thomas Munro、Peter Geoghegan、Tom Lane）
- 修复截断超过 1GB 的临时关系时文件描述符泄漏的问题（Andres Freund）
- 禁止在启用 `standby_mode` 的情况下启动独立后端（Michael Paquier）

  这不可能有任何用处，因为不会有 WAL 接收器进程来获取更多 WAL 数据；而且这可能导致未针对此种情况设计的代码出现不当行为。

- 不要在 libpq 中跨多个连接共享 SSL 上下文（Heikki Linnakangas）

  这导致了各种边界情况 bug，特别是当尝试对不同连接使用不同 SSL 参数时。

- 避免 libpq 中的边界情况内存泄漏（Tom Lane）

  报告的问题涉及 `PQreset()` 期间泄漏错误报告，但可能有相关情况。

- 使 ecpg 的 `--help` 和 `--version` 选项与其他可执行程序保持一致（Haribabu Kommi）
- 修复 `contrib/intarray/bench/bench.pl`，使其在给定 `-e` 选项时打印它执行的 `EXPLAIN` 的结果（Daniel Gustafsson）
- 防止已过时的动态时区缩写失效（Tom Lane）

  如果动态时区缩写与所引用时区中的任何条目都不匹配，就把它视为等效于该时区名称。这避免了当 IANA 从其时区数据库中移除缩写时出现意外失败——他们在 tzdata 2016f 版中就这样做过，而且将来可能再次这样做。后果不仅限于无法识别单个缩写；任何不匹配都会导致 `pg_timezone_abbrevs` 视图完全失败。

- 更新时区数据文件到 tzdata 版本 2016h，涉及巴勒斯坦和土耳其的夏令时法规变更，加上土耳其和俄罗斯一些地区的历史修正。南极洲、前苏联和斯里兰卡的一些时区改用数字缩写。

  IANA 时区数据库以前为所有时区提供文本缩写，有时编造在当地居民中几乎或完全没有使用的缩写。他们正在逆转这一政策，改为在没有证据表明实际使用英文缩写的区域使用数字 UTC 偏移。至少目前，PostgreSQL 将继续接受这种被移除的缩写用于时间戳输入。但它们不会显示在 `pg_timezone_names` 视图中，也不用于输出。

  在此更新中，`AMT` 不再显示为表示亚美尼亚时间。因此我们把 `Default` 缩写集改为把它解释为亚马逊时间，即 UTC-4 而不是 UTC+4。

E.2. 版本 9.1.23

发行日期
2016-08-11

本次发布包含对 9.1.22 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

PostgreSQL 社区将于 2016 年 9 月停止对 9.1.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.2.1. 迁移到版本 9.1.23

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.2.2. 变更

- 修复嵌套 `CASE-WHEN` 表达式可能被错误评估的问题（Heikki Linnakangas、Michael Paquier、Tom Lane）

  出现在另一个 `CASE` 的测试值子表达式中的 `CASE` 表达式，可能会混淆其自身的测试值是否为空。此外，内联实现 `CASE` 表达式所用相等操作符的 SQL 函数时，可能向该 SQL 函数体内 `CASE` 表达式中调用的函数传递错误的测试值。如果测试值的数据类型不同，可能导致崩溃；而且此类情况可被滥用以泄露服务器内存的部分内容。（CVE-2016-5423）

- 修复客户端程序对数据库名和角色名中特殊字符的处理（Noah Misch、Nathan Bossart、Michael Paquier）

  vacuumdb 及其他客户端程序中的许多地方，可能会被包含双引号或反斜杠的数据库名和角色名弄混。收紧引用规则以使其安全。同时确保当 conninfo 字符串被用作这些程序的数据库名参数时，全程都被正确地如此对待。

  修复 psql 的 `\connect` 和 `\password` 命令中对成对双引号的处理，使其与文档一致。

  在 psql 的 `\connect` 命令中引入新的 `-reuse-previous` 选项，以显式控制是否复用先前连接的连接参数。（若不使用该选项，则与以前一样，根据数据库名是否看起来像 conninfo 字符串来决定。）这使 pg_dumpall 脚本能够安全地处理包含特殊字符的数据库名。

  pg_dumpall 现在拒绝处理包含回车或换行的数据库名和角色名，因为在 Windows 上安全地引用这些字符似乎不切实际。将来我们可能会在服务器端拒绝此类名称，
  但尚未采取这一步骤。

  这些被视为安全修复，因为包含特殊字符的恶意构造的对象名，可能在超级用户下次执行 pg_dumpall 或其他例行维护操作时被用于以超级用户权限执行命令。（CVE-2016-5424）

- 修复对嵌套组合值应用 `IS NULL`/`IS NOT NULL` 时的边界错误行为（Andrew Gierth、Tom Lane）

  SQL 标准规定，对于全空值的行，`IS NULL` 应返回 TRUE（因此 `ROW(NULL,NULL) IS NULL` 得到 TRUE），但这并不意味着递归适用（因此 `ROW(NULL, ROW(NULL,NULL)) IS NULL` 得到 FALSE）。核心执行器对此处理正确，但某些规划器优化把该测试当作递归的（因而在两种情况下都产生 TRUE），而 `contrib/postgres_fdw` 可能产生类似错误行为的远程查询。

- 使 `inet` 和 `cidr` 数据类型正确拒绝包含过多冒号分隔字段的 IPv6 地址（Tom Lane）

- 防止 `close_ps()`（`point ## lseg` 操作符）在输入坐标为 NaN 时崩溃（Tom Lane）

  使其返回 NULL 而不是崩溃。

- 修复 `to_number()` 中几处一字节缓冲区过读（Peter Eisentraut）

在几种情况下，`to_number()` 函数会从输入字符串中多读取一个字符。如果输入恰好位于内存末尾附近，则有很小的崩溃可能。

- 避免在经过 `heap_update()` 的高开销路径中出现不安全的中间状态（Masahiko Sawada、Andres Freund）

  以前，这些情况会锁定目标元组（通过设置其 XMAX），但不为该动作记录 WAL，因此如果页面被刷出到磁盘，随后在元组更新完成之前发生数据库崩溃，就有数据完整性问题的风险。

- 避免在 `VACUUM` 期间消耗事务 ID（Alexander Korotkov）

  `VACUUM` 中的某些情况会不必要地导致向当前事务分配 XID。通常这可以忽略不计，但如果已接近 XID 回卷限制，在防回卷清理期间消耗更多 XID 就非常糟糕了。

- 避免在`VACUUM FREEZE`期间取消热备查询（Simon Riggs, Álvaro Herrera）

  在原本空闲的主库服务器上执行`VACUUM FREEZE`可能导致其备库服务器上的查询被不必要地取消。

- 当手动 `ANALYZE` 指定列列表时，不要重置表的 `changes_since_analyze` 计数器（Tom Lane）

  如果只分析部分列，就不应阻止其他列的例行自动分析。

- 修复 `ANALYZE` 对包含大量空值的唯一或近似唯一列的 `n_distinct` 高估（Tom Lane）

  空值可能被当作它们本身就是不同的值来计数，导致某些类型查询中规划器的严重误估。

- 防止自动清理为同一共享目录启动多个工作进程（Álvaro Herrera）

  通常这不是什么大问题，因为清理本来就不会花很长时间；但在目录严重膨胀的情况下，
  这可能导致除一个工作进程外的所有工作进程都在无用地等待，
  而不是对其他表做有用的工作。

- 修复 `contrib/btree_gin`，使其正确处理 `bigint` 的最小可能值（Peter Eisentraut）

- 让 libpq 能正确解码来自未来版本服务器的服务器版本号（Peter Eisentraut）

  计划在 9.6 之后的版本中把服务器版本号从三段式改为两段式。确保 `PQserverVersion()` 在此类情况下返回正确的值。

- 修复 ecpg 中 `unsigned long long` 数组元素的代码（Michael Meskes）

- 使 pg_basebackup 接受 `-Z 0` 表示不压缩（Fujii Masao）

- 恢复 `pg_ctl start -w` 的旧启发式超时（Tom Lane）

  9.1.20 采用的新方法在启用 `silent_mode` 时不工作，因此改回旧方法。

- 修复构建 AIX 共享库的 makefile 规则，使其在并行 make 下安全（Noah Misch）

- 修复 TAP 测试和 MSVC 脚本，使其在构建目录路径名包含空格时也能工作（Michael Paquier、Kyotaro Horiguchi）

- 使回归测试在丹麦语和威尔士语区域设置下安全（Jeff Janes、Tom Lane）

  修改了一些会触发这些区域设置特殊排序规则的测试数据。

- 更新我们的时区代码副本，以匹配 IANA 的 tzcode 2016c 版（Tom Lane）

  这是为了应对时区数据文件未来预期的变更。它还修复了处理特殊时区方面的一些边界 bug。

- 把时区数据文件更新到 tzdata 2016f 版，涉及克麦罗沃和新西伯利亚的夏令时法规变更，以及阿塞拜疆、白俄罗斯和摩洛哥的历史数据修正。

E.3. 版本 9.1.22

发行日期
2016-05-12

本次发布包含对 9.1.21 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

PostgreSQL 社区将于 2016 年 9 月停止对 9.1.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.3.1. 迁移到版本 9.1.22

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.3.2. 变更

- 在调用 OpenSSL 之前清空 OpenSSL 错误队列，而不是假设它已经为空；并确保调用之后也让它保持为空（Peter Geoghegan、Dave Vitek、Peter Eisentraut）

  此变更可防止在单个进程内存在多条使用 OpenSSL 的连接、且相关代码并非全部遵循相同的错误队列清空规则时出现问题。特别是当客户端应用并发使用 libpq 的 SSL 连接和 PHP、Python 或 Ruby 的 OpenSSL 封装库的 SSL 连接时，已有失败报告。如果扩展模块建立外向 SSL 连接，服务器内部也可能出现类似问题。

- 修复左连接右侧包含全连接时规划器出现 “failed to build any *N*-way joins” 错误的问题（Tom Lane）

- 修复 `to_timestamp()` 中 `TH`、`th` 和 `Y,YYY` 格式代码可能的错误行为（Tom Lane）

  这些代码可能越过输入字符串末尾，导致后续格式代码读取垃圾数据。

- 修复对这样的规则和视图的转储：*value operator* `ANY (`*array*`)` 结构的 *array* 参数是子 SELECT（Tom Lane）

- 使 pg_regress 在设置了 `PGCTLTIMEOUT` 环境变量时使用其中的启动超时（Tom Lane）

  这是为了与最近添加到 pg_ctl 的行为保持一致；它方便了在慢速机器上进行自动化测试。

- 修复 pg_upgrade 对仅包含一个操作符类的操作符族正确恢复扩展成员关系的问题（Tom Lane）

  在这种情况下，操作符族会被恢复到新数据库中，但不再被标记为扩展的一部分。
  这没有直接的坏影响，但会导致后续 pg_dump 运行产生在恢复时引发（无害）错误的输出。

- 把内部函数 `strtoi()` 重命名为 `strtoint()`，以避免与 NetBSD 库函数冲突（Thomas Munro）
- 修复 Windows 上 `bind()` 和 `listen()` 系统调用错误的报告（Tom Lane）
- 减少使用 Microsoft Visual Studio 构建时编译器输出的冗长程度（Christian Ullrich）
- 避免可能不安全地使用 Windows 的 `FormatMessage()` 函数（Christian Ullrich）

  在适当之处使用 `FORMAT_MESSAGE_IGNORE_INSERTS` 标志。目前不知道这里存在实际的 bug，但小心谨慎似乎是明智的。
- 把时区数据文件更新到 tzdata 2016d 版，涉及俄罗斯和委内瑞拉的夏令时法规变更。新增时区名称 `Europe/Kirov` 和 `Asia/Tomsk`，以反映这些地区现在与相邻地区有不同的时区历史。

E.4. 版本 9.1.21

发行日期
2016-03-31

本次发布包含对 9.1.20 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.4.1. 迁移到版本 9.1.21

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.4.2. 变更

- 修复索引化 `ROW()` 比较中对 NULL 索引条目处理不正确的问题（Tom Lane）

  使用诸如 `ROW(a, b) > ROW('x', 'y')` 的行比较进行索引搜索时，会在到达 `b` 列的 NULL 条目时停止，忽略了 `b` 列可能有与更靠后的 `a` 值关联的非 NULL 值这一事实。
- 避免因重命名文件前后未充分调用 `fsync()` 而导致的不太可能的数据丢失场景（Michael Paquier、Tomas Vondra、Andres Freund）
- 正确处理服务器启动期间 `pg_subtrans` 接近 XID 回卷的情况（Jeff Janes）
- 修复因多次尝试释放 `localeconv()` 输出字符串而导致的边界崩溃（Tom Lane）
- 修复 `ispell` 词典词缀文件的解析（Tom Lane）

  如果词缀文件包含任何在大小写转换期间字节长度会变化的字符（例如土耳其语 UTF8 区域设置中的 `I`），代码就可能出错。
- 避免使用 `sscanf()` 解析 `ispell` 词典文件（Artur Zakirov）

  这规避了 FreeBSD 衍生平台（包括 macOS）上的一个可移植性问题。
- 避免在旧版 Windows（7SP1/2008R2SP1 之前）上配合支持 AVX2 的 CPU 和使用 Visual Studio 2013 构建的 Postgres 时崩溃（Christian Ullrich）

这是对 Visual Studio 2013 运行时库中一个 bug 的变通方案，微软已表示不会在该版本中修复它。

- 修复 psql 的 Tab 补全逻辑，使其正确处理多字节字符（Kyotaro Horiguchi、Robert Haas）
- 修复 psql 对 `SECURITY LABEL` 的 Tab 补全（Tom Lane）

  在 `SECURITY LABEL` 之后按 TAB 可能导致崩溃或提供不合适的关键字。

- 使 pg_ctl 在命令行未指定时接受来自 `PGCTLTIMEOUT` 环境变量的等待超时（Noah Misch）

  这方便了较慢的 buildfarm 成员的测试，允许它们全局指定比正常更长的 postmaster 启动和关闭超时。

- 修复 pg_ctl 中对 Windows 服务状态的不正确测试（Manuel Mathar）

  上一轮小版本发布试图修复 pg_ctl，使其正确判断是否把日志消息发送到 Windows 事件日志，但把测试条件弄反了。

- 修复 pgbench 对 `-C` 与 `-M prepared` 选项组合的处理（Tom Lane）
- 在 PL/Perl 中，把空的 Postgres 数组正确转换为空的 Perl 数组（Alex Hunsaker）
- 使 PL/Python 能处理不是有效 Python 标识符的函数名（Jim Nasby）
- 修复 `contrib/pgstattuple` 的 `pgstatindex()` 函数返回统计信息中的多处错误（Tom Lane）
- 在 MSVC 构建中移除对 `psed` 的依赖，因为核心 Perl 已不再提供它（Michael Paquier、Andrew Dunstan）
- 把时区数据文件更新到 tzdata 2016c 版，涉及阿塞拜疆、智利、海地、巴勒斯坦和俄罗斯（阿尔泰、阿斯特拉罕、基洛夫、萨哈林、乌里扬诺夫斯克地区）的夏令时法规变更，以及立陶宛、摩尔多瓦和俄罗斯（加里宁格勒、萨马拉、伏尔加格勒）的历史数据修正。

E.5. 版本 9.1.20

发行日期
2016-02-11

本次发布包含对 9.1.19 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.5.1. 迁移到版本 9.1.20

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.5.2. 变更

- 修复正则表达式中的无限循环和缓冲区越界问题（Tom Lane）

方括号表达式中非常大的字符范围在某些情况下可能导致无限循环，
在另一些情况下可能导致内存覆写。（CVE-2016-0773）

- 如果postmaster.pid文件被移除则执行立即关闭（Tom Lane）

  postmaster 现在大约每分钟检查一次postmaster.pid是否仍然存在且仍包含自己的 PID。如果不是，它会执行立即关闭，就像收到SIGQUIT一样。此变更的主要目的是确保失败的 buildfarm 运行无需人工干预即可被清理；同时也可在 DBA 强行删除postmaster.pid并启动新 postmaster 时限制不良影响。

- 在SERIALIZABLE事务隔离模式下，插入期间的竞争条件可能导致漏掉串行化异常（Kevin Grittner, Thomas Munro）

- 修复对非日志关系执行ALTER TABLE ... SET TABLESPACE时未发出适当 WAL 记录的问题（Michael Paquier、Andres Freund）

  尽管关系的数据是非日志的，但移动操作本身必须被记录，
  否则在备库提升为主库后该关系将不可访问。

- 修复崩溃恢复结束时非日志关系可能被错误初始化的问题（Andres Freund, Michael Paquier）

- 修复ALTER COLUMN TYPE以正确重建继承的检查约束（Tom Lane）

- 修复REASSIGN OWNED以正确更改组合类型的所有权（Álvaro Herrera）

- 修复REASSIGN OWNED和ALTER OWNER在更改数据类型、外部数据包装器或外部服务器的所有者时正确更新已授权权限列表的问题（Bruce Momjian, Álvaro Herrera）

- 修复REASSIGN OWNED忽略外部用户映射而不是报错的问题（Álvaro Herrera）

- 当 GIN 索引的内部统计信息严重过时时，增加更多针对规划器错误代价估计的防护（Tom Lane）

- 使规划器能够处理由索引顾问插件建议的假设性 GIN 索引（Julien Rouhaud）

- 修复ROW()和VALUES()列表中整行 Var 的转储问题（Tom Lane）

- 修复numeric除法中可能出现的内部溢出（Dean Rasheed）

- 修复正则表达式先行约束中括号内限制的执行（Tom Lane）

  根据手册，先行约束不允许包含反向引用，且其中的括号始终被视为非捕获括号。
  但代码未能正确处理括号子表达式内的这些情况，会给出意外的结果。

- 将正则表达式转换为索引扫描边界时，
  可能从包含向前看约束的正则表达式生成错误边界（Tom Lane）

- 修复正则表达式编译器使其处理约束弧构成的环（Tom Lane）

  为 CVE-2007-4772 添加的代码既不完整——未能处理涉及多个状态的环——又不正确——可能导致断言失败（不过在非断言构建中似乎没有不良后果）。多状态环会使编译器一直运行，
  直到查询被取消或达到状态数过多的错误条件。

- 改进正则表达式编译器中的内存用量统计（Tom Lane）

这使代码在一些以前耗费不合理时间和内存量的情况下发出“regular expression is too complex”错误。

- 改进正则表达式编译器的性能（Tom Lane）
- 使 `log_line_prefix` 中的 `%h` 和 `%r` 转义对因 `log_connections` 而发出的消息也生效（Tom Lane）

  以前，`%h`/`%r` 要等到新会话发出“connection received”日志消息之后才开始生效；现在它们对该消息也生效。

- 在 Windows 上，确保不需要共享内存映射句柄的子进程关闭该句柄（Tom Lane、Amit Kapila）

  只要 `logging_collector` 处于开启状态，这一疏忽就会导致无法从崩溃中恢复。

- 修复 Windows 上非阻塞模式下可能无法检测套接字 EOF 的问题（Tom Lane）

  目前尚不完全清楚此问题是否会在 9.5 之前的分支中发生，但如果发生，其症状将是 walsender 进程无限等待而不是察觉到连接丢失。

- 避免 SSPI 身份验证期间泄漏令牌句柄（Christian Ullrich）
- 在 psql 中，确保当终端窗口大小改变时更新 libreadline 所认知的屏幕尺寸（Merlin Moncure）

  此前，如果窗口在查询输出期间被调整大小，libreadline 不会察觉，导致之后输入多行查询时出现奇怪的行为。

- 修复 psql 的 `\det` 命令，使其以与其他可能带模式限定模式的 `\d` 命令相同的方式解释模式参数（Reece Hart）
- 避免在以下情况下psql的`\c`命令可能崩溃：先前连接是通过 Unix 套接字建立的，而命令指定了新的主机名和相同的用户名（Tom Lane）
- 在 `pg_ctl start -w` 中，直接测试子进程状态而不是依赖启发式方法（Tom Lane、Michael Paquier）

  此前，pg_ctl 依赖一个假设：新的 postmaster 总会在五秒内创建 `postmaster.pid`。但在高负载系统上这可能失败，导致 pg_ctl 错误地报告 postmaster 启动失败。

  除 Windows 外，此变更还意味着紧接另一个此类命令之后执行的 `pg_ctl start -w` 现在会可靠地失败，而此前如果在第一条命令后两秒内执行则会报告成功。

- 在 `pg_ctl start -w` 中，不再尝试使用通配监听地址连接 postmaster（Kondo Yuta）

  在 Windows 上，如果 `listen_addresses` 被设置为 `0.0.0.0` 或 `::`，pg_ctl 将无法检测到 postmaster 的启动，因为它会试图原样使用该值作为要连接的地址，而这行不通。现在改为假定应分别使用 `127.0.0.1` 或 `::1`。

- 在 Windows 上的 pg_ctl 中，通过检查服务状态来决定输出发送到哪里，而不是检查标准输出是否为终端（Michael Paquier）
- 在 pg_dump 和 pg_basebackup 中，采用 GNU 约定来处理超过 8GB 的 tar 归档成员（Tom Lane）

`tar` 文件格式的 POSIX 标准不允许归档成员文件超过 8GB，但大多数现代 tar 实现支持解决此问题的扩展。采用这一扩展后，pg_dump 使用 `-Ft` 时不再在数据超过 8GB 的表上失败，pg_basebackup 也能处理大于 8GB 的文件。此外，还修复了一些可移植性问题，它们可能在某些平台上导致 4GB 到 8GB 之间的成员失败。这些问题有可能因备份文件不可读而造成不可恢复的数据丢失。

- 修复 pg_dump 处理扩展成员对象时的若干边界 bug（Tom Lane）
- 使 pg_dump 把视图的触发器标记为需要在其规则之后处理，以防止并行 pg_restore 期间可能出现的失败（Tom Lane）
- 确保在 pg_dump 中正确引用关系选项值（Kouhei Sutou、Tom Lane）

  不是简单标识符或数字的 reloption 值可能因 pg_dump 发出的 CREATE 语句中的语法错误而导致转储/重载失败。核心 PostgreSQL 目前支持的任何 reloption 都不存在此问题，但扩展可能允许导致该问题的 reloption。
- 修复pg_upgrade的文件复制代码在 Windows 上正确处理错误的问题（Bruce Momjian）
- 在 pgbench 中加入防护，防止计算脚本指定的除法或取模运算符期间的边界溢出情况（Fabien Coelho、Michael Paquier）
- 防止某些 PL/Java 参数被非超级用户设置（Noah Misch）

  此变更缓解了 PL/Java 的一个安全 bug（CVE-2016-0766），该 bug 在 PL/Java 中已通过把这些参数标记为仅超级用户可用来修复。为了保护那些更新 PostgreSQL 比 PL/Java 更频繁的站点，核心代码现在也识别这些参数。
- 改进libpq对内存耗尽情况的处理（Michael Paquier, Amit Kapila, Heikki Linnakangas）
- 修复ecpg生成的`typedef`语句中参数顺序错误的问题（Michael Meskes）
- 在ecpg的`PGTYPESnumeric_from_double()`中使用`%g`而非`%f`格式（Tom Lane）
- 修复 ecpg 提供的头文件，使其不包含从预处理指令行延续到下一行的注释（Michael Meskes）

  这样的注释会被 ecpg 拒绝。尚不清楚是否应该更改 ecpg 本身。
- 确保`contrib/pgcrypto`的`crypt()`函数可以被查询取消中断（Andreas Karlsson）
- 接受晚于 2.5.x 的flex版本（Tom Lane、Michael Paquier）

  由于 flex 2.6.0 已发布，需要调整构建脚本中的版本检查。
- 将我们的`missing`脚本安装到 PGXS 构建能找到的位置（Jim Nasby）

  这使得在缺少bison等构建工具的机器上进行 PGXS 构建时能有合理的行为。
- 确保 `dynloader.h` 包含在 MSVC 构建安装的头文件中（Bruce Momjian、Michael Paquier）
- 新增变体回归测试期望输出文件，以匹配当前libxml2的行为（Tom Lane）

  libxml2对 CVE-2015-7499 的修复导致它在某些过去会输出错误上下文报告的情况下不再输出。这似乎是一个缺陷，但我们可能不得不在一段时间内接受它，因此做了变通处理。

- 把时区数据文件更新到 tzdata 2016a 版，涉及开曼群岛、梅特拉卡特拉和外贝加尔边疆区的夏令时法规变更，以及巴基斯坦的历史数据修正。

E.6. 版本 9.1.19

发行日期
2015-10-08

本次发布包含对 9.1.18 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.6.1. 迁移到版本 9.1.19

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.6.2. 变更

- 修复`contrib/pgcrypto`以检测并报告过短的`crypt()`盐值（Josh Kupershmidt）

  某些无效的盐值参数会导致服务器崩溃或泄露少量服务器内存字节。
  我们尚未排除通过安排让泄露字节中出现机密信息实施攻击的可行性，
  但这种攻击似乎不太可能。（CVE-2015-5288）

- 修复属于外部子事务的门户（游标）失败后的子事务清理问题（Tom Lane、Michael Paquier）

  在外部子事务游标中执行的函数可能因引用在内部子事务中创建的关系而导致断言失败或崩溃。

- 修复关系插入关系缓存“init 文件”的问题（Tom Lane）

  最近几个小版本中一个补丁的疏忽导致`pg_trigger_tgrelid_tgname_index`被遗漏在 init 文件之外。后续会话检测到这一点后认定 init 文件已损坏并静默忽略它，导致会话启动时间显著退化。除修复该缺陷外，还加入了一些防护，使未来任何类似错误更明显。

- 避免向 SPI 查询结果插入大量元组时的 O(N^2) 行为（Neil Conway）

- 改进存在大量未读通知时`LISTEN`的启动时间（Matt Newell）

- 回移植 9.3 时代新增的逐资源所有者锁缓存（Jeff Janes）

  当 pg_dump 尝试转储大量表时，这能显著提升性能。

- 默认禁用 SSL 重新协商（Michael Paquier、Andres Freund）

  虽然使用 SSL 重新协商在理论上是个好主意，但实践中我们看到了太多缺陷，无论是底层 OpenSSL 库中还是我们对它的使用中。重新协商将在 9.5 及更高版本中完全移除。在较旧的分支中，只是把`ssl_renegotiation_limit`的默认值改为零（禁用）。

- 降低`*_freeze_max_age`参数的最小值（Andres Freund）

  这主要是为了使相关行为的测试不那么耗时，但对磁盘空间有限的安装也可能有价值。

- 将wal_buffers的最大值限制为 2GB 以避免服务器崩溃（Josh Berkus）
- 修复numeric值相乘时罕见的内部溢出（Dean Rasheed）
- 防御涉及记录类型、范围类型、json、jsonb、tsquery、ltxtquery和query_int的难以触发的栈溢出（Noah Misch）
- 修复日期时间输入中DOW和DOY的处理（Greg Stark）

  这些记号本不应在日期时间值中使用，但过去它们导致的是难懂的内部错误消息，而不是“无效的输入语法”。
- 在正则表达式匹配中增加更多查询取消检查（Tom Lane）
- 为正则表达式、SIMILAR TO和LIKE匹配增加递归深度保护（Tom Lane）

  合适的搜索模式加上较低的栈深度限制可能导致栈越界崩溃。
- 修复正则表达式执行中潜在的无限循环（Tom Lane）

  一个表面上能匹配零长度字符串、但实际上因反向引用而不匹配的搜索模式可能导致无限循环。
- 修复正则表达式编译中的内存不足失败（Andreas Seltenreich）
- 修复正则表达式执行期间低概率的内存泄漏（Tom Lane）
- 修复事务中止期间锁清理中罕见的内存不足失败（Tom Lane）
- 修复在使用较小work_mem设置的元组存储时出现“排序期间意外的内存不足情况”错误的问题（Tom Lane）
- 修复qsort中极低概率的栈越界（Tom Lane）
- 修复在使用较大work_mem设置的哈希连接中出现“无效的内存分配请求大小”失败的问题（Tomas Vondra, Tom Lane）
- 修复若干规划器缺陷（Tom Lane）

  这些错误可能导致给出错误结果的错误查询计划、启用断言的构建中的断言失败，或奇怪的规划器错误，例如“could not devise a query plan for the given query”、“could not find pathkey item to sort”、“plan should not reference subplan's variable”或“failed to assign all NestLoopParams to plan nodes”。感谢 Andreas Seltenreich 和 Piotr Stefaniak 通过模糊测试暴露了这些问题。
- 在所有受支持的分支中使用模糊路径代价平局判定规则（Tom Lane）

  此变更旨在避免当候选计划选项的估计代价实际相同时出现平台特定的行为。
- 确保在 postmaster 启动时移除备库提升触发文件（Michael Paquier、Fujii Masao）

  这可以防止这些文件出现在用于初始化新备库服务器的数据库备份中时发生不需要的提升。
- 在 postmaster 关闭期间，确保在移除postmaster.pid文件之前移除各套接字的锁文件并关闭监听套接字（Tom Lane）

  这可以避免外部脚本在pg_ctl stop一返回就尝试启动新 postmaster 时出现竞争条件失败。
- 修复 postmaster 对崩溃恢复期间启动进程崩溃的处理（Tom Lane）

如果在崩溃恢复周期中，启动进程在尚未恢复数据库一致性时崩溃，
我们会尝试启动一个新的启动进程，而这通常只会再次崩溃，导致无限循环。

- 当我们尝试向某个自动清理工作进程发送信号而它已经消失时，不再打印`WARNING`，
  并降低此类信号的日志详细程度（Tom Lane）
- 防止服务器时钟被大幅回拨时自动清理启动器睡眠时间过长（Álvaro Herrera）
- 确保 GIN 索引待插入列表的清理可被取消请求中断（Jeff Janes）
- 允许重用 GIN 索引中的全零页（Heikki Linnakangas）

  崩溃之后可能遗留这样的页。
- 修复一个差一错误，该错误会导致 subtrans/multixact 截断中出现本可避免的“表面回卷”警告（Thomas Munro）
- 修复PL/pgSQL错误上下文消息中`CONTINUE`和`MOVE`语句类型误报的问题（Pavel Stehule, Tom Lane）
- 修复PL/Perl以正确处理非ASCII错误消息文本（Alex Hunsaker）
- 修复PL/Python在返回`record`结果的字符串表示时崩溃的问题（Tom Lane）
- 修复PL/Tcl中一些未检查`malloc()`调用失败的位置（Michael Paquier, Álvaro Herrera）
- 在 `contrib/isn` 中，修复以 979 开头的 ISBN-13 号码的输出（Fabien Coelho）

  以 979（但不以 9790）开头的 EAN 被视为 ISBN，但它们必须以新的 13 位格式而非 10 位格式打印。
- 改进libpq对内存耗尽情况的处理（Michael Paquier, Heikki Linnakangas）
- 修复ecpg中的内存泄漏和缺失的内存不足检查（Michael Paquier）
- 修复 psql 中对数字输出的区域感知格式化代码（Tom Lane）

  `\pset numericlocale on` 所调用的格式化代码在一些不常见的情况下会出错，例如带指数但没有小数点的数字。它还可能破坏 `money` 数据类型已经区域化的输出。
- 防止在没有当前连接时psql的`\c`命令崩溃（Noah Misch）
- 修复pg_dump目录输出格式中默认zlib压缩级别的选择问题（Andrew Dunstan）
- 确保pg_dump以tar格式输出运行期间创建的临时文件不被全局可读（Michael Paquier）
- 修复pg_dump和pg_upgrade以支持`postgres`或`template1`
  数据库位于非默认表空间的情况（Marti Raudsepp, Bruce Momjian）
- 修复 pg_dump 从过于陈旧而没有特定权限类型的服务器转储时对对象权限的处理（Tom Lane）

  从 7.3 之前的服务器转储函数或过程语言时，pg_dump 会生成撤销所有者可授予权限、转而把所有权限授予 `PUBLIC` 的 `GRANT/REVOKE` 命令。由于涉及的权限只是 `USAGE` 和 `EXECUTE`，这不构成安全问题，但这确实是对旧系统行为令人惊讶的表示。已修复为在这些情况下保持默认权限状态不变。
- 修复pg_dump以转储 shell 类型（Tom Lane）

shell 类型（即尚未完全定义的类型）用处不大，但pg_dump仍应转储它们。

- 修复pg_dump及其他客户端程序中的若干小内存泄漏（Michael Paquier）
- 修复 PPC 硬件的自旋锁汇编代码使其与AIX本机汇编器兼容（Tom Lane）

  如果gcc被配置为使用本机汇编器（这正变得越来越普遍），则用gcc构建无法工作。
- 在AIX上，测试`-qlonglong`编译器选项，而不是直接假定它可以安全使用（Noah Misch）
- 在AIX上，使用`-Wl,-brtllib`链接选项以允许符号在运行时解析（Noah Misch）

  Perl 5.8.0 及更高版本依赖这一能力。
- 由于编译器缺陷，在使用 32 位xlc编译时避免使用内联函数（Noah Misch）
- 必要时使用`librt`提供`sched_yield()`，某些 Solaris 版本即是如此（Oskari Saarenmaa）
- 修复 Windows`install.bat`脚本以处理包含空格的目标目录名（Heikki Linnakangas）
- 使PostgreSQL版本号的数字形式（例如`90405`）以名为`VERSION_NUM`的变量方便地提供给扩展 Makefile 使用（Michael Paquier）
- 将时区数据文件更新到tzdata发行版 2015g，以反映开曼群岛、斐济、摩尔多瓦、摩洛哥、诺福克岛、朝鲜、土耳其和乌拉圭的夏令时法规变更。为加拿大北落基山地区新增了时区名`America/Fort_Nelson`。

E.7. 版本 9.1.18

发行日期
2015-06-12

本次发布包含对 9.1.17 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.7.1. 迁移到版本 9.1.18

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.7.2. 变更

- 修复罕见的未能使关系缓存 init 文件失效的问题（Tom Lane）

  在并发活动时序恰好不对的情况下，对系统目录执行`VACUUM FULL`可能无法更新用于避免新会话缓存加载工作的“init 文件”。这将导致后续会话完全无法访问该目录。这是一个非常古老的缺陷，但它极难触发，直到最近才出现可复现的案例。
- 避免传入会话与`CREATE/DROP DATABASE`之间的死锁（Tom Lane）

  在作为`DROP DATABASE`命令目标或`CREATE DATABASE`命令模板的数据库中启动的新会话，可能导致该命令等待五秒后失败，即使新会话本会在此之前退出。

E.8. 版本 9.1.17

发行日期
2015-06-04

本次发布包含对 9.1.16 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.8.1. 迁移到版本 9.1.17

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.16 的版本升级，请参见第 E.9 节。

E.8.2. 变更

- 避免崩溃重启期间对数据目录执行`fsync`时失败（Abhijit Menon-Sen、Tom Lane）

  在前几个小版本中，我们加入了一个补丁，在崩溃后对数据目录中的所有内容执行`fsync`。遗憾的是，它对任何错误情况的响应都是失败，从而阻止服务器启动，即使问题相当无害。例如，数据目录中一个不可写的文件在某些平台上会阻止重启；但把 SSL 证书文件设为服务器不可写是很常见的做法。我们修改了这一行为：权限失败被完全忽略，其他类型的失败会被记录但不会阻止继续。

- 移除configure中对在OpenBSD上链接线程版libpython的禁止性检查（Tom Lane）

  这一限制原本要防止的失败在当前的OpenBSD版本上似乎已不再是问题。

- 允许 libpq 使用高于 v1 的 TLS 协议版本（Noah Misch）

  长期以来，libpq 的代码只允许使用 TLS v1 这一种 SSL 协议。随着较新的 TLS 版本日益普及，现在允许它与服务器协商双方共同支持的最高 TLS 版本。（PostgreSQL 服务器早已支持这种协商，因此服务器端无需更改。）这是对 9.4.0 中已发布变更的回移植。

E.9. 版本 9.1.16

发行日期
2015-05-22

本次发布包含对 9.1.15 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.9.1. 迁移到版本 9.1.16

对于运行 9.1.X 的用户，不需要进行转储/恢复。

但是，如果你使用`contrib/citext`的`regexp_matches()`函数，请参见下面相关的变更日志条目。

不过，如果你是从早于 9.1.14 的版本升级，请参见第 E.11 节。

E.9.2. 变更

- 避免客户端恰在认证超时到期前断开时可能出现的崩溃（Benkocs Norbert Attila）

  如果超时中断在会话关闭序列进行到一半时触发，SSL 相关状态会被释放两次，通常导致崩溃并对其他会话造成服务拒绝。实验表明，未经认证的远程攻击者能够相当稳定地触发此缺陷，因此按安全问题处理。（CVE-2015-3165）

- 改进对系统调用失败的检测（Noah Misch）

  我们替换实现的`snprintf()`未检查底层系统库调用报告的错误；
  可能被遗漏的主要情况是内存耗尽。在最坏情况下，
  由于我们的代码假定缓冲区已被覆写而实际未覆写，这可能导致信息暴露。此外，
  还有少数地方对安全相关的其他系统库函数调用未检查失败。

  仍有可能某些对`*printf()`族函数的调用在恰逢内存耗尽错误时存在信息泄露风险。
  我们判断风险不大，但将继续在此领域进行分析。（CVE-2015-3166）

- 在`contrib/pgcrypto`中，统一将解密失败报告为“Wrong key or corrupt data”（Noah Misch）

  此前，某些使用错误密钥解密的情况可能报告其他错误消息文本。已经证明，
  这种错误报告的差异可能帮助攻击者从其他系统恢复密钥。虽然尚不清楚`pgcrypto`
  的具体行为是否同样可被利用，但使用统一消息来避免风险似乎更好。（CVE-2015-3167）

- 修复`contrib/citext`的`regexp_matches()`函数的错误声明（Tom Lane）

  这些函数应当像它们所包装的核心函数一样返回`setof text[]`；但它们被错误地声明为只返回`text[]`。这一错误有两个后果：第一，如果没有匹配，得到的是标量 null 结果，而正确的应该是空集（零行）。第二，`g`标志实际上被忽略了，
  因为即使有多个匹配也只会得到一个结果数组。

  虽然后一种行为显然是缺陷，但可能有应用程序依赖前一种行为；因此在PostgreSQL 9.5 之前默认不会更改函数声明。在 9.5 之前的分支中，旧行为存在于`citext`扩展的版本 1.0 中，而我们在版本 1.1 中提供了修正后的声明（不默认安装）。要在 9.5 之前的分支中采用修复，请在安装了`citext`的每个数据库中执行`ALTER EXTENSION citext UPDATE TO '1.1'`。（如果需要撤销，也可以“更新”回 1.0。）请注意，任一更新方向都需要删除并重新创建所有使用`citext`的`regexp_matches()`函数的视图或规则。

- 修复 HOT 更新后延迟排他约束检查不正确的问题（Tom Lane）

  如果同一事务中稍后对一条可能违反延迟排他约束的新行进行了 HOT 更新（即没有索引列发生变化，且该行可以存回同一表页），那么最终执行检查时将报告排他约束被违反，
  即使新行最初与之冲突的行已被删除。

- 防止反连接（NOT EXISTS 连接）与其他外连接之间的不当重排序（Tom Lane）

  已知规划器中的这一疏忽会导致“could not find RelOptInfo for given relids”错误，但有时错误的查询计划似乎有可能通过该一致性检查并导致静默错误的查询输出。

- 修复外连接计划节点中子表达式匹配不正确的问题（Tom Lane）

  此前，如果文本相同的非严格子表达式在外连接的上、下方都被使用，
  规划器可能试图重用在连接下方计算的值，但这是不正确的，因为对于未匹配的外部行，
  执行器会将该值强制置为 NULL。

- 修复 GEQO 规划器以应对其连接顺序启发式的失败（Tom Lane）

  已知这一疏忽会在涉及 `LATERAL` 的查询中导致 “failed to join all relations together” 错误，其他情况下也可能发生。

- 修复`max_prepared_transactions`过小时启动阶段可能出现的死锁（Heikki Linnakangas）

- 时间线切换后不再归档无用的预分配 WAL 文件（Heikki Linnakangas）

- 避免 “cannot GetMultiXactIdMembers() during recovery” 错误（Álvaro Herrera）

- 崩溃之后递归地对数据目录执行 `fsync()`（Abhijit Menon-Sen、Robert Haas）

  这确保了若随后不久再发生崩溃时数据仍保持一致。（要有问题的话，第二次崩溃必须是系统级崩溃，而不只是数据库崩溃。）

- 修复自动清理启动器在收到 SIGTERM 之后发生错误时可能无法关闭的问题（Álvaro Herrera）

- 处理 `LockBufferForCleanup()` 中的意外信号（Andres Freund）

  这一疏忽可能导致有关 “multiple backends attempting to wait for pincount 1” 的虚假错误。

- 就用户而言是只读的事务，在提交期间不再等待 WAL 刷写或同步复制（Andres Freund）

  此前，由于 HOT 页清理而写入了 WAL 的事务在提交时可能出现延迟，导致不良影响，例如所有同步副本都停机时会话在启动时卡住。还观察到使用同步复制时会话卡在追赶中断处理中；此修复也将解决该问题。

- 修复在临时表上操作哈希索引时崩溃的问题（Heikki Linnakangas）

- 修复哈希索引桶分裂期间（若其他进程正在并发修改该索引）可能失败的问题（Tom Lane）

- 在分析索引表达式期间检查中断（Jeff Janes）

  `ANALYZE` 会多次执行索引表达式；如果此类表达式中存在慢速函数，就希望能在此循环结束之前取消 `ANALYZE`。

- 确保在通过`SELECT FOR UPDATE`、`UPDATE`或`DELETE`锁定行之后发生`READ COMMITTED`复查时，外部表的`tableoid`被正确报告（Etsuro Fujita）

- 在外部服务器用户映射的对象描述字符串中加入目标服务器的名称（Álvaro Herrera）

- 建议在使用 Kerberos/GSSAPI/SSPI 认证时将 `include_realm` 设置为 1（Stephen Frost）

  不这样设置的话，来自不同领域的同名用户无法区分。目前这只是文档变更，但在 PostgreSQL 9.5 中它将成为默认设置。

- 移除将 IPv4 `pg_hba.conf` 条目匹配到 IPv4-in-IPv6 地址的代码（Tom Lane）

  这段应急代码于 2003 年添加，起因是当时的某些 Linux 内核会把 IPv4 连接报告为 IPv4-in-IPv6 地址。然而该逻辑在 9.0 中被意外破坏。此后没有任何现场投诉，说明它已不再需要。现在反而有报告称被破坏的代码在某些系统上导致崩溃，因此我们选择直接移除而不是修复它。（如果选择修复，将会使 IPv4 `pg_hba.conf` 条目的实际含义发生微妙且可能涉及安全性的变化，在次版本中这样做似乎并不妥当。）

- 在`IDENTIFY_SYSTEM`复制命令中报告 WAL 刷写位置而非插入位置（Heikki Linnakangas）

这避免了pg_receivexlog可能出现的启动失败。

- 在 Windows 上关闭服务期间，定期向服务控制管理器发送状态更新，以防止它过早终止服务；并确保pg_ctl会等待关闭完成（Krystian Bigaj）

- 降低使用libpq非阻塞模式时网络死锁的风险（Heikki Linnakangas）

  在发送大量数据时，定期排空输入缓冲区很重要，
  以防服务器已发送足够的响应数据导致其在输出上阻塞。（典型场景是服务器在`COPY FROM STDIN`期间发送一连串 NOTICE 消息。）这在普通阻塞模式下工作正常，但在非阻塞模式下则不然。我们修改了libpq，使其在可能时顺便排空输入，
  但要完全防御此问题需要应用程序配合：应用程序应同时监视套接字的读就绪和写就绪条件，
  并确保在读就绪时调用`PQconsumeInput()`。

- 修复ecpg中的数组处理（Michael Meskes）

- 修复psql以合理处理作为`\connect`第一个参数的 URI 和 conninfo 串（David Fetter、Andrew Dunstan, Álvaro Herrera）

  这一语法已被接受（但未写入文档）很长时间，
  但此前某些参数可能取自旧连接而非给定字符串，这被认为是不合适的。

- 抑制psql在某些平台上关于退出时未能写入`~/.psql_history`的错误抱怨（Tom Lane）

  这一不当行为源自针对非常老（2006 年之前）的libedit版本中一个缺陷的变通方案。我们通过移除该变通方案修复了问题，但这会使仍在使用此类libedit版本的用户出现类似失败。
  建议：升级该库，或改用libreadline。

- 修复pg_dump判定哪些转换是不应转储的系统提供转换的规则（Tom Lane）

- 修复pg_dump在`-Fd`下未遵循`-Z`压缩级别选项的问题（Michael Paquier）

- 使pg_dump在选择转储顺序时考虑扩展配置表之间的外键关系（Gilles Darold、Michael Paquier、Stephen Frost）

  这一疏忽可能导致生成的转储无法重新加载，因为外键约束会被瞬时违反。

- 修复对仅由`VALUES(...)`构成但带有列别名的视图的转储（Tom Lane）

- 在pg_upgrade中，强制新集簇使用时间线 1（Bruce Momjian）

  此变更可防止因错误抱怨缺少 WAL 历史文件而导致的升级失败。

- 在pg_upgrade中，继续之前先检查是否存在不当不可连接的数据库（Bruce Momjian）

- 在pg_upgrade中，在生成的`delete_old_cluster`脚本中对目录路径正确加引号（Bruce Momjian）

- 在pg_upgrade中正确保留数据库级冻结信息（Bruce Momjian）

  这一疏忽可能导致`postgres`和`template1`数据库中的表出现缺少 clog 文件的错误。

- 在 Windows 上以受限权限运行pg_upgrade和pg_resetxlog，使它们在由管理员运行时不会失败（Muhammad Asif Naeem）

- 改进initdb和pg_basebackup扫描目录时对`readdir()`失败的处理（Marco Nenciarini）

- 修复`contrib/intarray`中缓慢的排序算法（Tom Lane）
- 修复 Sparc V8 机器上的编译失败（Rob Rowan）
- 将时区数据文件更新到tzdata发行版 2015d，以反映埃及、蒙古和巴勒斯坦的夏令时法规变更，以及加拿大和智利的历史变更。同时采用 America/Adak 时区修订后的时区缩写（HST/HDT 而非 HAST/HADT）。

E.10. 版本 9.1.15

发行日期
2015-02-05

本次发布包含对 9.1.14 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.10.1. 迁移到版本 9.1.15

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.14 的版本升级，请参见第 E.11 节。

E.10.2. 变更

- 修复`to_char()`中的缓冲区越界（Bruce Momjian）

  当`to_char()`处理要求数字位数很多的数字格式模板时，
  PostgreSQL会读取越过缓冲区末尾。当处理恶意构造的时间戳格式模板时，
  PostgreSQL会写入越过缓冲区末尾。两种情况都可能导致服务器崩溃。
  我们尚未排除导致权限提升攻击的可能性，尽管这种攻击似乎不太可能。（CVE-2015-0241）

- 修复替换实现的`*printf()`函数中的缓冲区越界（Tom Lane）

  PostgreSQL包含`printf`及相关函数的替换实现。当代码格式化浮点数（转换说明符`e`、`E`、`f`、`F`、`g`或`G`）且请求的精度超过约 500 时，会发生栈缓冲区越界。这会导致服务器崩溃，我们尚未排除导致权限提升攻击的可能性。数据库用户可以通过`to_char()`SQL 函数触发这种缓冲区越界。虽然这是唯一受影响的核心PostgreSQL功能，但使用 printf 族函数的扩展模块也可能存在风险。

  此问题主要影响 Windows 上的PostgreSQL。PostgreSQL在系统实现足够完善时会使用系统实现，而其他现代平台上正是如此。（CVE-2015-0242）

- 修复`contrib/pgcrypto`中的缓冲区越界（Marko Tiikkaja、Noah Misch）

  `pgcrypto`
  模块中内存大小跟踪的错误导致了栈缓冲区越界和对未初始化内存内容的不当依赖。
  缓冲区越界情况可能导致服务器崩溃，我们尚未排除导致权限提升攻击的可能性。（CVE-2015-0243）

- 修复出错后前端/后端协议可能失去同步的问题（Heikki Linnakangas）

  如果服务器在从客户端读取协议消息中途发生任何错误，它可能失去同步，
  错误地把消息数据的一部分解释为新的协议消息。

能够在命令参数中提交恶意构造二进制数据的攻击者可能借此注入自己的 SQL 命令。语句超时和查询取消是触发此场景最可能的错误来源。
尤其容易受影响的是既使用超时又将任意用户构造数据作为二进制查询参数提交的应用程序。
禁用语句超时会降低但不能消除被利用的风险。感谢 Emil Lenngren 报告此问题。（CVE-2015-0244）

- 修复通过约束违反错误消息的信息泄露（Stephen Frost）

  某些服务器错误消息会显示违反约束（如唯一约束）的列的值。
  如果用户对该表的所有列没有`SELECT`权限，这可能意味着暴露用户不应看到的值。
  我们调整了代码，使值仅在来自 SQL 命令本身或用户可以查询到时才显示。（CVE-2014-8161）

- 锁定 Windows 上回归测试的临时安装（Noah Misch）

  使用 SSPI 认证，只允许启动测试套件的操作系统用户连接。这在 Windows 上关闭了此前已在其他平台上关闭的同一漏洞，即其他用户可能连接到测试 postmaster。（CVE-2014-0067）

- 避免在使用 `ALTER DATABASE SET TABLESPACE` 把数据库移到新表空间、随后不久又移回原表空间时可能出现的数据损坏（Tom Lane）

- 避免当事务内的 `ANALYZE` 被回滚时损坏表（Andres Freund、Tom Lane、Michael Paquier）

  如果失败的事务此前删除了表上的最后一个索引、规则或触发器，
  该表会处于损坏状态：相关的 `pg_class` 标志本应设置却未被设置。

- 确保在 `CREATE DATABASE` 或 `ALTER DATABASE SET TABLESPACE` 期间正确复制不记录日志的表（Pavan Deolasee、Andres Freund）

- 修复 `DROP` 的依赖搜索，使其正确处理表的列在其所属表之前被递归访问的情况（Petr Jelinek、Tom Lane）

  目前已知这种情况只在扩展同时创建一种数据类型和使用该数据类型的表时出现。
  有问题的代码可能会拒绝 `DROP EXTENSION`，除非指定 `CASCADE`，而这本不应该是必需的。

- 修复 EvalPlanQual 处理中使用已释放内存的问题（Tom Lane）

  在`READ COMMITTED`模式下，锁定或更新最近更新过的行的查询可能因此缺陷崩溃。

- 修复在子表上使用部分索引时 `SELECT FOR UPDATE` 的规划问题（Kyotaro Horiguchi）

  在 `READ COMMITTED` 模式下，`SELECT FOR UPDATE` 在重新检查最近更新的行以确认它是否仍满足查询的 `WHERE` 条件时，还必须重新检查部分索引的 `WHERE` 条件。如果索引属于继承子表，就会遗漏这一要求，因此有可能错误地返回已不再满足查询条件的行。

- 修复 `SELECT FOR UPDATE` 在极端情况下可能把同一行返回两次、并且可能漏掉其他行的 corner case（Tom Lane）

  在 `READ COMMITTED` 模式下，扫描继承树的 `SELECT FOR UPDATE` 可能错误地返回来自前一个子表的行，而不是本应从后一个子表返回的行。

- 拒绝 `FOREIGN KEY` 声明的被引用列列表中出现重复的列名（David Rowley）

  这一限制符合 SQL 标准。此前我们并未显式拒绝这种情况，但之后代码会以看起来很怪异的错误失败。

- 修复 `numeric` 值求大整数次幂时的缺陷（Tom Lane）

  之前的代码可能得到错误答案，或者在意识到结果必然溢出之前消耗过多的时间和内存。

- 在 `numeric_recv()` 中，按照值的 `dscale` 字段截去会被隐藏的小数位（Tom Lane）

  `numeric` 值的显示标度（`dscale`）绝不应小于非零小数位的位数；但显然至少存在一个有缺陷的客户端应用，它发送的二进制 `numeric` 值违反了这一点。这会导致奇怪的行为，因为多余的数字位虽然不会被打印出来，却会被算术运算考虑在内。
  风险最小的修复方法似乎是在接收时把这类“隐藏的”数字位截去，
  使值确实与它打印出来的一致。

- 拒绝超出范围的数字时区规范（Tom Lane）

  超过 +/- 168 小时（一周）的简单数字时区规范此前会被接受，但随后可能在某些操作中导致空指针解引用崩溃。如此大的 UTC 偏移没有实际使用场景，因此予以拒绝。

- 修复 `tsquery @> tsquery` 操作符中的缺陷（Heikki Linnakangas）

  两个不同的词如果 CRC 相同，就会被认为匹配。此外，如果第二个操作数的词条比第一个多，就会被假定不包含于第一个；这是错误的，因为它可能包含重复词条。

- 增强 ispell 词典对有问题的词缀文件的防御（Tom Lane）

- 允许词典（thesaurus）包含超过 64K 个短语（David Boutin）

  之前的代码在词典过大时会崩溃，因此这被视为一个可回移植的缺陷修复，而不是新特性。

- 修复`xpath()`中的命名空间处理（Ali Akbar）

  此前，如果命名空间声明附在输入`xml`值的祖先元素上而非被返回的特定元素上，`xpath()`调用产生的`xml`值就不会包含命名空间声明。我们将祖先声明传播下来，
  使结果在孤立考虑时也是正确的。

- 修复嵌套 append 关系（例如 `UNION ALL` 子查询中的继承表）的规划器问题（Tom Lane）

- 当 GiST 索引元组放不进一个页面时干净地报错，而不是陷入无限递归（Andrew Gierth）

- 把设有逐表 `cost_limit` 和/或 `cost_delay` 设置的表排除在自动清理的全局代价平衡规则之外（Álvaro Herrera）

  先前的行为实质上忽略了这些逐表设置，这并非本意。现在，
  拥有这类设置的表将按照这些设置进行清理，独立于其他自动清理工作进程的状况。
  这可能导致比以前更重的总 I/O 负载，因此应当重新审视这类设置是否合理。

- 避免在自动清理名义上被关闭时仍然进行全面的自动清理（Tom Lane）

  即使自动清理名义上被关闭，我们仍会启动自动清理工作进程，去清理有 XID 回卷风险的表。
  然而，这样的工作进程随后会继续清理目标数据库中所有满足常规自动清理阈值的表。
  这往好里说是相当出乎意料，往坏里说会延误对回卷威胁的响应。
  现已修复：如果自动清理被关闭，工作进程只执行防回卷清理，不做任何其他工作。

- 在崩溃恢复期间，
  确保不记录日志的关系被重写为空并在恢复被视为完成之前同步到磁盘（Abhijit Menon-Sen、Andres Freund）

  这可以避免数据库崩溃恢复之后不记录日志的关系中残留垃圾数据的情形。

- 修复热备查询与重放整页镜像之间的竞态条件（Heikki Linnakangas）

  这一失误可能导致热备中正在执行的查询出现瞬时错误。

- 修复若干恢复逻辑不当忽略 `COMMIT/ABORT PREPARED` WAL 记录的情况（Heikki Linnakangas）

  最值得注意的疏漏是 `recovery_target_xid` 无法用于在两阶段提交处停止。

- 避免为时间线历史文件创建不必要的 `.ready` 标记文件（Fujii Masao）

- 修复当使用空的预备语句且 `log_statement` 设置为 `mod` 或 `ddl` 时可能出现的空指针解引用（Fujii Masao）

- 将“pgstat wait timeout”警告消息改为 LOG 级别，并重新措辞使其更易理解（Tom Lane）

  这条消息最初被认为基本属于不可能发生的情况，但它在我们较慢的 buildfarm 成员上出现得足够频繁，成了干扰。现已将其降为 LOG 级别，并在措辞上多下了一点功夫：现在它的内容为“using stale statistics instead of current ones because stats collector is not responding”。

- 修复 SPARC 自旋锁实现，确保当 CPU 以非 TSO 一致性模式运行时（如某些非 Solaris 内核所做的那样）其正确性（Andres Freund）

- 当 macOS 的`setlocale()`在 postmaster 内启动多余的线程时发出警告（Noah Misch）

- 修复 `PQconnectdbParams()` 中重复 `dbname` 参数的处理（Alex Shulgin）

  如果第一个 `dbname` 出现处包含要展开的连接串或 URI，就会出现意外行为。

- 确保 libpq 在遇到意外的套接字 EOF 时报告合适的错误消息（Marko Tiikkaja、Tom Lane）

  取决于内核行为，当服务器意外关闭套接字时，libpq 可能返回空的错误字符串而不是有用的信息。

- 在 `PQreset()` 期间清除任何旧的错误消息（Heikki Linnakangas）

  如果反复调用 `PQreset()` 而连接始终无法重建，失败连接尝试的错误消息会在 `PGconn` 的错误字符串中不断累积。

- 正确处理 libpq 解析连接选项期间的内存不足情况（Alex Shulgin、Heikki Linnakangas）

- 修复 ecpg 版 `ParseDateTime()` 中的数组越界（Michael Paquier）

- 在 initdb 中，当指定的密码文件为空时给出更清晰的错误消息（Mats Erik Andersson）

- 修复 psql 的 `\s` 命令，使其与 libedit 良好配合，并新增分页器支持（Stepan Rutz、Tom Lane）

  使用 libedit 而非 readline 时，`\s` 会以一种相当难读的编码格式打印命令历史，而且在较新的 libedit 版本上可能完全失败。现已改为由我们自己打印历史而不是让库来做。一个令人愉快的副作用是，在合适的时候会使用分页器。

  此补丁还修复了一个在使用 libedit 保存命令历史时换行编码不一致的缺陷。旧版 psql 写出的多行历史条目在此补丁后可以干净地读取，但反过来未必可行，具体取决于所涉及的 libedit 版本。

- 改进 psql 特殊变量解析的一致性（Tom Lane）

允许 ECHO_HIDDEN 和 ON_ERROR_ROLLBACK 使用 on 和 off 的变体拼写（如 1/0）。对 COMP_KEYWORD_CASE、ECHO、ECHO_HIDDEN、HISTCONTROL、ON_ERROR_ROLLBACK 和 VERBOSITY 的无法识别的值发出警告。对所有这些变量的所有值进行不区分大小写的识别；此前大小写敏感与不敏感的行为混杂在一起。

- 修复 psql 扩展模式显示在使用 border = 3 且 linestyle = ascii 或 unicode 时的一致性问题（Stephen Frost）
- 改进当数据库中大量存在同一对对象之间的多条依赖路径时 pg_dump 的性能（Tom Lane）
- 修复并行恢复纯模式转储时可能出现的死锁（Robert Haas、Tom Lane）
- 修复 pg_dump --binary-upgrade 在零列组合类型上的核心转储（Rushabh Lathia）
- 防止由pg_basebackup -x/-X创建的 WAL 文件在备库提升后被再次归档（Andres Freund）
- 修复 contrib/citext 的从非打包升级脚本（Tom Lane）
- 修复 contrib/pageinspect 的 get_raw_page() 中的块号检查（Tom Lane）

  错误的检查逻辑可能阻止访问非主关系分支中的某些页面。
- 修复 contrib/pgcrypto 的 pgp_sym_decrypt()，使其在消息长度比 2 的幂小 6 时不再失败（Marko Tiikkaja）
- 修复 contrib/pg_test_fsync 中的文件描述符泄漏（Jeff Janes）

  这可能导致 Windows 上无法删除临时文件。
- 安全地处理 contrib/tablefunc 的 connectby() 中的意外查询结果，尤其是 NULL（Michael Paquier）

  此前 connectby() 遇到 NULL 键值时会崩溃。现在它会打印该行但不再进一步递归。
- 避免 contrib/xml2 的 xslt_process() 中可能出现的崩溃（Mark Simonetti）

  libxslt 似乎对资源释放顺序存在未记载的依赖；重新排列我们的调用以避免崩溃。
- 为一些 contrib I/O 函数标记正确的易变性属性（Tom Lane）

  先前过于保守的标记在正常使用中无关紧要，
  但可能导致优化问题或拒绝有效的索引表达式定义。由于影响不大，
  我们只是调整了扩展模块脚本中的函数定义，未更改版本号。
- 大量清理 Coverity 静态代码分析器的警告（Andres Freund、Tatsuo Ishii、Marko Kreen、Tom Lane、Michael Paquier）

  这些变更大多只是表面性的，但在某些情况下修复了极端情况的缺陷，
  例如内存不足失败后崩溃而不是正确报告错误。据信它们都不构成安全问题。
- 在构建期间检测不兼容的 OpenLDAP 版本（Noah Misch）

  对于 2.4.24 到 2.4.31（含）之间的 OpenLDAP 版本，PostgreSQL 后端进程可能在退出时崩溃。现在会根据编译时的 OpenLDAP 版本号在 configure 期间发出警告，并在 contrib/dblink 回归测试中测试该崩溃场景。
- 在非 MSVC 的 Windows 构建中，确保安装的 libpq.dll 带有执行权限（Noah Misch）

- 使 pg_regress 成功退出时删除其创建的任何临时安装（Tom Lane）

  这使 `make check-world` 期间的磁盘空间用量大幅减少，因为该序列会创建大量临时安装。

- 支持随时间改变 UTC 偏移的时区缩写（Tom Lane）

  此前，PostgreSQL 假定时区缩写（如 `EST`）关联的 UTC 偏移在任何特定区域设置的用法中永不改变。然而这一假设在现实世界中并不成立，因此引入了让区域缩写表示有时会变化的 UTC 偏移的能力。更新了时区缩写定义文件，在自 1970 年以来缩写的 UTC 偏移发生过变化的时区区域设置中使用此功能（依据 IANA 时区数据库）。在此类时区中，PostgreSQL 现在会根据给定日期为缩写关联正确的 UTC 偏移。

- 更新时区缩写列表（Tom Lane）

  把 CST（中国标准时间）加入列表。移除 ADT 作为“阿拉伯夏令时”的引用——这个缩写自 2007 年起已不再使用；因此，声称它与“大西洋夏令时”冲突似乎没有特别的帮助。修正 CKT（库克群岛）、FJT 和 FJST（斐济）完全错误的 GMT 偏移；此前我们甚至把它们放到了日期变更线的错误一侧。

- 把时区数据文件更新为 tzdata 版本 2015a。

  IANA 时区数据库为所有澳大利亚时区采用了 `AxST`/`AxDT` 形式的缩写，以反映他们认为的当地当前多数用法。这些名称与其他地方的用法不冲突（唯一例外是表示 Acre 夏令时的 ACST，该用法自 1994 年起已废弃）。因此，我们把这些名称采纳进“默认”时区缩写集。“澳大利亚”缩写集现在只包含 CST、EAST、EST、SAST、SAT 和 WST，这些被认为主要是历史用法。注意，在“默认”缩写集中，SAST 也已改为南非标准时间。

  此外，新增时区缩写 SRET（Asia/Srednekolymsk）和 XJT（Asia/Urumqi），并对西萨摩亚使用 WSST/WSDT。同时，智利、墨西哥、特克斯和凯科斯群岛（America/Grand_Turk）以及斐济的夏令时法规有变更。还为巴布亚新几内亚部分地区新增了时区 Pacific/Bougainville。另外还有大量针对历史（1970 年以前）时区数据的修正。

E.11. 版本 9.1.14

发行日期
2014-07-24

本次发布包含对 9.1.13 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.11.1. 迁移到版本 9.1.14

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了某些 GiST 索引的索引损坏问题。请参见下面第一条变更日志，以确认你的安装是否受影响以及应采取的步骤。

不过，如果你是从早于 9.1.11 的版本升级，请参见第 E.14 节。

E.11.2. 变更

- 正确初始化 `contrib/btree_gist` 在 `bit` 列上索引的填充字节（Heikki Linnakangas）

这一错误可能导致本应比较相等的值不被视为相等，从而产生错误的查询结果。在 `bit` 或 `bit varying` 列上建有 GiST 索引的用户，在安装此更新后应对这些索引执行 `REINDEX`。

- 在删除 GIN 列表页时防止残缺页（torn page）（Heikki Linnakangas）

  此修复可防止在页面更新写入磁盘期间发生系统崩溃时可能出现的索引损坏。

- 在从 WAL 重放更新时不清除 GiST 索引页的右链接（Heikki Linnakangas）

  这一错误可能导致热备中执行的 GiST 索引扫描出现瞬时错误的结果。

- 修复hot_standby_feedback被动态关闭时反馈状态的处理（Simon Riggs）

- 修复嵌套调用 `ReceiveSharedInvalidMessages` 期间缓存失效可能不正确的问题（Andres Freund）

- 修复对从带继承子表的表读取的子查询执行 `UNION ALL` 时规划器报 “could not find pathkey item to sort” 失败的问题（Tom Lane）

- 当子查询的目标列表中存在集合返回函数时，不再假定其输出是唯一的（David Rowley）

  这一疏忽可能导致诸如 `WHERE x IN (SELECT y, generate_series(1,10) FROM t GROUP BY y)` 之类的构造被错误优化。

- 修复未能对结构化类型的组合元素中的字段执行解 TOAST 的错误（Tom Lane）

  这纠正了 TOAST 指针可能未经解引用就被复制到其他表的情况。如果原始数据随后被删除，当这个已悬空的指针被使用时，会导致诸如 “missing chunk number 0 for toast value ...” 之类的错误。

- 修复对 Append 计划节点输出的整行引用导致 “record type has not been registered” 失败的问题（Tom Lane）

- 修复在回卷游标时调用用户定义函数可能崩溃的问题（Tom Lane）

- 修复为 `FROM` 中的函数求值参数时查询级内存泄漏的问题（Tom Lane）

- 修复正则表达式处理中的会话级内存泄漏（Tom Lane、Arthur O'Dwyer、Greg Stark）

- 修复 `hungarian.stop` 中的数据编码错误（Tom Lane）

- 防止在default_with_oids为真时创建带 OIDS 的外部表（Etsuro Fujita）

- 修复对在当前事务中插入、随后被一个现已回滚的子事务删除的行的活跃性检查（Andres Freund）

  如果同一事务中稍后执行了 `CREATE INDEX` 或 `CLUSTER`，这可能导致问题（至少是虚假警告，最坏是无限循环）。

- 在 `PREPARE TRANSACTION` 期间清除 `pg_stat_activity.xact_start`（Andres Freund）

  在 `PREPARE` 之后，发起会话已不在事务中，因此不应继续显示事务开始时间。

- 修复 `REASSIGN OWNED` 对文本搜索对象失败的问题（Álvaro Herrera）

- 在 postmaster 启动期间阻塞信号（Tom Lane）

  这确保了 postmaster 在例如启动过程中收到 SIGINT 时能够正确清理自身。

- 修复处理 `pg_hba.conf` 中指定主机名而非 IP 地址的条目时的客户端主机名查找（Tom Lane）

  确保反向 DNS 查找失败会被报告，而不是只是静默地不匹配这些条目。同时确保每个连接只做一次反向 DNS 查找尝试，而不是每个主机名条目一次——如果查找尝试失败，此前发生的就是后者。

- 保护 `make check` 期间启动的临时 postmaster 的 Unix 域套接字（Noah Misch）

  正如我们此前在 CVE-2014-0067 中指出的，任何能够访问套接字文件的本地用户都可以以服务器的引导超级用户身份连接，进而以运行测试的操作系统用户身份执行任意代码。
  此变更通过把服务器的套接字放到 `/tmp` 下一个临时的、模式为 0700 的子目录中来防御这一风险。但在不支持 Unix 套接字的平台（尤其是 Windows）上风险仍然存在，因为那时临时 postmaster 必须接受本地 TCP 连接。

  此变更的一个有用副作用是简化了覆盖 `DEFAULT_PGSOCKET_DIR` 的构建中的 `make check` 测试。常用的非默认值（如 `/var/run/postgresql`）往往不允许构建用户写入，此前需要一些变通办法，现在不再需要了。

- 修复表空间创建 WAL 重放在 Windows 上无法工作的问题（MauMau）

- 修复 Windows 上套接字创建失败的检测（Bruce Momjian）

- 在 Windows 上，允许新会话从配置文件获取 PGC_BACKEND 参数（例如log_connections）的值（Amit Kapila）

  此前，如果启动后在文件中更改了这类参数，更改不会生效。

- 在 Windows 上正确地为可执行文件路径名加引号（Nikhil Deshpande）

  如果安装路径同时包含空格和 `@` 符号，这一疏忽可能导致 initdb 和 pg_upgrade 在 Windows 上失败。

- 修复 macOS 上 libpython 的链接（Tom Lane）

  我们先前使用的方法在 Xcode 5.0 及之后版本提供的 Python 库上可能失败。

- 当服务器持续以快于客户端吸收速度的速率发送数据时，避免 libpq 中的缓冲区膨胀（Shin-ichi Morita、Tom Lane）

  libpq 可能被迫不断增大其输入缓冲区，直到内存耗尽（这会被误导性地报告为“lost synchronization with server”）。在正常情况下，数据持续以快于 `recv()` 循环吸收速度的速率传输是相当牵强的，但在客户端因调度器约束被人为减速时曾观察到这种情况。

- 确保 libpq 中的 LDAP 查找尝试按预期超时（Laurenz Albe）

- 修复 ecpg 在 `char *` 数组作为返回多行的 FETCH 语句目标时的行为，以及其他一些数组处理修复（Ashutosh Bapat）

- 修复 pg_restore 对旧式大对象注释的处理（Tom Lane）

  如果归档文件由 9.0 之前版本的 pg_dump 生成且包含较多大对象注释，直接恢复到数据库通常会失败。

- 在 `contrib/pgcrypto` 函数中，确保返回前清除栈变量中的敏感信息（Marko Kreen）
- 在 `contrib/uuid-ossp` 中，跨调用缓存 OSSP UUID 库的状态（Tom Lane）

  这提高了 UUID 生成的效率，并在有 `/dev/urandom` 的平台上减少了从中抽取的熵量。
- 把时区数据文件更新为 tzdata 版本 2014e for DST law changes in Crimea, Egypt, and Morocco。

E.12. 版本 9.1.13

发行日期
2014-03-20

本次发布包含对 9.1.12 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.12.1. 迁移到版本 9.1.13

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.11 的版本升级，请参见第 E.14 节。

E.12.2. 变更

- 无条件恢复 GIN 元页以避免残缺页风险（Heikki Linnakangas）

  虽然这一疏忽理论上可能导致索引损坏，但实际上不太可能造成任何问题，因为 GIN 元页的活动部分小于标准的 512 字节磁盘扇区。
- 避免在接收 `NOTIFY` 消息期间检查事务提交状态时的竞态条件（Marko Tiikkaja）

  这防止了这样一种场景：足够快的客户端可能在通知者所做的数据库更新对接收者可见之前就响应通知。
- 允许正则表达式操作符被查询取消请求提前终止（Tom Lane）

  这防止了病态正则表达式长时间不可中断地锁死服务器进程的场景。
- 移除试图允许 `OVERLAPS` 接受单元素行参数的错误代码（Joshua Yanovski）

  这段代码从未正确工作过，而且该情况既未被 SQL 标准规定也未被记录，因此移除它似乎比修复它更好。
- 避免在反解析规则或视图时获取超出 `AccessShareLock` 的锁（Dean Rasheed）

  这一疏忽导致 pg_dump 对规则中作为 `INSERT`/`UPDATE`/`DELETE` 命令目标的表意外获取了 `RowExclusiveLock` 锁。虽然通常无害，但它可能干扰试图在这些表上获取例如 `ShareLock` 的并发事务。
- 改进规划期间索引端点探测的性能（Tom Lane）

此变更修复了当索引末尾存在大量尚未提交的行时的显著性能问题，
这对于基于顺序赋值（如时间戳或序列生成的标识符）的索引是常见情况。

- 修复客户端为 pg_receivexlog 时 walsender 无法干净关闭的问题（Fujii Masao）
- 修复崩溃后是否可以立即允许热备连接的测试（Heikki Linnakangas）
- 在报告非 `ERROR` 消息期间防止中断（Tom Lane）

  这可以防止因递归进入 `syslog()` 而导致的罕见服务器进程冻结，或许还能防止其他相关问题。
- 修复 PL/Perl 返回组合结果（包括多 OUT 参数情况）时的内存泄漏（Alex Hunsaker）
- 防止在较新的 Windows 版本上间歇性出现 “could not reserve shared memory region” 失败（MauMau）
- 把时区数据文件更新为 tzdata 版本 2014a for DST law changes in Fiji and Turkey, plus historical changes in Israel and Ukraine。

E.13. 版本 9.1.12

发行日期
2014-02-20

本次发布包含对 9.1.11 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.13.1. 迁移到版本 9.1.12

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.11 的版本升级，请参见第 E.14 节。

E.13.2. 变更

- 加固 `GRANT ... WITH ADMIN OPTION` 的限制（Noah Misch）

  不带 `ADMIN OPTION` 授予角色本应阻止被授予者向该角色添加或移除成员，但这一限制很容易通过先执行 `SET ROLE` 绕过。其安全影响主要在于角色成员可以撤销其他人的访问权限，这违背了授权者的意愿。未经批准添加角色成员的影响相对较小，
  因为不配合的角色成员本来就可以通过创建视图或 `SECURITY DEFINER` 函数把自己的大部分权限提供给他人。（CVE-2014-0060）
- 防止通过手工调用 PL 验证器函数进行权限提升（Andres Freund）

  PL 验证器函数的主要角色是在 `CREATE FUNCTION` 期间被隐式调用，但它们也是普通 SQL 函数，用户可以显式调用。对实际以其他语言编写的函数调用验证器此前未被检查，可能被用于权限提升。修复方法是在每个验证器函数中加入对权限检查函数的调用。
  非核心过程语言如果自带验证器函数，也需要做同样的修改。（CVE-2014-0061）
- 避免在表和索引 DDL 期间进行多次名称查找（Robert Haas、Andres Freund）

如果由于并发活动导致名称查找得出不同结论，我们可能会对不同的表执行 DDL 的不同部分。至少在 `CREATE INDEX` 的情况下，这可被用来让权限检查针对与索引创建不同的表执行，从而造成权限提升攻击。（CVE-2014-0062）

- 防止长日期时间字符串造成缓冲区越界（Noah Misch）

  `MAXDATELEN` 常量对于 `interval` 类型可能的最长值来说太小，导致 `interval_out()` 中可能出现缓冲区越界。虽然日期时间输入函数在避免缓冲区越界方面更谨慎，但该限制仍短到导致它们拒绝某些有效输入，例如包含很长时区名的输入。ecpg 库存在这些漏洞以及它自身的一些漏洞。（CVE-2014-0063）

- 防止大小计算中的整数溢出导致缓冲区越界（Noah Misch、Heikki Linnakangas）

  若干函数（大多是类型输入函数）在计算分配大小时未检查溢出。如果发生溢出，就会分配过小的缓冲区并被写越界。（CVE-2014-0064）

- 防止固定大小缓冲区越界（Peter Eisentraut、Jozef Mlich）

  使用 `strlcpy()` 及相关函数，以明确保证固定大小缓冲区不会被越界写入。与前面几项不同，这些情况是否真的构成现实问题尚不清楚，因为在大多数情况下，输入字符串的大小似乎已有先前约束。尽管如此，消除所有此类 Coverity 警告似乎是明智的。（CVE-2014-0065）

- 避免在 `crypt()` 返回 NULL 时崩溃（Honza Horak、Bruce Momjian）

  `crypt()` 可能返回 NULL 的场景相对较少，但如果发生，`contrib/chkpass` 就会崩溃。一个可能出问题的实际情况是 libc 被配置为拒绝执行未获批准的哈希算法（例如“FIPS 模式”）。（CVE-2014-0066）

- 在回归测试说明中记录 `make check` 的风险（Noah Misch、Tom Lane）

  由于 `make check` 启动的临时服务器使用“trust”认证，同一台机器上的其他用户可以以数据库超级用户身份连接它，进而可能利用启动测试的操作系统用户的权限。未来版本可能会在测试流程中加入防止此风险的变更，但需要先进行一些公开讨论。因此目前只是警告：当同一台机器上存在不受信任的用户时，不要使用 `make check`。（CVE-2014-0067）

- 修复当关系的某些段不是完整大小时 WAL 记录可能被错误重放的问题（Greg Stark、Tom Lane）

  WAL 更新可能被应用到错误的页面，甚至可能偏后很多页。除了损坏数据外，已观察到这一错误导致备库相比主库显著“膨胀”，因为更新被应用到了远超文件末尾应在位置的地方。这种失败模式在崩溃恢复期间似乎不构成重大风险，只在初始同步从快速变化的主库基础备份创建的备库时出现。

- 修复判断恢复何时达到一致性的缺陷（Tomonari Katsumata、Heikki Linnakangas）

  某些情况下，WAL 重放会错误地在重放开始时就断定数据库已一致，从而可能在数据库真正一致之前允许热备查询。也可能出现其他症状，例如“PANIC: WAL contains references to invalid pages”。

- 修复在热备模式下重放 `VACUUM` 操作时对 btree 索引页的不当锁定（Andres Freund、Heikki Linnakangas、Tom Lane）

  这一错误可能导致“PANIC: WAL contains references to invalid pages”失败。

- 确保对 GIN 索引非叶页面的插入在合适时写入整页 WAL 记录（Heikki Linnakangas）

  先前的代码在系统崩溃期间发生部分页写入时存在索引损坏风险。

- 当 `pause_at_recovery_target` 和 `recovery_target_inclusive` 都被设置时，确保目标记录在暂停之前而不是之后被应用（Heikki Linnakangas）

- 修复服务器进程退出期间的竞态条件（Robert Haas）

  确保信号处理程序不会在进程的 `MyProc` 指针不再有效后尝试使用它。

- 修复 walsender 关闭逻辑和 walreceiver SIGHUP 信号处理程序中的竞态条件（Tom Lane）

- 修复错误报告逻辑中对 `errno` 的不安全引用（Christian Kruse）

  这通常会导致奇怪的行为，例如 `HINT` 字段缺失或不恰当。

- 修复服务器启动期间过早使用 `ereport()` 可能导致的崩溃（Tom Lane）

  我们在实际中见到的主要情况是：如果服务器在一个它没有读取权限的目录中被启动，就会崩溃。

- 在 OpenSSL 套接字写函数中正确清除重试标志（Alexander Kukushkin）

  这一遗漏可能导致 SSL 加密连接意外断开后服务器锁死。

- 修复对包含转义的 Unicode 标识符（`U&"..."` 语法）的长度检查（Tom Lane）

  如果标识符的转义形式过长，就会为这类标识符打印虚假的截断警告，但标识符在反转义之后实际上并不需要截断。

- 允许作为类型名的关键字用于角色列表（Stephen Frost）

  先前的补丁允许这类关键字在角色标识符等位置不加引号使用；但漏掉了允许角色标识符列表的情况，例如 `DROP ROLE`。

- 修复 `EXISTS(SELECT * FROM zero_column_table)` 的解析器崩溃（Tom Lane）

- 修复嵌套子选择（例如 `WHERE (... x IN (SELECT ...) ...) IN (SELECT ...)`）因无效计划而可能崩溃的问题（Tom Lane）

- 确保即使表列中的所有值都“过宽”，`ANALYZE` 也会为该列创建统计信息（Tom Lane）

  `ANALYZE` 有意在直方图和最常用值计算中省略非常宽的值，但在所有采样条目都过宽的情况下没有做出合理的处理。

- 在 `ALTER TABLE ... SET TABLESPACE` 中，允许不经权限检查就使用数据库的默认表空间（Stephen Frost）

  `CREATE TABLE` 一直允许这种用法，但 `ALTER TABLE` 没有照做。

- 修复当 `CASE` 的某些分支返回集合而另一些不返回时的“cannot accept a set”错误（Tom Lane）

- 修复 pgstat 函数中对全零客户端地址的检查（Kevin Grittner）

- 修复文本搜索解析器可能错误分类多字节字符的问题（Tom Lane）

在 C 区域设置下使用多字节编码时，非 ASCII 字符可能被错误分类。在 Cygwin 上，非 C 区域设置也可能失败。

- 修复 `plainto_tsquery()` 中可能的不当行为（Heikki Linnakangas）

  在复制重叠内存区域时使用 `memmove()` 而不是 `memcpy()`。尚无实际造成问题的现场报告，但这无疑是有风险的。

- 修复 `pg_start_backup()` 和 `pg_stop_backup()` 中权限检查的位置（Andres Freund、Magnus Hagander）

  先前的代码可能在不该访问系统目录时尝试访问。

- 在区域检查方面接受 `SHIFT_JIS` 作为编码名（Tatsuo Ishii）

- 修复 `PQhost()` 在 Windows 上的不当行为（Fujii Masao）

  如果未指定主机，它应返回 `localhost`。

- 改进 libpq 和 psql 对 `COPY TO STDOUT/FROM STDIN` 期间失败的错误处理（Tom Lane）

  这尤其修复了 9.2 及更高版本中，如果服务器连接在 `COPY FROM STDIN` 期间丢失可能发生的无限循环。该场景的变体在旧版本或其他客户端应用中也可能出现。

- 修复 pg_basebackup 详细模式下文件名可能打印错误的问题（Magnus Hagander）

- 避免在基础备份中把 PGDATA 内的表空间包含两次（Dimitri Fontaine、Magnus Hagander）

- 修复 ecpg 中未对齐的描述符（MauMau）

- 在 ecpg 中正确处理连接参数中缺少主机名的情况（Michael Meskes）

- 修复 `contrib/dblink` 连接启动中的性能退化（Joe Conway）

  在客户端和服务器编码一致时避免一次不必要的往返。

- 修复 `contrib/isn` 中 ISMN 值校验位计算错误的问题（Fabien Coelho）

- 确保仅安装客户端代码的流程按文档说明工作（Peter Eisentraut）

- 在 Mingw 和 Cygwin 构建中，把 libpq DLL 安装到 `bin` 目录（Andrew Dunstan）

  这与 MSVC 构建长期以来的做法一致。它应该能解决 psql 等程序因找不到 DLL 而无法启动的问题。

- 在 Cygwin 构建中避免使用已弃用的 `dllwrap` 工具（Marco Atzeri）

- 不再生成纯文本的 `HISTORY` 和 `src/test/regress/README` 文件（Tom Lane）

  这些文本文件与主文档的 HTML 和 PDF 格式重复。维护它们的麻烦远远超过纯文本格式可能的读者。发行版压缩包仍会包含这些名字的文件，但它们只是提示读者查阅主文档的存根。纯文本 `INSTALL` 文件仍将维护，因为它可以说还有使用场景。

- 把时区数据文件更新为 tzdata 版本 2013i for DST law changes in Jordan and historical changes in Cuba。

  此外，时区 `Asia/Riyadh87`、`Asia/Riyadh88` 和 `Asia/Riyadh89` 已被移除，因为 IANA 不再维护它们，而且它们从未代表过实际的民用计时实践。

E.14. 版本 9.1.11

发行日期
2013-12-05

本次发布包含对 9.1.10 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.14.1. 迁移到版本 9.1.11

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了若干潜在的数据损坏问题。请参见下面前三条变更日志，以确认你的安装是否受影响以及可以采取什么措施。

不过，如果你是从早于 9.1.9 的版本升级，请参见第 E.16 节。

E.14.2. 变更

- 修复 `VACUUM` 判断是否可以更新 `relfrozenxid` 的测试（Andres Freund）

  某些情况下，`VACUUM`（手工或自动）可能错误地推进表的 `relfrozenxid` 值，使元组逃过冻结，导致这些行在经过 2^31 个事务后变得不可见。由于需要多次错误推进才会发生实际丢失，数据丢失的概率相当低，但并非为零。从 9.0.4、8.4.8 或更早版本升级的用户不受影响，但之后的所有版本都包含该缺陷。

  升级后，在把 `vacuum_freeze_table_age` 设为零的情况下对所有数据库中的所有表执行清理，可以缓解此问题。这会修复任何潜伏的损坏，但无法修复所有既有的数据错误。不过，如果安装在其生命周期内执行的事务少于 2^31 个更新（可用 `SELECT txid_current() < 2^31` 检查），则在执行此清理之后即可认为该安装是安全的。

- 修复热备启动期间 `pg_clog` 和 `pg_subtrans` 的初始化（Andres Freund、Heikki Linnakangas）

  这一缺陷会把已提交事务标记为未提交，从而在备库开始接受热备查询的那一刻造成数据丢失。除非备库启动时主库自上次检查点以来执行了大量更新事务，否则这种损坏的可能性很小。症状包括行缺失、本应删除的行仍可见，以及已更新行的过时版本与新版本同时可见。

  该缺陷是在 9.3.0、9.2.5、9.1.10 和 9.0.14 版本中引入的。只运行过更早版本的备库没有风险。建议曾运行过任何含缺陷版本的备库在升级后从主库重新克隆（例如用新的基础备份）。

- 在 WAL 重放期间截断 `pg_multixact` 内容（Andres Freund）

  这避免了备库中磁盘空间消耗不断增加。

- 修复 GIN 索引 posting 树页面删除中的竞态条件（Heikki Linnakangas）

  这可能导致瞬时的错误结果或查询失败。

- 避免展平 `SELECT` 列表中包含包裹在子 `SELECT` 中的易变函数的子查询（Tom Lane）

  这避免了因易变函数被额外求值而导致的意外结果。

- 修复规划器对外连接中嵌套的非简单变量子查询输出的处理（Tom Lane）

  这一错误可能为涉及 `JOIN` 语法中多级子查询的查询生成错误的计划。

- 修复继承树的优化 MIN()/MAX() 计划生成错误（Tom Lane）

  当 MIN()/MAX() 的参数是表达式而不是简单变量时，规划器可能失败。

- 修复临时文件被过早删除的问题（Andres Freund）
- 修复规则打印中可能读取越过内存末尾的问题（Peter Eisentraut）
- 修复 `int2vector` 和 `oidvector` 值的数组切片（Tom Lane）

  这类表达式现在会被隐式提升为常规的 `int2` 或 `oid` 数组。

- 修复使用 SQL 标准的简单 GMT 偏移时区时的错误行为（Tom Lane）

  某些情况下，系统会使用简单 GMT 偏移值，而本应使用选择简单偏移之前一直生效的常规时区设置。此变更还使 `timeofday` 函数遵循简单 GMT 偏移时区。

- 防止记录 Windows 错误码翻译时可能出现的不当行为（Tom Lane）
- 正确地为 pg_ctl 生成的命令行加引号（Naoya Anzai、Tom Lane）

  此修复仅适用于 Windows。

- 修复当源数据库通过 `ALTER DATABASE SET` 设置 `default_transaction_read_only` 时 pg_dumpall 无法工作的问题（Kevin Grittner）

  此前，生成的脚本在恢复期间会失败。

- 让 ecpg 以大小写敏感方式搜索带引号的游标名（Zoltán Böszörményi）
- 修复 ecpg 对声明为 `varchar` 的变量列表的处理（Zoltán Böszörményi）
- 让 `contrib/lo` 防御不正确的触发器定义（Marc Cousin）
- 把时区数据文件更新为 tzdata 版本 2013h for DST law changes in Argentina, Brazil, Jordan, Libya, Liechtenstein, Morocco, and Palestine。Also, new timezone abbreviations WIB, WIT, WITA for Indonesia.

E.15. 版本 9.1.10

发行日期
2013-10-10

本次发布包含对 9.1.9 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.15.1. 迁移到版本 9.1.10

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.9 的版本升级，请参见第 E.16 节。

E.15.2. 变更

- 防止尝试对标识符做大小写折叠时损坏多字节字符（Andrew Dunstan）

  PostgreSQL 只在使用单字节服务器编码时才对非 ASCII 字符做大小写折叠。

- 修复 `wal_level = hot_standby` 时后台写入进程中的检查点内存泄漏（Naoya Anzai）
- 修复由 `lo_open()` 失败导致的内存泄漏（Heikki Linnakangas）
- 修复 `work_mem` 使用超过 24GB 内存时的内存过度提交缺陷（Stephen Frost）
- 可序列化快照修复（Kevin Grittner、Heikki Linnakangas）
- 修复使用 SSL 时 libpq 中的死锁缺陷（Stephen Frost）
- 修复多线程 libpq 应用中可能的 SSL 状态损坏（Nick Phillips、Stephen Frost）
- 正确计算包含大量 NULL 值的布尔列的行数估计（Andrew Gierth）

  此前，`col IS NOT TRUE` 和 `col IS NOT FALSE` 这类测试在估计计划代价时没有正确考虑 NULL 值。

- 防止把 `WHERE` 子句下推到不安全的 `UNION/INTERSECT` 子查询中（Tom Lane）

  `SELECT` 列表中包含集合返回函数或易变函数的 `UNION` 或 `INTERSECT` 子查询可能被不当优化，导致运行时错误或错误的查询结果。

- 修复罕见的 “failed to locate grouping columns” 规划器失败情况（Tom Lane）
- 修复带已删除列的外部表的 pg_dump 转储问题（Andrew Dunstan）

  此前这类情况可能导致 pg_upgrade 错误。

- 重新排序 pg_dump 对扩展相关规则和事件触发器的处理（Joe Conway）
- 如果通过 `pg_dump -t` 或 `-n` 指定，则强制转储扩展表（Joe Conway）
- 改进视图转储代码对被引用表中已删除列的处理（Tom Lane）
- 修复目录归档下 `pg_restore -l` 显示正确格式名的问题（Fujii Masao）
- 正确记录使用 `UNIQUE` 和 `PRIMARY KEY` 语法创建的索引注释（Andres Freund）

  这修复了一个并行 pg_restore 失败问题。

- 在干净切换之前正确保证 WAL 文件的传输（Fujii Masao）

  此前，流复制连接可能在所有 WAL 文件都已在备库上重放之前关闭。

- 修复恢复期间 WAL 段的时间线处理（Mitsumasa Kondo、Heikki Linnakangas）

  备库恢复期间的 WAL 文件回收可能导致恢复过早完成，造成数据丢失。

- 修复 `REINDEX TABLE` 和 `REINDEX DATABASE`，使其正确重新验证约束并把已失效的索引标记为有效（Noah Misch）

  `REINDEX INDEX` 一直工作正常。

- 修复并发执行 `CREATE INDEX CONCURRENTLY` 操作时可能出现的死锁（Tom Lane）
- 修复 `regexp_matches()` 对零长度匹配的处理（Jeevan Chalke）

  此前，像 '^' 这样的零长度匹配可能返回过多的匹配。
- 修复过于复杂的正则表达式导致崩溃的问题（Heikki Linnakangas）
- 修复反向引用与非贪婪量词组合时正则表达式匹配失败的问题（Jeevan Chalke）
- 除非启用了函数体检查，否则防止 `CREATE FUNCTION` 检查 `SET` 变量（Tom Lane）
- 允许 `ALTER DEFAULT PRIVILEGES` 在不需要 CREATE 权限的情况下对模式操作（Tom Lane）
- 放宽对查询中所用关键字的限制（Tom Lane）

  具体而言，放宽了角色名、语言名、`EXPLAIN` 和 `COPY` 选项以及 `SET` 值的关键字限制。这使得 `COPY ... (FORMAT BINARY)` 能按预期工作；此前 `BINARY` 需要加引号。
- 修复 `pgp_pub_decrypt()`，使其可用于带口令的密钥（Marko Kreen）
- 让 pg_upgrade 使用 `pg_dump --quote-all-identifiers`，以避免版本之间关键字变更带来的问题（Tom Lane）
- 移除无索引表清理期间罕见的不准确警告（Heikki Linnakangas）
- 确保当 `VACUUM ANALYZE` 的文件截断尝试因锁冲突被取消时，仍然执行 ANALYZE 阶段（Kevin Grittner）
- 避免在预备查询中执行事务控制命令（例如 `ROLLBACK`）时可能出现的失败（Tom Lane）
- 确保浮点数据输入在所有平台上都接受标准的“infinity”（无穷大）写法（Tom Lane）

  C99 标准规定允许的写法为 `inf`、`+inf`、`-inf`、`infinity`、`+infinity` 和 `-infinity`。即使平台的 `strtod` 函数不支持，也确保我们能识别这些写法。
- 扩展行与记录和数组比较的能力（Rafal Rzepecki、Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2013d for DST law changes in Israel, Morocco, Palestine, and Paraguay。Also, historical zone data corrections for Macquarie Island.

E.16. 版本 9.1.9

发行日期
2013-04-04

本次发布包含对 9.1.8 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.16.1. 迁移到版本 9.1.9

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了 GiST 索引管理中的若干错误。安装此更新后，建议对满足下述一个或多个条件的 GiST 索引执行 `REINDEX`。

不过，如果你是从早于 9.1.6 的版本升级，请参见第 E.19 节。

E.16.2. 变更

- 修复服务器命令行开关解析不安全的问题（Mitsumasa Kondo、Kyotaro Horiguchi）

  包含以“-”开头的数据库名的连接请求可能被精心构造，
  用于破坏或摧毁服务器数据目录中的文件，即使该请求最终被拒绝也是如此。（CVE-2013-1899）

- 在每个 postmaster 子进程中重置 OpenSSL 随机状态（Marko Kreen）

  这避免了另一种情况：另一数据库用户可能相对容易地猜出 `contrib/pgcrypto` 函数生成的随机数。只有当 postmaster 配置了 `ssl` = `on` 而大多数连接并不使用 SSL 加密时，这一风险才比较显著。（CVE-2013-1900）

- 让 REPLICATION 权限检查测试当前用户而不是已认证用户（Noah Misch）

  无特权数据库用户可能利用这一失误调用 `pg_start_backup()` 或 `pg_stop_backup()`，从而可能干扰例程备份的创建。（CVE-2013-1901）

- 修复 GiST 索引在不合适的情况下使用“模糊”几何比较的问题（Alexander Korotkov）

  核心几何类型使用“模糊”相等进行比较，但 `gist_box_same` 必须做精确比较，否则使用它的 GiST 索引可能变得不一致。安装此更新后，用户应对 `box`、`polygon`、`circle` 或 `point` 列上的所有 GiST 索引执行 `REINDEX`，因为它们都使用 `gist_box_same`。

- 修复对可变宽度数据类型（即 `text`、`bytea`、`bit` 和 `numeric` 列）使用 `contrib/btree_gist` 的 GiST 索引中错误的范围并集和惩罚逻辑（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，还会造成无用的索引膨胀。
  建议用户在安装此更新后对这类索引执行 `REINDEX`。

- 修复多列索引的 GiST 页面分裂代码中的缺陷（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，
  还会使索引的搜索效率不必要地低下。建议用户在安装此更新后对多列 GiST 索引执行 `REINDEX`。

- 修复 `gist_point_consistent`，使其一致地处理模糊性（Alexander Korotkov）

  `point` 列上 GiST 索引的索引扫描有时会得出与顺序扫描不同的结果，因为 `gist_point_consistent` 与底层操作符代码在是做精确比较还是模糊比较上不一致。

- 修复 WAL 重放中的缓冲区泄漏（Heikki Linnakangas）

  这一缺陷可能导致重放期间出现“incorrect local pin count”错误，使恢复无法进行。

- 修复 `DELETE RETURNING` 中的竞态条件（Tom Lane）

  在合适的条件下，`DELETE RETURNING` 可能尝试从当前进程已不再持有任何 pin 的共享缓冲区中取数据。如果其他进程在此期间更改了该缓冲区，就会导致垃圾 `RETURNING` 输出，甚至崩溃。

- 修复正则表达式编译中的无限循环风险（Tom Lane、Don Porter）
- 修复正则表达式编译中潜在的空指针解引用（Tom Lane）
- 修复 `to_char()`，在适当的地方只使用 ASCII 大小写折叠规则（Tom Lane）

  这修复了一些本应与区域设置无关的模板模式在土耳其区域下错误处理“`I`”和“`i`”的问题。
- 修复时间戳 `1999-12-31 24:00:00` 被不当拒绝的问题（Tom Lane）
- 修复单个事务先执行 `UNLISTEN` 再执行 `LISTEN` 时的逻辑错误（Tom Lane）

  会话最终完全没有监听通知事件，尽管在这种情况下它显然应该监听。
- 修复在向被另一视图依赖的视图添加列之后规划器可能崩溃的问题（Tom Lane）
- 移除无用的“picksplit doesn't support secondary split”日志消息（Josh Hansen、Tom Lane）

  这条消息似乎是为预期中从未编写、而且可能永远不会编写的代码而添加的，因为 GiST 对二次分裂的默认处理其实相当好。所以不再为此打扰最终用户。
- 修复会话最后几个事务的提交/中止计数可能未能发送到统计收集器的问题（Tom Lane）
- 消除 PL/Perl 的 `spi_prepare()` 函数中的内存泄漏（Alex Hunsaker、Tom Lane）
- 修复 pg_dumpall 对包含“`=`”的数据库名的处理（Heikki Linnakangas）
- 避免在给定不正确的连接串时 pg_dump 崩溃（Heikki Linnakangas）
- 让 pg_dump 和 pg_upgrade 忽略无效索引（Michael Paquier、Bruce Momjian）

  转储无效索引可能在恢复时造成问题，
  例如当索引创建失败的原因是它试图强制执行表数据不满足的唯一性条件时。此外，
  如果索引创建实际上仍在进行中，把它视为未提交的 DDL 变更似乎是合理的，pg_dump 本来也不应转储它。pg_upgrade 现在也跳过无效索引而不是失败。
- 在 pg_basebackup 中，备份表空间时只包含当前服务器版本的子目录（Heikki Linnakangas）
- 在 pg_basebackup 和 pg_receivexlog 中加入服务器版本检查，使其在版本组合无法工作时干净地失败（Heikki Linnakangas）
- 修复 `contrib/pg_trgm` 的 `similarity()` 函数，使其对无三元组的字符串返回零（Tom Lane）

  此前它由于内部除以零而返回 `NaN`。
- 把时区数据文件更新为 tzdata 版本 2013b for DST law changes in Chile, Haiti, Morocco, Paraguay, and some Russian areas。Also, historical zone data corrections for numerous places.

  此外，更新时区缩写文件以反映俄罗斯和其他地方的近期变更：`CHOT`、`GET`、`IRKT`、`KGT`、`KRAT`、`MAGT`、`MAWT`、`MSK`、`NOVT`、`OMST`、`TKT`、`VLAT`、`WST`、`YAKT`、`YEKT` 现在遵循其当前含义，同时把 `VOLT`（Europe/Volgograd）和 `MIST`（Antarctica/Macquarie）加入默认缩写列表。

E.17. 版本 9.1.8

发行日期
2013-02-07

本次发布包含对 9.1.7 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.17.1. 迁移到版本 9.1.8

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.6 的版本升级，请参见第 E.19 节。

E.17.2. 变更

- 防止从 SQL 执行 `enum_recv`（Tom Lane）

  该函数声明有误，允许一条简单的 SQL 命令使服务器崩溃。原则上攻击者可能利用它检查服务器内存的内容。感谢 Sumit Soni（通过 Secunia SVCRP）报告此问题。（CVE-2013-0255）

- 修复 WAL 重放期间检测何时达到一致数据库状态的多个问题（Fujii Masao、Heikki Linnakangas、Simon Riggs、Andres Freund）

- 在截断关系文件时更新最小恢复点（Heikki Linnakangas）

  数据一旦被丢弃，在时间线上更早的点停止恢复就不再安全。

- 修复更改恢复目标时间线后 WAL 段的回收问题（Heikki Linnakangas）

- 修复热备模式下缺失的取消操作（Noah Misch、Simon Riggs）

  取消冲突热备查询的需要有时会被遗漏，使这些查询看到不一致的数据。

- 防止恢复暂停功能在用户能够连接之前就暂停（Tom Lane）

- 修复 SQL 文法，允许对子 SELECT 结果做下标或字段选择（Tom Lane）

- 修复繁忙负载下自动清理截断的性能问题（Jan Wieck）

  截断表末尾的空页需要排他锁，
  但自动清理被编码为在出现冲突锁请求时失败（并释放表锁）。在负载下，
  截断很可能永远不会发生，导致表膨胀。修复方法是先执行部分截断，释放锁，
  然后尝试重新获取锁并继续。
  此修复还大幅缩短了冲突请求到达后自动清理释放锁的平均时间。

- 防止扫描 `pg_tablespace` 时的竞态条件（Stephen Frost、Tom Lane）

  如果 `pg_tablespace` 条目存在并发更新，`CREATE DATABASE` 和 `DROP DATABASE` 可能行为异常。

- 防止 `DROP OWNED` 试图删除整个数据库或表空间（Álvaro Herrera）

为安全起见，这些对象的所有权必须重新分配，而不是删除。

- 修复 `vacuum_freeze_table_age` 实现中的错误（Andres Freund）

  在已存在超过 `vacuum_freeze_min_age` 个事务的安装中，这一失误使自动清理无法使用部分表扫描，而总是进行全表扫描。

- 防止 `RowExpr` 或 `XmlExpr` 被二次解析分析时的不当行为（Andres Freund、Tom Lane）

  这一失误在 `CREATE TABLE LIKE INCLUDING INDEXES` 等上下文中可能被用户看到。

- 增强对哈希表大小计算中整数溢出的防御（Jeff Davis）
- 修复服务器崩溃后未能忽略遗留临时表的问题（Tom Lane）
- 拒绝 `to_date()` 中超出范围的日期（Hitoshi Harada）
- 修复 `pg_extension_config_dump()`，使其正确处理扩展更新情况（Tom Lane）

  该函数现在会替换目标表的任何现有条目，使其可以在扩展更新脚本中使用。

- 修复 PL/Python 对用作多个表上触发器的函数的处理（Andres Freund）
- 确保在 Windows 上非 ASCII 提示符字符串被转换到正确的代码页（Alexander Law、Noah Misch）

  这一缺陷影响 psql 和其他一些客户端程序。

- 修复未连接数据库时 psql 的 `\?` 命令可能崩溃的问题（Meng Qingzhong）
- 修复 pg_basebackup 运行期间关系文件被删除时可能出错的问题（Heikki Linnakangas）
- 让 pg_dump 在热备服务器上运行时排除不记录日志的表的数据（Magnus Hagander）

  由于数据在备库服务器上不可用，这本来就会失败，因此最方便的做法似乎是自动假定 `--no-unlogged-table-data`。

- 修复 pg_upgrade，使其安全处理无效索引（Bruce Momjian）
- 修复 libpq 的 `PQprintTuples` 中的一字节缓冲区越界（Xi Wang）

  这个古老的函数没有被 PostgreSQL 本身在任何地方使用，但仍可能被某些客户端代码使用。

- 让 ecpglib 正确使用翻译后的消息（Chen Huajun）
- 在 MSVC 上正确安装 ecpg_compat 和 pgtypes 库（Jiang Guiqing）
- 如果系统未提供 `isinf()`，则在 libecpg 中包含我们自己的版本（Jiang Guiqing）
- 重新排列 configure 对系统提供函数的测试，使其不被 libedit/libreadline 的虚假导出所欺骗（Christoph Berg）
- 确保 Windows 构建号随时间递增（Magnus Hagander）
- 让 pgxs 在为 Windows 交叉编译时以正确的 `.exe` 后缀构建可执行文件（Zoltan Boszormenyi）

- 新增时区缩写 `FET`（Tom Lane）

  它现在用于一些东欧时区。

E.18. 版本 9.1.7

发行日期
2012-12-06

本次发布包含对 9.1.6 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.18.1. 迁移到版本 9.1.7

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.6 的版本升级，请参见第 E.19 节。

E.18.2. 变更

- 修复与 `CREATE INDEX CONCURRENTLY` 相关的多个缺陷（Andres Freund、Tom Lane）

  修复 `CREATE INDEX CONCURRENTLY`，使其在更改索引 `pg_index` 行的状态时使用就地更新。这防止了可能导致并发会话错过更新目标索引（从而产生损坏的并发创建索引）的竞态条件。

  另外，修复各种其他操作，确保它们忽略由失败的 `CREATE INDEX CONCURRENTLY` 命令产生的无效索引。其中最重要的是 `VACUUM`，因为在采取纠正措施修复或删除无效索引之前，很容易就有自动清理在该表上启动。

- 修复 WAL 重放期间的缓冲区锁定（Tom Lane）

  WAL 重放代码在重放影响多个页面的 WAL 记录时，对缓冲区加锁不够谨慎。这可能导致热备查询瞬时看到不一致的状态，造成错误结果或意外失败。

- 修复 GIN 索引 WAL 生成逻辑中的一个错误（Tom Lane）

  如果发生残缺页故障，这可能导致索引损坏。

- 把热备服务器提升为正常运行时，正确移除启动进程的虚拟 XID 锁（Simon Riggs）

  这一疏忽可能阻止后续执行某些操作，例如 `CREATE INDEX CONCURRENTLY`。

- 避免备库模式下虚假的 “out-of-sequence timeline ID” 错误（Heikki Linnakangas）

- 防止 postmaster 在收到关闭信号后继续启动新的子进程（Tom Lane）

  这一失误可能导致关闭耗时比应有的更长，甚至在没有额外用户操作的情况下永远无法完成。

- 避免内存不足时内部哈希表损坏（Hitoshi Harada）

- 防止已删除表的文件描述符在事务结束之后仍保持打开（Tom Lane）

这应能减少早已删除的表继续占用磁盘空间的问题。

- 防止当新子进程无法为其闩锁创建管道时整个数据库崩溃并重启（Tom Lane）

  虽然新进程必须失败，但没有充分理由强制整个数据库重启，因此避免这样做。这提高了内核文件描述符几乎耗尽时的健壮性。

- 修复外连接之上非严格等值子句的规划（Tom Lane）

  规划器可能从把非严格构造与别的值等同的子句中推导出错误的约束，例如当 `foo` 来自外连接的可空侧时的 `WHERE COALESCE(foo, 0) = 0`。

- 修复在继承树上对索引优化的 `MIN/MAX` 使用 `SELECT DISTINCT` 的问题（Tom Lane）

  在这种因素组合下，规划器会报 “failed to re-find MinMaxAggInfo record” 失败。

- 改进规划器从等价类证明排他约束的能力（Tom Lane）

- 修复哈希子计划中的部分行匹配，使其正确处理跨类型情况（Tom Lane）

  这影响多列 `NOT IN` 子计划，例如 `WHERE (a, b) NOT IN (SELECT x, y FROM ...)`，当 `b` 和 `y` 分别是 `int4` 和 `int8` 时。根据所涉及的具体数据类型，这一失误会导致错误答案或崩溃。

- 在为 `AFTER ROW UPDATE/DELETE` 触发器重新获取旧元组时获取缓冲区锁（Andres Freund）

  在非常罕见的情况下，这一疏忽可能导致向触发器 `WHEN` 条件或外键强制触发器的预检查逻辑传递不正确的数据。这可能导致崩溃，或者在是否触发触发器上做出错误决定。

- 修复`ALTER COLUMN TYPE`以正确重建继承的检查约束（Tom Lane）

  这在 8.4 之前的版本中工作正常，现在在 8.4 及更高版本中也工作正常。

- 修复 `ALTER EXTENSION SET SCHEMA` 未能把一些从属对象移入新模式的问题（Álvaro Herrera、Dimitri Fontaine）

- 修复 `REASSIGN OWNED` 对表空间授权的处理（Álvaro Herrera）

- 忽略视图中系统列的不正确 `pg_attribute` 条目（Tom Lane）

  视图没有任何系统列。然而，我们在把表转换为视图时忘记移除这类条目。这一问题在 9.3 及更高版本中已正确修复，但在以前的分支中，我们需要防御已经错误转换的视图。

- 修复规则打印，使其正确转储 `INSERT INTO` *`table`* `DEFAULT VALUES`（Tom Lane）

- 防止当查询中有过多 `UNION/INTERSECT/EXCEPT` 子句时栈溢出（Tom Lane）

- 防止用最小可能整数值除以 -1 时依赖平台的失败（Xi Wang、Tom Lane）

- 修复日期解析中可能越过字符串末尾的访问（Hitoshi Harada）

- 修复当 XID 回卷发生在检查点期间且 `wal_level` 为 `hot_standby` 时 XID 纪元（epoch）未能推进的问题（Tom Lane、Andres Freund）

  虽然这一失误对 PostgreSQL 本身没有特别影响，但它对依赖 `txid_current()` 及相关函数的应用不利：TXID 值看起来会倒退。

- 修复 `pg_stat_replication.sync_state` 在页边界处的显示（Kyotaro Horiguchi）
- 当 Unix 域套接字的路径名长度超过平台特定限制时，生成可理解的错误消息（Tom Lane、Andrew Dunstan）

  此前这会导致相当无用的结果，例如“Non-recoverable failure in name resolution”。
- 修复向客户端发送组合列值时的内存泄漏（Tom Lane）
- 使 pg_ctl 读取 `postmaster.pid` 文件更健壮（Heikki Linnakangas）

  这修复了竞态条件和可能的文件描述符泄漏。
- 修复当呈现错误编码的数据且 `client_encoding` 设置为仅客户端编码（如 SJIS）时 psql 可能崩溃的问题（Jiang Guiqing）
- 让 pg_dump 把 `SEQUENCE SET` 项转储到归档的数据节而不是前数据节（Tom Lane）

  此变更修复了被标记为扩展配置表的序列的转储。
- 修复 pg_dump 以 `tar` 输出格式生成的 `restore.sql` 脚本中的缺陷（Tom Lane）

  该脚本在表名包含大写字符时会直接失败。此外，让该脚本既能以常规 COPY 模式也能以 `--inserts` 模式恢复数据。
- 修复 pg_restore，使其接受符合 POSIX 的 `tar` 文件（Brian Weaver、Tom Lane）

  pg_dump 的 `tar` 输出模式的原始编码生成的文件并不完全符合 POSIX 标准。这在 9.3 版本中已得到纠正。此补丁更新了以前的分支，使它们同时接受错误格式和纠正后的格式，希望借此避免 9.3 发布时的兼容性问题。
- 修复 pg_basebackup 生成的 `tar` 文件，使其符合 POSIX（Brian Weaver、Tom Lane）
- 修复 pg_resetxlog，使其在给定数据目录相对路径时正确定位 `postmaster.pid`（Tom Lane）

  这一失误可能导致 pg_resetxlog 没有注意到有活动 postmaster 正在使用该数据目录。
- 修复 libpq 的 `lo_import()` 和 `lo_export()` 函数，使其正确报告文件 I/O 错误（Tom Lane）
- 修复 ecpg 对嵌套结构指针变量的处理（Muhammad Usama）
- 修复 ecpg 的 `ecpg_get_data` 函数，使其正确处理数组（Michael Meskes）
- 让 `contrib/pageinspect` 的 btree 页面检查函数在检查页面时持有缓冲区锁（Tom Lane）
- 确保扩展的 `make install` 会创建 `extension` 安装目录（Cédric Villemain）

  此前，如果扩展的 Makefile 设置了 `MODULEDIR`，就会漏掉这一步。
- 修复 pgxs 在 AIX 上构建可加载模块的支持（Tom Lane）

  在原始源码树之外构建模块在 AIX 上不工作。
- 把时区数据文件更新为 tzdata 版本 2012j for DST law changes in Cuba, Israel, Jordan, Libya, Palestine, Western Samoa, and portions of Brazil。

E.19. 版本 9.1.6

发行日期
2012-09-24

本次发布包含对 9.1.5 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.19.1. 迁移到版本 9.1.6

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，你可能需要执行 `REINDEX` 操作，以从下面第一条变更日志条目所述的数据损坏缺陷的影响中恢复。

不过，如果你是从早于 9.1.4 的版本升级，请参见第 E.21 节。

E.19.2. 变更

- 修复 WAL 重放期间共享缓冲区的持久性标记（Jeff Davis）

  这一失误可能导致缓冲区在检查点期间未被写出，
  如果服务器之后在这些缓冲区从未写出的情况下崩溃，就会造成数据损坏。
  任何服务器在崩溃恢复之后都可能发生损坏，但在备库从库服务器上发生的可能性显著更高，
  因为它们执行多得多的 WAL 重放。btree 和 GIN 索引损坏的概率较低。表“可见性映射”损坏的概率要高得多。幸运的是，可见性映射在 9.1 中不是关键数据，因此这种损坏在 9.1 安装中最坏的后果只是清理的瞬时低效。表数据本身不会因此缺陷损坏。

  虽然尚无此缺陷导致的索引损坏在实际环境中发生的已知案例，但作为预防措施，
  建议生产安装在升级到 9.1.6 后的方便时间对所有 btree 和 GIN 索引执行 `REINDEX`。

  此外，如果你打算就地升级到 9.2.X，建议在升级之前在把 `vacuum_freeze_table_age` 设为零的情况下对所有表执行 `VACUUM`。这将确保可见性映射中任何残留的错误数据在 9.2.X 依赖它之前被纠正。可以调整 `vacuum_cost_delay` 来降低清理的性能影响，代价是需要更长时间完成。

- 修复规划器对执行器参数的分配，并修复执行器对 CTE 计划节点的重新扫描逻辑（Tom Lane）

  这些错误可能导致多次扫描同一 `WITH` 子查询的查询给出错误答案。

- 修复把 `default_transaction_isolation` 设为 `serializable` 时的不当行为（Kevin Grittner、Tom Lane、Heikki Linnakangas）

  症状包括 Windows 上进程启动时崩溃，以及热备操作中崩溃。

- 改进对涉及前缀（即 *`word`*`:*` 模式）的文本搜索查询的选择率估计（Tom Lane）

- 改进 GiST 索引的页面分裂决策（Alexander Korotkov、Robert Haas、Tom Lane）

  多列 GiST 索引可能因此错误而遭受意外膨胀。

- 修复级联权限撤销，使其在权限仍被持有时停止（Tom Lane）

如果我们撤销某个角色 *X* 的授予选项，但 *X* 仍通过别人的授予持有该选项，我们就不应从 *X* 曾授予该权限的角色 *Y* 那里递归撤销相应权限。

- 不允许扩展包含被指派给它的模式（Thom Brown）

  这种情况会造成循环依赖，使 pg_dump 及可能的其他东西困惑。它对人类也容易造成困惑，因此予以禁止。

- 改进热备配置错误的错误消息（Gurjeet Singh）
- 让 configure 探测 `mbstowcs_l`（Tom Lane）

  这修复了某些 AIX 版本上的构建失败。

- 修复使用 PL/Perl 时对 `SIGFPE` 的处理（Andres Freund）

  Perl 会把进程的 `SIGFPE` 处理程序重置为 `SIG_IGN`，这可能导致之后崩溃。在初始化 PL/Perl 之后恢复正常的 Postgres 信号处理程序。

- 防止递归 PL/Perl 函数在执行期间被重定义时 PL/Perl 崩溃（Tom Lane）
- 规避 PL/Perl 中可能的错误优化（Tom Lane）

  一些 Linux 发行版包含错误版本的 `pthread.h`，导致 PL/Perl 中的编译代码不正确，当一个 PL/Perl 函数调用另一个会抛出错误的函数时会导致崩溃。

- 修复 `contrib/pg_trgm` 的 `LIKE` 模式分析代码中的缺陷（Fujii Masao）

  如果模式包含 `LIKE` 转义字符，使用三元组索引的 `LIKE` 查询可能产生错误结果。

- 修复 pg_upgrade 在 Windows 上对行尾的处理（Andrew Dunstan）

  此前，pg_upgrade 可能在函数体等地方添加或删除回车符。

- 在 Windows 上，让 pg_upgrade 在其生成的脚本中使用反斜杠路径分隔符（Andrew Dunstan）
- 移除 pg_upgrade 对 pg_config 的不必要依赖（Peter Eisentraut）
- 把时区数据文件更新为 tzdata 版本 2014e，以反映克里米亚、埃及和摩洛哥的夏令时法规变更。

E.20. 版本 9.1.5

发行日期
2012-08-17

本次发布包含对 9.1.4 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.20.1. 迁移到版本 9.1.5

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.4 的版本升级，请参见第 E.21 节。

E.20.2. 变更

- 防止通过 XML 实体引用访问外部文件/URL（Noah Misch、Tom Lane）

  `xml_parse()` 会尝试按需获取外部文件或 URL 来解析 XML 值中的 DTD 和实体引用，从而使无特权数据库用户能够尝试以数据库服务器的权限获取数据。虽然外部数据不会直接返回给用户，但如果数据不能解析为有效 XML，其部分内容可能暴露在错误消息中；而且无论如何，仅仅能够检查文件是否存在对攻击者就可能有用。（CVE-2012-3489）

- 防止通过 `contrib/xml2` 的 `xslt_process()` 访问外部文件/URL（Peter Eisentraut）

  libxslt 提供通过样式表命令读写文件和 URL 的能力，从而使无特权数据库用户能够以数据库服务器的权限读写数据。现通过正确使用 libxslt 的安全选项禁用该能力。（CVE-2012-3488）

  此外，移除了 `xslt_process()` 从外部文件/URL 获取文档和样式表的能力。虽然这是一个有记载的“特性”，但长期以来被认为是个坏主意。CVE-2012-3489 的修复破坏了该能力，与其花费精力修复它，我们干脆将其移除。

- 防止 btree 索引页被过早回收（Noah Misch）

  当我们允许只读事务跳过分配 XID 时，引入了已删除 btree 页在只读事务仍在访问它时就被回收的可能性。这会导致错误的索引搜索结果。由于时序要求，
  这种错误在实际中发生的概率似乎很低，但仍应修复。

- 修复新创建或重置序列的崩溃安全缺陷（Tom Lane）

  如果在新创建或重置的序列上执行了 `ALTER SEQUENCE`，然后恰好对它调用了一次 `nextval()`，随后服务器崩溃，WAL 重放会把序列恢复到看起来没有执行过 `nextval()` 的状态，从而使下一个 `nextval()` 调用再次返回第一个序列值。这种情况尤其可能出现在 `serial` 列上，因为创建 serial 列的序列包含一个 `ALTER SEQUENCE OWNED BY` 步骤。

- 修复 `enum` 类型值比较中的竞态条件（Robert Haas、Tom Lane）

  当遇到当前查询启动之后添加的枚举值时，比较可能失败。

- 修复 `txid_current()`，使其在非热备下报告正确的纪元（epoch）（Heikki Linnakangas）

  这修复了上一个小版本引入的退化。

- 防止选择不合适的复制连接作为同步备库（Fujii Masao）

  主库可能不当地选择 pg_receivexlog 或 pg_basebackup 等伪服务器作为同步备库，然后无限期等待它们。

- 修复主库事务有大量子事务时热备启动的缺陷（Andres Freund）

  这一失误导致报告为“out-of-order XID insertion in KnownAssignedXids”的失败。

- 确保 `pg_start_backup()` 之后对 `backup_label` 文件执行 fsync（Dave Kerr）

- 修复 walsender 进程中的超时处理（Tom Lane）

  WAL 发送端后台进程疏于建立 SIGALRM 处理程序，这意味着在某些应发生超时的极端情况下它们会永远等待。

- 在 walwriter 每次后台刷写之后唤醒 walsender（Andres Freund、Simon Riggs）

  当负载只包含异步提交的事务时，这能大幅减少复制延迟。

- 修复 `LISTEN/NOTIFY`，使其更好地应对 I/O 问题（例如磁盘空间耗尽）（Tom Lane）

  写入失败之后，所有后续发送更多 `NOTIFY` 消息的尝试都会失败，并显示诸如 “Could not read from file "pg_notify/*nnnn*" at offset *nnnnn*: Success” 的消息。

- 只允许被直接阻塞的进程自动取消自动清理（Tom Lane）

  原始编码在某些情况下可能允许不一致的行为；特别是自动清理可能在不足 `deadlock_timeout` 宽限期之后就被取消。

- 改进自动清理取消的日志记录（Robert Haas）

- 修复日志收集器，使 `log_truncate_on_rotation` 在服务器启动后的第一次日志轮转期间就能工作（Tom Lane）

- 修复附加到嵌套集合操作（`UNION/INTERSECT/EXCEPT`）上的 `WITH`（Tom Lane）

- 确保对子查询的整行引用不包含任何多余的 `GROUP BY` 或 `ORDER BY` 列（Tom Lane）

- 修复 `ALTER TABLE ... ADD CONSTRAINT USING INDEX` 期间生成的依赖（Tom Lane）

  该命令为索引留下了一个多余的 `pg_depend` 条目，可能干扰后续操作，尤其是对被索引列之一的 `ALTER TABLE ... ALTER COLUMN TYPE`。

- 修复 `REASSIGN OWNED`，使其能用于扩展（Alvaro Herrera）

- 禁止在 `CREATE TABLE` 期间复制 `CHECK` 约束和索引定义中的整行引用（Tom Lane）

  这种情况可能出现在带 `LIKE` 或 `INHERITS` 的 `CREATE TABLE` 中。被复制的整行变量被错误地标记为原表而不是新表的行类型。对 `LIKE` 拒绝这种情况似乎是合理的，因为行类型之后很可能分化。对 `INHERITS`，理想情况下应该允许它，并隐式强制转换为父表的行类型；但这需要的工作量超出了回移植的安全范围。

- 修复 `ARRAY(SELECT ...)` 子查询中的内存泄漏（Heikki Linnakangas、Tom Lane）

- 修复规划器，使其向操作符选择率估计器传递正确的排序规则（Tom Lane）

  此前任何核心选择率估计函数都不需要它，但第三方代码可能需要。

- 修复从正则表达式中提取公共前缀的问题（Tom Lane）

  代码可能被带量词的带括号子表达式（例如 `^(foo)?bar`）混淆。这会导致对此类模式搜索的索引优化不正确。

- 修复解析 `interval` 常量中带符号的 *hh*:*mm* 和 *hh*:*mm*:*ss* 字段的缺陷（Amit Kapila、Tom Lane）

- 修复 pg_dump，使其更好地处理包含部分 `GROUP BY` 列表的视图（Tom Lane）

  只在 `GROUP BY` 中列出主键列、却把其他表列当作已分组的视图，会被标记为依赖于该主键。pg_dump 对这类主键依赖处理不当会导致转储顺序不佳，往好里说恢复效率低下，往坏里说可能导致并行 pg_restore 彻底失败。

- 在 PL/Perl 中，避免在 SQL_ASCII 编码下设置 UTF8 标志（Alex Hunsaker、Kyotaro Horiguchi、Alvaro Herrera）
- 在 PL/Python 中把 Python Unicode 字符串转换为服务器编码时，使用 Postgres 而不是 Python 的编码转换函数（Jan Urbanski）

  这避免了一些极端情况的问题，尤其是 Python 并不支持 Postgres 的所有编码。一个显著的功能变化是：如果服务器编码是 SQL_ASCII，你将得到该字符串的 UTF-8 表示；此前字符串中的任何非 ASCII 字符都会导致错误。
- 修复 PL/Python 中 PostgreSQL 编码到 Python 编码的映射（Jan Urbanski）
- 修复 `contrib/xml2` 的 `xslt_process()` 中的错误报告（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2012e，以反映摩洛哥和托克劳的夏令时法规变更

E.21. 版本 9.1.4

发行日期
2012-06-04

本次发布包含对 9.1.3 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.21.1. 迁移到版本 9.1.4

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你使用 `citext` 数据类型，并且是通过运行 pg_upgrade 从前一个大版本升级的，应执行 `CREATE EXTENSION citext FROM unpackaged`，以避免 `citext` 操作中与排序规则相关的失败。如果恢复来自包含 `citext` 数据类型的 9.1 之前数据库的转储，也需要这样做。如果你在升级到 9.1.4 之前已经运行过 `CREATE EXTENSION` 命令，则需要按下面第三条变更日志条目的说明手工更新系统目录。

不过，如果你是从早于 9.1.2 的版本升级，请参见第 E.23 节。

E.21.2. 变更

- 修复 `contrib/pgcrypto` 的 DES `crypt()` 函数中不正确的口令变换（Solar Designer）

  如果口令串包含字节值 `0x80`，口令的其余部分会被忽略，使口令比看起来弱得多。修复后，字符串的其余部分会被正确纳入 DES 哈希。因此，受此缺陷影响的任何已存储口令值将不再匹配，所以可能需要更新已存储的值。（CVE-2012-2143）
- 忽略过程语言调用处理程序的 `SECURITY DEFINER` 和 `SET` 属性（Tom Lane）

  把这类属性应用于调用处理程序可能导致服务器崩溃。（CVE-2012-2655）
- 让 `contrib/citext` 的升级脚本修复 `citext` 数组以及 `citext` 之上的域的排序规则（Tom Lane）

  9.1.2 提供了对从 9.1 之前安装升级或重载的数据库中 `citext` 列和索引排序规则的修复，但该修复不完整：它遗漏了对数组以及 `citext` 之上的域的处理。本版本扩展了该模块的升级脚本

以处理这些情况。和以前一样，如果你已经运行过升级脚本，
则需要改为手工运行排序规则更新命令。具体做法参见 9.1.2 发行说明。

- 允许 `timestamp` 输入中的数字时区偏移距 UTC 最多 16 小时（Tom Lane）

  一些历史时区的偏移超过此前 15 小时的限制。这可能导致转储的数据值在重载期间被拒绝。

- 修复时间戳转换，
  使其在给定时间恰好是当前时区最后一次夏令时转换时间时能正确处理（Tom Lane）

  这一疏忽已存在很长时间，但此前未被注意到，
  因为大多数使用夏令时的时区被假定有无限的未来夏令时转换序列。

- 修复 `text` 到 `name` 以及 `char` 到 `name` 的转换，使其在多字节编码下正确执行字符串截断（Karl Schnaitter）

- 修复 `to_tsquery()` 中的内存复制缺陷（Heikki Linnakangas）

- 确保 `txid_current()` 在热备中执行时报告正确的纪元（epoch）（Simon Riggs）

- 修复规划器对子查询内外层 PlaceHolderVar 的处理（Tom Lane）

  该缺陷涉及引用来自外层查询外连接可空侧变量的子 SELECT。在 9.1 中，受此缺陷影响的查询会报 “ERROR: Upper-level PlaceHolderVar found where not expected” 失败。但在 9.0 和 8.4 中，你会静默地得到可能错误的答案，因为传入子查询的值不会在该为空时变为空。

- 修复输出列不是简单变量的 `UNION ALL` 子查询的规划（Tom Lane）

  由于对 “MergeAppend child's targetlist doesn't match MergeAppend” 错误的一次误导性修复，这类情况的规划在 9.1 中明显变差。现已撤销该修复并改用其他方法。

- 修复 `pg_attribute` 非常大会话启动缓慢的问题（Tom Lane）

  如果 `pg_attribute` 超过 `shared_buffers` 的四分之一，会话启动期间有时需要的缓存重建代码就会触发同步扫描逻辑，使其耗时比正常情况多数倍。如果许多新会话同时启动，
  问题尤为严重。

- 确保顺序扫描以合理的频率检查查询取消（Merlin Moncure）

  遇到许多不含存活元组的连续页面的扫描在此期间不会响应中断。

- 确保 `PGSemaphoreLock()` 的 Windows 实现在返回前清除 `ImmediateInterruptOK`（Tom Lane）

  这一疏忽意味着同一查询稍后收到的查询取消中断可能在不安全的时刻被接受，
  后果不可预测且不好。

- 在打印视图或规则时安全地显示整行变量（Abbas Butt、Tom Lane）

  涉及有歧义名称（即该名称既可能是查询的表名也可能是列名）
  的极端情况此前以有歧义的方式打印，存在视图或规则在转储和重载后被不同解释的风险。
  现通过附加无操作转换来避免歧义情况。

- 修复 `COPY FROM`，使其正确处理对应无效编码的空值标记串（Tom Lane）

  诸如 `E'\\0'` 这样的空值标记串本应有效，过去也确实有效，但该情况在 8.4 中被破坏。

- 修复对包含 `RETURNING` 子句的可写 CTE 的 `EXPLAIN VERBOSE`（Tom Lane）

- 修复 PREPARE TRANSACTION，使其在存在咨询锁时正确工作（Tom Lane）

  历史上，PREPARE TRANSACTION 只是忽略会话持有的任何会话级咨询锁，但这一情况在 9.1 中被意外破坏。

- 修复不记录日志的表的截断（Robert Haas）

- 在 search_path 的非交互式赋值期间忽略缺失的模式（Tom Lane）

  这使 9.1 的行为与较旧的分支重新一致。此前，9.1 会对来自 ALTER DATABASE SET 等处的 search_path 设置中提到的不存在模式抛出错误。

- 修复扩展脚本中所用临时表或瞬态表的缺陷（Tom Lane）

  这包括扩展更新脚本中的重写型 ALTER TABLE 等情况，因为它在后台使用了瞬态表。

- 确保自动清理工作进程正确执行栈深度检查（Heikki Linnakangas）

  此前，由自动 ANALYZE 调用的函数中的无限递归可能导致工作进程崩溃。

- 修复日志收集器，使其在高负载下不丢失日志连贯性（Andrew Dunstan）

  此前，如果收集器过于忙碌，可能无法重新组装大消息。

- 修复日志收集器，确保它在收到 SIGHUP 后会重新开始文件轮转（Tom Lane）

- 修复 GiST 索引中的 “too many LWLocks taken” 失败（Heikki Linnakangas）

- 修复 GIN 索引的 WAL 重放逻辑，使其在索引随后被删除时不失败（Tom Lane）

- 在崩溃后正确检测预备事务的 SSI 冲突（Dan Ports）

- 提交只修改了临时表的事务时避免同步复制延迟（Heikki Linnakangas）

  在这种情况下，事务的提交记录不需要刷到备库服务器，但部分代码不知道这一点，仍然等待它发生。

- 修复 pg_basebackup 中的错误处理（Thomas Ogrisegg、Fujii Masao）

- 修复 walsender，使其在连接终止时不进入忙循环（Fujii Masao）

- 修复 PL/pgSQL 的 RETURN NEXT 命令中的内存泄漏（Joe Conway）

- 修复 PL/pgSQL 的 GET DIAGNOSTICS 命令在目标是函数第一个变量时的问题（Tom Lane）

- 确保 PL/Perl 为 _TD 变量加上包限定（Alex Hunsaker）

  该缺陷导致触发器调用在嵌套于更改当前包的函数调用中时失败。

- 修复返回组合类型的 PL/Python 函数，使其接受字符串作为结果值（Jan Urbanski）

  这一情况被 9.1 中允许以其他格式（如字典）提供组合结果值的新增内容意外破坏。

- 修复 psql 扩展显示（\x）模式中可能越过内存末尾的访问（Peter Eisentraut）

- 修复数据库包含大量对象时 pg_dump 的若干性能问题（Jeff Janes、Tom Lane）

  如果数据库包含许多模式、许多对象处于依赖循环中、或有许多被拥有的序列，pg_dump 可能变得非常慢。

- 修复 pg_restore 读取目录格式归档时的内存和文件描述符泄漏（Peter Eisentraut）
- 修复 pg_upgrade 处理存储在非默认表空间的数据库包含集簇默认表空间中的表的情况（Bruce Momjian）
- 在 ecpg 中，修复罕见的内存泄漏以及 `sqlca_t` 结构之后一个字节可能被覆写的问题（Peter Eisentraut）
- 修复 `contrib/dblink` 的 `dblink_exec()`，使其在出错时不泄漏临时数据库连接（Tom Lane）
- 修复 `contrib/dblink`，使其在错误消息中报告正确的连接名（Kyotaro Horiguchi）
- 修复 `contrib/vacuumlo`，使其在删除许多大对象时使用多个事务（Tim Lewis、Robert Haas、Tom Lane）

  此变更避免了需要删除许多对象时超出 `max_locks_per_transaction`。该行为可以通过新的 `-l`（limit）选项调整。
- 把时区数据文件更新为 tzdata 版本 2012c for DST law changes in Antarctica, Armenia, Chile, Cuba, Falkland Islands, Gaza, Haiti, Hebron, Morocco, Syria, and Tokelau Islands; also historical corrections for Canada。

E.22. 版本 9.1.3

发行日期
2012-02-27

本次发布包含对 9.1.2 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.22.1. 迁移到版本 9.1.3

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.1.2 的版本升级，请参见第 E.23 节。

E.22.2. 变更

- 让 `CREATE TRIGGER` 要求对触发器函数有执行权限（Robert Haas）

  这一缺失的检查可能允许另一用户通过把触发器函数安装到自己拥有的表上，以伪造的输入数据执行该触发器函数。这只对标记为 `SECURITY DEFINER` 的触发器函数有影响，因为其他情况下触发器函数本来就以表所有者身份运行。（CVE-2012-0866）
- 移除 SSL 证书中通用名长度的任意限制（Heikki Linnakangas）

  libpq 和服务器此前都会把从 SSL 证书提取的通用名截断到 32 字节。通常这最坏只会造成意外的验证失败，但在一些相当难以置信的场景下，它可能允许一个证书持有者冒充另一个。受害者必须恰好有 32 字节长的通用名，而攻击者必须说服受信任的 CA 颁发一个通用名以该字符串为前缀的证书。冒充服务器还需要额外的攻击手段来重定向客户端连接。（CVE-2012-0867）

- 把 pg_dump 注释中写入的名称里的换行符转换为空格（Robert Haas）

  pg_dump 此前对其输出脚本中 SQL 注释内写出的对象名清理不慎。包含换行符的名称至少会使脚本语法错误。恶意构造的对象名可能在脚本重载时构成 SQL 注入风险。（CVE-2012-0868）

- 修复插入与清理并发导致 btree 索引损坏的问题（Tom Lane）

  插入导致的索引页分裂有时会使并发运行的 `VACUUM` 漏删本应删除的索引项。在相应的表行被删除后，悬空的索引项会导致错误（例如 “could not read block N in file ...” ），或者更糟：当不相关的行被重新插入到现已空闲的表位置后，查询会静默地得到错误结果。该缺陷自 8.2 起就存在，但发生频率极低，直到现在才被诊断出来。如果你有理由怀疑它曾在你的数据库中发生，对受影响的索引重建索引即可修复。

- 修复 WAL 重放期间共享缓冲区的瞬时清零（Tom Lane）

  重放逻辑有时会把一个共享缓冲区清零再重新填充，使其内容瞬时无效。在热备模式下，这可能导致并行执行的查询看到垃圾数据。这可能引发各种症状，但最常见的好像是 “invalid memory alloc request size” 。

- 修复 `READ COMMITTED` 重新检查中对修改数据的 `WITH` 子计划的处理（Tom Lane）

  在 `READ COMMITTED` 模式下，如果父 `UPDATE` 或 `DELETE` 命令因并发更新而需要在一行或多行上重新求值，包含 `INSERT`/`UPDATE`/`DELETE` 的 `WITH` 子句就会崩溃。

- 修复 SSI 事务清理中的一个极端情况（Dan Ports）

  在结束一个读写可序列化事务时，如果其余所有活动的可序列化事务都是只读的，就可能发生崩溃。

- 修复 postmaster，使其在热备崩溃后尝试重启（Tom Lane）

  一个逻辑错误导致：如果有后端进程在热备模式下运行时崩溃，postmaster 会终止而不是尝试重启集簇。

- 修复 `CLUSTER`/`VACUUM FULL` 对最近更新的行所拥有的 toast 值的处理（Tom Lane）

  这一疏忽可能导致在这些命令之一执行期间，toast 表的索引报 “duplicate key value violates unique constraint” 错误。

- 更改表所有者时，不仅更新表级权限，也更新逐列权限（Tom Lane）

  未做到这一点意味着，此前授予的所有列权限仍显示为由旧所有者授予。这导致新所有者和超级用户都无法撤销这些现在无法追溯到表所有者的权限。

- 在 `REASSIGN OWNED` 中支持外部数据包装器和外部服务器（Alvaro Herrera）

  如果该命令需要更改任何这类对象的所有权，会报 “unexpected classid” 错误。

- 允许 `ALTER USER/DATABASE SET` 中某些设置取不存在的值（Heikki Linnakangas）

  允许把 `default_text_search_config`、`default_tablespace` 和 `temp_tablespaces` 设为未知的名称。这是因为设置要在其中使用的另一个数据库中它们可能是已知的；对表空间的情况，则是因为表空间可能尚未创建。此前 `search_path` 已被认识到有同样的问题，这些设置现在的行为与它一致。

- 修复 `INSERT` 表达式中的 `COLLATE` 导致的 “unsupported node type” 错误（Tom Lane）

- 避免在提交后删除表文件出现问题时崩溃（Tom Lane）

  删除表应只在事务提交后才删除底层磁盘文件。如果那时失败（例如因为文件权限错误），代码本应只发出警告消息并继续，因为此时中止事务已为时过晚。这一逻辑从 8.4 起被破坏，导致这类情况引发 PANIC 和无法重启的数据库。

- 从 `DROP TABLESPACE` 的 WAL 重放期间的错误中恢复（Tom Lane）

  重放会尝试删除表空间的目录，
  但这可能因各种原因失败（例如这些目录的所有权或权限不正确）。此前重放代码会 panic，使数据库在没有人工干预的情况下无法重启。记录问题并继续似乎更好，
  因为删除目录失败唯一后果只是浪费一些磁盘空间。

- 修复为热备记录 AccessExclusiveLock 时的竞态条件（Simon Riggs）

  有时锁会被记录为由“事务零”持有。至少已知这会在从库服务器上产生断言失败，
  还可能是更严重问题的原因。

- 在 WAL 重放期间正确跟踪 OID 计数器，即使它发生回卷（Tom Lane）

  此前 OID 计数器会一直停留在高值，直到系统退出重放模式。其实际后果通常为零，但在某些场景下，已提升为主库的备库服务器在需要 OID 值时，可能要很长时间才能把 OID 计数器推进到合理的值。

- 防止在崩溃恢复开始时发出误导性的“consistent recovery state reached”日志消息（Heikki Linnakangas）

- 修复 `pg_stat_replication.replay_location` 的初始值（Fujii Masao）

  此前，在至少重放一条 WAL 记录之前，显示的值是错误的。

- 修复带 `*` 的正则表达式反向引用（Tom Lane）

  代码实际上并不强制精确字符串匹配，
  而是接受任何满足反向引用符号所引用的模式子表达式的字符串。

  类似问题仍困扰着嵌入在更大带量词表达式中（而非量词直接作用对象）的反向引用。
  这将在未来的 PostgreSQL 版本中解决。

- 修复最近引入的 `inet/cidr` 值处理中的内存泄漏（Heikki Linnakangas）

  PostgreSQL 2011 年 12 月版本中的一个补丁导致这些操作中的内存泄漏，在此类列上构建 btree 索引等场景中可能相当显著。

- 修复规划器通过 `UNION ALL` 下推索引表达式限制的能力（Tom Lane）

  这类优化在 9.1.2 中被针对另一个问题的修复意外禁用。

- 修复继承表上 `UPDATE/DELETE` 中引用的 `WITH` 子句的规划（Tom Lane）

  该缺陷导致“could not find plan for CTE”失败。

- 修复 GIN 代价估计，使其处理 `column IN (...)` 索引条件（Marti Raudsepp）

  如果这类条件可用于 GIN 索引，这一疏忽通常会导致崩溃。

- 防止在会话带着打开的失败事务退出时断言失败（Tom Lane）

该缺陷对未启用断言的常规构建没有影响。

- 修复 SQL 语言函数中 `CREATE TABLE AS/SELECT INTO` 之后的悬空指针（Tom Lane）

  大多数情况下这只会在启用断言的构建中导致断言失败，但似乎可能有更糟的后果。

- 避免 Windows 上 syslogger 中文件句柄的双重关闭（MauMau）

  通常这个错误不可见，但在调试版 Windows 上运行时会导致异常。

- 修复 plpgsql 中与 I/O 转换相关的内存泄漏（Andres Freund、Jan Urbanski、Tom Lane）

  某些操作会一直泄漏内存直到当前函数结束。

- 规避 perl 的 SvPVutf8() 函数中的缺陷（Andrew Dunstan）

  该函数在收到类型通配或某些只读对象（如 `$^V`）时崩溃。让 plperl 避免把它们传给它。

- 在 pg_dump 中，如果扩展本身未被转储，则不转储其配置表的内容（Tom Lane）

- 改进 pg_dump 对继承表列的处理（Tom Lane）

  pg_dump 错误处理了子列与父列默认表达式不同的情况。如果默认值在文本上与父列相同但实际不同（例如因为模式搜索路径差异），它不会被识别为不同，
  因此转储和恢复后子列会被允许继承父列的默认值。子列为 `NOT NULL` 而父列不是的情况也可能被微妙地错误恢复。

- 修复 pg_restore 直接写数据库模式对 INSERT 式表数据的处理（Tom Lane）

  当使用 2011 年 9 月或 12 月版本的 pg_restore 时，对以 `--inserts` 或 `--column-inserts` 选项制作的归档文件直接恢复到数据库会失败，这是对另一个问题的修复中的一个疏忽所致。归档文件本身没有问题，文本模式输出也正常。

- 让 pg_upgrade 处理 plpython 共享库的重命名（Bruce Momjian）

  由于这一疏忽，升级包含 plpython 的 9.1 之前数据库会失败。

- 允许 pg_upgrade 处理包含 `regclass` 列的表（Bruce Momjian）

  由于 pg_upgrade 现在注意保留 `pg_class` OID，这一限制不再有任何理由存在。

- 让 libpq 在查找 SSL 客户端证书文件时忽略 `ENOTDIR` 错误（Magnus Hagander）

  即使用户的主目录被设置为 `/dev/null` 之类，这也能建立 SSL 连接（虽然没有证书）。

- 修复 ecpg 的 SQLDA 区域中更多的字段对齐问题（Zoltan Boszormenyi）

- 允许 ecpg 的 `DEALLOCATE` 语句中使用 `AT` 选项（Michael Meskes）

  支持它的基础设施已存在一段时间，但由于疏忽仍有一个错误检查拒绝这种情况。

- 在 ecpg 中定义 varchar 结构时不使用变量名（Michael Meskes）

- 修复 `contrib/auto_explain` 的 JSON 输出模式，使其产生有效的 JSON（Andrew Dunstan）

  输出在顶层使用了方括号，而本应使用花括号。

- 修复 `contrib/intarray` 的 `int[] & int[]` 操作符中的错误（Guillaume Lelarge）

  如果两个输入数组共有的最小整数是 1，且任一数组中有更小的值，那么 1 会被错误地从结果中遗漏。

- 修复 `contrib/pgcrypto` 的 `encrypt_iv()` 和 `decrypt_iv()` 中的错误检测（Marko Kreen）

  这些函数未能报告某些类型的无效输入错误，而是对错误输入返回随机垃圾值。

- 修复 `contrib/test_parser` 中的一字节缓冲区越界（Paul Guyot）

  代码会尝试多读一个字节，这在极端情况下会崩溃。由于 `contrib/test_parser` 只是示例代码，这本身不是安全问题，但坏的示例代码仍然是坏的。

- 在 ARM 上如果可用则使用 `__sync_lock_test_and_set()` 实现自旋锁（Martin Pitt）

  该函数取代了我们此前对 `SWPB` 指令的使用，该指令已被弃用且在 ARMv6 及更高版本上不可用。报告表明旧代码在较新的 ARM 板上不会以明显方式失败，而只是不对并发访问互锁，导致多进程操作中的离奇失败。

- 在使用接受该选项的 gcc 版本构建时使用 `-fexcess-precision=standard` 选项（Andrew Dunstan）

  这防止了较新版本 gcc 产生"创造性"结果的各种场景。

- 允许在 FreeBSD 上使用线程化 Python（Chris Rees）

  我们的 configure 脚本此前认为这种组合无法工作；但 FreeBSD 修复了问题，因此移除该错误检查。

- 允许 MinGW 构建使用标准命名的 OpenSSL 库（Tomasz Ostrowski）

E.23. 版本 9.1.2

发行日期
2011-12-05

本次发布包含对 9.1.1 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.23.1. 迁移到版本 9.1.2

对于运行 9.1.X 的用户，不需要进行转储/恢复。

不过，`information_schema.referential_constraints` 视图的定义中被发现存在一个长期错误。如果你依赖该视图的正确结果，应按下面第一条变更日志条目的说明替换其定义。

此外，如果你使用 `citext` 数据类型，并且是通过运行 pg_upgrade 从前一个大版本升级的，应执行 `CREATE EXTENSION citext FROM unpackaged`，以避免 `citext` 操作中与排序规则相关的失败。如果恢复来自包含 `citext` 数据类型的 9.1 之前数据库的转储，也需要这样做。如果你在升级到 9.1.2 之前已经运行过 `CREATE EXTENSION` 命令，则需要按下面第二条变更日志条目的说明手工更新系统目录。

E.23.2. 变更

- 修复 `information_schema.referential_constraints` 视图中的缺陷（Tom Lane）

  该视图在把外键约束与它依赖的主键或唯一键约束匹配时不够谨慎。
  这可能导致某个外键约束完全不被显示、被显示多次，
  或者声称它依赖于一个并非其实际依赖的约束。

  由于该视图定义是由 initdb 安装的，仅靠升级不会修复问题。如果需要修复现有安装，可以（以超级用户身份）删除 `information_schema` 模式，然后通过载入 *SHAREDIR*/`information_schema.sql` 重新创建它。（如果不确定 *SHAREDIR* 在哪里，运行 `pg_config --sharedir`。）这必须在每个要修复的数据库中重复执行。

- 让 `contrib/citext` 的升级脚本修复 `citext` 列和索引的排序规则（Tom Lane）

  在从 9.1 之前服务器执行 pg_upgrade 期间，或者当包含 `citext` 类型的 9.1 之前转储被载入 9.1 服务器时，现有的 `citext` 列和索引不会被正确标记为可排序数据类型。这导致这些列上的操作失败，报诸如 “could not determine which collation to use for string comparison” 的错误。此变更允许它们在 `CREATE EXTENSION citext FROM unpackaged` 期间由把 `citext` 模块升级为正规 9.1 扩展的同一脚本修复。

  如果你有一个此前升级的、正受此问题困扰的数据库，而且你已经运行过 `CREATE EXTENSION` 命令，可以（以超级用户身份）手工运行 *SHAREDIR*`/extension/citext--unpackaged--1.0.sql` 末尾的 `UPDATE` 命令。（如果不确定 *SHAREDIR* 在哪里，运行 `pg_config --sharedir`。）如果不确定，重复执行也没有害处。

- 修复连接到返回标量的函数输出的 `UPDATE` 或 `DELETE` 期间可能崩溃的问题（Tom Lane）

  只有当目标行被并发更新时才可能崩溃，因此该问题只是间歇性出现。

- 修复 GIN 索引更新 WAL 记录的错误重放（Tom Lane）

  这可能导致崩溃后或热备服务器上瞬时找不到索引项。不过，该问题会在该索引下一次 `VACUUM` 时被修复。

- 修复 `CREATE TABLE dest AS SELECT * FROM src` 或 `INSERT INTO dest SELECT * FROM src` 期间与 TOAST 相关的数据损坏（Tom Lane）

  如果表曾被 `ALTER TABLE ADD COLUMN` 修改过，在某些极端情况下，尝试把它的数据原样复制到另一个表可能产生损坏的结果。该问题只会在 8.4 及更高版本中以这种确切形式出现，但我们也给更早的版本打了补丁，以防有其他代码路径可能触发同一缺陷。

- 修复热备启动期间可能的失败（Simon Riggs）

- 当初始快照不完整时更快地启动热备（Simon Riggs）

- 修复从过期 syscache 条目访问 toast 表时的竞态条件（Tom Lane）

  典型症状是诸如 “missing chunk number 0 for toast value NNNNN in pg_toast_2619” 的瞬时错误，其中引用的 toast 表总是属于某个系统目录。

- 跟踪函数对参数默认值表达式中所用项的依赖（Tom Lane）

  此前，被引用的对象可以在未删除或修改函数的情况下被删除，导致函数被使用时行为异常。注意，仅仅安装此更新不会修复缺失的依赖条目；要做到这一点，

之后需要对每个这样的函数执行 `CREATE OR REPLACE`。如果你的函数默认值依赖于非内建对象，建议这样做。

- 修复 nestloop 连接中占位符变量的错误管理（Tom Lane）

  已知该缺陷会导致“variable not found in subplan target list”规划器错误，并且在涉及外连接时可能导致错误的查询输出。

- 修复按涉及聚合的表达式排序的窗口函数（Tom Lane）

  此前它们可能报“could not find pathkey item to sort”规划器错误。

- 修复“MergeAppend child's targetlist doesn't match MergeAppend”规划器错误（Tom Lane）

- 修复对同时具有可排序和不可排序输入的操作符的索引匹配（Tom Lane）

  在 9.1.0 中，左输入类型不可排序而右输入类型可排序的可索引操作符不会被识别为匹配左列的索引。例如 `hstore ? text` 操作符。

- 允许内联带多个 OUT 参数的集合返回 SQL 函数（Tom Lane）

- 不要为连接消除信任延迟唯一索引（Tom Lane、Marti Raudsepp）

  延迟唯一约束在事务内可能不成立，因此假设它成立可能给出错误的查询结果。

- 让 `DatumGetInetP()` 解包带 1 字节头的 inet datum，并新增一个不做此事的宏 `DatumGetInetPP()`（Heikki Linnakangas）

  此变更不影响任何核心代码，但可能避免期望 `DatumGetInetP()` 按常规约定产生已解包 datum 的附加代码中的崩溃。

- 改进 `money` 类型输入和输出的区域支持（Tom Lane）

  除了不支持所有标准 `lc_monetary` 格式化选项外，输入和输出函数还不一致，这意味着在某些区域设置下转储的 `money` 值无法重新读取。

- 不让 `transform_null_equals` 影响 `CASE foo WHEN NULL ...` 构造（Heikki Linnakangas）

  `transform_null_equals` 只应影响用户直接写的 `foo = NULL` 表达式，而不是这种形式的 `CASE` 在内部生成的相等性检查。

- 更改外键触发器的创建顺序，以更好地支持自引用外键（Tom Lane）

  对于引用自身表的级联外键，行更新会作为一个事件同时触发 `ON UPDATE` 触发器和 `CHECK` 触发器。`ON UPDATE` 触发器必须先执行，否则 `CHECK` 会检查行的非最终状态，并可能抛出不恰当的错误。然而这些触发器的触发顺序由其名称决定，而由于触发器使用遵循“RI_ConstraintTrigger_NNNN”约定的自动生成名称，
  它们通常按创建顺序排序。彻底的修复需要修改该约定（我们将在 9.2 中做），但在现有版本中更改它似乎有风险。因此此补丁只更改触发器的创建顺序。
  遇到这类错误的用户应删除并重新创建外键约束，使其触发器处于正确顺序。

- 修复 `IF EXISTS` 在 `DROP OPERATOR FAMILY` 中的正确工作（Robert Haas）

- 禁止从扩展自己的脚本内删除该扩展（Tom Lane）

这防止了扩展依赖管理不正确时的怪异行为。

- 不把自动生成的类型标记为扩展成员（Robert Haas）

  关系行类型和自动生成的数组类型不需要在 `pg_depend` 中有自己的扩展成员条目，创建这类条目会使扩展升级复杂化。

- 在 `CREATE EXTENSION` 期间应对无效的既有 `search_path` 设置（Tom Lane）

- 避免在跟踪缓冲区分配率时发生浮点下溢（Greg Matthews）

  虽然这本身无害，但在某些平台上会导致烦人的内核日志消息。

- 防止自动清理事务以可序列化模式运行（Tom Lane）

  自动清理此前使用集簇范围的默认事务隔离级别，但它没有必要使用高于 READ COMMITTED 的级别，而使用 SERIALIZABLE 可能给其他进程带来不必要的延迟。

- 确保 walsender 进程及时响应 SIGTERM（Magnus Hagander）

- 把 `postmaster.opts` 排除在基础备份之外（Magnus Hagander）

- 在 Windows 下启动子进程时保留配置文件名和行号值（Tom Lane）

  此前它们不会在 `pg_settings` 视图中正确显示。

- 修复 ecpg 的 SQLDA 区域中不正确的字段对齐（Zoltan Boszormenyi）

- 在 psql 的命令历史中保留命令内的空行（Robert Haas）

  例如，如果从字符串字面量中删除了空行，先前的行为可能导致问题。

- 避免 pg_dump 中平台特定的无限循环（Steve Singer）

- 修复 pg_dump 纯文本输出格式的压缩（Adrian Klaver、Tom Lane）

  历史上，pg_dump 把不带 `-F` 开关的 `-Z` 理解为应输出纯文本输出的 gzip 压缩版本。现已恢复该行为。

- 修复 pg_dump，使其转储自动生成类型（如表行类型）之间的用户定义转换（Tom Lane）

- 修复 pg_dump 中外部服务器名称漏加引号的问题（Tom Lane）

- 对 pg_upgrade 的若干修复（Bruce Momjian）

  正确处理排他约束、避免 Windows 上的失败、不再对 8.4 数据库中不匹配的 toast 表名报错。

- 在 PL/pgSQL 中，允许外部表定义行类型（Alexander Soudakov）

- 修正 PL/Perl 函数结果的转换（Alex Hunsaker、Tom Lane）

  恢复 9.1 之前的行为：返回 `void` 的 PL/Perl 函数忽略其最后一条 Perl 语句的结果值；9.1.0 在该语句返回引用时会抛出错误。同时，确保可以为组合类型返回字符串值，只要该字符串符合该类型的输入格式。此外，当函数声明的结果类型不是数组或组合类型时，
  对分别尝试返回 Perl 数组或哈希的做法抛出错误。（9.1 之前的版本在这种情况下会返回诸如 `ARRAY(0x221a9a0)` 或 `HASH(0x221aa90)` 之类毫无用处的字符串。）

- 确保 PL/Perl 字符串总是正确地以 UTF8 编码（Amit Khandekar、Alex Hunsaker）

- 使用首选版本的 xsubpp 构建 PL/Perl，而不一定是操作系统的主副本（David Wheeler、Alex Hunsaker）
- 在 PL/Python 异常中正确传播 SQLSTATE（Mika Eloranta、Jan Urbanski）
- 不为构建所针对的 Python 大版本之外的其他版本安装 PL/Python 扩展文件（Peter Eisentraut）
- 更改所有 `contrib` 扩展脚本文件，使其在被喂给 psql 时报告有用的错误消息（Andrew Dunstan、Tom Lane）

  这应有助于让人们了解使用 `CREATE EXTENSION` 加载这些文件的新方法。大多数情况下，直接载入脚本本来就会失败，只是错误消息更难理解。
- 修复 `contrib/dict_int` 和 `contrib/dict_xsyn` 中的错误编码（Tom Lane）

  一些函数错误地假定 `palloc()` 返回的内存保证被清零。
- 把 `contrib/sepgsql` 测试从常规回归测试机制中移除（Tom Lane）

  由于这些测试的设置需要 root 权限，自动运行不现实。改为手工方式，并提供一个测试脚本来协助。
- 修复 `contrib/unaccent` 配置文件解析中的若干错误（Tom Lane）
- 让 `pgstatindex()` 及时响应查询取消中断（Robert Haas）
- 修复 macOS 启动脚本中日志文件名的不正确引号（Sidar Lopez）
- 撤销无意中启用的 `WAL_DEBUG`（Robert Haas）

  幸运的是，就调试工具而言，这个相当廉价；但它并不打算默认启用，因此撤销。
- 确保 VPATH 构建正确安装所有服务器头文件（Peter Eisentraut）
- 缩短详细错误消息中报告的文件名（Peter Eisentraut）

  常规构建一直只报告包含错误消息调用的 C 文件的名称，但 VPATH 构建此前报告绝对路径名。
- 修复中美洲 Windows 时区名的解释（Tom Lane）

  把 “Central America Standard Time” 映射为 `CST6` 而不是 `CST6CDT`，因为中美洲各地通常不实行夏令时。
- 把时区数据文件更新为 tzdata 版本 2011n for DST law changes in Brazil, Cuba, Fiji, Palestine, Russia, and Samoa; also historical corrections for Alaska and British East Africa。

E.24. 版本 9.1.1

发行日期
2011-09-26

本次发布包含对 9.1.0 的多方面修复。有关 9.1 大版本中新特性的信息，请参见第 E.25 节。.

E.24.1. 迁移到版本 9.1.1

对于运行 9.1.X 的用户，不需要进行转储/恢复。

E.24.2. 变更

- 让 `pg_options_to_table` 对没有值的选项返回 NULL（Tom Lane）

  此前这类情况会导致服务器崩溃。

- 修复 GiST 索引扫描结束时的内存泄漏（Tom Lane）

  执行许多独立 GiST 索引扫描的命令（例如对已包含许多行的表验证新的基于 GiST 的排他约束）会因此泄漏瞬时需要大量内存。

- 修复 `CREATE TEMPORARY TABLE` 中对 `pg_temp` 模式的显式引用（Robert Haas）

  这过去是允许的，但在 9.1.0 中失败。

E.25. 版本 9.1

发行日期
2011-09-12

E.25.1. 概述

本版本表明 PostgreSQL 正凭借 PostgreSQL 独有的全新突破性功能，超越传统关系数据库的功能集。9.0 引入的流复制特性通过新增同步复制选项、流式备份和监控改进得到显著增强。主要增强包括：

- 允许同步复制
- 新增对外部表的支持
- 新增逐列排序规则支持
- 新增扩展，简化 PostgreSQL 附加功能的打包
- 新增真正的可序列化隔离级别
- 支持在 `CREATE TABLE` 中使用 `UNLOGGED` 选项的不记录日志的表
- 允许在 `WITH` 子句中使用数据修改命令（`INSERT`/`UPDATE`/`DELETE`）
- 为 GiST 索引新增最近邻（按操作符排序）搜索
- 新增 `SECURITY LABEL` 命令以及对 SELinux 权限控制的支持

- 更新 PL/Python 服务器端语言

上述各项在下面的章节中有更详细的说明。

E.25.2. 迁移到版本 9.1

对于希望从任何早期版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复，或者使用 pg_upgrade。

版本 9.1 包含许多可能影响与以前版本兼容性的变更。请注意以下不兼容之处：

E.25.2.1. 字符串

- 把 `standard_conforming_strings` 的默认值改为 on（Robert Haas）

  默认情况下，反斜杠现在是字符串字面量中的普通字符，而不是转义字符。
  此变更消除了一个与 SQL 标准长期存在的不兼容。`escape_string_warning` 多年来一直对这种用法发出警告。`E''` 字符串是在字符串中嵌入反斜杠转义的正确方式，不受此变更影响。

> **警告**
>
> 此变更可能破坏未预料到它并按旧规则自行做字符串转义的应用。后果可能严重到引入 SQL 注入安全漏洞。务必测试暴露于不可信输入的应用，确保它们正确处理文本字符串中的单引号和反斜杠。

E.25.2.2. 类型转换

- 禁止对组合类型使用函数式和属性式数据类型转换（Tom Lane）

  例如，禁止 *`composite_value`*`.text` 和 `text(`*`composite_value`*`)`。无意中使用这种语法常常导致缺陷报告；虽然这不是缺陷，但回到拒绝这类表达式似乎更好。
  当确实想要转换整个组合值时，`CAST` 和 `::` 语法仍然可用。

- 收紧对基于数组的域的转换检查（Tom Lane）

  当域基于数组类型时，允许“透过”域类型访问数组元素，
  包括对域值做下标操作来获取或赋值元素。对这种域值的元素赋值（例如通过 `UPDATE ... SET domaincol[5] = ...`）现在会重新检查域类型的约束，而此前会跳过检查。

E.25.2.3. 数组

- 更改 `string_to_array()`，使其对零长度字符串返回空数组（Pavel Stehule）

  此前这返回空值。

- 更改 `string_to_array()`，使 `NULL` 分隔符把字符串拆分为字符（Pavel Stehule）

  此前这返回空值。

E.25.2.4. 对象修改

- 修复对 BEFORE/AFTER 触发器的不当检查（Tom Lane）

触发器现在可以在三种情况下触发：某动作的 BEFORE、AFTER 或 INSTEAD OF。触发器函数作者应验证其逻辑在三种情况下都表现正常。

- 要求超级用户或 CREATEROLE 权限才能为角色设置注释（Tom Lane）

E.25.2.5. 服务器设置

- 更改 pg_last_xlog_receive_location()，使其永不倒退（Fujii Masao）

  此前，当流复制重启时，pg_last_xlog_receive_location() 的值可能倒退。

- 让复制连接的日志记录遵循 log_connections（Magnus Hagander）

  此前复制连接总是被记录。

E.25.2.6. PL/pgSQL Server-Side Language

- 更改 PL/pgSQL 的无参数 RAISE 命令，使其可被所附加的异常块捕获（Piyush Newe）

  此前代码块中的 RAISE 总是作用于所附加的异常块，因此在同一作用域内无法捕获。

- 调整 PL/pgSQL 的错误行号代码，使其与其他过程语言一致（Pavel Stehule）

  此前 PL/pgSQL 会忽略（不计数）函数体开头的空行。由于这与其他所有语言不一致，该特殊情况已被移除。

- 让 PL/pgSQL 对 IN 和 OUT 参数名冲突报错（Tom Lane）

  此前不会检测冲突，该名称会静默地只指向 OUT 参数。

- PL/pgSQL 变量的类型修饰符现在对 SQL 解析器可见（Tom Lane）

  附加到 PL/pgSQL 变量的类型修饰符（如 varchar 长度限制）此前只在赋值期间强制执行，其他用途一律忽略。这类变量现在的行为将更像是用相同修饰符声明的表列。
  预计在大多数情况下这不会造成可见差异，但可能给 PL/pgSQL 函数发出的某些 SQL 命令带来细微变化。

E.25.2.7. 附加模块

- 所有 contrib 模块现在都通过 CREATE EXTENSION 安装，而不是手工调用其 SQL 脚本（Dimitri Fontaine、Tom Lane）

  要更新包含某 contrib 模块 9.0 版本的现有数据库，使用 CREATE EXTENSION ... FROM unpackaged 把现有 contrib 模块的对象包装成扩展。从 9.0 之前的版本更新时，先用旧的卸载脚本删除该 contrib 模块的对象，然后使用 CREATE EXTENSION。

E.25.2.8. 其他不兼容之处

- 让 pg_stat_reset() 重置所有数据库级统计信息（Tomas Vondra）

  一些 pg_stat_database 计数器此前没有被重置。

- 修正一些 information_schema.triggers 列名，使其与新的 SQL 标准名称一致（Dean Rasheed）

- 把 ECPG 游标名视为大小写不敏感（Zoltan Boszormenyi）

E.25.3. 变更

下面是 PostgreSQL 9.1 与上一个主版本之间变更的详细说明。

E.25.3.1. 服务器

E.25.3.1.1. 性能

- 支持在 `CREATE TABLE` 中使用 `UNLOGGED` 选项的不记录日志的表（Robert Haas）

  这类表比常规表提供更好的更新性能，
  但不是崩溃安全的：服务器崩溃时其内容会被自动清除。其内容也不会传播到复制从库。

- 允许把 `FULL OUTER JOIN` 实现为哈希连接，并允许对 `LEFT OUTER JOIN` 或 `RIGHT OUTER JOIN` 的任一侧做哈希（Tom Lane）

  此前 `FULL OUTER JOIN` 只能实现为合并连接，而 `LEFT OUTER JOIN` 和 `RIGHT OUTER JOIN` 只能对连接的可空侧做哈希。这些变更提供了额外的查询优化可能性。

- 合并重复的 fsync 请求（Robert Haas、Greg Smith）

  这大幅改进了重写负载下的性能。

- 改进 `commit_siblings` 的性能（Greg Smith）

  这使 `commit_siblings` 的使用开销更低。

- 降低大型 ispell 词典的内存需求（Pavel Stehule、Tom Lane）

- 避免“盲写”之后数据文件保持打开（Alvaro Herrera）

  这修复了后端可能在文件被删除后很久仍持有其打开状态、阻止内核回收磁盘空间的场景。

E.25.3.1.2. 优化器

- 允许继承表扫描返回有意义的排序结果（Greg Stark、Hans-Jurgen Schonig、Robert Haas、Tom Lane）

  这允许对继承表上使用 `ORDER BY`、`LIMIT` 或 `MIN/MAX` 的查询做更好的优化。

- 改进 GIN 索引扫描的代价估计（Teodor Sigaev）

- 改进聚合和窗口函数的代价估计（Tom Lane）

E.25.3.1.3. 认证

- 在 `pg_hba.conf` 中支持主机名和主机后缀（例如 `.example.com`）（Peter Eisentraut）

  此前只支持主机 IP 地址和 CIDR 值。

- 在 `pg_hba.conf` 的 host 列中支持关键字 `all`（Peter Eisentraut）

  此前人们为此使用 `0.0.0.0/0` 或 `::/0`。

- 在不支持 Unix 套接字连接的平台上拒绝 `pg_hba.conf` 中的 `local` 行（Magnus Hagander）

  此前这类行会被静默忽略，这可能令人意外。这使该行为更像其他不受支持的情况。

- 允许使用 GSSAPI 通过 SSPI 向服务器认证（Christian Ullrich）

  具体而言，这允许基于 Unix 的 GSSAPI 客户端与 Windows 服务器进行 SSPI 认证。

- 本地套接字上的 `ident` 认证现在称为 `peer`（Magnus Hagander）

  为了向后兼容仍接受旧名称，但由于这两种方法根本不同，为它们采用不同的名称似乎更好。

- 重写 peer 认证，避免使用凭据控制消息（Tom Lane）

  此变更使 peer 认证代码更简单、性能更好。但它要求平台提供 `getpeereid` 函数或等效的套接字操作。据我们所知，唯一此前 peer 认证可用而现在不可用的平台是 5.0 之前的 NetBSD。

E.25.3.1.4. 监控

- 为重启点和检查点的日志记录添加细节，它由 `log_checkpoints` 控制（Fujii Masao、Greg Smith）

  新增细节包括 WAL 文件和同步活动。

- 新增 `log_file_mode`，控制日志收集器创建的日志文件的权限（Martin Pihlak）

- 把 syslog 日志的默认最大行长降低到 900 字节加前缀（Noah Misch）

  这避免了在有 1KB（而非更常见的 2KB）长度限制的 syslog 实现上长日志行被截断。

E.25.3.1.5. 统计视图

- 为 `pg_stat_activity` 新增 `client_hostname` 列（Peter Eisentraut）

  此前只报告客户端地址。

- 新增 `pg_stat_xact_*` 统计函数和视图（Joel Jacobson）

  它们类似于数据库范围的统计计数器视图，但只反映当前事务的计数。

- 在数据库级和后台写入进程统计视图中新增上次重置时间（Tomas Vondra）

- 在 `pg_stat_*_tables` 视图中新增显示清理和分析操作次数的列（Magnus Hagander）

- 为 `pg_stat_bgwriter` 新增 `buffers_backend_fsync` 列（Greg Smith）

  这个新列统计后端对缓冲区执行 fsync 的次数。

E.25.3.1.6. 服务器设置

- 提供 `wal_buffers` 的自动调整（Greg Smith）

  默认情况下，`wal_buffers` 的值现在根据 `shared_buffers` 的值自动选择。

- 提高 `deadlock_timeout`、`log_min_duration_statement` 和 `log_autovacuum_min_duration` 的最大值（Peter Eisentraut）

这些参数此前的最大值都只有约 35 分钟。现在允许大得多的值。

E.25.3.2. 复制和恢复

E.25.3.2.1. 流复制和连续归档

- 允许同步复制（Simon Riggs、Fujii Masao）

  这允许主服务器等待备库把事务信息写入磁盘后再确认提交。
  每次只有一个备库可以充当同步备库的角色，由 `synchronous_standby_names` 设置控制。
  可以使用 `synchronous_commit` 设置按事务启用或禁用同步复制。

- 新增协议支持，使用流复制网络连接向备库服务器发送文件系统备份（Magnus Hagander、Heikki Linnakangas）

  这避免了搭建备库服务器时手工传输文件系统备份的要求。

- 新增 `replication_timeout` 设置（Fujii Masao、Heikki Linnakangas）

  空闲时间超过 `replication_timeout` 间隔的复制连接将被自动终止。此前，失败的连接通常要等到 TCP 超时才会被检测到，这在许多情况下长得不方便。

- 新增命令行工具pg_basebackup，用于创建新的备库服务器或数据库备份（Magnus Hagander）

- 为角色新增复制权限（Magnus Hagander）

  这是一个用于流复制的只读权限。它允许非超级用户角色用于复制连接。
  此前只有超级用户可以发起复制连接；超级用户默认仍拥有此权限。

E.25.3.2.2. 复制监控

- 新增系统视图 `pg_stat_replication`，显示 WAL 发送端进程的活动（Itagaki Takahiro、Simon Riggs）

  它报告所有已连接备库服务器的状态。

- 新增监控函数 `pg_last_xact_replay_timestamp()`（Fujii Masao）

  它返回主库生成备库上已应用的最新的提交或中止记录的时间。

E.25.3.2.3. 热备

- 新增配置参数 `hot_standby_feedback`，使备库可以推迟主库上旧行版本的清理（Simon Riggs）

  这有助于避免取消备库上的长时间运行查询。

- 新增系统视图 `pg_stat_database_conflicts`，显示已被取消的查询及其原因（Magnus Hagander）

  取消可能因表空间被删除、锁超时、旧快照、被固定的缓冲区和死锁而发生。

- 为 `pg_stat_database` 新增 `conflicts` 计数（Magnus Hagander）

  这是该数据库中发生的冲突数量。

- 提高 max_standby_archive_delay 和 max_standby_streaming_delay 的最大值

  这些参数此前的最大值都只有约 35 分钟。现在允许大得多的值。

- 新增错误码 ERRCODE_T_R_DATABASE_DROPPED，用于报告因数据库被删除导致的恢复冲突（Tatsuo Ishii）

  这对连接池软件很有用。

E.25.3.2.4. 恢复控制

- 新增控制流复制重放的函数（Simon Riggs）

  新函数为 pg_xlog_replay_pause()、pg_xlog_replay_resume() 和状态函数 pg_is_xlog_replay_paused()。

- 新增 recovery.conf 设置 pause_at_recovery_target，用于在目标处暂停恢复（Simon Riggs）

  这允许查询恢复服务器以检查恢复点是否是期望的那个。

- 新增使用 pg_create_restore_point() 创建命名还原点的能力（Jaime Casanova）

  这些命名还原点可以通过新的 recovery.conf 设置 recovery_target_name 指定为恢复目标。

- 允许备库恢复自动切换到新时间线（Heikki Linnakangas）

  现在备库服务器会定期扫描归档目录以发现新时间线。

- 新增 restart_after_crash 设置，用于在后端崩溃后禁用自动服务器重启（Robert Haas）

  这允许外部集簇管理软件控制数据库服务器是否重启。

- 允许 recovery.conf 使用与 postgresql.conf 相同的引号行为（Dimitri Fontaine）

  此前所有值都必须加引号。

E.25.3.3. 查询

- 新增真正的可序列化隔离级别（Kevin Grittner、Dan Ports）

  此前，请求可序列化隔离只保证整个事务使用单个 MVCC 快照，这允许某些已记载的异常。旧的快照隔离行为仍可通过请求 REPEATABLE READ 隔离级别获得。

- 允许在 WITH 子句中使用数据修改命令（INSERT/UPDATE/DELETE）（Marko Tiikkaja、Hitoshi Harada）

  这些命令可以使用 RETURNING 把数据传递给包含它们的查询。

- 允许把 WITH 子句附加到 INSERT、UPDATE、DELETE 语句（Marko Tiikkaja、Hitoshi Harada）

- 当在 GROUP BY 子句中指定了主键时，允许查询目标列表中出现非 GROUP BY 列（Peter Eisentraut）

SQL 标准允许这种行为，而且由于有主键，结果是无歧义的。

- 允许在 `UNION/INTERSECT/EXCEPT` 子句中使用关键字 `DISTINCT`（Tom Lane）

  `DISTINCT` 是默认行为，因此使用该关键字是冗余的，但 SQL 标准允许。

- 修复带规则的普通查询，使其使用与 `EXPLAIN ANALYZE` 相同的快照行为（Marko Tiikkaja）

  此前 `EXPLAIN ANALYZE` 对涉及规则的查询使用的快照时机略有不同。`EXPLAIN ANALYZE` 的行为被认为更合逻辑。

E.25.3.3.1. 字符串

- 新增逐列排序规则支持（Peter Eisentraut、Tom Lane）

  此前排序规则（文本字符串的排序顺序）只能在创建数据库时选择。现在可以通过 SQL 标准的 `COLLATE` 子句逐列、逐域、逐索引或逐表达式设置排序规则。

E.25.3.4. 对象操作

- 新增扩展，简化 PostgreSQL 附加功能的打包（Dimitri Fontaine、Tom Lane）

  扩展由新的 `CREATE/ALTER/DROP EXTENSION` 命令控制。这取代了对添加到 PostgreSQL 安装的对象进行分组的临时方法。

- 新增对外部表的支持（Shigeru Hanada、Robert Haas、Jan Urbanski、Heikki Linnakangas）

  这允许像使用 PostgreSQL 原生存储的数据一样使用存储在数据库之外的数据。但外部表目前是只读的。

- 允许通过 `ALTER TYPE` 向现有枚举类型添加新值（Andrew Dunstan）

- 新增 `ALTER TYPE ... ADD/DROP/ALTER/RENAME ATTRIBUTE`（Peter Eisentraut）

  这允许修改组合类型。

E.25.3.4.1. `ALTER` Object

- 为类型化表上的 `ALTER TYPE` 操作新增 `RESTRICT/CASCADE`（Peter Eisentraut）

  这控制 `ADD/DROP/ALTER/RENAME ATTRIBUTE` 的级联行为。

- 支持 `ALTER TABLE` *`name`* `{OF | NOT OF}` *`type`*（Noah Misch）

  该语法允许把独立表变成类型化表，或把类型化表变成独立表。

- 为 `ALTER ... SET SCHEMA` 命令中的更多对象类型新增支持（Dimitri Fontaine）

  该命令现在支持转换、操作符、操作符类、操作符族、文本搜索配置、文本搜索词典、文本搜索解析器和文本搜索模板。

E.25.3.4.2. `CREATE/ALTER TABLE`

- 新增 `ALTER TABLE ... ADD UNIQUE/PRIMARY KEY USING INDEX`（Gurjeet Singh）

这允许使用现有唯一索引（包括并发创建的唯一索引）来定义主键或唯一约束。

- 允许 `ALTER TABLE` 添加未经验证的外键（Simon Riggs）

  新选项名为 `NOT VALID`。约束的状态之后可以改为 `VALIDATED` 并执行验证检查。这两者相结合，使你能够以对读写操作影响最小的方式添加外键。

- 允许 `ALTER TABLE ... SET DATA TYPE` 在适当情况下避免重写表（Noah Misch、Robert Haas）

  例如，把 `varchar` 列转换为 `text` 不再需要重写表。但增大 `varchar` 列的长度约束仍需要重写表。

- 新增 `CREATE TABLE IF NOT EXISTS` 语法（Robert Haas）

  这允许在表已存在时创建表而不报错。

- 修复当两个后端同时尝试向同一个表添加继承子表时可能出现的 “tuple concurrently updated” 错误（Robert Haas）

  `ALTER TABLE` 现在对父表持有更强的锁，使会话无法同时尝试更新它。

E.25.3.4.3. 对象权限

- 新增 `SECURITY LABEL` 命令（KaiGai Kohei）

  这允许为对象分配安全标签。

E.25.3.5. 工具操作

- 新增事务级咨询锁（Marko Tiikkaja）

  它们类似于现有的会话级咨询锁，但这类锁会在事务结束时自动释放。

- 让 `TRUNCATE ... RESTART IDENTITY` 以事务方式重启序列（Steve Singer）

  此前，如果后端在提交时截断活动与提交完成之间崩溃，计数器可能失去同步。

E.25.3.5.1. COPY

- 为 `COPY TO/FROM` 新增 `ENCODING` 选项（Hitoshi Harada、Itagaki Takahiro）

  这允许单独指定 `COPY` 文件的编码，而不是跟随客户端编码。

- 新增双向 `COPY` 协议支持（Fujii Masao）

  这目前只被流复制使用。

E.25.3.5.2. EXPLAIN

- 让 `EXPLAIN VERBOSE` 显示 `FunctionScan` 节点中的函数调用表达式（Tom Lane）

E.25.3.5.3. VACUUM

- 为 `VACUUM FULL VERBOSE` 和 `CLUSTER VERBOSE` 的输出添加更多细节（Itagaki Takahiro）

新信息包括存活和死亡元组计数，以及 `CLUSTER` 是否使用索引重建。

- 防止自动清理在无法获得表锁时等待（Robert Haas）

  它会在稍后再次尝试清理该表。

E.25.3.5.4. CLUSTER

- 当看起来更廉价时，允许 `CLUSTER` 对表排序而不是扫描索引（Leonardo Francalanci）

E.25.3.5.5. 索引

- 为 GiST 索引新增最近邻（按操作符排序）搜索（Teodor Sigaev、Tom Lane）

  这允许 GiST 索引在带 `LIMIT` 的查询中快速返回最接近的 *N* 个值。例如

  ```
  SELECT * FROM places ORDER BY location <-> point '(101,456)' LIMIT
  10;
  ```

  找出最接近给定目标点的十个地点。

- 允许 GIN 索引索引空值和空字符串（Tom Lane）

  这允许完整的 GIN 索引扫描，并修复了 GIN 扫描会失败的各种极端情况。

- 让 GIN 索引更好地识别重复的搜索条目（Tom Lane）

  这降低了索引扫描的代价，尤其是在能避免不必要的全索引扫描的情况下。

- 修复 GiST 索引，使其完全崩溃安全（Heikki Linnakangas）

  此前有少数情况需要执行 `REINDEX`（你会被告知）。

E.25.3.6. 数据类型

- 允许 `numeric` 在常见情况下使用更紧凑的两字节头（Robert Haas）

  此前所有 `numeric` 值都使用四字节头；此变更节省了磁盘存储。

- 新增 `money` 除以 `money` 的支持（Andy Balholm）

- 允许 `void` 类型的二进制 I/O（Radoslaw Smogura）

- 改进几何操作符的斜边计算（Paul Matthews）

  这避免了不必要的溢出，而且可能更精确。

- 支持对数组值做哈希（Tom Lane）

  这提供了额外的查询优化可能性。

- 除非组合类型的所有列类型都可排序，否则不把该组合类型视为可排序（Tom Lane）

  如果查询可以在不排序的情况下实现，这避免了运行时可能出现的 “could not identify a comparison function” 失败。此外，`ANALYZE` 也不会尝试对这类组合类型的列使用不适当的统计收集方法。

E.25.3.6.1. 类型转换

- 新增 money 与 numeric 之间转换的支持（Andy Balholm）
- 新增从 int4 和 int8 到 money 转换的支持（Joey Adams）
- 如果表是类型化表，允许把表的行类型转换为该表的超类型（Peter Eisentraut）

  这类似于允许把行类型转换为超表的行类型的现有机制。

E.25.3.6.2. XML

- 新增 XML 函数 XMLEXISTS 和 xpath_exists()（Mike Fowler）

  它们用于 XPath 匹配。

- 新增 XML 函数 xml_is_well_formed()、xml_is_well_formed_document() 和 xml_is_well_formed_content()（Mike Fowler）

  它们检查输入是否为格式正确的 XML。它们提供了此前只能在已弃用的 contrib/xml2 模块中使用的功能。

E.25.3.7. 函数

- 新增 SQL 函数 format(text, ...)，其行为类似于 C 的 printf()（Pavel Stehule、Robert Haas）

  它目前支持字符串、SQL 字面量和 SQL 标识符的格式。

- 新增字符串函数 concat()、concat_ws()、left()、right() 和 reverse()（Pavel Stehule）

  这些改进了与其他数据库产品的兼容性。

- 新增用于读取二进制文件的函数 pg_read_binary_file()（Dimitri Fontaine、Itagaki Takahiro）
- 新增函数 pg_read_file() 的单参数版本，用于读取整个文件（Dimitri Fontaine、Itagaki Takahiro）
- 新增 array_to_string() 和 string_to_array() 的三参数形式，用于控制空值处理（Pavel Stehule）

E.25.3.7.1. 对象信息函数

- 新增函数 pg_describe_object()（Alvaro Herrera）

  该函数用于根据 pg_class OID、对象 OID 和子对象 ID 获取描述对象的人类可读字符串。它可用于帮助解释 pg_depend 的内容。

- 更新内建操作符及其底层函数的注释（Tom Lane）

  旨在通过关联操作符使用的函数现在被这样注释。

- 新增变量 quote_all_identifiers，强制为 EXPLAIN 以及 pg_get_viewdef() 等系统目录函数中的所有标识符加引号（Robert Haas）

这使得把模式导出到具有不同引号规则的工具和其他数据库更加容易。

- 为 `information_schema.sequences` 系统视图新增列（Peter Eisentraut）

  此前虽然视图存在，但关于序列参数的列未实现。

- 允许在 `has_table_privilege()` 及相关函数中使用 `public` 作为伪角色名（Alvaro Herrera）

  这允许检查公共权限。

E.25.3.7.2. 函数和触发器创建

- 支持视图上的 `INSTEAD OF` 触发器（Dean Rasheed）

  此特性可用于实现完全可更新的视图。

E.25.3.8. 服务器端语言

E.25.3.8.1. PL/pgSQL Server-Side Language

- 为 PL/pgSQL 新增 `FOREACH IN ARRAY`（Pavel Stehule）

  这比此前遍历数组值元素的方法更高效也更易读。

- 允许在同一位置能够捕获 `RAISE ERROR` 的地方捕获无参数的 `RAISE`（Piyush Newe）

  先前的编码会从包含活动异常处理程序的块抛出错误。新行为与其他 DBMS 产品更一致。

E.25.3.8.2. PL/Perl Server-Side Language

- 允许 PL/Perl 函数接受通用 record 参数（Andrew Dunstan）

  PL/Perl 函数现在可以声明为接受 `record` 类型。其行为与任何命名的组合类型相同。

- 把 PL/Perl 的数组参数转换为 Perl 数组（Alexey Klyukin、Alex Hunsaker）

  字符串表示仍然可用。

- 把 PL/Perl 的组合类型参数转换为 Perl 哈希（Alexey Klyukin、Alex Hunsaker）

  字符串表示仍然可用。

E.25.3.8.3. PL/Python Server-Side Language

- 为 PL/Python 新增表函数支持（Jan Urbanski）

  PL/Python 现在可以返回多个 `OUT` 参数和记录集。

- 为 PL/Python 新增验证器（Jan Urbanski）

  这允许在创建函数时对 PL/Python 函数做语法检查。

- 允许 PL/Python 中 SQL 查询的异常（Jan Urbanski）

  这允许从 PL/Python 异常块访问 SQL 生成的异常错误码。

- 为 PL/Python 新增显式子事务（Jan Urbanski）
- 为 PL/Python 新增字符串加引号函数（Jan Urbanski）

  这些函数是 `plpy.quote_ident`、`plpy.quote_literal` 和 `plpy.quote_nullable`。
- 为 PL/Python 错误添加回溯信息（Jan Urbanski）
- 用 `PLy_elog` 报告迭代器的 PL/Python 错误（Jan Urbanski）
- 修复 Python 3 的异常处理（Jan Urbanski）

  此前在 Python 3 下 `plpy` 中没有异常类。

E.25.3.9. 客户端应用程序

- 由于createlang和droplang现在只是调用扩展命令，将其标记为已弃用（Tom Lane）

E.25.3.9.1. psql

- 新增 psql 命令 `\conninfo`，用于显示当前连接信息（David Christensen）
- 新增 psql 命令 `\sf`，用于显示函数定义（Pavel Stehule）
- 新增 psql 命令 `\dL`，用于列出语言（Fernando Ike）
- 为 psql 的 `\dn`（列出模式）命令新增 `S`（“system”）选项（Tom Lane）

  不带 `S` 的 `\dn` 现在会隐藏系统模式。
- 允许 psql 的 `\e` 和 `\ef` 命令接受行号以定位编辑器中的光标（Pavel Stehule）

  该行号根据 `PSQL_EDITOR_LINENUMBER_ARG` 环境变量传递给编辑器。
- 让 psql 默认根据操作系统区域设置客户端编码（Heikki Linnakangas）

  这只在未设置 `PGCLIENTENCODING` 环境变量时发生。
- 让 `\d` 区分唯一索引和唯一约束（Josh Kupershmidt）
- 让 `\dt+` 在连接 9.0 或更高版本服务器时报告 `pg_table_size` 而不是 `pg_relation_size`（Bernd Helmle）

  这是更有用的表大小度量，但注意它与此前同一显示中报告的值并不相同。
- 更多 Tab 补全支持（Itagaki Takahiro、Pavel Stehule、Andrey Popp、Christoph Berg、David Fetter、Josh Kupershmidt）

E.25.3.9.2. pg_dump

- 为pg_dump和pg_dumpall新增选项 `--quote-all-identifiers`，强制为所有标识符加引号（Robert Haas）
- 为 pg_dump 新增 `directory` 格式（Joachim Wieland、Heikki Linnakangas）

  它在内部类似于 pg_dump 的 `tar` 格式。

E.25.3.9.3. pg_ctl

- 修复 pg_ctl，使其不再错误地报告服务器未运行（Bruce Momjian）

  此前当服务器在运行但 pg_ctl 无法认证时可能发生这种情况。

- 改进 pg_ctl start 的“等待”（`-w`）选项（Bruce Momjian、Tom Lane）

  等待模式现在明显更健壮。它不会因非默认的 postmaster 端口号、非默认的 Unix 域套接字位置、权限问题或过期的 postmaster 锁文件而困惑。

- 为 pg_ctl 新增 `promote` 选项，用于把备库切换为主库（Fujii Masao）

E.25.3.10. Development Tools

E.25.3.10.1. libpq

- 新增 libpq 连接选项 `client_encoding`，其行为类似于 `PGCLIENTENCODING` 环境变量（Heikki Linnakangas）

  值 `auto` 根据操作系统区域设置客户端编码。

- 新增函数 `PQlibVersion()`，返回 libpq 库版本（Magnus Hagander）

  libpq 此前已有返回服务器版本的 `PQserverVersion()`。

- 允许使用 libpq 的客户端在通过 Unix 域套接字连接时检查服务器进程的用户名，使用新的 `requirepeer` 连接选项（Peter Eisentraut）

  PostgreSQL 此前已允许服务器在通过 Unix 域套接字连接时检查客户端用户名。

- 为 libpq 新增 `PQping()` 和 `PQpingParams()`（Bruce Momjian、Tom Lane）

  这些函数允许在不尝试打开新会话的情况下检测服务器状态。

E.25.3.10.2. ECPG

- 允许 ECPG 即使在 `WHERE CURRENT OF` 子句中也接受动态游标名（Zoltan Boszormenyi）
- 让 ecpglib 以 15 位精度写 `double` 值，而不是此前的 14 位（Akira Kurosawa）

E.25.3.11. 构建选项

- 对接受它的 HP-UX C 编译器使用 `+Olibmerrno` 编译标志（Ibrar Ahmed）

  这避免了较新 HP 平台上数学库调用可能的不当行为。

E.25.3.11.1. Makefile

- 改进的并行 make 支持（Peter Eisentraut）

  这使编译更快。此外，`make -k` 现在工作得更一致。

- 要求 GNU make 3.80 或更新版本（Peter Eisentraut）

  由于并行 make 的改进，这是必要的。

- 新增 `make maintainer-check` 目标（Peter Eisentraut）

  该目标执行各种既不适合构建也不适合回归测试的源代码检查。目前包括：duplicate_oids、SGML 语法和制表符检查、NLS 语法检查。

- 在 `contrib` 中支持 `make check`（Peter Eisentraut）

  此前只有 `make installcheck` 可用，但现在支持在临时安装中测试。顶层的 `make check-world` 目标现在包括以这种方式测试 `contrib`。

E.25.3.11.2. Windows

- 在 Windows 上，允许pg_ctl把服务注册为自动启动或按需启动（Quan Zongliang）
- 新增在 Windows 上收集崩溃转储的支持（Craig Ringer、Magnus Hagander）

  现在可以由非调试的 Windows 二进制生成 minidumps，并用标准调试工具分析。

- 支持使用 MinGW64 编译器构建（Andrew Dunstan）

  这允许通过交叉编译即使在非 Windows 平台上也能构建 64 位 Windows 二进制。

E.25.3.12. 源代码

- 修订 GUC 变量赋值钩子的 API（Tom Lane）

  此前赋值钩子的职能现在拆分到检查钩子和赋值钩子之间，前者可以失败而后者不应失败。此变更会影响定义自定义 GUC 参数的附加模块。

- 在源代码中新增闩锁以支持等待事件（Heikki Linnakangas）
- 集中化数据修改权限检查逻辑（KaiGai Kohei）
- 为一致性补齐缺失的 `get_object_oid()` 函数（Robert Haas）
- 通过移除冲突的关键字，改进使用 C++ 编译器编译附加模块的能力（Tom Lane）
- 新增对 DragonFly BSD 的支持（Rumko）
- 为后端使用公开 `quote_literal_cstr()`（Robert Haas）
- 以默认编码运行回归测试（Peter Eisentraut）

  回归测试此前总是以 `SQL_ASCII` 编码运行。

- 新增 src/tools/git_changelog 以取代 cvs2cl 和 pgcvslog（Robert Haas、Tom Lane）
- 向 `src/tools` 添加 git-external-diff 脚本（Bruce Momjian）

  它用于从 git 生成上下文差异。

- 改进对使用 Clang 构建的支持（Peter Eisentraut）

E.25.3.12.1. 服务器钩子

- 新增用于检查权限的源代码钩子（Robert Haas、Stephen Frost）

- 新增供安全框架使用的对象创建后函数钩子（KaiGai Kohei）
- 新增客户端认证钩子（KaiGai Kohei）

E.25.3.13. 附加模块

- 修改 `contrib` 模块和过程语言，使其通过新的扩展机制安装（Tom Lane、Dimitri Fontaine）
- 新增 `contrib/file_fdw` 外部数据包装器（Shigeru Hanada）

  使用这个外部数据包装器的外部表可以以非常类似于 `COPY` 的方式读取平面文件。
- 为 `contrib/pg_trgm` 和 `contrib/btree_gist` 新增最近邻搜索支持（Teodor Sigaev）
- 为 `contrib/btree_gist` 新增不相等搜索支持（Jeff Davis）
- 修复 `contrib/fuzzystrmatch` 的 `levenshtein()` 函数，使其处理多字节字符（Alexander Korotkov）
- 为 `contrib/sslinfo` 新增 `ssl_cipher()` 和 `ssl_version()` 函数（Robert Haas）
- 修复 `contrib/intarray` 和 `contrib/hstore`，使其对索引的空数组给出一致结果（Tom Lane）

  此前使用索引的空数组查询可能返回与使用顺序扫描不同的结果。
- 让 `contrib/intarray` 在多维数组上正确工作（Tom Lane）
- 在 `contrib/intarray` 中，避免在实际不存在空值的场合报错抱怨存在空值（Tom Lane）
- 在 `contrib/intarray` 中，修复包含操作符对空数组的行为（Tom Lane）

  空数组现在被正确地视为包含于任何其他数组。
- 移除 `contrib/xml2` 对 `xslt_process()` 可处理的 *`parameter`*=*`value`* 对数量的任意限制（Pavel Stehule）

  此前的限制是 10。
- 在 `contrib/pageinspect` 中，修复 heap_page_item 以 32 位值返回 infomask（Alvaro Herrera）

  这避免了返回负值（那令人困惑）。底层值是 16 位无符号整数。

E.25.3.13.1. 安全

- 新增 `contrib/sepgsql`，使权限检查与 SELinux 接口（KaiGai Kohei）

  它使用新的 `SECURITY LABEL` 机制。
- 新增 contrib 模块 `auth_delay`（KaiGai Kohei）

  它使服务器在返回认证失败前暂停；其设计目的是加大暴力破解口令攻击的难度。
- 新增 `dummy_seclabel` contrib 模块（KaiGai Kohei）

它用于权限回归测试。

E.25.3.13.2. 性能

- 为 `contrib/pg_trgm` 新增 `LIKE` 和 `ILIKE` 索引搜索支持（Alexander Korotkov）
- 为 `contrib/fuzzystrmatch` 新增针对小距离优化的函数 `levenshtein_less_equal()`（Alexander Korotkov）
- 改进 `contrib/seg` 列上索引查找的性能（Alexander Korotkov）
- 改进pg_upgrade在多关系数据库下的性能（Bruce Momjian）
- 为 `contrib/pgbench` 新增报告逐语句延迟的标志（Florian Pflug）

E.25.3.13.3. fsync 测试

- 把 `src/tools/test_fsync` 移到 `contrib/pg_test_fsync`（Bruce Momjian、Tom Lane）
- 为 `contrib/pg_test_fsync` 新增 `O_DIRECT` 支持（Bruce Momjian）

  这与 `wal_sync_method` 对 `O_DIRECT` 的使用相匹配。
- 为 `contrib/pg_test_fsync` 新增新测试（Bruce Momjian）

E.25.3.14. 文档

- 广泛的ECPG文档改进（Satoshi Nagayasu）
- 广泛的校对和文档改进（Thom Brown、Josh Kupershmidt、Susanne Ebrecht）
- 为 `exit_on_error` 添加文档（Robert Haas）

  该参数使会话在任何错误时退出。
- 为 `pg_options_to_table()` 添加文档（Josh Berkus）

  该函数以可读的形式显示表存储选项。
- 记录可以使用 `(compositeval).*` 语法访问所有组合类型字段（Peter Eisentraut）
- 记录 `translate()` 会删除 `from` 中没有对应 `to` 字符的字符（Josh Kupershmidt）
- 合并 `CREATE CONSTRAINT TRIGGER` 和 `CREATE TRIGGER` 的文档（Alvaro Herrera）
- 集中化权限和升级文档（Bruce Momjian）
- 为 Solaris 10 添加内核调优文档（Josh Berkus）

  此前只记录了 Solaris 9 的内核调优。
- 在 `HISTORY` 文件中一致地处理非 ASCII 字符（Peter Eisentraut）

  虽然 `HISTORY` 文件是英文的，但我们必须处理贡献者姓名中的非 ASCII 字母。这些现在被音译，以便在不假设字符集的情况下也相当易读。

E.26. 版本 9.0.23

发行日期
2015-10-08

本次发布包含对 9.0.22 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

这预计是 PostgreSQL 9.0.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.26.1. 迁移到版本 9.0.23

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.18 的版本升级，请参见第 E.31 节。

E.26.2. 变更

- 修复`contrib/pgcrypto`以检测并报告过短的`crypt()`盐值（Josh Kupershmidt）

  某些无效的盐值参数会导致服务器崩溃或泄露少量服务器内存字节。我们尚未排除通过安排让泄露字节中出现机密信息实施攻击的可行性，但这种攻击似乎不太可能。（CVE-2015-5288）

- 修复属于外部子事务的门户（游标）失败后的子事务清理问题（Tom Lane、Michael Paquier）

  在外部子事务游标中执行的函数可能因引用在内部子事务中创建的关系而导致断言失败或崩溃。

- 修复关系插入关系缓存“init 文件”的问题（Tom Lane）

  最近几个小版本中一个补丁的疏忽导致`pg_trigger_tgrelid_tgname_index`被遗漏在 init 文件之外。后续会话检测到这一点后认定 init 文件已损坏并静默忽略它，导致会话启动时间显著退化。除修复该缺陷外，还加入了一些防护，使未来任何类似错误更明显。

- 避免向 SPI 查询结果插入大量元组时的 O(N^2) 行为（Neil Conway）

- 改进存在大量未读通知时`LISTEN`的启动时间（Matt Newell）

- 默认禁用 SSL 重新协商（Michael Paquier、Andres Freund）

  虽然使用 SSL 重新协商在理论上是个好主意，但实践中我们看到了太多缺陷，无论是底层 OpenSSL 库中还是我们对它的使用中。重新协商将在 9.5 及更高版本中完全移除。在较旧的分支中，只是把`ssl_renegotiation_limit`的默认值改为零（禁用）。

- 降低`*_freeze_max_age`参数的最小值（Andres Freund）

  这主要是为了使相关行为的测试不那么耗时，但对磁盘空间有限的安装也可能有价值。

- 将`wal_buffers`的最大值限制为 2GB 以避免服务器崩溃（Josh Berkus）

- 修复`numeric`值相乘时罕见的内部溢出（Dean Rasheed）

- 防御涉及记录类型、范围类型、`json`、`jsonb`、`tsquery`、`ltxtquery`和`query_int`的难以触发的栈溢出（Noah Misch）
- 修复日期时间输入中`DOW`和`DOY`的处理（Greg Stark）

  这些记号本不应在日期时间值中使用，但过去它们导致的是难懂的内部错误消息，而不是“无效的输入语法”。
- 在正则表达式匹配中增加更多查询取消检查（Tom Lane）
- 为正则表达式、`SIMILAR TO`和`LIKE`匹配增加递归深度保护（Tom Lane）

  合适的搜索模式加上较低的栈深度限制可能导致栈越界崩溃。
- 修复正则表达式执行中潜在的无限循环（Tom Lane）

  一个表面上能匹配零长度字符串、但实际上因反向引用而不匹配的搜索模式可能导致无限循环。
- 修复正则表达式编译中的内存不足失败（Andreas Seltenreich）
- 修复正则表达式执行期间低概率的内存泄漏（Tom Lane）
- 修复事务中止期间锁清理中罕见的内存不足失败（Tom Lane）
- 修复在使用较小`work_mem`设置的元组存储时出现“排序期间意外的内存不足情况”错误的问题（Tom Lane）
- 修复`qsort`中极低概率的栈越界（Tom Lane）
- 修复在使用较大`work_mem`设置的哈希连接中出现“无效的内存分配请求大小”失败的问题（Tomas Vondra, Tom Lane）
- 修复若干规划器缺陷（Tom Lane）

  这些错误可能导致给出错误结果的错误查询计划、启用断言的构建中的断言失败，或奇怪的规划器错误，例如“could not devise a query plan for the given query”、“could not find pathkey item to sort”、“plan should not reference subplan's variable”或“failed to assign all NestLoopParams to plan nodes”。感谢 Andreas Seltenreich 和 Piotr Stefaniak 通过模糊测试暴露了这些问题。
- 在所有受支持的分支中使用模糊路径代价平局判定规则（Tom Lane）

  此变更旨在避免当候选计划选项的估计代价实际相同时出现平台特定的行为。
- 在 postmaster 关闭期间，确保在移除`postmaster.pid`文件之前移除各套接字的锁文件并关闭监听套接字（Tom Lane）

  这可以避免外部脚本在`pg_ctl stop`一返回就尝试启动新 postmaster 时出现竞争条件失败。
- 修复 postmaster 对崩溃恢复期间启动进程崩溃的处理（Tom Lane）

  如果在崩溃恢复周期中，启动进程在尚未恢复数据库一致性时崩溃，我们会尝试启动一个新的启动进程，而这通常只会再次崩溃，导致无限循环。
- 当我们尝试向某个自动清理工作进程发送信号而它已经消失时，不再打印`WARNING`，并降低此类信号的日志详细程度（Tom Lane）

- 防止服务器时钟被大幅回拨时自动清理启动器睡眠时间过长（Álvaro Herrera）
- 确保 GIN 索引待插入列表的清理可被取消请求中断（Jeff Janes）
- 允许重用 GIN 索引中的全零页（Heikki Linnakangas）

  崩溃之后可能遗留这样的页。
- 修复一个差一错误，该错误会导致 subtrans/multixact 截断中出现本可避免的“表面回卷”警告（Thomas Munro）
- 修复PL/pgSQL错误上下文消息中`CONTINUE`和`MOVE`语句类型误报的问题（Pavel Stehule, Tom Lane）
- 修复PL/Tcl中一些未检查`malloc()`调用失败的位置（Michael Paquier, Álvaro Herrera）
- 改进libpq对内存耗尽情况的处理（Michael Paquier, Heikki Linnakangas）
- 修复ecpg中的内存泄漏和缺失的内存不足检查（Michael Paquier）
- 修复 psql 中对数字输出的区域感知格式化代码（Tom Lane）

  `\pset numericlocale on` 所调用的格式化代码在一些不常见的情况下会出错，例如带指数但没有小数点的数字。它还可能破坏 `money` 数据类型已经区域化的输出。
- 防止在没有当前连接时psql的`\c`命令崩溃（Noah Misch）
- 确保pg_dump以tar格式输出运行期间创建的临时文件不被全局可读（Michael Paquier）
- 修复pg_dump和pg_upgrade以支持`postgres`或`template1`数据库位于非默认表空间的情况（Marti Raudsepp, Bruce Momjian）
- 修复 pg_dump 从过于陈旧而没有特定权限类型的服务器转储时对对象权限的处理（Tom Lane）

  从 7.3 之前的服务器转储函数或过程语言时，pg_dump 会生成撤销所有者可授予权限、转而把所有权限授予 `PUBLIC` 的 `GRANT/REVOKE` 命令。由于涉及的权限只是 `USAGE` 和 `EXECUTE`，这不构成安全问题，但这确实是对旧系统行为令人惊讶的表示。已修复为在这些情况下保持默认权限状态不变。
- 修复pg_dump以转储 shell 类型（Tom Lane）

  shell 类型（即尚未完全定义的类型）用处不大，但pg_dump仍应转储它们。
- 修复 PPC 硬件的自旋锁汇编代码使其与AIX本机汇编器兼容（Tom Lane）

  如果gcc被配置为使用本机汇编器（这正变得越来越普遍），则用gcc构建无法工作。
- 在AIX上，测试`-qlonglong`编译器选项，而不是直接假定它可以安全使用（Noah Misch）
- 在AIX上，使用`-Wl,-brtllib`链接选项以允许符号在运行时解析（Noah Misch）

  Perl 5.8.0 及更高版本依赖这一能力。
- 由于编译器缺陷，在使用 32 位xlc编译时避免使用内联函数（Noah Misch）
- 必要时使用`librt`提供`sched_yield()`，某些 Solaris 版本即是如此（Oskari Saarenmaa）

- 修复 Windows`install.bat`脚本以处理包含空格的目标目录名（Heikki Linnakangas）
- 使PostgreSQL版本号的数字形式（例如`90405`）以名为`VERSION_NUM`的变量方便地提供给扩展 Makefile 使用（Michael Paquier）
- 将时区数据文件更新到tzdata发行版 2015g，以反映开曼群岛、斐济、摩尔多瓦、摩洛哥、诺福克岛、朝鲜、土耳其和乌拉圭的夏令时法规变更。为加拿大北落基山地区新增了时区名`America/Fort_Nelson`。

E.27. 版本 9.0.22

发行日期
2015-06-12

本次发布包含对 9.0.21 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

PostgreSQL 社区将于 2015 年 9 月停止对 9.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.27.1. 迁移到版本 9.0.22

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.18 的版本升级，请参见第 E.31 节。

E.27.2. 变更

- 修复罕见的未能使关系缓存 init 文件失效的问题（Tom Lane）

  在并发活动时序恰好不对的情况下，对系统目录执行`VACUUM FULL`可能无法更新用于避免新会话缓存加载工作的“init 文件”。这将导致后续会话完全无法访问该目录。这是一个非常古老的缺陷，但它极难触发，直到最近才出现可复现的案例。

- 避免传入会话与`CREATE/DROP DATABASE`之间的死锁（Tom Lane）

  在作为`DROP DATABASE`命令目标或`CREATE DATABASE`命令模板的数据库中启动的新会话，可能导致该命令等待五秒后失败，即使新会话本会在此之前退出。

E.28. 版本 9.0.21

发行日期
2015-06-04

本次发布包含对 9.0.20 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

PostgreSQL 社区将于 2015 年 9 月停止对 9.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.28.1. 迁移到版本 9.0.21

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.18 的版本升级，请参见第 E.31 节。

E.28.2. 变更

- 避免崩溃重启期间对数据目录执行`fsync`时失败（Abhijit Menon-Sen、Tom Lane）

  在前几个小版本中，我们加入了一个补丁，在崩溃后对数据目录中的所有内容执行`fsync`。遗憾的是，它对任何错误情况的响应都是失败，从而阻止服务器启动，即使问题相当无害。例如，数据目录中一个不可写的文件在某些平台上会阻止重启；但把 SSL 证书文件设为服务器不可写是很常见的做法。我们修改了这一行为：权限失败被完全忽略，其他类型的失败会被记录但不会阻止继续。

- 移除configure中对在OpenBSD上链接线程版libpython的禁止性检查（Tom Lane）

  这一限制原本要防止的失败在当前的OpenBSD版本上似乎已不再是问题。

- 允许 libpq 使用高于 v1 的 TLS 协议版本（Noah Misch）

  长期以来，libpq 的代码只允许使用 TLS v1 这一种 SSL 协议。随着较新的 TLS 版本日益普及，现在允许它与服务器协商双方共同支持的最高 TLS 版本。（PostgreSQL 服务器早已支持这种协商，因此服务器端无需更改。）这是对 9.4.0 中已发布变更的回移植。

E.29. 版本 9.0.20

发行日期
2015-05-22

本次发布包含对 9.0.19 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

PostgreSQL 社区将于 2015 年 9 月停止对 9.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.29.1. 迁移到版本 9.0.20

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.18 的版本升级，请参见第 E.31 节。

E.29.2. 变更

- 避免客户端恰在认证超时到期前断开时可能出现的崩溃（Benkocs Norbert Attila）

  如果超时中断在会话关闭序列进行到一半时触发，SSL 相关状态会被释放两次，通常导致崩溃并对其他会话造成服务拒绝。实验表明，未经认证的远程攻击者能够相当稳定地触发此缺陷，因此按安全问题处理。（CVE-2015-3165）

- 改进对系统调用失败的检测（Noah Misch）

我们替换实现的`snprintf()`未检查底层系统库调用报告的错误；
可能被遗漏的主要情况是内存耗尽。在最坏情况下，
由于我们的代码假定缓冲区已被覆写而实际未覆写，这可能导致信息暴露。此外，
还有少数地方对安全相关的其他系统库函数调用未检查失败。

仍有可能某些对`*printf()`族函数的调用在恰逢内存耗尽错误时存在信息泄露风险。
我们判断风险不大，但将继续在此领域进行分析。（CVE-2015-3166）

- 在`contrib/pgcrypto`中，统一将解密失败报告为“Wrong key or corrupt data”（Noah Misch）

  此前，某些使用错误密钥解密的情况可能报告其他错误消息文本。已经证明，
  这种错误报告的差异可能帮助攻击者从其他系统恢复密钥。虽然尚不清楚`pgcrypto`
  的具体行为是否同样可被利用，但使用统一消息来避免风险似乎更好。（CVE-2015-3167）

- 修复 HOT 更新后延迟排他约束检查不正确的问题（Tom Lane）

  如果同一事务中稍后对一条可能违反延迟排他约束的新行进行了 HOT 更新（即没有索引列发生变化，且该行可以存回同一表页），那么最终执行检查时将报告排他约束被违反，
  即使新行最初与之冲突的行已被删除。

- 防止反连接（NOT EXISTS 连接）与其他外连接之间的不当重排序（Tom Lane）

  已知规划器中的这一疏忽会导致“could not find RelOptInfo for given relids”错误，但有时错误的查询计划似乎有可能通过该一致性检查并导致静默错误的查询输出。

- 修复外连接计划节点中子表达式匹配不正确的问题（Tom Lane）

  此前，如果文本相同的非严格子表达式在外连接的上、下方都被使用，
  规划器可能试图重用在连接下方计算的值，但这是不正确的，因为对于未匹配的外部行，
  执行器会将该值强制置为 NULL。

- 修复 GEQO 规划器以应对其连接顺序启发式的失败（Tom Lane）

  已知这一疏忽会在涉及 `LATERAL` 的查询中导致“failed to join all relations together”错误，
  其他情况下也可能发生。

- 修复`max_prepared_transactions`过小时启动阶段可能出现的死锁（Heikki Linnakangas）

- 时间线切换后不再归档无用的预分配 WAL 文件（Heikki Linnakangas）

- 避免“cannot GetMultiXactIdMembers() during recovery”错误（Álvaro Herrera）

- 崩溃之后递归地对数据目录执行 `fsync()`（Abhijit Menon-Sen、Robert Haas）

  这确保了若随后不久再发生崩溃时数据仍保持一致。（要有问题的话，
  第二次崩溃必须是系统级崩溃，而不只是数据库崩溃。）

- 修复自动清理启动器在收到 SIGTERM 之后发生错误时可能无法关闭的问题（Álvaro Herrera）

- 处理 `LockBufferForCleanup()` 中的意外信号（Andres Freund）

  这一疏忽可能导致有关“multiple backends attempting to wait for pincount 1”的虚假错误。

- 就用户而言是只读的事务，在提交期间不再等待 WAL 刷写或同步复制（Andres Freund）

此前，由于 HOT 页清理而写入了 WAL 的事务在提交时可能出现延迟，导致不良影响，例如所有同步副本都停机时会话在启动时卡住。还观察到使用同步复制时会话卡在追赶中断处理中；此修复也将解决该问题。

- 修复在临时表上操作哈希索引时崩溃的问题（Heikki Linnakangas）
- 修复哈希索引桶分裂期间（若其他进程正在并发修改该索引）可能失败的问题（Tom Lane）
- 在分析索引表达式期间检查中断（Jeff Janes）

  `ANALYZE` 会多次执行索引表达式；如果此类表达式中存在慢速函数，就希望能在此循环结束之前取消 `ANALYZE`。

- 在外部服务器用户映射的对象描述字符串中加入目标服务器的名称（Álvaro Herrera）
- 建议在使用 Kerberos/GSSAPI/SSPI 认证时将 `include_realm` 设置为 1（Stephen Frost）

  不这样设置的话，来自不同领域的同名用户无法区分。目前这只是文档变更，但在 PostgreSQL 9.5 中它将成为默认设置。

- 移除将 IPv4 `pg_hba.conf` 条目匹配到 IPv4-in-IPv6 地址的代码（Tom Lane）

  这段应急代码于 2003 年添加，起因是当时的某些 Linux 内核会把 IPv4 连接报告为 IPv4-in-IPv6 地址。然而该逻辑在 9.0 中被意外破坏。此后没有任何现场投诉，说明它已不再需要。现在反而有报告称被破坏的代码在某些系统上导致崩溃，因此我们选择直接移除而不是修复它。（如果选择修复，将会使 IPv4 `pg_hba.conf` 条目的实际含义发生微妙且可能涉及安全性的变化，在次版本中这样做似乎并不妥当。）

- 在 Windows 上关闭服务期间，定期向服务控制管理器发送状态更新，以防止它过早终止服务；并确保pg_ctl会等待关闭完成（Krystian Bigaj）
- 降低使用libpq非阻塞模式时网络死锁的风险（Heikki Linnakangas）

  在发送大量数据时，定期排空输入缓冲区很重要，
  以防服务器已发送足够的响应数据导致其在输出上阻塞。（典型场景是服务器在`COPY FROM STDIN`期间发送一连串 NOTICE 消息。）这在普通阻塞模式下工作正常，但在非阻塞模式下则不然。我们修改了libpq，使其在可能时顺便排空输入，
  但要完全防御此问题需要应用程序配合：应用程序应同时监视套接字的读就绪和写就绪条件，
  并确保在读就绪时调用`PQconsumeInput()`。

- 修复ecpg中的数组处理（Michael Meskes）
- 修复psql以合理处理作为`\connect`第一个参数的 URI 和 conninfo 串（David Fetter、Andrew Dunstan, Álvaro Herrera）

  这一语法已被接受（但未写入文档）很长时间，
  但此前某些参数可能取自旧连接而非给定字符串，这被认为是不合适的。

- 抑制psql在某些平台上关于退出时未能写入`~/.psql_history`的错误抱怨（Tom Lane）

  这一不当行为源自针对非常老（2006 年之前）的libedit版本中一个缺陷的变通方案。我们通过移除该变通方案修复了问题，但这会使仍在使用此类libedit版本的用户出现类似失败。
  建议：升级该库，或改用libreadline。

- 修复pg_dump判定哪些转换是不应转储的系统提供转换的规则（Tom Lane）
- 修复对仅由`VALUES(...)`构成但带有列别名的视图的转储（Tom Lane）

- 在pg_upgrade中，强制新集簇使用时间线 1（Bruce Momjian）

  此变更可防止因错误抱怨缺少 WAL 历史文件而导致的升级失败。

- 在pg_upgrade中，继续之前先检查是否存在不当不可连接的数据库（Bruce Momjian）
- 在pg_upgrade中，在生成的`delete_old_cluster`脚本中对目录路径正确加引号（Bruce Momjian）
- 在pg_upgrade中正确保留数据库级冻结信息（Bruce Momjian）

  这一疏忽可能导致`postgres`和`template1`数据库中的表出现缺少 clog 文件的错误。

- 在 Windows 上以受限权限运行pg_upgrade和pg_resetxlog，使它们在由管理员运行时不会失败（Muhammad Asif Naeem）
- 修复`contrib/intarray`中缓慢的排序算法（Tom Lane）
- 修复 Sparc V8 机器上的编译失败（Rob Rowan）
- 将时区数据文件更新到tzdata发行版 2015d，以反映埃及、蒙古和巴勒斯坦的夏令时法规变更，以及加拿大和智利的历史变更。同时采用 America/Adak 时区修订后的时区缩写（HST/HDT 而非 HAST/HADT）。

E.30. 版本 9.0.19

发行日期
2015-02-05

本次发布包含对 9.0.18 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.30.1. 迁移到版本 9.0.19

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.18 的版本升级，请参见第 E.31 节。

E.30.2. 变更

- 修复`to_char()`中的缓冲区越界（Bruce Momjian）

  当`to_char()`处理要求数字位数很多的数字格式模板时，
  PostgreSQL会读取越过缓冲区末尾。当处理恶意构造的时间戳格式模板时，
  PostgreSQL会写入越过缓冲区末尾。两种情况都可能导致服务器崩溃。
  我们尚未排除导致权限提升攻击的可能性，尽管这种攻击似乎不太可能。（CVE-2015-0241）

- 修复替换实现的`*printf()`函数中的缓冲区越界（Tom Lane）

  PostgreSQL包含`printf`及相关函数的替换实现。当代码格式化浮点数（转换说明符`e`、`E`、`f`、`F`、`g`或`G`）且请求的精度超过约 500 时，会发生栈缓冲区越界。这会导致服务器崩溃，我们尚未排除导致权限提升攻击的可能性。数据库用户可以通过`to_char()`SQL 函数触发这种缓冲区越界。虽然这是唯一受影响的核心PostgreSQL功能，但使用 printf 族函数的扩展模块也可能存在风险。

此问题主要影响 Windows 上的PostgreSQL。PostgreSQL在系统实现足够完善时会使用系统实现，而其他现代平台上正是如此。（CVE-2015-0242）

- 修复`contrib/pgcrypto`中的缓冲区越界（Marko Tiikkaja、Noah Misch）

  `pgcrypto`
  模块中内存大小跟踪的错误导致了栈缓冲区越界和对未初始化内存内容的不当依赖。
  缓冲区越界情况可能导致服务器崩溃，我们尚未排除导致权限提升攻击的可能性。（CVE-2015-0243）

- 修复出错后前端/后端协议可能失去同步的问题（Heikki Linnakangas）

  如果服务器在从客户端读取协议消息中途发生任何错误，它可能失去同步，
  错误地把消息数据的一部分解释为新的协议消息。
  能够在命令参数中提交恶意构造二进制数据的攻击者可能借此注入自己的 SQL 命令。语句超时和查询取消是触发此场景最可能的错误来源。
  尤其容易受影响的是既使用超时又将任意用户构造数据作为二进制查询参数提交的应用程序。
  禁用语句超时会降低但不能消除被利用的风险。感谢 Emil Lenngren 报告此问题。（CVE-2015-0244）

- 修复通过约束违反错误消息的信息泄露（Stephen Frost）

  某些服务器错误消息会显示违反约束（如唯一约束）的列的值。
  如果用户对该表的所有列没有`SELECT`权限，这可能意味着暴露用户不应看到的值。
  我们调整了代码，使值仅在来自 SQL 命令本身或用户可以查询到时才显示。（CVE-2014-8161）

- 锁定 Windows 上回归测试的临时安装（Noah Misch）

  使用 SSPI 认证，只允许启动测试套件的操作系统用户连接。这在 Windows 上关闭了此前已在其他平台上关闭的同一漏洞，即其他用户可能连接到测试 postmaster。（CVE-2014-0067）

- 避免在使用 `ALTER DATABASE SET TABLESPACE` 把数据库移到新表空间、随后不久又移回原表空间时可能出现的数据损坏（Tom Lane）

- 避免当事务内的 `ANALYZE` 被回滚时损坏表（Andres Freund、Tom Lane、Michael Paquier）

  如果失败的事务此前删除了表上的最后一个索引、规则或触发器，
  该表会处于损坏状态：相关的 `pg_class` 标志本应设置却未被设置。

- 修复 EvalPlanQual 处理中使用已释放内存的问题（Tom Lane）

  在`READ COMMITTED`模式下，锁定或更新最近更新过的行的查询可能因此缺陷崩溃。

- 修复在子表上使用部分索引时 `SELECT FOR UPDATE` 的规划问题（Kyotaro Horiguchi）

  在 `READ COMMITTED` 模式下，`SELECT FOR UPDATE` 在重新检查最近更新的行以确认它是否仍满足查询的 `WHERE` 条件时，还必须重新检查部分索引的 `WHERE` 条件。如果索引属于继承子表，就会遗漏这一要求，因此有可能错误地返回已不再满足查询条件的行。

- 修复 `SELECT FOR UPDATE` 在极端情况下可能把同一行返回两次、并且可能漏掉其他行的 corner case（Tom Lane）

  在 `READ COMMITTED` 模式下，扫描继承树的 `SELECT FOR UPDATE` 可能错误地返回来自前一个子表的行，而不是本应从后一个子表返回的行。

- 拒绝 `FOREIGN KEY` 声明的被引用列列表中出现重复的列名（David Rowley）

  这一限制符合 SQL 标准。此前我们并未显式拒绝这种情况，但之后代码会以看起来很怪异的错误失败。

- 修复 `numeric` 值求大整数次幂时的缺陷（Tom Lane）

  之前的代码可能得到错误答案，或者在意识到结果必然溢出之前消耗过多的时间和内存。

- 在 `numeric_recv()` 中，按照值的 `dscale` 字段截去会被隐藏的小数位（Tom Lane）

  `numeric` 值的显示标度（`dscale`）绝不应小于非零小数位的位数；但显然至少存在一个有缺陷的客户端应用，它发送的二进制 `numeric` 值违反了这一点。这会导致奇怪的行为，因为多余的数字位虽然不会被打印出来，却会被算术运算考虑在内。
  风险最小的修复方法似乎是在接收时把这类“隐藏的”数字位截去，
  使值确实与它打印出来的一致。

- 拒绝超出范围的数字时区规范（Tom Lane）

  超过 +/- 168 小时（一周）的简单数字时区规范此前会被接受，但随后可能在某些操作中导致空指针解引用崩溃。如此大的 UTC 偏移没有实际使用场景，因此予以拒绝。

- 修复 `tsquery @> tsquery` 操作符中的缺陷（Heikki Linnakangas）

  两个不同的词如果 CRC 相同，就会被认为匹配。此外，如果第二个操作数的词条比第一个多，就会被假定不包含于第一个；这是错误的，因为它可能包含重复词条。

- 增强 ispell 词典对有问题的词缀文件的防御（Tom Lane）

- 允许词典（thesaurus）包含超过 64K 个短语（David Boutin）

  之前的代码在词典过大时会崩溃，因此这被视为一个可回移植的缺陷修复，而不是新特性。

- 修复`xpath()`中的命名空间处理（Ali Akbar）

  此前，如果命名空间声明附在输入`xml`值的祖先元素上而非被返回的特定元素上，`xpath()`调用产生的`xml`值就不会包含命名空间声明。我们将祖先声明传播下来，
  使结果在孤立考虑时也是正确的。

- 修复嵌套 append 关系（例如 `UNION ALL` 子查询中的继承表）的规划器问题（Tom Lane）

- 当 GiST 索引元组放不进一个页面时干净地报错，而不是陷入无限递归（Andrew Gierth）

- 把设有逐表 `cost_limit` 和/或 `cost_delay` 设置的表排除在自动清理的全局代价平衡规则之外（Álvaro Herrera）

  先前的行为实质上忽略了这些逐表设置，这并非本意。现在，
  拥有这类设置的表将按照这些设置进行清理，独立于其他自动清理工作进程的状况。
  这可能导致比以前更重的总 I/O 负载，因此应当重新审视这类设置是否合理。

- 避免在自动清理名义上被关闭时仍然进行全面的自动清理（Tom Lane）

  即使自动清理名义上被关闭，我们仍会启动自动清理工作进程，去清理有 XID 回卷风险的表。
  然而，这样的工作进程随后会继续清理目标数据库中所有满足常规自动清理阈值的表。
  这往好里说是相当出乎意料，往坏里说会延误对回卷威胁的响应。
  现已修复：如果自动清理被关闭，工作进程只执行防回卷清理，不做任何其他工作。

- 修复热备查询与重放整页镜像之间的竞态条件（Heikki Linnakangas）

这一失误可能导致热备中正在执行的查询出现瞬时错误。

- 修复若干恢复逻辑不当忽略 `COMMIT/ABORT PREPARED` WAL 记录的情况（Heikki Linnakangas）

  最值得注意的疏漏是 `recovery_target_xid` 无法用于在两阶段提交处停止。

- 避免为时间线历史文件创建不必要的 `.ready` 标记文件（Fujii Masao）

- 修复当使用空的预备语句且 `log_statement` 设置为 `mod` 或 `ddl` 时可能出现的空指针解引用（Fujii Masao）

- 将“pgstat wait timeout”警告消息改为 LOG 级别，并重新措辞使其更易理解（Tom Lane）

  这条消息最初被认为基本属于不可能发生的情况，但它在我们较慢的 buildfarm 成员上出现得足够频繁，成了干扰。现已将其降为 LOG 级别，并在措辞上多下了一点功夫：现在它的内容为“using stale statistics instead of current ones because stats collector is not responding”。

- 修复 SPARC 自旋锁实现，确保当 CPU 以非 TSO 一致性模式运行时（如某些非 Solaris 内核所做的那样）其正确性（Andres Freund）

- 警告 OS X 的 `setlocale()` 在 postmaster 内启动多余线程的情况（Noah Misch）

- 修复 `PQconnectdbParams()` 中重复 `dbname` 参数的处理（Alex Shulgin）

  如果第一个 `dbname` 出现处包含要展开的连接串或 URI，就会出现意外行为。

- 确保 libpq 在遇到意外的套接字 EOF 时报告合适的错误消息（Marko Tiikkaja、Tom Lane）

  取决于内核行为，当服务器意外关闭套接字时，libpq 可能返回空的错误字符串而不是有用的信息。

- 在 `PQreset()` 期间清除任何旧的错误消息（Heikki Linnakangas）

  如果反复调用 `PQreset()` 而连接始终无法重建，失败连接尝试的错误消息会在 `PGconn` 的错误字符串中不断累积。

- 正确处理 libpq 解析连接选项期间的内存不足情况（Alex Shulgin、Heikki Linnakangas）

- 修复 ecpg 版 `ParseDateTime()` 中的数组越界（Michael Paquier）

- 在 initdb 中，当指定的密码文件为空时给出更清晰的错误消息（Mats Erik Andersson）

- 修复 psql 的 `\s` 命令，使其与 libedit 良好配合，并新增分页器支持（Stepan Rutz、Tom Lane）

  使用 libedit 而非 readline 时，`\s` 会以一种相当难读的编码格式打印命令历史，而且在较新的 libedit 版本上可能完全失败。现已改为由我们自己打印历史而不是让库来做。一个令人愉快的副作用是，在合适的时候会使用分页器。

  此补丁还修复了一个在使用 libedit 保存命令历史时换行编码不一致的缺陷。旧版 psql 写出的多行历史条目在此补丁后可以干净地读取，但反过来未必可行，具体取决于所涉及的 libedit 版本。

- 改进 psql 特殊变量解析的一致性（Tom Lane）

允许 ECHO_HIDDEN 和 ON_ERROR_ROLLBACK 使用 on 和 off 的变体拼写（如 1/0）。对 COMP_KEYWORD_CASE、ECHO、ECHO_HIDDEN、HISTCONTROL、ON_ERROR_ROLLBACK 和 VERBOSITY 的无法识别的值发出警告。对所有这些变量的所有值进行不区分大小写的识别；此前大小写敏感与不敏感的行为混杂在一起。

- 修复 psql 扩展模式显示在使用 border = 3 且 linestyle = ascii 或 unicode 时的一致性问题（Stephen Frost）
- 修复并行恢复纯模式转储时可能出现的死锁（Robert Haas、Tom Lane）
- 修复 pg_dump --binary-upgrade 在零列组合类型上的核心转储（Rushabh Lathia）
- 修复 contrib/pageinspect 的 get_raw_page() 中的块号检查（Tom Lane）

  错误的检查逻辑可能阻止访问非主关系分支中的某些页面。

- 修复 contrib/pgcrypto 的 pgp_sym_decrypt()，使其在消息长度比 2 的幂小 6 时不再失败（Marko Tiikkaja）
- 安全地处理 contrib/tablefunc 的 connectby() 中的意外查询结果，尤其是 NULL（Michael Paquier）

  此前 connectby() 遇到 NULL 键值时会崩溃。现在它会打印该行但不再进一步递归。

- 避免 contrib/xml2 的 xslt_process() 中可能出现的崩溃（Mark Simonetti）

  libxslt 似乎对资源释放顺序存在未记载的依赖；重新排列我们的调用以避免崩溃。

- 大量清理 Coverity 静态代码分析器的警告（Andres Freund、Tatsuo Ishii、Marko Kreen、Tom Lane、Michael Paquier）

  这些变更大多只是表面性的，但在某些情况下修复了极端情况的缺陷，
  例如内存不足失败后崩溃而不是正确报告错误。据信它们都不构成安全问题。

- 在构建期间检测不兼容的 OpenLDAP 版本（Noah Misch）

  对于 2.4.24 到 2.4.31（含）之间的 OpenLDAP 版本，PostgreSQL 后端进程可能在退出时崩溃。现在会根据编译时的 OpenLDAP 版本号在 configure 期间发出警告，并在 contrib/dblink 回归测试中测试该崩溃场景。

- 在非 MSVC 的 Windows 构建中，确保安装的 libpq.dll 带有执行权限（Noah Misch）
- 使 pg_regress 成功退出时删除其创建的任何临时安装（Tom Lane）

  这使 make check-world 期间的磁盘空间用量大幅减少，因为该序列会创建大量临时安装。

- 支持随时间改变 UTC 偏移的时区缩写（Tom Lane）

  此前，PostgreSQL 假定时区缩写（如 EST）关联的 UTC 偏移在任何特定区域设置的用法中永不改变。然而这一假设在现实世界中并不成立，因此引入了让区域缩写表示有时会变化的 UTC 偏移的能力。更新了时区缩写定义文件，在自 1970 年以来缩写的 UTC 偏移发生过变化的时区区域设置中使用此功能（依据 IANA 时区数据库）。在此类时区中，PostgreSQL 现在会根据给定日期为缩写关联正确的 UTC 偏移。

- 更新时区缩写列表（Tom Lane）

把 CST（中国标准时间）加入列表。移除 ADT 作为“阿拉伯夏令时”的引用——这个缩写自 2007 年起已不再使用；因此，声称它与“大西洋夏令时”冲突似乎没有特别的帮助。修正 CKT（库克群岛）、FJT 和 FJST（斐济）完全错误的 GMT 偏移；此前我们甚至把它们放到了日期变更线的错误一侧。

- 把时区数据文件更新为 tzdata 版本 2015a。

  IANA 时区数据库为所有澳大利亚时区采用了 `AxST`/`AxDT` 形式的缩写，以反映他们认为的当地当前多数用法。这些名称与其他地方的用法不冲突（唯一例外是表示 Acre 夏令时的 ACST，该用法自 1994 年起已废弃）。因此，我们把这些名称采纳进“默认”时区缩写集。“澳大利亚”缩写集现在只包含 CST、EAST、EST、SAST、SAT 和 WST，这些被认为主要是历史用法。注意，在“默认”缩写集中，SAST 也已改为南非标准时间。

  此外，新增时区缩写 SRET（Asia/Srednekolymsk）和 XJT（Asia/Urumqi），并对西萨摩亚使用 WSST/WSDT。同时，智利、墨西哥、特克斯和凯科斯群岛（America/Grand_Turk）以及斐济的夏令时法规有变更。还为巴布亚新几内亚部分地区新增了时区 Pacific/Bougainville。另外还有大量针对历史（1970 年以前）时区数据的修正。

E.31. 版本 9.0.18

发行日期
2014-07-24

本次发布包含对 9.0.17 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.31.1. 迁移到版本 9.0.18

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了某些 GiST 索引的索引损坏问题。请参见下面第一条变更日志，以确认你的安装是否受影响以及应采取的步骤。

不过，如果你是从早于 9.0.15 的版本升级，请参见第 E.34 节。

E.31.2. 变更

- 正确初始化 `contrib/btree_gist` 在 `bit` 列上索引的填充字节（Heikki Linnakangas）

  这一错误可能导致本应比较相等的值不被视为相等，从而产生错误的查询结果。在 `bit` 或 `bit varying` 列上建有 GiST 索引的用户，在安装此更新后应对这些索引执行 `REINDEX`。

- 在删除 GIN 列表页时防止残缺页（torn page）（Heikki Linnakangas）

  此修复可防止在页面更新写入磁盘期间发生系统崩溃时可能出现的索引损坏。

- 在从 WAL 重放更新时不清除 GiST 索引页的右链接（Heikki Linnakangas）

  这一错误可能导致热备中执行的 GiST 索引扫描出现瞬时错误的结果。

- 修复嵌套调用 `ReceiveSharedInvalidMessages` 期间缓存失效可能不正确的问题（Andres Freund）

- 当子查询的目标列表中存在集合返回函数时，不再假定其输出是唯一的（David Rowley）

  这一疏忽可能导致诸如 `WHERE x IN (SELECT y, generate_series(1,10) FROM t GROUP BY y)` 之类的构造被错误优化。

- 修复未能对结构化类型的组合元素中的字段执行解 TOAST 的错误（Tom Lane）

  这纠正了 TOAST 指针可能未经解引用就被复制到其他表的情况。如果原始数据随后被删除，当这个已悬空的指针被使用时，会导致诸如 “missing chunk number 0 for toast value ...” 之类的错误。

- 修复对 Append 计划节点输出的整行引用导致 “record type has not been registered” 失败的问题（Tom Lane）

- 修复在回卷游标时调用用户定义函数可能崩溃的问题（Tom Lane）

- 修复为 `FROM` 中的函数求值参数时查询级内存泄漏的问题（Tom Lane）

- 修复正则表达式处理中的会话级内存泄漏（Tom Lane、Arthur O'Dwyer、Greg Stark）

- 修复 `hungarian.stop` 中的数据编码错误（Tom Lane）

- 修复对在当前事务中插入、随后被一个现已回滚的子事务删除的行的活跃性检查（Andres Freund）

  如果同一事务中稍后执行了 `CREATE INDEX` 或 `CLUSTER`，这可能导致问题（至少是虚假警告，最坏是无限循环）。

- 在 `PREPARE TRANSACTION` 期间清除 `pg_stat_activity.xact_start`（Andres Freund）

  在 `PREPARE` 之后，发起会话已不在事务中，因此不应继续显示事务开始时间。

- 修复 `REASSIGN OWNED` 对文本搜索对象失败的问题（Álvaro Herrera）

- 在 postmaster 启动期间阻塞信号（Tom Lane）

  这确保了 postmaster 在例如启动过程中收到 SIGINT 时能够正确清理自身。

- 保护 `make check` 期间启动的临时 postmaster 的 Unix 域套接字（Noah Misch）

  正如我们此前在 CVE-2014-0067 中指出的，任何能够访问套接字文件的本地用户都可以以服务器的引导超级用户身份连接，进而以运行测试的操作系统用户身份执行任意代码。
  此变更通过把服务器的套接字放到 `/tmp` 下一个临时的、模式为 0700 的子目录中来防御这一风险。但在不支持 Unix 套接字的平台（尤其是 Windows）上风险仍然存在，因为那时临时 postmaster 必须接受本地 TCP 连接。

  此变更的一个有用副作用是简化了覆盖 `DEFAULT_PGSOCKET_DIR` 的构建中的 `make check` 测试。常用的非默认值（如 `/var/run/postgresql`）往往不允许构建用户写入，此前需要一些变通办法，现在不再需要了。

- 修复表空间创建 WAL 重放在 Windows 上无法工作的问题（MauMau）

- 修复 Windows 上套接字创建失败的检测（Bruce Momjian）

- 在 Windows 上，允许新会话从配置文件获取 PGC_BACKEND 参数（例如log_connections）的值（Amit Kapila）

此前，如果启动后在文件中更改了这类参数，更改不会生效。

- 在 Windows 上正确地为可执行文件路径名加引号（Nikhil Deshpande）

  如果安装路径同时包含空格和 `@` 符号，这一疏忽可能导致 initdb 和 pg_upgrade 在 Windows 上失败。

- 修复 OS X 上 libpython 的链接（Tom Lane）

  我们先前使用的方法在 Xcode 5.0 及之后版本提供的 Python 库上可能失败。

- 当服务器持续以快于客户端吸收速度的速率发送数据时，避免 libpq 中的缓冲区膨胀（Shin-ichi Morita、Tom Lane）

  libpq 可能被迫不断增大其输入缓冲区，直到内存耗尽（这会被误导性地报告为“lost synchronization with server”）。在正常情况下，数据持续以快于 `recv()` 循环吸收速度的速率传输是相当牵强的，但在客户端因调度器约束被人为减速时曾观察到这种情况。

- 确保 libpq 中的 LDAP 查找尝试按预期超时（Laurenz Albe）

- 修复 ecpg 在 `char *` 数组作为返回多行的 FETCH 语句目标时的行为，以及其他一些数组处理修复（Ashutosh Bapat）

- 修复 pg_restore 对旧式大对象注释的处理（Tom Lane）

  如果归档文件由 9.0 之前版本的 pg_dump 生成且包含较多大对象注释，直接恢复到数据库通常会失败。

- 在 `contrib/pgcrypto` 函数中，确保返回前清除栈变量中的敏感信息（Marko Kreen）

- 在 `contrib/uuid-ossp` 中，跨调用缓存 OSSP UUID 库的状态（Tom Lane）

  这提高了 UUID 生成的效率，并在有 `/dev/urandom` 的平台上减少了从中抽取的熵量。

- 把时区数据文件更新为 tzdata 版本 2014e for DST law changes in Crimea, Egypt, and Morocco。

E.32. 版本 9.0.17

发行日期
2014-03-20

本次发布包含对 9.0.16 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.32.1. 迁移到版本 9.0.17

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.15 的版本升级，请参见第 E.34 节。

E.32.2. 变更

- 无条件恢复 GIN 元页以避免残缺页风险（Heikki Linnakangas）

虽然这一疏忽理论上可能导致索引损坏，但实际上不太可能造成任何问题，因为 GIN 元页的活动部分小于标准的 512 字节磁盘扇区。

- 避免在接收 `NOTIFY` 消息期间检查事务提交状态时的竞态条件（Marko Tiikkaja）

  这防止了这样一种场景：足够快的客户端可能在通知者所做的数据库更新对接收者可见之前就响应通知。

- 允许正则表达式操作符被查询取消请求提前终止（Tom Lane）

  这防止了病态正则表达式长时间不可中断地锁死服务器进程的场景。

- 移除试图允许 `OVERLAPS` 接受单元素行参数的错误代码（Joshua Yanovski）

  这段代码从未正确工作过，而且该情况既未被 SQL 标准规定也未被记录，因此移除它似乎比修复它更好。

- 避免在反解析规则或视图时获取超出 `AccessShareLock` 的锁（Dean Rasheed）

  这一疏忽导致 pg_dump 对规则中作为 `INSERT`/`UPDATE`/`DELETE` 命令目标的表意外获取了 `RowExclusiveLock` 锁。虽然通常无害，但它可能干扰试图在这些表上获取例如 `ShareLock` 的并发事务。

- 改进规划期间索引端点探测的性能（Tom Lane）

  此变更修复了当索引末尾存在大量尚未提交的行时的显著性能问题，
  这对于基于顺序赋值（如时间戳或序列生成的标识符）的索引是常见情况。

- 修复崩溃后是否可以立即允许热备连接的测试（Heikki Linnakangas）

- 在报告非 `ERROR` 消息期间防止中断（Tom Lane）

  这可以防止因递归进入 `syslog()` 而导致的罕见服务器进程冻结，或许还能防止其他相关问题。

- 防止在较新的 Windows 版本上间歇性出现 “could not reserve shared memory region” 失败（MauMau）

- 把时区数据文件更新为 tzdata 版本 2014a for DST law changes in Fiji and Turkey, plus historical changes in Israel and Ukraine。

E.33. 版本 9.0.16

发行日期
2014-02-20

本次发布包含对 9.0.15 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.33.1. 迁移到版本 9.0.16

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.15 的版本升级，请参见第 E.34 节。

E.33.2. 变更

- 加固 `GRANT ... WITH ADMIN OPTION` 的限制（Noah Misch）

  不带 `ADMIN OPTION` 授予角色本应阻止被授予者向该角色添加或移除成员，但这一限制很容易通过先执行 `SET ROLE` 绕过。其安全影响主要在于角色成员可以撤销其他人的访问权限，这违背了授权者的意愿。未经批准添加角色成员的影响相对较小，
  因为不配合的角色成员本来就可以通过创建视图或 `SECURITY DEFINER` 函数把自己的大部分权限提供给他人。（CVE-2014-0060）

- 防止通过手工调用 PL 验证器函数进行权限提升（Andres Freund）

  PL 验证器函数的主要角色是在 `CREATE FUNCTION` 期间被隐式调用，但它们也是普通 SQL 函数，用户可以显式调用。对实际以其他语言编写的函数调用验证器此前未被检查，可能被用于权限提升。修复方法是在每个验证器函数中加入对权限检查函数的调用。
  非核心过程语言如果自带验证器函数，也需要做同样的修改。（CVE-2014-0061）

- 避免在表和索引 DDL 期间进行多次名称查找（Robert Haas、Andres Freund）

  如果由于并发活动导致名称查找得出不同结论，我们可能会对不同的表执行 DDL 的不同部分。至少在 `CREATE INDEX` 的情况下，这可被用来让权限检查针对与索引创建不同的表执行，从而造成权限提升攻击。（CVE-2014-0062）

- 防止长日期时间字符串造成缓冲区越界（Noah Misch）

  `MAXDATELEN` 常量对于 `interval` 类型可能的最长值来说太小，导致 `interval_out()` 中可能出现缓冲区越界。虽然日期时间输入函数在避免缓冲区越界方面更谨慎，但该限制仍短到导致它们拒绝某些有效输入，例如包含很长时区名的输入。ecpg 库存在这些漏洞以及它自身的一些漏洞。（CVE-2014-0063）

- 防止大小计算中的整数溢出导致缓冲区越界（Noah Misch、Heikki Linnakangas）

  若干函数（大多是类型输入函数）在计算分配大小时未检查溢出。如果发生溢出，
  就会分配过小的缓冲区并被写越界。（CVE-2014-0064）

- 防止固定大小缓冲区越界（Peter Eisentraut、Jozef Mlich）

  使用 `strlcpy()` 及相关函数，以明确保证固定大小缓冲区不会被越界写入。与前面几项不同，这些情况是否真的构成现实问题尚不清楚，因为在大多数情况下，
  输入字符串的大小似乎已有先前约束。尽管如此，消除所有此类 Coverity 警告似乎是明智的。（CVE-2014-0065）

- 避免在 `crypt()` 返回 NULL 时崩溃（Honza Horak、Bruce Momjian）

  `crypt()` 可能返回 NULL 的场景相对较少，但如果发生，`contrib/chkpass` 就会崩溃。一个可能出问题的实际情况是 libc 被配置为拒绝执行未获批准的哈希算法（例如“FIPS 模式”）。（CVE-2014-0066）

- 在回归测试说明中记录 `make check` 的风险（Noah Misch、Tom Lane）

  由于 `make check` 启动的临时服务器使用“trust”认证，同一台机器上的其他用户可以以数据库超级用户身份连接它，进而可能利用启动测试的操作系统用户的权限。
  未来版本可能会在测试流程中加入防止此风险的变更，但需要先进行一些公开讨论。

因此目前只是警告：当同一台机器上存在不受信任的用户时，不要使用 `make check`。（CVE-2014-0067）

- 修复当关系的某些段不是完整大小时 WAL 记录可能被错误重放的问题（Greg Stark、Tom Lane）

  WAL 更新可能被应用到错误的页面，甚至可能偏后很多页。除了损坏数据外，已观察到这一错误导致备库相比主库显著“膨胀”，因为更新被应用到了远超文件末尾应在位置的地方。
  这种失败模式在崩溃恢复期间似乎不构成重大风险，
  只在初始同步从快速变化的主库基础备份创建的备库时出现。

- 修复判断恢复何时达到一致性的缺陷（Tomonari Katsumata、Heikki Linnakangas）

  某些情况下，WAL 重放会错误地在重放开始时就断定数据库已一致，从而可能在数据库真正一致之前允许热备查询。也可能出现其他症状，例如“PANIC: WAL contains references to invalid pages”。

- 修复在热备模式下重放 `VACUUM` 操作时对 btree 索引页的不当锁定（Andres Freund、Heikki Linnakangas、Tom Lane）

  这一错误可能导致“PANIC: WAL contains references to invalid pages”失败。

- 确保对 GIN 索引非叶页面的插入在合适时写入整页 WAL 记录（Heikki Linnakangas）

  先前的代码在系统崩溃期间发生部分页写入时存在索引损坏风险。

- 修复服务器进程退出期间的竞态条件（Robert Haas）

  确保信号处理程序不会在进程的 `MyProc` 指针不再有效后尝试使用它。

- 修复错误报告逻辑中对 `errno` 的不安全引用（Christian Kruse）

  这通常会导致奇怪的行为，例如 `HINT` 字段缺失或不恰当。

- 修复服务器启动期间过早使用 `ereport()` 可能导致的崩溃（Tom Lane）

  我们在实际中见到的主要情况是：如果服务器在一个它没有读取权限的目录中被启动，
  就会崩溃。

- 在 OpenSSL 套接字写函数中正确清除重试标志（Alexander Kukushkin）

  这一遗漏可能导致 SSL 加密连接意外断开后服务器锁死。

- 修复对包含转义的 Unicode 标识符（`U&"..."` 语法）的长度检查（Tom Lane）

  如果标识符的转义形式过长，就会为这类标识符打印虚假的截断警告，
  但标识符在反转义之后实际上并不需要截断。

- 允许作为类型名的关键字用于角色列表（Stephen Frost）

  先前的补丁允许这类关键字在角色标识符等位置不加引号使用；
  但漏掉了允许角色标识符列表的情况，例如 `DROP ROLE`。

- 修复嵌套子选择（例如 `WHERE (... x IN (SELECT ...) ...) IN (SELECT ...)`）因无效计划而可能崩溃的问题（Tom Lane）

- 确保即使表列中的所有值都“过宽”，`ANALYZE` 也会为该列创建统计信息（Tom Lane）

ANALYZE 有意在直方图和最常用值计算中省略非常宽的值，但在所有采样条目都过宽的情况下没有做出合理的处理。

- 在 ALTER TABLE ... SET TABLESPACE 中，允许不经权限检查就使用数据库的默认表空间（Stephen Frost）

  CREATE TABLE 一直允许这种用法，但 ALTER TABLE 没有照做。

- 修复当 CASE 的某些分支返回集合而另一些不返回时的 “cannot accept a set” 错误（Tom Lane）
- 修复 pgstat 函数中对全零客户端地址的检查（Kevin Grittner）
- 修复文本搜索解析器可能错误分类多字节字符的问题（Tom Lane）

  在 C 区域设置下使用多字节编码时，非 ASCII 字符可能被错误分类。在 Cygwin 上，非 C 区域设置也可能失败。

- 修复 plainto_tsquery() 中可能的不当行为（Heikki Linnakangas）

  在复制重叠内存区域时使用 memmove() 而不是 memcpy()。尚无实际造成问题的现场报告，但这无疑是有风险的。

- 在区域检查方面接受 SHIFT_JIS 作为编码名（Tatsuo Ishii）
- 修复 PQhost() 在 Windows 上的不当行为（Fujii Masao）

  如果未指定主机，它应返回 localhost。

- 改进 libpq 和 psql 对 COPY TO STDOUT/FROM STDIN 期间失败的错误处理（Tom Lane）

  这尤其修复了 9.2 及更高版本中，如果服务器连接在 COPY FROM STDIN 期间丢失可能发生的无限循环。该场景的变体在旧版本或其他客户端应用中也可能出现。

- 修复 ecpg 中未对齐的描述符（MauMau）
- 在 ecpg 中正确处理连接参数中缺少主机名的情况（Michael Meskes）
- 修复 contrib/dblink 连接启动中的性能退化（Joe Conway）

  在客户端和服务器编码一致时避免一次不必要的往返。

- 修复 contrib/isn 中 ISMN 值校验位计算错误的问题（Fabien Coelho）
- 确保仅安装客户端代码的流程按文档说明工作（Peter Eisentraut）
- 在 Mingw 和 Cygwin 构建中，把 libpq DLL 安装到 bin 目录（Andrew Dunstan）

  这与 MSVC 构建长期以来的做法一致。它应该能解决 psql 等程序因找不到 DLL 而无法启动的问题。

- 在 Cygwin 构建中避免使用已弃用的 dllwrap 工具（Marco Atzeri）
- 不再生成纯文本的 HISTORY 和 src/test/regress/README 文件（Tom Lane）

  这些文本文件与主文档的 HTML 和 PDF 格式重复。维护它们的麻烦远远超过纯文本格式可能的读者。发行版压缩包仍会包含这些名字的文件，但它们只是提示读者查阅主文档的存根。纯文本 INSTALL 文件仍将维护，因为它可以说还有使用场景。

- 把时区数据文件更新为 tzdata 版本 2013i for DST law changes in Jordan and historical changes in Cuba。

  此外，时区 `Asia/Riyadh87`、`Asia/Riyadh88` 和 `Asia/Riyadh89` 已被移除，因为 IANA 不再维护它们，而且它们从未代表过实际的民用计时实践。

E.34. 版本 9.0.15

发行日期
2013-12-05

本次发布包含对 9.0.14 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.34.1. 迁移到版本 9.0.15

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了若干潜在的数据损坏问题。请参见下面前三条变更日志，
以确认你的安装是否受影响以及可以采取什么措施。

不过，如果你是从早于 9.0.13 的版本升级，请参见第 E.36 节。

E.34.2. 变更

- 修复 `VACUUM` 判断是否可以更新 `relfrozenxid` 的测试（Andres Freund）

  某些情况下，`VACUUM`（手工或自动）可能错误地推进表的 `relfrozenxid` 值，使元组逃过冻结，导致这些行在经过 2^31 个事务后变得不可见。由于需要多次错误推进才会发生实际丢失，数据丢失的概率相当低，但并非为零。从 9.0.4、8.4.8 或更早版本升级的用户不受影响，但之后的所有版本都包含该缺陷。

  升级后，在把 `vacuum_freeze_table_age` 设为零的情况下对所有数据库中的所有表执行清理，可以缓解此问题。这会修复任何潜伏的损坏，但无法修复所有既有的数据错误。不过，
  如果安装在其生命周期内执行的事务少于 2^31 个更新（可用 `SELECT txid_current() < 2^31` 检查），则在执行此清理之后即可认为该安装是安全的。

- 修复热备启动期间 `pg_clog` 和 `pg_subtrans` 的初始化（Andres Freund、Heikki Linnakangas）

  这一缺陷会把已提交事务标记为未提交，
  从而在备库开始接受热备查询的那一刻造成数据丢失。
  除非备库启动时主库自上次检查点以来执行了大量更新事务，否则这种损坏的可能性很小。
  症状包括行缺失、本应删除的行仍可见，以及已更新行的过时版本与新版本同时可见。

  该缺陷是在 9.3.0、9.2.5、9.1.10 和 9.0.14 版本中引入的。只运行过更早版本的备库没有风险。建议曾运行过任何含缺陷版本的备库在升级后从主库重新克隆（例如用新的基础备份）。

- 在 WAL 重放期间截断 `pg_multixact` 内容（Andres Freund）

  这避免了备库中磁盘空间消耗不断增加。

- 修复 GIN 索引 posting 树页面删除中的竞态条件（Heikki Linnakangas）

这可能导致瞬时的错误结果或查询失败。

- 避免展平 `SELECT` 列表中包含包裹在子 `SELECT` 中的易变函数的子查询（Tom Lane）

  这避免了因易变函数被额外求值而导致的意外结果。

- 修复规划器对外连接中嵌套的非简单变量子查询输出的处理（Tom Lane）

  这一错误可能为涉及 `JOIN` 语法中多级子查询的查询生成错误的计划。

- 修复临时文件被过早删除的问题（Andres Freund）
- 修复规则打印中可能读取越过内存末尾的问题（Peter Eisentraut）
- 修复 `int2vector` 和 `oidvector` 值的数组切片（Tom Lane）

  这类表达式现在会被隐式提升为常规的 `int2` 或 `oid` 数组。

- 修复使用 SQL 标准的简单 GMT 偏移时区时的错误行为（Tom Lane）

  某些情况下，系统会使用简单 GMT 偏移值，而本应使用选择简单偏移之前一直生效的常规时区设置。此变更还使 `timeofday` 函数遵循简单 GMT 偏移时区。

- 防止记录 Windows 错误码翻译时可能出现的不当行为（Tom Lane）
- 正确地为 pg_ctl 生成的命令行加引号（Naoya Anzai、Tom Lane）

  此修复仅适用于 Windows。

- 修复当源数据库通过 `ALTER DATABASE SET` 设置 `default_transaction_read_only` 时 pg_dumpall 无法工作的问题（Kevin Grittner）

  此前，生成的脚本在恢复期间会失败。

- 修复 ecpg 对声明为 `varchar` 的变量列表的处理（Zoltán Böszörményi）
- 让 `contrib/lo` 防御不正确的触发器定义（Marc Cousin）
- 把时区数据文件更新为 tzdata 版本 2013h for DST law changes in Argentina, Brazil, Jordan, Libya, Liechtenstein, Morocco, and Palestine。Also, new timezone abbreviations WIB, WIT, WITA for Indonesia.

E.35. 版本 9.0.14

发行日期
2013-10-10

本次发布包含对 9.0.13 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.35.1. 迁移到版本 9.0.14

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.13 的版本升级，请参见第 E.36 节。

E.35.2. 变更

- 防止尝试对标识符做大小写折叠时损坏多字节字符（Andrew Dunstan）

  PostgreSQL 只在使用单字节服务器编码时才对非 ASCII 字符做大小写折叠。

- 修复 `wal_level = hot_standby` 时后台写入进程中的检查点内存泄漏（Naoya Anzai）
- 修复由 `lo_open()` 失败导致的内存泄漏（Heikki Linnakangas）
- 修复 `work_mem` 使用超过 24GB 内存时的内存过度提交缺陷（Stephen Frost）
- 修复使用 SSL 时 libpq 中的死锁缺陷（Stephen Frost）
- 修复多线程 libpq 应用中可能的 SSL 状态损坏（Nick Phillips、Stephen Frost）
- 正确计算包含大量 NULL 值的布尔列的行数估计（Andrew Gierth）

  此前，`col IS NOT TRUE` 和 `col IS NOT FALSE` 这类测试在估计计划代价时没有正确考虑 NULL 值。

- 防止把 `WHERE` 子句下推到不安全的 `UNION/INTERSECT` 子查询中（Tom Lane）

  `SELECT` 列表中包含集合返回函数或易变函数的 `UNION` 或 `INTERSECT` 子查询可能被不当优化，导致运行时错误或错误的查询结果。

- 修复罕见的 “failed to locate grouping columns” 规划器失败情况（Tom Lane）
- 改进视图转储代码对被引用表中已删除列的处理（Tom Lane）
- 正确记录使用 `UNIQUE` 和 `PRIMARY KEY` 语法创建的索引注释（Andres Freund）

  这修复了一个并行 pg_restore 失败问题。

- 修复 `REINDEX TABLE` 和 `REINDEX DATABASE`，使其正确重新验证约束并把已失效的索引标记为有效（Noah Misch）

  `REINDEX INDEX` 一直工作正常。

- 修复并发执行 `CREATE INDEX CONCURRENTLY` 操作时可能出现的死锁（Tom Lane）
- 修复 `regexp_matches()` 对零长度匹配的处理（Jeevan Chalke）

  此前，像 '^' 这样的零长度匹配可能返回过多的匹配。

- 修复过于复杂的正则表达式导致崩溃的问题（Heikki Linnakangas）
- 修复反向引用与非贪婪量词组合时正则表达式匹配失败的问题（Jeevan Chalke）
- 除非启用了函数体检查，否则防止 `CREATE FUNCTION` 检查 `SET` 变量（Tom Lane）
- 允许 `ALTER DEFAULT PRIVILEGES` 在不需要 CREATE 权限的情况下对模式操作（Tom Lane）
- 放宽对查询中所用关键字的限制（Tom Lane）

具体而言，放宽了角色名、语言名、`EXPLAIN` 和 `COPY` 选项以及 `SET` 值的关键字限制。这使得 `COPY ... (FORMAT BINARY)` 能按预期工作；此前 `BINARY` 需要加引号。

- 修复 `pgp_pub_decrypt()`，使其可用于带口令的密钥（Marko Kreen）
- 移除无索引表清理期间罕见的不准确警告（Heikki Linnakangas）
- 确保当 `VACUUM ANALYZE` 的文件截断尝试因锁冲突被取消时，仍然执行 ANALYZE 阶段（Kevin Grittner）
- 避免在预备查询中执行事务控制命令（例如 ROLLBACK）时可能出现的失败（Tom Lane）
- 确保浮点数据输入在所有平台上都接受标准的“infinity”（无穷大）写法（Tom Lane）

  C99 标准规定允许的写法为 `inf`、`+inf`、`-inf`、`infinity`、`+infinity` 和 `-infinity`。即使平台的 `strtod` 函数不支持，也确保我们能识别这些写法。
- 扩展行与记录和数组比较的能力（Rafal Rzepecki、Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2013d for DST law changes in Israel, Morocco, Palestine, and Paraguay。Also, historical zone data corrections for Macquarie Island.

E.36. 版本 9.0.13

发行日期
2013-04-04

本次发布包含对 9.0.12 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.36.1. 迁移到版本 9.0.13

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了 GiST 索引管理中的若干错误。安装此更新后，建议对满足下述一个或多个条件的 GiST 索引执行 `REINDEX`。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.36.2. 变更

- 修复服务器命令行开关解析不安全的问题（Mitsumasa Kondo、Kyotaro Horiguchi）

  包含以“-”开头的数据库名的连接请求可能被精心构造，
  用于破坏或摧毁服务器数据目录中的文件，即使该请求最终被拒绝也是如此。（CVE-2013-1899）
- 在每个 postmaster 子进程中重置 OpenSSL 随机状态（Marko Kreen）

  这避免了另一种情况：另一数据库用户可能相对容易地猜出 `contrib/pgcrypto` 函数生成的随机数。只有当 postmaster 配置了 `ssl = on` 而大多数连接并不使用 SSL 加密时，这一风险才比较显著。（CVE-2013-1900）

- 修复 GiST 索引在不合适的情况下使用“模糊”几何比较的问题（Alexander Korotkov）

  核心几何类型使用“模糊”相等进行比较，但 `gist_box_same` 必须做精确比较，否则使用它的 GiST 索引可能变得不一致。安装此更新后，用户应对 `box`、`polygon`、`circle` 或 `point` 列上的所有 GiST 索引执行 `REINDEX`，因为它们都使用 `gist_box_same`。

- 修复对可变宽度数据类型（即 `text`、`bytea`、`bit` 和 `numeric` 列）使用 `contrib/btree_gist` 的 GiST 索引中错误的范围并集和惩罚逻辑（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，还会造成无用的索引膨胀。建议用户在安装此更新后对这类索引执行 `REINDEX`。

- 修复多列索引的 GiST 页面分裂代码中的缺陷（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，还会使索引的搜索效率不必要地低下。建议用户在安装此更新后对多列 GiST 索引执行 `REINDEX`。

- 修复 `gist_point_consistent`，使其一致地处理模糊性（Alexander Korotkov）

  `point` 列上 GiST 索引的索引扫描有时会得出与顺序扫描不同的结果，因为 `gist_point_consistent` 与底层操作符代码在是做精确比较还是模糊比较上不一致。

- 修复 WAL 重放中的缓冲区泄漏（Heikki Linnakangas）

  这一缺陷可能导致重放期间出现“incorrect local pin count”错误，使恢复无法进行。

- 修复 `DELETE RETURNING` 中的竞态条件（Tom Lane）

  在合适的条件下，`DELETE RETURNING` 可能尝试从当前进程已不再持有任何 pin 的共享缓冲区中取数据。如果其他进程在此期间更改了该缓冲区，就会导致垃圾 `RETURNING` 输出，甚至崩溃。

- 修复正则表达式编译中的无限循环风险（Tom Lane、Don Porter）

- 修复正则表达式编译中潜在的空指针解引用（Tom Lane）

- 修复 `to_char()`，在适当的地方只使用 ASCII 大小写折叠规则（Tom Lane）

  这修复了一些本应与区域设置无关的模板模式在土耳其区域下错误处理“`I`”和“`i`”的问题。

- 修复时间戳 `1999-12-31 24:00:00` 被不当拒绝的问题（Tom Lane）

- 修复单个事务先执行 `UNLISTEN` 再执行 `LISTEN` 时的逻辑错误（Tom Lane）

  会话最终完全没有监听通知事件，尽管在这种情况下它显然应该监听。

- 移除无用的“picksplit doesn't support secondary split”日志消息（Josh Hansen、Tom Lane）

  这条消息似乎是为预期中从未编写、而且可能永远不会编写的代码而添加的，因为 GiST 对二次分裂的默认处理其实相当好。所以不再为此打扰最终用户。

- 修复会话最后几个事务的提交/中止计数可能未能发送到统计收集器的问题（Tom Lane）

- 消除 PL/Perl 的 `spi_prepare()` 函数中的内存泄漏（Alex Hunsaker、Tom Lane）

- 修复 pg_dumpall 对包含“`=`”的数据库名的处理（Heikki Linnakangas）

- 避免在给定不正确的连接串时 pg_dump 崩溃（Heikki Linnakangas）
- 让 pg_dump 和 pg_upgrade 忽略无效索引（Michael Paquier、Bruce Momjian）

  转储无效索引可能在恢复时造成问题，
  例如当索引创建失败的原因是它试图强制执行表数据不满足的唯一性条件时。此外，
  如果索引创建实际上仍在进行中，把它视为未提交的 DDL 变更似乎是合理的，pg_dump 本来也不应转储它。pg_upgrade 现在也跳过无效索引而不是失败。
- 修复 `contrib/pg_trgm` 的 `similarity()` 函数，使其对无三元组的字符串返回零（Tom Lane）

  此前它由于内部除以零而返回 `NaN`。
- 把时区数据文件更新为 tzdata 版本 2013b for DST law changes in Chile, Haiti, Morocco, Paraguay, and some Russian areas。Also, historical zone data corrections for numerous places.

  此外，更新时区缩写文件以反映俄罗斯和其他地方的近期变更：`CHOT`、`GET`、`IRKT`、`KGT`、`KRAT`、`MAGT`、`MAWT`、`MSK`、`NOVT`、`OMST`、`TKT`、`VLAT`、`WST`、`YAKT`、`YEKT` 现在遵循其当前含义，同时把 `VOLT`（Europe/Volgograd）和 `MIST`（Antarctica/Macquarie）加入默认缩写列表。

E.37. 版本 9.0.12

发行日期
2013-02-07

本次发布包含对 9.0.11 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.37.1. 迁移到版本 9.0.12

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.37.2. 变更

- 防止从 SQL 执行 `enum_recv`（Tom Lane）

  该函数声明有误，允许一条简单的 SQL 命令使服务器崩溃。原则上攻击者可能利用它检查服务器内存的内容。感谢 Sumit Soni（通过 Secunia SVCRP）报告此问题。（CVE-2013-0255）
- 修复 WAL 重放期间检测何时达到一致数据库状态的多个问题（Fujii Masao、Heikki Linnakangas、Simon Riggs、Andres Freund）
- 在截断关系文件时更新最小恢复点（Heikki Linnakangas）

  数据一旦被丢弃，在时间线上更早的点停止恢复就不再安全。
- 修复热备模式下缺失的取消操作（Noah Misch、Simon Riggs）

取消冲突热备查询的需要有时会被遗漏，使这些查询看到不一致的数据。

- 修复 SQL 文法，允许对子 SELECT 结果做下标或字段选择（Tom Lane）
- 修复繁忙负载下自动清理截断的性能问题（Jan Wieck）

  截断表末尾的空页需要排他锁，
  但自动清理被编码为在出现冲突锁请求时失败（并释放表锁）。在负载下，
  截断很可能永远不会发生，导致表膨胀。修复方法是先执行部分截断，释放锁，
  然后尝试重新获取锁并继续。
  此修复还大幅缩短了冲突请求到达后自动清理释放锁的平均时间。

- 防止扫描 `pg_tablespace` 时的竞态条件（Stephen Frost、Tom Lane）

  如果 `pg_tablespace` 条目存在并发更新，`CREATE DATABASE` 和 `DROP DATABASE` 可能行为异常。

- 防止 `DROP OWNED` 试图删除整个数据库或表空间（Álvaro Herrera）

  为安全起见，这些对象的所有权必须重新分配，而不是删除。

- 修复 `vacuum_freeze_table_age` 实现中的错误（Andres Freund）

  在已存在超过 `vacuum_freeze_min_age` 个事务的安装中，这一失误使自动清理无法使用部分表扫描，而总是进行全表扫描。

- 防止 `RowExpr` 或 `XmlExpr` 被二次解析分析时的不当行为（Andres Freund、Tom Lane）

  这一失误在 `CREATE TABLE LIKE INCLUDING INDEXES` 等上下文中可能被用户看到。

- 增强对哈希表大小计算中整数溢出的防御（Jeff Davis）
- 拒绝 `to_date()` 中超出范围的日期（Hitoshi Harada）
- 确保在 Windows 上非 ASCII 提示符字符串被转换到正确的代码页（Alexander Law、Noah Misch）

  这一缺陷影响 psql 和其他一些客户端程序。

- 修复未连接数据库时 psql 的 `\?` 命令可能崩溃的问题（Meng Qingzhong）
- 修复 pg_upgrade，使其安全处理无效索引（Bruce Momjian）
- 修复 libpq 的 `PQprintTuples` 中的一字节缓冲区越界（Xi Wang）

  这个古老的函数没有被 PostgreSQL 本身在任何地方使用，但仍可能被某些客户端代码使用。

- 让 ecpglib 正确使用翻译后的消息（Chen Huajun）
- 在 MSVC 上正确安装 ecpg_compat 和 pgtypes 库（Jiang Guiqing）
- 如果系统未提供 `isinf()`，则在 libecpg 中包含我们自己的版本（Jiang Guiqing）
- 重新排列 configure 对系统提供函数的测试，使其不被 libedit/libreadline 的虚假导出所欺骗（Christoph Berg）
- 确保 Windows 构建号随时间递增（Magnus Hagander）

- 让 pgxs 在为 Windows 交叉编译时以正确的 `.exe` 后缀构建可执行文件（Zoltan Boszormenyi）
- 新增时区缩写 `FET`（Tom Lane）

  它现在用于一些东欧时区。

E.38. 版本 9.0.11

发行日期
2012-12-06

本次发布包含对 9.0.10 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.38.1. 迁移到版本 9.0.11

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.38.2. 变更

- 修复与 `CREATE INDEX CONCURRENTLY` 相关的多个缺陷（Andres Freund、Tom Lane）

  修复 `CREATE INDEX CONCURRENTLY`，使其在更改索引 `pg_index` 行的状态时使用就地更新。这防止了可能导致并发会话错过更新目标索引（从而产生损坏的并发创建索引）的竞态条件。

  另外，修复各种其他操作，确保它们忽略由失败的 `CREATE INDEX CONCURRENTLY` 命令产生的无效索引。其中最重要的是 `VACUUM`，因为在采取纠正措施修复或删除无效索引之前，很容易就有自动清理在该表上启动。

- 修复 WAL 重放期间的缓冲区锁定（Tom Lane）

  WAL 重放代码在重放影响多个页面的 WAL 记录时，对缓冲区加锁不够谨慎。这可能导致热备查询瞬时看到不一致的状态，造成错误结果或意外失败。

- 修复 GIN 索引 WAL 生成逻辑中的一个错误（Tom Lane）

  如果发生残缺页故障，这可能导致索引损坏。

- 把热备服务器提升为正常运行时，正确移除启动进程的虚拟 XID 锁（Simon Riggs）

  这一疏忽可能阻止后续执行某些操作，例如 `CREATE INDEX CONCURRENTLY`。

- 避免备库模式下虚假的 “out-of-sequence timeline ID” 错误（Heikki Linnakangas）
- 防止 postmaster 在收到关闭信号后继续启动新的子进程（Tom Lane）

  这一失误可能导致关闭耗时比应有的更长，甚至在没有额外用户操作的情况下永远无法完成。

- 避免内存不足时内部哈希表损坏（Hitoshi Harada）

- 修复外连接之上非严格等值子句的规划（Tom Lane）

  规划器可能从把非严格构造与别的值等同的子句中推导出错误的约束，例如当 `foo` 来自外连接的可空侧时的 `WHERE COALESCE(foo, 0) = 0`。

- 改进规划器从等价类证明排他约束的能力（Tom Lane）

- 修复哈希子计划中的部分行匹配，使其正确处理跨类型情况（Tom Lane）

  这影响多列 `NOT IN` 子计划，例如 `WHERE (a, b) NOT IN (SELECT x, y FROM ...)`，当 `b` 和 `y` 分别是 `int4` 和 `int8` 时。根据所涉及的具体数据类型，这一失误会导致错误答案或崩溃。

- 在为 `AFTER ROW UPDATE/DELETE` 触发器重新获取旧元组时获取缓冲区锁（Andres Freund）

  在非常罕见的情况下，这一疏忽可能导致向外键强制触发器的预检查逻辑传递不正确的数据。这可能导致崩溃，或者在是否触发触发器上做出错误决定。

- 修复`ALTER COLUMN TYPE`以正确重建继承的检查约束（Tom Lane）

  这在 8.4 之前的版本中工作正常，现在在 8.4 及更高版本中也工作正常。

- 修复 `REASSIGN OWNED` 对表空间授权的处理（Álvaro Herrera）

- 忽略视图中系统列的不正确 `pg_attribute` 条目（Tom Lane）

  视图没有任何系统列。然而，我们在把表转换为视图时忘记移除这类条目。这一问题在 9.3 及更高版本中已正确修复，但在以前的分支中，我们需要防御已经错误转换的视图。

- 修复规则打印，使其正确转储 `INSERT INTO table DEFAULT VALUES`（Tom Lane）

- 防止当查询中有过多 `UNION/INTERSECT/EXCEPT` 子句时栈溢出（Tom Lane）

- 防止用最小可能整数值除以 -1 时依赖平台的失败（Xi Wang、Tom Lane）

- 修复日期解析中可能越过字符串末尾的访问（Hitoshi Harada）

- 修复当 XID 回卷发生在检查点期间且 `wal_level` 为 `hot_standby` 时 XID 纪元（epoch）未能推进的问题（Tom Lane、Andres Freund）

  虽然这一失误对 PostgreSQL 本身没有特别影响，但它对依赖 `txid_current()` 及相关函数的应用不利：TXID 值看起来会倒退。

- 当 Unix 域套接字的路径名长度超过平台特定限制时，生成可理解的错误消息（Tom Lane、Andrew Dunstan）

  此前这会导致相当无用的结果，例如“Non-recoverable failure in name resolution”。

- 修复向客户端发送组合列值时的内存泄漏（Tom Lane）

- 使 pg_ctl 读取 `postmaster.pid` 文件更健壮（Heikki Linnakangas）

  这修复了竞态条件和可能的文件描述符泄漏。

- 修复当呈现错误编码的数据且 `client_encoding` 设置为仅客户端编码（如 SJIS）时 psql 可能崩溃的问题（Jiang Guiqing）

- 修复 pg_dump 以 `tar` 输出格式生成的 `restore.sql` 脚本中的缺陷（Tom Lane）

  该脚本在表名包含大写字符时会直接失败。此外，让该脚本既能以常规 COPY 模式也能以 `--inserts` 模式恢复数据。

- 修复 pg_restore，使其接受符合 POSIX 的 `tar` 文件（Brian Weaver、Tom Lane）

  pg_dump 的 `tar` 输出模式的原始编码生成的文件并不完全符合 POSIX 标准。这在 9.3 版本中已得到纠正。此补丁更新了以前的分支，使它们同时接受错误格式和纠正后的格式，希望借此避免 9.3 发布时的兼容性问题。

- 修复 pg_resetxlog，使其在给定数据目录相对路径时正确定位 `postmaster.pid`（Tom Lane）

  这一失误可能导致 pg_resetxlog 没有注意到有活动 postmaster 正在使用该数据目录。

- 修复 libpq 的 `lo_import()` 和 `lo_export()` 函数，使其正确报告文件 I/O 错误（Tom Lane）

- 修复 ecpg 对嵌套结构指针变量的处理（Muhammad Usama）

- 修复 ecpg 的 `ecpg_get_data` 函数，使其正确处理数组（Michael Meskes）

- 让 `contrib/pageinspect` 的 btree 页面检查函数在检查页面时持有缓冲区锁（Tom Lane）

- 修复 pgxs 在 AIX 上构建可加载模块的支持（Tom Lane）

  在原始源码树之外构建模块在 AIX 上不工作。

- 把时区数据文件更新为 tzdata 版本 2012j for DST law changes in Cuba, Israel, Jordan, Libya, Palestine, Western Samoa, and portions of Brazil。

E.39. 版本 9.0.10

发行日期
2012-09-24

本次发布包含对 9.0.9 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.39.1. 迁移到版本 9.0.10

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.39.2. 变更

- 修复规划器对执行器参数的分配，并修复执行器对 CTE 计划节点的重新扫描逻辑（Tom Lane）

  这些错误可能导致多次扫描同一 `WITH` 子查询的查询给出错误答案。

- 改进 GiST 索引的页面分裂决策（Alexander Korotkov、Robert Haas、Tom Lane）

多列 GiST 索引可能因此错误而遭受意外膨胀。

- 修复级联权限撤销，使其在权限仍被持有时停止（Tom Lane）

  如果我们撤销某个角色 *X* 的授予选项，但 *X* 仍通过别人的授予持有该选项，我们就不应从 *X* 曾授予该权限的角色 *Y* 那里递归撤销相应权限。

- 改进热备配置错误的错误消息（Gurjeet Singh）

- 修复使用 PL/Perl 时对 `SIGFPE` 的处理（Andres Freund）

  Perl 会把进程的 `SIGFPE` 处理程序重置为 `SIG_IGN`，这可能导致之后崩溃。在初始化 PL/Perl 之后恢复正常的 Postgres 信号处理程序。

- 防止递归 PL/Perl 函数在执行期间被重定义时 PL/Perl 崩溃（Tom Lane）

- 规避 PL/Perl 中可能的错误优化（Tom Lane）

  一些 Linux 发行版包含错误版本的 `pthread.h`，导致 PL/Perl 中的编译代码不正确，当一个 PL/Perl 函数调用另一个会抛出错误的函数时会导致崩溃。

- 修复 pg_upgrade 在 Windows 上对行尾的处理（Andrew Dunstan）

  此前，pg_upgrade 可能在函数体等地方添加或删除回车符。

- 在 Windows 上，让 pg_upgrade 在其生成的脚本中使用反斜杠路径分隔符（Andrew Dunstan）

- 把时区数据文件更新为 tzdata 版本 2014e，以反映克里米亚、埃及和摩洛哥的夏令时法规变更。

E.40. 版本 9.0.9

发行日期
2012-08-17

本次发布包含对 9.0.8 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.40.1. 迁移到版本 9.0.9

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.40.2. 变更

- 防止通过 XML 实体引用访问外部文件/URL（Noah Misch、Tom Lane）

  `xml_parse()` 会尝试按需获取外部文件或 URL 来解析 XML 值中的 DTD 和实体引用，从而使无特权数据库用户能够尝试以数据库服务器的权限获取数据。虽然外部数据不会直接返回给用户，但如果数据不能解析为有效 XML，其部分内容可能暴露在错误消息中；而且无论如何，仅仅能够检查文件是否存在对攻击者就可能有用。（CVE-2012-3489）

- 防止通过 `contrib/xml2` 的 `xslt_process()` 访问外部文件/URL（Peter Eisentraut）

  libxslt 提供通过样式表命令读写文件和 URL 的能力，从而使无特权数据库用户能够以数据库服务器的权限读写数据。现通过正确使用 libxslt 的安全选项禁用该能力。（CVE-2012-3488）

  此外，移除了 `xslt_process()` 从外部文件/URL 获取文档和样式表的能力。虽然这是一个有记载的“特性”，但长期以来被认为是个坏主意。CVE-2012-3489 的修复破坏了该能力，与其花费精力修复它，我们干脆将其移除。

- 防止 btree 索引页被过早回收（Noah Misch）

  当我们允许只读事务跳过分配 XID 时，引入了已删除 btree 页在只读事务仍在访问它时就被回收的可能性。这会导致错误的索引搜索结果。由于时序要求，
  这种错误在实际中发生的概率似乎很低，但仍应修复。

- 修复新创建或重置序列的崩溃安全缺陷（Tom Lane）

  如果在新创建或重置的序列上执行了 `ALTER SEQUENCE`，然后恰好对它调用了一次 `nextval()`，随后服务器崩溃，WAL 重放会把序列恢复到看起来没有执行过 `nextval()` 的状态，从而使下一个 `nextval()` 调用再次返回第一个序列值。这种情况尤其可能出现在 `serial` 列上，因为创建 serial 列的序列包含一个 `ALTER SEQUENCE OWNED BY` 步骤。

- 修复 `txid_current()`，使其在非热备下报告正确的纪元（epoch）（Heikki Linnakangas）

  这修复了上一个小版本引入的退化。

- 修复主库事务有大量子事务时热备启动的缺陷（Andres Freund）

  这一失误导致报告为“out-of-order XID insertion in KnownAssignedXids”的失败。

- 确保 `pg_start_backup()` 之后对 `backup_label` 文件执行 fsync（Dave Kerr）

- 修复 walsender 进程中的超时处理（Tom Lane）

  WAL 发送端后台进程疏于建立 SIGALRM 处理程序，这意味着在某些应发生超时的极端情况下它们会永远等待。

- 回移植压缩 fsync 请求队列的 9.1 改进（Robert Haas）

  这改进了检查点期间的性能。9.1 的变更已经过足够的实际测试，看起来可以安全回移植。

- 修复 `LISTEN/NOTIFY`，使其更好地应对 I/O 问题（例如磁盘空间耗尽）（Tom Lane）

  写入失败之后，所有后续发送更多 `NOTIFY` 消息的尝试都会失败，并显示诸如“Could not read from file "pg_notify/*nnnn*" at offset *nnnnn*: Success”的消息。

- 只允许被直接阻塞的进程自动取消自动清理（Tom Lane）

  原始编码在某些情况下可能允许不一致的行为；特别是自动清理可能在不足 `deadlock_timeout` 宽限期之后就被取消。

- 改进自动清理取消的日志记录（Robert Haas）

- 修复日志收集器，使 `log_truncate_on_rotation` 在服务器启动后的第一次日志轮转期间就能工作（Tom Lane）

- 修复附加到嵌套集合操作（`UNION/INTERSECT/EXCEPT`）上的 `WITH`（Tom Lane）

- 确保对子查询的整行引用不包含任何多余的 `GROUP BY` 或 `ORDER BY` 列（Tom Lane）
- 禁止在 `CREATE TABLE` 期间复制 `CHECK` 约束和索引定义中的整行引用（Tom Lane）

  这种情况可能出现在带 `LIKE` 或 `INHERITS` 的 `CREATE TABLE` 中。被复制的整行变量被错误地标记为原表而不是新表的行类型。对 `LIKE` 拒绝这种情况似乎是合理的，因为行类型之后很可能分化。对 `INHERITS`，理想情况下应该允许它，并隐式强制转换为父表的行类型；但这需要的工作量超出了回移植的安全范围。
- 修复 `ARRAY(SELECT ...)` 子查询中的内存泄漏（Heikki Linnakangas、Tom Lane）
- 修复从正则表达式中提取公共前缀的问题（Tom Lane）

  代码可能被带量词的带括号子表达式（例如 `^(foo)?bar`）混淆。这会导致对此类模式搜索的索引优化不正确。
- 修复解析 `interval` 常量中带符号的 *hh*:*mm* 和 *hh*:*mm*:*ss* 字段的缺陷（Amit Kapila、Tom Lane）
- 在 PL/Python 中把 Python Unicode 字符串转换为服务器编码时，使用 Postgres 而不是 Python 的编码转换函数（Jan Urbanski）

  这避免了一些极端情况的问题，尤其是 Python 并不支持 Postgres 的所有编码。一个显著的功能变化是：如果服务器编码是 SQL_ASCII，你将得到该字符串的 UTF-8 表示；此前字符串中的任何非 ASCII 字符都会导致错误。
- 修复 PL/Python 中 PostgreSQL 编码到 Python 编码的映射（Jan Urbanski）
- 修复 `contrib/xml2` 的 `xslt_process()` 中的错误报告（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2012e，以反映摩洛哥和托克劳的夏令时法规变更

E.41. 版本 9.0.8

发行日期
2012-06-04

本次发布包含对 9.0.7 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.41.1. 迁移到版本 9.0.8

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.41.2. 变更

- 修复 `contrib/pgcrypto` 的 DES `crypt()` 函数中不正确的口令变换（Solar Designer）

  如果口令串包含字节值 `0x80`，口令的其余部分会被忽略，使口令比看起来弱得多。修复后，字符串的其余部分会被正确纳入 DES 哈希。因此，受此缺陷影响的任何已存储口令值将不再匹配，所以可能需要更新已存储的值。（CVE-2012-2143）

- 忽略过程语言调用处理程序的 `SECURITY DEFINER` 和 `SET` 属性（Tom Lane）

  把这类属性应用于调用处理程序可能导致服务器崩溃。（CVE-2012-2655）

- 允许 `timestamp` 输入中的数字时区偏移距 UTC 最多 16 小时（Tom Lane）

  一些历史时区的偏移超过此前 15 小时的限制。这可能导致转储的数据值在重载期间被拒绝。

- 修复时间戳转换，
  使其在给定时间恰好是当前时区最后一次夏令时转换时间时能正确处理（Tom Lane）

  这一疏忽已存在很长时间，但此前未被注意到，
  因为大多数使用夏令时的时区被假定有无限的未来夏令时转换序列。

- 修复 `text` 到 `name` 以及 `char` 到 `name` 的转换，使其在多字节编码下正确执行字符串截断（Karl Schnaitter）

- 修复 `to_tsquery()` 中的内存复制缺陷（Heikki Linnakangas）

- 确保 `txid_current()` 在热备中执行时报告正确的纪元（epoch）（Simon Riggs）

- 修复规划器对子查询内外层 PlaceHolderVar 的处理（Tom Lane）

  该缺陷涉及引用来自外层查询外连接可空侧变量的子 SELECT。在 9.1 中，受此缺陷影响的查询会报 “ERROR: Upper-level PlaceHolderVar found where not expected” 失败。但在 9.0 和 8.4 中，你会静默地得到可能错误的答案，因为传入子查询的值不会在该为空时变为空。

- 修复 `pg_attribute` 非常大会话启动缓慢的问题（Tom Lane）

  如果 `pg_attribute` 超过 `shared_buffers` 的四分之一，会话启动期间有时需要的缓存重建代码就会触发同步扫描逻辑，使其耗时比正常情况多数倍。如果许多新会话同时启动，
  问题尤为严重。

- 确保顺序扫描以合理的频率检查查询取消（Merlin Moncure）

  遇到许多不含存活元组的连续页面的扫描在此期间不会响应中断。

- 确保 `PGSemaphoreLock()` 的 Windows 实现在返回前清除 `ImmediateInterruptOK`（Tom Lane）

  这一疏忽意味着同一查询稍后收到的查询取消中断可能在不安全的时刻被接受，
  后果不可预测且不好。

- 在打印视图或规则时安全地显示整行变量（Abbas Butt、Tom Lane）

  涉及有歧义名称（即该名称既可能是查询的表名也可能是列名）
  的极端情况此前以有歧义的方式打印，存在视图或规则在转储和重载后被不同解释的风险。
  现通过附加无操作转换来避免歧义情况。

- 修复 `COPY FROM`，使其正确处理对应无效编码的空值标记串（Tom Lane）

  诸如 `E'\\0'` 这样的空值标记串本应有效，过去也确实有效，但该情况在 8.4 中被破坏。

- 确保自动清理工作进程正确执行栈深度检查（Heikki Linnakangas）

  此前，由自动 `ANALYZE` 调用的函数中的无限递归可能导致工作进程崩溃。

- 修复日志收集器，使其在高负载下不丢失日志连贯性（Andrew Dunstan）

此前，如果收集器过于忙碌，可能无法重新组装大消息。

- 修复日志收集器，确保它在收到 SIGHUP 后会重新开始文件轮转（Tom Lane）
- 修复 GIN 索引的 WAL 重放逻辑，使其在索引随后被删除时不失败（Tom Lane）
- 修复 PL/pgSQL 的 `RETURN NEXT` 命令中的内存泄漏（Joe Conway）
- 修复 PL/pgSQL 的 `GET DIAGNOSTICS` 命令在目标是函数第一个变量时的问题（Tom Lane）
- 修复 psql 扩展显示（`\x`）模式中可能越过内存末尾的访问（Peter Eisentraut）
- 修复数据库包含大量对象时 pg_dump 的若干性能问题（Jeff Janes、Tom Lane）

  如果数据库包含许多模式、许多对象处于依赖循环中、或有许多被拥有的序列，pg_dump 可能变得非常慢。
- 修复 pg_upgrade 处理存储在非默认表空间的数据库包含集簇默认表空间中的表的情况（Bruce Momjian）
- 在 ecpg 中，修复罕见的内存泄漏以及 `sqlca_t` 结构之后一个字节可能被覆写的问题（Peter Eisentraut）
- 修复 `contrib/dblink` 的 `dblink_exec()`，使其在出错时不泄漏临时数据库连接（Tom Lane）
- 修复 `contrib/dblink`，使其在错误消息中报告正确的连接名（Kyotaro Horiguchi）
- 修复 `contrib/vacuumlo`，使其在删除许多大对象时使用多个事务（Tim Lewis、Robert Haas、Tom Lane）

  此变更避免了需要删除许多对象时超出 `max_locks_per_transaction`。该行为可以通过新的 `-l`（limit）选项调整。
- 把时区数据文件更新为 tzdata 版本 2012c for DST law changes in Antarctica, Armenia, Chile, Cuba, Falkland Islands, Gaza, Haiti, Hebron, Morocco, Syria, and Tokelau Islands; also historical corrections for Canada。

E.42. 版本 9.0.7

发行日期
2012-02-27

本次发布包含对 9.0.6 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.42.1. 迁移到版本 9.0.7

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.6 的版本升级，请参见第 E.43 节。

E.42.2. 变更

- 让 `CREATE TRIGGER` 要求对触发器函数有执行权限（Robert Haas）

  这一缺失的检查可能允许另一用户通过把触发器函数安装到自己拥有的表上，以伪造的输入数据执行该触发器函数。这只对标记为 `SECURITY DEFINER` 的触发器函数有影响，因为其他情况下触发器函数本来就以表所有者身份运行。（CVE-2012-0866）

- 移除 SSL 证书中通用名长度的任意限制（Heikki Linnakangas）

  libpq 和服务器此前都会把从 SSL 证书提取的通用名截断到 32 字节。通常这最坏只会造成意外的验证失败，但在一些相当难以置信的场景下，它可能允许一个证书持有者冒充另一个。受害者必须恰好有 32 字节长的通用名，而攻击者必须说服受信任的 CA 颁发一个通用名以该字符串为前缀的证书。冒充服务器还需要额外的攻击手段来重定向客户端连接。（CVE-2012-0867）

- 把 pg_dump 注释中写入的名称里的换行符转换为空格（Robert Haas）

  pg_dump 此前对其输出脚本中 SQL 注释内写出的对象名清理不慎。包含换行符的名称至少会使脚本语法错误。恶意构造的对象名可能在脚本重载时构成 SQL 注入风险。（CVE-2012-0868）

- 修复插入与清理并发导致 btree 索引损坏的问题（Tom Lane）

  插入导致的索引页分裂有时会使并发运行的 `VACUUM` 漏删本应删除的索引项。在相应的表行被删除后，悬空的索引项会导致错误（例如 “could not read block N in file ...” ），或者更糟：当不相关的行被重新插入到现已空闲的表位置后，查询会静默地得到错误结果。该缺陷自 8.2 起就存在，但发生频率极低，直到现在才被诊断出来。如果你有理由怀疑它曾在你的数据库中发生，对受影响的索引重建索引即可修复。

- 修复 WAL 重放期间共享缓冲区的瞬时清零（Tom Lane）

  重放逻辑有时会把一个共享缓冲区清零再重新填充，使其内容瞬时无效。在热备模式下，这可能导致并行执行的查询看到垃圾数据。这可能引发各种症状，但最常见的好像是 “invalid memory alloc request size” 。

- 修复 postmaster，使其在热备崩溃后尝试重启（Tom Lane）

  一个逻辑错误导致：如果有后端进程在热备模式下运行时崩溃，postmaster 会终止而不是尝试重启集簇。

- 修复 `CLUSTER`/`VACUUM FULL` 对最近更新的行所拥有的 toast 值的处理（Tom Lane）

  这一疏忽可能导致在这些命令之一执行期间，toast 表的索引报 “duplicate key value violates unique constraint” 错误。

- 更改表所有者时，不仅更新表级权限，也更新逐列权限（Tom Lane）

  未做到这一点意味着，此前授予的所有列权限仍显示为由旧所有者授予。这导致新所有者和超级用户都无法撤销这些现在无法追溯到表所有者的权限。

- 在 `REASSIGN OWNED` 中支持外部数据包装器和外部服务器（Alvaro Herrera）

  如果该命令需要更改任何这类对象的所有权，会报 “unexpected classid” 错误。

- 允许 `ALTER USER/DATABASE SET` 中某些设置取不存在的值（Heikki Linnakangas）

允许把 default_text_search_config、default_tablespace 和 temp_tablespaces 设为未知的名称。这是因为设置要在其中使用的另一个数据库中它们可能是已知的；对表空间的情况，则是因为表空间可能尚未创建。此前 search_path 已被认识到有同样的问题，这些设置现在的行为与它一致。

- 避免在提交后删除表文件出现问题时崩溃（Tom Lane）

  删除表应只在事务提交后才删除底层磁盘文件。如果那时失败（例如因为文件权限错误），代码本应只发出警告消息并继续，因为此时中止事务已为时过晚。这一逻辑从 8.4 起被破坏，导致这类情况引发 PANIC 和无法重启的数据库。

- 从 DROP TABLESPACE 的 WAL 重放期间的错误中恢复（Tom Lane）

  重放会尝试删除表空间的目录，
  但这可能因各种原因失败（例如这些目录的所有权或权限不正确）。此前重放代码会 panic，使数据库在没有人工干预的情况下无法重启。记录问题并继续似乎更好，
  因为删除目录失败唯一后果只是浪费一些磁盘空间。

- 修复为热备记录 AccessExclusiveLock 时的竞态条件（Simon Riggs）

  有时锁会被记录为由“事务零”持有。至少已知这会在从库服务器上产生断言失败，
  还可能是更严重问题的原因。

- 在 WAL 重放期间正确跟踪 OID 计数器，即使它发生回卷（Tom Lane）

  此前 OID 计数器会一直停留在高值，直到系统退出重放模式。其实际后果通常为零，但在某些场景下，已提升为主库的备库服务器在需要 OID 值时，可能要很长时间才能把 OID 计数器推进到合理的值。

- 防止在崩溃恢复开始时发出误导性的“consistent recovery state reached”日志消息（Heikki Linnakangas）

- 修复 pg_stat_replication.replay_location 的初始值（Fujii Masao）

  此前，在至少重放一条 WAL 记录之前，显示的值是错误的。

- 修复带 * 的正则表达式反向引用（Tom Lane）

  代码实际上并不强制精确字符串匹配，
  而是接受任何满足反向引用符号所引用的模式子表达式的字符串。

  类似问题仍困扰着嵌入在更大带量词表达式中（而非量词直接作用对象）的反向引用。
  这将在未来的 PostgreSQL 版本中解决。

- 修复最近引入的 inet/cidr 值处理中的内存泄漏（Heikki Linnakangas）

  PostgreSQL 2011 年 12 月版本中的一个补丁导致这些操作中的内存泄漏，在此类列上构建 btree 索引等场景中可能相当显著。

- 修复 SQL 语言函数中 CREATE TABLE AS/SELECT INTO 之后的悬空指针（Tom Lane）

  大多数情况下这只会在启用断言的构建中导致断言失败，但似乎可能有更糟的后果。

- 避免 Windows 上 syslogger 中文件句柄的双重关闭（MauMau）

  通常这个错误不可见，但在调试版 Windows 上运行时会导致异常。

- 修复 plpgsql 中与 I/O 转换相关的内存泄漏（Andres Freund、Jan Urbanski、Tom Lane）

  某些操作会一直泄漏内存直到当前函数结束。

- 改进 pg_dump 对继承表列的处理（Tom Lane）

  pg_dump 错误处理了子列与父列默认表达式不同的情况。如果默认值在文本上与父列相同但实际不同（例如因为模式搜索路径差异），它不会被识别为不同，
  因此转储和恢复后子列会被允许继承父列的默认值。子列为 `NOT NULL` 而父列不是的情况也可能被微妙地错误恢复。

- 修复 pg_restore 直接写数据库模式对 INSERT 式表数据的处理（Tom Lane）

  当使用 2011 年 9 月或 12 月版本的 pg_restore 时，对以 `--inserts` 或 `--column-inserts` 选项制作的归档文件直接恢复到数据库会失败，这是对另一个问题的修复中的一个疏忽所致。归档文件本身没有问题，文本模式输出也正常。

- 允许 pg_upgrade 处理包含 `regclass` 列的表（Bruce Momjian）

  由于 pg_upgrade 现在注意保留 `pg_class` OID，这一限制不再有任何理由存在。

- 让 libpq 在查找 SSL 客户端证书文件时忽略 `ENOTDIR` 错误（Magnus Hagander）

  即使用户的主目录被设置为 `/dev/null` 之类，这也能建立 SSL 连接（虽然没有证书）。

- 修复 ecpg 的 SQLDA 区域中更多的字段对齐问题（Zoltan Boszormenyi）

- 允许 ecpg 的 `DEALLOCATE` 语句中使用 `AT` 选项（Michael Meskes）

  支持它的基础设施已存在一段时间，但由于疏忽仍有一个错误检查拒绝这种情况。

- 在 ecpg 中定义 varchar 结构时不使用变量名（Michael Meskes）

- 修复 `contrib/auto_explain` 的 JSON 输出模式，使其产生有效的 JSON（Andrew Dunstan）

  输出在顶层使用了方括号，而本应使用花括号。

- 修复 `contrib/intarray` 的 `int[] & int[]` 操作符中的错误（Guillaume Lelarge）

  如果两个输入数组共有的最小整数是 1，且任一数组中有更小的值，那么 1 会被错误地从结果中遗漏。

- 修复 `contrib/pgcrypto` 的 `encrypt_iv()` 和 `decrypt_iv()` 中的错误检测（Marko Kreen）

  这些函数未能报告某些类型的无效输入错误，而是对错误输入返回随机垃圾值。

- 修复 `contrib/test_parser` 中的一字节缓冲区越界（Paul Guyot）

  代码会尝试多读一个字节，这在极端情况下会崩溃。由于 `contrib/test_parser` 只是示例代码，这本身不是安全问题，但坏的示例代码仍然是坏的。

- 在 ARM 上如果可用则使用 `__sync_lock_test_and_set()` 实现自旋锁（Martin Pitt）

  该函数取代了我们此前对 `SWPB` 指令的使用，该指令已被弃用且在 ARMv6 及更高版本上不可用。报告表明旧代码在较新的 ARM 板上不会以明显方式失败，而只是不对并发访问互锁，导致多进程操作中的离奇失败。

- 在使用接受该选项的 gcc 版本构建时使用 `-fexcess-precision=standard` 选项（Andrew Dunstan）

  这防止了较新版本 gcc 产生"创造性"结果的各种场景。

- 允许在 FreeBSD 上使用线程化 Python（Chris Rees）

  我们的 configure 脚本此前认为这种组合无法工作；但 FreeBSD 修复了问题，因此移除该错误检查。

E.43. 版本 9.0.6

发行日期
2011-12-05

本次发布包含对 9.0.5 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.43.1. 迁移到版本 9.0.6

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，`information_schema.referential_constraints` 视图的定义中被发现存在一个长期错误。如果你依赖该视图的正确结果，应按下面第一条变更日志条目的说明替换其定义。

不过，如果你是从早于 9.0.4 的版本升级，请参见第 E.45 节。

E.43.2. 变更

- 修复 `information_schema.referential_constraints` 视图中的缺陷（Tom Lane）

  该视图在把外键约束与它依赖的主键或唯一键约束匹配时不够谨慎。
  这可能导致某个外键约束完全不被显示、被显示多次，
  或者声称它依赖于一个并非其实际依赖的约束。

  由于该视图定义是由 initdb 安装的，仅靠升级不会修复问题。如果需要修复现有安装，可以（以超级用户身份）删除 `information_schema` 模式，然后通过载入 *SHAREDIR*`/information_schema.sql` 重新创建它。（如果不确定 *SHAREDIR* 在哪里，运行 `pg_config --sharedir`。）这必须在每个要修复的数据库中重复执行。

- 修复连接到返回标量的函数输出的 `UPDATE` 或 `DELETE` 期间可能崩溃的问题（Tom Lane）

  只有当目标行被并发更新时才可能崩溃，因此该问题只是间歇性出现。

- 修复 GIN 索引更新 WAL 记录的错误重放（Tom Lane）

  这可能导致崩溃后或热备服务器上瞬时找不到索引项。不过，该问题会在该索引下一次 `VACUUM` 时被修复。

- 修复 `CREATE TABLE dest AS SELECT * FROM src` 或 `INSERT INTO dest SELECT * FROM src` 期间与 TOAST 相关的数据损坏（Tom Lane）

如果表曾被 `ALTER TABLE ADD COLUMN` 修改过，在某些极端情况下，尝试把它的数据原样复制到另一个表可能产生损坏的结果。该问题只会在 8.4 及更高版本中以这种确切形式出现，但我们也给更早的版本打了补丁，以防有其他代码路径可能触发同一缺陷。

- 修复热备启动期间可能的失败（Simon Riggs）
- 当初始快照不完整时更快地启动热备（Simon Riggs）
- 修复从过期 syscache 条目访问 toast 表时的竞态条件（Tom Lane）

  典型症状是诸如 “missing chunk number 0 for toast value NNNNN in pg_toast_2619” 的瞬时错误，其中引用的 toast 表总是属于某个系统目录。

- 跟踪函数对参数默认值表达式中所用项的依赖（Tom Lane）

  此前，被引用的对象可以在未删除或修改函数的情况下被删除，导致函数被使用时行为异常。注意，仅仅安装此更新不会修复缺失的依赖条目；要做到这一点，之后需要对每个这样的函数执行 `CREATE OR REPLACE`。如果你的函数默认值依赖于非内建对象，建议这样做。

- 允许内联带多个 OUT 参数的集合返回 SQL 函数（Tom Lane）
- 不要为连接消除信任延迟唯一索引（Tom Lane、Marti Raudsepp）

  延迟唯一约束在事务内可能不成立，因此假设它成立可能给出错误的查询结果。

- 让 `DatumGetInetP()` 解包带 1 字节头的 inet datum，并新增一个不做此事的宏 `DatumGetInetPP()`（Heikki Linnakangas）

  此变更不影响任何核心代码，但可能避免期望 `DatumGetInetP()` 按常规约定产生已解包 datum 的附加代码中的崩溃。

- 改进 `money` 类型输入和输出的区域支持（Tom Lane）

  除了不支持所有标准 `lc_monetary` 格式化选项外，输入和输出函数还不一致，这意味着在某些区域设置下转储的 `money` 值无法重新读取。

- 不让 `transform_null_equals` 影响 `CASE foo WHEN NULL ...` 构造（Heikki Linnakangas）

  `transform_null_equals` 只应影响用户直接写的 `foo = NULL` 表达式，而不是这种形式的 `CASE` 在内部生成的相等性检查。

- 更改外键触发器的创建顺序，以更好地支持自引用外键（Tom Lane）

  对于引用自身表的级联外键，行更新会作为一个事件同时触发 `ON UPDATE` 触发器和 `CHECK` 触发器。`ON UPDATE` 触发器必须先执行，否则 `CHECK` 会检查行的非最终状态，并可能抛出不恰当的错误。然而这些触发器的触发顺序由其名称决定，而由于触发器使用遵循 “RI_ConstraintTrigger_NNNN” 约定的自动生成名称，它们通常按创建顺序排序。彻底的修复需要修改该约定（我们将在 9.2 中做），但在现有版本中更改它似乎有风险。因此此补丁只更改触发器的创建顺序。遇到这类错误的用户应删除并重新创建外键约束，使其触发器处于正确顺序。

- 避免在跟踪缓冲区分配率时发生浮点下溢（Greg Matthews）

  虽然这本身无害，但在某些平台上会导致烦人的内核日志消息。

- 在 Windows 下启动子进程时保留配置文件名和行号值（Tom Lane）

  此前它们不会在 `pg_settings` 视图中正确显示。

- 修复 ecpg 的 SQLDA 区域中不正确的字段对齐（Zoltan Boszormenyi）
- 在 psql 的命令历史中保留命令内的空行（Robert Haas）

  例如，如果从字符串字面量中删除了空行，先前的行为可能导致问题。

- 修复 pg_dump，使其转储自动生成类型（如表行类型）之间的用户定义转换（Tom Lane）
- 对 pg_upgrade 的若干修复（Bruce Momjian）

  正确处理排他约束、避免 Windows 上的失败、不再对 8.4 数据库中不匹配的 toast 表名报错。

- 使用首选版本的 xsubpp 构建 PL/Perl，而不一定是操作系统的主副本（David Wheeler、Alex Hunsaker）
- 修复 `contrib/dict_int` 和 `contrib/dict_xsyn` 中的错误编码（Tom Lane）

  一些函数错误地假定 `palloc()` 返回的内存保证被清零。

- 修复 `contrib/unaccent` 配置文件解析中的若干错误（Tom Lane）
- 让 `pgstatindex()` 及时响应查询取消中断（Robert Haas）
- 修复 Mac OS X 启动脚本中日志文件名的不正确引号（Sidar Lopez）
- 确保 VPATH 构建正确安装所有服务器头文件（Peter Eisentraut）
- 缩短详细错误消息中报告的文件名（Peter Eisentraut）

  常规构建一直只报告包含错误消息调用的 C 文件的名称，但 VPATH 构建此前报告绝对路径名。

- 修复中美洲 Windows 时区名的解释（Tom Lane）

  把 “Central America Standard Time” 映射为 `CST6` 而不是 `CST6CDT`，因为中美洲各地通常不实行夏令时。

- 把时区数据文件更新为 tzdata 版本 2011n for DST law changes in Brazil, Cuba, Fiji, Palestine, Russia, and Samoa; also historical corrections for Alaska and British East Africa。

E.44. 版本 9.0.5

发行日期
2011-09-26

本次发布包含对 9.0.4 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.44.1. 迁移到版本 9.0.5

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你是从早于 9.0.4 的版本升级，请参见第 E.45 节。

E.44.2. 变更

- 修复对系统目录执行 `VACUUM FULL` 或 `CLUSTER` 之后的目录缓存失效（Tom Lane）

  某些情况下，系统目录行被重定位到别处不会被并发服务器进程识别，
  如果它们随后尝试更新该行，就可能发生目录损坏。最坏的结果可能是一张表完全丢失。

- 修复 sinval 重置处理期间错误的操作顺序，并确保 TOAST OID 在系统目录中被保留（Tom Lane）

  这些失误可能导致对系统目录执行 `VACUUM FULL` 或 `CLUSTER` 之后出现瞬时失败。

- 修复对不确定的 HOT 更新元组建立索引的缺陷（Tom Lane）

  这些缺陷可能导致重建系统目录索引之后索引损坏。据信它们不影响用户索引。

- 修复 GiST 索引页面分裂处理中的多个缺陷（Heikki Linnakangas）

  发生概率很低，但它们可能导致索引损坏。

- 修复 `tsvector_concat()` 中可能的缓冲区越界（Tom Lane）

  该函数可能低估其结果所需的内存量，导致服务器崩溃。

- 修复 `xml_recv` 处理 “standalone” 参数时的崩溃（Tom Lane）

- 让 `pg_options_to_table` 对没有值的选项返回 NULL（Tom Lane）

  此前这类情况会导致服务器崩溃。

- 避免在 `ANALYZE` 和 SJIS-2004 编码转换中可能越过内存末尾的访问（Noah Misch）

  这修复了一些概率极低的服务器崩溃场景。

- 保护 `pg_stat_reset_shared()` 免受 NULL 输入（Magnus Hagander）

- 修复在子事务内检测到恢复冲突死锁时可能的失败（Tom Lane）

- 避免热备期间回收 btree 索引页时的虚假冲突（Noah Misch、Simon Riggs）

- 如果恢复结束时 WAL 接收端仍在运行则将其关闭（Heikki Linnakangas）

  postmaster 此前在这种情况下会 panic，但这实际上是合法的情况。

- 修复 relcache init 文件失效中的竞态条件（Tom Lane）

  存在一个时间窗口，新的后端进程可能读到过期的 init 文件，却错过本应告知它数据已过期的失效消息。结果是目录访问中的离奇失败，通常在启动稍后阶段报 “could not read block 0 in file ...” 。

- 修复 GiST 索引扫描结束时的内存泄漏（Tom Lane）

  执行许多独立 GiST 索引扫描的命令（例如对已包含许多行的表验证新的基于 GiST 的排他约束）会因此泄漏瞬时需要大量内存。

- 修复当必须对传入的命令串做编码转换且 `LISTEN` 处于活动状态时的内存泄漏（Tom Lane）

- 修复支持可持有游标和 plpgsql 的 `RETURN NEXT` 命令的 tuplestore 中不正确的内存核算（可能导致内存膨胀）（Tom Lane）
- 修复同时存在 `BEFORE` 和 `AFTER` 触发器时的触发器 `WHEN` 条件（Tom Lane）

  如果同一更新曾触发过 `BEFORE ROW` 触发器，`AFTER ROW UPDATE` 触发器的 `WHEN` 条件求值可能崩溃。
- 修复构造大型有损位图时的性能问题（Tom Lane）
- 修复唯一列的连接选择率估计（Tom Lane）

  这修复了一个可能错误估计连接结果大小的规划器启发式规则。
- 修复只出现在子 SELECT 目标列表中的嵌套 PlaceHolderVar 表达式（Tom Lane）

  这一失误可能导致外连接的输出错误地显示为 NULL。
- 允许规划器假定空的父表真的是空的（Tom Lane）

  通常为了规划目的，空表被假定具有某个最小大小；但对继承层次的父表而言，这一启发式规则似乎弊大于利，因为父表往往永久为空。
- 允许嵌套的 `EXISTS` 查询被正确优化（Tom Lane）
- 修复数组和路径创建函数，确保填充字节为零（Tom Lane）

  这避免了一些规划器认为语义相等的常量不相等的情况，这种情况会导致糟糕的优化。
- 修复 `EXPLAIN`，使其处理内层索引扫描子计划中的门控 Result 节点（Tom Lane）

  这一疏忽的常见症状是“bogus varno”错误。
- 修复 btree 对 *`indexedcol`* `IS NULL` 条件的预处理（Dean Rasheed）

  如果与同一索引列上任何其他类型的 btree 可索引条件组合，该条件是不可满足的。这种情况在 9.0.0 及更高版本中处理不正确，导致本应没有输出的查询产生输出。
- 规避破坏 WAL 重放的 gcc 4.6.0 缺陷（Tom Lane）

  这可能导致服务器崩溃后已提交事务的丢失。
- 修复视图中 `VALUES` 的转储缺陷（Tom Lane）
- 禁止对序列执行 `SELECT FOR UPDATE/SHARE`（Tom Lane）

  这种操作不像预期那样工作，可能导致失败。
- 修复 `VACUUM`，使其总是更新 `pg_class.reltuples/relpages`（Tom Lane）

  这修复了一些自动清理可能对何时清理表做出越来越差决策的场景。
- 防御计算哈希表大小时的整数溢出（Tom Lane）
- 修复 `CLUSTER` 可能尝试访问已移除的 TOAST 数据的情况（Tom Lane）
- 修复初始认证事务期间的过早超时失败（Tom Lane）

- 修复“peer”认证使用凭据控制消息时的可移植性缺陷（Tom Lane）
- 修复需要多次往返时的 SSPI 登录（Ahmed Shinwari、Magnus Hagander）

  该问题的典型症状是 SSPI 登录期间的“The function requested is not supported”错误。
- 修复向 `postgresql.conf` 添加自定义变量类的新变量时的失败（Tom Lane）
- 如果 `pg_hba.conf` 包含 `hostssl` 但 SSL 被禁用则抛出错误（Tom Lane）

  这被认定比此前静默忽略这类行的行为更用户友好。
- 修复 `DROP OWNED BY` 试图移除序列上的默认权限时的失败（Shigeru Hanada）
- 修复 `pg_srand48` 种子初始化中的笔误（Andres Freund）

  这导致未能使用所提供种子的所有位。该函数在大多数平台上不使用（只在没有 `srandom` 的平台上使用），而且无论如何看待，随机性低于预期的种子带来的潜在安全暴露似乎都很小。
- 避免 `LIMIT` 与 `OFFSET` 值之和超过 2^63 时的整数溢出（Heikki Linnakangas）
- 为 `generate_series()` 的 `int4` 和 `int8` 版本添加溢出检查（Robert Haas）
- 修复 `to_char()` 中移除尾随零的问题（Marti Raudsepp）

  在使用 `FM` 且小数点后没有数字位的格式中，小数点左侧的零可能被错误移除。
- 修复 `pg_size_pretty()`，避免接近 2^63 的输入溢出（Tom Lane）
- 放宽 plpgsql 对记录值中 typmod 匹配的检查（Tom Lane）

  过于严格的检查可能导致丢弃本应保留的长度修饰符。
- 在 initdb 期间正确处理区域名中的引号（Heikki Linnakangas）

  这种情况可能出现在某些 Windows 区域设置下，例如“People's Republic of China”。
- 在 pg_upgrade 中，避免转储孤儿临时表（Bruce Momjian）

  这防止了表 OID 分配在新旧安装之间失去同步的情况。
- 修复 pg_upgrade，使其在从 8.3 升级期间保留 toast 表的 relfrozenxid（Bruce Momjian）

  不这样做可能导致升级后 `pg_clog` 文件被过早删除。
- 修复 pg_upgrade 的 `-l`（日志）选项在 Windows 上工作的问题（Bruce Momjian）
- 在 pg_ctl 中，支持 Windows 上服务注册的静默模式（MauMau）
- 修复 psql 在从其他文件执行 `COPY` 期间脚本文件行号计数的问题（Tom Lane）
- 修复 pg_restore 直接写数据库模式对 `standard_conforming_strings` 的处理（Tom Lane）

  当从以 `standard_conforming_strings` 设为 `on` 制作的归档文件直接恢复到数据库服务器时，pg_restore 可能发出不正确的命令。
- 对并行 pg_restore 不支持的情况更加用户友好（Tom Lane）

此变更确保这类情况在任何恢复动作发生之前就被检测并报告。

- 修复 libpq 的 LDAP 服务查找代码中的越界写入和内存泄漏（Albe Laurenz）
- 在 libpq 中，避免使用非阻塞 I/O 和 SSL 连接时的失败（Martin Pihlak、Tom Lane）
- 改进 libpq 对连接启动期间失败的处理（Tom Lane）

  特别是，对 SSL 连接启动期间服务器报告 `fork()` 失败的响应现在更合理。
- 改进 libpq 对 SSL 失败的错误报告（Tom Lane）
- 修复 `PQsetvalue()`，避免向最初来自服务器查询的 `PGresult` 添加新元组时可能崩溃（Andrew Chernow）
- 让 ecpglib 以 15 位精度写 `double` 值（Akira Kurosawa）
- 在 ecpglib 中，确保出错后恢复 `LC_NUMERIC` 设置（Michael Meskes）
- 应用上游对 blowfish 有符号字符缺陷的修复（CVE-2011-2483）（Tom Lane）

  `contrib/pg_crypto` 的 blowfish 加密代码在 char 为有符号的平台（大多数平台如此）上可能给出错误结果，导致加密口令比应有的强度更弱。
- 修复 `contrib/seg` 中的内存泄漏（Heikki Linnakangas）
- 修复 `pgstatindex()`，使其对空索引给出一致结果（Tom Lane）
- 允许使用 perl 5.14 构建（Alex Hunsaker）
- 修复构建和安装文件路径包含空格的各种问题（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2011i for DST law changes in Canada, Egypt, Russia, Samoa, and South Sudan。

E.45. 版本 9.0.4

发行日期
2011-04-18

本次发布包含对 9.0.3 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.45.1. 迁移到版本 9.0.4

对于运行 9.0.X 的用户，不需要进行转储/恢复。

不过，如果你的安装是通过运行 pg_upgrade 从前一个大版本升级的，应采取措施防止因 pg_upgrade 一个现已修复的缺陷导致可能的数据丢失。推荐的解决方案是对所有 TOAST 表执行 `VACUUM FREEZE`。更多信息见 http://wiki.postgresql.org/wiki/20110408pg_upgrade_fix[3]。

[3] http://wiki.postgresql.org/wiki/20110408pg_upgrade_fix

E.45.2. 变更

- 修复 pg_upgrade 对 TOAST 表的处理（Bruce Momjian）

  在 pg_upgrade 期间，TOAST 表的 `pg_class.relfrozenxid` 值未被正确复制到新安装。这之后可能导致 `pg_clog` 文件在仍被需要用于验证 TOAST 表中的元组时被丢弃，导致 “could not access status of transaction” 失败。

  这一错误对曾用 pg_upgrade 升级的安装构成显著的数据丢失风险。此补丁纠正了未来使用 pg_upgrade 时的问题，但本身不能治愈已用带缺陷版本的 pg_upgrade 处理过的安装中的问题。

- 抑制错误的 “PD_ALL_VISIBLE flag was incorrectly set” 警告（Heikki Linnakangas）

  `VACUUM` 有时会在实际有效的情况下发出此警告。

- 为热备冲突情况使用更好的 SQLSTATE 错误码（Tatsuo Ishii、Simon Riggs）

  所有可重试的冲突错误现在都具有指示可以重试的错误码。此外，因主库上数据库被删除而关闭会话现在报告为 `ERRCODE_DATABASE_DROPPED` 而不是 `ERRCODE_ADMIN_SHUTDOWN`，使连接池能够正确处理这种情况。

- 防止启动进程与 bgwriter 进程交互时的间歇性挂起（Simon Riggs）

  这影响非热备情况下的恢复。

- 禁止在组合类型中包含其自身（Tom Lane）

  这防止了服务器在处理该组合类型时无限递归的场景。虽然这种结构有一些可能的用途，但它们似乎不足以抵消确保其始终安全工作所需的努力。

- 避免目录缓存初始化期间的潜在死锁（Nikhil Sontakke）

  某些情况下，缓存加载代码会在锁定索引的目录之前获取系统索引上的共享锁。这可能与尝试以另一种更标准的顺序获取排他锁的进程死锁。

- 修复当目标元组存在并发更新时 `BEFORE ROW UPDATE` 触发器处理中的悬空指针问题（Tom Lane）

  已观察到该缺陷在尝试执行 `UPDATE RETURNING ctid` 时导致间歇性的 “cannot extract system attribute from virtual tuple” 失败。发生更严重错误（例如为更新后的元组生成错误索引项）的概率非常小。

- 当表存在未决的延迟触发器事件时禁止 `DROP TABLE`（Tom Lane）

  此前 `DROP` 会照常执行，导致触发器最终触发时报 “could not open relation with OID nnn” 错误。

- 允许 “replication” 作为 `pg_hba.conf` 中的用户名（Andrew Dunstan）

  “replication” 在数据库名列中是特殊的，但它被误认为在用户名列中也是特殊的。

- 防止 GEQO 优化期间由恒假 WHERE 条件触发的崩溃（Tom Lane）

- 改进规划器对半连接和反连接情况的处理（Tom Lane）

- 修复子 SELECT 中 `SELECT FOR UPDATE` 的处理（Tom Lane）

该缺陷通常导致 “cannot extract system attribute from virtual tuple” 错误。

- 修复文本搜索的选择率估计，使其考虑 NULL（Jesper Krogh）
- 修复 get_actual_variable_range()，使其支持由索引顾问插件注入的假设索引（Gurjeet Singh）
- 修复涉及数组切片的 PL/Python 内存泄漏（Daniel Popowich）
- 允许 libpq 的 SSL 初始化在用户主目录不可用时成功（Tom Lane）

  如果 SSL 模式不要求根证书文件，就没有必要失败。此变更将行为恢复为 9.0 之前的样子。
- 修复 libpq，使其对 `conninfo_array_parse` 中检测到的错误返回有用的错误消息（Joseph Adams）

  一个笔误导致库向应用返回 NULL 而不是包含错误消息的 `PGconn` 结构。
- 修复 ecpg 预处理器对 float 常量的处理（Heikki Linnakangas）
- 修复并行 pg_restore，使其正确处理 POST_DATA 项上的注释（Arnd Hannemann）
- 修复 pg_restore，使其能处理 TOC 文件中的长行（超过 1KB）（Tom Lane）
- 加入更多防护，防止因过于激进的编译器优化导致的除零崩溃（Aurelien Jarno）
- 支持在 FreeBSD 和 OpenBSD 的 MIPS 上使用 dlopen()（Tom Lane）

  此前有一个硬编码的假设：该系统函数在这些系统的 MIPS 硬件上不可用。现在改用编译期测试，因为较新的版本已经有了它。
- 修复 HP-UX 上的编译失败（Heikki Linnakangas）
- 避免在进程启动极早期尝试写入 Windows 控制台时崩溃（Rushabh Lathia）
- 支持使用 MinGW 64 位编译器为 Windows 构建（Andrew Dunstan）
- 修复 Windows 上 libintl 的版本不兼容问题（Hiroshi Inoue）
- 修复 Windows 构建脚本中 xcopy 的用法，使其在 Windows 7 下正确工作（Andrew Dunstan）

  这只影响构建脚本，不影响安装或使用。
- 修复 pg_regress 在 Cygwin 上使用的路径分隔符（Andrew Dunstan）
- 把时区数据文件更新为 tzdata 版本 2011f for DST law changes in Chile, Cuba, Falkland Islands, Morocco, Samoa, and Turkey; also historical corrections for South Australia, Alaska, and Hawaii。

E.46. 版本 9.0.3

发行日期
2011-01-31

本次发布包含对 9.0.2 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.46.1. 迁移到版本 9.0.3

对于运行 9.0.X 的用户，不需要进行转储/恢复。

E.46.2. 变更

- 在退出 walreceiver 之前，确保所有已接收的 WAL 都已 fsync 到磁盘（Heikki Linnakangas）

  否则备库服务器可能重放一些未同步的 WAL，如果系统恰好在那时崩溃，可想而知会导致数据损坏。

- 避免 walreceiver 中过多的 fsync 活动（Heikki Linnakangas）

- 让 `ALTER TABLE` 在需要时重新验证唯一性和排他约束（Noah Misch）

  这在 9.0 中被一个本意是抑制 `VACUUM FULL` 和 `CLUSTER` 期间重新验证的变更破坏，但该变更无意中也影响了 `ALTER TABLE`。

- 修复对表不完全相同的继承树执行 `UPDATE` 时 EvalPlanQual 的问题（Tom Lane）

  表行类型的任何变化（包括只存在于部分子表中的已删除列）都会混淆 EvalPlanQual 代码，导致行为异常甚至崩溃。由于 EvalPlanQual 只在同一行的并发更新期间执行，该问题只是间歇性出现。

- 避免 `EXPLAIN` 尝试显示简单形式 `CASE` 表达式时的失败（Tom Lane）

  如果 `CASE` 的测试表达式是常量，规划器可能把 `CASE` 简化为令表达式显示代码困惑的形式，导致 “unexpected CASE WHEN clause” 错误。

- 修复对现有下标范围之前的数组切片赋值的问题（Tom Lane）

  如果新增下标与第一个既有下标之间存在间隙，
  代码会错误计算需要从旧数组的空值位图复制多少条目，可能导致数据损坏或崩溃。

- 避免规划器对非常久远的日期值出现意外的转换溢出（Tom Lane）

  `date` 类型支持的日期范围比 `timestamp` 类型能表示的更宽，但规划器假定它总是可以不受惩罚地把 date 转换为 timestamp。

- 修复数组包含空条目时 PL/Python 的崩溃（Alex Hunsaker）

- 移除 ecpg 对定义数组维度的常量的固定长度限制（Michael Meskes）

- 修复对包含 `... & !(subexpression) | ...` 的 `tsquery` 值的错误解析（Tom Lane）

  包含这种操作符组合的查询没有被正确执行。同样的错误也存在于 `contrib/intarray` 的 `query_int` 类型和 `contrib/ltree` 的 `ltxtquery` 类型中。

- 修复 `contrib/intarray` 的 `query_int` 类型输入函数中的缓冲区越界（Apple）

  该缺陷构成安全风险，因为函数的返回地址可能被覆写。感谢 Apple Inc 的安全团队报告此问题并提供修复。（CVE-2010-4015）

- 修复 `contrib/seg` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

这可能导致 `seg` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。（这与上一次更新中在 `contrib/cube` 中修复的缺陷相同。）

E.47. 版本 9.0.2

发行日期
2010-12-16

本次发布包含对 9.0.1 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.47.1. 迁移到版本 9.0.2

对于运行 9.0.X 的用户，不需要进行转储/恢复。

E.47.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复热备重放期间的 “too many KnownAssignedXids” 错误（Heikki Linnakangas）

- 修复热备期间锁获取中的竞态条件（Simon Riggs）

- 避免热备期间不必要的冲突（Simon Riggs）

  这修复了一些重放被认为与备库查询冲突（导致重放延迟或可能取消查询）但实际上并不存在真正冲突的情况。

- 修复 GIN 索引 WAL 重放逻辑中的若干缺陷（Tom Lane）

  这可能导致复制期间的 “bad buffer id: 0” 失败或索引内容损坏。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 修复在创建主数据库集簇后立即启用流复制时的极端情况缺陷（Heikki Linnakangas）

- 修复多个工作进程长期保持活动时自动清理工作进程的持续变慢（Tom Lane）

  如果自动清理工作进程处理了足够多的表，其有效 `vacuum_cost_limit` 可能降到接近零，导致其运行极慢。

- 修复自动清理启动器中的长期内存泄漏（Alvaro Herrera）

- 避免在事务之外尝试报告即将发生的事务回卷情况时失败（Tom Lane）

  这一疏忽阻碍了事务回卷过于接近后的恢复，因为数据库启动处理会失败。

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

  给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

  在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 修复早期连接处理期间的错误检查（Tom Lane）

  某些情况下对子进程过多的检查被跳过，可能在尝试把新子进程添加到固定大小数组时导致 postmaster 崩溃。

- 改进窗口函数的效率（Tom Lane）

  某些需要提前读取大量元组、但 `work_mem` 足够大可以全部保存在内存中的情况意外地慢。`percent_rank()`、`cume_dist()` 和 `ntile()` 尤其受此问题影响。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）

- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 补齐 `DROP OWNED BY` 中缺失的移除用户所拥有的外部数据包装器/服务器权限的支持（Heikki Linnakangas）

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复内联参数列表包含对可内联用户函数引用的集合返回函数时的崩溃（Tom Lane）

- 修复当 `ORDER BY`、`LIMIT`、`FOR UPDATE` 或 `WITH` 附加到 `INSERT ... VALUES` 的 `VALUES` 部分时的正确行为（Tom Lane）

- 把 `OFF` 关键字改为非保留（Heikki Linnakangas）

  这避免了在 PL/pgSQL 中使用 `off` 作为变量名的问题。这在 9.0 之前可行，但现在被破坏了，因为 PL/pgSQL 现在把所有核心保留字视为保留。

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

  规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 修复比较整行 Var 时的“could not find pathkey item to sort”规划器失败（Tom Lane）

- 修复当连接接受（`accept()` 或紧随其后的调用之一）失败且 postmaster 编译时带 GSSAPI 支持时的 postmaster 崩溃（Alexander Chernikov）

- 在收到来自 RADIUS 认证服务器的无效响应包之后重试（Magnus Hagander）

  这修复了一个低风险的潜在拒绝服务条件。

- 修复 `log_temp_files` 处于活动状态时临时文件漏删的问题（Tom Lane）

  如果尝试发出日志消息时出错，就不会执行删除，导致临时文件累积。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

  该缺陷影响多个不同的几何距离测量操作符。

- 修复 ecpg 中事务状态的错误计算（Itagaki Takahiro）
- 修复 psql Unicode 转义支持的错误（Tom Lane）
- 加快归档包含许多大对象（blob）时并行 pg_restore 的速度（Tom Lane）
- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）
- 修复 PL/pgSQL 对不存在列情况的错误报告（Tom Lane）

  从 9.0 起，它有时会报“missing FROM-clause entry for table foo”，而更合适的消息应是“record foo has no field bar”。

- 修复 PL/Python，使其在给元组字段赋值时遵循 typmod（即长度或精度限制）（Tom Lane）

  这修复了一个自 8.4 以来的退化。

- 修复 PL/Python 对集合返回函数的处理（Jan Urbanski）

  在生成集合结果的迭代器内尝试调用 SPI 函数会失败。

- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `cube` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。

- `contrib/dblink` 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）
- 修复 `contrib/pgcrypto` 在缺少公钥时可能的核心转储（Marti Raudsepp）
- 修复 `contrib/pg_upgrade` 中的缓冲区越界（Hernan Gonzalez）
- 修复 `contrib/xml2` 的 XPath 查询函数中的内存泄漏（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.48. 版本 9.0.1

发行日期
2010-10-04

本次发布包含对 9.0.0 的多方面修复。有关 9.0 大版本中新特性的信息，请参见第 E.49 节。.

E.48.1. 迁移到版本 9.0.1

对于运行 9.0.X 的用户，不需要进行转储/恢复。

E.48.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

  感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 改进 `pg_get_expr()` 的安全修复，使该函数仍可用于子 SELECT 的输出（Tom Lane）

- 修复占位符求值位置不正确的问题（Tom Lane）

  当外连接的内侧是输出列表中含非严格表达式的子 SELECT 时，该缺陷可能导致本应为空的查询输出非空。

- 修复连接消除对占位符表达式的处理（Tom Lane）

- 修复 `UNION ALL` 成员关系可能被重复扫描的问题（Tom Lane）

- 防止取消监听后 ProcessIncomingNotify() 中的无限循环（Jeff Davis）

- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）

- 重新允许输入公元 0001-01-01 之前的儒略日期（Tom Lane）

  像 `'J100000'::date` 这样的输入在 8.4 之前可行，但被新增的错误检查无意中破坏。

- 让 psql 在自动提交关闭模式下把 `DISCARD ALL` 识别为不应包裹在事务块中的命令（Itagaki Takahiro）

- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）

E.49. 版本 9.0

发行日期
2010-09-20

E.49.1. 概述

这个 PostgreSQL 版本添加了人们多年来请求的特性，如易于使用的复制、批量权限修改工具以及匿名代码块。虽然过去的大版本在范围上偏保守，但本版本展现出提供 PostgreSQL 新老用户都会拥抱的功能的大胆新愿望。这一切都是在极少不兼容的情况下完成的。主要增强包括：

- 基于日志传送的内建复制。
  这一进步由两个特性组成：流复制（允许通过网络连接把连续归档（WAL）文件流式传输到备库服务器）和热备（允许连续归档备库服务器执行只读查询）。
  最终效果是支持一主多只读从库。
- 更简便的数据库对象权限管理。`GRANT/REVOKE IN SCHEMA` 支持对现有对象批量修改权限，而 `ALTER DEFAULT PRIVILEGES` 允许控制未来创建的对象的权限。大对象（BLOB）现在也支持权限管理。
- 广泛增强的存储过程支持。`DO` 语句支持临时或“匿名”代码块。函数现在可以使用命名参数调用。PL/pgSQL 现在默认安装，PL/Perl 和 PL/Python 在多个方面得到增强，包括对 Python3 的支持。
- 对 64 位 Windows 的完整支持。
- 更高级的报表查询，包括额外的窗口选项（`PRECEDING` 和 `FOLLOWING`）以及控制值被馈送给聚合函数的顺序的能力。
- 新的触发器特性，包括符合 SQL 标准的逐列触发器和条件触发器执行。
- 可延迟的唯一约束。现在无需技巧即可对唯一键做批量更新。
- 排他约束。它们提供唯一约束的泛化版本，允许强制执行复杂条件。
- 新增和增强的安全特性，包括 RADIUS 认证、LDAP 认证改进，以及用于测试口令强度的新 contrib 模块 `passwordcheck`。
- `LISTEN/NOTIFY` 特性的新高性能实现。未决事件现在存储在基于内存的队列而不是表中。此外，每个事件可以随附一个“载荷”字符串，而不是像以前那样只传递事件名。
- `VACUUM FULL` 的新实现。该命令现在重写整个表和索引，而不是移动单个行来压缩空间。它在大多数情况下明显更快，而且不再导致索引膨胀。
- 新 contrib 模块 `pg_upgrade`，支持从 8.3 或 8.4 就地升级到 9.0。
- 针对特定类型查询的多项性能增强，包括消除不必要的连接。
  这有助于优化一些自动生成的查询，例如对象关系映射器（ORM）生成的查询。
- `EXPLAIN` 增强。输出现在可采用 JSON、XML 或 YAML 格式，并包含缓冲区利用情况和其他以前无法获得的数据。
- `hstore` 改进，包括新函数和更大的数据容量。

上述各项在下面的章节中有更详细的说明。

E.49.2. 迁移到版本 9.0

对于希望从任何早期版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复，或者使用 pg_upgrade。

版本 9.0 包含一些为支持新特性和代码质量改进而有选择地破坏向后兼容性的变更。特别是，大量使用 PL/pgSQL、时间点恢复（PITR）或温备的用户应测试其应用，因为这些领域有轻微的用户可见变化。请注意以下不兼容之处：

E.49.2.1. 服务器设置

- 移除服务器参数 `add_missing_from`（多年来默认关闭）（Tom Lane）
- 移除服务器参数 `regex_flavor`（多年来默认为 `advanced`）（Tom Lane）
- `archive_mode` 现在只影响 `archive_command`；新设置 `wal_level` 影响预写式日志的内容（Heikki Linnakangas）
- `log_temp_files` 现在默认文件大小单位为千字节（Robert Haas）

E.49.2.2. 查询

- 查询父表时，不再对作为查询一部分被扫描的子表做单独的权限检查（Peter Eisentraut）

  SQL 标准规定了这种行为，而且实践中它也比此前同时检查每个子表和父表权限的行为方便得多。

E.49.2.3. 数据类型

- `bytea` 输出现在默认以十六进制格式显示（Peter Eisentraut）

  如有兼容需要，可使用服务器参数 `bytea_output` 选择传统输出格式。

- 数组输入现在只把普通 ASCII 空白字符视为可能可忽略；它永远不会忽略非 ASCII 字符，即使按某些区域设置它们是空白（Tom Lane）

  这避免了一些数组值可能因服务器区域设置不同而被不同解释的极端情况。

- 改进 `SIMILAR TO` 模式和 SQL 风格 `substring()` 模式的标准符合性（Tom Lane）

  这包括把 `?` 和 `{...}` 视为模式元字符（它们此前是普通字面字符）；这对应 SQL:2008 新增的特性。此外，`^` 和 `$` 现在被视为普通字面字符；此前它们被视为元字符，就好像模式遵循 POSIX 而不是 SQL 规则。另外，在 SQL 标准 `substring()` 中，用括号嵌套不再干扰子串的捕获。另外，方括号表达式（字符类）的处理现在更符合标准。

- 按 SQL 标准拒绝位串 3 参数 `substring()` 中的负长度值（Tom Lane）
- 让 `date_trunc` 在降低秒的小数部分精度时截断而不是舍入（Tom Lane）

  对于基于整数的日期/时间，代码一直如此。现在基于浮点的日期/时间行为类似。

E.49.2.4. 对象重命名

- 当子表从多个不相关的父表继承同一列时，收紧 `RENAME` 期间列名一致性的强制（KaiGai Kohei）
- 不再在底层表列被重命名时自动重命名索引和索引列（Tom Lane）

  管理员仍可以手工重命名这类索引和列。此变更需要更新 JDBC 驱动（可能还有其他驱动），以便重命名后仍能正确识别唯一索引。

- `CREATE OR REPLACE FUNCTION` 不再能更改函数参数的声明名称（Pavel Stehule）

  为了避免在命名参数调用中产生歧义，
  不再允许更改现有函数声明中输入参数的别名（但仍可为此前未命名的参数指定名称）。
  现在必须 `DROP` 并重建函数才能做到这一点。

E.49.2.5. PL/pgSQL

- 当变量名与查询中使用的列名冲突时，PL/pgSQL 现在抛出错误（Tom Lane）

  此前的行为是把有歧义的名称优先绑定到 PL/pgSQL 变量而不是查询列，这常常导致令人惊讶的错误行为。抛出错误可以轻松检测歧义情况。
  虽然建议修改遇到此类错误的函数以消除冲突，但如有必要，可以通过配置参数 `plpgsql.variable_conflict` 或逐函数选项 `#variable_conflict` 恢复旧行为。

- PL/pgSQL 不再允许与某些 SQL 保留字匹配的变量名（Tom Lane）

  这是使 PL/pgSQL 解析器与核心 SQL 解析器更紧密对齐的结果。如有必要，可以为变量名加双引号来避开此限制。

- PL/pgSQL 现在要求组合结果的列不仅匹配基本类型，还要匹配预期的类型修饰符（Pavel Stehule、Tom Lane）

  例如，如果结果类型的某列声明为 `NUMERIC(30,2)`，则在该列中返回其他精度的 `NUMERIC` 不再被接受。以前的版本疏于检查类型修饰符，因此会允许实际上不符合声明限制的结果行。

- PL/pgSQL 现在更一致地处理选择到组合字段的情况（Tom Lane）

  此前，像 `SELECT ... INTO rec.fld FROM ...` 这样的语句即使记录字段 *fld* 是组合类型，也被视为标量赋值。现在它被视为记录赋值，与 `INTO` 目标是组合类型的常规变量时相同。因此，要赋给字段各子字段的值应写成 `SELECT` 列表的单独列，而不是像以前版本那样写成 `ROW(...)` 构造。

  如果需要以同时适用于 9.0 和以前版本的方式做到这一点，可以写成类似 `rec.fld := ROW(...) FROM ...`。

- 移除 PL/pgSQL 的 `RENAME` 声明（Tom Lane）

  用 `ALIAS` 取代 `RENAME`，它现在可以为任何变量创建别名，而不再像以前那样只能用于美元符号参数名（如 `$1`）。

E.49.2.6. 其他不兼容之处

- 废弃把 `=>` 用作操作符名（Robert Haas）

  为了支持 SQL 标准的命名函数参数记法，未来版本的 PostgreSQL 可能会完全拒绝此操作符名。目前它仍被允许，但定义此类操作符时会发出警告。

- 移除对没有可用 64 位整数数据类型的平台的支持（Tom Lane）

  据信所有仍受支持的平台都有可用的 64 位整数数据类型。

E.49.3. 变更

版本 9.0 拥有前所未有的数量的重大新特性，以及 200 多项增强、改进、新命令、新函数和其他变更。

E.49.3.1. 服务器

E.49.3.1.1. 连续归档和流复制

PostgreSQL 现有的备库服务器能力得到扩展，既支持在备库服务器上执行只读查询，又大幅减少主库与备库之间的延迟。对许多用户而言，这将成为一种有用且低管理成本的复制形式，无论用于高可用还是水平扩展。

- 允许备库服务器接受只读查询（Simon Riggs、Heikki Linnakangas）

  此特性称为热备。有新的 `postgresql.conf` 和 `recovery.conf` 设置控制此特性，并有详尽的文档。

- 允许把预写式日志（WAL）数据流式传输到备库服务器（Fujii Masao、Heikki Linnakangas）

  此特性称为流复制。此前 WAL 数据只能以整个 WAL 文件（通常每个 16 MB）为单位发送到备库服务器。流复制消除了这种低效，使主库上的更新能以极小延迟传播到备库服务器。有新的 `postgresql.conf` 和 `recovery.conf` 设置控制此特性，并有详尽的文档。

- 新增 `pg_last_xlog_receive_location()` 和 `pg_last_xlog_replay_location()`，可用于监控备库服务器的 WAL 活动（Simon Riggs、Fujii Masao、Heikki Linnakangas）

E.49.3.1.2. 性能

- 允许通过 `ALTER TABLESPACE ... SET/RESET` 为顺序和随机页面代价估计（`seq_page_cost`/`random_page_cost`）设置逐表空间的值（Robert Haas）

- 改进连接查询中 EvalPlanQual 重新检查的性能和可靠性（Tom Lane）

  涉及连接的 `UPDATE`、`DELETE` 和 `SELECT FOR UPDATE/SHARE` 查询在遇到刚更新的行时现在表现得好得多。

- 改进表在同一事务中较早前被创建或截断时 `TRUNCATE` 的性能（Tom Lane）

- 改进查找继承子表的性能（Tom Lane）

E.49.3.1.3. 优化器

- 移除不必要的外连接（Robert Haas）

  内侧唯一且未被连接以上引用的外连接是不必要的，因此现在会被移除。
  这将加速许多自动生成的查询，例如对象关系映射器（ORM）创建的查询。

- 允许 `IS NOT NULL` 限制使用索引（Tom Lane）

  这在包含大量空值的索引中查找 `MAX()`/`MIN()` 值时特别有用。

- 改进优化器关于何时使用物化节点、以及 `DISTINCT` 用排序还是哈希的选择（Tom Lane）

- 改进优化器对涉及 `boolean <>` 操作符的表达式的等价检测（Tom Lane）

E.49.3.1.4. GEQO

- 让 GEQO 每次规划查询时使用相同的随机种子（Andres Freund）

虽然遗传查询优化器（GEQO）仍然随机选择计划，
但它现在对相同查询总是选择相同的随机计划，从而给出更一致的性能。可以修改 `geqo_seed` 来试验其他计划。

- 改进 GEQO 计划选择（Tom Lane）

  这避免了罕见的 “failed to make a valid plan” 错误，还应能改进规划速度。

E.49.3.1.5. 优化器统计

- 改进 `ANALYZE`，使其支持继承树统计信息（Tom Lane）

  这对分区表特别有用。但自动清理尚不会在子表变化时自动重新分析父表。

- 改进自动清理对何时需要重新分析的检测（Tom Lane）

- 改进优化器对大于/小于比较的估计（Tom Lane）

  在查找大于/小于比较的统计信息时，如果比较值位于第一个或最后一个直方图桶中，
  则使用索引（如果可用）获取该列当前的实際最小值或最大值。
  这大幅改进了对数据范围端点附近比较值的估计准确性，
  尤其是当范围因新数据的添加而不断变化时。

- 允许使用 `ALTER TABLE` 设置不同值数量统计信息（Robert Haas）

  这允许用户覆盖某列不同值的估计数量或百分比。此统计信息通常由 `ANALYZE` 计算，但估计可能不佳，尤其是在行数非常多的表上。

E.49.3.1.6. 认证

- 新增对 RADIUS（远程认证拨入用户服务）认证的支持（Magnus Hagander）

- 允许 LDAP（轻量目录访问协议）认证以 “搜索/绑定” 模式运行（Robert Fleming、Magnus Hagander）

  这允许先查找用户，然后系统使用为该用户返回的 DN（可分辨名称）。

- 为 `pg_hba.conf` 新增 `samehost` 和 `samenet` 指定（Stef Walter）

  它们分别匹配服务器的 IP 地址和子网地址。

- 把受信任的 SSL 根证书名称传递给客户端，使客户端能返回合适的客户端证书（Craig Ringer）

E.49.3.1.7. 监控

- 新增客户端设置应用名的能力，它显示在 `pg_stat_activity` 中（Dave Page）

  这允许管理员按源应用刻画数据库流量并排查问题。

- 为 `log_line_prefix` 新增 SQLSTATE 选项（`%e`）（Guillaume Smet）

  这允许用户按错误码编号汇总错误和消息的统计信息。

- 以 UTF16 编码写入 Windows 事件日志（Itagaki Takahiro）

  现在 Windows 上的 PostgreSQL 日志消息有了真正的多语言支持。

E.49.3.1.8. 统计计数器

- 新增 `pg_stat_reset_shared('bgwriter')`，用于重置后台写入进程的集簇级共享统计信息（Greg Smith）
- 新增 `pg_stat_reset_single_table_counters()` 和 `pg_stat_reset_single_function_counters()`，允许重置单个表和函数的统计计数器（Magnus Hagander）

E.49.3.1.9. 服务器设置

- 允许基于数据库/角色组合设置配置参数（Alvaro Herrera）

  此前只能按数据库和按角色设置，不能组合。所有角色和数据库设置现在存储在新的 `pg_db_role_setting` 系统目录中。新的 psql 命令 `\drds` 显示这些设置。传统系统视图 `pg_roles`、`pg_shadow` 和 `pg_user` 不显示组合设置，因此不再完全表示用户或数据库的配置。

- 新增服务器参数 `bonjour`，控制启用 Bonjour 的服务器是否通过 Bonjour 宣告自身（Tom Lane）

  默认为关闭，即不宣告。这使打包者可以分发启用 Bonjour 的构建，而不必担心个别用户不想要此特性。

- 新增服务器参数 `enable_material`，控制优化器对物化节点的使用（Robert Haas）

  默认为开启。关闭时，优化器不会纯粹出于性能原因添加物化节点，
  但在正确性需要时仍会使用。

- 更改服务器参数 `log_temp_files`，使其默认文件大小单位为千字节（Robert Haas）

  此前如果未指定单位，此设置按字节解释。

- 在重新加载 `postgresql.conf` 时记录参数值的变更（Peter Eisentraut）

  这使管理员和安全人员可以审计数据库设置的变更，而且对检查 `postgresql.conf` 编辑的效果也非常方便。

- 对自定义服务器参数正确强制超级用户权限（Tom Lane）

  非超级用户不再能对服务器当前未知的参数发出 `ALTER ROLE/DATABASE SET`。这使服务器能够正确检查仅超级用户参数只由超级用户设置。此前，`SET` 会被允许，然后在会话启动时被忽略，使仅超级用户自定义参数远不如应有的有用。

E.49.3.2. 查询

- 在应用 `LIMIT` 之后执行 `SELECT FOR UPDATE/SHARE` 处理，使返回的行数总是可预测（Tom Lane）

  此前，并发事务所做的更改可能导致 `SELECT FOR UPDATE` 意外返回少于 `LIMIT` 指定的行数。`FOR UPDATE` 与 `ORDER BY` 组合仍可能产生令人惊讶的结果，但可以通过把 `FOR UPDATE` 放在子查询中来纠正。

- 允许混用传统与 SQL 标准的 `LIMIT/OFFSET` 语法（Tom Lane）
- 扩展窗口函数支持的帧选项（Hitoshi Harada）

  帧现在可以以 `CURRENT ROW` 开始，并且支持 `ROWS` *n* `PRECEDING/FOLLOWING` 选项。

- 让 `SELECT INTO` 和 `CREATE TABLE AS` 在其命令标签中向客户端返回行数（Boszormenyi Zoltan）

  这可以省去一整次到客户端的往返，允许在不需要额外 `COUNT` 查询的情况下计算结果数和分页。

E.49.3.2.1. Unicode 字符串

- 在 `U&` 字符串和标识符中支持 Unicode 代理对（双 16 位表示）（Peter Eisentraut）
- 在 `E'...'` 字符串中支持 Unicode 转义（Marko Kreen）

E.49.3.3. 对象操作

- 通过推迟刷盘加速 `CREATE DATABASE`（Andres Freund、Greg Stark）
- 只允许对表、视图和组合类型的列加注释，不允许对索引和 TOAST 表等其他关系类型（Tom Lane）
- 允许创建不含任何值的枚举类型（Bruce Momjian）
- 让存储类型为 `MAIN` 的列值保留在主堆页上，除非行放不进一个页面（Kevin Grittner）

  此前 `MAIN` 值会被强制移出到 TOAST 表，直到行大小小于页面大小的四分之一。

E.49.3.3.1. `ALTER TABLE`

- 为 `ALTER TABLE DROP COLUMN` 和 `ALTER TABLE DROP CONSTRAINT` 实现 `IF EXISTS`（Andres Freund）
- 允许重写表的 `ALTER TABLE` 命令跳过 WAL 日志记录（Itagaki Takahiro）

  这类操作要么产生表的新副本要么回滚，因此可以跳过 WAL 归档，除非运行在连续归档模式下。这减少了 I/O 开销并改进了性能。

- 修复由非表所有者执行 `ALTER TABLE *table* ADD COLUMN *col* serial` 时的失败（Tom Lane）

E.49.3.3.2. `CREATE TABLE`

- 为 `CREATE TABLE ... LIKE` 命令新增复制 `COMMENTS` 和 `STORAGE` 设置的支持（Itagaki Takahiro）
- 为 `CREATE TABLE ... LIKE` 命令新增复制所有属性的快捷方式（Itagaki Takahiro）
- 新增 SQL 标准的 `CREATE TABLE ... OF *type*` 命令（Peter Eisentraut）

  这允许创建匹配现有组合类型的表。命令中还可以指定额外的约束和默认值。

E.49.3.3.3. 约束

- 新增可延迟唯一约束（Dean Rasheed）

  这使批量更新（例如 `UPDATE tab SET col = col + 1`）能够在有唯一索引或标记为主键的列上可靠工作。如果约束指定为 `DEFERRABLE`，它将在语句末尾而不是每行更新后检查。约束检查还可以推迟到当前事务末尾，使这类更新可以分散到多个 SQL 命令中。

- 新增排他约束（Jeff Davis）

  排他约束通过允许任意比较操作符（而不仅仅是相等）来泛化唯一约束。它们通过 `CREATE TABLE CONSTRAINT ... EXCLUDE` 子句创建。排他约束最常见的用途是指定列项不得重叠，而不仅仅是不相等。这对时间段和其他范围以及数组都很有用。此特性增强了许多日历、时间管理和科学应用的数据完整性检查。

- 改进唯一约束违规错误消息，报告导致失败的值（Itagaki Takahiro）

  例如，唯一约束违规现在可能报告 `Key (x)=(2) already exists`。

E.49.3.3.4. 对象权限

- 新增使用新的 `GRANT/REVOKE IN SCHEMA` 子句在整个模式内批量修改权限的能力（Petr Jelinek）

  这简化了对象权限的管理，并使为应用数据安全利用数据库角色更加容易。

- 新增 `ALTER DEFAULT PRIVILEGES` 命令，用于控制稍后创建的对象的权限（Petr Jelinek）

  这大幅简化了复杂数据库应用中对象权限的分配。可以为表、视图、序列和函数设置默认权限。默认值可以按模式分配，也可以数据库范围分配。

- 新增用 `GRANT/REVOKE` 控制大对象（BLOB）权限的能力（KaiGai Kohei）

  此前任何数据库用户都可以读取或修改任何大对象。现在可以逐大对象授予和撤销读写权限，并且会跟踪大对象的所有权。

E.49.3.4. 工具操作

- 让 `LISTEN/NOTIFY` 把未决事件存储在内存队列而不是系统表中（Joachim Wieland）

  这大幅改进了性能，同时保留了事务支持和保证送达的现有特性。

- 允许 `NOTIFY` 向监听者传递可选的“载荷”字符串（Joachim Wieland）

  这大幅提升了 `LISTEN/NOTIFY` 作为通用事件队列系统的实用性。

- 允许对所有逐数据库系统目录执行 `CLUSTER`（Tom Lane）

  共享目录仍然不能聚簇。

E.49.3.4.1. COPY

- 接受 `COPY ... CSV FORCE QUOTE *`（Itagaki Takahiro）

  现在 `*` 可用作 `FORCE QUOTE` 子句中“所有列”的简写。

- 新增允许在括号内指定选项的 `COPY` 语法（Robert Haas、Emmanuel Cecchet）

  这为未来的 `COPY` 选项提供了更大的灵活性。仍支持旧语法，但只用于既有选项。

E.49.3.4.2. EXPLAIN

- 允许 `EXPLAIN` 以 XML、JSON 或 YAML 格式输出（Robert Haas、Greg Sabino Mullane）

  新的输出格式易于机器读取，支持开发分析 `EXPLAIN` 输出的新工具。

- 新增 `BUFFERS` 选项，用于在 `EXPLAIN ANALYZE` 期间报告查询缓冲区使用情况（Itagaki Takahiro）

  这允许对单个查询做更好的分析。缓冲区使用情况不再报告在 log_statement_stats 及相关设置的输出中。

- 在 `EXPLAIN` 输出中添加哈希使用信息（Robert Haas）

- 新增允许在括号内指定选项的 `EXPLAIN` 语法（Robert Haas）

  这为未来的 `EXPLAIN` 选项提供了更大的灵活性。仍支持旧语法，但只用于既有选项。

E.49.3.4.3. VACUUM

- 更改 `VACUUM FULL`，使其重写整个表并重建其索引，而不是移动单个行来压缩空间（Itagaki Takahiro、Tom Lane）

  以前的方法通常更慢并导致索引膨胀。注意新方法在 `VACUUM FULL` 期间会暂时使用更多磁盘空间；可能多达表及其索引通常占用空间的两倍。

- 新增允许在括号内指定选项的 `VACUUM` 语法（Itagaki Takahiro）

  这为未来的 `VACUUM` 选项提供了更大的灵活性。仍支持旧语法，但只用于既有选项。

E.49.3.4.4. 索引

- 允许在 `CREATE INDEX` 中省略索引名以自动命名索引（Tom Lane）

- 默认情况下，多列索引现在以所有列命名；索引表达式列现在基于其表达式命名（Tom Lane）

- 重建共享系统目录的索引现在完全是事务性且崩溃安全的（Tom Lane）

  此前，重建共享索引只允许在独立模式下进行，
  而且操作期间崩溃可能使索引处于比之前更糟的状态。

- 为 GiST 新增 `point_ops` 操作符类（Teodor Sigaev）

  此特性允许对 `point` 列建立 GiST 索引。该索引可用于多种类型的查询，例如 *`point <@ polygon`*（点在多边形内）。这应能使许多 PostGIS 查询更快。

- 在 GIN 索引创建中使用红黑二叉树（Teodor Sigaev）

  红黑树是自平衡的。这避免了输入以非随机顺序出现时的变慢。

E.49.3.5. 数据类型

- 允许 `bytea` 值以十六进制记法书写（Peter Eisentraut）

  服务器参数 `bytea_output` 控制 `bytea` 输出使用十六进制还是传统格式。Libpq 的 `PQescapeByteaConn()` 函数在连接 PostgreSQL 9.0 或更新服务器时自动使用十六进制格式。但 9.0 之前的 libpq 版本无法正确处理来自更新服务器的十六进制格式。

  新的十六进制格式与更多使用二进制数据的应用直接兼容，
  使它们无需额外转换即可存储和获取。它的读写也明显快于传统格式。

- 允许把服务器参数 extra_float_digits 增大到 `3`（Tom Lane）

此前 `extra_float_digits` 的最大设置为 `2`。有些情况需要 3 位数字才能精确转储和恢复 `float4` 值。pg_dump 现在在从允许它的服务器转储时将使用设置 3。

- 收紧对 `int2vector` 值的输入检查（Caleb Welton）

E.49.3.5.1. Full Text Search

- 在 `synonym` 词典中新增前缀支持（Teodor Sigaev）
- 新增过滤词典（Teodor Sigaev）

  过滤词典允许先修改词元再传递给后续词典。

- 允许电子邮件地址词元中包含下划线（Teodor Sigaev）
- 使用更符合标准的规则解析 URL 词元（Tom Lane）

E.49.3.6. 函数

- 允许函数调用提供参数名并与函数定义中的命名参数匹配（Pavel Stehule）

  例如，如果函数定义为接受参数 `a` 和 `b`，则可以用 `func(a := 7, b := 12)` 或 `func(b := 12, a := 7)` 调用它。

- 支持在 UTF-8 服务器编码下进行区域特定的正则表达式处理（Tom Lane）

  区域特定的正则表达式功能包括不区分大小写的匹配和区域特定的字符类。此前，这些特性只有当数据库使用单字节服务器编码（如 LATIN1）时才能对非 ASCII 字符正确工作。在 UTF-8 之外的多字节编码中它们仍会行为不当。

- 在 `to_char()` 中新增科学计数法支持（`EEEE` 规范）（Pavel Stehule、Brendan Jurd）
- 让 `to_char()` 在 `Y`、`YY` 和 `YYY` 规范中遵循 `FM`（填充模式）（Bruce Momjian、Tom Lane）

  `YYYY` 此前已遵循它。

- 修复 `to_char()`，使其在 Windows 上以正确的编码输出本地化的数字和货币字符串（Hiroshi Inoue、Itagaki Takahiro、Bruce Momjian）
- 修正多边形的“重叠”和“包含”运算的计算（Teodor Sigaev）

  多边形 `&&`（重叠）操作符此前只检查两个多边形的边界框是否重叠。现在它做更正确的检查。多边形 `@>` 和 `<@`（包含/被包含）操作符此前检查一个多边形的所有顶点是否都包含在另一个中；这对某些非凸多边形可能错误地报告“真”。现在它们检查一个多边形的所有线段都包含在另一个中。

E.49.3.6.1. 聚合

- 允许聚合函数使用 `ORDER BY`（Andrew Gierth）

  例如，现在支持：`array_agg(a ORDER BY b)`。这对输入值顺序很重要的聚合很有用，并消除了使用非标准子查询来确定排序的需要。

- 多参数聚合函数现在可以使用 `DISTINCT`（Andrew Gierth）
- 新增聚合函数 `string_agg()`，用于把值组合成单个字符串（Pavel Stehule）

- 如果聚合转换函数未标记为 `STRICT`，以 `DISTINCT` 调用的聚合函数现在会被传递 NULL 值（Andrew Gierth）

  例如，`agg(DISTINCT x)` 可能会把 NULL 的 `x` 值传递给 `agg()`。这与非 `DISTINCT` 情况下的行为更一致。

E.49.3.6.2. 位串

- 为 `bit` 字符串新增 `get_bit()` 和 `set_bit()` 函数，与 `bytea` 的对应函数保持一致（Leonardo F）
- 为 `bit` 字符串和 `bytea` 实现 `OVERLAY()`（替换）（Leonardo F）

E.49.3.6.3. 对象信息函数

- 新增 `pg_table_size()` 和 `pg_indexes_size()`，为 `pg_relation_size()` 函数提供更用户友好的接口（Bernd Helmle）
- 新增用于序列权限检查的 `has_sequence_privilege()`（Abhijit Menon-Sen）
- 更新 information_schema 视图以符合 SQL:2008（Peter Eisentraut）
- 让 `information_schema` 视图正确显示 `char` 和 `varchar` 列的最大八位字节长度（Peter Eisentraut）
- 加速 `information_schema` 权限视图（Joachim Wieland）

E.49.3.6.4. 函数和触发器创建

- 支持使用 `DO` 语句执行匿名代码块（Petr Jelinek、Joshua Tolley、Hannu Valtonen）

  这允许执行服务器端代码而无需创建和删除临时函数定义。
  可以用用户有权定义函数的任何语言执行代码。

- 实现符合 SQL 标准的逐列触发器（Itagaki Takahiro）

  这类触发器只在指定列受查询影响（例如出现在 `UPDATE` 的 `SET` 列表中）时触发。

- 为 `CREATE TRIGGER` 新增 `WHEN` 子句，允许控制触发器是否触发（Itagaki Takahiro）

  虽然同类检查总是可以在触发器内部执行，但在外部 `WHEN` 子句中执行可能有性能收益。

E.49.3.7. 服务器端语言

- 为 `CREATE LANGUAGE` 新增 `OR REPLACE` 子句（Tom Lane）

  这有助于在语言尚不存在时选择性地安装它，而且在 PL/pgSQL 默认安装的现在特别有用。

E.49.3.7.1. PL/pgSQL Server-Side Language

- 默认安装 PL/pgSQL（Bruce Momjian）

  如果管理员对其可用性有安全或性能方面的顾虑，仍可以把它从特定数据库中移除。

- 改进对 PL/pgSQL 变量名与函数内查询所用标识符冲突情况的处理（Tom Lane）

现在的默认行为是在存在冲突时抛出错误，以避免令人惊讶的行为。可以通过配置参数 `plpgsql.variable_conflict` 或逐函数选项 `#variable_conflict` 修改它，允许使用变量或查询提供的列。无论如何，PL/pgSQL 不再尝试在语法上无效的位置替换变量。

- 让 PL/pgSQL 使用主词法分析器而不是自己的版本（Tom Lane）

  这确保了对主系统行为（如字符串转义等细节）的准确跟踪。一些用户可见的细节（例如 PL/pgSQL 中被视为保留的关键字集合）因此而改变。

- 避免对无效的记录引用抛出不必要的错误（Tom Lane）

  现在只有当引用实际被获取时才抛出错误，而不是在到达外层表达式时就抛出。例如，许多人曾在触发器中尝试这样做：

  ```
  if TG_OP = 'INSERT' and NEW.col1 = ... then
  ```

  这现在将按预期工作。

- 改进 PL/pgSQL 处理带已删除列的行类型的能力（Pavel Stehule）

- 允许在 PL/pgSQL 函数内为输入参数赋值（Steve Prentice）

  此前输入参数被视为声明为 `CONST`，因此函数代码不能更改它们的值。此限制已被移除，以简化从没有等效限制的其他 DBMS 移植函数。输入参数现在就像初始化为传入值的局部变量。

- 改进 PL/pgSQL 的错误位置报告（Tom Lane）

- 为 PL/pgSQL 的 `MOVE FORWARD/BACKWARD` 新增 *`count`* 和 `ALL` 选项（Pavel Stehule）

- 允许 PL/pgSQL 的 `WHERE CURRENT OF` 使用游标变量（Tom Lane）

- 允许 PL/pgSQL 的 `OPEN` *`cursor`* `FOR EXECUTE` 使用参数（Pavel Stehule、Itagaki Takahiro）

  这通过新的 `USING` 子句实现。

E.49.3.7.2. PL/Perl Server-Side Language

- 新增 PL/Perl 函数：`quote_literal()`、`quote_nullable()`、`quote_ident()`、`encode_bytea()`、`decode_bytea()`、`looks_like_number()`、`encode_array_literal()`、`encode_array_constructor()`（Tim Bunce）

- 新增服务器参数 `plperl.on_init`，用于指定 PL/Perl 初始化函数（Tim Bunce）

  还提供 `plperl.on_plperl_init` 和 `plperl.on_plperlu_init`，分别用于可信或不可信语言的特定初始化。

- 支持 PL/Perl 中的 `END` 块（Tim Bunce）

  `END` 块目前不允许数据库访问。

- 允许 PL/Perl 中使用 `use strict`（Tim Bunce）

  Perl `strict` 检查还可以通过新的服务器参数 `plperl.use_strict` 全局启用。

- 允许 PL/Perl 中使用 `require`（Tim Bunce）

它基本上是测试模块是否已加载，如果没有则生成错误。
它不允许加载管理员未通过初始化参数预加载的模块。

- 如果使用 Perl 5.10 或更高版本，允许 PL/Perl 中使用 `use feature`（Tim Bunce）
- 验证 PL/Perl 返回值在服务器编码中有效（Andrew Dunstan）

E.49.3.7.3. PL/Python Server-Side Language

- 在 PL/Python 中新增 Unicode 支持（Peter Eisentraut）

  字符串会根据需要自动与服务器编码相互转换。

- 改进 PL/Python 的 `bytea` 支持（Caleb Welton）

  传入 PL/Python 的 `bytea` 值现在表示为二进制而不是 PostgreSQL 的 `bytea` 文本格式。包含空字节的 `bytea` 值现在也能从 PL/Python 正确输出。布尔、整数和浮点值的传递也得到了改进。

- 支持在 PL/Python 中使用数组作为参数和返回值（Peter Eisentraut）
- 改进 SQL 域到 Python 类型的映射（Peter Eisentraut）
- 为 PL/Python 新增 Python 3 支持（Peter Eisentraut）

  新的服务器端语言名为 `plpython3u`。它不能与 Python 2 服务器端语言在同一会话中使用。

- 改进 PL/Python 的错误位置和异常报告（Peter Eisentraut）

E.49.3.8. 客户端应用程序

- 为 `vacuumdb` 新增 `--analyze-only` 选项，只分析不清理（Bruce Momjian）

E.49.3.8.1. psql

- 新增对 psql 变量值作为 SQL 字符串或标识符加引号/转义的支持（Pavel Stehule、Robert Haas）

  例如，`:'var'` 会产生加了引号并正确转义为字面字符串的 `var` 值，而 `:"var"` 会产生加了引号并转义为标识符的值。

- 让 psql 忽略其读取的脚本文件中前导的 UTF-8 编码 Unicode 字节顺序标记（Itagaki Takahiro）

  这在客户端编码为 UTF-8 时启用。它改进了与某些（主要在 Windows 上）坚持插入此类标记的编辑器的兼容性。

- 修复 `psql --file -`，使其正确遵循 `--single-transaction`（Bruce Momjian）
- 避免两个 psql 会话并发运行时 psql 的命令行历史相互覆盖（Tom Lane）
- 改进 psql 的 Tab 补全支持（Itagaki Takahiro）
- 只要启用了 `\timing` 就显示其输出，无论是否处于“安静”模式（Peter Eisentraut）

E.49.3.8.1.1. psql Display

- 改进 psql 中换行列的显示（Roger Leigh）

此行为现在是默认。以前的格式可以通过 `\pset linestyle old-ascii` 使用。

- 允许 psql 通过 `\pset linestyle unicode` 使用精致的 Unicode 画线字符（Roger Leigh）

E.49.3.8.1.2. psql \d Commands

- 让 `\d` 显示继承自指定父表的子表（Damien Clochard）

  `\d` 只显示子表数量，而 `\d+` 显示所有子表的名称。

- 在 `\d index_name` 中显示索引列的定义（Khee Chin）

  该定义对表达式索引很有用。

- 只在 `\d+` 而不是 `\d` 中显示视图的定义查询（Peter Eisentraut）

  始终包含查询被认为过于冗长。

E.49.3.8.2. pg_dump

- 让 pg_dump/pg_restore 的 `--clean` 也移除大对象（Itagaki Takahiro）
- 修复 pg_dump，使其在启用 `standard_conforming_strings` 时正确转储大对象（Tom Lane）

  以前的编码在转储到归档文件再由 pg_restore 生成脚本输出时可能失败。

- pg_restore 现在在生成脚本输出时以十六进制格式发出大对象数据（Tom Lane）

  如果脚本随后被加载到 9.0 之前的服务器，这可能导致兼容性问题。解决办法是改为直接恢复到服务器。

- 允许 pg_dump 转储附加在组合类型列上的注释（Taro Minowa (Higepon)）
- 让 pg_dump `--verbose` 在文本输出模式下输出 pg_dump 和服务器版本（Jim Cox、Tom Lane）

  这些在自定义输出模式下早已提供。

- pg_restore 现在会在开关和可选文件名之后还有命令行参数时抱怨（Tom Lane）

  此前它会静默忽略任何这类参数。

E.49.3.8.3. pg_ctl

- 允许在系统重启期间安全地用 pg_ctl 启动 postmaster（Tom Lane）

  此前，pg_ctl 的父进程可能基于过期的 postmaster 锁文件被误识别为运行中的 postmaster，导致数据库瞬时启动失败。

- 赋予 pg_ctl 初始化数据库的能力（通过调用 initdb）（Zdenek Kotala）

E.49.3.9. Development Tools

E.49.3.9.1. libpq

- 新增 libpq 函数 `PQconnectdbParams()` 和 `PQconnectStartParams()`（Guillaume Lelarge）

这些函数类似于 `PQconnectdb()` 和 `PQconnectStart()`，只是它们接受以空结尾的连接选项数组，而不是要求在单个字符串中提供所有选项。

- 新增 libpq 函数 `PQescapeLiteral()` 和 `PQescapeIdentifier()`（Robert Haas）

  这些函数返回适当加引号和转义的 SQL 字符串字面量和标识符。调用者不需要像 `PQescapeStringConn()` 那样预分配字符串结果。

- 新增对每用户服务文件（`.pg_service.conf`）的支持，它在全站服务文件之前检查（Peter Eisentraut）
- 如果找不到指定的 libpq 服务则正确报告错误（Peter Eisentraut）
- 在 libpq 中新增 TCP keepalive 设置（Tollef Fog Heen、Fujii Masao、Robert Haas）

  TCP 连接的服务器端此前已支持 keepalive 设置。

- 在提供替代方法的平台上，避免 libpq 中为阻塞和取消阻塞 `SIGPIPE` 而进行的额外系统调用（Jeremy Kerr）
- 当 `.pgpass` 提供的口令失败时，在错误消息中说明口令来源（Bruce Momjian）
- 加载客户端证书文件中给出的所有 SSL 证书（Tom Lane）

  这改进了对间接签名 SSL 证书的支持。

E.49.3.9.2. ecpg

- 为 ecpg 新增 SQLDA（SQL 描述符区域）支持（Boszormenyi Zoltan）
- 为 ecpg 新增 `DESCRIBE` [`OUTPUT`] 语句（Boszormenyi Zoltan）
- 新增返回当前事务状态的 ECPGtransactionStatus 函数（Bernd Helmle）
- 在 ecpg Informix 兼容模式中新增 `string` 数据类型（Boszormenyi Zoltan）
- 允许 ecpg 无限制地使用 `new` 和 `old` 变量名（Michael Meskes）
- 允许 ecpg 在 `free()` 中使用变量名（Michael Meskes）
- 让 `ecpg_dynamic_type()` 对非 SQL3 数据类型返回零（Michael Meskes）

  此前它返回数据类型 OID 的负值。但这可能与有效的类型 OID 混淆。

- 在已有 64 位 `long` 的平台上支持 `long long` 类型（Michael Meskes）

E.49.3.9.2.1. ecpg Cursors

- 在 ecpg 的本机模式中新增范围外游标支持（Boszormenyi Zoltan）

  这允许 `DECLARE` 使用 `OPEN` 被调用时不在作用域内的变量。此机制在 ecpg 的 Informix 兼容模式中早已存在。

- 允许 ecpg 中的动态游标名（Boszormenyi Zoltan）
- 允许 ecpg 在 `FETCH` 和 `MOVE` 中使用噪声词 `FROM` 和 `IN`（Boszormenyi Zoltan）

E.49.3.10. 构建选项

- 默认启用客户端线程安全（Bruce Momjian）

  线程安全选项可以用 `configure --disable-thread-safety` 禁用。

- 新增对控制 Linux 内存不足终止器的支持（Alex Hunsaker、Tom Lane）

  既然 `/proc/self/oom_adj` 允许禁用 Linux 内存不足（OOM）终止器，建议对 postmaster 禁用 OOM 终止。之后可能希望对 postmaster 的子进程重新启用 OOM 终止。新的编译期选项 `LINUX_OOM_ADJ` 允许对子进程重新激活终止器。

E.49.3.10.1. Makefile

- 新的 `Makefile` 目标 `world`、`install-world` 和 `installcheck-world`（Andrew Dunstan）

  它们类似于现有的 `all`、`install` 和 `installcheck` 目标，但它们还构建 HTML 文档、构建并测试 `contrib`，并测试服务器端语言和 ecpg。

- 为 PGXS Makefile 新增数据和文档安装位置控制（Mark Cave-Ayland）

- 新增 Makefile 规则，把 PostgreSQL 文档构建为单个 HTML 文件或单个纯文本文件（Peter Eisentraut、Bruce Momjian）

E.49.3.10.2. Windows

- 支持在 64 位 Windows 上编译并以 64 位模式运行（Tsutomu Yamada、Magnus Hagander）

  这允许在 Windows 上使用大型共享内存。

- 支持使用 Visual Studio 2008 构建服务器（Magnus Hagander）

E.49.3.11. 源代码

- 以子目录树而非发行版压缩包内的 tar 归档文件的形式分发预构建文档（Peter Eisentraut）

  例如，预构建的 HTML 文档现在位于 `doc/src/sgml/html/`；手册页的打包方式类似。

- 使服务器词法分析器可重入（Tom Lane）

  这是 PL/pgSQL 使用该词法分析器所需要的。

- 改进内存分配的速度（Tom Lane、Greg Stark）

- 用户定义的约束触发器现在在 `pg_constraint` 和 `pg_trigger` 中都有条目（Tom Lane）

  由于此变更，`pg_constraint.pgconstrname` 现在是冗余的，已被移除。

- 新增系统目录列 `pg_constraint.conindid` 和 `pg_trigger.tgconstrindid`，以更好地记录约束强制执行中索引的使用（Tom Lane）

- 允许使用单个操作系统信号向后台进程传达多个条件（Fujii Masao）

  这允许新增特性而不受平台特定的信号条件数量限制。

- 改进源代码测试覆盖率，包括 `contrib`、PL/Python 和 PL/Perl（Peter Eisentraut、Andrew Dunstan）
- 移除系统表引导时对平面文件的使用（Tom Lane、Alvaro Herrera）

  这改进了使用大量角色或数据库时的性能，并消除了一些可能的失败条件。
- 自动为“引导”目录生成 `pg_attribute` 的初始内容（John Naylor）

  这大幅简化了对这些目录的更改。
- 把 `INSERT`/`UPDATE`/`DELETE` 操作的处理从 `execMain.c` 中拆分出来（Marko Tiikkaja）

  更新现在在单独的 ModifyTable 节点中执行。此变更是未来改进的必要基础设施。
- 简化 psql SQL 帮助文本的翻译（Peter Eisentraut）
- 缩短某些文件名的长度，使发行版压缩包中所有文件路径都少于 100 个字符（Tom Lane）

  某些解压程序对更长的文件路径有问题。
- 新增 `ERRCODE_INVALID_PASSWORD` `SQLSTATE` 错误码（Bruce Momjian）
- 经作者许可，移除少数剩余的个人源代码版权声明（Bruce Momjian）

  个人版权声明无关紧要，但社区偶尔不得不回答关于它们的问题。
- 新增关于以非持久模式运行 PostgreSQL 以改进性能的文档节（Bruce Momjian）
- 重构 HTML 文档 `Makefile` 规则，使其依赖检查正确工作，避免不必要的重建（Peter Eisentraut）
- 使用 DocBook XSL 样式表而非 Docbook2X 构建手册页（Peter Eisentraut）

  这改变了构建手册页所需的工具集。
- 改进 PL/Perl 代码结构（Tim Bunce）
- 改进 PL/Perl 的错误上下文报告（Alexey Klyukin）

E.49.3.11.1. 新构建要求

注意，从发行版压缩包构建时这些要求不适用，因为压缩包已包含这些程序用来构建的文件。

- 要求 Autoconf 2.63 来构建 configure（Peter Eisentraut）
- 要求 Flex 2.5.31 或更高版本以从 CVS 检出构建（Tom Lane）
- 要求 Perl 5.8 或更高版本以从 CVS 检出构建（John Naylor、Andrew Dunstan）

E.49.3.11.2. 可移植性

- 为 Bonjour 使用更现代的 API（Tom Lane）

  Bonjour 支持现在要求 OS X 10.3 或更高版本。旧 API 已被 Apple 弃用。
- 新增 SuperH 架构的自旋锁支持（Nobuhiro Iwamatsu）
- 允许非 GCC 编译器在支持时使用内联函数（Kurt Harriman）

- 移除对没有可用 64 位整数数据类型的平台的支持（Tom Lane）
- 重构 `LDFLAGS` 的使用，使其在平台间更一致（Tom Lane）

  `LDFLAGS` 现在用于链接可执行文件和共享库，并且在链接可执行文件时附加 `LDFLAGS_EX`，链接共享库时附加 `LDFLAGS_SL`。

E.49.3.11.3. 服务器编程

- 使后端头文件可以安全地包含在 C++ 中（Kurt Harriman、Peter Eisentraut）

  这些变更移除了此前使 C++ 难以用于后端代码的关键字冲突。但是，在为后端函数使用 C++ 时仍有其他复杂之处。在适当的地方仍需要 `extern "C" { }`，内存管理和错误处理仍是有问题的。

- 新增 `AggCheckCallContext()`，用于检测 C 函数是否被作为聚合调用（Hitoshi Harada）
- 更改 `SearchSysCache()` 及相关函数的调用约定，以避免硬编码缓存键的最大数量（Robert Haas）

  现有调用目前仍可工作，但如果不转换为新风格，预计会在 9.1 或更高版本中中断。

- 要求对 `fastgetattr()` 和 `heap_getattr()` 后端宏的调用提供非 NULL 的第四个参数（Robert Haas）
- 自定义 typanalyze 函数不应再依赖 `VacAttrStats.attr` 来确定将要传递给它们的数据类型（Tom Lane）

  此变更是为了允许对存储类型与底层列数据类型不同的索引列收集统计信息。
  有新的字段告知实际被分析的数据类型。

E.49.3.11.4. 服务器钩子

- 新增用于处理 ColumnRef 和 ParamRef 节点的解析器钩子（Tom Lane）
- 新增 ProcessUtility 钩子，使可加载模块可以控制工具命令（Itagaki Takahiro）

E.49.3.11.5. 二进制升级支持

- 新增 `contrib/pg_upgrade` 以支持就地升级（Bruce Momjian）

  这避免了升级到 PostgreSQL 新大版本时转储/重载数据库的要求，从而把停机时间降低了数量级。它支持从 PostgreSQL 8.3 和 8.4 升级到 9.0。

- 新增在二进制升级期间保留关系 `relfilenode` 值的支持（Bruce Momjian）
- 新增在二进制升级期间保留 `pg_type` 和 `pg_enum` OID 的支持（Bruce Momjian）
- 把表空间内的数据文件移到 PostgreSQL 版本特定的子目录中（Bruce Momjian）

  这简化了二进制升级。

E.49.3.12. 附加模块

- 为 `contrib/pgbench` 新增多线程选项（`-j`）（Itagaki Takahiro）

  这允许 pgbench 使用多个 CPU，降低 pgbench 本身成为测试瓶颈的风险。

- 为 contrib/pgbench 新增 \shell 和 \setshell 元命令（Michael Paquier）
- contrib/dict_xsyn 的新特性（Sergey Karpov）

  新选项为 matchorig、matchsynonyms 和 keepsynonyms。
- 新增全文词典 contrib/unaccent（Teodor Sigaev）

  这个过滤词典移除字母上的重音符号，使跨多种语言的全文搜索容易得多。
- 为 contrib/dblink 新增 dblink_get_notify()（Marcus Kempe）

  这允许 dblink 中的异步通知。
- 改进 contrib/dblink 对已删除列的处理（Tom Lane）

  这影响 dblink_build_sql_insert() 及相关函数。这些函数现在按逻辑而非物理列编号对列计数。
- 大幅提高 contrib/hstore 的数据长度限制，并新增 B-tree 和哈希支持，使 hstore 列上可以进行 GROUP BY 和 DISTINCT 操作（Andrew Gierth）

  还新增了函数和操作符。这些改进使 hstore 成为嵌入 PostgreSQL 的全功能键值存储。
- 新增 contrib/passwordcheck，用于支持站点特定的口令强度策略（Laurenz Albe）

  应修改此模块的源代码以实现站点特定的口令策略。
- 新增 contrib/pg_archivecleanup 工具（Simon Riggs）

  它设计用于 archive_cleanup_command 服务器参数，用于删除不再需要的归档文件。
- 在 contrib/auto_explain 输出中添加查询文本（Andrew Dunstan）
- 为 contrib/pg_stat_statements 新增缓冲区访问计数器（Itagaki Takahiro）
- 更新 contrib/start-scripts/linux，使其使用 /proc/self/oom_adj 禁用 Linux 内存不足（OOM）终止器（Alex Hunsaker、Tom Lane）

E.50. 版本 8.4.22

发行日期
2014-07-24

本次发布包含对 8.4.21 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

这预计是 PostgreSQL 8.4.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.50.1. 迁移到版本 8.4.22

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了某些 GiST 索引的索引损坏问题。请参见下面第一条变更日志，以确认你的安装是否受影响以及应采取的步骤。

不过，如果你是从earlier than 8.4.19 的版本升级，请参见第 E.53 节。

E.50.2. 变更

- 正确初始化 `contrib/btree_gist` 在 `bit` 列上索引的填充字节（Heikki Linnakangas）

  这一错误可能导致本应比较相等的值不被视为相等，从而产生错误的查询结果。在 `bit` 或 `bit varying` 列上建有 GiST 索引的用户，在安装此更新后应对这些索引执行 `REINDEX`。

- 在删除 GIN 列表页时防止残缺页（torn page）（Heikki Linnakangas）

  此修复可防止在页面更新写入磁盘期间发生系统崩溃时可能出现的索引损坏。

- 修复嵌套调用 `ReceiveSharedInvalidMessages` 期间缓存失效可能不正确的问题（Andres Freund）

- 当子查询的目标列表中存在集合返回函数时，不再假定其输出是唯一的（David Rowley）

  这一疏忽可能导致诸如 `WHERE x IN (SELECT y, generate_series(1,10) FROM t GROUP BY y)` 之类的构造被错误优化。

- 修复未能对结构化类型的组合元素中的字段执行解 TOAST 的错误（Tom Lane）

  这纠正了 TOAST 指针可能未经解引用就被复制到其他表的情况。如果原始数据随后被删除，当这个已悬空的指针被使用时，会导致诸如 “missing chunk number 0 for toast value ...” 之类的错误。

- 修复对 Append 计划节点输出的整行引用导致 “record type has not been registered” 失败的问题（Tom Lane）

- 修复在回卷游标时调用用户定义函数可能崩溃的问题（Tom Lane）

- 修复为 `FROM` 中的函数求值参数时查询级内存泄漏的问题（Tom Lane）

- 修复正则表达式处理中的会话级内存泄漏（Tom Lane、Arthur O'Dwyer、Greg Stark）

- 修复 `hungarian.stop` 中的数据编码错误（Tom Lane）

- 修复对在当前事务中插入、随后被一个现已回滚的子事务删除的行的活跃性检查（Andres Freund）

  如果同一事务中稍后执行了 `CREATE INDEX` 或 `CLUSTER`，这可能导致问题（至少是虚假警告，最坏是无限循环）。

- 在 `PREPARE TRANSACTION` 期间清除 `pg_stat_activity.xact_start`（Andres Freund）

  在 `PREPARE` 之后，发起会话已不在事务中，因此不应继续显示事务开始时间。

- 修复 `REASSIGN OWNED` 对文本搜索对象失败的问题（Álvaro Herrera）

- 在 postmaster 启动期间阻塞信号（Tom Lane）

  这确保了 postmaster 在例如启动过程中收到 SIGINT 时能够正确清理自身。

- 保护 `make check` 期间启动的临时 postmaster 的 Unix 域套接字（Noah Misch）

  正如我们此前在 CVE-2014-0067 中指出的，任何能够访问套接字文件的本地用户都可以以服务器的引导超级用户身份连接，进而以运行测试的操作系统用户身份执行任意代码。
  此变更通过把服务器的套接字放到 `/tmp` 下一个临时的、模式为 0700 的子目录中来防御这一风险。但在不支持 Unix 套接字的平台（尤其是 Windows）上风险仍然存在，因为那时临时 postmaster 必须接受本地 TCP 连接。

  此变更的一个有用副作用是简化了覆盖 `DEFAULT_PGSOCKET_DIR` 的构建中的 `make check` 测试。常用的非默认值（如 `/var/run/postgresql`）往往不允许构建用户写入，此前需要一些变通办法，现在不再需要了。

- 在 Windows 上，允许新会话从配置文件获取 PGC_BACKEND 参数（例如log_connections）的值（Amit Kapila）

  此前，如果启动后在文件中更改了这类参数，更改不会生效。

- 在 Windows 上正确地为可执行文件路径名加引号（Nikhil Deshpande）

  如果安装路径同时包含空格和 `@` 符号，这一疏忽可能导致 initdb 和 pg_upgrade 在 Windows 上失败。

- 修复 OS X 上 libpython 的链接（Tom Lane）

  我们先前使用的方法在 Xcode 5.0 及之后版本提供的 Python 库上可能失败。

- 当服务器持续以快于客户端吸收速度的速率发送数据时，避免 libpq 中的缓冲区膨胀（Shin-ichi Morita、Tom Lane）

  libpq 可能被迫不断增大其输入缓冲区，直到内存耗尽（这会被误导性地报告为“lost synchronization with server”）。在正常情况下，数据持续以快于 `recv()` 循环吸收速度的速率传输是相当牵强的，但在客户端因调度器约束被人为减速时曾观察到这种情况。

- 确保 libpq 中的 LDAP 查找尝试按预期超时（Laurenz Albe）

- 修复 pg_restore 对旧式大对象注释的处理（Tom Lane）

  如果归档文件由 9.0 之前版本的 pg_dump 生成且包含较多大对象注释，直接恢复到数据库通常会失败。

- 在 `contrib/pgcrypto` 函数中，确保返回前清除栈变量中的敏感信息（Marko Kreen）

- 在 `contrib/uuid-ossp` 中，跨调用缓存 OSSP UUID 库的状态（Tom Lane）

  这提高了 UUID 生成的效率，并在有 `/dev/urandom` 的平台上减少了从中抽取的熵量。

- 把时区数据文件更新为 tzdata 版本 2014e for DST law changes in Crimea, Egypt, and Morocco。

E.51. 版本 8.4.21

发行日期
2014-03-20

本次发布包含对 8.4.20 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

PostgreSQL 社区将于 2014 年 7 月停止对 8.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.51.1. 迁移到版本 8.4.21

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.19 的版本升级，请参见第 E.53 节。

E.51.2. 变更

- 无条件恢复 GIN 元页以避免残缺页风险（Heikki Linnakangas）

  虽然这一疏忽理论上可能导致索引损坏，但实际上不太可能造成任何问题，因为 GIN 元页的活动部分小于标准的 512 字节磁盘扇区。

- 允许正则表达式操作符被查询取消请求提前终止（Tom Lane）

  这防止了病态正则表达式长时间不可中断地锁死服务器进程的场景。

- 移除试图允许 `OVERLAPS` 接受单元素行参数的错误代码（Joshua Yanovski）

  这段代码从未正确工作过，而且该情况既未被 SQL 标准规定也未被记录，因此移除它似乎比修复它更好。

- 避免在反解析规则或视图时获取超出 `AccessShareLock` 的锁（Dean Rasheed）

  这一疏忽导致 pg_dump 对规则中作为 `INSERT`/`UPDATE`/`DELETE` 命令目标的表意外获取了 `RowExclusiveLock` 锁。虽然通常无害，但它可能干扰试图在这些表上获取例如 `ShareLock` 的并发事务。

- 在报告非 `ERROR` 消息期间防止中断（Tom Lane）

  这可以防止因递归进入 `syslog()` 而导致的罕见服务器进程冻结，或许还能防止其他相关问题。

- 把时区数据文件更新为 tzdata 版本 2014a for DST law changes in Fiji and Turkey, plus historical changes in Israel and Ukraine。

E.52. 版本 8.4.20

发行日期
2014-02-20

本次发布包含对 8.4.19 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

PostgreSQL 社区将于 2014 年 7 月停止对 8.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.52.1. 迁移到版本 8.4.20

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.19 的版本升级，请参见第 E.53 节。

E.52.2. 变更

- 加固 `GRANT ... WITH ADMIN OPTION` 的限制（Noah Misch）

  不带 `ADMIN OPTION` 授予角色本应阻止被授予者向该角色添加或移除成员，但这一限制很容易通过先执行 `SET ROLE` 绕过。其安全影响主要在于角色成员可以撤销其他人的访问权限，这违背了授权者的意愿。未经批准添加角色成员的影响相对较小，
  因为不配合的角色成员本来就可以通过创建视图或 `SECURITY DEFINER` 函数把自己的大部分权限提供给他人。（CVE-2014-0060）

- 防止通过手工调用 PL 验证器函数进行权限提升（Andres Freund）

  PL 验证器函数的主要角色是在 `CREATE FUNCTION` 期间被隐式调用，但它们也是普通 SQL 函数，用户可以显式调用。对实际以其他语言编写的函数调用验证器此前未被检查，可能被用于权限提升。修复方法是在每个验证器函数中加入对权限检查函数的调用。
  非核心过程语言如果自带验证器函数，也需要做同样的修改。（CVE-2014-0061）

- 避免在表和索引 DDL 期间进行多次名称查找（Robert Haas、Andres Freund）

  如果由于并发活动导致名称查找得出不同结论，我们可能会对不同的表执行 DDL 的不同部分。至少在 `CREATE INDEX` 的情况下，这可被用来让权限检查针对与索引创建不同的表执行，从而造成权限提升攻击。（CVE-2014-0062）

- 防止长日期时间字符串造成缓冲区越界（Noah Misch）

  `MAXDATELEN` 常量对于 `interval` 类型可能的最长值来说太小，导致 `interval_out()` 中可能出现缓冲区越界。虽然日期时间输入函数在避免缓冲区越界方面更谨慎，但该限制仍短到导致它们拒绝某些有效输入，例如包含很长时区名的输入。ecpg 库存在这些漏洞以及它自身的一些漏洞。（CVE-2014-0063）

- 防止大小计算中的整数溢出导致缓冲区越界（Noah Misch、Heikki Linnakangas）

  若干函数（大多是类型输入函数）在计算分配大小时未检查溢出。如果发生溢出，
  就会分配过小的缓冲区并被写越界。（CVE-2014-0064）

- 防止固定大小缓冲区越界（Peter Eisentraut、Jozef Mlich）

  使用 `strlcpy()` 及相关函数，以明确保证固定大小缓冲区不会被越界写入。与前面几项不同，这些情况是否真的构成现实问题尚不清楚，因为在大多数情况下，
  输入字符串的大小似乎已有先前约束。尽管如此，消除所有此类 Coverity 警告似乎是明智的。（CVE-2014-0065）

- 避免在 `crypt()` 返回 NULL 时崩溃（Honza Horak、Bruce Momjian）

  `crypt()` 可能返回 NULL 的场景相对较少，但如果发生，`contrib/chkpass` 就会崩溃。一个可能出问题的实际情况是 libc 被配置为拒绝执行未获批准的哈希算法（例如“FIPS 模式”）。（CVE-2014-0066）

- 在回归测试说明中记录 `make check` 的风险（Noah Misch、Tom Lane）

由于 `make check` 启动的临时服务器使用“trust”认证，同一台机器上的其他用户可以以数据库超级用户身份连接它，进而可能利用启动测试的操作系统用户的权限。
未来版本可能会在测试流程中加入防止此风险的变更，但需要先进行一些公开讨论。
因此目前只是警告：当同一台机器上存在不受信任的用户时，不要使用 `make check`。（CVE-2014-0067）

- 修复当关系的某些段不是完整大小时 WAL 记录可能被错误重放的问题（Greg Stark、Tom Lane）

  WAL 更新可能被应用到错误的页面，甚至可能偏后很多页。除了损坏数据外，已观察到这一错误导致备库相比主库显著“膨胀”，因为更新被应用到了远超文件末尾应在位置的地方。
  这种失败模式在崩溃恢复期间似乎不构成重大风险，
  只在初始同步从快速变化的主库基础备份创建的备库时出现。

- 确保对 GIN 索引非叶页面的插入在合适时写入整页 WAL 记录（Heikki Linnakangas）

  先前的代码在系统崩溃期间发生部分页写入时存在索引损坏风险。

- 修复服务器进程退出期间的竞态条件（Robert Haas）

  确保信号处理程序不会在进程的 `MyProc` 指针不再有效后尝试使用它。

- 修复错误报告逻辑中对 `errno` 的不安全引用（Christian Kruse）

  这通常会导致奇怪的行为，例如 `HINT` 字段缺失或不恰当。

- 修复服务器启动期间过早使用 `ereport()` 可能导致的崩溃（Tom Lane）

  我们在实际中见到的主要情况是：如果服务器在一个它没有读取权限的目录中被启动，
  就会崩溃。

- 在 OpenSSL 套接字写函数中正确清除重试标志（Alexander Kukushkin）

  这一遗漏可能导致 SSL 加密连接意外断开后服务器锁死。

- 修复对包含转义的 Unicode 标识符（`U&"..."` 语法）的长度检查（Tom Lane）

  如果标识符的转义形式过长，就会为这类标识符打印虚假的截断警告，
  但标识符在反转义之后实际上并不需要截断。

- 修复嵌套子选择（例如 `WHERE (... x IN (SELECT ...) ...) IN (SELECT ...)`）因无效计划而可能崩溃的问题（Tom Lane）

- 确保即使表列中的所有值都“过宽”，`ANALYZE` 也会为该列创建统计信息（Tom Lane）

  `ANALYZE` 有意在直方图和最常用值计算中省略非常宽的值，但在所有采样条目都过宽的情况下没有做出合理的处理。

- 在 `ALTER TABLE ... SET TABLESPACE` 中，允许不经权限检查就使用数据库的默认表空间（Stephen Frost）

  `CREATE TABLE` 一直允许这种用法，但 `ALTER TABLE` 没有照做。

- 修复当 `CASE` 的某些分支返回集合而另一些不返回时的“cannot accept a set”错误（Tom Lane）

- 修复 pgstat 函数中对全零客户端地址的检查（Kevin Grittner）

- 修复文本搜索解析器可能错误分类多字节字符的问题（Tom Lane）

  在 C 区域设置下使用多字节编码时，非 ASCII 字符可能被错误分类。在 Cygwin 上，非 C 区域设置也可能失败。

- 修复 `plainto_tsquery()` 中可能的不当行为（Heikki Linnakangas）

  在复制重叠内存区域时使用 `memmove()` 而不是 `memcpy()`。尚无实际造成问题的现场报告，但这无疑是有风险的。

- 在区域检查方面接受 `SHIFT_JIS` 作为编码名（Tatsuo Ishii）

- 修复 `PQhost()` 在 Windows 上的不当行为（Fujii Masao）

  如果未指定主机，它应返回 `localhost`。

- 改进 libpq 和 psql 对 `COPY TO STDOUT/FROM STDIN` 期间失败的错误处理（Tom Lane）

  这尤其修复了 9.2 及更高版本中，如果服务器连接在 `COPY FROM STDIN` 期间丢失可能发生的无限循环。该场景的变体在旧版本或其他客户端应用中也可能出现。

- 修复 ecpg 中未对齐的描述符（MauMau）

- 在 ecpg 中正确处理连接参数中缺少主机名的情况（Michael Meskes）

- 修复 `contrib/dblink` 连接启动中的性能退化（Joe Conway）

  在客户端和服务器编码一致时避免一次不必要的往返。

- 修复 `contrib/isn` 中 ISMN 值校验位计算错误的问题（Fabien Coelho）

- 确保仅安装客户端代码的流程按文档说明工作（Peter Eisentraut）

- 在 Mingw 和 Cygwin 构建中，把 libpq DLL 安装到 `bin` 目录（Andrew Dunstan）

  这与 MSVC 构建长期以来的做法一致。它应该能解决 psql 等程序因找不到 DLL 而无法启动的问题。

- 不再生成纯文本的 `HISTORY` 和 `src/test/regress/README` 文件（Tom Lane）

  这些文本文件与主文档的 HTML 和 PDF 格式重复。维护它们的麻烦远远超过纯文本格式可能的读者。发行版压缩包仍会包含这些名字的文件，但它们只是提示读者查阅主文档的存根。纯文本 `INSTALL` 文件仍将维护，因为它可以说还有使用场景。

- 把时区数据文件更新为 tzdata 版本 2013i for DST law changes in Jordan and historical changes in Cuba。

  此外，时区 `Asia/Riyadh87`、`Asia/Riyadh88` 和 `Asia/Riyadh89` 已被移除，因为 IANA 不再维护它们，而且它们从未代表过实际的民用计时实践。

E.53. 版本 8.4.19

发行日期
2013-12-05

本次发布包含对 8.4.18 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.53.1. 迁移到版本 8.4.19

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，此发行版纠正了一个潜在的数据损坏问题。请参阅下面第一条变更记录，以确定你的安装是否受到影响以及如果受到影响可以采取什么措施。

不过，如果你是从earlier than 8.4.17 的版本升级，请参见第 E.55 节。

E.53.2. 变更

- 修复 `VACUUM` 判断是否可以更新 `relfrozenxid` 的测试（Andres Freund）

  某些情况下，`VACUUM`（手工或自动）可能错误地推进表的 `relfrozenxid` 值，使元组逃过冻结，导致这些行在经过 2^31 个事务后变得不可见。由于需要多次错误推进才会发生实际丢失，数据丢失的概率相当低，但并非为零。从 8.4.8 或更早版本升级的用户不受影响，但之后的所有版本都包含该缺陷。

  升级后，在把 `vacuum_freeze_table_age` 设为零的情况下对所有数据库中的所有表执行清理，可以缓解此问题。这会修复任何潜伏的损坏，但无法修复所有既有的数据错误。不过，如果安装在其生命周期内执行的事务少于 2^31 个更新（可用 `SELECT txid_current() < 2^31` 检查），则在执行此清理之后即可认为该安装是安全的。

- 修复 GIN 索引 posting 树页面删除中的竞态条件（Heikki Linnakangas）

  这可能导致瞬时的错误结果或查询失败。

- 避免展平 `SELECT` 列表中包含包裹在子 `SELECT` 中的易变函数的子查询（Tom Lane）

  这避免了因易变函数被额外求值而导致的意外结果。

- 修复规划器对外连接中嵌套的非简单变量子查询输出的处理（Tom Lane）

  这一错误可能为涉及 `JOIN` 语法中多级子查询的查询生成错误的计划。

- 修复临时文件被过早删除的问题（Andres Freund）

- 修复规则打印中可能读取越过内存末尾的问题（Peter Eisentraut）

- 修复 `int2vector` 和 `oidvector` 值的数组切片（Tom Lane）

  这类表达式现在会被隐式提升为常规的 `int2` 或 `oid` 数组。

- 修复使用 SQL 标准的简单 GMT 偏移时区时的错误行为（Tom Lane）

  某些情况下，系统会使用简单 GMT 偏移值，而本应使用选择简单偏移之前一直生效的常规时区设置。此变更还使 `timeofday` 函数遵循简单 GMT 偏移时区。

- 防止记录 Windows 错误码翻译时可能出现的不当行为（Tom Lane）

- 正确地为 pg_ctl 生成的命令行加引号（Naoya Anzai、Tom Lane）

  此修复仅适用于 Windows。

- 修复当源数据库通过 `ALTER DATABASE SET` 设置 `default_transaction_read_only` 时 pg_dumpall 无法工作的问题（Kevin Grittner）

  此前，生成的脚本在恢复期间会失败。

- 修复 ecpg 对声明为 `varchar` 的变量列表的处理（Zoltán Böszörményi）

- 让 `contrib/lo` 防御不正确的触发器定义（Marc Cousin）

- 把时区数据文件更新为 tzdata 版本 2013h for DST law changes in Argentina, Brazil, Jordan, Libya, Liechtenstein, Morocco, and Palestine。Also, new timezone abbreviations WIB, WIT, WITA for Indonesia.

E.54. 版本 8.4.18

发行日期
2013-10-10

本次发布包含对 8.4.17 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.54.1. 迁移到版本 8.4.18

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.17 的版本升级，请参见第 E.55 节。

E.54.2. 变更

- 防止尝试对标识符做大小写折叠时损坏多字节字符（Andrew Dunstan）

  PostgreSQL 只在使用单字节服务器编码时才对非 ASCII 字符做大小写折叠。

- 修复由 `lo_open()` 失败导致的内存泄漏（Heikki Linnakangas）

- 修复 `work_mem` 使用超过 24GB 内存时的内存过度提交缺陷（Stephen Frost）

- 修复使用 SSL 时 libpq 中的死锁缺陷（Stephen Frost）

- 正确计算包含大量 NULL 值的布尔列的行数估计（Andrew Gierth）

  此前，`col IS NOT TRUE` 和 `col IS NOT FALSE` 这类测试在估计计划代价时没有正确考虑 NULL 值。

- 防止把 `WHERE` 子句下推到不安全的 `UNION/INTERSECT` 子查询中（Tom Lane）

  `SELECT` 列表中包含集合返回函数或易变函数的 `UNION` 或 `INTERSECT` 子查询可能被不当优化，导致运行时错误或错误的查询结果。

- 修复罕见的 “failed to locate grouping columns” 规划器失败情况（Tom Lane）

- 改进视图转储代码对被引用表中已删除列的处理（Tom Lane）

- 修复并发执行 `CREATE INDEX CONCURRENTLY` 操作时可能出现的死锁（Tom Lane）
- 修复 `regexp_matches()` 对零长度匹配的处理（Jeevan Chalke）

  此前，像 '^' 这样的零长度匹配可能返回过多的匹配。
- 修复过于复杂的正则表达式导致崩溃的问题（Heikki Linnakangas）
- 修复反向引用与非贪婪量词组合时正则表达式匹配失败的问题（Jeevan Chalke）
- 除非启用了函数体检查，否则防止 `CREATE FUNCTION` 检查 `SET` 变量（Tom Lane）
- 修复 `pgp_pub_decrypt()`，使其可用于带口令的密钥（Marko Kreen）
- 移除无索引表清理期间罕见的不准确警告（Heikki Linnakangas）
- 避免在预备查询中执行事务控制命令（例如 `ROLLBACK`）时可能出现的失败（Tom Lane）
- 确保浮点数据输入在所有平台上都接受标准的“infinity”（无穷大）写法（Tom Lane）

  C99 标准规定允许的写法为 `inf`、`+inf`、`-inf`、`infinity`、`+infinity` 和 `-infinity`。即使平台的 `strtod` 函数不支持，也确保我们能识别这些写法。
- 扩展行与记录和数组比较的能力（Rafal Rzepecki、Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2013d for DST law changes in Israel, Morocco, Palestine, and Paraguay。Also, historical zone data corrections for Macquarie Island.

E.55. 版本 8.4.17

发行日期
2013-04-04

本次发布包含对 8.4.16 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.55.1. 迁移到版本 8.4.17

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，本版本纠正了 GiST 索引管理中的若干错误。安装此更新后，建议对满足下述一个或多个条件的 GiST 索引执行 `REINDEX`。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.55.2. 变更

- 在每个 postmaster 子进程中重置 OpenSSL 随机状态（Marko Kreen）

  这避免了另一种情况：另一数据库用户可能相对容易地猜出 `contrib/pgcrypto` 函数生成的随机数。只有当 postmaster 配置了 `ssl = on` 而大多数连接并不使用 SSL 加密时，这一风险才比较显著。（CVE-2013-1900）

- 修复 GiST 索引在不合适的情况下使用“模糊”几何比较的问题（Alexander Korotkov）

  核心几何类型使用“模糊”相等进行比较，但 `gist_box_same` 必须做精确比较，否则使用它的 GiST 索引可能变得不一致。安装此更新后，用户应对 `box`、`polygon`、`circle` 或 `point` 列上的所有 GiST 索引执行 `REINDEX`，因为它们都使用 `gist_box_same`。

- 修复对可变宽度数据类型（即 `text`、`bytea`、`bit` 和 `numeric` 列）使用 `contrib/btree_gist` 的 GiST 索引中错误的范围并集和惩罚逻辑（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，还会造成无用的索引膨胀。建议用户在安装此更新后对这类索引执行 `REINDEX`。

- 修复多列索引的 GiST 页面分裂代码中的缺陷（Tom Lane）

  这些错误可能导致不一致的索引：某些存在的键搜索不到，还会使索引的搜索效率不必要地低下。建议用户在安装此更新后对多列 GiST 索引执行 `REINDEX`。

- 修复正则表达式编译中的无限循环风险（Tom Lane、Don Porter）

- 修复正则表达式编译中潜在的空指针解引用（Tom Lane）

- 修复 `to_char()`，在适当的地方只使用 ASCII 大小写折叠规则（Tom Lane）

  这修复了一些本应与区域设置无关的模板模式在土耳其区域下错误处理“`I`”和“`i`”的问题。

- 修复时间戳 `1999-12-31 24:00:00` 被不当拒绝的问题（Tom Lane）

- 移除无用的“picksplit doesn't support secondary split”日志消息（Josh Hansen、Tom Lane）

  这条消息似乎是为预期中从未编写、而且可能永远不会编写的代码而添加的，因为 GiST 对二次分裂的默认处理其实相当好。所以不再为此打扰最终用户。

- 修复会话最后几个事务的提交/中止计数可能未能发送到统计收集器的问题（Tom Lane）

- 消除 PL/Perl 的 `spi_prepare()` 函数中的内存泄漏（Alex Hunsaker、Tom Lane）

- 修复 pg_dumpall 对包含“`=`”的数据库名的处理（Heikki Linnakangas）

- 避免在给定不正确的连接串时 pg_dump 崩溃（Heikki Linnakangas）

- 让 pg_dump 忽略无效索引（Michael Paquier）

  转储无效索引可能在恢复时引起问题，例如当索引创建失败的原因是它试图强制执行表中数据不满足的唯一性条件时。另外，如果索引创建实际上仍在进行中，把它视为未提交的 DDL 变更是合理的，pg_dump 本来就不会转储它。

- 修复 `contrib/pg_trgm` 的 `similarity()` 函数，使其对无三元组的字符串返回零（Tom Lane）

  此前它由于内部除以零而返回 `NaN`。

- 把时区数据文件更新为 tzdata 版本 2013b for DST law changes in Chile, Haiti, Morocco, Paraguay, and some Russian areas。Also, historical zone data corrections for numerous places.

此外，更新时区缩写文件以反映俄罗斯和其他地方的近期变更：CHOT、GET、IRKT、KGT、KRAT、MAGT、MAWT、MSK、NOVT、OMST、TKT、VLAT、WST、YAKT、YEKT 现在遵循其当前含义，同时把 VOLT（Europe/Volgograd）和 MIST（Antarctica/Macquarie）加入默认缩写列表。

E.56. 版本 8.4.16

发行日期
2013-02-07

本次发布包含对 8.4.15 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.56.1. 迁移到版本 8.4.16

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.56.2. 变更

- 防止从 SQL 执行 enum_recv（Tom Lane）

  该函数声明有误，允许一条简单的 SQL 命令使服务器崩溃。原则上攻击者可能利用它检查服务器内存的内容。感谢 Sumit Soni（通过 Secunia SVCRP）报告此问题。（CVE-2013-0255）

- 在截断关系文件时更新最小恢复点（Heikki Linnakangas）

  数据一旦被丢弃，在时间线上更早的点停止恢复就不再安全。

- 修复 SQL 文法，允许对子 SELECT 结果做下标或字段选择（Tom Lane）

- 防止扫描 pg_tablespace 时的竞态条件（Stephen Frost、Tom Lane）

  如果 pg_tablespace 条目存在并发更新，CREATE DATABASE 和 DROP DATABASE 可能行为异常。

- 防止 DROP OWNED 试图删除整个数据库或表空间（Álvaro Herrera）

  为安全起见，这些对象的所有权必须重新分配，而不是删除。

- 修复 vacuum_freeze_table_age 实现中的错误（Andres Freund）

  在已存在超过 vacuum_freeze_min_age 个事务的安装中，这一失误使自动清理无法使用部分表扫描，而总是进行全表扫描。

- 防止 RowExpr 或 XmlExpr 被二次解析分析时的不当行为（Andres Freund、Tom Lane）

  这一失误在 CREATE TABLE LIKE INCLUDING INDEXES 等上下文中可能被用户看到。

- 增强对哈希表大小计算中整数溢出的防御（Jeff Davis）

- 拒绝 to_date() 中超出范围的日期（Hitoshi Harada）

- 确保在 Windows 上非 ASCII 提示符字符串被转换到正确的代码页（Alexander Law、Noah Misch）

  这一缺陷影响 psql 和其他一些客户端程序。

- 修复未连接数据库时 psql 的 `\?` 命令可能崩溃的问题（Meng Qingzhong）

- 修复 libpq 的 `PQprintTuples` 中的一字节缓冲区越界（Xi Wang）

  这个古老的函数没有被 PostgreSQL 本身在任何地方使用，但仍可能被某些客户端代码使用。

- 让 ecpglib 正确使用翻译后的消息（Chen Huajun）

- 在 MSVC 上正确安装 ecpg_compat 和 pgtypes 库（Jiang Guiqing）

- 重新排列 configure 对系统提供函数的测试，使其不被 libedit/libreadline 的虚假导出所欺骗（Christoph Berg）

- 确保 Windows 构建号随时间递增（Magnus Hagander）

- 让 pgxs 在为 Windows 交叉编译时以正确的 `.exe` 后缀构建可执行文件（Zoltan Boszormenyi）

- 新增时区缩写 `FET`（Tom Lane）

  它现在用于一些东欧时区。

E.57. 版本 8.4.15

发行日期
2012-12-06

本次发布包含对 8.4.14 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.57.1. 迁移到版本 8.4.15

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.57.2. 变更

- 修复与 `CREATE INDEX CONCURRENTLY` 相关的多个缺陷（Andres Freund、Tom Lane）

  修复 `CREATE INDEX CONCURRENTLY`，使其在更改索引 `pg_index` 行的状态时使用就地更新。这防止了可能导致并发会话错过更新目标索引（从而产生损坏的并发创建索引）的竞态条件。

  另外，修复各种其他操作，确保它们忽略由失败的 `CREATE INDEX CONCURRENTLY` 命令产生的无效索引。其中最重要的是 `VACUUM`，因为在采取纠正措施修复或删除无效索引之前，很容易就有自动清理在该表上启动。

- 避免内存不足时内部哈希表损坏（Hitoshi Harada）
- 修复外连接之上非严格等值子句的规划（Tom Lane）

  规划器可能从把非严格构造与别的值等同的子句中推导出错误的约束，例如当 `foo` 来自外连接的可空侧时的 `WHERE COALESCE(foo, 0) = 0`。
- 改进规划器从等价类证明排他约束的能力（Tom Lane）
- 修复哈希子计划中的部分行匹配，使其正确处理跨类型情况（Tom Lane）

  这影响多列 `NOT IN` 子计划，例如 `WHERE (a, b) NOT IN (SELECT x, y FROM ...)`，当 `b` 和 `y` 分别是 `int4` 和 `int8` 时。根据所涉及的具体数据类型，这一失误会导致错误答案或崩溃。
- 在为 `AFTER ROW UPDATE/DELETE` 触发器重新获取旧元组时获取缓冲区锁（Andres Freund）

  在非常罕见的情况下，这一疏忽可能导致向外键强制触发器的预检查逻辑传递不正确的数据。这可能导致崩溃，或者在是否触发触发器上做出错误决定。
- 修复`ALTER COLUMN TYPE`以正确重建继承的检查约束（Tom Lane）

  这在 8.4 之前的版本中工作正常，现在在 8.4 及更高版本中也工作正常。
- 修复 `REASSIGN OWNED` 对表空间授权的处理（Álvaro Herrera）
- 忽略视图中系统列的不正确 `pg_attribute` 条目（Tom Lane）

  视图没有任何系统列。然而，我们在把表转换为视图时忘记移除这类条目。这一问题在 9.3 及更高版本中已正确修复，但在以前的分支中，我们需要防御已经错误转换的视图。
- 修复规则打印，使其正确转储 `INSERT INTO *table* DEFAULT VALUES`（Tom Lane）
- 防止当查询中有过多 `UNION/INTERSECT/EXCEPT` 子句时栈溢出（Tom Lane）
- 防止用最小可能整数值除以 -1 时依赖平台的失败（Xi Wang、Tom Lane）
- 修复日期解析中可能越过字符串末尾的访问（Hitoshi Harada）
- 当 Unix 域套接字的路径名长度超过平台特定限制时，生成可理解的错误消息（Tom Lane、Andrew Dunstan）

  此前这会导致相当无用的结果，例如“Non-recoverable failure in name resolution”。
- 修复向客户端发送组合列值时的内存泄漏（Tom Lane）
- 使 pg_ctl 读取 `postmaster.pid` 文件更健壮（Heikki Linnakangas）

  这修复了竞态条件和可能的文件描述符泄漏。
- 修复当呈现错误编码的数据且 `client_encoding` 设置为仅客户端编码（如 SJIS）时 psql 可能崩溃的问题（Jiang Guiqing）
- 修复 pg_dump 以 `tar` 输出格式生成的 `restore.sql` 脚本中的缺陷（Tom Lane）

  该脚本在表名包含大写字符时会直接失败。此外，让该脚本既能以常规 COPY 模式也能以 `--inserts` 模式恢复数据。

- 修复 pg_restore，使其接受符合 POSIX 的 `tar` 文件（Brian Weaver、Tom Lane）

  pg_dump 的 `tar` 输出模式的原始编码生成的文件并不完全符合 POSIX 标准。这在 9.3 版本中已得到纠正。此补丁更新了以前的分支，使它们同时接受错误格式和纠正后的格式，希望借此避免 9.3 发布时的兼容性问题。

- 修复 pg_resetxlog，使其在给定数据目录相对路径时正确定位 `postmaster.pid`（Tom Lane）

  这一失误可能导致 pg_resetxlog 没有注意到有活动 postmaster 正在使用该数据目录。

- 修复 libpq 的 `lo_import()` 和 `lo_export()` 函数，使其正确报告文件 I/O 错误（Tom Lane）

- 修复 ecpg 对嵌套结构指针变量的处理（Muhammad Usama）

- 让 `contrib/pageinspect` 的 btree 页面检查函数在检查页面时持有缓冲区锁（Tom Lane）

- 修复 pgxs 在 AIX 上构建可加载模块的支持（Tom Lane）

  在原始源码树之外构建模块在 AIX 上不工作。

- 把时区数据文件更新为 tzdata 版本 2012j for DST law changes in Cuba, Israel, Jordan, Libya, Palestine, Western Samoa, and portions of Brazil。

E.58. 版本 8.4.14

发行日期
2012-09-24

本次发布包含对 8.4.13 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.58.1. 迁移到版本 8.4.14

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.58.2. 变更

- 修复规划器对执行器参数的分配，并修复执行器对 CTE 计划节点的重新扫描逻辑（Tom Lane）

  这些错误可能导致多次扫描同一 `WITH` 子查询的查询给出错误答案。

- 改进 GiST 索引的页面分裂决策（Alexander Korotkov、Robert Haas、Tom Lane）

  多列 GiST 索引可能因此错误而遭受意外膨胀。

- 修复级联权限撤销，使其在权限仍被持有时停止（Tom Lane）

  如果我们撤销某个角色 *X* 的授予选项，但 *X* 仍通过别人的授予持有该选项，我们就不应从 *X* 曾授予该权限的角色 *Y* 那里递归撤销相应权限。

- 修复使用 PL/Perl 时对 `SIGFPE` 的处理（Andres Freund）

  Perl 会把进程的 `SIGFPE` 处理程序重置为 `SIG_IGN`，这可能导致之后崩溃。在初始化 PL/Perl 之后恢复正常的 Postgres 信号处理程序。

- 防止递归 PL/Perl 函数在执行期间被重定义时 PL/Perl 崩溃（Tom Lane）

- 规避 PL/Perl 中可能的错误优化（Tom Lane）

  一些 Linux 发行版包含错误版本的 `pthread.h`，导致 PL/Perl 中的编译代码不正确，当一个 PL/Perl 函数调用另一个会抛出错误的函数时会导致崩溃。

- 把时区数据文件更新为 tzdata 版本 2014e，以反映克里米亚、埃及和摩洛哥的夏令时法规变更。

E.59. 版本 8.4.13

发行日期
2012-08-17

本次发布包含对 8.4.12 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.59.1. 迁移到版本 8.4.13

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.59.2. 变更

- 防止通过 XML 实体引用访问外部文件/URL（Noah Misch、Tom Lane）

  `xml_parse()` 会尝试按需获取外部文件或 URL 来解析 XML 值中的 DTD 和实体引用，从而使无特权数据库用户能够尝试以数据库服务器的权限获取数据。虽然外部数据不会直接返回给用户，但如果数据不能解析为有效 XML，其部分内容可能暴露在错误消息中；而且无论如何，仅仅能够检查文件是否存在对攻击者就可能有用。（CVE-2012-3489）

- 防止通过 `contrib/xml2` 的 `xslt_process()` 访问外部文件/URL（Peter Eisentraut）

  libxslt 提供通过样式表命令读写文件和 URL 的能力，从而使无特权数据库用户能够以数据库服务器的权限读写数据。现通过正确使用 libxslt 的安全选项禁用该能力。（CVE-2012-3488）

  此外，移除了 `xslt_process()` 从外部文件/URL 获取文档和样式表的能力。虽然这是一个有记载的“特性”，但长期以来被认为是个坏主意。CVE-2012-3489 的修复破坏了该能力，与其花费精力修复它，我们干脆将其移除。

- 防止 btree 索引页被过早回收（Noah Misch）

  当我们允许只读事务跳过分配 XID 时，引入了已删除 btree 页在只读事务仍在访问它时就被回收的可能性。这会导致错误的索引搜索结果。由于时序要求，
  这种错误在实际中发生的概率似乎很低，但仍应修复。

- 修复新创建或重置序列的崩溃安全缺陷（Tom Lane）

  如果在新创建或重置的序列上执行了 `ALTER SEQUENCE`，然后恰好对它调用了一次 `nextval()`，随后服务器崩溃，WAL 重放会把序列恢复到看起来没有执行过 `nextval()` 的状态，从而使下一个 `nextval()` 调用再次返回第一个序列值。这种情况尤其可能出现在 `serial` 列上，因为创建 serial 列的序列包含一个 `ALTER SEQUENCE OWNED BY` 步骤。

- 确保 `pg_start_backup()` 之后对 `backup_label` 文件执行 fsync（Dave Kerr）

- 回移植压缩 fsync 请求队列的 9.1 改进（Robert Haas）

  这改进了检查点期间的性能。9.1 的变更已经过足够的实际测试，看起来可以安全回移植。

- 只允许被直接阻塞的进程自动取消自动清理（Tom Lane）

  原始编码在某些情况下可能允许不一致的行为；特别是自动清理可能在不足 `deadlock_timeout` 宽限期之后就被取消。

- 改进自动清理取消的日志记录（Robert Haas）

- 修复日志收集器，使 `log_truncate_on_rotation` 在服务器启动后的第一次日志轮转期间就能工作（Tom Lane）

- 修复附加到嵌套集合操作（`UNION/INTERSECT/EXCEPT`）上的 `WITH`（Tom Lane）

- 确保对子查询的整行引用不包含任何多余的 `GROUP BY` 或 `ORDER BY` 列（Tom Lane）

- 禁止在 `CREATE TABLE` 期间复制 `CHECK` 约束和索引定义中的整行引用（Tom Lane）

  这种情况可能出现在带 `LIKE` 或 `INHERITS` 的 `CREATE TABLE` 中。被复制的整行变量被错误地标记为原表而不是新表的行类型。对 `LIKE` 拒绝这种情况似乎是合理的，因为行类型之后很可能分化。对 `INHERITS`，理想情况下应该允许它，并隐式强制转换为父表的行类型；但这需要的工作量超出了回移植的安全范围。

- 修复 `ARRAY(SELECT ...)` 子查询中的内存泄漏（Heikki Linnakangas、Tom Lane）

- 修复从正则表达式中提取公共前缀的问题（Tom Lane）

  代码可能被带量词的带括号子表达式（例如 `^(foo)?bar`）混淆。这会导致对此类模式搜索的索引优化不正确。

- 修复解析 `interval` 常量中带符号的 *hh*:*mm* 和 *hh*:*mm*:*ss* 字段的缺陷（Amit Kapila、Tom Lane）

- 修复 `contrib/xml2` 的 `xslt_process()` 中的错误报告（Tom Lane）

- 把时区数据文件更新为 tzdata 版本 2012e，以反映摩洛哥和托克劳的夏令时法规变更

E.60. 版本 8.4.12

发行日期
2012-06-04

本次发布包含对 8.4.11 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.60.1. 迁移到版本 8.4.12

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.60.2. 变更

- 修复 `contrib/pgcrypto` 的 DES `crypt()` 函数中不正确的口令变换（Solar Designer）

  如果口令串包含字节值 `0x80`，口令的其余部分会被忽略，使口令比看起来弱得多。修复后，字符串的其余部分会被正确纳入 DES 哈希。因此，受此缺陷影响的任何已存储口令值将不再匹配，所以可能需要更新已存储的值。（CVE-2012-2143）

- 忽略过程语言调用处理程序的 `SECURITY DEFINER` 和 `SET` 属性（Tom Lane）

  把这类属性应用于调用处理程序可能导致服务器崩溃。（CVE-2012-2655）

- 允许 `timestamp` 输入中的数字时区偏移距 UTC 最多 16 小时（Tom Lane）

  一些历史时区的偏移超过此前 15 小时的限制。这可能导致转储的数据值在重载期间被拒绝。

- 修复时间戳转换，
  使其在给定时间恰好是当前时区最后一次夏令时转换时间时能正确处理（Tom Lane）

  这一疏忽已存在很长时间，但此前未被注意到，
  因为大多数使用夏令时的时区被假定有无限的未来夏令时转换序列。

- 修复 `text` 到 `name` 以及 `char` 到 `name` 的转换，使其在多字节编码下正确执行字符串截断（Karl Schnaitter）

- 修复 `to_tsquery()` 中的内存复制缺陷（Heikki Linnakangas）

- 修复规划器对子查询内外层 PlaceHolderVar 的处理（Tom Lane）

  该缺陷涉及引用来自外层查询外连接可空侧变量的子 SELECT。在 9.1 中，受此缺陷影响的查询会报 “ERROR: Upper-level PlaceHolderVar found where not expected” 失败。但在 9.0 和 8.4 中，你会静默地得到可能错误的答案，因为传入子查询的值不会在该为空时变为空。

- 修复 `pg_attribute` 非常大会话启动缓慢的问题（Tom Lane）

  如果 `pg_attribute` 超过 `shared_buffers` 的四分之一，会话启动期间有时需要的缓存重建代码就会触发同步扫描逻辑，使其耗时比正常情况多数倍。如果许多新会话同时启动，
  问题尤为严重。

- 确保顺序扫描以合理的频率检查查询取消（Merlin Moncure）

  遇到许多不含存活元组的连续页面的扫描在此期间不会响应中断。

- 确保 `PGSemaphoreLock()` 的 Windows 实现在返回前清除 `ImmediateInterruptOK`（Tom Lane）

  这一疏忽意味着同一查询稍后收到的查询取消中断可能在不安全的时刻被接受，
  后果不可预测且不好。

- 在打印视图或规则时安全地显示整行变量（Abbas Butt、Tom Lane）

  涉及有歧义名称（即该名称既可能是查询的表名也可能是列名）的极端情况此前以有歧义的方式打印，存在视图或规则在转储和重载后被不同解释的风险。现通过附加无操作转换来避免歧义情况。

- 修复 `COPY FROM`，使其正确处理对应无效编码的空值标记串（Tom Lane）

  诸如 `E'\\0'` 这样的空值标记串本应有效，过去也确实有效，但该情况在 8.4 中被破坏。

- 确保自动清理工作进程正确执行栈深度检查（Heikki Linnakangas）

  此前，由自动 `ANALYZE` 调用的函数中的无限递归可能导致工作进程崩溃。

- 修复日志收集器，使其在高负载下不丢失日志连贯性（Andrew Dunstan）

  此前，如果收集器过于忙碌，可能无法重新组装大消息。

- 修复日志收集器，确保它在收到 SIGHUP 后会重新开始文件轮转（Tom Lane）
- 修复 GIN 索引的 WAL 重放逻辑，使其在索引随后被删除时不失败（Tom Lane）
- 修复 PL/pgSQL 的 `RETURN NEXT` 命令中的内存泄漏（Joe Conway）
- 修复 PL/pgSQL 的 `GET DIAGNOSTICS` 命令在目标是函数第一个变量时的问题（Tom Lane）
- 修复 psql 扩展显示（`\x`）模式中可能越过内存末尾的访问（Peter Eisentraut）
- 修复数据库包含大量对象时 pg_dump 的若干性能问题（Jeff Janes、Tom Lane）

  如果数据库包含许多模式、许多对象处于依赖循环中、或有许多被拥有的序列，pg_dump 可能变得非常慢。

- 修复 `contrib/dblink` 的 `dblink_exec()`，使其在出错时不泄漏临时数据库连接（Tom Lane）
- 修复 `contrib/dblink`，使其在错误消息中报告正确的连接名（Kyotaro Horiguchi）
- 把时区数据文件更新为 tzdata 版本 2012c for DST law changes in Antarctica, Armenia, Chile, Cuba, Falkland Islands, Gaza, Haiti, Hebron, Morocco, Syria, and Tokelau Islands; also historical corrections for Canada。

E.61. 版本 8.4.11

发行日期
2012-02-27

本次发布包含对 8.4.10 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.61.1. 迁移到版本 8.4.11

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.10 的版本升级，请参见第 E.62 节。

E.61.2. 变更

- 让 `CREATE TRIGGER` 要求对触发器函数有执行权限（Robert Haas）

  这一缺失的检查可能允许另一用户通过把触发器函数安装到自己拥有的表上，以伪造的输入数据执行该触发器函数。这只对标记为 `SECURITY DEFINER` 的触发器函数有影响，因为其他情况下触发器函数本来就以表所有者身份运行。（CVE-2012-0866）

- 移除 SSL 证书中通用名长度的任意限制（Heikki Linnakangas）

  libpq 和服务器此前都会把从 SSL 证书提取的通用名截断到 32 字节。通常这最坏只会造成意外的验证失败，但在一些相当难以置信的场景下，它可能允许一个证书持有者冒充另一个。受害者必须恰好有 32 字节长的通用名，而攻击者必须说服受信任的 CA 颁发一个通用名以该字符串为前缀的证书。冒充服务器还需要额外的攻击手段来重定向客户端连接。（CVE-2012-0867）

- 把 pg_dump 注释中写入的名称里的换行符转换为空格（Robert Haas）

  pg_dump 此前对其输出脚本中 SQL 注释内写出的对象名清理不慎。包含换行符的名称至少会使脚本语法错误。恶意构造的对象名可能在脚本重载时构成 SQL 注入风险。（CVE-2012-0868）

- 修复插入与清理并发导致 btree 索引损坏的问题（Tom Lane）

  插入导致的索引页分裂有时会使并发运行的 `VACUUM` 漏删本应删除的索引项。在相应的表行被删除后，悬空的索引项会导致错误（例如 “could not read block N in file ...” ），或者更糟：当不相关的行被重新插入到现已空闲的表位置后，查询会静默地得到错误结果。该缺陷自 8.2 起就存在，但发生频率极低，直到现在才被诊断出来。如果你有理由怀疑它曾在你的数据库中发生，对受影响的索引重建索引即可修复。

- 更改表所有者时，不仅更新表级权限，也更新逐列权限（Tom Lane）

  未做到这一点意味着，此前授予的所有列权限仍显示为由旧所有者授予。这导致新所有者和超级用户都无法撤销这些现在无法追溯到表所有者的权限。

- 允许 `ALTER USER/DATABASE SET` 中某些设置取不存在的值（Heikki Linnakangas）

  允许把 `default_text_search_config`、`default_tablespace` 和 `temp_tablespaces` 设为未知的名称。这是因为设置要在其中使用的另一个数据库中它们可能是已知的；对表空间的情况，则是因为表空间可能尚未创建。此前 `search_path` 已被认识到有同样的问题，这些设置现在的行为与它一致。

- 避免在提交后删除表文件出现问题时崩溃（Tom Lane）

  删除表应只在事务提交后才删除底层磁盘文件。如果那时失败（例如因为文件权限错误），代码本应只发出警告消息并继续，因为此时中止事务已为时过晚。这一逻辑从 8.4 起被破坏，导致这类情况引发 PANIC 和无法重启的数据库。

- 在 WAL 重放期间正确跟踪 OID 计数器，即使它发生回卷（Tom Lane）

  此前 OID 计数器会一直停留在高值，直到系统退出重放模式。其实际后果通常为零，但在某些场景下，已提升为主库的备库服务器在需要 OID 值时，可能要很长时间才能把 OID 计数器推进到合理的值。

- 修复带 `*` 的正则表达式反向引用（Tom Lane）

代码实际上并不强制精确字符串匹配，
而是接受任何满足反向引用符号所引用的模式子表达式的字符串。

类似问题仍困扰着嵌入在更大带量词表达式中（而非量词直接作用对象）的反向引用。
这将在未来的 PostgreSQL 版本中解决。

- 修复最近引入的 `inet/cidr` 值处理中的内存泄漏（Heikki Linnakangas）

  PostgreSQL 2011 年 12 月版本中的一个补丁导致这些操作中的内存泄漏，在此类列上构建 btree 索引等场景中可能相当显著。

- 修复 SQL 语言函数中 `CREATE TABLE AS/SELECT INTO` 之后的悬空指针（Tom Lane）

  大多数情况下这只会在启用断言的构建中导致断言失败，但似乎可能有更糟的后果。

- 避免 Windows 上 syslogger 中文件句柄的双重关闭（MauMau）

  通常这个错误不可见，但在调试版 Windows 上运行时会导致异常。

- 修复 plpgsql 中与 I/O 转换相关的内存泄漏（Andres Freund、Jan Urbanski、Tom Lane）

  某些操作会一直泄漏内存直到当前函数结束。

- 改进 pg_dump 对继承表列的处理（Tom Lane）

  pg_dump 错误处理了子列与父列默认表达式不同的情况。如果默认值在文本上与父列相同但实际不同（例如因为模式搜索路径差异），它不会被识别为不同，
  因此转储和恢复后子列会被允许继承父列的默认值。子列为 `NOT NULL` 而父列不是的情况也可能被微妙地错误恢复。

- 修复 pg_restore 直接写数据库模式对 INSERT 式表数据的处理（Tom Lane）

  当使用 2011 年 9 月或 12 月版本的 pg_restore 时，对以 `--inserts` 或 `--column-inserts` 选项制作的归档文件直接恢复到数据库会失败，这是对另一个问题的修复中的一个疏忽所致。归档文件本身没有问题，文本模式输出也正常。

- 允许 ecpg 的 `DEALLOCATE` 语句中使用 `AT` 选项（Michael Meskes）

  支持它的基础设施已存在一段时间，但由于疏忽仍有一个错误检查拒绝这种情况。

- 修复 `contrib/intarray` 的 `int[] & int[]` 操作符中的错误（Guillaume Lelarge）

  如果两个输入数组共有的最小整数是 1，且任一数组中有更小的值，那么 1 会被错误地从结果中遗漏。

- 修复 `contrib/pgcrypto` 的 `encrypt_iv()` 和 `decrypt_iv()` 中的错误检测（Marko Kreen）

  这些函数未能报告某些类型的无效输入错误，而是对错误输入返回随机垃圾值。

- 修复 `contrib/test_parser` 中的一字节缓冲区越界（Paul Guyot）

  代码会尝试多读一个字节，这在极端情况下会崩溃。由于 `contrib/test_parser` 只是示例代码，这本身不是安全问题，但坏的示例代码仍然是坏的。

- 在 ARM 上如果可用则使用 `__sync_lock_test_and_set()` 实现自旋锁（Martin Pitt）

该函数取代了我们此前对 `SWPB` 指令的使用，该指令已被弃用且在 ARMv6 及更高版本上不可用。报告表明旧代码在较新的 ARM 板上不会以明显方式失败，而只是不对并发访问互锁，导致多进程操作中的离奇失败。

- 在使用接受该选项的 gcc 版本构建时使用 `-fexcess-precision=standard` 选项（Andrew Dunstan）

  这防止了较新版本 gcc 产生"创造性"结果的各种场景。

- 允许在 FreeBSD 上使用线程化 Python（Chris Rees）

  我们的 configure 脚本此前认为这种组合无法工作；但 FreeBSD 修复了问题，因此移除该错误检查。

E.62. 版本 8.4.10

发行日期
2011-12-05

本次发布包含对 8.4.9 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.62.1. 迁移到版本 8.4.10

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，`information_schema.referential_constraints` 视图的定义中被发现存在一个长期错误。如果你依赖该视图的正确结果，应按下面第一条变更日志条目的说明替换其定义。

不过，如果你是从earlier than 8.4.8 的版本升级，请参见第 E.64 节。

E.62.2. 变更

- 修复 `information_schema.referential_constraints` 视图中的缺陷（Tom Lane）

  该视图在把外键约束与它依赖的主键或唯一键约束匹配时不够谨慎。
  这可能导致某个外键约束完全不被显示、被显示多次，
  或者声称它依赖于一个并非其实际依赖的约束。

  由于该视图定义是由 initdb 安装的，仅靠升级不会修复问题。如果需要修复现有安装，可以（以超级用户身份）删除 `information_schema` 模式，然后通过载入 *SHAREDIR*`/information_schema.sql` 重新创建它。（如果不确定 *SHAREDIR* 在哪里，运行 `pg_config --sharedir`。）这必须在每个要修复的数据库中重复执行。

- 修复 GIN 索引更新 WAL 记录的错误重放（Tom Lane）

  这可能导致崩溃后或热备服务器上瞬时找不到索引项。不过，该问题会在该索引下一次 `VACUUM` 时被修复。

- 修复 `CREATE TABLE dest AS SELECT * FROM src` 或 `INSERT INTO dest SELECT * FROM src` 期间与 TOAST 相关的数据损坏（Tom Lane）

如果表曾被 `ALTER TABLE ADD COLUMN` 修改过，在某些极端情况下，尝试把它的数据原样复制到另一个表可能产生损坏的结果。该问题只会在 8.4 及更高版本中以这种确切形式出现，但我们也给更早的版本打了补丁，以防有其他代码路径可能触发同一缺陷。

- 修复从过期 syscache 条目访问 toast 表时的竞态条件（Tom Lane）

  典型症状是诸如 “missing chunk number 0 for toast value NNNNN in pg_toast_2619” 的瞬时错误，其中引用的 toast 表总是属于某个系统目录。

- 跟踪函数对参数默认值表达式中所用项的依赖（Tom Lane）

  此前，被引用的对象可以在未删除或修改函数的情况下被删除，导致函数被使用时行为异常。注意，仅仅安装此更新不会修复缺失的依赖条目；要做到这一点，之后需要对每个这样的函数执行 `CREATE OR REPLACE`。如果你的函数默认值依赖于非内建对象，建议这样做。

- 允许内联带多个 OUT 参数的集合返回 SQL 函数（Tom Lane）

- 让 `DatumGetInetP()` 解包带 1 字节头的 inet datum，并新增一个不做此事的宏 `DatumGetInetPP()`（Heikki Linnakangas）

  此变更不影响任何核心代码，但可能避免期望 `DatumGetInetP()` 按常规约定产生已解包 datum 的附加代码中的崩溃。

- 改进 `money` 类型输入和输出的区域支持（Tom Lane）

  除了不支持所有标准 `lc_monetary` 格式化选项外，输入和输出函数还不一致，这意味着在某些区域设置下转储的 `money` 值无法重新读取。

- 不让 `transform_null_equals` 影响 `CASE foo WHEN NULL ...` 构造（Heikki Linnakangas）

  `transform_null_equals` 只应影响用户直接写的 `foo = NULL` 表达式，而不是这种形式的 `CASE` 在内部生成的相等性检查。

- 更改外键触发器的创建顺序，以更好地支持自引用外键（Tom Lane）

  对于引用自身表的级联外键，行更新会作为一个事件同时触发 `ON UPDATE` 触发器和 `CHECK` 触发器。`ON UPDATE` 触发器必须先执行，否则 `CHECK` 会检查行的非最终状态，并可能抛出不恰当的错误。然而这些触发器的触发顺序由其名称决定，而由于触发器使用遵循 “RI_ConstraintTrigger_NNNN” 约定的自动生成名称，它们通常按创建顺序排序。彻底的修复需要修改该约定（我们将在 9.2 中做），但在现有版本中更改它似乎有风险。因此此补丁只更改触发器的创建顺序。遇到这类错误的用户应删除并重新创建外键约束，使其触发器处于正确顺序。

- 避免在跟踪缓冲区分配率时发生浮点下溢（Greg Matthews）

  虽然这本身无害，但在某些平台上会导致烦人的内核日志消息。

- 在 Windows 下启动子进程时保留配置文件名和行号值（Tom Lane）

  此前它们不会在 `pg_settings` 视图中正确显示。

- 在 psql 的命令历史中保留命令内的空行（Robert Haas）

  例如，如果从字符串字面量中删除了空行，先前的行为可能导致问题。

- 修复 pg_dump，使其转储自动生成类型（如表行类型）之间的用户定义转换（Tom Lane）
- 使用首选版本的 xsubpp 构建 PL/Perl，而不一定是操作系统的主副本（David Wheeler、Alex Hunsaker）
- 修复 `contrib/dict_int` 和 `contrib/dict_xsyn` 中的错误编码（Tom Lane）

  一些函数错误地假定 `palloc()` 返回的内存保证被清零。
- 让 `pgstatindex()` 及时响应查询取消中断（Robert Haas）
- 确保 VPATH 构建正确安装所有服务器头文件（Peter Eisentraut）
- 缩短详细错误消息中报告的文件名（Peter Eisentraut）

  常规构建一直只报告包含错误消息调用的 C 文件的名称，但 VPATH 构建此前报告绝对路径名。
- 修复中美洲 Windows 时区名的解释（Tom Lane）

  把“Central America Standard Time”映射为 `CST6` 而不是 `CST6CDT`，因为中美洲各地通常不实行夏令时。
- 把时区数据文件更新为 tzdata 版本 2011n for DST law changes in Brazil, Cuba, Fiji, Palestine, Russia, and Samoa; also historical corrections for Alaska and British East Africa。

E.63. 版本 8.4.9

发行日期
2011-09-26

本次发布包含对 8.4.8 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.63.1. 迁移到版本 8.4.9

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.4.8 的版本升级，请参见第 E.64 节。

E.63.2. 变更

- 修复对不确定的 HOT 更新元组建立索引的缺陷（Tom Lane）

  这些缺陷可能导致重建系统目录索引之后索引损坏。据信它们不影响用户索引。
- 修复 GiST 索引页面分裂处理中的多个缺陷（Heikki Linnakangas）

  发生概率很低，但它们可能导致索引损坏。
- 修复 `tsvector_concat()` 中可能的缓冲区越界（Tom Lane）

  该函数可能低估其结果所需的内存量，导致服务器崩溃。

- 修复 `xml_recv` 处理 “standalone” 参数时的崩溃（Tom Lane）
- 让 `pg_options_to_table` 对没有值的选项返回 NULL（Tom Lane）

  此前这类情况会导致服务器崩溃。
- 避免在 `ANALYZE` 和 SJIS-2004 编码转换中可能越过内存末尾的访问（Noah Misch）

  这修复了一些概率极低的服务器崩溃场景。
- 防止启动进程与 bgwriter 进程交互时的间歇性挂起（Simon Riggs）

  这影响非热备情况下的恢复。
- 修复 relcache init 文件失效中的竞态条件（Tom Lane）

  存在一个时间窗口，新的后端进程可能读到过期的 init 文件，却错过本应告知它数据已过期的失效消息。结果是目录访问中的离奇失败，通常在启动稍后阶段报 “could not read block 0 in file ...” 。
- 修复 GiST 索引扫描结束时的内存泄漏（Tom Lane）

  执行许多独立 GiST 索引扫描的命令（例如对已包含许多行的表验证新的基于 GiST 的排他约束）会因此泄漏瞬时需要大量内存。
- 修复支持可持有游标和 plpgsql 的 `RETURN NEXT` 命令的 tuplestore 中不正确的内存核算（可能导致内存膨胀）（Tom Lane）
- 修复构造大型有损位图时的性能问题（Tom Lane）
- 修复唯一列的连接选择率估计（Tom Lane）

  这修复了一个可能错误估计连接结果大小的规划器启发式规则。
- 修复只出现在子 SELECT 目标列表中的嵌套 PlaceHolderVar 表达式（Tom Lane）

  这一失误可能导致外连接的输出错误地显示为 NULL。
- 允许嵌套的 `EXISTS` 查询被正确优化（Tom Lane）
- 修复数组和路径创建函数，确保填充字节为零（Tom Lane）

  这避免了一些规划器认为语义相等的常量不相等的情况，这种情况会导致糟糕的优化。
- 修复 `EXPLAIN`，使其处理内层索引扫描子计划中的门控 Result 节点（Tom Lane）

  这一疏忽的常见症状是 “bogus varno” 错误。
- 规避破坏 WAL 重放的 gcc 4.6.0 缺陷（Tom Lane）

  这可能导致服务器崩溃后已提交事务的丢失。
- 修复视图中 `VALUES` 的转储缺陷（Tom Lane）
- 禁止对序列执行 `SELECT FOR UPDATE/SHARE`（Tom Lane）

  这种操作不像预期那样工作，可能导致失败。
- 修复 `VACUUM`，使其总是更新 `pg_class.reltuples/relpages`（Tom Lane）

这修复了一些自动清理可能对何时清理表做出越来越差决策的场景。

- 防御计算哈希表大小时的整数溢出（Tom Lane）
- 修复 `CLUSTER` 可能尝试访问已移除的 TOAST 数据的情况（Tom Lane）
- 修复“peer”认证使用凭据控制消息时的可移植性缺陷（Tom Lane）
- 修复需要多次往返时的 SSPI 登录（Ahmed Shinwari、Magnus Hagander）

  该问题的典型症状是 SSPI 登录期间的“The function requested is not supported”错误。
- 如果 `pg_hba.conf` 包含 `hostssl` 但 SSL 被禁用则抛出错误（Tom Lane）

  这被认定比此前静默忽略这类行的行为更用户友好。
- 修复 `pg_srand48` 种子初始化中的笔误（Andres Freund）

  这导致未能使用所提供种子的所有位。该函数在大多数平台上不使用（只在没有 `srandom` 的平台上使用），而且无论如何看待，随机性低于预期的种子带来的潜在安全暴露似乎都很小。
- 避免 `LIMIT` 与 `OFFSET` 值之和超过 2^63 时的整数溢出（Heikki Linnakangas）
- 为 `generate_series()` 的 `int4` 和 `int8` 版本添加溢出检查（Robert Haas）
- 修复 `to_char()` 中移除尾随零的问题（Marti Raudsepp）

  在使用 `FM` 且小数点后没有数字位的格式中，小数点左侧的零可能被错误移除。
- 修复 `pg_size_pretty()`，避免接近 2^63 的输入溢出（Tom Lane）
- 放宽 plpgsql 对记录值中 typmod 匹配的检查（Tom Lane）

  过于严格的检查可能导致丢弃本应保留的长度修饰符。
- 在 initdb 期间正确处理区域名中的引号（Heikki Linnakangas）

  这种情况可能出现在某些 Windows 区域设置下，例如“People's Republic of China”。
- 修复 pg_upgrade，使其在从 8.3 升级期间保留 toast 表的 relfrozenxid（Bruce Momjian）

  不这样做可能导致升级后 `pg_clog` 文件被过早删除。
- 在 pg_ctl 中，支持 Windows 上服务注册的静默模式（MauMau）
- 修复 psql 在从其他文件执行 `COPY` 期间脚本文件行号计数的问题（Tom Lane）
- 修复 pg_restore 直接写数据库模式对 `standard_conforming_strings` 的处理（Tom Lane）

  当从以 `standard_conforming_strings` 设为 `on` 制作的归档文件直接恢复到数据库服务器时，pg_restore 可能发出不正确的命令。
- 对并行 pg_restore 不支持的情况更加用户友好（Tom Lane）

  此变更确保这类情况在任何恢复动作发生之前就被检测并报告。
- 修复 libpq 的 LDAP 服务查找代码中的越界写入和内存泄漏（Albe Laurenz）

- 在 libpq 中，避免使用非阻塞 I/O 和 SSL 连接时的失败（Martin Pihlak、Tom Lane）
- 改进 libpq 对连接启动期间失败的处理（Tom Lane）

  特别是，对 SSL 连接启动期间服务器报告 `fork()` 失败的响应现在更合理。
- 改进 libpq 对 SSL 失败的错误报告（Tom Lane）
- 修复 `PQsetvalue()`，避免向最初来自服务器查询的 `PGresult` 添加新元组时可能崩溃（Andrew Chernow）
- 让 ecpglib 以 15 位精度写 `double` 值（Akira Kurosawa）
- 在 ecpglib 中，确保出错后恢复 `LC_NUMERIC` 设置（Michael Meskes）
- 应用上游对 blowfish 有符号字符缺陷的修复（CVE-2011-2483）（Tom Lane）

  `contrib/pg_crypto` 的 blowfish 加密代码在 char 为有符号的平台（大多数平台如此）上可能给出错误结果，导致加密口令比应有的强度更弱。
- 修复 `contrib/seg` 中的内存泄漏（Heikki Linnakangas）
- 修复 `pgstatindex()`，使其对空索引给出一致结果（Tom Lane）
- 允许使用 perl 5.14 构建（Alex Hunsaker）
- 更新 configure 脚本探测系统函数是否存在的方法（Tom Lane）

  我们在 8.3 和 8.2 中使用的 autoconf 版本可能被执行链接时优化的编译器欺骗。
- 修复构建和安装文件路径包含空格的各种问题（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2011i for DST law changes in Canada, Egypt, Russia, Samoa, and South Sudan。

E.64. 版本 8.4.8

发行日期
2011-04-18

本次发布包含对 8.4.7 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.64.1. 迁移到版本 8.4.8

对于运行 8.4.X 的用户，不需要进行转储/恢复。

不过，如果你的安装是通过运行 pg_upgrade 从前一个大版本升级的，应采取措施防止因 pg_upgrade 一个现已修复的缺陷导致可能的数据丢失。推荐的解决方案是对所有 TOAST 表执行 `VACUUM FREEZE`。更多信息见 http://wiki.postgresql.org/wiki/20110408pg_upgrade_fix[4]。

不过，如果你是从earlier than 8.4.2 的版本升级，请参见第 E.70 节。

[4] http://wiki.postgresql.org/wiki/20110408pg_upgrade_fix

E.64.2. 变更

- 修复 pg_upgrade 对 TOAST 表的处理（Bruce Momjian）

  在 pg_upgrade 期间，TOAST 表的 `pg_class.relfrozenxid` 值未被正确复制到新安装。这之后可能导致 `pg_clog` 文件在仍被需要用于验证 TOAST 表中的元组时被丢弃，导致 “could not access status of transaction” 失败。

  这一错误对曾用 pg_upgrade 升级的安装构成显著的数据丢失风险。此补丁纠正了未来使用 pg_upgrade 时的问题，但本身不能治愈已用带缺陷版本的 pg_upgrade 处理过的安装中的问题。

- 抑制错误的 “PD_ALL_VISIBLE flag was incorrectly set” 警告（Heikki Linnakangas）

  `VACUUM` 有时会在实际有效的情况下发出此警告。

- 禁止在组合类型中包含其自身（Tom Lane）

  这防止了服务器在处理该组合类型时无限递归的场景。虽然这种结构有一些可能的用途，但它们似乎不足以抵消确保其始终安全工作所需的努力。

- 避免目录缓存初始化期间的潜在死锁（Nikhil Sontakke）

  某些情况下，缓存加载代码会在锁定索引的目录之前获取系统索引上的共享锁。这可能与尝试以另一种更标准的顺序获取排他锁的进程死锁。

- 修复当目标元组存在并发更新时 `BEFORE ROW UPDATE` 触发器处理中的悬空指针问题（Tom Lane）

  已观察到该缺陷在尝试执行 `UPDATE RETURNING ctid` 时导致间歇性的 “cannot extract system attribute from virtual tuple” 失败。发生更严重错误（例如为更新后的元组生成错误索引项）的概率非常小。

- 当表存在未决的延迟触发器事件时禁止 `DROP TABLE`（Tom Lane）

  此前 `DROP` 会照常执行，导致触发器最终触发时报 “could not open relation with OID nnn” 错误。

- 防止 GEQO 优化期间由恒假 WHERE 条件触发的崩溃（Tom Lane）
- 改进规划器对半连接和反连接情况的处理（Tom Lane）
- 修复文本搜索的选择率估计，使其考虑 NULL（Jesper Krogh）
- 改进 PL/pgSQL 处理带已删除列的行类型的能力（Pavel Stehule）

  这是先前在 9.0 中所做修复的向后移植。

- 修复涉及数组切片的 PL/Python 内存泄漏（Daniel Popowich）
- 修复 pg_restore，使其能处理 TOC 文件中的长行（超过 1KB）（Tom Lane）
- 加入更多防护，防止因过于激进的编译器优化导致的除零崩溃（Aurelien Jarno）
- 支持在 FreeBSD 和 OpenBSD 的 MIPS 上使用 dlopen()（Tom Lane）

  此前有一个硬编码的假设：该系统函数在这些系统的 MIPS 硬件上不可用。现在改用编译期测试，因为较新的版本已经有了它。

- 修复 HP-UX 上的编译失败（Heikki Linnakangas）
- 修复 Windows 上 libintl 的版本不兼容问题（Hiroshi Inoue）
- 修复 Windows 构建脚本中 xcopy 的用法，使其在 Windows 7 下正确工作（Andrew Dunstan）

  这只影响构建脚本，不影响安装或使用。
- 修复 pg_regress 在 Cygwin 上使用的路径分隔符（Andrew Dunstan）
- 把时区数据文件更新为 tzdata 版本 2011f for DST law changes in Chile, Cuba, Falkland Islands, Morocco, Samoa, and Turkey; also historical corrections for South Australia, Alaska, and Hawaii。

E.65. 版本 8.4.7

发行日期
2011-01-31

本次发布包含对 8.4.6 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.65.1. 迁移到版本 8.4.7

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.4.2 的版本升级，请参阅第 E.70 节。

E.65.2. 变更

- 避免 `EXPLAIN` 尝试显示简单形式 `CASE` 表达式时的失败（Tom Lane）

  如果 `CASE` 的测试表达式是常量，规划器可能把 `CASE` 简化为令表达式显示代码困惑的形式，导致 “unexpected CASE WHEN clause” 错误。
- 修复对现有下标范围之前的数组切片赋值的问题（Tom Lane）

  如果新增下标与第一个既有下标之间存在间隙，
  代码会错误计算需要从旧数组的空值位图复制多少条目，可能导致数据损坏或崩溃。
- 避免规划器对非常久远的日期值出现意外的转换溢出（Tom Lane）

  `date` 类型支持的日期范围比 `timestamp` 类型能表示的更宽，但规划器假定它总是可以不受惩罚地把 date 转换为 timestamp。
- 修复启用 `standard_conforming_strings` 时 pg_restore 对大对象（BLOB）的文本输出（Tom Lane）

  虽然直接恢复到数据库时工作正确，但如果要求 pg_restore 生成 SQL 文本输出且源数据库中已启用 `standard_conforming_strings`，字符串转义是不正确的。
- 修复对包含 `... & !(subexpression) | ...` 的 `tsquery` 值的错误解析（Tom Lane）

包含这种操作符组合的查询没有被正确执行。同样的错误也存在于 `contrib/intarray` 的 `query_int` 类型和 `contrib/ltree` 的 `ltxtquery` 类型中。

- 修复 `contrib/intarray` 的 `query_int` 类型输入函数中的缓冲区越界（Apple）

  该缺陷构成安全风险，因为函数的返回地址可能被覆写。感谢 Apple Inc 的安全团队报告此问题并提供修复。（CVE-2010-4015）

- 修复 `contrib/seg` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `seg` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。（这与上一次更新中在 `contrib/cube` 中修复的缺陷相同。）

E.66. 版本 8.4.6

发行日期
2010-12-16

本次发布包含对 8.4.5 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.66.1. 迁移到版本 8.4.6

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.4.2 的版本升级，请参阅第 E.70 节。

E.66.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复 GIN 索引 WAL 重放逻辑中的若干缺陷（Tom Lane）

  这可能导致复制期间的 “bad buffer id: 0” 失败或索引内容损坏。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 修复多个工作进程长期保持活动时自动清理工作进程的持续变慢（Tom Lane）

  如果自动清理工作进程处理了足够多的表，其有效 `vacuum_cost_limit` 可能降到接近零，导致其运行极慢。

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

  在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 修复早期连接处理期间的错误检查（Tom Lane）

  某些情况下对子进程过多的检查被跳过，可能在尝试把新子进程添加到固定大小数组时导致 postmaster 崩溃。

- 改进窗口函数的效率（Tom Lane）

  某些需要提前读取大量元组、但 `work_mem` 足够大可以全部保存在内存中的情况意外地慢。`percent_rank()`、`cume_dist()` 和 `ntile()` 尤其受此问题影响。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）

- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复当 `ORDER BY`、`LIMIT`、`FOR UPDATE` 或 `WITH` 附加到 `INSERT ... VALUES` 的 `VALUES` 部分时的正确行为（Tom Lane）

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

  规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 修复当连接接受（`accept()` 或紧随其后的调用之一）失败且 postmaster 编译时带 GSSAPI 支持时的 postmaster 崩溃（Alexander Chernikov）

- 修复 `log_temp_files` 处于活动状态时临时文件漏删的问题（Tom Lane）

  如果尝试发出日志消息时出错，就不会执行删除，导致临时文件累积。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

  该缺陷影响多个不同的几何距离测量操作符。

- 修复 ecpg 中事务状态的错误计算（Itagaki Takahiro）

- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）

- 修复 PL/Python 对集合返回函数的处理（Jan Urbanski）

  在生成集合结果的迭代器内尝试调用 SPI 函数会失败。

- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

这可能导致 cube 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 REINDEX。

- contrib/dblink 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）
- 修复 contrib/pgcrypto 在缺少公钥时可能的核心转储（Marti Raudsepp）
- 修复 contrib/xml2 的 XPath 查询函数中的内存泄漏（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.67. 版本 8.4.5

发行日期
2010-10-04

本次发布包含对 8.4.4 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.67.1. 迁移到版本 8.4.5

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.4.2 的版本升级，请参阅第 E.70 节。

E.67.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 SECURITY DEFINER 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

  感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 pg_get_expr()，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 在 Windows 上把退出码 128（ERROR_WAIT_NO_CHILDREN）视为非致命（Magnus Hagander）

  高负载下，Windows 进程有时会在启动时以此错误码失败。此前 postmaster 把这视为 panic 情况并重启整个数据库，但这似乎反应过度。

- 修复占位符求值位置不正确的问题（Tom Lane）

  当外连接的内侧是输出列表中含非严格表达式的子 SELECT 时，该缺陷可能导致本应为空的查询输出非空。

- 修复 `UNION ALL` 成员关系可能被重复扫描的问题（Tom Lane）

- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，会出现此错误。

- 修复对引用视图或子查询且出现在嵌套子查询中的整行 Var 的错误处理（Tom Lane）

- 修复跨类型 `IN` 比较的错误处理（Tom Lane）

  如果规划器试图用先排序、再去重、再普通连接的计划实现 `IN` 连接，这可能导致失败。

- 修复 `tsvector` 列的 `ANALYZE` 统计信息计算（Jan Urbanski）

  原来的编码可能产生不正确的统计信息，导致之后糟糕的计划选择。

- 改进规划器对 `array_agg()`、`string_agg()` 及类似聚合函数所用内存的估计（Hitoshi Harada）

  此前严重的低估可能因不当选择哈希聚合计划而导致内存不足失败。

- 修复未将缓存计划标记为瞬态的失败（Tom Lane）

  如果在某个被引用表正在执行 `CREATE INDEX CONCURRENTLY` 期间准备了计划，应当在索引可供使用后重新规划。这此前没有可靠地发生。

- 在偶有报告的 btree 失败情形中把 PANIC 降为 ERROR，并在产生的错误消息中提供更多细节（Tom Lane）

  这应能改进系统对损坏索引的健壮性。

- 修复 GIN 索引部分匹配查询的错误搜索逻辑（Tom Lane）

  涉及多个 GIN 索引条件的 AND/OR 组合的情形并不总是给出正确答案，而且有时比必要的慢得多。

- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）

- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）

- 修复子事务回滚期间待处理触发器事件列表可能损坏的问题（Tom Lane）

  这可能导致崩溃或触发器错误触发。

- 修复对按引用传递的函数结果计算哈希时可能的失败（Tao Ma、Tom Lane）

- 改进归并连接对连接列中 NULL 的处理（Tom Lane）

  如果排序顺序使 NULL 排在高位，归并连接现在可以在到达第一个 NULL 时完全停止。

- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 避免 walwriter 进程持有打开的旧 WAL 段（Magnus Hagander、Heikki Linnakangas）

  以前的编码会阻止删除不再需要的段。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）

- 防止误解 TOAST 表的部分指定关系选项（Itagaki Takahiro）

  特别是，如果为表设置了任何其他 reloption，`fillfactor` 会被读为零，导致严重膨胀。

- 修复 `ALTER TABLE ... ADD CONSTRAINT` 中的继承计数跟踪（Robert Haas）

- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）

- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）

- 改进 `CREATE INDEX` 对所提议索引表达式是否不可变的检查（Tom Lane）

- 修复 `REASSIGN OWNED` 以处理操作符类和操作符族（Asko Tiidumaa）

- 修复比较两个空 `tsquery` 值时可能的核心转储（Tom Lane）

- 修复 `LIKE` 对包含 `%` 后跟 _ 的模式的处理（Tom Lane）

  我们此前修复过它，但仍有一些被错误处理的情形。

- 重新允许输入公元 0001-01-01 之前的儒略日期（Tom Lane）

  像 `'J100000'::date` 这样的输入在 8.4 之前可行，但被新增的错误检查无意中破坏。

- 修复 PL/pgSQL：如果在迭代某游标的 `FOR` 循环内关闭该游标，抛出错误而不是崩溃（Heikki Linnakangas）

- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）

- 在 libpq 中，修复同时指定 `host` 和 `hostaddr` 情形下的完整 SSL 证书验证（Tom Lane）

- 让 psql 在自动提交关闭模式下把 `DISCARD ALL` 识别为不应包裹在事务块中的命令（Itagaki Takahiro）

- 修复 pg_dump 处理 SQL/MED 对象的一些问题（Tom Lane）

  特别是，pg_dump 由非超级用户运行时总会失败，这并非本意。

- 改进 pg_dump 和 pg_restore 对不可寻址归档文件的处理（Tom Lane、Robert Haas）

  这对并行恢复的正常运作很重要。

- 改进并行 pg_restore 应对选择性恢复（`-L` 选项）的能力（Tom Lane）

  如果 `-L` 文件指定了非默认的恢复顺序，原来的代码倾向于失败。

- 修复 ecpg 以正确处理来自 `RETURNING` 子句的数据（Michael Meskes）
- 修复 ecpg 中的一些内存泄漏（Zoltan Boszormenyi）
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 为 `contrib/hstore` 新增 `hstore(text, text)` 函数（Robert Haas）

  此函数是现已废弃的 `=>` 操作符的推荐替代。向后移植它使得面向未来的代码可以用于较旧的服务器版本。注意该补丁只有在 `contrib/hstore` 于特定数据库中安装或重新安装之后才生效。用户也可以改为手工执行 `CREATE FUNCTION` 命令。

- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

- 让 Windows 的“N. Central Asia Standard Time”时区映射到 Asia/Novosibirsk 而不是 Asia/Almaty（Magnus Hagander）

  Microsoft 在 KB976098 的时区更新中改变了此区域的 DST 行为。Asia/Novosibirsk 更匹配其新行为。

E.68. 版本 8.4.4

发行日期
2010-05-17

本次发布包含对 8.4.3 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.68.1. 迁移到版本 8.4.4

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.4.2 的版本升级，请参阅第 E.70 节。

E.68.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

近期的事态发展使我们确信 Safe.pm 太不安全，不能依靠它使 plperl 可信。此变更完全移除了对 Safe.pm 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 plperl 中以自然的方式使用 Perl 的 strict 编译指令，Perl 的 $a 和 $b 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 pltcl_modules 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 pltcl_modules 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 pltclu 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 修复 ALTER ... SET TABLESPACE 的 WAL 重放期间的数据损坏（Tom）

  当 archive_mode 开启时，ALTER ... SET TABLESPACE 生成的 WAL 记录的重放逻辑不正确。它可能把数据写到错误的位置，导致可能无法恢复的数据损坏。数据损坏会在备机从属上被观察到；如果在提交 ALTER 之后、下一个检查点之前发生数据库崩溃和恢复，主机上也可能发生。

- 修复在重建 relcache 条目期间收到缓存重置消息时可能的崩溃（Heikki）

  此错误是在 8.4.3 中修复相关失败时引入的。

- 在为函数运行语言验证器时应用逐函数 GUC 设置（Itagaki Takahiro）

  这避免了函数代码在没有该设置时无效所导致的失败；例如 search_path 不正确时 SQL 函数可能无法解析。

- 当 constraint_exclusion = partition 时，对继承的 UPDATE 和 DELETE 目标表做约束排除（Tom）

  由于疏忽，此设置此前只在 SELECT 命令中触发约束排除检查。

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 ALTER USER ... RESET ALL，或为他拥有的数据库运行 ALTER DATABASE ... RESET ALL，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，ALTER 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 CONTEXT 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 修复 recovery_end_command 中 %r 参数的错误处理（Heikki）

  该值总是出来为零。

- 确保归档进程尽快响应 archive_command 的变更（Tom）

- 修复 pl/pgsql 的 CASE 语句在 case 表达式是不返回行的查询时不失败（Tom）

- 为现代 Perl 版本更新 pl/perl 的 ppport.h（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 在 ecpg 中正确处理空字符串连接参数（Michael）
- 防止 psql 在展开引用自身的变量时无限递归（Tom）
- 修复 psql 的 `\copy`，不在 `\copy (select ...)` 内的点周围添加空格（Tom）

  在数字字面量的小数点周围添加空格会导致语法错误。
- 避免 psql 在与 `client_encoding` 不匹配的区域设置环境中运行时的格式化失败（Tom）
- 修复使用 `contrib/intarray` 操作符的不可满足查询产生不必要的 “GIN 索引不支持整索引扫描” 错误（Tom）
- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）
- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。
- 避免 Windows 上 syslogger 进程可能的崩溃（Heikki）
- 更健壮地处理 Windows 注册表中不完整的时区信息（Magnus）
- 更新已知 Windows 时区名称集合（Magnus）
- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

  另外，把 `PKST`（巴基斯坦夏令时）加入默认时区缩写集合。

E.69. 版本 8.4.3

发行日期
2010-03-15

本次发布包含对 8.4.2 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.69.1. 迁移到版本 8.4.3

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.4.2 的版本升级，请参阅第 E.70 节。

E.69.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 修复后端启动期间可能的死锁（Tom）
- 修复因未干净地处理 relcache 重载期间的错误而可能的崩溃（Tom）
- 修复因使用指向缓存计划的悬空指针而可能的崩溃（Tatsuo）
- 修复因过于激进地使 `ROLLBACK` 的缓存计划失效而可能的崩溃（Tom）
- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）
- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）
- 修复 GIST 索引页分裂的恢复结束清理期间发出的不正确 WAL 数据（Yoichi Hirai）

  如果在完成未完成的 GIST 插入之后、在恢复结束清理期间崩溃，这会导致索引损坏，或者更可能在 WAL 重放期间出错。
- 修复 GIN 索引 WAL 重做清理方法中的缺陷（Heikki）
- 修复 GIN 索引搜索中扫描键的不正确比较（Teodor）
- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。
- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）
- 修复若干病态缓慢的正则表达式匹配情形（Tom）
- 修复试图内联返回包含已删除列的组合类型集合的 SQL 函数时出现的缺陷（Tom）
- 修复试图更新组合类型数组列中某元素的字段时的缺陷（Tom）
- 避免 `EXPLAIN` 需要打印 FieldStore 或赋值 ArrayRef 表达式时的失败（Tom）

  既然 `EXPLAIN VERBOSE` 试图打印计划节点目标列表，这些情形现在可能出现。
- 避免某些情况下未加修饰的字面字符串出现在 `UNION/INTERSECT/EXCEPT` 内的子查询中时不必要的强制转换失败（Tom）

  这修复了某些在 8.4 之前可用的情形的回归。
- 避免某些情况下整行 Var 的行类型包含已删除列时不希望出现的行类型兼容性检查失败（Tom）
- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）
- 总是把目录 ID 传递给 `CREATE FOREIGN DATA WRAPPER` 中指定的选项验证器函数（Martin Pihlak）
- 修复更多临时文件泄漏的情形（Heikki）

  这纠正了上一个小版本中引入的问题。失败的一种情形是返回集合的 plpgsql 函数在另一个函数的异常处理器中被调用。

- 新增对 `FULL JOIN ON FALSE` 的支持（Tom）

  这防止了某些现在可被简化为常量假连接条件的查询相对 8.4 之前版本的回归。

- 改进布尔变量情形的约束排除处理，特别是使排除带有 “bool_column = false” 约束的分区成为可能（Tom）

- 防止把 `INOUT` 转换当作表示二进制兼容（Heikki）

- 在警告无法授予或撤销列级权限时在消息中包含列名（Stephen Frost）

  这比以前更有用，并有助于防止 `REVOKE` 产生多条消息时的困惑，那些消息此前看起来像是重复的。

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）

- 修复 `SSL_read` 或 `SSL_write` 失败而未设置 `errno` 时可能的无限循环（Tom）

  据报道这在某些 Windows 版本的 openssl 上可能发生。

- 不允许在本地连接上使用 GSSAPI 认证，因为它需要主机名才能正确工作（Magnus）

- 保护 ecpg 免遭应用程序意外释放字符串（Michael）

- 让 ecpg 在连接消失时报告正确的 SQLSTATE（Michael）

- 修复 psql `\d` 输出中单元格内容的翻译（Heikki）

- 修复 psql 的 `numericlocale` 选项，使其不在 latex 和 troff 输出格式中格式化不应格式化的字符串（Heikki）

- 修复 psql 中每查询的小内存泄漏（Tom）

- 当同时指定 `ON_ERROR_STOP` 和 `--single-transaction` 且隐含的 `COMMIT` 期间发生错误时，让 psql 返回正确的退出状态（3）（Bruce）

- 修复 pg_dump 对外部服务器权限的输出（Heikki）

- 修复并行 pg_restore 中因依赖 ID 超出范围而可能的崩溃（Tom）

- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）

- 修复从 PL/PerlU 调用 PL/Perl 函数或相反时可能的失败（Tim Bunce）

- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）

- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。

- 防止对在失败的事务或子事务内创建的门户运行 `ExecutorEnd`（Tom）

已知这会在使用 `contrib/auto_explain` 时引起问题。

- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）
- 允许 `contrib/ltree` 操作中使用零维数组（Tom）

  这种情形此前被作为错误拒绝，但把它与零元素数组同样处理更方便。特别是，这避免了当 `ltree` 操作应用于 `ARRAY(SELECT ...)` 的结果而子查询不返回行时不必要的失败。
- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）
- 让 `contrib/xml2` 在 Windows 上的构建更健壮（Andrew）
- 修复 Windows 信号处理中的竞态条件（Radu Ilie）

  此缺陷的一个已知症状是 `pg_listener` 中的行可能在重负载下被丢弃。
- 让 configure 脚本在 C 编译器不提供可用的 64 位整数数据类型时报告失败（Tom）

  这种情形已经损坏了一段时间，似乎不再值得支持，因此改为在 configure 时拒绝它。
- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.70. 版本 8.4.2

发行日期
2009-12-14

本次发布包含对 8.4.1 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.70.1. 迁移到版本 8.4.2

对于运行 8.4.X 的用户，不需要进行转储/恢复。不过，如果你有任何哈希索引，应当在更新到 8.4.2 后对它们执行 `REINDEX`，以修复可能的损坏。

E.70.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复哈希索引损坏（Tom）

  8.4 中让哈希索引按哈希值排序保存条目的变更没有更新桶分裂和压缩例程以保持该顺序。因此应用这两种操作之一可能导致索引永久损坏，即搜索可能找不到实际存在的条目。
  为处理此问题，建议在安装此更新后对你可能拥有的任何哈希索引执行 `REINDEX`。

- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 避免在空同义词词典上崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。
- 修复试图对内部依赖对象执行 `DROP` 时的崩溃（Tom）
- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 确保预备事务所持有的共享元组级锁不被忽略（Heikki）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复轮转到新 CSV 日志文件时 syslogger 进程的内存泄漏（Tom）
- 修复重新解析 `pg_hba.conf` 时 postmaster 的内存泄漏（Tom）
- 修复 Windows 权限降级逻辑（Jesse Morris）

  这修复了数据库在 Windows 上启动失败的一些情形，这些情形常伴随误导性错误消息，例如“找不到匹配的 postgres 可执行文件”。
- 让主查询中的 `FOR UPDATE/SHARE` 不传播到 `WITH` 查询中（Tom）

  例如，在

  ```
  WITH w AS (SELECT * FROM foo) SELECT * FROM w, bar ... FOR UPDATE
  ```

  中，`FOR UPDATE` 现在将影响 `bar` 而不影响 `foo`。这比原来 8.4 的行为更有用也更一致，原来的行为试图把 `FOR UPDATE` 传播到 `WITH` 查询中，但由于各种实现限制总是失败。这也遵循 `WITH` 查询如同独立于主查询一样执行的设计规则。
- 修复一个 `WITH RECURSIVE` 查询紧邻另一个之内时的缺陷（Tom）
- 修复哈希索引中的并发缺陷（Tom）

  并发插入可能导致索引扫描瞬时报告错误结果。
- 修复 GiST 索引页分裂的不正确逻辑：当分裂依赖于索引的非首列时（Paul Ramsey）
- 修复启用 `fastupdate` 的多列 GIN 索引搜索结果错误的问题（Teodor）
- 修复 GIN 索引 WAL 条目创建中的缺陷（Tom）

  这些缺陷在 `full_page_writes` 开启时被掩盖，但在它关闭时，如果下一个检查点之前发生崩溃，WAL 重放失败是必然的。
- 如果回收或删除旧 WAL 文件在检查点结束时失败，不要报错退出（Heikki）

更好的做法是把该问题视为非致命并允许检查点完成。未来的检查点会重试删除。
这类问题在正常运行中不应出现，但已知可能由设计不当的 Windows 反病毒和备份软件引起。

- 确保 WAL 文件不会在 Windows 上被重复归档（Heikki）

  这是其他进程干扰删除不再需要的文件时可能出现的另一种症状。

- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 提高 GSSAPI 和 SSPI 认证方法的最大认证令牌（Kerberos 票据）大小（Ian Turner）

  虽然旧的 2000 字节限制对 Unix Kerberos 实现绰绰有余，Windows 域控制器签发的票据可能大得多。

- 确保域约束在 `ARRAY[...]::domain`（域基于数组类型）这类构造中得到强制执行（Heikki）

- 修复某些以组合类型列作为外键的情形下的外键逻辑（Tom）

- 确保游标的快照在创建后不被修改（Alvaro）

  如果同一事务中稍后的操作修改了游标本应返回的数据，这可能导致游标交付错误结果。

- 修复 `CREATE TABLE`，使其正确合并来自不同继承父表的默认表达式（Tom）

  这以前可用，但在 8.4 中损坏了。

- 重新启用对序列的访问统计信息收集（Akira Kurosawa）

  这以前可用，但在 8.3 中损坏了。

- 修复 `CREATE OR REPLACE FUNCTION` 期间所有权依赖的处理（Tom）

- 修复对 `WHERE` *x*=*x* 条件的不正确处理（Tom）

  某些情况下它们可能被当作冗余而忽略，但它们并非冗余——它们等价于 *x* `IS NOT NULL`。

- 修复对文本上相同的易变表达式用哈希聚合实现 `DISTINCT` 时的错误计划构造（Tom）

- 修复易变 `SELECT DISTINCT ON` 表达式的断言失败（Tom）

- 修复 `ts_stat()`，使其在空 `tsvector` 值上不失败（Tom）

- 让文本搜索解析器接受 XML 属性中的下划线（Peter）

- 修复 `xml` 二进制输入的编码处理（Heikki）

  如果 XML 头未指定编码，现在默认采用 UTF-8；以前的处理不一致。

- 修复从 `plperl` 调用 `plperlu` 或相反时的缺陷（Tom）

  内层函数的错误退出可能因未能为外层函数重新选择正确的 Perl 解释器而导致崩溃。

- 修复重定义 PL/Perl 函数时的会话期内存泄漏（Tom）
- 确保 Perl 数组在由返回集合的 PL/Perl 函数返回时被正确转换为 PostgreSQL 数组（Andrew Dunstan、Abhijit Menon-Sen）

  对非返回集合的函数这早已正确工作。
- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 修复 ecpg 中 `DECLARE CURSOR` 语句内注释的问题（Michael）
- 修复 ecpg，使其不把最近新增的关键字当作保留字（Tom）

  这影响关键字 `CALLED`、`CATALOG`、`DEFINER`、`ENUM`、`FOLLOWING`、`INVOKER`、`OPTIONS`、`PARTITION`、`PRECEDING`、`RANGE`、`SECURITY`、`SERVER`、`UNBOUNDED` 和 `WRAPPER`。
- 重新允许 psql 的 `\df` 函数名参数中使用正则表达式特殊字符（Tom）
- 在 `contrib/fuzzystrmatch` 中，修正非默认代价下 `levenshtein` 距离的计算（Marcin Mank）
- 在 `contrib/pg_standby` 中，禁用在 Windows 上用信号触发故障转移（Fujii Masao）

  这从未有过任何用处，因为 Windows 没有 Unix 式信号，但最近的变更使它实际上崩溃。
- 在 `contrib/vacuumdb` 生成的 `VACUUM` 命令中把 `FREEZE` 和 `VERBOSE` 选项按正确顺序排列（Heikki）
- 修复 `contrib/dblink` 遇到错误时可能的连接泄漏（Tatsuhito Kasahara）
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。
- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 更新时区缩写文件以符合当前现实（Joachim Wieland）

  这包括把 `IDT` 加入默认时区缩写集合。
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.71. 版本 8.4.1

发行日期
2009-09-09

本次发布包含对 8.4 的多方面修复。有关 8.4 大版本中新特性的信息，请参见第 E.72 节。.

E.71.1. 迁移到版本 8.4.1

对于运行 8.4.X 的用户，不需要进行转储/恢复。

E.71.2. 变更

- 修复归档恢复结束时的 WAL 页头初始化（Heikki）

  这可能导致后续归档恢复中无法处理 WAL。

- 修复“恢复期间不能创建新的 WAL 条目”错误（Tom）

- 修复崩溃后可能使已过期行可见的问题（Tom）

  此缺陷涉及页状态位在服务器崩溃后可能未被正确设置。

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 把对已装载的可装载模块的 `LOAD` 变为无操作（Tom）

  此前 `LOAD` 会试图卸载并重新装载模块，但这不安全也不太有用。

- 让窗口函数的 `PARTITION BY` 和 `ORDER BY` 项总是被解释为简单表达式（Tom）

  在 8.4.0 中这些列表按照用于顶层 `GROUP BY` 和 `ORDER BY` 列表的规则解析。但这不符合 SQL 标准，而且可能导致循环。

- 修复半连接规划中的若干错误（Tom）

  这导致某些 `IN` 或 `EXISTS` 与另一个连接一起使用的情形下查询结果错误。

- 修复对处于外连接之内的子查询的整行引用的处理（Tom）

  例子是 `SELECT COUNT(ss.*) FROM ... LEFT JOIN (SELECT ...) ss ON ...`。这里，对于空扩展的连接行，`ss.*` 会被当作 `ROW(NULL,NULL,...)`，这与简单 NULL 不同。现在它被当作简单 NULL。

- 修复 Windows 共享内存分配代码（Tsutomu Yamada、Magnus）

  此缺陷导致常被报告的“无法重新附加到共享内存”错误消息。

- 修复 plperl 的区域设置处理（Heikki）

  此缺陷可能导致调用 plperl 函数时服务器的区域设置改变，进而导致数据损坏。

- 修复 reloptions 的处理，确保设置一个选项不会强制其他选项取默认值（Itagaki Takahiro）

- 确保“快速关闭”请求会强制终止打开的会话，即使“智能关闭”已在进行中（Fujii Masao）

- 避免 `GROUP BY` 查询中 `array_agg()` 的内存泄漏（Tom）

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

它以前被当作 `'th'`（小写）处理。

- 对 `time` 和 `time with time zone` 输入，在 `EXTRACT(second)` 和 `EXTRACT(milliseconds)` 的结果中包含小数部分（Tom）

  这在浮点日期时间配置下一直正常，但在整数日期时间代码中损坏了。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）
- 改进索引扫描中处理 TOAST 值时的性能（Tom）

  这对 PostGIS[5] 特别有用。

- 修复一个使 `commit_delay` 失效的笔误（Jeff Janes）
- 如果服务器以安静模式启动，把早期启动消息输出到 `postmaster.log`（Tom）

  此前这类错误消息被丢弃，导致调试困难。

- 移除已翻译的 FAQ（Peter）

  它们现在位于 wiki[6]。主 FAQ 此前已被移到 wiki。

- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）
- 修复 pg_dump 的 `--binary-upgrade` 模式中的若干错误（Bruce、Tom）

  `pg_dump --binary-upgrade` 由 pg_migrator 使用。

- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）
- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）
- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）
- 绕过 gcc 在某些平台上导致“浮点异常”而非“除以零”的缺陷（Tom）
- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Mauritius。

E.72. 版本 8.4

发行日期
2009-07-01

E.72.1. 概述

经过多年开发，PostgreSQL 在许多领域已经功能完备。此发行版展示了添加特性的针对性方法（例如认证、监控、空间复用），并增加了后续 SQL 标准中定义的能力。主要的增强领域是：

[5] http://postgis.net/
[6] http://wiki.postgresql.org/wiki/FAQ

- 窗口函数
- 公共表表达式和递归查询
- 函数的默认参数和可变参数
- 并行恢复
- 列权限
- 逐数据库的区域设置
- 改进的哈希索引
- 改进 `EXISTS` 和 `NOT EXISTS` 查询的连接性能
- 更易用的热备
- 自动调整空闲空间映射的大小
- 可见性映射（大幅降低变化缓慢的表的清理开销）
- 版本感知的 psql（反斜杠命令可针对较旧的服务器工作）
- 支持用 SSL 证书进行用户认证
- 逐函数的运行时统计
- 在 psql 中轻松编辑函数
- 新增 contrib 模块：pg_stat_statements、auto_explain、citext、btree_gin

上述各项在下面的章节中有更详细的说明。

E.72.2. 迁移到版本 8.4

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

E.72.2.1. 常规

- 默认使用 64 位整数日期时间（Neil Conway）

  此前这由 configure 的 `--enable-integer-datetimes` 选项选择。要保留旧行为，请用 `--disable-integer-datetimes` 构建。

- 移除 ipcclean 工具命令（Bruce）

  该工具只在少数平台上有效。用户应改用其操作系统的工具。

E.72.2.2. 服务器设置

- 把 `log_min_messages` 的默认设置改为 `warning`（以前是 `notice`），以减少日志文件量（Tom）
- 把 `max_prepared_transactions` 的默认设置改为零（以前是 5）（Tom）

- 让 debug_print_parse、debug_print_rewritten 和 debug_print_plan 的输出以 LOG 消息级别出现，而不是以前的 DEBUG1（Tom）
- 让 debug_pretty_print 默认为 on（Tom）
- 移除 explain_pretty_print 参数（不再需要）（Tom）
- 让 log_temp_files 仅超级用户可设置，与其他日志选项一样（Simon Riggs）
- 移除当 log_filename 中没有 % 转义时自动追加纪元时间戳的行为（Robert Haas）

  此变更是因为一些用户想要固定的日志文件名，以便与外部日志轮转工具配合使用。
- 从 recovery.conf 中移除 log_restartpoints；改用 log_checkpoints（Simon）
- 移除 krb_realm 和 krb_server_hostname；它们现在改在 pg_hba.conf 中设置（Magnus）
- pg_hba.conf 中也有重大变更，如下所述。

E.72.2.3. 查询

- 更改 TRUNCATE 和 LOCK，使其应用于指定表的子表（Peter）

  这些命令现在接受阻止处理子表的 ONLY 选项；如果需要旧行为，必须使用此选项。
- SELECT DISTINCT 和 UNION/INTERSECT/EXCEPT 不再总是产生排序的输出（Tom）

  此前，这类查询总是通过排序/唯一处理（即先排序再删除相邻重复项）来删除重复行。现在它们可以用哈希实现，这不会产生排序的输出。如果应用程序依赖输出按排序顺序，推荐的解决办法是添加 ORDER BY 子句。作为短期变通，可以通过禁用 enable_hashagg 恢复以前的行为，但那是非常耗费性能的解决办法。不过 SELECT DISTINCT ON 从不使用哈希，因此其行为不变。
- 强制子表从父表继承 CHECK 约束（Alex Hunsaker、Nikhil Sontakke、Tom）

  此前可以从子表删除这样的约束，使违反约束的行在扫描父表时可见。这被认为不一致，也违背 SQL 标准。
- 不允许负的 LIMIT 或 OFFSET 值，而不是把它们当作零（Simon）
- 不允许在事务块之外使用 LOCK TABLE（Tom）

  这样的操作毫无用处，因为锁会被立即释放。
- 序列现在包含一个额外的 start_value 列（Zoltan Boszormenyi）

  这支持 ALTER SEQUENCE ... RESTART。

E.72.2.4. 函数和操作符

- 让 numeric 零的分数次幂返回 0 而不是抛出错误，并让 numeric 零的零次幂返回 1 而不是出错（Bruce）

  这与 float8 长期的行为一致。
- 允许浮点值的一元减产生负零（Tom）

更改后的行为更符合 IEEE 标准。

- 如果转义字符是 `LIKE` 模式的最后一个字符（即它没有可转义的内容），抛出错误（Tom）

  此前这样的转义字符被静默忽略，因此可能掩盖应用程序逻辑错误。

- 移除以前用于 `LIKE` 索引比较的 `~=~` 和 `~<>~` 操作符（Tom）

  模式索引现在使用常规的相等操作符。

- `xpath()` 现在把它的参数不做任何更改地传递给 libxml（Andrew）

  这意味着 XML 参数必须是格式良好的 XML 文档。以前的编码试图允许 XML 片段，但效果不好。

- 让 `xmlelement()` 像格式化内容值一样格式化属性值（Peter）

  此前属性值按照常规 SQL 输出行为格式化，这有时与 XML 规则相悖。

- 重写使用 libxml 的函数的内存管理（Tom）

  此变更应能避免在 PL/Perl 和其他附加代码中使用 libxml 的一些兼容性问题。

- 为哈希函数采用更快的算法（Kenneth Marshall，基于 Bob Jenkins 的工作）

  许多内建哈希函数现在在小端和大端平台上产生不同的结果。

E.72.2.4.1. 时间函数和操作符

- `DateStyle` 不再控制 `interval` 输出格式化；改为有一个新变量 `IntervalStyle`（Ron Mayer）

- 改进 `timestamp` 和 `interval` 输出中小数秒处理的一致性（Ron Mayer）

  这可能导致显示的小数位数与以前不同，或者四舍五入而不是截断。

- 让 `to_char()` 的本地化月/日名称依赖于 `LC_TIME` 而不是 `LC_MESSAGES`（Euler Taveira de Oliveira）

- 让 `to_date()` 和 `to_timestamp()` 更一致地对无效输入报告错误（Brendan Jurd）

  以前的版本常常忽略或静默误读与格式字符串不匹配的输入。这类情形现在将导致错误。

- 修复 `to_timestamp()`，使其不要求上/下午（`AM`/`PM`）和纪元（`BC`/`AD`）格式标志的大/小写匹配（Brendan Jurd）

  例如，输入值 `ad` 现在匹配格式字符串 `AD`。

E.72.3. 变更

下面你将找到 PostgreSQL 8.4 与上一个主版本之间变更的详细说明。

E.72.3.1. 性能

- 改进优化器统计信息计算（Jan Urbanski、Tom）

  特别是，全文搜索操作符的估计得到极大改进。

- 允许 SELECT DISTINCT 和 UNION/INTERSECT/EXCEPT 使用哈希（Tom）

  这意味着这类查询不再自动产生排序的输出。

- 创建半连接和反连接的显式概念（Tom）

  此工作把我们以前对 IN (SELECT ...) 子句的临时处理形式化，并将其扩展到 EXISTS 和 NOT EXISTS 子句。它应能显著改善 EXISTS 和 NOT EXISTS 查询的规划。一般而言，逻辑上等价的 IN 和 EXISTS 子句现在应有相近的性能，而以前 IN 常常胜出。

- 改进外连接之下子查询的优化（Tom）

  此前，
  如果子查询或视图出现在外连接的可空侧且其结果列表中包含非严格表达式（例如常量），
  就无法很好地优化。

- 通过使用 Boyer-Moore-Horspool 搜索改进 text_position() 及相关函数的性能（David Rowley）

  这对长搜索模式特别有帮助。

- 通过仅在请求时写入统计信息收集文件来降低其写入的 I/O 负载（Martin Pihlak）

- 改进批量插入的性能（Robert Haas、Simon）

- 把 default_statistics_target 的默认值从 10 提高到 100（Greg Sabino Mullane、Tom）

  最大值也从 1000 提高到 10000。

- 在涉及继承或 UNION ALL 的查询中默认执行 constraint_exclusion 检查（Tom）

  新增了 constraint_exclusion 设置 partition 来指定此行为。

- 允许位图索引扫描的 I/O 预读（Greg Stark）

  预读量由 effective_io_concurrency 控制。此特性仅在内核支持 posix_fadvise() 时可用。

- 在 FROM 子句中内联简单的返回集合的 SQL 函数（Richard Rowell）

- 通过为外部关系中特别常见的连接键值提供特殊情形，改进多批哈希连接的性能（Bryce Cutt、Ramon Lawrence）

- 通过抑制 “物理 tlist” 优化减少多批哈希连接中的临时数据量（Michael Henderson、Ramon Lawrence）

- 避免 CREATE INDEX CONCURRENTLY 期间等待空闲在事务中的会话（Simon）

- 改进共享缓存失效的性能（Tom）

E.72.3.2. 服务器

E.72.3.2.1. 设置

- 把许多 postgresql.conf 设置转换为枚举值，使 pg_settings 能显示有效值（Magnus）

- 新增 `cursor_tuple_fraction` 参数，控制规划器假定会被获取的游标行的比例（Robert Hell）
- 允许 `postgresql.conf` 中自定义变量类名称包含下划线（Tom）

E.72.3.2.2. 认证与安全

- 移除对（不安全的）`crypt` 认证方法的支持（Magnus）

  这实际上使 PostgreSQL 7.2 之前的客户端库过时，因为它们不再有任何可用的非明文口令方法。

- 支持 `pg_ident.conf` 中的正则表达式（Magnus）
- 允许不重启 postmaster 即可更改 Kerberos/GSSAPI 参数（Magnus）
- 支持服务器证书文件中的 SSL 证书链（Andrew Gierth）

  包含完整证书链使客户端能够验证证书，而无需在本地存储中拥有所有中间 CA 证书——商业 CA 常是这种情形。

- 对 `MD5` 认证与 `db_user_namespace` 同时启用的组合报告适当的错误消息（Bruce）

E.72.3.2.3. `pg_hba.conf`

- 把所有认证选项改为使用 `name=value` 语法（Magnus）

  这对 `ldap`、`pam` 和 `ident` 认证方法做了不兼容的变更。所有使用这些方法的 `pg_hba.conf` 条目都需要用新格式重写。

- 移除 `ident sameuser` 选项，改为在未指定 usermap 时以该行为为默认（Magnus）
- 允许所有外部认证方法使用 usermap 参数（Magnus）

  此前 usermap 只在 `ident` 认证中受支持。

- 新增 `clientcert` 选项，控制是否请求客户端证书（Magnus）

  此前这由服务器数据目录中是否存在根证书文件控制。

- 新增 `cert` 认证方法，允许通过 SSL 证书进行用户认证（Magnus）

  此前 SSL 证书只能验证客户端拥有某证书的访问权，而不能认证用户。

- 允许在 `pg_hba.conf` 中指定 `krb5`、`gssapi` 和 `sspi` 的 realm 以及 `krb5` 主机设置（Magnus）

  它们覆盖 `postgresql.conf` 中的设置。

- 为 `krb5`、`gssapi` 和 `sspi` 方法新增 `include_realm` 参数（Magnus）

  这允许使用 usermap 把来自不同 realm 的相同用户名认证为不同的数据库用户。

- 装载时完整解析 `pg_hba.conf`，以便立即报告错误（Magnus）

  此前，文件中的大多数错误直到客户端尝试连接时才会被发现，
  因此一个有错误的文件可能使系统不可用。按照新行为，如果重载期间检测到错误，
  则拒绝坏文件，postmaster 继续使用其旧副本。

- 显示 `pg_hba.conf` 中的所有解析错误，而不是在第一个之后中止（Selena Deckelmann）
- 支持在 Solaris 上通过 Unix 域套接字进行 `ident` 认证（Garick Hamlin）

E.72.3.2.4. 连续归档

- 为 `pg_start_backup()` 提供一个选项，强制其隐含的检查点尽快完成（Tom）

  默认行为避免过多的 I/O 消耗，但如果没有并发查询活动在进行，那就没有意义。

- 让 `pg_stop_backup()` 等待已修改的 WAL 文件被归档（Simon）

  这保证备份在 `pg_stop_backup()` 完成时是有效的。

- 启用归档时，在关闭时轮转最后一个 WAL 段，使所有事务可以立即归档（Guillaume Smet、Heikki）
- 当连续归档基础备份正在进行时，延迟“智能”关闭（Laurenz Albe）
- 如果请求“快速”关闭，则取消正在进行的连续归档基础备份（Laurenz Albe）
- 允许 `recovery.conf` 布尔变量接受与 `postgresql.conf` 布尔变量相同的字符串值范围（Bruce）

E.72.3.2.5. 监控

- 新增 `pg_conf_load_time()`，报告 PostgreSQL 配置文件上次装载的时间（George Gensure）
- 新增 `pg_terminate_backend()` 以安全地终止后端（`SIGTERM` 信号也可行）（Tom、Bruce）

  虽然一直可以用 `SIGTERM` 终止单个后端，但这此前被认为不受支持；对该情形的测试发现了一些现已修复的缺陷。

- 新增跟踪用户定义函数调用次数和运行时间的能力（Martin Pihlak）

  函数统计信息出现在新系统视图 `pg_stat_user_functions` 中。跟踪由新参数 `track_functions` 控制。

- 通过新参数 `track_activity_query_size` 允许指定 `pg_stat_activity` 中查询字符串的最大大小（Thomas Lee）
- 增加发送到 syslog 的最大行长度，希望以此改进性能（Tom）
- 新增只读配置变量 `segment_size`、`wal_block_size` 和 `wal_segment_size`（Bernd Helmle）
- 报告死锁时，把死锁涉及的所有查询的文本报告给服务器日志（Itagaki Takahiro）
- 新增 `pg_stat_get_activity(pid)` 函数，返回特定进程 id 的信息（Magnus）
- 允许通过 `stats_temp_directory` 指定服务器统计信息文件的位置（Magnus）

  这允许把统计信息文件放在 RAM 驻留目录中以降低 I/O 需求。在启动/关闭时，该文件被复制到其传统位置（`$PGDATA/global/`），因此跨重启保留。

E.72.3.3. 查询

- 新增对 `WINDOW` 函数的支持（Hitoshi Harada）
- 新增对 `WITH` 子句（CTE）的支持，包括 `WITH RECURSIVE`（Yoshiyuki Asaba、Tatsuo Ishii、Tom）
- 新增 `TABLE` 命令（Peter）

  `TABLE tablename` 是 `SELECT * FROM tablename` 的 SQL 标准简写。
- 允许在指定 `SELECT`（或 `RETURNING`）列输出标签时省略 `AS`（Hiroshi Saito）

  只要列标签不是任何 PostgreSQL 关键字，这就能工作；否则仍需要 `AS`。
- 支持 `SELECT` 结果列表中的返回集合的函数，包括通过 tuplestore 返回结果的函数（Tom）

  特别是，这意味着用 PL/pgSQL 和其他 PL 语言编写的函数现在可以这样调用。
- 支持聚合和分组查询输出中的返回集合的函数（Tom）
- 允许 `SELECT FOR UPDATE/SHARE` 在继承树上工作（Tom）
- 为 SQL/MED 新增基础设施（Martin Pihlak、Peter）

  目前还没有远程或外部 SQL/MED 能力，但此变更为管理 `dblink` 和 `plproxy` 等模块的连接信息提供了一个标准化且面向未来的系统。
- 当被引用的模式、函数、操作符或操作符类被修改时使缓存计划失效（Martin Pihlak、Tom）

  这改进了系统对即时 DDL 变更的响应能力。
- 允许组合类型的比较，并允许匿名组合类型的数组（Tom）

  这允许诸如 `row(1, 1.1) = any (array[row(7, 7.7), row(1, 1.0)])` 的构造。这在递归查询中特别有用。
- 新增对用码点指定 Unicode 字符串字面量和标识符的支持，例如 `U&'d\0061t\+000061'`（Peter）
- 拒绝字符串字面量和 `COPY` 数据中的 `\000`（Tom）

  此前这被接受，但效果是终止字符串内容。
- 改进解析器报告错误位置的能力（Tom）

  现在许多此前无法定位的语义错误（如数据类型不匹配）也会报告错误位置。

E.72.3.3.1. TRUNCATE

- 支持语句级 `ON TRUNCATE` 触发器（Simon）
- 为 `TRUNCATE TABLE` 新增 `RESTART/CONTINUE IDENTITY` 选项（Zoltan Boszormenyi）

  序列的起始值可以通过 `ALTER SEQUENCE START WITH` 更改。
- 允许 `TRUNCATE tab1, tab1` 成功（Bruce）

- 新增单独的 TRUNCATE 权限（Robert Haas）

E.72.3.3.2. EXPLAIN

- 让 EXPLAIN VERBOSE 显示每个计划节点的输出列（Tom）

  此前 EXPLAIN VERBOSE 输出查询计划的内部表示。（该行为现在可通过 debug_print_plan 获得。）

- 让 EXPLAIN 用各自的标签标识子计划和 initplan（Tom）
- 让 EXPLAIN 遵循 debug_print_plan（Tom）
- 允许对 CREATE TABLE AS 使用 EXPLAIN（Peter）

E.72.3.3.3. LIMIT/OFFSET

- 允许 LIMIT 和 OFFSET 中的子查询（Tom）
- 为 LIMIT/OFFSET 能力新增 SQL 标准语法（Peter）

  即 OFFSET num {ROW|ROWS} FETCH {FIRST|NEXT} [num] {ROW|ROWS} ONLY。

E.72.3.4. 对象操作

- 新增对列级权限的支持（Stephen Frost、KaiGai Kohei）
- 重构多对象 DROP 操作以减少对 CASCADE 的需要（Alex Hunsaker）

  例如，如果表 B 依赖于表 A，命令 DROP TABLE A，B 不再需要 CASCADE 选项。

- 通过确保在开始删除对象依赖之前获取锁，修复并发 DROP 命令的各种问题（Tom）
- 改进 DROP 命令期间依赖的报告（Tom）
- 按照 SQL 标准为 CREATE TABLE AS 新增 WITH [NO] DATA 子句（Peter、Tom）
- 新增对用户定义 I/O 转换转换的支持（Heikki）
- 允许 CREATE AGGREGATE 使用 internal 转换数据类型（Tom）
- 为 CREATE TYPE 新增 LIKE 子句（Tom）

  这简化了创建与现有类型使用相同内部表示的数据类型。

- 允许为用户定义基本类型指定类型类别和“首选”状态（Tom）

  这允许对用户定义类型的强制转换行为有更多控制。

- 允许 CREATE OR REPLACE VIEW 向视图末尾添加列（Robert Haas）

E.72.3.4.1. ALTER

- 新增 ALTER TYPE RENAME（Petr Jelinek）
- 新增 ALTER SEQUENCE ... RESTART（不带参数），把序列重置为其初始值（Zoltan Boszormenyi）

- 修改 `ALTER TABLE` 语法，允许对表、索引、序列和视图的所有合理组合（Tom）

  此变更允许以下新语法：

  - `ALTER SEQUENCE OWNER TO`
  - `ALTER VIEW ALTER COLUMN SET/DROP DEFAULT`
  - `ALTER VIEW OWNER TO`
  - `ALTER VIEW SET SCHEMA`

  这里并没有实际的新功能，但以前你必须使用 `ALTER TABLE` 来做这些事情，这容易引起混淆。

- 新增对语法 `ALTER TABLE ... ALTER COLUMN ... SET DATA TYPE` 的支持（Peter）

  这是已支持功能的 SQL 标准语法。

- 让 `ALTER TABLE SET WITHOUT OIDS` 重写表以物理移除 `OID` 值（Tom）

  另外，新增 `ALTER TABLE SET WITH OIDS` 以重写表来添加 `OID`。

E.72.3.4.2. 数据库操作

- 当未提交的预备事务是原因时，改进 `CREATE/DROP/RENAME DATABASE` 失败的报告（Tom）
- 把 `LC_COLLATE` 和 `LC_CTYPE` 改为逐数据库的设置（Radek Strnad、Heikki）

  这使排序规则类似于编码——后者一直可以逐数据库配置。

- 改进对数据库编码、排序规则（`LC_COLLATE`）和字符分类（`LC_CTYPE`）匹配的检查（Heikki、Tom）

  特别注意，新数据库的编码和区域设置只有在从 `template0` 复制时才能更改。这防止可能复制与设置不匹配的数据。

- 新增 `ALTER DATABASE SET TABLESPACE`，把数据库移到新表空间（Guillaume Lelarge、Bernd Helmle）

E.72.3.5. 工具操作

- 为 `CLUSTER` 命令和 clusterdb 新增 `VERBOSE` 选项（Jim Cox）
- 降低记录待处理触发器事件的内存需求（Tom）

E.72.3.5.1. 索引

- 显著改进构建和访问哈希索引的速度（Tom Raney、Shreya Bhargava）

  这使哈希索引有时可以比 btree 索引更快。但哈希索引仍然不是崩溃安全的。

- 让哈希索引只存储哈希码，而不是被索引列的完整值（Xiao Meng）

  这大幅减小长索引值下哈希索引的大小，改进了性能。

- 实现 GIN 索引的快速更新选项（Teodor、Oleg）

此选项以搜索速度的小幅损失大幅改进更新速度。

- `xxx_pattern_ops` 索引现在可用于简单相等比较，而不仅用于 `LIKE`（Tom）

E.72.3.5.2. 全文索引

- 移除在全文本索引上进行 GIN 加权查找时必须使用 `@@@` 的要求（Tom、Teodor）

  可以改用常规的 `@@` 文本搜索操作符。

- 为 `@@` 文本搜索操作新增优化器选择率函数（Jan Urbanski）
- 允许全文搜索中的前缀匹配（Teodor Sigaev、Oleg Bartunov）
- 支持多列 GIN 索引（Teodor Sigaev）
- 改进对尼泊尔语和天城文的支持（Teodor）

E.72.3.5.3. VACUUM

- 在单独的逐关系“分叉”文件中跟踪空闲空间（Heikki）

  `VACUUM` 发现的空闲空间现在记录在 `*_fsm` 文件中，而不是固定大小的共享内存区域。`max_fsm_pages` 和 `max_fsm_relations` 设置已被移除，大幅简化了空闲空间管理。

- 新增可见性映射来跟踪不需要清理的页（Heikki）

  这允许 `VACUUM` 在只有表的一部分需要清理时避免扫描整表。可见性映射存储在逐关系的“分叉”文件中。

- 新增 `vacuum_freeze_table_age` 参数，控制 `VACUUM` 何时应忽略可见性映射并做全表扫描以冻结元组（Heikki）
- 更仔细地跟踪事务快照（Alvaro）

  这改进了 `VACUUM` 在存在长运行事务时回收空间的能力。

- 新增在 `CREATE TABLE` 中指定逐关系 autovacuum 和 TOAST 参数的能力（Alvaro、Euler Taveira de Oliveira）

  Autovacuum 选项以前存储在系统表中。

- 为 vacuumdb 新增 `--freeze` 选项（Bruce）

E.72.3.6. 数据类型

- 为文本搜索同义词词典新增 `CaseSensitive` 选项（Simon）
- 改进 `NUMERIC` 除法的精度（Tom）
- 新增 `int2` 与 `int8` 的基本算术操作符（Tom）

  这消除了某些情况下显式转换的需要。

- 允许 `UUID` 输入在每四个数字后接受可选的连字符（Robert Haas）
- 允许 `on/off` 作为布尔数据类型的输入（Itagaki Takahiro）

- 允许 `numeric` 类型输入字符串中 `NaN` 周围有空格（Sam Mason）

E.72.3.6.1. 时间数据类型

- 拒绝年份 `0 BC` 以及 `000` 和 `0000`（Tom）

  此前它们被解释为 `1 BC`。（注意：年份 `0` 和 `00` 仍被假定为 2000 年。）

- 把 `SGT`（新加坡时间）加入已知时区缩写的默认列表（Tom）

- 支持 `infinity` 和 `-infinity` 作为 `date` 类型的值（Tom）

- 使 `interval` 字面量的解析更符合标准（Tom、Ron Mayer）

  例如，`INTERVAL '1' YEAR` 现在做它本应做的事。

- 允许 `interval` 的小数秒精度指定在 `second` 关键字之后，以符合 SQL 标准（Tom）

  此前精度必须指定在 `interval` 关键字之后。（为向后兼容，该语法仍受支持，但已废弃。）数据类型定义现在将使用标准格式输出。

- 支持 IS0 8601 `interval` 语法（Ron Mayer、Kevin Grittner）

  例如，`INTERVAL 'P1Y2M3DT4H5M6.7S'` 现在受支持。

- 新增 `IntervalStyle` 参数，控制 `interval` 值的输出方式（Ron Mayer）

  有效值为：`postgres`、`postgres_verbose`、`sql_standard`、`iso_8601`。
  此设置还控制只有部分字段带正/负标志的负 `interval` 输入的处理。

- 改进 `timestamp` 和 `interval` 输出中小数秒处理的一致性（Ron Mayer）

E.72.3.6.2. 数组

- 改进应用于 `ARRAY[]` 构造的转换的处理，例如 `ARRAY[...]::integer[]`（Brendan Jurd）

  此前 PostgreSQL 试图在不参考后续转换的情况下确定 `ARRAY[]` 构造的数据类型。这在许多情况下会不必要地失败，特别是当 `ARRAY[]` 构造为空或只包含 `NULL` 等含糊条目时。现在会参考转换来确定数组元素必须的类型。

- 让 SQL 语法的 `ARRAY` 维度可选，以匹配 SQL 标准（Peter）

- 新增 `array_ndims()`，返回数组的维数（Robert Haas）

- 新增 `array_length()`，返回数组指定维度的长度（Jim Nasby、Robert Haas、Peter Eisentraut）

- 新增聚合函数 `array_agg()`，把所有聚合值作为单个数组返回（Robert Haas、Jeff Davis、Peter）

- 新增 `unnest()`，把数组转换为单独的行值（Tom）

  它与 `array_agg()` 相反。

- 新增 `array_fill()`，创建用某值初始化的数组（Pavel Stehule）

- 新增 `generate_subscripts()`，简化生成数组下标范围（Pavel Stehule）

E.72.3.6.3. Wide-Value Storage (TOAST)

- 对短至 32 字节的值考虑 TOAST 压缩（以前是 256 字节）（Greg Stark）
- 要求至少 25% 的空间节省才使用 TOAST 压缩（以前小值为 20%，大值为有任何节省即可）（Greg）
- 改进对混合大小可 TOAST 字段的行的 TOAST 启发式，使我们倾向于把大值推出行外并且不必要地压缩小值（Greg、Tom）

E.72.3.7. 函数

- 文档说明 `setseed()` 允许 `-1` 到 `1` 的值（不只是 `0` 到 `1`），并强制有效范围（Kris Jurka）
- 新增服务器端函数 `lo_import(filename, oid)`（Tatsuo）
- 新增 `quote_nullable()`，行为类似 `quote_literal()`，但对空参数返回字符串 `NULL`（Brendan Jurd）
- 改进全文搜索的 `headline()` 函数，允许提取多个文本片段（Sushant Sinha）
- 新增 `suppress_redundant_updates_trigger()` 触发器函数，避免非数据更改更新的开销（Andrew）
- 新增 `div(numeric, numeric)`，执行不舍入的 `numeric` 除法（Tom）
- 新增 `generate_series()` 的 `timestamp` 和 `timestamptz` 版本（Hitoshi Harada）

E.72.3.7.1. 对象信息函数

- 实现 `current_query()`，供需要知道当前运行查询的函数使用（Tomas Doran）
- 新增 `pg_get_keywords()`，返回解析器关键字列表（Dave Page）
- 新增 `pg_get_functiondef()`，查看函数的定义（Abhijit Menon-Sen）
- 当反解析不含变量的表达式时，允许 `pg_get_expr()` 的第二个参数为零（Tom）
- 修改 `pg_relation_size()` 以使用 `regclass`（Heikki）

  `pg_relation_size(data_type_name)` 不再有效。
- 向 `pg_settings` 输出新增 `boot_val` 和 `reset_val` 列（Greg Smith）
- 向 `pg_settings` 输出为在配置文件中设置的变量新增源文件名和行号列（Magnus、Alvaro）

  出于安全原因，这些列只有超级用户可见。
- 新增对 `CURRENT_CATALOG`、`CURRENT_SCHEMA`、`SET CATALOG`、`SET SCHEMA` 的支持（Peter）

  它们为现有特性提供 SQL 标准语法。
- 新增 `pg_typeof()`，返回任意值的数据类型（Brendan Jurd）

- 让 `version()` 返回服务器是 32 位还是 64 位二进制的信息（Bruce）
- 修复信息模式列 `is_insertable_into` 和 `is_updatable` 的行为使其一致（Peter）
- 改进信息模式 `datetime_precision` 列的行为（Peter）

  这些列现在对 `date` 列显示零，对未声明精度的 `time`、`timestamp` 和 `interval` 显示 6（默认精度），而不是像以前那样显示空值。
- 把剩余的内建返回集合的函数转换为使用 `OUT` 参数（Jaime Casanova）

  这使调用这些函数无需指定列列表成为可能：`pg_show_all_settings()`、`pg_lock_status()`、`pg_prepared_xact()`、`pg_prepared_statement()`、`pg_cursor()`
- 让 `pg_*_is_visible()` 和 `has_*_privilege()` 函数对无效 OID 返回 `NULL`，而不是报告错误（Tom）
- 扩展 `has_*_privilege()` 函数，允许一次调用查询多个权限的 OR（Stephen Frost、Tom）
- 新增 `has_column_privilege()` 和 `has_any_column_privilege()` 函数（Stephen Frost、Tom）

E.72.3.7.2. 函数创建

- 支持可变参数函数（参数数量可变的函数）（Pavel Stehule）

  只有尾部参数可以是可选的，且它们必须都是同一数据类型。
- 支持函数参数的默认值（Pavel Stehule）
- 新增 `CREATE FUNCTION ... RETURNS TABLE` 子句（Pavel Stehule）
- 允许 SQL 语言函数返回 `INSERT/UPDATE/DELETE` 的 `RETURNING` 子句的输出（Tom）

E.72.3.7.3. PL/pgSQL 服务器端语言

- 支持 `EXECUTE USING`，便于把数据值插入动态查询字符串（Pavel Stehule）
- 允许用 `FOR` 循环遍历游标的结果（Pavel Stehule）
- 支持 `RETURN QUERY EXECUTE`（Pavel Stehule）
- 改进 `RAISE` 命令（Pavel Stehule）
  - 支持 `DETAIL` 和 `HINT` 字段
  - 支持指定 `SQLSTATE` 错误码
  - 支持异常名参数
  - 允许异常块中不带参数的 `RAISE` 重新抛出当前错误
- 允许在 `EXCEPTION` 列表中指定 `SQLSTATE` 码（Pavel Stehule）

  这对处理自定义 `SQLSTATE` 码很有用。
- 支持 `CASE` 语句（Pavel Stehule）

- 让 RETURN QUERY 设置特殊的 FOUND 和 GET DIAGNOSTICS 的 ROW_COUNT 变量（Pavel Stehule）
- 让 FETCH 和 MOVE 设置 GET DIAGNOSTICS 的 ROW_COUNT 变量（Andrew Gierth）
- 让不带标签的 EXIT 总是退出最内层循环（Tom）

  此前，如果存在比任何循环都更贴近嵌套的 BEGIN 块，它会退出那个块。新行为与 Oracle(TM) 匹配，也与我们自己文档此前的说法一致。
- 使字符串字面量和嵌套块注释的处理与主 SQL 解析器的处理一致（Tom）

  特别是，RAISE 中的格式字符串现在与任何其他字符串字面字面量一样工作，包括受 standard_conforming_strings 约束。此变更还修复了 standard_conforming_strings 开启时有效命令会失败的其他情形。
- 避免同一函数在不同异常块嵌套深度被调用时的内存泄漏（Tom）

E.72.3.8. 客户端应用程序

- 修复 pg_ctl restart 以保留命令行参数（Bruce）
- 新增 -w/--no-password 选项，在所有具有 -W/--password 选项的工具中阻止口令提示（Peter）
- 移除 createdb、createuser、dropdb、dropuser 的 -q（安静）选项（Peter）

  这些选项自 PostgreSQL 8.3 起就没有效果。

E.72.3.8.1. psql

- 移除冗长的启动横幅；现在只建议 help（Joshua Drake）
- 让 help 显示常用反斜杠命令（Greg Sabino Mullane）
- 新增 \pset format wrapped 模式，把输出折行到屏幕宽度，或者在设置了 \pset columns 时也适用于文件/管道输出（Bryce Nesbitt）
- 允许 \pset 中布尔值的所有受支持拼法，而不只是 on 和 off（Bruce）

  此前，除 “off” 外的任何字符串都被静默地当作 true。psql 现在会对无法识别的拼法抱怨（但仍把它们当作 true）。
- 对宽输出使用分页器（Bruce）
- 要求单字母反斜杠命令与其第一个参数之间有一个空格（Bernd Helmle）

  这消除了一个历史性的歧义来源。
- 改进对模式限定和带引号标识符的 Tab 补全支持（Greg Sabino Mullane）
- 为 \timing 新增可选的 on/off 参数（David Fetter）
- 在多行上显示访问控制权限（Brendan Jurd、Andreas Scherbaum）
- 让 \l 显示数据库访问权限（Andrew Gilligan）

- 让 `\l+` 显示数据库大小（如果权限允许）（Andrew Gilligan）
- 新增 `\ef` 命令来编辑函数定义（Abhijit Menon-Sen）

E.72.3.8.2. psql \d* commands

- 让没有模式参数的 `\d*` 命令只在指定 `S` 修饰符时显示系统对象（Greg Sabino Mullane、Bruce）

  以前的行为在不同的 `\d` 变体之间不一致，而且在大多数情况下没有提供只查看用户对象的简便方法。

- 改进 `\d*` 命令，使其能用于较旧的 PostgreSQL 服务器版本（可追溯到 7.4），而不仅是当前服务器版本（Guillaume Lelarge）
- 让 `\d` 显示引用所选表的外键约束（Kenneth D'Souza）
- 让 `\d` 对序列显示其列值（Euler Taveira de Oliveira）
- 向 `\d+` 显示中添加列存储类型和其他关系选项（Gregory Stark、Euler Taveira de Oliveira）
- 在 `\dt+` 输出中显示关系大小（Dickson S. Guedes）
- 在 `\dT+` 中显示 `enum` 类型的可能值（David Fetter）
- 允许 `\dC` 接受通配符模式，匹配转换涉及的任一数据类型（Tom）
- 向 `\df` 的输出新增函数类型列，并新增只列出所选类型函数的选项（David Fetter）
- 让 `\df` 不隐藏接受或返回 `cstring` 类型的函数（Tom）

  此前这类函数被隐藏，因为它们大多是数据类型 I/O 函数，被认为无趣。默认隐藏系统函数的新策略使这个瑕疵不再必要。

E.72.3.8.3. pg_dump

- 为 pg_dump/pg_dumpall/pg_restore 新增 `--no-tablespaces` 选项，使转储可以恢复到表空间布局不匹配的集群（Gavin Roy）
- 移除 pg_dump 和 pg_dumpall 的 `-d` 和 `-D` 选项（Tom）

  这些选项太容易与其他 PostgreSQL 客户端应用程序中选择数据库名的选项混淆。功能仍然可用，但现在必须拼出长选项名 `--inserts` 或 `--column-inserts`。

- 移除 pg_dump 和 pg_dumpall 的 `-i`/`--ignore-version` 选项（Tom）

  使用此选项不会抛出错误，但它没有效果。移除此选项是因为版本检查对安全是必要的。

- 在转储和恢复期间禁用 `statement_timeout`（Joshua Drake）
- 新增 pg_dump/pg_dumpall 选项 `--lock-wait-timeout`（David Gould）

  这允许转储在指定时间内无法获得共享锁时失败。

- 重排 pg_dump 的 `--data-only` 输出，把被外键引用的表排在引用表之前转储（Tom）

  这允许在外键已存在时加载数据。如果循环引用使安全排序不可能，则发出 `NOTICE`。

- 允许 pg_dump、pg_dumpall 和 pg_restore 使用指定的角色（Benedek László）
- 允许 pg_restore 使用多个并发连接执行恢复（Andrew）

  并发连接数由选项 `--jobs` 控制。这只对自定义格式归档受支持。

E.72.3.9. 编程工具

E.72.3.9.1. libpq

- 允许在导入大对象时指定 `OID`，通过新函数 `lo_import_with_oid()`（Tatsuo）
- 新增“事件”支持（Andrew Chernow、Merlin Moncure）

  这新增了注册回调来管理与 `PGconn` 和 `PGresult` 对象关联的私有数据的能力。
- 改进错误处理，允许多条错误消息作为多行错误报告返回（Magnus）
- 让 `PQexecParams()` 及相关函数对空查询返回 `PGRES_EMPTY_QUERY`（Tom）

  它们以前返回 `PGRES_COMMAND_OK`。
- 记录如何避免 Windows 上 `WSACleanup()` 的开销（Andrew Chernow）
- 不依赖 Kerberos 票据来确定默认数据库用户名（Magnus）

  此前，支持 Kerberos 的 libpq 构建会把任何可用 Kerberos 票据中的主体名用作默认数据库用户名，即使连接没有使用 Kerberos 认证。这被认为不一致且令人困惑。现在默认用户名的确定方式在有或没有 Kerberos 时相同。但注意，使用 Kerberos 认证时数据库用户名仍必须与票据匹配。

E.72.3.9.2. libpq SSL (Secure Sockets Layer) support

- 修复 SSL 连接的证书验证（Magnus）

  libpq 现在支持在建立 SSL 连接时同时验证证书和服务器名称。如果没有可用于验证的根证书，SSL 连接将失败。`sslmode` 参数用于启用证书验证并设置检查级别。默认仍是不做任何验证，允许连接到启用 SSL 的服务器而不要求客户端上有根证书。
- 支持通配符服务器证书（Magnus）

  如果证书 CN 以 `*` 开头，匹配主机名时它将被视为通配符，允许同一证书用于多个服务器。
- 允许指定客户端证书的文件位置（Mark Woodward、Alvaro、Magnus）
- 新增 `PQinitOpenSSL` 函数，允许对 OpenSSL/libcrypto 初始化有更多控制（Andrew Chernow）
- 当没有数据库连接保持打开时，让 libpq 注销其 OpenSSL 回调（Bruce、Magnus、Russell Smith）

  这对卸载 libpq 库的应用程序是必需的，否则无效的 OpenSSL 回调将残留。

E.72.3.9.3. ecpg

- 新增消息的本地化支持（Euler Taveira de Oliveira）

- ecpg 解析器现在从服务器解析器自动生成（Michael）

  此前 ecpg 解析器是手工维护的。

E.72.3.9.4. Server Programming Interface (SPI)

- 新增对外线参数的单次使用计划的支持（Tom）
- 为 `SPI_execute()` 新增 `SPI_OK_REWRITTEN` 返回码（Heikki）

  这在命令被重写为另一种类型的命令时使用。

- 从 `executor/spi.h` 中移除不必要的包含（Tom）

  使用 SPI 的模块如果依赖 `spi.h` 替它们包含内容，可能需要添加一些 `#include` 行。

E.72.3.10. 构建选项

- 更新构建系统以使用 Autoconf 2.61（Peter）
- 源代码构建需要 GNU bison（Peter）

  这实际上多年来已是必需，但现在不再有声称支持其他解析器工具的基础设施。

- 新增 pg_config 的 `--htmldir` 选项（Peter）
- 在服务器内部按值传递 `float4`（Zoltan Boszormenyi）

  新增 configure 选项 `--disable-float4-byval` 以使用旧行为。使用旧式（版本 0）调用约定并传递或返回 `float4` 值的外部 C 函数将被此变更破坏，因此如果有这样的函数且不想更新它们，可能需要该 configure 选项。

- 在 64 位平台上，服务器内部按值传递 `float8`、`int8` 及相关数据类型（Zoltan Boszormenyi）

  新增 configure 选项 `--disable-float8-byval` 以使用旧行为。如上所述，此变更可能破坏旧式外部 C 函数。

- 新增 configure 选项 `--with-segsize`、`--with-blocksize`、`--with-wal-blocksize`、`--with-wal-segsize`（Zdenek Kotala、Tom）

  这简化了对几个常量的构建期控制，这些常量以前只能通过编辑 `pg_config_manual.h` 更改。

- 允许在 Solaris 2.5 上进行线程化构建（Bruce）
- 在 Solaris 上使用系统的 `getopt_long()`（Zdenek Kotala、Tom）

  这使选项处理更符合 Solaris 用户的期望。

- 新增对 Linux 上 Sun Studio 编译器的支持（Julius Stroffek）
- 把主版本号附加到后端 gettext 域，把 `soname` 主版本号附加到库的 gettext 域（Peter）

  这简化了多个版本的并行安装。

- 新增用 gcov 进行代码覆盖率测试的支持（Michelle Caisse）

- 允许在 Mingw 和 Cygwin 上进行树外构建（Richard Evans）
- 修复把 Mingw 用作交叉编译源平台的问题（Peter）

E.72.3.11. 源代码

- 支持 64 位时区数据文件（Heikki）

  这新增了对 2038 年以后的夏令时（DST）计算的支持。

- 弃用平台 `time_t` 数据类型的使用（Tom）

  一些平台已迁移到 64 位 `time_t`，一些没有，而 Windows 拿不定主意。定义 `pg_time_t`，其含义与 `time_t` 相同，但始终为 64 位（除非平台没有 64 位整数类型），并在所有模块 API 和磁盘上数据格式中使用该类型。

- 修复交叉编译时处理时区数据库的缺陷（Richard Evans）
- 一步链接后端目标文件，而不是分阶段（Peter）
- 改进 gettext 支持以允许更好地翻译复数（Peter）
- 为 PL 语言新增消息翻译支持（Alvaro、Peter）
- 新增更多 DTrace 探针（Robert Lor）
- 在 Mac OS X Leopard 和其他非 Solaris 平台上启用 DTrace 支持（Robert Lor）
- 通过提供通用函数来简化和标准化 C 字符串与 `text` datum 之间的转换（Brendan Jurd、Tom）
- 清理 `include/catalog/` 头文件，使前端程序可以包含它们而不必包含 `postgres.h`（Zdenek Kotala）
- 让 `name` 按 char 对齐，并抑制索引中 `name` 条目的零填充（Tom）
- 如果动态装载的代码执行 `exit()`，更好地恢复（Tom）
- 新增钩子让插件监控执行器（Itagaki Takahiro）
- 新增钩子，允许覆盖规划器的统计信息查找行为（Simon Riggs）
- 为自定义共享内存需求新增 `shmem_startup_hook()`（Tom）
- 把索引访问方法 `amgetmulti` 入口点替换为 `amgetbitmap`，并扩展 `amgettuple` 的 API 以支持运行时确定操作符有损性（Heikki、Tom、Teodor）

  GIN 和 GiST 操作符类 `consistent` 函数的 API 也被扩展了。

- 新增 GIN 索引中部分匹配搜索的支持（Teodor Sigaev、Oleg Bartunov）
- 把 `pg_class` 列 `reltriggers` 替换为布尔 `relhastriggers`（Simon）

  同时移除未使用的 `pg_class` 列 `relukeys`、`relfkeys` 和 `relrefs`。

- 向 `pg_class` 新增 `relistemp` 列，便于识别临时表（Tom）
- 把平台 FAQ 移入主文档（Peter）

- 防止解析器输入文件在有冲突的状态下构建（Peter）
- 新增对 KOI8U（乌克兰语）编码的支持（Peter）
- 新增日语消息翻译（Japan PostgreSQL Users Group）

  这以前作为独立项目维护。
- 修复在 MSVC 构建的系统上设置 LC_MESSAGES 的问题（Hiroshi Inoue、Hiroshi Saito、Magnus）

E.72.3.12. 附加模块

- 新增 contrib/auto_explain，对超过指定时长的查询自动运行 EXPLAIN（Itagaki Takahiro、Tom）
- 新增 contrib/btree_gin，允许 GIN 索引处理更多数据类型（Oleg、Teodor）
- 新增 contrib/citext，提供不区分大小写、多字节感知的文本数据类型（David Wheeler）
- 新增 contrib/pg_stat_statements，用于服务器范围的语句执行统计跟踪（Itagaki Takahiro）
- 为 contrib/pgbench 新增时长和查询模式选项（Itagaki Takahiro）
- 让 contrib/pgbench 使用表名 pgbench_accounts、pgbench_branches、pgbench_history 和 pgbench_tellers，而不只是 accounts、branches、history 和 tellers（Tom）

  这是为了降低运行 pgbench 时意外破坏真实数据的风险。
- 修复 contrib/pgstattuple 以处理超过 20 亿页的表和索引（Tatsuhito Kasahara）
- 在 contrib/fuzzystrmatch 中，新增允许用户指定插入、删除和替换代价的 Levenshtein 字符串距离函数版本（Volkan Yazici）
- 让 contrib/ltree 支持多字节编码（laser）
- 使 contrib/dblink 能使用存储在 SQL/MED 目录中的连接信息（Joe Conway）
- 改进 contrib/dblink 对来自远程服务器错误的报告（Joe Conway）
- 让 contrib/dblink 设置 client_encoding 以匹配本地数据库的编码（Joe Conway）

  这防止与使用不同编码的远程数据库通信时的编码问题。
- 确保 contrib/dblink 使用用户提供的口令，而不是意外地从服务器的 .pgpass 文件获取（Joe Conway）

  这是一个小的安全增强。
- 为 contrib/pageinspect 新增 fsm_page_contents()（Heikki）
- 修改 get_raw_page() 以支持空闲空间映射（*_fsm）文件。同时更新 contrib/pg_freespacemap。
- 为 contrib/pg_trgm 新增多字节编码支持（Teodor）

- 重写 `contrib/intagg` 以使用新函数 `array_agg()` 和 `unnest()`（Tom）
- 让 `contrib/pg_standby` 在故障转移前恢复所有可用 WAL（Fujii Masao、Simon、Heikki）

  为使这安全工作，现在需要在 `recovery.conf` 中设置新的 `recovery_end_command` 选项，以便在故障转移后清理触发器文件。pg_standby 将不再自行删除触发器文件。
- 现在 `contrib/pg_standby` 的 `-l` 选项是无操作，因为使用符号链接不安全（Simon）

E.73. 版本 8.3.23

发行日期
2013-02-07

本次发布包含对 8.3.22 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

这预计是 PostgreSQL 8.3.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.73.1. 迁移到版本 8.3.23

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.73.2. 变更

- 防止从 SQL 执行 `enum_recv`（Tom Lane）

  该函数声明有误，允许一条简单的 SQL 命令使服务器崩溃。原则上攻击者可能利用它检查服务器内存的内容。感谢 Sumit Soni（通过 Secunia SVCRP）报告此问题。（CVE-2013-0255）
- 修复 SQL 文法，允许对子 SELECT 结果做下标或字段选择（Tom Lane）
- 防止扫描 `pg_tablespace` 时的竞态条件（Stephen Frost、Tom Lane）

  如果 `pg_tablespace` 条目存在并发更新，`CREATE DATABASE` 和 `DROP DATABASE` 可能行为异常。
- 防止 `DROP OWNED` 试图删除整个数据库或表空间（Álvaro Herrera）

  为安全起见，这些对象的所有权必须重新分配，而不是删除。
- 防止 `RowExpr` 或 `XmlExpr` 被二次解析分析时的不当行为（Andres Freund、Tom Lane）

  这一失误在 `CREATE TABLE LIKE INCLUDING INDEXES` 等上下文中可能被用户看到。
- 增强对哈希表大小计算中整数溢出的防御（Jeff Davis）
- 确保在 Windows 上非 ASCII 提示符字符串被转换到正确的代码页（Alexander Law、Noah Misch）

这一缺陷影响 psql 和其他一些客户端程序。

- 修复未连接数据库时 psql 的 `\?` 命令可能崩溃的问题（Meng Qingzhong）
- 修复 libpq 的 `PQprintTuples` 中的一字节缓冲区越界（Xi Wang）

  这个古老的函数没有被 PostgreSQL 本身在任何地方使用，但仍可能被某些客户端代码使用。
- 重新排列 configure 对系统提供函数的测试，使其不被 libedit/libreadline 的虚假导出所欺骗（Christoph Berg）
- 确保 Windows 构建号随时间递增（Magnus Hagander）
- 让 pgxs 在为 Windows 交叉编译时以正确的 `.exe` 后缀构建可执行文件（Zoltan Boszormenyi）
- 新增时区缩写 `FET`（Tom Lane）

  它现在用于一些东欧时区。

E.74. 版本 8.3.22

发行日期
2012-12-06

本次发布包含对 8.3.21 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

PostgreSQL 社区将于 2013 年 2 月停止对 8.3.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.74.1. 迁移到版本 8.3.22

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.74.2. 变更

- 修复与 `CREATE INDEX CONCURRENTLY` 相关的多个缺陷（Andres Freund、Tom Lane）

  修复 `CREATE INDEX CONCURRENTLY`，使其在更改索引 `pg_index` 行的状态时使用就地更新。这防止了可能导致并发会话错过更新目标索引（从而产生损坏的并发创建索引）的竞态条件。

  另外，修复各种其他操作，确保它们忽略由失败的 `CREATE INDEX CONCURRENTLY` 命令产生的无效索引。其中最重要的是 `VACUUM`，因为在采取纠正措施修复或删除无效索引之前，很容易就有自动清理在该表上启动。
- 避免内存不足时内部哈希表损坏（Hitoshi Harada）
- 修复外连接之上非严格等值子句的规划（Tom Lane）

规划器可能从把非严格构造与别的值等同的子句中推导出错误的约束，例如当 `foo` 来自外连接的可空侧时的 `WHERE COALESCE(foo, 0) = 0`。

- 改进规划器从等价类证明排他约束的能力（Tom Lane）
- 修复哈希子计划中的部分行匹配，使其正确处理跨类型情况（Tom Lane）

  这影响多列 `NOT IN` 子计划，例如 `WHERE (a, b) NOT IN (SELECT x, y FROM ...)`，当 `b` 和 `y` 分别是 `int4` 和 `int8` 时。根据所涉及的具体数据类型，这一失误会导致错误答案或崩溃。

- 在为 `AFTER ROW UPDATE/DELETE` 触发器重新获取旧元组时获取缓冲区锁（Andres Freund）

  在非常罕见的情况下，这一疏忽可能导致向外键强制触发器的预检查逻辑传递不正确的数据。这可能导致崩溃，或者在是否触发触发器上做出错误决定。

- 修复 `REASSIGN OWNED` 对表空间授权的处理（Álvaro Herrera）
- 忽略视图中系统列的不正确 `pg_attribute` 条目（Tom Lane）

  视图没有任何系统列。然而，我们在把表转换为视图时忘记移除这类条目。这一问题在 9.3 及更高版本中已正确修复，但在以前的分支中，我们需要防御已经错误转换的视图。

- 修复规则打印，使其正确转储 `INSERT INTO` *`table`* `DEFAULT VALUES`（Tom Lane）
- 防止当查询中有过多 `UNION/INTERSECT/EXCEPT` 子句时栈溢出（Tom Lane）
- 防止用最小可能整数值除以 -1 时依赖平台的失败（Xi Wang、Tom Lane）
- 修复日期解析中可能越过字符串末尾的访问（Hitoshi Harada）
- 当 Unix 域套接字的路径名长度超过平台特定限制时，生成可理解的错误消息（Tom Lane、Andrew Dunstan）

  此前这会导致相当无用的结果，例如“Non-recoverable failure in name resolution”。

- 修复向客户端发送组合列值时的内存泄漏（Tom Lane）
- 使 pg_ctl 读取 `postmaster.pid` 文件更健壮（Heikki Linnakangas）

  这修复了竞态条件和可能的文件描述符泄漏。

- 修复当呈现错误编码的数据且 `client_encoding` 设置为仅客户端编码（如 SJIS）时 psql 可能崩溃的问题（Jiang Guiqing）
- 修复 pg_dump 以 `tar` 输出格式生成的 `restore.sql` 脚本中的缺陷（Tom Lane）

  该脚本在表名包含大写字符时会直接失败。此外，让该脚本既能以常规 COPY 模式也能以 `--inserts` 模式恢复数据。

- 修复 pg_restore，使其接受符合 POSIX 的 `tar` 文件（Brian Weaver、Tom Lane）

  pg_dump 的 `tar` 输出模式的原始编码生成的文件并不完全符合 POSIX 标准。这在 9.3 版本中已得到纠正。此补丁更新了以前的分支，使它们同时接受错误格式和纠正后的格式，希望借此避免 9.3 发布时的兼容性问题。

- 修复 pg_resetxlog，使其在给定数据目录相对路径时正确定位 `postmaster.pid`（Tom Lane）

  这一失误可能导致 pg_resetxlog 没有注意到有活动 postmaster 正在使用该数据目录。

- 修复 libpq 的 `lo_import()` 和 `lo_export()` 函数，使其正确报告文件 I/O 错误（Tom Lane）

- 修复 ecpg 对嵌套结构指针变量的处理（Muhammad Usama）

- 让 `contrib/pageinspect` 的 btree 页面检查函数在检查页面时持有缓冲区锁（Tom Lane）

- 修复 pgxs 在 AIX 上构建可加载模块的支持（Tom Lane）

  在原始源码树之外构建模块在 AIX 上不工作。

- 把时区数据文件更新为 tzdata 版本 2012j for DST law changes in Cuba, Israel, Jordan, Libya, Palestine, Western Samoa, and portions of Brazil。

E.75. 版本 8.3.21

发行日期
2012-09-24

本次发布包含对 8.3.20 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

PostgreSQL 社区将于 2013 年 2 月停止对 8.3.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.75.1. 迁移到版本 8.3.21

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.75.2. 变更

- 改进 GiST 索引的页面分裂决策（Alexander Korotkov、Robert Haas、Tom Lane）

  多列 GiST 索引可能因此错误而遭受意外膨胀。

- 修复级联权限撤销，使其在权限仍被持有时停止（Tom Lane）

  如果我们撤销某个角色 X 的授予选项，但 X 仍通过别人的授予持有该选项，我们就不应从 X 曾授予该权限的角色 Y 那里递归撤销相应权限。

- 修复使用 PL/Perl 时对 `SIGFPE` 的处理（Andres Freund）

  Perl 会把进程的 `SIGFPE` 处理程序重置为 `SIG_IGN`，这可能导致之后崩溃。在初始化 PL/Perl 之后恢复正常的 Postgres 信号处理程序。

- 防止递归 PL/Perl 函数在执行期间被重定义时 PL/Perl 崩溃（Tom Lane）
- 规避 PL/Perl 中可能的错误优化（Tom Lane）

  一些 Linux 发行版包含错误版本的 `pthread.h`，导致 PL/Perl 中的编译代码不正确，当一个 PL/Perl 函数调用另一个会抛出错误的函数时会导致崩溃。
- 把时区数据文件更新为 tzdata 版本 2014e，以反映克里米亚、埃及和摩洛哥的夏令时法规变更。

E.76. 版本 8.3.20

发行日期
2012-08-17

本次发布包含对 8.3.19 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

PostgreSQL 社区将于 2013 年 2 月停止对 8.3.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.76.1. 迁移到版本 8.3.20

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.76.2. 变更

- 防止通过 XML 实体引用访问外部文件/URL（Noah Misch、Tom Lane）

  `xml_parse()` 会尝试按需获取外部文件或 URL 来解析 XML 值中的 DTD 和实体引用，从而使无特权数据库用户能够尝试以数据库服务器的权限获取数据。虽然外部数据不会直接返回给用户，但如果数据不能解析为有效 XML，其部分内容可能暴露在错误消息中；而且无论如何，仅仅能够检查文件是否存在对攻击者就可能有用。（CVE-2012-3489）
- 防止通过 `contrib/xml2` 的 `xslt_process()` 访问外部文件/URL（Peter Eisentraut）

  libxslt 提供通过样式表命令读写文件和 URL 的能力，从而使无特权数据库用户能够以数据库服务器的权限读写数据。现通过正确使用 libxslt 的安全选项禁用该能力。（CVE-2012-3488）

  此外，移除了 `xslt_process()` 从外部文件/URL 获取文档和样式表的能力。虽然这是一个有记载的“特性”，但长期以来被认为是个坏主意。CVE-2012-3489 的修复破坏了该能力，与其花费精力修复它，我们干脆将其移除。
- 防止 btree 索引页被过早回收（Noah Misch）

  当我们允许只读事务跳过分配 XID 时，引入了已删除 btree 页在只读事务仍在访问它时就被回收的可能性。这会导致错误的索引搜索结果。由于时序要求，
  这种错误在实际中发生的概率似乎很低，但仍应修复。
- 修复新创建或重置序列的崩溃安全缺陷（Tom Lane）

如果在新创建或重置的序列上执行了 ALTER SEQUENCE，然后恰好对它调用了一次 nextval()，随后服务器崩溃，WAL 重放会把序列恢复到看起来没有执行过 nextval() 的状态，从而使下一个 nextval() 调用再次返回第一个序列值。这种情况尤其可能出现在 serial 列上，因为创建 serial 列的序列包含一个 ALTER SEQUENCE OWNED BY 步骤。

- 确保 pg_start_backup() 之后对 backup_label 文件执行 fsync（Dave Kerr）
- 回移植压缩 fsync 请求队列的 9.1 改进（Robert Haas）

  这改进了检查点期间的性能。9.1 的变更已经过足够的实际测试，看起来可以安全回移植。

- 只允许被直接阻塞的进程自动取消自动清理（Tom Lane）

  原始编码在某些情况下可能允许不一致的行为；特别是自动清理可能在不足 deadlock_timeout 宽限期之后就被取消。

- 改进自动清理取消的日志记录（Robert Haas）
- 修复日志收集器，使 log_truncate_on_rotation 在服务器启动后的第一次日志轮转期间就能工作（Tom Lane）
- 确保对子查询的整行引用不包含任何多余的 GROUP BY 或 ORDER BY 列（Tom Lane）
- 禁止在 CREATE TABLE 期间复制 CHECK 约束和索引定义中的整行引用（Tom Lane）

  这种情况可能出现在带 LIKE 或 INHERITS 的 CREATE TABLE 中。被复制的整行变量被错误地标记为原表而不是新表的行类型。对 LIKE 拒绝这种情况似乎是合理的，因为行类型之后很可能分化。对 INHERITS，理想情况下应该允许它，并隐式强制转换为父表的行类型；但这需要的工作量超出了回移植的安全范围。

- 修复 ARRAY(SELECT ...) 子查询中的内存泄漏（Heikki Linnakangas、Tom Lane）
- 修复从正则表达式中提取公共前缀的问题（Tom Lane）

  代码可能被带量词的带括号子表达式（例如 ^(foo)?bar）混淆。这会导致对此类模式搜索的索引优化不正确。

- 修复 contrib/xml2 的 xslt_process() 中的错误报告（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2012e，以反映摩洛哥和托克劳的夏令时法规变更

E.77. 版本 8.3.19

发行日期
2012-06-04

本次发布包含对 8.3.18 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.77.1. 迁移到版本 8.3.19

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.77.2. 变更

- 修复 `contrib/pgcrypto` 的 DES `crypt()` 函数中不正确的口令变换（Solar Designer）

  如果口令串包含字节值 `0x80`，口令的其余部分会被忽略，使口令比看起来弱得多。修复后，字符串的其余部分会被正确纳入 DES 哈希。因此，受此缺陷影响的任何已存储口令值将不再匹配，所以可能需要更新已存储的值。（CVE-2012-2143）

- 忽略过程语言调用处理程序的 `SECURITY DEFINER` 和 `SET` 属性（Tom Lane）

  把这类属性应用于调用处理程序可能导致服务器崩溃。（CVE-2012-2655）

- 允许 `timestamp` 输入中的数字时区偏移距 UTC 最多 16 小时（Tom Lane）

  一些历史时区的偏移超过此前 15 小时的限制。这可能导致转储的数据值在重载期间被拒绝。

- 修复时间戳转换，
  使其在给定时间恰好是当前时区最后一次夏令时转换时间时能正确处理（Tom Lane）

  这一疏忽已存在很长时间，但此前未被注意到，
  因为大多数使用夏令时的时区被假定有无限的未来夏令时转换序列。

- 修复 `text` 到 `name` 以及 `char` 到 `name` 的转换，使其在多字节编码下正确执行字符串截断（Karl Schnaitter）

- 修复 `to_tsquery()` 中的内存复制缺陷（Heikki Linnakangas）

- 修复 `pg_attribute` 非常大会话启动缓慢的问题（Tom Lane）

  如果 `pg_attribute` 超过 `shared_buffers` 的四分之一，会话启动期间有时需要的缓存重建代码就会触发同步扫描逻辑，使其耗时比正常情况多数倍。如果许多新会话同时启动，
  问题尤为严重。

- 确保顺序扫描以合理的频率检查查询取消（Merlin Moncure）

  遇到许多不含存活元组的连续页面的扫描在此期间不会响应中断。

- 确保 `PGSemaphoreLock()` 的 Windows 实现在返回前清除 `ImmediateInterruptOK`（Tom Lane）

  这一疏忽意味着同一查询稍后收到的查询取消中断可能在不安全的时刻被接受，
  后果不可预测且不好。

- 在打印视图或规则时安全地显示整行变量（Abbas Butt、Tom Lane）

  涉及有歧义名称（即该名称既可能是查询的表名也可能是列名）
  的极端情况此前以有歧义的方式打印，存在视图或规则在转储和重载后被不同解释的风险。
  现通过附加无操作转换来避免歧义情况。

- 确保自动清理工作进程正确执行栈深度检查（Heikki Linnakangas）

  此前，由自动 `ANALYZE` 调用的函数中的无限递归可能导致工作进程崩溃。

- 修复日志收集器，使其在高负载下不丢失日志连贯性（Andrew Dunstan）

此前，如果收集器过于忙碌，可能无法重新组装大消息。

- 修复日志收集器，确保它在收到 SIGHUP 后会重新开始文件轮转（Tom Lane）
- 修复 PL/pgSQL 的 `GET DIAGNOSTICS` 命令在目标是函数第一个变量时的问题（Tom Lane）
- 修复数据库包含大量对象时 pg_dump 的若干性能问题（Jeff Janes、Tom Lane）

  如果数据库包含许多模式、许多对象处于依赖循环中、或有许多被拥有的序列，pg_dump 可能变得非常慢。

- 修复 `contrib/dblink` 的 `dblink_exec()`，使其在出错时不泄漏临时数据库连接（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2012c for DST law changes in Antarctica, Armenia, Chile, Cuba, Falkland Islands, Gaza, Haiti, Hebron, Morocco, Syria, and Tokelau Islands; also historical corrections for Canada。

E.78. 版本 8.3.18

发行日期
2012-02-27

本次发布包含对 8.3.17 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.78.1. 迁移到版本 8.3.18

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，如果你是从earlier than 8.3.17 的版本升级，请参见第 E.79 节。

E.78.2. 变更

- 让 `CREATE TRIGGER` 要求对触发器函数有执行权限（Robert Haas）

  这一缺失的检查可能允许另一用户通过把触发器函数安装到自己拥有的表上，以伪造的输入数据执行该触发器函数。这只对标记为 `SECURITY DEFINER` 的触发器函数有影响，因为其他情况下触发器函数本来就以表所有者身份运行。（CVE-2012-0866）

- 把 pg_dump 注释中写入的名称里的换行符转换为空格（Robert Haas）

  pg_dump 此前对其输出脚本中 SQL 注释内写出的对象名清理不慎。包含换行符的名称至少会使脚本语法错误。恶意构造的对象名可能在脚本重载时构成 SQL 注入风险。（CVE-2012-0868）

- 修复插入与清理并发导致 btree 索引损坏的问题（Tom Lane）

  插入导致的索引页分裂有时会使并发运行的 `VACUUM` 漏删本应删除的索引项。在相应的表行被删除后，悬空的索引项会导致错误（例如 “could not read block N in file ...” ），或者更糟：当不相关的行被重新插入到现已空闲的表位置后，查询会静默地得到错误结果。该缺陷自 8.

2 起就存在，但发生频率极低，直到现在才被诊断出来。如果你有理由怀疑它曾在你的数据库中发生，对受影响的索引重建索引即可修复。

- 允许 `ALTER USER/DATABASE SET` 中某些设置取不存在的值（Heikki Linnakangas）

  允许把 `default_text_search_config`、`default_tablespace` 和 `temp_tablespaces` 设为未知的名称。这是因为设置要在其中使用的另一个数据库中它们可能是已知的；对表空间的情况，则是因为表空间可能尚未创建。此前 `search_path` 已被认识到有同样的问题，这些设置现在的行为与它一致。

- 在 WAL 重放期间正确跟踪 OID 计数器，即使它发生回卷（Tom Lane）

  此前 OID 计数器会一直停留在高值，直到系统退出重放模式。其实际后果通常为零，但在某些场景下，已提升为主库的备库服务器在需要 OID 值时，可能要很长时间才能把 OID 计数器推进到合理的值。

- 修复带 `*` 的正则表达式反向引用（Tom Lane）

  代码实际上并不强制精确字符串匹配，
  而是接受任何满足反向引用符号所引用的模式子表达式的字符串。

  类似问题仍困扰着嵌入在更大带量词表达式中（而非量词直接作用对象）的反向引用。
  这将在未来的 PostgreSQL 版本中解决。

- 修复最近引入的 `inet/cidr` 值处理中的内存泄漏（Heikki Linnakangas）

  PostgreSQL 2011 年 12 月版本中的一个补丁导致这些操作中的内存泄漏，在此类列上构建 btree 索引等场景中可能相当显著。

- 避免 Windows 上 syslogger 中文件句柄的双重关闭（MauMau）

  通常这个错误不可见，但在调试版 Windows 上运行时会导致异常。

- 修复 plpgsql 中与 I/O 转换相关的内存泄漏（Andres Freund、Jan Urbanski、Tom Lane）

  某些操作会一直泄漏内存直到当前函数结束。

- 改进 pg_dump 对继承表列的处理（Tom Lane）

  pg_dump 错误处理了子列与父列默认表达式不同的情况。如果默认值在文本上与父列相同但实际不同（例如因为模式搜索路径差异），它不会被识别为不同，
  因此转储和恢复后子列会被允许继承父列的默认值。子列为 `NOT NULL` 而父列不是的情况也可能被微妙地错误恢复。

- 修复 pg_restore 直接写数据库模式对 INSERT 式表数据的处理（Tom Lane）

  当使用 2011 年 9 月或 12 月版本的 pg_restore 时，对以 `--inserts` 或 `--column-inserts` 选项制作的归档文件直接恢复到数据库会失败，这是对另一个问题的修复中的一个疏忽所致。归档文件本身没有问题，文本模式输出也正常。

- 修复 `contrib/intarray` 的 `int[] & int[]` 操作符中的错误（Guillaume Lelarge）

  如果两个输入数组共有的最小整数是 1，且任一数组中有更小的值，那么 1 会被错误地从结果中遗漏。

- 修复 `contrib/pgcrypto` 的 `encrypt_iv()` 和 `decrypt_iv()` 中的错误检测（Marko Kreen）

这些函数未能报告某些类型的无效输入错误，而是对错误输入返回随机垃圾值。

- 修复 `contrib/test_parser` 中的一字节缓冲区越界（Paul Guyot）

  代码会尝试多读一个字节，这在极端情况下会崩溃。由于 `contrib/test_parser` 只是示例代码，这本身不是安全问题，但坏的示例代码仍然是坏的。

- 在 ARM 上如果可用则使用 `__sync_lock_test_and_set()` 实现自旋锁（Martin Pitt）

  该函数取代了我们此前对 `SWPB` 指令的使用，该指令已被弃用且在 ARMv6 及更高版本上不可用。报告表明旧代码在较新的 ARM 板上不会以明显方式失败，而只是不对并发访问互锁，导致多进程操作中的离奇失败。

- 在使用接受该选项的 gcc 版本构建时使用 `-fexcess-precision=standard` 选项（Andrew Dunstan）

  这防止了较新版本 gcc 产生"创造性"结果的各种场景。

- 允许在 FreeBSD 上使用线程化 Python（Chris Rees）

  我们的 configure 脚本此前认为这种组合无法工作；但 FreeBSD 修复了问题，因此移除该错误检查。

E.79. 版本 8.3.17

发行日期
2011-12-05

本次发布包含对 8.3.16 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.79.1. 迁移到版本 8.3.17

对于运行 8.3.X 的用户，不需要进行转储/恢复。

不过，`information_schema.referential_constraints` 视图的定义中被发现存在一个长期错误。如果你依赖该视图的正确结果，应按下面第一条变更日志条目的说明替换其定义。

不过，如果你是从earlier than 8.3.8 的版本升级，请参见第 E.88 节。

E.79.2. 变更

- 修复 `information_schema.referential_constraints` 视图中的缺陷（Tom Lane）

  该视图在把外键约束与它依赖的主键或唯一键约束匹配时不够谨慎。
  这可能导致某个外键约束完全不被显示、被显示多次，
  或者声称它依赖于一个并非其实际依赖的约束。

  由于该视图定义是由 initdb 安装的，仅靠升级不会修复问题。如果需要修复现有安装，可以（以超级用户身份）删除 `information_schema` 模式，然后通过载入 *SHAREDIR/*

information_schema.sql 重新创建它。（如果不确定 *SHAREDIR* 在哪里，运行 pg_config --sharedir。）这必须在每个要修复的数据库中重复执行。

- 修复 CREATE TABLE dest AS SELECT * FROM src 或 INSERT INTO dest SELECT * FROM src 期间与 TOAST 相关的数据损坏（Tom Lane）

  如果表曾被 ALTER TABLE ADD COLUMN 修改过，在某些极端情况下，尝试把它的数据原样复制到另一个表可能产生损坏的结果。该问题只会在 8.4 及更高版本中以这种确切形式出现，但我们也给更早的版本打了补丁，以防有其他代码路径可能触发同一缺陷。

- 修复从过期 syscache 条目访问 toast 表时的竞态条件（Tom Lane）

  典型症状是诸如 “missing chunk number 0 for toast value NNNNN in pg_toast_2619” 的瞬时错误，其中引用的 toast 表总是属于某个系统目录。

- 让 DatumGetInetP() 解包带 1 字节头的 inet datum，并新增一个不做此事的宏 DatumGetInetPP()（Heikki Linnakangas）

  此变更不影响任何核心代码，但可能避免期望 DatumGetInetP() 按常规约定产生已解包 datum 的附加代码中的崩溃。

- 改进 money 类型输入和输出的区域支持（Tom Lane）

  除了不支持所有标准 lc_monetary 格式化选项外，输入和输出函数还不一致，这意味着在某些区域设置下转储的 money 值无法重新读取。

- 不让 transform_null_equals 影响 CASE foo WHEN NULL ... 构造（Heikki Linnakangas）

  transform_null_equals 只应影响用户直接写的 foo = NULL 表达式，而不是这种形式的 CASE 在内部生成的相等性检查。

- 更改外键触发器的创建顺序，以更好地支持自引用外键（Tom Lane）

  对于引用自身表的级联外键，行更新会作为一个事件同时触发 ON UPDATE 触发器和 CHECK 触发器。ON UPDATE 触发器必须先执行，否则 CHECK 会检查行的非最终状态，并可能抛出不恰当的错误。然而这些触发器的触发顺序由其名称决定，而由于触发器使用遵循 “RI_ConstraintTrigger_NNNN” 约定的自动生成名称，它们通常按创建顺序排序。彻底的修复需要修改该约定（我们将在 9.2 中做），但在现有版本中更改它似乎有风险。因此此补丁只更改触发器的创建顺序。
  遇到这类错误的用户应删除并重新创建外键约束，使其触发器处于正确顺序。

- 避免在跟踪缓冲区分配率时发生浮点下溢（Greg Matthews）

  虽然这本身无害，但在某些平台上会导致烦人的内核日志消息。

- 在 psql 的命令历史中保留命令内的空行（Robert Haas）

  例如，如果从字符串字面量中删除了空行，先前的行为可能导致问题。

- 修复 pg_dump，使其转储自动生成类型（如表行类型）之间的用户定义转换（Tom Lane）

- 使用首选版本的 xsubpp 构建 PL/Perl，而不一定是操作系统的主副本（David Wheeler、Alex Hunsaker）

- 修复 contrib/dict_int 和 contrib/dict_xsyn 中的错误编码（Tom Lane）

一些函数错误地假定 `palloc()` 返回的内存保证被清零。

- 让 `pgstatindex()` 及时响应查询取消中断（Robert Haas）
- 确保 VPATH 构建正确安装所有服务器头文件（Peter Eisentraut）
- 缩短详细错误消息中报告的文件名（Peter Eisentraut）

  常规构建一直只报告包含错误消息调用的 C 文件的名称，但 VPATH 构建此前报告绝对路径名。

- 修复中美洲 Windows 时区名的解释（Tom Lane）

  把“Central America Standard Time”映射为 `CST6` 而不是 `CST6CDT`，因为中美洲各地通常不实行夏令时。

- 把时区数据文件更新为 tzdata 版本 2011n for DST law changes in Brazil, Cuba, Fiji, Palestine, Russia, and Samoa; also historical corrections for Alaska and British East Africa。

E.80. 版本 8.3.16

发行日期
2011-09-26

本次发布包含对 8.3.15 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.80.1. 迁移到版本 8.3.16

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.80.2. 变更

- 修复对不确定的 HOT 更新元组建立索引的缺陷（Tom Lane）

  这些缺陷可能导致重建系统目录索引之后索引损坏。据信它们不影响用户索引。

- 修复 GiST 索引页面分裂处理中的多个缺陷（Heikki Linnakangas）

  发生概率很低，但它们可能导致索引损坏。

- 修复 `tsvector_concat()` 中可能的缓冲区越界（Tom Lane）

  该函数可能低估其结果所需的内存量，导致服务器崩溃。

- 修复 `xml_recv` 处理“standalone”参数时的崩溃（Tom Lane）
- 避免在 `ANALYZE` 和 SJIS-2004 编码转换中可能越过内存末尾的访问（Noah Misch）

  这修复了一些概率极低的服务器崩溃场景。

- 修复 relcache init 文件失效中的竞态条件（Tom Lane）

  存在一个时间窗口，新的后端进程可能读到过期的 init 文件，却错过本应告知它数据已过期的失效消息。结果是目录访问中的离奇失败，通常在启动稍后阶段报 “could not read block 0 in file ...” 。

- 修复 GiST 索引扫描结束时的内存泄漏（Tom Lane）

  执行许多独立 GiST 索引扫描的命令（例如对已包含许多行的表验证新的基于 GiST 的排他约束）会因此泄漏瞬时需要大量内存。

- 修复构造大型有损位图时的性能问题（Tom Lane）
- 修复数组和路径创建函数，确保填充字节为零（Tom Lane）

  这避免了一些规划器认为语义相等的常量不相等的情况，这种情况会导致糟糕的优化。

- 规避破坏 WAL 重放的 gcc 4.6.0 缺陷（Tom Lane）

  这可能导致服务器崩溃后已提交事务的丢失。

- 修复视图中 `VALUES` 的转储缺陷（Tom Lane）
- 禁止对序列执行 `SELECT FOR UPDATE/SHARE`（Tom Lane）

  这种操作不像预期那样工作，可能导致失败。

- 防御计算哈希表大小时的整数溢出（Tom Lane）
- 修复 `CLUSTER` 可能尝试访问已移除的 TOAST 数据的情况（Tom Lane）
- 修复 “peer” 认证使用凭据控制消息时的可移植性缺陷（Tom Lane）
- 修复需要多次往返时的 SSPI 登录（Ahmed Shinwari、Magnus Hagander）

  该问题的典型症状是 SSPI 登录期间的 “The function requested is not supported” 错误。

- 修复 `pg_srand48` 种子初始化中的笔误（Andres Freund）

  这导致未能使用所提供种子的所有位。该函数在大多数平台上不使用（只在没有 `srandom` 的平台上使用），而且无论如何看待，随机性低于预期的种子带来的潜在安全暴露似乎都很小。

- 避免 `LIMIT` 与 `OFFSET` 值之和超过 2^63 时的整数溢出（Heikki Linnakangas）
- 为 `generate_series()` 的 `int4` 和 `int8` 版本添加溢出检查（Robert Haas）
- 修复 `to_char()` 中移除尾随零的问题（Marti Raudsepp）

  在使用 `FM` 且小数点后没有数字位的格式中，小数点左侧的零可能被错误移除。

- 修复 `pg_size_pretty()`，避免接近 2^63 的输入溢出（Tom Lane）
- 在 pg_ctl 中，支持 Windows 上服务注册的静默模式（MauMau）
- 修复 psql 在从其他文件执行 `COPY` 期间脚本文件行号计数的问题（Tom Lane）
- 修复 pg_restore 直接写数据库模式对 `standard_conforming_strings` 的处理（Tom Lane）

当从以 `standard_conforming_strings` 设为 `on` 制作的归档文件直接恢复到数据库服务器时，pg_restore 可能发出不正确的命令。

- 修复 libpq 的 LDAP 服务查找代码中的越界写入和内存泄漏（Albe Laurenz）
- 在 libpq 中，避免使用非阻塞 I/O 和 SSL 连接时的失败（Martin Pihlak、Tom Lane）
- 改进 libpq 对连接启动期间失败的处理（Tom Lane）

  特别是，对 SSL 连接启动期间服务器报告 `fork()` 失败的响应现在更合理。
- 改进 libpq 对 SSL 失败的错误报告（Tom Lane）
- 让 ecpglib 以 15 位精度写 `double` 值（Akira Kurosawa）
- 在 ecpglib 中，确保出错后恢复 `LC_NUMERIC` 设置（Michael Meskes）
- 应用上游对 blowfish 有符号字符缺陷的修复（CVE-2011-2483）（Tom Lane）

  `contrib/pg_crypto` 的 blowfish 加密代码在 char 为有符号的平台（大多数平台如此）上可能给出错误结果，导致加密口令比应有的强度更弱。
- 修复 `contrib/seg` 中的内存泄漏（Heikki Linnakangas）
- 修复 `pgstatindex()`，使其对空索引给出一致结果（Tom Lane）
- 允许使用 perl 5.14 构建（Alex Hunsaker）
- 更新 configure 脚本探测系统函数是否存在的方法（Tom Lane）

  我们在 8.3 和 8.2 中使用的 autoconf 版本可能被执行链接时优化的编译器欺骗。
- 修复构建和安装文件路径包含空格的各种问题（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2011i for DST law changes in Canada, Egypt, Russia, Samoa, and South Sudan。

E.81. 版本 8.3.15

发行日期
2011-04-18

本次发布包含对 8.3.14 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.81.1. 迁移到版本 8.3.15

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.81.2. 变更

- 禁止在组合类型中包含其自身（Tom Lane）

这防止了服务器在处理该组合类型时无限递归的场景。虽然这种结构有一些可能的用途，但它们似乎不足以抵消确保其始终安全工作所需的努力。

- 避免目录缓存初始化期间的潜在死锁（Nikhil Sontakke）

  某些情况下，缓存加载代码会在锁定索引的目录之前获取系统索引上的共享锁。这可能与尝试以另一种更标准的顺序获取排他锁的进程死锁。

- 修复当目标元组存在并发更新时 `BEFORE ROW UPDATE` 触发器处理中的悬空指针问题（Tom Lane）

  已观察到该缺陷在尝试执行 `UPDATE RETURNING ctid` 时导致间歇性的 “cannot extract system attribute from virtual tuple” 失败。发生更严重错误（例如为更新后的元组生成错误索引项）的概率非常小。

- 当表存在未决的延迟触发器事件时禁止 `DROP TABLE`（Tom Lane）

  此前 `DROP` 会照常执行，导致触发器最终触发时报 “could not open relation with OID nnn” 错误。

- 修复涉及数组切片的 PL/Python 内存泄漏（Daniel Popowich）
- 修复 pg_restore，使其能处理 TOC 文件中的长行（超过 1KB）（Tom Lane）
- 加入更多防护，防止因过于激进的编译器优化导致的除零崩溃（Aurelien Jarno）
- 支持在 FreeBSD 和 OpenBSD 的 MIPS 上使用 dlopen()（Tom Lane）

  此前有一个硬编码的假设：该系统函数在这些系统的 MIPS 硬件上不可用。现在改用编译期测试，因为较新的版本已经有了它。

- 修复 HP-UX 上的编译失败（Heikki Linnakangas）
- 修复 Windows 上 libintl 的版本不兼容问题（Hiroshi Inoue）
- 修复 Windows 构建脚本中 xcopy 的用法，使其在 Windows 7 下正确工作（Andrew Dunstan）

  这只影响构建脚本，不影响安装或使用。

- 修复 pg_regress 在 Cygwin 上使用的路径分隔符（Andrew Dunstan）
- 把时区数据文件更新为 tzdata 版本 2011f for DST law changes in Chile, Cuba, Falkland Islands, Morocco, Samoa, and Turkey; also historical corrections for South Australia, Alaska, and Hawaii。

E.82. 版本 8.3.14

发行日期
2011-01-31

本次发布包含对 8.3.13 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.82.1. 迁移到版本 8.3.14

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.82.2. 变更

- 避免 `EXPLAIN` 尝试显示简单形式 `CASE` 表达式时的失败（Tom Lane）

  如果 `CASE` 的测试表达式是常量，规划器可能把 `CASE` 简化为令表达式显示代码困惑的形式，导致 “unexpected CASE WHEN clause” 错误。

- 修复对现有下标范围之前的数组切片赋值的问题（Tom Lane）

  如果新增下标与第一个既有下标之间存在间隙，
  代码会错误计算需要从旧数组的空值位图复制多少条目，可能导致数据损坏或崩溃。

- 避免规划器对非常久远的日期值出现意外的转换溢出（Tom Lane）

  `date` 类型支持的日期范围比 `timestamp` 类型能表示的更宽，但规划器假定它总是可以不受惩罚地把 date 转换为 timestamp。

- 修复启用 `standard_conforming_strings` 时 pg_restore 对大对象（BLOB）的文本输出（Tom Lane）

  虽然直接恢复到数据库时工作正确，但如果要求 pg_restore 生成 SQL 文本输出且源数据库中已启用 `standard_conforming_strings`，字符串转义是不正确的。

- 修复对包含 `... & !(subexpression) | ...` 的 `tsquery` 值的错误解析（Tom Lane）

  包含这种操作符组合的查询没有被正确执行。同样的错误也存在于 `contrib/intarray` 的 `query_int` 类型和 `contrib/ltree` 的 `ltxtquery` 类型中。

- 修复 `contrib/intarray` 的 `query_int` 类型输入函数中的缓冲区越界（Apple）

  该缺陷构成安全风险，因为函数的返回地址可能被覆写。感谢 Apple Inc 的安全团队报告此问题并提供修复。（CVE-2010-4015）

- 修复 `contrib/seg` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `seg` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。（这与上一次更新中在 `contrib/cube` 中修复的缺陷相同。）

E.83. 版本 8.3.13

发行日期
2010-12-16

本次发布包含对 8.3.12 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.83.1. 迁移到版本 8.3.13

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.83.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复 GIN 索引 WAL 重放逻辑中的若干缺陷（Tom Lane）

  这可能导致复制期间的“bad buffer id: 0”失败或索引内容损坏。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 修复多个工作进程长期保持活动时自动清理工作进程的持续变慢（Tom Lane）

  如果自动清理工作进程处理了足够多的表，其有效 `vacuum_cost_limit` 可能降到接近零，导致其运行极慢。

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

  给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

  在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）

- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复当 `ORDER BY`、`LIMIT`、`FOR UPDATE` 或 `WITH` 附加到 `INSERT ... VALUES` 的 `VALUES` 部分时的正确行为（Tom Lane）

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

  规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 修复当连接接受（`accept()` 或紧随其后的调用之一）失败且 postmaster 编译时带 GSSAPI 支持时的 postmaster 崩溃（Alexander Chernikov）

- 修复 `log_temp_files` 处于活动状态时临时文件漏删的问题（Tom Lane）

  如果尝试发出日志消息时出错，就不会执行删除，导致临时文件累积。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

  该缺陷影响多个不同的几何距离测量操作符。

- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）

- 修复 PL/Python 对集合返回函数的处理（Jan Urbanski）

  在生成集合结果的迭代器内尝试调用 SPI 函数会失败。

- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `cube` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。

- `contrib/dblink` 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）

- 修复 `contrib/pgcrypto` 在缺少公钥时可能的核心转储（Marti Raudsepp）

- 修复 `contrib/xml2` 的 XPath 查询函数中的内存泄漏（Tom Lane）

- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.84. 版本 8.3.12

发行日期
2010-10-04

本次发布包含对 8.3.11 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.84.1. 迁移到版本 8.3.12

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.84.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一

点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）

- 在 Windows 上把退出码 128（`ERROR_WAIT_NO_CHILDREN`）视为非致命（Magnus Hagander）

  高负载下，Windows 进程有时会在启动时以此错误码失败。此前 postmaster 把这视为 panic 情况并重启整个数据库，但这似乎反应过度。

- 修复 Append 索引扫描中对非严格 OR 连接子句的错误使用（Tom Lane）

  这是 8.4 中一个修复的向后移植，该修复此前在 8.3 分支中被遗漏。这纠正了 8.3.8 中引入的一个错误，当内部关系是继承树或 `UNION ALL` 子查询时，该错误可能导致外连接的结果不正确。

- 修复 `UNION ALL` 成员关系可能被重复扫描的问题（Tom Lane）

- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。

- 修复未将缓存计划标记为瞬态的失败（Tom Lane）

  如果在某个被引用表正在执行 `CREATE INDEX CONCURRENTLY` 期间准备了计划，应当在索引可供使用后重新规划。这此前没有可靠地发生。

- 在偶有报告的 btree 失败情形中把 PANIC 降为 ERROR，并在产生的错误消息中提供更多细节（Tom Lane）

  这应能改进系统对损坏索引的健壮性。

- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）

- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）

- 修复对按引用传递的函数结果计算哈希时可能的失败（Tao Ma、Tom Lane）

- 改进归并连接对连接列中 NULL 的处理（Tom Lane）

  如果排序顺序使 NULL 排在高位，归并连接现在可以在到达第一个 NULL 时完全停止。

- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 避免 walwriter 进程持有打开的旧 WAL 段（Magnus Hagander、Heikki Linnakangas）

  以前的编码会阻止删除不再需要的段。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）

- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）

- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）

- 修复 `REASSIGN OWNED` 以处理操作符类和操作符族（Asko Tiidumaa）

- 修复比较两个空 `tsquery` 值时可能的核心转储（Tom Lane）

- 修复 `LIKE` 对包含 `%` 后跟 _ 的模式的处理（Tom Lane）

  我们此前修复过它，但仍有一些被错误处理的情形。

- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）

- 让 psql 在自动提交关闭模式下把 `DISCARD ALL` 识别为不应包裹在事务块中的命令（Itagaki Takahiro）

- 修复 ecpg 以正确处理来自 `RETURNING` 子句的数据（Michael Meskes）

- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）

- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）

- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）

- 为 `contrib/hstore` 新增 `hstore(text, text)` 函数（Robert Haas）

  此函数是现已废弃的 `=>` 操作符的推荐替代。向后移植它使得面向未来的代码可以用于较旧的服务器版本。注意该补丁只有在 `contrib/hstore` 于特定数据库中安装或重新安装之后才生效。用户也可以改为手工执行 `CREATE FUNCTION` 命令。

- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）

- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

- 让 Windows 的“N. Central Asia Standard Time”时区映射到 Asia/Novosibirsk 而不是 Asia/Almaty（Magnus Hagander）

  Microsoft 在 KB976098 的时区更新中改变了此区域的 DST 行为。Asia/Novosibirsk 更匹配其新行为。

E.85. 版本 8.3.11

发行日期
2010-05-17

本次发布包含对 8.3.10 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.85.1. 迁移到版本 8.3.11

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.85.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 修复在重建 relcache 条目期间收到缓存重置消息时可能的崩溃（Heikki）

  此错误是在 8.3.10 中修复相关失败时引入的。

- 在为函数运行语言验证器时应用逐函数 GUC 设置（Itagaki Takahiro）

  这避免了函数代码在没有该设置时无效所导致的失败；例如 `search_path` 不正确时 SQL 函数可能无法解析。

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 确保归档进程尽快响应 `archive_command` 的变更（Tom）

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）
- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）
- 防止 psql 在展开引用自身的变量时无限递归（Tom）
- 修复 psql 的 `\copy`，不在 `\copy (select ...)` 内的点周围添加空格（Tom）

  在数字字面量的小数点周围添加空格会导致语法错误。
- 修复使用 `contrib/intarray` 操作符的不可满足查询产生不必要的 “GIN 索引不支持整索引扫描” 错误（Tom）
- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）
- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。
- 避免 Windows 上 syslogger 进程可能的崩溃（Heikki）
- 更健壮地处理 Windows 注册表中不完整的时区信息（Magnus）
- 更新已知 Windows 时区名称集合（Magnus）
- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

  另外，把 `PKST`（巴基斯坦夏令时）加入默认时区缩写集合。

E.86. 版本 8.3.10

发行日期
2010-03-15

本次发布包含对 8.3.9 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.86.1. 迁移到版本 8.3.10

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.86.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。
- 修复后端启动期间可能的死锁（Tom）

- 修复因未干净地处理 relcache 重载期间的错误而可能的崩溃（Tom）
- 修复因使用指向缓存计划的悬空指针而可能的崩溃（Tatsuo）
- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）
- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）
- 修复 GIST 索引页分裂的恢复结束清理期间发出的不正确 WAL 数据（Yoichi Hirai）

  如果在完成未完成的 GIST 插入之后、在恢复结束清理期间崩溃，这会导致索引损坏，或者更可能在 WAL 重放期间出错。

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）
- 修复若干病态缓慢的正则表达式匹配情形（Tom）
- 修复 `xml` 处理中因草率内存管理导致的若干崩溃（Tom）

  这是最先应用于 8.4 的变更的向后移植。已知 8.3 代码有缺陷，但新代码差异足够大，在获得一些实际测试之前不想向后移植它。

- 修复试图更新组合类型数组列中某元素的字段时的缺陷（Tom）
- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）
- 修复更多临时文件泄漏的情形（Heikki）

  这纠正了上一个小版本中引入的问题。失败的一种情形是返回集合的 plpgsql 函数在另一个函数的异常处理器中被调用。

- 改进布尔变量情形的约束排除处理，特别是使排除带有“bool_column = false”约束的分区成为可能（Tom）
- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 修复 `SSL_read` 或 `SSL_write` 失败而未设置 `errno` 时可能的无限循环（Tom）

  据报道这在某些 Windows 版本的 openssl 上可能发生。

- 不允许在本地连接上使用 GSSAPI 认证，因为它需要主机名才能正确工作（Magnus）
- 让 ecpg 在连接消失时报告正确的 SQLSTATE（Michael）

- 修复 psql 的 `numericlocale` 选项，使其不在 latex 和 troff 输出格式中格式化不应格式化的字符串（Heikki）
- 当同时指定 `ON_ERROR_STOP` 和 `--single-transaction` 且隐含的 `COMMIT` 期间发生错误时，让 psql 返回正确的退出状态（3）（Bruce）
- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）
- 修复从 PL/PerlU 调用 PL/Perl 函数或相反时可能的失败（Tim Bunce）
- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）
- 允许 `contrib/ltree` 操作中使用零维数组（Tom）

  这种情形此前被作为错误拒绝，但把它与零元素数组同样处理更方便。特别是，这避免了当 `ltree` 操作应用于 `ARRAY(SELECT ...)` 的结果而子查询不返回行时不必要的失败。
- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）
- 让 `contrib/xml2` 在 Windows 上的构建更健壮（Andrew）
- 修复 Windows 信号处理中的竞态条件（Radu Ilie）

  此缺陷的一个已知症状是 `pg_listener` 中的行可能在重负载下被丢弃。
- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.87. 版本 8.3.9

发行日期
2009-12-14

本次发布包含对 8.3.8 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.87.1. 迁移到版本 8.3.9

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.8 的版本升级，请参阅第 E.88 节。

E.87.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。

- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。

- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 避免在空同义词词典上崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。

- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。

- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 确保预备事务所持有的共享元组级锁不被忽略（Heikki）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复轮转到新 CSV 日志文件时 syslogger 进程的内存泄漏（Tom）
- 修复 Windows 权限降级逻辑（Jesse Morris）

  这修复了数据库在 Windows 上启动失败的一些情形，这些情形常伴随误导性错误消息，例如“找不到匹配的 postgres 可执行文件”。

- 修复 GiST 索引页分裂的不正确逻辑：当分裂依赖于索引的非首列时（Paul Ramsey）
- 如果回收或删除旧 WAL 文件在检查点结束时失败，不要报错退出（Heikki）

  更好的做法是把该问题视为非致命并允许检查点完成。未来的检查点会重试删除。
  这类问题在正常运行中不应出现，但已知可能由设计不当的 Windows 反病毒和备份软件引起。

- 确保 WAL 文件不会在 Windows 上被重复归档（Heikki）

  这是其他进程干扰删除不再需要的文件时可能出现的另一种症状。

- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 提高 GSSAPI 和 SSPI 认证方法的最大认证令牌（Kerberos 票据）大小（Ian Turner）

  虽然旧的 2000 字节限制对 Unix Kerberos 实现绰绰有余，Windows 域控制器签发的票据可能大得多。

- 重新启用对序列的访问统计信息收集（Akira Kurosawa）

  这以前可用，但在 8.3 中损坏了。

- 修复 `CREATE OR REPLACE FUNCTION` 期间所有权依赖的处理（Tom）
- 修复对 `WHERE` *x*=*x* 条件的不正确处理（Tom）

某些情况下它们可能被当作冗余而忽略，但它们并非冗余——它们等价于 *x* `IS NOT NULL`。

- 让文本搜索解析器接受 XML 属性中的下划线（Peter）
- 修复 `xml` 二进制输入的编码处理（Heikki）

  如果 XML 头未指定编码，现在默认采用 UTF-8；以前的处理不一致。
- 修复从 `plperl` 调用 `plperlu` 或相反时的缺陷（Tom）

  内层函数的错误退出可能因未能为外层函数重新选择正确的 Perl 解释器而导致崩溃。
- 修复重定义 PL/Perl 函数时的会话期内存泄漏（Tom）
- 确保 Perl 数组在由返回集合的 PL/Perl 函数返回时被正确转换为 PostgreSQL 数组（Andrew Dunstan、Abhijit Menon-Sen）

  对非返回集合的函数这早已正确工作。
- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 在 `contrib/pg_standby` 中，禁用在 Windows 上用信号触发故障转移（Fujii Masao）

  这从未有过任何用处，因为 Windows 没有 Unix 式信号，但最近的变更使它实际上崩溃。
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。
- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 更新时区缩写文件以符合当前现实（Joachim Wieland）

  这包括把 `IDT` 和 `SGT` 加入默认时区缩写集合。
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.88. 版本 8.3.8

发行日期
2009-09-09

本次发布包含对 8.3.7 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.88.1. 迁移到版本 8.3.8

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.3.8 后对它们执行 `REINDEX`。另外，如果你是从早于 8.3.5 的版本升级，请参阅第 E.91 节。

E.88.2. 变更

- 修复 Windows 共享内存分配代码（Tsutomu Yamada、Magnus）

  此缺陷导致常被报告的“无法重新附加到共享内存”错误消息。

- 在 `pg_start_backup()` 期间强制 WAL 段切换（Heikki）

  这避免了可能使基础备份不可用的极端情形。

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 把对已装载的可装载模块的 `LOAD` 变为无操作（Tom）

  此前 `LOAD` 会试图卸载并重新装载模块，但这不安全也不太有用。

- 不允许在 LDAP 认证期间使用空口令（Magnus）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复与从 Sort 或 Materialize 计划节点的输出获取整行值相关的缺陷（Tom）

- 防止 `synchronize_seqscans` 改变可滚动游标和 `WITH HOLD` 游标的结果（Tom）

- 回退规划器中当 AND 或 OR 列表中的子句超过 100 个时禁用部分索引和约束排除优化的变更（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'`/`'HH12'` 下当作大写序数后缀（Heikki）

  它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 修复模式包含 `%_` 时 `LIKE` 的行为（Tom）

- 正确地对像 `00:12:57.9999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 XML 操作中的内存泄漏（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 确保“快速关闭”请求会强制终止打开的会话，即使“智能关闭”已在进行中（Fujii Masao）
- 避免当输入值（几乎）按排序顺序时批量插入 GIN 索引的性能退化（Tom）
- 在 PL/pgSQL 的某些上下文中正确强制执行 NOT NULL 域约束（Tom）
- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）
- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）
- 改进 pg_dump 在有许多大对象时的效率（Tamas Vincze）
- 用 `SIGUSR1` 而不是 `SIGQUIT` 作为 pg_standby 的故障转移信号（Heikki）
- 让 pg_standby 的 `maxretries` 选项按文档描述的那样工作（Fujii Masao）
- 让 `contrib/hstore` 在键或值太长而无法放入其数据结构时抛出错误，而不是静默截断（Andrew Gierth）
- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）
- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）
- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）
- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.89. 版本 8.3.7

发行日期
2009-03-16

本次发布包含对 8.3.6 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.89.1. 迁移到版本 8.3.7

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.5 的版本升级，请参阅第 E.91 节。

E.89.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的任何错误都可能在试图报告该错误时导致无限递归。因此，
  解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `xpath()`，使其仅在必要时修改路径表达式，并在必要时做更合理的尝试（Andrew）

  SQL 标准暗示 `xpath` 应能处理文档片段形式的数据，但 libxml 不支持这一点，而且按照 XPath 标准这是否合理也确实不清楚。`xpath` 试图通过同时修改数据和路径表达式来绕过这一不匹配，但该修改有缺陷，可能导致有效搜索失败。现在 `xpath` 检查数据是否确实是格式良好的文档，如果是，则不加更改地把数据和路径表达式交给 libxml。否则，使用一种较不容易失败的另一种修改方法。

  > **注意**
  >
  > 新的修改方法仍不能 100% 令人满意，而且似乎不可能有真正的解决方案。因此应把这个补丁视为权宜之计，以避免不必要地破坏现有应用程序。PostgreSQL 8.4 很可能干脆拒绝在非格式良好文档的数据上使用 `xpath`。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）
- 修复 C 区域设置与多字节编码一起使用时文本搜索可能失败的问题（Teodor）

  在 `wchar_t` 比 `int` 窄的平台（特别是 Windows）上可能崩溃。

- 修复文本搜索解析器处理包含多个 `@` 字符的类电子邮件字符串时的极端低效（Heikki）
- 修复较大子查询的输出列表中带子 `SELECT` 的规划器问题（Tom）

  此缺陷的已知症状是“定位分组列失败”错误，且与所涉及的数据类型有关；
  但也可能存在其他问题。

- 修复带隐式强制转换的 `CASE WHEN` 的反编译（Tom）

  在试图检查或转储视图时，这一错误在启用断言的构建中可能导致断言失败，
  在其他情况下可能导致“意外的 CASE WHEN 子句”错误消息。

- 修复 TOAST 表行类型所有者可能的错误分配（Tom）

  如果 `CLUSTER` 或重写形式的 `ALTER TABLE` 由表所有者之外的人执行，表的 TOAST 表的 `pg_type` 条目最终会被标记为那个人所拥有。这不会立即引起问题，因为普通数据库操作不检查 TOAST 行类型的权限。但是，如果后来试图删除发出该命令的角色（在 8.1 或 8.2 中），或者在那之后 pg_dump 发出“数据类型的所有者似乎无效”警告（在 8.3 中），则可能导致意外失败。

- 更改 `UNLISTEN`，如果当前会话从未执行过任何 `LISTEN` 命令则快速退出（Tom）

  大多数时候这不是特别有用的优化，但由于 `DISCARD ALL` 会调用 `UNLISTEN`，以前的编码对大量使用 `DISCARD ALL` 的应用程序造成了显著的性能问题。

- 修复 PL/pgSQL：不再把字符串中任何位置的 `INSERT` 之后的 `INTO` 当作 INTO-变量子句，而不仅是开头；特别是不要对 `CREATE RULE` 内的 `INSERT INTO` 失败（Tom）
- 在块退出时完全清理 PL/pgSQL 错误状态变量（Ashesh Vashi 和 Dave Page）

  这对 PL/pgSQL 本身不是问题，但这一疏漏可能导致 PL/pgSQL 调试器在检查函数状态时崩溃。

- 在 Windows 上重试失败的 `CallNamedPipe()` 调用（Steve Marshall、Magnus）

  该函数有时似乎会瞬时失败；我们此前把任何失败都当作硬错误，这可能扰乱 `LISTEN/NOTIFY` 以及其他操作。

- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.90. 版本 8.3.6

发行日期
2009-02-02

本次发布包含对 8.3.5 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.90.1. 迁移到版本 8.3.6

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.5 的版本升级，请参阅第 E.91 节。

E.90.2. 变更

- 让 `DISCARD ALL` 除了释放已有的一切外还释放咨询锁（Tom）

  这被认为是最合适的行为。但这可能影响现有应用程序。

- 修复整索引 GiST 扫描使其正确工作（Teodor）

  如果表在 GiST 索引上聚簇，此错误可能导致行丢失。

- 修复 `xmlconcat(NULL)` 的崩溃（Peter）

- 修复 `ispell` 词典在高位字符被用作标志时可能的崩溃（Teodor）

  已知一部广泛可用的挪威语词典会这样做，其他词典可能存在同样的情况。

- 修复 pg_dump 对组合类型输出的错误排序（Tom）

  最可能的问题是用户定义的操作符类在需要它们的索引或视图之后才被转储。

- 改进 `headline()` 函数中对 URL 的处理（Teodor）

- 改进 `headline()` 函数对超长标题的处理（Teodor）

- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）

- 修复 PL/pgSQL 中执行的语句被重写为另一种语句时可能的断言失败，例如 `INSERT` 被重写为 `UPDATE`（Heikki）

- 确保数据类型输入函数有可用快照（Tom）

这主要影响声明了涉及用户定义稳定或不可变函数的 `CHECK` 约束的域。如果没有设置快照，这类函数通常会失败。

- 让使用 SPI 的函数更安全地用于数据类型 I/O 中；特别是用于域检查约束（Tom）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复有时使 `ALTER TABLE ENABLE/DISABLE RULE` 无法被活动会话识别的问题（Tom）
- 修复使 `UPDATE RETURNING tableoid` 返回零而不是正确 OID 的问题（Tom）
- 允许声明为接受 `ANYARRAY` 的函数用于该类型的 `pg_statistic` 列（Tom）

  这以前可用，但在 8.3 中被无意破坏。
- 修复传递等式应用于外连接子句时规划器对选择率的错误估计（Tom）

  这可能导致诸如 `... from a left join b on a.a1 = b.b1 where a.a1 = 42 ...` 之类的查询得到糟糕的计划。
- 改进优化器对长 `IN` 列表的处理（Tom）

  此变更避免了在启用约束排除时在这类列表上浪费大量时间。
- 防止 GIN 索引构建期间的同步扫描（Tom）

  因为 GIN 针对按 TID 递增顺序插入元组做了优化，选择使用同步扫描可能使构建减慢三倍或更多。
- 确保可保持游标的内容不依赖于 TOAST 表的内容（Tom）

  此前，游标结果中的大字段值可能被表示为 TOAST 指针，如果被引用的表在游标被读取之前被删除，或者大值被删除随后被清理掉，这就会失败。普通游标不会发生这种情况，但对于保持到创建事务之后的游标则可能。
- 修复返回集合的函数在未读完整个结果就被终止时的内存泄漏（Tom）
- 修复数据库编码不是 UTF-8 时 XML 函数中的编码转换问题（Tom）
- 修复 `contrib/dblink` 的 `dblink_get_result(text,bool)` 函数（Joe）
- 修复 `contrib/sslinfo` 函数可能的垃圾输出（Tom）
- 修复 `contrib/tsearch2` 兼容触发器在一条命令中被触发多次时的错误行为（Teodor）
- 修复 autovacuum 中可能的错误信号（Heikki）
- 支持在 Windows 7 beta 上作为服务运行（Dave 和 Magnus）
- 修复 ecpg 对 varchar 结构的处理（Michael）
- 修复 configure 脚本，使其在无法获得 PL/Perl 的链接信息时正确报告失败（Andrew）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.91. 版本 8.3.5

发行日期
2008-11-03

本次发布包含对 8.3.4 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.91.1. 迁移到版本 8.3.5

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.1 的版本升级，请参阅第 E.95 节。另外，如果你运行的是之前的 8.3.X 版本，建议在升级后对所有 GiST 索引执行 `REINDEX`。

E.91.2. 变更

- 修复 GiST 索引因删除后把错误的索引条目标记为“死”而损坏的问题（Teodor）

  这会导致索引搜索找不到本应找到的行。损坏的索引可以用 `REINDEX` 修复。

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复 `bytea` 到 XML 映射中可能的崩溃（Michael McMaster）

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）

- 改进 *`expression`* `IN` (*`expression-list`*) 查询的优化（Tom，想法来自 Robert Haas）

  右侧有查询变量的情形在 8.2.x 和 8.3.x 中的处理效率低于以前的版本。此修复为此类情形恢复了 8.1 的行为。

- 修复当子 `SELECT` 出现在 `FROM` 中的函数调用、多行 `VALUES` 列表或 `RETURNING` 列表中时规则查询的错误展开（Tom）

  此问题通常的症状是“无法识别的节点类型”错误。

- 修复 GiST 索引 `IS NULL` 搜索重扫描期间的断言失败（Teodor）

- 修复哈希聚合计划重扫描期间的内存泄漏（Neil）

- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）

- 在 `CREATE DATABASE` 开始复制文件之前强制检查点（Heikki）

  这防止了源数据库中文件最近被删除时可能的失败。

- 防止用 `ALTER SET TABLESPACE` 把表移到另一个表空间时 `relfilenode` 号可能的冲突（Heikki）

  该命令试图重用现有文件名，而不是选取目标目录中已知未使用的文件名。

- 修复单个查询项匹配文本首词时文本搜索标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 让 `ILIKE` 即使在字符被转义时也不区分大小写地比较（Andrew）
- 确保语句日志正确处理 `DISCARD`（Tom）
- 修复 PITR 恢复期间最后完成事务时间的错误日志记录（Tom）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。
- 把 `SessionReplicationRole` 标记为 `PGDLLIMPORT`，使其能在 Windows 上被 Slony 使用（Magnus）
- 修复使用 libpq 的 `gsslib` 参数时的小内存泄漏（Magnus）

  参数字符串占用的空间在连接关闭时未被释放。
- 确保在需要时把 libgssapi 链接到 libpq（Markus Schaaf）
- 修复 ecpg 对 `CREATE ROLE` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 确保 `pg_control` 以二进制模式打开（Itagaki Takahiro）

  pg_controldata 和 pg_resetxlog 此前做得不正确，因此在 Windows 上可能失败。
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.92. 版本 8.3.4

发行日期
2008-09-22

本次发布包含对 8.3.3 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.92.1. 迁移到版本 8.3.4

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.1 的版本升级，请参阅第 E.95 节。

E.92.2. 变更

- 修复 btree WAL 恢复代码中的缺陷（Heikki）

如果 WAL 在页分裂操作中途结束，恢复就会失败。

- 修复 HOT 页修剪可能使用错误截断 XID 的问题（Alvaro）

  此错误在 `VACUUM` 会查询的系统目录中造成损坏风险：死元组版本可能被过早移除。这对实际数据库操作的影响很小，因为系统在检查目录时不遵循 MVCC 规则，但可能导致 pg_dump 或其他客户端程序的瞬时错误输出。

- 修复 `datfrozenxid` 可能的错误计算（Alvaro）

  此错误或许能解释最近一些无法删除旧 `pg_clog` 数据的报告。

- 修复 `pg_class` 被重新索引后的错误 HOT 更新（Tom）

  如果 `REINDEX TABLE pg_class` 之后在同一会话中执行 `ALTER TABLE RENAME` 或 `ALTER TABLE SET SCHEMA` 命令，`pg_class` 可能损坏。

- 修复遗漏的“组合 cid”情形（Karl Schnaitter）

  此错误使行对某个事务错误地不可见，
  而这些行在该事务中曾被多个全部中止的子事务删除过。

- 防止 autovacuum 在它正在检查的表恰好在错误时刻被删除时崩溃（Alvaro）

- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

  这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。

- 修复 GiST 索引扫描期间可能的元组重复输出（Teodor）

- 当任一表被修改时从头重新生成外键检查查询（Tom）

  此前 8.3 会试图重新规划查询，但基于以前生成的查询文本。如果表或列被重命名，这就会失败。

- 修复视图包含简单 `UNION ALL` 构造时遗漏的权限检查（Heikki）

  被引用表的权限被正确检查，但视图本身的权限没有。

- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  据信这种情况在 8.3 中不可能发生，但在以前的版本中可能，因此检查似乎是明智的。

- 修复 `DROP OWNED` 期间可能的重复删除（Tom）

  这通常会导致奇怪的错误，例如“关系 NNN 的缓存查找失败”。

- 修复 XML 操作中的多个内存泄漏（Kris Jurka、Tom）

- 修复 `xmlserialize()`，使其对不可接受的目标数据类型正确抛出错误（Tom）

- 修复几处在文本搜索配置文件解析中错误处理多字节字符的地方（Tom）

  配置文件中出现的某些字符总是会导致“编码的字节序列无效”失败。

- 为文本搜索配置文件中报告的所有错误提供文件名和行号位置（Tom）

- 修复 `AT TIME ZONE`：先尝试把时区参数解释为时区缩写，仅在失败时才尝试完整时区名，而不是像以前那样相反（Tom）

  时间戳输入函数一直按此顺序解析有歧义的时区名。让 `AT TIME ZONE` 也这样做改进了一致性，并修复了 8.1 中引入的兼容性缺陷：在有歧义的情形下，我们现在的行为与 8.0 及更早版本相同，因为在旧版本中 `AT TIME ZONE` 只接受缩写。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 防止显示带单位的配置参数时单位转换期间的整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 允许 `pg_hba.conf` 中 LDAP URL 的后缀部分包含空格（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复可能把 `IS NULL` 测试不当下推到外连接之下的规划器缺陷（Tom）

  这由上层 `OR` 子句的所有分支中对同一关系的 `IS NULL` 测试触发。

- 修复嵌套子查询表达式的规划器缺陷（Tom）

  如果外层子查询不直接依赖父查询而内层依赖，外层的值可能不会为新的父查询行重新计算。

- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/pgSQL 在 `FOR` 循环的目标变量是包含组合类型字段的记录时不失败（Tom）
- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）
- 改进 `PQescapeBytea()` 的性能（Rudolf Leitgeb）
- 在 Windows 上，通过阻止 libpq 在每次系统调用中发送超过 64kB 来绕过 Microsoft 的一个缺陷（Magnus）
- 修复 ecpg 在 `SET` 命令中正确处理变量（Michael）
- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）
- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）
- 修复 pg_standby 中 WAL 文件截断点计算的错误（Simon）
- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.93. 版本 8.3.3

发行日期
2008-06-12

此发行版在 8.3.2 的基础上包含一个严重和一个次要的修复。有关 8.3 主版本新特性的信息，请参阅第 E.96 节。

E.93.1. 迁移到版本 8.3.3

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.1 的版本升级，请参阅第 E.95 节。

E.93.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

- 让 `ALTER AGGREGATE ... OWNER TO` 更新 `pg_shdepend`（Tom）

  如果聚合后来涉及 `DROP OWNED` 或 `REASSIGN OWNED` 操作，这一疏漏可能导致问题。

E.94. 版本 8.3.2

发行日期
从未发布

本次发布包含对 8.3.1 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.94.1. 迁移到版本 8.3.2

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.3.1 的版本升级，请参阅第 E.95 节。

E.94.2. 变更

- 修复在 Windows 上使用 UTF-8 数据库编码和不同客户端编码时发生的 `ERRORDATA_STACK_SIZE exceeded` 崩溃（Tom）

- 修复 `recovery_command` 参数中 `%r` 宏的归档截断点计算错误（Simon）

  如果热备脚本依赖 `%r` 决定何时丢弃 WAL 段文件，这可能导致数据丢失。

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复 `REASSIGN OWNED`，使其也适用于过程语言（Alvaro）

- 修复 `SELECT FOR UPDATE/SHARE` 作为顶层操作非 `SELECT` 的查询中的子查询时的问题（Tom）
- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）
- 修复 `pg_get_ruledef()`，显示 `UPDATE` 或 `DELETE` 目标表上附加的别名（如果有）（Tom）
- 恢复 8.3 之前的行为：TidScan 计划中使用的 TID 的块号超出范围时静默不匹配任何行（Tom）

  8.3.0 和 8.3.1 改为抛出错误。
- 修复可能导致 `too many LWLocks taken` 失败的 GIN 缺陷（Teodor）
- 修复 `tsquery` 损坏的 GiST 比较函数（Teodor）
- 修复 `tsvector_update_trigger()` 和 `ts_stat()`，使其接受其预期工作类型之上的域（Tom）
- 修复不支持枚举数据类型作为外键的问题（Tom）
- 避免解压损坏数据时可能的崩溃（Zdenek Kotala）
- 修复延迟取消链接与 `DROP DATABASE` 之间的竞态条件（Heikki）

  最坏情况下这可能导致删除新数据库中恰好获得与最近删除的表相同 OID 的新建表；当然这是概率极低的情形。
- 修复两处后端 SIGTERM 退出可能在共享内存中留下损坏状态的地方（Tom）

  如果 SIGTERM 用于一起关闭整个数据库集簇，这两种情形都不太重要，但如果有人试图对单个后端发 SIGTERM，就有问题。
- 修复当 x 和 y 数据类型不同时为 x `IN (SELECT` y `FROM ...)` 子句生成的错误计划可能导致的崩溃；并确保从 y 的类型到 x 的转换有损时行为在语义上正确（Tom）
- 修复阻止规划器把已知 Param 值当作常数替换的疏漏（Tom）

  此错误部分禁用了 8.3.0 和 8.3.1 中未命名扩展查询语句的优化：特别是 LIKE 模式作为参数传递时 LIKE 到索引扫描的优化永远不会被应用，依赖参数值的约束排除也不起作用。
- 修复可索引的 `MIN` 或 `MAX` 聚合与 `DISTINCT` 或 `ORDER BY` 一起使用时规划器失败的问题（Tom）
- 修复规划器，确保它从不对向 Sort 节点供数据的计划节点使用“物理 tlist”（Tom）

  这导致排序不得不搬运比实际需要更多的数据，因为未使用的列值被包含在排序数据中。
- 避免不必要的查询字符串复制（Tom）

  这修复了 8.3.0 中引入的当大量命令作为单个查询字符串提交时的性能问题。
- 让 `TransactionIdIsCurrentTransactionId()` 在检查子事务 XID 时使用二分搜索而不是线性搜索（Heikki）

这修复了 8.3.0 在某些情形下明显慢于早期版本的问题。

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）
- 修复几个数据类型输入函数（特别是 `array_in()`），它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。
- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。
- 防止取消为防止 XID 回卷而启动的自动清理（Alvaro）
- 改进 `ANALYZE` 对不确定元组（由尚未提交的事务插入或删除的元组）的处理，使其报告给统计收集器的计数更可能正确（Pavan Deolasee）
- 修复 initdb，拒绝其 `--xlogdir`（`-X`）选项的相对路径（Tom）
- 让 psql 把制表符打印为适当数量的空格，而不是像 8.3.0 和 8.3.1 那样打印 `\x09`（Bruce）
- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴和阿根廷/圣路易斯的 DST 法规变更）
- 为 ecpglib 新增 `ECPGget_PGconn()` 函数（Michael）
- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）
- 修复 ecpg 中续行标记的处理（Michael）
- 修复 `contrib/cube` 函数可能的崩溃（Tom）
- 修复 `contrib/xml2` 的 `xpath_table()` 函数在输入查询返回 NULL 值时的核心转储（Tom）
- 修复 `contrib/xml2` 的 makefile，使其不覆盖 `CFLAGS`，并使其针对 libxslt 是否存在正确自动配置（Tom）

E.95. 版本 8.3.1

发行日期
2008-03-17

本次发布包含对 8.3.0 的多方面修复。有关 8.3 大版本中新特性的信息，请参见第 E.96 节。.

E.95.1. 迁移到版本 8.3.1

对于运行 8.3.X 的用户，不需要进行转储/恢复。不过，如果你受到下述 Windows 区域设置问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.95.2. 变更

- 修复把不同字符组合视为相等的 Windows 区域设置下的字符串比较（Tom）

  此修复仅在 Windows 上且仅在使用 UTF-8 数据库编码时适用。所有其他情形的相同修复两年多前就已完成，但使用 UTF-8 的 Windows 走的是未更新的单独代码路径。如果你使用的区域设置把某些不同的字符串视为相等，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 修复 `VACUUM FULL` 的极端情形缺陷（Tom）

  8.2 中引入了不同系统目录上并发 `VACUUM FULL` 操作之间的潜在死锁。这已被纠正。8.3 使情况更糟，因为死锁可能发生在临界代码区内，使它成为 PANIC 而不只是 ERROR。

  此外，在清理系统目录中途失败的 `VACUUM FULL` 可能导致并发数据库会话中的缓存损坏。

  8.3 中引入的另一个 `VACUUM FULL` 缺陷在处理不含活元组的页时可能导致崩溃或内存不足报告。

- 修复涉及 `character` 或 `bit` 列的外键检查的错误行为（Tom）

  如果引用列是不同但兼容的类型（例如 `varchar`），约束的强制执行就不正确。

- 避免空操作外键检查中不必要的死锁失败（Stephan Szabo、Tom）

- 修复重新规划预备查询时可能的核心转储（Tom）

  此缺陷只影响协议级 prepare 操作，不影响 SQL `PREPARE`，因此往往只出现在大量使用预备语句的 JDBC、DBI 和其他客户端驱动上。

- 修复重新规划调用使用 SPI 的函数的查询时可能的失败（Tom）

- 修复涉及不同数据类型列的按行比较失败（Tom）

- 修复长期存在的 `LISTEN`/`NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 不允许在预备事务内使用 `LISTEN` 和 `UNLISTEN`（Tom）

  这以前是允许的，但尝试这样做会有各种不愉快的后果，特别是只要 `UNLISTEN` 未提交，发起的后端就无法退出。

- 不允许在预备事务内删除临时表（Heikki）

  8.1 正确地禁止了这一点，但该检查在 8.2 和 8.3 中被无意破坏。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）
- 修复 `tsquery` 值的错误比较（Teodor）
- 修复单字节编码中 `LIKE` 对非 ASCII 字符的错误行为（Rolf Jentsch）
- 禁用 `xmlvalidate`（Tom）

  此函数本应在 8.3 发布前移除，但被无意留在了源代码中。它带来小的安全风险，因为非特权用户可以用它读取服务器可访问的任何文件的前几个字符。
- 修复返回集合函数某些用法中的内存泄漏（Neil）
- 让 `encode(`*`bytea`*`, 'escape')` 把所有高位字节值转换为 `\`*`nnn`* 八进制转义序列（Tom）

  这对于在数据库编码为多字节时避免编码问题是必要的。此变更可能给期望 `encode` 特定结果的应用程序带来兼容性问题。
- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。
- 修复 `ALTER OWNER` 某些变体中的“无法识别的节点类型”错误（Tom）
- 避免 `CREATE TABLE LIKE INCLUDING INDEXES` 中的表空间权限错误（Tom）
- 确保锁等待中止时清除 `pg_stat_activity.waiting` 标志（Tom）
- 修复 Windows Vista 上进程权限的处理（Dave、Magnus）

  特别是，此修复允许以 Administrator 用户启动服务器。
- 更新时区数据文件到 tzdata 发行版 2008a（特别是智利最近的变更）；调整时区缩写 `VET`（委内瑞拉）为 UTC-4:30 而不是 UTC-4:00（Tom）
- 修复 ecpg 的数组问题（Michael）
- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 `pg_ctl start -w` 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。
- 使用 `-fwrapv` 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。
- 支持用 MSVC 构建 `contrib/uuid-ossp`（Hiroshi Saito）

E.96. 版本 8.3

发行日期
2008-02-04

E.96.1. 概述

凭借重要的新功能和性能增强，此发行版代表了 PostgreSQL 的重大飞跃。这得益于一个不断壮大的社区，它显著加快了开发步伐。此发行版增加了以下主要特性：

- 全文搜索被集成到核心数据库系统中
- 支持 SQL/XML 标准，包括新操作符和 `XML` 数据类型
- 枚举数据类型（`ENUM`）
- 组合类型的数组
- 通用唯一标识符（`UUID`）数据类型
- 新增对 `NULL` 排在最前还是最后的控制
- 可更新游标
- 服务器配置参数现在可以逐函数设置
- 用户定义类型现在可以带类型修饰符
- 当表定义更改或统计信息更新时自动重新规划缓存查询
- 日志和统计信息收集的大量改进
- 支持在 Windows 上使用安全服务提供程序接口（SSPI）进行认证
- 支持多个并发 autovacuum 进程及其他 autovacuum 改进
- 允许整个 PostgreSQL 发行版用 Microsoft Visual C++ 编译

主要性能改进如下所列。这些增强大多是自动的，不需要用户更改或调优：

- 异步提交延迟事务提交期间对 WAL 的写入
- 检查点写入可以分散到更长的时间段，以平滑每个检查点期间的 I/O 峰值
- 仅堆元组（HOT）加速大多数 `UPDATE` 和 `DELETE` 的空间复用
- 即时后台写入器策略改进磁盘写入效率
- 对只读事务使用非持久事务 ID 减少开销和 `VACUUM` 需求
- 逐字段和逐行的存储开销已被降低
- 大型顺序扫描不再驱逐频繁使用的缓存页
- 并发的大型顺序扫描现在可以共享磁盘读取
- `ORDER BY ... LIMIT` 可以不排序完成

上述各项在下面的章节中有更详细的说明。

E.96.2. 迁移到版本 8.3

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

E.96.2.1. 常规

- 非字符数据类型不再自动转换为 `TEXT`（Peter、Tom）

  此前，如果把非字符值提供给要求 `text` 输入的操作符或函数，它会被自动转换为 `text`（对大多数但非全部内建数据类型）。这不再发生：现在所有非字符串类型都需要显式转换为 `text`。例如，这些表达式以前可用：

  ```
  substr(current_date, 1, 4)
  23 LIKE '2%'
  ```

  但现在分别会报“函数不存在”和“操作符不存在”错误。请改用显式转换：

  ```
  substr(current_date::text, 1, 4)
  23::text LIKE '2%'
  ```

  （当然，也可以使用更冗长的 `CAST()` 语法。）此变更的原因是这些自动转换太常引起令人惊讶的行为。例如，在以前的版本中，这个表达式被接受但做的不是预期的事：

  ```
  current_date < 2017-11-17
  ```

  这实际上是在把日期与整数比较，本应被拒绝（现在也是）——但在存在自动转换时，两边都被转换为 `text` 并做了文本比较，因为没有其他 `<` 操作符能匹配该表达式时 `text < text` 操作符可以匹配。

  `char(n)` 和 `varchar(n)` 类型仍然自动转换为 `text`。此外，只要至少一个输入是字符串类型，到 `text` 的自动转换对连接（`||`）操作符的输入仍然有效。

- `contrib/tsearch2` 的全文搜索特性已移入核心服务器，语法有少量变化

  `contrib/tsearch2` 现在包含一个兼容接口。

- `ARRAY(SELECT ...)` 在 `SELECT` 不返回行时现在返回空数组而不是 NULL（Tom）

- 基本数据类型的数组类型名不再总是基本类型名加下划线前缀

  旧的命名约定在可能时仍被沿用，但应用程序代码不应再依赖它。请改用新的 `pg_type.typarray` 列来标识与给定类型关联的数组数据类型。

- `ORDER BY ... USING` *`operator`* 现在必须使用 btree 操作符类中定义的小于或大于 *`operator`*

  添加此限制是为了防止不一致的结果。

- `SET LOCAL` 的更改现在持续到最外层事务结束，除非回滚（Tom）

  此前 `SET LOCAL` 的效果在子事务提交（`RELEASE SAVEPOINT` 或退出 PL/pgSQL 异常块）后丢失。

- 在事务块中被拒绝的命令现在在多语句查询字符串中也被拒绝（Tom）

  例如，`"BEGIN; DROP DATABASE; COMMIT"` 现在即使作为单个查询消息提交也会被拒绝。

- 事务块外的 `ROLLBACK` 现在发出 `NOTICE` 而不是 `WARNING`（Bruce）

- 防止 NOTIFY/LISTEN/UNLISTEN 接受模式限定的名称（Bruce）

  此前这些命令接受 schema.relation 但忽略模式部分，这令人困惑。

- ALTER SEQUENCE 不再影响序列的 currval() 状态（Tom）

- 外键现在对于跨数据类型引用必须匹配可索引的条件（Tom）

  这改进了语义一致性并有助于避免性能问题。

- 把对象大小函数限制为有合理权限查看此类信息的用户（Tom）

  例如，pg_database_size() 现在需要 CONNECT 权限（默认授予所有人）。pg_tablespace_size() 需要表空间中的 CREATE 权限，或者如果该表空间是数据库的默认表空间则允许。

- 移除未记录的 !!=（not in）操作符（Tom）

  NOT IN (SELECT ...) 是执行此操作的正确方式。

- 内部哈希函数现在分布更均匀（Tom）

  如果应用程序代码使用 PostgreSQL 内部哈希函数计算并存储哈希值，必须重新生成这些哈希值。

- 处理变长数据值的 C 代码约定已更改（Greg Stark、Tom）

  必须使用新的 SET_VARSIZE() 宏来设置生成的 varlena 值的长度。此外，可能需要在更多情形下展开（“去 TOAST”）输入值。

- 连续归档不再把每个成功的归档操作报告给服务器日志，除非使用 DEBUG 级别（Simon）

E.96.2.2. 配置参数

- 管理服务器参数的大量变更

  bgwriter_lru_percent、bgwriter_all_percent、bgwriter_all_maxpages、stats_start_collector 和 stats_reset_on_server_start 被移除。redirect_stderr 改名为 logging_collector。stats_command_string 改名为 track_activities。stats_block_level 和 stats_row_level 合并为 track_counts。新的布尔配置参数 archive_mode 控制归档。Autovacuum 的默认设置已更改。

- 移除 stats_start_collector 参数（Tom）

  现在总是启动收集器进程，除非 UDP 套接字创建失败。

- 移除 stats_reset_on_server_start 参数（Tom）

  移除它是因为 pg_stat_reset() 可用于此目的。

- 注释掉 postgresql.conf 中的参数现在使其恢复默认值（Joachim Wieland）

  此前，注释掉条目使参数值保持不变直到下一次服务器重启。

E.96.2.3. 字符编码

- 新增更多对无效编码数据的检查（Andrew）

此变更堵住了字面反斜杠转义字符串处理和 `COPY` 转义处理中存在的一些漏洞。现在对去转义后的字符串重新检查结果是否产生了无效的多字节字符。

- 不允许与服务器区域设置不一致的数据库编码（Tom）

  在大多数平台上，`C` 区域设置是唯一能用于任何数据库编码的区域设置。其他区域设置隐含特定编码，如果数据库编码不同就会行为异常。（典型症状包括文本排序顺序错误以及 `upper()` 或 `lower()` 结果错误。）服务器现在拒绝创建编码不兼容的数据库。

- 确保 `chr()` 不能创建无效编码的值（Andrew）

  在 UTF8 编码的数据库中，`chr()` 的参数现在被视为 Unicode 码点。在其他多字节编码中，`chr()` 的参数必须指定 7 位 ASCII 字符。不再接受零。`ascii()` 已相应调整。

- 调整 `convert()` 的行为以确保编码有效性（Andrew）

  `convert()` 的两参数形式已被移除。三参数形式现在接受 `bytea` 作为第一个参数并返回 `bytea`。为弥补功能损失，新增了三个函数：

  - `convert_from(bytea, name)` 返回 `text`——把第一个参数从指定编码转换为数据库编码
  - `convert_to(text, name)` 返回 `bytea`——把第一个参数从数据库编码转换为指定编码
  - `length(bytea, name)` 返回 `integer`——给出第一个参数按指定编码的字符长度

- 移除 `convert(argument USING conversion_name)`（Andrew）

  其行为不符合 SQL 标准。

- 把 JOHAB 编码改为仅客户端（Tatsuo）

  JOHAB 作为服务器端编码不安全。

E.96.3. 变更

下面你将找到 PostgreSQL 8.3 与上一个主版本之间变更的详细说明。

E.96.3.1. 性能

- 异步提交延迟事务提交期间对 WAL 的写入（Simon）

  此特性显著提高了短的数据修改事务的性能。缺点是由于磁盘写入被延迟，
  如果数据库或操作系统在数据写入磁盘之前崩溃，已提交的数据将丢失。
  此特性对能接受一些数据丢失的应用程序有用。与关闭 `fsync` 不同，使用异步提交不会使数据库一致性面临风险；最坏的情况是崩溃后最后几个报告已提交的事务可能实际上没有提交。
  通过关闭 `synchronous_commit` 启用此特性（如果某些事务关键而另一些不关键，可以逐会话或逐事务进行）。可以调整 `wal_writer_delay` 来控制事务实际到达磁盘前的最大延迟。

- 检查点写入可以分散到更长的时间段，以平滑每个检查点期间的 I/O 峰值（Itagaki Takahiro 和 Heikki Linnakangas）

  此前所有修改过的缓冲区在检查点期间被尽快强制写盘，造成降低服务器性能的 I/O 峰值。新方法把检查点期间的磁盘写分散开，降低峰值 I/O 使用。（用户请求的检查点和关闭检查点仍然尽快写入。）

- 仅堆元组（HOT）加速大多数 `UPDATE` 和 `DELETE` 的空间复用（Pavan Deolasee，想法来自许多其他人）

  `UPDATE` 和 `DELETE` 会留下死元组，失败的 `INSERT` 也是。此前只有 `VACUUM` 能回收死元组占用的空间。有了 HOT，如果没有对索引列做更改，死元组空间可以在 `INSERT` 或 `UPDATE` 时自动回收。这带来更一致的性能。此外，HOT 避免添加重复的索引条目。

- 即时后台写入器策略改进磁盘写入效率（Greg Smith、Itagaki Takahiro）

  这大幅减少了对后台写入器的手动调优需求。

- 逐字段和逐行的存储开销已被降低（Greg Stark、Heikki Linnakangas）

  数据值短于 128 字节的变长数据类型的存储将减少 3 到 6 字节。例如，两个相邻的 `char(1)` 字段现在用 4 字节而不是 16 字节。行头也比以前短 4 字节。

- 对只读事务使用非持久事务 ID 减少开销和 `VACUUM` 需求（Florian Pflug）

  非持久事务 ID 不递增全局事务计数器。因此，它们降低了 `pg_clog` 的负载，并延长了为防止事务 ID 回卷而强制清理之间的时间间隔。还做了其他应能改进并发性的性能改进。

- 避免在只读命令之后递增命令计数器（Tom）

  此前每个事务有 2^{32}（40 亿）条命令的硬限制。现在只有实际更改数据库的命令才计数，因此虽然该限制仍存在，但应该明显不那么碍事了。

- 创建专门的 WAL 写入器进程来分担后端的工作（Simon）

- 跳过 `CLUSTER` 和 `COPY` 不必要的 WAL 写入（Simon）

  除非启用了 WAL 归档，系统现在避免为 `CLUSTER` 写 WAL，只是在命令结束时对表执行 `fsync()`。如果表在同一事务中创建，`COPY` 也同样处理。

- 大型顺序扫描不再驱逐频繁使用的缓存页（Simon、Heikki、Tom）

- 并发的大型顺序扫描现在可以共享磁盘读取（Jeff Davis）

  这是通过在表中间（另一个顺序扫描已在进行的地方）
  开始新的顺序扫描并回绕到开头来完成来实现的。这会影响未指定 `ORDER BY` 的查询返回行的顺序。必要时可以用 `synchronize_seqscans` 配置参数禁用它。

- `ORDER BY ... LIMIT` 可以不排序完成（Greg Stark）

  这是通过顺序扫描表并只跟踪“前 N”候选行而不是对整表做完全排序来完成的。这在没有匹配索引且 `LIMIT` 不大时很有用。

- 对后端发送给统计收集器的消息设置速率限制（Tom）

  这降低了短事务的开销，但有时可能增加统计汇总前的延迟。

- 改进有许多 NULL 情形下的哈希连接性能（Tom）

- 加快非精确数据类型匹配情形下的操作符查找（Tom）

E.96.3.2. 服务器

- Autovacuum 现在默认启用（Alvaro）

做了若干更改以消除启用 autovacuum 的缺点，从而证明默认值变更的合理性。还修改了几个其他 autovacuum 参数的默认值。

- 支持多个并发 autovacuum 进程（Alvaro、Itagaki Takahiro）

  这允许多个清理并发运行。这防止大表的清理延迟较小表的清理。

- 当表定义更改或统计信息更新时自动重新规划缓存查询（Tom）

  此前，引用临时表的 PL/pgSQL 函数在临时表于函数调用之间被删除并重建时会失败，除非使用 `EXECUTE`。此改进修复了该问题及许多相关问题。

- 新增 `temp_tablespaces` 参数，控制临时表和文件的表空间（Jaime Casanova、Albert Cervera、Bernd Helmle）

  此参数定义要使用的表空间列表。这使 I/O 负载可以分散到多个表空间。每次创建临时对象时随机选择一个表空间。临时文件不再存储在逐数据库的 `pgsql_tmp/` 目录中，而是存储在逐表空间的目录中。

- 把临时表的 TOAST 表放在名为 `pg_toast_temp_`*`nnn`* 的特殊模式中（Tom）

  这使低层代码能把这些表识别为临时的，从而支持各种优化，例如不写 WAL 日志更改以及使用本地而非共享缓冲区访问。这还修复了后端意外持有临时 TOAST 表打开文件引用的缺陷。

- 修复新连接请求持续不断可能无限期延迟 postmaster 完成关闭或崩溃重启的问题（Tom）

- 通过防止重用已删除表的 relfilenode 直到下一个检查点之后，防范概率极低的数据丢失情形（Heikki）

- 修复 `CREATE CONSTRAINT TRIGGER`，把旧式外键触发器定义转换为常规外键约束（Tom）

  这将便于移植从 7.3 之前数据库延续下来的外键约束，如果它们从未用 `contrib/adddepend` 转换过。

- 修复 `DEFAULT NULL` 以覆盖继承的默认值（Tom）

  `DEFAULT NULL` 以前被视为噪声短语，但它应该（现在也确实）覆盖否则会从父表或域继承的非空默认值。

- 新增编码 EUC_JIS_2004 和 SHIFT_JIS_2004（Tatsuo）

  这些新编码可以与 UTF-8 相互转换。

- 把服务器启动日志消息从“数据库系统就绪”改为“数据库系统准备好接受连接”，并调整其时机

  该消息现在只在 postmaster 真正准备好接受连接时出现。

E.96.3.3. 监控

- 新增 `log_autovacuum_min_duration` 参数，支持可配置的 autovacuum 活动日志（Simon、Alvaro）

- 新增 `log_lock_waits` 参数来记录锁等待（Simon）

- 新增 `log_temp_files` 参数来记录临时文件使用（Bill Moran）

- 新增 `log_checkpoints` 参数以改进检查点的日志记录（Greg Smith、Heikki）
- `log_line_prefix` 现在在所有进程中支持 `%s` 和 `%c` 转义（Andrew）

  此前这些转义只对用户会话有效，对后台数据库进程无效。
- 新增 `log_restartpoints` 来控制时间点恢复重启点的日志记录（Simon）
- 现在在恢复结束和每个被记录的重启点记录最后事务结束时间（Simon）
- Autovacuum 现在在 `pg_stat_activity` 中报告其活动开始时间（Tom）
- 允许服务器日志输出为逗号分隔值（CSV）格式（Arul Shaji、Greg Smith、Andrew Dunstan）

  CSV 格式的日志文件可以轻松装入数据库表以供后续分析。
- 使用 PostgreSQL 提供的时区支持来格式化服务器日志中显示的时间戳（Tom）

  这避免了本地化时区名编码错误这一 Windows 特有的问题。新增了 `log_timezone` 参数，独立于客户端可见的 `timezone` 参数控制日志消息中使用的时区。
- 新系统视图 `pg_stat_bgwriter` 显示后台写入器活动的统计信息（Magnus）
- 为 `pg_stat_database` 新增数据库范围元组统计的列（Magnus）
- 为 `pg_stat_activity` 新增 `xact_start`（事务开始时间）列（Neil）

  这使识别长运行事务更容易。
- 为 `pg_stat_all_tables` 及相关视图新增 `n_live_tuples` 和 `n_dead_tuples` 列（Glen Parker）
- 把 `stats_block_level` 和 `stats_row_level` 参数合并为单一参数 `track_counts`，它控制发送给统计收集器进程的所有消息（Tom）
- 把 `stats_command_string` 参数改名为 `track_activities`（Tom）
- 修复活元组和死元组的统计计数，使其认识到已提交和中止的事务有不同的影响（Tom）

E.96.3.4. 认证

- 支持在 Windows 上使用安全服务提供程序接口（SSPI）进行认证（Magnus）
- 支持 GSSAPI 认证（Henry Hotz、Magnus）

  应优先于本机 Kerberos 认证使用它，因为 GSSAPI 是行业标准。
- 支持全局 SSL 配置文件（Victor Wagner）
- 新增 `ssl_ciphers` 参数，控制接受的 SSL 密码（Victor Wagner）
- 新增 Kerberos realm 参数 `krb_realm`（Magnus）

E.96.3.5. Write-Ahead Log (WAL) and Continuous Archiving

- 把事务 WAL 记录中记录的时间戳从 time_t 改为 TimestampTz 表示（Tom）

这为 WAL 提供了亚秒分辨率，对时间点恢复很有用。

- 减少热备服务器所需的 WAL 磁盘空间（Simon）

  此变更允许热备服务器把最早仍需要的 WAL 文件名传递给恢复脚本，从而自动删除不再需要的 WAL 文件。这是通过 `recovery.conf` 的 `restore_command` 参数中的 `%r` 完成的。

- 新的布尔配置参数 `archive_mode` 控制归档（Simon）

  此前把 `archive_command` 设为空字符串会关闭归档。现在 `archive_mode` 独立于 `archive_command` 控制归档的开关。这对临时停止归档很有用。

E.96.3.6. 查询

- 全文搜索被集成到核心数据库系统（Teodor、Oleg）

  文本搜索得到改进，移入核心代码，现在默认安装。`contrib/tsearch2` 现在包含一个兼容接口。

- 新增对 `NULL` 排在最前还是最后的控制（Teodor、Tom）

  语法为 `ORDER BY ... NULLS FIRST/LAST`。

- 允许索引的逐列升序/降序（`ASC/DESC`）排序选项（Teodor、Tom）

  此前混合使用 `ASC/DESC` 说明符的 `ORDER BY` 查询无法完全使用索引。现在，如果索引创建时带有匹配的 `ASC/DESC` 说明，这类情形可以完全使用索引。索引内的 `NULL` 排序顺序也可以控制。

- 允许 `col IS NULL` 使用索引（Teodor）

- 可更新游标（Arul Shaji、Tom）

  这消除了为 `UPDATE` 或 `DELETE` 游标返回的行而引用主键的需要。语法为 `UPDATE/DELETE WHERE CURRENT OF`。

- 允许游标中使用 `FOR UPDATE`（Arul Shaji、Tom）

- 创建一个通用机制，通过调用数据类型的 I/O 函数，支持每个数据类型与标准字符串类型（`TEXT`、`VARCHAR`、`CHAR`）之间的转换（Tom）

  此前，这类转换只对有专门函数的类型可用。这些新转换在转字符串方向只允许赋值转换，另一方向只允许显式转换，因此不应产生令人惊讶的行为。

- 当所有输入都是同一域类型时，允许 `UNION` 及相关构造返回域类型（Tom）

  此前，输出会被认为属于域的基本类型。

- 允许在使用两种不同数据类型时进行有限哈希（Tom）

  如果数据类型有兼容的哈希函数，这允许在涉及跨数据类型比较的情形中使用哈希连接、哈希索引、哈希子计划和哈希聚合。目前，跨数据类型哈希支持存在于 `smallint/integer/bigint` 以及 `float4/float8`。

- 改进优化器检测 `WHERE` 子句中变量何时相等的逻辑（Tom）

这允许归并连接以降序排序工作，并改进对冗余排序列的识别。

- 在大多数表被约束排除的大继承树情形下改进规划性能（Tom）

E.96.3.7. 对象操作

- 组合类型的数组（David Fetter、Andrew、Tom）

  除显式声明的组合类型的数组外，现在还支持常规表和视图的行类型的数组，但不包括系统目录、序列和 TOAST 表的行类型。

- 服务器配置参数现在可以逐函数设置（Tom）

  例如，函数现在可以设置自己的 `search_path`，以防止运行时存在不同 `search_path` 时出现意外行为。安全定义者函数应设置 `search_path` 以避免安全漏洞。

- `CREATE/ALTER FUNCTION` 现在支持 `COST` 和 `ROWS` 选项（Tom）

  `COST` 允许指定函数调用的代价。`ROWS` 允许指定返回集合函数返回的平均行数。这些值由优化器在选择最佳计划时使用。

- 实现 `CREATE TABLE LIKE ... INCLUDING INDEXES`（Trevor Hardcastle、Nikhil Sontakke、Neil）

- 允许 `CREATE INDEX CONCURRENTLY` 忽略其他数据库中的事务（Simon）

- 新增 `ALTER VIEW ... RENAME TO` 和 `ALTER SEQUENCE ... RENAME TO`（David Fetter、Neil）

  此前这只能通过 `ALTER TABLE ... RENAME TO` 完成。

- 让 `CREATE/DROP/RENAME DATABASE` 在失败前短暂等待冲突的后端退出（Tom）

  这提高了这些命令成功的可能性。

- 允许使用配置参数成组停用触发器和规则，以用于复制（Jan）

  这允许复制系统成组禁用触发器和重写规则，而无需直接修改系统目录。该行为由 `ALTER TABLE` 和新参数 `session_replication_role` 控制。

- 用户定义类型现在可以带类型修饰符（Teodor、Tom）

  这允许用户定义类型带修饰符，例如 `ssnum(7)`。此前只有内建数据类型可以有修饰符。

E.96.3.8. 工具命令

- 非超级用户的数据库所有者现在默认可以向其数据库添加可信过程语言（Jeremy Drake）

  虽然这相当安全，但一些管理员可能希望撤销该权限。它由 `pg_pltemplate.tmpldbacreate` 控制。

- 允许把会话的当前参数设置用作未来会话的默认值（Tom）

  这通过 `CREATE/ALTER FUNCTION`、`ALTER DATABASE` 或 `ALTER ROLE` 中的 `SET ... FROM CURRENT` 完成。

- 实现新命令 `DISCARD ALL`、`DISCARD PLANS`、`DISCARD TEMPORARY`、`CLOSE ALL` 和 `DEALLOCATE ALL`（Marko Kreen、Neil）

  这些命令简化了把数据库会话重置为初始状态，对连接池软件特别有用。

- 使 `CLUSTER` MVCC 安全（Heikki Linnakangas）

  此前，`CLUSTER` 会丢弃所有已提交为死的元组，即使按照 MVCC 可见性规则仍有事务应当能看到它们。

- 新增 `CLUSTER` 语法：`CLUSTER` *`table`* `USING` *`index`*（Holger Schurig）

  旧的 `CLUSTER` 语法仍受支持，但新形式被认为更合逻辑。

- 修复 `EXPLAIN`，使其能更准确地显示复杂计划（Tom）

  对子计划输出的引用现在总是正确显示，而不是在复杂情形使用 `?column`*`N`*`?`。

- 限制删除用户时报告的信息量（Alvaro）

  此前，删除（或试图删除）拥有许多对象的用户会产生列出所有这些对象的大型 `NOTICE` 或 `ERROR` 消息；这给一些客户端应用程序带来问题。现在消息长度受限，但完整列表仍会发送到服务器日志。

E.96.3.9. 数据类型

- 支持 SQL/XML 标准，包括新操作符和 `XML` 数据类型（Nikolay Samokhvalov、Pavel Stehule、Peter）

- 枚举数据类型（`ENUM`）（Tom Dunstan）

  此特性为有小的固定允许值集合的字段提供便利支持。创建 `ENUM` 类型的例子是 `CREATE TYPE mood AS ENUM ('sad', 'ok', 'happy')`。

- 通用唯一标识符（`UUID`）数据类型（Gevik Babakhani、Neil）

  这与 RFC 4122 紧密匹配。

- 把 `MONEY` 数据类型拓宽到 64 位（D'Arcy Cain）

  这大幅增加了支持的 `MONEY` 值的范围。

- 修复 `float4`/`float8`，使其一致地处理 `Infinity` 和 `NAN`（非数字）（Bruce）

  以前的代码在区分 `Infinity` 与溢出条件方面不一致。

- 允许 `boolean` 值输入时有前导和尾随空白（Neil）

- 防止 `COPY` 用数字和小写字母作为分隔符（Tom）

E.96.3.10. 函数

- 新增正则表达式函数 `regexp_matches()`、`regexp_split_to_array()` 和 `regexp_split_to_table()`（Jeremy Drake、Neil）

  这些函数提供正则表达式子表达式的提取，并允许用 POSIX 正则表达式拆分字符串。

- 新增用于大对象截断的 `lo_truncate()`（Kris Jurka）
- 为 `float8` 数据类型实现 `width_bucket()`（Neil）
- 新增 `pg_stat_clear_snapshot()`，丢弃当前事务期间收集的统计快照（Tom）

  事务中第一次请求统计信息时会取一个在事务期间不变的统计快照。
  此函数允许丢弃该快照并在下一次统计查询期间装载新快照。这对被限制在单个事务中的 PL/pgSQL 函数特别有用。
- 为 `EXTRACT()` 和 `date_part()` 新增 `isodow` 选项（Bruce）

  它返回星期几，星期日为七。（`dow` 返回星期日为零。）
- 为 `to_char()`、`to_date()` 和 `to_timestamp()` 新增 `ID`（ISO 星期几）和 `IDDD`（ISO 年积日）格式代码（Brendan Jurd）
- 让 `to_timestamp()` 和 `to_date()` 对可能变宽的字段假定 `TM`（修剪）选项（Bruce）

  这与 Oracle 的行为匹配。
- 修复 `to_date()`/`to_timestamp()` 的 `D`（非 ISO 星期几）字段中的差一转换错误（Bruce）
- 让 `setseed()` 返回 void，而不是无用的整数值（Neil）
- 为 `NUMERIC` 新增哈希函数（Neil）

  这允许 `NUMERIC` 列使用哈希索引和基于哈希的计划。
- 改进 `LIKE`/`ILIKE` 的效率，特别是对 UTF-8 等多字节字符集（Andrew、Itagaki Takahiro）
- 让 `currtid()` 函数要求目标表上的 `SELECT` 权限（Tom）
- 新增若干 `txid_*()` 函数来查询活动事务 ID（Jan）

  这对各种复制解决方案很有用。

E.96.3.11. PL/pgSQL 服务器端语言

- 新增可滚动游标支持，包括 `FETCH` 中的方向控制（Pavel Stehule）
- 为了与后端的 `FETCH` 命令一致，允许在 PL/pgSQL 的 `FETCH` 语句中用 `IN` 替代 `FROM`（Pavel Stehule）
- 为 PL/pgSQL 新增 `MOVE`（Magnus、Pavel Stehule、Neil）
- 实现 `RETURN QUERY`（Pavel Stehule、Neil）

  这为想返回查询结果的 PL/pgSQL 返回集合函数新增了便利语法。`RETURN QUERY` 比围绕 `RETURN NEXT` 的循环更容易也更高效。
- 允许函数参数名用函数名限定（Tom）

  例如 `myfunc.myvar`。这在变量名可能与列名匹配的查询中指定变量时特别有用。
- 使用块标签限定变量正确工作（Tom）

此前，外层块标签可能意外干扰内层记录或行引用的识别。

- 收紧对 FOR 循环 STEP 值的要求（Tom）

  防止非正的 STEP 值，并处理循环溢出。

- 改进报告语法错误位置时的准确性（Tom）

E.96.3.12. 其他服务器端语言

- 允许 PL/Perl spi_prepare() 的类型名参数除了 pg_type 中的名称外还可以是数据类型别名（Andrew）
- 允许 PL/Python plpy.prepare() 的类型名参数除了 pg_type 中的名称外还可以是数据类型别名（Andrew）
- 允许 PL/Tcl spi_prepare 的类型名参数除了 pg_type 中的名称外还可以是数据类型别名（Andrew）
- 使 PL/PythonU 能在 Python 2.5 上编译（Marko Kreen）
- 在兼容的 Python 版本（Python 2.3 及更高版本）中支持真正的 PL/Python 布尔类型（Marko Kreen）
- 修复启用线程的 libtcl 在后端内产生多个线程导致的 PL/Tcl 问题（Steve Marshall、Paul Bayer、Doug Knight）

  这造成了各种各样的麻烦。

E.96.3.13. psql

- 在 \d 输出中单独列出已禁用的触发器（Brendan Jurd）
- 在 \d 模式中，始终按字面匹配 $（Tom）
- 在 \da 输出中显示聚合返回类型（Greg Sabino Mullane）
- 把函数的易变性状态添加到 \df+ 的输出中（Neil）
- 新增 \prompt 功能（Chad Wagner）
- 允许 \pset、\t 和 \x 指定 on 或 off，而不是仅仅切换（Chad Wagner）
- 新增 \sleep 功能（Jan）
- 为 \copy 启用 \timing 输出（Andrew）
- 改进 Windows 上 \timing 的分辨率（Itagaki Takahiro）
- 在每个反斜杠命令后刷新 \o 输出（Tom）
- 正确检测和报告读取 -f 输入文件期间的错误（Peter）
- 移除 -u 选项（此选项早已废弃）（Tom）

E.96.3.14. pg_dump

- 为 pg_dumpall 新增 `--tablespaces-only` 和 `--roles-only` 选项（Dave Page）
- 为 pg_dumpall 新增输出文件选项（Dave Page）

  这主要在 Windows 上有用，在那里子 pg_dump 进程的输出重定向不起作用。
- 允许 pg_dumpall 接受初始连接数据库名而不是默认的 `template1`（Dave Page）
- 在 `-n` 和 `-t` 开关中，始终按字面匹配 `$`（Tom）
- 改进数据库有数千个对象时的性能（Tom）
- 移除 `-u` 选项（此选项早已废弃）（Tom）

E.96.3.15. 其他客户端应用程序

- 在 initdb 中，允许指定 `pg_xlog` 目录的位置（Euler Taveira de Oliveira）
- 在支持的操作系统上启用 pg_regress 中的服务器核心转储生成（Andrew）
- 为 pg_ctl 新增 `-t`（超时）参数（Bruce）

  这控制 pg_ctl 等待服务器启动或关闭的时间。此前超时被硬编码为 60 秒。
- 新增 pg_ctl 选项来控制服务器核心转储的生成（Andrew）
- 允许用 Control-C 取消 clusterdb、reindexdb 和 vacuumdb（Itagaki Takahiro、Magnus）
- 抑制 createdb、createuser、dropdb 和 dropuser 的命令标签输出（Peter）

  `--quiet` 选项被忽略并将在 8.4 中移除。对所有数据库操作时的进度消息现在发送到 stdout 而不是 stderr，因为它们实际上不是错误。

E.96.3.16. libpq

- 如果 `PQsetdbLogin()` 的 `dbName` 参数包含等号，把它解释为 `conninfo` 字符串（Andrew）

  这允许仍在使用 `PQsetdbLogin()` 的客户端程序使用 `conninfo` 字符串。
- 支持全局 SSL 配置文件（Victor Wagner）
- 新增环境变量 `PGSSLKEY` 来控制 SSL 硬件密钥（Victor Wagner）
- 新增用于大对象截断的 `lo_truncate()`（Kris Jurka）
- 新增 `PQconnectionNeedsPassword()`，如果服务器要求口令但未提供则返回真（Joe Conway、Tom）

  如果在连接尝试失败后它返回真，客户端应用程序应提示用户输入口令。
  过去应用程序必须检查特定的错误消息字符串来决定是否需要口令；该方法现已废弃。
- 新增 `PQconnectionUsedPassword()`，如果提供的口令确实被使用则返回真（Joe Conway、Tom）

这在某些需要知道用户提供的口令是否确实有效的安全上下文中很有用。

E.96.3.17. ecpg

- 使用 V3 前端/后端协议（Michael）

  这新增了对服务器端预备语句的支持。

- 在 Windows 上使用本机线程而不是 pthreads（Magnus）
- 改进 ecpglib 的线程安全（Itagaki Takahiro）
- 让 ecpg 库只导出必要的 API 符号（Michael）

E.96.3.18. Windows Port

- 允许整个 PostgreSQL 发行版用 Microsoft Visual C++ 编译（Magnus 等）

  这使基于 Windows 的开发者可以使用熟悉的开发和调试工具。用 Visual C++ 制作的 Windows 可执行文件也可能比用其他工具集制作的具有更好的稳定性和性能。仅客户端的 Visual C++ 构建脚本已被移除。

- 大幅降低 postmaster 有许多子进程时的内存使用（Magnus）
- 允许由管理用户启动回归测试（Magnus）
- 新增本机共享内存实现（Magnus）

E.96.3.19. Server Programming Interface (SPI)

- 在 SPI 中新增游标相关功能（Pavel Stehule）

  允许访问与游标相关的规划选项，并新增 `FETCH/MOVE` 例程。

- 允许通过 `SPI_execute` 执行游标命令（Tom）

  宏 `SPI_ERROR_CURSOR` 仍然存在但永远不会被返回。

- SPI 计划指针现在声明为 `SPIPlanPtr` 而不是 `void *`（Tom）

  这不会破坏应用程序代码，但建议切换以帮助捕获简单的编程错误。

E.96.3.20. 构建选项

- 新增 configure 选项 `--enable-profiling` 以启用代码剖析（只与 gcc 一起工作）（Korry Douglas 和 Nikhil Sontakke）
- 新增 configure 选项 `--with-system-tzdata` 以使用操作系统的时区数据库（Peter）
- 修复 PGXS，使扩展可以针对 pg_config 程序不在 `PATH` 最前面的 PostgreSQL 安装构建（Tom）
- 构建 SGML 文档时支持 `gmake draft`（Bruce）

  除非使用 `draft`，文档构建现在会在必要时重复进行以确保索引是最新的。

E.96.3.21. 源代码

- 把宏 `DLLIMPORT` 改名为 `PGDLLIMPORT`，以避免与定义 `DLLIMPORT` 的第三方包含（如 Tcl）冲突（Magnus）
- 创建“操作符族”以改进涉及跨数据类型比较的查询的规划（Tom）
- 更新 GIN `extractQuery()` API，允许发出没有任何东西能满足查询的信号（Teodor）
- 把 `NAMEDATALEN` 定义从 `postgres_ext.h` 移到 `pg_config_manual.h`（Peter）
- 在所有平台上提供 `strlcpy()` 和 `strlcat()`，并替换易出错的 `strncpy()`、`strncat()` 等用法（Peter）
- 创建钩子，让外部插件监控（甚至替换）规划器并为假想情形创建计划（Gurjeet Singh、Tom）
- 创建函数变量 `join_search_hook`，让插件覆盖规划器的连接搜索顺序部分（Julius Stroffek）
- 为 Renesas 的 M32R 处理器新增 `tas()` 支持（Kazuhiro Inaoka）
- `quote_identifier()` 和 pg_dump 不再对按语法属于非保留的关键字加引号（Tom）
- 更改 `NUMERIC` 数据类型的磁盘表示，使 `sign_dscale` 字位于权重之前（Tom）
- 在 Darwin >= 6.0（即 OS X 10.2 及以上）上使用 SYSV 信号量而不是 POSIX 信号量（Chris Marcellino）
- 新增缩写词和NFS文档节（Bruce）
- 文档现在记录 "Postgres" 是 "PostgreSQL" 的可接受别名（Peter）
- 新增关于在服务器停机时防止数据库服务器欺骗的文档（Bruce）

E.96.3.22. 附加模块

- 把 `contrib` 的 `README` 内容移入 PostgreSQL 主文档（Albert Cervera i Areny）
- 新增用于低层页检查的 `contrib/pageinspect` 模块（Simon、Heikki）
- 新增用于控制热备操作的 `contrib/pg_standby` 模块（Simon）
- 新增 `contrib/uuid-ossp` 模块，用 OSSP UUID 库生成 `UUID` 值（Peter）

  用 configure 的 `--with-ossp-uuid` 激活。它利用了新的内建 `UUID` 类型。
- 新增 `contrib/dict_int`、`contrib/dict_xsyn` 和 `contrib/test_parser` 模块，提供附加文本搜索词典模板和解析器的示例（Sergey Karpov）
- 允许 contrib/pgbench 设置填充因子（Pavan Deolasee）
- 为 contrib/pgbench 的 `-l` 添加时间戳（Greg Smith）
- 为 `contrib/pgbuffercache` 新增使用计数统计（Greg Smith）
- 为 `contrib/hstore` 新增 GIN 支持（Teodor）

- 为 `contrib/pg_trgm` 新增 GIN 支持（Guillaume Smet、Teodor）
- 更新 `contrib/start-scripts` 中的 OS/X 启动脚本（Mark Cotner、David Fetter）
- 把 `pgrowlocks()` 和 `dblink_get_pkey()` 限制为对目标表有 `SELECT` 权限的用户（Tom）
- 把 `contrib/pgstattuple` 函数限制为超级用户（Tom）
- `contrib/xml2` 已废弃，计划在 8.4 中移除（Peter）

  核心 PostgreSQL 中的新 XML 支持取代了此模块。

E.97. 版本 8.2.23

发行日期
2011-12-05

本次发布包含对 8.2.22 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

这预计是 PostgreSQL 8.2.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.97.1. 迁移到版本 8.2.23

对于运行 8.2.X 的用户，不需要进行转储/恢复。

不过，`information_schema.referential_constraints` 视图的定义中被发现存在一个长期错误。如果你依赖该视图的正确结果，应按下面第一条变更日志条目的说明替换其定义。

不过，如果你是从earlier than 8.2.14 的版本升级，请参见第 E.106 节。

E.97.2. 变更

- 修复 `information_schema.referential_constraints` 视图中的缺陷（Tom Lane）

  该视图在把外键约束与它依赖的主键或唯一键约束匹配时不够谨慎。
  这可能导致某个外键约束完全不被显示、被显示多次，
  或者声称它依赖于一个并非其实际依赖的约束。

  由于该视图定义是由 initdb 安装的，仅靠升级不会修复问题。如果需要修复现有安装，可以（以超级用户身份）删除 `information_schema` 模式，然后通过载入 *SHAREDIR*`/information_schema.sql` 重新创建它。（如果不确定 *SHAREDIR* 在哪里，运行 `pg_config --sharedir`。）这必须在每个要修复的数据库中重复执行。

- 修复 `CREATE TABLE dest AS SELECT * FROM src` 或 `INSERT INTO dest SELECT * FROM src` 期间与 TOAST 相关的数据损坏（Tom Lane）

  如果表曾被 `ALTER TABLE ADD COLUMN` 修改过，在某些极端情况下，尝试把它的数据原样复制到另一个表可能产生损坏的结果。该问题只会在 8.4 及更高版本中以这种确切形式出现，但我们也给更早的版本打了补丁，以防有其他代码路径可能触发同一缺陷。

- 修复从过期 syscache 条目访问 toast 表时的竞态条件（Tom Lane）

典型症状是诸如“missing chunk number 0 for toast value NNNNN in pg_toast_2619”的瞬时错误，其中引用的 toast 表总是属于某个系统目录。

- 改进 money 类型输入和输出的区域支持（Tom Lane）

  除了不支持所有标准 lc_monetary 格式化选项外，输入和输出函数还不一致，这意味着在某些区域设置下转储的 money 值无法重新读取。

- 不让 transform_null_equals 影响 CASE foo WHEN NULL ... 构造（Heikki Linnakangas）

  transform_null_equals 只应影响用户直接写的 foo = NULL 表达式，而不是这种形式的 CASE 在内部生成的相等性检查。

- 更改外键触发器的创建顺序，以更好地支持自引用外键（Tom Lane）

  对于引用自身表的级联外键，行更新会作为一个事件同时触发 ON UPDATE 触发器和 CHECK 触发器。ON UPDATE 触发器必须先执行，否则 CHECK 会检查行的非最终状态，并可能抛出不恰当的错误。然而这些触发器的触发顺序由其名称决定，而由于触发器使用遵循“RI_ConstraintTrigger_NNNN”约定的自动生成名称，它们通常按创建顺序排序。彻底的修复需要修改该约定（我们将在 9.2 中做），但在现有版本中更改它似乎有风险。因此此补丁只更改触发器的创建顺序。
  遇到这类错误的用户应删除并重新创建外键约束，使其触发器处于正确顺序。

- 在 psql 的命令历史中保留命令内的空行（Robert Haas）

  例如，如果从字符串字面量中删除了空行，先前的行为可能导致问题。

- 使用首选版本的 xsubpp 构建 PL/Perl，而不一定是操作系统的主副本（David Wheeler、Alex Hunsaker）

- 让 pgstatindex() 及时响应查询取消中断（Robert Haas）

- 确保 VPATH 构建正确安装所有服务器头文件（Peter Eisentraut）

- 缩短详细错误消息中报告的文件名（Peter Eisentraut）

  常规构建一直只报告包含错误消息调用的 C 文件的名称，但 VPATH 构建此前报告绝对路径名。

- 修复中美洲 Windows 时区名的解释（Tom Lane）

  把“Central America Standard Time”映射为 CST6 而不是 CST6CDT，因为中美洲各地通常不实行夏令时。

- 把时区数据文件更新为 tzdata 版本 2011n for DST law changes in Brazil, Cuba, Fiji, Palestine, Russia, and Samoa; also historical corrections for Alaska and British East Africa。

E.98. 版本 8.2.22

发行日期
2011-09-26

本次发布包含对 8.2.21 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

PostgreSQL 社区将于 2011 年 12 月停止对 8.2.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.98.1. 迁移到版本 8.2.22

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.98.2. 变更

- 修复 GiST 索引页面分裂处理中的多个缺陷（Heikki Linnakangas）

  发生概率很低，但它们可能导致索引损坏。

- 避免 `ANALYZE` 中可能访问内存末尾之外（Noah Misch）

  这修复了概率极低的服务器崩溃情形。

- 修复 relcache init 文件失效中的竞态条件（Tom Lane）

  存在一个时间窗口，新的后端进程可能读到过期的 init 文件，却错过本应告知它数据已过期的失效消息。结果是目录访问中的离奇失败，通常在启动稍后阶段报 “could not read block 0 in file ...” 。

- 修复 GiST 索引扫描结束时的内存泄漏（Tom Lane）

  执行许多独立 GiST 索引扫描的命令（例如对已包含许多行的表验证新的基于 GiST 的排他约束）会因此泄漏瞬时需要大量内存。

- 修复构造大型有损位图时的性能问题（Tom Lane）

- 修复数组和路径创建函数，确保填充字节为零（Tom Lane）

  这避免了一些规划器认为语义相等的常量不相等的情况，这种情况会导致糟糕的优化。

- 规避破坏 WAL 重放的 gcc 4.6.0 缺陷（Tom Lane）

  这可能导致服务器崩溃后已提交事务的丢失。

- 修复视图中 `VALUES` 的转储缺陷（Tom Lane）

- 禁止对序列执行 `SELECT FOR UPDATE/SHARE`（Tom Lane）

  这种操作不像预期那样工作，可能导致失败。

- 防御计算哈希表大小时的整数溢出（Tom Lane）

- 修复 “peer” 认证使用凭据控制消息时的可移植性缺陷（Tom Lane）

- 修复 `pg_srand48` 种子初始化中的笔误（Andres Freund）

  这导致未能使用所提供种子的所有位。该函数在大多数平台上不使用（只在没有 `srandom` 的平台上使用），而且无论如何看待，随机性低于预期的种子带来的潜在安全暴露似乎都很小。

- 避免 `LIMIT` 与 `OFFSET` 值之和超过 2^63 时的整数溢出（Heikki Linnakangas）
- 为 `generate_series()` 的 `int4` 和 `int8` 版本添加溢出检查（Robert Haas）
- 修复 `to_char()` 中移除尾随零的问题（Marti Raudsepp）

  在使用 `FM` 且小数点后没有数字位的格式中，小数点左侧的零可能被错误移除。
- 修复 `pg_size_pretty()`，避免接近 2^63 的输入溢出（Tom Lane）
- 修复 psql 在从其他文件执行 `COPY` 期间脚本文件行号计数的问题（Tom Lane）
- 修复 pg_restore 直接写数据库模式对 `standard_conforming_strings` 的处理（Tom Lane）

  当从以 `standard_conforming_strings` 设为 `on` 制作的归档文件直接恢复到数据库服务器时，pg_restore 可能发出不正确的命令。
- 修复 libpq 的 LDAP 服务查找代码中的越界写入和内存泄漏（Albe Laurenz）
- 在 libpq 中，避免使用非阻塞 I/O 和 SSL 连接时的失败（Martin Pihlak、Tom Lane）
- 改进 libpq 对连接启动期间失败的处理（Tom Lane）

  特别是，对 SSL 连接启动期间服务器报告 `fork()` 失败的响应现在更合理。
- 让 ecpglib 以 15 位精度写 `double` 值（Akira Kurosawa）
- 应用上游对 blowfish 有符号字符缺陷的修复（CVE-2011-2483）（Tom Lane）

  `contrib/pg_crypto` 的 blowfish 加密代码在 char 为有符号的平台（大多数平台如此）上可能给出错误结果，导致加密口令比应有的强度更弱。
- 修复 `contrib/seg` 中的内存泄漏（Heikki Linnakangas）
- 修复 `pgstatindex()`，使其对空索引给出一致结果（Tom Lane）
- 允许使用 perl 5.14 构建（Alex Hunsaker）
- 更新 configure 脚本探测系统函数是否存在的方法（Tom Lane）

  我们在 8.3 和 8.2 中使用的 autoconf 版本可能被执行链接时优化的编译器欺骗。
- 修复构建和安装文件路径包含空格的各种问题（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2011i for DST law changes in Canada, Egypt, Russia, Samoa, and South Sudan。

E.99. 版本 8.2.21

发行日期
2011-04-18

本次发布包含对 8.2.20 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.99.1. 迁移到版本 8.2.21

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.99.2. 变更

- 避免目录缓存初始化期间的潜在死锁（Nikhil Sontakke）

  某些情况下，缓存加载代码会在锁定索引的目录之前获取系统索引上的共享锁。这可能与尝试以另一种更标准的顺序获取排他锁的进程死锁。

- 修复当目标元组存在并发更新时 `BEFORE ROW UPDATE` 触发器处理中的悬空指针问题（Tom Lane）

  已观察到该缺陷在尝试执行 `UPDATE RETURNING ctid` 时导致间歇性的 “cannot extract system attribute from virtual tuple” 失败。发生更严重错误（例如为更新后的元组生成错误索引项）的概率非常小。

- 当表存在未决的延迟触发器事件时禁止 `DROP TABLE`（Tom Lane）

  此前 `DROP` 会照常执行，导致触发器最终触发时报 “could not open relation with OID nnn” 错误。

- 修复涉及数组切片的 PL/Python 内存泄漏（Daniel Popowich）

- 修复 pg_restore，使其能处理 TOC 文件中的长行（超过 1KB）（Tom Lane）

- 加入更多防护，防止因过于激进的编译器优化导致的除零崩溃（Aurelien Jarno）

- 支持在 FreeBSD 和 OpenBSD 的 MIPS 上使用 dlopen()（Tom Lane）

  此前有一个硬编码的假设：该系统函数在这些系统的 MIPS 硬件上不可用。现在改用编译期测试，因为较新的版本已经有了它。

- 修复 HP-UX 上的编译失败（Heikki Linnakangas）

- 修复 pg_regress 在 Cygwin 上使用的路径分隔符（Andrew Dunstan）

- 把时区数据文件更新为 tzdata 版本 2011f for DST law changes in Chile, Cuba, Falkland Islands, Morocco, Samoa, and Turkey; also historical corrections for South Australia, Alaska, and Hawaii。

E.100. 版本 8.2.20

发行日期
2011-01-31

本次发布包含对 8.2.19 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.100.1. 迁移到版本 8.2.20

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.100.2. 变更

- 避免 `EXPLAIN` 尝试显示简单形式 `CASE` 表达式时的失败（Tom Lane）

  如果 `CASE` 的测试表达式是常量，规划器可能把 `CASE` 简化为令表达式显示代码困惑的形式，导致 “unexpected CASE WHEN clause” 错误。

- 修复对现有下标范围之前的数组切片赋值的问题（Tom Lane）

  如果新增下标与第一个既有下标之间存在间隙，
  代码会错误计算需要从旧数组的空值位图复制多少条目，可能导致数据损坏或崩溃。

- 避免规划器对非常久远的日期值出现意外的转换溢出（Tom Lane）

  `date` 类型支持的日期范围比 `timestamp` 类型能表示的更宽，但规划器假定它总是可以不受惩罚地把 date 转换为 timestamp。

- 修复启用 `standard_conforming_strings` 时 pg_restore 对大对象（BLOB）的文本输出（Tom Lane）

  虽然直接恢复到数据库时工作正确，但如果要求 pg_restore 生成 SQL 文本输出且源数据库中已启用 `standard_conforming_strings`，字符串转义是不正确的。

- 修复对包含 `... & !(subexpression) | ...` 的 `tsquery` 值的错误解析（Tom Lane）

  包含这种操作符组合的查询没有被正确执行。同样的错误也存在于 `contrib/intarray` 的 `query_int` 类型和 `contrib/ltree` 的 `ltxtquery` 类型中。

- 修复 `contrib/intarray` 的 `query_int` 类型输入函数中的缓冲区越界（Apple）

  该缺陷构成安全风险，因为函数的返回地址可能被覆写。感谢 Apple Inc 的安全团队报告此问题并提供修复。（CVE-2010-4015）

- 修复 `contrib/seg` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `seg` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。（这与上一次更新中在 `contrib/cube` 中修复的缺陷相同。）

E.101. 版本 8.2.19

发行日期
2010-12-16

本次发布包含对 8.2.18 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.101.1. 迁移到版本 8.2.19

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.101.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复 GIN 索引 WAL 重放逻辑中的若干缺陷（Tom Lane）

  这可能导致复制期间的“bad buffer id: 0”失败或索引内容损坏。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

  给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

  在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）

- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复当 `ORDER BY`、`LIMIT`、`FOR UPDATE` 或 `WITH` 附加到 `INSERT ... VALUES` 的 `VALUES` 部分时的正确行为（Tom Lane）

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

  规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

该缺陷影响多个不同的几何距离测量操作符。

- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）
- 修复 PL/Python 对集合返回函数的处理（Jan Urbanski）

  在生成集合结果的迭代器内尝试调用 SPI 函数会失败。
- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `cube` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。
- `contrib/dblink` 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）
- 修复 `contrib/pgcrypto` 在缺少公钥时可能的核心转储（Marti Raudsepp）
- 修复 `contrib/xml2` 的 XPath 查询函数中的内存泄漏（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.102. 版本 8.2.18

发行日期
2010-10-04

本次发布包含对 8.2.17 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.102.1. 迁移到版本 8.2.18

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.102.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复 Windows 共享内存分配代码（Tsutomu Yamada、Magnus Hagander）

  此缺陷导致常被报告的“无法重新附加到共享内存”错误消息。
  这是不久前应用于较新分支的修复的向后移植。

- 在 Windows 上把退出码 128（`ERROR_WAIT_NO_CHILDREN`）视为非致命（Magnus Hagander）

  高负载下，Windows 进程有时会在启动时以此错误码失败。此前 postmaster 把这视为 panic 情况并重启整个数据库，但这似乎反应过度。

- 修复 `UNION ALL` 成员关系可能被重复扫描的问题（Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。

- 在偶有报告的 btree 失败情形中把 PANIC 降为 ERROR，并在产生的错误消息中提供更多细节（Tom Lane）

  这应能改进系统对损坏索引的健壮性。

- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）
- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）
- 修复对按引用传递的函数结果计算哈希时可能的失败（Tao Ma、Tom Lane）
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）
- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）
- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）
- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）

- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 为 `contrib/hstore` 新增 `hstore(text, text)` 函数（Robert Haas）

  此函数是现已废弃的 `=>` 操作符的推荐替代。向后移植它使得面向未来的代码可以用于较旧的服务器版本。注意该补丁只有在 `contrib/hstore` 于特定数据库中安装或重新安装之后才生效。用户也可以改为手工执行 `CREATE FUNCTION` 命令。
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。
- 让 Windows 的 “N. Central Asia Standard Time” 时区映射到 Asia/Novosibirsk 而不是 Asia/Almaty（Magnus Hagander）

  Microsoft 在 KB976098 的时区更新中改变了此区域的 DST 行为。Asia/Novosibirsk 更匹配其新行为。

E.103. 版本 8.2.17

发行日期
2010-05-17

本次发布包含对 8.2.16 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.103.1. 迁移到版本 8.2.17

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.103.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）
- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，

因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 修复在重建 relcache 条目期间收到缓存重置消息时可能的崩溃（Heikki）

  此错误是在 8.2.16 中修复相关失败时引入的。

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 防止 psql 在展开引用自身的变量时无限递归（Tom）

- 修复 psql 的 `\copy`，不在 `\copy (select ...)` 内的点周围添加空格（Tom）

  在数字字面量的小数点周围添加空格会导致语法错误。

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

- 避免 Windows 上 syslogger 进程可能的崩溃（Heikki）

- 更健壮地处理 Windows 注册表中不完整的时区信息（Magnus）

- 更新已知 Windows 时区名称集合（Magnus）

- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

  另外，把 `PKST`（巴基斯坦夏令时）加入默认时区缩写集合。

E.104. 版本 8.2.16

发行日期
2010-03-15

本次发布包含对 8.2.15 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.104.1. 迁移到版本 8.2.16

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.104.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 修复后端启动期间可能的死锁（Tom）

- 修复因未干净地处理 relcache 重载期间的错误而可能的崩溃（Tom）

- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）

- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）

- 修复 GIST 索引页分裂的恢复结束清理期间发出的不正确 WAL 数据（Yoichi Hirai）

  如果在完成未完成的 GIST 插入之后、在恢复结束清理期间崩溃，这会导致索引损坏，或者更可能在 WAL 重放期间出错。

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）

- 修复若干病态缓慢的正则表达式匹配情形（Tom）

- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）

- 修复更多临时文件泄漏的情形（Heikki）

  这纠正了上一个小版本中引入的问题。失败的一种情形是返回集合的 plpgsql 函数在另一个函数的异常处理器中被调用。

- 改进布尔变量情形的约束排除处理，特别是使排除带有“bool_column = false”约束的分区成为可能（Tom）

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）

- 修复 `SSL_read` 或 `SSL_write` 失败而未设置 `errno` 时可能的无限循环（Tom）

  据报道这在某些 Windows 版本的 openssl 上可能发生。

- 修复 psql 的 `numericlocale` 选项，使其不在 latex 和 troff 输出格式中格式化不应格式化的字符串（Heikki）

- 当同时指定 `ON_ERROR_STOP` 和 `--single-transaction` 且隐含的 `COMMIT` 期间发生错误时，让 psql 返回正确的退出状态（3）（Bruce）

- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）

- 修复从 PL/PerlU 调用 PL/Perl 函数或相反时可能的失败（Tim Bunce）

- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）

- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。

- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）

- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）

- 让 `contrib/xml2` 在 Windows 上的构建更健壮（Andrew）

- 修复 Windows 信号处理中的竞态条件（Radu Ilie）

  此缺陷的一个已知症状是 `pg_listener` 中的行可能在重负载下被丢弃。

- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.105. 版本 8.2.15

发行日期
2009-12-14

本次发布包含对 8.2.14 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.105.1. 迁移到版本 8.2.15

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.14 的版本升级，请参阅第 E.106 节。

E.105.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。

- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。

- 修复后端启动时缓存初始化期间可能的崩溃（Tom）

- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。

- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。

- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）

- 确保预备事务所持有的共享元组级锁不被忽略（Heikki）

- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）

- 修复 GiST 索引页分裂的不正确逻辑：当分裂依赖于索引的非首列时（Paul Ramsey）

- 如果回收或删除旧 WAL 文件在检查点结束时失败，不要报错退出（Heikki）

  更好的做法是把该问题视为非致命并允许检查点完成。未来的检查点会重试删除。
  这类问题在正常运行中不应出现，但已知可能由设计不当的 Windows 反病毒和备份软件引起。

- 确保 WAL 文件不会在 Windows 上被重复归档（Heikki）

  这是其他进程干扰删除不再需要的文件时可能出现的另一种症状。

- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 修复 `CREATE OR REPLACE FUNCTION` 期间所有权依赖的处理（Tom）

- 修复从 `plperl` 调用 `plperlu` 或相反时的缺陷（Tom）

  内层函数的错误退出可能因未能为外层函数重新选择正确的 Perl 解释器而导致崩溃。

- 修复重定义 PL/Perl 函数时的会话期内存泄漏（Tom）

- 确保 Perl 数组在由返回集合的 PL/Perl 函数返回时被正确转换为 PostgreSQL 数组（Andrew Dunstan、Abhijit Menon-Sen）

  对非返回集合的函数这早已正确工作。

- 修复 PL/Python 异常处理中的罕见崩溃（Peter）

- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。

- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）

- 更新时区缩写文件以符合当前现实（Joachim Wieland）

  这包括把 `IDT` 和 `SGT` 加入默认时区缩写集合。

- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.106. 版本 8.2.14

发行日期
2009-09-09

本次发布包含对 8.2.13 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.106.1. 迁移到版本 8.2.14

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.2.14 后对它们执行 `REINDEX`。另外，如果你是从早于 8.2.11 的版本升级，请参阅第 E.109 节。

E.106.2. 变更

- 在 `pg_start_backup()` 期间强制 WAL 段切换（Heikki）

  这避免了可能使基础备份不可用的极端情形。

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 把对已装载的可装载模块的 `LOAD` 变为无操作（Tom）

  此前 `LOAD` 会试图卸载并重新装载模块，但这不安全也不太有用。

- 不允许在 LDAP 认证期间使用空口令（Magnus）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复与从 Sort 或 Materialize 计划节点的输出获取整行值相关的缺陷（Tom）

- 回退规划器中当 AND 或 OR 列表中的子句超过 100 个时禁用部分索引和约束排除优化的变更（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）
- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。
- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）
- 正确地对像 `00:12:57.999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）
- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）
- 避免当输入值（几乎）按排序顺序时批量插入 GIN 索引的性能退化（Tom）
- 在 PL/pgSQL 的某些上下文中正确强制执行 NOT NULL 域约束（Tom）
- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）
- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）
- 让 `contrib/hstore` 在键或值太长而无法放入其数据结构时抛出错误，而不是静默截断（Andrew Gierth）
- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）
- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）
- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）
- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.107. 版本 8.2.13

发行日期
2009-03-16

本次发布包含对 8.2.12 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.107.1. 迁移到版本 8.2.13

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.11 的版本升级，请参阅第 E.109 节。

E.107.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

此变更扩展了上两个小版本中针对相关失败情形所做的修复。
以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的任何错误都可能在试图报告该错误时导致无限递归。因此，
解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）

- 修复 C 区域设置与多字节编码一起使用时 `contrib/tsearch2` 可能失败的问题（Teodor）

  在 `wchar_t` 比 `int` 窄的平台（特别是 Windows）上可能崩溃。

- 修复 `contrib/tsearch2` 解析器处理包含多个 `@` 字符的类电子邮件字符串时的极端低效（Heikki）

- 修复带隐式强制转换的 `CASE WHEN` 的反编译（Tom）

  在试图检查或转储视图时，这一错误在启用断言的构建中可能导致断言失败，
  在其他情况下可能导致“意外的 CASE WHEN 子句”错误消息。

- 修复 TOAST 表行类型所有者可能的错误分配（Tom）

  如果 `CLUSTER` 或重写形式的 `ALTER TABLE` 由表所有者之外的人执行，表的 TOAST 表的 `pg_type` 条目最终会被标记为那个人所拥有。这不会立即引起问题，因为普通数据库操作不检查 TOAST 行类型的权限。但是，如果后来试图删除发出该命令的角色（在 8.1 或 8.2 中），或者在那之后 pg_dump 发出“数据类型的所有者似乎无效”警告（在 8.3 中），则可能导致意外失败。

- 修复 PL/pgSQL：不再把字符串中任何位置的 `INSERT` 之后的 `INTO` 当作 INTO-变量子句，而不仅是开头；特别是不要对 `CREATE RULE` 内的 `INSERT INTO` 失败（Tom）

- 在块退出时完全清理 PL/pgSQL 错误状态变量（Ashesh Vashi 和 Dave Page）

  这对 PL/pgSQL 本身不是问题，但这一疏漏可能导致 PL/pgSQL 调试器在检查函数状态时崩溃。

- 在 Windows 上重试失败的 `CallNamedPipe()` 调用（Steve Marshall、Magnus）

  该函数有时似乎会瞬时失败；我们此前把任何失败都当作硬错误，这可能扰乱 `LISTEN/NOTIFY` 以及其他操作。

- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.108. 版本 8.2.12

发行日期
2009-02-02

本次发布包含对 8.2.11 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.108.1. 迁移到版本 8.2.12

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.11 的版本升级，请参阅第 E.109 节。

E.108.2. 变更

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 修复 PL/pgSQL 中执行的语句被重写为另一种语句时可能的断言失败，例如 `INSERT` 被重写为 `UPDATE`（Heikki）
- 确保数据类型输入函数有可用快照（Tom）

  这主要影响声明了涉及用户定义稳定或不可变函数的 `CHECK` 约束的域。如果没有设置快照，这类函数通常会失败。
- 让使用 SPI 的函数更安全地用于数据类型 I/O 中；特别是用于域检查约束（Tom）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复使 `UPDATE RETURNING tableoid` 返回零而不是正确 OID 的问题（Tom）
- 修复传递等式应用于外连接子句时规划器对选择率的错误估计（Tom）

  这可能导致诸如 `... from a left join b on a.a1 = b.b1 where a.a1 = 42 ...` 之类的查询得到糟糕的计划。
- 改进优化器对长 `IN` 列表的处理（Tom）

  此变更避免了在启用约束排除时在这类列表上浪费大量时间。
- 确保可保持游标的内容不依赖于 TOAST 表的内容（Tom）

  此前，游标结果中的大字段值可能被表示为 TOAST 指针，如果被引用的表在游标被读取之前被删除，或者大值被删除随后被清理掉，这就会失败。普通游标不会发生这种情况，但对于保持到创建事务之后的游标则可能。
- 修复返回集合的函数在未读完整个结果就被终止时的内存泄漏（Tom）
- 修复 `contrib/dblink` 的 `dblink_get_result(text,bool)` 函数（Joe）
- 修复 `contrib/sslinfo` 函数可能的垃圾输出（Tom）
- 修复 configure 脚本，使其在无法获得 PL/Perl 的链接信息时正确报告失败（Andrew）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.109. 版本 8.2.11

发行日期
2008-11-03

本次发布包含对 8.2.10 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.109.1. 迁移到版本 8.2.11

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.7 的版本升级，请参阅第 E.113 节。另外，如果你运行的是之前的 8.2.X 版本，建议在升级后对所有 GiST 索引执行 `REINDEX`。

E.109.2. 变更

- 修复 GiST 索引因删除后把错误的索引条目标记为“死”而损坏的问题（Teodor）

  这会导致索引搜索找不到本应找到的行。损坏的索引可以用 `REINDEX` 修复。

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）

- 改进 *expression* `IN` (*expression-list*) 查询的优化（Tom，想法来自 Robert Haas）

  右侧有查询变量的情形在 8.2.x 和 8.3.x 中的处理效率低于以前的版本。此修复为此类情形恢复了 8.1 的行为。

- 修复当子 `SELECT` 出现在 `FROM` 中的函数调用、多行 `VALUES` 列表或 `RETURNING` 列表中时规则查询的错误展开（Tom）

  此问题通常的症状是“无法识别的节点类型”错误。

- 修复哈希聚合计划重扫描期间的内存泄漏（Neil）

- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）

- 防止用 `ALTER SET TABLESPACE` 把表移到另一个表空间时 `relfilenode` 号可能的冲突（Heikki）

  该命令试图重用现有文件名，而不是选取目标目录中已知未使用的文件名。

- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）

- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）

- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
唯一可能的后果是错误的错误指示。

- 修复 ecpg 对 `CREATE ROLE` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 确保 `pg_control` 以二进制模式打开（Itagaki Takahiro）

  pg_controldata 和 pg_resetxlog 此前做得不正确，因此在 Windows 上可能失败。
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.110. 版本 8.2.10

发行日期
2008-09-22

本次发布包含对 8.2.9 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.110.1. 迁移到版本 8.2.10

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.7 的版本升级，请参阅第 E.113 节。

E.110.2. 变更

- 修复 btree WAL 恢复代码中的缺陷（Heikki）

  如果 WAL 在页分裂操作中途结束，恢复就会失败。
- 修复 `datfrozenxid` 可能的错误计算（Alvaro）

  此错误或许能解释最近一些无法删除旧 `pg_clog` 数据的报告。
- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

  这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。
- 修复 GiST 索引扫描期间可能的元组重复输出（Teodor）
- 修复视图包含简单 `UNION ALL` 构造时遗漏的权限检查（Heikki）

  被引用表的权限被正确检查，但视图本身的权限没有。
- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  `ALTER COLUMN TYPE` 之后重用先前缓存的计划可能产生这种情形。该检查防范随之而来的数据损坏和/或崩溃。

- 修复 `DROP OWNED` 期间可能的重复删除（Tom）

  这通常会导致奇怪的错误，例如“关系 NNN 的缓存查找失败”。

- 修复 `AT TIME ZONE`：先尝试把时区参数解释为时区缩写，仅在失败时才尝试完整时区名，而不是像以前那样相反（Tom）

  时间戳输入函数一直按此顺序解析有歧义的时区名。让 `AT TIME ZONE` 也这样做改进了一致性，并修复了 8.1 中引入的兼容性缺陷：在有歧义的情形下，我们现在的行为与 8.0 及更早版本相同，因为在旧版本中 `AT TIME ZONE` 只接受缩写。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）

- 防止显示带单位的配置参数时单位转换期间的整数溢出（Tom）

- 改进向 syslog 写很长日志消息的性能（Tom）

- 允许 `pg_hba.conf` 中 LDAP URL 的后缀部分包含空格（Tom）

- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）

- 修复嵌套子查询表达式的规划器缺陷（Tom）

  如果外层子查询不直接依赖父查询而内层依赖，外层的值可能不会为新的父查询行重新计算。

- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/pgSQL 在 `FOR` 循环的目标变量是包含组合类型字段的记录时不失败（Tom）

- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）

- 在 Windows 上，通过阻止 libpq 在每次系统调用中发送超过 64kB 来绕过 Microsoft 的一个缺陷（Magnus）

- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）

- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）

- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.111. 版本 8.2.9

发行日期
2008-06-12

此发行版在 8.2.8 的基础上包含一个严重和一个次要的修复。有关 8.2 主版本新特性的信息，请参阅第 E.120 节。

E.111.1. 迁移到版本 8.2.9

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.7 的版本升级，请参阅第 E.113 节。

E.111.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

- 让 `ALTER AGGREGATE ... OWNER TO` 更新 `pg_shdepend`（Tom）

  如果聚合后来涉及 `DROP OWNED` 或 `REASSIGN OWNED` 操作，这一疏漏可能导致问题。

E.112. 版本 8.2.8

发行日期
从未发布

本次发布包含对 8.2.7 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.112.1. 迁移到版本 8.2.8

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.2.7 的版本升级，请参阅第 E.113 节。

E.112.2. 变更

- 修复在 Windows 上使用 UTF-8 数据库编码和不同客户端编码时发生的 `ERRORDATA_STACK_SIZE exceeded` 崩溃（Tom）

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）

- 修复 `pg_get_ruledef()`，显示 `UPDATE` 或 `DELETE` 目标表上附加的别名（如果有）（Tom）

- 修复可能导致 `too many LWLocks taken` 失败的 GIN 缺陷（Teodor）

- 避免解压损坏数据时可能的崩溃（Zdenek Kotala）

- 修复两处后端 SIGTERM 退出可能在共享内存中留下损坏状态的地方（Tom）

  如果 SIGTERM 用于一起关闭整个数据库集簇，这两种情形都不太重要，但如果有人试图对单个后端发 SIGTERM，就有问题。

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数（特别是 `array_in()`），它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。

- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。

- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴和阿根廷/圣路易斯的 DST 法规变更）

- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）

- 修复 `contrib/tsearch2` 的 `tsquery` 类型损坏的 GiST 比较函数（Teodor）

- 修复 `contrib/cube` 函数可能的崩溃（Tom）

- 修复 `contrib/xml2` 的 `xpath_table()` 函数在输入查询返回 NULL 值时的核心转储（Tom）

- 修复 `contrib/xml2` 的 makefile，使其不覆盖 `CFLAGS`（Tom）

- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

E.113. 版本 8.2.7

发行日期
2008-03-17

本次发布包含对 8.2.6 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.113.1. 迁移到版本 8.2.7

对于运行 8.2.X 的用户，不需要进行转储/恢复。不过，如果你受到下述 Windows 区域设置问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.113.2. 变更

- 修复把不同字符组合视为相等的 Windows 区域设置下的字符串比较（Tom）

  此修复仅在 Windows 上且仅在使用 UTF-8 数据库编码时适用。所有其他情形的相同修复两年多前就已完成，但使用 UTF-8 的 Windows 走的是未更新的单独代码路径。如果你使用的区域设置把某些不同的字符串视为相等，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 修复不同系统目录上并发 `VACUUM FULL` 操作之间的潜在死锁（Tom）

- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 不允许在预备事务内使用 `LISTEN` 和 `UNLISTEN`（Tom）

  这以前是允许的，但尝试这样做会有各种不愉快的后果，特别是只要 `UNLISTEN` 未提交，发起的后端就无法退出。

- 不允许在预备事务内删除临时表（Heikki）

  8.1 正确地禁止了这一点，但该检查在 8.2 中被无意破坏。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）

- 修复返回集合函数某些用法中的内存泄漏（Neil）

- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。

- 修复 `ALTER OWNER` 某些变体中的“无法识别的节点类型”错误（Tom）

- 确保锁等待中止时清除 `pg_stat_activity.waiting` 标志（Tom）

- 修复 Windows Vista 上进程权限的处理（Dave、Magnus）

  特别是，此修复允许以 Administrator 用户启动服务器。

- 更新时区数据文件到 tzdata 发行版 2008a（特别是智利最近的变更）；调整时区缩写 `VET`（委内瑞拉）为 UTC-4:30 而不是 UTC-4:00（Tom）

- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 `pg_ctl start -w` 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。

- 使用 `-fwrapv` 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。

- 正确强制执行超过 `INT_MAX` 微秒（约 35 分钟）的 `statement_timeout` 值（Tom）

此缺陷只影响使用 `--enable-integer-datetimes` 的构建。

- 修复常量折叠简化子查询时“意外的 PARAM_SUBLINK ID”规划器错误（Tom）
- 修复 `IS NULL` 和 `NOT` 表达式的约束排除处理中的逻辑错误（Tom）

  规划器有时会因 NULL 结果的可能性而排除本不应排除的分区。
- 修复“无法构建任何 N 路连接”规划器错误的另一个原因（Tom）

  这可能发生在无子句连接需要在连接子句可被利用之前被强制的情形。
- 修复外连接规划中不正确的常量传播（Tom）

  规划器有时可能错误地断定变量可被约束为等于常量，导致错误的查询结果。
- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

  显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。
- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.114. 版本 8.2.6

发行日期
2008-01-07

此发行版包含 8.2.5 之后的多种修复，包括对重要安全问题的修复。有关 8.2 主版本新特性的信息，请参阅第 E.120 节。

E.114.1. 迁移到版本 8.2.6

对于运行 8.2.X 的用户，不需要进行转储/恢复。

E.114.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  8.2.5 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 修复 GIN 索引 WAL 重放中的缺陷（Teodor）
- 修复 `maintenance_work_mem` 为 4GB 或更多时 GIN 索引构建无法正确工作的问题（Tom）
- 更新时区数据文件到 tzdata 发行版 2007k（特别是阿根廷最近的变更）（Tom）
- 改进规划器在非 C 区域设置下对 LIKE/正则表达式估计的处理（Tom）
- 修复深层外连接嵌套的规划速度问题以及可能糟糕的连接顺序选择（Tom）
- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）
- 让 `CREATE TABLE ... SERIAL` 和 `ALTER SEQUENCE ... OWNED BY` 不更改序列的 `currval()` 状态（Tom）
- 保留被 `ALTER TABLE ... ALTER COLUMN TYPE` 重建的索引的表空间和存储参数（Tom）
- 让归档恢复总是开始新的 WAL 时间线，而不是只在使用恢复停止时间时（Simon）

  这避免了试图覆盖最后一个 WAL 段现有归档副本的极端情形风险，而且看起来比原来的定义更简单干净。

- 当表太小而不适用时，让 `VACUUM` 不使用全部 `maintenance_work_mem`（Alvaro）
- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）
- 让 `corr()` 对负相关值返回正确结果（Neil）
- 修复超过 68 年的 interval 在 `extract(epoch from interval)` 中的溢出（Tom）
- 修复 PL/Perl 在可信函数中使用 UTF-8 正则表达式时不失败（Andrew）
- 修复 PL/Perl 在平台 Perl 把 `bool` 类型定义为 `int` 而非 `char` 时的问题（Tom）

  虽然理论上这可能在任何地方发生，但没有标准 Perl 构建这样做.....直到 Mac OS X 10.5。

- 修复 PL/Python 在 64 位机器上与 Python 2.5 正确工作（Marko Kreen）
- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）
- 修复 pg_dump，使其正确处理默认表达式与父表不同的继承子表（Tom）
- 修复 `PGPASSFILE` 指向非普通文件时 libpq 的崩溃（Martin Pitt）
- ecpg 解析器修复（Michael）
- 让 `contrib/pgcrypto` 防御在超过 128 位密钥上失败的 OpenSSL 库——至少某些 Solaris 版本是这种情形（Marko Kreen）

- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）
- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。
- 更新 `gettimeofday` 配置检查，使 PostgreSQL 能在更新版本的 MinGW 上构建（Magnus）

E.115. 版本 8.2.5

发行日期
2007-09-17

本次发布包含对 8.2.4 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.115.1. 迁移到版本 8.2.5

对于运行 8.2.X 的用户，不需要进行转储/恢复。

E.115.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 修复涉及域上之域情形的 `ALTER DOMAIN ADD CONSTRAINT`（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复外连接的一些规划器问题，特别是 `t1 LEFT JOIN t2 WHERE t2.col IS NULL` 的糟糕大小估计（Tom）
- 允许 `interval` 数据类型接受仅由毫秒或微秒组成的输入（Neil）
- 允许 `timestamp` 输入中时区名出现在年份之前（Tom）
- `/contrib/tsearch2` 所用 GIN 索引的修复（Teodor）
- 加快 rtree 索引插入（Teodor）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复日志，使使用 syslogger 进程时日志消息永不交错（Andrew）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 修复某些外键极端情形的错误处理（Tom）

- 修复 stddev_pop(numeric) 和 var_pop(numeric)（Tom）
- 防止 REINDEX 和 CLUSTER 因试图处理其他会话的临时表而失败（Alvaro）
- 更新时区数据库规则，特别是新西兰即将到来的变更（Tom）
- Windows 套接字和信号量改进（Magnus）
- 让 pg_ctl -w 在 Windows 服务模式下正确工作（Dave Page）
- 修复在 Windows 上使用 MIT Kerberos 时的内存分配缺陷（Magnus）
- 因可能的编码不匹配，在 Windows 上抑制日志时间戳中的时区名（%Z）（Tom）
- 作为安全措施，要求使用 /contrib/dblink 的非超级用户只使用口令认证（Joe）
- 出于安全原因，把 /contrib/pgstattuple 函数限制为超级用户（Tom）
- 不让 /contrib/intarray 试图使其 GIN 操作符类成为默认（这会在转储/恢复时引起问题）（Tom）

E.116. 版本 8.2.4

发行日期
2007-04-23

此发行版包含 8.2.3 之后的多种修复，包括一个安全修复。有关 8.2 主版本新特性的信息，请参阅第 E.120 节。

E.116.1. 迁移到版本 8.2.4

对于运行 8.2.X 的用户，不需要进行转储/恢复。

E.116.2. 变更

- 支持在 search_path 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 search_path 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 CREATE FUNCTION。

- 通过强制在每个后端中重载来修复 Windows 的 shared_preload_libraries（Korry Douglas）
- 修复 to_char()，使其正确地对本地化的日或月名做大/小写转换（Pavel Stehule）
- /contrib/tsearch2 崩溃修复（Teodor）
- 要求 COMMIT PREPARED 在事务预备时所在的同一数据库中执行（Heikki）
- 允许 pg_dump 在 Windows 上做超过 2GB 的二进制备份（Magnus）

- 新的繁体（台湾）中文 FAQ（Zhou Daojing）
- 防止统计收集器过于频繁地写磁盘（Tom）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 修复使用数组类型的域中的缺陷（Tom）
- 修复 `pg_dump`，使其在不转储所属表时也能用 `-t` 转储 serial 列的序列（Tom）
- 规划器修复，包括改进外连接和位图扫描选择逻辑（Tom）
- 修复 PL/pgSQL 函数试图在 `EXCEPTION` 块内 `RETURN` 时可能的错误答案或崩溃（Tom）
- 修复哈希索引扩容期间的 PANIC（Tom）
- 修复 POSIX 风格时区规范以遵循新的美国 DST 规则（Tom）

E.117. 版本 8.2.3

发行日期
2007-02-07

此发行版包含 8.2.2 之后的两个修复。有关 8.2 主版本新特性的信息，请参阅第 E.120 节。

E.117.1. 迁移到版本 8.2.3

对于运行 8.2.X 的用户，不需要进行转储/恢复。

E.117.2. 变更

- 移除约束和函数索引中对类型长度的过度严格检查（Tom）
- 修复优化，使子查询中的 MIN/MAX 再次能使用索引（Tom）

E.118. 版本 8.2.2

发行日期
2007-02-05

此发行版包含 8.2.1 之后的多种修复，包括一个安全修复。有关 8.2 主版本新特性的信息，请参阅第 E.120 节。

E.118.1. 迁移到版本 8.2.2

对于运行 8.2.X 的用户，不需要进行转储/恢复。

E.118.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  这些漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，以及更改表列的数据类型（CVE-2007-0555、CVE-2007-0556）。这些错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复不再那么罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 修复 Borland C 编译脚本（L Bayuk）

- 正确处理以 `00` 结尾年份的 `to_char('CC')`（Tom）

  2000 年属于二十世纪，不是二十一世纪。

- `/contrib/tsearch2` 本地化改进（Tatsuo、Teodor）

- 修复 `information_schema.key_column_usage` 视图中不正确的权限检查（Tom）

  症状是“OID 为 nnnnn 的关系不存在”错误。要不经 `initdb` 获得此修复，用 `CREATE OR REPLACE VIEW` 安装 `share/information_schema.sql` 中的修正定义。注意需要在每个数据库中这样做。

- 改进有大量表的数据库的 `VACUUM` 性能（Tom）

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 修复使用 `ROW` 不等条件的索引搜索可能不正确的结果（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

- 修复 Windows 上因试图 fsync 已删除文件而发生的虚假“权限被拒绝”失败（Magnus、Tom）

- 修复可能导致统计收集器在 Windows 上挂起的缺陷（Magnus）

  这进而导致 autovacuum 不工作。

- 修复已在使用中的 PL/pgSQL 函数被更新时可能的崩溃（Tom）

- 改进 PL/pgSQL 对域类型的处理（Sergiy Vyshnevetskiy、Tom）

- 修复 PL/pgSQL 异常块处理中可能的错误（Tom）

E.119. 版本 8.2.1

发行日期
2007-01-08

本次发布包含对 8.2 的多方面修复。有关 8.2 大版本中新特性的信息，请参见第 E.120 节。.

E.119.1. 迁移到版本 8.2.1

对于运行 8.2 的用户，不需要进行转储/恢复。

E.119.2. 变更

- 修复 `SELECT ... LIMIT ALL`（以及 `LIMIT NULL`）的崩溃（Tom）
- 若干 `/contrib/tsearch2` 修复（Teodor）
- 在 Windows 上，使来自操作系统的日志消息使用 ASCII 编码（Hiroshi Saito）

  这修复了操作系统与数据库服务器编码不匹配时的转换问题。
- 修复用 `win32.mak` 在 Windows 上链接 pg_dump 的问题（Hiroshi Saito）
- 修复外连接查询的规划器错误（Tom）
- 修复涉及子 SELECT 的查询中的若干问题（Tom）
- 修复子事务中止期间 SPI 中可能的崩溃（Tom）

  这影响所有 PL 函数，因为它们都使用 SPI。
- 改进 PDF 文档的构建速度（Peter）
- 重新加入 JST（日本）时区缩写（Tom）
- 改进与索引扫描相关的优化决策（Tom）
- 让 psql 像以前一样打印多字节组合字符，而不是输出为 `\u`（Tom）
- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。
- 让 pg_dumpall 在从 8.2 之前的服务器转储时假定数据库有公共 `CONNECT` 权限（Tom）

  这保留了以前的行为：只要 `pg_hba.conf` 允许，任何人都可以连接到数据库。

E.120. 版本 8.2

发行日期
2006-12-05

E.120.1. 概述

此发行版新增了许多用户要求的功能和性能改进，包括：

- 查询语言增强，包括 `INSERT/UPDATE/DELETE RETURNING`、多行 `VALUES` 列表以及 `UPDATE/DELETE` 中可选的目标表别名

- 创建索引而不阻塞并发的 INSERT/UPDATE/DELETE 操作
- 许多查询优化改进，包括对外连接重排序的支持
- 改进的排序性能和更低的内存使用
- 更高效的锁定和更好的并发性
- 更高效的清理
- 更易管理的热备服务器
- 新的表和索引 FILLFACTOR 支持
- 监控、日志和性能调优的新增内容
- 对创建和删除对象的更多控制
- 可以为已有表定义和移除表继承关系
- COPY TO 可以复制任意 SELECT 语句的输出
- 数组改进，包括数组中的空值
- 聚合函数改进，包括多输入聚合和 SQL:2003 统计函数
- 许多 contrib/ 改进

E.120.2. 迁移到版本 8.2

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

- 把 escape_string_warning 默认设为 on（Bruce）

  如果在非转义（非 E''）字符串中使用反斜杠转义，这会发出警告。

- 更改行构造器语法（ROW(...)），使列表元素 foo.* 展开为其成员字段的列表，而不是像以前那样创建嵌套的行类型字段（Tom）

  新行为有用得多，因为它允许例如触发器用 IF row(new.*) IS DISTINCT FROM row(old.*) 检查数据更改。省略 .* 仍可获得旧行为。

- 让按行比较遵循 SQL 标准语义，并允许它们用于索引扫描（Tom）

  此前，行 = 和 <> 比较遵循标准，但 < <= > >= 不遵循。现在行比较可以用作匹配行值的多列索引的索引约束。

- 让行 IS [NOT] NULL 测试遵循 SQL 标准语义（Tom）

  以前的行为在 IS NULL 的简单情形下符合标准，但 IS NOT NULL 在任一行字段非空时就返回真，而标准说只有在所有字段都非空时才应返回真。

- 让 SET CONSTRAINT 只影响一个约束（Kris Jurka）

在以前的版本中，SET CONSTRAINT 修改所有名称匹配的约束。在此版本中，模式搜索路径用于只修改第一个匹配的约束。也支持模式指定。这更接近 SQL 标准。

- 出于安全原因移除表的 RULE 权限（Tom）

  从此版本起，只有表的所有者能为表创建或修改规则。为了向后兼容，GRANT/REVOKE RULE 仍被接受，但它不做任何事情。

- 数组比较改进（Tom）

  现在数组维度也被比较。

- 更改数组连接以符合文档行为（Tom）

  这改变了以前连接会修改数组下界的行为。

- 使 postmaster 和postgres的命令行选项一致（Peter）

  这允许 postmaster 不用 -o 就能向每个后端传递参数。注意某些选项现在只有长形式，因为存在冲突的单字母选项。

- 废弃 postmaster 符号链接的使用（Peter）

  postmaster 和 postgres 命令现在行为完全相同，由命令行选项决定行为。postmaster 符号链接为兼容性保留，但已不再真正需要。

- 更改 log_duration，即使查询未被输出也输出（Tom）

  在以前的版本中，log_duration 只在查询先前出现在日志中时才打印。

- 让 to_char(time) 和 to_char(interval) 把 HH 和 HH12 当作 12 小时间隔

  大多数应用程序应使用 HH24，除非想要 12 小时显示。

- 在从 INET 到 CIDR 的转换中将未掩码的位清零（Tom）

  这确保转换后的值对 CIDR 确实有效。

- 移除 australian_timezones 配置变量（Joachim Wieland）

  此变量已被配置时区缩写的更通用机制取代。

- 改进嵌套循环索引扫描的代价估计（Tom）

  这可能消除设置不切实际的小 random_page_cost 值的需要。如果你一直使用很小的 random_page_cost，请重新检查你的测试用例。

- 更改 pg_dump 的 -n 和 -t 选项的行为。（Greg Sabino Mullane）

  详情请参阅 pg_dump 手册页。

- 更改libpq的 PQdsplen() 以返回有用的值（Martijn van Oosterhout）

- 把libpq的 PQgetssl() 声明为返回 void * 而不是 SSL *（Martijn van Oosterhout）

  这允许应用程序使用该函数而不包含 OpenSSL 头文件。

- C 语言可装载模块现在必须包含 `PG_MODULE_MAGIC` 宏调用以进行版本兼容性检查（Martijn van Oosterhout）
- 出于安全考虑，PL/PerlU 函数使用的模块不再对 PL/Perl 函数可用（Andrew）

> **注意**
>
> 这也意味着数据不能再在 PL/Perl 函数与 PL/PerlU 函数之间共享。某些 Perl 安装编译时未带允许单个进程中存在多个解释器的正确标志。在这种情形下，PL/Perl 和 PL/PerlU 不能在单个后端中同时使用。解决办法是获得支持多解释器的 Perl 安装。

- 在 `contrib/xml2/` 中，把 `xml_valid()` 改名为 `xml_is_well_formed()`（Tom）

  `xml_valid()` 将为向后兼容保留，但其行为将在未来版本中改为做模式检查。
- 移除 `contrib/ora2pg/`，现位于 http://www.samse.fr/GPL/ora2pg
- 移除已迁移到 PgFoundry 的 contrib 模块：`adddepend`、`dbase`、`dbmirror`、`fulltextindex`、`mac`、`userlock`
- 移除被废弃的 contrib 模块：`mSQL-interface`、`tips`
- 移除 QNX 和 BEOS 移植（Bruce）

  这些移植不再有活跃的维护者。

E.120.3. 变更

下面你将找到 PostgreSQL 8.2 与上一个主版本之间变更的详细说明。

E.120.3.1. 性能改进

- 允许规划器在某些情形下重排外连接（Tom）

  在以前的版本中，外连接总是按照查询中书写的顺序求值。
  此变更允许查询优化器在能确定连接顺序可以在不改变查询含义的情况下更改时，
  考虑重排外连接。这对涉及多个外连接或内外连接混合的查询可带来可观的性能差异。
- 改进 `IN`（表达式列表）子句的效率（Tom）
- 改进排序速度并降低内存使用（Simon、Tom）
- 改进子事务性能（Alvaro、Itagaki Takahiro、Tom）
- 为表和索引创建新增 `FILLFACTOR`（ITAGAKI Takahiro）

  这在每个表或索引页中留出额外空闲空间，使数据库增长时性能得以改进。
  这对维护聚簇特别有价值。
- 提高 `shared_buffers` 和 `max_fsm_pages` 的默认值（Andrew）
- 通过把锁管理器表分成若干节来改进锁定性能（Tom）

  这使锁定更细粒度，减少争用。

- 降低顺序扫描的锁定要求（Qingqing Zhou）
- 降低创建和销毁数据库所需的锁定（Tom）
- 改进优化器对 `LIKE`、`ILIKE` 和正则表达式操作的选择率估计（Tom）
- 改进到继承表和 `UNION ALL` 视图连接的规划（Tom）
- 允许把约束排除应用于继承的 `UPDATE` 和 `DELETE` 查询（Tom）

  `SELECT` 已经遵循约束排除。
- 改进常量 `WHERE` 子句的规划，例如只依赖于从外层查询级别继承的变量的条件（Tom）
- 协议级未命名预备语句对每组 `BIND` 值重新规划（Tom）

  这改进了性能，因为可以在计划中使用确切的参数值。
- 加快 B-Tree 索引的清理（Heikki Linnakangas、Tom）
- 避免 `VACUUM` 期间对无索引表的额外扫描（Greg Stark）
- 改进多列 GiST 索引（Oleg、Teodor）
- 在 B-Tree 页分裂前移除死索引条目（Junji Teramoto）

E.120.3.2. 服务器变更

- 允许强制切换到新事务日志文件（Simon、Tom）

  这对保持热备从属服务器与主服务器同步很有价值。事务日志文件切换现在也会在 `pg_stop_backup()` 期间自动发生。这确保恢复所需的全部事务日志文件可以立即归档。
- 新增 WAL 信息函数（Simon）

  新增查询当前事务日志插入点以及从 `pg_stop_backup()` 和相关函数显示的十六进制 WAL 位置确定 WAL 文件名的函数。
- 改进 WAL 重放期间崩溃后的恢复（Simon）

  服务器现在在 WAL 恢复期间做周期性检查点，因此如果发生崩溃，未来的 WAL 恢复会缩短。这还消除了热备服务器崩溃后重放自基础备份以来整个日志的需要。
- 改进长期 WAL 重放的可靠性（Heikki、Simon、Tom）

  此前，由于 XID 回卷，试图前滚超过 20 亿个事务行不通。这意味着热备服务器必须定期从新的基础备份重载。
- 新增 `archive_timeout`，按给定间隔强制事务日志文件切换（Simon）

  这为热备服务器强制了最大复制延迟。
- 新增本机 LDAP 认证（Magnus Hagander）

  这对不支持 PAM 的平台（如 Windows）特别有用。
- 新增 `GRANT CONNECT ON DATABASE`（Gevik Babakhani）

这提供了对数据库访问的 SQL 级控制。它作为现有 pg_hba.conf 控制之上的附加过滤器工作。

- 新增对 SSL 证书吊销列表（CRL）文件的支持（Libor Hohoš）

  服务器和 libpq 现在都识别 CRL 文件。

- GiST 索引现在可聚簇（Teodor）

- 移除 autovacuum 例行的服务器日志条目（Bruce）

  pg_stat_activity 现在显示 autovacuum 活动。

- 在单个表内而不是整个数据库内跟踪最大 XID 年龄（Alvaro）

  这通过避免不必要的 VACUUM 降低了防止事务 ID 回卷所涉及的开销。

- 为统计收集器新增上次清理和分析的时间戳列（Larry Rosenman）

  这些值现在出现在 pg_stat_*_tables 系统视图中。

- 改进统计监控的性能，特别是 stats_command_string（Tom、Bruce）

  既然其开销已最小化，此版本默认启用 stats_command_string。这意味着 pg_stat_activity 现在默认显示所有活动查询。

- 为 pg_stat_activity 新增 waiting 列（Tom）

  这使 pg_stat_activity 能显示 ps 显示中包含的所有信息。

- 新增配置参数 update_process_title，控制 ps 显示是否随每条命令更新（Bruce）

  在更新 ps 显示代价高昂的平台上，关闭它并只依靠 pg_stat_activity 获取状态信息可能是值得的。

- 允许在配置设置中指定单位（Peter）

  例如，现在可以把 shared_buffers 设为 32MB，而不用心算换算大小。

- 新增对 postgresql.conf 中include 指令的支持（Joachim Wieland）

- 改进协议级 prepare/bind/execute 消息的日志记录（Bruce、Tom）

  这类日志现在显示语句名、绑定参数值以及被执行查询的文本。此外，当由 log_min_error_statement 启用时，查询文本也被正确包含在记录的错误消息中。

- 防止把 max_stack_depth 设为不安全的值

  在能确定实际内核栈深度限制的平台（大多数平台）上，确保 max_stack_depth 的初始默认值是安全的，并拒绝把它设为不安全的大值的企图。

- 在更多情形中启用查询中错误位置的高亮（Tom）

  服务器现在能为某些语义错误（如无法识别的列名）报告具体的错误位置，而不像以前那样只针对基本语法错误。

- 修复 VACUUM 中的“未能重新找到父键”错误（Tom）

- 在服务器重启期间清除 `pg_internal.init` 缓存文件（Simon）

  这避免了缓存文件在 PITR 恢复后可能包含陈旧数据的风险。

- 修复 `VACUUM` 截断跨越 GB 边界的大关系时的竞态条件（Tom）
- 修复导致行级锁上不必要死锁错误的缺陷（Tom）
- 修复影响多 GB 哈希索引的缺陷（Tom）
- 每个后端进程现在都是自己的进程组长（Tom）

  这允许查询取消中止从后端或归档/恢复进程调用的子进程。

E.120.3.3. 查询变更

- 新增 `INSERT`/`UPDATE`/`DELETE` 的 `RETURNING`（Jonah Harris、Tom）

  这允许这些命令返回值，例如为新行计算的序列键。在 `UPDATE` 的情形，返回行更新后版本的值。

- 按照 SQL 标准，新增对多行 `VALUES` 子句的支持（Joe、Tom）

  这允许 `INSERT` 插入多行常量，或查询用常量生成结果集。例如 `INSERT ... VALUES (...), (...), ....` 和 `SELECT * FROM (VALUES (...), (...), ....) AS alias(f1, ...)`。

- 允许 `UPDATE` 和 `DELETE` 为目标表使用别名（Atsushi Ogawa）

  SQL 标准不允许这些命令中使用别名，但许多数据库系统为了书写方便仍允许。

- 允许 `UPDATE` 用值列表设置多列（Susanne Ebrecht）

  这基本上是成对分配列和值的简写。语法为 `UPDATE tab SET (column, ...) = (val, ...)`。

- 让行比较按标准工作（Tom）

  <、<=、>、>= 形式现在按字典序比较行，即比较第一个元素，如果相等就比较第二个元素，依此类推。此前它们展开为对所有元素的 AND 条件，这既不标准也不太有用。

- 为 `TRUNCATE` 新增 `CASCADE` 选项（Joachim Wieland）

  这使 `TRUNCATE` 自动包含通过外键引用指定表的所有表。虽然方便，但这是危险工具——请谨慎使用！

- 支持在同一条 `SELECT` 命令中使用 `FOR UPDATE` 和 `FOR SHARE`（Tom）
- 新增 `IS NOT DISTINCT FROM`（Pavel Stehule）

  此操作符类似于相等（`=`），但当左右操作数都是 `NULL` 时结果为真，只有一个是时为假，而不是在这些情形下产生 `NULL`。

- 改进 `UNION`/`INTERSECT`/`EXCEPT` 使用的长度输出（Tom）

  当所有对应列的定义长度相同时，结果使用该长度而不是通用长度。

- 让 `ILIKE` 在多字节编码下工作（Tom）

在内部，`ILIKE` 现在先调用 `lower()` 再使用 `LIKE`。特定于区域设置的正则表达式模式在这些编码下仍不工作。

- 使 `standard_conforming_strings` 可以被设为 `on`（Kevin Grittner）

  这允许禁用字符串中的反斜杠转义，使 PostgreSQL 更符合标准。为了向后兼容默认值为 `off`，但未来版本将默认为 `on`。

- 不平展目标列表中包含 `volatile` 函数的子查询（Jaime Casanova）

  这防止了因 `volatile` 函数（如 `random()` 或 `nextval()`）被多次求值导致的意外行为。在存在被不必要地标记为 `volatile` 的函数时，它可能导致性能退化。

- 新增系统视图 `pg_prepared_statements` 和 `pg_cursors`，显示预备语句和打开的游标（Joachim Wieland、Neil）

  这在连接池设置中非常有用。

- 支持 `EXPLAIN` 和 `EXECUTE` 中的门户参数（Tom）

  这例如允许 JDBC 的 `?` 参数在这些命令中工作。

- 如果未指定 SQL 级 `PREPARE` 的参数，则从查询内容推断其类型（Neil）

  协议级的 `PREPARE` 已经这样做。

- 允许 `LIMIT` 和 `OFFSET` 超过二十亿（Dhanaraj M）

E.120.3.4. 对象操作变更

- 为 `CREATE TABLE AS` 新增 `TABLESPACE` 子句（Neil）

  这允许为新表指定表空间。

- 为 `CREATE TABLE AS` 新增 `ON COMMIT` 子句（Neil）

  这允许临时表在事务提交时被截断或删除。默认行为是表保留到会话结束。

- 为 `CREATE TABLE LIKE` 新增 `INCLUDING CONSTRAINTS`（Greg Stark）

  这允许把 `CHECK` 约束轻松复制到新表。

- 允许创建占位符（shell）类型（Martijn van Oosterhout）

  shell 类型声明只创建类型名，不指定类型的任何细节。创建 shell 类型是有用的，因为它允许更干净地声明类型的输入/输出函数，这些函数必须在类型可以“真正”定义之前存在。语法为 `CREATE TYPE` *`typename`*。

- 聚合函数现在支持多个输入参数（Sergey Koposov、Tom）

- 新增聚合创建语法（Tom）

  新语法为 `CREATE AGGREGATE` *`aggname`* `(`*`input_type`*`) (`*`parameter_list`*`)`。它更自然地支持新的多参数聚合功能。以前的语法仍受支持。

- 新增 `ALTER ROLE PASSWORD NULL`，移除先前设置的角色口令（Peter）

- 为许多对象类型新增 `DROP` 对象 `IF EXISTS`（Andrew）

  这允许对不存在的对象执行 `DROP` 操作而不产生错误。

- 新增 `DROP OWNED`，删除角色拥有的所有对象（Alvaro）

- 新增 `REASSIGN OWNED`，重新分配角色拥有的所有对象的所有权（Alvaro）

  它与上面的 `DROP OWNED` 一起便于删除角色。

- 新增 `GRANT ON SEQUENCE` 语法（Bruce）

  这是为设置序列特定的权限而添加的。为向后兼容，对序列的 `GRANT ON TABLE` 仍受支持。

- 为序列新增 `USAGE` 权限，只允许 `currval()` 和 `nextval()`，不允许 `setval()`（Bruce）

  `USAGE` 权限允许对序列访问更细粒度的控制。授予 `USAGE` 允许用户递增序列，但防止他们用 `setval()` 把序列设为任意值。

- 新增 `ALTER TABLE [ NO ] INHERIT`（Greg Stark）

  这允许动态调整继承，而不只是在表创建和销毁时。当用继承实现表分区时这非常有价值。

- 允许注释全局对象被全局存储（Kris Jurka）

  此前，附加到数据库的注释存储在各个数据库中，使其无效，
  而且完全没有为角色或表空间的注释做准备。此变更新增了共享目录 `pg_shdescription`，
  并把数据库、角色和表空间的注释存储在其中。

E.120.3.5. 工具命令变更

- 新增允许创建索引而不阻塞对表并发写入的选项（Greg Stark、Tom）

  新语法为 `CREATE INDEX CONCURRENTLY`。默认行为仍是在创建索引期间阻止表修改。

- 提供咨询锁定功能（Abhijit Menon-Sen、Tom）

  这是为取代以前 /contrib/userlock 中的内容而设计的新锁定 API。userlock 代码现在位于 pgfoundry。

- 允许 `COPY` 转储 `SELECT` 查询（Zoltan Boszormenyi、Karel Zak）

  这允许 `COPY` 转储任意 SQL 查询。语法为 `COPY (SELECT ...) TO`。

- 让 `COPY` 命令返回包含复制行数的命令标签（Volkan YAZICI）

- 允许 `VACUUM` 使行过期而不受其他并发 `VACUUM` 操作影响（Hannu Krossing、Alvaro、Tom）

- 让initdb检测操作系统区域设置并相应设置默认 `DateStyle`（Peter）

  这使安装的 `postgresql.conf` 中 `DateStyle` 值更可能符合期望。

- 减少 initdb 显示的进度消息数量（Tom）

E.120.3.6. 日期/时间变更

- 允许 `timestamp` 输入值中使用完整时区名（Joachim Wieland）

例如 '2006-05-24 21:11 America/New_York'::timestamptz。

- 支持可配置的时区缩写（Joachim Wieland）

  可以通过配置参数 timezone_abbreviations 选择所需的时区缩写集合。

- 新增 pg_timezone_abbrevs 和 pg_timezone_names 视图以显示受支持的时区（Magnus Hagander）

- 新增 clock_timestamp()、statement_timestamp() 和 transaction_timestamp()（Bruce）

  clock_timestamp() 是当前墙上时钟时间，statement_timestamp() 是当前语句到达服务器的时刻，transaction_timestamp() 是 now() 的别名。

- 允许 to_char() 打印本地化的月名和日名（Euler Taveira de Oliveira）

- 允许 to_char(time) 和 to_char(interval) 输出 AM/PM 规格（Bruce）

  interval 和 time 被视为 24 小时周期，例如 25 hours 被认为是 AM。

- 新增函数 justify_interval()，调整 interval 单位（Mark Dilger）

- 允许距 GMT 最多 14:59 的时区偏移

  基里巴斯使用 GMT+14，所以我们最好接受它。

- interval 计算改进（Michael Glaesemann、Bruce）

E.120.3.7. 其他数据类型和函数变更

- 允许数组包含 NULL 元素（Tom）

- 允许给与现有条目不连续的数组元素赋值（Tom）

  中间的数组位置将用空值填充。这符合 SQL 标准。

- 新增用于数组子集比较的内建操作符（@>、<@、&&）（Teodor、Tom）

  这些操作符可以用 GiST 或 GIN 索引为许多数据类型建立索引。

- 新增对 INET/CIDR 值的便捷算术操作（Stephen R. van den Berg）

  新操作符为 &（与）、|（或）、~（非）、inet + int8、inet - int8 和 inet - inet。

- 新增 SQL:2003 的新聚合函数（Neil）

  新函数为 var_pop()、var_samp()、stddev_pop() 和 stddev_samp()。var_samp() 和 stddev_samp() 只是现有聚合 variance() 和 stddev() 的改名。后两个名字仍可为向后兼容使用。

- 新增 SQL:2003 统计聚合（Sergey Koposov）

  新函数：regr_intercept()、regr_slope()、regr_r2()、corr()、covar_samp()、covar_pop()、regr_avgx()、regr_avgy()、regr_sxy()、regr_sxx()、regr_syy()、regr_count()。

- 允许域基于其他域（Tom）
- 在所有地方正确强制执行域的 `CHECK` 约束（Neil、Tom）

  例如，声明为返回域类型的用户定义函数的结果现在会对照域的约束检查。
  这堵住了域实现中一个重要的漏洞。
- 修复转储重命名的 `SERIAL` 列的问题（Tom）

  修复方法是在转储 `SERIAL` 列时显式指定其 `DEFAULT` 和序列元素，并在重载时用新的 `ALTER SEQUENCE OWNED BY` 命令重建 `SERIAL` 列。这也允许删除 `SERIAL` 列规格。
- 新增服务器端休眠函数 `pg_sleep()`（Joachim Wieland）
- 为 `tid`（元组 id）数据类型新增所有比较操作符（Mark Kirkwood、Greg Stark、Tom）

E.120.3.8. PL/pgSQL 服务器端语言变更

- 为触发器参数新增 `TG_table_name` 和 `TG_table_schema`（Andrew）

  `TG_relname` 现已废弃。其他 PL 的触发器参数也做了类似变更。
- 允许 `FOR` 语句向标量以及记录和行类型返回值（Pavel Stehule）
- 为 `FOR` 循环新增 `BY` 子句，控制迭代增量（Jaime Casanova）
- 为 `SELECT INTO` 新增 `STRICT`（Matt Miller）

  为了与 Oracle PL/SQL 兼容，当 `SELECT` 返回的行多于或少于一行时，`STRICT` 模式抛出异常。

E.120.3.9. PL/Perl 服务器端语言变更

- 为触发器参数新增 `table_name` 和 `table_schema`（Adam Sjøgren）
- 新增预备查询（Dmitry Karasik）
- 把 `$_TD` 触发器数据改为全局变量（Andrew）

  此前它是词法作用域的，导致意外的共享冲突。
- 出于安全原因，在独立解释器中运行 PL/Perl 和 PL/PerlU（Andrew）

  因此它们不能再共享数据或已装载的模块。此外，如果 Perl 编译时未带允许多解释器的必需标志，任何给定后端进程中只能使用这两种语言之一。

E.120.3.10. PL/Python 服务器端语言变更

- 命名参数除了通过 `args[]` 数组传递外，还作为普通变量传递（Sven Suursoho）
- 为触发器参数新增 `table_name` 和 `table_schema`（Andrew）
- 允许返回组合类型和结果集（Sven Suursoho）
- 以 `list`、`iterator` 或 `generator` 返回结果集（Sven Suursoho）

- 允许函数返回 `void`（Neil）
- 现在支持 Python 2.5（Tom）

E.120.3.11. psql Changes

- 新增命令 `\password`，用客户端口令加密更改角色口令（Peter）
- 允许 `\c` 连接到新的主机和端口号（David、Volkan YAZICI）
- 为 `\l+` 新增表空间显示（Philip Yarra）
- 改进 `\df` 反斜杠命令，使其包含函数的参数名和模式（`OUT` 或 `INOUT`）（David Fetter）
- 支持二进制 `COPY`（Andreas Pflug）
- 新增在单个事务中运行整个会话的选项（Simon）

  使用选项 `-1` 或 `--single-transaction`。
- 支持用游标自动分批获取 `SELECT` 结果（Chris Mair）

  这通过 `\set FETCH_COUNT` *n* 启用。此特性允许在 psql 中获取大结果集而不试图把整个结果集缓存在内存中。
- 使多行值在正确的列上对齐（Martijn van Oosterhout）

  包含换行符的字段值现在以更可读的方式显示。
- 把多行语句保存为单个条目，而不是一次一行（Sergey E. Koposov）

  这使用上箭头召回查询更容易。（这在 Windows 上不可用，因为该平台使用操作系统自带的本机命令行编辑。）
- 把行计数器改为 64 位，使其能处理超过二十亿行的文件（David Fetter）
- 为 `INSERT/UPDATE/DELETE RETURNING` 同时报告返回的数据和命令状态标签（Tom）

E.120.3.12. pg_dump Changes

- 允许 pg_dump 复杂选择要包含或排除的对象（Greg Sabino Mullane）

  pg_dump 现在支持多个 `-n`（模式）和 `-t`（表）选项，并新增 `-N` 和 `-T` 选项来排除对象。此外，这些开关的参数现在可以是通配表达式而不是单个对象名，例如 `-t 'foo*'`，模式也可以是 `-t` 或 `-T` 开关的一部分，例如 `-t schema1.table1`。
- 新增pg_restore的 `--no-data-for-failed-tables` 选项，在表创建失败（即表已存在）时抑制数据装载（Martin Pitt）
- 新增pg_restore选项，在单个事务中运行整个会话（Simon）

  使用选项 `-1` 或 `--single-transaction`。

E.120.3.13. libpq Changes

- 新增用于加密口令的 `PQencryptPassword()`（Tom）

这允许为 `ALTER ROLE ... PASSWORD` 等命令发送预加密的口令。

- 新增函数 `PQisthreadsafe()`（Bruce）

  这允许应用程序查询库的线程安全状态。

- 新增 `PQdescribePrepared()`、`PQdescribePortal()` 及相关函数，返回有关先前预备的语句和打开的游标的信息（Volkan YAZICI）

- 允许从 `pg_service.conf` 进行 LDAP 查找（Laurenz Albe）

- 允许 `~/.pgpass` 中的主机名匹配默认套接字目录（Bruce）

  空白主机名继续匹配任何 Unix 套接字连接，但此新增允许条目特定于机器上多个 postmaster 之一。

E.120.3.14. ecpg Changes

- 允许 `SHOW` 把结果放入变量（Joachim Wieland）

- 新增 `COPY TO STDOUT`（Joachim Wieland）

- 新增回归测试（Joachim Wieland、Michael）

- 大量源代码清理（Joachim Wieland、Michael）

E.120.3.15. Windows Port

- 允许 MSVC 编译 PostgreSQL 服务器（Magnus、Hiroshi Saito）

- 为工具命令和pg_dump新增 MSVC 支持（Hiroshi Saito）

- 新增对 Windows 代码页 `1253`、`1254`、`1255` 和 `1257` 的支持（Kris Jurka）

- 启动时丢弃权限，使服务器可以从管理账户启动（Magnus）

- 稳定性修复（Qingqing Zhou、Magnus）

- 新增本机信号量实现（Qingqing Zhou）

  以前的代码模仿 SysV 信号量。

E.120.3.16. 源代码变更

- 新增 GIN（通用倒排索引）索引访问方法（Teodor、Oleg）

- 移除 R-tree 索引（Tom）

  Rtree 已用 GiST 重新实现。除其他差异外，这意味着 rtree 索引现在通过预写日志（WAL）支持崩溃恢复。

- 减少不必要链接到后端的库（Martijn van Oosterhout、Tom）

- 新增 configure 标志，允许优先使用 libedit 而不是 GNU readline（Bruce）

  使用 configure 的 `--with-libedit-preferred`。

- 允许安装到包含空格的目录（Peter）
- 改进重新定位安装目录的能力（Tom）
- 新增对用 Solaris 编译器的 Solaris x86_64 的支持（Pierre Girard、Theo Schlossnagle、Bruce）
- 新增 DTrace 支持（Robert Lor）
- 新增 `PG_VERSION_NUM`，供想要在 C 中用 > 和 < 比较测试后端版本的第三方应用程序使用（Bruce）
- 新增独立于 `BLCKSZ` 的 `XLOG_BLCKSZ`（Mark Wong）
- 新增 `LWLOCK_STATS` 定义以报告锁定活动（Tom）
- 对未知的 configure 选项发出警告（Martijn van Oosterhout）
- 新增对“插件”库的服务器支持，可用于调试和性能测量等附加任务（Korry Douglas）

  这由两个特性组成：一张允许分别装载的共享库通信的“会合变量”表，以及新的配置参数 `local_preload_libraries`，
  它允许把库装载到特定会话而无需客户端应用程序的显式配合。这允许外部附加组件实现诸如 PL/pgSQL 调试器之类的特性。
- 把现有配置参数 `preload_libraries` 改名为 `shared_preload_libraries`（Tom）

  这是为了与 `local_preload_libraries` 相比更清晰。
- 新增配置参数 `server_version_num`（Greg Sabino Mullane）

  它类似 `server_version`，但是整数，例如 `80200`。这使应用程序更容易做版本检查。
- 新增配置参数 `seq_page_cost`（Tom）
- 把回归测试脚本重新实现为 C 程序（Magnus、Tom）
- 允许可装载模块分配共享内存和轻量锁（Marc Munro）
- 新增动态装载库的自动初始化和终结（Ralf Engelschall、Tom）

  如果库定义了这样的符号，就会调用新函数 `_PG_init()` 和 `_PG_fini()`。因此我们不再需要在 `shared_preload_libraries` 中指定初始化函数；可以假定库使用了 `_PG_init()` 约定。
- 为所有共享目标文件新增 `PG_MODULE_MAGIC` 头块（Martijn van Oosterhout）

  魔块防止可装载目标文件与服务器之间的版本不匹配。
- 新增 AIX 的共享库支持（Laurenz Albe）
- 新增 XML 文档节（Bruce）

E.120.3.17. 附加模块变更

- 重大 tsearch2 改进（Oleg、Teodor）

  - 多字节编码支持，包括 UTF8
  - 查询重写支持
  - 改进的排名函数
  - 同义词词典支持
  - Ispell 词典现在识别 MySpell 格式（OpenOffice 使用）
  - GIN 支持
- 新增包含 Pgadmin 管理函数的 adminpack 模块（Dave）

  这些函数提供默认 PostgreSQL 服务器中没有的额外文件系统访问例程。
- 新增 sslinfo 模块（Victor Wagner）

  报告当前连接 SSL 证书的信息。
- 新增 pgrowlocks 模块（Tatsuo）

  它显示指定表的行锁定信息。
- 新增 hstore 模块（Oleg、Teodor）
- 新增 isn 模块，取代 isbn_issn（Jeremy Kronuz）

  新实现支持 EAN13、UPC、ISBN（书籍）、ISMN（音乐）和 ISSN（期刊）。
- 为 pgstattuple 新增索引信息函数（ITAGAKI Takahiro、Satoshi Nagayasu）
- 新增显示空闲空间映射信息的 pg_freespacemap 模块（Mark Kirkwood）
- pgcrypto 现在具有所有计划的功能（Marko Kreen）
  - 在 pgcrypto 中包含 iMath 库，使公钥加密函数始终可用。
  - 新增 OpenBSD 代码中缺失的 SHA224 算法。
  - 在较旧的 OpenSSL 上激活 SHA224/256/384/512 哈希的内建代码，使这些算法始终可用。
  - 新函数 gen_random_bytes()，返回密码学上强的随机数。适用于生成加密密钥。
  - 移除 digest_exists()、hmac_exists() 和 cipher_exists() 函数。
- cube 模块改进（Joshua Reich）

  新函数为 `cube(float[])`、`cube(float[], float[])` 和 `cube_subset(cube, int4[])`。
- 为 dblink 新增异步查询能力（Kai Londenberg、Joe Conway）
- 新增用于数组子集比较的操作符（`@>`、`<@`、`&&`）（Tom）

  各种 contrib 包已经为它们的数据类型提供了这些操作符，但命名不一致。我们现在为核心代码和所有具有此功能的 contrib 包添加了命名一致的数组子集比较操作符。（旧名称仍可用，但已废弃。）

- 为所有有安装脚本的 contrib 包新增卸载脚本（David、Josh Drake）

E.121. 版本 8.1.23

发行日期
2010-12-16

本次发布包含对 8.1.22 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

这预计是 PostgreSQL 8.1.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.121.1. 迁移到版本 8.1.23

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅第 E.126 节。

E.121.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

  给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

  在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）

- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

  该缺陷影响多个不同的几何距离测量操作符。

- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）

- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `cube` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。

- `contrib/dblink` 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）

- 修复 `contrib/pgcrypto` 在缺少公钥时可能的核心转储（Marti Raudsepp）

- 修复 `contrib/xml2` 的 XPath 查询函数中的内存泄漏（Tom Lane）

- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.122. 版本 8.1.22

发行日期
2010-10-04

本次发布包含对 8.1.21 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

PostgreSQL 社区将于 2010 年 11 月停止对 8.1.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.122.1. 迁移到版本 8.1.22

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅第 E.126 节。

E.122.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。
- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）
- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）
- 修复对按引用传递的函数结果计算哈希时可能的失败（Tao Ma、Tom Lane）
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。
- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。
- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）
- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）
- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）
- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

E.123. 版本 8.1.21

发行日期
2010-05-17

本次发布包含对 8.1.20 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.123.1. 迁移到版本 8.1.21

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅第 E.126 节。

E.123.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 防止 psql 在展开引用自身的变量时无限递归（Tom）

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

E.124. 版本 8.1.20

发行日期
2010-03-15

本次发布包含对 8.1.19 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.124.1. 迁移到版本 8.1.20

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅第 E.126 节。

E.124.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）

- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）

- 修复若干病态缓慢的正则表达式匹配情形（Tom）

- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）

- 修复更多临时文件泄漏的情形（Heikki）

  这纠正了上一个小版本中引入的问题。失败的一种情形是返回集合的 plpgsql 函数在另一个函数的异常处理器中被调用。

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 修复 psql 的 `numericlocale` 选项，使其不在 latex 和 troff 输出格式中格式化不应格式化的字符串（Heikki）
- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）
- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）
- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）
- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.125. 版本 8.1.19

发行日期
2009-12-14

本次发布包含对 8.1.18 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.125.1. 迁移到版本 8.1.19

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅第 E.126 节。

E.125.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。

- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。

- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 确保预备事务所持有的共享元组级锁不被忽略（Heikki）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 修复 `CREATE OR REPLACE FUNCTION` 期间所有权依赖的处理（Tom）
- 确保 Perl 数组在由返回集合的 PL/Perl 函数返回时被正确转换为 PostgreSQL 数组（Andrew Dunstan、Abhijit Menon-Sen）

  对非返回集合的函数这早已正确工作。

- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。

- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.126. 版本 8.1.18

发行日期
2009-09-09

本次发布包含对 8.1.17 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.126.1. 迁移到版本 8.1.18

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.1.18 后对它们执行 `REINDEX`。另外，如果你是从早于 8.1.15 的版本升级，请参阅第 E.129 节。

E.126.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

  它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）

- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）

- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.127. 版本 8.1.17

发行日期
2009-03-16

本次发布包含对 8.1.16 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.127.1. 迁移到版本 8.1.17

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.15 的版本升级，请参阅第 E.129 节。

E.127.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，
  解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）

- 修复带隐式强制转换的 `CASE WHEN` 的反编译（Tom）

  在试图检查或转储视图时，这一错误在启用断言的构建中可能导致断言失败，
  在其他情况下可能导致“意外的 CASE WHEN 子句”错误消息。

- 修复 TOAST 表行类型所有者可能的错误分配（Tom）

  如果 `CLUSTER` 或重写形式的 `ALTER TABLE` 由表所有者之外的人执行，表的 TOAST 表的 `pg_type` 条目最终会被标记为那个人所拥有。这不会立即引起问题，因为普通数据库操作不检查 TOAST 行类型的权限。但是，如果后来试图删除发出该命令的角色（在 8.1 或 8.2 中），或者在那之后 pg_dump 发出“数据类型的所有者似乎无效”警告（在 8.3 中），则可能导致意外失败。

- 在块退出时完全清理 PL/pgSQL 错误状态变量（Ashesh Vashi 和 Dave Page）

  这对 PL/pgSQL 本身不是问题，但这一疏漏可能导致 PL/pgSQL 调试器在检查函数状态时崩溃。

- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.128. 版本 8.1.16

发行日期
2009-02-02

本次发布包含对 8.1.15 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.128.1. 迁移到版本 8.1.16

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.15 的版本升级，请参阅第 E.129 节。

E.128.2. 变更

- 修复 autovacuum 中的崩溃（Alvaro）

  该崩溃只在为防止事务回卷而清理整个数据库之后发生，这意味着它很少发生且难以追查。

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 确保可保持游标的内容不依赖于 TOAST 表的内容（Tom）

  此前，游标结果中的大字段值可能被表示为 TOAST 指针，如果被引用的表在游标被读取之前被删除，或者大值被删除随后被清理掉，这就会失败。普通游标不会发生这种情况，但对于保持到创建事务之后的游标则可能。

- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 修复 configure 脚本，使其在无法获得 PL/Perl 的链接信息时正确报告失败（Andrew）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.129. 版本 8.1.15

发行日期
2008-11-03

本次发布包含对 8.1.14 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.129.1. 迁移到版本 8.1.15

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。另外，如果你运行的是之前的 8.1.X 版本，建议在升级后对所有 GiST 索引执行 `REINDEX`。

E.129.2. 变更

- 修复 GiST 索引因删除后把错误的索引条目标记为“死”而损坏的问题（Teodor）

  这会导致索引搜索找不到本应找到的行。损坏的索引可以用 `REINDEX` 修复。

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）
- 修复当子 `SELECT` 出现在 `FROM` 中的函数调用、多行 `VALUES` 列表或 `RETURNING` 列表中时规则查询的错误展开（Tom）

  此问题通常的症状是“无法识别的节点类型”错误。
- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）
- 防止用 `ALTER SET TABLESPACE` 把表移到另一个表空间时 `relfilenode` 号可能的冲突（Heikki）

  该命令试图重用现有文件名，而不是选取目标目录中已知未使用的文件名。
- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。
- 修复 ecpg 对 `CREATE ROLE` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.130. 版本 8.1.14

发行日期
2008-09-22

本次发布包含对 8.1.13 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.130.1. 迁移到版本 8.1.14

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.130.2. 变更

- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。

- 修复 GiST 索引扫描期间可能的元组重复输出（Teodor）
- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  `ALTER COLUMN TYPE` 之后重用先前缓存的计划可能产生这种情形。该检查防范随之而来的数据损坏和/或崩溃。

- 修复 `AT TIME ZONE`：先尝试把时区参数解释为时区缩写，仅在失败时才尝试完整时区名，而不是像以前那样相反（Tom）

  时间戳输入函数一直按此顺序解析有歧义的时区名。让 `AT TIME ZONE` 也这样做改进了一致性，并修复了 8.1 中引入的兼容性缺陷：在有歧义的情形下，我们现在的行为与 8.0 及更早版本相同，因为在旧版本中 `AT TIME ZONE` 只接受缩写。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复嵌套子查询表达式的规划器缺陷（Tom）

  如果外层子查询不直接依赖父查询而内层依赖，外层的值可能不会为新的父查询行重新计算。

- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/pgSQL 在 `FOR` 循环的目标变量是包含组合类型字段的记录时不失败（Tom）
- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）
- 修复 PL/Python 使其与 Python 2.5 一起工作

  这是 8.2 开发周期中所做修复的向后移植。

- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）
- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）
- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.131. 版本 8.1.13

发行日期
2008-06-12

此发行版在 8.1.12 的基础上包含一个严重和一个次要的修复。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.131.1. 迁移到版本 8.1.13

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.131.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

- 让 `ALTER AGGREGATE ... OWNER TO` 更新 `pg_shdepend`（Tom）

  如果聚合后来涉及 `DROP OWNED` 或 `REASSIGN OWNED` 操作，这一疏漏可能导致问题。

E.132. 版本 8.1.12

发行日期
从未发布

本次发布包含对 8.1.11 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.132.1. 迁移到版本 8.1.12

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.132.2. 变更

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 ORDER BY 和 DISTINCT 表达式不匹配。

- 修复正则表达式子串匹配（substring(*string* from *pattern*)）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 substring('foo' from 'foo(bar)?')。由于 (bar) 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 foo）。

- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴、阿根廷/圣路易斯和智利的 DST 法规变更）

- 修复 ecpg 的 PGTYPEStimestamp_sub() 函数结果不正确的问题（Michael）

- 修复 contrib/xml2 的 xpath_table() 函数在输入查询返回 NULL 值时的核心转储（Tom）

- 修复 contrib/xml2 的 makefile，使其不覆盖 CFLAGS（Tom）

- 修复 DatumGetBool 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

- 修复长期存在的 LISTEN/NOTIFY 竞态条件（Tom）

  极少数情况下，刚执行过 LISTEN 的会话可能收不到通知，即使按理应当收到，因为观察到执行 NOTIFY 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 LISTEN 命令的事务如果去查看，将看不到该 LISTEN 在 pg_listener 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 不允许在预备事务内使用 LISTEN 和 UNLISTEN（Tom）

  这以前是允许的，但尝试这样做会有各种不愉快的后果，特别是只要 UNLISTEN 未提交，发起的后端就无法退出。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）

- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。

- 修复 ALTER OWNER 某些变体中的“无法识别的节点类型”错误（Tom）

- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 pg_ctl start -w 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。

- 使用 -fwrapv 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。

- 修复 ORDER BY 和 GROUP BY 中常量表达式的显示（Tom）

显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.133. 版本 8.1.11

发行日期
2008-01-07

此发行版包含 8.1.10 之后的多种修复，包括对重要安全问题的修复。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

这是 PostgreSQL 社区为 Windows 制作二进制包的最后一个 8.1.X 版本。建议 Windows 用户迁移到 8.2.X 或更高版本，因为 8.2.X 中有一些不便于向后移植的 Windows 特定修复。8.1.X 在其他平台上将继续受支持。

E.133.1. 迁移到版本 8.1.11

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.133.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  8.1.10 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 更新时区数据文件到 tzdata 发行版 2007k（特别是阿根廷最近的变更）（Tom）
- 改进规划器在非 C 区域设置下对 LIKE/正则表达式估计的处理（Tom）
- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）
- 保留被 `ALTER TABLE ... ALTER COLUMN TYPE` 重建的索引的表空间（Tom）
- 让归档恢复总是开始新的 WAL 时间线，而不是只在使用恢复停止时间时（Simon）

  这避免了试图覆盖最后一个 WAL 段现有归档副本的极端情形风险，而且看起来比原来的定义更简单干净。
- 当表太小而不适用时，让 `VACUUM` 不使用全部 `maintenance_work_mem`（Alvaro）
- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）
- 修复超过 68 年的 interval 在 `extract(epoch from interval)` 中的溢出（Tom）
- 修复 PL/Perl 在可信函数中使用 UTF-8 正则表达式时不失败（Andrew）
- 修复 PL/Perl 在平台 Perl 把 `bool` 类型定义为 `int` 而非 `char` 时的问题（Tom）

  虽然理论上这可能在任何地方发生，但没有标准 Perl 构建这样做.....直到 Mac OS X 10.5。
- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）
- 修复 pg_dump，使其正确处理默认表达式与父表不同的继承子表（Tom）
- 修复 `PGPASSFILE` 指向非普通文件时 libpq 的崩溃（Martin Pitt）
- ecpg 解析器修复（Michael）
- 让 `contrib/pgcrypto` 防御在超过 128 位密钥上失败的 OpenSSL 库——至少某些 Solaris 版本是这种情形（Marko Kreen）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）
- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.134. 版本 8.1.10

发行日期
2007-09-17

本次发布包含对 8.1.9 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.134.1. 迁移到版本 8.1.10

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.134.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 允许 `interval` 数据类型接受仅由毫秒或微秒组成的输入（Neil）
- 加快 rtree 索引插入（Teodor）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复日志，使使用 syslogger 进程时日志消息永不交错（Andrew）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 修复某些外键极端情形的错误处理（Tom）
- 防止 `REINDEX` 和 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 更新时区数据库规则，特别是新西兰即将到来的变更（Tom）
- Windows 套接字改进（Magnus）
- 因可能的编码不匹配，在 Windows 上抑制日志时间戳中的时区名（`%Z`）（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.135. 版本 8.1.9

发行日期
2007-04-23

此发行版包含 8.1.8 之后的多种修复，包括一个安全修复。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.135.1. 迁移到版本 8.1.9

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.135.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）
- 要求 `COMMIT PREPARED` 在事务预备时所在的同一数据库中执行（Heikki）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 规划器修复，包括改进外连接和位图扫描选择逻辑（Tom）
- 修复哈希索引扩容期间的 PANIC（8.1.6 中引入的缺陷）（Tom）
- 修复 POSIX 风格时区规范以遵循新的美国 DST 规则（Tom）

E.136. 版本 8.1.8

发行日期
2007-02-07

此发行版包含 8.1.7 之后的一个修复。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.136.1. 迁移到版本 8.1.8

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.136.2. 变更

- 移除约束和函数索引中对类型长度的过度严格检查（Tom）

E.137. 版本 8.1.7

发行日期
2007-02-05

此发行版包含 8.1.6 之后的多种修复，包括一个安全修复。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.137.1. 迁移到版本 8.1.7

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.137.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  这些漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，以及更改表列的数据类型（CVE-2007-0555、CVE-2007-0556）。这些错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 改进有大量表的数据库的 `VACUUM` 性能（Tom）

- 修复 autovacuum，避免在不可连接的数据库中留下非持久事务 ID（Alvaro）

  此缺陷只影响 8.1 分支。

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

- 修复 Windows 上因试图 fsync 已删除文件而发生的虚假“权限被拒绝”失败（Magnus、Tom）

- 修复已在使用中的 PL/pgSQL 函数被更新时可能的崩溃（Tom）

E.138. 版本 8.1.6

发行日期
2007-01-08

本次发布包含对 8.1.5 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.138.1. 迁移到版本 8.1.6

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.138.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 pg_restore 处理包含带注释的大对象（blob）的 tar 格式备份（Tom）

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）

- 在服务器重启期间清除 `pg_internal.init` 缓存文件（Simon）

这避免了缓存文件在 PITR 恢复后可能包含陈旧数据的风险。

- 修复 `VACUUM` 截断跨越 GB 边界的大关系时的竞态条件（Tom）
- 修复导致行级锁上不必要死锁错误的缺陷（Tom）
- 修复影响多 GB 哈希索引的缺陷（Tom）
- 修复 Windows 信号处理中可能的死锁（Teodor）
- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）
- 修复 ecpg 连接期间的内存泄漏（Michael）
- 修复 Darwin（OS X）编译（Tom）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。
- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。
- 更新时区数据库

  这特别影响澳大利亚和加拿大的夏令时规则。

E.139. 版本 8.1.5

发行日期
2006-10-16

本次发布包含对 8.1.4 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.139.1. 迁移到版本 8.1.5

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.139.2. 变更

- 不允许在 `UPDATE` 命令中使用聚合函数，子 SELECT 中除外（Tom）

  这种聚合的行为不可预测，而且在 8.1.X 中可能导致崩溃，因此已被禁用。SQL 标准也不允许这样做。
- 修复把无类型字面值当作 ANYARRAY 时的核心转储
- 修复扩展查询协议中执行 `COMMIT` 或 `ROLLBACK` 时持续时间日志记录的核心转储

- 修复查询包含返回多行的 SQL 函数时 AFTER 触发器的错误处理（Tom）
- 修复 `ALTER TABLE ... TYPE`，对 `USING` 子句重新检查 `NOT NULL`（Tom）
- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。
- 修复 `AM/PM` 格式的 `to_timestamp()`（Bruce）
- 修复 autovacuum 决定是否需要 `ANALYZE` 的计算（Alvaro）
- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- ecpg 的大量健壮性修复（Joachim Wieland）
- 修复 /contrib/dbmirror 中的反斜杠转义
- /contrib/dblink 和 /contrib/tsearch2 的小修复
- 哈希表和位图索引扫描的效率改进（Tom）
- 修复 Windows 上统计收集的不稳定（Tom、Andrew）
- 修复 `statement_timeout` 在 Win32 上使用正确的单位（Bruce）

  在以前的 Win32 8.1.X 版本中，延迟偏差了 100 倍。
- MSVC 和 Borland C++ 编译器的修复（Hiroshi Saito）
- AIX 和 Intel 编译器的修复（Tom）
- 修复连续归档中的罕见缺陷（Tom）

E.140. 版本 8.1.4

发行日期
2006-05-23

此发行版包含 8.1.3 之后的多种修复，包括针对极其严重安全问题的补丁。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.140.1. 迁移到版本 8.1.4

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐

的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescape StringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.140.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复 pgcrypto 中的弱密钥选择（Marko Kreen）

  fortuna PRNG 重播种逻辑中的错误可能在某些情况下导致 `pgp_sym_encrypt()` 选择可预测的会话密钥。这只影响不使用 OpenSSL 的构建。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`win866_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 让 autovacuum 在 `pg_stat_activity` 中可见（Alvaro）

- 禁用 `full_page_writes`（Tom）

  某些情况下，关闭 `full_page_writes` 会导致崩溃恢复失败。适当的修复将出现在 8.2 中；目前只是禁用它。

- 各种规划器修复，特别是位图索引扫描和 MIN/MAX 优化（Tom）

- 修复归并连接中的错误优化（Tom）

  外连接有时可能发出未匹配行的多个副本。

- 修复在同一事务中使用并修改 plpgsql 函数导致的崩溃

- 修复 B-Tree 索引被截断情形的 WAL 重放

- 修复涉及 `|` 的模式的 `SIMILAR TO`（Tom）

- 修复 `SELECT INTO` 和 `CREATE TABLE AS`，在默认表空间而不是基目录中创建表（Kris Jurka）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 改进 qsort 性能（Dann Corbit）

  目前此代码只在 Solaris 上使用。
- 修复 x86 系统上的 OS/X Bonjour（Ashley Clark）
- 修复各种小的内存泄漏
- 修复某些 Win32 系统上的口令提示问题（Robert Kinberg）
- 改进 pg_dump 对域默认值的处理
- 修复 pg_dumpall 合理处理同名的用户和组（只在从 8.1 之前的服务器转储时可能）（Tom）

  用户和组将被合并为带 `LOGIN` 权限的单个角色。此前合并后的角色没有 `LOGIN` 权限，使其不能作为用户使用。
- 修复 pg_restore 的 `-n`，使其按文档工作（Tom）

E.141. 版本 8.1.3

发行日期
2006-02-14

此发行版包含 8.1.2 之后的多种修复，包括一个非常严重的安全问题。有关 8.1 主版本新特性的信息，请参阅第 E.144 节。

E.141.1. 迁移到版本 8.1.3

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅第 E.142 节。

E.141.2. 变更

- 修复允许任何已登录用户把 `SET ROLE` 设为任何其他数据库用户 id 的缺陷（CVE-2006-0553）

  由于有效性检查不充分，用户可以利用 `SET ROLE` 通常在错误后用来恢复先前角色设置的特殊情形。这例如允许普通用户获得超级用户状态。权限提升风险只存在于 8.1.0-8.1.2。但在直到 7.3 的所有版本中，`SET SESSION AUTHORIZATION` 都有一个相关缺陷，如果服务器编译时启用了断言（这不是默认），非特权用户可使服务器崩溃。感谢 Akio Ishida 报告此问题。
- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 8.0.4、7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 和 pg_subtrans 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复缓存失效消息恰好在错误时刻到达时可能导致崩溃的情形（Tom）
- 为预备语句中的 `UNKNOWN` 参数正确检查 `DOMAIN` 约束（Neil）
- 确保 `ALTER COLUMN TYPE` 按正确顺序处理 `FOREIGN KEY`、`UNIQUE` 和 `PRIMARY KEY` 约束（Nakano Yoshihisa）
- 修复以允许恢复有对自定义操作符或操作符类的跨模式引用的转储（Tom）
- 允许 pg_restore 在 `COPY` 失败后正确继续；此前它试图把剩余的 `COPY` 数据当作 SQL 命令（Stephen Frost）
- 修复未指定数据目录时 pg_ctl 的 `unregister` 崩溃（Magnus）
- 修复 libpq 的 `PQprint` HTML 标签（Christoph Zwerschke）
- 修复 AMD64 和 PPC 上 ecpg 的崩溃（Neil）
- 允许在函数结果类型声明中同时使用 `SETOF` 和 `%TYPE`
- 修复 PL/python 参数传递期间发生错误时的正确恢复（Neil）
- 修复 `plperl_return_next` 的内存泄漏（Neil）
- 修复 PL/perl 在 Win32 上对区域设置的处理以与后端匹配（Andrew）
- 各种优化器修复（Tom）
- 修复 Win32 上 `postgresql.conf` 中 `log_min_messages` 设为 `DEBUG3` 或更高时的崩溃（Bruce）
- 修复 Win32、Cygwin、OS X、AIX 的 pgxs `-L` 库路径指定（Bruce）
- 在检查 Win32 管理员权限时检查 SID 是否启用（Magnus）
- 正确拒绝超出范围的日期输入（Kris Jurka）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）
- 通过避免每数据行一次内核调用，改进经 libpq 的 `COPY IN` 速度（Alon Goldshuv）
- 改进 `/contrib/tsearch2` 索引创建的速度（Tom）

E.142. 版本 8.1.2

发行日期
2006-01-09

本次发布包含对 8.1.1 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.142.1. 迁移到版本 8.1.2

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.142.2. 变更

- 修复 Windows 代码，使 ShmemBackendArray 没有更多空间时 postmaster 继续而不是退出（Magnus）

  如果过多连接请求几乎同时到达，以前的行为可能导致拒绝服务情形。这只适用于 Windows 移植。

- 修复 8.0 中引入的缺陷，它可能使 ReadBuffer 返回已使用过的页当作新页，可能导致最近提交的数据丢失（Tom）

- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）

- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 允许更灵活地重新定位安装目录（Tom）

  以前的版本只有在所有安装目录路径除最后一个组件外都相同时才支持重新定位。

- 防止因使用 `ISO-8859-5` 和 `ISO-8859-9` 编码导致的崩溃（Tatsuo）

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）

- 修复 COPY CSV 模式把任何 `\.` 视为复制数据结束的缺陷

  新代码要求 `\.` 单独出现在一行上，如文档所述。

- 让 COPY CSV 模式给字面数据值 `\.` 加引号，以确保它不会被解释为数据结束标记（Bruce）

- 对返回 `RECORD` 的函数的各种修复（Tom）

- 修复 `postgresql.conf` 的处理，使没有换行符的最后一行被正确处理（Tom）

- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复处理表达式索引时 autovacuum 的崩溃

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.143. 版本 8.1.1

发行日期
2005-12-12

本次发布包含对 8.1.0 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.144 节。.

E.143.1. 迁移到版本 8.1.1

对于运行 8.1.X 的用户，不需要进行转储/恢复。

E.143.2. 变更

- 修复外连接条件的错误优化（Tom）
- 修复涉及被优化器平展的子查询的情形下报告列名错误的问题（Tom）
- 修复涉及 CHECK 约束、TOAST 列和索引的场景中的更新失败（Tom）
- 修复从错误恢复后 bgwriter 的问题（Tom）

  发现后台写入器在写错误后泄漏缓冲区固定。虽然本身不致命，但这可能导致后续 VACUUM 命令的神秘阻塞。
- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败
- `/contrib/tsearch2` 和 `/contrib/ltree` 修复（Teodor）
- 修复需要单词重排的语言（如土耳其语）中翻译错误消息的问题；还有输出字符串意外截断和 bigint 最小可能值显示错误的问题（Andrew、Tom）

  这些问题只出现在使用我们的 `port/snprintf.c` 代码的平台上，包括指定了 `--enable-nls` 的 BSD 变体，可能还有其他平台。此外，根据所使用的 `libintl` 版本，翻译错误消息问题的另一种形式可能出现在 Windows 上。
- 重新允许 `to_char(time)` 和 `to_char(interval)` 使用 `AM`/`PM`、`HH`、`HH12` 和 `D` 格式说明符。（`to_char(interval)` 或许应该用 `HH24`。）（Bruce）
- AIX、HPUX 和 MSVC 编译修复（Tom、Hiroshi Saito）
- 优化器改进（Tom）
- 在 Windows NO_SYSTEM_RESOURCES 错误后重试文件读写（Qingqing Zhou）
- 防止 autovacuum 在分析表达式索引期间崩溃（Alvaro）
- 修复 ON COMMIT DELETE ROWS 临时表的问题
- 修复触发器更改 SELECT DISTINCT 查询输出时的问题
- 新增关于如何迁移无效 `UTF-8` 字节序列的 8.1.0 发行说明条目（Paul Lindner）

E.144. 版本 8.1

发行日期
2005-11-08

E.144.1. 概述

此发行版的主要变更：

Improve concurrent access to the shared buffer cache (Tom)

对共享缓冲区缓存的访问被认定为显著的可伸缩性问题，特别是在多 CPU 系统上。在此版本中，缓冲区管理器中加锁的方式被彻底改造，以减少锁争用并改进可伸缩性。
缓冲区管理器还被改为使用“时钟扫描”替换策略。

Allow index scans to use an intermediate in-memory bitmap (Tom)

在以前的版本中，对表做查找只能使用单个索引。有了此特性，如果查询有 `WHERE tab.col1 = 4 and tab.col2 = 9`，而 `col1` 和 `col2` 上没有多列索引，但 `col1` 上有一个索引、`col2` 上有另一个索引，则可以搜索两个索引并在内存中合并结果，然后只对同时匹配 `col1` 和 `col2` 限制的行做堆获取。这在有大量非结构化查询、不可能创建匹配所有可能访问条件的索引的环境中非常有用。位图扫描即使对单个索引也有用，因为它减少了所需的随机访问量；
位图索引扫描对于检索整表中相当大的部分是高效的，而普通索引扫描则不是。

Add two-phase commit (Heikki Linnakangas, Alvaro, Tom)

两阶段提交允许在多台计算机上"预备"事务，
一旦所有计算机都成功预备了各自的事务（无一失败），所有事务就可以提交。
即使机器在预备后崩溃，预备的事务也可以在机器重启后提交。新语法包括 `PREPARE TRANSACTION` 和 `COMMIT/ROLLBACK PREPARED`。还新增了系统视图 `pg_prepared_xacts`。

Create a new role system that replaces users and groups (Stephen Frost)

角色是用户和组的组合。像用户一样，它们可以有登录能力；像组一样，
角色可以拥有其他角色作为成员。角色基本上消除了用户和组的区别。例如，角色可以：

- （可选地）拥有登录能力
- 拥有对象
- 持有数据库对象的访问权限
- 从它是成员的其他角色继承权限

用户登录到角色后，就获得该登录角色加上任何继承角色的能力，并可以用 `SET ROLE` 切换到她是成员的其他角色。此特性是 SQL 标准角色概念的推广。此变更还用支持角色的新目录 `pg_authid` 和 `pg_auth_members` 取代了 `pg_shadow` 和 `pg_group`。旧表被重新定义为新角色表上的只读视图。

Automatically use indexes for `MIN()` and `MAX()` (Tom)

在以前的版本中，让 `MIN()` 或 `MAX()` 使用索引的唯一方法是把查询改写为 `SELECT col FROM tab ORDER BY col LIMIT 1`。现在索引的使用自动发生。

Move `/contrib/pg_autovacuum` into the main server (Alvaro)

把 autovacuum 集成到服务器中使其可以随数据库服务器同步自动启动和停止，并允许从 `postgresql.conf` 配置 autovacuum。

Add shared row level locks using `SELECT ... FOR SHARE` (Alvaro)

虽然 PostgreSQL 的 MVCC 锁定使 `SELECT` 从不被写入者阻塞，因此典型操作不需要共享行锁，但共享锁对需要共享行锁定的应用程序很有用。
特别是这降低了参照完整性检查施加的锁定要求。

Add dependencies on shared objects, specifically roles (Alvaro)

依赖机制的这一扩展防止角色在其仍拥有数据库对象时被删除。
此前可能通过删除所有者而意外地使对象“孤立”。虽然这可以恢复，但既混乱又令人不快。

Improve performance for partitioned tables (Simon)

新的 `constraint_exclusion` 配置参数避免在约束表明子表中不存在匹配行时查找子表。

这支持一种基本的表分区。如果子表存储独立的键范围并用适当的 `CHECK` 约束强制执行，当约束保证子表中不存在匹配行时，优化器将跳过对子表的访问。

E.144.2. 迁移到版本 8.1

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

8.0 发行版宣布 interval 的 `to_char()` 函数将在 8.1 中移除。但由于没有提出更好的 API，`to_char(interval)` 在 8.1 中得到增强并将保留在服务器中。

请注意以下不兼容之处：

- `add_missing_from` 现在默认为 false（Neil）

  默认情况下，如果查询中使用的表没有 `FROM` 引用，我们现在会产生错误。旧行为仍可用，但必须把参数设为 'true' 才能获得。

  如果转储包含用隐式 `FROM` 语法创建的视图或规则，可能需要把 `add_missing_from` 设为 true 才能装载现有转储文件。这应当只是一次性的麻烦，因为 PostgreSQL 8.1 会把这样的视图和规则转换为标准的显式 `FROM` 语法。因此后续转储不会有此问题。

- 使 `float4`/`float8`/`oid` 输入零长度字符串（`''`）抛出错误，而不是把它当作零（Neil）

  此变更与当前整数的零长度字符串处理一致。此变更的时间表在 8.0 中已宣布。

- `default_with_oids` 现在默认为 false（Neil）

  此选项设为 false 后，用户创建的表不再有 OID 列，除非在 `CREATE TABLE` 中指定 `WITH OIDS`。虽然 OID 在 PostgreSQL 的所有版本中都存在，但它们的用途有限，因为它们只有四字节长，而且计数器在所有已安装数据库之间共享。唯一标识行的首选方式是通过序列和 `SERIAL` 类型，它们自 PostgreSQL 6.4 起就受支持。

- 新增 E'' 语法，使最终普通字符串能按字面处理反斜杠（Bruce）

  目前 PostgreSQL 把字符串字面量中的反斜杠当作引入特殊转义序列（如 \n 或 \010）处理。虽然这便于输入特殊值，但它非标准且使从其他数据库移植应用程序更困难。因此，PostgreSQL 项目计划移除字符串中反斜杠的特殊含义。为了向后兼容和满足想要特殊反斜杠处理的用户，创建了新的字符串语法。
  新字符串语法的写法是在启动字符串的单引号前紧挨着写 E，例如 E'hi\n'。虽然此版本不改变字符串中反斜杠的处理，但它确实新增了帮助用户为未来版本迁移应用程序的配置参数：

  - standard_conforming_strings——此版本是否按字面处理普通字符串中的反斜杠?

  - escape_string_warning——对普通（非 E）字符串中的反斜杠发出警告

  standard_conforming_strings 的值是只读的。应用程序可以检索该值以了解反斜杠如何被处理。（参数的存在也可被视为支持 E'' 字符串语法的标志。）在未来版本中，standard_conforming_strings 将为 true，意味着非 E 字符串中的反斜杠将按字面处理。为此变更做准备：在需要特殊反斜杠处理的地方使用 E'' 字符串，并打开 escape_string_warning 以找出其他需要转换为使用 E'' 的字符串。此外，用两个单引号（''）在字符串中嵌入字面单引号，而不要用 PostgreSQL 支持的反斜杠单引号（\'）语法。前者符合标准且不需要使用 E'' 字符串语法。也可以使用 $$ 字符串语法，它不特殊处理反斜杠。

- 让 REINDEX DATABASE 重新索引数据库中的所有索引（Tom）

  此前 REINDEX DATABASE 只重新索引系统表。新行为看起来更直观。新命令 REINDEX SYSTEM 提供只重新索引系统表的旧功能。

- 只读大对象描述符现在遵循 MVCC 快照语义

  当大对象以 INV_READ（而非 INV_WRITE）打开时，从描述符读取的数据现在将反映调用 lo_open() 的查询所用事务快照时刻大对象状态的“快照”。要获得总是返回最新已提交数据的旧行为，请在 lo_open() 的模式标志中包含 INV_WRITE。

- 为序列函数的参数添加适当的依赖（Tom）

  在以前的版本中，传递给 nextval()、currval() 和 setval() 的序列名存储为简单文本字符串，这意味着重命名或删除 DEFAULT 子句中使用的序列会使子句无效。此版本把所有新创建的序列函数参数存储为内部 OID，使它们能跟踪序列重命名，并添加防止不当删除序列的依赖信息。它还使这样的 DEFAULT 子句不受模式重命名和搜索路径更改的影响。

  有些应用程序可能依赖旧的运行时查找序列名的行为。仍然可以通过把参数显式转换为 text 来做到，例如 nextval('myseq'::text)。

  装载到 8.1 的 8.1 之前数据库转储将使用旧的基于文本的表示，因此不会有 OID 存储参数的特性。但可以更新包含基于文本的 DEFAULT 子句的数据库。首先，把下面这个查询保存到文件中，例如 fixseq.sql：

```
SELECT  'ALTER TABLE ' ||
   pg_catalog.quote_ident(n.nspname) || '.' ||
   pg_catalog.quote_ident(c.relname) ||
   ' ALTER COLUMN ' || pg_catalog.quote_ident(a.attname) ||
   ' SET DEFAULT ' ||
   regexp_replace(d.adsrc,
                  $$val\(\(('[^']*')::text\)::regclass$$,
                  $$val(\1$$,
                  'g') ||
```

```
   ';'
FROM    pg_namespace n, pg_class c, pg_attribute a, pg_attrdef d
WHERE   n.oid = c.relnamespace AND
   c.oid = a.attrelid AND
   a.attrelid = d.adrelid AND
   a.attnum = d.adnum AND
   d.adsrc ~ $$val\(\('[^']*'::text\)::regclass$$;
```

然后，对数据库运行该查询以找出需要哪些调整，例如对数据库 `db1`：

```
psql -t -f fixseq.sql db1
```

这会显示把数据库转换为较新的基于 OID 的表示所需的 `ALTER TABLE` 命令。如果这些命令看起来合理，运行以下命令更新数据库：

```
psql -t -f fixseq.sql db1 | psql -e db1
```

此过程必须在每个要更新的数据库中重复进行。

- 在 psql 中，把未加引号的 `\{digit}+` 序列当作八进制（Bruce）

  在以前的版本中，`\{digit}+` 序列被当作十进制，只有 `\0{digit}+` 被当作八进制。此变更是为了保持一致。

- 移除前缀和后缀 `%` 和 `^` 操作符的语法产生式（Tom）

  它们从未被记录，而且使取模操作符（`%`）与负数一起使用变得复杂。

- 让多边形的 `&<` 和 `&>` 与 box 的"over"操作符一致（Tom）

- `CREATE LANGUAGE` 可以忽略提供的参数而采用 `pg_pltemplate` 中的信息（Tom）

  定义了新系统目录 `pg_pltemplate`，携带过程语言首选定义的信息（例如是否有验证器函数）。当该目录中存在正要创建的语言的条目时，`CREATE LANGUAGE` 将忽略除语言名之外的所有参数而改用目录信息。采取此措施是因为旧转储文件装载过时语言定义的问题日益增多。从 8.1 起，pg_dump 将把过程语言定义只转储为 `CREATE LANGUAGE` *`name`*，依赖装载时存在模板条目。我们预期这将是更面向未来的表示。

- 让 `pg_cancel_backend(int)` 返回 `boolean` 而不是 `integer`（Neil）

- 一些用户在向 8.1.X 装载 UTF-8 数据时遇到问题。这是因为以前的版本允许把无效 UTF-8 字节序列输入数据库，而此版本只正确接受有效的 UTF-8 序列。纠正转储文件的一种方法是运行命令 `iconv -c -f UTF-8 -t UTF-8 -o cleanfile.sql dumpfile.sql`。`-c` 选项移除无效字符序列。两个文件的 diff 会显示无效的序列。`iconv` 把整个输入文件读入内存，因此可能需要用 split 把转储拆成多个较小的文件处理。

E.144.3. 附加变更

下面你将找到 PostgreSQL 8.1 与上一个主版本之间额外变更的详细说明。

E.144.3.1. 性能改进

- 改进 GiST 和 R-tree 索引性能（Neil）

- 改进优化器，包括哈希连接的自动调整大小（Tom）

- 彻底改造若干领域的内部 API
- 把 WAL 记录的 CRC 从 64 位改为 32 位（Tom）

  我们认定计算 64 位 CRC 的额外代价显著，而可靠性收益太小不值得。
- 防止在 WAL 页中写入大的空隙（Tom）
- 改进 SMP 机器（特别是 Opteron）上的自旋锁行为（Tom）
- 允许多列索引使用非连续的索引列（Tom）

  例如，这允许 a,b,c 列上的索引用于 `WHERE a = 4 and c = 10` 的查询。
- 跳过 `CREATE TABLE AS` / `SELECT INTO` 的 WAL 日志（Simon）

  由于 `CREATE TABLE AS` 期间的崩溃会导致表在恢复期间被删除，没有理由在装载表时写 WAL 日志。（但如果启用了 WAL 归档，仍会记录。）
- 允许并发 GiST 索引访问（Teodor、Oleg）
- 新增配置参数 `full_page_writes`，控制向 WAL 写整页（Bruce）

  为防止部分磁盘写损坏数据库，PostgreSQL 在检查点后首次修改每个数据库磁盘页时把它的完整副本写入 WAL。此选项为提高速度关闭该功能。在使用电池供电的磁盘缓存（不会发生部分页写）时可以安全使用。
- 当 `wal_sync_method` 使用 `O_SYNC` 时，如果可用则使用 `O_DIRECT`（Itagaki Takahiro）

  `O_DIRECT` 使磁盘写绕过内核缓存，对 WAL 写入而言这改进了性能。
- 改进 `COPY FROM` 的性能（Alon Goldshuv）

  这是通过以更大的块读取 `COPY` 输入而不是逐字符读取实现的。
- 改进 `COUNT()`、`SUM`、`AVG()`、`STDDEV()` 和 `VARIANCE()` 的性能（Neil、Tom）

E.144.3.2. 服务器变更

- 防止事务 ID（XID）回卷导致的问题（Tom）

  当事务计数器接近回卷点时，服务器现在会发出警告。如果计数器过于接近回卷，服务器将停止接受查询。这确保在执行所需清理之前数据不会丢失。
- 修复 OID 计数器回卷后对象 ID（OID）与现有系统对象冲突的问题（Tom）
- 新增关于 `VACUUM` 期间需要增大 `max_fsm_relations` 和 `max_fsm_pages` 的警告（Ron Mayer）
- 新增 `temp_buffers` 配置参数，允许用户确定临时表访问的本地缓冲区大小（Tom）
- 为 `pg_stat_activity` 新增会话开始时间和客户端 IP 地址（Magnus）
- 为位图扫描调整 `pg_stat` 视图（Tom）

  某些字段的含义略有变化。
- 增强 `pg_locks` 视图（Tom）

- 记录客户端 PREPARE 和 EXECUTE 的查询（Simon）
- 允许在 postgresql.conf 中指定 Kerberos 名和用户名的大小写敏感性（Magnus）
- 新增配置参数 krb_server_hostname，使服务器主机名可以被指定为服务主体的一部分（Todd Kover）

  如果未设置，可以使用与 keytab 中条目匹配的任何服务主体。这是此版本新的 Kerberos 匹配行为。
- 为 log_line_prefix 新增毫秒时间戳（%m）和远程主机（%h）选项（Ed L.）
- 为 GiST 索引新增 WAL 日志（Teodor、Oleg）

  GiST 索引现在对崩溃和时间点恢复是安全的。
- 执行 pg_stop_backup() 时移除旧的 *.backup 文件（Bruce）

  这防止 pg_xlog/ 中存在大量 *.backup 文件。
- 新增控制 TCP/IP keep-alive 空闲时间、间隔和次数的配置参数（Oliver Jowett）

  可以更改这些值以更快检测丢失的客户端连接。
- 新增逐用户和逐数据库的连接限制（Petr Jelinek）

  使用 ALTER USER 和 ALTER DATABASE，现在可以对以特定用户身份或到特定数据库并发连接的最大会话数施加限制。把限制设为零禁用用户或数据库连接。
- 允许 64 位机器上超过 2GB 的共享内存和逐后端工作内存（Koichi Suzuki）
- 新系统目录 pg_pltemplate 允许覆盖转储文件中过时的过程语言定义（Tom）

E.144.3.3. 查询变更

- 新增临时视图（Koju Iijima、Neil）
- 修复没有任何聚合函数或 GROUP BY 的 HAVING，使查询返回单个组（Tom）

  此前这种情形会把 HAVING 子句当作 WHERE 子句。这不符合规范。
- 新增 USING 子句，允许为 DELETE 指定额外的表（Euler Taveira de Oliveira、Neil）

  在以前的版本中，没有明确的方法为 DELETE 语句指定用于连接的额外表。UPDATE 已有用于此目的的 FROM 子句。
- 新增对后端和 ecpg 字符串中 \x 十六进制转义的支持（Bruce）

  这就像标准的 C \x 转义语法。八进制转义此前已支持。
- 新增 BETWEEN SYMMETRIC 查询语法（Pavel Stehule）

  此特性允许 BETWEEN 比较而不要求第一个值小于第二个值。例如 2 BETWEEN [ASYMMETRIC] 3 AND 1 返回假，而 2 BETWEEN SYMMETRIC 3 AND 1 返回真。BETWEEN ASYMMETRIC 此前已支持。
- 为 SELECT ... FOR UPDATE/SHARE 新增 NOWAIT 选项（Hans-Juergen Schoenig）

虽然 statement_timeout 配置参数允许取消耗时超过一定时间的查询，但 NOWAIT 选项允许在 SELECT ... FOR UPDATE/SHARE 命令不能立即获得行锁时立刻取消查询。

E.144.3.4. 对象操作变更

- 跟踪共享对象的依赖（Alvaro）

  PostgreSQL 允许全局表（用户、数据库、表空间）引用多个数据库中的信息。此新增为全局表添加了依赖信息，因此例如用户所有权可以跨数据库跟踪，
  在任何数据库中拥有某些东西的用户不能再被删除。数据库本地对象的依赖跟踪此前已存在。

- 允许对象所有者执行有限的 ALTER OWNER 命令（Stephen Frost）

  以前的版本只允许超级用户更改对象所有者。现在，
  如果执行命令的用户拥有该对象并能以新所有者身份创建它（即该用户是新所有角色的成员且该角色具有重新创建对象所需的 CREATE 权限），就可以转移所有权。

- 为某些对象类型（表、函数、类型）新增 ALTER 对象 SET SCHEMA 能力（Bernd Helmle）

  这允许把对象移到不同的模式。

- 新增 ALTER TABLE ENABLE/DISABLE TRIGGER 来禁用触发器（Satoshi Nagayasu）

E.144.3.5. 工具命令变更

- 允许 TRUNCATE 用单条命令截断多个表（Alvaro）

  由于参照完整性检查，不允许截断作为参照完整性约束一部分的表。使用此新功能，
  如果参照完整性约束涉及的两个表在同一条 TRUNCATE 命令中被截断，就可以用 TRUNCATE 截断这样的表。

- 在 COPY CSV 模式中正确处理回车和换行（Andrew）

  在 8.0 版本中，CSV COPY TO 中的回车和换行的处理方式不一致。（这记录在 TODO 列表中。）

- 新增 COPY WITH CSV HEADER，允许 COPY 的第一行为头行（Andrew）

  这允许处理把列名放在数据文件首行的常见 CSV 用法。对 COPY TO，首行包含列名；对 COPY FROM，首行被忽略。

- 在 Windows 上，EXPLAIN ANALYZE 显示更好的亚秒精度（Magnus）

- 为 EXPLAIN ANALYZE 新增触发器时长显示（Tom）

  以前的版本把触发器执行时间计入总执行时间，但没有单独显示。
  现在可以看到每个触发器花费多少时间。

- 新增对 COPY 中 \x 十六进制转义的支持（Sergey Ten）

  以前的版本只支持八进制转义。

- 让 SHOW ALL 包含变量描述（Matthias Schmidt）

  SHOW varname 仍然只显示变量的值，不包含描述。

- 让 initdb 创建名为 postgres 的新标准数据库，并把工具转换为用 postgres 而不是 template1 做标准查找（Dave）

  在以前的版本中，template1 既被用作 createuser 等工具的默认连接，又被用作新数据库的模板。这导致 CREATE DATABASE 有时失败，因为如果其他人在模板数据库中就无法创建新数据库。有了此变更，默认连接数据库现在是 postgres，意味着 CREATE DATABASE 期间有人正在使用 template1 的可能性小得多。

- 通过把 /contrib/reindexdb 移入服务器来创建新的 reindexdb 命令行工具（Euler Taveira de Oliveira）

E.144.3.6. 数据类型和函数变更

- 为数组类型新增 MAX() 和 MIN() 聚合（Koju Iijima）

- 修复 to_date() 和 to_timestamp() 在同时使用 CC 和 YY 字段时合理行事（Karel Zak）

  如果格式说明包含 CC 而年规格为 YYY 或更长，忽略 CC。如果年规格为 YY 或更短，把 CC 解释为上一个世纪。

- 新增 md5(bytea)（Abhijit Menon-Sen）

  md5(text) 此前已存在。

- 新增对 numeric ^ numeric 的支持，基于 power(numeric, numeric)

  该函数此前已存在，但没有为它分配操作符。

- 修复 NUMERIC 取模，在计算期间正确截断商（Bruce）

  在以前的版本中，大值的取模有时因商的四舍五入而返回负结果。

- 新增函数 lastval()（Dennis Björklund）

  lastval() 是 currval() 的简化版本。它根据当前会话最近执行的 nextval() 或 setval() 调用自动确定正确的序列名。

- 新增 to_timestamp(DOUBLE PRECISION)（Michael Glaesemann）

  把自 1970 年以来的 Unix 秒数转换为 TIMESTAMP WITH TIMEZONE。

- 新增 pg_postmaster_start_time() 函数（Euler Taveira de Oliveira、Matthias Schmidt）

- 允许在 AT TIME ZONE 中完整使用时区名，而不只是以前可用的短列表（Magnus）

  此前 AT TIME ZONE 只支持预定义的时区名列表。现在可以使用任何受支持的时区名，例如：

  ```
  SELECT CURRENT_TIMESTAMP AT TIME ZONE 'Europe/London';
  ```

  在上面的查询中，使用的时区根据所提供日期当时有效的夏令时规则调整。

- 新增 GREATEST() 和 LEAST() 可变参数函数（Pavel Stehule）

  这些函数接受可变数量的参数，返回参数中的最大或最小值。

- 新增 pg_column_size()（Mark Kirkwood）

它返回列的存储大小，该列可能被压缩过。

- 新增 `regexp_replace()`（Atsushi Ogawa）

  这允许像 sed 那样做正则表达式替换。可选的标志参数允许选择全局（全部替换）和不区分大小写模式。

- 修复 interval 的除法和乘法（Bruce）

  以前的版本有时返回没有道理的结果，例如 `'4 months'::interval / 5` 返回 `'1 mon -6 days'`。

- 修复 timestamp、time 和 interval 输出的舍入行为（Tom）

  这修复了秒字段会显示为 `60` 而不是进位到更高字段的一些情形。

- 为 `interval` 类型新增单独的天字段，使一天间隔可与 24 小时间隔区分（Michael Glaesemann）

  包含夏令时调整的天不是 24 小时长，而通常是 23 或 25 小时。此变更创建了“多少天”间隔与“多少小时”间隔之间的概念区别。向时间戳加 `1 day` 现在即使中间发生夏令时调整，也在第二天给出相同的本地时间，而加 `24 hours` 在这种情况下会给出不同的本地时间。例如，按美国 DST 规则：

  ```
  '2005-04-03 00:00:00-05' + '1 day' = '2005-04-04 00:00:00-04'
  '2005-04-03 00:00:00-05' + '24 hours' = '2005-04-04 01:00:00-04'
  ```

- 新增 `justify_days()` 和 `justify_hours()`（Michael Glaesemann）

  这两个函数分别把天调整为适当的整月数加天数，把小时调整为适当的整天数加小时数。

- 把 `/contrib/dbsize` 移入后端，并重命名其中一些函数（Dave Page、Andreas Pflug）

  - `pg_tablespace_size()`
  - `pg_database_size()`
  - `pg_relation_size()`
  - `pg_total_relation_size()`
  - `pg_size_pretty()`

  `pg_total_relation_size()` 包括索引和 TOAST 表。

- 新增对集群目录只读文件访问的函数（Dave Page、Andreas Pflug）

  - `pg_stat_file()`
  - `pg_read_file()`
  - `pg_ls_dir()`

- 新增 `pg_reload_conf()`，强制重新装载配置文件（Dave Page、Andreas Pflug）

- 新增 `pg_rotate_logfile()`，强制轮转服务器日志文件（Dave Page、Andreas Pflug）

- 更改 `pg_stat_*` 视图以包含 TOAST 表（Tom）

E.144.3.7. 编码和区域设置变更

- 重命名一些编码使其更一致并遵循国际标准（Bruce）
  - `UNICODE` 现在是 `UTF8`
  - `ALT` 现在是 `WIN866`
  - `WIN` 现在是 `WIN1251`
  - `TCVN` 现在是 `WIN1258`

  原名仍然有效。
- 新增对 `WIN1252` 编码的支持（Roland Volkmann）
- 新增对四字节 `UTF8` 字符的支持（John Hansen）

  此前只支持一、二、三字节的 `UTF8` 字符。这对支持某些中文字符集特别重要。
- 允许 `EUC_JP` 和 `SJIS` 之间的直接转换以改进性能（Atsushi Ogawa）
- 允许 UTF8 编码在 Windows 上工作（Magnus）

  这是通过把 UTF8 映射到 Windows 本机 UTF16 实现完成的。

E.144.3.8. 常规服务器端语言变更

- 修复 `ALTER LANGUAGE RENAME`（Sergey Yatskevich）
- 允许通过 `ALTER FUNCTION` 修改函数特性，如严格性和易变性（Neil）
- 把函数参数的最大数量提高到 100（Tom）
- 允许 SQL 和 PL/pgSQL 函数使用 `OUT` 和 `INOUT` 参数（Tom）

  `OUT` 是函数返回值的另一种方式。除了使用 `RETURN`，还可以通过给声明为 `OUT` 或 `INOUT` 的参数赋值来返回值。这在某些情形下书写更简单，特别是需要返回多个值时。虽然以前的版本也可以从函数返回多个值，但这大幅简化了该过程。
  （该特性将在未来版本中扩展到其他服务器端语言。）
- 把语言处理函数移入 `pg_catalog` 模式

  这使按需删除 public 模式变得更容易。
- 为 SPI 新增 `SPI_getnspname()`（Neil）

E.144.3.9. PL/pgSQL 服务器端语言变更

- 彻底改造 PL/pgSQL 函数的内存管理（Neil）

  每个函数的分析树现在存储在单独的内存上下文中。这使该内存在不再需要时可以轻松回收。
- 在 `CREATE FUNCTION` 时检查函数语法，而不是在运行时（Neil）

此前，大多数语法错误只在函数执行时才被报告。

- 允许 `OPEN` 打开非 `SELECT` 查询，如 `EXPLAIN` 和 `SHOW`（Tom）
- 不再要求函数发出 `RETURN` 语句（Tom）

  这是新增的 `OUT` 和 `INOUT` 功能的副产品。当不需要提供函数返回值时可以省略 `RETURN`。
- 为 PL/pgSQL 的 `EXECUTE` 语句新增可选的 `INTO` 子句（Pavel Stehule、Neil）
- 让 `CREATE TABLE AS` 设置 `ROW_COUNT`（Tom）
- 定义 `SQLSTATE` 和 `SQLERRM`，返回当前异常的 `SQLSTATE` 和错误消息（Pavel Stehule、Neil）

  这些变量只在异常块内定义。
- 允许 `RAISE` 语句的参数为表达式（Pavel Stehule、Neil）
- 新增循环 `CONTINUE` 语句（Pavel Stehule、Neil）
- 允许块和循环标签（Pavel Stehule）

E.144.3.10. PL/Perl 服务器端语言变更

- 允许高效返回大结果集（Abhijit Menon-Sen）

  这允许函数使用 `return_next()` 来避免在内存中构建整个结果集。
- 允许逐行获取查询结果（Abhijit Menon-Sen）

  这允许函数使用 `spi_query()` 和 `spi_fetchrow()` 来避免在内存中累积整个结果集。
- 如果服务器编码为 `UTF8`，强制 PL/Perl 把字符串当作 `UTF8` 处理（David Kamholz）
- 为 PL/Perl 新增验证器函数（Andrew）

  这使语法错误可以在定义时而不是执行时报告。
- 当函数返回数组类型时允许 PL/Perl 返回 Perl 数组（Andrew）

  这基本上把 PostgreSQL 数组映射到 Perl 数组。
- 允许 Perl 非致命警告生成 `NOTICE` 消息（Andrew）
- 允许启用 Perl 的 `strict` 模式（Andrew）

E.144.3.11. psql Changes

- 新增 `\set ON_ERROR_ROLLBACK`，允许事务中的语句出错而不影响事务的其余部分（Greg Sabino Mullane）

  这基本上是通过把每条语句包装在子事务中实现的。
- 新增对 psql 变量中 `\x` 十六进制字符串的支持（Bruce）

  八进制转义此前已支持。

- 新增对 `troff -ms` 输出格式的支持（Roger Leigh）
- 允许历史文件的位置由 `HISTFILE` 控制（Andreas Seltenreich）

  这允许配置逐数据库的历史存储。
- 防止 `\x`（扩展模式）影响 `\d tablename` 的输出（Neil）
- 为 psql 新增 `-L` 选项以记录会话（Lorne Sunley）

  添加此选项是因为某些操作系统没有简单的命令行活动记录功能。
- 让 `\d` 显示索引的表空间（Qingqing Zhou）
- 允许 psql 帮助（`\h`）对正确的帮助信息做最佳猜测（Greg Sabino Mullane）

  这使用户只需在语法错误的查询前加上 `\h` 就能获得支持语法的帮助。此前必须删除命令名之外的任何附加查询文本才能使用 `\h`。
- 新增 `\pset numericlocale`，允许以区域感知的格式输出数字（Eugen Nedelcu）

  例如，使用 `C` 区域设置时 `100000` 会输出为 `100,000.0`，而欧洲区域设置可能输出为 `100.000,0`。
- 让启动横幅在服务器版本号与 psql 版本号不同时同时显示两者（Bruce）

  此外，如果服务器和 psql 来自不同的主版本，会显示警告。

E.144.3.12. pg_dump Changes

- 为 pg_restore 新增 `-n` / `--schema` 开关（Richard van den Berg）

  这允许只恢复指定模式中的对象。
- 允许 pg_dump 即使在文本模式也转储大对象（Tom）

  有了此变更，大对象现在总是被转储；以前的 `-b` 开关是无操作。
- 允许 pg_dump 转储大对象的一致快照（Tom）
- 转储大对象的注释（Tom）
- 为 pg_dump 新增 `--encoding`（Magnus Hagander）

  这允许以不同于服务器编码的编码转储数据库。
  当把转储传送到不同编码的机器时这很有价值。
- 过程语言依靠 `pg_pltemplate`（Tom）

  如果过程语言的调用处理器在 `pg_catalog` 模式中，pg_dump 不转储该处理器。而是只用 `CREATE LANGUAGE` *`name`* 转储该语言，依赖 `pg_pltemplate` 目录在装载时提供该语言的创建参数。

E.144.3.13. libpq Changes

- 新增 `PGPASSFILE` 环境变量，指定口令文件的文件名（Andrew）

- 新增 `lo_create()`，类似于 `lo_creat()` 但允许指定大对象的 OID（Tom）
- 让 libpq 在 `malloc()` 失败时一致地向客户端应用程序返回错误（Neil）

E.144.3.14. 源代码变更

- 修复 pgxs 以支持针对重新定位的安装构建
- 新增使用 Intel 编译器的 Itanium 处理器自旋锁支持（Vikram Kalsi）
- 新增 Windows 的 Kerberos 5 支持（Magnus）
- 新增中文 FAQ（laser@pgsqldb.com）
- 把 Rendezvous 改名为 Bonjour 以与 OS/X 特性改名一致（Bruce）
- 新增 Darwin 上对 `fsync_writethrough` 的支持（Chris Campbell）
- 精简服务器、优化器和锁系统内部的信息传递（Tom）
- 允许用 MSVC 编译 pg_config（Andrew）

  这是用 MSVC 构建 DBD::Pg 所必需的。
- 移除对 Kerberos V4 的支持（Magnus）

  Kerberos 4 有安全漏洞且不再维护。
- 代码清理（由 EnterpriseDB 执行 Coverity 静态分析）
- 修改 `postgresql.conf`，使用文档默认的 `on/off` 而不是 `true/false`（Bruce）
- 增强 pg_config，使其能报告更多构建期值（Tom）
- 允许在 Windows 上把 libpq 构建为线程安全（Dave Page）
- 允许在 Windows 上使用 IPv6 连接（Andrew）
- 新增关于 I/O 子系统可靠性的服务器管理文档（Bruce）
- 把私有声明从 `gist.h` 移到 `gist_private.h`（Neil）

  在以前的版本中，`gist.h` 既包含公共 GiST API（供 GiST 索引实现的作者使用），也包含 GiST 自身实现使用的一些私有声明。后者已被移到单独的文件 `gist_private.h`。大多数 GiST 索引实现应不受影响。
- 彻底改造 GiST 内存管理（Neil）

  GiST 方法现在总是在短生命周期的内存上下文中调用。因此，通过 `palloc()` 分配的内存将被自动回收，GiST 索引实现不需要通过 `pfree()` 手动释放分配的内存。

E.144.3.15. 附加模块变更

- 新增 `/contrib/pg_buffercache` contrib 模块（Mark Kirkwood）

  它显示缓冲区缓存的内容，用于调试和性能调优。

- 移除 /contrib/array，因为它已过时（Tom）
- 清理 /contrib/lo 模块（Tom）
- 把 /contrib/findoidjoins 移到 /src/tools（Tom）
- 从 /contrib/cube 移除 <<、>>、&< 和 &> 操作符

  这些操作符没有用处。
- 改进 /contrib/btree_gist（Janko Richter）
- 改进 /contrib/pgbench（Tomoaki Sato、Tatsuo）

  现在有一套用用户提供的 SQL 命令脚本进行测试的设施，而不是只有硬编码的命令序列。
- 改进 /contrib/pgcrypto（Marko Kreen）
  - 实现 OpenPGP 对称密钥和公钥加密

    支持 RSA 和 Elgamal 两种公钥算法。
  - 独立构建：包含 SHA256/384/512 哈希、Fortuna PRNG
  - OpenSSL 构建：支持 3DES，OpenSSL < 0.9.7 时使用内建 AES
  - 从 configure 结果获取构建参数（OpenSSL、zlib）

    不再需要编辑 Makefile。
  - 移除对 libmhash 和 libmcrypt 的支持

E.145. 版本 8.0.26

发行日期
2010-10-04

本次发布包含对 8.0.25 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

这预计是 PostgreSQL 8.0.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.145.1. 迁移到版本 8.0.26

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅第 E.149 节。

E.145.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 SECURITY DEFINER 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一

点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。

- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）
- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）
- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）
- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

E.146. 版本 8.0.25

发行日期
2010-05-17

本次发布包含对 8.0.24 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 8.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.146.1. 迁移到版本 8.0.25

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅第 E.149 节。

E.146.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 防止 psql 在展开引用自身的变量时无限递归（Tom）

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

E.147. 版本 8.0.24

发行日期
2010-03-15

本次发布包含对 8.0.23 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 8.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.147.1. 迁移到版本 8.0.24

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅第 E.149 节。

E.147.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）

- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 -1 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）

- 修复若干病态缓慢的正则表达式匹配情形（Tom）

- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）
- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）
- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）
- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.148. 版本 8.0.23

发行日期
2009-12-14

本次发布包含对 8.0.22 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.148.1. 迁移到版本 8.0.23

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅第 E.149 节。

E.148.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

这可能发生在规划器对哈希连接结果大小的估计极大时。

- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。
- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。
- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.149. 版本 8.0.22

发行日期
2009-09-09

本次发布包含对 8.0.21 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.149.1. 迁移到版本 8.0.22

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.0.22 后对它们执行 `REINDEX`。另外，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.149.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）
- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）
- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

  它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.99999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）

- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）

- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.150. 版本 8.0.21

发行日期
2009-03-16

本次发布包含对 8.0.20 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.150.1. 迁移到版本 8.0.21

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.150.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，

解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）
- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.151. 版本 8.0.20

发行日期
2009-02-02

本次发布包含对 8.0.19 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.151.1. 迁移到版本 8.0.20

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.151.2. 变更

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.152. 版本 8.0.19

发行日期
2008-11-03

本次发布包含对 8.0.18 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.152.1. 迁移到版本 8.0.19

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.152.2. 变更

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
  修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）
- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）
- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。

- 修复 ecpg 对 `CREATE USER` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.153. 版本 8.0.18

发行日期
2008-09-22

本次发布包含对 8.0.17 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.153.1. 迁移到版本 8.0.18

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.153.2. 变更

- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

  这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。

- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  `ALTER COLUMN TYPE` 之后重用先前缓存的计划可能产生这种情形。该检查防范随之而来的数据损坏和/或崩溃。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）
- 修复 PL/Python 使其与 Python 2.5 一起工作

  这是 8.2 开发周期中所做修复的向后移植。

- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）
- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）
- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.154. 版本 8.0.17

发行日期
2008-06-12

此发行版在 8.0.16 的基础上包含一个严重的缺陷修复。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.154.1. 迁移到版本 8.0.17

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.154.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

E.155. 版本 8.0.16

发行日期

从未发布

本次发布包含对 8.0.15 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.155.1. 迁移到版本 8.0.16

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.155.2. 变更

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。

- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。

- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴、阿根廷/圣路易斯和智利的 DST 法规变更）

- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）

- 修复 `contrib/xml2` 的 `xpath_table()` 函数在输入查询返回 NULL 值时的核心转储（Tom）

- 修复 `contrib/xml2` 的 makefile，使其不覆盖 `CFLAGS`（Tom）

- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）

- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。

- 修复 `ALTER OWNER` 某些变体中的“无法识别的节点类型”错误（Tom）

- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 `pg_ctl start -w` 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。

- 使用 `-fwrapv` 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。

- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

  显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.156. 版本 8.0.15

发行日期
2008-01-07

此发行版包含 8.0.14 之后的多种修复，包括对重要安全问题的修复。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

这是 PostgreSQL 社区为 Windows 制作二进制包的最后一个 8.0.X 版本。建议 Windows 用户迁移到 8.2.X 或更高版本，因为 8.2.X 中有一些不便于向后移植的 Windows 特定修复。8.0.X 在其他平台上将继续受支持。

E.156.1. 迁移到版本 8.0.15

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.156.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  8.0.14 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 更新时区数据文件到 tzdata 发行版 2007k（特别是阿根廷最近的变更）（Tom）

- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）

- 保留被 `ALTER TABLE ... ALTER COLUMN TYPE` 重建的索引的表空间（Tom）

- 让归档恢复总是开始新的 WAL 时间线，而不是只在使用恢复停止时间时（Simon）

  这避免了试图覆盖最后一个 WAL 段现有归档副本的极端情形风险，而且看起来比原来的定义更简单干净。

- 当表太小而不适用时，让 `VACUUM` 不使用全部 `maintenance_work_mem`（Alvaro）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）

- 修复 PL/Perl 在平台 Perl 把 `bool` 类型定义为 `int` 而非 `char` 时的问题（Tom）

  虽然理论上这可能在任何地方发生，但没有标准 Perl 构建这样做.....直到 Mac OS X 10.5。

- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）

- 修复 pg_dump，使其正确处理默认表达式与父表不同的继承子表（Tom）

- ecpg 解析器修复（Michael）

- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）

- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）

- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）

- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.157. 版本 8.0.14

发行日期
2007-09-17

本次发布包含对 8.0.13 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.157.1. 迁移到版本 8.0.14

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.157.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复日志，使使用 syslogger 进程时日志消息永不交错（Andrew）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 修复某些外键极端情形的错误处理（Tom）
- 防止 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 更新时区数据库规则，特别是新西兰即将到来的变更（Tom）
- Windows 套接字改进（Magnus）
- 因可能的编码不匹配，在 Windows 上抑制日志时间戳中的时区名（`%Z`）（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.158. 版本 8.0.13

发行日期
2007-04-23

此发行版包含 8.0.12 之后的多种修复，包括一个安全修复。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.158.1. 迁移到版本 8.0.13

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.158.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 修复哈希索引扩容期间的 PANIC（8.0.10 中引入的缺陷）（Tom）
- 修复 POSIX 风格时区规范以遵循新的美国 DST 规则（Tom）

E.159. 版本 8.0.12

发行日期
2007-02-07

此发行版包含 8.0.11 之后的一个修复。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.159.1. 迁移到版本 8.0.12

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.159.2. 变更

- 移除约束和函数索引中对类型长度的过度严格检查（Tom）

E.160. 版本 8.0.11

发行日期
2007-02-05

此发行版包含 8.0.10 之后的多种修复，包括一个安全修复。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.160.1. 迁移到版本 8.0.11

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.160.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  这些漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，以及更改表列的数据类型（CVE-2007-0555、CVE-2007-0556）。这些错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.161. 版本 8.0.10

发行日期
2007-01-08

本次发布包含对 8.0.9 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.161.1. 迁移到版本 8.0.10

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.161.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）

- 修复 `VACUUM` 截断跨越 GB 边界的大关系时的竞态条件（Tom）

- 修复影响多 GB 哈希索引的缺陷（Tom）

- 修复 Windows 信号处理中可能的死锁（Teodor）

- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）

- 修复 ecpg 连接期间的内存泄漏（Michael）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。
- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。
- 更新时区数据库

  这特别影响澳大利亚和加拿大的夏令时规则。

E.162. 版本 8.0.9

发行日期
2006-10-16

本次发布包含对 8.0.8 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.162.1. 迁移到版本 8.0.9

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.162.2. 变更

- 修复在规则 WHERE 表达式中引用 `NEW` 行值时的崩溃（Tom）
- 修复把无类型字面值当作 ANYARRAY 时的核心转储
- 修复查询包含返回多行的 SQL 函数时 AFTER 触发器的错误处理（Tom）
- 修复 `ALTER TABLE ... TYPE`，对 `USING` 子句重新检查 `NOT NULL`（Tom）
- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。
- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- ecpg 的大量健壮性修复（Joachim Wieland）
- 修复 /contrib/dbmirror 中的反斜杠转义
- 修复 Win32 上统计收集的不稳定（Tom、Andrew）
- AIX 和 Intel 编译器的修复（Tom）

E.163. 版本 8.0.8

发行日期
2006-05-23

此发行版包含 8.0.7 之后的多种修复，包括针对极其严重安全问题的补丁。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.163.1. 迁移到版本 8.0.8

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.163.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 修复有时导致 OR 索引扫描漏掉本应返回的行的缺陷

- 修复 btree 索引被截断情形的 WAL 重放
- 修复涉及 | 的模式的 SIMILAR TO（Tom）
- 修复 SELECT INTO 和 CREATE TABLE AS，在默认表空间而不是基目录中创建表（Kris Jurka）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 修复 Intel Mac 上的 Bonjour（Ashley Clark）
- 修复各种小的内存泄漏
- 修复某些 Win32 系统上的口令提示问题（Robert Kinberg）

E.164. 版本 8.0.7

发行日期
2006-02-14

本次发布包含对 8.0.6 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.164.1. 迁移到版本 8.0.7

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅第 E.165 节。

E.164.2. 变更

- 修复 SET SESSION AUTHORIZATION 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 8.0.4、7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 和 pg_subtrans 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复缓存失效消息恰好在错误时刻到达时可能导致崩溃的情形（Tom）
- 为预备语句中的 UNKNOWN 参数正确检查 DOMAIN 约束（Neil）
- 确保 ALTER COLUMN TYPE 按正确顺序处理 FOREIGN KEY、UNIQUE 和 PRIMARY KEY 约束（Nakano Yoshihisa）
- 修复以允许恢复有对自定义操作符或操作符类的跨模式引用的转储（Tom）
- 允许 pg_restore 在 COPY 失败后正确继续；此前它试图把剩余的 COPY 数据当作 SQL 命令（Stephen Frost）

- 修复未指定数据目录时 pg_ctl 的 `unregister` 崩溃（Magnus）
- 修复 AMD64 和 PPC 上 ecpg 的崩溃（Neil）
- 修复 PL/python 参数传递期间发生错误时的正确恢复（Neil）
- 修复 PL/perl 在 Win32 上对区域设置的处理以与后端匹配（Andrew）
- 修复 Win32 上 `postgresql.conf` 中 `log_min_messages` 设为 `DEBUG3` 或更高时的崩溃（Bruce）
- 修复 Win32、Cygwin、OS X、AIX 的 pgxs `-L` 库路径指定（Bruce）
- 在检查 Win32 管理员权限时检查 SID 是否启用（Magnus）
- 正确拒绝超出范围的日期输入（Kris Jurka）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.165. 版本 8.0.6

发行日期
2006-01-09

本次发布包含对 8.0.5 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.165.1. 迁移到版本 8.0.6

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅第 E.168 节。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.165.2. 变更

- 修复 Windows 代码，使 ShmemBackendArray 没有更多空间时 postmaster 继续而不是退出（Magnus）

  如果过多连接请求几乎同时到达，以前的行为可能导致拒绝服务情形。这只适用于 Windows 移植。

- 修复 8.0 中引入的缺陷，它可能使 ReadBuffer 返回已使用过的页当作新页，可能导致最近提交的数据丢失（Tom）
- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）
- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 允许更灵活地重新定位安装目录（Tom）

  以前的版本只有在所有安装目录路径除最后一个组件外都相同时才支持重新定位。

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）

- 对返回 `RECORD` 的函数的各种修复（Tom）

- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.166. 版本 8.0.5

发行日期
2005-12-12

本次发布包含对 8.0.4 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.166.1. 迁移到版本 8.0.5

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅第 E.168 节。

E.166.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- 修复从错误恢复后 bgwriter 的问题（Tom）

  发现后台写入器在写错误后泄漏缓冲区固定。虽然本身不致命，但这可能导致后续 VACUUM 命令的神秘阻塞。

- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败

- `/contrib/ltree` 修复（Teodor）

- AIX 和 HPUX 编译修复（Tom）

- 在 Windows NO_SYSTEM_RESOURCES 错误后重试文件读写（Qingqing Zhou）

- 修复 `log_line_prefix` 包含 `%i` 时的间歇性失败
- 修复 Windows 上长脚本的 psql 性能问题（Merlin Moncure）
- 修复 `pg_group` 平面文件更新的缺失
- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。
- 把时区初始化推迟到 `postmaster.pid` 创建之后

  这避免让期待 pid 文件快速出现的启动脚本困惑。
- 防止表被删除后 pg_autovacuum 的核心转储
- 修复对子查询结果的整行引用（`foo.*`）的问题

E.167. 版本 8.0.4

发行日期
2005-10-04

本次发布包含对 8.0.3 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.167.1. 迁移到版本 8.0.4

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅第 E.168 节。

E.167.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。
- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

  在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。
- 在提交 `CREATE DATABASE` 前强制检查点

  这应能修复最近报告的 `CREATE DATABASE` 之后不久发生崩溃时“索引不是 btree”失败的报告。
- 修正 `COPY` 中只读事务测试的方向

  以前的代码禁止 `COPY TO`，而它应禁止的是 `COPY FROM`。
- 处理 `COPY` CSV 模式输入中连续的嵌入换行符

- 修复年末附近日期的 `date_trunc(week)`
- 修复只引用内侧关系的外连接 ON 子句的规划问题
- 对 `x FULL JOIN y ON true` 极端情形的进一步修复
- 修复对 `x IN (SELECT DISTINCT ...)` 及相关情形的过度优化
- 修复因考虑不周的“模糊”代价比较导致小 `LIMIT` 值查询的错误规划
- 让 `array_in` 和 `array_recv` 更严格地验证其 OID 参数
- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题
- 改进日期时间解析的健壮性
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- 改进 MIPS 和 M68K 自旋锁代码
- postmaster 启动期间不要试图打开超过 `max_files_per_process` 个文件
- 各种内存泄漏修复
- 各种可移植性改进
- 更新时区数据文件
- 改进 Windows 上 DLL 装载失败的处理
- 改进 Windows 上的随机数生成
- 让 `psql -f filename` 在打开文件失败时返回非零退出码
- 更改 pg_dump 以更可靠地处理继承的检查约束
- 修复 Windows 上 pg_restore 的口令提示
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`
- 修复 PL/Perl 的 `%_SHARED`，使其真正共享
- 修复 `contrib/pg_autovacuum`，允许超过 2000 秒的休眠间隔
- 更新 `contrib/tsearch2` 以使用当前的 Snowball 代码

E.168. 版本 8.0.3

发行日期
2005-05-09

此发行版包含 8.0.2 之后的多种修复，包括几个与安全相关的问题。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.168.1. 迁移到版本 8.0.3

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，它是处理在 8.0.X 系统目录初始内容中发现的两个重要安全问题的可能方法之一。用 8.0.3 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

较大的安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）

较小的问题是 `contrib/tsearch2` 模块创建了几个被错误声明为返回 `internal` 的函数，而它们并不接受 `internal` 参数。这破坏了所有使用 `internal` 参数的函数的类型安全。

强烈建议所有安装通过 initdb 或按照下述手工修复程序修复这些错误。这些错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行 7.4.8 发行说明中所示的相同手工修复程序。

E.168.2. 变更

- 更改编码函数签名以防止误用
- 更改 `contrib/tsearch2` 以避免不安全地使用 `INTERNAL` 函数结果
- 防范 `record_out` 的错误第二参数
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

  此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化

- 防止 `COALESCE(NULL,NULL)` 的崩溃
- 修复 libpq 的 Borland makefile
- 修复 `contrib/btree_gist` 的 `timetz` 类型（Teodor）
- 让 `pg_ctl` 检查在 `postmaster.pid` 中找到的 PID，看它是否仍是活进程
- 修复因添加转储时间戳导致的 `pg_dump`/`pg_restore` 问题
- 修复事务提交期间物化可保持游标与触发延迟触发器之间的相互作用
- 修复返回按引用传递数据类型的 SQL 函数中的内存泄漏

E.169. 版本 8.0.2

发行日期
2005-04-07

本次发布包含对 8.0.1 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.171 节。.

E.169.1. 迁移到版本 8.0.2

对于运行 8.0.* 的用户，不需要进行转储/恢复。此发行版更新了 PostgreSQL 库的主版本号，因此如果某些用户应用程序找不到正确编号的共享库，可能需要重新链接它们。

E.169.2. 变更

- 递增所有接口库的主版本号（Bruce）

  这本应在 8.0.0 中完成。需要它才能使 7.4.X 版本的 PostgreSQL 客户端应用程序（如 psql）能与 8.0.X 应用程序在同一台机器上使用。这可能需要重新链接使用这些库的用户应用程序。

- 新增仅 Windows 的 `wal_sync_method` 设置 `fsync_writethrough`（Magnus、Bruce）

  此设置使 PostgreSQL 在写 WAL 时穿透磁盘驱动器的任何写缓存。此行为以前称为 `fsync`，但因它在其他平台上的行为与 `fsync` 大不相同而被改名。

- 在 Windows 上启用 `wal_sync_method` 的 `open_datasync` 设置，并使其成为该平台的默认（Magnus、Bruce）

  由于默认不再是 `fsync_writethrough`，如果磁盘驱动器启用了写缓存，断电期间可能丢失数据。要在 Windows 上关闭写缓存，请在设备管理器中选择驱动器属性，然后选择策略。

- 新缓存管理算法 2Q 取代 ARC（Tom）

  这是为了避免 ARC 的一项待决美国专利。对某些负载，2Q 代码可能比 ARC 慢几个百分点。更好的缓存管理算法将出现在 8.1 中。

- 规划器调整以改进新建表上的行为（Tom）
- 允许 plpgsql 给初始为 `NULL` 的数组的元素赋值（Tom）

此前数组会保持 NULL，但现在它变成单元素数组。主 SQL 引擎在 8.0 中已改为以这种方式处理空数组值的 UPDATE，但 plpgsql 中的类似情形被遗漏了。

- 把 plpython 函数体中的 \r\n 和 \r 转换为 \n（Michael Fuhr）

  这防止在 Windows 或 Mac 客户端上编写 plpython 代码时出现语法错误。

- 允许 SPI 游标处理返回行的工具命令，如 EXPLAIN（Tom）
- 修复 ALTER TABLE SET WITHOUT OIDS 之后 CLUSTER 的失败（Tom）
- 降低 ALTER TABLE ADD COLUMN 的内存使用（Neil）
- 修复 ALTER LANGUAGE RENAME（Tom）
- 记录 pg_ctl 仅 Windows 的 register 和 unregister 选项（Magnus）
- 确保后端关闭期间完成的操作被统计收集器计数

  这有望解决关于 pg_autovacuum 不够频繁地清理系统目录的报告——它没有被通知后端退出期间临时表删除导致的目录删除。

- 把 Windows 上配置参数 log_destination 的默认改为 eventlog（Magnus）

  默认情况下，在 Windows 上运行的服务器现在将把日志输出发送到 Windows 事件记录器而不是标准错误。

- 使 Kerberos 认证在 Windows 上工作（Magnus）
- 允许没有 CREATEDB 权限标志的超级用户执行 ALTER DATABASE RENAME（Tom）
- 修改 CREATE 和 DROP DATABASE 的 WAL 日志条目，不指定绝对路径（Tom）

  这允许在数据库位置可能不同的另一台机器上做时间点恢复。注意 CREATE TABLESPACE 在这种情形下仍有风险。

- 修复后端带着创建过表并对其打开游标的未决事务退出时的崩溃（Tom）
- 修复 array_map()，使其能调用 PL 函数（Tom）
- 若干 contrib/tsearch2 和 contrib/btree_gist 修复（Teodor）
- 修复某些平台上一些 contrib/pgcrypto 函数的崩溃（Marko Kreen）
- 修复 contrib/intagg 在 64 位平台上的问题（Tom）
- 修复 ecpg 解析 CREATE 语句的缺陷（Michael）
- 绕过 powerpc 和 amd64 上的 gcc 缺陷，它导致 ecpg 中的问题（Christof Petig）
- 当区域设置为 C 时不使用区域感知版本的 upper()、lower() 和 initcap()（Bruce）

  这使这些函数能在区域设置为 C 时对非 7 位数据报错的平台上工作。

- 修复 quote_ident()，给匹配关键字的名称加引号（Tom）
- 修复 to_date() 在同时使用 CC 和 YY 字段时合理行事（Karel）

- 防止 `to_char(interval)` 在给定零月 interval 时失败（Tom）
- 修复 `date_trunc('week')` 返回错误星期的问题（Bruce）

  `date_trunc('week')` 在某些年份对一月的前几天返回错误的年份。
- 为 `INET` 数据类型的 `D` 类地址使用正确的默认掩码长度（Tom）

E.170. 版本 8.0.1

发行日期
2005-01-31

此发行版包含 8.0.0 之后的多种修复，包括几个与安全相关的问题。有关 8.0 主版本新特性的信息，请参阅第 E.171 节。

E.170.1. 迁移到版本 8.0.1

对于运行 8.0.0 的用户，不需要进行转储/恢复。

E.170.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。
- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。
- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 让 `ALTER TABLE ADD COLUMN` 在所有情形下强制执行域约束
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。
- 改进分组聚合查询的规划
- `ROLLBACK TO` *`savepoint`* 关闭自保存点以来创建的游标
- 修复 Windows 上后端栈大小不足的问题
- 在 Windows 上避免 SHGetSpecialFolderPath()（Magnus）

- 修复作为 Windows 服务运行 pg_autovacuum 的一些问题（Dave Page）
- pg_dump/pg_restore 的多个小缺陷修复
- 修复 ecpg 在 typedef 中使用命名结构时的段错误（Michael）

E.171. 版本 8.0

发行日期
2005-01-19

E.171.1. 概述

此发行版的主要变更：

Microsoft Windows Native Server

这是第一个作为服务器原生运行于 Microsoft Windows® 上的 PostgreSQL 版本。它可以作为 Windows 服务运行。此版本支持基于 NT 的 Windows 版本，如 Windows 2000 SP4、Windows XP 和 Windows 2003。更老的版本如 Windows 95、Windows 98 和 Windows ME 不受支持，因为这些操作系统没有支持 PostgreSQL 的基础设施。已创建单独的安装程序项目以简化在 Windows 上的安装——参见http://www.postgresql.org/ftp/win32/。

虽然在整个发布周期中都做了测试，但 Windows 移植没有 PostgreSQL 在 Unix 平台上多年生产环境使用的积累。因此应像对待新产品一样谨慎对待它。

以前的版本需要 Unix 仿真工具包 Cygwin 才能在 Windows 操作系统上运行服务器。PostgreSQL 在 Windows 上的本机客户端支持已有多年。

Savepoints

保存点允许中止事务的特定部分而不影响事务的其余部分。以前的版本没有这种能力；事务内的语句失败除了中止整个事务外无法恢复。
此特性对需要在复杂事务内进行错误恢复的应用程序作者很有价值。

Point-In-Time Recovery

在以前的版本中，除了从以前的备份恢复或使用备用复制服务器外，无法从磁盘驱动器故障中恢复。时间点恢复允许对服务器做连续备份。
你可以恢复到故障点或过去的某个事务。

Tablespaces

表空间允许管理员为各个表、索引和数据库的存储选择不同的文件系统。
这改进了性能和对磁盘空间使用的控制。以前的版本使用 initlocation 和手工符号链接管理来完成此类任务。

Improved Buffer Management, `CHECKPOINT`, `VACUUM`

此版本有更智能的缓冲区替换策略，能更好地利用可用共享缓冲区并改进性能。
清理和检查点的性能影响也减轻了。

Change Column Types

现在可以用 `ALTER TABLE` 更改列的数据类型。

New Perl Server-Side Language

新版本的 plperl 服务器端语言现在支持持久共享存储区、触发器、返回记录和记录数组以及访问数据库的 SPI 调用。

Comma-separated-value (CSV) support in `COPY`

`COPY` 现在可以读写逗号分隔值文件。它还能灵活解释非标准的引号和分隔字符。

E.171.2. 迁移到版本 8.0

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

- 在 `READ COMMITTED` 序列化模式下，易变函数现在能看到并发事务提交到函数内每条语句开始处为止的结果，而不是到调用函数的交互命令开始处为止。
- 声明为 `STABLE` 或 `IMMUTABLE` 的函数总是使用调用查询的快照，因此看不到调用查询开始之后采取的动作的效果，无论在其自身事务还是其他事务中。这样的函数也必须是只读的，即它不能使用 `SELECT` 之外的任何 SQL 命令。
- 非延迟的 `AFTER` 触发器现在在触发查询完成后立即触发，而不是在当前交互命令完成时。当触发查询发生在函数内时这就有区别：触发器在函数继续其下一个操作之前被调用。
- 服务器配置参数 `virtual_host` 和 `tcpip_socket` 已被更通用的参数 `listen_addresses` 取代。此外，服务器现在默认监听 `localhost`，这在许多场景中消除了对 `-i` postmaster 开关的需要。
- 服务器配置参数 `SortMem` 和 `VacuumMem` 已改名为 `work_mem` 和 `maintenance_work_mem`，以更好地反映其用途。原名在 `SET` 和 `SHOW` 中仍受支持。
- 服务器配置参数 `log_pid`、`log_timestamp` 和 `log_source_port` 已被更通用的参数 `log_line_prefix` 取代。
- 服务器配置参数 `syslog` 已被更合逻辑的 `log_destination` 变量取代，以控制日志输出目的地。
- 服务器配置参数 `log_statement` 已更改，可以有选择地只记录数据库修改或数据定义语句。服务器配置参数 `log_duration` 现在只在 `log_statement` 打印查询时打印。
- 服务器配置参数 `max_expr_depth` 已被 `max_stack_depth` 取代，后者测量物理栈大小而不是表达式嵌套深度。这有助于防止因递归函数导致的栈溢出造成的会话终止。
- `length()` 函数不再统计 `CHAR(n)` 值中的尾随空格。
- 把整数转换为 `BIT(N)` 现在选取整数最右边的 N 位，而不是像以前那样最左边 N 位。
- 更新 NULL 数组值的元素或切片现在产生非空数组结果，即只包含被赋值位置的数组。
- 数组输入值的语法检查已大幅收紧。
  以前在奇怪位置被允许并产生奇怪结果的垃圾现在会导致错误。空字符串元素值现在必须写成

""，而不是什么也不写。还更改了数组元素周围空白的行为：尾随空白现在被忽略，与前导空白（一直被忽略）对称。

- 整数算术运算中的溢出现在被检测并报告为错误。
- 与单字节 "char" 数据类型相关的算术操作符已被移除。
- extract() 函数（也称 date_part）现在为 BC 日期返回正确的年份。此前它返回比正确年份小一的值。该函数现在还为千年和世纪返回正确的值。
- CIDR 值现在必须使其未掩码的位为零。例如，我们不再允许 204.248.199.1/31 作为 CIDR 值。这类值本就不应被 PostgreSQL 接受，现在将被拒绝。
- EXECUTE 现在返回与被执行语句匹配的完成标签。
- psql 的 \copy 命令现在读写查询的 stdin/stdout，而不是 psql 的 stdin/stdout。以前的行为可通过新的 pstdin/pstdout 参数访问。
- JDBC 客户端接口已从核心发行版中移除，现托管于http://jdbc.postgresql.org。
- Tcl 客户端接口也被移除。现在有几个 Tcl 接口托管于http://gborg.postgresql.org。
- 服务器现在使用自己的时区数据库，而不是操作系统提供的。这将在所有平台上提供一致的行为。在大多数情况下，时区行为应没有明显差异，只是 SET/SHOW TimeZone 使用的时区名可能与你的平台提供的不同。
- Configure 的线程选项不再要求用户运行测试或编辑配置文件；线程选项现在自动检测。'
- 既然表空间已经实现，initlocation 已被移除。
- 用户定义 GiST 索引的 API 已更改。Union 和 PickSplit 方法现在传递指向特殊 GistEntryVector 结构的指针，而不是 bytea。

E.171.3. 废弃特性

PostgreSQL 行为的某些方面已被确定为次优。为了向后兼容，它们在 8.0 中没有被移除，但被认为已废弃，将在下一个主版本中移除。

- 8.1 版本将移除 interval 的 to_char() 函数。
- 服务器现在对传递给 oid/float4/float8 数据类型的空字符串发出警告，但仍像以前一样把它们解释为零。在下一个主版本中，空字符串将被视为这些数据类型的无效输入。
- 默认情况下，PostgreSQL 8.0 及更早版本中的表创建时带 OID。在下一个版本中，情况将不再如此：要创建包含 OID 的表，必须指定 WITH OIDS 子句或设置 default_with_oids 配置参数。鼓励用户在表需要 OID 时显式指定 WITH OIDS，以与 PostgreSQL 的未来版本兼容。

E.171.4. 变更

下面你将找到 8.0 版本与上一个主版本之间变更的详细说明。

E.171.4.1. 性能改进

- 支持跨数据类型的索引使用（Tom）

  在此变更之前，如果数据类型不完全匹配，许多查询不会使用索引。此改进使索引使用更直观一致。

- 改进缓存的新缓冲区替换策略（Jan）

  以前的版本使用最近最少使用（LRU）缓存把最近引用的页保留在内存中。LRU 算法不考虑特定缓存条目被访问的次数，因此大表扫描可能驱逐有用的缓存页。
  新缓存算法使用四个独立的列表来跟踪最近使用和最常使用的缓存页，
  并根据负载动态优化其替换。这应能带来对共享缓冲区缓存高效得多的利用。
  过去测试过共享缓冲区大小的管理员应使用此新缓存替换策略重新测试。

- 新增周期性写脏缓冲区的子进程以减少检查点写入（Jan）

  在以前的版本中，
  每隔几分钟运行一次的检查点进程会把所有脏缓冲区写入操作系统的缓冲区缓存，
  然后把所有脏的操作系统缓冲区刷到磁盘。这导致周期性的磁盘使用尖峰，常常损害性能。
  新代码使用后台写入器以稳定的节奏细流式写磁盘，因此检查点要写到磁盘的脏页少得多。
  此外，新代码不发出全局 `sync()` 调用，而是只对自上次检查点以来写入的文件执行 `fsync()`。这应能改进性能并使检查点期间的退化最小化。

- 新增延长清理时间以降低性能影响的能力（Jan）

  在繁忙系统上，`VACUUM` 执行许多 I/O 请求，可能损害其他用户的性能。此版本允许放慢 `VACUUM` 以降低对其他用户的影响，尽管这会增加 `VACUUM` 的总时长。

- 改进重复键的 B-tree 索引性能（Dmitry Tkach、Tom）

  这改进了索引中存在许多重复值时扫描索引的方式。

- 规划时使用动态生成的表大小估计（Tom）

  此前规划器使用最后一次 `VACUUM` 或 `ANALYZE` 看到的值来估计表大小，包括物理表大小（页数）和行数。现在，当前物理表大小从内核获得，行数通过把表大小乘以最后一次 `VACUUM` 或 `ANALYZE` 看到的行密度（每页行数）来估计。在表大小自上次维护命令以来变化显著的情形下，这应能产生更可靠的估计。

- 改进 `OR` 子句的索引使用（Tom）

  这允许优化器在有许多 OR 子句的语句中使用索引，这些语句过去不会被索引。它还可以使用多列索引，其中第一列被指定而第二列是 `OR` 子句的一部分。

- 改进部分索引子句的匹配（Tom）

  服务器现在更善于在涉及复杂 `WHERE` 子句的查询中使用部分索引。

- 改进 GEQO 优化器的性能（Tom）

  GEQO 优化器用于规划涉及许多表的查询（默认十二个或更多）。此版本加快了查询分析方式，以减少花在优化上的时间。

- 其他优化器改进

  这里没有篇幅列出所做的所有小改进，但许多特殊情形比以前的版本工作得更好。

- 改进 C 函数的查找速度（Tom）

  此版本使用哈希表查找动态装载的 C 函数的信息。这改进了它们的速度，使其几乎与内建在服务器可执行文件中的函数一样快。

- 新增类型特定的 `ANALYZE` 统计能力（Mark Cave-Ayland）

此特性允许为非标准数据类型生成统计信息时有更多灵活性。

- ANALYZE 现在收集表达式索引的统计信息（Tom）

  表达式索引（也称函数索引）允许用户不仅索引列，还索引表达式和函数调用的结果。
  有了此版本，优化器可以收集和使用有关表达式索引内容的统计信息。
  这将大幅改进表达式索引相关查询的规划质量。

- ANALYZE 的新两阶段采样方法（Manfred Koizar）

  当表中不同区域有效行的密度差异很大时，这给出更好的统计信息。

- 加快 TRUNCATE（Tom）

  这挽回了一些 7.4 中观察到的性能损失，同时仍保持 TRUNCATE 事务安全。

E.171.4.2. 服务器变更

- 新增 WAL 文件归档和时间点恢复（Simon Riggs）
- 新增表空间，使管理员能控制磁盘布局（Gavin）
- 新增内建日志轮转程序（Andreas Pflug）

  现在可以方便地记录服务器消息，而不依赖 syslog 或外部日志轮转程序。

- 新增只读服务器配置参数以显示服务器编译期设置：block_size、integer_datetimes、max_function_args、max_identifier_length、max_index_keys（Joe）
- 使 sameuser、samegroup 和 all 的加引号能移除这些词在 pg_hba.conf 中的特殊含义（Andrew）
- 在默认 pg_hba.conf 中为 localhost 使用更清晰的 IPv6 名称 ::1/128（Andrew）
- 在 pg_hba.conf 示例中使用 CIDR 格式（Andrew）
- 把服务器配置参数 SortMem 和 VacuumMem 改名为 work_mem 和 maintenance_work_mem（旧名仍受支持）（Tom）

  此变更是为了澄清索引和外键创建等批量操作使用 maintenance_work_mem，而 work_mem 用于查询执行期间使用的工作区。

- 允许用服务器配置 log_disconnections 记录会话断开（Andrew）
- 新增服务器配置参数 log_line_prefix，允许控制每行日志中发出的信息（Andrew）

  可用信息包括用户名、数据库名、远程 IP 地址和会话开始时间。

- 移除服务器配置参数 log_pid、log_timestamp、log_source_port；功能被 log_line_prefix 取代（Andrew）
- 用统一的 listen_addresses 参数取代 virtual_host 和 tcpip_socket 参数（Andrew、Tom）

  virtual_host 只能指定监听单个 IP 地址。listen_addresses 允许指定多个地址。

- 默认监听 localhost，这在许多场景中消除了对 `-i` postmaster 开关的需要（Andrew）

  监听 localhost（`127.0.0.1`）不会打开新的安全漏洞，但允许 Windows 和 JDBC 等不支持本地套接字的配置无需特殊调整即可工作。

- 移除 `syslog` 服务器配置参数，并新增更合逻辑的 `log_destination` 变量来控制日志输出位置（Magnus）

- 更改服务器配置参数 `log_statement`，取值 `all`、`mod`、`ddl` 或 `none`，以选择记录哪些查询（Bruce）

  这允许管理员只记录数据定义更改或只记录数据修改语句。

- 一些与日志相关的配置参数以前可由普通用户调整，但只能向“更详细”方向。
  现在对它们的处理更严格：只有超级用户能设置。但超级用户可以用 `ALTER USER` 为非超级用户提供这些值的逐用户设置。此外，超级用户现在还可以通过 `PGOPTIONS` 设置仅超级用户配置参数的值。

- 允许把配置文件放在数据目录之外（mlw）

  默认情况下，配置文件保存在集群的顶级目录。有了此新增，
  配置文件可以放在数据目录之外，便于管理。

- 只在首次执行时规划预备查询，使常数可用于统计（Oliver Jowett）

  预备语句一次规划查询并多次执行。虽然预备查询避免了每次使用时重新规划的开销，
  但计划质量因不知道查询中将使用的确切参数而受损。在此版本中，
  未命名预备语句的规划推迟到首次执行，并把该执行的实际参数值用作优化提示。
  这允许使用外线参数传递而不带来性能损失。

- 允许 `DECLARE CURSOR` 接受参数（Oliver Jowett）

  现在在带参数的 `Parse` 消息中发出 `DECLARE CURSOR` 是有用的。`Bind` 时发送的参数值将被代入游标查询的执行。

- 修复 `inet` 和 `cidr` 数据类型的哈希连接和聚合（Tom）

  7.4 版本对混合 `inet` 和 `cidr` 值的哈希处理不正确。（此缺陷在以前的版本中不存在，因为它们不会尝试哈希这两种数据类型。）

- 让 `log_duration` 只在 `log_statement` 打印查询时打印（Ed L.）

E.171.4.3. 查询变更

- 新增保存点（嵌套事务）（Alvaro）

- 不支持的隔离级别现在被接受并提升到最接近的受支持级别（Peter）

  SQL 规范规定，如果数据库不支持特定隔离级别，应使用下一个更严格的级别。此变更遵循该建议。

- 允许 `BEGIN WORK` 像 `START TRANSACTION` 那样指定事务隔离级别（Bruce）

- 修复规则生成的查询类型与最初提交的查询不同情形下的表权限检查（Tom）

- 实现美元引用以简化单引号的使用（Andrew、Tom、David Fetter）

在以前的版本中，由于必须用单引号引用函数体，
函数文本中单引号的使用需要用两个单引号或其他易错的记法。有了此版本，
我们新增了用"美元引用"引用文本块的能力。
在不同嵌套级别使用不同引用定界符的能力大幅简化了正确引用的任务，
特别是在复杂函数中。美元引用可用于任何需要引用文本的地方。

- 让 `CASE val WHEN compval1 THEN ...` 只求值 `val` 一次（Tom）

  `CASE` 不再多次求值被测试的表达式。当表达式复杂或易变时这有好处。

- 在计算聚合查询的目标列表之前先测试 `HAVING`（Tom）

  修复诸如 `SELECT SUM(win)/SUM(lose) ... GROUP BY ... HAVING SUM(lose) > 0` 之类情形的不当失败。这本应工作，但此前可能因除以零而失败。

- 用 `max_stack_depth` 参数（以栈大小千字节计）取代 `max_expr_depth` 参数（Tom）

  这为防止因失控递归函数而崩溃提供了相当可靠的防御。我们现在直接测量执行栈的大小，
  而不是测量表达式嵌套的深度。

- 允许任意行表达式（Tom）

  此版本允许 SQL 表达式包含任意组合类型，即行值。它还允许函数更容易地接受行作为参数并返回行值。

- 允许 `LIKE`/`ILIKE` 用作行和子查询比较中的操作符（Fabien Coelho）

- 避免标识符和关键字中的基本 ASCII 字母的区域特定大小写转换（Tom）

  这解决了包含 `I` 和 `i` 的词被破坏的“土耳其问题”。7 位 ASCII 集之外字符的折叠仍是区域感知的。

- 改进语法错误报告（Fabien、Tom）

  语法错误报告比以前更有用。

- 更改 `EXECUTE`，使其返回与被执行语句匹配的完成标签（Kris Jurka）

  以前的版本对任何 `EXECUTE` 调用都返回 `EXECUTE` 标签。在此版本中，返回的标签将反映被执行的命令。

- 避免在规则列表中发出 `NATURAL CROSS JOIN`（Tom）

  这样的子句在逻辑上没有意义，但在某些情形下规则反编译器以前会产生这种语法。

E.171.4.4. 对象操作变更

- 为转换、编码转换、语言、操作符类和大对象新增 `COMMENT ON`（Christopher）

- 新增服务器配置参数 `default_with_oids`，控制表默认是否带 `OID` 创建（Neil）

  这允许管理员控制 `CREATE TABLE` 命令默认创建带或不带 `OID` 列的表。（注意：`default_with_oids` 当前的出厂默认设置为 `TRUE`，但未来版本中默认将变为 `FALSE`。）

- 为 `CREATE TABLE AS` 新增 `WITH` / `WITHOUT OIDS` 子句（Neil）

- 允许 `ALTER TABLE DROP COLUMN` 删除 `OID` 列（`ALTER TABLE SET WITHOUT OIDS` 仍有效）（Tom）
- 允许组合类型作为表列（Tom）
- 允许带默认值和 `NOT NULL` 约束的 `ALTER ... ADD COLUMN`；按 SQL 规范工作（Rod）

  现在 `ADD COLUMN` 可以创建初始不用 NULL 填充而是用指定默认值填充的列。
- 新增 `ALTER COLUMN TYPE` 来更改列的类型（Rod）

  现在可以在不删除并重新添加列的情况下更改列的数据类型。
- 允许在单条 `ALTER TABLE` 命令中有多个 `ALTER` 动作（Rod）

  这对重写表的 `ALTER` 命令（包括 `ALTER COLUMN TYPE` 和带默认值的 `ADD COLUMN`）特别有用。通过把 `ALTER` 命令分组，表只需重写一次。
- 允许 `ALTER TABLE` 添加 `SERIAL` 列（Tom）

  这是新列可指定默认值这一新能力的自然结果。
- 允许更改聚合、编码转换、数据库、函数、操作符、操作符类、模式、类型和表空间的所有者（Christopher、Euler Taveira de Oliveira）

  此前这需要直接修改系统表。
- 允许把临时对象创建限制为 `SECURITY DEFINER` 函数（Sean Chittenden）
- 新增 `ALTER TABLE ... SET WITHOUT CLUSTER`（Christopher）

  在此版本之前，除了修改系统表外没有办法清除自动聚簇规格。
- 约束/索引/`SERIAL` 名称现在为 *`table_column_type`* 并附加数字以保证在模式内唯一（Tom）

  SQL 规范规定此类名称应在模式内唯一。
- 新增 `pg_get_serial_sequence()`，返回 `SERIAL` 列的序列名（Christopher）

  这使自动脚本能可靠地找到 `SERIAL` 序列名。
- 当主/外键数据类型不匹配需要昂贵查找时发出警告
- 新 `ALTER INDEX` 命令，允许在表空间之间移动索引（Gavin）
- 让 `ALTER TABLE OWNER` 也更改依赖序列的所有权（Alvaro）

E.171.4.5. 工具命令变更

- 允许 `CREATE SCHEMA` 创建触发器、索引和序列（Neil）
- 为 `CREATE RULE` 新增 `ALSO` 关键字（Fabien Coelho）

  这允许在创建规则时添加 `ALSO`，与 `INSTEAD` 规则相对。
- 为 `LOCK` 新增 `NOWAIT` 选项（Tatsuo）

这允许 `LOCK` 命令在必须等待所请求的锁时失败。

- 允许 `COPY` 读写逗号分隔值（CSV）文件（Andrew、Bruce）
- 当 `COPY` 分隔符与 NULL 字符串冲突时报错（Bruce）
- `GRANT/REVOKE` 行为更贴近 SQL 规范
- 避免 `CREATE INDEX` 与 `CHECKPOINT` 之间的锁定冲突（Tom）

  在 7.3 和 7.4 中，长时间运行的 B-tree 索引构建可能阻塞并发 `CHECKPOINT` 完成，从而因 WAL 日志无法回收导致 WAL 膨胀。

- 数据库范围的 `ANALYZE` 不再跨表持有锁（Tom）

  这降低了与其他想独占锁定表的后端发生死锁的可能性。要获得此变更的好处，不要在事务块（`BEGIN` 块）内执行数据库范围的 `ANALYZE`；它必须能为每个表提交并开始新事务。

- `REINDEX` 不再独占锁定索引的父表

  索引本身仍被独占锁定，但如果表的读者不使用正被重建的特定索引，他们可以继续。

- 用户被重命名时擦除 MD5 用户口令（Bruce）

  PostgreSQL 用 MD5 加密口令时把用户名用作盐。当用户名被更改时，盐将不再与存储的 MD5 口令匹配，因此存储的口令变得无用。在此版本中会生成通知并清除口令。之后如果用户要能用口令登录，必须分配新口令。

- pg_ctl 新增用于 Windows 的 `kill` 选项（Andrew）

  Windows 没有向后端发送信号的 `kill` 命令，因此把此能力添加到了 pg_ctl。

- 信息模式改进
- 为 initdb 新增 `--pwfile` 选项，使初始口令可由 GUI 工具设置（Magnus）
- 在 initdb 中检测区域设置/编码不匹配（Peter）
- 为 pg_ctl 新增 `register` 命令来注册 Windows 操作系统服务（Dave Page）

E.171.4.6. 数据类型和函数变更

- 更完整的组合类型（行类型）支持（Tom）

  组合值可以用在许多以前只有标量值可用的地方。

- 把非矩形数组值作为错误拒绝（Joe）

  此前 `array_in` 会静默构建令人惊讶的结果。

- 整数算术运算中的溢出现在被检测（Tom）
- 与单字节 `"char"` 数据类型相关的算术操作符已被移除。

  此前，解析器会在许多情形下选择这些操作符，而“无法选择操作符”错误会更合适，例如 `null * null`。如果确实想对 `"char"` 列做算术，可以显式把它转换为整数。

- 数组输入值的语法检查大幅收紧（Joe）

  以前在奇怪位置被允许并产生奇怪结果的垃圾现在导致 `ERROR`，例如右花括号之后的非空白。

- 空字符串数组元素值现在必须写成 `""`，而不是什么也不写（Joe）

  此前两种写空字符串元素值的方式都被允许，但现在要求带引号的空字符串。
  完全什么都不出现的情形在未来某个版本中可能会被视为 NULL 元素值。

- 数组元素的尾随空白现在被忽略（Joe）

  此前前导空白被忽略，但元素值与分隔符或右花括号之间的尾随空白是有意义的。
  现在尾随空白也被忽略。

- 当下界不为一时以显式数组界限输出数组值（Joe）
- 接受 `YYYY-monthname-DD` 作为日期字符串（Tom）
- 让 `netmask` 和 `hostmask` 函数返回最大长度的掩码长度（Tom）
- 更改阶乘函数使其返回 `numeric`（Gavin）

  返回 `numeric` 使阶乘函数能用于更大范围的输入值。

- `to_char`/`to_date()` 日期转换改进（Kurt Roeckx、Fabien Coelho）
- 让 `length()` 无视 `CHAR(n)` 中的尾随空格（Gavin）

  此变更是为了改进一致性：尾随空格在 `CHAR(n)` 数据中语义上无关紧要，因此不应被 `length()` 统计。

- 对传递给 `OID`/`float4`/`float8` 数据类型的空字符串发出警告（Neil）

  8.1 将改为抛出错误。

- 允许 `int2`/`int4`/`int8`/`float4`/`float8` 输入例程中的前导或尾随空白（Neil）
- 更好地支持 `float4`/`float8` 中的 IEEE `Infinity` 和 `NaN` 值（Neil）

  它们现在应在所有支持 IEEE 兼容浮点运算的平台上工作。

- 为 `date_trunc()` 新增 `week` 选项（Robert Creager）
- 修复 `to_char` 对 `1 BC` 的处理（此前它返回 `1 AD`）（Bruce）
- 修复 BC 日期的 `date_part(year)`（此前它返回比正确年份小一的值）（Bruce）
- 修复 `date_part()`，使其返回正确的千年和世纪（Fabien Coelho）

  与这类事物的标准推算相比，以前的版本中世纪和千年的结果数字错误且起始年份错误。

- 为符合标准，新增 `ceiling()` 作为 `ceil()` 的别名，`power()` 作为 `pow()` 的别名（Neil）
- 更改 `ln()`、`log()`、`power()` 和 `sqrt()`，使其按 SQL:2003 的规定在某些错误条件下发出正确的 `SQLSTATE` 错误码（Neil）
- 新增 SQL:2003 定义的 `width_bucket()` 函数（Neil）
- 新增 `generate_series()` 函数以简化数字集合的处理（Joe）

- 修复 `upper/lower/initcap()` 函数使其支持多字节编码（Tom）
- 新增布尔和按位整数 `AND/OR` 聚合（Fabien Coelho）
- 新增返回客户端和服务器网络地址的会话信息函数（Sean Chittenden）
- 新增确定闭合路径面积的函数（Sean Chittenden）
- 新增向其他后端发送取消请求的函数（Magnus）
- 新增 `interval` 加 `datetime` 操作符（Tom）

  相反的顺序（`datetime` 加 `interval`）此前已支持，但两者都是 SQL 标准要求的。
- 把整数转换为 `BIT(N)` 选取整数最右边的 N 位（Tom）

  在以前的版本中选取的是最左边 N 位，但这被认为没有用处，更不用说与从 bit 到 int 的转换不一致了。
- 要求 `CIDR` 值的所有未掩码位为零（Kevin Brintnall）

E.171.4.7. 服务器端语言变更

- 在 `READ COMMITTED` 序列化模式下，易变函数现在能看到并发事务提交到函数内每条语句开始处为止的结果，而不是到调用函数的交互命令开始处为止。
- 声明为 `STABLE` 或 `IMMUTABLE` 的函数总是使用调用查询的快照，因此看不到调用查询开始之后采取动作的效果，无论在其自身事务还是其他事务中。这样的函数也必须是只读的，即不能使用 `SELECT` 之外的任何 SQL 命令。把函数声明为 `STABLE` 或 `IMMUTABLE` 而不是 `VOLATILE` 有可观的性能收益。
- 非延迟的 `AFTER` 触发器现在在触发查询完成后立即触发，而不是在当前交互命令完成时。当触发查询发生在函数内时这就有区别：触发器在函数继续其下一个操作之前被调用。例如，如果函数向表插入新行，任何非延迟外键检查都在函数继续之前进行。
- 允许函数参数带名称声明（Dennis Björklund）

  这允许对函数做更好的文档记录。名称是否真起作用取决于所用的具体函数语言。
- 允许在函数中引用 PL/pgSQL 参数名（Dennis Björklund）

  这基本上为每个命名参数创建了自动别名。
- 在创建时对 PL/pgSQL 函数做最简语法检查（Tom）

  这使我们能更早捕获简单的语法错误。
- PL/pgSQL 中对组合类型（行和记录变量）的更多支持

  例如，现在可以把行类型变量作为单个变量传递给另一个函数。
- PL/pgSQL 变量的默认值现在可以引用先前声明的变量
- 改进 PL/pgSQL FOR 循环的解析（Tom）

  解析现在由 `".."` 的存在驱动，而不是 `FOR` 变量的数据类型。这对正确的函数没有区别，但在出错时应产生更易理解的错误消息。

- PL/Perl 服务器端语言的重大改造（Command Prompt、Andrew Dunstan）
- 在 PL/Tcl 中，SPI 命令现在在子事务中运行。如果发生错误，子事务被清理，错误被报告为普通 Tcl 错误，可以用 `catch` 捕获。此前无法捕获此类错误。
- PL/pgSQL 接受 `ELSEIF`（Neil）

  此前 PL/pgSQL 只允许 `ELSIF`，但许多人习惯把此关键字拼写为 `ELSEIF`。

E.171.4.8. psql Changes

- 改进 psql 对数据库对象的信息显示（Christopher）
- 允许 psql 在 `\du` 和 `\dg` 中显示组成员身份（Markus Bertheau）
- 防止 psql 的 `\dn` 显示临时模式（Bruce）
- 允许 psql 处理文件名的波浪号用户展开（Zach Irmen）
- 允许 psql 通过 readline 显示精致提示符，包括颜色（Reece Hart、Chet Ramey）
- 让 psql 的 `\copy` 完全匹配 `COPY` 命令语法（Tom）
- 显示语法错误的位置（Fabien Coelho、Tom）
- 向 psql 的 `\d` 显示中添加 `CLUSTER` 信息（Bruce）
- 更改 psql 的 `\copy stdin/stdout`，使其从命令输入/输出读取（Bruce）
- 新增 `pstdin`/`pstdout`，从 psql 的 `stdin`/`stdout` 读取（Mark Feit）
- 新增全局 psql 配置文件 `psqlrc.sample`（Bruce）

  这允许用一个中心文件存储全局 psql 启动命令。
- 让 psql 的 `\d+` 指示表是否有 `OID` 列（Neil）
- 在 Windows 上，psql 读取文件时使用二进制模式，使 control-Z 不被视为文件结束
- 让 `\dn+` 显示模式的权限和描述（Dennis Björklund）
- 改进 Tab 补全支持（Stefan Kaltenbrunn、Greg Sabino Mullane）
- 允许用大写或小写设置布尔设置（Michael Paesold）

E.171.4.9. pg_dump Changes

- 使用依赖信息改进 pg_dump 的可靠性（Tom）

  这应能解决相关对象有时以错误顺序转储的长期问题。
- 让 pg_dump 尽可能按字母顺序输出对象（Tom）

  这应使识别转储文件之间的更改更容易。
- 允许 pg_restore 忽略某些 SQL 错误（Fabien Coelho）

这使 pg_restore 的行为类似于把 pg_dump 输出脚本送入 psql 的结果。在大多数情况下，忽略错误继续前进是最有用的做法。还添加了 pg_restore 选项以提供出错即退出的旧行为。

- pg_restore 的 `-l` 显示现在包含对象的模式名
- pg_dump 文本输出中新增开始/结束标记（Bruce）
- 为详细模式下的 pg_dump/pg_dumpall 新增开始/停止时间（Bruce）
- 允许在 pg_dumpall 中使用大多数 pg_dump 选项（Christopher）
- 让 pg_dump 默认使用 `ALTER OWNER` 而不是 `SET SESSION AUTHORIZATION`（Christopher）

E.171.4.10. libpq 变更

- 使 libpq 的 `SIGPIPE` 处理线程安全（Bruce）
- 新增 `PQmbdsplen()`，返回字符的显示长度（Tatsuo）
- 为 SSL 和 Kerberos 连接添加线程锁定（Manfred Spraul）
- 允许 `PQoidValue()`、`PQcmdTuples()` 和 `PQoidStatus()` 用于 `EXECUTE` 命令（Neil）
- 新增 `PQserverVersion()`，提供对服务器版本号更便捷的访问（Greg Sabino Mullane）
- 新增 `PQprepare/PQsendPrepared()` 函数，支持预备语句而不必指定其参数的数据类型（Abhijit Menon-Sen）
- 许多 ECPG 改进，包括 `SET DESCRIPTOR`（Michael）

E.171.4.11. 源代码变更

- 允许数据库服务器原生运行于 Windows（Claudio、Magnus、Andrew）
- 把 shell 脚本命令转换为 C 版本以支持 Windows（Andrew）
- 创建扩展 makefile 框架（Fabien Coelho、Peter）

  这简化了在原始源代码树之外构建扩展的任务。

- 支持可重新定位的安装（Bruce）

  已安装文件（如 `/share` 目录）的目录路径现在相对于可执行文件的实际位置计算，因此安装树可以移动到其他地方而无需重新配置和重新构建。

- 用 `--with-docdir` 选择文档的安装位置；还允许 `--infodir`（Peter）
- 新增 `--without-docdir` 以阻止安装文档（Peter）
- 升级到 DocBook V4.2 SGML（Peter）
- 新的 `PostgreSQL` CVS 标签（Marc）

  这是为了使组织更容易管理自己的 PostgreSQL CVS 仓库副本。来自主仓库的文件版本戳不会因向复制的仓库检入或从中检出而被破坏。

- 澄清锁定代码（Manfred Koizar）
- 缓冲区管理器清理（Neil）
- 把平台测试与 CPU 自旋锁代码解耦（Bruce、Tom）
- 为 gcc 在 PA-RISC 上新增内联 test-and-set 代码（ViSolve、Tom）
- 改进 i386 自旋锁代码（Manfred Spraul）
- 清理自旋锁汇编代码以避免较新 gcc 版本的警告（Tom）
- 从源代码树移除 JDBC；现为独立项目
- 移除 libpgtcl 客户端接口；现为独立项目
- 更准确地估计内存和文件描述符的使用（Tom）
- Mac OS X 启动脚本的改进（Ray A.）
- 新的 `fsync()` 测试程序（Bruce）
- 重大文档改进（Neil、Peter）
- 移除 pg_encoding；不再需要
- 移除 pg_id；不再需要
- 移除 initlocation；不再需要
- 自动检测线程标志（不再需要手工测试）（Bruce）
- 使用 Olson 的公有领域 timezone 库（Magnus）
- 启用线程时，Unixware 上的后端可执行文件也使用线程标志（Bruce）

  Unixware 不能在同一可执行文件中混用线程化和非线程化目标文件，因此所有内容都必须编译为线程化。
- psql 现在使用 flex 生成的词法分析器处理命令字符串
- 重新实现后端各处使用的链表数据结构（Neil）

  这通过使列表追加和长度操作更高效来改进性能。
- 允许动态装载的模块创建自己的服务器配置参数（Thomas Hallgren）
- 新的巴西葡萄牙语版 FAQ（Euler Taveira de Oliveira）
- 新增法语 FAQ（Guillaume Lelarge）
- 新的用于 Windows 日志的 pgevent
- 使 libpq 和 ECPG 在 OS X 上构建为正确的共享库（Tom）

E.171.4.12. 附加模块变更

- `contrib/dblink` 的彻底改造（Joe）
- `contrib/dbmirror` 改进（Steven Singer）

- 新增 `contrib/xml2`（John Gray、Torchbox）
- 更新了 `contrib/mysql`
- 新版本的 `contrib/btree_gist`（Teodor）
- 新的 `contrib/trgm`，PostgreSQL 的三字母组匹配（Teodor）
- 许多 `contrib/tsearch2` 改进（Teodor）
- 为 `contrib/fuzzystrmatch` 添加双变音位（Andrew）
- 允许 `contrib/pg_autovacuum` 作为 Windows 服务运行（Dave Page）
- 为 `contrib/dbsize` 新增函数（Andreas Pflug）
- 移除 `contrib/pg_logger`：被集成日志子进程取代
- 移除 `contrib/rserv`：被多个独立项目取代

E.172. 版本 7.4.30

发行日期
2010-10-04

本次发布包含对 7.4.29 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

这预计是 PostgreSQL 7.4.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.172.1. 迁移到版本 7.4.30

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅第 E.176 节。

E.172.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

  感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，会出现此错误。
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）

E.173. 版本 7.4.29

发行日期
2010-05-17

本次发布包含对 7.4.28 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 7.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.173.1. 迁移到版本 7.4.29

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅第 E.176 节。

E.173.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）
- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，

因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

E.174. 版本 7.4.28

发行日期
2010-03-15

本次发布包含对 7.4.27 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 7.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.174.1. 迁移到版本 7.4.28

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅第 E.176 节。

E.174.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复若干病态缓慢的正则表达式匹配情形（Tom）
- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。
- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）

E.175. 版本 7.4.27

发行日期
2009-12-14

本次发布包含对 7.4.26 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.175.1. 迁移到版本 7.4.27

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅第 E.176 节。

E.175.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。
- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）

- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）

E.176. 版本 7.4.26

发行日期
2009-09-09

本次发布包含对 7.4.25 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.176.1. 迁移到版本 7.4.26

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 7.4.26 后对它们执行 `REINDEX`。另外，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.176.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.9999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）
- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

E.177. 版本 7.4.25

发行日期
2009-03-16

本次发布包含对 7.4.24 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.177.1. 迁移到版本 7.4.25

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.177.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，
  解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）
- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.178. 版本 7.4.24

发行日期
2009-02-02

本次发布包含对 7.4.23 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.178.1. 迁移到版本 7.4.24

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.178.2. 变更

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 修复 `to_char()` 处理 `TH` 格式代码中的缺陷（Andreas Scherbaum）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）

E.179. 版本 7.4.23

发行日期
2008-11-03

本次发布包含对 7.4.22 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.179.1. 迁移到版本 7.4.23

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.179.2. 变更

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
  修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。

- 修复 ecpg 对 `CREATE USER` 的解析（Michael）

E.180. 版本 7.4.22

发行日期
2008-09-22

本次发布包含对 7.4.21 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.180.1. 迁移到版本 7.4.22

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.180.2. 变更

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。
- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）

E.181. 版本 7.4.21

发行日期
2008-06-12

此发行版在 7.4.20 的基础上包含一个严重的缺陷修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.181.1. 迁移到版本 7.4.21

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.181.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最

近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

E.182. 版本 7.4.20

发行日期

从未发布

本次发布包含对 7.4.19 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.182.1. 迁移到版本 7.4.20

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.182.2. 变更

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）
- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。
- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。
- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）
- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。
- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。
- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.183. 版本 7.4.19

发行日期
2008-01-07

此发行版包含 7.4.18 之后的多种修复，包括对重要安全问题的修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.183.1. 迁移到版本 7.4.19

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.183.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  7.4.18 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）

- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）
- ecpg 解析器修复（Michael）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）
- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.184. 版本 7.4.18

发行日期
2007-09-17

此发行版包含 7.4.17 之后的修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.184.1. 迁移到版本 7.4.18

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.184.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 防止 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.185. 版本 7.4.17

发行日期
2007-04-23

此发行版包含 7.4.16 之后的修复，包括一个安全修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.185.1. 迁移到版本 7.4.17

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.185.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 修复哈希索引扩容期间的 PANIC（7.4.15 中引入的缺陷）（Tom）

E.186. 版本 7.4.16

发行日期
2007-02-05

此发行版包含 7.4.15 之后的多种修复，包括一个安全修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.186.1. 迁移到版本 7.4.16

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.186.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  该漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，或更改 SQL 函数中使用的表列的数据类型（CVE-2007-0555）。此错误很容易被利用来导致后端崩溃，
  而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）
- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）
- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.187. 版本 7.4.15

发行日期
2007-01-08

本次发布包含对 7.4.14 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.187.1. 迁移到版本 7.4.15

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.187.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）
- 修复影响多 GB 哈希索引的缺陷（Tom）
- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。

- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。

E.188. 版本 7.4.14

发行日期
2006-10-16

本次发布包含对 7.4.13 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.188.1. 迁移到版本 7.4.14

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.188.2. 变更

- 修复把无类型字面值当作 ANYARRAY 时的核心转储

- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。

- 修复 psql 的 `\d` 命令模式匹配的极端情形

- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）

- 修复 /contrib/dbmirror 中的反斜杠转义

- 针对美国 DST 法规的最近变更调整回归测试

E.189. 版本 7.4.13

发行日期
2006-05-23

此发行版包含 7.4.12 之后的多种修复，包括针对极其严重安全问题的补丁。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.189.1. 迁移到版本 7.4.13

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.189.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用

程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）
- 修复有时导致 OR 索引扫描漏掉本应返回的行的缺陷
- 修复 btree 索引被截断情形的 WAL 重放
- 修复涉及 `|` 的模式的 `SIMILAR TO`（Tom）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 修复 Intel Mac 上的 Bonjour（Ashley Clark）
- 修复各种小的内存泄漏

E.190. 版本 7.4.12

发行日期
2006-02-14

本次发布包含对 7.4.11 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.190.1. 迁移到版本 7.4.12

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅第 E.191 节。

E.190.2. 变更

- 修复 `SET SESSION AUTHORIZATION` 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 为预备语句中的 `UNKNOWN` 参数正确检查 `DOMAIN` 约束（Neil）
- 修复以允许恢复有对自定义操作符跨模式引用的转储（Tom）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.191. 版本 7.4.11

发行日期
2006-01-09

本次发布包含对 7.4.10 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.191.1. 迁移到版本 7.4.11

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅第 E.194 节。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.191.2. 变更

- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）
- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。
- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。
- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）
- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。
- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.192. 版本 7.4.10

发行日期
2005-12-12

本次发布包含对 7.4.9 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.192.1. 迁移到版本 7.4.10

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅第 E.194 节。

E.192.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败
- `/contrib/ltree` 修复（Teodor）
- AIX 和 HPUX 编译修复（Tom）
- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。

- 防止表被删除后 pg_autovacuum 的核心转储

E.193. 版本 7.4.9

发行日期
2005-10-04

本次发布包含对 7.4.8 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.193.1. 迁移到版本 7.4.9

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅第 E.194 节。

E.193.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。

- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

  在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。

- 修正 `COPY` 中只读事务测试的方向

  以前的代码禁止 `COPY TO`，而它应禁止的是 `COPY FROM`。

- 修复只引用内侧关系的外连接 ON 子句的规划问题
- 对 `x FULL JOIN y ON true` 极端情形的进一步修复
- 让 `array_in` 和 `array_recv` 更严格地验证其 OID 参数
- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题

- 改进日期时间解析的健壮性
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- postmaster 启动期间不要试图打开超过 `max_files_per_process` 个文件
- 各种内存泄漏修复
- 各种可移植性改进
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`
- 更新 `contrib/tsearch2` 以使用当前的 Snowball 代码

E.194. 版本 7.4.8

发行日期
2005-05-09

此发行版包含 7.4.7 之后的多种修复，包括几个与安全相关的问题。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.194.1. 迁移到版本 7.4.8

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，它是处理在 7.4.X 系统目录初始内容中发现的两个重要安全问题的可能方法之一。用 7.4.8 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

较大的安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）

较小的问题是 `contrib/tsearch2` 模块创建了几个被错误声明为返回 `internal` 的函数，而它们并不接受 `internal` 参数。这破坏了所有使用 `internal` 参数的函数的类型安全。

强烈建议所有安装通过 initdb 或按照下述手工修复程序修复这些错误。这些错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行以下程序。作为数据库超级用户，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[3] = 'internal'::regtype
WHERE pronamespace = 11 AND pronargs = 5
     AND proargtypes[2] = 'cstring'::regtype;
-- The command should report having updated 90 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

然后，如果你安装了 `contrib/tsearch2`，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[0] = 'internal'::regtype
WHERE oid IN (
   'dex_init(text)'::regprocedure,
   'snb_en_init(text)'::regprocedure,
   'snb_ru_init(text)'::regprocedure,
   'spell_init(text)'::regprocedure,
   'syn_init(text)'::regprocedure
);
-- The command should report having updated 5 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

如果此命令失败并显示类似 “function "dex_init(text)" does not exist” 的消息，则要么该数据库中未安装 `tsearch2`，要么你已经做过此更新。

上述程序必须在安装的每个数据库中执行，包括 `template1`，理想情况下也包括 `template0`。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。`template1` 可以像任何其他数据库一样修复，但修复 `template0` 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 `template0` 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.194.2. 变更

- 更改编码函数签名以防止误用
- 更改 `contrib/tsearch2` 以避免不安全地使用 `INTERNAL` 函数结果
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。

- 确保后端关闭期间完成的操作被统计收集器计数

  这有望解决关于 pg_autovacuum 不够频繁地清理系统目录的报告——它没有被通知后端退出期间临时表删除导致的目录删除。

- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 防止 `COALESCE(NULL,NULL)` 的崩溃
- 修复 `array_map`，使其正确调用 PL 函数
- 修复 `ALTER DATABASE RENAME` 中的权限检查
- 修复 `ALTER LANGUAGE RENAME`
- 让 `RemoveFromWaitQueue` 清理自身

  这修复了一个锁管理错误，它只在事务被踢出锁等待（通常由查询取消）且锁持有者随后在极窄的窗口内释放锁时可见。

- 修复 `INSERT ... SELECT` 中出现无类型参数的问题
- 修复 `ALTER TABLE SET WITHOUT OIDS` 之后 `CLUSTER` 的失败

E.195. 版本 7.4.7

发行日期
2005-01-31

此发行版包含 7.4.6 之后的多种修复，包括几个与安全相关的问题。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.195.1. 迁移到版本 7.4.7

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.195.2. 变更

- 不允许非超级用户使用 `LOAD`

在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。

- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 plperl 处理元组字段中引号的问题
- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示
- 让 age(timestamptz) 用本地时区而不是 GMT 计算

E.196. 版本 7.4.6

发行日期
2004-10-22

本次发布包含对 7.4.5 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.196.1. 迁移到版本 7.4.6

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.196.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 `make_oidjoins_check` 在 `/tmp` 中使用临时文件

这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

- 防止强制后端关闭重新发出先前的命令结果

  极少数情况下，客户端可能认为其最后一条命令成功了，
  而它实际上已被强制数据库关闭中止。

- 修复 `pg_stat_get_backend_idset` 中的缺陷

  这可能导致某些系统统计视图行为异常。

- 修复 postmaster 中的小内存泄漏

- 修复“预期两个交换的表都有 TOAST 表”缺陷

  这可能发生在如 ALTER TABLE DROP COLUMN 之后再 CLUSTER 的情形。

- 防止 `pg_ctl restart` 多次添加 `-D`

- 修复 GiST 索引中 NULL 值的问题

- `::` 不再被解释为 ECPG prepare 语句中的变量

E.197. 版本 7.4.5

发行日期
2004-08-18

此发行版在 7.4.4 的基础上包含一个严重的缺陷修复。有关 7.4 主版本新特性的信息，请参阅第 E.202 节。

E.197.1. 迁移到版本 7.4.5

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.197.2. 变更

- 修复并发 B-tree 索引插入期间可能的崩溃

  此补丁修复了并发插入 B-tree 索引可能导致服务器 panic 的罕见情形。不会造成永久损坏，但仍值得重新发布。该缺陷在 7.4 之前的版本中不存在。

E.198. 版本 7.4.4

发行日期
2004-08-16

本次发布包含对 7.4.3 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.198.1. 迁移到版本 7.4.4

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.198.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 在求值聚合计划的结果列表之前检查 HAVING 限制
- 避免会话的当前用户 ID 被删除时的崩溃
- 修复零行情形的哈希 crosstab（Joe）
- 在外键中重命名列后强制更新缓存
- 正确美化 UNION 查询的打印
- 让 psql 在 COPY IN 中正确处理 `\r\n` 换行
- pg_dump 对带 grant 选项的 ACL 处理不正确
- 修复 OS X 和 Solaris 的线程支持
- 更新了 JDBC 驱动（build 215），包含各种修复
- ECPG 修复
- 翻译更新（多位贡献者）

E.199. 版本 7.4.3

发行日期
2004-06-14

本次发布包含对 7.4.2 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.199.1. 迁移到版本 7.4.3

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.199.2. 变更

- 修复使用非哈希聚合时的临时内存泄漏（Tom）

- ECPG 修复，包括一些针对 Informix 兼容性的（Michael）
- 修复线程安全编译的问题，特别是 Solaris（Bruce）
- 修复使用旧网络协议时 COPY IN 结束时的错误（ljb）
- pg_autovacuum 的几个重要修复，包括大表、无符号 oid、稳定性、临时表和调试模式的修复（Matthew T. O'Connor）
- 修复在 NetBSD 和 BSD/OS 上读取 tar 格式转储的问题（Bruce）
- 若干 JDBC 修复
- 修复 last_value 等于重启值时的 ALTER SEQUENCE RESTART（Tom）
- 修复嵌套子查询未重新计算的问题（Tom）
- 修复 LIMIT/OFFSET 中非常量表达式的问题
- 支持无连接子句的 FULL JOIN，如 X FULL JOIN Y ON TRUE（Tom）
- 修复另一个零列表缺陷（Tom）
- 改进子查询中 GROUP BY 子句里非限定标识符的处理（Tom）

  子查询内的选择列表别名现在优先于来自外层查询级别的名称。
- 反编译规则时不生成“NATURAL CROSS JOIN”（Tom）
- 为二进制 COPY 中的无效字段长度添加检查（Tom）

  这修复了一个难以利用的安全漏洞。
- 避免 `ANALYZE` 与 `LISTEN/NOTIFY` 之间的锁定冲突
- 大量翻译更新（多位贡献者）

E.200. 版本 7.4.2

发行日期
2004-03-08

本次发布包含对 7.4.1 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.200.1. 迁移到版本 7.4.2

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，作为合并 7.4.X 系统目录初始内容中发现的两个错误的修复的最简单方法，它可能是可取的。用 7.4.2 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

两个错误中较严重的是数据类型 `anyarray` 的对齐标签错误；这是一个问题，因为 `pg_statistic` 系统目录使用 `anyarray` 列。当规划涉及双对齐列（如 `float8` 和 `timestamp`）

上 WHERE 子句的查询时，错误标记可能导致规划器误估甚至崩溃。强烈建议所有安装通过 initdb 或按照下述手工修复程序修复此错误。

较小的错误是系统视图 pg_settings 应被标记为具有公共更新访问权，以允许用 UPDATE pg_settings 替代 SET。这也可以通过 initdb 或手工修复，但除非你想用 UPDATE pg_settings，否则不必修复。

如果你不想做 initdb，以下程序可以修复 pg_statistic。作为数据库超级用户，执行：

```
-- clear out old data in pg_statistic:
DELETE FROM pg_statistic;
VACUUM pg_statistic;
-- this should update 1 row:
UPDATE pg_type SET typalign = 'd' WHERE oid = 2277;
-- this should update 6 rows:
UPDATE pg_attribute SET attalign = 'd' WHERE atttypid = 2277;
--
-- At this point you MUST start a fresh backend to avoid a crash!
--
-- repopulate pg_statistic:
ANALYZE;
```

这可以在运行中的数据库上完成，但要注意在重新填充 pg_statistic 之前，被更改数据库中运行的所有后端都必须重启才是安全的。

要修复 pg_settings 错误，只需执行：

```
GRANT SELECT, UPDATE ON pg_settings TO PUBLIC;
```

上述程序必须在安装的每个数据库中执行，包括 template1，理想情况下也包括 template0。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。template1 可以像任何其他数据库一样修复，但修复 template0 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 template0 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.200.2. 变更

7.4.2 版本包含 7.3.6 版本中的所有修复，外加以下修复：

- 修复可能导致优化器崩溃的 pg_statistics 对齐缺陷

  有关此问题的详情见上文。

- 允许非超级用户更新 pg_settings

- 修复几个优化器缺陷，其中大多数导致 “在子计划目标列表中未找到变量” 错误

- 避免大型多索引扫描启动期间的内存不足失败
- 修复可能导致 `COPY IN` 期间“内存不足”错误的多字节问题
- 修复从无 OID 的表 `SELECT INTO` / `CREATE TABLE AS` 的问题
- 修复并行测试期间 `alter_table` 回归测试的问题
- 修复达到打开文件限制的问题，特别是在 OS X 上（Tom）
- 对土耳其语区域设置问题的部分修复

  initdb 现在能在土耳其语区域设置下成功，但仍有一些与 `i/I` 问题相关的不便。
- 让 pg_dump 在恢复时设置客户端编码
- 其他次要的 pg_dump 修复
- 允许 ecpg 再次使用 C 关键字作为列名（Michael）
- 为 ecpg 的 `SELECT/INSERT/UPDATE/DELETE` 添加 `WHENEVER NOT_FOUND`（Michael）
- 修复调用返回集合函数的查询导致的 ecpg 崩溃（Michael）
- 各种其他 ecpg 修复（Michael）
- Borland 编译器的修复
- 线程构建改进（Bruce）
- 各种其他构建修复
- 各种 JDBC 修复

E.201. 版本 7.4.1

发行日期
2003-12-22

本次发布包含对 7.4 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.202 节。.

E.201.1. 迁移到版本 7.4.1

对于运行 7.4 的用户，不需要进行转储/恢复。

如果你想在信息模式中安装这些修复，需要把它重新装载到数据库中。这可以通过运行 `initdb` 初始化新集群来完成，或者在安装新版本后，以超级用户身份在 psql 中对每个数据库（理想情况下包括 `template1`）运行以下 SQL 命令序列：

```
DROP SCHEMA information_schema CASCADE;
\i /usr/local/pgsql/share/information_schema.sql
```

在第二条命令中替换你的安装路径。

E.201.2. 变更

- 修复 ECPG 中 `CREATE SCHEMA` 解析的缺陷（Michael）
- 修复 `--enable-thread-safety` 与 `--with-perl` 一起使用时的编译错误（Peter）
- 修复使用哈希连接的子查询（Tom）

  某些使用哈希连接的子查询会因不当共享的结构而崩溃。
- 修复空闲空间映射压缩缺陷（Tom）

  这修复了空闲空间映射的压缩可能导致数据库服务器关闭的缺陷。
- 修复 libpq 的 Borland 编译器构建（Bruce）
- 修复 `netmask()` 和 `hostmask()`，使其返回最大长度的掩码长度（Tom）

  修复这些函数使其返回与 7.4 之前版本一致的值。
- 若干 `contrib/pg_autovacuum` 修复

  修复包括不当的变量初始化、`TRUNCATE` 后缺失的清理以及长清理的时长计算溢出。
- 允许 `contrib/cube` 在 Cygwin 下编译（Jason Tishler）
- 修复 Solaris 在未定义口令时使用口令文件的问题（Tom）

  修复 Solaris 上因在未定义口令时使用任何类型的口令认证导致的崩溃。
- JDBC 线程问题修复及其他修复
- 修复 `bytea` 索引查找（Joe）
- 修复信息模式对位数据类型的支持（Peter）
- 强制在从 WAL 恢复期间开启 zero_damaged_pages
- 防止“变量不在子计划目标列表中”的一些晦涩情形
- 使 `PQescapeBytea` 与 `byteaout` 相互一致（Joe）
- 对大于 0x7e 的字节转义 `bytea` 输出（Joe）

  如果 `bytea` 输出和输入使用不同的客户端编码，`bytea` 值可能被不同的编码破坏。此修复转义了所有可能受影响的字节。
- 为 dblink 的 `get_tuple_of_interest()` 添加缺失的 `SPI_finish()` 调用（Joe）
- 新的捷克语 FAQ
- 修复信息模式视图 `constraint_column_usage` 对外键的支持（Peter）
- ECPG 修复（Michael）
- 修复多个 `IN` 子查询及子查询内连接的缺陷（Tom）
- 允许 `COUNT('x')` 工作（Tom）

- 把 Informix 兼容的 ECPG 头文件安装到单独的目录（Peter）

  一些 Informix 兼容的 ECPG 头文件名与操作系统头文件冲突。通过把它们安装在自己的目录中，减少了名称冲突。

- 修复 SSL 内存泄漏（Neil）

  此发行版修复了 7.4 中 SSL 未释放其分配的全部内存的缺陷。

- 防止 `pg_service.conf` 把服务名用作默认数据库名（Bruce）
- 修复 FreeBSD 上的本地 ident 认证（Tom）

E.202. 版本 7.4

发行日期
2003-11-17

E.202.1. 概述

此发行版的主要变更：

`IN` / `NOT IN` subqueries are now much more efficient

在以前的版本中，`IN`/`NOT IN` 子查询通过顺序扫描子查询寻找匹配来与上层查询连接。7.4 的代码使用与普通连接相同的复杂技术，因此快得多。`IN` 现在通常与等价的 `EXISTS` 子查询一样快或更快；这颠倒了适用于以前版本的传统观念。

Improved `GROUP BY` processing by using hash buckets

在以前的版本中，要分组的行必须先排序。7.4 的代码可以不做排序就完成 `GROUP BY`，方法是把结果累积到每组一个条目的哈希表中。但如果估计哈希表太大放不进 `sort_mem`，它仍会使用排序技术。

New multikey hash join capability

在以前的版本中，哈希连接只能在单键上发生。此版本允许多列哈希连接。

Queries using the explicit `JOIN` syntax are now better optimized

以前的版本只按语法隐含的顺序求值使用显式 `JOIN` 语法的查询。7.4 允许对这些查询做完全优化，即优化器考虑所有可能的连接顺序并选择最高效的。
但外连接仍必须遵循声明的顺序。

Faster and more powerful regular expression code

整个正则表达式模块已被 Henry Spencer 的新版本（最初为 Tcl 编写）取代。该代码大幅改进了性能并支持多种风格的正则表达式。

Function-inlining for simple SQL functions

简单 SQL 函数现在可以通过把其 SQL 包含在主查询中来内联。这通过消除每次调用的开销改进了性能。这意味着简单 SQL 函数现在的行为像宏。

Full support for IPv6 connections and IPv6 address data types

以前的版本只允许 IPv4 连接，IP 数据类型也只支持 IPv4 地址。此版本在这两方面都增加了完整的 IPv6 支持。

Major improvements in SSL performance and reliability

几位非常熟悉 SSL API 的人彻底改造了我们的 SSL 代码，以改进 SSL 密钥协商和错误恢复。

Make free space map efficiently reuse empty index pages, and other free space management improvements

在以前的版本中，因行被删除而变空的 B-tree 索引页只能被索引值与该页上最初索引的行相似的行重用。在 7.4 中，`VACUUM` 记录空索引页并允许它们被未来的任何索引行重用。

SQL-standard information schema

信息模式提供了访问数据库中定义的模式对象信息的标准化且稳定的方式。

Cursors conform more closely to the SQL standard

`FETCH` 和 `MOVE` 命令已被彻底改造，以更贴近 SQL 标准。

Cursors can exist outside transactions

这些游标也称为可保持游标。

New client-to-server protocol

新协议增加了错误码、更多状态信息、更快的启动、对二进制数据传输的更好支持、与 SQL 命令分离的参数值、协议级的预备语句以及从 `COPY` 失败更干净的恢复。服务器和客户端仍支持较旧的协议。

libpq and ECPG applications are now fully thread-safe

虽然以前的 libpq 版本已支持线程，但此版本通过修复数据库连接启动期间使用的一些非线程安全代码改进了线程安全。必须使用 `configure` 选项 `--enable-thread-safety` 才能启用此特性。

New version of full-text indexing

`contrib/tsearch2` 中提供了新的全文索引套件。

New autovacuum tool

`contrib/autovacuum` 中的新 autovacuum 工具监控数据库统计表中的 `INSERT`/`UPDATE`/`DELETE` 活动，并在需要时自动清理表。

Array handling has been improved and moved into the server core

数组的许多限制已被移除，数组的行为更像完全受支持的数据类型。

E.202.2. 迁移到版本 7.4

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

- 服务器端的自动提交设置被移除，并在客户端应用程序和语言中重新实现。
  服务器端自动提交给想要控制自己自动提交行为的语言和应用程序带来了太多问题，
  因此自动提交从服务器中移除并酌情添加到各个客户端 API。
- 此版本中错误消息的措辞发生了实质变化。投入了大量努力使消息更一致、更面向用户。
  如果你的应用程序试图通过解析错误消息来检测不同的错误条件，
  强烈建议改用新的错误码机制。
- 使用显式 `JOIN` 语法的内连接可能行为不同，因为它们现在得到了更好的优化。
- 为清晰起见，许多服务器配置参数已被重命名，主要是与日志相关的那些。
- `FETCH 0` 或 `MOVE 0` 现在不做任何事情。在以前的版本中，`FETCH 0` 会获取所有剩余行，`MOVE 0` 会移到游标末尾。
- `FETCH` 和 `MOVE` 现在返回实际获取/移动的行数，如果位于游标开头/末尾则返回零。以前的版本会返回传给命令的行数，而不是实际获取或移动的行数。
- `COPY` 现在能处理使用回车或回车/换行行尾序列的文件。数据值中不再接受字面回车和换行；请改用 `\r` 和 `\n`。
- 从 `char(n)` 转换为 `varchar(n)` 或 `text` 时现在会修剪尾随空格。这本来就是大多数人一直期望的。
- 数据类型 `float(p)` 现在用二进制位而不是十进制位度量 *p*。新行为遵循 SQL 标准。
- 有歧义的日期值现在必须匹配 `datestyle` 设置指定的顺序。在以前的版本中，即使 `datestyle` 指定日在前，`10/20/03` 的日期规格也会被解释为十月的日期。7.4 将在日期规格对当前 `datestyle` 设置无效时抛出错误。
- 函数 `oidrand`、`oidsrand` 和 `userfntest` 已被移除。这些函数被认定不再有用。
- 指定随时间变化的日期/时间值的字符串字面量（如 `'now'` 或 `'today'`）在列默认表达式中不再按预期工作；它们现在使表创建时刻成为默认值，而不是插入时刻。应改用 `now()`、`current_timestamp` 或 `current_date` 等函数。

  在以前的版本中，有特殊代码使 `'now'` 之类的字符串在 `INSERT` 时而不是表创建时被解释，但这一变通没有覆盖所有情形。7.4 版本现在要求用 `now()` 或 `current_timestamp()` 等函数正确定义默认值。这些在所有情形下都能工作。
- 操作符名中不再允许美元符号（`$`）。它可以用作标识符中的非首字符。
  这是为了改进与其他数据库系统的兼容性，并避免参数占位符（`$n`）
  紧邻操作符书写时的语法问题。

E.202.3. 变更

下面你将找到 7.4 版本与上一个主版本之间变更的详细说明。

E.202.3.1. 服务器操作变更

- 允许 IPv6 服务器连接（Nigel Kukard、Johan Jordaan、Bruce、Tom、Kurt Roeckx、Andrew Dunstan）
- 修复 SSL 以干净地处理错误（Nathan Mueller）

  在以前的版本中，某些 SSL API 错误报告未被正确处理。此版本修复了这些问题。

- SSL 协议安全和性能改进（Sean Chittenden）

  SSL 密钥重协商过于频繁，导致 SSL 性能不佳。此外，初始密钥处理也得到了改进。

- 检测到死锁时打印锁信息（Tom）

  这使调试死锁情形更容易。

- 定期更新 `/tmp` 套接字的修改时间以避免它们被删除（Tom）

  这应有助于防止清理 `/tmp` 目录的管理脚本删除服务器套接字文件。

- 为 Mac OS X 启用 PAM（Aaron Hillegass）

- 使 B-tree 索引完全 WAL 安全（Tom）

  在以前的版本中，某些罕见情形下服务器崩溃可能导致 B-tree 索引损坏。此版本消除了最后几个罕见情形。

- 允许 B-tree 索引压缩和空页重用（Tom）

- 修复首个根页分裂期间索引查找不一致的问题（Tom）

  在以前的版本中，当单页索引分裂为两页时，
  有一小段时间另一个数据库会话可能看不到某个索引条目。此版本修复了这一罕见失败情形。

- 改进空闲空间映射的分配逻辑（Tom）

- 在服务器重启之间保留空闲空间信息（Tom）

  在以前的版本中，postmaster 停止时空闲空间映射不被保存，因此新启动的服务器没有空闲空间信息。此版本保存空闲空间映射，并在服务器重启时重新装载它。

- 为 `pg_stat_activity` 添加开始时间（Neil）

- 检测损坏磁盘页的新代码；用 `zero_damaged_pages` 清除（Tom）

- 新的客户端/服务器协议：更快、无用户名长度限制、允许从 `COPY` 干净退出（Tom）

- 为客户端/服务器协议添加事务状态、表 ID、列 ID（Tom）

- 为客户端/服务器协议添加二进制 I/O（Tom）

- 移除服务器端自动提交设置；移至客户端应用程序（Tom）

- 新的错误消息措辞、错误码和三级错误细节（Tom、Joe、Peter）

E.202.3.2. 性能改进

- 为 `GROUP BY` 聚合添加哈希（Tom）

- 使嵌套循环连接更聪明地使用多列索引（Tom）

- 允许多键哈希连接（Tom）

- 改进常量折叠（Tom）

- 新增内联简单 SQL 函数的能力（Tom）

- 降低使用复杂函数的查询的内存使用（Tom）

  在以前的版本中，返回已分配内存的函数直到查询完成才释放它。
  此版本允许在函数调用完成时释放函数分配的内存，减少了函数使用的总内存。

- 改进 GEQO 优化器性能（Tom）

  此版本修复了 GEQO 优化器管理潜在查询路径方式中的几个低效之处。

- 允许 `IN`/`NOT IN` 通过哈希表处理（Tom）
- 改进 `NOT IN (subquery)` 的性能（Tom）
- 允许大多数 `IN` 子查询作为连接处理（Tom）
- 模式匹配操作可以使用索引而不受区域设置影响（Peter）

  非 ASCII 区域设置无法使用标准索引做 `LIKE` 比较。此版本增加了一种为 `LIKE` 创建特殊索引的方法。

- 允许 postmaster 用 `preload_libraries` 预装载库（Joe）

  对于需要很长时间装载的共享库，此选项使库可以在 postmaster 中预装载并被所有数据库会话继承。

- 改进优化器代价计算，特别是对子查询（Tom）
- 当子查询 `ORDER BY` 与上层查询匹配时避免排序（Tom）
- 从 `WHERE a.x = b.y AND b.y = 42` 推出 `a.x = 42`（Tom）
- 允许复杂连接上的哈希/归并连接（Tom）
- 允许更多数据类型的哈希连接（Tom）
- 允许显式内连接的连接优化，可用 `join_collapse_limit` 禁用（Tom）
- 新增参数 `from_collapse_limit`，控制子查询到连接的转换（Tom）
- 使用来自 Tcl 的更快更强大的正则表达式代码（Henry Spencer、Tom）
- 在优化器中使用位图关系集（Tom）
- 改进连接启动时间（Tom）

  新的客户端/服务器协议启动数据库会话需要更少的网络包。

- 改进触发器/约束的性能（Stephan）
- 改进 `col IN (const, const, const, ...)` 的速度（Tom）
- 修复罕见情形下损坏的哈希索引（Tom）
- 改进哈希索引的并发和速度（Tom）

  以前的版本哈希索引性能不佳，特别是在高并发情形下。此版本修复了这一点，
  开发组有兴趣了解比较 B-tree 与哈希索引性能的报告。

- 把共享缓冲区对齐到 32 字节边界以改进复制速度（Manfred Spraul）

某些 CPU 在地址按 32 字节对齐时执行更快的数据复制。

- 为更好的性能重新实现了 `numeric` 数据类型（Tom）

  `numeric` 以前以 100 为基数存储。新代码使用 10000 为基数，性能显著提升。

E.202.3.3. 服务器配置变更

- 把服务器参数 `server_min_messages` 改名为 `log_min_messages`（Bruce）

  这是为了使大多数控制服务器日志的参数以 `log_` 开头。

- 把 `show_*_stats` 改名为 `log_*_stats`（Bruce）
- 把 `show_source_port` 改名为 `log_source_port`（Bruce）
- 把 `hostname_lookup` 改名为 `log_hostname`（Bruce）
- 新增 `checkpoint_warning`，对过度检查点发出警告（Bruce）

  在以前的版本中，很难确定检查点是否发生得过于频繁。
  此特性在过度检查点发生时向服务器日志添加警告。

- 新的本地化只读服务器参数（Tom）
- 把调试服务器日志消息改为以 `DEBUG` 而不是 `LOG` 输出（Bruce）
- 防止服务器日志变量被非超级用户关闭（Bruce）

  这是一个安全特性，使非超级用户不能禁用管理员启用的日志。

- `log_min_messages`/`client_min_messages` 现在控制 `debug_*` 输出（Bruce）

  这集中了客户端调试信息，使所有调试输出都可以发送到客户端或服务器日志。

- 新增 Mac OS X Rendezvous 服务器支持（Chris Campbell）

  这允许 Mac OS X 主机向网络查询可用的 PostgreSQL 服务器。

- 新增用 `log_min_duration_statement` 只打印慢语句的能力（Christopher）

  这是常被要求的调试特性，允许管理员在服务器日志中只看到慢查询。

- 允许 `pg_hba.conf` 接受 CIDR 格式的网络掩码（Andrew Dunstan）

  这允许管理员把主机 IP 地址和网络掩码字段合并为 `pg_hba.conf` 中的单个 CIDR 字段。

- 新的只读参数 `is_superuser`（Tom）
- 新参数 `log_error_verbosity`，控制错误细节（Tom）

  它与新的错误报告特性一起工作，提供提示、文件名和行号等额外错误信息。

- `postgres --describe-config` 现在转储服务器配置变量（Aizaz Ahmed、Peter）

  此选项对需要知道配置变量名及其最小值、最大值、默认值和描述的管理工具很有用。

- 为 pg_settings 新增列：context、type、source、min_val、max_val（Joe）
- 如果可能，把默认 shared_buffers 设为 1000、max_connections 设为 100（Tom）

  以前的版本默认 64 个共享缓冲区，以便 PostgreSQL 甚至能在很老的系统上启动。此版本测试平台允许的共享内存量并在可能时选择更合理的默认值。当然，
  仍鼓励用户评估其资源负载并相应地设定 shared_buffers 的大小。
- 新的 pg_hba.conf 记录类型 hostnossl，阻止 SSL 连接（Jon Jensen）

  在以前的版本中，如果客户端和服务器都支持 SSL，就没有办法阻止 SSL 连接。此选项提供了该能力。
- 移除参数 geqo_random_seed（Tom）
- 新增服务器参数 regex_flavor，控制正则表达式处理（Tom）
- 让 pg_ctl 更好地处理非标准端口（Greg）

E.202.3.4. 查询变更

- 新的 SQL 标准信息模式（Peter）
- 新增只读事务（Peter）
- 在外键违反消息中打印键名和值（Dmitry Tkach）
- 允许用户在 pg_stat_activity 中看到自己的查询（Kevin Brown）

  在以前的版本中，只有超级用户能用 pg_stat_activity 看到查询字符串。现在普通用户可以看到自己的查询字符串。
- 修复子查询中的聚合以符合 SQL 标准（Tom）

  SQL 标准规定，如果嵌套子查询中出现的聚合函数的参数只包含外层查询变量，则它属于外层查询。以前的 PostgreSQL 版本没有正确处理这个细节。
- 新增阻止自动添加查询中引用的表的选项（Nigel J. Andrews）

  默认情况下，查询中提到的表如果不在 FROM 子句中会被自动添加到其中。这与历史 POSTGRES 行为兼容但违背 SQL 标准。此选项允许选择标准兼容的行为。
- 允许 UPDATE ... SET col = DEFAULT（Rod）

  这允许 UPDATE 把列设为其声明的默认值。
- 允许在 LIMIT/OFFSET 中使用表达式（Tom）

  在以前的版本中，LIMIT/OFFSET 只能使用常数，不能用表达式。
- 实现 CREATE TABLE AS EXECUTE（Neil、Peter）

E.202.3.5. 对象操作变更

- 使 CREATE SEQUENCE 语法更符合 SQL:2003（Neil）
- 新增语句级触发器（Neil）

虽然这允许触发器在语句结束时触发，但不允许触发器访问该语句修改的所有行。此能力计划在未来版本中提供。

- 为域新增检查约束（Rod）

  这通过允许域使用检查约束大幅提高了域的用处。

- 新增 `ALTER DOMAIN`（Rod）

  这允许操纵现有域。

- 修复几个零列表缺陷（Tom）

  PostgreSQL 支持零列表。这修复了使用此类表时发生的各种缺陷。

- 让 `ALTER TABLE ... ADD PRIMARY KEY` 添加非空约束（Rod）

  在以前的版本中，`ALTER TABLE ... ADD PRIMARY` 会添加唯一索引但不添加非空约束。此版本修复了这一点。

- 新增 `ALTER TABLE ... WITHOUT OIDS`（Rod）

  这允许控制新行和更新的行是否有 OID 列。这对节省存储空间最有用。

- 新增 `ALTER SEQUENCE`，修改最小值、最大值、增量、缓存、循环值（Rod）

- 新增 `ALTER TABLE ... CLUSTER ON`（Alvaro Herrera）

  此命令被 `pg_dump` 用来记录每个先前聚簇过的表的聚簇列。该信息被数据库范围的 cluster 用来聚簇所有先前聚簇过的表。

- 改进域的自动类型转换（Rod、Tom）

- 允许标识符中的美元符号，但不能作为首字符（Tom）

- 不允许操作符名中的美元符号，使 `x=$1` 可行（Tom）

- 允许用 `LIKE` *`subtable`* 复制表模式，还有 SQL:2003 特性 `INCLUDING DEFAULTS`（Rod）

- 为 `GRANT` 新增 `WITH GRANT OPTION` 子句（Peter）

  这使 `GRANT` 能授予其他用户在对象上授予权限的能力。

E.202.3.6. 工具命令变更

- 为 `CREATE TABLE` 的临时表新增 `ON COMMIT` 子句（Gavin）

  这增加了在事务提交时删除表或删除所有行的能力。

- 允许用 `WITH HOLD` 在事务之外使用游标（Neil）

  在以前的版本中，游标在创建它的事务结束时被删除。现在可以用 `WITH HOLD` 选项创建游标，这允许在创建事务提交后继续访问它们。

- `FETCH 0` 和 `MOVE 0` 现在不做任何事情（Bruce）

  在以前的版本中，`FETCH 0` 获取所有剩余行，`MOVE 0` 移到游标末尾。

- 按照 SQL 标准，让 FETCH 和 MOVE 返回获取/移动的行数，如果位于游标开头/末尾则返回零（Bruce）

  在以前的版本中，FETCH 和 MOVE 返回的行数没有准确反映处理的行数。

- 正确处理游标的 SCROLL，否则报告错误（Neil）

  对某些类型的查询允许随机访问（前向和后向滚动）需要一些额外的工作才能做到。
  如果创建游标时指定了 SCROLL，将执行这些额外工作。此外，如果游标以 NO SCROLL 创建，则不允许随机访问。

- 为 FETCH 和 MOVE 实现 SQL 兼容的选项 FIRST、LAST、ABSOLUTE *n*、RELATIVE *n*（Tom）

- 允许对 DECLARE CURSOR 使用 EXPLAIN（Tom）

- 允许 CLUSTER 默认使用标记为预聚簇的索引（Alvaro Herrera）

- 允许 CLUSTER 聚簇所有表（Alvaro Herrera）

  这允许用单条命令重新聚簇数据库中所有先前聚簇过的表。

- 禁止对部分索引使用 CLUSTER（Tom）

- 允许 COPY 文件中的 DOS 和 Mac 行尾（Bruce）

- 不允许字面回车作为数据值，反斜杠-回车和 \r 仍被允许（Bruce）

- COPY 变更（二进制、\.）（Tom）

- 从 COPY 失败中干净恢复（Tom）

- 防止 COPY 中可能的内存泄漏（Tom）

- 使 TRUNCATE 事务安全（Rod）

  TRUNCATE 现在可以在事务内使用。如果事务中止，TRUNCATE 所做的更改会自动回滚。

- 允许预备/绑定 FETCH 和 EXPLAIN 等工具命令（Tom）

- 新增 EXPLAIN EXECUTE（Neil）

- 通过减少 WAL 流量改进索引上的 VACUUM 性能（Tom）

- 函数索引已被推广为表达式上的索引（Tom）

  在以前的版本中，函数索引只支持应用于一个或多个列名的简单函数。
  此版本允许任何类型的标量表达式。

- 让 SHOW TRANSACTION ISOLATION 与 SET TRANSACTION ISOLATION 的输入匹配（Tom）

- 让对非本地数据库的 COMMENT ON DATABASE 生成警告而不是错误（Rod）

  数据库注释存储在数据库本地表中，因此对数据库的注释必须存储在每个数据库中。

- 改进 LISTEN/NOTIFY 的可靠性（Tom）

- 允许 REINDEX 可靠地重新索引非共享系统目录索引（Tom）

  这允许重新索引系统表而不需要独立会话，这在以前的版本中是必需的。
  现在唯一需要独立会话来重新索引的是全局系统表 pg_database、pg_shadow 和 pg_group。

E.202.3.7. 数据类型和函数变更

- 新服务器参数 extra_float_digits，控制浮点数的精度显示（Pedro Ferreira、Tom）

  这控制导致回归测试问题的输出精度。

- 允许 +1300 作为数字时区说明符，用于 FJST（Tom）

- 移除很少使用的函数 oidrand、oidsrand 和 userfntest（Neil）

- 把 md5() 函数加入主服务器，此前已在 contrib/pgcrypto 中（Joe）

  MD5 函数是常被要求的。需要更复杂的加密能力时，请使用 contrib/pgcrypto。

- 扩大 timestamp 的日期范围（John Cochran）

- 更改 EXTRACT(EPOCH FROM timestamp)，使 timestamp without time zone 被假定处于本地时间而不是 GMT（Tom）

- 以防操作系统不防止除以零，捕获它（Tom）

- 把 numeric 数据类型内部改为以 10000 为基数（Tom）

- 新函数 hostmask()（Greg Wickham）

- to_char() 和 to_timestamp() 的修复（Karel）

- 允许接受任意参数数据类型并返回任意数据类型的函数，使用 anyelement 和 anyarray（Joe）

  这允许创建能用于任何数据类型的函数。

- 数组现在可以指定为 ARRAY[1,2,3]、ARRAY[['a','b'],['c','d']] 或 ARRAY[ARRAY[ARRAY[2]]]（Joe）

- 允许数组的正确比较，包括 ORDER BY 和 DISTINCT 支持（Joe）

- 允数组列上的索引（Joe）

- 允许用 || 连接数组（Joe）

- 允许 WHERE 限定 *expr op* ANY/SOME/ALL (*array_expr*)（Joe）

  这允许数组表现得像值列表，用于诸如 SELECT * FROM tab WHERE col IN (array_val) 之类的用途。

- 新数组函数 array_append、array_cat、array_lower、array_prepend、array_to_string、array_upper、string_to_array（Joe）

- 允许用户定义的聚合使用多态函数（Joe）

- 允许给空数组赋值（Joe）
- 允许 `time`、`timestamp` 和 `interval` 输入值的秒字段为 60（Tom）

  闰秒需要六十秒的值。
- 允许 `cidr` 数据类型转换为 `text`（Tom）
- 不允许 SET TIMEZONE 中的无效时区名
- 当 `char` 转换为 `varchar` 或 `text` 时修剪尾随空格（Tom）
- 让 `float(p)` 用二进制位而不是十进制位度量精度 *p*（Tom）
- 为 `inet` 和 `cidr` 数据类型添加 IPv6 支持（Michael Graff）
- 新增 `family()` 函数，报告地址是 IPv4 还是 IPv6（Michael Graff）
- 让 `SHOW datestyle` 生成与 `SET datestyle` 所用类似的输出（Tom）
- 让 `EXTRACT(TIMEZONE)` 和 `SET/SHOW TIME ZONE` 遵循 SQL 对时区偏移符号的约定，即正值为 UTC 以东（Tom）
- 修复 `date_trunc('quarter', ...)`（Böjthe Zoltán）

  以前的版本对此函数调用返回不正确的值。
- 使 `initcap()` 更兼容 Oracle（Mike Nolan）

  `initcap()` 现在把任何非字母数字字符之后出现的字母大写，而不只是空白之后。
- 对非 ISO-8601 格式的日期值只允许 `datestyle` 字段顺序（Greg）
- 新增 `datestyle` 值 `MDY`、`DMY` 和 `YMD` 来设置输入字段顺序；为向后兼容保留 `US` 和 `European`（Tom）
- 像 `'now'` 或 `'today'` 这样的字符串字面量将不再可用作列默认值。请改用 `now()`、`current_timestamp` 等函数。（预备语句需要此变更）（Tom）
- 在 `min()`/`max()` 中把 NaN 当作大于任何其他值（Tom）

  在大多数用途中 NaN 已排在普通数值之后，但 `min()` 和 `max()` 没有正确处理。
- 防止 interval 抑制 `:00` 秒的显示
- 新函数 `pg_get_triggerdef(prettyprint)` 和 `pg_conversion_is_visible()`（Christopher）
- 允许时间指定为 `040506` 或 `0405`（Tom）
- 输入日期顺序现在必须是 `YYYY-MM-DD`（4 位年份）或匹配 `datestyle`
- 让 `pg_get_constraintdef` 支持唯一、主键和检查约束（Christopher）

E.202.3.8. 服务器端语言变更

- 防止对零行记录变量使用 `RETURN NEXT` 时 PL/pgSQL 崩溃（Tom）

- 让 PL/Python 的 `spi_execute` 接口正确处理空值（Andrew Bosma）
- 允许 PL/pgSQL 不用 `%ROWTYPE` 声明组合类型的变量（Tom）
- 修复 PL/Python 的 `_quote()` 函数以处理大整数
- 把 PL/Python 改为不可信语言，现名 `plpythonu`（Kevin Jacobs、Tom）

  Python 语言不再支持受限执行环境，因此删除了可信版 PL/Python。如果这种情况改变，将重新添加可供非超级用户使用的 PL/Python 版本。
- 允许多态 PL/pgSQL 函数（Joe、Tom）
- 允许多态 SQL 函数（Joe）
- 改进了 PL/pgSQL 的编译函数缓存机制，完全支持多态（Joe）
- 在 PL/pgSQL 中新增表示函数实际返回类型的新参数 `$0`（Joe）
- 允许 PL/Tcl 和 PL/Python 在多个表上使用同一触发器（Tom）
- 修复 PL/Tcl 的 `spi_prepare`，使其接受参数类型列表中完全限定的类型名（Jan）

E.202.3.9. psql 变更

- 新增 `\pset pager always`，总是使用分页器（Greg）

  这强制使用分页器，即使行数少于屏幕高度。这对跨多个屏幕行的行很有价值。
- 改进 Tab 补全（Rod、Ross Reedstrom、Ian Barwick）
- 把 `\?` 帮助重排为分组（Harald Armin Massa、Bruce）
- 新增列出模式、转换和编码转换的反斜杠命令（Christopher）
- `\encoding` 现在根据服务器参数 `client_encoding` 变化（Tom）

  在以前的版本中，`\encoding` 不知道用 `SET client_encoding` 所做的编码更改。
- 把编辑器缓冲区保存到 readline 历史（Ross）

  当用 `\e` 编辑查询时，结果被保存到 readline 历史中，可用上箭头检索。
- 改进 `\d` 显示（Christopher）
- 增强 HTML 模式使其更符合标准（Greg）
- 新的 `\set AUTOCOMMIT off` 能力（Tom）

  这取代了被移除的服务器参数 `autocommit`。
- 新的 `\set VERBOSITY`，控制错误细节（Tom）

  这控制新的错误报告细节。
- 新的提示符转义序列 `%x`，显示事务状态（Tom）
- psql 的长选项现在在所有平台上可用

E.202.3.10. pg_dump 变更

- 多个 pg_dump 修复，包括 tar 格式和大对象
- 允许 pg_dump 转储特定的模式（Neil）
- 让 pg_dump 保留列存储特性（Christopher）

  这保留 `ALTER TABLE ... SET STORAGE` 信息。
- 让 pg_dump 保留 `CLUSTER` 特性（Christopher）
- 让 pg_dumpall 用 `GRANT/REVOKE` 转储数据库级权限（Tom）
- 允许 pg_dumpall 支持 pg_dump 的选项 `-a`、`-s`、`-x`（Tom）
- 防止 pg_dump 把命令行上指定的标识符转为小写（Tom）
- pg_dump 选项 `--use-set-session-authorization` 和 `--no-reconnect` 现在无作用，所有转储都使用 `SET SESSION AUTHORIZATION`

  pg_dump 不再重新连接来切换用户，而是始终使用 `SET SESSION AUTHORIZATION`。这将减少恢复期间的口令提示。
- pg_dump 的长选项现在在所有平台上可用

  PostgreSQL 现在包含自己的长选项处理例程。

E.202.3.11. libpq 变更

- 新增用于在 Windows 上释放内存的函数 `PQfreemem`，建议用于 `NOTIFY`（Bruce）

  Windows 要求库中分配的内存由同一库中的函数释放，因此 `free()` 不能释放 libpq 分配的内存。`PQfreemem` 是释放 libpq 内存的正确方式，特别是在 Windows 上，也建议在其他平台上使用。
- 记录服务能力并添加示例文件（Bruce）

  这允许客户端在客户端机器上的中心文件中查找连接信息。
- 让 `PQsetdbLogin` 与 `PQconnectdb` 有相同的默认值（Tom）
- 允许 libpq 在结果集过大时干净地失败（Tom）
- 改进函数 `PQunescapeBytea` 的性能（Ben Lamb）
- 用 `configure` 选项 `--enable-thread-safety` 允许线程安全的 libpq（Lee Kindness、Philip Yarra）
- 允许函数 `pqInternalNotice` 接受格式串和参数而不只是预先格式化的消息（Tom、Sean Chittenden）
- 用 `sslmode` 值 `disable`、`allow`、`prefer` 和 `require` 控制 SSL 协商（Jon Jensen）
- 允许新的错误码和文本级别（Tom）
- 允许访问查询结果的基础表和列（Tom）

这对想知道与特定结果集关联的基础表和列名的查询构建器应用程序很有帮助。

- 允许访问当前事务状态（Tom）
- 新增把二进制数据直接传递给服务器的能力（Tom）
- 新增函数 `PQexecPrepared` 和 `PQsendQueryPrepared`，执行先前预备语句的绑定/执行（Tom）

E.202.3.12. JDBC 变更

- 允许可更新结果集上的 `setNull`
- 允许预备语句上的 `executeBatch`（Barry）
- 支持 SSL 连接（Barry）
- 处理结果集中的模式名（Paul Sorenson）
- 新增 refcursor 支持（Nic Ferrier）

E.202.3.13. 其他接口变更

- 防止 libpgtcl 关闭期间可能的内存泄漏或核心转储（Tom）
- 为 ECPG 新增 Informix 兼容性（Michael）

  这允许 ECPG 处理使用某些 Informix 扩展编写的嵌入式 C 程序。
- 为 ECPG 新增定长的 `decimal` 类型，用于 Informix（Michael）
- 用 `configure` 选项 `--enable-thread-safety` 允许线程安全的嵌入式 SQL 程序（Lee Kindness、Bruce）

  这允许多个线程同时访问数据库。
- 把 Python 客户端 PyGreSQL 移到http://www.pygresql.org（Marc）

E.202.3.14. 源代码变更

- 消除对单独的平台几何回归结果文件的需要（Tom）
- 改进的 PPC 锁定原语（Reinhard Max）
- 新函数 `palloc0`，分配并清零内存（Bruce）
- 修复 s390x CPU（64 位）的锁定代码（Tom）
- 允许 OpenBSD 使用本地 ident 凭据（William Ahern）
- 使查询计划树对执行器只读（Tom）
- 新增 Darwin 启动脚本（David Wheeler）
- 允许 libpq 用 Borland C++ 编译器编译（Lester Godwin、Karl Waclawek）

- 需要时使用我们自己的 `getopt_long()` 版本（Peter）
- 把管理脚本转换为 C（Peter）
- 如果从 CVS 构建，现在需要 Bison >= 1.85 来构建 PostgreSQL 语法
- 把文档合并为一本书（Peter）
- 新增 Windows 兼容函数（Bruce）
- 允许客户端接口在 MinGW 下编译（Bruce）
- 新的用于错误报告的 `ereport()` 函数（Tom）
- 支持 Linux 上的 Intel 编译器（Peter）
- 改进 Linux 启动脚本（Slawomir Sudnik、Darko Prenosil）
- 新增对 AMD Opteron 和 Itanium 的支持（Jeffrey W. Baker、Bruce）
- 从 `configure` 移除 `--enable-recode` 选项

  既然有了 `CREATE CONVERSION`，这不再需要。
- 如果未找到自旋锁代码则产生编译错误（Bruce）

  没有自旋锁代码的平台现在将编译失败，而不是静默使用信号量。此失败可以用新的 `configure` 选项禁用。

E.202.3.15. 附加模块变更

- 把 dbmirror 许可证改为 BSD
- 改进 earthdistance（Bruno Wolff III）
- pgcrypto 的可移植性改进（Marko Kreen）
- 防止 xml 中的崩溃（John Gray、Michael Richards）
- 更新 oracle
- 更新 mysql
- 更新 cube（Bruno Wolff III）
- 更新 earthdistance 以使用 cube（Bruno Wolff III）
- 更新 btree_gist（Oleg）
- 新的 tsearch2 全文搜索模块（Oleg、Teodor）
- 为 tablefuncs 新增基于哈希的 crosstab 函数（Joe）
- 为 tablefuncs 新增用来对 `connectby()` 兄弟节点排序的 serial 列（Nabil Sayegh、Joe）
- 为 dblink 新增命名持久连接（Shridhar Daithanka）

- 新的 pg_autovacuum 允许自动 `VACUUM`（Matthew T. O'Connor）
- 让 pgbench 遵循环境变量 `PGHOST`、`PGPORT`、`PGUSER`（Tatsuo）
- 改进 intarray（Teodor Sigaev）
- 改进 pgstattuple（Rod）
- 修复 fuzzystrmatch 中 `metaphone()` 的缺陷
- 改进 adddepend（Rod）
- 更新 spi/timetravel（Böjthe Zoltán）
- 修复 dbase 的 `-s` 选项并改进非 ASCII 处理（Thomas Behr、Márcio Smiderle）
- 移除 array 模块，因为其功能现已默认包含（Joe）

E.203. 版本 7.3.21

发行日期
2008-01-07

此发行版包含 7.3.20 之后的多种修复，包括对重要安全问题的修复。

这预计是 PostgreSQL 7.3.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.203.1. 迁移到版本 7.3.21

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.203.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  7.3.20 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.204. 版本 7.3.20

发行日期
2007-09-17

此发行版包含 7.3.19 之后的修复。

E.204.1. 迁移到版本 7.3.20

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.204.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.205. 版本 7.3.19

发行日期
2007-04-23

此发行版包含 7.3.18 之后的修复，包括一个安全修复。

E.205.1. 迁移到版本 7.3.19

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.205.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）

E.206. 版本 7.3.18

发行日期
2007-02-05

此发行版包含 7.3.17 之后的多种修复，包括一个安全修复。

E.206.1. 迁移到版本 7.3.18

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.206.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  该漏洞涉及更改 SQL 函数中使用的表列的数据类型（CVE-2007-0555）。此错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.207. 版本 7.3.17

发行日期
2007-01-08

此发行版包含 7.3.16 之后的多种修复。

E.207.1. 迁移到版本 7.3.17

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.207.2. 变更

- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。

- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。

E.208. 版本 7.3.16

发行日期
2006-10-16

此发行版包含 7.3.15 之后的多种修复。

E.208.1. 迁移到版本 7.3.16

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.208.2. 变更

- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- 向后移植 7.4 自旋锁代码，以改进性能并更好地支持 64 位架构
- 修复 libpq 中与 SSL 相关的内存泄漏
- 修复 /contrib/dbmirror 中的反斜杠转义
- 针对美国 DST 法规的最近变更调整回归测试

E.209. 版本 7.3.15

发行日期
2006-05-23

此发行版包含 7.3.14 之后的多种修复，包括针对极其严重安全问题的补丁。

E.209.1. 迁移到版本 7.3.15

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.209.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）

- 修复各种小的内存泄漏

E.210. 版本 7.3.14

发行日期
2006-02-14

此发行版包含 7.3.13 之后的多种修复。

E.210.1. 迁移到版本 7.3.14

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅第 E.211 节。

E.210.2. 变更

- 修复 `SET SESSION AUTHORIZATION` 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复以允许恢复有对自定义操作符跨模式引用的转储（Tom）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.211. 版本 7.3.13

发行日期
2006-01-09

此发行版包含 7.3.12 之后的多种修复。

E.211.1. 迁移到版本 7.3.13

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅第 E.214 节。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.211.2. 变更

- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）
- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.212. 版本 7.3.12

发行日期
2005-12-12

此发行版包含 7.3.11 之后的多种修复。

E.212.1. 迁移到版本 7.3.12

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅第 E.214 节。

E.212.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- `/contrib/ltree` 修复（Teodor）
- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。

- 防止表被删除后 pg_autovacuum 的核心转储

E.213. 版本 7.3.11

发行日期
2005-10-04

此发行版包含 7.3.10 之后的多种修复。

E.213.1. 迁移到版本 7.3.11

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅第 E.214 节。

E.213.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。

- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。

- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- 各种内存泄漏修复
- 各种可移植性改进
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`

E.214. 版本 7.3.10

发行日期
2005-05-09

此发行版包含 7.3.9 之后的多种修复，包括几个与安全相关的问题。

E.214.1. 迁移到版本 7.3.10

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，它是处理在 7.3.X 系统目录初始内容中发现的严重安全问题的可能方法之一。用 7.3.10 的 initdb 做转储/initdb/重载序列会自动纠正此问题。

安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）强烈建议所有安装通过 initdb 或按照下述手工修复程序修复此错误。该错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行以下程序。作为数据库超级用户，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[3] = 'internal'::regtype
WHERE pronamespace = 11 AND pronargs = 5
     AND proargtypes[2] = 'cstring'::regtype;
-- The command should report having updated 90 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

上述程序必须在安装的每个数据库中执行，包括 `template1`，理想情况下也包括 `template0`。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。`template1` 可以像任何其他数据库一样修复，但修复 `template0` 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 `template0` 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.214.2. 变更

- 更改编码函数签名以防止误用
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

  此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。
- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化

E.215. 版本 7.3.9

发行日期
2005-01-31

此发行版包含 7.3.8 之后的多种修复，包括几个与安全相关的问题。

E.215.1. 迁移到版本 7.3.9

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.215.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。

- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 plperl 处理元组字段中引号的问题
- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示

E.216. 版本 7.3.8

发行日期
2004-10-22

此发行版包含 7.3.7 之后的多种修复。

E.216.1. 迁移到版本 7.3.8

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.216.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 make_oidjoins_check 在 /tmp 中使用临时文件

  这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

E.217. 版本 7.3.7

发行日期
2004-08-16

此发行版在 7.3.6 的基础上包含一个关键修复和一些次要项。

E.217.1. 迁移到版本 7.3.7

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.217.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 移除 tsearch 中不对称的词处理（Teodor）

- 在 pg_dump 转储 CAST 时正确地对函数名加模式限定

E.218. 版本 7.3.6

发行日期
2004-03-02

此发行版包含 7.3.5 之后的多种修复。

E.218.1. 迁移到版本 7.3.6

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.218.2. 变更

- 回退规则权限检查中的错误变更

7.3.3 中为修复规则权限检查极端情形而应用的补丁结果在许多并非极端的情形中禁用了与规则相关的权限检查。这例如会允许用户向他们本无权插入的视图插入。因此我们回退了 7.3.3 的补丁。原始缺陷将在 8.0 中修复。

- 修复 GetNewTransactionId() 中错误的操作顺序

  此缺陷可能导致磁盘空间耗尽条件下失败，包括释放磁盘空间后也无法重启。

- 确保即使提供了外部 CFLAGS 值，configure 也选择 -fno-strict-aliasing

  在某些平台上，用 -fstrict-aliasing 构建会导致缺陷。

- 让 pg_restore 正确处理 64 位 off_t

  此缺陷妨碍了从超过 4 GB 的归档文件正确恢复。

- 让 contrib/dblink 不假定本地和远程类型 OID 匹配（Joe）
- 正确引用 connectby() 的 start_with 参数（Joe）
- 当 plpgsql 函数的行类型参数为 NULL 时不崩溃
- 避免在规划 LIKE 操作时的极端情形生成无效字符编码序列
- 确保 text_position() 在多字节情形不能扫描越过源字符串末尾（Korea PostgreSQL Users' Group）
- 修复 bytea 列上 LIKE 操作的索引优化和选择率估计（Joe）

E.219. 版本 7.3.5

发行日期
2003-12-03

这包含 7.3.4 之后的多种修复。

E.219.1. 迁移到版本 7.3.5

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.219.2. 变更

- 强制在从 WAL 恢复期间开启 zero_damaged_pages
- 防止“变量不在子计划目标列表中”的一些晦涩情形
- 强制统计进程从共享内存脱离，确保更干净的关闭
- 使 PQescapeBytea 与 byteaout 相互一致（Joe）
- 为 dblink 的 get_tuple_of_interest() 添加缺失的 SPI_finish() 调用（Joe）

- 修复规则重写 INSERT 时可能的外键违反（Jan）
- 支持 PL/Tcl 的 spi_prepare 命令中限定类型名（Jan）
- 让 pg_dump 处理位于 pg_catalog 中的过程语言处理器
- 让 pg_dump 处理自定义操作符类在另一个模式中的情形
- 让 pg_dump 正确转储二进制兼容的转换（Jan）
- 修复把包含子查询的表达式插入规则体的问题
- 修复 clusterdb 脚本中不正确的参数处理（Anand Ranganathan）
- 修复 plpython 触发器中已删除列的问题
- 修复 to_char() 读取越过输入字符串末尾的问题（Karel）
- 修复 GB18030 映射错误（Tatsuo）
- 修复 SSL 错误处理和异步 SSL I/O 的几个问题
- 移除 JDBC 中把值列表绑定到单个参数的能力（防止可能的 SQL 注入攻击）
- 修复 HAVE_INT64_TIMESTAMP 代码路径中的一些错误
- 修复与首个根页分裂并行的 btree 搜索的极端情形

E.220. 版本 7.3.4

发行日期
2003-07-24

这包含 7.3.3 之后的多种修复。

E.220.1. 迁移到版本 7.3.4

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.220.2. 变更

- 修复 2000 年之前日期的时间戳到日期转换的损坏
- 防止服务器启动失败的罕见可能（Tom）
- 修复 interval 到 time 转换中的缺陷（Tom）
- 在 pg_dump 的几处添加约束名（Rod）
- 改进多参数函数的性能（Tom）
- 修复 to_ascii() 缓冲区溢出（Tom）

- 防止恢复数据库注释时抛出错误（Tom）
- 绕过某些 Solaris 版本中存在缺陷的 strxfrm()（Tom）
- 正确转义 jdbc setObject() 字符串以改进安全（Barry）

E.221. 版本 7.3.3

发行日期
2003-05-22

此发行版包含 7.3.2 版本的多种修复。

E.221.1. 迁移到版本 7.3.3

对于运行 7.3.* 版本的用户，不需要进行转储/恢复。

E.221.2. 变更

- 修复崩溃后 StartUpID 有时不正确的计算
- 避免一个事务中大量延迟触发器的缓慢（Stephan）
- 当 `UPDATE` 未更改外键值时不锁定被引用行（Jan）
- 在 Sparc 上使用 `-fPIC` 而不是 `-fpic`（Tom Callaway）
- 修复 contrib/reindexdb 缺乏模式感知的问题
- 修复 contrib/intarray 对零元素结果数组的错误（Teodor）
- 确保 createuser 脚本在 control-C 时退出（Oliver）
- 修复已删除列的类型本身也被删除时的错误
- `CHECKPOINT` 在非关键步骤失败时不再导致数据库 panic
- 接受 timestamp、time、interval 输入值秒字段中的 60
- 当 `TIMESTAMP`、`TIME` 或 `INTERVAL` 精度过大时发出通知而不是错误
- 修复 `abstime-to-time` 转换函数（除非 initdb，否则不应用修复）
- 修复 `timestampt_izone` 的 pg_proc 条目（除非 initdb，否则不应用修复）
- 让 `EXTRACT(EPOCH FROM timestamp without time zone)` 把输入视为本地时间
- 如果事务早些时候更改了时区，`'now'::timestamptz` 给出错误答案
- 带时区 time 的 `HAVE_INT64_TIMESTAMP` 代码覆写了其输入
- 接受 `GLOBAL TEMP/TEMPORARY` 作为 `TEMPORARY` 的同义词

- 避免外键触发器中不当的模式权限检查失败
- 修复 `SET DEFAULT` 动作的外键触发器中的缺陷
- 修复 `UPDATE` 和 `DELETE` 触发器行获取中不正确的 time-qual 检查
- `ALTER TABLE ADD COLUMN` 中外键子句被解析但被忽略
- 修复处理函数已存在时 createlang 脚本的损坏
- 修复 pg_dump、COPY、ANALYZE 等处对零列表的行为异常
- 修复 `func_error()` 对包含 '%' 的类型名的行为异常
- 修复 `replace()` 对包含 '%' 的字符串的行为异常
- 包含某些多字节字符的正则表达式模式失败
- 在连接大小估计中在更多情形正确考虑 `NULL`
- 避免与 `isblank()` 函数或宏的系统定义冲突
- 修复 EUC_TW 转换中未能转换大码点值的问题（Tatsuo）
- 修复 `SSL_read`/`SSL_write` 调用的错误恢复
- 不提前做类型强制表达式的常量折叠
- 读入任何页后立即验证页头字段
- 修复未命名连接中对未分组变量的不正确检查
- 修复 `to_ascii` 中的缓冲区溢出（Guido Notari）
- contrib/ltree 修复（Teodor）
- 修复 char 为无符号的机器上死锁检测的核心转储
- 避免多路索引扫描耗尽缓冲区（7.3 中引入的缺陷）
- 修复规划器的选择率估计函数以正确处理域
- 修复 dbmirror 内存分配缺陷（Steven Singer）
- 防止 `ln(numeric)` 因舍入误差无限循环
- 如果存在多个相等的 GROUP BY 项，`GROUP BY` 会混乱
- 修复继承的 `UPDATE`/`DELETE` 引用另一个继承表时的坏计划
- 防止在不完整（部分或不存储 NULL）的索引上聚簇
- 如果关闭请求在启动期间到达，在正确时刻处理它
- 修复临时索引中的左链接（可能使反向扫描漏掉条目）
- 修复 postgresql.conf 中 client_encoding 设置的不正确处理（Tatsuo）

- 修复 Async_NotifyHandler 运行后不响应 `pg_ctl stop -m fast` 的问题
- 修复规则包含多个同类型语句情形的 SPI
- 修复规则查询中检查错误类型访问权限的问题
- 修复 `CREATE RULE` 中 `EXCEPT` 的问题
- 防止删除带 serial 列的临时表的问题
- 修复复杂视图中 replace_vars_with_subplan_refs 的失败
- 修复单字节编码中正则表达式的缓慢（Tatsuo）
- 允许 `CREATE CAST` 和 `DROP CAST` 中的限定类型名
- 接受 `SETOF type[]`，以前必须写成 `SETOF _type`
- 修复某些带过程语言情形下 pg_dump 的核心转储
- 为可移植性在 pg_dump 输出中强制 ISO datestyle（Oliver）
- pg_dump 未能处理 `lo_read` 的错误返回（Oleg Drokin）
- pg_dumpall 对没有成员的组失败（Nick Eskelinen）
- pg_dumpall 未能识别 --globals-only 开关
- 如果指定了 -X disable-triggers，pg_restore 未能恢复 blob
- 修复 plpgsql 中的函数内内存泄漏
- 如果给定错误参数，pltcl 的 `elog` 命令转储核心（Ian Harding）
- plpython 使用了错误的 `atttypmod` 值（Brad McLean）
- 修复 Python 接口中布尔值的不当引用（D'Arcy）
- 为 JDBC 的 PGConnection 接口添加 `addDataType()` 方法
- 修复 JDBC 可更新 ResultSet 的各种问题（Shawn Green）
- 修复 JDBC DatabaseMetaData 的各种问题（Kris Jurka、Peter Royal）
- 修复 JDBC 中解析表 ACL 的问题
- JDBC 中字符集转换问题的更好错误消息

E.222. 版本 7.3.2

发行日期
2003-02-04

此发行版包含 7.3.1 版本的多种修复。

E.222.1. 迁移到版本 7.3.2

对于运行 7.3.* 版本的用户，不需要进行转储/恢复。

E.222.2. 变更

- 恢复 CREATE TABLE AS / SELECT INTO 中 OID 列的创建
- 修复转储带注释视图时 pg_dump 的核心转储
- 正确转储 DEFERRABLE/INITIALLY DEFERRED 约束
- 修复子表列编号与父表不同时的 UPDATE
- 提高 max_fsm_relations 的默认值
- 修复单行查询游标反向获取时的问题
- 让 SELECT DISTINCT 查询游标的反向获取正确工作
- 修复装载包含 contrib/lo 用法的 pg_dump 文件的问题
- 修复全数字用户名的问题
- 修复 libpgtcl 断开连接期间可能的内存泄漏和核心转储
- 让 plpython 的 spi_execute 命令正确处理空值（Andrew Bosma）
- 调整 plpython 错误报告使其回归测试再次通过
- 与 bison 1.875 一起工作
- 正确处理 plpgsql %type 中的大小写混合名（Neil）
- 修复执行被规则重写的查询时 pltcl 的核心转储
- 修复数组下标越界（据 Yichen Xie 报告）
- 把浮点情形的 MAX_TIME_PRECISION 从 13 减到 10
- 在逐数据库和逐用户设置中正确折叠变量名的大小写
- 修复 SELECT INTO 记录不返回行时 plpgsql RETURN NEXT 的核心转储
- 修复 python 客户端接口中对 pg_type.typprtlen 的过时使用
- 正确处理 JDBC 驱动中时间戳的小数秒
- 改进 JDBC 中 getImportedKeys() 的性能
- 使共享库符号链接在 HPUX 上按标准工作（Giles）
- 修复 timestamp、time、interval 不一致的舍入行为
- SSL 协商修复（Nathan Mueller）

- 使 libpq 的 ~/.pgpass 特性在用 PQconnectDB 连接时工作
- 更新 my2pg、ora2pg
- 翻译更新
- 在 contrib/lo 中添加 lo 与 oid 类型之间的转换
- fastpath 代码现在检查调用函数的权限

E.223. 版本 7.3.1

发行日期
2002-12-18

此发行版包含 7.3 版本的多种修复。

E.223.1. 迁移到版本 7.3.1

对于运行 7.3 版本的用户，不需要进行转储/恢复。但应注意，主 PostgreSQL 接口库 libpq 在此版本有了新的主版本号，某些情况下可能需要重新编译客户端代码。

E.223.2. 变更

- 修复客户端/服务器编码不匹配时 COPY TO 的核心转储（Tom）
- 允许 pg_dump 与 7.2 之前的服务器一起工作（Philip）
- contrib/adddepend 修复（Tom）
- 修复删除逐用户/逐数据库配置设置的问题（Tom）
- contrib/vacuumlo 修复（Tom）
- 即使 pg_shadow 包含 MD5 口令也允许 'password' 加密（Bruce）
- contrib/dbmirror 修复（Steven Singer）
- 优化器修复（Tom）
- contrib/tsearch 修复（Teodor Sigaev、Magnus）
- 允许区域设置名为大小写混合（Nicolai Tufar）
- 递增 libpq 库的主版本号（Bruce）
- pg_hba.conf 错误报告修复（Bruce、Neil）
- 新增 SCO Openserver 5.0.4 为受支持平台（Bruce）
- 防止 EXPLAIN 使服务器崩溃（Tom）

- SSL 修复（Nathan Mueller）
- 防止通过 ALTER TABLE 创建组合列（Tom）

E.224. 版本 7.3

发行日期
2002-11-27

E.224.1. 概述

此发行版的主要变更：

Schemas

模式允许用户在独立的命名空间中创建对象，因此两个人或应用程序可以拥有同名的表。还有用于共享表的 public 模式。可以通过移除 public 模式上的权限来限制表/索引的创建。

Drop Column

PostgreSQL 现在支持 `ALTER TABLE ... DROP COLUMN` 功能。

Table Functions

返回多行和/或多列的函数现在比以前好用得多。可以在 `SELECT` 的 `FROM` 子句中调用这样的“表函数”，把其输出当作表。此外，PL/pgSQL 函数现在可以返回集合。

Prepared Queries

PostgreSQL 现在支持预备查询，以改进性能。

Dependency Tracking

PostgreSQL 现在记录对象依赖，这在许多方面带来了改进。`DROP` 语句现在接受 `CASCADE` 或 `RESTRICT` 来控制是否也删除依赖对象。

Privileges

函数和过程语言现在有权限，函数可以定义为以其创建者的权限运行。

Internationalization

多字节和区域设置支持现在总是启用。

Logging

多种日志选项得到了增强。

Interfaces

大量接口已被移到 http://gborg.postgresql.org，在那里可以独立开发和发布。

Functions/Identifiers

默认情况下，函数现在最多可接受 32 个参数，标识符最长可达 63 字节。此外，`OPAQUE` 现已废弃：有特定的“伪数据类型”来表示函数参数和结果类型中 `OPAQUE` 以前的每种含义。

E.224.2. 迁移到版本 7.3

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。如果你的应用程序检查系统目录，由于 7.3 中引入了模式，还需要额外的更改；更多信息参见http://developer.postgresql.org/~momjian/upgrade_tips_7.3。

请注意以下不兼容之处：

- 不再支持 6.3 之前的客户端。
- `pg_hba.conf` 现在有用户名列和额外特性。现有文件需要调整。
- 几个 `postgresql.conf` 日志参数已被重命名。
- `LIMIT #,#` 已被禁用；请使用 `LIMIT # OFFSET #`。
- 带列列表的 `INSERT` 语句必须为每个指定的列提供值。例如 `INSERT INTO tab (col1, col2) VALUES ('val1')` 现在无效。如果 `INSERT` 没有列列表，仍允许提供比预期少的列。
- `serial` 列不再自动 `UNIQUE`；因此不会自动创建索引。
- 已中止事务内的 `SET` 命令现在会被回滚。
- `COPY` 不再把缺失的尾随列视为空。需要指定所有列。（不过，可以通过在 `COPY` 命令中指定列列表达到类似效果。）
- 数据类型 `timestamp` 现在等价于 `timestamp without time zone`，而不是 `timestamp with time zone`。
- 装载到 7.3 的 7.3 之前数据库不会有 `serial` 列、唯一约束和外键的新对象依赖。参见 `contrib/adddepend/` 目录中的详细说明和添加此类依赖的脚本。
- 不再允许空字符串（`''`）作为整数字段的输入。以前它被静默解释为 0。

E.224.3. 变更

E.224.3.1. 服务器操作

- 新增显示锁的 pg_locks 视图（Neil）
- 口令协商内存分配的安全修复（Neil）
- 移除对版本 0 FE/BE 协议的支持（PostgreSQL 6.2 及更早）（Tom）
- 为超级用户保留最后几个后端槽，新增参数 superuser_reserved_connections 来控制（Nigel J. Andrews）

E.224.3.2. 性能

- 通过只调用一次 localtime() 改进启动（Tom）

- 把系统目录信息缓存在平面文件中以加快启动（Tom）
- 改进索引信息的缓存（Tom）
- 优化器改进（Tom、Fernando Nasser）
- 目录缓存现在存储失败的查找（Tom）
- 哈希函数改进（Neil）
- 改进查询词法化和网络处理的性能（Peter）
- 大对象恢复的速度改进（Mario Weilguni）
- 首次查找时标记过期索引条目，节省后续堆获取（Tom）
- 避免过多的 NULL 位图填充（Manfred Koizar）
- 为 Solaris 添加 BSD 许可的 qsort()，以提升性能（Bruce）
- 把每行开销减少四字节（Manfred Koizar）
- 修复 GEQO 优化器缺陷（Neil Conway）
- 使 WITHOUT OID 实际每行节省四字节（Manfred Koizar）
- 新增 default_statistics_target 变量，指定 ANALYZE 桶数（Neil）
- 临时表使用本地缓冲区缓存，因此没有 WAL 开销（Tom）
- 改进大表上空闲空间映射的性能（Stephen Marshall、Tom）
- 改进的 WAL 写并发（Tom）

E.224.3.3. 权限

- 新增函数和过程语言上的权限（Peter）
- 为 CREATE DATABASE 新增 OWNER，使超级用户能代表非特权用户创建数据库（Gavin Sherry、Tom）
- 新增对象权限位 EXECUTE 和 USAGE（Tom）
- 新增 SET SESSION AUTHORIZATION DEFAULT 和 RESET SESSION AUTHORIZATION（Tom）
- 允许函数以其所有者的权限执行（Peter）

E.224.3.4. 服务器配置

- 服务器日志消息现在标记为 LOG 而不是 DEBUG（Bruce）
- 为 pg_hba.conf 添加用户列（Bruce）
- 让 log_connections 在日志文件中输出两行（Tom）
- 从 postgresql.conf 移除 debug_level，现为 server_min_messages（Bruce）

- 新的 ALTER DATABASE/USER ... SET 命令，用于逐用户/数据库初始化（Peter）
- 新参数 server_min_messages 和 client_min_messages，控制哪些消息发送到服务器日志或客户端应用程序（Bruce）
- 允许 pg_hba.conf 指定以逗号分隔的用户/数据库列表、以 + 为前缀的组名和以 @ 为前缀的文件名（Bruce）
- 移除辅助口令文件能力和 pg_password 工具（Bruce）
- 新增变量 db_user_namespace，用于数据库本地用户名（Bruce）
- SSL 改进（Bear Giles）
- 使存储口令的加密成为默认（Bruce）
- 允许通过调用 pg_stat_reset() 重置 pg_statistics（Christopher）
- 新增 log_duration 参数（Bruce）
- 把 debug_print_query 改名为 log_statement（Bruce）
- 把 show_query_stats 改名为 show_statement_stats（Bruce）
- 新增参数 log_min_error_statement，在出错时把命令打印到日志（Gavin）

E.224.3.5. 查询

- 使游标不敏感，即其内容不变（Tom）
- 禁用 LIMIT #,# 语法；现在只支持 LIMIT # OFFSET #（Bruce）
- 把标识符长度增加到 63（Neil、Bruce）
- UNION 修复：合并不同长度的 3 列以上（Tom）
- 为 INSERT 新增 DEFAULT 关键字，例如 INSERT ... (..., DEFAULT, ...)（Rod）
- 允许视图用 ALTER COLUMN ... SET DEFAULT 拥有默认值（Neil）
- 对列列表未提供所有列值的 INSERT 失败，例如 INSERT INTO tab (col1, col2) VALUES ('val1')；（Rod）
- 连接别名的修复（Tom）
- FULL OUTER JOIN 的修复（Tom）
- 改进无效标识符和位置的报告（Tom、Gavin）
- 修复 OPEN cursor(args)（Tom）
- 允许在视图中使用 'ctid' 以及 currtid(viewname)（Hiroshi）
- 带 UNION 的 CREATE TABLE AS 的修复（Tom）
- SQL99 语法改进（Thomas）

- 新增 statement_timeout 变量来取消查询（Bruce）
- 允许用 PREPARE/EXECUTE 预备查询（Neil）
- 允许 FOR UPDATE 出现在 LIMIT/OFFSET 之后（Bruce）
- 新增变量 autocommit（Tom、David Van Wie）

E.224.3.6. 对象操作

- 使 CREATE DATABASE 中的等号可选（Gavin Sherry）
- 让 ALTER TABLE OWNER 也更改索引所有权（Neil）
- 新 ALTER TABLE tabname ALTER COLUMN colname SET STORAGE 控制 TOAST 存储和压缩（John Gray）
- 新增模式支持，CREATE/DROP SCHEMA（Tom）
- 为临时表创建模式（Tom）
- 新增用于模式搜索的变量 search_path（Tom）
- 新增 ALTER TABLE SET/DROP NOT NULL（Christopher）
- 新 CREATE FUNCTION 易变性级别（Tom）
- 使规则名只在每个表内唯一（Tom）
- 为 DROP RULE 和 COMMENT ON RULE 新增 'ON tablename' 子句（Tom）
- 新增 ALTER TRIGGER RENAME（Joe）
- 新 current_schema() 和 current_schemas() 查询函数（Tom）
- 允许函数返回多行（表函数）（Joe）
- 为一致性使 CREATE DATABASE 中的 WITH 可选（Bruce）
- 新增对象依赖跟踪（Rod、Tom）
- 为 DROP 命令新增 RESTRICT/CASCADE（Rod）
- 为非 CHECK 约束新增 ALTER TABLE DROP（Rod）
- 删除带 SERIAL 的表时自动销毁序列（Rod）
- 如果列被外键使用则阻止删除列（Rod）
- 对象被删除时自动删除约束/函数（Rod）
- 新增 CREATE/DROP OPERATOR CLASS（Bill Studenmund、Tom）
- 新增 ALTER TABLE DROP COLUMN（Christopher、Tom、Hiroshi）
- 防止继承列被删除或重命名（Alvaro Herrera）
- 修复外键约束在中间数据库状态不报错（Stephan）

- 把列或表重命名传播到外键约束
- 新增 CREATE OR REPLACE VIEW（Gavin、Neil、Tom）
- 新增 CREATE OR REPLACE RULE（Gavin、Neil、Tom）
- 让规则按字母顺序执行，返回更可预测的值（Tom）
- 触发器现在按字母顺序触发（Tom）
- 新增 /contrib/adddepend 来处理 7.3 之前的对象依赖（Rod）
- 插入/更新值时允许更好的转换（Tom）

E.224.3.7. 工具命令

- 让 COPY TO 把嵌入的回车和换行输出为 \r 和 \n（Tom）
- 允许 COPY FROM 中的 DELIMITER 为 8 位干净（Tatsuo）
- 让 pg_dump 使用 ALTER TABLE ADD PRIMARY KEY 以提升性能（Neil）
- 禁用多语句规则中的方括号（Bruce）
- 禁止在函数内调用 VACUUM（Bruce）
- 允许 dropdb 和其他脚本使用带空格的标识符（Bruce）
- 把数据库注释更改限制在当前数据库
- 允许对操作符添加注释，独立于底层函数（Rod）
- 回退已中止事务中的 SET 命令（Tom）
- EXPLAIN 现在作为查询输出（Tom）
- 在 EXPLAIN 中显示条件表达式和排序键（Tom）
- 新增 'SET LOCAL var = value'，为单个事务设置配置变量（Tom）
- 允许 ANALYZE 在事务中运行（Bruce）
- 用新 WITH 子句改进 COPY 语法，保持向后兼容（Bruce）
- 修复 pg_dump 在非 ASCII 转储中一致输出标签（Bruce）
- 使转储文件中的外键约束更清晰（Rod）
- 新增 COMMENT ON CONSTRAINT（Rod）
- 允许 COPY TO/FROM 指定列名（Brent Verner）
- 把 UNIQUE 和 PRIMARY KEY 约束转储为 ALTER TABLE（Rod）
- 让 SHOW 输出查询结果（Joe）
- 对短 COPY 行产生失败而不是填充 NULL（Neil）

- 修复 CLUSTER 以保留所有表属性（Alvaro Herrera）
- 新 pg_settings 表，用于查看/修改 GUC 设置（Joe）
- 为 pg_dump 输出添加智能引用和可移植性改进（Peter）
- 把 serial 列转储为 SERIAL（Tom）
- 为 pg_dump 启用大文件支持（>2G）（Peter、Philip Warner、Bruce）
- 不允许对涉及参照约束的表使用 TRUNCATE（Rod）
- 让 TRUNCATE 也自动截断关系的 toast 表（Tom）
- 新增 clusterdb 工具，基于先前的 CLUSTER 操作自动聚簇整个数据库（Alvaro Herrera）
- 彻底改造 pg_dumpall（Peter）
- 允许对 TOAST 表执行 REINDEX（Tom）
- 按 SQL99 实现 START TRANSACTION（Neil）
- 修复页分裂影响批量删除时的罕见索引损坏（Tom）
- 修复继承的 ALTER TABLE ... ADD COLUMN（Alvaro Herrera）

E.224.3.8. 数据类型和函数

- 修复 factorial(0) 使其返回 1（Bruce）
- 日期/时间/时区改进（Thomas）
- 数组切片提取的修复（Tom）
- 修复 extract/date_part 为 timestamp 报告正确的微秒（Tatsuo）
- 允许 text_substr() 和 bytea_substr() 更高效地读取 TOAST 值（John Gray）
- 新增域支持（Rod）
- 使 WITHOUT TIME ZONE 成为 TIMESTAMP 和 TIME 数据类型的默认（Thomas）
- 允许用 configure 的 --enable-integer-datetimes 为日期/时间类型使用 64 位整数的替代存储方案（Thomas）
- 让 timezone(timestamptz) 返回 timestamp 而不是字符串（Thomas）
- 允许 1BC 之前日期的日期/时间类型带小数秒（Thomas）
- 把 timestamp 数据类型的精度限制为 6 位小数（Thomas）
- 把时区转换函数从 timetz() 改为 timezone()（Thomas）
- 新增配置变量 datestyle 和 timezone（Tom）
- 新增 OVERLAY()，允许替换字符串中的子串（Thomas）
- 新增 SIMILAR TO（Thomas、Tom）

- 新增正则表达式 SUBSTRING(string FROM pat FOR escape)（Thomas）
- 新增 LOCALTIME 和 LOCALTIMESTAMP 函数（Thomas）
- 用 CREATE TYPE typename AS (column) 新增命名组合类型（Joe）
- 允许在表别名子句中定义组合类型（Joe）
- 新增 API 以简化 C 语言表函数的创建（Joe）
- 从与标准不匹配的 SQL99 函数调用中移除 ODBC 兼容的空括号（Thomas）
- 允许 macaddr 数据类型接受无分隔符的 12 个十六进制数字（Mike Wyer）
- 新增 CREATE/DROP CAST（Peter）
- 新增 IS DISTINCT FROM 操作符（Thomas）
- 新增 SQL99 TREAT() 函数，CAST() 的同义词（Thomas）
- 新增 pg_backend_pid()，输出后端 pid（Bruce）
- 新增 IS OF / IS NOT OF 类型谓词（Thomas）
- 允许不完全指定长度的位串常量（Thomas）
- 允许 8 字节整数与位串之间的转换（Thomas）
- 实现十六进制字面量到位串字面量的转换（Thomas）
- 允许表函数出现在 FROM 子句中（Joe）
- 把函数参数的最大数量增加到 32（Bruce）
- 不再自动为 SERIAL 列创建索引（Tom）
- 新增 current_database()（Rod）
- 修复 cash_words() 使其不溢出缓冲区（Tom）
- 新增函数 replace()、split_part()、to_hex()（Joe）
- 修复 bytea 作为右侧参数的 LIKE（Joe）
- 防止 SELECT cash_out(2) 导致的崩溃（Tom）
- 修复 to_char(1,'FM999.99') 使其返回句点（Karel）
- 修复返回 OPAQUE 的触发器/类型/语言函数，使其返回正确的类型（Tom）

E.224.3.9. 国际化

- 新增编码：韩语（JOHAB）、泰语（WIN874）、越南语（TCVN）、阿拉伯语（WIN1256）、简体中文（GBK）、韩语（UHC）（Eiji Tokuya）
- 默认启用区域设置支持（Peter）
- 新增区域设置变量（Peter）

- 在 PQescapeBytea/PQunescapeBytea 中为多字节转义 >= 0x7f 的字节（Tatsuo）
- 为正则表达式字符类添加区域感知
- 默认启用多字节支持（Tatsuo）
- 新增 GB18030 多字节支持（Bill Huang）
- 新增 CREATE/DROP CONVERSION，允许可装载的编码（Tatsuo、Kaori）
- 新增 pg_conversion 表（Tatsuo）
- 新增 SQL99 CONVERT() 函数（Tatsuo）
- pg_dumpall、pg_controldata 和 pg_resetxlog 现在支持国家语言（Peter）
- 新增和更新的翻译

E.224.3.10. 服务器端语言

- 允许递归 SQL 函数（Peter）
- 更改 PL/Tcl 构建以使用配置的编译器和 Makefile.shlib（Peter）
- 彻底改造 PL/pgSQL 的 FOUND 变量以更兼容 Oracle（Neil、Tom）
- 允许 PL/pgSQL 处理带引号的标识符（Tom）
- 允许返回集合的 PL/pgSQL 函数（Neil）
- 使 PL/pgSQL 感知模式（Joe）
- 移除一些内存泄漏（Nigel J. Andrews、Tom）

E.224.3.11. psql

- 为兼容 7.2.0，psql \connect 不再把数据库名转为小写（Tom）
- 新增 psql \timing 来为用户查询计时（Greg Sabino Mullane）
- 让 psql \d 显示索引信息（Greg Sabino Mullane）
- 新 psql \dD 显示域（Jonathan Eisler）
- 允许 psql 显示视图上的规则（Paul ?）
- psql 变量替换的修复（Tom）
- 允许 psql \d 显示临时表结构（Tom）
- 允许 psql \d 显示外键（Rod）
- 修复 \? 以遵循 \pset pager（Bruce）
- 让 psql 在启动时报告其版本号（Tom）
- 允许 \copy 指定列名（Tom）

E.224.3.12. libpq

- 新增 ~/.pgpass 来存储主机/用户口令组合（Alvaro Herrera）
- 为 libpq 新增 PQunescapeBytea() 函数（Patrick Welche）
- 修复通过非阻塞连接发送大查询的问题（Bernhard Herzog）
- 修复 libpq 在 Win9X 上使用定时器的问题（David Ford）
- 允许 libpq notify 处理标识符长度不同的服务器（Tom）
- 把 libpq 的 PQescapeString() 和 PQescapeBytea() 添加到 Windows（Bruce）
- 修复非阻塞连接的 SSL 问题（Jack Bates）
- 新增 libpq 连接超时参数（Denis A Ustimenko）

E.224.3.13. JDBC

- 允许 JDBC 用 JDK 1.4 编译（Dave）
- 新增 JDBC 3 支持（Barry）
- 允许 JDBC 通过在连接 URL 中添加 ?loglevel=X 来设置 loglevel（Barry）
- 新增打印版本号的 Driver.info() 消息（Barry）
- 新增可更新结果集（Raghu Nidagal、Dave）
- 新增对可调用语句的支持（Paul Bethe）
- 新增查询取消能力
- 新增 refresh row（Dave）
- 修复多字节服务器的 MD5 加密处理（Jun Kawai）
- 新增对预备语句的支持（Barry）

E.224.3.14. 其他接口

- 修复 ECPG 关于单引号中八进制数的缺陷（Michael）
- 把 src/interfaces/libpgeasy 移到 http://gborg.postgresql.org（Marc、Bruce）
- 改进 Python 接口（Elliot Lee、Andrew Johnson、Greg Copeland）
- 新增 libpgtcl 连接关闭事件（Gerhard Hintermayer）
- 把 src/interfaces/libpq++ 移到 http://gborg.postgresql.org（Marc、Bruce）
- 把 src/interfaces/odbc 移到 http://gborg.postgresql.org（Marc）
- 把 src/interfaces/libpgeasy 移到 http://gborg.postgresql.org（Marc、Bruce）
- 把 src/interfaces/perl5 移到 http://gborg.postgresql.org（Marc、Bruce）

- 从主树移除 src/bin/pgaccess，现位于 http://www.pgaccess.org（Bruce）
- 为 libpgtcl 新增 pg_on_connection_loss 命令（Gerhard Hintermayer、Tom）

E.224.3.15. 源代码

- 并行 make 的修复（Peter）
- 链接 Tcl 的 AIX 修复（Andreas Zeugswetter）
- 允许 PL/Perl 在 Cygwin 下构建（Jason Tishler）
- 改进 MIPS 编译（Peter、Oliver Elphick）
- 要求 Autoconf 2.53 版本（Peter）
- configure 默认要求 readline 和 zlib（Peter）
- 允许 Solaris 使用紧密共享内存（ISM）以提升性能（Scott Brunza、P.J. Josh Rovero）
- 编译时总是启用 syslog，移除 --enable-syslog 选项（Tatsuo）
- 编译时总是启用多字节，移除 --enable-multibyte 选项（Tatsuo）
- 编译时总是启用区域设置，移除 --enable-locale 选项（Peter）
- Win9x DLL 创建的修复（Magnus Naeslund）
- 修复 WAL 代码在 Windows、BeOS 上使用 link() 的问题（Jason Tishler）
- 把 sys/types.h 添加到 c.h，从主文件移除（Peter、Bruce）
- 修复 SMP 机器上的 AIX 挂起（Tomoyuki Niijima）
- AIX SMP 挂起修复（Tomoyuki Niijima）
- 修复较新 glibc 库上 1970 年之前的日期处理（Tom）
- 修复 PowerPC SMP 锁定（Tom）
- 防止使用 gcc -ffast-math（Peter、Tom）
- 开发者构建现在要求 Bison >= 1.50
- Kerberos 5 支持现在能用 Heimdal 构建（Peter）
- 在用户指南中新增列出 SQL 特性的附录（Thomas）
- 改进可装载模块链接以使用 RTLD_NOW（Tom）
- 新错误级别 WARNING、INFO、LOG、DEBUG[1-5]（Bruce）
- 新 src/port 目录存放替代的 libc 函数（Peter、Bruce）
- 用于模式的新 pg_namespace 系统目录（Tom）
- 为模式新增 pg_class.relnamespace（Tom）

- 为模式新增 pg_type.typnamespace（Tom）
- 为模式新增 pg_proc.pronamespace（Tom）
- 重构聚合使其有 pg_proc 条目（Tom）
- 系统关系现在有自己的命名空间，不需要 pg_* 测试（Fernando Nasser）
- 把 TOAST 索引名改为 *_index 而不是 *_idx（Neil）
- 为操作符、操作符类新增命名空间（Tom）
- 为服务器控制文件新增额外检查（Thomas）
- 新波兰语 FAQ（Marcin Mazurek）
- 新增 Posix 信号量支持（Tom）
- 记录重新索引的需要（Bruce）
- 重命名一些内部标识符以简化 Windows 编译（Jan、Katherine Ward）
- 新增关于计算磁盘空间的文档（Bruce）
- 从 GUC 移除 KSQO（Bruce）
- 修复 rtree 中的内存泄漏（Kenneth Been）
- 为一致性修改一些错误消息（Bruce）
- 移除未使用的系统表列（Peter）
- 在适当处使系统列为 NOT NULL（Tom）
- 清理 sprintf 的使用，改用 snprintf()（Neil、Jukka Holappa）
- 移除 OPAQUE 并创建特定子类型（Tom）
- 数组内部处理的清理（Joe、Tom）
- 不允许 pg_atoi('')（Bruce）
- 移除参数 wal_files，因为 WAL 文件现在被回收（Bruce）
- 为堆页新增版本号（Tom）

E.224.3.16. 附加模块

- 允许 /contrib/array 中的 inet 数组（Neil）
- GiST 修复（Teodor Sigaev、Neil）
- 升级 /contrib/mysql
- 新增 /contrib/dbsize，无需清理即可显示表大小（Peter）
- 新增 /contrib/intagg，整数聚合例程（mlw）

- 改进 /contrib/oid2name（Neil、Bruce）
- 改进 /contrib/tsearch（Oleg、Teodor Sigaev）
- /contrib/rserver 的清理（Alexey V. Borzov）
- 更新 /contrib/oracle 转换工具（Gilles Darold）
- 更新 /contrib/dblink（Joe）
- 改进 /contrib/vacuumlo 支持的选项（Mario Weilguni）
- /contrib/intarray 的改进（Oleg、Teodor Sigaev、Andrey Oktyabrski）
- 新增 /contrib/reindexdb 工具（Shaun Thomas）
- 为 /contrib/isbn_issn 新增索引（Dan Weston）
- 新增 /contrib/dbmirror（Steven Singer）
- 改进 /contrib/pgbench（Neil）
- 新增 /contrib/tablefunc 表函数示例（Joe）
- 新增用于树结构的 /contrib/ltree 数据类型（Teodor Sigaev、Oleg Bartunov）
- 把 /contrib/pg_controldata、pg_resetxlog 移入主树（Bruce）
- /contrib/cube 的修复（Bruno Wolff）
- 改进 /contrib/fulltextindex（Christopher）

E.225. 版本 7.2.8

发行日期
2005-05-09

此发行版包含 7.2.7 之后的多种修复，包括一个与安全相关的问题。

E.225.1. 迁移到版本 7.2.8

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.225.2. 变更

- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。

- 修复关系扩展与 VACUUM 之间的竞态条件

这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
没有已知案例造成超过断言失败的后果。

- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump 正确转储包含 `%` 的索引名和触发器名（Neil）
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）

E.226. 版本 7.2.7

发行日期
2005-01-31

此发行版包含 7.2.6 之后的多种修复，包括几个与安全相关的问题。

E.226.1. 迁移到版本 7.2.7

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.226.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，
  而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示

E.227. 版本 7.2.6

发行日期
2004-10-22

此发行版包含 7.2.5 之后的多种修复。

E.227.1. 迁移到版本 7.2.6

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.227.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，
  这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 make_oidjoins_check 在 /tmp 中使用临时文件

  这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

- 更新到更新版本的 Bison

E.228. 版本 7.2.5

发行日期
2004-08-16

此发行版包含 7.2.4 之后的多种修复。

E.228.1. 迁移到版本 7.2.5

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.228.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 修复与首个根页分裂并行的 btree 搜索的极端情形

- 修复 `to_ascii` 中的缓冲区溢出（Guido Notari）

- 修复 char 为无符号的机器上死锁检测的核心转储

- 修复 Async_NotifyHandler 运行后不响应 `pg_ctl stop -m fast` 的问题
- 修复 pg_dump 中的内存泄漏
- 避免与 `isblank()` 函数或宏的系统定义冲突

E.229. 版本 7.2.4

发行日期
2003-01-30

此发行版包含 7.2.3 版本的多种修复，包括防止可能数据丢失的修复。

E.229.1. 迁移到版本 7.2.4

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.229.2. 变更

- 修复 VACUUM "No one parent tuple was found" 错误的一些额外情形
- 禁止在函数内调用 VACUUM（Bruce）
- 确保在标记检查点完成之前 pg_clog 更新已同步到磁盘
- 避免大型哈希连接期间的整数溢出
- 使 GROUP 命令在 pg_group.grolist 大到需要 TOAST 时也能工作
- 修复日期时间表中的错误；一些时区名未被识别
- 修复 circle_poly()、path_encode()、path_add() 中的整数溢出（Neil）
- 修复 lseg_eq()、lseg_ne()、lseg_center() 中长期存在的逻辑错误

E.230. 版本 7.2.3

发行日期
2002-10-01

此发行版包含 7.2.2 版本的多种修复，包括防止可能数据丢失的修复。

E.230.1. 迁移到版本 7.2.3

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.230.2. 变更

- 防止压缩事务日志的可能丢失（Tom）
- 防止非超级用户推进最近的清理信息（Tom）
- 处理较新版本 glibc 中 1970 年之前的日期值（Tom）
- 修复服务器关闭期间可能的挂起
- 防止 SMP PPC 机器上的自旋锁挂起（Tomoyuki Niijima）
- 修复 pg_dump 正确转储 FULL JOIN USING（Tom）

E.231. 版本 7.2.2

发行日期
2002-08-23

此发行版包含 7.2.1 版本的多种修复。

E.231.1. 迁移到版本 7.2.2

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.231.2. 变更

- 允许在 PL/pgSQL 中 EXECUTE "CREATE TABLE AS ... SELECT"（Tom）
- 修复压缩事务日志 id 回卷的问题（Tom）
- 修复 PQescapeBytea/PQunescapeBytea 使其处理 > 0x7f 的字节（Tatsuo）
- 修复 psql 和 pg_dump 以不存在的长选项调用时的崩溃（Tatsuo）
- 修复调用几何操作符时的崩溃（Tom）
- 允许 OPEN cursor(args)（Tom）
- rtree_gist 索引构建的修复（Teodor）
- 转储用户定义聚合的修复（Tom）
- contrib/intarray 修复（Oleg）
- 修复使用括号的复杂 UNION/EXCEPT/INTERSECT 查询（Tom）
- pg_convert 的修复（Tatsuo）
- 长 DATA 字符串崩溃的修复（Thomas、Neil）
- repeat()、lpad()、rpad() 与长字符串的修复（Neil）

E.232. 版本 7.2.1

发行日期
2002-03-21

此发行版包含 7.2 版本的多种修复。

E.232.1. 迁移到版本 7.2.1

对于运行 7.2 版本的用户，不需要进行转储/恢复。

E.232.2. 变更

- 确保序列计数器在崩溃后不会倒退（Tom）
- 修复 pgaccess 汉字转换键绑定（Tatsuo）
- 优化器改进（Tom）
- Cash I/O 改进（Tom）
- 新俄语 FAQ
- 缺失 AuthBlockSig 的编译修复（Heiko）
- 附加时区和时区修复（Thomas）
- 允许 psql \connect 处理大小写混合的数据库和用户名（Tom）
- 即使有 ON INSERT 规则在命令完成时也返回正确的 OID（Tom）
- 允许 COPY FROM 使用 8 位 DELIMITERS（Tatsuo）
- 修复 extract/date_part 对毫秒/微秒的缺陷（Tatsuo）
- 改进多个不同长度 UNION 的处理（Tom）
- contrib/btree_gist 改进（Teodor Sigaev）
- contrib/tsearch 词典改进，附加安装步骤见 README.tsearch（Thomas T. Thai、Teodor Sigaev）
- 数组下标处理的修复（Tom）
- 允许在 PL/pgSQL 中 EXECUTE "CREATE TABLE AS ... SELECT"（Tom）

E.233. 版本 7.2

发行日期
2002-02-04

E.233.1. 概述

此发行版改进了 PostgreSQL 在高流量应用程序中的使用。

此发行版的主要变更：

VACUUM

清理不再锁定表，因此允许在清理期间正常的用户访问。新 `VACUUM FULL` 命令通过锁定表并收缩表的磁盘副本执行旧式清理。

Transactions

超过四十亿事务的安装不再有问题。

OIDs

OID 现在是可选的。用户现在可以在 OID 使用过多的情形下创建不带 OID 的表。

Optimizer

系统现在在 `ANALYZE` 期间计算直方图列统计，使优化器的选择好得多。

Security

新的 MD5 加密选项允许更安全地存储和传输口令。Linux 和 BSD 系统上有新的 Unix 域套接字认证选项。

Statistics

管理员可以使用新的表访问统计模块获取表和索引使用的细粒度信息。

Internationalization

程序和库消息现在可以用多种语言显示。

E.233.2. 迁移到版本 7.2

希望从任何以前的版本迁移数据的用户，需要使用 `pg_dump` 进行转储/恢复。

请注意以下不兼容之处：

- `VACUUM` 命令的语义在此版本中已更改。你可能需要相应地更新维护程序。
- 在此版本中，使用 `= NULL` 的比较总是返回假（更准确地说是 NULL）。以前的版本自动把此语法转换为 `IS NULL`。旧行为可以用 `postgresql.conf` 参数重新启用。
- `pg_hba.conf` 和 `pg_ident.conf` 配置现在只在收到 SIGHUP 信号后重新装载，而不是每次连接时。
- 函数 `octet_length()` 现在返回未压缩的数据长度。
- 日期/时间值 `'current'` 不再可用。你需要重写你的应用程序。
- `timestamp()`、`time()` 和 `interval()` 函数不再可用。不要用 `timestamp()`，请改用 `timestamp 'string'` 或 `CAST`。

`SELECT ... LIMIT #,#` 语法将在下一个版本中移除。你应把查询改为使用独立的 LIMIT 和 OFFSET 子句，例如 `LIMIT 10 OFFSET 20`。

E.233.3. 变更

E.233.3.1. 服务器操作

- 在单独的目录中创建临时文件（Bruce）
- postmaster 启动时删除孤立的临时文件（Bruce）
- 为一些系统表添加唯一索引（Tom）
- 系统表操作符重组（Oleg Bartunov、Teodor Sigaev、Tom）
- 把 pg_log 改名为 pg_clog（Tom）
- 启用 SIGTERM、SIGQUIT 来终止后端（Jan）
- 移除后端数量的编译期限制（Tom）
- 信号量资源失败时更好的清理（Tatsuo、Tom）
- 允许安全的事务 ID 回卷（Tom）
- 从一些系统表移除 OID（Tom）
- 移除"触发数据更改违反"错误检查（Tom）
- 预备/保存计划的 SPI 门户创建（Jan）
- 允许 SPI 列函数用于系统列（Tom）
- 长值压缩改进（Tom）
- 表、索引访问的统计收集器（Jan）
- 把超长序列名截断为合理的值（Tom）
- 以毫秒测量事务时间（Thomas）
- 修复 TID 顺序扫描（Hiroshi）
- 超级用户 ID 现在固定为 1（Peter E）
- pg_ctl 新 "reload" 选项（Tom）

E.233.3.2. 性能

- 优化器改进（Tom）
- 优化器的新直方图列统计（Tom）
- 重用预写日志文件而不是丢弃它们（Tom）
- 缓存改进（Tom）

- IS NULL、IS NOT NULL 优化器改进（Tom）
- 改进锁管理器以减少锁争用（Tom）
- 为索引访问支持函数保留 relcache 条目（Tom）
- 使 NUMERIC 中的 NaN 和无穷大有更好的选择率（Tom）
- R-tree 性能改进（Kenneth Been）
- B-tree 分裂更高效（Tom）

E.233.3.3. 权限

- 把 UPDATE、DELETE 权限改为彼此独立（Peter E）
- 新 REFERENCES、TRIGGER 权限（Peter E）
- 允许一次对多个用户 GRANT/REVOKE（Peter E）
- 新 has_table_privilege() 函数（Joe Conway）
- 允许非超级用户清理数据库（Tom）
- 新 SET SESSION AUTHORIZATION 命令（Peter E）
- 修复新建表上权限修改的缺陷（Tom）
- 禁止非超级用户访问 pg_statistic，新增用户可访问的视图（Tom）

E.233.3.4. 客户端认证

- 在认证前 fork postmaster 以防止挂起（Peter E）
- 在 Linux、*BSD 上通过 Unix 域套接字新增 ident 认证（Helge Bahmann、Oliver Elphick、Teodor Sigaev、Bruce）
- 新增使用 MD5 加密的口令认证方法（Bruce）
- 允许用 MD5 加密存储的口令（Bruce）
- PAM 认证（Dominic J. Eidson）
- 只在启动和 SIGHUP 时装载 pg_hba.conf 和 pg_ident.conf（Bruce）

E.233.3.5. 服务器配置

- 一些时区缩写解释为澳大利亚而不是北美，现在可在运行时设置（Bruce）
- 设置默认事务隔离级别的新参数（Peter E）
- 启用把 "expr = NULL" 转换为 "expr IS NULL" 的新参数，默认关闭（Peter E）
- 控制 VACUUM 内存使用的新参数（Tom）
- 设置客户端认证超时的新参数（Tom）
- 设置最大打开文件数的新参数（Tom）

E.233.3.6. 查询

- INSERT 规则添加的语句现在在 INSERT 之后执行（Jan）
- 阻止目标列表中不加修饰的关系名（Bruce）
- ORDER BY 中 NULL 现在排在所有正常值之后（Tom）
- 新 IS UNKNOWN、IS NOT UNKNOWN 布尔测试（Tom）
- 新 SHARE UPDATE EXCLUSIVE 锁模式（Tom）
- 新 EXPLAIN ANALYZE 命令，显示运行时间和行数（Martijn van Oosterhout）
- 修复 LIMIT 与子查询的问题（Tom）
- 修复推入子查询的 LIMIT、DISTINCT ON（Tom）
- 修复嵌套的 EXCEPT/INTERSECT（Tom）

E.233.3.7. 模式操作

- 修复临时表中的 SERIAL（Bruce）
- 允许临时序列（Bruce）
- 序列现在内部使用 int8（Tom）
- 新 SERIAL8 创建带序列的 int8 列，默认仍为 SERIAL4（Tom）
- 用 WITHOUT OIDS 使 OID 可选（Tom）
- 为 CREATE TYPE 新增 %TYPE 语法（Ian Lance Taylor）
- 为 CHECK 约束新增 ALTER TABLE / DROP CONSTRAINT（Christopher Kings-Lynne）
- 新 CREATE OR REPLACE FUNCTION 更改现有函数（保留函数 OID）（Gavin Sherry）
- 新增 ALTER TABLE / ADD [UNIQUE | PRIMARY]（Christopher Kings-Lynne）
- 允许视图中重命名列
- 让 ALTER TABLE / RENAME COLUMN 更新索引的列名（Brent Verner）
- 修复继承表的 ALTER TABLE / ADD CONSTRAINT ... CHECK（Stephan Szabo）
- ALTER TABLE RENAME 正确更新外键触发器参数（Brent Verner）
- DROP AGGREGATE 和 COMMENT ON AGGREGATE 现在接受 aggtype（Tom）
- 为 SQL 函数新增自动返回类型数据转换（Tom）
- 允许 GiST 索引处理 NULL 和多键索引（Oleg Bartunov、Teodor Sigaev、Tom）
- 启用部分索引（Martijn van Oosterhout）

E.233.3.8. 工具命令

- 新增 RESET ALL、SHOW ALL（Marko Kreen）

- CREATE/ALTER USER/GROUP 现在允许选项任意顺序（Vince）
- 新增 LOCK A, B, C 功能（Neil Padgett）
- CREATE/ALTER USER 的新 ENCRYPTED/UNENCRYPTED 选项（Bruce）
- 新的轻量 VACUUM 不锁定表；旧语义以 VACUUM FULL 提供（Tom）
- 禁用视图上的 COPY TO/FROM（Bruce）
- COPY DELIMITERS 字符串必须恰好一个字符（Tom）
- 关于索引元组少于堆的 VACUUM 警告现在只在适当时出现（Martijn van Oosterhout）
- 修复 CREATE INDEX 的权限检查（Tom）
- 禁止不当使用 CREATE/DROP INDEX/TRIGGER/VIEW（Tom）

E.233.3.9. 数据类型和函数

- SUM()、AVG()、COUNT() 现在内部使用 int8 以提升速度（Tom）
- 新增 convert()、convert2()（Tatsuo）
- 新函数 bit_length()（Peter E）
- 使 CHAR(n)/VARCHAR(n) 中的 "n" 表示字符数而不是字节数（Tatsuo）
- CHAR()、VARCHAR() 现在拒绝过长的字符串（Peter E）
- BIT VARYING 现在拒绝过长的位串（Peter E）
- BIT 现在拒绝与声明大小不匹配的位串（Peter E）
- INET、CIDR 文本转换函数（Alex Pilosov）
- INET、CIDR 操作符 << 和 <<= 可索引（Alex Pilosov）
- Bytea \### 现在要求有效的三位八进制数
- Bytea 比较改进，现在支持 =、<>、>、>=、< 和 <=
- Bytea 现在支持 B-tree 索引
- Bytea 现在支持 LIKE、LIKE...ESCAPE、NOT LIKE、NOT LIKE...ESCAPE
- Bytea 现在支持连接
- 新 bytea 函数：position、substring、trim、btrim 和 length
- 新 encode() 函数模式 "escaped"，在最小转义的 bytea 与 text 之间转换
- 新增 pg_database_encoding_max_length()（Tatsuo）
- 新增 pg_client_encoding() 函数（Tatsuo）
- now() 返回毫秒精度的时间（Thomas）
- 新 TIMESTAMP WITHOUT TIMEZONE 数据类型（Thomas）

- 新增带 "T" 的 ISO 日期/时间规格，yyyy-mm-ddThh:mm:ss（Thomas）
- 新 xid/int 比较函数（Hiroshi）
- 为 TIME、TIMESTAMP 和 INTERVAL 数据类型新增精度（Thomas）
- 修改类型强制逻辑，先尝试二进制兼容函数（Tom）
- 新 encode() 函数默认安装（Marko Kreen）
- 改进的 to_*() 转换函数（Karel Zak）
- 优化单字节编码下的 LIKE/ILIKE（Tatsuo）
- contrib/pgcrypto 中的新函数：crypt()、hmac()、encrypt()、gen_salt()（Marko Kreen）
- 修正 translate() 函数的描述（Bruce）
- 为 SET TIME ZONE 新增 INTERVAL 参数（Thomas）
- 新增 INTERVAL YEAR TO MONTH（等）语法（Thomas）
- 优化单字节编码下的长度函数（Tatsuo）
- 修复 path_inter、path_distance、path_length、dist_ppath 以处理闭合路径（Curtis Barrett、Tom）
- octet_length(text) 现在返回未压缩的长度（Tatsuo、Bruce）
- 处理日期/时间字面量中的 "July" 全名（Greg Sabino Mullane）
- 一些 datatype() 函数调用现在的求值方式不同
- 新增对儒略和 ISO 时间规格的支持（Thomas）

E.233.3.10. 国际化

- psql、pg_dump、libpq 和服务器的国家语言支持（Peter E）
- 中文（简体、繁体）、捷克语、法语、德语、匈牙利语、俄语、瑞典语的消息翻译（Peter E、Serguei A. Mokhov、Karel Zak、Weiping He、Zhenbang Wei、Kovacs Zoltan）
- 使 trim、ltrim、rtrim、btrim、lpad、rpad、translate 感知多字节（Tatsuo）
- 新增 LATIN5,6,7,8,9,10 支持（Tatsuo）
- 新增 ISO 8859-5,6,7,8 支持（Tatsuo）
- 更正 LATIN5 为 ISO-8859-9 而不是 ISO-8859-5（Tatsuo）
- 使 mic2ascii() 不感知非 ASCII（Tatsuo）
- 拒绝无效的多字节字符序列（Tatsuo）

E.233.3.11. PL/pgSQL

- 现在 SELECT 循环使用门户，允许超大结果集（Jan）

- CURSOR 和 REFCURSOR 支持（Jan）
- 现在可以返回打开的游标（Jan）
- 新增 ELSEIF（Klaus Reger）
- 改进 PL/pgSQL 错误报告，包括错误位置（Tom）
- 为兼容允许游标声明中的 IS 或 FOR 关键字（Bruce）
- SELECT ... FOR UPDATE 的修复（Tom）
- 返回多行的 PERFORM 的修复（Tom）
- 让 PL/pgSQL 使用服务器的类型强制代码（Tom）
- 内存泄漏修复（Jan、Tom）
- 使尾部分号可选（Tom）

E.233.3.12. PL/Perl

- 新不可信 PL/Perl（Alex Pilosov）
- 即使 libperl 不是共享的，PL/Perl 现在也能在一些平台上构建（Peter E）

E.233.3.13. PL/Tcl

- 现在报告 errorInfo（Vsevolod Lobko）
- 新增 spi_lastoid 函数（bob@redivi.com）

E.233.3.14. PL/Python

- ……是新的（Andrew Bosma）

E.233.3.15. psql

- \d 按唯一、主键分组显示索引（Christopher Kings-Lynne）
- 允许反斜杠命令中的尾部分号（Greg Sabino Mullane）
- 如果可能从 /dev/tty 读取口令
- 更改用户和数据库时强制新的口令提示（Tatsuo、Tom）
- 为 Unicode 格式化正确的列数（Patrice）

E.233.3.16. libpq

- 新函数 PQescapeString()，转义命令字符串中的引号（Florian Weimer）
- 新函数 PQescapeBytea()，转义二进制字符串以用作 SQL 字符串字面量

E.233.3.17. JDBC

- 返回 INSERT 的 OID（Ken K）

- 处理更多数据类型（Ken K）
- 处理字符串中的单引号和换行（Ken K）
- 处理 NULL 变量（Ken K）
- 时区处理的修复（Barry Lind）
- 改进的 Druid 支持
- 允许非多字节服务器使用 8 位字符（Barry Lind）
- 支持 BIT、BINARY 类型（Ned Wolpert）
- 降低内存使用（Michael Stephens、Dave Cramer）
- 更新 DatabaseMetaData（Peter E）
- 新增 DatabaseMetaData.getCatalogs()（Peter E）
- 编码修复（Anders Bengtsson）
- Get/setCatalog 方法（Jason Davies）
- DatabaseMetaData.getColumns() 现在返回列默认值（Jason Davies）
- DatabaseMetaData.getColumns() 性能改进（Jeroen van Vianen）
- 一些 JDBC1 和 JDBC2 的合并（Anders Bengtsson）
- 事务性能改进（Barry Lind）
- 数组修复（Greg Zoller）
- 序列化添加
- 批处理修复（Rene Pijlman）
- ExecSQL 方法重组（Anders Bengtsson）
- GetColumn() 修复（Jeroen van Vianen）
- 修复 isWriteable() 函数（Rene Pijlman）
- 更好地通过 JDBC2 一致性测试（Rene Pijlman）
- 新增 bytea 类型能力（Barry Lind）
- 新增 isNullable()（Rene Pijlman）
- JDBC 日期/时间测试套件修复（Liam Stewart）
- SELECT 'id' AS xxx FROM table 的修复（Dave Cramer）
- 修复 DatabaseMetaData 以正确显示精度（Mark Lillywhite）
- 新 getImported/getExported 键（Jason Davies）
- MD5 口令加密支持（Jeremy Wohl）

- 修复为真正使用类型缓存（Ned Wolpert）

E.233.3.18. ODBC

- 移除查询大小限制（Hiroshi）
- 移除文本字段大小限制（Hiroshi）
- 多字节模式下 SQLPrimaryKeys 的修复（Hiroshi）
- 允许 ODBC 过程调用（Hiroshi）
- 改进布尔处理（Aidan Mountford）
- 大多数配置选项现在可通过 DSN 设置（Hiroshi）
- 多字节、性能修复（Hiroshi）
- 允许驱动与 iODBC 或 unixODBC 一起使用（Peter E）
- MD5 口令加密支持（Bruce）
- 为 odbc.sql 添加更多兼容函数（Peter E）

E.233.3.19. ECPG

- 实现 EXECUTE ... INTO（Christof Petig）
- 多行描述符支持（如 CARDINALITY）（Christof Petig）
- GRANT 参数的修复（Lee Kindness）
- 修复 INITIALLY DEFERRED 缺陷
- 各种缺陷修复（Michael、Christof Petig）
- 指示符变量数组的自动分配（int *ind_p=NULL）
- 字符串数组的自动分配（char **foo_pp=NULL）
- 修复 ECPGfree_auto_mem
- 所有有外部链接的函数名现在都以 ECPG 为前缀
- 结构数组的修复（Michael）

E.233.3.20. 其他接口

- Python 修复 fetchone()（Gerhard Haring）
- 在适当处在 Tcl 中使用 UTF、Unicode（Vsevolod Lobko、Reinhard Max）
- 新增 Tcl COPY TO/FROM（ljb）
- 防止 pg_dump 输出默认索引操作符类（Tom）
- 修复 libpgeasy 内存泄漏（Bruce）

E.233.3.21. 构建和安装

- Configure、动态装载器和共享库修复（Peter E）
- QNX 4 移植的修复（Bernd Tegge）
- Cygwin 和 Windows 移植的修复（Jason Tishler、Gerhard Haring、Dmitry Yurtaev、Darko Prenosil、Mikhail Terekhov）
- Windows 套接字通信失败的修复（Magnus、Mikhail Terekhov）
- Hurd 编译修复（Oliver Elphick）
- BeOS 修复（Cyril Velter）
- 移除 configure --enable-unicode-conversion，现由多字节启用（Tatsuo）
- AIX 修复（Tatsuo、Andreas）
- 修复并行 make（Peter E）
- 把 SQL 语言手册页安装到操作系统特定的目录（Peter E）
- 把 config.h 改名为 pg_config.h（Peter E）
- 重新组织头文件的安装布局（Peter E）

E.233.3.22. 源代码

- 移除 SEP_CHAR（Bruce）
- 新 GUC 钩子（Tom）
- 合并 GUC 和命令行处理（Marko Kreen）
- 移除 EXTEND INDEX（Martijn van Oosterhout、Tom）
- 新的缩进 java 代码的 pgjindent 工具（Bruce）
- 在 C++ 下编译时移除 true/false 的定义（Leandro Fanzone、Tom）
- pgindent 修复（Bruce、Tom）
- 在适当处用 strcmp() 取代 strcasecmp()（Peter E）
- Dynahash 可移植性改进（Tom）
- 在自旋锁结构中添加 'volatile' 用法
- 改进信号处理逻辑（Tom）

E.233.3.23. 附加模块

- 新 contrib/rtree_gist（Oleg Bartunov、Teodor Sigaev）
- 新 contrib/tsearch 全文索引（Oleg、Teodor Sigaev）
- 新增用于远程数据库访问的 contrib/dblink（Joe Conway）

- contrib/ora2pg Oracle 转换工具（Gilles Darold）
- contrib/xml XML 转换工具（John Gray）
- contrib/fulltextindex 修复（Christopher Kings-Lynne）
- 新 contrib/fuzzystrmatch，含 levenshtein 和 metaphone，合并了 soundex（Joe Conway）
- 为 contrib/intarray 新增布尔查询、二分搜索和修复（Oleg Bartunov）
- 新 pg_upgrade 工具（Bruce）
- 新增 pg_resetxlog 选项（Bruce、Tom）

E.234. 版本 7.1.3

发行日期
2001-08-15

E.234.1. 迁移到版本 7.1.3

对于运行 7.1.X 的用户，不需要进行转储/恢复。

E.234.2. 变更

移除大事务未使用的 WAL 段（Tom）
多动作规则修复（Tom）
PL/pgSQL 内存分配修复（Jan）
VACUUM 缓冲区修复（Tom）
回归测试修复（Tom）
pg_dump 对视图上 GRANT/REVOKE/注释及用户定义类型的修复（Tom）
修复带 DISTINCT ON 或 LIMIT 的子查询（Tom）
BeOS 修复
禁用视图上的 COPY TO/FROM（Tom）
Cygwin 构建（Jason Tishler）

E.235. 版本 7.1.2

发行日期
2001-05-11

这包含 7.1.1 之后的一个修复。

E.235.1. 迁移到版本 7.1.2

对于运行 7.1.X 的用户，不需要进行转储/恢复。

E.235.2. 变更

修复返回零行时的 PL/pgSQL SELECT
修复 psql 反斜杠核心转储
参照完整性权限修复
优化器修复
pg_dump 清理

E.236. 版本 7.1.1

发行日期
2001-05-05

这包含 7.1 之后的多种修复。

E.236.1. 迁移到版本 7.1.1

对于运行 7.1 的用户，不需要进行转储/恢复。

E.236.2. 变更

修复 numeric MODULO 操作符 (Tom)
pg_dump 修复 (Philip)
pg_dump 可转储 7.0 数据库 (Philip)
readline 4.2 修复 (Peter E)
JOIN 修复 (Tom)
AIX、MSWIN、VAX、N32K 修复 (Tom)
多字节修复 (Tom)
Unicode 修复 (Tatsuo)
优化器改进 (Tom)
修复函数中的整行 (Tom)
修复 pg_ctl 及带空格的选项字符串 (Peter E)
ODBC 修复 (Hiroshi)
EXPLAIN 现在可接受字符串参数 (Thomas)
Python 修复 (Darcy)

E.237. 版本 7.1

发行日期
2001-04-13

此发行版专注于移除 PostgreSQL 代码中存在多年的限制。

此发行版的主要变更：

Write-ahead Log (WAL)

为了在操作系统崩溃时保持数据库一致性，以前的 PostgreSQL 版本强制在每次事务提交前把所有数据修改写到磁盘。有了 WAL，只需把一个日志文件刷到磁盘，大幅改进了性能。如果你在以前的版本中一直用 -F 禁用磁盘刷写，可能要考虑停止使用它。

TOAST

TOAST——以前的版本有编译期行长度限制，通常为 8k - 32k。这一限制使长文本字段的存储很困难。有了 TOAST，任意长度的长行都能以良好的性能存储。

Outer Joins

我们现在支持外连接。不再需要用 UNION/NOT IN 变通实现外连接。我们使用 SQL92 外连接语法。

Function Manager

以前的 C 函数管理器不能正确处理空值，也不支持 64 位 CPU（Alpha）。新的函数管理器可以。你可以继续使用旧的自定义函数，
但将来可能想重写它们以使用新的函数管理器调用接口。

Complex Queries

以前版本不支持的大量复杂查询现在可以工作。视图、聚合、UNION、LIMIT、游标、
子查询和继承表的许多组合现在都能正常工作。继承表现在默认被访问。FROM 中的子查询现在受支持。

E.237.1. 迁移到版本 7.1

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

E.237.2. 变更

```
缺陷修复
---------
大量多字节/Unicode/区域设置修复 (Tatsuo 等)
更可靠的 ALTER TABLE RENAME (Tom)
Kerberos V 修复 (David Wragg)
修复目标列表带子查询的 INSERT INTO...SELECT (Tom)
在标准错误上提示用户名/口令 (Bruce)
大对象 inv_read/inv_write 修复 (Tom)
to_char()、to_date()、to_ascii() 和 to_timestamp() 的修复 (Karel、
   Daniel Baldoni)
防止查询表达式泄漏内存 (Tom)
允许更新数组元素 (Tom)
取消时唤醒锁等待者 (Hiroshi)
修复使用哈希连接时游标的罕见崩溃 (Tom)
修复回滚事务中的 DROP TABLE/INDEX (Hiroshi)
修复启用 MULTIBYTE 时 \l+ 导致 psql 崩溃的问题 (Peter E)
修复 CREATE VIEW 期间规则名截断的问题 (Ross Reedstrom)
修复 PL/perl (Alex Kapranoff)
禁止视图上的 LOCK (Mark Hollomon)
禁止视图上的 INSERT/UPDATE/DELETE (Mark Hollomon)
禁止视图上的 DROP RULE、CREATE INDEX、TRUNCATE (Mark Hollomon)
允许 PL/pgSQL 接受非 ASCII 标识符 (Tatsuo)
```

允许视图正确处理 GROUP BY、聚合、DISTINCT (Tom)
修复 TRUNCATE 命令的罕见失败 (Tom)
允许 UNION/INTERSECT/EXCEPT 与 ALL、子查询、视图、
DISTINCT、ORDER BY、SELECT...INTO 一起使用 (Tom)
修复中止事务期间的解析器失败 (Tom)
允许临时关系正确清理索引 (Bruce)
修复 VACUUM 在同页移动行的问题 (Tom)
修改 pg_dump 以更好地处理 template1 中的用户定义项 (Philip)
允许 VIEW 中使用 LIMIT (Tom)
要求游标 FETCH 遵循 LIMIT (Tom)
允许继承列上的 PRIMARY/FOREIGN KEY 定义 (Stephan)
允许子查询中的 ORDER BY、LIMIT (Tom)
允许 CREATE RULE 中的 UNION (Tom)
使 ALTER/DROP TABLE 可回滚 (Vadim、Tom)
把 initdb 的排序规则存储在 pg_control 中使排序规则不可更改 (Tom)
修复带规则的 INSERT...SELECT (Tom)
修复视图和子查询内的 FOR UPDATE (Tom)
修复 OVERLAPS 操作符在 NULL 方面符合 SQL92 规范 (Tom)
修复 lpad() 和 rpad() 以处理长度小于输入字符串的情形 (Tom)
修复某些规则中 NOTIFY 的使用 (Tom)
彻底改造 btree 代码 (Tom)
修复 PL/pgSQL 变量中 NOT NULL 的使用 (Tom)
彻底改造 GIST 代码 (Oleg)
修复 CLUSTER 以保留约束和列默认值 (Tom)
改进的死锁检测处理 (Tom)
允许表中有多个 SERIAL 列 (Tom)
防止偶发索引损坏 (Vadim)

增强

添加 OUTER JOIN (Tom)
函数管理器彻底改造 (Tom)
允许对索引 ALTER TABLE RENAME (Tom)
改进 CLUSTER (Tom)
改进更多平台的 ps 状态显示 (Peter E、Marc)
改进 CREATE FUNCTION 失败消息 (Ross)
JDBC 改进 (Peter、Travis Bauer、Christopher Cain、William Webber、
Gunnar)
大一统配置方案/GUC。现在许多选项可以在
data/postgresql.conf、postmaster/postgres 标志或 SET 命令中设置 (Peter
E)
改进文件描述符缓存的处理 (Tom)
关于自动创建表别名条目的新警告代码 (Bruce)
彻底改造 initdb 流程 (Tom、Peter E)
继承表的彻底改造；继承表现在默认被访问；
新 ONLY 关键字可阻止 (Chris Bitmead、Tom)
ODBC 清理/改进 (Nick Gorham、Stephan Szabo、Zoltan Kovacs、
Michael Fork)
允许重命名临时表 (Tom)
彻底改造内存管理器上下文 (Tom)
pg_dumpall 使用 CREATE USER 或 CREATE GROUP 而不是 COPY (Peter E)
彻底改造 pg_dump (Philip Warner)
允许 pg_hba.conf 辅助口令文件只指定用户名 (Peter E)
创建临时表时允许 TEMPORARY 或 TEMP 关键字 (Bruce)

新内存泄漏检查器 (Karel)
新 SET SESSION CHARACTERISTICS (Thomas)
允许嵌套块注释 (Thomas)
添加 WITHOUT TIME ZONE 类型限定符 (Thomas)
新 ALTER TABLE ADD CONSTRAINT (Stephan)
INTEGER 聚合使用 NUMERIC 累加器 (Tom)
彻底改造聚合代码 (Tom)
新 VARIANCE 和 STDDEV() 聚合
改进 pg_dump 的依赖排序 (Philip)
新 pg_restore 命令 (Philip)
新 pg_dump tar 输出选项 (Philip)
新 pg_dump 大对象支持 (Philip)
LIKE 的新 ESCAPE 选项 (Thomas)
新不区分大小写的 LIKE—ILIKE (Thomas)
允许函数索引使用二进制兼容类型 (Tom)
允许 SQL 函数用于更多上下文 (Tom)
新 pg_config 工具 (Peter E)
新 PL/pgSQL EXECUTE 命令，允许动态 SQL 和工具语句
(Jan)
新 PL/pgSQL GET DIAGNOSTICS 语句用于 SPI 值访问 (Jan)
新 quote_identifiers() 和 quote_literal() 函数 (Jan)
新 ALTER TABLE table OWNER TO user 命令 (Mark Hollomon)
允许 FROM 中的子查询，即 FROM (SELECT ...) [AS] alias (Tom)
把 PyGreSQL 更新到 3.1 版 (D'Arcy)
以 OID 命名的文件存储表 (Vadim)
新 SQL 函数 setval(seq,val,bool) 供 pg_dump 使用 (Philip)
要求用 DROP VIEW 删除视图，不再用 DROP TABLE (Mark)
允许 DROP VIEW view1, view2 (Mark)
允许 DROP INDEX、DROP RULE 和 DROP TYPE 中的多个对象 (Tom)
允许与 Unicode 的自动转换 (Tatsuo、Eiji)
新 /contrib/pgcrypto 哈希函数 (Marko Kreen)
新 pg_dumpall --globals-only 选项 (Peter E)
WAL 的新 CHECKPOINT 命令，创建新 WAL 日志文件 (Vadim)
新 AT TIME ZONE 语法 (Thomas)
允许配置 Unix 域套接字的位置 (David J. MacKenzie)
允许 postmaster 监听特定 IP 地址 (David J. MacKenzie)
允许通过前导斜杠在主机名中指定套接字路径名
(David J. MacKenzie)
允许 CREATE DATABASE 指定模板数据库 (Tom)
新工具把 MySQL 模式转储转换为 SQL92 和 PostgreSQL (Thomas)
新 /contrib/rserv 复制工具包 (Vadim)
COPY BINARY 的新文件格式 (Tom)
新 /contrib/oid2name 把数字文件映射到表名 (B Palmer)
新 "idle in transaction" ps 状态消息 (Marc)
更新到 pgaccess 0.98.7 (Constantin Teodorescu)
pg_ctl 关闭时现在默认 -w (等待)，新 -l (日志) 选项
向 pg_dump 添加初步的依赖检查 (Philip)

类型

修复 INET/CIDR 类型排序并添加新函数 (Tom)
使 OID 表现为无符号类型 (Tom)
允许 BIGINT 作为 INT8 的同义词 (Peter E)
新 int2 和 int8 比较操作符 (Tom)

新 BIT 和 BIT VARYING 类型 (Adriaan Joubert、Tom、Peter E)
因 TOAST, CHAR() 不再比 VARCHAR() 快 (Tom)
新 GIST seg/cube 示例 (Gene Selkov)
改进的 round(numeric) 处理 (Tom)
修复 CIDR 输出格式化 (Tom)
新 CIDR abbrev() 函数 (Tom)

性能

预写日志 (WAL) 以更少性能开销提供崩溃恢复
    (Vadim)
VACUUM 的 ANALYZE 阶段不再独占锁定表 (Bruce)
减少文件寻道 (Denis Perchine)
改进重复键的 BTREE 代码 (Tom)
把所有大对象存储在单个表中 (Denis Perchine、Tom)
改进内存分配性能 (Karel、Tom)

源代码

新函数管理器调用约定 (Tom)
SGI 可移植性修复 (David Kaelbling)
新 configure --enable-syslog 选项 (Peter E)
新 BSDI README (Bruce)
configure 脚本移到顶级而不是 /src (Peter E)
Makefile/配置/编译彻底改造 (Peter E)
新 configure --with-python 选项 (Peter E)
Solaris 清理 (Peter E)
彻底改造 /contrib Makefile (Karel)
新 OpenSSL 配置选项 (Magnus、Peter E)
AIX 修复 (Andreas)
QNX 修复 (Maurizio)
新 heap_open()、heap_openr() API (Tom)
移除冒号和分号操作符 (Thomas)
视图的新 pg_class.relkind 值 (Mark Hollomon)
把 ichar() 改名为 chr() (Karel)
btrim()、ascii()、chr()、repeat() 的新文档 (Karel)
NT/Cygwin 修复 (Pete Forman)
AIX 移植修复 (Andreas)
新 BeOS 移植 (David Reid、Cyril Velter)
把校对者的更改添加到文档 (Addison-Wesley、Bruce)
新 Alpha 自旋锁代码 (Adriaan Joubert、Compaq)
UnixWare 移植彻底改造 (Peter E)
新 Darwin/Mac OS X 移植 (Peter Bierman、Bruce Hartzler)
新 FreeBSD Alpha 移植 (Alfred)
彻底改造共享内存段 (Tom)
添加 IBM S/390 支持 (Neale Ferguson)
把 macmanuf 移到 /contrib (Larry Rosenman)
Syslog 改进 (Larry Rosenman)
不含用户添加内容的新 template0 数据库 (Tom)
新 /contrib/cube 和 /contrib/seg GIST 示例代码 (Gene Selkov)
允许用 NetBSD 的 libedit 代替 readline (Peter)
改进的汇编语言源代码格式 (Bruce)
新 contrib/pg_logger
createdb 的新 --template 选项

新 contrib/pg_control 工具 (Oliver)
新 FreeBSD 工具 ipc_check、start-scripts/freebsd

E.238. 版本 7.0.3

发行日期
2000-11-11

这包含 7.0.2 之后的多种修复。

E.238.1. 迁移到版本 7.0.3

对于运行 7.0.* 的用户，不需要进行转储/恢复。

E.238.2. 变更

Jdbc 修复 (Peter)
大对象修复 (Tom)
修复 COPY WITH OIDS 泄漏的倾向 (Tom)
修复反向索引扫描 (Tom)
修复 SELECT ... FOR UPDATE 使其检查重复键 (Hiroshi)
为 configure 添加 --enable-syslog (Marc)
修复罕见情形下后端退出时中止事务的问题 (Tom)
启用多字节时 psql \l+ 的修复 (Tatsuo)
允许 PL/pgSQL 接受非 ASCII 标识符 (Tatsuo)
使 vacuum 总是刷新缓冲区 (Tom)
修复以允许在等待锁时取消 (Hiroshi)
用户认证代码中内存分配问题的修复 (Tom)
移除对 int4out() 的虚假使用 (Tom)
COALESCE 或 BETWEEN 中多个子查询的修复 (Tom)
某些情形下堆打开时触发器失败的修复 (Jeroen van Vianen)
修复不相等的选择率错误 (Tom)
修复 strcmp() 的错误使用 (Tom)
修复存储管理器访问文件末尾之外条目的缺陷 (Tom)
修复以在所有 smgr elog 消息中包含内核 errno 消息 (Tom)
构建时 '.' 不在 PATH 中的修复 (SL Baur)
文件描述符耗尽错误的修复 (Tom)
修复以使 pg_dump 转储函数的 'iscachable' 标志 (Tom)
Append 节点目标列表中子查询的修复 (Tom)
归并连接计划的修复 (Tom)
修复带索引关系上的 TRUNCATE 失败 (Tom)
避免写错误时数据库范围重启 (Hiroshi)
修复 nodeMaterial 通过重算输出以遵循 chgParam (Tom)
修复当行版本的源和目的地位于同一页时 VACUUM 移动更新行版本链的问题 (Tom)
修复 user.c CommandCounterIncrement (Tom)
to_char() 中 AM/PM 边界问题的修复 (Karel Zak)
修复 TIME 聚合处理 (Tom)
修复 to_char() 以避免对 NULL 输入核心转储 (Tom)
缓冲区修复 (Tom)
修复向 char() 数据类型插入/复制较长多字节字符串的问题 (Tatsuo)

修复中止时后端崩溃的问题（Tom）

E.239. 版本 7.0.2

发行日期
2000-06-05

这是 7.0.1 的重新打包，附加了文档。

E.239.1. 迁移到版本 7.0.2

对于运行 7.* 的用户，不需要进行转储/恢复。

E.239.2. 变更

已把文档添加到压缩包。

E.240. 版本 7.0.1

发行日期
2000-06-01

这是 7.0 的清理版本。

E.240.1. 迁移到版本 7.0.1

对于运行 7.0 的用户，不需要进行转储/恢复。

E.240.2. 变更

修复许多 CLUSTER 失败（Tom）
允许 ALTER TABLE RENAME 用于索引（Tom）
修复 plpgsql 处理 datetime->timestamp 和 timespan->interval（Bruce）
新的 configure --with-setproctitle 开关以使用 setproctitle()（Marc、Bruce）
修复 6.5.3 起 ResultSet 的差一错误等
jdbc ResultSet 修复（Joseph Shraibman）
优化器调优（Tom）
为 pgaccess 修复 create user
修复 UNLISTEN 失败
IRIX 修复（David Kaelbling）
QNX 修复（Andreas Kardos）
降低 COPY IN 锁级别（Tom）
把 libpqeasy 改为使用 PQconnectdb() 风格参数（Bruce）
修复 pg_dump 处理 OID 索引（Tom）
修复小内存泄漏（Tom）

Solaris 上 createdb/dropdb 的修复 (Tatsuo)
非阻塞连接修复 (Alfred Perlstein)
修复 RENAME TABLE 失败后的不当恢复 (Tom)
把 pg_ident.conf.sample 复制到安装的 /lib 目录 (Bruce)
新增 SJIS UDC (NEC selection IBM 汉字) 支持 (Eiji Tokuya)
修复过长的 syslog 消息 (Tatsuo)
修复带引号索引名过长的问题 (Tom)
JDBC ResultSet.getTimestamp() 修复 (Gregory Krasnow & Floyd Marinescu)
ecpg 变更 (Michael)

E.241. 版本 7.0

发行日期
2000-05-08

此发行版包含许多领域的改进，展示了 PostgreSQL 的持续成长。7.0 的改进和修复比以往任何版本都多。开发者确信这是迄今最好的版本；我们尽力只发布扎实的版本，这个版本也不例外。

此发行版的主要变更：

Foreign Keys

外键现在已实现，PARTIAL MATCH 外键除外。许多用户一直要求此特性，我们很高兴提供它。

Optimizer Overhaul

延续一年前开始的工作，优化器得到改进，允许更好的查询计划选择，性能更快且内存使用更少。

Updated psql

我们的交互式终端监视器 psql 已更新，新增多种特性。详情见 psql 手册页。

Join Syntax

SQL92 连接语法现在受支持，但此版本只支持 INNER JOIN。JOIN、NATURAL JOIN、JOIN/USING 和 JOIN/ON 都可用，列相关名也可用。

E.241.1. 迁移到版本 7.0

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。对于从 6.5.* 升级的用户，可以改用 pg_upgrade 升级到此版本；但完整的转储/重装安装始终是最可靠的升级方法。

新版本需要考虑的接口和兼容性问题包括：

- 日期/时间类型 datetime 和 timespan 已被 SQL92 定义的 timestamp 和 interval 取代。虽然我们做了一些努力来简化过渡，允许 PostgreSQL 识别废弃的类型名并转换为新类型名，但这一机制对你的现有应用程序不可能是完全透明的。
- 优化器在查询代价估计方面得到显著改进。在某些情况下，由于优化器为首选计划做出更好的选择，查询时间会缩短。

但在少数情形下（通常涉及病态的数据分布），查询时间可能上升。如果你处理大量数据，可能想检查查询以验证性能。

- JDBC 和 ODBC 接口已升级和扩展。
- 字符串函数 CHAR_LENGTH 现在是原生函数。以前的版本把它转换为对 LENGTH 的调用，这可能与实现 LENGTH 的其他类型（如几何类型）产生歧义。

E.241.2. 变更

```
缺陷修复
---------
防止函数调用超过最大参数数量 (Tom)
改进 CASE 构造 (Tom)
修复 SELECT coalesce(f1,0) FROM int4_tbl GROUP BY f1 (Tom)
修复 SELECT sentence.words[0] FROM sentence GROUP BY sentence.words[0]
 (Tom)
修复 GROUP BY 扫描缺陷 (Tom)
SQL 语法处理的改进 (Tom)
修复涉及 INSERT ... SELECT ... 的视图的问题 (Tom)
修复 SELECT a/2, a/2 FROM test_missing_target GROUP BY a/2 (Tom)
修复 INSERT ... SELECT 中的子查询 (Tom)
防止 INSERT ... SELECT ... ORDER BY (Tom)
大于 2GB 关系的修复，包括 vacuum
改进系统表变更向其他后端的传播 (Tom)
改进用户表变更向其他后端的传播 (Tom)
修复复杂情形下临时表的处理 (Bruce、Tom)
允许在表打开时锁定表，改进并发可靠性 (Tom)
在 pg_dump 中正确引用序列名 (Ross J. Reedstrom)
防止他人访问时 DROP DATABASE
若无行被处理则防止 GROUP BY 返回任何行 (Tom)
修复无行匹配 WHERE 时 SELECT COUNT(1) FROM table WHERE ... (Tom)
修复 pg_upgrade 使其适用于 MVCC (Tom)
修复 SELECT ... WHERE x IN (SELECT ... HAVING SUM(x) > 1) (Tom)
修复 "f1 datetime DEFAULT 'now'" (Tom)
修复 DEFAULT 中使用 CURRENT_DATE 的问题 (Tom)
允许纯注释行，也允许 ;;; 行。 (Tom)
改进磁盘写入失败、磁盘满后的恢复 (Hiroshi)
修复 FROM 中提到表但未连接的情形 (Tom)
允许无聚合函数的 HAVING 子句 (Tom)
修复 "--" 注释且无尾随换行的问题，如 perl 接口中所见
改进 pg_dump 失败错误报告 (Bruce)
允许排序和哈希超过 2GB 文件大小 (Tom)
修复 pg_dump 转储继承规则的问题 (Tom)
修复 NULL 处理比较的问题 (Tom)
修复失败 CREATE/DROP 命令导致的不一致状态 (Hiroshi)
修复带连字符的数据库名的问题
防止 DROP INDEX 干扰其他后端 (Tom)
修复 verify_password() 中的文件描述符泄漏
修复 "Unable to identify an operator =$" 问题
修复 ODBC 在启用 CommLog 和 Debug 时不再段错误 (Dirk Niggemann)
修复递归 exit 调用 (Massimo)
修复超长时区的问题 (Jeroen van Vianen)
使 pg_dump 保留主键信息 (Peter E)
```

防止带单引号的数据库名 (Peter E)
防止事务内 DROP DATABASE (Peter E)
ecpg 内存泄漏修复 (Stephen Birch)
修复 SELECT null::text、SELECT int4fac(null) 和 SELECT 2 + (null) (Tom)
Y2K 时间戳修复 (Massimo)
修复 VACUUM 'HEAP_MOVED_IN was not expected' 错误 (Tom)
修复含空格表/列名的视图的问题 (Tom)
防止索引上的权限 (Peter E)
修复生成错误时自旋锁卡住的问题 (Hiroshi)
修复 Linux 上的 ipcclean
修复 NULL 约束条件的处理 (Tom)
修复 odbc 驱动中的内存泄漏 (Nick Gorham)
修复 UNION 表上的权限检查 (Tom)
修复以允许 SELECT 'a' LIKE 'a' (Tom)
修复 SELECT 1 + NULL (Tom)
CHAR 的修复
修复 numeric 类型上的 log() (Tom)
废弃 ':' 和 ';' 操作符
允许清理临时表
不允许与新列同名的继承列
磁盘空间耗尽时恢复或强制失败 (Hiroshi)
修复 AS 列与结果列匹配的 INSERT INTO ... SELECT
修复 INSERT ... SELECT ... GROUP BY 按目标列而非源列分组的问题 (Tom)
修复带 INSERT 的 CREATE TABLE test (a char(5) DEFAULT text '', b int4) (Tom)
修复带 LIMIT 的 UNION
修复 CREATE TABLE x AS SELECT 1 UNION SELECT 2
修复 CREATE TABLE test(col char(2) DEFAULT user)
修复 CREATE TABLE ... DEFAULT 中的类型不匹配
修复 SELECT * FROM pg_class where oid in (0,-1)
修复 SELECT COUNT('asdf') FROM pg_class WHERE oid=12
防止能创建数据库的用户修改 pg_database 表 (Peter E)
修复 btree 在键 > 1/2 (页 - 开销) 时给出有用的 elog (Tom)
修复向 DECIMAL(4,4) 字段插入 0.0 的问题 (Tom)

增强

基于 SQL3/SQL98 的新 CLI 接口头文件 sqlcli.h
移除查询长度的所有限制，行长度限制仍存在 (Tom)
把 jdbc 协议更新到 2.0 (Jens Glaser <jens@jens.de>)
添加 TRUNCATE 命令以快速截断关系 (Mike Mascari)
修复以给超级用户和 createdb 用户适当的更新目录权限 (Peter E)
允许 ecpg bool 变量具有 NULL 值 (Christof)
对无 NULL 指示符变量的 NULL 值发出 ecpg 错误 (Christof)
允许 ^C 取消 COPY 命令 (Massimo)
添加 SET FSYNC 和 SHOW PG_OPTIONS 命令 (Massimo)
动态装载 C 函数的函数名重载 (Frankpitt)
向 libpq++ 添加 CmdTuples() (Vince)
新 CREATE CONSTRAINT TRIGGER 和 SET CONSTRAINTS 命令 (Jan)
允许 CREATE FUNCTION/WITH 子句用于所有语言类型
configure --enable-debug 添加 -g (Peter E)
configure --disable-debug 移除 -g (Peter E)
允许更复杂的默认表达式 (Tom)
第一个真正可用的 FOREIGN KEY 约束触发器功能 (Jan)

添加 FOREIGN KEY ... MATCH FULL ... ON DELETE CASCADE (Jan)
添加 FOREIGN KEY ... MATCH <unspecified> 参照动作 (Don Baccus)
允许对 ctid (物理堆位置) 的 WHERE 限制 (Hiroshi)
把 pginterface 从 contrib 移到 interface 目录并改名为 pgeasy (Bruce)
更改 pgeasy connectdb() 参数顺序 (Bruce)
要求 SELECT DISTINCT 目标列表包含所有 ORDER BY 列 (Tom)
添加 Oracle 的 COMMENT ON 命令 (Mike Mascari <mascarim@yahoo.com>)
libpq 的 PQsetNoticeProcessor 函数现在返回先前的钩子 (Peter E)
防止 PQsetNoticeProcessor 被设为 NULL (Peter E)
使 COPY 中的 USING 可选 (Bruce)
允许目标列表中的子查询 (Tom)
允许比较操作符左侧的子查询 (Tom)
新并行回归测试 (Jan)
更改后端侧 COPY 使写文件的权限为 644 而不是 666 (Tom)
强制 PGDATA 目录权限安全，即使它已存在 (Tom)
添加 psql LASTOID 变量以返回最后插入的 oid (Peter E)
允许并发 vacuum 并移除 pg_vlock vacuum 锁文件 (Tom)
为 vacuum 添加权限检查 (Peter E)
新的允许异步连接的 libpq 函数：PQconnectStart()、PQconnectPoll()、PQresetStart()、PQresetPoll()、PQsetenvStart()、PQsetenvPoll()、PQsetenvAbort (Ewan Mellor)
新 libpq PQsetenv() 函数 (Ewan Mellor)
create/alter user 扩展 (Peter E)
$PGDATA 下新的 postmaster.pid 和 postmaster.opts (Tatsuo)
新 create/drop user/db 脚本 (Peter E)
psql 重大改造 (Peter E)
向 libpq 接口添加 const (Peter E)
新 libpq 函数 PQoidValue (Peter E)
显示导致 GROUP BY 问题的特定非聚合 (Tom)
使 pg_shadow 的更改重建 pg_pwd 文件 (Peter E)
添加 aggregate(DISTINCT ...) (Tom)
允许用标志控制 NULL 的 COPY 输入/输出 (Peter E)
使 postgres 用户默认有口令 (Peter E)
添加 CREATE/ALTER/DROP GROUP (Peter E)
所有管理脚本现在支持 --long 选项 (Peter E、Karel)
Vacuumdb 脚本现在支持 --all 选项 (Peter E)
ecpg 新的可移植 FETCH 语法
添加 ecpg 的 EXEC SQL IFDEF、EXEC SQL IFNDEF、EXEC SQL ELSE、EXEC SQL ELIF
和 EXEC SQL ENDIF 指令
添加 pg_ctl 脚本以控制后端启动 (Tatsuo)
添加 postmaster.opts.default 文件以存储启动标志 (Tatsuo)
允许 --with-mb=SQL_ASCII
把索引键的最大数量增加到 16 (Bruce)
把函数参数的最大数量增加到 16 (Bruce)
允许配置索引键和参数的最大数量 (Bruce)
允许非特权用户更改自己的口令 (Peter E)
启用口令认证；新用户必须使用 (Peter E)
不允许删除拥有数据库的用户 (Peter E)
把 initdb 选项 --with-mb 改为 --enable-multibyte
添加 initdb 提示超级用户口令的选项 (Peter E)
允许复杂类型转换如 col::numeric(9,2) 和 col::int2::float8 (Tom)
更新 initdb、initlocation、pg_dump、ipcclean 的用户界面 (Peter E)
新 pg_char_to_encoding() 和 pg_encoding_to_char() 函数 (Tatsuo)

libpq 非阻塞模式 (Alfred Perlstein)
改进未指定长度转换中类型的转换
新 plperl 内部编程语言 (Mark Hollomon)
允许 COPY IN 读取不以换行结尾的文件 (Tom)
指示长标识符何时被截断 (Tom)
允许聚合使用类型等价 (Peter E)
添加 Oracle 的 to_char()、to_date()、to_datetime()、to_timestamp()、to_number()
转换函数 (Karel Zak <zakkr@zf.jcu.cz>)
添加 SELECT DISTINCT ON (expr [, expr ...]) targetlist ... (Tom)
检查以确保 ORDER BY 与 DISTINCT 操作兼容 (Tom)
向 ODBC 添加 NUMERIC 和 int8 类型
改进 Append、Group、Agg、Unique 的 EXPLAIN 结果 (Tom)
添加 ALTER TABLE ... ADD FOREIGN KEY (Stephan Szabo)
允许 PL/pgSQL 中的 SELECT .. FOR UPDATE (Hiroshi)
即使到达 EOF 也启用反向顺序扫描 (Hiroshi)
添加布尔值及 >= 和 <= 的 btree 索引 (Don Baccus)
COPY FROM 失败时打印当前行号 (Massimo)
识别 POSIX 时区，如 "PST+8" 和 "GMT-8" (Thomas)
添加 DEC 作为 DECIMAL 的同义词 (Thomas)
添加 SESSION_USER 作为 SQL92 关键字，同 CURRENT_USER (Thomas)
实现 SQL92 列别名（也称相关名）(Thomas)
实现 SQL92 连接语法 (Thomas)
使 INTERVAL 保留字可用作列标识符 (Thomas)
实现 REINDEX 命令 (Hiroshi)
接受聚合函数 SUM(ALL col) 中的 ALL (Tom)
防止 GROUP BY 使用列别名 (Tom)
新 psql \encoding 选项 (Tatsuo)
允许 PQrequestCancel() 在等待锁状态下终止 (Hiroshi)
允许在所有情形下对负数取负
添加 ecpg 描述符 (Christof、Michael)
允许 CREATE VIEW v AS SELECT f1::char(8) FROM tbl
允许带长度的转换，如 foo::char(8)
新 libpq 函数 PQsetClientEncoding()、PQclientEncoding() (Tatsuo)
添加对 SJIS 用户定义字符的支持 (Tatsuo)
支持更大的视图/规则
使 libpq 的 PQconndefaults() 线程安全 (Tom)
禁用 // 作为注释以符合 ANSI，应使用 -- (Tom)
允许视图上的列别名 CREATE VIEW name (collist)
修复带子查询的视图的问题 (Tom)
允许 UPDATE table SET fld = (SELECT ...) (Tom)
SET 命令选项不再需要引号
把 pgaccess 更新到 0.98.6
新 SET SEED 命令
新 pg_options.sample 文件
新 SET FSYNC 命令 (Massimo)
允许创建表时 pg_descriptions
允许创建类型、列和函数时 pg_descriptions
允许 psql \copy 使用分隔符 (Peter E)
允许 psql 把 null 打印为与 "" 不同的 [null] (Peter E)

类型

大量数组修复 (Tom)

允许裸列名作为数组下标 (Tom)
改进 int 和 float 常量的类型转换 (Tom)
清理 int8 输入、范围检查和类型转换 (Tom)
修复 SELECT timespan('21:11:26'::time) (Tom)
netmask('x.x.x.x/0') 为 255.255.255.255 而不是 0.0.0.0 (Oleg Sharoiko)
在 NUMERIC 上添加 btree 索引 (Jan)
含 NUL 字符大对象的 Perl 修复 (Douglas Thomson)
大对象的 ODBC 修复 (free)
修复 cidr 数据类型的索引
修复以太网 MAC 地址 (macaddr 类型) 比较的问题
修复计算中发生溢出时日期/时间类型的问题 (Tom)
允许 int8 上的数组 (Peter E)
修复 NUMERIC 类型 (如 NUMERIC(4,4)) 的舍入/溢出问题 (Tom)
允许 NUMERIC 数组
修复 NUMERIC ceil() 和 floor() 函数中的缺陷 (Tom)
使 char_length()/octet_length 包含尾随空白 (Tom)
使 abstime/reltime 使用 int4 而不是 time_t (Peter E)
用于压缩文本字段的新 lztext 数据类型
修改处理 int 和 float 常量强制的代码 (Tom)
开始编写实现 BIT 和 BIT VARYING 类型的新代码 (Adriaan Joubert)
NUMERIC 现在接受科学记数法 (Tom)
NUMERIC 到 int4 舍入 (Tom)
正确地把 float4/8 转换为 NUMERIC (Tom)
允许与 NUMERIC 的类型转换 (Thomas)
使 ISO 日期风格 (2000-02-16 09:33) 成为默认 (Thomas)
添加 NATIONAL CHAR [VARYING] (Thomas)
允许 NUMERIC round 和 trunc 接受负标度 (Tom)
新 TIME WITH TIME ZONE 类型 (Thomas)
在 time 类型上添加 MAX()/MIN() (Thomas)
为 int8 添加 abs()、mod()、fac() (Thomas)
把函数改名为 float8 的 round()、sqrt()、cbrt()、pow() (Thomas)
为 float8 添加超越数学函数 (如 sin()、acos()) (Thomas)
为 NUMERIC 类型添加 exp() 和 ln()
把 NUMERIC power() 改名为 pow() (Thomas)
改进的 TRANSLATE() 函数 (Edwin Ramirez、Tom)
允许 X=-Y 操作符 (Tom)
允许 SELECT float8(COUNT(*))/(SELECT COUNT(*) FROM t) FROM t GROUP BY f1; (Tom)
允许 LOCALE 在正则表达式搜索中使用索引 (Tom)
允许创建函数索引时使用默认类型

性能

防止大量 AND 和 OR 的指数级空间消耗 (Tom)
收集系统列的属性选择率值 (Tom)
降低聚合的内存使用 (Tom)
修复 LIKE 优化以在多字节编码下使用索引 (Tom)
修复 r-tree 索引优化器选择率 (Thomas)
改进优化器选择率计算和函数 (Tom)
优化存在许多相等键情形的 btree 搜索 (Tom)
只在有索引时启用快速 LIKE 索引处理 (Tom)
重用带重复项索引页上的空闲空间 (Tom)
改进哈希连接处理 (Tom)
结果已排序时防止降序排序 (Hiroshi)

允许索引扫描查询限定的交换 (Tom)
在需要 ORDER BY/GROUP BY 的情形优先索引扫描 (Tom)
为性能把大内存请求按固定大小块分配 (Tom)
通过减少内存分配请求修复 vacuum 的性能 (Tom)
实现常量表达式化简 (Bernard Frankpitt、Tom)
用辅助列确定索引扫描的起点 (Hiroshi)
防止内部排序时四倍的磁盘空间使用 (Tom)
通过调用更少函数加快排序 (Tom)
创建与所有系统缓存匹配的系统索引 (Bruce、Hiroshi)
使系统缓存使用系统索引 (Bruce)
使所有系统索引唯一 (Bruce)
改进 pg_statistics 管理以提升 VACUUM 速度 (Tom)
降低后端缓存刷新频率 (Tom、Hiroshi)
COPY 现在重用先前的内存分配，改进性能 (Tom)
改进优化代价估计 (Tom)
改进范围查询 x > lowbound AND x < highbound 的优化器估计 (Tom)
在适当处使用 DNF 而不是 CNF (Tom、Taral)
进一步清理 OR-of-AND WHERE 子句 (Tom)
使用 OR 子句中的索引 (x = 1 AND y = 2) OR (x = 2 AND y = 4) (Tom)
随机索引页访问的更智能优化器计算 (Tom)
控制优化器代价的新 SET 变量 (Tom)
基于 LIMIT、OFFSET 和 EXISTS 限定优化查询 (Tom)
减少优化器连接路径的内部管理以提速 (Tom)
子查询大幅提速 (Tom)
未禁用 fsync 时减少 fsync 写 (Tom)
改进的 LIKE 优化器估计 (Tom)
防止 SELECT-only 查询中的 fsync (Vadim)
使索引创建使用 psort 代码，因为它现在更快 (Tom)
允许创建 > 1GB 的排序临时表

源代码树变更

修复 linux PPC 编译
新通用表达式树遍历子例程 (Tom)
把 form() 改为 varargform() 以防止可移植性问题
改进 Alpha 上大整数的范围检查
清理 /include 目录中的 #include (Bruce)
添加检查 include 的脚本 (Bruce)
从 *.c 文件移除不需要的 #include (Bruce)
更改 #include 以酌情使用 <> 和 "" (Bruce)
启用 libpq 的 Windows 编译
来自 Uncle George <gatgul@voicenet.com> 的 Alpha 自旋锁修复
优化器数据结构的彻底改造 (Tom)
修复 cygipc 库 (Yutaka Tanida)
允许 pgsql 在更新的 Cygwin 快照上工作 (Dan)
新目录版本号 (Tom)
添加 Linux ARM
把 heap_replace 改名为 heap_update
为 QNX 更新 (Dr. Andreas Kardos)
新平台特定的回归处理 (Tom)
把 oid8 改名为 oidvector、int28 改名为 int2vector (Bruce)
把所有 yacc 和 lex 文件包含进发行版 (Peter E.)
移除 lextest，不再需要 (Peter E)
修复 Windows 上的 libpq 和 psql (Magnus)

内部把 datetime 和 timespan 改为 timestamp 和 interval (Thomas)
修复 BSD/OS 上的 plpgsql
向回归测试添加 SQL_ASCII 测试用例 (Tatsuo)
configure --with-mb 现已废弃 (Tatsuo)
NT 修复
NetBSD 修复 (Johnny C. Lam <lamj@stat.cmu.edu>)
Alpha 编译修复
新多字节编码

E.242. 版本 6.5.3

发行日期
1999-10-13

这基本上是 6.5.2 的清理版本。我们添加了 6.5.2 中缺失的新 PgAccess，并安装了 NT 特定的修复。

E.242.1. 迁移到版本 6.5.3

对于运行 6.5.* 的用户，不需要进行转储/恢复。

E.242.2. 变更

更新了 pgaccess 0.98 版本
NT 特定补丁
修复继承表上规则的转储

E.243. 版本 6.5.2

发行日期
1999-09-15

这基本上是 6.5.1 的清理版本。我们修复了 6.5.1 用户报告的多种问题。

E.243.1. 迁移到版本 6.5.2

对于运行 6.5.* 的用户，不需要进行转储/恢复。

E.243.2. 变更

子查询+CASE 修复 (Tom)
为 solaris_i386 和 solaris_sparc 移植添加 SHLIB_LINK 设置 (Daren Sefcik)
WHERE 连接子句中 CASE 的修复 (Tom)
修复 BTScan 中止 (Tom)
修复对冗余 UNIQUE 和 PRIMARY KEY 索引的检查 (Thomas)
改进为可检查多列约束 (Thomas)

修复启用 MB 时 Windows 的制作问题 (Hiroki Kataoka)
允许 BSD yacc 和 bison 编译 pl 代码 (Bruce)
修复 SET NAMES 的功能
int8 修复 (Thomas)
修复 vacuum 的内存消耗 (Hiroshi、Tatsuo)
降低 vacuum 的总内存消耗 (Tom)
修复 timestamp(datetime)
规则反编译缺陷修复 (Tom)
修复 mkMakefile.tcldefs.sh.in 和 mkMakefile.tkdefs.sh.in 中的引号问题 (Tom)
这是为重用 vacuum 释放的索引页空间 (Vadim)
为 pg_dump 记录 -x (Bruce)
修复规则反编译器中的一元操作符 (Tom)
在 mdtruncate() 期间注释掉多余段的 FileUnlink (Tom)
来自 Yu Cao 的 IRIX 链接修复
修复 LIKE 中的逻辑错误：到达模式末尾而文本未结束时不应返回 LIKE_ABORT (Tom)
修复事务中止期间堆内存分配的不当清理 (Tom)
更新 pgaccess 0.98 版本

E.244. 版本 6.5.1

发行日期
1999-07-15

这基本上是 6.5 的清理版本。我们修复了 6.5 用户报告的多种问题。

E.244.1. 迁移到版本 6.5.1

对于运行 6.5 的用户，不需要进行转储/恢复。

E.244.2. 变更

添加 NT README 文件
linux_ppc、IRIX、linux_alpha、OpenBSD、alpha 的可移植性修复
移除 QUERY_LIMIT，使用 SELECT...LIMIT
修复继承上的 EXPLAIN (Tom)
补丁允许对多段表进行清理 (Hiroshi)
R-Tree 优化器选择率修复 (Tom)
ACL 文件描述符泄漏修复 (Atsushi Ogawa)
新表达式子树代码 (Tom)
只读事务避免磁盘写 (Vadim)
修复最后事务中止时临时表的移除 (Bruce)
防止创建过大的行 (Bruce)
plpgsql 修复
允许 32k - 64k 的端口号 (Bruce)
添加 ^ 优先级 (Bruce)
把名为 pg_temp 的排序文件改名为 pg_sorttemp (Bruce)
修复时间值中的微秒 (Tom)
教程源码清理
新 linux_m68k 移植

修复某些情形下 NULL 的排序（Tom）
修复共享库依赖（Tom）
修复影响子查询中 GROUP BY 的故障（Tom）
修复一些编译器警告（Tomoaki Nishiyama）
新增 Win1250（捷克语）支持（Pavel Behal）

E.245. 版本 6.5

发行日期
1999-06-09

此版本标志着开发团队对从 Berkeley 继承的源代码的掌握迈出了一大步。你会看到我们现在能轻松添加重大特性，这要归功于我们全球开发团队不断扩大的规模和日益丰富的经验。

以下是较显著变更的简要总结：

多版本并发控制（MVCC）

这移除了我们旧的表级锁定，代之以优于大多数商业数据库系统的锁定系统。
在传统系统中，每个被修改的行都被锁定直到提交，阻止其他用户读取。MVCC 利用 PostgreSQL 天生的多版本特性，允许读者在写入者活动期间继续读取一致的数据。写入者继续使用紧凑的 pg_log 事务系统。这一切都不需要像传统数据库系统那样为每行分配锁。因此，基本上我们不再受简单表级锁定的限制；我们有了比行级锁定更好的东西。

Hot backups from pg_dump

pg_dump 利用新的 MVCC 特性，在数据库保持在线并可用于查询的同时给出一致的数据库转储/备份。

Numeric data type

我们现在有了真正的数字数据类型，精度由用户指定。

Temporary tables

临时表保证在数据库会话内名称唯一，并在会话退出时销毁。

New SQL features

我们现在支持 CASE、INTERSECT 和 EXCEPT 语句。新增了 LIMIT/OFFSET、SET TRANSACTION ISOLATION LEVEL、SELECT ... FOR UPDATE 以及改进的 LOCK TABLE 命令。

Speedups

得益于团队内的多种才能，我们继续加快 PostgreSQL 的速度。我们加速了内存分配、优化、表连接和行传输例程。

Ports

我们继续扩大移植列表，这次包括 Windows NT/ix86 和 NetBSD/arm32。

Interfaces

大多数接口有了新版本，现有功能得到改进。

Documentation

文档中到处都有新的和更新的材料。SGI 和 AIX 平台有了新的 FAQ。教程有 Stefan Simkovics 编写的 SQL 入门信息。用户指南有涵盖 postmaster 和更多工具程序的参考页，新附录包含日期/时间行为的细节。管理员指南有 Tom Lane 编写的故障排除新章节。程序员指南有 Stefan 编写的查询处理描述，以及通过匿名 CVS 和 CVSup 获取 PostgreSQL 源代码树的细节。

E.245.1. 迁移到版本 6.5

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。pg_upgrade 不能用于升级到此版本，因为表的磁盘结构相对于以前的版本已更改。

新的多版本并发控制（MVCC）特性在多用户环境中可能给出略为不同的行为。
请阅读并理解以下章节，以确保你的现有应用程序能给出你需要的行为。

E.245.1.1. 多版本并发控制

由于 6.5 中的读者不锁定数据（无论事务隔离级别如何），一个事务读取的数据可能被另一个事务覆写。换句话说，如果 `SELECT` 返回了一行，并不意味着该行在被返回时（即语句或事务开始之后的某个时刻）确实存在，也不意味着该行在当前事务提交或回滚之前受到保护，
不被并发事务删除或更新。

为确保某行确实存在并保护它不被并发更新，必须使用 `SELECT FOR UPDATE` 或适当的 `LOCK TABLE` 语句。从以前的 PostgreSQL 版本和其他环境移植应用程序时应考虑到这一点。

如果你在使用 `contrib/refint.*` 触发器做参照完整性，请记住上述内容。现在需要额外的技术。一种方法是：如果事务要更新/删除主键，使用 `LOCK parent_table IN SHARE ROW EXCLUSIVE MODE` 命令；如果事务要更新/插入外键，使用 `LOCK parent_table IN SHARE MODE` 命令。

> **注意**
>
> 注意，如果你以 SERIALIZABLE 模式运行事务，则必须在执行事务中的任何 DML 语句（`SELECT/INSERT/DELETE/UPDATE/FETCH/COPY_TO`）之前执行上述 `LOCK` 命令。

当实现读取脏（未提交）数据的能力（无论隔离级别）和真正的参照完整性后，
这些不便将消失。

E.245.2. 变更

```
缺陷修复
---------
修复 text<->float8 和 text<->float4 转换函数 (Thomas)
修复创建带大小写混合约束表的问题 (Billy)
更改 exp()/pow() 行为以在下溢/上溢时报错 (Jan)
修复 pg_dump -z 中的缺陷
内存越界清理 (Tatsuo)
修复 lo_import 崩溃 (Tatsuo)
调整数据类型名的处理以抑制双引号 (Thomas)
```

用类型强制匹配列和 DEFAULT (Thomas)
修复死锁，使其在一秒睡眠后只检查一次 (Bruce)
聚合和 PL/pgsql 的修复 (Hiroshi)
修复子查询崩溃 (Vadim)
修复 libpq 函数 PQfnumber 和大小写不敏感名称的问题 (Bahman Rafatjoo)
修复大对象中间写入、无额外块和内存消耗的问题 (Tatsuo)
修复 pg_dump -d 或 -D 以及 INSERT 中特殊字符的引号问题
修复 dynahash 的严重问题 (Tom)
修复 INET/CIDR 可移植性问题
修复 ALTER TABLE ADD COLUMN 中选择率错误的问题 (Bruce)
修复执行器使不同列类型的归并连接可用 (Tom)
修复 Alpha OR 选择率缺陷
修复 OR 索引选择率问题 (Bruce)
修复 \d 为 char()/varchar() 显示正确长度的问题 (Ryan)
修复教程代码 (Clark)
改进 destroyuser 检查 (Oliver)
修复 Kerberos 的问题 (Rodney McDuff)
修复存在脏缓冲区时删除数据库的问题 (Bruce)
修复使 sequence nextval() 可区分大小写的问题 (Bruce)
修复 !!= 操作符
销毁数据库文件前丢弃缓冲区 (Bruce)
修复执行器两次求值函数的情形 (Tatsuo)
允许 sequence nextval 动作区分大小写 (Bruce)
修复优化器对负数索引不工作的问题 (Bruce)
修复执行器中 fjIsNull 的内存泄漏
修复聚合内存泄漏 (Erik Riedel)
允许含连字符的用户名授予权限
清理 inet 类型中的 NULL
清理系统表缺陷 (Tom)
修复 PAGER 和 \? 命令的问题 (Masaaki Sakaida)
把默认多段文件大小限制降为 1GB (Peter)
修复 CREATE OPERATOR 的转储 (Tom)
修复游标反向扫描 (Hiroshi Inoue)
修复使用 \i 时的 COPY FROM STDIN (Tom)
修复子查询在表达式内比较的问题 (Jan)
修复返回行时的错误报告处理 (Tom)
修复数组类型引用的问题 (Tom、Jan)
防止 UPDATE SET oid (Jan)
修复 pg_dump 使 -t 选项可处理大小写敏感的表名
特殊情形 GROUP BY 的修复 (Tom、Jan)
修复失败查询的内存泄漏 (Tom)
DEFAULT 现在支持大小写混合的标识符 (Tom)
修复表和索引 DROP/RENAME 的多段使用 (Ole Gjerde)
禁用 pg_dump 同时使用 -o 和 -d 选项 (Bruce)
允许 pg_dump 正确转储组权限 (Bruce)
修复 INSERT INTO table SELECT * FROM table2 中的 GROUP BY (Jan)
修复视图中的计算 (Jan)
修复数组索引上的聚合 (Tom)
修复 DEFAULT 处理值中需要过多引号的单引号的问题
修复非超级用户导入/导出大对象的安全问题 (Tom)
创建表的事务回滚现在正确清理 (Tom)
修复以允许长表名和列名生成正确的序列名 (Tom)

增强

```
------------
添加 "vacuumdb" 工具
通过更好的内存分配加速 libpq (Tom)
EXPLAIN 显示所有使用的索引 (Tom)
实现 CASE、COALESCE、NULLIF 表达式 (Thomas)
新 pg_dump 表输出格式 (Constantin)
添加字符串 min()/max() 函数 (Thomas)
把新类型强制技术扩展到聚合 (Thomas)
新 moddatetime contrib (Terry)
更新到 pgaccess 0.96 (Constantin)
为单字节 "char" 类型添加例程 (Thomas)
改进的 substr() 函数 (Thomas)
改进的多字节处理 (Tatsuo)
多版本并发控制/MVCC (Vadim)
新 Serialized 模式 (Vadim)
修复超过 2GB 表的问题 (Peter)
新 SET TRANSACTION ISOLATION LEVEL (Vadim)
新 LOCK TABLE IN ... MODE (Vadim)
更新 ODBC 驱动 (Byron)
新 NUMERIC 数据类型 (Jan)
新 SELECT FOR UPDATE (Vadim)
处理输入值的 "NaN" 和 "Infinity" (Jan)
改进的日期/年份处理 (Thomas)
改进的后端连接处理 (Magnus)
日志文件的新选项 ELOG_TIMESTAMPS 和 USE_SYSLOG (Massimo)
新 TCL_ARRAYS 选项 (Massimo)
新 INTERSECT 和 EXCEPT (Stefan)
用于主键跟踪的新 pg_index.indisprimary (D'Arcy)
允许在创建前删除表的新 pg_dump 选项 (Brook)
行输出例程加速 (Tom)
新 READ COMMITTED 隔离级别 (Vadim)
新 TEMP 表/索引 (Bruce)
结果已排序时防止排序 (Jan)
新内存分配优化 (Jan)
允许 psql 执行 \p\g (Bruce)
允许多个规则动作 (Jan)
添加 LIMIT/OFFSET 功能 (Jan)
改进连接大量表时的优化器 (Bruce)
来自 S. Simkovics 硕士论文的新 SQL 介绍 (Stefan、Thomas)
来自 S. Simkovics 硕士论文的后端处理新介绍 (Stefan)
改进的 int8 支持 (Ryan Bradetich、Thomas、Tom)
int8 与 text/varchar 类型之间转换的新例程 (Thomas)
新浓密计划，其中连接元表 (Bruce)
默认启用右手查询 (Bruce)
允许在 configure 时设置可靠的最大后端数
      (--with-maxbackends 和 postmaster 开关 (-N backends) ) (Tom)
因优化器提速，GEQO 默认现为 10 表 (Tom)
为 MS-SQL 可移植性允许 NULL=Var (Michael、Bruce)
修改 contrib check_primary_key() 使其可为 "automatic" 或
 "dependent" (Anand)
允许 psql 对视图的 \d 显示查询 (Ryan)
LIKE 提速 (Bruce)
Ecpg 修复/特性，参见 src/interfaces/ecpg/ChangeLog 文件 (Michael)
JDBC 修复/特性，参见 src/interfaces/jdbc/CHANGELOG (Peter)
```

使 % 操作符具有与 / 相同的优先级 (Bruce)
添加新 postgres -O 选项以允许系统表结构更改 (Bruce)
更新 contrib/pginterface/findoidjoins 脚本 (Tom)
带索引已删行 vacuum 的大幅提速 (Vadim)
允许非 SQL 函数根据参数运行不同版本 (Tom)
添加 -E 选项显示 \dt 等发送的实际查询 (Masaaki Sakaida)
在 psql 启动横幅中添加版本号 (Masaaki Sakaida)
新 contrib/vacuumlo 移除未被引用的大对象 (Peter)
新表大小初始化使未清理的表性能更好 (Tom)
改进连接被拒绝时的错误消息 (Tom)
支持 char() 和 varchar() 字段数组 (Massimo)
哈希代码的彻底改造以提高可靠性和性能 (Tom)
更新到 PyGreSQL 2.4 (D'Arcy)
更改调试选项使 -d4 和 -d5 产生不同的节点显示 (Jan)
新 pg_options: pretty_plan、pretty_parse、pretty_rewritten (Jan)
更好的系统表访问优化统计 (Tom)
更好地处理非默认块大小 (Massimo)
改进 GEQO 优化器内存消耗 (Tom)
UNION 现在支持对不在目标列表中的列 ORDER BY (Jan)
libpq++ 的重大改进 (Vince Vielhaber)
pg_dump 现在默认使用 -z (ACL) (Bruce)
后端缓存和内存提速 (Tom)
让 pg_dump 在一个快照事务中完成所有工作 (Vadim)
修复大对象内存泄漏和 pg_dump 问题 (Tom)
INET 类型现在比较时遵循网络掩码
使 VACUUM ANALYZE 只使用读锁 (Vadim)
允许 UNION 上的视图 (Jan)
pg_dump 现在能在活动数据库上生成一致快照 (Vadim)

源代码树变更

改进移植匹配 (Tom)
SunOS 可移植性修复
添加 Windows NT 后端移植并启用动态装载 (Magnus 和 Daniel Horak)
新移植到运行 Linux 的 Cobalt Qube (Mips) (Tatsuo)
移植到 NetBSD/m68k (Mr. Mutsuki Nakajima)
移植到 NetBSD/sun3 (Mr. Mutsuki Nakajima)
移植到 NetBSD/macppc (Toshimi Aoki)
修复 tcl/tk 配置 (Vince)
移除规则查询的 CURRENT 关键字 (Jan)
NT 动态装载现在可用 (Daniel Horak)
添加 ARM32 支持 (Andrew McMurry)
更好地支持 HP-UX 11 和 UnixWare
改进文件处理使其更统一，防止文件描述符泄漏 (Tom)
plpgsql 的新安装命令 (Jan)

E.246. 版本 6.4.2

发行日期
1998-12-20

6.4.1 版本打包不当。此版本还包含一个额外的缺陷修复。

E.246.1. 迁移到版本 6.4.2

对于运行 6.4.* 的用户，不需要进行转储/恢复。

E.246.2. 变更

修复某些平台上日期时间常量的问题（Thomas）

E.247. 版本 6.4.1

发行日期
1998-12-18

这基本上是 6.4 的清理版本。我们修复了 6.4 用户报告的多种问题。

E.247.1. 迁移到版本 6.4.1

对于运行 6.4 的用户，不需要进行转储/恢复。

E.247.2. 变更

新增 pg_dump -N 标志以强制标识符周围加双引号。这是默认（Thomas）
修复 where 子句中 NOT 导致崩溃的问题（Bruce）
EXPLAIN VERBOSE 核心转储修复（Vadim）
修复 Linux 上的共享库问题
修复表存在性测试以允许表名中的大小写混合和空白（Thomas）
修复几个 pg_dump 缺陷
Configure 更好地匹配 template/.similar 条目（Tom）
把内建函数名从 SPI_* 改为 spi_*
OR WHERE 子句修复（Vadim）
大小写混合表名的修复（Billy）
contrib/linux/postgres.init.csh/sh 修复（Thomas）
libpq 内存越界修复
SunOS 修复（Tom）
更改 exp() 行为在下溢时报错（Thomas）
pg_dump 对内存泄漏、继承约束、布局变更的修复
把 pgaccess 更新到 0.93
64 位平台的原型修复
多字节修复（Tatsuo）
新 ecpg 手册页
内存越界修复（Tatsuo）
lo_import() 崩溃修复（Bruce）
更好地搜索安装程序（Tom）
时区修复（Tom）
HP-UX 修复（Tom）
用隐式类型强制匹配 DEFAULT 值（Thomas）

添加帮助处理单字节（内部）字符类型的例程（Thomas）
Windows 下 libpq 编译修复（Magnus）
升级到 PyGreSQL 2.2（D'Arcy）

E.248. 版本 6.4

发行日期
1998-10-30

此版本有许多新特性和改进。感谢我们的开发者和维护者，
自上一个版本以来系统的几乎每个方面都得到了关注。以下是简要而不完整的总结：

- 得益于 Jan Wieck 在重写规则系统中的大量新代码，视图和规则现在可以工作了。他还为程序员指南写了相关章节。
- Jan 还贡献了第二个过程语言 PL/pgSQL，与他在上个版本贡献的最初的 PL/pgTCL 过程语言相伴。
- 我们有 Tatsuo Ishii 提供的可选多字节字符集支持，以补充现有的区域设置支持。
- 客户端/服务器通信得到了清理，感谢 Tom Lane，对异步消息和中断的支持更好了。
- 解析器现在执行自动类型强制，以把参数匹配到可用的操作符和函数，
  并把列和表达式与目标列匹配。这使用了支持 PostgreSQL 类型可扩展特性的通用机制。用户指南有涵盖此主题的新章节。
- 新增了三种数据类型。其中两种，inet 和 cidr，支持各种形式的 IP 网络、子网和机器寻址。某些平台上现在有 8 字节整数类型可用。详情见用户指南中的数据类型章节。第四种类型 serial 现在被解析器支持，作为 int4 类型、序列和唯一索引的混合体。
- 新增了更多 SQL92 兼容的语法特性，包括 INSERT DEFAULT VALUES
- 自动配置和安装系统得到了一些关注，应比以往在更多平台上更健壮。

E.248.1. 迁移到版本 6.4

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 或 pg_dumpall 进行转储/恢复。

E.248.2. 变更

缺陷修复

修复 PQsetdb/PQfinish 中的小内存泄漏（Bryan）
移除 char2-16 数据类型，使用 char/varchar（Darren）
Pqfn 现在处理 NOTICE 消息（Anders）
降低多后端时自旋锁的忙等开销（dg）
卡死自旋锁检测（dg）
修复 "ISO 风格" timespan 的解码和编码（Thomas）

修复事务回滚后删除表的问题 (Vadim)
更改错误消息并移除无效的更新消息 (Vadim)
修复 COPY 数组检查的问题
修复 SELECT 1 UNION SELECT NULL
修复大对象调用中的缓冲区泄漏 (Pascal)
把 owner 从 oid 类型改为 int4 (Bruce)
修复 Oracle 兼容函数 btrim()、ltrim() 和 rtrim() 中的缺陷
修复共享失效缓存溢出 (Massimo)
防止失败的 COPY 中的文件描述符泄漏 (Bruce)
修复 libpgtcl 的 pg_select 中的内存泄漏 (Constantin)
修复超过 8 字符的用户名/口令问题 (Tom)
修复后端处理异步 NOTIFY 的问题 (Tom)
修复许多错误的系统表条目 (Tom)

增强

升级 ecpg 和 ecpglib，参见 src/interfaces/ecpc/ChangeLog (Michael)
在 EXPLAIN 中显示使用的索引 (Zeugswetter)
EXPLAIN 调用规则系统并显示被重写查询的计划 (Jan)
通过 configure 使许多数据类型和函数感知多字节 (Tatsuo)
新 configure --with-mb 选项 (Tatsuo)
新 initdb --pgencoding 选项 (Tatsuo)
新 createdb -E 多字节选项 (Tatsuo)
Select version(); 现在返回 PostgreSQL 版本 (Jeroen)
libpq 现在允许异步客户端 (Tom)
允许从客户端取消后端查询 (Tom)
psql 现在可用 Control-C 取消查询 (Tom)
libpq 用户不必发送哑查询即可获取 NOTIFY 消息 (Tom)
NOTIFY 现在发送发送者的 PID，可以判断是否是自己的 (Tom)
PGresult 结构现在包含相关的错误消息（如果有） (Tom)
定义 date_part() 的 "tz_hour" 和 "tz_minute" 参数 (Thomas)
添加 varchar 和 bpchar 之间转换的例程 (Thomas)
添加允许把 varchar 和 bpchar 调整到目标列大小的例程 (Thomas)
添加位标志以支持数据检索中的时区小时和分钟 (Thomas)
允许有效浮点数的更多变体（如 ".1"、"1e6"） (Thomas)
修复带前导空格的一元负号解析 (Thomas)
按 SQL92 规范实现 TIMEZONE_HOUR、TIMEZONE_MINUTE (Thomas)
检查并正确忽略 FOREIGN KEY 列约束 (Thomas)
按 SQL92 规范定义 USER 为 CURRENT_USER 的同义词 (Thomas)
启用 HAVING 子句但其他地方尚无修复。
使 "char" 类型成为 "char(1)" 的同义词（实际实现为 bpchar） (Thomas)
为 DEFAULT 子句处理保存指定的字符串类型 (Thomas)
强制涉及不同数据类型的操作 (Thomas)
允许对不同类型列的某些索引使用 (Thomas)
添加自动类型转换能力 (Thomas)
清理大对象，使文件在打开时被截断 (Peter)
Readline 清理 (Tom)
允许 psql \f \ 用空格作为分隔符 (Bruce)
把 pg_attribute.atttypmod 传递给前端以获取列字段长度 (Tom、Bruce)
/contrib 中的 Msql 兼容库 (Aldrin)
移除 ORDER/GROUP BY 子句标识符必须
包含在目标列表中的要求 (David)
转换列以匹配 UNION 子句中的列 (Thomas)
移除 fork()/exec() 只做 fork() (Bruce)

Jdbc 清理 (Peter)
在 ps 命令行上显示后端状态 (只在某些平台上有效) (Bruce)
Pg_hba.conf 的 database 字段现在有 sameuser 选项
使 lo_unlink 接受 oid 参数而不是 int4
为无法处理我们宏的编译器新增 DISABLE_COMPLEX_MACRO (Bruce)
Libpgtcl 现在把 NOTIFY 作为 Tcl 事件处理，不必发送哑查询 (Tom)
libpgtcl 清理 (Tom)
向 libpgtcl 的 pg_result 命令添加 -error 选项 (Tom)
新区域设置补丁，参见 docs/README/locale (Oleg)
修复 pg_dump 使 CONSTRAINT 和 CHECK 语法正确 (ccb)
新 contrib/lo 代码用于移除孤立大对象 (Peter)
多字节特性的新 psql 命令 "SET CLIENT_ENCODING TO 'encoding'",
参见 /doc/README.mb (Tatsuo)
contrib/noupdate 代码用于撤销列上的更新权限
libpq 现在可以在 Windows 上编译 (Magnus)
在 libpq 中添加 PQsetdbLogin()
新 8 字节整数类型，由 configure 检查操作系统支持 (Thomas)
更好地支持带引号的表/列名 (Thomas)
在 pg_dump 中用双引号包围表和列名 (Thomas)
PQreset() 现在可用于口令 (Tom)
处理 GROUP BY 目标列表列号越界的情形 (David)
允许子查询中的 UNION
向 \d? 命令添加屏幕自动调整大小 (Bruce)
用 UNION 在一个查询中显示所有 \d? 结果 (Bruce)
添加 \d? 字段搜索特性 (Bruce)
Pg_dump 发出更少的 \connect 请求 (Tom)
使 pg_dump -z 标志更好地工作并在手册页中记录 (Tom)
添加对子查询和 union 完全支持的 HAVING 子句 (Stephan)
contrib/fulltextindex 中的全文索引例程 (Maarten)
事务 id 现在存储在共享内存中 (Vadim)
发出 COPY 命令时的新 PGCLIENTENCODING (Tatsuo)
支持 SQL92 语法 "SET NAMES" (Tatsuo)
支持 LATIN2-5 (Tatsuo)
添加 UNICODE 回归测试用例 (Tatsuo)
锁管理器清理，LLL 的新锁定模式 (Vadim)
允许 OR 子句使用索引 (Bruce)
允许 "SELECT NULL ORDER BY 1;"
Explain VERBOSE 打印计划，现在还把计划美化打印到
postmaster 日志文件 (Bruce)
向 \d 命令添加索引显示 (Bruce)
允许对函数 GROUP BY (David)
大对象的新 pg_class.relkind (Bruce)
把 libpq NOTICE 消息发送到其他位置的新方法 (Tom)
psql 的新 \w 写入命令 (Bruce)
新 /contrib/findoidjoins 扫描 oid 列以发现连接关系 (Bruce)
检查含常量的限制子句的有效索引时
允许考虑二进制兼容的索引 (Thomas)
/contrib/isbn_issn 中的新 ISBN/ISSN 代码
允许 NOT LIKE、IN、NOT IN、BETWEEN 和 NOT BETWEEN 约束 (Thomas)
新重写系统修复了规则和视图的许多问题 (Jan)

* 关系上的规则可用
* insert/update/delete 上的事件限定可用
* 新的 OLD 变量引用 CURRENT，CURRENT 未来将被移除
* 更新规则可在规则限定/动作中引用 NEW 和 OLD

```
        * 视图上的 insert/update/delete 规则可用
        * 现在支持多个规则动作，用括号包围
        * 普通用户可以在有 RULE 权限的表上创建视图/规则
        * 规则和视图继承创建者的权限
        * 没有列级别的规则
        * 没有 UPDATE NEW/OLD 规则
        * 新 pg_tables、pg_indexes、pg_rules 和 pg_views 系统视图
        * SELECT 规则只有单个动作
        * 彻底的重写改造，也许在 6.5
        * 处理子查询
        * 处理视图上的聚合
        * 处理 insert into select from view 可用
系统索引现在是多键的 (Bruce)
移除 Oidint2、oidint4 和 oidname 类型 (Bruce)
更多系统表查找使用系统缓存 (Bruce)
backend/pl 中的新后端编程语言 PL/pgSQL (Jan)
新 SERIAL 数据类型，自动创建序列/索引 (Thomas)
无需重新编译即可启用断言检查 (Massimo)
用户锁增强 (Massimo)
设置序列值的新 setval() 命令 (Massimo)
启动时若无 postmaster 运行则自动移除 unix 套接字文件 (Massimo)
条件跟踪包 (Massimo)
新 UNLISTEN 命令 (Massimo)
psql 和 libpq 现在可用 win32.mak 在 Windows 下编译 (Magnus)
Lo_read 不再存储尾随 NULL (Bruce)
标识符现在内部截断为 31 字符 (Bruce)
Createuser 选项现在可在命令行上使用
添加 64 位整数支持代码，configure 已测试，int8 类型 (Thomas)
防止失败 COPY 的文件描述符泄漏 (Bruce)
新 pg_upgrade 命令 (Bruce)
更新的 /contrib 目录 (Massimo)
新 CREATE TABLE DEFAULT VALUES 语句可用 (Thomas)
新 INSERT INTO TABLE DEFAULT VALUES 语句可用 (Thomas)
新 DECLARE 和 FETCH 特性 (Thomas)
libpq 的内部结构现在不导出 (Tom)
允许多达 8 键的索引 (Bruce)
移除 ARCHIVE 关键字，它不再使用 (Thomas)
pg_dump -n 标志以抑制标识符周围的引号
禁用视图的系统列 (Jan)
用于网络地址的新 INET 和 CIDR 类型 (TomH、Paul)
psql 输出中不再有双引号
pg_dump 现在转储视图 (Terry)
新 SET QUERY_LIMIT (Tatsuo、Jan)

源代码树变更
--------------------
/contrib 清理 (Jun)
内联一些每行都调用的小函数 (Bruce)
Alpha/linux 修复
HP-UX 清理 (Tom)
多字节回归测试 (Soonmyung.)
从 configure 移除 --disabled 选项
定义 PGDOC 默认使用 POSTGRESDIR
使回归可选
```

移除 pgindent 的额外花括号代码 (Bruce)
添加 bsdi 共享库支持 (Bruce)
新 --without-CXX 支持 configure 选项 (Brook)
新 FAQ_CVS
更新 tools/backend 中的后端流程图 (Bruce)
把 atttypmod 从 int16 改为 int32 (Bruce、Tom)
为没有 Getrusage() 的平台提供修复 (Tom)
把 PQconnectdb、PGUSER、PGPASSWORD 添加到 libpq 手册页
NS32K 平台修复 (Phil Nelson、John Buller)
SCO 7/UnixWare 2.x 修复 (Billy 等)
Sparc/Solaris 2.5 修复 (Ryan)
Pgbuiltin.3 已过时，移到 doc 文件 (Thomas)
更多的文档 (Thomas)
Nextstep 支持 (Jacek)
Aix 支持 (David)
pginterface 手册页 (Bruce)
所有共享库都有版本号
把所有操作系统特定的共享库定义合并到一个文件
更智能的 TCL/TK 配置检查 (Billy)
更智能的 perl 配置 (Brook)
未找到 install 脚本时 configure 使用随附的 install-sh (Tom)
用于共享库配置的新 Makefile.shlib (Tom)

E.249. 版本 6.3.2

发行日期
1998-04-07

这是 6.3.x 的缺陷修复版本。有关 6.3 版本系列新特性更完整的摘要，请参阅 6.3 版本的发行说明。

摘要：

- 修复某些平台（包括 Linux）的自动配置支持。
- 正确处理 BETWEEN 和 LIKE 子句左侧的函数调用。

对于运行 6.3 或 6.3.1 的用户，不需要进行转储/恢复。只需做一次 make distclean、make 和 make install。最后一步应在 postmaster 未运行时执行。你应重新链接任何使用 PostgreSQL 库的自定义应用程序。

对于从 6.3 之前安装的升级，请参阅 6.3 版本的安装和迁移说明。

E.249.1. 变更

对 tcl/tk 的配置检测改进 (Brook Milligan、Alvin)
手册页改进 (Bruce)
BETWEEN 和 LIKE 修复 (Thomas)
修复 pg_dump 所用 psql \connect 的问题 (Oliver Elphick)
新 odbc 驱动

pgaccess 0.86 版
移除 qsort，现使用 libc 版本并清理 (Jeroen)
修复检测到的缓冲区越界 (Maurice Gittens)
修复 libpgtcl 中的缓冲区越界 (Randy Kunkee)
修复带 DISTINCT 或 ORDER BY 的 UNION (Bruce)
gettimeofday configure 检查 (Doug Winterburn)
修复 "indexes not used" 缺陷 (Vadim)
文档补充 (Thomas)
修复后端内存泄漏 (Bruce)
libreadline 清理 (Erwan MAS)
移除 DISTDIR (Bruce)
Makefile 依赖清理 (Jeroen van Vianen)
ASSERT 修复 (Bruce)

E.250. 版本 6.3.1

发行日期
1998-03-23

摘要：

- 对多字节字符集的额外支持。
- 修复混合端序客户端和服务器的字节顺序。
- 对允许的 SQL 语法做小的更新。
- 改进安装时配置的自动检测。

对于运行 6.3 的用户，不需要进行转储/恢复。只需做一次 make distclean、make 和 make install。最后一步应在 postmaster 未运行时执行。你应重新链接任何使用 PostgreSQL 库的自定义应用程序。

对于从 6.3 之前安装的升级，请参阅 6.3 版本的安装和迁移说明。

E.250.1. 变更

ecpg 清理/修复，现为 1.1 版 (Michael Meskes)
pg_user 清理 (Bruce)
pg_dump 和 tclsh 的大对象修复 (alvin)
修复多个相邻下划线的 LIKE
修复内建函数的重定义 (Thomas)
ultrix4 清理
升级到 pg_access 0.83
更新 CLUSTER 手册页
多字节字符集支持，参见 doc/README.mb (Tatsuo)
configure --with-pgport 修复
pg_ident 修复
大对象位图修复 (Massimo)
修复所有 postgres 用户的 Postgres 大小写问题

修复 CREATE TRIGGER 中的对齐错误
pg_dump 的国家语言支持 (Fabien.Mabille)
关于把 max vector 消息长度乘以 3 的查询
为 libpq 添加自定义标题
修复 libpq 中断时的内存泄漏 (D. Sears)
修复 pg_proc 定义中的问题
pg_dump 修复 (Bruce)
修复 isnull 缺陷 (José Soares)
透明的多字节转换
上下文和位置错误消息 (Ian Burrell)
修复小内存泄漏
拆除 libpq 中的所有代码重复
乱序对象修复
新索引估计器
启用单个动作规则重写 (Jan)
外连接语法现在生成查询计划 (Thomas)
>= 与 <= 与 = 的 btree 索引
服务器对客户端编码转换指令 (Tatsuo)
对 EUC_TW 转换的 BIG5 修复 (Tatsuo)

E.251. 版本 6.3

发行日期

1998-03-01

此版本有许多新特性和改进。以下是简要而不完整的总结：

- 许多新的 SQL 特性，包括完整的 SQL92 子查询能力（只缺目标列表子查询，其余一切都有了）。
- 支持用客户端环境变量指定时区和日期样式。
- 客户端/服务器连接的套接字接口。现在这是默认，所以你可能需要用 -i 标志启动 postmaster。
- 更好的口令授权机制。默认表权限已更改。
- 旧式时间旅行已被移除。性能得到改进。

注意

Bruce Momjian 写了以下说明来介绍新版本。

我想提一些 6.3 的普遍问题。这里只是一句话说不完的大项。关于全部特性列表，请参阅发行说明末尾的详细变更列表。

首先，我们现在有了子查询。既然有了它们，我想指出：没有子查询的 SQL 是一种非常受限的语言。子查询是绝大多数严肃的商业应用所必需的，我们终于在特性列表中勾掉了它。

我们对此特别自豪，因为这一代码里程碑意味着 PostgreSQL 的查询处理器现在完全准备就绪。注意，外部连接现在可以使用子查询语法实现。我们准备了一个脚本在 contrib 目录中来做这件事，它会把 UNION 连接重写为子查询。

其次，6.3 默认使用 Unix 域套接字而不是 TCP/IP。要启用来自其他机器的连接，必须使用新的 postmaster -i 选项，当然还要编辑 `pg_hba.conf`。此外，由于这个原因，`pg_hba.conf` 的格式已经更改。

第三，`char()` 字段现在允许比 `varchar()` 或 `text` 更快的访问。具体来说，`text` 和 `varchar()` 在访问该类型第一列之后的任何列时有性能损失。`char()` 过去也有这种访问损失，但现在没有了。这可能提示你重新设计一些表，特别是如果你有定义为 `varchar()` 或 `text` 的短字符列。这项和其他变更使 6.3 比以前的版本更快。

我们现在可以独立于任何 Unix 文件定义口令。有新的 SQL USER 命令。更多信息见管理员指南。还有一个新表 pg_shadow，用于存储用户信息和用户口令，默认只有 postgres 超级用户可以 SELECT。pg_user 现在是 pg_shadow 的视图，可被 PUBLIC SELECT。你的应用程序应继续使用 pg_user 而无需更改。

用户创建的表现在默认不再对 PUBLIC 授予 SELECT 权限。这样做是因为 ANSI 标准要求这样做。如果需要，你可以通过 GRANT 语句更改这些权限。

我们还有真正的死锁检测代码。不再有六十秒超时。新的锁定代码实现了 SQL2 事务隔离级别。

以前的版本中有很多关于文档不足的抱怨。Thomas 在许多新文档上投入了大量精力。包括新的数据库管理员指南、入门指南、程序员指南和教程。文档树的 v6.3 版本有 8.6MB，而这个版本的文档源有 7.2MB。

出于性能原因，时间旅行已被移除，但可以用触发器实现（参见 `pgsql/contrib/spi/README`）。请查看新的 \d 命令以了解类型、操作符等。此外，视图现在有自己的权限，不再基于底层表，因此必须单独设置视图的权限。查看 `/pgsql/interfaces` 了解与 PostgreSQL 交互的一些新方式。

这是第一个真正需要为现有用户做解释的版本。在许多方面，这是必要的，因为版本 6.3 为很多问题提供了最终的解决方案，这些问题以前只能通过低效的变通方法处理。
这是强烈建议的升级；它包含许多重要的新特性和对现有特性的修复。

E.251.1. 迁移到版本 6.3

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 或 pg_dumpall 进行转储/恢复。

E.251.2. 变更

```
缺陷修复
---------
修复被 MOVE 实现破坏的二进制游标 (Vadim)
修复 tcl 库崩溃 (Jan)
修复数组处理的问题，来自 Gerhard Hintermayer
修复 acl 错误并移除重复的 pqtrace (Bruce)
修复 psql \e 对空文件的问题 (Bruce)
修复 varchar() 字段上的 textcat (Bruce)
修复 DBT Sendproc (Zeugswetter Andres)
修复 vacuum analyze 语法问题 (Bruce)
```

修复国际标识符的问题 (Tatsuo)
修复继承表上的聚合 (Bruce)
修复 substr() 的越界数据问题
修复 select 1=1 or 2=2、select 1=1 and 2=2 和 select sum(2+2) (Bruce)
修复无 tty 输出以显示状态结果。-q 选项仍会关闭它 (Bruce)
修复 count(*)、带视图和多表的聚合以及 sum(3) (Bruce)
修复 cluster (Bruce)
修复 PQtrace 多次启动/停止的问题 (Bruce)
修复各种锁定问题，如较新的锁等待者先于较旧者
获得锁、写者等待锁时读者不共享锁、等待的写者
没有优先于等待的读者 (Bruce)
修复从外部文件执行查询时 psql 的崩溃 (James)
修复多列 order by 且第一列含 NULL 值的问题 (Jeroen)
为 float8 和 int4 使用正确的哈希表支持函数 (Thomas)
重新启用 CREATE OPERATOR 语句中的 JOIN= 选项 (Thomas)
更改布尔操作符的优先级以符合预期行为 (Thomas)
对过大整数生成 elog(ERROR) (Bruce)
允许约束子句中的多参数函数 (Thomas)
检查布尔输入字面量 'true'、'false'、'yes'、'no'、'1'、'0',
无法识别时抛出 elog(ERROR) (Thomas)
重要的大对象修复
修复 GROUP BY 显示重复项的问题 (Vadim)
修复 MergeJoin 中的索引扫描 (Vadim)

增强

带 EXISTS、IN、ALL、ANY 关键字的子查询 (Vadim、Bruce、Thomas)
新用户手册 (Thomas 等)
通过内联一些频繁调用的函数提速
真正的死锁检测，不再超时 (Bruce)
添加 SQL92 "常量" CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP、
CURRENT_USER (Thomas)
修改约束语法以符合 SQL92 (Thomas)
用索引实现 SQL92 PRIMARY KEY 和 UNIQUE 子句 (Thomas)
识别 FOREIGN KEY 的 SQL92 语法。抛出 elog 通知 (Thomas)
允许 NOT NULL UNIQUE 约束子句（以前各自单独允许） (Thomas)
允许非常量的 PostgreSQL 风格转换 ("::") (Thomas)
添加对 SQL3 TRUE 和 FALSE 布尔常量的支持 (Thomas)
支持 IS TRUE/IS FALSE/IS NOT TRUE/IS NOT FALSE 的 SQL92 语法 (Thomas)
允许更短的布尔字面量字符串（如 "t"、"tr"、"tru") (Thomas)
允许 SQL92 分隔标识符 (Thomas)
实现 SQL92 二进制和十六进制字符串解码 (b'10' 和 x'1F') (Thomas)
支持字面量字符串类型强制的 SQL92 语法
(如 "DATETIME 'now'") (Thomas)
添加 int2、int4 和 OID 类型与 text 之间的转换 (Thomas)
构建索引时使用共享锁 (Vadim)
事务块内为用户查询分配的内存在该查询完成后释放，
此功能在 <= 6.2.1 中被关闭 (Vadim)
新 SQL 语句 CREATE PROCEDURAL LANGUAGE (Jan)
新 PostgreSQL 过程语言 (PL) 后端接口 (Jan)
把 pg_dump -H 选项改名为 -h (Bruce)
添加 Java 对口令和欧洲日期的支持 (Peter)
为 LIKE 和 ~、!~ 操作使用索引 (Bruce)
添加 datetime 和 timespan 的哈希函数 (Thomas)

移除时间旅行 (Vadim、Bruce)
为 \d 和 \z 添加分页并修复 \i (Bruce)
向后端和前端库添加 Unix 域套接字支持 (Goran)
实现 CREATE DATABASE/WITH LOCATION 和 initlocation 工具 (Thomas)
允许更多 SQL92 和/或 PostgreSQL 保留字作为列标识符 (Thomas)
增强对 SQL92 SET TIME ZONE... 的支持 (Thomas)
SET/SHOW/RESET TIME ZONE 使用 TZ 后端环境变量 (Thomas)
实现 SET keyword = DEFAULT 和 SET TIME ZONE DEFAULT (Thomas)
启用使用 TZ 环境变量的 SET TIME ZONE (Thomas)
把 PGDATESTYLE 环境变量添加到前端和后端初始化 (Thomas)
添加 PGTZ、PGCOSTHEAP、PGCOSTINDEX、PGRPLANS、PGGEQO
前端库初始化环境变量 (Thomas)
回归测试时区用 "setenv PGTZ PST8PDT" 自动设置 (Thomas)
添加 pg_description 表以保存表、列、操作符、类型和
聚合的信息 (Bruce)
把系统表/索引名 16 字符的限制增加到 32 字符 (Bruce)
重命名系统索引 (Bruce)
向 SET DATESTYLE 添加 'GERMAN' 选项 (Thomas)
定义带 "hh:mm:ss" 字段的 "ISO 风格" timespan 输出格式 (Thomas)
允许 delta time 的小数值 (如 '2.5 days') (Thomas)
更仔细地验证 delta time 的数字输入 (Thomas)
实现把年积日作为 date_part() 的可能输入 (Thomas)
定义 timespan_finite() 和 text_timespan() 函数 (Thomas)
移除归档内容 (Bruce)
允许有独立于系统口令文件的 pg_password 认证数据库 (Todd)
转储 ACL、GRANT、REVOKE 权限 (Matt)
定义 text、varchar 和 bpchar 字符串长度函数 (Thomas)
修复继承的查询处理和代价计算 (Bruce)
实现 CREATE TABLE/AS SELECT (SELECT/INTO 的替代) (Thomas)
允许约束中的 NOT、IS NULL、IS NOT NULL (Thomas)
实现 SELECT 的 UNION (Bruce)
向 INSERT 添加 UNION、GROUP、DISTINCT (Bruce)
varchar() 在磁盘上只存储必要的字节 (Bruce)
修复 BLOB 的问题 (Peter)
JDBC 超大补丁……变更列表见 README_6.3 (Peter)
从 PQconnectdb() 移除未使用的 "option"
新 LOCK 命令和描述死锁的锁手册页 (Bruce)
添加新 psql \da、\dd、\df、\do、\dS 和 \dT 命令 (Bruce)
增强 psql \z 以显示序列 (Bruce)
在 psql \d 表中显示 NOT NULL 和 DEFAULT (Bruce)
新 psql .psqlrc 文件启动 (Andrew)
修改 contrib/linux 中的示例启动脚本以显示 syslog (Thomas)
contrib/ip_and_mac 中 IP 和 MAC 地址的新类型 (TomH)
contrib/unixdate 中与日期/时间类型的 Unix 系统时间转换 (Thomas)
contrib 内容的更新 (Massimo)
向 DBD::Pg 添加 Unix 套接字支持 (Goran)
新 python 接口 (PyGreSQL 2.0) (D'Arcy)
新前端/后端协议有版本号和网络字节序 (Phil)
pg_hba.conf 的安全特性得到增强和记录，大量清理 (Phil)
CHAR() 现在比 VARCHAR() 或 TEXT 访问更快
ecpg 嵌入式 SQL 预处理器
降低系统列开销 (Vadmin)
移除 pg_time 表 (Vadim)
添加 pg_type 属性以标识需要长度的类型 (bpchar、varchar)

COPY 命令失败时添加对问题行的报告
允许视图权限与底层表分开设置。
为安全起见，酌情对视图使用 GRANT/REVOKE (Jan)
表现在没有默认的 GRANT SELECT TO PUBLIC。你必须
显式授予此类权限。
清理教程示例 (Darren)

源代码树变更

添加新的 html 开发工具和 /tools/backend 中的流程图
修复 SCO 编译
Stratus 计算机移植 (Robert Gillies)
为 BSD44_derived & i386_solaris 添加 shlib 支持
使 configure 更加自动化 (Brook)
添加检查回归测试结果的脚本
把解析器函数拆分为更小的文件并分组 (Bruce)
把 heap_create 改名为 heap_create_and_catalog，把 heap_creatr
改名为 heap_create() (Bruce)
Sparc/Linux 的锁定补丁 (TomS)
移除 PORTNAME 并重组移植特定内容 (Marc)
添加优化器 README 文件 (Bruce)
移除优化器中的一些递归并清理其中一些代码 (Bruce)
修复 NetBSD 锁定 (Henry)
修复 libptcl 的制作 (Tatsuo)
AIX 补丁 (Darren)
把 IS TRUE、IS FALSE、... 更改为使用 "=" 的表达式而不是
对 istrue() 或 isfalse() 的函数调用以允许优化 (Thomas)
各种 NetBSD/Sparc 相关修复 (TomH)
Alpha linux 锁定 (Travis、Ryan)
把 elog(WARN) 改为 elog(ERROR) (Bruce)
FreeBSD 的 FAQ (Marc)
把 PostODBC 源代码树纳入我们的标准发行版 (Marc)
HP/UX 10 与 9 的小补丁 (Stan)
新 pg_attribute.atttypmod 用于 varchar 长度等类型特定信息 (Bruce)
UnixWare 补丁 (Billy)
自旋锁 asm 的新 i386 'lock' (Billy)
移除对多路复用后端的支持
开始 OpenBSD 移植
开始 AUX 移植
开始 Cygnus 移植
向回归套件添加字符串函数 (Thomas)
扩展一些以前截断为 16 字符的函数名 (Thomas)
移除不需要的 malloc() 调用并用 palloc() 替换 (Bruce)

E.252. 版本 6.2.1

发行日期
1997-10-17

6.2.1 是 6.2 之上的缺陷修复和易用性版本。

摘要：

- 按照 SQL92，允许字符串跨行。
- 包含在表更新时插入用户名的示例触发器函数。

这是 6.2 之上的次要缺陷修复版本。从 6.2 之前的系统升级需要完整的转储/重载。请参阅 6.2 版本的发行说明了解操作说明。

E.252.1. 从版本 6.2 迁移到版本 6.2.1

这是次要缺陷修复版本。从 6.2 版本不需要转储/重载，但从 6.2 之前的任何版本需要。

从 6.2 版本升级时，如果你选择转储/重载，你会发现 avg(money) 现在计算正确。所有其他缺陷修复在更新可执行文件后生效。

避免转储/重载的另一种方法是从 `psql` 使用以下 SQL 命令更新现有的系统表：

```
update pg_aggregate set aggfinalfn = 'cash_div_flt8'
 where aggname = 'avg' and aggbasetype = 790;
```

这需要对每个现有数据库执行，包括 template1。

E.252.2. 变更

允许 `TIME` 和 `TYPE` 作为列名 (`Thomas`)
允许更大范围的 `true/false` 布尔值 (`Thomas`)
支持 `"now"` 和 `"current"` 的输出 (`Thomas`)
正确处理 `INSERT NULL` 与 `DEFAULT` (`Vadim`)
修复缓冲区管理器中关系引用计数的问题 (`Vadim`)
像 `ANSI` 一样允许字符串跨行 (`Thomas`)
修复带 `ORDER BY` 的反向游标 (`Vadim`)
修复 `avg(cash)` 计算 (`Thomas`)
修复 `ORDER/GROUP BY` 中重复指定列的问题 (`Vadim`)
记录返回受影响行数的新 `libpq` 函数 `PQcmdTuples` (`Bruce`)
用于在 `INSERT/UPDATE` 时插入用户名的触发器函数 (`Brook Milligan`)

E.253. 版本 6.2

发行日期
1997-10-02

希望从以前的 PostgreSQL 版本迁移数据的用户，需要进行转储/恢复。

E.253.1. 从版本 6.1 迁移到版本 6.2

此迁移需要完整转储 6.1 数据库并在 6.2 中恢复。

注意，应使用 6.2 的 `pg_dump` 和 `pg_dumpall` 工具来转储 6.1 数据库。如果使用 6.2 之前版本的这些工具，把 template0 复制到 template1 数据库的步骤会失败。

E.253.2. Migration from version 1.*x* to version 6.2

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.253.3. 变更

```
缺陷修复
---------
修复 pg_dump 对继承、序列、归档表的问题 (Bruce)
修复 Solaris 上因移位、无符号和坏原型
        导致的溢出编译错误 (Diab Jerius)
修复几何线段算术中的缺陷（错误的交点计算） (Thomas)
在端点处检查几何相交以避免舍入难看 (Thomas)
捕获无效的删除尝试 (Vadim)
把时间函数名改得更一致 (Michael Reifenberg)
检查除零 (Michael Reifenberg)
修复使命令更改/插入的行对命令本身可见的很久以来的缺陷
        （因此曾出现已更新行的多次更新等） (Vadim)
修复 SELECT null, 'fail' FROM pg_am (Patrick)
现在允许 SELECT NULL as EMPTY_FIELD (Patrick)
从 contrib/pginterface 移除不需要的信号代码
修复 OR (where x != 1 or x isnull 不返回 x 为 NULL 的行) (Vadim)
修复 time_cmp 函数 (Vadim)
修复 WHERE 子句中第一个参数非属性的
       函数的处理 (Vadim)
修复条目顺序与目标列表顺序不同时的
       GROUP BY (Vadim)
修复 pg_dump 对无 sfunc1 聚合的处理 (Vadim)

增强
------------
默认遗传优化器 GEQO 参数现在是 8 (Bruce)
允许在目标列表中使用函数内带聚合的参数 (Vadim)
添加 JDBC 驱动作为接口 (Adrian & Peter)
pg_password 工具
返回 INSERT/UPDATE/DELETE 等插入/影响的行数 (Vadim)
用 CREATE TRIGGER 实现触发器 (SQL3) (Vadim)
SPI (服务器编程接口) 允许在 C 函数内执行查询 (Vadim)
实现 NOT NULL (SQL92) (Robson Paniago de Miranda)
包含用于字符串处理、外连接和并集的保留字 (Thomas)
用独占状态实现扩展注释 ("/* ... */") (Thomas)
添加 "//" 单行注释 (Bruce)
移除操作符名中字符的一些限制 (Thomas)
实现表的 DEFAULT 和 CONSTRAINT (SQL92) (Vadim & Thomas)
添加文本连接操作符和函数 (SQL92) (Thomas)
支持 WITH TIME ZONE 语法 (SQL92) (Thomas)
支持 INTERVAL unit TO unit 语法 (SQL92) (Thomas)
定义类型 DOUBLE PRECISION、INTERVAL、CHARACTER
       和 CHARACTER VARYING (SQL92) (Thomas)
定义类型 FLOAT(p) 和基础的 DECIMAL(p,s)、NUMERIC(p,s) (SQL92) (Thomas)
定义 EXTRACT()、POSITION()、SUBSTRING() 和 TRIM() (SQL92) (Thomas)
定义 CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP (SQL92) (Thomas)
为 UNION、HAVING、INNER 和 OUTER JOIN 添加语法和警告 (SQL92) (Thomas)
```

添加更多保留字，主要为符合 SQL92 (Thomas)
允许 timespan/reltime 类型的 hh:mm:ss 时间输入 (Thomas)
为 lseg、path、polygon 添加 center() 例程 (Thomas)
为 circle-polygon、polygon-polygon 添加 distance() 例程 (Thomas)
用轴交叉算法显式检查多边形内包含的点和多边形 (Thomas)
添加 circle-box 转换例程 (Thomas)
合并不同几何数据类型的冲突操作符 (Thomas)
把距离操作符 "<===>" 替换为 "<->" (Thomas)
把"上方"操作符 "!^" 替换为 ">^", "下方"操作符 "!|" 替换为 "<^" (Thomas)
添加两端文本修剪、子串和字符串位置例程 (Thomas)
添加转换例程 circle(box) 和 poly(circle) (Thomas)
允许内部排序存储在内存中而不是文件中 (Bruce & Vadim)
允许内部相同类型上的函数和操作符成功 (Bruce)
通过性能分析加快后端启动 (Bruce)
内联频繁调用的函数以提升性能 (Bruce)
减少 open() 调用 (Bruce)
psql: 为 \h 和 \? 添加 PAGER, \C 修复
修复无 tty 时 psql 分页器的问题 (Bruce)
新 entab 工具 (Bruce)
用于参照完整性的一般触发器函数 (Vadim)
用于时间旅行的一般触发器函数 (Vadim)
用于 AUTOINCREMENT/IDENTITY 特性的一般触发器函数 (Vadim)
MOVE 实现 (Vadim)

源代码树变更

HP-UX 10 补丁 (Vladimir Turin)
添加 SCO 支持 (Daniel Harris)
MkLinux 补丁 (Tatsuo Ishii)
把几何 box 术语从 "length" 改为 "width" (Thomas)
废弃几何代码中的临时未存储斜率字段 (Thomas)
从 INSTALL 移除重启说明 (Bruce)
先在 /usr/ucb 中查找 install (Bruce)
修复 c++ 复制示例代码 (Thomas)
在 psql 手册页添加 -o (Bruce)
防止 relname 未分配字符串长度被复制进数据库 (Bruce)
清理 NAMEDATALEN 的使用 (Bruce)
修复输出中超过 15 字符的 pg_proc 名 (Bruce)
添加 strNcpy() 函数 (Bruce)
移除一些不必要的 (void) 转换 (Bruce)
新的 interfaces 目录 (Marc)
用对 fd.c 函数的调用替换 fopen() 调用 (Bruce)
尽可能把函数设为 static (Bruce)
把未使用的函数包在 #ifdef NOT_USED 中 (Bruce)
移除时间戳支持中对 difftime() 的调用以修复 SunOS (Bruce & Thomas)
针对 Digital Unix 的变更
pg_dumpall 的可移植性修复 (Bruce)
把 pg_attribute.attnvals 改名为 attdispersion (Bruce)
"intro/unix" 手册页现为 "pgintro" (Bruce)
"built-in" 手册页现为 "pgbuiltin" (Bruce)
"drop" 手册页现为 "drop_table" (Bruce)
添加 "create_trigger"、"drop_trigger" 手册页 (Thomas)
添加约束回归测试 (Vadim & Thomas)
添加注释语法回归测试 (Thomas)

添加 PGINDENT 和支持程序 (Bruce)
对所有 *.c 和 *.h 文件运行 PGINDENT 的大规模提交 (Bruce)
文件移到 /src/tools 目录 (Bruce)
SPI 和触发器编程指南 (Vadim & D'Arcy)

E.254. 版本 6.1.1

发行日期
1997-07-22

E.254.1. 从版本 6.1 迁移到版本 6.1.1

这是次要缺陷修复版本。从 6.1 版本不需要转储/重载，但从 6.1 之前的任何版本需要。更多细节请参阅 6.1 的发行说明。

E.254.2. 变更

修复带选项的 SET (Thomas)
允许 pg_dump/pg_dumpall 保留所有表/对象的所有权 (Bruce)
新 psql \connect 选项允许只更改用户名而不更改数据库
修复 initdb --debug 选项 (Yoshihiko Ichikawa))
lextest 清理 (Bruce)
哈希修复 (Vadim)
修复日期/时间的月份边界算术 (Thomas)
修复某些移植的时区夏令时处理 (Thomas、Bruce、Tatsuo)
timestamp 彻底改造为使用标准函数 (Thomas)
日期/时间例程的其他代码清理 (Thomas)
psql 的 \d 现在不区分大小写 (Bruce)
psql 的反斜杠命令现在可以带尾部分号 (Bruce)
修复使用 \g 时 psql 的内存泄漏 (Bruce)
与服务器通信的字节序处理重大修复 (Thomas、Tatsuo)
Solaris 汇编器和头文件的修复 (Yoshihiko Ichikawa)
允许用户名中使用下划线 (Bruce)
pg_dumpall 现在返回正确状态，可移植性修复 (Bruce)

E.255. 版本 6.1

发行日期
1997-06-08

回归测试已经为 PostgreSQL 6.1 版本做了适配和大量修改。

新增了三种数据类型（datetime、timespan 和 circle），已加入到 PostgreSQL 本地类型集中。点、框、路径和多边形在各数据类型间的输出格式已统一。misc.out 中的多边形输出只相对原始回归输出做了抽查。

PostgreSQL 6.1 引入了使用遗传算法的新替代优化器。当查询包含多个限定条件或多个表时（优化器可选择求值顺序），这些算法会给查询结果的顺序引入随机行为。
已有几个回归测试被修改为显式排序结果，因此对优化器的选择不敏感。
少数回归测试针对本质上无序的数据类型（如点和时间间隔），涉及这些类型的测试显式地用 `set geqo to 'off'` 和 `reset geqo` 括起来。

数组说明符（原子值周围的花括号）
的解释似乎在最初的回归测试生成之后的某个时候发生了变化。当前的 `./expected/*.out` 文件反映了这一新解释，它可能并不正确！

float8 回归测试至少在某些平台上失败。这是由于 `pow()` 和 `exp()` 实现的差异以及上溢/下溢条件所用的信号机制不同，不是 PostgreSQL 的缺陷。

random 测试中的“random”结果会使“random”测试被标记为“failed”，因为回归测试是用简单的 diff 评估的。不过，“random”在我的测试机（Linux/gcc/i686）上似乎不会产生随机结果。仔细检查输出会发现输出是有效的。

E.255.1. 迁移到版本 6.1

此迁移需要完整转储 6.0 数据库并在 6.1 中恢复。

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.255.2. 变更

```
缺陷修复
---------
库例程中的包长度检查
锁管理器优先级补丁
检查 float8 的下溢/上溢 (Bruce)
多表连接修复 (Vadim)
SIGPIPE 崩溃修复 (Darren)
大对象修复 (Sven)
允许 btree 索引处理 NULL (Vadim)
时区修复 (D'Arcy)
无行时 select SUM(x) 可返回 NULL (Thomas)
内部优化器、执行器缺陷修复 (Vadim)
修复 < 或 <= 内层循环无行的问题 (Vadim)
防止连接索引子句重新交换 (Vadim)
修复多表的连接子句 (Vadim)
修复数组的哈希和哈希连接 (Vadim)
修复 abstime 类型的 btree (Vadim)
大对象修复 (Raymond)
修复哈希索引中的缓冲区泄漏 (Vadim)
修复 rtree 以用于内层扫描 (Vadim)
修复 gist 以用于内层扫描并清理 (Vadim、Andrea)
避免不必要的本地缓冲区分配 (Vadim、Massimo)
修复事务中止时本地缓冲区泄漏 (Vadim)
修复文件管理器内存泄漏并清理 (Vadim、Massimo)
修复存储管理器内存泄漏 (Vadim)
修复 btree 重复项处理 (Vadim)
修复 vacuum 导致已删行复活的问题 (Vadim)
修复 SELECT varchar()/char() INTO TABLE 生成零长度字段的问题 (Bruce)
用 Purify 修复许多 psql、pg_dump 和 libpq 内存泄漏 (Igor)
```

增强

属性优化统计 (Bruce)
快得多的新 btree 批量装载代码 (Paul)
批量装载代码加入 BTREE UNIQUE (Vadim)
新锁调试代码 (Massimo)
libpg++ 的大规模变更 (Leo)
新 GEQO 优化器加速多表优化 (Martin)
对唯一键非唯一插入的新 WARN 消息 (Marc)
update x=-3 (无空格) 现在有效 (Bruce)
移除标识符的大小写敏感处理 (Bruce、Thomas、Dan)
调试后端现在美化打印树 (Darren)
新 Oracle 字符函数 (Edmund)
新明文口令函数 (Dan)
"no such class or insufficient privilege" 拆分为不同的消息 (Dan)
新 ANSI timestamp 函数 (Dan)
新 ANSI Time 和 Date 类型 (Thomas)
在后端中移动大数据块 (Martin)
多列 btree 索引 (Vadim)
新 SET var TO value 命令 (Martin)
读取时更新事务状态 (Dan)
字符类型的新区域设置 (Oleg)
新 SEQUENCE 序列号生成器 (Vadim)
现在可以使用 GROUP BY 函数 (Vadim)
重新组织回归测试 (Thomas、Marc)
新优化器操作权重 (Vadim)
新 psql \z 授予/许可选项 (Marc)
新 MONEY 数据类型 (D'Arcy、Thomas)
tcp 套接字通信速度改进 (Vadim)
VACUUM 的属性统计及特定列新选项 (Vadim)
许多几何类型改进 (Thomas、Keith)
附加回归测试 (Thomas)
新 datestyle 变量 (Thomas、Vadim、Martin)
更多用于排序的比较操作符 (Thomas)
新转换函数 (Thomas)
新更紧凑的 btree 格式 (Vadim)
允许 pg_dumpall 保留数据库所有权 (Bruce)
新 SET GEQO=# 和 R_PLANS 变量 (Vadim)
旧 (!GEQO) 优化器可使用右侧计划 (Vadim)
SQL 解析器类型检查改进 (Bruce)
新 SET、SHOW、RESET 命令 (Thomas、Vadim)
新 \connect database USER 选项
新 destroydb -i 选项 (Igor)
新 \dt 和 \di psql 命令 (Darren)
SELECT "\n" 现在转义换行 (A. Duursma)
从旧格式转换的新几何转换函数 (Thomas)

源代码树变更

新配置脚本 (Marc)
添加 readline 配置选项 (Marc)
移除操作系统特定的配置选项 (Marc)
新的操作系统特定模板文件 (Marc)

不再需要编辑 Makefile.global (Marc)
重新安排头文件 (Marc)
nextstep 补丁 (Gregor Hoffleit)
移除 Windows 特定代码 (Bruce)
移除 postmaster -e 选项，现在只有 postgres -e 选项 (Bruce)
合并前后端中重复的库代码 (Martin)
现在可与 eBones、国际 Kerberos 一起工作 (Jun)
更多共享库支持
c++ 头文件清理 (Bruce)
对有缺陷的 flex 发出警告 (Bruce)
DG/UX、Ultrix、IRIX、AIX 可移植性修复

E.256. 版本 6.0

发行日期
1997-01-29

希望从以前的 PostgreSQL 版本迁移数据的用户，需要进行转储/恢复。

E.256.1. 从版本 1.09 迁移到版本 6.0

此迁移需要完整转储 1.09 数据库并在 6.0 中恢复。

E.256.2. 从 1.09 之前版本迁移到版本 6.0

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.256.3. 变更

缺陷修复

ALTER TABLE 缺陷—运行中的 postgres 进程需要重新读取表定义
允许对一个表或整个数据库运行 vacuum (Bruce)
数组修复
修复数组内存写的越界 (Kurt)
修复难以捉摸的 btree 范围/非范围缺陷 (Dan)
修复某些类型（如 time 和 date）上的哈希索引
修复 pg_log 大小爆炸的问题
修复 lo_export() 的权限 (Bruce)
修复未初始化的内存读取 (Kurt)
修复 ALTER TABLE ... char(3) 缺陷 (Bruce)
修复了几个小的内存泄漏
修复 EXPLAIN 的选项处理并更改了 full_path 选项名
修复组 acl 权限的输出
内存泄漏（用 Purify 等工具查找并消灭）(Kurt)
规则系统的小改进
NOTIFY 修复
新的运行检查断言
彻底改造 parser/analyze 代码以正确报告错误并提高速度

Pg_dump -d 现在正确处理 NULL (Bruce)
防止 SELECT NULL 使服务器崩溃 (Bruce)
INSERT ... SELECT 列不匹配时正确报告错误
插入列名不正确时正确报告错误
psql \g filename 现在可用 (Bruce)
修复 psql 一行多语句多输出的问题
移除重复的系统 OID
SELECT * INTO TABLE . GROUP/ORDER BY 在表已存在时给出 unlink 错误 (Bruce)
几项对使后端崩溃查询的修复
插入字符串中起始引号的错误 (Bruce)
提交空查询现在返回空状态，而不只是 " " 查询 (Bruce)

增强

添加 EXPLAIN 手册页 (Bruce)
添加 UNIQUE 索引能力 (Dan)
添加主机名/用户级访问控制，而不只是主机名和用户
为 <> 添加 != 同义词 (Bruce)
允许 "select oid,* from table"
允许 BY、ORDER BY 按编号或按非别名的 table.column 指定列 (Bruce)
允许从前端 COPY (Bryan)
允许 GROUP BY 使用别名列名 (Bruce)
允许真正的压缩，而不只是同一页上的重用 (Vadim)
允许安装配置选项自动添加所有本地用户 (Bryan)
允许 libpq 区分文本值 '' 和 null (Bruce)
允许有 createdb 权限的非 postgres 用户 destroydb
允许限制谁可以创建 C 函数 (Bryan)
允许限制谁可以做后端 COPY (Bryan)
可以收缩表、pg_time 和 pg_log (Vadim & Erich)
更改调试级别 2 为只打印查询，更改调试标题布局 (Bruce)
把十进制常量的默认表示从 float4 改为 float8 (Bruce)
postmaster 启动时现在设置欧洲日期格式
找不到精确大小写时执行小写函数名
聚合/GROUP 处理的修复，允许 'select sum(func(x),sum(x+y) from z'
Gist 现在包含在发行版中 (Marc)
本地用户的 Ident 认证 (Bryan)
实现 BETWEEN 限定符 (Bruce)
实现 IN 限定符 (Bruce)
libpq 有 PQgetisnull() (Bruce)
libpq++ 改进
initdb 的新选项 (Bryan)
Pg_dump 允许转储 OID (Bruce)
Pg_dump 为提高速度在表装载后创建索引 (Bruce)
Pg_dumpall 转储所有数据库和用户表
Pginterface 对 NULL 值的补充 (Bruce)
防止以 root 运行 postmaster
psql \h 和 \? 现在可读 (Bruce)
psql 允许行内任意位置的反斜杠和分号 (Bruce)
psql 更改查询中或引号中行的命令提示符 (Bruce)
psql char(3) 现在在 \d 输出中显示为 (bp)char (Bruce)
psql 返回码现在更准确 (Bryan?)
psql 更新的帮助语法 (Bruce)
重新检查并修复 vacuum (Vadim)
减小回归 diff 的大小，移除时区名差异 (Bruce)

移除编译期参数以支持二进制发行 (Bryan)
反转 HBA 掩码的含义 (Bryan)
本地用户的安全认证 (Bryan)
加快 vacuum (Vadim)
Vacuum 现在有 VERBOSE 选项 (Bruce)

源代码树变更

所有函数现在都有与调用比较的原型
允许从 Makefile.global 轻松禁用断言 (Bruce)
把代码中使用的 oid 常量改为 #define 名称
解耦 sparc 和 solaris 定义 (Kurt)
Gcc -Wall 干净编译，警告只来自无法修复的构造
主要的头文件重组/精简 (Marc)
编译失败时 make 现在停止 (Bryan)
Makefile 重构 (Bryan、Marc)
把 bsdi_2_1 合并到 bsdi (Bruce)
移除 Monitor 程序
从 Postgres95 改名为 PostgreSQL
新 config.h 文件 (Marc、Bryan)
PG_VERSION 现在设为 6.0 并被 postmaster 使用
可移植性补充，包括 Ultrix、DG/UX、AIX 和 Solaris
减少 #define 的数量，集中化 #define
移除系统表中的重复 OID (Dan)
移除重复的系统目录信息或报告不匹配 (Dan)
移除了许多操作系统特定的 #define
重构目标文件的生成/位置 (Bryan、Marc)
重构移植特定文件的位置 (Bryan、Marc)
纠正未使用/未初始化的变量

E.257. 版本 1.09

发行日期
1996-11-04

抱歉，我们没有记录 1.02 到 1.09 的变更。6.0 中列出的一些变更实际上包含在 1.09 版本中。

E.258. 版本 1.02

发行日期
1996-08-01

E.258.1. 从版本 1.02 迁移到版本 1.02.1

这是 1.02.1 的新迁移文件。它包括 'copy' 变更和一个把旧 ASCII 文件转换为新格式的脚本。

> **注意**
>
> 以下说明是为了帮助希望把数据库从 Postgres95 1.01 和 1.02 迁移到 Postgres95 1.02.1 的用户。
>
> 如果你全新开始使用 Postgres95 1.02.1 而不需要迁移旧数据库，则无需执行任何操作。

为了把较旧的 Postgres95 1.01 或 1.02 版本数据库升级到 1.02.1 版本，需要以下步骤：

1. 启动新的 1.02.1 postmaster

2. 把 1.02.1 的新内建函数和操作符添加到 1.01 或 1.02 数据库。方法是用新的 1.02.1 服务器运行你自己的 1.01 或 1.02 数据库，并应用文件末尾随附的查询。这可以通过 `psql` 轻松完成。如果你的 1.01 或 1.02 数据库命名为 `testdb`，并且你已从本文件末尾剪下命令并保存到 `addfunc.sql`：

   ```
   % psql testdb -f addfunc.sql
   ```

   升级 1.02 数据库的用户在执行文件中的最后两条语句时会得到一个警告，因为它们已存在于 1.02 中。这不必担心。

E.258.2. 转储/重载程序

如果你试图重新装载用以前版本生成的 pg_dump 或文本模式 `copy tablename to stdout` 输出，需要先对 ASCII 文件运行随附的 `sed` 脚本再装入数据库。旧格式用 '.' 作为数据结束，而 '\.' 现在是数据结束标记。此外，空字符串现在装填为 '' 而不是 NULL。完整细节参见 copy 手册页。

```
sed 's/^\.$/\\./g' <in_file >out_file
```

如果你在装载较旧的二进制备份或非 stdout 副本，则没有数据结束字符，因此无需转换。

```
-- following lines added by agc to reflect the case-insensitive
-- regexp searching for varchar (in 1.02), and bpchar (in 1.02.1)
create operator ~* (leftarg = bpchar, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = bpchar, rightarg = text, procedure =
 texticregexne);
create operator ~* (leftarg = varchar, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = varchar, rightarg = text, procedure =
 texticregexne);
```

E.258.3. 变更

```
源代码维护和开发
  * 全球志愿者团队
  * 源代码树现在在 ftp.ki.net 的 CVS 中

增强
  * psql（及底层 libpq 库）现在有更多输出格式化选项，
    包括 HTML
```

```
 * pg_dump 现在输出模式和/或数据，并有许多提升完整性的修复。
 * 管理 shell 脚本中用 psql 代替 monitor。
   monitor 将在下个版本中废弃。
 * 日期/时间函数增强
 * NULL 插入/更新/比较修复/增强
 * TCL/TK 库和 shell 修复为同时支持 tck7.4/tk4.0 和 tcl7.5/tk4.1

缺陷修复（几乎多到无法列举）
 * 索引
 * 存储管理
 * 解引用前检查 NULL 指针
 * Makefile 修复

新移植
 * 新增 SolarisX86 移植
 * 新增 BSD/OS 2.1 移植
 * 新增 DG/UX 移植
```

E.259. 版本 1.01

发行日期
1996-02-23

E.259.1. 从版本 1.0 迁移到版本 1.01

以下说明是为了帮助希望把数据库从 Postgres95 1.0 迁移到 Postgres95 1.01 的用户。

如果你全新开始使用 Postgres95 1.01 而不需要迁移旧数据库，则无需执行任何操作。

为了在用 Postgres95 1.0 版本创建的数据库上运行 Postgres95 1.01 版本，需要以下步骤：

1. 把 `src/Makefile.global` 中 `NAMEDATALEN` 的定义改为 16，把 `OIDNAMELEN` 的定义改为 20

2. 决定是否要使用基于主机的认证。

   a. 如果要用，你必须在顶级数据目录（通常是 `$PGDATA` 的值）中创建名为 `pg_hba` 的文件。`src/libpq/pg_hba` 给出了示例语法。

   b. 如果不想使用基于主机的认证，可以注释掉这一行：

      ```
      HBA = 1
      ```

      （位于 `src/Makefile.global`）

      注意，基于主机的认证默认开启，如果你不执行上述 A 或 B 步骤，默认的 pg_hba 文件将拒绝所有到 Postgres 的连接，因此请执行上面的 A 或 B，或删除 pg_hba 文件。

3. 编译并安装 1.01，但不要执行 `initdb` 步骤。

4. 在做任何其他事情之前，终止你的 1.0 postmaster，并备份现有的 `$PGDATA` 目录。

5. 把 PGDATA 环境变量指向你的 1.0 数据库，但把路径设置为使用 1.01 二进制文件。
6. 把文件 $PGDATA/PG_VERSION 从 5.0 改为 5.1
7. 启动新的 1.01 postmaster
8. 把 1.01 的新内建函数和操作符添加到 1.0 数据库。方法是用新的 1.01 服务器运行你自己的 1.0 数据库，并应用随附保存在 1.0_to_1.01.sql 文件中的查询。这可以通过 psql 轻松完成。如果你的 1.0 数据库名为 testdb：

   ```
   % psql testdb -f 1.0_to_1.01.sql
   ```

   然后执行以下命令（从此处剪切粘贴）：

   ```
   -- add builtin functions that are new to 1.01

   create function int4eqoid (int4, oid) returns bool as 'foo'
   language 'internal';
   create function oideqint4 (oid, int4) returns bool as 'foo'
   language 'internal';
   create function char2icregexeq (char2, text) returns bool as 'foo'
   language 'internal';
   create function char2icregexne (char2, text) returns bool as 'foo'
   language 'internal';
   create function char4icregexeq (char4, text) returns bool as 'foo'
   language 'internal';
   create function char4icregexne (char4, text) returns bool as 'foo'
   language 'internal';
   create function char8icregexeq (char8, text) returns bool as 'foo'
   language 'internal';
   create function char8icregexne (char8, text) returns bool as 'foo'
   language 'internal';
   create function char16icregexeq (char16, text) returns bool as
    'foo'
   language 'internal';
   create function char16icregexne (char16, text) returns bool as
    'foo'
   language 'internal';
   create function texticregexeq (text, text) returns bool as 'foo'
   language 'internal';
   create function texticregexne (text, text) returns bool as 'foo'
   language 'internal';

   -- add builtin functions that are new to 1.01

   create operator = (leftarg = int4, rightarg = oid, procedure =
    int4eqoid);
   create operator = (leftarg = oid, rightarg = int4, procedure =
    oideqint4);
   create operator ~* (leftarg = char2, rightarg = text, procedure =
    char2icregexeq);
   create operator !~* (leftarg = char2, rightarg = text, procedure =
    char2icregexne);
   create operator ~* (leftarg = char4, rightarg = text, procedure =
    char4icregexeq);
   ```

```
create operator !~* (leftarg = char4, rightarg = text, procedure =
 char4icregexne);
create operator ~* (leftarg = char8, rightarg = text, procedure =
 char8icregexeq);
create operator !~* (leftarg = char8, rightarg = text, procedure =
 char8icregexne);
create operator ~* (leftarg = char16, rightarg = text, procedure =
 char16icregexeq);
create operator !~* (leftarg = char16, rightarg = text, procedure
 = char16icregexne);
create operator ~* (leftarg = text, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = text, rightarg = text, procedure =
 texticregexne);
```

E.259.2. 变更

不兼容变更:
* 只要用户按照 MIGRATION_from_1.0_to_1.01 文件中概述的步骤操作，1.01 与 1.0 数据库向后兼容。
  如果未采取这些步骤，1.01 与 1.0 数据库不兼容。

增强:
* 为 libpq 添加 PQdisplayTuples() 并更改 monitor 和 psql 使用它
* 新增 NeXT 移植（需要 SysVIPC 实现）
* 新增 CAST .. AS 语法
* 为 libpq 添加 PQtrace()/PQuntrace()
* 新增为表字段设置默认值的能力
* 新增了出色的 QNX 移植
* 新增了许多新的内置函数
* 在 tclsh 中支持字符串数组

E.260. 版本 1.0

发行日期
1995-09-05

E.260.1. 变更

版权变更:
* Postgres 1.0 的版权已放宽为可自由修改并可用于任何目的。
  请阅读 COPYRIGHT 文件。感谢 Michael Stonebraker 教授使之成为可能。

不兼容变更:
* 日期格式必须为 MM-DD-YYYY（或使用欧式风格时的 DD-MM-YYYY）。这遵循 SQL-92 规范。
* "double precision"（浮点数）而不是 "float8" 字面量
  （float8 现在自动提升为 double precision）
* 直接调用内部日期/时间例程（例如 abstimein）的输入现在产生错误

（内部例程继续存在，但从 SQL 解释器调用它们现在给出错误）
* 在 "create" 语句中，现在允许（且仅允许）为列加上标签

新特性：
* 为 SELECT 结果提供终端监控器 psql（附带 readline 支持）
* 为列增加唯一的行标识符，使用新的类型名 "oid"
* 服务器同时支持 SQL 和 POSTQUEL 语法
* 无重复字段名限制；可以有任意数量的同名列
* DELETE 语句可在 RSELECT 语句中使用子句
* 若启用了访问控制，则所有用户需要认证访问

E.261. Postgres95 Release 0.03

发行日期
1995-07-21

E.261.1. 变更

不兼容变更：
* BETA-0.3 与以前版本创建的数据库不兼容
（由于系统目录和索引结构的变更）。
* 双引号（"）作为字符串字面量的引号字符已废弃；
需要把它们转换为单引号（'）。
* 聚合名（如 int4sum）已按照 SQL 标准改名（如 sum）。
* CHANGE ACL 语法被 GRANT/REVOKE 语法取代。
* 浮点字面量（如 3.14）现在为 float4 类型（以前版本中是 float8）；
如果你依赖它是 float8 类型，可能需要做类型转换。如果忽略类型转换
而把浮点字面量赋给 float8 类型的字段，可能会存储不正确的值！
* LIBPQ 已彻底改造，使前端应用程序
可以连接多个后端
* pg_user 中的 usesysid 字段从 int2 改为 int4，
以允许更大范围的 Unix 用户 id。
* netbsd/freebsd/bsd 操作系统移植已合并为
单一的 BSD44_derived 移植。（感谢 Alistair Crooks）

SQL 标准符合性（以下详述使 postgres95
更符合 SQL-92 标准的变更）：
* 以下 SQL 类型现在是内建的：smallint、int(eger)、float、real、
char(N)、varchar(N)、date 和 time。

以下是对现有 postgres 类型的别名：
```
smallint -> int2
integer, int -> int4
float, real  -> float4
```
char(N) 和 varchar(N) 实现为截断的文本类型。此外，
char(N) 做空白填充。
* 单引号（'）用于引用字符串字面量；''（除
\' 外）也支持作为在字符串中插入单引号的手段
* 使用 SQL 标准聚合名（MAX、MIN、AVG、SUM、COUNT）
（此外，聚合现在可以重载，即你可以定义自己的 MAX 聚合来接受用户定义的类型。）

```
 * 移除 CHANGE ACL。添加 GRANT/REVOKE 语法。
   - 可以使用 "GROUP" 关键字向组授予权限。
       例如:
               GRANT SELECT ON foobar TO GROUP my_group;
       也支持关键字 'PUBLIC' 表示所有用户。

       一次只能向一个用户或组授予或撤销权限。

       不支持 "WITH GRANT OPTION"。只有类所有者可以更改
       访问控制
   - 默认访问控制是给用户只读访问。
     必须显式向用户授予插入/更新权限。要更改这一点，请修改
               src/backend/utils/acl.h
     中定义 ACL_WORLD_DEFAULT 的那一行

缺陷修复:
 * 修复了空表聚合不运行的缺陷。现在，
   对空表运行的聚合将返回聚合的初始条件。因此，空表的 COUNT 现在正确返回 0。
   空表的 MAX/MIN 将返回值为 NULL 的一行。
 * 允许在 monitor 中使用 \;
 * LISTEN/NOTIFY 异步通知机制现在可用
 * 规则动作体中的 NOTIFY 现在可用
 * 哈希索引可用，访问方法总体上性能应更好。
   大 btree 索引的创建应快得多。(感谢 Paul
   Aoki)

其他变更和增强:
 * 新增 EXPLAIN 语句用于解释查询执行计划
   (例如 "EXPLAIN SELECT * FROM EMP" 打印该查询的执行计划)。
 * WARN 和 NOTICE 消息不再带时间戳。要打开错误消息的时间戳，请取消注释
   src/backend/utils/elog.h 中的这一行:
       /* define ELOG_TIMESTAMPS */
 * 发生访问控制违规时，将给出消息
       "Either no such class or insufficient privilege"。
   这与找不到类时返回的消息相同。这可以打消非特权用户
   猜测特权类是否存在的念头。
 * 做了一些用户不可见的额外系统目录变更。

libpgtcl 变更:
 * 向 "pg_result" tcl 命令添加了 -oid 选项。
   pg_result -oid 返回最后插入行的 oid。如果
   最后一条命令不是 INSERT，则 pg_result -oid 返回 ""。
 * 大对象接口以 pg_lo* tcl 命令提供:
   pg_lo_open、pg_lo_close、pg_lo_creat 等。

可移植性增强和新移植:
 * flex/lex 的问题已澄清。现在，你应该可以在任何平台上使用
   flex 代替 lex。我们不再根据你使用的平台假设
   你使用什么词法分析器。
 * 现在支持 Linux-ELF 移植。已测试各种配置：已知以下配置可用:
       kernel 1.2.10, gcc 2.6.3, libc 4.7.2, flex 2.5.2, bison 1.24
   全部为 ELF 格式，

新工具:
```

```
 * ipcclean 添加到发行版中
   通常不需要运行 ipcclean，但如果你的后端崩溃
   并留下悬空的共享内存段，ipcclean 会
   帮你清理。

新文档：
 * 用户手册已修订并添加了 libpq 文档。
```

E.262. Postgres95 Release 0.02

发行日期
1995-05-25

E.262.1. 变更

```
不兼容变更：
* 创建数据库的 SQL 语句是 'CREATE DATABASE' 而不是 'CREATEDB'。
  类似地，删除数据库是 'DROP DATABASE' 而不是 'DESTROYDB'。
  但可执行文件名 'createdb' 和 'destroydb' 保持不变。

新工具：
* pgperl - Postgres95 的 Perl (4.036) 接口
* pg_dump - 转储 postgres 数据库的工具
* 启动 Postgres95 服务器的 initdb 工具

高级特性：
* 打开数据库的异步 Postgres 信号量 I/O
* 支持基于主机的访问控制文件 postmaster/pg_hba.conf
```

E.263. Postgres95 Release 0.01

发行日期
1995-05-01

初始版本。

附录 F. 额外提供的模块与扩展

本附录包含关于 PostgreSQL 发布的 `contrib` 目录中各模块的信息。其中包括移植工具、分析实用程序，以及不属于 PostgreSQL 核心系统的插件特性——之所以不纳入核心，主要是因为它们面向有限的用户群，或者过于实验性，不适合进入主源码树。但这并不妨碍它们发挥作用。

从源码发布构建时，除非构建"world"目标（见步骤 2），否则这些模块不会自动构建。你可以用以下命令构建并安装所有这些模块：

```
gmake
gmake install
```

在已完成配置的源码树的 `contrib` 目录中运行；或者，要只构建并安装某一个选定的模块，就在该模块的子目录中执行相同的操作。许多模块带有回归测试，可以通过运行以下命令执行：

```
gmake installcheck
```

前提是已有一个 PostgreSQL 服务器在运行。（注意，这里不支持 `gmake check`；你必须有一个可用的数据库服务器来执行这些测试，而且必须已经构建并安装了要测试的模块。）

如果你使用的是预打包版本的 PostgreSQL，这些组件通常会作为单独的子包提供，例如 `postgresql-contrib`。

许多模块提供新的用户定义函数、操作符或类型。安装代码后，要使用其中某个模块，就需要在数据库系统中注册新的 SQL 对象。在PostgreSQL9.1 及后续版本中，这可通过执行CREATE EXTENSION命令完成。在一个新建的数据库中，你可以直接执行：

```
CREATE EXTENSION module_name;
```

该命令必须由数据库超级用户执行。它只会在当前数据库中注册新的 SQL 对象，因此需要在每个希望使用该模块功能的数据库中运行它。另一种做法是在数据库`template1`中运行它，这样该扩展默认会被复制到随后创建的数据库中。

许多模块允许你把它们的对象安装到你选择的模式中。为此，可在 `CREATE EXTENSION` 命令中加上 `SCHEMA schema_name`。默认情况下，这些对象会被放在你当前的创建目标模式中，通常是 `public`。

如果你的数据库是从PostgreSQL9.1 之前的版本通过转储和重新载入升级而来的，并且其中一直在使用该模块的 9.1 之前版本，则应改为执行：

```
CREATE EXTENSION module_name FROM unpackaged;
```

这会将该模块在 9.1 之前版本中的对象更新为一个正规的扩展对象。今后该模块的更新将由ALTER EXTENSION管理。有关扩展更新的更多信息，请参见第 35.15 节。

F.1. adminpack

adminpack提供了一些支持函数，pgAdmin和其他管理工具可以使用它们来提供额外功能，例如远程管理服务器日志文件。所有这些函数都只允许超级用户使用。

表 F.1中所示的函数提供对服务器所在机器上文件的写访问。(See also the functions in 表 9.62, which provide read-only access.) Only files within the database cluster directory can be accessed, but either a relative or absolute path is allowable.

表 F.1. adminpack 函数

Name	Return Type	Description
pg_catalog.pg_file_write(filename text, data text, append boolean)	bigint	Write, or append to, a text file
pg_catalog.pg_file_rename(oldname text, newname text [, archivename text])	boolean	Rename a file
pg_catalog.pg_file_unlink(filename text)	boolean	Remove a file
pg_catalog.pg_logdir_ls()	setof record	List the log files in the log_directory directory

pg_file_write将指定的*data*写入由*filename*指定的文件中。如果*append*为 false，则该文件必须尚不存在。如果*append*为 true，则该文件可以已经存在，并且若已存在则会向其追加内容。返回写入的字节数。

pg_file_rename重命名文件。如果省略*archivename*或其值为空值，则它只是将*oldname*重命名为*newname*（后者必须尚不存在）。如果提供了*archivename*，它会先将*newname*重命名为*archivename*（后者必须尚不存在），然后再将*oldname*重命名为*newname*。如果第二个重命名步骤失败，它会在报告错误之前尝试将*archivename*再改回*newname*。成功时返回 true；如果源文件不存在或不可写，则返回 false；其他情况会抛出错误。

pg_file_unlink删除指定的文件。成功时返回 true；如果指定的文件不存在，或者unlink()调用失败，则返回 false；其他情况会抛出错误。

pg_logdir_ls返回log_directory目录中所有日志文件的起始时间戳和路径名。要使用此函数，log_filename参数必须保持默认设置（postgresql-%Y-%m-%d_%H%M%S.log）。

表 F.2中所示的函数已被废弃，不应在新应用中使用；instead use those shown in 表 9.56 and 表 9.62. These functions are provided in adminpack only for compatibility with old versions of pgAdmin.

表 F.2. 已弃用的adminpack函数

Name	Return Type	Description
pg_catalog.pg_file_read(filename text, offset bigint, nbytes bigint)	text	Alternate name for pg_read_file()

Name	Return Type	Description
`pg_catalog.pg_file_length(filename text)`	`bigint`	Same as `size` column returned by `pg_stat_file()`
`pg_catalog.pg_logfile_rotate()`	`integer`	Alternate name for `pg_rotate_logfile()`, but note that it returns integer 0 or 1 rather than boolean

F.2. auth_delay

`auth_delay`会使服务器在报告认证失败之前短暂暂停，
从而增加对数据库密码进行暴力破解攻击的难度。注意，它并不能阻止拒绝服务攻击，
甚至可能加剧此类攻击，因为在报告认证失败前处于等待状态的进程仍会占用连接槽位。

为了使该模块正常工作，必须在`postgresql.conf`中通过shared_preload_libraries加载它。

F.2.1. 配置参数

`auth_delay.milliseconds` (`int`)

在报告认证失败之前等待的毫秒数。默认值为 0。

为了在 `postgresql.conf` 文件中设置这些参数，你需要把 `auth_delay` 加入custom_variable_classes。典型用法可能是：

```
# postgresql.conf
shared_preload_libraries = 'auth_delay'

custom_variable_classes = 'auth_delay'
auth_delay.milliseconds = '500'
```

F.2.2. 作者

KaiGai Kohei `<kaigai@ak.jp.nec.com>`

F.3. auto_explain

auto_explain模块提供了一种自动记录慢语句执行计划的方法，而无须手工运行EXPLAIN。这对于在大型应用中追查未优化的查询尤其有帮助。

该模块不提供可从 SQL 访问的函数。要使用它，只需把它加载到服务器中。你可以把它加载到单个会话中：

```
LOAD 'auto_explain';
```

（要这样做，必须是超级用户。）更典型的用法是，在 postgresql.conf 的shared_preload_libraries中包含 auto_explain，从而把它预加载到所有会话中。这样，无论查询何时意外变慢，都可以追踪到它们。当然，这要付出额外开销。

F.3.1. 配置参数

有若干配置参数可控制auto_explain的行为。注意，其默认行为是什么都不做，因此如果希望得到任何结果，至少必须设置auto_explain.log_min_duration。

auto_explain.log_min_duration (integer)

auto_explain.log_min_duration是会导致记录该语句计划的最小语句执行时间，单位为毫秒。将其设为 0 会记录所有计划。-1（默认值）会禁用计划记录。例如，如果将其设为250ms，则所有运行 250ms 或更久的语句都会被记录。只有超级用户可以更改此设置。

auto_explain.log_analyze (boolean)

auto_explain.log_analyze使得在记录执行计划时输出EXPLAIN ANALYZE，而不只是输出EXPLAIN。该参数默认关闭。只有超级用户可以更改此设置。

> **注意**
>
> 当该参数开启时，所有已执行语句都会对每个计划节点计时，无论它们是否实际运行得足够久而最终被记录。这可能对性能造成极其负面的影响。

auto_explain.log_verbose (boolean)

auto_explain.log_verbose 使得在记录执行计划时输出EXPLAIN VERBOSE，而不只是输出EXPLAIN。该参数默认关闭。只有超级用户可以更改此设置。

auto_explain.log_buffers (boolean)

auto_explain.log_buffers 使得在记录执行计划时输出EXPLAIN (ANALYZE, BUFFERS)，而不只是输出EXPLAIN。该参数默认关闭。只有超级用户可以更改此设置。除非设置了 auto_explain.log_analyze，否则该参数没有效果。

auto_explain.log_format (enum)

auto_explain.log_format选择要使用的EXPLAIN输出格式。允许的值有text、xml、json和yaml。默认值为 text。只有超级用户可以更改此设置。

auto_explain.log_nested_statements (boolean)

auto_explain.log_nested_statements使嵌套语句（在函数内部执行的语句）也会被纳入记录考虑范围。关闭该参数时，只记录顶层查询计划。该参数默认关闭。只有超级用户可以更改此设置。

为了能在 postgresql.conf 文件中设置这些参数，你需要把 auto_explain 加入custom_variable_classes。典型用法如下：

```
# postgresql.conf
shared_preload_libraries = 'auto_explain'

custom_variable_classes = 'auto_explain'
auto_explain.log_min_duration = '3s'
```

F.3.2. 示例

```
postgres=# LOAD 'auto_explain';
postgres=# SET auto_explain.log_min_duration = 0;
postgres=# SELECT count(*)
           FROM pg_class, pg_index
           WHERE oid = indrelid AND indisunique;
```

这可能产生如下日志输出：

```
LOG:  duration: 3.651 ms  plan:
  Query Text: SELECT count(*)
              FROM pg_class, pg_index
              WHERE oid = indrelid AND indisunique;
  Aggregate  (cost=16.79..16.80 rows=1 width=0) (actual time=3.626..3.
627 rows=1 loops=1)
    ->  Hash Join  (cost=4.17..16.55 rows=92 width=0) (actual time=3.
349..3.594 rows=92 loops=1)
          Hash Cond: (pg_class.oid = pg_index.indrelid)
          ->  Seq Scan on pg_class  (cost=0.00..9.55 rows=255 width=4)
 (actual time=0.016..0.140 rows=255 loops=1)
          ->  Hash  (cost=3.02..3.02 rows=92 width=4) (actual time=3.
238..3.238 rows=92 loops=1)
                Buckets: 1024  Batches: 1  Memory Usage: 4kB
                ->  Seq Scan on pg_index  (cost=0.00..3.02 rows=92
 width=4) (actual time=0.008..3.187 rows=92 loops=1)
                      Filter: indisunique
```

F.3.3. 作者

Takahiro Itagaki <itagaki.takahiro@oss.ntt.co.jp>

F.4. btree_gin

btree_gin提供了 GIN 操作符类，为以下数据类型实现 B-树等价行为：int2、int4、int8、float4、float8、timestamp with time zone、timestamp without time zone、time with time zone、time without time zone、date、interval、oid、money、"char"、varchar、text、bytea、bit、varbit、macaddr、inet和cidr。

一般来说，这些操作符类的性能不会优于等价的标准 B-树索引方法，而且它们缺少标准 B-树代码的一项主要特性：强制唯一性的能力。不过，它们对于 GIN 测试以及作为开发其他 GIN 操作符类的基础很有用。此外，对于同时测试一个可由 GIN 建索引的列和一个可由 B-树建索引的列的查询，创建一个使用这些操作符类之一的多列 GIN 索引，可能比创建两个必须通过位图 AND 运算组合的独立索引更高效。

F.4.1. 用法示例

```
CREATE TABLE test (a int4);
-- create index
CREATE INDEX testidx ON test USING gin (a);
-- query
SELECT * FROM test WHERE a < 10;
```

F.4.2. 作者

Teodor Sigaev (<teodor@stack.net>) 和 Oleg Bartunov (<oleg@sai.msu.su>) 。参阅http://www.sai.msu.su/~megera/oddmuse/index.cgi/Gin。

F.5. btree_gist

btree_gist提供了 GiST 索引操作符类，可为以下数据类型实现与 B-树等价的行为：int2, int4, int8, float4, float8, numeric, timestamp with time zone, timestamp without time zone, time with time zone, time without time zone, date, interval, oid, money, char, varchar, text, bytea, bit, varbit, macaddr, inet, 和 cidr。

一般来说，这些操作符类的性能不会优于对应的标准 B-树索引方法，而且它们缺少标准 B-树实现的一项主要特性：强制唯一性的能力。不过，正如下文所述，它们提供了一些 B-树索引所不具备的其他特性。另外，当需要多列 GiST 索引，而其中某些列的数据类型只能用 GiST 建立索引、其他列只是简单数据类型时，这些操作符类就很有用。最后，这些操作符类对于 GiST 测试以及作为开发其他 GiST 操作符类的基础也很有用。

除典型的 B-树搜索操作符之外，btree_gist还为<>（“不等于”）提供索引支持。这在与下文描述的排他约束结合使用时可能很有用。

此外，对于那些具有自然距离度量的数据类型，btree_gist定义了距离操作符<->，并为使用该操作符的最近邻搜索提供 GiST 索引支持。为以下类型提供了距离操作符：int2、int4、int8、float4、float8、timestamp with time zone、timestamp without time zone、time without time zone、date、interval、oid和money。

F.5.1. 用法示例

一个使用btree_gist代替btree的简单示例：

```
CREATE TABLE test (a int4);
-- create index
CREATE INDEX testidx ON test USING gist (a);
-- query
SELECT * FROM test WHERE a < 10;
-- nearest-neighbor search: find the ten entries closest to "42"
SELECT *, a <-> 42 AS dist FROM test ORDER BY a <-> 42 LIMIT 10;
```

使用排他约束来强制执行这样一条规则：动物园中的一个笼子只能容纳一种动物：

```
=> CREATE TABLE zoo (
  cage   INTEGER,
  animal TEXT,
  EXCLUDE USING gist (cage WITH =, animal WITH <>)
);

=> INSERT INTO zoo VALUES(123, 'zebra');
INSERT 0 1
=> INSERT INTO zoo VALUES(123, 'zebra');
INSERT 0 1
=> INSERT INTO zoo VALUES(123, 'lion');
ERROR:  conflicting key value violates exclusion constraint "zoo_cage_
animal_excl"
DETAIL:  Key (cage, animal)=(123, lion) conflicts with existing key (
cage, animal)=(123, zebra).
=> INSERT INTO zoo VALUES(124, 'lion');
INSERT 0 1
```

F.5.2. 作者

Teodor Sigaev（<teodor@stack.net>）、Oleg Bartunov（<oleg@sai.msu.su>）和 Janko Richter（<jankorichter@yahoo.de>）。更多信息见http://www.sai.msu.su/~megera/postgres/gist/。

F.6. chkpass

此模块实现了用于存储加密密码的数据类型chkpass。
每个密码在输入时都会自动转换为加密形式，并且始终以加密形式存储。比较时，
只需与明文密码比较，比较函数会在比较之前将其加密。

代码中预留了在判定密码容易被破解时报告错误的机制。不过，
目前这只是一个不做任何事情的占位实现。

如果在输入字符串前加上冒号，它就会被视为已加密的密码，直接存储而不再加密。
这样就可以输入以前加密过的密码。

输出时会在前面加上冒号。这样就可以转储并重新载入密码，而不必重新加密。
如果需要不带冒号的加密密码，请使用raw()函数。这样就可以将此类型与 Apache 的Auth_PostgreSQL模块等工具一起使用。

加密使用标准 Unix 函数crypt()，因此也受该函数所有常见限制的影响，尤其是只考虑密码的前八个字符。

请注意，chkpass数据类型不能建立索引。

使用示例：

```
test=# create table test (p chkpass);
CREATE TABLE
test=# insert into test values ('hello');
INSERT 0 1
test=# select * from test;
       p
----------------
 :dVGkpXdOrE3ko
(1 row)

test=# select raw(p) from test;
      raw
---------------
 dVGkpXdOrE3ko
(1 row)

test=# select p = 'hello' from test;
 ?column?
----------
 t
(1 row)

test=# select p = 'goodbye' from test;
 ?column?
----------
 f
(1 row)
```

F.6.1. 作者

D'Arcy J.M. Cain（<darcy@druid.net>）

F.7. citext – 大小写不敏感的字符串类型

citext模块提供一种大小写不敏感的字符串类型 citext。本质上，它在比较值时会在内部调用 lower。除此之外，它的行为几乎与 text完全相同。

F.7.1. 原理

在PostgreSQL中进行大小写不敏感匹配的标准方法，一直是在比较值时使用lower函数，例如：

```
SELECT * FROM tab WHERE lower(col) = LOWER(?);
```

这种做法效果尚可，但有几个缺点：

- 它会让 SQL 语句变得冗长，而且你必须时刻记得同时对列和查询值调用 lower。
- 除非你创建一个使用lower的函数索引，否则它不会使用索引。
- 如果把列声明为UNIQUE或PRIMARY KEY，隐式生成的索引仍然是大小写敏感的。因此，它既无法用于大小写不敏感搜索，也不能以大小写不敏感的方式强制唯一性。

citext数据类型允许你在 SQL 查询中省去对 lower的调用，并允许主键不区分大小写。与text一样，citext也与区域设置相关，这意味着大写字符和小写字符如何匹配取决于数据库 LC_CTYPE设置的规则。同样，这种行为与在查询中使用 lower完全一致。但由于这是由数据类型透明地完成的，因此你不必在查询中额外记住任何特殊处理。

F.7.2. 如何使用

下面是一个简单的使用示例：

```
CREATE TABLE users (
    nick CITEXT PRIMARY KEY,
    pass TEXT   NOT NULL
);

INSERT INTO users VALUES ( 'larry',  md5(random()::text) );
INSERT INTO users VALUES ( 'Tom',    md5(random()::text) );
INSERT INTO users VALUES ( 'Damian', md5(random()::text) );
INSERT INTO users VALUES ( 'NEAL',   md5(random()::text) );
INSERT INTO users VALUES ( 'Bjørn',  md5(random()::text) );

SELECT * FROM users WHERE nick = 'Larry';
```

这个SELECT语句仍会返回一个元组，尽管nick列中存的是larry，而查询条件写的是Larry。

F.7.3. 字符串比较行为

citext在比较时会先把每个字符串转换为小写（就像调用了lower一样），
然后再按常规方式比较结果。因此，举例来说，如果两个字符串经过lower后得到的结果相同，它们就会被视为相等。

为了尽可能贴近大小写不敏感排序规则的行为，一些字符串处理操作符和函数都提供了 citext 专用版本。例如，当应用于citext时，正则表达式操作符~和~*
表现相同：它们都会以大小写不敏感的方式匹配。!~和 !~*也是如此，LIKE操作符 ~~和~~*、

以及 !~~和!~~*也一样。如果你希望进行大小写敏感匹配，可以把这些操作符的参数转换为text。

同样地，如果这些函数的参数是citext，它们也会以大小写不敏感方式进行匹配：

- regexp_matches()
- regexp_replace()
- regexp_split_to_array()
- regexp_split_to_table()
- replace()
- split_part()
- strpos()
- translate()

对于正则表达式函数，如果你想按大小写敏感方式匹配，可以指定“c”标志来强制大小写敏感匹配。否则，若要获得大小写敏感行为，就必须在调用这些函数之前先把值转换为text。

F.7.4. 限制

- citext的大小写折叠行为依赖于数据库的 LC_CTYPE设置，因此它如何比较值是在创建数据库时确定的。按照 Unicode 标准中的定义，它并不是真正意义上的大小写不敏感。实际上，这意味着只要你对当前排序规则满意，通常也会对 citext的比较结果满意。但是，如果数据库中存有多种语言的数据，而排序规则只适用于其中某一种语言，那么其他语言的用户可能会发现查询结果并不符合预期。
- 从PostgreSQL 9.1 起，你可以为 citext列或数据值附加COLLATE说明。当前，citext操作符在比较已完成大小写折叠的字符串时，会遵从非默认的 COLLATE说明；但最初转换为小写这一步，始终仍按数据库的 LC_CTYPE设置执行（也就是说，等同于给出了COLLATE "default"）。未来的发行版中，这一点可能会改变，从而使这两步都遵循输入的COLLATE说明。
- citext不如text高效，因为操作符函数和 B-树比较函数必须复制数据，并将数据转换为小写后才能进行比较。不过，在需要大小写不敏感匹配时，它仍然比使用lower略高效一些。
- 如果你在某些场景下需要大小写敏感比较，而在另一些场景下又需要大小写不敏感比较，那么citext并不会帮上太多忙。标准做法是使用 text类型，并在需要大小写不敏感比较时手工调用 lower；如果只是偶尔需要这类比较，这种做法是完全可行的。如果你大多数时候都需要大小写不敏感行为，而只是在少数场景下需要大小写敏感比较，那么可以考虑把数据存储为citext，并在需要大小写敏感比较时显式地把列转换为 text。无论哪种情况，如果你希望这两类搜索都足够快，就都需要建立两个索引。
- 包含citext操作符的模式必须位于当前 search_path中（通常是public）；如果不在，调用的将是普通的、大小写敏感的text操作符。

F.7.5. 作者

David E. Wheeler <david@kineticode.com>

灵感来自 Donald Fraser 最初编写的`citext`模块。

F.8. cube – 多维立方体数据类型

该模块实现了用于表示多维立方体的cube数据类型。

F.8.1. 语法

表 F.3展示了cube类型合法的外部表示。*x*、*y*等表示浮点数。

表 F.3. cube 外部表示

x	一个一维点（或者长度为零的一维区间）
(*x*)	同上
x1,*x2*,...,*xn*	n 维空间中的一个点，在内部表示为零体积立方体
(*x1*,*x2*,...,*xn*)	同上
(*x*),(*y*)	一个从*x*开始、到*y*结束的一维区间，反过来也可以；顺序无关紧要
[(*x*),(*y*)]	同上
(*x1*,...,*xn*),(*y1*,...,*yn*)	一个 n 维立方体，由一对对角相对的角点表示
[(*x1*,...,*xn*),(*y1*,...,*yn*)]	同上

立方体的两个对角点以何种顺序输入都无关紧要。必要时，cube函数会自动交换这些值，以创建统一的“左下角 – 右上角”内部表示。

空白会被忽略，因此 [(*x*),(*y*)]与 [(*x*), (*y*)] 相同。

F.8.2. 精度

值在内部以 64 位浮点数存储。这意味着有效数字超过大约 16 位的数值会被截断。

F.8.3. 用法

cube模块为cube值提供了一个 GiST 索引操作符类。GiST 操作符类支持的操作符如表 F.4所示。

表 F.4. cube GiST 操作符

操作符	描述
a = b	立方体 a 和 b 相同。
a && b	立方体 a 和 b 重叠。
a @> b	立方体 a 包含立方体 b。
a <@ b	立方体 a 包含在立方体 b 中。

（在 PostgreSQL 8.2 之前，包含操作符 @> 和 <@ 分别称为 @ 和 ~。这些名称仍然可用，但已弃用，最终将被删除。请注意，旧名称与核心几何数据类型以前采用的约定正好相反！）

也提供了标准的 B-树操作符，例如

操作符	描述
[a, b] < [c, d]	小于

操作符	描述
`[a, b] > [c, d]`	大于

这些操作符除了排序之外，在实际用途上没有太大意义。它们首先比较 (a) 与 (c)，如果相等，再比较 (b) 与 (d)。这在大多数情况下能得到相当好的排序结果，如果你希望对此类型使用 ORDER BY，这会很有用。

表 F.5展示了可用函数。

表 F.5. cube 函数

`cube(float8) returns cube`	创建一个一维立方体，其两个坐标相同。`cube(1) == '(1)'`
`cube(float8, float8) returns cube`	创建一个一维立方体。`cube(1,2) == '(1),(2)'`
`cube(float8[]) returns cube`	根据数组定义的坐标创建一个零体积立方体。`cube(ARRAY[1,2]) == '(1,2)'`
`cube(float8[], float8[]) returns cube`	创建一个立方体，其右上角和左下角坐标由这两个数组定义，这两个数组必须等长。`cube('{1,2}'::float[], '{3,4}'::float[]) == '(1,2),(3,4)'`
`cube(cube, float8) returns cube`	通过向现有立方体增加一个维度来创建新的立方体，新坐标的两部分取相同的值。这可用于根据计算得到的值逐步构建立方体。`cube('(1)',2) == '(1,2),(1,2)'`
`cube(cube, float8, float8) returns cube`	通过向现有立方体增加一个维度来创建新的立方体。这可用于根据计算得到的值逐步构建立方体。`cube('(1,2)',3,4) == '(1,3),(2,4)'`
`cube_dim(cube) returns int`	返回立方体的维数
`cube_ll_coord(cube, int) returns double`	返回立方体左下角的第 n 个坐标值
`cube_ur_coord(cube, int) returns double`	返回立方体右上角的第 n 个坐标值
`cube_is_point(cube) returns bool`	如果立方体是一个点，也就是定义它的两个角相同，则返回真。
`cube_distance(cube, cube) returns double`	返回两个立方体之间的距离。如果两个立方体都是点，这就是普通的距离函数。
`cube_subset(cube, int[]) returns cube`	根据数组中给出的维度索引列表，从现有立方体创建一个新立方体。可用于找出单个维度的左下角（LL）和右上角（UR）坐标，例如，`cube_subset(cube('(1,3,5),(6,7,8)'), ARRAY[2]) = '(3),(7)'`。也可用于删除维度，或者按需要重新排列维度，例如，`cube_subset(cube('(1,3,5),(6,7,8)'), ARRAY[3,2,1,1]) = '(5, 3, 1, 1),(8, 7, 6, 6)'`。
`cube_union(cube, cube) returns cube`	生成两个立方体的并集
`cube_inter(cube, cube) returns cube`	生成两个立方体的交集
`cube_enlarge(cube c, double r, int n) returns cube`	将立方体至少 n 个维度的大小增加指定半径。如果半径为负，则改为缩小立方体。此函数对于创建点周围的边界框、以搜索附近点很有用。所有已定义的维度都会按半径 r 改变。左下角（LL）坐标减小 r，右上角（UR）坐标增大 r。如果某个左下角坐标增大到超

	过对应的右上角坐标（这只可能在 r < 0 时发生），则将两个坐标都设为它们的平均值。如果 n 大于已定义维度数且立方体正在扩大（r >= 0），则额外坐标以 0 作为基准。

F.8.4. 默认规则

下面这个并集：

```
select cube_union('(0,5,2),(2,3,1)', '0');
cube_union
--------------------
(0, 0, 0),(2, 5, 2)
(1 row)
```

并不违背常识，下面这个交集也是如此：

```
select cube_inter('(0,-1),(1,1)', '(-2),(2)');
cube_inter
-------------
(0, 0),(1, 0)
(1 row)
```

在所有对不同维度立方体执行的二元操作中，都假定维度较低的那个是一个笛卡尔投影，也就是说，在字符串表示中省略的坐标位置上补 0。上面的例子等价于：

```
cube_union('(0,5,2),(2,3,1)','(0,0,0),(0,0,0)');
cube_inter('(0,-1),(1,1)','(-2,0),(2,0)');
```

下面的包含谓词使用的是点语法，但实际上第二个参数在内部表示为一个 box。这种语法使我们无须单独定义点类型以及用于 (box,point) 谓词的函数。

```
select cube_contains('(0,0),(1,1)', '0.5,0.5');
cube_contains
--------------
t
(1 row)
```

F.8.5. 注意

有关用法示例，请参见回归测试`sql/cube.sql`。

为了避免用户轻易把事情弄坏，立方体的维数上限被设为 100。如果需要更大的值，可在`cubedata.h`中修改该限制。

F.8.6. 致谢

原作者：Gene Selkov, Jr. `<selkovjr@mcs.anl.gov>`，阿贡国家实验室数学与计算机科学部。

我首先要感谢 Joe Hellerstein 教授（http://db.cs.berkeley.edu/jmh/），他为我阐明了 GiST (http://gist.cs.berkeley.edu/) 的要旨；也感谢他曾经的学生 Andy Dong (http://best.me.berkeley.edu/~adong/) 为 Illustra 编写了示例，http://best.berkeley.edu/~adong/rtree/index.html。我同样感谢过去和现在所有的 Postgres 开发者，他们使我得以创造自己的世界并在其中不受打扰地生活。我还要感谢阿贡实验室以及美国能源部，多年来始终如一地支持我的数据库研究。

Bruno Wolff III <bruno@wolff.to> 在 2002 年 8 月和 9 月对该软件包做了一些小更新，包括将精度从单精度改为双精度，并增加了一些新函数。

Joshua Reich <josh@root.net> 在 2006 年 7 月做了进一步更新。这些更新包括加入`cube(float8[], float8[])`，并清理代码，使其使用 V1 调用协议而不是已废弃的 V0 协议。

F.9. dblink

`dblink` 是一个支持在数据库会话内连接到其他 PostgreSQL 数据库的模块。

dblink_connect

dblink_connect — opens a persistent connection to a remote database

大纲

```
dblink_connect(text connstr) returns text
dblink_connect(text connname, text connstr) returns text
```

描述

`dblink_connect()` 建立到远程 PostgreSQL 数据库的连接。要联系的服务器和数据库通过标准的 libpq 连接字符串标识。可以为该连接指定一个名称。可同时打开多个命名连接，但同一时间只允许存在一个未命名连接。该连接会一直保留到被关闭或数据库会话结束。

连接字符串也可以是现有外部服务器的名称。定义对应的外部数据包装器时，建议使用 `postgresql_fdw_validator`。请参见下面的示例，以及：CREATE FOREIGN DATA WRAPPER, CREATE SERVER, CREATE USER MAPPING

参数

connname

用于该连接的名称；如果省略，则打开一个未命名连接，并替换现有的未命名连接。

connstr

libpq 风格的连接信息串，例如 `hostaddr=127.0.0.1 port=5432 dbname=mydb user=postgres password=mypasswd`。详见第 31.1 节中的 `PQconnectdb`。

返回值

返回状态，总是 `OK`（因为任何错误都会导致该函数抛出错误，而不是返回）。

注解

只有超级用户才能使用 `dblink_connect` 创建不使用密码认证的连接。如果非超级用户需要这种能力，请改用 `dblink_connect_u`。

选择包含等号的连接名称并不明智，因为这会导致在其他 `dblink` 函数中与连接信息字符串产生混淆风险。

示例

```
SELECT dblink_connect('dbname=postgres');
 dblink_connect
----------------
 OK
(1 row)

SELECT dblink_connect('myconn', 'dbname=postgres');
 dblink_connect
----------------
```

```
 OK
(1 row)

-- FOREIGN DATA WRAPPER functionality
-- Note: local connection must require password authentication for
 this to work properly
--       Otherwise, you will receive the following error from dblink_
connect():
--       ---------------------------------------------------------------
---------
--       ERROR:  password is required
--       DETAIL:  Non-superuser cannot connect if the server does not
 request a password.
--       HINT:  Target server's authentication method must be changed.
CREATE USER dblink_regression_test WITH PASSWORD 'secret';
CREATE FOREIGN DATA WRAPPER postgresql VALIDATOR postgresql_fdw_
validator;
CREATE SERVER fdtest FOREIGN DATA WRAPPER postgresql OPTIONS (hostaddr
 '127.0.0.1', dbname 'contrib_regression');

CREATE USER MAPPING FOR dblink_regression_test SERVER fdtest OPTIONS (
user 'dblink_regression_test', password 'secret');
GRANT USAGE ON FOREIGN SERVER fdtest TO dblink_regression_test;
GRANT SELECT ON TABLE foo TO dblink_regression_test;

\set ORIGINAL_USER :USER
\c - dblink_regression_test
SELECT dblink_connect('myconn', 'fdtest');
 dblink_connect
----------------
 OK
(1 row)

SELECT * FROM dblink('myconn','SELECT * FROM foo') AS t(a int, b text,
 c text[]);
 a  | b |       c
----+---+---------------
  0 | a | {a0,b0,c0}
  1 | b | {a1,b1,c1}
  2 | c | {a2,b2,c2}
  3 | d | {a3,b3,c3}
  4 | e | {a4,b4,c4}
  5 | f | {a5,b5,c5}
  6 | g | {a6,b6,c6}
  7 | h | {a7,b7,c7}
  8 | i | {a8,b8,c8}
  9 | j | {a9,b9,c9}
 10 | k | {a10,b10,c10}
(11 rows)

\c - :ORIGINAL_USER
REVOKE USAGE ON FOREIGN SERVER fdtest FROM dblink_regression_test;
REVOKE SELECT ON TABLE foo FROM  dblink_regression_test;
DROP USER MAPPING FOR dblink_regression_test SERVER fdtest;
```

```
DROP USER dblink_regression_test;
DROP SERVER fdtest;
DROP FOREIGN DATA WRAPPER postgresql;
```

dblink_connect_u

dblink_connect_u — opens a persistent connection to a remote database, insecurely

大纲

```
dblink_connect_u(text connstr) returns text
dblink_connect_u(text connname, text connstr) returns text
```

描述

`dblink_connect_u()` 与 `dblink_connect()` 相同，不同之处在于它允许非超级用户使用任意认证方法进行连接。

如果远程服务器选择的认证方法不涉及密码，则可能发生身份冒充以及随之而来的权限提升，因为该会话看起来会像是由运行本地 PostgreSQL 服务器的那个用户发起的。此外，即使远程服务器确实要求密码，密码也有可能来自服务器环境，例如属于服务器用户的 `~/.pgpass` 文件。这不仅带来身份冒充的风险，也可能将密码暴露给不可信的远程服务器。因此，`dblink_connect_u()` 在初始安装时会撤销 `PUBLIC` 的全部权限，从而除了超级用户之外无法调用它。在某些情况下，可能适合向被认为可信的特定用户授予 `dblink_connect_u()`的`EXECUTE` 权限，但必须谨慎操作。还建议服务器用户的任何 `~/.pgpass` 文件不要包含指定通配主机名的记录。

更多细节请参见 `dblink_connect()`。

dblink_disconnect

dblink_disconnect — closes a persistent connection to a remote database

大纲

```
dblink_disconnect() returns text
dblink_disconnect(text connname) returns text
```

描述

`dblink_disconnect()` 关闭先前由 `dblink_connect()` 打开的连接。不带参数的形式会关闭未命名连接。

参数

connname

要关闭的命名连接名称。

返回值

返回状态，总是 `OK`（因为任何错误都会导致该函数抛出错误，而不是返回）。

示例

```
SELECT dblink_disconnect();
 dblink_disconnect
-------------------
 OK
(1 row)

SELECT dblink_disconnect('myconn');
 dblink_disconnect
-------------------
 OK
(1 row)
```

dblink

dblink — executes a query in a remote database

大纲

```
dblink(text connname, text sql [, bool fail_on_error]) returns setof
 record
dblink(text connstr, text sql [, bool fail_on_error]) returns setof
 record
dblink(text sql [, bool fail_on_error]) returns setof record
```

描述

`dblink` 在远程数据库中执行查询（通常是 `SELECT`，但也可以是任意返回行的 SQL 语句）。

当提供两个 `text` 参数时，首先会将第一个参数作为持久连接名称查找；如果找到，就在该连接上执行命令。如果没有找到，则把第一个参数视为像 `dblink_connect` 那样的连接信息字符串，并且只在本命令执行期间建立所指明的连接。

参数

connname

要使用的连接名称；省略该参数则使用未命名连接。

connstr

如前面对 `dblink_connect` 所描述的连接信息字符串。

sql

要在远程数据库中执行的 SQL 查询，例如 `select * from foo`。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数不返回任何行。

返回值

该函数返回查询产生的行。由于 `dblink` 可以与任意查询一起使用，因此它被声明为返回 `record`，而不是指定某个特定的列集。这意味着必须在调用查询中指定预期的列集 – 否则 PostgreSQL 无法知道应该期待什么。以下是一个示例：

```
SELECT *
    FROM dblink('dbname=mydb', 'select proname, prosrc from pg_proc')
      AS t1(proname name, prosrc text)
    WHERE proname LIKE 'bytea%';
```

`FROM` 子句中的“别名”部分必须指定该函数将返回的列名和类型。（在别名中指定列名实际上是标准 SQL 语法，但指定列类型则是 PostgreSQL 的扩展。）这样系统才能在尝试执行该函数之前，就知道 `*` 应展开成什么，以及 `WHERE` 子句中的 `proname` 指的是什么。运行时，如果远程数据库返回的实际查询结果与 `FROM` 子句中列出的列数不同，就会抛出错误。不过，列名不必匹

配，dblink 也不要求类型完全一致。只要返回的数据字符串可以作为 FROM 子句中声明的列类型的有效输入，它就会成功。

注解

dblink 会在向本地系统返回任何结果行之前，先获取整个远程查询结果。如果查询预计会返回大量行，最好用 dblink_open 将其作为游标打开，然后每次提取可管理数量的行。

将预定义查询用于 dblink 的一种方便方式是创建视图。这样就可以把列类型信息隐藏在视图中，而不必在每个查询里都重复写出。例如，

```
CREATE VIEW myremote_pg_proc AS
  SELECT *
    FROM dblink('dbname=postgres', 'select proname, prosrc from pg_
proc')
    AS t1(proname name, prosrc text);

SELECT * FROM myremote_pg_proc WHERE proname LIKE 'bytea%';
```

示例

```
SELECT * FROM dblink('dbname=postgres', 'select proname, prosrc from
 pg_proc')
  AS t1(proname name, prosrc text) WHERE proname LIKE 'bytea%';
  proname   |   prosrc
------------+------------
 byteacat   | byteacat
 byteaeq    | byteaeq
 bytealt    | bytealt
 byteale    | byteale
 byteagt    | byteagt
 byteage    | byteage
 byteane    | byteane
 byteacmp   | byteacmp
 bytealike  | bytealike
 byteanlike | byteanlike
 byteain    | byteain
 byteaout   | byteaout
(12 rows)

SELECT dblink_connect('dbname=postgres');
 dblink_connect
----------------
 OK
(1 row)

SELECT * FROM dblink('select proname, prosrc from pg_proc')
  AS t1(proname name, prosrc text) WHERE proname LIKE 'bytea%';
  proname   |   prosrc
------------+------------
 byteacat   | byteacat
 byteaeq    | byteaeq
 bytealt    | bytealt
 byteale    | byteale
```

```
 byteagt    | byteagt
 byteage    | byteage
 byteane    | byteane
 byteacmp   | byteacmp
 bytealike  | bytealike
 byteanlike | byteanlike
 byteain    | byteain
 byteaout   | byteaout
(12 rows)

SELECT dblink_connect('myconn', 'dbname=regression');
 dblink_connect
----------------
 OK
(1 row)

SELECT * FROM dblink('myconn', 'select proname, prosrc from pg_proc')
  AS t1(proname name, prosrc text) WHERE proname LIKE 'bytea%';
  proname   |   prosrc
------------+------------
 bytearecv  | bytearecv
 byteasend  | byteasend
 byteale    | byteale
 byteagt    | byteagt
 byteage    | byteage
 byteane    | byteane
 byteacmp   | byteacmp
 bytealike  | bytealike
 byteanlike | byteanlike
 byteacat   | byteacat
 byteaeq    | byteaeq
 bytealt    | bytealt
 byteain    | byteain
 byteaout   | byteaout
(14 rows)
```

dblink_exec

dblink_exec — executes a command in a remote database

大纲

```
dblink_exec(text connname, text sql [, bool fail_on_error]) returns
 text
dblink_exec(text connstr, text sql [, bool fail_on_error]) returns
 text
dblink_exec(text sql [, bool fail_on_error]) returns text
```

描述

`dblink_exec` 在远程数据库中执行命令（即任何不返回行的 SQL 语句）。

当提供两个 `text` 参数时，首先会将第一个参数作为持久连接名称查找；如果找到，就在该连接上执行命令。如果没有找到，则把第一个参数视为像 `dblink_connect` 那样的连接信息字符串，并且只在本命令执行期间建立所指明的连接。

参数

connname

要使用的连接名称；省略该参数则使用未命名连接。

connstr

如前面对 `dblink_connect` 所描述的连接信息字符串。

sql

要在远程数据库中执行的 SQL 命令，例如 `insert into foo values(0, 'a', '{"a0", "b0","c0"}')`。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数的返回值会被设为 `ERROR`。

返回值

返回状态，可以是命令的状态字符串，也可以是 `ERROR`。

示例

```
SELECT dblink_connect('dbname=dblink_test_standby');
 dblink_connect
----------------
 OK
(1 row)

SELECT dblink_exec('insert into foo values(21,''z'',''{"a0","b0","c0"}
'');');
```

```
   dblink_exec
-----------------
 INSERT 943366 1
(1 row)

SELECT dblink_connect('myconn', 'dbname=regression');
 dblink_connect
----------------
 OK
(1 row)

SELECT dblink_exec('myconn', 'insert into foo values(21,''z'',''{"a0",
"b0","c0"}'');');
   dblink_exec
------------------
 INSERT 6432584 1
(1 row)

SELECT dblink_exec('myconn', 'insert into pg_class values (''foo'')',
false);
NOTICE:  sql error
DETAIL:  ERROR:  null value in column "relnamespace" violates not-null
 constraint

 dblink_exec
-------------
 ERROR
(1 row)
```

dblink_open

dblink_open — opens a cursor in a remote database

大纲

```
dblink_open(text cursorname, text sql [, bool fail_on_error]) returns
 text
dblink_open(text connname, text cursorname, text sql [, bool fail_on_
error]) returns text
```

描述

`dblink_open()` 在远程数据库中打开游标。随后可以使用 `dblink_fetch()` 和 `dblink_close()` 操作该游标。

参数

connname

要使用的连接名称；省略该参数则使用未命名连接。

cursorname

要分配给该游标的名称。

sql

要在远程数据库中执行的 `SELECT` 语句，例如 `select * from pg_class`。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数的返回值会被设为 `ERROR`。

返回值

返回状态，可以是 `OK` 或 `ERROR`。

注解

由于游标只能在事务中持续存在，如果远程端当前尚未处于事务中，`dblink_open` 会在远程端启动一个显式事务块（`BEGIN`）。当执行匹配的 `dblink_close` 时，该事务会再次关闭。注意，如果在 `dblink_open` 和 `dblink_close` 之间用 `dblink_exec` 修改了数据，随后又发生错误，或者在调用 `dblink_close` 之前使用了 `dblink_disconnect`，所做的更改将会丢失，因为该事务会被中止。

示例

```
SELECT dblink_connect('dbname=postgres');
 dblink_connect
----------------
 OK
```

```
(1 row)

SELECT dblink_open('foo', 'select proname, prosrc from pg_proc');
 dblink_open
-------------
 OK
(1 row)
```

dblink_fetch

dblink_fetch — returns rows from an open cursor in a remote database

大纲

```
dblink_fetch(text cursorname, int howmany [, bool fail_on_error])
 returns setof record
dblink_fetch(text connname, text cursorname, int howmany [, bool fail_
on_error]) returns setof record
```

描述

`dblink_fetch` 从先前由 `dblink_open` 建立的游标中提取行。

参数

connname

要使用的连接名称；省略该参数则使用未命名连接。

cursorname

要从中提取数据的游标名称。

howmany

要检索的最大行数。从当前游标位置开始，向前提取接下来的 *howmany* 行。一旦游标到达末尾，就不会再产生更多行。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数不返回任何行。

返回值

该函数返回从游标提取的行。要使用此函数，需要像前面讨论 `dblink` 时那样，指定预期的列集。

注解

如果 `FROM` 子句中指定的返回列数与远程游标实际返回的列数不一致，就会抛出错误。在这种情况下，远程游标仍会像没有发生该错误一样前进相同的行数。远程 `FETCH` 完成之后，本地查询中发生的任何其他错误也是如此。

示例

```
SELECT dblink_connect('dbname=postgres');
 dblink_connect
----------------
 OK
(1 row)
```

```
SELECT dblink_open('foo', 'select proname, prosrc from pg_proc where
 proname like ''bytea%''');
 dblink_open
-------------
 OK
(1 row)

SELECT * FROM dblink_fetch('foo', 5) AS (funcname name, source text);
 funcname |  source
----------+----------
 byteacat | byteacat
 byteacmp | byteacmp
 byteaeq  | byteaeq
 byteage  | byteage
 byteagt  | byteagt
(5 rows)

SELECT * FROM dblink_fetch('foo', 5) AS (funcname name, source text);
 funcname  |  source
-----------+-----------
 byteain   | byteain
 byteale   | byteale
 bytealike | bytealike
 bytealt   | bytealt
 byteane   | byteane
(5 rows)

SELECT * FROM dblink_fetch('foo', 5) AS (funcname name, source text);
  funcname  |   source
------------+------------
 byteanlike | byteanlike
 byteaout   | byteaout
(2 rows)

SELECT * FROM dblink_fetch('foo', 5) AS (funcname name, source text);
 funcname | source
----------+--------
(0 rows)
```

dblink_close

dblink_close — closes a cursor in a remote database

大纲

```
dblink_close(text cursorname [, bool fail_on_error]) returns text
dblink_close(text connname, text cursorname [, bool fail_on_error])
 returns text
```

描述

dblink_close 关闭先前由 dblink_open 打开的游标。

参数

connname

要使用的连接名称；省略该参数则使用未命名连接。

cursorname

要关闭的游标名。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数的返回值会被设为 ERROR。

返回值

返回状态，可以是 OK 或 ERROR。

注解

如果 dblink_open 启动了显式事务块，并且这是此连接上最后一个仍然打开的游标，dblink_close 将发出对应的 COMMIT。

示例

```
SELECT dblink_connect('dbname=postgres');
 dblink_connect
----------------
 OK
(1 row)

SELECT dblink_open('foo', 'select proname, prosrc from pg_proc');
 dblink_open
-------------
 OK
(1 row)

SELECT dblink_close('foo');
```

```
 dblink_close
--------------
 OK
(1 row)
```

dblink_get_connections

dblink_get_connections — returns the names of all open named dblink connections

大纲

```
dblink_get_connections() returns text[]
```

描述

`dblink_get_connections` 返回包含所有已打开命名 `dblink` 连接名称的数组。

返回值

返回包含连接名称的文本数组；如果没有连接，则返回 NULL。

示例

```
SELECT dblink_get_connections();
```

dblink_error_message

dblink_error_message — gets last error message on the named connection

大纲

```
dblink_error_message(text connname) returns text
```

描述

`dblink_error_message` 获取给定连接上最近一次的远程错误消息。

参数

connname

要使用的连接名称。

返回值

返回最后一条错误消息；如果此连接上没有发生过错误，则返回空字符串。

示例

```
SELECT dblink_error_message('dtest1');
```

dblink_send_query

dblink_send_query — sends an async query to a remote database

大纲

```
dblink_send_query(text connname, text sql) returns int
```

描述

`dblink_send_query` 发送要异步执行的查询，也就是不会立即等待结果。该连接上不能已经有正在进行的异步查询。

异步查询成功发送后，可以用 `dblink_is_busy` 检查完成状态，并最终通过 `dblink_get_result` 收集结果。也可以使用 `dblink_cancel_query` 尝试取消活动中的异步查询。

参数

connname

要使用的连接名称。

sql

要在远程数据库中执行的 SQL 语句，例如`select * from pg_class`。

返回值

如果查询已成功发送，则返回 1；否则返回 0。

示例

```
SELECT dblink_send_query('dtest1', 'SELECT * FROM foo WHERE f1 < 3');
```

dblink_is_busy

dblink_is_busy — checks if connection is busy with an async query

大纲

```
dblink_is_busy(text connname) returns int
```

描述

`dblink_is_busy` 测试是否有异步查询正在进行。

参数

connname

要检查的连接名称。

返回值

如果连接正忙则返回 1，否则返回 0。如果此函数返回 0，则可以保证 `dblink_get_result` 不会阻塞。

示例

```
SELECT dblink_is_busy('dtest1');
```

dblink_get_notify

dblink_get_notify — retrieve async notifications on a connection

大纲

```
dblink_get_notify() returns setof (notify_name text, be_pid int, extra
 text)
dblink_get_notify(text connname) returns setof (notify_name text, be_
pid int, extra text)
```

描述

`dblink_get_notify` 检索未命名连接上的通知；如果指定了命名连接，则检索该连接上的通知。要通过 dblink 接收通知，必须先使用 `dblink_exec` 执行 `LISTEN`。详见LISTEN和NOTIFY。

参数

connname

要检索通知的命名连接名称。

返回值

返回 `setof (notify_name text, be_pid int, extra text)`；如果没有，则返回空集。

示例

```
SELECT dblink_exec('LISTEN virtual');
 dblink_exec
-------------
 LISTEN
(1 row)

SELECT * FROM dblink_get_notify();
 notify_name | be_pid | extra
-------------+--------+-------
(0 rows)

NOTIFY virtual;
NOTIFY

SELECT * FROM dblink_get_notify();
 notify_name | be_pid | extra
-------------+--------+-------
 virtual     |   1229 |
(1 row)
```

dblink_get_result

dblink_get_result — gets an async query result

大纲

```
dblink_get_result(text connname [, bool fail_on_error]) returns setof
 record
```

描述

`dblink_get_result` 收集此前用 `dblink_send_query` 发送的异步查询结果。如果该查询尚未完成，`dblink_get_result` 将等待直至其完成。

参数

connname

要使用的连接名称。

fail_on_error

如果为真（省略时默认），则连接远程端抛出的错误也会在本地抛出。如果为假，远程端错误会在本地报告为一个 NOTICE，并且该函数不返回任何行。

返回值

对于异步查询（即返回行的 SQL 语句），该函数返回查询产生的行。要使用此函数，需要像前面讨论 `dblink` 时那样，指定预期的列集。

对于异步命令（即不返回行的 SQL 语句），该函数返回一个只有一个 text 列的单行，其中包含命令的状态字符串。即便如此，仍需要在调用查询的 `FROM` 子句中指定结果只有一个 text 列。

注解

如果 `dblink_send_query` 返回 1，则必须调用此函数。每发送一个查询都必须调用一次，并且还要额外再调用一次以获得空集结果，之后该连接才能再次使用。

示例

```
contrib_regression=# SELECT dblink_connect('dtest1', 'dbname=contrib_
regression');
 dblink_connect
----------------
 OK
(1 row)

contrib_regression=# SELECT * FROM
contrib_regression-# dblink_send_query('dtest1', 'select * from foo
 where f1 < 3') AS t1;
 t1
----
  1
```

```
(1 row)

contrib_regression=# SELECT * FROM dblink_get_result('dtest1') AS t1(
f1 int, f2 text, f3 text[]);
 f1 | f2 |     f3
----+----+------------
  0 | a  | {a0,b0,c0}
  1 | b  | {a1,b1,c1}
  2 | c  | {a2,b2,c2}
(3 rows)

contrib_regression=# SELECT * FROM dblink_get_result('dtest1') AS t1(
f1 int, f2 text, f3 text[]);
 f1 | f2 | f3
----+----+----
(0 rows)

contrib_regression=# SELECT * FROM
contrib_regression-# dblink_send_query('dtest1', 'select * from foo
 where f1 < 3; select * from foo where f1 > 6') AS t1;
 t1
----
  1
(1 row)

contrib_regression=# SELECT * FROM dblink_get_result('dtest1') AS t1(
f1 int, f2 text, f3 text[]);
 f1 | f2 |     f3
----+----+------------
  0 | a  | {a0,b0,c0}
  1 | b  | {a1,b1,c1}
  2 | c  | {a2,b2,c2}
(3 rows)

contrib_regression=# SELECT * FROM dblink_get_result('dtest1') AS t1(
f1 int, f2 text, f3 text[]);
 f1 | f2 |      f3
----+----+---------------
  7 | h  | {a7,b7,c7}
  8 | i  | {a8,b8,c8}
  9 | j  | {a9,b9,c9}
 10 | k  | {a10,b10,c10}
(4 rows)

contrib_regression=# SELECT * FROM dblink_get_result('dtest1') AS t1(
f1 int, f2 text, f3 text[]);
 f1 | f2 | f3
----+----+----
(0 rows)
```

dblink_cancel_query

dblink_cancel_query — cancels any active query on the named connection

大纲

```
dblink_cancel_query(text connname) returns text
```

描述

`dblink_cancel_query` 尝试取消命名连接上正在进行的任何查询。注意，这不一定会成功（例如远程查询可能已经结束）。取消请求只会提高该查询很快失败的概率。
仍必须完成正常的查询协议，例如调用 `dblink_get_result`。

参数

connname

要使用的连接名称。

返回值

如果取消请求已发送，则返回 `OK`；如果失败，则返回错误消息文本。

示例

```
SELECT dblink_cancel_query('dtest1');
```

dblink_get_pkey

dblink_get_pkey — returns the positions and field names of a relation's primary key fields

大纲

```
dblink_get_pkey(text relname) returns setof dblink_pkey_results
```

描述

`dblink_get_pkey` 提供有关本地数据库中某个关系主键的信息。这有时有助于生成要发送到远程数据库的查询。

参数

relname

本地关系的名称，例如 `foo` 或 `myschema.mytab`。如果该名称是大小写混合的或包含特殊字符，则应包含双引号，例如 `"FooBar"`；如果没有引号，字符串将被折叠为小写形式。

返回值

为每个主键字段返回一行；如果该关系没有主键，则不返回任何行。结果行类型定义如下：

```
CREATE TYPE dblink_pkey_results AS (position int, colname text);
```

`position` 列只是从 1 到 *N* 递增；它表示该字段在主键中的序号，而不是在表列中的序号。

示例

```
CREATE TABLE foobar (
    f1 int,
    f2 int,
    f3 int,
    PRIMARY KEY (f1, f2, f3)
);
CREATE TABLE

SELECT * FROM dblink_get_pkey('foobar');
 position | colname
----------+---------
        1 | f1
        2 | f2
        3 | f3
(3 rows)
```

dblink_build_sql_insert

dblink_build_sql_insert — builds an INSERT statement using a local tuple, replacing the primary key field values with alternative supplied values

大纲

```
dblink_build_sql_insert(text relname,
                        int2vector primary_key_attnums,
                        integer num_primary_key_atts,
                        text[] src_pk_att_vals_array,
                        text[] tgt_pk_att_vals_array) returns text
```

描述

`dblink_build_sql_insert` 可用于将本地表有选择地复制到远程数据库。它根据主键从本地表中选取一行，然后构建一条会复制该行的 SQL `INSERT` 命令，但主键值会替换为最后一个参数中的值。（若要精确复制该行，只需为最后两个参数指定相同的值。）

参数

relname

本地关系的名称，例如 `foo` 或 `myschema.mytab`。如果该名称是大小写混合的或包含特殊字符，则应包含双引号，例如 `"FooBar"`；如果没有引号，字符串将被折叠为小写形式。

primary_key_attnums

主键字段的属性编号（从 1 开始），例如 `1 2`。

num_primary_key_atts

主键字段数量。

src_pk_att_vals_array

用于查找本地元组的主键字段值。每个字段都以文本形式表示。
如果没有本地行具有这些主键值，则抛出错误。

tgt_pk_att_vals_array

放入结果 `INSERT` 命令中的主键字段值。每个字段都以文本形式表示。

返回值

以文本形式返回所请求的 SQL 语句。

注解

从 PostgreSQL 9.0 起，*primary_key_attnums* 中的属性编号会被解释为逻辑列号，对应于列在 `SELECT * FROM relname` 中的位置。早期版本将这些编号解释为物理列位置。如果在表的生命周期中，指示列左侧的某些列已被删除，两者就会有差异。

示例

```
SELECT dblink_build_sql_insert('foo', '1 2', 2, '{"1", "a"}', '{"1",
 "b''a"}');
            dblink_build_sql_insert
--------------------------------------------------
 INSERT INTO foo(f1,f2,f3) VALUES('1','b''a','1')
(1 row)
```

dblink_build_sql_delete

dblink_build_sql_delete — builds a DELETE statement using supplied values for primary key field values

大纲

```
dblink_build_sql_delete(text relname,
                        int2vector primary_key_attnums,
                        integer num_primary_key_atts,
                        text[] tgt_pk_att_vals_array) returns text
```

描述

`dblink_build_sql_delete` 可用于将本地表有选择地复制到远程数据库。它会构建一条 SQL `DELETE` 命令，用来删除具有给定主键值的行。

参数

relname

本地关系的名称，例如 `foo` 或 `myschema.mytab`。如果该名称是大小写混合的或包含特殊字符，则应包含双引号，例如 `"FooBar"`；如果没有引号，字符串将被折叠为小写形式。

primary_key_attnums

主键字段的属性编号（从 1 开始），例如 `1 2`。

num_primary_key_atts

主键字段数量。

tgt_pk_att_vals_array

用于结果 `DELETE` 命令中的主键字段值。每个字段都以文本形式表示。

返回值

以文本形式返回所请求的 SQL 语句。

注解

从 PostgreSQL 9.0 起，*primary_key_attnums* 中的属性编号会被解释为逻辑列号，对应于列在 `SELECT * FROM relname` 中的位置。早期版本将这些编号解释为物理列位置。如果在表的生命周期中，指示列左侧的某些列已被删除，两者就会有差异。

示例

```
SELECT dblink_build_sql_delete('"MyFoo"', '1 2', 2, '{"1", "b"}');
           dblink_build_sql_delete
----------------------------------------------
 DELETE FROM "MyFoo" WHERE f1='1' AND f2='b'
(1 row)
```

dblink_build_sql_update

dblink_build_sql_update — builds an UPDATE statement using a local tuple, replacing the primary key field values with alternative supplied values

大纲

```
dblink_build_sql_update(text relname,
                        int2vector primary_key_attnums,
                        integer num_primary_key_atts,
                        text[] src_pk_att_vals_array,
                        text[] tgt_pk_att_vals_array) returns text
```

描述

`dblink_build_sql_update` 可用于将本地表有选择地复制到远程数据库。它根据主键从本地表中选取一行，然后构建一条会复制该行的 SQL `UPDATE` 命令，但主键值会替换为最后一个参数中的值。（若要精确复制该行，只需为最后两个参数指定相同的值。）`UPDATE` 命令总是对该行的所有字段赋值 – 它与 `dblink_build_sql_insert` 的主要区别在于，前者假定目标行已经存在于远程表中。

参数

relname

本地关系的名称，例如 `foo` 或 `myschema.mytab`。如果该名称是大小写混合的或包含特殊字符，则应包含双引号，例如 `"FooBar"`；如果没有引号，字符串将被折叠为小写形式。

primary_key_attnums

主键字段的属性编号（从 1 开始），例如 `1 2`。

num_primary_key_atts

主键字段数量。

src_pk_att_vals_array

用于查找本地元组的主键字段值。每个字段都以文本形式表示。
如果没有本地行具有这些主键值，则抛出错误。

tgt_pk_att_vals_array

用于结果 `UPDATE` 命令中的主键字段值。每个字段都以文本形式表示。

返回值

以文本形式返回所请求的 SQL 语句。

注解

从 PostgreSQL 9.0 起，*primary_key_attnums* 中的属性编号会被解释为逻辑列号，对应于列在 `SELECT * FROM relname` 中的位置。早期版本将这些编号解释为物理列位置。如果在表的生命周期中，指示列左侧的某些列已被删除，两者就会有差异。

示例

```
SELECT dblink_build_sql_update('foo', '1 2', 2, '{"1", "a"}', '{"1",
 "b"}');
                   dblink_build_sql_update
-------------------------------------------------------------
 UPDATE foo SET f1='1',f2='b',f3='1' WHERE f1='1' AND f2='b'
(1 row)
```

F.10. dict_int – 用于整数的示例全文搜索词典

dict_int是一个全文搜索附加词典模板的示例。
引入这个示例词典是为了控制整数（有符号和无符号）的索引，使这类数字能够被索引，
同时又避免唯一词的数量过度增长，因为那会严重影响搜索性能。

F.10.1. 配置

该词典接受两个选项：

- maxlen参数指定整数词中允许的最大数字位数。默认值为 6。
- rejectlong参数指定超长整数应被截断还是忽略。如果rejectlong为false（默认值），
  词典将返回该整数的前maxlen位数字。如果rejectlong为true，
  词典会将超长整数视为停用词，因此不会为其建立索引。注意，
  这也意味着无法搜索这类整数。

F.10.2. 用法

安装dict_int扩展后，会创建一个文本搜索模板 intdict_template以及一个基于该模板、使用默认参数的词典 intdict。你可以修改这些参数，例如：

```
mydb# ALTER TEXT SEARCH DICTIONARY intdict (MAXLEN = 4, REJECTLONG =
 true);
ALTER TEXT SEARCH DICTIONARY
```

或者基于该模板创建新的词典。

要测试该词典，可以尝试：

```
mydb# select ts_lexize('intdict', '12345678');
 ts_lexize
-----------
 {123456}
```

但在实际使用中，通常需要像第 12 章所述那样，将它包含到某个文本搜索配置中。可能类似如下：

```
ALTER TEXT SEARCH CONFIGURATION english
    ALTER MAPPING FOR int, uint WITH intdict;
```

F.11. dict_xsyn – 示例同义词全文检索词典

dict_xsyn（扩展同义词词典）是全文检索的一个附加词典模板示例。
这种词典类型会将词替换为其同义词组，因此可以使用某个词的任一同义词来搜索该词。

F.11.1. 配置

该dict_xsyn词典接受以下选项：

- matchorig控制词典是否接受原词。默认为true。
- matchsynonyms控制词典是否接受同义词。默认为false。
- keeporig控制词典输出中是否包含原词。默认为true。
- keepsynonyms控制词典输出中是否包含同义词。默认为true。
- rules是包含同义词列表的文件的基本名。该文件必须存放在$SHAREDIR/tsearch_data/（其中$SHAREDIR表示PostgreSQL安装的共享数据目录）中。文件名必须以.rules结尾（这一后缀不应写入rules参数）。

规则文件具有下列格式：

- 每一行表示某个单词的一组同义词，且该单词位于行首。同义词之间用空白分隔，例如：

  ```
  word syn1 syn2 syn3
  ```

- 井号（#）是注释分隔符。它可以出现在一行中的任意位置，该行余下的内容都会被跳过。

示例可参见安装在$SHAREDIR/tsearch_data/中的 xsyn_sample.rules。

F.11.2. 用法

安装dict_xsyn扩展会使用默认参数创建一个全文检索模板 xsyn_template和一个基于它的词典xsyn。你可以修改这些参数，例如：

```
mydb# ALTER TEXT SEARCH DICTIONARY xsyn (RULES='my_rules', KEEPORIG=
false);
ALTER TEXT SEARCH DICTIONARY
```

或者基于该模板创建新的词典。

要测试该词典，可以尝试：

```
mydb=# SELECT ts_lexize('xsyn', 'word');
      ts_lexize
-----------------------
 {syn1,syn2,syn3}

mydb# ALTER TEXT SEARCH DICTIONARY xsyn (RULES='my_rules', KEEPORIG=
true);
ALTER TEXT SEARCH DICTIONARY

mydb=# SELECT ts_lexize('xsyn', 'word');
```

```
      ts_lexize
-----------------------
 {word,syn1,syn2,syn3}

mydb# ALTER TEXT SEARCH DICTIONARY xsyn (RULES='my_rules', KEEPORIG=
false, MATCHSYNONYMS=true);
ALTER TEXT SEARCH DICTIONARY

mydb=# SELECT ts_lexize('xsyn', 'syn1');
      ts_lexize
-----------------------
 {syn1,syn2,syn3}

mydb# ALTER TEXT SEARCH DICTIONARY xsyn (RULES='my_rules', KEEPORIG=
true, MATCHORIG=false, KEEPSYNONYMS=false);
ALTER TEXT SEARCH DICTIONARY

mydb=# SELECT ts_lexize('xsyn', 'syn1');
      ts_lexize
-----------------------
 {word}
```

但实际使用通常需要把它纳入第 12 章中描述的全文检索配置。可能会像这样：

```
ALTER TEXT SEARCH CONFIGURATION english
    ALTER MAPPING FOR word, asciiword WITH xsyn, english_stem;
```

F.12. dummy_seclabel

dummy_seclabel模块的存在只是为了支持对SECURITY LABEL语句的回归测试。它不适合在生产环境中使用。

F.12.1. 缘由

SECURITY LABEL语句允许用户为数据库对象分配安全标签；但是，只有当可装载模块明确允许时才能分配安全标签，因此提供这个模块是为了能够进行正常的回归测试。

打算在生产环境中使用的安全标签提供程序通常依赖于特定于平台的特性，例如SELinux。这个模块是平台无关的，因此更适合用于回归测试。

F.12.2. 用法

下面是一个简单的用法示例：

```
# postgresql.conf
shared_preload_libraries = 'dummy_seclabel'

postgres=# CREATE TABLE t (a int, b text);
CREATE TABLE
postgres=# SECURITY LABEL ON TABLE t IS 'classified';
SECURITY LABEL
```

dummy_seclabel模块只提供四个硬编码的标签：unclassified、classified、secret和top secret。它不允许将任何其他字符串用作安全标签。

这些标签并不用于实施访问控制。它们只用于检查SECURITY LABEL语句是否按预期工作。

F.12.3. 作者

KaiGai Kohei <kaigai@ak.jp.nec.com>

F.13. earthdistance – 计算大圆距离

earthdistance 模块提供了两种不同的方法，用于计算地球表面上的大圆距离。首先介绍的一种依赖于 cube 模块（在安装 earthdistance 之前必须先安装该模块）。第二种方法基于内置的 point 数据类型，使用经度和纬度作为坐标。

在这个模块中，地球被假定为完全球形的。（如果这种假设对你而言精度不足，可以参见 PostGIS[1] 项目。）

F.13.1. 基于立方体的地球距离

数据以立方体形式存储，而这些立方体实际上是点（两个角相同），使用 3 个坐标分别表示距地心的 x、y、z 距离。这里还提供了一个基于 cube 的域 earth，其中包含约束检查，用以确认该值满足这些限制，并且与地球实际表面足够接近。

地球半径取自 earth() 函数，其单位为米。不过，只要修改这一个函数，你就可以让该模块改用其他单位，或者使用你认为更合适的半径值。

这个模块在天文数据库中也有用途。天文学家大概会希望修改 earth()，使其返回半径 180/pi()，这样距离单位就是度。

这里提供了一些函数，用于按经纬度（以度为单位）输入、输出经纬度、
计算两点之间的大圆距离，以及方便地指定一个可用于索引搜索的边界框。

所提供的函数见表 F.6。

表 F.6. 基于立方体的地球距离函数

Function	Returns	Description
earth()	float8	Returns the assumed radius of the Earth.
sec_to_gc(float8)	float8	Converts the normal straight line (secant) distance between between two points on the surface of the Earth to the great circle distance between them.
gc_to_sec(float8)	float8	Converts the great circle distance between two points on the surface of the Earth to the normal straight line (secant) distance between them.
ll_to_earth(float8, float8)	earth	Returns the location of a point on the surface of the Earth given its latitude (argument 1) and longitude (argument 2) in degrees.
latitude(earth)	float8	Returns the latitude in degrees of a point on the surface of the Earth.
longitude(earth)	float8	Returns the longitude in degrees of a point on the surface of the Earth.

[1] http://postgis.net/

Function	Returns	Description
`earth_distance(earth, earth)`	`float8`	Returns the great circle distance between two points on the surface of the Earth.
`earth_box(earth, float8)`	`cube`	Returns a box suitable for an indexed search using the cube `@>` operator for points within a given great circle distance of a location. Some points in this box are further than the specified great circle distance from the location, so a second check using `earth_distance` should be included in the query.

F.13.2. 基于点的地球距离

该模块的第二部分依赖于将地球上的位置表示为 `point` 类型的值，其中第一个分量被视为以度为单位的经度，第二个分量被视为以度为单位的纬度。之所以采用 (longitude, latitude) 而不是相反的顺序，是因为经度更接近 x 轴的直观概念，而纬度更接近 y 轴。

这一部分只提供了一个操作符，见表 F.7。

表 F.7. 基于点的地球距离操作符

Operator	Returns	Description
`point <@> point`	`float8`	Gives the distance in statute miles between two points on the Earth's surface.

注意，与该模块基于 `cube` 的部分不同，这里的单位是固定的：修改 `earth()` 函数不会影响该操作符的结果。

经度/纬度表示的一个缺点是，你需要小心靠近两极以及接近经度 +/- 180 度时的边界情况。基于 `cube` 的表示可以避免这些不连续性。

F.14. file_fdw

file_fdw 模块提供外部数据包装器 file_fdw，它可用于访问服务器文件系统中的数据文件。数据文件必须采用 COPY FROM 可读取的格式; see COPY for details.

使用该包装器创建的外部表可以具有以下选项：

filename

指定要读取的文件。必需。必须是绝对路径名。

format

指定文件的格式，与COPY的FORMAT选项相同。

header

指定文件是否包含标题行，与COPY的HEADER选项相同。

delimiter

指定文件的定界符字符，与COPY的DELIMITER选项相同。

quote

指定文件的引用字符，与COPY的QUOTE选项相同。

escape

指定文件的转义字符，与COPY的ESCAPE选项相同。

null

指定文件的空值串，与COPY的NULL选项相同。

encoding

指定文件的编码，与 COPY 的 ENCODING 选项相同。

file_fdw 目前不支持 COPY 的 OIDS、FORCE_QUOTE 和 FORCE_NOT_NULL 选项。

这些选项只能为外部表指定，不能在 file_fdw 外部数据包装器的选项中指定，也不能在使用该包装器的服务器或用户映射的选项中指定。

出于安全原因，更改表级选项需要超级用户权限：只有超级用户才能决定读取哪个文件。
原则上可以允许非超级用户更改其他选项，但目前尚不支持。

对于使用file_fdw的外部表，EXPLAIN会显示待读取的文件名。除非指定COSTS OFF，否则还会显示文件大小（以字节为单位）。

F.15. fuzzystrmatch – 确定字符串的相似性和距离

fuzzystrmatch模块提供多个函数，用于确定字符串的相似性和距离。

> **小心**
>
> 目前，soundex、metaphone、dmetaphone 和 dmetaphone_alt 函数不能很好地处理多字节编码（例如 UTF-8）。

F.15.1. Soundex

Soundex 系统是一种通过将发音相近的名字转换为相同代码来进行匹配的方法。它最初在 1880 年、1900 年和 1910 年的美国人口普查中使用。注意，Soundex 对非英语名字并不是很有用。

fuzzystrmatch模块提供两个用于处理 Soundex 代码的函数：

```
soundex(text) returns text
difference(text, text) returns int
```

soundex函数将字符串转换为其 Soundex 代码。difference函数将两个字符串转换为各自的 Soundex 代码，然后返回代码中相同位置上匹配的个数。由于 Soundex 代码有四个字符，因此结果范围为 0 到 4，其中 0 表示没有匹配，4 表示完全匹配。（因此，这个函数的命名并不贴切，similarity 会是更好的名字。）

下面是一些用法示例：

```
SELECT soundex('hello world!');

SELECT soundex('Anne'), soundex('Ann'), difference('Anne', 'Ann');
SELECT soundex('Anne'), soundex('Andrew'), difference('Anne',
 'Andrew');
SELECT soundex('Anne'), soundex('Margaret'), difference('Anne',
 'Margaret');

CREATE TABLE s (nm text);

INSERT INTO s VALUES ('john');
INSERT INTO s VALUES ('joan');
INSERT INTO s VALUES ('wobbly');
INSERT INTO s VALUES ('jack');

SELECT * FROM s WHERE soundex(nm) = soundex('john');

SELECT * FROM s WHERE difference(s.nm, 'john') > 2;
```

F.15.2. Levenshtein

该函数计算两个字符串之间的 Levenshtein 距离：

```
levenshtein(text source, text target, int ins_cost, int del_cost, int
 sub_cost) returns int
```

```
levenshtein(text source, text target) returns int
levenshtein_less_equal(text source, text target, int ins_cost, int
 del_cost, int sub_cost, int max_d) returns int
levenshtein_less_equal(text source, text target, int max_d) returns
 int
```

`source` 和 `target` 都可以是任意非空字符串，最大长度为 255 字节。代价参数分别指定字符插入、删除或替换的代价。可以像该函数的第二种形式那样省略代价参数；此时它们都默认为 1。`levenshtein_less_equal` 是 Levenshtein 函数在距离值较小时的加速版本。如果实际距离小于等于 max_d，`levenshtein_less_equal` 会返回准确的距离值；否则该函数返回某个大于 max_d 的值。

示例：

```
test=# SELECT levenshtein('GUMBO', 'GAMBOL');
 levenshtein
-------------
           2
(1 row)

test=# SELECT levenshtein('GUMBO', 'GAMBOL', 2,1,1);
 levenshtein
-------------
           3
(1 row)

test=# SELECT levenshtein_less_equal('extensive', 'exhaustive',2);
 levenshtein_less_equal
------------------------
                      3
(1 row)

test=# SELECT levenshtein_less_equal('extensive', 'exhaustive',4);
 levenshtein_less_equal
------------------------
                      4
(1 row)
```

F.15.3. Metaphone

Metaphone 与 Soundex 一样，基于为输入字符串构造一个代表性代码的思想。如果两个字符串具有相同的代码，则认为它们相似。

该函数计算输入字符串的 Metaphone 代码：

```
metaphone(text source, int max_output_length) returns text
```

`source` 必须是非空字符串，最大长度为 255 字节。`max_output_length` 设置输出 Metaphone 代码的最大长度；如果超过该长度，输出会被截断为该长度。

示例：

```
test=# SELECT metaphone('GUMBO', 4);
 metaphone
-----------
```

```
 KM
(1 row)
```

F.15.4. Double Metaphone

Double Metaphone 系统会为给定输入字符串计算两个“近音”代码，即一个“主”代码和一个“备选”代码。在大多数情况下，它们相同，但尤其对于非英语名字，
这两个代码可能会因发音不同而略有差异。这些函数分别计算主代码和备选代码：

```
dmetaphone(text source) returns text
dmetaphone_alt(text source) returns text
```

输入字符串没有长度限制。

示例：

```
test=# select dmetaphone('gumbo');
 dmetaphone
------------
 KMP
(1 row)
```

F.16. hstore

此模块实现了hstore数据类型，用于在单个 PostgreSQL值中存储一组键/值对。这在多种场景中都很有用，例如属性很多但很少查看的行，或者半结构化数据。键和值都只是文本字符串。

F.16.1. hstore 外部表示

用于输入和输出的hstore文本表示包含零个或多个以逗号分隔的 *key* => *value* 对。一些示例：

```
k => v
foo => bar, baz => whatever
"1-a" => "anything at all"
```

键/值对的顺序并不重要（而且在输出时可能不会按原样重现）。键/值对之间或 => 号周围的空白会被忽略。包含空白、逗号、= 或 > 的键和值必须用双引号括起来。要在键或值中包含双引号或反斜线，请用反斜线转义。

每个hstore中的键都是唯一的。如果声明的hstore 带有重复键，则在该hstore中只会存储其中一个，而且无法保证保留的是哪一个：

```
SELECT 'a=>1,a=>2'::hstore;
  hstore
----------
 "a"=>"1"
```

值（但键不能）可以是 SQL NULL。例如：

```
key => NULL
```

NULL关键字不区分大小写。若要将NULL 视为普通字符串“NULL”，请用双引号括起来。

> **注意**
>
> 请注意，当hstore文本格式用于输入时，它会在任何必需的加引号或转义之前应用。如果通过参数传递一个 hstore字面量，则不需要额外处理。但如果将其作为带引号的字面量常量传递，那么其中的单引号字符以及（取决于 standard_conforming_strings配置参数的设置）反斜线字符都需要被正确转义。关于字符串常量的处理，见第 4.1.2.1 节。

在输出时，即使严格来说并非必需，键和值也总是带有双引号。

F.16.2. hstore 操作符和函数

hstore模块提供的操作符见表 F.8，函数见表 F.9。

表 F.8. hstore 操作符

Operator	Description	Example	Result
hstore -> text	get value for key (NULL if not present)	'a=>x, b=>y'::hstore -> 'a'	x
hstore -> text[]	get values for keys (NULL if not present)	'a=>x, b=>y, c=>z'::hstore -> ARRAY['c','a']	{"z","x"}

Operator	Description	Example	Result
`text => text`	make single-pair `hstore`	`'a' => 'b'`	`"a"=>"b"`
`hstore || hstore`	concatenate `hstore`s	`'a=>b, c=>d'::hstore || 'c=>x, d=>q'::hstore`	`"a"=>"b", "c"=>"x", "d"=>"q"`
`hstore ? text`	does `hstore` contain key?	`'a=>1'::hstore ? 'a'`	`t`
`hstore ?& text[]`	does `hstore` contain all specified keys?	`'a=>1,b=>2'::hstore ?& ARRAY['a','b']`	`t`
`hstore ?| text[]`	does `hstore` contain any of the specified keys?	`'a=>1,b=>2'::hstore ?| ARRAY['b','c']`	`t`
`hstore @> hstore`	does left operand contain right?	`'a=>b, b=>1, c=>NULL'::hstore @> 'b=>1'`	`t`
`hstore <@ hstore`	is left operand contained in right?	`'a=>c'::hstore <@ 'a=>b, b=>1, c=>NULL'`	`f`
`hstore - text`	delete key from left operand	`'a=>1, b=>2, c=>3'::hstore - 'b'::text`	`"a"=>"1", "c"=>"3"`
`hstore - text[]`	delete keys from left operand	`'a=>1, b=>2, c=>3'::hstore - ARRAY['a','b']`	`"c"=>"3"`
`hstore - hstore`	delete matching pairs from left operand	`'a=>1, b=>2, c=>3'::hstore - 'a=>4, b=>2'::hstore`	`"a"=>"1", "c"=>"3"`
`record #= hstore`	replace fields in `record` with matching values from `hstore`	see Examples section	
`%% hstore`	convert `hstore` to array of alternating keys and values	`%% 'a=>foo, b=>bar'::hstore`	`{a,foo,b,bar}`
`%# hstore`	convert `hstore` to two-dimensional key/value array	`%# 'a=>foo, b=>bar'::hstore`	`{{a,foo},{b,bar}}`

注意

在 PostgreSQL 8.2 之前，包含操作符 `@>` 和 `<@` 分别称为 `@` 和 `~`。这些名称仍然可用，但已弃用，最终将被删除。请注意，旧名称与核心几何数据类型以前采用的约定正好相反！

注意

`=>` 操作符已弃用，未来的发行版中可能会将其删除。请改用 `hstore(text, text)` 函数。

表 F.9. hstore Functions

Function	Return Type	Description	Example	Result
hstore(record)	hstore	construct an hstore from a record or row	hstore(ROW(1, 2))	f1=>1,f2=>2
hstore(text[])	hstore	construct an hstore from an array, which may be either a key/value array, or a two-dimensional array	hstore(ARRAY['a','1','b','2']) || hstore(ARRAY[['c','3'],['d','4']])	a=>1, b=>2, c=>3, d=>4
hstore(text[], text[])	hstore	construct an hstore from separate key and value arrays	hstore(ARRAY['a','b'], ARRAY['1','2'])	"a"=>"1","b"=>"2"
hstore(text, text)	hstore	make single-item hstore	hstore('a', 'b')	"a"=>"b"
akeys(hstore)	text[]	get hstore's keys as an array	akeys('a=>1,b=>2')	{a,b}
skeys(hstore)	setof text	get hstore's keys as a set	skeys('a=>1,b=>2')	a b
avals(hstore)	text[]	get hstore's values as an array	avals('a=>1,b=>2')	{1,2}
svals(hstore)	setof text	get hstore's values as a set	svals('a=>1,b=>2')	1 2
hstore_to_array(hstore)	text[]	get hstore's keys and values as an array of altern ating keys and values	hstore_to_array('a=>1,b=>2')	{a,1,b,2}
hstore_to_matrix(hstore)	text[]	get hstore's keys and values as a two-dimensional array	hstore_to_matrix('a=>1,b=>2')	{{a,1},{b,2}}
slice(hstore, text[])	hstore	extract a subset of an hstore	slice('a=>1,b=>2,c=>3'::hstore, ARRAY['b','c','x'])	"b"=>"2", "c"=>"3"
each(hstore)	setof(key text, value text)	get hstore's keys and values as a set	select * from each('a=>1,b=>2')	key | value -----+------- a | 1 b | 2
exist(hstore, text)	boolean	does hstore contain key?	exist('a=>1','a')	t
defined(hstore,text)	boolean	does hstore contain non-NULL value for key?	defined('a=>NULL','a')	f
delete(hstore, text)	hstore	delete pair with matching key	delete('a=>1,b=>2','b')	"a"=>"1"

Function	Return Type	Description	Example	Result
delete(hstore, text[])	hstore	delete pairs with matching keys	delete('a=>1, b=>2,c=>3', ARRAY['a','b'])	"c"=>"3"
delete(hstore, hstore)	hstore	delete pairs matching those in the second argument	delete('a=>1, b=>2','a=>4,b= >2'::hstore)	"a"=>"1"
populate_ record(record, hstore)	record	replace fields in record with matching values from hstore	see Examples section	

注意

populate_record函数的第一个参数实际声明为anyelement，而不是record，但它会在运行时出错并拒绝非记录类型。

F.16.3. 索引

hstore支持针对@>、?、?&和?|操作符的 GiST 和 GIN 索引。例如：

```
CREATE INDEX hidx ON testhstore USING GIST (h);

CREATE INDEX hidx ON testhstore USING GIN (h);
```

hstore也支持用于=操作符的 btree或hash索引。这允许hstore列被声明为 UNIQUE，或者用于GROUP BY、ORDER BY或DISTINCT表达式。hstore值的排序顺序本身并没有特别实用的意义，但这些索引可能适合用于等值查找。可按如下方式为=比较创建索引：

```
CREATE INDEX hidx ON testhstore USING BTREE (h);

CREATE INDEX hidx ON testhstore USING HASH (h);
```

F.16.4. 示例

添加一个键，或用新值更新现有键：

```
UPDATE tab SET h = h || ('c' => '3');
```

删除一个键：

```
UPDATE tab SET h = delete(h, 'k1');
```

将record转换为hstore：

```
CREATE TABLE test (col1 integer, col2 text, col3 text);
INSERT INTO test VALUES (123, 'foo', 'bar');

SELECT hstore(t) FROM test AS t;
                   hstore
```

```
----------------------------------------------
 "col1"=>"123", "col2"=>"foo", "col3"=>"bar"
(1 row)
```

将hstore转换为预定义的record类型：

```
CREATE TABLE test (col1 integer, col2 text, col3 text);

SELECT * FROM populate_record(null::test,
                              '"col1"=>"456", "col2"=>"zzz"');
 col1 | col2 | col3
------+------+------
  456 | zzz  |
(1 row)
```

使用hstore中的值修改现有记录：

```
CREATE TABLE test (col1 integer, col2 text, col3 text);
INSERT INTO test VALUES (123, 'foo', 'bar');

SELECT (r).* FROM (SELECT t #= '"col3"=>"baz"' AS r FROM test t) s;
 col1 | col2 | col3
------+------+------
  123 | foo  | baz
(1 row)
```

F.16.5. 统计信息

由于hstore类型本身比较宽松，它可能包含大量不同的键。检查键是否合法是应用程序的任务。下面的示例展示了检查键并获取统计信息的几种技术。

简单示例：

```
SELECT * FROM each('aaa=>bq, b=>NULL, ""=>1');
```

使用表：

```
SELECT (each(h)).key, (each(h)).value INTO stat FROM testhstore;
```

在线统计信息：

```
SELECT key, count(*) FROM
  (SELECT (each(h)).key FROM testhstore) AS stat
  GROUP BY key
  ORDER BY count DESC, key;
    key    | count
-----------+-------
 line      |   883
 query     |   207
 pos       |   203
 node      |   202
 space     |   197
 status    |   195
 public    |   194
 title     |   190
```

```
 org       |   189
....................
```

F.16.6. 兼容性

自 PostgreSQL 9.0 起，`hstore`使用了与更早版本不同的内部表示。这不会妨碍转储/恢复升级，因为文本表示（即转储中使用的表示）没有改变。

在进行二进制升级时，通过让新代码识别旧格式数据，维持了向上兼容性。
这会在处理尚未被新代码修改过的数据时带来轻微的性能损失。可以通过执行如下 `UPDATE`语句，强制升级表列中的所有值：

```
UPDATE tablename SET hstorecol = hstorecol || '';
```

另一种方式是：

```
ALTER TABLE tablename ALTER hstorecol TYPE hstore USING hstorecol ||
 '';
```

使用`ALTER TABLE`方法需要对表加独占锁，但不会因旧行版本而导致表膨胀。

F.16.7. 作者

Oleg Bartunov `<oleg@sai.msu.su>`，俄罗斯莫斯科，莫斯科大学

Teodor Sigaev `<teodor@sigaev.ru>`，俄罗斯莫斯科，Delta-Soft Ltd.

Andrew Gierth `<andrew@tao11.riddles.org.uk>`，英国，对本模块作了额外增强

F.17. intagg – 整数聚合器和枚举器

intagg 模块提供一个整数聚合器和一个枚举器。由于已经有内置函数提供了其能力的超集，所以 intagg 现已过时。不过，该模块仍作为这些内置函数的兼容性包装器提供。

F.17.1. 函数

聚合器是聚合函数int_array_aggregate(integer)，它会生成一个整数数组，
其中恰好包含输入给它的那些整数。这是对array_agg的包装器，
后者对任意数组类型都能做同样的事情。

枚举器是函数int_array_enum(integer[])，它返回setof integer。本质上，它是聚合器的逆操作：给定一个整数数组，将其展开为一组行。这是对unnest的包装器，
后者对任意数组类型都能做同样的事情。

F.17.2. 使用示例

许多数据库系统有一对多表的概念。这种表通常位于两个带索引的表之间，例如：

```
CREATE TABLE left (id INT PRIMARY KEY, ...);
CREATE TABLE right (id INT PRIMARY KEY, ...);
CREATE TABLE one_to_many(left INT REFERENCES left, right INT
 REFERENCES right);
```

它通常按如下方式使用：

```
SELECT right.* from right JOIN one_to_many ON (right.id = one_to_many.
right)
  WHERE one_to_many.left = item;
```

这将返回左侧表中某个条目在右侧表中对应的所有项。这是 SQL 中非常常见的一种构造。

不过，如果one_to_many表中的条目非常多，这种方法就会变得相当繁琐。通常，
对于左侧表中的某个特定条目，右侧表中的每个对应项都要执行一次索引扫描并取回一行。
如果你的系统非常动态，那就没有太多办法。不过，如果有一部分数据相当静态，
可以借助聚合器创建一个汇总表。

```
CREATE TABLE summary AS
  SELECT left, int_array_aggregate(right) AS right
  FROM one_to_many
  GROUP BY left;
```

这样会创建一个表，其中左侧每个条目对应一行，并附带一个由右侧各项组成的数组。不过，
如果没有某种使用该数组的方法，它就几乎没有用处；这正是数组枚举器存在的原因。
你可以这样做：

```
SELECT left, int_array_enum(right) FROM summary WHERE left = item;
```

上面这个使用int_array_enum的查询，产生的结果与下面的查询相同：

```
SELECT left, right FROM one_to_many WHERE left = item;
```

区别在于，针对汇总表的查询只需要从表中取出一行，而直接查询one_to_many
则必须对每个对应项都进行索引扫描并取回一行。

在某个系统上，EXPLAIN显示，某个查询的代价从 8488 降到了 329。原始查询是一个涉及one_to_many表的连接，后来被替换为：

```
SELECT right, count(right) FROM
  ( SELECT left, int_array_enum(right) AS right
    FROM summary JOIN (SELECT left FROM left_table WHERE left = item)
 AS lefts
         ON (summary.left = lefts.left)
  ) AS list
  GROUP BY right
  ORDER BY count DESC;
```

F.18. intarray

`intarray`模块提供了一些有用的函数和操作符，用于操作不含 NULL 的整数数组。它还支持使用其中某些操作符进行索引搜索。

如果所提供的数组中包含任何 NULL 元素，所有这些操作都会抛出错误。

这些操作中有许多只对一维数组才有意义。尽管它们也接受更高维度的输入数组，但数据会被视为按存储顺序排列的线性数组。

F.18.1. intarray 函数和操作符

`intarray`模块提供的函数列在表 F.10中，操作符列在表 F.11中。

表 F.10. intarray 函数

Function	Return Type	Description	Example	Result
`icount(int[])`	`int`	number of elements in array	`icount('{1,2,3}'::int[])`	`3`
`sort(int[], text dir)`	`int[]`	sort array – *dir* must be `asc` or `desc`	`sort('{1,2,3}'::int[], 'desc')`	`{3,2,1}`
`sort(int[])`	`int[]`	sort in ascending order	`sort(array[11,77,44])`	`{11,44,77}`
`sort_asc(int[])`	`int[]`	sort in ascending order		
`sort_desc(int[])`	`int[]`	sort in descending order		
`uniq(int[])`	`int[]`	remove adjacent duplicates	`uniq(sort('{1,2,3,2,1}'::int[]))`	`{1,2,3}`
`idx(int[], int item)`	`int`	index of first element matching *item* (0 if none)	`idx(array[11,22,33,22,11], 22)`	`2`
`subarray(int[], int start, int len)`	`int[]`	portion of array starting at position *start*, *len* elements	`subarray('{1,2,3,2,1}'::int[], 2, 3)`	`{2,3,2}`
`subarray(int[], int start)`	`int[]`	portion of array starting at position *start*	`subarray('{1,2,3,2,1}'::int[], 2)`	`{2,3,2,1}`
`intset(int)`	`int[]`	make single-element array	`intset(42)`	`{42}`

表 F.11. intarray 操作符

Operator	Returns	Description
`int[] && int[]`	`boolean`	overlap – `true` if arrays have at least one common element
`int[] @> int[]`	`boolean`	contains – `true` if left array contains right array

Operator	Returns	Description
`int[] <@ int[]`	`boolean`	contained – `true` if left array is contained in right array
`# int[]`	`int`	number of elements in array
`int[] # int`	`int`	index (same as `idx` function)
`int[] + int`	`int[]`	push element onto array (add it to end of array)
`int[] + int[]`	`int[]`	array concatenation (right array added to the end of left one)
`int[] - int`	`int[]`	remove entries matching right argument from array
`int[] - int[]`	`int[]`	remove elements of right array from left
`int[] | int`	`int[]`	union of arguments
`int[] | int[]`	`int[]`	union of arrays
`int[] & int[]`	`int[]`	intersection of arrays
`int[] @@ query_int`	`boolean`	`true` if array satisfies query (see below)
`query_int ~~ int[]`	`boolean`	`true` if array satisfies query (commutator of `@@`)

（在 PostgreSQL 8.2 之前，包含操作符 `@>` 和 `<@` 分别称为 `@` 和 `~`。这些名称仍然可用，但已弃用，最终将被删除。请注意，旧名称与核心几何数据类型以前采用的约定正好相反！）

操作符`&&`、`@>`和 `<@`等价于PostgreSQL内置的同名操作符，不同之处在于它们只适用于不包含 NULL 的整数数组，而内置操作符适用于任何数组类型。这一限制使它们在很多情况下比内置操作符更快。

`@@`和`~~`操作符用于测试数组是否满足某个查询，该查询表示为专用数据类型`query_int` 的一个值。一个查询由若干整数值组成，这些值会与数组元素进行检查，并且可通过操作符`&`（AND）、`|`（OR）和`!`（NOT）组合起来。必要时可以使用括号。例如，查询`1&(2|3)`可匹配包含 1 且还包含 2 或 3 之一的数组。

F.18.2. 索引支持

`intarray`为`&&`、`@>`、`<@`和`@@`操作符以及常规数组相等运算提供索引支持。

提供了两个 GiST 索引操作符类：`gist__int_ops`（默认使用）适用于小型到中型数据集，而`gist__intbig_ops`使用更大的签名，
更适合为大型数据集建立索引（即包含大量不同数组值的列）。实现使用带有内置有损压缩的 RD 树数据结构。

另有一个非默认的 GIN 操作符类`gin__int_ops`，支持相同的操作符。

作为经验法则，GIN 索引的搜索速度快于 GiST 索引，但构建或更新较慢；因此 GIN 更适合静态数据，而 GiST 更适合经常更新的数据。

F.18.3. 示例

```
-- a message can be in one or more "sections"
```

```
CREATE TABLE message (mid INT PRIMARY KEY, sections INT[], ...);

-- create specialized index
CREATE INDEX message_rdtree_idx ON message USING GIST (sections gist__
int_ops);

-- select messages in section 1 OR 2 - OVERLAP operator
SELECT message.mid FROM message WHERE message.sections && '{1,2}';

-- select messages in sections 1 AND 2 - CONTAINS operator
SELECT message.mid FROM message WHERE message.sections @> '{1,2}';

-- the same, using QUERY operator
SELECT message.mid FROM message WHERE message.sections @@ '1&2'::
query_int;
```

F.18.4. 基准测试

源代码目录`contrib/intarray/bench`包含一个基准测试套件。运行方式如下：

```
cd .../bench
createdb TEST
psql TEST < ../_int.sql
./create_test.pl | psql TEST
./bench.pl
```

`bench.pl`脚本有很多选项，在不带任何参数运行时会显示这些选项。

F.18.5. 作者

所有工作都由 Teodor Sigaev（`<teodor@sigaev.ru>`）和 Oleg Bartunov（`<oleg@sai.msu.su>`）完成。更多信息请见http://www.sai.msu.su/~megera/postgres/gist/。Andrey Oktyabrski 在添加新函数和新操作方面也做出了很大贡献。

F.19. isn

`isn`模块为以下国际产品编号标准提供数据类型：EAN13、UPC、ISBN（图书）、ISMN（音乐）和 ISSN（连续出版物）。输入这些编号时，会依据一份硬编码的前缀列表进行校验；输出时，这份前缀列表也会用于给编号加上连字符。由于新的前缀会不时被分配，
这份前缀列表可能已经过时。希望该模块未来版本能从一个或多个表中获取前缀列表，
以便用户按需更新；但目前，这份列表只能通过修改源代码并重新编译来更新。或者，
该模块未来版本也可能取消前缀校验和连字符支持。

F.19.1. 数据类型

表 F.12显示了`isn`模块提供的数据类型。

表 F.12. `isn` Data Types

Data Type	Description
`EAN13`	European Article Numbers, always displayed in the EAN13 display format
`ISBN13`	International Standard Book Numbers to be displayed in the new EAN13 display format
`ISMN13`	International Standard Music Numbers to be displayed in the new EAN13 display format
`ISSN13`	International Standard Serial Numbers to be displayed in the new EAN13 display format
`ISBN`	International Standard Book Numbers to be displayed in the old short display format
`ISMN`	International Standard Music Numbers to be displayed in the old short display format
`ISSN`	International Standard Serial Numbers to be displayed in the old short display format
`UPC`	Universal Product Codes

几点说明：

1. ISBN13、ISMN13 和 ISSN13 编号都是 EAN13 编号。
2. EAN13 编号并不总是 ISBN13、ISMN13 或 ISSN13（不过其中有些确实是）。
3. 某些 ISBN13 编号可以显示为 ISBN。
4. 某些 ISMN13 编号可以显示为 ISMN。
5. 某些 ISSN13 编号可以显示为 ISSN。
6. UPC 编号是 EAN13 编号的一个子集（它们基本上就是去掉首位 `0` 的 EAN13 编号）。
7. 所有 UPC、ISBN、ISMN 和 ISSN 编号都可以表示为 EAN13 编号。

在内部，所有这些类型都使用同一种表示形式（一个 64 位整数），并且彼此可以互换。之所以提供多种类型，是为了控制显示格式，
并对原本应表示某一特定类型编号的输入执行更严格的有效性检查。

只要可能，ISBN、ISMN 和 ISSN 类型都会显示编号的短版本（ISxN 10）；对于不能用短版本表示的编号，则显示为 ISxN 13 格式。EAN13、ISBN13、ISMN13 和 ISSN13 类型则始终显示 ISxN 的长版本（EAN13）。

F.19.2. 类型转换

isn模块提供以下几对类型转换：

- ISBN13 <=> EAN13
- ISMN13 <=> EAN13
- ISSN13 <=> EAN13
- ISBN <=> EAN13
- ISMN <=> EAN13
- ISSN <=> EAN13
- UPC <=> EAN13
- ISBN <=> ISBN13
- ISMN <=> ISMN13
- ISSN <=> ISSN13

当从EAN13转换到其他类型时，会在运行时检查该值是否落在另一类型的取值范围内；
如果不在，就会抛出错误。其他类型转换都只是简单地重新标记，因此总会成功。

F.19.3. 函数和操作符

isn模块提供标准比较操作符，以及对所有这些数据类型的 B-树和哈希索引支持。此外，它还提供了一些专用函数，如表 F.13所示。在该表中，isn表示该模块提供的任意一种数据类型。

表 F.13. isn Functions

Function	Returns	Description
isn_weak(boolean)	boolean	Sets the weak input mode (returns new setting)
isn_weak()	boolean	Gets the current status of the weak mode
make_valid(isn)	isn	Validates an invalid number (clears the invalid flag)
is_valid(isn)	boolean	Checks for the presence of the invalid flag

弱模式用于允许向表中插入无效数据。这里的“无效”指的是校验位错误，而不是缺少数字。

为什么会需要使用弱模式呢？例如，手头可能有一大批 ISBN 编号，数量多到难免会有一些编号因为某些奇怪的原因而带有错误的校验位（也许这些编号是从印刷清单扫描得到的，而 OCR 把数字识别错了；也许这些编号是人工录入的.....谁知道呢）。总之，
可能想把这些混乱情况清理干净，但同时仍希望先把所有编号都装入数据库，
并借助外部工具在数据库中定位无效编号，以便核对信息并更容易完成校验；例如，
可能会想把表中所有无效编号都查询出来。

当在弱模式下向表中插入无效编号时，实际插入的是校验位已更正的编号，但显示时会在末尾附加一个感叹号（!），例如0-11-000322-5!。可以用is_valid函数检查这个无效标记，并用make_valid函数清除它。

即使未启用弱模式，也可以通过在编号末尾附加!字符来强制插入无效编号。

另一个特殊功能是，在输入时可以用?代替校验位，系统会自动插入正确的校验位。

F.19.4. 示例

```
--Using the types directly:
SELECT isbn('978-0-393-04002-9');
SELECT isbn13('0901690546');
SELECT issn('1436-4522');

--Casting types:
-- note that you can only cast from ean13 to another type when the
-- number would be valid in the realm of the target type;
-- thus, the following will NOT work: select isbn(ean13(
'0220356483481'));
-- but these will:
SELECT upc(ean13('0220356483481'));
SELECT ean13(upc('220356483481'));

--Create a table with a single column to hold ISBN numbers:
CREATE TABLE test (id isbn);
INSERT INTO test VALUES('9780393040029');

--Automatically calculate check digits (observe the '?'):
INSERT INTO test VALUES('220500896?');
INSERT INTO test VALUES('978055215372?');

SELECT issn('3251231?');
SELECT ismn('979047213542?');

--Using the weak mode:
SELECT isn_weak(true);
INSERT INTO test VALUES('978-0-11-000533-4');
INSERT INTO test VALUES('9780141219307');
INSERT INTO test VALUES('2-205-00876-X');
SELECT isn_weak(false);

SELECT id FROM test WHERE NOT is_valid(id);
UPDATE test SET id = make_valid(id) WHERE id = '2-205-00876-X!';

SELECT * FROM test;

SELECT isbn13(id) FROM test;
```

F.19.5. 参考文献

实现本模块所需的信息收集自多个站点，包括：

- http://www.isbn-international.org/

- http://www.issn.org/
- http://www.ismn-international.org/
- http://www.wikipedia.org/

The prefixes used for hyphenation were also compiled from:

- http://www.gs1.org/productssolutions/idkeys/support/prefix_list.html
- http://www.isbn-international.org/en/identifiers.html
- http://www.ismn-international.org/ranges.html

Care was taken during the creation of the algorithms and they were meticulously verified against the suggested algorithms in the official ISBN, ISMN, ISSN User Manuals.

F.19.6. 作者

Germán Méndez Bravo (Kronuz), 2004 - 2006

该模块受到了 Garrett A. Wollman 的`isbn_issn`代码的启发。

F.20. lo – 管理大对象

`lo`模块为管理大对象（也称为 LO 或 BLOB）提供支持，其中包括数据类型`lo`和触发器`lo_manage`。

F.20.1. 原理

JDBC 驱动的一个问题是（ODBC 驱动也受此影响），其规范假定对 BLOB（二进制大对象）的引用存储在表内，并且如果该表项被更改，相关的 BLOB 就会从数据库中删除。

但在目前的PostgreSQL中并非如此。大对象被视为独立对象；表中的一个项可以通过 OID 引用某个大对象，但也可能有多个表项引用同一个大对象 OID，因此系统不会仅因你更改或删除了其中一个表项就删除该大对象。

这对于PostgreSQL专用应用没有问题，但使用 JDBC 或 ODBC 的标准代码不会删除这些对象，从而产生孤立对象，也就是不再被任何内容引用、只是占用磁盘空间的对象。

`lo`模块允许通过把触发器附加到包含 LO 引用列的表上来解决这个问题。该触发器本质上就是在你删除或修改引用大对象的值时调用`lo_unlink`。使用这个触发器时，
你实际上是假定：凡是出现在受该触发器控制的列中的大对象，在数据库中都只有一个引用！

该模块还提供了数据类型`lo`，它实际上只是`oid`类型的一个域。
这有助于区分保存大对象引用的数据库列与保存其他对象 OID 的列。使用该触发器并不要求必须使用`lo`类型，但用它来标识数据库中哪些列表示由该触发器管理的大对象，可能会更方便。
还有传言说，如果 BLOB 列不用`lo`，ODBC 驱动会感到困惑。

F.20.2. 如何使用

下面是一个简单的用法示例：

```
CREATE TABLE image (title TEXT, raster lo);

CREATE TRIGGER t_raster BEFORE UPDATE OR DELETE ON image
    FOR EACH ROW EXECUTE PROCEDURE lo_manage(raster);
```

对于每个将保存指向大对象的唯一引用的列，创建一个 `BEFORE UPDATE OR DELETE` 触发器，并将该列名作为唯一的触发器参数。如果同一张表中需要多个 `lo` 列，就为每一列分别创建一个触发器，并记得为同一张表上的每个触发器指定不同的名称。

F.20.3. 限制

- 删除表时，其中包含的任何对象仍会变成孤立对象，因为这种情况下不会执行触发器。
  为避免这种情况，可以先执行`DELETE FROM *table*`，再执行`DROP TABLE`。

  `TRUNCATE`也有同样的风险。

  如果你已经有，或者怀疑有，孤立的大对象，请参见vacuumlo模块来帮助清理它们。
  偶尔运行vacuumlo，作为对`lo_manage`触发器的补充保障，是个不错的主意。

- 某些前端可能会创建自己的表，但不会创建相应的触发器。此外，
  用户也可能不记得（或根本不知道）要创建这些触发器。

F.20.4. 作者

Peter Mount <peter@retep.org.uk>

F.21. ltree

该模块实现了数据类型 `ltree`，用于表示存储在层次化树状结构中的数据标签。它还提供了丰富的标签树搜索能力。

F.21.1. 定义

标签是由字母数字字符和下划线组成的序列（例如，在 C 区域设置下，允许的字符为 `A-Za-z0-9_`）。标签长度必须少于 256 字节。

示例：`42`、`Personal_Services`

标签路径是由点号分隔的零个或多个标签组成的序列，例如 `L1.L2.L3`，表示从层次树根节点到某个特定节点的一条路径。标签路径的长度必须小于 65Kb，但最好保持在 2Kb 以下。实践中这并不是一个大的限制；例如，DMOZ 目录（http://www.dmoz.org）中最长的标签路径约为 240 字节。

示例：`Top.Countries.Europe.Russia`

`ltree`模块提供了几种数据类型：

- `ltree`存储一个标签路径。
- `lquery`表示一种用于匹配`ltree`值的、类似正则表达式的模式。
  一个简单单词会匹配路径中的相应标签。星号（`*`）匹配零个或多个标签。例如：

`foo`	精确匹配标签路径*foo*
`*.foo.*`	匹配任何包含标签*foo*的标签路径
`*.foo`	匹配最后一个标签为*foo*的任意标签路径

  星号还可以带量词，以限制它们能够匹配的标签数量：

`*{n}`	精确匹配 *n* 个标签
`*{n,}`	至少匹配 *n* 个标签
`*{n,m}`	至少匹配 *n* 个、但不超过 *m* 个标签
`*{,m}`	至多匹配 *m* 个标签 — 与下式相同： `*{0,m}`

  在`lquery`中，有几个修饰符可以放在非星号标签的末尾，使其不只匹配完全相同的标签：

`@`	不区分大小写地匹配，例如 *a@* 可匹配 *A*
`*`	匹配以此前缀开头的任意标签，例如 *foo** 可匹配 *foobar*
`%`	匹配标签起始处由下划线分隔的单词

  修饰符`%`的行为稍微复杂一些。它尝试匹配单词，而不是整个标签。例如，`foo_bar%`可匹配`foo_bar_baz`，但不能匹配`foo_barbaz`。如果与`*`组合使用，则前缀匹配会分别作用于每个单词，例如`foo_bar%*`可匹配`foo1_bar2_baz`，但不能匹配`foo1_br2_baz`。

  此外，还可以写出多个可能带修饰符的非星号项，并用`|`（OR）分隔，以匹配其中任意一项；也可以在非星号组前加上`!`（NOT），以匹配不符合这些备选项中任意一项的标签。

  下面是一个带注释的示例，使用`lquery`：

  ```
  Top.*{0,2}.sport*@.!football|tennis.Russ*|Spain
  ```

```
a.  b.     c.      d.                e.
```

此查询将匹配满足以下条件的任意标签路径：

a. 以标签`Top`开头

b. 接下来，在下一个条件之前有零到两个标签

c. 然后是一个以前缀`sport`开头的标签，且匹配时不区分大小写

d. 接着有一个不匹配`football`或`tennis`的标签

e. 最后以一个以`Russ`开头的标签，或精确匹配`Spain`的标签结束。

- `ltxtquery`表示一种用于匹配`ltree`值的、类似全文检索的模式。一个`ltxtquery`值包含单词，末尾还可以带有修饰符`@`、`*`、`%`；这些修饰符与它们在`lquery`中的含义相同。单词可以通过`&`（AND）、`|`（OR）、`!`（NOT）以及圆括号组合。它与`lquery`的关键区别在于，`ltxtquery`匹配单词时不考虑它们在标签路径中的位置。

  下面是一个`ltxtquery`示例：

  ```
  Europe & Russia*@ & !Transportation
  ```

  它将匹配包含标签`Europe`以及任意以`Russia`开头（不区分大小写）的标签的路径，但不匹配包含标签`Transportation`的路径。这些单词在路径中的位置并不重要。另外，当使用`%`时，该单词可以匹配标签中任意由下划线分隔的单词，而不考虑其位置。

注意：`ltxtquery`允许在符号之间出现空白，而`ltree`和`lquery`不允许。

F.21.2. 操作符和函数

类型`ltree`具有常见的比较操作符 `=`、`<>`、`<`、`>`、`<=`、`>=`。比较时采用树遍历顺序，其中节点的子节点按标签文本排序。此外，还提供了表 F.14中所示的专用操作符。

表 F.14. `ltree` Operators

Operator	Returns	Description
`ltree @> ltree`	`boolean`	is left argument an ancestor of right (or equal)?
`ltree <@ ltree`	`boolean`	is left argument a descendant of right (or equal)?
`ltree ~ lquery`	`boolean`	does `ltree` match `lquery`?
`lquery ~ ltree`	`boolean`	does `ltree` match `lquery`?
`ltree ? lquery[]`	`boolean`	does `ltree` match any `lquery` in array?
`lquery[] ? ltree`	`boolean`	does `ltree` match any `lquery` in array?
`ltree @ ltxtquery`	`boolean`	does `ltree` match `ltxtquery`?
`ltxtquery @ ltree`	`boolean`	does `ltree` match `ltxtquery`?
`ltree || ltree`	`ltree`	concatenate `ltree` paths
`ltree || text`	`ltree`	convert text to `ltree` and concatenate

Operator	Returns	Description
`text || ltree`	`ltree`	convert text to `ltree` and concatenate
`ltree[] @> ltree`	`boolean`	does array contain an ancestor of `ltree`?
`ltree <@ ltree[]`	`boolean`	does array contain an ancestor of `ltree`?
`ltree[] <@ ltree`	`boolean`	does array contain a descen dant of `ltree`?
`ltree @> ltree[]`	`boolean`	does array contain a descen dant of `ltree`?
`ltree[] ~ lquery`	`boolean`	does array contain any path matching `lquery`?
`lquery ~ ltree[]`	`boolean`	does array contain any path matching `lquery`?
`ltree[] ? lquery[]`	`boolean`	does `ltree` array contain any path matching any `lquery`?
`lquery[] ? ltree[]`	`boolean`	does `ltree` array contain any path matching any `lquery`?
`ltree[] @ ltxtquery`	`boolean`	does array contain any path matching `ltxtquery`?
`ltxtquery @ ltree[]`	`boolean`	does array contain any path matching `ltxtquery`?
`ltree[] ?@> ltree`	`ltree`	first array entry that is an ancestor of `ltree`; NULL if none
`ltree[] ?<@ ltree`	`ltree`	first array entry that is a descen dant of `ltree`; NULL if none
`ltree[] ?~ lquery`	`ltree`	first array entry that matches `lquery`; NULL if none
`ltree[] ?@ ltxtquery`	`ltree`	first array entry that matches `ltxtquery`; NULL if none

操作符`<@`、`@>`、`@`和`~`都有对应的 `^<@`、`^@>`、`^@`、`^~`变体，它们除了不使用索引之外完全相同。这些变体仅对测试有用。

可用函数见表 F.15。

表 F.15. `ltree` Functions

Function	Return Type	Description	Example	Result
`subltree(ltree, int start, int end)`	`ltree`	subpath of `ltree` from position *`start`* to position *`end`*-1 (counting from 0)	`subltree('Top.Child1.Child2',1,2)`	`Child1`
`subpath(ltree, int offset, int len)`	`ltree`	subpath of `ltree` starting at position *`offset`*, length *`len`*. If *`offset`* is negati	`subpath('Top.Child1.Child2',0,2)`	`Top.Child1`

Function	Return Type	Description	Example	Result
		ve, subpath starts that far from the end of the path. If *len* is negative, leaves that many labels off the end of the path.		
subpath(ltree, int offset)	ltree	subpath of ltree starting at position *offset*, extending to end of path. If *offset* is negative, subpath starts that far from the end of the path.	subpath('Top.Child1.Child2',1)	Child1.Child2
nlevel(ltree)	integer	number of labels in path	nlevel('Top.Child1.Child2')	3
index(ltree a, ltree b)	integer	position of first occurrence of *b* in *a*; -1 if not found	index('0.1.2.3.5.4.5.6.8.5.6.8','5.6')	6
index(ltree a, ltree b, int offset)	integer	position of first occurrence of *b* in *a*, searching starting at *offset*; negative *offset* means start - *offset* labels from the end of the path	index('0.1.2.3.5.4.5.6.8.5.6.8','5.6',-4)	9
text2ltree(text)	ltree	cast text to ltree		
ltree2text(ltree)	text	cast ltree to text		
lca(ltree, ltree, ...)	ltree	lowest common ancestor, i.e., longest common prefix of paths (up to 8 arguments supported)	lca('1.2.2.3','1.2.3.4.5.6')	1.2
lca(ltree[])	ltree	lowest common ancestor, i.e., longest common prefix of paths	lca(array['1.2.2.3'::ltree,'1.2.3'])	1.2

F.21.3. 索引

ltree支持几种能够加速所示操作符的索引类型：

- ltree上的 B-树索引：<、<=、=、>=、>
- ltree上的 GiST 索引：<、<=、=、>=、>、@>、<@、@、~、?

创建此类索引的示例：

```
CREATE INDEX path_gist_idx ON test USING GIST (path);
```

- `ltree[]`上的 GiST 索引：`ltree[] <@ ltree`、`ltree @> ltree[]`、`@`、`~`、`?`

  创建此类索引的示例：

  ```
  CREATE INDEX path_gist_idx ON test USING GIST (array_path);
  ```

  注意：这种索引类型是有损的。

F.21.4. 示例

本示例使用下列数据（在源代码发行包中的 `contrib/ltree/ltreetest.sql`文件里也能找到）：

```
CREATE TABLE test (path ltree);
INSERT INTO test VALUES ('Top');
INSERT INTO test VALUES ('Top.Science');
INSERT INTO test VALUES ('Top.Science.Astronomy');
INSERT INTO test VALUES ('Top.Science.Astronomy.Astrophysics');
INSERT INTO test VALUES ('Top.Science.Astronomy.Cosmology');
INSERT INTO test VALUES ('Top.Hobbies');
INSERT INTO test VALUES ('Top.Hobbies.Amateurs_Astronomy');
INSERT INTO test VALUES ('Top.Collections');
INSERT INTO test VALUES ('Top.Collections.Pictures');
INSERT INTO test VALUES ('Top.Collections.Pictures.Astronomy');
INSERT INTO test VALUES ('Top.Collections.Pictures.Astronomy.Stars');
INSERT INTO test VALUES ('Top.Collections.Pictures.Astronomy.
Galaxies');
INSERT INTO test VALUES ('Top.Collections.Pictures.Astronomy.
Astronauts');
CREATE INDEX path_gist_idx ON test USING gist(path);
CREATE INDEX path_idx ON test USING btree(path);
```

现在，我们有一个表`test`，其中的数据描述了下图所示的层次结构：

```
                        Top
                      /   |  \
             Science Hobbies Collections
                 /       |              \
        Astronomy   Amateurs_Astronomy Pictures
           /  \                            |
Astrophysics  Cosmology                Astronomy
                                        /  |    \
                                 Galaxies Stars Astronauts
```

我们可以做继承查询：

```
ltreetest=> SELECT path FROM test WHERE path <@ 'Top.Science';
                path
------------------------------------
 Top.Science
```

```
 Top.Science.Astronomy
 Top.Science.Astronomy.Astrophysics
 Top.Science.Astronomy.Cosmology
(4 rows)
```

下面是一些路径匹配的示例：

```
ltreetest=> SELECT path FROM test WHERE path ~ '*.Astronomy.*';
                     path
----------------------------------------------
 Top.Science.Astronomy
 Top.Science.Astronomy.Astrophysics
 Top.Science.Astronomy.Cosmology
 Top.Collections.Pictures.Astronomy
 Top.Collections.Pictures.Astronomy.Stars
 Top.Collections.Pictures.Astronomy.Galaxies
 Top.Collections.Pictures.Astronomy.Astronauts
(7 rows)

ltreetest=> SELECT path FROM test WHERE path ~ '*.!pictures@.*.
Astronomy.*';
                path
-------------------------------------
 Top.Science.Astronomy
 Top.Science.Astronomy.Astrophysics
 Top.Science.Astronomy.Cosmology
(3 rows)
```

下面是一些全文检索示例：

```
ltreetest=> SELECT path FROM test WHERE path @ 'Astro*% & !pictures@';
                path
-------------------------------------
 Top.Science.Astronomy
 Top.Science.Astronomy.Astrophysics
 Top.Science.Astronomy.Cosmology
 Top.Hobbies.Amateurs_Astronomy
(4 rows)

ltreetest=> SELECT path FROM test WHERE path @ 'Astro* & !pictures@';
                path
-------------------------------------
 Top.Science.Astronomy
 Top.Science.Astronomy.Astrophysics
 Top.Science.Astronomy.Cosmology
(3 rows)
```

使用函数构造路径：

```
ltreetest=> SELECT subpath(path,0,2)||'Space'||subpath(path,2) FROM
 test WHERE path <@ 'Top.Science.Astronomy';
                 ?column?
-------------------------------------------
 Top.Science.Space.Astronomy
 Top.Science.Space.Astronomy.Astrophysics
```

```
 Top.Science.Space.Astronomy.Cosmology
(3 rows)
```

可以通过创建一个 SQL 函数，在路径的指定位置插入标签来简化这一操作：

```
CREATE FUNCTION ins_label(ltree, int, text) RETURNS ltree
    AS 'select subpath($1,0,$2) || $3 || subpath($1,$2);'
    LANGUAGE SQL IMMUTABLE;

ltreetest=> SELECT ins_label(path,2,'Space') FROM test WHERE path <@
 'Top.Science.Astronomy';
                ins_label
-------------------------------------------
 Top.Science.Space.Astronomy
 Top.Science.Space.Astronomy.Astrophysics
 Top.Science.Space.Astronomy.Cosmology
(3 rows)
```

F.21.5. 作者

全部工作均由 Teodor Sigaev（`<teodor@stack.net>`）和 Oleg Bartunov（`<oleg@sai.msu.su>`）完成。更多信息见http://www.sai.msu.su/~megera/postgres/gist/。作者谨感谢 Eugeny Rodichev 的有益讨论。欢迎提出意见和缺陷报告。

F.22. oid2name

oid2name 是一个帮助管理员检查 PostgreSQL 所用文件结构的实用程序。要使用它，你需要熟悉数据库文件结构，这在第 55 章中有说明。

> **注意**
>
> “oid2name” 这个名称是历史遗留，实际上颇具误导性，
> 因为多数时候你真正关心的是表的 filenode 编号（也就是数据库目录中可见的文件名）。
> 务必理解表 OID 与表 filenode 之间的区别。

F.22.1. Overview

oid2name 连接到目标数据库，并提取 OID、filenode 和/或表名信息。你也可以让它显示数据库 OID 或表空间 OID。

F.22.2. oid2name Options

oid2name accepts the following command-line arguments:

`-o` *`oid`*

显示 OID 为 *`oid`* 的表的信息

`-f` *`filenode`*

显示文件节点为 *`filenode`* 的表的信息

`-t` *`tablename_pattern`*

显示与 *`tablename_pattern`* 匹配的表的信息

`-s`

显示表空间 OID

`-S`

包含系统对象（位于`information_schema`、`pg_toast`和`pg_catalog`模式中的对象）

`-i`

在列表中包含索引和序列

`-x`

显示每个对象的更多信息：表空间名、模式名和 OID

`-q`

省略标题（适用于脚本）

-d *database*

要连接的数据库

-H *host*

数据库服务器的主机

-p *port*

数据库服务器的端口

-U *username*

用于连接的用户名

-P *password*

密码（已弃用 – 将其放在命令行上会带来安全风险）

若要显示特定的表，可使用-o、-f和/或-t来选择要显示哪些表。-o接受一个 OID，-f接受一个 filenode，而-t接受一个表名（实际上是一个LIKE模式，因此你可以使用诸如foo%之类的形式）。这些选项可以按需重复使用，最终的列表将包含匹配任意一个选项的所有对象。
但请注意，这些选项只能显示由-d指定的数据库中的对象。

如果没有给出-o、-f或-t，但给出了-d，那么它将列出-d指定数据库中的所有表。
在这种模式下，-S和-i选项控制要列出哪些对象。

如果连-d也没有给出，它将显示数据库 OID 列表。另外，你也可以给出-s来获取表空间列表。

F.22.3. 示例

```
$ # what's in this database server, anyway?
$ oid2name
All databases:
    Oid  Database Name  Tablespace
----------------------------------
  17228       alvherre  pg_default
  17255     regression  pg_default
  17227      template0  pg_default
      1      template1  pg_default

$ oid2name -s
All tablespaces:
     Oid  Tablespace Name
-------------------------
    1663       pg_default
    1664        pg_global
  155151         fastdisk
  155152          bigdisk

$ # OK, let's look into database alvherre
$ cd $PGDATA/base/17228

$ # get top 10 db objects in the default tablespace, ordered by size
```

```
$ ls -lS * | head -10
-rw-------  1 alvherre alvherre 136536064 sep 14 09:51 155173
-rw-------  1 alvherre alvherre  17965056 sep 14 09:51 1155291
-rw-------  1 alvherre alvherre   1204224 sep 14 09:51 16717
-rw-------  1 alvherre alvherre    581632 sep  6 17:51 1255
-rw-------  1 alvherre alvherre    237568 sep 14 09:50 16674
-rw-------  1 alvherre alvherre    212992 sep 14 09:51 1249
-rw-------  1 alvherre alvherre    204800 sep 14 09:51 16684
-rw-------  1 alvherre alvherre    196608 sep 14 09:50 16700
-rw-------  1 alvherre alvherre    163840 sep 14 09:50 16699
-rw-------  1 alvherre alvherre    122880 sep  6 17:51 16751

$ # I wonder what file 155173 is ...
$ oid2name -d alvherre -f 155173
From database "alvherre":
  Filenode  Table Name
----------------------
    155173    accounts

$ # you can ask for more than one object
$ oid2name -d alvherre -f 155173 -f 1155291
From database "alvherre":
  Filenode     Table Name
-------------------------
    155173       accounts
   1155291  accounts_pkey

$ # you can mix the options, and get more details with -x
$ oid2name -d alvherre -t accounts -f 1155291 -x
From database "alvherre":
  Filenode     Table Name      Oid  Schema  Tablespace
--------------------------------------------------------
    155173       accounts   155173  public  pg_default
   1155291  accounts_pkey  1155291  public  pg_default

$ # show disk space for every db object
$ du [0-9]* |
> while read SIZE FILENODE
> do
>   echo "$SIZE       `oid2name -q -d alvherre -i -f $FILENODE`"
> done
16            1155287  branches_pkey
16            1155289  tellers_pkey
17561            1155291  accounts_pkey
...

$ # same, but sort by size
$ du [0-9]* | sort -rn | while read SIZE FN
> do
>   echo "$SIZE   `oid2name -q -d alvherre -f $FN`"
> done
133466             155173    accounts
17561            1155291  accounts_pkey
1177              16717  pg_proc_proname_args_nsp_index
```

```
...

$ # If you want to see what's in tablespaces, use the pg_tblspc
 directory
$ cd $PGDATA/pg_tblspc
$ oid2name -s
All tablespaces:
     Oid  Tablespace Name
-------------------------
    1663       pg_default
    1664        pg_global
  155151         fastdisk
  155152          bigdisk

$ # what databases have objects in tablespace "fastdisk"?
$ ls -d 155151/*
155151/17228/  155151/PG_VERSION

$ # Oh, what was database 17228 again?
$ oid2name
All databases:
    Oid  Database Name  Tablespace
----------------------------------
  17228       alvherre  pg_default
  17255     regression  pg_default
  17227      template0  pg_default
      1      template1  pg_default

$ # Let's see what objects does this database have in the tablespace.
$ cd 155151/17228
$ ls -l
total 0
-rw-------  1 postgres postgres 0 sep 13 23:20 155156

$ # OK, this is a pretty small table ... but which one is it?
$ oid2name -d alvherre -f 155156
From database "alvherre":
  Filenode  Table Name
----------------------
    155156         foo
```

F.22.4. Limitations

oid2name要求数据库服务器正在运行，且系统目录未损坏。因此，它对于从数据库灾难性损坏中恢复的帮助非常有限。

F.22.5. 作者

B. Palmer <bpalmer@crimelabs.net>

F.23. pageinspect

pageinspect模块提供了一组函数，允许你在底层检查数据库页的内容，这对调试很有帮助。所有这些函数都只能由超级用户使用。

F.23.1. 函数

get_raw_page(relname text, fork text, blkno int) returns bytea

get_raw_page读取指定关系的指定块，并以bytea值的形式返回其副本。这样就能获得该块在单一时刻的一致副本。*fork*应当为'main'（主数据分支）、'fsm'（空闲空间映射）、'vm'（可见性映射）或'init'（初始化分支）。

get_raw_page(relname text, blkno int) returns bytea

这是get_raw_page的简写形式，用于从主分支读取。等价于 get_raw_page(relname, 'main', blkno)

page_header(page bytea) returns record

page_header显示PostgreSQL所有堆页和索引页共有的字段。

应将用 get_raw_page 获得的页映像作为参数传入。例如：

```
test=# SELECT * FROM page_header(get_raw_page('pg_class', 0));
    lsn    | tli | flags | lower | upper | special | pagesize |
 version | prune_xid
-----------+-----+-------+-------+-------+---------+----------+----
-----+-----------
 0/24A1B50 |   1 |     1 |   232 |   368 |    8192 |     8192 |
   4 |         0
```

返回的列对应于 PageHeaderData 结构体中的字段。详细信息见 src/include/storage/bufpage.h。

heap_page_items(page bytea) returns setof record

heap_page_items显示堆页上的所有行指针。对于正在使用的行指针，还会显示元组头。无论这些元组在复制原始页时是否对某个 MVCC 快照可见，都会显示出来。

应将用 get_raw_page 获得的堆页映像作为参数传入。例如：

```
test=# SELECT * FROM heap_page_items(get_raw_page('pg_class', 0));
```

关于返回字段的解释，见 src/include/storage/itemid.h 和 src/include/access/htup.h。

bt_metap(relname text) returns record

bt_metap返回 B-树索引元页的信息。例如：

```
test=# SELECT * FROM bt_metap('pg_cast_oid_index');
-[ RECORD 1 ]-----
magic     | 340322
version   | 2
```

```
root      | 1
level     | 0
fastroot  | 1
fastlevel | 0
```

`bt_page_stats(relname text, blkno int) returns record`

`bt_page_stats`返回 B-树索引单个页面的概要信息。例如：

```
test=# SELECT * FROM bt_page_stats('pg_cast_oid_index', 1);
-[ RECORD 1 ]-+-----
blkno         | 1
type          | l
live_items    | 256
dead_items    | 0
avg_item_size | 12
page_size     | 8192
free_size     | 4056
btpo_prev     | 0
btpo_next     | 0
btpo          | 0
btpo_flags    | 3
```

`bt_page_items(relname text, blkno int) returns setof record`

`bt_page_items`返回 B-树索引页中所有项的详细信息。例如：

```
test=# SELECT * FROM bt_page_items('pg_cast_oid_index', 1);
 itemoffset |  ctid   | itemlen | nulls | vars |    data
------------+---------+---------+-------+------+-------------
          1 | (0,1)   |      12 | f     | f    | 23 27 00 00
          2 | (0,2)   |      12 | f     | f    | 24 27 00 00
          3 | (0,3)   |      12 | f     | f    | 25 27 00 00
          4 | (0,4)   |      12 | f     | f    | 26 27 00 00
          5 | (0,5)   |      12 | f     | f    | 27 27 00 00
          6 | (0,6)   |      12 | f     | f    | 28 27 00 00
          7 | (0,7)   |      12 | f     | f    | 29 27 00 00
          8 | (0,8)   |      12 | f     | f    | 2a 27 00 00
```

`fsm_page_contents(page bytea) returns text`

`fsm_page_contents`显示 FSM 页的内部节点结构。输出是一个多行字符串，页内二叉树中的每个节点对应一行。只打印非零节点。还会打印所谓的 “next” 指针，
它指向下一个将从该页返回的槽位。

参见`src/backend/storage/freespace/README`，了解 FSM 页面结构的更多信息。

F.24. passwordcheck

每当通过CREATE ROLE或ALTER ROLE设置用户密码时，`passwordcheck` 模块都会检查密码强度。如果密码被认为过弱，就会被拒绝，该命令也会因错误而终止。

要启用该模块，请将 `'$libdir/passwordcheck'` 加入 `postgresql.conf` 中的shared_preload_libraries，然后重启服务器。

可以通过修改源代码使该模块适应你的需求。例如，可以使用 CrackLib[2] 来检查密码，这只需要在 `Makefile` 中取消两行注释并重新构建该模块。（由于许可原因，我们不能默认包含CrackLib。）如果不使用 CrackLib，该模块会对密码强度施加一些简单规则，而这些规则也可以按需修改或扩展。

小心

为了防止未加密的密码通过网络传输、被写入服务器日志，
或以其他方式被数据库管理员窃取，PostgreSQL允许用户提供预加密密码。
很多客户端程序都会利用这一能力，在将密码发送给服务器之前先行加密。

这就限制了 `passwordcheck` 模块的作用，因为在这种情况下，它只能尝试猜测密码。因此，如果安全要求很高，并不建议使用 `passwordcheck`。与其依赖数据库内部的密码，不如使用 Kerberos 之类的外部认证方法（见第 19 章），那样更安全。

另外，也可以修改 `passwordcheck` 以拒绝预加密密码，但强迫用户以明文设置密码本身也会带来安全风险。

[2] http://sourceforge.net/projects/cracklib/

F.25. pg_archivecleanup

pg_archivecleanup设计为在服务器作为备库运行时用作 `archive_cleanup_command`，以清理 WAL 文件归档（见第 25.2 节）。pg_archivecleanup也可作为独立程序使用，以清理 WAL 文件归档。

pg_archivecleanup features include:

- 用 C 语言编写，因此移植性很好，易于安装
- 源代码易于修改，并且专门划分出了若干区域，供你按自身需要修改

F.25.1. Usage

要配置备库使用 pg_archivecleanup，请将以下内容放入其 `recovery.conf` 配置文件中：

```
archive_cleanup_command = 'pg_archivecleanup archivelocation %r'
```

其中，*`archivelocation`* 是应从中移除 WAL 段文件的目录。

当在archive_cleanup_command中使用时，逻辑上早于`%r`参数值的所有 WAL 文件都会从*`archivelocation`*中移除。这样既能将需要保留的文件数量降到最低，
又能保留崩溃后重启能力。如果*`archivelocation`*只是该特定备库的临时暂存区，
那么使用此参数是合适的；但以下情况下则不适合使用：*`archivelocation`*打算作为长期 WAL 归档区域，或者多个备库正在从同一归档位置恢复。

The full syntax of pg_archivecleanup's command line is

```
pg_archivecleanup [ option ... ] archivelocation restartwalfile
```

作为独立程序使用时，逻辑上早于 `restartwalfile` 的所有 WAL 文件都会从 *`archivelocation`* 中移除。在此模式下，如果你指定的是 `.backup` 文件名，则只会使用文件名前缀作为 `restartwalfile`。这允许你无误地删除某个特定基础备份之前归档的所有 WAL 文件。例如，下例会移除所有早于 WAL 文件名 `000000010000003700000010` 的文件：

```
pg_archivecleanup -d archive 000000010000003700000010.00000020.backup

pg_archivecleanup:  keep WAL file "archive/000000010000003700000010"
 and later
pg_archivecleanup:  removing file "archive/00000001000000370000000F"
pg_archivecleanup:  removing file "archive/00000001000000370000000E"
```

pg_archivecleanup 假定 *`archivelocation`* 是一个服务器所属用户可读写的目录。

F.25.2. pg_archivecleanup 选项

pg_archivecleanup 接受以下命令行参数：

`-d`

在 `stderr` 上输出大量调试日志。

F.25.3. 示例

在 Linux 或 Unix 系统上，你可以这样使用：

```
archive_cleanup_command = 'pg_archivecleanup -d /mnt/standby/archive
 %r 2>>cleanup.log'
```

其中归档目录实际上位于备库上，因此 `archive_command` 通过 NFS 访问它，但这些文件对备库来说是本地的。这将会：

- 在 `cleanup.log` 中产生调试输出
- 从归档目录中移除不再需要的文件

F.25.4. Supported Server Versions

pg_archivecleanup 设计为在作为独立工具使用时与 PostgreSQL 8.0 及更高版本配合工作，或在作为归档清理命令使用时与 PostgreSQL 9.0 及更高版本配合工作。

F.25.5. 作者

Simon Riggs <simon@2ndquadrant.com>

F.26. pgbench

pgbench是一个用于对PostgreSQL执行基准测试的简单程序。它会反复执行同一组 SQL 命令，也可在多个并发数据库会话中运行，然后计算平均事务速率（每秒事务数）。默认情况下，pgbench测试的是一个大体上基于 TPC-B 的场景，每个事务包含五条SELECT、UPDATE和INSERT命令。不过，通过编写自己的事务脚本文件，也很容易测试其他场景。

下面是 pgbench 的典型输出：

```
transaction type: TPC-B (sort of)
scaling factor: 10
query mode: simple
number of clients: 10
number of threads: 1
number of transactions per client: 1000
number of transactions actually processed: 10000/10000
tps = 85.184871 (including connections establishing)
tps = 85.296346 (excluding connections establishing)
```

前六行报告了一些最重要的参数设置。
下一行报告已完成的事务数和预期的事务数（后者就是客户端数与每个客户端的事务数的乘积）；除非运行在完成前失败，否则这两个数应该相等。（在 -T 模式下，只打印实际的事务数。）最后两行报告每秒事务数，分别计入和不计入启动数据库会话的时间。

F.26.1. 概述

默认的类 TPC-B 事务测试要求预先建立特定的表。应使用 -i（初始化）选项调用 pgbench 来创建并填充这些表。（测试自定义脚本时不需要这一步，但需要自行完成测试所需的准备工作。）初始化命令如下：

```
pgbench -i [ other-options ] dbname
```

其中 *dbname* 是要在其中进行测试的已创建数据库的名称。（可能还需要 -h、-p 和/或 -U 选项来指定如何连接数据库服务器。）

> **小心**
>
> pgbench -i会创建四个表pgbench_accounts、pgbench_branches、pgbench_history和pgbench_tellers，并销毁任何已存在的同名表。
> 如果数据库中已经存在这些名称的表，请务必改用其他数据库！

At the default “scale factor” of 1, the tables initially contain this many rows:

```
table                    # of rows
---------------------------------
pgbench_branches         1
pgbench_tellers          10
pgbench_accounts         100000
pgbench_history          0
```

你可以（而且对于大多数用途来说，也许应该）使用 -s（缩放因子）选项增加行数。此时还可以使用 -F（填充因子）选项。

完成必要的设置之后，就可以用不带 `-i` 的命令运行基准测试，即

`pgbench [ options ] dbname`

几乎在所有情况下，都需要一些选项才能做出有意义的测试。最重要的选项是 `-c`（客户端数量）、`-t`（事务数量）、`-T`（时间限制）和 `-f`（指定自定义脚本文件）。完整列表见下文。

第 F.26.2 节给出数据库初始化期间使用的选项，第 F.26.3 节给出运行基准测试期间使用的选项，而第 F.26.4 节给出两种情况下都有用的选项。

F.26.2. pgbench 初始化选项

pgbench 接受以下命令行初始化参数：

`-i`

进入初始化模式所必需。

`-F fillfactor`

以给定的 fillfactor 创建`pgbench_accounts`、`pgbench_tellers`和 `pgbench_branches`表。默认值为 100。

`-s scale_factor`

将生成的行数乘以缩放因子。例如，`-s 100` 会在 `pgbench_accounts` 表中创建 10,000,000 行。默认值为 1。

F.26.3. pgbench Benchmarking Options

pgbench 接受以下命令行基准测试参数：

`-c clients`

模拟的客户端数量，也就是并发数据库会话的数量。默认值为 1。

`-C`

为每个事务建立一个新的连接，而不是每个客户端会话只建立一次。
这有助于测量连接开销。

`-d`

打印调试输出。

`-D varname=value`

为自定义脚本定义一个变量（见下文）。允许使用多个 `-D` 选项。

`-f filename`

从 *filename* 读取事务脚本。详见下文。`-N`、`-S` 和 `-f` 互斥。

`-j threads`

pgbench 内的工作线程数。在多 CPU 机器上使用多个线程会有帮助。客户端数必须时线程数的整数倍，因为每个线程都会分到相同数量的客户端会话来管理。默认值为 1。

`-l`

将每个事务所耗时间写入日志文件。详见下文。

`-M` *`querymode`*

用于向服务器提交查询的协议：

- `simple`：使用简单查询协议。
- `extended`：使用扩展查询协议。
- `prepared`：使用带预备语句的扩展查询协议。

The default is simple query protocol. (See 第 46 章 for more information.)

`-n`

运行测试前不执行清理。如果你运行的自定义测试场景不包含标准表 `pgbench_accounts`、`pgbench_branches`、`pgbench_history` 和 `pgbench_tellers`，则此选项是必需的。

`-N`

不更新 `pgbench_tellers` 和 `pgbench_branches`。这会避免在这些表上发生更新争用，但也会让该测试场景更不像 TPC-B。

`-r`

在基准测试完成后，报告每条命令的平均语句延迟（从客户端视角看到的执行时间）。详情见下文。

`-s` *`scale_factor`*

在 pgbench 的输出中报告指定的缩放因子。对于内置测试，这不是必需的；正确的缩放因子会通过统计 `pgbench_branches` 表中的行数检测出来。但是，在测试自定义基准（`-f` 选项）时，除非使用此选项，否则缩放因子将报告为 1。

`-S`

执行只读 SELECT 事务，而不是类 TPC-B 测试。

`-t` *`transactions`*

每个客户端运行的事务数。默认值为 10。

`-T` *`seconds`*

让测试运行这么多秒，而不是每个客户端固定数量的事务。`-t` 和 `-T` 互斥。

`-v`

运行测试前清理全部四个标准表。若既不指定 `-n` 也不指定 `-v`，pgbench 将清理 `pgbench_tellers` 和 `pgbench_branches` 表，并截断 `pgbench_history`。

F.26.4. pgbench Common Options

pgbench 接受以下通用命令行参数：

`-h` *`hostname`*

数据库服务器的主机名。

`-p` *`port`*

数据库服务器的端口号。

`-U` *`login`*

用于连接的用户名

F.26.5. 在 pgbench 中实际执行的“事务”是什么？

默认的事务脚本在每个事务中发出七条命令：

1. `BEGIN;`
2. `UPDATE pgbench_accounts SET abalance = abalance + :delta WHERE aid = :aid;`
3. `SELECT abalance FROM pgbench_accounts WHERE aid = :aid;`
4. `UPDATE pgbench_tellers SET tbalance = tbalance + :delta WHERE tid = :tid;`
5. `UPDATE pgbench_branches SET bbalance = bbalance + :delta WHERE bid = :bid;`
6. `INSERT INTO pgbench_history (tid, bid, aid, delta, mtime) VALUES ( :tid, :bid, :aid, :delta, CURRENT_TIMESTAMP);`
7. `END;`

如果指定`-N`，事务中将不包含第 4 步和第 5 步。如果指定`-S`，则只执行`SELECT`。

F.26.6. 自定义脚本

pgbench 支持运行自定义基准场景，方法是用从文件读取的事务脚本（`-f` 选项）替换（前文所述的）默认事务脚本。在这种情况下，一次“事务”算作脚本文件的一次执行。
甚至可以指定多个脚本（多个 `-f` 选项），此时每次客户端会话开始新事务时，都会随机选择其中一个脚本。

脚本文件的格式是每行一条 SQL 命令；不支持跨多行的 SQL 命令。空行以及以`--`开头的行会被忽略。脚本文件的行还可以是“元命令”，它们由pgbench自身解释，详见下文。

脚本文件有一个简单的变量替换设施。变量可以通过命令行 `-D` 选项（前文已说明）或下文说明的元命令来设置。除了由 `-D` 命令行选项预设的变量之外，变量 `scale` 会被预设为当前缩放因子。变量一经设置，就可以通过写成 `:`*`variablename`* 把它的值插入 SQL 命令中。运行多个客户端会话时，每个会话都有自己的一组变量。

脚本文件中的元命令以反斜线（`\`）开头。元命令的参数以空白分隔。支持的元命令如下：

`\set` *`varname operand1`* `[` *`operator operand2`* `]`

将变量*`varname`*设置为一个计算出的整数值。每个*`operand`*要么是一个整数常量，
要么是一个对具有整数值的变量的`:`*`variablename`*引用。*`operator`*可以是`+`、`-`、`*`或`/`。

示例：

```
\set ntellers 10 * :scale
```

\setrandom *varname min max*

将变量*varname*设置为介于*min*和*max*之间（含边界）的一个随机整数值。
每个界限既可以是整数常量，也可以是指向某个整数值变量的:*variablename*引用。

示例：

```
\setrandom aid 1 :naccounts
```

\sleep *number* [us | ms | s]

让脚本执行休眠指定时长，单位可以是微秒（us）、毫秒（ms）或秒（s）。
如果省略单位，则默认为秒。*number*可以是整数常量，也可以是引用了整数值变量的:*variablename*。

示例：

```
\sleep 10 ms
```

\setshell *varname command* [*argument* ...]

将变量*varname*设置为 shell 命令*command*的结果。该命令必须通过标准输出返回一个整数值。

*argument*可以是文本常量，也可以是指向任意类型变量的:*variablename*引用。
如果要使用以冒号开头的*argument*，需要在*argument*开头再添加一个冒号。

示例：

```
\setshell variable_to_be_assigned command literal_argument :
variable ::literal_starting_with_colon
```

\shell *command* [*argument* ...]

与\setshell相同，但结果会被忽略。

示例：

```
\shell command literal_argument :variable ::literal_starting_with_
colon
```

作为一个示例，内置的类 TPC-B 事务的全部定义是：

```
\set nbranches :scale
\set ntellers 10 * :scale
\set naccounts 100000 * :scale
\setrandom aid 1 :naccounts
\setrandom bid 1 :nbranches
\setrandom tid 1 :ntellers
\setrandom delta -5000 5000
BEGIN;
UPDATE pgbench_accounts SET abalance = abalance + :delta WHERE aid = :
aid;
```

```
SELECT abalance FROM pgbench_accounts WHERE aid = :aid;
UPDATE pgbench_tellers SET tbalance = tbalance + :delta WHERE tid = :
tid;
UPDATE pgbench_branches SET bbalance = bbalance + :delta WHERE bid = :
bid;
INSERT INTO pgbench_history (tid, bid, aid, delta, mtime) VALUES (:
tid, :bid, :aid, :delta, CURRENT_TIMESTAMP);
END;
```

该脚本允许事务的每次迭代都引用不同的随机选中行。
（这个示例也说明了为什么每个客户端会话都必须拥有自己的变量 – 否则它们就无法彼此独立地访问不同的行。）

F.26.7. 逐事务日志记录

使用 `-l` 选项时，pgbench 会把每个事务所耗时间写入日志文件。日志文件名为 `pgbench_log.`*`nnn`*，其中 *`nnn`* 是 pgbench 进程的 PID。如果 `-j` 选项为 2 或更高（创建多个工作线程），每个线程都会有自己的日志文件。第一个工作线程的日志文件名与标准单工作线程情形相同。其余工作线程的额外日志文件名为 `pgbench_log.`*`nnn`*`.`*`mmm`*，其中 *`mmm`* 是从 1 开始的各工作线程顺序编号。

日志格式如下：

`client_id transaction_no time file_no time_epoch time_us`

其中，*`time`* 是事务经过的总时间，单位为微秒，*`file_no`* 标识使用的脚本文件（在通过 `-f` 指定多个脚本时很有用），而 *`time_epoch`*/*`time_us`* 分别是 Unix 纪元格式的时间戳和以微秒计的偏移量（适合用来生成带小数秒的 ISO 8601 时间戳），表示事务完成的时间。

示例输出如下：

```
 0 199 2241 0 1175850568 995598
 0 200 2465 0 1175850568 998079
 0 201 2513 0 1175850569 608
 0 202 2038 0 1175850569 2663
```

F.26.8. 逐语句延迟

使用`-r`选项时，pgbench会收集每个客户端执行的每条语句所经过的事务时间。
基准测试完成后，它会报告这些值的平均值，称为每条语句的延迟。

对于默认脚本，输出与下面类似：

```
starting vacuum...end.
transaction type: TPC-B (sort of)
scaling factor: 1
query mode: simple
number of clients: 10
number of threads: 1
number of transactions per client: 1000
number of transactions actually processed: 10000/10000
tps = 618.764555 (including connections establishing)
tps = 622.977698 (excluding connections establishing)
statement latencies in milliseconds:
```

```
        0.004386        \set nbranches 1 * :scale
        0.001343        \set ntellers 10 * :scale
        0.001212        \set naccounts 100000 * :scale
        0.001310        \setrandom aid 1 :naccounts
        0.001073        \setrandom bid 1 :nbranches
        0.001005        \setrandom tid 1 :ntellers
        0.001078        \setrandom delta -5000 5000
        0.326152        BEGIN;
        0.603376        UPDATE pgbench_accounts SET abalance =
 abalance + :delta WHERE aid = :aid;
        0.454643        SELECT abalance FROM pgbench_accounts WHERE
 aid = :aid;
        5.528491        UPDATE pgbench_tellers SET tbalance = tbalance
 + :delta WHERE tid = :tid;
        7.335435        UPDATE pgbench_branches SET bbalance =
 bbalance + :delta WHERE bid = :bid;
        0.371851        INSERT INTO pgbench_history (tid, bid, aid,
 delta, mtime) VALUES (:tid, :bid, :aid, :delta, CURRENT_TIMESTAMP);
        1.212976        END;
```

如果指定了多个脚本文件，则会分别为每个脚本文件报告平均值。

注意，为逐语句延迟计算收集额外的计时信息会带来一定开销。这会拖慢平均执行速度，并降低计算出的 TPS。减速幅度在很大程度上取决于平台和硬件。比较启用和未启用延迟报告时的平均 TPS 值，是判断这一计时开销是否显著的好方法。

F.26.9. 良好实践

很容易用pgbench得出完全没有意义的数字。下面给出一些有助于获得有用结果的准则。

首先，绝不要相信任何只运行几秒钟的测试。使用 `-t` 或 `-T` 选项让运行至少持续几分钟，以平均掉噪声。在某些情况下，可能需要数小时才能获得可重现的数字。最好把测试多运行几次，看看你的数字是否可重现。

对于默认的类 TPC-B 测试场景，初始化缩放因子（`-s`）至少应与你打算测试的最大客户端数（`-c`）一样大；否则你测到的将主要是更新争用。`pgbench_branches` 表中只有 `-s` 行，而每个事务都要更新其中一行，因此 `-c` 值超过 `-s` 无疑会导致大量事务阻塞等待其他事务。

默认测试场景还会对表初始化后的时间长短非常敏感：表中死元组和无效空间的累积会改变结果。要理解这些结果，必须跟踪更新总数以及何时发生清理。如果启用了自动清理，它可能会给测得的性能带来不可预测的变化。

pgbench的一个局限是：在尝试测试大量客户端会话时，它自己也可能成为瓶颈。可以通过在与数据库服务器不同的机器上运行pgbench来缓解这一点，不过网络延迟必须足够低。甚至可以在多台客户端机器上同时运行多个pgbench实例，对同一台数据库服务器施压。

F.27. pg_buffercache

pg_buffercache 模块提供了一种手段，可实时检查共享缓冲区缓存中正在发生的情况。它还提供了一种底层方式，可出于测试目的从中逐出数据。

该模块提供一个 C 函数pg_buffercache_pages，它返回一组记录；另外还提供一个视图pg_buffercache，该视图封装此函数以便使用。

默认情况下，这两个对象的公共访问权限都已被撤销，以防存在潜在的安全问题。

F.27.1. pg_buffercache 视图

该视图公开的列定义见表 F.16。

表 F.16. pg_buffercache 列

Name	Type	References	Description
bufferid	integer		ID, in the range 1.. shared_buffers
relfilenode	oid	pg_class.relfilenode	Filenode number of the relation
reltablespace	oid	pg_tablespace.oid	Tablespace OID of the relation
reldatabase	oid	pg_database.oid	Database OID of the relation
relblocknumber	bigint		Page number within the relation
relforknumber	smallint		Fork number within the relation
isdirty	boolean		Is the page dirty?
usagecount	smallint		Page LRU count

共享缓存中的每个缓冲区都有一行记录。未使用的缓冲区除了 bufferid 之外，其余字段均显示为 null。共享系统目录显示为属于数据库 0。

由于缓存由所有数据库共享，其中通常会有不属于当前数据库的关系页。这意味着某些行在 pg_class 中可能没有可匹配的行，甚至还可能出现错误的连接结果。如果要与 pg_class 进行连接，最好将连接限制在 reldatabase 等于当前数据库 OID 或 0 的那些行上。

访问 pg_buffercache 视图时，会在足够长的时间内持有内部缓冲区管理器锁，以复制该视图将要显示的全部缓冲区状态数据。这确保了该视图产生一致的结果集，
同时又不会对正常缓冲区活动造成超出必要时长的阻塞。尽管如此，如果频繁读取该视图，
仍可能对数据库性能产生一定影响。

F.27.2. 示例输出

```
regression=# SELECT c.relname, count(*) AS buffers
             FROM pg_buffercache b INNER JOIN pg_class c
             ON b.relfilenode = pg_relation_filenode(c.oid) AND
                b.reldatabase IN (0, (SELECT oid FROM pg_database
```

```
                                        WHERE datname = current_
database()))
             GROUP BY c.relname
             ORDER BY 2 DESC
             LIMIT 10;

             relname              | buffers
---------------------------------+---------
 tenk2                           |     345
 tenk1                           |     141
 pg_proc                         |      46
 pg_class                        |      45
 pg_attribute                    |      43
 pg_class_relname_nsp_index      |      30
 pg_proc_proname_args_nsp_index  |      28
 pg_attribute_relid_attnam_index |      26
 pg_depend                       |      22
 pg_depend_reference_index       |      20
(10 rows)
```

F.27.3. 作者

Mark Kirkwood <markir@paradise.net.nz>

设计建议：Neil Conway <neilc@samurai.com>

调试建议：Tom Lane <tgl@sss.pgh.pa.us>

F.28. pgcrypto

pgcrypto模块为PostgreSQL提供密码学函数。

F.28.1. 通用哈希函数

F.28.1.1. digest()

```
digest(data text, type text) returns bytea
digest(data bytea, type text) returns bytea
```

计算给定*data*的二进制哈希值。*type*是要使用的算法。标准算法包括md5、sha1、sha224、sha256、sha384和sha512。如果pgcrypto是在有 OpenSSL 的情况下构建的，则还有更多算法可用，详见表 F.20。

如果想把摘要表示为十六进制字符串，可以对结果使用encode()。例如：

```
CREATE OR REPLACE FUNCTION sha1(bytea) returns text AS $$
    SELECT encode(digest($1, 'sha1'), 'hex')
$$ LANGUAGE SQL STRICT IMMUTABLE;
```

F.28.1.2. hmac()

```
hmac(data text, key text, type text) returns bytea
hmac(data bytea, key text, type text) returns bytea
```

使用密钥*key*为*data*计算基于散列的消息认证码（HMAC）。*type*与digest()中的相同。

这与digest()类似，但只有知道密钥时才能重新计算该哈希。
这可以防止有人篡改数据后再同时修改哈希使之匹配。

如果密钥大于哈希块大小，则会先对其进行哈希，并将结果用作密钥。

F.28.2. 密码哈希函数

函数crypt()和gen_salt()是专门为密码哈希设计的。crypt()负责执行哈希，而gen_salt()负责为其准备算法参数。

crypt() 中的算法在以下方面不同于 MD5 或 SHA1 之类通常的哈希算法：

1. 它们很慢。由于处理的数据量很小，这是使暴力破解密码变得困难的唯一办法。
2. 它们使用一个称为salt的随机盐值，这样使用相同密码的用户也会得到不同的哈希结果。这也为逆向求解该算法增加了一层额外防护。
3. 它们会在结果中包含算法类型，这样用不同算法哈希的密码就能共存。
4. 其中一些是自适应的 – 这意味着当计算机变快时，你可以把算法调得更慢，而不引入与现有密码的不兼容性。

表 F.17列出了crypt()函数支持的算法。

表 F.17. crypt() 支持的算法

Algorithm	Max Password Length	Adaptive?	Salt Bits	Description
bf	72	yes	128	Blowfish-based, variant 2a
md5	unlimited	no	48	MD5-based crypt
xdes	8	yes	24	Extended DES
des	8	no	12	Original UNIX crypt

F.28.2.1. crypt()

```
crypt(password text, salt text) returns text
```

计算*password*的一个 crypt(3) 风格哈希。存储新密码时，需要使用gen_salt()生成新的*salt*值。校验密码时，把已存储的哈希值作为*salt*传入，并测试结果是否与已存储值匹配。

设置一个新密码的示例：

```
UPDATE ... SET pswhash = crypt('new password', gen_salt('md5'));
```

Example of authentication:

```
SELECT pswhash = crypt('entered password', pswhash) FROM ... ;
```

如果输入的密码正确，这会返回 true。

F.28.2.2. gen_salt()

```
gen_salt(type text [, iter_count integer ]) returns text
```

生成一个供crypt()使用的新随机盐值字符串。该盐值字符串还会告诉crypt()应使用哪种算法。

*type*参数指定 hash 算法。接受的类型有：des、xdes、md5和bf。

*iter_count*参数允许用户为支持该参数的算法指定迭代次数。次数越高，对密码进行 hash 所需的时间越长，从而破解它所需时间也越长。不过，如果次数过高，计算一个 hash 可能需要数年时间 – 这显然不切实际。若省略*iter_count*参数，则使用默认迭代次数。允许的*iter_count*值取决于算法，如表 F.18所示。

表 F.18. crypt() 的迭代计数

Algorithm	Default	Min	Max
xdes	725	1	16777215
bf	6	4	31

对xdes算法还有额外的限制：迭代计数必须是一个奇数。

为了选择合适的迭代次数，可以考虑原始 DES crypt 在当时硬件上的设计速度是每秒 4 次 hash。低于每秒 4 次 hash 可能会影响可用性，而高于每秒 100 次 hash 则很可能过快。

表 F.19概述了不同 hash 算法之间的相对速度差异。该表展示了在 8 字符密码上尝试所有字符组合所需的时间，假定密码只包含小写字母，或者包含大小写字母和数字。在`crypt-bf`条目中，斜杠后的数字是`gen_salt`的*`iter_count`*参数值。

表 F.19. Hash 算法速度

Algorithm	Hashes/sec	For `[a-z]`	For `[A-Za-z0-9]`
`crypt-bf/8`	28	246 years	251322 years
`crypt-bf/7`	57	121 years	123457 years
`crypt-bf/6`	112	62 years	62831 years
`crypt-bf/5`	211	33 years	33351 years
`crypt-md5`	2681	2.6 years	2625 years
`crypt-des`	362837	7 days	19 years
`sha1`	590223	4 days	12 years
`md5`	2345086	1 day	3 years

注意：

- 所用机器为 1.5GHz Pentium 4。
- `crypt-des`和`crypt-md5`算法的数字取自 John the Ripper v1.6.38 的 `-test` 输出。
- `md5` 的数据来自 mdcrack 1.2。
- `sha1`的数字来自 lcrack-20031130-beta。
- `crypt-bf`的数字是使用一个简单程序测得的，该程序循环处理 1000 个 8 字符密码。这样可以展示不同迭代次数下的速度。作为参考：`john -test` 对 `crypt-bf/5` 给出的结果是 213 次循环/秒。（结果上的极小差异与这样一个事实一致：`crypt-bf`在`pgcrypto`中的实现与 John the Ripper 使用的是同一套实现。）

请注意，“尝试所有组合”并不是现实中的做法。通常密码破解是借助词典完成的，其中包含常见单词及其各种变体。因此，即使是稍微有点像单词的密码，被破解的速度也可能远快于上表所示；而一个 6 字符、不像单词的密码则可能逃过破解，也可能不会。

F.28.3. PGP 加密函数

这里的函数实现 OpenPGP（RFC 4880）标准的加密部分。同时支持对称密钥加密和公钥加密。

一个加密的 PGP 消息由两个部分，或称两个包组成：

- 包含会话密钥的包 – 该会话密钥要么由对称密钥加密，要么由公钥加密。
- 包含用会话密钥加密的数据的包。

当使用对称密钥（即密码）加密时：

1. 给定密码使用 String2Key (S2K) 算法进行哈希。这与`crypt()`算法颇为类似 – 故意设计得较慢，并带有随机盐值 – 但它产生的是一个全长度的二进制密钥。
2. 如果请求单独的会话密钥，则会生成一个新的随机密钥。否则，S2K 密钥将直接用作会话密钥。

3. 如果直接使用 S2K 密钥，那么会话密钥包中只写入 S2K 设置。否则，会话密钥会先用 S2K 密钥加密，再放入会话密钥包。

当使用公钥加密时：

1. 会生成一个新的随机会话密钥。
2. 该密钥会用公钥加密，并放入会话密钥包中。

无论哪种情况，要加密的数据都会按如下步骤处理：

1. 可选的数据处理包括：压缩、转换为 UTF-8 和转换行结束符，三者可任意组合。
2. 数据前面会加上一个随机字节块。这相当于使用随机 IV。
3. 追加对随机前缀和数据计算得到的 SHA-1 哈希值。
4. 然后用会话密钥加密所有这些内容，并放入数据包。

F.28.3.1. pgp_sym_encrypt()

```
pgp_sym_encrypt(data text, psw text [, options text ]) returns bytea
pgp_sym_encrypt_bytea(data bytea, psw text [, options text ]) returns
 bytea
```

使用对称 PGP 密码*psw*加密*data*。*options*参数可以包含下文所述的选项设置。

F.28.3.2. pgp_sym_decrypt()

```
pgp_sym_decrypt(msg bytea, psw text [, options text ]) returns text
pgp_sym_decrypt_bytea(msg bytea, psw text [, options text ]) returns
 bytea
```

解密一个经过对称密钥加密的 PGP 消息。

`bytea`数据不能用`pgp_sym_decrypt`解密。这是为了避免输出无效字符数据。
若原始数据本来是文本，则使用`pgp_sym_decrypt_bytea`解密也没有问题。

*options*参数可以包含下文所述的选项设置。

F.28.3.3. pgp_pub_encrypt()

```
pgp_pub_encrypt(data text, key bytea [, options text ]) returns bytea
pgp_pub_encrypt_bytea(data bytea, key bytea [, options text ]) returns
 bytea
```

使用 PGP 公钥*key*加密*data*。向该函数提供私钥会报错。

*options*参数可以包含下文所述的选项设置。

F.28.3.4. pgp_pub_decrypt()

```
pgp_pub_decrypt(msg bytea, key bytea [, psw text [, options text ]])
 returns text
pgp_pub_decrypt_bytea(msg bytea, key bytea [, psw text [, options
 text ]]) returns bytea
```

解密经过公钥加密的消息。*key*必须是与加密时所用公钥对应的私钥。如果私钥受密码保护，则必须在*psw*中给出密码。如果没有密码但想指定选项，则需要传入空密码。

bytea数据不能用pgp_pub_decrypt解密。这是为了避免输出无效字符数据。若原始数据本来是文本，则使用pgp_pub_decrypt_bytea解密也没有问题。

*options*参数可以包含下文所述的选项设置。

F.28.3.5. pgp_key_id()

```
pgp_key_id(bytea) returns text
```

pgp_key_id提取 PGP 公钥或私钥的密钥 ID。如果传入的是加密消息，则返回用于加密该数据的密钥 ID。

它可以返回两个特殊的密钥 ID：

- SYMKEY

  该消息是用对称密钥加密的。

- ANYKEY

  该消息是用公钥加密的，但密钥 ID 已被移除。这意味着你需要尝试自己的所有私钥，看看哪一个能解密它。pgcrypto本身不会生成这样的消息。

注意，不同的密钥可能具有相同的 ID。这种情况虽然罕见，但属于正常情况。客户端应用此时应该尝试使用每一个密钥解密，以判断哪个匹配 – 就像处理ANYKEY时一样。

F.28.3.6. armor(), dearmor()

```
armor(data bytea) returns text
dearmor(data text) returns bytea
```

这些函数将二进制数据封装/解封装为 PGP ASCII-armor 格式，它本质上就是带 CRC 和附加格式信息的 Base64。

F.28.3.7. PGP 函数的选项

这些选项的命名方式与 GnuPG 类似。选项值应写在等号后面；各选项之间用逗号分隔。例如：

```
pgp_sym_encrypt(data, psw, 'compress-algo=1, cipher-algo=aes256')
```

除convert-crlf外，所有选项都只适用于加密函数。解密函数会从 PGP 数据中获取这些参数。

最值得关注的选项可能是compress-algo和unicode-mode。
其余选项应该都具有合理的默认值。

F.28.3.7.1. cipher-algo

使用哪种密码算法。

Values: bf, aes128, aes192, aes256 (OpenSSL-only: 3des, cast5)
Default: aes128
Applies to: pgp_sym_encrypt, pgp_pub_encrypt

F.28.3.7.2. compress-algo

使用哪种压缩算法。仅当PostgreSQL在编译时包含 zlib 时才可用。

Values:
 0 - no compression
 1 - ZIP compression
 2 - ZLIB compression (= ZIP plus meta-data and block CRCs)
Default: 0
Applies to: pgp_sym_encrypt, pgp_pub_encrypt

F.28.3.7.3. compress-level

压缩程度。级别越高，压缩后越小，但速度也越慢。0 表示禁用压缩。

Values: 0, 1-9
Default: 6
Applies to: pgp_sym_encrypt, pgp_pub_encrypt

F.28.3.7.4. convert-crlf

是否在加密时将`\n`转换为`\r\n`，并在解密时将`\r\n`转换为`\n`。RFC 4880 规定文本数据应使用`\r\n`换行符存储。使用此选项可获得完全符合 RFC 的行为。

Values: 0, 1
Default: 0
Applies to: pgp_sym_encrypt, pgp_pub_encrypt, pgp_sym_decrypt, pgp_pub_decrypt

F.28.3.7.5. disable-mdc

不使用 SHA-1 保护数据。使用此选项唯一合理的理由是兼容早于 RFC 4880 加入 SHA-1 保护包的古老 PGP 产品。较新的 gnupg.org 和 pgp.com 软件都能很好地支持它。

Values: 0, 1
Default: 0
Applies to: pgp_sym_encrypt, pgp_pub_encrypt

F.28.3.7.6. sess-key

使用单独的会话密钥。公钥加密总是使用单独的会话密钥；这个选项用于对称密钥加密，因为后者默认直接使用 S2K 密钥。

Values: 0, 1
Default: 0
Applies to: pgp_sym_encrypt

F.28.3.7.7. s2k-mode

使用哪一种 S2K 算法。

Values:
 0 - Without salt. Dangerous!
 1 - With salt but with fixed iteration count.
 3 - Variable iteration count.
Default: 3

Applies to: pgp_sym_encrypt

F.28.3.7.8. s2k-digest-algo

要在 S2K 计算中使用哪种摘要算法。

Values: md5, sha1
Default: sha1
Applies to: pgp_sym_encrypt

F.28.3.7.9. s2k-cipher-algo

使用哪种密码算法来加密单独的会话密钥。

Values: bf, aes, aes128, aes192, aes256
Default: use cipher-algo
Applies to: pgp_sym_encrypt

F.28.3.7.10. unicode-mode

是否将文本数据在数据库内部编码和 UTF-8 之间相互转换。如果数据库已经是 UTF-8，则不会发生转换，但消息会被标记为 UTF-8。如果不启用该选项，则不会进行这种标记。

Values: 0, 1
Default: 0
Applies to: pgp_sym_encrypt, pgp_pub_encrypt

F.28.3.8. 使用 GnuPG 生成 PGP 密钥

要生成新密钥：

```
gpg --gen-key
```

首选的密钥类型是“DSA 和 Elgamal”。

对于 RSA 加密，你必须先创建一个仅用于签名的 DSA 或 RSA 主密钥，然后使用`gpg --edit-key`添加 RSA 加密子密钥。

要列出密钥：

```
gpg --list-secret-keys
```

要以 ASCII-armor 格式导出公钥：

```
gpg -a --export KEYID > public.key
```

要以 ASCII-armor 格式导出私钥：

```
gpg -a --export-secret-keys KEYID > secret.key
```

在把这些密钥交给 PGP 函数之前，需要先用`dearmor()`处理它们。或者，如果你能处理二进制数据，也可以从命令中去掉`-a`。

更多细节见`man gpg`、GNU Privacy Handbook[3]以及http://www.gnupg.org上的其他文档。

[3] http://www.gnupg.org/gph/en/manual.html

F.28.3.9. PGP 代码的限制

- 不支持签名。这也意味着不会检查加密子密钥是否属于主密钥。
- 不支持使用加密密钥作为主密钥。由于通常不鼓励这种做法，这不应成为问题。
- 不支持多个子密钥。这看起来可能是个问题，因为这在实践中相当常见。另一方面，你不应将常规的 GPG/PGP 密钥用于`pgcrypto`，而应新建一套密钥，因为其使用场景相当不同。

F.28.4. 原始加密函数

这些函数只是对数据应用密码算法；它们不具备 PGP 加密的任何高级特性。因此存在一些严重问题：

1. 它们直接把用户提供的密钥用作密码算法的密钥。
2. 它们不提供任何完整性检查，无法判断加密数据是否被修改。
3. 它们希望用户自己管理所有加密参数，甚至是 IV。
4. 它们无法处理文本。

因此，在引入 PGP 加密之后，不建议使用原始加密函数。

```
encrypt(data bytea, key bytea, type text) returns bytea
decrypt(data bytea, key bytea, type text) returns bytea

encrypt_iv(data bytea, key bytea, iv bytea, type text) returns bytea
decrypt_iv(data bytea, key bytea, iv bytea, type text) returns bytea
```

对数据进行加密/解密，所用加密方法由 *type* 指定。*type* 字符串的语法是：

```
algorithm [ - mode ] [ /pad: padding ]
```

其中 *algorithm* 是以下之一：

- `bf` – Blowfish
- `aes` – AES（Rijndael-128）

and *mode* is one of:

- `cbc` – 下一个块依赖于前一个块（默认）
- `ecb` – 每个块单独加密（仅用于测试）

and *padding* is one of:

- `pkcs` – 数据长度可以任意（默认）
- `none` – 数据必须是密码块大小的整数倍

因此，例如下面两种写法是等价的：

```
encrypt(data, 'fooz', 'bf')
```

```
encrypt(data, 'fooz', 'bf-cbc/pad:pkcs')
```

在`encrypt_iv`和`decrypt_iv`中，*`iv`*参数是 CBC 和 CFB 模式的初始值；对于 ECB，它会被忽略。如果长度不等于块大小，则会被截断或用零填充。在不带该参数的函数中，它默认为全零。

F.28.5. 随机数据函数

```
gen_random_bytes(count integer) returns bytea
```

返回*`count`*个具有密码学安全强度的随机字节。一次最多可提取 1024 个字节，以避免耗尽随机数生成器池。

F.28.6. 注意事项

F.28.6.1. 配置

`pgcrypto`会根据 PostgreSQL 主`configure`脚本的探测结果自行配置。影响它的选项有`--with-zlib`和`--with-openssl`。

如果编译时包含 zlib，PGP 加密函数就能够在加密前压缩数据。

如果编译时包含 OpenSSL，则会有更多算法可用。同时公钥加密函数也会更快，因为 OpenSSL 具有更加优化的 BIGNUM 函数。

表 F.20. 有无 OpenSSL 时的功能概要

Functionality	Built-in	With OpenSSL
MD5	yes	yes
SHA1	yes	yes
SHA224/256/384/512	yes	yes (Note 1)
Other digest algorithms	no	yes (Note 2)
Blowfish	yes	yes
AES	yes	yes (Note 3)
DES/3DES/CAST5	no	yes
Raw encryption	yes	yes
PGP Symmetric encryption	yes	yes
PGP Public-Key encryption	yes	yes

注意：

1. SHA2 算法是在 OpenSSL 0.9.8 版中加入的。对于更早的版本，`pgcrypto`将使用内建代码。
2. OpenSSL 支持的任何摘要算法都会自动被采用。加密算法则无法如此，必须显式添加支持。
3. AES 是在 OpenSSL 0.9.7 版中引入的。对于更早的版本，`pgcrypto`将使用内建代码。

F.28.6.2. NULL 处理

按照 SQL 的标准，只要任一参数为 NULL，所有函数都返回 NULL。这在使用不慎时可能带来安全风险。

F.28.6.3. 安全性限制

所有`pgcrypto`函数都在数据库服务器内部运行。这意味着所有数据和密码都会在`pgcrypto`与客户端应用之间以明文传输。因此，你必须：

1. 使用本地连接或 SSL 连接。
2. 同时信任系统管理员和数据库管理员。

如果做不到，最好在客户端应用内部执行密码学操作。

该实现无法抵御侧信道攻击[4]。例如，对于给定大小的密文，`pgcrypto`解密函数完成所需的时间会因具体密文不同而变化。

F.28.6.4. 有用的阅读材料

- http://www.gnupg.org/gph/en/manual.html

  GNU 隐私手册。

- http://www.openwall.com/crypt/

  介绍 crypt-blowfish 算法。

- http://www.stack.nl/~galactus/remailers/passphrase-faq.html

  如何选择一个好的密码。

- http://world.std.com/~reinhold/diceware.html

  选择密码的有趣思路。

- http://www.interhack.net/people/cmcurtin/snake-oil-faq.html

  介绍好的和不好的密码学。

F.28.6.5. 技术参考资料

- http://www.ietf.org/rfc/rfc4880.txt

  OpenPGP 消息格式。

- http://www.ietf.org/rfc/rfc1321.txt

  MD5 消息摘要算法。

- http://www.ietf.org/rfc/rfc2104.txt

  HMAC：用于消息认证的密钥散列。

- http://www.usenix.org/events/usenix99/provos.html

  crypt-des、crypt-md5 和 bcrypt 算法的比较。

- http://csrc.nist.gov/cryptval/des.htm

[4] http://en.wikipedia.org/wiki/Side-channel_attack

DES、3DES 和 AES 的标准。

- http://en.wikipedia.org/wiki/Fortuna_(PRNG)

  Fortuna CSPRNG 的描述。

- http://jlcooke.ca/random/

  Jean-Luc Cooke 基于 Fortuna 的 Linux `/dev/random` 驱动。

- http://research.cyber.ee/~lipmaa/crypto/

  密码学资料链接汇集。

F.28.7. 作者

Marko Kreen `<markokr@gmail.com>`

`pgcrypto`使用了来自以下来源的代码：

Algorithm	Author	Source origin
DES crypt	David Burren and others	FreeBSD libcrypt
MD5 crypt	Poul-Henning Kamp	FreeBSD libcrypt
Blowfish crypt	Solar Designer	www.openwall.com
Blowfish cipher	Simon Tatham	PuTTY
Rijndael cipher	Brian Gladman	OpenBSD sys/crypto
MD5 and SHA1	WIDE Project	KAME kame/sys/crypto
SHA256/384/512	Aaron D. Gifford	OpenBSD sys/crypto
BIGNUM math	Michael J. Fromberger	dartmouth.edu/~sting/sw/imath

F.29. pg_freespacemap

pg_freespacemap模块提供了一种检查空闲空间映射（FSM）的方法。它提供了一个名为pg_freespace的函数，更准确地说，是两个重载函数。
这些函数显示FSM中为给定页面或关系中所有页面记录的值。

默认情况下，这些函数的公共访问权限已被撤销，以防存在潜在的安全问题。

F.29.1. 函数

pg_freespace(rel regclass IN, blkno bigint IN) returns int2

返回该关系中由blkno指定页面上按 FSM 记录的空闲空间量。

pg_freespace(rel regclass IN, blkno OUT bigint, avail OUT int2)

显示该关系中每个页面上按 FSM 记录的空闲空间量。返回一组(blkno bigint, avail int2)元组，其中关系中的每个页面对应一个元组。

存储在空闲空间映射中的值并不精确。它们会被舍入到BLCKSZ的 1/256 精度（在默认 BLCKSZ 下为 32 字节），并且不会随着元组的插入和更新而完全保持最新。

对于索引，被跟踪的是完全未使用的页面，而不是页面内部的空闲空间。因此，
这些值本身并无实际意义，只能表明某个页面是满的还是空的。

注意：该接口在 8.4 版中发生了变化，以反映同一版本中引入的新 FSM 实现。

F.29.2. 示例输出

```
postgres=# SELECT * FROM pg_freespace('foo');
 blkno | avail
-------+-------
     0 |     0
     1 |     0
     2 |     0
     3 |    32
     4 |   704
     5 |   704
     6 |   704
     7 |  1216
     8 |   704
     9 |   704
    10 |   704
    11 |   704
    12 |   704
    13 |   704
    14 |   704
    15 |   704
    16 |   704
    17 |   704
    18 |   704
    19 |  3648
(20 rows)
```

```
postgres=# SELECT * FROM pg_freespace('foo', 7);
 pg_freespace
--------------
         1216
(1 row)
```

F.29.3. 作者

最初版本由 Mark Kirkwood <markir@paradise.net.nz> 编写。在 8.4 版中，Heikki Linnakangas <heikki@enterprisedb.com> 为适应新的 FSM 实现对其进行了重写。

F.30. pgrowlocks – 显示表的行锁信息

pgrowlocks 模块提供了一个函数，用于显示指定表的行锁信息。

F.30.1. 概述

```
pgrowlocks(text) returns setof record
```

参数为表名。结果是一个记录集合，其中表中每个被锁定的行对应结果中的一行。输出列见表 F.21。

表 F.21. pgrowlocks 输出列

Name	Type	Description
locked_row	tid	Tuple ID (TID) of locked row
lock_type	text	Shared for shared lock, or Exclusive for exclusive lock
locker	xid	Transaction ID of locker, or multixact ID if multi-transaction
multi	boolean	True if locker is a multi-transa ction
xids	xid[]	Transaction IDs of lockers (more than one if multi-transa ction)
pids	integer[]	Process IDs of locking backends (more than one if multi-transa ction)

pgrowlocks 会对目标表获取 AccessShareLock，并逐行读取以收集行锁信息。对于大表，这样做速度并不快。请注意：

1. 如果整张表被其他人独占锁定，pgrowlocks 将被阻塞。
2. pgrowlocks 并不保证给出一个自一致快照。在其执行期间，可能会有新的行锁被获取，也可能有旧的锁被释放。

pgrowlocks 不显示被锁定行的内容。如果你想同时查看这些行的内容，可以这样做：

```
SELECT * FROM accounts AS a, pgrowlocks('accounts') AS p
  WHERE p.locked_row = a.ctid;
```

不过要注意，（截至 PostgreSQL 8.3）这样的查询效率会很低。

F.30.2. 示例输出

```
test=# SELECT * FROM pgrowlocks('t1');
 locked_row | lock_type | locker | multi |   xids    |     pids
------------+-----------+--------+-------+-----------+---------------
      (0,1) | Shared    |     19 | t     | {804,805} | {29066,29068}
      (0,2) | Shared    |     19 | t     | {804,805} | {29066,29068}
      (0,3) | Exclusive |    804 | f     | {804}     | {29066}
      (0,4) | Exclusive |    804 | f     | {804}     | {29066}
```

```
(4 rows)
```

F.30.3. 作者

Tatsuo Ishii

F.31. pg_standby

pg_standby 支持创建“温备”数据库服务器。它既是可用于生产环境的程序，也是可定制的模板，便于在有需要时进行特定修改。

pg_standby 设计为一个会等待的 `restore_command`，将标准归档恢复转变为温备运行需要这样的命令。此外还需要其他配置，服务器的主手册中对此均有介绍（参见第 25.2 节）。

pg_standby features include:

- 用 C 语言编写，因此移植性很好，易于安装
- 源代码易于修改，并且专门划分出了若干区域，供你按自身需要修改
- 已在 Linux 和 Windows 上经过测试

F.31.1. Usage

要配置备库使用 pg_standby，请将以下内容放入其 `recovery.conf` 配置文件：

```
restore_command = 'pg_standby archiveDir %f %p %r'
```

其中，*archiveDir* 是恢复 WAL 段文件时所读取的目录。

The full syntax of pg_standby's command line is

```
pg_standby [ option ... ] archivelocation nextwalfile xlogfilepath
 [ restartwalfile ]
```

在 `restore_command` 中使用时，应为 `nextwalfile` 和 `xlogfilepath` 分别指定 *%f* 和 *%p* 宏，以提供恢复所需的实际文件和路径。

如果指定了 *restartwalfile*（通常使用 `%r` 宏），就会从 *archivelocation* 中删除逻辑上位于该文件之前的所有 WAL 文件。这样既能尽量减少需要保留的文件数量，又能保留崩溃后重启的能力。如果 *archivelocation* 是专供此备库使用的临时暂存区，就适合使用此参数；但如果 *archivelocation* 用作长期 WAL 归档区，就不应使用。

pg_standby 假定 *archivelocation* 是服务器所属用户可读的目录。如果指定了 *restartwalfile*（或 `-k`），*archivelocation* 目录还必须可写。

主库发生故障时，切换到“温备”数据库服务器有以下两种方式：

智能故障切换

在智能故障切换中，服务器会先应用归档中所有可用的 WAL 文件，再正式启动。即使备库已经落后，这也能做到零数据丢失；但如果尚未应用的 WAL 很多，备库可能需要很长时间才能就绪。要触发智能故障切换，请创建包含单词 `smart` 的触发文件，或者直接创建一个空的触发文件。

快速故障切换

在快速故障切换中，服务器会立即正式启动。归档中所有尚未应用的 WAL 文件都会被忽略，这些文件中的全部事务都会丢失。要触发快速故障切换，请创建触发文件，
并在其中写入单词 `fast`。也可以配置 pg_standby，使其在指定时间间隔内没有出现新 WAL 文件时，自动执行快速故障切换。

F.31.2. pg_standby Options

pg_standby accepts the following command-line arguments:

`-c`

使用 `cp` 或 `copy` 命令从归档恢复 WAL 文件。这是唯一受支持的行为，因此该选项没有用处。

`-d`

在 `stderr` 上输出大量调试日志。

`-k`

从 *archivelocation* 中删除文件，使归档中当前文件之前保留的 WAL 文件不超过该数量。零（默认值）表示不从 *archivelocation* 中删除任何文件。如果指定了 *restartwalfile*，此参数会被静默忽略，因为那种指定方式能更准确地确定正确的归档截断点。从 PostgreSQL 8.3 起，此参数已弃用；指定 *restartwalfile* 参数更安全也更高效。设置得太小可能导致删除备库服务器重启仍需要的文件，而设置得太大又会浪费归档空间。

`-r` *maxretries*

设置复制命令失败后的最大重试次数（默认为 3）。每次失败后，都会等待 *sleeptime* * *num_retries*，因此等待时间会逐次增加。默认情况下，会依次等待 5 秒、10 秒、15 秒，然后才向备库服务器报告失败。这会被解释为恢复结束，从而使备库完全启动。

`-s` *sleeptime*

设置在两次测试要恢复的 WAL 文件是否已出现在归档之间休眠的秒数（最多 60，默认 5）。默认设置不一定就是推荐设置；相关讨论参见第 25.2 节。

`-t` *triggerfile*

指定触发文件；一旦该文件出现，就会引发故障切换。建议使用带有明确结构的文件名，例如 `/tmp/pgsql.trigger.5432`，以免同一系统上存在多个服务器时，混淆被触发的是哪个服务器。

`-w` *maxwaittime*

设置等待下一个 WAL 文件的最长秒数，超过此时间后执行快速故障切换。零（默认值）表示永远等待。默认设置不一定就是推荐设置；相关讨论参见第 25.2 节。

F.31.3. 示例

在 Linux 或 Unix 系统上，可以使用：

```
archive_command = 'cp %p .../archive/%f'

restore_command = 'pg_standby -d -s 2 -t /tmp/pgsql.trigger.5442 .../
archive %f %p %r 2>>standby.log'

recovery_end_command = 'rm -f /tmp/pgsql.trigger.5442'
```

这里，归档目录实际位于备库上，因此 `archive_command` 通过 NFS 访问该目录，但这些文件对备库而言是本地文件（因此可以使用 `ln`）。此配置会：

- 将调试输出写入 `standby.log`
- 每隔 2 秒检查一次下一个 WAL 文件是否可用
- 仅在名为 `/tmp/pgsql.trigger.5442` 的触发文件出现时停止等待，并根据其内容执行故障切换
- 恢复结束时删除触发文件
- 从归档目录中移除不再需要的文件

在 Windows 上，可以使用：

```
archive_command = 'copy %p ...\\archive\\%f'

restore_command = 'pg_standby -d -s 5 -t C:\pgsql.trigger.5442 ...
\archive %f %p %r 2>>standby.log'

recovery_end_command = 'del C:\pgsql.trigger.5442'
```

注意，需要双写反斜杠的设置是 `archive_command`，而无需双写的设置是 `restore_command` 或 `recovery_end_command`。此配置会：

- 使用 `copy` 命令从归档恢复 WAL 文件
- 将调试输出写入 `standby.log`
- 每隔 5 秒检查一次下一个 WAL 文件是否可用
- 仅在名为 `C:\pgsql.trigger.5442` 的触发文件出现时停止等待，并根据其内容执行故障切换
- 恢复结束时删除触发文件
- 从归档目录中移除不再需要的文件

Windows 上的 `copy` 命令会在文件复制完成之前设置最终文件大小，这通常会让 pg_standby 产生误判。因此，pg_standby 一旦看到正确的文件大小，还会等待 `sleeptime` 秒。GNUWin32 的 `cp` 只有在文件复制完成后才设置文件大小。

由于 Windows 示例在两端都使用 `copy`，任一服务器或两个服务器都可能通过网络访问归档目录。

F.31.4. Supported Server Versions

pg_standby 设计为与 PostgreSQL 8.2 及更高版本配合使用。

PostgreSQL 8.3 提供了 `%r` 宏，用于让 pg_standby 知道需要保留的最后一个文件。使用 PostgreSQL 8.2 时，如果需要清理归档，就必须使用 `-k` 选项。该选项在 8.3 中仍然可用，但已弃用。

PostgreSQL 8.4 提供了 `recovery_end_command` 选项。如果没有此选项，遗留的触发文件可能带来危险。

F.31.5. 作者

Simon Riggs `<simon@2ndquadrant.com>`

F.32. pg_stat_statements

pg_stat_statements模块提供了一种机制，用于跟踪服务器执行的所有 SQL 语句的执行统计信息。

由于该模块需要额外的共享内存，因此必须通过在postgresql.conf中将pg_stat_statements加入shared_preload_libraries来加载该模块。
这意味着添加或移除此模块都需要重启服务器。

F.32.1. pg_stat_statements 视图

该模块收集的统计信息可通过名为 pg_stat_statements 的系统视图获取。该视图为每个不同的查询文本、数据库 ID 和用户 ID 各包含一行（最多达到该模块能够跟踪的不同语句数量上限）。视图的列见表 F.22。

表 F.22. pg_stat_statements 列

Name	Type	References	Description
userid	oid	pg_authid.oid	OID of user who executed the statement
dbid	oid	pg_database.oid	OID of database in which the statement was executed
query	text		Text of the statement (up to track_activity_query_size bytes)
calls	bigint		Number of times executed
total_time	double precision		Total time spent in the statement, in seconds
rows	bigint		Total number of rows retrieved or affected by the statement
shared_blks_hit	bigint		Total number of shared blocks hits by the statement
shared_blks_read	bigint		Total number of shared blocks reads by the statement
shared_blks_written	bigint		Total number of shared blocks writes by the statement
local_blks_hit	bigint		Total number of local blocks hits by the statement
local_blks_read	bigint		Total number of local blocks reads by the statement
local_blks_written	bigint		Total number of local blocks writes by the statement

Name	Type	References	Description
temp_blks_read	bigint		Total number of temp blocks reads by the statement
temp_blks_written	bigint		Total number of temp blocks writes by the statement

该视图以及函数pg_stat_statements_reset只有在通过安装pg_stat_statements扩展而专门安装了它们的数据库中才可用。不过，只要pg_stat_statements模块被装载到服务器中，无论视图是否存在，它都会跟踪该服务器上所有数据库的统计信息。

出于安全原因，非超级用户不允许查看其他用户执行的查询的文本。不过，只要该视图已经安装在其数据库中，他们仍然可以查看统计信息。

注意，只要文本相同，语句就被视为相同的语句，无论语句中使用的任何离线参数的值如何。使用离线参数有助于把语句分组，从而可能使统计信息更有用。

F.32.2. 函数

pg_stat_statements_reset() returns void

pg_stat_statements_reset会丢弃到目前为止由pg_stat_statements收集的所有统计信息。默认情况下，只有超级用户可以执行此函数。

F.32.3. 配置参数

pg_stat_statements.max (integer)

pg_stat_statements.max 是该模块跟踪的语句最大数量（即 pg_stat_statements 视图中的最大行数）。如果观察到的不同语句超过该数量，则执行次数最少的语句信息会被丢弃。默认值为 1000。该参数只能在服务器启动时设置。

pg_stat_statements.track (enum)

pg_stat_statements.track 控制该模块统计哪些语句。指定 top 可跟踪顶层语句（即直接由客户端发出的语句），指定 all 还会跟踪嵌套语句（例如在函数内调用的语句），而指定 none 则会禁用。默认值为 top。只有超级用户可以更改此设置。

pg_stat_statements.track_utility (boolean)

pg_stat_statements.track_utility 控制该模块是否跟踪实用命令。实用命令是除 SELECT、INSERT、UPDATE 和 DELETE 之外的所有命令。默认值为 on。只有超级用户可以更改此设置。

pg_stat_statements.save (boolean)

pg_stat_statements.save 指定是否在服务器关闭后保留语句统计信息。如果它为 off，则在关闭时不会保存统计信息，并且在服务器启动时也不会重新载入这些统计信息。默认值为 on。该参数只能在 postgresql.conf 文件中或在服务器命令行上设置。

该模块需要约 pg_stat_statements.max * track_activity_query_size字节的额外共享内存。注意，只要该模块被装载，就会消耗这些内存，即使 pg_stat_statements.track 被设为 none。

为了能在 `postgresql.conf` 文件中设置这些参数，你需要把 `pg_stat_statements` 加入custom_variable_classes。典型用法如下：

```
# postgresql.conf
shared_preload_libraries = 'pg_stat_statements'

custom_variable_classes = 'pg_stat_statements'
pg_stat_statements.max = 10000
pg_stat_statements.track = all
```

F.32.4. 示例输出

```
bench=# SELECT pg_stat_statements_reset();

$ pgbench -i bench
$ pgbench -c10 -t300 -M prepared bench

bench=# \x
bench=# SELECT query, calls, total_time, rows, 100.0 * shared_blks_
hit /
               nullif(shared_blks_hit + shared_blks_read, 0) AS hit_
percent
          FROM pg_stat_statements ORDER BY total_time DESC LIMIT 5;
-[ RECORD 1 ]-----------------------------------------------------------
------------
query       | UPDATE pgbench_branches SET bbalance = bbalance + $1
 WHERE bid = $2;
calls       | 3000
total_time  | 9.60900100000002
rows        | 2836
hit_percent | 99.9778970000200936
-[ RECORD 2 ]-----------------------------------------------------------
------------
query       | UPDATE pgbench_tellers SET tbalance = tbalance + $1
 WHERE tid = $2;
calls       | 3000
total_time  | 8.015156
rows        | 2990
hit_percent | 99.9731126579631345
-[ RECORD 3 ]-----------------------------------------------------------
------------
query       | copy pgbench_accounts from stdin
calls       | 1
total_time  | 0.310624
rows        | 100000
hit_percent | 0.30395136778115501520
-[ RECORD 4 ]-----------------------------------------------------------
------------
query       | UPDATE pgbench_accounts SET abalance = abalance + $1
 WHERE aid = $2;
calls       | 3000
total_time  | 0.271741999999997
rows        | 3000
hit_percent | 93.7968855088209426
```

```
-[ RECORD 5 ]-----------------------------------------------------------
------------
query       | alter table pgbench_accounts add primary key (aid)
calls       | 1
total_time  | 0.08142
rows        | 0
hit_percent | 34.4947735191637631
```

F.32.5. Author

Takahiro Itagaki <itagaki.takahiro@oss.ntt.co.jp>

F.33. pgstattuple – 获取元组级统计信息

`pgstattuple` 模块提供了多种用于获取元组级统计信息的函数。

F.33.1. 函数

`pgstattuple(text) returns record`

`pgstattuple` 返回一个关系的物理长度、“死”元组所占百分比以及其他信息。这可以帮助用户判断是否需要执行清理。参数是目标关系的名称（可选带模式限定）。例如：

```
test=> SELECT * FROM pgstattuple('pg_catalog.pg_proc');
-[ RECORD 1 ]------+-------
table_len          | 458752
tuple_count        | 1470
tuple_len          | 438896
tuple_percent      | 95.67
dead_tuple_count   | 11
dead_tuple_len     | 3157
dead_tuple_percent | 0.69
free_space         | 8932
free_percent       | 1.95
```

输出列在表 F.23中描述。

表 F.23. `pgstattuple` 输出列

Column	Type	Description
`table_len`	`bigint`	Physical relation length in bytes
`tuple_count`	`bigint`	Number of live tuples
`tuple_len`	`bigint`	Total length of live tuples in bytes
`tuple_percent`	`float8`	Percentage of live tuples
`dead_tuple_count`	`bigint`	Number of dead tuples
`dead_tuple_len`	`bigint`	Total length of dead tuples in bytes
`dead_tuple_percent`	`float8`	Percentage of dead tuples
`free_space`	`bigint`	Total free space in bytes
`free_percent`	`float8`	Percentage of free space

`pgstattuple` 只会在该关系上获取读锁。因此，结果并不代表一个瞬时快照；并发更新会影响结果。

如果 `HeapTupleSatisfiesDirty` 返回 false，`pgstattuple` 就会将该元组判定为“死”。

`pgstattuple(oid) returns record`

它与 `pgstattuple(text)` 相同，区别在于目标是用 OID 指定的关系。

`pgstatindex(text) returns record`

`pgstatindex` 返回一条记录，显示有关 B-树索引的信息。例如：

```
test=> SELECT * FROM pgstatindex('pg_cast_oid_index');
-[ RECORD 1 ]------+------
version            | 2
tree_level         | 0
index_size         | 16384
root_block_no      | 1
internal_pages     | 0
leaf_pages         | 1
empty_pages        | 0
deleted_pages      | 0
avg_leaf_density   | 54.27
leaf_fragmentation | 0
```

输出列如下：

列	类型	描述
`version`	`integer`	B-树版本号
`tree_level`	`integer`	根页所在的树层级
`index_size`	`bigint`	索引总大小（字节）
`root_block_no`	`bigint`	根页的位置（若无则为零）
`internal_pages`	`bigint`	“内部”（上层）页的数量
`leaf_pages`	`bigint`	叶子页数量
`empty_pages`	`bigint`	空页数量
`deleted_pages`	`bigint`	已删除页数量
`avg_leaf_density`	`float8`	叶子页平均密度
`leaf_fragmentation`	`float8`	叶子页碎片化程度

报告的 `index_size` 通常会比 `internal_pages + leaf_pages + empty_pages + deleted_pages` 所对应的页数多 1 页，因为它还包含索引的元页。

与 `pgstattuple` 一样，结果是逐页累积得到的，因此不应期望它表示整个索引的瞬时快照。

`pg_relpages(text) returns bigint`

`pg_relpages` 返回该关系中的页数。

F.33.2. 作者

Tatsuo Ishii 和 Satoshi Nagayasu

F.34. pg_test_fsync

pg_test_fsync 旨在让你大致了解在你的特定系统上哪种wal_sync_method最快，并在出现已识别的 I/O 问题时提供诊断信息。However, differences shown by pg_test_fsync might not make any difference in real database throughput, especially since many database servers are not speed-limited by their transaction logs.

F.34.1. Usage

```
pg_test_fsync [options]
```

pg_test_fsync 接受以下命令行选项：

`-f`
`--filename`

指定用于写入测试数据的文件名。该文件应位于 `pg_xlog`目录所在或将要放置的同一文件系统中。（`pg_xlog`包含WAL文件。）默认值是当前目录中的`pg_test_fsync.out`。

`-o`
`--ops-per-test`

指定每次测试的操作数。每次测试的操作数越多，测试的准确性越高，但运行时间也越长。默认值为 2000。

F.34.2. 作者

Bruce Momjian `<bruce@momjian.us>`

F.35. pg_trgm

pg_trgm 模块提供函数和操作符，用于基于三元组匹配确定 ASCII 字母数字文本的相似度，同时还提供支持快速搜索相似字符串的索引操作符类。

F.35.1. 三字符组（Trigram 或 Trigraph）概念

三字符组是一组从字符串中取出的三个连续字符。
我们可以通过统计两个字符串共享的三字符组数量来度量它们的相似度。
这个简单的思想在度量许多自然语言中词的相似度时都非常有效。

> **注意**
>
> 在确定字符串所包含的三元组集合时，认为字符串前面有两个空格，后面有一个空格。例如，字符串“cat”的三元组集合是“ c”、“ ca”、“cat”和“at ”.

F.35.2. 函数和操作符

pg_trgm模块提供的函数列在表 F.24中，操作符列在表 F.25中。

表 F.24. pg_trgm函数

Function	Returns	Description
similarity(text, text)	real	Returns a number that indicates how similar the two arguments are. The range of the result is zero (indicating that the two strings are completely dissimilar) to one (indicating that the two strings are identical).
show_trgm(text)	text[]	Returns an array of all the trigrams in the given string. (In practice this is seldom useful except for debugging.)
show_limit()	real	Returns the current similarity threshold used by the % operator. This sets the minimum similarity between two words for them to be considered similar enough to be misspellings of each other, for example.
set_limit(real)	real	Sets the current similarity threshold that is used by the % operator. The threshold must be between 0 and 1 (default is 0.3). Returns the same value passed in.

表 F.25. pg_trgm操作符

Operator	Returns	Description
`text % text`	`boolean`	Returns `true` if its arguments have a similarity that is greate r than the current similarity threshold set by `set_limit`.
`text <-> text`	`real`	Returns the "distance" betwee n the arguments, that is one minus the `similarity()` value.

F.35.3. 索引支持

`pg_trgm` 模块提供了 GiST 和 GIN 索引操作符类，允许你在文本列上建立索引，用于极快的相似度搜索。These index types support the above-described similarity operators, and additionally support trigram-based index searches for `LIKE` and `ILIKE` queries. (These indexes do not support equality nor simple comparison operators, so you may need a regular B-tree index too.)

示例：

```
CREATE TABLE test_trgm (t text);
CREATE INDEX trgm_idx ON test_trgm USING gist (t gist_trgm_ops);
```

或者

```
CREATE INDEX trgm_idx ON test_trgm USING gin (t gin_trgm_ops);
```

此时，你已经在`t`列上有了一个可用于相似度搜索的索引。典型查询如下：

```
SELECT t, similarity(t, 'word') AS sml
  FROM test_trgm
  WHERE t % 'word'
  ORDER BY sml DESC, t;
```

这将返回文本列中所有与以下词足够相似的值：*word*，按从最佳匹配到最差匹配的顺序排序。即使在非常大的数据集上，索引也会让这一操作保持高效。

上述查询的一种变体是：

```
SELECT t, t <-> 'word' AS dist
  FROM test_trgm
  ORDER BY dist LIMIT 10;
```

GiST 索引可以相当高效地实现这一点，但 GIN 索引不能。当只需要少量最接近的匹配项时，它通常会优于第一种写法。

从PostgreSQL9.1 起，这些索引类型还支持以下操作的索引搜索：`LIKE`以及`ILIKE`，例如：

```
SELECT * FROM test_trgm WHERE t LIKE '%foo%bar';
```

索引搜索的工作方式是从搜索字符串中提取三字符组，然后在索引中查找这些三字符组。搜索字符串中包含的三字符组越多，索引搜索就越有效。与基于 B-树的搜索不同，搜索字符串不需要在左端锚定。

GiST 和 GIN 索引之间如何取舍，取决于二者各自的相对性能特征；相关讨论见其他章节。作为经验法则，GIN 索引的搜索速度快于 GiST 索引，但构建或更新较慢；因此 GIN 更适合静态数据，而 GiST 更适合经常更新的数据。

F.35.4. 文本检索集成

与全文索引结合使用时，三字符组匹配是非常有用的工具。尤其是，
它有助于识别那些因拼写错误而无法被全文检索机制直接匹配的输入词。

第一步是生成一个辅助表，其中包含文档中的全部唯一词：

```
CREATE TABLE words AS SELECT word FROM
        ts_stat('SELECT to_tsvector(''simple'', bodytext) FROM
 documents');
```

其中`documents`是一个表，包含我们希望搜索的文本字段`bodytext`。之所以对`to_tsvector`函数使用`simple`配置，而不是使用特定语言的配置，是因为我们需要原始的（未经词干提取的）词列表。

接下来，在词列上创建一个三字符组索引：

```
CREATE INDEX words_idx ON words USING gin (word gin_trgm_ops);
```

现在，可以使用与前面示例类似的`SELECT`查询，为用户搜索词中拼错的单词提供拼写建议。
一个有用的附加测试是要求选出的词长度也与该拼错单词相近。

> **注意**
>
> 由于`words`表是作为一张独立的静态表生成的，因此需要定期重新生成，
> 以便与文档集合保持大致同步。通常没有必要让它始终保持精确同步。

F.35.5. 参考

GiST 开发站点 http://www.sai.msu.su/~megera/postgres/gist/

Tsearch2 开发站点 http://www.sai.msu.su/~megera/postgres/gist/tsearch/V2/

F.35.6. 作者

Oleg Bartunov `<oleg@sai.msu.su>`，俄罗斯莫斯科，莫斯科大学

Teodor Sigaev `<teodor@sigaev.ru>`，俄罗斯莫斯科，Delta-Soft Ltd.

文档：Christopher Kings-Lynne

本模块由俄罗斯莫斯科的 Delta-Soft Ltd. 赞助。

F.36. pg_upgrade

pg_upgrade（以前称为 pg_migrator）允许将存储在 PostgreSQL 数据文件中的数据升级到更新的 PostgreSQL主版本，而无需执行主版本升级通常所需的数据转储/重新装载，例如从 8.4.7 升级到 PostgreSQL 的当前主版本。次版本升级不需要使用它，例如从 9.0.1 升级到 9.0.4。

PostgreSQL 主版本会定期增加新特性，这些特性常常会改变系统表的布局，但内部数据存储格式很少改变。pg_upgrade利用这一点，
通过创建新的系统表并直接重用旧的用户数据文件来快速完成升级。
如果未来某个主版本改变了数据存储格式，致使旧数据格式无法读取，那么 pg_upgrade就不能用于这类升级。（社区会尽量避免这种情况。）

pg_upgrade会尽最大努力确保新旧集簇在二进制上兼容，例如会检查兼容的编译时设置，包括 32/64 位二进制程序。任何外部模块也同样必须二进制兼容，这一点很重要，但 pg_upgrade 无法检查。

F.36.1. 支持的版本

pg_upgrade 支持从 8.3.X 及更高版本升级到当前 PostgreSQL主版本，包括快照版和 alpha 版。

F.36.2. pg_upgrade 选项

pg_upgrade 接受下列命令行参数：

`-b` *`old_bindir`*
`--old-bindir=`*`old_bindir`*

旧集簇的可执行文件目录；环境变量 `OLDBINDIR`

`-B` *`new_bindir`*
`--new-bindir=`*`new_bindir`*

新集簇的可执行文件目录；环境变量 `NEWBINDIR`

`-c`
`--check`

只检查集簇，不更改任何数据

`-d` *`old_datadir`*
`--old-datadir=`*`old_datadir`*

旧集簇的数据目录；环境变量 `OLDDATADIR`

`-D` *`new_datadir`*
`--new-datadir=`*`new_datadir`*

新集簇的数据目录；环境变量 `NEWDATADIR`

`-g`
`--debug`

启用调试

`-G` *`debug_filename`*
`--debugfile=`*`debug_filename`*

将调试活动输出到文件

`-k`
`--link`

使用硬链接而不是将文件复制到新集簇

`-l` *`log_filename`*
`--logfile=`*`log_filename`*

将会话活动记录到文件

`-p` *`old_port_number`*
`--old-port=`*`old_portnum`*

旧集簇的端口号；环境变量 `PGPORT`

`-P` *`new_port_number`*
`--new-port=`*`new_portnum`*

新集簇的端口号；环境变量 `PGPORT`

`-u` *`user_name`*
`--user=`*`user_name`*

集簇超级用户名称；环境变量 `PGUSER`

`-v`
`--verbose`

启用详细输出

`-V`
`--version`

显示版本信息，然后退出

`-?`
`-h`
`--help`

显示帮助，然后退出

F.36.3. 升级步骤

1. **移动旧集簇（可选）**

   如果你使用的是带版本号的安装目录，例如 `/opt/PostgreSQL/8.4`，则无需移动旧集簇。一键安装程序都使用带版本号的安装目录。

   如果你的安装目录不是带版本号的，例如 `/usr/local/pgsql`，就必须移动当前的 PostgreSQL 安装目录，以免它干扰新的 PostgreSQL 安装。当前的 PostgreSQL 服务器关闭后，就可以安全地重命名 PostgreSQL 安装目录。假设旧目录是 `/usr/local/pgsql`，你可以这样做：

   ```
   mv /usr/local/pgsql /usr/local/pgsql.old
   ```

以上命令会重命名该目录。

2. **对于源码安装，编译新版本**

   使用与旧集簇兼容的 `configure` 标志编译新的 PostgreSQL 源码。pg_upgrade 会在开始升级前检查 `pg_controldata`，以确保所有设置都兼容。

3. **安装新的 PostgreSQL 二进制文件**

   安装新服务器的二进制文件和支持文件。两个集簇可以使用相同的端口号（通常是 5432），因为新旧集簇不会同时运行。

   对于源码安装，如果你希望把新服务器安装到自定义位置，可以使用 `prefix` 变量：

   ```
   gmake prefix=/usr/local/pgsql.new install
   ```

4. **安装 pg_upgrade 和 pg_upgrade_support**

   把 pg_upgrade 二进制和 pg_upgrade_support 库安装到新的 PostgreSQL 集簇中。

5. **初始化新的 PostgreSQL 集簇**

   使用 `initdb` 初始化新集簇。这里同样要使用与旧集簇匹配的兼容 `initdb` 标志。许多预构建的安装程序会自动执行该步骤。无需启动新集簇。

6. **安装自定义共享对象文件**

   将旧集簇使用的所有自定义共享对象文件（或 DLL）安装到新集簇中，例如 `pgcrypto.so`，无论它们来自 `contrib` 还是其他来源。不要安装模式定义，例如 `pgcrypto.sql`，因为这些会从旧集簇升级过来。

7. **调整认证**

   `pg_upgrade` 会多次连接旧服务器和新服务器，因此你可能希望将认证设置为 `peer`（在 `pg_hba.conf` 中），或者使用 `~/.pgpass` 文件（见第 31.14 节）。

8. **停止两个服务器**

   确保两个数据库服务器都已停止。例如，在 Unix 上使用：

   ```
   pg_ctl -D /opt/PostgreSQL/8.4 stop
   pg_ctl -D /opt/PostgreSQL/9.0 stop
   ```

   或者在 Windows 上使用正确的服务名：

   ```
   NET STOP postgresql-8.4
   NET STOP postgresql-9.0
   ```

   或

   `NET STOP pgsql-8.3`（`PostgreSQL 8.3` 及更早版本使用不同的服务名）

9. **运行 pg_upgrade**

   始终运行新服务器的 pg_upgrade 二进制，而不是旧服务器的。pg_upgrade 需要指定新旧集簇的数据目录和可执行文件（`bin`）目录。你还可以指定用户和端口值，
   以及是否希望链接数据而不是复制（默认）。

如果使用链接模式，升级将快得多（无需复制文件），但一旦在升级后启动新集簇，就无法再访问旧集簇。链接模式还要求新旧集簇的数据目录位于同一文件系统中。完整选项列表请参阅 `pg_upgrade --help`。

对于 Windows 用户，必须以管理员账户登录，然后以 `postgres` 用户身份启动一个 shell，并设置正确的路径：

```
RUNAS /USER:postgres "CMD.EXE"
SET PATH=%PATH%;C:\Program Files\PostgreSQL\9.0\bin;
```

然后给目录参数加上引号，再运行 pg_upgrade，例如：

```
pg_upgrade.exe
        --old-datadir "C:/Program Files/PostgreSQL/8.4/data"
        --new-datadir "C:/Program Files/PostgreSQL/9.0/data"
        --old-bindir "C:/Program Files/PostgreSQL/8.4/bin"
        --new-bindir "C:/Program Files/PostgreSQL/9.0/bin"
```

启动之后，`pg_upgrade` 会验证两个集簇是否兼容，然后执行升级。即使旧服务器仍在运行，也可以使用 `pg_upgrade --check` 只执行检查。`pg_upgrade --check` 还会概述升级后需要手工调整的内容。`pg_upgrade` 要求对当前目录有写权限。

显然，升级期间不应有人访问集簇。考虑为旧集簇和新集簇使用非默认端口号（例如 50432），以避免升级期间意外的客户端连接。

如果在恢复数据库模式时发生错误，`pg_upgrade` 将退出，你必须按照下文步骤 14所述回退到旧集簇。若要再次尝试 `pg_upgrade`，你需要修改旧集簇，使 pg_upgrade 的模式恢复步骤能够成功完成。如果问题出在某个 contrib 模块，而该模块并未用于存储用户数据，则你可能需要先从旧集簇卸载这个 contrib 模块，并在升级后将其安装到新集簇。

10. **恢复 `pg_hba.conf`**

    如果修改过 `pg_hba.conf`，请恢复其原始设置。

11. **升级后处理**

    如果需要任何升级后处理，pg_upgrade 在完成时会发出警告。它还会生成必须由管理员运行的脚本文件。这些脚本会连接到每个需要升级后处理的数据库。
    每个脚本应通过以下方式运行：

    ```
    psql --username postgres --file script.sql postgres
    ```

    这些脚本可以按任意顺序运行，运行完毕后即可删除。

    > **小心**
    >
    > 一般来说，在重建脚本运行完成之前访问其中引用的表是不安全的；
    > 这样做可能产生错误结果或较差的性能。未在重建脚本中引用的表则可以立即访问。

12. **统计信息**

    由于优化器统计信息不会被 `pg_upgrade` 传送，在升级结束时你会被指示运行一个命令重新生成这些信息。

13. **删除旧集簇**

    对升级结果满意后，可以运行 `pg_upgrade` 完成时提到的脚本删除旧集簇的数据目录。还可以删除旧的安装目录（例如 `bin`、`share`）。

14. **恢复到旧集簇**

    如果在运行 `pg_upgrade` 之后希望回退到旧集簇，有以下几种选择：

    - 如果你运行 `pg_upgrade` 时使用了 `--check`，旧集簇未做任何修改，可以随时重新使用它。
    - 如果你运行 `pg_upgrade` 时使用了 `--link`，数据文件由新旧集簇共享。如果你已经启动过新集簇，新服务器已写入这些共享文件，再使用旧集簇就不安全了。
    - 如果你运行 `pg_upgrade` 时没有使用 `--link`，或者没有启动过新服务器，那么旧集簇除了在 `$PGDATA/global/pg_control` 以及可能还有表空间目录上追加了 `.old` 后缀之外未被修改。要重新使用旧集簇，请移除 `$PGDATA/global/pg_control` 上的 `.old` 后缀，并且如果是升级到 8.4 或更早版本，还要删除升级创建的表空间目录并移除表空间目录名上的 `.old` 后缀；然后就可以重启旧集簇。

F.36.4. 从 PostgreSQL 8.3 升级时的限制

从 PostgreSQL 8.3 升级存在从更高 PostgreSQL 版本升级时所没有的额外限制。例如，如果有用户列定义为下列类型，pg_upgrade 将无法用于从 8.3 升级：

- `tsquery` 数据类型
- `name` 数据类型且不是第一列

你必须删除所有这样的列并手工升级它们。

如果数据库中安装了`ltree` contrib 模块，pg_upgrade 将无法工作。

如果出现下列情况，pg_upgrade 将要求重建表：

- 某个用户列的数据类型为`tsvector`

如果出现下列情况，pg_upgrade 将要求重新索引：

- 某个索引为 hash 或 GIN 类型
- 某个索引使用了`bpchar_pattern_ops`

此外，在PostgreSQL 8.3 之后，默认的日期时间存储格式改为了整数。pg_upgrade 会检查新旧集簇使用的日期时间存储格式是否匹配。请确保你的新集簇在构建时使用了 configure 选项`--disable-integer-datetimes`。

对于 Windows 用户，请注意，由于一键安装程序和 MSI 安装程序使用的整数日期时间设置不同，只能从一键发行版的 8.3 版本升级到一键发行版的 8.4 或更高版本。无法从 MSI 安装程序升级到一键安装程序。

F.36.5. 注解

pg_upgrade不支持升级包含下列`reg*` OID 引用系统数据类型的数据库：`regproc`、`regprocedure`、`regoper`、`regoperator`、`regconfig`和`regdictionary`。（`regtype`可以升级。）

所有失败、重建和重新索引的情形，只要影响你的安装，都会被 pg_upgrade 报告；用于重建表和索引的升级后脚本会自动生成。

为了测试部署，请创建旧集簇的纯模式副本，插入虚拟数据，然后对其升级。

如果你想使用链接模式，又不希望旧集簇在新集簇启动后被修改，
可以先制作旧集簇的一个副本，然后用链接模式升级该副本。要制作一个有效的旧集簇副本，
可以在服务器运行时用 `rsync` 创建旧集簇的“脏”副本，然后关闭旧服务器并再次运行 `rsync`，
用所有变更更新副本使其保持一致。

F.37. seg – 用于线段或浮点区间的数据类型

该模块实现了一个用于表示线段或浮点区间的`seg`数据类型。`seg`
可以表示区间端点中的不确定性，因此特别适合表示实验室测量结果。

F.37.1. 原理

测量值的几何形态通常比数值连续体中的一个点更复杂。一次测量通常是该连续体上的一段，
其边界多少有些模糊。测量结果之所以会呈现为区间，一方面是由于不确定性和随机性，
另一方面也是因为被测值本身可能天然就是表示某种状态的区间，例如蛋白质的稳定温度范围。

凭常识也能看出，把这类数据存储为区间比存储为一对数字更方便。实际上，在大多数应用中，
这样做甚至更高效。

再顺着这一常识往下想，边界的模糊性说明，使用传统数值数据类型会造成一定的信息损失。
设想一下：你的仪器读数是 6.50，而你把这个读数输入到数据库。取出来时会得到什么？请看：

```
test=> select 6.50 :: float8 as "pH";
 pH
---
6.5
(1 row)
```

在测量领域，6.50 与 6.5 并不相同。有时这种差别至关重要。实验人员通常会记下（并发表）
他们认为可靠的那些数位。6.50 实际上是一个模糊区间，它包含在更大、也更模糊的区间 6.5 之
中；它们共享的特征（大概）只有中心点。我们当然不希望这类不同的数据项看起来却一样。

结论是什么？最好能有一种专门的数据类型，能够以任意可变的精度记录区间边界。
这里所谓“可变”，是指每个数据元素都记录其自身的精度。

看看这个：

```
test=> select '6.25 .. 6.50'::seg as "pH";
          pH
------------
6.25 .. 6.50
(1 row)
```

F.37.2. 语法

区间的外部表示由一个或两个浮点数通过范围操作符（`..` 或 `...`）连接而成。另一种写法是指定
中心点再加减一个偏差值。还可以存储可选的确定性指示符（<、>或 ~）。（不过，所有内置操
作符都会忽略这些确定性指示符。）表 F.26概述了允许的表示形式；表 F.27给出了一些示例。

在表 F.26中，*x*、*y*和*delta*表示浮点数。*x*和*y*前面可以带确定性指示符，而*delta*不可以。

表 F.26. seg 外部表示

x	单个值（零长度区间）
x `..` *y*	从*x*到*y*的区间
x `(+-)` *delta*	从*x* - *delta*到 *x* + *delta*的区间
x `..`	下界为*x*、上界开放的区间

`.. x`	上界为*x*、下界开放的区间

表 F.27. 合法 seg 输入示例

`5.0`	创建一个零长度线段（如果你愿意，也可以把它看成一个点）
`~5.0`	创建一个零长度线段，并在数据中记录`~`。`~`会被`seg`操作忽略，但会作为注释保留下来。
`<5.0`	在 5.0 处创建一个点。`<`会被忽略，但会作为注释保留下来。
`>5.0`	在 5.0 处创建一个点。`>`会被忽略，但会作为注释保留下来。
`5(+-)0.3`	创建区间`4.7 .. 5.3`。注意，`(+-)`记法不会被保留。
`50 ..`	所有大于或等于 50 的值
`.. 0`	所有小于或等于 0 的值
`1.5e-2 .. 2E-2`	创建区间`0.015 .. 0.02`
`1 ... 2`	与`1...2`、`1 .. 2`或`1..2`相同（范围操作符两侧的空格会被忽略）

由于`...`操作符在数据源中被广泛使用，因此允许把它作为 `..`操作符的另一种写法。不幸的是，这会带来解析歧义：无法确定`0...23`中的上界是`23`还是 `0.23`。解决办法是要求`seg`输入中的所有数字在小数点前至少有一位数字。

作为合理性检查，`seg`会拒绝下界大于上界的区间，例如 `5 .. 2`。

F.37.3. 精度

`seg`值在内部存储为一对 32 位浮点数。这意味着有效位数超过 7 位的数字会被截断。

有效位数不超过 7 位的数字会保留其原始精度。也就是说，如果你的查询返回 0.00，你可以确信末尾的零不是格式化造成的假象，而是反映了原始数据的精度。前导零的个数不影响精度：值 0.0067 被认为只有 2 位有效数字。

F.37.4. 用法

`seg`模块包含适用于`seg`值的一个 GiST 索引操作符类。该 GiST 操作符类支持的操作符见表 F.28。

表 F.28. seg GiST 操作符

操作符	描述
`[a, b] << [c, d]`	[a, b] 完全位于 [c, d] 的左侧。也就是说，当 b < c 时，[a, b] << [c, d] 为真，否则为假。
`[a, b] >> [c, d]`	[a, b] 完全位于 [c, d] 的右侧。也就是说，当 a > d 时，[a, b] >> [c, d] 为真，否则为假。
`[a, b] &< [c, d]`	重叠或位于左侧 – 也可以更好地理解为“不会延伸到右侧”。当 b <= d 时为真。
`[a, b] &> [c, d]`	重叠或位于右侧 – 也可以更好地理解为“不会延伸到左侧”。当 a >= c 时为真。
`[a, b] = [c, d]`	相同 – 线段 [a, b] 和 [c, d] 相同，即 a = c 且 b = d。

操作符	描述
`[a, b] && [c, d]`	线段 [a, b] 和 [c, d] 重叠。
`[a, b] @> [c, d]`	线段 [a, b] 包含线段 [c, d]，即 a <= c 且 b >= d。
`[a, b] <@ [c, d]`	线段 [a, b] 包含于 [c, d] 中，即 a >= c 且 b <= d。

（在 PostgreSQL 8.2 之前，包含操作符 `@>` 和 `<@` 分别称为 `@` 和 `~`。这些名称仍然可用，但已弃用，最终将被删除。请注意，旧名称与核心几何数据类型以前采用的约定正好相反！）

还提供标准的 B-树操作符，例如

操作符	描述
`[a, b] < [c, d]`	小于
`[a, b] > [c, d]`	大于

这些操作符除了排序之外，在实际用途上没有太大意义。它们首先比较 (a) 与 (c)，如果相等，再比较 (b) 与 (d)。这在大多数情况下会产生相当好的排序效果；如果希望对这种类型使用 ORDER BY，这会很有用。

F.37.5. 注意

有关用法示例，请参见回归测试`sql/seg.sql`。

把`(+-)`转换为常规范围的机制，在确定边界的有效位数时并不完全准确。例如，
如果结果区间包含 10 的幂，它会在下边界上多加一位：

```
postgres=> select '10(+-)1'::seg as seg;
      seg
---------
9.0 .. 11             -- should be: 9 .. 11
```

R 树索引的性能很大程度上取决于输入值的初始顺序。按`seg`列对输入表排序，可能会很有帮助；示例见脚本 `sort-segments.pl`。

F.37.6. 致谢

原作者：Gene Selkov, Jr. `<selkovjr@mcs.anl.gov>`，阿贡国家实验室数学与计算机科学部。

我首先要感谢 Joe Hellerstein 教授 (http://db.cs.berkeley.edu/jmh/)，他为我阐明了 GiST (http://gist.cs.berkeley.edu/) 的要旨。我同样感谢过去和现在所有的 Postgres 开发者，他们使我得以创造自己的世界并在其中不受打扰地生活。我还要感谢阿贡实验室以及美国能源部，多年来始终如一地支持我的数据库研究。

F.38. sepgsql – 基于 SELinux 标签的强制访问控制（MAC）安全模块

`sepgsql` 是一个可加载模块，支持基于 SELinux 安全策略和安全标签的强制访问控制（MAC）。

> **警告**
>
> 当前实现存在重大限制，并不会对所有操作实施强制访问控制。详见第 F.38.6 节。

F.38.1. 概述

该模块与 SELinux 集成，在 PostgreSQL 通常提供的安全检查之外再增加一层安全检查。从 SELinux 的角度看，该模块使 PostgreSQL 能够充当用户空间对象管理器。由 DML 查询发起的每一次表或函数访问，都会依据系统安全策略进行检查。这种检查是对 PostgreSQL 常规 SQL 权限检查的补充。

SELinux 的访问控制决策通过安全标签作出，安全标签以 `system_u:object_r:sepgsql_table_t:s0` 这样的字符串表示。每个访问控制决策都涉及两个标签：尝试执行操作的主体标签，以及要执行该操作的对象标签。由于这些标签可以应用到任意类型的对象，数据库中存储对象的访问控制决策也可以与文件等其他类型对象一样，遵循同一套一般性标准；使用该模块时也确实如此。这样的设计旨在让集中式安全策略能够保护信息资产，而不受这些资产具体存储方式的影响。

SECURITY LABEL语句允许为数据库对象分配安全标签。

F.38.2. 安装

`sepgsql` 只能用于启用了 SELinux 的 Linux 2.6.28 或更高版本。它在任何其他平台上都不可用。You will also need libselinux 2.0.93 or higher and selinux-policy 3.9.13 or higher (although some distributions may backport the necessary rules into older policy versions).

可以使用 `sestatus` 命令检查 SELinux 的状态。典型显示如下：

```
$ sestatus
SELinux status:                 enabled
SELinuxfs mount:                /selinux
Current mode:                   enforcing
Mode from config file:          enforcing
Policy version:                 24
Policy from config file:        targeted
```

如果 SELinux 被禁用或尚未安装，则必须先安装并配置它，之后才能安装本模块。

要构建此模块，请将选项 `--with-selinux` 加入 PostgreSQL 的 `configure` 命令中。还要确保构建时已经安装 `libselinux-devel` RPM。

要使用此模块，必须将 `sepgsql` 包含在shared_preload_libraries参数中，该参数位于 `postgresql.conf` 内。如果以其他任何方式加载该模块，它都无法正确工作。模块加载后，应在每个数据库中执行 `sepgsql.sql`。这会安装安全标签管理所需的函数，并分配初始安全标签。

下面的示例展示了如何初始化一个全新的数据库集簇，并安装 sepgsql 函数和安全标签。请根据实际安装情况调整其中的路径：

```
$ export PGDATA=/path/to/data/directory
$ initdb
$ vi $PGDATA/postgresql.conf
  change
    #shared_preload_libraries = ''                # (change requires
 restart)
  to
    shared_preload_libraries = 'sepgsql'          # (change requires
 restart)
$ for DBNAME in template0 template1 postgres; do
    postgres --single -F -c exit_on_error=true $DBNAME \
      </usr/local/pgsql/share/contrib/sepgsql.sql >/dev/null
  done
```

请注意，具体会看到以下通知中的哪些，取决于所使用的 libselinux 和 selinux-policy 版本，可能是部分，也可能是全部：

```
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 33 has invalid
 object type db_blobs
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 36 has invalid
 object type db_language
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 37 has invalid
 object type db_language
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 38 has invalid
 object type db_language
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 39 has invalid
 object type db_language
/etc/selinux/targeted/contexts/sepgsql_contexts:  line 40 has invalid
 object type db_language
```

这些消息没有害处，应予忽略。

如果安装过程无错误完成，此时就可以正常启动服务器。

F.38.3. 回归测试

由于 SELinux 的特性，运行 sepgsql 的回归测试需要额外执行若干配置步骤，其中部分步骤必须以 root 身份完成。普通的 make check 或 make installcheck 命令不会运行这些回归测试；必须完成配置，然后手工调用测试脚本。这些测试必须在已配置的 PostgreSQL 构建树的 contrib/sepgsql 目录中运行。虽然需要构建树，但这些测试是为已安装的服务器设计的，也就是说，它们类似于 make installcheck 而不是 make check。

首先，在一个正常工作的数据库中设置 sepgsql，具体请按照第 F.38.2 节中的说明操作。注意，当前操作系统用户必须能够不经密码认证就以超级用户身份连接数据库。

其次，为回归测试构建并安装策略包。sepgsql-regtest 策略是一个特殊用途的策略包，它提供回归测试期间允许的一组规则。它应当从策略源文件 sepgsql-regtest.te 构建，使用 SELinux 提供的 Makefile 通过 make 完成。You will need to locate the appropriate Makefile on your system; the path shown below is only an example. Once built, install this policy package using the semodule command, which loads supplied policy packages into the kernel. If the package is correctly installed, semodule -l should list sepgsql-regtest as an available policy package:

```
$ cd .../contrib/sepgsql
$ make -f /usr/share/selinux/devel/Makefile
$ sudo semodule -u sepgsql-regtest.pp
$ sudo semodule -l | grep sepgsql
sepgsql-regtest 1.03
```

第三，打开 `sepgsql_regression_test_mode`。为了你系统的安全，默认不启用 `sepgsql-regtest` 中的所有规则。The `sepgsql_regression_test_mode` parameter enables the rules needed to launch the regression tests. It can be turned on using the `setsebool` command:

```
$ sudo setsebool sepgsql_regression_test_mode on
$ getsebool sepgsql_regression_test_mode
sepgsql_regression_test_mode --> on
```

第四，确认 shell 正在 `unconfined_t` 域中运行：

```
$ id -Z
unconfined_u:unconfined_r:unconfined_t:s0-s0:c0.c1023
```

如有需要，请参阅第 F.38.7 节，了解如何调整工作域。

最后，运行回归测试脚本：

```
$ ./test_sepgsql
```

该脚本会尝试验证是否已正确完成全部配置步骤，然后运行 `sepgsql` 模块的回归测试。

完成测试后，建议关闭 `sepgsql_regression_test_mode` 参数：

```
$ sudo setsebool sepgsql_regression_test_mode off
```

也可以选择彻底移除 `sepgsql-regtest` 策略：

```
$ sudo semodule -r sepgsql-regtest
```

F.38.4. GUC 参数

`sepgsql.permissive` (`boolean`)

该参数使 `sepgsql` 无论系统设置如何都以宽容模式运行。默认值为关闭。该参数只能在 `postgresql.conf` 文件中或服务器命令行上设置。

当该参数打开时，`sepgsql` 会以宽容模式运行，即使 SELinux 整体处于强制模式也是如此。该参数主要用于测试。

`sepgsql.debug_audit` (`boolean`)

该参数会在不考虑系统策略设置的情况下启用审计消息输出。默认值为关闭，这意味着消息将按系统设置输出。

SELinux 的安全策略本身也有规则来控制是否记录特定的访问。默认情况下，访问违例会被记录，而被允许的访问不会被记录。

该参数会在不考虑系统策略的情况下，强制打开所有可能的日志记录。

F.38.5. 特性

F.38.5.1. 受控对象类

SELinux 的安全模型把所有访问控制规则描述为主体实体（通常是数据库的客户端）与客体实体（例如数据库对象）之间的关系，二者各由一个安全标签标识。If access to an unlabelled object is attempted, the object is treated as if it were assigned the label `unlabeled_t`.

目前，`sepgsql` 允许把安全标签指派给模式、表、列、序列、视图和函数。When `sepgsql` is in use, security labels are automatically assigned to supported database objects at creation time. This label is called a default security label, and is decided according to the system security policy, which takes as input the creator's label and the label assigned to the new object's parent object.

新数据库对象基本上会继承父对象的安全标签，
但如果安全策略中存在称为类型转换规则的特殊规则，则可能会应用不同的标签。对于模式，
父对象是当前数据库；对于表、序列、视图和函数，父对象是其所在模式；对于列，
父对象是其所在表。

F.38.5.2. DML 权限

对于表，根据语句的种类，对所有被引用的目标表检查 `db_table:select`、`db_table:insert`、`db_table:update` 或 `db_table:delete`；此外，还对每条 DML 语句中引用的表检查 `db_table:select` is also checked for all the tables that contain the columns referenced in the `WHERE` or `RETURNING` clause, as a data source of `UPDATE`, and so on. For example, consider:

```
UPDATE t1 SET x = 2, y = md5sum(y) WHERE z = 100;
```

In this case we must have `db_table:select` in addition to `db_table:update`, because `t1.a` is referenced within the `WHERE` clause. Column-level permissions will also be checked for each referenced column.

对于列，不仅在用 `SELECT` 读取的列上检查 `db_column:select`，在其他 DML 语句引用的列上也做检查。Of course, it also checks `db_column:update` or `db_column:insert` on columns being modified by `UPDATE` or `INSERT`.

```
UPDATE t1 SET x = 2, y = md5sum(y) WHERE z = 100;
```

在这个例子中，对被更新的列 `t1.x` 检查 `db_column:update`；对既被更新又被引用的列 `t1.y` 检查 `db_column:{select update}`；对列 `t1.z` 检查 `db_column:select`，因为它只在 `WHERE` 子句中被引用。此外还会在表级检查 `db_table:{select update}`。

对于序列，当使用 `SELECT` 引用序列对象时，会检查 `db_sequence:get_value`；但请注意，当前不会检查执行相应函数（如 `lastval()`）的权限。

对于视图，会先检查 `db_view:expand`，然后对由该视图展开得到的各个对象分别检查所需的其他权限。

对于函数，虽然定义了 `db_procedure:{execute}`，但在这个版本中不进行检查。

客户端必须被允许访问所有被引用的表和列，
即使这些表和列最初来自随后被展开的视图也是如此，
这样我们才能在不受表内容引用方式影响的情况下，应用一致的访问控制规则。

默认数据库权限系统允许数据库超级用户使用 DML 命令修改系统目录，也允许引用或修改 TOAST 表。启用 `sepgsql` 后，这些操作都会被禁止。

F.38.5.3. DDL 权限

当执行SECURITY LABEL时，会在被重新标记的对象上用其旧安全标签检查 `setattr` 和 `relabelfrom`，然后用所提供的新安全标签检查 `relabelto`。

在安装了多个标签提供者且用户尝试设置一个不由 SELinux 管理的安全标签的情况下，这里只应检查 `setattr`。由于实现上的限制，目前没有这样做。

F.38.5.4. 受信任过程

可信过程类似于安全定义者函数或 set-uid 命令。SELinux 提供了一种特性，允许可信代码使用与客户端不同的安全标签运行，通常用于提供对敏感数据的高度受控访问（例如可以省略某些行，或降低所存储值的精度）。
一个函数是否充当可信过程由它的安全标签和操作系统安全策略控制。例如：

```
postgres=# CREATE TABLE customer (
               cid     int primary key,
               cname   text,
               credit  text
           );
CREATE TABLE
postgres=# SECURITY LABEL ON COLUMN customer.credit
               IS 'system_u:object_r:sepgsql_secret_table_t:s0';
SECURITY LABEL
postgres=# CREATE FUNCTION show_credit(int) RETURNS text
             AS 'SELECT regexp_replace(credit, ''-[0-9]+$'', ''-
xxxx'', ''g'')
                        FROM customer WHERE cid = $1'
           LANGUAGE sql;
CREATE FUNCTION
postgres=# SECURITY LABEL ON FUNCTION show_credit(int)
               IS 'system_u:object_r:sepgsql_trusted_proc_exec_t:s0';
SECURITY LABEL
```

上述操作应由管理用户执行。

```
postgres=# SELECT * FROM customer;
ERROR:  SELinux: security policy violation
postgres=# SELECT cid, cname, show_credit(cid) FROM customer;
 cid | cname  |     show_credit
-----+--------+---------------------
   1 | taro   | 1111-2222-3333-xxxx
   2 | hanako | 5555-6666-7777-xxxx
(2 rows)
```

在这种情况下，普通用户不能直接引用 `customer.credit`，但受信任过程 `show_credit` 允许他打印客户信用卡号，并对其中部分数字做掩码处理。

F.38.5.5. 杂项

我们一律拒绝LOAD命令，因为装入任意模块都可能轻易绕过安全策略的强制执行。

F.38.6. 限制

Data Definition Language (DDL) Permissions

由于实现上的限制，不检查 DDL 权限。

Data Control Language (DCL) Permissions

由于实现上的限制，不检查 DCL 权限。

Row-level access control

PostgreSQL 不支持行级访问；因此，`sepgsql` 也不支持。

Covert channels

`sepgsql` 不会试图隐藏某个对象的存在，即使用户无权引用它也是如此。例如，即便我们无法取得一个不可见对象的内容，也仍然可以从主键冲突、
外键违例等结果中推断出它的存在。绝密表的存在无法隐藏；我们只希望隐藏其内容。

F.38.7. 外部资源

SE-PostgreSQL Introduction[5]

该 wiki 页面提供了简要概述，并介绍了安全设计、体系结构、管理以及未来特性。

Fedora SELinux User Guide[6]

该文档提供了在系统上管理 SELinux 所需的广泛知识。它主要聚焦于 Fedora，但并不限于 Fedora。

Fedora SELinux FAQ[7]

该文档回答了关于 SELinux 的常见问题。它主要聚焦于 Fedora，但并不限于 Fedora。

F.38.8. 作者

KaiGai Kohei `<kaigai@ak.jp.nec.com>`

[5] http://wiki.postgresql.org/wiki/SEPostgreSQL
[6] http://docs.fedoraproject.org/selinux-user-guide/
[7] http://docs.fedoraproject.org/selinux-faq

F.39. spi

spi 模块提供了几个使用 SPI 和触发器的可行示例。虽然这些函数本身就有一定价值，但作为可按自己目的修改的示例，它们更有用。The functions are general enough to be used with any table, but you have to specify table and field names (as described below) while creating a trigger.

下文描述的每组函数都作为一个可单独安装的扩展提供。

F.39.1. refint – 用于实现引用完整性的函数

`check_primary_key()`和`check_foreign_key()`用于检查外键约束。（当然，这项功能早已被内置的外键机制取代，不过该模块作为示例仍然有用。）

`check_primary_key()` 检查引用表。使用时，在引用其他表的表上创建一个使用该函数的 `BEFORE INSERT OR UPDATE` 触发器。触发器参数依次为：引用表中构成外键的列名、被引用表名，以及被引用表中构成主键/唯一键的列名。若要处理多个外键，请为每个引用分别创建一个触发器。

`check_foreign_key()` 检查被引用表。使用时，在被其他表引用的表上创建一个使用该函数的 `BEFORE DELETE OR UPDATE` 触发器。触发器参数依次为：该函数需要检查的引用表数量、发现引用键时采取的动作（`cascade` 表示删除引用行，`restrict` 表示若存在引用键则中止事务，`setnull` 表示将引用键字段设为空）、被引用表中构成主键/唯一键的列名，然后是引用表名和列名（按第一个参数指定的引用表数量重复提供）。请注意，主键/唯一键列应标记为 NOT NULL，并具有唯一索引。

`refint.example` 中有示例。

F.39.2. timetravel – 实现时间旅行的函数

很久以前，PostgreSQL曾内置时间旅行功能，为每个元组保存插入和删除时间。可以使用这些函数模拟这一功能。要使用这些函数，必须为表添加两个`abstime`类型的列，以存储元组插入的时间（start_date）和更改/删除的时间（stop_date）：

```
CREATE TABLE mytab (
        ...             ...
        start_date      abstime,
        stop_date       abstime
        ...             ...
);
```

这些列可以任意命名，不过在这里我们将它们称为 start_date 和 stop_date。

插入新行时，start_date 通常应设置为当前时间，stop_date 应设置为`infinity`。如果插入的数据在这些列中包含空值，触发器会自动将其替换为上述值。一般来说，只有在重新载入转储数据时，才应向这些列中插入显式的非空数据。

stop_date 等于`infinity`的元组“当前有效”，可以修改。stop_date 为有限值的元组则不能再修改，触发器会阻止这种操作。（如果需要这样做，可以按下文所示关闭时间旅行。）

对于可修改的行，更新时只会将被更新元组的 stop_date 改为当前时间，并插入一个包含修改后数据的新元组。新元组的 start_date 将设置为当前时间，stop_date 设置为`infinity`。

删除操作实际上不会删除元组，而只会将其 stop_date 设置为当前时间。

要查询“当前有效”的元组，请在查询的 WHERE 条件中加入`stop_date = 'infinity'`。（你可能希望将其放入视图中。）类似地，通过为 start_date 和 stop_date 设置合适的条件，也可以查询在过去任何时间有效的元组。

`timetravel()`是支持此行为的通用触发器函数。在每个使用时间旅行的表上，使用此函数创建一个`BEFORE INSERT OR UPDATE OR DELETE`触发器。指定两个触发器参数：start_date 和 stop_date 列的实际名称。还可以选择指定一到三个额外参数，它们必须引用`text`类型的列。触发器会在 INSERT 时将当前用户名存入其中的第一列，在 UPDATE 时存入第二列，在 DELETE 时存入第三列。

`set_timetravel()`允许为表打开或关闭时间旅行。`set_timetravel('mytab', 1)`会为表`mytab`打开 TT。`set_timetravel('mytab', 0)`会为表`mytab`关闭 TT。两种情况都会报告原来的状态。在 TT 关闭时，可以自由修改 start_date 和 stop_date 列。请注意，开关状态仅作用于当前数据库会话；新会话总是从所有表的 TT 都打开的状态开始。

`get_timetravel()`返回表的 TT 状态，而不改变它。

`timetravel.example`中有一个示例。

F.39.3. autoinc – 用于字段自动递增的函数

`autoinc()`是一个将序列的下一个值存入整数字段的触发器。它与内置的“序列列”功能有一些重叠，但并不相同：`autoinc()`会覆盖插入时试图为该字段指定其他值的操作，而且还可以选择在更新时也递增该字段。

使用时，创建一个使用该函数的 `BEFORE INSERT`（或者可选的 `BEFORE INSERT OR UPDATE`）触发器。指定两个触发器参数：要修改的整型列名，以及提供这些值的序列对象名。（实际上，如果你希望更新多个自动递增列，也可以指定任意多组这样的名称对。）

`autoinc.example` 中有一个示例。

F.39.4. insert_username – 用于跟踪谁修改了表的函数

`insert_username()` 是一个将当前用户名存入文本字段的触发器。它有助于跟踪表中某一特定行最后一次是由谁修改的。

使用时，创建一个使用该函数的`BEFORE INSERT`和/或`UPDATE`触发器。指定一个触发器参数：要修改的文本列名。

`insert_username.example` 中有一个示例。

F.39.5. moddatetime – 用于跟踪最后修改时间的函数

`moddatetime()` 是一个将当前时间存入 `timestamp` 字段的触发器。它有助于跟踪表中某一特定行最后一次被修改的时间。

使用时，创建一个使用该函数的 `BEFORE UPDATE` 触发器。指定一个触发器参数：要修改的列名。该列必须是 `timestamp` 或 `timestamp with time zone` 类型。

`moddatetime.example` 中有一个示例。

F.40. sslinfo

`sslinfo` 模块提供关于当前客户端在连接到 PostgreSQL 时所提供 SSL 证书的信息。如果当前连接没有使用 SSL，则该模块没有用处（大多数函数将返回 NULL）。

除非安装时使用`--with-openssl`配置，否则此扩展根本无法构建。

F.40.1. 提供的函数

`ssl_is_used() returns boolean`

如果当前到服务器的连接使用 SSL，则返回 true，否则返回 false。

`ssl_version() returns text`

返回 SSL 连接使用的协议名称（例如 SSLv2、SSLv3 或 TLSv1）。

`ssl_cipher() returns text`

返回 SSL 连接所用密码的名称（例如 DHE-RSA-AES256-SHA）。

`ssl_client_cert_present() returns boolean`

如果当前客户端已向服务器提供了有效的 SSL 客户端证书，则返回 true，否则返回 false。（服务器可能配置为要求客户端证书，也可能没有这样配置。）

`ssl_client_serial() returns numeric`

返回当前客户端证书的序列号。
证书序列号与证书颁发者的组合可保证唯一标识一个证书（但不能保证唯一标识其所有者 — 所有者应当定期更换其密钥，并从颁发者处获取新证书）。

因此，如果你运行自己的 CA，并且只允许服务器接受由该 CA 签发的证书，那么序列号就是识别用户的最可靠手段（尽管并不便于记忆）。

`ssl_client_dn() returns text`

返回当前客户端证书的完整主题，并将字符数据转换为当前数据库编码。
假定如果你在证书名称中使用了非 ASCII 字符，你的数据库也能表示这些字符。如果你的数据库使用 SQL_ASCII 编码，名称中的非 ASCII 字符将表示为 UTF-8 序列。

结果类似于 `/CN=Somebody /C=Some country/O=Some organization`。

`ssl_issuer_dn() returns text`

返回当前客户端证书的完整颁发者名称，并将字符数据转换为当前数据库编码。
编码转换的处理方式与 `ssl_client_dn` 相同。

该函数的返回值与证书序列号的组合可唯一标识该证书。

只有在服务器的`root.crt`文件中包含多个受信任的 CA 证书，或者此 CA 颁发了一些中间证书颁发机构证书时，此函数才真正有用。

`ssl_client_dn_field(fieldname text) returns text`

该函数返回证书主题中指定字段的值；如果该字段不存在，则返回 NULL。字段名是字符串常量，它们使用 OpenSSL 对象数据库转换为 ASN1 对象标识符。可接受的值如下：

```
commonName (alias CN)
surname (alias SN)
name
givenName (alias GN)
countryName (alias C)
localityName (alias L)
stateOrProvinceName (alias ST)
organizationName (alias O)
organizationUnitName (alias OU)
title
description
initials
postalCode
streetAddress
generationQualifier
description
dnQualifier
x500UniqueIdentifier
pseudonym
role
emailAddress
```

除 `commonName` 外，所有这些字段都是可选的。究竟包含哪些字段、哪些字段不包含，完全取决于你的 CA 策略。不过，这些字段的含义由 X.500 和 X.509 标准严格定义，因此你不能随意赋予它们任意含义。

`ssl_issuer_field(fieldname text) returns text`

与 `ssl_client_dn_field` 相同，只是针对证书颁发者而不是证书主题。

F.40.2. 作者

Victor Wagner `<vitus@cryptocom.ru>`, Cryptocom LTD

Cryptocom OpenSSL 开发组的电子邮件：`<openssl@cryptocom.ru>`

F.41. tablefunc

`tablefunc` 模块包含多种返回表（即多行结果）的函数。这些函数本身很有用，也可作为如何编写返回多行的 C 函数的示例。

F.41.1. 提供的函数

表 F.29总结了 `tablefunc` 模块提供的函数。

表 F.29. `tablefunc`函数

Function	Returns	Description
`normal_rand(int numvals, float8 mean, float8 stddev)`	`setof float8`	Produces a set of normally distributed random values
`crosstab(text sql)`	`setof record`	Produces a "pivot table" containing row names plus *N* value columns, where *N* is determined by the row type specified in the calling query
`crosstab`*N*`(text sql)`	`setof table_crosstab_`*N*	Produces a "pivot table" containing row names plus *N* value columns. `crosstab2`, `crosstab3`, and `crosstab4` are predefined, but you can create additional `crosstab`*N* functions as described below
`crosstab(text source_sql, text category_sql)`	`setof record`	Produces a "pivot table" with the value columns specified by a second query
`crosstab(text sql, int N)`	`setof record`	这是 `crosstab(text)` 的过时版本。参数 *N* 现已被忽略，因为值列的数量始终由调用查询决定。
`connectby(text relname, text keyid_fld, text parent_keyid_fld [, text orderby_fld ], text start_with, int max_depth [, text branch_delim ])`	`setof record`	Produces a representation of a hierarchical tree structure

F.41.1.1. `normal_rand`

```
normal_rand(int numvals, float8 mean, float8 stddev) returns setof
 float8
```

`normal_rand` 生成一组正态分布（高斯分布）的随机值。

`numvals` 是该函数要返回的值的个数。*`mean`* 是这些值所服从的正态分布的均值，*`stddev`* 是这些值所服从的正态分布的标准偏差。

例如，下面这个调用请求生成 1000 个值，均值为 5，标准偏差为 3：

```
test=# SELECT * FROM normal_rand(1000, 5, 3);
```

```
     normal_rand
----------------------
     1.56556322244898
     9.10040991424657
     5.36957140345079
   -0.369151492880995
    0.283600703686639
       .
       .
       .
     4.82992125404908
     9.71308014517282
     2.49639286969028
(1000 rows)
```

F.41.1.2. crosstab(text)

```
crosstab(text sql)
crosstab(text sql, int N)
```

该crosstab函数用于产生“透视”显示形式，其中数据沿页面横向排列，而不是纵向排列。例如，我们可能有这样的数据

```
row1    val11
row1    val12
row1    val13
...
row2    val21
row2    val22
row2    val23
...
```

并希望将它们显示成这样

```
row1    val11   val12   val13   ...
row2    val21   val22   val23   ...
...
```

该crosstab函数接受一个文本参数，该参数是一个 SQL 查询，用于产生按第一种方式格式化的原始数据，并生成一个按第二种方式格式化的表。

sql 参数是一个产生源数据集的 SQL 语句。该语句必须返回一个 `row_name` 列、一个 `category` 列和一个 `value` 列。*N* 是一个过时参数，即使提供也会被忽略（以前它必须与输出值列的个数匹配，但现在这一点由调用查询决定）。

例如，给出的查询可能会产生如下结果：

```
 row_name    cat    value
----------+-------+-------
  row1      cat1    val1
  row1      cat2    val2
  row1      cat3    val3
  row1      cat4    val4
  row2      cat1    val5
  row2      cat2    val6
```

```
  row2      cat3    val7
  row2      cat4    val8
```

该crosstab 函数声明的返回类型为 setof record，因此必须在 FROM 子句中定义输出列的实际名称和类型，该子句属于调用它的 SELECT 语句。例如：

```
SELECT * FROM crosstab('...') AS ct(row_name text, category_1 text,
 category_2 text);
```

此示例生成的集合类似于：

```
           <== value  columns  ==>
 row_name   category_1   category_2
----------+------------+------------
  row1        val1         val2
  row2        val5         val6
```

FROM 子句必须把输出定义为一个 row_name 列（其数据类型与 SQL 查询的第一列结果相同），其后跟着 N 个 value 列（它们的数据类型都与 SQL 查询的第三列结果相同）。可以按需设置任意数量的输出值列，输出列的名称也可自行决定。

crosstab 函数会为输入行中具有相同 row_name 值的每个连续分组生成一行输出。它会用这些行中的 value 字段从左到右填充输出的 value 列。如果某个分组中的行数少于输出 value 列的个数，多余的输出列会填充为空值；如果行数更多，多出来的输入行会被跳过。

实际上，SQL 查询应始终指定 ORDER BY 1,2，以确保输入行按正确顺序排列，也就是使具有相同 row_name 的值聚在一起，并在行内正确排序。注意，crosstab 本身并不会关注查询结果的第二列；该列只是为了排序而存在，用来控制第三列的值在页面上横向排列的顺序。

下面是一个完整示例：

```
CREATE TABLE ct(id SERIAL, rowid TEXT, attribute TEXT, value TEXT);
INSERT INTO ct(rowid, attribute, value) VALUES('test1','att1','val1');
INSERT INTO ct(rowid, attribute, value) VALUES('test1','att2','val2');
INSERT INTO ct(rowid, attribute, value) VALUES('test1','att3','val3');
INSERT INTO ct(rowid, attribute, value) VALUES('test1','att4','val4');
INSERT INTO ct(rowid, attribute, value) VALUES('test2','att1','val5');
INSERT INTO ct(rowid, attribute, value) VALUES('test2','att2','val6');
INSERT INTO ct(rowid, attribute, value) VALUES('test2','att3','val7');
INSERT INTO ct(rowid, attribute, value) VALUES('test2','att4','val8');

SELECT *
FROM crosstab(
  'select rowid, attribute, value
   from ct
   where attribute = ''att2'' or attribute = ''att3''
   order by 1,2')
AS ct(row_name text, category_1 text, category_2 text, category_3
 text);

 row_name | category_1 | category_2 | category_3
----------+------------+------------+------------
 test1    | val2       | val3       |
 test2    | val6       | val7       |
(2 rows)
```

可以定义一个在其定义中固定了所需输出行类型的自定义 crosstab 函数，以避免每次都必须写出用于定义输出列的 FROM 子句。下一节会介绍这种做法。另一种可能性是在视图定义中嵌入所需的 FROM 子句。

F.41.1.3. crosstab*N*(text)

crosstab*N*(text sql)

crosstab*N* 函数展示了如何为通用的 crosstab 函数设置自定义包装器，这样就不必在调用的 SELECT 查询中写出列名和类型。tablefunc 模块包含 crosstab2、crosstab3 和 crosstab4，其输出行类型定义为：

```
CREATE TYPE tablefunc_crosstab_N AS (
    row_name TEXT,
    category_1 TEXT,
    category_2 TEXT,
        .
        .
        .
    category_N TEXT
);
```

因此，当输入查询产生类型为 text 的 row_name 列和 value 列，并且需要 2、3 或 4 个输出值列时，这些函数可以直接使用。除此之外，它们在其他方面的行为与上文描述的通用 crosstab 函数完全相同。

例如，前一节中的示例也可以这样写：

```
SELECT *
FROM crosstab3(
  'select rowid, attribute, value
   from ct
   where attribute = ''att2'' or attribute = ''att3''
   order by 1,2');
```

这些函数主要用于说明。也可以基于底层的 crosstab() 函数创建自己的返回类型和函数。做法有两种：

- 创建一个描述所需输出列的复合类型，类似于 contrib/tablefunc/tablefunc--1.0.sql 中的示例。然后定义一个具有唯一名称的函数，该函数接受单个 text 参数并返回 setof your_type_name，但底层链接到同一个 crosstab C 函数。例如，如果源数据生成的行名是 text，值是 float8，而希望有 5 个值列：

  ```
  CREATE TYPE my_crosstab_float8_5_cols AS (
      my_row_name text,
      my_category_1 float8,
      my_category_2 float8,
      my_category_3 float8,
      my_category_4 float8,
      my_category_5 float8
  );

  CREATE OR REPLACE FUNCTION crosstab_float8_5_cols(text)
      RETURNS setof my_crosstab_float8_5_cols
  ```

```
    AS '$libdir/tablefunc','crosstab' LANGUAGE C STABLE STRICT;
```

- 使用 OUT 参数隐式定义返回类型。同一个示例也可以这样实现：

```
CREATE OR REPLACE FUNCTION crosstab_float8_5_cols(
    IN text,
    OUT my_row_name text,
    OUT my_category_1 float8,
    OUT my_category_2 float8,
    OUT my_category_3 float8,
    OUT my_category_4 float8,
    OUT my_category_5 float8)
  RETURNS setof record
  AS '$libdir/tablefunc','crosstab' LANGUAGE C STABLE STRICT;
```

F.41.1.4. crosstab(text, text)

```
crosstab(text source_sql, text category_sql)
```

crosstab 单参数形式的主要限制在于，它会把同一组中的所有值一视同仁，并把每个值插入第一个可用列中。如果希望值列对应于特定的数据类别，而某些分组可能没有某些类别的数据，这种形式就不太适用了。crosstab 的双参数形式通过提供一个与输出列对应的显式类别列表来处理这种情况。

source_sql 是一个产生源数据集的 SQL 语句。该语句必须返回一个 row_name 列、一个 category 列和一个 value 列。它也可以有一个或多个“extra”列。row_name 列必须位于第一列。category 列和 value 列必须是最后两列，并且顺序必须如此。位于 row_name 和 category 之间的任何列都被视为“extra”。对于具有相同 row_name 值的所有行，这些“extra”列预期应当相同。

例如，*source_sql*可能会产生如下结果：

```
SELECT row_name, extra_col, cat, value FROM foo ORDER BY 1;

 row_name    extra_col   cat    value
----------+------------+-----+---------
  row1         extra1    cat1    val1
  row1         extra1    cat2    val2
  row1         extra1    cat4    val4
  row2         extra2    cat1    val5
  row2         extra2    cat2    val6
  row2         extra2    cat3    val7
  row2         extra2    cat4    val8
```

*category_sql*是一个生成类别集合的 SQL 语句。该语句必须只返回一列。它必须至少生成一行，否则会产生错误。此外，它不得生成重复值，否则会产生错误。*category_sql*
可以是这样的：

```
SELECT DISTINCT cat FROM foo ORDER BY 1;
    cat
  -------
    cat1
    cat2
    cat3
    cat4
```

该crosstab 函数声明的返回类型为 setof record，因此必须在 FROM 子句中定义输出列的实际名称和类型，该子句属于调用它的 SELECT 语句。例如：

```
SELECT * FROM crosstab('...', '...')
    AS ct(row_name text, extra text, cat1 text, cat2 text, cat3 text,
 cat4 text);
```

这将产生如下结果：

```
                  <==  value  columns   ==>
row_name   extra   cat1   cat2   cat3   cat4
---------+-------+------+------+------+------
  row1     extra1  val1   val2          val4
  row2     extra2  val5   val6   val7   val8
```

FROM 子句必须定义数量和数据类型都正确的输出列。如果 *source_sql* 查询结果有 *N* 列，那么其中前 *N*-2 列必须与前 *N*-2 个输出列相匹配。其余输出列必须具有 *source_sql* 查询结果最后一列的数据类型，并且这些输出列的数量必须与 *category_sql* 查询结果的行数完全相同。

crosstab 函数会为输入行中具有相同 row_name 值的每个连续分组生成一行输出。输出的 row_name 列以及任何 “extra” 列都从该分组的第一行复制而来。输出的 value 列则使用那些具有匹配 category 值的行中的 value 字段来填充。如果某一行的 category 与 *category_sql* 查询的任何输出都不匹配，那么它的 value 会被忽略。那些匹配类别未出现在该分组任何输入行中的输出列会填充为空值。

实际上，*source_sql* 查询应始终指定 ORDER BY 1，以确保具有相同 row_name 的值被聚在一起。不过，分组内类别的顺序并不重要。另外，必须确保 *category_sql* 查询输出的顺序与指定的输出列顺序一致。

下面给出两个完整示例：

```
create table sales(year int, month int, qty int);
insert into sales values(2007, 1, 1000);
insert into sales values(2007, 2, 1500);
insert into sales values(2007, 7, 500);
insert into sales values(2007, 11, 1500);
insert into sales values(2007, 12, 2000);
insert into sales values(2008, 1, 1000);

select * from crosstab(
  'select year, month, qty from sales order by 1',
  'select m from generate_series(1,12) m'
) as (
  year int,
  "Jan" int,
  "Feb" int,
  "Mar" int,
  "Apr" int,
  "May" int,
  "Jun" int,
  "Jul" int,
  "Aug" int,
  "Sep" int,
  "Oct" int,
  "Nov" int,
```

```
  "Dec" int
);
 year | Jan  | Feb  | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct |
 Nov  | Dec
------+------+------+-----+-----+-----+-----+-----+-----+-----+-----+-
-----+------
 2007 | 1000 | 1500 |     |     |     |     | 500 |     |     |     |
 1500 | 2000
 2008 | 1000 |      |     |     |     |     |     |     |     |     |
      |
(2 rows)

CREATE TABLE cth(rowid text, rowdt timestamp, attribute text, val
 text);
INSERT INTO cth VALUES('test1','01 March 2003','temperature','42');
INSERT INTO cth VALUES('test1','01 March 2003','test_result','PASS');
INSERT INTO cth VALUES('test1','01 March 2003','volts','2.6987');
INSERT INTO cth VALUES('test2','02 March 2003','temperature','53');
INSERT INTO cth VALUES('test2','02 March 2003','test_result','FAIL');
INSERT INTO cth VALUES('test2','02 March 2003','test_startdate','01
 March 2003');
INSERT INTO cth VALUES('test2','02 March 2003','volts','3.1234');

SELECT * FROM crosstab
(
  'SELECT rowid, rowdt, attribute, val FROM cth ORDER BY 1',
  'SELECT DISTINCT attribute FROM cth ORDER BY 1'
)
AS
(
       rowid text,
       rowdt timestamp,
       temperature int4,
       test_result text,
       test_startdate timestamp,
       volts float8
);
 rowid |          rowdt           | temperature | test_result |
 test_startdate      | volts
-------+--------------------------+-------------+-------------+-------
-------------------+--------
 test1 | Sat Mar 01 00:00:00 2003 |          42 | PASS        |
                   | 2.6987
 test2 | Sun Mar 02 00:00:00 2003 |          53 | FAIL        | Sat
 Mar 01 00:00:00 2003 | 3.1234
(2 rows)
```

可以创建预定义函数，以避免在每个查询中都写出结果列的名称和类型。
请参见前一节中的示例。这种形式的 `crosstab` 所对应的底层 C 函数名为 `crosstab_hash`。

F.41.1.5. connectby

```
connectby(text relname, text keyid_fld, text parent_keyid_fld
          [, text orderby_fld ], text start_with, int max_depth
```

```
          [, text branch_delim ])
```

connectby 函数会显示存储在表中的层次数据。该表必须有一个唯一标识各行的键字段，以及一个引用每一行的父行（如果有）的父键字段。connectby 可以显示从任意一行开始向下展开的子树。

表 F.30解释了这些参数。

表 F.30. connectby参数

Parameter	Description
relname	Name of the source relation
keyid_fld	Name of the key field
parent_keyid_fld	Name of the parent-key field
orderby_fld	Name of the field to order siblings by (optional)
start_with	Key value of the row to start at
max_depth	Maximum depth to descend to, or zero for unlimited depth
branch_delim	String to separate keys with in branch output (optional)

键字段和父键字段可以是任意数据类型，但它们必须是同一类型。注意，
无论键字段的类型是什么，*start_with* 值都必须作为文本字符串输入。

connectby 函数被声明为返回 setof record，因此输出列的实际名称和类型必须在调用 SELECT 语句的 FROM 子句中定义，例如：

```
SELECT * FROM connectby('connectby_tree', 'keyid', 'parent_keyid',
 'pos', 'row2', 0, '~')
    AS t(keyid text, parent_keyid text, level int, branch text, pos
 int);
```

前两个输出列用于当前行的键和其父行的键，它们必须与该表键字段的类型匹配。
第三个输出列表示树中的深度，必须是 integer 类型。如果给出了 *branch_delim* 参数，下一个输出列就是分支路径显示，必须是 text 类型。最后，如果给出了 *orderby_fld* 参数，最后一个输出列就是一个序列号，必须是 integer 类型。

“branch” 输出列显示了到达当前行所经过的键路径。各键之间用指定的 *branch_delim* 字符串分隔。如果不需要分支显示，则在输出列列表中同时省略 *branch_delim* 参数和 branch 列即可。

如果同一父节点下各同级节点的顺序很重要，可以包含 *orderby_fld* 参数来指定按哪个字段对同级节点排序。该字段可以是任何可排序的数据类型。当且仅当指定了 *orderby_fld* 时，输出列列表才必须包含最后那个整数类型的序列号列。

表名和字段名参数会原样复制到 connectby 在内部生成的 SQL 查询中。因此，如果名称是大小写混合的，或者包含特殊字符，就应包含双引号。还可能需要对表名进行模式限定。

在大表中，除非父键字段上建有索引，否则性能会很差。

重要的是，*branch_delim* 字符串不要出现在任何键值中，否则 connectby 可能会错误地报告无限递归错误。注意，如果没有提供 *branch_delim*，为了递归检测会使用默认值 ~。

下面是一个示例：

```
CREATE TABLE connectby_tree(keyid text, parent_keyid text, pos int);

INSERT INTO connectby_tree VALUES('row1',NULL, 0);
INSERT INTO connectby_tree VALUES('row2','row1', 0);
INSERT INTO connectby_tree VALUES('row3','row1', 0);
INSERT INTO connectby_tree VALUES('row4','row2', 1);
INSERT INTO connectby_tree VALUES('row5','row2', 0);
INSERT INTO connectby_tree VALUES('row6','row4', 0);
INSERT INTO connectby_tree VALUES('row7','row3', 0);
INSERT INTO connectby_tree VALUES('row8','row6', 0);
INSERT INTO connectby_tree VALUES('row9','row5', 0);

-- with branch, without orderby_fld (order of results is not
 guaranteed)
SELECT * FROM connectby('connectby_tree', 'keyid', 'parent_keyid',
 'row2', 0, '~')
 AS t(keyid text, parent_keyid text, level int, branch text);
 keyid | parent_keyid | level |       branch
-------+--------------+-------+---------------------
 row2  |              |     0 | row2
 row4  | row2         |     1 | row2~row4
 row6  | row4         |     2 | row2~row4~row6
 row8  | row6         |     3 | row2~row4~row6~row8
 row5  | row2         |     1 | row2~row5
 row9  | row5         |     2 | row2~row5~row9
(6 rows)

-- without branch, without orderby_fld (order of results is not
 guaranteed)
SELECT * FROM connectby('connectby_tree', 'keyid', 'parent_keyid',
 'row2', 0)
 AS t(keyid text, parent_keyid text, level int);
 keyid | parent_keyid | level
-------+--------------+-------
 row2  |              |     0
 row4  | row2         |     1
 row6  | row4         |     2
 row8  | row6         |     3
 row5  | row2         |     1
 row9  | row5         |     2
(6 rows)

-- with branch, with orderby_fld (notice that row5 comes before row4)
SELECT * FROM connectby('connectby_tree', 'keyid', 'parent_keyid',
 'pos', 'row2', 0, '~')
 AS t(keyid text, parent_keyid text, level int, branch text, pos int);
 keyid | parent_keyid | level |       branch        | pos
-------+--------------+-------+---------------------+-----
 row2  |              |     0 | row2                |   1
 row5  | row2         |     1 | row2~row5           |   2
 row9  | row5         |     2 | row2~row5~row9      |   3
 row4  | row2         |     1 | row2~row4           |   4
 row6  | row4         |     2 | row2~row4~row6      |   5
 row8  | row6         |     3 | row2~row4~row6~row8 |   6
```

```
(6 rows)

-- without branch, with orderby_fld (notice that row5 comes before
 row4)
SELECT * FROM connectby('connectby_tree', 'keyid', 'parent_keyid',
 'pos', 'row2', 0)
 AS t(keyid text, parent_keyid text, level int, pos int);
 keyid | parent_keyid | level | pos
-------+--------------+-------+-----
 row2  |              |     0 |   1
 row5  | row2         |     1 |   2
 row9  | row5         |     2 |   3
 row4  | row2         |     1 |   4
 row6  | row4         |     2 |   5
 row8  | row6         |     3 |   6
(6 rows)
```

F.41.2. 作者

Joe Conway

F.42. test_parser

test_parser是一个全文检索自定义解析器的示例。它并不做任何特别有用的事情，但可以作为开发你自己的解析器的起点。

test_parser识别由空白分隔的单词，并且只返回两种词元类型：

```
mydb=# SELECT * FROM ts_token_type('testparser');
 tokid | alias |  description
-------+-------+---------------
     3 | word  | Word
    12 | blank | Space symbols
(2 rows)
```

这些词元编号是特意选择为与默认解析器的编号兼容的。这使我们可以使用它的headline()函数，从而让示例保持简单。

F.42.1. 用法

安装test_parser扩展会创建一个文本搜索解析器testparser。它没有可供用户配置的参数。

你可以用如下方式测试这个解析器：

```
mydb=# SELECT * FROM ts_parse('testparser', 'That''s my first own
 parser');
 tokid | token
-------+--------
     3 | That's
    12 |
     3 | my
    12 |
     3 | first
    12 |
     3 | own
    12 |
     3 | parser
```

实际使用时需要建立一个使用该解析器的文本搜索配置。例如：

```
mydb=# CREATE TEXT SEARCH CONFIGURATION testcfg ( PARSER =
 testparser );
CREATE TEXT SEARCH CONFIGURATION

mydb=# ALTER TEXT SEARCH CONFIGURATION testcfg
mydb-#   ADD MAPPING FOR word WITH english_stem;
ALTER TEXT SEARCH CONFIGURATION

mydb=#  SELECT to_tsvector('testcfg', 'That''s my first own parser');
          to_tsvector
-------------------------------
 'that':1 'first':3 'parser':5
(1 row)
```

```
mydb=# SELECT ts_headline('testcfg', 'Supernovae stars are the
 brightest phenomena in galaxies',
mydb(#                    to_tsquery('testcfg', 'star'));
                           ts_headline
-----------------------------------------------------------------
 Supernovae <b>stars</b> are the brightest phenomena in galaxies
(1 row)
```

F.43. tsearch2

tsearch2 模块为那些在文本搜索于 8.3 版集成进 PostgreSQL 核心之前使用过 tsearch2 的应用提供向后兼容的文本搜索功能。

F.43.1. 可移植性问题

尽管内置的文本搜索特性以 tsearch2 为基础，并且与它大体相似，但仍存在许多细微差异，会给现有应用带来可移植性问题：

- 一些函数的名称发生了变化，例如 `rank` 改为 `ts_rank`。替代的 `tsearch2` 模块提供了使用旧名称的别名。
- 内置文本搜索的数据类型和函数都存在于系统模式 `pg_catalog` 中。在使用 tsearch2 的安装中，这些对象通常位于 `public` 模式，不过也有些用户选择把它们放在自己单独的模式里。因此，无论哪种情况，对这些对象的显式模式限定引用都会失败。替代的 `tsearch2` 模块提供了存储在 `public`（必要时也可放在其他模式）中的别名对象，使这类引用仍能工作。
- 内置文本搜索特性中没有“当前解析器”或“当前词典”的概念，只有当前搜索配置（由 `default_text_search_config` 参数设置）的概念。虽然当前解析器和当前词典只被用于调试目的的函数使用，但在某些情况下这仍可能构成移植障碍。替代的 `tsearch2` 模块模拟了这些额外的状态变量，并提供用于设置和获取它们的向后兼容函数。

还有一些问题是替代的 `tsearch2` 模块没有解决的，因而无论如何都需要修改应用代码：

- 旧的 `tsearch2` 触发器函数允许其参数列表中的项是函数名，这些函数会在文本数据被转换为 `tsvector` 格式之前对其调用。这项功能因为属于安全漏洞而被移除，因为无法保证被调用的函数就是原本想要调用的那个。如果数据在建立索引之前必须经过处理，
推荐的做法是编写一个自定义触发器自行完成这项工作。
- 文本搜索配置信息已被移入核心系统目录中，这些目录与 tsearch2 使用的表明显不同。任何检查或修改那些表的应用都需要调整。
- 如果应用使用了任何自定义文本搜索配置，就需要使用新的文本搜索配置 SQL 命令在核心目录中重建它们。替代的 `tsearch2` 模块对此提供了一点支持：它使旧的一组 tsearch2 配置表可以装载到 PostgreSQL 8.3 中。（如果没有该模块，配置数据将无法装载，因为 `regprocedure` 列中的值无法解析为函数。）虽然这些配置表实际上不会做任何事情，但至少在 8.3 中设置等效的自定义配置时可以查阅它们的内容。
- 不支持旧的 `reset_tsearch()` 和 `get_covers()` 函数。
- 替代的 `tsearch2` 模块没有定义任何别名操作符，而是完全依赖内置操作符。只有当应用使用了显式模式限定的操作符名称时这才会成为问题，而这种情况非常少见。

F.43.2. 转换 8.3 之前的安装

更新一个使用 tsearch2 的 8.3 之前安装的推荐方法是：

1. 按通常方式从旧安装中做一个转储，但务必不要使用 pg_dump 或 pg_dumpall 的 `-c`（`--clean`）选项。
2. 在新安装中创建空数据库，并将替代的 `tsearch2` 模块安装到每个要使用文本搜索的数据库中。这必须在装载转储数据之前完成！如果你的旧安装把 tsearch2 对象放在了 `public` 以外的模式中，请务必调整 `CREATE EXTENSION` 命令，使替代对象创建在同一模式中。

3. 装载转储数据。由于无法重新创建原始的 tsearch2 对象，会报告相当多的错误。这些错误可以忽略，但这意味着你不能在单个事务中恢复转储（例如，不能使用 pg_restore 的 `-1` 开关）。

4. 检查恢复出的 tsearch2 配置表（`pg_ts_cfg` 等）的内容，并按需创建等效的内置文本搜索配置。从中提取出所有有用信息之后，就可以删除旧的配置表。

5. 测试你的应用。

之后你也许希望把应用中对别名文本搜索对象的引用改名，以便最终卸载替代的 `tsearch2` 模块。

F.43.3. 参考文献

Tsearch2 开发站点 http://www.sai.msu.su/~megera/postgres/gist/tsearch/V2/

F.44. unaccent

unaccent是一个文本检索词典，它会从词位中移除重音符号（变音符号）。
它是一个过滤词典，也就是说，它的输出总会被传递给下一个词典（如果有），
这不同于词典的通常行为。这使得全文检索能够以不区分重音的方式处理文本。

unaccent当前的实现还不能作为thesaurus词典的正规化词典使用。

F.44.1. 配置

unaccent词典接受下列选项：

- RULES是包含转换规则列表的文件的基名。该文件必须存放在$SHAREDIR/tsearch_data/中（其中$SHAREDIR表示PostgreSQL安装的共享数据目录）。其名称必须以.rules结尾，但在RULES参数中不应包含这个后缀。

规则文件的格式如下：

- 每一行表示一对字符，由一个带重音的字符和一个不带重音的字符组成。
  前者会被转换成后者。例如：

```
À        A
Á        A
Â        A
Ã        A
Ä        A
Å        A
Æ        A
```

一个更完整并且对大多数欧洲语言都可直接使用的示例可见于unaccent.rules。
安装unaccent模块时，该文件会被安装到$SHAREDIR/tsearch_data/中。

F.44.2. 用法

安装unaccent扩展会创建一个文本检索模板unaccent以及一个基于该模板的词典unaccent。
unaccent词典的默认参数设置是RULES='unaccent'，因此它可立即配合标准的unaccent.rules文件使用。如果愿意，也可以修改这个参数，例如

```
mydb=# ALTER TEXT SEARCH DICTIONARY unaccent (RULES='my_rules');
```

或者基于该模板创建新的词典。

要测试该词典，可以尝试：

```
mydb=# select ts_lexize('unaccent','Hôtel');
 ts_lexize
-----------
 {Hotel}
(1 row)
```

下面的示例展示了如何将unaccent词典插入到文本检索配置中：

```
mydb=# CREATE TEXT SEARCH CONFIGURATION fr ( COPY = french );
mydb=# ALTER TEXT SEARCH CONFIGURATION fr
```

```
        ALTER MAPPING FOR hword, hword_part, word
        WITH unaccent, french_stem;
mydb=# select to_tsvector('fr','Hôtels de la Mer');
    to_tsvector
-------------------
 'hotel':1 'mer':4
(1 row)

mydb=# select to_tsvector('fr','Hôtel de la Mer') @@ to_tsquery('fr',
'Hotels');
 ?column?
----------
 t
(1 row)

mydb=# select ts_headline('fr','Hôtel de la Mer',to_tsquery('fr',
'Hotels'));
      ts_headline
------------------------
 <b>Hôtel</b> de la Mer
(1 row)
```

F.44.3. 函数

`unaccent()`函数会从给定字符串中移除重音符号（变音符号）。从本质上说，它是对`unaccent`这一类型词典的一个包装器，但也可以在常规文本检索环境之外使用。

`unaccent([`*`dictionary`*`, ] `*`string`*`) returns text`

例如：

```
SELECT unaccent('unaccent', 'Hôtel');
SELECT unaccent('Hôtel');
```

F.45. uuid-ossp

`uuid-ossp`模块提供一些函数，可使用几种标准算法之一生成通用唯一标识符（UUID）。它还提供一些函数，用于生成某些特殊的 UUID 常量。

此模块依赖于 OSSP UUID 库，其网址为http://www.ossp.org/pkg/lib/uuid/。

F.45.1. uuid-ossp 函数

表 F.31显示了可用于生成 UUID 的函数。相关标准 ITU-T Rec. X.667、ISO/IEC 9834-8:2005 和 RFC 4122 规定了生成 UUID 的四种算法，由版本号 1、3、4 和 5 标识。（不存在版本 2 算法。）每种算法可能适用于不同的一组应用。

表 F.31. 用于生成 UUID 的函数

Function	Description
`uuid_generate_v1()`	此函数生成第 1 版 UUID。这需要使用计算机的 MAC 地址和时间戳。请注意，这类 UUID 会泄露创建该标识符的计算机身份及创建时间，因此可能不适合某些对安全敏感的应用。
`uuid_generate_v1mc()`	此函数生成第 1 版 UUID，但使用随机多播 MAC 地址，而不是计算机的真实 MAC 地址。
`uuid_generate_v3(namespace uuid, name text)`	此函数在给定命名空间中使用指定输入名称生成第 3 版 UUID。命名空间应是表 F.32中所示`uuid_ns_*()`函数生成的特殊常量之一。（理论上它可以是任意 UUID。）名称是所选命名空间中的一个标识符。 例如： `SELECT uuid_generate_v3(uuid_ns_url(), 'http://www.postgresql.org');` 名称参数会先做 MD5 哈希，因此无法从生成的 UUID 中反推出明文内容。通过这种方法生成 UUID 不包含随机因素，也不依赖于环境，因此结果可重现。
`uuid_generate_v4()`	此函数生成第 4 版 UUID，它完全由随机数派生。
`uuid_generate_v5(namespace uuid, name text)`	此函数生成第 5 版 UUID，其工作方式类似于第 3 版 UUID，但使用 SHA-1 作为哈希方法。由于 SHA-1 被认为比 MD5 更安全，应优先使用第 5 版而不是第 3 版。

表 F.32. 返回 UUID 常量的函数

`uuid_nil()`	一个“nil” UUID 常量，不会作为真实 UUID 出现。
`uuid_ns_dns()`	指定 UUID 的 DNS 命名空间的常量。
`uuid_ns_url()`	指定 UUID 的 URL 命名空间的常量。
`uuid_ns_oid()`	指定 UUID 的 ISO 对象标识符（OID）命名空间的常量。（这涉及 ASN.1 OID，与PostgreSQL使用的 OID 无关。）
`uuid_ns_x500()`	指定 UUID 的 X.500 可辨识名称（DN）命名空间的常量。

F.45.2. 作者

Peter Eisentraut <peter_e@gmx.net>

F.46. vacuumlo

vacuumlo是一个简单的实用程序，用于从 PostgreSQL数据库中移除“孤立的”大对象。孤立的大对象（LO）是指其 OID 没有出现在该数据库任何 `oid`或`lo`数据列中的 LO。

如果你使用这个程序，也可能会对lo模块中的 `lo_manage`触发器感兴趣。`lo_manage`有助于尽量避免一开始就创建出孤立 LO。

F.46.1. 用法

```
vacuumlo [options] database [database2 ... databaseN]
```

命令行上给出的所有数据库都会被处理。可用的选项包括：

`-v`

输出大量进度消息。

`-n`

不删除任何内容，只显示将要执行的操作。

`-l` *`limit`*

每个事务最多移除 *`limit`* 个大对象（默认值为 1000）。由于服务器会为每个被移除的 LO 获取一个锁，在单个事务中移除过多 LO 有可能超出max_locks_per_transaction。如果你希望在单个事务中完成全部移除，请把该限制设为 0。

`-U` *`username`*

用于连接的用户名。

`-w`
`--no-password`

绝不发出密码提示。如果服务器要求密码认证，而又无法通过 `.pgpass` 文件之类的其他方式获得密码，则连接尝试将失败。这个选项在批处理作业和脚本中很有用，
因为这些场景下没有用户可以输入密码。

`-W`

强制 vacuumlo 在连接数据库之前提示输入密码。

这个选项并非必不可少，因为如果服务器要求密码认证，vacuumlo 会自动提示输入密码。不过，vacuumlo 会浪费一次连接尝试，才知道服务器需要密码。在某些情况下，为了避免这次额外的连接尝试，提前指定 `-W` 是值得的。

`-h` *`hostname`*

Database server's host.

`-p` *`port`*

数据库服务器端口。

F.46.2. 方法

首先，vacuumlo 构建一个临时表，其中包含所选数据库中所有大对象的 OID。

然后，它会扫描数据库中所有类型为 `oid` 或 `lo` 的列，并从临时表中移除匹配的项。（注意：只会考虑名称正好为这两种的类型；特别是，基于这些类型定义的域不会被考虑。）

临时表中剩余的项标识出孤立的 LO。这些 LO 会被移除。

F.46.3. 作者

Peter Mount <peter@retep.org.uk>

F.47. xml2

`xml2` 模块提供 XPath 查询与 XSLT 功能。

F.47.1. 弃用说明

从 PostgreSQL 8.3 起，核心服务器就提供了基于 SQL/XML 标准的 XML 相关功能。该功能覆盖了 XML 语法检查和 XPath 查询，也就是本模块所做的事情，而且能力更多；不过，两者的 API 完全不兼容。计划在未来的 PostgreSQL 版本中移除此模块，转而采用较新的标准 API，因此建议你尝试迁移应用程序。如果你发现本模块的某些功能在较新的 API 中还没有以足够完善的形式提供，请向 `<pgsql-hackers@postgresql.org>` 说明你的问题，以便补齐这一不足。

F.47.2. 函数说明

表 F.33显示了该模块提供的函数。这些函数提供简单直接的 XML 解析和 XPath 查询功能。所有参数的类型都是 `text`，为简洁起见，此处不再显示。

表 F.33. 函数

Function	Returns	Description
`xml_is_well_formed(document)`	`bool`	此函数解析其参数中的文档文本，如果文档是良构 XML，则返回 true。（注意：在 PostgreSQL 8.2 之前，此函数名为 `xml_valid()`。由于 XML 中有效性和良构性具有不同含义，那是个错误的名称。旧名称仍然可用，但已被弃用。）
`xpath_string(document, query)`	`text`	这些函数计算所提供文档上的 XPath 查询，并将结果转换为指定类型。
`xpath_number(document, query)`	`float4`	
`xpath_bool(document, query)`	`bool`	

`xpath_nodeset(document, query, toptag, itemtag)`	`text`	该函数计算文档上的查询，并将结果封装在 XML 标签中。如果结果有多个值，输出将如下所示：
		
		`<toptag>`
		`<itemtag>Value 1` (可以是 XML 片段) `</itemtag>`
		`<itemtag>Value 2....</itemtag>`
		`</toptag>`
		
		如果`toptag`或`itemtag`为空字符串，就会省略相应的标签。

`xpath_nodeset(document, query)`	`text`	与 `xpath_nodeset(document, query, toptag, itemtag)` 相同，但结果省略这两个标签。
`xpath_nodeset(document, query, itemtag)`	`text`	与`xpath_nodeset(document, query, toptag, itemtag)`类似，但结果省略 toptag。

Function	Returns	Description
`xpath_list(document, query, separator)`	`text`	此函数返回由指定分隔符分隔的多个值，例如，如果分隔符为`,`，则返回 `Value 1,Value 2,Value 3`。
`xpath_list(document, query)`	`text`	This is a wrapper for the above function that uses `,` as the separator.

F.47.3. xpath_table

```
xpath_table(text key, text document, text relation, text xpaths, text
 criteria) returns setof record
```

`xpath_table` 是一个表函数，它对一组文档中的每个文档执行一组 XPath 查询，并将结果作为表返回。原始文档表中的主键字段会作为结果的第一列返回，因此结果集可以方便地用于连接。参数说明见表 F.34。

表 F.34. xpath_table 参数

Parameter	Description
key	“键” 字段的名称 – 这只是一个作为输出表第一列使用的字段，也就是说，它用于标识每个输出行来自哪条记录（关于多个值，见下文注释）
document	包含 XML 文档的字段名
relation	包含这些文档的表或视图名
xpaths	一个或多个 XPath 表达式，用 `|` 分隔
criteria	WHERE 子句的内容。这一项不能省略，因此如果你想处理该表或视图中的所有行，请使用 `true` 或 `1=1`

这些参数（XPath 字符串除外）只是直接替换进一个普通的 SQL SELECT 语句中，因此你有一定的灵活性 – 该语句为

```
SELECT <key>, <document> FROM <relation> WHERE <criteria>
```

因此，在这些特定位置上，任何合法内容都可以使用。该 SELECT 的结果必须恰好返回两列（除非你试图为 key 或 document 列出多个字段，否则都会如此）。请注意，这种简化做法要求你验证任何用户提供的值，以避免 SQL 注入攻击。

该函数必须用在 `FROM` 表达式中，并带有一个 `AS` 子句来指定输出列，例如

```
SELECT * FROM
xpath_table('article_id',
            'article_xml',
            'articles',
            '/article/author|/article/pages|/article/title',
            'date_entered > ''2003-01-01'' ')
AS t(article_id integer, author text, page_count integer, title text);
```

`AS` 子句定义了输出表中各列的名称和类型。第一列是 “键” 字段，其余各列对应 XPath 查询。如果 XPath 查询多于结果列，多出的查询将被忽略。如果结果列多于 XPath 查询，多出的列将为 NULL。

注意，这个示例把 page_count 结果列定义为整数。该函数内部只处理字符串表示，因此当你在输出中声明整数时，它会取 XPath 结果的字符串表示，并使用 PostgreSQL 输入函数将其转换为整数（或者 AS 子句要求的任何其他类型）。如果做不到这一点 – 例如结果为空 – 就会报错，因此如果你认为数据可能有问题，最好还是把列类型设为 text。

调用该函数的 SELECT 语句不一定非得只是 SELECT * – 它可以按名称引用输出列，或者将它们与其他表连接。该函数会生成一个虚拟表，你可以对它执行任何想要的操作（例如聚合、连接、排序等）。因此，更复杂一点的例子还可以是：

```
SELECT t.title, p.fullname, p.email
FROM xpath_table('article_id', 'article_xml', 'articles',
                 '/article/title|/article/author/@id',
                 'xpath_string(article_xml,''/article/@date'') >
 ''2003-03-20'' ')
       AS t(article_id integer, title text, author_id integer),
     tblPeopleInfo AS p
WHERE t.author_id = p.person_id;
```

当然，出于方便，你也可以把这一切封装到一个视图中。

F.47.3.1. 多值结果

xpath_table 函数假定每个 XPath 查询的结果都可能有多个值，因此该函数返回的行数可能不同于输入文档的数量。返回的第一行包含每个查询的第一个结果，
第二行包含每个查询的第二个结果。如果某个查询的值少于其他查询，
则会在相应位置返回空值。

在某些情况下，用户会知道某个 XPath 查询只会返回单个结果（例如唯一文档标识符）– 如果它与返回多个结果的 XPath 查询一起使用，这个单值结果只会出现在结果的第一行。解决办法是把键字段作为与一个更简单的 XPath 查询结果进行连接的条件之一。例如：

```
CREATE TABLE test (
    id int PRIMARY KEY,
    xml text
);

INSERT INTO test VALUES (1, '<doc num="C1">
<line num="L1"><a>1</a><b>2</b><c>3</c></line>
<line num="L2"><a>11</a><b>22</b><c>33</c></line>
</doc>');

INSERT INTO test VALUES (2, '<doc num="C2">
<line num="L1"><a>111</a><b>222</b><c>333</c></line>
<line num="L2"><a>111</a><b>222</b><c>333</c></line>
</doc>');

SELECT * FROM
  xpath_table('id','xml','test',
              '/doc/@num|/doc/line/@num|/doc/line/a|/doc/line/b|/doc/
line/c',
              'true')
  AS t(id int, doc_num varchar(10), line_num varchar(10), val1 int,
 val2 int, val3 int)
WHERE id = 1 ORDER BY doc_num, line_num
```

```
 id | doc_num | line_num | val1 | val2 | val3
----+---------+----------+------+------+------
  1 | C1      | L1       |    1 |    2 |    3
  1 |         | L2       |   11 |   22 |   33
```

要让 `doc_num` 在每一行中都出现，解决办法是调用两次 `xpath_table` 并连接结果：

```
SELECT t.*,i.doc_num FROM
  xpath_table('id', 'xml', 'test',
              '/doc/line/@num|/doc/line/a|/doc/line/b|/doc/line/c',
              'true')
    AS t(id int, line_num varchar(10), val1 int, val2 int, val3 int),
  xpath_table('id', 'xml', 'test', '/doc/@num', 'true')
    AS i(id int, doc_num varchar(10))
WHERE i.id=t.id AND i.id=1
ORDER BY doc_num, line_num;

 id | line_num | val1 | val2 | val3 | doc_num
----+----------+------+------+------+---------
  1 | L1       |    1 |    2 |    3 | C1
  1 | L2       |   11 |   22 |   33 | C1
(2 rows)
```

F.47.4. XSLT 函数

如果安装了 libxslt，则可使用下列函数：

F.47.4.1. `xslt_process`

```
xslt_process(text document, text stylesheet, text paramlist) returns
 text
```

这个函数将 XSL 样式表应用到文档上，并返回转换后的结果。`paramlist` 是转换时要使用的参数赋值列表，以 `a=1,b=2` 的形式指定。注意，参数解析非常简单：参数值中不能包含逗号！

此外还有 `xslt_process` 的双参数版本，它不会向转换传递任何参数。

F.47.5. 作者

John Gray `<jgray@azuli.co.uk>`

本模块由 Torchbox Ltd. (www.torchbox.com) 赞助开发。It has the same BSD licence as PostgreSQL.

附录 G. 外部项目

PostgreSQL 是一个复杂的软件项目，对其进行管理并非易事。我们发现，许多对 PostgreSQL 的增强若脱离核心项目单独开发，往往能更高效地完成。

G.1. 客户端接口

基础PostgreSQL发行版中只包含两个客户端接口：

- 之所以包含 libpq，是因为它是主要的 C 语言接口，而且许多其他客户端接口都构建在它之上。
- 之所以包含 ECPG，是因为它依赖服务器端 SQL 语法，因此对 PostgreSQL 自身的变更非常敏感。

其他所有语言接口都属于外部项目，并单独发布。表 G.1列出了其中一些项目。请注意，其中有些软件包可能并非按与PostgreSQL相同的许可发布。有关各语言接口的更多信息，包括许可条款，请参阅其网站和文档。

表 G.1. 由外部维护的客户端接口

Name	Language	Comments	Website
DBD::Pg	Perl	Perl DBI driver	http://search.cpan.org/dist/DBD-Pg/
JDBC	JDBC	Type 4 JDBC driver	http://jdbc.postgresql.org/
libpqxx	C++	New-style C++ interface	http://pqxx.org/
Npgsql	.NET	.NET data provider	http://www.npgsql.org/
pgtclng	Tcl		http://sourceforge.net/projects/pgtclng/
psqlODBC	ODBC	ODBC driver	https://odbc.postgresql.org/
psycopg	Python	DB API 2.0-compliant	http://initd.org/psycopg/

G.2. 管理工具

PostgreSQL有多种可用的管理工具。最流行的是pgAdmin III[1]，此外还有若干商业工具。

G.3. 过程语言

基础 PostgreSQL 发行版中包含若干过程语言：PL/pgSQL、PL/Tcl、PL/Perl 以及 PL/Python。

此外，还有一些过程语言是在核心PostgreSQL发行版之外开发和维护的。表 G.2列出了其中一些软件包。请注意，其中有些项目可能并非按与PostgreSQL相同的许可发布。有关各过程语言的更多信息，包括许可信息，请参阅其网站和文档。

[1] http://www.pgadmin.org/

表 G.2. 由外部维护的过程语言

Name	Language	Website
PL/Java	Java	https://github.com/tada/pljava
PL/PHP	PHP	https://public.commandprompt.com/projects/plphp
PL/Py	Python	http://python.projects.postgresql.org/
PL/R	R	http://www.joeconway.com/plr/
PL/Ruby	Ruby	http://raa.ruby-lang.org/project/pl-ruby/
PL/Scheme	Scheme	http://plscheme.projects.postgresql.org/
PL/sh	Unix shell	https://github.com/petere/plsh

G.4. 扩展

PostgreSQL 的设计目标之一，就是使其易于扩展。因此，载入数据库的扩展可以像内置特性一样工作。随源代码一同提供的 `contrib/` 目录中包含若干扩展，它们在附录 F中有所描述。其他扩展则是独立开发的，例如 PostGIS[2]。甚至 PostgreSQL 的复制解决方案也可以在外部开发。例如，Slony-I[3] 就是一个流行的主库/备库复制解决方案，它是独立于核心项目开发的。

[2] http://postgis.net/
[3] http://www.slony.info

附录 H. 源代码仓库

PostgreSQL的源代码使用Git版本控制系统进行存储和管理。官方提供了主仓库的公开镜像；主仓库发生任何变更后，该镜像都会在一分钟内更新。

我们的 wiki http://wiki.postgresql.org/wiki/Working_with_Git上有一些关于使用 Git 的讨论。

注意，从源代码仓库构建PostgreSQL需要相当新的bison、flex和Perl版本。从发行版压缩包构建时不需要这些工具，因为它们所构建的文件已包含在压缩包中。其他工具要求与第 15 章中所示相同。

H.1. 通过 Git 获取源代码

使用Git时，你将在本地机器上创建整个代码仓库的一个副本，因此即使离线也能访问全部历史和分支。这是开发或测试补丁最快且最灵活的方式。

Git

1. 你需要安装Git，可以从http://git-scm.com获取。许多系统默认已安装了较新版本的Git，或者可以在其软件包分发系统中找到。

2. 要开始使用 Git 仓库，请克隆官方镜像：

   ```
   git clone git://git.postgresql.org/git/postgresql.git
   ```

   这会把整个仓库复制到你的本地机器，因此可能需要一段时间才能完成，尤其是在网络连接较慢的情况下。文件将被放置在当前目录的一个新子目录`postgresql`中。

   如果防火墙阻止了对 Git 协议的访问，也可以通过 HTTP 协议访问 Git 镜像。只需把 URL 前缀改为`http`，例如：

   ```
   git clone http://git.postgresql.org/git/postgresql.git
   ```

   HTTP 协议的效率低于 Git 协议，因此使用起来会更慢。

3. 每当想要获取系统的最新更新时，`cd`进入该仓库目录并运行：

   ```
   git fetch
   ```

Git除了获取源代码外还能做更多事情。欲了解更多信息，请查阅Git的手册页，或访问http://git-scm.com网站。

附录 I. 文档

PostgreSQL有四种主要的文档格式：

- 纯文本，用于安装前信息
- HTML，用于在线浏览和参考
- PDF 或 PostScript，用于打印
- 手册页，用于快速参考。

另外，若干纯文本`README`文件可以在 PostgreSQL源代码树各处找到，用于说明各种实现方面的问题。

HTML文档和手册页属于标准发行包的一部分，并会默认安装。PDF 和 PostScript 格式的文档可单独下载。

I.1. DocBook

文档源码使用DocBook编写，这是一种表面上类似于HTML的标记语言。
这两种语言都是标准通用标记语言（SGML）的应用，后者本质上是一种描述其他语言的语言。
下文会同时使用 DocBook 和SGML这两个术语，但从技术上讲，它们不能互换。

DocBook允许作者指定技术文档的结构和内容，而无需操心展示细节。
文档样式定义这些内容如何被渲染成若干最终形式之一。DocBook 由 OASIS group[1]维护。官方 DocBook 站点[2]提供了很好的入门和参考文档，以及一本可供在线阅读的完整 O'Reilly 图书。NewbieDoc Docbook Guide[3]对初学者非常有帮助。FreeBSD Documentation Project[4]也使用 DocBook，并提供了一些很有价值的信息，其中包括若干值得参考的风格指南。

I.2. 工具集

以下工具用于处理文档。其中一些可能是可选的，具体见说明。

DocBook DTD[5]

这是 DocBook 自身的定义。我们当前使用 4.2 版；不能使用更高或更低的版本。你需要 DocBook DTD 的 SGML 变体，但要构建手册页，还需要同一版本的 XML 变体。

ISO 8879 character entities[6]

DocBook 需要这些实体，但由于它们由 ISO 维护，因此单独分发。

DocBook DSSSL Stylesheets[7]

它们包含将 DocBook 源文件转换成其他格式（例如 HTML）的处理指令。

[1] http://www.oasis-open.org
[2] http://www.oasis-open.org/docbook/
[3] http://newbiedoc.sourceforge.net/metadoc/docbook-guide.html
[4] http://www.freebsd.org/docproj/docproj.html
[5] http://www.oasis-open.org/docbook/
[6] http://www.oasis-open.org/cover/ISOEnts.zip
[7] http://wiki.docbook.org/DocBookDssslStylesheetDocs

DocBook XSL Stylesheets[8]

这是另一种将 DocBook 转换为其他格式的样式表。我们目前用它生成手册页，以及可选的 HTMLHelp。你也可以用这套工具链生成 HTML 或 PDF 输出，但 PostgreSQL 官方发行版为此使用 DSSSL 样式表。

OpenJade[9]

这是处理 SGML 的基础软件包。它包含一个 SGML 解析器、一个 DSSSL 处理器（即使用 DSSSL 样式表将 SGML 转换为其他格式的程序），以及若干相关工具。Jade 现在由 Open Jade 小组维护，而不再由 James Clark 维护。

Libxslt[10] for `xsltproc`

这是与 XSLT 样式表配合使用的处理工具（就像 `jade` 是与 DSSSL 样式表配合使用的处理工具一样）。

JadeTeX[11]

如果你愿意，还可以安装 JadeTeX，用 TeX 作为 Jade 的格式化后端。JadeTeX 可以创建 PostScript 或 PDF 文件（后者带书签）。

不过，JadeTeX 的输出质量不如 RTF 后端。问题特别多的地方是表格，以及各种垂直和水平间距的瑕疵。另外，也没有机会对结果进行手工润色。

我们已经记录了多种安装处理文档所需工具的方法，下面将加以介绍。
这些工具也可能还有其他打包发行形式。请将软件包状态报告到文档邮件列表，
我们会把这些信息补充到这里。

I.2.1. Linux RPM 安装

大多数发行版厂商都在其发行版中提供了用于 DocBook 处理的完整 RPM 集合。安装时查找“SGML”选项，或者查找下列软件包：`sgml-common`、`docbook`、`stylesheets`、`openjade`（或 `jade`）。可能还需要 `sgml-tools`。如果你的发行版厂商没有提供这些软件包，那么应该可以使用其他足够兼容的厂商提供的软件包。

I.2.2. FreeBSD 安装

FreeBSD Documentation Project 本身就大量使用 DocBook，因此 FreeBSD 提供了完整的文档工具“ports”也就不足为奇了。要在 FreeBSD 上构建文档，需要安装以下 ports。

- `textproc/sp`
- `textproc/openjade`
- `textproc/docbook-xsl`
- `textproc/docbook-xsl`
- `textproc/libxslt`

[8] http://wiki.docbook.org/DocBookXslStylesheets
[9] http://openjade.sourceforge.net
[10] http://xmlsoft.org/XSLT/
[11] http://jadetex.sourceforge.net

/usr/ports/print 中的一些内容（tex、jadetex）可能也值得安装。

有可能 ports 没有更新 /usr/local/share/sgml/catalog.ports 中的主 catalog 文件，或者其中顺序不正确。请确保该文件开头包含以下几行：

```
CATALOG "openjade/catalog"
CATALOG "iso8879/catalog"
CATALOG "docbook/dsssl/modular/catalog"
CATALOG "docbook/4.2/catalog"
```

如果不想编辑该文件，也可以把环境变量 SGML_CATALOG_FILES 设置为以冒号分隔的 catalog 文件列表（例如上面那些）。

更多关于 FreeBSD 文档工具的信息，请参阅FreeBSD Documentation Project 的说明[12]。

I.2.3. Debian 软件包

Debian GNU/Linux 提供了一整套文档工具的软件包。要安装，只需使用：

```
apt-get install docbook docbook-dsssl docbook-xsl openjade1.3 xsltproc
```

I.2.4. 从源码手动安装

手动安装 DocBook 工具的过程有些复杂，因此如果有预构建的软件包，就使用它们。这里只描述一种标准配置，使用合理且标准的安装路径，不使用任何“花哨”功能。有关细节，应查阅各软件包的文档，并阅读SGML入门资料。

I.2.4.1. 安装 OpenJade

1. OpenJade 的安装采用 GNU 风格的./configure; make; make install构建过程。详细信息可在 OpenJade 源码发行包中找到。简要来说：

   ```
   ./configure --enable-default-catalog=/usr/local/share/sgml/catalog
   make
   make install
   ```

   务必记住“默认目录文件”的存放位置，后面会用到它。也可以不指定，但这样以后每次使用jade时，都必须设置环境变量SGML_CATALOG_FILES，使其指向该文件。（如果 OpenJade 已经安装，而你希望在本地安装工具链的其余部分，也可以采用这种方法。）

> **注意**
>
> 一些用户报告说，使用 openjade 1.4devel 构建 PDF 时会遇到段错误，并出现类似下面这样的消息：
>
> ```
> openjade:./stylesheet.dsl:664:2:E: flow object not accepted
> by port; only display flow objects accepted
> make: *** [postgres-A4.tex-pdf] Segmentation fault
> ```
>
> 降级到 openjade 1.3 应该可以消除这个错误。

[12] http://www.freebsd.org/doc/en_US.ISO8859-1/books/fdp-primer/tools.html

2. 此外，还应该把dsssl目录中的文件 dsssl.dtd、fot.dtd、style-sheet.dtd和catalog 安装到某个位置，例如 /usr/local/share/sgml/dsssl。最简单的做法可能是复制整个目录：

   ```
   cp -R dsssl /usr/local/share/sgml
   ```

3. 最后，创建文件 /usr/local/share/sgml/catalog，并向其中加入这一行：

   ```
   CATALOG "dsssl/catalog"
   ```

   （这是对安装到步骤 2中的文件的相对路径引用。如果你选择了不同的安装布局，务必作相应调整。）

I.2.4.2. 安装DocBook DTD工具包

1. 获取DocBook V4.2 发行包[13]。

2. 创建目录/usr/local/share/sgml/docbook-4.2并切换到该目录。
   （具体位置无关紧要，但这个位置在本文所采用的布局中是合理的。）

   ```
   $ mkdir /usr/local/share/sgml/docbook-4.2
   $ cd /usr/local/share/sgml/docbook-4.2
   ```

3. 解压归档包：

   ```
   $ unzip -a ...../docbook-4.2.zip
   ```

   （归档包会将文件解压到当前目录。）

4. 编辑文件/usr/local/share/sgml/catalog（或者安装时告诉 jade 的其他位置），并加入类似下面的一行：

   ```
   CATALOG "docbook-4.2/docbook.cat"
   ```

5. 下载ISO 8879 字符实体归档包[14]，将其解压，然后把文件放入存放 DocBook 文件的同一目录：

   ```
   $ cd /usr/local/share/sgml/docbook-4.2
   $ unzip ...../ISOEnts.zip
   ```

6. 在包含 DocBook 和 ISO 文件的目录中运行以下命令：

   ```
   perl -pi -e 's/iso-(.*).gml/ISO\1/g' docbook.cat
   ```

   （这会修正 DocBook 目录文件中使用的名称与 ISO 字符实体文件的实际名称之间的混淆。）

I.2.4.3. 安装 DocBook DSSSL 样式表

要安装样式表，先解压发行包，再把它移到合适的位置，例如 /usr/local/share/sgml。（归档会自动创建一个子目录。）

```
$ gunzip docbook-dsssl-1.xx.tar.gz
$ tar -C /usr/local/share/sgml -xf docbook-dsssl-1.xx.tar
```

[13] http://www.docbook.org/sgml/4.2/docbook-4.2.zip
[14] http://www.oasis-open.org/cover/ISOEnts.zip

/usr/local/share/sgml/catalog 中通常的目录项也可以加上：

```
CATALOG "docbook-dsssl-1.xx/catalog"
```

由于样式表变化相当频繁，而且有时尝试其他版本是有益的，PostgreSQL不使用这个目录项。关于如何改为选择样式表，参见第 I.2.5 节。

I.2.4.4. 安装JadeTeX

要安装和使用JadeTeX，你需要可用的 TeX和LaTeX2e安装，包括受支持的tools和 graphics宏包、Babel、AMS 字体和 AMS-LaTeX、PSNFSS扩展及“35 种字体”配套包、用于生成 PostScript的dvips程序，以及宏包fancyhdr、hyperref、minitoc、url和ot2enc。所有这些都可以在邻近的 CTAN 站点[15]上找到。TeX基础系统的安装远远超出了本介绍的范围。任何能运行TeX的系统都应该有可用的二进制软件包。

在将JadeTeX用于 PostgreSQL文档源之前，你需要增大 TeX内部数据结构的尺寸。相关细节可以在 JadeTeX的安装说明中找到。

完成之后，就可以安装JadeTeX：

```
$ gunzip jadetex-xxx.tar.gz
$ tar xf jadetex-xxx.tar
$ cd jadetex
$ make install
$ mktexlsr
```

最后两条需要以root身份执行。

I.2.5. 通过configure检测

在构建文档之前，需要运行configure脚本，就像构建PostgreSQL程序本身时一样。检查运行即将结束时的输出，它应该类似于：

```
checking for onsgmls... onsgmls
checking for openjade... openjade
checking for DocBook V4.2... yes
checking for DocBook stylesheets... /usr/share/sgml/docbook/
stylesheet/dsssl/modular
checking for collateindex.pl... /usr/bin/collateindex.pl
checking for xsltproc... xsltproc
checking for osx... osx
```

如果onsgmls和nsgmls都没有找到，后续一些测试会被跳过。nsgmls是 Jade 软件包的一部分。如果没有自动找到这些程序，可以向 configure 传递环境变量JADE和NSGMLS以指明它们的位置。如果没有找到“DocBook V4.2”，则说明 DocBook DTD 工具包没有安装到 Jade 能找到的位置，或者目录文件没有正确设置。请参阅上面的安装提示。DocBook 样式表会在若干个较为标准的位置中查找，但如果你把它们放在其他地方，则应该设置环境变量DOCBOOKSTYLE指向该位置，然后重新运行configure。

I.3. 构建文档

一切设置妥当后，切换到doc/src/sgml目录，并运行后续各小节中介绍的某个命令来构建文档。（记得使用 GNU make。）

[15] http://www.ctan.org

I.3.1. HTML

To build the HTML version of the documentation:

```
doc/src/sgml$ gmake html
```

这也是默认目标。输出位于子目录 `html` 中。

要创建合适的索引，构建过程可能会经历多个相同的阶段。如果你不关心索引，只想校对输出，可以使用 `draft`：

```
doc/src/sgml$ gmake draft
```

要把文档构建为单个 HTML 页面，使用：

```
doc/src/sgml$ gmake postgres.html
```

I.3.2. 手册页

我们使用 DocBook XSL 样式表将 DocBook `refentry` 页面转换为适合手册页的 *roff 输出。与 HTML 版本类似，手册页也以 tar 归档包形式分发。要创建手册页，请使用以下命令：

```
cd doc/src/sgml
gmake man
```

I.3.3. Print Output via JadeTeX

如果你想使用 JadeTex 生成文档的可打印版本，可以使用下列命令之一：

- To generate PostScript via DVI in A4 format:

  ```
  doc/src/sgml$ gmake postgres-A4.ps
  ```

  In U.S. letter format:

  ```
  doc/src/sgml$ gmake postgres-US.ps
  ```

- To make a PDF:

  ```
  doc/src/sgml$ gmake postgres-A4.pdf
  ```

  or:

  ```
  doc/src/sgml$ gmake postgres-US.pdf
  ```

  （当然，你也可以用 PostScript 生成 PDF 版本，但如果直接生成 PDF，它将带有超链接和其他增强特性。）

使用 JadeTeX 构建 PostgreSQL 文档时，你很可能需要增大 TeX 的一些内部参数。这些参数可以在文件`texmf.cnf`中设置。撰写本文时，以下设置可用：

```
hash_extra.jadetex  = 200000
hash_extra.pdfjadetex  = 200000
pool_size.jadetex = 2000000
pool_size.pdfjadetex = 2000000
string_vacancies.jadetex = 150000
```

```
string_vacancies.pdfjadetex = 150000
max_strings.jadetex = 300000
max_strings.pdfjadetex = 300000
save_size.jadetex = 15000
save_size.pdfjadetex = 15000
```

I.3.4. 超宽文本

有时文本会超出打印边距，在极端情况下甚至超出打印页面，例如未折行的文本、过宽的表格。过宽的文本会在 TeX 日志输出文件（例如 `postgres-US.log`或`postgres-A4.log`）中产生“Overfull hbox”消息。一英寸有 72 个点，因此任何报告为超出 72 点以上宽度的内容都可能放不进打印页面（假定边距为一英寸）。要找到导致溢出的SGML文本，
可在溢出消息上方找到提到的第一个页码，例如`[50 ###]`（第 50 页），然后在PDF 文件中查看其后一页（例如第 51 页），看到溢出文本后相应调整 SGML即可。

I.3.5. 通过RTF生成打印输出

你也可以把PostgreSQL文档转换为 RTF并用办公套件做一些小的格式修正，从而生成可打印版本。视具体办公套件的能力而定，随后可以把文档转换为 PostScript 或 PDF。下面的过程以 Applixware 为例说明。

> **注意**
>
> 当前版本的PostgreSQL文档似乎会触发 OpenJade 的某个错误，或者超出其大小限制。如果RTF版本的构建过程长时间挂起且输出文件大小仍为 0，你可能是遇到了这个问题。（不过请记住，正常构建也需要 5 到 10 分钟，所以不要太早中止。）

Applixware RTF 清理

OpenJade没有为正文文本指定默认样式。过去，
这个未确诊的问题导致目录生成过程极其漫长。不过，在 Applixware方面的大力帮助下，症状已得到诊断，并且有了可用的变通方法。

1. 输入以下命令生成RTF版本：

   ```
   doc/src/sgml$ make postgres.rtf
   ```

2. 修复 RTF 文件，使其正确指定所有样式，特别是默认样式。如果文档包含 `refentry`小节，还必须替换把前一段落与当前段落绑定的格式提示，改为把当前段落与后一段落绑定。`doc/src/sgml` 中提供了一个实用程序`fixrtf`来完成这些修复：

   ```
   doc/src/sgml$ ./fixrtf --refentry postgres.rtf
   ```

   该脚本会添加`{\s0 Normal;}`作为文档的第 0 号样式。按照Applixware的说法，RTF 标准不允许添加隐式的第 0 号样式，不过 Microsoft Word 恰好能处理这种情况。对于修复`refentry`小节，脚本会把`\keepn` 标记替换为`\keep`。

3. 在Applixware Words中打开一个新文档，然后导入RTF文件。

4. 使用Applixware生成新目录（ToC）。

   a. 从第一行第一个字符开始到最后一行最后一个字符为止，选中现有的 ToC 行。

b. 使用Tools → Book Building → Create Table of Contents构建新的 ToC。选择让 ToC 包含前三级标题。这会用Applixware原生的 ToC 替换从 RTF 导入的现有行。

c. 使用Format → Style 调整 ToC 格式，依次选择三种 ToC 样式，并调整`First` 和`Left`的缩进。使用以下数值：

样式	首行缩进（英寸）	左缩进（英寸）
TOC-Heading 1	0.4	0.4
TOC-Heading 2	0.8	0.8
TOC-Heading 3	1.2	1.2

5. 在整个文档中完成以下工作：
   - 调整分页。
   - 调整表格列宽。
6. 用正确的值替换 ToC 中 Examples 和 Figures 部分右对齐的页码。这只需要几分钟。
7. (可选的) 如果索引节为空，从文档中删除它。
8. 重新生成并调整目录。
   a. 选中 ToC 域。
   b. 选择Tools → Book Building → Create Table of Contents。
   c. 通过选择Tools → Field Editing → Unprotect 解除 ToC 的保护。
   d. 删除 ToC 中的第一行，那是 ToC 自身的条目。
9. 将文档保存为Applixware Words原生格式，以便日后更容易地进行最后时刻的编辑。
10. 把文档 “打印” 到 PostScript 格式的文件。

I.3.6. 纯文本文件

安装说明也以纯文本形式分发，以便在没有更好的阅读工具时使用。`INSTALL` 文件对应第 15 章，并针对不同语境作了少量调整。要重新生成该文件，请切换到 `doc/src/sgml` 目录，然后输入 **`gmake INSTALL`**。

过去，发行说明和回归测试说明也曾以纯文本形式分发，但现已停止这种做法。

I.3.7. 语法检查

构建文档可能非常耗时。但有一种方法可以只检查文档文件的语法是否正确，这只需要几秒钟：

```
doc/src/sgml$ gmake check
```

I.4. 文档编写

SGML和DocBook的开源编写工具并不多。
最常用的工具组合是Emacs/XEmacs编辑器加上适当的编辑模式。在某些系统上，
典型的完整安装会包含这些工具。

I.4.1. Emacs/PSGML

PSGML是编辑SGML文档最常用、也最强大的模式。正确配置后，可以使用Emacs插入标签并检查标记一致性。它也可以用于HTML。下载、安装说明和详细文档请参阅PSGML 网站[16]。

使用PSGML时有一点需要注意：它的作者假定主SGML DTD目录是`/usr/local/lib/sgml`。如果像本章示例一样使用`/usr/local/share/sgml`，就必须作相应调整，可以设置`SGML_CATALOG_FILES`环境变量，也可以定制PSGML的安装配置（其手册介绍了具体方法）。

将以下内容放入`~/.emacs`环境文件中（将路径名调整为适合你系统的值）：

```
; ********** for SGML mode (psgml)

(setq sgml-omittag t)
(setq sgml-shorttag t)
(setq sgml-minimize-attributes nil)
(setq sgml-always-quote-attributes t)
(setq sgml-indent-step 1)
(setq sgml-indent-data t)
(setq sgml-parent-document nil)
(setq sgml-exposed-tags nil)
(setq sgml-catalog-files '("/usr/local/share/sgml/catalog"))

(autoload 'sgml-mode "psgml" "Major mode to edit SGML files." t )
```

并在同一文件中为SGML加入一项，添加到现有的变量定义中。该变量是`auto-mode-alist`：

```
(setq
  auto-mode-alist
  '(("\\.sgml$" . sgml-mode)
   ))
```

使用PSGML时，你可能会发现，处理这些分别保存书籍各部分的文件，有一种方便的方式：在编辑时插入适当的`DOCTYPE`声明。例如，当前这个源码文件是一章附录，因此可以将它指定为 DocBook 文档的 “appendix” 实例，把第一行写成这样：

```
<!DOCTYPE appendix PUBLIC "-//OASIS//DTD DocBook V4.2//EN">
```

这样，所有读取SGML的工具都能正确处理该文档，并且可以用以下命令验证文档：`nsgmls -s docguide.sgml`。（但在构建整套文档之前，需要移除这一行。）

I.4.2. 其他 Emacs 模式

GNU Emacs自带另一种SGML模式，功能不如PSGML强大，但更容易理解，也更轻量。它还提供语法高亮（font lock），这可能很有帮助。`src/tools/editors/emacs.samples`包含此模式的示例设置。

Norm Walsh 提供了一种专用于 DocBook 的主模式[17]，也带有 font-lock 和若干减少输入量的功能。

[16] http://www.lysator.liu.se/projects/about_psgml.html
[17] http://nwalsh.com/emacs/docbookide/index.html

I.5. 风格指南

I.5.1. 参考页

参考页应遵循标准布局。这能让用户更快找到所需信息，
也能促使作者记录命令的所有相关方面。不仅希望PostgreSQL参考页彼此保持一致，
也希望它们与操作系统及其他软件包提供的参考页保持一致。因此制定了下列准则。
它们在很大程度上与多个操作系统所建立的类似准则保持一致。

Reference pages that describe executable commands should contain the following sections, in this order. Sections that do not apply can be omitted. Additional top-level sections should only be used in special circumstances; often that information belongs in the“Usage”section.

名称

This section is generated automatically. It contains the command name and a half-sentence summary of its functionality.

概要

本节包含命令的语法图。概要通常不应列出每一个命令行选项；这些内容会在下文说明。
相反，它应列出命令行的主要组成部分，例如输入文件和输出文件的位置。

描述

用若干段解释该命令的作用。

选项

用列表描述每个命令行选项。如果选项很多，可以使用小节。

退出状态

如果程序以 0 表示成功、以非零表示失败，那么你不需要记录这一点。如果不同的非零退出代码有各自的含义，就在这里列出它们。

用法

描述程序的任何子语言或运行时接口。如果程序不是交互式的，通常可以省略本节。否则，
本节就是用于描述运行时特性的总括性小节。适当时可以使用小节。

环境

列出程序可能使用的所有环境变量。尽量完整；即便像 `SHELL` 这样看似微不足道的变量，也可能让用户感兴趣。

文件

列出程序可能会隐式访问的任何文件。也就是说，
不要列出在命令行上指定的输入和输出文件，而应列出配置文件等内容。

诊断

解释程序可能产生的任何不寻常输出。不要去列出所有可能的错误消息。这项工作量很大，
而在实践中用处不大。不过，例如如果错误消息有一种用户可以解析的标准格式，
那么这里就是解释它的地方。

注解

放不进其他地方的任何内容，尤其包括错误、实现缺陷、安全性考虑和兼容性问题。

示例

示例

历史

如果该程序的历史上有一些重要里程碑，可以在这里列出。通常本节可以省略。

另见

交叉引用按以下顺序列出：其他 PostgreSQL 命令参考页、PostgreSQL SQL 命令参考页、对 PostgreSQL 手册的引用、其他参考页（例如操作系统或其他软件包）、其他文档。同一组中的条目按字母顺序排列。

描述 SQL 命令的参考页应包含下列小节：名称、概要、描述、参数、输出、注解、示例、兼容性、历史、另见。参数小节类似于选项小节，但在决定列出命令的哪些子句时有更大的自由度。只有当命令返回的内容不是默认的命令完成标签时，才需要输出小节。
兼容性小节应解释该命令在多大程度上符合 SQL 标准，或者它与哪种其他数据库系统兼容。SQL 命令的另见小节应先列出 SQL 命令，再列出对程序的交叉引用。

附录 J. 缩略词

这是在PostgreSQL文档和有关PostgreSQL的讨论中常用的缩略词列表。

ANSI

美国国家标准协会[1]

API

应用程序编程接口[2]

ASCII

美国信息交换标准代码[3]

BKI

后端接口

CA

证书颁发机构[4]

CIDR

无类域间路由[5]

CPAN

Perl综合档案网络[6]

CRL

证书吊销列表[7]

CSV

逗号分隔值[8]

CTE

公共表表达式

CVE

常见漏洞与暴露[9]

[1] http://en.wikipedia.org/wiki/American_National_Standards_Institute
[2] http://en.wikipedia.org/wiki/API
[3] http://en.wikipedia.org/wiki/Ascii
[4] http://en.wikipedia.org/wiki/Certificate_authority
[5] http://en.wikipedia.org/wiki/Classless_Inter-Domain_Routing
[6] http://www.cpan.org/
[7] http://en.wikipedia.org/wiki/Certificate_revocation_list
[8] http://en.wikipedia.org/wiki/Comma-separated_values
[9] http://cve.mitre.org/

DBA

数据库管理员[10]

DBI

数据库接口（Perl）[11]

DBMS

数据库管理系统[12]

DDL

数据定义语言[13]，例如SQL命令`CREATE TABLE`、`ALTER USER`

DML

数据操纵语言[14]，例如SQL命令`INSERT`、`UPDATE`、`DELETE`

DST

夏令时[15]

ECPG

用于 PostgreSQL 的嵌入式 C

ESQL

嵌入式SQL[16]

FAQ

常见问题[17]

FSM

空闲空间映射

GEQO

遗传查询优化器

GIN

通用倒排索引

GiST

通用搜索树

[10] http://en.wikipedia.org/wiki/Database_administrator
[11] http://dbi.perl.org/
[12] http://en.wikipedia.org/wiki/Dbms
[13] http://en.wikipedia.org/wiki/Data_Definition_Language
[14] http://en.wikipedia.org/wiki/Data_Manipulation_Language
[15] http://en.wikipedia.org/wiki/Daylight_saving_time
[16] http://en.wikipedia.org/wiki/Embedded_SQL
[17] http://en.wikipedia.org/wiki/FAQ

Git

Git[18]

GMT

格林尼治标准时间[19]

GSSAPI

通用安全服务应用程序接口[20]

GUC

Grand Unified Configuration，即负责处理服务器配置的PostgreSQL子系统

HBA

基于主机的认证

HOT

堆内元组[21]

IEC

国际电工委员会[22]

IEEE

电气和电子工程师学会[23]

IPC

进程间通信[24]

ISO

国际标准化组织[25]

ISSN

国际标准连续出版物编号[26]

JDBC

Java数据库连接[27]

[18] http://en.wikipedia.org/wiki/Git_(software)
[19] http://en.wikipedia.org/wiki/GMT
[20] http://en.wikipedia.org/wiki/Generic_Security_Services_Application_Program_Interface
[21] http://git.postgresql.org/gitweb?p=postgresql.git;a=blob;f=src/backend/access/heap/README.HOT;hb=HEAD
[22] http://en.wikipedia.org/wiki/International_Electrotechnical_Commission
[23] http://standards.ieee.org/
[24] http://en.wikipedia.org/wiki/Inter-process_communication
[25] http://www.iso.org/iso/home.htm
[26] http://en.wikipedia.org/wiki/Issn
[27] http://en.wikipedia.org/wiki/Java_Database_Connectivity

LDAP

轻量级目录访问协议[28]

MSVC

Microsoft Visual C[29]

MVCC

多版本并发控制

NLS

本地语言支持[30]

ODBC

开放数据库连接[31]

OID

对象标识符

OLAP

联机分析处理[32]

OLTP

联机事务处理[33]

ORDBMS

对象关系数据库管理系统[34]

PAM

可插拔认证模块[35]

PGSQL

PostgreSQL

PGXS

PostgreSQL扩展系统

PID

进程标识符[36]

[28] http://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol
[29] http://en.wikipedia.org/wiki/Visual_C++
[30] http://en.wikipedia.org/wiki/Internationalization_and_localization
[31] http://en.wikipedia.org/wiki/Open_Database_Connectivity
[32] http://en.wikipedia.org/wiki/Olap
[33] http://en.wikipedia.org/wiki/OLTP
[34] http://en.wikipedia.org/wiki/ORDBMS
[35] http://en.wikipedia.org/wiki/Pluggable_Authentication_Modules
[36] http://en.wikipedia.org/wiki/Process_identifier

PITR

时间点恢复（持续归档）

PL

过程语言（服务器端）

POSIX

可移植操作系统接口[37]

RDBMS

关系数据库管理系统[38]

RFC

征求意见稿[39]

SGML

标准通用标记语言[40]

SPI

服务器编程接口

SQL

结构化查询语言[41]

SRF

集合返回函数

SSH

安全外壳协议[42]

SSL

安全套接字层[43]

SSPI

安全支持提供程序接口[44]

SYSV

Unix System V[45]

[37] http://en.wikipedia.org/wiki/POSIX
[38] http://en.wikipedia.org/wiki/Relational_database_management_system
[39] http://en.wikipedia.org/wiki/Request_for_Comments
[40] http://en.wikipedia.org/wiki/SGML
[41] http://en.wikipedia.org/wiki/SQL
[42] http://en.wikipedia.org/wiki/Secure_Shell
[43] http://en.wikipedia.org/wiki/Secure_Sockets_Layer
[44] http://msdn.microsoft.com/en-us/library/aa380493%28VS.85%29.aspx
[45] http://en.wikipedia.org/wiki/System_V

TCP/IP

传输控制协议（TCP）/ 互联网协议（IP）[46]

TID

元组标识符

TOAST

超长属性存储技术

TPC

事务处理性能委员会[47]

URL

统一资源定位符[48]

UTC

协调世界时[49]

UTF

Unicode 转换格式[50]

UTF8

8 位 Unicode 转换格式[51]

UUID

通用唯一标识符

WAL

预写式日志

XID

事务 ID

XML

可扩展标记语言[52]

[46] http://en.wikipedia.org/wiki/Transmission_Control_Protocol
[47] http://www.tpc.org/
[48] http://en.wikipedia.org/wiki/URL
[49] http://en.wikipedia.org/wiki/Coordinated_Universal_Time
[50] http://www.unicode.org/
[51] http://en.wikipedia.org/wiki/Utf8
[52] http://en.wikipedia.org/wiki/XML

参考书目

这里列出了一些关于 SQL 和 PostgreSQL 的参考文献与读物。

来自最初的 POSTGRES 开发团队的一些白皮书和技术报告，可在加州大学伯克利分校计算机系的网站[1]上获取。

SQL参考书

SQL 特性的参考文本。

[bowman01] The Practical SQL Handbook. Bowman et al, 2001. Using SQL Variants. Fourth Edition. Judith Bowman、Sandra Emerson和Marcy Darnovsky. ISBN 0-201-70309-2. 2001. Addison-Wesley Professional. 版权 © 2001.

[date97] A Guide to the SQL Standard. Date and Darwen, 1997. A user's guide to the standard database language SQL. Fourth Edition. C. J. Date和Hugh Darwen. ISBN 0-201-96426-0. 1997. Addison-Wesley. 版权 © 1997 Addison-Wesley Longman, Inc..

[date04] An Introduction to Database Systems. Date, 2004. Eighth Edition. C. J. Date. ISBN 0-321-19784-4. 2003. Addison-Wesley. 版权 © 2004 Pearson Education, Inc..

[elma04] Fundamentals of Database Systems. Fourth Edition. Ramez Elmasri和Shamkant Navathe. ISBN 0-321-12226-7. 2003. Addison-Wesley. 版权 © 2004.

[melt93] Understanding the New SQL. Melton and Simon, 1993. A complete guide. Jim Melton和Alan R. Simon. ISBN 1-55860-245-3. 1993. Morgan Kaufmann. 版权 © 1993 Morgan Kaufmann Publishers, Inc..

[ull88] Principles of Database and Knowledge. Base Systems. Ullman, 1988. Jeffrey D. Ullman. Volume 1. Computer Science Press. 1988.

PostgreSQL相关文档

本节收录相关文档。

[sim98] Enhancement of the ANSI SQL Implementation of PostgreSQL. Simkovics, 1998. Stefan Simkovics.

该文讨论了 SQL 的历史和语法，并描述了将 `INTERSECT` 和 `EXCEPT` 构造加入 PostgreSQL 的过程。它是作者在维也纳技术大学、在 O. Univ. Prof. Dr. Georg Gottlob 和 Univ. Ass. Mag. Katrin Seyr 支持下完成的硕士学位论文。

November 29, 1998. Department of Information Systems, Vienna University of Technology. Vienna, Austria.

[yu95] The Postgres95. User Manual. Yu and Chen, 1995. A. Yu和J. Chen. . Sept. 5, 1995. University of California. Berkeley, California.

[fong] The design and implementation of the POSTGRES query optimizer [2]. Zelaine Fong. University of California, Berkeley, Computer Science Department.

[1] http://db.cs.berkeley.edu/papers/
[2] http://db.cs.berkeley.edu/papers/UCB-MS-zfong.pdf

会议论文和期刊文章

本节收录文章和简报。

[olson93] Partial indexing in POSTGRES: research project. Olson, 1993. Nels Olson. 1993. UCB Engin T7.49.1993 O676. University of California. Berkeley, California.

[ong90] "A Unified Framework for Version Modeling Using Production Rules in a Database System". Ong and Goh, 1990. L. Ong和J. Goh. ERL Technical Memorandum M90/33. April, 1990. University of California. Berkely, California.

[rowe87] " The POSTGRES data model [3]". Rowe and Stonebraker, 1987. L. Rowe和M. Stonebraker. VLDB Conference, Sept. 1987.

[seshadri95] "Generalized Partial Indexes (cached version) [4] ". Seshardri, 1995. P. Seshadri和A. Swami. Eleventh International Conference on Data Engineering, 6-10 March 1995. 1995. Cat. No.95CH35724. IEEE Computer Society Press. Los Alamitos, California. 420-7.

[ston86] " The design of POSTGRES [5]". Stonebraker and Rowe, 1986. M. Stonebraker和L. Rowe. ACM-SIGMOD Conference on Management of Data, May 1986.

[ston87a] "The design of the POSTGRES. rules system". Stonebraker, Hanson, Hong, 1987. M. Stonebraker、E. Hanson和C. H. Hong. IEEE Conference on Data Engineering, Feb. 1987.

[ston87b] " The design of the POSTGRES storage system [6]". Stonebraker, 1987. M. Stonebraker. VLDB Conference, Sept. 1987.

[ston89] " A commentary on the POSTGRES rules system [7]". Stonebraker et al, 1989. M. Stonebraker、M. Hearst和S. Potamianos. SIGMOD Record 18(3). Sept. 1989.

[ston89b] " The case for partial indexes [8]". Stonebraker, M, 1989b. M. Stonebraker. SIGMOD Record 18(4). 4-11. Dec. 1989.

[ston90a] " The implementation of POSTGRES [9]". Stonebraker, Rowe, Hirohama, 1990. M. Stonebraker、L. A. Rowe和M. Hirohama. Transactions on Knowledge and Data Engineering 2(1). IEEE. March 1990.

[ston90b] " On Rules, Procedures, Caching and Views in Database Systems [10]". Stonebraker et al, ACM, 1990. M. Stonebraker、A. Jhingran、J. Goh和S. Potamianos. ACM-SIGMOD Conference on Management of Data, June 1990.

[3] http://db.cs.berkeley.edu/papers/ERL-M87-13.pdf
[4] http://citeseer.ist.psu.edu/seshadri95generalized.html
[5] http://db.cs.berkeley.edu/papers/ERL-M85-95.pdf
[6] http://db.cs.berkeley.edu/papers/ERL-M87-06.pdf
[7] http://db.cs.berkeley.edu/papers/ERL-M89-82.pdf
[8] http://db.cs.berkeley.edu/papers/ERL-M89-17.pdf
[9] http://db.cs.berkeley.edu/papers/ERL-M90-34.pdf
[10] http://db.cs.berkeley.edu/papers/ERL-M90-36.pdf

索引

符号

$，40
$libdir，795
$libdir/plugins，，1292
*，94
.pgpass，605
.pg_service.conf，605
::，44
_PG_fini，795
_PG_init，795
下标，40
不可重复读，319
不带时区的时间，113
不带时区的时间戳，113
串行化异常，322
主机名，557
事务 ID
  回卷，481
事务日志 (见 WAL)
事务隔离，319
事务隔离级别，319
  可串行化，322
  可重复读，321
  读已提交，320
二进制串
  串接，168
  长度，170
二进制数据，111
  函数，168
交叉编译，356
交叉连接，
任意精度数字，106
休眠，203
位串
  函数，171
  常量，35
  数据类型，129
位图扫描，275
信号
  后端进程，253
信号量，381
值表达式，39
全文搜索，283
  数据类型，129
全文检索
  函数和操作符，129
公共表表达式 (见 WITH)
共享内存，381
共享库，361，804
关键词
  语法，31
关闭，389
内存上下文
  在 SPI 中，1007
函数
  位置记法，50
  命名记法，51
  在FROM子句中，90
  混合记法，51
  调用，41
  调用中的类型解析，266
函数依赖，93
分组，92
列引用，40
删除，82
别名
  在FROM子句中，89
  在选择列表中，95
  查询中的表名，12
加密，392
区域设置，378，468
升级，389
协方差
  总体，235
  样本，235
参数
  语法，40
双精度，107
反斜线转义，33
可串行化，322
可串行化快照隔离，319
可重复读，321
右连接，
合取，149
同步提交，540
名称
  语法，31
否定，149
咨询锁，327
回归截距，235
回归斜率，235
回归测试，359
圆括号，40
备份，254
复合类型
  构造器，47
外连接，
多版本并发控制，319
大小写敏感性
  SQL 命令，31
子查询，46，90，237
字段选择，41
字符，110
字符串

串接，156
常量，33
数据类型，110
长度，157
字符集，472
安装，349
在 Windows 上，372
密码，459
超级用户的，378
对象标识符
数据类型，145
左连接，
布尔
操作符 (见操作符，逻辑)
带时区的时间，113，116
带时区的时间戳，113，117
常量，32
平均值，233
并发，319
幻读，319
序列，225
延迟，203
异步提交，540
引号
与标识符，32
转义，33
引导
期间启动服务器，379
排序，96
排序规则，470
插入，81
操作符
优先级，38
用户定义的，822
语法，37
调用，41
调用中的类型解析，263
逻辑，149
操作符族，279，833
操作符类，279，827
数字
常量，35
数字（数据类型），106
数据区域 (见数据库集簇)
数据库
创建权限，459
数据库集簇，377
数据类型，104
enumerated (enum)，123
分类，263
常量，36
数字，105
转换，262
数组
构造器，46
用户定义类型的，822
整数，106
文件系统挂载点，378
文本搜索，283
函数和操作符，129
数据类型，129
索引，313
文档
全文搜索，283
方差
总体，235
样本，235
无时区的时间，116
无时区的时间戳，117
日常维护，479
日志写到哪里，418
日期
当前，201
时区
转换，201
时区数据，
时间
当前，201
时间点恢复，486
时间跨度，113
更新，82
服务器日志
日志文件维护，485
服务器欺骗，392
条件表达式，227
析取，149
查询，84
标准差，235
总体，235
样本，235
标签 (见别名)
标识符
语法，31
长度，31
标量 (见表达式)
格式化，185
检查点，541
模式匹配，172
正则表达式，173，174
(参见模式匹配)
与区域设置，469
死锁，326
比较
子查询结果行，237
按行，240
操作符，149
注释
在 SQL 中，37

浮点，107
清理，479
版本
  兼容性，389
登录权限，459
目标列表，856
相关性，235
磁盘使用情况，536
磁盘空间，480
磁盘驱动器，543
空值
  在 DISTINCT 中，96
  在 PL/Perl 中，943
  在 PL/Python 中，959
  比较，150
窗口函数
  invocation，43
  内置，236
  执行顺序，94
端口，557
类型 (见数据类型)
类型转换，36
索引，272
  B-树，273
  基于表达式，277
  GIN，274
    文本搜索，313
  GiST，273
    文本搜索，313
  hash，273
  与 ORDER BY，275
  唯一，276
  多列，274
  检查使用情况，281
  组合多个索引，275
  部分，277
  锁，329
线性回归，234
线段，126
统计，234
统计信息
  规划器的，481
维护，479
网络文件系统，378
网络附加存储（NAS）(见网络文件系统)
美元引用，34
聚合函数
  内置，233
  调用，42
脏读，319
自动清理
  一般信息，483
自然连接，86
范围表，855
行比较，240
行类型
  构造器，47
表函数，90
表表达式，84
表达式
  计算的顺序，49
  语法，39
角色
  创建权限，459
  发起复制的权限，459
触发器
  用于更新一个派生的 tsvector 列，296
  触发器函数的参数，847
词元，31
读已提交，320
谓词锁，322
超级用户，459
转义字符串语法，33
输入函数，819
输出函数，819
连接，10，85
  交叉，
  右，
  外，
  外连接，11
  左，
  自然，86
  自连接，12
通知处理器，596
通知接收器，596
配置
  服务器
    函数，253
重启点，542
重复，95
重建索引，484
重载
  操作符，822
锁，323
  咨询，327
错误消息，568
长度
  二进制串 (见二进制串，长度)
  字符串 (见字符串，长度)
间隔，113
集合交，96
集合差，96
集合并，96
集合操作，96
集合返回函数
  函数，242
非数字
  double precision，108

数字（数据类型），107
非空约束，55
非阻塞连接，561

A

abbrev，208
ABORT，1028
abs，152
acos，156
administration tools
externally maintained，2430
adminpack，2260
age，192
aggregate function，12
user-defined，817
AIX
IPC 配置，
在其上安装，363
ALL，237，240
allow_system_table_mods配置参数，
ALTER AGGREGATE，1029
ALTER COLLATION，1031
ALTER CONVERSION，1032
ALTER DATABASE，1033
ALTER DEFAULT PRIVILEGES，1035
ALTER DOMAIN，1038
ALTER EXTENSION，1041
ALTER FOREIGN DATA WRAPPER，1044
ALTER FOREIGN TABLE，1046
ALTER FUNCTION，1049
ALTER GROUP，1052
ALTER INDEX，1053
ALTER LANGUAGE，1055
ALTER LARGE OBJECT，1056
ALTER OPERATOR，1057
ALTER OPERATOR CLASS，1059
ALTER OPERATOR FAMILY，1060
ALTER ROLE，459，1064
ALTER SCHEMA，1067
ALTER SEQUENCE，1068
ALTER SERVER，1071
ALTER TABLE，1072
ALTER TABLESPACE，1082
ALTER TEXT SEARCH CONFIGURATION，1084
ALTER TEXT SEARCH DICTIONARY，1086
ALTER TEXT SEARCH PARSER，1088
ALTER TEXT SEARCH TEMPLATE，1089
ALTER TRIGGER，1090
ALTER TYPE，1091
ALTER USER，1094
ALTER USER MAPPING，1095
ALTER VIEW，1096
ANALYZE，481，1098
AND（操作符），149
anonymous code blocks，1231
any，147
ANY，234，237，240
anyarray，147
anyelement，147
anyenum，147
anynonarray，147
applicable role，735
application_name配置参数，
archive_cleanup_command recovery parameter，
archive_command配置参数，
archive_mode配置参数，
archive_timeout配置参数，
area，205
ARRAY，46
结果类型的确定，269
array，134
accessing，136
constant，135
declaration，134
I/O，141
modifying，138
searching，140
array_agg，233
array_append，231
array_cat，231
array_dims，231
array_fill，231
array_length，231
array_lower，231
array_ndims，231
array_nulls配置参数，
array_prepend，231
array_to_string，231
array_upper，231
ascii，158
asin，156
AT TIME ZONE，201
atan，156
atan2，156
authentication_timeout配置参数，
auth_delay，2262
auth_delay.milliseconds configuration parameter，
auto-increment (见 serial)
autocommit
bulk-loading data，339
psql，
autovacuum
configuration parameters，427
autovacuum_analyze_scale_factor配置参数，

autovacuum_analyze_threshold配置参数，
autovacuum_freeze_max_age配置参数，
autovacuum_max_workers配置参数，
autovacuum_naptime配置参数，
autovacuum_vacuum_cost_delay配置参数，
autovacuum_vacuum_cost_limit配置参数，
autovacuum_vacuum_scale_factor配置参数，
autovacuum_vacuum_threshold配置参数，
autovacuum配置参数，
auto_explain，2263
auto_explain.log_analyze configuration parameter，
auto_explain.log_buffers configuration parameter，
auto_explain.log_format configuration parameter，
auto_explain.log_min_duration configuration parameter，
auto_explain.log_nested_statements configuration parameter，
auto_explain.log_verbose configuration parameter，
avg，233

B

B-树 (见索引)
backslash_quote配置参数，
backup，486
base type，778
BEGIN，1100
BETWEEN，150
BETWEEN SYMMETRIC，150
bgwriter_delay配置参数，
bgwriter_lru_maxpages配置参数，
bgwriter_lru_multiplier配置参数，
bigint，36，106
bigserial，108
bison，350
bitmap scan，
bit_and，233
bit_length，157
bit_or，233
BLOB (见 large object)
block_size配置参数，
bonjour_name配置参数，
bonjour配置参数，
Boolean
  数据类型，122
bool_and，233
bool_or，233
box，206
box (data type)，126
broadcast，208
BSD/OS
  IPC 配置，
  shared library，
btree_gin，2265
btree_gist，2266
btrim，158，169
bytea，111
bytea_output配置参数，

C

C，556，632
C++，817
canceling
  SQL command，586
CASCADE
  with DROP，79
  外键动作，59
CASE，227
  结果类型的确定，269
cast
  I/O conversion，1124
cbrt，152
ceil，152
ceiling，152
center，205
Certificate，456
char，110
character set， ，
character varying，110
char_length，157
check constraint，54
CHECKPOINT，1102
checkpoint_completion_target配置参数，
checkpoint_segments配置参数，
checkpoint_timeout配置参数，
checkpoint_warning配置参数，
check_function_bodies配置参数，
chkpass，2268
chr，159
cid，145
cidr，127
circle，127，206
citext，2269
client authentication，443
  timeout during，
client_encoding配置参数，
client_min_messages配置参数，
clock_timestamp，192
CLOSE，1103
cluster
  of databases (见 database cluster)

CLUSTER，1104
clusterdb，1368
clustering，499
cmax，61
cmin，61
COALESCE，229
COLLATE，45
collation
  in PL/pgSQL，887
  in SQL functions，792
column，7，52
  adding，62
  removing，62
  renaming，64
  system column，60
column data type
  changing，63
col_description，251
comment
  关于数据库对象，251
COMMENT，1106
COMMIT，1110
COMMIT PREPARED，1111
commit_delay配置参数，
commit_siblings配置参数，
compiling
  libpq applications，610
composite type，142，778
  constant，143
computed field，784
concat，159
concat_ws，159
configuration
  of recovery
    of a standby server，516
  of the server，397
configure，351
config_file配置参数，
connection service file，605
constraint，54
  adding，62
  check，54
  exclusion，60
  foreign key，58
  name，54
  NOT NULL，55
  primary key，57
  removing，63
  unique，56
constraint exclusion，76，
constraint_exclusion配置参数，
CONTINUE
  in PL/pgSQL，901
continuous archiving，486
control file，838
convert，159
convert_from，159
convert_to，159
COPY，9，1112
  with libpq，589
corr，235
cos，156
cot，156
count，233
covar_pop，235
covar_samp，235
cpu_index_tuple_cost配置参数，
cpu_operator_cost配置参数，
cpu_tuple_cost配置参数，
CREATE AGGREGATE，1121
CREATE CAST，1124
CREATE COLLATION，1128
CREATE CONVERSION，1130
CREATE DATABASE，463，1132
CREATE DOMAIN，1135
CREATE EXTENSION，1137
CREATE FOREIGN DATA WRAPPER，1139
CREATE FOREIGN TABLE，1141
CREATE FUNCTION，1143
CREATE GROUP，1150
CREATE INDEX，1151
CREATE LANGUAGE，1157
CREATE OPERATOR，1160
CREATE OPERATOR CLASS，1163
CREATE OPERATOR FAMILY，1166
CREATE ROLE，458，1167
CREATE RULE，1172
CREATE SCHEMA，1175
CREATE SEQUENCE，1177
CREATE SERVER，1180
CREATE TABLE，7，1182
CREATE TABLE AS，1195
CREATE TABLESPACE，466，1198
CREATE TEXT SEARCH CONFIGURATION，1200
CREATE TEXT SEARCH DICTIONARY，1201
CREATE TEXT SEARCH PARSER，1203
CREATE TEXT SEARCH TEMPLATE，1205
CREATE TRIGGER，1206
CREATE TYPE，1211
CREATE USER，1218
CREATE USER MAPPING，1219
CREATE VIEW，1220
createdb，3，464，1371
createlang，1374
createuser，458，1377
cstring，147
ctid，61
CTID，862

cube（扩展），2272
cume_dist，236
current_catalog，245
current_database，245
current_date，193
current_query，245
current_schema，245
current_schemas，245
current_setting，253
current_time，193
current_timestamp，193
current_user，245
currval，225
cursor
  CLOSE，1103
  DECLARE，1224
  FETCH，1277
  in PL/pgSQL，906
  MOVE，1296
  showing the query plan，1272
cursor_tuple_fraction配置参数，
custom_variable_classes配置参数，
Cygwin
  在其上安装，366

D

data partitioning，499
data type
  base，778
  composite，778
  internal organization，796
  type cast，44
  user-defined，819
database，463
  creating，3
database activity
  monitoring，519
database cluster，7
data_directory配置参数，
date，113，115
  constants，117
  输出格式，118
    (参见 formatting)
DateStyle配置参数，
date_part，193，196
date_trunc，193，200
dblink，2276
db_user_namespace配置参数，
deadlock
  timeout during，
deadlock_timeout配置参数，
DEALLOCATE，1223
debug_assertions配置参数，
debug_deadlocks配置参数，
debug_pretty_print配置参数，
debug_print_parse配置参数，
debug_print_plan配置参数，
debug_print_rewritten配置参数，
decimal (见 numeric)
DECLARE，1224
decode，160，170
decode_bytea
  in PL/Perl，
default value，53
  changing，63
default_statistics_target配置参数，
default_tablespace配置参数，
default_text_search_config配置参数，
default_transaction_deferrable配置参数，
default_transaction_isolation配置参数，
default_transaction_read_only配置参数，
default_with_oids配置参数，
deferrable transaction
  setting，1348
  setting default，
degrees，152
DELETE，14，82，1227
dense_rank，236
diameter，205
dict_int，2307
dict_xsyn，2308
Digital UNIX (见 Tru64 UNIX)
DISCARD，1230
DISTINCT，10，95
div，153
DO，1231
DROP AGGREGATE，1233
DROP CAST，1234
DROP COLLATION，1235
DROP CONVERSION，1236
DROP DATABASE，465，1237
DROP DOMAIN，1238
DROP EXTENSION，1239
DROP FOREIGN DATA WRAPPER，1240
DROP FOREIGN TABLE，1241
DROP FUNCTION，1242
DROP GROUP，1244
DROP INDEX，1245
DROP LANGUAGE，1246
DROP OPERATOR，1247
DROP OPERATOR CLASS，1249
DROP OPERATOR FAMILY，1251
DROP OWNED，1253
DROP ROLE，458，1254
DROP RULE，1255
DROP SCHEMA，1256
DROP SEQUENCE，1257

DROP SERVER，1258
DROP TABLE，8，1259
DROP TABLESPACE，1260
DROP TEXT SEARCH CONFIGURATION，1261
DROP TEXT SEARCH DICTIONARY，1262
DROP TEXT SEARCH PARSER，1263
DROP TEXT SEARCH TEMPLATE，1264
DROP TRIGGER，1265
DROP TYPE，1266
DROP USER，1267
DROP USER MAPPING，1268
DROP VIEW，1269
dropdb，466，1381
droplang，1383
dropuser，458，1386
DTD，133
DTrace，357，528
dummy_seclabel，2310
duplicate，10
dynamic loading，，795
dynamic_library_path，795
dynamic_library_path配置参数，

E

earthdistance，2311
ECPG，632
ecpg，1389
effective_cache_size配置参数，
effective_io_concurrency配置参数，
elog，1584
  in PL/Perl，
  in PL/Tcl，
  在 PL/Python 中，968
embedded SQL
  in C，632
enabled role，750
enable_bitmapscan配置参数，
enable_hashagg配置参数，
enable_hashjoin配置参数，
enable_indexscan配置参数，
enable_material配置参数，
enable_mergejoin配置参数，
enable_nestloop配置参数，
enable_seqscan配置参数，
enable_sort配置参数，
enable_tidscan配置参数，
encode，160，170
encode_array_constructor
  in PL/Perl，
encode_array_literal
  in PL/Perl，
encode_bytea
  in PL/Perl，
encode_typed_literal
  in PL/Perl，
encryption
  for specific columns，2359
END，1270
enumerated types，123
enum_first，203
enum_last，203
enum_range，203
environment variable，603
ereport，1584
error codes
  libpq，
  list of，1667
escape_string_warning配置参数，
escaping strings
  in libpq，580
every，233
EXCEPT，96
exceptions
  in PL/pgSQL，904
exclusion constraint，60
EXECUTE，1271
EXISTS，237
EXIT
  in PL/pgSQL，900
exit_on_error配置参数，
exp，153
EXPLAIN，331，1272
extending SQL，778
extension，837
  externally maintained，2431
external_pid_file配置参数，
extract，193，196
extra_float_digits配置参数，

F

failover，499
false，122
family，209
fast path，587
fdw_handler，147
FETCH，1277
field
  computed，784
file_fdw，2313
first_value，236
flex，350
float4 (见 real)
float8 (见双精度)
floating-point
  display，
floor，153

foreign data，78
foreign data wrapper
  handler for，1599
foreign key，16，58
foreign table，78
format，160
  use in PL/pgSQL，892
format_type，249
Free Space Map，1633
FreeBSD
  IPC 配置，
  shared library，
  启动脚本，379
from_collapse_limit配置参数，
FSM (见 Free Space Map)
fsync配置参数，
full_page_writes配置参数，
function，149
  default values for arguments，787
  internal，794
  named parameter，785
  output parameter，785
  polymorphic，779
  RETURNS TABLE，790
  user-defined，779
    in C，795
    in SQL，780
  variadic，786
  with SETOF，788
fuzzystrmatch，2314

G

generate_series，242
generate_subscripts，243
genetic query optimization，
GEQO (见 genetic query optimization)
geqo_effort配置参数，
geqo_generations配置参数，
geqo_pool_size配置参数，
geqo_seed配置参数，
geqo_selection_bias配置参数，
geqo_threshold配置参数，
geqo配置参数，
get_bit，170
get_byte，170
get_current_ts_config，211
GIN (见索引)
gin_fuzzy_search_limit配置参数，
GiST (见索引)
global data
  in PL/Python，963
  in PL/Tcl，936
GRANT，64，1281
GREATEST，230
  结果类型的确定，269
GROUP BY，13，92
GSSAPI，451
GUID，132

H

hash (见索引)
has_any_column_privilege，247
has_column_privilege，247
has_database_privilege，247
has_foreign_data_wrapper_privilege，247
has_function_privilege，247
has_language_privilege，247
has_schema_privilege，247
has_sequence_privilege，247
has_server_privilege，247
has_tablespace_privilege，247
has_table_privilege，247
HAVING，13，93
hba_file配置参数，
height，205
hierarchical database，7
high availability，499
history
  of PostgreSQL，xxxiii
host，209
hostmask，209
Hot Standby，499
hot_standby_feedback配置参数，
hot_standby配置参数，
HP-UX
  IPC 配置，
  shared library，
  在其上安装，366
hstore，2317

I

ident，453
ident_file配置参数，
IFNULL，229
ignore_system_indexes配置参数，
IMMUTABLE，793
IN，237，240
include
  配置文件中的，397
index
  building concurrently，1153
  for user-defined data type，827
  GIN，1625
  GiST，1616
index scan，
inet_client_addr，245

inet_client_port，245
inet_server_addr，245
inet_server_port，245
inet（数据类型），127
information schema，734
inheritance，21，68，
initcap，160
initdb，377，1455
Initialization Fork，1634
INSERT，8，81，1287
instr，928
int2 (见 smallint)
int4 (见整数)
int8 (见 bigint)
intagg，2323
intarray，2325
integer，36
integer_datetimes配置参数，
interfaces
  externally maintained，2430
internal，147
INTERSECT，96
interval，120
  输出格式，121
    (参见 formatting)
IntervalStyle配置参数，
IRIX
  installation on，367
  shared library，
IS DISTINCT FROM，150，240
IS DOCUMENT，219
IS FALSE，151
IS NOT DISTINCT FROM，150，240
IS NOT FALSE，151
IS NOT NULL，150
IS NOT TRUE，151
IS NOT UNKNOWN，151
IS NULL，150，
IS TRUE，151
IS UNKNOWN，151
isclosed，205
isfinite，194
isn，2328
ISNULL，150
isopen，205
is_array_ref
  in PL/Perl，

J

join
  controlling the order，337
join_collapse_limit配置参数，
justify_days，194
justify_hours，194
justify_interval，195

K

Kerberos，452
key word
  list of，1680
krb_caseins_users配置参数，
krb_server_keyfile配置参数，
krb_srvname配置参数，

L

lag，236
language_handler，147
large object，622
lastval，225
last_value，236
lc_collate配置参数，
lc_ctype配置参数，
lc_messages配置参数，
lc_monetary配置参数，
lc_numeric配置参数，
lc_time配置参数，
LDAP，354，454
LDAP connection parameter lookup，605
ldconfig，361
lead，236
LEAST，230
  结果类型的确定，269
left，160
length，160，170，205，211
length(tsvector)，
lex，350
libedit，349
libperl，350
libpq，556
libpq-fe.h，556，565
libpq-int.h，565
libpython，350
library finalization function，795
library initialization function，795
LIKE，172
  与区域设置，469
LIMIT，97
Linux
  IPC 配置，384
  shared library，
  启动脚本，379
LISTEN，1290
listen_addresses配置参数，
ln，153
lo，2332
LOAD，1292

load balancing，499
localtime，195
localtimestamp，195
local_preload_libraries配置参数，
LOCK，324，1293
lock
  monitoring，528
log，153
log shipping，499
logging_collector配置参数，
log_autovacuum_min_duration配置参数，
log_btree_build_stats配置参数，
log_checkpoints配置参数，
log_connections配置参数，
log_destination配置参数，
log_directory配置参数，
log_disconnections配置参数，
log_duration配置参数，
log_error_verbosity配置参数，
log_executor_stats配置参数，
log_filename配置参数，
log_file_mode配置参数，
log_hostname配置参数，
log_line_prefix配置参数，
log_lock_waits配置参数，
log_min_duration_statement配置参数，
log_min_error_statement配置参数，
log_min_messages配置参数，
log_parser_stats配置参数，
log_planner_stats配置参数，
log_rotation_age配置参数，
log_rotation_size配置参数，
log_statement_stats配置参数，
log_statement配置参数，
log_temp_files配置参数，
log_timezone配置参数，
log_truncate_on_rotation配置参数，
looks_like_number
  in PL/Perl，
loop
  in PL/pgSQL，900
lower，157
  与区域设置，469
lo_close，625
lo_compat_privileges配置参数，
lo_creat，622，625
lo_create，623，625
lo_export，623，625
lo_import，623，625
lo_import_with_oid，623
lo_lseek，624
lo_open，624
lo_read，624
lo_tell，624
lo_truncate，625
lo_unlink，625，625
lo_write，624
lpad，161
lseg，125，206
ltree，2334
ltrim，161

M

Mac OS X
  IPC 配置，
  shared library，
macaddr（数据类型），128
MAC地址 (见 macaddr)
magic block，795
maintenance_work_mem配置参数，
make，349
MANPATH，362
masklen，209
max，233
max_connections配置参数，
max_files_per_process配置参数，
max_function_args配置参数，
max_identifier_length配置参数，
max_index_keys配置参数，
max_locks_per_transaction配置参数，
max_pred_locks_per_transaction配置参数，

max_prepared_transactions配置参数，
max_stack_depth配置参数，
max_standby_archive_delay配置参数，
max_standby_streaming_delay配置参数，
max_wal_senders配置参数，
md5，161，170
MD5，451
min，233
MinGW
  在其上安装，368
mod，153
monitoring
  database activity，519
MOVE，1296
MVCC，319

N

name
  qualified，65
  unqualified，66
NaN (见非数字)
NetBSD
  IPC 配置，
  shared library，
  启动脚本，379

netmask，209
network，209
  数据类型，127
nextval，225
NFS (见网络文件系统)
non-durable，341
nonblocking connection，582
NOT IN，237，240
notation
  functions，50
notice processing
  in libpq，596
NOTIFY，1298
  in libpq，588
NOTNULL，150
NOT（操作符），149
now，195
npoints，205
nth_value，236
ntile，236
null value
  with check constraints，55
  default value，53
  in libpq，578
  with unique constraints，57
NULLIF，229
numeric，36
numnode，211，
NVL，229

O

object-oriented database，7
obj_description，251
octet_length，157，169
OFFSET，97
OID
  libpq 中的，579
  列，60
oid，145
oid2name，2341
ONLY，85
opaque，147
OpenBSD
  IPC 配置，
  shared library，
  启动脚本，379
OpenSSL，
  (参见 SSL)
operator，149
Oracle
  porting from PL/SQL to PL/pgSQL，925
ORDER BY，10，96
  与区域设置，469
ordering operator，835
OR（操作符），149
OVER clause，43
overlay，157，169
overloading
  functions，792
owner，64

P

pageinspect，2345
palloc，804
PAM，354，456
partitioning，71
password
  authentication，451
password file，605
passwordcheck，2347
password_encryption配置参数，
path，206
  for schemas，
PATH，362
path (data type)，126
patterns
  in psql and pg_dump，1437
pause_at_recovery_target recovery parameter，

pclose，205
peer，454
percent_rank，236
performance，331
perl，351
Perl，942
permission (见 privilege)
pfree，804
PGAPPNAME，604
pgbench，2350
PGcancel，586
PGCLIENTENCODING，604
PGconn，556
PGCONNECT_TIMEOUT，604
pgcrypto，2359
PGDATA，377
PGDATABASE，603
PGDATESTYLE，604
PGEventProc，599
PGGEQO，604
PGGSSLIB，604
PGHOST，603
PGHOSTADDR，603
PGKRBSRVNAME，604
PGLOCALEDIR，605
PGOPTIONS，604
PGPASSFILE，604

PGPASSWORD，604
PGPORT，603
PGREALM，604
PGREQUIREPEER，604
PGREQUIRESSL，604
PGresult，572
pgrowlocks，2372
PGSERVICE，604
PGSERVICEFILE，604
PGSSLCERT，604
PGSSLCRL，604
PGSSLKEY，604
PGSSLMODE，604
PGSSLROOTCERT，604
pgstattuple，2381
PGSYSCONFDIR，605
PGTZ，604
PGUSER，604
pgxs，842
pg_advisory_lock，260
pg_advisory_lock_shared，260
pg_advisory_unlock，260
pg_advisory_unlock_all，260
pg_advisory_unlock_shared，260
pg_advisory_xact_lock，260
pg_advisory_xact_lock_shared，260
pg_aggregate，1485
pg_am，1486
pg_amop，1488
pg_amproc，1489
pg_archivecleanup，2348
pg_attrdef，1489
pg_attribute，1490
pg_authid，1492
pg_auth_members，1493
pg_available_extensions，1535
pg_available_extension_versions，1535
pg_backend_pid，245
pg_basebackup，1391
pg_buffercache，2357
pg_cancel_backend，253
pg_cast，1494
pg_class，1495
pg_client_encoding，161
pg_collation，1500
pg_collation_is_visible，249
pg_column_size，256
pg_config，1395
  用于 libpq，611
  用于用户定义的 C 函数，804
pg_conf_load_time，246
pg_constraint，1498
pg_controldata，1458
pg_conversion，1501
pg_conversion_is_visible，249
pg_create_restore_point，254
pg_ctl，377，379，1459
pg_current_xlog_insert_location，254
pg_current_xlog_location，254
pg_cursors，1536
pg_database，465，1501
pg_database_size，256
pg_db_role_setting，1503
pg_default_acl，1503
pg_depend，1504
pg_describe_object，249
pg_description，1505
pg_dump，1398
pg_dumpall，1407
  在升级期间使用，391
pg_enum，1506
pg_extension，1506
pg_extension_config_dump，840
pg_file_rename，2260
pg_file_unlink，2260
pg_file_write，2260
pg_foreign_data_wrapper，1507
pg_foreign_server，1508
pg_foreign_table，1508
pg_freespacemap，2370
pg_function_is_visible，249
pg_get_constraintdef，249
pg_get_expr，249
pg_get_functiondef，249
pg_get_function_arguments，249
pg_get_function_identity_arguments，249
pg_get_function_result，249
pg_get_indexdef，249
pg_get_keywords，249
pg_get_ruledef，249
pg_get_serial_sequence，249
pg_get_triggerdef，249
pg_get_userbyid，249
pg_get_viewdef，249
pg_group，1537
pg_has_role，247
pg_hba.conf，443
pg_ident.conf，449
pg_index，1508
pg_indexes，1537
pg_indexes_size，256
pg_inherits，1511
pg_is_in_recovery，255
pg_is_other_temp_schema，246
pg_is_xlog_replay_paused，256
pg_language，1511
pg_largeobject，1512
pg_largeobject_metadata，1513

pg_last_xact_replay_timestamp，255
pg_last_xlog_receive_location，255
pg_last_xlog_replay_location，255
pg_listening_channels，245
pg_locks，1537
pg_logdir_ls，2260
pg_ls_dir，258
pg_my_temp_schema，246
pg_namespace，1513
pg_notify，1299
pg_opclass，1513
pg_opclass_is_visible，249
pg_operator，1514
pg_operator_is_visible，249
pg_opfamily，1515
pg_options_to_table，249
pg_pltemplate，1515
pg_postmaster_start_time，246
pg_prepared_statements，1540
pg_prepared_xacts，1541
pg_proc，1516
pg_read_binary_file，258
pg_read_file，258
pg_relation_filenode，257
pg_relation_filepath，257
pg_relation_size，256
pg_reload_conf，253
pg_restore，1412
pg_rewrite，1519
pg_roles，1541
pg_rotate_logfile，253
pg_rules，1542
pg_seclabel，1520
pg_seclabels，1543
pg_service.conf，605
pg_settings，1543
pg_shadow，1545
pg_shdepend，1521
pg_shdescription，1522
pg_size_pretty，256
pg_sleep，203
pg_standby，2374
pg_start_backup，254
pg_statio_all_indexes，522
pg_statio_all_sequences，522
pg_statio_all_tables，522
pg_statio_sys_indexes，522
pg_statio_sys_sequences，522
pg_statio_sys_tables，522
pg_statio_user_indexes，522
pg_statio_user_sequences，522
pg_statio_user_tables，522
pg_statistic，336，1522
pg_stats，336，1546
pg_stat_activity，521
pg_stat_all_indexes，522
pg_stat_all_tables，521
pg_stat_bgwriter，521
pg_stat_database，521
pg_stat_database_conflicts，521
pg_stat_file，258
pg_stat_replication，521
pg_stat_statements，2377
pg_stat_sys_indexes，522
pg_stat_sys_tables，522
pg_stat_user_functions，523
pg_stat_user_indexes，522
pg_stat_user_tables，522
pg_stat_xact_all_tables，522
pg_stat_xact_sys_tables，522
pg_stat_xact_user_functions，523
pg_stat_xact_user_tables，522
pg_stop_backup，254
pg_switch_xlog，254
pg_tables，1548
pg_tablespace，1524
pg_tablespace_databases，249
pg_tablespace_size，256
pg_table_is_visible，249
pg_table_size，256
pg_terminate_backend，253
pg_test_fsync，2383
pg_timezone_abbrevs，1548
pg_timezone_names，1549
pg_total_relation_size，256
pg_trgm，2384
pg_trigger，1525
pg_try_advisory_lock，260
pg_try_advisory_lock_shared，260
pg_try_advisory_xact_lock，260
pg_try_advisory_xact_lock_shared，260
pg_ts_config，1526
pg_ts_config_is_visible，249
pg_ts_config_map，1527
pg_ts_dict，1527
pg_ts_dict_is_visible，249
pg_ts_parser，1528
pg_ts_parser_is_visible，249
pg_ts_template，1528
pg_ts_template_is_visible，249
pg_type，1528
pg_typeof，249
pg_type_is_visible，249
pg_upgrade，2387
pg_user，1549
pg_user_mapping，1534
pg_user_mappings，1550
pg_views，1550

pg_xlogfile_name，254
pg_xlogfile_name_offset，254
pg_xlog_replay_pause，256
pg_xlog_replay_resume，256
pi，154
PIC，804
PID
  确定服务器进程的 PID
    在 libpq 中，568
PITR，486
PITR standby，499
PL/Perl，942
PL/PerlU，952
PL/pgSQL，881
PL/Python，956
PL/SQL (Oracle)
  porting to PL/pgSQL，925
PL/Tcl，935
plainto_tsquery，212，289
plperl.on_init configuration parameter，
plperl.on_plperlu_init configuration parameter，
plperl.on_plperl_init configuration parameter，
plperl.use_strict configuration parameter，
plpgsql.variable_conflict configuration parameter，921
point，125，207
polygon，126，207
polymorphic function，779
polymorphic type，779
popen，205
port配置参数，
position，157，169
POSTGRES，xxxiii
postgres，3，378，463，1466
postgres 用户，377
Postgres95，xxxiii
postgresql.conf，397
postmaster，1473
post_auth_delay配置参数，
power，154
PQbackendPID，568
PQbinaryTuples，577
  with COPY，589
PQcancel，586
PQclear，575
PQclientEncoding，592
PQcmdStatus，579
PQcmdTuples，579
PQconndefaults，563
PQconnectdb，560
PQconnectdbParams，556
PQconnectionNeedsPassword，568
PQconnectionUsedPassword，569
PQconnectPoll，561
PQconnectStart，561
PQconnectStartParams，561
PQconninfoFree，594
PQconninfoParse，564
PQconsumeInput，585
PQcopyResult，595
PQdb，566
PQdescribePortal，572
PQdescribePrepared，572
PQencryptPassword，594
PQendcopy，592
PQerrorMessage，568
PQescapeBytea，582
PQescapeByteaConn，581
PQescapeIdentifier，580
PQescapeLiteral，580
PQescapeString，581
PQescapeStringConn，581
PQexec，569
PQexecParams，569
PQexecPrepared，571
PQfformat，577
  with COPY，589
PQfinish，564
PQfireResultCreateEvents，594
PQflush，586
PQfmod，577
PQfn，587
PQfname，576
PQfnumber，576
PQfreeCancel，586
PQfreemem，594
PQfsize，577
PQftable，576
PQftablecol，576
PQftype，577
PQgetCancel，586
PQgetCopyData，590
PQgetisnull，578
PQgetlength，578
PQgetline，591
PQgetlineAsync，591
PQgetResult，584
PQgetssl，569
PQgetvalue，577
PQhost，566
PQinitOpenSSL，609
PQinitSSL，609
PQinstanceData，600
PQisBusy，585
PQisnonblocking，586
PQisthreadsafe，610

PQlibVersion，596
(参见 PQserverVersion)
PQmakeEmptyPGresult，594
PQnfields，575
用于 COPY，589
PQnotifies，588
PQnparams，578
PQntuples，575
PQoidStatus，580
PQoidValue，579
PQoptions，566
PQparameterStatus，567
PQparamtype，578
PQpass，566
PQping，565
PQpingParams，565
PQport，566
PQprepare，571
PQprint，578
PQprotocolVersion，568
PQputCopyData，589
PQputCopyEnd，590
PQputline，591
PQputnbytes，592
PQregisterEventProc，600
PQrequestCancel，587
PQreset，564
PQresetPoll，564
PQresetStart，564
PQresStatus，573
PQresultAlloc，595
PQresultErrorField，574
PQresultErrorMessage，573
PQresultInstanceData，600
PQresultSetInstanceData，600
PQresultStatus，572
PQsendDescribePortal，584
PQsendDescribePrepared，584
PQsendPrepare，583
PQsendQuery，583
PQsendQueryParams，583
PQsendQueryPrepared，583
PQserverVersion，568
PQsetClientEncoding，593
PQsetdb，561
PQsetdbLogin，561
PQsetErrorVerbosity，593
PQsetInstanceData，600
PQsetnonblocking，585
PQsetNoticeProcessor，596
PQsetNoticeReceiver，596
PQsetResultAttrs，595
PQsetvalue，595
PQsocket，568
PQstatus，566
PQtrace，593
PQtransactionStatus，567
PQtty，566
PQunescapeBytea，582
PQuntrace，593
PQuser，566
PREPARE，1300
PREPARE TRANSACTION，1302
prepared statements
creating，1300
executing，1271
removing，1223
showing the query plan，1272
preparing a query
in PL/pgSQL，922
in PL/Python，964
in PL/Tcl，938
pre_auth_delay配置参数，
primary key，57
primary_conninfo recovery parameter，
privilege，64
querying，246
with rules，873
for schemas，67
with views，873
procedural language，878
externally maintained，2430
handler for，1596
protocol
frontend-backend，1552
ps
to monitor activity，519
psql，5，1419
Python，956

Q

qualified name，65
query，9
query plan，331
query tree，855
querytree，212，
quote_all_identifiers配置参数，
quote_ident，161
in PL/Perl，
use in PL/pgSQL，892
quote_literal，161
in PL/Perl，
use in PL/pgSQL，892
quote_nullable，162
in PL/Perl，
use in PL/pgSQL，892

R

radians，154
radius，205
RADIUS，456
RAISE，911
random，155
random_page_cost配置参数，
rank，236
read-only transaction
  setting，1348
  setting default，
readline，349
real，107
REASSIGN OWNED，1304
record，147
recovery.conf，516
recovery_end_command recovery parameter，
recovery_target_inclusive recovery parameter，
recovery_target_name recovery parameter，
recovery_target_time recovery parameter，
recovery_target_timeline recovery parameter，
recovery_target_xid recovery parameter，
rectangle，126
referential integrity，16，58
regclass，145
regconfig，145
regdictionary，145
regexp_matches，162，174
regexp_replace，162，174
regexp_split_to_array，162，174
regexp_split_to_table，163，174
regoper，145
regoperator，145
regproc，145
regprocedure，145
regression tests，544
regr_avgx，235
regr_avgy，235
regr_count，235
regr_intercept，235
regr_r2，235
regr_slope，235
regr_sxx，235
regr_sxy，235
regr_syy，235
regtype，145
REINDEX，1305
reindexdb，1448
relation，7
relational database，7
RELEASE SAVEPOINT，1308
repeat，163
replace，163
replication，499
replication_timeout配置参数，
reporting errors
  in PL/pgSQL，911
RESET，1309
restart_after_crash配置参数，
restore_command recovery parameter，
RESTRICT
  with DROP，79
  外键动作，59
RETURN NEXT
  in PL/pgSQL，896
RETURN QUERY
  in PL/pgSQL，896
RETURNING INTO
  in PL/pgSQL，890
reverse，163
REVOKE，64，1310
right，163
role，458
  applicable，735
  enabled，750
  membership in，460
ROLLBACK，1314
rollback
  psql，
ROLLBACK PREPARED，1315
ROLLBACK TO SAVEPOINT，1316
round，154
row，7，52
ROW，47
row estimation
  planner，1640
row type，142
row_number，236
rpad，163
rtrim，164
rule，855
  and views，856
  for DELETE，863
  for INSERT，863
  for SELECT，857
  compared with triggers，875
  for UPDATE，863

S

SAVEPOINT，1318
savepoints

defining，1318
releasing，1308
rolling back，1316
schema，65，463
creating，65
current，66
public，66
removing，66
当前，245
SCO
安装，368
SCO OpenServer
IPC 配置，385
search path，66
当前，245
search_path，66
search_path配置参数，
SECURITY LABEL，1320
seg，2393
segment_size配置参数，
SELECT，9，84，1322
选择列表，94
SELECT INTO，1338
in PL/pgSQL，890
sepgsql，2396
sepgsql.debug_audit configuration parameter，
sepgsql.permissive configuration parameter，
sequence
and serial type，108
sequential scan，
seq_page_cost配置参数，
serial，108
serial4，108
serial8，108
server log，418
server_encoding配置参数，
server_version_num配置参数，
server_version配置参数，
session_replication_role配置参数，
session_user，245
SET，253，1340
SET CONSTRAINTS，1343
SET ROLE，1344
SET SESSION AUTHORIZATION，1346
SET TRANSACTION，1348
SET XML OPTION，
setseed，155
setval，225
setweight，212，
set_bit，170
set_byte，171
set_config，253
set_masklen，209
shared_buffers配置参数，
shared_preload_libraries，816
shared_preload_libraries配置参数，
SHMMAX，382
shobj_description，251
SHOW，253，1350
SIGHUP，397，447，450
SIGINT，389
sign，154
significant digits，
SIGQUIT，389
SIGTERM，389
silent_mode配置参数，
SIMILAR TO，173
sin，156
sliced bread (见 TOAST)
smallint，106
Solaris
IPC 配置，386
shared library，
启动脚本，380
在其上安装，369
SOME，234，237，240
SPI，969
examples，2402
SPI_connect，970
SPI_copytuple，1011
spi_cursor_close
in PL/Perl，
SPI_cursor_close，998
SPI_cursor_fetch，994
SPI_cursor_find，993
SPI_cursor_move，995
SPI_cursor_open，989
SPI_cursor_open_with_args，990
SPI_cursor_open_with_paramlist，992
SPI_exec，977
SPI_execp，988
SPI_execute，974
SPI_execute_plan，986
SPI_execute_plan_with_paramlist，987
SPI_execute_with_args，978
spi_exec_prepared
in PL/Perl，
spi_exec_query
in PL/Perl，
spi_fetchrow
in PL/Perl，
SPI_finish，971
SPI_fname，1000
SPI_fnumber，1001
spi_freeplan
in PL/Perl，

SPI_freeplan，1016
SPI_freetuple，1014
SPI_freetuptable，1015
SPI_getargcount，983
SPI_getargtypeid，984
SPI_getbinval，1003
SPI_getnspname，1007
SPI_getrelname，1006
SPI_gettype，1004
SPI_gettypeid，1005
SPI_getvalue，1002
SPI_is_cursor_plan，985
spi_lastoid，
SPI_modifytuple，1013
SPI_palloc，1008
SPI_pfree，1010
SPI_pop，973
spi_prepare
  in PL/Perl，
SPI_prepare，980
SPI_prepare_cursor，981
SPI_prepare_params，982
SPI_push，972
spi_query
  in PL/Perl，
spi_query_prepared
  in PL/Perl，
SPI_repalloc，1009
SPI_returntuple，1012
SPI_saveplan，999
SPI_scroll_cursor_fetch，996
SPI_scroll_cursor_move，997
split_part，164
SQL/CLI，1705
SQL/Foundation，1705
SQL/Framework，1705
SQL/JRT，1705
SQL/MED，1705
SQL/OLB，1705
SQL/PSM，1705
SQL/Schemata，1705
SQL/XML，1705
sql_inheritance配置参数，
sqrt，155
ssh，395
SSI，319
SSL，393，606
  在 libpq 中，559
  在libpq中，569
sslinfo，2404
ssl_ciphers配置参数，
ssl_renegotiation_limit配置参数，
ssl配置参数，
SSPI，452
STABLE，793
standard_conforming_strings配置参数，
standby server，499
standby_mode recovery parameter，
START TRANSACTION，1352
statement_timeout配置参数，
statement_timestamp，195
statistics，520
  of the planner，336
stats_temp_directory配置参数，
stddev，235
stddev_pop，235
stddev_samp，235
STONITH，499
storage parameters，1189
Streaming Replication，499
string (见字符串)
strings
  backslash quotes，
  escape warning，
  standard conforming，
string_agg，234
string_to_array，231
strip，212，
strpos，164
subquery，13
substr，164
substring，157，169，173，174
sum，234
superuser，5
superuser_reserved_connections配置参数，

suppress_redundant_updates_trigger，260
synchronize_seqscans配置参数，
Synchronous Replication，499
synchronous_commit配置参数，
synchronous_standby_names配置参数，
syntax
  SQL，31
syslog_facility配置参数，
syslog_ident配置参数，
system catalog
  schema，68

T

table，7，52
  creating，52
  inheritance，68
  modifying，61
  partitioning，71
  removing，53
  renaming，64
TABLE command，1322

tablefunc，2406
tableoid，61
tablespace，466
  default，
  temporary，
tan，156
Tcl，935
tcp_keepalives_count配置参数，
tcp_keepalives_idle配置参数，
tcp_keepalives_interval配置参数，
template0，464
template1，464，464
temp_buffers配置参数，
temp_tablespaces配置参数，
test，544
test_parser，2416
text，110，210
threads
  with libpq，610
tid，145
time，113，116
  constants，117
  输出格式，118
    (参见 formatting)
time zone，118，
  input abbreviations，1677
time zone names，
timelines，486
timeofday，195
timeout
  client authentication，
  deadlock，
timestamp，113，117
timestamptz，113
timezone_abbreviations配置参数，
timezone配置参数，
TOAST，1632
  and user-defined types，822
  per-column storage settings，
  versus large objects，622
to_ascii，164
to_char，185
  与区域设置，469
to_date，186
to_hex，164
to_number，186
to_timestamp，186
to_tsquery，212，289
to_tsvector，212，288
trace_locks配置参数，
trace_lock_oidmin配置参数，
trace_lock_table配置参数，
trace_lwlocks配置参数，
trace_notify配置参数，
trace_recovery_messages配置参数，
trace_sort配置参数，
trace_userlocks配置参数，
track_activities配置参数，
track_activity_query_size配置参数，
track_counts配置参数，
track_functions配置参数，
transaction，17
transaction isolation level
  setting，1348
  setting default，
transaction_timestamp，195
transform_null_equals配置参数，
translate，164
trigger，147，846
  in C，848
  in PL/pgSQL，913
  in PL/Python，963
  in PL/Tcl，939
  compared with rules，875
trigger_file recovery parameter，
trim，158，169
Tru64 UNIX
  shared library，
true，122
trunc，155，210
TRUNCATE，1353
trusted
  PL/Perl，952
tsearch2，2418
tsquery（数据类型），131
tsvector 连接，
tsvector_update_trigger，213
tsvector_update_trigger_column，213
tsvector（数据类型），130
ts_debug，214，309
ts_headline，212，292
ts_lexize，214，313
ts_parse，214，311
ts_rank，213，
ts_rank_cd，213，
ts_rewrite，213，295
ts_stat，215，297
ts_token_type，214，312
txid_current，252
txid_current_snapshot，252
txid_snapshot_xip，252
txid_snapshot_xmax，252
txid_snapshot_xmin，252
txid_visible_in_snapshot，252
type
  polymorphic，779
type cast，44

U

UESCAPE，32，34
unaccent，2420，2421
Unicode 转义
  在字符串常量中，34
  在标识符中，32
UNION，96
  结果类型的确定，269
unique constraint，56
Unix 域套接字，557
UnixWare
  IPC 配置，386
  shared library，
  安装，368
unix_socket_directory配置参数，
unix_socket_group配置参数，
unix_socket_permissions配置参数，
UNLISTEN，1355
unnest，231
unqualified name，66
UPDATE，14，82，1357
update_process_title配置参数，
upper，158
  与区域设置，469
user，245，458
  current，245
user mapping，78
User name maps，449
UUID，132，355
uuid-ossp，2422

V

VACUUM，1361
vacuumdb，1451
vacuumlo，2424
vacuum_cost_delay配置参数，
vacuum_cost_limit配置参数，
vacuum_cost_page_dirty配置参数，
vacuum_cost_page_hit配置参数，
vacuum_cost_page_miss配置参数，
vacuum_defer_cleanup_age配置参数，
vacuum_freeze_min_age配置参数，
vacuum_freeze_table_age配置参数，
VALUES，98，1364
  结果类型的确定，269
varchar，110
variadic function，786
variance，235
var_pop，235
var_samp，235
version，5，246
view，16
  implementation through rules，856
  updating，867
Visibility Map，1633
VM (见 Visibility Map)
void，147
VOLATILE，793
volatility
  functions，793
VPATH，351

W

WAL，538
wal_block_size配置参数，
wal_buffers配置参数，
wal_debug配置参数，
wal_keep_segments配置参数，
wal_level配置参数，
wal_receiver_status_interval配置参数，
wal_segment_size配置参数，
wal_sender_delay配置参数，
wal_sync_method配置参数，
wal_writer_delay配置参数，
warm standby，499
WHERE，91
WHILE
  in PL/pgSQL，901
width，205
width_bucket，155
window function，19
WITH
  in SELECT，1322
  在 SELECT 中，99
witness server，499
work_mem配置参数，

X

xid，145
xmax，61
xmin，61
XML，132
XML export，222
XML option，
XML 选项，133
xml2，2426
xmlagg，218，234
xmlbinary配置参数，
xmlcomment，215
xmlconcat，215
xmlelement，216
XMLEXISTS，219
xmlforest，217
xmloption配置参数，
xmlparse，133
xmlpi，218

xmlroot，218
xmlserialize，133
xml_is_well_formed，220
xml_is_well_formed_content，220
xml_is_well_formed_document，220
XPath，221
xpath_exists，221

Y

yacc，350

Z

zero_damaged_pages配置参数，
zlib，349，357