PostgreSQL 8.1.23 手册

PostgreSQL 全球开发组

翻译：冯若航[1]，Pigsty 项目组[2]

[1]https://vonng.com
[2]https://pigsty.cc

PostgreSQL 8.1.23 手册

PostgreSQL 全球开发组
翻译：冯若航[1]，Pigsty 项目组[2]
版权 © 1996-2005 PostgreSQL 全球开发组

法律声明

PostgreSQL 版权所有 © 1996-2005 PostgreSQL 全球开发组，并按照下文所述加利福尼亚大学的许可条款分发。

Postgres95 版权所有 © 1994-5 加利福尼亚大学董事会。

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

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

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

[1] https://vonng.com
[2] https://pigsty.cc

目录

前言 xxiv
1. 什么是PostgreSQL? xxiv
2. PostgreSQL简史 xxv
2.1. 伯克利 POSTGRES 项目 xxv
2.2. Postgres95 xxv
2.3. PostgreSQL xxvi
3. 约定 xxvi
4. 更多信息 xxvi
5. 缺陷报告指南 xxvii
5.1. 识别缺陷 xxvii
5.2. 报告哪些内容 xxviii
5.3. 到哪里报告缺陷 xxix
I. 教程 1
1. 从头开始 3
1.1. 安装 3
1.2. 架构基础 3
1.3. 创建数据库 3
1.4. 访问数据库 5
2. SQL语言 7
2.1. 简介 7
2.2. 概念 7
2.3. 创建一个新表 7
2.4. 向表中填充行 8
2.5. 查询表 9
2.6. 表之间的连接 10
2.7. 聚合函数 12
2.8. 更新 14
2.9. 删除 14
3. 高级特性 15
3.1. 简介 15
3.2. 视图 15
3.3. 外键 15
3.4. 事务 16
3.5. 继承 18
3.6. 小结 19
II. SQL 语言 20
4. SQL语法 25
4.1. 词法结构 25
4.2. 值表达式 31
5. 数据定义 39
5.1. 表基础 39
5.2. 默认值 40
5.3. 约束 41
5.4. 系统列 47
5.5. 修改表 48
5.6. 权限 50
5.7. 模式 51
5.8. 继承 55
5.9. 分区 58
5.10. 其他数据库对象 64
5.11. 依赖跟踪 64

6. 数据操纵 66
6.1. 插入数据 66
6.2. 更新数据 67
6.3. 删除数据 67
7. 查询 69
7.1. 概述 69
7.2. 表表达式 69
7.3. 选择列表 78
7.4. 组合查询 79
7.5. 行排序 80
7.6. LIMIT和OFFSET 81
8. 数据类型 82
8.1. 数字类型 83
8.2. 货币类型 86
8.3. 字符类型 86
8.4. 二进制数据类型 88
8.5. 日期/时间类型 89
8.6. 布尔类型 95
8.7. 几何类型 96
8.8. 网络地址类型 98
8.9. 位串类型 100
8.10. 数组 101
8.11. 复合类型 108
8.12. 对象标识符类型 112
8.13. 伪类型 113
9. 函数和操作符 114
9.1. 逻辑操作符 114
9.2. 比较操作符 114
9.3. 数学函数和操作符 116
9.4. 字符串函数和操作符 121
9.5. 二进制串函数和操作符 130
9.6. 位串函数和操作符 132
9.7. 模式匹配 133
9.8. 数据类型格式化函数 143
9.9. 日期/时间函数和操作符 148
9.10. 几何函数和操作符 159
9.11. 网络地址函数和操作符 162
9.12. 序列操作函数 164
9.13. 条件表达式 166
9.14. 数组函数和操作符 168
9.15. 聚合函数 170
9.16. 子查询表达式 172
9.17. 行和数组比较 175
9.18. 集合返回函数 176
9.19. 系统信息函数 178
9.20. 系统管理函数 183
10. 类型转换 186
10.1. 概述 186
10.2. 操作符 187
10.3. 函数 190
10.4. 值存储 192
10.5. UNION、CASE及相关结构 193
11. 索引 195

11.1. 简介 195
11.2. 索引类型 195
11.3. 多列索引 197
11.4. 组合多个索引 197
11.5. 唯一索引 198
11.6. 表达式索引 198
11.7. 部分索引 199
11.8. 操作符类 201
11.9. 检查索引使用情况 202
12. 并发控制 204
12.1. 介绍 204
12.2. 事务隔离 204
12.3. 显式锁定 207
12.4. 应用级别的数据一致性检查 210
12.5. 锁定和索引 210
13. 性能提示 212
13.1. 使用EXPLAIN 212
13.2. 规划器使用的统计信息 216
13.3. 用显式JOIN子句控制规划器 218
13.4. 填充一个数据库 219
III. 服务器管理 222
14. 安装说明 226
14.1. 简要说明 226
14.2. 需求 226
14.3. 获取源代码 228
14.4. 如果你要升级 228
14.5. 安装过程 229
14.6. 安装后设置 234
14.7. 受支持的平台 236
15. Windows上仅安装客户端 241
16. 操作系统环境 243
16.1. The PostgreSQL User Account 243
16.2. 创建一个数据库集簇 243
16.3. 启动数据库服务器 244
16.4. 管理内核资源 247
16.5. 关闭服务器 253
16.6. 加密选项 253
16.7. 使用 SSL 的安全 TCP/IP 连接 254
16.8. Secure TCP/IP Connections with SSH Tunnels 255
17. 服务器配置 257
17.1. 设置参数 257
17.2. 文件位置 258
17.3. 连接和认证 259
17.4. 资源消耗 261
17.5. 预写式日志 265
17.6. 查询规划 267
17.7. 错误报告和日志 271
17.8. 运行时统计 275
17.9. 自动清理 276
17.10. 客户端连接默认值 277
17.11. 锁管理 280
17.12. 版本和平台兼容性 280
17.13. 预置选项 282

17.14. 自定义选项 283
17.15. 开发者选项 283
17.16. 短选项 284
18. 数据库角色和权限 286
18.1. 数据库角色 286
18.2. 角色属性 287
18.3. 权限 288
18.4. 角色成员资格 288
18.5. 函数和触发器 290
19. 管理数据库 291
19.1. 概述 291
19.2. 创建数据库 291
19.3. 模板数据库 292
19.4. 数据库配置 293
19.5. 删除数据库 293
19.6. 表空间 294
20. 客户端认证 296
20.1. pg_hba.conf 文件 296
20.2. 认证方法 301
20.3. 认证问题 304
21. Localization 305
21.1. Locale Support 305
21.2. Character Set Support 307
22. 日常数据库维护任务 313
22.1. 日常清理 313
22.2. 日常重建索引 317
22.3. 日志文件维护 318
23. 备份和恢复 319
23.1. SQL转储 319
23.2. 文件系统级备份 321
23.3. 在线备份和时间点恢复（PITR） 322
23.4. 版本间的迁移 328
24. 监控数据库活动 330
24.1. 标准 Unix 工具 330
24.2. 统计收集器 331
24.3. 查看锁 335
25. 监控磁盘使用情况 336
25.1. 确定磁盘使用情况 336
25.2. 磁盘已满故障 337
26. 可靠性与预写式日志 338
26.1. 可靠性 338
26.2. 预写式日志（WAL） 338
26.3. WAL 配置 339
26.4. WAL 内部机制 340
27. 回归测试 341
27.1. 运行测试 341
27.2. 测试结果评估 342
27.3. 变体比较文件 344
IV. 客户端接口 346
28. libpq - C 库 350
28.1. 数据库连接控制函数 350
28.2. 连接状态函数 355
28.3. 命令执行函数 359

28.4. 异步命令处理 ... 370
28.5. 取消进行中的查询 ... 373
28.6. 快速路径接口 ... 374
28.7. 异步通知 ... 375
28.8. 与COPY命令相关的函数 ... 376
28.9. 控制函数 ... 380
28.10. 通知处理 ... 381
28.11. 环境变量 ... 382
28.12. 密码文件 ... 383
28.13. 连接服务文件 ... 383
28.14. SSL 支持 ... 383
28.15. 在线程化程序中的行为 ... 384
28.16. 构建libpq程序 ... 384
28.17. 示例程序 ... 385
29. 大对象 ... 396
29.1. 历史 ... 396
29.2. 实现特性 ... 396
29.3. 客户端接口 ... 396
29.4. 服务器端函数 ... 399
29.5. 示例程序 ... 399
30. ECPG - C 中的嵌入式 SQL ... 405
30.1. 概念 ... 405
30.2. 连接到数据库服务器 ... 405
30.3. 关闭一个连接 ... 406
30.4. 运行 SQL 命令 ... 407
30.5. 选择一个连接 ... 407
30.6. 使用主变量 ... 408
30.7. 动态 SQL ... 410
30.8. 使用描述符区域 ... 411
30.9. 错误处理 ... 413
30.10. 包含文件 ... 418
30.11. 处理嵌入式 SQL 程序 ... 418
30.12. 库函数 ... 419
30.13. 内部 ... 419
31. 信息模式 ... 422
31.1. 模式 ... 422
31.2. 数据类型 ... 422
31.3. information_schema_catalog_name ... 422
31.4. applicable_roles ... 423
31.5. check_constraints ... 423
31.6. column_domain_usage ... 423
31.7. column_privileges ... 424
31.8. column_udt_usage ... 424
31.9. columns ... 425
31.10. constraint_column_usage ... 427
31.11. constraint_table_usage ... 427
31.12. data_type_privileges ... 428
31.13. domain_constraints ... 428
31.14. domain_udt_usage ... 429
31.15. domains ... 429
31.16. element_types ... 431
31.17. enabled_roles ... 433
31.18. key_column_usage ... 433

31.19. parameters 433
31.20. referential_constraints 435
31.21. role_column_grants 435
31.22. role_routine_grants 436
31.23. role_table_grants 436
31.24. role_usage_grants 437
31.25. routine_privileges 437
31.26. routines 438
31.27. schemata 440
31.28. sql_features 441
31.29. sql_implementation_info 441
31.30. sql_languages 442
31.31. sql_packages 442
31.32. sql_sizing 442
31.33. sql_sizing_profiles 443
31.34. table_constraints 443
31.35. table_privileges 444
31.36. tables 444
31.37. triggers 445
31.38. usage_privileges 446
31.39. view_column_usage 446
31.40. view_table_usage 447
31.41. views 447
V. 服务器编程 448
32. 扩展SQL 452
32.1. 可扩展性如何运作 452
32.2. PostgreSQL 类型系统 452
32.3. 用户定义的函数 453
32.4. 查询语言（SQL）函数 454
32.5. 函数重载 462
32.6. 函数易变性分类 462
32.7. 过程语言函数 464
32.8. 内部函数 464
32.9. C 语言函数 464
32.10. 用户定义的聚合 486
32.11. 用户定义类型 488
32.12. 用户定义操作符 491
32.13. 操作符优化信息 492
32.14. 索引扩展接口 496
33. 触发器 504
33.1. 触发器行为概述 504
33.2. 数据更改的可见性 505
33.3. 用 C 编写触发器函数 505
33.4. 一个完整的示例 508
34. 规则系统 512
34.1. 查询树 512
34.2. 视图和规则系统 513
34.3. INSERT、UPDATE 和 DELETE 上的规则 520
34.4. 规则和权限 529
34.5. 规则和命令状态 531
34.6. 规则与触发器 531
35. Procedural Languages 534
35.1. 安装过程语言 534

36. PL/pgSQL - SQL 过程语言 536
36.1. 概述 536
36.2. PL/pgSQL 开发提示 538
36.3. PL/pgSQL的结构 540
36.4. 声明 541
36.5. 表达式 545
36.6. 基本语句 546
36.7. 控制结构 550
36.8. 游标 558
36.9. 错误和消息 562
36.10. 触发器函数 562
36.11. 从Oracle PL/SQL 移植 568
37. PL/Tcl - Tcl 过程语言 578
37.1. 概述 578
37.2. PL/Tcl 函数和参数 578
37.3. PL/Tcl 中的数据值 579
37.4. PL/Tcl 中的全局数据 579
37.5. 从 PL/Tcl 访问数据库 580
37.6. PL/Tcl 中的触发器函数 582
37.7. 模块与`unknown`命令 583
37.8. Tcl 过程名 584
38. PL/Perl - Perl 过程语言 585
38.1. PL/Perl 函数和参数 585
38.2. 从 PL/Perl 访问数据库 588
38.3. PL/Perl 中的数据值 590
38.4. PL/Perl 中的全局值 590
38.5. 受信任与不受信任的 PL/Perl 591
38.6. PL/Perl 触发器 592
38.7. 限制与缺失特性 593
39. PL/Python - Python 过程语言 594
39.1. PL/Python 函数 594
39.2. 触发器函数 595
39.3. 数据库访问 595
40. 服务器编程接口 597
40.1. 接口函数 597
40.2. 接口支持函数 618
40.3. 内存管理 626
40.4. 数据更改的可见性 635
40.5. 示例 635
VI. 参考 639
I. SQL 命令 643
ABORT 646
ALTER AGGREGATE 647
ALTER CONVERSION 649
ALTER DATABASE 650
ALTER DOMAIN 652
ALTER FUNCTION 655
ALTER GROUP 658
ALTER INDEX 659
ALTER LANGUAGE 661
ALTER OPERATOR 662
ALTER OPERATOR CLASS 663
ALTER ROLE 664

ALTER SCHEMA 667
ALTER SEQUENCE 668
ALTER TABLE 670
ALTER TABLESPACE 677
ALTER TRIGGER 678
ALTER TYPE 679
ALTER USER 680
ANALYZE 681
BEGIN 683
CHECKPOINT 685
CLOSE 686
CLUSTER 687
COMMENT 689
COMMIT 692
COMMIT PREPARED 693
COPY 694
CREATE AGGREGATE 702
CREATE CAST 705
CREATE CONSTRAINT TRIGGER 708
CREATE CONVERSION 709
CREATE DATABASE 711
CREATE DOMAIN 713
CREATE FUNCTION 715
CREATE GROUP 720
CREATE INDEX 721
CREATE LANGUAGE 724
CREATE OPERATOR 727
CREATE OPERATOR CLASS 730
CREATE ROLE 733
CREATE RULE 737
CREATE SCHEMA 739
CREATE SEQUENCE 741
CREATE TABLE 744
CREATE TABLE AS 753
CREATE TABLESPACE 755
CREATE TRIGGER 757
CREATE TYPE 760
CREATE USER 765
CREATE VIEW 766
DEALLOCATE 768
DECLARE 769
DELETE 772
DROP AGGREGATE 774
DROP CAST 775
DROP CONVERSION 776
DROP DATABASE 777
DROP DOMAIN 778
DROP FUNCTION 779
DROP GROUP 781
DROP INDEX 782
DROP LANGUAGE 783
DROP OPERATOR 784
DROP OPERATOR CLASS 786
DROP ROLE 787

DROP RULE 788
DROP SCHEMA 789
DROP SEQUENCE 790
DROP TABLE 791
DROP TABLESPACE 792
DROP TRIGGER 793
DROP TYPE 794
DROP USER 795
DROP VIEW 796
END 797
EXECUTE 798
EXPLAIN 799
FETCH 802
GRANT 806
INSERT 811
LISTEN 813
LOAD 814
LOCK 815
MOVE 817
NOTIFY 818
PREPARE 820
PREPARE TRANSACTION 822
REINDEX 824
RELEASE SAVEPOINT 827
RESET 828
REVOKE 829
ROLLBACK 832
ROLLBACK PREPARED 833
ROLLBACK TO SAVEPOINT 834
SAVEPOINT 836
SELECT 838
SELECT INTO 849
SET 851
SET CONSTRAINTS 854
SET ROLE 855
SET SESSION AUTHORIZATION 857
SET TRANSACTION 859
SHOW 861
START TRANSACTION 863
TRUNCATE 864
UNLISTEN 865
UPDATE 866
VACUUM 869
II. PostgreSQL 客户端应用 872
clusterdb 873
createdb 875
createlang 878
createuser 880
dropdb 884
droplang 886
dropuser 888
ecpg 890
pg_config 892
pg_dump 895

pg_dumpall 901
pg_restore 905
psql 911
reindexdb 935
vacuumdb 938
III. PostgreSQL 服务器端应用 941
initdb 942
ipcclean 945
pg_controldata 946
pg_ctl 947
pg_resetxlog 951
postgres 953
postmaster 957
VII. 内部 962
41. PostgreSQL 内部概述 966
41.1. 查询的路径 966
41.2. 连接是如何建立的 966
41.3. 解析器阶段 967
41.4. PostgreSQL 规则系统 968
41.5. 规划器/优化器 968
41.6. 执行器 969
42. 系统目录 971
42.1. 概述 971
42.2. pg_aggregate 972
42.3. pg_am 972
42.4. pg_amop 973
42.5. pg_amproc 974
42.6. pg_attrdef 974
42.7. pg_attribute 974
42.8. pg_authid 976
42.9. pg_auth_members 977
42.10. pg_autovacuum 977
42.11. pg_cast 978
42.12. pg_class 979
42.13. pg_constraint 980
42.14. pg_conversion 981
42.15. pg_database 982
42.16. pg_depend 983
42.17. pg_description 984
42.18. pg_index 984
42.19. pg_inherits 985
42.20. pg_language 986
42.21. pg_largeobject 986
42.22. pg_listener 987
42.23. pg_namespace 987
42.24. pg_opclass 987
42.25. pg_operator 988
42.26. pg_pltemplate 989
42.27. pg_proc 989
42.28. pg_rewrite 991
42.29. pg_shdepend 992
42.30. pg_statistic 993
42.31. pg_tablespace 994

42.32. pg_trigger 994
42.33. pg_type 995
42.34. 系统视图 999
42.35. pg_group 999
42.36. pg_indexes 1000
42.37. pg_locks 1000
42.38. pg_prepared_xacts 1002
42.39. pg_roles 1002
42.40. pg_rules 1003
42.41. pg_settings 1003
42.42. pg_shadow 1004
42.43. pg_stats 1004
42.44. pg_tables 1006
42.45. pg_user 1006
42.46. pg_views 1007
43. 前端/后端协议 1008
43.1. 概述 1008
43.2. 消息流 1009
43.3. 消息数据类型 1018
43.4. 消息格式 1019
43.5. 错误和通知消息的字段 1033
43.6. 协议 2.0 以来的变更总结 1034
44. PostgreSQL 编码约定 1036
44.1. 格式 1036
44.2. 在服务器内部报告错误 1036
44.3. 错误消息风格指南 1038
45. 本地语言支持 1043
45.1. 给翻译者 1043
45.2. 给程序员 1045
46. 编写过程语言调用处理器 1048
47. 遗传查询优化器 1050
47.1. 将查询处理看成是一个复杂的优化问题 1050
47.2. 遗传算法 1050
47.3. PostgreSQL 中的遗传查询优化（GEQO） 1051
47.4. 进一步阅读 1052
48. 索引访问方法接口定义 1053
48.1. 索引的目录条目 1053
48.2. 索引访问方法函数 1054
48.3. 索引扫描 1056
48.4. 索引锁定注意事项 1057
48.5. 索引唯一性检查 1058
48.6. 索引代价估算函数 1058
49. GiST 索引 1061
49.1. 简介 1061
49.2. 可扩展性 1061
49.3. 实现 1061
49.4. 示例 1062
49.5. 崩溃恢复 1062
50. 数据库物理存储 1064
50.1. 数据库文件布局 1064
50.2. TOAST 1065
50.3. 数据库页面布局 1066
51. BKI 后端接口 1069

51.1. BKI 文件格式 1069
51.2. BKI 命令 1069
51.3. 引导 BKI 文件的结构 1070
51.4. 示例 1070
52. 规划器如何使用统计信息 1072
52.1. 行估计示例 1072
VIII. 附录 1076
A. PostgreSQL错误代码 1086
B. 日期/时间支持 1093
B.1. 日期/时间输入解释 1093
B.2. 日期/时间关键字 1094
B.3. 历法的历史 1108
C. SQL 关键字 1110
D. SQL 符合性 1129
D.1. 支持的特性 1130
D.2. 不支持的特性 1140
E. 发布说明 1147
E.1. 版本 8.1.23 1147
E.2. 版本 8.1.22 1148
E.3. 版本 8.1.21 1150
E.4. 版本 8.1.20 1151
E.5. 版本 8.1.19 1152
E.6. 版本 8.1.18 1153
E.7. 版本 8.1.17 1154
E.8. 版本 8.1.16 1155
E.9. 版本 8.1.15 1156
E.10. 版本 8.1.14 1157
E.11. 版本 8.1.13 1158
E.12. 版本 8.1.12 1159
E.13. 版本 8.1.11 1161
E.14. 版本 8.1.10 1162
E.15. 版本 8.1.9 1163
E.16. 版本 8.1.8 1164
E.17. 版本 8.1.7 1164
E.18. 版本 8.1.6 1165
E.19. 版本 8.1.5 1166
E.20. 版本 8.1.4 1167
E.21. 版本 8.1.3 1169
E.22. 版本 8.1.2 1170
E.23. 版本 8.1.1 1172
E.24. 版本 8.1 1173
E.25. 版本 8.0.26 1186
E.26. 版本 8.0.25 1188
E.27. 版本 8.0.24 1189
E.28. 版本 8.0.23 1190
E.29. 版本 8.0.22 1191
E.30. 版本 8.0.21 1192
E.31. 版本 8.0.20 1193
E.32. 版本 8.0.19 1193
E.33. 版本 8.0.18 1194
E.34. 版本 8.0.17 1195
E.35. 版本 8.0.16 1196
E.36. 版本 8.0.15 1197
E.37. 版本 8.0.14 1199

E.38. 版本 8.0.13 1199
E.39. 版本 8.0.12 1200
E.40. 版本 8.0.11 1200
E.41. 版本 8.0.10 1201
E.42. 版本 8.0.9 1202
E.43. 版本 8.0.8 1203
E.44. 版本 8.0.7 1204
E.45. 版本 8.0.6 1205
E.46. 版本 8.0.5 1206
E.47. 版本 8.0.4 1207
E.48. 版本 8.0.3 1208
E.49. 版本 8.0.2 1210
E.50. 版本 8.0.1 1212
E.51. 版本 8.0 1213
E.52. 版本 7.4.30 1227
E.53. 版本 7.4.29 1228
E.54. 版本 7.4.28 1229
E.55. 版本 7.4.27 1230
E.56. 版本 7.4.26 1231
E.57. 版本 7.4.25 1232
E.58. 版本 7.4.24 1232
E.59. 版本 7.4.23 1233
E.60. 版本 7.4.22 1234
E.61. 版本 7.4.21 1234
E.62. 版本 7.4.20 1235
E.63. 版本 7.4.19 1236
E.64. 版本 7.4.18 1237
E.65. 版本 7.4.17 1237
E.66. 版本 7.4.16 1238
E.67. 版本 7.4.15 1239
E.68. 版本 7.4.14 1239
E.69. 版本 7.4.13 1240
E.70. 版本 7.4.12 1241
E.71. 版本 7.4.11 1242
E.72. 版本 7.4.10 1242
E.73. 版本 7.4.9 1243
E.74. 版本 7.4.8 1244
E.75. 版本 7.4.7 1246
E.76. 版本 7.4.6 1247
E.77. 版本 7.4.5 1248
E.78. 版本 7.4.4 1248
E.79. 版本 7.4.3 1249
E.80. 版本 7.4.2 1250
E.81. 版本 7.4.1 1252
E.82. 版本 7.4 1254
E.83. 版本 7.3.21 1269
E.84. 版本 7.3.20 1270
E.85. 版本 7.3.19 1270
E.86. 版本 7.3.18 1271
E.87. 版本 7.3.17 1271
E.88. 版本 7.3.16 1272
E.89. 版本 7.3.15 1272
E.90. 版本 7.3.14 1273
E.91. 版本 7.3.13 1274

E.92. 版本 7.3.12 1275
E.93. 版本 7.3.11 1275
E.94. 版本 7.3.10 1276
E.95. 版本 7.3.9 1277
E.96. 版本 7.3.8 1278
E.97. 版本 7.3.7 1279
E.98. 版本 7.3.6 1279
E.99. 版本 7.3.5 1280
E.100. 版本 7.3.4 1281
E.101. 版本 7.3.3 1282
E.102. 版本 7.3.2 1284
E.103. 版本 7.3.1 1286
E.104. 版本 7.3 1287
E.105. 版本 7.2.8 1299
E.106. 版本 7.2.7 1300
E.107. 版本 7.2.6 1300
E.108. 版本 7.2.5 1301
E.109. 版本 7.2.4 1302
E.110. 版本 7.2.3 1302
E.111. 版本 7.2.2 1303
E.112. 版本 7.2.1 1304
E.113. 版本 7.2 1304
E.114. 版本 7.1.3 1315
E.115. 版本 7.1.2 1315
E.116. 版本 7.1.1 1316
E.117. 版本 7.1 1316
E.118. 版本 7.0.3 1321
E.119. 版本 7.0.2 1322
E.120. 版本 7.0.1 1322
E.121. 版本 7.0 1323
E.122. 版本 6.5.3 1330
E.123. 版本 6.5.2 1330
E.124. 版本 6.5.1 1331
E.125. 版本 6.5 1332
E.126. 版本 6.4.2 1336
E.127. 版本 6.4.1 1337
E.128. 版本 6.4 1338
E.129. 版本 6.3.2 1342
E.130. 版本 6.3.1 1343
E.131. 版本 6.3 1344
E.132. 版本 6.2.1 1348
E.133. 版本 6.2 1349
E.134. 版本 6.1.1 1352
E.135. 版本 6.1 1352
E.136. 版本 6.0 1355
E.137. 版本 1.09 1357
E.138. 版本 1.02 1357
E.139. 版本 1.01 1359
E.140. 版本 1.0 1361
E.141. Postgres95 版本 0.03 1362
E.142. Postgres95 版本 0.02 1364
E.143. Postgres95 版本 0.01 1364
F. 源代码仓库 1365
F.1. 通过 Git 获取源代码 1365

G. 文档 1366
G.1. DocBook 1366
G.2. 工具集 1366
G.3. 构建文档 1370
G.4. 文档编写 1373
G.5. 风格指南 1374
H. 外部项目 1376
H.1. 外部开发的接口 1376
H.2. 扩展 1377
参考书目 1378
索引 1380

插图清单

47.1. 遗传算法的结构图 1051

表格清单

4.1. 操作符优先级（递减） 30
8.1. 数据类型 82
8.2. 数字类型 83
8.3. 货币类型 86
8.4. 字符类型 86
8.5. 特殊字符类型 88
8.6. 二进制数据类型 88
8.7. bytea输入转义字节 89
8.8. bytea输出转义字节 89
8.9. 日期/时间类型 90
8.10. 日期输入 91
8.11. 时间输入 92
8.12. 时区输入 92
8.13. 特殊日期/时间输入 94
8.14. 日期/时间输出风格 94
8.15. 日期顺序习惯 94
8.16. 几何类型 96
8.17. 网络地址类型 98
8.18. cidr类型输入示例 99
8.19. 对象标识符类型 112
8.20. 伪类型 113
9.1. 比较操作符 114
9.2. 数学操作符 116
9.3. 数学函数 117
9.4. 三角函数 120
9.5. SQL字符串函数和操作符 121
9.6. 其他字符串函数 123
9.7. 内置转换 127
9.8. SQL二进制串函数和操作符 130
9.9. 其他二进制串函数 131
9.10. 位串操作符 132
9.11. 正则表达式匹配操作符 135
9.12. 正则表达式原子 136
9.13. 正则表达式量词 137
9.14. 正则表达式约束 138
9.15. 正则表达式字符项转义 139
9.16. 正则表达式类简写转义 140
9.17. 正则表达式约束转义 140
9.18. 正则表达式反向引用 140
9.19. ARE 嵌入选项字母 141
9.20. 格式化函数 144
9.21. 用于日期/时间格式化的模板模式 144
9.22. 用于日期/时间格式化的模板模式修饰符 146
9.23. 用于数值格式化的模板模式 147
9.24. to_char示例 147
9.25. 日期/时间操作符 149
9.26. 日期/时间函数 149
9.27. AT TIME ZONE 变体 157
9.28. 几何操作符 159
9.29. 几何函数 160
9.30. 几何类型转换函数 161

9.31. cidr 和 inet 操作符 162
9.32. cidr 和 inet 函数 163
9.33. macaddr 函数 164
9.34. 序列函数 164
9.35. array 操作符 169
9.36. array 函数 169
9.37. 聚合函数 170
9.38. 序列生成函数 177
9.39. 会话信息函数 178
9.40. 访问权限查询函数 179
9.41. 模式可见性查询函数 181
9.42. 系统目录信息函数 181
9.43. 注释信息函数 182
9.44. 配置设定函数 183
9.45. 服务器信号函数 183
9.46. 备份控制函数 184
9.47. 数据库对象尺寸函数 184
9.48. 通用文件访问函数 185
12.1. 标准 SQL 事务隔离级别 204
16.1. System V IPC参数 247
16.2. 影响PostgreSQL共享内存用量的配置参数 251
17.1. 短选项对照 284
21.1. PostgreSQL 字符集 307
21.2. 客户端/服务器字符集转换 310
24.1. 标准统计视图 332
24.2. 统计访问函数 333
31.1. information_schema_catalog_name 列 423
31.2. applicable_roles 列 423
31.3. check_constraints 列 423
31.4. column_domain_usage 列 423
31.5. column_privileges 列 424
31.6. column_udt_usage 列 424
31.7. columns 列 425
31.8. constraint_column_usage 列 427
31.9. constraint_table_usage 列 428
31.10. data_type_privileges 列 428
31.11. domain_constraints 列 428
31.12. domain_udt_usage 列 429
31.13. domains 列 429
31.14. element_types 列 431
31.15. enabled_roles 列 433
31.16. key_column_usage 列 433
31.17. parameters 列 433
31.18. referential_constraints 列 435
31.19. role_column_grants 列 435
31.20. role_routine_grants 列 436
31.21. role_table_grants 列 436
31.22. role_usage_grants 列 437
31.23. routine_privileges 列 437
31.24. routines 列 438
31.25. schemata 列 440
31.26. sql_features 列 441

31.27. sql_implementation_info 列 441
31.28. sql_languages 列 442
31.29. sql_packages 列 442
31.30. sql_sizing 列 443
31.31. sql_sizing_profiles 列 443
31.32. table_constraints 列 443
31.33. table_privileges 列 444
31.34. tables 列 444
31.35. triggers 列 445
31.36. usage_privileges 列 446
31.37. view_column_usage 列 446
31.38. view_table_usage 列 447
31.39. views 列 447
32.1. 内置 SQL 类型的等价 C 类型 467
32.2. B-树策略 497
32.3. 哈希策略 497
32.4. R 树策略 497
32.5. B-树支持函数 498
32.6. 哈希支持函数 498
32.7. R 树支持函数 498
32.8. GiST 支持函数 499
42.1. 系统目录 971
42.2. pg_aggregate 列 972
42.3. pg_am 列 972
42.4. pg_amop列 973
42.5. pg_amproc列 974
42.6. pg_attrdef 列 974
42.7. pg_attribute 列 975
42.8. pg_authid 列 976
42.9. pg_auth_members 列 977
42.10. pg_autovacuum列 977
42.11. pg_cast列 978
42.12. pg_class 列 979
42.13. pg_constraint 列 981
42.14. pg_conversion 列 982
42.15. pg_database 列 982
42.16. pg_depend 列 983
42.17. pg_description 列 984
42.18. pg_index 列 984
42.19. pg_inherits 列 985
42.20. pg_language 列 986
42.21. pg_largeobject 列 986
42.22. pg_listener 列 987
42.23. pg_namespace 列 987
42.24. pg_opclass 列 988
42.25. pg_operator 列 988
42.26. pg_pltemplate 列 989
42.27. pg_proc 列 990
42.28. pg_rewrite 列 991
42.29. pg_shdepend 列 992
42.30. pg_statistic 列 993
42.31. pg_tablespace 列 994

42.32. pg_trigger 列 994
42.33. pg_type 列 995
42.34. 系统视图 999
42.35. pg_group 列 999
42.36. pg_indexes 列 1000
42.37. pg_locks 列 1000
42.38. pg_prepared_xacts 列 1002
42.39. pg_roles 列 1002
42.40. pg_rules 列 1003
42.41. pg_settings 列 1003
42.42. pg_shadow 列 1004
42.43. pg_stats 列 1005
42.44. pg_tables 列 1006
42.45. pg_user 列 1006
42.46. pg_views 列 1007
50.1. PGDATA 的内容 1064
50.2. Page Layout 1066
50.3. PageHeaderData Layout 1067
50.4. HeapTupleHeaderData Layout 1068
A.1. PostgreSQL错误代码 1086
B.1. 月份名称 1094
B.2. 星期名称 1094
B.3. 日期/时间字段修饰符 1095
B.4. 输入用的时区缩写 1095
B.5. 输入用的澳大利亚时区缩写 1098
B.6. 用于设置 timezone 的时区名 1098
C.1. SQL 关键字 1110

范例清单

8.1. 使用字符类型 87
8.2. 使用boolean类型 96
8.3. 使用位串类型 100
10.1. 指数操作符类型解析 188
10.2. 字符串连接操作符类型解析 188
10.3. 绝对值与取反操作符类型解析 189
10.4. round 函数参数类型解析 191
10.5. substr 函数类型解析 191
10.6. character 的存储类型转换 192
10.7. 未充分指定类型的 UNION 中的类型解析 193
10.8. 简单 UNION 中的类型解析 193
10.9. 次序对调的 UNION 中的类型解析 193
11.1. 建立一个部分索引以排除常见值 199
11.2. 建立一个部分索引以排除不感兴趣的值 200
11.3. 建立一个部分唯一索引 201
20.1. pg_hba.conf 记录示例 299
20.2. 一个 pg_ident.conf 文件示例 303
28.1. libpq示例程序 1 385
28.2. libpq示例程序 2 388
28.3. libpq Example Program 3 390
29.1. 用 libpq 操作大对象的示例程序 399
35.1. 手工安装 PL/pgSQL 535
36.1. UPDATE/INSERT 与异常 558
36.2. 一个 PL/pgSQL 触发器函数 563
36.3. 一个用于审计的 PL/pgSQL 触发器函数 564
36.4. 一个维护汇总表的 PL/pgSQL 触发器函数 565
36.5. 把一个简单的函数从 PL/SQL 移植到 PL/pgSQL 568
36.6. 从 PL/SQL 移植一个创建另一个函数的函数到 PL/pgSQL 569
36.7. 从 PL/SQL 移植一个带有字符串操作以及 OUT 参数的过程到 PL/pgSQL 571
36.8. 从 PL/SQL 移植一个过程到 PL/pgSQL 572

前言

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

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

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

1. 什么是PostgreSQL?

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

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

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

此外，PostgreSQL还允许用户以多种方式扩展，例如添加新的：

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

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

[1] http://s2k-ftp.CS.Berkeley.EDU:8000/postgres/postgres.html

2. PostgreSQL简史

如今名为PostgreSQL的对象关系数据库管理系统，源自加州大学伯克利分校编写的POSTGRES软件包。经过十多年的发展，PostgreSQL如今已成为当今最先进的开源数据库。

2.1. 伯克利 POSTGRES 项目

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

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

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

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

2.2. Postgres95

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

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

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

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

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

2.3. PostgreSQL

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

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

自那以后 PostgreSQL 发生了哪些变化，可见附录 E。

3. 约定

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

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

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

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

4. 更多信息

除了本文档，也就是本书之外，还有其他关于PostgreSQL的资源：

FAQs

FAQ 列表持续更新地收录了对常见问题的解答。

READMEs

大多数附加包都提供了`README`文件。

网站

PostgreSQL 网站[5]提供最新版本的详细信息，以及其他能让你在工作中或日常使用 PostgreSQL时更加高效的信息。

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

邮件列表

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

你自己！

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

5. 缺陷报告指南

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

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

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

5.1. 识别缺陷

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

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

这里的“程序”指的是任何可执行程序，不只是后端服务器。

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

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

5.2. 报告哪些内容

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

以下各项应包含在每一个缺陷报告中：

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

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

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

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

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

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

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

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

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

- 任何偏离安装说明的做法。

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

  如果你的版本早于 8.1.23，我们几乎肯定会建议你升级。每个新版本都会包含大量缺陷修复和改进，因此你在旧版 PostgreSQL中遇到的缺陷，很可能已经被修复。对于使用旧版 PostgreSQL的站点，我们只能提供有限支持；如果你需要超出这一范围的帮助，请考虑购买商业支持合同。

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

不要担心你的缺陷报告会变得很长；这很正常。第一次就把所有内容都报告出来，总比让我们反复追问细节要好。另一方面，如果你的输入文件非常大，先问问是否有人愿意看一看，也是合理的。这里有一篇文章[6]，概述了更多关于报告缺陷的建议。

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

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

5.3. 到哪里报告缺陷

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

另一种方法是填写项目网站[7]上提供的缺陷报告网页表单。通过这种方式提交的缺陷报告，也会被发送到 `<pgsql-bugs@postgresql.org>`邮件列表。

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

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

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

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

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

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

注意

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

部分 I. 教程

欢迎阅读PostgreSQL教程。下面几章面向初次接触 PostgreSQL、关系数据库概念或 SQL 语言中任一方面的读者，对它们作简明的入门介绍。我们只假定读者具备一些计算机使用的一般常识，不要求具备特定的 Unix 或编程经验。本教程旨在让读者通过动手实践了解PostgreSQL系统的一些重要方面。
它并不试图对所涵盖的主题作全面论述。

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

目录

1. 从头开始 3
1.1. 安装 3
1.2. 架构基础 3
1.3. 创建数据库 3
1.4. 访问数据库 5
2. SQL语言 7
2.1. 简介 7
2.2. 概念 7
2.3. 创建一个新表 7
2.4. 向表中填充行 8
2.5. 查询表 9
2.6. 表之间的连接 10
2.7. 聚合函数 12
2.8. 更新 14
2.9. 删除 14
3. 高级特性 15
3.1. 简介 15
3.2. 视图 15
3.3. 外键 15
3.4. 事务 16
3.5. 继承 18
3.6. 小结 19

第 1 章 从头开始

1.1. 安装

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

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

如果你要自己安装PostgreSQL，请参阅第 14 章中的安装说明，待安装完成后再回到本指南。务必仔细遵循关于设置适当环境变量的小节。

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

1.2. 架构基础

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

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

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

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

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

1.3. 创建数据库

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

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

要创建一个新数据库（本例中命名为 mydb），请使用以下命令：

```
$createdb mydb
```

这条命令应该产生如下响应：

```
CREATE DATABASE
```

如果是这样，这一步就成功了，你可以跳过本节余下的内容。

如果你看到类似下面这样的信息：

```
createdb: command not found
```

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

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

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

另一种响应可能是这样：

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

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

另一种响应可能是这样：

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

其中提到了你自己的登录名。如果管理员没有为你创建 PostgreSQL 用户帐号，就会发生这种情况。（PostgreSQL用户帐号与操作系统用户帐号是分开的。）如果你是管理员，请参阅第 18 章了解如何创建帐号。你需要切换成安装 PostgreSQL 时所用的操作系统用户（通常是 postgres），才能创建第一个用户帐号。也可能分配给你的 PostgreSQL 用户名与操作系统用户名不同；在这种情况下，你需要使用 -U 选项，或者设置 PGUSER 环境变量，来指定你的 PostgreSQL 用户名。

如果你有用户帐号，但它没有创建数据库所需的权限，那么你会看到下面的信息：

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

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

[1] 这之所以可行，是因为PostgreSQL用户名与操作系统用户帐号是分开的。当你连接到数据库时，可以选择以哪个PostgreSQL用户名连接；如果不指定，默认就会使用你当前操作系统帐号的名称。恰好总会有一个PostgreSQL用户帐号，

你也可以创建其他名称的数据库。PostgreSQL允许你在一个给定站点上创建任意数量的数据库。数据库名的首字符必须是字母，并且长度限制为 63 字节。一种方便的做法是创建一个与当前用户名同名的数据库。许多工具都将该数据库名视为默认值，因此这样可以少打一些字。
要创建这个数据库，只需输入：

```
$createdb
```

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

```
$dropdb mydb
```

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

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

1.4. 访问数据库

一旦你创建了数据库，你就可以通过以下方式访问它：

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

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

```
$psql mydb
```

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

在psql中，你将看到下面的欢迎信息：

```
Welcome to psql 8.1.23, the PostgreSQL interactive terminal.

Type:  \copyright for distribution terms
       \h for help with SQL commands
       \? for help with psql commands
       \g or terminate with semicolon to execute query
       \q to quit

mydb=>
```

最后一行也可能是：

```
mydb=#
```

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

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

其名称与启动服务器的操作系统用户相同，而该用户也总是有权创建数据库。除了直接以该用户身份登录之外，你还可以在各处使用 -U选项来选择连接时使用的PostgreSQL用户名。

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

```
mydb=>SELECT version();
                                 version
 ----------------------------------------------------------------------
---
 PostgreSQL 8.1.23 on i586-pc-linux-gnu, compiled by GCC 2.96, 32-bit
(1 row)

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

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

psql程序有一些不属于 SQL 命令的内部命令。它们以反斜线“\”开头。其中一些命令已在欢迎信息中列出。例如，你可以输入下面的命令来查看各种PostgreSQL SQL命令的语法帮助：

```
mydb=>\h
```

要退出psql，输入：

```
mydb=>\q
```

psql将退出并返回到命令 shell。（要了解更多内部命令，可输入\?，在psql提示符下执行。）psql的完整功能见psql。如果PostgreSQL 安装正确，你还可以在操作系统 shell 提示符下输入man psql来查看文档。在本教程中，我们不会明确使用这些功能，但在有帮助时你可以自行使用它们。

第 2 章 SQL语言

2.1. 简介

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

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

本手册中的示例也可以在PostgreSQL源代码发行包的 src/tutorial/目录中找到。要使用这些源文件，请先切换到该目录并运行 make：

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

这会创建脚本并编译包含用户定义函数和类型的 C 文件。（为此你必须使用 GNU make——在你的系统上它可能叫别的名字，通常是gmake。）然后，要开始本教程，请执行以下操作：

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

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

2.2. 概念

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

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

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

2.3. 创建一个新表

你可以通过指定表名以及各列的列名和类型来创建一个新表：

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

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

在 SQL 命令中可以随意使用空白字符（即空格、制表符和换行符）。这意味着你可以按与上面不同的方式对齐该命令，甚至把它全部写在一行上。两个连字符（“--”）表示注释。

从它们开始直到行尾的内容都会被忽略。SQL 对关键字和标识符不区分大小写，除非用双引号括起标识符以保留大小写（上面的例子没有这么做）。

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

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

第二个示例将存储城市及其关联的地理位置：

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

`point`类型就是PostgreSQL特有数据类型的一个例子。

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

```
DROP TABLE tablename;
```

2.4. 向表中填充行

`INSERT`语句用于向表中填充行：

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

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

`point`类型要求一个坐标对作为输入，如下所示：

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

到目前为止所用的语法要求你记住列的顺序。另一种语法允许你显式列出列：

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

如果愿意，你可以按不同顺序列出列，甚至省略某些列，例如降水量未知时：

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

许多开发者认为，与依赖隐含顺序相比，显式列出列是一种更好的风格。

请把上面展示的所有命令都输入一遍，这样你在下面各节中才有一些数据可供操作。

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

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

其中，源文件必须可以在后端服务器机器上访问，而不是客户端机器，因为后端服务器直接读取该文件。有关COPY命令的更多信息，请参见COPY。

2.5. 查询表

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

```
SELECT * FROM weather;
```

这里*是“所有列”的缩写。[1]因此，使用下面的语句也会得到相同结果：

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

输出应该是：

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

你可以在选择列表中编写表达式，而不仅仅是简单的列引用。例如，你可以这样做：

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

这样应该得到：

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

注意这里AS子句如何用于重新命名输出列。（AS子句是可选的。）

查询可以被“限定”，方法是添加一个指定所需行的WHERE子句。WHERE子句包含一个布尔（真值）表达式，只有使该布尔表达式为真的行才会被返回。在限定条件中可以使用常见的布尔操作符（AND、OR和NOT）。例如，下面的查询检索 San Francisco 在雨天的天气记录：

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

结果：

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

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

你可以要求查询结果按排序后的顺序返回：

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

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

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

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

你可以要求从查询结果中去除重复行：

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

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

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

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

2.6. 表之间的连接

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

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

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

下面的查询可以完成这一点：

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

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

注意结果集的两点：

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

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

练习：. 试着确定省略`WHERE`子句时，这条查询的含义。

由于所有列的名称各不相同，解析器自动确定了它们所属的表，
不过在连接查询中完全限定列名是一种良好的风格：

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

目前所见的连接查询，也可以采用下面这种形式编写：

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

这种语法不如上面的语法常用，不过这里展示它是为了帮助理解后续主题。

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

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

     city      | temp_lo | temp_hi | prcp |    date    |     name
 | location
---------------+---------+---------+------+------------+--------------
-+-----------
```

```
 Hayward       |      37 |      54 |       | 1994-11-29 |
  |
 San Francisco |      46 |      50 | 0.25 | 1994-11-27 | San Francisco
 | (-194,53)
 San Francisco |      43 |      57 |    0 | 1994-11-29 | San Francisco
 | (-194,53)
(3 rows)
```

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

练习：. 还有右外连接和全外连接。试着找出它们能做什么。

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

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

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

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

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

你会经常遇到这种缩写方式。

2.7. 聚合函数

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

例如，我们可以用下面的语句找出所有记录中最低温度读数的最高值：

```
SELECT max(temp_lo) FROM weather;

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

如果我们想知道该读数出现在哪个城市（或哪些城市），可能会尝试：

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

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

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

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

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

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

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

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

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

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

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

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

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

❶ LIKE操作符执行模式匹配，详细说明见第 9.7 节。

理解聚合与SQL的WHERE和 HAVING子句之间的相互作用非常重要。WHERE和HAVING的根本区别在于：WHERE在分组和聚合计算之前选择输入行（因此它控制哪些行进入聚合计算），而

HAVING在分组和聚合计算之后选择分组行。因此，WHERE子句中不能包含聚合函数；试图用聚合来决定哪些行应该成为聚合的输入是没有意义的。另一方面，HAVING子句总是包含聚合函数。（严格地说，也允许编写不使用聚合的HAVING子句，但那很少有用。同样的条件在WHERE阶段使用会更高效。）

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

2.8. 更新

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

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

看看数据的新状态：

```
SELECT * FROM weather;

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

2.9. 删除

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

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

所有属于 Hayward 的天气记录都会被删除。

```
SELECT * FROM weather;

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

对如下形式的语句必须保持警惕：

```
DELETE FROM tablename;
```

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

第 3 章 高级特性

3.1. 简介

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

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

3.2. 视图

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

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

SELECT * FROM myview;
```

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

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

3.3. 外键

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

新的表声明如下：

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

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

```
        date      date
);
```

现在试着插入一条无效记录：

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

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

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

3.4. 事务

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.5. 继承

继承是面向对象数据库中的一个概念。它为数据库设计打开了一些有趣的新可能性。

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

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

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

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

就查询而言，这样做还行，但一旦需要更新多行，它就会变得很麻烦。

更好的解决办法是：

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

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

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

例如，下面的查询会找出所有海拔超过500英尺的城市名称，其中也包括州首府：

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

返回：

```
   name    | altitude
-----------+-----------
```

```
 Las Vegas |      2174
 Mariposa  |      1953
 Madison   |       845
(3 rows)
```

另一方面，下面的查询会找出所有海拔500英尺或更高且不是州首府的城市：

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

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

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

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

3.6. 小结

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

如果你觉得还需要更多入门材料，请访问 PostgreSQL 官方网站[1]，其中有指向更多资源的链接。

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

部分 II. SQL 语言

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

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

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

目录

4. SQL语法 25
4.1. 词法结构 25
4.1.1. 标识符和关键词 25
4.1.2. 常量 26
4.1.3. 操作符 29
4.1.4. 特殊字符 29
4.1.5. 注释 30
4.1.6. 词法优先级 30
4.2. 值表达式 31
4.2.1. 列引用 32
4.2.2. 位置参数 32
4.2.3. 下标 32
4.2.4. 字段选择 33
4.2.5. 操作符调用 33
4.2.6. 函数调用 33
4.2.7. 聚合表达式 34
4.2.8. 类型转换 34
4.2.9. 标量子查询 35
4.2.10. 数组构造器 35
4.2.11. 行构造器 36
4.2.12. 表达式计算规则 37
5. 数据定义 39
5.1. 表基础 39
5.2. 默认值 40
5.3. 约束 41
5.3.1. 检查约束 41
5.3.2. 非空约束 42
5.3.3. 唯一约束 43
5.3.4. 主键 44
5.3.5. 外键 45
5.4. 系统列 47
5.5. 修改表 48
5.5.1. 增加列 48
5.5.2. 移除列 49
5.5.3. 增加约束 49
5.5.4. 移除约束 49
5.5.5. 更改列的默认值 50
5.5.6. 修改列的数据类型 50
5.5.7. 重命名列 50
5.5.8. 重命名表 50
5.6. 权限 50
5.7. 模式 51
5.7.1. 创建模式 52
5.7.2. 公共模式 53
5.7.3. 模式搜索路径 53
5.7.4. 模式和权限 54
5.7.5. 系统目录模式 54
5.7.6. 使用方式 55
5.7.7. 可移植性 55
5.8. 继承 55
5.8.1. 注意事项 57

5.9. 分区 58
5.9.1. 概述 58
5.9.2. 实现分区 59
5.9.3. 分区和约束排除 62
5.10. 其他数据库对象 64
5.11. 依赖跟踪 64
6. 数据操纵 66
6.1. 插入数据 66
6.2. 更新数据 67
6.3. 删除数据 67
7. 查询 69
7.1. 概述 69
7.2. 表表达式 69
7.2.1. FROM子句 69
7.2.2. WHERE子句 75
7.2.3. GROUP BY和HAVING子句 76
7.3. 选择列表 78
7.3.1. 选择列表项 78
7.3.2. 列标签 78
7.3.3. DISTINCT 79
7.4. 组合查询 79
7.5. 行排序 80
7.6. LIMIT和OFFSET 81
8. 数据类型 82
8.1. 数字类型 83
8.1.1. 整数类型 83
8.1.2. 任意精度数值 84
8.1.3. 浮点类型 84
8.1.4. serial 类型 85
8.2. 货币类型 86
8.3. 字符类型 86
8.4. 二进制数据类型 88
8.5. 日期/时间类型 89
8.5.1. 日期/时间输入 91
8.5.2. 日期/时间输出 94
8.5.3. 时区 95
8.5.4. 内部实现 95
8.6. 布尔类型 95
8.7. 几何类型 96
8.7.1. 点 97
8.7.2. 线段 97
8.7.3. 方框 97
8.7.4. 路径 97
8.7.5. 多边形 98
8.7.6. 圆 98
8.8. 网络地址类型 98
8.8.1. inet 99
8.8.2. cidr 99
8.8.3. inet 与 cidr 99
8.8.4. macaddr 100
8.9. 位串类型 100
8.10. 数组 101
8.10.1. 数组类型的声明 101

8.10.2. 数组值输入 101
8.10.3. 访问数组 103
8.10.4. 修改数组 104
8.10.5. 在数组中搜索 106
8.10.6. 数组输入和输出语法 107
8.11. 复合类型 108
8.11.1. 复合类型的声明 108
8.11.2. 构造复合值 109
8.11.3. 访问复合类型 110
8.11.4. 修改组合值 110
8.11.5. 复合类型的输入和输出语法 111
8.12. 对象标识符类型 112
8.13. 伪类型 113
9. 函数和操作符 114
9.1. 逻辑操作符 114
9.2. 比较操作符 114
9.3. 数学函数和操作符 116
9.4. 字符串函数和操作符 121
9.5. 二进制串函数和操作符 130
9.6. 位串函数和操作符 132
9.7. 模式匹配 133
9.7.1. LIKE 133
9.7.2. SIMILAR TO 正则表达式 134
9.7.3. POSIX正则表达式 135
9.8. 数据类型格式化函数 143
9.9. 日期/时间函数和操作符 148
9.9.1. EXTRACT, date_part 153
9.9.2. date_trunc 156
9.9.3. AT TIME ZONE 157
9.9.4. 当前日期/时间 157
9.10. 几何函数和操作符 159
9.11. 网络地址函数和操作符 162
9.12. 序列操作函数 164
9.13. 条件表达式 166
9.13.1. CASE 166
9.13.2. COALESCE 168
9.13.3. NULLIF 168
9.13.4. GREATEST and LEAST 168
9.14. 数组函数和操作符 168
9.15. 聚合函数 170
9.16. 子查询表达式 172
9.16.1. EXISTS 172
9.16.2. IN 172
9.16.3. NOT IN 173
9.16.4. ANY/SOME 173
9.16.5. ALL 174
9.16.6. 按行比较 174
9.17. 行和数组比较 175
9.17.1. IN 175
9.17.2. NOT IN 175
9.17.3. ANY/SOME（数组） 176
9.17.4. ALL（数组） 176
9.17.5. 按行比较 176

9.18. 集合返回函数 176
9.19. 系统信息函数 178
9.20. 系统管理函数 183
10. 类型转换 186
10.1. 概述 186
10.2. 操作符 187
10.3. 函数 190
10.4. 值存储 192
10.5. UNION、CASE及相关结构 193
11. 索引 195
11.1. 简介 195
11.2. 索引类型 195
11.3. 多列索引 197
11.4. 组合多个索引 197
11.5. 唯一索引 198
11.6. 表达式索引 198
11.7. 部分索引 199
11.8. 操作符类 201
11.9. 检查索引使用情况 202
12. 并发控制 204
12.1. 介绍 204
12.2. 事务隔离 204
12.2.1. 读已提交隔离级别 205
12.2.2. 可串行化隔离级别 205
12.3. 显式锁定 207
12.3.1. 表级锁 207
12.3.2. 行级锁 209
12.3.3. 死锁 209
12.4. 应用级别的数据一致性检查 210
12.5. 锁定和索引 210
13. 性能提示 212
13.1. 使用EXPLAIN 212
13.2. 规划器使用的统计信息 216
13.3. 用显式JOIN子句控制规划器 218
13.4. 填充一个数据库 219
13.4.1. 禁用自动提交 219
13.4.2. 使用COPY 220
13.4.3. 移除索引 220
13.4.4. 移除外键约束 220
13.4.5. 增加maintenance_work_mem 220
13.4.6. 增加checkpoint_segments 220
13.4.7. 事后运行ANALYZE 220
13.4.8. 关于pg_dump的一些注记 221

第 4 章 SQL语法

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

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

4.1. 词法结构

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

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

此外，注释也可以出现在 SQL 输入中。它们不是词元，实际上等价于空白。

例如，下面是一个（语法上）合法的SQL输入：

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

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

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

4.1.1. 标识符和关键词

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

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

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

标识符和关键词的名称不区分大小写。因此：

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

也可以等价地写成：

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

一种常见约定是关键词使用大写、名称使用小写，例如：

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

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

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

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

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

4.1.2. 常量

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

4.1.2.1. 字符串常量

SQL 中的字符串常量是由单引号（`'`）括起的任意字符序列，例如 `'This is a string'`。在字符串常量中写入单引号字符的符合标准方法，是写两个相邻的单引号，例如 `'Dianne''s horse'`。PostgreSQL 还允许用一个反斜线（`\'`）转义单引号。但是，未来版本的 PostgreSQL 将不允许这样做，因此使用反斜线的应用应转换为上面所述的符合标准的方法。

PostgreSQL的另一个扩展是支持 C 风格的反斜线转义：`\b`是退格、`\f`是换页、`\n`是换行、`\r`是回车、`\t`是制表符。此外还支持`\`*digits*，其中*digits*表示一个八进制字节值，以及`\x`*hexdigits*，其中*hexdigits*表示一个十六进制字节值。
（你有责任确保所创建的字节序列是服务器字符集编码中的有效字符。）
跟随在一个反斜线后面的任何其他字符被当做其字面意思。因此，
要在字符串常量中包含一个反斜线，请写两个反斜线。

> **注意**
>
> 虽然普通字符串目前支持 C 风格的反斜线转义，但未来的版本会对这种用法产生警告，并最终把反斜线当作符合标准的字面字符对待。
> 指定转义处理的正确方法是使用转义字符串语法来表明需要转义处理。
> 转义字符串语法是通过在字符串前面紧接着写一个字母`E`（大写或小写形式）来指定的，例如`E'\041'`。这种方法在PostgreSQL的所有未来版本中都能使用。

编码为零的字符不能出现在字符串常量中。

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

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

等同于：

```
SELECT 'foobar';
```

但是：

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

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

4.1.2.2. 美元引用的字符串常量

指定字符串常量的标准语法通常很方便，但当所需字符串包含很多单引号或反斜线时，
就可能难以理解，因为每个这样的字符都必须写成双份。为了在这种情况下让查询更易读，
PostgreSQL提供了另一种编写字符串常量的方法，称为“美元引用”。
美元引用的字符串常量由一个美元符号（`$`）、一个由零个或多个字符组成的可选“标签”、
另一个美元符号、构成字符串内容的任意字符序列、一个美元符号、
与该美元引用开头相同的标签，以及一个美元符号组成。例如，对于字符串“Dianne's horse”，
下面是使用美元引用指定它的两种不同方法：

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

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

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

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

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

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

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

美元引用不是 SQL 标准的一部分，但在书写复杂字符串字面量时，它往往比符合标准的单引号语法更方便。当要表示的字符串常量位于其他常量内部时，它尤其有用，
这种情况在过程函数定义中很常见。如果使用单引号语法，
上一个示例中的每个反斜线都必须写成四个反斜线；

在解析原始字符串常量时它们会缩减为两个反斜线，
而在函数执行期间重新解析内层字符串常量时又会变成一个。

4.1.2.3. 位串常量

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

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

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

4.1.2.4. 数字常量

数字常量接受下列一般形式：

digitsdigits.[*digits*][e[+-]*digits*]
[*digits*].*digits*[e[+-]*digits*]
*digits*e[+-]*digits*

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

这些是合法数字常量的示例：

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

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

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

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

这些实际上只是下面将要讨论的一般类型转换记法的特例。

4.1.2.5. 其他类型的常量

任意类型的常量都可以用下列记法之一输入：

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

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

字符串常量可以使用常规 SQL 记法或美元引用书写。

也可以用类似函数调用的语法来指定类型强制转换：

typename ('*string*')

但并非所有类型名都可以这样使用，详见第 4.2.8 节。

如第 4.2.8 节中所述，`::`、`CAST()` 以及函数调用语法也可以用来指定任意表达式的运行时类型转换。为避免语法歧义，*type* '*string*' 语法只能用来指定简单字面常量的类型。*type* '*string*' 语法的另一个限制是不能用于数组类型；要指定数组常量的类型，请使用 `::` 或 `CAST()`。

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

4.1.3. 操作符

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

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

不过，操作符名称也有一些限制：

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

  ~ ! @ # % ^ & | ` ?

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

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

4.1.4. 特殊字符

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

- 美元符号（`$`）后跟数字时，用来表示函数定义体或预备语句中的位置参数。在其他上下文中，美元符号可以是标识符的一部分，也可以是美元引用字符串常量的一部分。
- 圆括号（`()`）具有它们通常的含义，用来对表达式分组并确定运算优先级。在某些情况中，圆括号被要求作为一个特定 SQL 命令的固定语法的一部分。
- 方括号（`[]`）被用来选择一个数组中的元素。更多关于数组的信息见第 8.10 节。
- 逗号（`,`）被用在某些语法结构中来分割一个列表的元素。

- 分号（`;`）结束一个 SQL 命令。它不能出现在一个命令中间的任何位置，除了在一个字符串常量中或者一个带引号的标识符中。
- 冒号（`:`）被用来从数组中选择“切片”（见第 8.10 节）。在某些 SQL 的“方言”（例如嵌入式 SQL）中，冒号被用来作为变量名的前缀。
- 星号（`*`）在某些上下文中用来表示表行或复合值的所有字段。当它被用作`COUNT`聚合函数的参数时，还有一种特殊含义。
- 句点（`.`）被用在数字常量中，并且被用来分割模式、表和列名。

4.1.5. 注释

注释是一个以双连字符开始并延伸到行尾的任意字符序列，例如：

```
-- 这是一条标准 SQL 注释
```

另外，也可以使用 C 风格注释块：

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

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

在进一步进行语法分析之前，注释会从输入流中移除，并实际被替换为空白。

4.1.6. 词法优先级

表 4.1展示了 PostgreSQL 中操作符的优先级和结合性。大多数操作符具有相同的优先级，并且是左结合的。操作符的优先级和结合性是固定写在解析器中的。这可能导致不直观的行为；例如，布尔操作符 `<` 和 `>` 与布尔操作符 `<=` 和 `>=` 的优先级不同。此外，在组合使用二元和一元操作符时，有时需要添加圆括号。例如：

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

将被解析为：

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

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

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

这是为可扩展性付出的代价。

表 4.1. 操作符优先级（递减）

操作符/元素	结合性	描述
`.`	左	表/列名分隔符
`::`	左	PostgreSQL-风格的类型转换
`[ ]`	左	数组元素选择

操作符/元素	结合性	描述
-	右	一元负号
^	左	求幂
* / %	左	乘、除、取模
+ -	左	加、减
IS		IS TRUE、IS FALSE、IS UNKNOWN、IS NULL
ISNULL		空值测试
NOTNULL		非空测试
（任意其他）	左	所有其他内建和用户定义的操作符
IN		集合成员关系
BETWEEN		范围包含
OVERLAPS		时间区间重叠
LIKE ILIKE SIMILAR		字符串模式匹配
< >		小于、大于
=	右	相等、赋值
NOT	右	逻辑否定
AND	左	逻辑合取
OR	左	逻辑析取

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

当一个模式限定的操作符名被用在OPERATOR语法中时，如下面的示例：

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

OPERATOR 结构在优先级上会被视为表 4.1中为“任意其他”操作符所示的优先级。无论OPERATOR()中指定的是哪个具体操作符，这一点都成立。

4.2. 值表达式

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

一个值表达式是下列之一：

- 一个常量或字面值。
- 一个列引用。
- 在一个函数定义体或预备语句中的一个位置参数引用。
- 一个下标表达式。
- 一个字段选择表达式。

- 一个操作符调用。
- 一个函数调用。
- 一个聚合表达式。
- 一个类型转换。
- 一个标量子查询。
- 一个数组构造器。
- 一个行构造器。
- 另一个放在圆括号中的值表达式，用于对子表达式分组并覆盖优先级。

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

我们已经在第 4.1.2 节中讨论过常量。下面的小节会讨论剩下的选项。

4.2.1. 列引用

一个列可以以下面的形式被引用：

correlation.columnname

*correlation*是一个表的名字（可能带有模式名限定），或者是在`FROM`子句中为某个表定义的别名，或者是关键词`NEW`或`OLD`之一。（`NEW`和`OLD`只能出现在重写规则中，而其他关联名称可以在任何 SQL 语句中使用。）如果该列名在当前查询所用的所有表中都是唯一的，则关联名称和分隔用的句点都可以省略（另见第 7 章）。

4.2.2. 位置参数

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

$*number*

例如，考虑一个函数`dept`的定义：

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

这里`$1`引用函数被调用时第一个函数参数的值。

4.2.3. 下标

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

expression[*subscript*]

或者抽取出多个相邻元素（一个“数组切片”）：

expression[*lower_subscript*:*upper_subscript*]

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

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

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

最后一个示例中的圆括号是必需的。详见第 8.10 节。

4.2.4. 字段选择

如果一个表达式得到一个复合类型（行类型）的值，那么可以抽取该行的指定字段：

expression.*fieldname*

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

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

（因此，限定列引用实际上只是字段选择语法的一种特例。）

4.2.5. 操作符调用

对于一次操作符调用，有三种可能的语法：

expressionoperatorexpression（二元中缀操作符）
operatorexpression（一元前缀操作符）
expressionoperator（一元后缀操作符）

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

OPERATOR(*schema*.*operatorname*)

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

4.2.6. 函数调用

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

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

例如，下面会计算 2 的平方根：

```
sqrt(2)
```

内置函数的列表在第 9 章中。其他函数可以由用户增加。

4.2.7. 聚合表达式

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

```
aggregate_name (expression)
aggregate_name (ALL expression)
aggregate_name (DISTINCT expression)
aggregate_name ( * )
```

其中，*aggregate_name* 是先前定义的聚合（可以带模式名限定），*expression* 是任何本身不包含聚合表达式的值表达式。

第一种形式的聚合表达式在给定表达式得出非空值的所有输入行上调用聚合。（实际上，是否忽略空值取决于聚合函数本身 – 但所有标准聚合都会忽略。）第二种形式和第一种相同，因为`ALL`是默认选项。第三种形式为输入行中表达式的所有非空可区分值调用聚合。
最后一种形式为每一个输入行调用一次聚合，无论值是否为空；因为没有特定的输入值被指定，它通常只对于`count()`聚合函数有用。

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

预定义的聚合函数在第 9.15 节中描述。其他聚合函数可以由用户增加。

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

当一个聚合表达式出现在一个子查询中（见第 4.2.9 节和第 9.16 节），聚合通常在该子查询的行上被计算。
但是如果该聚合的参数只包含外层变量则会产生一个异常：该聚合则属于最近的那个外层，并且会在那个查询的行上被计算。
该聚合表达式从整体上则是对其所出现于的子查询的一种外层引用，
并且在那个子查询的任意一次计算中都作为一个常量。只出现在结果列表或`HAVING`子句的限制适用于该聚合所属的查询层次。

4.2.8. 类型转换

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

```
CAST ( expression AS type )
expression::type
```

`CAST`语法遵从 SQL，而用`::`的语法是PostgreSQL的历史用法。

当类型转换被应用于一种已知类型的值表达式时，它表示一次运行时类型转换。
只有在已经定义了合适的类型转换操作时，该类型转换才会成功。注意，
这与常量上的类型转换（如第 4.1.2.5 节中所示）略有不同。应用于未修饰字符串字面量的类型转换，表示为一个字面常量值赋予初始类型，
因此它对任意类型都能成功（前提是该字符串字面量的内容符合该数据类型的输入语法）。

如果一个值表达式必须产生的类型没有歧义（例如当它被赋给一个表列时），
通常可以省略显式类型转换，在这种情况下系统会自动应用类型转换。但是，
自动类型转换只会对那些在系统目录中被标记为“可以隐式应用”的类型转换执行。
其他类型转换必须使用显式类型转换语法来调用。
施加这一限制的目的是防止出人意料的转换被悄无声息地应用。

还可以用类似函数调用的语法来指定类型转换：

typename (*expression*)

不过，这只对那些名称本身也可作为函数名使用的类型有效。例如，double precision 不能用这种方式，但等效的 float8 可以。另外，由于语法冲突，名称 interval、time 和 timestamp 只有在使用双引号引用时才能采用这种写法。因此，函数风格的类型转换语法会带来不一致性，在新应用中很可能应该避免使用。（函数风格的语法事实上只是一次函数调用。
当两种标准类型转换语法之一被用来做一次运行时转换时，
它会在内部调用一个已注册的函数来执行该转换。按照惯例，
这些转换函数与它们的输出类型同名，因此“函数风格的语法”
无非是对底层转换函数的一次直接调用。显然，可移植的应用不应当依赖于这一点。）

4.2.9. 标量子查询

标量子查询是一个放在圆括号中的普通 SELECT 查询，它恰好返回一行一列（关于如何编写查询，见第 7 章）。该 SELECT 查询会被执行，其返回的单个值会用在外围值表达式中。把一个返回多于一行或多于一列的查询当作标量子查询来用是错误的。
（但如果在某次具体执行中该子查询没有返回任何行，则不算错误；
其标量结果会被视为空值。）该子查询可以引用外围查询中的变量，
这些变量在该子查询的每一次求值期间都会作为常量。有关其他涉及子查询的表达式，另见第 9.16 节。

例如，下列语句会寻找每个州中最大的城市人口：

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

4.2.10. 数组构造器

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

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

数组元素类型是成员表达式的公共类型，其确定规则与 UNION 或 CASE 结构相同（见第 10.5 节）。

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

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

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

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

由于多维数组必须是长方形的，因此同一级别的内部构造器必须生成维度完全相同的子数组。

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

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

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

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

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

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

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

用 ARRAY 构造出来的数组值，其下标总是从 1 开始。关于数组的更多信息，见第 8.10 节。

4.2.11. 行构造器

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

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

当列表中有多个表达式时，关键字 ROW 是可选的。

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

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

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

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

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

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

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

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

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

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

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

SELECT ROW(a, b, c) IS NOT NULL FROM table;
```

更多详情见第 9.17 节。行构造器也可以与子查询结合使用，详见第 9.16 节。

4.2.12. 表达式计算规则

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

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

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

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

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

注意这和一些编程语言中布尔操作符从左至右的“短路”不同。

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

当有必要强制计算顺序时，可以使用一个CASE结构（见第 9.13 节）。例如，在一个WHERE子句中使用下面的方法尝试避免除零是不可靠的：

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

但是这是安全的：

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

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

第 5 章 数据定义

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

5.1. 表基础

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

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

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

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

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

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

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

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

（`numeric`类型可以存储小数部分，这对金额很常见。）

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

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

如果不再需要某个表，可以使用`DROP TABLE`命令删除它。例如：

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

尝试删除不存在的表会报错。不过，在 SQL 脚本文件中，人们常常在创建每个表之前无条件地尝试删除它，并忽略可能出现的错误消息。

如果我们需要修改一个已经存在的表，请参考本章稍后的第 5.5 节。

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

5.2. 默认值

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

如果没有显式指定默认值，则默认值是空值。这是合理的，因为空值表示未知数据。

在表定义中，默认值写在列的数据类型之后。例如：

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

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

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

其中`nextval()`函数从序列对象（见第 9.12 节）中依次取值。这种写法非常常见，因此还有一种专门的简写形式：

```
CREATE TABLE products (
```

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

SERIAL简写会在第 8.1.4 节进一步讨论。

5.3. 约束

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

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

5.3.1. 检查约束

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

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

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

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

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

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

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

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

```
    CHECK (price > discounted_price)
);
```

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

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

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

甚至是：

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

这只是口味的问题。

表约束也可以用与列约束相同的方法来指定名称：

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

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

检查约束对于提升分区表的性能也很有用。详情见第 5.9 节。

5.3.2. 非空约束

非空约束只是指定某列不能取空值。语法示例如下：

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

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

当然，一列可以有多个约束。只需将约束逐个写出：

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

顺序并不重要，也不一定决定检查约束的顺序。

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

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

然后在需要的地方插入`NOT`关键字。

> **提示**
>
> 在大多数数据库设计中，多数列都应标记为非空。

5.3.3. 唯一约束

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

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

写成表约束时则是：

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

如果唯一约束引用一组列，把这些列用逗号分隔列出：

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

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

我们可以按通常的方式为唯一约束指定名称：

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

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

5.3.4. 主键

从技术上讲，主键约束只是唯一约束和非空约束的组合。因此，
下面的两个表定义接受相同的数据：

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

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

主键也可以包含多于一个列，其语法和唯一约束相似：

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

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

主键表示一个列或一组列可以用作表中行的唯一标识符。（这是主键定义的直接结果。注意，唯一约束本身并不能提供唯一标识符，因为它不排除空值。）
这对于文档目的和客户端应用都很有用。例如，一个允许修改行值的 GUI 应用可能需要知道一个表的主键，以便能唯一地标识行。

一个表最多只能有一个主键（不过它可以有多个唯一和非空约束）。
关系数据库理论要求每一个表都要有一个主键。PostgreSQL 并未强制要求这一点，但通常最好遵循它。

5.3.5. 外键

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

例如我们有一个使用过多次的产品表：

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

再假设有一个表存储这些产品的订单。我们希望确保订单表只包含实际存在的产品的订单。
因此，我们在订单表中定义一个引用产品表的外键约束：

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

现在就不可能创建包含不存在于产品表中的`product_no`值的订单。

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

我们也可以把上述命令简写为：

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

因为如果缺少列的列表，则被引用表的主键将被用作被引用列。

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

```
CREATE TABLE t1 (
  a integer PRIMARY KEY,
  b integer,
```

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

当然，被约束列的数量和类型应该匹配被引用列的数量和类型。

你可以按常规方式为外键约束指定自己的名称。

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

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

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

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

注意，最后一个表中的主键与外键有重叠。

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

- 不允许删除被引用的产品
- 同时删除订单
- 其他处理方式？

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

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

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

CREATE TABLE order_items (
```

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

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

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

更多关于更新和删除数据的信息请见第 6 章。

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

5.4. 系统列

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

oid

行的对象标识符（对象 ID）。只有使用WITH OIDS创建表，或设置了default_with_oids配置变量，才存在此列。该列的类型是oid（与列同名）；关于该类型的更多信息，参见第 8.12 节。

tableoid

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

xmin

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

cmin

插入事务中的命令标识符（从0开始）。

xmax

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

`cmax`

删除事务中的命令标识符，或者为0。

`ctid`

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

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

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

事务 ID也是 32 位的量。在一个长期运行的数据库中，事务 ID 可能会回卷。只要采取适当的维护措施，这并不是致命问题，详见第 22 章。不过，从长期来看（超过十亿个事务）依赖事务 ID 的唯一性是不明智的。

命令标识符也是 32 位的量。这为单个事务中的SQL命令数设置了一个硬上限：2^{32}（40 亿）。在实践中，这个限制并不是问题——注意，这里限制的是SQL命令的数量，而不是处理的行数。

5.5. 修改表

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

利用这些命令，我们可以：

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

所有这些动作都由ALTER TABLE命令执行。

5.5.1. 增加列

要添加一列，可以使用这样的命令：

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

新列最初会填入给定的默认值（如果没有指定DEFAULT子句，则填入空值）。

也可以同时为该列定义约束，使用常规语法：

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

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

5.5.2. 移除列

要删除一列，使用如下命令：

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

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

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

关于这个操作背后的一般性机制请见第 5.11 节。

5.5.3. 增加约束

添加约束时使用表约束语法。例如：

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

非空约束不能写成表约束，添加时应使用以下语法：

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

系统会立即检查约束，因此只有表中的数据满足约束，才能将其添加。

5.5.4. 移除约束

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

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

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

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

除非空约束之外，所有类型的约束都可以用相同方式删除。要删除非空约束，使用：

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

（请记住，非空约束没有名称。）

5.5.5. 更改列的默认值

要为一个列设置一个新默认值，使用命令：

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

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

要移除任何默认值，使用：

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

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

5.5.6. 修改列的数据类型

为了将一个列转换为一种不同的数据类型，使用如下命令：

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

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

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

5.5.7. 重命名列

要重命名一个列：

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

5.5.8. 重命名表

要重命名一个表：

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

5.6. 权限

创建数据库对象时，你就成为它的所有者。默认情况下，
只有对象的所有者才能对该对象做任何操作。为了让其他用户能使用它，必须授予权限。
（不过，具有超级用户属性的用户总是可以访问任何对象。）

有不同种类的权限：`SELECT`、`INSERT`、`UPDATE`、`DELETE`、`RULE`、`REFERENCES`、`TRIGGER`、`CREATE`、`TEMPORARY`、`EXECUTE`和`USAGE`。
适用于特定对象的权限取决于对象的类型（表、函数等）。关于

PostgreSQL支持的不同权限类型的完整信息，请参阅GRANT参考页。
后续的各节和各章还将展示这些权限是如何使用的。

修改或销毁对象的权利始终只属于该对象的拥有者。

> **注意**
>
> 要更改表、索引、序列或视图的所有者，可使用ALTER TABLE命令。针对其他对象类型也有相应的ALTER命令。

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

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

要把权限授予一个组，可以使用这样的语法：

```
GRANT SELECT ON accounts TO GROUP staff;
```

特殊的“用户”名PUBLIC可用于向系统中的每个用户授予权限。用ALL代替某个具体权限，
会授予与该对象类型相关的全部权限。

要撤销权限，使用名称贴切的REVOKE命令：

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

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

通常，只有对象的拥有者（或超级用户）可以授予或撤销对象上的权限。不过，
也可以在授予权限时“附带授权选项”，使接收者有权进一步将该权限授予其他人。
如果之后撤销了授权选项，所有从该接收者直接或经过授权链取得此权限的用户都会失去权限。
详见GRANT和REVOKE参考页。

5.7. 模式

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

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

一个数据库包含一个或多个命名的模式，模式中又包含表。模式还包含其他类型的命名对象，
包括数据类型、函数和操作符。在同一个模式中，同一类型的两个对象不能有相同的名称。
此外，表、序列、索引、视图、物化视图和外部表共享同一个名字空间，

因此例如当它们位于同一模式中时，索引和表就必须具有不同的名称。
相同的对象名可以在不同模式中重复使用而不发生冲突；例如，`schema1`和`myschema`
都可以包含名为`mytable`的表。与数据库不同，模式并不是被严格隔离的：只要拥有相应权限，
用户就可以访问其所连接数据库中任意模式里的对象。

下面是一些使用模式的原因：

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

模式类似于操作系统层的目录，但是模式不能嵌套。

5.7.1. 创建模式

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

```
CREATE SCHEMA myschema;
```

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

schema.table

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

实际上，也可以使用更通用的语法

database.schema.table

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

因此，如果要在一个新模式中创建一个表，可用：

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

要删除一个为空的模式（其中的所有对象已经被删除），可用：

```
DROP SCHEMA myschema;
```

要删除一个模式以及其中包含的所有对象，可用：

```
DROP SCHEMA myschema CASCADE;
```

有关于此的更一般的机制请参见第 5.11 节。

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

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

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

以pg_开头的模式名被保留用于系统目的，所以不能被用户所创建。

5.7.2. 公共模式

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

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

以及：

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

5.7.3. 模式搜索路径

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

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

要显示当前搜索路径，可使用下面的命令：

```
SHOW search_path;
```

在默认设置下这将返回：

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

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

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

要把新模式放在搜索路径中，我们可以使用：

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

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

```
DROP TABLE mytable;
```

同样，由于myschema是路径中的第一个元素，新对象默认也会创建在其中。

我们也可以这样写：

```
SET search_path TO myschema;
```

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

其他操作模式搜索路径的方法请见第 9.19 节。

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

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

这是为了避免句法歧义。例如：

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

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

5.7.4. 模式和权限

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

也可以允许用户在他人的模式中创建对象。为此，需要授予该模式上的CREATE权限。注意，
默认情况下，每个人都具有模式public上的CREATE和USAGE权限。
这使所有能够连接到某个数据库的用户都可以在它的public模式中创建对象。
如果不希望允许这样做，可以撤销这一权限：

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

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

5.7.5. 系统目录模式

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

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

5.7.6. 使用方式

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

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

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

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

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

5.7.7. 可移植性

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

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

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

5.8. 继承

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

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

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

CREATE TABLE capitals (
```

```
    state           char(2)
) INHERITS (cities);
```

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

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

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

对于来自PostgreSQL教程（见第 2.1 节）的示例数据，它将返回：

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

另一方面，下面的查询将找到所有高度超过 500 尺且不是州首府的城市：

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

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

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

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

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

将会返回：

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

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

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

将会返回：

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

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

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

我们也许会希望数据能以某种方式被路由到capitals表中，但这不会发生：INSERT 总是向指定的表中插入。在某些情况下，可以通过使用一个规则（见第 34 章）将插入动作重定向。但是这对上面的情况并没有帮助，因为cities表根本就不包含state列，因而这个命令会在触发规则之前就被拒绝。

可以在继承层次中的表上定义检查约束。父表上的所有检查约束都将自动被它的所有子表继承。但其他类型的约束不会被继承。

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

表继承目前只能使用CREATE TABLE语句来定义。相关的CREATE TABLE AS语句不允许指定继承。目前没有办法通过添加继承链接把一个现有表变为子表。类似地，一旦定义了继承链接，除了完全删除该表之外，也没有办法从子表中移除继承链接。只要还有子表存在，父表就不能被删除。如果希望移除一个表和它的所有后代，一种简单的方法是使用CASCADE 选项删除父表。

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

5.8.1. 注意事项

表访问权限不会被自动继承。因此，试图访问父表的用户要么必须对所有子表也拥有执行相同操作的权限，要么必须使用ONLY记法。向现有的继承层次中添加新的子表时，请注意为其授予所有需要的权限。

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

- 如果我们把cities.name声明为UNIQUE或PRIMARY KEY，这并不能阻止capitals表中出现与cities中城市同名的行。而且这些重复行默认还会出现在针对cities的查询结果中。

事实上，默认情况下capitals根本没有唯一约束，因此它可以包含多行同名记录。
你当然可以给capitals添加唯一约束，但这仍无法阻止相对于cities的重复。

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

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

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

已弃用

在PostgreSQL之前的版本中，默认行为是不在查询中包含子表。
这种做法被发现容易出错，而且也违反了 SQL 标准。在旧语法中，要在查询中包含子表，
需在表名后附加*。例如：

```
SELECT * from cities*;
```

你仍然可以通过附加*来显式指定扫描子表，也可以通过写ONLY来显式指定不扫描子表。
但从 7.1 版开始，不带修饰的表名的默认行为是同时扫描其子表，而在此之前默认行为并非如此。要获得旧的默认行为，可以禁用sql_inheritance配置选项。

5.9. 分区

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

5.9.1. 概述

分区是指将逻辑上的一个大表拆分成较小的物理部分。分区可以带来以下好处：

- 对于某些类型的查询，查询性能可以得到显著提升。

- 更新性能也可以得到提升，因为表的每个部分的索引都比针对整个数据集的索引要小。
当索引不再容易容纳在内存中时，对索引的读和写操作都会需要越来越多的磁盘访问。

- 如果在分区设计中考虑到了这种需求，就可以通过简单地移除分区来完成批量删除。DROP TABLE远快于批量DELETE，更不用说随之而来的VACUUM开销。

- 很少使用的数据可以被迁移到便宜且较慢的存储介质上。

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

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

在PostgreSQL中可以实现以下几种分区形式：

范围分区

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

列表分区

通过明确列出每个分区包含哪些键值来对表进行分区。

目前不支持哈希分区。

5.9.2. 实现分区

要设置一个分区表，可以按以下步骤操作：

1. 创建所有分区都将继承的“主”表。

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

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

   典型示例如下：

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

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

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

   这是错误的，因为无法明确键值 200 属于哪个分区。

   注意，范围分区和列表分区在语法上没有区别；这些术语只是描述性的。
4. 对于每个分区，在键列上创建索引，以及其他所需的索引。（键索引不是严格必需的，
   但在大多数场景中是有帮助的。如果希望键值唯一，
   则应始终为每个分区创建唯一或主键约束。）
5. 可以选择定义一个规则或触发器，把对主表的修改重定向到适当的分区。
6. 确认constraint_exclusion配置参数在`postgresql.conf`中已启用。如果没有启用，
   查询将无法按预期得到优化。

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

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

```
);
```

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

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

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

   ```
   CREATE TABLE measurement_yy04mm02 ( ) INHERITS (measurement);
   CREATE TABLE measurement_yy04mm03 ( ) INHERITS (measurement);
   ...
   CREATE TABLE measurement_yy05mm11 ( ) INHERITS (measurement);
   CREATE TABLE measurement_yy05mm12 ( ) INHERITS (measurement);
   CREATE TABLE measurement_yy06mm01 ( ) INHERITS (measurement);
   ```

   每个分区本身都是完整的表，但它们从measurement表继承其定义。

   这解决了我们的一个问题：删除旧数据。每个月，我们只需对最旧的子表执行DROP TABLE，并为新月份的数据创建一个新的子表。
3. 我们必须提供互不重叠的表约束。与其像上面那样只创建分区表，表创建脚本其实应该是：

   ```
    CREATE TABLE measurement_yy04mm02 (
        CHECK ( logdate >= DATE '2004-02-01' AND logdate < DATE '2004
   -03-01' )
    ) INHERITS (measurement);
    CREATE TABLE measurement_yy04mm03 (
        CHECK ( logdate >= DATE '2004-03-01' AND logdate < DATE '2004
   -04-01' )
    ) INHERITS (measurement);
    ...
    CREATE TABLE measurement_yy05mm11 (
        CHECK ( logdate >= DATE '2005-11-01' AND logdate < DATE '2005
   -12-01' )
    ) INHERITS (measurement);
    CREATE TABLE measurement_yy05mm12 (
        CHECK ( logdate >= DATE '2005-12-01' AND logdate < DATE '2006
   -01-01' )
    ) INHERITS (measurement);
    CREATE TABLE measurement_yy06mm01 (
        CHECK ( logdate >= DATE '2006-01-01' AND logdate < DATE '2006
   -02-01' )
    ) INHERITS (measurement);
   ```
4. 我们可能还需要在键列上创建索引：

   ```
    CREATE INDEX measurement_yy04mm02_logdate ON measurement_yy04mm02
    (logdate);
    CREATE INDEX measurement_yy04mm03_logdate ON measurement_yy04mm03
    (logdate);
   ...
    CREATE INDEX measurement_yy05mm11_logdate ON measurement_yy05mm11
    (logdate);
   ```

```
CREATE INDEX measurement_yy05mm12_logdate ON measurement_yy05mm12
(logdate);
CREATE INDEX measurement_yy06mm01_logdate ON measurement_yy06mm01
(logdate);
```

我们此时选择不添加更多索引。

5. 如果数据只会添加到最新的分区，我们可以设置一条非常简单的插入规则。
   我们必须每个月重新定义它，使其始终指向当前分区。

```
CREATE OR REPLACE RULE measurement_current_partition AS
ON INSERT TO measurement
DO INSTEAD
    INSERT INTO measurement_yy06mm01 VALUES ( NEW.city_id,
                                              NEW.logdate,
                                              NEW.peaktemp,
                                              NEW.unitsales );
```

我们也可能希望插入数据时由服务器自动定位该行应添加到的分区。
这可以通过下面这样一组更复杂的规则来完成。

```
CREATE RULE measurement_insert_y2004m02 AS
ON INSERT TO measurement WHERE
    ( logdate >= DATE '2004-02-01' AND logdate < DATE '2004-03
-01' )
DO INSTEAD
    INSERT INTO measurement_yy04mm02 VALUES ( NEW.city_id,
                                              NEW.logdate,
                                              NEW.peaktemp,
                                              NEW.unitsales );
...
CREATE RULE measurement_insert_y2005m12 AS
ON INSERT TO measurement WHERE
    ( logdate >= DATE '2005-12-01' AND logdate < DATE '2006-01
-01' )
DO INSTEAD
    INSERT INTO measurement_yy05mm12 VALUES ( NEW.city_id,
                                              NEW.logdate,
                                              NEW.peaktemp,
                                              NEW.unitsales );
CREATE RULE measurement_insert_yy06mm01 AS
ON INSERT TO measurement WHERE
    ( logdate >= DATE '2006-01-01' AND logdate < DATE '2006-02
-01' )
DO INSTEAD
    INSERT INTO measurement_yy06mm01 VALUES ( NEW.city_id,
                                              NEW.logdate,
                                              NEW.peaktemp,
                                              NEW.unitsales );
```

注意，每条规则的`WHERE`子句都与其分区的`CHECK`约束完全匹配。

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

以下注意事项适用：

- 目前没有办法验证所有CHECK约束是否互斥。这需要数据库设计者自行小心。
- 目前还没有简单的方法指定不许把行插入到主表中。主表上的CHECK (false)约束会被所有子表继承，因此不能用于这一目的。一种可能的做法是在主表上设置一个总是引发错误的ON INSERT触发器。（或者，也可以用这样的触发器把数据重定向到适当的子表，而不是像上面建议的那样使用一组规则。）

也可以使用UNION ALL视图来安排分区：

```
CREATE VIEW measurement AS
          SELECT * FROM measurement_yy04mm02
UNION ALL SELECT * FROM measurement_yy04mm03
...
UNION ALL SELECT * FROM measurement_yy05mm11
UNION ALL SELECT * FROM measurement_yy05mm12
UNION ALL SELECT * FROM measurement_y2004m01;
```

不过，以这种方式定义的分区表目前不支持约束排除。此外，重新创建视图的需要给数据集分区的添加和删除增加了一个额外步骤。

5.9.3. 分区和约束排除

约束排除是一种查询优化技术，可提高按上述方式定义的分区表的性能。例如：

```
SET constraint_exclusion = on;
SELECT count(*) FROM measurement WHERE logdate >= DATE '2006-01-01';
```

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

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

```
SET constraint_exclusion = off;
EXPLAIN SELECT count(*) FROM measurement WHERE logdate >= DATE '2006
-01-01';

                                           QUERY PLAN
-----------------------------------------------------------------------
-------------------------
 Aggregate  (cost=158.66..158.68 rows=1 width=0)
   ->  Append  (cost=0.00..151.88 rows=2715 width=0)
         ->  Seq Scan on measurement  (cost=0.00..30.38 rows=543
 width=0)
               Filter: (logdate >= '2006-01-01'::date)
         ->  Seq Scan on measurement_yy04mm02 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2006-01-01'::date)
         ->  Seq Scan on measurement_yy04mm03 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2006-01-01'::date)
...
         ->  Seq Scan on measurement_yy05mm12 measurement  (cost=0.00.
.30.38 rows=543 width=0)
```

```
               Filter: (logdate >= '2006-01-01'::date)
         ->  Seq Scan on measurement_yy06mm01 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2006-01-01'::date)
```

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

```
SET constraint_exclusion = on;
EXPLAIN SELECT count(*) FROM measurement WHERE logdate >= DATE '2006
-01-01';
                                          QUERY PLAN
-----------------------------------------------------------------------
-------------------------
 Aggregate  (cost=63.47..63.48 rows=1 width=0)
   ->  Append  (cost=0.00..60.75 rows=1086 width=0)
         ->  Seq Scan on measurement  (cost=0.00..30.38 rows=543
 width=0)
               Filter: (logdate >= '2006-01-01'::date)
         ->  Seq Scan on measurement_yy06mm01 measurement  (cost=0.00.
.30.38 rows=543 width=0)
               Filter: (logdate >= '2006-01-01'::date)
```

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

以下注意事项适用：

- 只有查询的WHERE子句包含常量时，约束排除才能有效果。参数化查询不会被优化，因为规划器无法知道参数值在运行时会选择哪些分区。出于同样的原因，必须避免CURRENT_DATE这类“稳定”函数。把分区键与另一个表的列进行连接也不会被优化。

- 避免在CHECK约束中使用跨数据类型的比较，因为规划器目前无法证明这类条件为假。例如，下面的约束在x是integer列时可行，但在x是bigint列时不行：

  ```
  CHECK ( x = 1 )
  ```

  对于bigint列，必须使用如下形式的约束：

  ```
  CHECK ( x = 1::bigint )
  ```

  这个问题并不限于bigint数据类型——只要常量的默认数据类型与被比较列的数据类型不匹配，就可能出现。所提供查询中的跨数据类型比较通常没有问题，只是不能出现在CHECK条件中。

- 约束排除期间会考虑主表所有分区上的全部约束，因此大量分区很可能显著增加查询规划时间。

- 别忘了你仍需要分别在各个分区上运行ANALYZE。像这样的命令：

  ```
  ANALYZE measurement;
  ```

  只会处理主表。

- 对主表执行的UPDATE和DELETE命令目前不会执行约束排除。

5.10. 其他数据库对象

表是关系数据库结构中的核心对象，因为它们承载着数据。
但它们并不是数据库中唯一存在的对象。还可以创建许多其他种类的对象，
以便让数据的使用和管理更加高效或方便。本章不会讨论这些对象，但这里给出一个列表，
让你了解有哪些可能性：

- 视图
- 函数和操作符
- 数据类型和域
- 触发器和重写规则

这些主题的详细信息请见第 V 部分。

5.11. 依赖跟踪

当我们创建一个涉及到很多具有外键约束、视图、触发器、函数等的表的复杂数据库结构时，
我们隐式地创建了一张对象之间的依赖关系网。例如，
具有一个外键约束的表依赖于它所引用的表。

为了保证整个数据库结构的完整性，PostgreSQL确保我们无法删除仍然被其他对象依赖的对象。
例如，尝试删除第 5.3.5 节中考虑过的产品表，由于有订单表依赖于它，会导致一个如下的错误消息：

```
DROP TABLE products;

NOTICE:  constraint orders_product_no_fkey on table orders depends on
 table products
ERROR:  cannot drop table products because other objects depend on it
HINT:  Use DROP ... CASCADE to drop the dependent objects too.
```

该错误消息包含了一个有用的提示：如果我们不想一个一个去删除所有的依赖对象，
我们可以执行：

```
DROP TABLE products CASCADE;
```

这样所有的依赖对象将被移除。在这种情况下，它不会移除订单表，只会移除外键约束。
（如果想检查`DROP ... CASCADE`会做什么，可以不带`CASCADE`运行`DROP`并阅读`NOTICE`消息。）

PostgreSQL中所有 DROP 命令都支持指定`CASCADE`。当然，可能出现的依赖关系形态会随着对象类型不同而变化。你也可以写`RESTRICT`来代替`CASCADE`，从而得到默认行为，
也就是阻止删除任何被其他对象依赖的对象。

> **注意**
>
> 根据 SQL 标准，必须指定`RESTRICT`或`CASCADE`二者之一。实际上，没有数据库系统强制执行这条规则，但不同系统的默认行为是`RESTRICT`还是`CASCADE`，各有不同。

注意

在升级过程中，来自PostgreSQL 7.3 之前版本的外键约束依赖和 serial 列依赖不会被维护或创建。从 7.3 之前的数据库升级时，所有其他类型的依赖都会被正确创建。

第 6 章 数据操纵

前一章讨论了如何创建用于保存数据的表和其他结构。现在该把数据填入表中了。
本章介绍如何插入、更新和删除表数据。
我们还将介绍在特定事件发生时实现自动数据变更的方法：触发器和重写规则。
下一章则终于会讲到如何从数据库中取回那些久寻不得的数据。

6.1. 插入数据

创建表时，其中不包含任何数据。在数据库真正派上用场之前，首先要做的就是插入数据。
从概念上说，数据一次插入一行。当然也可以插入多行，但不能插入少于一行的数据。
即使只知道部分列值，也必须创建完整的一行。

要创建新行，可使用INSERT命令。该命令需要提供表名和列值。例如，考虑第 5 章中的 products 表：

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

插入一行的示例命令如下：

```
INSERT INTO products VALUES (1, 'Cheese', 9.99);
```

数据值按列在表中出现的顺序列出，并以逗号分隔。通常，数据值会是字面量（常量），
但也允许使用标量表达式。

上述语法的缺点在于你必须知道表中列的顺序。为了避免这一点，你也可以显式列出列。例如，
下面两条命令的效果都与上面的命令相同：

```
INSERT INTO products (product_no, name, price) VALUES (1, 'Cheese', 9.
99);
INSERT INTO products (name, price, product_no) VALUES ('Cheese', 9.99,
 1);
```

许多用户认为，总是列出列名是一种好习惯。

如果你没有所有列的值，可以省略其中一些列。在这种情况下，这些列将填入各自的默认值。
例如：

```
INSERT INTO products (product_no, name) VALUES (1, 'Cheese');
INSERT INTO products VALUES (1, 'Cheese');
```

第二种形式是PostgreSQL 的一个扩展。它从左到右用给出的值填充各列，其余列则取默认值。

为求清晰，你也可以显式请求默认值，既可以针对单个列，也可以针对整行：

```
INSERT INTO products (product_no, name, price) VALUES (1, 'Cheese',
 DEFAULT);
INSERT INTO products DEFAULT VALUES;
```

> **提示**
>
> 要进行“批量装载”（即一次插入大量数据），可以考虑使用COPY命令。
> 它不像INSERT命令那样灵活，但效率更高。关于如何改善批量装载性能的更多信息，
> 请参阅第 13.4 节。

6.2. 更新数据

修改数据库中已有的数据称为更新。你可以更新单独的行、表中的所有行，
或者所有行中的一个子集。每一列都可以单独更新；其他列不受影响。

要执行更新，你需要三部分信息：

1. 要更新的表名和列名
2. 列的新值
3. 要更新哪一行（或哪些行）

回想一下第 5 章中的内容，SQL 通常并不为行提供唯一标识符。因此，并不总能直接指定要更新哪一行。取而代之的是，你要指定一行必须满足哪些条件才会被更新。
只有当表中存在主键时（不论你是否声明了它），
才能通过选择一个匹配主键的条件来可靠地定位单独的行。
图形化数据库访问工具正是依赖这一事实来允许你逐行更新。

例如，这条命令把所有价格为 5 的产品的价格更新为 10：

```
UPDATE products SET price = 10 WHERE price = 5;
```

这可能更新零行、一行或多行。尝试执行一个匹配不到任何行的更新并不算错误。

让我们仔细看看这条命令。首先是关键字 `UPDATE`，后面跟着表名。和平常一样，表名可以用模式限定；否则就在搜索路径中查找。接着是关键字 `SET`，后面跟着列名、一个等号以及新的列值。新列值可以是任意标量表达式，而不仅仅是常量。例如，如果你想把所有产品的价格提高 10%，可以使用：

```
UPDATE products SET price = price * 1.10;
```

如你所见，用于新值的表达式也可以引用该行中的现有值。我们还省略了`WHERE`子句。
如果省略它，就意味着更新表中的所有行。如果给出了它，则只有匹配 `WHERE` 条件的那些行才会被更新。请注意，`SET` 子句中的等号表示赋值，而 `WHERE` 子句中的等号表示比较，但这不会造成歧义。当然，`WHERE` 条件不必是等值测试。还有许多其他操作符可用（见第 9 章）。不过，该表达式必须计算出一个布尔结果。

你可以在一条`UPDATE`命令中更新多个列，只需在 `SET`子句中列出多个赋值。例如：

```
UPDATE mytable SET a = 5, b = 3, c = 1 WHERE a > 0;
```

6.3. 删除数据

到目前为止，我们已经解释了如何向表中添加数据以及如何修改数据。
接下来要讨论的是如何删除不再需要的数据。正如添加数据只能按整行进行一样，
从表中删除数据也只能按整行进行。前一节已经说明，SQL 不提供直接寻址单独行的方法。因此，只能通过指定待删除行必须满足的条件来删除它们。如果表中有主键，

那么你可以指定精确的一行；但你也可以删除匹配某个条件的一组行，
或者一次删除表中的所有行。

可以使用DELETE命令删除行；其语法与`UPDATE`命令非常相似。例如，要删除 products 表中所有价格为 10 的行，可以使用：

```
DELETE FROM products WHERE price = 10;
```

如果你只是写：

```
DELETE FROM products;
```

那么表中的所有行都会被删除！程序员须当心。

第 7 章 查询

前面的章节解释了如何创建表、如何用数据填充它们，以及如何操纵这些数据。
现在我们终于可以讨论如何从数据库中检索数据了。

7.1. 概述

从数据库检索数据的过程或用于检索数据的命令称为查询。在 SQL 中，SELECT命令用于指定查询。SELECT命令的一般语法为：

```
SELECT select_list FROM table_expression [sort_specification]
```

以下各节将详细描述选择列表、表表达式和排序说明。

一种简单查询的形式如下：

```
SELECT * FROM table1;
```

假设存在一个名为table1的表，则这条命令会从table1中检索所有行以及所有列。
（具体的检索方式取决于客户端应用。例如，psql程序会在屏幕上显示一个 ASCII 形式的表格，而客户端库则会提供从查询结果中提取单个值的函数。）选择列表说明*
表示表表达式所提供的全部列。选择列表也可以只选择部分可用列，或者利用这些列进行计算。
例如，如果table1有名为a、b和c的列（还可能有其他列），那么可以写出下面的查询：

```
SELECT a, b + c FROM table1;
```

（假设b和c都是数值数据类型）。有关更多细节请参见第 7.3 节。

FROM table1是一种特别简单的表表达式：它只读取一个表。通常，表表达式可以是由基本表、连接和子查询组成的复杂结构。不过，你也可以完全省略表表达式，把SELECT
命令当成计算器来使用：

```
SELECT 3 * 4;
```

如果选择列表中的表达式会返回变化的结果，这种用法就更有用了。例如，
你可以这样调用函数：

```
SELECT random();
```

7.2. 表表达式

表表达式用于计算出一个表。表表达式包含一个FROM子句，后面可以根据需要跟上WHERE、GROUP BY和HAVING子句。最简单的表表达式只是引用磁盘上的一个表，即所谓的基本表；但也可以使用更复杂的表达式，以多种方式修改或组合基本表。

表表达式中可选的WHERE、GROUP BY和HAVING子句指定了一个连续转换的流水线，这些转换作用于由FROM子句派生出的表。所有这些转换都会生成一个虚拟表，
该表提供的各行会传递给选择列表，以计算查询的输出行。

7.2.1. FROM子句

此FROM Clause从逗号分隔的表引用列表所给出的一个或多个其他表推导出一个表。

```
FROM table_reference [, table_reference [, ...]]
```

表引用可以是表名（可能带模式限定），也可以是子查询、表连接等推导表，
或它们的复杂组合。如果FROM子句中列出了多个表引用，它们会被交叉连接（见下文）
以形成中间虚拟表，随后该表可以用WHERE，GROUP BY和HAVING子句对它进行变换，最终得到
整个表表达式的结果。

当表引用命名的是一张表的超表（supertable）时，除非在表名前加上关键字ONLY，
否则该表引用不仅会产生该表中的行，还会产生其所有子表后继中的行。不过，
这种引用只会产生该命名表中出现的列 – 子表中新增的列会被忽略。

7.2.1.1. 连接表

一个连接表是根据特定的连接类型的规则从两个其它表（真实表或生成表）中派生的表。
目前支持内连接、外连接和交叉连接。

连接类型

Cross join

T1 CROSS JOIN *T2*

对*T1*的每一行与*T2*的每一行的组合，派生表将包含这样一行：它由所有*T1*
里面的列后面跟着所有*T2*里面的列构成。如果两个表分别有 N 和 M 行，连接表将有 N * M
行。

FROM *T1* CROSS JOIN *T2*等效于FROM *T1*,*T2*。它也等效于FROM *T1* INNER JOIN
T2 ON TRUE（见下文）。

Qualified joins

T1 { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN *T2*
ON *boolean_expression*
T1 { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN *T2* USING (
join column list)
T1 NATURAL { [INNER] | { LEFT | RIGHT | FULL } [OUTER] } JOIN *T2*

INNER和OUTER在所有形式中都是可选的。INNER是默认值；LEFT、RIGHT和FULL
表示外连接。

连接条件在ON或USING子句中指定，或者用关键字NATURAL隐含地指定。
连接条件决定来自两个源表中的哪些行是“匹配”的，这些我们将在后文详细解释。

ON子句是最通用的连接条件形式：它接受一个与WHERE
子句中所用种类相同的布尔值表达式。如果一对分别来自*T1*和*T2*的行在ON
表达式上的求值结果为真，那么它们就是匹配的。

USING是一种简写记法：它接受一个逗号分隔的列名列表，
这些列名必须是两个被连接表共有的，
并为每一对这样的列构造指定相等性比较的连接条件。此外，JOIN USING的输出会为每一
对相等的输入列各生成一个列，后面跟着两个表的所有其他列。因此，USING (a, b, c)等
效于ON (t1.a = t2.a AND t1.b = t2.b AND t1.c = t2.c)，区别在于：如果使用
ON，结果中a、b、c各会出现两个列，而使用USING时每个列只有一个。

最后，NATURAL是USING的一种简写形式：它会形成一个USING列表，
其中恰好包含那些同时出现在两个输入表中的列名。和USING一样，
这些列在输出表中只出现一次。

限定连接有以下几种类型：

INNER JOIN

对于 T1 的每一行 R1，生成的连接表都有一行对应 T2 中的每一个满足和 R1 的连接条件的行。

LEFT OUTER JOIN

首先，执行一次内连接。然后，为 T1 中每一个无法在连接条件上匹配 T2 里任何一行的行返回一个连接行，该连接行中 T2 的列用空值补齐。因此，生成的连接表里为来自 T1 的每一行都至少包含一行。

RIGHT OUTER JOIN

首先，执行一次内连接。然后，为 T2 中每一个无法在连接条件上匹配 T1 里任何一行的行返回一个连接行，该连接行中 T1 的列用空值补齐。因此，生成的连接表里为来自 T2 的每一行都至少包含一行。

FULL OUTER JOIN

首先，执行一次内连接。然后，为 T1 中每一个无法在连接条件上匹配 T2 里任何一行的行返回一个连接行，该连接行中 T2 的列用空值补齐。同样，为 T2 中每一个无法在连接条件上匹配 T1 里任何一行的行返回一个连接行，该连接行中 T1 的列用空值补齐。

所有类型的连接都可以被链在一起或者嵌套：*T1*和*T2*之一或两者都可以是连接表。在`JOIN`子句周围可以使用圆括号来控制连接顺序。如果不使用圆括号，`JOIN`子句会从左至右嵌套。

综合起来看，假设我们有表`t1`：

```
 num | name
-----+------
   1 | a
   2 | b
   3 | c
```

和`t2`：

```
 num | value
-----+-------
   1 | xxx
   3 | yyy
   5 | zzz
```

然后我们用不同的连接方式可以获得各种结果：

```
=>SELECT * FROM t1 CROSS JOIN t2;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   1 | a    |   3 | yyy
   1 | a    |   5 | zzz
   2 | b    |   1 | xxx
   2 | b    |   3 | yyy
   2 | b    |   5 | zzz
   3 | c    |   1 | xxx
   3 | c    |   3 | yyy
   3 | c    |   5 | zzz
(9 rows)
```

```
=> SELECT * FROM t1 INNER JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   3 | c    |   3 | yyy
(2 rows)

=> SELECT * FROM t1 INNER JOIN t2 USING (num);
 num | name | value
-----+------+-------
   1 | a    | xxx
   3 | c    | yyy
(2 rows)

=> SELECT * FROM t1 NATURAL INNER JOIN t2;
 num | name | value
-----+------+-------
   1 | a    | xxx
   3 | c    | yyy
(2 rows)

=> SELECT * FROM t1 LEFT JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |   3 | yyy
(3 rows)

=> SELECT * FROM t1 LEFT JOIN t2 USING (num);
 num | name | value
-----+------+-------
   1 | a    | xxx
   2 | b    |
   3 | c    | yyy
(3 rows)

=> SELECT * FROM t1 RIGHT JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   3 | c    |   3 | yyy
     |      |   5 | zzz
(3 rows)

=> SELECT * FROM t1 FULL JOIN t2 ON t1.num = t2.num;
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |   3 | yyy
     |      |   5 | zzz
(4 rows)
```

用ON指定的连接条件也可以包含与连接不直接相关的条件。这种功能可能对某些查询很有用，但是需要我们仔细想清楚。例如：

```
=>SELECT * FROM t1 LEFT JOIN t2 ON t1.num = t2.num AND t2.value =
 'xxx';
 num | name | num | value
-----+------+-----+-------
   1 | a    |   1 | xxx
   2 | b    |     |
   3 | c    |     |
(3 rows)
```

7.2.1.2. 表和列别名

你可以给表以及复杂的表引用指定一个临时名字，以便在查询的其余部分引用该派生表。这被称为表别名。

要创建一个表别名，我们可以写：

```
FROM table_reference AS alias
```

或者

```
FROM table_reference alias
```

AS关键字只是一些噪音。*alias*可以是任意标识符。

表别名的典型应用是给长表名赋予比较短的标识符，好让连接子句更易读。例如：

```
SELECT * FROM some_very_long_table_name s JOIN another_fairly_long_
name a ON s.id = a.num;
```

对于当前查询而言，别名成为该表引用的新名称–此后不再可能用原来的名称引用该表。因此：

```
SELECT * FROM my_table AS m WHERE my_table.a > 5;
```

这不是有效的 SQL 语法。实际会发生的（这是PostgreSQL对标准的一个扩展）是向FROM子句隐式添加一个表引用，因此该查询会像下面这样写一样被处理：

```
SELECT * FROM my_table AS m, my_table AS my_table WHERE my_table.a >
 5;
```

这将导致一个交叉连接，而这通常不是你想要的。

表别名主要用于简化符号，但是当把一个表连接到它自身时必须使用别名，例如：

```
SELECT * FROM my_table AS a CROSS JOIN my_table AS b ...
```

此外，当表引用是子查询时也必须使用别名（参见第 7.2.1.3 节）。

圆括弧用于解决歧义。下面的语句会把别名b赋给连接的结果，这与前一个示例不同：

```
SELECT * FROM (my_table AS a CROSS JOIN my_table) AS b ...
```

另外一种给表指定别名的形式是给表的列赋予临时名字，就像给表本身指定别名一样：

```
FROM table_reference [AS] alias ( column1 [, column2 [, ...]] )
```

如果指定的列别名比表里实际的列少，那么剩下的列就没有被重命名。
这种语法对于自连接或子查询特别有用。

如果用这些形式中的任何一种给一个JOIN子句的输出附加了一个别名，那么该别名就在JOIN的作用下隐去了其原始的名字。例如：

```
SELECT a.* FROM my_table AS a JOIN your_table AS b ON ...
```

是合法 SQL，但是：

```
SELECT a.* FROM (my_table AS a JOIN your_table AS b ON ...) AS c
```

是不合法的：表别名a在别名c外面是看不到的。

7.2.1.3. 子查询

指定派生表的子查询必须用圆括号括起来，并且必须赋予一个表别名。（参见第 7.2.1.2 节。）例如：

```
FROM (SELECT * FROM table1) AS alias_name
```

这个示例等效于FROM table1 AS alias_name。更有趣的情况是在子查询里面有分组或聚合的时候，子查询不能被简化为一个简单的连接。

7.2.1.4. 表函数

表函数是那些生成行集合的函数，这些行可以由基础类型（标量类型）组成，
也可以由复合数据类型（表行）组成。它们在查询的FROM子句中用法类似于表、视图或子查询。
表函数返回的列可以像表、视图或子查询的列一样，出现在SELECT、JOIN或WHERE子句中。

如果一个表函数返回基础数据类型，唯一的结果列名与该函数名相同。
如果该函数返回一个复合类型，结果列会得到与该类型的各个属性相同的名称。

表函数可以在FROM子句中被赋予别名，但也可以不使用别名。如果一个函数在FROM子句中使用时没有别名，该函数名将被用作结果表的表名。

示例：

```
CREATE TABLE foo (fooid int, foosubid int, fooname text);

CREATE FUNCTION getfoo(int) RETURNS SETOF foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT * FROM getfoo(1) AS t1;

SELECT * FROM foo
    WHERE foosubid IN (
                        SELECT foosubid
                        FROM getfoo(foo.fooid) z
                        WHERE z.fooid = foo.fooid
                      );

CREATE VIEW vw_getfoo AS SELECT * FROM getfoo(1);

SELECT * FROM vw_getfoo;
```

有时候，定义一个能够根据调用方式返回不同列集合的表函数很有用。为支持这一点，表函数可以被声明为返回伪类型record。当此类函数在查询中使用时，必须在查询本身中指定预期的行结构，这样系统才能知道如何分析和规划该查询。考虑下面的示例：

```
SELECT *
    FROM dblink('dbname=mydb', 'SELECT proname, prosrc FROM pg_proc')
      AS t1(proname name, prosrc text)
    WHERE proname LIKE 'bytea%';
```

dblink函数执行一个远程查询（见contrib/dblink）。它被声明为返回record，因为它可能用于任意类型的查询。实际的列集必须在调用它的查询中指定，这样分析器才知道像*这样的写法应当扩展成什么。

7.2.2. WHERE子句

此WHERE Clause的语法是：

```
WHERE search_condition
```

其中，*search_condition*可以是任意值表达式（见第 4.2 节），只要其返回值的类型为boolean。

在完成FROM子句的处理之后，生成的虚拟表中的每一行都会根据搜索条件进行检查。如果条件结果为真，该行就会保留在输出表中；否则（也就是说，如果结果为假或空）就会被丢弃。搜索条件通常至少会引用FROM子句生成的表中的某一列；这并非必须，但若不引用这些列，WHERE子句通常就没什么用了。

> **注意**
>
> 内连接的连接条件既可以写在WHERE子句也可以写在JOIN子句里。例如，这些表表达式是等效的：
>
> ```
> FROM a, b WHERE a.id = b.id AND b.val > 5
> ```
>
> 和：
>
> ```
> FROM a INNER JOIN b ON (a.id = b.id) WHERE b.val > 5
> ```
>
> 或者可能还有：
>
> ```
> FROM a NATURAL JOIN b WHERE b.val > 5
> ```
>
> 选择其中哪一种主要只是风格问题。FROM子句中的JOIN语法虽然属于 SQL 标准，但在其他 SQL 数据库管理系统中可能没有那么容易移植。对于外连接无论如何都没有选择：它们必须在FROM子句中完成。外连接的ON或USING子句与WHERE条件并不等价，因为它除了会在最终结果中移除行之外，还会为不匹配的输入行增加结果行。

下面是一些WHERE子句的例子：

```
SELECT ... FROM fdt WHERE c1 > 5

SELECT ... FROM fdt WHERE c1 IN (1, 2, 3)
```

```
SELECT ... FROM fdt WHERE c1 IN (SELECT c1 FROM t2)

SELECT ... FROM fdt WHERE c1 IN (SELECT c3 FROM t2 WHERE c2 = fdt.c1 +
 10)

SELECT ... FROM fdt WHERE c1 BETWEEN (SELECT c3 FROM t2 WHERE c2 =
 fdt.c1 + 10) AND 100

SELECT ... FROM fdt WHERE EXISTS (SELECT c1 FROM t2 WHERE c2 > fdt.c1)
```

fdt是由FROM子句推导得到的表。不满足WHERE子句搜索条件的行会从fdt中消除。注意，这里将标量子查询用作值表达式。与其他查询一样，子查询也可以使用复杂的表表达式。还要注意子查询中对fdt的引用方式。将c1限定为fdt.c1仅在c1也是子查询推导输入表中的列名时才有必要。不过，即使不必限定，限定列名仍能使含义更清晰。本例说明了外层查询的列命名作用域如何扩展到其内部查询。

7.2.3. GROUP BY和HAVING子句

通过WHERE过滤之后，派生出来的输入表还可以使用GROUP BY子句进行分组，并通过HAVING子句排除某些分组行。

```
SELECT select_list
    FROM ...
    [WHERE ...]
    GROUP BY grouping_column_reference [, grouping_column_reference].
..
```

此GROUP BY Clause用于把表中所有列出列的值均相同的行分到同一组。列的列出顺序没有影响。其效果是把具有相同值的每一组行合并为一个分组行，由它代表组内的所有行。这样可以消除输出中的冗余，或者计算应用于各组的聚合。例如：

```
=>SELECT * FROM test1;
 x | y
---+---
 a | 3
 c | 2
 b | 5
 a | 1
(4 rows)

=> SELECT x FROM test1 GROUP BY x;
 x
---
 a
 b
 c
(3 rows)
```

在第二个查询里，我们不能写成SELECT * FROM test1 GROUP BY x，因为列y没有一个单一值可以与每个分组对应。被分组的列可以在选择列表中引用是因为它们在每个组都有单一的值。

通常，如果一个表被分了组，那么那些没有用于分组的列就不能被引用，除非在聚合表达式中。一个用聚合表达式的示例是：

```
=>SELECT x, sum(y) FROM test1 GROUP BY x;
 x | sum
---+-----
 a |   4
 b |   5
 c |   2
(3 rows)
```

这里的sum是一个聚合函数，它在整个组上计算出一个单一值。
有关可用的聚合函数的更多信息可以在第 9.15 节中找到。

> **提示**
>
> 不带聚合表达式的分组实际上是在计算某一列的非重复值集合。这也可以用DISTINCT子句实现（参阅第 7.3.3 节）。

这里是另外一个示例：它计算每种产品的总销售额（而不是所有产品的总销售额）：

```
SELECT product_id, p.name, (sum(s.units) * p.price) AS sales
    FROM products p LEFT JOIN sales s USING (product_id)
    GROUP BY product_id, p.name, p.price;
```

在这个示例里，列product_id、p.name和p.price必须在GROUP BY子句里，因为它们都在查询的选择列表里被引用到。（根据 products 表的具体设计，name 和 price 可能完全依赖于产品 ID，因此理论上这些额外的分组并非必要，不过这一点尚未实现。）列s.units不必在GROUP BY列表里，因为它只是在一个聚合表达式（sum(...)）里使用，它代表一种产品的销售额。对于每种产品，这个查询都返回一行关于该产品所有销售额的汇总。

在严格的 SQL 里，GROUP BY只能对源表的列进行分组，但PostgreSQL把这个扩展为也允许GROUP BY去根据选择列表中的列分组。也允许对值表达式进行分组，而不仅是简单的列名。

如果一个表已经用GROUP BY子句分了组，而你只对其中某些组感兴趣，那么就可以使用HAVING子句。它很像WHERE子句，用于从结果中删除某些组。其语法是：

```
SELECT select_list FROM ... [WHERE ...] GROUP BY ... HAVING boolean_
expression
```

在HAVING
子句中的表达式可以引用分组的表达式和未分组的表达式（后者必须涉及一个聚合函数）。

示例：

```
=>SELECT x, sum(y) FROM test1 GROUP BY x HAVING sum(y) > 3;
 x | sum
---+-----
 a |   4
 b |   5
(2 rows)

=> SELECT x, sum(y) FROM test1 GROUP BY x HAVING x < 'c';
 x | sum
---+-----
 a |   4
```

```
 b |   5
(2 rows)
```

再次，一个更现实的示例：

```
SELECT product_id, p.name, (sum(s.units) * (p.price - p.cost)) AS
 profit
    FROM products p LEFT JOIN sales s USING (product_id)
    WHERE s.date > CURRENT_DATE - INTERVAL '4 weeks'
    GROUP BY product_id, p.name, p.price, p.cost
    HAVING sum(p.price * s.units) > 5000;
```

在上面的示例里，`WHERE`
子句通过一个未分组的列来选择数据行（该表达式仅对最近四周内发生的销售为真），而`HAVING`子句则把输出限制为总销售收入超过 5000 的组。请注意，聚合表达式不必在查询的所有部分都完全相同。

7.3. 选择列表

如前面几节所示，`SELECT`命令中的表表达式会构造出一个中间虚拟表，
其中可能涉及组合表和视图、消除行、分组等操作。这个表随后交由选择列表处理。
选择列表决定中间表中的哪些列会实际输出。

7.3.1. 选择列表项

最简单的选择列表是`*`，它会输出表表达式生成的所有列。否则，
选择列表就是一个逗号分隔的值表达式列表（定义见第 4.2 节）。例如，它可以是一个列名列表：

```
SELECT a, b, c FROM ...
```

列名`a`、`b`和`c`要么是`FROM`子句中所引用表的实际列名，要么是像第 7.2.1.2 节所述为它们指定的别名。选择列表中可用的名字空间与`WHERE`子句相同；如果使用了分组，则与`HAVING`
子句相同。

如果超过一个表有同样的列名，那么你还必须给出表名字，如：

```
SELECT tbl1.a, tbl2.a, tbl1.b FROM ...
```

在使用多个表时，要求一个特定表的所有列也是有用的：

```
SELECT tbl1.*, tbl2.a FROM ...
```

（另请参阅第 7.2.2 节。）

如果在选择列表中使用任意值表达式，那么从概念上说，它会在返回表中增加一个新的虚拟列。
该值表达式会对每个结果行计算一次，并用该行中的值替换其中的列引用。不过，
选择列表中的表达式并不一定要引用`FROM`子句表表达式中的列；例如，
它也可以是任意常量算术表达式。

7.3.2. 列标签

选择列表中的项可以被赋予名字以供进一步处理。此处的“进一步处理”
指的是可选的排序说明和客户端应用（例如，用于显示的列标题）。例如：

```
SELECT a AS value, b + c AS sum FROM ...
```

如果没有使用AS指定输出列名，系统就会分配一个默认列名。对于简单的列引用，它就是被引用列的名字；对于函数调用，它就是函数的名字；对于复杂表达式，系统则会生成一个通用名称。

> **注意**
>
> 输出列的命名和在FROM子句里的命名是不一样的（参阅第 7.2.1.2 节）。它实际上允许你对同一个列命名两次，但是在选择列表中分配的名字是要传递下去的名字。

7.3.3. DISTINCT

在处理完选择列表之后，结果表还可以选择性地删除重复行。我们可以直接在SELECT后面写上DISTINCT关键字来指定：

```
SELECT DISTINCT select_list ...
```

（如果不用DISTINCT，则可以用ALL关键字来显式指定保留所有行这一默认行为。）

显然，如果两行至少在一个列值上不同，就认为它们是不同的。空值在这种比较中被认为是相等的。

另外，我们还可以用任意表达式来判断什么行可以被认为是可区分的：

```
SELECT DISTINCT ON (expression [, expression ...]) select_list ...
```

这里的*expression*是一个对所有行求值的任意值表达式。如果某一组行在这些表达式上的值都相等，那么它们就会被视为重复行，因此只保留该组中的第一行。请注意，某一组里的“第一行”是不可预测的，除非查询在足够多的列上进行了排序，以保证到达DISTINCT过滤器的行顺序是唯一的。（DISTINCT ON的处理发生在ORDER BY排序之后。）

DISTINCT ON子句不是 SQL 标准的一部分，有时会被认为风格不佳，因为它的结果可能具有不确定性。通过审慎使用GROUP BY以及FROM中的子查询，可以避免使用这种结构，但它往往是最方便的替代方案。

7.4. 组合查询

两个查询的结果可以用集合操作并、交、差进行组合。语法是

```
query1 UNION [ALL] query2
query1 INTERSECT [ALL] query2
query1 EXCEPT [ALL] query2
```

*query1*和*query2*都是可以使用以上所有特性的查询。集合操作也可以嵌套和链式使用，例如

```
query1 UNION query2 UNION query3
```

它会被当作：

```
(query1 UNION query2) UNION query3
```

来执行。

UNION有效地把*query2*的结果附加到*query1*
的结果上（不过我们不能保证这就是这些行实际被返回的顺序）。此外，
它将删除结果中所有重复的行，就像DISTINCT做的那样，除非你使用了UNION ALL。

INTERSECT返回那些同时存在于*query1*和*query2*的结果中的行，除非声明了INTERSECT ALL，否则所有重复行都被消除。

EXCEPT返回所有在*query1*结果中但不在*query2*结果中的行。
（这有时被称为两个查询的差。）同样，除非使用了EXCEPT ALL，否则重复行都会被消除。

为了计算两个查询的并、交、差，这两个查询必须是“并操作兼容的”，
也就意味着它们都返回同样数量的列，并且对应的列有兼容的数据类型，如第 10.5 节中描述的那样。

7.5. 行排序

当一个查询生成输出表之后（也就是处理完选择列表之后），还可以选择对其排序。
如果没有选择排序，行将按未指定的顺序返回。
这时的实际顺序取决于扫描和连接计划类型以及磁盘上的物理顺序，但绝不能依赖这些因素。
只有显式选择排序步骤，才能保证特定的输出顺序。

ORDER BY子句指定了排序顺序：

```
SELECT select_list
    FROM table_expression
    ORDER BY column1 [ASC | DESC] [, column2 [ASC | DESC] ...]
```

*column1*等指的是选择列表中的列。它们既可以是一个列的输出名（见第 7.3.2 节），也可以是一个列的编号。一些例子：

```
SELECT a, b FROM table1 ORDER BY a;
SELECT a + b AS sum, c FROM table1 ORDER BY sum;
SELECT a, sum(b) FROM table1 GROUP BY a ORDER BY 1;
```

作为对 SQL 标准的一个扩展，PostgreSQL还允许按任意表达式排序：

```
SELECT a, b FROM table1 ORDER BY a + b;
```

还允许引用FROM子句中未出现在选择列表里的列名：

```
SELECT a FROM table1 ORDER BY b;
```

但这些扩展在涉及UNION、INTERSECT或EXCEPT的查询中不起作用，并且不可移植到其他 SQL 数据库。

每个列说明后面都可以选择性地写上ASC或DESC，以指定升序或降序排序方向。ASC是默认值。
升序会把较小的值排在前面，而“较小”是由<操作符定义的。类似地，降序则由>操作符定义。[1]

如果指定了多个排序列，
那么后面的项将用于对那些在前面排序列所确定的顺序下相等的行继续排序。

[1] 事实上，PostgreSQL会使用该列数据类型的默认 B-树操作符类来决定ASC和DESC的排序顺序。按惯例，数据类型会被设置成由<和>操作符来对应这种排序顺序，不过用户定义数据类型的设计者也可以选择不同的做法。

7.6. LIMIT和OFFSET

LIMIT和OFFSET允许仅检索查询其余部分所生成行的一部分：

```
SELECT select_list
    FROM table_expression
    [LIMIT { number | ALL }] [OFFSET number]
```

如果给出了一个限制计数，那么会返回数量不超过该限制的行（但可能更少些，因为查询本身可能生成的行数就比较少）。LIMIT ALL的效果和省略LIMIT子句一样。

OFFSET说明在开始返回结果行之前要跳过多少行。OFFSET 0的效果与省略OFFSET子句相同。如果OFFSET和LIMIT都出现，那么在开始数要返回的LIMIT行之前，会先跳过OFFSET行。

如果使用LIMIT，那么用一个ORDER BY子句把结果行约束成一个唯一的顺序是很重要的。否则你就会拿到一个不可预料的该查询的行的子集。你要的可能是第十到第二十行，但以什么顺序的第十到第二十？除非你指定了ORDER BY，否则顺序是不知道的。

查询优化器在生成查询计划时会考虑LIMIT，因此如果你给定LIMIT和OFFSET，那么你很可能收到不同的规划（产生不同的行顺序）。因此，使用不同的LIMIT/OFFSET值选择查询结果的不同子集将生成不一致的结果，除非你用ORDER BY强制一个可预测的顺序。这并非bug，这是一个很自然的结果，因为 SQL 没有许诺把查询的结果按照任何特定的顺序发出，除非用了ORDER BY来约束顺序。

被OFFSET子句跳过的行仍然需要在服务器内部计算；因此，较大的OFFSET可能效率不高。

第 8 章 数据类型

PostgreSQL 提供了丰富的原生数据类型。用户可以使用CREATE TYPE命令向 PostgreSQL添加新类型。

表 8.1展示了所有内置的通用数据类型。“别名”列中列出的多数可选名称，都是 PostgreSQL出于历史原因在内部使用的名称。此外，还有一些内部使用或已废弃的类型也可用，
但未在此列出。

表 8.1. 数据类型

名字	别名	描述
bigint	int8	有符号的8字节整数
bigserial	serial8	自动递增的8字节整数
bit [(*n*)]		定长位串
bit varying [(*n*)]	varbit	变长位串
boolean	bool	逻辑布尔值（真/假）
box		平面上的矩形框
bytea		二进制数据（“字节数组”）
character varying [(*n*)]	varchar [(*n*)]	变长字符串
character [(*n*)]	char [(*n*)]	定长字符串
cidr		IPv4或IPv6网络地址
circle		平面上的圆
date		日历日期（年、月、日）
double precision	float8	双精度浮点数
inet		IPv4或IPv6主机地址
integer	int, int4	有符号4字节整数
interval [(*p*)]		时间段
line		平面上的无限长的线
lseg		平面上的线段
macaddr		MAC 地址
money		货币数量
numeric [(*p*, *s*)]	decimal [(*p*, *s*)]	可选择精度的精确数字
path		平面上的几何路径
point		平面上的几何点
polygon		平面上的封闭几何路径
real	float4	单精度浮点数
smallint	int2	有符号2字节整数
serial	serial4	自动递增的4字节整数
text		变长字符串
time [(*p*)] [without time zone]		一天中的时间（无时区）
time [(*p*)] with time zone	timetz	一天中的时间，包括时区

名字	别名	描述
timestamp [(*p*)] [without time zone]		日期和时间（无时区）
timestamp [(*p*)] with time zone	timestamptz	日期和时间，包括时区

兼容性

下列类型（或它们的这些拼写形式）由SQL规定：bit、bit varying、boolean、char、character varying、character、varchar、date、double precision、integer、interval、numeric、decimal、real、smallint、time（有时区或无时区）、timestamp（有时区或无时区）。

每种数据类型都有一种由其输入和输出函数决定的外部表示。
许多内置类型的外部格式都很直观。不过，也有一些类型是 PostgreSQL所特有的，例如几何路径；还有一些类型可能存在多种可能的格式，例如日期/时间类型。
有些输入和输出函数并不可逆，也就是说，输出函数的结果与原始输入相比可能会丢失精度。

8.1. 数字类型

数字类型包括 2、4、8 字节的整数，4、8 字节的浮点数，以及可选精度的小数。表 8.2列出了所有可用类型。

表 8.2. 数字类型

名字	存储大小	描述	范围
smallint	2字节	小范围整数	-32768 到 +32767
integer	4字节	整数的常用选择	-2147483648 到 +2147483647
bigint	8字节	大范围整数	-9223372036854775808 到 9223372036854775807
decimal	可变	用户指定精度，精确	无限制
numeric	可变	用户指定精度，精确	无限制
real	4字节	可变精度，不精确	6位十进制精度
double precision	8字节	可变精度，不精确	15位十进制精度
serial	4字节	自动递增的整数	1到2147483647
bigserial	8字节	自动递增的大整数	1到9223372036854775807

数字类型常量的语法在第 4.1.2 节里描述。数字类型有一整套对应的算术操作符和函数。相关信息请参考第 9 章。下面的几节详细描述这些类型。

8.1.1. 整数类型

类型 smallint、integer 和 bigint 用于存储不同范围的整数，也就是没有小数部分的数。试图存储超出允许范围的值会导致错误。

常用的类型是integer，因为它提供了在范围、存储空间和性能之间的最佳平衡。
一般只有在磁盘空间紧张的时候才使用smallint类型。而bigint类型只有在integer的范围不够用时才应使用，因为后者（integer）无疑更快。

bigint类型并非在所有平台上都能正确工作，因为它依赖于编译器对八字节整数的支持。在没有这种支持的机器上，bigint的行为与integer相同（但仍占用八个字节的存储空间）。不过，我们不知道有任何合理的平台属于这种情况。

SQL只规定了整数类型integer（或int）、smallint和bigint。类型int2、int4和int8都是扩展，也在某些其他SQL数据库系统中使用。

8.1.2. 任意精度数值

类型numeric可以存储最多 1000 位精度的数字并精确地执行计算。我们特别建议将它用于货币金额和其他要求精确的数量。不过，numeric类型上的算术运算比整数类型或者下一节描述的浮点数类型要慢很多。

下面使用如下术语：numeric的小数位数是小数部分中十进制数字的数量，即小数点右侧的位数。numeric的精度是整个数中有效数字的总数，即小数点两侧的数字位数之和。因此，数值 23.5141 的精度为 6，小数位数为 4。整数可以视为小数位数为零。

可以配置numeric列的最大精度和最大小数位数。要声明numeric类型的列，使用以下语法：

NUMERIC(*precision*, *scale*)

精度必须为正数，小数位数必须为零或正数。也可以使用：

NUMERIC(*precision*)

这会选择小数位数为 0。指定：

NUMERIC

而不指定任何精度或小数位数，会创建一个可存储任意精度和小数位数数值的列，直到达到实现对精度的限制。这种列不会把输入值强制转换为某个特定的小数位数，而声明了小数位数的numeric列会将输入值强制转换为该小数位数。（SQL标准要求默认小数位数为 0，即强制转换为整数精度。我们认为这没有多少用处。如果你关心可移植性，请始终显式指定精度和小数位数。）

如果要存储的值的小数位数大于该列声明的小数位数，系统会把该值舍入到指定的小数位数。然后，如果小数点左侧的位数超过了声明的精度减去声明的小数位数，就会报错。

数值在物理上存储时不会保留多余的前导零或尾随零。因此，列上声明的精度和小数位数只是最大值，而不是固定分配的空间（从这个意义上说，numeric更像varchar(*n*)，而不像char(*n*)）。实际存储需求是每四个十进制数字组占两个字节，再加上 5 到 8 字节的开销。

除了普通数值外，numeric类型还允许特殊值NaN，表示“非数”。对NaN的任何运算都会产生另一个NaN。在 SQL 命令中将该值写为常量时，必须加上引号，例如UPDATE table SET x = 'NaN'。输入时，字符串NaN的识别不区分大小写。

类型decimal和numeric是等效的。两种类型都是SQL标准的一部分。

8.1.3. 浮点类型

数据类型real和double precision是不精确的、变精度的数字类型。实际上，在底层处理器、操作系统和编译器支持的范围内，这两种类型通常实现了IEEE 754 二进制浮点算术标准（分别对应单精度和双精度）。

所谓不精确，是指某些值无法被精确转换为内部格式，只能以近似值存储，
因此存储并再取出一个值时，可能会看到轻微差异。
如何处理这类误差以及它们在计算中如何传播，是数学和计算机科学中的一个独立领域，
这里不再展开，只强调以下几点：

- 如果你要求准确的存储和计算（例如计算货币金额），应使用`numeric`类型。
- 如果你想用这些类型进行任何重要的复杂计算，尤其是依赖边界情形（无穷大、下溢）特定行为的计算，那么应该仔细评估其实现。
- 比较两个浮点值是否相等，未必总能得到符合预期的结果。

在大多数平台上，`real`类型的范围至少为 1E-37 到 1E+37，精度至少为 6 位十进制数字。`double precision`类型的范围通常约为 1E-307 到 1E+308，精度至少为 15 位数字。过大或过小的值会导致错误。如果输入数字的精度过高，可能会发生舍入。过于接近零且无法表示为与零不同的数值，会导致下溢错误。

除了普通数值外，浮点类型还有几个特殊值：

`Infinity-InfinityNaN`

它们分别表示 IEEE 754 的特殊值“无穷大”，“负无穷大”和“非数字”。（在浮点运算不遵循 IEEE 754 的机器上，这些值可能无法按预期工作。）在 SQL 命令中将这些值写为常量时，必须加上引号，例如`UPDATE table SET x = 'Infinity'`。输入时，这些字符串的识别不区分大小写。

PostgreSQL 也支持 SQL 标准记法 `float` 和 `float(p)` 来指定不精确数值类型。这里，*p* 指定可接受的最小二进制精度位数。PostgreSQL 将 `float(1)` 到 `float(24)` 视为选择 `real` 类型，而 `float(25)` 到 `float(53)` 则选择 `double precision`。*p* 超出允许范围会报错。未指定精度的 `float` 视为 `double precision`。

> **注意**
>
> 在PostgreSQL 7.4 之前，`float(p)`中的精度被理解为这么多位十进制数字。现在这一点已被纠正，以符合 SQL 标准，即精度以二进制位来衡量。假定`real`和`double precision`的尾数分别恰好有 24 位和 53 位，对于 IEEE 标准浮点实现而言是正确的。在非 IEEE 平台上可能会有少许偏差，但为简单起见，所有平台都使用相同的*p*取值范围。

8.1.4. serial 类型

`serial`和`bigserial`并不是真正的数据类型，
它们只是为了创建唯一标识符列而提供的记法便利（类似于其他一些数据库支持的`AUTO_INCREMENT`属性）。在当前实现中，下列语句：

```
CREATE TABLE tablename (
    colname SERIAL
);
```

等效于指定：

```
CREATE SEQUENCE tablename_colname_seq;
CREATE TABLE tablename (
    colname integer DEFAULT nextval('tablename_colname_seq') NOT NULL
);
```

这样，我们就创建了一个整型列，并安排它的默认值由一个序列产生器分配。这里应用了NOT NULL约束，以确保不会插入空值。在大多数情况下，你可能还会想附加一个UNIQUE或PRIMARY KEY约束，以防止意外插入重复值，但这不是自动的。

> **注意**
>
> 在PostgreSQL 7.3 之前，serial隐含UNIQUE。现在这不再是自动的。如果希望一个 serial 列具有唯一约束或作为主键，现在必须像对待任何其他数据类型一样显式指定。

要向 serial 列插入序列中的下一个值，应指定让serial列使用其默认值。这既可以通过在 INSERT 语句的列表中省略该列来实现，也可以通过使用 DEFAULT 关键字来实现。

类型名serial和serial4是等价的：二者都会创建integer列。类型名bigserial和serial8的工作方式相同，只是它们创建的是bigint列。
如果预计表在其生命周期内会使用超过 2^{31} 个标识符，就应使用bigserial。

为 serial 列创建的序列会在其所属列被删除时自动删除，并且不能以其他方式删除。（在 PostgreSQL 7.3 之前的版本中情况并非如此。注意，对于通过重新装载 7.3 之前数据库的转储而创建的序列，这种自动删除关联不会发生；转储文件不包含建立依赖链接所需的信息。）此外，序列和列之间的这种依赖只针对serial列本身建立。
如果任何其他列引用了该序列（可能是通过手动调用nextval函数），
当序列被删除时它们将被破坏。以这种方式使用serial列的序列被认为是不良做法；
如果希望多个列从同一个序列产生器取值，应将序列创建为独立对象。

8.2. 货币类型

> **注意**
>
> money类型已被弃用。请改用numeric或decimal类型，并结合to_char函数。

money类型以固定的小数精度存储货币金额；参见表 8.3。输入支持多种格式，包括整数字面量和浮点数字面量，以及“典型的”货币格式，例如'$1,000.00'。输出通常也采用后一种形式，但会受到区域设置影响。

表 8.3. 货币类型

名字	存储大小	描述	范围
money	8 字节	货币额	-92233720368547758.08 到 +92233720368547758.07

8.3. 字符类型

表 8.4. 字符类型

名字	描述
character varying(*n*),varchar(*n*)	有长度限制的变长
character(*n*),char(*n*)	定长，空白填充
text	无限长度的变长

表 8.4显示了在PostgreSQL里可用的一般用途的字符类型。

SQL 定义了两种主要字符类型：character varying(*n*) 和 character(*n*)，其中 *n* 是正整数。这两种类型都可以存储长度最多为 *n* 个字符的字符串。试图把更长的字符串存入这些类型的列时会报错，除非超出的字符全部都是空格，在这种情况下字符串会被截断到最大长度。（这一多少有些奇怪的例外是 SQL 标准要求的。）如果要存储的字符串短于声明长度，character 类型的值会用空格填充；character varying 类型则只是简单地存储较短的字符串。

如果显式把一个值转换为 character varying(*n*) 或 character(*n*)，那么超长的值会被截断为 *n* 个字符而不会报错。（这同样是 SQL 标准要求的。）

> **注意**
>
> 在PostgreSQL 7.2 之前，无论显式还是隐式转换的场合，超长的字符串总是被截断而不报错。

varchar(*n*) 和 char(*n*) 分别是 character varying(*n*) 和 character(*n*) 的别名。character 若不带长度说明则等同于 character(1)。如果 character varying 不带长度说明，则该类型接受任意长度的字符串。这是 PostgreSQL 的扩展。

此外，PostgreSQL 还提供 text 类型，用于存储任意长度的字符串。虽然 text 类型不在 SQL 标准中，但其他若干 SQL 数据库管理系统也提供了它。

character类型的值在物理存储时会在右侧用空格填充到指定宽度*n*，并且也会按这种形式存储和显示。但是，这些填充空格在语义上被视为不重要。比较两个character类型的值时，尾随空格会被忽略；把character值转换为其他字符串类型之一时，尾随空格也会被移除。请注意，在character varying和text值中，尾随空格确实具有语义意义。

短字符串（最长 126 字节）的存储需求是 1 个字节，再加上实际字符串本身；对 character 而言，这其中还包括填充的空格。更长的字符串则需要 4 个字节的额外开销，而不是 1 个字节。长字符串会被系统自动压缩，因此在磁盘上的实际占用可能更小。非常长的值还会被存储在后台表中，以免影响对较短列值的快速访问。无论如何，可存储的最长字符串大约为 1 GB。（数据类型声明中 *n* 允许的最大值比这还小。更改这个限制并没有意义，因为在多字节字符编码下，字符数和字节数可能差异很大。如果你想存储没有明确上限的长字符串，应使用 text 或未指定长度的 character varying，而不是随意给出一个长度上限。）

> **提示**
>
> 这三种类型之间没有性能差别，除了使用空格填充类型时会占用更多存储空间，以及在写入带长度约束的列时需要少量额外周期来检查长度。
> 虽然在某些其他数据库系统中，character(*n*) 可能有一定性能优势，但在 PostgreSQL 中并不存在这种优势。在大多数情况下，应优先使用 text 或 character varying。

关于字符串常量的语法，请参见第 4.1.2.1 节；关于可用的操作符和函数，请参见第 9 章。数据库字符集决定了用于存储文本值的字符集；有关字符集支持的更多信息，请参阅第 21.2 节。

例 8.1. 使用字符类型

```
CREATE TABLE test1 (a character(4));
```

```
INSERT INTO test1 VALUES ('ok');
SELECT a, char_length(a) FROM test1; -- ❶
  a   | char_length
------+-------------
 ok   |           2

CREATE TABLE test2 (b varchar(5));
INSERT INTO test2 VALUES ('ok');
INSERT INTO test2 VALUES ('good      ');
INSERT INTO test2 VALUES ('too long');
ERROR:  value too long for type character varying(5)
INSERT INTO test2 VALUES ('too long'::varchar(5)); -- explicit
 truncation
SELECT b, char_length(b) FROM test2;
   b   | char_length
-------+-------------
 ok    |           2
 good  |           5
 too l |           5
```

❶ 函数`char_length`在第 9.4 节中讨论。

PostgreSQL 中还有两种固定长度字符类型，如表 8.5所示。`name` 类型仅用于在内部系统目录中存储标识符，并非供一般用户使用。它的长度目前定义为 64 字节（63 个可用字符加结束符），但应使用常量 `NAMEDATALEN` 来引用。这个长度是在编译时设定的（因此可以针对特殊用途调整）；默认最大长度在未来版本中可能会变化。类型 `"char"`（注意带引号）不同于 `char(1)`，因为它只使用 1 个字节存储。它在系统目录中被用作一种简单的枚举类型。

表 8.5. 特殊字符类型

名字	存储大小	描述
`"char"`	1字节	单字节内部类型
`name`	64字节	用于对象名的内部类型

8.4. 二进制数据类型

`bytea`数据类型允许存储二进制串，参见表 8.6。

表 8.6. 二进制数据类型

名字	存储尺寸	描述
`bytea`	1或4字节外加真正的二进制串	变长二进制串

二进制串是八位组（或字节）的序列。二进制串与字符串有两点区别。首先，
二进制串明确允许存储值为零的字节以及其他“不可打印”的字节（通常指 32 到 126 范围之外的字节）。而字符串不允许零字节，
也不允许那些按数据库所选字符集编码看属于非法的其他字节值或字节序列。其次，
对二进制串的操作处理的是实际字节，而字符串的处理则取决于区域设置。简单地说，
二进制串适合存储程序员视为“裸字节”的数据，而字符串适合存储文本。

在SQL语句中把`bytea`值用作字符串常量的一部分时，某些值的字节
必须转义（但所有字节值都可以转义）。一般而言，要转义一个字节，

需要把它转换为三位八进制值并在前面加上两个反斜线。表 8.7展示了必须转义的字符，并给出了适用的替代转义序列。

表 8.7. bytea输入转义字节

十进制字节值	描述	转义输入表示	示例	输出表示
0	零字节	'\\000'	SELECT '\\000'::bytea;	\000
39	单引号	'\''或'\\047'	SELECT '\''::bytea;	'
92	反斜线	'\\\\'或'\\134'	SELECT '\\\\'::bytea;	\\
0 到 31 以及 127 到 255	“不可打印”字节	'*xxx*'（八进制值）	SELECT '\\001'::bytea;	\001

是否必须转义这些“不可打印”字节，会因区域设置不同而有所差异。在某些情况下，你可以不转义它们。请注意，表 8.7中每个示例的结果都恰好是一个字节长，尽管零字节和反斜线的输出表示多于一个字符。

表 8.7中之所以需要多个反斜线，是因为写成字符串常量的输入串在PostgreSQL服务器中必须经过两个解析阶段。每一对中的第一个反斜线会被字符串常量解析器解释为转义字符并因此被消耗，留下该对中的第二个反斜线。剩下的反斜线随后被bytea输入函数识别为三位八进制值的开始，或者对另一个反斜线的转义。例如，以'\\001'形式传给服务器的字符串常量，在经过字符串常量解析器之后变成\001。然后\001被送到bytea输入函数，在那里它被转换为十进制值为 1 的单个字节。注意，撇号字符不会被bytea特殊对待，因此它遵循字符串常量的正常规则。（另见第 4.1.2.1 节。）

Bytea字节在输出时也会被转义。通常，每个“不可打印”字节都会被转换为等价的三位八进制值并冠以一个反斜线。多数“可打印”字节按其在客户端字符集中的标准表示输出。十进制值为 92（反斜线）的字节有一种特殊的替代输出表示。细节见表 8.8。

表 8.8. bytea输出转义字节

十进制字节值	描述	转义输出表示	示例	输出结果
92	反斜线	\\	SELECT '\\134'::bytea;	\\
0 到 31 以及 127 到 255	“不可打印”字节	*xxx*（八进制值）	SELECT '\\001'::bytea;	\001
32 到 126	“可打印”字节	客户端字符集表示	SELECT '\\176'::bytea;	~

取决于你所使用的 PostgreSQL 前端，在转义和反转义 bytea 串时可能还需要做额外工作。例如，如果你的接口会自动转换换行和回车，那么你可能还需要对它们进行转义。

SQL标准定义了一种不同的二进制串类型，叫做BLOB或者BINARY LARGE OBJECT。其输入格式和bytea不同，但是提供的函数和操作符大多一样。

8.5. 日期/时间类型

PostgreSQL支持完整的一组SQL日期和时间类型，如表 8.9所示。这些数据类型上可用的操作见第 9.9 节。

表 8.9. 日期/时间类型

名字	存储大小	描述	低值	高值	精度
timestamp [(*p*)] [without time zone]	8字节	日期和时间	4713 BC	5874897 AD	1微秒 / 14位
timestamp [(*p*)] with time zone	8字节	日期和时间，带时区	4713 BC	5874897 AD	1微秒 / 14位
interval [(*p*)]	12字节	时间间隔	-178000000年	178000000年	1微秒 / 14位
date	4字节	仅日期	4713 BC	5874897 AD	1日
time [(*p*)] [without time zone]	8字节	仅一天中的时间	00:00:00	24:00:00	1微秒 / 14位
time [(*p*)] with time zone	12字节	仅一天中的时间，带时区	00:00:00+1359	24:00:00-1359	1微秒 / 14位

> **注意**
>
> 在 PostgreSQL 7.3 之前的版本中，仅写timestamp等效于timestamp with time zone。为了符合 SQL 标准，这一行为已被改变。

time、timestamp 和 interval 都接受一个可选精度值 *p*，用来指定在秒字段中保留多少位小数。默认情况下，对精度没有显式上界。对于 timestamp 和 interval 类型，*p* 的允许范围是 0 到 6。

> **注意**
>
> 当 timestamp 值存储为双精度浮点数（目前的默认方式）时，有效精度上界可能小于 6。timestamp 值存储为相对于 2000-01-01 午夜之前或之后的秒数。对于距 2000-01-01 几年之内的日期可以达到微秒精度，但对更远的日期精度会下降。而当 timestamp 值存储为八字节整数（一个编译期选项）时，在整个取值范围内都可以得到微秒精度。不过，八字节整数时间戳的日期范围比上表所列更有限：从公元前 4713 年到公元 294276 年。同一个编译期选项还决定 time 和 interval 值存储为浮点数还是八字节整数。在浮点存储的情况下，大的 interval 值会随着间隔长度的增大而降低精度。

对于 time 类型，采用八字节整数存储时，*p* 的允许范围是 0 到 6；采用浮点存储时，允许范围是 0 到 10。

time with time zone 类型由 SQL 标准定义，但其定义具有一些会让人怀疑其实用性的特性。在大多数情况下，date、time、timestamp without time zone 和 timestamp with time zone 的组合，就足以提供任何应用所需的完整日期/时间功能。

类型abstime和reltime是内部使用的低精度类型。不推荐在应用中使用这些类型；
这些内部类型可能在将来的版本中消失。

8.5.1. 日期/时间输入

日期和时间输入几乎接受任何合理的格式，包括 ISO 8601、与 SQL 兼容的格式、传统 POSTGRES 格式等。对于某些格式，日期输入中日、月、年的顺序可能存在歧义，因此支持指定这些字段的预期顺序。将DateStyle参数设置为 `MDY`，表示采用月-日-年解释；设置为 `DMY` 表示采用日-月-年解释；而 `YMD` 表示采用年-月-日解释。

PostgreSQL 在处理日期/时间输入方面，比 SQL 标准要求的更灵活。有关日期/时间输入的精确解析规则，以及可识别的文本字段（包括月份、星期几和时区），请参见附录 B。

请记住，任何日期或时间字面值输入都必须像文本字符串一样用单引号括起来。
更多信息请参见第 4.1.2.5 节。SQL 要求使用下列语法：

```
type [ (p) ] 'value'
```

其中 *p* 是可选的精度说明，给出秒字段中保留的小数位数。精度可用于 `time`、`timestamp` 和 `interval` 类型，允许的取值范围见上文。如果在常量声明中没有指定精度，则默认采用该字面值本身的精度。

8.5.1.1. 日期

表 8.10显示了`date`类型可能的输入方式。

表 8.10. 日期输入

示例	描述
1999-01-08	ISO 8601；任何模式下的1月8日（推荐格式）
January 8, 1999	在任何`datestyle`输入模式下都无歧义
1/8/1999	`MDY`模式中的1月8日；`DMY`模式中的8月1日
1/18/1999	`MDY`模式中的1月18日；在其他模式中被拒绝
01/02/03	`MDY`模式中的2003年1月2日；`DMY`模式中的2003年2月1日；`YMD`模式中的2001年2月3日
1999-Jan-08	任何模式下的1月8日
Jan-08-1999	任何模式下的1月8日
08-Jan-1999	任何模式下的1月8日
99-Jan-08	`YMD`模式中的1月8日，否则错误
08-Jan-99	1月8日，除了在`YMD`模式中错误
Jan-08-99	1月8日，除了在`YMD`模式中错误
19990108	ISO 8601；任何模式中的1999年1月8日
990108	ISO 8601；任何模式中的1999年1月8日
1999.008	年和一年中的日子
J2451187	儒略日
January 8, 99 BC	公元前99年

8.5.1.2. 时间

一天中的时间类型包括 `time [ (p) ] without time zone` 和 `time [ (p) ] with time zone`。单独写 `time` 等效于 `time without time zone`。

这些类型的有效输入由一个一天中的时间，加上一个可选时区组成（见表 8.11和表 8.12）。如果在 `time without time zone` 的输入中指定了时区，它会被静默忽略。

表 8.11. 时间输入

示例	描述
`04:05:06.789`	ISO 8601
`04:05:06`	ISO 8601
`04:05`	ISO 8601
`040506`	ISO 8601
`04:05 AM`	和04:05一样，AM并不影响值
`04:05 PM`	和16:05一样，输入的小时必须为 <= 12
`04:05:06.789-8`	ISO 8601
`04:05:06-08:00`	ISO 8601
`04:05-08:00`	ISO 8601
`040506-08`	ISO 8601
`04:05:06 PST`	按名称指定的时区

表 8.12. 时区输入

示例	描述
`PST`	太平洋标准时间
`-8:00`	PST的ISO-8601偏移
`-800`	PST的ISO-8601偏移
`-8`	PST的ISO-8601偏移
`zulu`	UTC的军方缩写
`z`	`zulu`的缩写形式

可识别的输入时区名称列表请参阅附录 B。

8.5.1.3. 时间戳

时间戳类型的有效输入由一个日期和时间的串接组成，后面跟着一个可选时区，以及一个可选的 `AD` 或 `BC`（另外，`AD/BC` 也可以出现在时区前面，但这种顺序并不推荐）。因此：

```
1999-01-08 04:05:06
```

和：

```
1999-01-08 04:05:06 -8:00
```

都是遵循 ISO 8601 标准的有效值。另外，广泛使用的下列格式：

```
January 8 04:05:06 1999 PST
```

也被支持。

按照SQL标准，`timestamp without time zone`和`timestamp with time zone`字面量的区别在于，时间后是否有 “+” 或 “-” 符号及其后的时区偏移。因此，按照该标准，

```
TIMESTAMP '2004-10-19 10:23:54'
```

是timestamp without time zone，而

```
TIMESTAMP '2004-10-19 10:23:54+02'
```

是timestamp with time zone。PostgreSQL在确定字符串字面量的类型之前，从不检查其内容，因此会把上述两者都视为timestamp without time zone。为确保字面量被视为timestamp with time zone，应为它显式指定正确类型：

```
TIMESTAMP WITH TIME ZONE '2004-10-19 10:23:54+02'
```

若字面量已经被确定为timestamp without time zone，PostgreSQL会静默忽略任何时区标记。也就是说，所得值来自输入值中的日期/时间字段，不会根据时区调整。

对于timestamp with time zone，内部存储的值始终使用 UTC（协调世界时，传统上称为格林尼治标准时间，GMT）。对于显式指定时区的输入值，会使用该时区适当的偏移将它转换为 UTC。如果输入字符串中没有说明时区，则假定它属于系统timezone参数指定的时区，并使用timezone时区的偏移将它转换为 UTC。

当输出一个 timestamp with time zone 值时，它总会从 UTC 转换到当前 timezone 时区，并显示为该时区的本地时间。若要查看其他时区的时间，可以修改 timezone，或者使用 AT TIME ZONE 构造（见第 9.9.3 节）。

在 timestamp without time zone 和 timestamp with time zone 之间转换时，通常假定 timestamp without time zone 值应被解释为，或输出为，timezone 本地时间。要为该转换指定不同的时区，可以使用 AT TIME ZONE。

8.5.1.4. 间隔

interval值可以用下列语法书写：

[@] *quantity unit* [*quantity unit*...] [*direction*]

其中：*quantity*是一个数字（可带符号）；*unit*是microsecond、millisecond、second、minute、hour、day、week、month、year、decade、century、millennium，或者这些单位的缩写或复数形式；*direction*可以是ago或为空。at 符号（@）是可选的噪音。不同单位的数量会带着适当的符号被隐式地累加起来。

天、小时、分钟和秒的数量可以不带显式的单位标记来指定。例如，'1 12:59:10'的读法与 '1 day 12 hours 59 min 10 sec'相同。

可选的亚秒精度*p*应当在 0 到 6 之间，默认为输入字面量的精度。

在内部，interval值以月、天和秒的形式存储。这样做是因为一个月的天数不定，而且在涉及夏令时调整时一天可能有 23 或 25 个小时。由于间隔通常由常量字符串或 timestamp相减创建，这种存储方法在大多数情况下都能良好工作。可以使用函数justify_days和 justify_hours来调整超出其正常周期的天和小时。

8.5.1.5. 特殊值

为了方便起见，PostgreSQL 支持若干特殊的日期/时间输入值，如表 8.13所示。infinity 和 -infinity在系统内部有特殊表示，并且输出时会保持不变；其余值则只是记法上的简写，在读取时会被转换成普通日期/时间值。（特别是，now 及相关字符串在被读取后会立刻转换成某个特定的时间值。）这些值在 SQL 命令中作为常量使用时，都必须用单引号括起来。

表 8.13. 特殊日期/时间输入

输入字符串	有效类型	描述
epoch	date, timestamp	1970-01-01 00:00:00+00（Unix系统时间0）
infinity	timestamp	晚于所有其他时间戳
-infinity	timestamp	早于所有其他时间戳
now	date, time, timestamp	当前事务的开始时间
today	date、timestamp	今天午夜
tomorrow	date、timestamp	明天午夜
yesterday	date、timestamp	昨天午夜
allballs	time	00:00:00.00 UTC

以下与 SQL 兼容的函数也可用于获取相应数据类型的当前时间值：CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP、LOCALTIME 和 LOCALTIMESTAMP。后四个函数接受可选的亚秒级精度说明。（参见第 9.9.4 节。）请注意，这些是 SQL 函数，不会在日期/时间输入字符串中被识别。

8.5.2. 日期/时间输出

日期/时间类型的输出格式可以设为四种样式之一：ISO 8601、SQL（Ingres）、传统的 POSTGRES 和 German，这通过命令SET datestyle设置。默认是ISO格式。（SQL标准要求使用 ISO 8601 格式。之所以存在名为“SQL”的输出格式，只是历史原因。）表 8.14展示了各种输出样式的示例。date和time类型的输出当然只包含给定示例中相应的日期部分或时间部分。

表 8.14. 日期/时间输出风格

样式说明	描述	示例
ISO	ISO 8601/SQL 标准	1997-12-17 07:37:16-08
SQL	传统样式	12/17/1997 07:37:16.00 PST
POSTGRES	original style	Wed Dec 17 07:37:16 1997 PST
German	地区样式	17.12.1997 07:37:16.00 PST

SQL和POSTGRES风格中，如果指定了 DMY 字段顺序，“日”将出现在“月”之前，否则“月”出现在“日”之前（有关该设置如何影响输入值的解释，请参考第 8.5.1 节）。表 8.15给出了示例。

表 8.15. 日期顺序习惯

datestyle Setting	输入顺序	输出示例
SQL, DMY	*day*/*month*/*year*	17/12/1997 15:37:16.00 CET
SQL, MDY	*month*/*day*/*year*	12/17/1997 07:37:16.00 PST
Postgres, DMY	*day*/*month*/*year*	Wed 17 Dec 07:37:16 1997 PST

interval的输出与输入格式类似，但诸如 century或week这样的单位会被转换为年和天，而ago会被转换为适当的符号。在 ISO 模式下，输出形如：

```
[ quantity unit [ ... ] ] [ days ] [ hours:minutes:seconds ]
```

用户可以通过SET datestyle命令、postgresql.conf配置文件中的DateStyle参数，或者服务器端或客户端上的PGDATESTYLE环境变量来选择日期/时间样式。格式化函数to_char（见第 9.8 节）也可用，它是格式化日期/时间输出的一种更灵活的方式。

8.5.3. 时区

时区及其约定不仅受地球几何形状影响，也受政治决定影响。世界各地的时区在 20 世纪逐渐趋于标准化，但仍然容易发生任意变化，尤其是夏令时规则方面。PostgreSQL目前支持的夏令时规则覆盖 1902 到 2038 年这段时间（对应传统 Unix 系统时间的完整范围）。超出该范围的时间将被视为处于所选时区的“标准时间”，无论它们处于一年中的哪个时段。

PostgreSQL 努力在典型用法上与 SQL 标准定义保持兼容。不过，SQL 标准在日期和时间类型及其能力方面存在一些奇怪的混搭。两个显而易见的问题是：

- 尽管 `date` 类型不能有关联的时区，`time` 类型却可以。但现实世界中的时区如果不同时关联日期和时间，几乎没有意义，因为偏移量可能会随着夏令时切换而在一年中发生变化。
- 默认时区被指定为相对于 UTC 的一个固定数值偏移。因此，在跨越 DST 边界做日期/时间算术时，根本无法适应夏令时变化。

为了克服这些困难，我们建议在使用时区时采用同时包含日期和时间的日期/时间类型。我们不建议使用 `time with time zone` 类型（尽管 PostgreSQL 出于兼容旧应用以及遵循 SQL 标准的考虑而支持它）。PostgreSQL 对于任何只包含日期或时间的类型，都会假定其使用本地时区。

在系统内部，所有带时区的日期和时间都以 UTC 存储。显示给客户端之前，它们会被转换为由timezone配置参数指定的本地时间。

timezone配置参数可以在`postgresql.conf`文件中设置，也可以通过第 17 章中说明的其他标准方式设置。另外，还有一些特殊的设置方法：

- 如果没有在`postgresql.conf`中或以 postmaster 命令行开关的方式指定`timezone`，服务器会尝试使用`TZ`环境变量的值作为默认时区。如果`TZ`没有定义，或者不是PostgreSQL已知的任何时区名称，服务器会通过检查 C 库函数`localtime()`的行为来确定操作系统的默认时区。默认时区会选取PostgreSQL已知时区中最接近的匹配。
- SQL命令`SET TIME ZONE`用于设置会话的时区。它是`SET TIMEZONE TO`的另一种写法，语法上更符合 SQL 规范。
- libpq客户端使用`PGTZ`环境变量在连接时向服务器发送`SET TIME ZONE`命令。

有关可用时区的列表，请参阅附录 B。

8.5.4. 内部实现

PostgreSQL在所有日期/时间计算中使用儒略日。这带来一个有用的特性：在假定一年长度为 365.2425 天的前提下，能够正确计算从公元前 4713 年到遥远未来的日期。

19 世纪之前的日期约定读起来很有意思，但它们并不足够一致，不值得写进日期/时间处理器中。

8.6. 布尔类型

PostgreSQL提供标准的SQL类型`boolean`。`boolean`类型只能处于两种状态之一：“true”或“false”。第三种状态“unknown”由SQL空值表示。

“真”状态的有效字面值有：

```
TRUE
't'
'true'
'y'
'yes'
'1'
```

对于“假”状态，可以使用下列值：

```
FALSE
'f'
'false'
'n'
'no'
'0'
```

前导和尾随的空白会被忽略，大小写也无关紧要。关键字 `TRUE` 和 `FALSE` 是首选的（符合 SQL）用法。

例 8.2. 使用boolean类型

```
CREATE TABLE test1 (a boolean, b text);
INSERT INTO test1 VALUES (TRUE, 'sic est');
INSERT INTO test1 VALUES (FALSE, 'non est');
SELECT * FROM test1;
 a |    b
---+---------
 t | sic est
 f | non est

SELECT * FROM test1 WHERE a;
 a |    b
---+---------
 t | sic est
```

例 8.2表明，`boolean`值在输出时使用字母`t`和`f`。

`boolean`使用 1 字节存储。

8.7. 几何类型

几何数据类型表示二维的空间物体。表 8.16展示了PostgreSQL中可以用的几何类型。最基础的类型是点，它是所有其他类型的基础。

表 8.16. 几何类型

名字	存储尺寸	表示	描述
`point`	16字节	平面上的点	(x,y)
`line`	32字节	无限直线	{A,B,C}
`lseg`	32字节	有限线段	((x1,y1),(x2,y2))
`box`	32字节	矩形框	((x1,y1),(x2,y2))
`path`	16+16n字节	封闭路径（类似于多边形）	((x1,y1),...)

名字	存储尺寸	表示	描述
path	16+16n字节	开放路径	[(x1,y1),...]
polygon	40+16n字节	多边形（类似于封闭路径）	((x1,y1),...)
circle	24字节	圆	<(x,y),r>（中心点和半径）

我们提供了丰富的函数和操作符来进行各种几何操作，例如缩放、平移、旋转以及求交等，
详见第 9.10 节。

8.7.1. 点

点是几何类型的基本二维构造块。用下面任一种语法描述point类型的值：

```
( x , y )
  x , y
```

其中*x*和*y*分别是坐标，都是浮点数。

点使用第一种语法输出。

8.7.2. 线段

线段（lseg）用成对的点表示。lseg类型的值用下面任一种语法指定：

```
[ ( x1 , y1 ) , ( x2 , y2 ) ]
( ( x1 , y1 ) , ( x2 , y2 ) )
  ( x1 , y1 ) , ( x2 , y2 )
    x1 , y1   ,   x2 , y2
```

where (*x1*,*y1*) 和 (*x2*,*y2*) 是线段的两个端点。

线段使用第一种语法输出。

8.7.3. 方框

方框用其对角的点对表示。box类型的值使用下面任一种语法指定：

```
( ( x1 , y1 ) , ( x2 , y2 ) )
  ( x1 , y1 ) , ( x2 , y2 )
    x1 , y1   ,   x2 , y2
```

其中 (*x1*,*y1*) 和 (*x2*,*y2*) 是方框的对角点。

方框使用第二种语法输出。

在输入时可以提供任意两个对角，但是值将根据需要被重新排列，
并依次存储为右上角和左下角。

8.7.4. 路径

路径由一系列连接的点组成。路径可能是开放的，
也就是认为列表中第一个点和最后一个点没有被连接起来；也可能是封闭的，
这时认为第一个和最后一个点被连接起来。

path类型的值用下面任一种语法指定：

```
[ ( x1 , y1 ) , ... , ( xn , yn ) ]
( ( x1 , y1 ) , ... , ( xn , yn ) )
  ( x1 , y1 ) , ... , ( xn , yn )
  ( x1 , y1   , ... ,   xn , yn )
    x1 , y1   , ... ,   xn , yn
```

其中的点是组成路径的线段的端点。方括弧（[]）表示一个开放的路径，圆括弧（()）表示一个封闭的路径。如第三种到第五种语法所示，当最外面的圆括号被省略时，路径将被假定为封闭。

路径的输出视情况使用第一种或第二种语法。

8.7.5. 多边形

多边形由点的列表（多边形的顶点）表示。多边形与封闭路径非常相似，但存储方式不同，并且有自己的支持例程。

polygon类型的值用下列任一种语法指定：

```
( ( x1 , y1 ) , ... , ( xn , yn ) )
  ( x1 , y1 ) , ... , ( xn , yn )
  ( x1 , y1   , ... ,   xn , yn )
    x1 , y1   , ... ,   xn , yn
```

其中的点是组成多边形边界的线段的端点。

多边形的输出使用第一种语法。

8.7.6. 圆

圆由一个圆心和一个半径代表。circle类型的值用下面任一种语法指定：

```
< ( x , y ) , r >
( ( x , y ) , r )
  ( x , y ) , r
    x , y   , r
```

其中(x,y)是圆心，而r是圆的半径。

圆的输出用第一种语法。

8.8. 网络地址类型

PostgreSQL 提供了用于存储 IPv4、IPv6 和 MAC 地址的数据类型，如表 8.17所示。使用这些数据类型来存储网络地址，比使用纯文本类型更好，因为它们提供了输入错误检查以及专门的操作符和函数（见第 9.11 节）。

表 8.17. 网络地址类型

名字	存储尺寸	描述
cidr	7或19字节	IPv4和IPv6网络
inet	7或19字节	IPv4和IPv6主机以及网络
macaddr	8字节	MAC地址

在对 inet 或 cidr 数据类型排序时，IPv4 地址总是排在 IPv6 地址之前，即使后者是封装或映射 IPv4 地址所得的 IPv6 地址，例如 ::10.2.3.4 或 ::ffff:10.4.3.2。

8.8.1. inet

inet 在一个字段中保存 IPv4 或 IPv6 主机地址，以及可选的子网信息。子网由主机地址中包含的网络地址位数（即“网络掩码”）表示。如果网络掩码为 32 且地址是 IPv4，那么该值表示的不是子网，而只是单个主机。在 IPv6 中，地址长度为 128 位，因此 128 位表示一个唯一的主机地址。请注意，如果你只想接受网络地址，应使用 cidr 类型，而不是 inet。

该类型的输入格式为 *address/y*，其中 *address* 是 IPv4 或 IPv6 地址，而 *y* 是网络掩码中的位数。如果省略 */y* 部分，则 IPv4 的网络掩码取 32，IPv6 的网络掩码取 128，因此该值表示单个主机。显示时，如果 */y* 部分表示的是单个主机，则它不会被显示出来。

8.8.2. cidr

cidr 类型保存 IPv4 或 IPv6 网络说明。输入和输出格式遵循无类别域间路由（CIDR）约定。指定网络的格式为 *address/y*，其中 *address* 是网络的 IPv4 或 IPv6 表示，而 *y* 是网络掩码中的位数。如果省略 *y*，则会按照旧式的分类网络编号系统假设来计算，但至少会大到足以覆盖输入中写出的所有八位组。若指定的网络地址在所给网络掩码右侧仍有置位比特，则会报错。

表 8.18展示了一些示例。

表 8.18. cidr类型输入示例

cidr输入	cidr输出	abbrev(cidr)
192.168.100.128/25	192.168.100.128/25	192.168.100.128/25
192.168/24	192.168.0.0/24	192.168.0/24
192.168/25	192.168.0.0/25	192.168.0.0/25
192.168.1	192.168.1.0/24	192.168.1/24
192.168	192.168.0.0/24	192.168.0/24
128.1	128.1.0.0/16	128.1/16
128	128.0.0.0/16	128.0/16
128.1.2	128.1.2.0/24	128.1.2/24
10.1.2	10.1.2.0/24	10.1.2/24
10.1	10.1.0.0/16	10.1/16
10	10.0.0.0/8	10/8
10.1.2.3/32	10.1.2.3/32	10.1.2.3/32
2001:4f8:3:ba::/64	2001:4f8:3:ba::/64	2001:4f8:3:ba::/64
2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1
::ffff:1.2.3.0/120	::ffff:1.2.3.0/120	::ffff:1.2.3/120
::ffff:1.2.3.0/128	::ffff:1.2.3.0/128	::ffff:1.2.3.0/128

8.8.3. inet 与 cidr

inet 和 cidr 两种数据类型的本质区别在于：inet 接受网络掩码右侧仍含有非零位的值，而 cidr 不接受。

> **提示**
>
> 如果你不喜欢 `inet` 或 `cidr` 值的输出格式，可以试试 `host`、`text` 和 `abbrev` 函数。

8.8.4. macaddr

`macaddr`类型存储 MAC 地址，例如以太网卡硬件地址（尽管 MAC 地址还用于其它用途）。可以接受下列格式的输入：

```
'08002b:010203'
'08002b-010203'
'0800.2b01.0203'
'08-00-2b-01-02-03'
'08:00:2b:01:02:03'
```

这些示例指定的都是同一个地址。十六进制数字 `a` 到 `f` 可使用大小写任意形式。输出总是采用上面展示的最后一种形式。

PostgreSQL源码分发中的 `contrib/mac`目录包含一些工具，可用于把 MAC 地址映射到硬件制造商名称。

8.9. 位串类型

位串是由 1 和 0 组成的字符串。它们可用于存储或展示位掩码。SQL 中有两种位类型：`bit(n)` 和 `bit varying(n)`，其中 *n* 是正整数。

`bit` 类型的数据长度必须与 *n* 完全一致；试图存储更短或更长的位串都会报错。`bit varying` 数据则是长度上限为 *n* 的变长类型，更长的串会被拒绝。不带长度的 `bit` 等效于 `bit(1)`，而不带长度说明的 `bit varying` 表示长度不受限。

> **注意**
>
> 如果显式地将一个位串值转换为 `bit(n)`，它会在右侧被截断或补零，直到长度恰好为 *n* 位，且不会报错。类似地，如果显式地将一个位串值转换为 `bit varying(n)`，而其长度超过 *n* 位，则会在右侧被截断。

> **注意**
>
> 在PostgreSQL 7.2 之前，`bit`数据总是被静默地在右侧截断或补零，无论是否显式转换。为遵循SQL标准，这一行为已被改变。

关于位串常量的语法信息，请参见第 4.1.2.3 节。位逻辑操作符和字符串操作函数也可用；参见第 9.6 节。

例 8.3. 使用位串类型

```
CREATE TABLE test (a BIT(3), b BIT VARYING(5));
```

```
INSERT INTO test VALUES (B'101', B'00');
INSERT INTO test VALUES (B'10', B'101');
ERROR:  bit string length 2 does not match type bit(3)
INSERT INTO test VALUES (B'10'::bit(3), B'101');
SELECT * FROM test;
  a  |  b
-----+-----
 101 | 00
 100 | 101
```

8.10. 数组

PostgreSQL 允许将表列定义为变长多维数组。可以创建任何内置或用户定义的基础类型、枚举类型或复合类型的数组。尚不支持域的数组。

8.10.1. 数组类型的声明

为了说明数组类型的用法，我们创建下面这个表：

```
CREATE TABLE sal_emp (
    name            text,
    pay_by_quarter  integer[],
    schedule        text[][]
);
```

如上所示，数组数据类型通过在数组元素的数据类型名后附加方括号（`[]`）来命名。
上述命令将创建一个名为 `sal_emp` 的表，其中有一个 `text` 类型的列（`name`）、一个表示员工各季度薪资的一维 `integer` 数组（`pay_by_quarter`），以及一个表示员工每周日程安排的二维 `text` 数组（`schedule`）。

`CREATE TABLE` 的语法允许指定数组的确切大小，例如：

```
CREATE TABLE tictactoe (
    squares   integer[3][3]
);
```

不过，当前实现会忽略给出的任何数组大小限制，也就是说，其行为与未指定长度的数组相同。

当前实现也不会强制执行所声明的维度数。对于某一特定元素类型的数组，
无论大小或维度数如何，都会被视为同一种类型。因此，在 `CREATE TABLE` 中声明数组大小或维度数仅仅是文档说明；它不会影响运行时行为。

另一种符合 SQL 标准的语法可用于一维数组。`pay_by_quarter` 也可以定义为：

```
    pay_by_quarter  integer ARRAY[4],
```

这种语法要求用一个整数常量来表示数组大小。不过，和前面一样，PostgreSQL 并不强制执行大小限制。

8.10.2. 数组值输入

要把数组值写成字面常量，请将元素值放在花括号内，并用逗号分隔。（如果你了解 C，这与 C 中初始化结构体的语法有些类似。）你可以给任意元素值加上双引号；如果它包含逗号或花括号，则必须这样做。（更多细节见下文。）因此，数组常量的一般格式如下：

```
'{ val1 delim val2 delim ... }'
```

其中 *delim* 是该类型的分隔符字符，它记录在其 pg_type 条目中。在 PostgreSQL 发行版提供的标准数据类型里，除类型 box 使用分号（;）外，其余都使用逗号（,）。每个 *val* 要么是数组元素类型的常量，要么是一个子数组。数组常量的一个示例是：

```
'{{1,2,3},{4,5,6},{7,8,9}}'
```

这个常量是一个 3 x 3 的二维数组，由三个整数子数组构成。

（这类数组常量实际上只是第 4.1.2.5 节中讨论的通用类型常量的一种特例。该常量最初会被当作字符串处理，然后传递给数组输入转换例程。可能需要显式指定类型。）

现在我们来看几个 INSERT 语句。

```
INSERT INTO sal_emp
    VALUES ('Bill',
    '{10000, 10000, 10000, 10000}',
    '{{"meeting", "lunch"}, {"meeting"}}');
ERROR:  multidimensional arrays must have array expressions with
 matching dimensions
```

注意，多维数组的各个子数组在每个维度上的长度必须匹配。不匹配会导致一条错误报告。

```
INSERT INTO sal_emp
    VALUES ('Bill',
    '{10000, 10000, 10000, 10000}',
    '{{"meeting", "lunch"}, {"training", "presentation"}}');

INSERT INTO sal_emp
    VALUES ('Carol',
    '{20000, 25000, 25000, 25000}',
    '{{"breakfast", "consulting"}, {"meeting", "lunch"}}');
```

前两条插入语句的结果如下：

```
SELECT * FROM sal_emp;
 name  |      pay_by_quarter       |                  schedule
-------+---------------------------+-----------------------------------
---------
 Bill  | {10000,10000,10000,10000} | {{meeting,lunch},{training,
presentation}}
 Carol | {20000,25000,25000,25000} | {{breakfast,consulting},{meeting,
lunch}}
(2 rows)
```

当前数组实现的一个局限是，数组中的单个元素不能是 SQL 空值。整个数组可以被设置为 null，但不能让数组的一部分元素为 null 而另一部分不为 null。（将来这可能会改变。）

也可以使用 ARRAY 构造器语法：

```
INSERT INTO sal_emp
    VALUES ('Bill',
    ARRAY[10000, 10000, 10000, 10000],
    ARRAY[['meeting', 'lunch'], ['training', 'presentation']]);
```

```
INSERT INTO sal_emp
    VALUES ('Carol',
    ARRAY[20000, 25000, 25000, 25000],
    ARRAY[['breakfast', 'consulting'], ['meeting', 'lunch']]);
```

注意，数组元素是普通的 SQL 常量或表达式；例如，字符串字面值用单引号而不是双引号包围，而在数组字面量中则会用双引号。关于 `ARRAY` 构造器语法的更多细节见第 4.2.10 节。

8.10.3. 访问数组

现在，我们可以对该表执行一些查询。首先，演示如何访问数组中的单个元素。
下面的查询取回第二季度薪资发生变化的员工姓名：

```
SELECT name FROM sal_emp WHERE pay_by_quarter[1] <> pay_by_quarter[2];

 name
-------
 Carol
(1 row)
```

数组下标写在方括号内。默认情况下，PostgreSQL 对数组采用从 1 开始的编号约定，也就是说，一个有 *n* 个元素的数组从 `array[1]` 开始，到 `array[n]` 结束。

下面这个查询取回所有员工第三季度的薪资：

```
SELECT pay_by_quarter[3] FROM sal_emp;

 pay_by_quarter
----------------
          10000
          25000
(2 rows)
```

我们还可以访问数组的任意矩形切片，即子数组。数组切片通过在一个或多个数组维度上写成 *lower-bound*:*upper-bound* 的形式来表示。例如，下面这个查询取回 Bill 在一周中前两天里每天的首项日程：

```
SELECT schedule[1:2][1:1] FROM sal_emp WHERE name = 'Bill';

        schedule
------------------------
 {{meeting},{training}}
(1 row)
```

我们也可以写成

```
SELECT schedule[1:2][1] FROM sal_emp WHERE name = 'Bill';
```

得到相同的结果。只要任何一个下标写成 *lower*:*upper* 的形式，数组下标操作就总是被当作数组切片来处理。对于只指定了一个数值的下标，会假定其下界为 1，例如：

```
SELECT schedule[1:2][2] FROM sal_emp WHERE name = 'Bill';

                 schedule
-------------------------------------------
```

```
 {{meeting,lunch},{training,presentation}}
(1 row)
```

从数组当前边界之外取值会得到一个 SQL 空值，而不是错误。例如，如果 `schedule` 当前的维度是 `[1:3][1:2]`，那么引用 `schedule[3][3]` 会得到 NULL。类似地，使用错误数量的下标访问数组，得到的也是空值而不是错误。
取回一个完全位于当前边界之外的数组切片同样会得到一个 null 数组；但如果所请求的切片与数组边界仅部分重叠，那么它会被静默缩减为重叠区域。

任意数组值的当前维度都可以用 `array_dims` 函数取回：

```
SELECT array_dims(schedule) FROM sal_emp WHERE name = 'Carol';

 array_dims
------------
 [1:2][1:2]
(1 row)
```

`array_dims` 产生的是一个 `text` 结果，这对人工阅读比较方便，但对程序来说可能不太方便。也可以用 `array_upper` 和 `array_lower` 取回维度信息，它们分别返回指定数组维度的上界和下界：

```
SELECT array_upper(schedule, 1) FROM sal_emp WHERE name = 'Carol';

 array_upper
-------------
           2
(1 row)
```

8.10.4. 修改数组

数组值可以被整体替换：

```
UPDATE sal_emp SET pay_by_quarter = '{25000,25000,27000,27000}'
    WHERE name = 'Carol';
```

或者使用 `ARRAY` 表达式语法：

```
UPDATE sal_emp SET pay_by_quarter = ARRAY[25000,25000,27000,27000]
    WHERE name = 'Carol';
```

也可以更新数组中的单个元素：

```
UPDATE sal_emp SET pay_by_quarter[4] = 15000
    WHERE name = 'Bill';
```

或者更新其中一个切片：

```
UPDATE sal_emp SET pay_by_quarter[1:2] = '{27000,27000}'
    WHERE name = 'Carol';
```

已存储的数组值可以通过给与已有元素相邻的元素赋值，
或者给与已有数据相邻或重叠的切片赋值来扩展。例如，如果数组 `myarray` 当前有 4 个元素，那么在一次更新把值赋给 `myarray[5]` 之后，它将有 5 个元素。目前，以这种方式扩展只允许用于一维数组，不允许用于多维数组。

带下标的赋值也允许创建不使用从 1 开始下标的数组。例如，可以给 `myarray[-2:7]` 赋值，从而创建一个下标值范围为 -2 到 7 的数组。

新的数组值也可以使用连接操作符 `||` 构造：

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

SELECT ARRAY[5,6] || ARRAY[[1,2],[3,4]];
      ?column?
---------------------
 {{5,6},{1,2},{3,4}}
(1 row)
```

连接操作符允许把单个元素添加到一维数组的开头或末尾。它也接受两个 *N* 维数组，或者一个 *N* 维数组与一个 *N+1* 维数组。

当把单个元素添加到一维数组的开头或末尾时，结果数组会保留该数组操作数的下界下标。例如：

```
SELECT array_dims(1 || '[0:1]={2,3}'::int[]);
 array_dims
------------
 [0:2]
(1 row)

SELECT array_dims(ARRAY[1,2] || 3);
 array_dims
------------
 [1:3]
(1 row)
```

当连接两个维度数相同的数组时，结果会保留左侧操作数外层维度的下界下标。结果数组由左侧操作数的所有元素后跟右侧操作数的所有元素组成。例如：

```
SELECT array_dims(ARRAY[1,2] || ARRAY[3,4,5]);
 array_dims
------------
 [1:5]
(1 row)

SELECT array_dims(ARRAY[[1,2],[3,4]] || ARRAY[[5,6],[7,8],[9,0]]);
 array_dims
------------
 [1:5][1:2]
(1 row)
```

当把一个 *N* 维数组添加到一个 *N+1* 维数组的开头或末尾时，其结果与前面“单个元素与数组相连”的情况类似。每个 *N* 维子数组本质上都是该 *N+1* 维数组外层维度中的一个元素。例如：

```
SELECT array_dims(ARRAY[1,2] || ARRAY[[3,4],[5,6]]);
 array_dims
```

```
------------
 [1:3][1:2]
(1 row)
```

也可以使用 array_prepend、array_append 或 array_cat 函数来构造数组。前两个只支持一维数组，但 array_cat 支持多维数组。注意，与直接使用这些函数相比，更推荐使用上面讨论的连接操作符。实际上，这些函数的存在主要是为了在实现连接操作符时使用。不过，在创建用户定义聚合时，它们可能也有直接的用处。示例如下：

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

SELECT array_append(ARRAY[1,2], 3);
 array_append
--------------
 {1,2,3}
(1 row)

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

SELECT array_cat(ARRAY[[1,2],[3,4]], ARRAY[5,6]);
      array_cat
---------------------
 {{1,2},{3,4},{5,6}}
(1 row)

SELECT array_cat(ARRAY[5,6], ARRAY[[1,2],[3,4]]);
      array_cat
---------------------
 {{5,6},{1,2},{3,4}}
```

8.10.5. 在数组中搜索

要在数组中搜索某个值，就必须检查每一个值。如果你知道数组的大小，这可以手工完成。例如：

```
SELECT * FROM sal_emp WHERE pay_by_quarter[1] = 10000 OR
                            pay_by_quarter[2] = 10000 OR
                            pay_by_quarter[3] = 10000 OR
                            pay_by_quarter[4] = 10000;
```

不过，对于大型数组，这样很快就会变得繁琐；而在数组大小未知时，这也没有帮助。另一种方法见第 9.17 节。上面的查询可以改写为：

```
SELECT * FROM sal_emp WHERE 10000 = ANY (pay_by_quarter);
```

此外，如果要查找数组中所有值都等于 10000 的行，可以使用：

```
SELECT * FROM sal_emp WHERE 10000 = ALL (pay_by_quarter);
```

> **提示**
>
> 数组不是集合；搜索特定数组元素可能是数据库设计不当的信号。
> 可以考虑使用一个独立的表，让原本会成为数组元素的每个项各占一行。
> 这样会更容易搜索，而且在元素数量很多时通常也有更好的可伸缩性。

8.10.6. 数组输入和输出语法

数组值的外部文本表示由若干项构成，这些项会按照数组元素类型的 I/O 转换规则进行解释，再加上一些表示数组结构的修饰。修饰部分包括数组值外层的花括号（`{` 和 `}`），以及相邻项之间的分隔符字符。分隔符字符通常是逗号（`,`），但也可能是别的字符：它由数组元素类型的 `typdelim` 设置决定。在 PostgreSQL 发行版提供的标准数据类型中，除类型 `box` 使用分号（`;`）外，其余都使用逗号。在多维数组中，每个维度（行、平面、立方体等）都有自己的一层花括号，并且同一级别相邻的、由花括号括起的各项之间必须写出分隔符。

如果元素值是空字符串、包含花括号、分隔符字符、双引号、反斜线或空白，
数组输出例程就会在元素值外面加上双引号。元素值中嵌入的双引号和反斜线会用反斜线转义。
对于数字数据类型，可以安全地假定不会出现双引号；但对于文本数据类型，
则应准备好处理带引号和不带引号两种情况。（这是相对 7.2 之前 PostgreSQL 版本的一个行为变化。）

默认情况下，数组各维度的下界下标值都设为 1。要表示具有其他下界的数组，可以在写出数组内容之前显式指定数组下标范围。这种修饰由包围每个数组维度上下界的方括号（`[]`）构成，
中间以冒号（`:`）作为分隔符字符。数组维度修饰后面再跟一个等号（`=`）。例如：

```
SELECT f1[1][-2][3] AS e1, f1[1][-1][5] AS e2
 FROM (SELECT '[1:1][-2:-1][3:5]={{{1,2,3},{4,5,6}}}'::int[] AS f1) AS
 ss;

 e1 | e2
----+----
  1 |  6
(1 row)
```

只有当一个或多个下界不等于 1 时，数组输出例程才会在结果中包含显式维度信息。

这种语法也可以用来在数组字面值中指定非默认的数组下标。例如：

```
SELECT f1[1][-2][3] AS e1, f1[1][-1][5] AS e2
 FROM (SELECT '[1:1][-2:-1][3:5]={{{1,2,3},{4,5,6}}}'::int[] AS f1) AS
 ss;

 e1 | e2
----+----
  1 |  6
(1 row)
```

如前所述，在写数组值时，可以给任意单个数组元素加上双引号。
如果元素值不加引号就会让数组值解析器产生混淆，那么你必须这样做。例如，包含花括号、
逗号（或者该数据类型的分隔符字符）、双引号、反斜线，或者前后带有空白的元素值，
必须使用双引号。要在一个带双引号的数组元素值中放入双引号或反斜线，

需要在它前面加反斜线。或者，
你也可以用反斜线转义来保护所有原本会被当作数组语法的数据字符。

你可以在左花括号之前或右花括号之后添加空白。
也可以在任意单个项字符串之前或之后添加空白。在所有这些情况下，这些空白都会被忽略。
不过，双引号元素内部的空白，或者元素中夹在两个非空白字符之间的空白，不会被忽略。

注意

请记住，你在 SQL 命令中写下的内容会先被解释为字符串字面值，然后才被解释为数组。这会使你所需的反斜线数量加倍。例如，要插入一个包含一个反斜线和一个双引号的 `text` 数组值，你需要写成：

```
INSERT ... VALUES ('{"\\\\","\\""}');
```

转义字符串处理器会去掉一层反斜线，因此到达数组值解析器的内容形如 `{"\\","\""}`。而传给 `text` 数据类型输入例程的字符串则分别变成 `\` 和 `"`。（如果我们使用的是输入例程同样对反斜线做特殊处理的数据类型，例如 `bytea`，那么要在命令中写入八个反斜线才能让存储的数组元素中出现一个反斜线。）可以使用美元引用（见第 4.1.2.2 节）来避免双写反斜线。

提示

在 SQL 命令中写数组值时，`ARRAY` 构造器语法（见第 4.2.10 节）通常比数组字面量语法更容易使用。在 `ARRAY` 中，各个元素值的写法与它们不作为数组成员时完全相同。

8.11. 复合类型

复合类型表示一行或一条记录的结构；本质上它就是字段名及其数据类型的列表。
PostgreSQL允许像使用简单类型那样，在许多相同的场合使用复合类型。例如，
表中的一列可以声明为复合类型。

8.11.1. 复合类型的声明

下面是两个定义复合类型的简单示例：

```
CREATE TYPE complex AS (
    r       double precision,
    i       double precision
);

CREATE TYPE inventory_item AS (
    name            text,
    supplier_id     integer,
    price           numeric
);
```

这种语法与`CREATE TABLE`类似，但只能指定字段名和类型；目前还不能包含约束（例如`NOT NULL`）。注意，`AS`关键字必不可少；如果没有它，系统会认为你想使用另一种`CREATE TYPE`命令，并产生令人费解的语法错误。

定义了这些类型之后，我们可以用它们来创建表：

```
CREATE TABLE on_hand (
    item      inventory_item,
    count     integer
);

INSERT INTO on_hand VALUES (ROW('fuzzy dice', 42, 1.99), 1000);
```

或者创建函数：

```
CREATE FUNCTION price_extension(inventory_item, integer) RETURNS
 numeric
AS 'SELECT $1.price * $2' LANGUAGE SQL;

SELECT price_extension(item, 10) FROM on_hand;
```

每当创建一张表时，也会自动创建一个与该表同名的复合类型，用来表示表的行类型。例如，假如我们执行了：

```
CREATE TABLE inventory_item (
    name            text,
    supplier_id     integer REFERENCES suppliers,
    price           numeric CHECK (price > 0)
);
```

那么，与上文相同的`inventory_item`复合类型就会随之产生，而且可以像上文那样使用。不过，请注意当前实现的一个重要限制：由于复合类型本身不关联任何约束，表定义中的约束并不适用于表之外的复合类型值。（一种部分解决办法是使用域类型作为复合类型的成员。）

8.11.2. 构造复合值

要把组合值写成字面常量，请将各字段值放在圆括号内，并用逗号分隔。
你可以给任意字段值加双引号；如果它包含逗号或圆括号，则必须这样做。
（更多细节见下文。）因此，组合常量的一般格式如下：

```
'( val1 , val2 , ... )'
```

An example is:

```
'("fuzzy dice",42,1.99)'
```

它是上面定义的`inventory_item`类型的一个合法值。要让某个字段为 NULL，在该字段在列表中的位置上什么也不写即可。例如，下面的常量指定第三个字段为 NULL：

```
'("fuzzy dice",42,)'
```

如果你想要的是空串而不是 NULL，请写双引号：

```
'("",42,)'
```

这里第一个字段是非 NULL 的空串，第三个字段是 NULL。

（这些常量实际上只是第 4.1.2.5 节中讨论的通用类型常量的一个特例。该常量最初被当作字符串处理，然后传递给组合类型的输入转换例程。有时可能需要显式指定类型。）

ROW表达式语法也可以用来构造组合值。在多数情况下，它比字符串字面量语法简单得多，因为你不必担心多层引号的问题。我们在上面已经用过这种方法：

```
ROW('fuzzy dice', 42, 1.99)
ROW('', 42, NULL)
```

只要表达式中的字段多于一个，ROW 关键字实际上可以省略，因此上面的例子可以简化为：

```
('fuzzy dice', 42, 1.99)
('', 42, NULL)
```

ROW表达式语法的更多细节见第 4.2.11 节。

8.11.3. 访问复合类型

要访问组合列中的某个字段，可以写一个点号再加字段名，这很像通过表名选取字段。实际上，它与通过表名选取字段太像了，以至于你通常必须使用圆括号，以免让解析器混淆。例如，你可能尝试从示例表on_hand中选取一些子字段：

```
SELECT item.name FROM on_hand WHERE item.price > 9.99;
```

这不会起作用，因为根据 SQL 语法规则，名称item会被当成表名，而不是字段名。你必须写成这样：

```
SELECT (item).name FROM on_hand WHERE (item).price > 9.99;
```

或者，如果你还需要使用表名（例如在多表查询中），可以这样写：

```
SELECT (on_hand.item).name FROM on_hand WHERE (on_hand.item).price >
 9.99;
```

现在，加上括号的对象就会被正确解释为对item列的引用，然后就可以从中选出子字段。

无论何时从组合值中选择字段，都会遇到类似的语法问题。例如，要从一个返回组合值的函数结果中只选取一个字段，你需要这样写：

```
SELECT (my_func(...)).field FROM ...
```

如果没有额外的圆括号，这将生成一个语法错误。

8.11.4. 修改组合值

下面是一些插入和更新组合列时正确语法的示例。先看插入或更新整个列值的情况：

```
INSERT INTO mytab (complex_col) VALUES((1.1,2.2));

UPDATE mytab SET complex_col = ROW(1.1,2.2) WHERE ...;
```

第一个示例省略了ROW，第二个示例使用了它；两种写法都可以。

我们也可以更新组合列中的单个子字段：

```
UPDATE mytab SET complex_col.r = (complex_col).r + 1 WHERE ...;
```

注意，这里不需要（事实上也不能）给紧跟在SET后面的列名加圆括号，但在等号右侧的表达式中引用同一列时，则需要加圆括号。

我们也可以把子字段指定为INSERT的目标：

```
INSERT INTO mytab (complex_col.r, complex_col.i) VALUES(1.1, 2.2);
```

如果我们没有为该列的所有子字段提供值，其余子字段就会填充为 NULL 值。

8.11.5. 复合类型的输入和输出语法

组合值的外部文本表示由两部分组成：一部分是按照各字段类型的 I/O 转换规则解释的项，另一部分是表明组合结构的附加符号。这些附加符号包括包围整个值的圆括号（(和)），以及相邻项之间的逗号（,）。圆括号外部的空白会被忽略；但在圆括号内部，空白会被视为字段值的一部分，其是否有意义取决于该字段数据类型的输入转换规则。例如，在

```
'(  42)'
```

中，如果字段类型是 integer，则空白会被忽略；如果是 text，则不会被忽略。

如前所示，在写组合值时，你可以给任意单个字段值加双引号。如果字段值本身可能让组合值解析器混淆，则必须这样做。特别是，包含圆括号、逗号、双引号或反斜杠的字段必须用双引号括起来。要在带引号的组合字段值中写入双引号或反斜杠，需要在其前面加一个反斜杠。（另外，带双引号的字段值内部成对出现的双引号会被视为一个双引号字符，这与 SQL 字面字符串中单引号的规则类似。）或者，你也可以完全不使用引号，而改用反斜杠转义，保护所有原本会被当作组合语法的数据字符。

完全空的字段值（即逗号或圆括号之间一个字符也没有）表示 NULL。要写一个空字符串值而不是 NULL，可以写成""。

如果字段值是空字符串，或者包含圆括号、逗号、双引号、反斜杠或空白字符，复合类型输出例程就会在其周围加上双引号。（对空白字符这样做并非必需，但有助于提高可读性。）嵌入字段值中的双引号和反斜杠会被加倍。

> **注意**
>
> 请记住，你在 SQL 命令中写的内容会先被解释为字符串字面量，然后再被解释为组合值。这使你需要的反斜线数量翻倍（假定使用转义字符串语法）。例如，要在组合值中插入一个包含双引号和反斜线的text字段，你需要写：
>
> ```
> INSERT ... VALUES (E'("\\\"\\\\")');
> ```
>
> 字符串字面量处理器会去掉一层反斜线，因此到达组合值解析器的内容看起来像("\"\\")。相应地，传给text数据类型输入例程的字符串变成"\。（如果我们使用的是输入例程也特殊处理反斜线的类型，例如bytea，则要在命令中写多达八个反斜线才能在存储的组合字段中得到一个反斜线。）可以使用美元引用（见第 4.1.2.2 节）来避免双写反斜线。

> **提示**
>
> 在 SQL 命令中编写组合值时，ROW构造器语法通常比组合字面量语法更容易使用。在ROW中，各个字段值的写法与它们不是组合成员时完全相同。

8.12. 对象标识符类型

对象标识符（OID）在PostgreSQL内部被用作若干系统表的主键。除非建表时指定WITH OIDS，或启用配置变量default_with_oids，否则不会向用户创建的表添加 OID。类型oid表示一个对象标识符。还有若干oid的别名类型：regproc、regprocedure、regoper、regoperator、regclass和regtype。表 8.19给出了概要说明。

oid 类型目前实现为无符号 4 字节整数。因此，在大型数据库中，它不足以提供数据库范围内的唯一性，甚至在大型的单个表中也无法保证唯一性。因此，不推荐把用户创建表的 OID 列用作主键。OID 最好仅用于引用系统表。

oid 类型本身除比较之外几乎没有其他操作。不过，它可以转换为整数，然后再用标准整数操作符进行处理。（这样做时要注意有符号与无符号可能造成的混淆。）

OID 别名类型除了专门的输入和输出例程外，没有自己的操作。这些例程能够接受并显示系统对象的符号名称，而不是 oid 类型所使用的原始数值。别名类型简化了对象 OID 值的查找。例如，要查看与表 mytable 相关的 pg_attribute 行，可以写成：

```
SELECT * FROM pg_attribute WHERE attrelid = 'mytable'::regclass;
```

而不是：

```
SELECT * FROM pg_attribute
  WHERE attrelid = (SELECT oid FROM pg_class WHERE relname =
 'mytable');
```

虽然这样看起来也不算太糟，但其实仍然过于简化。如果不同模式中有多个名为 mytable 的表，就必须写一个复杂得多的子查询来选出正确的 OID。regclass 的输入转换器会按照模式路径设置来处理表查找，因此能够自动完成“正确的事情”。同样，把表的 OID 转换成 regclass，也很适合用来以符号形式显示数值 OID。

表 8.19. 对象标识符类型

名字	引用	描述	值示例
oid	任意	数字形式的对象标识符	564182
regproc	pg_proc	函数名称	sum
regprocedure	pg_proc	函数与参数类型	sum(int4)
regoper	pg_operator	操作符名称	+
regoperator	pg_operator	带参数类型的操作符	*(integer,integer) 或-(NONE,integer)
regclass	pg_class	关系名称	pg_type
regtype	pg_type	数据类型名称	integer

所有 OID 别名类型都接受模式限定名称；如果不加限定就无法在当前搜索路径中找到对象，输出时也会显示模式限定名称。regproc和regoper别名类型只接受唯一的（未重载的）输入名称，因此用途有限；对于大多数用途，regprocedure或regoperator更合适。
对于regoperator，通过将未使用的操作数写为NONE来标识一元操作符。

大多数 OID 别名类型还有一个附加特性，就是会创建依赖关系。如果这些类型的常量出现在存储的表达式中（例如列默认表达式或视图），它就会对被引用对象建立依赖。例如，
如果某列的默认表达式是 nextval('my_seq'::regclass)，PostgreSQL 就会知道该默认表达式依赖于序列 my_seq，因此系统不允许在尚未移除该默认表达式时删除序列。

系统使用的另一种标识符类型是xid，即事务（缩写为xact）标识符。这是系统列xmin和xmax的数据类型。事务标识符是 32 位的量。

系统使用的第三种标识符类型是 cid，也就是命令标识符。这是系统列 cmin 和 cmax 的数据类型。命令标识符同样是 32 位量。

系统使用的最后一种标识符类型是 tid，也就是元组标识符（行标识符）。这是系统列 ctid 的数据类型。一个元组 ID 是一对值（块号、块内元组索引），用来标识该行在其所属表中的物理位置。

（系统列的更多说明见第 5.4 节。）

8.13. 伪类型

PostgreSQL 的类型系统中包含一些特殊用途的条目，统称为伪类型。伪类型不能作为列数据类型使用，但可以用于声明函数的参数类型或结果类型。
每一种伪类型都适用于这样的场景：函数的行为并不对应于简单地接受或返回某个特定 SQL 数据类型的值。表 8.20列出了现有的伪类型。

表 8.20. 伪类型

名字	描述
any	表示一个函数可以接受任意输入数据类型。
anyarray	表示一个函数可以接受任意数组数据类型（参见第 32.2.5 节）
anyelement	表示一个函数可以接受任意数据类型（参见第 32.2.5 节）。
cstring	表示一个函数接受或返回一个以空字符结尾的 C 字符串。
internal	表示一个函数接受或返回一个服务器内部数据类型。
language_handler	表示过程语言调用处理器被声明为返回 language_handler。
record	标识一个接受或返回未指定行类型的函数。
trigger	触发器函数被声明为返回trigger。
void	表示一个函数不返回值。
opaque	一个已废弃的类型名称，以前用于上述所有用途。

用 C 编写的函数（无论是内置的还是动态装载的）都可以声明为接受或返回这些伪类型中的任意一种。函数作者必须自行确保，当伪类型被用作参数类型时，该函数的行为仍然安全。

用过程语言编写的函数，只有在其实现语言允许的情况下才能使用伪类型。目前，
各过程语言都禁止把伪类型用作参数类型，而只允许把void和record
用作结果类型（如果函数被用作触发器，则也允许trigger）。有些语言还支持使用anyarray和anyelement类型实现多态函数。

internal 伪类型用于声明那些只应由数据库系统内部调用、而不应被 SQL 查询直接调用的函数。如果一个函数至少有一个 internal 类型参数，那么它就不能从 SQL 中调用。为了保持这一限制的类型安全性，必须遵守这样一条编码规则：除非函数至少有一个 internal 类型参数，否则不要声明任何返回 internal 的函数。

第 9 章 函数和操作符

PostgreSQL为内置数据类型提供了大量函数和操作符。用户也可以按第 V 部分中的说明定义自己的函数和操作符。psql命令`\df`和`\do`可分别用于列出所有可用的函数和操作符。

如果你关心可移植性，请注意，本章描述的大多数函数和操作符，
除了最简单的算术和比较操作符以及少数明确标记的函数之外，都不是SQL标准规定的。不过，
其中一些扩展功能也存在于其他SQL数据库管理系统中，而且在很多情况下，
各实现之间的这类功能是兼容且一致的。本章也并非详尽无遗；
手册的相关章节中还有其他函数。

9.1. 逻辑操作符

常用的逻辑操作符有：

```
AND
OR
NOT
```

SQL使用三值的逻辑系统，包括真、假和 null，它表示“未知”。观察下面的真值表：

a	*b*	*a* AND *b*	*a* OR *b*
TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	FALSE	TRUE
TRUE	NULL	NULL	TRUE
FALSE	FALSE	FALSE	FALSE
FALSE	NULL	FALSE	NULL
NULL	NULL	NULL	NULL

a	NOT *a*
TRUE	FALSE
FALSE	TRUE
NULL	NULL

操作符`AND`和`OR`是可交换的，也就是说，你可以交换左右操作数而不影响结果。但请参见第 4.2.12 节获取有关子表达式计算顺序的更多信息。

9.2. 比较操作符

常见的比较操作符都可用，如表 9.1所示。

表 9.1. 比较操作符

Operator	描述
<	小于
>	大于
<=	小于或等于
>=	大于或等于
=	等于

Operator	描述
<> or !=	not equal

注意

!= 操作符在解析器阶段会转换为 <>。不可能实现行为不同的 != 和 <> 操作符。

比较操作符适用于所有相关的数据类型。所有比较操作符都是返回 `boolean` 类型值的二元操作符；类似 `1 < 2 < 3` 的表达式是无效的（因为没有用于比较布尔值和 `3` 的 < 操作符）。

除比较操作符之外，还可以使用特殊的`BETWEEN`结构：

```
a BETWEEN x AND y
```

等价于

```
a >= x AND a <= y
```

类似地，

```
a NOT BETWEEN x AND y
```

等价于

```
a < x OR a > y
```

除了把第一种形式在内部改写成第二种形式所需的CPU周期之外，两种形式之间没有任何区别。`BETWEEN SYMMETRIC`与`BETWEEN`相同，只是不要求`AND`左侧的参数小于或等于右侧的参数；正确的范围会自动确定。

要检查一个值是否为 null，使用下面的结构：

```
expression IS NULL
expression IS NOT NULL
```

或者等效但非标准的结构：

```
expression ISNULL
expression NOTNULL
```

不要写*expression* `= NULL`，因为`NULL`并不“等于”`NULL`。（null 值表示未知值，而我们并不知道两个未知值是否相等。）这一行为符合 SQL 标准。

提示

有些应用可能期望表达式*expression* `= NULL`在*expression*求值为 null 时返回真。我们强烈建议此类应用修改为遵循 SQL 标准。但是，如果无法这样修改，那么可以使用配置变量transform_null_equals。如果将其打开，PostgreSQL会把`x = NULL`子句转换成`x IS NULL`。这是PostgreSQL 6.5 到 7.1 版本中的默认行为。

当任一输入为 null 时，普通的比较操作符会得到 null（表示“未知”）。进行比较的另一种方式是使用`IS DISTINCT FROM`结构：

```
expression IS DISTINCT FROM expression
```

对于非 null 输入，它与<>操作符一样。不过，如果两个输入都为 null，它会返回假。而如果只有一个输入为 null，它会返回真。因此，它实际上把 null 当作一种普通数据值，而不是“未知”。

布尔值也可以使用下列结构进行测试：

```
expression IS TRUE
expression IS NOT TRUE
expression IS FALSE
expression IS NOT FALSE
expression IS UNKNOWN
expression IS NOT UNKNOWN
```

这些谓词总是返回真或假，从不返回空值，即使操作数为 null 也是如此。null 输入被当作逻辑值“未知”。请注意，`IS UNKNOWN`和`IS NOT UNKNOWN`实际上分别等同于`IS NULL`和`IS NOT NULL`，只是输入表达式必须是布尔类型。

9.3. 数学函数和操作符

PostgreSQL为很多类型提供了数学操作符。对于那些没有标准数学惯例的类型（如日期/时间类型），我们将在后续小节中描述实际的行为。

表 9.2列出了可用的数学操作符。

表 9.2. 数学操作符

Operator	描述	示例	Result
+	加法	2 + 3	5
-	减法	2 - 3	-1
*	乘法	2 * 3	6
/	除法（整数除法会截断结果）	4 / 2	2
%	取模（余数）	5 % 4	1
^	求幂（从左向右结合）	2.0 ^ 3.0	8
|/	平方根	|/ 25.0	5
||/	立方根	||/ 27.0	3
!	阶乘	5 !	120
!!	阶乘（前缀操作符）	!! 5	120
@	绝对值	@ -5.0	5
&	按位与	91 & 15	11
|	按位或	32 | 3	35
#	按位异或	17 # 5	20
~	按位非	~1	-2
<<	按位左移	1 << 4	16
>>	按位右移	8 >> 2	2

按位操作符仅适用于整数数据类型，而其他操作符适用于所有数值数据类型。
按位操作符也可用于位串类型 `bit` 和 `bit varying`，如表 9.10所示。

表 9.3列出了可用的数学函数。表中的 dp 表示 double precision。许多函数提供了参数类型不同的多种形式。除非另有说明，函数的每种形式都返回与其参数相同的数据类型。处理 double precision 数据的函数大多基于主机系统的 C 库实现；因此，其精度和边界情况下的行为可能因主机系统而异。

表 9.3. 数学函数

Function	abs(*x*)
Return Type	（与*x*相同）
描述	绝对值
示例	abs(-17.4)
Result	17.4
Function	cbrt(dp)
Return Type	dp
描述	cube root
示例	cbrt(27.0)
Result	3
Function	ceil(dp or numeric)
Return Type	（与输入相同）
描述	不小于参数的最小整数
示例	ceil(-42.8)
Result	-42
Function	ceiling(dp or numeric)
Return Type	（与输入相同）
描述	不小于参数的最小整数（ceil的别名）
示例	ceiling(-95.3)
Result	-95
Function	degrees(dp)
Return Type	dp
描述	弧度转换为角度
示例	degrees(0.5)
Result	28.6478897565412
Function	exp(dp or numeric)
Return Type	（与输入相同）
描述	指数函数
示例	exp(1.0)
Result	2.71828182845905

Function	`floor(dp or numeric)`
Return Type	（与输入相同）
描述	不大于参数的最大整数
示例	`floor(-42.8)`
Result	`-43`
Function	`ln(dp or numeric)`
Return Type	（与输入相同）
描述	自然对数
示例	`ln(2.0)`
Result	`0.693147180559945`
Function	`log(dp or numeric)`
Return Type	（与输入相同）
描述	以 10 为底的对数
示例	`log(100.0)`
Result	`2`
Function	`log(b numeric, x numeric)`
Return Type	`numeric`
描述	以 *b* 为底的对数
示例	`log(2.0, 64.0)`
Result	`6.0000000000`
Function	`mod(y, x)`
Return Type	（与参数类型相同）
描述	*y*/*x*的余数
示例	`mod(9,4)`
Result	`1`
Function	`pi()`
Return Type	`dp`
描述	“π” constant
示例	`pi()`
Result	`3.14159265358979`
Function	`power(a dp, b dp)`
Return Type	`dp`
描述	*a*的*b*次幂
示例	`power(9.0, 3.0)`
Result	`729`

Function	`power(a numeric, b numeric)`
Return Type	`numeric`
描述	*a*的*b*次幂
示例	`power(9.0, 3.0)`
Result	`729`
Function	`radians(dp)`
Return Type	`dp`
描述	角度转换为弧度
示例	`radians(45.0)`
Result	`0.785398163397448`
Function	`random()`
Return Type	`dp`
描述	介于 0.0 和 1.0 之间的随机值
示例	`random()`
Result	
Function	`round(dp or numeric)`
Return Type	（与输入相同）
描述	舍入到最接近的整数
示例	`round(42.4)`
Result	`42`
Function	`round(v numeric, s int)`
Return Type	`numeric`
描述	舍入到 *s* 位小数
示例	`round(42.4382, 2)`
Result	`42.44`
Function	`setseed(dp)`
Return Type	`int`
描述	为后续`random()`调用设置种子
示例	`setseed(0.54823)`
Result	`1177314959`
Function	`sign(dp or numeric)`
Return Type	（与输入相同）
描述	参数的符号（-1、0、+1）
示例	`sign(-8.4)`
Result	`-1`

Function	`sqrt(dp or numeric)`
Return Type	（与输入相同）
描述	平方根
示例	`sqrt(2.0)`
Result	`1.4142135623731`
Function	`trunc(dp or numeric)`
Return Type	（与输入相同）
描述	向零截断
示例	`trunc(42.8)`
Result	`42`
Function	`trunc(v numeric, s int)`
Return Type	`numeric`
描述	截断到 *s* 位小数
示例	`trunc(42.4382, 2)`
Result	`42.43`
Function	`width_bucket(op numeric, b1 numeric, b2 numeric, count int)`
Return Type	`int`
描述	返回 *operand* 在以 *b1* 为上界、*b2* 为下界、有 *count* 个桶的等深直方图中所属的桶编号
示例	`width_bucket(5.35, 0.024, 10.06, 5)`
Result	`3`

最后，表 9.4列出了可用的三角函数。所有三角函数都接受`double precision`类型的参数并返回该类型的值。

表 9.4. 三角函数

Function	描述
`acos(x)`	inverse cosine
`asin(x)`	inverse sine
`atan(x)`	inverse tangent
`atan2(x, y)`	*x*/*y*的反正切
`cos(x)`	cosine
`cot(x)`	cotangent
`sin(x)`	sine
`tan(x)`	tangent

9.4. 字符串函数和操作符

本节描述用于检查和操作字符串值的函数和操作符。这里的字符串包括 `character`、`character varying` 和 `text` 类型的值。除非另有说明，下面列出的所有函数都适用于这些类型，但使用 `character` 类型时，应注意自动填充空格可能带来的影响。通常，这里描述的函数也能用于非字符串类型的数据，方法是先把该数据转换为字符串表示。
一些函数也为位串类型提供了原生实现。

SQL定义了一些字符串函数，它们使用特殊语法，其中用某些关键字而不是逗号来分隔参数。详情见表 9.5。这些函数也使用常规的函数调用语法实现（见表 9.6）。

表 9.5. SQL字符串函数和操作符

Function	*string* `||` *string*
Return Type	`text`
描述	字符串串接
示例	`'Post' || 'greSQL'`
Result	`PostgreSQL`
Function	`bit_length(`*string*`)`
Return Type	`int`
描述	串中的位数
示例	`bit_length('jose')`
Result	`32`
Function	`char_length(`*string*`)` or `character_length(`*string*`)`
Return Type	`int`
描述	字符串中的字符数
示例	`char_length('jose')`
Result	`4`
Function	`convert(`*string* `using` *conversion_name*`)`
Return Type	`text`
描述	使用指定的转换名更改编码。可以用`CREATE CONVERSION`定义转换，也有一些预定义的转换名。可用的转换名见表 9.7。
示例	`convert('PostgreSQL' using iso_8859_1_to_utf8)`
Result	UTF8（Unicode，8 位）编码下的`'PostgreSQL'`
Function	`lower(`*string*`)`
Return Type	`text`
描述	范围的下界
示例	`lower('TOM')`
Result	`tom`

Function	`octet_length(string)`
Return Type	`int`
描述	串中的字节数
示例	`octet_length('jose')`
Result	`4`
Function	`overlay(string placing string from int [for int])`
Return Type	`text`
描述	Replace substring
示例	`overlay('Txxxxas' placing 'hom' from 2 for 4)`
Result	`Thomas`
Function	`position(substring in string)`
Return Type	`int`
描述	指定子串的位置
示例	`position('om' in 'Thomas')`
Result	`3`
Function	`substring(string [from int] [for int])`
Return Type	`text`
描述	Extract substring
示例	`substring('Thomas' from 2 for 3)`
Result	`hom`
Function	`substring(string from pattern)`
Return Type	`text`
描述	提取匹配 POSIX 正则表达式的子字符串。有关模式匹配的更多信息，参见第 9.7 节。
示例	`substring('Thomas' from '...$')`
Result	`mas`
Function	`substring(string from pattern for escape)`
Return Type	`text`
描述	提取匹配SQL正则表达式的子字符串。有关模式匹配的更多信息，参见第 9.7 节。
示例	`substring('Thomas' from '%#"o_a#"_' for '#')`
Result	`oma`
Function	`trim([leading | trailing | both] [characters] from string)`
Return Type	`text`
描述	移除*string*开始端/结束端/两端（默认为两端）仅含*characters*（默认为空格）的最长字符串
示例	`trim(both 'x' from 'xTomxx')`

Result	`Tom`
Function	`upper(string)`
Return Type	`text`
描述	范围的上界
示例	`upper('tom')`
Result	`TOM`

还有其他字符串操作函数，列在表 9.6中。其中一些用于内部实现表 9.5中列出的 SQL 标准字符串函数。

表 9.6. 其他字符串函数

Function	`ascii(text)`
Return Type	`int`
描述	参数第一个字符的ASCII码
示例	`ascii('x')`
Result	`120`
Function	`btrim(string text [, characters text])`
Return Type	`text`
描述	移除*string*开始端和结束端仅由*characters*中的字符（默认为空格）组成的最长字符串
示例	`btrim('xyxtrimyyx', 'xy')`
Result	`trim`
Function	`chr(int)`
Return Type	`text`
描述	具有给定ASCII代码的字符
示例	`chr(65)`
Result	`A`
Function	`convert(string text, [src_encoding name,] dest_encoding name)`
Return Type	`text`
描述	把字符串转换为*dest_encoding*编码。原始编码由*src_encoding*指定。如果省略*src_encoding*，则假定使用数据库编码。
示例	`convert( 'text_in_utf8', 'UTF8', 'LATIN1')`
Result	ISO 8859-1 编码表示的`text_in_utf8`
Function	`decode(string text, type text)`
Return Type	`bytea`
描述	解码*string*中先前用`encode`编码的二进制数据。参数类型与`encode`中的相同。
示例	`decode('MTIzAAE=', 'base64')`

Result	`123\000\001`
Function	`encode(data bytea, type text)`
Return Type	`text`
描述	把二进制数据编码为仅含ASCII的表示。支持的类型有：`base64`、`hex`、`escape`。
示例	`encode('123\\000\\001', 'base64')`
Result	`MTIzAAE=`
Function	`initcap(text)`
Return Type	`text`
描述	把每个词的第一个字母转换为大写，其余字母转换为小写。词是由非字母数字字符分隔的字母数字字符序列。
示例	`initcap('hi THOMAS')`
Result	`Hi Thomas`
Function	`length(string text)`
Return Type	`int`
描述	*`string`*中的字符数
示例	`length('jose')`
Result	`4`
Function	`lpad(string text, length int [, fill text])`
Return Type	`text`
描述	通过在前面重复字符*`fill`*（默认为空格）把*`string`*填充到长度*`length`*。如果*`string`*已经比*`length`*长，则将其截断（在右侧）。
示例	`lpad('hi', 5, 'xy')`
Result	`xyxhi`
Function	`ltrim(string text [, characters text])`
Return Type	`text`
描述	移除*`string`*开始端仅含*`characters`*（默认为空格）的最长字符串
示例	`ltrim('zzzytrim', 'xyz')`
Result	`trim`
Function	`md5(string text)`
Return Type	`text`
描述	计算*`string`*的 MD5 哈希，以十六进制返回结果
示例	`md5('abc')`
Result	`900150983cd24fb0 d6963f7d28e17f72`
Function	`pg_client_encoding()`
Return Type	`name`

描述	Current client encoding name
示例	`pg_client_encoding()`
Result	`SQL_ASCII`
Function	`quote_ident(string text)`
Return Type	`text`
描述	返回给定字符串的适当加引号形式，使其可以在SQL语句字符串中用作标识符。只在必要时才加引号（即，当字符串包含非标识符字符或会被大小写折叠时）。内嵌的引号会被适当地双写。
示例	`quote_ident('Foo bar')`
Result	`"Foo bar"`
Function	`quote_literal(string text)`
Return Type	`text`
描述	返回给定字符串的适当加引号形式，使其可以在SQL语句字符串中用作字符串字面量。内嵌的引号和反斜线会被适当地双写。
示例	`quote_literal('O\'Reilly')`
Result	`'O''Reilly'`
Function	`regexp_replace(string text, pattern text, replacement text [, flags text])`
Return Type	`text`
描述	替换匹配 POSIX 正则表达式的子字符串。有关模式匹配的更多信息，参见第 9.7 节。
示例	`regexp_replace('Thomas', '.[mN]a.', 'M')`
Result	`ThM`
Function	`repeat(string text, number int)`
Return Type	`text`
描述	将*string*重复指定次数（*number*）
示例	`repeat('Pg', 4)`
Result	`PgPgPgPg`
Function	`replace(string text, from text, to text)`
Return Type	`text`
描述	把*string*中子字符串*from*的所有出现替换为子字符串*to*
示例	`replace('abcdefabcdef', 'cd', 'XX')`
Result	`abXXefabXXef`
Function	`rpad(string text, length int [, fill text])`
Return Type	`text`
描述	通过追加字符*fill*（默认为一个空格）把*string*填充到长度*length*。如果*string*已经比*length*长，则将它截断。
示例	`rpad('hi', 5, 'xy')`

Result	`hixyx`
Function	`rtrim(string text [, characters text])`
Return Type	`text`
描述	移除*string*结束端仅含*characters*（默认为空格）的最长字符串
示例	`rtrim('trimxxxx', 'x')`
Result	`trim`
Function	`split_part(string text, delimiter text, field int)`
Return Type	`text`
描述	按*delimiter*分隔*string*，并返回给定字段（从 1 开始计数）
示例	`split_part('abc~@~def~@~ghi', '~@~', 2)`
Result	`def`
Function	`strpos(string, substring)`
Return Type	`int`
描述	指定子字符串的位置（与`position(substring in string)`相同，但注意参数顺序相反）
示例	`strpos('high', 'ig')`
Result	`2`
Function	`substr(string, from [, count])`
Return Type	`text`
描述	Extract substring (same as `substring(string from from for count)`)
示例	`substr('alphabet', 3, 2)`
Result	`ph`
Function	`to_ascii(text [, encoding])`
Return Type	`text`
描述	把*text*从其它编码转换为ASCII [a]
示例	`to_ascii('Karel')`
Result	`Karel`
Function	`to_hex(number int or bigint)`
Return Type	`text`
描述	把*number*转换为等价的十六进制表示
示例	`to_hex(2147483647)`
Result	`7fffffff`
Function	`translate(string text, from text, to text)`
Return Type	`text`
描述	*string*中任何与*from*集合中字符匹配的字符，都会被替换成*to*集合中的相应字符

示例	translate('12345', '14', 'ax')
Result	a23x5

[a] to_ascii函数只支持从LATIN1、LATIN2、LATIN9和WIN1250编码转换。

表 9.7. 内置转换

Conversion Name [a]	Source Encoding	Destination Encoding
ascii_to_mic	SQL_ASCII	MULE_INTERNAL
ascii_to_utf8	SQL_ASCII	UTF8
big5_to_euc_tw	BIG5	EUC_TW
big5_to_mic	BIG5	MULE_INTERNAL
big5_to_utf8	BIG5	UTF8
euc_cn_to_mic	EUC_CN	MULE_INTERNAL
euc_cn_to_utf8	EUC_CN	UTF8
euc_jp_to_mic	EUC_JP	MULE_INTERNAL
euc_jp_to_sjis	EUC_JP	SJIS
euc_jp_to_utf8	EUC_JP	UTF8
euc_kr_to_mic	EUC_KR	MULE_INTERNAL
euc_kr_to_utf8	EUC_KR	UTF8
euc_tw_to_big5	EUC_TW	BIG5
euc_tw_to_mic	EUC_TW	MULE_INTERNAL
euc_tw_to_utf8	EUC_TW	UTF8
gb18030_to_utf8	GB18030	UTF8
gbk_to_utf8	GBK	UTF8
iso_8859_10_to_utf8	LATIN6	UTF8
iso_8859_13_to_utf8	LATIN7	UTF8
iso_8859_14_to_utf8	LATIN8	UTF8
iso_8859_15_to_utf8	LATIN9	UTF8
iso_8859_16_to_utf8	LATIN10	UTF8
iso_8859_1_to_mic	LATIN1	MULE_INTERNAL
iso_8859_1_to_utf8	LATIN1	UTF8
iso_8859_2_to_mic	LATIN2	MULE_INTERNAL
iso_8859_2_to_utf8	LATIN2	UTF8
iso_8859_2_to_windows_1250	LATIN2	WIN1250
iso_8859_3_to_mic	LATIN3	MULE_INTERNAL
iso_8859_3_to_utf8	LATIN3	UTF8
iso_8859_4_to_mic	LATIN4	MULE_INTERNAL
iso_8859_4_to_utf8	LATIN4	UTF8
iso_8859_5_to_koi8_r	ISO_8859_5	KOI8
iso_8859_5_to_mic	ISO_8859_5	MULE_INTERNAL
iso_8859_5_to_utf8	ISO_8859_5	UTF8

Conversion Name [a]	Source Encoding	Destination Encoding
iso_8859_5_to_windows_1251	ISO_8859_5	WIN1251
iso_8859_5_to_windows_866	ISO_8859_5	WIN866
iso_8859_6_to_utf8	ISO_8859_6	UTF8
iso_8859_7_to_utf8	ISO_8859_7	UTF8
iso_8859_8_to_utf8	ISO_8859_8	UTF8
iso_8859_9_to_utf8	LATIN5	UTF8
johab_to_utf8	JOHAB	UTF8
koi8_r_to_iso_8859_5	KOI8	ISO_8859_5
koi8_r_to_mic	KOI8	MULE_INTERNAL
koi8_r_to_utf8	KOI8	UTF8
koi8_r_to_windows_1251	KOI8	WIN1251
koi8_r_to_windows_866	KOI8	WIN866
mic_to_ascii	MULE_INTERNAL	SQL_ASCII
mic_to_big5	MULE_INTERNAL	BIG5
mic_to_euc_cn	MULE_INTERNAL	EUC_CN
mic_to_euc_jp	MULE_INTERNAL	EUC_JP
mic_to_euc_kr	MULE_INTERNAL	EUC_KR
mic_to_euc_tw	MULE_INTERNAL	EUC_TW
mic_to_iso_8859_1	MULE_INTERNAL	LATIN1
mic_to_iso_8859_2	MULE_INTERNAL	LATIN2
mic_to_iso_8859_3	MULE_INTERNAL	LATIN3
mic_to_iso_8859_4	MULE_INTERNAL	LATIN4
mic_to_iso_8859_5	MULE_INTERNAL	ISO_8859_5
mic_to_koi8_r	MULE_INTERNAL	KOI8
mic_to_sjis	MULE_INTERNAL	SJIS
mic_to_windows_1250	MULE_INTERNAL	WIN1250
mic_to_windows_1251	MULE_INTERNAL	WIN1251
mic_to_windows_866	MULE_INTERNAL	WIN866
sjis_to_euc_jp	SJIS	EUC_JP
sjis_to_mic	SJIS	MULE_INTERNAL
sjis_to_utf8	SJIS	UTF8
tcvn_to_utf8	WIN1258	UTF8
uhc_to_utf8	UHC	UTF8
utf8_to_ascii	UTF8	SQL_ASCII
utf8_to_big5	UTF8	BIG5
utf8_to_euc_cn	UTF8	EUC_CN
utf8_to_euc_jp	UTF8	EUC_JP
utf8_to_euc_kr	UTF8	EUC_KR
utf8_to_euc_tw	UTF8	EUC_TW

Conversion Name [a]	Source Encoding	Destination Encoding
utf8_to_gb18030	UTF8	GB18030
utf8_to_gbk	UTF8	GBK
utf8_to_iso_8859_1	UTF8	LATIN1
utf8_to_iso_8859_10	UTF8	LATIN6
utf8_to_iso_8859_13	UTF8	LATIN7
utf8_to_iso_8859_14	UTF8	LATIN8
utf8_to_iso_8859_15	UTF8	LATIN9
utf8_to_iso_8859_16	UTF8	LATIN10
utf8_to_iso_8859_2	UTF8	LATIN2
utf8_to_iso_8859_3	UTF8	LATIN3
utf8_to_iso_8859_4	UTF8	LATIN4
utf8_to_iso_8859_5	UTF8	ISO_8859_5
utf8_to_iso_8859_6	UTF8	ISO_8859_6
utf8_to_iso_8859_7	UTF8	ISO_8859_7
utf8_to_iso_8859_8	UTF8	ISO_8859_8
utf8_to_iso_8859_9	UTF8	LATIN5
utf8_to_johab	UTF8	JOHAB
utf8_to_koi8_r	UTF8	KOI8
utf8_to_sjis	UTF8	SJIS
utf8_to_tcvn	UTF8	WIN1258
utf8_to_uhc	UTF8	UHC
utf8_to_windows_1250	UTF8	WIN1250
utf8_to_windows_1251	UTF8	WIN1251
utf8_to_windows_1252	UTF8	WIN1252
utf8_to_windows_1256	UTF8	WIN1256
utf8_to_windows_866	UTF8	WIN866
utf8_to_windows_874	UTF8	WIN874
windows_1250_to_iso_8859_2	WIN1250	LATIN2
windows_1250_to_mic	WIN1250	MULE_INTERNAL
windows_1250_to_utf8	WIN1250	UTF8
windows_1251_to_iso_8859_5	WIN1251	ISO_8859_5
windows_1251_to_koi8_r	WIN1251	KOI8
windows_1251_to_mic	WIN1251	MULE_INTERNAL
windows_1251_to_utf8	WIN1251	UTF8
windows_1251_to_windows_866	WIN1251	WIN866
windows_1252_to_utf8	WIN1252	UTF8
windows_1256_to_utf8	WIN1256	UTF8
windows_866_to_iso_8859_5	WIN866	ISO_8859_5

Conversion Name [a]	Source Encoding	Destination Encoding
windows_866_to_koi8_r	WIN866	KOI8
windows_866_to_mic	WIN866	MULE_INTERNAL
windows_866_to_utf8	WIN866	UTF8
windows_866_to_windows_1251	WIN866	WIN
windows_874_to_utf8	WIN874	UTF8

[a] 转换名称遵循标准命名规则：将源编码正式名称中的所有非字母数字字符替换为下划线，再接上 _to_，然后接上按相同方式处理的目标编码名称。因此，这些名称可能与惯用的编码名称不同。

9.5. 二进制串函数和操作符

本节描述用于检查和操作 bytea 类型值的函数和操作符。

SQL定义了一些具有特殊语法的二进制串函数，其中使用某些关键字而不是逗号来分隔参数。详情见表 9.8。其中一些函数也使用常规的函数调用语法实现（见表 9.9）。

表 9.8. SQL二进制串函数和操作符

Function	*string* || *string*
Return Type	bytea
描述	字符串串接
示例	'\\\\Post'::bytea || '\\047gres\\000'::bytea
Result	\\Post'gres\000
Function	octet_length(*string*)
Return Type	int
描述	二进制串中的字节数
示例	octet_length('jo\\000se'::bytea)
Result	5
Function	position(*substring* in *string*)
Return Type	int
描述	指定子串的位置
示例	position('\\000om'::bytea in 'Th\\000omas'::bytea)
Result	3
Function	substring(*string* [from int] [for int])
Return Type	bytea
描述	Extract substring
示例	substring('Th\\000omas'::bytea from 2 for 3)
Result	h\000o
Function	trim([both] *bytes* from *string*)
Return Type	bytea

描述	移除*string*开始端和结束端仅含*bytes*中的字节（默认为空格）组成的最长字符串
示例	`trim('\\000'::bytea from '\\000Tom\\000'::bytea)`
Result	`Tom`
Function	`get_byte(`*`string`*`,` *`offset`*`)`
Return Type	`int`
描述	从字符串中抽取字节
示例	`get_byte('Th\\000omas'::bytea, 4)`
Result	`109`
Function	`set_byte(`*`string`*`,` *`offset`*`,` *`newvalue`*`)`
Return Type	`bytea`
描述	在字符串中设置字节
示例	`set_byte('Th\\000omas'::bytea, 4, 64)`
Result	`Th\000o@as`
Function	`get_bit(`*`string`*`,` *`offset`*`)`
Return Type	`int`
描述	从字符串中抽取位
示例	`get_bit('Th\\000omas'::bytea, 45)`
Result	`1`
Function	`set_bit(`*`string`*`,` *`offset`*`,` *`newvalue`*`)`
Return Type	`bytea`
描述	在字符串中设置位
示例	`set_bit('Th\\000omas'::bytea, 45, 0)`
Result	`Th\000omAs`

还有一些二进制串处理函数可以使用，在表 9.9列出。其中有一些是在内部使用，用于实现表 9.8列出的 SQL 标准串函数。

表 9.9. 其他二进制串函数

Function	`btrim(`*`string`* `bytea,` *`bytes`* `bytea)`
Return Type	`bytea`
描述	移除*string*开始端和结束端仅由*bytes*中的字节组成的最长字符串
示例	`btrim('\\000trim\\000'::bytea, '\\000'::bytea)`
Result	`trim`
Function	`length(`*`string`*`)`
Return Type	`int`
描述	二进制串的长度

示例	`length('jo\\000se'::bytea)`
Result	`5`
Function	`md5(string)`
Return Type	`text`
描述	计算*string*的 MD5 哈希，以十六进制返回结果
示例	`md5('Th\\000omas'::bytea)`
Result	`8ab2d3c9689aaf18 b4958c334c82d8b1`
Function	`decode(string text, type text)`
Return Type	`bytea`
描述	解码*string*中先前用encode编码的二进制串。参数类型与encode中的相同。
示例	`decode('123\\000456', 'escape')`
Result	`123\000456`
Function	`encode(string bytea, type text)`
Return Type	`text`
描述	把二进制串编码为仅含ASCII的表示。支持的类型有：base64、hex、escape。
示例	`encode('123\\000456'::bytea, 'escape')`
Result	`123\000456`

9.6. 位串函数和操作符

本节描述用于检查和操作位串的函数和操作符，也就是 `bit` 和 `bit varying` 类型的值。除常用比较操作符外，还可使用表 9.10中的操作符。`&`、`|` 和 `#` 的位串操作数长度必须相同。移位时会保留字符串的原始长度，如示例所示。

表 9.10. 位串操作符

Operator	描述	示例	Result
`||`	串接	`B'10001' || B'011'`	`10001011`
`&`	按位与	`B'10001' & B'01101'`	`00001`
`|`	按位或	`B'10001' | B'01101'`	`11101`
`#`	按位异或	`B'10001' # B'01101'`	`11100`
`~`	按位非	`~ B'10001'`	`01110`
`<<`	按位左移	`B'10001' << 3`	`01000`
`>>`	按位右移	`B'10001' >> 2`	`00100`

以下 SQL 标准函数既适用于位串，也适用于字符串：`length`、`bit_length`、`octet_length`、`position`、`substring`。

此外，还可以在整数值与bit类型之间进行类型转换。例如：

```
44::bit(10)                     0000101100
44::bit(3)                      100
cast(-44 as bit(12))            111111010100
'1110'::bit(4)::integer         14
```

注意，仅转换为 “bit” 意味着转换为`bit(1)`，因此只会得到该整数的最低有效位。

> **注意**
>
> 在PostgreSQL 8.0 之前，把整数转换为`bit(n)`会复制该整数最左边的`n`位，而现在复制的是最右边的`n`位。此外，把整数转换为比该整数本身更宽的位串时，
> 会在左侧进行符号扩展。

9.7. 模式匹配

PostgreSQL提供了三种独立的模式匹配方法：传统的SQL `LIKE`操作符、较新的`SIMILAR TO`操作符（在 SQL:1999 中加入），以及POSIX风格的正则表达式。此外，还有一个模式匹配函数`substring`可用，它可以使用`SIMILAR TO`风格或 POSIX 风格的正则表达式。

> **提示**
>
> 如果你的模式匹配的要求超出了这些，请考虑用 Perl 或 Tcl 写一个用户定义的函数。

9.7.1. LIKE

string LIKE *pattern* [ESCAPE *escape-character*]
string NOT LIKE *pattern* [ESCAPE *escape-character*]

每一个*pattern*都定义了一个字符串集合。如果*string*包含在*pattern*
所表示的字符串集合中，那么`LIKE`表达式返回真（和预期的一样，如果`LIKE`返回真，那么`NOT LIKE`表达式返回假，反之亦然。一个等效的表达式是`NOT (`*string* `LIKE` *pattern*`)`）。

如果*pattern*不包含百分号或下划线，那么该模式只表示它本身的字符串；在这种情况下，
`LIKE`的行为就像等号操作符。*pattern*中的下划线（`_`）代表（匹配）任意单个字符；
百分号（`%`）匹配任意由零个或多个字符组成的序列。

一些示例：

```
'abc' LIKE 'abc'    true
'abc' LIKE 'a%'     true
'abc' LIKE '_b_'    true
'abc' LIKE 'c'      false
```

`LIKE`模式匹配总是覆盖整个字符串。因此，如果想要匹配字符串内任意位置上的一个序列，
该模式就必须以百分号开头并以百分号结尾。

要匹配字面量下划线或百分号而不是把它们当作通配符，*pattern*
中相应的字符前面必须带有转义字符。默认的转义字符是反斜线，但也可以使用`ESCAPE`
子句选择其他转义字符。要匹配转义字符本身，请写两个转义字符。

注意，反斜线在字符串常量中本来就有特殊含义，因此要写一个包含反斜线的模式常量，必须在
SQL 语句中写两个反斜线。这样，要写一个实际匹配字面量反斜线的模式，就意味着在语句中写

四个反斜线。可以通过ESCAPE选择不同的转义字符来避免这一点；这样反斜线对LIKE
就不再是特殊字符。（但反斜线对字符串常量解析器仍然是特殊字符，
因此要匹配一个反斜线仍需写两个。）

也可以通过写 ESCAPE '' 来选择不使用转义字符。这会禁用转义机制，从而无法关闭模式中下划线和百分号的特殊含义。

可以用关键字 ILIKE 代替 LIKE，使匹配根据当前区域设置忽略大小写。这不属于 SQL 标准，而是 PostgreSQL 的扩展。

操作符~~等效于LIKE，而~~*对应ILIKE。还有 !~~和!~~*操作符分别代表NOT LIKE和NOT ILIKE。所有这些操作符都是PostgreSQL特有的。

9.7.2. SIMILAR TO 正则表达式

string SIMILAR TO *pattern* [ESCAPE *escape-character*]
string NOT SIMILAR TO *pattern* [ESCAPE *escape-character*]

SIMILAR TO 操作符根据其模式是否匹配给定字符串返回真或假。它与 LIKE 类似，但按照 SQL 标准定义的正则表达式来解释模式。SQL 正则表达式是 LIKE 表示法和常见正则表达式表示法的一种奇特结合。

与LIKE类似，SIMILAR TO操作符只有在其模式匹配整个字符串时才算成功；这一点不同于普通正则表达式，后者可以匹配字符串的任意部分。与LIKE相同，SIMILAR TO也使用_和%作为通配符，分别匹配任意单个字符和任意字符串（分别类似于 POSIX 正则表达式中的.和.*）。

除了这些从LIKE借用的功能之外，SIMILAR TO支持下面这些从 POSIX 正则表达式借用的模式匹配元字符：

- |表示选择（两个候选之一）。
- *表示重复前面的项零次或更多次。
- +表示重复前面的项一次或更多次。
- 可以使用圆括号()把多个项组合成一个逻辑项。
- 一个方括号表达式[...]声明一个字符类，就像 POSIX 正则表达式一样。

注意，有界重复（?和{...}）没有提供，尽管它们存在于 POSIX 中。另外，点号（.）不是 SIMILAR TO 的一个元字符。

与LIKE一样，反斜杠会禁用这些元字符的特殊含义；也可以用 ESCAPE 指定其他转义字符。

下面是一些示例：

```
'abc' SIMILAR TO 'abc'      true
'abc' SIMILAR TO 'a'        false
'abc' SIMILAR TO '%(b|d)%'  true
'abc' SIMILAR TO '(b|c)%'   false
```

带三个参数的 substring 函数，即 substring(*string* from *pattern* for *escape-character*)，用于提取匹配 SQL 正则表达式模式的子串。与 SIMILAR TO 一样，指定模式必须匹配整个数据字符串，否则函数失败并返回空值。要指出成功时应返回的模式部分，模式中必须出现两次后跟双引号（"）的转义字符。
函数返回匹配这两个标记之间的模式部分的文本。

下面是一些示例：

```
substring('foobar' from '%#"o_b#"%' for '#')    oob
substring('foobar' from '#"o_b#"%' for '#')     NULL
```

9.7.3. POSIX正则表达式

表 9.11列出了所有可用于 POSIX 正则表达式模式匹配的操作符。

表 9.11. 正则表达式匹配操作符

Operator	描述	示例
~	匹配正则表达式，区分大小写	'thomas' ~ '.*thomas.*'
~*	匹配正则表达式，不区分大小写	'thomas' ~* '.*Thomas.*'
!~	不匹配正则表达式，区分大小写	'thomas' !~ '.*Thomas.*'
!~*	不匹配正则表达式，不区分大小写	'thomas' !~* '.*vadim.*'

POSIX正则表达式提供了比LIKE和SIMILAR TO操作符更强大的模式匹配方式。许多 Unix 工具，例如egrep、sed或awk，都使用与这里描述的模式匹配语言相似的语言。

正则表达式是一个字符序列，是定义一组字符串（一个正则集）的简写。
如果字符串属于正则表达式描述的正则集，就称该字符串匹配此正则表达式。与LIKE一样，
模式中的字符精确匹配字符串中的字符，除非该模式字符在正则表达式语言中有特殊含义 – 但正则表达式使用的特殊字符与LIKE不同。与LIKE模式不同，
正则表达式可以匹配字符串中的任意位置，除非显式将其锚定到字符串开头或末尾。

下面是一些示例：

```
'abc' ~ 'abc'    true
'abc' ~ '^a'     true
'abc' ~ '(b|d)'  true
'abc' ~ '^(b|c)' false
```

带两个参数的 substring 函数，即 substring(*string* from *pattern*)，用于提取匹配 POSIX 正则表达式模式的子串。如果没有匹配，它返回空值；否则返回匹配模式的那部分文本。但是，如果模式包含圆括号，则返回匹配第一个括号子表达式（左圆括号最先出现的那个）的文本。如果要在表达式内部使用圆括号而不触发这个例外，
可以在整个表达式外再加一对圆括号。如果需要在想要提取的子表达式之前使用圆括号，
请参见下文介绍的非捕获圆括号。

下面是一些示例：

```
substring('foobar' from 'o.b')     oob
substring('foobar' from 'o(.)b')   o
```

regexp_replace函数用新文本替换匹配 POSIX 正则表达式模式的子字符串。它的语法为regexp_replace(*source*,*pattern*,*replacement*[,*flags*])。如果没有与*pattern*匹配的内容，则原样返回*source*字符串。如果存在匹配，则返回把匹配子字符串替换成*replacement*字符串后的*source*字符串。*replacement*字符串中可以包含*n*（其中*n*为1到9），表示应插入与模式中第*n*个带圆括号子表达式匹配的源子字符串；还可以包含\&，
表示应插入与整个模式匹配的子字符串。如果需要在替换文本中放入字面量反斜线，请写\\。
（与通常一样，请记住把字面常量字符串中的反斜线双写。）*flags*
参数是一个可选的文本字符串，包含零个或多个单字母标志，用于改变函数的行为。标志i
指定不区分大小写的匹配，而标志g指定替换每一个匹配的子字符串而不是仅替换第一个。

下面是一些示例：

```
regexp_replace('foobarbaz', 'b..', 'X')
                                   fooXbaz
regexp_replace('foobarbaz', 'b..', 'X', 'g')
                                   fooXX
regexp_replace('foobarbaz', 'b(..)', 'X\\1Y', 'g')
                                   fooXarYXazY
```

PostgreSQL的正则表达式是使用 Henry Spencer 写的一个包来实现的。下面的正则表达式的大部分描述都是从他的手册页中逐字拷贝过来的。

9.7.3.1. 正则表达式细节

正则表达式（RE），在POSIX 1003.2 中定义，它有两种形式：扩展的RE或者是ERE（大概地说就是那些在`egrep`里的），基本的RE或者是BRE（大概地说就是那些在`ed`里的）。PostgreSQL支持两种形式，并且还实现了一些POSIX标准中没有但是在类似 Perl 或者 Tcl 这样的语言中得到广泛应用的一些扩展。使用了那些非POSIX扩展的RE叫高级RE，或者本文档里说的ARE。ARE 几乎完全是 ERE 的超集，但是 BRE 有几个符号上的不兼容（以及更多的限制）。我们首先描述 ARE 和 ERE 形式，描述那些只适用于 ARE 的特性，然后描述 BRE 的区别是什么。

> **注意**
>
> PostgreSQL接受的正则表达式形式可以通过设置regex_flavor运行时参数来选择。通常的设置是 `advanced`，但为了与 PostgreSQL 7.4 之前的版本向后兼容，也可以选择 `extended`。

一个正则表达式被定义为一个或更多分支，它们之间被 `|` 分隔。
只要能匹配其中一个分支的东西都能匹配正则表达式。

一个分支是零个或多个量化原子或约束，连接在一起。它匹配第一个的匹配项，
然后是第二个的匹配项，依此类推；一个空分支匹配空字符串。

一个量化原子是一个原子，后面可以跟一个量词。没有量词时，匹配一次原子所匹配的内容；
有量词时，按量词指定的次数匹配原子所匹配的内容。原子可以是表 9.12列出的任何一种形式。
可用量词及其含义见表 9.13。

一个约束匹配一个空串，但只是在满足特定条件下才匹配。
约束可以在能够使用原子的地方使用，只是它不能跟着量词。简单的约束在表 9.14里显示；更多的约束稍后描述。

表 9.12. 正则表达式原子

Atom	描述
`(` *re* `)`	（其中*re*是任意正则表达式）匹配*re*所匹配的内容，并记录该匹配，以备输出结果
`(?:` *re* `)`	同上，但匹配不被记录以供报告（一组“不捕获的”圆括号）（仅 ARE）
`.`	匹配任意单个字符
`[` *chars* `]`	一个方括号表达式，匹配*chars*中的任意一个（详见第 9.7.3.2 节）

Atom	描述
\ *k*	（其中*k*既不是字母也不是数字）把该字符视为普通字符并匹配它，例如，\\匹配反斜线字符
\ *c*	其中*c*是字母或数字（后面可能还有其他字符），这是一个转义，参见第 9.7.3.3 节（仅适用于 ARE；在 ERE 和 BRE 中，它匹配*c*）
{	如果后面跟着非数字字符，则匹配左花括号{；如果后面跟着数字，则是*bound*的开头（见下文）
x	其中*x*是一个没有其它意义的单个字符，则匹配该字符

RE 不能以反斜线（\）结尾。

> **注意**
>
> 请记住，反斜线（\）在PostgreSQL字符串常量中本来就有特殊含义。
> 要写一个包含反斜线的模式常量，必须在语句中写两个反斜线。

表 9.13. 正则表达式量词

Quantifier	Matches
*	一个由原子的 0 次或更多次匹配组成的序列
+	一个由原子的 1 次或更多次匹配组成的序列
?	一个由原子的 0 次或 1 次匹配组成的序列
{ *m* }	一个由原子的正好*m*次匹配组成的序列
{ *m* , }	一个由原子的*m*次或更多次匹配组成的序列
{ *m* , *n* }	一个由原子的从*m*次到*n*次（包括）匹配组成的序列；*m*不能超过*n*
*?	*的非贪婪版本
+?	+的非贪婪版本
??	?的非贪婪版本
{ *m* }?	{*m*}的非贪婪版本
{ *m* ,}?	{*m*,}的非贪婪版本
{ *m* , *n* }?	{*m*,*n*}的非贪婪版本

使用{ . . . }的形式被称作范围。一个范围内的数字*m*和*n*都是无符号十进制整数，允许的数值从 0 到 255（包含）。

非贪婪量词（仅适用于 ARE）与对应的普通（贪婪）量词匹配相同的可能内容，但优先选择最少的匹配次数，而不是最多的匹配次数。详见第 9.7.3.5 节。

> **注意**
>
> 一个量词不能紧跟在另外一个量词后面。量词不能作为表达式或者子表达式的开头，
> 也不能跟在^或者 | 后面。

表 9.14. 正则表达式约束

Constraint	描述
^	在字符串开头匹配
$	在字符串末尾匹配
(?= *re*)	正向先行在匹配*re*的子字符串开始的任何位置匹配（仅限 ARE）
(?! *re*)	负向先行在不匹配*re*的子字符串开始的任何位置匹配（仅限 ARE）

先行约束不能包含反向引用（参见第 9.7.3.3 节），其中的所有圆括号都视为非捕获圆括号。

9.7.3.2. 方括号表达式

方括号表达式是用[]括起来的字符列表。通常，
它匹配列表中的任意单个字符（但请参见下文）。如果列表以^开头，则匹配任意
不在列表剩余部分中的单个字符。如果列表中的两个字符用-分隔，
则表示排序序列中这两个字符之间的完整字符范围（包含两个端点）；例如，ASCII中的[0-9]
匹配任意十进制数字。两个范围共享一个端点是不合法的，例如a-c-e。
范围高度依赖排序序列，因此可移植程序应避免依赖它们。

想在列表中包含字面字符]，可以让它做列表的首字符（如果使用了^，需要放在其后）。
想在列表中包含字面字符-，可以让它做列表的首字符或者尾字符，或者一个范围的第二个端点。
想在列表中把字面字符-当做范围的起点，把它用[.和.]包围起来，
这样它就成为一个排序元素（见下文）。除了这些字符本身、一些用[的组合（见下段）
以及转义（只在 ARE 中有效）以外，所有其它特殊字符在方括号表达式里都失去它们的特殊含义。特别是，在 ERE 和 BRE 规则下\不是特殊的，但在 ARE 里，它是特殊的（引入一个转义）。

在一个方括号表达式里，一个排序元素（一个字符、
一个被当做一个单一字符排序的多字符序列或者一个表示上面两种情况的排序序列名称）包含在[.和.]里面的时候表示该排序元素的字符序列。该序列被当做该方括号列表的一个单一元素。
这允许一个包含多字符排序元素的方括号表达式去匹配多于一个字符，例如，
如果排序序列包含一个ch排序元素，那么 RE [[.ch.]]*c匹配chchcc的头五个字符。

> **注意**
>
> PostgreSQL当前不支持多字符排序元素。这些信息描述了将来可能有的行为。

在方括号表达式里，包围在[=和=]里的排序元素是一个等价类，
代表等效于那一个的所有排序元素的字符序列，包括它本身（如果没有其它等效排序元素，
那么就好像封装定界符是[.和 .]）。例如，如果o和^是一个等价类的成员，那么[[=o=]]、[[=^=]]和[o^]都是同义的。一个等价类不能是一个范围的端点。

在方括号表达式中，用 [: 和 :] 括起的字符类名表示属于该类的所有字符的列表。标准字符类名包括：alnum、alpha、blank、cntrl、digit、graph、lower、print、punct、space、upper、xdigit。它们表示 ctype 中定义的字符类。区域设置可以提供其他字符类。字符类不能用作范围的端点。

方括号表达式有两个特例：[[:<:]] 和 [[:>:]] 是约束，分别匹配单词开头和结尾的空字符串。单词定义为一个前后都没有单词字符的单词字符序列。单词字符是 alnum 字符（由 ctype 定义）或下划线。这是与 POSIX 1003.2 兼容但未由其规定的扩展，在打算移植到其他系统的软件中应谨慎使用。通常更适合使用下文介绍的约束转义；它们并不更标准，但更容易键入。

9.7.3.3. 正则表达式转义

转义是以\开头，后面跟着一个字母或数字字符的特殊序列。转义有好几种变体：字符输入转义、字符类简写转义、约束转义以及反向引用。在 ARE 里，如果一个\后面跟着一个字母或数字，但是并未组成一个合法的转义，那么它是非法的。在 ERE 中没有转义：在方括号表达式之外，一个后面跟着字母或数字字符的\只是表示该字符是一个普通的字符，而且在一个方括号表达式里，\是一个普通的字符（后者是 ERE 与 ARE 之间唯一的实际不兼容之处）。

字符输入转义便于在 RE 中指定不可打印或其他不便输入的字符，见表 9.15。

字符类简写转义用来提供一些常用的字符类简写。它们显示在表 9.16中。

约束转义是以转义形式书写的约束，在满足特定条件时匹配空字符串，见表 9.17。

反向引用（*n*）匹配数字*n*指定的被前面的圆括号子表达式匹配的同一个串（参阅表 9.18）。例如，`([bc])\1`匹配`bb`或者`cc`，但是不匹配`bc`或者`cb`。RE 中子表达式必须完全在反向引用前面。子表达式以它们的左圆括号的顺序编号。非捕获圆括号并不定义子表达式。

> **注意**
>
> 请记住，把模式作为 SQL 字符串常量输入时，转义的开头\需要双写。例如：
>
> `'123' ~ '^\\d{3}'` *true*

表 9.15. 正则表达式字符项转义

Escape	描述
`\a`	警告（响铃）字符，和 C 中一样
`\b`	退格，和 C 中一样
`\B`	反斜线（\）的同义词，用来减少双写反斜线
`\c` *X*	（其中*X*是任意字符）低序5位和*X*相同的字符，它的其他位都是零
`\e`	排序序列名称为`ESC`的字符；若不存在这样的字符，则使用八进制值为 033 的字符
`\f`	换页，和 C 中一样
`\n`	换行符，与 C 中相同
`\r`	回车，和 C 中一样
`\t`	水平制表符，和 C 中一样
`\u` *wxyz*	（其中*wxyz*正好是四个十六进制位）本地字节序中的 UTF16（Unicode 16 位）字符`U+`*wxyz*
`\U` *stuvwxyz*	（其中*stuvwxyz*正好是八个十六进制位）为假设性的 32 位 Unicode 扩展保留
`\v`	垂直制表符，和 C 中一样
`\x` *hhh*	（其中*hhh*是十六进制位的任意序列）十六进制值为`0x`*hhh*的字符（一个单一字符，不管用了多少个十六进制位）
`\0`	值为`0`（空字节）的字符
`\` *xy*	（其中*xy*正好是两个八进制位，并且不是一个反向引用）八进制值为`0`*xy*的字符

Escape	描述
\ *xyz*	（其中*xyz*正好是三个八进制位，并且不是一个反向引用）八进制值为0*xyz*的字符

十六进制位是0-9、a-f和A-F。八进制位是0-7。

字符输入转义总是被当作普通字符。例如，\135是 ASCII 中的]，但\135并不终止一个方括号表达式。

表 9.16. 正则表达式类简写转义

Escape	描述
\d	[[:digit:]]
\s	[[:space:]]
\w	[[:alnum:]_]（注意包括下划线）
\D	[^[:digit:]]
\S	[^[:space:]]
\W	[^[:alnum:]_]（注意包括下划线）

在方括号表达式中，\d、\s 和 \w 会失去其外层方括号，而 \D、\S 和 \W 则是非法的。（因此，例如 [a-c\d] 等同于 [a-c[:digit:]]。此外，等同于 [a-c^[:digit:]] 的 [a-c\D] 是非法的。）

表 9.17. 正则表达式约束转义

Escape	描述
\A	只在串开头匹配（与^的不同请参见第 9.7.3.5 节）
\m	只在一个词的开头匹配
\M	只在一个词的末尾匹配
\y	只在一个词的开头或末尾匹配
\Y	只在不属于单词开头或末尾的位置匹配
\Z	只在串的末尾匹配（与$的不同请参见第 9.7.3.5 节）

单词的定义与上文[[:<:]]和[[:>:]]的说明相同。方括号表达式中不允许使用约束转义。

表 9.18. 正则表达式反向引用

Escape	描述
\ *m*	(where *m* is a nonzero digit) 对第*m*个子表达式的反向引用
\ *mnn*	（其中*m*是一个非零数字，并且*nn*是后续的若干数字，并且十进制值*mnn*不大于此前已出现的捕获右圆括号数）一个到第*mnn*个子表达式的反向引用

注意

八进制字符输入转义与反向引用之间存在固有歧义，
按上文提到的启发式规则解决：前导零始终表示八进制转义。单个非零数字，
如果后面没有其他数字，始终视为反向引用。不以零开头的多位数字序列，

> 如果前面已有相应的子表达式（即该数字在反向引用的合法范围内），则视为反向引用，否则视为八进制转义。

9.7.3.4. 正则表达式元语法

除了上面描述的主要语法之外，还有几种特殊形式和杂项语法。

通常，所使用的 RE 的风格由`regex_flavor`决定。不过，这可以被一个引导前缀覆盖。如果 RE 以`***:`开头，则不管`regex_flavor`如何设置，RE 的余下部分都被当作 ARE。如果 RE 以`***=`开头，则 RE 的余下部分被当作一个字面字符串，其中所有字符都被认为是普通字符。

一个 ARE 可以以嵌入选项开头：一个序列`(?xyz)`（这里的*xyz*是一个或多个字母字符）声明影响剩余 RE 的选项。这些选项覆盖任何先前确定的选项（包括 RE 的语法种类和大小写敏感性）。可用的选项字母在表 9.19中显示。

表 9.19. ARE 嵌入选项字母

Option	描述
`b`	RE的剩余部分是一个BRE
`c`	大小写敏感的匹配（覆盖操作符类型）
`e`	RE的剩余部分是一个ERE
`i`	大小写不敏感的匹配（见第 9.7.3.5 节）（覆盖操作符类型）
`m`	`n`的历史原因的同义词
`n`	换行敏感的匹配（见第 9.7.3.5 节）
`p`	部分换行敏感的匹配（见第 9.7.3.5 节）
`q`	RE 的剩余部分按字面（“加引号”）解释，所有字符都视为普通字符
`s`	非换行敏感的匹配（默认）
`t`	紧凑语法（默认；见下文）
`w`	逆部分换行敏感（“怪异”）的匹配（见第 9.7.3.5 节）
`x`	expanded syntax (see below)

嵌入选项从结束该序列的`)`处开始生效。它们只能出现在 ARE 的开头（如果存在`***:`引导前缀，则位于该前缀之后）。

除了通常的紧凑 RE 语法（其中所有字符都有意义）之外，还有一种扩展语法，可以通过指定嵌入的`x`选项来使用。在扩展语法中，RE 中的空白字符会被忽略，同样被忽略的还有`#`与其后的换行符（或 RE 末尾）之间的所有字符。这使得复杂的 RE 可以分段并添加注释。此基本规则有三个例外：

- 空白字符或 `#` 的前面若有 `\`，该字符就会被保留。
- 方括号表达式里的空白或者`#`将被保留
- 在多字符符号里面不能出现空白和注释，例如`(?:`

在这里，空白字符包括空格、制表符、换行符，以及属于*space*字符类的任何字符。

最后，在 ARE 里，方括号表达式外面，序列`(?#ttt)`（其中*ttt*是任意不包含一个`)`的文本）是一个注释，它被完全忽略。同样，这样的东西是不允许出现在多字符符号的字符中间的，例如 `(`

?:。这种注释更像是一种历史产物而不是一种有用的设施，并且它们的使用已经被废弃；请使用扩展语法来替代。

如果指定了开头的***=引导前缀，那么这些元语法扩展都不能使用，因为这表示把用户输入作为按字面解释的字符串，而非 RE 处理。

9.7.3.5. 正则表达式匹配规则

在 RE 可以在给定串中匹配多于一个子串的情况下，RE 匹配串中最靠前的那个子串。如果 RE 可以匹配在那个位置开始的多个子串，要么是取最长的子串，要么是最短的，具体哪种，取决于 RE 是贪婪的还是非贪婪的。

一个 RE 是否贪婪取决于下面规则：

- 大多数原子以及所有约束，都没有贪婪属性（因为它们毕竟无法匹配长度不定的文本）。
- 在一个 RE 周围加上圆括号并不会改变其贪婪性。
- 带一个固定重复次数量词（{*m*}或者{*m*}?）的量化原子和原子自身具有同样的贪婪性（可能是没有）。
- 一个带其他普通的量词（包括{*m*,*n*}中*m*等于*n*的情况）的量化原子是贪婪的（首选最长匹配）。
- 一个带非贪婪量词（包括{*m*,*n*}?中*m*等于 *n*的情况）的量化原子是非贪婪的（首选最短匹配）。
- 一个分支 – 也就是说，一个没有顶级|操作符的 RE – 和它里面的第一个有贪婪属性的量化原子有着同样的贪婪性。
- 一个由|操作符连接起来的两个或者更多分支组成的 RE 总是贪婪的。

上面的规则所描述的贪婪属性不仅仅适用于独立的量化原子，而且也适用于包含量化原子的分支和整个 RE。这里的意思是，匹配是按照分支或者整个 RE 作为一个整体匹配最长或者最短的可能子串。一旦整个匹配的长度确定，那么匹配任意特定子表达式的部分就基于该子表达式的贪婪属性进行判断，在 RE 里面靠前的子表达式的优先级高于靠后的子表达式。

一个相应的示例：

```
SELECT SUBSTRING('XY1234Z', 'Y*([0-9]{1,3})');
结果: 123
SELECT SUBSTRING('XY1234Z', 'Y*?([0-9]{1,3})');
结果: 1
```

在第一个示例里，RE 作为整体是贪婪的，因为Y*是贪婪的。它可以匹配从Y开始的东西，并且它匹配从这个位置开始的最长的串，也就是，Y123。输出是这里的圆括号包围的部分，或者说是123。在第二个示例里，RE 总体上是一个非贪婪的 RE，因为Y*?是非贪婪的。它可以匹配从Y开始的最短的子串，也就是说Y1。子表达式[0-9]{1,3}是贪婪的，但是它不能修改总体匹配长度的决定；因此它被迫只匹配1。

简而言之，如果一个 RE 同时包含贪婪和非贪婪的子表达式，那么总的匹配长度要么是尽可能长，要么是尽可能短，这取决于给整个 RE 赋予的属性。给子表达式赋予的属性只影响在这个匹配里，各个子表达式相对于其他子表达式能“吃掉”多少内容。

量词{1,1}和{1,1}?可以分别用于在一个子表达式或者整个 RE 上强制贪婪或者非贪婪。

在决定更长或者更短的匹配时，匹配长度是以字符衡量的，而不是排序元素。一个空串会被认为比什么都不匹配长。例如：bb*匹配abbbc的中间三个字符；(week|wee)(

night|knights)匹配weeknights的所有十个字符；而(.*).*匹配 abc的时候，圆括号包围的子表达式匹配所有三个字符；当(a*)*被拿来匹配bc时，整个 RE 和圆括号子表达式都匹配一个空串。

如果指定不区分大小写的匹配，其效果近似于字母表中的所有大小写差别都消失了。
当存在大小写形式的字母作为普通字符出现在方括号表达式之外时，
实际上会转换为包含其大小写形式的方括号表达式，例如x变成[xX]。
当它出现在方括号表达式内部时，其所有大小写形式都会加入该表达式，例如[x]变成[xX]，[^x]变成[^xX]。

如果指定了换行敏感的匹配，.和使用^的方括号表达式将永远不会匹配换行字符（这样，
匹配就不会跨越行，除非 RE 显式安排了跨行匹配）并且^和$除了分别匹配串开头和结尾之外，
还将分别匹配换行后面和前面的空串。但是 ARE 转义\A和\Z仍然只匹配串的开头和结尾。

如果指定了部分换行敏感的匹配，那么它影响.和方括号表达式，
这个时候和换行敏感的匹配一样，但是不影响^和$。

如果指定了逆部分换行敏感匹配，那么它影响^和$，其作用和在换行敏感的匹配里一样，
但是不影响.和方括号表达式。这个并不是很有用，只是为了满足对称性而提供的。

9.7.3.6. 限制和兼容性

在这个实现里，对 RE 的长度没有特别的限制。但是，那些希望高移植性的程序应该避免使用长度超过 256 字节的 RE，因为 POSIX 兼容的实现可以拒绝接受这样的 RE。

ARE 与 POSIX ERE 实际不兼容的唯一特性是：\在方括号表达式中不会失去特殊含义。其他所有 ARE 特性所使用的语法，在 POSIX ERE 中都是非法的，或其效果未定义或未指定；引导前缀的***语法同样不属于 POSIX 的 BRE 或 ERE 语法。

许多 ARE 扩展借鉴自 Perl，但其中一些经过了整理和修改，也未实现少数 Perl 扩展。需要注意的不兼容之处包括\b、\B、不对末尾换行符作特殊处理、
取反的方括号表达式也受换行敏感匹配影响、先行约束中对圆括号及反向引用的限制，
以及采用最长或最短匹配而非首次匹配的语义。

ARE 与旧版 ERE 语法存在两点重要的不兼容，这里比较的是 7.4 之前的PostgreSQL：

- 在 ARE 中，\ 后跟字母或数字字符时，要么表示转义，要么是错误；而在以前的版本中，它只是书写该字母或数字字符的另一种方式。这应该不是什么大问题，
  因为以前的版本没有理由使用这样的序列。
- 在 ARE 中，\ 在 [] 内仍是特殊字符，因此方括号表达式内的字面 \ 必须写成 \\。

虽然这些差异对大多数应用程序来说不太可能造成问题，但如果需要，可以通过把regex_flavor设置为extended来避免它们。

9.7.3.7. 基本正则表达式

BRE 在几个方面和 ERE 不太一样。在 BRE 中，|、+和?都是普通字符并且没有与它们功能等价的东西。范围的定界符是\{和\}，而{和}本身是普通字符。嵌套的子表达式的圆括号是\(和\)，而(和)自身是普通字符。除非在 RE 开头或者是圆括号子表达式开头，^都是一个普通字符。除非在 RE 结尾或者是圆括号子表达式的结尾，$是一个普通字符。如果*出现在 RE 开头或者是圆括号封装的子表达式开头（前面可能有^），那么它是个普通字符。最后，可以用单数字的反向引用，\<和\>分别是[[:<:]]和[[:>:]]的同义词；在 BRE 中没有其它可用的转义。

9.8. 数据类型格式化函数

PostgreSQL格式化函数提供一套强大的工具用于把各种数据类型（日期/时间、整数、浮点数、数值）转换成格式化的字符串以及反过来从格式化的字符串转换成指定的数据类型。表 9.20列出了这些函数。这些函数都遵循一个公共的调用规范：第一个参数是待格式化的值，而第二个是一个定义输出或输入格式的模板。

还有一个单参数的`to_timestamp`函数；它接受一个`double precision`参数，并把 Unix 纪元（自 1970-01-01 00:00:00+00 起的秒数）转换为`timestamp with time zone`。（`Integer` Unix 纪元会被隐式转换为`double precision`。）

表 9.20. 格式化函数

Function	Return Type	描述	示例
`to_char(timestamp, text)`	`text`	将时间间隔转换为字符串	`to_char(current_timestamp, 'HH12:MI:SS')`
`to_char(interval, text)`	`text`	将时间间隔转换为字符串	`to_char(interval '15h 2m 12s', 'HH24:MI:SS')`
`to_char(int, text)`	`text`	将时间间隔转换为字符串	`to_char(125, '999')`
`to_char(double precision, text)`	`text`	将时间间隔转换为字符串	`to_char(125.8::real, '999D9')`
`to_char(numeric, text)`	`text`	将时间间隔转换为字符串	`to_char(-125.8, '999D99S')`
`to_date(text, text)`	`date`	convert string to date	`to_date('05 Dec 2000', 'DD Mon YYYY')`
`to_timestamp(text, text)`	`timestamp with time zone`	把字符串转换为时间戳	`to_timestamp('05 Dec 2000', 'DD Mon YYYY')`
`to_timestamp(double precision)`	`timestamp with time zone`	把 Unix 纪元转换为时间戳	`to_timestamp(200120400)`
`to_number(text, text)`	`numeric`	convert string to numeric	`to_number('12,454.8-', '99G999D9S')`

在一个输出模板串（对`to_char`）中，一些特定的模式会被识别，并被替换成来自待格式化值的、格式恰当的数据。任何不属于模板模式的文本都简单地照字面拷贝。同样，在一个输入模板串里（对除`to_char`以外的函数），模板模式标识输入数据串中要查看的部分以及要在其中找到的值。

表 9.21展示了可以用于格式化日期和时间值的模板模式。

表 9.21. 用于日期/时间格式化的模板模式

Pattern	描述
`HH`	一天中的小时（01-12）
`HH12`	一天中的小时（01-12）
`HH24`	一天中的小时（00-23）
`MI`	分钟（00-59）
`SS`	秒（00-59）
`MS`	毫秒（000-999）

Pattern	描述
US	微秒（000000-999999）
SSSS	自午夜起的秒数（0-86399）
AM or A.M. or PM or P.M.	上/下午指示符（大写）
am or a.m. or pm or p.m.	上/下午指示符（小写）
Y,YYY	带逗号的年（4 位或者更多位）
YYYY	年（4 位或者更多位）
YYY	年的最后 3 位数字
YY	年的最后 2 位数字
Y	年的最后 1 位数字
IYYY	ISO 8601 周编号方式的年（4 位或更多位）
IYY	ISO 8601 周编号方式的年的最后 3 位数字
IY	ISO 8601 周编号方式的年的最后 2 位数字
I	ISO 8601 周编号方式的年的最后 1 位数字
BC or B.C. or AD or A.D.	纪元指示符（大写）
bc or b.c. or ad or a.d.	纪元指示符（小写）
MONTH	大写的月份全称（空格补齐到 9 字符）
Month	首字母大写的月份全称（空格补齐到 9 字符）
month	小写的月份全称（空格补齐到 9 字符）
MON	简写的大写形式的月名（3 个字符）
Mon	简写的首字母大写形式的月名（3 个字符）
mon	简写的小写形式的月名（3 个字符）
MM	月份编号（01-12）
DAY	大写的星期全称（空格补齐到 9 字符）
Day	首字母大写的星期全称（空格补齐到 9 字符）
day	小写的星期全称（空格补齐到 9 字符）
DY	大写的星期简称（3 个字符）
Dy	首字母大写的星期简称（3 个字符）
dy	小写的星期简称（3 个字符）
DDD	一年中的第几天（001-366）
DD	月内日序数（01-31）
D	一周中的日（1-7；周日是 1）
W	月内周序数（1-5）（第一周从该月第一天开始）
WW	一年中的第几周（1-53）（第一周从该年第一天开始）
IW	ISO 8601 周编号年中的第几周（01-53；该年的第一个星期四位于第 1 周）
CC	世纪（2 位数）
J	儒略日（从公元前 4714 年 11 月 24 日 UTC 午夜开始的整数日数）
Q	季度
RM	以大写罗马数字表示的月份（I-XII；I=一月）

Pattern	描述
rm	以小写罗马数字表示的月份（i-xii；i=一月）
TZ	大写形式的时区名称
tz	小写形式的时区名称

修饰符可以被应用于模板模式来修改它们的行为。例如，FMMonth就是带着FM修饰符的Month模式。表 9.22展示了可用于日期/时间格式化的修饰符模式。

表 9.22. 用于日期/时间格式化的模板模式修饰符

修饰符	描述	示例
FM prefix	填充模式（抑制填充的空格和零）	FMMonth
TH suffix	大写序数后缀	DDTH
th suffix	小写序数后缀	DDth
FX prefix	固定格式的全局选项（见用法说明）	FX Month DD Day
SP suffix	spell mode (not implemented)	DDSP

日期/时间格式化的使用注意事项：

- FM抑制那些为了使模式输出为固定宽度而本会被添加的前导零和尾随空格。
- 除非使用FX选项，to_timestamp和to_date会跳过输入字符串中的多个空白。例如，to_timestamp('2000 JUN', 'YYYY MON')可以工作，但to_timestamp('2000 JUN', 'FXYYYY MON')会返回错误，因为to_timestamp期望只有一个空格。FX必须指定为模板中的第一项。
- to_char模板中允许使用普通文本，它们会被按字面输出。可以把一个子串放在双引号中，强制把它解释为字面文本，即使它包含模式关键字。例如，在'"Hello Year "YYYY'中，YYYY会被年份数据替换，但Year中的单个Y不会。
- 如果想在输出中出现双引号，必须在它前面加上反斜线，例如'\\"YYYY Month\\"'。（需要两个反斜线是因为反斜线在字符串常量中具有特殊含义。）
- 从字符串到timestamp或date的YYYY转换在处理多于 4 位的年份时有一个限制：必须在YYYY后面使用某个非数字字符或模板，否则年份总是被解释为 4 位数字。例如（对于年份20000）：to_date('200001131', 'YYYYMMDD')会被解释为 4 位年份；应在年份后使用非数字分隔符，例如to_date('20000-1131', 'YYYY-MMDD')或to_date('20000Nov31', 'YYYYMonDD')。
- 在从字符串到timestamp或date的转换中，如果存在YYY、YYYY或Y,YYY字段，则CC（世纪）字段被忽略。如果CC与YY或Y一起使用，则年份按(CC-1)*100+YY计算。
- 在从字符串到timestamp的转换中，毫秒（MS）或微秒（US）值用作小数点后的秒数部分。例如to_timestamp('12:3', 'SS:MS')不是 3 毫秒，而是 300 毫秒，因为转换把它算作 12 + 0.3 秒。这意味着对于格式SS:MS，输入值12:3、12:30和12:300指定的是相同的毫秒数。要得到三毫秒，必须使用12:003，转换把它算作 12 + 0.003 = 12.003 秒。

  下面是一个更复杂的例子：to_timestamp('15:12:02.020.001230', 'HH:MI:SS.MS.US')是 15 小时 12 分 2 秒 + 20 毫秒 + 1230 微秒 = 2.021230 秒。
- to_char的星期编号（见'D'格式化模式）与extract函数的编号不同。

- `to_char`(interval)把`HH`和`HH12`格式化为一天之内的小时数，而`HH24`可以输出超过一天的小时数，例如 >24。

表 9.23展示了可以用于格式化数值的模板模式。

表 9.23. 用于数值格式化的模板模式

Pattern	描述
`9`	数位（非有效位可以被省略）
`0`	数位（即便是非有效位也不会被省略）
`.` (period)	匹配任意单个字符
`,` (comma)	group (thousand) separator
`PR`	尖括号内的负值
`S`	紧贴数值的正负号（使用区域设置）
`L`	货币符号（使用区域设置）
`D`	decimal point (uses locale)
`G`	分组分隔符（使用区域设置）
`MI`	指定位置上的负号（如果数字 < 0）
`PL`	在指定位置的正号（如果数字 > 0）
`SG`	在指定位置的正/负号
`RN`	罗马数字（输入在 1 和 3999 之间）
`TH` or `th`	ordinal number suffix
`V`	移动指定位数（参阅注解）
`EEEE`	科学记数法（未实现）

数值格式化的使用注意事项：

- 用`SG`、`PL`或`MI`格式化的符号不固定在数字旁边；例如，`to_char(-12, 'MI9999')`产生`'-12'`，而`to_char(-12, 'S9999')`产生`'  -12'`。Oracle 的实现不允许把`MI`放在`9`之前，而是要求`9`在`MI`之前。
- `9`产生的值的位数与`9`的个数相同。如果没有可用的数字，则输出一个空格。
- `TH`不转换小于零的值，也不转换小数。
- `PL`、`SG` 和 `TH` 是 PostgreSQL 的扩展。
- `V`实际上把输入值乘以`10^n`，其中*n*是`V`后面数字的个数。`to_char`不支持把`V`与小数点组合使用（例如，不允许`99.9V99`）。

表 9.24展示了一些使用`to_char`函数的示例。

表 9.24. `to_char`示例

Expression	Result
`to_char(current_timestamp, 'Day, DD HH12:MI:SS')`	`'Tuesday  , 06  05:39:18'`
`to_char(current_timestamp, 'FMDay, FMDD  HH12:MI:SS')`	`'Tuesday, 6  05:39:18'`
`to_char(-0.1, '99.99')`	`'  -.10'`
`to_char(-0.1, 'FM9.99')`	`'-.1'`

Expression	Result
to_char(0.1, '0.9')	' 0.1'
to_char(12, '9990999.9')	' 0012.0'
to_char(12, 'FM9990999.9')	'0012.'
to_char(485, '999')	' 485'
to_char(-485, '999')	'-485'
to_char(485, '9 9 9')	' 4 8 5'
to_char(1485, '9,999')	' 1,485'
to_char(1485, '9G999')	' 1 485'
to_char(148.5, '999.999')	' 148.500'
to_char(148.5, 'FM999.999')	'148.5'
to_char(148.5, 'FM999.990')	'148.500'
to_char(148.5, '999D999')	' 148,500'
to_char(3148.5, '9G999D999')	' 3 148,500'
to_char(-485, '999S')	'485-'
to_char(-485, '999MI')	'485-'
to_char(485, '999MI')	'485 '
to_char(485, 'FM999MI')	'485'
to_char(485, 'PL999')	'+485'
to_char(485, 'SG999')	'+485'
to_char(-485, 'SG999')	'-485'
to_char(-485, '9SG99')	'4-85'
to_char(-485, '999PR')	'<485>'
to_char(485, 'L999')	'DM 485
to_char(485, 'RN')	' CDLXXXV'
to_char(485, 'FMRN')	'CDLXXXV'
to_char(5.2, 'FMRN')	'V'
to_char(482, '999th')	' 482nd'
to_char(485, '"Good number:"999')	'Good number: 485'
to_char(485.8, '"Pre:"999" Post:" .999')	'Pre: 485 Post: .800'
to_char(12, '99V999')	' 12000'
to_char(12.4, '99V999')	' 12400'
to_char(12.45, '99V9')	' 125'

9.9. 日期/时间函数和操作符

表 9.26展示了可用于处理日期/时间值的函数，其细节在随后的小节中描述。表 9.25演示了基本算术操作符（+、*等）的行为。而与格式化相关的函数，可以参考第 9.8 节。你应当熟悉第 8.5 节中的日期/时间数据类型的背景知识。

所有下文描述的接受time或timestamp
输入的函数和操作符实际上都有两种变体：一种接收time with time zone或timestamp

`with time zone`，另外一种接受`time without time zone`或者 `timestamp without time zone`。为了简化，这些变种没有被独立地展示。此外，+和*操作符都是可交换的操作符对（例如，date + integer 和 integer + date）；我们只显示每一对中的一个。

表 9.25. 日期/时间操作符

Operator	示例	Result
+	`date '2001-09-28' + integer '7'`	`date '2001-10-05'`
+	`date '2001-09-28' + interval '1 hour'`	`timestamp '2001-09-28 01:00:00'`
+	`date '2001-09-28' + time '03:00'`	`timestamp '2001-09-28 03:00:00'`
+	`interval '1 day' + interval '1 hour'`	`interval '1 day 01:00:00'`
+	`timestamp '2001-09-28 01:00' + interval '23 hours'`	`timestamp '2001-09-29 00:00:00'`
+	`time '01:00' + interval '3 hours'`	`time '04:00:00'`
-	`- interval '23 hours'`	`interval '-23:00:00'`
-	`date '2001-10-01' - date '2001-09-28'`	`integer '3'` (days)
-	`date '2001-10-01' - integer '7'`	`date '2001-09-24'`
-	`date '2001-09-28' - interval '1 hour'`	`timestamp '2001-09-27 23:00:00'`
-	`time '05:00' - time '03:00'`	`interval '02:00:00'`
-	`time '05:00' - interval '2 hours'`	`time '03:00:00'`
-	`timestamp '2001-09-28 23:00' - interval '23 hours'`	`timestamp '2001-09-28 00:00:00'`
-	`interval '1 day' - interval '1 hour'`	`interval '1 day -01:00:00'`
-	`timestamp '2001-09-29 03:00' - timestamp '2001-09-27 12:00'`	`interval '1 day 15:00:00'`
*	`900 * interval '1 second'`	`interval '00:15:00'`
*	`21 * interval '1 day'`	`interval '21 days'`
*	`double precision '3.5' * interval '1 hour'`	`interval '03:30:00'`
/	`interval '1 hour' / double precision '1.5'`	`interval '00:40:00'`

表 9.26. 日期/时间函数

Function	`age(timestamp, timestamp)`
Return Type	`interval`

描述	参数相减，产生一个使用年和月表示的“符号化”结果
示例	`age(timestamp '2001-04-10', timestamp '1957-06-13')`
Result	`43 years 9 mons 27 days`
Function	`age(timestamp)`
Return Type	`interval`
描述	从 `current_date`（午夜）减去
示例	`age(timestamp '1957-06-13')`
Result	`43 years 8 mons 3 days`
Function	`current_date`
Return Type	`date`
描述	今天的日期；见第 9.9.4 节
示例	
Result	
Function	`current_time`
Return Type	`time with time zone`
描述	当日的时刻；见第 9.9.4 节
示例	
Result	
Function	`current_timestamp`
Return Type	`timestamp with time zone`
描述	日期和时间；见第 9.9.4 节
示例	
Result	
Function	`date_part(text, timestamp)`
Return Type	`double precision`
描述	获取子字段（等价于 `extract`）；参见第 9.9.1 节
示例	`date_part('hour', timestamp '2001-02-16 20:38:40')`
Result	`20`
Function	`date_part(text, interval)`
Return Type	`double precision`
描述	获取子字段（等价于 `extract`）；参见第 9.9.1 节
示例	`date_part('month', interval '2 years 3 months')`
Result	`3`
Function	`date_trunc(text, timestamp)`
Return Type	`timestamp`

描述	截断到指定精度；另见第 9.9.2 节
示例	`date_trunc('hour', timestamp '2001-02-16 20:38:40')`
Result	`2001-02-16 20:00:00`
Function	`extract(field from timestamp)`
Return Type	`double precision`
描述	获取子字段；参见第 9.9.1 节
示例	`extract(hour from timestamp '2001-02-16 20:38:40')`
Result	`20`
Function	`extract(field from interval)`
Return Type	`double precision`
描述	获取子字段；参见第 9.9.1 节
示例	`extract(month from interval '2 years 3 months')`
Result	`3`
Function	`isfinite(timestamp)`
Return Type	`boolean`
描述	测试是否为有限时间戳（不等于无穷）
示例	`isfinite(timestamp '2001-02-16 21:28:30')`
Result	`true`
Function	`isfinite(interval)`
Return Type	`boolean`
描述	测试时间间隔是否有限
示例	`isfinite(interval '4 hours')`
Result	`true`
Function	`justify_hours(interval)`
Return Type	`interval`
描述	调整 interval，把 24 小时的时间段表示为天
示例	`justify_hours(interval '24 hours')`
Result	`1 day`
Function	`justify_days(interval)`
Return Type	`interval`
描述	调整 interval，把 30 天的时间段表示为月
示例	`justify_days(interval '30 days')`
Result	`1 month`

Function	`localtime`
Return Type	`time`
描述	当日的时刻；见第 9.9.4 节
示例	
Result	
Function	`localtimestamp`
Return Type	`timestamp`
描述	日期和时间；见第 9.9.4 节
示例	
Result	
Function	`now()`
Return Type	`timestamp with time zone`
描述	当前的日期和时间（等效于`current_timestamp`）；见第 9.9.4 节
示例	
Result	
Function	`timeofday()`
Return Type	`text`
描述	当前的日期和时间；见第 9.9.4 节
示例	
Result	

如果你同时使用`justify_hours`和`justify_days`，最好先用`justify_hours`，这样任何多出来的天数就会被包含在`justify_days`的计算中。

除了这些函数以外，还支持 SQL 操作符`OVERLAPS`：

```
(start1, end1) OVERLAPS (start2, end2)
(start1, length1) OVERLAPS (start2, length2)
```

当两个时间段（由其端点定义）重叠时，该表达式产生真；不重叠时产生假。端点可以指定为一对日期、时间或时间戳；也可以是一个日期、时间或时间戳后跟一个interval。

```
SELECT (DATE '2001-02-16', DATE '2001-12-21') OVERLAPS
       (DATE '2001-10-30', DATE '2002-10-30');
Result: true
SELECT (DATE '2001-02-16', INTERVAL '100 days') OVERLAPS
       (DATE '2001-10-30', DATE '2002-10-30');
Result: false
```

当把一个`interval`值加到某个`timestamp with time zone`值上（或从中减去一个`interval`值）时，天数部分会按指定的天数推进或退回`timestamp with time zone`的日期。当跨越夏令时变化时（会话时区设置为识别夏令时的时区），这意味着`interval '1 day'`不一定等于`interval '24 hours'`。例如，将会话时区设置为`CST7CDT`时，`timestamp`

with time zone '2005-04-02 12:00-07' + interval '1 day' 会产生timestamp with time zone '2005-04-03 12:00-06'，而把interval '24 hours'加到同一个初始timestamp with time zone上会产生timestamp with time zone '2005-04-03 13:00-06'，因为在时区CST7CDT中，2005-04-03 02:00发生了夏令时变更。

9.9.1. EXTRACT, date_part

EXTRACT(*field* FROM *source*)

extract函数从日期/时间值中提取年份、小时等子字段。*source*必须是timestamp、time或interval类型的值表达式。（date类型的表达式会转换为timestamp，因此也可以使用。）*field*是一个标识符或字符串，用于选择从源值中提取的字段。extract函数返回的值的类型为double precision。以下是有效的字段名称：

century

世纪

```
SELECT EXTRACT(CENTURY FROM TIMESTAMP '2000-12-16 12:21:13');
Result: 20
SELECT EXTRACT(CENTURY FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 21
```

第一个世纪从公元 0001-01-01 00:00:00 开始，尽管当时的人并不知道这一点。这个定义适用于所有使用格里高利历的国家。没有世纪编号 0，你从 -1 世纪直接到 1 世纪。如果对此有异议，请把你的意见寄到：罗马，圣彼得大教堂，教皇收。

PostgreSQL 8.0 之前的版本不遵循通常的世纪编号，而只是返回年份数字除以 100 的结果。

day

（月中的）日字段（1 - 31）

```
SELECT EXTRACT(DAY FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 16
```

decade

年份字段除以 10

```
SELECT EXTRACT(DECADE FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 200
```

dow

一周中的日（0 - 6；周日是 0）（仅用于timestamp值）

```
SELECT EXTRACT(DOW FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 5
```

注意，extract的星期编号与to_char(..., 'D')函数的编号不同。

doy

一年中的第几天（1 - 365/366）（仅用于timestamp值）

```
SELECT EXTRACT(DOY FROM TIMESTAMP '2001-02-16 20:38:40');
```

```
Result: 47
```

epoch

对date和timestamp值，是自 1970-01-01 00:00:00 UTC 以来的秒数（可以为负）；对interval值，是该 interval 中的总秒数

```
SELECT EXTRACT(EPOCH FROM TIMESTAMP WITH TIME ZONE '2001-02-16 20:38:40.12-08');
Result: 982384720

SELECT EXTRACT(EPOCH FROM INTERVAL '5 days 3 hours');
Result: 442800
```

下面说明如何把 epoch 值转换回时间戳：

```
SELECT TIMESTAMP WITH TIME ZONE 'epoch' + 982384720 * INTERVAL '1 second';
```

hour

小时域（0 - 23）

```
SELECT EXTRACT(HOUR FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 20
```

microseconds

秒字段（包含小数部分）乘以 1 000 000；注意这包含完整的秒

```
SELECT EXTRACT(MICROSECONDS FROM TIME '17:12:28.5');
Result: 28500000
```

millennium

千年域

```
SELECT EXTRACT(MILLENNIUM FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 3
```

1900 年代属于第二个千年。第三个千年始于 2001 年 1 月 1 日。

PostgreSQL 8.0 之前的版本不遵循通常的千年编号，而只是返回年份数字除以 1000 的结果。

milliseconds

秒字段，包括小数部分，乘以 1000。注意这包含完整的秒。

```
SELECT EXTRACT(MILLISECONDS FROM TIME '17:12:28.5');
Result: 28500
```

minute

分钟字段（0 - 59）

```
SELECT EXTRACT(MINUTE FROM TIMESTAMP '2001-02-16 20:38:40');
```

Result: 38

month

对timestamp值，是一年中的月份（1 - 12）；对interval值，是月数模 12（0 - 11）

```
SELECT EXTRACT(MONTH FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 2

SELECT EXTRACT(MONTH FROM INTERVAL '2 years 3 months');
Result: 3

SELECT EXTRACT(MONTH FROM INTERVAL '2 years 13 months');
Result: 1
```

quarter

该日期所在的一年中的季度（1 - 4）（仅用于timestamp值）

```
SELECT EXTRACT(QUARTER FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 1
```

second

秒字段（包含小数部分）（0 - 59[1]）

```
SELECT EXTRACT(SECOND FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 40

SELECT EXTRACT(SECOND FROM TIME '17:12:28.5');
Result: 28.5
```

timezone

相对 UTC 的时区偏移，以秒计。正值对应 UTC 以东的时区，负值对应 UTC 以西的时区。

timezone_hour

时区偏移的小时部分

timezone_minute

时区偏移的分钟部分

week

该日期在一年中的周号。按照定义（ISO 8601），一年的第 1 周包含该年的 1 月 4 日。（ISO-8601 的周从星期一开始。）换句话说，一年的第一个星期四属于该年的第 1 周。（仅用于timestamp值）

因此，1 月上旬的日期有可能属于上一年的第 52 或 53 周。例如，2005-01-01属于 2004 年的第 53 周，而2006-01-01属于 2005 年的第 52 周。

```
SELECT EXTRACT(WEEK FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 7
```

[1]如果操作系统实现了闰秒则为 60

year

年份字段。请记住没有公元 0 年，因此从公元（AD）年份减去公元前（BC）年份时应小心。

```
SELECT EXTRACT(YEAR FROM TIMESTAMP '2001-02-16 20:38:40');
Result: 2001
```

extract函数主要的用途是做计算性处理。对于用于显示的日期/时间值格式化，参阅第 9.8 节。

date_part函数仿照传统的Ingres实现，后者对应SQL标准的extract函数：

date_part('*field*', *source*)

注意，此处的*field*参数必须是字符串值，而不能是名称。date_part
的有效字段名与extract相同。

```
SELECT date_part('day', TIMESTAMP '2001-02-16 20:38:40');
Result: 16

SELECT date_part('hour', INTERVAL '4 hours 3 minutes');
Result: 4
```

9.9.2. date_trunc

date_trunc函数在概念上和用于数字的trunc函数类似。

date_trunc('*field*', *source*)

*source*是timestamp或interval类型的值表达式。（类型为date和time
的值会自动转换为timestamp或interval，分别对应这两种输入类型。）*field*
选择输入值的截断精度。返回值的类型是timestamp或interval，
所有小于所选精度的字段都设为零（日和月则设为一）。

*field*的有效值是：

```
microseconds
milliseconds
second
minute
hour
day
week
month
quarter
year
decade
century
millennium
```

示例：

```
SELECT date_trunc('hour', TIMESTAMP '2001-02-16 20:38:40');
结果: 2001-02-16 20:00:00

SELECT date_trunc('year', TIMESTAMP '2001-02-16 20:38:40');
结果: 2001-01-01 00:00:00
```

9.9.3. AT TIME ZONE

AT TIME ZONE结构允许把时间戳转换到不同的时区。表 9.27展示了它的各种变体。

表 9.27. AT TIME ZONE 变体

Expression	Return Type	描述
timestamp without time zone AT TIME ZONE *zone*	timestamp with time zone	将给定的不带时区时间戳视为指定时区中的时间
timestamp with time zone AT TIME ZONE *zone*	timestamp without time zone	将给定的带时区时间戳转换为新时区的时间，结果不带时区标识
time with time zone AT TIME ZONE *zone*	time with time zone	将给定的带时区时间转换为新时区的时间

在这些表达式里，所需的时区*zone*可以指定为文本字符串（例如'PST'），
也可以指定为一个间隔（例如INTERVAL '-08:00'）。在文本情况下，可用的时区名称是表 B.6或表 B.4中所示的那些。

示例（假定本地时区为PST8PDT）：

```
SELECT TIMESTAMP '2001-02-16 20:38:40' AT TIME ZONE 'MST';
结果: 2001-02-16 19:38:40-08
SELECT TIMESTAMP WITH TIME ZONE '2001-02-16 20:38:40-05' AT TIME ZONE
 'MST';
结果: 2001-02-16 18:38:40
```

第一个示例取一个不带时区的时间戳，把它解释为 MST 时间（UTC-7），然后转换成 PST（UTC-8）显示。第二个示例取一个以 EST（UTC-5）指定的时间戳，把它转换成 MST（UTC-7）的本地时间。

函数timezone(*zone*, *timestamp*)等效于符合 SQL 标准的结构*timestamp* AT TIME ZONE *zone*。

9.9.4. 当前日期/时间

下列函数可用于获取当前日期和/或时间：

```
CURRENT_DATE
CURRENT_TIME
CURRENT_TIMESTAMP
CURRENT_TIME (precision)
CURRENT_TIMESTAMP (precision)
LOCALTIME
LOCALTIMESTAMP
LOCALTIME (precision)
LOCALTIMESTAMP (precision)
```

CURRENT_TIME和CURRENT_TIMESTAMP返回带时区的值；LOCALTIME和LOCALTIMESTAMP返回不带时区的值。

CURRENT_TIME、CURRENT_TIMESTAMP、LOCALTIME和 LOCALTIMESTAMP可以有选择地接受一个精度参数，该精度会使结果的秒字段舍入到指定的小数位数。如果没有精度参数，
结果将给出可用的全部精度。

> **注意**
>
> 在PostgreSQL 7.2 之前，精度参数是没有实现的，结果总是以整数秒给出。

下面是一些示例：

```
SELECT CURRENT_TIME;
结果： 14:39:53.662522-05
SELECT CURRENT_DATE;
结果： 2001-12-23
SELECT CURRENT_TIMESTAMP;
结果： 2001-12-23 14:39:53.662522-05
SELECT CURRENT_TIMESTAMP(2);
结果： 2001-12-23 14:39:53.66-05
SELECT LOCALTIMESTAMP;
结果： 2001-12-23 14:39:53.662522
```

`now()`是PostgreSQL中与`CURRENT_TIMESTAMP`等价的传统函数。

重要的是要知道，`CURRENT_TIMESTAMP`及相关函数返回的是当前事务的开始时间；它们的值在事务运行的整个期间内都不改变。这被认为是一个特性：目的是为了允许一个事务在“当前”时间上有一致的概念，这样在同一个事务里的多个修改可以保持同样的时间戳。

> **注意**
>
> 其他数据库系统可能会更频繁地推进这些值。

还有函数`timeofday()`，它返回墙上时钟时间，并且在事务期间前进。由于历史原因，`timeofday()`返回一个`text`字符串而不是`timestamp`值：

```
SELECT timeofday();
结果: Sat Feb 17 19:07:32.000126 2001 EST
```

所有日期/时间数据类型也都接受特殊字面值`now`来指定当前日期和时间。因此，下面三种写法都返回相同的结果：

```
SELECT CURRENT_TIMESTAMP;
SELECT now();
SELECT TIMESTAMP 'now';  -- incorrect for use with DEFAULT
```

> **提示**
>
> 在创建表时指定`DEFAULT`子句的情况下，你不希望使用第三种形式。系统将在分析这个常量的时候把`now`转换为一个`timestamp`，这样需要默认值时就会得到创建表的时间！而前两种形式要到实际使用默认值的时候才被计算，因为它们是函数调用。因此它们可以给出每次插入行的时刻。

9.10. 几何函数和操作符

几何类型point、box、lseg、line、path、polygon和circle
有大量内置支持函数和操作符，如表 9.28、表 9.29和表 9.30中所示。

> **小心**
>
> 注意，“相同”操作符 ~= 表示 point、box、polygon 和 circle 类型通常意义上的相等。其中一些类型也有 = 操作符，但 = 仅比较面积是否相等。这些类型的其他标量比较操作符（<= 等）同样比较面积。

表 9.28. 几何操作符

Operator	描述	示例
+	Translation	box '((0,0),(1,1))' + point '(2.0,0)'
–	Translation	box '((0,0),(1,1))' - point '(2.0,0)'
*	Scaling/rotation	box '((0,0),(1,1))' * point '(2.0,0)'
/	Scaling/rotation	box '((0,0),(2,2))' / point '(2.0,0)'
#	相交的点或矩形框	'((1,-1),(-1,1))' # '((1,1),(-1,-1))'
#	相交的点或矩形框	# '((1,0),(0,1),(-1,0))'
@-@	长度或周长	@-@ path '((0,0),(1,0))'
@@	中心	@@ circle '((0,0),10)'
##	第二个操作数上距离第一个操作数最近的点	point '(0,0)' ## lseg '((2,0),(0,2))'
<->	两者之间的距离	circle '((0,0),1)' <-> circle '((5,0),1)'
&&	是否重叠？（有一个公共点即为真。）	box '((0,0),(1,1))' && box '((0,0),(2,2))'
<<	是否严格位于左侧？	circle '((0,0),1)' << circle '((5,0),1)'
>>	是否严格位于右侧？	circle '((5,0),1)' >> circle '((0,0),1)'
&<	是否未超出对方的右边界？	box '((0,0),(1,1))' &< box '((0,0),(2,2))'
&>	是否未超出对方的左边界？	box '((0,0),(3,3))' &> box '((0,0),(2,2))'
<<|	是否严格位于下方？	box '((0,0),(3,3))' <<| box '((3,4),(5,5))'
|>>	是否严格位于上方？	box '((3,4),(5,5))' |>> box '((0,0),(3,3))'
&<|	是否未超出对方的上边界？	box '((0,0),(1,1))' &<| box '((0,0),(2,2))'

Operator	描述	示例
`|&>`	是否未超出对方的下边界?	`box '((0,0),(3,3))' |&> box '((0,0),(2,2))'`
`<^`	是否位于下方（允许接触）?	`circle '((0,0),1)' <^ circle '((0,5),1)'`
`>^`	是否位于上方（允许接触）?	`circle '((0,5),1)' >^ circle '((0,0),1)'`
`?#`	是否相交?	`lseg '((-1,0),(1,0))' ?# box '((-2,-2),(2,2))'`
`?-`	是否水平?	`?- lseg '((-1,0),(1,0))'`
`?-`	是否水平?	`point '(1,0)' ?- point '(0,0)'`
`?|`	是否竖直?	`?| lseg '((-1,0),(1,0))'`
`?|`	是否竖直?	`point '(0,1)' ?| point '(0,0)'`
`?-|`	是否互相垂直?	`lseg '((0,0),(0,1))' ?-| lseg '((0,0),(1,0))'`
`?||`	是否平行?	`lseg '((-1,0),(1,0))' ?|| lseg '((-1,2),(1,2))'`
`~`	是否包含?	`circle '((0,0),2)' ~ point '(1,1)'`
`@`	是否位于内部或边界上?	`point '(1,1)' @ circle '((0,0),2)'`
`~=`	是否相同?	`polygon '((0,0),(1,1))' ~= polygon '((1,1),(0,0))'`

表 9.29. 几何函数

Function	Return Type	描述	示例
`area(object)`	`double precision`	面积	`area(box '((0,0),(1,1))')`
`center(object)`	`point`	中心	`center(box '((0,0),(1,2))')`
`diameter(circle)`	`double precision`	圆的直径	`diameter(circle '((0,0),2.0)')`
`height(box)`	`double precision`	矩形框的竖直尺寸	`height(box '((0,0),(1,1))')`
`isclosed(path)`	`boolean`	是否为闭合路径?	`isclosed(path '((0,0),(1,1),(2,0))')`
`isopen(path)`	`boolean`	是否为开放路径?	`isopen(path '[(0,0),(1,1),(2,0)]')`
`length(object)`	`double precision`	长度	`length(path '((-1,0),(1,0))')`
`npoints(path)`	`int`	点数	`npoints(path '[(0,0),(1,1),(2,0)]')`

Function	Return Type	描述	示例
npoints(polygon)	int	点数	npoints(polygon '((1,1),(0,0))')
pclose(path)	path	将路径转换为闭合路径	pclose(path '[(0,0),(1,1),(2,0)]')
popen(path)	path	将路径转换为开放路径	popen(path '((0,0),(1,1),(2,0))')
radius(circle)	double precision	圆的半径	radius(circle '((0,0),2.0)')
width(box)	double precision	矩形框的水平尺寸	width(box '((0,0),(1,1))')

表 9.30. 几何类型转换函数

Function	Return Type	描述	示例
box(circle)	box	将点转换为矩形框	box(circle '((0,0),2.0)')
box(point, point)	box	将点转换为矩形框	box(point '(0,0)', point '(1,1)')
box(polygon)	box	将点转换为矩形框	box(polygon '((0,0),(1,1),(2,0))')
circle(box)	circle	由圆心和半径构造圆	circle(box '((0,0),(1,1))')
circle(point, double precision)	circle	由圆心和半径构造圆	circle(point '(0,0)', 2.0)
circle(polygon)	circle	由圆心和半径构造圆	circle(polygon '((0,0),(1,1),(2,0))')
lseg(box)	lseg	由点构造线段	lseg(box '((-1,0),(1,0))')
lseg(point, point)	lseg	由点构造线段	lseg(point '(-1,0)', point '(1,0)')
path(polygon)	path	polygon to path	path(polygon '((0,0),(1,1),(2,0))')
point(double precision, double precision)	point	交集	point(23.4, -44.5)
point(box)	point	交集	point(box '((-1,0),(1,0))')
point(circle)	point	交集	point(circle '((0,0),2.0)')
point(lseg)	point	交集	point(lseg '((-1,0),(1,0))')
point(polygon)	point	交集	point(polygon '((0,0),(1,1),(2,0))')
polygon(box)	polygon	将圆转换为 12 个顶点的多边形	polygon(box '((0,0),(1,1))')

Function	Return Type	描述	示例
polygon(circle)	polygon	将圆转换为 12 个顶点的多边形	polygon(circle '((0,0),2.0)')
polygon(*npts*, circle)	polygon	圆转换为*npts*个点的多边形	polygon(12, circle '((0,0),2.0)')
polygon(path)	polygon	将圆转换为 12 个顶点的多边形	polygon(path '((0,0),(1,1),(2,0))')

可以把一个point当作下标为 0 和 1 的数组，访问它的两个数值分量。例如，如果t.p是一个point列，那么SELECT p[0] FROM t检索 X 坐标而 UPDATE t SET p[1] = ...改变 Y 坐标。同样，box或者lseg类型的值可以当作两个point值组成的数组看待。

area 函数适用于 box、circle 和 path 类型。对于 path 数据类型，只有当 path 中的点构成的路径不自相交时，area 函数才能工作。例如，path '((0,0),(0,1),(2,1),(2,2),(1,2),(1,0),(0,0))'::PATH 无法使用；但下面这个视觉上相同的 path '((0,0),(0,1),(1,1),(1,2),(2,2),(2,1),(1,1),(1,0),(0,0))'::PATH 可以使用。如果难以理解自相交与非自相交 path 的区别，可以把上述两条 path 并排画在方格纸上。

9.11. 网络地址函数和操作符

表 9.31列出了可用于 cidr 和 inet 类型的操作符。操作符<<、<<=、>>和>>=测试子网包含关系。它们只考虑两个地址的网络部分（忽略主机部分），
并判断一个网络是否与另一个网络相同，或是另一个网络的子网。

表 9.31. cidr 和 inet 操作符

Operator	描述	示例
<	小于	inet '192.168.1.5' < inet '192.168.1.6'
<=	小于或等于	inet '192.168.1.5' <= inet '192.168.1.5'
=	等于	inet '192.168.1.5' = inet '192.168.1.5'
>=	大于或等于	inet '192.168.1.5' >= inet '192.168.1.5'
>	大于	inet '192.168.1.5' > inet '192.168.1.4'
<>	不等于	inet '192.168.1.5' <> inet '192.168.1.4'
<<	是否严格位于左侧？	inet '192.168.1.5' << inet '192.168.1/24'
<<=	被包含于或等于	inet '192.168.1/24' <<= inet '192.168.1/24'
>>	是否严格位于右侧？	inet '192.168.1/24' >> inet '192.168.1.5'
>>=	包含或等于	inet '192.168.1/24' >>= inet '192.168.1/24'

表 9.32列出了可用于 cidr 和 inet 类型的函数。abbrev、host 和 text 函数主要用于提供其他显示格式。你可以用普通的类型转换语法把一个文本值转换为inet：inet(*expression*)或*colname*::inet。

表 9.32. cidr 和 inet 函数

Function	broadcast(inet)
Return Type	inet
描述	broadcast address for network
示例	broadcast('192.168.1.5/24')
Result	192.168.1.255/24
Function	host(inet)
Return Type	text
描述	以 text 形式抽取 IP 地址
示例	host('192.168.1.5/24')
Result	192.168.1.5
Function	masklen(inet)
Return Type	int
描述	抽取网络掩码长度
示例	masklen('192.168.1.5/24')
Result	24
Function	set_masklen(inet, int)
Return Type	inet
描述	设置 inet 值的网络掩码长度
示例	set_masklen('192.168.1.5/24', 16)
Result	192.168.1.5/16
Function	netmask(inet)
Return Type	inet
描述	构造网络的网络掩码
示例	netmask('192.168.1.5/24')
Result	255.255.255.0
Function	hostmask(inet)
Return Type	inet
描述	构造网络的主机掩码
示例	hostmask('192.168.23.20/30')
Result	0.0.0.3
Function	network(inet)
Return Type	cidr
描述	抽取地址的网络部分
示例	network('192.168.1.5/24')

Result	`192.168.1.0/24`
Function	`text(inet)`
Return Type	`text`
描述	以 text 形式抽取 IP 地址和网络掩码长度
示例	`text(inet '192.168.1.5')`
Result	`192.168.1.5/32`
Function	`abbrev(inet)`
Return Type	`text`
描述	以文本表示的缩略显示格式
示例	`abbrev(cidr '10.1.0.0/16')`
Result	`10.1/16`
Function	`family(inet)`
Return Type	`int`
描述	extract family of address; `4` for IPv4, `6` for IPv6
示例	`family('::1')`
Result	`6`

表 9.33列出了可用于 `macaddr` 类型的函数。函数 `trunc(macaddr)` 返回将最后 3 个字节置零的 MAC 地址。可以用余下的前缀确定制造商。源码发行版的`contrib/mac`目录中包含一些用于创建和维护这种关联表的工具。

表 9.33. macaddr 函数

Function	`trunc(macaddr)`
Return Type	`macaddr`
描述	把最后 3 个字节设置为零
示例	`trunc(macaddr '12:34:56:78:90:ab')`
Result	`12:34:56:00:00:00`

`macaddr`类型还支持用于字典序排序的标准关系操作符（`>`、`<=`等）。

9.12. 序列操作函数

本节描述PostgreSQL中用于操作序列对象（也称为序列生成器，或简称序列）的函数。
序列对象是用`CREATE SEQUENCE`创建的特殊单行表。序列对象通常用于为表中的行生成惟一标识符。在表 9.34中列出的序列函数，提供了简单的、多用户安全方法，用于从序列对象中获取连续的序列值。

表 9.34. 序列函数

Function	Return Type	描述
`nextval(regclass)`	`bigint`	推进序列并返回新值

Function	Return Type	描述
`currval(regclass)`	`bigint`	返回最近对指定序列调用 `nextval` 得到的值
`lastval()`	`bigint`	返回最近一次用`nextval`获取的值
`setval(regclass, bigint)`	`bigint`	设置序列的当前值
`setval(regclass, bigint, boolean)`	`bigint`	设置序列的当前值和`is_called`标志

序列函数要操作的序列由一个`regclass`参数指定，该参数就是序列在`pg_class`系统目录中的OID。不过，不必手动查找 OID，因为`regclass`数据类型的输入转换函数会替你完成这项工作。只需将序列名用单引号括起来，使其看上去像一个字面常量。为兼容普通SQL名称的处理方式，除非序列名被双引号括起，否则字符串会转换为小写。因此：

`nextval('foo')`	操作序列 *foo*
`nextval('FOO')`	操作序列 *foo*
`nextval('"Foo"')`	操作序列 *Foo*

必要时，序列名可以带模式限定：

`nextval('myschema.foo')`	操作 *myschema.foo*
`nextval('"myschema".foo')`	同上
`nextval('foo')`	在搜索路径中查找 *foo*

参见第 8.12 节以获取更多有关以下类型的信息：`regclass`。

> **注意**
>
> 在PostgreSQL 8.1 之前，序列函数的参数类型是`text`，而不是`regclass`；上述从文本字符串到 OID 值的转换会在每次调用时于运行期间执行。为保持向后兼容，这种能力仍然存在，但内部现在会在调用函数前，通过从`text`到`regclass`的隐式强制转换来处理。
>
> 将序列函数的参数写成不加修饰的字符串字面量时，它会成为以下类型的常量：`regclass`。由于它实际上只是一个 OID，即使后来发生重命名、模式变更等情况，它仍会指向最初标识的序列。这种“早绑定”行为通常适用于列默认值和视图中的序列引用。但有时你可能希望采用“后绑定”，在运行时解析序列引用。要获得后绑定行为，应强制将常量存储为`text`常量，而不是`regclass`：
>
> `nextval('foo'::text)` *foo*运行时被查找
>
> 请注意，后绑定是下列旧版本唯一支持的行为：PostgreSQL8.1 之前的版本，因此可能需要这样做来保留旧应用的语义。
>
> 当然，序列函数的参数既可以是常量，也可以是表达式。如果它是文本表达式，隐式强制转换就会导致运行时查找。

可用的序列函数有：

`nextval`

把序列对象推进到它的下一个值并返回该值。
这是以原子方式完成的：即使多个会话并发地执行`nextval`，
每个会话也会安全地获得一个不同的序列值。

`currval`

返回当前会话中`nextval`最近一次为该序列获取的值。
（如果在本会话中从未对该序列调用过`nextval`，则会报告错误。）
由于返回的是一个会话局部的值，无论其他会话在当前会话之后是否执行过`nextval`，
它都给出可预测的答案。

`lastval`

返回当前会话中`nextval`最近一次返回的值。该函数与`currval`相同，
只是它不以序列名为参数，而是获取当前会话中`nextval`最近使用的序列的值。
如果当前会话尚未调用过`nextval`，调用`lastval`是错误的。

`setval`

重置序列对象的计数值。双参数形式把序列的`last_value`字段设置为指定值，并把其`is_called`字段设置为`true`，这意味着下一次`nextval`将在返回值之前推进序列。
在三参数形式中，`is_called`可以设置为`true`或`false`。如果设置为`false`，下一次`nextval`将恰好返回指定的值，序列的推进从再下一次`nextval`开始。例如，

```
SELECT setval('foo', 42);           Next nextval will return 43
SELECT setval('foo', 42, true);     Same as above
SELECT setval('foo', 42, false);    Next nextval will return 42
```

`setval`返回的结果就是其第二个参数的值。

如果一个序列对象是用默认参数创建的，那么连续调用`nextval`将返回从 1 开始的连续值。通过在CREATE SEQUENCE命令中使用特殊参数，可以得到其他行为；更多信息见其命令参考页。

重要

为了避免阻塞从同一序列获取编号的并发事务，`nextval`操作永远不会被回滚；
也就是说，一旦取出了一个值，它就被视为已使用，即使执行`nextval`
的事务随后中止也是如此。
这意味着中止的事务可能会在已分配的值序列中留下未使用的“空洞”。`setval`
操作也永远不会被回滚。

9.13. 条件表达式

本节描述在PostgreSQL中可用的SQL兼容的条件表达式。

提示

如果你的需求超过这些条件表达式的能力，
你可能会希望用一种更富表现力的编程语言写一个存储过程。

9.13.1. CASE

SQL `CASE`表达式是一种通用的条件表达式，类似于其它编程语言中的 if/else 语句：

```
CASE WHEN condition THEN result
     [WHEN ...]
     [ELSE result]
END
```

CASE子句可以用于任何表达式有效的地方。*condition*是一个返回boolean结果的表达式。如果结果为真，那么CASE表达式的值就是紧随该条件之后的*result*。如果结果为假，那么以相同的方式搜寻任何随后的WHEN子句。如果没有WHEN *condition*为真，那么 case 表达式的值就是在ELSE子句里的*result*。如果省略了ELSE子句而且没有条件匹配，结果为 null。

示例：

```
SELECT * FROM test;

 a
---
 1
 2
 3

SELECT a,
       CASE WHEN a=1 THEN 'one'
            WHEN a=2 THEN 'two'
            ELSE 'other'
       END
    FROM test;

 a | case
---+-------
 1 | one
 2 | two
 3 | other
```

所有*result*表达式的数据类型都必须可以转换成单一的输出类型。参阅第 10.5 节获取细节。

下面这个“简单”形式的CASE表达式是上述通用形式的一个变种：

```
CASE expression
    WHEN value THEN result
    [WHEN ...]
    [ELSE result]
END
```

第一个*expression*会被计算，然后与所有在WHEN子句中的每一个*value*对比，直到找到一个相等的。如果没有找到匹配的，则返回在ELSE子句中的*result*（或者 null 值）。这类似于 C 里的switch语句。

上面的示例可以用简单CASE语法来写：

```
SELECT a,
       CASE a WHEN 1 THEN 'one'
              WHEN 2 THEN 'two'
              ELSE 'other'
       END
    FROM test;
```

```
 a | case
---+-------
 1 | one
 2 | two
 3 | other
```

CASE表达式并不计算任何无助于判断结果的子表达式。例如，
下面是一个可以避免被零除错误的方法：

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

9.13.2. COALESCE

```
COALESCE(value [, ...])
```

COALESCE函数返回参数中第一个不为 null 的值。只有所有参数都为 null 时，才返回 null。它常用于在检索数据以供显示时，用默认值替换 null 值。例如：

```
SELECT COALESCE(description, short_description, '(none)') ...
```

和CASE表达式一样，COALESCE只计算确定结果所需的参数；也就是说，在第一个不为 null 的参数右边的参数不会被计算。这个 SQL 标准函数提供了类似于NVL和IFNULL的能力，它们被用在某些其他数据库系统中。

9.13.3. NULLIF

```
NULLIF(value1, value2)
```

当*value1*和*value2*相等时，NULLIF返回 null。否则它返回*value1*。这些可以用于执行前文给出的COALESCE示例的逆操作：

```
SELECT NULLIF(value, '(none)') ...
```

在这个示例中，如果*value1*是(none)，将返回 null，否则返回*value1*。

9.13.4. GREATEST and LEAST

```
GREATEST(value [, ...])

LEAST(value [, ...])
```

GREATEST和LEAST函数从由任意数量的表达式组成的列表中选取最大值或最小值。
这些表达式都必须能转换为同一个数据类型，该类型将作为结果类型（详情参见第 10.5 节）。列表中的 NULL 值会被忽略。只有所有表达式的求值结果都为 NULL 时，结果才为 NULL。

请注意GREATEST和LEAST都未包含在 SQL 标准中，但却是很常见的扩展。某些其他数据库让它们在任何参数为 NULL 时返回 NULL，而不是在所有参数都为 NULL 时才返回 NULL。

9.14. 数组函数和操作符

表 9.35列出了可用于array类型的操作符。

表 9.35. array 操作符

Operator	描述	示例	Result
=	等于	ARRAY[1.1,2.1,3.1]::int[] = ARRAY[1,2,3]	t
<>	不等于	ARRAY[1,2,3] <> ARRAY[1,2,4]	t
<	小于	ARRAY[1,2,3] < ARRAY[1,2,4]	t
>	大于	ARRAY[1,4,3] > ARRAY[1,2,4]	t
<=	小于等于	ARRAY[1,2,3] <= ARRAY[1,2,3]	t
>=	大于等于	ARRAY[1,4,3] >= ARRAY[1,4,3]	t
||	串接	ARRAY[1,2,3] || ARRAY[4,5,6]	{1,2,3,4,5,6}
||	串接	ARRAY[1,2,3] || ARRAY[[4,5,6],[7,8,9]]	{{1,2,3},{4,5,6},{7,8,9}}
||	串接	3 || ARRAY[4,5,6]	{3,4,5,6}
||	串接	ARRAY[4,5,6] || 7	{4,5,6,7}

参阅第 8.10 节获取有关数组操作符行为的更多细节。

表 9.36展示了可以用于数组类型的函数。参阅第 8.10 节获取更多信息以及使用这些函数的示例。

表 9.36. array 函数

Function	array_cat(anyarray, anyarray)
Return Type	anyarray
描述	串接两个数组
示例	array_cat(ARRAY[1,2,3], ARRAY[4,5])
Result	{1,2,3,4,5}
Function	array_append(anyarray, anyelement)
Return Type	anyarray
描述	向数组末尾追加一个元素
示例	array_append(ARRAY[1,2], 3)
Result	{1,2,3}
Function	array_prepend(anyelement, anyarray)
Return Type	anyarray
描述	向数组开头添加一个元素
示例	array_prepend(1, ARRAY[2,3])

Result	`{1,2,3}`
Function	`array_dims(anyarray)`
Return Type	`text`
描述	返回数组维度的文本表示
示例	`array_dims(ARRAY[[1,2,3], [4,5,6]])`
Result	`[1:2][1:3]`
Function	`array_lower(anyarray, int)`
Return Type	`int`
描述	返回数组指定维度的下界
示例	`array_lower(array_prepend(0, ARRAY[1,2,3]), 1)`
Result	`0`
Function	`array_upper(anyarray, int)`
Return Type	`int`
描述	返回数组指定维度的上界
示例	`array_upper(ARRAY[1,2,3,4], 1)`
Result	`4`
Function	`array_to_string(anyarray, text)`
Return Type	`text`
描述	使用给定分隔符和可选的 null 替代字符串串接数组元素
示例	`array_to_string(ARRAY[1, 2, 3], '~^~')`
Result	`1~^~2~^~3`
Function	`string_to_array(text, text)`
Return Type	`text[]`
描述	使用给定分隔符和可选的 null 替代字符串将字符串拆分为数组元素
示例	`string_to_array('xx~^~yy~^~zz', '~^~')`
Result	`{xx,yy,zz}`

9.15. 聚合函数

聚合函数从一组输入值计算单个结果。表 9.37展示了内置的聚合函数。聚合函数的特殊语法注意事项在第 4.2.7 节中说明。更多入门信息请参见第 2.7 节。

表 9.37. 聚合函数

Function	Argument Type(s)	Return Type	描述
`avg(`*`expression`*`)`	`smallint`, `int`, `bigint`, `real`, `double`	整数类型参数为`numeric`，浮点参数为`double precision`，否则与参数数据类型相同	所有输入值的平均值（算术平均）

Function	Argument Type(s)	Return Type	描述
	precision, numeric, or interval		
bit_and(*expression*)	smallint、int、bigint 或 bit	与参数数据类型相同	所有非空输入值的按位与；没有非空输入时为 null
bit_or(*expression*)	smallint、int、bigint 或 bit	与参数数据类型相同	所有非空输入值的按位或；没有非空输入时为 null
bool_and(*expression*)	bool	bool	所有输入值都为真时返回真，否则返回假
bool_or(*expression*)	bool	bool	至少一个输入值为真时返回真，否则返回假
count(*)	任意	bigint	number of input rows
count(*expression*)	任意	bigint	*expression* 的值不为 null 的输入行数
every(*expression*)	bool	bool	等价于 bool_and
max(*expression*)	any array, numeric, string, or date/time type	same as argument type	所有输入值中*expression*的最大值
min(*expression*)	any array, numeric, string, or date/time type	same as argument type	所有输入值中*expression*的最小值
stddev(*expression*)	smallint, int, bigint, real, double precision, or numeric	double precision（浮点参数），否则 numeric	输入值的样本标准差
sum(*expression*)	smallint, int, bigint, real, double precision, numeric, interval, or money	smallint或int参数为bigint，bigint参数为numeric，否则与参数数据类型相同	所有输入值中*expression*的总和
variance(*expression*)	smallint, int, bigint, real, double precision, or numeric	double precision（浮点参数），否则 numeric	输入值的样本方差（样本标准差的平方）

需要注意，除了count之外，这些函数在没有选中任何行时都会返回空值。特别地，sum 在没有输入行时返回空值，而不是预期中的零。必要时，可以用coalesce 函数把空值替换成零。

注意

布尔聚合bool_and和bool_or对应于标准 SQL 聚合every和any或some。至于any 和some，标准语法似乎存在歧义：

```
SELECT b1 = ANY((SELECT b2 FROM t2 ...)) FROM t1 ...;
```

此处的ANY既可以被视为引入一个子查询，
也可以在该子查询返回一行布尔值时被视为聚合函数。因此，
不能将标准名称用于这些聚合。

注意

习惯于其他 SQL 数据库管理系统的用户，可能会对count聚合用于整个表时的性能感到失望。如下查询：

```
SELECT count(*) FROM sometable;
```

将被PostgreSQL以顺序扫描整个表的方式执行。

9.16. 子查询表达式

本节描述PostgreSQL中可用的SQL兼容的子查询表达式。
本节介绍的所有表达式形式都返回布尔值（真/假）结果。

9.16.1. EXISTS

EXISTS (*subquery*)

EXISTS的参数是一个任意的SELECT语句，或者说子查询。
系统对子查询进行运算以判断它是否返回行。如果它至少返回一行，那么
EXISTS的结果就为“真”；如果子查询没有返回行，那么EXISTS的结果是“假”。

子查询可以引用外层查询的变量，这些变量在该子查询的任何一次计算中都起常量的作用。

这个子查询通常只是运行到能判断它是否可以返回至少一行为止，而不是等到全部结束。
在这里写任何有副作用的子查询都是不明智的（例如调用序列函数）；
这些副作用是否发生是很难判断的。

因为结果只取决于是否会返回行，而不取决于这些行的内容，
所以这个子查询的输出列表通常是无关紧要的。一个常用的编码习惯是用EXISTS(SELECT 1 WHERE ...)的形式写所有的EXISTS测试。不过这条规则有例外，例如那些使用INTERSECT的子查询。

下面这个简单的示例类似在col2上的一次内连接，但是它为每个 tab1的行最多生成一行输出，即使存在多个匹配tab2的行也如此：

```
SELECT col1
FROM tab1
WHERE EXISTS (SELECT 1 FROM tab2 WHERE col2 = tab1.col2);
```

9.16.2. IN

expression IN (*subquery*)

右手边是一个圆括号括起来的子查询，它必须恰好返回一列。左手边表达式会被求值，
并与子查询结果逐行比较。如果找到任意相等的子查询行，那么IN的结果就是“真”。
如果没有找到相等行，那么结果就是“假”（包括子查询不返回任何行的情况）。

请注意，如果左侧表达式得到空值，或者右侧没有相等的值且至少有一行得到空值，则
IN结构的结果将是空值，而不是假。这符合 SQL 对空值布尔组合的一般规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor IN (*subquery*)

这种形式的IN左手边是一个行构造器，如第 4.2.11 节中所述。右手边是一个圆括号括起来的子查询，它必须返回与左手边行中表达式数量完全相同的列数。左手边表达式会被求值，并与子查询结果逐行比较。如果找到任意相等的子查询行，则IN的结果为“真”。如果没有找到相等行，那么结果为“假”（包括子查询不返回任何行的情况）。

通常，表达式或者子查询行里的空值是按照 SQL 布尔表达式的一般规则进行组合的。如果两个行对应的成员都非空并且相等，那么认为这两行相等；如果任意对应成员为非空且不等，那么这两行不等；否则这样的行比较的结果是未知（空值）。如果所有行的结果要么是不等，要么是空值，并且至少有一个空值，那么IN的结果是空值。

9.16.3. NOT IN

expression NOT IN (*subquery*)

右手边是一个圆括号括起来的子查询，它必须恰好返回一列。左手边表达式会被求值，并与子查询结果逐行比较。如果只找到不相等的子查询行（包括子查询不返回任何行的情况），那么NOT IN的结果是“真”。如果找到任何相等行，则结果为“假”。

请注意，如果左侧表达式得到空值，或者右侧没有相等的值且至少有一行得到空值，则NOT IN结构的结果将是空值，而不是真。这符合 SQL 对空值布尔组合的一般规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

row_constructor NOT IN (*subquery*)

这种形式的NOT IN左侧是一个行构造器，如第 4.2.11 节中所述。右侧是一个用圆括号括起来的子查询，必须返回与左侧行中表达式数量完全相同的列数。对左侧表达式求值后，将其按行与子查询结果的每一行比较。
如果只找到不相等的子查询行（包括子查询不返回任何行的情况），则NOT IN的结果为“真”。如果找到任何相等行，则结果为“假”。

通常，表达式或者子查询行里的空值是按照 SQL 布尔表达式的一般规则进行组合的。如果两个行对应的成员都非空并且相等，那么认为这两行相等；如果任意对应成员为非空且不等，那么这两行不等；否则这样的行比较的结果是未知（空值）。如果所有行的结果要么是不等，要么是空值，并且至少有一个空值，那么NOT IN的结果是空值。

9.16.4. ANY/SOME

expressionoperator ANY (*subquery*)
expression operator SOME (*subquery*)

右侧是一个用圆括号括起来的子查询，必须恰好返回一列。左侧表达式会被求值，并使用给定的*operator*与子查询结果的每一行进行比较。该操作符必须产生布尔结果。
如果得到任何真值结果，则ANY的结果为“真”。
如果没有找到真值结果（包括子查询不返回任何行的情况），则结果为“假”。

SOME是ANY的同义词。IN等价于= ANY。

注意，如果没有任何成功的比较，并且至少有一个右侧行的操作符结果为 null，则ANY结构的结果将是 null 而不是假。这符合 SQL 对空值布尔组合的正常规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

*row_constructor*operator ANY (*subquery*)
row_constructor operator SOME (*subquery*)

这种形式的ANY左侧是一个行构造器，如第 4.2.11 节所述。右侧是一个用括号括起来的子查询，它必须返回与左侧行中表达式数量完全相同的列数。左侧表达式会被求值，并使用给定的*operator*与子查询结果的每一行逐行比较。目前，行式ANY结构中只允许使用=和<>操作符。如果分别找到任何相等或不等的行，则ANY的结果为“真”。
如果没有找到这样的行（包括子查询不返回行的情况），则结果为“假”。

和通常一样，行中的空值按照 SQL 布尔表达式的正常规则进行组合。如果两个行的所有对应成员都非空且相等，则认为两行相等；如果任何对应成员非空且不相等，则两行不相等；
否则该行比较的结果为未知（null）。如果至少有一个行的结果为 null，则ANY的结果不可能是假；它将是真或 null。

9.16.5. ALL

*expression*operator ALL (*subquery*)

右侧是一个用圆括号括起来的子查询，必须恰好返回一列。左侧表达式会被求值，并使用给定的*operator*与子查询结果的每一行进行比较。该操作符必须产生布尔结果。
如果所有行都得到真（包括子查询不返回任何行的情况），则ALL的结果为“真”。
如果得到任何假值结果，则结果为“假”。

NOT IN等价于<> ALL。

注意，如果没有任何失败，但至少有一个右侧行的操作符结果为 null，则ALL结构的结果将是 null 而不是真。这符合 SQL 对空值布尔组合的正常规则。

和EXISTS一样，假定子查询一定会完整运行并不明智。

*row_constructor*operator ALL (*subquery*)

这种形式的ALL左侧是一个行构造器，如第 4.2.11 节所述。右侧是一个用括号括起来的子查询，它必须返回与左侧行中表达式数量完全相同的列数。左侧表达式会被求值，并使用给定的*operator*与子查询结果的每一行逐行比较。目前，行式ALL查询中只允许使用=和<>操作符。
如果所有子查询行分别相等或不相等，则ALL的结果为“真”
（包括子查询不返回任何行的情况）。如果发现任何一行分别不相等或相等，则结果为“假”。

和通常一样，行中的空值按照 SQL 布尔表达式的正常规则进行组合。如果两个行的所有对应成员都非空且相等，则认为两行相等；如果任何对应成员非空且不相等，则两行不相等；
否则该行比较的结果为未知（null）。如果至少有一个行的结果为 null，则ALL的结果不可能是真；它将是假或 null。

9.16.6. 按行比较

*row_constructor*operator (*subquery*)

左侧是一个行构造器，如第 4.2.11 节所述。右侧是一个用圆括号括起来的子查询，必须返回与左侧行中表达式数量完全相同的列数。此外，该子查询不能返回超过一行；如果它返回零行，则结果为空值。对左侧求值后，将所得的行与子查询返回的唯一一行进行比较。目前，
行式比较中只允许使用=和<>操作符。如果两行分别相等或不相等，则结果为“真”。

和通常一样，行中的空值按照 SQL 布尔表达式的正常规则进行组合。如果两个行的所有对应成员都非空且相等，则认为两行相等；如果任何对应成员非空且不相等，则两行不相等；否则该行比较的结果为未知（null）。

9.17. 行和数组比较

本节描述几个特殊的结构，用于在值的组之间进行多重比较。
这些形式语法上和前面一节的子查询形式相关，但是不涉及子查询。涉及数组子表达式的形式是PostgreSQL的扩展；其余形式是SQL兼容的。
本节介绍的所有表达式形式都返回布尔（Boolean）结果（真/假）。

9.17.1. IN

```
expression IN (value [, ...])
```

右侧是一个用圆括号括起来的标量表达式列表。
如果左侧表达式的结果等于任何一个右侧表达式的结果，则结果为“真”。这是

```
expression = value1
OR
expression = value2
OR
...
```

的简写形式。

注意，如果左侧表达式产生 null，或者不存在相等的右侧值且至少有一个右侧表达式产生 null，则`IN`结构的结果将是 null 而不是假。这符合 SQL 对空值布尔组合的正常规则。

9.17.2. NOT IN

```
expression NOT IN (value [, ...])
```

右侧是一个用圆括号括起来的标量表达式列表。
如果左侧表达式的结果与所有右侧表达式的结果都不相等，则结果为“真”。这是

```
expression <> value1
AND
expression <> value2
AND
...
```

的简写形式。

注意，如果左侧表达式产生 null，或者不存在相等的右侧值且至少有一个右侧表达式产生 null，则`NOT IN`结构的结果将是 null，而不是人们可能天真期望的真。这符合 SQL 对空值布尔组合的正常规则。

> **提示**
>
> `x NOT IN y`在任何情况下都等价于`NOT (x IN y)`。不过，对新手来说，使用`NOT IN`时null 值比使用`IN`时更容易造成困扰。最好尽可能用肯定的形式表达你的条件。

9.17.3. ANY/SOME（数组）

```
expressionoperator ANY (array expression)
expression operator SOME (array expression)
```

右侧是一个用圆括号括起来的表达式，它必须产生一个数组值。左侧表达式会被求值，并使用给定的*operator*与数组的每个元素进行比较。该操作符必须产生布尔结果。如果得到任何真值结果，则ANY的结果为“真”。如果没有找到真值结果（包括数组有零个元素的情况），则结果为“假”。

SOME是ANY的同义词。

9.17.4. ALL（数组）

```
expressionoperator ALL (array expression)
```

右侧是一个用圆括号括起来的表达式，它必须产生一个数组值。左侧表达式会被求值，并使用给定的*operator*与数组的每个元素进行比较。该操作符必须产生布尔结果。如果所有比较都得到真（包括数组有零个元素的情况），则ALL的结果为“真”。如果找到任何假值结果，则结果为“假”。

9.17.5. 按行比较

```
row_constructor
operator
row_constructor
```

两侧都是行构造器，如第 4.2.11 节所述。两个行值的字段数必须相同。对两侧分别求值后，按行进行比较。目前，行式比较中只允许使用=和<>操作符。如果两行分别相等或不相等，则结果为“真”。

和通常一样，行中的空值按照 SQL 布尔表达式的正常规则进行组合。如果两个行的所有对应成员都非空且相等，则认为两行相等；如果任何对应成员非空且不相等，则两行不相等；否则该行比较的结果为未知（null）。

```
row_constructor IS DISTINCT FROM row_constructor
```

这个结构与<>行比较相似，但是它对于空值输入不会得到空值。任何空值被认为和任何非空值不相等（有区别），并且任意两个空值被认为相等（无区别）。因此结果将总是为真或为假，永远不会是空值。

```
row_constructor IS NULL
row_constructor IS NOT NULL
```

这些结构测试一个行值是否为 null 或非 null。如果一个行值至少有一个字段不为 null，则认为它是非 null 的。

9.18. 集合返回函数

本节描述可能返回多于一行的函数。目前这一类函数只有序列生成函数，详见表 9.38。

表 9.38. 序列生成函数

Function	Argument Type	Return Type	描述
generate_series(*start*, *stop*)	int或bigint	setof int或setof bigint（与参数类型相同）	以 1 为步长，从 *start* 到 *stop* 生成一系列值
generate_series(*start*, *stop*, *step*)	int或bigint	setof int或setof bigint（与参数类型相同）	生成一系列值，从*start*到*stop*，步长为*step*

当*step*为正数时，如果*start*大于*stop*，则返回零行。反之，当*step*为负数时，如果*start*小于*stop*则返回零行。如果输入为NULL，也返回零行。如果*step*为零，则会报错。
下面是一些示例：

```
SELECT * FROM generate_series(2,4);
 generate_series
-----------------
                2
                3
                4
(3 rows)

SELECT * FROM generate_series(5,1,-2);
 generate_series
-----------------
                5
                3
                1
(3 rows)

SELECT * FROM generate_series(4,3);
 generate_series
-----------------
(0 rows)

-- 此示例使用日期加整数的操作符
SELECT current_date + s.a AS dates FROM generate_series(0,14,7) AS s(
a);
   dates
------------
 2004-02-05
 2004-02-12
 2004-02-19
(3 rows)

SELECT * FROM generate_series('2008-03-01 00:00'::timestamp,
                              '2008-03-04 12:00', '10 hours');
   generate_series
---------------------
 2008-03-01 00:00:00
 2008-03-01 10:00:00
 2008-03-01 20:00:00
 2008-03-02 06:00:00
 2008-03-02 16:00:00
```

```
 2008-03-03 02:00:00
 2008-03-03 12:00:00
 2008-03-03 22:00:00
 2008-03-04 08:00:00
(9 rows)
```

9.19. 系统信息函数

表 9.39列出了多个用于提取会话和系统信息的函数。

表 9.39. 会话信息函数

Name	Return Type	描述
current_database()	name	当前数据库名称
current_schema()	name	当前模式名称
current_schemas(boolean)	name[]	搜索路径中的模式名称，可选地包含隐式模式
current_user	name	当前执行上下文的用户名
inet_client_addr()	inet	远端连接地址
inet_client_port()	int	远端连接端口
inet_server_addr()	inet	本地连接地址
inet_server_port()	int	本地连接端口
session_user	name	会话用户名
pg_postmaster_start_time ()	timestamp with time zone	postmaster 启动时间
user	name	等价于 current_user
version()	text	PostgreSQL版本信息

session_user通常是发起当前数据库连接的用户；但超级用户可以用SET SESSION AUTHORIZATION更改这一设置。current_user是适用于权限检查的用户标识符。通常它等于会话用户，但可以通过SET ROLE更改。在执行带有SECURITY DEFINER属性的函数期间，它也会改变。用 Unix 的话说，会话用户是“真实用户”，而当前用户是“有效用户”。

> **注意**
>
> current_user、session_user和user
> 在SQL中具有特殊的语法地位：调用它们时不得带尾随的圆括号。

current_schema 返回搜索路径中第一个模式的名称（如果搜索路径为空，则返回空值）。在创建表或其他命名对象时，如果未指定目标模式，就会使用该模式。current_schemas(boolean) 返回当前搜索路径中所有模式名称的数组。布尔选项决定是否在返回的搜索路径中包含 pg_catalog 等隐式包含的系统模式。

> **注意**
>
> 可以在运行时修改搜索路径，命令如下：

```
SET search_path TO schema [, schema, ...]
```

inet_client_addr 返回当前客户端的 IP 地址，inet_client_port 返回端口号。inet_server_addr 返回服务器接受当前连接所用的 IP 地址，inet_server_port 返回端口号。如果当前连接通过 Unix 域套接字建立，这些函数都返回 NULL。

pg_postmaster_start_time返回postmaster启动时刻的timestamp with time zone值。

version返回一个描述PostgreSQL服务器版本的字符串。

表 9.40列出了一些函数，它们允许用户通过编程方式查询对象访问权限。关于权限的更多信息，请参见第 5.6 节。

表 9.40. 访问权限查询函数

Name	Return Type	描述
has_table_privilege(*user*, *table*, *privilege*)	boolean	用户是否具有表权限
has_table_privilege(*table*, *privilege*)	boolean	用户是否具有表权限
has_database_privilege(*user*, *database*, *privilege*)	boolean	用户是否具有数据库权限
has_database_privilege(*database*, *privilege*)	boolean	用户是否具有数据库权限
has_function_privilege(*user*, *function*, *privilege*)	boolean	用户是否具有函数权限
has_function_privilege(*function*, *privilege*)	boolean	用户是否具有函数权限
has_language_privilege(*user*, *language*, *privilege*)	boolean	用户是否具有语言权限
has_language_privilege(*language*, *privilege*)	boolean	用户是否具有语言权限
pg_has_role(*user*, *role*, *privilege*)	boolean	用户是否具有角色权限
pg_has_role(*role*, *privilege*)	boolean	用户是否具有角色权限
has_schema_privilege(*user*, *schema*, *privilege*)	boolean	用户是否具有模式权限
has_schema_privilege(*schema*, *privilege*)	boolean	用户是否具有模式权限
has_tablespace_privilege(*user*, *tablespace*, *privilege*)	boolean	用户是否具有表空间权限

Name	Return Type	描述
has_tablespace_privilege(*tablespace*, *privilege*)	boolean	用户是否具有表空间权限

has_table_privilege检查用户是否可以以某种方式访问表。用户可以通过名称或 OID（pg_authid.oid）指定；如果省略该参数，则假定使用current_user。表可以通过名称或 OID 指定。（因此，has_table_privilege实际上有六种变体，
可以通过参数的数量和类型来区分。）通过名称指定时，如有需要，可以用模式限定名称。
所需的访问权限类型由文本字符串指定，其值必须为SELECT、INSERT、UPDATE、DELETE、RULE、REFERENCES或TRIGGER之一。（不过字符串的大小写无关紧要。）一个例子是：

```
SELECT has_table_privilege('myschema.mytable', 'select');
```

has_database_privilege 检查用户是否可以以某种方式访问数据库。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE、TEMPORARY 或 TEMP（等同于TEMPORARY）的某种组合。

has_function_privilege检查用户是否可以以某种方式访问函数。其参数形式类似于has_table_privilege。通过文本字符串而不是 OID 指定函数时，允许的输入与regprocedure数据类型相同（参见第 8.12 节）。所需访问权限类型必须为EXECUTE。例如：

```
SELECT has_function_privilege('joeuser', 'myfunc(int, text)',
 'execute');
```

has_language_privilege 检查用户是否可以以某种方式访问过程语言。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 USAGE。

pg_has_role检查用户是否可以以某种方式访问角色。它的参数可能性与has_table_privilege类似。所需访问权限类型必须为MEMBER或USAGE的某种组合。MEMBER
表示直接或间接地属于该角色（即有权执行SET ROLE），而USAGE表示无需执行SET ROLE就能立即使用该角色的权限。

has_schema_privilege 检查用户是否可以以某种方式访问模式。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE 或 USAGE 的某种组合。

has_tablespace_privilege 检查用户是否可以以某种方式访问表空间。其参数形式与 has_table_privilege 类似。所需访问权限类型必须为 CREATE。

要测试用户是否持有某个权限的转授选项，可以在权限关键字后加上 WITH GRANT OPTION；例如'UPDATE WITH GRANT OPTION'。

表 9.41列出了判断某个特定对象是否可见的函数，其判断依据是当前模式搜索路径。例如，如果表所在的模式位于搜索路径中，并且在搜索路径的更前面没有同名表，就称该表可见。
这等价于说，可以只通过表名引用该表，而不必显式地用模式限定。要列出所有可见表的名称：

```
SELECT relname FROM pg_class WHERE pg_table_is_visible(oid);
```

pg_table_is_visible对表（或视图，或任何其他种类的pg_class项）执行检查。pg_type_is_visible、pg_function_is_visible、pg_operator_is_visible、pg_opclass_is_visible和pg_conversion_is_visible分别对类型（和域）、函数、
操作符、操作符类和转换执行同样的可见性检查。对于函数和操作符，
如果搜索路径中的一个对象在路径更早的位置没有同参数数据类型的同名对象，
则该对象在搜索路径中可见。对于操作符类，要考虑名字和相关的索引访问方法。

表 9.41. 模式可见性查询函数

Name	Return Type	描述
pg_table_is_visible(*table_oid*)	boolean	表是否在搜索路径中可见
pg_type_is_visible(*type_oid*)	boolean	类型（或域）是否在搜索路径中可见
pg_function_is_visible(*function_oid*)	boolean	函数是否在搜索路径中可见
pg_operator_is_visible(*operator_oid*)	boolean	操作符是否在搜索路径中可见
pg_opclass_is_visible(*opclass_oid*)	boolean	操作符类是否在搜索路径中可见
pg_conversion_is_visible(*conversion_oid*)	boolean	转换是否在搜索路径中可见

所有这些函数都需要用对象 OID 标识要检查的对象。如果想按名称测试对象，使用 OID 别名类型会很方便（regclass、regtype、regprocedure或regoperator），例如：

```
SELECT pg_type_is_visible('myschema.widget'::regtype);
```

注意，以这种方式测试一个不带模式限定的名称没有多大意义——如果该名称能够被识别，它必然是可见的。

表 9.42列出从系统目录中提取信息的函数。

表 9.42. 系统目录信息函数

Name	Return Type	描述
format_type(*type_oid*, *typemod*)	text	获取数据类型的 SQL 名称
pg_get_constraintdef(*constraint_oid*)	text	获取约束定义
pg_get_constraintdef(*constraint_oid*, *pretty_bool*)	text	获取约束定义
pg_get_expr(*expr_text*, *relation_oid*)	text	反编译表达式的内部形式，假定其中所有 Var 节点都引用第二个参数指定的关系
pg_get_expr(*expr_text*, *relation_oid*, *pretty_bool*)	text	反编译表达式的内部形式，假定其中所有 Var 节点都引用第二个参数指定的关系
pg_get_indexdef(*index_oid*)	text	获取索引的 CREATE INDEX 命令
pg_get_indexdef(*index_oid*, *column_no*, *pretty_bool*)	text	获取索引的CREATE INDEX命令，或在*column_no*不为零时仅获取一个索引列的定义
pg_get_ruledef(*rule_oid*)	text	获取规则的 CREATE RULE 命令
pg_get_ruledef(*rule_oid*, *pretty_bool*)	text	获取规则的 CREATE RULE 命令

Name	Return Type	描述
pg_get_serial_sequence(*table_name*, *column_name*)	text	获取serial或bigserial列所使用的序列的名称
pg_tablespace_databases(*tablespace_oid*)	setof oid	获取在该表空间中具有对象的数据库 OID 集合
pg_get_triggerdef(*trigger_oid*)	text	获取触发器的 CREATE [CONSTRAINT] TRIGGER 命令
pg_get_userbyid(*roleid*)	name	获取具有给定 ID 的角色名称
pg_get_viewdef(*view_name*)	text	获取视图或物化视图的底层 SELECT 命令（已弃用）
pg_get_viewdef(*view_name*, *pretty_bool*)	text	获取视图或物化视图的底层 SELECT 命令（已弃用）
pg_get_viewdef(*view_oid*)	text	获取视图底层的SELECT命令
pg_get_viewdef(*view_oid*, *pretty_bool*)	text	获取视图底层的SELECT命令

format_type返回由类型 OID 和可能的类型修饰符标识的数据类型的 SQL 名称。如果没有具体的修饰符，则为类型修饰符传递 NULL。

pg_get_constraintdef、pg_get_indexdef、pg_get_ruledef和pg_get_triggerdef分别重建约束、索引、规则或触发器的创建命令。（注意，这是通过反编译重建的，并非命令的原始文本。）pg_get_expr反编译单个表达式的内部形式，例如列的默认值。这在检查系统目录内容时可能很有用。pg_get_viewdef重建定义视图的SELECT查询。这些函数中的大多数有两种变体，其中一种可以选择对结果进行“美化打印”。美化打印的格式更易读，但默认格式更有可能被未来版本的PostgreSQL以相同方式解释；做转储时应避免使用美化打印。向美化打印参数传入false，所得结果与没有该参数的变体完全相同。

pg_get_serial_sequence获取与serial或bigserial列关联的序列的名称。该名称的格式适合传递给序列函数（见第 9.12 节）。如果该列没有关联的序列，则返回 NULL。

pg_get_userbyid根据给定的 OID 获取角色的名称。

pg_tablespace_databases用于检查表空间。它返回在该表空间中存有对象的数据库的 OID 集合。如果这个函数返回了任何行，则该表空间非空，不能被删除。要显示占用该表空间的具体对象，你需要连接到pg_tablespace_databases标识出的各个数据库，并查询它们的pg_class目录。

表 9.43中的函数用于提取此前通过COMMENT命令存储的注释。如果找不到与指定参数对应的注释，则返回空值。

表 9.43. 注释信息函数

Name	Return Type	描述
obj_description(*object_oid*, *catalog_name*)	text	获取数据库对象的注释
obj_description(*object_oid*)	text	获取数据库对象的注释（已废弃）
col_description(*table_oid*, *column_number*)	text	获取表列的注释

col_description返回表列的注释，列由所属表的 OID 和列号指定。由于列本身没有各自的OID，obj_description不能用于表列。

obj_description 的双参数形式返回数据库对象的注释，对象由其 OID 和所在系统目录的名称指定。例如，obj_description(123456,'pg_class') 会获取 OID 为 123456 的表的注释。obj_description 的单参数形式只需要对象 OID。该形式已被弃用，因为无法保证 OID 在不同系统目录之间唯一，因而可能返回错误的注释。

9.20. 系统管理函数

表 9.44展示了那些可以用于查询以及修改运行时配置参数的函数。

表 9.44. 配置设定函数

Name	Return Type	描述
current_setting(*setting_name*)	text	获取设置的当前值
set_config(*setting_name*, *new_value*, *is_local*)	text	设置参数并返回新值

函数current_setting返回设置*setting_name*的当前值。它对应于SQL命令SHOW。例如：

```
SELECT current_setting('datestyle');

 current_setting
-----------------
 ISO, MDY
(1 row)
```

set_config将参数*setting_name*设置为*new_value*。如果*is_local*为true，新值仅适用于当前事务。如果要让新值适用于当前会话，请改用false。该函数对应于 SQL 命令SET。例如：

```
SELECT set_config('log_statement_stats', 'off', false);

 set_config
------------
 off
(1 row)
```

表 9.45中展示的函数向其他服务器进程发送控制信号。这些函数仅限超级用户使用。

表 9.45. 服务器信号函数

Name	Return Type	描述
pg_cancel_backend(*pid* int)	boolean	取消后端的当前查询
pg_reload_conf()	boolean	使服务器进程重新加载配置文件
pg_rotate_logfile()	boolean	轮转服务器日志文件

这些函数在成功时都返回 true，否则返回 false。

pg_cancel_backend向由进程 ID 标识的后端进程发送一个查询取消（SIGINT）信号。活动后端的进程 ID 可以从pg_stat_activity视图的procpid列中找到，或者通过用ps列出服务器上的postgres进程得到。

pg_reload_conf 向postmaster发送 SIGHUP 信号，使所有服务器进程重新加载配置文件。

pg_rotate_logfile 向日志文件管理器发送信号，使其立即切换到新的输出文件。只有使用redirect_stderr进行日志记录时才有效，否则不存在日志文件管理器子进程。

表 9.46中列出的函数有助于进行在线备份。这些函数仅限超级用户使用。

表 9.46. 备份控制函数

Name	Return Type	描述
pg_start_backup(*label* text)	text	准备执行在线备份
pg_stop_backup()	text	结束在线备份

pg_start_backup接受单个参数，它是一个任意的用户自定义备份标签。（通常是备份转储文件将要保存的名称。）
该函数把一个备份标签文件写入数据库集簇的数据目录，然后以文本形式返回备份的起始 WAL 偏移。（用户不必关注这个结果值，但提供它只是以防它有用。）

pg_stop_backup 删除 pg_start_backup 创建的标签文件，并在 WAL 归档区域创建一个备份历史文件。历史文件包括传给 pg_start_backup 的标签、备份的起止 WAL 偏移以及备份的起止时间。返回值是备份的结束 WAL 偏移（同样可以忽略）。

有关正确使用这些函数的详细信息，参见第 23.3 节。

表 9.47中的函数计算数据库对象的磁盘空间用量。

表 9.47. 数据库对象尺寸函数

Name	Return Type	描述
pg_column_size(any)	int	存储某个值所使用的字节数（可能已压缩）
pg_tablespace_size(oid)	bigint	指定 OID 的表空间占用的磁盘空间
pg_tablespace_size(name)	bigint	指定名称的表空间占用的磁盘空间
pg_database_size(oid)	bigint	指定 OID 的数据库占用的磁盘空间
pg_database_size(name)	bigint	指定名称的数据库占用的磁盘空间
pg_relation_size(oid)	bigint	指定 OID 的表或索引占用的磁盘空间
pg_relation_size(text)	bigint	指定名称的表或索引占用的磁盘空间。表名可以用模式名限定
pg_total_relation_size(oid)	bigint	指定 OID 的表占用的磁盘空间总量，包括索引和 TOAST 数据
pg_total_relation_size(text)	bigint	指定名称的表占用的磁盘空间总量，包括索引和 TOAST 数据。表名可以用模式名限定
pg_size_pretty(bigint)	text	将以 64 位整数表示的字节大小转换为带有大小单位的人类可读格式

pg_column_size 显示存储任意单个数据值所用的空间。

pg_total_relation_size接受表或 TOAST 表的 OID 或名称，返回数据及所有关联索引和 TOAST 表的字节大小。

pg_tablespace_size和pg_database_size接受表空间或数据库的 OID 或名称，并返回其中占用的磁盘总空间。

pg_relation_size接受表、索引或 TOAST 表的 OID 或名称，并以字节为单位返回大小。

pg_size_pretty 可以将其他函数的结果格式化为便于阅读的形式，适当地使用 kB、MB、GB 或 TB 单位。

表 9.48中展示的函数提供对数据库服务器所在机器上文件的本地访问。只能访问数据库集簇目录和log_directory中的文件。集簇目录中的文件使用相对路径，日志文件使用与log_directory配置设置相匹配的路径。这些函数仅限超级用户使用。

表 9.48. 通用文件访问函数

Name	Return Type	描述
pg_ls_dir(*dirname* text)	setof text	列出目录内容。
pg_read_file(*filename* text, *offset* bigint, *length* bigint)	text	返回文本文件的内容。
pg_stat_file(*filename* text)	record	返回文件信息。

pg_ls_dir 返回指定目录中的所有名称，但不包括特殊项“.”和“..”。

pg_read_file从给定的*offset*开始返回文本文件的一部分，最多返回*length*字节（如果先到达文件末尾，则返回更少的字节）。如果*offset*为负，则它相对于文件末尾。

pg_stat_file返回一条记录，包含文件大小、最后访问时间戳、最后修改时间戳、最后文件状态更改时间戳（仅限 Unix 平台）、文件创建时间戳（仅限 Windows），以及一个boolean值，表示它是否是目录。典型用法包括：

```
SELECT * FROM pg_stat_file('filename');
SELECT (pg_stat_file('filename')).modification;
```

第 10 章 类型转换

SQL语句可能有意或无意地要求在同一表达式中混用不同的数据类型。
PostgreSQL提供了丰富的机制来求值这类混合类型表达式。

在很多情况下，用户并不需要理解类型转换机制的细节。不过，
PostgreSQL执行的隐式类型转换会影响查询结果。必要时，可以使用
显式类型转换来控制这些结果。

本章介绍PostgreSQL的类型转换机制和约定。
有关特定数据类型以及允许使用的函数和操作符的更多信息，请参阅第 8 章和第 9 章中的相关章节。

10.1. 概述

SQL是一种强类型语言。也就是说，
每个数据项都关联着一个决定其行为和允许用法的数据类型。
PostgreSQL拥有可扩展的类型系统，比其他SQL实现更通用也更灵活。因此，
PostgreSQL中的大多数类型转换行为都由通用规则控制，而不是依赖临时性的启发式规则。
这使得即使表达式中包含用户定义类型，也能够使用混合类型表达式。

PostgreSQL扫描器/解析器将词法元素分成五个基本类别：整数、非整数数字、字符串、
标识符和关键字。大多数非数值类型的常量会先被归类为字符串。
SQL语言定义允许给字符串指定类型名，而
PostgreSQL可以利用这一机制让解析器从一开始就沿着正确的路径处理。例如，查询：

```
SELECT text 'Origin' AS "label", point '(0,0)' AS "value";

 label  | value
--------+-------
 Origin | (0,0)
(1 row)
```

包含两个字面量，它们的类型分别是`text`和`point`。如果没有为某个字符串字面量指定类型，
那么它最初会被赋予占位类型`unknown`，并在后文所述的后续阶段解析。

在PostgreSQL解析器中，有四种基本的SQL结构需要采用不同的类型转换规则：

函数调用

PostgreSQL类型系统的很大一部分建立在一套丰富的函数之上。
函数可以有一个或多个参数。由于PostgreSQL支持函数重载，
单凭函数名无法唯一确定要调用哪个函数；
解析器必须根据所提供参数的数据类型选择正确的函数。

操作符

PostgreSQL允许表达式使用前缀和后缀一元（单参数）操作符，以及二元（双参数）
操作符。与函数一样，操作符也可以重载，因此同样存在选择正确操作符的问题。

值存储

SQL的`INSERT`和`UPDATE`语句会把表达式的结果放入表中。
语句中的表达式必须与目标列的类型匹配，并且在必要时转换为目标列类型。

UNION、CASE和相关结构

由于经UNION合并的SELECT语句的所有查询结果都必须出现在同一组列中，因此每个SELECT子句的结果类型必须彼此匹配，并转换为统一的一组类型。类似地，CASE结构中的结果表达式必须转换为某种公共类型，这样整个CASE表达式才有确定的输出类型。对于ARRAY构造以及GREATEST和LEAST函数，情况也是如此。

系统目录保存了哪些数据类型之间的转换（称为类型转换）是有效的，以及如何执行这些转换的信息。用户可以使用 CREATE CAST 命令添加额外的类型转换。（这通常与定义新数据类型一并完成。内置类型之间的类型转换集合经过精心设计，最好不要改动。）

解析器中提供了一条额外的启发式规则，以便对 SQL 标准类型的正确行为做出更好的猜测。系统定义了若干基本的类型分类：boolean、numeric、string、bitstring、datetime、timespan、geometric、network以及用户定义类型。除用户定义分类外，每个分类都有一个或多个首选类型，在存在歧义时会优先选择它们。在用户定义分类中，每个类型都是自己的首选类型。因此，当多个可能的内建类型并存时，有歧义的表达式（即存在多个候选解析方案的表达式）往往仍能被解析；但当存在多个用户定义类型可供选择时，则会引发错误。

所有类型转换规则都以以下几个原则为基础：

- 隐式转换绝不应产生令人意外或无法预测的结果。
- 解析器对其没有先验知识的用户定义类型，应当位于类型层次结构中“更高”的位置。在混合类型表达式中，原生类型总是应当被转换为用户定义类型（当然，仅在确需转换时）。
- 用户定义类型之间没有关联。目前，PostgreSQL除了针对内建类型的硬编码启发式规则，以及基于可用函数和类型转换的隐式关系之外，没有关于类型之间关系的信息可用。
- 如果查询不需要隐式类型转换，解析器和执行器就不应承担额外开销。也就是说，如果查询本身书写正确且类型已经匹配，那么执行时不应让解析器额外耗时，也不应在查询中引入不必要的隐式类型转换调用。

  此外，如果某个查询通常需要先对参数做隐式类型转换才能调用某个函数，而此后用户又定义了一个参数类型正确的新函数，那么解析器应改为使用这个新函数，而不再通过隐式转换去调用旧函数。

10.2. 操作符

操作符调用中所使用的具体操作符由下面的过程确定。注意，该过程会间接受到相关操作符优先级的影响。详见第 4.1.6 节。

操作符类型解析

1. 从系统目录pg_operator中选取要考虑的操作符。如果使用的是未带模式限定的操作符名（通常如此），则考虑当前搜索路径中可见且名称和参数个数匹配的操作符（见第 5.7.3 节）。如果给出了限定名，则只考虑指定模式中的操作符。

   - (可选的) 如果搜索路径中找到多个参数类型完全相同的操作符，只考虑路径中最早出现的那一个。参数类型不同的操作符则不受搜索路径位置影响，处于同等地位。

2. 检查是否存在一个恰好接受输入参数类型的操作符。如果存在（在被考虑的操作符集合中至多只有一个精确匹配），就使用它。

- (可选的) 如果二元操作符调用的一个参数是unknown类型，则在本次检查中假定它与另一个参数类型相同。其他涉及unknown的情况在这一步永远不会找到匹配。

3. 寻找最佳匹配。

   a. 丢弃那些输入类型不匹配，且也无法通过隐式类型转换实现匹配的候选操作符。为此，假定unknown字面量可以转换为任何类型。如果只剩一个候选操作符，就使用它；否则继续下一步。

   b. 遍历所有候选操作符，保留那些在输入类型上精确匹配数最多的候选项。（在这一步中，域被视为与其基础类型相同。）如果没有任何候选项存在精确匹配，则保留全部候选项。如果只剩一个候选操作符，就使用它；否则继续下一步。

   c. 遍历所有候选操作符，在需要类型转换的位置上，保留接受首选类型（属于输入数据类型的类型分类）的位置数最多的候选项。如果没有候选项接受首选类型，则保留全部候选项。如果只剩一个候选操作符，就使用它；否则继续下一步。

   d. 如果有任何输入参数是unknown，就检查剩余候选操作符在这些参数位置上接受哪些类型分类。在每个位置上，如果有任何候选项接受string分类，就选择该分类。（这种偏向字符串的做法是合适的，因为unknown类型的字面量看起来像字符串。）否则，如果所有剩余候选项都接受同一种类型分类，就选择该分类；否则失败，因为在缺少更多线索的情况下无法推断出正确选择。然后丢弃不接受所选类型分类的候选项。此外，如果有任何候选项接受该分类中的首选类型，就丢弃在该参数上接受非首选类型的候选项。

   e. 如果只剩一个候选操作符，就使用它。如果没有候选操作符或者不止一个候选操作符剩余，则失败。

下面是一些示例。

例 10.1. 指数操作符类型解析

目录中只定义了一个指数操作符，它接受类型为 double precision 的参数。扫描器对这个查询表达式的两个参数都赋予初始类型 integer：

```
SELECT 2 ^ 3 AS "exp";

 exp
-----
   8
(1 row)
```

因此，解析器会对两个操作数都做类型转换，该查询等价于：

```
SELECT CAST(2 AS double precision) ^ CAST(3 AS double precision) AS
 "exp";
```

例 10.2. 字符串连接操作符类型解析

处理字符串类型以及复杂的扩展类型时，常常会使用类似字符串的语法。未指定类型的字符串会与可能的操作符候选项进行匹配。

一个参数未指定类型的例子：

```
SELECT text 'abc' || 'def' AS "text and unknown";

 text and unknown
------------------
 abcdef
(1 row)
```

在这种情况下，解析器会查看是否存在一个两边参数都接受text的操作符。既然存在，它就会假定第二个参数应解释为text类型。

下面是对未指定类型进行连接的例子：

```
SELECT 'abc' || 'def' AS "unspecified";

 unspecified
-------------
 abcdef
(1 row)
```

这里对使用哪种类型没有初始提示，因为查询中没有指定任何类型。因此，解析器会查找所有候选操作符，并发现既有接受字符串分类输入的候选项，也有接受位串分类输入的候选项。由于在可用时优先选择字符串分类，所以会选中该分类，再使用字符串的首选类型text作为解析这些unknown类型字面量的具体类型。

例 10.3. 绝对值与取反操作符类型解析

PostgreSQL操作符目录中为前缀操作符@提供了多个条目，它们分别实现各种数值数据类型的绝对值操作。其中一个条目对应float8，它是数值分类中的首选类型。因此，当遇到一个unknown输入时，PostgreSQL会使用该条目：

```
SELECT @ '-4.5' AS "abs";
 abs
-----
 4.5
(1 row)
```

这里，系统在应用所选操作符之前，已经把unknown类型字面量隐式解析为float8。我们可以验证使用的确实是float8，而不是其他类型：

```
SELECT @ '-4.5e500' AS "abs";

ERROR:  "-4.5e500" is out of range for type double precision
```

另一方面，前缀操作符~（按位取反）只对整数数据类型定义，并没有为float8定义。因此，如果对~试一个类似的例子，会得到：

```
SELECT ~ '20' AS "negation";

ERROR:  operator is not unique: ~ "unknown"
HINT:  Could not choose a best candidate operator. You might need to
 add
explicit type casts.
```

这是因为系统无法决定几个可能的~操作符中应优先选择哪一个。
我们可以通过显式类型转换来帮助它：

```
SELECT ~ CAST('20' AS int8) AS "negation";

 negation
----------
      -21
(1 row)
```

10.3. 函数

函数调用所引用的具体函数按照下述过程确定。

函数类型解析

1. 从系统目录pg_proc中选取要考虑的函数。如果使用的是未带模式限定的函数名，
   则考虑当前搜索路径中可见且名称和参数个数匹配的函数（见第 5.7.3 节）。如果给出了限定名，则只考虑指定模式中的函数。

   - (可选的) 如果搜索路径中找到多个参数类型完全相同的函数，只考虑路径中最早出现的那一个。参数类型不同的函数则不受搜索路径位置影响，处于同等地位。

2. 检查是否存在一个恰好接受输入参数类型的函数。
   如果存在（在被考虑的函数集合中至多只有一个精确匹配），就使用它。（涉及unknown的情况在这一步永远找不到匹配。）

3. 如果没有找到精确匹配，就检查该函数调用是否看起来像一个平凡的类型转换请求。
   这种情况发生在：函数调用只有一个参数，且函数名与某个数据类型的（内部）名称相同。
   此外，函数参数必须是unknown类型字面量，
   或者是一种与该命名数据类型二进制兼容的类型。满足这些条件时，
   函数参数会被直接转换为该命名数据类型，而不会发生实际的函数调用。

4. 寻找最佳匹配。

   a. 丢弃那些输入类型不匹配，且也无法通过隐式类型转换实现匹配的候选函数。为此，
      假定unknown字面量可以转换为任何类型。如果只剩一个候选函数，就使用它；
      否则继续下一步。

   b. 遍历所有候选函数，保留那些在输入类型上精确匹配数最多的候选项。（在这一步中，
      域被视为与其基础类型相同。）如果没有任何候选项存在精确匹配，
      则保留全部候选项。如果只剩一个候选函数，就使用它；否则继续下一步。

   c. 遍历所有候选函数，在需要类型转换的位置上，
      保留接受首选类型（属于输入数据类型的类型分类）的位置数最多的候选项。
      如果没有候选项接受首选类型，则保留全部候选项。如果只剩一个候选函数，
      就使用它；否则继续下一步。

   d. 如果有任何输入参数是unknown，
      就检查剩余候选函数在这些参数位置上接受哪些类型分类。在每个位置上，
      如果有任何候选项接受string分类，就选择该分类。
      （这种偏向字符串的做法是合适的，因为unknown类型的字面量看起来像字符串。）
      否则，如果所有剩余候选项都接受同一种类型分类，就选择该分类；否则失败，
      因为在缺少更多线索的情况下无法推断出正确选择。

然后丢弃不接受所选类型分类的候选项。此外，
如果有任何候选项接受该分类中的首选类型，
就丢弃在该参数上接受非首选类型的候选项。

e. 如果只剩一个候选函数，就使用它。如果没有候选函数或者不止一个候选函数剩余，则失败。

注意，操作符类型解析和函数类型解析的“最佳匹配”规则是完全相同的。下面是一些示例。

例 10.4. round 函数参数类型解析

只有一个接受两个参数的round函数；它的第一个参数类型是numeric，第二个参数类型是integer。因此下面的查询会自动把类型为integer的第一个参数转换为numeric：

```
SELECT round(4, 4);

 round
--------
 4.0000
(1 row)
```

该查询实际上会被解析器改写成：

```
SELECT round(CAST (4 AS numeric), 4);
```

由于带小数点的数字常量最初会被赋予numeric类型，下面的查询不需要类型转换，
因此可能稍微更高效一些：

```
SELECT round(4.0, 4);
```

例 10.5. substr 函数类型解析

存在多个substr函数，其中一个接受text和integer。如果以未指定类型的字符串常量调用，系统会选择接受首选类型分类string参数的候选函数（也就是text类型的那个）。

```
SELECT substr('1234', 3);

 substr
--------
     34
(1 row)
```

如果字符串被声明为varchar类型，例如它来自表中的某一列，那么解析器会尝试把它转换为text：

```
SELECT substr(varchar '1234', 3);

 substr
--------
     34
(1 row)
```

解析器会把它改写成如下形式：

```
SELECT substr(CAST (varchar '1234' AS text), 3);
```

> **注意**
>
> 解析器从pg_cast目录得知text与varchar二进制兼容，
> 这意味着把其中一种类型传给接受另一种类型的函数时，无需做任何物理转换。因此，
> 这种情况下实际上不会插入类型转换调用。

如果用integer类型参数调用该函数，解析器会尝试把它转换为text：

```
SELECT substr(1234, 3);

 substr
--------
     34
(1 row)
```

它实际执行的是

```
SELECT substr(CAST (1234 AS text), 3);
```

这种自动转换之所以能够成功，是因为存在从integer到text的隐式可调用类型转换。

10.4. 值存储

要插入表中的值会按照以下步骤转换为目标列的数据类型。

值存储的类型转换

1. 检查是否与目标类型精确匹配。

2. 否则，尝试把表达式转换为目标类型。如果这两种类型之间存在已登记的类型转换，
   这就会成功。如果表达式是unknown类型字面量，
   则会把字符串字面量的内容送入目标类型的输入转换例程。

3. 检查目标类型是否有尺寸调整类型转换。尺寸调整类型转换是该类型到其自身的类型转换。
   如果在pg_cast目录中找到了这种转换，就在把表达式存入目标列之前先应用它。
   此类类型转换的实现函数总是额外接受一个integer参数，
   用来接收目标列的声明长度（实际上就是它的atttypmod值；不同数据类型对atttypmod
   的解释各不相同）。类型转换函数负责应用所有与长度相关的语义，例如长度检查或截断。

例 10.6. character 的存储类型转换

对于声明为character(20)的目标列，下面的语句表明存储值会被调整到正确的长度：

```
CREATE TABLE vv (v character(20));
INSERT INTO vv SELECT 'abc' || 'def';
SELECT v, length(v) FROM vv;

          v           | octet_length
----------------------+--------------
 abcdef               |           20
(1 row)
```

这里实际发生的是，两个unknown类型字面量默认都被解析为text，从而使||操作符被解析为text的连接操作。然后，操作符得到的text结果被转换为bpchar（“用空格填充的字符”，blank-padded char，即character数据类型的内部名称），以匹配目标列类型。（由于text和bpchar是二进制兼容的，这一步不会插入任何真实的函数调用。）最后，系统从系统目录中找到尺寸调整函数bpchar(bpchar, integer)，并把它应用到操作符结果以及存储列长度上。这个特定于类型的函数会执行所需的长度检查并补齐填充空格。

10.5. UNION、CASE及相关结构

SQL 的UNION结构必须让可能不同的类型彼此匹配，以形成单一结果集。该解析算法会分别应用到联合查询的每个输出列。INTERSECT和EXCEPT结构以与UNION相同的方式解析不同类型。CASE、ARRAY以及GREATEST和LEAST构造使用完全相同的算法来匹配其组成表达式并选择结果数据类型。

UNION、CASE及相关结构的类型解析

1. 如果所有输入都是unknown类型，则解析为text类型（字符串分类的首选类型）。否则，在选择结果类型时忽略unknown输入。
2. 如果非unknown输入并非全都属于同一类型分类，则失败。
3. 选择第一个属于该分类首选类型、或者允许所有非unknown输入都隐式转换为它的非unknown输入类型。
4. 把所有输入都转换为所选类型。

下面是一些示例。

例 10.7. 未充分指定类型的 UNION 中的类型解析

```
SELECT text 'a' AS "text" UNION SELECT 'b';

 text
------
 a
 b
(2 rows)
```

这里，unknown类型字面量'b'会被解析为text类型。

例 10.8. 简单 UNION 中的类型解析

```
SELECT 1.2 AS "numeric" UNION SELECT 1;

 numeric
---------
       1
     1.2
(2 rows)
```

字面量1.2的类型是numeric，而integer值1可以隐式转换为numeric，因此使用该类型。

例 10.9. 次序对调的 UNION 中的类型解析

```
SELECT 1 AS "real" UNION SELECT CAST('2.2' AS REAL);
```

```
 real
------
    1
  2.2
(2 rows)
```

这里，由于real类型不能隐式转换为integer，而integer可以隐式转换为real，因此UNION结果类型被解析为real。

第 11 章 索引

索引是增强数据库性能的一种常见方式。
索引允许数据库服务器比没有索引时更快地找到并检索特定行。
但索引也会给整个数据库系统带来额外开销，因此应当合理使用。

11.1. 简介

假设我们有一个类似如下的表：

```
CREATE TABLE test1 (
    id integer,
    content varchar
);
```

并且应用会发出很多如下形式的查询：

```
SELECT content FROM test1 WHERE id = constant;
```

如果没有预先准备，系统就必须逐行扫描整个`test1`表，以找出所有匹配项。如果`test1`中有很多行，而这种查询只会返回很少几行（甚至可能一行也没有，或者只有一行），这显然是低效的办法。但是如果系统被要求维护一个基于`id`列的索引，它就可以用更高效的方法定位匹配行。例如，它可能只需要在一棵搜索树中向下走几层。

大多数非虚构类书籍也采用了类似的方法：读者经常查找的术语和概念会被收集到书末按字母顺序排列的索引中。感兴趣的读者可以相对快速地浏览索引并翻到相应页面，而不必通读整本书才能找到自己感兴趣的内容。正如作者需要预判读者可能查找哪些条目一样，数据库程序员也需要预见哪些索引会有用。

如前所述，下列命令可用于在`id`列上创建索引：

```
CREATE INDEX test1_id_index ON test1 (id);
```

`test1_id_index`这个名字可以自由选择，但最好选一个以后仍能让你记住该索引用途的名字。

要删除索引，使用`DROP INDEX`命令。索引可以在任何时候添加到表上，也可以在任何时候从表上删除。

一旦创建了索引，就无需再进行额外干预：系统会在表被修改时更新索引，并且会在它认为这样做比顺序扫描表更高效时，在查询中使用该索引。不过，你可能仍需要定期运行`ANALYZE`命令来更新统计信息，以便查询规划器作出更有根据的决策。关于如何判断某个索引是否被使用，以及规划器何时和为何可能选择不使用索引，可参见第 13 章。

索引还可以让带有搜索条件的`UPDATE`和`DELETE`命令受益。索引也可以用于连接搜索。因此，定义在连接条件组成列上的索引也能显著加速带连接的查询。

索引创建之后，系统还必须使其与表保持同步。这会给数据操作增加开销。因此，那些在查询中很少使用或从不使用的索引应当被移除。

11.2. 索引类型

PostgreSQL提供了多种索引类型：B-树、R-树、Hash 和 GiST。每种索引类型都采用不同的算法，分别最适合不同类型的查询。默认情况下，`CREATE INDEX`命令创建的是 B-树索引，它适用于最常见的场景。

B-树能够处理可以按某种顺序排序的数据上的等值查询和范围查询。特别是，
只要已索引列参与了下列任一操作符的比较，PostgreSQL查询规划器就会考虑使用 B-树索引：

```
<
<=
=
>=
>
```

与这些操作符组合等价的构造，例如BETWEEN和IN，也可以通过 B-树索引搜索来实现。（但要注意，IS NULL并不等价于=，因此不能使用索引。）

如果模式是常量，并且锚定在字符串起始位置，优化器也可以对涉及模式匹配操作符LIKE和~的查询使用 B-树索引，例如col LIKE 'foo%'或col ~ '^foo'，但不能用于col LIKE '%bar'。不过，如果你的数据库没有使用 C 区域设置，就需要以一个特殊操作符类来创建该索引，才能支持模式匹配查询的索引化，详见下文第 11.8 节。B-树索引也可以用于ILIKE和~*，但前提是模式以非字母字符开头，也就是不会受到大小写转换影响的字符。

R-树索引适合对二维空间数据的查询。要创建一个 R-树索引，可使用如下形式的命令：

```
CREATE INDEX name ON table USING rtree (column);
```

只要已索引列参与的是使用下列任一操作符的比较，PostgreSQL查询规划器就会考虑使用 R-树索引：

```
<<
&<
&>
>>
<<|
&<|
|&>
|>>
~
@
~=
&&
```

（这些操作符的含义见第 9.10 节。）

Hash 索引只能处理简单的等值比较。只要已索引列参与的是使用=操作符的比较，查询规划器就会考虑使用 Hash 索引。使用以下命令创建 Hash 索引：

```
CREATE INDEX name ON table USING hash (column);
```

GiST 索引并不是某一种单独的索引，而是一种基础设施，可在其中实现许多不同的索引策略。因此，GiST 索引可使用哪些具体操作符，取决于所采用的索引策略（即操作符类）。PostgreSQL标准发布版中包含了与 R-树操作符类等价的 GiST 操作符类，还有许多其他 GiST 操作符类可在contrib集合或独立项目中获得。更多信息见第 49 章。

注意

测试表明，PostgreSQL的哈希索引性能并不比 B-树索引好，而且哈希索引的索引大小和构建时间都要差得多。此外，哈希索引操作目前不会写入 WAL 日志，因此如果发生过未写

> 入的更改，数据库崩溃后可能需要用REINDEX重建哈希索引。由于这些原因，
> 目前不建议使用哈希索引。
>
> 类似地，与 GiST 索引的等价操作相比，R-树索引似乎也没有任何性能优势。与哈希索引一样，它们不会写入 WAL 日志，数据库崩溃后可能需要重建。
>
> 虽然哈希索引的问题最终可能会被修复，但 R-树索引类型很可能在未来的发行版中被淘汰。鼓励用户将使用 R-树索引的应用迁移到 GiST 索引。

11.3. 多列索引

一个索引可以定义在表的多个列上。例如，如果你有一个如下形式的表：

```
CREATE TABLE test2 (
  major int,
  minor int,
  name varchar
);
```

（假设你把/dev目录保存在数据库里.....）并且经常发出如下查询：

```
SELECT name FROM test2 WHERE major = constant AND minor = constant;
```

那么将major和minor两列一起建一个索引可能是合适的，例如：

```
CREATE INDEX test2_mm_idx ON test2 (major, minor);
```

当前，只有 B-树和 GiST 索引类型支持多列索引。最多可以指定 32 列。（这个限制可以在构建 PostgreSQL 时修改；参见文件 pg_config_manual.h。）

多列 B-树索引可以用于涉及索引任意列子集的查询条件，但当对前导（最左）列存在约束时，索引效率最高。精确的规则是：对前导列的等值约束，
再加上第一个没有等值约束的列上的任何不等约束，会被用来限制索引扫描的范围。
这些列右侧列上的约束会在索引中进行检查，因此可以减少访问表本体的次数，
但不会缩小必须扫描的索引范围。例如，给定一个 (a, b, c) 上的索引和查询条件 WHERE a = 5 AND b >= 42 AND c < 77，索引必须从第一个 a = 5 且 b = 42 的条目扫描到最后一个 a = 5 的条目。c >= 77 的索引项会被排除，但仍然必须扫描这些项。原则上，这个索引也可以用于只对 b 和/或 c 有约束、而对 a 没有约束的查询，但必须扫描整个索引，因此在大多数情况下，规划器会倾向于顺序扫描表，而不是使用该索引。

多列 GiST 索引只有在对其前导列存在查询条件时才能使用。附加列上的条件会限制索引返回的项，但决定索引需要扫描多少内容的，最重要的仍是第一列上的条件。
如果第一列只有很少几个非重复值，即使其他列有很多非重复值，GiST 索引也会相对低效。

当然，每一列都必须配合适合该索引类型的操作符使用；涉及其他操作符的子句不会被考虑。

多列索引应谨慎使用。在大多数情况下，单列索引已经足够，而且更省空间、也更省时间。
除非表的使用方式极其程式化，否则超过三列的索引通常不会有帮助。
关于不同索引配置优缺点的讨论，还可参见第 11.4 节。

11.4. 组合多个索引

单次索引扫描只能利用那些在索引列上使用了该索引操作符类中操作符、并且彼此以AND连接的查询子句。例如，给定一个(a, b)上的索引，查询条件WHERE a = 5 AND b = 6可以使用该索引，而WHERE a = 5 OR b = 6这样的查询就不能直接使用该索引。

从 8.1 版开始，PostgreSQL能够组合多个索引（包括同一个索引的多次使用），以处理单次索引扫描无法实现的情况。系统可以在多个索引扫描之间形成AND和OR条件。例如，WHERE x = 42 OR x = 47 OR x = 53 OR x = 99这样的查询，可以拆分成对x上的一个索引进行四次独立扫描，每次扫描使用一个查询子句。然后将这些扫描的结果做 OR 运算以得到最终结果。另一个例子是，如果我们在x和y上分别有索引，那么WHERE x = 5 AND y = 6这样的查询，其一种可能实现方式就是分别用两个索引处理相应条件，然后将索引结果做 AND 运算，以确定结果行。

为了组合多个索引，系统会扫描每个所需索引，并在内存中准备一个位图，
给出满足该索引条件的表行位置。然后按查询需要对这些位图进行 AND 和 OR 运算。最后，再访问并返回实际的表行。表行是按物理顺序访问的，因为位图就是这样布局的；
这意味着原始索引中的任何排序都会丢失，所以如果查询带有ORDER BY子句，就需要单独的排序步骤。也正因为如此，再加上每多一次索引扫描都会增加额外时间，
规划器有时会选择使用简单的单一索引扫描，即使还有其他本来也可以利用的索引。

除了最简单的应用之外，通常会有多种可能有用的索引组合，
数据库开发者必须权衡决定提供哪些索引。有时多列索引最好，
但有时创建独立索引并依赖索引组合功能会更合适。例如，
如果你的工作负载包含一组查询：有时只涉及列x，有时只涉及列y，有时同时涉及两列，
那么你可以选择分别在x和y上创建两个独立索引，并依赖索引组合来处理同时使用两列的查询。
你也可以创建一个(x, y)上的多列索引。对于同时涉及两列的查询，这个索引通常会比索引组合更高效，但正如第 11.3 节中所讨论的，它对于只涉及y的查询几乎没有用处，因此不应只创建这一个索引。将这个多列索引与一个单独的y索引组合起来，能够取得相当好的效果。对于只涉及x的查询，多列索引也可以使用，但它会比单独的x索引更大，因此更慢。
最后一种选择是同时创建这三个索引，但这大概只有在该表被搜索的频率远高于被更新的频率，
并且三类查询都很常见时才合理。如果其中一种查询远没有另外两种常见，
那么你大概只需创建最适合常见查询类型的两个索引即可。

11.5. 唯一索引

索引还可以用于强制列值唯一，或者强制多个列组合值唯一。

```
CREATE UNIQUE INDEX name ON table (column [, ...]);
```

当前，只有 B-树索引可以声明为唯一。

当一个索引被声明为唯一时，就不允许存在多个具有相同索引值的表行。空值不被视为相等。
多列唯一索引只会拒绝那种在多行中所有索引列都相等的情况。

当为表定义唯一约束或主键时，PostgreSQL会自动创建一个唯一索引。
该索引覆盖构成主键或唯一约束的列（必要时是多列索引），并且正是它在强制执行该约束。

> **注意**
>
> 向表中添加唯一约束的首选方式是ALTER TABLE ... ADD CONSTRAINT。用索引来强制唯一约束可以被认为是一种不应被直接访问的实现细节。不过应当知道，
> 无需手工在唯一列上创建索引；那样做只会重复自动创建的索引。

11.6. 表达式索引

索引列不一定非得是底层表中的某一列，
也可以是一个由表中一列或多列计算得到的函数或标量表达式。
这个特性有助于基于计算结果来快速访问表。

例如，进行大小写不敏感比较的一种常见方式，是使用`lower`函数：

```
SELECT * FROM test1 WHERE lower(col1) = 'value';
```

如果为`lower(col1)`函数的结果定义了索引，这个查询就可以使用该索引：

```
CREATE INDEX test1_lower_col1_idx ON test1 (lower(col1));
```

如果我们把这个索引声明为`UNIQUE`，那么它不仅会阻止创建那些`col1`
值仅在大小写上不同的行，也会阻止创建`col1`值完全相同的行。因此，
表达式索引可用于强制那些无法定义为简单唯一约束的约束。

再例如，如果你经常执行如下查询：

```
SELECT * FROM people WHERE (first_name || ' ' || last_name) = 'John
 Smith';
```

那么创建一个如下索引可能是值得的：

```
CREATE INDEX people_names ON people ((first_name || ' ' || last_name)
);
```

正如第二个示例所示，`CREATE INDEX`命令的语法通常要求对索引表达式写上圆括号。当表达式只是一个函数调用时，则可以像第一个示例那样省略这些圆括号。

维护索引表达式的代价相对较高，因为在插入每一行时以及每次更新时都必须计算派生表达式。
不过，在使用索引搜索时，索引表达式不会被重新计算，因为它们已经存储在索引中。
在上面的两个示例里，系统看到的查询只是`WHERE indexedcolumn = 'constant'`，因此搜索速度与其他简单索引查询相同。因此，当检索速度比插入和更新速度更重要时，表达式索引会很有用。

11.7. 部分索引

部分索引是建立在表的一个子集上的索引；
这个子集由一个条件表达式定义（称为部分索引的谓词）。
索引中只包含满足该谓词的表行的项。部分索引是一种专门特性，但有几种场景下它会很有用。

使用部分索引的一个主要原因是避免索引常见值。
由于搜索常见值（即占全部表行百分之几以上的值）的查询反正也不会使用索引，
因此完全没有必要把这些行保留在索引里。这会减小索引尺寸，
从而加快那些确实会使用该索引的查询。它也会加快很多表更新操作，
因为索引并不需要在所有情况下都更新。例 11.1展示了这种思路的一种可能应用。

例 11.1. 建立一个部分索引以排除常见值

假设你把 Web 服务器访问日志存储在数据库中。大多数访问来自你所在组织的 IP 地址范围，但也有一些来自其他地方（例如使用拨号连接的员工）。如果你按 IP 搜索时主要关心外部访问，那么你大概没有必要为对应于组织内子网的 IP 范围建索引。

假设有一个如下表：

```
CREATE TABLE access_log (
    url varchar,
    client_ip inet,
    ...
);
```

要创建一个适合这个例子的部分索引，可使用如下命令：

```
CREATE INDEX access_log_client_ip_ix ON access_log (client_ip)
WHERE NOT (client_ip > inet '192.168.100.0' AND
           client_ip < inet '192.168.100.255');
```

一个能够使用此索引的典型查询如下：

```
SELECT *
FROM access_log
WHERE url = '/index.html' AND client_ip = inet '212.78.10.32';
```

以下查询无法使用此索引：

```
SELECT *
FROM access_log
WHERE client_ip = inet '192.168.100.23';
```

请注意，这类部分索引要求常见值事先可知，因此最适合用于数据分布不会变化的场景。
也可以偶尔重建这类索引以适应新的数据分布，但那会增加维护工作量。

部分索引的另一种可能用途，是把典型查询工作负载不感兴趣的值排除在索引之外，如例 11.2所示。这样会得到与上面相同的好处，但也意味着这些“不感兴趣”的值无法通过该索引访问，
即使在那种情况下索引扫描可能是有利的。显然，
为这种场景建立部分索引需要大量的谨慎和实验。

例 11.2. 建立一个部分索引以排除不感兴趣的值

如果你有一张表，其中同时包含已开账单和未开账单的订单，
而未开账单订单只占整张表的一小部分，却是最常被访问的那些行，
那么你可以通过只为未开账单的行创建索引来提升性能。创建该索引的命令如下：

```
CREATE INDEX orders_unbilled_index ON orders (order_nr)
    WHERE billed is not true;
```

一个可能会使用该索引的查询是：

```
SELECT * FROM orders WHERE billed is not true AND order_nr < 10000;
```

不过，这个索引也可以用于那些完全不涉及`order_nr`的查询，例如：

```
SELECT * FROM orders WHERE billed is not true AND amount > 5000.00;
```

这不像在`amount`列上建立部分索引那样高效，因为系统必须扫描整个索引。不过，
如果未开账单的订单相对较少，那么即使只是用这个部分索引来找出未开账单订单，
也可能是值得的。

请注意，下面这个查询不能使用该索引：

```
SELECT * FROM orders WHERE order_nr = 3501;
```

因为订单 3501 可能属于已开账单订单，也可能属于未开账单订单。

例 11.2也说明了：索引列和谓词中使用的列不必一致。PostgreSQL支持带任意谓词的部分索引，只要其中只涉及正在建立索引的那张表的列。不过要记住，谓词必须与那些希望从该索引受益的查询中使用的条件相匹配。更准确地说，只有当系统能够识别出查询的`WHERE`条件在数学上蕴含该索引的谓词时，部分索引才能用于该查询。PostgreSQL并没有一个复杂的定理证明器，来识别那些写法不同但数学上等价的表达式。（不仅构建这样一个通用定理证明器极其困难，而且它很可能也会慢到失去实际用途。）系统可以识别简单的不等式蕴含，例如“x < 1”蕴含“x < 2”；否则，谓词条件必须与查询`WHERE`条件的某一部分完全匹配，否则索引不会被识别为可用。

部分索引的第三种可能用途，甚至不要求索引被查询使用。这里的思路是像例 11.3那样，在表的一个子集上创建唯一索引。这样就能在满足索引谓词的那些行之间强制唯一性，而不会约束不满足谓词的行。

例 11.3. 建立一个部分唯一索引

假设我们有一张描述测试结果的表。我们希望确保对于给定的测试对象和目标组合，只有一条“成功”记录，但可以有任意多条“失败”记录。实现方法之一如下：

```
CREATE TABLE tests (
    subject text,
    target text,
    success boolean,
    ...
);

CREATE UNIQUE INDEX tests_success_constraint ON tests (subject,
 target)
    WHERE success;
```

当成功测试很少而失败测试很多时，这是一种特别高效的方法。

最后，部分索引还可以用来影响系统的查询计划选择。同样，分布异常的数据集可能导致系统在实际上不应该使用索引的时候选择使用它。在这种情况下，可以把索引建成对那个有问题的查询不可用。通常，PostgreSQL会对索引使用作出合理选择（例如，它会在检索常见值时避免使用索引，因此前面的例子实际上只是节省索引空间，而不是为了避免使用索引所必需的），如果出现明显错误的计划选择，那就应当提交 bug 报告。

请记住，建立部分索引意味着你至少与查询规划器一样了解情况，尤其是你知道索引何时可能是有利的。形成这种认识需要经验，以及对PostgreSQL中索引工作方式的理解。在大多数情况下，部分索引相比普通索引的优势都很小。

关于部分索引的更多信息可参见[ston89b]、[olson93]和[seshadri95]。

11.8. 操作符类

索引定义可以为索引的每一列指定一个操作符类。

```
CREATE INDEX name ON table (column opclass [, ...]);
```

操作符类确定索引针对该列使用哪些操作符。例如，建立在类型`int4`上的 B-树索引会使用`int4_ops`类；该操作符类包含针对以下类型的值的比较函数：`int4`。实际中，

列数据类型的默认操作符类通常就足够了。之所以需要操作符类，
主要是因为对于某些数据类型，可能存在不止一种有意义的索引行为。例如，
可能希望按绝对值或实部对复数数据类型排序。可以为该数据类型定义两个操作符类，
并在创建索引时选择合适的类。

除了默认操作符类之外，还有一些内置操作符类：

- 操作符类text_pattern_ops、varchar_pattern_ops、bpchar_pattern_ops和name_pattern_ops分别支持类型text、varchar、char和name上的 B-树索引。它们与默认操作符类的区别在于，值是严格按字符逐个比较，
  而不是按照区域设置相关的排序规则比较。因此，当服务器未使用标准“C”区域设置时，
  这些操作符类适用于涉及模式匹配表达式（LIKE或 POSIX 正则表达式）的查询。例如，你可以像这样为一个varchar列建立索引：

  ```
  CREATE INDEX test_index ON test_table (col varchar_pattern_ops);
  ```

  注意，如果你希望涉及普通比较的查询也能使用索引，
  那么还应该再创建一个使用默认操作符类的索引。这类查询不能使用*xxx*_pattern_ops操作符类。同一列上可以创建多个使用不同操作符类的索引。如果你使用的是 C 区域设置，则不需要*xxx*_pattern_ops操作符类，因为带默认操作符类的索引在 C 区域设置下就可以用于模式匹配查询。

下面的查询会显示所有已定义的操作符类：

```
SELECT am.amname AS index_method,
       opc.opcname AS opclass_name
    FROM pg_am am, pg_opclass opc
    WHERE opc.opcamid = am.oid
    ORDER BY index_method, opclass_name;
```

还可以扩展该查询，显示每个类中包含的所有操作符：

```
SELECT am.amname AS index_method,
       opc.opcname AS opclass_name,
       opr.oid::regoperator AS opclass_operator
    FROM pg_am am, pg_opclass opc, pg_amop amop, pg_operator opr
    WHERE opc.opcamid = am.oid AND
          amop.amopclaid = opc.oid AND
          amop.amopopr = opr.oid
    ORDER BY index_method, opclass_name, opclass_operator;
```

11.9. 检查索引使用情况

尽管PostgreSQL中的索引不需要维护或调优，
但检查真实查询工作负载到底实际使用了哪些索引，仍然很重要。
检查某个单独查询的索引使用情况，可以使用EXPLAIN命令；它在这方面的应用见第 13.1 节。也可以像第 24.2 节中描述的那样，在运行中的服务器上收集索引使用的总体统计信息。

很难给出一个确定应该创建哪些索引的通用流程。前面各节的示例已经展示了若干典型情形。
通常需要做大量实验。下面这部分会给出一些相关提示：

- 一定要先运行ANALYZE。这个命令会收集表中值分布的统计信息。估计查询会返回多少行，
  需要这些信息；而规划器要为每种可能的查询计划分配现实的代价，也需要这些信息。
  如果没有真实统计信息，系统就会假定一些默认值，而这些默认值几乎肯定并不准确。因此，
  在没有运行ANALYZE的情况下去检查应用的索引使用情况，基本上是徒劳的。

- 用真实数据做实验。使用测试数据来设置索引，只会告诉你测试数据需要什么索引，仅此而已。

  特别要避免使用非常小的测试数据集。虽然在 100000 行里选 1000 行可能适合用索引，但在 100 行里选 1 行几乎不适合，因为这 100 行大概都能放进一个磁盘页，而没有任何计划能比顺序取一个磁盘页更快。

  另外，在编造测试数据时也要小心；如果应用尚未投入生产，这往往无法避免。非常相似、完全随机、或者按排序顺序插入的值，都会让统计信息偏离真实数据应有的分布。

- 当索引没有被使用时，强制使用它们对测试会很有帮助。
  有一些运行时参数可以关闭不同的计划类型（见第 17.6.1 节）。例如，关闭顺序扫描（`enable_seqscan`）和嵌套循环连接（`enable_nestloop`）这两种最基本的计划，就会迫使系统使用不同的计划。如果系统仍然选择顺序扫描或嵌套循环连接，那么索引未被使用很可能有更根本的原因，例如查询条件并不匹配该索引。（前面各节已经解释了什么样的查询可以使用什么样的索引。）

- 如果强制使用索引后确实使用了索引，那么就有两种可能：要么系统是对的，使用索引确实不合适；要么查询计划的代价估计并未反映现实。因此，你应该分别在有索引和无索引的情况下对查询计时。`EXPLAIN ANALYZE`命令在这里会很有帮助。

- 如果发现代价估计是错的，那么又有两种可能。
  总代价是根据每个计划节点的逐行代价乘以该计划节点的选择率估计计算出来的。计划节点的估计代价可以通过运行时参数来调整（见第 17.6.2 节）。而选择率估计不准确，则是因为统计信息不足。可以尝试通过调整统计信息收集参数来改进这一点（见ALTER TABLE）。

  如果你无法把这些代价调得更合适，那么可能就不得不退而求其次，显式强制使用索引。你也可能希望联系PostgreSQL开发者来研究这个问题。

第 12 章 并发控制

本章描述 PostgreSQL 数据库系统在两个或更多会话试图同时访问同一数据时的行为。在这种情况下，目标是在保持严格数据完整性的同时，为所有会话提供高效访问。
每一个数据库应用开发者都应该熟悉本章涉及的主题。

12.1. 介绍

与传统数据库系统使用锁进行并发控制不同，
PostgreSQL通过使用多版本模型（多版本并发控制，MVCC）来维护数据一致性。这意味着，
在查询数据库时，每个事务看到的都是某个较早时刻的数据快照（即一个数据库版本），
而不受底层数据当前状态的影响。这样可以保护事务，使其看不到由（其他）
并发事务在同一数据行上执行更新所产生的不一致数据，从而为每个数据库会话提供事务隔离。

使用MVCC并发控制模型而不是加锁的主要优点在于，在MVCC中，为查询（读取）
数据而获得的锁不会与为写入数据而获得的锁冲突，因此读取不会阻塞写入，
写入也不会阻塞读取。

对于无法轻易适应MVCC行为的应用，PostgreSQL也提供表级和行级锁设施。不过，正确使用
MVCC通常会比加锁提供更好的性能。

12.2. 事务隔离

SQL标准根据并发事务之间必须防止的三个现象，定义了四个事务隔离级别。
这些不受欢迎的现象是：

脏读

一个事务读取了由另一个并发未提交事务写入的数据。

不可重复读

一个事务重新读取它之前读过的数据，
发现这些数据已经被另一个事务修改（且该事务已在初次读取之后提交）。

幻读

一个事务重新执行一个返回满足某个搜索条件的行集合的查询，
发现满足该条件的行集合由于另一个最近提交的事务而发生了变化。

四种事务隔离级别及其对应行为见表 12.1。

表 12.1. 标准 SQL 事务隔离级别

Isolation Level	Dirty Read	Nonrepeatable Read	Phantom Read
Read uncommitted	Possible	Possible	Possible
Read committed	Not possible	Possible	Possible
Repeatable read	Not possible	Not possible	Possible
Serializable	Not possible	Not possible	Not possible

在PostgreSQL中，你可以请求四种标准事务隔离级别中的任意一种。但在内部，
实际上只有两种不同的隔离级别，分别对应读已提交和可串行化这两个级别。
当你选择读未提交级别时，实际得到的是读已提交；而当你选择可重复读时，

实际得到的是可串行化，因此实际的隔离级别可能比你选择的更严格。这是 SQL 标准所允许的：这四个隔离级别只定义了哪些现象不允许发生，并没有定义哪些现象必须发生。PostgreSQL之所以只提供两种隔离级别，是因为要把标准隔离级别映射到多版本并发控制架构上，这是唯一合理的做法。可用隔离级别的行为将在下面各小节中详细说明。

要设置事务的事务隔离级别，请使用命令SET TRANSACTION。

12.2.1. 读已提交隔离级别

读已提交是PostgreSQL中默认的隔离级别。当事务使用这一隔离级别时，`SELECT` 查询只能看到查询开始之前已经提交的数据；它既看不到未提交的数据，也看不到查询执行期间并发事务提交的更改。实际上，`SELECT` 查询看到的是该查询开始运行瞬间的数据库快照。不过，`SELECT` 能看到其自身事务中先前执行的更新效果，即使这些更新尚未提交。还要注意，即使两个连续的`SELECT`命令位于同一个事务中，如果其他事务在第一个`SELECT`执行期间提交了更改，这两个命令也可能看到不同的数据。

`UPDATE`、`DELETE`、`SELECT FOR UPDATE` 和 `SELECT FOR SHARE` 命令在搜索目标行时的行为与 `SELECT` 相同：它们只会找到在命令开始时已经提交的目标行。不过，当找到这样的目标行时，它可能已经被其他并发事务更新、删除或者锁定。在这种情况下，即将执行更新的事务会等待第一个更新事务提交或回滚（如果它仍在进行中）。如果第一个更新事务回滚，那么它的作用将被撤销，第二个更新事务就可以继续更新最初找到的行。如果第一个更新事务提交了，那么若该行已被第一个更新者删除，第二个更新事务就会忽略该行；否则，第二个更新者将尝试在该行的已更新版本上应用自己的操作。该命令的搜索条件（`WHERE` 子句）会被重新计算，以判断该行的已更新版本是否仍然符合搜索条件。如果符合，则第二个更新者会基于该行的已更新版本继续执行操作。对于 `SELECT FOR UPDATE` 和 `SELECT FOR SHARE`，这意味着被锁住并返回给客户端的是该行的已更新版本。

由于上述规则，更新命令可能看到一个不一致的快照：它能够看到并发更新命令在它试图更新的同一行上的效果，却看不到这些命令对数据库中其他行的影响。这种行为使得读已提交模式不适合涉及复杂搜索条件的命令；不过，对于较简单的场景它恰到好处。例如，考虑以如下事务更新银行余额：

```
BEGIN;
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 12345;
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 7534;
COMMIT;
```

如果两个这样的事务并发地尝试更改账户 12345 的余额，我们显然希望第二个事务从该账户行的已更新版本开始。因为每个命令只影响一个预先确定的行，让它看到该行的已更新版本不会造成任何麻烦的不一致。

因为在读已提交模式中，每个新命令都从一个新快照开始，而该快照包含截至当时已提交的所有事务，所以同一事务中的后续命令无论如何都会看到并发事务已提交的效果。这里讨论的问题在于，在单个命令内部，我们是否能看到数据库的一个绝对一致视图。

读已提交模式所提供的部分事务隔离对于许多应用来说已经足够，而且这种模式既快速又容易使用。不过，它并不能满足所有场景。执行复杂查询和更新的应用，可能需要比读已提交模式所提供的更严格一致的数据库视图。

12.2.2. 可串行化隔离级别

可串行化隔离级别提供最严格的事务隔离。该级别模拟串行事务执行，
就像事务是一个接一个串行执行而不是并发执行一样。不过，
使用该级别的应用必须准备好在发生串行化失败时重试事务。

当事务使用可串行化级别时，SELECT查询只能看到事务开始之前已经提交的数据；
它既看不到未提交的数据，也看不到事务执行期间并发事务提交的更改。（不过，SELECT
能看到其自身事务中先前执行的更新效果，即使这些更新尚未提交。）这一点与读已提交不同：
SELECT看到的是事务开始时刻的快照，而不是事务中当前查询开始时刻的快照。因此，
单个事务中连续的SELECT命令看到相同的数据，也就是说，
它们看不到在本事务启动之后由其他事务提交的更改。
（这种行为对于报表类应用来说可能非常理想。）

UPDATE、DELETE、SELECT FOR UPDATE和SELECT FOR SHARE命令在搜索目标行方面的行为
与SELECT相同：它们只会找到事务开始时刻已提交的目标行。然而，这样的目标行在被找到时，
可能已经被另一个并发事务更新（或删除、锁定）。在这种情况下，
可串行化事务将等待第一个更新事务提交或回滚（如果仍在进行）。如果第一个更新事务回滚，
那么它的作用将被撤销，可串行化事务就可以继续更新最初找到的行。
但如果第一个更新事务提交了（并且确实更新或删除了该行，而不仅仅是锁定它），
那么可串行化事务将被回滚，并收到以下消息：

```
ERROR:  could not serialize access due to concurrent update
```

这是因为可串行化事务不能修改或锁定在它开始之后被其他事务更改过的行。

当应用收到这条错误消息时，应当中止当前事务，并从头重试整个事务。第二次执行时，
该事务会把先前已提交的更改视为其初始数据库视图的一部分，
因此以该行的新版本作为新事务更新的起点时，就不存在逻辑冲突。

注意，只有更新事务才可能需要重试；只读事务永远不会发生串行化冲突。

可串行化模式严格保证每个事务看到的都是完全一致的数据库视图。然而，
当并发更新使得串行执行的假象无法维持时，应用必须准备好重试事务。
由于重做复杂事务的代价可能相当可观，
只有当更新事务包含的逻辑复杂到在读已提交模式下可能给出错误结果时，
才推荐使用可串行化模式。最常见的情况是，
当一个事务要执行多条必须看到相同数据库视图的连续命令时，就需要使用可串行化模式。

12.2.2.1. 可串行化隔离与真正的可串行化

“可串行化”执行的直观含义（以及数学定义）是：任何两个成功提交的并发事务，
看起来都像严格串行地一个接一个执行过一样——尽管哪一个看起来先发生未必能事先预料。
必须认识到，仅仅禁止表 12.1中列出的那些不受欢迎的行为，并不足以保证真正的可串行性；事实上，PostgreSQL的可串行化模式并不保证这个意义上的可串行化执行。作为一个例子，考虑表mytab，它最初包含：

```
 class | value
-------+-------
     1 |    10
     1 |    20
     2 |   100
     2 |   200
```

假设可串行化事务 A 执行以下计算：

```
SELECT SUM(value) FROM mytab WHERE class = 1;
```

然后将结果（30）作为新行的value插入，且该行的class为 2。与此同时，可串行化事务 B 执行以下计算：

```
SELECT SUM(value) FROM mytab WHERE class = 2;
```

得到结果 300，并将它插入一个新行，该行的class为 1。然后两个事务都提交。上面列出的任何不受欢迎的行为都没有发生，然而我们得到的结果，却不可能由任何一种串行顺序产生。如果 A 先于 B 执行，B 计算出的总和将是 330 而不是 300；类似地，另一种执行顺序也会使 A 计算出不同的总和。

要保证真正的数学可串行性，数据库系统必须实现谓词锁，也就是说，一个事务不能插入或修改这样一个行：它会匹配另一个并发事务中某条查询的WHERE条件。例如，一旦事务 A 执行了查询SELECT ... WHERE class = 1，一个采用谓词锁的系统就会禁止事务 B 插入任何 class 为 1 的新行，直到 A 提交为止。[1]这样的锁系统实现起来很复杂，执行起来的代价也极其高昂，因为每个会话都必须了解每个并发事务执行的每条查询的细节。而且这笔巨大的开销大部分都被浪费掉了，因为在实践中，大多数应用并不会做那些可能造成问题的事情。（当然，上面的例子相当人为，也不太可能代表真实的软件。）据此，PostgreSQL没有实现谓词锁，而且据我们所知，其他生产型 DBMS 也都没有实现。

在非可串行化执行的可能性构成实际风险的场合，可以通过恰当地使用显式锁定来防止问题。进一步的讨论见后续各节。

12.3. 显式锁定

PostgreSQL提供多种锁模式来控制对表中数据的并发访问。当MVCC无法给出期望的行为时，这些模式可用于由应用控制的锁定。此外，大多数PostgreSQL命令会自动获取适当模式的锁，以确保被引用的表在命令执行期间不会以不兼容的方式被删除或修改。（例如，ALTER TABLE不能与同一张表上的其他操作并发执行。）

要查看数据库服务器中当前尚未释放的锁列表，可以使用pg_locks系统视图（第 42.37 节）。有关监控锁管理器子系统状态的更多信息，请参见第 24 章。

12.3.1. 表级锁

下面的列表给出了可用的锁模式，以及它们在 PostgreSQL 中被自动使用的场景。你也可以通过命令LOCK显式获取其中任意一种锁。请记住，这些锁模式全部都是表级锁，即使名称中包含“row”一词；这些名称只是历史遗留。在某种程度上，这些名称反映了各锁模式的典型用途 – 但它们的语义是完全相同的。一个锁模式与另一个锁模式真正的区别，只在于它与哪些锁模式冲突。两个事务不能在同一时刻在同一张表上持有相互冲突模式的锁。（不过，事务永远不会与自己冲突。例如，它可能先获取 ACCESS EXCLUSIVE 锁，随后又在同一张表上获取 ACCESS SHARE 锁。）不冲突的锁模式可以被多个事务同时持有。锁一旦被获取，就会一直持有到事务结束。特别要注意，有些锁模式与自身冲突（例如，一次只能有一个事务持有 ACCESS EXCLUSIVE 锁），而另一些锁模式并不与自身冲突（例如，可以有多个事务持有 ACCESS SHARE 锁）。

表级锁模式

ACCESS SHARE

只与 ACCESS EXCLUSIVE 锁模式冲突。

[1] 本质上，谓词锁系统是通过限制写入的内容来防止幻读的，而 MVCC 则是通过限制读取的内容来防止幻读。

SELECT和ANALYZE命令会在被引用的表上获取这种模式的锁。通常，任何只读取表而不修改它的查询都会获取这种锁模式。

ROW SHARE

与 EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。

SELECT FOR UPDATE 和 SELECT FOR SHARE 命令会在目标表上获取这种模式的锁（此外，对于引用到但没有用 FOR UPDATE/FOR SHARE 选取的其他表，还会获取 ACCESS SHARE 锁）。

ROW EXCLUSIVE

与SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE和ACCESS EXCLUSIVE锁模式冲突。

UPDATE、DELETE 和 INSERT 命令会在目标表上获取这种锁模式（此外，对任何其他被引用的表还会获取 ACCESS SHARE 锁）。通常，任何修改表中数据的命令都会获取这种锁模式。

SHARE UPDATE EXCLUSIVE

与 SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式用于保护表不受并发模式更改和 VACUUM 运行的影响。

由VACUUM（不带FULL）获取。

SHARE

与 ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式用于保护表不受并发数据更改的影响。

由CREATE INDEX获取。

SHARE ROW EXCLUSIVE

与ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE和ACCESS EXCLUSIVE锁模式冲突。

这种锁模式不会被任何 PostgreSQL 命令自动获取。

EXCLUSIVE

与 ROW SHARE、ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE 锁模式冲突。这种模式只允许并发的 ACCESS SHARE 锁，也就是说，只有对该表的读操作可以与持有这种锁模式的事务并行执行。

这种锁模式不会被任何PostgreSQL命令自动获取到用户表上。不过，在某些操作中，它会被获取到某些系统目录上。

ACCESS EXCLUSIVE

与所有模式的锁冲突（ACCESS SHARE、ROW SHARE、ROW EXCLUSIVE、SHARE UPDATE EXCLUSIVE、SHARE、SHARE ROW EXCLUSIVE、EXCLUSIVE 和 ACCESS EXCLUSIVE）。这种模式保证持有者是以任何方式访问该表的唯一事务。

由ALTER TABLE、DROP TABLE、REINDEX、CLUSTER和VACUUM FULL命令获取。这也是未显式指定模式的LOCK TABLE语句的默认锁模式。

> **提示**
>
> 只有 ACCESS EXCLUSIVE 锁才会阻塞 SELECT（不带 FOR UPDATE/SHARE）语句。

12.3.2. 行级锁

除了表级锁之外，还有行级锁，行级锁可以是排他锁或共享锁。当一行被更新或删除时，会自动获得该行上的排他行级锁。这种锁一直保持到事务提交或回滚为止。行级锁不影响数据查询；它们只会阻塞对同一行的写入者。

要在不实际修改某行的情况下获取该行上的排他行级锁，可以使用SELECT FOR UPDATE选择该行。注意，一旦获得了行级锁，事务就可以多次更新该行而无须担心冲突。

要获取某行上的共享行级锁，可以使用SELECT FOR SHARE选择该行。共享锁不会阻止其他事务获取同样的共享锁。然而，任何事务都不允许更新、删除或排他锁定一个已被其他事务持有共享锁的行。任何这样的尝试都将阻塞，直到那些共享锁被释放。

PostgreSQL 不会在内存中保存任何关于已修改行的信息，因此一次加锁的行数没有限制。不过，锁定一行可能会导致一次磁盘写；例如，SELECT FOR UPDATE 会修改被选中的行以标记它们已被锁定，因此会产生磁盘写入。

除了表级锁和行级锁之外，还使用页级共享/排他锁来控制对共享缓冲池中表页面的读写访问。这些锁会在取出或更新某一行后立即释放。应用开发者通常不需要关心页级锁，这里提到它们只是为了完整性。

12.3.3. 死锁

使用显式锁可能会增加死锁的发生概率。所谓死锁，是指两个（或更多）事务各自持有对方想要的锁。例如，如果事务 1 在表 A 上获得了排他锁，然后试图再获取表 B 上的排他锁，而事务 2 已经持有表 B 上的排他锁，现在又想获取表 A 上的排他锁，那么两者都无法继续进行。PostgreSQL 会自动检测死锁，并通过中止其中一个事务来解决这个问题，使其他事务得以继续完成。（具体会中止哪个事务很难预测，也不应依赖这种预测。）

还要注意，死锁也可能由于行级锁而发生（因此，即使没有使用显式锁，它们也可能出现）。考虑下面这种情况：两个并发事务都在修改一个表。第一个事务执行：

```
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 11111;
```

这会在指定账号的那一行上获取一个行级锁。然后第二个事务执行：

```
UPDATE accounts SET balance = balance + 100.00 WHERE acctnum = 22222;
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 11111;
```

第一条 UPDATE 语句成功地在指定行上获取了一个行级锁，因此它成功更新了那一行。然而，第二条 UPDATE 语句发现它试图更新的行已经被锁住了，于是它等待持有该锁的事务结束。此时，事务二正在等待事务一结束之后才能继续执行。现在，事务一执行：

```
UPDATE accounts SET balance = balance - 100.00 WHERE acctnum = 22222;
```

事务一试图在指定行上获取一个行级锁，但它做不到，因为事务二已经持有了这个锁。于是它必须等待事务二完成。这样一来，事务一被事务二阻塞，

而事务二又被事务一阻塞：死锁条件形成了。PostgreSQL 会检测到这种情况，并中止其中一个事务。

防止死锁的最好办法通常是避免它们出现，
也就是确保所有使用同一数据库的应用都以一致的顺序在多个对象上获取锁。在上面的例子中，
如果两个事务都是按相同顺序更新这些行，就不会发生死锁。
还应确保事务在某个对象上获取的第一个锁，就是该对象所需的最严格锁模式。
如果事先无法验证这一点，那么就应通过重试因死锁而中止的事务来动态处理死锁。

只要没有检测到死锁，寻求表级锁或行级锁的事务就会无限期等待冲突锁被释放。
这意味着让应用长时间保持事务打开并不是好主意（例如等待用户输入时）。

12.4. 应用级别的数据一致性检查

由于PostgreSQL中的读操作不加锁（无论事务隔离级别如何），
一个事务读取的数据仍可能被另一个并发事务覆盖。换句话说，某行被`SELECT`返回，
并不意味着该行在返回的那一刻（即当前查询开始之后的某个时刻）仍然是当前的。
该行可能已经被一个在本事务开始之后提交的事务修改或删除。即使该行“现在”仍然有效，
它也可能在当前事务提交或回滚之前被更改或删除。

换一种方式来看：每个事务看到的都是数据库内容的一个快照，
而并发执行的事务很可能看到不同的快照。因此，“现在”
这个概念本身在某种程度上就是没有明确定义的。如果各个客户端应用彼此隔离，
这通常不是大问题；但如果客户端能够通过数据库之外的渠道通信，就可能引起严重的混乱。

要确保某一行当前仍然有效，并保护它不受并发更新影响，就必须使用`SELECT FOR UPDATE`、`SELECT FOR SHARE`或适当的`LOCK TABLE`语句。（`SELECT FOR UPDATE`和`SELECT FOR SHARE`只锁定返回的行以防止并发更新，而`LOCK TABLE`会锁住整张表。）从其他环境向PostgreSQL迁移应用时，应当考虑这一点。（在 6.5 版之前，PostgreSQL使用的是读锁，因此在从 6.5 之前的PostgreSQL版本升级时，上面这个考虑也同样适用。）

在MVCC之下，全局有效性检查需要额外考虑。例如，
一个银行应用可能希望检查一个表中的贷方金额总和等于另一个表中的借方金额总和，
而这两个表都在被活跃更新。在读已提交模式下，比较两个连续的`SELECT sum(...)`命令的结果并不可靠，因为第二个查询很可能会包含第一个查询没有统计到的事务提交结果。
在单个可串行化事务中完成这两次求和，
只能准确反映在该可串行化事务开始之前已提交事务的效果——但等到结果交付时，
人们完全可能合理地怀疑这个答案是否仍然相关。
如果可串行化事务本身在尝试进行一致性检查之前已经应用了某些更改，
那么这种检查的意义就更加值得商榷，因为此时它包含了事务开始后的一部分而不是全部更改。
在这种情况下，谨慎的做法可能是锁定执行检查所需的所有表，
以获得当前真实状态的无可争议图景。`SHARE`模式（或更高）的锁能够保证，在被锁定的表中，
除了当前事务自身的更改之外，没有其他未提交更改。

注意，如果依赖显式锁定来防止并发更改，就应当使用读已提交模式，
或者在可串行化模式下小心地在执行查询之前先获取锁。
可串行化事务获得的锁能够保证没有其他修改该表的事务仍在运行，
但如果该事务看到的快照早于获取锁的时刻，
那么它看到的快照也可能早于表中某些现在已经提交的更改。
可串行化事务的快照实际上是在其第一条查询或数据修改命令（`SELECT`、`INSERT`、`UPDATE`、`DELETE`）开始时被冻结的，因此可以在快照冻结之前显式获取锁。

12.5. 锁定和索引

尽管PostgreSQL对表数据提供了非阻塞的读写访问，但对于
PostgreSQL中实现的每一种索引访问方法，目前并非都能提供非阻塞的读写访问。
各种索引类型的处理方式如下：

B-树和GiST索引

读/写访问使用短期的共享/排他页级锁。锁在每次索引行被取出或插入之后立即释放。
这些索引类型在不产生死锁条件的情况下提供了最高的并发度。

Hash 索引

读/写访问使用共享/排他的哈希桶级锁。锁在整个桶被处理完之后释放。
桶级锁比索引级锁提供更好的并发度，但由于锁的持有时间超过一次索引操作，
可能出现死锁。

R 树索引

读/写访问使用共享/排他的索引级锁。锁在整个命令完成之后释放。

目前，B-树索引为并发应用提供了最佳性能；由于它们还比 Hash 索引拥有更多特性，因此对于需要为标量数据建立索引的并发应用，推荐使用 B-树索引类型。处理非标量数据时，B-树就不再适用，此时应改用 GiST 索引。R 树索引已被弃用，很可能在未来的发行版中完全消失。

第 13 章 性能提示

查询性能可能受许多因素影响。其中一些因素可以由用户控制，
另一些则属于系统底层设计的基本特性。本章提供一些帮助理解和调优PostgreSQL性能的提示。

13.1. 使用EXPLAIN

PostgreSQL会为收到的每个查询制定一个查询计划。
选择与查询结构和数据特性相匹配的正确计划，对获得良好的性能至关重要，
因此系统内置了一个复杂的规划器来尽量选出好的计划。
可以使用EXPLAIN命令查看规划器为任意查询生成的查询计划。
读懂计划是一门值得专门教程详述的技艺，本节并非那样的教程，但这里会介绍一些基础知识。

查询计划的结构是一棵由计划节点组成的树。树的最底层是扫描节点，它们从表中返回原始行。
不同的表访问方法对应不同类型的扫描节点，例如顺序扫描、索引扫描和位图索引扫描。
如果查询需要对原始行执行连接、聚合、排序或其他操作，那么扫描节点“之上”
还会出现额外的节点来完成这些操作。同样，这些操作通常不止一种实现方式，
因此这里也会出现不同的节点类型。EXPLAIN会为计划树中的每个节点输出一行，
显示基本节点类型以及规划器对该计划节点执行代价的估计值。第一行（最顶层节点）
给出了整个计划的估计总执行代价；规划器力图最小化的正是这个数字。

这是一个最简单的例子，只为了展示输出的样子：[1]

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..458.00 rows=10000 width=244)
```

EXPLAIN引用的这些数字依次是（从左到右）：

- 估计启动代价（在输出扫描开始之前消耗的时间，例如排序节点中完成排序所需的时间）
- 估计总代价（如果所有行都被检索；不过它们也可能不被检索，例如带LIMIT
  子句的查询会在付出Limit计划节点的输入节点的全部代价之前停止）
- 该计划节点输出的估计行数（同样，仅在完全执行的情况下）
- 该计划节点输出行的估计平均宽度（以字节计）

这些开销以磁盘页读取为单位来衡量；也就是说，按照定义，1.0 等于一次顺序磁盘页读取。（也会对 CPU 开销进行估计；它们会使用一些相当武断的换算系数转换成磁盘页单位。如果你想试验这些系数，可以参见第 17.6.2 节中的运行时配置参数列表。）

需要理解的一点是，上层节点的开销包含了其所有子节点的开销。还要注意，
这个开销只反映规划器关心的内容。特别是，它没有考虑将结果行传输给客户端所消耗的时间，
而这在实际耗时中可能是重要因素；但规划器会忽略这些代价，
因为它无法通过改变计划来影响它们。（我们相信每个正确的计划都会输出相同的行集。）

行数输出（rows）这个数字有些容易误解，因为它不是计划节点处理或扫描的行数。
这个数字通常更小，反映的是在该节点上应用的任何WHERE子句条件的估计选择率。
理想情况下，顶层的行数估计应当接近查询实际返回、更新或删除的行数。

[1] 本节中的示例取自执行过VACUUM ANALYZE的回归测试数据库，使用的是 8.1 开发版本源码。如果自行尝试这些示例，通常应当能得到相近的结果，但估计代价和行计数可能略有差异，因为ANALYZE生成的统计信息来自随机采样而非精确计数。

回到我们的示例：

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..458.00 rows=10000 width=244)
```

这些数字的来源非常直接。如果执行：

```
SELECT relpages, reltuples FROM pg_class WHERE relname = 'tenk1';
```

可以看到tenk1有 358 个磁盘页和 10000 行。因此估计代价为 358 次页读取，按定义每次代价为 1.0，再加上 10000 * cpu_tuple_cost，其典型值为 0.01（可以用SHOW cpu_tuple_cost查看）。

现在修改查询，添加一个WHERE条件：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 7000;

                         QUERY PLAN
------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..483.00 rows=7033 width=244)
   Filter: (unique1 < 7000)
```

注意，EXPLAIN的输出表明，WHERE子句被作为一个“过滤器”条件应用。
这表示计划节点会对扫描到的每一行检查该条件，并且只输出满足条件的行。由于有了WHERE子句，估计输出行数减少了。但扫描仍需访问全部 10000 行，因此代价没有下降；实际上还略有上升，以反映检查WHERE条件所消耗的额外 CPU 时间。

这条查询实际选出的行数是 7000，但行数估计只是近似值。如果重复这个实验，很可能会得到略有不同的估计值。此外，由于ANALYZE生成的统计信息来自该表的随机采样，
这个估计值可能在每次执行ANALYZE之后发生变化。

现在，让条件更严格一些：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 100;

                                  QUERY PLAN
-----------------------------------------------------------------------
--------
 Bitmap Heap Scan on tenk1  (cost=2.37..232.35 rows=106 width=244)
   Recheck Cond: (unique1 < 100)
   ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37 rows=106
 width=0)
         Index Cond: (unique1 < 100)
```

这里规划器决定使用两步计划：子计划节点访问索引，找出匹配索引条件的行的位置，
然后上层计划节点再从表本身取出这些行。分别取出各行的代价远高于顺序读取，
但由于不必访问表的所有页，这仍然比顺序扫描便宜。（使用两层计划的原因是，
上层计划节点在读取行之前，会先把索引确定的行位置按物理顺序排列，
以尽量降低逐行读取的代价。节点名称中的“位图”就是执行这种排序的机制。）

如果WHERE条件的选择性足够强，规划器可能转而使用一种“简单”索引扫描计划：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 3;
```

```
                                    QUERY PLAN
----------------------------------------------------------------------
--------
 Index Scan using tenk1_unique1 on tenk1  (cost=0.00..10.00 rows=2
 width=244)
   Index Cond: (unique1 < 3)
```

在这种计划中，表行按索引顺序取出，因此读取代价更高，但行数很少，
不值得为排序行位置付出额外代价。对于仅取出一行的查询，以及带有与索引顺序匹配的ORDER BY条件的查询，最常见到这种计划类型。

再向WHERE子句添加另一个条件：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 3 AND stringu1 = 'xxx';

                                    QUERY PLAN
----------------------------------------------------------------------
--------
 Index Scan using tenk1_unique1 on tenk1  (cost=0.00..10.01 rows=1
 width=244)
   Index Cond: (unique1 < 3)
   Filter: (stringu1 = 'xxx'::name)
```

新增的条件stringu1 = 'xxx'降低了估计输出行数，却没有降低代价，因为仍然必须访问同一组行。注意，stringu1子句不能用作索引条件（因为该索引只建立在unique1列上）。
它只能作为过滤条件，应用到通过索引取出的行上。因此，代价实际上略有增加，
以反映这项额外检查。

如果在WHERE引用的多个列上分别建有索引，规划器可能选择将这些索引做 AND 或 OR 组合：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 100 AND unique2 > 9000;

                                      QUERY PLAN
----------------------------------------------------------------------
---------------
 Bitmap Heap Scan on tenk1  (cost=11.27..49.11 rows=11 width=244)
   Recheck Cond: ((unique1 < 100) AND (unique2 > 9000))
   ->  BitmapAnd  (cost=11.27..11.27 rows=11 width=0)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0)
               Index Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique2  (cost=0.00..8.65
 rows=1042 width=0)
               Index Cond: (unique2 > 9000)
```

但这需要访问两个索引，因此与只使用一个索引、把另一个条件作为过滤条件相比，
并不一定更划算。如果改变所涉及的范围，就会看到计划随之变化。

使用之前讨论的列，尝试连接两个表：

```
EXPLAIN SELECT *
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;
```

```
                                        QUERY PLAN
----------------------------------------------------------------------
----------------
 Nested Loop  (cost=2.37..553.11 rows=106 width=488)
   ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244)
         Recheck Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0)
               Index Cond: (unique1 < 100)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..3.01
 rows=1 width=244)
         Index Cond: (t2.unique2 = t1.unique2)
```

在这个嵌套循环连接中，外层（上层）扫描就是前面见过的那种位图索引扫描，其代价和行计数也相同，因为`WHERE`子句`unique1 < 100`正是在该节点上应用的。`t1.unique2 = t2.unique2`子句此时还无关，因此不会影响外层扫描的行计数。对于内层（下层）扫描，当前外层扫描行的`unique2`值被代入内层索引扫描，产生一个形如`t2.unique2 = constant`的索引条件。因此我们得到的内层扫描计划和代价，与执行`EXPLAIN SELECT * FROM tenk2 WHERE unique2 = 42`之类查询所得到的相同。循环节点的代价则建立在外层扫描代价之上，再加上每个外层行都要执行一次内侧扫描的代价（这里是 106 * 3.01），以及少量连接处理的 CPU 时间。

本例中，连接的输出行数等于两个扫描行数的乘积，但并非所有情况都如此，因为可能还有同时涉及两个表的`WHERE`子句，因此只能在连接处应用，不能应用到任一输入扫描。例如，如果我们加上`WHERE ... AND t1.hundred < t2.hundred`，它会减少连接节点的输出行数，但不会改变任何一个输入扫描。

查看不同计划的一种办法，是使用第 17.6.1 节中所述的启用/禁用开关，强制规划器放弃它认为最廉价的策略。（这是一种粗糙的工具，但很有用。另见第 13.3 节。）

```
SET enable_nestloop = off;
EXPLAIN SELECT *
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;

                                          QUERY PLAN
----------------------------------------------------------------------
--------------------
 Hash Join  (cost=232.61..741.67 rows=106 width=488)
   Hash Cond: (t2.unique2 = t1.unique2)
   ->  Seq Scan on tenk2 t2  (cost=0.00..458.00 rows=10000 width=244)
   ->  Hash  (cost=232.35..232.35 rows=106 width=244)
         ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244)
               Recheck Cond: (unique1 < 100)
               ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.
37 rows=106 width=0)
                     Index Cond: (unique1 < 100)
```

这个计划打算用与之前相同的索引扫描取出`tenk1`中那 100 行感兴趣的行，把它们存入一个内存中的哈希表，然后对`tenk2`做顺序扫描，对每一行`tenk2`在哈希表中探测`t1.unique2 = t2.unique2`的可能匹配。读取`tenk1`并建立哈希表的代价是哈希连接的启动代价，因为在开始读取`tenk2`之前不会有任何输出。连接的总时间估计中还包含一笔不小的 CPU 时间开销，用于对

哈希表探测 10000 次。但注意，我们不是按 10000 乘以 232.35 来计费的；在这种计划类型中，哈希表的建立只做一次。

可以使用EXPLAIN ANALYZE检查规划器代价估计的准确性。这个命令会实际执行查询，然后显示每个计划节点内累计的真实运行时间，以及普通EXPLAIN所显示的相同估计代价。例如，我们可能得到这样的结果：

```
EXPLAIN ANALYZE SELECT *
FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 100 AND t1.unique2 = t2.unique2;

                                                              QUERY PLAN
----------------------------------------------------------------------
-------------------------------------------------------------
 Nested Loop  (cost=2.37..553.11 rows=106 width=488) (actual time=1.
392..12.700 rows=100 loops=1)
   ->  Bitmap Heap Scan on tenk1 t1  (cost=2.37..232.35 rows=106
 width=244) (actual time=0.878..2.367 rows=100 loops=1)
         Recheck Cond: (unique1 < 100)
         ->  Bitmap Index Scan on tenk1_unique1  (cost=0.00..2.37
 rows=106 width=0) (actual time=0.546..0.546 rows=100 loops=1)
               Index Cond: (unique1 < 100)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..3.01
 rows=1 width=244) (actual time=0.067..0.078 rows=1 loops=100)
         Index Cond: (t2.unique2 = t1.unique2)
 Total runtime: 14.452 ms
```

注意，“actual time”值以真实时间的毫秒为单位，而“cost”估计则以任意的磁盘页读取单位表示，因此两者不太可能相互吻合。需要关注的是这些比值。

在某些查询计划中，一个子计划节点可能会执行多次。例如，上面那个嵌套循环计划中的内侧索引扫描会对外侧的每一行执行一次。在这种情况下，“loops”值报告的是该节点的总执行次数，而显示的 actual time 和 rows 值是每次执行的平均值。这样做是为了让这些数字更容易与代价估计的展示方式相比较。将它们乘以“loops”值，就能得到该节点实际消耗的总时间。

EXPLAIN ANALYZE显示的Total runtime包括执行器启动和关闭所需的时间，以及处理结果行所花费的时间，但不包括解析、重写和规划时间。对于SELECT查询，总运行时间通常只是比顶层计划节点报告的总时间略大一点。而对于INSERT、UPDATE和DELETE命令，总运行时间可能会大得多，因为它包含处理结果行所花费的时间。对于这些命令，顶层计划节点的时间基本上就是定位旧行和/或计算新行所花费的时间，但不包括做出这些更改所花费的时间。触发触发器（如果有）所花费的时间也不在顶层计划节点之内，而是按每个触发器单独显示。

值得注意的是，不应把EXPLAIN的结果外推到实际测试场景之外的情况；例如，不能假定在很小的表上得到的结果也适用于大型表。规划器的代价估计不是线性的，因此它可能会为更大或更小的表选择不同的计划。一个极端示例是：对于只占用一个磁盘页的表，无论索引是否可用，几乎总会得到顺序扫描计划。规划器认识到，无论如何处理该表都需要一次磁盘页读取，因此再额外读取页面去查看索引并没有价值。

13.2. 规划器使用的统计信息

如上一节所见，查询规划器需要估计查询将检索多少行，才能对查询计划做出良好选择。本节简要介绍系统用于这些估计的统计信息。

统计信息的一部分是各个表和索引的条目总数，以及各个表和索引占用的磁盘块数。
这些信息保存在表pg_class的reltuples和relpages列中。
可以使用类似下面的查询来查看：

```
SELECT relname, relkind, reltuples, relpages FROM pg_class WHERE
 relname LIKE 'tenk1%';

       relname        | relkind | reltuples | relpages
----------------------+---------+-----------+----------
 tenk1                | r       |     10000 |      358
 tenk1_hundred        | i       |     10000 |       30
 tenk1_thous_tenthous | i       |     10000 |       30
 tenk1_unique1        | i       |     10000 |       30
 tenk1_unique2        | i       |     10000 |       30
(5 rows)
```

这里可以看到，tenk1包含 10000 行，它的索引也一样，但索引比表小得多（这并不意外）。

出于效率考虑，reltuples和relpages不会实时更新，因此它们通常包含有些过时的值。
它们会在VACUUM、ANALYZE以及少数 DDL 命令（如CREATE INDEX）执行时被更新。独立的ANALYZE（即不作为VACUUM一部分执行的那种）由于不会读取表的每一行，生成的reltuples值是近似的。规划器会把它在pg_class中找到的值按当前物理表大小进行缩放，从而得到更接近实际情况的近似值。

大多数查询只会检索表中一部分行，因为它们通过WHERE子句限制了需要检查的行。因此，
规划器需要估算WHERE子句的选择度，也就是满足WHERE子句中各个条件的行所占比例。
完成这项任务所需的信息存储在pg_statistic系统目录中。pg_statistic
中的条目由ANALYZE和VACUUM ANALYZE命令更新，而且即使刚更新完，也始终只是近似值。

手动检查统计信息时，与其直接查看pg_statistic，不如查看它的视图pg_stats。pg_stats旨在让统计信息更易读。此外，pg_stats所有用户都可以读取，而pg_statistic
只有超级用户可以读取。
（这可以防止没有相应权限的用户通过统计信息获知其他用户表中的内容。pg_stats
视图被限制为只显示当前用户有权读取的表的相关行。）例如，可以执行：

```
SELECT attname, n_distinct, most_common_vals FROM pg_stats WHERE
 tablename = 'road';

 attname | n_distinct |

                                                                most_
common_vals


---------+------------+-----------------------------------------------
-----------------------------------------------------------------------
-----------------------------------------------------------------------
-----------------------------------------------------------------------
-----------------------------------------------------------------------
----------------------------------------------
 name    |  -0.467008 | {"I- 580                        Ramp","I- 880
                       Ramp","Sp Railroad                          ",
"I- 580                             ","I- 680
```

```
 Ramp","I- 80                          Ramp","14th
     St  ","5th                            St  ","Mission
         Blvd","I- 880                           "}
 thepath |         20 | {"[(-122.089,37.71),(-122.0886,37.711)]"}
(2 rows)
```

pg_stats的详细描述见第 42.43 节。

pg_statistic中存储的信息量，特别是每列most_common_vals中的最大项数和histogram_bounds数组的大小，可以使用ALTER TABLE SET STATISTICS命令按列设置，也可以通过设置配置变量default_statistics_target进行全局设置。目前默认上限是 10 项。提高这一上限可能让规划器做出更准确的估计，尤其是对于数据分布不规则的列；代价则是pg_statistic占用更多空间，并且计算估计值所需时间也会略有增加。相反，对于数据分布较简单的列，较低的上限可能已经足够。

13.3. 用显式JOIN子句控制规划器

可以在一定程度上用显式JOIN语法控制查询规划器。要理解这一点为何重要，先需要一些背景知识。

在一个简单的连接查询中，例如：

```
SELECT * FROM a, b, c WHERE a.id = b.id AND b.ref = c.id;
```

规划器可以自由地按任意顺序连接这些表。例如，它可以生成一个查询计划，先利用WHERE条件a.id = b.id将 A 连接到 B，然后再利用另一个WHERE条件把 C 连接到这个结果上。也可以先连接 B 和 C，再把 A 连接到所得结果上。甚至还可以先连接 A 和 C，再与 B 连接，但这样效率会很低，因为必须先形成 A 和 C 的完整笛卡尔积，而WHERE子句中并没有可用于优化这一连接的条件。（PostgreSQL执行器中的所有连接都发生在两个输入表之间，因此结果必须以这些形式之一逐步构造出来。）关键在于，这些不同的连接可能性在语义上是等价的，但执行代价可能相差极大。因此，规划器会探索它们，力图找出最高效的查询计划。

当查询只涉及两个或三个表时，需要考虑的连接顺序并不多。但可能的连接顺序数量会随着表数增加而呈指数增长。输入表超过十个左右之后，对所有可能性做穷举搜索实际上就不再可行，甚至六七个表也可能让规划耗时长得令人厌烦。输入表过多时，PostgreSQL规划器会从穷举搜索切换到一种遗传概率搜索，只考虑有限数量的可能性。（切换阈值由运行时参数geqo_threshold控制。）遗传搜索耗时更少，但并不一定能找到最优计划。

当查询涉及外连接时，规划器比处理普通（内）连接时拥有的自由度要小得多。例如，考虑：

```
SELECT * FROM a LEFT JOIN (b JOIN c ON (b.ref = c.id)) ON (a.id = b.
id);
```

尽管这个查询的约束表面上与前一个非常相似，但它们的语义不同，因为如果 A 中有某一行无法匹配 B 和 C 连接结果中的任何行，该行仍然必须被输出。因此这里规划器对连接顺序没有选择：它必须先连接 B 和 C，再把 A 连接到该结果上。相应地，这个查询比前一个查询需要更少的规划时间。

显式连接语法（INNER JOIN、CROSS JOIN或无修饰的JOIN）在语义上和FROM中列出输入关系是一样的，因此它不需要约束连接顺序。但可以指示PostgreSQL查询规划器将显式内JOIN当作带有连接顺序约束来处理。例如，下面三个查询在逻辑上是等价的：

```
SELECT * FROM a, b, c WHERE a.id = b.id AND b.ref = c.id;
```

```
SELECT * FROM a CROSS JOIN b CROSS JOIN c WHERE a.id = b.id AND b.ref
 = c.id;
SELECT * FROM a JOIN (b JOIN c ON (b.ref = c.id)) ON (a.id = b.id);
```

但如果告诉规划器遵循JOIN顺序，那么第二个和第三个查询在规划上会比第一个花费更少时间。对于只有三个表的连接来说，这种效果微不足道；但当表很多时，它可能非常关键。

要强制规划器遵循显式JOIN给出的连接顺序，可以将运行时参数join_collapse_limit设置为 1。（其他可能值见下文讨论。）

不必为了缩短搜索时间而完全约束连接顺序，因为可以在普通FROM列表的某一项中使用JOIN操作符。例如：

```
SELECT * FROM a CROSS JOIN b, c, d, e WHERE ...;
```

如果设置join_collapse_limit = 1，就会强制规划器先将 A 连接到 B，然后再与其他表连接，但不会进一步约束它的选择。在这个示例中，可能的连接顺序数量减少了 5 倍。

以这种方式约束规划器的搜索，是一种既可减少规划时间、
又可引导规划器生成更好查询计划的实用技巧。如果规划器默认选择了糟糕的连接顺序，
可以通过JOIN语法强制它采用更好的顺序，前提当然是确实知道哪个顺序更好。建议进行实验。

与此密切相关、同样会影响规划时间的另一个问题，是将子查询折叠进父查询。例如：

```
SELECT *
FROM x, y,
    (SELECT * FROM a, b, c WHERE something) AS ss
WHERE somethingelse;
```

这种情况可能出现在使用包含连接的视图时；该视图的SELECT规则会被插入到引用视图的位置，从而得到一个与上面非常相似的查询。通常，规划器会尝试把子查询折叠进父查询，得到：

```
SELECT * FROM x, y, a, b, c WHERE something AND somethingelse;
```

这通常会生成比单独规划子查询更好的计划。（例如，外层WHERE条件可能会先把 X 连接到 A，从而消除 A 中的大量行，也就避免了形成子查询完整逻辑输出的需要。）但与此同时，规划时间也增加了；这里把两个彼此独立的三路连接问题，替换成了一个五路连接问题。
由于可能性数量呈指数增长，这种差别可能非常大。为了避免陷入巨大的连接搜索问题，
如果折叠子查询会使父查询产生超过from_collapse_limit个FROM项，
规划器就会尝试通过停止提升子查询来规避这一点。可以通过调高或调低这个运行时参数，
在规划时间和计划质量之间做权衡。

from_collapse_limit和join_collapse_limit名称相似，
因为它们做的几乎是同一件事：一个控制规划器何时将子查询“平面化”，
另一个控制何时将显式内连接平面化。通常，要么把join_collapse_limit设置为与from_collapse_limit相同，这样显式连接与子查询的行为类似；要么把join_collapse_limit设置为 1，如果想使用显式连接控制连接顺序。不过，也可以把它们设置为不同的值，以更细致地调节规划时间与运行时间之间的平衡。

13.4. 填充一个数据库

初次填充数据库时，可能需要插入大量数据。本节给出一些使这一过程尽可能高效的建议。

13.4.1. 禁用自动提交

应关闭自动提交，只在最后提交一次。（在普通 SQL 中，这意味着开始时发出BEGIN，结束时发出COMMIT。某些客户端库可能会替调用方完成这件事，
在这种情况下需要确认它们确实会在所需的时候这样做。）如果允许每次插入都单独提交，
PostgreSQL就必须为每一行的加入执行大量额外工作。
把所有插入都放在一个事务中的另一个好处是：如果其中某一行插入失败，
那么此前插入的所有行都会被回滚，这样就不会留下部分装载的数据。

13.4.2. 使用COPY

使用COPY在一条命令中装载所有记录，而不是使用一系列INSERT命令。COPY
命令针对装载大量行做了优化；它不如INSERT灵活，但在大规模数据装载时开销显著更小。由于COPY是一条单独的命令，因此采用这种方法填充表时无须关闭自动提交。

如果不能使用COPY，那么用PREPARE创建一个预备INSERT语句，再按需多次执行EXECUTE
也会有所帮助。这样可以避免重复解析和规划INSERT的开销。

请注意，在装载大量行时，使用COPY几乎总是比使用INSERT更快，即使已经使用了PREPARE，
并把多次插入批量放入同一个事务中也是如此。

13.4.3. 移除索引

如果正在装载一个新创建的表，最快的方法是先创建表，用COPY批量装载数据，
然后再创建该表所需的索引。在已有数据的表上创建索引，
要比在每行装载时对索引做增量更新更快。

如果正在向现有表加入大量数据，那么删除索引、装载数据、再重建索引可能是更好的方案。
当然，在索引缺失期间，其他数据库用户的性能可能会下降。删除唯一索引之前也必须慎重，
因为唯一约束提供的错误检查会在索引缺失期间丧失。

13.4.4. 移除外键约束

与索引类似，“批量”检查外键约束比逐行检查更高效。因此，先删除外键约束、装载数据、
再重建约束可能很有用。同样，
这里也需要在装载速度与约束缺失期间失去错误检查之间做权衡。

13.4.5. 增加maintenance_work_mem

在装载大量数据时，临时增大maintenance_work_mem配置变量可以提升性能。
这个参数也有助于加速CREATE INDEX和ALTER TABLE ADD FOREIGN KEY命令。它对COPY
本身帮助不大，因此这个建议只有在采用前述一种或两种技巧时才有意义。

13.4.6. 增加checkpoint_segments

临时增大checkpoint_segments配置变量，也可以让大规模数据装载更快。这是因为向
PostgreSQL中装载大量数据，会导致检查点比平常更频繁地发生，而正常频率由checkpoint_timeout配置变量指定。每次发生检查点时，所有脏页都必须刷写到磁盘。
通过在批量装载期间临时增大checkpoint_segments，可以减少所需的检查点次数。

13.4.7. 事后运行ANALYZE

每当显著改变了表中数据的分布，都强烈建议运行ANALYZE。
这也包括向表中批量装载大量数据。运行ANALYZE（或VACUUM ANALYZE）可以确保规划器掌握

该表的最新统计信息。如果没有统计信息，或者统计信息已经过时，
规划器在生成查询计划时就可能做出糟糕决定，从而导致相关表性能不佳。

13.4.8. 关于pg_dump的一些注记

由pg_dump生成的转储脚本会自动应用上面若干条指导原则，但并非全部。
若要尽可能快速地还原pg_dump的转储，仍需手动做一些额外操作。（注意，这些要点适用于还原转储，而不是创建转储。在使用pg_restore从pg_dump归档文件加载时，
相关要点是一样的。）

默认情况下，pg_dump使用`COPY`；而当它生成完整的模式加数据转储时，
也会小心地先装载数据，再创建索引和外键。因此在这种情况下，
上述若干指导原则已经被自动处理。剩下需要做的，只是在载入转储脚本之前为`maintenance_work_mem`和`checkpoint_segments`设置适当的（即比正常值大的）值，然后在完成后运行`ANALYZE`。

仅包含数据的转储仍然会使用`COPY`，但它不会删除或重建索引，通常也不会处理外键。[2]因此，
在装载纯数据转储时，如果想采用这些技术，就需要自行负责删除并重建索引与外键。
装载数据期间增大`checkpoint_segments`仍然有益，但没有必要同时增大`maintenance_work_mem`；后者更适合留到之后手工重建索引和外键时再调大。
完成后也别忘了执行`ANALYZE`。

[2] 可以通过使用`-X disable-triggers`选项达到禁用外键的效果 – 但要注意，这样做是取消外键验证，而不仅仅是推迟它。因此如果使用该选项，就有可能插入坏数据。

部分 III. 服务器管理

本部分涵盖PostgreSQL管理员关心的主题，包括软件安装、服务器配置、用户和数据库管理以及维护任务。任何运行PostgreSQL服务器的人，即使仅供个人使用，尤其是在生产环境中，都应熟悉这些主题。

本部分的内容大体按照新用户应当阅读的顺序来安排。各章自成体系，也可按需单独阅读。
相关内容按主题单元以叙述方式展开。希望获得某个命令完整描述的读者，建议查阅第 VI 部分。

开头几章写得无需预备知识即可理解，因此需要自行搭建服务器的新用户可以从这里开始探索。
本部分其余内容涉及调优和管理；这些材料假定读者已经熟悉PostgreSQL数据库系统的一般用法。
建议读者再参阅第 I 部分和第 II 部分部分，以获取更多信息。

目录

14. 安装说明 226
14.1. 简要说明 226
14.2. 需求 226
14.3. 获取源代码 228
14.4. 如果你要升级 228
14.5. 安装过程 229
14.6. 安装后设置 234
14.6.1. 共享库 234
14.6.2. 环境变量 235
14.7. 受支持的平台 236
15. Windows上仅安装客户端 241
16. 操作系统环境 243
16.1. The PostgreSQL User Account 243
16.2. 创建一个数据库集簇 243
16.3. 启动数据库服务器 244
16.3.1. 服务器启动失败 245
16.3.2. 客户端连接问题 246
16.4. 管理内核资源 247
16.4.1. 共享内存和信号量 247
16.4.2. Resource Limits 251
16.4.3. Linux 内存过量分配 252
16.5. 关闭服务器 253
16.6. 加密选项 253
16.7. 使用 SSL 的安全 TCP/IP 连接 254
16.8. Secure TCP/IP Connections with SSH Tunnels 255
17. 服务器配置 257
17.1. 设置参数 257
17.2. 文件位置 258
17.3. 连接和认证 259
17.3.1. 连接设置 259
17.3.2. 安全和认证 260
17.4. 资源消耗 261
17.4.1. 内存 261
17.4.2. 空闲空间映射 262
17.4.3. 内核资源使用 263
17.4.4. 基于代价的清理延迟 263
17.4.5. 后台写入器 264
17.5. 预写式日志 265
17.5.1. 设置 265
17.5.2. 检查点 267
17.5.3. 归档 267
17.6. 查询规划 267
17.6.1. 规划器方法配置 267
17.6.2. 规划器代价常量 268
17.6.3. 遗传查询优化器 269
17.6.4. 其他规划器选项 270
17.7. 错误报告和日志 271
17.7.1. 日志记录到哪里 271
17.7.2. 什么时候记录日志 272
17.7.3. 记录哪些内容 274
17.8. 运行时统计 275

17.8.1. 统计监控 275
17.8.2. 查询和索引统计信息收集器 276
17.9. 自动清理 276
17.10. 客户端连接默认值 277
17.10.1. 语句行为 277
17.10.2. 区域设置和格式化 278
17.10.3. 其他默认值 279
17.11. 锁管理 280
17.12. 版本和平台兼容性 280
17.12.1. 以前的 PostgreSQL 版本 280
17.12.2. 平台和客户端兼容性 281
17.13. 预置选项 282
17.14. 自定义选项 283
17.15. 开发者选项 283
17.16. 短选项 284
18. 数据库角色和权限 286
18.1. 数据库角色 286
18.2. 角色属性 287
18.3. 权限 288
18.4. 角色成员资格 288
18.5. 函数和触发器 290
19. 管理数据库 291
19.1. 概述 291
19.2. 创建数据库 291
19.3. 模板数据库 292
19.4. 数据库配置 293
19.5. 删除数据库 293
19.6. 表空间 294
20. 客户端认证 296
20.1. `pg_hba.conf` 文件 296
20.2. 认证方法 301
20.2.1. trust 认证 301
20.2.2. 密码认证 301
20.2.3. Kerberos 认证 301
20.2.4. 基于 ident 的认证 302
20.2.5. PAM 认证 304
20.3. 认证问题 304
21. Localization 305
21.1. Locale Support 305
21.1.1. Overview 305
21.1.2. Behavior 306
21.1.3. Problems 306
21.2. Character Set Support 307
21.2.1. 支持的字符集 307
21.2.2. 设置字符集 309
21.2.3. 服务器和客户端之间的自动字符集转换 310
21.2.4. 进一步阅读 312
22. 日常数据库维护任务 313
22.1. 日常清理 313
22.1.1. 回收磁盘空间 313
22.1.2. 更新规划器统计信息 314
22.1.3. 防止事务 ID 回卷失败 315
22.1.4. 自动清理守护进程 316

22.2. 日常重建索引 317
22.3. 日志文件维护 318
23. 备份和恢复 319
23.1. SQL转储 319
23.1.1. 恢复转储 319
23.1.2. 使用pg_dumpall 320
23.1.3. 处理大型数据库 320
23.2. 文件系统级备份 321
23.3. 在线备份和时间点恢复（PITR） 322
23.3.1. 设置 WAL 归档 322
23.3.2. 进行基础备份 324
23.3.3. 使用在线备份进行恢复 325
23.3.4. 时间线 327
23.3.5. 注意事项 328
23.4. 版本间的迁移 328
24. 监控数据库活动 330
24.1. 标准 Unix 工具 330
24.2. 统计收集器 331
24.2.1. 统计信息收集配置 331
24.2.2. 查看收集到的统计信息 331
24.3. 查看锁 335
25. 监控磁盘使用情况 336
25.1. 确定磁盘使用情况 336
25.2. 磁盘已满故障 337
26. 可靠性与预写式日志 338
26.1. 可靠性 338
26.2. 预写式日志（WAL） 338
26.3. WAL 配置 339
26.4. WAL 内部机制 340
27. 回归测试 341
27.1. 运行测试 341
27.2. 测试结果评估 342
27.2.1. 错误消息差异 343
27.2.2. 区域差异 343
27.2.3. 日期和时间差异 343
27.2.4. 浮点差异 343
27.2.5. 行顺序差异 343
27.2.6. 栈深度不足 343
27.2.7. “random” 测试 344
27.3. 变体比较文件 344

第 14 章 安装说明

本章介绍如何从源代码发布包安装PostgreSQL。（如果你安装的是预打包的发行版，例如 RPM 或 Debian 软件包，请忽略本章，改为阅读打包者提供的说明。）

14.1. 简要说明

```
./configure
gmake
su
gmake install
adduser postgres
mkdir /usr/local/pgsql/data
chown postgres /usr/local/pgsql/data
su - postgres
/usr/local/pgsql/bin/initdb -D /usr/local/pgsql/data
/usr/local/pgsql/bin/postmaster -D /usr/local/pgsql/data >logfile 2>&1
 &
/usr/local/pgsql/bin/createdb test
/usr/local/pgsql/bin/psql test
```

长版本就是本章的其余部分。

14.2. 需求

一般来说，现代的 Unix 兼容平台应该能够运行 PostgreSQL。发布时经过具体测试的平台列于下方第 14.7 节。发行包的 `doc` 子目录中有若干平台特定的 FAQ 文档，遇到问题时可以查阅。

构建PostgreSQL需要下列软件包：

- 需要 GNU make；其他make程序不能用。GNU make通常安装在`gmake`名下；本文档一律用该名称称呼它。（在某些系统上GNU make就是默认的`make`工具。）要测试是否为GNU make，可输入

  `gmake --version`

  建议使用 3.79.1 或更高版本。

- 需要一个ISO/ANSI C 编译器。推荐使用较新版本的GCC，但已知PostgreSQL可以用各厂商的多种编译器构建。

- 首先需要tar来解开源码发布包，此外还需要gzip或bzip2之一。

- GNU Readline库（用于舒适的行编辑和命令历史检索）默认会被使用。如果不想使用它，必须为`configure`指定`--without-readline`选项。（在NetBSD上，`libedit` 库与Readline兼容，在找不到`libreadline`时会使用它。）如果你使用基于软件包的 Linux 发行版，注意如果你的发行版把二者分开的话，需要同时安装`readline`和`readline-devel`两个软件包。

- 默认使用zlib压缩库。如果不想使用它，必须为`configure`指定`--without-zlib`选项。使用该选项会禁用pg_dump和pg_restore对压缩归档的支持。

- 在Windows上构建PostgreSQL还需要额外的软件。你可以使用MinGW为基于NT的Windows版本（如 Windows XP 和 2003）构建PostgreSQL；详情见doc/FAQ_MINGW。也可以使用Cygwin构建PostgreSQL；见doc/FAQ_CYGWIN。
  基于Cygwin的构建可以在较老版本的Windows上工作，但如果有选择，我们推荐MinGW方法。虽然这些是仅有的被推荐用于完整构建的工具集，但也可以使用其他Windows工具集只构建 C 客户端库（libpq）和交互式终端（psql）。相关详情见第 15 章。

下列软件包是可选的。默认配置不需要它们，但在启用某些构建选项时会用到，如下所述。

- 要构建服务器编程语言PL/Perl，需要一个完整的Perl安装，包括libperl库和头文件。
  由于PL/Perl是一个共享库，libperl库在大多数平台上也必须是共享库。
  较新版本的Perl似乎默认如此，但较早版本不是，而且这无论如何都取决于在你的站点上安装Perl 的人所作的选择。如果你打算较多地使用PL/Perl，应确保Perl安装时启用了usemultiplicity选项（perl -V会显示是否如此）。

  如果你没有共享库但需要一个，构建期间会出现类似下面这样的消息来指出这一事实：

  ```
  *** Cannot build PL/Perl because libperl is not a shared library.
  *** You might have to rebuild your Perl installation.  Refer to
  *** the documentation for details.
  ```

  （如果你没有留意屏幕输出，只会发现PL/Perl库对象plperl.so或类似文件未被安装。）
  看到这条消息时，你必须手工重新构建并安装Perl，才能构建PL/Perl。在Perl的配置过程中，请选择生成共享库。

- 要构建PL/Python服务器编程语言，需要一个带有头文件和distutils模块的Python安装。distutils模块默认包含在 1.6 及之后的Python版本中；使用更早版本Python的用户需要自行安装它。

  由于PL/Python是一个共享库，libpython库在大多数平台上也必须是共享库。
  默认的Python安装并非如此。如果构建并安装之后存在一个名为plpython.so的文件（扩展名可能不同），说明一切正常。
  否则你应当看到过一条类似下面一闪而过的通知：

  ```
  *** Cannot build PL/Python because libpython is not a shared
   library.
  *** You might have to rebuild your Python installation.  Refer to
  *** the documentation for details.
  ```

  这意味着你必须重新构建（部分）Python安装来提供这个共享库。

  如果遇到问题，可为Python 2.3 或更高版本的 configure 加上--enable-shared标志。在某些操作系统上无须构建共享库，但你得让PostgreSQL的构建系统相信这一点。
  细节请查阅src/pl/plpython目录中的Makefile。

- 要构建PL/Tcl过程语言，当然需要一个 Tcl 安装。

- 要启用本地语言支持（NLS），即以英语之外的语言显示程序消息的能力，需要一个Gettext API的实现。某些操作系统内置了它（如Linux、NetBSD、Solaris），其他系统可以从http://developer.postgresql.org/~petere/bsd-gettext/下载附加软件包。
  如果你使用的是GNU C 库中的Gettext实现，还需要GNU Gettext软件包提供的一些实用程序。对其他实现则不需要。

- 如果要支持使用这些服务的认证或加密，需要Kerberos、OpenSSL和/或PAM。

如果你要从Git树而不是正式源码包构建，或者要做服务器开发，还需要下列软件包：

- 从 Git 检出构建时需要 GNU Flex和Bison，修改了实际的扫描器和解析器定义文件时也需要。如果需要它们，务必使用Flex 2.5.4 或更高版本以及Bison 1.875 或更高版本。其他yacc程序有时也能使用，但这样做需要额外的努力，不予推荐。其他lex程序则肯定不能用。

如果需要获取 GNU 软件包，可以从本地的 GNU 镜像站点获取（站点列表见http://www.gnu.org/order/ftp.html），或者从ftp://ftp.gnu.org/gnu/获取。

还要检查磁盘空间是否充足。编译期间，源码树需要约 65 MB，安装目录需要约 15 MB。空数据库集簇约占 25 MB；数据库所需空间大约是存储相同数据的纯文本文件的五倍。如果要运行回归测试，还会临时额外需要最多 90 MB。使用 `df` 命令检查可用磁盘空间。

14.3. 获取源代码

PostgreSQL 8.1.23 的源码可从 anonymous FTP from ftp://ftp.postgresql.org/pub/source/v8.1.23/postgresql-8.1.23.tar.gz. 其他下载选项见我们的网站：http://www.postgresql.org/download/。获取文件后，解开源码包：

`gunzip postgresql-8.1.23.tar.gztar xf postgresql-8.1.23.tar`

这会创建目录 `postgresql-8.1.23`（位于当前目录之下），其中是PostgreSQL的源码。在其余步骤之前请切换到该目录以继续安装流程。

14.4. 如果你要升级

PostgreSQL的内部数据存储格式随新版本而变化。因此，
如果你要升级的现有安装的版本号不是“8.1.x”，就必须按这里所述备份并恢复数据。
这些说明假定你现有的安装在`/usr/local/pgsql`目录下，数据区域在`/usr/local/pgsql/data`中。请酌情替换为你实际的路径。

1. 确保数据库在备份期间和备份之后不被更新。这不影响备份的完整性，
   但已更改的数据当然不会包含在内。必要时，可以修改文件`/usr/local/pgsql/data/pg_hba.conf`（或等价文件）中的权限，禁止除你之外的所有人访问。

2. 要备份数据库安装，输入：

   `pg_dumpall > outputfile`

   如果需要保留 OID（例如用作外键），运行pg_dumpall时请使用`-o`选项。

   做备份时，可以使用当前运行版本的pg_dumpall命令。不过，为获得最佳结果，
   请尽量使用PostgreSQL 8.1.23 的pg_dumpall命令，因为该版本包含对旧版本的缺陷修复和改进。虽然在你尚未安装新版本时这条建议看似不合常理，
   但如果打算让新旧两个版本并行安装，遵循它是可取的。
   这种情况下你可以先正常完成安装，之后再传输数据。这也能缩短停机时间。

3. 如果你把新版本安装在和老版本相同的位置，那么最迟在安装新文件之前，
   应关闭旧服务器：

   `pg_ctl stop`

   在开机时自动启动PostgreSQL的系统上，很可能有启动文件可以完成同样的事情。例如，在Red Hat Linux系统上可以发现

   `/etc/rc.d/init.d/postgresql stop`

   有效。

非常老的版本可能没有pg_ctl。如果找不到它或者它不能工作，可以找出老服务器的进程ID，例如输入

ps ax | grep postmaster

然后这样通知它停止：

kill -INT *processID*

4. 如果你安装在和老版本相同的位置，把旧安装移开也是好主意，以防遇到问题需要回退。可以使用类似下面的命令：

   mv /usr/local/pgsql /usr/local/pgsql.old

安装PostgreSQL 8.1.23 之后，创建新的数据库目录并启动新服务器。记住，这些命令必须以特殊的数据库用户账户登录执行（如果你是在升级，该账户已经存在）。

/usr/local/pgsql/bin/initdb -D /usr/local/pgsql/data/usr/local/pgsql/bin/postmaster -D /usr/local/pgsql/data

最后，用

/usr/local/pgsql/bin/psql -d postgres -f *outputfile*

恢复数据，使用新的psql。

更多讨论见第 23.4 节，无论如何都建议你阅读。

14.5. 安装过程

1. **Configuration**

   安装流程的第一步是为你的系统配置源码树并选择所需选项。这通过运行`configure`脚本来完成。默认安装只需输入

   ./configure

   该脚本会运行一系列测试来猜测各种系统相关变量的值并检测操作系统的某些特性，最后会在构建树中创建若干文件记录其发现。（如果希望把构建目录与源码树分开，也可以在源码树之外的目录中运行`configure`。）

   默认配置将构建服务器和实用程序，以及所有只需要 C 编译器的客户端应用程序和接口。默认情况下，所有文件都会安装到 `/usr/local/pgsql` 之下。

   可以通过向`configure`提供下列一个或多个命令行选项来自定义构建和安装过程：

   `--prefix=`*`PREFIX`*

   把所有文件安装在目录*`PREFIX`*而不是`/usr/local/pgsql`之下。实际文件会安装到各个子目录中；任何文件都不会直接安装到*`PREFIX`*目录本身。

   如有特殊需要，还可以用下列选项自定义各个子目录。不过，如果保持默认值，安装将是可重定位的，即安装之后可以移动该目录。（`man`和`doc`的位置不受此影响。）

   对可重定位安装，你可能想使用`configure`的`--disable-rpath`选项。此外，你还需要告诉操作系统如何找到共享库。

`--exec-prefix=EXEC-PREFIX`

可以把体系结构相关的文件安装在不同于*PREFIX*所设置的另一个前缀*EXEC-PREFIX*之下。这在多台主机共享体系结构无关文件时很有用。如果省略此选项，*EXEC-PREFIX*将被设置为等于*PREFIX*，体系结构相关和无关键将安装到同一个目录树下，这多半正是你想要的。

`--bindir=DIRECTORY`

指定可执行程序所在的目录。默认值为*EXEC-PREFIX*/bin，通常即/usr/local/pgsql/bin。

`--datadir=DIRECTORY`

已安装程序使用的只读数据文件所在的目录。默认值为*PREFIX*/share。注意这与数据库文件放在哪里无关。

`--sysconfdir=DIRECTORY`

各种配置文件所在的目录，默认为*PREFIX*/etc。

`--libdir=DIRECTORY`

安装库和动态可装载模块的位置。默认值为*EXEC-PREFIX*/lib。

`--includedir=DIRECTORY`

安装 C 和 C++ 头文件的目录。默认值为*PREFIX*/include。

`--mandir=DIRECTORY`

PostgreSQL附带的手册页将安装在该目录下各自的man*x*子目录中。默认值为*PREFIX*/man。

`--with-docdir=DIRECTORY`
`--without-docdir`

除“man”页之外的文档文件将安装到该目录。默认值为*PREFIX*/doc。如果指定了--without-docdir选项，make install将不会安装文档。这适用于对文档安装有特殊方法的打包脚本。

> **注意**
>
> 为了能把 PostgreSQL 安装到共享安装位置（例如 /usr/local/include），同时又不干扰系统其他部分的名字空间，我们做了特别处理。首先，除非完整展开后的目录名已经包含字符串“postgres”或“pgsql”，否则会自动把字符串“/postgresql”追加到 datadir、sysconfdir 和 docdir 上。例如，如果你选择 /usr/local 作为前缀，那么文档会安装到 /usr/local/doc/postgresql；但如果前缀是 /opt/postgres，那么它会安装到 /opt/postgres/doc。客户端接口的公共 C 头文件会安装到 includedir 中，并且不会污染名字空间。内部头文件和服务器头文件则会安装到 includedir

下的私有目录中。关于如何访问这些头文件，请参见各接口自己的文档。最后，如果有需要，也会在 `libdir` 下创建私有子目录，用于存放动态可加载模块。

`--with-includes=`*`DIRECTORIES`*

*`DIRECTORIES`*是一个冒号分隔的目录列表，这些目录将被加入编译器搜索头文件的路径列表。如果把可选软件包（如 GNU Readline）安装在非标准位置，就必须使用此选项，并且很可能还要使用相应的`--with-libraries`选项。

例如：`--with-includes=/opt/gnu/include:/usr/sup/include`。

`--with-libraries=`*`DIRECTORIES`*

*`DIRECTORIES`*是一个冒号分隔的搜索库的目录列表。如果软件包安装在非标准位置，你很可能需要使用此选项（以及相应的`--with-includes`选项）。

例如：`--with-libraries=/opt/gnu/lib:/usr/sup/lib`。

`--enable-nls[=`*`LANGUAGES`*`]`

启用本地语言支持（NLS），即以英语之外的语言显示程序消息的能力。*`LANGUAGES`*是一个可选的空格分隔的语言代码列表，指定你希望支持的语言，例如`--enable-nls='de fr'`。（该列表与实际提供的翻译集合的交集会自动计算。）如果不指定列表，则安装所有可用的翻译。

要使用此选项，需要一个 Gettext API 实现；参见上文。

`--with-pgport=`*`NUMBER`*

将 *`NUMBER`* 设为服务器和客户端的默认端口号。默认值是 5432。端口号以后始终都可以修改，但如果在这里指定，那么服务器和客户端都会编译进同一个默认值，这可能很方便。通常选择非默认值的唯一合理原因，是你打算在同一台机器上运行多个 PostgreSQL 服务器。

`--with-perl`

构建 PL/Perl 服务器端语言。

`--with-python`

构建 PL/Python 服务器端语言。

`--with-tcl`

构建 PL/Tcl 服务器端语言。

`--with-tclconfig=`*`DIRECTORY`*

Tcl 会安装文件 `tclConfig.sh`，其中包含构建与 Tcl 交互的模块所需的配置信息。通常会在已知位置自动找到此文件，但如果要使用其他版本的 Tcl，可以指定查找它的目录。

`--with-krb5`

构建时支持 Kerberos 5 认证。在许多系统上，Kerberos 系统并不安装在默认搜索的位置（如 `/usr/include`、`/usr/lib`），因此除了此选项，还必须使用 `--with-includes` 和 `--with-libraries`。`configure` 会检查所需的头文件和库，确保 Kerberos 安装满足要求，然后才会继续。

`--with-krb-srvnam=`*`NAME`*

Kerberos 服务主体的默认名称。默认值为 `postgres`。通常没有理由更改它。

`--with-openssl`

构建时支持 SSL（加密）连接。这要求已安装 OpenSSL 软件包。`configure` 会检查所需的头文件和库，确保 OpenSSL 安装满足要求，然后才会继续。

`--with-pam`

构建时支持 PAM（可插拔认证模块）。

`--without-readline`

禁止使用Readline库。这会禁用 psql 中的命令行编辑和历史记录，因此不建议使用。

`--with-bonjour`

构建时支持 Bonjour。这要求操作系统支持 Bonjour。建议在 Mac OS X 上启用。

`--enable-integer-datetimes`

对日期时间和时间间隔使用 64 位整数存储，而不是默认的浮点数存储。这缩小了可表示值的范围，但保证了整个范围内的微秒精度（更多信息参见第 8.5 节）。另请注意，整数日期时间代码比浮点代码新，我们时至今日仍会在其中发现缺陷。

`--disable-spinlocks`

即使 PostgreSQL 不支持该平台的 CPU 自旋锁，也允许构建成功。缺少自旋锁支持会导致性能不佳；因此，仅应在构建中止并告知平台缺少自旋锁支持时使用此选项。
如果在你的平台上构建 PostgreSQL 必须使用此选项，请向 PostgreSQL 开发者报告该问题。

`--enable-thread-safety`

使客户端库成为线程安全的。这允许 libpq 和 ECPG 程序中的并发线程安全地控制各自私有的连接句柄。此选项要求操作系统提供足够的线程支持。

`--without-zlib`

禁止使用 Zlib 库。这会禁用 pg_dump 和 pg_restore 对压缩归档的支持。此选项仅用于没有该库的罕见系统。

`--enable-debug`

将所有程序和库编译为带调试符号的版本。这意味着你可以在调试器中运行程序，
以分析问题。这会显著增大安装后的可执行文件大小，而且在非 GCC 编译器上，通常还会禁用编译器优化，从而导致变慢。不过，

保留这些符号对于处理可能出现的各种问题极其有帮助。目前，只有在你使用 GCC 的情况下，才建议在生产安装中使用该选项。但如果你在做开发工作或运行测试版，就应始终启用它。

`--enable-cassert`

在服务器中启用断言检查，用于测试许多“不可能发生”的条件。这对代码开发非常有价值，但这些测试多少会拖慢速度。此外，启用这些测试并不一定会增强服务器稳定性！断言检查并未按严重程度分类，因此即使某个 bug 相对无害，只要触发了断言失败，仍会导致服务器重启。该选项不建议用于生产环境，但如果你在做开发工作或运行测试版，就应当启用它。

`--enable-depend`

启用自动依赖跟踪。启用后，makefile 会在任何头文件被修改时，重新构建所有受影响的目标文件。如果你在做开发工作，这很有用；但如果你只是打算编译一次并安装，这只是额外的开销。目前该选项只在 GCC 下有效。

如果你希望使用不同于 `configure` 所选的 C 编译器，可以将环境变量 `CC` 设为所选程序。默认情况下，`configure` 会在 `gcc` 可用时选择它，否则选择平台默认编译器（通常为 `cc`）。类似地，如有需要，可以使用 `CFLAGS` 变量覆盖默认编译器标志。

可以在 `configure` 命令行上指定环境变量，例如：

```
./configure CC=/opt/bin/gcc CFLAGS='-O2 -pipe'
```

2. **构建**

   要开始构建，输入

   ```
   gmake
   ```

   （记住要使用 GNU make。）构建过程可能需要 5 分钟到半小时不等，具体取决于你的硬件。最后显示的一行应该是

   ```
   All of PostgreSQL is successfully made. Ready to install.
   ```

3. **回归测试**

   如果你想在安装之前测试新构建的服务器，可以在此时运行回归测试。回归测试是一套测试用例，用于验证 PostgreSQL 在你的机器上按开发者预期的方式运行。输入

   ```
   gmake check
   ```

   （以 root 身份无法运行；请以非特权用户身份执行。）第 27 章包含关于解释测试结果的详细信息。你可以在之后任何时候通过执行相同的命令来重复此测试。

4. **安装文件**

注意

如果你正在升级现有系统，并打算把新文件安装在旧文件之上，请务必按上文第 14.4 节所述，先备份数据并关闭旧服务器，然后再继续。

要安装 PostgreSQL，输入

gmake install

这会把文件安装到步骤 1中指定的目录中。请确保你有写入该区域的适当权限。通常你需要以 root 身份执行这一步。或者，你也可以提前创建目标目录，并安排授予适当的权限。

你可以用 `gmake install-strip` 代替 `gmake install`，以便在安装时剥离可执行文件和库中的符号。这会节省一些空间。如果你构建时启用了调试支持，剥离操作会实际移除调试支持，因此只应在不再需要调试时进行。`install-strip` 会尽力节省空间，但它并不能完全掌握如何从可执行文件中剥离每一个多余的字节，所以如果想尽可能节省所有磁盘空间，你还得做一些手工工作。

标准安装会提供客户端应用开发以及服务器端程序开发（如用 C 编写的自定义函数或数据类型）所需的全部头文件。（在 PostgreSQL 8.0 之前，后者需要单独的 `gmake install-all-headers` 命令，但现在这一步已并入标准安装。）

仅客户端安装：. 如果你只想安装客户端应用和接口库，可以使用这些命令：

gmake -C src/bin installgmake -C src/include installgmake -C src/interfaces installgmake -C doc install

在 Windows 上注册 eventlog：. 要在操作系统中注册 Windows eventlog 库，请在安装后执行以下命令：

regsvr32 *pgsql_library_directory*/pgevent.dll

这会创建事件查看器使用的注册表项。

卸载：. 要撤销安装，可使用命令 `gmake uninstall`。但是，这不会删除任何已创建的目录。

清理：. 安装完成后，你可以用命令 `gmake clean`从源码树中删除构建的文件，以腾出空间。这会保留 `configure` 程序生成的文件，以便你之后能用 `gmake` 重新构建所有内容。要把源码树重置为发布时的状态，可使用 `gmake distclean`。如果你要在同一个源码树中为多个平台构建，就必须这样做并为每次构建重新配置。（或者，为每个平台使用单独的构建树，使源码树保持未修改状态。）

如果你执行了构建，之后发现 `configure` 选项有误，或者你更改了 `configure` 所考察的任何内容（例如软件升级），那么最好在重新配置和重新构建之前先执行 `gmake distclean`。否则，你在配置选择上的更改可能无法传播到所有需要的地方。

14.6. 安装后设置

14.6.1. 共享库

在一些有共享库的系统上（大多数系统都有），你需要告诉系统如何找到新安装的共享库。不需要这样做的系统包括 BSD/OS、FreeBSD、HP-UX、IRIX、Linux、NetBSD、OpenBSD、Tru64 UNIX（原 Digital UNIX）和 Solaris。

设置共享库搜索路径的方法因平台而异，但最常见的方法是设置环境变量 `LD_LIBRARY_PATH`，方法如下。在 Bourne shell（`sh`，`ksh`，`bash`，`zsh`）：

```
LD_LIBRARY_PATH=/usr/local/pgsql/lib
export LD_LIBRARY_PATH
```

或者在 csh 或 tcsh：

```
setenv LD_LIBRARY_PATH /usr/local/pgsql/lib
```

请将 /usr/local/pgsql/lib 替换为你为 --libdir 指定的值，设置过程见步骤 1。应将这些命令放入 shell 启动文件，例如 /etc/profile 或 ~/.bash_profile。关于此方法相关注意事项的一些有用信息，参见http://xahlee.org/UnixResource_dir/_/ldpath.html。

在某些系统上，更好的做法可能是在构建之前设置环境变量 LD_RUN_PATH。

在 Cygwin 上，请把库目录加入 PATH，或者把 .dll 文件移到 bin 目录中。

如果不确定，请参考你系统的手册页（可能是 ld.so 或 rld）。如果你随后收到类似下面这样的消息

```
psql: error in loading shared libraries
libpq.so.2.1: cannot open shared object file: No such file or
 directory
```

那就说明这一步确实是必需的。到那时再处理即可。

如果你使用的是 BSD/OS、Linux 或 SunOS 4，并且拥有 root 权限，可以在安装后运行：

```
/sbin/ldconfig /usr/local/pgsql/lib
```

（或等价目录），使运行时链接器能更快地找到共享库。更多信息参见 ldconfig 的手册页。在 FreeBSD、NetBSD 和 OpenBSD 上，命令则是：

```
/sbin/ldconfig -m /usr/local/pgsql/lib
```

其他系统据知没有等价的命令。

14.6.2. 环境变量

如果你安装到了 /usr/local/pgsql 或其他默认不搜索程序的位置，就应该把 /usr/local/pgsql/bin（或你在步骤 1中为 --bindir 设置的值）加入 PATH。严格来说，这不是必需的，但这会让使用 PostgreSQL 方便得多。

为此，请把下面内容加到你的 shell 启动文件中，例如 ~/.bash_profile（如果你希望影响所有用户，则用 /etc/profile）：

```
PATH=/usr/local/pgsql/bin:$PATH
export PATH
```

如果你使用的是 csh 或 tcsh，则使用这条命令：

```
set path = ( /usr/local/pgsql/bin $path )
```

要让系统能找到 man 文档，你需要把如下所示的行加入 shell 启动文件，除非你安装到了默认会搜索的位置：

```
MANPATH=/usr/local/pgsql/man:$MANPATH
export MANPATH
```

环境变量 PGHOST 和 PGPORT 会告诉客户端应用数据库服务器的主机和端口，从而覆盖编译时内置的默认值。如果你准备从远端运行客户端应用，那么让每个打算使用数据库的用户都设置 PGHOST 会比较方便。不过这不是必需的；对于大多数客户端程序，也可以通过命令行选项传递这些设置。

14.7. 受支持的平台

PostgreSQL已被开发者社区验证可以在下面列出的平台上工作。
受支持的平台通常意味着PostgreSQL能按照本章说明构建和安装，
并且回归测试通过。“构建农场”条目指 PostgreSQL 构建农场[1]中处于活动状态的测试机器。显示较旧 PostgreSQL 版本的平台条目，是在 8.1 版本发布时未经过显式测试、但我们仍然预期可以工作的平台。

> **注意**
>
> 如果你在受支持的平台上安装时遇到问题，请写信给 `<pgsql-bugs@postgresql.org>`，而不是写信给这里列出的人。

OS	处理器	版本	报告	备注
AIX	PowerPC	8.1.0	构建农场 kookaburra (5.2, cc 6.0); asp (5.2, gcc 3.3.2)	see `doc/FAQ_AIX`, particularly if using AIX 5.3 ML3
AIX	RS6000	8.0.0	Hans-Jürgen Schönig (`<hs@cybertec.at>`), 2004-12-06	see `doc/FAQ_AIX`
BSD/OS	x86	8.1.0	Bruce Momjian (`<pgman@candle.pha.pa.us>`), 2005-10-26	4.3.1
Debian GNU/Linux	Alpha	8.1.0	构建农场 hare (3.1, gcc 3.3.4)	
Debian GNU/Linux	AMD64	8.1.0	构建农场 panda (sid, gcc 3.3.5)	
Debian GNU/Linux	ARM	8.1.0	构建农场 penguin (3.1, gcc 3.3.4)	
Debian GNU/Linux	Athlon XP	8.1.0	构建农场 rook (3.1, gcc 3.3.5)	
Debian GNU/Linux	IA64	7.4	Noèl Köthe (`<noel@debian.org>`), 2003-10-25	
Debian GNU/Linux	m68k	8.0.0	Noèl Köthe (`<noel@debian.org>`), 2004-12-09	sid
Debian GNU/Linux	MIPS	8.1.0	构建农场 otter (3.1, gcc 3.3.4)	
Debian GNU/Linux	MIPSEL	8.1.0	构建农场 lionfish (3.1, gcc 3.3.4); corgi (3.1, gcc 3.3.4)	

[1] http://www.pgbuildfarm.org/

OS	处理器	版本	报告	备注
Debian GNU/Linux	PA-RISC	8.1.0	构建农场 kingfisher (3.1, gcc 3.3.5)	
Debian GNU/Linux	PowerPC	8.0.0	Noèl Köthe (<noel@debian.org>), 2004-12-15	sid
Debian GNU/Linux	S/390	7.4	Noèl Köthe (<noel@debian.org>), 2003-10-25	
Debian GNU/Linux	Sparc	8.1.0	构建农场 dormouse (3.1, gcc 3.2.5; 64-bit)	
Debian GNU/Linux	x86	8.0.0	Peter Eisentraut (<peter_e@gmx.net>), 2004-12-06	3.1 (sarge), kernel 2.6
Fedora	AMD64	8.1.0	构建农场 viper (FC3, gcc 3.4.2)	
Fedora	x86	8.1.0	构建农场 thrush (FC1, gcc 3.3.2)	
FreeBSD	Alpha	7.4	Peter Eisentraut (<peter_e@gmx.net>), 2003-10-25	4.8
FreeBSD	AMD64	8.1.0	构建农场 platypus (5.2.1, gcc 3.3.3); dove (5.4, gcc 3.4.2)	
FreeBSD	x86	8.1.0	构建农场 octopus (4.11, gcc 2.95.4); flatworm (5.3, gcc 3.4.2); echidna (6, gcc 3.4.2); herring (6, Intel cc 7.1)	
Gentoo Linux	AMD64	8.1.0	构建农场 caribou (2.6.9, gcc 3.3.5)	
Gentoo Linux	IA64	8.1.0	构建农场 stoat (2.6, gcc 3.3)	
Gentoo Linux	PowerPC 64	8.1.0	构建农场 cobra (1.4.16, gcc 3.4.3)	
Gentoo Linux	x86	8.0.0	Paul Bort (<pbort@tmwsystems.com>), 2004-12-07	
HP-UX	IA64	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-15	11.23, `gcc` and `cc`; see `doc/FAQ_HPUX`
HP-UX	PA-RISC	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-15	10.20 and 11.23, `gcc` and `cc`; see `doc/FAQ_HPUX`

OS	处理器	版本	报告	备注
IRIX	MIPS	8.1.0	Kenneth Marshall (<ktm@is.rice.edu>), 2005-11-04	6.5, cc only
Mac OS X	PowerPC	8.1.0	构建农场 tuna (10.4.2, gcc 4.0); cuckoo (10.3.9, gcc 3.3); wallaroo (10.3.8, gcc 3.3)	
Mandrake Linux	x86	8.1.0	构建农场 shrew (10.0, gcc 3.3.2)	
NetBSD	arm32	7.4	Patrick Welche (<prlw1@newn.cam.ac.uk>), 2003-11-12	1.6ZE/acorn32
NetBSD	m68k	8.1.0	构建农场 osprey (2.0, gcc 3.3.3)	
NetBSD	Sparc	7.4.1	Peter Eisentraut (<peter_e@gmx.net>), 2003-11-26	1.6.1, 32-bit
NetBSD	x86	8.0.0	构建农场 canary, snapshot 2004-12-06 03:30:00	1.6
OpenBSD	Sparc	8.0.0	Chris Mair (<list@1006.org>), 2005-01-10	3.3
OpenBSD	Sparc64	8.1.0	构建农场 spoonbill (3.6, gcc 3.3.2)	compiler bug affects contrib/seg
OpenBSD	x86	8.0.0	构建农场 emu, snapshot 2004-12-06 11:35:03	3.6
Red Hat Linux	AMD64	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Red Hat Linux	IA64	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Red Hat Linux	PowerPC	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Red Hat Linux	PowerPC 64	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Red Hat Linux	S/390	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Red Hat Linux	S/390x	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4

OS	处理器	版本	报告	备注
Red Hat Linux	x86	8.1.0	Tom Lane (<tgl@sss.pgh.pa.us>), 2005-10-23	RHEL 4
Slackware Linux	x86	8.1.0	Sergey Koposov (<math@sai.msu.ru>), 2005-10-24	10.0
Solaris	Sparc	8.1.0	构建农场 buzzard (Solaris 10, gcc 3.3.2); Robert Lor (<Robert.Lor@sun.com>), 2005-11-04 (Solaris 9); Kenneth Marshall (<ktm@is.rice.edu>), 2005-10-28 (Solaris 8, gcc 3.4.3)	see doc/FAQ_Solaris
Solaris	x86	8.1.0	Robert Lor (<Robert.Lor@sun.com>), 2005-11-04 (Solaris 10)	see doc/FAQ_Solaris
SUSE Linux	AMD64	8.1.0	Josh Berkus (<josh@agliodbs.com>), 2005-10-23	SLES 9.3
SUSE Linux	IA64	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	SLES 9
SUSE Linux	PowerPC	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	SLES 9
SUSE Linux	PowerPC 64	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	SLES 9
SUSE Linux	S/390	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	SLES 9
SUSE Linux	S/390x	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	SLES 9
SUSE Linux	x86	8.0.0	Reinhard Max (<max@suse.de>), 2005-01-03	9.0, 9.1, 9.2, SLES 9
Tru64 UNIX	Alpha	8.1.0	Honda Shigehiro (<fwif0083@mb.infoweb.ne.jp>), 2005-11-01	5.0, cc 6.1-011
UnixWare	x86	8.1.0	构建农场 firefly (7.1.4, cc 4.2)	see doc/FAQ_SCO
Windows	x86	8.1.0	构建农场 loris (XP Pro, gcc 3.2.3); snake (Windows	see doc/FAQ_MINGW

OS	处理器	版本	报告	备注
			Server 2003, gcc 3.4.2)	
Windows with Cygwin	x86	8.1.0	构建农场 ferret (XP Pro, gcc 3.3.3)	see `doc/FAQ_CYGWIN`
Yellow Dog Linux	PowerPC	8.1.0	构建农场 carp (4.0, gcc 3.3.3)	

不受支持的平台：. 以下平台要么已知无法工作，要么只能在相当久远的早期版本中工作。我们把它们列在这里，是想让你知道只要给予一些关注，这些平台是可能受到支持的。

OS	Processor	Version	Reported	Remarks
BeOS	x86	7.2	Cyril Velter (`<cyril.velter@libertysurf.fr>`), 2001-11-29	needs updates to semaphore code
Linux	PlayStation 2	8.0.0	Chris Mair (`<list@1006.org>`), 2005-01-09	requires `--disable-spinlocks` (works, but slow)
NetBSD	Alpha	7.2	Thomas Thai (`<tom@minnesota.com>`), 2001-11-20	1.5W
NetBSD	MIPS	7.2.1	Warwick Hunter (`<whunter@agile.tv>`), 2002-06-13	1.5.3
NetBSD	PowerPC	7.2	Bill Studenmund (`<wrstuden@netbsd.org>`), 2001-11-28	1.5
NetBSD	VAX	7.1	Tom I. Helbekkmo (`<tih@kpnQwest.no>`), 2001-03-30	1.5
QNX 4 RTOS	x86	7.2	Bernd Tegge (`<tegge@repas-aeg.de>`), 2001-12-10	needs updates to semaphore code; see also `doc/FAQ_QNX4`
QNX RTOS v6	x86	7.2	Igor Kovalenko (`<Igor.Kovalenko@motorola.com>`), 2001-11-20	patches available in archives, but too late for 7.2
SCO OpenServer	x86	7.3.1	Shibashish Satpathy (`<shib@postmark.net>`), 2002-12-11	5.0.4, `gcc`; see also `doc/FAQ_SCO`
SunOS 4	Sparc	7.2	Tatsuo Ishii (`<t-ishii@sra.co.jp>`), 2001-12-04	

第 15 章 Windows上仅安装客户端

虽然Windows上完整的 PostgreSQL安装只能使用 MinGW或Cygwin构建，但 C 客户端库（libpq）和交互式终端（psql）可以使用其他 Windows 工具集编译。源码发行版中包含了用于Microsoft Visual C++ 和Borland C++的 Makefile。为其他配置手工编译这些库应当也是可能的。

> **提示**
>
> 推荐使用MinGW或 Cygwin。如果要使用这些工具集之一，请参见第 14 章。

要使用Microsoft Visual C++构建你在 Windows上所能构建的一切，请进入 `src`目录并输入命令

```
nmake /f win32.mak
```

这假定你的路径中已有Visual C++。

要使用Borland C++构建一切，请进入 `src`目录并输入命令

```
make -N -DCFG=Release /f bcc32.mak
```

将构建出下列文件：

`interfaces\libpq\Release\libpq.dll`

可动态链接的前端库

`interfaces\libpq\Release\libpqdll.lib`

用于把你的程序链接到`libpq.dll`的导入库

`interfaces\libpq\Release\libpq.lib`

前端库的静态版本

`bin\psql\Release\psql.exe`

PostgreSQL的交互式终端

通常你不需要安装任何客户端文件。你应该把`libpq.dll` 文件放在与应用程序的 .EXE 文件相同的目录中。只有在出于某种原因无法这样做时，才应把它安装到`WINNT\SYSTEM32`目录（在 Windows 95/98/ME 系统上则是`WINDOWS\SYSTEM`目录）。如果该文件是用安装程序安装的，安装时应利用文件中包含的`VERSIONINFO`资源进行版本检查，以确保不会覆盖更新版本的库。

如果你打算在这台机器上使用libpq进行开发，就必须把源码树的`src\include`和 `src\interfaces\libpq`子目录加到编译器设置的包含路径中。

要使用该库，你必须把`libpqdll.lib`文件加到你的项目中。（在 Visual C++ 中，只需在项目上点击右键并选择添加它即可。）

Microsoft的免费开发工具可以从http://msdn.microsoft.com/visualc/vctoolkit2003/下载。你还需要平台 SDK 中的`MSVCRT.lib`，可从http://www.microsoft.com/msdownload/platformsdk/sdkupdate/获取。你也可以从http://msdn.microsoft.com/netframework/downloads/updates/default.aspx下载.NET框架。安装之后，

工具包的二进制文件必须位于你的路径中，而且你可能需要添加`/lib:<libpath>`来指向`MSVCRT.lib`。免费的Borland C++ 编译器工具可以从http://www.borland.com/products/downloads/download_cbuilder.html#下载，并且需要类似的设置。

第 16 章 操作系统环境

本章讨论如何设置和运行数据库服务器，以及它与操作系统的交互。

16.1. The PostgreSQL User Account

与任何可从外部访问的服务器守护进程一样，建议使用单独的用户账户运行 PostgreSQL。该用户账户应该只拥有服务器管理的数据，不应与其他守护进程共用。（例如，使用 nobody 用户就不是好主意。）不建议将可执行文件安装为由该用户所有，因为系统一旦被攻破，就可能修改自己的二进制文件。

要在系统中添加一个 Unix 用户账户，请查找useradd或adduser命令。通常使用的用户名是 postgres，本书也都按此假定，不过你也可以使用其他名称。

16.2. 创建一个数据库集簇

在做任何事情之前，必须先在磁盘上初始化一个数据库存储区域。我们把它称为数据库集簇。（SQL 标准使用术语目录集簇（catalog cluster）。）数据库集簇是由一个正在运行的数据库服务器实例管理的一组数据库。初始化之后，数据库集簇中会包含一个名为postgres的数据库，作为供工具、用户和第三方应用使用的默认数据库。数据库服务器本身并不要求该postgres数据库存在，但许多外部工具程序都假定它存在。初始化期间，集簇中还会创建另一个数据库，名为template1。顾名思义，它会作为后续新建数据库的模板；不应将它用于实际工作。（关于在集簇中创建新数据库的信息，见第 19 章。）

从文件系统角度看，数据库集簇就是一个目录，所有数据都存储在其下。我们将它称为数据目录或数据区域。数据存放位置完全由你决定，没有默认值；常用位置包括 /usr/local/pgsql/data或/var/lib/pgsql/data。要初始化数据库集簇，请使用命令initdb，该命令随以下产品安装：PostgreSQL。数据库集簇在文件系统中的目标位置通过 -D 选项指定，例如：

```
$initdb -D /usr/local/pgsql/data
```

注意，必须先登录 PostgreSQL 用户账户，再执行此命令。该账户已在上一节介绍。

> **提示**
>
> 除了使用 -D 选项，也可以设置环境变量 PGDATA。

如果指定的目录尚不存在，initdb 会尝试创建它。但（如果你听从了我们的建议、创建了非特权账户）它很可能没有这样的权限。这种情况下，你应该自己（以 root 身份）创建该目录，并把所有者改为 PostgreSQL 用户。具体做法可能如下：

```
root# mkdir /usr/local/pgsql/data
root# chown postgres /usr/local/pgsql/data
root# su postgres
postgres$ initdb -D /usr/local/pgsql/data
```

如果数据目录看起来已经被初始化过，initdb 会拒绝运行。

由于数据目录包含存储在数据库中的全部数据，必须保护它免遭未经授权的访问。因此，initdb 会撤销除 PostgreSQL 用户以外所有人的访问权限。

不过，尽管目录内容是安全的，默认的客户端认证设置却允许任何本地用户连接到数据库，甚至成为数据库超级用户。如果你不信任其他本地用户，我们建议在运行initdb时使用-W、--pwprompt或--pwfile之一，为数据库超级用户指定密码。另外，指定-A md5或-A password，以避免使用默认的trust认证方式；或者在运行initdb之后、首次启动服务器之前修改生成的pg_hba.conf文件。（其他合理方法包括使用ident认证，或使用文件系统权限限制连接。更多信息见第 20 章。）

initdb 还会初始化数据库集簇的默认区域设置。通常，它会直接采用环境中的区域设置，并将其应用于初始化的数据库。可以为数据库指定不同的区域设置；更多信息见第 21.1 节。特定数据库集簇中使用的排序顺序由initdb设置，除非把所有数据转储出来、重新运行initdb、再重新装载数据，否则此后无法更改。使用 C 或 POSIX 以外的区域设置还会影响性能。因此，一开始就作出正确选择很重要。

initdb还会为数据库集簇设置默认字符集编码。通常应选择与区域设置相匹配的字符集编码。详见第 21.2 节。

16.3. 启动数据库服务器

在任何人访问数据库之前，必须先启动数据库服务器。数据库服务器程序名为 postmaster。其中，postmaster必须知道应到哪里查找它要使用的数据。这通过 -D 选项指定。因此，启动服务器最简单的方法是：

```
$ postmaster -D /usr/local/pgsql/data
```

这会让服务器在前台运行。必须先登录 PostgreSQL 用户账户再执行此操作。如果不使用 -D，服务器会尝试使用环境变量所指定的数据目录，该变量为 PGDATA。如果也未提供该变量，则会失败。

通常最好在后台启动 postmaster。做法是使用常见的 Unix shell 语法：

```
$ postmaster -D /usr/local/pgsql/data >logfile 2>&1 &
```

如上所示，把服务器的stdout和stderr输出保存到某个地方非常重要。这有助于审计，也有助于诊断问题。（关于日志文件处理的更深入讨论，见第 22.3 节。）

postmaster还接受许多其他命令行选项。
更多信息请见postmaster参考页以及下面的第 17 章。

这些 shell 语法很快就会让人觉得繁琐。因此提供了包装器程序pg_ctl来简化一些任务。例如：

```
pg_ctl start -l logfile
```

这会在后台启动服务器，并把输出写入指定的日志文件。这里的 -D 选项含义与在 postmaster 中相同。pg_ctl 还可用于停止服务器。

通常，你会希望在计算机启动时就启动数据库服务器。自动启动脚本依赖于操作系统。PostgreSQL 在 contrib/start-scripts 目录中附带了一些示例脚本。安装这些脚本需要 root 权限。

不同系统在引导时启动守护进程的惯例各不相同。许多系统有/etc/rc.local或/etc/rc.d/rc.local文件，其他系统则使用rc.d目录。无论采用哪种方式，服务器都必须由

PostgreSQL用户账户而不是 root或其他用户来启动。因此，你大概应该使用`su -c '...' postgres`这种形式来构造命令。例如：

```
su -c 'pg_ctl start -D /usr/local/pgsql/data -l serverlog' postgres
```

下面是一些与操作系统相关的补充建议。（在每种情况下，
请确保使用正确的安装目录和用户名；这里展示的是通用值。）

- 在FreeBSD上，请查看PostgreSQL源码发布包中的`contrib/start-scripts/freebsd`文件。
- 在OpenBSD上，把以下内容加入`/etc/rc.local`：

  ```
  if [ -x /usr/local/pgsql/bin/pg_ctl -a -x /usr/local/pgsql/bin/
  postmaster ]; then
      su - -c '/usr/local/pgsql/bin/pg_ctl start -l /var/postgresql/
  log -s' postgres
      echo -n ' postgresql'
  fi
  ```

- 在Linux系统上，可以把

  ```
  /usr/local/pgsql/bin/pg_ctl start -l logfile -D /usr/local/pgsql/
  data
  ```

  加入`/etc/rc.d/rc.local`，或者查看PostgreSQL源码发布包中的`contrib/start-scripts/linux`文件。

- 在NetBSD上，可以根据喜好使用FreeBSD或Linux的启动脚本。
- 在Solaris上，创建一个名为`/etc/init.d/postgresql`的文件，其中包含以下行：

  ```
  su - postgres -c "/usr/local/pgsql/bin/pg_ctl start -l logfile -D /
  usr/local/pgsql/data"
  ```

  然后在`/etc/rc3.d`中创建一个指向它的符号链接`S99postgresql`。

当`postmaster`运行时，其PID会保存在数据目录中的`postmaster.pid`文件里。
这可以用来防止多个`postmaster`进程在同一个数据目录上运行，也可以用来关闭`postmaster`进程。

16.3.1. 服务器启动失败

服务器启动失败有几种常见原因。请检查服务器日志文件，
或者手工启动服务器（不要重定向标准输出和标准错误），看看出现了什么错误消息。
下面将更详细地解释一些最常见的错误消息。

```
LOG:  could not bind IPv4 socket: Address already in use
HINT:  Is another postmaster already running on port 5432? If not,
 wait a few seconds and retry.
FATAL:  could not create TCP/IP listen socket
```

这通常就是字面上的意思：你试图在一个已有服务器运行的端口上再启动另一个`postmaster`。
不过，如果内核错误消息不是`Address already in use`或类似变体，也可能是别的问题。例如，试图在一个保留端口上启动一个`postmaster`，可能会得到类似下面的消息：

```
$ postmaster -p 666
LOG:  could not bind IPv4 socket: Permission denied
HINT:  Is another postmaster already running on port 666? If not, wait
 a few seconds and retry.
FATAL:  could not create TCP/IP listen socket
```

像下面这样的消息：

```
FATAL:  could not create shared memory segment: Invalid argument
DETAIL:  Failed system call was shmget(key=5440001, size=4011376640,
 03600).
```

可能表示内核对共享内存大小的限制小于 PostgreSQL 尝试创建的工作区（本例中为4011376640 字节）。也可能表示内核中根本没有配置 System V 风格的共享内存支持。作为临时解决办法，可以尝试以少于通常数量的缓冲区启动服务器（shared_buffers）。
最终还是需要重新配置内核，增大允许的共享内存大小。
如果尝试在同一台机器上启动多个服务器，而它们请求的总空间超出内核限制，
也可能看到这条消息。

像下面这样的错误：

```
FATAL:  could not create semaphores: No space left on device
DETAIL:  Failed system call was semget(5440126, 17, 03600).
```

并不意味着你已经用光了磁盘空间。它表示你的内核对System V信号量数量的限制，小于PostgreSQL想要创建的数量。和上面一样，你也许可以通过减少允许的连接数（max_connections）来暂时绕过这个问题，但最终还是应该提高内核限制。

如果收到“illegal system call”错误，很可能是内核根本不支持共享内存或信号量。在这种情况下，唯一的办法就是重新配置内核以启用这些功能。

关于配置System V IPC功能的细节请见第 16.4.1 节。

16.3.2. 客户端连接问题

客户端一侧可能出现的错误种类很多，并且依赖于具体应用，
但其中有一些可能直接与服务器的启动方式有关。除下面列出的几种情况外，
其他问题应查阅相应客户端应用的文档。

```
psql: could not connect to server: Connection refused
        Is the server running on host "server.joe.com" and accepting
        TCP/IP connections on port 5432?
```

这种常见的失败表示“我找不到可通信的服务器”。尝试 TCP/IP 通信时，它就会显示为上述形式。一个常见错误是忘记配置服务器以允许 TCP/IP 连接。

另一种情况是，尝试通过 Unix 域套接字与本地服务器通信时，会得到以下消息：

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

最后一行有助于确认客户端是否正尝试连接到正确的位置。如果那里实际上没有运行服务器，
内核错误消息通常会像示例那样是 `Connection refused` 或 `No such file or directory`。（要注意，在这里`Connection refused`并不表示服务器收到了你的连接请求并

拒绝了它；那种情况会产生另一条消息，如第 20.3 节所示。）其他错误消息，例如 `Connection timed out`，可能表示更底层的问题，例如网络不通。

16.4. 管理内核资源

一个大型PostgreSQL安装会很快耗尽各种操作系统资源限制。（在某些系统上，出厂默认值非常低，你甚至不需要真正“大型”的安装就会遇到问题。）如果你遇到过这类问题，请继续阅读。

16.4.1. 共享内存和信号量

共享内存和信号量统称为“System VIPC”（连同消息队列，但消息队列与PostgreSQL无关）。几乎所有现代操作系统都提供这些特性，但其中许多系统默认没有启用它们或尺寸不足，尤其是随着可用内存和数据库应用需求的增长。（对于Windows、QNX和BeOS端口，PostgreSQL为这些设施提供了自己的替代实现。）

完全缺少这些功能时，通常会在服务器启动时出现 Illegal system call 错误。在这种情况下，除了重新配置内核别无选择。没有它们，PostgreSQL 就无法工作。不过，在现代操作系统中，这种情况很少见。

当PostgreSQL超出各种硬性IPC限制之一时，服务器会拒绝启动，并应留下带有指导性的错误消息，说明问题所在以及应如何处理（另见第 16.3.1 节）。相关内核参数在不同系统上的命名基本一致，表 16.1给出了概览；不过，设置它们的方法却各不相同。下面给出一些平台上的建议。

表 16.1. System V IPC参数

名称	描述	合理的值
`SHMMAX`	共享内存段的最大尺寸（字节）	至少数 MB（见正文）
`SHMMIN`	共享内存段的最小尺寸（字节）	1
`SHMALL`	可用共享内存的总量（字节或页面）	如果是字节，同`SHMMAX`；如果是页面，为`ceil(SHMMAX/PAGE_SIZE)`
`SHMSEG`	每个进程的最大共享内存段数目	只需要 1 段，但是默认值高很多
`SHMMNI`	系统范围内的最大共享内存段数目	像`SHMSEG`外加其他应用的空间
`SEMMNI`	信号量标识符（即，集合）的最大数目	至少 `ceil(max_connections / 16)`
`SEMMNS`	系统范围内的最大信号量数目	`ceil(max_connections / 16) * 17`，并为其他应用保留空间
`SEMMSL`	每个集合中信号量的最大数目	至少 17
`SEMMAP`	信号量映射中的项数	见文本
`SEMVMX`	信号量的最大值	至少 1000（默认值常常是 32767，如非必要不要更改）

最重要的共享内存参数是`SHMMAX`，即以字节计的共享内存段的最大尺寸。如果`shmget`返回像Invalid argument这样的错误消息，很可能是超出了这个限制。所需共享内存段的尺寸取决于若干PostgreSQL配置参数，如表 16.2所示。作为临时解决办法，你可以调低其中某些设置来避免失败。作为粗略的近似，你可以把所需的段尺寸估计为 500 kB 加上表中所示的变量部分。（你可能得到的任何错误消息都会包含失败分配请求的确切尺寸。）虽然`SHMMAX`小到 1 MB 也能让PostgreSQL运行，但要获得可接受的性能，至少需要 4 MB。比较理想的设置是几十兆字节。

一些系统还对系统内共享内存的总量有限制（SHMALL）。
请确保它对PostgreSQL以及使用共享内存段的任何其他应用来说都足够大。注意，在许多系统上SHMALL以页而不是字节计。

不太可能出问题的是共享内存段的最小尺寸（SHMMIN），对PostgreSQL来说应该最多大约是500 kB（通常只是 1）。而系统范围（SHMMNI）或每个进程（SHMSEG）的最大共享内存段数目不太可能会导致问题，除非你的系统把它们设成零。

PostgreSQL为每个允许的连接（max_connections）使用一个信号量，每 16 个组成一组。每组还包含第 17 个信号量，其中存放一个“魔数”，用来检测与其他应用程序使用的信号量集的冲突。系统中的信号量最大数量由SEMMNS设置，因此它必须至少等于max_connections，再加上每 16 个允许连接一个的额外信号量（见表 16.1中的公式）。参数SEMMNI决定系统中同时可以存在的信号量集的数量上限，因此它必须至少为ceil(max_connections / 16)。降低允许的连接数是失败时的一种临时变通办法，这种失败通常表现为semget函数返回措辞令人困惑的No space left on device。

在某些情况下，可能还需要增大SEMMAP，使其至少与SEMMNS处于同一数量级。
这个参数定义信号量资源映射的大小，映射中每个连续的可用信号量块都需要占用一项。
每当一个信号量集合被释放时，它要么会并入与该释放块相邻的现有项，
要么会登记为一个新的映射项。如果映射已满，被释放的信号量就会丢失（直到重启）。因此，随着时间推移，信号量空间碎片化可能会导致可用信号量少于应有数量。

SEMMSL 参数决定一个信号量集中可以包含多少个信号量，对于PostgreSQL，它必须至少为17。

与“信号量撤销”有关的其他各种设置，如SEMMNU和SEMUME，不会影响PostgreSQL。

BSD/OS

共享内存. 默认只支持 4 MB 共享内存。请记住，共享内存是不可换页的，它被锁定在 RAM 中。要增加系统支持的共享内存量，可以在内核配置文件中加入类似下面的内容：

```
options "SHMALL=8192"
options "SHMMAX=\(SHMALL*PAGE_SIZE\)"
```

SHMALL以 4 kB 页计，因此值 1024 表示 4 MB 共享内存。所以上面的设置把最大共享内存区域增加到 32 MB。对于运行 4.3 或更高版本的用户，可能还需要把KERNEL_VIRTUAL_MB增加到默认值248之上。所有修改完成后，重新编译内核并重启。

对于运行 4.0 及更早版本的用户，可使用 bpatch 查找当前内核中的 sysptsize 值。该值在系统引导时动态计算：

```
$ bpatch -r sysptsize
0x9 = 9
```

然后，把 SYSPTSIZE 作为一个硬编码值加入内核配置文件。在用 bpatch 找到的值的基础上增加：每需要额外 4 MB 共享内存就加 1。

```
options "SYSPTSIZE=16"
```

sysptsize 无法通过 sysctl 修改。

Semaphores. 你可能还想同时增加信号量的数量；系统默认总数 60 只允许大约 50 个PostgreSQL连接。请在内核配置文件中设置你想要的值，例如：

```
options "SEMMNI=40"
options "SEMMNS=240"
```

FreeBSD

默认设置只适合小型安装（例如，默认的SHMMAX为 32 MB）。可以通过sysctl或loader接口进行修改。以下参数可以用sysctl设置：

```
$sysctl -w kern.ipc.shmall=32768$sysctl -w kern.ipc.shmmax=
134217728$sysctl -w kern.ipc.semmap=256
```

要让这些设置在重启后保持，请修改/etc/sysctl.conf。

就sysctl而言，其余的信号量设置是只读的，但可以在系统引导前通过loader提示符修改：

```
(loader)set kern.ipc.semmni=256(loader)set kern.ipc.semmns=512(
loader)set kern.ipc.semmnu=256
```

同样，这些设置也可以保存在/boot/loader.conf中，以便重启后保持。）

你可能还希望配置内核，把共享内存锁定在 RAM 中，防止其被换出到交换区。这可以通过 sysctl 设置 kern.ipc.shm_use_phys 来实现。

如果通过启用 sysctl 的 security.jail.sysvipc_allowed 在 FreeBSD jail 中运行，应让不同 jail 中的 postmaster 使用不同的操作系统用户。这能提高安全性，因为它可以防止非 root 用户干扰不同 jail 中的共享内存或信号量，也能让 PostgreSQL 的 IPC 清理代码正常工作。（在 FreeBSD 6.0 及更高版本中，IPC 清理代码无法正确检测其他 jail 中的进程，导致无法在不同 jail 中使用同一端口运行 postmaster。）

FreeBSD 4.0 之前的版本行为类似于 NetBSD和OpenBSD（见下文）。

NetBSD
OpenBSD

编译内核时需要启用选项SYSVSHM和SYSVSEM（默认已启用）。
共享内存的最大尺寸由选项SHMMAXPGS（以页计）决定。
下面展示了一个如何设置各参数的例子（OpenBSD使用option代替）：

```
options         SYSVSHM
options         SHMMAXPGS=4096
options         SHMSEG=256

options         SYSVSEM
options         SEMMNI=256
options         SEMMNS=512
options         SEMMNU=256
options         SEMMAP=256
```

你可能还希望配置内核，把共享内存锁定在 RAM 中，防止其被换出到交换区。这可以通过 sysctl 设置 kern.ipc.shm_use_phys 来实现。

HP-UX

默认设置通常足以满足普通安装的需求。在 HP-UX 10 上，SEMMNS 的出厂默认值为 128，对较大的数据库站点来说可能过低。

可以在 System Administration Manager（SAM）的 Kernel Configuration→Configurable Parameters 中设置 IPC 参数。完成后选择 Create A New Kernel。

Linux

默认设置只适合小型安装（默认的最大段尺寸为 32 MB）。不过，其余的默认值都相当充裕，通常不需要更改。最大段尺寸可以通过`sysctl`接口修改。例如，要允许 128 MB，并显式把共享内存总量的最大值设置为 2097152 页（默认值）：

```
$sysctl -w kernel.shmmax=134217728$sysctl -w kernel.shmall=2097152
```

此外，可以将这些设置保存在`/etc/sysctl.conf`中，使其在重启后仍然生效。

较旧的发行版可能没有 `sysctl` 程序，但可以通过操作 `/proc` 文件系统进行等效更改：

```
$echo 134217728 >/proc/sys/kernel/shmmax$echo 2097152 >/proc/sys/
kernel/shmall
```

MacOS X

在 OS X 10.2 及更早版本中，请编辑文件 `/System/Library/StartupItems/SystemTuning/SystemTuning` 并更改以下命令中的值：

```
sysctl -w kern.sysv.shmmax
sysctl -w kern.sysv.shmmin
sysctl -w kern.sysv.shmmni
sysctl -w kern.sysv.shmseg
sysctl -w kern.sysv.shmall
```

在 OS X 10.3 及更高版本中，这些命令已被移到 `/etc/rc` 中，必须在该文件中编辑。注意，`/etc/rc` 通常会被 OS X 系统更新（例如 10.3.6 到 10.3.7）覆盖，因此每次更新后可能都需要重新进行这些编辑。在所有版本中，都需要重启才能使更改生效。

在 OS X 10.3.9 及更高版本中，文件 `/etc/sysctl.conf` 允许共享内存设置在操作系统升级后保留，这是设置这些参数的推荐方法。使用该文件时，必须设置全部五个共享内存值，否则更改将被忽略。

在这个平台上，`SHMALL` 以 4KB 的页为单位，并且近期 OS X 版本会拒绝将 `SHMALL` 和 `SHMMAX` 设为非 4096 整数倍的值的尝试。

SCO OpenServer

在默认配置中，每个共享内存段只允许 512 kB。要提高此设置，首先切换到目录`/etc/conf/cf.d`。要显示`SHMMAX`的当前值，运行：

```
./configure -y SHMMAX
```

要为`SHMMAX`设置新值，运行：

```
./configure SHMMAX=value
```

其中*`value`*是你要使用的新值（以字节为单位）。设置`SHMMAX`后，重新构建内核：

```
./link_unix
```

然后重启。

AIX

至少从 5.1 版起，应该不需要对 `SHMMAX` 等参数进行任何特殊配置，因为它似乎已配置为允许将全部内存用作共享内存。这也是 DB/2 等其他数据库常用的配置。

不过，可能需要修改 `/etc/security/limits` 中的全局 `ulimit` 信息，因为文件大小（`fsize`）和文件数量（`nofiles`）的默认硬限制可能过低。

Solaris

至少在 2.6 版本中，共享内存段的默认最大尺寸对PostgreSQL来说太低。相关设置可以在`/etc/system`中修改，例如：

```
set shmsys:shminfo_shmmax=0x2000000
set semsys:seminfo_semmns=1024
```

要让更改生效，需要重启。

另见http://sunsite.uakom.sk/sunworldonline/swol-09-1997/swol-09-insidesolaris.html，其中介绍了Solaris下的共享内存。

UnixWare

在UnixWare 7 中，默认配置下共享内存段的最大尺寸只有 512 kB。要显示`SHMMAX`的当前值，请运行：

```
/etc/conf/bin/idtune -g SHMMAX
```

这会显示当前值、默认值、最小值和最大值。要为`SHMMAX`设置新值，请运行：

```
/etc/conf/bin/idtune SHMMAX value
```

其中*`value`*是你要使用的新值（以字节计）。设置`SHMMAX`之后，重新构建内核：

```
/etc/conf/bin/idbuild -B
```

and reboot.

表 16.2. 影响PostgreSQL共享内存用量的配置参数

名称	近似倍数（每个增量的字节数）
max_connections	400 + 220 * max_locks_per_transaction
max_prepared_transactions	600 + 220 * max_locks_per_transaction
shared_buffers	8300（假定 8K 的 `BLCKSZ`）
wal_buffers	8200（假定 8K 的 `BLCKSZ`）
max_fsm_relations	70
max_fsm_pages	6

16.4.2. Resource Limits

类 Unix 操作系统会施加多种资源限制，这些限制可能干扰 PostgreSQL 服务器的运行。其中尤其重要的是：每个用户可用的进程数限制、每个进程可打开文件数限制，以及每个进程可用内存量限制。每一种限制都有“硬”限制和“软”限制。实际生效的是软限制，但用户可以自行把它调高到不超过硬限制；硬限制则只能由 root 用户修改。系统调用`setrlimit`负责设置这些参数。shell 的内置命令`ulimit`（Bourne shell）或`limit`（csh）可用于在命令行控制资源限制。在 BSD 派生系统上，`/etc/login.conf`文件控制登录时设置的各种资源限制。详细信息请参阅操作系统文档。相关参数有`maxproc`、`openfiles`和`datasize`。例如：

```
default:\
```

```
...
        :datasize-cur=256M:\
        :maxproc-cur=256:\
        :openfiles-cur=256:\
...
```

（-cur表示软限制；把它改为-max则表示设置硬限制。）

内核还可能对某些资源施加系统范围的限制。

- 在Linux上，内核参数 /proc/sys/fs/file-max确定内核支持的最大打开文件数。可以向该文件写入一个不同的数字来修改它，也可以在/etc/sysctl.conf中添加相应赋值。
  每个进程的文件数上限是在内核编译时固定的；请参阅 /usr/src/linux/Documentation/proc.txt获取更多信息。

PostgreSQL服务器为每个连接使用一个进程，因此你至少应提供与允许连接数相同数量的进程，再加上系统其他部分所需的进程数。通常这不是问题，但如果你在一台机器上运行多个服务器，资源可能就会变得紧张。

打开文件数的出厂默认限制通常被设为“对共享环境友好”的值，
也就是允许许多用户在一台机器上共存，而不会占用不成比例的系统资源。
如果你在一台机器上运行很多服务器，这也许正合适；但在专用服务器上，
你可能会希望提高这个限制。

另一方面，有些系统允许单个进程打开非常多的文件；如果不止少数几个进程都这么做，
就很容易超过系统范围的限制。如果你遇到这种情况，又不想修改系统范围的限制，可以设置PostgreSQL的max_files_per_process配置参数来限制其打开文件的消耗。

16.4.3. Linux 内存过量分配

在 Linux 2.4 及更高版本中，默认的虚拟内存行为对 PostgreSQL 并非最佳。由于内核实现内存过量分配的方式，如果其他进程的内存需求导致系统耗尽虚拟内存，内核可能会终止 PostgreSQL 的 postmaster 进程。

如果发生这种情况，你会看到类似下面的内核消息（至于应到哪里查看此类消息，
请参阅你的系统文档和配置）：

```
Out of Memory: Killed process 12345 (postmaster).
```

这表示 postmaster 进程因内存压力而被终止。尽管现有数据库连接仍会继续正常运行，但新连接将不再被接受。要恢复服务，必须重启PostgreSQL。

避免该问题的一种办法，是让
PostgreSQL运行在一台你能确定不会被其他进程耗尽内存的机器上。

在 Linux 2.6 及更高版本中，一种更好的解决办法是修改内核的行为，使其不会“过量分配”内存。这可以通过sysctl选择严格的过量分配模式来完成：

```
sysctl -w vm.overcommit_memory=2
```

或者在以下文件中加入等效条目：/etc/sysctl.conf。你可能还希望修改相关设置 vm.overcommit_ratio。详细信息参见内核文档文件 Documentation/vm/overcommit-accounting。

据报告，某些厂商的 Linux 2.4 内核包含 2.6 版过量分配 sysctl 参数的早期实现。但是，在没有相关代码的 2.4 内核上将 vm.overcommit_memory 设为 2，会使情况更糟，而非更好。

建议在 2.4 安装上尝试之前，检查实际内核源码（参见文件 mm/mmap.c 中的函数 vm_enough_memory），确认内核支持的功能。存在 overcommit-accounting 文档文件，不能作为该功能已存在的证据。如有任何疑问，请咨询内核专家或内核供应商。

16.5. 关闭服务器

关闭数据库服务器有几种方式。通过向postmaster进程发送不同的信号来控制关闭的类型。

SIGTERM

收到SIGTERM后，服务器禁止新连接，但让现有会话正常结束其工作。
只有在所有会话正常终止之后，它才关闭。这就是智能关闭。

SIGINT

服务器禁止新连接，并向所有现有服务器进程发送SIGTERM，
使它们中止当前事务并立即退出。然后它等待所有服务器进程退出并最终关闭。这就是
快速关闭。

SIGQUIT

这就是立即关闭，postmaster进程将向所有子进程发送SIGQUIT并立即退出，
而不进行正常的关闭。子进程收到SIGQUIT后同样立即退出。
这将导致下次启动时需要恢复（通过重放 WAL 日志）。仅在紧急情况下才建议这样做。

pg_ctl程序提供了一个便于发送这些信号以关闭服务器的接口。

另外，你也可以用kill直接发送这些信号。可以用ps程序，或者从数据目录中的postmaster.pid文件找到postmaster进程的PID。例如，要执行一次快速关闭：

```
$ kill -INT `head -1 /usr/local/pgsql/data/postmaster.pid`
```

重要

最好不要使用 SIGKILL 关闭服务器。这样做会阻止服务器释放共享内存和信号量，因此在启动新服务器之前，可能必须手动释放它们。此外，SIGKILL 会终止 postmaster 进程，使其没有机会将信号转发给子进程，因此还必须手动逐个终止子进程。

16.6. 加密选项

PostgreSQL提供了多个层次的加密，并在防止数据因数据库服务器被盗、
不诚实的管理员或不安全网络等原因而泄露方面提供了很高的灵活性。为了保护医疗记录、
金融交易等敏感数据，也可能需要使用加密。

密码存储加密

默认情况下，数据库用户口令以 MD5 哈希的形式存储，因此管理员无法确定指派给用户的实际口令。如果客户端认证使用 MD5 加密，那么未加密的口令甚至不会临时出现在服务器上，因为客户端在通过网络发送之前就先对它做了 MD5 加密。

特定列加密

/contrib函数库pgcrypto允许某些字段以加密形式存储。如果只有部分数据是敏感的，
这将很有用。客户端提供解密密钥，数据在服务器上解密后发送给客户端。

当数据被解密并在服务器与客户端之间传输时，
解密后的数据和解密密钥会在服务器端短暂存在。
这就给那些可以完全访问数据库服务器的人（例如系统管理员）
提供了一个短暂的机会来截获密钥和数据。

数据分区加密

在 Linux 上，可以使用“回环设备”在文件系统之上叠加加密。这允许整个文件系统分区在磁盘上加密，由操作系统解密。在 FreeBSD 上，等价的设施称为基于 GEOM 的磁盘加密（gbde），许多其他操作系统也支持这一功能，包括 Windows。

这种机制可以防止在整台计算机或磁盘被盗时，从驱动器中直接读取未加密数据。
但它无法防御文件系统已挂载时的攻击，因为在挂载之后，
操作系统会提供数据的解密视图。不过，要挂载该文件系统，
就必须以某种方式把加密密钥提供给操作系统，
而有时这个密钥就存放在挂载该磁盘的主机上的某处。

跨网络加密口令

`MD5`认证方法在把口令发送给服务器之前，先在客户端对其进行双重加密。它先基于用户名做 MD5 加密，然后再基于建立数据库连接时服务器发送的随机盐加密。通过网络发送给服务器的正是这个双重加密后的值。双重加密不仅可防止口令被发现，
还可防止其他连接在之后利用同一个已加密的口令连接到数据库服务器。

跨网络加密数据

SSL 连接会加密通过网络发送的全部数据：密码、查询和返回的数据。`pg_hba.conf` 文件允许管理员指定哪些主机可以使用非加密连接（`host`），哪些主机必须使用 SSL 加密连接（`hostssl`）。此外，客户端可以指定仅通过 SSL 连接到服务器。也可以使用 Stunnel 或 SSH 加密传输。

SSL 主机认证

客户端和服务器都可以向对方提供 SSL 证书。这需要双方做一些额外配置，但它提供的认证强度高于单纯依赖密码。它可以防止某台计算机短暂冒充服务器，
只为读取客户端发送的密码；同时也有助于防御"中间人"（man in the middle）攻击，即某台位于客户端与服务器之间的计算机冒充服务器，读取并转发双方之间的全部数据。

客户端加密

如果服务器所在机器的系统管理员不可信，那么就有必要由客户端自行加密数据。
这样一来，未加密的数据就永远不会出现在数据库服务器上。
数据在发送给服务器之前就在客户端加密，而查询结果也必须在客户端解密后才能使用。

16.7. 使用 SSL 的安全 TCP/IP 连接

PostgreSQL原生支持使用SSL连接来加密客户端与服务器之间的通信，以提高安全性。
这要求客户端和服务器系统上都安装了OpenSSL，并且在构建
PostgreSQL时启用了该支持（见第 14 章）。

如果构建时启用了SSL支持，那么在`postgresql.conf`中将参数ssl设为`on`，就可以在启动PostgreSQL服务器时启用SSL。以SSL模式启动时，服务器会在数据目录中查找文件`server.key`和`server.crt`，它们必须分别包含服务器私钥和证书。这些文件必须正确设置，启用了SSL的服务器才能启动。如果私钥受口令保护，服务器将提示输入该口令，输入之前不会启动。

服务器会在同一个 TCP 端口上同时监听普通连接和SSL连接，并与每个连接进来的客户端协商是否使用SSL。默认情况下，这由客户端决定；
关于如何把服务器配置为要求某些或全部连接必须使用SSL，请参阅第 20.1 节。

关于如何创建服务器私钥和证书的细节，请参阅OpenSSL文档。自签名证书可用于测试，
但在生产环境中应使用由证书颁发机构（CA）（全局CA之一或本地 CA）签名的证书，以便客户端能验证服务器的身份。要快速创建一个自签名证书，可以使用下面的OpenSSL命令：

```
openssl req -new -text -out server.req
```

请填写`openssl`要求的信息。确保把本地主机名输入为“Common Name”；挑战密码可以留空。该程序会生成一个带口令保护的密钥；它不接受少于四个字符的口令。
要移除口令（如果希望服务器自动启动就必须这样做），请运行命令

```
openssl rsa -in privkey.pem -out server.key
rm privkey.pem
```

输入旧口令以解锁现有密钥。接着执行

```
openssl req -x509 -in server.req -text -key server.key -out server.crt
chmod og-rwx server.key
```

把证书转换为自签名证书，并把密钥和证书复制到服务器将会查找它们的位置。

如果需要验证客户端证书，请把你希望检查的CA的证书放在数据目录中的`root.crt`文件里。
该文件存在时，在 SSL 连接启动时会向客户端请求客户端证书，并且该证书必须由`root.crt`中现有的某个证书签名。

如果`root.crt`文件不存在，就不会请求或检查客户端证书。在这种模式下，SSL 提供通信安全，但不提供认证。

文件`server.key`、`server.crt`和`root.crt`只在服务器启动时检查；
因此更改它们后必须重启服务器才能生效。

16.8. Secure TCP/IP Connections with SSH Tunnels

可以使用SSH来加密客户端与PostgreSQL服务器之间的网络连接。如果配置得当，
即使客户端本身不支持 SSL，这种方法也能提供足够安全的网络连接。

首先确认在PostgreSQL服务器所在的同一台机器上有一个正常运行的SSH服务器，
并且你可以使用`ssh`以某个用户身份登录。
然后就可以在客户端机器上用如下命令建立安全隧道：

```
ssh -L 3333:foo.com:5432 joe@foo.com
```

`-L`参数中的第一个数字 3333 是隧道在本地的端口号；它可以自由选择。第二个数字 5432 是隧道的远端：你的服务器所使用的端口号。两个端口号之间的名称或 IP 地址，是你要连接的数据库服务器所在的主机。要通过这条隧道连接数据库服务器，只需连接本地机器上的 3333 端口：

```
psql -h localhost -p 3333 postgres
```

对数据库服务器来说，它看起来就像是用户`joe@foo.com`真实在连接，
并将应用为该用户和主机配置的认证方式。注意，服务器不会认为该连接是 SSL 加密的，因为实

际上SSH服务器与PostgreSQL服务器之间并没有加密。只要它们位于同一台机器上，这不会带来任何额外的安全风险。

要让隧道建立成功，你必须有权通过`ssh`以`joe@foo.com`身份连接，就像你要用`ssh`建立终端会话时一样。

提示

还有一些其他应用也能提供安全隧道，其思路与刚才描述的 SSH 方法类似。

第 17 章 服务器配置

有许多配置参数会影响数据库系统的行为。本章第一节将介绍如何与配置参数交互。
后续各节将详细讨论每个参数。

17.1. 设置参数

所有参数名都是大小写不敏感的。每个参数都接受四种类型之一的值：布尔、整数、
浮点数或字符串。布尔值可以被写成 on、off、true、false、yes、no、1、0（都是大小写不敏感的）或者这些值的任何无歧义前缀。

设置这些参数最基本的方法是编辑文件 postgresql.conf，它通常位于数据目录中。（initdb 会在那里安装一个默认副本。）该文件的内容可能如下所示：

```
# This is a comment
log_connections = yes
log_destination = 'syslog'
search_path = '"$user", public'
shared_buffers = 128MB
```

每行指定一个参数。名称和值之间的等号是可选的。
空白不重要（引号括起的参数值内部除外），空行会被忽略。井号（#）
表示该行余下部分是注释。不是简单标识符或数字的参数值必须用单引号括起。
要在参数值中嵌入单引号，可以写两个单引号（推荐）或使用反斜线转义单引号。
如果文件包含相同参数的多个条目，则忽略除最后一个之外的所有条目。

每当postmaster进程收到一个SIGHUP信号（最容易的发送方式是使用pg_ctl reload）时，
就会重新读取配置文件。postmaster进程还会把这个信号传播给所有正在运行的服务器进程，
这样现有的会话也能采用新值。另外，你可以直接向一个单一服务器进程发送该信号。
有些参数只能在服务器启动时设置，在配置文件中对这些条目的修改将被忽略，
直到服务器重启。

设置这些配置参数的第二种方法是把它们作为 postmaster 命令的命令行选项给出，例如：

```
postgres -c log_connections=yes -c log_destination='syslog'
```

命令行选项覆盖 postgresql.conf 中任何冲突的设置。注意这意味着你将无法通过编辑 postgresql.conf 即时更改该值，因此虽然命令行方法可能很方便，但日后可能会让你失去灵活性。

有时只给某个特定会话指定命令行选项很有用。为此可以在客户端使用环境变量 PGOPTIONS，把选项传递给 postmaster：

```
env PGOPTIONS='-c geqo=off' psql
```

（这适用于任何基于 libpq 的客户端应用，而不仅仅是 psql。）注意，对于在服务器启动时就固定、或者必须在 postgresql.conf 中指定的参数，这种方式无效。

此外，还可以把一组参数设置指派给某个用户或某个数据库。每当会话启动时，
都会加载相关用户和数据库的默认设置。分别使用ALTER USER和ALTER DATABASE命令来配置这些设置。每个数据库的设置覆盖从postmaster命令行或配置文件收到的任何内容，
而后者又被每用户的设置覆盖；两者都会被每会话的设置覆盖。

有些参数可以在各个SQL会话中用SET命令更改，例如：

```
SET ENABLE_SEQSCAN TO OFF;
```

如果允许通过SET设置，它将覆盖该参数所有其他来源的值。有些参数不能通过SET更改：例如，控制在不重启整个PostgreSQL服务器的情况下无法更改的行为的参数。此外，有些参数需要超级用户权限才能通过SET或ALTER更改。

SHOW命令允许查看所有参数的当前值。

虚拟表pg_settings（在第 42.41 节中描述）也允许显示和更新会话运行时参数。它等价于SHOW和SET，但用起来可能更方便，因为它可以与其他表连接，或者用任何想要的选择条件来查询。

17.2. 文件位置

除了已经提到过的postgresql.conf文件之外，PostgreSQL还使用另外两个手工编辑的配置文件，它们控制客户端认证（其使用在第 20 章中讨论）。默认情况下，所有三个配置文件都存放在数据库集簇的数据目录中。本节描述的参数允许配置文件放在别的地方（这么做可以简化管理，特别是如果配置文件被独立放置，可以很容易保证它得到恰当的备份）。

data_directory (string)

指定用于数据存储的目录。这个参数只能在服务器启动时设置。

config_file (string)

指定主服务器配置文件（通常叫postgresql.conf）。这个参数只能在 postmaster 命令行上设置。

hba_file (string)

指定基于主机认证配置文件（通常叫pg_hba.conf）。这个参数只能在服务器启动的时候设置。

ident_file (string)

指定用于ident认证的配置文件（通常叫pg_ident.conf）。这个参数只能在服务器启动的时候设置。

external_pid_file (string)

指定服务器应创建的额外进程 ID（PID）文件的名称，供postmaster管理程序使用。这个参数只能在服务器启动的时候设置。

在默认安装中不会显式设置以上参数。相反，命令行参数-D或者环境变量PGDATA指定数据目录，并且上述配置文件都能在数据目录中找到。

如果你想把配置文件放在别的地方而不是数据目录中，那么 postmaster 的 -D 命令行选项或者环境变量PGDATA必须指向包含配置文件的目录，并且postgresql.conf中（或者命令行上）的data_directory参数必须设置为数据目录的实际位置。请注意，data_directory将覆盖-D和PGDATA指定的数据目录位置，但是不覆盖配置文件的位置。

如果你愿意，可以使用选项config_file、hba_file和/或ident_file单独指定配置文件名称和位置。config_file只能在postmaster命令行上指定，但是其他参数可以在主配置文件中设置。如果所有三个参数外加data_directory被显式地设置，则不必指定-D或PGDATA。

在设置任何这些参数时，相对路径将被解释为相对于启动 postmaster 时所在目录的路径。

17.3. 连接和认证

17.3.1. 连接设置

listen_addresses (string)

指定服务器用于监听客户端应用连接的 TCP/IP 地址。此值采用以逗号分隔的主机名和/或数字 IP 地址列表的形式。特殊项*对应所有可用的 IP 接口。如果列表为空，服务器不会监听任何 IP 接口，此时只能通过 Unix 域套接字连接。默认值为 localhost，它只允许建立本地“回环”连接。此参数只能在服务器启动时设置。

port (integer)

服务器监听的 TCP 端口，默认是 5432。请注意，服务器监听的所有 IP 地址都使用同一个端口号。此参数只能在服务器启动时设置。

max_connections (integer)

决定数据库服务器允许的最大并发连接数。默认值通常是 100 个连接，但如果内核设置不支持（在 initdb 期间确定），则可能更少。这个参数只能在服务器启动时设置。

增加这个参数可能会导致PostgreSQL请求比你的操作系统默认配置所允许的更多的System V共享内存或信号量。如果有必要，关于如何调整那些参数的信息见第 16.4.1 节。

superuser_reserved_connections (integer)

决定为 PostgreSQL 超级用户连接保留多少个连接“槽”。同时活跃的连接数最多始终只能达到max_connections。当活跃并发连接数至少达到 max_connections 减去 superuser_reserved_connections 时，新连接将只接受超级用户，并且不再接受新的复制连接。

默认值是 3 个连接。该值必须小于 max_connections。这个参数只能在服务器启动时设置。

unix_socket_directory (string)

指定服务器用于监听来自客户端应用连接的 Unix 域套接字目录。默认值通常是 /tmp，但可以在构建时更改。这个参数只能在服务器启动时设置。

unix_socket_group (string)

设置 Unix 域套接字的所属组（套接字的所属用户总是启动服务器的用户）。与参数 unix_socket_permissions 组合，可以作为一种针对 Unix 域连接的附加访问控制机制。默认是一个空字符串，表示当前用户的默认组。这个参数只能在服务器启动时设置。

unix_socket_permissions (integer)

设置 Unix 域套接字的访问权限。Unix 域套接字使用通常的 Unix 文件系统权限集。参数值应是以 chmod 和 umask 系统调用所接受格式指定的数字权限模式。（要使用惯用的八进制格式，数字必须以 0（零）开头。）

默认权限是 0777，表示任何人都可以连接。合理的其他取值包括 0770（仅属主和所属组，另见 unix_socket_group）和 0700（仅属主）。（注意，对 Unix 域套接字而言，只有写权限起作用，因此设置或撤销读权限和执行权限没有意义。）

此访问控制机制独立于第 20 章中描述的机制。

此参数只能在服务器启动时设置。

`bonjour_name` (`string`)

指定 Bonjour 广播名。如果此参数设置为空字符串（默认值），则使用计算机名。如果服务器编译时未启用 Bonjour 支持，此参数会被忽略。此参数只能在服务器启动时设置。

`tcp_keepalives_idle` (`integer`)

在支持 `TCP_KEEPIDLE` 套接字选项的系统上，指定在一条空闲连接上两次发送保活（keepalive）报文之间间隔的秒数。值为零时使用系统默认值。如果不支持 `TCP_KEEPIDLE`，该参数必须为零。该参数对通过 Unix 域套接字建立的连接无效。

`tcp_keepalives_interval` (`integer`)

在支持 `TCP_KEEPINTVL` 套接字选项的系统上，指定在重传一个保活报文之前等待响应的秒数。值为零时使用系统默认值。如果不支持 `TCP_KEEPINTVL`，该参数必须为零。该参数对通过 Unix 域套接字建立的连接无效。

`tcp_keepalives_count` (`integer`)

在支持 `TCP_KEEPCNT` 套接字选项的系统上，指定在连接被视为已断开之前可以丢失多少个保活报文。值为零时使用系统默认值。如果不支持 `TCP_KEEPCNT`，该参数必须为零。该参数对通过 Unix 域套接字建立的连接无效。

17.3.2. 安全和认证

`authentication_timeout` (`integer`)

允许完成客户端认证的最长时间，以秒为单位。
如果一个客户端没有在这段时间里完成认证协议，服务器将关闭连接。
这样就避免了出问题的客户端无限制地占有一个连接。默认值是 1分钟（1 分钟）。这个参数只能在服务器命令行上或者在`postgresql.conf`文件中设置。

`ssl` (`boolean`)

启用SSL连接。使用前请阅读第 16.7 节。默认值是`off`。这个参数只能在服务器启动时设置。

`ssl_renegotiation_limit` (`integer`)

指定在一个SSL加密连接上可以流过多少数据之后将进行会话密钥的重新协商。
当可以检查大量流量时，重新协商能降低攻击者进行密码分析的机会，
但它也会带来很大的性能损失。该限制以发送和接收流量的总和来检查。
如果这个参数被设置为 0，则禁用。默认值为`512MB`重新协商。默认值为0。

> **注意**
>
> 2009 年 11 月之前的 SSL 库在使用 SSL 重新协商时不安全，原因是 SSL 协议中的一个漏洞。作为针对该漏洞的权宜修复，一些厂商发布的 SSL 库不具备重新协商能力。如果在客户端或服务器上正在使用这类库，应当禁用 SSL 重新协商。

`password_encryption` (`boolean`)

在CREATE USER或ALTER USER中指定密码，且既没有写`ENCRYPTED`也没有写`UNENCRYPTED`时，此参数决定密码是否被加密。默认值为`on`（加密密码）。

`krb_server_keyfile` (`string`)

设置服务器的Kerberos密钥文件的位置。详见第 20.2.3 节。这个参数只能在服务器启动时设置。

`krb_srvname` (`string`)

设置 Kerberos 服务名。详情见第 20.2.3 节。这个参数只能在服务器启动时设置。

`krb_server_hostname` (`string`)

设置服务主体的主机名部分。它与`krb_srvname`组合起来生成完整的服务主体，即`krb_srvname/krb_server_hostname@`REALM。

如果没有设置，默认为服务器的主机名。详见第 20.2.3 节。这个参数只能在服务器启动时设置。

`krb_caseins_users` (`boolean`)

设置是否应该以大小写不敏感的方式对待 Kerberos 和 GSSAPI 用户名。默认值是`off`（大小写敏感）。这个参数只能在服务器启动时设置。

`db_user_namespace` (`boolean`)

此参数启用各数据库独立的用户名。默认关闭。此参数只能在 postgresql.conf 文件中或在服务器命令行上设置。

如果启用此参数，你应以`username@dbname`的形式创建用户。
连接的客户端传入`username`时，会在用户名后附加`@`和数据库名，
然后由服务器查找此数据库专属的用户名。请注意，在 SQL 环境中创建名称包含`@`的用户时，需要用引号括起用户名。

启用此参数后，仍然可以创建普通的全局用户。只需在客户端指定用户名时附加`@`。
服务器查找用户名之前会去掉 `@`。

注意

此特性旨在作为找到完整解决方案之前的临时措施。届时将移除此选项。

17.4. 资源消耗

17.4.1. 内存

`shared_buffers` (`integer`)

设置数据库服务器使用的共享内存缓冲区数目。默认值通常是 1000，但如果你的内核设置不支持（在initdb期间确定），也可能更小。除非在构建服务器时选择了不同的`BLCKSZ`值，每个缓冲区为 8192 字节。这个设置必须至少为 16，并且至少是max_connections值的两

倍；不过，要获得良好的性能，通常需要明显高于最小值的设置。
生产安装推荐使用几千的数值。这个选项只能在服务器启动时设置。

增大这个参数可能导致PostgreSQL请求比操作系统默认配置允许的更多System V共享内存。必要时如何调整这些参数的信息参见第 16.4.1 节。

`temp_buffers`(`integer`)

设置每个数据库会话使用的临时缓冲区的最大数量。这些是会话本地的缓冲区，
仅用于访问临时表。默认值为 8 兆字节（8MB）。可以在单个会话内更改此设置，但必须在该会话首次使用临时表之前更改；此后尝试更改该值，对该会话不会产生影响。

会话会按需分配临时缓冲区，上限为 `temp_buffers`。对于实际不需要很多临时缓冲区的会话，将此参数设得较大时，开销仅为 `temp_buffers` 每增加一就多分配一个缓冲区描述符，约为 64 字节。不过，如果实际使用了某个缓冲区，还会为它额外消耗 8192 字节（一般而言为 `BLCKSZ` 字节）。

`max_prepared_transactions`(`integer`)

设置可同时处于“预备”状态的事务的最大数量（见PREPARE TRANSACTION）。将此参数设为零（默认值）会禁用预备事务功能。此参数只能在服务器启动时设置。

如果不打算使用预备事务，应将此参数设为零，以防意外创建预备事务。
如果使用预备事务，通常应将 `max_prepared_transactions` 设为不小于max_connections的值，以便每个会话都能有一个待处理的预备事务。

增加这个参数可能会导致PostgreSQL请求比你的操作系统默认配置所允许的更多的System V共享内存。如果有必要，关于如何调整那些参数的信息见第 16.4.1 节。

`work_mem`(`integer`)

指定内部排序操作和哈希表在写入临时磁盘文件之前可使用的内存量。默认值为 1 兆字节（1MB）。请注意，复杂查询可能并行执行多个排序或哈希操作，
每个操作在开始向临时文件写入数据之前，都可以使用此值指定的内存量。此外，
多个正在运行的会话也可能并发执行此类操作。因此，使用的总内存量可能是 `work_mem` 值的数倍；选择此值时必须考虑这一点。排序操作用于 `ORDER BY`、`DISTINCT` 和归并连接。
哈希表用于哈希连接、基于哈希的聚合、以及基于哈希的 `IN` 子查询处理。

`maintenance_work_mem`(`integer`)

指定维护操作（如 `VACUUM`、`CREATE INDEX` 和 `ALTER TABLE ADD FOREIGN KEY`）可使用的最大内存量。默认值为 16 兆字节（16MB）。由于一个数据库会话一次只能执行一个此类操作，而一个数据库系统通常也不会并发运行很多此类操作，
因此可以安全地将该值设得远大于 `work_mem`。更大的设置可能改善清理和恢复数据库转储的性能。

`max_stack_depth`(`integer`)

指定服务器执行栈的最大安全深度。
此参数的理想设置是内核强制的实际栈大小限制（由`ulimit -s`或本地等效设置确定），减去大约一兆字节的安全余量。需要安全余量是因为服务器中并非每个例程都检查栈深度，
而只在可能递归的关键例程（如表达式求值）中检查。
把这个参数设置得高于实际的内核限制，
将意味着失控的递归函数可能使单个后端进程崩溃。默认设置是 2048 KB（两兆字节），这个值较小，比较保守，不太可能导致崩溃。不过，对于执行复杂函数来说，它可能太小。

17.4.2. 空闲空间映射

这些参数控制共享空闲空间映射（FSM）的大小，它跟踪数据库中未使用空间的位置。空闲空间映射过小会导致数据库随时间推移消耗越来越多的磁盘空间，因为不在映射中的空闲空间无法被重新使用；当需要存储新数据时，PostgreSQL会转而向操作系统申请更多磁盘空间。数据库范围的`VACUUM VERBOSE`命令输出的最后几行有助于判断当前设置是否足够。如果当前设置过低，此类操作期间还会打印一条`NOTICE`消息。

增大这些参数可能导致PostgreSQL申请的System V共享内存超过操作系统默认配置所允许的数量。必要时如何调整这些参数，参见第 16.4.1 节。

`max_fsm_pages` (`integer`)

设置共享空闲空间映射中跟踪其空闲空间的磁盘页面的最大数量。每个页槽位消耗 6 字节共享内存。此设置必须至少为 16 * `max_fsm_relations`。默认值由 initdb 根据可用内存量选择，可在 20k 到 200k 页之间。此参数只能在服务器启动时设置。

`max_fsm_relations` (`integer`)

设置共享空闲空间映射中跟踪其空闲空间的关系（表和索引）的最大数量。每个槽位大约消耗 70 字节共享内存。默认值为一千个关系。此参数只能在服务器启动时设置。

17.4.3. 内核资源使用

`max_files_per_process` (`integer`)

设置每个服务器子进程允许同时打开的最大文件数量。默认值为一千个文件。如果内核强制实施了安全的每进程上限，就不必担心此设置。但在某些平台上（尤其是大多数 BSD 系统），内核允许单个进程打开的文件数量很大，如果很多进程都尝试打开这么多文件，就会远超系统实际能够支持的总量。如果遇到“Too many open files”（打开的文件过多）错误，可尝试减小此设置。此参数只能在服务器启动时设置。

`preload_libraries` (`string`)

这个变量指定在服务器启动时预装载的一个或多个共享库。可以选择为每个库调用一个无参数的初始化函数。要指定它，在库名后加上冒号和初始化函数的名称。例如`'$libdir/mylib:mylib_init'`将导致`mylib`被预装载并执行`mylib_init`。如果要装载多个库，用逗号分隔它们的名称。

如果找不到指定的库或初始化函数，服务器将启动失败。

PostgreSQL的过程语言库可以这样预装载，通常使用语法`'$libdir/plXXX:plXXX_init'`，其中`XXX`是`pgsql`、`perl`、`tcl`或`python`。

通过预装载共享库（并在适用时初始化它），可以避免首次使用该库时的库启动时间。不过，每个新服务器进程的启动时间可能会略有增加，即使该进程从未使用过这个库。因此，这个选项只推荐用于大多数会话都会用到的库。

17.4.4. 基于代价的清理延迟

执行VACUUM和ANALYZE命令期间，系统维护一个内部计数器，记录已执行的各种 I/O 操作的估算代价。当累计代价达到上限（由 `vacuum_cost_limit` 指定）时，执行该操作的进程会休眠一小段时间，时长由 `vacuum_cost_delay` 指定。随后重置计数器并继续执行。

此功能让管理员能够降低这些命令对并发数据库活动的 I/O 影响。在许多情况下，VACUUM 和 ANALYZE 等维护命令是否快速完成并不重要，但避免它们显著干扰系统执行其他数据库操作的能力通常很重要。基于代价的清理延迟为管理员提供了实现这一点的方法。

此功能默认禁用。要启用它，将 vacuum_cost_delay 变量设为非零值。

vacuum_cost_delay (integer)

超过代价上限后，进程将休眠的时长，单位为毫秒。默认值为零，表示禁用基于代价的清理延迟功能。正值会启用基于代价的清理。注意，在许多系统上，休眠延迟的有效分辨率为 10 毫秒；将 vacuum_cost_delay 设为不是 10 的倍数的值，可能与将它设为下一个更大的 10 的倍数效果相同。

vacuum_cost_page_hit (integer)

清理在共享缓冲区缓存中找到的缓冲区时所计入的估算代价。它表示锁定缓冲池、查找共享哈希表和扫描页内容的代价。默认值为 1。

vacuum_cost_page_miss (integer)

清理必须从磁盘读取的缓冲区时所计入的估算代价。它表示锁定缓冲池、查找共享哈希表、从磁盘读取所需数据块并扫描其内容所需的工作量。默认值为 10。

vacuum_cost_page_dirty (integer)

清理操作修改原本干净的数据块时所计入的估算代价。它表示再次将脏块刷盘所需的额外 I/O。默认值为 20。

vacuum_cost_limit (integer)

会使清理进程休眠的累计代价。默认值为 200。

注意

某些操作持有关键的锁，因此应尽快完成。这些操作期间不会发生基于代价的清理延迟，所以累计代价可能远超指定上限。为避免此时出现无益的长时间延迟，实际延迟按 vacuum_cost_delay * accumulated_balance / vacuum_cost_limit 计算，但最大不超过 vacuum_cost_delay * 4。

17.4.5. 后台写入器

从PostgreSQL 8.0 开始，有一个独立的服务器进程，称为后台写入器，它唯一的职责就是发出对“脏”共享缓冲区的写操作。这样设计的意图是，处理用户查询的服务器进程应当很少或从不需要等待写操作的发生，因为后台写入器会完成它。这种安排也降低了与检查点相关的性能损失。后台写入器会把脏页持续不断地少量刷写到磁盘，因此在检查点时刻到来时只需要强制写出很少的页面，而不是像以前那样在每个检查点发生一次脏缓冲区写风暴。不过，总体 I/O 负载会有净增加，因为一个反复变脏的页面以前可能每个检查点间隔只写出一次，而后台写入器可能在同一间隔内把它写出多次。在大多数情况下，持续的低压负载优于周期性的尖峰，但本小节讨论的参数可用于根据本地需求调整这一行为。

bgwriter_delay (integer)

指定后台写入器各轮活动之间的延迟。每一轮中，写入器会对一定数量的脏缓冲区发出写操作（由下面的参数控制），然后休眠bgwriter_

delay毫秒，再重复此过程。默认值为 200。注意，在许多系统上，休眠延迟的有效分辨率为 10 毫秒；将bgwriter_delay设为不是 10 的倍数的值，可能与将它设为下一个更大的 10 的倍数效果相同。这个选项只能在服务器启动时或在postgresql.conf文件中设置。

bgwriter_lru_percent (floating point)

为了降低服务器进程需要自己发出写操作的概率，
后台写入器会尝试写出很快就会被回收的缓冲区。每一轮中，它最多检查bgwriter_lru_percent比例的最接近被回收的缓冲区，并写出其中的脏缓冲区。默认值为 1.0（这是共享缓冲区总数的一个百分比）。这个选项只能在服务器启动时或在postgresql.conf文件中设置。

bgwriter_lru_maxpages (integer)

每一轮中，作为扫描即将被回收的缓冲区的结果，写出的缓冲区不会超过这个数目。
默认值为 5。这个选项只能在服务器启动时或在postgresql.conf文件中设置。

bgwriter_all_percent (floating point)

为了减少检查点时所需的工作量，后台写入器还会对整个缓冲池做循环扫描，
写出发现的脏缓冲区。每一轮中，它为此最多检查bgwriter_all_percent比例的缓冲区。默认值为 0.333（这是共享缓冲区总数的一个百分比）。在默认的bgwriter_delay设置下，这允许整个共享缓冲池大约每分钟被扫描一次。
这个选项只能在服务器启动时或在postgresql.conf文件中设置。

bgwriter_all_maxpages (integer)

每一轮中，作为扫描整个缓冲池的结果，写出的缓冲区不会超过这个数目。
（如果达到这个上限，扫描会停止，并在下一轮从下一个缓冲区继续。）默认值为 5。这个选项只能在服务器启动时或在postgresql.conf文件中设置。

较小的bgwriter_all_percent和bgwriter_all_maxpages
值会降低后台写入器造成的额外 I/O 负载，但会把更多工作留到检查点时刻完成。要减少检查点时的负载尖峰，可以增大这两个值。类似地，较小的bgwriter_lru_percent和bgwriter_lru_maxpages值会降低后台写入器造成的额外 I/O 负载，但会使服务器进程更有可能必须自己发出写操作，从而延迟交互式查询。要完全禁用后台写操作，可以把两个maxpages
值和/或两个percent值都设置为零。

17.5. 预写式日志

关于 WAL 与检查点调优的细节，另见第 26.3 节。

17.5.1. 设置

fsync (boolean)

如果打开这个参数，PostgreSQL服务器将尝试确保更新被物理地写入到磁盘，做法是发出fsync()系统调用或者使用多种等价的方法（见wal_sync_method）。
这保证了数据库集簇在一次操作系统或者硬件崩溃后能恢复到一个一致的状态。

不过，使用 fsync 会带来性能损失：提交一个事务时，PostgreSQL 必须等待操作系统把预写日志刷写到磁盘。当 fsync 被禁用时，操作系统可以尽其所能地缓冲、排序和延迟写操作。这可以显著提高性能。然而，如果系统崩溃，
最近提交的若干事务的结果可能部分或全部丢失。在最坏的情况下，

可能发生不可恢复的数据损坏。（数据库软件自身的崩溃在这里不是风险因素。只有操作系统层面的崩溃才会带来损坏的风险。）

由于存在这些风险，`fsync` 没有一个普遍正确的设置。有些管理员总是禁用 `fsync`，另一些管理员只在初始批量装载数据时才关闭它（此时一旦出错，有明确的重启点），还有一些管理员始终启用 `fsync`。默认值是启用 `fsync`，以获得最大的可靠性。如果你信任你的操作系统、你的硬件以及你的电力公司（或你的电池备份），可以考虑禁用 `fsync`。

fsync 只能在`postgresql.conf`文件中或在服务器命令行上设置。如果你关闭这个参数，请也考虑关闭full_page_writes。

`wal_sync_method`(`string`)

用于将 WAL 更新强制写入磁盘的方法。如果`fsync`关闭，此设置就没有作用，因为 WAL 文件更新根本不会被强制写入磁盘。可选值为：

- open_datasync（用`open()`选项`O_DSYNC`写 WAL 文件）
- `fdatasync`（在每次提交时调用`fdatasync()`）
- `fsync`（在每次提交时调用`fsync()`）
- `fsync_writethrough`（在每次提交时调用`fsync()`，强制穿透任何磁盘写缓存）
- `open_sync`（用`open()`选项`O_SYNC`写 WAL 文件）

不是在所有平台上都能使用所有这些选择。默认值是列表中第一个被平台支持的那个，不过`fdatasync`是 Linux 中的默认值。这个选项可以在服务器启动时或在`postgresql.conf`文件中设置。

`full_page_writes`(`boolean`)

当这个选项打开时，PostgreSQL服务器在检查点之后第一次修改磁盘页时，会把整个磁盘页的内容写入 WAL。这是必需的，因为在操作系统崩溃期间正在进行的页写入可能只完成了一部分，导致磁盘上出现新旧数据混合的页。WAL 中通常存储的行级变更数据不足以在崩溃后恢复中完全还原这样的页。存储完整页映像可以保证页能被正确还原，但代价是必须写入 WAL 的数据量增加。（因为 WAL 重放总是从检查点开始，所以只需在检查点后每页的第一次变更时这样做就足够了。因此，降低全页写代价的一种办法是增大检查点间隔参数。）

目前这个参数被忽略（被视为总是`on`），因为关闭它可能导致即使没有发生硬件或操作系统级错误也无法从崩溃中恢复。这将在未来某个版本中修复，或者该参数将被完全移除。

`wal_buffers`(`integer`)

用于 WAL 数据的共享内存中的内存量。默认值是 64 千字节（64kB）。这个设置只需要大到足以容纳一个典型事务产生的 WAL 数据量，因为数据在每次事务提交时都会被写出到磁盘。这个参数只能在服务器启动时设置。

增加这个参数可能会导致PostgreSQL请求比你的操作系统默认配置所允许的更多的System V共享内存。如果有必要，关于如何调整那些参数的信息见第 16.4.1 节。

`commit_delay`(`integer`)

把一个提交记录写入 WAL 缓冲区和把缓冲区刷出到磁盘之间的时间延迟，以微秒计。如果系统负载足够高，使得额外的事务在给定间隔内变为准备好提交，

一个非零的延迟可以允许多个事务只用一次fsync()系统调用提交。
但如果没有其他事务变为准备好提交，这个延迟就纯属浪费。因此，
只有当服务器进程写完它的提交记录的那一刻，至少有commit_siblings
个其他事务处于活动状态时，才会执行该延迟。默认值为零（无延迟）。

commit_siblings(integer)

执行commit_delay延迟前要求的并发活动事务的最小数目。
大一些的值会导致在延迟间隔期间更可能有至少另外一个事务准备好提交。
默认值是五个事务。

17.5.2. 检查点

checkpoint_segments(integer)

自动 WAL 检查点之间日志文件段的最大数量（每个段通常为 16 兆字节）。默认值为三个段。增加这个参数会增加崩溃恢复所需的时间。此参数只能在postgresql.conf
文件中或服务器命令行上设置。

checkpoint_timeout(integer)

自动 WAL 检查点之间的最长时间，以秒为单位。合理的范围在 30 秒到 1 小时之间。默认是 5 分钟（5 分钟）。增加这个参数的值可能会增加崩溃恢复所需的时间。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

checkpoint_warning(integer)

如果由填充检查点段文件导致的检查点发生得比这个秒数更接近，
就向服务器日志写一条消息（这建议应该增大checkpoint_segments）。默认值是 30 秒（30 秒）。零会禁用警告。此参数只能在postgresql.conf文件中或服务器命令行上设置。

17.5.3. 归档

archive_command(string)

用来归档 WAL 文件系列中一个已完成段的 shell 命令。如果它是空字符串（默认值），WAL 归档被禁用。字符串中的任何%p都会被替换为要归档文件的路径名，任何%f只被替换为文件名。（该路径名相对于服务器的工作目录，即集簇的数据目录。）使用%%
可以在命令中嵌入真正的%字符。更多信息见第 23.3.1 节。这个选项只能在服务器启动时或在postgresql.conf文件中设置。

当且仅当命令成功时它返回退出状态零这一点很重要。例如：

```
archive_command = 'cp -i %p /mnt/server/archivedir/%f </dev/null'
```

17.6. 查询规划

17.6.1. 规划器方法配置

这些配置参数提供了一种较为粗糙的方法，用于影响查询优化器所选择的查询计划。
如果优化器为某个特定查询选择的默认计划并不理想，
一种临时解决方案是使用这些配置参数强制优化器选择另一种计划。不过，
永久关闭其中某项设置很少是个好主意。
改善优化器所选计划质量的更好办法包括调整规划器代价常数（见规划器代价常量）、

手工运行ANALYZE、增加default_statistics_target配置参数的值，以及使用 `ALTER TABLE SET STATISTICS` 增加为特定列收集的统计信息量。

`enable_bitmapscan` (`boolean`)

启用或禁用查询规划器对位图扫描计划类型的使用。默认值为 `on`。

`enable_hashagg` (`boolean`)

启用或禁用查询规划器对哈希聚合计划类型的使用。默认值为 `on`。

`enable_hashjoin` (`boolean`)

启用或禁用查询规划器对哈希连接计划类型的使用。默认值为 `on`。

`enable_indexscan` (`boolean`)

启用或禁用查询规划器对索引扫描计划类型的使用。默认值为 `on`。

`enable_mergejoin` (`boolean`)

启用或禁用查询规划器对归并连接计划类型的使用。默认值为 `on`。

`enable_nestloop` (`boolean`)

允许或禁止查询规划器使用嵌套循环连接计划。它不可能完全禁止嵌套循环连接，但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是`on`。

`enable_seqscan` (`boolean`)

允许或禁止查询规划器使用顺序扫描计划类型。它不可能完全禁止顺序扫描，但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是`on`。

`enable_sort` (`boolean`)

允许或禁止查询规划器使用显式排序步骤。它不可能完全禁止显式排序，但是关闭这个变量将使得规划器尽可能优先使用其他方法。默认值是`on`。

`enable_tidscan` (`boolean`)

允许或禁止查询规划器使用TID扫描计划类型。默认值是`on`。

17.6.2. 规划器代价常量

注意

遗憾的是，没有明确的方法可以确定下文出现的这组“代价”变量的理想值。建议你进行试验并分享你的发现。

`effective_cache_size` (`floating point`)

设置规划器对单次索引扫描可用的有效磁盘缓存大小的假设。这个值会被纳入使用索引的代价估计；较高的值使索引扫描更可能被使用，较低的值使顺序扫描更可能被使用。设置这个参数时，应同时考虑PostgreSQL的共享缓冲区，以及内核磁盘缓存中将用于PostgreSQL数据文件的部分。

还应考虑预计使用不同索引的并发查询数，因为它们必须共享可用空间。这个参数不影响PostgreSQL分配的共享内存大小，也不会预留内核磁盘缓存；它仅用于估计。
该值以磁盘页为单位，通常每页为 8192 字节。默认值是 1000。

`random_page_cost` (`floating point`)

设置规划器对非顺序获取磁盘页代价的估计。它以顺序获取一页代价的倍数来衡量。
较高的值使规划器更倾向于使用顺序扫描，较低的值使规划器更倾向于使用索引扫描。
默认值是 4。

`cpu_tuple_cost` (`floating point`)

设置规划器对一次查询中处理每一行的代价估计。默认值是 0.01。

`cpu_index_tuple_cost` (`floating point`)

设置规划器对一次索引扫描中处理每一个索引项的代价估计。默认值是 0.005。

`cpu_operator_cost` (`floating point`)

设置规划器对`WHERE`子句中处理每个运算符所需代价的估计。
它以顺序获取一页代价的分数来衡量。默认值是 0.0025。

17.6.3. 遗传查询优化器

`geqo` (`boolean`)

允许或禁止遗传查询优化，它是一种试图不做穷举搜索来进行查询规划的算法。
默认是启用。`geqo_threshold`变量提供了一种更细粒度的方式，
可以对某些类别的查询禁用 GEQO。

`geqo_threshold` (`integer`)

只有当涉及的`FROM`项数量至少有这么多个的时候，才使用遗传查询优化（注意一个`JOIN`只被计为一个`FROM`项）。默认值是 12。对于更简单的查询，通常会使用普通的穷举搜索规划器，但是对于有很多表的查询穷举搜索会花很长时间，
通常比执行一个次优的计划带来的惩罚值还要长。因此，在查询尺寸上的一个阈值是管理GEQO 使用的一种方便的方法。

`geqo_effort` (`integer`)

控制 GEQO 中规划时间和查询计划质量之间的权衡。此变量必须是 1 到 10 之间的整数。默认值是 5。较大的值会增加查询规划所用的时间，但也会提高选择高效查询计划的可能性。

`geqo_effort`实际并不直接做任何事情；它只是被用来计算其他影响 GEQO 行为的变量（如下所述）的默认值。如果你愿意，你可以手工设置其他参数。

`geqo_pool_size` (`integer`)

控制 GEQO 使用的池尺寸，它就是遗传种群中的个体数目。它必须至少为 2，且有用的值通常在 100 到 1000 之间。如果它被设置为零（默认设置）则会基于`geqo_effort`和查询中表的数量选择一个合适的值。

`geqo_generations` (`integer`)

控制 GEQO 使用的代数，即算法的迭代次数。它必须至少为 1，通常有用的值与池大小处于相同范围。如果设置为零（默认设置），则根据`geqo_pool_size`选择合适的值。

`geqo_selection_bias`(`floating point`)

控制 GEQO 使用的选择偏好。选择偏好是种群中的选择压力。值可以是 1.5 到 2.0 之间，后者是默认值。

17.6.4. 其他规划器选项

`default_statistics_target`(`integer`)

为没有通过`ALTER TABLE SET STATISTICS`设置列相关目标的表列设置默认统计目标。更大的值增加了需要做`ANALYZE`的时间，但是可能会改善规划器的估计质量。默认值是 10。有关PostgreSQL查询规划器使用的统计信息的更多内容，请参考第 13.2 节。

`constraint_exclusion`(`boolean`)

允许或禁止查询规划器使用表约束来优化查询。默认值是`off`。

当这个参数为`on`时，规划器会比较查询条件与表的`CHECK`约束，
并且忽略扫描那些条件与约束相矛盾的表。例如：

```
CREATE TABLE parent(key integer, ...);
CREATE TABLE child1000(check (key between 1000 and 1999)) INHERITS(
parent);
CREATE TABLE child2000(check (key between 2000 and 2999)) INHERITS(
parent);
...
SELECT * FROM parent WHERE key = 2400;
```

在启用约束排除时，这个`SELECT`将完全不会扫描`child1000`，从而提高性能。

目前，`constraint_exclusion`默认是禁用的，因为约束检查的代价相对较高，
而且在许多情况下不会带来任何节省。
建议只在确实使用为利用该特性而设计的分区表时才启用它。

关于在PostgreSQL中使用约束排除和分区的更多信息请参阅第 5.9 节。

`from_collapse_limit`(`integer`)

如果生成的`FROM`列表不超过这么多项，规划器将把子查询融合到上层查询。
较小的值可以减少规划时间，但是可能会生成较差的查询计划。默认值是 8。通常应让它小于 GEQO geqo_threshold。

`join_collapse_limit`(`integer`)

只要最终得到的列表总项数不超过这个数目，规划器就会把显式内`JOIN`结构改写为`FROM`项的列表。在PostgreSQL 7.4 之前，通过`JOIN`结构指定的连接永远不会被查询规划器重新排序。查询规划器后来得到了改进，使得以这种形式书写的内连接可以被重新排序；
这个配置参数控制这种重排进行的程度。

> **注意**
>
> 目前，通过`JOIN`结构指定的外连接的顺序永远不会被查询规划器调整；因此，`join_collapse_limit`对这一行为没有影响。
> 规划器可能会在PostgreSQL的未来版本中改进为能对某些类别的外连接重新排序。

默认情况下，这个变量被设置为与`from_collapse_limit`相同的值，这适合大多数用途。把它设置为 1 可以禁止内`JOIN`的任何重新排序。因此，查询中指定的显式连接顺序就是关系实际连接的顺序。查询规划器并不总是选择最优的连接顺序；
高级用户可以选择临时把这个变量设置为 1，然后显式指定他们想要的连接顺序。把这个变量设置为 1 的另一个后果是，查询规划器的行为会更像PostgreSQL 7.3 的查询规划器，一些用户可能出于向后兼容的考虑觉得这很有用。

把这个变量设置为 1 到`from_collapse_limit`之间的值，可能有助于在规划时间和所选计划的质量之间做权衡（较高的值产生较好的计划）。

17.7. 错误报告和日志

17.7.1. 日志记录到哪里

`log_destination`(`string`)

PostgreSQL支持多种把日志记录到不同目的地的方法。在 Windows 上还支持eventlog。默认是只记录到stderr。这个选项以逗号分隔的列表形式设置所需的日志目的地。
有效值为stderr和syslog。这个选项可以在服务器启动时或在`postgresql.conf`配置文件中设置。

`redirect_stderr`(`boolean`)

这个选项允许捕获发送到stderr的消息并重定向到日志文件中。
这个选项与记录到stderr相结合，通常比记录到syslog更有用，
因为某些类型的消息可能不会出现在
syslog输出中（一个常见的例子是动态链接器的失败消息）。
这个选项只能在服务器启动时设置。

`log_directory`(`string`)

当`redirect_stderr`被启用时，这个参数决定日志文件将被在哪个目录下创建。
它可以被指定为一个绝对路径，也可以被指定为一个相对于集簇数据目录的相对路径。
这个参数只能在`postgresql.conf`文件中或在服务器命令行上设置。

`log_filename`(`string`)

当启用了`redirect_stderr`时，这个选项设置所创建日志文件的文件名。该值会被当作strftime模式处理，因此可以使用`%`转义来指定随时间变化的文件名。如果没有`%`转义，PostgreSQL将追加新日志文件开始记录时的纪元时间。例如，如果`log_filename`是`server_log`，那么对于 MST 2004 年 8 月 29 日星期日 19:02:33 开始的日志，选定的文件名将是`server_log.1093827753`。这个选项可以在服务器启动时或在`postgresql.conf`配置文件中设置。

`log_rotation_age`(`integer`)

当`redirect_stderr`被启用时，这个参数决定单个日志文件的最长使用时间。
经过指定的分钟数之后，将创建一个新的日志文件。
将这个参数设置为零将禁用基于时间的新日志文件创建。这个参数只能在`postgresql.conf`文件中或在服务器命令行上设置。

`log_rotation_size`(`integer`)

当`redirect_stderr`被启用时，这个参数决定一个个体日志文件的最大尺寸。
当指定千字节数的数据被写入一个日志文件后，将创建一个新的日志文件。

设置为零时将禁用基于大小创建新的日志文件。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

log_truncate_on_rotation (boolean)

当redirect_stderr被启用时，这个参数将导致PostgreSQL截断（覆盖而不是追加）任何已有的同名日志文件。不过，截断只在一个新文件由于基于时间的轮转被打开时发生，在服务器启动或基于尺寸的轮转时不会发生。如果被关闭，在所有情况下以前存在的文件将被追加。例如，使用这个设置和一个类似postgresql-%H.log的log_filename将导致产生 24 个每小时的日志文件，并且循环地覆盖它们。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

示例：保留7天的日志，每天一个日志文件，命名为server_log.Mon，server_log.Tue，等等，并自动用本周的日志覆盖上周的日志，将log_filename设置为server_log.%a，将log_truncate_on_rotation设置为on，将log_rotation_age设置为1440。

示例：要保留 24 小时的日志，每个小时一个日志文件，如果日志文件尺寸超过 1GB，也会提前轮转。可以这样做：将log_filename设置为server_log.%H%M、将log_truncate_on_rotation设置为on、将log_rotation_age设置为60并且将log_rotation_size设置为1000000。在log_filename中包含%M，可让按文件大小触发的轮转选用与整点初始文件名不同的新文件名。

syslog_facility (string)

当启用了向syslog记录时，这个参数决定要使用的syslog“设施”。你可以在LOCAL0、LOCAL1、LOCAL2、LOCAL3、LOCAL4、LOCAL5、LOCAL6、LOCAL7中选择，默认值是LOCAL0。还请参阅系统的syslog守护进程的文档。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。除非服务器编译时带有syslog支持，此参数不可用。

syslog_ident (string)

当启用了向syslog记录时，这个参数决定用来标识syslog中的PostgreSQL消息的程序名。默认值是postgres。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。除非服务器编译时带有syslog支持，此参数不可用。

17.7.2. 什么时候记录日志

client_min_messages (string)

控制将哪些消息级别发送给客户端。有效值是DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、LOG、NOTICE、WARNING、ERROR、FATAL、和PANIC。每个级别都包括其后的所有级别。级别越靠后，被发送的消息越少。默认值是NOTICE。注意，LOG在这里的排名与在 log_min_messages中不同。

log_min_messages (string)

控制将哪些消息级别写入服务器日志。有效值是DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、INFO、NOTICE、WARNING、ERROR、LOG、FATAL 和 PANIC。每个级别都包含它后面的所有级别。级别越靠后，发送到日志的消息越少。默认值是NOTICE。注意，LOG在这里的排名与在 client_min_messages中不同。只有超级用户可以更改此设置。

log_error_verbosity (string)

控制在服务器日志中记录的每条消息的详细程度。有效值为TERSE、DEFAULT和VERBOSE，它们依次在显示的消息中增加更多字段。只有超级用户能更改这个设置。

log_min_error_statement (string)

控制在服务器日志中记录哪些导致错误条件的SQL语句。
对于达到指定严重级别或更高级别的消息，其日志条目中会包含当前 SQL语句。有效值是DEBUG5、DEBUG4、DEBUG3、DEBUG2、DEBUG1、INFO、NOTICE、WARNING、ERROR、FATAL 和 PANIC。默认值是ERROR，表示导致错误、日志消息、致命错误或恐慌的语句将被记录。要有效地关闭失败语句的记录，可将此参数设置为PANIC。
只有超级用户可以更改此设置。

log_min_duration_statement (integer)

如果一条已完成语句的运行时间至少达到指定毫秒数，就记录其持续时间。
将此值设置为零会打印所有语句的持续时间。负一（默认值）禁用语句持续时间记录。
例如，如果设置为250，则会记录所有运行 250ms 或更长时间的 SQL 语句。启用此参数有助于发现应用中未优化的查询。此设置独立于log_statement和log_duration。
只有超级用户能更改这个设置。

silent_mode (boolean)

静默地运行服务器。如果设置了此参数，
服务器将自动在后台运行并脱离任何控制终端（效果与postmaster的-S选项相同）。
服务器的标准输出和标准错误被重定向到/dev/null，
因此发送给它们的所有消息都会丢失。除非选择了syslog日志或者启用了redirect_stderr，否则不鼓励使用此参数，因为它会使错误消息无法被看到。
这个参数只能在服务器启动时设置。

下面列出了这些设置中使用的各种消息严重级别：

DEBUG[1-5]

提供供开发者使用的信息。

INFO

提供用户隐式请求的信息，例如VACUUM VERBOSE的输出。

NOTICE

提供可能对用户有用的信息，例如长标识符截断提示。

WARNING

提供可能出现的问题的警告，例如在事务块外COMMIT。

ERROR

报告一个导致当前命令中断的错误。

LOG

报告管理员可能感兴趣的信息，例如检查点活动。

FATAL

报告一个导致当前会话中断的错误。

PANIC

报告一个导致所有数据库会话中断的错误。

17.7.3. 记录哪些内容

debug_print_parse (boolean)
debug_print_rewritten (boolean)
debug_print_plan (boolean)
debug_pretty_print (boolean)

这些参数将会让多种调试输出被发出。对每个被执行的查询，它们打印结果分析树、查询重写器输出或执行计划。debug_pretty_print会缩进这些显示，产生可读性更强但长得多的输出格式。要把这种输出真正发送到客户端或服务器日志，必须把client_min_messages或log_min_messages分别设为DEBUG1或更低。这些参数默认是关闭的。

log_connections (boolean)

记录每一次到服务器的连接尝试，以及客户端认证的成功完成。某些客户端程序（如psql）在判断是否需要密码时会尝试连接两次，因此可能会看到“connection received”消息重复。此参数只能在postgresql.conf文件中或在服务器命令行上设置。默认为关闭。

log_disconnections (boolean)

在会话终止时向服务器日志输出一行与log_connections类似的内容，并且包含会话的持续时间。默认情况下这是关闭的。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

log_duration (boolean)

导致每个满足log_statement的已完成语句的持续时间被记录。使用这个选项时，如果你没有使用syslog，建议用log_line_prefix记录 PID 或会话 ID，这样你就能通过进程 ID 或会话 ID 把语句消息与后面的持续时间消息关联起来。默认值是off。只有超级用户可以更改这个设置。

log_line_prefix (string)

这是个printf风格的字符串，在每条日志行的开头输出。默认值是一个空字符串。每个被识别的转义都按下述说明被替换——其他任何看起来像转义的内容都会被忽略。其他字符被直接复制到日志行。有些转义只被会话进程识别，不适用于 postmaster 等后台进程。Syslog会产生自己的时间戳和进程 ID 信息，因此如果你在使用syslog，可能不会想使用这些转义。这个选项只能在服务器启动时或在postgresql.conf配置文件中设置。

转义	效果	仅会话
%u	用户名	是
%d	数据库名	是
%r	远程主机名或 IP 地址，以及远程端口	是
%h	远程主机名或 IP 地址	是
%p	进程 ID	否
%t	时间戳（无毫秒）	否
%m	时间戳（含毫秒）	否
%i	命令标签：即产生该日志行的命令。	是
%c	会话 ID：每个会话的唯一标识符。它由两个 4 字节的十六进制	是

转义	效果	仅会话
	数（不带前导零）以点号分隔组成。这两个数字分别是会话启动时间和进程 ID，因此它也可以作为打印这些项的省空间方式。	
%l	每个进程的日志行号，从 1 开始	否
%s	会话启动时间戳	是
%x	事务 ID	是
%q	不产生任何输出，但告诉非会话进程在字符串的此处停止。会话进程忽略它。	否
%%	字面量%	否

log_statement (string)

控制哪些 SQL 语句被记录。有效值是 none（关闭）、ddl、mod和 all。ddl记录所有数据定义语句，例如CREATE、ALTER和 DROP语句。mod记录所有ddl语句，外加INSERT、UPDATE、DELETE、TRUNCATE和COPY FROM等数据修改语句。如果PREPARE和 EXPLAIN ANALYZE包含的命令属于合适的类型，它们也会被记录。

默认值为none。只有超级用户能更改这个设置。

注意

EXECUTE语句不被认为是ddl或mod语句。产生语法错误的语句不会被记录。将log_min_error_statement设置为error可以记录这种语句。

当函数以PL/pgSQL服务器端语言定义时，
该函数执行的查询只在函数于某个特定会话中首次被调用时才会被记录。这是因为PL/pgSQL会为函数中的 SQL 语句所产生的查询计划保留一个缓存。

log_hostname (boolean)

默认情况下，连接日志消息只显示连接主机的 IP 地址。打开这个参数将导致也记录主机名。注意根据你的主机名解析设置，这可能会导致不可忽视的性能开销。这个参数只能在postgresql.conf文件中或在服务器命令行上设置。

17.8. 运行时统计

17.8.1. 统计监控

log_statement_stats (boolean)
log_parser_stats (boolean)
log_planner_stats (boolean)
log_executor_stats (boolean)

对于每个查询，将各自模块的性能统计输出到服务器日志中。
这是一个简单的性能分析工具，类似于Unix getrusage()操作系统功能。log_statement_stats报告整个语句的统计信息，而其他选项报告每个模块的统计信息。log_statement_stats不能与任何单独模块选项一起启用。所有这些选项默认情况下都是禁用的。
只有超级用户才能更改这些设置。

17.8.2. 查询和索引统计信息收集器

stats_start_collector (boolean)

控制服务器是否应启动统计收集子进程。这个选项默认为 on，但如果你知道自己对收集统计信息不感兴趣，可以把它关掉。这个选项只能在服务器启动时设置。

stats_command_string (boolean)

启用对每个会话当前正在执行的命令及其开始执行时间的统计信息收集。
这个选项默认关闭。注意，即使启用，这些信息也不是对所有用户可见，
只对超级用户和被报告会话的属主用户可见；因此它不会构成安全风险。这些数据可以通过pg_stat_activity系统视图访问；更多信息参见第 24 章。

stats_block_level (boolean)

启用对数据库活动的块级统计信息的收集。这个选项默认被禁用。如果启用了这个选项，
产生的数据可以通过pg_stat和pg_statio系列系统视图访问；更多信息参见第 24 章。

stats_row_level (boolean)

启用对数据库活动的行级统计信息的收集。这个选项默认被禁用。如果启用了这个选项，
产生的数据可以通过pg_stat和pg_statio系列系统视图访问；更多信息参见第 24 章。

stats_reset_on_server_start (boolean)

如果为 on，收集到的统计信息会在服务器每次重启时清零。如果为 off，统计信息将跨服务器重启累积。默认值为off。这个选项只能在服务器启动时设置。

17.9. 自动清理

这些设置控制autovacuum特性的行为。更多信息见第 22.1.4 节。

autovacuum (boolean)

控制服务器是否应启动 autovacuum 子进程。默认情况下它是关闭的。stats_start_collector和stats_row_level也必须打开它才会启动。
这个选项可以在服务器启动时或在postgresql.conf文件中设置。

autovacuum_naptime (integer)

指定自动清理在任意给定数据库上各次运行之间的最小间隔。
在每一轮中后台进程检查数据库并根据需要为数据库中的表发出VACUUM和ANALYZE命令。
延迟以秒为单位。默认值为1分钟（1 分钟）。该参数只能在postgresql.conf文件或在服务器命令行上设置。

autovacuum_vacuum_threshold (integer)

指定能在一个表上触发VACUUM的被更新或被删除元组的最小数量。默认值为1000。
该选项可以在服务器启动时或在postgresql.conf文件中设置。对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

autovacuum_analyze_threshold (integer)

指定能在一个表上触发ANALYZE的被插入、被更新或被删除元组的最小数量。
默认值为500。该选项可以在服务器启动时或在postgresql.conf文件中设置。
对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

autovacuum_vacuum_scale_factor(floating point)

指定一个表尺寸的分数，在决定是否触发VACUUM时将它加到autovacuum_vacuum_threshold上。默认值为0.4。该选项可以在服务器启动时或在postgresql.conf文件中设置。对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

autovacuum_analyze_scale_factor(floating point)

指定一个表尺寸的分数，在决定是否触发ANALYZE时将它加到autovacuum_analyze_threshold上。默认值为0.2。该选项可以在服务器启动时或在postgresql.conf文件中设置。对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

autovacuum_vacuum_cost_delay(integer)

指定用于自动VACUUM操作中的代价延迟值。如果指定-1（默认值），则使用常规的vacuum_cost_delay值。对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

autovacuum_vacuum_cost_limit(integer)

指定用于自动VACUUM操作中的代价限制值。如果指定-1（默认值），则使用常规的vacuum_cost_limit值。对个别表，可以通过在pg_autovacuum中的条目来覆盖该设置。

17.10. 客户端连接默认值

17.10.1. 语句行为

search_path(string)

这个变量指定当一个对象（表、数据类型、函数等）被用一个无模式限定的简单名称引用时，搜索该对象时的模式顺序。当在不同模式中有同名对象时，将使用第一个在搜索路径中被找到的对象。一个不属于搜索路径中任何一个模式的对象只能通过用限定名（带点号）指定包含它的模式来引用。

search_path的值必须是一个逗号分隔的模式名列表。如果列表项之一是特殊值$user，则具有SESSION_USER返回的名字的模式将取代它（如果有这样一个模式。如果没有，$user会被忽略）。

系统目录模式pg_catalog总是被搜索，不管它是否在搜索路径中被提及。如果它在路径中被提及，那么它将被按照路径指定的顺序搜索。如果pg_catalog不在路径中，则它将在任何路径项之前被搜索。

同样，当前会话的临时表模式pg_temp_*nnn*（如果存在）也总是被搜索。可以通过使用别名pg_temp把它显式列在路径中。如果它没有被列在路径中，那么它会被最先搜索（甚至先于pg_catalog）。不过，临时模式只会用于关系（表、视图、序列等）和数据类型名称的搜索，永远不会用于函数或操作符名称的搜索。

当对象创建时没有指定一个特定目标模式，它们将被放置在搜索路径中列出的第一个模式中。如果搜索路径为空将报告一个错误。

这个参数的默认值是'$user, public'（其中第二部分在没有名为public的模式时会被忽略）。这支持共享使用一个数据库（其中没有用户拥有私有模式，所有人都共享使用public）、每用户私有模式以及两者的组合。通过全局或者按用户修改默认搜索路径设置，可以得到其他效果。

搜索路径的当前有效值可以通过SQL函数current_schemas()检查。它和检查search_path的值不太一样，因为current_schemas显示出现在search_path中的项如何被解析。

有关模式处理的更多信息，见第 5.7 节。

default_tablespace (string)

这个变量指定当一个CREATE命令没有显式指定一个表空间时，创建对象（表和索引）的默认表空间。

该值要么是一个表空间的名字，要么是一个指定使用当前数据库默认表空间的空字符串。如果该值和任何现有表空间的名字都不匹配，PostgreSQL将自动使用当前数据库的默认表空间。

有关表空间的更多的信息，请见第 19.6 节。

check_function_bodies (boolean)

这个参数通常为打开。当设置为off时，它禁用CREATE FUNCTION期间对函数体字符串的验证。在从转储恢复函数定义时，偶尔禁用验证有助于避免前向引用之类的问题。

default_transaction_isolation (string)

每个 SQL 事务都有一个隔离级别，可以是“读未提交”、“读已提交”、“可重复读”或者“可串行化”。这个参数控制每个新事务的默认隔离级别。默认是“读已提交”。

更多信息请参阅第 12 章和SET TRANSACTION。

default_transaction_read_only (boolean)

一个只读的 SQL 事务不能修改非临时表。这个参数控制每个新事务的默认只读状态。默认是off（读/写）。

更多信息请参考SET TRANSACTION。

statement_timeout (integer)

中止任何使用了超过指定毫秒数的语句，从客户端命令到达服务器时开始计时。如果log_min_error_statement被设置为ERROR或更低，语句如果超时也会被记录。一个零值（默认）将禁用超时。

17.10.2. 区域设置和格式化

DateStyle (string)

设置日期和时间值的显示格式，以及解释有歧义的日期输入值的规则。由于历史原因，这个变量包含两个独立的部分：输出格式声明（ISO、Postgres、SQL或German）、输入/输出的年/月/日顺序（DMY、MDY或YMD）。这些可以被独立设置或者一起设置。关键字Euro和European是DMY的同义词；关键字US、NonEuro和NonEuropean是MDY的同义词。详见第 8.5 节。内置默认值是ISO, MDY。

timezone (string)

设置用于显示和解释时间戳的时区。默认值是unknown，表示使用系统环境指定的时区。更多信息见第 8.5 节。

australian_timezones (boolean)

如果设为 on，ACST、CST、EST和SAT将被解释为澳大利亚时区，而不是北美/南美时区和星期六。默认值为off。

extra_float_digits (integer)

此参数调整浮点值显示的位数，包括float4、float8和几何数据类型。
参数值会加到标准位数（按情况使用FLT_DIG或DBL_DIG）上。此值最高可以设置为 3，以包含部分有效的数位；这对于转储需要精确还原的浮点数据特别有用。也可以设置为负数，以抑制不需要的数位。

client_encoding (string)

设置客户端编码（字符集）。默认使用数据库编码。

lc_messages (string)

设置消息显示的语言。可接受的值是系统相关的；详见第 21.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

在一些系统上，这个区域设置类别并不存在。仍然可以设置这个变量，
只是不会有任何效果。同样，所期望语言的翻译消息也可能不存在。在这种情况下，
你将仍然继续看到英文消息。

lc_monetary (string)

设置用于格式化货币量的区域设置，例如用to_char函数族。可接受的值是系统相关的；
详见第 21.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

lc_numeric (string)

设置用于格式化数字的区域设置，例如用to_char函数族。可接受的值是系统相关的；
详见第 21.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

lc_time (string)

设置用于格式化日期和时间值的区域设置。（目前，这个设置不做任何事情，
但将来可能会。）可接受的值是系统相关的；详见第 21.1 节。如果这个变量被设置为空字符串（默认），那么该值将以一种系统相关的方式从服务器的执行环境中继承。

17.10.3. 其他默认值

explain_pretty_print (boolean)

决定EXPLAIN VERBOSE显示详细查询树转储时使用缩进格式还是非缩进格式。默认值是on。

dynamic_library_path (string)

如果需要打开一个可以动态装载的模块并且在CREATE FUNCTION或LOAD命令中指定的文件名没有目录部分（即名字中不包含斜线），那么系统将搜索这个路径以查找所需的文件。

参数dynamic_library_path的值必须是由冒号（Windows上为分号）
分隔的绝对目录路径列表。如果某个列表元素以特殊字符串$libdir开头，

则会使用编译时确定的PostgreSQL软件包的库目录来替换`$libdir`；
该目录是标准PostgreSQL发行版所提供模块的安装位置。（使用`pg_config --pkglibdir`可以找出此目录的名称。）例如：

```
dynamic_library_path = '/usr/local/lib/postgresql:/home/my_project/lib:$libdir'
```

或者，在 Windows 环境中：

```
dynamic_library_path = 'C:\tools\postgresql;H:\my_project\lib;$libdir'
```

这个参数的默认值是`'$libdir'`。如果该值被设置为一个空字符串，则关闭自动路径搜索。

这个参数可以由超级用户在运行时更改，
但以这种方式进行的设置只会持续到客户端连接结束，
因此这种方法应该保留用于开发目的。推荐设置此参数的方法是在`postgresql.conf`配置文件中。

17.11. 锁管理

`deadlock_timeout` (`integer`)

指定在检查是否发生死锁之前等待锁的毫秒数。检查死锁相对昂贵，
因此服务器不会每次等待锁时都运行它。我们乐观地假设在生产应用程序中死锁并不常见，
所以在检查死锁之前只是等待一段时间。
增加此值会减少在不必要的死锁检查中浪费的时间，但会减慢实际死锁错误的报告速度。
默认值为一秒（1 秒），这可能是你在实践中想要的最小值。在负载较重的服务器上，你可能希望提高它。理想情况下，设置应超过你的典型事务时间，
以提高在等待者决定检查死锁之前释放锁的几率。只有超级用户能更改这个设置。

`max_locks_per_transaction` (`integer`)

共享锁表跟踪在`max_locks_per_transaction` * (max_connections + max_prepared_transactions）个对象（如表）上的锁。因此，在任何一个时刻，只有不超过这么多个可区分对象能够被锁住。这个参数控制为每个事务分配的对象锁的平均数量。
只要所有事务的锁都能放入锁表中，单个事务就可以锁住更多对象。
这不是能被锁住的行数，那个值是没有限制的。默认值 64 已经被历史证明是足够的，但是如果你有需要在一个事务中使用很多不同表的查询（例如查询一个有很多子表的父表），
你可能需要提高这个值。这个参数只能在服务器启动时设置。

增加这个参数可能会导致PostgreSQL请求比你的操作系统默认配置所允许的更多的System V共享内存。如果有必要，关于如何调整那些参数的信息见第 16.4.1 节。

17.12. 版本和平台兼容性

17.12.1. 以前的 PostgreSQL 版本

`add_missing_from` (`boolean`)

打开时，如果查询引用的表尚未出现在`FROM`子句中，会自动把它添加进去。这一行为不符合SQL 标准，许多人也不喜欢它，因为它可能掩盖错误（例如本应引用表的别名却引用了表本身）。默认值是`off`。为了与 8.1 之前默认允许这一行为的PostgreSQL版本兼容，可以启用这个变量。

注意，即使启用了此变量，查询每引用一个隐式FROM项，仍会发出一条警告消息。建议用户更新其应用，把查询引用的所有表都加到查询的FROM子句（对DELETE则是USING子句）中，从而不依赖这一行为。

regex_flavor (string)

正则表达式的“风格”可以设置为 advanced、extended 或 basic。默认值是advanced。extended 设置可用于与 PostgreSQL 7.4 之前的版本精确向后兼容。详见第 9.7.3.1 节。

sql_inheritance (boolean)

这个设置控制不带修饰的表引用是否被认为包括继承子表。默认值为 on，即子表被包括在内（因此，默认假定有 * 后缀）。如果设为 off，子表不被包括（因此，假定有ONLY前缀）。SQL 标准要求包括子表，因此off设置不符合标准，但提供它是为了与 7.1 之前版本的PostgreSQL 兼容。更多信息见第 5.8 节。

backslash_quote (string)

控制字符串字面量中的引号是否能够用\'表示。首选的、符合 SQL 标准的表示引号的方法是将其双写（''），但PostgreSQL历史上也接受\'。不过，使用\'会带来安全风险，因为在某些客户端字符集编码中，存在最后一字节在数值上等价于 ASCII \的多字节字符。如果客户端代码转义不当，就可能发生 SQL 注入攻击。让服务器拒绝引号看起来是被反斜线转义的查询，可以避免这种风险。backslash_quote的允许值是on（总是允许\'）、off（总是拒绝）和safe_encoding（仅当客户端编码不允许在多字节字符中出现 ASCII \时允许）。默认设置是safe_encoding。

default_with_oids (boolean)

控制在既未指定WITH OIDS也未指定WITHOUT OIDS时，CREATE TABLE和CREATE TABLE AS创建的新表是否包含 OID 列。它还决定SELECT INTO创建的表是否包含 OID。在PostgreSQL 8.1 中，default_with_oids默认被禁用；在此之前的PostgreSQL版本中，它默认是打开的。

在用户表中使用 OID 已被认为过时，因此大多数安装应让这个变量保持禁用状态。需要为特定表使用 OID 的应用应在创建该表时指定WITH OIDS。可以为了兼容不遵循这一行为的旧应用而启用这个变量。

escape_string_warning (boolean)

打开时，如果在普通字符串字面量中（'...'语法）出现了一个反斜线（\）并且 standard_conforming_strings为关闭，那么就会发出一个警告。默认值是off。希望使用反斜线作为转义符的应用应该被修改来使用转义字符串语法（E'...'），因为为了SQL 兼容性，普通字符串的默认行为将在将来的版本中改变。这个变量可以被启用来帮助检测将会受损的应用。

未来PostgreSQL版本的默认行为将会改变，因此建议尽快迁移应用。

17.12.2. 平台和客户端兼容性

transform_null_equals (boolean)

当打开时，形如*expr* = NULL（或NULL = *expr*）的表达式将被当做*expr* IS NULL，也就是说，如果*expr*计算结果为空值则返回真，否则返回假。正确的 SQL 标准兼容的*expr* = NULL行为总是返回空值（未知）。因此这个参数默认为off。

不过，在Microsoft Access里的过滤表单生成的查询似乎使用*expr* = NULL来测试空值，因此，如果你使用这个接口访问数据库，你可能想把这个选项打开。因为*expr* = NULL形式的表达式总是返回空值（使用 SQL 标准解释），它们不是非常有用并且在普通应用中也不常见，因此这个选项实际上没有什么危害。
但是新用户常常对涉及空值的表达式语义感到困惑，因此这个选项默认为关闭。

请注意这个选项只影响= NULL形式，而不影响其它比较操作符或者其它与一些涉及等值操作符的表达式在计算上等效的其他表达式（例如IN）。因此，
这个选项不能普遍修复错误的程序写法。

相关信息请见第 9.2 节。

17.13. 预置选项

下列“参数”是只读的，在编译或安装 PostgreSQL 时确定。因此，它们未列入示例 postgresql.conf 文件。这些选项报告 PostgreSQL 行为的各个方面，某些应用（特别是管理前端）可能对此感兴趣。

block_size (integer)

报告一个磁盘块的大小。它由编译服务器时BLCKSZ的值确定。默认值是 8192 字节。有些配置变量的含义（例如shared_buffers）会被block_size影响。详见第 17.4 节。

integer_datetimes (boolean)

报告PostgreSQL是否在构建时启用了对 64 位整数日期时间的支持。它通过在构建PostgreSQL时配置--enable-integer-datetimes来设置。默认值是off。

lc_collate (string)

报告对文本数据进行排序所使用的区域设置。更多信息参见第 21.1 节。该值在数据库集簇初始化时确定。

lc_ctype (string)

报告决定字符分类的区域设置。更多信息参见第 21.1 节。该值在数据库集簇初始化时确定。通常它与lc_collate相同，但特殊应用可能将它设置为不同的值。

max_function_args (integer)

报告函数参数的最大数量。它由编译服务器时的FUNC_MAX_ARGS值决定。默认值是 100 个参数。

max_identifier_length (integer)

报告标识符的最大长度。它由编译服务器时的NAMEDATALEN值减一决定。NAMEDATALEN的默认值是 64；因此max_identifier_length的默认值是 63 字节，在使用多字节编码时，这可能不足 63 个字符。

max_index_keys (integer)

报告索引键的最大数目。它由编译服务器时的INDEX_MAX_KEYS值决定。默认值是 32 个键。

server_encoding (string)

报告数据库的编码（字符集）。这是在数据库被创建时决定的。通常，客户端只需要关心client_encoding的值。

server_version (string)

报告服务器的版本号。它是由编译服务器时的PG_VERSION值决定的。

standard_conforming_strings (boolean)

控制普通字符串字面量（'...'）是否按照 SQL 标准把反斜线当普通文本。当前默认值为off，即PostgreSQL保持其把反斜线当作转义字符的历史行为。为了改进与 SQL 标准的兼容性，默认值将在将来的版本中改为on。
应用可以检查这个参数来判断字符串字面量如何被处理。
这个参数的存在也可以被当做转义字符串语法（E'...'）被支持的标志。
如果一个应用希望反斜线被当做转义字符，应该使用转义字符串语法。

17.14. 自定义选项

这个特性允许附加模块（例如过程语言）向PostgreSQL添加系统通常不认识的参数。
这样便能以标准方式配置扩展模块。

custom_variable_classes (string)

这个变量以逗号分隔列表的形式指定用于自定义变量的一个或多个类名。
自定义变量是一个PostgreSQL本身通常不知道、但被某个附加模块使用的变量。
这种变量的名称必须由一个类名、一个点号和一个变量名组成。custom_variable_classes指定一个特定安装中使用的所有类名。此参数只能在postgresql.conf文件中或在服务器命令行上设置。

在postgresql.conf中设置自定义变量的困难在于，该文件必须在附加模块被载入之前读取，
因此自定义变量通常会被当作未知变量而拒绝。当设置了custom_variable_classes时，
服务器将接受每个指定类中任意变量的定义。这些变量会被当作占位符，
在定义它们的模块被载入之前没有任何功能。当一个特定类的模块被载入时，
它会为其类名加入正确的变量定义，根据那些定义转换任何占位符值，
并对该类中剩下的任何无法识别的占位符发出警告。

下面是一个在使用自定义变量时postgresql.conf可能包含内容的例子：

```
custom_variable_classes = 'plr,plperl'
plr.path = '/usr/lib/R'
plperl.use_strict = true
plruby.use_strict = true        # generates error: unknown class name
```

17.15. 开发者选项

下列参数用于PostgreSQL源代码的开发工作，在某些情况下也用于辅助恢复严重损坏的数据库。
没有理由在生产数据库中使用它们。因此，它们被排除在示例postgresql.conf文件之外。
请注意，许多参数需要特殊的源代码编译标志才能起作用。

debug_assertions (boolean)

打开各种断言检查。这是一种调试辅助手段。如果你遇到了奇怪的问题或崩溃，
可以打开它，因为它可能暴露出编程错误。要使用这个参数，在构建PostgreSQL时必须定义宏USE_ASSERT_CHECKING（通过configure选项--enable-cassert完成）。注意，如果PostgreSQL构建时启用了断言，debug_assertions默认为on。

`pre_auth_delay` (`integer`)

如果非零，则在新服务器进程刚刚派生之后、执行认证过程之前，延迟指定的秒数。这旨在给开发者一个机会，用调试器附加到服务器进程上，跟踪认证过程中的异常行为。这个选项只能在 postgresql.conf 文件中设置。

`trace_notify` (`boolean`)

为`LISTEN`和`NOTIFY`命令生成大量调试输出。client_min_messages或log_min_messages必须是`DEBUG1`或者更低才能把这种输出分别发送到客户端或者服务器日志。

`trace_sort` (`boolean`)

如果开启，发出排序操作期间资源使用情况的信息。此参数只有在编译PostgreSQL时定义了`TRACE_SORT`宏的情况下才可用（不过，`TRACE_SORT`目前是默认定义的）。

`trace_locks` (`boolean`)
`trace_lwlocks` (`boolean`)
`trace_userlocks` (`boolean`)
`trace_lock_oidmin` (`boolean`)
`trace_lock_table` (`boolean`)
`debug_deadlocks` (`boolean`)
`log_btree_build_stats` (`boolean`)

其他各种代码跟踪和调试选项。

`wal_debug` (`boolean`)

打开时，输出与 WAL 相关的调试信息。只有在编译 PostgreSQL 时定义了 `WAL_DEBUG` 宏，该参数才可用。

`zero_damaged_pages` (`boolean`)

检测到一个损坏的页面头通常会导致PostgreSQL报告一个错误并中止当前命令。把`zero_damaged_pages`设置为 on 会让系统改为报告一个警告，把损坏的页面清零，然后继续处理。这种行为将破坏数据，即损坏页面上的所有行。但它允许你越过错误，从表中可能存在的任何未损坏页面中检索行。因此，如果由于硬件或软件错误发生了损坏，它对于恢复数据是有用的。一般来说，只有在你对从表的损坏页面中恢复数据已经不抱希望之后，才应该打开它。默认设置为`off`，并且只有超级用户能更改它。

17.16. 短选项

为了方便，某些参数还有单字母的命令行选项开关可用。它们在表 17.1中描述。

表 17.1. 短选项对照

短选项	等价形式
`-B` *x*	`shared_buffers =` *x*
`-d` *x*	`log_min_messages = DEBUG`*x*
`-F`	`fsync = off`
`-h` *x*	`listen_addresses =` *x*

短选项	等价形式
-i	listen_addresses = '*'
-k *x*	unix_socket_directory = *x*
-l	ssl = on
-N *x*	max_connections = *x*
-p *x*	port = *x*
-fb、-fh、-fi、-fm、-fn、-fs、-ft[a]	enable_bitmapscan = off、enable_hashjoin = off、enable_indexscan = off、enable_mergejoin = off、enable_nestloop = off、enable_seqscan = off、enable_tidscan = off
-s [a]	log_statement_stats = on
-S *x* [a]	work_mem = *x*
-tpa、-tpl、-te[a]	log_parser_stats = on、log_planner_stats = on、log_executor_stats = on

[a] 由于历史原因，这些选项必须通过-o postmaster选项传递给每个服务器进程，例如

```
$ postmaster -o '-S 1024 -s'
```

或者像上文说明的那样，从客户端通过PGOPTIONS传递。

第 18 章 数据库角色和权限

PostgreSQL使用角色这一概念来管理数据库访问权限。角色可以被看作数据库用户，
也可以被看作一组数据库用户，这取决于该角色的设置方式。
角色可以拥有数据库对象（例如表），并且可以把这些对象上的权限赋予其他角色，
以控制谁能访问哪些对象。此外，还可以把一个角色中的成员资格授予另一个角色，
从而允许成员角色使用它作为成员的那个角色被授予的权限。

角色这一概念涵盖了“用户”和“组”这两个概念。在 PostgreSQL 8.1 之前的版本中，用户和组是两类不同的实体，但现在只有角色。任意角色都可以充当用户、组，或者同时充当两者。

本章描述如何创建和管理角色，并介绍权限系统。
关于各种数据库对象类型和权限效果的更多信息，见第 5 章。

18.1. 数据库角色

数据库角色在概念上与操作系统用户完全分离。在实践中，
维持两者之间的对应关系可能比较方便，但这并非必需。
数据库角色在整个数据库集簇安装中是全局的（而非每个数据库各自独立）。要创建角色，
请使用CREATE ROLE SQL 命令：

```
CREATE ROLE name;
```

name 遵循 SQL 标识符规则：要么不加修饰且不含特殊字符，要么用双引号括起。（在实践中，通常还需要为该命令添加其他选项，例如 `LOGIN`。下文会给出更多细节。）要移除现有角色，请使用类似的DROP ROLE命令：

```
DROP ROLE name;
```

为了方便，提供了createuser和dropuser程序，作为这些 SQL 命令的包装器，可以从 shell 命令行调用：

```
createuser name
dropuser name
```

要确定现有角色的集合，请查看 `pg_roles` 系统目录，例如

```
SELECT rolname FROM pg_roles;
```

使用psql程序的 `\du` 元命令也可以列出现有角色。

为了完成数据库系统的初始引导，一个刚初始化的系统总会包含一个预定义角色。
该角色总是一个“超级用户”，并且除非另行指定了不同名称，否则它的名称与用`initdb`
初始化数据库集簇的操作系统用户相同。这个角色通常命名为`postgres`。要创建更多角色，
必须先以这个初始角色建立连接。

每个到数据库服务器的连接都是以某个特定角色的名义建立的，
而这个角色决定在该连接上发出命令时的初始访问权限。某个数据库连接要使用的角色名，
由发起连接请求的客户端以应用特定的方式指明。例如，`psql`
程序使用`-U`命令行选项来指明要以哪个角色连接。
很多应用默认假定使用当前操作系统用户的名字（包括`createuser`和`psql`）。因此，
在角色名和操作系统用户名之间维护对应关系通常很方便。

某个客户端连接能够以哪些数据库角色进行连接，由客户端认证设置决定，如第 20 章中所述。（因此，客户端并不一定限于只能以与其操作系统用户同名的角色连接，
正如一个人的登录名不必与其真实姓名一致。）
由于角色身份决定了已连接客户端可用的权限集合，所以在设置多用户环境时，
仔细配置这一点非常重要。

18.2. 角色属性

数据库角色可以具有若干定义其权限的属性，这些属性还会与客户端认证系统交互。

登录权限

只有具有LOGIN属性的角色才能用作数据库连接的初始角色名。具有LOGIN
属性的角色可以被视为与“数据库用户”等同。要创建具有登录权限的角色，可以使用：

```
CREATE ROLE name LOGIN;
CREATE USER name;
```

（CREATE USER等价于CREATE ROLE，只是CREATE USER默认假定LOGIN，而CREATE ROLE不是。）

超级用户状态

数据库超级用户绕过所有权限检查。这是一种危险的权限，不应草率使用；
最好以非超级用户角色完成大部分工作。要创建一个新的数据库超级用户，使用CREATE ROLE *name* SUPERUSER。必须以已经是超级用户的角色来执行此操作。

创建数据库

必须显式授予角色创建数据库的权限（超级用户除外，因为超级用户绕过所有权限检查）。
要创建这样的角色，使用CREATE ROLE *name* CREATEDB。

创建角色

必须显式授予角色创建更多角色的权限（超级用户除外）。要创建这样的角色，使用CREATE ROLE *name* CREATEROLE。具有CREATEROLE权限的角色还可以修改和删除其他角色，以及授予或撤销其中的成员资格。但是，要创建、修改、
删除超级用户角色或更改其成员资格，需要超级用户身份；CREATEROLE
并不足以做到这一点。

密码

只有当客户端认证方法要求用户在连接数据库时提供密码时，密码才有意义。password、md5和crypt认证方法使用密码。数据库密码与操作系统密码是分开的。
在创建角色时用CREATE ROLE *name* PASSWORD '*string*'.

角色的属性在创建之后可以用ALTER ROLE修改。细节见CREATE ROLE和ALTER ROLE命令的参考页。

提示

一种良好做法是创建具有CREATEDB和CREATEROLE权限但不是超级用户的角色，
然后使用此角色执行数据库和角色的所有日常管理工作。

> 这种方法可以避免在实际上并不需要超级用户权限的任务中以超级用户身份操作所带来的风险。

角色还可以为第 17 章中描述的许多运行时配置设置指定角色特定默认值。例如，如果出于某种原因你希望每次连接时都禁用索引扫描（提示：这不是个好主意），你可以使用：

```
ALTER ROLE myname SET enable_indexscan TO off;
```

这会保存该设置（但不会立即生效）。在该角色后续建立的连接中，它看起来就像在会话开始之前执行了`SET enable_indexscan TO off;`一样。你仍然可以在会话期间更改该设置；它只会作为默认值。要移除角色特定默认设置，使用`ALTER ROLE *rolename* RESET *varname*;`。注意，附加到没有`LOGIN`权限的角色上的角色特定默认值几乎没有用，因为它们永远不会生效。

18.3. 权限

创建对象时会为它指定一个所有者。所有者通常是执行创建语句的角色。对大多数类型的对象来说，初始状态是只有所有者（或超级用户）能对对象做任何操作。要让其他角色使用它，必须授予权限。权限有多种不同的类别：`SELECT`、`INSERT`、`UPDATE`、`DELETE`、`RULE`、`REFERENCES`、`TRIGGER`、`CREATE`、`TEMPORARY`、`EXECUTE`和 `USAGE`。关于PostgreSQL支持的不同权限类型的更多信息，见GRANT参考页。

使用`GRANT`命令来授予权限。例如，如果`joe`是一个已有角色，`accounts`是一个已有表，更新该表的权限可以这样授予：

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

特殊名称`PUBLIC`可用于向系统中的每个角色授予权限。在具体权限的位置写`ALL`，表示授予适用于该对象的所有权限。

要撤销权限，使用名称贴切的REVOKE命令：

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

对象所有者的特殊权限（即修改或销毁对象的权利）总是隐含在所有者身份中，不能被授予或撤销。但所有者可以选择撤销自己的普通权限，例如使表对他自己和其他人都变为只读。

可以用与对象类型相对应的`ALTER`命令把对象指派给新的所有者。超级用户总是可以这样做；普通角色只有在其既是对象的当前所有者（或所有者角色的成员）、又是新所有者角色的成员时才能这样做。

18.4. 角色成员资格

为了简化权限管理，把用户分组通常很方便：这样，权限就可以整体授予给一个组，或者从整个组中撤销。在PostgreSQL中，这通过创建一个表示该组的角色，然后把该组角色中的成员资格授予各个用户角色来实现。

要建立一个组角色，首先创建该角色：

```
CREATE ROLE name;
```

通常，作为组使用的角色不会有`LOGIN`属性，不过如果你愿意，也可以设置它。

组角色存在后，就可以使用GRANT和REVOKE命令添加和移除成员：

```
GRANT group_role TO role1, ... ;
REVOKE group_role FROM role1, ... ;
```

也可以把成员资格授予其他组角色（因为组角色与非组角色实际上没有任何区别）。
数据库不允许建立循环成员关系。此外，也不允许将角色的成员资格授予 PUBLIC。

角色的成员可以通过两种方式使用该组角色的权限。首先，组的每个成员都可以显式执行SET ROLE，以暂时“成为”该组角色。在这种状态下，数据库会话可以使用组角色的权限，
而不是原始登录角色的权限；创建的任何数据库对象也被视为由组角色而非登录角色拥有。
其次，具有 INHERIT 属性的成员角色可以自动使用其所属角色的权限。例如，假定我们已经执行了：

```
CREATE ROLE joe LOGIN INHERIT;
CREATE ROLE admin NOINHERIT;
CREATE ROLE wheel NOINHERIT;
GRANT admin TO joe;
GRANT wheel TO admin;
```

以角色 joe 连接后，数据库会话立即可以使用直接授予 joe 的权限，以及授予 admin 的所有权限，因为 joe “继承” admin 的权限。不过，授予 wheel 的权限不可用，因为即使 joe 间接属于 wheel，该成员资格也是通过 admin 获得的，而后者具有 NOINHERIT 属性。执行以下命令后：

```
SET ROLE admin;
```

会话将只能使用授予 admin 的权限，而不能使用授予 joe 的权限。执行以下命令后：

```
SET ROLE wheel;
```

会话将只能使用授予 wheel 的权限，而不能使用授予 joe 或 admin 的权限。使用以下任一命令可以恢复原始的权限状态：

```
SET ROLE joe;
SET ROLE NONE;
RESET ROLE;
```

注意

SET ROLE命令总是允许切换到原始登录角色直接或间接所属的任何角色。因此，在上例中，没有必要先成为admin再成为wheel。

注意

在 SQL 标准中，用户与角色有明确区分，用户不会自动继承权限，而角色会。行为在PostgreSQL中可以通过给用作 SQL 角色的角色赋予INHERIT属性、给用作 SQL 用户的角色赋予NOINHERIT属性来获得。不过，为了向后兼容 8.1 之前的版本（那时用户总是可以使用其所属组被授予的权限），PostgreSQL默认为所有角色赋予INHERIT属性。

角色属性LOGIN、SUPERUSER、CREATEDB和CREATEROLE可以被视为特殊权限，
但它们从不会像数据库对象上的普通权限那样被继承。要使用其中之一，必须实际执行SET ROLE

切换到拥有该属性的特定角色。继续上面的例子，我们完全可以选择把CREATEDB和CREATEROLE赋予admin角色。这样，一个以joe角色连接的会话并不会立即拥有这些权限，只有在执行SET ROLE admin之后才会拥有。

要销毁组角色，使用DROP ROLE：

```
DROP ROLE name;
```

组角色中的任何成员资格都会被自动撤销（但成员角色本身不受其他影响）。但要注意，组角色拥有的任何对象必须先被删除或重新指派给其他所有者；授予组角色的任何权限也必须先被撤销。

18.5. 函数和触发器

函数和触发器允许用户把代码插入到后端服务器中，而其他用户可能会无意间执行这些代码。因此这两种机制都使得用户能比较容易地以“特洛伊木马”的方式坑害他人。唯一真正的保护是严格控制谁可以定义函数。

函数在后端服务器进程中运行，并具有数据库服务器守护进程的操作系统权限。如果函数所用的编程语言允许不受检查的内存访问，就有可能更改服务器的内部数据结构。因此，这类函数可以做很多别的事情，其中就包括绕过任何系统访问控制。允许这种访问的函数语言被认为是“不受信任的”，而PostgreSQL只允许超级用户创建用这类语言编写的函数。

第 19 章 管理数据库

每个正在运行的PostgreSQL服务器实例都管理着一个或多个数据库。因此，数据库是组织SQL对象（“数据库对象”）的最高层级。本章描述数据库的属性，以及如何创建、管理和删除数据库。

19.1. 概述

数据库是一个命名的SQL对象（“数据库对象”）集合。通常，每个数据库对象（表、函数等）都属于且只属于一个数据库。（不过也有少量系统目录，例如`pg_database`，属于整个集簇，可以从集簇中的每个数据库访问。）更准确地说，数据库是模式的集合，而模式中包含表、函数等。因此完整的层次结构是：服务器、数据库、模式、表（或其他类型的对象，如函数）。

当连接到数据库服务器时，客户端必须在连接请求中指定要连接的数据库名。一个连接无法访问多个数据库。不过，一个应用可以不受限制地打开到同一数据库或其他数据库的多个连接。数据库在物理上是相互隔离的，访问控制在连接级别管理。因此，如果一个PostgreSQL服务器实例要容纳彼此应当分离、并且在大多数情况下互不知晓的项目或用户，建议把它们放到不同的数据库中。如果项目或用户是相互关联的，并且应该能够使用彼此的资源，则应当把它们放在同一个数据库中，但可能使用不同的模式。模式是纯逻辑结构，谁能访问什么由权限系统管理。更多关于管理模式的信息见第 5.7 节。

数据库使用`CREATE DATABASE`命令创建（见第 19.2 节），使用`DROP DATABASE`命令删除（见第 19.5 节）。要确定现有数据库的集合，可以检查系统目录`pg_database`，例如：

```
SELECT datname FROM pg_database;
```

psql程序的`\l`元命令和`-l`命令行选项也可用于列出现有数据库。

> **注意**
>
> SQL标准将数据库称为“目录”，但在实践中两者没有区别。

19.2. 创建数据库

要创建数据库，PostgreSQL服务器必须已经启动并正在运行（见第 16.3 节）。

数据库使用 SQL 命令CREATE DATABASE创建：

```
CREATE DATABASE name;
```

其中*name*遵循SQL标识符的通常规则。当前角色自动成为新数据库的所有者。数据库的所有者有权在以后删除它（这也会删除其中的所有对象，即使它们有不同的所有者）。

创建数据库是一项受限操作。关于如何授予该权限，见第 18.2 节。

由于执行`CREATE DATABASE`命令时必须先连接到数据库服务器，因此还会有一个问题：某个站点上的第一个数据库是如何创建的？第一个数据库总是由`initdb`命令在初始化数据存储区时创建的（见第 16.2 节）。这个数据库叫作`postgres`。因此，要创建第一个“普通”数据库，可以连接到`postgres`。

在初始化数据库集簇时，还会由initdb创建第二个数据库template1。
每当在集簇中创建一个新数据库时，新数据库本质上都是通过克隆template1得到的。
这意味着你在template1中所做的任何更改，都会传播到后来创建的所有数据库中。因此，把template1用于实际工作并不明智；但如果运用得当，这一特性也可以带来便利。
更多细节见第 19.3 节。

为了方便，还可以从 shell 中执行一个程序来创建新数据库，即createdb。

```
createdb dbname
```

createdb并没有什么特殊之处。它连接到postgres数据库并发出CREATE DATABASE命令，和上面描述的完全一样。关于调用细节见createdb参考页。注意，不带任何参数的createdb会创建一个与当前用户名同名的数据库，而这未必是你想要的。

> **注意**
>
> 第 20 章包含关于如何限制谁能连接到给定数据库的信息。

有时你会想为其他人创建数据库，并让他们成为新数据库的拥有者，
以便他们自行配置和管理它。要实现这一点，可使用下列命令之一：

```
CREATE DATABASE dbname OWNER rolename;
```

这是在 SQL 环境中使用；或者：

```
createdb -O rolename dbname
```

这是在 shell 中使用。只有超级用户才允许为其他人（即为一个你不是其成员的角色）创建数据库。

19.3. 模板数据库

CREATE DATABASE实际上是通过复制一个现有数据库来创建新数据库。默认情况下，它复制名为template1的标准系统数据库。因此，该数据库就是创建新数据库时使用的“模板”。
如果你向template1中添加对象，这些对象也会被复制到以后创建的用户数据库中。
这种行为允许对数据库中的标准对象集合进行站点本地修改。例如，如果你将过程语言PL/pgSQL安装到template1中，那么以后创建用户数据库时无需额外操作，它就会自动可用。

还有第二个标准系统数据库，名为template0。该数据库包含与template1
初始内容相同的数据，也就是说，只包含你的PostgreSQL版本预定义的标准对象。在initdb
之后，template0就不应再被修改。通过指示CREATE DATABASE复制template0而不是template1，你可以创建一个“纯净的”用户数据库，它不包含template1
中的任何站点本地附加对象。这在恢复pg_dump
转储时特别方便：转储脚本应当恢复到一个纯净的数据库中，
以确保重建出被转储数据库的正确内容，而不会与后来可能加入template1的对象发生冲突。

要通过复制template0创建数据库，可在 SQL 环境中使用：

```
CREATE DATABASE dbname TEMPLATE template0;
```

或在 shell 中使用：

```
createdb -T template0 dbname
```

可以创建额外的模板数据库，实际上也可以通过将集簇中任意数据库的名称指定为CREATE DATABASE的模板来复制它。不过，必须理解的是，这（至少目前）并不是一个通用的“COPY DATABASE”功能。特别要强调的是，在复制操作进行期间，源数据库必须是空闲的（没有正在进行的数据更改事务）。CREATE DATABASE会检查在操作开始时没有其他会话（它自己除外）连接到源数据库，但这并不能保证在复制过程中不会发生更改，否则将导致复制出来的数据库不一致。因此，我们建议把用作模板的数据库当作只读对待。

对每个数据库，pg_database中都有两个有用的标志列：datistemplate和datallowconn。datistemplate可设置为表明某个数据库将作为CREATE DATABASE的模板。如果设置了该标志，任何拥有CREATEDB权限的用户都可以克隆该数据库；如果未设置，则只有超级用户和数据库拥有者才能克隆它。如果datallowconn为假，则不允许建立到该数据库的新连接（但仅仅把该标志设为假并不会杀死已有会话）。为了防止对它的修改，template0数据库通常被标记为datallowconn = false。template0和template1始终都应标记为datistemplate = true。

在准备好模板数据库或对其进行任何更改之后，最好在该数据库中执行 VACUUM FREEZE。如果在同一数据库中没有其他打开的事务时执行此操作，就可以保证该数据库中的所有行都已被“冻结”，不会受到事务 ID 回卷问题的影响。这对于将被设置 datallowconn为假的数据库尤其重要，因为在这种数据库中将无法进行例行的维护性VACUUM。更多信息参见第 22.1.3 节。

注意

template1和template0除了template1这个名字是CREATE DATABASE默认源数据库名这一点之外，并无任何特殊地位。例如，可以删除template1，然后再从template0重新创建它，而不会有任何不良后果。如果有人不小心在template1里加入了许多杂项对象，这样做也许是可取的。

初始化数据库集簇时也会创建postgres数据库。这个数据库被用作用户和应用连接时的默认数据库。它只是template1的一个副本，因此在必要时也可以删除并重建。

19.4. 数据库配置

正如第 17 章所述，PostgreSQL服务器提供了大量运行时配置变量。你可以为其中许多设置指定数据库特定的默认值。

例如，如果你因为某种原因想在某个数据库中禁用GEQO优化器，通常要么必须对所有数据库都禁用它，要么必须确保每个连入的客户端都会执行SET geqo TO off。若要让这个设置在某个特定数据库中成为默认值，可以执行如下命令：

```
ALTER DATABASE mydb SET geqo TO off;
```

这会保存该设置（但不会立即生效）。以后连接到该数据库时，它看起来就像是在会话开始前刚执行过SET geqo TO off;一样。注意，用户仍然可以在自己的会话中更改该设置；它只会作为默认值。要撤销这样的设置，可使用ALTER DATABASE *dbname* RESET *varname*;。

19.5. 删除数据库

数据库使用DROP DATABASE命令删除：

```
DROP DATABASE name;
```

只有数据库拥有者或超级用户可以删除数据库。删除数据库会移除其中包含的全部对象。数据库的删除无法撤销。

当你连接到目标数据库时，不能执行DROP DATABASE命令。不过，你可以连接到任何其他数据库，包括template1数据库。若要删除某个集簇中的最后一个用户数据库，template1将是唯一的选择。

为了方便，还有一个 shell 程序可用于删除数据库，即dropdb。

```
dropdb dbname
```

（与createdb不同，默认操作并不是删除与当前用户名同名的数据库。）

19.6. 表空间

PostgreSQL中的表空间允许数据库管理员在文件系统中定义存放表示数据库对象的文件的位置。表空间一旦创建，在创建数据库对象时就可以通过名称引用它。

通过使用表空间，管理员可以控制PostgreSQL安装的磁盘布局。这至少有两方面的用途。首先，如果初始化集簇所在的分区或卷空间耗尽且无法扩展，可以在另一个分区上创建表空间并使用它，直到系统能够重新配置。

其次，表空间允许管理员利用对数据库对象使用模式的了解来优化性能。例如，一个使用非常频繁的索引可以放在速度非常快且高可用的磁盘上，例如昂贵的固态设备。同时，一个存放归档数据且很少使用或对性能并不关键的表，则可以放在成本更低、速度更慢的磁盘系统上。

要定义表空间，请使用CREATE TABLESPACE命令，例如：

```
CREATE TABLESPACE fastspace LOCATION '/mnt/sda1/postgresql/data';
```

该位置必须是一个已存在的、属于PostgreSQL操作系统用户的空目录。随后在该表空间中创建的所有对象都将存储在这个目录下的文件中。

> **注意**
>
> 通常来说，在每个逻辑文件系统上创建多个表空间意义不大，因为你无法控制单个文件在一个逻辑文件系统中的具体位置。不过，PostgreSQL并不强制这种限制，而且它实际上并不能直接感知你的系统中文件系统的边界。它只是在你指定的目录中存储文件。

表空间本身必须由数据库超级用户创建，但之后你可以允许普通数据库用户使用它。为此，需要向他们授予其上的CREATE权限。

表、索引以及整个数据库都可以被分配到特定的表空间。为此，对给定表空间拥有CREATE权限的用户必须把表空间名作为参数传给相应命令。例如，下列命令会在表空间space1中创建一个表：

```
CREATE TABLE foo(i int) TABLESPACE space1;
```

另一种方式是使用default_tablespace参数：

```
SET default_tablespace = space1;
CREATE TABLE foo(i int);
```

当default_tablespace被设置为非空字符串时，它会为未显式指定TABLESPACE子句的CREATE TABLE和CREATE INDEX命令隐含地补上该子句。

与数据库关联的表空间用于存储该数据库的系统目录，以及使用该数据库的服务器进程所创建的任何临时文件。此外，如果在创建对象时没有给出TABLESPACE子句（无论是显式给出还是通过default_tablespace给出），它还是在该数据库中创建的表和索引的默认表空间。如果创建数据库时没有为其指定表空间，它将使用其所复制模板数据库的同一个表空间。

initdb会自动创建两个表空间。pg_global表空间只用于共享系统目录。pg_default表空间是template1和template0数据库的默认表空间（因此，除非被TABLESPACE子句覆盖，而该子句可在CREATE DATABASE中指定，否则它也会是其他数据库的默认表空间）。

表空间一旦创建，只要请求的用户具有足够权限，就可以从任何数据库中使用它。这意味着，只有当使用该表空间的所有数据库中的所有对象都已被移除之后，才能删除该表空间。

要移除一个空表空间，使用DROP TABLESPACE命令。

要确定现有表空间的集合，可以检查pg_tablespace系统目录，例如

```
SELECT spcname FROM pg_tablespace;
```

psql程序的\db元命令也可用于列出现有的表空间。

PostgreSQL使用符号链接来简化表空间的实现。这意味着表空间只能在支持符号链接的系统上使用。

目录$PGDATA/pg_tblspc包含指向集簇中定义的每个非内建表空间的符号链接。尽管不推荐，但可以通过重新定义这些链接来手工调整表空间布局。两个警告：不要在服务器运行期间这样做；并且在重启服务器之后，要用新位置更新pg_tablespace目录。（如果不这样做，pg_dump将继续输出旧的表空间位置。）

第 20 章 客户端认证

当客户端应用连接到数据库服务器时，它会指定要以哪个 PostgreSQL 数据库用户名连接，这很像以某个特定用户身份登录 Unix 计算机。在 SQL 环境中，当前活动的数据库用户名决定了对数据库对象的访问权限 – 详见第 18 章。因此，必须限制哪些数据库用户能够连接。

> **注意**
>
> 如第 18 章中所述，PostgreSQL 实际上是以“角色”为单位进行权限管理的。在本章中，我们统一使用数据库用户来表示“拥有 `LOGIN` 权限的角色”。

认证是数据库服务器确认客户端身份的过程，
并据此决定是否允许该客户端应用（或者运行该客户端应用的用户）
以请求的数据库用户名进行连接。

PostgreSQL 提供了多种不同的客户端认证方法。用于认证特定客户端连接的方法可以根据（客户端）主机地址、数据库和用户来选择。

PostgreSQL 数据库用户名在逻辑上独立于服务器所在操作系统中的用户名。如果某台服务器的所有用户在该机器上也都有账号，
那么为他们分配与操作系统用户名一致的数据库用户名是有意义的。不过，
接受远程连接的服务器可能有许多数据库用户并没有本地操作系统账号，在这种情况下，
数据库用户名与操作系统用户名之间就不必存在任何对应关系。

20.1. pg_hba.conf 文件

客户端认证由一个配置文件控制，该文件按惯例命名为 `pg_hba.conf`，并存放在数据库集簇的数据目录中。（HBA 代表 host-based authentication，即基于主机的认证。）当数据目录由 `initdb` 初始化时，会安装一个默认的 `pg_hba.conf` 文件。不过，也可以把认证配置文件放在别处；请参见配置参数hba_file。

`pg_hba.conf` 文件的基本格式是一组记录，每行一条。空行会被忽略，`#` 注释字符之后的文本同样如此。记录不能跨行续写。每条记录由若干字段组成，字段之间用空格和/或制表符分隔。
如果字段值用引号括起来，就可以包含空白。

每条认证记录都指定一种连接类型、一个客户端 IP 地址范围（如果该连接类型需要）、一个数据库名、一个用户名，以及对匹配这些参数的连接要使用的认证方法。第一条同时匹配连接类型、客户端地址、请求数据库和用户名的记录会被用来执行认证。这里不存在“继续向后匹配”
或“后备”机制：如果选中某条记录而认证失败，就不会再考虑后续记录。
如果没有任何记录匹配，则拒绝访问。

记录可以采用以下七种格式之一

```
local      database  user  auth-method  [auth-option]
host       database  user  CIDR-address  auth-method  [auth-option]
hostssl    database  user  CIDR-address  auth-method  [auth-option]
hostnossl  database  user  CIDR-address  auth-method  [auth-option]
host       database  user  IP-address  IP-mask  auth-method  [auth-
option]
hostssl    database  user  IP-address  IP-mask  auth-method  [auth-
option]
```

`hostnossl` *`database`* *`user`* *`IP-address`* *`IP-mask`* *`auth-method`* [*`auth-option`*]

各字段的含义如下：

`local`

该记录匹配使用 Unix 域套接字的连接尝试。没有这种类型的记录时，Unix 域套接字连接将被禁止。

`host`

该记录匹配使用 TCP/IP 进行的连接尝试。`host` 记录既匹配 SSL 加密连接尝试，也匹配非 SSL 连接尝试。

> **注意**
>
> 除非服务器启动时为配置参数listen_addresses设置了恰当的值，否则无法进行远程 TCP/IP 连接，因为默认行为是只在本地环回地址 `localhost` 上监听 TCP/IP 连接。

`hostssl`

该记录匹配使用 TCP/IP 进行的连接尝试，但仅当连接使用 SSL 加密时才匹配。

要使用此选项，服务器编译时必须带有 SSL 支持。此外，还必须在服务器启动时通过设置ssl配置参数来启用 SSL（更多信息参见第 16.7 节）。

`hostnossl`

这种记录类型的行为与 `hostssl` 相反：它只匹配通过 TCP/IP 进行的、不使用 SSL 的连接尝试。

`database`

指定此记录匹配哪些数据库名。值 `all` 表示匹配所有数据库。值 `sameuser` 表示当请求的数据库与请求的用户同名时记录匹配。值 `samerole` 表示请求的用户必须是与请求数据库同名的角色的成员。（`samegroup` 是 `samerole` 已过时但仍被接受的拼法。）否则，这就是某个具体 PostgreSQL 数据库的名称。多个数据库名可以用逗号分隔提供。还可以在文件名前加 `@` 来指定一个包含数据库名的单独文件。

`user`

指定此记录匹配哪些数据库用户名。值 `all` 表示匹配所有用户。否则，它是某个具体数据库用户的名称，或以 `+` 开头的组名。（回想一下，PostgreSQL 中用户和组之间没有真正的区别；`+` 标记的实际含义是“匹配直接或间接属于该成员的任何角色”，而不带 `+` 标记的名称只匹配那个特定角色。）多个用户名可以用逗号分隔提供。还可以在文件名前加 `@` 来指定一个包含用户名的单独文件。

`CIDR-address`

指定此记录匹配的客户端机器 IP 地址范围。此字段包含一个标准点分十进制表示的 IP 地址和一个 CIDR 掩码长度。（IP 地址只能以数值形式指定，不能用域名或主机名。）掩码长度指示

客户端 IP 地址中必须匹配的高位位数。给定 IP 地址中此位之后的位必须为零。IP 地址、/ 与 CIDR 掩码长度之间不能有任何空白。

CIDR-address 的典型示例是：单个主机用 `172.20.143.89/32`，或网络用 `172.20.143.0/24`。要指定单个主机，IPv4 使用 CIDR 掩码 32，IPv6 使用 128。

以 IPv4 格式给出的 IP 地址会匹配拥有对应地址的 IPv6 连接，例如 `127.0.0.1` 会匹配 IPv6 地址 `::ffff:127.0.0.1`。以 IPv6 格式给出的条目只匹配 IPv6 连接，即使所表示的地址位于 IPv4-in-IPv6 范围内也是如此。注意，如果系统的 C 库不支持 IPv6 地址，以 IPv6 格式给出的条目将被拒绝。

此字段只适用于 `host`、`hostssl` 和 `hostnossl` 记录。

IP-address
IP-mask

这些字段可用作 *CIDR-address* 表示法的替代。此时指定的不是掩码长度，而是在单独一列中指定实际的掩码。例如，`255.0.0.0` 表示 IPv4 CIDR 掩码长度 8，`255.255.255.255` 表示 CIDR 掩码长度 32。

这些字段只适用于 `host`、`hostssl` 和 `hostnossl` 记录。

auth-method

指定当连接匹配此记录时所使用的认证方法。可选方法在此概述，细节见第 20.2 节。

`trust`

无条件地允许连接。此方法允许任何能连接到 PostgreSQL 数据库服务器的人以他们希望的任何 PostgreSQL 用户身份登录，无需密码或任何其他认证。详见第 20.2.1 节。

`reject`

无条件地拒绝连接。这有助于从一组主机中“过滤掉”某些主机。

`md5`

要求客户端提供经 MD5 加密的密码进行认证。详见第 20.2.2 节。

`crypt`

> **注意**
>
> 此选项只推荐用于与 7.2 之前的客户端通信。

要求客户端提供经`crypt()`加密的密码进行认证。现在推荐用`md5`取代`crypt`。详见第 20.2.2 节。

`password`

要求客户端提供未加密的密码进行认证。由于密码以明文形式在网络上发送，不应在不可信网络上使用此方法。详见第 20.2.2 节。

`krb5`

使用 Kerberos V5 认证用户。这只适用于 TCP/IP 连接。详见第 20.2.3 节。

`ident`

获取客户端的操作系统用户名（对 TCP/IP 连接，通过联系客户端上的 ident 服务器获取；对本地连接，从操作系统获取），并使用跟在`ident`
关键词后面的映射所指定的映射，
检查该用户是否被允许以其请求的数据库用户身份连接。详见第 20.2.4 节。

`pam`

使用操作系统提供的可插拔认证模块（PAM）服务进行认证。详见第 20.2.5 节。

`auth-option`

这个可选字段的含义取决于所选的认证方法。细节见下文。

由 `@` 构造引用的文件会被读取为名称列表，其中的名称可以用空白或逗号分隔。注释仍然用 `#` 引入，与 `pg_hba.conf` 中相同，并且允许嵌套的 `@` 构造。除非 `@` 后面的文件名是绝对路径，否则它会被视为相对于引用它的文件所在目录。

由于每次连接尝试都会按顺序检查 `pg_hba.conf` 记录，因此记录的顺序很重要。通常，靠前的记录使用较严格的连接匹配条件和较弱的认证方法，
靠后的记录使用较宽松的匹配条件和较强的认证方法。例如，可能希望对本地 TCP/IP 连接使用 `trust` 认证，而要求远程 TCP/IP 连接提供密码。此时，为来自 127.0.0.1 的连接指定 `trust` 认证的记录，应放在为更大范围的允许客户端 IP 地址指定密码认证的记录之前。

`pg_hba.conf` 文件在启动时以及主服务器进程（`postmaster`）收到 SIGHUP 信号时读取。如果在运行中的系统上编辑了该文件，需要通知 `postmaster`（用 `pg_ctl reload` 或 `kill -HUP`）让它重新读取该文件。

例 20.1展示了一些 `pg_hba.conf` 条目示例。有关不同认证方法的详细信息，参见下一节。

例 20.1. `pg_hba.conf` 记录示例

```
# Allow any user on the local system to connect to any database with
# any database user name using Unix-domain sockets (the default for
 local
# connections).
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
local   all             all                                     trust

# The same using local loopback TCP/IP connections.
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
host    all             all             127.0.0.1/32            trust

# The same as the previous line, but using a separate netmask column
#
# TYPE  DATABASE        USER            IP-ADDRESS      IP-MASK
      METHOD
host    all             all             127.0.0.1       255.255.255.
255     trust
```

```
# Allow any user from any host with IP address 192.168.93.x to connect
# to database "postgres" as the same user name that ident reports for
# the connection (typically the operating system user name).
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
host    postgres        all             192.168.93.0/24         ident
 sameuser

# Allow any user from host 192.168.12.10 to connect to database
# "postgres" if the user's password is correctly supplied.
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
host    postgres        all             192.168.12.10/32        md5

# In the absence of preceding "host" lines, these two lines will
# reject all connections from 192.168.54.1 (since that entry will be
# matched first), but allow Kerberos 5 connections from anywhere else
# on the Internet.  The zero mask causes no bits of the host IP
# address to be considered, so it matches any host.
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
host    all             all             192.168.54.1/32         reject
host    all             all             0.0.0.0/0               krb5

# Allow users from 192.168.x.x hosts to connect to any database, if
# they pass the ident check.  If, for example, ident says the user is
# "bryanh" and he requests to connect as PostgreSQL user "guest1", the
# connection is allowed if there is an entry in pg_ident.conf for map
# "omicron" that says "bryanh" is allowed to connect as "guest1".
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
host    all             all             192.168.0.0/16          ident
 omicron

# If these are the only three lines for local connections, they will
# allow local users to connect only to their own databases (databases
# with the same name as their database user name) except for
 administrators
# and members of role "support", who can connect to all databases.
 The file
# $PGDATA/admins contains a list of names of administrators.
 Passwords
# are required in all cases.
#
# TYPE  DATABASE        USER            CIDR-ADDRESS            METHOD
local   sameuser        all                                     md5
local   all             @admins                                 md5
local   all             +support                                md5

# The last two lines above can be combined into a single line:
local   all             @admins,+support                        md5

# The database column can also use lists and file names:
```

```
local   db1,db2,@demodbs  all                                    md5
```

20.2. 认证方法

下面几节会更详细地介绍这些认证方法。

20.2.1. trust 认证

当trust认证被指定时，PostgreSQL假设任何可以连接到服务器的人都被授权使用他们指定的任何数据库用户名（包括超级用户）访问数据库。当然，在database和 user列中设置的限制仍然适用。只有当在操作系统层对进入服务器的连接有足够保护时，才应该使用这种方法。

trust 认证对单用户工作站上的本地连接来说是合适且非常方便的。在多用户机器上，通常单用它并不合适。不过，如果用文件系统权限限制对服务器 Unix 域套接字文件的访问，即使在多用户机器上也可以使用 trust。为此，可按第 17.3 节中的说明设置 unix_socket_permissions（可能还有 unix_socket_group）配置参数。或者，也可以设置 unix_socket_directory 配置参数，把套接字文件放在一个经过适当限制的目录中。

设置文件系统权限只能有助于 Unix 套接字连接。本地 TCP/IP 连接不会被文件系统权限限制。因此，如果你想利用文件系统权限来控制本地安全，那么从pg_hba.conf中移除host ... 127.0.0.1 ...行，或者把它改为一个非trust认证方法。

只有当你信任由 pg_hba.conf 中指定 trust 的行所允许连接的每台机器上的每个用户时，trust 认证才适合用于 TCP/IP 连接。对来自 localhost（127.0.0.1）以外的任何 TCP/IP 连接使用 trust，通常都不合理。

20.2.2. 密码认证

基于密码的认证方法有md5、crypt和password。除了密码通过连接发送的方式——分别以 MD5 哈希、crypt 加密和明文发送——之外，这些方法的操作是相似的。一个限制是crypt方法不能用于pg_authid中已被加密存储的密码。

如果你担心密码“嗅探”攻击，那么应优先使用md5，只有在必须支持 7.2 之前的客户端时才使用crypt。明文password应当尽量避免，尤其是通过开放 Internet 的连接（除非你为连接使用了SSL、SSH或其他通信安全包装）。

PostgreSQL 数据库密码与操作系统用户密码是分开的。每个数据库用户的密码存储在 pg_authid 系统目录中。密码可以用 SQL 命令CREATE USER和ALTER USER管理，例如 **CREATE USER foo WITH PASSWORD 'secret'**。如果用户没有设置密码，存储的密码为空，该用户的密码认证将始终失败。

20.2.3. Kerberos 认证

Kerberos 是一种行业标准的、适合在公共网络上进行分布式计算的安全认证系统。Kerberos 系统的说明超出了本文档的范围；在最一般的形式下它可能相当复杂（但功能强大）。Kerberos FAQ[1] 或 MIT Kerberos 页面[2]是不错的入门起点。Kerberos 发行版有多个来源。

PostgreSQL 支持 Kerberos 版本 5。构建 PostgreSQL 时必须启用 Kerberos 支持；更多信息参见第 14 章。

[1] http://www.cmf.nrl.navy.mil/CCS/people/kenh/kerberos-faq.html
[2] http://web.mit.edu/kerberos/www/

PostgreSQL 的运行方式与普通的 Kerberos 服务相同。服务主体的名称为 *servicename/hostname@realm*。

servicename 可以在服务器端用krb_srvname配置参数设置，在客户端用 krbsrvname 连接参数设置。（另见第 28.1 节。）安装默认值可以在构建时用 ./configure --with-krb-srvnam=whatever 从默认的 postgres 改为其他值。在大多数环境中，此参数永远不需要更改。但是，在同一台主机上支持多个 PostgreSQL 安装时，就需要修改它。某些 Kerberos 实现可能还要求不同的服务名，例如 Microsoft Active Directory 要求服务名为大写（POSTGRES）。

hostname 是服务器机器的完全限定主机名。服务主体的 realm 是服务器机器的首选 realm。

客户端主体必须以PostgreSQL数据库用户名作为其第一个组成部分，例如pgusername/otherstuff@realm。目前，PostgreSQL不检查客户端的 realm；因此，如果启用了跨 realm 认证，那么任何能够与你的 realm 通信的任何 realm 中的任何主体都会被接受。

确保你的服务器 keytab 文件可被 PostgreSQL 服务器账户读取（最好只有该账户可读）。（另见第 16.1 节。）密钥文件的位置由krb_server_keyfile配置参数指定。默认值为 /usr/local/pgsql/etc/krb5.keytab（或构建时指定为 sysconfdir 的任何目录）。

keytab 文件由 Kerberos 软件生成；详情参见 Kerberos 文档。以下示例适用于兼容 MIT 的 Kerberos 5 实现：

```
kadmin% ank -randkey postgres/server.my.domain.orgkadmin% ktadd -k
 krb5.keytab postgres/server.my.domain.org
```

连接数据库时，请确保持有与请求的数据库用户名相匹配的主体票据。例如，对于数据库用户名 fred，主体 fred@EXAMPLE.COM 和 fred/users.example.com@EXAMPLE.COM 都可以用来向数据库服务器认证。

如果在 Apache Web 服务器上使用 mod_auth_kerb[3]和 mod_perl，则可以在 mod_perl 脚本中使用 AuthType KerberosV5SaveCredentials。这样即可通过 Web 安全地访问数据库，无需额外的密码。

20.2.4. 基于 ident 的认证

ident 认证方法的工作方式是获取客户端的操作系统用户名，然后使用一个列出允许的对应名称对的映射文件，确定允许的数据库用户名。确定客户端用户名是安全上的关键点，
其工作方式因连接类型而异，如下所述。

20.2.4.1. 通过 TCP/IP 的 Ident 认证

“标识协议”（Identification Protocol）在RFC 1413中描述。几乎所有类 Unix 操作系统都自带 ident 服务器，默认监听 TCP 端口 113。ident 服务器的基本功能是回答这样的问题：“从你的端口 *X* 连到我的端口 *Y* 的连接，是哪个用户发起的？”由于建立物理连接时，PostgreSQL 已知 *X* 和 *Y*，因此可以询问连接客户端所在主机上的 ident 服务器，理论上能够确定任意给定连接的操作系统用户。

这个过程的缺点在于它依赖客户端本身的可信性：如果客户端机器不可信或者已被攻破，
攻击者几乎可以在 113 端口上运行任何程序，并返回任意他选择的用户名。因此，这种认证方法只适用于封闭网络，在这类网络中每台客户端机器都受到严格控制，
而且数据库管理员与系统管理员之间保持密切协作。换句话说，你必须信任运行 ident 服务器的那台机器。请注意下面的警告：

[3]http://modauthkerb.sf.net

> 标识协议的本意不是作为一种授权或访问控制协议。
>
> —RFC 1413

有些 ident 服务器提供了一个非标准选项，会让返回的用户名被加密，而解密所需密钥只有发起连接机器的管理员才知道。将 ident 服务器与 PostgreSQL 配合使用时，绝不能启用这个选项，因为 PostgreSQL 无法解密返回的字符串，也就无法确定实际用户名。

20.2.4.2. 通过本地套接字的 Ident 认证

在支持对 Unix 域套接字使用 `SO_PEERCRED` 请求的系统（目前包括 Linux、FreeBSD、NetBSD、OpenBSD 和 BSD/OS）上，ident 认证也可以应用于本地连接。在这种情况下，使用 ident 认证不会增加任何安全风险；事实上，在这类系统上，它是本地连接的首选方案。

在不支持 `SO_PEERCRED` 请求的系统上，ident 认证只适用于 TCP/IP 连接。作为一种变通办法，可以指定 localhost 地址 127.0.0.1 并连接到该地址。此方法的可信程度取决于你对本地 ident 服务器的信任程度。

20.2.4.3. Ident 映射

使用基于 ident 的认证时，在确定了发起连接的操作系统用户的名称之后，PostgreSQL会检查该用户是否被允许以其请求的数据库用户身份连接。这由`pg_hba.conf`文件中跟在`ident`关键词后面的 ident 映射参数控制。有一个预定义的 ident 映射`sameuser`，它允许任何操作系统用户以同名的数据库用户身份连接（如果后者存在）。其他映射必须手工创建。

`sameuser`之外的 ident 映射定义在 ident 映射文件中，该文件默认名为`pg_ident.conf`，存储在集簇的数据目录中。（不过也可以把该映射文件放在其他地方，参见ident_file配置参数。）ident 映射文件包含如下一般形式的行：

`map-nameident-usernamedatabase-username`

注释和空白的处理方式与`pg_hba.conf`中相同。*`map-name`*是任意名称，将用在`pg_hba.conf`中引用该映射。另外两个字段指定哪个操作系统用户被允许以哪个数据库用户身份连接。
同一个*`map-name`*可以重复使用，在一个映射中指定更多用户映射。
对于一个操作系统用户可以对应多少个数据库用户，或者反过来，都没有限制。

`pg_ident.conf`文件在启动时以及主服务器进程（`postmaster`）收到SIGHUP信号时读取。如果你在一个运行中的系统上编辑了该文件，你需要通知`postmaster`重新读取它（使用`pg_ctl reload`或`kill -HUP`）。

一个可以与例 20.1中的`pg_hba.conf`文件配合使用的`pg_ident.conf`文件如例 20.2所示。在这个示例配置中，任何登录到 192.168 网络上某台机器的用户，如果其 Unix 用户名不是`bryanh`、`ann`或`robert`，将不会被授予访问权限。Unix 用户`robert`只有在他试图以PostgreSQL用户`bob`而非`robert`或其他任何人身份连接时才被允许访问。`ann`只被允许以`ann`身份连接。用户`bryanh`被允许以`bryanh`本人或`guest1`身份连接。

例 20.2. 一个 `pg_ident.conf` 文件示例

```
# MAPNAME     IDENT-USERNAME    PG-USERNAME

omicron       bryanh            bryanh
omicron       ann               ann
# bob has user name robert on these machines
omicron       robert            bob
# bryanh can also connect as guest1
```

```
omicron       bryanh            guest1
```

20.2.5. PAM 认证

这种认证方法的工作方式与 `password` 类似，只不过它使用 PAM（可插拔认证模块）作为认证机制。默认的 PAM 服务名为 `postgresql`。你也可以在`pg_hba.conf`文件中`pam`关键词后面提供自己的服务名。PAM 只用于验证用户名/密码对。因此，在使用 PAM 进行认证之前，用户必须已经存在于数据库中。关于 PAM 的更多信息，请阅读 Linux-PAM 页面[4]和 Solaris PAM 页面[5]。

20.3. 认证问题

认证失败及相关问题通常会表现为类似下面这样的错误消息：

```
FATAL:  no pg_hba.conf entry for host "123.123.123.123", user "andym",
 database "testdb"
```

这种情况通常意味着你已经成功联系到了服务器，但服务器不愿意接受你的连接。
正如消息所示，服务器拒绝了该连接请求，因为它没有在自己的 `pg_hba.conf` 配置文件中找到匹配项。

```
FATAL:  password authentication failed for user "andym"
```

这样的消息表示你已经联系到了服务器，而且服务器也愿意继续处理连接，
但前提是你必须先通过 `pg_hba.conf` 文件中指定的认证方法。请检查你提供的密码；如果错误消息提到了 Kerberos 或 ident 等认证类型，也请检查对应的软件配置。

```
FATAL:  user "andym" does not exist
```

指定的数据库用户不存在。

```
FATAL:  database "testdb" does not exist
```

你试图连接的数据库不存在。注意，如果你没有指定数据库名，
默认会使用数据库用户名作为数据库名，但这不一定是所需的数据库名。

> **提示**
>
> 服务器日志中可能包含比返回给客户端的更多认证失败信息。如果你不清楚失败原因，请检查服务器日志。

[4]http://www.kernel.org/pub/linux/libs/pam/
[5]http://www.sun.com/software/solaris/pam/

第 21 章 Localization

本章从管理员的角度描述可用的本地化特性。PostgreSQL通过两种途径支持本地化：

- 使用操作系统的区域特性来提供与区域相关的排序顺序、数字格式、翻译消息以及其他方面。
- 提供 PostgreSQL服务器中定义的多种字符集（包括多字节字符集），以支持存储各种语言的文本，并提供客户端与服务器之间的字符集转换。

21.1. Locale Support

区域设置支持是指应用程序在字母表、排序、数字格式等方面尊重文化偏好。
PostgreSQL使用服务器操作系统提供的标准 ISO C 和POSIX区域设置机制。更多信息请参考系统文档。

21.1.1. Overview

在使用`initdb`创建数据库集簇时，区域设置支持会自动初始化。默认情况下，`initdb`
会使用其执行环境中的区域设置来初始化数据库集簇，
因此如果你的系统已经设置为数据库集簇想要使用的区域设置，就无需额外操作。
如果想使用不同的区域设置（或者不确定系统当前设置了哪种区域设置），可以通过指定
`--locale`选项，精确告诉`initdb`使用哪种区域设置。例如：

```
initdb --locale=sv_SE
```

这个例子把区域设置为瑞典（`SE`）所讲的瑞典语（`sv`）。其他可能的设置有`en_US`（美国英语）和`fr_CA`（加拿大法语）。如果一个区域可以使用多种字符集，那么规范形如`cs_CZ.ISO8859-2`。系统上有哪些区域设置可用，以及它们以什么名称出现，
取决于操作系统供应商提供了什么以及实际安装了什么。在大多数系统上，命令 `locale -a`会给出所有可用区域设置的列表。

有时混合多个区域设置的规则会很有用，例如使用英语排序规则但使用西班牙语消息。为此，
存在一组只控制本地化规则某些方面的区域设置子类别：

`LC_COLLATE`	字符串排序顺序
`LC_CTYPE`	字符分类（什么算字母？它的大写等价形式是什么？）
`LC_MESSAGES`	消息的语言
`LC_MONETARY`	货币数量使用的格式
`LC_NUMERIC`	数字的格式
`LC_TIME`	日期和时间的格式

这些类别名会转换成`initdb`选项名，用于覆盖特定类别的区域设置选择。例如，
要将区域设置设为加拿大法语，但对货币格式采用美国规则，可以使用`initdb --locale=fr_CA --lc-monetary=en_US`。

如果希望系统的行为如同没有区域支持一样，可以使用特殊区域`C`或`POSIX`。

某些区域类别的性质决定了它们的值必须在数据库集簇的整个生命周期内保持固定。也就是说，
一旦`initdb`运行过，你就再也无法更改它们。`LC_COLLATE`和`LC_CTYPE`就属于这类类别。
它们影响索引的排序顺序，因此必须保持固定，否则文本列上的索引会损坏。
PostgreSQL通过记录`initdb`所看到的`LC_COLLATE`和`LC_CTYPE`值来强制这一点。
服务器启动时会自动采用这两个值。

其他区域类别可以在服务器运行期间随时按需更改，
方法是设置与区域类别同名的运行时配置变量（细节见第 17.10.2 节）。initdb所选择的值实际上只是写入了配置文件postgresql.conf，作为服务器启动时的默认值。如果你从postgresql.conf中删除这些赋值，服务器将从其执行环境继承设置。

请注意，服务器的区域设置行为由服务器看到的环境变量决定，
而不是由任何客户端的环境决定。因此，务必在启动服务器前配置好正确的区域设置。
由此带来的一个结果是，如果客户端和服务器使用不同的区域设置，
消息可能会因其来源不同而显示为不同语言。

> **注意**
>
> 这里所说的从执行环境继承区域设置，
> 在大多数操作系统上是指：对于某个给定的区域设置类别，例如排序规则，
> 会按以下顺序检查环境变量，直到找到一个已设置的变量为止：LC_ALL、LC_COLLATE（或相应类别对应的变量）、LANG。如果这些环境变量都没有设置，则区域设置默认为C。
>
> 某些消息本地化库还会查看环境变量LANGUAGE，
> 它会覆盖所有其他用于设置消息的语言的区域设置。如果有疑问，
> 请参考操作系统的文档，特别是关于gettext的文档，以获取更多信息。

若要让消息可以翻译为用户偏好的语言，构建时必须启用NLS。
这一选择与其他区域设置支持无关。

21.1.2. Behavior

区域设置影响下列 SQL 特性：

- 使用ORDER BY或标准比较操作符对文本数据进行排序的顺序
- LIKE子句使用索引的能力
- upper、lower和initcap函数
- to_char函数族

在PostgreSQL中使用C或 POSIX之外的区域设置，缺点是会影响性能。它会减慢字符处理速度，并阻止普通索引用于LIKE。因此，只有在确有需要时才应使用区域设置。

作为一种变通方法，为了让PostgreSQL在非 C 区域下能对LIKE子句使用索引，存在若干自定义操作符类。它们允许创建一个执行严格的逐字符比较、忽略区域比较规则的索引。
更多信息参见第 11.8 节。

21.1.3. Problems

如果区域设置支持没有像上面说明的那样工作，
请检查操作系统中的区域设置支持是否已经正确配置。要检查系统中安装了哪些区域设置，
可以使用 locale -a命令（如果操作系统提供该命令）。

请确认PostgreSQL实际使用的区域设置就是你预期的区域设置。LC_COLLATE和LC_CTYPE设置是在 initdb时确定的，不重复执行initdb就无法更改。其他区域设置，包括 LC_MESSAGES和LC_MONETARY，最初由启动服务器时的环境决定，但可以在运行中动态修改。
你可以使用SHOW 命令检查当前生效的区域设置。

源码发布包中的`src/test/locale`目录包含 PostgreSQL区域设置支持的测试套件。

那些通过解析错误消息文本来处理服务器端错误的客户端应用，显然会遇到问题，因为服务器消息可能使用不同语言。这类应用的作者应改用错误代码机制。

维护消息翻译目录需要许多志愿者持续投入，他们希望 PostgreSQL能够良好地支持他们偏好的语言。如果你所用语言的消息目前还不可用，或尚未完全翻译，我们将非常感谢你的协助。如果你愿意帮忙，请参考第 45 章，或向开发者邮件列表写信。

21.2. Character Set Support

PostgreSQL中的字符集支持允许你以多种字符集存储文本，包括 ISO 8859 系列之类的单字节字符集，以及 EUC（扩展 Unix 编码）、UTF-8 和 Mule 内部编码等多字节字符集。所有受支持的字符集都可以被客户端透明地使用，但其中少数不支持在服务器内部使用（即不能作为服务器端编码）。默认字符集是在使用`initdb`初始化PostgreSQL 数据库集簇时选定的。创建数据库时可以覆盖该设置，因此你可以拥有多个数据库，并让每个数据库使用不同的字符集。

21.2.1. 支持的字符集

表 21.1显示了PostgreSQL中可用的字符集。

表 21.1. PostgreSQL 字符集

名称	描述	语言	是否服务器端?	字节数/字符	别名
`BIG5`	Big Five	繁体中文	否	1-2	`WIN950`, `Windows950`
`EUC_CN`	扩展 Unix 编码-CN	简体中文	是	1-3	
`EUC_JP`	扩展 Unix 编码-JP	日语	是	1-3	
`EUC_KR`	扩展 Unix 编码-KR	韩语	是	1-3	
`EUC_TW`	扩展 Unix 编码-TW	繁体中文、台湾语	是	1-3	
`GB18030`	国家标准	中文	否	1-2	
`GBK`	扩展国家标准	简体中文	否	1-2	`WIN936`, `Windows936`
`ISO_8859_5`	ISO 8859-5, ECMA 113	拉丁语/西里尔语	是	1	
`ISO_8859_6`	ISO 8859-6, ECMA 114	拉丁语/阿拉伯语	是	1	
`ISO_8859_7`	ISO 8859-7, ECMA 118	拉丁语/希腊语	是	1	
`ISO_8859_8`	ISO 8859-8, ECMA 121	拉丁语/希伯来语	是	1	
`JOHAB`	JOHAB	韩语（谚文）	是	1-3	
`KOI8`	KOI8-R(U)	西里尔语	是	1	`KOI8R`
`LATIN1`	ISO 8859-1, ECMA 94	西欧	是	1	`ISO88591`

名称	描述	语言	是否服务器端?	字节数/字符	别名
LATIN2	ISO 8859-2, ECMA 94	中欧	是	1	ISO88592
LATIN3	ISO 8859-3, ECMA 94	南欧	是	1	ISO88593
LATIN4	ISO 8859-4, ECMA 94	北欧	是	1	ISO88594
LATIN5	ISO 8859-9, ECMA 128	土耳其语	是	1	ISO88599
LATIN6	ISO 8859-10, ECMA 144	北欧	是	1	ISO885910
LATIN7	ISO 8859-13	波罗的海	是	1	ISO885913
LATIN8	ISO 8859-14	凯尔特语	是	1	ISO885914
LATIN9	ISO 8859-15	带欧元符号和重音字符的LATIN1	是	1	ISO885915
LATIN10	ISO 8859-16, ASRO SR 14111	罗马尼亚语	是	1	ISO885916
MULE_INTERNAL	Mule 内部编码	多语种 Emacs	是	1-4	
SJIS	Shift JIS	日语	否	1-2	Mskanji, ShiftJIS ,WIN932, Windows932
SQL_ASCII	未指定（见正文）	任意	是	1	
UHC	统一韩语编码	韩语	否	1-2	WIN949, Windows949
UTF8	Unicode，8 位	所有	是	1-4	Unicode
WIN866	Windows CP866	西里尔语	是	1	ALT
WIN874	Windows CP874	泰语	是	1	
WIN1250	Windows CP1250	中欧	是	1	
WIN1251	Windows CP1251	西里尔语	是	1	WIN
WIN1252	Windows CP1252	西欧	是	1	
WIN1256	Windows CP1256	阿拉伯语	是	1	
WIN1258	Windows CP1258	越南语	是	1	ABC, TCVN, TCVN5712, VSCII

并非所有客户端API都支持上表中的全部字符集。例如，PostgreSQL JDBC 驱动就不支持 MULE_INTERNAL、LATIN6、LATIN8和LATIN10。

SQL_ASCII 设置的行为与其他设置有很大不同。当服务器字符集为 SQL_ASCII 时，服务器按 ASCII 标准解释字节值 0-127，而将字节值 128-255 视为不作解释的字符。设置为 SQL_ASCII

时，不会进行任何编码转换。因此，这一设置与其说是声明使用某种特定编码，不如说是声明不关心编码。在大多数情况下，如果需要处理任何非 ASCII 数据，使用 SQL_ASCII 都是不明智的，因为 PostgreSQL 将无法帮助转换或验证非 ASCII 字符。

21.2.2. 设置字符集

initdb为PostgreSQL集簇定义默认字符集（编码）。例如：

```
initdb -E EUC_JP
```

这会把默认字符集设为EUC_JP（日语的扩展 Unix 编码）。如果你喜欢更长的选项形式，也可以用--encoding代替-E。如果既没有给出-E也没有给出 --encoding，initdb会根据指定的或默认的区域设置，尝试确定应使用的合适编码。

你可以创建使用不同字符集的数据库：

```
createdb -E EUC_KR korean
```

这将创建一个名为korean的数据库，它使用 EUC_KR字符集。另一种实现方式是使用以下 SQL 命令：

```
CREATE DATABASE korean WITH ENCODING 'EUC_KR';
```

数据库的编码存储在系统目录pg_database中。可以使用 psql的-l选项或\l命令查看。

```
$ psql -l
            List of databases
   Database    |  Owner  |   Encoding
---------------+---------+---------------
 euc_cn        | t-ishii | EUC_CN
 euc_jp        | t-ishii | EUC_JP
 euc_kr        | t-ishii | EUC_KR
 euc_tw        | t-ishii | EUC_TW
 mule_internal | t-ishii | MULE_INTERNAL
 postgres      | t-ishii | EUC_JP
 regression    | t-ishii | SQL_ASCII
 template1     | t-ishii | EUC_JP
 test          | t-ishii | EUC_JP
 utf8          | t-ishii | UTF8
(9 rows)
```

重要

虽然你可以为数据库指定任何想要的编码，
但选择一个与所选区域设置所期望的编码不一致的编码是不明智的。LC_COLLATE和 LC_CTYPE设置隐含着一种特定的编码，依赖区域的操作（例如排序）
很可能会错误解释以不兼容编码存储的数据。

由于这些区域设置被initdb冻结，
在同一集簇的不同数据库中使用不同编码的表面灵活性，与其说是真实的，
不如说是理论上的。这些机制将来很可能会在 PostgreSQL的未来版本中被重新考虑。

安全使用多种编码的一种办法，是在initdb时把区域设置设为C或POSIX，
从而禁用任何真正的区域感知。

21.2.3. 服务器和客户端之间的自动字符集转换

PostgreSQL 支持在服务器和客户端之间，对某些字符集组合进行自动字符集转换。转换信息保存在系统目录 `pg_conversion` 中。PostgreSQL 提供了一些预定义转换，如表 21.2所示。可以使用 SQL 命令 `CREATE CONVERSION` 创建新的转换。

表 21.2. 客户端/服务器字符集转换

服务器字符集	可用的客户端字符集
BIG5	不支持作为服务器编码
EUC_CN	EUC_CN, MULE_INTERNAL, UTF8
EUC_JP	EUC_JP, MULE_INTERNAL, SJIS, UTF8
EUC_KR	EUC_KR, MULE_INTERNAL, UTF8
EUC_TW	EUC_TW, BIG5, MULE_INTERNAL, UTF8
GB18030	不支持作为服务器编码
GBK	不支持作为服务器编码
ISO_8859_5	ISO_8859_5, KOI8, MULE_INTERNAL, UTF8, WIN866, WIN1251
ISO_8859_6	ISO_8859_6, UTF8
ISO_8859_7	ISO_8859_7, UTF8
ISO_8859_8	ISO_8859_8, UTF8
JOHAB	JOHAB, UTF8
KOI8	KOI8, ISO_8859_5, MULE_INTERNAL, UTF8, WIN866, WIN1251
LATIN1	LATIN1, MULE_INTERNAL, UTF8
LATIN2	LATIN2, MULE_INTERNAL, UTF8, WIN1250
LATIN3	LATIN3, MULE_INTERNAL, UTF8
LATIN4	LATIN4, MULE_INTERNAL, UTF8
LATIN5	LATIN5, UTF8
LATIN6	LATIN6, UTF8
LATIN7	LATIN7, UTF8
LATIN8	LATIN8, UTF8
LATIN9	LATIN9, UTF8
LATIN10	LATIN10, UTF8
MULE_INTERNAL	MULE_INTERNAL, BIG5, EUC_CN, EUC_JP, EUC_KR, EUC_TW, ISO_8859_5, KOI8, LATIN1 to LATIN4, SJIS, WIN866, WIN1250, WIN1251
SJIS	不支持作为服务器编码
SQL_ASCII	任意（不会执行转换）
UHC	不支持作为服务器编码
UTF8	所有支持的编码
WIN866	WIN866, ISO_8859_5, KOI8, MULE_INTERNAL, UTF8, WIN1251
WIN874	WIN874, UTF8

服务器字符集	可用的客户端字符集
WIN1250	WIN1250, LATIN2, MULE_INTERNAL, UTF8
WIN1251	WIN1251, ISO_8859_5, KOI8, MULE_INTERNAL, UTF8, WIN866
WIN1252	WIN1252, UTF8
WIN1256	WIN1256, UTF8
WIN1258	WIN1258, UTF8

若要启用自动字符集转换，必须告诉PostgreSQL 你希望客户端使用哪种字符集（编码）。有几种方法可以做到这一点：

- 在psql中使用\encoding 命令。\encoding允许动态更改客户端编码。例如，要把编码改成SJIS，可以输入：

  ```
  \encoding SJIS
  ```

- 使用libpq函数。\encoding实际上就是为达到这一目的而调用 PQsetClientEncoding()。

  ```
  int PQsetClientEncoding(PGconn *conn, const char *encoding);
  ```

  其中*conn*是到服务器的连接，而*encoding*是你想使用的编码。如果该函数成功设置了编码，它返回 0，否则返回 -1。此连接的当前编码可以用下面的函数确定：

  ```
  int PQclientEncoding(const PGconn *conn);
  ```

  注意，它返回的是编码 ID，而不是像EUC_JP这样的符号字符串。要把编码 ID 转换为编码名，可以用：

  ```
  char *pg_encoding_to_char(int encoding_id);
  ```

- 使用SET client_encoding TO。可以使用以下 SQL 命令设置客户端编码：

  ```
  SET CLIENT_ENCODING TO 'value';
  ```

  也可以为此使用标准 SQL 语法SET NAMES：

  ```
  SET NAMES 'value';
  ```

  要查询当前客户端编码：

  ```
  SHOW client_encoding;
  ```

  要返回到默认编码：

  ```
  RESET client_encoding;
  ```

- 使用PGCLIENTENCODING。如果在客户端环境中定义了 PGCLIENTENCODING环境变量，那么连接服务器时会自动选择该客户端编码。
  （之后仍可以用上面提到的任何其他方法覆盖它。）

- 使用配置变量client_encoding。如果设置了 client_encoding变量，那么连接服务器时会自动选择该客户端编码。（之后仍可以用上面提到的任何其他方法覆盖它。）

如果某个特定字符无法完成转换，就会报错。例如，假设服务器使用 EUC_JP而客户端使用LATIN1，此时若返回了一些在LATIN1中没有表示形式的日语字符，就会出现错误。

如果客户端字符集定义为 `SQL_ASCII`，则无论服务器字符集是什么，都会禁用编码转换。与服务器一样，除非处理的全部是 ASCII 数据，否则使用 `SQL_ASCII` 是不明智的。

21.2.4. 进一步阅读

这些都是开始了解各种编码系统的良好资料。

http://www.i18ngurus.com/docs/984813247.html

关于字符集、编码和代码页的大量文档汇集。

ftp://ftp.ora.com/pub/examples/nutshell/ujip/doc/cjk.inf

其中对`EUC_JP`、`EUC_CN`、`EUC_KR`和`EUC_TW`的详细说明见第 3.2 节。

http://www.unicode.org/

Unicode Consortium 网站。

RFC 2044

UTF-8 在这里定义。

第 22 章 日常数据库维护任务

要让一个PostgreSQL服务器平稳运行，需要定期执行一些例行维护工作。
这里讨论的任务本质上是重复性的，可以很容易地用标准 Unix 工具（例如 cron 脚本）实现自动化。但建立合适的脚本并检查其是否成功执行，是数据库管理员的职责。

一个显而易见的维护任务，是按固定计划创建数据的备份副本。没有最近的备份，
在灾难（磁盘故障、火灾、误删关键表等）发生后就没有恢复的可能。
PostgreSQL提供的备份和恢复机制在第 23 章中有详细讨论。

另一大类维护任务是定期对数据库进行“清理”。这一活动在第 22.1 节中讨论。

另一样可能需要定期关注的东西是日志文件管理。这在第 22.3 节中讨论。

和某些其他数据库管理系统相比，PostgreSQL的维护工作量较小。不过，适当地关注这些任务，将大大有助于确保你愉快而高效地使用该系统。

22.1. 日常清理

PostgreSQL的`VACUUM`命令必须定期运行，原因有以下几个：

1. 回收被更新或删除的行所占用的磁盘空间。

2. 更新PostgreSQL查询规划器使用的数据统计信息。

3. 防止因事务 ID 回卷而丢失非常旧的数据。

出于上述各原因而执行的`VACUUM`操作，其频率和范围因各站点的需要而异。因此，
数据库管理员必须理解这些问题，并制定适当的维护策略。本节集中于解释高层问题；
有关命令语法等细节，见VACUUM参考页。

从PostgreSQL 7.2 开始，标准形式的`VACUUM`可以与正常的数据库操作（选择、插入、更新、删除，但不能修改表定义）并行运行。因此，例行清理不像在以前的版本中那样具有侵入性，
也就不那么需要刻意安排在一天中的低使用时段执行。

从PostgreSQL 8.0 开始，有一些配置参数可以调整，用来进一步降低后台清理对性能的影响。参见第 17.4.4 节。

PostgreSQL 8.1 中加入了一种自动执行所需`VACUUM`操作的机制。参见第 22.1.4 节。

22.1.1. 回收磁盘空间

在PostgreSQL的正常操作中，对某一行的`UPDATE`或`DELETE`不会立即移除该行的旧版本。
这种方法对于获得多版本并发控制（见第 12 章）的好处是必需的：当行版本仍可能对其他事务可见时，就不能删除它。但最终，过时或已删除的行版本将不再是任何事务关心的对象。
它占用的空间必须被回收以供新行重用，以避免磁盘空间需求的无限增长。这项工作由运行`VACUUM`完成。

显然，一个经常被更新或删除的表将需要比很少更新的表更频繁地清理。设置一些只`VACUUM`选定表的周期性cron任务、跳过那些已知不常更改的表，可能会有用。
只有当你同时拥有更新频繁的大型表和不常更新的大型表时，
这才可能带来帮助——清理一个小表的额外代价不值得操心。

VACUUM命令有两种变体。第一种形式称为“懒惰清理”或就叫VACUUM，它把表和索引中的过期数据标记为将来可重用；除非过期数据占据的空间位于表的末端并且可以轻松获得一个独占表锁，否则它不会试图回收这些过期数据占用的空间。位于文件开头或中间的未用空间不会导致文件缩短、空间归还给操作系统。VACUUM的这种变体可以与正常的数据库操作并发运行。

第二种形式是VACUUM FULL命令。它使用一种更激进的算法来回收被过期行版本消耗的空间。VACUUM FULL释放的任何空间都会立即归还给操作系统。遗憾的是，VACUUM命令的这种变体在VACUUM FULL处理每个表时会获取该表的独占锁。因此，频繁使用VACUUM FULL可能对并发数据库查询的性能产生极其负面的影响。

标准形式的VACUUM最适合用于维持磁盘空间使用相当平稳的稳态。如果你需要把磁盘空间归还给操作系统，可以使用VACUUM FULL——但释放掉的磁盘空间如果很快又得重新分配，这样做又有什么意义呢？对于维护被频繁更新的表，适度频繁的标准VACUUM运行优于不频繁的VACUUM FULL运行。

对大多数站点，推荐的做法是在一天中低使用时段每天安排一次面向整个数据库的VACUUM，并在必要时更频繁地清理那些更新特别频繁的表。（某些数据修改率极高的安装，甚至会每隔几分钟就对繁忙的表执行一次VACUUM。）如果你在一个集簇中有多个数据库，不要忘记对每一个数据库都执行VACUUM；程序vacuumdb可能会有帮助。

当你知道自己已经删除了一个表中的大多数行时，推荐使用VACUUM FULL，这样用VACUUM FULL更激进的方式可以大幅缩小表的稳态尺寸。在例行的以空间回收为目的的清理中，应使用普通的VACUUM而不是VACUUM FULL。

如果你有一个表，其全部内容会被定期删除，考虑使用TRUNCATE而不是DELETE后接VACUUM。TRUNCATE会立即移除表的全部内容，而不需要后续再执行VACUUM或VACUUM FULL来回收此时未使用的磁盘空间。

22.1.2. 更新规划器统计信息

PostgreSQL查询规划器依赖于有关表内容的统计信息，以便为查询生成良好的计划。这些统计信息由ANALYZE命令收集，它既可以单独调用，也可以作为VACUUM的一个可选步骤执行。拥有足够准确的统计信息很重要，否则糟糕的计划选择可能会降低数据库性能。

就像为了回收空间而进行的清理一样，频繁更新统计信息对于更新频繁的表比对于很少更新的表更有用。但即便是更新频繁的表，如果数据的统计分布变化不大，也未必需要更新统计信息。一个简单的经验法则是考虑表中各列的最小值和最大值变化了多少。例如，一个包含行更新时间的timestamp列，会随着行的插入和更新而持续增大其最大值；这样的列可能比例如存放网站页面 URL 的列更需要频繁更新统计信息。URL 列收到更改的频率可能一样高，但其值的统计分布变化大概相对缓慢。

可以对特定表，甚至仅对表中的特定列运行ANALYZE，因此如果你的应用需要，确实可以比其他统计更频繁地更新某些统计信息。然而在实践中，这个特性的用处值得怀疑。从PostgreSQL 7.2 开始，ANALYZE即使在大型表上也是一项相当快的操作，因为它使用对表行的统计随机抽样，而不是读取每一行。因此，定期对整个数据库运行它可能要简单得多。

提示

虽然按列微调ANALYZE的频率未必很有成效，但按列调整ANALYZE收集的统计信息详细程度可能是值得的。在WHERE

> 子句中被频繁使用且数据分布高度不规则的列，
> 可能需要比其他列更细粒度的数据直方图。参见ALTER TABLE SET STATISTICS。

对大多数站点，推荐的做法是在一天中低使用时段每天安排一次面向整个数据库的ANALYZE；这可以方便地与每夜的VACUUM结合进行。不过，
表统计信息变化相对缓慢的站点可能会发现这样做过了头，更低频率的ANALYZE运行就足够了。

22.1.3. 防止事务 ID 回卷失败

PostgreSQL的 MVCC 事务语义依赖于比较事务 ID（XID）数值的能力：插入 XID 大于当前事务 XID 的行版本处于“未来”，不应对当前事务可见。但事务 ID 的大小有限（撰写本文时为 32 位），长期运行的集簇（超过 40 亿个事务）会遭遇事务 ID 回卷：XID 计数器回卷到零，原本处于过去的事务突然看起来处于未来——这意味着它们的输出变得不可见。简而言之，灾难性的数据丢失。（实际上数据还在那里，但如果你无法访问它，
这并没有什么安慰作用。）

在PostgreSQL 7.2 之前，对付 XID 回卷的唯一手段是每至少 40 亿个事务重新执行一次initdb。这对于高流量站点当然不能令人满意，因此设计了一种更好的解决方案。
新方法允许服务器无限期运行，而不需要initdb或任何形式的重启。
其代价是这样一条维护要求：数据库中的每个表必须每至少 10 亿个事务清理一次。

实践中这并不是一个苛刻的要求，
但由于未能满足它的后果可能是彻底的数据丢失（而不只是浪费磁盘空间或性能变慢），
因此专门做了一些规定来帮助数据库管理员避免灾难。对于集簇中的每个数据库，
PostgreSQL会跟踪上一次数据库范围VACUUM的时间。当任何数据库接近 10 亿事务的危险水平时，系统将开始发出警告消息。如果不采取任何措施，它最终会关闭正常操作，
直到完成适当的手工维护。本节其余部分给出具体细节。

对付 XID 比较的新方法区分两个特殊的 XID，即 1 号和 2 号（BootstrapXID和FrozenXID）。这两个 XID 总是被认为比任何正常 XID 都旧。正常 XID（大于 2 的那些）使用模 2^{31}算术进行比较。这意味着对每个正常 XID 来说，都有 20 亿个“更旧”的 XID 和 20 亿个“更新”的 XID；换一种说法，正常 XID 空间是环形的，没有端点。因此，一旦一个行版本以某个特定的正常 XID 创建，那么无论我们谈论的是哪个正常 XID，在接下来的 20 亿个事务里，该行版本都会显得处于“过去”。如果行版本在超过 20 亿个事务之后仍然存在，它就会突然显得处于未来。为防止这一点，旧行版本必须在达到 20 亿个事务这个界限之前的某个时刻被重新指派 XIDFrozenXID。一旦被指派了这个特殊的 XID，无论回卷问题如何，它们对所有正常事务来说都处于“过去”，因此这样的行版本在被删除之前一直有效，无论时间多久。这种 XID 的重新指派由VACUUM处理。

VACUUM的常规策略是，把正常 XID 距今超过 10 亿个事务的所有行版本都重新指派为FrozenXID。这一策略会在原始插入 XID 不再可能令人感兴趣之前保留它。（实际上，大多数行版本可能从生到死都从未被“冻结”过。）按照这一策略，任何表上两次VACUUM
之间的最大安全间隔正好是 10 亿个事务：如果等待更久，上一次还不够旧、未被重新指派的某个行版本现在可能已超过 20 亿个事务的年龄，从而回卷到了未来——也就是对你而言丢失了。（当然，再经过 20 亿个事务它还会重新出现，但那于事无补。）

由于前述原因，本来就需要周期性地运行VACUUM，因此任何表不太可能长达 10 亿个事务都没有被清理。但为了帮助管理员确保满足这一约束，VACUUM在系统表pg_database中存储事务 ID 统计信息。特别地，任何一次数据库范围的VACUUM操作（即不指定具体表的VACUUM）完成时，都会更新该数据库pg_database行中的datfrozenxid列。该字段中存储的值就是那次VACUUM命令所使用的冻结截止 XID。比该截止 XID 更旧的所有正常 XID 都保证已在该数据库内被替换为FrozenXID。查看这些信息的一个便捷方法是执行查询：

```
SELECT datname, age(datfrozenxid) FROM pg_database;
```

age列度量从截止 XID 到当前事务 XID 的事务数。

按照标准的冻结策略，一个刚完成清理的数据库，其age列将从 10 亿开始。当age接近 20 亿时，必须再次对该数据库进行清理，以避免回卷失败的风险。推荐的做法是每个数据库至少每 5 亿个事务执行一次VACUUM，以留出充足的安全余量。为帮助遵守这一规则，任何一次数据库范围的VACUUM都会在存在pg_database条目显示出超过 15 亿事务的age时自动发出警告，例如：

```
play=# VACUUM;
WARNING:  database "mydb" must be vacuumed within 177009986
 transactions
HINT:  To avoid a database shutdown, execute a full-database VACUUM in
 "mydb".
VACUUM
```

如果忽略VACUUM发出的这些警告，那么当距回卷只剩不到 1000 万个事务时，PostgreSQL将开始在每次事务启动时发出类似上面的警告。如果这些警告也被忽略，那么当距回卷只剩不到 100 万个事务时，系统将关闭并拒绝执行任何新事务：

```
play=# select 2+2;
ERROR:  database is shut down to avoid wraparound data loss in
 database "mydb"
HINT:  Stop the postmaster and use a standalone backend to VACUUM in
 "mydb".
```

设置 100 万个事务的安全余量，是为了让管理员能够通过手工执行所需的VACUUM命令在不丢失数据的情况下恢复。然而，由于系统一旦进入安全关闭模式就不再执行命令，唯一的办法是停止 postmaster 并使用独立后端执行VACUUM。独立后端不强制执行关闭模式。关于使用独立后端的细节，见postgres参考页。

带FREEZE选项的VACUUM
使用更激进的冻结策略：只要行版本旧到被所有打开的事务都认为是好的，就会被冻结。
特别地，在一个除此之外空闲的数据库中执行VACUUM FREEZE，可以保证该数据库中的所有行版本都被冻结。因此，只要该数据库不以任何方式被修改，
它就不需要再进行后续的清理来避免事务 ID 回卷问题。这一技术被initdb用来准备template0数据库。它也应当用于准备任何要在pg_database中标记为datallowconn = false的用户创建数据库，因为对于无法连接的数据库，并没有方便的办法执行VACUUM。

> **警告**
>
> 在pg_database中被标记为datallowconn = false的数据库被假定为已正确冻结；自动警告和回卷保护关闭不会考虑此类数据库。因此，在把数据库标记为datallowconn = false之前，必须由你自己确保已正确冻结了它。

22.1.4. 自动清理守护进程

从PostgreSQL 8.1 开始，有了一个独立的可选服务器进程，称为自动清理守护进程，其目的是自动执行VACUUM和ANALYZE命令。启用后，自动清理守护进程会周期性地运行，
检查那些已经积累了大量插入、更新或删除元组的表。这些检查使用行级统计收集功能；因此，
除非stats_start_collector和stats_row_level都被设置为true，
否则无法使用自动清理守护进程。另外，在选取superuser_reserved_connections的值时，
为自动清理进程保留一个连接槽位也很重要。

自动清理守护进程在启用后，每隔autovacuum_naptime秒运行一次，
并确定要处理哪个数据库。任何接近事务 ID 回卷的数据库都会被立即处理。在这种情况下，自动清理会发出一次数据库范围的VACUUM调用；如果是一个模板数据库，则发出VACUUM FREEZE，然后终止。如果没有数据库满足这一标准，则选择最久未被自动清理处理过的那个数据库。
在这种情况下，会检查所选数据库中的每个表，并根据需要发出各自的VACUUM或ANALYZE
命令。

对于每个表，将根据两个条件来决定要应用哪些操作。如果自上次 VACUUM 以来失效的元组数超过“清理阈值”，就会清理该表。清理阈值定义为：

```
vacuum threshold = vacuum base threshold + vacuum scale factor *
 number of tuples
```

其中，清理基础阈值为autovacuum_vacuum_threshold，清理比例因子为autovacuum_vacuum_scale_factor，元组数为 pg_class.reltuples。失效元组数来自统计收集器；这是一个近似计数，由每次 UPDATE 和 DELETE 操作更新。（之所以只是近似值，是因为高负载下可能丢失部分信息。）对于分析操作，也使用了一个类似的条件：其阈值定义如下：

```
analyze threshold = analyze base threshold + analyze scale factor *
 number of tuples
```

该阈值会与自上次 ANALYZE 以来插入、更新或删除的元组总数进行比较。

默认阈值和尺度因子取自 postgresql.conf，但也可以通过在系统目录pg_autovacuum中创建条目来按表覆盖它们。如果某个特定表存在 pg_autovacuum 行，则应用该行指定的设置；否则使用全局设置。关于全局设置的更多细节，见第 17.9 节。

除了基础阈值和缩放因子之外，还有三个可以通过 pg_autovacuum 为每个表设置的参数。第一个是 pg_autovacuum.enabled，可以把它设置为false，指示自动清理守护进程完全跳过该特定表。这种情况下，只有在自动清理处理整个数据库时，它才会触及该表。另外两个参数，
即清理代价延迟（pg_autovacuum.vac_cost_delay）和清理代价限制（pg_autovacuum.vac_cost_limit），用于为基于代价的清理延迟特性设置表特定的值。

如果 pg_autovacuum 中的任何值被设置为负数，或者某个特定表在 pg_autovacuum 中根本没有对应的行，则使用 postgresql.conf 中相应的值。

目前还不支持以任何方式创建 pg_autovacuum 条目，只能手工向该目录执行 INSERT。这一特性在将来的发行版中会得到改进，而且该目录的定义也很可能会改变。

> **小心**
>
> pg_autovacuum 系统目录的内容目前不会被 pg_dump 和 pg_dumpall 工具保存到数据库转储中。如果你希望在转储/重新装载的循环中保留它们，请务必手工转储该目录。

22.2. 日常重建索引

在某些情况下，值得周期性地使用REINDEX命令来重建索引。

在 7.4 之前的 PostgreSQL 版本中，由于 B-树索引缺乏内部空间回收机制，经常需要定期重建索引以避免“索引膨胀”。任何索引键范围随时间推移而变化的情况——例如，
在一个最终会删除旧条目的表上按时间戳建的索引——都会导致膨胀，
因为键范围中不再需要的部分所对应的索引页不会被回收以供重用。随着时间的推移，
索引大小可能会变得远大于其中有用的数据量。

在 PostgreSQL 7.4 及之后的版本中，已经完全变空的索引页会被回收以供重用。不过，仍然可能出现空间利用低效：如果某个页面上的索引键除了少数几个之外都被删除，
该页面仍会保留分配。因此，在一种使用模式下，
如果每个键范围中的大多数但不是全部键最终都被删除，就会出现糟糕的空间利用。
膨胀的可能性不是无限的——最坏情况下每个索引页只有一个键——但对于有这种使用模式的索引，定期安排重建索引仍可能是值得的。

对于非 B-树索引中的膨胀潜力，人们还没有做过充分研究。在使用任何非 B-树索引类型时，定期监控索引的物理大小是个好主意。

此外，对 B-树索引而言，新近构建的索引比已经多次更新过的索引在访问时稍快一些，因为在新建的索引中，逻辑上相邻的页面通常在物理上也彼此相邻。（这一考虑不适用于非 B-树索引。）
即便只是为了提高访问速度，周期性重建索引也可能是值得的。

22.3. 日志文件维护

最好把数据库服务器的日志输出保存到某个地方，而不是简单地把它丢弃到 `/dev/null`。诊断问题时，日志输出极其宝贵。但是，日志输出往往十分庞大（尤其是在较高调试级别下），
因此你不会希望无限期保存它。你需要对日志文件进行“轮转”，
以便周期性地启用新的日志文件，并在合理时间后移除旧文件。

如果你只是把 `postmaster` 的 stderr 重定向到一个文件，那么虽然也能得到日志输出，但截断该日志文件的唯一办法是停止并重新启动服务器。如果你是在开发环境中使用 PostgreSQL，这也许可以接受，但几乎没有生产环境的 `postmaster` 会认为这种行为可以接受。

更好的做法是把 `postmaster` 的stderr输出发送给某种日志轮转程序。有一个内建的日志轮转程序，只需在`postgresql.conf`中把配置参数`redirect_stderr`设置为`true`即可使用。
该程序的控制参数在第 17.7.1 节中介绍。

也可以使用外部日志轮转程序，特别是已经在其他服务器软件上使用了某个此类程序时。例如，rotatelogs 工具包含在 Apache 发行版中，也可用于 PostgreSQL。为此，只需通过管道把 `postmaster` 的 stderr 输出传递给所需程序。如果使用 `pg_ctl` 启动服务器，那么 stderr 已经被重定向到 stdout，因此只需一个管道命令，例如：

```
pg_ctl start | rotatelogs /var/log/pgsql_log 86400
```

另一种生产级的管理日志输出的方法，是把它发送给syslog，让syslog处理文件轮转。为此，在`postgresql.conf`中把配置参数`log_destination`设置为`syslog`（仅记录到syslog）。
然后，每当你想强制syslog守护进程开始写一个新的日志文件时，可以向它发送`SIGHUP`信号。
如果想自动化日志轮转，可以把logrotate程序配置为与syslog的日志文件配合工作。

不过，在很多系统上，syslog 并不十分可靠，尤其是在日志消息很大时；它可能恰恰在你最需要消息的时候截断或丢弃它们。此外，在 Linux 上，syslog 会把每条消息都刷盘，导致性能较差。（你可以在 syslog 配置文件的文件名开头使用一个 - 来禁用这一行为。）

请注意，上面描述的所有方案都处理了按可配置间隔启动新日志文件的问题，
但它们并不处理删除那些已经没有用处的旧日志文件。你很可能还需要设置一个批处理任务，
定期删除旧日志文件。另一种可能是把轮转程序配置为循环覆盖旧日志文件。

第 23 章 备份和恢复

与任何保存重要数据的系统一样，PostgreSQL数据库也应定期备份。虽然其过程基本简单，但清楚理解其底层技术和前提假设非常重要。

备份PostgreSQL数据主要有三种根本不同的方法：

- SQL转储
- 文件系统级备份
- 在线备份

每种方法都有各自的优缺点，下面各节将依次讨论。

23.1. SQL转储

SQL 转储方法的思路是生成一个包含 SQL 命令的文本文件；将其重新送回服务器后，服务器就会把数据库重建为转储时的状态。PostgreSQL为此提供了工具程序pg_dump。该命令的基本用法是：

```
pg_dump dbname > outfile
```

如你所见，pg_dump把结果写到标准输出。我们将在下面看到这样做的用处。

pg_dump是一个普通的PostgreSQL客户端应用程序（尽管它实现得相当巧妙）。这意味着你可以从任何能够访问该数据库的远程主机执行这种备份过程。但请记住，pg_dump并不会以特殊权限运行。特别是，它必须对你想要备份的所有表具有读权限，因此实践中你几乎总是必须以数据库超级用户身份运行它。

要指定pg_dump应连接哪个数据库服务器，请使用命令行选项`-h host`和`-p port`。默认主机是本地主机，或者由环境变量`PGHOST`指定的主机。类似地，默认端口由环境变量`PGPORT`指定；如果未设置，则采用编译时内置的默认值。（方便的是，服务器通常也会使用同样的内置默认值。）

像任何其他PostgreSQL客户端应用程序一样，pg_dump默认会以与当前操作系统用户名相同的数据库用户名进行连接。若要覆盖这一点，可以指定`-U`选项，或者设置环境变量`PGUSER`。请记住，pg_dump连接同样受常规客户端认证机制约束（见第 20 章）。

由pg_dump创建的转储在内部是一致的，也就是说，在pg_dump运行期间对数据库的更新不会出现在转储中。pg_dump在工作期间不会阻塞数据库上的其他操作。（例外是那些需要独占锁的操作，例如`VACUUM FULL`。）

> **重要**
>
> 当你的数据库模式依赖于 OID（例如用作外键）时，就必须指示pg_dump同时转储 OID。为此，请使用`-o`命令行选项。

23.1.1. 恢复转储

由pg_dump创建的文本文件供psql程序读取。恢复转储的通用命令格式是

```
psql dbname < infile
```

其中*infile*就是你在pg_dump命令中作为*outfile*使用的文件。*dbname*
数据库不会被这条命令创建，因此你必须先从`template0`自行创建它，然后再执行psql（例如用`createdb -T template0 dbname`). psql 支持与pg_dump类似的用于指定要连接的数据库服务器和所用用户名的选项。更多信息见psql参考页。

不仅目标数据库必须在开始运行恢复之前已经存在，
被转储数据库中拥有对象或曾被授予对象权限的所有用户也必须已经存在。否则，
恢复将无法以原有所有权和/或权限重新创建这些对象。（有时这正是你想要的，但通常不是。）

pg_dump和psql能够写入或读取管道，
这使得可以把数据库直接从一台服务器转储到另一台服务器，例如：

```
pg_dump -h host1 dbname | psql -h host2 dbname
```

> **重要**
>
> pg_dump生成的转储是以`template0`为基准的。这意味着通过`template1`
> 增加的任何语言、过程等也都会被pg_dump转储。因此，
> 恢复时如果你使用的是定制过的`template1`，就必须像上面的例子那样，从`template0`
> 创建空数据库。

恢复完成后，最好对每个数据库运行ANALYZE，这样优化器就能拥有有用的统计信息。
一个简便的方法是运行`vacuumdb -a -z`对所有数据库做`VACUUM ANALYZE`；这等价于手动运行`VACUUM ANALYZE`。

关于如何高效地将大量数据装入PostgreSQL的建议，请参见第 13.4 节。

23.1.2. 使用pg_dumpall

上述机制在备份整个数据库集簇时显得笨拙而不合适。为此提供了pg_dumpall程序。pg_dumpall会备份给定集簇中的每个数据库，并保留用户和组等集簇范围的数据。
该命令的基本用法是：

```
pg_dumpall > outfile
```

得到的转储可以用psql恢复：

```
psql -f infile postgres
```

（实际上，你可以指定任意现有的数据库名作为起始，但如果是载入空集簇，通常应使用`postgres`。）恢复pg_dumpall转储时始终需要数据库超级用户权限，
因为恢复用户和组信息需要这种权限。

23.1.3. 处理大型数据库

由于PostgreSQL允许表大于系统上的最大文件大小，把这样一张表转储到文件可能会产生问题，
因为得到的文件很可能超过系统允许的最大尺寸。幸运的是，pg_dump可以把输出写到标准输出，因此你只需使用标准 Unix 工具就可以绕过这个潜在问题。

使用压缩转储。. 你可以使用自己喜欢的压缩程序，例如gzip。

```
pg_dump dbname | gzip > filename.gz
```

恢复时用

```
createdb dbname
gunzip -c filename.gz | psql dbname
```

或者

```
cat filename.gz | gunzip | psql dbname
```

使用split。. split命令允许你把输出分割成文件系统可接受大小的较小文件。例如，要切分为 1 MB 的块：

```
pg_dump dbname | split -b 1m - filename
```

恢复时用

```
createdb dbname
cat filename* | psql dbname
```

使用自定义转储格式。. 如果构建PostgreSQL所用的系统安装了zlib压缩库，自定义转储格式会在向输出文件写入数据时对其进行压缩。这会产生与使用gzip大致相同大小的转储文件，但额外的优点是其中的表可以有选择地恢复。下面的命令使用自定义转储格式转储数据库：

```
pg_dump -Fc dbname > filename
```

自定义格式的转储不是供psql执行的脚本，而必须通过pg_restore恢复。详细信息见pg_dump和pg_restore参考页。

23.2. 文件系统级备份

另一种备份策略是直接复制PostgreSQL用来存储数据库数据的文件；第 16.2 节解释了这些文件的位置。你可以使用自己喜欢的任何方法进行文件系统备份，例如：

```
tar -cf backup.tar /usr/local/pgsql/data
```

不过，这种方法有两个限制，使得它并不实用，至少也不如pg_dump方法：

1. 为了得到可用的备份，数据库服务器必须关闭。
   像禁止所有连接这样的折中办法不起作用（主要是因为tar和类似工具不会对文件系统在某一时刻的状态执行原子快照）。有关停止 postmaster 的信息见第 16.5 节。不言而喻，恢复数据前也必须关闭服务器。

2. 如果你已经深入了解数据库文件系统布局的细节，
   可能会想尝试仅通过某些表或数据库各自的文件或目录来备份或恢复它们。这是行不通的，因为这些文件中的信息离开记录所有事务提交状态的提交日志文件pg_clog/*就无法使用。一个表文件只有和这些信息一起才可用。当然，也不可能只恢复一个表以及相关的pg_clog数据，因为那会让数据库集簇中的其他所有表都无法使用。因此，文件系统备份只适用于整个数据库集簇的完整备份与恢复。

另一种文件系统备份方法是对数据目录制作一个“一致快照”，前提是文件系统支持该功能（并且你愿意相信其实现是正确的）。典型流程是：对包含数据库的卷制作一个“冻结快照”，然后从该快照把整个数据目录（不是其中一部分，见上文）复制到备份设备，再释放这个冻结快照。即使数据库服务器正在运行，这种方法也能工作。不过，以这种方式创建的备份会把数据库文件保存为一种仿佛服务器没有被正确关闭的状态；因此，当你基于备份数据启动数据库服务器时，它会认为前一个服务器实例发生了崩溃，并重放 WAL 日志。这不是问题；只要意识到这一点即可（并且务必把 WAL 文件也包含在备份中）。

如果数据库分布在多个文件系统上，就可能根本无法对所有卷获得完全同步的冻结快照。例如，如果数据文件和 WAL 日志位于不同磁盘，或者表空间位于不同文件系统上，就可能无法使用快照备份，因为这些快照必须同时获取。在这种情况下，在信任一致快照技术之前，一定要非常仔细地阅读文件系统文档。最安全的做法是把数据库服务器关闭足够长的时间，以建立所有冻结快照。

另一种选择是使用rsync进行文件系统备份。做法是先在数据库服务器运行时执行一次rsync，然后将数据库服务器只关闭足够再做一次rsync的时间。第二次rsync会比第一次快得多，因为它需要传输的数据相对较少，而最终结果会因为服务器已经关闭而保持一致。这种方法可以在最小停机时间下完成文件系统备份。

注意，文件系统备份通常会比 SQL 转储更大。（例如，pg_dump不需要转储索引内容，只需转储重建索引的命令。）

23.3. 在线备份和时间点恢复（PITR）

在任何时候，PostgreSQL都会维护一个预写式日志（WAL），它位于集簇数据目录的`pg_xlog/`子目录中。该日志记录对数据库数据文件所做的每一次修改。这个日志首先是为崩溃安全而存在：如果系统崩溃，可以通过“重放”自上次检查点以来的日志记录，将数据库恢复到一致状态。不过，日志的存在也使第三种数据库备份策略成为可能：我们可以把文件系统级备份与 WAL 文件备份结合起来。如果需要恢复，就先恢复文件系统备份，再重放已备份的 WAL 文件，使系统回到当前状态。与前两种方法相比，这种方法管理起来更复杂，但它有一些显著优点：

- 我们不需要一个完全一致的文件系统备份作为起点。备份中的任何内部不一致性都会通过日志重放纠正（这与崩溃恢复期间发生的事情并没有本质不同）。因此，我们不需要文件系统快照能力，只需要tar或类似的归档工具。

- 由于我们可以把任意长的一串 WAL 文件串联起来进行重放，只要持续归档 WAL 文件，就可以实现连续备份。这对于大型数据库尤其有价值，因为此时频繁进行完整备份可能并不方便。

- 不必把 WAL 记录一直重放到最后。我们可以在任意位置停止重放，并得到数据库在当时的一致快照。因此，这项技术支持时间点恢复：可以把数据库恢复到自基础备份之后任意时刻的状态。

- 如果我们持续把这一串 WAL 文件输送给另一台已经装载了同一基础备份文件的机器，就得到了一个“热后备”系统：我们随时都可以启用第二台机器，而它将拥有一个几乎是最新的数据库副本。

和普通文件系统备份技术一样，这种方法只能支持整个数据库集簇的恢复，而不支持其中某个子集的恢复。此外，它需要大量归档存储：基础备份可能很庞大，繁忙系统也会产生许多兆字节的、必须归档的 WAL 流量。尽管如此，在很多需要高可靠性的场景中，它仍是首选的备份技术。

要想利用在线备份成功恢复，你需要一串连续的已归档 WAL 文件，其时间至少要追溯到备份的开始时刻。因此，入门时应在进行第一次基础备份之前就建立并测试好归档 WAL 文件的流程。下面先讨论归档 WAL 文件的机制。

23.3.1. 设置 WAL 归档

从抽象角度看，运行中的 PostgreSQL 系统会产生一个无限延伸的 WAL 记录序列。在物理存储上，系统将该序列分为 WAL 段文件，通常每个为 16MB（但可以在构建 PostgreSQL 时修改段大小）。段文件使用数字名称，反映其在抽象 WAL 序列中的位置。不使

用 WAL 归档时，系统通常只创建少量段文件，然后通过将不再需要的段文件重命名为更大的段号来“回收”它们。系统假定，内容早于倒数第二个检查点的段文件已不再需要，可以回收。

在归档 WAL 数据时，我们需要在每个段文件写满后捕获其内容，并在该段文件被回收重用之前把数据保存到某处。根据应用场景和可用硬件的不同，“把数据保存到某处”
可以有很多不同的方法：可以把段文件复制到另一台机器上的 NFS 挂载目录，把它们写到磁带机中（确保你有办法识别每个文件的原始文件名），把它们批量打包后刻录到 CD 上，或者采用完全不同的方式。为了给数据库管理员提供灵活性，PostgreSQL尽量不对归档方式作任何假设。
相反，PostgreSQL允许管理员指定一个 shell 命令，以便把已完成的段文件复制到它应去的地方。该命令可以只是一次简单的cp调用，也可以调用一个复杂的 shell 脚本，全由你决定。

要使用的 shell 命令由archive_command配置参数指定，在实践中它总是会放在`postgresql.conf`文件中。在这个字符串中，`%p`会被替换为要归档文件的路径名，而`%f`只被替换为文件名。（该路径名相对于当前工作目录，即集簇的数据目录。）如果需要在命令中嵌入真正的`%`字符，请使用`%%`。最简单可用的命令形如：

```
archive_command = 'cp -i %p /mnt/server/archivedir/%f </dev/null'
```

它会把可归档的 WAL 段复制到目录`/mnt/server/archivedir`。（这只是一个例子，并非建议，而且未必在所有平台上都能工作。）

归档命令将以运行 PostgreSQL 服务器的同一用户身份执行。由于归档的一系列 WAL 文件实际上包含数据库中的全部内容，应确保归档数据不会被他人窥视；例如，
将其归档到不允许所属组或其他用户读取的目录。

归档命令必须当且仅当成功时才返回退出状态零。收到零状态后，PostgreSQL 会认为该文件已成功归档，并将其删除或回收。非零状态则告诉 PostgreSQL 该文件尚未归档；它会定期重试，直到成功。

通常应将归档命令设计为拒绝覆盖任何已存在的归档文件。这是一项重要的安全保护措施，
能在管理员操作失误时（例如将两台不同服务器的输出发送到同一归档目录）
维护归档的完整性。建议测试所拟定的归档命令，确保它确实不会覆盖已有文件，
并且在这种情况下返回非零状态。我们发现`cp -i`在某些平台上能正确做到这一点，但在另一些平台上不行。如果所选命令本身不能正确处理这种情况，
应该增加一个测试归档文件是否已存在的命令。例如，类似下面这样的写法在大多数 Unix 变体上都能正确工作：

```
archive_command = 'test ! -f .../%f && cp %p .../%f'
```

在设计归档方案时，请考虑如果归档命令因为某些环节需要操作员干预，
或者归档空间耗尽而反复失败，会发生什么。例如，
如果你在没有自动换带器的情况下向磁带写入数据，那么磁带满了以后，
在更换磁带之前将无法继续归档。
你应确保任何错误情况或对人工操作员的请求都能得到适当报告，以便问题能够比较快地解决。
在问题解决之前，`pg_xlog/`目录会继续堆积 WAL 段文件。

归档命令的速度并不重要，只要它能跟上服务器生成 WAL 数据的平均速度即可。即使归档过程稍有滞后，正常操作也会继续。如果归档明显落后，灾难发生时可能丢失的数据量就会增加。
这还意味着`pg_xlog/`目录会包含大量尚未归档的段文件，最终可能耗尽可用磁盘空间。
建议监控归档过程，确保它按你的预期工作。

在编写归档命令时，应假定待归档文件名最长可达 64 个字符，并且可以包含 ASCII 字母、数字和点号的任意组合。无需保留原始完整路径（`%p`），但必须保留文件名（`%f`）。

注意，虽然 WAL 归档允许你恢复对PostgreSQL数据库中数据所做的任何修改，但它不会恢复对配置文件（即`postgresql.conf`、`pg_hba.conf`和`pg_ident.conf`）的修改，

因为这些文件是手工编辑的，而不是通过 SQL 操作修改的。你可能希望把配置文件放在常规文件系统备份过程能够覆盖的位置。关于如何重定位配置文件，见第 17.2 节。

如果你关心能否一直恢复到当前时刻，可能还需要采取一些额外步骤，确保当前尚未填满的 WAL 段也被复制到别处保存。如果你的服务器产生的 WAL 流量很小（或者存在这样的低谷期），这一点尤其重要，因为一个 WAL 段文件可能要过很久才会被完全填满并准备好归档。一种可行的处理方法是设置一个cron作业，让它周期性地（比如每分钟一次）找出当前的 WAL 段文件并把它保存到安全的地方。这样，已归档的 WAL 段加上保存的当前段，就足以确保你总能恢复到距当前时间一分钟之内的状态。这一行为目前并未内置于PostgreSQL，因为我们不想让archive_command的定义复杂化——那将要求它跟踪同一 WAL 文件先后归档出的多个不同副本。archive_command只对已完成的 WAL 段调用。除非是重试一次失败，否则对任何给定的文件名它只会被调用一次。

23.3.2. 进行基础备份

制作基础备份的过程相当简单：

1. 确保 WAL 归档已启用并且正常工作。
2. 以超级用户身份连接到数据库，并发出命令：

   ```
   SELECT pg_start_backup('label');
   ```

   其中`label`是任意你想用来唯一标识这次备份操作的字符串。（一种好的做法是使用你打算存放备份转储文件的完整路径。）`pg_start_backup`会在集簇目录中创建一个备份标签文件，名为`backup_label`，其中包含关于备份的信息，包括开始时间和标签字符串。

   连接到集簇中的哪个数据库来发出这条命令无关紧要。你可以忽略函数返回的结果；但如果它报告错误，请先处理错误再继续。
3. 使用任何方便的文件系统备份工具（如tar或cpio）执行备份。在此期间既没有必要也不应该停止数据库的正常运行。
4. 再次以超级用户身份连接到数据库，并发出命令：

   ```
   SELECT pg_stop_backup();
   ```

   这条命令应当成功返回。
5. 备份期间使用的 WAL 段文件作为正常数据库活动的一部分归档完成后，备份就完成了。

有些文件系统备份工具在要复制的文件于复制期间发生变化时会发出警告或错误。对活动数据库制作基础备份时，这种情况是正常的，并不是错误。不过，你需要确保能把这类抱怨与真正的错误区分开。例如，某些版本的rsync会对“消失的源文件”返回单独的退出码，你可以编写驱动脚本把这种退出码当作非错误情况接受。另外，某些版本的GNU tar在文件于tar复制它期间被更改时也会视为错误。除了人工检查tar的输出消息之外，似乎没有很方便的办法把这种错误与其他类型的错误区分开。因此，GNU tar并不是制作基础备份的最佳工具。

不必担心`pg_start_backup`与实际备份开始之间、或备份结束与`pg_stop_backup`之间流逝的时间；延迟几分钟不会有任何损害。不过，你必须相当确定这些操作是按顺序进行的，并且没有重叠。

务必确认你的备份包含数据库集簇目录（例如`/usr/local/pgsql/data`）下的全部文件。如果你使用的表空间不位于该目录之下，也要记得把它们包括进来（并确保备份把符号链接作为链接归档，否则恢复时会破坏表空间）。

不过，可以从备份转储中省略集簇pg_xlog/子目录中的文件。这一小调整是值得的，因为它降低了恢复时出错的风险。如果pg_xlog/是指向集簇目录之外某个位置的符号链接，这就很容易安排——出于性能原因，这本来就是常见的设置。

要利用这份备份，你需要保留文件系统备份期间及其后生成的所有 WAL 段文件。为帮助你做到这一点，pg_stop_backup函数会立即创建一个备份历史文件并存入 WAL 归档区。该文件以文件系统备份所需的第一个 WAL 段文件命名。例如，如果起始 WAL 文件是000000010000123400005 5CD，则备份历史文件的名称将类似于000000010000123400005 5CD.007C9330.backup。（文件名的第二部分表示 WAL 文件内的一个精确位置，通常可以忽略。）一旦你安全归档了文件系统备份以及备份期间使用的 WAL 段文件（如备份历史文件所指定的），所有名称编号更小的已归档 WAL 段就不再需要用于恢复该文件系统备份，可以删除。不过，为绝对确保能恢复数据，应考虑保留几套备份。请记住，只有已完成的 WAL 段文件才会被归档，因此从运行pg_stop_backup到所有使文件系统备份保持一致所需的 WAL 段文件全部归档之间，会有一段延迟。

备份历史文件只是一个很小的文本文件。它包含你给pg_start_backup的标签字符串，以及备份的开始和结束时间与 WAL 段。如果你用标签来标识关联的转储文件，那么已归档的历史文件就足以告诉你应恢复哪个转储文件。

由于你必须保留自上一次基础备份以来的所有已归档 WAL 文件，基础备份之间的间隔通常应根据你愿意为已归档 WAL 文件投入多少存储空间来决定。你还应考虑在确实需要恢复时，你愿意花多长时间进行恢复，因为系统必须重放所有这些 WAL 段；如果距离上次基础备份已经过去很久，这可能会花费一些时间。

还值得注意的是，pg_start_backup函数会在数据库集簇目录中创建一个名为backup_label的文件，该文件由pg_stop_backup删除。这个文件当然会作为备份转储文件的一部分被归档。备份标签文件包含你给pg_start_backup的标签字符串、pg_start_backup运行的时间，以及起始 WAL 文件的名称。因此，在困惑时可以查看备份转储文件的内部，确定该转储文件究竟来自哪次备份会话。

postmaster 停止时也可以进行备份转储。在这种情况下，你显然无法使用pg_start_backup或pg_stop_backup，因此只能自己追踪各个备份转储的身份以及相关 WAL 文件最早需要追溯到哪里。通常最好还是遵循上面的在线备份过程。

23.3.3. 使用在线备份进行恢复

现在，最坏的情况发生了，你需要通过备份进行恢复。步骤如下：

1. 如果 postmaster 还在运行，就先停止它。

2. 如果你有足够空间，请把整个集簇数据目录以及所有表空间复制到临时位置，以备后续需要。注意，这一预防措施要求系统有足够空闲空间来保存现有数据库的两份副本。如果空间不够，至少也应保存集簇pg_xlog子目录中的内容，因为其中可能包含系统停机前尚未归档的 WAL 文件。

3. 删除集簇数据目录下以及所有正在使用的表空间根目录下的现有文件和子目录。

4. 从你的备份转储中恢复数据库文件。务必确保它们以正确的所有者（数据库系统用户，而不是 root！）和正确的权限恢复。如果使用了表空间，你可能还需要验证pg_tblspc/中的符号链接是否已正确恢复。

5. 删除pg_xlog/中现有的所有文件；这些文件来自文件系统备份，因此很可能已经过时而不是最新的。如果你根本没有归档pg_xlog/，那么就重新创建它，

并注意如果它原先是符号链接，就要重新把它设置成符号链接。此外，务必也重新创建子目录pg_xlog/archive_status/。

6. 如果你手头还有第 2 步中保存下来的未归档 WAL 段文件，请把它们复制到pg_xlog/中。（最好复制，而不是移动，这样一旦出问题需要重来时，你手里仍然保留着未修改的原始文件。）

7. 在集簇数据目录中创建恢复命令文件 recovery.conf（参见Recovery Settings）。也可以临时修改 pg_hba.conf，在确认恢复成功之前阻止普通用户连接。

8. 启动 postmaster。postmaster 会进入恢复模式，依次读取所需的已归档 WAL 文件。恢复过程完成后，postmaster 会将 recovery.conf 重命名为 recovery.done（防止以后意外重新进入恢复模式），然后开始正常的数据库操作。

9. 检查数据库内容，确认你已经恢复到期望状态。如果不是，就回到第 1 步。如果一切正常，就把pg_hba.conf恢复为正常设置，让用户重新连接。

整个过程的关键，是设置一个恢复配置文件，说明希望如何恢复，以及恢复到什么位置。可以使用 recovery.conf.sample（通常位于安装目录的 share/ 目录中）作为模板。必须在 recovery.conf 中指定 restore_command，它告诉 PostgreSQL 如何获取已归档的 WAL 段文件。与 archive_command 一样，它是一个 shell 命令字符串，可以包含 %f，该标记会被替换为所需日志文件的名称；还可以包含 %p，该标记会被替换为日志文件要复制到的路径名。（路径名相对于服务器的工作目录，即集簇的数据目录。）使用 %% 可以在命令中嵌入实际的 % 字符。最简单的可用命令类似于：

```
restore_command = 'cp /mnt/server/archivedir/%f %p'
```

这会复制先前归档的 WAL 段，来源目录为 /mnt/server/archivedir。当然，也可以使用复杂得多的命令，甚至可以使用要求操作人员挂载相应磁带的 shell 脚本。

重要的是，该命令在失败时必须返回非零退出状态。该命令将被请求归档中不存在的文件；当这样被请求时，它必须返回非零。这不是错误情况。还要注意，%p路径的基本名会与%f不同；不要指望它们可以互换。

在归档中找不到的 WAL 段将在pg_xlog/中查找；这允许使用最近的未归档段。但是，归档中可用的段将优先于pg_xlog/中的文件使用。取回归档文件时，系统不会覆盖pg_xlog/中的现有内容。

通常，恢复会处理完所有可用的 WAL 段，从而把数据库恢复到当前时间点（或者在可用 WAL 段所允许的情况下尽可能接近当前时间点）。但如果你想恢复到之前的某个时间点（比如初级 DBA 删掉你的主事务表之前的那一刻），只需在recovery.conf中指定所需的停止点即可。可以用日期/时间或特定事务 ID 的完成来指定停止点（即所谓的“恢复目标”）。截至撰写本文时，只有日期/时间选项非常实用，因为还没有工具能帮你精确确定应使用哪个事务 ID。

> **注意**
>
> 停止点必须晚于基础备份的结束时间，也就是pg_stop_backup的结束时间。你不能用某次基础备份恢复到该备份仍在进行中的时间点。（若要恢复到这样的时间点，必须回到更早的一次基础备份，再从那里向前重放日志。）

23.3.3.1. 恢复设置

这些设置只能在 recovery.conf 文件中设置，并且只在恢复期间有效。要进行任何后续的恢复，必须重新设置它们。一旦恢复开始，它们就不能再更改。

`restore_command`(`string`)

用于检索 WAL 文件系列中已归档段的 shell 命令。此参数是必需的。字符串中的 `%f`会被替换为要从归档中检索的文件名，`%p`会被替换为服务器上复制到的路径名。（该路径名是相对于服务器的工作目录，即集簇的数据目录的。）写入 `%%` 可以在命令中嵌入一个实际的 `%` 字符。

当且仅当命令成功时返回零退出状态，这一点很重要。命令会被要求提供归档中不存在的文件名；遇到这种情况时它必须返回非零值。示例：

```
restore_command = 'cp /mnt/server/archivedir/%f "%p"'
restore_command = 'copy /mnt/server/archivedir/%f "%p"'  # Windows
```

`recovery_target_time`(`timestamp`)

此参数指定恢复将要推进到的时间戳。`recovery_target_time`和recovery_target_xid中至多只能指定一个。默认是恢复到 WAL 日志的末尾。确切的停止点还受recovery_target_inclusive的影响。

`recovery_target_xid`(`string`)

此参数指定恢复将要推进到的事务 ID。请记住，虽然事务 ID 是在事务开始时顺序分配的，但事务可能以不同的数字顺序完成。将被恢复的是在指定事务之前（可选地包括该事务）提交的那些事务。`recovery_target_xid`和recovery_target_time中至多只能指定一个。默认是恢复到 WAL 日志的末尾。确切的停止点还受recovery_target_inclusive的影响。

`recovery_target_inclusive`(`boolean`)

指定我们是恰好在指定的恢复目标之后停止（`true`），还是恰好在恢复目标之前停止（`false`）。适用于recovery_target_time和recovery_target_xid中为本次恢复指定的那一个。它表示提交时间或 ID 恰好等于目标值的事务是否会包含在恢复中。默认值为 `true`。

`recovery_target_timeline`(`string`)

指定恢复到一条特定的时间线。默认是沿着做基础备份时当前的那条时间线恢复。只有在复杂的再次恢复情形下（你需要返回到一个本身就是在一次时间点恢复之后达到的状态），才需要设置此参数。相关讨论见第 23.3.4 节。

23.3.4. 时间线

能够将数据库恢复到过去某个时间点，也会带来一些类似科幻故事中时间旅行和平行宇宙的复杂情况。例如，假设在数据库原来的历史中，你在周二下午 5:15 删除了一张重要的表。你从容地取出备份，将数据库恢复到周二下午 5:14，然后重新投入运行。在数据库宇宙的这段历史中，你从未删除过那张表。但假设你后来发现这样做不太合适，希望回到原来历史中稍后的某个时刻。如果数据库恢复运行后覆盖了通往该时刻所需的某些 WAL 段文件，就无法回去了。因此，为了避免这种情况，需要区分时间点恢复后产生的一系列 WAL 记录与数据库原来历史中产生的记录。

为解决这一问题，PostgreSQL 引入了时间线的概念。每当归档恢复完成，系统都会创建一条新时间线，用于标识此次恢复之后产生的一系列 WAL 记录。时间线 ID 是 WAL 段文件名的一部分，因此新时间线不会覆盖旧时间线产生的 WAL 数据。实际上，可以归档许多不同的时间线。这个功能看似没什么用，却常常能救急。比如，你不能确定应该恢复到哪个时间点，需要反复尝试时间点恢复，直到找到脱离旧历史的最佳分支点。没有时间线，这个过程很快就会乱得无法管理。有了时间线，就可以恢复到任何先前的状态，包括早先已放弃的时间线分支中的状态。

每当创建一条新的时间线时，PostgreSQL都会创建一个“时间线历史”文件，
记录它是从哪条时间线、在何时分叉出来的。当从包含多条时间线的归档中恢复时，
这些历史文件对于系统选取正确的 WAL 段文件是必需的。因此，它们会像 WAL 段文件一样被归档到 WAL 归档区域。历史文件只是很小的文本文件，因此长期保存它们既便宜也合适（而段文件通常很大）。如果你愿意，还可以在历史文件中加入注释，记录创建这条时间线的方式和原因。当你因实验而积累出一批错综复杂的时间线时，这类注释会特别有价值。

默认情况下，恢复会沿着制作基础备份时的当前时间线进行。
如果希望恢复到某条子时间线（即返回到一次恢复尝试之后产生的某个状态），需要在 `recovery.conf` 中指定目标时间线 ID。不能恢复到在基础备份之前就已分支出去的时间线。

23.3.5. 注意事项

在撰写本文时，在线备份技术还存在若干局限。这些问题很可能会在未来版本中得到修复：

- 哈希索引和 R 树索引上的操作目前不会写入 WAL 日志，因此重放不会更新这些索引类型。建议的解决办法是，在恢复操作完成后对每个这样的索引手工执行`REINDEX`。

- 如果在进行基础备份时执行了`CREATE DATABASE`命令，而该`CREATE DATABASE`所复制的模板数据库又在基础备份尚未结束时被修改，
  那么恢复时可能会把这些修改也传播到新建数据库中。这当然并不理想。为避免这种风险，
  最好在进行基础备份时不要修改任何模板数据库。

- `CREATE TABLESPACE`命令会以字面绝对路径写入 WAL，因此重放时会按相同的绝对路径创建表空间。如果 WAL 在另一台机器上重放，这可能并不理想。即使 WAL 在同一台机器上、但重放到新的数据目录中，也可能有危险：重放仍会覆盖原表空间的内容。为避免此类潜在陷阱，最佳做法是在创建或删除表空间后重新执行一次基础备份。

还应注意，默认的WAL格式相当臃肿，因为它包含许多磁盘页面快照。
这些页面快照是为支持崩溃恢复而设计的，因为我们可能需要修复部分写入的磁盘页。
根据你的系统硬件和软件情况，部分写入的风险可能小到可以忽略；在这种情况下，可以通过 full_page_writes参数关闭页面快照，从而显著减少已归档 WAL 文件的总量。（在这样做之前，请先阅读第 26 章中的说明和警告。）关闭页面快照并不妨碍把 WAL 用于 PITR 操作。未来一个可能的开发方向，是在`full_page_writes`开启的情况下，通过去除不必要的页面副本来压缩归档 WAL 数据。在此之前，管理员可以考虑尽可能增大检查点间隔相关参数，以减少 WAL 中包含的页面快照数量。

23.4. 版本间的迁移

本节讨论如何把数据库数据从一个PostgreSQL版本迁移到更新的版本。
软件安装过程本身不是本节的主题；相关细节见第 14 章。

一般而言，PostgreSQL的内部数据存储格式在大版本（第一个点号后的数字变化）
之间可能发生变化。同一大版本下的小版本（第二个点号后的数字变化）不受此影响；
它们总是具有兼容的存储格式。例如，8.1.1、8.2.3 和 8.3 互不兼容，而 8.2.3 和 8.2.4 是兼容的。在兼容的版本之间更新时，只需替换可执行文件并复用磁盘上的数据目录即可。否则就需要备份数据并在新服务器上恢复。这必须使用pg_dump来完成；
文件系统级的备份方法显然不行。
系统中有检查防止你在不兼容版本的PostgreSQL上使用数据目录，
因此在数据目录上尝试启动错误版本的服务器不会造成太大损害。

建议使用新版本PostgreSQL的pg_dump和pg_dumpall程序，
以利用这些程序中可能已实现的增强。当前版本的转储程序可以读取从 7.0 起任何服务器版本的数据。

停机时间最短的做法是把新服务器安装到不同目录，
并让旧服务器和新服务器在不同端口上并行运行。然后可以用类似下面的命令：

```
pg_dumpall -p 5432 | psql -d postgres -p 6543
```

来传输数据。如果愿意，也可以使用中间文件。然后可以关闭旧服务器，
并用旧服务器使用的端口启动新服务器。务必确保在开始运行pg_
dumpall之后旧数据库不再被更新，否则会丢失那些更新。有关如何禁止访问，见第 20 章。

如果无法或不想并行运行两个服务器，可以在安装新版本之前先执行备份步骤，
然后关闭旧服务器、把旧版本移开、安装新版本、启动新服务器并恢复数据。例如：

```
pg_dumpall > backup
pg_ctl stop
mv /usr/local/pgsql /usr/local/pgsql.old
cd ~/postgresql-8.1.23
gmake install
initdb -D /usr/local/pgsql/data
postgres -D /usr/local/pgsql/data
psql -f backup postgres
```

有关启动和停止服务器的方式及其他细节，见第 16 章。安装说明会提示执行这些步骤的合适时机。

注意

把旧安装“移开”后，它可能不再能完全正常使用。
有些可执行程序中包含指向各种已安装程序和数据文件的绝对路径。这通常不是大问题，
但如果计划并行使用两个安装一段时间，应在构建时给它们分配不同的安装目录。
（这个问题在PostgreSQL 8.0 及更高版本中已得到解决，但对移动更旧的安装需要小心。）

在实践中，在完全切换之前，你可能希望先在新版本上测试客户端应用程序。
这也是同时安装新旧版本的另一个理由。

第 24 章 监控数据库活动

数据库管理员常常会问："系统现在正在做什么？"本章将讨论如何回答这个问题。

有几种工具可用于监控数据库活动和分析性能。
本章的大部分内容将用于描述PostgreSQL的统计收集器，但也不应忽略常见的 Unix 监控程序，例如ps、top、iostat和vmstat。此外，一旦确定了某个查询性能不佳，可能需要使用PostgreSQL的EXPLAIN命令进行进一步调查。第 13.1 节讨论了EXPLAIN以及理解单个查询行为的其他方法。

24.1. 标准 Unix 工具

在大多数平台上，PostgreSQL会修改由ps报告的命令标题，以便轻松识别各个服务器进程。
以下是一个显示示例：

```
$ ps auxww | grep ^postgres
postgres   960  0.0  1.1  6104 1480 pts/1    SN   13:17   0:00
 postgres -i
postgres   963  0.0  1.1  7084 1472 pts/1    SN   13:17   0:00
 postgres: stats buffer process
postgres   965  0.0  1.1  6152 1512 pts/1    SN   13:17   0:00
 postgres: stats collector process
postgres   998  0.0  2.3  6532 2992 pts/1    SN   13:18   0:00
 postgres: tgl runbug 127.0.0.1 idle
postgres  1003  0.0  2.4  6532 3128 pts/1    SN   13:19   0:00
 postgres: tgl regression [local] SELECT waiting
postgres  1016  0.1  2.4  6532 3080 pts/1    SN   13:19   0:00
 postgres: tgl regression [local] idle in transaction
```

（ps的适当调用方式随平台而异，显示内容的细节也是如此。此示例来自较新的 Linux 系统。）
这里列出的第一个进程是postmaster，即主服务器进程。
显示的命令参数与启动该进程时使用的参数相同。接下来的两个进程实现统计收集器，
下一节将详细描述。（如果你已把系统设置为不启动统计收集器，这两个进程将不存在。）
其余每个进程都是处理一个客户端连接的服务器进程。
每个这样的进程都将其命令行显示设置为以下形式：

```
postgres: user database host activity
```

在客户端连接的整个生命周期内，用户、数据库和连接来源主机这几项保持不变，
但活动指示符会改变。活动可以是idle（即等待客户端命令）、idle in transaction（在BEGIN块内等待客户端），或者命令类型名称，例如SELECT。此外，
服务器进程当前若正在等待另一服务器进程持有的锁，会附上waiting。在上面的示例中，
可以推断进程 1003 正在等待进程 1016 完成事务，从而释放某个锁。

> **提示**
>
> Solaris需要特殊处理。您必须使用/usr/ucb/ps，而不是/bin/ps。
> 您还必须使用两个w标志，而不仅仅是一个。此外，您对postmaster
> 命令的原始调用必须具有比每个服务器进程提供的更短的ps状态显示。
> 如果未满足这三个条件，每个服务器进程的ps输出都会显示原始的postmaster命令行。

24.2. 统计收集器

PostgreSQL的统计收集器是一个支持收集和报告服务器活动信息的子系统。目前，收集器可以按磁盘块和单行两种方式统计对表和索引的访问。它还支持确定其他服务器进程当前正在执行的确切命令。

24.2.1. 统计信息收集配置

由于收集统计信息会增加查询执行的开销，因此可以配置系统是否收集信息。这由通常在`postgresql.conf`中设置的配置参数控制（有关设置配置参数的详细信息，请参阅第 17 章）。

参数stats_start_collector必须设置为`true`，统计收集器才会被启动。这是默认且被推荐的设置，但如果你对统计信息不感兴趣、想榨取每一点开销，也可以把它关掉。（不过节省的开销可能很小。）注意这个选项不能在服务器运行期间更改。

参数stats_command_string、stats_block_level和stats_row_level控制实际发送给收集器的信息量，因而决定了运行时开销的大小。它们分别决定服务器进程是否把其当前命令字符串、磁盘块级的访问统计和行级的访问统计发送给收集器。通常这些参数会设置在`postgresql.conf`中，以便它们适用于所有服务器进程，但也可以在单个会话中使用SET命令打开或关闭它们。（为防止普通用户隐藏其活动不被管理员发现，只有超级用户才能使用`SET`更改这些参数。）

> **注意**
>
> 由于参数`stats_command_string`、`stats_block_level`和 `stats_row_level`默认为`false`，默认配置下收集的统计信息很少。启用这些配置变量中的一个或多个，都会显著增加统计收集器产生的有用数据量，但代价是额外的运行时开销。

24.2.2. 查看收集到的统计信息

若干预定义视图（列于表 24.1中）可用于展示统计收集的结果。另外，也可以使用底层的统计函数构建自定义视图。

在使用统计信息监控当前活动时，重要的是要认识到这些信息不会即时更新。每个服务器进程在即将进入空闲之前把新的块和行访问计数传送给收集器；因此仍在进行中的查询或事务不会影响显示的总数。此外，收集器本身最多每`PGSTAT_STAT_INTERVAL`毫秒（除非构建服务器时更改，否则为 500）发出一次新报告。因此，显示的信息滞后于实际活动。当前查询的信息会立即报告给收集器，但在它变得可见之前，仍要经历`PGSTAT_STAT_INTERVAL`的延迟。

另一个重要点是，当要求服务器进程显示这些统计信息中的任何一种时，它会先取回收集器进程最近发出的报告，然后在当前事务结束之前，对所有统计视图和函数一直使用这一快照。因此，只要你继续当前事务，统计信息就会显示静态信息。这是一个特性，而不是一个错误，因为它允许你对统计信息执行多个查询并对结果做关联，而不必担心数字在你眼前变化。但如果希望每个查询都看到新的结果，请确保在任何事务块之外执行这些查询。

表 24.1. 标准统计视图

View Name	Description
pg_stat_activity	每个服务器进程一行，显示数据库 OID、数据库名、进程ID、用户 OID、用户名、当前查询、当前查询开始执行的时间、进程启动的时间，以及客户端的地址和端口号。报告当前查询数据的各列只有在参数stats_command_string被打开时才可用。此外，除非查看该视图的用户是超级用户或与被报告进程的属主是同一用户，这些列读取为空值。（注意，由于收集器的报告延迟，当前查询只对长时间运行的查询是最新的。）
pg_stat_database	每个数据库一行，显示数据库 OID、数据库名、连接到该数据库的活动服务器进程数、该数据库中提交和回滚的事务数、读取的磁盘块总数，以及缓冲区命中总数（即因在缓冲缓存中找到块而避免的块读请求数）。
pg_stat_all_tables	当前数据库中的每个表（包括 TOAST 表）一行，显示表 OID、模式名和表名、发起的顺序扫描次数、顺序扫描取出的活行数、发起的索引扫描次数（针对属于该表的所有索引）、索引扫描取出的活行数，以及行插入、更新和删除的次数。
pg_stat_sys_tables	与pg_stat_all_tables相同，但只显示系统表。
pg_stat_user_tables	与pg_stat_all_tables相同，但只显示用户表。
pg_stat_all_indexes	当前数据库中的每个索引一行，显示表和索引 OID、模式名、表名和索引名、在该索引上发起的索引扫描次数、索引扫描返回的索引条目数，以及使用该索引的简单索引扫描取出的活表行数。（当存在指向过期堆行的索引条目时，后者可能会更少。）
pg_stat_sys_indexes	和pg_stat_all_indexes一样，但只显示系统表上的索引。
pg_stat_user_indexes	和pg_stat_all_indexes一样，但只显示用户表上的索引。
pg_statio_all_tables	当前数据库中的每个表（包括 TOAST 表）一行，显示表 OID、模式名和表名、从该表读取的磁盘块数、缓冲区命中数、该表所有索引的磁盘块读取数和缓冲区命中数、该表辅助 TOAST 表（若有）的磁盘块读取数和缓冲区命中数，以及 TOAST 表索引的磁盘块读取数和缓冲区命中数。
pg_statio_sys_tables	与pg_statio_all_tables相同，但只显示系统表。
pg_statio_user_tables	与pg_statio_all_tables相同，但只显示用户表。
pg_statio_all_indexes	当前数据库中的每个索引一行，显示表和索引 OID、模式名、表名和索引名，以及该索引中的磁盘块读取数和缓冲区命中数。
pg_statio_sys_indexes	和pg_statio_all_indexes一样，但只显示系统表上的索引。
pg_statio_user_indexes	和pg_statio_all_indexes一样，但只显示用户表上的索引。
pg_statio_all_sequences	当前数据库中的每个序列对象一行，显示序列 OID、模式名和序列名，以及该序列中的磁盘块读取数和缓冲区命中数。

View Name	Description
pg_statio_sys_sequences	和pg_statio_all_sequences一样，但只显示系统序列。（目前没有定义任何系统序列，因此该视图总是空的。）
pg_statio_user_sequences	和pg_statio_all_sequences一样，但只显示用户序列。

每个索引的统计信息对于判断哪些索引正在被使用以及它们有多有效尤其有用。

从PostgreSQL 8.1 开始，索引既可以直接使用，也可以通过“位图扫描”使用。在位图扫描中，多个索引的输出可以通过 AND 或 OR 规则组合；因此在使用位图扫描时，很难把单个堆行取出与特定索引关联起来。所以，位图扫描会递增它所用索引的pg_stat_all_indexes.idx_tup_read计数，并递增该表的pg_stat_all_tables.idx_tup_fetch计数，但不影响pg_stat_all_indexes.idx_tup_fetch。

> **注意**
>
> 在PostgreSQL 8.1 之前，idx_tup_read和idx_tup_fetch计数基本总是相等。现在即使不考虑位图扫描，它们也可能不同，因为idx_tup_read
> 统计从索引中检索的索引条目，而idx_tup_fetch统计从表中取出的活行；
> 如果使用索引取出过任何死行或尚未提交的行，后者就会更少。

pg_statio_视图主要用于判断缓冲区缓存的有效性。
如果实际磁盘读取次数远少于缓冲区命中次数，缓存就能满足大多数读取请求，
而无需调用内核。不过，这些统计信息并不能反映全部情况：由于PostgreSQL处理磁盘 I/O 的方式，不在PostgreSQL缓冲区缓存中的数据仍可能驻留在内核的 I/O 缓存中，因而仍可在不执行物理读取的情况下取得。对于希望更详细地了解PostgreSQL I/O 行为的用户，建议将PostgreSQL统计收集器与能够深入查看内核 I/O 处理情况的操作系统工具结合使用。

也可以通过编写查询来建立查看统计信息的其他方式，
这些查询使用与标准视图相同的底层统计访问函数。这些函数列于表 24.2中。每数据库访问函数以数据库 OID 为参数来标识要报告的数据库。每表和每索引函数以表或索引 OID 为参数。函数调用统计的函数以函数 OID 为参数。（注意，用这些函数只能看到当前数据库中的表、索引和函数。）每服务器进程访问函数以一个服务器进程编号为参数，其范围从 1 到当前活动的服务器进程数。

表 24.2. 统计访问函数

Function	Return Type	Description
pg_stat_get_db_numbackends(oid)	integer	数据库的活动服务器进程数
pg_stat_get_db_xact_commit(oid)	bigint	数据库中提交的事务
pg_stat_get_db_xact_rollback(oid)	bigint	数据库中回滚的事务
pg_stat_get_db_blocks_fetched(oid)	bigint	数据库的磁盘块读取请求数
pg_stat_get_db_blocks_hit(oid)	bigint	数据库在缓存中找到的磁盘块读取请求数
pg_stat_get_numscans(oid)	bigint	参数为表时是完成的顺序扫描数，参数为索引时是完成的索引扫描数

Function	Return Type	Description
pg_stat_get_tuples_returned(oid)	bigint	参数为表时是顺序扫描取出的活行数，参数为索引时是返回的索引条目数
pg_stat_get_tuples_fetched(oid)	bigint	参数为表时是位图扫描取出的表行数，参数为索引时是使用该索引的简单索引扫描取出的表行数
pg_stat_get_tuples_inserted(oid)	bigint	插入表中的行数
pg_stat_get_tuples_updated(oid)	bigint	表中更新的行数
pg_stat_get_tuples_deleted(oid)	bigint	从表中删除的行数
pg_stat_get_blocks_fetched(oid)	bigint	表或索引的磁盘块读取请求数
pg_stat_get_blocks_hit(oid)	bigint	表或索引在缓存中找到的磁盘块请求数
pg_stat_get_backend_idset()	setof integer	当前活动的服务器进程编号的集合（从 1 到活动的服务器进程数）。用法示例见正文。
pg_backend_pid()	integer	附属于当前会话的服务器进程的进程 ID
pg_stat_get_backend_pid(integer)	integer	给定服务器进程的进程 ID
pg_stat_get_backend_dbid(integer)	oid	给定服务器进程的数据库 ID
pg_stat_get_backend_userid(integer)	oid	给定服务器进程的用户 ID
pg_stat_get_backend_activity(integer)	text	给定服务器进程的活动命令（如果当前用户既不是超级用户也不是被查询会话的用户，或者stats_command_string未打开，则为空值）
pg_stat_get_backend_activity_start(integer)	timestamp with time zone	给定服务器进程当前执行的查询的启动时间（如果当前用户既不是超级用户也不是被查询会话的用户，或者stats_command_string未打开，则为空值）
pg_stat_get_backend_start(integer)	timestamp with time zone	给定服务器进程的启动时间（如果当前用户既不是超级用户也不是被查询会话的用户，则为空值）
pg_stat_get_backend_client_addr(integer)	inet	连接到给定服务器进程的客户端的 IP 地址。如果连接是通过 Unix 域套接字进行的，则为空值。如果当前用户既不是超级用户也不是被查询会话的用户，同样为空值
pg_stat_get_backend_client_port(integer)	integer	连接到给定服务器进程的客户端的 TCP 端口号。如果连接是通过 Unix 域套接字进行的，则为 -1。如果当前用户既不是超级用户也不是被查询会话的用户，则为空值

Function	Return Type	Description
pg_stat_reset()	boolean	重置当前收集到的所有统计信息

注意

blocks_fetched减去blocks_hit得到的是为该表、索引或数据库发出的内核read()调用数；但实际物理读取数通常由于内核级缓冲而更低。

函数pg_stat_get_backend_idset提供了一种为每个活动服务器进程生成一行的便捷方法。例如，要显示所有服务器进程的PID和当前查询：

```
SELECT pg_stat_get_backend_pid(s.backendid) AS procpid,
       pg_stat_get_backend_activity(s.backendid) AS current_query
    FROM (SELECT pg_stat_get_backend_idset() AS backendid) AS s;
```

24.3. 查看锁

监控数据库活动的另外一个有用的工具是pg_locks系统表。
这样就允许数据库管理员查看锁管理器中当前存在的锁的信息。例如，这个功能可以被用于：

- 查看当前存在的所有锁、在一个特定数据库中的关系上所有的锁、在一个特定关系上所有的锁，或者由一个特定PostgreSQL会话持有的所有的锁。
- 判断当前数据库中带有最多未授予锁的关系（它可能是数据库客户端之间的竞争来源）。
- 判断锁竞争给数据库总体性能带来的影响，以及锁竞争随着整个数据库流量的变化范围。

pg_locks视图的细节在第 42.37 节中。更多有关PostgreSQL的锁和管理并发性的信息，请参考第 12 章。

第 25 章 监控磁盘使用情况

本章讨论如何监控PostgreSQL数据库系统的磁盘使用情况。

25.1. 确定磁盘使用情况

每个表都有一个存放大部分数据的主堆磁盘文件。如果表有任何可能很宽的列，还会有一个与该表关联的TOAST文件，用于存储宽到无法舒适地放在主表中的值（见第 50.2 节）。如果存在TOAST表，其上将有一个索引。基表还可能有关联的索引。每个表和索引都存储在单独的磁盘文件中——如果文件超过 1 GB，则可能不止一个文件。这些文件的命名约定见第 50.1 节。

可以用三种方式监控磁盘空间：使用表 9.47中列出的 SQL 函数、使用VACUUM提供的信息，以及在命令行上使用contrib/oid2name中的工具。SQL 函数最容易使用，可以报告表、带索引和长值存储（TOAST）的表、数据库以及表空间的信息。

在最近进行过清理或分析的数据库上使用psql时，可以发出查询以查看任意表的磁盘使用情况：

```
SELECT relfilenode, relpages FROM pg_class WHERE relname = 'customer';

 relfilenode | relpages
-------------+----------
       16806 |       60
(1 row)
```

每个页通常为 8 KB。（请记住，relpages只会由VACUUM、ANALYZE以及少数 DDL 命令（如CREATE INDEX）更新。）如果你想直接检查该表的磁盘文件，relfilenode值就很有用。

要显示TOAST表使用的空间，可以使用如下查询：

```
SELECT relname, relpages
FROM pg_class,
     (SELECT reltoastrelid
      FROM pg_class
      WHERE relname = 'customer') AS ss
WHERE oid = ss.reltoastrelid OR
      oid = (SELECT reltoastidxid
             FROM pg_class
             WHERE oid = ss.reltoastrelid)
ORDER BY relname;

       relname        | relpages
----------------------+----------
 pg_toast_16806       |        0
 pg_toast_16806_index |        1
```

也可以轻松显示索引大小：

```
SELECT c2.relname, c2.relpages
FROM pg_class c, pg_class c2, pg_index i
WHERE c.relname = 'customer' AND
      c.oid = i.indrelid AND
```

```
      c2.oid = i.indexrelid
ORDER BY c2.relname;

       relname        | relpages
----------------------+----------
 customer_id_indexdex |       26
```

利用这些信息，你可以很容易地找出最大的表和索引：

```
SELECT relname, relpages
FROM pg_class
ORDER BY relpages DESC;

       relname        | relpages
----------------------+----------
 bigtable             |     3290
 customer             |     3144
```

你还可以使用 `contrib/oid2name` 来显示磁盘使用情况。示例见该目录中的`README.oid2name`。其中包含一个显示每个数据库磁盘使用情况的脚本。

25.2. 磁盘已满故障

数据库管理员最重要的磁盘监控任务，是确保磁盘不会被写满。
数据磁盘写满不会导致数据损坏，但可能会阻止有用的活动继续进行。如果保存 WAL 文件的磁盘被写满，则数据库服务器可能会 panic 并随后关闭。

如果无法通过删除其他内容来释放磁盘上的额外空间，
可以利用表空间将部分数据库文件移到其他文件系统上。有关更多信息，请参见第 19.6 节。

> **提示**
>
> 有些文件系统在接近写满时性能会明显变差，因此不要等到磁盘完全写满才采取措施。

如果你的系统支持按用户设置磁盘配额，
那么数据库自然也会受到运行服务器的那个用户所拥有配额的限制。
超出配额会带来与磁盘空间完全耗尽相同的不良影响。

第 26 章 可靠性与预写式日志

本章说明如何使用预写式日志，实现高效、可靠的运行。

26.1. 可靠性

可靠性是任何严肃数据库系统的重要属性，而 PostgreSQL 会尽一切可能保证运行可靠。可靠运行的一个方面是，已提交事务所记录的所有数据都应存储在不受断电、
操作系统故障和硬件故障影响的非易失性区域中（当然不包括该非易失性区域本身的故障）。
通常，只要能成功将数据写入计算机的永久存储设备（磁盘驱动器或等效设备），
就满足这一要求。实际上，即使一台计算机遭受致命损坏，只要磁盘驱动器得以幸存，
就可以将它们移到另一台硬件相近的计算机上，所有已提交事务仍将完好无损。

虽然周期性地强制将数据写到磁盘盘片似乎是个简单操作，但事实并非如此。
由于磁盘驱动器远慢于主存和 CPU，在计算机主存与磁盘盘片之间存在多层缓存。首先是操作系统的缓冲区缓存，它会缓存常用磁盘块并合并磁盘写入。幸运的是，
所有操作系统都为应用提供了强制将数据从缓冲区缓存写入磁盘的方法，而 PostgreSQL 会使用这些机制。（关于如何调整这一过程，见参数wal_sync_method。）

其次，磁盘驱动器控制器中可能还有缓存；这在 RAID 控制卡上尤为常见。其中一些缓存是直写式，也就是说写请求一到达就立刻传递给驱动器。另一些是回写式，
也就是说数据会在稍后的某个时间发送到驱动器。这类缓存可能构成可靠性隐患，
因为磁盘控制器缓存中的内存是易失性的，断电时会丢失内容。较好的控制卡配有
电池后备缓存，也就是卡上带有电池，可在系统断电时维持缓存供电。电力恢复后，
数据便会写入磁盘驱动器。

最后，大多数磁盘驱动器自身也带有缓存。有些是直写式，有些是回写式；
对于回写式驱动器缓存，同样存在与磁盘控制器缓存相同的数据丢失隐患。消费级 IDE 驱动器尤其可能含有在断电后无法保留内容的回写式缓存。

当操作系统向存储硬件发出写请求后，
它几乎做不了什么来确保数据已经到达真正的非易失性存储区域。相反，
管理员有责任确保所有存储组件都能保证数据完整性。
应避免使用写缓存没有电池后备的磁盘控制器。在驱动器层面，
如果驱动器无法保证在关闭前把数据写出，就应禁用回写缓存。

数据丢失的另一种风险来自磁盘盘片写操作本身。磁盘盘片被分为若干扇区，通常每个 512 字节。每一次物理读写都处理整个扇区。当写请求到达驱动器时，它可能是 512 字节、1024 字节或 8192 字节，而写入过程可能在任何时刻因断电失败，这就意味着某些 512 字节扇区写成功了，另一些却没有。为防范这类故障，PostgreSQL 会在修改磁盘上的实际页面之前，定期将整页镜像写入持久 WAL 存储。这样一来，在崩溃恢复时 PostgreSQL 就能从 WAL 恢复部分写入的页面。如果你使用了能阻止部分页面写入的后备电池磁盘控制器或文件系统软件（例如 Reiser4），可以通过full_page_writes参数关闭这种页面镜像。

WAL 的目标是确保数据库记录在被修改之前，日志已经先被写出；但如果磁盘驱动器实际上只是缓存了数据、尚未将其存储到磁盘上，却向内核谎报写入成功，这一目标就会被破坏。
在这种情况下，断电可能导致不可恢复的数据损坏。管理员应尽量确保保存 PostgreSQL WAL 文件的磁盘不会做出这种虚假报告。

26.2. 预写式日志（WAL）

预写式日志（WAL）是确保数据完整性的标准方法。大多数（即使不是全部）
事务处理书籍中都能找到详细描述。简而言之，WAL 的核心思想是，对数据文件（即表和索引所

在之处）的更改，必须在这些更改被记入日志之后才能写出；也就是说，必须在描述这些更改的 WAL 记录被刷入持久存储之后，再去写数据文件。如果遵循这个过程，就无需在每次事务提交时都把数据页刷盘，因为我们知道，一旦发生崩溃，
就可以利用日志恢复数据库：任何尚未应用到数据页上的更改，都可以根据 WAL 记录重做。（这就是前滚恢复，也称 REDO。）

使用 WAL 会显著减少磁盘写入次数，因为要保证事务已提交，通常只需把 WAL 文件刷入磁盘，而不必把该事务修改过的每个数据文件都刷盘。WAL 文件按顺序写入，因此同步 WAL 的代价远低于刷写数据页。对于处理大量小事务、且这些事务触及数据存储不同部分的服务器，这一点尤其明显。此外，当服务器正在处理大量并发的小事务时，对 WAL 文件执行一次 `fsync` 就可能足以提交多个事务。

WAL 还使在线备份和时间点恢复成为可能，如第 23.3 节所述。通过归档 WAL 数据，我们就能支持回退到可用 WAL 数据覆盖范围内的任意时刻：只需先安装数据库的一个较早物理备份，然后把 WAL 重放到目标时刻即可。更进一步说，这个物理备份不必是数据库状态的瞬时快照 – 即使备份是在一段时间内完成的，重放这段时间内的 WAL 也能修复任何内部不一致。

26.3. WAL 配置

有若干与 WAL 相关的配置参数会影响数据库性能。本节解释它们的用法。有关设置服务器配置参数的一般信息，请参见第 17 章。

检查点是事务序列中的一些时间点，
在这些点可以保证堆和索引数据文件已经更新了检查点之前写入的所有信息。在检查点时刻，
所有脏数据页都被刷写到磁盘，并向日志文件写入一条特殊的检查点记录。一旦发生崩溃，
崩溃恢复过程会查看最新的检查点记录，以确定日志中开始 REDO 操作的位置（称为重做记录）。在此之前对数据文件所做的更改都保证已在磁盘上。因此，检查点之后，
包含重做记录的日志段之前的日志段不再需要，可以循环使用或删除。（在进行WAL归档时，
日志段在循环使用或删除之前必须先归档。）

服务器的后台写进程会自动定期执行检查点。每隔checkpoint_segments个日志段，或每隔checkpoint_timeout秒（以先到者为准）就会创建一个检查点。默认设置分别是 3 个段和 300 秒（5 分钟）。也可以使用 SQL 命令`CHECKPOINT`强制创建检查点。

减小`checkpoint_segments`和/或`checkpoint_timeout`会使检查点更频繁地发生。
这可以让崩溃后的恢复更快（因为需要重做的工作更少）。不过，
必须与更频繁地刷写脏数据页所增加的代价相权衡。如果启用了full_page_writes（默认如此），还有另一个因素需要考虑。为保证数据页一致性，
每个检查点之后对数据页的第一次修改会把整页内容记入日志。在这种情况下，
更小的检查点间隔会增加写入 WAL 日志的数据量，部分抵消了使用更小间隔的目的，而且无论如何都会导致更多磁盘 I/O。

检查点的代价相当高：首先它要求写出所有当前脏缓冲区，其次如上所述它还会导致后续额外的 WAL 流量。因此，把检查点参数设置得足够高、使检查点不至于过于频繁是明智的。作为对检查点参数的简单检查，可以设置checkpoint_warning参数。
如果检查点发生的间隔小于`checkpoint_warning`秒，服务器日志中就会输出一条消息，
建议增大`checkpoint_segments`。偶尔出现这样的消息不必惊慌，但如果它频繁出现，
就应当增大检查点控制参数。如果没有把`checkpoint_segments`设置得足够高，大型`COPY`传输之类的批量操作可能导致出现多条这样的警告。

至少会有一个 WAL 段文件，通常不会多于 2 * `checkpoint_segments` + 1 个文件。每个段文件通常是 16 MB（不过构建服务器时可以改变这个大小）。可以用它来估算WAL的空间需求。通常，当日志段文件不再需要时，它们会被循环使用（改名为编号序列中的后续段）。
如果由于日志输出速率的短期峰值，段文件数超过了 2 * `checkpoint_segments` + 1，则不需要的段文件将被删除而不是循环使用，直到系统回到该限制之内。

有两个常用的内部WAL函数：LogInsert和LogFlush。LogInsert用于把新记录放入共享内存中的WAL缓冲区。如果没有存放新记录的空间，LogInsert就不得不把（移动到内核缓存中）一些已写满的WAL缓冲区写出。这是不希望发生的，因为LogInsert在每次数据库底层修改（例如行插入）时都会被调用，而且此时受影响的数据页上持有排他锁，所以该操作需要尽可能快。更糟的是，写出WAL缓冲区还可能强制创建新的日志段，这需要更多时间。通常，WAL缓冲区应当由LogFlush请求来写出并刷盘，该请求大部分在事务提交时发出，以确保事务记录被刷到永久存储。在日志输出量很大的系统上，LogFlush请求可能不够频繁，无法避免LogInsert不得不进行写操作。在这样的系统上，应当通过修改配置参数wal_buffers来增加WAL缓冲区的数量。WAL缓冲区的默认数量是 8。增大这个值会相应增加共享内存的使用。当启用了full_page_writes且系统非常繁忙时，调高这个值有助于平滑每个检查点之后紧随时段的响应时间。

commit_delay参数定义服务器进程在用LogInsert把提交记录写入日志之后、执行LogFlush之前，将休眠多少微秒。这一延迟让其他服务器进程可以把它们的提交记录也加入日志，从而所有这些记录可以通过一次日志同步一起刷盘。如果未启用fsync，或者当前处于活动事务中的其他会话少于commit_siblings个，则不会发生休眠；这避免了在其他会话不太可能很快提交时进行休眠。注意，在大多数平台上，休眠请求的分辨率是十毫秒，因此任何介于 1 到 10000 微秒之间的非零commit_delay设置效果都相同。这些参数的理想值尚不明确；鼓励进行实验。

wal_sync_method参数决定PostgreSQL以何种方式请求内核把WAL更新强制写到磁盘。就可靠性而言，所有选项都应当相同；不过，哪个选项最快是相当平台相关的。注意，如果关闭了fsync，此参数就无关紧要了。

启用wal_debug配置参数（前提是PostgreSQL编译时包含了相应支持）会把每次LogInsert和LogFlushWAL调用记录到服务器日志中。此选项将来可能被更通用的机制取代。

26.4. WAL 内部机制

WAL 会自动启用；管理员除了确保满足 WAL 文件的磁盘空间需求，并完成任何必要调优之外，不需要再做其他操作（见第 26.3 节）。

WAL 文件保存在数据目录下的 pg_xlog 目录中，由一组段文件组成，通常每段 16 MB。每段又被划分为若干页面，通常每页 8 KB。WAL 记录头定义在 access/xlog.h 中；记录内容则取决于被记录的事件类型。段文件以不断递增的数字命名，从 000000010000000000000000 开始。这些编号不会回绕，但要把可用的编号耗尽还需要非常、非常长的时间。

如果能将 WAL 放在与主数据库文件不同的磁盘上，会更有利。这可以通过把 pg_xlog 目录移动到其他位置（当然要在服务器关闭时进行），然后在主数据目录中的原位置创建一个指向新位置的符号链接来实现。

在完成检查点并刷写 WAL 之后，该检查点的位置会保存在文件 pg_control 中。因此，在恢复开始时，服务器会先读取 pg_control，再读取检查点记录；然后从检查点记录给出的 WAL 位置向前扫描并执行 REDO。由于检查点之后对数据页的第一次修改，会把整页内容保存在 WAL 中，因而自该检查点以来被修改的所有页面都会恢复到一致状态。

为处理 pg_control 损坏的情况，我们本应支持按相反顺序扫描现有 WAL 段 – 也就是从新到旧 –，以便找到最新检查点。这一功能尚未实现。pg_control 足够小（不足一个磁盘页），因此不会受部分写问题影响；到目前为止，也没有仅仅因为无法读取 pg_control 本身而导致数据库故障的报告。所以，尽管理论上它是个薄弱环节，pg_control 在实践中似乎不是问题。

第 27 章 回归测试

回归测试是针对PostgreSQL中 SQL 实现的一套全面测试。它们既测试标准 SQL 操作，也测试PostgreSQL的扩展能力。

27.1. 运行测试

回归测试既可以针对已经安装并正在运行的服务器执行，
也可以在构建树中使用临时安装来执行。此外，运行测试时还有“parallel”和“sequential”两种模式。顺序方式一次只运行一个测试脚本，而并行方式会启动多个服务器进程，
以并行方式运行一组测试。并行测试能够确认进程间通信和锁定是否正常工作。由于历史原因，
顺序测试通常针对已有安装运行，而并行方式通常针对临时安装运行，
但这并没有技术上的必然性。

要在构建完成后、安装之前运行回归测试，请在顶层目录中输入：

```
gmake check
```

（也可以切换到`src/test/regress`并在那里运行此命令。）
这会首先构建若干辅助文件（例如示例用户定义触发器函数），然后运行测试驱动脚本。
结束时应该会看到类似下面的结果：

```
=======================
 All 98 tests passed.
=======================
```

否则会显示哪些测试失败的说明。在假定“失败”代表严重问题之前，请先阅读下面的第 27.2 节。

由于这种测试方法会运行一个临时服务器，当你以 root 用户身份运行时它将无法工作（因为服务器不会以 root 身份启动）。如果你已经以 root 身份完成了构建，也不必从头再来。只需让回归测试目录可被其他用户写入，以该用户身份登录，然后重新启动测试即可。例如

```
root# chmod -R a+w src/test/regressroot# chmod -R a+w contrib/
spiroot# su - joeuserjoeuser$ cd top-level build directoryjoeuser
$ gmake check
```

（这里唯一可能的“安全风险”是其他用户可能在你不知情的情况下篡改回归测试的结果。
在管理用户权限时请运用常识。）

或者，在安装完成后再运行测试。

如果你把PostgreSQL配置为安装到某个已经存在旧版PostgreSQL安装的位置，
并且在安装新版本之前执行`gmake check`，可能会发现测试失败，因为新程序会尝试使用已安装的共享库。（典型症状是报出未定义符号。）如果你希望在覆盖旧安装之前运行测试，
就需要使用`configure --disable-rpath`构建。不过，不建议在最终安装中使用此选项。

并行回归测试会在你的用户 ID 下启动相当多的进程。目前最大并发度为二十个并行测试脚本，也就是六十个进程：每个测试脚本都有一个服务器进程、一个psql进程，而且psql通常还有一个shell 父进程。因此，如果你的系统对每个用户可创建的进程数施加限制，请确保该限制至少有七十五个左右，否则并行测试中可能会出现看似随机的失败。如果无法提高该限制，可以通过设置`MAX_CONNECTIONS`参数来降低并行度。例如：

```
gmake MAX_CONNECTIONS=10 check
```

这样最多同时运行十个测试。

在某些系统上，默认的与 Bourne 兼容的 shell（/bin/sh）在需要并行管理太多子进程时会出问题，这可能导致并行测试挂起或失败。在这种情况下，可以在命令行上指定另一个与 Bourne 兼容的 shell，例如：

```
gmake SHELL=/bin/ksh check
```

如果找不到没有问题的 shell，也可以像上面那样通过限制连接数来绕过该问题。

要在安装后运行测试 (见第 14 章)，请初始化一个数据区域并启动服务器如第 16 章所述，然后输入：

```
gmake installcheck
```

或者进行并行测试：

```
gmake installcheck-parallel
```

除非由PGHOST和PGPORT环境变量另行指定，测试将期望在本地主机和默认端口上连接服务器。

源码发布包中还包含针对可选过程语言以及部分contrib模块的回归测试。
目前这些测试只能针对已安装的服务器使用。要为所有已构建并安装的过程语言运行测试，
请切换到构建树的src/pl目录并输入：

```
gmake installcheck
```

也可以在src/pl的任一子目录中这样做，只为一种过程语言运行测试。要为所有带有测试的
contrib模块运行测试，请切换到构建树的contrib目录并输入：

```
gmake installcheck
```

contrib模块必须先构建并安装。也可以在contrib的某个子目录中这样做，
只为一个模块运行测试。

27.2. 测试结果评估

某些安装正确且功能完备的PostgreSQL实例，
可能会因为平台特有因素而在部分回归测试中“失败”，例如浮点表示或消息措辞不同。
目前这些测试是通过把输出与参考系统生成的输出做简单的diff比较来评估的，
因此结果会对细微的系统差异很敏感。报告某项测试“失败”时，
请务必检查预期结果与实际结果之间的差异；你可能会发现这些差异并不重要。尽管如此，
我们仍努力在所有受支持平台上维护准确的参考文件，因此原则上应期待所有测试都能通过。

回归测试的实际输出位于src/test/regress/results目录中的文件里。测试脚本使用diff
把每个输出文件与存储在src/test/regress/expected目录中的参考输出进行比较。
任何差异都会保存到src/test/regress/regression.diffs中供你检查。（如果愿意，
也可以自己运行diff。）

如果由于某种原因，某个特定平台在某个测试上产生了“失败”，
但对输出的检查令你相信结果是有效的，那么可以新增一个比较文件，
让后续测试运行不再报告该失败。详情见第 27.3 节。

27.2.1. 错误消息差异

某些回归测试包含故意构造的非法输入值。错误消息可能来自 PostgreSQL代码，也可能来自宿主平台的系统例程。在后一种情况下，消息会因平台而异，但应反映相近的信息。
这类消息差异会导致回归测试显示为“失败”，但可以通过检查确认其有效性。

27.2.2. 区域差异

如果你针对一个使用 C 以外排序规则顺序的区域设置初始化的已安装服务器运行测试，就可能因为排序顺序不同而出现差异，并导致后续失败。
回归测试套件通过提供备用结果文件来处理这个问题；这些文件已知可以覆盖大量区域设置。

27.2.3. 日期和时间差异

大多数日期和时间结果依赖于时区环境。参考文件是在时区 `PST8PDT`（加利福尼亚州伯克利）下生成的，如果测试未在该时区设置下运行，就会出现表面上的失败。
回归测试驱动器会将环境变量 `PGTZ`设为`PST8PDT`，这通常能够确保获得正确结果。

27.2.4. 浮点差异

某些测试涉及从表列计算 64 位浮点数（`double precision`）。已经观察到涉及`double precision`列数学函数的结果存在差异。`float8`和`geometry`测试尤其容易在不同平台之间、甚至在不同编译器优化设置下出现微小差异。
需要人工目测比较来确定这些差异的真实意义，它们通常出现在小数点右侧第 10 位。

某些系统把负零显示为`-0`，而另一些只显示 `0`。

某些系统对`pow()`和`exp()` 发出错误的方式，与当前PostgreSQL代码所预期的机制不同。

27.2.5. 行顺序差异

你可能会看到这样的差异：同样的行在输出中的顺序与预期文件中的顺序不同。
在大多数情况下，严格说来这并不是缺陷。大多数回归测试脚本并没有细致到为每一个`SELECT`都使用`ORDER BY`，因此按 SQL 规范，它们的结果行顺序并没有良好定义。实际上，由于我们看到的是同一软件在同一数据上执行相同查询，通常在所有平台上都会得到相同的结果顺序，
所以缺少`ORDER BY`并不是问题。不过，有些查询确实会表现出跨平台的顺序差异。针对已安装服务器测试时，非 C 区域设置或非默认参数设置，例如自定义的`work_mem`值或规划器代价参数，也可能导致顺序差异。

因此，如果你看到顺序差异，一般无需担心，除非查询确实包含 `ORDER BY`而你的结果违反了它。不过，仍请报告该问题，这样我们可以为那个特定查询加上`ORDER BY`，以在后续版本中消除这种虚假的“失败”。

你可能会好奇，为什么我们不显式地为所有回归测试查询排序，从而一劳永逸地解决这个问题。
原因在于，那样反而会降低回归测试的价值，
因为测试会倾向于覆盖能产生有序结果的查询计划类型，
而排除那些不能产生有序结果的计划类型。

27.2.6. 栈深度不足

如果`errors`测试在`select infinite_recurse()`命令处导致服务器崩溃，说明该平台的进程栈大小限制小于max_stack_depth参数所指示的值。这可以通过在更高的栈大小限制下运行

postmaster 来解决（在使用默认的max_stack_depth值时建议 4MB）。如果无法做到，另一种办法是减小max_stack_depth的值。

27.2.7. “random” 测试

random测试脚本旨在产生随机结果。在少数情况下，这会导致随机回归测试失败。输入：

```
diff results/random.out expected/random.out
```

应该只产生一行或几行差异。除非随机测试反复失败，否则不必担心。

27.3. 变体比较文件

由于某些测试天生会产生依赖环境的结果，我们提供了指定备用“预期”结果文件的方法。
每个回归测试都可以有多个比较文件，用来展示不同平台上的可能结果。
对于每个测试应使用哪个比较文件，有两种相互独立的判定机制。

第一种机制允许针对特定平台选择比较文件。有一个映射文件 src/test/regress/resultmap，用于定义每个平台应使用哪个比较文件。
要为某个特定平台消除虚假的测试“失败”，你需要先选择或创建一个变体结果文件，然后在 resultmap 文件中增加一行。

映射文件中的每一行都采用如下形式：

```
testname/platformpattern=comparisonfilename
```

测试名就是相应回归测试模块的名称。平台模式是 Unix 工具expr风格的模式（也就是在开头隐含^锚点的正则表达式）。它会与config.guess打印出的平台名进行比较，后跟 :gcc或:cc，这取决于你使用的是 GNU 编译器还是系统的原生编译器（在两者有差别的系统上）。比较文件名则是替代结果比较文件的基名。

例如：某些系统会把极小的浮点数值解释为零，而不是报告下溢错误。这会导致float8 回归测试出现一些差异。因此，我们提供了一个变体比较文件float8-small-is-zero.out，其中包含这些系统上的预期结果。为了消除虚假的“失败”消息，在OpenBSD平台上，resultmap包含以下内容：

```
float8/i.86-.*-openbsd=float8-small-is-zero
```

如果一台机器上的config.guess输出匹配i.86-.*-openbsd，就会触发该规则。
resultmap中的其他行会为其他适用的平台选择变体比较文件。

第二种用于选择变体比较文件的机制更加自动化：它只是在多个提供的比较文件中选取“最佳匹配”。回归测试驱动脚本既考虑某个测试的标准比较文件*testname*.out，也考虑名为*testname*_*digit*.out 的变体文件（其中*digit*可以是任一单个数字 0-9）。如果其中任何一个文件完全匹配，就认为测试通过；否则，会使用生成最短 diff 的那个文件来创建失败报告。（如果resultmap为该特定测试包含一项，那么基础*testname*就是 resultmap中给出的替代名称。）

例如，对于char测试，比较文件 char.out包含在C和 POSIX区域设置中应当出现的结果，而文件 char_1.out则包含在许多其他区域设置中出现的排序结果。

最佳匹配机制原本是为了解决依赖区域设置的结果问题，
但它也可以用于任何仅凭平台名无法轻易预测测试结果的场景。该机制的一个局限是，
测试驱动器无法判断哪一个变体对当前环境才真正“正确”；它只会选择看起来最合适的那个。

因此，
最安全的做法是只对那些你愿意在所有上下文中都视为同样有效的变体结果使用这种机制。

部分 IV. 客户端接口

本部分描述随PostgreSQL一起发布的客户端编程接口。这些章节彼此独立，可以分别阅读。
还有许多面向客户端程序的外部编程接口是单独发布的，它们各自也附带文档（附录 H列出了一些较常用的接口）。本部分的读者应熟悉如何使用SQL操作和查询数据库（见第 II 部分），当然还应熟悉自己所选的编程语言。

目录

28. libpq - C 库 350
28.1. 数据库连接控制函数 350
28.2. 连接状态函数 355
28.3. 命令执行函数 359
28.3.1. 主要函数 359
28.3.2. 检索查询结果信息 364
28.3.3. 检索其他结果信息 367
28.3.4. 用于在 SQL 命令中嵌入字符串的转义 368
28.3.5. 用于在 SQL 命令中嵌入二进制串的转义 369
28.4. 异步命令处理 370
28.5. 取消进行中的查询 373
28.6. 快速路径接口 374
28.7. 异步通知 375
28.8. 与COPY命令相关的函数 376
28.8.1. 发送COPY数据的函数 377
28.8.2. 接收COPY数据的函数 378
28.8.3. COPY的已废弃函数 378
28.9. 控制函数 380
28.10. 通知处理 381
28.11. 环境变量 382
28.12. 密码文件 383
28.13. 连接服务文件 383
28.14. SSL 支持 383
28.15. 在线程化程序中的行为 384
28.16. 构建libpq程序 384
28.17. 示例程序 385
29. 大对象 396
29.1. 历史 396
29.2. 实现特性 396
29.3. 客户端接口 396
29.3.1. 创建一个大对象 396
29.3.2. 导入一个大对象 397
29.3.3. 导出一个大对象 397
29.3.4. 打开一个现有的大对象 397
29.3.5. 向大对象写入数据 398
29.3.6. 从大对象读取数据 398
29.3.7. 在大对象中定位 398
29.3.8. 获取大对象的当前位置 398
29.3.9. 关闭一个大对象描述符 398
29.3.10. 移除一个大对象 399
29.4. 服务器端函数 399
29.5. 示例程序 399
30. ECPG - C 中的嵌入式 SQL 405
30.1. 概念 405
30.2. 连接到数据库服务器 405
30.3. 关闭一个连接 406
30.4. 运行 SQL 命令 407
30.5. 选择一个连接 407
30.6. 使用主变量 408
30.6.1. 概述 408
30.6.2. 声明节 408

30.6.3. SELECT INTO与FETCH INTO 409
30.6.4. 指示符 410
30.7. 动态 SQL 410
30.8. 使用描述符区域 411
30.9. 错误处理 413
30.9.1. 设置回调 413
30.9.2. sqlca 414
30.9.3. SQLSTATE 与 SQLCODE 415
30.10. 包含文件 418
30.11. 处理嵌入式 SQL 程序 418
30.12. 库函数 419
30.13. 内部 419
31. 信息模式 422
31.1. 模式 422
31.2. 数据类型 422
31.3. information_schema_catalog_name 422
31.4. applicable_roles 423
31.5. check_constraints 423
31.6. column_domain_usage 423
31.7. column_privileges 424
31.8. column_udt_usage 424
31.9. columns 425
31.10. constraint_column_usage 427
31.11. constraint_table_usage 427
31.12. data_type_privileges 428
31.13. domain_constraints 428
31.14. domain_udt_usage 429
31.15. domains 429
31.16. element_types 431
31.17. enabled_roles 433
31.18. key_column_usage 433
31.19. parameters 433
31.20. referential_constraints 435
31.21. role_column_grants 435
31.22. role_routine_grants 436
31.23. role_table_grants 436
31.24. role_usage_grants 437
31.25. routine_privileges 437
31.26. routines 438
31.27. schemata 440
31.28. sql_features 441
31.29. sql_implementation_info 441
31.30. sql_languages 442
31.31. sql_packages 442
31.32. sql_sizing 442
31.33. sql_sizing_profiles 443
31.34. table_constraints 443
31.35. table_privileges 444
31.36. tables 444
31.37. triggers 445
31.38. usage_privileges 446
31.39. view_column_usage 446

31.40. `view_table_usage` 447
31.41. `views` 447

第 28 章 libpq - C 库

libpq是PostgreSQL的C应用程序编程接口。libpq是一组库函数，客户端程序可用它们向PostgreSQL后端服务器发送查询并接收查询结果。

libpq也是其他几个PostgreSQL应用程序接口的底层引擎，包括为 C++、Perl、Python、Tcl 和ECPG编写的接口。因此，对于这些包的用户，libpq行为的某些方面也很重要。特别是，第 28.11 节、第 28.12 节和第 28.14 节描述了任何使用libpq的应用程序的用户都能观察到的行为。

本章末尾（第 28.17 节）包含一些简短程序，展示如何编写使用libpq的程序。源代码发行包的`src/test/examples`目录中还提供了几个完整的libpq应用程序示例。

使用libpq的客户端程序必须包含头文件`libpq-fe.h`，并且必须与libpq库链接。

28.1. 数据库连接控制函数

下列函数处理与PostgreSQL后端服务器建立连接。一个应用可以同时打开多个后端连接。（这样做的一个原因是访问多个数据库。）每个连接由一个`PGconn`对象表示，该对象从函数`PQconnectdb`或`PQsetdbLogin`获得。注意，除非可能连分配`PGconn`对象的内存都不够，这些函数将总是返回一个非空的对象指针。在通过连接对象发送查询之前，应该调用`PQstatus`函数检查连接是否已成功建立。

`PQconnectdb`

开启一个到数据库服务器的新连接。

```
PGconn *PQconnectdb(const char *conninfo);
```

这个函数使用从字符串`conninfo`中得到的参数开启一个新的数据库连接。与下面的`PQsetdbLogin`不同，参数集合可以在不改变函数签名的情况下扩展，因此对于新应用，最好使用这个函数（或者相应的非阻塞函数`PQconnectStart`和`PQconnectPoll`）。

传入的字符串可以为空以使用全部默认参数，也可以包含一个或多个用空白分隔的参数设置。每一个参数设置的形式都是`keyword = value`。等号周围的空白是可选的。要写一个空值或一个包含空白的值，将它用单引号包围，例如`keyword = 'a value'`。值中的单引号和反斜线必须用一个反斜线转义，即`\'`和`\\`。

当前能被识别的参数关键字如下：

`host`

要连接的主机名。如果主机名以斜杠开头，则指定的是 Unix 域通信，而非 TCP/IP 通信；此值是存放套接字文件的目录名。当未指定 `host` 或其值为空时，默认连接到 `/tmp`（或构建 PostgreSQL 时指定的套接字目录）中的 Unix 域套接字。在没有 Unix 域套接字的机器上，默认连接到 `localhost`。

`hostaddr`

要连接的主机的数字 IP 地址。应采用标准的 IPv4 地址格式，例如 `172.28.40.9`。如果机器支持 IPv6，也可以使用 IPv6 地址。只要此参数指定了非空字符串，就始终使用 TCP/IP 通信。

使用hostaddr而非host让应用可以避免一次主机名查找，这对于有时间约束的应用可能很重要。不过，Kerberos 和 GSSAPI 认证需要主机名。因此适用下列规则：如果指定了host而没有指定hostaddr，会发生一次主机名查找。如果指定了hostaddr而没有指定host，则hostaddr的值给出远端地址；使用 Kerberos 时，会发生一次反向名称查询以获得用于 Kerberos 的主机名。如果同时指定了host和hostaddr，则hostaddr的值给出远端地址；host的值会被忽略，除非使用 Kerberos，此时该值用于 Kerberos 认证。（注意，如果传给libpq的主机名不是位于hostaddr的那台机器的名字，认证很可能会失败。）另外，在~/.pgpass中是用host而不是hostaddr来标识连接的（见第 28.12 节）。

如果既没有主机名也没有主机地址，libpq 会使用本地 Unix 域套接字连接；在没有 Unix 域套接字的机器上，则会尝试连接到 localhost。

port

连接到服务器主机的端口号，或者Unix域连接的套接字文件名扩展。

dbname

数据库名称。默认为与用户名相同。

user

建立连接所用的 PostgreSQL 用户名。默认与运行应用程序的操作系统用户名相同。

password

服务器要求密码认证时所使用的密码。

connect_timeout

连接时的最长等待时间，以秒为单位（写成十进制整数字符串）。零或未指定表示无限等待。不建议使用小于2秒的超时时间。

options

发送给服务器的命令行选项。

tty

忽略此参数（以前用于指定服务器调试输出的发送位置）。

sslmode

这个选项决定是否与服务器协商SSL连接以及协商的优先级。有四种模式：disable只尝试不加密的SSL连接；allow会协商，先尝试非SSL连接，失败后再尝试SSL连接；prefer（默认值）会协商，先尝试SSL连接，失败后再尝试普通的非SSL连接；require只尝试SSL连接。

如果PostgreSQL编译时没有包含 SSL 支持，使用选项require将导致错误，而选项allow和prefer会被接受，但libpq实际上不会尝试SSL连接。

requiressl

此选项已弃用，请改用 sslmode 设置。

如果设置为1，则需要与服务器建立SSL连接（这相当于`sslmode require`）。libpq将拒绝连接，如果服务器不接受 SSL连接。如果设置为0（默认值），libpq将与服务器协商连接类型（相当于`sslmode prefer`）。此选项仅在 PostgreSQL 编译时启用了 SSL 支持的情况下可用。

`krbsrvname`

使用 Kerberos 5 或 GSSAPI 认证时所用的 Kerberos 服务名。这必须与服务器配置中指定的服务名匹配，Kerberos 认证才能成功。（另请参见第 20.2.3 节。）

`service`

用于额外参数的服务名称。它指定了`pg_service.conf`中保存额外连接参数的服务名称。这允许应用程序只指定一个服务名称，以便可以集中维护连接参数。参见第 28.13 节。

如果某个参数未指定，则会检查相应的环境变量（见第 28.11 节）。如果环境变量也未设置，则使用上面指出的内建默认值。

`PQsetdbLogin`

开启一个到数据库服务器的新连接。

```
PGconn *PQsetdbLogin(const char *pghost,
                     const char *pgport,
                     const char *pgoptions,
                     const char *pgtty,
                     const char *dbName,
                     const char *login,
                     const char *pwd);
```

这是 `PQconnectdb` 的前身，使用固定的一组参数。除缺失参数始终采用默认值之外，功能相同。对于要使用默认值的任意固定参数，请传入 `NULL` 或空字符串。

`PQsetdb`

开启一个到数据库服务器的新连接。

```
PGconn *PQsetdb(char *pghost,
                char *pgport,
                char *pgoptions,
                char *pgtty,
                char *dbName);
```

这是一个调用`PQsetdbLogin`的宏，其中为*`login`*和*`pwd`*参数使用空指针。提供它是为了向后兼容非常老的程序。

`PQconnectStart`
`PQconnectPoll`

以非阻塞的方式建立一个到数据库服务器的连接。

```
PGconn *PQconnectStart(const char *conninfo);


PostgresPollingStatusType PQconnectPoll(PGconn *conn);
```

这两个函数被用来开启一个到数据库服务器的连接，
这样你的应用的执行线程不会因为远程的I/O而被阻塞。这种方法的要点在于等待 I/O 完成可以在应用的主循环中发生，而不是在PQconnectdb中，
并且因此应用能够把这种操作和其他动作并行处理。

在PQconnectStart中，数据库连接使用从字符串conninfo中取得的参数创建，
这与前面描述的PQconnectdb相同。

只要满足若干限制，PQconnectStart和PQconnectPoll都不会阻塞：

- 必须恰当地使用 hostaddr 和 host 参数，以确保不会进行名称和反向名称查询。详细信息请参见上文PQconnectdb中这些参数的说明。
- 如果你调用PQtrace，确保接收追踪输出的流对象不会阻塞。
- 如后文所述，你要确保在调用PQconnectPoll之前，套接字处于合适的状态。

要开始无阻塞的连接请求，可调用conn = PQconnectStart("*connection_info_string*")。如果conn为空，则libpq无法分配一个新的PGconn结构体。否则，一个有效的PGconn指针会被返回（不过还没有表示一个到数据库的有效连接）。从PQconnectStart返回后，调用status = PQstatus(conn)。如果status等于CONNECTION_BAD，则PQconnectStart失败。

如果PQconnectStart成功，下一个阶段是轮询libpq，这样它能够继续进行连接序列。
使用PQsocket(conn)来获得该数据库连接底层的套接字描述符。
这样循环：如果PQconnectPoll(conn)上一次返回PGRES_POLLING_READING，
等到该套接字准备好读取（按照select()、poll()或类似的系统函数所指示的）。
则再次调用PQconnectPoll(conn)。反之，如果PQconnectPoll(conn)
上一次返回PGRES_POLLING_WRITING，等到该套接字准备好写入，则再次调用
PQconnectPoll(conn)。如果你还没有调用过PQconnectPoll，即刚刚调用过
PQconnectStart之后，行为就像是它上次返回了PGRES_POLLING_WRITING。
持续这个循环直到PQconnectPoll(conn)返回PGRES_POLLING_FAILED
指示连接过程已经失败，或者返回PGRES_POLLING_OK指示连接已经被成功地建立。

在连接过程中的任何时刻，都可以通过调用PQstatus来检查连接状态。
如果该调用返回CONNECTION_BAD，则连接过程已经失败；如果返回CONNECTION_OK，
则连接已就绪。通过以下函数的返回值也同样可以检测这两种状态：PQconnectPoll，
该函数已在上文介绍。在异步连接过程中（也仅在此过程中）还可能出现其他状态。
它们指明连接过程的当前阶段，例如可以用来向用户提供反馈。这些状态如下：

CONNECTION_STARTED

等待连接被建立。

CONNECTION_MADE

连接 OK，等待发送。

CONNECTION_AWAITING_RESPONSE

等待来自服务器的一个回应。

CONNECTION_AUTH_OK

收到认证，等待后端启动结束。

CONNECTION_SSL_STARTUP

协商 SSL 加密。

CONNECTION_SETENV

协商环境驱动的参数设置。

注意，尽管这些常量将保留（为了保持兼容性），应用绝不应该依赖它们以特定顺序出现、依赖它们全部出现，或者依赖状态总是这些已文档化的值之一。应用可以这样做：

```
switch(PQstatus(conn))
{
        case CONNECTION_STARTED:
            feedback = "Connecting...";
            break;

        case CONNECTION_MADE:
            feedback = "Connected to server...";
            break;
.
.
.
        default:
            feedback = "Connecting...";
}
```

在使用PQconnectPoll时，连接参数connect_timeout会被忽略：判断是否超时是应用的责任。除此之外，PQconnectStart后面跟着PQconnectPoll循环等效于PQconnectdb。

注意如果PQconnectStart返回一个非空的指针，你必须在用完它之后调用PQfinish来处理该结构体和任何相关的内存块。即使连接尝试失败或被放弃时也必须完成这些工作。

PQconndefaults

返回默认连接选项。

```
PQconninfoOption *PQconndefaults(void);

typedef struct
{
    char   *keyword;   /* 该选项的关键词 */
    char   *envvar;    /* 后备环境变量名 */
    char   *compiled;  /* 编译时设置的后备默认值 */
    char   *val;       /* 选项的当前值，或者 NULL */
    char   *label;     /* 连接对话框中字段的标签 */
    char   *dispchar;  /* 指示如何在连接对话框中显示此字段。可取值：
                          ""          显示输入的值
                          "*"         密码字段 – 隐藏值
                          "D"         调试选项 – 默认不显示 */
    int     dispsize;  /* 对话框中的字段宽度，以字符计 */
} PQconninfoOption;
```

返回一个连接选项数组。这可以用来确定所有可能的PQconnectdb选项和它们的当前默认值。返回值指向一个PQconninfoOption结构体的数组，

该数组以一个包含空keyword指针的条目结束。如果无法分配内存，则返回空指针。
注意当前默认值（val字段）将依赖于环境变量和其他上下文。
调用者必须把连接选项数据当作只读对待。

在处理完选项数组后，把它交给PQconninfoFree释放。如果没有这么做，每次调用PQconndefaults都会导致一小部分内存泄漏。

PQfinish

关闭与服务器的连接。同时释放PGconn对象使用的内存。

```
void PQfinish(PGconn *conn);
```

注意，即使与服务器的连接尝试失败（由PQstatus指示），应用也应当调用PQfinish来释放PGconn对象使用的内存。不能在调用PQfinish之后再使用PGconn指针。

PQreset

重置与服务器的通信通道。

```
void PQreset(PGconn *conn);
```

此函数将关闭与服务器的连接，并尝试重新建立到同一服务器的新连接，
使用之前使用过的所有参数。这可能有助于在工作连接丢失后的错误恢复。

PQresetStart
PQresetPoll

以非阻塞方式重置与服务器的通信通道。

```
int PQresetStart(PGconn *conn);

PostgresPollingStatusType PQresetPoll(PGconn *conn);
```

这些函数会关闭与服务器的连接，并尝试重新建立到同一服务器的新连接，
使用之前使用过的所有参数。如果原本可用的连接丢失，这可以用于错误恢复。
它们与上文的 PQreset 不同之处在于采用非阻塞方式。它们受到与 PQconnectStart 和 PQconnectPoll 相同的限制。

要开始重置连接，请调用 PQresetStart。如果返回 0，表示重置失败。如果返回 1，则使用 PQresetPoll 轮询重置过程，方式与使用 PQconnectPoll 建立连接完全相同。

28.2. 连接状态函数

这些函数可用于查询现有数据库连接对象的状态。

提示

编写 libpq 应用程序时，应注意维护 PGconn 的抽象。请使用下述访问函数获取 PGconn 的内容。避免直接引用 PGconn 结构的字段，因为这些字段将来可能改变。（从 PostgreSQL 6.4 版开始，PGconn 背后的 struct 定义甚至不再在 libpq-fe.h 中提供。如果你有直接访问 PGconn 字段的旧代码，可以通过同时包含 libpq-int.h 来继续使用它，但我们建议你尽快修改代码。）

以下函数返回建立连接时确定的参数值。这些值在PGconn对象的生命周期内保持不变。

PQdb

返回该连接的数据库名。

```
char *PQdb(const PGconn *conn);
```

PQuser

返回该连接的用户名。

```
char *PQuser(const PGconn *conn);
```

PQpass

返回该连接的密码。

```
char *PQpass(const PGconn *conn);
```

PQhost

返回连接的服务器主机名。可能是主机名、IP 地址或者一个目录路径（如果通过 Unix 套接字连接）。

```
char *PQhost(const PGconn *conn);
```

PQport

返回连接的端口。

```
char *PQport(const PGconn *conn);
```

PQtty

返回该连接用于调试的TTY。（此功能已过时，因为服务器不再使用TTY设置，但为保持向后兼容仍保留此函数。）

```
char *PQtty(const PGconn *conn);
```

PQoptions

返回连接请求中传递的命令行选项。

```
char *PQoptions(const PGconn *conn);
```

下列函数返回的状态数据会随着在PGconn对象上执行的操作而变化。

PQstatus

返回该连接的状态。

```
ConnStatusType PQstatus(const PGconn *conn);
```

该状态可以是一系列值之一。不过，其中只有两个在一个异步连接过程之外可见：CONNECTION_OK和CONNECTION_BAD。一个到数据库的完好连接的状态为CONNECTION_OK。一个失败的连接尝试则由状态CONNECTION_BAD表示。通常，一个 OK 状态将一直保持到PQfinish，但是一次通信失败可能导致该状态过早地改变为CONNECTION_BAD。在那种情况下，该应用可以通过调用PQreset尝试恢复。

关于其他可能会被返回的状态代码，请见PQconnectStart和PQconnectPoll的条目。

PQtransactionStatus

返回服务器的当前事务内状态。

```
PGTransactionStatusType PQtransactionStatus(const PGconn *conn);
```

该状态可能是PQTRANS_IDLE（当前空闲）、PQTRANS_ACTIVE（一个命令运行中）、PQTRANS_INTRANS（空闲，处于一个有效的事务块中）或者PQTRANS_INERROR（空闲，处于一个失败的事务块中）。如果该连接异常，将会报告PQTRANS_UNKNOWN。
只有当一个查询已经被发送给服务器并且还没有完成时，才会报告PQTRANS_ACTIVE。

> **小心**
>
> 在使用参数autocommit被设置为 off 的PostgreSQL 7.3 服务器时，PQtransactionStatus会给出不正确的结果。服务器端的 autocommit 特性已被废弃，在之后的服务器版本中不存在。

PQparameterStatus

查找服务器某个参数的当前设置。

```
const char *PQparameterStatus(const PGconn *conn, const char *paramName);
```

服务器会在连接启动时，以及某些参数值发生变化时，自动报告这些参数值。PQparameterStatus可用于查询这些设置。如果已知该参数，则返回其当前值；如果未知，则返回NULL。

截至当前发行版本会报告的参数包括server_version、server_encoding、client_encoding、is_superuser、session_authorization、DateStyle、TimeZone、integer_datetimes和standard_conforming_strings。（server_encoding、TimeZone和integer_datetimes在 8.0 之前的发行中不被报告；standard_conforming_strings在 8.1 之前的发行中不被报告。）注意server_version、server_encoding和integer_datetimes在启动后不能改变。

3.0 之前协议的服务器不报告参数设置，但libpq包含了为server_version和client_encoding获取值的逻辑。鼓励应用使用PQparameterStatus而不是ad hoc代码来确定这些值。（但要注意，在 3.0 之前的连接上，连接启动之后通过SET修改client_encoding不会被PQparameterStatus反映。）对于server_version，另见PQserverVersion，它以一种更容易比较的数字形式返回该信息。

如果没有报告standard_conforming_strings的值，应用可以假定它是false，即在字符串字面量中反斜线被当作转义符。此外，
这个参数的存在可以被当作接受转义字符串语法（E'...'）的一个标志。

返回的指针虽然被声明为const，但实际上指向与PGconn结构体关联的可变存储。
不能假定该指针在执行其他查询后仍然有效。

PQprotocolVersion

查询正在使用的前端/后端协议。

```
int PQprotocolVersion(const PGconn *conn);
```

应用程序可以使用此函数判断是否支持某些特性。目前，可能的值为 2（协议 2.0）、3（协议 3.0）或零（连接异常）。连接启动完成后，协议版本不会改变，但理论上可能在连接重置期间改变。通常，与 PostgreSQL 7.4 或更高版本的服务器通信时使用协议 3.0；7.4 之前的服务器仅支持协议 2.0。（协议 1.0 已过时，libpq 不支持它。）

`PQserverVersion`

返回一个表示后端版本的整数。

```
int PQserverVersion(const PGconn *conn);
```

应用可能会使用这个函数来判断它们连接到的数据库服务器的版本。
这个数字是这样形成的：将主版本、次版本和修订版本号分别转换成两位十进制数，
然后把它们拼接在一起。例如，版本7.4.2将被返回为70402，而版本8.
1将被返回为80100（不显示前导零）。如果连接无效则返回零。

`PQerrorMessage`

返回连接上的一个操作最近产生的错误消息。

```
char *PQerrorMessage(const PGconn *conn);
```

几乎所有 libpq 函数在失败时都会设置一条供 `PQerrorMessage` 返回的消息。注意，按照 libpq 的约定，非空的 `PQerrorMessage` 结果可能包含多行，并以换行符结尾。调用者不应直接释放该结果；当关联的 `PGconn` 句柄被传给 `PQfinish` 时，结果会被释放。不能假定在对 `PGconn` 结构体执行其他操作后，结果字符串仍保持不变。

`PQsocket`

获取与服务器相连的套接字的文件描述符编号。有效描述符大于或等于 0。结果为 -1 表示当前没有打开服务器连接（在普通操作期间这将不会改变，但是在连接设置或重置期间可能改变）。

```
int PQsocket(const PGconn *conn);
```

`PQbackendPID`

返回处理这个连接的后端进程的进程ID（PID）。

```
int PQbackendPID(const PGconn *conn);
```

后端 PID 可用于调试，也可与 `NOTIFY` 消息进行比较（消息包含发出通知的后端进程的 PID）。注意，该 PID 属于在数据库服务器主机上运行的进程，而非本地主机上的进程！

`PQgetssl`

返回在连接中使用的SSL结构体，如果未使用SSL，则返回NULL。

```
SSL *PQgetssl(const PGconn *conn);
```

该结构体可以用来验证加密级别、检查服务器证书等。有关此结构体的信息，请参阅OpenSSL文档。

必须定义 `USE_SSL` 才能得到这个函数正确的原型。这样做还会自动包含来自OpenSSL的`ssl.h`。

28.3. 命令执行函数

与数据库服务器的连接成功建立后，此处描述的函数用于执行 SQL 查询和命令。

28.3.1. 主要函数

`PQexec`

提交一个命令给服务器并且等待结果。

```
PGresult *PQexec(PGconn *conn, const char *command);
```

返回一个`PGresult`指针或者可能是一个空指针。
除了内存不足或严重错误（例如无法将命令发送到服务器）之外，通常都会返回非空指针。
如果返回了空指针，应把它当作一个`PGRES_FATAL_ERROR`结果对待。使用`PQerrorMessage`可以获取有关这类错误的更多信息。

命令串中允许包含多条 SQL 命令（用分号分隔）。在单次`PQexec`调用中发送的多个查询会在单个事务中处理，除非查询串中包含显式的`BEGIN/COMMIT`命令把它分成多个事务。但注意，返回的`PGresult`结构只描述该串中最后一条被执行的命令的结果。如果其中一条命令失败，对串的处理就到此为止，返回的`PGresult`描述该错误情况。

`PQexecParams`

提交一个命令给服务器并且等待结果，它可以在 SQL 命令文本之外独立地传递参数。

```
PGresult *PQexecParams(PGconn *conn,
                       const char *command,
                       int nParams,
                       const Oid *paramTypes,
                       const char * const *paramValues,
                       const int *paramLengths,
                       const int *paramFormats,
                       int resultFormat);
```

`PQexecParams`与`PQexec`相似，
但是提供了额外的功能：参数值可以与命令字符串分开指定，
并且可以以文本或二进制格式请求查询结果。`PQexecParams` 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

如果使用了参数，它们在命令字符串中被引用为`$1`、`$2`等。*`nParams`*是所提供参数的数量；它是数组*`paramTypes[]`*、*`paramValues[]`*、*`paramLengths[]`*和*`paramFormats[]`*的长度。（当*`nParams`*为零时，数组指针可以为`NULL`。）*`paramTypes[]`*按 OID 指定要赋给参数符号的数据类型。如果*`paramTypes`*为`NULL`，或数组中某个特定元素为零，
服务器将以与处理无类型文字字符串相同的方式为参数符号指定数据类型。*`paramValues[]`*指定参数的实际值。此数组中的空指针表示对应的参数为 null；否则该指针指向一个以零结尾的文本字符串（文本格式）或服务器所预期格式的二进制数据（二进制格式）。*`paramLengths[]`*指定二进制格式参数的实际数据长度。对 null 参数和文本格式参数会忽略它。当没有二进制参数时，数组指针可以为空。*`paramFormats[]`*
指定参数是文本（在数组中放一个零）还是二进制（在数组中放一个一）。
如果数组指针为空，则所有参数都被当作文本。*`resultFormat`*
为零时以文本格式获取结果，为一时以二进制格式获取结果。
（目前还没有办法让不同的结果列使用不同的格式，尽管底层协议支持这一点。）

PQexecParams 相对于 PQexec 的主要优点是可以将参数值与命令字符串分开，从而避免繁琐且容易出错的加引号和转义工作。和PQexec不同，PQexecParams至多允许在给定串中出现一个 SQL 命令（其中可以有分号，但是不能有超过一个非空命令）。这是底层协议的一个限制，但是有助于抵抗 SQL 注入攻击。

> **提示**
>
> 通过 OID 指定参数类型较为繁琐，尤其是在不希望将具体 OID 值写死在程序中时。不过，即使服务器无法自行确定参数类型，或者推断出的类型与你所需的不同，
> 也可以避免直接指定 OID。在 SQL 命令文本中，为参数符号添加显式类型转换，即可指定要发送的数据类型。例如：
>
> ```
> SELECT * FROM mytable WHERE x = $1::bigint;
> ```
>
> 这会强制将参数 $1 当作 bigint，而默认情况下会为它分配与 x 相同的类型。以二进制格式发送参数值时，强烈建议采用这种方式，或直接指定类型的数值 OID，来明确决定参数类型。因为二进制格式的冗余比文本格式更少，
> 服务器发现类型不匹配错误的机会也更少。

PQprepare

提交一个请求用给定参数创建一个预备语句并且等待完成。

```
PGresult *PQprepare(PGconn *conn,
                    const char *stmtName,
                    const char *query,
                    int nParams,
                    const Oid *paramTypes);
```

PQprepare 创建一个预备语句，供随后使用 PQexecPrepared 执行。此功能允许将被反复使用的命令只解析和规划一次，而不是每次执行时都重复进行。PQprepare 仅在使用协议 3.0 及更高版本的连接中受支持，使用协议 2.0 时会失败。

该函数从*query*串创建一个名为*stmtName*的预备语句，该串必须包含一个单一 SQL 命令。*stmtName*可以是""来创建一个未命名语句，
在这种情况下任何已存在的未命名语句将被自动替换。否则，
如果语句名称已经在当前会话中被定义，则是一种错误。如果使用了任何参数，
它们在查询中以$1、$2等引用。*nParams* 是在数组 *paramTypes[]* 中预先指定了类型的参数数量（当*nParams*为零时，该数组指针可以是NULL）。*paramTypes[]*通过 OID 指定要赋予给参数符号的数据类型。如果*paramTypes*是NULL或者该数组中任何特定元素为零，
服务器会用对待未指定类型的字符串字面量的方式为参数符号推测一种数据类型。还有，
查询能够使用编号高于*nParams*的参数符号，它们的数据类型也会被自动推测。

正如PQexec一样，结果通常是一个PGresult对象，其内容代表服务器端成功或失败。
返回空指针表示内存不足，或者根本无法发送命令。关于错误的更多信息请见PQerrorMessage。

目前，没有办法确定未在*paramTypes[]*中指定类型的任何参数实际被推测出的数据类型。
这是libpq的一个缺失，将来可能会补上。

用于PQexecPrepared的预备语句也可以通过执行 SQLPREPARE语句来创建。（但PQprepare更灵活，因为它不要求预先指定参数类型。）此外，虽然没有libpq函数可删除预备语句，
但可以使用 SQLDEALLOCATE语句来完成。

PQexecPrepared

发送一个请求来执行一个带有给定参数的预备语句，并等待结果。

```
PGresult *PQexecPrepared(PGconn *conn,
                         const char *stmtName,
                         int nParams,
                         const char * const *paramValues,
                         const int *paramLengths,
                         const int *paramFormats,
                         int resultFormat);
```

PQexecPrepared类似于PQexecParams，但它通过已创建的预备语句的名称指定要执行的命令，而非提供查询字符串。此功能使重复使用的命令只需解析和规划一次，而不必在每次执行时都进行这些工作。该语句必须事先在当前会话中创建为预备语句。PQexecPrepared 仅支持使用协议 3.0 及更高版本的连接；使用协议 2.0 时会失败。

参数与PQexecParams相同，只是给出了预备语句的名称而不是查询字符串，并且*paramTypes[]*参数不存在（因为在创建预备语句时已确定了参数类型）。

PGresult结构封装了服务器返回的结果。libpq应用程序员应该小心地维护PGresult的抽象。使用下面的访问器函数来获取PGresult的内容。避免直接引用PGresult结构的字段，因为它们在未来可能会改变。

PQresultStatus

返回该命令的结果状态。

```
ExecStatusType PQresultStatus(const PGresult *res);
```

PQresultStatus可以返回下列值之一：

PGRES_EMPTY_QUERY

发送给服务器的字符串为空。

PGRES_COMMAND_OK

一个不返回数据的命令成功完成。

PGRES_TUPLES_OK

一个返回数据的命令（例如SELECT或者SHOW）成功完成。

PGRES_COPY_OUT

从服务器复制出数据的传输开始。

PGRES_COPY_IN

复制数据到服务器的传输开始。

PGRES_BAD_RESPONSE

无法理解服务器的响应。

PGRES_NONFATAL_ERROR

发生了一次非致命错误（一个提示或警告）。

PGRES_FATAL_ERROR

发生了一次致命错误。

如果结果状态是PGRES_TUPLES_OK，那么可以使用下面描述的函数来检索查询返回的行。注意，一个恰好检索到零行的SELECT命令仍然显示为PGRES_TUPLES_OK。PGRES_COMMAND_OK用于永远不会返回行的命令（INSERT、UPDATE等）。一个PGRES_EMPTY_QUERY响应可能表示客户端软件中的一个 bug。

一个状态为PGRES_NONFATAL_ERROR的结果将不会被PQexec或者其他查询执行函数直接返回，这类结果将被传递给通知处理器（见第 28.10 节）。

PQresStatus

将 PQresultStatus 返回的枚举值转换为描述该状态码的字符串常量。调用者不应释放此结果。

```
char *PQresStatus(ExecStatusType status);
```

PQresultErrorMessage

返回与命令关联的错误消息；如果没有错误，则返回空字符串。

```
char *PQresultErrorMessage(const PGresult *res);
```

如果发生了错误，返回的字符串会以换行符结尾。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

紧跟着一个PQexec 或 PQgetResult调用，PQerrorMessage（在连接上）将返回与PQresultErrorMessage相同的字符串（在结果上）。不过，一个PGresult 将保持它的错误消息直到被销毁，而连接的错误消息将在后续操作被执行时被更改。当你想要知道与一个特定PGresult相关的状态，使用PQresultErrorMessage。而当你想要知道连接上最后一个操作的状态，使用PQerrorMessage。

PQresultErrorField

返回错误报告中的单个字段。

```
char *PQresultErrorField(const PGresult *res, int fieldcode);
```

fieldcode 是错误字段标识符，参见下文列出的符号。如果 PGresult 不是错误或警告结果，或者不包含指定字段，则返回 NULL。字段值通常不含末尾换行符。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

可以使用以下字段代码：

PG_DIAG_SEVERITY

严重性。字段的内容是ERROR、FATAL或PANIC（在一个错误消息中）。或者是WARNING、NOTICE、DEBUG、INFO或LOG（在一个提示消息中）。或者是其中之一的一个本地化翻译。总是存在。

PG_DIAG_SQLSTATE

错误的 SQLSTATE 代码。SQLSTATE 代码标识所发生错误的类型；前端应用可以用它针对特定的数据库错误执行特定操作（例如错误处理）。可能的 SQLSTATE 代码列表见附录 A。这个字段不可本地化，并且总是存在。

PG_DIAG_MESSAGE_PRIMARY

主要的人类可读的错误消息（通常是一行）。总是存在。

PG_DIAG_MESSAGE_DETAIL

细节：一个可选的次级错误消息，它携带了关于问题的更多细节。可能有多行。

PG_DIAG_MESSAGE_HINT

提示：一个关于如何处理该问题的可选建议。
它与细节的区别在于它提供了建议（可能不合适）而不是确切事实。可能有多行。

PG_DIAG_STATEMENT_POSITION

包含一个十进制整数的字符串，它表示一个错误游标位置，
该位置是原始语句字符串的索引。第一个字符的索引是 1，位置以字符计算而不是以字节计算。

PG_DIAG_INTERNAL_POSITION

这被定义为与PG_DIAG_STATEMENT_POSITION字段相同，
但是它被用在游标位置引用一个内部产生的命令而不是客户端提交的命令时。
当这个字段出现时，PG_DIAG_INTERNAL_QUERY字段将总是出现。

PG_DIAG_INTERNAL_QUERY

一个失败的内部产生的命令的文本。例如，这可能是由一个 PL/pgSQL 函数发出的 SQL 查询。

PG_DIAG_CONTEXT

指示错误发生的上下文。
当前这包括活动过程语言函数的调用栈追踪以及内部生成的查询。追踪是每行一项，
最近的排在最前面。

PG_DIAG_SOURCE_FILE

报告错误的源代码所在的文件名。

PG_DIAG_SOURCE_LINE

报告错误的源代码行号。

PG_DIAG_SOURCE_FUNCTION

报告错误的源代码函数的名字。

客户端负责按自身需要格式化所显示的信息，尤其应在需要时将长行折行。
错误消息字段中的换行符应当视为段落分隔，而非行分隔。

libpq 内部产生的错误包含严重性和主要消息，但通常没有其他字段。使用 3.0 之前协议的服务器返回的错误包含严重性和主要消息，有时还包含详细消息，但没有其他字段。

注意，错误字段只对PGresult对象有效，对PGconn对象无效。没有PQerrorField函数。

PQclear

释放与 PGresult 关联的存储空间。每个命令结果在不再需要时都应通过 PQclear 释放。

```
void PQclear(PGresult *res);
```

你可以在需要时一直保留PGresult对象；它不会在你发出新命令时消失，甚至在关闭连接后也不会消失。要销毁它，你必须调用PQclear。否则应用程序会发生内存泄漏。

PQmakeEmptyPGresult

用给定的状态，构造一个空PGresult对象。

```
PGresult *PQmakeEmptyPGresult(PGconn *conn, ExecStatusType
status);
```

这是libpq内部用于分配并初始化一个空PGresult对象的函数。如果无法分配内存，此函数返回NULL。将它导出供外部调用，是因为一些应用需要自行生成结果对象，特别是带有错误状态的对象。如果*conn* 不为 null，并且*status*表示错误，指定连接的当前错误消息会被复制到PGresult中。注意，最终应对该对象调用PQclear，就像处理libpq自身返回的PGresult一样。

28.3.2. 检索查询结果信息

这些函数用于从表示成功查询结果的PGresult对象（即状态为PGRES_TUPLES_OK的对象）中提取信息。对于其他状态值的对象，这些函数会将结果视为零行、零列。

PQntuples

返回查询结果中的行（元组）数。因为它返回一个整数结果，在 32 位操作系统上大的结果集可能会导致返回值溢出。

```
int PQntuples(const PGresult *res);
```

PQnfields

返回查询结果中每一行的列（字段）数。

```
int PQnfields(const PGresult *res);
```

PQfname

返回给定列号对应的列名。列号从 0 开始。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

```
char *PQfname(const PGresult *res,
              int column_number);
```

如果列号超出范围，将返回NULL。

PQfnumber

返回与给定列名相关联的列号。

```
int PQfnumber(const PGresult *res,
              const char *column_name);
```

如果给定的名字不匹配任何列，将返回 -1。

给定名称按 SQL 命令中的标识符处理，即除非用双引号引用，否则会转换为小写。例如，对于以下 SQL 命令生成的查询结果：

```
SELECT 1 AS FOO, 2 AS "BAR";
```

会得到以下结果：

```
PQfname(res, 0)              foo
PQfname(res, 1)              BAR
PQfnumber(res, "FOO")        0
PQfnumber(res, "foo")        0
PQfnumber(res, "BAR")        -1
PQfnumber(res, "\"BAR\"")    1
```

`PQftable`

返回给定列所取自的表的 OID。列号从 0 开始。

```
Oid PQftable(const PGresult *res,
             int column_number);
```

如果列号超出范围、指定的列不是对表列的简单引用，或者使用 3.0 之前的协议，则返回 `InvalidOid`。可以查询系统表 `pg_class`，确定所引用的具体表。

包含 libpq 头文件后，将定义类型 `Oid` 和常量 `InvalidOid`。它们都属于某种整数类型。

`PQftablecol`

返回指定查询结果列所对应的表列在表中的列号。查询结果的列号从 0 开始，而表列的编号非零。

```
int PQftablecol(const PGresult *res,
                int column_number);
```

如果列号超出范围、指定的列不是对表列的简单引用，或者使用 3.0 之前的协议，则返回零。

`PQfformat`

返回表示给定列格式的格式代码。列号从 0 开始。

```
int PQfformat(const PGresult *res,
              int column_number);
```

格式代码零表示文本数据，格式代码一表示二进制数据。（其他代码保留供将来定义。）

`PQftype`

返回与给定列号相关联的数据类型。被返回的整数是该类型的内部 OID 号。列号从 0 开始。

```
Oid PQftype(const PGresult *res,
            int column_number);
```

可以查询系统表 `pg_type` 来获取各种数据类型的名称和属性。内置数据类型的 OID 定义在源码树的 `src/include/catalog/pg_type.h` 文件中。

`PQfmod`

返回与给定列号相关联的列的类型修饰符。列号从 0 开始。

```
int PQfmod(const PGresult *res,
           int column_number);
```

修饰符值的含义由数据类型决定，通常表示精度或大小限制。值 -1 表示“没有可用信息”。大多数数据类型不使用修饰符，此时该值始终为 -1。

PQfsize

返回给定列号对应列的大小，以字节计。列号从 0 开始。

```
int PQfsize(const PGresult *res,
            int column_number);
```

PQfsize 返回数据库行中为该列分配的空间，即服务器内部表示该数据类型所需的大小。（因此，它对客户端用处不大。）负值表示该数据类型是变长类型。

PQbinaryTuples

如果PGresult包含二进制数据，返回 1。如果包含的是文本数据，返回 0。

```
int PQbinaryTuples(const PGresult *res);
```

此函数已弃用（与 COPY 配合使用的情况除外），因为单个 PGresult 可能在部分列中包含文本数据，而在其他列中包含二进制数据。推荐使用 PQfformat。只有结果中的所有列都采用二进制格式（格式 1）时，PQbinaryTuples 才返回 1。

PQgetvalue

返回 PGresult 中某一行的单个字段值。行号和列号从 0 开始。调用者不应直接释放结果；当关联的 PGresult 句柄被传给 PQclear 时，结果会被释放。

```
char *PQgetvalue(const PGresult *res,
                 int row_number,
                 int column_number);
```

对于文本格式的数据，PQgetvalue 返回字段值的字符串表示，以零字节结尾。对于二进制格式的数据，返回值采用该数据类型的 typsend 和 typreceive 函数所决定的二进制表示。（这种情况下，值后面实际上也有一个零字节，但通常没有用处，
因为值本身很可能包含零字节。）

如果字段值为 null，则返回空字符串。关于如何区分 null 值与空字符串值，参见 PQgetisnull。

PQgetvalue 返回的指针指向属于 PGresult 结构体的存储空间。不应修改它所指向的数据；如果需要在 PGresult 结构体的生命周期结束后继续使用这些数据，就必须显式地将数据复制到其他存储空间。

PQgetisnull

检查字段是否为 null 值。行号和列号从 0 开始。

```
int PQgetisnull(const PGresult *res,
                int row_number,
                int column_number);
```

如果字段为 null，此函数返回 1；如果包含非 null 值，则返回 0。（注意，对于 null 字段，PQgetvalue 返回空字符串，而非空指针。）

PQgetlength

返回字段值的实际长度，以字节计。行号和列号从 0 开始。

```
int PQgetlength(const PGresult *res,
                int row_number,
                int column_number);
```

这是该数据值的实际数据长度，即 PQgetvalue 所指对象的大小。对于文本格式的数据，它与 strlen() 的结果相同。对于二进制格式，这是必不可少的信息。注意，不应依赖 PQfsize 来获取实际数据长度。

PQprint

将所有行输出到指定的输出流，并可选择输出列名。

```
void PQprint(FILE *fout,      /* 输出流 */
             const PGresult *res,
             const PQprintOpt *po);
typedef struct
{
    pqbool  header;      /* 打印输出字段标题和行数 */
    pqbool  align;       /* 填充并对齐字段 */
    pqbool  standard;    /* 旧的简陋格式 */
    pqbool  html3;       /* 输出 HTML 表格 */
    pqbool  expanded;    /* 展开表格 */
    pqbool  pager;       /* 需要时使用分页器输出 */
    char    *fieldSep;   /* 字段分隔符 */
    char    *tableOpt;   /* 用于 HTML 表格元素的属性 */
    char    *caption;    /* HTML 表格标题 */
    char    **fieldName; /* 以空指针结尾的替换字段名数组 */
} PQprintOpt;
```

psql 以前用此函数打印查询结果，现在已不再使用。注意，此函数假定所有数据都是文本格式。

28.3.3. 检索其他结果信息

这些函数被用来从PGresult对象中抽取其他信息。

PQcmdStatus

返回来自于产生PGresult的 SQL 命令的命令状态标签。

```
char *PQcmdStatus(PGresult *res);
```

通常这就是该命令的名称，但是它可能包括额外数据，例如已被处理的行数。
调用者不应该直接释放该结果。它将在相关的PGresult句柄被传递给PQclear之后被释放。

PQcmdTuples

返回受该 SQL 命令影响的行数。

```
char *PQcmdTuples(PGresult *res);
```

此函数返回一个字符串，其中包含产生该 PGresult 的 SQL 语句所影响的行数。此函数只能在执行 SELECT、INSERT、UPDATE、DELETE、MOVE 或 FETCH 语句之后使用，也可以在对包含 INSERT、UPDATE 或 DELETE 语句的预备查询执行 EXECUTE 之后使用。如果产生 PGresult 的是其他命令，PQcmdTuples 将返回空字符串。调用者不应直接释放返回值；当关联的 PGresult 句柄被传给 PQclear 时，返回值会被释放。

PQoidValue

如果该SQL命令是一个正好将一行插入到具有 OID 的表的INSERT，或者是一个包含合适INSERT语句的预备查询的EXECUTE，这个函数返回被插入行的 OID。否则，这个函数返回InvalidOid。如果被INSERT语句影响的表不包含 OID，这个函数也将返回InvalidOid。

```
Oid PQoidValue(const PGresult *res);
```

PQoidStatus

如果该SQL命令是一个恰好插入一行的INSERT，或者是一个由合适的INSERT构成的预备语句的EXECUTE，返回一个带有被插入行 OID 的字符串。（如果INSERT没有恰好插入一行，或者目标表没有 OID，该字符串将为0。）如果该命令不是一个INSERT，返回一个空字符串。

```
char *PQoidStatus(const PGresult *res);
```

此函数已被弃用，推荐使用PQoidValue。它不是线程安全的。

28.3.4. 用于在 SQL 命令中嵌入字符串的转义

PQescapeStringConn为了让一个字符串可用于 SQL 命令而对它进行转义。当在 SQL 命令中把数据值作为字符串字面量插入时，这很有用。一些字符（例如引号和反斜杠）必须经过转义，才不会被 SQL 解析器解释成特殊含义。PQescapeStringConn执行这种操作。

> **提示**
>
> 在处理来自不可信来源的字符串时，正确转义尤其重要。
> 否则就会有安全风险：你很容易受到“SQL 注入”攻击，不期望的 SQL 命令可能会被送入数据库。

注意，当一个数据值被作为PQexecParams或相关例程中的一个独立参数传递时，没有必要做转义而且做转义也不正确。

```
size_t PQescapeStringConn(PGconn *conn,
                          char *to, const char *from, size_t length,
                          int *error);
```

PQescapeStringConn把*from*字符串的转义版本写入*to*缓冲区，对特殊字符进行转义使它们不会造成任何危害，并添加一个终止的零字节。包围PostgreSQL字符串字面量所需的单引号不包含在结果字符串中；它们应在插入转义结果的 SQL 命令中提供。*from*参数指向待转义字符串的首字符，*length*参数给出该字符串的字节数。输入不必以零字节结尾，末尾零字节也不应计入*length*。（如果在处理完*length*个字节之前遇到末尾零字节，PQescapeStringConn会在该字节处停止；这一行为类似于strncpy。）*to*必须指向一个缓冲区，其容量至少为*length*的两倍加一个字节，否则行为未定义。如果*to*与*from*字符串重叠，行为同样未定义。

如果*error*参数不是NULL，则成功时将*error设为零，出错时设为非零。
目前唯一可能的错误是源字符串中存在无效的多字节编码。出错时仍会生成输出字符串，
但预计服务器会因其格式错误而拒绝它。发生错误时，无论*error*是否为NULL，都会在*conn*
对象中保存相应消息。

PQescapeStringConn返回写到*to*的字节数，不包括终止的零字节。

```
size_t PQescapeString (char *to, const char *from, size_t length);
```

PQescapeString是PQescapeStringConn的一个较老的、已弃用的版本；区别在于它不接受*conn*或*error*参数。因此，它无法根据连接属性（例如字符编码）调整行为，
可能给出错误的结果。此外，它也无法报告错误情况。

PQescapeString
可以在同一时刻仅使用一个PostgreSQL连接的单线程客户端程序中安全使用（在这种情况下，
它能够“在内部”取得所需信息）。在其他情形下，它存在安全隐患，应改用PQescapeString
Conn。

28.3.5. 用于在 SQL 命令中嵌入二进制串的转义

PQescapeByteaConn

对二进制数据进行转义，使其能够在 SQL 命令中用作 bytea 类型的值。与 PQescape StringConn 一样，这仅用于将数据直接插入 SQL 命令字符串的情况。

```
unsigned char *PQescapeByteaConn(PGconn *conn,
                                 const unsigned char *from,
                                 size_t from_length,
                                 size_t *to_length);
```

当某些字节值被用作一个 SQL 语句中的 bytea 字面量的一部分时，它们必须被转义（但所有字节值都可以被转义）。一般而言，要转义一个字节，
需要把它转换为与该字节值相等的三位八进制数，并在前面加上一个或两个反斜线。
单引号（'）和反斜线（\）字符有特殊的替代转义序列。更多信息见第 8.4 节。PQescape ByteaConn 执行这种操作，只转义最少必需的字节。

from 参数指向待转义字符串的首字节，*from_length* 参数给出该二进制字符串的字节数。（末尾零字节既不需要，也不计入长度。）*to_length* 参数指向用于保存转义后字符串长度的变量。该结果字符串长度包含结果末尾的零字节。

PQescapeByteaConn 返回 *from*参数所指二进制字符串的转义版本，存放在通过 malloc( ) 分配的内存中。结果不再需要时，应使用 PQfreemem() 释放该内存。返回字符串中的所有特殊字符都已替换，以便 PostgreSQL 字符串字面量解析器和 bytea 输入函数正确处理。还会添加一个末尾零字节。包围 PostgreSQL 字符串字面量所需的单引号不包含在结果字符串中。

在发生错误时，将返回一个空指针，并且一个合适的错误消息被存储在*conn*对象中。当前，
唯一可能的错误是没有足够的内存用于结果串。

PQescapeBytea

PQescapeBytea 是 PQescapeByteaConn 的旧版本，现已弃用。

```
unsigned char *PQescapeBytea(const unsigned char *from,
```

```
                                size_t from_length,
                                size_t *to_length);
```

与 PQescapeByteaConn 唯一的区别是，PQescapeBytea 不接受 PGconn 参数。因此，它无法根据连接属性（尤其是是否启用了标准符合字符串）调整行为，可能给出错误的结果。此外，它也无法在失败时返回错误消息。

PQescapeBytea 可以在同一时刻仅使用一个 PostgreSQL 连接的单线程客户端程序中安全使用（在这种情况下，它能够“在内部”取得所需信息）。在其他情形下，它存在安全隐患，应改用 PQescapeByteaConn。

PQunescapeBytea

将二进制数据的字符串表示转换为二进制数据，这是PQescapeBytea的逆操作。
以文本格式取得bytea数据时需要此操作；以二进制格式取得时则不需要。

```
unsigned char *PQunescapeBytea(const unsigned char *from, size_t
 *to_length);
```

*from*参数指向一个字符串，例如对bytea列调用PQgetvalue时返回的字符串。
PQunescapeBytea将这个字符串表示转换为二进制表示。它返回指向通过malloc()
分配的缓冲区的指针，出错时返回NULL，并将缓冲区大小存入*to_length*。
不再需要结果时，必须使用PQfreemem释放它。

此转换并不完全是PQescapeBytea的逆操作，因为从PQgetvalue
收到的字符串并非经过“转义”的形式。具体而言，这意味着无需考虑字符串引号，
因此也不需要PGconn参数。

PQfreemem

释放libpq分配的内存。

```
void PQfreemem(void *ptr);
```

释放 libpq 分配的内存，特别是 PQescapeByteaConn、PQescapeBytea、PQunescapeBytea 和 PQnotifies 分配的内存。Microsoft Windows 需要这个函数，因为它无法跨 DLL 释放内存，除非使用多线程 DLL（VC6 中的/MD）。在其他平台上，此函数与标准库函数 free() 相同。

28.4. 异步命令处理

PQexec函数足以满足普通同步应用程序提交命令的需要。不过，它有两个缺点，
对某些用户可能很重要：

- PQexec会等待命令完成。应用可能有其他工作要做（例如维护一个用户界面），
  这种情况下它不会想要阻塞等待响应。
- 因为客户端应用的执行在它等待结果时会被挂起，
  对于应用来说很难决定要不要尝试取消正在进行的命令（这可以在一个信号处理器中完成，
  但别无他法）。
- PQexec只能返回一个PGresult结构体。如果提交的命令串包含多个SQL命令，除了最后一个
  PGresult之外都会被PQexec丢弃。

不喜欢这些限制的应用可以改用PQexec构建所基于的底层函数：PQsendQuery和PQgetResult。还有PQsendQueryParams、PQsendPrepare和PQsendQueryPrepared，它们可以与PQgetResult一起使用来复制PQexecParams、PQprepare和PQexecPrepared的功能 respectively.

PQsendQuery

向服务器提交命令，不等待结果。命令发送成功时返回 1，否则返回 0（此时可使用 PQerrorMessage 获取更多失败信息）。

```
int PQsendQuery(PGconn *conn, const char *command);
```

成功调用 PQsendQuery 后，应调用 PQgetResult 一次或多次来获取结果。在 PQgetResult 返回空指针、表明命令已完成之前，不得在同一连接上再次调用 PQsendQuery。

PQsendQueryParams

向服务器提交命令及独立指定的参数，不等待结果。

```
int PQsendQueryParams(PGconn *conn,
                      const char *command,
                      int nParams,
                      const Oid *paramTypes,
                      const char * const *paramValues,
                      const int *paramLengths,
                      const int *paramFormats,
                      int resultFormat);
```

该函数等价于 PQsendQuery，但查询参数可以与查询字符串分开指定。函数参数的处理方式与 PQexecParams 相同。与 PQexecParams 一样，它不能用于协议 2.0 的连接，并且查询字符串中只允许包含一条命令。

PQsendPrepare

发送按给定参数创建预备语句的请求，不等待完成。

```
int PQsendPrepare(PGconn *conn,
                  const char *stmtName,
                  const char *query,
                  int nParams,
                  const Oid *paramTypes);
```

这是 PQprepare 的异步版本：请求发送成功时返回 1，否则返回 0。调用成功后，再调用 PQgetResult，确定服务器是否成功创建了预备语句。函数参数的处理方式与 PQprepare 相同。与 PQprepare 一样，它不能用于协议 2.0 的连接。

PQsendQueryPrepared

发送使用给定参数执行预备语句的请求，不等待结果。

```
int PQsendQueryPrepared(PGconn *conn,
                        const char *stmtName,
                        int nParams,
                        const char * const *paramValues,
                        const int *paramLengths,
                        const int *paramFormats,
```

```
                                int resultFormat);
```

该函数类似于 PQsendQueryParams，但通过已创建的预备语句的名称指定要执行的命令，而非提供查询字符串。函数参数的处理方式与 PQexecPrepared 相同。与 PQexecPrepared 一样，它不能用于协议 2.0 的连接。

PQgetResult

等待先前的PQsendQuery、PQsendQueryParams、PQsendPrepare或PQsendQueryPrepared调用产生的下一个结果，并返回该结果。当命令执行完毕且不再有其他结果时，返回空指针。

```
PGresult *PQgetResult(PGconn *conn);
```

必须反复调用 PQgetResult，直到它返回空指针，表明命令已经完成。（如果当前没有正在执行的命令，调用 PQgetResult 会立即返回空指针。）对于 PQgetResult 返回的非空指针，应使用前文介绍的 PGresult 访问函数处理相应结果。使用完毕后，不要忘记调用 PQclear 释放每个结果对象。注意，只有存在正在执行的命令，且所需响应数据尚未被 PQconsumeInput 读取时，PQgetResult 才会阻塞。

使用PQsendQuery和PQgetResult解决了PQexec
的一个问题：如果命令字符串包含多个SQL命令，就能分别获取这些命令的结果。
（这也支持一种简单的重叠处理方式：客户端可以处理某条命令的结果，
同时服务器继续处理同一命令字符串中后面的查询。）然而，调用PQgetResult
仍会导致客户端阻塞直到服务器完成下一条SQL命令。
通过恰当使用另外两个函数可以避免这一点：

PQconsumeInput

如果服务器的输入可用，消费它。

```
int PQconsumeInput(PGconn *conn);
```

PQconsumeInput 通常返回 1，表示“没有错误”；发生问题时则返回 0（此时可查看 PQerrorMessage）。注意，返回值并不说明是否实际读取了输入数据。调用 PQconsumeInput 后，应用程序可以检查 PQisBusy 和/或 PQnotifies，以确定其状态是否发生变化。

即使应用程序尚未准备好处理结果或通知，也可以调用 PQconsumeInput。此函数会读取可用数据并将其保存在缓冲区中，从而清除 select() 的可读就绪指示。因此，应用程序可以用 PQconsumeInput 立即清除 select() 的就绪条件，随后在合适的时候检查结果。

PQisBusy

如果命令仍在忙碌，则返回 1，意味着PQgetResult会阻塞等待输入。返回 0 则表示可以调用PQgetResult，并保证不会阻塞。

```
int PQisBusy(PGconn *conn);
```

PQisBusy本身将不会尝试从服务器读取数据，因此必须先调用PQconsumeInput，
否则繁忙状态将永远不会结束。

使用这些函数的典型应用程序会在主循环中通过 select() 或 poll() 等待需要响应的各种条件。其中一个条件是服务器有可读取的输入；对于 select()，这意味着 PQsocket 标识的文件描述符上有可读数据。主循环检测到输入就绪时，应调用 PQconsumeInput 读取输入，然后

调用 PQisBusy。如果 PQisBusy 返回假（0），就可以接着调用 PQgetResult。还可以调用 PQnotifies 检测 NOTIFY 消息（见第 28.7 节）。

一个使用PQsendQuery/PQgetResult
的客户端也可以尝试取消一个正在被服务器处理的命令，见第 28.5 节。但是，不管PQcancel的返回值是什么，应用都必须继续使用PQgetResult进行正常的结果读取序列。
一次成功的取消只会导致命令比不取消时更快终止。

使用上述函数可以避免在等待数据库服务器输入时阻塞。不过，
应用程序仍可能在等待向服务器发送输出时阻塞。这种情况较少见，但发送很长的 SQL 命令或数据值时可能发生。（如果应用程序通过 COPY IN 发送数据，发生的可能性则大得多。）为了防止这种情况，实现完全非阻塞的数据库操作，可以使用以下附加函数。

PQsetnonblocking

设置连接的非阻塞状态。

```
int PQsetnonblocking(PGconn *conn, int arg);
```

如果*arg*为1，则将连接状态设置为非阻塞，如果 *arg*为0，则设置为阻塞。如果成功返回0，出错返回-1。

在非阻塞状态下，对 PQsendQuery、PQputline、PQputnbytes 和 PQendcopy 的调用不会阻塞；如果需要再次调用，它们会返回错误。

请注意，PQexec不遵守非阻塞模式；如果调用它，它将以阻塞方式执行。

PQisnonblocking

返回数据库连接的阻塞状态。

```
int PQisnonblocking(const PGconn *conn);
```

如果连接设置为非阻塞模式，则返回1，如果为阻塞，则返回0。

PQflush

尝试将发送队列中的输出数据发送到服务器。成功（或发送队列为空）时返回 0；因某种原因失败时返回 -1；如果尚未能发送队列中的全部数据，则返回 1（这种情况只可能发生在非阻塞连接上）。

```
int PQflush(PGconn *conn);
```

在非阻塞连接上发送命令或数据后，应调用 PQflush。如果返回 1，就等待套接字变为可读或可写。套接字可写时，再次调用 PQflush；可读时，先调用 PQconsumeInput，再调用 PQflush。重复上述步骤，直到 PQflush 返回 0。（必须检查套接字是否可读，并用 PQconsumeInput 读完输入，因为服务器可能在尝试向客户端发送数据时阻塞，例如发送 NOTICE 消息；在客户端读取这些数据之前，服务器不会读取客户端发送的数据。）当 PQflush 返回 0 后，等待套接字变为可读，再按前述方法读取响应。

28.5. 取消进行中的查询

客户端应用程序可以使用本节介绍的函数，请求取消服务器仍在处理的命令。

PQgetCancel

创建一个数据结构，其中包含取消通过特定数据库连接发出的命令所需的信息。

```
PGcancel *PQgetCancel(PGconn *conn);
```

PQgetCancel基于一个PGconn连接对象创建PGcancel对象。如果给定的*conn*为NULL或无效连接，则返回NULL。PGcancel是不透明结构体，不应由应用程序直接访问；它只能传给PQcancel或PQfreeCancel。

PQfreeCancel

释放由以下函数创建的数据结构：PQgetCancel。

```
void PQfreeCancel(PGcancel *cancel);
```

PQfreeCancel 释放之前由 PQgetCancel 创建的数据对象。

PQcancel

请求服务器停止处理当前命令。

```
int PQcancel(PGcancel *cancel, char *errbuf, int errbufsize);
```

返回值为 1 表示取消请求成功发送，为 0 表示未成功发送。如果未成功发送，*errbuf*将填充解释性错误消息。*errbuf*必须是大小为*errbufsize*的字符数组（推荐大小为 256 字节）。

不过，成功发送请求并不保证请求会产生效果。如果取消生效，当前命令会提前终止并返回错误结果。如果取消失败（例如服务器已经处理完该命令），则不会产生任何可见结果。

如果*errbuf*是信号处理程序中的局部变量，则PQcancel可以安全地从信号处理程序中调用。对于PQcancel而言，PGcancel对象是只读的，因此也可以从与操作PGconn对象的线程不同的线程中调用。

PQrequestCancel

PQrequestCancel是以下函数的已弃用变体：PQcancel。

```
int PQrequestCancel(PGconn *conn);
```

请求服务器放弃当前命令的处理。它直接作用于PGconn对象，失败时会把错误消息存储到PGconn对象中（可通过PQerrorMessage获取）。虽然功能相同，但这种方法在多线程程序或信号处理程序中并不安全，因为它可能覆盖PGconn中的错误消息，从而破坏当前连接上正在进行的操作。

28.6. 快速路径接口

PostgreSQL提供一种快速路径接口来向服务器发送简单的函数调用。

> **提示**
>
> 此接口已显得有些过时，因为可以通过建立定义函数调用的预备语句，获得相近的性能和更多功能。随后使用二进制格式传输参数和结果来执行该语句，就可以替代快速路径函数调用。

函数PQfn请求通过快速路径接口执行服务器函数。

```
PGresult *PQfn(PGconn *conn,
               int fnid,
               int *result_buf,
               int *result_len,
               int result_is_int,
               const PQArgBlock *args,
               int nargs);

typedef struct
{
    int len;
    int isint;
    union
    {
        int *ptr;
        int integer;
    } u;
} PQArgBlock;
```

fnid 参数是要执行函数的 OID。*args* 和 *nargs* 指定传给函数的参数，必须与函数声明中的参数列表匹配。参数结构体的 *isint* 字段为真时，*u.integer* 值会以指定长度的整数发送到服务器，该长度必须是 2 或 4 字节，并会进行适当的字节序转换。*isint* 为假时，位于 **u.ptr* 的指定数量字节会原样发送；数据必须符合服务器对该函数参数数据类型的二进制传输格式要求。（将 u.ptr 声明为 int * 是历史原因；将其视为 void * 更合适。）*result_buf* 指向用于存放函数返回值的缓冲区。调用者必须事先分配足够空间来保存返回值，这里不会检查！实际结果长度以字节为单位，返回到 *result_len* 指向的整数中。如果预期结果是 2 或 4 字节整数，将 *result_is_int* 设为 1，否则设为 0。将 *result_is_int* 设为 1 后，libpq 会按需转换字节序，使结果成为适合客户端机器的 `int` 值；注意，无论是哪种允许的结果大小，传入 *result_buf 的都是 4 字节整数。*result_is_int* 为 0 时，服务器发送的二进制格式字节串会原样返回。（此时，将 result_buf 视为 void * 更合适。）

`PQfn` 总是返回有效的 `PGresult` 指针。使用结果前应检查其状态。不再需要结果时，调用者负责使用 `PQclear` 释放 `PGresult`。

注意，使用此接口时无法处理 NULL 参数、NULL 结果以及集合值结果。

28.7. 异步通知

PostgreSQL通过`LISTEN`和`NOTIFY`命令提供异步通知。客户端会话可使用`LISTEN`命令注册其感兴趣的特定通知通道（也可以用`UNLISTEN`命令停止监听）。当任何会话执行带有该通道名的`NOTIFY`命令时，所有监听该通道的会话都会异步收到通知。

libpq应用把`LISTEN`、`UNLISTEN`和`NOTIFY`命令作为普通 SQL 命令提交。随后通过调用`PQnotifies`来检测 NOTIFY 消息的到达。

函数 `PQnotifies` 从已收到但尚未处理的服务器通知消息列表中返回下一条通知。没有待处理通知时返回空指针。一旦 `PQnotifies` 返回一条通知，该通知就被视为已处理，并从通知列表中移除。

```
PGnotify *PQnotifies(PGconn *conn);

typedef struct pgNotify
{
    char *relname;              /* 通知条件名称 */
```

```
    int  be_pid;                /* 服务器进程 ID */
    char *extra;                /* 通知参数 */
} PGnotify;
```

处理完 PQnotifies 返回的 PGnotify 对象后，务必使用 PQfreemem 释放它。只需释放 PGnotify 指针；relname 和 extra 字段并非独立分配。（目前，extra字段未被使用，并且总是指向一个空串。）

> **注意**
>
> 在PostgreSQL 6.4 及之后的版本中，be_pid是发送通知的服务器进程的 PID；而在更早的版本中，它总是你自己的服务器进程的PID。

例 28.2给出了一个示例程序展示异步通知的使用。

PQnotifies实际上并不从服务器读取数据；它只是返回之前已被其他libpq函数读取的消息。在较早版本的libpq中，及时收到NOTIFY消息的唯一方法是不断提交命令，哪怕是空命令，然后在每次PQexec后检查PQnotifies。虽然这种方法仍然有效，但由于浪费处理能力，现已弃用。

当没有需要执行的实际命令时，一种更好的检查NOTIFY消息的方法是调用PQconsumeInput，然后检查PQnotifies。你可以使用select()等待服务器数据到达，这样在无事可做时就不会浪费CPU资源（参见PQsocket以获得可传给select()的文件描述符）。注意不管是用PQsendQuery/PQgetResult提交命令还是简单地使用PQexec，这种方法都能正常工作。不过，你应该记住在每次PQgetResult 或 PQexec之后检查PQnotifies，看看在命令的处理过程中是否有通知到达。

28.8. 与COPY命令相关的函数

PostgreSQL 的 COPY 命令提供了选项，可以通过 libpq 使用的网络连接读取或写入数据。本节介绍的函数允许应用程序通过提供或接收复制数据来使用这一能力。

整体流程如下：应用程序先通过 PQexec 或等效函数发出 SQL COPY 命令。如果命令没有错误，响应就是一个 PGresult 对象，其状态码为 PGRES_COPY_OUT 或 PGRES_COPY_IN，取决于指定的复制方向。应用程序随后应使用本节函数接收或发送数据行。数据传输完成后，会返回另一个 PGresult 对象，表示传输成功或失败：成功时状态为 PGRES_COMMAND_OK，出现问题时为 PGRES_FATAL_ERROR。此时可以通过 PQexec 继续发出 SQL 命令。（COPY 操作进行期间，不能在同一连接上执行其他 SQL 命令。）

如果一个COPY命令是通过PQexec在一个可能包含额外命令的字符串中发出的，那么应用在完成COPY序列之后必须继续用PQgetResult取得结果。只有在PQgetResult返回NULL时，我们才能确信PQexec的命令字符串已经处理完毕，并且可以安全地发出更多命令。

只有从 PQexec 或 PQgetResult 获得 PGRES_COPY_OUT 或 PGRES_COPY_IN 结果状态后，才应调用本节函数。

带有上述某个状态值的 PGresult 对象，还会携带关于即将开始的 COPY 操作的附加数据。这些数据可以通过下列函数获取，这些函数也用于查询结果：

PQnfields

返回要复制的列（字段）的数量。

`PQbinaryTuples`

0表示整体复制格式为文本（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式为二进制。有关更多信息，请参见COPY。

`PQfformat`

返回与复制操作的每列关联的格式代码（0表示文本，1表示二进制）。
当整体复制格式为文本时，每列的格式代码始终为零，
但二进制格式可以支持文本和二进制列。（但是，在当前的 `COPY` 实现中，只有二进制列出现在二进制复制中；因此，每列格式目前始终与整体格式匹配。）

> **注意**
>
> 这些附加数据值仅在使用协议 3.0 时可用。使用协议 2.0 时，这些函数都返回 0。

28.8.1. 发送COPY数据的函数

这些函数用于在 `COPY FROM STDIN` 期间发送数据。如果连接不处于 `COPY_IN` 状态，调用它们会失败。

`PQputCopyData`

在`COPY_IN`状态中向服务器发送数据。

```
int PQputCopyData(PGconn *conn,
                  const char *buffer,
                  int nbytes);
```

将指定 *buffer* 中长度为 *nbytes* 的 `COPY` 数据传输到服务器。数据发送成功时返回 1；因尝试会阻塞而未发送时返回零（仅可能发生在连接处于非阻塞模式时）；发生错误时返回 -1。（返回 -1 时，可用 `PQerrorMessage` 获取详细信息。返回零时，应等待可写就绪后重试。）

应用程序可以将 `COPY` 数据流分成任意方便大小的数据块，逐块装入缓冲区。发送时，这些数据块的边界没有语义含义。数据流内容必须符合 `COPY` 命令预期的数据格式；详见COPY。

`PQputCopyEnd`

在`COPY_IN`状态中向服务器发送数据结束的指示。

```
int PQputCopyEnd(PGconn *conn,
                 const char *errormsg);
```

如果 *errormsg* 为 `NULL`，则成功结束 `COPY_IN` 操作。如果 *errormsg* 不为 `NULL`，则强制 `COPY` 失败，并将 *errormsg* 指向的字符串用作错误消息。（但不应假定服务器一定会返回这条完全相同的错误消息，因为服务器可能已经因自身原因使 `COPY` 失败。还要注意，在使用 3.0 之前协议的连接上，强制失败选项不起作用。）

终止数据发送成功时返回 1；因尝试会阻塞而未发送时返回零（仅可能发生在连接处于非阻塞模式时）；发生错误时返回 -1。（返回 -1 时，可用 `PQerrorMessage` 获取详细信息。返回零时，应等待可写就绪后重试。）

成功调用 `PQputCopyEnd` 后，调用 `PQgetResult` 获取 `COPY` 命令的最终结果状态。可以按通常方式等待该结果就绪，然后恢复正常操作。

28.8.2. 接收COPY数据的函数

这些函数用于在COPY TO STDOUT的过程中接收数据。如果连接不在COPY_OUT状态，那么调用它们将会失败。

PQgetCopyData

在COPY_OUT状态下从服务器接收数据。

```
int PQgetCopyData(PGconn *conn,
                  char **buffer,
                  int async);
```

在 COPY 期间尝试从服务器获取下一行数据。每次总是返回一个完整数据行；如果只有部分行可用，则不返回。成功返回数据行时，会分配一块内存保存数据。*buffer* 参数必须为非 NULL。**buffer* 会被设置为指向所分配的内存；如果没有返回缓冲区，则设为 NULL。非 NULL 的结果缓冲区在不再需要时应使用 PQfreemem 释放。

成功返回一行时，返回值是该行的数据字节数，始终大于零。
返回的字符串总是以零字节结尾，不过这可能仅对文本 COPY 有用。返回零表示 COPY 仍在进行，但尚无可用行（仅在 *async* 为真时可能发生）。返回 -1 表示 COPY 已完成；返回 -2 表示发生了错误（可用 PQerrorMessage 查看原因）。

当 *async* 为真（非零）时，PQgetCopyData 不会阻塞等待输入；如果 COPY 仍在进行，但没有完整行可用，则返回零。（此时，应等待读就绪，先调用 PQconsumeInput，再调用 PQgetCopyData。）当 *async* 为假（零）时，PQgetCopyData 会阻塞，直到数据可用或操作完成。

在 PQgetCopyData 返回 -1 后，调用 PQgetResult 获取 COPY 命令的最终结果状态。可以按通常方式等待该结果就绪，然后恢复正常操作。

28.8.3. COPY的已废弃函数

这些函数使用较旧的方式处理 COPY。虽然仍然可用，但由于错误处理欠佳、检测数据结束的方式不便，而且缺少对二进制或非阻塞传输的支持，已被弃用。

PGgetline

将服务器传来的、以换行符结尾的一行字符读入大小为 *length* 的字符串缓冲区。

```
int PQgetline(PGconn *conn,
              char *buffer,
              int length);
```

此函数最多将 *length*-1 个字符复制到缓冲区，并将末尾的换行符转换为零字节。PQgetline 在输入结束时返回 EOF，读完一整行时返回 0，缓冲区已满但尚未读到末尾换行符时返回 1。

注意，应用程序必须检查新读入的一行是否仅由 \. 两个字符组成，这表示服务器已发送完 COPY 命令的结果。如果可能收到长度超过 *length*-1 个字符的行，必须确保正确识别 \. 行，例如不能把长数据行的末尾误当作终止行。

PQgetlineAsync

以非阻塞方式将服务器传来的一行 COPY 数据读入缓冲区。

```
int PQgetlineAsync(PGconn *conn,
                   char *buffer,
                   int bufsize);
```

此函数类似于 PQgetline，但可用于必须异步读取 COPY 数据的应用程序，即读取时不阻塞。发出 COPY 命令并收到 PGRES_COPY_OUT 响应后，应用程序应调用 PQconsumeInput 和 PQgetlineAsync，直到检测到数据结束信号。

与 PQgetline 不同，此函数会负责检测数据结束。

每次调用时，如果 libpq 的输入缓冲区中有完整数据行，PQgetlineAsync 就会返回数据；否则，要等该行剩余部分到达后才返回数据。识别到复制数据结束标记时返回 -1，没有可用数据时返回 0，否则返回正数，表示返回的数据字节数。返回 -1 后，调用者必须接着调用 PQendcopy，然后恢复正常处理。

返回的数据不会跨越数据行边界。只要可能，每次就返回一整行；
但如果调用者提供的缓冲区太小，容不下服务器发送的一行，则只返回部分行。
对于文本数据，可检查最后返回的字节是否为 \n，以判断是否返回了完整行。（对于二进制 COPY，则必须实际解析 COPY 数据格式才能作出相同判断。）返回的字符串不以零字节结尾。（如果要自行添加末尾的零字节，务必将传入的 *bufsize* 设置为比实际可用空间少一字节。）

PQputline

向服务器发送以零字节结尾的字符串。成功时返回 0，无法发送字符串时返回 EOF。

```
int PQputline(PGconn *conn,
              const char *string);
```

连续调用 PQputline 发送的 COPY 数据流，与 PQgetlineAsync 返回的数据格式相同。不过，应用程序不必在每次 PQputline 调用中恰好发送一个数据行；每次发送部分行或多行也可以。

> **注意**
>
> 在 PostgreSQL 协议 3.0 之前，应用程序必须显式发送由 \. 两个字符组成的最后一行，告知服务器应用程序已发送完 COPY 数据。虽然这种方式仍然有效，但已被弃用，\. 的特殊含义预计会在未来版本中移除。发送完实际数据后，调用 PQendcopy 即可。

PQputnbytes

向服务器发送不以零字节结尾的字符串。成功时返回 0，无法发送字符串时返回 EOF。

```
int PQputnbytes(PGconn *conn,
                const char *buffer,
                int nbytes);
```

此函数与 PQputline 完全相同，只是直接指定了要发送的字节数，因此数据缓冲区不必以零字节结尾。发送二进制数据时可使用此函数。

PQendcopy

与服务器同步。

```
int PQendcopy(PGconn *conn);
```

此函数会等待服务器完成复制。应在使用 PQputline 向服务器发送最后一个字符串后，或使用 PQgetline 从服务器接收最后一个字符串后调用它。必须调用此函数，否则服务器与客户端会“失去同步”。函数返回后，服务器便准备好接收下一条 SQL 命令。成功完成时返回 0，否则返回非零值。（返回非零值时，可用 PQerrorMessage 获取详细信息。）

使用 PQgetResult 时，收到 PGRES_COPY_OUT 结果后，应用程序应反复调用 PQgetline，并在看到终止行后调用 PQendcopy。随后应回到 PQgetResult 循环，直到 PQgetResult 返回空指针。类似地，收到 PGRES_COPY_IN 结果后，应连续调用 PQputline，再调用 PQendcopy，然后回到 PQgetResult 循环。这样可以保证嵌在一系列 SQL 命令中的 COPY 命令正确执行。

旧的应用很可能会通过PQexec提交一个COPY命令并且假定事务在PQendcopy之后完成。只有在COPY是命令字符串中唯一的SQL命令时才能正确工作。

28.9. 控制函数

这些函数控制 libpq 行为的各种细节。

PQsetErrorVerbosity

设置以下函数所返回消息的详细程度：PQerrorMessage和PQresultErrorMessage。

```
typedef enum
{
    PQERRORS_TERSE,
    PQERRORS_DEFAULT,
    PQERRORS_VERBOSE
} PGVerbosity;

PGVerbosity PQsetErrorVerbosity(PGconn *conn, PGVerbosity
 verbosity);
```

PQsetErrorVerbosity设置详细程度模式，并返回该连接先前的设置。在TERSE模式下，返回的消息只包括严重性、主要文本和位置，通常一行就能容纳。
默认模式生成的消息除上述字段外，还包括详细信息、
提示或上下文字段（这些内容可能跨越多行）。VERBOSE模式包括所有可用字段。
改变详细程度不会影响已有PGresult对象中的消息，只影响随后创建的对象。

PQtrace

启用客户端/服务器通信跟踪，将跟踪信息写入调试文件流。

```
void PQtrace(PGconn *conn, FILE *stream);
```

> **注意**
>
> 在 Windows 上，如果 libpq 库与应用程序使用不同的标志编译，调用此函数会导致应用程序崩溃，因为 FILE 指针的内部表示不同。具体而言，库与所有使用它的应用程序都应使用相同的多线程/单线程、发布/调试和静态/动态标志。

PQuntrace

禁用由 PQtrace 启动的跟踪。

```
void PQuntrace(PGconn *conn);
```

28.10. 通知处理

服务器生成的通知和警告消息不由查询执行函数返回，因为它们并不表示查询失败。
这些消息会传递给通知处理函数，处理函数返回后，执行正常继续。
默认通知处理函数将消息打印到 stderr，应用程序也可以提供自己的处理函数来替换这一行为。

由于历史原因，通知处理有两个级别，称为通知接收器和通知处理器。
通知接收器的默认行为是格式化通知并且将一个字符串传递给通知处理器来打印。不过，
如果一个应用选择提供自己的通知接收器，
它通常会忽略通知处理器层并且在通知接收器中完成所有工作。

函数PQsetNoticeReceiver 为一个连接对象设置或者检查当前的通知接收器。相似地，PQsetNoticeProcessor 设置或检查当前的通知处理器。

```
typedef void (*PQnoticeReceiver) (void *arg, const PGresult *res);

PQnoticeReceiver
PQsetNoticeReceiver(PGconn *conn,
                    PQnoticeReceiver proc,
                    void *arg);

typedef void (*PQnoticeProcessor) (void *arg, const char *message);

PQnoticeProcessor
PQsetNoticeProcessor(PGconn *conn,
                     PQnoticeProcessor proc,
                     void *arg);
```

这些函数中的每一个会返回之前的通知接收器或处理器函数指针，并且设置新值。
如果你提供了一个空函数指针，将不会采取任何动作，只会返回当前指针。

从服务器收到通知或警告消息，或者 libpq 内部生成这类消息时，便会调用通知接收器函数。消息以状态为 PGRES_NONFATAL_ERROR 的 PGresult 对象传给该函数。（接收器可使用 PQresultErrorField 提取单个字段，或使用 PQresultErrorMessage 获取完整的预格式化消息。）传给 PQsetNoticeReceiver 的同一个 void 指针也会传入。（必要时，可通过该指针访问应用程序特有的状态。）

默认的通知接收器会简单地抽取消息（使用PQresultErrorMessage）
并且将它传递给通知处理器。

通知处理器负责处理以文本形式提供的通知或警告消息。
传给它的参数包括消息字符串（含末尾的换行符），以及先前传给 PQsetNoticeProcessor 的同一个 void 指针。（必要时，可通过该指针访问应用程序特有的状态。）

默认的通知处理器很简单：

```
static void
defaultNoticeProcessor(void *arg, const char *message)
{
    fprintf(stderr, "%s", message);
}
```

设置通知接收器或处理器后，应当预期：只要 PGconn 对象或由它生成的 PGresult 对象仍存在，就可能调用该函数。创建 PGresult 时，会将 PGconn 当前的通知处理指针复制到 PGresult 中，供 PQgetvalue 等函数在需要时使用。

28.11. 环境变量

如果调用代码没有直接指定值，下列环境变量可用来选择默认的连接参数值，它们将被 PQconnectdb、PQsetdbLogin和 PQsetdb使用。例如，这些变量有助于避免在简单的客户端应用中硬编码数据库连接信息。

- PGHOST设置数据库服务器名。如果它以斜线开头，则指定使用 Unix 域通信而不是 TCP/IP 通信；该值是存放套接字文件的目录名（在默认安装配置中它是/tmp）。
- PGHOSTADDR指定数据库服务器的数字 IP 地址。这可以替代或者与PGHOST一起设置，以避免 DNS 查找开销。关于这两个参数的相互作用，见上文PQconnectdb下对这些参数的文档。

  当PGHOST和PGHOSTADDR都未设置时，默认行为是使用本地 Unix 域套接字连接；在没有 Unix 域套接字的机器上，libpq将尝试连接到localhost。
- PGPORT设置用于与PostgreSQL服务器通信的 TCP 端口号或 Unix 域套接字文件扩展名。
- PGDATABASE设置PostgreSQL数据库名。
- PGUSER设置用于连接数据库的用户名。
- PGPASSWORD设置当服务器要求密码认证时所使用的密码。出于安全原因，不建议使用这个环境变量（某些操作系统允许非 root 用户通过ps查看进程环境变量）；可以考虑改用~/.pgpass文件（见第 28.12 节）。
- PGPASSFILE指定用于查找的密码文件名。如果未设置，默认为~/.pgpass（见第 28.12 节）。
- PGSERVICE设置要在pg_service.conf中查找的服务名。这提供了一种设置所有参数的简写方式。
- PGREALM设置与PostgreSQL一起使用的 Kerberos realm（如果它与本地 realm 不同）。如果设置了PGREALM，libpq应用将尝试与该 realm 的服务器进行认证，并使用单独的票据文件以避免与本地票据文件冲突。只有当服务器选择了 Kerberos 认证时才会使用这个环境变量。
- PGOPTIONS设置PostgreSQL服务器的额外运行时选项。
- PGSSLMODE决定是否或以何种优先级与服务器协商一个安全的SSL连接。有四种模式：disable将只尝试一次非SSL连接；allow将协商，首先尝试一次非SSL连接，失败后再尝试一次SSL连接；prefer（默认）将协商，首先尝试一次SSL连接，失败后再尝试一次常规的非SSL连接；require将只尝试一次SSL连接。如果PostgreSQL编译时未启用 SSL 支持，使用选项require会导致错误，而选项allow和prefer将被接受，但libpq实际上不会尝试建立SSL连接。
- PGREQUIRESSL设置连接是否必须通过SSL建立。如果设置为“1”，当服务器不接受SSL连接时，libpq将拒绝连接（等效于sslmode为prefer）。此选项已弃用，请改用sslmode设置，并且它只有在PostgreSQL编译时启用了 SSL 支持的情况下才可用。
- PGKRBSRVNAME设置使用 Kerberos 5 或 GSSAPI 认证时所用的 Kerberos 服务名。
- PGCONNECT_TIMEOUT设置libpq尝试连接PostgreSQL服务器时等待的最大秒数。如果未设置或设置为零，libpq将无限期等待。不建议把超时设置为小于 2 秒。

下列环境变量可用来为每个PostgreSQL会话指定默认行为。（另见ALTER USER和ALTER DATABASE命令了解按用户或按数据库设置默认行为的方式。）

- PGDATESTYLE设置日期/时间表示的默认风格（等同于SET datestyle TO ...）。
- PGTZ设置默认的时区（等同于SET timezone TO ...）。
- PGCLIENTENCODING设置默认的客户端字符集编码（等同于SET client_encoding TO ...）。
- PGGEQO为遗传查询优化器设置默认模式（等同于SET geqo TO ...）。

关于这些环境变量的正确取值的信息，请参考SQL命令SET。

下列环境变量决定libpq的内部行为；它们会覆盖编译在程序中的默认值。

- PGSYSCONFDIR设置包含pg_service.conf文件的目录。
- PGLOCALEDIR设置包含用于消息国际化的locale文件的目录。

28.12. 密码文件

用户主目录中的.pgpass文件或PGPASSFILE引用的文件可保存密码，
供连接需要密码且尚未通过其他方式指定密码时使用。在 Microsoft Windows 上，文件名为 %APPDATA%\postgresql\pgpass.conf（其中%APPDATA%
指用户配置文件中的应用数据子目录）。

这个文件应该包含以下格式的行：

hostname:*port*:*database*:*username*:*password*

前四个字段中的每一个都可以是一个字面值或者*（匹配任意值）。
将与当前连接参数匹配的第一行中的密码字段用作密码。（因此，在使用通配符时，
应把更特定的条目放在前面。）如果某个条目需要包含:或\，用\转义该字符。
主机名localhost同时匹配来自本机的host（TCP）和local（Unix 域套接字）连接。

在 Unix 系统上，.pgpass文件上的权限必须不允许所有人或组内访问，可以用chmod 0600 ~/.pgpass这样的命令实现。如果权限没有这么严格，该文件将被忽略。在微软 Windows 上，该文件被假定存储在一个安全的目录中，因此不会进行特别的权限检查。

28.13. 连接服务文件

连接服务文件允许 libpq 连接参数与一个单一服务名称关联。之后在一个 libpq 连接中就可以指定那个服务名称，与其相关的设置将被使用。这允许在不重新编译 libpq 应用的前提下修改连接参数。服务名称也可以使用PGSERVICE环境变量来指定。

要使用这一功能，把share/pg_service.conf.sample复制为 etc/pg_service.conf并编辑该文件，加入服务名和参数。该文件也可用于仅客户端安装。文件的位置还可以通过PGSYSCONFDIR 环境变量指定。

28.14. SSL 支持

PostgreSQL原生支持使用SSL连接来加密客户端/服务器通信以提高安全性。
关于服务器端SSL功能的细节，参见第 16.7 节。

如果服务器要求一个客户端证书，libpq将发送存储在用户主目录文件~/.postgresql/postgresql.crt中的证书。还必须存在一个配套的私钥文件~/.postgresql/postgresql.key，并且它不能允许所有用户读取。（在 Microsoft Windows 上这些文件名为%APPDATA%\postgresql\postgresql.crt和%APPDATA%\postgresql\postgresql.key。）

如果用户主目录中存在文件~/.postgresql/root.crt，libpq将使用其中存储的证书列表来验证服务器的证书。（在 Microsoft Windows 上该文件名为%APPDATA%\postgresql\root.crt。）如果服务器不提供证书，SSL 连接将失败；因此要使用这个特性，服务器还必须有root.crt文件。

如果你在自己的应用程序内部（除了在libpq内部之外）也使用了SSL，可以用PQinitSSL(int)告诉libpq，SSL库已经由你的应用程序初始化过了。

28.15. 在线程化程序中的行为

如果在构建 PostgreSQL 发行版时使用了 configure 命令行选项 --enable-thread-safety，则 libpq 是可重入且线程安全的。此外，编译应用程序代码时，可能还需要使用额外的编译器命令行选项。有关如何构建支持线程的应用程序，请参阅系统文档，或查看 src/Makefile.global 中的 PTHREAD_CFLAGS 和 PTHREAD_LIBS。

一项限制是：两个线程不能同时操作同一个 PGconn 对象。尤其不能通过同一个连接对象从不同线程并发发出命令。（如果需要并发执行命令，请使用多个连接。）

PGresult对象在创建后是只读的，因此可以在线程之间自由传递。

已弃用的函数 PQrequestCancel 和 PQoidStatus 不是线程安全的，不应在多线程程序中使用。PQrequestCancel 可以用 PQcancel 替代。PQoidStatus 可以用 PQoidValue 替代。

如果你在应用程序中使用 Kerberos（除了libpq内部之外），则需要在 Kerberos 调用周围加锁，因为 Kerberos 函数不是线程安全的。可参考libpq源代码中的PQregisterThreadLock函数，它提供了一种在libpq与应用程序之间协作加锁的方法。

如果你在多线程应用程序中遇到问题，请运行 src/tools/thread 目录中的程序，检查你的平台是否存在线程不安全的函数。configure 会运行这个程序，但对于二进制发行版，你自己的库可能与编译这些二进制文件时所用的库不一致。

28.16. 构建libpq程序

要构建（即编译和链接）一个使用 libpq 的程序，需要完成以下所有步骤：

- 包含 libpq-fe.h 头文件：

  ```
  #include <libpq-fe.h>
  ```

  如果没有这样做，编译器通常会给出类似以下内容的错误消息：

  ```
  foo.c: In function `main':
  foo.c:34: `PGconn' undeclared (first use in this function)
  foo.c:35: `PGresult' undeclared (first use in this function)
  foo.c:54: `CONNECTION_BAD' undeclared (first use in this function)
  foo.c:68: `PGRES_COMMAND_OK' undeclared (first use in this function)
  foo.c:95: `PGRES_TUPLES_OK' undeclared (first use in this function)
  ```

- 向编译器提供 `-I*directory*` 选项，指定 PostgreSQL 头文件的安装目录。（某些情况下，编译器默认会搜索该目录，此时可以省略这个选项。）例如，编译命令行可以是：

  ```
  cc -c -I/usr/local/pgsql/include testprog.c
  ```

  如果使用 makefile，请将该选项添加到 `CPPFLAGS` 变量中：

  ```
  CPPFLAGS += -I/usr/local/pgsql/include
  ```

  如果你的程序可能由其他用户编译，那么你不应该像那样硬编码目录位置。你可以运行工具`pg_config`在本地系统上找出头文件在哪里：

  ```
  $ pg_config --includedir
  /usr/local/include
  ```

  未向编译器指定正确的选项将导致类似以下内容的错误消息：

  ```
  testlibpq.c:8:22: libpq-fe.h: No such file or directory
  ```

- 链接最终程序时，指定 `-lpq` 选项以链接 libpq 库，同时指定 `-L*directory*` 选项，告知编译器 libpq 库所在的目录。（同样，编译器默认会搜索某些目录。）为获得最大的可移植性，请将 `-L` 选项放在 `-lpq` 选项之前。例如：

  ```
  cc -o testprog testprog1.o testprog2.o -L/usr/local/pgsql/lib -lpq
  ```

  你也可以使用`pg_config`找出库目录：

  ```
  $ pg_config --libdir
  /usr/local/pgsql/lib
  ```

  指出这一部分问题的错误消息可能看起来像：

  ```
  testlibpq.o: In function `main':
  testlibpq.o(.text+0x60): undefined reference to `PQsetdbLogin'
  testlibpq.o(.text+0x71): undefined reference to `PQstatus'
  testlibpq.o(.text+0xa4): undefined reference to `PQerrorMessage'
  ```

  这意味着你忘记了 `-lpq`。

  ```
  /usr/bin/ld: cannot find -lpq
  ```

  这意味着你忘记了`-L`选项或者没有指定正确的目录。

如果你的代码引用了头文件`libpq-int.h`，而你又不愿修改代码以避免使用它，那么从 PostgreSQL 7.2 开始，这个文件将位于*includedir*`/postgresql/internal/libpq-int.h`，因此你需要在编译器命令行中加入适当的`-I`选项。

28.17. 示例程序

这些示例和其他示例可在源代码发行包的`src/test/examples`目录中找到。

例 28.1. libpq示例程序 1

```
/*
 * testlibpq.c
 *
```

```
 *      Test the C version of libpq, the PostgreSQL frontend library.
 */
#include &lt;stdio.h&gt;
#include &lt;stdlib.h&gt;
##KEEP##

static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}

int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    int         nFields;
    int         i,
                j;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * Our test case here involves using a cursor, for which we must
 be inside
     * a transaction block.  We could do the whole thing with a single
     * PQexec() of "select * from pg_database", but that's too trivial
 to make
```

```
     * a good example.
     */

    /* Start a transaction block */
    res = PQexec(conn, "BEGIN");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "BEGIN command failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /*
     * Should PQclear PGresult whenever it is no longer needed to
 avoid memory
     * leaks
     */
    PQclear(res);

    /*
     * Fetch rows from pg_database, the system catalog of databases
     */
    res = PQexec(conn, "DECLARE myportal CURSOR FOR select * from pg_
database");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "DECLARE CURSOR failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }
    PQclear(res);

    res = PQexec(conn, "FETCH ALL in myportal");
    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
        fprintf(stderr, "FETCH ALL failed: %s", PQerrorMessage(conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /* first, print out the attribute names */
    nFields = PQnfields(res);
    for (i = 0; i < nFields; i++)
        printf("%-15s", PQfname(res, i));
    printf("\n\n");

    /* next, print out the rows */
    for (i = 0; i < PQntuples(res); i++)
    {
        for (j = 0; j < nFields; j++)
            printf("%-15s", PQgetvalue(res, i, j));
        printf("\n");
```

```
    }

    PQclear(res);

    /* close the portal ... we don't bother to check for errors ... */
    res = PQexec(conn, "CLOSE myportal");
    PQclear(res);

    /* end the transaction */
    res = PQexec(conn, "END");
    PQclear(res);

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

例 28.2. libpq示例程序 2

```
/*
 * testlibpq2.c
 *      Test of the asynchronous notification interface
 *
 * Start this program, then from psql in another window do
 *   NOTIFY TBL2;
 * Repeat four times to get this program to exit.
 *
 * Or, if you want to get fancy, try this:
 * populate a database with the following commands
 * (provided in src/test/examples/testlibpq2.sql):
 *
 *   CREATE TABLE TBL1 (i int4);
 *
 *   CREATE TABLE TBL2 (i int4);
 *
 *   CREATE RULE r1 AS ON INSERT TO TBL1 DO
 *     (INSERT INTO TBL2 VALUES (new.i); NOTIFY TBL2);
 *
 * and do this four times:
 *
 *   INSERT INTO TBL1 VALUES (10);
 */
#include &lt;stdio.h&gt;
#include &lt;stdlib.h&gt;
#include &lt;string.h&gt;
#include &lt;errno.h&gt;
#include &lt;sys/time.h&gt;
##KEEP##

static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
```

```
    exit(1);
}

int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    PGnotify   *notify;
    int         nnotifies;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * Issue LISTEN command to enable notifications from the rule's
 NOTIFY.
     */
    res = PQexec(conn, "LISTEN TBL2");
    if (PQresultStatus(res) != PGRES_COMMAND_OK)
    {
        fprintf(stderr, "LISTEN command failed: %s", PQerrorMessage(
conn));
        PQclear(res);
        exit_nicely(conn);
    }

    /*
     * should PQclear PGresult whenever it is no longer needed to
 avoid memory
     * leaks
```

```
     */
    PQclear(res);

    /* Quit after four notifies are received. */
    nnotifies = 0;
    while (nnotifies < 4)
    {
        /*
         * Sleep until something happens on the connection.  We use
 select(2)
         * to wait for input, but you could also use poll() or similar
         * facilities.
         */
        int         sock;
        fd_set      input_mask;

        sock = PQsocket(conn);

        if (sock < 0)
            break;              /* shouldn't happen */

        FD_ZERO(&input_mask);
        FD_SET(sock, &input_mask);

        if (select(sock + 1, &input_mask, NULL, NULL, NULL) < 0)
        {
            fprintf(stderr, "select() failed: %s\n", strerror(errno));
            exit_nicely(conn);
        }

        /* Now check for input */
        PQconsumeInput(conn);
        while ((notify = PQnotifies(conn)) != NULL)
        {
            fprintf(stderr,
                    "ASYNC NOTIFY of '%s' received from backend pid %d
\n",
                    notify->relname, notify->be_pid);
            PQfreemem(notify);
            nnotifies++;
        }
    }

    fprintf(stderr, "Done.\n");

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

例 28.3. libpq Example Program 3

```
/*
```

```
 * testlibpq3.c
 *      Test out-of-line parameters and binary I/O.
 *
 * Before running this, populate a database with the following
 commands
 * (provided in src/test/examples/testlibpq3.sql):
 *
 * CREATE TABLE test1 (i int4, t text, b bytea);
 *
 * INSERT INTO test1 values (1, 'joe''s place', '\\000\\001\\002\\003\
\004');
 * INSERT INTO test1 values (2, 'ho there', '\\004\\003\\002\\001\
\000');
 *
 * The expected output is:
 *
 * tuple 0: got
 *  i = (4 bytes) 1
 *  t = (11 bytes) 'joe's place'
 *  b = (5 bytes) \000\001\002\003\004
 *
 * tuple 0: got
 *  i = (4 bytes) 2
 *  t = (8 bytes) 'ho there'
 *  b = (5 bytes) \004\003\002\001\000
 */
#include &lt;stdio.h&gt;
#include &lt;stdlib.h&gt;
#include &lt;string.h&gt;
#include &lt;sys/types.h&gt;
##KEEP##

/* for ntohl/htonl */
#include &lt;netinet/in.h&gt;
#include &lt;arpa/inet.h&gt;


static void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}

/*
 * This function prints a query result that is a binary-format fetch
 from
 * a table defined as in the comment above.  We split it out because
 the
 * main() function uses it twice.
 */
static void
show_binary_results(PGresult *res)
{
```

```
   int         i,
               j;
   int         i_fnum,
               t_fnum,
               b_fnum;

   /* Use PQfnumber to avoid assumptions about field order in result
*/
   i_fnum = PQfnumber(res, "i");
   t_fnum = PQfnumber(res, "t");
   b_fnum = PQfnumber(res, "b");

   for (i = 0; i < PQntuples(res); i++)
   {
       char       *iptr;
       char       *tptr;
       char       *bptr;
       int         blen;
       int         ival;

       /* Get the field values (we ignore possibility they are null!)
*/
       iptr = PQgetvalue(res, i, i_fnum);
       tptr = PQgetvalue(res, i, t_fnum);
       bptr = PQgetvalue(res, i, b_fnum);

       /*
        * The binary representation of INT4 is in network byte order,
which
        * we'd better coerce to the local byte order.
        */
       ival = ntohl(*((uint32_t *) iptr));

       /*
        * The binary representation of TEXT is, well, text, and since
libpq
        * was nice enough to append a zero byte to it, it'll work
just fine
        * as a C string.
        *
        * The binary representation of BYTEA is a bunch of bytes,
which could
        * include embedded nulls so we have to pay attention to field
length.
        */
       blen = PQgetlength(res, i, b_fnum);

       printf("tuple %d: got\n", i);
       printf(" i = (%d bytes) %d\n",
              PQgetlength(res, i, i_fnum), ival);
       printf(" t = (%d bytes) '%s'\n",
              PQgetlength(res, i, t_fnum), tptr);
       printf(" b = (%d bytes) ", blen);
       for (j = 0; j < blen; j++)
```

```
            printf("\\%03o", bptr[j]);
        printf("\n\n");
    }
}

int
main(int argc, char **argv)
{
    const char *conninfo;
    PGconn     *conn;
    PGresult   *res;
    const char *paramValues[1];
    int         paramLengths[1];
    int         paramFormats[1];
    uint32_t    binaryIntVal;

    /*
     * If the user supplies a parameter on the command line, use it as
 the
     * conninfo string; otherwise default to setting dbname=postgres
 and using
     * environment variables or defaults for all other connection
 parameters.
     */
    if (argc > 1)
        conninfo = argv[1];
    else
        conninfo = "dbname = postgres";

    /* Make a connection to the database */
    conn = PQconnectdb(conninfo);

    /* Check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) != CONNECTION_OK)
    {
        fprintf(stderr, "Connection to database failed: %s",
                PQerrorMessage(conn));
        exit_nicely(conn);
    }

    /*
     * The point of this program is to illustrate use of PQexecParams(
) with
     * out-of-line parameters, as well as binary transmission of data.
     *
     * This first example transmits the parameters as text, but
 receives the
     * results in binary format.  By using out-of-line parameters we
 can
     * avoid a lot of tedious mucking about with quoting and escaping,
 even
     * though the data is text.  Notice how we don't have to do
 anything
```

```
     * special with the quote mark in the parameter value.
     */

    /* Here is our out-of-line parameter value */
    paramValues[0] = "joe's place";

    res = PQexecParams(conn,
                       "SELECT * FROM test1 WHERE t = $1",
                       1,       /* one param */
                       NULL,    /* let the backend deduce param type
*/
                       paramValues,
                       NULL,    /* don't need param lengths since text
*/
                       NULL,    /* default to all text params */
                       1);      /* ask for binary results */

    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
        fprintf(stderr, "SELECT failed: %s", PQerrorMessage(conn));
        PQclear(res);
        exit_nicely(conn);
    }

    show_binary_results(res);

    PQclear(res);

    /*
     * In this second example we transmit an integer parameter in
binary
     * form, and again retrieve the results in binary form.
     *
     * Although we tell PQexecParams we are letting the backend deduce
     * parameter type, we really force the decision by casting the
parameter
     * symbol in the query text.  This is a good safety measure when
sending
     * binary parameters.
     */

    /* Convert integer value "2" to network byte order */
    binaryIntVal = htonl((uint32_t) 2);

    /* Set up parameter arrays for PQexecParams */
    paramValues[0] = (char *) &binaryIntVal;
    paramLengths[0] = sizeof(binaryIntVal);
    paramFormats[0] = 1;        /* binary */

    res = PQexecParams(conn,
                       "SELECT * FROM test1 WHERE i = $1::int4",
                       1,       /* one param */
                       NULL,    /* let the backend deduce param type
*/
```

```
                        paramValues,
                        paramLengths,
                        paramFormats,
                        1);      /* ask for binary results */

    if (PQresultStatus(res) != PGRES_TUPLES_OK)
    {
        fprintf(stderr, "SELECT failed: %s", PQerrorMessage(conn));
        PQclear(res);
        exit_nicely(conn);
    }

    show_binary_results(res);

    PQclear(res);

    /* close the connection to the database and cleanup */
    PQfinish(conn);

    return 0;
}
```

第 29 章 大对象

PostgreSQL提供一种大对象机制，允许以流式方式访问存储在专用大对象结构中的用户数据。在处理大到无法方便地整体操作的数据值时，这种流式访问非常有用。

本章介绍PostgreSQL大对象数据的实现，以及相应的编程接口和查询语言接口。本章示例使用 libpq C 库，但大多数PostgreSQL原生编程接口都支持等效的功能。其他接口也可能在内部使用大对象接口来为大值提供通用支持，这里不作说明。

29.1. 历史

PostgreSQL 的间接前身 POSTGRES 4.2 支持三种标准的大对象实现方式：作为 POSTGRES 服务器外部的文件、作为由 POSTGRES 服务器管理的外部文件，以及作为存储在 POSTGRES 数据库内部的数据。这在用户中引起了相当大的混乱。因此，PostgreSQL 只保留了把大对象作为数据库内部存储数据的方式。尽管这样访问速度较慢，但它提供了更严格的数据完整性。由于历史原因，这种存储方案被称为 Inversion 大对象（Inversion large objects）。（你会偶尔看到 Inversion 一词被用来表示与大对象相同的含义。）从 PostgreSQL 7.1 开始，所有大对象都放在一个名为 `pg_largeobject` 的系统表中。

PostgreSQL 7.1 引入了一种（绰号为“TOAST”的）机制，允许数据值远大于单个页面。这使得大对象机制在一定程度上已经过时。大对象机制仍然保留的一个优势是，它允许的值大小可达 2 GB，而经过 TOAST 处理的字段最大只能到 1 GB。此外，大对象可以比普通数据字段更容易地逐段进行操作，因此实际的限制差别很大。

29.2. 实现特性

大对象实现会把大对象拆分成若干“块”，并将这些块存储在数据库行中。一个 B-树索引保证在执行随机访问读写时，能够快速找到正确的块号。

29.3. 客户端接口

本节描述 PostgreSQL 客户端接口库为访问大对象所提供的功能。使用这些函数对大对象进行的所有操作必须发生在一个 SQL 事务块内。（从 PostgreSQL 6.5 开始，这一要求被强制执行；不过在更早的版本中它一直是一项隐含要求，忽略它会导致错误行为。）PostgreSQL 的大对象接口是仿照 Unix 文件系统接口设计的，提供了与 `open`、`read`、`write`、`lseek` 等相对应的操作。

在 libpq 中使用大对象接口的客户端应用应包含头文件 `libpq/libpq-fs.h` 并与 libpq 库链接。

29.3.1. 创建一个大对象

The function

```
Oid lo_creat(PGconn *conn, int mode);
```

创建一个新的大对象。返回值是分配给新大对象的 OID，失败时为 InvalidOid（零）。自 PostgreSQL 8.1 起，*mode* 未被使用且被忽略；不过，为了与更早版本的向后兼容，最好将其设置为 `INV_READ`、`INV_WRITE` 或 `INV_READ | INV_WRITE`。（这些符号常量定义在头文件 `libpq/libpq-fs.h` 中。）

例如：

```
inv_oid = lo_creat(conn, INV_READ|INV_WRITE);
```

The function

```
Oid lo_create(PGconn *conn, Oid lobjId);
```

也会创建一个新的大对象。要分配的 OID 可以由 *lobjId* 指定；如果指定了它，而该 OID 已经被某个大对象使用，则会失败。如果 *lobjId* 是 InvalidOid（零），则 `lo_create` 会分配一个未使用的 OID（其行为与 `lo_creat` 相同）。返回值是分配给新大对象的 OID，失败时为 InvalidOid（零）。

`lo_create`是PostgreSQL 8.1 新增的；如果将该函数用于更早版本的服务器，它会失败并返回 InvalidOid。

例如：

```
inv_oid = lo_create(conn, desired_oid);
```

29.3.2. 导入一个大对象

要把一个操作系统文件导入为大对象，调用

```
Oid lo_import(PGconn *conn, const char *filename);
```

filename 指定要作为大对象导入的操作系统文件名。返回值是分配给新大对象的 OID，失败时为 InvalidOid（零）。注意，该文件是由客户端接口库读取的，而不是由服务器读取的；因此它必须存在于客户端文件系统中，并且对客户端应用可读。

29.3.3. 导出一个大对象

要把一个大对象导出到操作系统文件，调用

```
int lo_export(PGconn *conn, Oid lobjId, const char *filename);
```

lobjId 参数指定要导出的大对象的 OID，*filename* 参数指定该文件的操作系统文件名。注意，该文件是由客户端接口库写入的，而不是由服务器写入的。成功时返回 1，失败时返回 -1。

29.3.4. 打开一个现有的大对象

要打开一个现有的大对象进行读取或写入，调用

```
int lo_open(PGconn *conn, Oid lobjId, int mode);
```

lobjId 参数指定要打开的大对象的 OID。*mode* 位控制该对象将以读取（`INV_READ`）、写入（`INV_WRITE`）或两者兼有的方式打开。（这些符号常量定义在头文件 `libpq/libpq-fs.h` 中。）大对象必须先创建才能打开。`lo_open` 返回一个（非负的）大对象描述符，供后续在 `lo_read`、`lo_write`、`lo_lseek`、`lo_tell` 和 `lo_close` 中使用。该描述符只在当前事务持续期间有效。失败时返回 -1。

服务器当前不区分 `INV_WRITE` 和 `INV_READ | INV_WRITE` 这两种模式：在这两种情况下都允许通过该描述符读取。不过，这两种模式与单独使用 `INV_READ` 存在一个重要区别：使用 `INV_READ` 时，不能通过该描述符写入，而且从该描述符读取到的数据会反映执行 `lo_`

open 时活动事务快照中的大对象内容，而不受本事务或其他事务之后写入的影响。对于以 INV_WRITE 打开的描述符，读取返回的数据会反映其他已提交事务的所有写入以及当前事务的写入。这类似于普通 SQL SELECT 命令在 SERIALIZABLE 与 READ COMMITTED 事务模式下的行为差异。

例如：

```
inv_fd = lo_open(conn, inv_oid, INV_READ|INV_WRITE);
```

29.3.5. 向大对象写入数据

The function

```
int lo_write(PGconn *conn, int fd, const char *buf, size_t len);
```

把 *buf* 中的 *len* 个字节写入大对象描述符 *fd*。*fd* 参数必须是先前由 lo_open 返回的。返回值是实际写入的字节数。发生错误时，返回值为负。

29.3.6. 从大对象读取数据

The function

```
int lo_read(PGconn *conn, int fd, char *buf, size_t len);
```

从大对象描述符 *fd* 中读取 *len* 个字节到 *buf* 中。*fd* 参数必须是先前由 lo_open 返回的。返回值是实际读取的字节数。发生错误时，返回值为负。

29.3.7. 在大对象中定位

要更改与大对象描述符关联的当前读或写位置，调用

```
int lo_lseek(PGconn *conn, int fd, int offset, int whence);
```

该函数将由 *fd* 标识的大对象描述符的当前位置指针移动到由 *offset* 指定的新位置。*whence* 的有效值是 SEEK_SET（从对象起始处定位）、SEEK_CUR（从当前位置定位）以及 SEEK_END（从对象末尾定位）。返回值是新的位置指针，出错时为 -1。

29.3.8. 获取大对象的当前位置

要获取大对象描述符的当前读或写位置，调用

```
int lo_tell(PGconn *conn, int fd);
```

发生错误时，返回值为负。

29.3.9. 关闭一个大对象描述符

可以通过调用

```
int lo_close(PGconn *conn, int fd);
```

来关闭一个大对象描述符，其中 *fd* 是由 lo_open 返回的大对象描述符。成功时，lo_close 返回零。出错时，返回值为负。

任何在事务结束时仍保持打开的大对象描述符都会被自动关闭。

29.3.10. 移除一个大对象

要从数据库中移除一个大对象，调用

```
int lo_unlink(PGconn *conn, Oid lobjId);
```

lobjId 参数指定要移除的大对象的 OID。成功时返回 1，失败时返回 -1。

29.4. 服务器端函数

与前面介绍的每个客户端函数相对应，都有可以从 SQL 调用的服务端函数；实际上，大多数客户端函数都只是等价服务端函数的接口。通过 SQL 命令调用起来真正方便的有 `lo_creat`、`lo_create`、`lo_unlink`、`lo_import` 和 `lo_export`。下面是它们的使用示例：

```
CREATE TABLE image (
    name            text,
    raster          oid
);

SELECT lo_creat(-1);       -- returns OID of new, empty large object

SELECT lo_create(43213);   -- attempts to create large object with OID
 43213

SELECT lo_unlink(173454);  -- deletes large object with OID 173454

INSERT INTO image (name, raster)
    VALUES ('beautiful image', lo_import('/etc/motd'));

SELECT lo_export(image.raster, '/tmp/motd') FROM image
    WHERE name = 'beautiful image';
```

服务器端 `lo_import` 和 `lo_export` 函数的行为与客户端对应函数有很大不同。这两个函数使用运行数据库服务器的操作系统用户的权限，读写服务器文件系统中的文件，
因此仅限超级用户使用。相比之下，客户端导入和导出函数使用客户端程序的权限，
读写客户端文件系统中的文件。客户端函数可以被任何 PostgreSQL 用户使用。

29.5. 示例程序

例 29.1是一个示例程序，它展示了如何使用libpq中的大对象接口。程序中有些部分已经被注释掉，但仍保留在源代码中供读者参考。该程序也可以在源代码发行包的`src/test/examples/testlo.c`中找到。

例 29.1. 用 libpq 操作大对象的示例程序

```
/*--------------------------------------------------------------
 *
 * testlo.c--
 *    test using large objects with libpq
 *
 * Copyright (c) 1994, Regents of the University of California
```

```
 *
 *-------------------------------------------------------------
 */
#include <stdio.h>
#include "libpq-fe.h"
#include "libpq/libpq-fs.h"

#define BUFSIZE          1024

/*
 * importFile
 *    import file "in_filename" into database as large object "lobj
Oid"
 *
 */
Oid
importFile(PGconn *conn, char *filename)
{
    Oid         lobjId;
    int         lobj_fd;
    char        buf[BUFSIZE];
    int         nbytes,
                tmp;
    int         fd;

    /*
     * open the file to be read in
     */
    fd = open(filename, O_RDONLY, 0666);
    if (fd < 0)
    {                           /* error */
        fprintf(stderr, "cannot open unix file %s\n", filename);
    }

    /*
     * create the large object
     */
    lobjId = lo_creat(conn, INV_READ | INV_WRITE);
    if (lobjId == 0)
        fprintf(stderr, "cannot create large object\n");

    lobj_fd = lo_open(conn, lobjId, INV_WRITE);

    /*
     * read in from the Unix file and write to the inversion file
     */
    while ((nbytes = read(fd, buf, BUFSIZE)) > 0)
    {
        tmp = lo_write(conn, lobj_fd, buf, nbytes);
        if (tmp < nbytes)
            fprintf(stderr, "error while reading large object\n");
    }

    (void) close(fd);
```

```
    (void) lo_close(conn, lobj_fd);

    return lobjId;
}

void
pickout(PGconn *conn, Oid lobjId, int start, int len)
{
    int         lobj_fd;
    char       *buf;
    int         nbytes;
    int         nread;

    lobj_fd = lo_open(conn, lobjId, INV_READ);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    lo_lseek(conn, lobj_fd, start, SEEK_SET);
    buf = malloc(len + 1);

    nread = 0;
    while (len - nread > 0)
    {
        nbytes = lo_read(conn, lobj_fd, buf, len - nread);
        buf[nbytes] = ' ';
        fprintf(stderr, ">>> %s", buf);
        nread += nbytes;
    }
    free(buf);
    fprintf(stderr, "\n");
    lo_close(conn, lobj_fd);
}

void
overwrite(PGconn *conn, Oid lobjId, int start, int len)
{
    int         lobj_fd;
    char       *buf;
    int         nbytes;
    int         nwritten;
    int         i;

    lobj_fd = lo_open(conn, lobjId, INV_WRITE);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    lo_lseek(conn, lobj_fd, start, SEEK_SET);
    buf = malloc(len + 1);
```

```
    for (i = 0; i < len; i++)
        buf[i] = 'X';
    buf[i] = ' ';

    nwritten = 0;
    while (len - nwritten > 0)
    {
        nbytes = lo_write(conn, lobj_fd, buf + nwritten, len -
 nwritten);
        nwritten += nbytes;
    }
    free(buf);
    fprintf(stderr, "\n");
    lo_close(conn, lobj_fd);
}

/*
 * exportFile
 *    export large object "lobjOid" to file "out_filename"
 *
 */
void
exportFile(PGconn *conn, Oid lobjId, char *filename)
{
    int         lobj_fd;
    char        buf[BUFSIZE];
    int         nbytes,
                tmp;
    int         fd;

    /*
     * open the large object
     */
    lobj_fd = lo_open(conn, lobjId, INV_READ);
    if (lobj_fd < 0)
    {
        fprintf(stderr, "cannot open large object %d\n",
                lobjId);
    }

    /*
     * open the file to be written to
     */
    fd = open(filename, O_CREAT | O_WRONLY, 0666);
    if (fd < 0)
    {                           /* error */
        fprintf(stderr, "cannot open unix file %s\n",
                filename);
    }

    /*
     * read in from the inversion file and write to the Unix file
     */
```

```
    while ((nbytes = lo_read(conn, lobj_fd, buf, BUFSIZE)) > 0)
    {
        tmp = write(fd, buf, nbytes);
        if (tmp < nbytes)
        {
            fprintf(stderr, "error while writing %s\n",
                    filename);
        }
    }

    (void) lo_close(conn, lobj_fd);
    (void) close(fd);

    return;
}

void
exit_nicely(PGconn *conn)
{
    PQfinish(conn);
    exit(1);
}

int
main(int argc, char **argv)
{
    char       *in_filename,
               *out_filename;
    char       *database;
    Oid         lobjOid;
    PGconn     *conn;
    PGresult   *res;

    if (argc != 4)
    {
        fprintf(stderr, "Usage: %s database_name in_filename out_
filename\n",
                argv[0]);
        exit(1);
    }

    database = argv[1];
    in_filename = argv[2];
    out_filename = argv[3];

    /*
     * set up the connection
     */
    conn = PQsetdb(NULL, NULL, NULL, NULL, database);

    /* check to see that the backend connection was successfully made
 */
    if (PQstatus(conn) == CONNECTION_BAD)
    {
```

```
        fprintf(stderr, "Connection to database '%s' failed.\n",
 database);
        fprintf(stderr, "%s", PQerrorMessage(conn));
        exit_nicely(conn);
    }

    res = PQexec(conn, "begin");
    PQclear(res);

    printf("importing file %s\n", in_filename);
/*  lobjOid = importFile(conn, in_filename); */
    lobjOid = lo_import(conn, in_filename);
/*
    printf("as large object %d.\n", lobjOid);

    printf("picking out bytes 1000-2000 of the large object\n");
    pickout(conn, lobjOid, 1000, 1000);

    printf("overwriting bytes 1000-2000 of the large object with X's
\n");
    overwrite(conn, lobjOid, 1000, 1000);
*/

    printf("exporting large object to file %s\n", out_filename);
/*    exportFile(conn, lobjOid, out_filename); */
    lo_export(conn, lobjOid, out_filename);

    res = PQexec(conn, "end");
    PQclear(res);
    PQfinish(conn);
    exit(0);
}
```

第 30 章 ECPG - C 中的嵌入式 SQL

这一章描述了用于PostgreSQL的嵌入式SQL包。它由 Linus Tolke（<linus@epact.se>）和 Michael Meskes（<meskes@postgresql.org>）编写。最初它是为了与C一起工作而编写的。它也能与C++配合，但是它还不识别所有的C++结构。

这份文档还很不完整。不过，由于这一接口是标准化的，因此可以在许多关于 SQL 的资料中找到补充信息。

30.1. 概念

嵌入式 SQL 程序由普通编程语言编写的代码（这里是 C）与位于特殊标记区段中的 SQL 命令混合组成。要构建该程序，首先要将源代码交给嵌入式 SQL 预处理器，它会把源代码转换成普通 C 程序，之后就可以由 C 编译器处理。

与其他在 C 代码中处理SQL命令的方法相比，嵌入式SQL有其优势。首先，它能处理与你的C程序变量之间来回传递信息的繁琐工作。其次，程序中的 SQL 代码会在构建时接受语法正确性检查。第三，C 中的嵌入式SQL由SQL标准规定，并且得到许多其他SQL数据库系统的支持。PostgreSQL 的实现被设计为尽可能贴合该标准，因此通常可以比较容易地把为其他 SQL 数据库编写的嵌入式SQL程序移植到PostgreSQL。

如前所述，为嵌入式 SQL 接口编写的程序，就是插入了特殊代码的普通 C 程序，这些特殊代码用于执行数据库相关操作，其形式始终如下：

```
EXEC SQL ...;
```

这些语句在语法上占据一条 C 语句的位置。根据具体语句的不同，它们可以出现在全局层级，也可以出现在函数内部。嵌入式 SQL 语句遵循普通 SQL 代码的大小写规则，而不是 C 的规则。此外，它们还允许 SQL 标准中规定的嵌套 C 风格注释。但程序中的 C 部分仍遵循 C 标准，不接受嵌套注释。

下列小节解释了所有嵌入式 SQL 语句。

30.2. 连接到数据库服务器

使用以下语句连接数据库：

```
EXEC SQL CONNECT TO target [AS connection-name] [USER user-name];
```

其中，*target* 可以按以下方式指定：

- *dbname*[@*hostname*][:*port*]
- tcp:postgresql://*hostname*[:*port*][/*dbname*][?*options*]
- unix:postgresql://*hostname*[:*port*][/*dbname*][?*options*]
- 包含上述某种形式的 SQL 字符串常量
- 对包含上述某种形式的字符变量的引用（见示例）
- DEFAULT

如果直接写出连接目标（即不通过变量引用），且未用引号括起该值，就会应用普通 SQL 的大小写不敏感规则。此时，也可以按需分别用双引号括起各个参数。实际使用中，

采用单引号括起的字符串字面值或变量引用，可能更不容易出错。连接目标 `DEFAULT` 会以默认用户名发起到默认数据库的连接，此时不能单独指定用户名或连接名。

也有不同的方法来指定用户名：

- *username*
- *username/password*
- *username* IDENTIFIED BY *password*
- *username* USING *password*

如上所述，参数*username*以及*password*可以是一个 SQL 标识符、一个 SQL 字符串常量或者一个对字符变量的引用。

*connection-name*被用来在一个程序中处理多个连接。如果一个程序只使用一个连接，则可以省略它。最近被打开的连接将成为当前连接，当一个 SQL 语句要被执行时，将默认使用它（见这一章稍后的部分）。

以下是一些 `CONNECT` 语句的示例：

```
EXEC SQL CONNECT TO mydb@sql.mydomain.com;

EXEC SQL CONNECT TO 'unix:postgresql://sql.mydomain.com/mydb' AS
 myconnection USER john;

EXEC SQL BEGIN DECLARE SECTION;
const char *target = "mydb@sql.mydomain.com";
const char *user = "john";
EXEC SQL END DECLARE SECTION;
 ...
EXEC SQL CONNECT TO :target USER :user;
```

最后一种形式使用了上面提到的字符变量引用。后续章节会说明，如何在 C 变量前加冒号，以便在 SQL 语句中使用它们。

请注意，连接目标的格式并未在 SQL 标准中规定。因此，如果你想开发可移植的应用，最好采用类似上面最后一个示例的方法，把连接目标字符串封装在某个地方。

30.3. 关闭一个连接

要关闭连接，请使用以下语句：

```
EXEC SQL DISCONNECT [connection];
```

其中，*connection* 可以按以下方式指定：

- *connection-name*
- `DEFAULT`
- `CURRENT`
- `ALL`

如果未指定连接名，则关闭当前连接。

应用程序始终显式断开其打开的每个连接，是良好的编程习惯。

30.4. 运行 SQL 命令

任何 SQL 命令都可以在一个嵌入式 SQL 应用中被运行。下面是一些在嵌入式 SQL 应用中运行 SQL 命令的例子。

创建一个表：

```
EXEC SQL CREATE TABLE foo (number integer, ascii char(16));
EXEC SQL CREATE UNIQUE INDEX num1 ON foo(number);
EXEC SQL COMMIT;
```

插入行：

```
EXEC SQL INSERT INTO foo (number, ascii) VALUES (9999, 'doodad');
EXEC SQL COMMIT;
```

删除行：

```
EXEC SQL DELETE FROM foo WHERE number = 9999;
EXEC SQL COMMIT;
```

单行选择：

```
EXEC SQL SELECT foo INTO :FooBar FROM table1 WHERE ascii = 'doodad';
```

使用游标选择：

```
EXEC SQL DECLARE foo_bar CURSOR FOR
    SELECT number, ascii FROM foo
    ORDER BY ascii;
EXEC SQL OPEN foo_bar;
EXEC SQL FETCH foo_bar INTO :FooBar, DooDad;
...
EXEC SQL CLOSE foo_bar;
EXEC SQL COMMIT;
```

更新：

```
EXEC SQL UPDATE foo
    SET ascii = 'foobar'
    WHERE number = 9999;
EXEC SQL COMMIT;
```

`:`*`something`*形式的词元是主变量，即它们指向 C 程序中的变量。其说明见第 30.6 节。

在默认模式下，只有发出`EXEC SQL COMMIT`时语句才会提交。嵌入式 SQL 接口也支持事务自动提交（类似于libpq的行为），可以通过`ecpg`的`-t`命令行选项（见下文）或者`EXEC SQL SET AUTOCOMMIT TO ON`语句启用。在自动提交模式中，除非位于显式事务块内，每条命令都会自动提交。这种模式也可以通过`EXEC SQL SET AUTOCOMMIT TO OFF`显式关闭。

30.5. 选择一个连接

上一节中展示的 SQL 语句在当前连接（即最近打开的连接）上执行。如果一个应用需要管理多个连接，有两种处理方式。

第一种方式是显式地为每一个 SQL 语句选择一个连接，例如：

```
EXEC SQL AT connection-name SELECT ...;
```

如果应用需要以混合的顺序使用多个连接，这种方式特别合适。

如果你的应用使用多个线程执行，它们不能并发地共享一个连接。
你必须显式地控制对连接的访问（使用互斥量）或者为每个线程使用一个连接。
如果每个线程使用自己的连接，将需要使用 AT 子句来指定该线程将使用的连接。

第二种方式是执行一个语句来切换当前的连接。该语句是：

```
EXEC SQL SET CONNECTION connection-name;
```

如果有很多语句要在同一个连接上执行，这种方式尤其方便。

30.6. 使用主变量

在第 30.4 节中，你已经看到如何从嵌入式 SQL 程序中执行 SQL 语句。其中一些语句只使用固定值，既无法把用户提供的值插入语句中，也无法让程序处理查询返回的值。
这类语句在实际应用中并不真正有用。本节将详细解释如何使用一种称为主变量的简单机制，在 C 程序与嵌入式 SQL 语句之间传递数据。

30.6.1. 概述

在嵌入式 SQL 中，在 C 程序和 SQL 语句之间传递数据特别简单。我们不必让程序把数据拼接进语句中去，那样会带来诸如正确为值加引号等各种复杂问题；只需在 SQL 语句中写出 C 变量的名称，并在前面加上一个冒号即可。例如：

```
EXEC SQL INSERT INTO sometable VALUES (:v1, 'foo', :v2);
```

该语句引用了两个名为`v1`和`v2`的 C 变量，同时还使用了一个普通的 SQL 字符串常量，以说明你并不局限于只使用某一种数据。

这种在 SQL 语句中插入 C 变量的方式可以用在 SQL 语句中每一个应该出现值表达式的地方。在 SQL 环境中，我们把对 C 变量的引用称为主变量。

30.6.2. 声明节

要把数据从程序传给数据库（例如作为查询参数），或者把数据从数据库传回程序，
用来承载这些数据的 C 变量必须声明在经过特殊标记的区段中，这样嵌入式 SQL 预处理器才能识别它们。

这个区段始于：

```
EXEC SQL BEGIN DECLARE SECTION;
```

并结束于：

```
EXEC SQL END DECLARE SECTION;
```

在这两行之间，必须是正常的 C 变量声明，例如：

```
int   x;
char  foo[16], bar[16];
```

一个程序中可以根据需要放置任意多个声明节。

这些声明也会作为普通 C 变量声明写入输出文件，因此无需再次声明它们。不准备在 SQL 命令中使用的变量可以照常在这些特殊声明节之外声明。

结构体或联合的定义也必须写在DECLARE声明节中。否则预处理器无法处理这些类型，因为它不知道它们的定义。

特殊类型VARCHAR会被转换成每个变量的一个命名的struct。像这样的声明：

```
VARCHAR var[180];
```

会被转换为：

```
struct varchar_var { int len; char arr[180]; } var;
```

这个结构适合与varchar类型的 SQL 数据交互。

30.6.3. SELECT INTO与FETCH INTO

现在你应该能够把程序产生的数据传递到一个 SQL 命令中了。但是怎么检索一个查询的结果呢？为此，嵌入式 SQL 提供了常规命令SELECT和FETCH的特殊变体。这些命令有一个特殊的INTO子句，它指定被检索到的值要被存储在哪些主变量中。

这里是一个例子：

```
/*
 * 假定有下列表:
 * CREATE TABLE test1 (a int, b varchar(50));
 */

EXEC SQL BEGIN DECLARE SECTION;
int v1;
VARCHAR v2;
EXEC SQL END DECLARE SECTION;

 ...

EXEC SQL SELECT a, b INTO :v1, :v2 FROM test;
```

那么INTO子句出现在选择列表和FROM子句之间。选择列表中的元素数量必须和INTO后面列表（也被称为目标列表）的元素数量相等。

这里有一个使用命令FETCH的例子：

```
EXEC SQL BEGIN DECLARE SECTION;
int v1;
VARCHAR v2;
EXEC SQL END DECLARE SECTION;

 ...

EXEC SQL DECLARE foo CURSOR FOR SELECT a, b FROM test;

 ...
```

```
do
{
    ...
    EXEC SQL FETCH NEXT FROM foo INTO :v1, :v2;
    ...
} while (...);
```

这里INTO子句出现在所有正常子句的后面。

30.6.4. 指示符

上面这两种方法每次都只能检索单行数据。如果需要处理可能包含多于一行的结果集，就需要使用游标，如第二个示例所示。

上面的示例没有处理空值。实际上，如果读取示例从数据库取得空值，就会抛出错误。要向数据库传入空值或从数据库取得空值，需要在每个包含数据的主变量后，再指定第二个主变量。第二个主变量称为指示符，其中包含一个标志，表示该数据是否为空值；如果为空值，就会忽略实际主变量的值。下面的示例能正确处理空值的读取：

```
EXEC SQL BEGIN DECLARE SECTION;
VARCHAR val;
int val_ind;
EXEC SQL END DECLARE SECTION:

 ...

EXEC SQL SELECT b INTO :val :val_ind FROM test1;
```

指示符变量 val_ind 在数据非空时为零，在数据为空值时为负数。

指示符有另一种功能：如果指示符值为正，它表示值不为空，但是当它被存储在主变量中时已被截断。

30.7. 动态 SQL

在很多情况中，一个应用必须要执行的特定 SQL 语句在编写该应用时就已知。不过在某些情况中，SQL 语句在运行时构造或者由一个外部来源提供。这样你就不能直接把 SQL 语句嵌入到 C 源代码，不过有一种功能允许你调用在一个字符串变量中提供的任意 SQL 语句。

执行一个任意 SQL 语句的最简单方法是使用命令EXECUTE IMMEDIATE。例如：

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "CREATE TABLE test1 (...);";
EXEC SQL END DECLARE SECTION;

EXEC SQL EXECUTE IMMEDIATE :stmt;
```

你不能用这种方法执行检索数据的语句（例如SELECT）。

执行任意 SQL 语句的一种更强大的方法是将语句预备一次，然后在每次需要时执行该预备语句。也可以预备一个一般化的语句，然后通过替换参数执行它的特定版本。在预备语句时，在你想要稍后替换参数的地方写上问号。例如：

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "INSERT INTO test1 VALUES(?, ?);";
```

```
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE mystmt FROM :stmt;
 ...
EXEC SQL EXECUTE mystmt USING 42, 'foobar';
```

如果你正在执行的语句返回值，那么加上一个INTO子句：

```
EXEC SQL BEGIN DECLARE SECTION;
const char *stmt = "SELECT a, b, c FROM test1 WHERE a > ?";
int v1, v2;
VARCHAR v3;
EXEC SQL END DECLARE SECTION;

EXEC SQL PREPARE mystmt FROM :stmt;
 ...
EXEC SQL EXECUTE mystmt INTO v1, v2, v3 USING 37;
```

一个EXECUTE命令可以带INTO子句、USING子句、两者都带或两者都不带。

当你不再需要该预备语句时，你应该释放它：

```
EXEC SQL DEALLOCATE PREPARE name;
```

30.8. 使用描述符区域

SQL 描述符区域是一种处理SELECT、FETCH或 DESCRIBE 语句结果的高级方法。SQL 描述符区域将一行数据及其元数据项组合到一个数据结构中。在执行动态 SQL 语句时，结果列的性质可能无法提前获知，此时元数据特别有用。PostgreSQL 提供两种使用描述符区域的方式：命名 SQL 描述符区域和 C 结构体 SQLDA。

一个命名 SQL 描述符区域由一个头部以及一个或多个条目描述符区域构成，头部包含与整个描述符相关的信息，而条目描述符区域则描述结果行中的每一列。

在使用 SQL 描述符区域之前，需要先分配一个：

```
EXEC SQL ALLOCATE DESCRIPTOR identifier;
```

identifier 会作为该描述符区域的“变量名”。当不再需要该描述符时，应当释放它：

```
EXEC SQL DEALLOCATE DESCRIPTOR identifier;
```

要使用一个描述符区域，把它指定为INTO子句的存储目标（而不是列出主变量）：

```
EXEC SQL FETCH NEXT FROM mycursor INTO SQL DESCRIPTOR mydesc;
```

如果结果集为空，该描述符区域仍然会包含查询的元数据，即字段的名称。

那么如何从描述符区域中取得数据呢？你可以把描述符区域当作一个包含命名字段的结构。
要从头部取得某个字段的值并存入一个主变量，可使用以下命令：

```
EXEC SQL GET DESCRIPTOR name :hostvar = field;
```

目前只定义了一个头部字段：*COUNT*，
它表示存在多少个条目描述符区域（即结果中包含多少列）。主变量需要是整数类型。
要从条目描述符区域取得一个字段，可使用以下命令：

`EXEC SQL GET DESCRIPTOR name VALUE num :hostvar = field;`

*num*可以是一个整数字面量或者一个包含整数的宿主变量。可能的字段有：

`CARDINALITY` (integer)

结果集中的行数

`DATA`

实际的数据项（因此，这个字段的数据类型取决于查询）

`DATETIME_INTERVAL_CODE` (integer)

?

`DATETIME_INTERVAL_PRECISION` (integer)

没有实现

`INDICATOR` (integer)

指示符（表示一个空值或者一个值截断）

`KEY_MEMBER` (integer)

没有实现

`LENGTH` (integer)

以字符计的数据长度

`NAME` (string)

列名

`NULLABLE` (integer)

没有实现

`OCTET_LENGTH` (integer)

以字节计的数据字符表示的长度

`PRECISION` (integer)

精度（用于类型`numeric`）

`RETURNED_LENGTH` (integer)

以字符计的数据长度

`RETURNED_OCTET_LENGTH` (integer)

以字节计的数据字符表示的长度

`SCALE` (integer)

小数位数（用于类型 `numeric`）

`TYPE` (integer)

列的数据类型的数字编码

30.9. 错误处理

本节描述如何在嵌入式 SQL 程序中处理异常情况和警告。为此有几种可以同时使用的机制。

30.9.1. 设置回调

捕获错误和警告的一种简单方法是指定一个动作，在特定条件发生时执行。一般形式如下：

```
EXEC SQL WHENEVER condition action;
```

*condition*可以是以下值之一：

`SQLERROR`

只要在 SQL 语句执行期间发生一个错误就调用指定的动作。

`SQLWARNING`

只要在 SQL 语句执行期间发生一个警告就调用指定的动作。

`NOT FOUND`

只要一个 SQL 语句检索或者影响零行就调用指定的动作（这种情况不是一个错误，但是你可能需要特别地处理它）。

*action*可以是以下值之一：

`CONTINUE`

这实际上表示该情况被忽略。这是默认值。

```
GOTO label
GO TO label
```

跳到指定的标签（使用一个 C `goto`语句）。

`SQLPRINT`

把一个消息打印到标准错误。这对简单程序或原型开发很有用。消息的细节无法配置。

`STOP`

调用`exit(1)`终止程序。

`BREAK`

执行 C 语句`break`。只应被用在循环或`switch`语句中。

```
CALL name (args)
DO name (args)
```

使用指定参数调用指定的 C 函数。

SQL 标准只规定了以下动作：`CONTINUE` 和 `GOTO`（以及 `GO TO`）。

这里有一个可能会用在简单程序中的例子。当一个警告发生时它打印一个简单消息，而发生一个错误时它会中止程序：

```
EXEC SQL WHENEVER SQLWARNING SQLPRINT;
EXEC SQL WHENEVER SQLERROR STOP;
```

语句`EXEC SQL WHENEVER`是 SQL 预处理器的指令，而不是 C 语句。它设置的错误或警告动作适用于源代码中出现在处理器设置位置之后的所有嵌入式 SQL 语句，除非在第一个`EXEC SQL WHENEVER`与触发该条件的 SQL 语句之间，为同一条件设置了其他动作。这与 C 程序的控制流无关。因此，下面两个 C 程序片段都不会产生预期效果：

```
/*
 * 错误
 */
int main(int argc, char *argv[])
{
    ...
    if (verbose) {
        EXEC SQL WHENEVER SQLWARNING SQLPRINT;
    }
    ...
    EXEC SQL SELECT ...;
    ...
}

/*
 * 错误
 */
int main(int argc, char *argv[])
{
    ...
    set_error_handler();
    ...
    EXEC SQL SELECT ...;
    ...
}

static void set_error_handler(void)
{
    EXEC SQL WHENEVER SQLERROR STOP;
}
```

30.9.2. sqlca

为了支持更强大的错误处理，嵌入式 SQL 接口提供了一个名为`sqlca`（SQL 通信区域）的全局变量，其结构如下：

```
struct
{
    char sqlcaid[8];
    long sqlabc;
    long sqlcode;
```

```
    struct
    {
        int sqlerrml;
        char sqlerrmc[70];
    } sqlerrm;
    char sqlerrp[8];
    long sqlerrd[6];
    char sqlwarn[8];
    char sqlstate[5];
} sqlca;
```

（在一个多线程程序中，每一个线程会自动得到它自己的sqlca副本。这和对于标准 C 全局变量errno的处理相似。）

sqlca覆盖了警告和错误。如果执行一个语句时发生了多个警告或错误，那么sqlca将只包含关于最后一个的信息。

如果在上一个SQL语句中没有产生错误，sqlca.sqlcode将为 0 并且sqlca.sqlstate将为"00000"。如果发生一个警告或错误，则sqlca.sqlcode将为负并且sqlca.sqlstate将不为"00000"。一个正的sqlca.sqlcode表示一种无害的情况，例如上一个查询返回零行。sqlcode和sqlstate是两种不同的错误编码方案，详见下文。

如果上一个 SQL 语句成功，则sqlca.sqlerrd[1]包含所处理行的 OID（如果适用），sqlca.sqlerrd[2]包含处理或返回的行数（如果适用于该命令）。

发生错误或警告时，sqlca.sqlerrm.sqlerrmc包含描述该错误的字符串。字段sqlca.sqlerrm.sqlerrml包含存储在sqlca.sqlerrm.sqlerrmc中的错误消息的长度（即strlen()的结果，对于 C 程序员通常没有多少用处）。注意，有些消息太长，无法存入定长的sqlerrmc数组，因此会被截断。

发生警告时，sqlca.sqlwarn[2]被设为W（其他情况下，它被设为非W的值）。如果sqlca.sqlwarn[1]被设为W，说明某个值在存入主变量时被截断了。如果其他任一元素被设置为指示警告，sqlca.sqlwarn[0]就会被设为W。

字段sqlcaid、sqlabc、sqlerrp，以及sqlerrd和sqlwarn的其余元素，目前不包含有用信息。

SQL 标准中没有定义sqlca结构体，但是在一些其他的 SQL 数据库系统中都有实现。在核心上这些定义都相似，但是如果你想要编写可移植的应用，那么你应该仔细研究不同的实现。

30.9.3. SQLSTATE 与 SQLCODE

字段sqlca.sqlstate和sqlca.sqlcode采用两种不同的错误编码方案。两者都源自 SQL 标准，但SQLCODE在 SQL-92 标准中已被标记为弃用，并在后续版本中被删除。因此，强烈建议新应用使用SQLSTATE。

SQLSTATE是一个包含五个字符的数组。这五个字符由数字或大写字母组成，表示各种错误和警告条件的代码。SQLSTATE采用分层编码方案：前两个字符表示条件所属的大类，后三个字符表示该大类中的子类。代码00000表示成功状态。大部分SQLSTATE代码由 SQL 标准定义。PostgreSQL服务器原生支持SQLSTATE错误码，因此在所有应用中统一使用这种错误编码方案，可以实现较高的一致性。更多信息见附录 A。

已弃用的错误编码方案SQLCODE使用一个简单的整数。0 表示成功，正值表示成功并有附加信息，负值表示错误。SQL 标准只定义了正值 +100，表示上一个命令返回或影响了零行，没有定义

具体的负值。因此，这种方案的可移植性较差，也不采用分层编码。历史上，PostgreSQL的嵌入式 SQL 处理器为自身使用分配了一些特定的SQLCODE值，下文列出了它们的数值和符号名称。注意，这些值不能移植到其他 SQL 实现。为了便于将应用迁移到SQLSTATE方案，还列出了对应的SQLSTATE。不过，两种方案之间不存在一对一或一对多的映射关系（实际上是多对多），因此每次都应查阅附录 A中的完整SQLSTATE列表。

下面是这些被分配的 SQLCODE 值：

-12 (ECPG_OUT_OF_MEMORY)

表示你的虚拟内存已被耗尽。（SQLSTATE YE001）

-200 (ECPG_UNSUPPORTED)

表示预处理器生成了该库无法识别的内容。可能是所用的预处理器与库的版本不兼容（SQLSTATE YE002）。

-201 (ECPG_TOO_MANY_ARGUMENTS)

这表示命令指定了超过该命令预期数量的主变量（SQLSTATE 07001 或 07002）。

-202 (ECPG_TOO_FEW_ARGUMENTS)

这表示命令指定的主变量数量低于该命令的预期（SQLSTATE 07001 或 07002）。

-203 (ECPG_TOO_MANY_MATCHES)

这意味着一个查询已经返回了多个行，但是该语句只准备存储一个结果行（例如，因为指定的变量不是数组）（SQLSTATE 21000）。

-204 (ECPG_INT_FORMAT)

主变量是类型int而数据库中的数据是一种不同的类型并且含有一个不能被解释为int的值。该库使用strtol()进行这种转换（SQLSTATE 42804）。

-205 (ECPG_UINT_FORMAT)

主变量是类型unsigned int而数据库中的数据是一种不同的类型并且含有一个不能被解释为unsigned int的值。该库使用strtoul()进行这种转换（SQLSTATE 42804）。

-206 (ECPG_FLOAT_FORMAT)

主变量是类型float而数据库中的数据是另一种类型并且含有一个不能被解释为float的值。该库使用strtod()进行这种转换（SQLSTATE 42804）。

-211 (ECPG_CONVERT_BOOL)

这表示主变量是类型bool而数据库中的数据既不是't'也不是'f'（SQLSTATE 42804）。

-212 (ECPG_EMPTY)

发送给PostgreSQL服务器的语句是空的（通常在一个嵌入式 SQL 程序中不会发生，因此它可能指向一个内部错误）（SQLSTATE YE002）。

-213 (ECPG_MISSING_INDICATOR)

返回了一个空值，但未提供空值指示符变量（SQLSTATE 22002）。

-214 (ECPG_NO_ARRAY)

在要求一个数组的地方使用了一个普通变量（SQLSTATE 42804）。

-215 (ECPG_DATA_NOT_ARRAY)

在一个要求数组值的地方数据库返回了一个普通变量（SQLSTATE 42804）。

-220 (ECPG_NO_CONN)

程序尝试访问一个不存在的连接（SQLSTATE 08003）。

-221 (ECPG_NOT_CONN)

程序尝试访问一个存在的连接但是它没有打开（这是一个内部错误）（SQLSTATE YE002）。

-230 (ECPG_INVALID_STMT)

尝试使用的语句尚未预备（SQLSTATE 26000）。

-240 (ECPG_UNKNOWN_DESCRIPTOR)

未找到指定的描述符。尝试使用的语句尚未预备（SQLSTATE 33000）。

-241 (ECPG_INVALID_DESCRIPTOR_INDEX)

指定的描述符超出范围（SQLSTATE 07009）。

-242 (ECPG_UNKNOWN_DESCRIPTOR_ITEM)

请求了一个无效的描述符条目（这是一个内部错误）（SQLSTATE YE002）。

-243 (ECPG_VAR_NOT_NUMERIC)

执行动态语句时，数据库返回了数值，而主变量不是数值类型（SQLSTATE 07006）。

-244 (ECPG_VAR_NOT_CHAR)

执行动态语句时，数据库返回了非数值的值，而主变量是数值类型（SQLSTATE 07006）。

-400 (ECPG_PGSQL)

PostgreSQL服务器导致了某个错误。该消息包含来自PostgreSQL服务器的错误消息。

-401 (ECPG_TRANS)

PostgreSQL服务器通知我们不能启动、提交或回滚事务（SQLSTATE 08007）。

-402 (ECPG_CONNECT)

到数据库的连接尝试没有成功（SQLSTATE 08001）。

100 (ECPG_NOT_FOUND)

这是一种无害情况，它表示上一个命令检索或者处理了零行，
或者你到达了游标的末尾（SQLSTATE 02000）。

30.10. 包含文件

要在你的嵌入式 SQL 程序中包含一个外部文件，可以使用：

```
EXEC SQL INCLUDE filename;
```

嵌入式 SQL 预处理器会寻找一个名为*filename*.h的文件，对它进行预处理，并把结果包含到生成的 C 输出中。这样，被包含文件中的嵌入式 SQL 语句就能被正确处理。

注意，这与

```
#include <filename.h>
```

不同，因为这个文件不会经过 SQL 命令预处理。自然，你可以继续使用 C 的#include指令来包含其他头文件。

> **注意**
>
> 被包含文件的名字是区分大小写的，尽管EXEC SQL INCLUDE命令的其余部分遵循通常的 SQL 大小写敏感性规则。

30.11. 处理嵌入式 SQL 程序

现在你已经对如何构造嵌入式 SQL C 程序有所了解了，你可能希望知道如何编译它们。在编译之前，你需要让该文件通过嵌入式SQL C预处理器，它会把你用到的SQL转换成特殊的函数调用。在编译之后，你必须链接一个包含所需函数的特殊库。这些函数从参数中取得信息、使用libpq执行SQL命令并且把结果放在指定的参数中用来输出。

该预处理器程序被称作ecpg并且被包括在一个正常的PostgreSQL安装中。嵌入式 SQL 程序通常带有扩展名.pgc。如果你有一个程序文件prog1.pgc，
你可以调用下面的命令对它进行预处理：

```
ecpg prog1.pgc
```

这将创建一个文件prog1.c。如果你的输入文件不遵循建议的命名模式，你可以用
-o选项显式地指定输出文件。

预处理过的文件可以被正常地编译，例如：

```
cc -c prog1.c
```

生成的 C 源文件会包含PostgreSQL安装中的头文件。因此，如果PostgreSQL安装在默认搜索路径之外，就必须在编译命令行中添加类似-I/usr/local/pgsql/include的选项。

要链接嵌入式 SQL 程序，需要链接libecpg库，例如：

```
cc -o myprog prog1.o prog2.o ... -lecpg
```

同样，可能需要在该命令行中添加类似-L/usr/local/pgsql/lib的选项。

如果你使用make来管理一个大工程的构建过程，把下面的隐式规则包括在你的 makefile 中将会很方便：

```
ECPG = ecpg
```

```
%.c: %.pgc
        $(ECPG) $<
```

ecpg命令的完整语法可见ecpg。

如果构建 ecpg 库时使用了 configure 的 --enable-thread-safety 命令行选项，它就是线程安全的。（编译客户端代码时，你可能还需要使用其他线程相关的命令行选项。）

30.12. 库函数

libecpg库主要包含用于实现嵌入式 SQL 命令功能的“隐藏”函数，也有一些适合直接调用的函数。注意，直接调用这些函数会使代码失去可移植性。

- 调用ECPGdebug(int *on*, FILE **stream*)时，如果第一个参数非零，就会启用调试日志。日志写入*stream*，包含代入全部输入变量后的所有SQL语句，以及PostgreSQL服务器返回的结果。这对于查找SQL语句中的错误很有用。

  注意

  在 Windows 上，如果ecpg库和应用是使用不同标志编译的，这个函数调用将会使应用崩溃，因为FILE指针的内部表示不同。特别地，库和使用库的应用应该使用相同的多线程/单线程、发行/调试以及静态/动态标志。

- 如果你连接到了一个数据库，ECPGstatus(int *lineno*, const char* *connection_name*)返回真，否则返回假。如果使用的是一个单一连接，*connection_name*可以为NULL。

30.13. 内部

这一节解释ECPG在内部如何工作。这些信息有时有助于用户理解如何使用ECPG。

ecpg写入输出的前四行是固定的：其中两行是注释，另外两行是与库接口所必需的包含行。随后预处理器会通读整个文件并写出输出，通常它只是把所有内容原样回显到输出中。

当它看到EXEC SQL语句时，就会介入并对其进行改写。命令以EXEC SQL开始，以;结束。其间的所有内容都会被视为一条SQL语句，并进行变量替换解析。

当某个符号以冒号（:）开头时，就会发生变量替换。系统会在此前于EXEC SQL DECLARE声明节中声明的变量里查找同名变量。

库中最重要的函数是ECPGdo，负责执行大多数命令。它接收可变数量的参数，参数数量很容易达到 50 个左右；我们希望这在任何平台上都不会造成问题。

参数是：

一个行号

这是原始行的行号，只用于错误消息。

一个字符串

这是将要发出的SQL命令。输入变量会对它进行修改，也就是那些在编译时尚未知晓、但要填入命令中的变量。字符串中的?表示变量应当放置的位置。

输入变量

每个输入变量都会生成十个参数（见下文）。

ECPGt_EOIT

一个表明后面没有更多输入变量的`enum`值。

输出变量

每个输出变量都会生成十个参数（见下文）。这些变量将由该函数填充。

ECPGt_EORT

一个表明后面没有更多变量的`enum`值。

对于每一个作为SQL命令一部分的变量，该函数得到十个参数：

1. 用特殊符号表示的类型。
2. 一个值的指针或者一个指针的指针。
3. 如果变量是一个`char`或者`varchar`，这是它的尺寸。
4. 数组中元素的数量（用于数组获取）。
5. 数组中下一个元素的偏移量（用于数组获取）。
6. 作为一个特殊符号的指示符变量的类型。
7. 一个指示符变量的指针。
8. 0
9. 指示符数组中的元素数量（用于数组取数）。
10. 到指示符数组中下一个元素的偏移量（用于数组取数）。

注意并非所有 SQL 命令都被以这种方式对待。例如，一个打开游标语句：

```
EXEC SQL OPEN cursor;
```

它不会被复制到输出中。相反，游标的`DECLARE`命令会被放在`OPEN`命令的位置上，
因为真正打开游标的正是它。

这里有一个完整的例子，它描述了一个文件`foo.pgc`
的预处理器输出（对预处理器的每一个特定版本细节可能不同）：

```
EXEC SQL BEGIN DECLARE SECTION;
int index;
int result;
EXEC SQL END DECLARE SECTION;
...
EXEC SQL SELECT res INTO :result FROM mytable WHERE index = :index;
```

会被转换成：

```
/* Processed by ecpg (2.6.0) */
```

```
/* These two include files are added by the preprocessor */
#include <ecpgtype.h>;
#include <ecpglib.h>;

/* 声明节开始 */

#line 1 "foo.pgc"

 int index;
 int result;
/* 声明节结束 */
...
ECPGdo(__LINE__, NULL, "SELECT res FROM mytable WHERE index = ?     ",
        ECPGt_int,&(index),1L,1L,sizeof(int),
        ECPGt_NO_INDICATOR, NULL , 0L, 0L, 0L, ECPGt_EOIT,
        ECPGt_int,&(result),1L,1L,sizeof(int),
        ECPGt_NO_INDICATOR, NULL , 0L, 0L, 0L, ECPGt_EORT);
#line 147 "foo.pgc"
```

（这里的缩进是为了可读性而添加的，并非是预处理器做的处理）。

第 31 章 信息模式

信息模式由一组视图组成，其中包含当前数据库中定义的对象的信息。信息模式由 SQL 标准定义，因此可以预期它具有可移植性并保持稳定；而系统目录则不同，它们是 PostgreSQL 特有的，并且是围绕实现方面的考虑建模的。不过，信息模式视图不包含 PostgreSQL 特有功能的信息；要查询这些信息，需要查询系统目录或其他 PostgreSQL 特有视图。

31.1. 模式

信息模式本身是一个名为`information_schema`的模式。这个模式自动存在于所有数据库中。这个模式的拥有者是集簇中的初始数据库用户，并且该用户自然地拥有这个模式上的所有权限，包括删除它的能力（但是这样节省的空间是很小的）。

默认情况下，信息模式不在模式搜索路径中，因此你需要使用限定名访问其中的所有对象。因为信息模式中的某些对象的名称是可能出现在用户应用中的一般名称，如果你想把该信息模式放在路径中，你应该小心。

31.2. 数据类型

信息模式视图的列使用在信息模式中定义的特殊数据类型。

这些类型为：

`cardinal_number`

一种非负整数。

`character_data`

一种字符串（没有指定最大长度）。

`sql_identifier`

一种字符串。该类型用于 SQL 标识符，其他任何文本数据则使用类型`character_data`。

`time_stamp`

在类型`timestamp with time zone`之上的一个域。

信息模式中的每一列都具有这四种类型之一。

在信息模式中，布尔（真/假）数据由包含`YES`或`NO`的`character_data`类型列来表示。（信息模式是在`boolean`类型被纳入 SQL 标准之前设计出来的，因此需要采用这一约定来保持信息模式向后兼容。）

31.3. information_schema_catalog_name

`information_schema_catalog_name` 是一个表，它始终只包含一行一列，用于存放当前数据库的名称（按 SQL 术语称为当前目录）。

表 31.1. information_schema_catalog_name 列

名称	数据类型	描述
catalog_name	sql_identifier	包含这个信息模式的数据库名称

31.4. applicable_roles

视图applicable_roles标识当前用户是其成员的所有组。（角色和组是同一个东西。）
一般来说，最好改用enabled_roles视图；另见那里的说明。

表 31.2. applicable_roles 列

名称	数据类型	描述
grantee	sql_identifier	总是当前用户的名称
role_name	sql_identifier	组名
is_grantable	character_data	适用于PostgreSQL中不可用的特性

31.5. check_constraints

视图check_constraints包含所有检查约束，无论它们是定义在表上还是定义在域上，
只要它们归当前启用角色所有即可显示（表或域的拥有者就是约束的拥有者）。

表 31.3. check_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含约束的数据库名称(总是当前数据库)
constraint_schema	sql_identifier	包含约束的模式的名称
constraint_name	sql_identifier	约束的名称
check_clause	character_data	该检查约束的检查表达式

31.6. column_domain_usage

视图column_domain_usage
标识所有使用定义在当前数据库中并且被一个当前启用的角色拥有的域的列（表列或视图列）。

表 31.4. column_domain_usage 列

名称	数据类型	描述
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称
column_name	sql_identifier	列名称

31.7. column_privileges

视图column_privileges标识所有授予给当前用户或者由当前用户授予的列权限。对每一个列、授予者、被授予者的组合只有一行。授予给组的权限在视图role_column_grants中标识。

在PostgreSQL中，你只能对整个表授予权限，而不能对单个列授权。因此，这个视图包含与table_privileges相同的信息，只是通过每个适当表中每一列一行的方式来表示，但它只覆盖列粒度可能的权限类型：SELECT、INSERT、UPDATE、REFERENCES。如果你想让你的应用适应将来可能的发展，在涉及这些权限类型时，用这个视图代替table_privileges通常是正确的选择。

表 31.5. column_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
table_catalog	sql_identifier	包含该列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该列的表所在的模式名称
table_name	sql_identifier	包含该列的表名称
column_name	sql_identifier	列名称
privilege_type	character_data	权限类型：SELECT、INSERT、UPDATE或 REFERENCES
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO

注意，grantee列不区分用户和组。如果你的用户和组同名，很遗憾无法区分它们。PostgreSQL的未来版本可能会禁止用户和组同名。

31.8. column_udt_usage

视图column_udt_usage标识所有使用被一个当前启用的角色拥有的数据类型的列。注意在PostgreSQL中，内置数据类型的行为和用户定义的类型相似，因此它们也被包括在这里。详见第 31.9 节。

表 31.6. column_udt_usage 列

名称	数据类型	描述
udt_catalog	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的模式名称
udt_name	sql_identifier	该列数据类型（如果适用，底层的域类型）的名称
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称

名称	数据类型	描述
column_name	sql_identifier	列名称

31.9. columns

视图columns包含数据库中有关所有表列（或视图列）的信息。系统列（oid等）不被包括在内。只有那些当前用户能够访问（作为拥有者或具有某些权限）的列才被显示。

表 31.7. columns 列

名称	数据类型	描述
table_catalog	sql_identifier	包含表的数据库的名称(总是当前数据库)
table_schema	sql_identifier	包含表的模式的名称
table_name	sql_identifier	表的名称
column_name	sql_identifier	列名称
ordinal_position	cardinal_number	该列在表内的顺序位置（从 1 开始计）
column_default	character_data	列的默认表达式（如果当前用户不是包含该列的表的属主，则为空值）
is_nullable	character_data	如果该列可以为空，则为YES，否则为NO。一个非空约束是让一列成为不能为空的方法，但还有其他方法。
data_type	character_data	如果该列的数据类型是一种内置类型，则为该列的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况下该类型被标识在udt_name和相关列中）。如果该列基于一个域，这一列引用该域底层的类型（该列被标识在domain_name和相关列中）。
character_maximum_length	cardinal_number	如果data_type标识字符类型或位串类型，这里是声明的最大长度；如果未声明最大长度，或者对于所有其他数据类型，则为空。
character_octet_length	cardinal_number	如果data_type标识字符类型，这里是一个数据值可能的最大字节长度（以字节计）（这一个数值不必让 PostgreSQL 用户关心）；对于所有其他数据类型则为空。
numeric_precision	cardinal_number	如果data_type标识一种数字类型，这列包含这个属性类型的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。

名称	数据类型	描述
numeric_precision_radix	cardinal_number	如果data_type标识一种数字类型，这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。对于所有其他数据类型，这一列为空。
numeric_scale	cardinal_number	如果data_type标识一种准确数字类型，这列包含这个属性类型的（声明的或隐式的）比例。标度表示小数点右侧的有效数字位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
datetime_precision	cardinal_number	如果data_type标识一种日期、时间或时间间隔类型，则为声明的精度；对所有其他数据类型或未声明精度时为空。
interval_type	character_data	尚未实现
interval_precision	character_data	还未被实现
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
domain_catalog	sql_identifier	如果该列有一个域类型，这里是该域所在的数据库名（总是当前数据库），否则为空。
domain_schema	sql_identifier	如果该列有一个域类型，这里是该域所在的模式名，否则为空。
domain_name	sql_identifier	如果该列有一个域类型，这里是该域的名称，否则为空。
udt_catalog	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该列数据类型（如果适用，底层的域类型）被定义的模式名称
udt_name	sql_identifier	该列数据类型（如果适用，底层的域类型）的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该列的数据类型描述符的一个标识符，在从属于该表的数据类型标识符之中唯一。这主要用于与这类标

名称	数据类型	描述
		识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。
is_self_referencing	character_data	对应 PostgreSQL 不支持的特性

因为在 SQL 中有多种方式定义数据类型，而PostgreSQL还包含额外的方式来定义数据类型，它们在信息模式中的表示可能有点困难。列data_type应该标识列的底层内置类型。在PostgreSQL中，这表示定义在系统目录模式pg_catalog中的类型。如果应用能够特别地（例如以不同方式格式化数字类型或使用精度列中的数据）处理众所周知的内置类型，这列可能会有用。列udt_name、udt_schema和udt_catalog总是标识列的底层数据类型，即使该列是基于一个域的（因为PostgreSQL对待内置类型和用户定义类型的方式是一样的，内置类型也出现在这里。这是 SQL 标准的一种扩展）。如果一个应用想要根据该类型以不同的方式处理数据，就应该使用这些列，因为在那种情况下即使该列真地基于一个域也没有关系。如果该列是基于一个域，该域的标识被存储在列domain_name、domain_schema和domain_catalog。如果你想要把列和它们相关的数据类型配对并且把域视作单独的类型，你可以写coalesce(domain_name, udt_name)等等。

31.10. constraint_column_usage

视图constraint_column_usage标识当前数据库中被某种约束使用的所有列。只显示那些位于当前启用角色所拥有表中的列。对于检查约束，此视图标识检查表达式中使用的列。对于非空约束，此视图标识定义该约束的列。对于外键约束，此视图标识外键所引用的列。对于唯一约束或主键约束，此视图标识受该约束限制的列。

表 31.8. constraint_column_usage 列

名称	数据类型	描述
table_catalog	sql_identifier	包含被某个约束使用的列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被某个约束使用的列的表所在的模式名称
table_name	sql_identifier	包含被某个约束使用的列的表名称
column_name	sql_identifier	包含被某个约束使用的列名称
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称

31.11. constraint_table_usage

视图constraint_table_usage标识在当前数据库中被某个约束使用的所有表（这与视图table_constraints不同，它标识哪些表约束定义在哪些表上）。对于一个外键约束，这个视图标识该外键引用的表。对于一个唯一或主键约束，这个视图仅标识该约束属于的表。检查约束和非空约束不被包括在这个视图中。

表 31.9. constraint_table_usage 列

名称	数据类型	描述
table_catalog	sql_identifier	包含被某个约束使用的表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被某个约束使用的表的模式名称
table_name	sql_identifier	包含被某个约束使用的表名称
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称

31.12. data_type_privileges

视图data_type_privileges
标识当前用户能够访问（作为被描述对象的拥有者或者具有其上的某种权限）
的所有数据类型描述符。只要一个数据类型被用在一个表列、
一个域或一个函数（作为参数或返回类型）
就会生成一个数据类型描述符并且在那个实例中存储一些有关该数据类型如何被使用的信息（例如，声明的最大长度，如果适用）。每一个数据类型描述符被赋予一个任意的标识符，
它在被赋予给一个对象（表、域、函数）的数据类型描述符中唯一。
这个视图对于应用可能没什么用，但是它被用于定义信息模式中的一些其他视图。

表 31.10. data_type_privileges 列

名称	数据类型	描述
object_catalog	sql_identifier	包含该被描述对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	包含该被描述对象的模式名称
object_name	sql_identifier	被描述对象的名称
object_type	character_data	被描述对象的类型：TABLE（从属于表的一列的数据类型描述符）、DOMAIN（从属于域的数据类型描述符）、ROUTINE（从属于函数的一个参数或返回数据类型的数据类型描述符）。
dtd_identifier	sql_identifier	数据类型描述符的标识符，它在同一对象的数据类型描述符之间唯一。

31.13. domain_constraints

视图domain_constraints包含所有属于当前数据库中定义的域的约束。

表 31.11. domain_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）

名称	数据类型	描述
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称
is_deferrable	character_data	如果该约束是可延迟的，则为YES，否则为NO
initially_deferred	character_data	如果该约束是可延迟的且初始就被延迟，则为YES，否则为NO

31.14. domain_udt_usage

视图domain_udt_usage标识所有基于被一个当前启用的角色拥有的数据类型的域。
注意在PostgreSQL中，内置数据类型的行为相似于用户定义的类型，因此它们也被包括在这里。

表 31.12. domain_udt_usage 列

名称	数据类型	描述
udt_catalog	sql_identifier	该域数据类型被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该域数据类型被定义的模式名称
udt_name	sql_identifier	该域数据类型的名称
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称

31.15. domains

视图domains包含当前数据库中定义的所有域。

表 31.13. domains 列

名称	数据类型	描述
domain_catalog	sql_identifier	包含该域的数据库名称（总是当前数据库）
domain_schema	sql_identifier	包含该域的模式名称
domain_name	sql_identifier	域名称
data_type	character_data	该域的数据类型如果是一种内置类型，这里是该域的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	如果该域有一个字符或位串类型，这里是声明的最大长度；如果没有

名称	数据类型	描述
		声明最大长度，则对于所有其他数据类型为空。
character_octet_length	cardinal_number	如果该域有一个字符类型，这里是一个数据值可能的最大字节长度（以字节计）（这一个数值不必让PostgreSQL 用户关心）；对其他所有数据类型则为空。
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
numeric_precision	cardinal_number	如果该域有一种数字类型，这列包含这个域类型的（声明的或隐式的）精度。精度指示了有效位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
numeric_precision_radix	cardinal_number	如果该域有一种数字类型，这一列指示numeric_precision和numeric_scale列中的值是基于什么来表示。该值为 2 或 10。对于所有其他数据类型，这一列为空。
numeric_scale	cardinal_number	如果该域有一种准确数字类型，这列包含这个域类型的（声明的或隐式的）比例。标度表示小数点右侧的有效数字位数。它可以按照列numeric_precision_radix中指定的被表示为十进制（基于 10）或二进制（基于 2）。对于所有其他数据类型，这一列为空。
datetime_precision	cardinal_number	如果该域的类型是日期、时间或时间间隔类型，则为声明的精度；对所有其他数据类型或未声明精度时为空。
interval_type	character_data	尚未实现
interval_precision	character_data	还未被实现
domain_default	character_data	该域的默认表达式
udt_catalog	sql_identifier	该域数据类型被定义的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该域数据类型被定义的模式名称
udt_name	sql_identifier	该域数据类型的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该域的数据类型描述符的一个标识符，在从属于该域的数据类型标识符之中唯一（这不重要，因为一个域只包含一个数据类型描述符）。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。

31.16. element_types

视图element_types包含数组元素的数据类型描述符。当一个表列、复合类型属性、域、函数参数或函数返回值被定义为一种数组类型时，相应的信息模式视图只在列data_type中包含ARRAY。要获得该数组元素类型的信息，你可以连接该相应的视图和这个视图。例如，要显示一个表的列及其数据类型和数组元素类型（如果适用），你可以：

```
SELECT c.column_name, c.data_type, e.data_type AS element_type
FROM information_schema.columns c LEFT JOIN information_schema.
element_types e
     ON ((c.table_catalog, c.table_schema, c.table_name, 'TABLE', c.
dtd_identifier)
       = (e.object_catalog, e.object_schema, e.object_name, e.object_
type, e.array_type_identifier))
WHERE c.table_schema = '...' AND c.table_name = '...'
ORDER BY c.ordinal_position;
```

这个视图只包括当前用户能够访问（作为拥有者或具有某些权限）的对象。

表 31.14. element_types 列

名称	数据类型	描述
object_catalog	sql_identifier	包含使用被描述的数组的对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	包含使用被描述的数组的对象的模式名称
object_name	sql_identifier	使用被描述的模式的对象名称
object_type	character_data	使用被描述的数组的对象的类型：TABLE（被一个表的某一列使用的数组）、DOMAIN（被该域使用的数组）、ROUTINE（被该函数的一个参数或返回数据类型使用的数组）。
array_type_identifier	sql_identifier	被描述的数组的数据类型描述符的标识符。使用这个去与其他信息模式视图的dtd_identifier列连接。
data_type	character_data	如果数组元素的数据类型是内置类型，这里是数组元素的数据类型，否则为USER-DEFINED（在那种

名称	数据类型	描述
		情况下，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的数组元素数据类型
domain_default	character_data	尚未实现
udt_catalog	sql_identifier	元素的数据类型所在的数据库名称（总是当前数据库）
udt_schema	sql_identifier	定义元素的数据类型的模式的名称
udt_name	sql_identifier	元素的数据类型名
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该元素的数据类型描述符的标识符。当前没有用。

31.17. enabled_roles

视图enabled_roles标识当前用户是其成员的所有组。（角色和组是同一个东西。）
这个视图与applicable_roles
的区别在于：将来可能会出现一种在会话期间启用和禁用组的机制。
届时这个视图标识的是当前已启用的那些组。

表 31.15. enabled_roles 列

名称	数据类型	描述
role_name	sql_identifier	组名

31.18. key_column_usage

视图key_column_usage标识当前数据库中所有被某种唯一、主键或外键约束限制的列。
检查约束不被包括在这个视图中。此视图中只显示包含在当前用户拥有的表中的列。

表 31.16. key_column_usage 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
table_catalog	sql_identifier	包含被这个约束限制的列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被这个约束限制的列的表所在的模式名称
table_name	sql_identifier	包含被这个约束限制的列的表的名称
column_name	sql_identifier	被这个约束限制的列的名称
ordinal_position	cardinal_number	该列在约束键中的顺序位置（从 1 开始计数）

31.19. parameters

视图parameters包含当前数据库中所有函数的参数的有关信息。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的函数才会被显示。

表 31.17. parameters 列

名称	数据类型	描述
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 31.26 节。
ordinal_position	cardinal_number	该参数在函数参数列表中的顺序位置（从 1 开始计数）

名称	数据类型	描述
parameter_mode	character_data	总是IN，表示输入参数（将来可能会有其他参数模式）
is_result	character_data	对应 PostgreSQL 不支持的特性
as_locator	character_data	对应 PostgreSQL 不支持的特性
parameter_name	sql_identifier	参数名称，如果参数没有名称则为空
data_type	character_data	该参数的数据类型如果是一种内置类型，这里是该参数的数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在udt_name和相关列中）。
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
udt_catalog	sql_identifier	该参数的数据类型所在的数据库名称（总是当前数据库）
udt_schema	sql_identifier	该参数的数据类型所在的模式名称
udt_name	sql_identifier	该参数的数据类型的名称
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该参数的数据类型描述符的一个标识符，在从属于该函数的数据类型标识符之中唯一（这不重要，因

名称	数据类型	描述
		为一个域只包含一个数据类型描述符）。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。

31.20. referential_constraints

视图referential_constraints
包含当前数据库中属于一个当前启用角色所拥有的表的所有引用（外键）约束。

表 31.18. referential_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含约束的数据库名称(总是当前数据库)
constraint_schema	sql_identifier	包含约束的模式的名称
constraint_name	sql_identifier	约束的名称
unique_constraint_catalog	sql_identifier	包含该外键约束所引用的唯一或主键约束的数据库名称（总是当前数据库）
unique_constraint_schema	sql_identifier	包含该外键约束所引用的唯一或主键约束的模式名称
unique_constraint_name	sql_identifier	包含该外键约束所引用的唯一或主键约束的名称
match_option	character_data	外键约束的匹配选项：FULL、PARTIAL或NONE。
update_rule	character_data	外键约束的更新规则：CASCADE、SET NULL、SET DEFAULT、RESTRICT或 NO ACTION。
delete_rule	character_data	外键约束的删除规则：CASCADE、SET NULL、SET DEFAULT、RESTRICT或 NO ACTION。

31.21. role_column_grants

视图role_column_grants标识所有在列上授予的权限，
这些权限的授予者或者被授予者是当前已启用的角色。更多信息可以在column_privileges中找到。

表 31.19. role_column_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
table_catalog	sql_identifier	包含该列的表所在的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该列的表所在的模式名称
table_name	sql_identifier	包含该列的表名称

名称	数据类型	描述
column_name	sql_identifier	列名称
privilege_type	character_data	权限类型：SELECT、INSERT、UPDATE或 REFERENCES
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO

31.22. role_routine_grants

视图role_routine_grants
标识所有授予给一个当前启用的角色或者被一个当前启用的角色授予的权限。对于每一种函数、授予者和被授予者的组合，这里都有一行。更多信息可以在routine_privileges中找到。

表 31.20. role_routine_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 31.26 节。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）
privilege_type	character_data	总是为EXECUTE（函数唯一的权限类型）
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO

31.23. role_table_grants

视图role_table_grants标识所有在表或视图上授予的权限，
这些权限的授予者或者被授予者是当前已启用的角色。更多信息可以在table_privileges中找到。

表 31.21. role_table_grants 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称

名称	数据类型	描述
privilege_type	character_data	该权限的类型：SELECT、DELETE、INSERT、UPDATE、REFERENCES、RULE或TRIGGER
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO
with_hierarchy	character_data	对应 PostgreSQL 不支持的特性

31.24. role_usage_grants

视图role_usage_grants
旨在标识授予给一个当前启用的角色或者被一个当前启用的角色授予的在多种对象上的USAGE权限。在PostgreSQL中，这目前只适用于域，而且由于域在PostgreSQL中没有真正的权限，这个视图是空的。更多信息可以在usage_privileges中找到。将来，这个视图可能会包含更多有用的信息。

表 31.22. role_usage_grants 列

名称	数据类型	描述
grantor	sql_identifier	将来：授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
object_catalog	sql_identifier	包含该对象的数据库名称（总是当前数据库）
object_schema	sql_identifier	包含该对象的模式名称
object_name	sql_identifier	对象的名称
object_type	character_data	将来：该对象的类型
privilege_type	character_data	总是USAGE
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO

31.25. routine_privileges

视图routine_privileges标识所有授予给当前用户或者由当前用户授予的函数权限。对于每一种函数、授予者和被授予者的组合，这里都有一行。授予给组的权限在视图role_routine_grants中标识。

表 31.23. routine_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	函数的“特定名称”。详见第 31.26 节。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)

名称	数据类型	描述
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）
privilege_type	character_data	总是为EXECUTE（函数唯一的权限类型）
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO

注意，grantee列不区分用户和组。如果你的用户和组同名，很遗憾无法区分它们。PostgreSQL的未来版本可能会禁止用户和组同名。

31.26. routines

视图routines包含当前数据库中所有的函数。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的函数才会被显示。

表 31.24. routines 列

名称	数据类型	描述
specific_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
specific_schema	sql_identifier	包含函数的模式的名称
specific_name	sql_identifier	该函数的“专用名”。这是一个在模式中唯一标识该函数的名称，即使该函数真正的名称已经被重载。专用名的格式尚未被定义，它应当仅被用来与指定例程名称的其他实例进行比较。
routine_catalog	sql_identifier	包含该函数的数据库名称(总是当前数据库)
routine_schema	sql_identifier	包含函数的模式的名称
routine_name	sql_identifier	该函数的名称（在重载的情况下可能会重复）
routine_type	character_data	总是FUNCTION（将来可能会有其他类型的例程。）
module_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
module_schema	sql_identifier	对应 PostgreSQL 不支持的特性
module_name	sql_identifier	对应 PostgreSQL 不支持的特性
udt_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
udt_schema	sql_identifier	对应 PostgreSQL 不支持的特性
udt_name	sql_identifier	对应 PostgreSQL 不支持的特性
data_type	character_data	该函数的返回数据类型如果是一种内置类型，这里是该数据类型；如果是某种数组（此种情况见视图element_types），则为ARRAY；否则为USER-DEFINED（此种情况中，该类型被标识在type_udt_name和相关列中）。

名称	数据类型	描述
character_maximum_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_octet_length	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
collation_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
collation_schema	sql_identifier	对应 PostgreSQL 不支持的特性
collation_name	sql_identifier	对应 PostgreSQL 不支持的特性
numeric_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_precision_radix	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
numeric_scale	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
datetime_precision	cardinal_number	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_type	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
interval_precision	character_data	总是为空，因为这种信息不适用于PostgreSQL中的返回数据类型
type_udt_catalog	sql_identifier	该函数的返回数据类型所在的数据库名（总是当前数据库）。
type_udt_schema	sql_identifier	该函数的返回数据类型所在的模式名。
type_udt_name	sql_identifier	该函数的返回数据类型的名字。
scope_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
scope_schema	sql_identifier	对应 PostgreSQL 不支持的特性
scope_name	sql_identifier	对应 PostgreSQL 不支持的特性
maximum_cardinality	cardinal_number	总是为空，因为 PostgreSQL 中数组的最大基数始终不受限制
dtd_identifier	sql_identifier	该函数返回数据类型的数据类型描述符的一个标识符，在从属于该函数的数据类型标识符之中唯一。这主要用于与这类标识符的其他实例进行连接（该标识符的指定格式没有被定义并且不保证在未来的版本中保持相同）。
routine_body	character_data	如果该函数是一个 SQL 函数，则为SQL，否则为EXTERNAL。
routine_definition	character_data	函数的源文本（如果当前用户不是该函数的属主，则为空值）（根据SQL 标准，只有routine_body为SQL时这一列才适用。但是在PostgreSQL中，它将会包含该函数被创建时所指定的任何语言。

名称	数据类型	描述
external_name	character_data	如果这个函数是一个 C 函数，则为该函数的外部名称（链接符号），否则为空（这会产生和显示在routine_definition中相同的值）。
external_language	character_data	该函数所用的语言
parameter_style	character_data	总是GENERAL（SQL 标准定义了其他参数风格，但在PostgreSQL中不可用）
is_deterministic	character_data	如果该函数被声明为不变（在 SQL 标准中被称为确定性的），则为YES，否则为NO（你不能通过该信息模式查询在PostgreSQL中可用的其他易变性级别）。
sql_data_access	character_data	总是MODIFIES，表示该函数可能修改 SQL 数据。这种信息对PostgreSQL没有用处。
is_null_call	character_data	如果该函数在任一参数为空时自动返回空值，则为YES，否则为NO。
sql_path	character_data	对应 PostgreSQL 不支持的特性
schema_level_routine	character_data	总是YES（反例是一个用户定义类型的方法，这是在PostgreSQL不可用的一种特性）。
max_dynamic_result_sets	cardinal_number	对应 PostgreSQL 不支持的特性
is_user_defined_cast	character_data	对应 PostgreSQL 不支持的特性
is_implicitly_invocable	character_data	对应 PostgreSQL 不支持的特性
security_type	character_data	如果该函数以当前用户的权限运行，则为INVOKER；如果该函数以定义它的用户的权限运行，则为DEFINER。
to_sql_specific_catalog	sql_identifier	对应 PostgreSQL 不支持的特性
to_sql_specific_schema	sql_identifier	对应 PostgreSQL 不支持的特性
to_sql_specific_name	sql_identifier	对应 PostgreSQL 不支持的特性
as_locator	character_data	对应 PostgreSQL 不支持的特性

31.27. schemata

视图schemata包含当前数据库中属于当前启用角色的所有模式。

表 31.25. schemata 列

名称	数据类型	描述
catalog_name	sql_identifier	该模式所在的数据库名称（总是当前数据库）
schema_name	sql_identifier	模式的名称
schema_owner	sql_identifier	模式拥有者的名称
default_character_set_catalog	sql_identifier	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
default_character_set_schema	sql_identifier	对应 PostgreSQL 不支持的特性
default_character_set_name	sql_identifier	对应 PostgreSQL 不支持的特性
sql_path	character_data	对应 PostgreSQL 不支持的特性

31.28. sql_features

表sql_features包含的信息指示了哪些 SQL 标准中定义的正式特性被PostgreSQL所支持。这和附录 D中的信息一样。这里你也能找到一些额外的背景信息。

表 31.26. sql_features 列

名称	数据类型	描述
feature_id	character_data	特性的标识符字符串
feature_name	character_data	特性的描述性名称
sub_feature_id	character_data	该子特性的标识符字符串，或者如果不是一个子特性则为一个长度为零的字符串
sub_feature_name	character_data	该子特性的描述性名称，或者如果不是一个子特性则为一个长度为零的字符串
is_supported	character_data	如果当前版本的 PostgreSQL 完全支持该特性，则为YES，否则为NO
is_verified_by	character_data	总是为空，因为PostgreSQL开发组没有对特性的一致性执行正式的测试
comments	character_data	可能会是关于该特性被支持状态的一段注释

31.29. sql_implementation_info

表sql_implementation_info包含的信息指示剩下的由 SQL 标准实现定义的多个方面。这类信息主要用来在 ODBC 接口的情境中使用；其它接口的用户可能将发现这类信息用处不大。由于这个原因，个体实现信息项没有在这里描述，你将会在 ODBC 接口的描述中找到它们。

表 31.27. sql_implementation_info 列

名称	数据类型	描述
implementation_info_id	character_data	实现信息项的标识符字符串
implementation_info_name	character_data	实现信息项的描述性名称
integer_value	cardinal_number	实现信息项的整数值；如果该值在列character_value中给出，则本列为空
character_value	character_data	实现信息项的字符值；如果该值在列integer_value中给出，则本列为空
comments	character_data	可能是从属于该实现信息项的一段注释

31.30. sql_languages

表sql_languages为PostgreSQL支持的每一种 SQL 语言绑定包含一行。PostgreSQL支持直接 SQL 和 C 中的嵌入式 SQL；从此表中只能了解到这些信息。

表 31.28. sql_languages 列

名称	数据类型	描述
sql_language_source	character_data	语言定义来源的名称；总是ISO 9075，即 SQL 标准
sql_language_year	character_data	sql_language_source引用的标准获批的年份；当前为2003
sql_language_conformance	character_data	该语言绑定的标准符合性级别。对于 ISO 9075:2003，此值总是CORE。
sql_language_integrity	character_data	总是为空（此值与 SQL 标准的较早版本有关。）
sql_language_implementation	character_data	总是为空
sql_language_binding_style	character_data	语言绑定方式，为DIRECT或EMBEDDED
sql_language_programming_language	character_data	如果绑定方式为EMBEDDED，则为所用编程语言，否则为空。PostgreSQL只支持 C 语言。

31.31. sql_packages

表sql_packages包含PostgreSQL支持 SQL 标准所定义的哪些功能包的信息。功能包的背景信息请参阅附录 D。

表 31.29. sql_packages 列

名称	数据类型	描述
feature_id	character_data	功能包的标识符字符串
feature_name	character_data	功能包的描述性名称
is_supported	character_data	如果当前版本的PostgreSQL完全支持此功能包，则为YES，否则为NO
is_verified_by	character_data	总是为空，因为PostgreSQL开发组没有对特性的一致性执行正式的测试
comments	character_data	关于此功能包支持状态的可选注释

31.32. sql_sizing

表sql_sizing包含有关PostgreSQL中多种尺寸限制和最大值的信息。这类信息主要用来在 ODBC 接口的情境中使用；其它接口的用户可能将发现这类信息用处不大。由于这个原因，个体实现信息项没有在这里描述，你将会在 ODBC 接口的描述中找到它们。

表 31.30. sql_sizing 列

名称	数据类型	描述
sizing_id	cardinal_number	尺寸项的标识符
sizing_name	character_data	尺寸项的描述性名称
supported_value	cardinal_number	尺寸项的值，如果尺寸是不受限制或不能确定的则为 0，如果尺寸项适用的特性不受支持则为空
comments	character_data	可能是从属于尺寸项的一段注释

31.33. sql_sizing_profiles

表sql_sizing_profiles包含 SQL 标准的各类配置规范所要求的sql_sizing值的信息。PostgreSQL不跟踪任何 SQL 配置规范，因此此表为空。

表 31.31. sql_sizing_profiles 列

名称	数据类型	描述
sizing_id	cardinal_number	尺寸项的标识符
sizing_name	character_data	尺寸项的描述性名称
profile_id	character_data	配置规范的标识符字符串
required_value	cardinal_number	SQL 配置规范对该尺寸限制项要求的值；如果配置规范对此项不设限制，则为 0；如果配置规范不要求此项所适用的任何功能，则为空
comments	character_data	关于此配置规范中该尺寸限制项的可选注释

31.34. table_constraints

视图table_constraints包含当前用户拥有或具有某些权限的表的所有约束。

表 31.32. table_constraints 列

名称	数据类型	描述
constraint_catalog	sql_identifier	包含该约束的数据库名称（总是当前数据库）
constraint_schema	sql_identifier	包含该约束的模式名称
constraint_name	sql_identifier	约束的名称
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
constraint_type	character_data	约束类型：CHECK（包括非空约束）、FOREIGN KEY、PRIMARY KEY 或 UNIQUE
is_deferrable	character_data	如果该约束是可延迟的，则为 YES，否则为NO

名称	数据类型	描述
initially_deferred	character_data	如果该约束是可延迟的且初始就被延迟，则为YES，否则为NO

31.35. table_privileges

视图table_privileges标识所有在表或视图上授予给当前用户或者由当前用户授予的权限。对每一个表、授予者和被授予者的组合都有一行。授予给组的权限在视图role_table_grants中标识。

表 31.33. table_privileges 列

名称	数据类型	描述
grantor	sql_identifier	授予该权限的角色名称
grantee	sql_identifier	将来：该权限被授予到的角色名称
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
privilege_type	character_data	该权限的类型：SELECT、DELETE、INSERT、UPDATE、REFERENCES、RULE或TRIGGER
is_grantable	character_data	如果该权限是可授予的，则为YES，否则为NO
with_hierarchy	character_data	对应 PostgreSQL 不支持的特性

注意，grantee列不区分用户和组。如果你的用户和组同名，很遗憾无法区分它们。PostgreSQL的未来版本可能会禁止用户和组同名。

31.36. tables

视图tables包含定义在当前数据库中的所有表和视图。
只有那些当前用户能够访问（作为拥有者或具有某些权限）的表和视图才会被显示。

表 31.34. tables 列

名称	数据类型	描述
table_catalog	sql_identifier	包含该表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该表的模式名称
table_name	sql_identifier	表的名称
table_type	character_data	Type of the table: BASE TABLE for a persistent base table (the normal table type), VIEW for a view, or LOCAL TEMPORARY for a temporary table
self_referencing_column_name	sql_identifier	对应 PostgreSQL 不支持的特性
reference_generation	character_data	对应 PostgreSQL 不支持的特性

名称	数据类型	描述
user_defined_type_catalog	sql_identifier	对应 PostgreSQL 中不可用的特性
user_defined_type_schema	sql_identifier	对应 PostgreSQL 中不可用的特性
user_defined_type_name	sql_identifier	对应 PostgreSQL 中不可用的特性

31.37. triggers

视图triggers包含当前数据库中当前用户拥有的表上定义的所有触发器。（表的属主就是触发器的属主。）

表 31.35. triggers 列

名称	数据类型	描述
trigger_catalog	sql_identifier	包含该触发器的数据库名称（总是当前数据库）
trigger_schema	sql_identifier	包含该触发器的模式名称
trigger_name	sql_identifier	触发器的名称
event_manipulation	character_data	触发该触发器的事件（INSERT、UPDATE或DELETE）
event_object_catalog	sql_identifier	包含触发器所在的表的数据库名称（总是当前数据库）
event_object_schema	sql_identifier	包含触发器所在的表的模式名称
event_object_table	sql_identifier	触发器所在的表的名称
action_order	cardinal_number	尚未实现
action_condition	character_data	对应 PostgreSQL 中不可用的特性
action_statement	character_data	触发器执行的语句（当前总是EXECUTE PROCEDURE *function*(...)）
action_orientation	character_data	标识触发器是对每个被处理的行触发一次还是为每个语句触发一次（ROW或STATEMENT）
condition_timing	character_data	触发器触发的时间（BEFORE或AFTER）
condition_reference_old_table	sql_identifier	对应 PostgreSQL 不支持的特性
condition_reference_new_table	sql_identifier	对应 PostgreSQL 不支持的特性

PostgreSQL中的触发器有两点与 SQL 标准不兼容，这会影响在该信息模式中的表示。第一，在PostgreSQL中触发器的名字是局限于每个表的，而不是独立于模式对象。因此可能在一个模式中会有重复的触发器名称，只要它们属于不同的表（trigger_catalog和trigger_schema才真正标识了触发器被定义在哪个表上）。第二，在PostgreSQL中触发器可以被定义为在多个事件上触发（例如ON INSERT OR UPDATE），而在SQL 标准中只允许一个。如果一个触发器被定义为在多个事件上触发，它在信息模式中被表示为多行，每一行对应于一类事件。作为这两个问题的结果，视图triggers的主键实际上是(trigger_catalog, trigger_schema, trigger_name, event_object_table,

event_manipulation)，而不是(trigger_catalog, trigger_schema, trigger_name)（这是 SQL 标准指定的）。尽管如此，如果你以符合 SQL 标准（在模式中触发器名称唯一并且每个触发器只能有一种事件类型）的方式定义你的触发器，这将不会影响你。

31.38. usage_privileges

视图usage_privileges
旨在标识授予给一个当前启用的角色或者被一个当前启用的角色授予的在多种对象上的USAGE权限。在PostgreSQL中，这目前只适用于域，而且由于域在PostgreSQL中没有真正的权限，这个视图对所有域显示隐式授予给PUBLIC的USAGE权限。将来，这个视图可能会包含更多有用的信息。

表 31.36. usage_privileges 列

名称	数据类型	描述
grantor	sql_identifier	当前被设置为该对象拥有者的名称
grantee	sql_identifier	当前总是PUBLIC
object_catalog	sql_identifier	包含该对象的数据库的名称（总是当前数据库）
object_schema	sql_identifier	包含该对象的模式名称
object_name	sql_identifier	该对象的名称
object_type	character_data	当前总是DOMAIN
privilege_type	character_data	总是USAGE
is_grantable	character_data	当前总是NO

31.39. view_column_usage

视图view_column_usage标识视图查询表达式（即定义该视图的 SELECT 语句）中使用的所有列。只有当包含该列的表归当前已启用角色所有时，该列才会出现在这个视图中。

> **注意**
>
> 系统表列不被包括。在某个时候这应该会被修复。

表 31.37. view_column_usage 列

名称	数据类型	描述
view_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
view_schema	sql_identifier	包含该视图的模式名称
view_name	sql_identifier	视图的名称
table_catalog	sql_identifier	被视图所使用的列所属表的数据库名称（总是当前数据库）
table_schema	sql_identifier	被视图所使用的列所属表的模式名称
table_name	sql_identifier	被视图所使用的列所属表的名称

名称	数据类型	描述
column_name	sql_identifier	被该视图所使用的列名称

31.40. view_table_usage

视图view_table_usage标识视图查询表达式（即定义该视图的 SELECT 语句）中使用的所有表。只有归当前已启用角色所有的表才会出现在这个视图中。

> **注意**
>
> 系统表没有被包括。这应当会在某个时候被修复。

表 31.38. view_table_usage 列

名称	数据类型	描述
view_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
view_schema	sql_identifier	包含该视图的模式名称
view_name	sql_identifier	视图的名称
table_catalog	sql_identifier	包含被该视图所使用的表的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含被该视图所使用的表的模式名称
table_name	sql_identifier	包含被该视图所使用的表的名称

31.41. views

视图views包含定义在当前数据库中的所有视图。
只有当前用户能够访问（作为拥有者或具有某些权限）的视图才会被显示。

表 31.39. views 列

名称	数据类型	描述
table_catalog	sql_identifier	包含该视图的数据库名称（总是当前数据库）
table_schema	sql_identifier	包含该视图的模式名称
table_name	sql_identifier	视图的名称
view_definition	character_data	定义视图的查询表达式（如果当前用户不是该视图的属主，则为空值）
check_option	character_data	对应 PostgreSQL 不支持的特性
is_updatable	character_data	尚未实现
is_insertable_into	character_data	尚未实现

部分 V. 服务器编程

本部分介绍如何使用用户定义函数、数据类型、触发器等扩展服务器功能。这些属于高级主题，
通常应在读者已经理解其他有关PostgreSQL的用户文档之后再来学习。
本部分后面的章节介绍PostgreSQL发行版中提供的服务器端编程语言，
以及与服务器端编程有关的一般问题。在深入学习服务器端编程的内容之前，至少应先阅读第 32 章的前几节（涵盖函数）。

目录

32. 扩展SQL 452
32.1. 可扩展性如何运作 452
32.2. PostgreSQL 类型系统 452
32.2.1. 基础类型 452
32.2.2. 复合类型 452
32.2.3. 域 452
32.2.4. 伪类型 453
32.2.5. 多态类型 453
32.3. 用户定义的函数 453
32.4. 查询语言（SQL）函数 454
32.4.1. 基本类型上的 SQL 函数 454
32.4.2. 组合类型上的 SQL 函数 455
32.4.3. 带输出参数的函数 458
32.4.4. 作为表源的 SQL 函数 459
32.4.5. 返回集合的 SQL 函数 460
32.4.6. 多态 SQL 函数 461
32.5. 函数重载 462
32.6. 函数易变性分类 462
32.7. 过程语言函数 464
32.8. 内部函数 464
32.9. C 语言函数 464
32.9.1. 动态载入 464
32.9.2. C 语言函数中的基础类型 465
32.9.3. C 函数的调用约定版本 0 468
32.9.4. C 函数的调用约定版本 1 470
32.9.5. 编写代码 472
32.9.6. Compiling and Linking Dynamically-Loaded Functions 473
32.9.7. 扩展构建基础设施 475
32.9.8. C 函数中的复合类型参数 477
32.9.9. 从 C 函数返回行（复合类型） 478
32.9.10. 从 C 函数返回集合 480
32.9.11. 多态参数和返回类型 485
32.10. 用户定义的聚合 486
32.11. 用户定义类型 488
32.12. 用户定义操作符 491
32.13. 操作符优化信息 492
32.13.1. COMMUTATOR 492
32.13.2. NEGATOR 493
32.13.3. RESTRICT 493
32.13.4. JOIN 494
32.13.5. HASHES 494
32.13.6. MERGES (SORT1, SORT2, LTCMP, GTCMP) 495
32.14. 索引扩展接口 496
32.14.1. 索引方法和操作符类 496
32.14.2. 索引方法策略 497
32.14.3. 索引方法支持例程 498
32.14.4. 示例 499
32.14.5. 跨数据类型的操作符类 501
32.14.6. 系统对操作符类的依赖 502
32.14.7. 操作符类的特殊特性 502
33. 触发器 504

33.1. 触发器行为概述 504
33.2. 数据更改的可见性 505
33.3. 用 C 编写触发器函数 505
33.4. 一个完整的示例 508
34. 规则系统 512
34.1. 查询树 512
34.2. 视图和规则系统 513
34.2.1. SELECT 规则如何工作 514
34.2.2. 非 SELECT 语句中的视图规则 518
34.2.3. PostgreSQL 中视图的能力 519
34.2.4. 更新视图 520
34.3. INSERT、UPDATE 和 DELETE 上的规则 520
34.3.1. 更新规则如何工作 520
34.3.2. 与视图的协作 524
34.4. 规则和权限 529
34.5. 规则和命令状态 531
34.6. 规则与触发器 531
35. Procedural Languages 534
35.1. 安装过程语言 534
36. PL/pgSQL - SQL 过程语言 536
36.1. 概述 536
36.1.1. 使用PL/pgSQL的优点 537
36.1.2. 支持的参数和结果数据类型 537
36.2. PL/pgSQL 开发提示 538
36.2.1. 引号的处理 538
36.3. PL/pgSQL的结构 540
36.4. 声明 541
36.4.1. 函数参数的别名 541
36.4.2. 复制类型 543
36.4.3. 行类型 544
36.4.4. 记录类型 544
36.4.5. RENAME 544
36.5. 表达式 545
36.6. 基本语句 546
36.6.1. 赋值 546
36.6.2. SELECT INTO 546
36.6.3. 执行没有结果的表达式或查询 547
36.6.4. 什么也不做 547
36.6.5. 执行动态命令 548
36.6.6. 获取结果状态 549
36.7. 控制结构 550
36.7.1. 从函数返回 550
36.7.2. 条件语句 551
36.7.3. 简单循环 553
36.7.4. 遍历查询结果 555
36.7.5. 捕获错误 556
36.8. 游标 558
36.8.1. 声明游标变量 558
36.8.2. 打开游标 559
36.8.3. 使用游标 560
36.9. 错误和消息 562
36.10. 触发器函数 562
36.11. 从Oracle PL/SQL 移植 568

36.11.1. 移植示例 568
36.11.2. 其他要关注的事项 574
36.11.3. 附录 574
37. PL/Tcl - Tcl 过程语言 578
37.1. 概述 578
37.2. PL/Tcl 函数和参数 578
37.3. PL/Tcl 中的数据值 579
37.4. PL/Tcl 中的全局数据 579
37.5. 从 PL/Tcl 访问数据库 580
37.6. PL/Tcl 中的触发器函数 582
37.7. 模块与`unknown`命令 583
37.8. Tcl 过程名 584
38. PL/Perl - Perl 过程语言 585
38.1. PL/Perl 函数和参数 585
38.2. 从 PL/Perl 访问数据库 588
38.3. PL/Perl 中的数据值 590
38.4. PL/Perl 中的全局值 590
38.5. 受信任与不受信任的 PL/Perl 591
38.6. PL/Perl 触发器 592
38.7. 限制与缺失特性 593
39. PL/Python - Python 过程语言 594
39.1. PL/Python 函数 594
39.2. 触发器函数 595
39.3. 数据库访问 595
40. 服务器编程接口 597
40.1. 接口函数 597
40.2. 接口支持函数 618
40.3. 内存管理 626
40.4. 数据更改的可见性 635
40.5. 示例 635

第 32 章 扩展SQL

在接下来的各节中，我们将讨论如何通过增加以下内容来扩展 PostgreSQL SQL 查询语言：

- 函数（从第 32.3 节开始）
- 聚合（从第 32.10 节开始）
- 数据类型（从第 32.11 节开始）
- 操作符（从第 32.12 节开始）
- 用于索引的操作符类（从第 32.14 节开始）

32.1. 可扩展性如何运作

PostgreSQL 之所以具有可扩展性，是因为其运作由系统目录驱动。如果你熟悉标准的关系数据库系统，就会知道它们把有关数据库、表、
列等的信息存储在通常所说的系统目录中（有些系统把这称为数据字典）。
这些目录对用户而言看起来就像普通表一样，但 DBMS 会在其中保存自己的内部管理信息。PostgreSQL 与标准关系数据库系统的一个关键差别是，PostgreSQL 在目录中存储的信息要多得多：不仅有关于表和列的信息，还有关于数据类型、函数、访问方法等的信息。
这些表可以由用户修改，而 PostgreSQL 又是基于这些表来运行的，这意味着 PostgreSQL 可以由用户扩展。相比之下，传统数据库系统只能通过修改源代码中的硬编码过程，或加载由 DBMS 供应商专门编写的模块来扩展。

此外，PostgreSQL 服务器还能通过动态加载把用户编写的代码纳入自身。也就是说，用户可以指定一个实现了新类型或新函数的目标代码文件（例如共享库），而 PostgreSQL 会在需要时加载它。把用 SQL 编写的代码加入服务器就更为简单了。这种能够“即时”修改自身行为的能力，使 PostgreSQL 特别适合用于新应用和新存储结构的快速原型设计。

32.2. PostgreSQL 类型系统

PostgreSQL 数据类型可以分为基础类型、复合类型、域和伪类型。

32.2.1. 基础类型

基础类型是在 SQL 语言层之下实现的类型（通常使用 C 等低级语言），例如 `int4`。它们通常对应于所谓的抽象数据类型。PostgreSQL 只能通过用户提供的函数操作这些类型，对其行为的理解也仅限于用户所描述的内容。基础类型进一步分为标量类型和数组类型。对于每个标量类型，都会自动创建相应的数组类型，以存储由该标量类型组成的可变大小数组。

32.2.2. 复合类型

用户每次创建表时，都会创建复合类型（也称行类型）。还可以使用CREATE TYPE定义不关联任何表的“独立”复合类型。复合类型就是一个带有对应字段名的类型列表。
复合类型的值是由字段值组成的一行或一条记录。用户可以在 SQL 查询中访问各个组成字段。关于复合类型的更多信息，参见第 8.11 节。

32.2.3. 域

域基于某一种特定的基础类型，并且在很多场合下都可以与其基础类型互换。不过，
域可以带有约束，把其合法值限制在底层基础类型所允许值的一个子集内。

可以使用 SQL 命令CREATE DOMAIN创建域。本章不讨论域的创建和使用。

32.2.4. 伪类型

有少数几种“伪类型”用于特殊目的。伪类型不能作为表的列或复合类型的属性出现，但它们可以用来声明函数的参数类型和结果类型。这为在类型系统内部识别特殊类别的函数提供了一种机制。表 8.20列出了现有的伪类型。

32.2.5. 多态类型

两种特别值得关注的伪类型是 `anyelement` 和 `anyarray`，它们统称为多态类型。使用这些类型声明的函数称为多态函数。多态函数可以操作多种不同的数据类型，具体类型由某次调用时实际传入的数据类型决定。

多态参数和结果相互关联，在解析调用多态函数的查询时，会被解析为某种具体的数据类型。每个声明为 `anyelement` 的位置（参数或返回值）都可以具有任意具体的实际数据类型，但在同一次调用中，它们必须具有相同的实际类型。每个声明为 `anyarray` 的位置可以具有任意数组数据类型，但同样必须全部具有相同的类型。此外，如果某些位置声明为 `anyarray`，另一些位置声明为 `anyelement`，那么 `anyarray` 位置上的实际数组类型，其元素类型必须与 `anyelement` 位置上的类型相同。

因此，当多个参数位置被声明为多态类型时，其总体效果就是只允许某些实际参数类型组合。例如，声明为 `equal(anyelement, anyelement)` 的函数可以接受任意两个输入值，只要它们属于同一数据类型。

如果函数的返回值声明为多态类型，那么至少有一个参数位置也必须是多态的，并且实际传入的参数数据类型决定该次调用的实际结果类型。例如，如果尚不存在数组下标机制，可以将实现下标访问的函数定义为 `subscript(anyarray, integer) returns anyelement`。这一声明将实际的第一个参数约束为数组类型，并允许解析器根据第一个参数的实际类型推导正确的结果类型。

32.3. 用户定义的函数

PostgreSQL提供四种函数：

- 查询语言函数（用SQL编写的函数）（第 32.4 节）
- 过程语言函数（例如，用PL/pgSQL或PL/Tcl编写的函数）（第 32.7 节）
- 内部函数（第 32.8 节）
- C 语言函数（第 32.9 节）

每一类函数可以采用基础类型、复合类型或者它们的组合作为参数。此外，每一类函数可以返回一个基础类型或一个复合类型。函数也能被定义成返回基础类型或复合类型值的集合。

很多类函数可以接受或者返回特定的伪类型（例如，多态类型），但是可用的功能因函数类别而异。详情可以参考每一种函数的描述。

定义SQL函数最容易，因此我们将从讨论它们开始。大部分SQL函数的概念也能用到其他类型的函数上。

在这一章中，查看CREATE FUNCTION命令的参考页有助于更好地理解示例。这章中的一些示例可以在PostgreSQL源代码发布的`src/tutorial`目录中的`funcs.sql`和`funcs.c`中找到。

32.4. 查询语言（SQL）函数

SQL 函数执行由任意 SQL 语句组成的一个列表，并返回列表中最后一个查询的结果。在简单（非集合）情况下，将返回最后一个查询结果的第一行。（请记住，除非使用 ORDER BY，否则多行结果中的“第一行”并没有明确定义。）如果最后一个查询完全没有返回任何行，则返回空值。

或者，一个 SQL 函数可以通过指定函数的返回类型为SETOF *sometype*被声明为返回一个集合。在这种情况下，最后一个查询的结果的所有行会被返回。下文将给出进一步的细节。

SQL 函数的函数体必须是一个由分号分隔的 SQL 语句列表。最后一条语句后的分号是可选的。除非该函数被声明为返回 void，否则最后一条语句必须是 SELECT。

SQL 语言中的任何命令集合都可以打包在一起并定义为函数。除了 SELECT 查询之外，这些命令还可以包括数据修改查询（INSERT、UPDATE 和 DELETE），以及其他 SQL 命令。（唯一的例外是不能把 BEGIN、COMMIT、ROLLBACK 或 SAVEPOINT 命令放进 SQL 函数。）不过，最后一条命令必须是 SELECT，并且返回与函数声明返回类型相符的结果。或者，如果你想定义一个执行动作但没有有用返回值的 SQL 函数，也可以把它定义为返回 void。在这种情况下，函数体不能以 SELECT 结尾。例如，下面这个函数会删除 emp 表中薪资为负的行：

```
CREATE FUNCTION clean_emp() RETURNS void AS '
    DELETE FROM emp
        WHERE salary < 0;
' LANGUAGE SQL;

SELECT clean_emp();

 clean_emp
-----------

(1 row)
```

CREATE FUNCTION 命令的语法要求把函数体写成一个字符串常量。对于字符串常量，通常使用美元引用最方便（见第 4.1.2.2 节）。如果你选择使用常规的单引号字符串常量语法，那么必须对函数体中使用的单引号（'）和反斜线（\）进行转义，通常是把它们写成双份（见第 4.1.2.1 节）。

SQL 函数的参数在函数体中使用语法 $*n* 引用：$1 引用第一个参数，$2 引用第二个参数，依此类推。如果参数是组合类型，则可以用点号记法（例如 $1.name）访问该参数的属性。这些参数只能用作数据值，而不能用作标识符。因此，举例来说，这样写是合理的：

```
INSERT INTO mytable VALUES ($1);
```

但这样写行不通：

```
INSERT INTO $1 VALUES (42);
```

32.4.1. 基本类型上的 SQL 函数

最简单的 SQL 函数没有参数，只是返回一个基础类型，例如 integer：

```
CREATE FUNCTION one() RETURNS integer AS $$
    SELECT 1 AS result;
$$ LANGUAGE SQL;
```

```
-- Alternative syntax for string literal:
CREATE FUNCTION one() RETURNS integer AS '
    SELECT 1 AS result;
' LANGUAGE SQL;

SELECT one();

 one
-----
   1
```

注意我们为该函数的结果在函数体内定义了一个列别名（名为result），
但是这个列别名在函数以外是不可见的。因此，结果被标记为one而不是result。

定义接受基本类型作为参数的 SQL 函数几乎同样容易。在下面的示例中，注意我们如何在函数中把参数引用为 $1 和 $2。

```
CREATE FUNCTION add_em(integer, integer) RETURNS integer AS $$
    SELECT $1 + $2;
$$ LANGUAGE SQL;

SELECT add_em(1, 2) AS answer;

 answer
--------
      3
```

下面是一个更有用的函数，它可以用来从银行账户中扣款：

```
CREATE FUNCTION tf1 (integer, numeric) RETURNS integer AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1;
    SELECT 1;
$$ LANGUAGE SQL;
```

用户可以这样执行该函数，从 17 号账户中扣除 $100.00：

```
SELECT tf1(17, 100.0);
```

实际上我们可能希望从该函数得到比常数 1 更有用的结果，因此一个更可能的定义是：

```
CREATE FUNCTION tf1 (integer, numeric) RETURNS numeric AS $$
    UPDATE bank
        SET balance = balance - $2
        WHERE accountno = $1;
    SELECT balance FROM bank WHERE accountno = $1;
$$ LANGUAGE SQL;
```

它会调整余额并返回新余额。

32.4.2. 组合类型上的 SQL 函数

在编写参数为组合类型的函数时，我们必须不仅指定想要哪个参数（如上面用 $1 和 $2 那样），还要指定该参数的所需属性（字段）。例如，假设 emp 是一个包含员工数据的表，因而也是该表每一行的组合类型的名称。下面的函数 double_salary 计算某人薪资翻倍后会是多少：

```
CREATE TABLE emp (
    name        text,
    salary      numeric,
    age         integer,
    cubicle     point
);

INSERT INTO emp VALUES ('Bill', 4200, 45, '(2,1)');

CREATE FUNCTION double_salary(emp) RETURNS numeric AS $$
    SELECT $1.salary * 2 AS salary;
$$ LANGUAGE SQL;

SELECT name, double_salary(emp.*) AS dream
    FROM emp
    WHERE emp.cubicle ~= point '(2,1)';

 name | dream
------+-------
 Bill |  8400
```

注意这里用 `$1.salary` 语法来选取参数行值中的一个字段。还要注意，调用时的 `SELECT` 命令使用 `*` 将表的当前整行取作一个复合值。该表行也可以仅用表名来引用：

```
SELECT name, double_salary(emp) AS dream
    FROM emp
    WHERE emp.cubicle ~= point '(2,1)';
```

但这种用法已被弃用，因为它很容易让人混淆。

有时候即时构造一个复合参数值会很方便。这可以用`ROW`构造器完成。例如，
我们可以调整被传递给函数的数据：

```
SELECT name, double_salary(ROW(name, salary*1.1, age, cubicle)) AS
 dream
    FROM emp;
```

也可以构建一个返回复合类型的函数。这是一个返回单一`emp`行的函数示例：

```
CREATE FUNCTION new_emp() RETURNS emp AS $$
    SELECT text 'None' AS name,
        1000.0 AS salary,
        25 AS age,
        point '(2,2)' AS cubicle;
$$ LANGUAGE SQL;
```

在这个示例中，我们为每一个属性指定了一个常量值，但是可以用任何计算来替换这些常量。

定义该函数时有两点重要注意事项：

- 查询中的选择列表顺序必须与列在该复合类型所关联的表中出现的顺序完全相同。（系统不考虑像上面这样指定的列名。）
- 你必须对表达式进行类型转换，使之与复合类型的定义匹配，否则会得到这样的错误：

  ```
  ERROR:  function declared to return emp returns varchar instead of
   text at column 1
  ```

定义同一个函数的另一种方法是：

```
CREATE FUNCTION new_emp() RETURNS emp AS $$
    SELECT ROW('None', 1000.0, 25, '(2,2)')::emp;
$$ LANGUAGE SQL;
```

这里我们编写了一个 `SELECT`，它只返回具有正确复合类型的单列。在这种情况下，这样做并没有更好，但在某些情况下它是一种方便的替代方法 – 例如，如果我们需要通过调用另一个返回所需复合值的函数来计算结果。

我们可以用两种方式之一直接调用这个函数：

```
SELECT new_emp();

         new_emp
--------------------------
 (None,1000.0,25,"(2,2)")

SELECT * FROM new_emp();

 name | salary | age | cubicle
------+--------+-----+---------
 None | 1000.0 |  25 | (2,2)
```

第二种方式在第 32.4.4 节中有更完整的描述。

当你使用返回复合类型的函数时，可能只需要取其结果中的一个字段（属性）。
可以使用下面这样的语法：

```
SELECT (new_emp()).name;

 name
------
 None
```

这里需要额外的括号，以免解析器产生歧义。如果不加括号，结果会是这样：

```
SELECT new_emp().name;
ERROR:  syntax error at or near "." at character 17
LINE 1: SELECT new_emp().name;
                         ^
```

另一种选择是使用函数记法来提取属性。简单的解释是：记法 `attribute(table)` 和 `table.attribute` 可以互换使用。

```
SELECT name(new_emp());

 name
------
 None

-- This is the same as:
-- SELECT emp.name AS youngster FROM emp WHERE emp.age < 30;

SELECT name(emp) AS youngster FROM emp WHERE age(emp) < 30;
```

```
 youngster
-----------
 Sam
 Andy
```

> **提示**
>
> 函数记法与属性记法之间的等价性，使得可以用组合类型上的函数来模拟“计算字段”。例如，利用前面定义的 `double_salary(emp)`，我们可以写
>
> ```
> SELECT emp.name, emp.double_salary FROM emp;
> ```
>
> 使用这种写法的应用不需要直接了解 `double_salary` 并不是表的真实列。（也可以用视图来模拟计算字段。）

另一种使用函数返回复合类型的方法是将结果传递给另一个接受正确行类型作为输入的函数：

```
CREATE FUNCTION getname(emp) RETURNS text AS $$
    SELECT $1.name;
$$ LANGUAGE SQL;

SELECT getname(new_emp());
 getname
---------
 None
(1 row)
```

使用返回复合类型的函数的又一种方式，是把它作为表函数调用，如第 32.4.4 节所述。

32.4.3. 带输出参数的函数

描述函数结果的另一种方式是用输出参数来定义它，如下例所示：

```
CREATE FUNCTION add_em (IN x int, IN y int, OUT sum int)
AS 'SELECT $1 + $2'
LANGUAGE SQL;

SELECT add_em(3,7);
 add_em
--------
     10
(1 row)
```

这与第 32.4.1 节中所示的 `add_em` 版本没有本质区别。输出参数的真正价值在于，它们为定义返回若干列的函数提供了一种便捷方式。例如，

```
CREATE FUNCTION sum_n_product (x int, y int, OUT sum int, OUT product
 int)
AS 'SELECT $1 + $2, $1 * $2'
LANGUAGE SQL;
```

```
 SELECT * FROM sum_n_product(11,42);
 sum | product
-----+---------
  53 |     462
(1 row)
```

这里本质上发生的事情是：我们为该函数的结果创建了一个匿名复合类型。
上面的示例与下面这种写法的最终效果相同：

```
CREATE TYPE sum_prod AS (sum int, product int);

CREATE FUNCTION sum_n_product (int, int) RETURNS sum_prod
AS 'SELECT $1 + $2, $1 * $2'
LANGUAGE SQL;
```

但通常不必额外定义一个独立的复合类型会更方便。注意，
附在输出参数上的名称并非只是装饰，它们决定了这个匿名复合类型的列名。
（如果省略输出参数名称，系统会自行选择一个名称。）

在从 SQL 调用这样一个函数时，输出参数不会被包括在调用参数列表中。这是因为PostgreSQL只考虑输入参数来定义函数的调用签名。
这也意味着在为诸如删除函数等目的引用该函数时只有输入参数有关系。
我们可以用下面的命令之一删除上述函数

```
DROP FUNCTION sum_n_product (x int, y int, OUT sum int, OUT product
 int);
DROP FUNCTION sum_n_product (int, int);
```

参数可以被标记为`IN`（默认）、`OUT`或`INOUT`。`INOUT`
参数既充当输入参数（调用参数列表的一部分），又充当输出参数（结果记录类型的一部分）。

32.4.4. 作为表源的 SQL 函数

所有的 SQL 函数都可以被用在查询的`FROM`子句中，但是对于返回复合类型的函数特别有用。如果函数被定义为返回一种基础类型，该表函数会产生一个单列表。
如果该函数被定义为返回一种复合类型，该表函数会为该复合类型的每一个属性产生一列。

以下是一个示例：

```
CREATE TABLE foo (fooid int, foosubid int, fooname text);
INSERT INTO foo VALUES (1, 1, 'Joe');
INSERT INTO foo VALUES (1, 2, 'Ed');
INSERT INTO foo VALUES (2, 1, 'Mary');

CREATE FUNCTION getfoo(int) RETURNS foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT *, upper(fooname) FROM getfoo(1) AS t1;

 fooid | foosubid | fooname | upper
-------+----------+---------+-------
     1 |        1 | Joe     | JOE
```

```
(1 row)
```

如示例所示，我们可以像操作普通表的列一样操作函数结果中的列。

注意我们只从函数得到了一行。这是因为我们没有使用SETOF。这会在下一节中介绍。

32.4.5. 返回集合的 SQL 函数

当一个 SQL 函数被声明为返回SETOF *sometype*时，该函数的最后一个 SELECT 查询会被执行完，并且它输出的每一行都会被作为结果集的一个元素返回。

这一特性通常是在查询的 FROM 子句中调用函数时使用的。在这种情况下，函数返回的每一行都会成为查询所见表中的一行。例如，假设表 foo 与上文相同，我们写：

```
CREATE FUNCTION getfoo(int) RETURNS SETOF foo AS $$
    SELECT * FROM foo WHERE fooid = $1;
$$ LANGUAGE SQL;

SELECT * FROM getfoo(1) AS t1;
```

那么会得到：

```
 fooid | foosubid | fooname
-------+----------+---------
     1 |        1 | Joe
     1 |        2 | Ed
(2 rows)
```

当前，返回集合的函数也可以在查询的 select 列表中被调用。对于查询本身生成的每一行，该返回集合的函数都会被调用，并为其结果集中的每一个元素生成一行输出。但请注意，
这一能力已被弃用，未来的发行版中可能会将其移除。下面是一个从 select 列表返回集合的函数示例：

```
CREATE FUNCTION listchildren(text) RETURNS SETOF text AS $$
    SELECT name FROM nodes WHERE parent = $1
$$ LANGUAGE SQL;

SELECT * FROM nodes;
   name    | parent
-----------+--------
 Top       |
 Child1    | Top
 Child2    | Top
 Child3    | Top
 SubChild1 | Child1
 SubChild2 | Child1
(6 rows)

SELECT listchildren('Top');
 listchildren
--------------
 Child1
 Child2
 Child3
(3 rows)
```

```
SELECT name, listchildren(name) FROM nodes;
  name  | listchildren
--------+--------------
 Top    | Child1
 Top    | Child2
 Top    | Child3
 Child1 | SubChild1
 Child1 | SubChild2
(5 rows)
```

在最后一个 `SELECT` 中，注意 `Child2`、`Child3` 等没有对应的输出行。这是因为 `listchildren` 对这些参数返回空集，因此不生成任何结果行。

32.4.6. 多态 SQL 函数

SQL 函数可以声明为接受和返回多态类型 `anyelement` 和 `anyarray`。关于多态函数的更详细解释，见第 32.2.5 节。下面是一个多态函数 `make_array`，它由两个任意数据类型的元素构建一个数组：

```
CREATE FUNCTION make_array(anyelement, anyelement) RETURNS anyarray AS
 $$
    SELECT ARRAY[$1, $2];
$$ LANGUAGE SQL;

SELECT make_array(1, 2) AS intarray, make_array('a'::text, 'b') AS
 textarray;
 intarray | textarray
----------+-----------
 {1,2}    | {a,b}
(1 row)
```

注意类型转换 `'a'::text` 的使用是为了指定该参数的类型是 `text`。如果该参数只是一个字符串字面量，就必须这样做，否则它会被当作 `unknown` 类型，并且 `unknown` 的数组也不是一种合法的类型。如果没有该类型转换，将得到这样的错误：

```
ERROR:  could not determine "anyarray"/"anyelement" type because input
 has type "unknown"
```

允许存在具有固定返回类型的多态参数，但反过来则不允许。例如：

```
CREATE FUNCTION is_greater(anyelement, anyelement) RETURNS boolean AS
 $$
    SELECT $1 > $2;
$$ LANGUAGE SQL;

SELECT is_greater(1, 2);
 is_greater
------------
 f
(1 row)

CREATE FUNCTION invalid_func() RETURNS anyelement AS $$
    SELECT 1;
```

```
$$ LANGUAGE SQL;
ERROR:  cannot determine result data type
DETAIL:  A function returning "anyarray" or "anyelement" must have at
 least one argument of either type.
```

多态也可以和带输出参数的函数一起使用。例如：

```
CREATE FUNCTION dup (f1 anyelement, OUT f2 anyelement, OUT f3
 anyarray)
AS 'select $1, array[$1,$1]' LANGUAGE sql;

SELECT * FROM dup(22);
 f2 |   f3
----+---------
 22 | {22,22}
(1 row)
```

32.5. 函数重载

可以用同一个 SQL 名称定义多个函数，只要它们接受的参数不同即可。换句话说，函数名可以被重载。当一个查询被执行时，
服务器将根据所提供参数的数据类型和数量来决定要调用哪个函数。
重载也可用来模拟具有可变参数个数（最大个数有限）的函数。

在创建一个重载函数家族时，应该小心不要创建歧义。例如，给定函数：

```
CREATE FUNCTION test(int, real) RETURNS ...
CREATE FUNCTION test(smallint, double precision) RETURNS ...
```

对于`test(1, 1.5)`这样的输入就无法立刻清楚地知道应该调用哪个函数。当前实现的解决规则在第 10 章中有描述，但是设计一个依赖于这种行为的系统是不明智的。

具有单个复合类型参数的函数，通常不应与该类型的任何属性（字段）同名。回想一下，`attribute(table)`被视为等价于 `table.attribute`。如果在复合类型上的函数与复合类型的属性之间出现歧义，系统总会优先选择属性。可以通过给函数名加上模式限定（即 `schema.func(table)`）来覆盖这一选择，但更好的做法仍然是避免使用会冲突的名字。

在重载 C 语言函数时有一个额外的约束：重载函数家族中的每一个函数的 C 名称必须与其他所有函数的 C 名称不同，不管是内部的还是动态载入的。如果这条规则被违背，该行为将不可移植。你可能会得到一个运行时链接器错误，或者这些函数之一将被调用（通常是内部的那一个）。
SQL `CREATE FUNCTION`命令的`AS`子句的另一种形式可以把 SQL 函数名和 C 源代码中的函数名分离。例如：

```
CREATE FUNCTION test(int) RETURNS int
    AS 'filename', 'test_1arg'
    LANGUAGE C;
CREATE FUNCTION test(int, int) RETURNS int
    AS 'filename', 'test_2arg'
    LANGUAGE C;
```

这里的 C 函数名称反映了很多种可能的习惯之一。

32.6. 函数易变性分类

每一个函数都有一个易变性分类，可能的类别是 VOLATILE，STABLE 或 IMMUTABLE。VOLATILE 会作为默认类别，条件是CREATE FUNCTION命令没有指定类别。易变性分类是给优化器的关于该函数行为的一种承诺：

- 一个VOLATILE函数可以做任何事情，包括修改数据库。在使用相同的参数连续调用时，它能返回不同的结果。优化器不会对这类函数的行为做任何假定。使用易变函数的查询会在需要该函数值的每一行上重新对它求值。

- 一个STABLE函数不能修改数据库并且被确保对一个语句中的所有行用给定的相同参数返回相同的结果。这种分类允许优化器把该函数的多个调用优化成一个调用。特别是，在一个索引扫描条件中使用包含这样一个函数的表达式是安全的（因为一次索引扫描只会计算一次比较值，而不是为每一行都计算一次，在一个索引扫描条件中不能使用 VOLATILE函数）。

- 一个IMMUTABLE函数不能修改数据库并且被确保用相同的参数永远返回相同的结果。这种分类允许优化器在一个查询用常量参数调用该函数时提前计算该函数。例如，一个 SELECT ... WHERE x = 2 + 2这样的查询可以被简化为 SELECT ... WHERE x = 4，因为整数加法操作符底层的函数被标记为IMMUTABLE。

为了最好的优化结果，你应该把函数标记为对它们合法的易变性分类中最严格的那种。

任何带有副作用的函数必须被标记为VOLATILE，这样对它的调用就不能被优化掉。甚至如果一个函数的值在一个查询中会变化，即使它没有副作用也需要被标记为VOLATILE。这样的示例有random()、currval()、timeofday()等。

在考虑先规划然后立即执行的简单交互式查询时，在STABLE和 IMMUTABLE分类间的区别相对较小：一个函数是在规划时只执行一次还是在查询执行开始期间只执行一次没有太大关系。但是如果计划被保存下来然后在后面被重用，区别就大了。如果把一个实际上并非不可变的函数标记为IMMUTABLE，就可能在规划期间过早将它折叠成常量，导致在后续重用该计划时继续使用陈旧的值。在使用预备语句，或使用会缓存执行计划的函数语言（如 PL/pgSQL）时，这会带来严重问题。

由于 MVCC 的快照行为（见第 12 章），只包含SELECT命令的函数可以安全地标记为STABLE，即使它查询的表可能正被并发查询修改。PostgreSQL会使用为调用查询建立的快照来执行STABLE函数，因而该函数在整个查询期间看到的都是数据库的固定视图。还要注意current_timestamp家族的函数有资格被标记为稳定，因为它们的值在一个事务中不会改变。

对 IMMUTABLE 函数中的 SELECT 也采用同样的快照行为。一般来说，在 IMMUTABLE 函数里查询数据库表并不明智，因为一旦表内容发生变化，就会破坏其不变性。不过，PostgreSQL 并不会强制禁止这样做。

一种常见的错误是当一个函数的结果依赖于一个配置参数时把它标记为 IMMUTABLE。例如，一个操纵时间戳的函数有可能结果依赖于timezone设置。为了安全起见，这类函数应该被标记为 STABLE。

注意

在 PostgreSQL 8.0 版之前，STABLE 和 IMMUTABLE 函数不能修改数据库这项要求并未被系统强制执行。8.0 版通过要求这些类别的 SQL 函数和过程语言函数不包含 SELECT 之外的 SQL 命令来强制执行它。（这并不是一个完全万无一失的测试，因为这类函数仍可能调

用修改数据库的 VOLATILE 函数。如果你这样做，会发现 STABLE 或 IMMUTABLE 函数不会注意到被调用函数所应用的数据库更改。）

32.7. 过程语言函数

PostgreSQL允许用除 SQL 和 C 之外的语言编写用户定义的函数。这些语言通常被称为过程语言（PL）。过程语言并不内置在PostgreSQL服务器中，它们通过可装载模块提供。更多信息请见第 35 章以及接下来的章节。

32.8. 内部函数

内部函数由 C 编写并且已经被静态链接到 PostgreSQL 服务器中。该函数定义的“主体”指定该函数的 C 语言名称，它不必与为 SQL 使用而声明的名称相同。（出于向后兼容的原因，也接受空主体，此时会认为 C 语言函数名与 SQL 函数名相同。）

通常，所有存在于服务器中的内部函数都在数据库集簇的初始化（initdb）期间被声明，但是用户可以使用 CREATE FUNCTION为一个内部函数创建额外的别名。在CREATE FUNCTION中用语言名internal来声明内部函数。例如，要为 sqrt函数创建一个别名：

```
CREATE FUNCTION square_root(double precision) RETURNS double precision
    AS 'dsqrt'
    LANGUAGE internal
    STRICT;
```

（大部分内部函数应该被声明为“严格”）。

注意

并非所有“预定义”函数在上述意义上都是“内部”函数。有些预定义的函数由 SQL 编写。

32.9. C 语言函数

用户定义的函数可以用 C 编写（或者可以与 C 兼容的语言，例如 C++）。这类函数被编译成动态载入对象（也被称为共享库）并且由服务器在需要时载入。动态载入是把“C语言”函数和“内部”函数区分开的特性 – 两者真正的编码习惯实际上是一样的（因此，标准内部函数库为用户定义的 C 函数提供了丰富的编码示例）。

目前 C 函数使用两种不同的调用约定。较新的“版本 1”调用约定通过为函数编写一个PG_FUNCTION_INFO_V1()宏来指示，如下文所示。没有这样的宏则表示这是一个旧式（“版本 0”）函数。无论哪种情况，CREATE FUNCTION中指定的语言名称都是C。旧式函数由于可移植性问题和功能缺乏，现在已被弃用，但出于兼容性原因仍然支持。

32.9.1. 动态载入

在一个会话中第一次调用一个特定可载入目标文件中的用户定义函数时，动态载入器会把那个目标文件载入到内存以便该函数被调用。因此用户定义的 C 函数的CREATE FUNCTION必须为该函数指定两块信息：可载入目标文件的名称，以及要在该目标文件中调用的特定函数的 C 名称（链接符号）。如果没有显式指定 C 名称，则它被假定为和 SQL 函数名相同。

下面的算法被用来基于CREATE FUNCTION 命令中给定的名称来定位共享目标文件：

1. 如果名称是一个绝对路径，则载入给定的文件。
2. 如果该名称以字符串$libdir开始，那么这一部分会被 PostgreSQL包的库目录名（在编译时确定）替换。
3. 如果该名称不包含目录部分，会在配置变量dynamic_library_path指定的路径中搜索该文件。
4. 否则（在该路径中没找到该文件，或者它包含一个非绝对目录），动态载入器将尝试接受给定的名称，这很可能会失败（依赖当前工作目录是不可靠的）。

如果这个序列不起作用，会把平台相关的共享库文件名扩展（通常是 .so）追加到给定的名称并且再次尝试上述的过程。如果还是失败，则载入失败。

用于运行PostgreSQL服务器的用户 ID 必须能够通过要载入文件的路径。常见的错误是把文件或更高层的目录变得对postgres用户不可读或者不可执行。

在任何情况下，CREATE FUNCTION命令中给定的文件名会被原封不动地记录在系统目录中，这样如果需要再次载入该文件则会应用同样的过程。

> **注意**
>
> PostgreSQL不会自动编译 C 函数。在从CREATE FUNCTION命令中引用目标文件之前，它必须先被编译好。更多信息请见第 32.9.6 节。

在第一次使用之后，动态载入目标文件会保留在内存中。在同一个会话中，后续对该文件中函数的调用只需付出一次很小的符号表查找开销。如果需要强制重新载入一个目标文件（例如在重新编译之后），可以使用LOAD命令，或者开启一个新的会话。

建议通过相对于 $libdir 的路径，或者通过动态库路径来定位共享库。这样一来，如果新安装位于不同的位置，版本升级会更简单。$libdir 实际代表的目录可以通过命令 pg_config --pkglibdir 查出。

在PostgreSQL 7.2 版之前，CREATE FUNCTION中只能指定目标文件的精确绝对路径。现在不推荐这种做法，因为它会使函数定义变得不必要地不可移植。最好只指定共享库名，不带路径和扩展名，而让搜索机制来提供这些信息。

32.9.2. C 语言函数中的基础类型

要了解如何编写 C 语言函数，你需要了解 PostgreSQL如何在内部表达基础类型以及如何与函数传递它们。在内部，PostgreSQL把基础类型视为一块“内存数据块”。你为该类型定义的用户自定义函数，决定了 PostgreSQL 如何操作它。也就是说，PostgreSQL 只负责把数据存到磁盘、再从磁盘取回，而数据的输入、处理和输出则依赖你定义的这些函数。

基础类型可以有三种内部格式之一：

- 传值，定长
- 传引用，定长
- 传引用，变长

传值类型在长度上只能是 1、2 或 4 字节（如果你的机器上 sizeof(Datum) 是 8，则还有 8 字节）。你应当小心地定义你的类型，使其在所有架构上都具有相同的尺寸（字节）。例如，long 类型很危险，因为它在某些机器上是 4 字节但在另一些机器上是 8 字节；而 int 类型在大多数 Unix 机器上都是 4 字节。在 Unix 机器上 int4 类型的一种合理实现可以是：

```
/* 4-byte integer, passed by value */
typedef int int4;
```

另一方面，任意尺寸的定长类型都可以通过引用传递。例如，这里有一种 PostgreSQL类型的实现示例：

```
/* 16 字节结构，传引用 */
typedef struct
{
    double  x, y;
} Point;
```

在PostgreSQL函数中传进或传出这种类型时，只能使用指向这种类型的指针。
要返回这样一种类型的值，用 palloc分配正确的内存量，然后填充分配好的内存，并且返回一个指向该内存的指针（还有，如果只想返回与具有相同数据类型的一个输入参数相同的值，可以跳过额外的palloc并且返回指向该输入值的指针）。

最后，所有变长类型也必须按引用传递。所有变长类型都必须以一个正好 4 字节的长度字段开始，而要存储在该类型中的所有数据都必须位于紧随该长度字段之后的内存中。
长度字段包含该结构的总长度，也就是说，它包含长度字段本身的尺寸。

> **警告**
>
> 绝不要修改通过引用传递的输入值的内容。如果这样做很可能会破坏磁盘上的数据，因为给出的指针可能直接指向一个磁盘缓冲区。这条规则唯一的例外在第 32.10 节中有解释。

作为一个例子，我们可以把类型 text 定义为如下形式：

```
typedef struct {
    int4 length;
    char data[1];
} text;
```

显然，这里声明的 data 字段不足以容纳所有可能的字符串。由于在 C 中无法声明变长结构体，我们依赖于这样的知识：C 编译器不会对数组下标做范围检查。我们只需分配必要数量的空间，然后就像数组声明为正确长度那样去访问它。（这是一个常见的技巧，你可以在许多关于 C 的教科书中读到。）

在操作变长类型时，我们必须小心分配正确数量的内存，并正确设置长度字段。例如，
如果我们想在一个 text 结构体中存储 40 字节，我们可以使用这样的代码片段：

```
#include "postgres.h"
...
char buffer[40]; /* our source data */
...
text *destination = (text *) palloc(VARHDRSZ + 40);
destination->length = VARHDRSZ + 40;
```

```
memcpy(destination->data, buffer, 40);
...
```

VARHDRSZ 和 sizeof(int4) 一样，但是用宏 VARHDRSZ 来引用变长类型的额外开销的尺寸被认为是比较好的风格。

表 32.1说明了在编写使用 PostgreSQL 内置类型的 C 语言函数时，哪种 C 类型对应哪种 SQL 类型。“定义文件”列给出了需要包含的头文件，以获取类型定义。（实际的定义可能位于所列文件包含的其他文件中。建议用户坚持使用已定义的接口。）请注意，
在任何源文件中都应该始终首先包含postgres.h，因为它声明了很多你反正都会用到的内容。

表 32.1. 内置 SQL 类型的等价 C 类型

SQL 类型	C 类型	定义文件
abstime	AbsoluteTime	utils/nabstime.h
boolean	bool	postgres.h（可能是编译器内置）
box	BOX*	utils/geo_decls.h
bytea	bytea*	postgres.h
"char"	char	（编译器内置）
character	BpChar*	postgres.h
cid	CommandId	postgres.h
date	DateADT	utils/date.h
smallint (int2)	int2 or int16	postgres.h
int2vector	int2vector*	postgres.h
integer (int4)	int4 or int32	postgres.h
real (float4)	float4*	postgres.h
double precision (float8)	float8*	postgres.h
interval	Interval*	utils/timestamp.h
lseg	LSEG*	utils/geo_decls.h
name	Name	postgres.h
oid	Oid	postgres.h
oidvector	oidvector*	postgres.h
path	PATH*	utils/geo_decls.h
point	POINT*	utils/geo_decls.h
regproc	regproc	postgres.h
reltime	RelativeTime	utils/nabstime.h
text	text*	postgres.h
tid	ItemPointer	storage/itemptr.h
time	TimeADT	utils/date.h
time with time zone	TimeTzADT	utils/date.h
timestamp	Timestamp*	utils/timestamp.h
tinterval	TimeInterval	utils/nabstime.h
varchar	VarChar*	postgres.h
xid	TransactionId	postgres.h

了解了基础类型所有可能的结构后，就可以看一些实际函数示例。

32.9.3. C 函数的调用约定版本 0

我们先介绍“旧式”调用约定——尽管这种方法现在已被弃用，但入门时更容易上手。在版本 0 方法中，C 函数的参数和结果就按普通的 C 风格声明，只是要注意使用上文所示的每种 SQL 数据类型的 C 表示。

下面是一些示例：

```
#include "postgres.h"
#include <string.h>
#include "utils/geo_decls.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

/* by value */

int
add_one(int arg)
{
    return arg + 1;
}

/* by reference, fixed length */

float8 *
add_one_float8(float8 *arg)
{
    float8    *result = (float8 *) palloc(sizeof(float8));

    *result = *arg + 1.0;

    return result;
}

Point *
makepoint(Point *pointx, Point *pointy)
{
    Point     *new_point = (Point *) palloc(sizeof(Point));

    new_point->x = pointx->x;
    new_point->y = pointy->y;

    return new_point;
}

/* by reference, variable length */

text *
copytext(text *t)
{
```

```
    /*
     * VARSIZE is the total size of the struct in bytes.
     */
    text *new_t = (text *) palloc(VARSIZE(t));
    VARATT_SIZEP(new_t) = VARSIZE(t);
    /*
     * VARDATA is a pointer to the data region of the struct.
     */
    memcpy((void *) VARDATA(new_t), /* destination */
           (void *) VARDATA(t),     /* source */
           VARSIZE(t)-VARHDRSZ);    /* how many bytes */
    return new_t;
}

text *
concat_text(text *arg1, text *arg2)
{
    int32 new_text_size = VARSIZE(arg1) + VARSIZE(arg2) - VARHDRSZ;
    text *new_text = (text *) palloc(new_text_size);

    VARATT_SIZEP(new_text) = new_text_size;
    memcpy(VARDATA(new_text), VARDATA(arg1), VARSIZE(arg1)-VARHDRSZ);
    memcpy(VARDATA(new_text) + (VARSIZE(arg1)-VARHDRSZ),
           VARDATA(arg2), VARSIZE(arg2) - VARHDRSZ);
    return new_text;
}
```

假定上述代码已经写在文件`funcs.c`中并编译为共享对象，我们可以用类似下面的命令向PostgreSQL定义这些函数：

```
CREATE FUNCTION add_one(integer) RETURNS integer
     AS 'DIRECTORY/funcs', 'add_one'
     LANGUAGE C STRICT;

-- note overloading of SQL function name "add_one"
CREATE FUNCTION add_one(double precision) RETURNS double precision
     AS 'DIRECTORY/funcs', 'add_one_float8'
     LANGUAGE C STRICT;

CREATE FUNCTION makepoint(point, point) RETURNS point
     AS 'DIRECTORY/funcs', 'makepoint'
     LANGUAGE C STRICT;

CREATE FUNCTION copytext(text) RETURNS text
     AS 'DIRECTORY/funcs', 'copytext'
     LANGUAGE C STRICT;

CREATE FUNCTION concat_text(text, text) RETURNS text
     AS 'DIRECTORY/funcs', 'concat_text'
     LANGUAGE C STRICT;
```

这里*DIRECTORY*代表共享库文件所在的目录（例如 PostgreSQL的教程目录，其中包含本节示例所用的代码）。（更好的风格是在`AS`子句中只写`'funcs'`，前提是已把 *DIRECTORY*加入搜索路径。无论哪种情况，都可以省略共享库的系统特定扩展名，通常是 `.so`或`.sl`。）

注意我们把函数声明为“strict”（严格），意思是如果任何输入值为空，系统会自动假定结果为空。这样做可以避免在函数代码中检查空输入。否则，我们就必须显式检查空值，即检查每个传引用参数是否为空指针。（对于传值参数，我们甚至没有办法检查！）

尽管这种调用约定使用简单，但可移植性不好；在某些体系结构上，以这种方式传递小于`int`的数据类型会有问题。另外，也没有返回空结果的简单方法，除了把函数声明为严格函数之外，也没有其他处理空参数的简单方法。下面介绍的版本 1 约定克服了这些缺点。

32.9.4. C 函数的调用约定版本 1

版本-1 的调用约定依赖于宏来屏蔽传递参数和结果的大部分复杂性。版本-1 函数的 C 声明总是：

```
Datum funcname(PG_FUNCTION_ARGS)
```

此外，宏调用：

```
PG_FUNCTION_INFO_V1(funcname);
```

必须出现在同一个源文件中（按惯例会正好写在该函数本身之前）。这种宏调用不是`internal`语言函数所需要的，因为 PostgreSQL会假定所有内部函数都使用版本-1 调用约定。不过，对于动态载入函数是必需的。

在版本-1 函数中，每一个实参都使用对应于该参数数据类型的`PG_GETARG_`*`xxx`*`()`宏取得，结果要用对应于返回类型的`PG_RETURN_`*`xxx`*`()`宏返回。`PG_GETARG_`*`xxx`*`()`的参数是要取得的函数参数的编号，从零开始计。`PG_RETURN_`*`xxx`*`()`的参数是实际要返回的值。

下面我们用版本-1 风格编写与上面相同的函数：

```
#include "postgres.h"
#include <string.h>
#include "fmgr.h"
#include "utils/geo_decls.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

/* by value */

PG_FUNCTION_INFO_V1(add_one);

Datum
add_one(PG_FUNCTION_ARGS)
{
    int32   arg = PG_GETARG_INT32(0);

    PG_RETURN_INT32(arg + 1);
}

/* by reference, fixed length */

PG_FUNCTION_INFO_V1(add_one_float8);
```

```
Datum
add_one_float8(PG_FUNCTION_ARGS)
{
    /* The macros for FLOAT8 hide its pass-by-reference nature. */
    float8   arg = PG_GETARG_FLOAT8(0);

    PG_RETURN_FLOAT8(arg + 1.0);
}

PG_FUNCTION_INFO_V1(makepoint);

Datum
makepoint(PG_FUNCTION_ARGS)
{
    /* Here, the pass-by-reference nature of Point is not hidden. */
    Point     *pointx = PG_GETARG_POINT_P(0);
    Point     *pointy = PG_GETARG_POINT_P(1);
    Point     *new_point = (Point *) palloc(sizeof(Point));

    new_point->x = pointx->x;
    new_point->y = pointy->y;

    PG_RETURN_POINT_P(new_point);
}

/* by reference, variable length */

PG_FUNCTION_INFO_V1(copytext);

Datum
copytext(PG_FUNCTION_ARGS)
{
    text     *t = PG_GETARG_TEXT_P(0);
    /*
     * VARSIZE is the total size of the struct in bytes.
     */
    text     *new_t = (text *) palloc(VARSIZE(t));
    VARATT_SIZEP(new_t) = VARSIZE(t);
    /*
     * VARDATA is a pointer to the data region of the struct.
     */
    memcpy((void *) VARDATA(new_t), /* destination */
           (void *) VARDATA(t),     /* source */
           VARSIZE(t)-VARHDRSZ);    /* how many bytes */
    PG_RETURN_TEXT_P(new_t);
}

PG_FUNCTION_INFO_V1(concat_text);

Datum
concat_text(PG_FUNCTION_ARGS)
{
    text  *arg1 = PG_GETARG_TEXT_P(0);
    text  *arg2 = PG_GETARG_TEXT_P(1);
```

```
    int32 new_text_size = VARSIZE(arg1) + VARSIZE(arg2) - VARHDRSZ;
    text *new_text = (text *) palloc(new_text_size);

    VARATT_SIZEP(new_text) = new_text_size;
    memcpy(VARDATA(new_text), VARDATA(arg1), VARSIZE(arg1)-VARHDRSZ);
    memcpy(VARDATA(new_text) + (VARSIZE(arg1)-VARHDRSZ),
           VARDATA(arg2), VARSIZE(arg2) - VARHDRSZ);
    PG_RETURN_TEXT_P(new_text);
}
```

这些函数的CREATE FUNCTION命令与版本 0 的等价形式相同。

乍看之下，版本 1 的编码约定似乎只是无谓的故弄玄虚。不过，它们确实提供了许多改进，因为宏可以隐藏不必要的细节。例如，在编码add_one_float8时，我们不再需要知道float8是传引用类型。另一个例子是，变长类型的GETARG宏允许更有效地取得 “toasted”（压缩或外置）值。

版本-1 函数的一大改进是对空输入和结果的更好处理。宏PG_ARGISNULL(*n*)允许一个函数测试是否每一个输入为空值（当然，只需要在没有声明为 “strict” 的函数中这样做）。和PG_GETARG_*xxx*()宏一样，输入参数也是从零开始计数。
注意应该在验证了一个参数不是空值之后才执行PG_GETARG_*xxx*()。要返回空值结果，应执行PG_RETURN_NULL()，它对严格的以及非严格的函数都有用。

新式接口中提供的其他选项是PG_GETARG_*xxx*()宏的两个变种。其中的第一种是PG_GETARG_*xxx*_COPY()，
它确保返回的指定参数的拷贝可以被安全地写入（通常的宏有时会返回一个指向表中物理存储值的指针，不能写入该值。使用PG_GETARG_*xxx*_COPY()宏可以保证得到一个可写的结果）。
第二种变种PG_GETARG_*xxx*_SLICE()宏有三个参数。第一个是函数参数的编号（如上文）。
第二个和第三个是要被返回的段的偏移量和长度。偏移量从零开始计算，
而负值的长度则表示要求返回该值的剩余部分。当大型值的存储类型为 “external” 时，
这些宏提供了访问这些大型值的部分内容的更有效方法（列的存储类型可以使用ALTER TABLE *tablename* ALTER COLUMN *colname* SET STORAGE *storagetype*来指定。*storagetype*取plain、external、extended或者main）。

最后，版本-1 的函数调用约定可以返回集合结果（第 32.9.10 节）、实现触发器函数（第 33 章）和过程语言调用处理器（第 46 章）。版本-1 代码也比版本-0 更可移植，因为它不违反 C 标准中关于函数调用协议的限制。更多细节可见源代码发布中的src/backend/utils/fmgr/README。

32.9.5. 编写代码

在开始更高级的话题之前，我们应该讨论一下用于 PostgreSQL C 语言函数的编码规则。虽然有可能把不是 C 编写的函数载入到 PostgreSQL中，但即使能够做到，通常也很困难，因为其他语言（例如 C++、FORTRAN 或者 Pascal）通常不会遵循和 C 相同的调用约定。也就是说，其他语言不会以同样的方式在函数之间传递参数以及返回值。由于这个原因，我们会假定你的 C 语言函数确实是用 C 编写的。

编写和构建 C 语言函数的基本规则如下：

- 使用 pg_config --includedir-server 查出 PostgreSQL 服务器头文件在你的系统上（或你的用户将要运行的系统上）安装于何处。这个选项是PostgreSQL 7.2 新增的。对于PostgreSQL 7.1，你应当使用选项--includedir。（如果遇到未知选项，pg_config将以非零状态退出。）对于 7.1 之前的版本，你只能猜测，但由于那是在现行调用约定引入之前，你不太可能还想要支持那些版本。

- 分配内存时，使用 PostgreSQL 提供的 `palloc` 和 `pfree`，而不是相应的 C 库函数 `malloc` 和 `free`。用 `palloc` 分配的内存会在每个事务结束时自动释放，从而避免内存泄漏。
- 总是使用 `memset` 将结构体的所有字节清零。如果不这么做，就很难支持哈希索引或哈希连接，因为那时你必须只挑出数据结构中真正有意义的位来计算哈希值。
  即使你初始化了结构体的所有字段，
  结构体中仍可能存在包含垃圾值的对齐填充字节（也就是结构体中的空洞）。
- PostgreSQL 的大多数内部类型都在 `postgres.h` 中声明，而函数管理器接口（`PG_FUNCTION_ARGS` 等）位于 `fmgr.h` 中，因此至少需要包含这两个文件。出于可移植性考虑，最好把 `postgres.h` 放在最前面，先于任何其他系统或用户头文件。包含 `postgres.h` 时，也会顺带为你包含 `elog.h` 和 `palloc.h`。
- 目标文件中定义的符号名不能彼此冲突，也不能与 PostgreSQL 服务器可执行文件中定义的符号冲突。如果你收到这类错误消息，就必须重命名相关函数或变量。
- 为了让你的代码能够被 PostgreSQL 动态装入，编译和链接时总是需要特殊的选项。关于如何在特定操作系统上完成这件事，详见第 32.9.6 节。

32.9.6. Compiling and Linking Dynamically-Loaded Functions

在你能够使用以 C 编写的 PostgreSQL 扩展函数之前，必须以特殊方式对它们进行编译和链接，以生成一个可由服务器动态装载的文件。更准确地说，需要创建一个共享库。

若想了解本节未涵盖的信息，你应阅读操作系统的文档，特别是 C 编译器 `cc` 和链接编辑器 `ld` 的手册页。此外，PostgreSQL 源代码在 `contrib` 目录中包含若干可用的示例。不过，如果你依赖这些示例，就会使你的模块依赖于 PostgreSQL 源代码是否可用。

创建共享库通常与链接可执行文件类似：先把源文件编译为目标文件，
再把目标文件链接在一起。目标文件需要以位置无关代码（PIC）形式生成。从概念上讲，
这意味着当它们被可执行文件装载时，可以放在内存中的任意位置。
（面向可执行文件的目标文件通常不会这样编译。）链接共享库的命令中也包含一些特殊标志，
用来把它与链接可执行文件的命令区分开来（至少理论上如此，
某些系统上的实际做法要丑陋得多）。

在下面的示例中，我们假定你的源代码位于文件 `foo.c` 中，并将创建共享库 `foo.so`。除非另有说明，中间目标文件名为 `foo.o`。共享库可以包含多个目标文件，但这里我们只使用一个。

BSD/OS

生成 PIC 的编译器选项是 `-fpic`。创建共享库时使用的链接器选项是 `-shared`。

```
gcc -fpic -c foo.c
ld -shared -o foo.so foo.o
```

这从 BSD/OS 4.0 版起适用。

FreeBSD

生成 PIC 的编译器选项是 `-fpic`。创建共享库时使用的编译器选项是 `-shared`。

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

这从 FreeBSD 3.0 版起适用。

HP-UX

生成 PIC 的系统编译器选项是 `+z`。使用 GCC 时则是 `-fpic`。用于共享库的链接器选项是 `-b`。因此：

```
cc +z -c foo.c
```

or:

```
gcc -fpic -c foo.c
```

and then:

```
ld -b -o foo.sl foo.o
```

与大多数其他系统不同，HP-UX 使用 `.sl` 作为共享库扩展名。

IRIX

PIC 是默认行为，无须特殊的编译器选项。生成共享库时使用的链接器选项是 `-shared`。

```
cc -c foo.c
ld -shared -o foo.so foo.o
```

Linux

生成 PIC 的编译器选项是 `-fpic`。在某些平台的某些情况下，如果 `-fpic` 不起作用，则必须使用 `-fPIC`。更多信息请参阅 GCC 手册。创建共享库的编译器选项是 `-shared`。完整示例如下：

```
cc -fpic -c foo.c
cc -shared -o foo.so foo.o
```

MacOS X

下面是一个示例。它假定开发者工具已经安装。

```
cc -c foo.c
cc -bundle -flat_namespace -undefined suppress -o foo.so foo.o
```

NetBSD

生成 PIC 的编译器选项是 `-fpic`。对于 ELF 系统，使用带 `-shared` 选项的编译器来链接共享库。在较旧的非 ELF 系统上，则使用 `ld -Bshareable`。

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

OpenBSD

生成 PIC 的编译器选项是 `-fpic`。链接共享库使用 `ld -Bshareable`。

```
gcc -fpic -c foo.c
```

```
ld -Bshareable -o foo.so foo.o
```

Solaris

使用 Sun 编译器时，生成 PIC 的编译器选项是 `-KPIC`；使用 GCC 时则是 `-fpic`。要链接共享库，两种编译器都使用编译器选项 `-G`，或者在使用 GCC 时改用 `-shared`。

```
cc -KPIC -c foo.c
cc -G -o foo.so foo.o
```

or

```
gcc -fpic -c foo.c
gcc -G -o foo.so foo.o
```

Tru64 UNIX

PIC 是默认值，因此编译命令就是常规的那条。链接时需要使用带特殊选项的`ld`。

```
cc -c foo.c
ld -shared -expect_unresolved '*' -o foo.so foo.o
```

使用 GCC 代替系统编译器时过程相同；不需要特殊选项。

UnixWare

生成 PIC 的编译器选项，对于 SCO 编译器是 `-K PIC`，而 `-fpic` 则用于 GCC。链接共享库时，编译器选项对于 SCO 编译器是 `-G`，对于 GCC 则是 `-shared`。

```
cc -K PIC -c foo.c
cc -G -o foo.so foo.o
```

或者

```
gcc -fpic -c foo.c
gcc -shared -o foo.so foo.o
```

提示

如果这些内容对你来说过于复杂，可以考虑使用 GNU Libtool[1]，它通过统一接口隐藏了平台差异。

生成的共享库文件随后就可以装载到 PostgreSQL 中。在向 `CREATE FUNCTION` 命令指定文件名时，必须给出共享库文件名，而不是中间目标文件名。请注意，系统标准的共享库扩展名（通常是 `.so` 或 `.sl`）可以在 `CREATE FUNCTION` 命令中省略，并且通常也应省略，以获得最佳可移植性。

关于服务器期望在何处找到共享库文件，请回头参见第 32.9.1 节。

32.9.7. 扩展构建基础设施

[1] http://www.gnu.org/software/libtool/

如果你正在考虑分发你的 PostgreSQL 扩展模块，为它们搭建一个可移植的构建系统可能相当困难。因此，PostgreSQL 安装为扩展提供了一套构建基础设施，称为 PGXS，使得简单的扩展模块可以直接针对已安装的服务器构建。注意，这套基础设施并不是一个通用的构建系统框架，不能用来构建所有与 PostgreSQL 交互的软件；它只是为简单的服务器扩展模块自动化了常用的构建规则。对于更复杂的软件包，你需要编写自己的构建系统。

要为你的扩展使用这套基础设施，你必须编写一个简单的 makefile。在该 makefile 中，你需要设置一些变量，并在最后包含全局的 PGXS makefile。下面是一个示例，它构建一个名为 `isbn_issn` 的扩展模块，该模块由一个共享库、一个 SQL 脚本和一个文档文本文件组成：

```
MODULES = isbn_issn
DATA_built = isbn_issn.sql
DOCS = README.isbn_issn

PGXS := $(shell pg_config --pgxs)
include $(PGXS)
```

最后两行应始终相同。在文件的前面部分，可以给变量赋值或添加自定义的 make 规则。

可以设置下列变量：

`MODULES`

要从同名源文件构建的共享库对象列表（列表中不要包含库后缀）

`DATA`

要安装到 *`prefix`*`/share/contrib` 的任意文件

`DATA_built`

要安装到 *`prefix`*`/share/contrib` 的任意文件，但它们需要先被构建

`DOCS`

要安装到 *`prefix`*`/doc/contrib` 下的任意文件

`SCRIPTS`

要安装到 *`prefix`*`/bin` 的脚本文件（不是二进制文件）

`SCRIPTS_built`

要安装到 *`prefix`*`/bin` 的脚本文件（不是二进制文件），但它们需要先被构建

`REGRESS`

回归测试用例列表（不带后缀）

`PROGRAM`

要构建的二进制程序（在 `OBJS` 中列出目标文件）

`MODULE_big`

要从多个源文件构建的共享对象（在 `OBJS` 中列出目标文件）

`EXTRA_CLEAN`

要在 `make clean` 中额外移除的文件

`PG_CPPFLAGS`

将被追加到 `CPPFLAGS`

`PG_LIBS`

将被追加到 `PROGRAM` 链接行

`SHLIB_LINK`

将被追加到 `MODULE_big` 链接行

把这个 makefile 命名为 `Makefile`，并放在保存扩展的目录中。然后你就可以执行 `make` 进行编译，再执行 `make install` 安装你的模块。该扩展会针对你的 path 中找到的第一个 `pg_config` 所对应的 PostgreSQL 安装进行编译和安装。

32.9.8. C 函数中的复合类型参数

复合类型没有像 C 结构体那样的固定布局。复合类型的实例可能包含空值字段。此外，继承层次中的复合类型可能具有和同一继承层次中其他成员不同的字段。因此，PostgreSQL提供了函数接口以便从 C 访问复合类型的字段。

假设我们想编写一个函数来回答如下查询：

```
SELECT name, c_overpaid(emp, 1500) AS overpaid
    FROM emp
    WHERE name = 'Bill' OR name = 'Sam';
```

使用版本 0 调用约定，我们可以将 `c_overpaid` 定义为：

```
#include "postgres.h"
#include "executor/executor.h"  /* for GetAttributeByName() */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

bool
c_overpaid(HeapTupleHeader t, /* the current row of emp */
           int32 limit)
{
    bool isnull;
    int32 salary;

    salary = DatumGetInt32(GetAttributeByName(t, "salary", &isnull));
    if (isnull)
        return false;
    return salary > limit;
}
```

按照版本-1 编码，上面的函数写成这样：

```
#include "postgres.h"
#include "executor/executor.h"  /* for GetAttributeByName() */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

PG_FUNCTION_INFO_V1(c_overpaid);

Datum
c_overpaid(PG_FUNCTION_ARGS)
{
    HeapTupleHeader  t = PG_GETARG_HEAPTUPLEHEADER(0);
    int32            limit = PG_GETARG_INT32(1);
    bool isnull;
    Datum salary;

    salary = GetAttributeByName(t, "salary", &isnull);
    if (isnull)
        PG_RETURN_BOOL(false);
    /* Alternatively, we might prefer to do PG_RETURN_NULL() for null
 salary. */

    PG_RETURN_BOOL(DatumGetInt32(salary) > limit);
}
```

`GetAttributeByName` 是 PostgreSQL 的一个系统函数，用于从指定行中取出属性。它有三个参数：传入函数的 `HeapTupleHeader` 类型参数、所需属性的名称，以及一个用于指示该属性是否为 null 的返回参数。`GetAttributeByName` 返回一个 `Datum` 值，你可以用适当的 `Datum Get*XXX*()` 宏把它转换为正确的数据类型。注意，如果 null 标志被设置，那么返回值本身没有意义；在尝试对结果做任何处理之前，务必先检查这个 null 标志。

也有`GetAttributeByNum`函数，它可以用目标属性的列号而不是属性名来选择目标属性。

下面的命令声明 SQL 中的`c_overpaid`：

```
CREATE FUNCTION c_overpaid(emp, integer) RETURNS boolean
    AS 'DIRECTORY/funcs', 'c_overpaid'
    LANGUAGE C STRICT;
```

注意我们用了`STRICT`，这样我们不需要检查输入参数是否为 NULL。

32.9.9. 从 C 函数返回行（复合类型）

要从 C 语言函数中返回一行或一个复合类型值，可以使用一套特殊的 API，它通过一组宏和函数隐藏了构造复合数据类型时的大部分复杂性。要使用这套 API，源文件中必须包含：

```
#include "funcapi.h"
```

构造复合数据值（下文简称“元组”）有两种方式：一种是从 Datum 值数组构造，另一种是从 C 字符串数组构造，这些字符串会传给该元组各列数据类型的输入转换函数。无论采用哪种方式，首先都需要获取或构造描述该元组结构的 `TupleDesc`。处理 Datum 时，需要把 `TupleDesc` 传给 `BlessTupleDesc`，然后为每一行调用 `heap_form_tuple`。处理 C 字符串时，则要把 `TupleDesc` 传给 `TupleDescGetAttInMetadata`，然后为每一行调用 `BuildTupleFromCStrings`。对于返回元组集合的函数，这些准备步骤可以在第一次调用函数时一次性完成。

有一些辅助函数可以用来设置所需的 TupleDesc。在大多数返回复合值的函数中，推荐的做法是调用：

```
TypeFuncClass get_call_result_type(FunctionCallInfo fcinfo,
                                   Oid *resultTypeId,
                                   TupleDesc *resultTupleDesc)
```

传入调用函数本身收到的同一个 fcinfo 结构体（这当然要求使用版本 1 调用约定）。resultTypeId 可以指定为 NULL，也可以指定为一个本地变量的地址，用于接收函数结果类型的 OID。resultTupleDesc 应当是一个本地 TupleDesc 变量的地址。检查返回结果是否为 TYPEFUNC_COMPOSITE；如果是，resultTupleDesc 就会被填入所需的 TupleDesc。（如果不是，可以报告一个类似 “function returning record called in context that cannot accept type record” 的错误。）

> **提示**
>
> get_call_result_type能够解析一个多态函数结果的实际类型，
> 因此不仅在返回复合类型的函数中，在返回标量多态结果的函数中它也是非常有用的。
> resultTypeId输出主要用于返回多态标量的函数。

> **注意**
>
> get_call_result_type有一个兄弟 get_expr_result_type，它被用来解析被表示为一棵表达式树的函数调用的输出类型。
> 在尝试从函数自身外部确定结果类型时可以用它。也有一个get_func_result_type，当只有函数的 OID 可用时可以用它。不过这些函数无法处理被声明为返回record的函数，并且get_func_result_type无法解析多态类型，因此你应该优先使用get_call_result_type。

更早、现在已被弃用的获取TupleDesc的函数有：

```
TupleDesc RelationNameGetTupleDesc(const char *relname)
```

它可以为一个指定名称的关系的行类型得到TupleDesc，还有：

```
TupleDesc TypeGetTupleDesc(Oid typeoid, List *colaliases)
```

可以基于一个类型 OID 得到TupleDesc。这可以被用来为一种基本或者复合类型获得TupleDesc。不过，对于返回record的函数它不起作用，并且它无法解析多态类型。

一旦有了一个TupleDesc，如果计划处理 Datum，可以调用：

```
TupleDesc BlessTupleDesc(TupleDesc tupdesc)
```

如果计划处理 C 字符串，可调用：

```
AttInMetadata *TupleDescGetAttInMetadata(TupleDesc tupdesc)
```

如果正在编写一个返回集合的函数，你可以把这些函数的结果保存在 FuncCallContext结构体中 – 分别使用 tuple_desc或者attinmeta字段。

在处理 Datum 时，使用

```
HeapTuple heap_form_tuple(TupleDesc tupdesc, Datum *values, bool
 *isnull)
```

来用 Datum 形式的用户数据构建一个`HeapTuple`。

在处理 C 字符串时，使用

```
HeapTuple BuildTupleFromCStrings(AttInMetadata *attinmeta, char
 **values)
```

来用 C 字符串形式的用户数据构建一个`HeapTuple`。`values`是一个 C 字符串数组，每一个元素是返回行的一个属性。每一个 C 字符串应该是该属性数据类型的输入函数所期望的格式。为了对一个属性返回空值，*`values`*数组中对应的指针应该被设置为`NULL`。
对于你返回的每一行都需要再次调用这个函数。

一旦已经构建了一个要从函数中返回的元组，它必须被转换成一个 `Datum`。使用

```
HeapTupleGetDatum(HeapTuple tuple)
```

可把一个`HeapTuple`转换成合法的 Datum。如果你只想返回一行，那么这个`Datum`可以被直接返回，在一个集合返回函数中它也可以被当做当前的返回值。

下一节中会有一个示例。

32.9.10. 从 C 函数返回集合

还有一种特殊的 API，为从 C 语言函数返回集合（多行）提供支持。集合返回函数必须遵循版本 1 调用约定。此外，与前述一样，源文件必须包含`funcapi.h`。

集合返回函数（SRF）为它返回的每一项被调用一次。因此，SRF 必须保存足够的状态，以记住它先前在做什么，并在每次调用时返回下一项。`FuncCallContext` 结构体就是用来帮助控制这一过程的。在函数内部，`fcinfo->flinfo->fn_extra` 用于在调用之间保存指向 `FuncCallContext` 的指针。

```
typedef struct
{
    /*
     * Number of times we've been called before
     *
     * call_cntr is initialized to 0 for you by SRF_FIRSTCALL_INIT(),
 and
     * incremented for you every time SRF_RETURN_NEXT() is called.
     */
    uint32 call_cntr;

    /*
     * OPTIONAL maximum number of calls
     *
     * max_calls is here for convenience only and setting it is
 optional.
     * If not set, you must provide alternative means to know when the
     * function is done.
```

```
     */
    uint32 max_calls;

    /*
     * OPTIONAL pointer to result slot
     *
     * This is obsolete and only present for backwards compatibility,
viz,
     * user-defined SRFs that use the deprecated TupleDescGetSlot().
     */
    TupleTableSlot *slot;

    /*
     * OPTIONAL pointer to miscellaneous user-provided context
information
     *
     * user_fctx is for use as a pointer to your own data to retain
     * arbitrary context information between calls of your function.
     */
    void *user_fctx;

    /*
     * OPTIONAL pointer to struct containing attribute type input
metadata
     *
     * attinmeta is for use when returning tuples (i.e., composite
data types)
     * and is not used when returning base data types. It is only
needed
     * if you intend to use BuildTupleFromCStrings() to create the
return
     * tuple.
     */
    AttInMetadata *attinmeta;

    /*
     * memory context used for structures that must live for multiple
calls
     *
     * multi_call_memory_ctx is set by SRF_FIRSTCALL_INIT() for you,
and used
     * by SRF_RETURN_DONE() for cleanup. It is the most appropriate
memory
     * context for any memory that is to be reused across multiple
calls
     * of the SRF.
     */
    MemoryContext multi_call_memory_ctx;

    /*
     * OPTIONAL pointer to struct containing tuple description
     *
     * tuple_desc is for use when returning tuples (i.e., composite
data types)
```

```
     * and is only needed if you are going to build the tuples with
     * heap_form_tuple() rather than with BuildTupleFromCStrings().
 Note that
     * the TupleDesc pointer stored here should usually have been run
 through
     * BlessTupleDesc() first.
     */
    TupleDesc tuple_desc;

} FuncCallContext;
```

SRF 会使用若干自动操纵 `FuncCallContext` 结构体（并期望通过 `fn_extra` 找到它）的函数和宏。使用：

```
SRF_IS_FIRSTCALL()
```

来判断你的函数是第一次被调用还是后续调用。在第一次调用时（只在第一次调用时）使用：

```
SRF_FIRSTCALL_INIT()
```

来初始化 `FuncCallContext`。在每次函数调用时，包括第一次，使用：

```
SRF_PERCALL_SETUP()
```

来正确设置对 `FuncCallContext` 的使用，并清除上一次处理遗留的任何已返回数据。

如果你的函数有数据要返回，使用：

```
SRF_RETURN_NEXT(funcctx, result)
```

把它返回给调用者（`result`必须是类型`Datum`，
可以是一个单一值或者按上文所述准备好的元组）。最后，当函数完成了数据返回后，可使用：

```
SRF_RETURN_DONE(funcctx)
```

来清理并且结束SRF。

调用 SRF 时当前所处的内存上下文是一个瞬时上下文，它会在两次调用之间被清空。这意味着，你不必对用 `palloc` 分配的所有东西都调用 `pfree`，因为它们反正会自动释放。不过，如果你需要分配在多次调用之间持续存在的数据结构，就必须把它们放到别处。对于任何需要一直存活到 SRF 运行结束的数据，`multi_call_memory_ctx` 所指向的内存上下文就是合适的位置。在大多数情况下，这意味着你应当在做首次调用初始化时切换到 `multi_call_memory_ctx`。

一个完整的伪代码示例：

```
Datum
my_set_returning_function(PG_FUNCTION_ARGS)
{
    FuncCallContext  *funcctx;
    Datum             result;
    further declarations as needed

    if (SRF_IS_FIRSTCALL())
    {
        MemoryContext oldcontext;
```

```
        funcctx = SRF_FIRSTCALL_INIT();
        oldcontext = MemoryContextSwitchTo(funcctx->multi_call_memory_
ctx);
        /* 这里是一次性设置代码： */
        user code
        if returning composite
            build TupleDesc, and perhaps AttInMetadata
        endif returning composite
        user code
        MemoryContextSwitchTo(oldcontext);
    }

    /* 这里是每一次都要做的设置代码： */
    user code
    funcctx = SRF_PERCALL_SETUP();
    user code

    /* 这里只是一种测试是否执行完的方法： */
    if (funcctx->call_cntr < funcctx->max_calls)
    {
        /* 这里返回另一个项： */
        user code
        obtain result Datum
        SRF_RETURN_NEXT(funcctx, result);
    }
    else
    {
        /* 这里已经完成了项的返回，只需要做清理： */
        user code
        SRF_RETURN_DONE(funcctx);
    }
}
```

一个返回复合类型的简单 SRF 的完整示例如下：

```
PG_FUNCTION_INFO_V1(retcomposite);

Datum
retcomposite(PG_FUNCTION_ARGS)
{
    FuncCallContext     *funcctx;
    int                  call_cntr;
    int                  max_calls;
    TupleDesc            tupdesc;
    AttInMetadata       *attinmeta;

    /* stuff done only on the first call of the function */
    if (SRF_IS_FIRSTCALL())
    {
        MemoryContext   oldcontext;

        /* create a function context for cross-call persistence */
        funcctx = SRF_FIRSTCALL_INIT();
```

```
        /* switch to memory context appropriate for multiple function
 calls */
        oldcontext = MemoryContextSwitchTo(funcctx->multi_call_memory_
ctx);

        /* total number of tuples to be returned */
        funcctx->max_calls = PG_GETARG_UINT32(0);

        /* Build a tuple descriptor for our result type */
        if (get_call_result_type(fcinfo, NULL, &tupdesc) != TYPEFUNC_
COMPOSITE)
            ereport(ERROR,
                    (errcode(ERRCODE_FEATURE_NOT_SUPPORTED),
                     errmsg("function returning record called in
 context "
                            "that cannot accept type record")));

        /*
         * generate attribute metadata needed later to produce tuples
 from raw
         * C strings
         */
        attinmeta = TupleDescGetAttInMetadata(tupdesc);
        funcctx->attinmeta = attinmeta;

        MemoryContextSwitchTo(oldcontext);
    }

    /* stuff done on every call of the function */
    funcctx = SRF_PERCALL_SETUP();

    call_cntr = funcctx->call_cntr;
    max_calls = funcctx->max_calls;
    attinmeta = funcctx->attinmeta;

    if (call_cntr < max_calls)    /* do when there is more left to
 send */
    {
        char       **values;
        HeapTuple    tuple;
        Datum        result;

        /*
         * Prepare a values array for building the returned tuple.
         * This should be an array of C strings which will
         * be processed later by the type input functions.
         */
        values = (char **) palloc(3 * sizeof(char *));
        values[0] = (char *) palloc(16 * sizeof(char));
        values[1] = (char *) palloc(16 * sizeof(char));
        values[2] = (char *) palloc(16 * sizeof(char));

        snprintf(values[0], 16, "%d", 1 * PG_GETARG_INT32(1));
```

```
        snprintf(values[1], 16, "%d", 2 * PG_GETARG_INT32(1));
        snprintf(values[2], 16, "%d", 3 * PG_GETARG_INT32(1));

        /* build a tuple */
        tuple = BuildTupleFromCStrings(attinmeta, values);

        /* make the tuple into a datum */
        result = HeapTupleGetDatum(tuple);

        /* clean up (this is not really necessary) */
        pfree(values[0]);
        pfree(values[1]);
        pfree(values[2]);
        pfree(values);

        SRF_RETURN_NEXT(funcctx, result);
    }
    else    /* do when there is no more left */
    {
        SRF_RETURN_DONE(funcctx);
    }
}
```

在 SQL 中声明此函数的一种方法是：

```
CREATE TYPE __retcomposite AS (f1 integer, f2 integer, f3 integer);

CREATE OR REPLACE FUNCTION retcomposite(integer, integer)
    RETURNS SETOF __retcomposite
    AS 'filename', 'retcomposite'
    LANGUAGE C IMMUTABLE STRICT;
```

另一种方式是使用 OUT 参数：

```
CREATE OR REPLACE FUNCTION retcomposite(IN integer, IN integer,
    OUT f1 integer, OUT f2 integer, OUT f3 integer)
    RETURNS SETOF record
    AS 'filename', 'retcomposite'
    LANGUAGE C IMMUTABLE STRICT;
```

请注意，采用这种方法时，函数的输出类型在形式上是一个匿名的 `record` 类型。

源码发布中的 `contrib/tablefunc` 目录包含更多返回集合函数的示例。

32.9.11. 多态参数和返回类型

C 语言函数可以声明为接受和返回多态类型 `anyelement` 和 `anyarray`。关于多态函数的更详细解释，见第 32.2.5 节。当函数参数或者返回类型被定义为多态类型时，函数的编写者无法提前知道会用什么数据类型调用该函数或者该函数需要返回什么数据类型。`fmgr.h` 中提供了两种例程，允许版本-1 的 C 函数发现其参数的实际数据类型以及它要返回的类型。这些例程被称为 `get_fn_expr_rettype(FmgrInfo *flinfo)` 和 `get_fn_expr_argtype(FmgrInfo *flinfo, int argnum)`。它们返回结果或者参数的类型的 OID，或者当该信息不可用时返回 `InvalidOid`。结构体 `flinfo` 通常被当做 `fcinfo->flinfo` 访问。参数 `argnum` 则是从零开始计。`get_call_result_type` 也可被用作 `get_fn_expr_rettype` 的替代。

例如，假设我们想要写一个接收一个任意类型元素并且返回一个该类型的一维数组的函数：

```
PG_FUNCTION_INFO_V1(make_array);
Datum
make_array(PG_FUNCTION_ARGS)
{
    ArrayType  *result;
    Oid         element_type = get_fn_expr_argtype(fcinfo->flinfo, 0);
    Datum       element;
    bool        isnull;
    int16       typlen;
    bool        typbyval;
    char        typalign;
    int         ndims;
    int         dims[MAXDIM];
    int         lbs[MAXDIM];

    if (!OidIsValid(element_type))
        elog(ERROR, "could not determine data type of input");

    /* 得到提供的元素 */
    element = PG_GETARG_DATUM(0);

    /* 只有一个维度 */
    ndims = 1;
    /* 和一个元素 */
    dims[0] = 1;
    /* 且下界是 1 */
    lbs[0] = 1;

    /* 得到该元素类型所需的信息 */
    get_typlenbyvalalign(element_type, &typlen, &typbyval, &typalign);

    /* 现在构建数组 */
    result = construct_md_array(&element, ndims, dims, lbs,
                                element_type, typlen, typbyval,
 typalign);

    PG_RETURN_ARRAYTYPE_P(result);
}
```

下面的命令在 SQL 中声明了 `make_array` 函数：

```
CREATE FUNCTION make_array(anyelement) RETURNS anyarray
    AS 'DIRECTORY/funcs', 'make_array'
    LANGUAGE C STRICT;
```

注意这里使用了`STRICT`；这一点很关键，因为代码没有费力去测试空输入。

32.10. 用户定义的聚合

PostgreSQL 中的聚合函数是依据状态值和状态转移函数来定义的。也就是说，聚合可以定义为一种状态，在处理每个输入项时修改该状态。要定义一个新的聚合函数，
需要为状态值选定一种数据类型、为状态选定一个初始值，并指定一个状态转移函数。

状态转移函数只是普通函数，也可以在聚合的上下文之外使用。
如果聚合的期望结果不同于运行状态值中需要保存的数据，还可以指定一个最终函数。

因此，除了聚合用户所看到的参数和结果数据类型之外，还存在一种内部状态值数据类型，
它可能既不同于参数类型，也不同于结果类型。

如果我们定义一个不使用最终函数的聚合，
那么得到的就是一个对每一行的列值进行逐步计算的聚合。sum 就是这种聚合的一个示例。sum 从零开始，并始终把当前行的值加到运行总和中。例如，如果我们希望让 sum 聚合能用于复数数据类型，那么只需要该数据类型的加法函数。聚合定义如下：

```
CREATE AGGREGATE complex_sum (
    sfunc = complex_add,
    basetype = complex,
    stype = complex,
    initcond = '(0,0)'
);

SELECT complex_sum(a) FROM test_complex;

 complex_sum
-------------
 (34,53.9)
```

（实际上，我们会直接把这个聚合命名为sum，并依靠 PostgreSQL 来判断对 complex 类型的列应用哪种 sum。）

如果没有非空输入值，上面这个 sum 的定义会返回零（初始状态条件）。也许我们希望在这种情况下返回 null —— SQL 标准就期望 sum 这样表现。只需省略 initcond 短语，使初始状态条件为 null，就可以做到这一点。通常这意味着 sfunc 需要检查空状态条件输入。但对于 sum 以及其他一些简单聚合（如 max 和 min），只需把第一个非空输入值插入状态变量，然后从第二个非空输入值开始应用转移函数即可。如果初始条件为 null 且转移函数被标记为“strict”（即不为空输入调用），PostgreSQL 会自动完成这项工作。

“strict”转移函数的另一项默认行为是，只要遇到空值输入，就会保留先前的状态值不变。
因此，空值会被忽略。如果你需要对空值输入采用其他行为，就不要把转移函数声明为 strict，而应在函数代码中自行检测空值输入并完成所需处理。

avg（平均值）是更复杂的聚合示例。它需要两个运行状态：输入的总和与输入的个数。
最终结果由二者相除得到。平均值通常用双元素数组作为状态值来实现。例如，avg(float8) 的内置实现如下所示：

```
CREATE AGGREGATE avg (
    sfunc = float8_accum,
    basetype = float8,
    stype = float8[],
    finalfunc = float8_avg,
    initcond = '{0,0}'
);
```

聚合函数可以使用多态的状态转移函数或最终函数，这样同一组函数就能用于实现多个聚合。
关于多态函数的解释，请参见第 32.2.5 节。再进一步，聚合函数自身也可以指定为具有多态基类型和状态类型，从而让同一个聚合定义服务于多种输入数据类型。下面是一个多态聚合的示例：

```
CREATE AGGREGATE array_accum (
```

```
    sfunc = array_append,
    basetype = anyelement,
    stype = anyarray,
    initcond = '{}'
);
```

对于任何给定的聚合调用，其实际状态类型都是一种数组类型，
该数组以实际输入类型为元素类型。

下面是以两种不同的实际数据类型作为参数时的输出：

```
SELECT attrelid::regclass, array_accum(attname)
    FROM pg_attribute
    WHERE attnum > 0 AND attrelid = 'pg_tablespace'::regclass
    GROUP BY attrelid;

   attrelid    |              array_accum
---------------+---------------------------------------
 pg_tablespace | {spcname,spcowner,spclocation,spcacl}
(1 row)

SELECT attrelid::regclass, array_accum(atttypid)
    FROM pg_attribute
    WHERE attnum > 0 AND attrelid = 'pg_tablespace'::regclass
    GROUP BY attrelid;

   attrelid    |   array_accum
---------------+-----------------
 pg_tablespace | {19,26,25,1034}
(1 row)
```

用 C 编写的函数可以通过检查函数调用的“上下文”是否被传入了一个`AggState`节点，来检测自己是否作为聚合转移函数或最终函数被调用，例如：

```
        if (fcinfo->context && IsA(fcinfo->context, AggState))
```

检查这一点的一个原因是，如果结果为真，那么左输入一定是临时转移值，
因此可以安全地原地修改，而不必分配一个新副本。示例见`int8inc()`。（这是函数可以安全地修改按引用传递的输入的唯一情况。）

更多细节请参见CREATE AGGREGATE命令。

32.11. 用户定义类型

如第 32.2 节中所述，PostgreSQL可以扩展以支持新的数据类型。本节描述如何定义新的基础类型，也就是在SQL语言层之下定义的数据类型。创建新的基础类型需要用底层语言（通常是 C）实现操作该类型的函数。

本节中的示例位于源码分发包的`src/tutorial`目录中的 `complex.sql`和`complex.c`。关于如何运行这些示例，请参见该目录中的`README`文件。

用户定义类型必须始终有输入和输出函数。
这些函数决定该类型以字符串形式出现（供用户输入和向用户输出）的方式，
以及该类型在内存中如何组织。输入函数接受一个以空字符结尾的字符串作为参数，

并返回该类型的内部（内存中）表示。输出函数接受该类型的内部表示作为参数，
并返回一个以空字符结尾的字符串。如果我们希望该类型除了存储之外还能做别的事情，
就必须提供额外的函数来实现我们希望该类型支持的各种操作。

假设我们要定义一种表示复数的类型complex。在内存中表示复数的一种自然方式是下面这个 C 结构体：

```
typedef struct Complex {
    double      x;
    double      y;
} Complex;
```

我们需要把它做成按引用传递的类型，因为它太大了，无法放进单个 Datum值中。

作为该类型的外部字符串表示，我们选择形如(x,y)的字符串。

输入和输出函数通常不难编写，尤其是输出函数。但在定义该类型的外部字符串表示时，
请记住，你最终必须为该表示编写一个完整而健壮的解析器作为输入函数。例如：

```
PG_FUNCTION_INFO_V1(complex_in);

Datum
complex_in(PG_FUNCTION_ARGS)
{
    char       *str = PG_GETARG_CSTRING(0);
    double      x,
                y;
    Complex    *result;

    if (sscanf(str, " ( %lf , %lf )", &x, &y) != 2)
        ereport(ERROR,
                (errcode(ERRCODE_INVALID_TEXT_REPRESENTATION),
                 errmsg("invalid input syntax for complex: \"%s\"",
                        str)));

    result = (Complex *) palloc(sizeof(Complex));
    result->x = x;
    result->y = y;
    PG_RETURN_POINTER(result);
}
```

输出函数可以简单地写成：

```
PG_FUNCTION_INFO_V1(complex_out);

Datum
complex_out(PG_FUNCTION_ARGS)
{
    Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
    char       *result;

    result = (char *) palloc(100);
    snprintf(result, 100, "(%g,%g)", complex->x, complex->y);
    PG_RETURN_CSTRING(result);
}
```

应当注意让输入函数和输出函数互为逆运算。如果不是这样，
当你需要把数据转储到文件中再读回时，就会遇到严重问题。这在涉及浮点数时尤为常见。

可选地，用户定义类型还可以提供二进制输入和输出例程。二进制 I/O 通常比文本 I/O 更快，但可移植性较差。与文本 I/O 一样，外部二进制表示的精确定义完全由你决定。大多数内置数据类型都尽量提供与机器无关的二进制表示。对于`complex`，我们将借助类型`float8`的二进制 I/O 转换器：

```
PG_FUNCTION_INFO_V1(complex_recv);

Datum
complex_recv(PG_FUNCTION_ARGS)
{
    StringInfo  buf = (StringInfo) PG_GETARG_POINTER(0);
    Complex    *result;

    result = (Complex *) palloc(sizeof(Complex));
    result->x = pq_getmsgfloat8(buf);
    result->y = pq_getmsgfloat8(buf);
    PG_RETURN_POINTER(result);
}

PG_FUNCTION_INFO_V1(complex_send);

Datum
complex_send(PG_FUNCTION_ARGS)
{
    Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
    StringInfoData buf;

    pq_begintypsend(&buf);
    pq_sendfloat8(&buf, complex->x);
    pq_sendfloat8(&buf, complex->y);
    PG_RETURN_BYTEA_P(pq_endtypsend(&buf));
}
```

要定义`complex`类型，我们需要在创建该类型之前先创建用户定义的 I/O 函数：

```
CREATE FUNCTION complex_in(cstring)
    RETURNS complex
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_out(complex)
    RETURNS cstring
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_recv(internal)
   RETURNS complex
   AS 'filename'
   LANGUAGE C IMMUTABLE STRICT;

CREATE FUNCTION complex_send(complex)
   RETURNS bytea
```

```
   AS 'filename'
   LANGUAGE C IMMUTABLE STRICT;
```

注意，输入和输出函数的声明必须引用尚未定义的类型。这是允许的，
但会产生一些可以忽略的警告消息。输入函数必须放在前面。

最后，我们可以声明该数据类型：

```
CREATE TYPE complex (
   internallength = 16,
   input = complex_in,
   output = complex_out,
   receive = complex_recv,
   send = complex_send,
   alignment = double
);
```

当你定义一种新的基本类型时，PostgreSQL 会自动提供对该类型数组的支持。由于历史原因，
该数组类型的名称与基本类型相同，只是在前面加一个下划线字符（_）。

一旦该数据类型存在，我们就可以声明额外的函数，为该数据类型提供有用的操作。
随后可以在这些函数之上定义操作符；如果需要，
还可以创建操作符类以支持该数据类型的索引。这些附加层会在后续各节中讨论。

如果你的数据类型的值（在内部形式上）可能超过几百字节，就应该让该数据类型支持
TOAST（见第 50.2 节）。为此，内部表示必须遵循可变长度数据的标准布局：前四个字节必须是
一个`int32`，其中包含该 datum 的总字节长度（包括其自身）。操作该数据类型的 C 函数必须注
意使用 `PG_DETOAST_DATUM`对传给它们的任何已 TOAST 的值先做解包处理。（这一细节通常通
过定义特定于该类型的 `GETARG`宏来隐藏。）然后，在执行 `CREATE TYPE`命令时，把内部长度指
定为 `variable`并选择合适的存储选项。

更多细节见CREATE TYPE命令的说明。

32.12. 用户定义操作符

每个操作符都是对执行实际工作的底层函数调用的一种“语法糖”；因此，
你必须先创建底层函数，才能创建操作符。不过，操作符并不仅仅是语法糖，
因为它还携带一些额外信息，可帮助查询规划器优化使用该操作符的查询。
下一节将专门解释这些附加信息。

PostgreSQL支持左一元、右一元和二元操作符。操作符可以被重载；也就是说，
同一个操作符名可以用于不同的操作符，而这些操作符具有不同数量和类型的操作数。
执行查询时，系统会根据提供的操作数数量和类型确定应调用哪个操作符。

下面是创建一个用于将两个复数相加的操作符的示例。假定我们已经创建了类型`complex`
的定义（见第 32.11 节）。首先需要一个完成实际工作的函数，然后就可以定义操作符：

```
CREATE FUNCTION complex_add(complex, complex)
    RETURNS complex
    AS 'filename', 'complex_add'
    LANGUAGE C IMMUTABLE STRICT;

CREATE OPERATOR + (
    leftarg = complex,
```

```
    rightarg = complex,
    procedure = complex_add,
    commutator = +
);
```

现在我们就可以执行下面这样的查询：

```
SELECT (a + b) AS c FROM test_complex;

        c
-----------------
 (5.2,6.05)
 (133.42,144.95)
```

上面展示了如何创建一个二元操作符。要创建一元操作符，只需省略 leftarg（左一元操作符）或 rightarg（右一元操作符）之一。procedure 子句和参数子句是 CREATE OPERATOR 中唯一必需的项。示例中出现的 commutator 子句，则是给查询优化器的一个可选提示。关于 commutator 以及其他优化器提示的更多细节，会在下一节说明。

32.13. 操作符优化信息

PostgreSQL 的操作符定义可以包含若干可选子句，用来告诉系统该操作符行为的一些有用信息。只要适用，就应当提供这些子句，因为它们可以显著加快使用该操作符的查询执行速度。但如果你提供了这些信息，就必须确保它们是正确的！错误地使用优化子句可能导致服务器进程崩溃、输出出现隐蔽错误，或者引发其他糟糕后果。如果你拿不准，完全可以省略某个优化子句；唯一的后果，只是查询可能会比本来需要的更慢。

未来版本的 PostgreSQL 可能会加入更多优化子句。这里描述的是 8.1.23 版本所理解的全部子句。

32.13.1. COMMUTATOR

如果给出 COMMUTATOR 子句，它指定一个与正在定义的操作符互为交换子的操作符。若对所有可能的输入值 x、y，都有 (x A y) 等于 (y B x)，则称操作符 A 是操作符 B 的交换子。注意，B 也同样是 A 的交换子。例如，对某种特定数据类型而言，< 和 > 通常互为交换子，而操作符 + 通常与其自身可交换。但操作符 - 通常并不与任何操作符可交换。

一个可交换操作符的左操作数类型，与其交换子的右操作数类型相同，反之亦然。因此，PostgreSQL 只需知道交换子操作符的名称，就可以查找出该交换子，而这也正是 COMMUTATOR 子句中所需提供的全部内容。

对于将用于索引和连接子句的操作符，提供交换子信息至关重要，因为这样查询优化器才能把这类子句“翻转”成不同计划类型所需的形式。例如，考虑这样一个查询，其 WHERE 子句形如 tab1.x = tab2.y，其中 tab1.x 和 tab2.y 都是某种用户定义类型，并假设 tab2.y 上建有索引。除非优化器能够确定如何把该子句翻转成 tab2.y = tab1.x，否则它就无法生成索引扫描，因为索引扫描机制要求传给它的操作符左侧必须是已建立索引的列。PostgreSQL 不会仅凭假设就认为这种转换有效；= 操作符的创建者必须通过为该操作符标记交换子信息，明确声明这种转换是有效的。

定义一个与自身可交换的操作符时，直接定义即可。但定义一对可交换操作符时，情况就稍微复杂一些：第一个要定义的操作符如何引用另一个尚未定义的操作符呢？这个问题有两种解决办法：

- 一种办法是在定义第一个操作符时省略 COMMUTATOR 子句，然后在第二个操作符的定义中提供该子句。由于 PostgreSQL 知道可交换操作符是成对出现的，因此它在看到第二个定义时，会自动回头补全第一个定义中缺少的 COMMUTATOR 子句。
- 另一种更直接的办法是在两个定义中都包含 COMMUTATOR 子句。当 PostgreSQL 处理第一个定义并发现 COMMUTATOR 引用了一个不存在的操作符时，系统会在系统目录中为该操作符建立一个占位项。这个占位项只有操作符名称、左右操作数类型和结果类型包含有效数据，因为这就是 PostgreSQL 此时能够推断出的全部信息。第一个操作符的目录项将链接到这个占位项。之后，当你定义第二个操作符时，系统会用第二个定义中的附加信息更新该占位项。如果在占位操作符补全之前尝试使用它，就只会得到一条错误消息。

32.13.2. NEGATOR

如果给出 NEGATOR 子句，它指定一个与正在定义的操作符互为求反器的操作符。若操作符 A 和 B 都返回布尔结果，并且对所有可能的输入 x、y，都有 (x A y) 等于 NOT (x B y)，则称 A 是 B 的求反器。注意，B 也同样是 A 的求反器。例如，对大多数数据类型而言，< 和 >= 构成一对求反器。一个操作符永远不可能合法地成为其自身的求反器。

与交换子不同，一对一元操作符完全可能合法地被标记为彼此的求反器；这意味着对所有 x，都有 (A x) 等于 NOT (B x)，或者对右一元操作符有等价的关系。

一个操作符的求反器必须与待定义操作符具有相同的左操作数类型和/或右操作数类型，因此与 COMMUTATOR 一样，在 NEGATOR 子句中只需给出操作符名称即可。

提供求反器对查询优化器很有帮助，因为它允许把 NOT (x = y) 这样的表达式简化为 x <> y。这种情况比你想象得更常见，因为 NOT 操作可能会作为其他重排的结果被插入进来。

可以使用上面解释的定义交换子对的相同方法，来定义成对的求反器操作符。

32.13.3. RESTRICT

如果给出 RESTRICT 子句，它指定该操作符的限制选择率估算函数。（注意，这里是函数名，而不是操作符名。）RESTRICT 子句只对返回 boolean 的二元操作符有意义。限制选择率估算器的作用，是针对当前操作符和某个特定常量值，猜测一张表中有多少比例的行会满足如下形式的 WHERE 子句条件：

```
column OP constant
```

这会帮助优化器大致了解具有这种形式的 WHERE 子句将淘汰多少行。（你可能会问：如果常量在左边会怎样？嗯，这正是 COMMUTATOR 的作用之一.....）

编写新的限制选择率估算函数远远超出了本章的范围，不过幸运的是，对于你自己的很多操作符，通常都可以直接使用系统提供的某个标准估算器。标准的限制选择率估算器如下：

eqsel 用于 =
neqsel 用于 <>
scalarltsel 用于 < 或 <=
scalargtsel 用于 > 或 >=

这些分类可能看起来有点奇怪，但仔细想想就会发现它们是有道理的。= 通常只会接受表中很小一部分行；<> 通常只会拒绝很小一部分行。< 接受的比例取决于给定常量落在该表列的值范围中的什么位置（恰好，这正是 ANALYZE 收集并提供给选择率估算器的信息）。对于相同的比较常量，<= 接受的比例会比 < 略大，但二者足够接近，不值得区分，尤其是无论如何我们很可能也只能做出粗略猜测。类似的说法也适用于 > 和 >=。

对于选择率非常高或非常低的操作符，即使它们实际上并不是真正的相等或不等比较，
你也常常可以勉强使用 eqsel 或 neqsel。例如，几何类型中的近似相等操作符就使用 eqsel，
其依据是它们通常只会匹配表中很小一部分项。

对于那些能够以某种合理方式转换为数值标量、从而可进行范围比较的数据类型，你可以使用 scalarltsel 和 scalargtsel。如果可能的话，请把该数据类型加入函数 convert_to_scalar()（位于 src/backend/utils/adt/selfuncs.c）所理解的范围中。（最终，这个函数应被通过 pg_type 系统目录某一列标识的、按数据类型划分的函数所取代；但这件事目前还没有发生。）如果不这样做，系统仍然可以工作，
但优化器的估算效果就不会像本可达到的那样好。

在 src/backend/utils/adt/geo_selfuncs.c 中，还为几何操作符提供了其他选择率估算函数：areasel、positionsel 和 contsel。截至目前，这些函数都还只是桩实现，但你也许仍会想使用它们（或者更好的是，改进它们）。

32.13.4. JOIN

如果给出 JOIN 子句，它指定该操作符的连接选择率估算函数。（注意，这里是函数名，而不是操作符名。）JOIN 子句只对返回 boolean 的二元操作符有意义。连接选择率估算器的作用，是针对当前操作符，猜测两张表中有多少比例的行对会满足如下形式的 WHERE 子句条件：

```
table1.column1 OP table2.column2
```

与 RESTRICT 子句一样，这会极大地帮助优化器判断在若干可能的连接顺序中，哪一种预计工作量最小。

与前面一样，本章不会尝试解释如何编写连接选择率估算函数，
而只是建议你在适用时使用某个标准估算器：

eqjoinsel 用于 =
neqjoinsel 用于 <>
scalarltjoinsel 用于 < 或 <=
scalargtjoinsel 用于 > 或 >=
areajoinsel 用于基于二维面积的比较
positionjoinsel 用于基于二维位置的比较
contjoinsel 用于基于二维包含关系的比较

32.13.5. HASHES

如果存在 HASHES 子句，它会告知系统：在基于该操作符的连接中，可以使用哈希连接方法。HASHES 只对返回 boolean 的二元操作符有意义；在实践中，该操作符最好是某种数据类型上的相等关系。

哈希连接背后的假设是：只有当左值和右值被哈希到同一个哈希码时，连接操作符才有可能返回 true。如果两个值被放进不同的哈希桶，连接过程就根本不会去比较它们，
这实际上隐含假定连接操作符的结果必定为 false。因此，对于那些并不表示相等关系的操作符，指定 HASHES 永远没有意义。

若要被标记为 HASHES，该连接操作符必须出现在某个哈希索引操作符类中。创建操作符时并不会强制这一点，因为那时引用它的操作符类当然还不存在。但如果没有这样的操作符类存在，
运行时尝试在哈希连接中使用该操作符就会失败。
系统需要通过该操作符类来查找与该操作符输入数据类型相对应的、
特定于数据类型的哈希函数。当然，在创建该操作符类之前，你还必须先提供合适的哈希函数。

准备哈希函数时应格外小心，因为它有一些依赖机器的方式可能会导致结果不正确。例如，如果你的数据类型是某种结构体，其中可能包含无意义的填充位，那就不能简单地把整个结构体传给 hash_any。（除非你编写了其他操作符和函数，以确保这些未使用位始终为零，而这正是推荐的策略。）再比如，在符合 IEEE 浮点标准的机器上，负零和正零是不同的值（位模式不同），但它们被定义为比较相等。如果某个浮点值可能包含负零，就需要采取额外步骤，确保它生成与正零相同的哈希值。

> **注意**
>
> 作为可哈希连接操作符底层实现的函数，必须标记为 immutable 或 stable。如果它是 volatile，系统就永远不会尝试把该操作符用于哈希连接。

> **注意**
>
> 如果某个可哈希连接操作符的底层函数被标记为 strict，那么该函数还必须是完备的：也就是说，对于任意两个非空输入，它都应返回 true 或 false，而绝不能返回 null。如果不遵循这条规则，对 IN 操作进行哈希优化时可能会产生错误结果。（具体来说，IN 可能会在按标准本应返回 null 的地方返回 false；或者抛出一条错误，抱怨它没有为 null 结果做好准备。）

32.13.6. MERGES (SORT1, SORT2, LTCMP, GTCMP)

如果存在 MERGES 子句，它会告知系统：在基于该操作符的连接中，可以使用归并连接方法。MERGES 只对返回 boolean 的二元操作符有意义；在实践中，该操作符最好是某种数据类型或某对数据类型上的相等关系。

归并连接的基本思想，是先把左表和右表分别排序，然后同步扫描它们。因此，两种数据类型都必须能够被完全排序，而连接操作符必须只能在那对值位于排序次序中“同一位置”时才成功。实际效果上，这意味着连接操作符必须表现得像相等比较一样。但与哈希连接不同（哈希连接要求左右数据类型最好相同，或者至少按位等价），只要两种不同的数据类型在逻辑上兼容，就可以对它们进行归并连接。例如，smallint 与 integer 之间的相等操作符就是可归并连接的。我们只需要能把两种数据类型都带入逻辑兼容排序序列的排序操作符即可。

归并连接的执行要求系统能够识别出与该归并连接相等操作符相关的四个操作符：左操作数数据类型的小于比较、右操作数数据类型的小于比较、两种数据类型之间的小于比较，以及两种数据类型之间的大于比较。（如果可归并连接的操作符具有两个不同的操作数数据类型，这实际上是四个不同的操作符；但当两个操作数类型相同时，三个小于操作符是同一个操作符。）可以通过名称分别指定这些操作符，即分别使用 SORT1、SORT2、LTCMP 和 GTCMP 选项。在指定 MERGES 时，如果省略了其中任何一个，系统将分别填入默认名称 <、<、<、>。此外，如果这四个操作符选项中任何一个出现，就会认为隐含了 MERGES，因此可以只指定其中一部分，而让系统补全其余部分。

这四个比较操作符的操作数数据类型可以从可归并连接操作符的操作数类型推断出来，因此与 COMMUTATOR 一样，这些子句中只需给出操作符名称即可。除非你选用的操作符名称比较特殊，否则只需写 MERGES 并让系统补全细节就足够了。（与 COMMUTATOR 和 NEGATOR 一样，如果你碰巧先定义了相等操作符，系统也能够建立占位操作符条目。）

对于你标记为可归并连接的操作符，还有一些附加限制。这些限制目前不会被 CREATE OPERATOR 检查，但如果其中任何一条不成立，在使用该操作符时可能会出错：

- 一个可归并连接的相等操作符必须拥有一个可归并连接的交换子（如果两个操作数数据类型相同，则为它自身；如果不同，则应是相关的相等操作符）。
- 如果存在一个可归并连接的操作符关联某两个数据类型 A 和 B，又存在另一个可归并连接的操作符关联 B 和任何第三种数据类型 C，那么 A 和 C 也必须有一个可归并连接的操作符；换句话说，可归并连接操作符的存在必须是可传递的。
- 如果你指定的四个比较操作符没有以兼容的方式对数据值排序，运行时会出现奇怪的结果。

注意

作为可归并连接操作符底层实现的函数，必须标记为 immutable 或 stable。如果它是 volatile，系统就永远不会尝试把该操作符用于归并连接。

注意

在 PostgreSQL 7.3 之前的版本中，没有 MERGES 这个简写形式：要创建一个可归并连接的操作符，必须显式地同时写出 SORT1 和 SORT2。而且，当时也不存在 LTCMP 和 GTCMP 选项；那两个操作符的名称被硬编码为 < 和 >。

32.14. 索引扩展接口

到目前为止所描述的过程使我们能够定义新类型、新函数以及新操作符。然而，我们还不能在一种新数据类型的列上定义索引。为此，必须为该新数据类型定义一个操作符类。本节稍后将用一个示例说明这一概念：为 B-树索引方法定义一个新的操作符类，以便按绝对值升序存储和排序复数。

注意

在PostgreSQL 7.3 之前的版本中，要创建用户定义的操作符类，必须手工向系统目录pg_amop、pg_amproc和pg_opclass中添加条目。这种做法现在已被弃用，取而代之的是使用CREATE OPERATOR CLASS，后者是一种更简单、更不易出错的创建所需目录条目的方式。

32.14.1. 索引方法和操作符类

pg_am 表为每个索引方法（内部称为访问方法）保存一行。对表进行常规访问的支持内置于 PostgreSQL 中，但所有索引方法都在 pg_am 中描述。可以定义所需的接口例程，然后在 pg_am 中创建一行，从而添加新的索引方法 – 但这超出了本章的范围（参见第 48 章）。

索引方法的例程并不直接知道它将处理哪些数据类型。相反，一个操作符类标识了索引方法在处理特定数据类型时需要使用的那组操作。之所以称为操作符类，是因为它指定的一项内容就是可与索引一起使用的 WHERE 子句操作符集合（也就是能被转换成索引扫描条件的操作符）。操作符类还可以指定索引方法内部操作所需的一些支持函数，但这些函数并不直接对应任何可与索引一起使用的 WHERE 子句操作符。

可以为同一种数据类型和索引方法定义多个操作符类。
这样就能为一种数据类型定义多套索引语义。例如，一个 B-树索引要求为其处理的每一种数据类型定义一种排序顺序。对于复数数据类型，也许既需要一个按复数绝对值排序的 B-树操作符类，也需要另一个按实部排序的操作符类，等等。通常，其中一个操作符类会被视为最常用，
并标记为该数据类型在该索引方法上的默认操作符类。

同一个操作符类名可以用于多个不同的索引方法（例如，B-树和哈希索引方法都有名为 `int4_ops` 的操作符类），但每一个这样的类都是独立实体，必须分别定义。

32.14.2. 索引方法策略

与操作符类关联的操作符通过“策略号”来标识，
用以表示每个操作符在其操作符类上下文中的语义。例如，B-
树对键施加了严格的从小到大的顺序，因此像“小于”和“大于等于”这样的操作符，对 B-树来说就很重要。由于 PostgreSQL 允许用户定义操作符，PostgreSQL 不能仅凭操作符名称（例如 `<` 或 `>=`）就判断它是哪一类比较。取而代之的是，索引方法定义了一组“策略”，可以把它们看成是广义的操作符。每个操作符类都会说明，对于某种特定数据类型和某种索引语义解释，
每一种策略分别对应哪个实际操作符。

B-树索引方法定义了五种策略，如表 32.2所示。

表 32.2. B-树策略

Operation	Strategy Number
less than	1
less than or equal	2
equal	3
greater than or equal	4
greater than	5

哈希索引只支持等值比较，因此它们只使用一种策略，如表 32.3所示。

表 32.3. 哈希策略

Operation	Strategy Number
equal	1

R 树索引表示二维空间中的关系。它们使用十二种策略，如表 32.4所示。其中四个是真正的二维测试（重叠、相同、包含、被包含），四个只考虑 X 方向，另外四个则在 Y 方向上提供相同测试。

表 32.4. R 树策略

Operation	Strategy Number
严格位于左侧	1
不延伸到右侧	2
重叠	3
不延伸到左侧	4
严格位于右侧	5
相同	6

Operation	Strategy Number
包含	7
被包含	8
不延伸到上方	9
严格位于下方	10
严格位于上方	11
不延伸到下方	12

GiST 索引更加灵活：它们根本没有固定的策略集合。相反，每个特定 GiST 操作符类中负责“一致性”检查的支持例程会按自己的方式解释策略号。

注意，所有策略操作符都返回布尔值。实际上，所有被定义为索引方法策略的操作符都必须返回 `boolean` 类型，因为要与索引配合使用，它们必须出现在 `WHERE` 子句的顶层。

顺便说一下，`pg_am` 中的 `amorderstrategy`列表明该索引方法是否支持有序扫描。零表示不支持；如果支持，`amorderstrategy`就是与排序操作符对应的策略号。例如，B-树的 `amorderstrategy` = 1，也就是它的“小于”策略号。

32.14.3. 索引方法支持例程

仅靠策略信息通常不足以让系统知道如何使用索引。实际上，
索引方法还需要额外的支持例程才能工作。例如，B-树索引方法必须能够比较两个键，
并判断其中一个是大于、等于还是小于另一个。类似地，R 树索引方法必须能够计算矩形的交集、并集和尺寸。这些操作并不对应 SQL 命令条件中使用的操作符；它们是索引方法内部使用的管理例程。

与策略一样，操作符类会标识对于给定的数据类型和语义解释，
应由哪些具体函数承担这些角色。索引方法定义它需要的函数集合，
而操作符类则会通过为函数分配由索引方法规定的“支持函数号”来标识正确的函数。

B-tree 要求一个支持函数，如表 32.5所示。

表 32.5. B-树支持函数

函数	支持号
比较两个键，并返回一个小于零、等于零或大于零的整数，用以表示第一个键是小于、等于还是大于第二个键	1

哈希索引同样只需要一个支持函数，如表 32.6所示。

表 32.6. 哈希支持函数

函数	支持号
计算一个键的哈希值	1

R 树索引要求三个支持函数，如表 32.7所示。

表 32.7. R 树支持函数

函数	支持号
union	1

函数	支持号
intersection	2
size	3

GiST 索引要求七个支持函数，如表 32.8所示。

表 32.8. GiST 支持函数

函数	支持号
consistent	1
union	2
compress	3
decompress	4
penalty	5
picksplit	6
equal	7

与策略操作符不同，支持函数返回的是特定索引方法所期望的数据类型；例如，对 B-tree 的比较函数来说，就是一个有符号整数。

32.14.4. 示例

现在我们已经了解了这些基本思想，下面给出先前承诺的创建新操作符类示例。
（这个可运行示例位于源码发布包中的`src/tutorial/complex.c`和`src/tutorial/complex.sql`。）该操作符类封装了一组按绝对值顺序对复数排序的操作符，
因此我们把它命名为`complex_abs_ops`。首先，我们需要一组操作符。
定义操作符的过程已经在第 32.12 节中讨论过。对于 B-树上的操作符类，我们需要如下操作符：

- 绝对值小于（策略 1）
- 绝对值小于等于（策略 2）
- 绝对值等于（策略 3）
- 绝对值大于等于（策略 4）
- 绝对值大于（策略 5）

定义一组相关比较操作符时，最不容易出错的方式是先编写 B-树比较支持函数，再把其他函数写成围绕该支持函数的一行包装器函数。这样可以降低在边界情况下得到不一致结果的概率。
按照这种方法，我们首先编写：

```
#define Mag(c)  ((c)->x*(c)->x + (c)->y*(c)->y)

static int
complex_abs_cmp_internal(Complex *a, Complex *b)
{
    double      amag = Mag(a),
                bmag = Mag(b);

    if (amag < bmag)
        return -1;
    if (amag > bmag)
        return 1;
    return 0;
```

```
}
```

现在，小于函数如下所示：

```
PG_FUNCTION_INFO_V1(complex_abs_lt);

Datum
complex_abs_lt(PG_FUNCTION_ARGS)
{
    Complex    *a = (Complex *) PG_GETARG_POINTER(0);
    Complex    *b = (Complex *) PG_GETARG_POINTER(1);

    PG_RETURN_BOOL(complex_abs_cmp_internal(a, b) < 0);
}
```

其他四个函数的区别只在于它们如何比较内部函数的结果与 0。

接下来，在 SQL 中声明这些函数，以及基于这些函数的操作符：

```
CREATE FUNCTION complex_abs_lt(complex, complex) RETURNS bool
    AS 'filename', 'complex_abs_lt'
    LANGUAGE C IMMUTABLE STRICT;

CREATE OPERATOR < (
   leftarg = complex, rightarg = complex, procedure = complex_abs_lt,
   commutator = > , negator = >= ,
   restrict = scalarltsel, join = scalarltjoinsel
);
```

必须指定正确的交换子和求反器操作符，以及合适的限制选择率与连接选择率函数，
否则优化器无法有效使用索引。注意，小于、等于和大于这几种情况应使用不同的选择率函数。

这里还有几点值得注意：

- 只能有一个名为 = 且两个操作数都为 `complex` 类型的操作符。在这个例子里，我们并没有任何其他 = 操作符可用于 `complex`；但如果我们是在构造一种实际使用的数据类型，可能会希望 = 表示复数的普通相等，而不是绝对值相等。在那种情况下，我们就需要为 `complex_abs_eq` 选用其他操作符名。
- 尽管 PostgreSQL 能处理 SQL 名称相同但参数数据类型不同的函数，C 却只能处理给定名称的一个全局函数。因此，我们不应该把 C 函数简单命名成 `abs_eq` 之类。通常，在 C 函数名中包含数据类型名称是个好习惯，这样就不会与其他数据类型的函数发生冲突。
- 我们原本也可以把该函数的 SQL 名称取为 `abs_eq`，并依靠 PostgreSQL 通过参数数据类型把它与其他同名 SQL 函数区分开。为了让示例保持简单，这里我们让 C 层和 SQL 层的函数使用相同的名称。

下一步是注册 B-树要求的支持例程。实现该例程的 C 示例代码与操作符函数位于同一个文件中。
该函数的声明如下：

```
CREATE FUNCTION complex_abs_cmp(complex, complex)
    RETURNS integer
    AS 'filename'
    LANGUAGE C IMMUTABLE STRICT;
```

现在我们已经有了所需的操作符和支持例程，就可以最终创建操作符类：

```
CREATE OPERATOR CLASS complex_abs_ops
    DEFAULT FOR TYPE complex USING btree AS
        OPERATOR        1       < ,
        OPERATOR        2       <= ,
        OPERATOR        3       = ,
        OPERATOR        4       >= ,
        OPERATOR        5       > ,
        FUNCTION        1       complex_abs_cmp(complex, complex);
```

这样就完成了！现在应该可以在 `complex` 列上创建并使用 B-树索引。

我们本来也可以把操作符项写得更详细一些，例如：

```
        OPERATOR        1       < (complex, complex) ,
```

但是当操作符接受的数据类型与该操作符类所服务的数据类型相同时，就没有必要这样写。

上述示例假定你希望把这个新操作符类设为 `complex` 数据类型的默认 B-树操作符类。如果不是这样，只需省去 `DEFAULT` 这个词。

32.14.5. 跨数据类型的操作符类

到目前为止，我们一直隐含地假定一个操作符类只处理一种数据类型。
虽然某个特定的索引列当然只能有一种数据类型，
但对把被索引列与另一种数据类型的值进行比较的操作建立索引往往也很有用。目前 B-树和 GiST 索引方法支持这一点。

B-树要求每个操作符的左操作数是被索引的数据类型，但右操作数可以是另一种类型。
此时必须存在一个具有匹配签名的支持函数。例如，`bigint`（`int8`）
类型的内建操作符类允许与`int4`和 `int2`进行跨类型比较。可以用下面的定义复制出它：

```
CREATE OPERATOR CLASS int8_ops
DEFAULT FOR TYPE int8 USING btree AS
  -- standard int8 comparisons
  OPERATOR 1 < ,
  OPERATOR 2 <= ,
  OPERATOR 3 = ,
  OPERATOR 4 >= ,
  OPERATOR 5 > ,
  FUNCTION 1 btint8cmp(int8, int8) ,

  -- cross-type comparisons to int2 (smallint)
  OPERATOR 1 < (int8, int2) ,
  OPERATOR 2 <= (int8, int2) ,
  OPERATOR 3 = (int8, int2) ,
  OPERATOR 4 >= (int8, int2) ,
  OPERATOR 5 > (int8, int2) ,
  FUNCTION 1 btint82cmp(int8, int2) ,

  -- cross-type comparisons to int4 (integer)
  OPERATOR 1 < (int8, int4) ,
  OPERATOR 2 <= (int8, int4) ,
  OPERATOR 3 = (int8, int4) ,
  OPERATOR 4 >= (int8, int4) ,
  OPERATOR 5 > (int8, int4) ,
```

```
FUNCTION 1 btint84cmp(int8, int4) ;
```

注意，此定义“重载”了操作符策略号和支持函数号。
只要同一编号的每个实例具有不同的右操作数数据类型，就允许这样做（仅限 B-树操作符类）。
非跨类型的实例是该操作符类的默认或主要操作符。

GiST 索引不允许重载策略号或支持函数号，但仍然可以通过为每个需要支持的操作符分配不同的策略号，来实现支持多种右操作数数据类型的效果。consistent
支持函数必须根据策略号判断自己该做什么，并且必须准备好接受相应数据类型的比较值。

32.14.6. 系统对操作符类的依赖

PostgreSQL利用操作符类来从多方面推断操作符的属性，而不仅仅是判断它们能否用于索引。
因此，即便你并不打算为自己的数据类型列建立索引，也可能会想创建操作符类。

特别地，ORDER BY和DISTINCT等 SQL 特性要求对值的比较和排序。为了在用户定义的数据类型上实现这些特性，PostgreSQL会为数据类型查找默认 B-树操作符类。这个操作符类的“相等”成员定义了用于GROUP BY和DISTINCT的值的等值概念，而该操作符类施加的排序顺序定义了默认的ORDER BY顺序。

用户定义类型的数组比较也依赖于该类型默认 B-树操作符类所定义的语义。

如果一种数据类型没有默认的 B-树操作符类，系统就会查找默认的哈希操作符类。但由于这类操作符类只提供等值语义，因此在实践中它只足以支持数组相等比较。

如果某种数据类型没有默认操作符类，而你又试图将这些 SQL 特性用于该数据类型，就会得到类似“无法识别排序操作符”这样的错误。

> **注意**
>
> 在版本 7.4 以前的PostgreSQL中，排序和分组操作将隐式地使用名为=、<以及>的操作符。新的依赖于默认操作符类的行为避免了对具有特定名字的操作符行为作出任何假设。

32.14.7. 操作符类的特殊特性

还有两个操作符类的特殊特性我们尚未讨论，主要是因为它们对最常用的索引方法没有用处。

通常，把一个操作符声明为操作符类（或操作符族）的成员，
意味着索引方法能够正好检索出满足使用该操作符的 WHERE 条件的行集。例如：

```
SELECT * FROM table WHERE integer_column < 4;
```

可以由整数列上的 B-tree 索引精确满足。但也存在这样的情况：索引只是查找匹配行的不精确向导。例如，如果一个 R 树索引只为几何对象存储包围盒，那么它就无法精确满足测试多边形等非矩形对象之间重叠的 WHERE 条件。不过，我们可以用索引找出包围盒与目标对象的包围盒重叠的对象，然后只对索引找到的对象做精确的重叠测试。如果适用这种情形，
就称该索引对该操作符是“有损”的。这种情况下，我们可以在 CREATE OPERATOR CLASS命令的OPERATOR 子句中加上RECHECK。如果索引保证返回所有需要的行（外加可能的一些额外行，它们可以通过执行原始操作符调用来排除），那么RECHECK就是有效的。

再次考虑只在索引中存储多边形等复杂对象的包围盒的情况。此时，
在索引条目中存储整个多边形没有多少价值，不如只存储一个更简单的对象，其类型为box。
这种情况由STORAGE选项表达，该选项位于CREATE OPERATOR CLASS：可以写成如下形式：

```
CREATE OPERATOR CLASS polygon_ops
    DEFAULT FOR TYPE polygon USING gist AS
        ...
        STORAGE box;
```

目前，只有 GiST 索引方法支持与列数据类型不同的STORAGE类型。GiST 的compress和decompress支持函数在使用STORAGE时必须处理数据类型转换。

第 33 章 触发器

本章提供有关编写触发器函数的一般信息。触发器函数可以用大多数可用的过程语言编写，包括 PL/pgSQL（第 36 章）、PL/Tcl（第 37 章）、PL/Perl（第 38 章）和 PL/Python（第 39 章）。读完本章后，你还应查阅所用过程语言对应的章节，了解用该语言编写触发器的语言相关细节。

也可以用 C 编写触发器函数，不过大多数人会觉得使用某种过程语言更容易。目前还不能用普通 SQL 函数语言编写触发器函数。

33.1. 触发器行为概述

触发器是这样一种规定：每当执行某种特定类型的操作时，
数据库就应自动执行一个特定的函数。触发器可以定义为在任何 `INSERT`、`UPDATE` 或 `DELETE` 操作之前或之后执行，执行方式既可以是每个被修改的行一次，也可以是每条 SQL 语句一次。一旦触发器事件发生，触发器函数就会在恰当的时机被调用以处理该事件。

必须先定义触发器函数，之后才能创建触发器本身。
触发器函数必须声明为不接受参数并返回类型 `trigger` 的函数。（触发器函数通过特殊传入的 `TriggerData` 结构体接收输入，而不是通过普通函数参数。）

创建出合适的触发器函数后，就可以使用CREATE TRIGGER创建触发器。同一个触发器函数可以供多个触发器使用。

PostgreSQL 同时提供每行触发器和每语句触发器。对于每行触发器，触发它的语句所影响的每一行都会调用一次触发器函数。相反，每语句触发器只会在执行相应语句时调用一次，
不管该语句影响了多少行。特别是，即使某个语句一行也没有影响，
仍会执行所有适用的每语句触发器。这两类触发器有时也分别称为行级触发器和语句级触发器。

触发器还分为之前触发器和之后触发器。语句级 BEFORE 触发器自然在语句开始做任何事之前触发，而语句级 AFTER 触发器在语句的最终末尾触发。行级 BEFORE 触发器在对某个特定行执行操作之前立即触发，而行级 AFTER 触发器在语句末尾触发（但早于任何语句级 AFTER 触发器）。

由每语句触发器调用的触发器函数应始终返回 `NULL`。由每行触发器调用的触发器函数则可以根据需要向调用它的执行器返回一个表行（即类型 `HeapTuple` 的值）。在操作之前触发的行级触发器可以作出如下选择：

- 它可以返回 `NULL`，以跳过对当前行的操作。这会指示执行器不要执行引发该触发器的行级操作（即对某个具体表行的插入或修改）。
- 仅对于行级 `INSERT` 和 `UPDATE` 触发器，返回的行会成为将要插入的行，或替换正在更新的行。这使得触发器函数可以修改即将插入或更新的行。

不打算引发上述任何一种行为的行级 BEFORE 触发器，必须注意把传入的行原样作为结果返回（也就是说，对于 `INSERT` 和 `UPDATE` 触发器返回 `NEW` 行，对于 `DELETE` 触发器返回 `OLD` 行）。

对于在操作之后触发的行级触发器，其返回值会被忽略，因此它们可以返回 `NULL`。

如果为同一关系上的同一事件定义了多个触发器，它们会按触发器名称的字母顺序触发。对于 BEFORE 触发器，每个触发器返回的、可能已经被修改过的行都会成为下一个触发器的输入。如果任何 BEFORE 触发器返回 `NULL`，则会放弃对该行执行该操作，并且后续触发器（针对该行）都不会再触发。

通常，行级 BEFORE 触发器用于检查或修改将要插入或更新的数据。例如，BEFORE 触发器可以用来把当前时间写入某个 `timestamp` 列，或者检查该行的两个元素是否一致。行级

AFTER 触发器最适合用于把更新传播到其他表，或对其他表执行一致性检查。之所以这样分工，是因为 AFTER 触发器可以确定自己看到的是该行的最终值，而 BEFORE 触发器不能；它之后可能还会有其他 BEFORE 触发器触发。如果没有特定理由把触发器做成 BEFORE 或 AFTER，BEFORE 的情形效率更高，因为有关该操作的信息不必保存到语句结束。

如果触发器函数执行 SQL 命令，那么这些命令可能会再次触发触发器。这就是所谓的级联触发器。对级联层数没有直接限制。级联还有可能导致同一个触发器被递归调用；例如，一个 `INSERT` 触发器可能执行一条向同一个表再插入一行的命令，从而导致该 `INSERT` 触发器再次触发。因此，在这种场景下，避免无限递归是触发器编写者自己的责任。

在定义触发器时，可以为它指定参数。在触发器定义中包含参数，
是为了让需求相似的不同触发器能够调用同一个函数。举例来说，可以有一个通用触发器函数，
它接受两个列名作为参数，把当前用户写入其中一个列，把当前时间戳写入另一个列。
只要编写得当，这个触发器函数就应独立于它所作用的具体表。因此，
同一个函数可用于任何具有适当列的表上的 `INSERT` 事件，例如自动跟踪某个事务表中记录的创建。如果把它定义成 `UPDATE` 触发器，还可以用来跟踪最近更新事件。trigger.

每种支持触发器的编程语言都有自己的方法，使触发器输入数据可供触发器函数使用。
这些输入数据包括触发器事件的类型（例如 `INSERT` 或 `UPDATE`），以及 `CREATE TRIGGER` 中列出的任何参数。对于行级触发器，输入数据还包括 `INSERT` 和 `UPDATE` 触发器的 `NEW` 行，以及/或者 `UPDATE` 和 `DELETE` 触发器的 `OLD` 行。语句级触发器目前无法检查该语句修改的各个行。

33.2. 数据更改的可见性

如果你在触发器函数中执行 SQL 命令，而这些命令访问的是触发器所作用的表，那么你需要了解数据可见性规则，因为它们决定了这些 SQL 命令能否看到导致触发器触发的数据更改。简而言之：

- 语句级触发器遵循简单的可见性规则：语句所做的任何更改对在语句之前调用的语句级触发器都不可见，而所有修改对语句级 AFTER 触发器都可见。
- 导致触发器触发的数据更改（插入、更新或删除）对于在行级 BEFORE 触发器中执行的 SQL 命令自然是不可见的，因为它尚未发生。
- 然而，在行级 BEFORE 触发器中执行的 SQL 命令会看到同一外层命令中先前已处理各行上的数据更改结果。这里必须小心，因为这些更改事件的顺序通常不可预测；一条影响多行的 SQL 命令可能按任意顺序访问这些行。
- 当行级 AFTER 触发器被触发时，外层命令所做的所有数据更改都已经完成，并且对被调用的触发器函数可见。

有关数据可见性规则的更多信息可见第 40.4 节。第 33.4 节中的示例展示了这些规则。

33.3. 用 C 编写触发器函数

本节说明触发器函数接口的底层细节。这些信息仅在用 C 编写触发器函数时才需要。如果你使用更高层语言，这些细节会由系统代为处理。在多数情况下，在用 C 编写触发器之前，应先考虑使用过程语言。每种过程语言的文档都会说明如何用该语言编写触发器。

触发器函数必须使用“版本 1”函数管理器接口。

当函数由触发器管理器调用时，不会向它传递任何普通参数，但会传递一个指向 `TriggerData` 结构体的“上下文”指针。C 函数可以通过执行下列宏来检查自己是否由触发器管理器调用：

```
CALLED_AS_TRIGGER(fcinfo)
```

它会展开为：

```
((fcinfo)->context != NULL && IsA((fcinfo)->context, TriggerData))
```

如果该宏返回真，就可以安全地把 `fcinfo->context` 转换为 `TriggerData *` 类型，并使用其所指向的 `TriggerData` 结构体。函数绝不能修改 `TriggerData` 结构体本身，也不能修改它所指向的任何数据。

`struct TriggerData` is defined in `commands/trigger.h`:

```
typedef struct TriggerData
{
    NodeTag       type;
    TriggerEvent  tg_event;
    Relation      tg_relation;
    HeapTuple     tg_trigtuple;
    HeapTuple     tg_newtuple;
    Trigger      *tg_trigger;
    Buffer        tg_trigtuplebuf;
    Buffer        tg_newtuplebuf;
} TriggerData;
```

其中各成员的定义如下：

`type`

始终是 `T_TriggerData`。

`tg_event`

描述调用该函数的事件。可以使用以下宏来检查 `tg_event`：

`TRIGGER_FIRED_BEFORE(tg_event)`

如果触发器在操作之前触发，则返回真。

`TRIGGER_FIRED_AFTER(tg_event)`

如果触发器在操作之后触发，则返回真。

`TRIGGER_FIRED_FOR_ROW(tg_event)`

如果触发器是因行级事件触发，则返回真。

`TRIGGER_FIRED_FOR_STATEMENT(tg_event)`

如果触发器是因语句级事件触发，则返回真。

`TRIGGER_FIRED_BY_INSERT(tg_event)`

如果触发器是由 `INSERT` 命令触发，则返回真。

`TRIGGER_FIRED_BY_UPDATE(tg_event)`

如果触发器是由 `UPDATE` 命令触发，则返回真。

`TRIGGER_FIRED_BY_DELETE(tg_event)`

如果触发器是由 `DELETE` 命令触发，则返回真。

tg_relation

指向一个描述该触发器所针对关系的结构体。有关此结构体的细节见 utils/rel.h。其中最值得关注的是 tg_relation->rd_att（该关系元组的描述符）和 tg_relation->rd_rel->relname（关系名；其类型不是 char* 而是 NameData；如果你需要该名称的副本，可以用 SPI_getrelname(tg_relation) 得到 char*）。

tg_trigtuple

指向触发该触发器的那一行。这是正在被插入、更新或删除的行。如果该触发器因 INSERT 或 DELETE 而触发，那么如果你不想用另一行替换该行（对 INSERT 而言）或跳过该操作，就应当从函数中返回它。

tg_newtuple

如果触发器因 UPDATE 而触发，则指向该行的新版本；如果因 INSERT 或 DELETE 而触发，则为 NULL。如果事件是 UPDATE，而你不想用另一行替换该行或跳过该操作，就必须从函数中返回它。

tg_trigger

指向一个 Trigger 类型的结构体，该结构体定义在 utils/rel.h 中：

```
typedef struct Trigger
{
    Oid         tgoid;
    char       *tgname;
    Oid         tgfoid;
    int16       tgtype;
    char        tgenabled;
    bool        tgisconstraint;
    Oid         tgconstrrelid;
    Oid         tgconstraint;
    bool        tgdeferrable;
    bool        tginitdeferred;
    int16       tgnargs;
    int16       tgnattr;
    int16      *tgattr;
    char      **tgargs;
} Trigger;
```

其中 tgname 是触发器的名称，tgnargs 是 tgargs 中参数的数量，而 tgargs 是一个指针数组，指向 CREATE TRIGGER 语句中指定的参数。其他成员仅供内部使用。

tg_trigtuplebuf

包含 tg_trigtuple 的缓冲区；如果没有这样的元组，或者它未存储在磁盘缓冲区中，则为 InvalidBuffer。

tg_newtuplebuf

包含 tg_newtuple 的缓冲区；如果没有这样的元组，或者它未存储在磁盘缓冲区中，则为 InvalidBuffer。

触发器函数必须返回一个 HeapTuple 指针或一个 NULL 指针（不是 SQL 空值，也就是说不要把 *isNull* 设为真）。如果你不想修改正在处理的行，就要小心地根据情况返回适当的 tg_trigtuple 或 tg_newtuple。

33.4. 一个完整的示例

这里给出一个非常简单的、用 C 编写的触发器函数示例。（用过程语言编写的触发器示例可见各过程语言的文档。）

函数 `trigf` 会报告表 `ttest` 中的行数，并在命令试图向列 `x` 中插入空值时跳过实际操作。（这样，这个触发器就起到了非空约束的作用，但不会中止事务。）

首先，表定义：

```
CREATE TABLE ttest (
    x integer
);
```

下面是触发器函数的源代码：

```
#include "postgres.h"
#include "executor/spi.h"       /* this is what you need to work with
 SPI */
#include "commands/trigger.h"   /* ... and triggers */

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

extern Datum trigf(PG_FUNCTION_ARGS);

PG_FUNCTION_INFO_V1(trigf);

Datum
trigf(PG_FUNCTION_ARGS)
{
    TriggerData *trigdata = (TriggerData *) fcinfo->context;
    TupleDesc   tupdesc;
    HeapTuple   rettuple;
    char       *when;
    bool        checknull = false;
    bool        isnull;
    int         ret, i;

    /* make sure it's called as a trigger at all */
    if (!CALLED_AS_TRIGGER(fcinfo))
        elog(ERROR, "trigf: not called by trigger manager");

    /* tuple to return to executor */
    if (TRIGGER_FIRED_BY_UPDATE(trigdata->tg_event))
        rettuple = trigdata->tg_newtuple;
    else
        rettuple = trigdata->tg_trigtuple;

    /* check for null values */
    if (!TRIGGER_FIRED_BY_DELETE(trigdata->tg_event)
        && TRIGGER_FIRED_BEFORE(trigdata->tg_event))
        checknull = true;
```

```
    if (TRIGGER_FIRED_BEFORE(trigdata->tg_event))
        when = "before";
    else
        when = "after ";

    tupdesc = trigdata->tg_relation->rd_att;

    /* connect to SPI manager */
    if ((ret = SPI_connect()) < 0)
        elog(ERROR, "trigf (fired %s): SPI_connect returned %d", when,
 ret);

    /* get number of rows in table */
    ret = SPI_exec("SELECT count(*) FROM ttest", 0);

    if (ret < 0)
        elog(ERROR, "trigf (fired %s): SPI_exec returned %d", when,
 ret);

    /* count(*) returns int8, so be careful to convert */
    i = DatumGetInt64(SPI_getbinval(SPI_tuptable->vals[0],
                                    SPI_tuptable->tupdesc,
                                    1,
                                    &isnull));

    elog (INFO, "trigf (fired %s): there are %d rows in ttest", when,
 i);

    SPI_finish();

    if (checknull)
    {
        SPI_getbinval(rettuple, tupdesc, 1, &isnull);
        if (isnull)
            rettuple = NULL;
    }

    return PointerGetDatum(rettuple);
}
```

编译完源代码后，声明函数和触发器：

```
CREATE FUNCTION trigf() RETURNS trigger
    AS 'filename'
    LANGUAGE C;

CREATE TRIGGER tbefore BEFORE INSERT OR UPDATE OR DELETE ON ttest
    FOR EACH ROW EXECUTE PROCEDURE trigf();

CREATE TRIGGER tafter AFTER INSERT OR UPDATE OR DELETE ON ttest
    FOR EACH ROW EXECUTE PROCEDURE trigf();
```

现在可以测试该触发器的行为：

```
=> INSERT INTO ttest VALUES (NULL);
INFO:  trigf (fired before): there are 0 rows in ttest
INSERT 0 0

-- Insertion skipped and AFTER trigger is not fired

=> SELECT * FROM ttest;
 x
---
(0 rows)

=> INSERT INTO ttest VALUES (1);
INFO:  trigf (fired before): there are 0 rows in ttest
INFO:  trigf (fired after ): there are 1 rows in ttest
                                       ^^^^^^^^
                             remember what we said about visibility.
INSERT 167793 1
vac=> SELECT * FROM ttest;
 x
---
 1
(1 row)

=> INSERT INTO ttest SELECT x * 2 FROM ttest;
INFO:  trigf (fired before): there are 1 rows in ttest
INFO:  trigf (fired after ): there are 2 rows in ttest
                                       ^^^^^^
                             remember what we said about visibility.
INSERT 167794 1
=> SELECT * FROM ttest;
 x
---
 1
 2
(2 rows)

=> UPDATE ttest SET x = NULL WHERE x = 2;
INFO:  trigf (fired before): there are 2 rows in ttest
UPDATE 0
=> UPDATE ttest SET x = 4 WHERE x = 2;
INFO:  trigf (fired before): there are 2 rows in ttest
INFO:  trigf (fired after ): there are 2 rows in ttest
UPDATE 1
vac=> SELECT * FROM ttest;
 x
---
 1
 4
(2 rows)

=> DELETE FROM ttest;
INFO:  trigf (fired before): there are 2 rows in ttest
INFO:  trigf (fired before): there are 1 rows in ttest
INFO:  trigf (fired after ): there are 0 rows in ttest
```

```
INFO:  trigf (fired after ): there are 0 rows in ttest
                                       ^^^^^^
                             remember what we said about visibility.
DELETE 2
=> SELECT * FROM ttest;
 x
---
(0 rows)
```

更复杂的示例可见 `src/test/regress/regress.c` 和 `contrib/spi`。

第 34 章 规则系统

本章讨论PostgreSQL中的规则系统。产生式规则系统在概念上很简单，
但在实际使用时涉及许多微妙之处。

有些其他数据库系统定义了主动型数据库规则，它们通常表现为存储过程和触发器。在PostgreSQL中，这类功能同样可以用函数和触发器实现。

规则系统（更准确地说，查询重写规则系统）与存储过程和触发器完全不同。它会修改查询，
使其将规则纳入考虑，然后把修改后的查询交给查询规划器进行规划和执行。它非常强大，
可用于查询语言过程、视图和版本等许多用途。关于这一规则系统的理论基础及其能力，另见[ston90b]和[ong90]。

34.1. 查询树

要理解规则系统如何工作，必须先知道它在何时被调用，以及它的输入和输出是什么。

规则系统位于解析器和规划器之间。它接收解析器的输出，即一棵查询树，
以及用户定义的重写规则；这些规则本身也是带有一些附加信息的查询树。
它会生成零棵或多棵查询树作为结果。因此，
它的输入和输出始终都是解析器本身也可能产生的东西，
所以它看到的任何内容基本上都可以表示为一个SQL语句。

那么什么是查询树？它是SQL语句的一种内部表示，其中构成语句的各个部分被分别存储。
如果你设置了配置参数`debug_print_parse`、`debug_print_rewritten`或`debug_print_plan`，这些查询树就可以显示在服务器日志中。规则动作也以查询树的形式存储在系统目录`pg_rewrite`中。它们的格式不像日志输出，但包含的却是完全相同的信息。

阅读原始查询树需要一些经验。但由于查询树的SQL表示已足以理解规则系统，
本章不会讲解如何阅读它们。

在阅读本章中查询树的 SQL 表示时，需要能够辨认语句在查询树结构中被拆分成了哪些部分。查询树的组成部分是

命令类型

这是一个简单的值，用来说明是哪一种命令（`SELECT`、`INSERT`、`UPDATE`、`DELETE`）
产生了该查询树。

范围表

范围表是查询中使用的关系列表。在 `SELECT` 语句中，它们就是关键字 `FROM` 后面给出的关系。

每个范围表项标识一个表或视图，并说明在查询的其他部分以哪个名称来称呼它。
在查询树中，范围表项是按编号而不是按名称引用的，因此即使像 SQL 语句中那样出现重名，这里也无关紧要。这种情况可能出现在规则的范围表被合并之后。
本章中的示例不会涉及这种情形。

结果关系

这是范围表中的一个索引，用来标识查询结果应写入哪个关系。

`SELECT` 查询通常没有结果关系。`SELECT INTO` 这一特殊情况与 `CREATE TABLE` 后接 `INSERT ... SELECT` 几乎相同，这里不再单独讨论。

对于 INSERT、UPDATE 和 DELETE 命令，结果关系就是要让修改生效的表（或视图！）。

目标列表

目标列表是定义查询结果的表达式列表。对于 SELECT，这些表达式构成查询的最终输出。它们对应于关键字 SELECT 和 FROM 之间的表达式。（* 只是某个关系全部列名的缩写。解析器会把它展开为各个独立列，因此规则系统永远不会看到它。）

DELETE 命令不需要目标列表，因为它们不产生任何结果。实际上，规划器会向空目标列表加入一个特殊的 CTID 项，但这发生在规则系统之后，将在后面讨论；对规则系统而言，目标列表是空的。

对于 INSERT 命令，目标列表描述的是将要进入结果关系的新行。它由 VALUES 子句中的表达式，或 INSERT ... SELECT 里 SELECT 子句中的表达式组成。重写过程的第一步会为原始命令未赋值但带有默认值的列补上目标列表项。其余列（既没有给定值也没有默认值）将由规划器填入常量空值表达式。

对于 UPDATE 命令，目标列表描述要替换旧行的新行。在规则系统中，它只包含命令中 SET column = expression 部分的表达式。规划器会通过插入把旧行值复制到新行的表达式来处理缺失列。与 DELETE 一样，它也会加入一个特殊的 CTID 项。

目标列表中的每一项都包含一个表达式，它可以是常量值、
指向范围表中某个关系列的变量、一个参数，或者由函数调用、常量、变量、
操作符等构成的表达式树。

条件

查询的条件是一个表达式，与目标列表项中的表达式很相似。该表达式的结果值是布尔值，用来说明是否应对最终结果行执行相应操作（INSERT、UPDATE、DELETE 或 SELECT）。它对应于 SQL 语句的 WHERE 子句。

连接树

查询的连接树显示了 FROM 子句的结构。对于 SELECT ... FROM a, b, c 这样的简单查询，连接树只是 FROM 项的一个列表，因为允许按任意顺序连接它们。但当使用 JOIN 表达式，特别是外连接时，就必须按连接所显示的顺序进行连接。在这种情况下，连接树展示的是 JOIN 表达式的结构。与特定 JOIN 子句相关的限制（来自 ON 或 USING 表达式）会作为条件表达式附着在相应的连接树节点上。把顶层 WHERE 表达式也存储为附着在顶层连接树项上的一个条件，会很方便。因此，连接树实际上同时表示 SELECT 的 FROM 和 WHERE 子句。

其他

查询树的其他部分，如 ORDER BY 子句，这里不作关注。规则系统在应用规则时会替换其中某些项，但这与规则系统的基本原理关系不大。

34.2. 视图和规则系统

在 PostgreSQL 中，视图通过规则系统实现。实际上，以下命令：

```
CREATE VIEW myview AS SELECT * FROM mytab;
```

与下面这两条命令基本没有区别：

```
CREATE TABLE myview (same column list as mytab);
```

```
CREATE RULE "_RETURN" AS ON SELECT TO myview DO INSTEAD
    SELECT * FROM mytab;
```

因为这正是 `CREATE VIEW` 命令在内部所做的事情。这会带来一些副作用。其中之一是，在 PostgreSQL 系统目录中，视图的信息与表的信息完全相同。因此，对解析器而言，表和视图完全没有区别。它们是同一种东西：关系。

34.2.1. SELECT 规则如何工作

`ON SELECT`规则会在所有查询上作为最后一步应用，即使给出的命令是`INSERT`、`UPDATE`或`DELETE`也一样。它们与其他命令类型上的规则在语义上不同，因为它们是就地修改查询树，而不是创建新的查询树。因此我们先讨论`SELECT`规则。

目前，一个`ON SELECT`规则中只能有一个动作，而且它必须是一个无条件、带有`INSTEAD`的`SELECT`动作。之所以有这个限制，是为了让规则足够安全，从而能够向普通用户开放；它也把`ON SELECT`规则限制为像视图那样工作。

本章的示例是两个进行一些计算的连接视图，以及另外一些依次使用它们的视图。最初两个视图中的一个会在后面通过为`INSERT`、`UPDATE`和`DELETE`操作添加规则来定制，从而最终得到一个在行为上像真正的表、但又带有某些特殊功能的视图。作为入门示例，这并不算简单，因此会让理解变得更难一些。但与其使用许多可能令人混淆的不同示例，不如用一个示例逐步覆盖这里讨论的全部要点。

对于这个示例，我们需要一个小的 `min` 函数，它返回两个整数值中较小的一个。可以这样创建它：

```
CREATE FUNCTION min(integer, integer) RETURNS integer AS $$
    SELECT CASE WHEN $1 < $2 THEN $1 ELSE $2 END
$$ LANGUAGE SQL STRICT;
```

在前两节对规则系统的描述中，我们需要用到如下真实表：

```
CREATE TABLE shoe_data (
    shoename   text,          -- primary key
    sh_avail   integer,       -- available number of pairs
    slcolor    text,          -- preferred shoelace color
    slminlen   real,          -- minimum shoelace length
    slmaxlen   real,          -- maximum shoelace length
    slunit     text           -- length unit
);

CREATE TABLE shoelace_data (
    sl_name    text,          -- primary key
    sl_avail   integer,       -- available number of pairs
    sl_color   text,          -- shoelace color
    sl_len     real,          -- shoelace length
    sl_unit    text           -- length unit
);

CREATE TABLE unit (
    un_name    text,          -- primary key
    un_fact    real           -- factor to transform to cm
);
```

如你所见，它们表示的是鞋店数据。

这些视图的创建方式如下：

```
CREATE VIEW shoe AS
    SELECT sh.shoename,
           sh.sh_avail,
           sh.slcolor,
           sh.slminlen,
           sh.slminlen * un.un_fact AS slminlen_cm,
           sh.slmaxlen,
           sh.slmaxlen * un.un_fact AS slmaxlen_cm,
           sh.slunit
      FROM shoe_data sh, unit un
     WHERE sh.slunit = un.un_name;

CREATE VIEW shoelace AS
    SELECT s.sl_name,
           s.sl_avail,
           s.sl_color,
           s.sl_len,
           s.sl_unit,
           s.sl_len * u.un_fact AS sl_len_cm
      FROM shoelace_data s, unit u
     WHERE s.sl_unit = u.un_name;

CREATE VIEW shoe_ready AS
    SELECT rsh.shoename,
           rsh.sh_avail,
           rsl.sl_name,
           rsl.sl_avail,
           min(rsh.sh_avail, rsl.sl_avail) AS total_avail
      FROM shoe rsh, shoelace rsl
     WHERE rsl.sl_color = rsh.slcolor
       AND rsl.sl_len_cm >= rsh.slminlen_cm
       AND rsl.sl_len_cm <= rsh.slmaxlen_cm;
```

创建 shoelace 视图的 CREATE VIEW 命令（这是我们这里最简单的一个例子）会创建一个关系 shoelace，并在 pg_rewrite 中创建一项，说明只要查询的范围表中引用了关系 shoelace，就必须应用一条重写规则。该规则没有规则条件（稍后在讨论非 SELECT 规则时再谈，因为目前 SELECT 规则不能有规则条件），并且它是 INSTEAD 规则。注意，规则条件与查询条件不是一回事。这里我们的规则动作带有一个查询条件。规则动作本身是一棵查询树，它是视图创建命令中 SELECT 语句的一个副本。

> **注意**
>
> 你在pg_rewrite项中看到的、用于NEW和OLD的两个额外范围表项（在打印出的查询树中，它们因历史原因被命名为*NEW*和*OLD*），与SELECT规则无关。

现在我们填充 unit，shoe_data 和 shoelace_data，然后对视图运行一个简单查询：

```
INSERT INTO unit VALUES ('cm', 1.0);
INSERT INTO unit VALUES ('m', 100.0);
```

```
INSERT INTO unit VALUES ('inch', 2.54);

INSERT INTO shoe_data VALUES ('sh1', 2, 'black', 70.0, 90.0, 'cm');
INSERT INTO shoe_data VALUES ('sh2', 0, 'black', 30.0, 40.0, 'inch');
INSERT INTO shoe_data VALUES ('sh3', 4, 'brown', 50.0, 65.0, 'cm');
INSERT INTO shoe_data VALUES ('sh4', 3, 'brown', 40.0, 50.0, 'inch');

INSERT INTO shoelace_data VALUES ('sl1', 5, 'black', 80.0, 'cm');
INSERT INTO shoelace_data VALUES ('sl2', 6, 'black', 100.0, 'cm');
INSERT INTO shoelace_data VALUES ('sl3', 0, 'black', 35.0 , 'inch');
INSERT INTO shoelace_data VALUES ('sl4', 8, 'black', 40.0 , 'inch');
INSERT INTO shoelace_data VALUES ('sl5', 4, 'brown', 1.0 , 'm');
INSERT INTO shoelace_data VALUES ('sl6', 0, 'brown', 0.9 , 'm');
INSERT INTO shoelace_data VALUES ('sl7', 7, 'brown', 60 , 'cm');
INSERT INTO shoelace_data VALUES ('sl8', 1, 'brown', 40 , 'inch');

SELECT * FROM shoelace;

 sl_name   | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
-----------+----------+----------+--------+---------+-----------
 sl1       |        5 | black    |     80 | cm      |        80
 sl2       |        6 | black    |    100 | cm      |       100
 sl7       |        7 | brown    |     60 | cm      |        60
 sl3       |        0 | black    |     35 | inch    |      88.9
 sl4       |        8 | black    |     40 | inch    |     101.6
 sl8       |        1 | brown    |     40 | inch    |     101.6
 sl5       |        4 | brown    |      1 | m       |       100
 sl6       |        0 | brown    |    0.9 | m       |        90
(8 rows)
```

这是你能在这些视图上执行的最简单的SELECT，因此我们借此机会说明视图规则的基础。SELECT * FROM shoelace由解析器解释后，会生成如下查询树：

```
SELECT shoelace.sl_name, shoelace.sl_avail,
       shoelace.sl_color, shoelace.sl_len,
       shoelace.sl_unit, shoelace.sl_len_cm
  FROM shoelace shoelace;
```

然后它会被交给规则系统。规则系统遍历范围表，检查其中的关系是否有相应规则。在处理shoelace的范围表项时（到目前为止只有这一个），它会找到带有如下查询树的_RETURN规则：

```
SELECT s.sl_name, s.sl_avail,
       s.sl_color, s.sl_len, s.sl_unit,
       s.sl_len * u.un_fact AS sl_len_cm
  FROM shoelace *OLD*, shoelace *NEW*,
       shoelace_data s, unit u
 WHERE s.sl_unit = u.un_name;
```

为了展开视图，重写器只需创建一个子查询范围表条目，其中包含规则的动作查询树，然后用这个范围表条目替换原先引用视图的条目。
所得的重写后查询树几乎等同于输入以下语句的结果：

```
SELECT shoelace.sl_name, shoelace.sl_avail,
       shoelace.sl_color, shoelace.sl_len,
```

```
       shoelace.sl_unit, shoelace.sl_len_cm
  FROM (SELECT s.sl_name,
               s.sl_avail,
               s.sl_color,
               s.sl_len,
               s.sl_unit,
               s.sl_len * u.un_fact AS sl_len_cm
          FROM shoelace_data s, unit u
         WHERE s.sl_unit = u.un_name) shoelace;
```

不过有一点不同：子查询的范围表中有两个额外条目，shoelace *OLD* 和 shoelace *NEW*。这些条目不直接参与查询，因为子查询的连接树或目标列表并未引用它们。重写器用它们保存原先引用视图的范围表条目中的访问权限检查信息。这样，即使重写后的查询没有直接使用视图，执行器仍会检查用户是否具有访问该视图所需的权限。

这就是应用的第一条规则。规则系统接着会检查顶层查询中剩余的范围表项（本例中已经没有了），并递归检查新增子查询中的范围表项，看它们是否引用了视图。（但它不会展开*OLD*或*NEW*，否则就会出现无限递归！）在这个例子中，shoelace_data和unit都没有重写规则，因此重写到此结束，上面的结果就是交给规划器的最终结果。

现在我们想写一个查询，找出商店里目前哪些鞋子有颜色和长度都匹配的鞋带，并且完全匹配的总双数大于等于二。

```
SELECT * FROM shoe_ready WHERE total_avail >= 2;

 shoename | sh_avail | sl_name | sl_avail | total_avail
----------+----------+---------+----------+-------------
 sh1      |        2 | sl1     |        5 |           2
 sh3      |        4 | sl7     |        7 |           4
(2 rows)
```

这一次解析器的输出是如下查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM shoe_ready shoe_ready
 WHERE shoe_ready.total_avail >= 2;
```

首先应用的是用于 shoe_ready 视图的规则，它得到如下查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM (SELECT rsh.shoename,
               rsh.sh_avail,
               rsl.sl_name,
               rsl.sl_avail,
               min(rsh.sh_avail, rsl.sl_avail) AS total_avail
          FROM shoe rsh, shoelace rsl
         WHERE rsl.sl_color = rsh.slcolor
           AND rsl.sl_len_cm >= rsh.slminlen_cm
           AND rsl.sl_len_cm <= rsh.slmaxlen_cm) shoe_ready
 WHERE shoe_ready.total_avail >= 2;
```

类似地，shoe 和 shoelace 的规则会被替换进子查询的范围表中，最终得到一棵三层查询树：

```
SELECT shoe_ready.shoename, shoe_ready.sh_avail,
       shoe_ready.sl_name, shoe_ready.sl_avail,
       shoe_ready.total_avail
  FROM (SELECT rsh.shoename,
               rsh.sh_avail,
               rsl.sl_name,
               rsl.sl_avail,
               min(rsh.sh_avail, rsl.sl_avail) AS total_avail
          FROM (SELECT sh.shoename,
                       sh.sh_avail,
                       sh.slcolor,
                       sh.slminlen,
                       sh.slminlen * un.un_fact AS slminlen_cm,
                       sh.slmaxlen,
                       sh.slmaxlen * un.un_fact AS slmaxlen_cm,
                       sh.slunit
                  FROM shoe_data sh, unit un
                 WHERE sh.slunit = un.un_name) rsh,
               (SELECT s.sl_name,
                       s.sl_avail,
                       s.sl_color,
                       s.sl_len,
                       s.sl_unit,
                       s.sl_len * u.un_fact AS sl_len_cm
                  FROM shoelace_data s, unit u
                 WHERE s.sl_unit = u.un_name) rsl
         WHERE rsl.sl_color = rsh.slcolor
           AND rsl.sl_len_cm >= rsh.slminlen_cm
           AND rsl.sl_len_cm <= rsh.slmaxlen_cm) shoe_ready
 WHERE shoe_ready.total_avail > 2;
```

事实证明，规划器会把这棵树折叠成两层的查询树：最底层的 SELECT 命令会被“上拉”到中间的 SELECT 中，因为不需要单独处理它们。但中间的 SELECT 会保持与顶层分离，因为它包含聚合函数。如果把那些也上拉，就会改变顶层 SELECT 的行为，而这不是我们想要的。不过，折叠查询树属于一种优化，重写系统本身无须关心。

> **注意**
>
> 目前规则系统对视图规则没有任何递归停止机制（其他类型的规则才有）。这并不会造成太大问题，因为要把查询推入无限循环（使服务器进程不断膨胀，直到达到内存上限）的唯一途径，是先创建一些表，然后手工用 CREATE RULE 设置视图规则，让其中一个视图选择自另一个视图，而后者又选择自前者。如果使用 CREATE VIEW，这种情况永远不会发生，因为在执行第一个 CREATE VIEW 时，第二个关系尚不存在，因此第一个视图无法从第二个视图中选择数据。

34.2.2. 非 SELECT 语句中的视图规则

上面关于视图规则的描述没有涉及查询树的两个细节：命令类型和结果关系。事实上，视图规则并不需要这些信息。

SELECT的查询树与其他任何命令的查询树之间只有少数差别。显然，它们的命令类型不同；对于SELECT之外的命令，结果关系会指向结果应写入的那个范围表项。除此之外，
其余部分完全相同。因此，假设有两个表t1和t2，都具有列a和b，那么下面两条语句的查询树：

```
SELECT t2.b FROM t1, t2 WHERE t1.a = t2.a;

UPDATE t1 SET b = t2.b WHERE t1.a = t2.a;
```

几乎是一样的。特别是：

- 范围表中包含表t1和t2的项。
- 目标列表中都包含一个变量，该变量指向表t2的范围表项中的列b。
- 条件表达式会比较两个范围表项中的列a是否相等。
- 连接树都表示t1与t2之间的一次简单连接。

结果是，这两个查询树都会产生相似的执行计划：它们都是对这两个表的连接。对于 UPDATE，规划器会把 t1 中缺失的列补入目标列表，最终查询树会变成：

```
UPDATE t1 SET a = t1.a, b = t2.b WHERE t1.a = t2.a;
```

因而，执行器在这个连接上运行时会产生与下面语句完全相同的结果集：

```
SELECT t1.a, t2.b FROM t1, t2 WHERE t1.a = t2.a;
```

但在 UPDATE 中有个小问题：执行器并不关心它所做的连接的结果将被用于什么。它只是生成一个行结果集。一个是 SELECT 命令，另一个是 UPDATE 命令，这一差别是在执行器的调用者中处理的。调用者仍然知道（通过查看查询树）这是一个 UPDATE，也知道这个结果应写入表 t1。但问题在于：其中哪一行应当被新行替换？

为了解决这个问题，UPDATE（以及 DELETE）语句的目标列表中会额外加入一项：当前元组 ID（CTID）。这是一个系统列，包含该行所在的文件块号以及在块中的位置。已知表之后，
就可以利用 CTID 取回要更新的 t1 原始行。把 CTID 加入目标列表后，查询实际上会变成：

```
SELECT t1.a, t2.b, t1.ctid FROM t1, t2 WHERE t1.a = t2.a;
```

现在还要引入 PostgreSQL 的另一个细节。旧表行不会被覆盖，这也是 ROLLBACK 之所以很快的原因。在 UPDATE 中，新结果行会被插入表中（去掉 CTID 之后），而 CTID 所指向的旧行，其行头中的 cmax 和 xmax 项会被设置为当前命令计数器和当前事务 ID。于是旧行被隐藏起来，事务提交后，清理器（vacuum）最终就可以真正移除它。

知道了这些以后，我们就可以用完全相同的方式把视图规则应用到任何命令上。没有区别。

34.2.3. PostgreSQL 中视图的能力

上文演示了规则系统如何把视图定义整合进原始查询树。在第二个示例中，
从一个视图发出的简单SELECT最终生成了一棵四表连接的查询树（unit
以不同名称使用了两次）。

用规则系统实现视图的好处在于：规划器能够在单棵查询树中同时看到哪些表必须被扫描、
这些表之间的关系、来自视图的限制条件，以及原始查询自身的条件。
即使原始查询本身已经是对若干视图的连接，情况也仍然如此。
规划器必须决定执行查询的最佳路径，而它掌握的信息越多，这个决定通常就越好。
PostgreSQL实现的规则系统能够保证，到那个阶段为止，
关于该查询的全部可用信息都已经集中在这里。

34.2.4. 更新视图

如果把一个视图指定为 INSERT、UPDATE 或 DELETE 的目标关系，会发生什么？在完成上述替换之后，我们得到的查询树中，结果关系会指向一个子查询范围表项。这是行不通的，因此重写器一旦发现自己产生了这样的结果，就会抛出一个错误。

要改变这种情况，我们可以定义规则来修改这类命令的行为。这是下一节的主题。

34.3. INSERT、UPDATE 和 DELETE 上的规则

定义在INSERT、UPDATE和DELETE上的规则与前几节描述的视图规则有明显的不同。首先，它们的CREATE RULE命令允许更多：

- 它们可以没有动作。
- 它们可以有多个动作。
- 它们可以是INSTEAD或ALSO（缺省）。
- 伪关系NEW和OLD可以派上用场。
- 它们可以有规则条件。

第二，它们不是就地修改查询树，而是创建零棵或多棵新查询树，并且可能丢弃原始查询树。

34.3.1. 更新规则如何工作

记住以下语法：

```
CREATE [ OR REPLACE ] RULE name AS ON event
    TO table [ WHERE condition ]
    DO [ ALSO | INSTEAD ] { NOTHING | command | ( command ; command .
.. ) }
```

在后文中，更新规则是指定义在INSERT、UPDATE或DELETE上的规则。

当查询树的结果关系和命令类型分别等于CREATE RULE命令中给出的对象和事件时，规则系统就会应用更新规则。对于更新规则，规则系统会创建一个查询树列表，初始时该列表为空。动作可以有零个（关键字NOTHING）、一个或多个。为简化说明，我们先看只有一个动作的规则。这个规则可以有条件，也可以没有条件；它还可以是INSTEAD或ALSO（默认值）。

什么是规则条件？它是一种限制，用来说明何时执行规则动作、何时不执行。这个条件只能引用伪关系NEW和/或OLD，它们基本上表示作为对象给出的那个关系，只是带有特殊含义。

因此，对于只有一个动作的规则，有以下四种情况会产生相应的查询树。

没有条件，有ALSO

规则动作的查询树，再附加原始查询树的条件

没有条件，但有INSTEAD

规则动作的查询树，再附加原始查询树的条件

给出了条件，有ALSO

规则动作的查询树，再附加规则条件和原始查询树的条件

给出了条件，有INSTEAD

规则动作的查询树，再附加规则条件和原始查询树的条件；
以及附加了规则条件取反后的原始查询树

最后，如果规则是 ALSO，就把未经更改的原始查询树加入列表。由于只有带条件的 INSTEAD 规则已经加入了原始查询树，因此，对于只有一个动作的规则，最终会得到一棵或两棵输出查询树。

对于ON INSERT规则，原始查询（如果未被INSTEAD抑制）会先于规则添加的任何动作执行。这样一来，这些动作就能看到被插入的行。但对ON UPDATE和ON DELETE规则，原始查询会在规则添加的动作之后执行。这就保证了这些动作能够看到将被更新或删除的行；否则，
动作可能什么也做不了，因为它们找不到符合条件的行。

由规则动作生成的查询树会再次送入重写系统，随后可能还会有更多规则被应用，
从而产生更多或更少的查询树。因此，一个规则的动作必须具有不同的命令类型，
或者具有与该规则自身不同的结果关系。否则，这种递归过程就会陷入无限循环。
（规则的递归展开会被检测出来，并作为错误报告。）

pg_rewrite系统目录中动作里的查询树只是模板。由于它们可以引用NEW和OLD的范围表项，
因此在使用前必须进行一些替换。对任何NEW的引用，
都会先在原始查询的目标列表中查找相应项。如果找到了，就用该项的表达式替换该引用。
否则，NEW就与OLD含义相同（对于UPDATE），或者被替换为一个空值（对于INSERT）。
任何对OLD的引用，都会被替换为对结果关系对应范围表项的引用。

在系统完成应用更新规则后，它再应用视图规则到生成的查询树上。
视图无法插入新的更新动作，所以没有必要向视图重写的输出应用更新规则。

34.3.1.1. 第一个规则：逐步分析

假设我们想跟踪shoelace_data关系中sl_avail列的变化。因此，
我们建立一个日志表和一条规则，使其在shoelace_data上执行UPDATE时，
有条件地写入一条日志记录。

```
CREATE TABLE shoelace_log (
    sl_name    text,          -- shoelace changed
    sl_avail   integer,       -- new available value
    log_who    text,          -- who did it
    log_when   timestamp      -- when
);

CREATE RULE log_shoelace AS ON UPDATE TO shoelace_data
    WHERE NEW.sl_avail <> OLD.sl_avail
    DO INSERT INTO shoelace_log VALUES (
                                    NEW.sl_name,
                                    NEW.sl_avail,
                                    current_user,
                                    current_timestamp
                                );
```

现在有人执行了：

```
UPDATE shoelace_data SET sl_avail = 6 WHERE sl_name = 'sl7';
```

然后我们查看日志表：

```
SELECT * FROM shoelace_log;

 sl_name | sl_avail | log_who | log_when
---------+----------+---------+----------------------------------
 sl7     |        6 | Al      | Tue Oct 20 16:14:45 1998 MET DST
(1 row)
```

这正是我们预期的结果。后台发生的事情如下。解析器创建了查询树：

```
UPDATE shoelace_data SET sl_avail = 6
  FROM shoelace_data shoelace_data
 WHERE shoelace_data.sl_name = 'sl7';
```

此时存在一条ON UPDATE规则log_shoelace，它带有如下规则条件表达式：

```
NEW.sl_avail <> OLD.sl_avail
```

其动作是：

```
INSERT INTO shoelace_log VALUES (
       *NEW*.sl_name, *NEW*.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*;
```

（这看起来有点奇怪，因为通常你不能写INSERT ... VALUES ... FROM。这里的FROM子句只是为了表明查询树中存在用于*NEW*和*OLD*的范围表项。之所以需要这些项，是为了让INSERT命令查询树中的变量能够引用它们。）

该规则是一条带条件的ALSO规则，因此规则系统必须返回两棵查询树：修改后的规则动作，以及原始查询树。第 1 步中，原始查询的范围表会并入规则动作的查询树，得到：

```
INSERT INTO shoelace_log VALUES (
       *NEW*.sl_name, *NEW*.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*,
       shoelace_data shoelace_data;
```

第 2 步将规则条件加进去，因此结果集被限制为sl_avail发生变化的行：

```
INSERT INTO shoelace_log VALUES (
       *NEW*.sl_name, *NEW*.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*,
       shoelace_data shoelace_data
 WHERE *NEW*.sl_avail <> *OLD*.sl_avail;
```

（这看起来更奇怪，因为INSERT ... VALUES同样没有WHERE子句，但规划器和执行器处理它并无困难。反正它们本来也需要为INSERT ... SELECT支持相同功能。）

第 3 步把原始查询树的条件加进去，把结果集进一步限制为只有那些原始查询会触及的行：

```
INSERT INTO shoelace_log VALUES (
```

```
       *NEW*.sl_name, *NEW*.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*,
       shoelace_data shoelace_data
 WHERE *NEW*.sl_avail <> *OLD*.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

第 4 步把NEW引用替换为原始查询树中的目标列表项，或者替换为结果关系中相应的变量引用：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*,
       shoelace_data shoelace_data
 WHERE 6 <> *OLD*.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

第 5 步把OLD引用替换为结果关系引用：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data *NEW*, shoelace_data *OLD*,
       shoelace_data shoelace_data
 WHERE 6 <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

至此就完成了。由于规则是ALSO，我们还要输出原始查询树。简而言之，
规则系统输出的是一个包含两棵查询树的列表，它们对应于以下语句：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, 6,
       current_user, current_timestamp )
  FROM shoelace_data
 WHERE 6 <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';

UPDATE shoelace_data SET sl_avail = 6
 WHERE sl_name = 'sl7';
```

它们会按这个顺序执行，而这正是该规则想要达到的效果。

上述替换以及附加的条件能够保证：如果原始查询是下面这样，就不会写入任何日志记录：

```
UPDATE shoelace_data SET sl_color = 'green'
 WHERE sl_name = 'sl7';
```

在这种情况下，原始查询树不包含sl_avail的目标列表项，因此NEW.sl_avail
会被shoelace_data.sl_avail替换。于是，规则生成的额外命令是：

```
INSERT INTO shoelace_log VALUES (
       shoelace_data.sl_name, shoelace_data.sl_avail,
       current_user, current_timestamp )
  FROM shoelace_data
 WHERE shoelace_data.sl_avail <> shoelace_data.sl_avail
   AND shoelace_data.sl_name = 'sl7';
```

而该条件永远不可能为真。

如果原始查询修改多行，这种机制同样能够正常工作。例如，假设有人发出如下命令：

```
UPDATE shoelace_data SET sl_avail = 0
 WHERE sl_color = 'black';
```

实际上会更新四行（sl1、sl2、sl3和sl4）。但sl3本来就已经是sl_avail = 0。在这种情况下，原始查询树的条件不同，因此规则会产生额外的查询树：

```
INSERT INTO shoelace_log
SELECT shoelace_data.sl_name, 0,
       current_user, current_timestamp
  FROM shoelace_data
 WHERE 0 <> shoelace_data.sl_avail
   AND shoelace_data.sl_color = 'black';
```

这棵查询树必然会插入三条新的日志记录。这完全正确。

到这里就能看出，为什么原始查询树最后执行至关重要。如果先执行UPDATE，那么所有行都已经被设为零，记日志的INSERT就找不到任何满足0 <> shoelace_data.sl_avail的行了。

34.3.2. 与视图的协作

为了防止有人像前面提到的那样对视图关系执行INSERT、UPDATE或DELETE，一种简单的办法是让那些查询树直接被丢弃。因此我们可以创建如下规则：

```
CREATE RULE shoe_ins_protect AS ON INSERT TO shoe
    DO INSTEAD NOTHING;
CREATE RULE shoe_upd_protect AS ON UPDATE TO shoe
    DO INSTEAD NOTHING;
CREATE RULE shoe_del_protect AS ON DELETE TO shoe
    DO INSTEAD NOTHING;
```

如果现在某人尝试对视图关系shoe执行这些操作，规则系统就会应用这些规则。由于这些规则没有动作而且是INSTEAD，生成的查询树列表将为空；整个查询也就什么都不会做，因为规则系统处理完后已经没有任何内容可供优化或执行。

另一种更完善的做法，是创建一些规则，把查询树重写成在真实表上执行正确操作的查询树。要在视图shoelace上做到这一点，我们创建下列规则：

```
CREATE RULE shoelace_ins AS ON INSERT TO shoelace
    DO INSTEAD
    INSERT INTO shoelace_data VALUES (
           NEW.sl_name,
           NEW.sl_avail,
           NEW.sl_color,
           NEW.sl_len,
           NEW.sl_unit
    );

CREATE RULE shoelace_upd AS ON UPDATE TO shoelace
    DO INSTEAD
    UPDATE shoelace_data
```

```
       SET sl_name = NEW.sl_name,
           sl_avail = NEW.sl_avail,
           sl_color = NEW.sl_color,
           sl_len = NEW.sl_len,
           sl_unit = NEW.sl_unit
     WHERE sl_name = OLD.sl_name;

CREATE RULE shoelace_del AS ON DELETE TO shoelace
    DO INSTEAD
    DELETE FROM shoelace_data
     WHERE sl_name = OLD.sl_name;
```

现在假定商店每隔一段时间就会收到一包鞋带，随附一份很长的部件清单。
但你不想每次都手工更新 shoelace 视图。因此，我们建立两个小表：一个用于插入部件清单中的项目，另一个则采用一个特殊技巧。创建命令如下：

```
CREATE TABLE shoelace_arrive (
    arr_name    text,
    arr_quant   integer
);

CREATE TABLE shoelace_ok (
    ok_name     text,
    ok_quant    integer
);

CREATE RULE shoelace_ok_ins AS ON INSERT TO shoelace_ok
    DO INSTEAD
    UPDATE shoelace
       SET sl_avail = sl_avail + NEW.ok_quant
     WHERE sl_name = NEW.ok_name;
```

现在可以将部件清单中的数据填入 shoelace_arrive 表：

```
SELECT * FROM shoelace_arrive;

 arr_name | arr_quant
----------+-----------
 sl3      |        10
 sl6      |        20
 sl8      |        20
(3 rows)
```

快速查看一下当前数据：

```
SELECT * FROM shoelace;

 sl_name  | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
----------+----------+----------+--------+---------+-----------
 sl1      |        5 | black    |     80 | cm      |        80
 sl2      |        6 | black    |    100 | cm      |       100
 sl7      |        6 | brown    |     60 | cm      |        60
 sl3      |        0 | black    |     35 | inch    |      88.9
 sl4      |        8 | black    |     40 | inch    |     101.6
 sl8      |        1 | brown    |     40 | inch    |     101.6
```

```
 sl5      |        4 | brown    |      1 | m       |       100
 sl6      |        0 | brown    |    0.9 | m       |        90
(8 rows)
```

现在把到货的鞋带入库：

```
INSERT INTO shoelace_ok SELECT * FROM shoelace_arrive;
```

然后检查结果：

```
SELECT * FROM shoelace ORDER BY sl_name;

 sl_name  | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
----------+----------+----------+--------+---------+-----------
 sl1      |        5 | black    |     80 | cm      |        80
 sl2      |        6 | black    |    100 | cm      |       100
 sl7      |        6 | brown    |     60 | cm      |        60
 sl4      |        8 | black    |     40 | inch    |     101.6
 sl3      |       10 | black    |     35 | inch    |      88.9
 sl8      |       21 | brown    |     40 | inch    |     101.6
 sl5      |        4 | brown    |      1 | m       |       100
 sl6      |       20 | brown    |    0.9 | m       |        90
(8 rows)

SELECT * FROM shoelace_log;

 sl_name | sl_avail | log_who| log_when
---------+----------+--------+----------------------------------
 sl7     |        6 | Al     | Tue Oct 20 19:14:45 1998 MET DST
 sl3     |       10 | Al     | Tue Oct 20 19:25:16 1998 MET DST
 sl6     |       20 | Al     | Tue Oct 20 19:25:16 1998 MET DST
 sl8     |       21 | Al     | Tue Oct 20 19:25:16 1998 MET DST
(4 rows)
```

从一个`INSERT ... SELECT`到这些结果，中间要经历相当长的一段过程。本章最后再来描述这一查询树转换的细节。首先，看一下解析器的输出：

```
INSERT INTO shoelace_ok
SELECT shoelace_arrive.arr_name, shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok;
```

现在应用第一条规则`shoelace_ok_ins`，它会把这一输出转换成：

```
UPDATE shoelace
   SET sl_avail = shoelace.sl_avail + shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace
 WHERE shoelace.sl_name = shoelace_arrive.arr_name;
```

同时会丢弃针对`shoelace_ok`的原始`INSERT`。这个重写后的查询会再次交给规则系统，而接下来应用的规则`shoelace_upd`会生成：

```
UPDATE shoelace_data
   SET sl_name = shoelace.sl_name,
       sl_avail = shoelace.sl_avail + shoelace_arrive.arr_quant,
```

```
       sl_color = shoelace.sl_color,
       sl_len = shoelace.sl_len,
       sl_unit = shoelace.sl_unit
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace *OLD*,
       shoelace *NEW*, shoelace_data shoelace_data
 WHERE shoelace.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = shoelace.sl_name;
```

这同样是一条INSTEAD规则，因此前一个查询树会被丢弃。注意，这个查询仍然使用视图shoelace。但规则系统尚未完成，因此会继续在其上应用_RETURN规则，于是得到：

```
UPDATE shoelace_data
   SET sl_name = s.sl_name,
       sl_avail = s.sl_avail + shoelace_arrive.arr_quant,
       sl_color = s.sl_color,
       sl_len = s.sl_len,
       sl_unit = s.sl_unit
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace *OLD*,
       shoelace *NEW*, shoelace_data shoelace_data,
       shoelace *OLD*, shoelace *NEW*,
       shoelace_data s, unit u
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name;
```

最后，规则log_shoelace被应用，生成额外的查询树：

```
INSERT INTO shoelace_log
SELECT s.sl_name,
       s.sl_avail + shoelace_arrive.arr_quant,
       current_user,
       current_timestamp
  FROM shoelace_arrive shoelace_arrive, shoelace_ok shoelace_ok,
       shoelace_ok old, shoelace_ok new,
       shoelace shoelace, shoelace *OLD*,
       shoelace *NEW*, shoelace_data shoelace_data,
       shoelace *OLD*, shoelace *NEW*,
       shoelace_data s, unit u,
       shoelace_data *OLD*, shoelace_data *NEW*
       shoelace_log shoelace_log
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name
   AND (s.sl_avail + shoelace_arrive.arr_quant) <> s.sl_avail;
```

到这里，规则系统已经没有更多规则可用，并返回生成的查询树。

因此，我们最终得到两棵查询树，它们等效于以下SQL语句：

```
INSERT INTO shoelace_log
SELECT s.sl_name,
       s.sl_avail + shoelace_arrive.arr_quant,
       current_user,
```

```
       current_timestamp
  FROM shoelace_arrive shoelace_arrive, shoelace_data shoelace_data,
       shoelace_data s
 WHERE s.sl_name = shoelace_arrive.arr_name
   AND shoelace_data.sl_name = s.sl_name
   AND s.sl_avail + shoelace_arrive.arr_quant <> s.sl_avail;

UPDATE shoelace_data
   SET sl_avail = shoelace_data.sl_avail + shoelace_arrive.arr_quant
  FROM shoelace_arrive shoelace_arrive,
       shoelace_data shoelace_data,
       shoelace_data s
 WHERE s.sl_name = shoelace_arrive.sl_name
   AND shoelace_data.sl_name = s.sl_name;
```

结果是：来自一个关系的数据被插入到另一个关系中，这个插入被改写为对第三个关系的更新，再被改写为对第四个关系的更新外加在第五个关系中记录该更新，
最后整个过程被化简为两个查询。

这里有个稍显难看的小细节。观察这两个查询会发现，shoelace_data
关系在范围表中出现了两次，而实际上完全可以缩减成一次。规划器不会处理这一点，
因此规则系统为INSERT输出的执行计划会是

```
Nested Loop
  ->  Merge Join
        ->  Seq Scan
              ->  Sort
                    ->  Seq Scan on s
        ->  Seq Scan
              ->  Sort
                    ->  Seq Scan on shoelace_arrive
  ->  Seq Scan on shoelace_data
```

而省略那个额外的范围表项则会得到

```
Merge Join
  ->  Seq Scan
        ->  Sort
              ->  Seq Scan on s
  ->  Seq Scan
        ->  Sort
              ->  Seq Scan on shoelace_arrive
```

这会在日志表中生成完全相同的项。因此，规则系统导致了对shoelace_data
表的一次完全不必要的额外扫描。而在UPDATE中，同样的冗余扫描还会再发生一次。不过，
能让这一切总体上工作起来，已经是一项相当艰巨的工作。

现在我们来做一个关于PostgreSQL规则系统及其威力的最后演示。
假设你向数据库中加入了一些颜色特别的鞋带：

```
INSERT INTO shoelace VALUES ('sl9', 0, 'pink', 35.0, 'inch', 0.0);
INSERT INTO shoelace VALUES ('sl10', 1000, 'magenta', 40.0, 'inch', 0.
0);
```

我们想建立一个视图，用来检查哪些shoelace项在颜色上与任何鞋子都不匹配。这个视图是：

```
CREATE VIEW shoelace_mismatch AS
    SELECT * FROM shoelace WHERE NOT EXISTS
        (SELECT shoename FROM shoe WHERE slcolor = sl_color);
```

它的输出是：

```
SELECT * FROM shoelace_mismatch;

 sl_name | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
---------+----------+----------+--------+---------+-----------
 sl9     |        0 | pink     |     35 | inch    |      88.9
 sl10    |     1000 | magenta  |     40 | inch    |     101.6
```

现在希望将没有库存的不匹配鞋带从数据库中删除。为了给 PostgreSQL 增加一点难度，我们不直接删除，而是再创建一个视图：

```
CREATE VIEW shoelace_can_delete AS
    SELECT * FROM shoelace_mismatch WHERE sl_avail = 0;
```

然后这样执行：

```
DELETE FROM shoelace WHERE EXISTS
    (SELECT * FROM shoelace_can_delete
             WHERE sl_name = shoelace.sl_name);
```

Voilà：

```
SELECT * FROM shoelace;

 sl_name | sl_avail | sl_color | sl_len | sl_unit | sl_len_cm
---------+----------+----------+--------+---------+-----------
 sl1     |        5 | black    |     80 | cm      |        80
 sl2     |        6 | black    |    100 | cm      |       100
 sl7     |        6 | brown    |     60 | cm      |        60
 sl4     |        8 | black    |     40 | inch    |     101.6
 sl3     |       10 | black    |     35 | inch    |      88.9
 sl8     |       21 | brown    |     40 | inch    |     101.6
 sl10    |     1000 | magenta  |     40 | inch    |     101.6
 sl5     |        4 | brown    |      1 | m       |       100
 sl6     |       20 | brown    |    0.9 | m       |        90
(9 rows)
```

作用于一个视图上的`DELETE`，其子查询条件总共使用了四个嵌套/连接的视图，
其中一个视图自身又带有一个包含视图的子查询条件，并且还用到了计算得到的视图列，
最终仍会被重写成单棵查询树，从真正的表中删除所请求的数据。

在现实世界中，大概只有很少的场景会需要这样的构造。但知道它确实能工作，总归让人安心。

34.4. 规则和权限

由于PostgreSQL规则系统会重写查询，因此可能访问原始查询中并未使用的其他表或视图。
使用更新规则时，这甚至可能包括对表的写访问。

重写规则没有单独的所有者。关系（表或视图）
的所有者会自动成为为其定义的重写规则的所有者。

PostgreSQL规则系统改变了默认访问控制系统的行为。
由于规则而使用的关系都会根据规则所有者的权限进行检查，而不是调用规则的用户。
这意味着用户只需要对其查询中明确命名的表/视图具有所需的权限。

例如，某个用户有一份电话号码列表，其中一部分是私人的，另一部分对办公室秘书有用。
该用户可以这样构造：

```
CREATE TABLE phone_data (person text, phone text, private boolean);
CREATE VIEW phone_number AS
    SELECT person, CASE WHEN NOT private THEN phone END AS phone
    FROM phone_data;
GRANT SELECT ON phone_number TO assistant;
```

除了该用户本人（以及数据库超级用户）之外，没有人可以访问phone_data表。但由于GRANT的存在，秘书可以对phone_number视图执行SELECT。规则系统会把对phone_number的SELECT重写成对phone_data的SELECT。由于该用户是phone_number的所有者，因此也是规则的所有者，对phone_data的读访问会按照该用户的权限进行检查，于是查询被允许。同时，对phone_number本身的访问检查仍会执行，但这是针对调用用户进行的，因此除了用户本人和秘书之外，没有其他人能使用它。

权限是逐条规则检查的。因此，目前秘书是唯一能看到公开电话号码的人。
但秘书还可以再建立一个视图，并把该视图授予公众访问。这样，
任何人都可以通过秘书的视图看到phone_number中的数据。
秘书做不到的是创建一个直接访问phone_data的视图。（其实秘书可以创建，
但它不会起作用，因为每次访问都会在权限检查时被拒绝。）而且，一旦用户发现秘书开放了其phone_number视图，用户就可以撤销秘书的访问权限。那样一来，
对秘书视图的任何访问都会立即失败。

这种逐条规则的检查看起来似乎像是一个安全漏洞，但实际上并不是。因为即使不这样工作，
秘书也可以建立一个与phone_number具有相同列的表，每天把数据复制进去。那样一来，
这些就是秘书自己的数据，秘书同样可以把访问权限授予任何人。GRANT
的含义本来就是“我信任你”。如果某个你信任的人做了上面的事，那么该重新考虑这份信任，
并使用REVOKE了。

需要注意的是，虽然视图可以用前文展示的技术来隐藏某些列的内容，
但它们不能被用来可靠地隐藏不可见行中的数据。例如，下面这个视图就是不安全的：

```
CREATE VIEW phone_number AS
    SELECT person, phone FROM phone_data WHERE phone NOT LIKE '412%';
```

这个视图看起来似乎很安全，因为规则系统会把对 phone_number 的任何 SELECT 重写为对 phone_data 的 SELECT，并附加条件，要求只处理 phone 不以 412 开头的项。但是，如果用户可以创建自己的函数，就不难让规划器在 NOT LIKE 表达式之前先执行用户定义函数。

```
CREATE FUNCTION tricky(text, text) RETURNS bool AS $$
BEGIN
    RAISE NOTICE '% => %', $1, $2;
    RETURN true;
END
$$ LANGUAGE plpgsql COST 0.0000000000000000000001;
SELECT * FROM phone_number WHERE tricky(person, phone);
```

phone_data 表中的每个人和电话号码都会被打印为一条 NOTICE，因为规划器会选择先执行廉价的 tricky 函数，再执行更昂贵的 NOT LIKE。即使禁止用户定义新函数，内置函数也可以用于类似攻击。（例如，类型转换函数会将其输入包含在它们产生的错误消息中。）

类似的考虑也适用于更新规则。在上一节的示例中，示例数据库中那些表的所有者可以把shoelace视图上的SELECT、INSERT、UPDATE和DELETE权限授予其他用户，但对shoelace_log只授予SELECT权限。写日志记录的规则动作仍会成功执行，因此其他用户可以看到日志记录。但他们不能伪造记录，也不能操纵或删除已有记录。在这个例子里，不存在通过说服规划器改变操作顺序来破坏规则的可能性，因为唯一引用shoelace_log的规则是一条无条件的INSERT。在更复杂的场景中，情况未必如此。

34.5. 规则和命令状态

PostgreSQL服务器会为收到的每条命令返回一个命令状态字符串，例如INSERT 149592 1。在没有规则参与时，这很简单；但当查询被规则重写时会发生什么呢？

规则对命令状态的影响如下：

- 如果该查询没有无条件的 INSTEAD 规则，那么原始给出的查询就会照常执行，并返回它自身的命令状态。（但请注意，如果存在任何带条件的 INSTEAD 规则，那么这些规则条件的取反会被加到原始查询上。这可能会减少它处理的行数，从而影响报告的状态。）

- 如果该查询存在任意无条件的 INSTEAD 规则，那么原始查询将完全不会执行。在这种情况下，服务器会返回由 INSTEAD 规则（带条件或不带条件）插入的、最后一条且与原始查询具有相同命令类型（INSERT、UPDATE 或 DELETE）的查询的命令状态。如果没有任何规则添加出满足这些条件的查询，那么返回的命令状态会显示原始查询类型，并且把行计数和 OID 字段都置零。

（这一系统是在 PostgreSQL 7.3 中确立的。在更早的版本中，存在规则时命令状态可能会显示不同的结果。）

在第二种情况下，程序员可以通过为所需的INSTEAD规则指定一个在活动规则中按字母顺序排在最后的规则名，来确保由它设置命令状态，因为它会最后被应用。

34.6. 规则与触发器

许多能够用触发器完成的事情，同样也可以用PostgreSQL规则系统实现。不能用规则实现的内容之一是某些约束，尤其是外键。你可以放置一条带条件的规则：当某列的值没有出现在另一张表中时，把命令重写成NOTHING。但那样做会悄无声息地丢弃数据，这并不是好主意。如果需要检查值的有效性，并在值无效时生成错误消息，就必须使用触发器。

另一方面，在视图上由INSERT触发的触发器可以做到与规则相同的事情：把数据放到别处，并抑制对视图的插入。但在 UPDATE或DELETE上它做不到这一点，因为视图关系中没有可以被扫描的真实数据，触发器因而永远不会被调用。只有规则才能解决问题。

对于两者都能实现的事情，哪一种更好取决于数据库的使用方式。触发器会对每个受影响的行触发一次，而规则会修改查询，或生成额外的查询。因此，如果一条语句会影响很多行，那么发出一条额外命令的规则往往会比触发器更快，因为触发器必须为每一行调用一次，并反复执行其操作。不过，触发器方法在概念上要比规则方法简单得多，也更容易让新手写对。

下面通过一个例子说明，在一种具体情形中选择规则或触发器会有什么不同。这里有两个表：

```
CREATE TABLE computer (
    hostname        text,    -- indexed
```

```
    manufacturer    text     -- indexed
);

CREATE TABLE software (
    software        text,    -- indexed
    hostname        text     -- indexed
);
```

两个表都有数千行，并且 `hostname` 上的索引都是唯一索引。规则或触发器应实现这样一个约束：从 `software` 中删除引用了已删除计算机的行。触发器会使用以下命令：

```
DELETE FROM software WHERE hostname = $1;
```

由于从 `computer` 中删除的每一行都会调用触发器，因此触发器可以为这条命令准备并保存执行计划，并通过参数传入 `hostname` 值。规则则会写成：

```
CREATE RULE computer_del AS ON DELETE TO computer
    DO DELETE FROM software WHERE hostname = OLD.hostname;
```

现在我们来看不同类型的删除。对于下面这种情况：

```
DELETE FROM computer WHERE hostname = 'mypc.local.net';
```

`computer`表会通过索引扫描（很快），由触发器发出的命令也会使用索引扫描（同样很快）。规则产生的额外查询是：

```
DELETE FROM software WHERE computer.hostname = 'mypc.local.net'
                       AND software.hostname = computer.hostname;
```

由于已经建立了合适的索引，规划器将创建一个规划

```
Nestloop
  ->  Index Scan using comp_hostidx on computer
  ->  Index Scan using soft_hostidx on software
```

因此，触发器实现和规则实现之间的速度差异不会太大。

在下一个删除场景中，我们想去掉全部 2000 台`hostname`以`*OLD*`开头的计算机。有两种命令可以做到这件事。其中一种是：

```
DELETE FROM computer WHERE hostname >= 'old'
                       AND hostname <  'ole'
```

规则添加出来的命令将是：

```
DELETE FROM software WHERE computer.hostname >= 'old' AND computer.
hostname < 'ole'
                       AND software.hostname = computer.hostname;
```

其计划为：

```
Hash Join
  ->  Seq Scan on software
  ->  Hash
    ->  Index Scan using comp_hostidx on computer
```

另一个可能的命令是：

```
DELETE FROM computer WHERE hostname ~ '^old';
```

它会让规则所添加的命令得到下面这个执行计划：

```
Nestloop
  ->  Index Scan using comp_hostidx on computer
  ->  Index Scan using soft_hostidx on software
```

这说明，当多个条件表达式通过AND组合在一起时，规划器并没有意识到，对computer中hostname的限制同样也可以用于software上的索引扫描，而在该命令的正则表达式版本里它却能做到。触发器会对需要删除的 2000 台旧计算机中的每一台各调用一次，这意味着对computer做一次索引扫描，并对software做 2000 次索引扫描。规则实现则用两条使用索引的命令完成这件事。不过，在顺序扫描这种情况下，规则是否仍然更快，还取决于software表的总体大小。即使所有索引块很快都会进入缓存，通过 SPI 管理器执行来自触发器的 2000 条命令仍然要耗费一些时间。

我们要看的最后一个命令是：

```
DELETE FROM computer WHERE manufacturer = 'bim';
```

同样，这也可能导致从computer中删除很多行。因此，触发器同样会通过执行器运行很多条命令。规则生成的命令则是：

```
DELETE FROM software WHERE computer.manufacturer = 'bim'
                       AND software.hostname = computer.hostname;
```

这个命令的计划又将是在两个索引扫描上的嵌套循环，只不过使用了computer上的另一个索引：

```
Nestloop
  ->  Index Scan using comp_manufidx on computer
  ->  Index Scan using soft_hostidx on software
```

在上述任何一种情况下，规则系统产生的额外命令都或多或少与该命令影响的行数无关。

总结来说，只有当规则动作导致了规模很大且条件很差的连接，而规划器又无能为力时，规则才会明显慢于触发器。

第 35 章 Procedural Languages

PostgreSQL允许使用 SQL 和 C 之外的其他语言编写用户定义的函数。这些语言统称为过程语言（PL）。对于用过程语言编写的函数，
数据库服务器并不内置关于如何解释函数源文本的知识。相反，
这项任务会交给一个了解该语言细节的专门的调用处理器。该调用处理器既可以自行完成解析、语法分析、执行等全部工作，也可以在
PostgreSQL与某种现有编程语言实现之间充当“粘合剂”。与其他任何 C 函数一样，调用处理器本身也是一个被编译进共享对象并按需装载的 C 语言函数。

标准PostgreSQL发行版当前提供四种过程语言：PL/pgSQL（第 36 章）、PL/Tcl（第 37 章）、PL/Perl（第 38 章）以及 PL/Python（第 39 章）。用户还可以自行定义其他语言。开发新过程语言的基础知识见第 46 章。

另有一些可用的过程语言并未包含在核心发行版中。附录 H提供了查找它们的信息。

35.1. 安装过程语言

过程语言必须在每个要使用它的数据库中“安装”。不过，安装在数据库`template1`
中的过程语言会自动在随后创建的所有数据库中可用，因为它们在`template1`中的条目会由`CREATE DATABASE`复制。因此，数据库管理员可以决定哪些数据库提供哪些语言，并且如果愿意，也可以让某些语言默认可用。

对于标准发行版自带的语言，只需执行 `CREATE LANGUAGE` *`language_name`*，即可将该语言安装到当前数据库中。此外，也可以从 shell 命令行使用程序createlang来完成这项工作。例如，要把语言 PL/pgSQL 安装到数据库 `template1`，可以使用：

```
createlang plpgsql template1
```

下文所述的手工过程只建议用于安装 `CREATE LANGUAGE` 不认识的自定义语言。

手工安装过程语言

在数据库中安装过程语言需要四个步骤，且必须由数据库超级用户执行。（对于 `CREATE LANGUAGE` 已认识的语言，可以省略第二和第三步，因为这些步骤会自动完成 if needed.)

1. 该语言调用处理器的共享对象必须先编译好并安装到合适的库目录中。
   这与构建和安装含有普通用户定义 C 函数的模块的方法相同；见第 32.9.6 节。该语言调用处理器往往还会依赖一个提供实际编程语言引擎的外部库；如果是这样，该库也必须安装。

2. 必须用如下命令声明该调用处理器：

   ```
   CREATE FUNCTION handler_function_name()
       RETURNS language_handler
       AS 'path-to-shared-object'
       LANGUAGE C;
   ```

   特殊返回类型`language_handler`会告诉数据库系统，该函数返回的不是某种已定义的SQL数据类型，并且不能在 SQL语句中直接使用。

3. (可选的) 可选地，语言调用处理器可以提供一个“验证器”函数，用于在不实际执行的情况下检查函数定义是否正确。如果存在验证器函数，`CREATE FUNCTION` 就会调用它。如果该语言提供了验证器函数，可用类似下面的命令声明它：

```
CREATE FUNCTION validator_function_name(oid)
    RETURNS void
    AS 'path-to-shared-object'
    LANGUAGE C;
```

4. 该 PL 必须用如下命令声明：

```
CREATE [TRUSTED] [PROCEDURAL] LANGUAGE language-name
    HANDLER handler_function_name
    [VALIDATOR validator_function_name] ;
```

可选关键字 `TRUSTED` 表示该语言不会授予用户原本不具备的数据访问能力。受信任的语言是为普通数据库用户（即没有超级用户权限的用户）设计的，允许他们安全地创建函数和触发器过程。由于 PL 函数是在数据库服务器内部执行的，`TRUSTED` 标记只应赋予那些不允许访问数据库服务器内部或文件系统的语言。语言 PL/pgSQL、PL/Tcl 和 PL/Perl 被认为是受信任的；而语言 PL/TclU、PL/PerlU 和 PL/PythonU 旨在提供无限制的功能，因此不应被标记为受信任的。

例 35.1展示了以 PL/pgSQL为例时，手工安装过程是如何进行的。

例 35.1. 手工安装 PL/pgSQL

下面的命令告诉数据库服务器到哪里查找 PL/pgSQL语言调用处理器函数的共享对象：

```
CREATE FUNCTION plpgsql_call_handler() RETURNS language_handler AS
    '$libdir/plpgsql' LANGUAGE C;
```

PL/pgSQL 具有验证器函数，因此我们也声明它：

```
CREATE FUNCTION plpgsql_validator(oid) RETURNS void AS
    '$libdir/plpgsql' LANGUAGE C;
```

下面的命令：

```
CREATE TRUSTED PROCEDURAL LANGUAGE plpgsql
    HANDLER plpgsql_call_handler
    VALIDATOR plpgsql_validator;
```

随后指定应为以下函数和触发器过程调用前面声明的函数：其语言属性为`plpgsql`。

在默认的PostgreSQL安装中，PL/pgSQL语言的调用处理器会被构建并安装到“库”目录中。如果在构建时配置了Tcl支持，那么 PL/Tcl和PL/TclU 的调用处理器会被构建并安装到该库目录中，但这两种语言本身默认并不会安装到任何数据库中。同样，如果在构建时配置了 Perl 支持，就会构建并安装 PL/Perl和PL/PerlU 的调用处理器；如果在构建时配置了 Python 支持，则会安装 PL/PythonU的调用处理器，但这些语言默认也不会安装到任何数据库中。

第 36 章 PL/pgSQL - SQL 过程语言

PL/pgSQL 是一种用于 PostgreSQL 数据库系统的可载入过程语言。PL/pgSQL 的设计目标，是创建一种具备以下特性的可载入过程语言：

- 可用于创建函数和触发器函数，
- 为SQL语言增加控制结构，
- 可以执行复杂计算，
- 继承所有用户定义的类型、函数和操作符，
- 可定义为受服务器信任的语言，
- 便于使用。

除了用户定义类型的输入/输出转换和计算函数之外，凡是能用 C 语言函数定义的东西，也都可以用 PL/pgSQL 完成。例如，可以创建复杂的条件计算函数，随后用它们定义操作符，或者把它们用于索引表达式。

36.1. 概述

PL/pgSQL 调用处理器会在函数第一次被调用时（在每个会话内）解析函数的源文本，并生成一棵内部二进制指令树。这棵指令树完整地翻译了 PL/pgSQL 的语句结构，但函数中使用的各个 SQL 表达式和 SQL 命令并不会被立即翻译。

当每个表达式和 SQL 命令在函数中第一次被使用时，PL/pgSQL 解释器会创建一个预备执行计划（使用 SPI 管理器的 `SPI_prepare` 和 `SPI_saveplan` 函数）。之后再遇到该表达式或命令时，将复用这个预备计划。因此，一个包含大量条件代码、可能需要为许多语句生成执行计划的函数，
只会准备和保存数据库连接生命周期内真正用到的那些计划。这可以显著减少为 PL/pgSQL 函数中的语句解析并生成执行计划所需的总时间。一个缺点是，特定表达式或命令中的错误，
可能要到执行到函数的那一部分时才会被发现。

一旦 PL/pgSQL 为函数中的某条命令生成了执行计划，它就会在数据库连接的整个生命周期内复用该计划。通常这有利于性能，但如果你动态地修改数据库模式，就可能带来一些问题。例如：

```
CREATE FUNCTION populate() RETURNS integer AS $$
DECLARE
    -- declarations
BEGIN
    PERFORM my_function();
END;
$$ LANGUAGE plpgsql;
```

如果你执行上面的函数，它会为 `PERFORM` 语句生成的执行计划引用 `my_function()` 的 OID。此后，如果你删除并重新创建 `my_function()`，那么 `populate()` 将再也找不到 `my_function()`。你将不得不重新创建 `populate()`，或者至少启动一个新的数据库会话，使其重新被编译。避免这一问题的另一种方法是，在更新 `my_function()` 的定义时使用 `CREATE OR REPLACE FUNCTION`（函数被“替换”时，其 OID 不会改变）。

由于 PL/pgSQL 以这种方式保存执行计划，直接出现在 PL/pgSQL 函数中的 SQL 命令在每次执行时必须引用相同的表和列；也就是说，不能把参数用作 SQL 命令中表或列的名称。要绕开这一限

制，可以使用 PL/pgSQL 的 `EXECUTE` 语句构造动态命令——代价是每次执行都要构造一个新的执行计划。

> **注意**
>
> PL/pgSQL 的 `EXECUTE` 语句与 PostgreSQL 服务器支持的EXECUTE SQL 语句无关。服务器的 `EXECUTE` 语句不能用在 PL/pgSQL 函数中（也用不着它）。

36.1.1. 使用PL/pgSQL的优点

SQL 是 PostgreSQL 以及大多数其他关系数据库使用的查询语言。它具有可移植性，也容易学习。但每条 SQL 语句都必须由数据库服务器单独执行。

这意味着客户端应用必须把每个查询发送给数据库服务器，等待其处理完成，接收并处理结果，做一些计算，然后再向服务器发送更多查询。这一切都会产生进程间通信开销；
如果客户端与数据库服务器不在同一台机器上，还会额外产生网络开销。

借助 PL/pgSQL，你可以把一段计算逻辑和一系列查询放在数据库服务器内部执行。这样既具备了过程语言的能力，也保留了 SQL 的易用性，同时还能显著减少客户端/服务器通信开销。

- 消除了客户端和服务器之间额外的往返通信
- 客户端不需要的中间结果不必在服务器与客户端之间编组或传输
- 可以避免多轮查询解析

与不使用存储函数的应用相比，这可能带来相当可观的性能提升。

此外，通过 PL/pgSQL 你可以使用 SQL 的全部数据类型、操作符和函数。

36.1.2. 支持的参数和结果数据类型

用 PL/pgSQL 编写的函数，可以接受服务器支持的任意标量或数组数据类型作为参数，也可以返回这些类型中的任意一种。它们还可以接受或返回任何按名称指定的复合类型（行类型）。
也可以把 PL/pgSQL 函数声明为返回 `record`，这表示结果是一种行类型，其列由调用查询中的指定内容决定，如第 7.2.1.4 节所述。

PL/pgSQL 函数也可以声明为接受和返回多态类型 `anyelement` 和 `anyarray`。多态函数处理的实际数据类型可以在不同调用之间变化，如第 32.2.5 节所述。示例见第 36.4.1 节。

PL/pgSQL 函数还可以声明为返回它能够返回单个实例的任意数据类型的“集合”（或表）。这类函数通过对结果集中每个需要的元素执行 `RETURN NEXT` 来生成输出。

最后，如果某个 PL/pgSQL 函数没有有意义的返回值，可以将其声明为返回 `void`。

> **注意**
>
> PL/pgSQL目前还没有对域类型的完整支持：它把域当作底层标量类型同样对待。
> 这意味着与域关联的约束不会被强制执行。这对函数参数来说不是问题，但如果你把一个PL/pgSQL函数声明为返回域类型，就存在隐患。

PL/pgSQL 函数也可以声明为使用输出参数，而不是显式指定返回类型。这并没有为该语言增加新的基本能力，但通常会更方便，尤其是在需要返回多个值时。

具体示例见第 36.4.1 节和第 36.7.1 节。

36.2. PL/pgSQL 开发提示

在 PL/pgSQL 中进行开发的一种好方法，是用你喜欢的文本编辑器编写函数，并在另一个窗口里使用 psql 载入和测试这些函数。如果你采用这种方式，最好使用 `CREATE OR REPLACE FUNCTION` 来编写函数。这样只需重新载入该文件，就可以更新函数定义。例如：

```
CREATE OR REPLACE FUNCTION testfunc(integer) RETURNS integer AS $$
          ....
$$ LANGUAGE plpgsql;
```

在运行 psql 期间，你可以用下面的命令载入或者重载这样一个函数定义文件：

```
\i filename.sql
```

并且接着立即发出 SQL 命令来测试该函数。

在 PL/pgSQL 中进行开发的另一种好方法是使用便于过程语言开发的 GUI 数据库访问工具。这类工具的一个例子是 pgAdmin，当然还有其他工具。这些工具通常提供诸如转义单引号、更便于重新创建和调试函数之类的便捷特性。

36.2.1. 引号的处理

一个 PL/pgSQL 函数的代码在一个 `CREATE FUNCTION` 中被指定为一个字符串字面量。如果你用通常的方式把该字符串写在单引号中间，那么该函数体中的任何单引号都必须被双写；
同样任何反斜线也必须被双写。双写引号本身就很繁琐，
并且在更复杂的情况中代码会变得完全无法理解，
因为你很容易发现你需要半打或者更多相邻的引号。
我们推荐你转而把函数体写成一个“美元引用”的字符串字面量（见第 4.1.2.2 节）。在美元引用方法中，你从不需要双写任何引号。
但是要注意为你需要的每一层嵌套选择一个不同的美元引用定界符。例如，你可能把 `CREATE FUNCTION` 命令写成：

```
CREATE OR REPLACE FUNCTION testfunc(integer) RETURNS integer AS $PROC$
          ....
$PROC$ LANGUAGE plpgsql;
```

在这里面，你可以在 SQL 命令中为简单字符串使用引号并且用 `$$` 来界定被你组装成字符串的 SQL 命令片段。如果你需要引用包括 `$$` 的文本，你可以使用 `$Q$` 等等。

下表展示了在不使用美元引用时，编写引号必须采取的做法。
在把旧式的非美元引用代码改写成更易理解的形式时，它可能会有所帮助。

1 个引号

用来开始和结束函数体，例如：

```
CREATE FUNCTION foo() RETURNS integer AS '
          ....
' LANGUAGE plpgsql;
```

在使用单引号包围的函数体中的任何位置，引号标记都必须成对出现。

2 个引号

用于函数体内的字符串字面量，例如：

```
a_output := ''Blah'';
SELECT * FROM users WHERE f_name=''foobar'';
```

采用美元引用的方法时，你只需要写：

```
a_output := 'Blah';
SELECT * FROM users WHERE f_name='foobar';
```

这恰好就是 PL/pgSQL 解析器在两种情况中会看到的。

4 个引号

当你在函数内的一个字符串常量中需要一个单引号时，例如：

```
a_output := a_output || '' AND name LIKE ''''foobar'''' AND xyz''
```

实际会被追加到 `a_output` 的值将是：`AND name LIKE 'foobar' AND xyz`。

采用美元引用的方法时，你可以写：

```
a_output := a_output || $$ AND name LIKE 'foobar' AND xyz$$
```

要小心在这周围的任何美元引用定界符都不能只是 `$$`。

6 个引号

当在函数体内的一个字符串中的一个单引号与该字符串常量末尾相邻，例如：

```
a_output := a_output || '' AND name LIKE ''''foobar''''''
```

此时被追加到 `a_output` 的值将是：`AND name LIKE 'foobar'`.

采用美元引用的方法时，它变成：

```
a_output := a_output || $$ AND name LIKE 'foobar'$$
```

10 个引号

当你想要在一个字符串常量中使用两个单引号（这需要 8 个引号），并且它们与该字符串常量的末尾相邻（还需 2 个）时。你可能只在编写生成其他函数的函数时才需要这样做，例如例 36.6。例如：

```
a_output := a_output || '' if v_'' ||
    referrer_keys.kind || '' like '''''''''
    || referrer_keys.key_string || '''''''''
    then return ''''''  || referrer_keys.referrer_type
    || ''''''; end if;'';
```

此时 `a_output` 的值将是：

```
if v_... like ''...'' then return ''...''; end if;
```

采用美元引用的方法时，它变成：

```
a_output := a_output || $$ if v_$$ || referrer_keys.kind || $$ like
 '$$
    || referrer_keys.key_string || $$'
    then return '$$  || referrer_keys.referrer_type
    || $$'; end if;$$;
```

这里我们假定只需要把单引号放进 a_output，因为它在使用前会被重新加引号。

另一种变通做法是用反斜线而不是双写来转义函数体中的引号标记。使用这种方法时，你会发现自己写的是\'\'而不是''''。有些人觉得这样更容易记，有些人不这么认为。

36.3. PL/pgSQL的结构

PL/pgSQL 是一种块结构语言。函数定义的完整文本必须是一个块。块的定义如下：

```
[ <<label>> ]
[ DECLARE
    declarations ]
BEGIN
    statements
END [ label ];
```

块中的每个声明和每条语句都以分号结束。如上所示，嵌套在其他块中的块，必须在 END 之后写分号；但是结束整个函数体的最后一个 END 不需要分号。

所有关键字都不区分大小写。标识符除非用双引号括起，否则会像普通 SQL 命令中那样被隐式转换为小写。

PL/pgSQL 代码中的注释与普通 SQL 相同。双连字符（--）开始一段持续到行尾的注释。/* 开始一段块注释，它会持续到匹配的 */。块注释不能嵌套，但双横线注释可以被包在块注释中，而且双横线可以隐藏块注释定界符 /* 和 */。

块的语句部分中的任意语句都可以是一个子块。子块可用于逻辑分组，或者把变量局部化到一小组语句中。

块前面的声明部分中声明的变量，在每次进入该块时都会被初始化为它们的默认值，而不仅仅是每次函数调用时初始化一次。例如：

```
CREATE FUNCTION somefunc() RETURNS integer AS $$
DECLARE
    quantity integer := 30;
BEGIN
    RAISE NOTICE 'Quantity here is %', quantity;  -- Quantity here is
 30
    quantity := 50;
    --
    -- 创建一个子块
    --
    DECLARE
        quantity integer := 80;
    BEGIN
        RAISE NOTICE 'Quantity here is %', quantity;  -- Quantity here
 is 80
    END;
```

```
    RAISE NOTICE 'Quantity here is %', quantity;  -- Quantity here is
 50

    RETURN quantity;
END;
$$ LANGUAGE plpgsql;
```

需要特别注意，不要把BEGIN/END在PL/pgSQL中用于分组语句的用法，
与事务控制的数据库命令混淆。PL/pgSQL中的BEGIN/END只用于分组，不会开始或结束事务。
函数和触发器函数总是在外层查询建立的事务中执行 – 它们不能开始或提交该事务，否则就没有可供其执行的上下文。但是，包含EXCEPTION
子句的块实际上会形成一个可回滚而不影响外层事务的子事务。详见第 36.7.5 节。

36.4. 声明

在一个块中使用的所有变量，都必须在该块的声明部分声明。（唯一的例外是 FOR 循环的循环变量会被自动声明为整数变量。）

PL/pgSQL 变量可以是任意 SQL 数据类型，例如 integer、varchar 和 char。

这里是变量声明的一些示例：

```
user_id integer;
quantity numeric(5);
url varchar;
myrow tablename%ROWTYPE;
myfield tablename.columnname%TYPE;
arow RECORD;
```

变量声明的一般语法是：

```
name [ CONSTANT ] type [ NOT NULL ] [ { DEFAULT | := } expression ];
```

如果给出了 DEFAULT 子句，它指定进入该块时赋给变量的初始值。如果没有给出 DEFAULT 子句，则变量被初始化为 SQL 空值。CONSTANT 选项阻止给变量赋值，因此它的值在整个块的持续期间保持不变。如果指定了 NOT NULL，赋予空值会导致运行时错误。所有声明为 NOT NULL 的变量都必须指定一个非空默认值。

变量的默认值会在每次进入该块时重新计算并赋给该变量，而不是每次函数调用只计算一次。
因此，例如把 now() 赋给一个 timestamp 类型变量，会使该变量得到当前函数调用时的时间，而不是函数预编译时的时间。

例如：

```
quantity integer DEFAULT 32;
url varchar := 'http://mysite.com';
user_id CONSTANT integer := 10;
```

36.4.1. 函数参数的别名

传递给函数的参数使用标识符 $1、$2 等命名。也可以为这些 $*n* 形式的参数名声明别名，以提高可读性。之后既可以使用别名，也可以使用数字标识符来引用参数值。

有两种方式可以创建别名。首选方式是在 CREATE FUNCTION 命令中直接为参数命名。例如：

```
CREATE FUNCTION sales_tax(subtotal real) RETURNS real AS $$
BEGIN
    RETURN subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

另一种方式（在 PostgreSQL 8.0 之前这是唯一可用的方式）是使用声明语法显式声明别名

name ALIAS FOR $*n*;

同样的示例，按这种写法如下：

```
CREATE FUNCTION sales_tax(real) RETURNS real AS $$
DECLARE
    subtotal ALIAS FOR $1;
BEGIN
    RETURN subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

更多一些示例：

```
CREATE FUNCTION instr(varchar, integer) RETURNS integer AS $$
DECLARE
    v_string ALIAS FOR $1;
    index ALIAS FOR $2;
BEGIN
    -- some computations using v_string and index here
END;
$$ LANGUAGE plpgsql;


CREATE FUNCTION concat_selected_fields(in_t sometablename) RETURNS
 text AS $$
BEGIN
    RETURN in_t.f1 || in_t.f3 || in_t.f5 || in_t.f7;
END;
$$ LANGUAGE plpgsql;
```

当 PL/pgSQL 函数声明了输出参数时，输出参数也会像普通输入参数一样获得 $*n* 名称和可选别名。输出参数本质上是一个初始值为 NULL 的变量，应在函数执行期间给它赋值。该参数的最终值就是返回值。例如，销售税的示例也可以这样写：

```
CREATE FUNCTION sales_tax(subtotal real, OUT tax real) AS $$
BEGIN
    tax := subtotal * 0.06;
END;
$$ LANGUAGE plpgsql;
```

注意这里省略了 RETURNS real – 当然也可以写上，但那只是冗余。

当需要返回多个值时，输出参数尤其有用。下面是一个简单示例：

```
CREATE FUNCTION sum_n_product(x int, y int, OUT sum int, OUT prod int)
 AS $$
```

```
BEGIN
    sum := x + y;
    prod := x * y;
END;
$$ LANGUAGE plpgsql;
```

如第 32.4.3 节所述，这实际上会为函数结果创建一个匿名记录类型。如果写了 RETURNS 子句，它必须是 RETURNS record。

当 PL/pgSQL 函数的返回类型被声明为多态类型（anyelement 或 anyarray）时，会创建一个特殊参数 $0。它的数据类型就是该函数的实际返回类型，由实际输入类型推导而来（见第 32.2.5 节）。这样，函数就能像第 36.4.2 节所示的那样访问自己的实际返回类型。$0 被初始化为空值，并且可以由函数修改，因此如果需要，可以用它来保存返回值，不过这并不是必须的。$0 也可以被赋予别名。例如，下面的函数适用于任何带有 + 操作符的数据类型：

```
CREATE FUNCTION add_three_values(v1 anyelement, v2 anyelement, v3
 anyelement)
RETURNS anyelement AS $$
DECLARE
    result ALIAS FOR $0;
BEGIN
    result := v1 + v2 + v3;
    RETURN result;
END;
$$ LANGUAGE plpgsql;
```

把一个或多个输出参数声明为 anyelement 或 anyarray，也可以达到同样的效果。在这种情况下，不使用特殊参数 $0，输出参数本身就承担相同作用。例如：

```
CREATE FUNCTION add_three_values(v1 anyelement, v2 anyelement, v3
 anyelement,
                                 OUT sum anyelement)
AS $$
BEGIN
    sum := v1 + v2 + v3;
END;
$$ LANGUAGE plpgsql;
```

36.4.2. 复制类型

variable%TYPE

%TYPE提供变量或表列的数据类型。可以用它声明将保存数据库值的变量。例如，假设有一个名为user_id的列，位于users表中。要声明一个数据类型与users.user_id相同的变量，可以写：

```
user_id users.user_id%TYPE;
```

使用 %TYPE 的好处是，你不必知道所引用结构的实际数据类型；更重要的是，如果被引用项的数据类型将来发生变化（例如把 user_id 的类型从 integer 改成 real），你可能就不需要修改函数定义。

%TYPE 在多态函数中特别有价值，因为内部变量所需的数据类型可能在不同调用之间变化。可以把 %TYPE 应用到函数参数或结果占位符上，以创建合适的变量。

36.4.3. 行类型

nametable_name%ROWTYPE;
name composite_type_name;

复合类型的变量称为行变量（或行类型变量）。只要查询的列集合与该变量声明的类型相匹配，这种变量就可以保存 SELECT 或 FOR 查询结果中的整行。行值的各个字段可以使用通常的点号记法访问，例如 rowvar.field。

行变量既可以通过 *table_name*%ROWTYPE 记法声明为与现有表或视图的行具有相同类型，也可以通过给出某个复合类型的名称来声明。（由于每个表都有一个同名的关联复合类型，所以在 PostgreSQL 中实际上写不写 %ROWTYPE 并无区别；不过带 %ROWTYPE 的形式可移植性更好。）

函数参数也可以是复合类型（完整的表行）。在这种情况下，相应的标识符 $*n* 就是一个行变量，并且可以从中选取字段，例如 $1.user_id。

在行类型变量中，只能访问表行中用户定义的列，不能访问 OID 或其他系统列（因为该行可能来自视图）。对于char(*n*)这样的数据类型，行类型中的字段会继承表字段的大小或精度。

下面是一个使用复合类型的示例。table1 和 table2 是已经存在的表，它们至少包含下面提到的字段：

```
CREATE FUNCTION merge_fields(t_row table1) RETURNS text AS $$
DECLARE
    t2_row table2%ROWTYPE;
BEGIN
    SELECT * INTO t2_row FROM table2 WHERE ... ;
    RETURN t_row.f1 || t2_row.f3 || t_row.f5 || t2_row.f7;
END;
$$ LANGUAGE plpgsql;

SELECT merge_fields(t.*) FROM table1 t WHERE ... ;
```

36.4.4. 记录类型

name RECORD;

记录变量与行类型变量类似，但没有预定义结构。它会在 SELECT 或 FOR 命令为其赋值时采用相应行的实际结构。记录变量的内部结构在每次被赋值时都可能变化。
其后果是：在记录变量第一次被赋值之前，它没有任何子结构，
任何试图访问其中字段的行为都会引发运行时错误。

注意，RECORD 并不是真正的数据类型，它只是一个占位符。还需要认识到，PL/pgSQL 函数被声明为返回 record，与记录变量并不是完全相同的概念，尽管这样的函数可能会用记录变量保存结果。这两种情况下，在编写函数时都不知道实际的行结构；但对于返回 record 的函数，实际结构会在解析调用查询时确定，而记录变量的行结构则可以在运行过程中随时变化。

36.4.5. RENAME

RENAME *oldname* TO *newname*;

使用 RENAME 声明可以更改变量、记录或行的名称。这主要在触发器过程中需要用别的名称引用 NEW 或 OLD 时有用。另请参见 ALIAS。

示例：

```
RENAME id TO user_id;
RENAME this_var TO that_var;
```

> **注意**
>
> RENAME 从 PostgreSQL 7.3 起似乎已损坏。修复它的优先级较低，因为 ALIAS 已经覆盖了 RENAME 的大部分实际用途。

36.5. 表达式

PL/pgSQL 语句中使用的所有表达式都由服务器的主 SQL 执行器处理。实际上，像下面这样的查询

```
SELECT expression
```

是通过 SPI 管理器执行的。在求值之前，出现的 PL/pgSQL 变量标识符会被替换为参数，而变量的实际值则通过参数数组传递给执行器。这使得 SELECT 的查询计划只需准备一次，随后就可以重复用于后续的求值。

PostgreSQL 主解析器所做的求值，对常量值的解释有一些副作用。具体来说，下面这两个函数的行为有所不同：

```
CREATE FUNCTION logfunc1(logtxt text) RETURNS timestamp AS $$
    BEGIN
        INSERT INTO logtable VALUES (logtxt, 'now');
        RETURN 'now';
    END;
$$ LANGUAGE plpgsql;
```

和

```
CREATE FUNCTION logfunc2(logtxt text) RETURNS timestamp AS $$
    DECLARE
        curtime timestamp;
    BEGIN
        curtime := 'now';
        INSERT INTO logtable VALUES (logtxt, curtime);
        RETURN curtime;
    END;
$$ LANGUAGE plpgsql;
```

在logfunc1的例子中，PostgreSQL主解析器在为INSERT准备计划时就知道字符串'now'应被解释为timestamp，因为logtable的目标列就是该类型。这样，'now'会在INSERT完成规划时被转换为一个常量，然后在该会话的整个生命周期内logfunc1的所有调用中都使用它。毫无疑问，这并不是程序员想要的结果。

在logfunc2的例子中，PostgreSQL主解析器不知道'now'应转换成什么类型，因此它返回一个包含字符串now的text类型数据值。在随后向局部变量curtime赋值时，PL/pgSQL解释器通过调用text_out和timestamp_in函数进行转换，把这个字符串强制转换为timestamp类型。这样，计算得到的时间戳在每次执行时都会更新，正符合程序员的期望。

记录变量的可变本性在此方面带来一个问题。当记录变量的字段被用于表达式或语句中时，各字段的数据类型在同一表达式的多次调用之间不得改变，因为表达式将使用第一次到达时出现的数据类型进行规划。在编写处理多张表事件的触发器过程时，请牢记这一点。（必要时可以用EXECUTE来绕过这一问题。）

36.6. 基本语句

在本节和后续各节中，我们描述PL/pgSQL能够明确理解的所有语句类型。凡是不被识别为这些类型之一的内容，都被认为是 SQL 命令，并在替换其中使用的所有PL/pgSQL变量后被发送到主数据库引擎执行。因此，例如 SQL 命令INSERT、UPDATE和DELETE可以被认为是PL/pgSQL的语句，但这里没有专门列出它们。

36.6.1. 赋值

给变量或行/记录字段赋一个值的写法是：

```
identifier := expression;
```

如前所述，这种语句中的表达式是通过向主数据库引擎发送一条 SQL SELECT 命令来求值的。表达式必须产生单一值。

如果该表达式的结果数据类型不匹配变量的数据类型，或者变量具有特定的大小/精度（例如char(20)），结果值将由 PL/pgSQL解释器使用结果类型的输出函数和变量类型的输入函数隐式转换。注意如果结果值的字符串形式无法被输入函数所接受，这可能会导致由输入函数产生的运行时错误。

例如：

```
tax := subtotal * 0.06;
my_record.user_id := 20;
```

36.6.2. SELECT INTO

产生多列（但只有一行）结果的SELECT命令，其结果可以赋给一个记录变量、行类型变量或标量变量列表。方法是通过：

```
SELECT INTO target select_expressions FROM ...;
```

其中*target*可以是一个记录变量、一个行变量，或者一个逗号分隔的简单变量及记录/行字段列表。*select_expressions*和命令的其余部分与普通 SQL 中的写法相同。

注意，这与PostgreSQL对SELECT INTO的通常解释完全不同，后者的INTO目标是一个新创建的表。如果想在PL/pgSQL函数内部根据SELECT结果创建表，请使用CREATE TABLE ... AS SELECT语法。

如果使用行变量或变量列表作为目标，选出的值必须与目标的结构完全匹配，否则会发生运行时错误。当目标是记录变量时，它会自动将自身配置为查询结果列的行类型。

除了INTO子句以外，SELECT语句与普通 SQLSELECT命令相同，可以使用其全部能力。

INTO子句几乎可以出现在SELECT语句的任何位置。习惯上它要么像上面那样紧接在SELECT之后，要么紧接在FROM之前——也就是*select_expressions*列表之前或之后。

如果查询返回零行，会把空值赋给目标。如果查询返回多行，第一行会被赋给目标，其余的行将被丢弃。（注意，除非使用了ORDER BY，否则“第一行”并没有明确的定义。）

可以在SELECT INTO语句之后检查特殊的FOUND变量（见第 36.6.6 节），以确定赋值是否成功，即查询是否至少返回了一行。例如：

```
SELECT INTO myrec * FROM emp WHERE empname = myname;
IF NOT FOUND THEN
    RAISE EXCEPTION 'employee % not found', myname;
END IF;
```

要测试记录/行的结果是否为空，可以使用IS NULL条件。但是，没有办法判断是否还有额外的行被丢弃。下面是一个处理未返回行情况的示例：

```
DECLARE
    users_rec RECORD;
BEGIN
    SELECT INTO users_rec * FROM users WHERE user_id=3;

    IF users_rec.homepage IS NULL THEN
        -- user entered no homepage, return "http://"
        RETURN 'http://';
    END IF;
END;
```

36.6.3. 执行没有结果的表达式或查询

有时我们会希望对一个表达式或查询求值但丢弃结果（典型情况是调用一个有有用副作用但没有有用结果值的函数）。要在PL/pgSQL中做到这一点，可以使用PERFORM语句：

```
PERFORM query;
```

这会执行*query*并丢弃结果。编写*query*的方式与编写 SQLSELECT命令相同，只需把开头的关键字SELECT替换为PERFORM。PL/pgSQL变量会照常被替换到查询中。另外，如果查询至少产生了一行，特殊变量FOUND会被设为真；如果未产生行，则设为假。

> **注意**
>
> 我们可能期望不带INTO子句的SELECT能实现这个结果，但当前唯一被接受的方式是PERFORM。

一个示例：

```
PERFORM create_mv('cs_session_page_requests_mv', my_query);
```

36.6.4. 什么也不做

有时一个什么也不做的占位语句也很有用。例如，它能够指示 if/then/else 链中故意留出的空分支。可以使用NULL语句达到这个目的：

```
NULL;
```

例如，下面的两段代码是等价的：

```
BEGIN
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN
        NULL;  -- 忽略错误
END;

BEGIN
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN  -- 忽略错误
END;
```

究竟使用哪一种取决于各人的喜好。

> **注意**
>
> 在 Oracle 的 PL/SQL 中，不允许出现空语句列表，并且因此在这种情况下必须使用 `NULL` 语句。而 PL/pgSQL 允许你什么也不写。

36.6.5. 执行动态命令

很多时候你将想要在PL/pgSQL函数中产生动态命令，
也就是每次执行中会涉及到不同表或不同数据类型的命令。
PL/pgSQL通常对命令缓存计划的尝试在这种情境下无法工作。要处理这一类问题，提供了`EXECUTE`语句：

```
EXECUTE command-string [ INTO target ];
```

其中*`command-string`*是一个表达式，其求值结果（类型为 `text`）是包含待执行命令的字符串，而*`target`*是一个记录变量、一个行变量或者一个逗号分隔的简单变量以及记录/行字段的列表。

特别要注意，在计算得到的命令字符串上不会做PL/pgSQL变量的替换。
变量的值必须在构造命令字符串时被插入其中。

与PL/pgSQL中的所有其他命令不同，通过`EXECUTE`
语句运行的命令不会在会话生命周期内只准备和保存一次。相反，
每次运行该语句时都会重新准备命令。命令字符串可以在函数内动态创建，
以对不同的表和列执行操作。

`INTO`子句指定一个`SELECT`命令的结果应该被赋值到哪里。如果提供了一个行变量或变量列表，
它必须完全匹配该`SELECT`命令结果的结构（当一个记录变量被提供时，
它会自动把它自己配置为匹配结果结构）。如果返回多个行，只有第一个行会被赋值给`INTO`
变量。如果没有返回行，NULL 会被赋值给`INTO`变量。如果没有指定`INTO`子句，`SELECT`命令的结果将被丢弃。

目前不支持在`EXECUTE`中使用`SELECT INTO`。

在使用动态命令时，经常需要处理单引号的转义。
我们推荐在函数体中使用美元引用来引用固定文本。（如果你有未使用美元引用的旧代码，
请参阅第 36.2.1 节中的概述；在把这类代码转换成更合理的写法时，它会帮你省下一些工夫。）

将被插入到构造出的查询中的动态值需要被小心地处理，因为它们自身可能包含引号字符。
一个示例（这假设你对函数整体使用了美元引用，因此引号标记不需要被双写）：

```
EXECUTE 'UPDATE tbl SET '
        || quote_ident(colname)
        || ' = '
        || quote_literal(newvalue)
        || ' WHERE key = '
        || quote_literal(keyvalue);
```

这个示例展示了quote_ident和quote_literal函数的用法。为了安全起见，
包含列和表标识符的表达式应被传给quote_ident。
包含在构造出的命令中应作为字符串字面量出现的值的表达式，应被传给quote_literal。
这两个函数都会采取适当措施，分别返回用双引号或单引号括起来的输入文本，
并正确转义其中嵌入的特殊字符。

请注意美元符号引用只对引用固定文本有用。
把上面的示例写成下面这样将是一个非常糟糕的主意：

```
EXECUTE 'UPDATE tbl SET '
        || quote_ident(colname)
        || ' = $$'
        || newvalue
        || '$$ WHERE key = '
        || quote_literal(keyvalue);
```

因为如果newvalue的内容碰巧含有$$，那么这段代码就会出问题。
同样的问题也适用于你选择的任何其他美元符号引用定界符。因此，
要想安全地引用事先不知道的文本，必须使用quote_literal。

动态命令和EXECUTE的一个更大的示例可以在例 36.6中找到，它会构建并且执行一个CREATE FUNCTION命令来定义一个新的函数。

36.6.6. 获取结果状态

有好几种方法可以判断一条命令的效果。第一种方法是使用 GET DIAGNOSTICS 命令，其形式如下：

```
GET DIAGNOSTICS variable = item [ , ... ];
```

这条命令允许检索系统状态指示符。每个 *item* 是一个关键字，它标识一个要赋给指定变量（该变量应具有正确的数据类型来接收它）的状态值。当前可用的状态项是 ROW_COUNT（发送到 SQL 引擎的最后一条 SQL 命令所处理的行数）和 RESULT_OID（最近一条 SQL 命令插入的最后一行的 OID）。注意，RESULT_OID 只有在向包含 OID 的表执行 INSERT 命令之后才有用。

一个示例：

```
GET DIAGNOSTICS integer_var = ROW_COUNT;
```

第二种确定命令效果的方法是检查名为 FOUND 的特殊变量，其类型为 boolean。在每次 PL/pgSQL 函数调用开始时，FOUND 都为假。以下各类语句会设置它：

- 如果 SELECT INTO 语句赋予了一行，则把 FOUND 设为真；如果没有返回行，则设为假。
- 如果 PERFORM 语句产生（并丢弃）了一行或多行，则把 FOUND 设为真；如果没有产生行，则设为假。

- 如果 UPDATE、INSERT 和 DELETE 语句至少影响了一行，则把 FOUND 设为真；如果没有影响行，则设为假。
- 如果 FETCH 语句返回了一行，则把 FOUND 设为真；如果没有返回行，则设为假。
- 如果 FOR 语句迭代了一次或多次，则把 FOUND 设为真，否则设为假。这适用于 FOR 语句的所有三种变体（整数 FOR 循环、记录集 FOR 循环和动态记录集 FOR 循环）。FOUND 是在 FOR 循环退出时按上述方式设置的；在循环执行期间，FOUND 不会被 FOR 语句本身修改，尽管循环体内其他语句的执行可能会改变它。

FOUND 是每个 PL/pgSQL 函数中的局部变量；对它的任何修改都只影响当前函数。

36.7. 控制结构

控制结构可能是PL/pgSQL中最有用的（以及最重要）的部分了。利用PL/pgSQL的控制结构，你可以以非常灵活而且强大的方法操纵PostgreSQL的数据。

36.7.1. 从函数返回

有两个命令让我们能够从函数中返回数据：RETURN和RETURN NEXT。

36.7.1.1. RETURN

RETURN *expression*;

带有一个表达式的RETURN用于终止函数并把*expression*的值返回给调用者。
这种形式被用于不返回集合的PL/pgSQL函数。

如果函数返回的是标量类型，任何表达式都可以使用。
表达式结果会按照赋值部分的说明自动转换为函数的返回类型。但要返回一个复合（行）值，你必须把记录变量或行变量写为 *expression*。

如果你声明带输出参数的函数，那么就只需要写不带表达式的RETURN。
输出参数变量的当前值将被返回。

如果你声明函数返回void，一个RETURN语句可以被用来提前退出函数；但是不要在RETURN后面写一个表达式。

一个函数的返回值不能是未定义。如果控制到达了函数最顶层块的末尾而没有碰到一个RETURN语句，那么会发生一个运行时错误。不过，这个限制不适用于带输出参数的函数以及返回void的函数。在这些情况中，如果顶层的块结束，将自动执行一个RETURN语句。

36.7.1.2. RETURN NEXT

RETURN NEXT *expression*;

当 PL/pgSQL 函数被声明为返回 SETOF *sometype* 时，返回过程会略有不同。在这种情况下，要返回的各个项通过一系列 RETURN NEXT 命令指定，最后再用一个不带参数的 RETURN 命令表明函数已经执行完毕。RETURN NEXT 可用于标量和复合数据类型；对于复合结果类型，会返回完整的结果“表”。

RETURN NEXT 实际上不会从函数中返回 – 它只是把表达式的值保存起来。然后会继续执行PL/pgSQL函数中的下一条语句。随着后继的RETURN NEXT命令的执行，结果集就建立起来了。最后一个RETURN（应该没有参数）会导致控制退出该函数（或者你可以让控制到达函数的结尾）。

如果函数是用输出参数声明的，只需写不带表达式的 `RETURN NEXT`。每次执行时，输出参数变量的当前值都会被保存，最终作为结果的一行返回。注意，为了创建带输出参数的返回集合函数，当有多个输出参数时，必须把函数声明为返回 `SETOF record`；当只有一个类型为 *sometype* 的输出参数时，必须声明为返回 `SETOF` *sometype*。

使用`RETURN NEXT`的函数应当按照下面的方式调用：

```
CREATE TABLE foo (fooid INT, foosubid INT, fooname TEXT);
INSERT INTO foo VALUES (1, 2, 'three');
INSERT INTO foo VALUES (4, 5, 'six');

CREATE OR REPLACE FUNCTION getAllFoo() RETURNS SETOF foo AS
$BODY$
DECLARE
    r foo%rowtype;
BEGIN
    FOR r IN SELECT * FROM foo
    WHERE fooid > 0
    LOOP
        -- can do some processing here
        RETURN NEXT r; -- return current row of SELECT
    END LOOP;
    RETURN;
END
$BODY$
LANGUAGE 'plpgsql' ;

SELECT * FROM getallfoo();
```

也就是说，该函数必须作为表源在`FROM`子句中使用。

> **注意**
>
> 如上所述，目前 PL/pgSQL 的 `RETURN NEXT` 实现在从函数返回之前会把整个结果集都保存起来。这意味着如果一个PL/pgSQL函数生成一个非常大的结果集，性能可能会很差：数据将被写到磁盘上以避免内存耗尽，但是函数本身在整个结果集都生成之前不会退出。将来的PL/pgSQL版本可能会允许用户定义没有这种限制的集合返回函数。目前，数据开始被写入到磁盘的时机由配置变量work_mem控制。拥有足够内存来存储大型结果集的管理员可以考虑增大这个参数。

36.7.2. 条件语句

`IF` 语句让你能够根据特定条件执行命令。PL/pgSQL 有五种形式的 `IF`：

- `IF ... THEN`
- `IF ... THEN ... ELSE`
- `IF ... THEN ... ELSE IF`
- `IF ... THEN ... ELSIF ... THEN ... ELSE`

- IF ... THEN ... ELSEIF ... THEN ... ELSE

36.7.2.1. IF-THEN

```
IF boolean-expression THEN
    statements
END IF;
```

IF-THEN 语句是 IF 的最简单形式。如果条件为真，在 THEN 和 END IF 之间的语句将被执行。否则，将忽略它们。

示例：

```
IF v_user_id <> 0 THEN
    UPDATE users SET email = v_email WHERE user_id = v_user_id;
END IF;
```

36.7.2.2. IF-THEN-ELSE

```
IF boolean-expression THEN
    statements
ELSE
    statements
END IF;
```

IF-THEN-ELSE 语句对 IF-THEN 进行了增加，它让你能够指定一组在条件求值为假时应该被执行的语句。

示例：

```
IF parentid IS NULL OR parentid = ''
THEN
    RETURN fullname;
ELSE
    RETURN hp_true_filename(parentid) || '/' || fullname;
END IF;

IF v_count > 0 THEN
    INSERT INTO users_count (count) VALUES (v_count);
    RETURN 't';
ELSE
    RETURN 'f';
END IF;
```

36.7.2.3. IF-THEN-ELSE IF

IF语句可以嵌套，如下例所示：

```
IF demo_row.sex = 'm' THEN
    pretty_sex := 'man';
ELSE
    IF demo_row.sex = 'f' THEN
        pretty_sex := 'woman';
    END IF;
END IF;
```

使用这种形式时，你实际上是把一个IF语句嵌套在外层IF语句的ELSE部分中。这样，每个嵌套的IF都需要一个END IF语句，外层的IF-ELSE也需要一个。这是可行的，但当要检查的备选情况很多时就变得繁琐了。因此有了下一种形式。

36.7.2.4. IF-THEN-ELSIF-ELSE

```
IF boolean-expression THEN
    statements
[ ELSIF boolean-expression THEN
    statements
[ ELSIF boolean-expression THEN
    statements
    ...]]
[ ELSE
    statements ]
END IF;
```

IF-THEN-ELSIF-ELSE提供了一种在一条语句中检查多种备选情况的更方便的方法。在功能上它等价于嵌套的IF-THEN-ELSE-IF-THEN命令，但只需要一个END IF。

这里有一个示例：

```
IF number = 0 THEN
    result := 'zero';
ELSIF number > 0 THEN
    result := 'positive';
ELSIF number < 0 THEN
    result := 'negative';
ELSE
    -- hmm, the only other possibility is that number is null
    result := 'NULL';
END IF;
```

36.7.2.5. IF-THEN-ELSEIF-ELSE

ELSEIF是ELSIF的一个别名。

36.7.3. 简单循环

使用 LOOP、EXIT、CONTINUE、WHILE 和 FOR 语句，可以让你的 PL/pgSQL 函数重复执行一系列命令。

36.7.3.1. LOOP

```
[ <<label>> ]
LOOP
    statements
END LOOP [ label ];
```

LOOP 定义了一个无条件循环，它会无限重复，直到被 EXIT 或 RETURN 语句终止。可选的 *label* 可以供嵌套循环中的 EXIT 和 CONTINUE 语句使用，以指定这些语句作用的是哪一个循环。

36.7.3.2. EXIT

```
EXIT [ label ] [ WHEN expression ];
```

如果没有给出 *label*，那么最内层的循环会被终止，然后跟在 END LOOP 后面的语句会被执行。如果给出了 *label*，那么它必须是当前或者更高层的嵌套循环或者语句块的标签。然后该命名循环或块就会被终止，并且控制会转移到该循环/块相应的 END 之后的语句上。

如果指定了 WHEN，则只有在 *expression* 为真时才会退出循环。否则，控制传递到 EXIT 之后的语句。

EXIT 可以与所有类型的循环一起使用；它并不限于在无条件循环中使用。当与 BEGIN 块一起使用时，EXIT 会把控制转交给该块结束后的下一条语句。

示例：

```
LOOP
    -- some computations
    IF count > 0 THEN
        EXIT;  -- exit loop
    END IF;
END LOOP;

LOOP
    -- some computations
    EXIT WHEN count > 0;  -- same result as previous example
END LOOP;

<<ablock>>
BEGIN
    -- some computations
    IF stocks > 100000 THEN
        EXIT;  -- causes exit from the BEGIN block
    END IF;
    -- computations here will be skipped when stocks > 100000
END;
```

36.7.3.3. CONTINUE

```
CONTINUE [ label ] [ WHEN expression ];
```

如果没有给出 *label*，最内层循环的下一次迭代会开始。也就是，循环体中剩余的所有语句将被跳过，并且控制会返回到循环控制表达式（如果有）来决定是否需要另一次循环迭代。如果 *label* 存在，它指定应该继续执行的循环的标签。

如果指定了 WHEN，则只有在 *expression* 为真时才会开始循环的下一次迭代。否则，控制传递到 CONTINUE 之后的语句。

CONTINUE 可以与所有类型的循环一起使用；它并不限于在无条件循环中使用。

示例：

```
LOOP
    -- some computations
    EXIT WHEN count > 100;
    CONTINUE WHEN count < 50;
    -- some computations for count IN [50 .. 100]
END LOOP;
```

36.7.3.4. WHILE

```
[ <<label>> ]
WHILE expression LOOP
    statements
END LOOP [ label ];
```

只要条件表达式的计算结果为真，WHILE 语句就会重复一个语句序列。在每次进入到循环体之前都会检查该表达式。

例如：

```
WHILE amount_owed > 0 AND gift_certificate_balance > 0 LOOP
    -- some computations here
END LOOP;

WHILE NOT done LOOP
    -- some computations here
END LOOP;
```

36.7.3.5. FOR（整型变体）

```
[ <<label>> ]
FOR name IN [ REVERSE ] expression .. expression LOOP
    statements
END LOOP [ label ];
```

这种形式的 FOR 会创建一个在一个整数范围上迭代的循环。变量 *name* 会自动定义为类型 integer 并且只在循环内存在。给出范围上下界的两个表达式在进入循环的时候计算一次。迭代步长通常为 1，但在指定了 REVERSE 时为 -1。

整数 FOR 循环的一些示例：

```
FOR i IN 1..10 LOOP
    -- i will take on the values 1,2,3,4,5,6,7,8,9,10 within the loop
END LOOP;

FOR i IN REVERSE 10..1 LOOP
    -- i will take on the values 10,9,8,7,6,5,4,3,2,1 within the loop
END LOOP;
```

如果下界大于上界（或者在 REVERSE 情况下是小于），循环体根本不会被执行。而且不会抛出任何错误。

36.7.4. 遍历查询结果

使用另一种形式的 FOR 循环，可以遍历查询的结果并相应地处理这些数据。语法是：

```
[ <<label>> ]
FOR record_or_row IN query LOOP
    statements
END LOOP [ label ];
```

记录或行变量会被依次赋予*query*（它必须是一个SELECT命令）产生的每一行，
并且循环体对每一行都执行一次。这里有一个示例：

```
CREATE FUNCTION cs_refresh_mviews() RETURNS integer AS $$
DECLARE
    mviews RECORD;
BEGIN
    PERFORM cs_log('Refreshing materialized views...');

    FOR mviews IN SELECT * FROM cs_materialized_views ORDER BY sort_
key LOOP

        -- Now "mviews" has one record from cs_materialized_views

        PERFORM cs_log('Refreshing materialized view '
                   || quote_ident(mviews.mv_name) || ' ...');
        EXECUTE 'TRUNCATE TABLE ' || quote_ident(mviews.mv_name);
        EXECUTE 'INSERT INTO '
                   || quote_ident(mviews.mv_name) || ' '
                   || mviews.mv_query;
    END LOOP;

    PERFORM cs_log('Done refreshing materialized views.');
    RETURN 1;
END;
$$ LANGUAGE plpgsql;
```

如果循环被 EXIT 语句终止，最后一次赋予的行值在循环之后仍然可以访问。

FOR-IN-EXECUTE 语句是另一种遍历行的方式：

```
[ <<label>> ]
FOR record_or_row IN EXECUTE text_expression LOOP
    statements
END LOOP [ label ];
```

它与前一种形式类似，只是源SELECT语句被指定为一个字符串表达式，在每次进入 FOR 循环时都会重新求值和重新规划。
这让程序员可以在预先规划的查询的速度与动态查询的灵活性之间做出选择，就像使用普通的 EXECUTE 语句一样。

> **注意**
>
> PL/pgSQL 解析器目前通过检查 IN 和 LOOP 之间、任何括号之外是否出现 .. 来区分两类 FOR 循环（整数循环和查询结果循环）。如果没有看到 ..，就假定该循环是遍历行的循环。因此，如果把 .. 打错，很可能得到类似“遍历行的循环的循环变量必须是记录或行变量或者标量变量列表”这样的报错，而不是人们预想的简单语法错误。

36.7.5. 捕获错误

默认情况下，PL/pgSQL 函数中发生的任何错误都会中止函数及其外围事务的执行。你可以使用带有 EXCEPTION 子句的 BEGIN 块来捕获错误并从中恢复。其语法是在普通 BEGIN 块语法上的扩展：

```
[ <<label>> ]
[ DECLARE
```

```
    declarations ]
BEGIN
    statements
EXCEPTION
    WHEN condition [ OR condition ... ] THEN
        handler_statements
    [ WHEN condition [ OR condition ... ] THEN
          handler_statements
      ... ]
END;
```

如果没有发生错误，这种形式的块只是简单地执行所有 *statements*，并且接着控制转到 END 之后的下一个语句。但是如果在该 *statements* 内发生了一个错误，则会放弃对 *statements* 的进一步处理，然后控制会转到 EXCEPTION 列表。系统会在列表中寻找匹配所发生错误的第一个 *condition*。如果找到一个匹配，则执行对应的 *handler_statements*，并且接着把控制转到 END 之后的下一个语句。如果没有找到匹配，该错误就会传播出去，就好像根本没有 EXCEPTION 一样：错误可以被一个带有 EXCEPTION 的外围块捕捉，如果没有这样的块则中止该函数的处理。

condition 的名字可以是附录 A中显示的任何名字。一个分类名匹配其中所有的错误。特殊的条件名 OTHERS 匹配除了 QUERY_CANCELED 之外的所有错误类型（虽然可能但通常并不明智，还是可以用名字捕获 QUERY_CANCELED）。条件名是大小写无关的。

如果在选中的 *handler_statements* 内发生了新的错误，那么它不能被这个 EXCEPTION 子句捕获，而是被传播出去。一个外层的 EXCEPTION 子句可以捕获它。

当一个错误被 EXCEPTION 子句捕获时，PL/pgSQL 函数的局部变量会保持错误发生时的值，但是该块中所有对持久数据库状态的改变都会被回滚。例如，考虑这个片段：

```
INSERT INTO mytab(firstname, lastname) VALUES('Tom', 'Jones');
BEGIN
    UPDATE mytab SET firstname = 'Joe' WHERE lastname = 'Jones';
    x := x + 1;
    y := x / 0;
EXCEPTION
    WHEN division_by_zero THEN
        RAISE NOTICE 'caught division_by_zero';
        RETURN x;
END;
```

当控制到达对 y 的赋值时，它会以 division_by_zero 错误失败。该错误会被 EXCEPTION 子句捕获。RETURN 语句返回的值将是 x 递增后的值，但 UPDATE 命令的效果已被回滚。不过，块之前的 INSERT 命令没有被回滚，因此最终结果是数据库中包含的是 Tom Jones 而不是 Joe Jones。

> **提示**
>
> 包含 EXCEPTION 子句的块，其进入和退出的开销明显高于没有该子句的块。因此，不要在没有必要的情况下使用 EXCEPTION。

在异常处理器中，SQLSTATE 变量包含与所引发异常对应的错误代码（可能错误代码的列表参见表 A.1）。SQLERRM 变量包含与该异常关联的错误消息。这些变量在异常处理器之外是未定义的。

例 36.1. UPDATE/INSERT 与异常

本例使用异常处理来按需执行 UPDATE 或 INSERT：

```
CREATE TABLE db (a INT PRIMARY KEY, b TEXT);

CREATE FUNCTION merge_db(key INT, data TEXT) RETURNS VOID AS
$$
BEGIN
    LOOP
        -- first try to update the key
        UPDATE db SET b = data WHERE a = key;
        IF found THEN
            RETURN;
        END IF;
        -- not there, so try to insert the key
        -- if someone else inserts the same key concurrently,
        -- we could get a unique-key failure
        BEGIN
            INSERT INTO db(a,b) VALUES (key, data);
            RETURN;
        EXCEPTION WHEN unique_violation THEN
            -- do nothing, and loop to try the UPDATE again
        END;
    END LOOP;
END;
$$
LANGUAGE plpgsql;

SELECT merge_db(1, 'david');
SELECT merge_db(1, 'dennis');
```

36.8. 游标

和一次执行整个查询不同，可以建立一个游标来封装该查询，
并且接着一次读取该查询结果的一些行。
这样做的原因之一是在结果中包含大量行时避免内存不足（不过，PL/pgSQL 用户通常不需要担心这些，因为 FOR 循环在内部会自动使用一个游标来避免内存问题）。一种更有趣的用法是返回一个函数已经创建的游标的引用，允许调用者读取行。
这提供了一种有效的方法从函数中返回大型行集。

36.8.1. 声明游标变量

所有在 PL/pgSQL 中对游标的访问都会通过游标变量，它总是特殊的数据类型 refcursor。创建游标变量的一种方法是把它声明为一个类型为 refcursor 的变量。另外一种方法是使用游标声明语法，通常是：

name CURSOR [(*arguments*)] FOR *query*;

（为了兼容 Oracle，FOR 也可以写成 IS。）*arguments* 如果指定，它就是一个由 *name datatype* 对组成的逗号分隔列表，这些名字会在给定查询中被参数值替换。实际替换这些名字的值会在打开游标时提供。

Some examples：

```
DECLARE
    curs1 refcursor;
    curs2 CURSOR FOR SELECT * FROM tenk1;
    curs3 CURSOR (key integer) IS SELECT * FROM tenk1 WHERE unique1 =
 key;
```

所有这三个变量都是 `refcursor` 类型，但是第一个可以用于任何查询，而第二个已经被绑定了一个完全指定的查询，并且最后一个被绑定了一个参数化查询。（游标被打开时，`key` 将被一个整数参数值替换。）变量 `curs1` 被称为未绑定，因为它没有被绑定到任何特定查询。

36.8.2. 打开游标

在能够使用游标检索行之前，必须先将其打开（这等效于 SQL 命令 `DECLARE CURSOR`）。PL/pgSQL 有三种形式的 `OPEN` 命令，其中两种用于未绑定游标变量，另一种用于已绑定游标变量。

36.8.2.1. OPEN FOR *query*

```
OPEN unbound_cursor FOR query;
```

该游标变量会被打开，并被赋予要执行的指定查询。该游标不能已经处于打开状态，
并且它必须已被声明为未绑定游标变量（即，一个简单的 `refcursor` 变量）。该查询必须是 `SELECT`，或其他会返回行的命令（例如 `EXPLAIN`）。该查询会以与 PL/pgSQL 中其他 SQL 命令相同的方式处理：替换 PL/pgSQL 变量名，并缓存查询计划以备后续重用

一个示例：

```
OPEN curs1 FOR SELECT * FROM foo WHERE key = mykey;
```

36.8.2.2. OPEN FOR EXECUTE

```
OPEN unbound_cursor FOR EXECUTE query_string;
```

该游标变量会被打开，并被赋予要执行的指定查询。该游标不能已经处于打开状态，
并且必须已被声明为未绑定游标变量（即，一个简单的 `refcursor` 变量）。该查询以字符串表达式的形式给出，这一点与 `EXECUTE` 命令相同。像往常一样，这提供了灵活性，因此查询可以随每次运行而不同。

一个示例：

```
OPEN curs1 FOR EXECUTE 'SELECT * FROM ' || quote_ident($1);
```

36.8.2.3. 打开已绑定游标

```
OPEN bound_cursor [ ( argument_values ) ];
```

这种形式的 `OPEN` 用于打开一个在声明时就已经绑定查询的游标变量。该游标不能已经处于打开状态。当且仅当该游标被声明为接收参数时，才必须提供实际参数值表达式列表。
这些值会被替换到查询中。已绑定游标的查询计划始终被视为可缓存；在这种情况下没有与 `EXECUTE` 对应的形式。

示例：

```
OPEN curs2;
OPEN curs3(42);
```

36.8.3. 使用游标

一旦一个游标已经被打开，那么就可以用这里描述的语句操作它。

这些操作不必发生在最初打开该游标的同一个函数中。你可以从函数中返回一个 `refcursor` 值，让调用者来操作该游标。（在内部，`refcursor` 值只是一个所谓 portal 的字符串名称，该 portal 包含了该游标的活动查询。这个名称可以被传递、赋给其他 `refcursor` 变量等等，而不会干扰该 portal。）

所有 portal 都会在事务结束时被隐式关闭。因此，`refcursor` 值只能在事务结束之前用于引用一个打开的游标。

36.8.3.1. FETCH

```
FETCH cursor INTO target;
```

`FETCH` 从游标中检索下一行到目标中，目标可以是一个行变量、记录变量或者逗号分隔的简单变量列表，就像 `SELECT INTO` 一样。与 `SELECT INTO` 一样，可以检查特殊变量 `FOUND` 来看是否获得了一行。

示例：

```
FETCH curs1 INTO rowvar;
FETCH curs2 INTO foo, bar, baz;
FETCH LAST FROM curs3 INTO x, y;
FETCH RELATIVE -2 FROM curs4 INTO x;
```

36.8.3.2. CLOSE

```
CLOSE cursor;
```

`CLOSE` 关闭打开游标所基于的 portal。这样就可以在事务结束之前提前释放资源，或者释放该游标变量以便再次打开。

一个示例：

```
CLOSE curs1;
```

36.8.3.3. 返回游标

PL/pgSQL 函数可以向调用者返回游标。这对于返回多行或多列，特别是非常大的结果集时很有用。要做到这一点，函数需要打开游标，
并把游标名返回给调用者（或者直接使用调用者指定或已知的 portal 名称来打开游标）。随后调用者就可以从该游标中提取行。游标既可以由调用者关闭，也会在事务结束时自动关闭。

游标使用的 portal 名称既可以由程序员指定，也可以自动生成。要指定 portal 名称，只需在打开 `refcursor` 变量之前给它赋一个字符串值。`OPEN` 会把该 `refcursor` 变量的字符串值用作底层 portal 的名称。不过，如果 `refcursor` 变量为 null，`OPEN` 就会自动生成一个与任何现有 portal 都不冲突的名称，并把它赋回给 `refcursor` 变量。

> **注意**
>
> 已绑定游标变量会被初始化为表示其名称的字符串值，
> 因此除非程序员在打开游标之前通过赋值覆盖它，否则 portal 名称与游标变量名相同。而

> 未绑定游标变量在初始时默认为空值，因此除非被覆盖，
> 否则它会得到一个自动生成的唯一名称。

下面的示例展示了由调用者提供游标名称的一种方式：

```
CREATE TABLE test (col text);
INSERT INTO test VALUES ('123');

CREATE FUNCTION reffunc(refcursor) RETURNS refcursor AS '
BEGIN
    OPEN $1 FOR SELECT col FROM test;
    RETURN $1;
END;
' LANGUAGE plpgsql;

BEGIN;
SELECT reffunc('funccursor');
FETCH ALL IN funccursor;
COMMIT;
```

下面的示例使用了自动游标名称生成：

```
CREATE FUNCTION reffunc2() RETURNS refcursor AS '
DECLARE
    ref refcursor;
BEGIN
    OPEN ref FOR SELECT col FROM test;
    RETURN ref;
END;
' LANGUAGE plpgsql;
BEGIN;
SELECT reffunc2();

      reffunc2
--------------------
 <unnamed cursor 1>
(1 row)

FETCH ALL IN "<unnamed cursor 1>";
COMMIT;
```

下面的示例展示了从单个函数返回多个游标的一种方式：

```
CREATE FUNCTION myfunc(refcursor, refcursor) RETURNS SETOF refcursor
 AS $$
BEGIN
    OPEN $1 FOR SELECT * FROM table_1;
    RETURN NEXT $1;
    OPEN $2 FOR SELECT * FROM table_2;
    RETURN NEXT $2;
END;
$$ LANGUAGE plpgsql;
```

```
-- need to be in a transaction to use cursors.
BEGIN;

SELECT * FROM myfunc('a', 'b');

FETCH ALL FROM a;
FETCH ALL FROM b;
COMMIT;
```

36.9. 错误和消息

使用 RAISE 语句报告消息并抛出错误。

```
RAISE level 'format' [, expression [, ...]];
```

可能的级别有DEBUG、LOG、INFO、NOTICE、WARNING 和 EXCEPTION。EXCEPTION抛出一个错误（这通常会中止当前事务）；其他级别只生成不同优先级的消息。
某个特定优先级的消息是报告给客户端、写入服务器日志，还是两者都做，由log_min_messages和client_min_messages配置变量控制。更多信息见第 17 章。

在格式字符串内，%会被替换为下一个可选参数的字符串表示。写成%%可以输出一个字面的%。参数可以是简单变量或表达式，但格式必须是简单的字符串字面量。

在这个例子中，v_job_id的值将替换字符串中的%：

```
RAISE NOTICE 'Calling cs_create_job(%)', v_job_id;
```

这个例子将用给定的错误消息中止事务：

```
RAISE EXCEPTION 'Nonexistent ID --> %', user_id;
```

目前，RAISE EXCEPTION不管带什么消息调用，总是生成相同的 SQLSTATE 码P0001。可以用EXCEPTION ... WHEN RAISE_EXCEPTION THEN ...捕获这个异常，但没有办法区分一个RAISE和另一个。

36.10. 触发器函数

PL/pgSQL 可以被用来定义触发器函数。触发器函数用 CREATE FUNCTION 命令创建，它被声明为一个没有参数并且返回类型为 trigger 的函数。注意，即便该函数准备接收一些在 CREATE TRIGGER 中指定的参数 – 这类参数通过 TG_ARGV 传递（如下所述），也必须把它声明为没有参数。

当一个 PL/pgSQL 函数作为触发器被调用时，会在顶层块中自动创建一些特殊变量。它们是：

NEW

数据类型为 RECORD；该变量保存行级触发器中用于 INSERT/UPDATE 操作的新数据行。在语句级触发器中该变量为 NULL（因此该变量在语句级触发器中为 NULL）。

OLD

数据类型为 RECORD；该变量保存行级触发器中用于 UPDATE/DELETE 操作的旧数据行。在语句级触发器中该变量为 NULL。

TG_NAME

数据类型为 name；该变量包含实际触发的触发器的名称。

TG_WHEN

数据类型为 text；根据触发器定义，其值为字符串 BEFORE 或 AFTER。

TG_LEVEL

数据类型为 text；根据触发器定义，其值为字符串 ROW 或 STATEMENT。

TG_OP

数据类型为 text；表示触发器对应操作的字符串：INSERT、UPDATE 或 DELETE，表示触发器是由哪个操作触发的。

TG_RELID

数据类型为 oid；导致触发器调用的表的对象 ID。

TG_RELNAME

数据类型为 name；导致触发器调用的表名。

TG_NARGS

数据类型为 integer；CREATE TRIGGER 语句中给触发器函数的参数个数。

TG_ARGV[]

数据类型为 text 的数组；CREATE TRIGGER 语句中的参数。索引从 0 开始计。无效索引（小于 0 或大于等于 tg_nargs）会产生空值。

一个触发器函数必须返回 NULL 或者是一个与触发器为之引发的表结构完全相同的记录/行值。

BEFORE 行级触发器可以返回 null，以通知触发器管理器跳过该行后续的操作（也就是说，不再触发后续触发器，并且不会对该行执行 INSERT/UPDATE/DELETE）。如果返回非 null 值，则操作会继续进行，并使用该行值。返回一个不同于原始 NEW 的行值会改变即将插入或更新的行（但对 DELETE 没有直接影响）。若要修改将被存储的行，可以直接替换 NEW 中的单个值并返回修改后的 NEW，或者构造一个完整的新记录/行来返回。

一个 AFTER 行级触发器或一个 BEFORE 或 AFTER 语句级触发器的返回值总是会被忽略，因此也可以返回 null。不过，任何这些类型的触发器可能仍会通过抛出一个错误来中止整个操作。

例 36.2展示了 PL/pgSQL 中一个触发器函数的示例。

例 36.2. 一个 PL/pgSQL 触发器函数

这个示例触发器保证：任何时候一个行在表中被插入或更新时，
当前用户名和时间也会被标记在该行中。
并且它会检查给出了一个雇员的姓名以及薪水是一个正值。

```
CREATE TABLE emp (
    empname text,
    salary integer,
    last_date timestamp,
```

```
    last_user text
);

CREATE FUNCTION emp_stamp() RETURNS trigger AS $emp_stamp$
    BEGIN
        -- Check that empname and salary are given
        IF NEW.empname IS NULL THEN
            RAISE EXCEPTION 'empname cannot be null';
        END IF;
        IF NEW.salary IS NULL THEN
            RAISE EXCEPTION '% cannot have null salary', NEW.empname;
        END IF;

        -- Who works for us when she must pay for it?
        IF NEW.salary < 0 THEN
            RAISE EXCEPTION '% cannot have a negative salary', NEW.
empname;
        END IF;

        -- Remember who changed the payroll when
        NEW.last_date := current_timestamp;
        NEW.last_user := current_user;
        RETURN NEW;
    END;
$emp_stamp$ LANGUAGE plpgsql;

CREATE TRIGGER emp_stamp BEFORE INSERT OR UPDATE ON emp
    FOR EACH ROW EXECUTE PROCEDURE emp_stamp();
```

另一种记录对表的改变的方法涉及到创建一个新表来为每一个发生的插入、更新或删除保持一行。这种方法可以被认为是对一个表的改变的审计。例 36.3展示了 PL/pgSQL 中一个审计触发器函数的示例。

例 36.3. 一个用于审计的 PL/pgSQL 触发器函数

这个示例触发器保证 `emp` 表上的任何插入、更新或删除一行的动作都被记录（即审计）在 `emp_audit` 表中。当前时间和用户名会被记录到行中，还有在其上执行的操作类型。

```
CREATE TABLE emp (
    empname           text NOT NULL,
    salary            integer
);

CREATE TABLE emp_audit(
    operation         char(1)   NOT NULL,
    stamp             timestamp NOT NULL,
    userid            text      NOT NULL,
    empname           text      NOT NULL,
    salary integer
);

CREATE OR REPLACE FUNCTION process_emp_audit() RETURNS TRIGGER AS
 $emp_audit$
    BEGIN
```

```
        --
        -- Create a row in emp_audit to reflect the operation
 performed on emp,
        -- make use of the special variable TG_OP to work out the
 operation.
        --
        IF (TG_OP = 'DELETE') THEN
            INSERT INTO emp_audit SELECT 'D', now(), user, OLD.*;
            RETURN OLD;
        ELSIF (TG_OP = 'UPDATE') THEN
            INSERT INTO emp_audit SELECT 'U', now(), user, NEW.*;
            RETURN NEW;
        ELSIF (TG_OP = 'INSERT') THEN
            INSERT INTO emp_audit SELECT 'I', now(), user, NEW.*;
            RETURN NEW;
        END IF;
        RETURN NULL; -- result is ignored since this is an AFTER
 trigger
    END;
$emp_audit$ LANGUAGE plpgsql;

CREATE TRIGGER emp_audit
AFTER INSERT OR UPDATE OR DELETE ON emp
    FOR EACH ROW EXECUTE PROCEDURE process_emp_audit();
```

触发器的一种用法是维护另一个表的汇总表。所得的汇总可以替代原表用于某些查询 – 运行时间往往大幅缩短。这种技术常用于数据仓库，其中被测量或观察数据（称为事实表）的表可能极其庞大。例 36.4展示了 PL/pgSQL 中一个触发器函数的示例，它为数据仓库中的一个事实表维护汇总表。

例 36.4. 一个维护汇总表的 PL/pgSQL 触发器函数

这里详述的模式部分基于 Ralph Kimball 的 The Data Warehouse Toolkit 中的 Grocery Store 示例。

```
--
-- Main tables - time dimension and sales fact.
--
CREATE TABLE time_dimension (
    time_key                    integer NOT NULL,
    day_of_week                 integer NOT NULL,
    day_of_month                integer NOT NULL,
    month                       integer NOT NULL,
    quarter                     integer NOT NULL,
    year                        integer NOT NULL
);
CREATE UNIQUE INDEX time_dimension_key ON time_dimension(time_key);

CREATE TABLE sales_fact (
    time_key                    integer NOT NULL,
    product_key                 integer NOT NULL,
    store_key                   integer NOT NULL,
    amount_sold                 numeric(12,2) NOT NULL,
    units_sold                  integer NOT NULL,
```

```
    amount_cost                 numeric(12,2) NOT NULL
);
CREATE INDEX sales_fact_time ON sales_fact(time_key);

--
-- Summary table - sales by time.
--
CREATE TABLE sales_summary_bytime (
    time_key                    integer NOT NULL,
    amount_sold                 numeric(15,2) NOT NULL,
    units_sold                  numeric(12) NOT NULL,
    amount_cost                 numeric(15,2) NOT NULL
);
CREATE UNIQUE INDEX sales_summary_bytime_key ON sales_summary_bytime(
time_key);

--
-- Function and trigger to amend summarized column(s) on UPDATE,
 INSERT, DELETE.
--
CREATE OR REPLACE FUNCTION maint_sales_summary_bytime() RETURNS
 TRIGGER
AS $maint_sales_summary_bytime$
    DECLARE
        delta_time_key          integer;
        delta_amount_sold       numeric(15,2);
        delta_units_sold        numeric(12);
        delta_amount_cost       numeric(15,2);
    BEGIN

        -- Work out the increment/decrement amount(s).
        IF (TG_OP = 'DELETE') THEN

            delta_time_key = OLD.time_key;
            delta_amount_sold = -1 * OLD.amount_sold;
            delta_units_sold = -1 * OLD.units_sold;
            delta_amount_cost = -1 * OLD.amount_cost;

        ELSIF (TG_OP = 'UPDATE') THEN

            -- forbid updates that change the time_key -
            -- (probably not too onerous, as DELETE + INSERT is how
 most
            -- changes will be made).
            IF ( OLD.time_key != NEW.time_key) THEN
                RAISE EXCEPTION 'Update of time_key : % -> % not
 allowed',
                                                        OLD.time_key,
 NEW.time_key;
            END IF;

            delta_time_key = OLD.time_key;
            delta_amount_sold = NEW.amount_sold - OLD.amount_sold;
            delta_units_sold = NEW.units_sold - OLD.units_sold;
```

```
            delta_amount_cost = NEW.amount_cost - OLD.amount_cost;

        ELSIF (TG_OP = 'INSERT') THEN

            delta_time_key = NEW.time_key;
            delta_amount_sold = NEW.amount_sold;
            delta_units_sold = NEW.units_sold;
            delta_amount_cost = NEW.amount_cost;

        END IF;


        -- Insert or update the summary row with the new values.
        <<insert_update>>
        LOOP
            UPDATE sales_summary_bytime
                SET amount_sold = amount_sold + delta_amount_sold,
                    units_sold = units_sold + delta_units_sold,
                    amount_cost = amount_cost + delta_amount_cost
                WHERE time_key = delta_time_key;

            EXIT insert_update WHEN found;

            BEGIN
                INSERT INTO sales_summary_bytime (
                            time_key,
                            amount_sold,
                            units_sold,
                            amount_cost)
                    VALUES (
                            delta_time_key,
                            delta_amount_sold,
                            delta_units_sold,
                            delta_amount_cost
                           );

                EXIT insert_update;

            EXCEPTION
                WHEN UNIQUE_VIOLATION THEN
                    -- do nothing
            END;
        END LOOP insert_update;

        RETURN NULL;

    END;
$maint_sales_summary_bytime$ LANGUAGE plpgsql;

CREATE TRIGGER maint_sales_summary_bytime
AFTER INSERT OR UPDATE OR DELETE ON sales_fact
    FOR EACH ROW EXECUTE PROCEDURE maint_sales_summary_bytime();

INSERT INTO sales_fact VALUES(1,1,1,10,3,15);
```

```
INSERT INTO sales_fact VALUES(1,2,1,20,5,35);
INSERT INTO sales_fact VALUES(2,2,1,40,15,135);
INSERT INTO sales_fact VALUES(2,3,1,10,1,13);
SELECT * FROM sales_summary_bytime;
DELETE FROM sales_fact WHERE product_key = 1;
SELECT * FROM sales_summary_bytime;
UPDATE sales_fact SET units_sold = units_sold * 2;
SELECT * FROM sales_summary_bytime;
```

36.11. 从Oracle PL/SQL 移植

这一节解释了 PostgreSQL 的 PL/pgSQL 语言和 Oracle 的 PL/SQL 语言之间的差别，用以帮助那些从 Oracle® 向 PostgreSQL 移植应用的人。

PL/pgSQL 在许多方面都与 PL/SQL 类似。它是一种具有块结构的命令式语言，所有变量都必须声明。赋值、循环和条件语句也都很相似。在从 PL/SQL 移植到 PL/pgSQL 时，应当记住以下主要差异：

- 在PostgreSQL中，参数没有默认值。
- 不能使用与函数中引用的列同名的参数名。Oracle 允许这样做，只要用 `function_name.paramater_name` 限定参数名即可。
- 在PostgreSQL中可以重载函数名。这常被用来弥补缺少默认参数的问题。
- 在 PL/pgSQL 中不需要使用游标，直接把查询放到 `FOR` 语句中即可。（见例 36.6。）
- 在 PostgreSQL 中，函数体必须写成字符串字面量。因此你需要使用美元符引用或者转义函数体中的单引号（见第 36.2.1 节）。
- 应该用模式把函数组织成不同的分组，而不是用包。
- 因为没有包，所以也没有包级变量。这一点有时会带来不便。
  你可以改为在临时表中保存会话级状态。

36.11.1. 移植示例

例 36.5展示了如何把一个简单的函数从 PL/SQL 移植到 PL/pgSQL。

例 36.5. 把一个简单的函数从 PL/SQL 移植到 PL/pgSQL

下面是一个 Oracle PL/SQL 函数：

```
CREATE OR REPLACE FUNCTION cs_fmt_browser_version(v_name varchar,
                                                  v_version varchar)
RETURN varchar IS
BEGIN
    IF v_version IS NULL THEN
        RETURN v_name;
    END IF;
    RETURN v_name || '/' || v_version;
END;
/
show errors;
```

让我们过一遍这个函数并且看看与 PL/pgSQL 相比有什么样的不同：

- 在函数原型中（不是函数体中）的 `RETURN` 关键字在 PostgreSQL 中变成了 `RETURNS`。还有，`IS` 变成了 `AS`，并且你还需要增加一个 `LANGUAGE` 子句，因为 PL/pgSQL 并非唯一可用的函数语言。
- 在 PostgreSQL 中，函数体被认为是一个字符串字面量，所以你需要使用引号或者美元引用定界符包围它。这代替了 Oracle 方法中的用于终止的 `/`。
- 在 PostgreSQL 中没有 `show errors` 命令，并且也不需要这个命令，因为错误是自动报告的。

这个函数被移植到 PostgreSQL 后看起来会是这样：

```
CREATE OR REPLACE FUNCTION cs_fmt_browser_version(v_name varchar,
                                                  v_version varchar)
RETURNS varchar AS $$
BEGIN
    IF v_version IS NULL THEN
        RETURN v_name;
    END IF;
    RETURN v_name || '/' || v_version;
END;
$$ LANGUAGE plpgsql;
```

例 36.6展示了如何移植一个会创建另一个函数的函数，以及如何处理随之而来的引号问题。

例 36.6. 从 PL/SQL 移植一个创建另一个函数的函数到 PL/pgSQL

下面的过程从 `SELECT` 语句读取行，并将结果写入 `IF` 语句，从而构造一个大型函数，以提高效率。请特别注意游标和 `FOR` 循环方面的差异。

这是 Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_update_referrer_type_proc IS
    CURSOR referrer_keys IS
        SELECT * FROM cs_referrer_keys
        ORDER BY try_order;
    func_cmd VARCHAR(4000);
BEGIN
    func_cmd := 'CREATE OR REPLACE FUNCTION cs_find_referrer_type(v_
host IN VARCHAR,
                 v_domain IN VARCHAR, v_url IN VARCHAR) RETURN VARCHAR
 IS BEGIN';

    FOR referrer_key IN referrer_keys LOOP
        func_cmd := func_cmd ||
          ' IF v_' || referrer_key.kind
          || ' LIKE ''' || referrer_key.key_string
          || ''' THEN RETURN ''' || referrer_key.referrer_type
          || '''; END IF;';
    END LOOP;

    func_cmd := func_cmd || ' RETURN NULL; END;';
```

```
    EXECUTE IMMEDIATE func_cmd;
END;
/
show errors;
```

这个函数在 PostgreSQL 中的最终版本如下：

```
CREATE OR REPLACE FUNCTION cs_update_referrer_type_proc() RETURNS void
 AS $func$
DECLARE
    referrer_keys CURSOR IS
        SELECT * FROM cs_referrer_keys
        ORDER BY try_order;
    func_body text;
    func_cmd text;
BEGIN
    func_body := 'BEGIN';

    FOR referrer_key IN referrer_keys LOOP
        func_body := func_body ||
          ' IF v_' || referrer_key.kind
          || ' LIKE ' || quote_literal(referrer_key.key_string)
          || ' THEN RETURN ' || quote_literal(referrer_key.referrer_
type)
          || '; END IF;' ;
    END LOOP;

    func_body := func_body || ' RETURN NULL; END;';

    func_cmd :=
      'CREATE OR REPLACE FUNCTION cs_find_referrer_type(v_host
 varchar,
                                                          v_domain
 varchar,
                                                          v_url varchar)
        RETURNS varchar AS '
      || quote_literal(func_body)
      || ' LANGUAGE plpgsql;' ;

    EXECUTE func_cmd;
END;
$func$ LANGUAGE plpgsql;
```

注意，这里单独构造函数体，然后将其传给 `quote_literal`，使其中的每个引号都变成两个。这种技术是必需的，
因为不能安全地使用美元引用来定义新函数：我们无法确定会插入什么字符串，其来源是 `referrer_key.key_string` 字段。（这里假定 `referrer_key.kind` 可信，其值总是 `host`、`domain` 或 `url`，但是 `referrer_key.key_string` 可能是任何内容，尤其可能包含美元符号。）这个函数实际上改进了 Oracle 原版：当 `referrer_key.key_string` 或 `referrer_key.referrer_type` 中包含引号时，它也不会生成有问题的代码。

例 36.7展示了如何移植一个带有 `OUT` 参数和字符串处理的函数。PostgreSQL 没有内置的 `instr` 函数，但是你可以用其它函数的组合来创建一个。在第 36.11.3 节中有一个 `instr` 的 PL/pgSQL 实现，你可以用它让你的移植变得更容易。

例 36.7. 从 PL/SQL 移植一个带有字符串操作以及 OUT 参数的过程到 PL/pgSQL

下面的 Oracle PL/SQL 过程被用来解析一个 URL 并且返回一些元素（主机、路径和查询）。

这是 Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_parse_url(
    v_url IN VARCHAR,
    v_host OUT VARCHAR,  -- This will be passed back
    v_path OUT VARCHAR,  -- This one too
    v_query OUT VARCHAR) -- And this one
IS
    a_pos1 INTEGER;
    a_pos2 INTEGER;
BEGIN
    v_host := NULL;
    v_path := NULL;
    v_query := NULL;
    a_pos1 := instr(v_url, '//');

    IF a_pos1 = 0 THEN
        RETURN;
    END IF;
    a_pos2 := instr(v_url, '/', a_pos1 + 2);
    IF a_pos2 = 0 THEN
        v_host := substr(v_url, a_pos1 + 2);
        v_path := '/';
        RETURN;
    END IF;

    v_host := substr(v_url, a_pos1 + 2, a_pos2 - a_pos1 - 2);
    a_pos1 := instr(v_url, '?', a_pos2 + 1);

    IF a_pos1 = 0 THEN
        v_path := substr(v_url, a_pos2);
        RETURN;
    END IF;

    v_path := substr(v_url, a_pos2, a_pos1 - a_pos2);
    v_query := substr(v_url, a_pos1 + 1);
END;
/
show errors;
```

下面是一个可能的 PL/pgSQL 版本：

```
CREATE OR REPLACE FUNCTION cs_parse_url(
    v_url IN VARCHAR,
    v_host OUT VARCHAR,  -- This will be passed back
    v_path OUT VARCHAR,  -- This one too
    v_query OUT VARCHAR) -- And this one
AS $$
DECLARE
    a_pos1 INTEGER;
    a_pos2 INTEGER;
```

```
BEGIN
    v_host := NULL;
    v_path := NULL;
    v_query := NULL;
    a_pos1 := instr(v_url, '//');

    IF a_pos1 = 0 THEN
        RETURN;
    END IF;
    a_pos2 := instr(v_url, '/', a_pos1 + 2);
    IF a_pos2 = 0 THEN
        v_host := substr(v_url, a_pos1 + 2);
        v_path := '/';
        RETURN;
    END IF;

    v_host := substr(v_url, a_pos1 + 2, a_pos2 - a_pos1 - 2);
    a_pos1 := instr(v_url, '?', a_pos2 + 1);

    IF a_pos1 = 0 THEN
        v_path := substr(v_url, a_pos2);
        RETURN;
    END IF;

    v_path := substr(v_url, a_pos2, a_pos1 - a_pos2);
    v_query := substr(v_url, a_pos1 + 1);
END;
$$ LANGUAGE plpgsql;
```

这个函数可以这样使用：

```
SELECT * FROM cs_parse_url('http://foobar.com/query.cgi?baz');
```

例 36.8展示了如何移植一个使用了大量 Oracle 专有特性的过程。

例 36.8. 从 PL/SQL 移植一个过程到 PL/pgSQL

Oracle 版本：

```
CREATE OR REPLACE PROCEDURE cs_create_job(v_job_id IN INTEGER) IS
    a_running_job_count INTEGER;
    PRAGMA AUTONOMOUS_TRANSACTION;1
BEGIN
    LOCK TABLE cs_jobs IN EXCLUSIVE MODE;2

    SELECT count(*) INTO a_running_job_count FROM cs_jobs WHERE end_
stamp IS NULL;

    IF a_running_job_count > 0 THEN
        COMMIT; -- free lock3
        raise_application_error(-20000,
                 'Unable to create a new job: a job is currently
 running.');
    END IF;
```

```
    DELETE FROM cs_active_job;
    INSERT INTO cs_active_job(job_id) VALUES (v_job_id);

    BEGIN
        INSERT INTO cs_jobs (job_id, start_stamp) VALUES (v_job_id,
 sysdate);
    EXCEPTION
        WHEN dup_val_on_index THEN NULL; -- don't worry if it already
 exists
    END;
    COMMIT;
END;
/
show errors
```

这样的过程可以很容易地转换为 PostgreSQL 中返回 `void` 的函数。这个过程尤其值得关注，因为它能说明以下几点：

1. PostgreSQL 中没有 `PRAGMA` 语句。
2. 如果在 PL/pgSQL 中执行 `LOCK TABLE`，锁要等到调用事务结束时才会释放。
3. 不能在 PL/pgSQL 函数中发出 `COMMIT`。函数运行在某个外层事务之中，因此 `COMMIT` 意味着终止函数的执行。不过，在这个特定情形中本来也不需要提交，因为抛出错误时，`LOCK TABLE` 获得的锁就会被释放。

我们可以像这样把这个过程移植到 PL/pgSQL：

```
CREATE OR REPLACE FUNCTION cs_create_job(v_job_id integer) RETURNS
 void AS $$
DECLARE
    a_running_job_count integer;
BEGIN
    LOCK TABLE cs_jobs IN EXCLUSIVE MODE;

    SELECT count(*) INTO a_running_job_count FROM cs_jobs WHERE end_
stamp IS NULL;

    IF a_running_job_count > 0 THEN
        RAISE EXCEPTION 'Unable to create a new job: a job is
 currently running';❶
    END IF;

    DELETE FROM cs_active_job;
    INSERT INTO cs_active_job(job_id) VALUES (v_job_id);

    BEGIN
        INSERT INTO cs_jobs (job_id, start_stamp) VALUES (v_job_id,
 now());
    EXCEPTION
        WHEN unique_violation THEN ❷
            -- don't worry if it already exists
    END;
END;
$$ LANGUAGE plpgsql;
```

❶ `RAISE` 的语法与 Oracle 的类似语句差别很大。

2 PL/pgSQL 支持的异常名与 Oracle 的不同。内置异常名的集合要大得多（见附录 A）。目前没有办法声明用户自定义的异常名，但你可以在抛出时使用自己选择的 SQLSTATE 值。

这个过程与 Oracle 等价版本在功能上的主要区别是，对 `cs_jobs` 表的排他锁会一直保持到调用事务完成。此外，如果调用者随后中止（例如由于错误），该过程的效果也会被回滚。

36.11.2. 其他要关注的事项

本节说明在把 Oracle PL/SQL 函数移植到 PostgreSQL 时需要注意的其他一些事项。

36.11.2.1. 异常后隐式回滚

在 PL/pgSQL 中，当异常被 `EXCEPTION` 子句捕获后，自该块的 `BEGIN` 以来所做的所有数据库更改都会被自动回滚。也就是说，这种行为等效于你在 Oracle 中使用下面的代码所得到的效果：

```
BEGIN
    SAVEPOINT s1;
    ... code here ...
EXCEPTION
    WHEN ... THEN
        ROLLBACK TO s1;
        ... code here ...
    WHEN ... THEN
        ROLLBACK TO s1;
        ... code here ...
END;
```

如果你正在移植一个以这种方式使用 `SAVEPOINT` 和 `ROLLBACK TO` 的 Oracle 过程，工作会比较简单：只要省略 `SAVEPOINT` 和 `ROLLBACK TO` 即可。如果你的 Oracle 过程以不同方式使用 `SAVEPOINT` 和 `ROLLBACK TO`，那就需要认真思考了。

36.11.2.2. EXECUTE

PL/pgSQL 中的 `EXECUTE` 与 PL/SQL 中的版本工作方式相似，但必须记得按照第 36.6.5 节中的说明使用 `quote_literal` 和 `quote_ident`。像 `EXECUTE 'SELECT * FROM $1';` 这样的写法如果不借助这些函数，将无法可靠地工作。

36.11.2.3. 优化 PL/pgSQL 函数

PostgreSQL 提供了两种函数创建修饰符来优化执行：“volatility”（对于给定的相同参数，函数是否总是返回相同的结果）以及“strictness”（如果任何参数为空值，函数是否返回空值）。详见CREATE FUNCTION参考页。

使用这些优化属性时，你的 `CREATE FUNCTION` 语句可能看起来像这样：

```
CREATE FUNCTION foo(...) RETURNS integer AS $$
...
$$ LANGUAGE plpgsql STRICT IMMUTABLE;
```

36.11.3. 附录

这一节包含了一组 Oracle 兼容的 `instr` 函数代码，你可以用它来简化你的移植工作。

```
--
```

```
-- instr functions that mimic Oracle's counterpart
-- Syntax: instr(string1, string2, [n], [m]) where [] denotes optional
 parameters.
--
-- Searches string1 beginning at the nth character for the mth
 occurrence
-- of string2.  If n is negative, search backwards.  If m is not
 passed,
-- assume 1 (search starts at first character).
--

CREATE FUNCTION instr(varchar, varchar) RETURNS integer AS $$
DECLARE
    pos integer;
BEGIN
    pos:= instr($1, $2, 1);
    RETURN pos;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;


CREATE FUNCTION instr(string varchar, string_to_search varchar, beg_
index integer)
RETURNS integer AS $$
DECLARE
    pos integer NOT NULL DEFAULT 0;
    temp_str varchar;
    beg integer;
    length integer;
    ss_length integer;
BEGIN
    IF beg_index > 0 THEN
        temp_str := substring(string FROM beg_index);
        pos := position(string_to_search IN temp_str);

        IF pos = 0 THEN
            RETURN 0;
        ELSE
            RETURN pos + beg_index - 1;
        END IF;
    ELSE
        ss_length := char_length(string_to_search);
        length := char_length(string);
        beg := length + beg_index - ss_length + 2;

        WHILE beg > 0 LOOP
            temp_str := substring(string FROM beg FOR ss_length);
            pos := position(string_to_search IN temp_str);

            IF pos > 0 THEN
                RETURN beg;
            END IF;

            beg := beg - 1;
```

```
        END LOOP;

        RETURN 0;
    END IF;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;


CREATE FUNCTION instr(string varchar, string_to_search varchar,
                      beg_index integer, occur_index integer)
RETURNS integer AS $$
DECLARE
    pos integer NOT NULL DEFAULT 0;
    occur_number integer NOT NULL DEFAULT 0;
    temp_str varchar;
    beg integer;
    i integer;
    length integer;
    ss_length integer;
BEGIN
    IF beg_index > 0 THEN
        beg := beg_index;
        temp_str := substring(string FROM beg_index);

        FOR i IN 1..occur_index LOOP
            pos := position(string_to_search IN temp_str);

            IF i = 1 THEN
                beg := beg + pos - 1;
            ELSE
                beg := beg + pos;
            END IF;

            temp_str := substring(string FROM beg + 1);
        END LOOP;

        IF pos = 0 THEN
            RETURN 0;
        ELSE
            RETURN beg;
        END IF;
    ELSE
        ss_length := char_length(string_to_search);
        length := char_length(string);
        beg := length + beg_index - ss_length + 2;

        WHILE beg > 0 LOOP
            temp_str := substring(string FROM beg FOR ss_length);
            pos := position(string_to_search IN temp_str);

            IF pos > 0 THEN
                occur_number := occur_number + 1;

                IF occur_number = occur_index THEN
```

```
                    RETURN beg;
                END IF;
            END IF;

            beg := beg - 1;
        END LOOP;

        RETURN 0;
    END IF;
END;
$$ LANGUAGE plpgsql STRICT IMMUTABLE;
```

第 37 章 PL/Tcl - Tcl 过程语言

PL/Tcl 是 PostgreSQL 数据库系统的一种可加载过程语言，可以使用 Tcl 语言[1]编写函数和触发器函数。

37.1. 概述

PL/Tcl提供了函数编写者在C语言中所具备的大部分能力，但会施加少量限制，并额外提供Tcl可用的强大字符串处理库。

其中一个颇具吸引力的良性限制是，所有内容都在 Tcl 解释器的安全上下文中执行。除了安全 Tcl 受限的命令集之外，只提供了少数通过 SPI 访问数据库以及通过`elog()`发出消息的命令。与 C 函数不同，PL/Tcl 不提供访问数据库服务器内部机制的方法，也不能在PostgreSQL服务器进程的权限下获得操作系统级访问。因此，可以信任非特权数据库用户使用这种语言；
它不会授予他们无限制的权限。

另一项值得注意的实现限制是，Tcl 函数不能用来为新数据类型创建输入/输出函数。

有时需要编写不受安全 Tcl 限制的 Tcl 函数。例如，可能希望某个 Tcl 函数能够发送电子邮件。为处理这类情况，PL/Tcl 提供了一个名为`PL/TclU`的变体（其中 U 表示不受信任的 Tcl）。除了使用完整的 Tcl 解释器之外，它与前者是完全相同的语言。如果使用PL/TclU，就必须将其安装为一种不受信任的过程语言，这样就只有数据库超级用户才能在其中创建函数。
PL/TclU函数的编写者必须确保该函数不会被用来做任何不希望发生的事情，
因为它能够执行任何以数据库管理员身份登录的用户所能做的事情。

如果在安装过程的配置步骤中指定了 Tcl 支持，则 PL/Tcl 和 PL/TclU 调用处理器的共享对象代码会自动构建并安装到 PostgreSQL 的库目录中。要在某个特定数据库中安装 PL/Tcl 和/或 PL/TclU，请使用 `createlang` 程序，例如 `createlang pltcl *dbname*` 或 `createlang pltclu *dbname*`。

37.2. PL/Tcl 函数和参数

要用PL/Tcl语言创建函数，可使用标准的CREATE FUNCTION语法：

```
CREATE FUNCTION funcname (argument-types) RETURNS return-type AS $$
    # PL/Tcl 函数体
$$ LANGUAGE pltcl;
```

PL/TclU的写法相同，只是语言必须指定为`pltclu`。

函数体就是一段 Tcl 脚本。调用函数时，参数值以变量 `$1` ... `$n` 的形式传入 Tcl 脚本。Tcl 代码按通常方式使用 `return` 语句返回结果。

例如，一个返回两个整数值中较大值的函数可以定义为：

```
CREATE FUNCTION tcl_max(integer, integer) RETURNS integer AS $$
    if {$1 > $2} {return $1}
    return $2
$$ LANGUAGE pltcl STRICT;
```

注意`STRICT`子句，它让我们不必考虑空值输入：如果传入的是空值，函数根本不会被调用，
而是会自动返回空值结果。

[1] http://www.tcl.tk/

在非严格函数中，如果某个参数的实际值为空值，对应的$*n*变量会被设置为空串。
要检测某个特定参数是否为空值，可使用函数argisnull。例如，假设我们希望在tcl_max的两个参数中一个为空值、一个非空值时返回非空值参数，而不是返回空值：

```
CREATE FUNCTION tcl_max(integer, integer) RETURNS integer AS $$
    if {[argisnull 1]} {
        if {[argisnull 2]} { return_null }
        return $2
    }
    if {[argisnull 2]} { return $1 }
    if {$1 > $2} {return $1}
    return $2
$$ LANGUAGE pltcl;
```

如上所示，要从 PL/Tcl 函数返回空值，请执行return_null。无论函数是否为严格函数，都可以这样做。

复合类型参数会作为 Tcl 数组传递给函数。数组元素名就是该复合类型的属性名。如果传入行中的某个属性为空值，它就不会出现在数组中。下面是一个示例：

```
CREATE TABLE employee (
    name text,
    salary integer,
    age integer
);

CREATE FUNCTION overpaid(employee) RETURNS boolean AS $$
    if {200000.0 < $1(salary)} {
        return "t"
    }
    if {$1(age) < 30 && 100000.0 < $1(salary)} {
        return "t"
    }
    return "f"
$$ LANGUAGE pltcl;
```

目前尚不支持返回复合类型的结果值，也不支持返回集合。

PL/Tcl目前对域类型还没有完整支持：它把域当作底层标量类型同样对待。
这意味着与该域关联的约束不会被强制执行。对于函数参数来说这不是问题，但如果把PL/Tcl 函数声明为返回域类型，就会带来隐患。

37.3. PL/Tcl 中的数据值

提供给 PL/Tcl 函数代码的参数值，只是将输入参数转换成文本形式（就像用SELECT语句将它们显示出来一样）。反过来，return命令会接受任何字符串，
只要它是该函数声明返回类型的可接受输入格式。因此，在 PL/Tcl 函数内部，所有值都只是文本字符串。

37.4. PL/Tcl 中的全局数据

有时需要在一次函数调用结束后保留某些全局数据，供下一次调用继续使用，
或者在不同函数之间共享全局数据。在 PL/Tcl 中这很容易做到，但必须理解其中的一些限制。

出于安全原因，PL/Tcl 会针对每个 SQL 角色，在该角色各自独立的 Tcl 解释器中执行其调用的函数。这样可以防止一个用户无意或恶意地干扰另一用户的 PL/Tcl 函数行为。每个这样的解释器都为任何“global” Tcl 变量维护各自的值。因此，当且仅当两个 PL/Tcl 函数由同一个 SQL 角色执行时，它们才会共享同一组全局变量。在某些应用中，一个会话会在多个 SQL 角色下执行代码（例如通过`SECURITY DEFINER`函数、使用`SET ROLE`等），这时你可能需要显式采取一些步骤，确保 PL/Tcl 函数能够共享数据。为此，应确保需要互相通信的函数由同一用户拥有，并将它们标记为`SECURITY DEFINER`。当然，你必须小心，确保这类函数不会被用来执行任何非预期操作。

一个会话中使用的所有 PL/TclU 函数都在同一个 Tcl 解释器中执行，而这个解释器当然不同于 PL/Tcl 函数所使用的解释器。因此，PL/TclU 函数之间会自动共享全局数据。这不被视为安全风险，因为所有 PL/TclU 函数都在相同的信任级别上执行，也就是数据库超级用户的级别。

为帮助防止 PL/Tcl 函数无意间彼此干扰，每个函数都可通过`upvar`命令访问一个全局数组。该变量的全局名称是函数的内部名称，局部名称是`GD`。建议将`GD`用于保存函数的持久私有数据。
只有当你明确希望某些值在多个函数之间共享时，才应使用常规的 Tcl 全局变量。（注意`GD`数组只在某个特定解释器内部是全局的，因此它们不会绕过上文提到的安全限制。）

下面`spi_execp`的示例展示了`GD`的用法。

37.5. 从 PL/Tcl 访问数据库

在 PL/Tcl 函数体中可以使用下列命令来访问数据库：

`spi_exec ?-count` *`n`*`? ?-array` *`name`*`?` *`command`* `?`*`loop-body`*`?`

执行以字符串形式给出的 SQL 命令。命令出错时会引发错误。否则，`spi_exec` 的返回值是该命令处理的行数（选出、插入、更新或删除的行），如果命令是工具语句则返回零。
此外，如果命令是 `SELECT` 语句，则所选列的值会按下文所述放入 Tcl 变量中。

可选的 `-count` 值告诉 `spi_exec` 此命令最多处理多少行，其效果类似于将查询设为游标后执行 `FETCH` *`n`*。

如果命令是 `SELECT` 语句，则结果列的值会放入以列名命名的 Tcl 变量中。如果给定了 `-array` 选项，则列值会存储在指定的关联数组元素中，列名用作数组索引。

如果命令是 `SELECT` 语句且未给出 *`loop-body`* 脚本，则只会把结果的第一行存入 Tcl 变量中；其余行如果存在，会被忽略。如果查询没有返回任何行，则不会进行任何存储。（这种情况可以通过检查 `spi_exec` 的结果来发现。）例如：

```
spi_exec "SELECT count(*) AS cnt FROM pg_proc"
```

会把 Tcl 变量 `$cnt` 设置为 `pg_proc` 系统目录中的行数。

如果给出了可选的 *`loop-body`* 参数，它就是一段 Tcl 脚本，对查询结果中的每一行执行一次。（如果给定的命令不是 `SELECT`，则忽略 *`loop-body`*。）在每次迭代之前，当前行各列的值会被存入 Tcl 变量。例如：

```
spi_exec -array C "SELECT * FROM pg_class" {
    elog DEBUG "have table $C(relname)"
}
```

会为 `pg_class` 的每一行打印一条日志消息。这一特性的工作方式类似于其他 Tcl 循环构造；特别是 `continue` 和 `break` 在循环体中按通常方式工作。

如果查询结果的某一列为空值，则其目标变量会被“unset”（取消设置），而不是被设置。

spi_prepare*querytypelist*

准备并保存一个查询计划以供后续执行。保存的计划会在当前会话的整个生命周期内保留。

查询可以使用参数，也就是在实际执行该计划时要提供的值的占位符。在查询字符串中，用符号 $1 ... $*n* 引用参数。如果查询使用了参数，则必须以 Tcl 列表的形式给出各参数类型的名称。（如果未使用参数，则为 *typelist* 写一个空列表。）目前，参数类型必须用系统表 pg_type中所示的内部类型名来标识；例如应写int4而不是 integer。

spi_prepare 的返回值是一个查询 ID，供后续调用 spi_execp 时使用。示例参见 spi_execp。

spi_execp ?-count *n*? ?-array *name*? ?-nulls *string*? *queryid* ?*value-list*? ?*loop-body*?

执行先前用 spi_prepare 准备好的查询。*queryid* 是 spi_prepare 返回的 ID。如果查询引用了参数，则必须提供 *value-list*。这是参数实际值构成的 Tcl 列表。该列表的长度必须与先前提供给 spi_prepare 的参数类型列表相同。如果查询没有参数，则省略 *value-list*。

可选的 -nulls 值是由空格和 'n' 字符组成的字符串，用来告诉 spi_execp 哪些参数是空值。如果给出，它的长度必须与 *value-list* 完全相同。如果未给出，则所有参数值都视为非空值。

除了指定查询及其参数的方式之外，spi_execp 的工作方式与 spi_exec 完全相同。-count、-array 和 *loop-body* 选项的含义相同，结果值也相同。

下面是使用已准备计划的 PL/Tcl 函数示例：

```
CREATE FUNCTION t1_count(integer, integer) RETURNS integer AS $$
    if {![ info exists GD(plan) ]} {
        # prepare the saved plan on the first call
        set GD(plan) [ spi_prepare \
                "SELECT count(*) AS cnt FROM t1 WHERE num >= \$1
 AND num <= \$2" \
                [ list int4 int4 ] ]
    }
    spi_execp -count 1 $GD(plan) [ list $1 $2 ]
    return $cnt
$$ LANGUAGE pltcl;
```

我们需要在传给 spi_prepare 的查询字符串中加入反斜杠，以确保 $*n* 标记会原样传递给 spi_prepare，而不会被 Tcl 执行变量替换。

spi_lastoid

如果最后一次 spi_exec 或 spi_execp 执行的是单行 INSERT，且被修改的表包含 OID，则返回所插入行的 OID。（否则返回零。）

quote*string*

将给定字符串中的所有单引号和反斜杠字符都加倍。这可用于安全地为那些要插入到传给 spi_exec 或 spi_prepare 的 SQL 命令中的字符串加引号。例如，考虑如下 SQL 命令字符串：

```
"SELECT '$val' AS ret"
```

其中 Tcl 变量 `val` 的实际内容是 `doesn't`。这会得到最终命令字符串：

```
SELECT 'doesn't' AS ret
```

这会在 `spi_exec` 或 `spi_prepare` 期间导致解析错误。要使其正常工作，提交的命令应当包含：

```
SELECT 'doesn''t' AS ret
```

它在 PL/Tcl 中可以这样构成：

```
"SELECT '[ quote $val ]' AS ret"
```

`spi_execp` 的一个优点是，你不必像这样为参数值加引号，因为参数永远不会被当作 SQL 命令字符串的一部分来解析。

`elog`*`levelmsg`*

发出日志或错误消息。可用级别包括 `DEBUG`、`LOG`、`INFO`、`NOTICE`、`WARNING`、`ERROR` 和 `FATAL`。`ERROR` 会引发错误条件；如果外围 Tcl 代码没有捕获它，该错误就会传播到调用查询，导致当前事务或子事务中止。这实际上与 Tcl 的 `error` 命令相同。`FATAL` 会中止事务并导致当前会话关闭。（在 PL/Tcl 函数中使用这个错误级别可能并没有什么充分理由，但为了完整性仍提供它。）其他级别只会生成不同优先级的消息。
某个特定优先级的消息是报告给客户端、写入服务器日志，还是两者都做，由log_min_messages和client_min_messages配置变量控制。参见第 17 章了解更多信息。

37.6. PL/Tcl 中的触发器函数

可以用 PL/Tcl 编写触发器函数。PostgreSQL 要求作为触发器调用的函数必须声明为无参数、返回类型为 `trigger` 的函数。

触发器管理器的信息通过下列变量传入过程体：

`$TG_name`

`CREATE TRIGGER` 语句中触发器的名称。

`$TG_relid`

导致触发器过程被调用的表的对象 ID。

`$TG_relatts`

表列名构成的 Tcl 列表，其前面带有一个空列表元素。因此，使用 Tcl 的 `lsearch` 命令在该列表中查找列名时，返回的元素编号会从 1 开始表示第一列，这与 PostgreSQL 中列的惯常编号方式一致。（空列表元素也会出现在已删除列的位置上，这样其右侧列的属性编号仍然正确。）

`$TG_when`

根据触发器事件的类型，为字符串 `BEFORE` 或 `AFTER`。

`$TG_level`

根据触发器事件的类型，为字符串 `ROW` 或 `STATEMENT`。

`$TG_op`

字符串 `INSERT`、`UPDATE` 或 `DELETE`，取决于触发器事件的类型。

$NEW

对于 INSERT 或 UPDATE 动作，这是一个包含新表行值的关联数组；对于 DELETE 则为空。该数组以列名为索引。值为空值的列不会出现在数组中。对于语句级触发器，不会设置这个变量。

$OLD

对于 UPDATE 或 DELETE 动作，这是一个包含旧表行值的关联数组；对于 INSERT 则为空。该数组以列名为索引。值为空值的列不会出现在数组中。对于语句级触发器，不会设置这个变量。

$args

一个 Tcl 列表，包含 CREATE TRIGGER 语句中给出的过程参数。这些参数也可以在过程体中以 $1 ... $*n* 的形式访问。

触发器过程的返回值可以是字符串 OK 或 SKIP，也可以是 array get Tcl 命令返回的列表。如果返回值为 OK，触发该触发的操作（INSERT/UPDATE/DELETE）将正常进行。SKIP 告诉触发器管理器静默地取消针对该行的操作。如果返回的是列表，则告诉 PL/Tcl 向触发器管理器返回一个修改后的行，用该行代替 $NEW 中给出的行被插入。（这只对 INSERT 和 UPDATE 有效。）不用说，所有这些只在触发器是 BEFORE 且为 FOR EACH ROW 时才有意义；否则返回值会被忽略。

下面是一个简单的触发器函数示例，强制用表中的一个整数值记录该行的更新次数。
新插入的行将此值初始化为 0，以后每次更新操作都将其递增。

```
CREATE FUNCTION trigfunc_modcount() RETURNS trigger AS $$
    switch $TG_op {
        INSERT {
            set NEW($1) 0
        }
        UPDATE {
            set NEW($1) $OLD($1)
            incr NEW($1)
        }
        default {
            return OK
        }
    }
    return [array get NEW]
$$ LANGUAGE pltcl;

CREATE TABLE mytab (num integer, description text, modcnt integer);

CREATE TRIGGER trig_mytab_modcount BEFORE INSERT OR UPDATE ON mytab
    FOR EACH ROW EXECUTE PROCEDURE trigfunc_modcount('modcnt');
```

注意，触发器函数本身不知道列名，列名通过触发器参数提供。这样，
同一个触发器函数就可以在不同表上复用。

37.7. 模块与unknown命令

PL/Tcl 支持在使用时自动装载 Tcl 代码。它会识别一张特殊的表 pltcl_modules，该表被假定包含若干 Tcl 代码模块。如果这张表存在，模块unknown会从该表取出，并在数据库会话中第一

次执行 PL/Tcl 函数之前立即装载到 Tcl 解释器中。（如果一个会话中使用了多个 Tcl 解释器，则对每个解释器分别执行这一过程；参见第 37.4 节。）

虽然unknown模块实际上可以包含你需要的任何初始化脚本，但它通常会定义一个 Tcl 的unknown过程，每当 Tcl 无法识别被调用的过程名时，就会调用该过程。PL/Tcl的这一过程的标准版本会尝试在pltcl_modules中找到一个能够定义所需过程的模块。如果找到了，就把它装载到解释器中，然后允许继续执行最初尝试的过程调用。辅助表pltcl_modfuncs提供了哪个模块定义了哪些函数的索引，使查找相当快速。

PostgreSQL发行版中包含用于维护这些表的支持脚本：pltcl_loadmod、pltcl_listmod、pltcl_delmod，以及标准unknown模块的源代码，位于share/unknown.pltcl。要支持自动装载机制，最初必须把这个模块装载到每个数据库中。

表pltcl_modules和pltcl_modfuncs 必须对所有用户可读，但最好只让数据库管理员拥有并可以写入它们。作为一种安全预防措施，除非pltcl_modules由超级用户拥有，否则 PL/Tcl 会忽略它（因而不尝试装载unknown模块）。不过，如果你足够信任其他用户，可以把这张表的 UPDATE 权限授予他们。

37.8. Tcl 过程名

在PostgreSQL中，只要参数个数或参数类型不同，就可以复用同一个函数名。不过，Tcl 要求所有过程名都必须不同。PL/Tcl 处理这一问题的方式是：在内部 Tcl 过程名中包含系统表 pg_proc 中的函数对象 ID 作为其名称的一部分。因此，名称相同但参数类型不同的PostgreSQL函数，也会对应不同的 Tcl 过程。这通常不是 PL/Tcl 程序员需要关心的事情，但在调试时可能会看见。

第 38 章 PL/Perl - Perl 过程语言

PL/Perl 是一种可加载的过程语言，可以用 Perl 编程语言[1]编写 PostgreSQL 函数。

使用 PL/Perl 的主要好处，是可以在存储函数中使用 Perl 提供的多种“字符串处理”操作符和函数。与使用 PL/pgSQL 提供的字符串函数和控制结构相比，用 Perl 解析复杂字符串可能更容易。

要在某个特定数据库中安装 PL/Perl，使用 `createlang plperl dbname`。

> **提示**
>
> 如果把某种语言安装到`template1`中，之后创建的所有数据库都会自动安装该语言。

> **注意**
>
> 使用源码包的用户必须在安装过程中专门启用 PL/Perl 的构建（更多信息见第 14 章）。使用二进制包的用户可能会在单独的子包中找到 PL/Perl。

38.1. PL/Perl 函数和参数

要用 PL/Perl 语言创建一个函数，可使用标准的CREATE FUNCTION语法：

```
CREATE FUNCTION funcname (argument-types) RETURNS return-type AS $$
    # PL/Perl 函数体
$$ LANGUAGE plperl;
```

函数的主体就是普通的 Perl 代码。事实上，PL/Perl 的粘合代码会把它包裹在一个 Perl 子程序中。PL/Perl 函数必须总是返回一个标量值。如下文所述，可以通过返回引用来返回更复杂的结构（数组、记录和集合）。绝不要返回列表。

> **注意**
>
> 在 Perl 中使用命名嵌套子程序是有危险的，特别是当它们引用外层作用域中的词法变量时。因为 PL/Perl 函数被包装成一个子程序，任何放在其中的命名子程序都会成为嵌套子程序。一般来说，创建通过代码引用（coderef）调用的匿名子程序会安全得多。
> 更多细节见 `perldiag` 手册页。

`CREATE FUNCTION` 命令的语法要求函数体写成一个字符串常量。通常对该字符串常量使用美元引用（见第 4.1.2.2 节）最方便。如果选择使用常规的单引号字符串常量语法，就必须对函数体中用到的单引号（`'`）和反斜线（`\`）进行转义，通常是把它们写成双份（见第 4.1.2.1 节）。

参数和结果的处理与其他 Perl 子程序一样：参数通过 `@_` 传入，结果值通过 `return` 返回，或者作为函数中最后一个被求值的表达式返回。

例如，一个返回两个整数值中较大值的函数可以定义为：

[1] http://www.perl.com

```
CREATE FUNCTION perl_max (integer, integer) RETURNS integer AS $$
    if ($_[0] > $_[1]) { return $_[0]; }
    return $_[1];
$$ LANGUAGE plperl;
```

如果一个 SQL 空值被传给一个函数，在 Perl 中该参数值将呈现为“未定义”。上述函数定义对于空值输入的行为并不理想（实际上，它会把它们当作零）。我们可以在函数定义中添加 STRICT，让 PostgreSQL 采取更合理的做法：如果传入空值，函数将根本不会被调用，而是自动返回空值结果。另一种方式是在函数体中检查未定义输入。例如，假设我们希望在 perl_max 的两个参数中一个为空值、另一个非空值时，返回非空值参数而不是空值：

```
CREATE FUNCTION perl_max (integer, integer) RETURNS integer AS $$
    my ($x,$y) = @_;
    if (! defined $x) {
        if (! defined $y) { return undef; }
        return $y;
    }
    if (! defined $y) { return $x; }
    if ($x > $y) { return $x; }
    return $y;
$$ LANGUAGE plperl;
```

如上所示，要从 PL/Perl 函数返回 SQL 空值，就返回一个未定义值。无论函数是否为严格函数，都可以这样做。

Perl 可以把 PostgreSQL 数组返回为对 Perl 数组的引用。这里有一个示例：

```
CREATE OR REPLACE function returns_array()
RETURNS text[][] AS $$
    return [['a"b','c,d'],['e\\f','g']];
$$ LANGUAGE plperl;

select returns_array();
```

复合类型参数被作为哈希的引用传递给函数。哈希的键是复合类型的属性名。这里是一个示例：

```
CREATE TABLE employee (
    name text,
    basesalary integer,
    bonus integer
);

CREATE FUNCTION empcomp(employee) RETURNS integer AS $$
    my ($emp) = @_;
    return $emp->{basesalary} + $emp->{bonus};
$$ LANGUAGE plperl;

SELECT name, empcomp(employee.*) FROM employee;
```

PL/Perl 函数也可以用相同的方法返回复合类型：返回一个具有所需属性的哈希的引用。例如：

```
CREATE TYPE testrowperl AS (f1 integer, f2 text, f3 text);

CREATE OR REPLACE FUNCTION perl_row() RETURNS testrowperl AS $$
    return {f2 => 'hello', f1 => 1, f3 => 'world'};
```

```
$$ LANGUAGE plperl;

SELECT * FROM perl_row();
```

在声明的结果数据类型中，凡是哈希中未出现的列都将作为空值返回。

PL/Perl 函数还可以返回标量类型或复合类型的集合。通常会希望逐行返回，这既能缩短启动时间，也能避免把整个结果集排入内存。如下所示，可以用 `return_next` 实现这一点。注意，在最后一次 `return_next` 之后，必须写上 `return` 或者（更好）`return undef`。

```
CREATE OR REPLACE FUNCTION perl_set_int(int)
RETURNS SETOF INTEGER AS $$
    foreach (0..$_[0]) {
        return_next($_);
    }
    return undef;
$$ LANGUAGE plperl;

SELECT * FROM perl_set_int(5);

CREATE OR REPLACE FUNCTION perl_set()
RETURNS SETOF testrowperl AS $$
    return_next({ f1 => 1, f2 => 'Hello', f3 => 'World' });
    return_next({ f1 => 2, f2 => 'Hello', f3 => 'PostgreSQL' });
    return_next({ f1 => 3, f2 => 'Hello', f3 => 'PL/Perl' });
    return undef;
$$ LANGUAGE plperl;
```

对于较小的结果集，可以返回一个数组引用，其中对简单类型、
数组类型和复合类型分别包含标量、数组引用或哈希引用。
下面是把整个结果集作为数组引用返回的几个简单示例：

```
CREATE OR REPLACE FUNCTION perl_set_int(int) RETURNS SETOF INTEGER AS
 $$
    return [0..$_[0]];
$$ LANGUAGE plperl;

SELECT * FROM perl_set_int(5);

CREATE OR REPLACE FUNCTION perl_set() RETURNS SETOF testrowperl AS $$
    return [
        { f1 => 1, f2 => 'Hello', f3 => 'World' },
        { f1 => 2, f2 => 'Hello', f3 => 'PostgreSQL' },
        { f1 => 3, f2 => 'Hello', f3 => 'PL/Perl' }
    ];
$$ LANGUAGE plperl;

SELECT * FROM perl_set();
```

PL/Perl 目前对域类型没有完整支持：它把域同其底层标量类型一样对待。这意味着与域关联的约束不会被强制执行。这对函数参数来说不是问题，但如果你把一个 PL/Perl 函数声明为返回域类型，就会存在隐患。

如果希望在代码中使用 `strict` 编译指示，最简单的方法是把 `SET plperl.use_strict` 设为真。该参数影响后续编译的 PL/Perl 函数，但不会影响当前会话中已经编译的函数。要在 PL/Perl

装载之前设置该参数，必须先把“plperl”加入 postgresql.conf 中的custom_variable_classes列表。

使用 strict 编译指示的另一种方法是把：

```
use strict;
```

放在函数体中。

38.2. 从 PL/Perl 访问数据库

在 Perl 函数中访问数据库本身，可以通过下文描述的函数 spi_exec_query，也可以通过一个实验性模块 DBD::PgSPI[2]（也可以从 CPAN 镜像站点[3]获取）。该模块提供了一个名为 $pg_dbh 的、符合 DBI 的数据库句柄，可用普通的 DBI 语法执行查询。

PL/Perl 提供了三个附加的 Perl 命令：

```
spi_exec_query(query [, max-rows])
spi_exec_query(command)
spi_query(command)
spi_fetchrow(command)
```

spi_exec_query 执行 SQL 命令，并返回整个行集；返回值是对一个数组的引用，数组元素为哈希引用。只有在确定结果集会比较小时，才应使用此命令。下面是一个查询（SELECT 命令）的示例，指定了可选的最大行数：

```
$rv = spi_exec_query('SELECT * FROM my_table', 5);
```

这会返回最多 5 行，来自表 my_table。如果 my_table 包含列 my_column，就可以取得结果中第 $i 行的该列值，方法如下：

```
$foo = $rv->{rows}[$i]->{my_column};
```

要取得 SELECT 查询返回的总行数，可以这样做：

```
$nrows = $rv->{processed}
```

下面是使用另一种命令类型的示例：

```
$query = "INSERT INTO my_table VALUES (1, 'test')";
$rv = spi_exec_query($query);
```

可以这样访问命令状态（例如 SPI_OK_INSERT）：

```
$res = $rv->{status};
```

要获取受影响的行数，可使用：

```
$nrows = $rv->{processed};
```

这里是一个完整的示例：

```
CREATE TABLE test (
    i int,
    v varchar
```

[2]http://www.cpan.org/modules/by-module/DBD/APILOS/
[3]http://www.cpan.org/SITES.html

```
);

INSERT INTO test (i, v) VALUES (1, 'first line');
INSERT INTO test (i, v) VALUES (2, 'second line');
INSERT INTO test (i, v) VALUES (3, 'third line');
INSERT INTO test (i, v) VALUES (4, 'immortal');

CREATE OR REPLACE FUNCTION test_munge() RETURNS SETOF test AS $$
    my $rv = spi_exec_query('select i, v from test;');
    my $status = $rv->{status};
    my $nrows = $rv->{processed};
    foreach my $rn (0 .. $nrows - 1) {
        my $row = $rv->{rows}[$rn];
        $row->{i} += 200 if defined($row->{i});
        $row->{v} =~ tr/A-Za-z/a-zA-Z/ if (defined($row->{v}));
        return_next($row);
    }
    return undef;
$$ LANGUAGE plperl;

SELECT * FROM test_munge();
```

spi_query 和 spi_fetchrow 需要配合使用，适用于结果集可能很大，或者希望在行到达时立即返回的情况。spi_fetchrow 只能与 spi_query 一起使用。下面的示例演示了它们的配合方式：

```
CREATE TYPE foo_type AS (the_num INTEGER, the_text TEXT);

CREATE OR REPLACE FUNCTION lotsa_md5 (INTEGER) RETURNS SETOF foo_
type AS $$
    use Digest::MD5 qw(md5_hex);
    my $file = '/usr/share/dict/words';
    my $t = localtime;
    elog(NOTICE, "opening file $file at $t" );
    open my $fh, '<', $file # ooh, it's a file access!
        or elog(ERROR, "Can't open $file for reading: $!");
    my @words = <$fh>;
    close $fh;
    $t = localtime;
    elog(NOTICE, "closed file $file at $t");
    chomp(@words);
    my $row;
    my $sth = spi_query("SELECT * FROM generate_series(1,$_[0]) AS
 b(a)");
    while (defined ($row = spi_fetchrow($sth))) {
        return_next({
            the_num => $row->{a},
            the_text => md5_hex($words[rand @words])
        });
    }
    return;
$$ LANGUAGE plperlu;

SELECT * from lotsa_md5(500);
```

`elog(level, msg)`

发出日志消息或错误信息。可用级别包括 `DEBUG`、`LOG`、`INFO`、`NOTICE`、`WARNING`以及`ERROR`。`ERROR` 会引发错误条件；如果周围的 Perl 代码没有捕获它，错误就会传播到调用查询，导致当前事务或子事务被中止。这实质上等同于 Perl 的 `die` 命令。其他级别只会生成不同优先级的消息。某一优先级的消息是报告给客户端、写入服务器日志，还是两者兼有，由配置变量log_min_messages和client_min_messages控制。详见第 17 章。

38.3. PL/Perl 中的数据值

传给 PL/Perl 函数代码的参数值，就是转换为文本形式的输入参数（就像它们被 `SELECT` 语句显示出来一样）。反过来，`return`命令会接受任何可作为函数所声明返回类型之可接受输入格式的字符串。因此，在 PL/Perl 函数内部，所有值都只是文本字符串。

38.4. PL/Perl 中的全局值

可以使用全局哈希 `%_SHARED` 在当前会话的整个生命周期内跨函数调用存储数据，包括代码引用。

这是共享数据的一个简单示例：

```
CREATE OR REPLACE FUNCTION set_var(name text, val text) RETURNS text
 AS $$
    if ($_SHARED{$_[0]} = $_[1]) {
        return 'ok';
    } else {
        return "can't set shared variable $_[0] to $_[1]";
    }
$$ LANGUAGE plperl;

CREATE OR REPLACE FUNCTION get_var(name text) RETURNS text AS $$
    return $_SHARED{$_[0]};
$$ LANGUAGE plperl;

SELECT set_var('sample', 'Hello, PL/Perl!  How's tricks?');
SELECT get_var('sample');
```

下面是一个稍复杂一些的、使用代码引用的示例：

```
CREATE OR REPLACE FUNCTION myfuncs() RETURNS void AS $$
    $_SHARED{myquote} = sub {
        my $arg = shift;
        $arg =~ s/(['\\])/\\$1/g;
        return "'$arg'";
    };
$$ LANGUAGE plperl;

SELECT myfuncs(); /* 初始化函数 */

/* 创建一个使用加引号函数的函数 */

CREATE OR REPLACE FUNCTION use_quote(TEXT) RETURNS text AS $$
```

```
    my $text_to_quote = shift;
    my $qfunc = $_SHARED{myquote};
    return &$qfunc($text_to_quote);
$$ LANGUAGE plperl;
```

（上面的代码也可以替换为单行 `return  $_SHARED{myquote}->($_[0]);`，但代价是可读性会变差。）

出于安全原因，PL/Perl 会在每个 SQL 角色各自独立的 Perl 解释器中执行该角色调用的函数。这可以防止一个用户意外或恶意地干扰另一个用户的 PL/Perl 函数行为。每个这样的解释器都有自己的 `%_SHARED` 变量值和其他全局状态。因此，当且仅当两个 PL/Perl 函数都由同一 SQL 角色执行时，它们才会共享 `%_SHARED` 的值。在某些应用中，一个会话可能会在多个 SQL 角色下执行代码（通过 `SECURITY  DEFINER` 函数、使用 `SET  ROLE` 等），这时可能需要显式采取措施，确保 PL/Perl 函数能够通过 `%_SHARED` 共享数据。为此，要确保需要相互通信的函数由同一用户拥有，并将它们标记为 `SECURITY  DEFINER`。当然，必须注意不要让这些函数被用于任何非预期用途。

38.5. 受信任与不受信任的 PL/Perl

通常，PL/Perl 会被安装为一种名为 `plperl` 的“受信任的”编程语言。在这种设置下，为了保持安全性，某些 Perl 操作会被禁用。一般来说，受限制的是那些与环境交互的操作，包括文件句柄操作、`require` 和 `use`（针对外部模块）。它无法像 C 函数那样访问数据库服务器进程的内部，也无法以服务器进程的权限获取操作系统级访问。因此，
可以允许任何无特权的数据库用户使用这种语言。

下面示例中的函数将无法工作，因为出于安全原因不允许它做文件操作：

```
CREATE FUNCTION badfunc() RETURNS integer AS $$
    my $tmpfile = "/tmp/badfile";
    open my $fh, '>', $tmpfile
        or elog(ERROR, qq{Could not open the file "$tmpfile": $!});
    print $fh "Testing writing to a file\n";
    close $fh or elog(ERROR, qq{Could not close the file "$tmpfile":
 $!});
    return 1;
$$ LANGUAGE plperl;
```

这个函数的创建会失败，因为验证器会捕捉到它使用了禁用的操作。

有时需要编写不受这些限制的 Perl 函数。例如，可能需要一个能发送邮件的 Perl 函数。为处理这类情况，也可以把 PL/Perl 安装成一种“不受信任的”语言（通常称为 PL/PerlU）。在这种情况下，完整的 Perl 语言都可用。如果用 `createlang` 程序安装该语言，则语言名 `plperlu` 会选择不受信任的 PL/Perl 变体。

PL/PerlU 函数的编写者必须注意，函数不能被用于任何非预期用途，因为它能够执行以数据库管理员身份登录的用户所能做的任何事情。请注意，
数据库系统只允许数据库超级用户用不受信任的语言创建函数。

如果上述函数是由超级用户使用 `plperlu` 语言创建的，执行就会成功。

> **注意**
>
> 虽然 PL/Perl 函数会为每个 SQL 角色在独立的 Perl 解释器中运行，但给定会话中执行的所有 PL/PerlU 函数都运行在同一个 Perl 解释器中（而且这个解释器并不是任何一个用于

PL/Perl 函数的解释器）。这允许 PL/PerlU 函数自由共享数据，但 PL/Perl 与 PL/PerlU 函数之间不能进行通信。

注意

Perl 无法在单个进程内支持多个解释器，除非构建时使用了适当的标志，即 `usemultiplicity` 或 `useithreads` 之一。（除非确实需要使用线程，否则更推荐 `usemultiplicity`。更多细节见 perlembed 手册页。）如果 PL/Perl 使用的是未按这种方式构建的 Perl 副本，那么每个会话中就只能有一个 Perl 解释器，因此任一会话只能执行 PL/PerlU 函数，或者执行全部由同一 SQL 角色调用的 PL/Perl 函数。

38.6. PL/Perl 触发器

PL/Perl 可用于编写触发器函数。在触发器函数中，哈希引用 `$_TD` 包含有关当前触发器事件的信息。当前触发器事件的信息。`$_TD` 哈希引用的字段包括：

`$_TD->{new}{foo}`

列 `foo` 的 `NEW` 值

`$_TD->{old}{foo}`

列 `foo` 的 `OLD` 值

`$_TD->{name}`

被调用的触发器名称

`$_TD->{event}`

触发器事件：`INSERT`、`UPDATE`、`DELETE` 或 `UNKNOWN`

`$_TD->{when}`

触发器被调用的时机：`BEFORE`、`AFTER` 或 `UNKNOWN`

`$_TD->{level}`

触发器级别：`ROW`、`STATEMENT` 或 `UNKNOWN`

`$_TD->{relid}`

触发该触发器的表的 OID

`$_TD->{relname}`

触发该触发器的表的名称。

`$_TD->{argc}`

触发器函数的参数数目

`@{$_TD->{args}}`

触发器函数的参数。如果 `$_TD->{argc}` 为 0，则该字段不存在。

触发器可以返回下列之一：

`return;`

执行语句

`"SKIP"`

不执行语句

`"MODIFY"`

指示触发器函数修改了 `NEW` 行

下面的触发器函数示例演示了上述部分内容：

```
CREATE TABLE test (
    i int,
    v varchar
);

CREATE OR REPLACE FUNCTION valid_id() RETURNS trigger AS $$
    if (($_TD->{new}{i} >= 100) || ($_TD->{new}{i} <= 0)) {
        return "SKIP";    # 跳过 INSERT/UPDATE 命令
    } elsif ($_TD->{new}{v} ne "immortal") {
        $_TD->{new}{v} .= "(modified by trigger)";
        return "MODIFY";  # 修改行并执行 INSERT/UPDATE 命令
    } else {
        return;           # 执行 INSERT/UPDATE 命令
    }
$$ LANGUAGE plperl;

CREATE TRIGGER test_valid_id_trig
    BEFORE INSERT OR UPDATE ON test
    FOR EACH ROW EXECUTE PROCEDURE valid_id();
```

38.7. 限制与缺失特性

PL/Perl 目前仍缺少下列特性，但欢迎为此作出贡献。

- PL/Perl 函数不能直接调用彼此（因为它们是 Perl 内部的匿名子程序）。
- SPI 尚未完全实现。
- 如果使用 `spi_exec_query` 取回非常大的数据集，应注意这些数据都会进入内存。可以像前面所示那样，通过使用 `spi_query`/`spi_fetchrow` 来避免这种情况。

  如果集合返回函数通过 `return` 把一个大型行集返回给 PostgreSQL，也会出现类似问题。前面已经说明过，可以改为对每一行使用 `return_next`，从而避免这个问题。

第 39 章 PL/Python - Python 过程语言

PL/Python 过程语言允许用 Python 语言[1]编写 PostgreSQL 函数。

要在某个特定数据库中安装 PL/Python，使用 `createlang plpythonu dbname`。

> **提示**
>
> 如果把某种语言安装到`template1`中，之后创建的所有数据库都会自动安装该语言。

从 PostgreSQL 7.4 起，PL/Python 只能作为一种“不受信任的”语言使用（这意味着它不提供任何机制来限制用户能在其中做什么）。因此其名称被改为 `plpythonu`。如果将来能在 Python 中开发出新的安全执行机制，受信任的变体 `plpython` 可能会重新出现。

> **注意**
>
> 源码包用户必须在安装过程中专门启用 PL/Python 的构建（更多信息请参阅安装说明）。二进制包用户则可能会在单独的子包中找到 PL/Python。

39.1. PL/Python 函数

PL/Python 中的函数通过常用的CREATE FUNCTION语法声明。例如：

```
CREATE FUNCTION myfunc(text) RETURNS text
    AS 'return args[0]'
    LANGUAGE plpythonu;
```

作为函数体给出的 Python 代码会被转换成一个 Python 函数。例如，上面的定义会得到

```
def __plpython_procedure_myfunc_23456():
        return args[0]
```

这里假定 23456 是 PostgreSQL 分配给该函数的 OID。

如果你不提供返回值，Python 会返回默认的 `None`。PL/Python 会把 Python 的 `None` 转换为 SQL 空值。

PostgreSQL 函数的参数可以在全局 `args` 列表中获得。在 `myfunc` 示例中，`args[0]` 包含作为 text 参数传入的内容。对于`myfunc2(text, integer)`，`args[0]` 将包含 `text` 参数，`args[1]` 包含 `integer` 参数。

全局字典 `SD` 可用于在函数调用之间存储数据。这个变量是私有的静态数据。全局字典 `GD` 是公共数据，同一会话中的所有 Python 函数都可以使用。请谨慎使用。

每个函数在 Python 解释器中都有自己独立的执行环境，因此 `myfunc` 的全局数据和函数参数对 `myfunc2` 不可用。例外是上面提到的 `GD` 字典中的数据。

[1] http://www.python.org

39.2. 触发器函数

当函数被用作触发器时，字典 `TD` 包含与触发器相关的值。触发行位于 `TD["new"]` 和/或 `TD["old"]` 中，取决于触发器事件。`TD["event"]` 以字符串形式包含事件（`INSERT`、`UPDATE`、`DELETE` 或 `UNKNOWN`）。`TD["when"]` 包含 `BEFORE`、`AFTER` 和 `UNKNOWN` 之一。`TD["level"]` 包含 `ROW`、`STATEMENT` 和 `UNKNOWN` 之一。`TD["name"]` 包含触发器的名称，`TD["relid"]` 包含触发器所作用的表的 OID。如果 `CREATE TRIGGER` 命令包含参数，它们可以在 `TD["args"][0]` 到 `TD["args"][(n-1)]` 中获得。

如果 `TD["when"]` 是 `BEFORE`，你可以从 Python 函数返回 `None` 或 `"OK"` 表示行未修改，返回 `"SKIP"` 中止该事件，或返回 `"MODIFY"` 表示你修改了该行。

39.3. 数据库访问

PL/Python 语言模块会自动导入一个名为 `plpy` 的 Python 模块。该模块中的函数和常量在 Python 代码中可以通过 `plpy.foo` 这样的形式访问。目前 `plpy` 实现了函数 `plpy.debug(msg)`、`plpy.log(msg)`、`plpy.info(msg)`、`plpy.notice(msg)`、`plpy.warning(msg)`、`plpy.error(msg)` 和 `plpy.fatal(msg)`。`plpy.error` 和 `plpy.fatal` 实际上会抛出一个 Python 异常；如果该异常未被捕获，会传播到调用查询之外，导致当前事务或子事务中止。`raise plpy.ERROR(msg)` 和 `raise plpy.FATAL(msg)` 分别等效于调用 `plpy.error` 和 `plpy.fatal`。其余函数只生成不同优先级别的消息。特定优先级别的消息是报告给客户端、写入服务器日志还是两者都做，由log_min_messages和client_min_messages配置变量控制。更多信息见第 17 章。

此外，`plpy` 模块提供了两个函数：`execute` 和 `prepare`。使用查询字符串和一个可选的限制参数调用 `plpy.execute` 会执行该查询，并把结果在一个结果对象中返回。结果对象模拟了列表或字典对象。可以通过行号和列名访问结果对象。它还有这些附加方法：`nrows` 返回查询返回的行数，`status` 是 `SPI_execute()` 的返回值。结果对象可以被修改。

例如：

```
rv = plpy.execute("SELECT * FROM my_table", 5)
```

会从 `my_table` 中返回最多 5 行。如果 `my_table` 有一列名为 `my_column`，则可以这样访问它：

```
foo = rv[i]["my_column"]
```

第二个函数 `plpy.prepare` 为查询准备执行计划。若查询中有参数引用，则它以查询字符串和参数类型列表作为参数。例如：

```
plan = plpy.prepare("SELECT last_name FROM my_users WHERE first_name =
 $1", [ "text" ])
```

`text` 是你要传给 `$1` 的变量的类型。准备好语句之后，使用函数 `plpy.execute` 运行它：

```
rv = plpy.execute(plan, [ "name" ], 5)
```

第三个参数是限制，它是可选的。

当你使用 PL/Python 模块准备一个计划时，该计划会被自动保存。关于这意味着什么，请阅读 SPI 文档（第 40 章）。为了在函数调用之间有效地利用这一点，需要使用持久存储字典 `SD` 或 `GD`（见第 39.1 节）。例如：

```
CREATE FUNCTION usesavedplan() RETURNS trigger AS $$
    if SD.has_key("plan"):
        plan = SD["plan"]
    else:
        plan = plpy.prepare("SELECT 1")
        SD["plan"] = plan
    # rest of function
$$ LANGUAGE plpythonu;
```

第 40 章 服务器编程接口

服务器编程接口（SPI）使用户定义 C 函数的编写者能够在其函数中运行 SQL 命令。SPI 是一组接口函数，用于简化对解析器、规划器和执行器的访问。SPI 还负责一部分内存管理工作。

> **注意**
>
> 可用的过程语言提供了多种从过程中执行 SQL 命令的方法。这些设施大多基于 SPI，因此本文档对这些语言的用户也会有帮助。

为避免误解，本文用“函数”指SPI接口函数，而用“过程”指使用SPI的用户定义 C 函数。

注意，如果通过 SPI 调用的某条命令失败，控制不会返回到你的过程。相反，执行该过程的事务或子事务会被回滚。（考虑到 SPI 函数大多记录了错误返回约定，这一点可能看起来有些意外。不过，这些约定只适用于在 SPI 函数自身内部检测到的错误。）如果在可能失败的 SPI 调用外围建立自己的子事务，则可以在出错后重新取得控制权。

SPI 函数在成功时返回非负结果（要么直接作为整数返回值，要么如后文所述存放在全局变量 `SPI_result` 中）。发生错误时，则返回负值或 `NULL`。

使用 SPI 的源代码文件必须包含头文件 `executor/spi.h`。

40.1. 接口函数

SPI_connect

SPI_connect — 将一个过程连接到 SPI 管理器

大纲

```
int SPI_connect(void)
```

描述

SPI_connect 会为某次过程调用打开到 SPI 管理器的连接。如果要通过 SPI 执行命令，就必须调用此函数。不过，有些 SPI 辅助函数可以在未连接的过程中调用。

如果你的过程已经处于连接状态，SPI_connect将返回错误代码→ SPI_ERROR_CONNECT。当某个直接调用了SPI_connect的过程又直接调用另一个调用SPI_connect的过程时，就会发生这种情况。虽然当通过 SPI 调用的 SQL 命令调用另一个使用 SPI 的函数时，允许对 SPI 管理器的递归调用，但直接嵌套调用SPI_connect和SPI_finish是被禁止的。（但请参见SPI_push和SPI_pop。）

返回值

SPI_OK_CONNECT

成功时

SPI_ERROR_CONNECT

发生错误时

SPI_finish

SPI_finish — 将一个过程与 SPI 管理器断开

大纲

```
int SPI_finish(void)
```

描述

`SPI_finish` 关闭到 SPI 管理器的现有连接。在完成当前这次过程调用所需的 SPI 操作后，必须调用此函数。不过，如果通过 `elog(ERROR)` 中止了事务，就不必担心是否显式调用它；这种情况下 SPI 会自行清理。

如果在没有有效连接的情况下调用`SPI_finish`，它将返回`SPI_ERROR_UNCONNECTED`。这并不会带来根本性问题；它只表示 SPI 管理器无事可做。

返回值

`SPI_OK_FINISH`

如果正确地断开连接

`SPI_ERROR_UNCONNECTED`

如果从未连接的过程中调用

SPI_push

SPI_push — 压入 SPI 栈以允许递归使用 SPI

大纲

```
void SPI_push(void)
```

描述

SPI_push应当在执行另一个自身可能希望使用 SPI 的过程之前调用。在SPI_push之后，SPI 不再处于“连接”状态，除非重新执行SPI_connect，否则 SPI 函数调用将被拒绝。这确保了你的过程的 SPI 状态与你所调用的另一个过程的状态之间的清晰分离。在另一个过程返回之后，调用SPI_pop恢复对你自己 SPI 状态的访问。

注意，SPI_execute及相关函数在把控制交还给 SQL 执行引擎之前，会自动完成相当于SPI_push的操作，因此在使用这些函数时无需为此操心。只有当你直接调用可能包含SPI_connect调用的任意代码时，才需要发出SPI_push和SPI_pop。

SPI_pop

SPI_pop — 弹出 SPI 栈以从递归 SPI 使用中返回

大纲

```
void SPI_pop(void)
```

描述

`SPI_pop`从 SPI 调用栈中弹出先前的环境。参见`SPI_push`。

SPI_execute

SPI_execute — 执行一个命令

大纲

```
int SPI_execute(const char * command, bool read_only, long count)
```

描述

SPI_execute 执行指定的 SQL 命令，并最多检索 *count* 行。如果 *read_only* 为 true，该命令必须是只读的，且执行开销会略有降低。

此函数只能从已连接的过程中调用。

如果 *count* 为零，则该命令会针对其适用的所有行执行。如果 *count* 大于 0，则该命令要执行的行数会受到限制（很像一个 LIMIT 子句）。例如：

```
SPI_execute("INSERT INTO foo SELECT * FROM bar", false, 5);
```

最多只允许向该表中插入 5 行。

可以在一个字符串中传递多条命令。SPI_execute 返回最后执行的那条命令的结果。*count* 限制会分别作用于每条命令，但不适用于规则生成的隐藏命令。

当 *read_only* 为 false 时，SPI_execute 会递增命令计数器，并在执行字符串中的每条命令前计算新的快照。如果当前事务隔离级别是 SERIALIZABLE，这个快照实际上不会变化；但在 READ COMMITTED 模式下，更新快照会让每条命令都能看到其他会话中新近提交事务的结果。这对于修改数据库的命令获得一致行为至关重要。

当 *read_only* 为 true 时，SPI_execute 不会更新快照和命令计数器，并且只允许命令字符串中出现普通的 SELECT 命令。这些命令会使用外围查询先前建立的快照来执行。由于消除了每条命令的额外开销，这种执行模式比读写模式略快。它还允许构造真正
稳定的函数：由于连续执行都会使用同一个快照，结果也就不会发生变化。

在同一个使用 SPI 的函数中混合只读命令和读写命令通常并不明智，因为只读查询看不到读写查询所做的数据库更新，这可能导致非常令人困惑的行为。

（最后一条）命令实际执行所处理的行数，会通过全局变量 SPI_processed 返回（除非函数的返回值是 SPI_OK_UTILITY）。如果函数的返回值是 SPI_OK_SELECT，则可以通过全局指针 SPITupleTable *SPI_tuptable 访问结果行。

结构 SPITupleTable 定义如下：

```
typedef struct
{
    MemoryContext tuptabcxt;    /* 结果表的内存上下文 */
    uint32      alloced;        /* 已分配的 vals 数量 */
    uint32      free;           /* 空闲的 vals 数量 */
    TupleDesc   tupdesc;        /* 行描述符 */
    HeapTuple  *vals;           /* 行 */
} SPITupleTable;
```

vals 是一个指向行的指针数组。（有效项数为 SPI_processed。）tupdesc 是一个行描述符，可以传给处理行的 SPI 函数。tuptabcxt，alloced 和 free 是内部字段，不供 SPI 调用者使用。

SPI_finish 会释放当前过程调用期间分配的全部 SPITupleTable。如果某个结果表已经不再需要，也可以提前调用 SPI_freetuptable 释放它。

参数

const char * *command*

包含待执行命令的字符串

bool *read_only*

true 表示只读执行

long *count*

要处理或返回的最大行数

返回值

如果命令执行成功，则返回下列（非负）值之一：

SPI_OK_SELECT

执行了 SELECT（但不是 SELECT INTO）

SPI_OK_SELINTO

执行了 SELECT INTO

SPI_OK_DELETE

执行了 DELETE

SPI_OK_INSERT

执行了 INSERT

SPI_OK_UPDATE

执行了 UPDATE

SPI_OK_UTILITY

执行了工具命令（例如 CREATE TABLE）

出错时，返回下列负值之一：

SPI_ERROR_ARGUMENT

如果*command*为NULL或 *count*小于 0

SPI_ERROR_COPY

尝试执行了 COPY TO stdout 或 COPY FROM stdin

SPI_ERROR_CURSOR

试图执行DECLARE、CLOSE或FETCH命令

SPI_ERROR_TRANSACTION

尝试执行了事务控制命令（BEGIN、COMMIT、ROLLBACK、SAVEPOINT、PREPARE TRANSACTION、COMMIT PREPARED、ROLLBACK PREPARED 及其各种变体）

SPI_ERROR_OPUNKNOWN

命令类型未知（理论上不应发生）

SPI_ERROR_UNCONNECTED

如果从未连接的过程中调用

注解

函数 SPI_execute、SPI_exec、SPI_execute_plan 和 SPI_execp 都会更改 SPI_processed 和 SPI_tuptable（只更改指针，而不更改结构体内容）。如果需要在后续调用之后继续访问 SPI_execute 或相关函数的结果表，请把这两个全局变量保存到过程局部变量中。

SPI_exec

SPI_exec — 以读/写模式执行一个命令

大纲

```
int SPI_exec(const char * command, long count)
```

描述

`SPI_exec`和 `SPI_execute`相同，但后者的 *`read_only`*参数的值总是取 `false`。

参数

`const char * command`

包含待执行命令的字符串

`long count`

要处理或返回的最大行数

返回值

见`SPI_execute`。

SPI_prepare

SPI_prepare — 为命令准备一个计划，但暂不执行

大纲

`void * SPI_prepare(const char * command, int nargs, Oid * argtypes)`

描述

`SPI_prepare` 为指定命令创建并返回一个执行计划，但不执行该命令。该函数只能从已连接的过程调用。

当同一条或相似命令需要反复执行时，只做一次规划可能是有利的。`SPI_prepare` 把命令字符串转换成一个执行计划，之后可以使用 `SPI_execute_plan` 反复执行。

预备命令可以通过在普通命令原本应写常量的位置写入参数（`$1`、`$2` 等）而得到泛化。这些参数的实际值会在调用 `SPI_execute_plan` 时指定。这样，预备语句就能适用于比无参形式更广泛的场景。

`SPI_prepare` 返回的计划只能在当前这次过程调用中使用，因为 `SPI_finish` 会释放为计划分配的内存。但可以使用函数 `SPI_saveplan` 把计划保存得更久。

参数

`const char * command`

命令字符串

`int nargs`

输入参数的数量（`$1`、`$2` 等）

`Oid * argtypes`

一个数组指针，它指向的数组包含参数的数据类型的 OID

返回值

`SPI_prepare` 返回一个指向执行计划的非空指针。发生错误时会返回 `NULL`，并将 `SPI_result` 设为 `SPI_execute` 所使用的那些错误码之一；但如果 *command* 为 `NULL`，或者 *nargs* 小于 0，或者 *nargs* 大于 0 且 *argtypes* 为 `NULL`，则会将其设置为 `SPI_ERROR_ARGUMENT`。

注解

使用参数有一个不利之处：由于规划器不知道将为参数提供什么值，
它做出的规划选择可能比在所有常量都可见的普通命令下更差。

SPI_getargcount

SPI_getargcount — 返回由 `SPI_prepare` 准备的语句所需参数个数

大纲

```
int SPI_getargcount(void * plan)
```

描述

`SPI_getargcount` 返回执行由 `SPI_prepare` 准备的计划所需的参数个数。

参数

`void * plan`

执行计划（由 `SPI_prepare` 返回）

返回值

该*plan*的期望参数个数；如果 *plan*为`NULL`，则为`SPI_ERROR_ARGUMENT`

SPI_getargtypeid

SPI_getargtypeid — 返回由 `SPI_prepare` 准备的语句中某个参数的数据类型 OID

大纲

`Oid SPI_getargtypeid(void * plan, int argIndex)`

描述

`SPI_getargtypeid` 返回表示由 `SPI_prepare` 准备的计划的第 *`argIndex`* 个参数类型 ID 的 OID。第一个参数的索引为零。

参数

`void * plan`

执行计划（由 `SPI_prepare` 返回）

`int argIndex`

参数的索引，从零开始

返回值

给定索引处参数的类型 ID；如果 *`plan`*为`NULL`，或者 *`argIndex`*小于 0 或不小于该 *`plan`*声明的参数个数，则为`SPI_ERROR_ARGUMENT`

SPI_is_cursor_plan

SPI_is_cursor_plan — 如果由 `SPI_prepare` 准备的计划可以传给 `SPI_cursor_open` 使用，则返回 `true`

大纲

`bool SPI_is_cursor_plan(void * plan)`

描述

如果由 `SPI_prepare` 准备的计划可以作为参数传给 `SPI_cursor_open`，则 `SPI_is_cursor_plan` 返回 `true`；否则返回 `false`。判定标准是 *plan* 必须表示单条命令，并且该命令是一条不带 `INTO` 子句的 `SELECT`。

参数

`void * plan`

执行计划（由 `SPI_prepare` 返回）

返回值

返回`true`或`false`，表示 *plan*能否产生游标；如果 *plan*为`NULL`，则为`SPI_ERROR_ARGUMENT`

SPI_execute_plan

SPI_execute_plan — 执行由 SPI_prepare 准备好的语句

大纲

```
int SPI_execute_plan(void * plan, Datum * values, const char * nulls,
                     bool read_only, long count)
```

描述

SPI_execute_plan 执行由 SPI_prepare 准备的计划。*read_only* 和 *count* 的含义与 SPI_execute 中相同。

参数

void * *plan*

执行计划（由 SPI_prepare 返回）

Datum * *values*

一个实际参数值的数组。必须与该计划的参数个数等长。

const char * *nulls*

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。n 表示空值（*values* 中的对应项会被忽略）；空格表示非空值（*values* 中的对应项有效）。

如果 *nulls* 为 NULL，则 SPI_execute_plan 假定没有参数为空值。

bool *read_only*

true 表示只读执行

long *count*

要处理或返回的最大行数

返回值

返回值和SPI_execute相同，还有下列额外可能的错误（负值）结果：

SPI_ERROR_ARGUMENT

如果*plan*为NULL 或者非法，或者*count*小于 0

SPI_ERROR_PARAM

如果*values*为NULL，但 *plan*在准备时带有参数

成功时，SPI_processed 和 SPI_tuptable 的设置方式与 SPI_execute 相同。

注解

如果预备计划所引用的对象（表、函数等）在会话期间被删除，那么对该计划执行 `SPI_execute_plan` 的结果将是不可预知的。

SPI_execp

SPI_execp — 以读/写模式执行一个语句

大纲

```
int SPI_execp(void * plan, Datum * values, const char * nulls,
 long count)
```

描述

SPI_execp 与 SPI_execute_plan 相同，只不过后者的 *read_only* 参数固定为 false。

参数

void * *plan*

执行计划（由 SPI_prepare 返回）

Datum * *values*

一个实际参数值的数组。必须与该计划的参数个数等长。

const char * *nulls*

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。n 表示空值（*values* 中的对应项会被忽略）；空格表示非空值（*values* 中的对应项有效）。

如果 *nulls* 为 NULL，则 SPI_execp 假定没有参数为空值。

long *count*

要处理或返回的最大行数

返回值

见 SPI_execute_plan。

成功时，SPI_processed 和 SPI_tuptable 的设置方式与 SPI_execute 相同。

SPI_cursor_open

SPI_cursor_open — 使用由 `SPI_prepare` 创建的计划建立一个游标

大纲

```
Portal SPI_cursor_open(const char * name, void * plan,
                       Datum * values, const char * nulls,
                       bool read_only)
```

描述

`SPI_cursor_open` 建立一个游标（在内部是一个 portal），该游标将执行由 `SPI_prepare` 准备好的一个计划。参数具有和 `SPI_execute_plan` 的相应参数相同的含义。

使用游标而不是直接执行该计划有两个好处。首先，可以一次只取出少量结果行，从而避免对返回大量行的查询耗尽内存。其次，一个 portal 可以比当前的过程存活更久（实际上，它可以一直存活到当前事务结束）。把 portal 的名称返回给该过程的调用者，就提供了一种把行集作为结果返回的方法。

参数

`const char * name`

portal 的名称，或为`NULL`以让系统选择名称

`void * plan`

执行计划（由 `SPI_prepare` 返回）

`Datum * values`

一个实际参数值的数组。必须与该计划的参数个数等长。

`const char * nulls`

一个描述哪些参数为空值的数组。必须与该计划的参数个数等长。`n` 表示空值（*values* 中的对应项会被忽略）；空格表示非空值（*values* 中的对应项有效）。

如果 *nulls* 为 `NULL`，则 `SPI_cursor_open` 假定没有参数为空值。

`bool read_only`

`true` 表示只读执行

返回值

指向包含该游标的 portal 的指针；出错时为`NULL`。

SPI_cursor_find

SPI_cursor_find — 用名称查找一个现有的游标

大纲

```
Portal SPI_cursor_find(const char * name)
```

描述

`SPI_cursor_find` 按名称查找现有的 portal。这主要用于解析由其他函数以文本形式返回的游标名。

参数

`const char * name`

该 portal 的名称

返回值

带有指定名称的 portal 的指针，如果没有找到就是 `NULL`

SPI_cursor_fetch

SPI_cursor_fetch — 从一个游标取出一些行

大纲

`void SPI_cursor_fetch(Portal portal, bool forward, long count)`

描述

`SPI_cursor_fetch`从一个游标取得一些行。这等效于 SQL 命令`FETCH`。

参数

`Portal portal`

包含该游标的 portal

`bool forward`

为真表示向前获取，为假表示向后获取

`long count`

要取得的最大行数

返回值

成功时，`SPI_processed` 和 `SPI_tuptable` 的设置方式与 `SPI_execute` 相同。

SPI_cursor_move

SPI_cursor_move — 移动一个游标

大纲

```
void SPI_cursor_move(Portal portal, bool forward, long count)
```

描述

`SPI_cursor_move`跳过一个游标中的一些行。这等效于 SQL 命令`MOVE`。

参数

`Portal portal`

包含该游标的 portal

`bool forward`

为真表示前移，为假表示后移

`long count`

要移动的最大行数

SPI_cursor_close

SPI_cursor_close — 关闭一个游标

大纲

`void SPI_cursor_close(Portal portal)`

描述

`SPI_cursor_close`关闭一个之前创建的游标并且释放它的 portal 存储。

所有打开的游标会在事务结束时自动被关闭。只有在希望尽快释放资源时，才需要调用 `SPI_cursor_close`。

参数

`Portal portal`

包含该游标的 portal

SPI_saveplan

SPI_saveplan — 保存一个计划

大纲

`void * SPI_saveplan(SPIPlanPtr plan)`

描述

`SPI_saveplan` 会把传入的计划（由 `SPI_prepare` 准备）保存到不会被 `SPI_finish` 或事务管理器释放的内存中，并返回指向所保存计划的指针。这让你能够在当前会话后续的过程调用中重用预备计划。你可以把返回的指针保存在一个局部变量中。无论是在准备计划，还是在 `SPI_execute_plan` 中使用一个已准备好的计划时，都要始终检查这个指针是否为 `NULL`。

参数

`void * plan`

要保存的计划

返回值

指向所保存计划的指针；如果不成功则为 `NULL`。出错时，`SPI_result` 会被设置为：

`SPI_ERROR_ARGUMENT`

如果 *plan* 为 `NULL` 或无效

`SPI_ERROR_UNCONNECTED`

如果从未连接的过程中调用

注解

如果预备计划所引用的对象（表、函数等）之一在会话期间被删除，那么 `SPI_execute_plan` 对该计划的执行结果将不可预测。

40.2. 接口支持函数

这里介绍的函数提供了一个接口，用于从 `SPI_execute` 及其他 SPI 函数返回的结果集中提取信息。

本节介绍的所有函数都可以在已连接和未连接的过程中使用。

SPI_fname

SPI_fname — 为指定的列号确定列名

大纲

```
char * SPI_fname(TupleDesc rowdesc, int colnumber)
```

描述

`SPI_fname` 返回指定列列名的一个副本。（不再需要时，可以用 `pfree` 释放它。）

参数

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

返回值

列名；如果 *`colnumber`* 超出范围则返回 `NULL`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_fnumber

SPI_fnumber — 根据指定列名确定列号

大纲

```
int SPI_fnumber(TupleDesc rowdesc, const char * colname)
```

描述

`SPI_fnumber` 返回指定列名对应的列号。

如果 *colname* 指向的是系统列（例如 `oid`），则会返回相应的负列号。调用者应当通过检查返回值是否恰好等于 `SPI_ERROR_NOATTRIBUTE` 来判断错误；除非本就要拒绝系统列，否则测试结果是否小于等于 0 并不正确。

参数

`TupleDesc` *rowdesc*

输入行描述

`const char *` *colname*

列名

返回值

列号（从 1 开始计），如果没有找到所提到的列名则返回 `SPI_ERROR_NOATTRIBUTE`。

SPI_getvalue

SPI_getvalue — 返回指定列的字符串值

大纲

```
char * SPI_getvalue(HeapTuple row, TupleDesc rowdesc, int colnumber)
```

描述

`SPI_getvalue`返回指定列的值的字符串表示。

结果保存在由 `palloc` 分配的内存中返回。（不再需要时，可以用 `pfree` 释放该内存。）

参数

`HeapTuple` *`row`*

要检查的输入行

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

返回值

列值；如果该列为空值、*`colnumber`* 超出范围（`SPI_result` 被设置为 `SPI_ERROR_NOATTRIBUTE`）或者没有输出函数可用（`SPI_result` 被设置为 `SPI_ERROR_NOOUTFUNC`）则返回 `NULL`。

SPI_getbinval

SPI_getbinval — 返回指定列的二进制值

大纲

```
Datum SPI_getbinval(HeapTuple row, TupleDesc rowdesc, int colnumber,
                    bool * isnull)
```

描述

`SPI_getbinval` 以内部形式（即 `Datum` 类型）返回指定列的值。

此函数不会为该 datum 分配新空间。对于传引用的数据类型，返回值将是指向传入行内部数据的指针。

参数

`HeapTuple` *`row`*

要检查的输入行

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

`bool *` *`isnull`*

列中是否为空值的标志

返回值

返回该列的二进制值。如果该列为空值，则 *`isnull`* 指向的变量会被设为真，否则设为假。

出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_gettype

SPI_gettype — 返回指定列的数据类型名称

大纲

```
char * SPI_gettype(TupleDesc rowdesc, int colnumber)
```

描述

`SPI_gettype`返回该指定列的数据类型名称的副本（当你不再需要该副本后，可以使用`pfree`释放它）。

参数

`TupleDesc` *`rowdesc`*

输入行描述

`int` *`colnumber`*

列号（从 1 开始计）

返回值

指定列的数据类型名称；出错时返回 `NULL`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_gettypeid

SPI_gettypeid — 返回指定列的数据类型 OID

大纲

`Oid SPI_gettypeid(TupleDesc rowdesc, int colnumber)`

描述

`SPI_gettypeid`返回该指定列的数据类型的 OID。

参数

`TupleDesc rowdesc`

输入行描述

`int colnumber`

列号（从 1 开始计）

返回值

指定列数据类型的 OID；出错时返回 `InvalidOid`。出错时，`SPI_result` 会被设置为 `SPI_ERROR_NOATTRIBUTE`。

SPI_getrelname

SPI_getrelname — 返回指定关系的名称

大纲

```
char * SPI_getrelname(Relation rel)
```

描述

`SPI_getrelname`返回该指定关系的名称的副本（当你不再需要该副本后，可以使用`pfree` 释放它）。

参数

`Relation rel`

输入关系

返回值

指定关系的名称。

SPI_getnspname

SPI_getnspname — 返回指定关系所属的名字空间

大纲

```
char * SPI_getnspname(Relation rel)
```

描述

SPI_getnspname 返回指定 Relation 所属名字空间名称的一个副本。这等价于该关系的模式名。用完该返回值后，应调用 pfree 释放它。

参数

Relation *rel*

输入关系

返回值

指定关系的名字空间的名称。

40.3. 内存管理

PostgreSQL 在内存上下文中分配内存。内存上下文为管理那些在许多不同位置创建、且生命周期各不相同的内存分配提供了便捷手段。销毁某个上下文时，会释放其中分配的全部内存。因此，不必为了避免内存泄漏而跟踪每个单独对象；只需管理数量相对较少的上下文即可。palloc 及相关函数都从“当前”上下文中分配内存。

SPI_connect 创建新的内存上下文并将其设为当前上下文。SPI_finish 则恢复先前的当前上下文，并销毁由 SPI_connect 创建的上下文。这些操作可以确保在过程内部所做的临时内存分配会在过程退出时被回收，从而避免内存泄漏。

不过，如果过程需要返回位于已分配内存中的对象（例如传引用数据类型的值），就不能使用 palloc 来分配这块内存，至少在连接到 SPI 的期间不能这么做。否则，该对象会在 SPI_finish 时被释放，过程也就无法可靠工作。解决办法是用 SPI_palloc 为返回对象分配内存。SPI_palloc 在“上层执行器上下文”中分配内存，也就是调用 SPI_connect 时的当前内存上下文；这正是从过程返回值最合适的上下文。

如果在过程尚未连接到 SPI 时调用SPI_palloc，它的行为与普通的palloc相同。在过程连接到 SPI 管理器之前，当前内存上下文就是上层执行器上下文，因此过程通过palloc或 SPI 辅助函数所做的所有分配都位于这个上下文中。

调用 SPI_connect 时，会把该过程的私有上下文（由 SPI_connect 创建）设为当前上下文。所有通过 palloc、repalloc 或 SPI 辅助函数分配的内存（但SPI_copytuple、SPI_returntuple、SPI_modifytuple 和SPI_palloc除外）都位于这个上下文中。当过程通过 SPI_finish 与 SPI 管理器断开连接时，当前上下文会恢复为上层执行器上下文，而在该过程内存上下文中分配的所有内存都会被释放，之后就不能再使用。

本节介绍的所有函数既可由已连接的过程使用，也可由未连接的过程使用。在未连接的过程中，它们的行为与底层普通服务器函数（palloc等）相同。

SPI_palloc

SPI_palloc — 在上层执行器上下文中分配内存

大纲

```
void * SPI_palloc(Size size)
```

描述

`SPI_palloc` 在上层执行器上下文中分配内存。

参数

`Size size`

要分配的存储空间大小（以字节计）

返回值

指向具有指定大小的新存储空间的指针

SPI_repalloc

SPI_repalloc — 在上层执行器上下文中重分配内存

大纲

```
void * SPI_repalloc(void * pointer, Size size)
```

描述

`SPI_repalloc`改变之前用`SPI_palloc` 分配的内存段的大小。

这个函数如今与普通的`repalloc`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`void * pointer`

指向要改变的现有存储空间的指针

`Size size`

要分配的存储空间大小（以字节计）

返回值

指向具有指定大小的新存储空间的指针，现有区域的内容会被复制到其中

SPI_pfree

SPI_pfree — 在上层执行器上下文中释放内存

大纲

`void SPI_pfree(void * pointer)`

描述

`SPI_pfree`释放之前使用 `SPI_palloc`或者 `SPI_repalloc`分配的内存。

这个函数如今与普通的`pfree`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`void * pointer`

指向要释放的现有存储空间的指针

SPI_copytuple

SPI_copytuple — 在上层执行器上下文中创建一行的副本

大纲

`HeapTuple SPI_copytuple(HeapTuple row)`

描述

`SPI_copytuple` 在上层执行器上下文中创建一行的副本。它通常用于从触发器中返回修改后的行。对于声明为返回复合类型的函数，应改用 `SPI_returntuple`。

参数

`HeapTuple row`

要复制的行

返回值

复制后的行；仅当*tuple*为 `NULL` 时才返回 `NULL`

SPI_returntuple

SPI_returntuple — 准备把一个元组返回为一个 Datum

大纲

`HeapTupleHeader SPI_returntuple(HeapTuple row, TupleDesc rowdesc)`

描述

`SPI_returntuple` 在上层执行器上下文中创建一行的副本，并以行类型 `Datum` 的形式返回。返回的指针只需在返回前用 `PointerGetDatum` 转换为 `Datum` 即可。

请注意，它应用于声明为返回复合类型的函数，而不用于触发器；
触发器中返回修改后的行应使用 `SPI_copytuple`。

参数

`HeapTuple row`

要复制的行

`TupleDesc rowdesc`

行描述符（若要获得最佳缓存效果，每次都传入同一个描述符）

返回值

指向复制后行的 `HeapTupleHeader`；仅当*row* 或 *rowdesc*为 `NULL` 时才返回 `NULL`

SPI_modifytuple

SPI_modifytuple — 通过替换给定行中选定的字段来创建一个行

大纲

```
HeapTuple SPI_modifytuple(Relation rel, HeapTuple row, int ncols,
                          int * colnum, Datum * values, const char
 * nulls)
```

描述

SPI_modifytuple 通过为选定列替换新值、并在其他位置复制原行的列来创建一行新记录。输入行本身不会被修改。新行会在上层执行器上下文中返回。

参数

Relation *rel*

仅用作该行的行描述符来源。（传递关系而不是行描述符是个设计缺陷。）

HeapTuple *row*

要被修改的行

int *ncols*

数组 *colnum* 中列号的数量

int * *colnum*

一个数组，包含要被修改的列的编号（列号从 1 开始）

Datum * *values*

指定列的新值

const char * *Nulls*

哪些新值为空值（如果有的话；格式参见 SPI_execute_plan）

返回值

修改后的新行，分配在上层执行器上下文中；仅当*row* 为 NULL 时才返回 NULL

出错时，SPI_result被设置如下：

SPI_ERROR_ARGUMENT

如果*rel*为NULL，或者 *row*为NULL，或者*ncols* 小于等于 0，或者*colnum*为NULL，或者*values*为NULL。

SPI_ERROR_NOATTRIBUTE

如果*colnum*包含一个无效的列号（小于等于 0 或者大于 *row*中的列数）。

SPI_freetuple

SPI_freetuple — 释放在上层执行器上下文中分配的一个行

大纲

```
void SPI_freetuple(HeapTuple row)
```

描述

`SPI_freetuple`释放之前在上层执行器上下文中分配的一行。

这个函数如今与普通的`heap_freetuple`已无区别。保留它只是为了与现有代码保持向后兼容。

参数

`HeapTuple` *`row`*

要释放的行

SPI_freetuptable

SPI_freetuptable — 释放由 `SPI_execute` 或类似函数创建的结果行集

大纲

```
void SPI_freetuptable(SPITupleTable * tuptable)
```

描述

`SPI_freetuptable` 释放由先前的 SPI 命令执行函数（例如 `SPI_execute`）创建的结果行集。因此，调用它时经常会直接把全局变量 `SPI_tupletable` 作为参数。

如果某个使用 SPI 的过程需要执行多条命令，并且不想将前面命令的结果一直保留到过程结束，那么这个函数就很有用。注意，任何尚未释放的结果行集都会在 `SPI_finish` 时被释放。

参数

`SPITupleTable * tuptable`

要释放的行集的指针

SPI_freeplan

SPI_freeplan — 释放一个先前保存的计划

大纲

`int SPI_freeplan(void * plan)`

描述

`SPI_freeplan` 释放先前由 `SPI_prepare` 返回或由 `SPI_saveplan` 保存的命令执行计划。

参数

`void * plan`

要释放的计划的指针

返回值

如果 *plan* 为 `NULL` 或无效，返回 `SPI_ERROR_ARGUMENT`

40.4. 数据更改的可见性

下列规则决定了使用 SPI 的函数（以及其他任何 C 函数）中数据更改的可见性：

- 在某条 SQL 命令执行期间，该命令所做的任何数据更改对该命令本身都是不可见的。例如，在

  ```
  INSERT INTO a SELECT * FROM a;
  ```

  中，被插入的行对`SELECT`部分不可见。

- 某条命令 C 所做的更改，对所有在 C 之后启动的命令都可见，无论这些命令是在 C 内部（即 C 执行期间）启动，还是在 C 完成之后启动。

- 在由 SQL 命令调用的函数中（无论是普通函数还是触发器）通过 SPI 执行的命令，究竟遵循上述哪条规则，取决于传给 SPI 的读写标志。只读模式下执行的命令遵循第一条规则：它们看不到调用它们的命令所做的更改。
  读写模式下执行的命令遵循第二条规则：它们能够看到截至当前为止的所有更改。

- 所有标准过程语言都会根据函数的易变性属性设置 SPI 的读写模式。`STABLE` 和 `IMMUTABLE` 函数中的命令会以只读模式执行，而 `VOLATILE` 函数中的命令则以读写模式执行。虽然 C 函数作者能够违背这一约定，但通常并不是好主意。

下一节给出了一个展示这些规则如何应用的示例。

40.5. 示例

本节给出一个非常简单的 SPI 用法示例。过程 `execq` 以一条 SQL 命令作为第一个参数、以行数计数作为第二个参数，使用 `SPI_exec` 执行该命令，并返回该命令实际处理的行数。更复杂的 SPI 示例可以在源码树的 `src/test/regress/regress.c` 以及 `contrib/spi` 中找到。

```
#include "postgres.h"

#include "executor/spi.h"
#include "utils/builtins.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

int execq(text *sql, int cnt);

int
execq(text *sql, int cnt)
{
    char *command;
    int ret;
    int proc;

    /* Convert given text object to a C string */
    command = text_to_cstring(sql);

    SPI_connect();

    ret = SPI_exec(command, cnt);

    proc = SPI_processed;
    /*
     * If this is a SELECT and some rows were fetched,
     * then the rows are printed via elog(INFO).
     */
    if (ret == SPI_OK_SELECT && SPI_processed > 0)
    {
        TupleDesc tupdesc = SPI_tuptable->tupdesc;
        SPITupleTable *tuptable = SPI_tuptable;
        char buf[8192];
        int i, j;

        for (j = 0; j < proc; j++)
        {
            HeapTuple tuple = tuptable->vals[j];

            for (i = 1, buf[0] = 0; i <= tupdesc->natts; i++)
                snprintf(buf + strlen (buf), sizeof(buf) - strlen(buf)
, " %s%s",
                         SPI_getvalue(tuple, tupdesc, i),
                         (i == tupdesc->natts) ? " " : " |");
            elog (INFO, "EXECQ: %s", buf);
        }
    }

    SPI_finish();
    pfree(command);

    return (proc);
```

```
}
```

（此函数使用版本 0 调用约定，以使示例更易于理解。在实际应用中，你应该使用新的版本 1 接口。）

把该函数编译成共享库之后，可以这样声明它：

```
CREATE FUNCTION execq(text, integer) RETURNS integer
    AS 'filename'
    LANGUAGE C;
```

以下是一个示例会话：

```
=> SELECT execq('CREATE TABLE a (x integer)', 0);
 execq
-------
     0
(1 row)

=> INSERT INTO a VALUES (execq('INSERT INTO a VALUES (0)', 0));
INSERT 0 1
=> SELECT execq('SELECT * FROM a', 0);
INFO:  EXECQ:  0    -- 由 execq 插入
INFO:  EXECQ:  1    -- 由 execq 返回并被外层 INSERT 插入

 execq
-------
     2
(1 row)

=> SELECT execq('INSERT INTO a SELECT x + 2 FROM a', 1);
 execq
-------
     1
(1 row)

=> SELECT execq('SELECT * FROM a', 10);
INFO:  EXECQ:  0
INFO:  EXECQ:  1
INFO:  EXECQ:  2    -- 0 + 2，按指定只插入一行

 execq
-------
     3              -- 10 只是最大值，3 才是真正的行数
(1 row)

=> DELETE FROM a;
DELETE 3
=> INSERT INTO a VALUES (execq('SELECT * FROM a', 0) + 1);
INSERT 0 1
=> SELECT * FROM a;
 x
---
 1                  -- a 中没有行 (0) + 1
(1 row)
```

```
=> INSERT INTO a VALUES (execq('SELECT * FROM a', 0) + 1);
INFO:  EXECQ:  1
INSERT 0 1
=> SELECT * FROM a;
 x
---
 1
 2                  -- a 中原有一行 + 1
(2 rows)

-- 下面演示数据更改可见性规则:

=> INSERT INTO a SELECT execq('SELECT * FROM a', 0) * x FROM a;
INFO:  EXECQ:  1
INFO:  EXECQ:  2
INFO:  EXECQ:  1
INFO:  EXECQ:  2
INFO:  EXECQ:  2
INSERT 0 2
=> SELECT * FROM a;
 x
---
 1
 2
 2                  -- 2 行 * 1 (第一行中的 x)
 6                  -- 3 行 (前面的 2 行加上刚插入的 1 行) * 2 (第二行中的
 x)
(4 rows)               ^^^^^^
                       不同调用中 execq() 可见的行
```

部分 VI. 参考

本参考中的各条目旨在以适当的篇幅，对各自的主题给出权威、完整且正式的总结。关于如何使用PostgreSQL的更多信息，例如叙述性说明、教程或示例，可在本书的其他部分找到。
另见各参考页列出的交叉引用。

这些参考条目也提供传统的“man”手册页形式。

目录

I. SQL 命令 643
ABORT 646
ALTER AGGREGATE 647
ALTER CONVERSION 649
ALTER DATABASE 650
ALTER DOMAIN 652
ALTER FUNCTION 655
ALTER GROUP 658
ALTER INDEX 659
ALTER LANGUAGE 661
ALTER OPERATOR 662
ALTER OPERATOR CLASS 663
ALTER ROLE 664
ALTER SCHEMA 667
ALTER SEQUENCE 668
ALTER TABLE 670
ALTER TABLESPACE 677
ALTER TRIGGER 678
ALTER TYPE 679
ALTER USER 680
ANALYZE 681
BEGIN 683
CHECKPOINT 685
CLOSE 686
CLUSTER 687
COMMENT 689
COMMIT 692
COMMIT PREPARED 693
COPY 694
CREATE AGGREGATE 702
CREATE CAST 705
CREATE CONSTRAINT TRIGGER 708
CREATE CONVERSION 709
CREATE DATABASE 711
CREATE DOMAIN 713
CREATE FUNCTION 715
CREATE GROUP 720
CREATE INDEX 721
CREATE LANGUAGE 724
CREATE OPERATOR 727
CREATE OPERATOR CLASS 730
CREATE ROLE 733
CREATE RULE 737
CREATE SCHEMA 739
CREATE SEQUENCE 741
CREATE TABLE 744
CREATE TABLE AS 753
CREATE TABLESPACE 755
CREATE TRIGGER 757
CREATE TYPE 760
CREATE USER 765

CREATE VIEW 766
DEALLOCATE 768
DECLARE 769
DELETE 772
DROP AGGREGATE 774
DROP CAST 775
DROP CONVERSION 776
DROP DATABASE 777
DROP DOMAIN 778
DROP FUNCTION 779
DROP GROUP 781
DROP INDEX 782
DROP LANGUAGE 783
DROP OPERATOR 784
DROP OPERATOR CLASS 786
DROP ROLE 787
DROP RULE 788
DROP SCHEMA 789
DROP SEQUENCE 790
DROP TABLE 791
DROP TABLESPACE 792
DROP TRIGGER 793
DROP TYPE 794
DROP USER 795
DROP VIEW 796
END 797
EXECUTE 798
EXPLAIN 799
FETCH 802
GRANT 806
INSERT 811
LISTEN 813
LOAD 814
LOCK 815
MOVE 817
NOTIFY 818
PREPARE 820
PREPARE TRANSACTION 822
REINDEX 824
RELEASE SAVEPOINT 827
RESET 828
REVOKE 829
ROLLBACK 832
ROLLBACK PREPARED 833
ROLLBACK TO SAVEPOINT 834
SAVEPOINT 836
SELECT 838
SELECT INTO 849
SET 851
SET CONSTRAINTS 854
SET ROLE 855
SET SESSION AUTHORIZATION 857
SET TRANSACTION 859
SHOW 861

START TRANSACTION 863
TRUNCATE 864
UNLISTEN 865
UPDATE 866
VACUUM 869
II. PostgreSQL 客户端应用 872
clusterdb 873
createdb 875
createlang 878
createuser 880
dropdb 884
droplang 886
dropuser 888
ecpg 890
pg_config 892
pg_dump 895
pg_dumpall 901
pg_restore 905
psql 911
reindexdb 935
vacuumdb 938
III. PostgreSQL 服务器端应用 941
initdb 942
ipcclean 945
pg_controldata 946
pg_ctl 947
pg_resetxlog 951
postgres 953
postmaster 957

SQL 命令

本部分包含PostgreSQL支持的SQL命令的参考信息。这里的 “SQL” 指的是一般意义上的该语言；每条命令的标准符合性和兼容性信息可在相应的参考页中找到。

目录

ABORT 646
ALTER AGGREGATE 647
ALTER CONVERSION 649
ALTER DATABASE 650
ALTER DOMAIN 652
ALTER FUNCTION 655
ALTER GROUP 658
ALTER INDEX 659
ALTER LANGUAGE 661
ALTER OPERATOR 662
ALTER OPERATOR CLASS 663
ALTER ROLE 664
ALTER SCHEMA 667
ALTER SEQUENCE 668
ALTER TABLE 670
ALTER TABLESPACE 677
ALTER TRIGGER 678
ALTER TYPE 679
ALTER USER 680
ANALYZE 681
BEGIN 683
CHECKPOINT 685
CLOSE 686
CLUSTER 687
COMMENT 689
COMMIT 692
COMMIT PREPARED 693
COPY 694
CREATE AGGREGATE 702
CREATE CAST 705
CREATE CONSTRAINT TRIGGER 708
CREATE CONVERSION 709
CREATE DATABASE 711
CREATE DOMAIN 713
CREATE FUNCTION 715
CREATE GROUP 720
CREATE INDEX 721
CREATE LANGUAGE 724
CREATE OPERATOR 727
CREATE OPERATOR CLASS 730
CREATE ROLE 733
CREATE RULE 737
CREATE SCHEMA 739
CREATE SEQUENCE 741

CREATE TABLE 744
CREATE TABLE AS 753
CREATE TABLESPACE 755
CREATE TRIGGER 757
CREATE TYPE 760
CREATE USER 765
CREATE VIEW 766
DEALLOCATE 768
DECLARE 769
DELETE 772
DROP AGGREGATE 774
DROP CAST 775
DROP CONVERSION 776
DROP DATABASE 777
DROP DOMAIN 778
DROP FUNCTION 779
DROP GROUP 781
DROP INDEX 782
DROP LANGUAGE 783
DROP OPERATOR 784
DROP OPERATOR CLASS 786
DROP ROLE 787
DROP RULE 788
DROP SCHEMA 789
DROP SEQUENCE 790
DROP TABLE 791
DROP TABLESPACE 792
DROP TRIGGER 793
DROP TYPE 794
DROP USER 795
DROP VIEW 796
END 797
EXECUTE 798
EXPLAIN 799
FETCH 802
GRANT 806
INSERT 811
LISTEN 813
LOAD 814
LOCK 815
MOVE 817
NOTIFY 818
PREPARE 820
PREPARE TRANSACTION 822
REINDEX 824
RELEASE SAVEPOINT 827
RESET 828
REVOKE 829
ROLLBACK 832
ROLLBACK PREPARED 833
ROLLBACK TO SAVEPOINT 834
SAVEPOINT 836
SELECT 838
SELECT INTO 849

SET 851
SET CONSTRAINTS 854
SET ROLE 855
SET SESSION AUTHORIZATION 857
SET TRANSACTION 859
SHOW 861
START TRANSACTION 863
TRUNCATE 864
UNLISTEN 865
UPDATE 866
VACUUM 869

ABORT

ABORT — 中止当前事务

大纲

```
ABORT [ WORK | TRANSACTION ]
```

描述

`ABORT`回滚当前事务，并丢弃该事务所做的全部更新。这个命令在行为上与标准 SQL 命令ROLLBACK完全相同，只是出于历史原因而保留。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用COMMIT可成功地终止一个事务。

在事务块之外发出`ABORT`会产生一条警告，除此之外没有效果。

示例

要中止所有更改：

```
ABORT;
```

兼容性

这个命令是PostgreSQL扩展，只是出于历史原因而保留。`ROLLBACK`是与之等效的标准 SQL 命令。

另见

BEGIN, COMMIT, ROLLBACK

ALTER AGGREGATE

ALTER AGGREGATE — 更改一个聚合函数的定义

大纲

```
ALTER AGGREGATE name ( type ) RENAME TO new_name
ALTER AGGREGATE name ( type ) OWNER TO new_owner
ALTER AGGREGATE name ( type ) SET SCHEMA new_schema
```

描述

`ALTER AGGREGATE`更改聚合函数的定义。

要使用`ALTER AGGREGATE`，必须拥有该聚合函数。要更改聚合函数的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在聚合函数所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建聚合函数完成的操作。不过，超级用户无论如何都可以更改任何聚合函数的所有权。）

参数

name

现有聚合函数的名称（可以用模式限定）。

type

聚合函数的参数数据类型，如果函数接受任意数据类型，则为`*`。

new_name

聚合函数的新名称。

new_owner

聚合函数的新拥有者。

new_schema

聚合函数的新模式。

示例

要把用于类型`integer`的聚合函数 `myavg`重命名为`my_average`：

```
ALTER AGGREGATE myavg(integer) RENAME TO my_average;
```

要把用于类型`integer`的聚合函数 `myavg`的拥有者改为`joe`：

```
ALTER AGGREGATE myavg(integer) OWNER TO joe;
```

要把用于`float8`的聚合`myavg`移动到模式`myschema`中：

```
ALTER AGGREGATE myavg(float8) SET SCHEMA myschema;
```

兼容性

在 SQL 标准中没有`ALTER AGGREGATE`语句。

另见

CREATE AGGREGATE, DROP AGGREGATE

ALTER CONVERSION

ALTER CONVERSION — 更改一个转换的定义

大纲

```
ALTER CONVERSION name RENAME TO newname
ALTER CONVERSION name OWNER TO newowner
```

描述

`ALTER CONVERSION`更改一个转换的定义。

要使用`ALTER CONVERSION`，必须拥有该转换。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在转换所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建转换完成的操作。不过，超级用户无论如何都可以更改任何转换的所有权。）

参数

name

一个现有转换的名称（可以是模式限定的）。

newname

转换的新名称。

newowner

转换的新拥有者。

示例

要把转换`iso_8859_1_to_utf8`重命名为`latin1_to_unicode`：

```
ALTER CONVERSION iso_8859_1_to_utf8 RENAME TO latin1_to_unicode;
```

要把转换`iso_8859_1_to_utf8`的拥有者改成`joe`：

```
ALTER CONVERSION iso_8859_1_to_utf8 OWNER TO joe;
```

兼容性

在 SQL 标准中没有`ALTER CONVERSION`语句。

另见

CREATE CONVERSION, DROP CONVERSION

ALTER DATABASE

ALTER DATABASE — 更改一个数据库

大纲

```
ALTER DATABASE name [ [ WITH ] option [ ... ] ]

where option can be:

    CONNECTION LIMIT connlimit

ALTER DATABASE name SET parameter { TO | = } { value | DEFAULT }
ALTER DATABASE name RESET parameter

ALTER DATABASE name RENAME TO newname

ALTER DATABASE name OWNER TO new_owner
```

描述

`ALTER DATABASE`更改一个数据库的属性。

第一种形式更改某些数据库级设置（详见下文）。
只有数据库拥有者或者超级用户才能更改这些设置。

第二种和第三种形式会更改某个PostgreSQL数据库的运行时配置变量的会话默认值。
此后每当在该数据库中启动一个新会话时，指定的值就会成为会话默认值。
数据库特定的默认值会覆盖`postgresql.conf`中的设置，或者从`postmaster`
命令行接收到的设置。只有数据库拥有者或者超级用户才能更改该数据库的会话默认值。
某些变量不能以这种方式设置，或者只能由超级用户设置。

第四种形式更改数据库的名称。只有数据库拥有者或者超级用户可以重命名数据库；
非超级用户拥有者还必须具有`CREATEDB`权限。
当前连接的数据库不能被重命名（如果需要这样做，请连接到另一个数据库）。

第五种形式更改数据库的所有者。要更改所有者，必须拥有该数据库，
并且还必须是新所有者角色的直接或间接成员，并且必须拥有`CREATEDB`权限。（请注意，
超级用户会自动拥有所有这些权限。）

参数

name

要修改其属性的数据库名称。

connlimit

这个数据库允许多少并发连接。-1 表示没有限制。

parameter
value

将此数据库在指定参数上的会话默认值设为给定值。如果 *value*是DEFAULT，或者等效地使用了 RESET，则数据库特定设置会被移除，因此新会话将继承系统范围的默认设置。使用RESET ALL可清除所有数据库特定设置。

关于允许的参数名和值的更多信息，见SET和第 17 章。

newname

数据库的新名称。

new_owner

数据库的新拥有者。

注解

也可以把会话默认值绑定到特定用户，而不是数据库；见ALTER USER。如果发生冲突，用户特定设置会覆盖数据库特定设置。

示例

要在数据库test中默认禁用索引扫描：

```
ALTER DATABASE test SET enable_indexscan TO off;
```

兼容性

ALTER DATABASE语句是一个PostgreSQL扩展。

参见

CREATE DATABASE, DROP DATABASE, SET

ALTER DOMAIN

ALTER DOMAIN — 更改一个域的定义

大纲

```
ALTER DOMAIN name
    { SET DEFAULT expression | DROP DEFAULT }
ALTER DOMAIN name
    { SET | DROP } NOT NULL
ALTER DOMAIN name
    ADD domain_constraint
ALTER DOMAIN name
    DROP CONSTRAINT constraint_name [ RESTRICT | CASCADE ]
ALTER DOMAIN name
    OWNER TO new_owner
ALTER DOMAIN name
    SET SCHEMA new_schema
```

描述

`ALTER DOMAIN`更改一个现有域的定义。有以下几种子形式：

SET/DROP DEFAULT

这些形式为域设置或移除默认值。注意，默认值只适用于后续的`INSERT`命令；它们不会影响已存在于使用该域的表中的行。

SET/DROP NOT NULL

这些形式更改域是被标记为允许 NULL 值还是拒绝 NULL 值。只有当使用该域的列中不包含空值时，才能执行`SET NOT NULL`。

ADD *domain_constraint*

这种形式使用与CREATE DOMAIN相同的语法向域添加一个新约束。只有当所有使用该域的列都满足新约束时，这一操作才会成功。

DROP CONSTRAINT

这种形式删除域上的约束。

OWNER

这种形式将域的拥有者更改为指定用户。

SET SCHEMA

这种形式更改域所在的模式。与该域关联的任何约束也会一并移动到新模式中。

要使用`ALTER DOMAIN`，必须拥有该域。要更改域的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在域所在模式上拥有`CREATE`权限。

（这些限制确保更改所有者不会执行任何无法通过删除并重新创建域完成的操作。不过，超级用户无论如何都可以更改任何域的所有权。）

参数

name

要修改的现有域的名称（可以是模式限定的）。

domain_constraint

该域的新约束。

constraint_name

要删除的现有约束的名称。

CASCADE

自动删除依赖于该约束的对象。

RESTRICT

如果存在任何依赖对象，则拒绝删除该约束。这是默认行为。

new_owner

域的新拥有者的用户名。

new_schema

域的新模式。

示例

向域添加一个NOT NULL约束：

```
ALTER DOMAIN zipcode SET NOT NULL;
```

从域中移除一个NOT NULL约束：

```
ALTER DOMAIN zipcode DROP NOT NULL;
```

向域添加一个检查约束：

```
ALTER DOMAIN zipcode ADD CONSTRAINT zipchk CHECK (char_length(VALUE) =
 5);
```

从域中移除一个检查约束：

```
ALTER DOMAIN zipcode DROP CONSTRAINT zipchk;
```

将域移动到另一个模式中：

```
ALTER DOMAIN zipcode SET SCHEMA customers;
```

兼容性

`ALTER DOMAIN` 符合 SQL 标准，但 `OWNER` 和 `SET SCHEMA` 变体是 PostgreSQL 扩展。

另见

CREATE DOMAIN, DROP DOMAIN

ALTER FUNCTION

ALTER FUNCTION — 更改函数的定义

大纲

```
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    action [, ... ] [ RESTRICT ]
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    RENAME TO new_name
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    OWNER TO new_owner
ALTER FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    SET SCHEMA new_schema

where action is one of:

    CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
    IMMUTABLE | STABLE | VOLATILE
    [ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER
```

描述

`ALTER FUNCTION`更改函数的定义。

要使用`ALTER FUNCTION`，必须拥有该函数。要更改函数的模式，还必须在新模式上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，
并且该角色必须在函数所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建函数完成的操作。不过，
超级用户无论如何都可以更改任何函数的所有权。）

参数

name

现有函数的名称（可以用模式限定）。

argmode

参数的模式：`IN`、`OUT`或 `INOUT`。如果省略，默认为`IN`。注意`ALTER FUNCTION` 实际上并不关心任何`OUT`参数，因为确定函数身份只需要输入参数。因此，只列出`IN`和 `INOUT`参数就足够了。

argname

参数的名称。注意`ALTER FUNCTION`实际上并不关心参数名，因为确定函数身份只需要参数数据类型。

argtype

函数参数的数据类型（如果有，可以用模式限定）。

new_name

该函数的新名称。

new_owner

该函数的新拥有者。注意，如果该函数被标记为 `SECURITY DEFINER`，随后它将以新拥有者的身份执行。

new_schema

该函数的新模式。

`CALLED ON NULL INPUT`
`RETURNS NULL ON NULL INPUT`
`STRICT`

`CALLED ON NULL INPUT`将该函数改为在其部分或全部参数为 null 时仍会被调用。`RETURNS NULL ON NULL INPUT`或 `STRICT`将该函数改为只要任一参数为 null 就不调用，而是自动假定结果为 null。详见CREATE FUNCTION。

`IMMUTABLE`
`STABLE`
`VOLATILE`

将该函数的易变性更改为指定的设置。详见CREATE FUNCTION。

`[ EXTERNAL ] SECURITY INVOKER`
`[ EXTERNAL ] SECURITY DEFINER`

更改该函数是否为安全性定义者。关键字`EXTERNAL` 为了符合 SQL 标准而被忽略。关于这一能力的更多信息，参见CREATE FUNCTION。

`RESTRICT`

为符合 SQL 标准而被忽略。

示例

要将类型`integer`的函数`sqrt` 重命名为`square_root`：

```
ALTER FUNCTION sqrt(integer) RENAME TO square_root;
```

要将类型`integer`的函数`sqrt` 的拥有者改为`joe`：

```
ALTER FUNCTION sqrt(integer) OWNER TO joe;
```

要将类型`integer`的函数`sqrt` 的模式改为`maths`：

```
ALTER FUNCTION sqrt(integer) SET SCHEMA maths;
```

兼容性

该语句与 SQL 标准中的`ALTER FUNCTION`语句部分兼容。标准允许修改函数的更多属性，但不提供重命名函数、将函数设为安全性定义者，或者更改函数的拥有者、模式或易变性的能力。标准还要求使用`RESTRICT`关键字，而它在 PostgreSQL中是可选的。

另见

CREATE FUNCTION, DROP FUNCTION

ALTER GROUP

ALTER GROUP — 更改角色名称或成员资格

大纲

```
ALTER GROUP groupname ADD USER username [, ... ]
ALTER GROUP groupname DROP USER username [, ... ]

ALTER GROUP groupname RENAME TO newname
```

描述

`ALTER GROUP`更改用户组的属性。这是一个已废弃的命令，但出于向后兼容的考虑仍然接受它，因为组（以及用户）已经被更一般的角色概念所取代。

前两个变体将用户加入一个组，或将其从组中移除。（为此目的，任何角色都可以充当“用户”或“组”。）这些变体实际上等效于对名为“组”的角色授予或撤销成员资格，因此更推荐使用GRANT或REVOKE。

第三个变体更改组的名称。这完全等同于使用ALTER ROLE重命名角色。

参数

groupname

要修改的组（角色）的名称。

username

要加入该组或从该组中移除的用户（角色）。这些用户必须已经存在；`ALTER GROUP`不会创建或删除用户。

newname

该组的新名称。

示例

向组中添加用户：

```
ALTER GROUP staff ADD USER karl, john;
```

从组中移除用户：

```
ALTER GROUP workers DROP USER beth;
```

兼容性

SQL 标准中没有`ALTER GROUP`语句。

另见

GRANT, REVOKE, ALTER ROLE

ALTER INDEX

ALTER INDEX — 更改索引的定义

大纲

```
ALTER INDEX name RENAME TO new_name
ALTER INDEX name SET TABLESPACE tablespace_name
```

描述

`ALTER INDEX`更改现有索引的定义。有几种子形式：

`RENAME`

`RENAME`形式更改索引的名称。存储的数据不受影响。

`SET TABLESPACE`

这种形式把索引的表空间更改为指定的表空间，
并把与该索引关联的数据文件移动到新表空间。另见CREATE TABLESPACE。

参数

name

要修改的现有索引名称（可以是模式限定的）。

new_name

该索引的新名称。

tablespace_name

该索引将被移动到的表空间。

注解

这些操作也可以使用ALTER TABLE完成。实际上，`ALTER INDEX`只是`ALTER TABLE`中适用于索引的那些形式的别名。

以前曾有一种`ALTER INDEX OWNER`变体，但现在会被忽略（同时发出警告）。索引的拥有者不能不同于其所属表的拥有者。更改表的拥有者时，索引的拥有者也会自动更改。

不允许更改系统目录索引的任何部分。

示例

要重命名现有索引：

```
ALTER INDEX distributors RENAME TO suppliers;
```

要将索引移到另一个表空间：

```
ALTER INDEX distributors SET TABLESPACE fasttablespace;
```

兼容性

`ALTER INDEX`是一种 PostgreSQL扩展。

ALTER LANGUAGE

ALTER LANGUAGE — 更改一种过程语言的定义

大纲

`ALTER LANGUAGE` *`name`* `RENAME TO` *`newname`*

描述

`ALTER LANGUAGE`更改一种语言的定义。唯一的功能是重命名该语言。只有超级用户才能重命名语言。

参数

`name`

语言的名称

`newname`

该语言的新名称

兼容性

SQL 标准中没有`ALTER LANGUAGE`语句。

另见

CREATE LANGUAGE, DROP LANGUAGE

ALTER OPERATOR

ALTER OPERATOR — 更改一个操作符的定义

大纲

```
ALTER OPERATOR name ( { lefttype | NONE } , { righttype | NONE } )
 OWNER TO newowner
```

描述

`ALTER OPERATOR`更改一个操作符的定义。目前唯一可用的功能是更改操作符的拥有者。

要使用`ALTER OPERATOR`，必须拥有该操作符。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在操作符所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建操作符完成的操作。不过，超级用户无论如何都可以更改任何操作符的所有权。）

参数

name

现有操作符的名称（可带模式限定）。

lefttype

操作符左操作数的数据类型；如果操作符没有左操作数，写作`NONE`。

righttype

操作符右操作数的数据类型；如果操作符没有右操作数，写作`NONE`。

newowner

操作符的新所有者。

示例

更改自定义操作符的限制和连接选择率估计函数，操作符为`a && b`，类型为`int[]`：

```
ALTER OPERATOR && (_int4, _int4) SET (RESTRICT = _int_contsel, JOIN =
 _int_contjoinsel);
```

兼容性

在 SQL 标准中没有`ALTER OPERATOR`语句。

另见

CREATE OPERATOR, DROP OPERATOR

ALTER OPERATOR CLASS

ALTER OPERATOR CLASS — 更改一个操作符类的定义

大纲

```
ALTER OPERATOR CLASS name USING index_method RENAME TO newname
ALTER OPERATOR CLASS name USING index_method OWNER TO newowner
```

描述

`ALTER OPERATOR CLASS`更改一个操作符类的定义。

要使用`ALTER OPERATOR CLASS`，必须拥有该操作符类。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在操作符类所在模式上拥有`CREATE`权限。
（这些限制确保更改所有者不会执行任何无法通过删除并重新创建操作符类完成的操作。不过，超级用户无论如何都可以更改任何操作符类的所有权。）

参数

name

一个现有操作符类的名称（可以是模式限定的）。

index_method

该操作符类所对应的索引方法的名称。

newname

该操作符类的新名称。

newowner

该操作符类的新拥有者。

兼容性

SQL 标准中没有`ALTER OPERATOR CLASS`语句。

另见

CREATE OPERATOR CLASS, DROP OPERATOR CLASS

ALTER ROLE

ALTER ROLE — 更改一个数据库角色

大纲

```
ALTER ROLE name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'

ALTER ROLE name RENAME TO newname

ALTER ROLE name SET parameter { TO | = } { value | DEFAULT }
ALTER ROLE name RESET parameter
```

描述

`ALTER ROLE`更改 PostgreSQL角色的属性。

概要中列出的此命令的第一种形式可以更改CREATE ROLE中可指定的许多角色属性。（涵盖了所有可能的属性，但没有用于添加或移除成员资格的选项；请使用GRANT和REVOKE完成此操作。）命令中未提及的属性保留其以前的设置。数据库超级用户可以更改任何角色的任何这些设置。拥有`CREATEROLE`权限的角色可以更改任何这些设置，但只能针对非超级用户角色。普通角色只能更改自己的密码。

第二种形式更改角色的名称。数据库超级用户可以重命名任何角色。拥有`CREATEROLE`权限的角色可以重命名非超级用户角色。当前会话用户不能被重命名。（如果需要这样做，请以其他用户连接。）由于`MD5`加密的密码使用角色名作为密码学盐，重命名角色会清除其密码（如果密码使用`MD5`加密）。

第三和第四种形式更改角色的某个配置变量的会话默认值。每当该角色随后启动新会话时，指定的值就成为会话默认值，覆盖`postgresql.conf`中的设置或从`postmaster`命令行接收的设置。（对于没有`LOGIN`权限的角色，会话默认值不起作用。）普通角色可以更改自己的会话默认值。超级用户可以更改任何人的会话默认值。具有`CREATEROLE`权限的角色可以更改非超级用户角色的默认值。某些配置变量不能用这种方式设置，或者只有超级用户发出命令时才能设置。

参数

name

将要修改其属性的角色名称。

SUPERUSER
NOSUPERUSER
CREATEDB
NOCREATEDB
CREATEROLE
NOCREATEROLE
CREATEUSER
NOCREATEUSER
INHERIT
NOINHERIT
LOGIN
NOLOGIN
CONNECTION LIMIT *connlimit*
PASSWORD *password*
ENCRYPTED
UNENCRYPTED
VALID UNTIL '*timestamp*'

这些子句更改最初由CREATE ROLE设置的属性，更多信息请参见该参考页。

newname

角色的新名称。

parameter
value

将该角色在指定配置参数上的会话默认值设为给定值。如果 *value*是DEFAULT，或者等效地使用了RESET，则会移除角色特定变量设置，
因此该角色在新会话中将继承系统范围的默认设置。使用 RESET ALL可清除所有角色特定设置。

关于允许的参数名称和值详见SET和第 17 章。

注解

使用CREATE ROLE添加新角色，使用DROP ROLE移除角色。

ALTER ROLE不能更改角色的成员资格。请使用GRANT和REVOKE完成此操作。

也可以把会话默认值绑定到特定数据库，而不是角色；详见ALTER DATABASE。如果发生冲突，角色特定设置会覆盖数据库特定设置。

示例

更改角色的密码：

```
ALTER ROLE davide WITH PASSWORD 'hu8jmn3';
```

更改密码的失效日期，并指定该密码应在比 UTC快 1 小时的时区中，于 2015 年 5 月 4 日中午失效：

```
ALTER ROLE chris VALID UNTIL 'May 4 12:00:00 2015 +1';
```

使密码永久有效：

```
ALTER ROLE fred VALID UNTIL 'infinity';
```

授予角色创建其他角色和新数据库的能力：

```
ALTER ROLE miriam CREATEROLE CREATEDB;
```

为角色指定maintenance_work_mem参数的非默认设置：

```
ALTER ROLE worker_bee SET maintenance_work_mem = 100000;
```

兼容性

`ALTER ROLE`语句是 PostgreSQL的扩展。

另见

CREATE ROLE, DROP ROLE, SET

ALTER SCHEMA

ALTER SCHEMA — 更改一个模式的定义

大纲

```
ALTER SCHEMA name RENAME TO newname
ALTER SCHEMA name OWNER TO newowner
```

描述

`ALTER SCHEMA`更改一个模式的定义。

要使用`ALTER SCHEMA`，必须拥有该模式。要重命名模式，还必须在数据库上拥有`CREATE`权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且必须在数据库上拥有`CREATE`权限。（请注意，超级用户会自动拥有所有这些权限。）

参数

name

现有模式的名称。

newname

该模式的新名称。新名称不能以`pg_`开头，因为这类名称是为系统模式保留的。

newowner

该模式的新拥有者。

兼容性

SQL 标准中没有`ALTER SCHEMA`语句。

另见

CREATE SCHEMA, DROP SCHEMA

ALTER SEQUENCE

ALTER SEQUENCE — 更改序列发生器的定义

大纲

```
ALTER SEQUENCE name [ INCREMENT [ BY ] increment ]
    [ MINVALUE minvalue | NO MINVALUE ] [ MAXVALUE maxvalue | NO
 MAXVALUE ]
    [ RESTART [ WITH ] start ] [ CACHE cache ] [ [ NO ] CYCLE ]
ALTER SEQUENCE name SET SCHEMA new_schema
```

描述

`ALTER SEQUENCE`更改现有序列发生器的参数。在 `ALTER SEQUENCE`命令中未明确设置的参数，会保留其原有设置。

要使用`ALTER SEQUENCE`，必须拥有该序列。要更改序列的模式，还必须在新模式上拥有`CREATE`权限。

参数

name

要更改的序列名称（可以是模式限定的）。

increment

子句`INCREMENT BY` *increment*是可选的。正值会创建升序序列，负值会创建降序序列。如果未指定，则保留原有的增量值。

minvalue
`NO MINVALUE`

可选子句`MINVALUE` *minvalue*确定序列可生成的最小值。如果指定了`NO MINVALUE`，则升序序列和降序序列将分别使用默认值 1 和 $-2^{63}-1$。如果两种选项都未指定，则保留当前最小值。

maxvalue
`NO MAXVALUE`

可选子句`MAXVALUE` *maxvalue*确定序列的最大值。如果指定了`NO MAXVALUE`，则升序序列和降序序列将分别使用默认值 $2^{63}-1$ 和 -1。如果两种选项都未指定，则保留当前最大值。

start

可选子句`RESTART WITH` *start*更改序列的当前值。

cache

`CACHE` *cache*子句启用预分配序列号并将其存储在内存中，以便更快访问。最小值为 1（一次只能生成一个值，即不缓存）。如果未指定，则保留旧的缓存值。

`CYCLE`

可选的`CYCLE`关键字可用于启用序列回绕：当递增或递减序列分别达到*maxvalue*或*minvalue*时回绕。如果达到限制，下一个生成的数将分别为*minvalue*或*maxvalue*。

`NO CYCLE`

如果指定了可选关键字`NO CYCLE`，则当序列达到其最大值后，任何对 `nextval`的调用都会返回错误。如果既未指定`CYCLE`也未指定 `NO CYCLE`，则保留原有的循环行为。

new_schema

该序列的新模式。

示例

将一个名为`serial`的序列重启到 105：

```
ALTER SEQUENCE serial RESTART WITH 105;
```

注解

为了避免阻塞从同一序列获取数字的并发事务，`ALTER SEQUENCE`永远不会被回滚；更改会立即生效且不可逆转。

`ALTER SEQUENCE`不会立即影响除当前后端之外、已经预分配（缓存）了序列值的其他后端中的 `nextval`结果。它们会先用完所有缓存值，然后才会注意到序列参数已经改变。当前后端则会立即受到影响。

`ALTER TABLE`的一些变体也可以用于序列；例如，要重命名序列，可使用`ALTER TABLE RENAME`。

兼容性

`ALTER SEQUENCE`符合SQL标准，但 `SET SCHEMA`变体是PostgreSQL扩展。

ALTER TABLE

ALTER TABLE — 更改一个表的定义

大纲

```
ALTER TABLE [ ONLY ] name [ * ]
    action [, ... ]
ALTER TABLE [ ONLY ] name [ * ]
    RENAME [ COLUMN ] column TO new_column
ALTER TABLE name
    RENAME TO new_name
ALTER TABLE name
    SET SCHEMA new_schema

where action is one of:

    ADD [ COLUMN ] column type [ column_constraint [ ... ] ]
    DROP [ COLUMN ] column [ RESTRICT | CASCADE ]
    ALTER [ COLUMN ] column TYPE type [ USING expression ]
    ALTER [ COLUMN ] column SET DEFAULT expression
    ALTER [ COLUMN ] column DROP DEFAULT
    ALTER [ COLUMN ] column { SET | DROP } NOT NULL
    ALTER [ COLUMN ] column SET STATISTICS integer
    ALTER [ COLUMN ] column SET STORAGE { PLAIN | EXTERNAL | EXTENDED
 | MAIN }
    ADD table_constraint
    DROP CONSTRAINT constraint_name [ RESTRICT | CASCADE ]
    DISABLE TRIGGER [ trigger_name | ALL | USER ]
    ENABLE TRIGGER [ trigger_name | ALL | USER ]
    CLUSTER ON index_name
    SET WITHOUT CLUSTER
    SET WITHOUT OIDS
    OWNER TO new_owner
    SET TABLESPACE new_tablespace
```

描述

`ALTER TABLE`更改一个现有表的定义。它有若干种子形式：

`ADD COLUMN`

这种形式使用与CREATE TABLE相同的语法向表添加一个新列。

`DROP COLUMN`

这种形式从表中删除一个列。涉及该列的索引和表约束也将被自动删除。
如果表外的任何对象依赖于该列（例如外键引用或视图），则需要指定 `CASCADE`。

`ALTER COLUMN TYPE`

这种形式更改表中某一列的类型。
涉及该列的索引和简单表约束会通过重新解析最初提供的表达式，

自动转换为使用新的列类型。可选的USING子句指定如何根据旧值计算新列值；如果省略，默认转换与从旧数据类型到新数据类型的赋值转换相同。
如果旧类型到新类型之间不存在隐式转换或赋值转换，则必须提供USING子句。

SET/DROP DEFAULT

这些形式为列设置或移除默认值。默认值只会应用于后续的 INSERT 命令；它们不会导致表中已有的行发生变化。也可以为视图创建默认值，这种情况下，在应用视图的 ON INSERT 规则之前，默认值会被插入到视图上的 INSERT 语句中。

SET/DROP NOT NULL

这些形式更改列是否被标记为允许空值，或拒绝空值。只有当列不包含空值时，才能使用SET NOT NULL。

SET STATISTICS

该形式为后续ANALYZE操作设置每列的统计信息收集目标。目标可以设置在 0 到 1000 范围内；也可以将其设置为 -1，以恢复使用系统默认的统计目标（default_statistics_target）。
有关PostgreSQL查询规划器使用统计信息的更多信息，请参见第 13.2 节。

SET STORAGE

该形式为列设置存储模式。这控制该列是内联保存还是保存在补充表中，
以及是否压缩数据。对于integer等定长值，必须使用PLAIN，
数据以内联且未压缩的形式保存。MAIN用于内联的可压缩数据。EXTERNAL
用于外部保存的未压缩数据，EXTENDED用于外部保存的压缩数据。对于支持非PLAIN
存储的大多数数据类型，EXTENDED是默认值。使用EXTERNAL会使text和bytea
列上的子字符串操作更快，但会增加存储空间。请注意，SET STORAGE本身不会更改表中的任何内容，它只会设置今后更新表时采用的策略。更多信息请参见第 50.2 节。

ADD *table_constraint*

这种形式使用与CREATE TABLE相同的语法向一个表增加一个新约束。

DROP CONSTRAINT

该形式从表中删除指定的约束。

DISABLE/ENABLE TRIGGER

这些形式禁用或启用属于表的触发器。已禁用的触发器仍为系统所知，
但在触发事件发生时不会执行。对于延迟触发器，会在事件发生时检查启用状态，
而不是在实际执行触发器函数时检查。可以按名称指定单个触发器，
或指定表上的所有触发器，或仅指定用户触发器，
从而禁用或启用触发器（最后一种选项不包括内部生成的约束触发器，
例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器）。
禁用或启用内部生成的约束触发器需要超级用户权限；应谨慎操作，
因为如果不执行触发器，当然无法保证约束的完整性。

CLUSTER

该形式为将来的CLUSTER操作选择默认索引。它实际上不会对表重新聚簇。

SET WITHOUT CLUSTER

该形式从表中移除最近使用的CLUSTER索引规范。
这会影响将来的聚簇操作（这些操作未指定索引）。

`SET WITHOUT OIDS`

该形式从表中移除oid系统列。这完全等同于`DROP COLUMN oid RESTRICT`，但如果表中已经没有oid列，则不会报错。

注意，没有任何一种`ALTER TABLE`的变体允许在 OID 被移除之后重新为表恢复它们。

`OWNER`

该形式把表、序列或视图的所有者更改为指定用户。

`SET TABLESPACE`

这种形式把表的表空间更改为指定的表空间，并把与该表关联的数据文件移动到新表空间。表上的索引（如果有）不会被移动；但可以用额外的 `SET TABLESPACE` 命令单独移动。另见CREATE TABLESPACE。

`RENAME`

`RENAME` 形式更改表（或索引、序列或视图）的名称，或者更改表中某一列的名称。存储的数据不受影响。

`SET SCHEMA`

该形式把表移动到另一个模式。相关索引、约束以及 SERIAL 列的序列也会一并移动。

对单个表执行操作的 ALTER TABLE 所有形式，除了`RENAME`和`SET SCHEMA`之外，都可以组合成一个列表，一起应用多项更改。例如，可以在一条命令中添加多列和/或更改多列的类型。这对大表尤其有用，因为只需对表执行一次遍历。

要使用 `ALTER TABLE`，必须拥有该表。要更改表的模式，还必须在新模式上拥有 `CREATE` 权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在表所在模式上拥有 `CREATE` 权限。（这些限制确保更改所有者不会执行任何无法通过删除并重新创建该表完成的操作。不过，超级用户无论如何都可以更改任何表的所有权。）

参数

name

要更改的现有表的名称（可带模式限定）。如果指定了`ONLY`，则只更改该表。如果未指定`ONLY`，则更改该表及其所有后代表（如果有）。可以在表名后附加`*`来表示要更改后代表，但在当前版本中，这是默认行为。（在 7.1 之前的版本中，`ONLY`是默认行为。可以通过更改配置参数sql_inheritance来改变默认行为。）

column

新列或现有列的名称。

new_column

现有列的新名称。

new_name

表的新名称。

type

新列的数据类型，或现有列的新数据类型。

table_constraint

表的新约束。

constraint_name

新约束或现有约束的名称。

CASCADE

自动删除依赖于被删除列或约束的对象（例如引用该列的视图）。

RESTRICT

如果存在任何依赖对象，则拒绝删除列或约束。这是默认行为。

trigger_name

要禁用或启用的单个触发器的名称。

ALL

禁用或启用属于表的所有触发器。（如果触发器中有内部生成的约束触发器，
例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器，则需要超级用户权限。）

USER

禁用或启用属于表的所有触发器，但内部生成的约束触发器除外，
例如用于实现外键约束或可延迟唯一性约束和排他约束的触发器。

index_name

现有索引的名称。

new_owner

表的新所有者的用户名。

new_tablespace

表将被移动到的表空间名称。

new_schema

表将被移动到的模式名称。

注意

关键字COLUMN只是噪声，可以省略。

使用ADD COLUMN添加列时，表中的所有现有行都会使用该列的默认值初始化（如果未指定DEFAULT子句，则使用 NULL）。如果没有 DEFAULT 子句，这只是一次元数据更改，不需要立即更新表数据；添加的 NULL 值会在读取时提供。

增加一个带非空默认值的列或更改一个现有列的类型将要求重写整个表及其索引。
这对于大型表可能需要相当长的时间，并且会暂时需要两倍的磁盘空间。

添加CHECK或NOT NULL约束需要扫描表，以验证现有行满足约束，但不需要重写表。

允许在单条ALTER TABLE命令中指定多项更改，主要是因为这样可以将多次表扫描或重写合并为一次遍历。

DROP COLUMN形式不会从物理上删除列，而只是使其对 SQL 操作不可见。表后续的插入和更新操作会为该列存储空值。因此，删除列的速度很快，但不会立即减少表在磁盘上的大小，
因为被删除列占用的空间不会被回收。随着现有行被更新，空间会逐渐回收。

ALTER TYPE要求重写整个表这一事实有时反而是一个优点，因为重写过程会清除表中的任何死空间。例如，要立即回收一个被删除列占用的空间，最快的方式是：

```
ALTER TABLE table ALTER COLUMN anycol TYPE anytype;
```

其中anycol是表中任意一个剩余的列，而anytype是该列已有的类型。
这不会对表产生任何语义上可见的改变，但该命令会强制重写，从而清除不再有用的数据。

ALTER TYPE的USING选项实际上可以指定任何涉及该行旧值的表达式；也就是说，它既可以引用正在转换的列，也可以引用其他列。这使得使用ALTER TYPE语法完成非常通用的转换成为可能。正因为这种灵活性，USING表达式不会应用到列的默认值（如果有）上，
因为其结果可能不是默认值所要求的常量表达式。这意味着，
当从旧类型到新类型不存在隐式或赋值类型转换时，即便提供了USING子句，ALTER TYPE也可能仍然无法转换默认值。在这种情况下，可以先用DROP DEFAULT删除默认值，执行ALTER TYPE，然后再用SET DEFAULT添加一个合适的新默认值。类似的考虑也适用于涉及该列的索引和约束。

如果表有任何后代表，则不允许只在父表中添加、重命名或更改列的类型，
或重命名继承的约束，而不对后代表执行相同操作。也就是说，ALTER TABLE ONLY会被拒绝。这确保后代表始终拥有与父表匹配的列。

只有当某个后代表中的列既不是从其他父表继承而来，也从未有过该列的独立定义时，递归的DROP COLUMN操作才会移除该后代表中的此列。非递归的DROP COLUMN（即ALTER TABLE ONLY ... DROP COLUMN）永远不会移除任何后代列，而只会把它们标记为独立定义，而非继承得到。

TRIGGER、CLUSTER、OWNER 和 TABLESPACE 操作永远不会递归到后代表；也就是说，它们的行为始终如同指定了 ONLY。添加约束时，只有 CHECK 约束可以递归，而且对这类约束必须递归。

不允许更改系统目录表的任何部分。

有关有效参数的进一步说明，请参见CREATE TABLE。第 5 章中还有关于继承的更多信息。

示例

要添加类型为varchar的列到表中：

```
ALTER TABLE distributors ADD COLUMN address varchar(30);
```

要从表中删除一列：

```
ALTER TABLE distributors DROP COLUMN address RESTRICT;
```

要在一个操作中更改两个现有列的类型：

```
ALTER TABLE distributors
    ALTER COLUMN address TYPE varchar(80),
    ALTER COLUMN name TYPE varchar(100);
```

要把一个包含 Unix 时间戳的整数列改为 `timestamp with time zone`，并通过`USING`子句完成转换：

```
ALTER TABLE foo
    ALTER COLUMN foo_timestamp TYPE timestamp with time zone
    USING
        timestamp with time zone 'epoch' + foo_timestamp * interval '1
 second';
```

要重命名一个现有列：

```
ALTER TABLE distributors RENAME COLUMN address TO city;
```

重命名一个现有的表：

```
ALTER TABLE distributors RENAME TO suppliers;
```

为一列增加一个非空约束：

```
ALTER TABLE distributors ALTER COLUMN street SET NOT NULL;
```

从一列移除一个非空约束：

```
ALTER TABLE distributors ALTER COLUMN street DROP NOT NULL;
```

要向一个表添加一个检查约束：

```
ALTER TABLE distributors ADD CONSTRAINT zipchk CHECK (char_length(
zipcode) = 5);
```

要从一个表及其所有后代移除一个检查约束：

```
ALTER TABLE distributors DROP CONSTRAINT zipchk;
```

为一个表增加一个外键约束：

```
ALTER TABLE distributors ADD CONSTRAINT distfk FOREIGN KEY (address)
 REFERENCES addresses (address);
```

为一个表增加一个（多列）唯一约束：

```
ALTER TABLE distributors ADD CONSTRAINT dist_id_zipcode_key UNIQUE (
dist_id, zipcode);
```

为一个表增加一个自动命名的主键约束，注意一个表只能拥有一个主键：

```
ALTER TABLE distributors ADD PRIMARY KEY (dist_id);
```

把一个表移动到一个不同的表空间：

```
ALTER TABLE distributors SET TABLESPACE fasttablespace;
```

把一个表移动到一个不同的模式：

```
ALTER TABLE myschema.distributors SET SCHEMA yourschema;
```

兼容性

`ADD`、`DROP`和`SET DEFAULT`形式符合 SQL 标准。其他形式是PostgreSQL对 SQL 标准的扩展。此外，在单条`ALTER TABLE`命令中指定多个操作的能力也是一种扩展。

`ALTER TABLE DROP COLUMN`可以被用来删除一个表的唯一的列，从而留下一个零列的表。这是一种 SQL 的扩展，SQL 中不允许零列的表。

ALTER TABLESPACE

ALTER TABLESPACE — 更改一个表空间的定义

大纲

```
ALTER TABLESPACE name RENAME TO newname
ALTER TABLESPACE name OWNER TO newowner
```

描述

`ALTER TABLESPACE`可以用于更改表空间的定义。

要使用`ALTER TABLESPACE`，必须拥有该表空间。要更改所有者，还必须是新所有者角色的直接或间接成员。（请注意，超级用户会自动拥有这些权限。）

参数

name

现有表空间的名称。

newname

该表空间的新名称。新名称不能以`pg_`开头，因为这类名称保留给系统表空间使用。

newowner

该表空间的新拥有者。

示例

将表空间`index_space`重命名为`fast_raid`：

```
ALTER TABLESPACE index_space RENAME TO fast_raid;
```

更改表空间`index_space`的拥有者：

```
ALTER TABLESPACE index_space OWNER TO mary;
```

兼容性

在 SQL 标准中没有 `ALTER TABLESPACE`语句。

另见

CREATE TABLESPACE, DROP TABLESPACE

ALTER TRIGGER

ALTER TRIGGER — 更改触发器的定义

大纲

```
ALTER TRIGGER name ON table RENAME TO newname
```

描述

`ALTER TRIGGER`更改现有触发器的属性。`RENAME`子句更改给定触发器的名称，但不以其他方式更改触发器定义。

若要更改触发器的属性，你必须拥有该触发器所作用的表。

参数

name

要修改的现有触发器的名称。

table

该触发器所作用的表的名称。

newname

该触发器的新名称。

注解

临时启用或禁用触发器的能力由ALTER TABLE提供，而不是由`ALTER TRIGGER`提供，因为`ALTER TRIGGER`没有便捷的方式来表达一次启用或禁用表中所有触发器的选项。

示例

要重命名一个现有触发器：

```
ALTER TRIGGER emp_stamp ON emp RENAME TO emp_track_chgs;
```

兼容性

`ALTER TRIGGER`是 PostgreSQL 对 SQL 标准的扩展。

参见

ALTER TABLE

ALTER TYPE

ALTER TYPE — 更改类型的定义

大纲

```
ALTER TYPE name OWNER TO new_owner
ALTER TYPE name SET SCHEMA new_schema
```

描述

`ALTER TYPE`更改一个现有类型的定义。目前唯一可用的功能是更改类型的所有者和模式。

要使用 `ALTER TYPE`，必须拥有该类型。要更改类型的模式，还必须在新模式上拥有 `CREATE` 权限。要更改所有者，还必须是新所有者角色的直接或间接成员，并且该角色必须在类型所在模式上拥有 `CREATE` 权限。（这些限制确保更改所有者不会执行任何无法通过删除并重新创建该类型完成的操作。不过，超级用户无论如何都可以更改任何类型的所有权。）

参数

name

要修改的现有类型的名称（可以是模式限定的）。

new_owner

该类型新的拥有者的用户名。

new_schema

该类型的新模式。

示例

要将用户定义类型`email`的拥有者改为 `joe`：

```
ALTER TYPE email OWNER TO joe;
```

要将用户定义类型`email`所在的模式改为 `customers`：

```
ALTER TYPE email SET SCHEMA customers;
```

兼容性

SQL 标准中没有 `ALTER TYPE` 语句。

ALTER USER

ALTER USER — 更改数据库角色

大纲

```
ALTER USER name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'

ALTER USER name RENAME TO newname

ALTER USER name SET parameter { TO | = } { value | DEFAULT }
ALTER USER name RESET parameter
```

描述

`ALTER USER`现在是ALTER ROLE的别名，更多信息请参见该命令。

兼容性

`ALTER USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义交由具体实现决定。

另见

ALTER ROLE

ANALYZE

ANALYZE — 收集数据库的统计信息

大纲

ANALYZE [VERBOSE] [*table* [(*column* [, ...])]]

描述

ANALYZE收集数据库中各表内容的统计信息，并将结果存储到 pg_statistic系统表中。随后，查询规划器会使用这些统计信息来帮助确定查询的最高效执行计划。

如果不带参数，ANALYZE 会检查当前数据库中的每个表。带有参数时，ANALYZE 只检查该表。还可以提供列名列表，这时只会收集这些列的统计信息。

参数

VERBOSE

启用进度消息显示。

table

要分析的特定表的名称（可带模式限定）。默认为当前数据库中的所有表。

column

要分析的特定列的名称。默认为所有列。

输出

指定VERBOSE时，ANALYZE会输出进度消息，指示当前正在处理哪个表，
同时还会打印这些表的各种统计信息。

注解

最好定期运行 ANALYZE，或者在对表内容做出重大更改之后立即运行。准确的统计信息有助于规划器选择最合适的查询计划，从而提高查询处理速度。
一个常见的策略是在每天使用率较低的时段运行一次VACUUM和 ANALYZE。

与VACUUM FULL不同，ANALYZE 只需获取目标表上的读锁，因此可以与该表上的其他活动并行运行。

ANALYZE收集的统计信息通常包括每列中某些高频值的列表，
以及显示每列大致数据分布的直方图。如果ANALYZE认为它们没有意义（例如，
唯一键列中没有高频值），或者列的数据类型不支持相应的操作符，
则可能省略其中一项或两项。更多统计信息见第 22 章。

对于大型表，ANALYZE会对表内容进行随机采样，而不是检查每一行。这使得即使是很大的表，也能在较短时间内完成分析。不过要注意，这些统计信息只是近似值，

并且即使实际表内容没有变化，每次运行ANALYZE时统计信息也会略有变化。
这可能导致EXPLAIN中显示的规划器估算代价略有变化。在少数情况下，
这种非确定性会导致规划器在运行ANALYZE之后改用不同的查询计划。要避免这种情况，
可以按下文所述提高ANALYZE收集的统计信息量。

可以通过调整配置变量default_statistics_target来控制分析程度，也可以针对单个列使用ALTER TABLE ... ALTER COLUMN ... SET STATISTICS设置每列的统计信息目标（参见ALTER TABLE）。目标值会设置高频值列表中的最大条目数，以及直方图中的最大桶数。默认目标值是 100，但可以把它调高或调低，以在规划器估算精度、ANALYZE耗费的时间以及pg_statistic占用的空间之间作出权衡。特别地，
将统计信息目标设置为零会禁用该列的统计信息收集。对于那些从不出现在查询 WHERE、GROUP BY或ORDER BY子句中的列，这样做可能很有用，因为规划器不会用到这些列的统计信息。

被分析列中最大的统计信息目标决定了为了生成统计信息而需要采样的表行数。
增加该目标会导致执行ANALYZE所需的时间和空间按比例增加。

兼容性

SQL 标准中没有ANALYZE语句。

BEGIN

BEGIN — 开始一个事务块

大纲

```
BEGIN [ WORK | TRANSACTION ] [ transaction_mode [, ...] ]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
```

描述

`BEGIN`启动一个事务块，也就是说，在 `BEGIN`命令之后的所有语句都会在同一个事务中执行，直到显式给出COMMIT或ROLLBACK为止。默认情况下（不使用`BEGIN`），
PostgreSQL以“自动提交”模式执行事务，也就是说，每条语句都在自己的事务中执行，
并在语句结束时隐式提交（如果执行成功，否则回滚）。

在事务块中执行语句通常会更快，因为事务的启动和提交需要大量的 CPU 和磁盘活动。在进行多个相关更改时，
把多条语句放在一个事务中执行也有助于保证一致性：其他会话将无法看到那些相关更新尚未全部完成时的中间状态。

如果指定了隔离级别、读写模式或可延迟模式，新事务就会具有这些特征，就像执行了SET TRANSACTION一样。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

关于本语句其他参数含义的信息，参见SET TRANSACTION。

注解

START TRANSACTION与`BEGIN`具有相同功能。

使用COMMIT或ROLLBACK来结束事务块。

如果在已经处于事务块中时发出`BEGIN`，将会产生一条警告消息。事务状态不会受影响。
若要在事务块中实现事务嵌套，请使用保存点（参见SAVEPOINT）。

出于向后兼容的考虑，连续多个*transaction_modes*之间的逗号可以省略。

示例

开始一个事务块：

```
BEGIN;
```

兼容性

BEGIN是PostgreSQL的一种语言扩展。它等价于 SQL 标准命令START TRANSACTION，后者的参考页中包含更多兼容性信息。

另外，在嵌入式 SQL 中，关键字BEGIN用于不同的目的。在移植数据库应用程序时，应当谨慎处理事务语义。

另见

COMMIT, ROLLBACK, START TRANSACTION, SAVEPOINT

CHECKPOINT

CHECKPOINT — 强制执行一次事务日志检查点

大纲

```
CHECKPOINT
```

描述

预写式日志（WAL）会每隔一段时间在事务日志中设置一个检查点。（要调整自动检查点间隔，请参见运行时配置选项checkpoint_segments和checkpoint_timeout。）`CHECKPOINT` 命令会在发出该命令时强制立即执行检查点，而不等待预定的检查点。

检查点是事务日志序列中的一个位置，到这一位置时，所有数据文件都已更新，以反映日志中的信息。所有数据文件都会被刷出到磁盘。有关 WAL 系统的更多信息，参见第 26 章。

只有超级用户才能调用 `CHECKPOINT`。该命令并非设计用于正常运行期间。

兼容性

`CHECKPOINT`命令是 PostgreSQL的一种语言扩展。

CLOSE

CLOSE — 关闭一个游标

大纲

```
CLOSE name
```

描述

CLOSE释放与一个打开的游标相关联的资源。游标关闭后，不允许再对其执行任何后续操作。当不再需要游标时，应将其关闭。

每个不可保持的打开游标都会在事务通过COMMIT或 ROLLBACK结束时被隐式关闭。可保持游标会在创建它的事务通过ROLLBACK中止时被隐式关闭。如果创建它的事务成功提交，则该可保持游标会一直保持打开状态，直到显式执行 CLOSE，或者客户端断开连接。

参数

name

要关闭的打开游标的名称。

注解

PostgreSQL没有显式的OPEN游标语句；游标在声明时即被视为打开。请使用DECLARE语句来声明游标。

示例

关闭游标liahona：

```
CLOSE liahona;
```

兼容性

CLOSE完全符合 SQL 标准。

另见

DECLARE, FETCH, MOVE

CLUSTER

CLUSTER — 按照一个索引对表进行聚簇

大纲

```
CLUSTER indexname ON tablename
CLUSTER tablename
CLUSTER
```

描述

`CLUSTER` 指示 PostgreSQL 按照 *indexname* 指定的索引，对 *tablename* 指定的表进行聚簇。该索引必须已经定义在 *tablename* 上。

当一个表被聚簇时，它会根据索引信息在物理上重新排序。
聚簇是一次性操作：之后如果表再被更新，这些更改不会再次被聚簇。也就是说，
系统不会试图按照索引顺序存储新行或更新后的行。如果需要，
可以定期再次执行该命令来重新聚簇。

当一个表被聚簇时，PostgreSQL 会记住该表是按哪个索引聚簇的。形式 `CLUSTER` *tablename* 会使用与之前相同的索引对表重新聚簇。

不带任何参数的 `CLUSTER` 会对当前数据库中属于调用用户的所有表重新执行聚簇；如果由超级用户调用，则会对所有表重新执行聚簇。（从未聚簇过的表不包括在内。）这种形式的 `CLUSTER` 不能在事务或函数内调用。

当对一个表执行聚簇时，会在该表上获取 `ACCESS EXCLUSIVE` 锁。这会阻止任何其他数据库操作（包括读和写）在 `CLUSTER` 完成前访问该表。

参数

indexname

索引的名称。

tablename

表的名称（可能是模式限定的）。

注解

`CLUSTER`会丢失元组的全部可见性信息，这使得在`CLUSTER`命令完成之前获取的任何快照看来，该表都是空的。这使得`CLUSTER`不适合在访问被聚簇表的事务与`CLUSTER`
并发运行的应用中使用。这一点在可串行化事务中最为明显，
因为它们在事务开始时只取一个快照，但读已提交事务也会受到影响。

当你在表中随机访问单行时，数据在表中的实际顺序并不重要。但是，
如果访问某些数据比其他数据更频繁，并且存在将这些数据放在一起的索引，使用`CLUSTER`
就会有益。如果要从表中获取一段索引值范围，或者获取有多行匹配的单个索引值，`CLUSTER`
会有帮助，因为一旦索引确定第一个匹配行所在的表页，
其他匹配行很可能已在同一个堆页面中，从而减少磁盘访问并加快查询。

使用索引扫描时，会创建一个表的临时副本，其中表数据按索引顺序排列。
还会为表上的每个索引创建临时副本。因此，
所需的磁盘空闲空间至少应等于表大小与索引大小之和。

由于CLUSTER会记住聚簇信息，你可以第一次手工聚簇想要聚簇的表，然后设置一个类似于VACUUM的定时事件，让这些表被周期性地重新聚簇。

因为规划器会记录有关表中数据顺序的统计信息，建议在新近聚簇过的表上运行ANALYZE。
否则，规划器可能会选择很差的查询计划。

还有另一种聚簇数据的方法。CLUSTER命令通过使用你指定的索引扫描原始表来对它重新排序。
这对于大型表可能很慢，因为行是按索引顺序从堆中取得的，而如果堆表是无序的，
条目就散布在随机页面中，因此每移动一行都要检索一个磁盘页。（PostgreSQL有缓存，
但一个大表的绝大部分放不进缓存。）另一种聚簇表的方法是使用：

```
CREATE TABLE newtable AS
    SELECT columnlist FROM table ORDER BY columnlist;
```

它使用ORDER BY子句中PostgreSQL的排序代码产生想要的顺序；对于无序数据，这通常比索引扫描快得多。然后你删除旧表，使用ALTER TABLE ... RENAME把*newtable*改名为旧表的名字，并重建该表的索引。不过，这种方法不保留 OID、约束、外键关系、被授予的权限以及表的其他附属属性–所有这些项都必须手工重建。

示例

按照索引 emp_ind 对表 emp 进行聚簇：

```
CLUSTER emp_ind ON emp;
```

使用之前用过的同一个索引对 emp 表进行聚簇：

```
CLUSTER emp;
```

对数据库中已聚簇过的所有表执行聚簇：

```
CLUSTER;
```

兼容性

SQL 标准中没有 CLUSTER 语句。

另见

clusterdb

COMMENT

COMMENT — 定义或修改对象的注释

大纲

```
COMMENT ON
{
  TABLE object_name |
  COLUMN table_name.column_name |
  AGGREGATE agg_name (agg_type) |
  CAST (sourcetype AS targettype) |
  CONSTRAINT constraint_name ON table_name |
  CONVERSION object_name |
  DATABASE object_name |
  DOMAIN object_name |
  FUNCTION func_name ( [ [ argmode ] [ argname ] argtype [, ...] ] ) |
  INDEX object_name |
  LARGE OBJECT large_object_oid |
  OPERATOR op (leftoperand_type, rightoperand_type) |
  OPERATOR CLASS object_name USING index_method |
  [ PROCEDURAL ] LANGUAGE object_name |
  RULE rule_name ON table_name |
  SCHEMA object_name |
  SEQUENCE object_name |
  TRIGGER trigger_name ON table_name |
  TYPE object_name |
  VIEW object_name
} IS 'text'
```

描述

COMMENT存储数据库对象的注释。

要修改一个注释，可以对同一对象发出一条新的COMMENT命令。
每个对象只存储一个注释字符串。要移除一个注释，用NULL代替文本字符串。对象被删除时，其注释也会被自动删除。

可以很容易地用psql的命令 \dd、\d+和\l+检索注释。其他用于检索注释的用户界面，也可以建立在 psql使用的同一组内置函数之上，即 obj_description和col_description（见表 9.43）。

参数

object_name
table_name.column_name
agg_name
constraint_name
func_name
op
rule_name
trigger_name

要添加注释的对象的名称。表、聚合、域、函数、索引、操作符、操作符类、序列、类型和视图的名称可以带模式限定。

agg_type

聚合函数的参数数据类型，如果函数接受任意数据类型，则为*。

sourcetype

类型转换的源数据类型的名称。

targettype

类型转换的目标数据类型的名称。

argmode

函数参数的模式：要么是IN、OUT，要么是INOUT。如果省略，默认为 IN。注意 COMMENT ON FUNCTION 实际上并不关心 OUT 参数，因为确定函数的标识只需要输入参数。因此只列出 IN 和 INOUT 参数就足够了。

argname

函数参数的名称。注意 COMMENT ON FUNCTION 实际上并不关心参数名称，因为确定函数的标识只需要参数的数据类型。

argtype

大对象的 OID。

large_object_oid

大对象的 OID。

PROCEDURAL

这是一个噪声词。

text

新的注释，以字符串字面量形式书写；也可以写 NULL 来移除该注释。

注解

对数据库的注释只能在该数据库中创建，并且只在该数据库中可见，在其他数据库中不可见。

目前注释没有任何安全机制：任何连接到数据库的用户都能看到该数据库中对象的全部注释（不过只有超级用户才能修改他们不拥有的对象的注释）。因此，
不要把安全攸关的信息放在注释里。

示例

为表mytable附加一条注释：

```
COMMENT ON TABLE mytable IS 'This is my table.';
```

再将其移除：

```
COMMENT ON TABLE mytable IS NULL;
```

更多示例：

```
COMMENT ON AGGREGATE my_aggregate (double precision) IS 'Computes
 sample variance';
COMMENT ON CAST (text AS int4) IS 'Allow casts from text to int4';
COMMENT ON COLUMN my_table.my_column IS 'Employee ID number';
COMMENT ON CONVERSION my_conv IS 'Conversion to UTF8';
COMMENT ON DATABASE my_database IS 'Development Database';
COMMENT ON DOMAIN my_domain IS 'Email Address Domain';
COMMENT ON FUNCTION my_function (timestamp) IS 'Returns Roman
 Numeral';
COMMENT ON INDEX my_index IS 'Enforces uniqueness on employee ID';
COMMENT ON LANGUAGE plpython IS 'Python support for stored
 procedures';
COMMENT ON LARGE OBJECT 346344 IS 'Planning document';
COMMENT ON OPERATOR ^ (text, text) IS 'Performs intersection of two
 texts';
COMMENT ON OPERATOR - (NONE, text) IS 'This is a prefix operator on
 text';
COMMENT ON OPERATOR CLASS int4ops USING btree IS '4 byte integer
 operators for btrees';
COMMENT ON RULE my_rule ON my_table IS 'Logs updates of employee
 records';
COMMENT ON SCHEMA my_schema IS 'Departmental data';
COMMENT ON SEQUENCE my_sequence IS 'Used to generate primary keys';
COMMENT ON TABLE my_schema.my_table IS 'Employee Information';
COMMENT ON TEXT SEARCH DICTIONARY swedish IS 'Snowball stemmer for
 Swedish language';
COMMENT ON TRIGGER my_trigger ON my_table IS 'Used for RI';
COMMENT ON TYPE complex IS 'Complex number data type';
COMMENT ON VIEW my_view IS 'View of departmental costs';
```

兼容性

SQL 标准中没有COMMENT命令。

COMMIT

COMMIT — 提交当前事务

大纲

```
COMMIT [ WORK | TRANSACTION ]
```

描述

`COMMIT`提交当前事务。该事务所做的所有更改都会对其他会话可见，并且如果发生崩溃，也能保证其持久性。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用ROLLBACK中止事务。

在事务之外发出 `COMMIT` 不会造成危害，但会产生一条警告消息。

示例

提交当前事务，并使所有更改永久生效：

```
COMMIT;
```

兼容性

SQL 标准只规定了 `COMMIT` 和 `COMMIT WORK` 这两种形式。在其他方面，此命令完全符合标准。

另见

BEGIN, ROLLBACK

COMMIT PREPARED

COMMIT PREPARED — 提交一个先前为两阶段提交而预备的事务

大纲

```
COMMIT PREPARED transaction_id
```

描述

`COMMIT PREPARED`提交一个处于预备状态的事务。

参数

transaction_id

要提交的事务的事务标识符。

注解

要提交预备事务，执行者必须是最初执行该事务的同一用户，或者是超级用户；但不必处在执行该事务的同一会话中。

这个命令不能在事务块内执行。该预备事务会被立即提交。

当前所有处于预备状态的事务都列在 `pg_prepared_xacts` 系统视图中。

示例

提交事务标识符为`foobar`的事务：

```
COMMIT PREPARED 'foobar';
```

另见

PREPARE TRANSACTION, ROLLBACK PREPARED

COPY

COPY — 在文件和表之间复制数据

大纲

```
COPY tablename [ ( column [, ...] ) ]
    FROM { 'filename' | STDIN }
    [ [ WITH ]
          [ BINARY ]
          [ OIDS ]
          [ DELIMITER [ AS ] 'delimiter' ]
          [ NULL [ AS ] 'null string' ]
          [ CSV [ HEADER ]
                [ QUOTE [ AS ] 'quote' ]
                [ ESCAPE [ AS ] 'escape' ]
                [ FORCE NOT NULL column [, ...] ]

COPY tablename [ ( column [, ...] ) ]
    TO { 'filename' | STDOUT }
    [ [ WITH ]
          [ BINARY ]
          [ OIDS ]
          [ DELIMITER [ AS ] 'delimiter' ]
          [ NULL [ AS ] 'null string' ]
          [ CSV [ HEADER ]
                [ QUOTE [ AS ] 'quote' ]
                [ ESCAPE [ AS ] 'escape' ]
                [ FORCE QUOTE column [, ...] ]
```

描述

`COPY`在 PostgreSQL表与标准文件系统文件之间传输数据。`COPY TO`将表的内容复制到文件，而`COPY FROM`则将数据从文件复制到表中（追加到表中已有的数据之后）。

如果指定了列列表，`COPY`将只在指定列与文件之间复制数据。如果表中存在未列入列列表的列，`COPY FROM`将为这些列插入默认值。

带文件名的`COPY`会指示 PostgreSQL 服务器直接从文件读取或向文件写入。该文件必须可由服务器访问，并且其名称必须从服务器的视角指定。当指定`STDIN`或`STDOUT`时，数据通过客户端与服务器之间的连接传输。

参数

tablename

一个现有表的名称（可以是模式限定的）。

column

要复制的可选列列表。如果没有指定列列表，则会使用所有列。

filename

输入或输出文件的绝对路径名。Windows 用户可能需要把路径名中用到的反斜线都写成双份。

STDIN

指定输入来自客户端应用程序。

STDOUT

指定输出发送到客户端应用程序。

BINARY

令所有数据以二进制格式而不是文本格式存储或读取。在二进制模式下不能指定 DELIMITER、NULL 或 CSV 选项。

OIDS

指定复制每一行的 OID。（如果为没有 OID 的表指定了 OIDS，就会报错。）

delimiter

分隔文件中每行内各列的单个字符。文本模式下默认为制表符，CSV 模式下默认为逗号。

null string

指定表示一个空值的字符串。文本格式中默认是 \N（反斜线-N），CSV格式中默认是一个未加引用的空串。在你不想区分空值和空串的情况下，
即使在文本格式中你也可能更喜欢空串。

> **注意**
>
> 在使用COPY FROM时，任何匹配此字符串的数据项都会被存储为空值，因此应确保这里使用的字符串与 COPY TO时使用的相同。

CSV

选择逗号分隔值（CSV）模式。

HEADER

指定文件包含一个标题行，其中列出文件中每一列的名称。输出时，
第一行包含表中的列名；输入时，第一行会被忽略。

quote

指定 CSV 模式中的引号字符。默认是双引号。

escape

指定在 CSV 模式中出现于数据字符值中的 QUOTE 字符之前的字符。默认为 QUOTE 值（通常是双引号）。

FORCE QUOTE

在 CSV COPY TO 模式中，强制对每个指定列中的所有非 NULL 值使用引号。NULL 输出永远不会加引号。

FORCE NOT NULL

在 CSV COPY FROM 模式中，将每个指定列处理为好像加了引号、因而不会是 NULL 值。对于 CSV 模式中的默认空值串（''），这会使缺失值被输入为零长度的字符串。

注解

COPY只能用于普通表，不能用于视图。

BINARY关键字使所有数据以二进制格式而不是文本格式存储/读取。它比普通文本模式略快，但二进制格式的文件在跨机器体系结构和PostgreSQL版本之间的可移植性较差。

你必须对COPY TO读取其值的表具有 SELECT权限，并对COPY FROM 插入其值的表具有INSERT权限。

COPY 命令中命名的文件由服务器直接读取或写入，而不是由客户端应用进行。因此，这些文件必须位于数据库服务器机器上或可由其访问，而不是客户端上。它们必须可由 PostgreSQL 用户（服务器运行时使用的用户 ID）访问并可读或可写，而不是由客户端访问。带有文件名的 COPY 只允许数据库超级用户使用，因为它允许读取或写入服务器有权访问的任何文件。

不要把 COPY 与 psql 的指令 \copy 混淆。\copy 会调用 COPY FROM STDIN 或 COPY TO STDOUT，然后在 psql 客户端可访问的文件中获取/存储数据。因此，使用 \copy 时，文件的可访问性和访问权限取决于客户端而不是服务器。

建议在COPY中使用的文件名始终指定为绝对路径。对于COPY TO，服务器会强制这一点；但对于 COPY FROM，你仍可选择从使用相对路径指定的文件中读取。该路径将相对于服务器进程的工作目录（位于数据目录下某处）而非客户端的工作目录进行解释。

COPY FROM将调用目标表上的任何触发器和检查约束。但是它不会调用规则。

COPY的输入和输出会受 DateStyle影响。为确保数据能移植到其他可能使用非默认 DateStyle设置的PostgreSQL 安装中，使用COPY TO前应将 DateStyle设置为ISO。

COPY 会在遇到第一个错误时停止操作。对于 COPY TO，这应当不会导致问题；但对于 COPY FROM，目标表此时已经接收了前面的行。这些行不可见也不可访问，但仍占用磁盘空间。如果在一次大型复制操作已经进行很久后才失败，可能会浪费大量磁盘空间。可以调用 VACUUM 回收这些浪费的空间。

文件格式

文本格式

当COPY不带BINARY或CSV选项使用时，读取或写入的是一个文本文件，
其中表中的每一行对应文件中的一行。每行中的列由分隔符字符隔开。
列值本身是由各属性数据类型的输出函数生成或可被其输入函数接受的字符串。
对于为空值的列，会使用指定的空值串代替。
如果输入文件中的任何一行包含的列数多于或少于预期，COPY FROM就会报错。如果指定了 OIDS，OID 会作为第一列读取或写入，位于用户数据列之前。

数据结束可以表示为只包含反斜线加点号（\.）的一行。从文件读取时，不需要数据结束标记，因为文件结束已足够；只有在使用 3.0 之前版本的客户端协议、向客户端应用程序复制数据或从其复制数据时，才需要该标记。

在COPY数据中，可以使用反斜线字符（\）
来转义那些原本可能被当作行或列分隔符的数据字符。特别是，
如果下列字符作为列值的一部分出现，那么它们前面必须加一个反斜线：反斜线本身、换行、
回车以及当前分隔符字符。

COPY TO输出指定的空值串时不会添加任何反斜线；相反，COPY FROM会在去除反斜线之前先将输入与空值串进行匹配。因此，像\N这样的空值串不会与实际的数据值\N混淆，
因为后者会表示为\\N。

COPY FROM识别下列特殊的反斜线序列：

序列	表示
\b	退格 (ASCII 8)
\f	换页 (ASCII 12)
\n	新行 (ASCII 10)
\r	回车 (ASCII 13)
\t	制表 (ASCII 9)
\v	纵向制表 (ASCII 11)
digits	反斜线后跟一到三个八进制数字表示该数字代码对应的字节
\x*digits*	反斜线加x后跟一到两个十六进制数字表示该数字代码对应的字节

目前，COPY TO从不会输出八进制或十六进制数字反斜线序列，但对这些控制字符确实会使用上表列出的其他序列。

任何上表中未提到的其他反斜线字符都将表示其自身。不过，要注意不要不必要地添加反斜线，
因为那可能意外地产生与数据结束标记（\.）或空值串（默认是\N）匹配的字符串。
这些字符串会在进行任何其他反斜线处理之前先被识别出来。

强烈建议生成COPY数据的应用将数据中的换行和回车分别转换为\n和\r序列。目前，
仍然可以用反斜线加回车表示数据回车，用反斜线加换行表示数据换行。不过，
未来版本可能不再接受这些表示方式。如果COPY文件在不同机器之间传输（例如从 Unix 到 Windows，或反之），这些表示方式也非常容易被破坏。

COPY TO会用 Unix 风格的换行（“\n”）结束每一行。运行在 Microsoft Windows 上的服务器则会输出回车/换行（“\r\n”），但这只适用于复制到服务器文件的COPY；为保证跨平台一致性，COPY TO STDOUT总是发送“\n”，与服务器平台无关。COPY FROM能够处理以换行、回车或回车/换行结束的行。为减少本应是数据的未加反斜线新行或回车带来的风险，如果输入中的行结束符并不一致，COPY FROM将会报错。

CSV 格式

这种格式选项用于导入和导出许多其他程序（如电子表格）使用的逗号分隔值（CSV）
文件格式。不同于 PostgreSQL标准文本格式使用的转义规则，它会生成并识别通用的 CSV转义机制。

每条记录中的值由DELIMITER字符分隔。如果某个值包含分隔符字符、QUOTE字符、NULL
字符串、回车或换行字符，那么整个值都会以前后各一个QUOTE字符包围，并且该值内每次出现

QUOTE字符或ESCAPE 字符之前都会加上转义字符。还可以使用FORCE QUOTE在输出特定列的非NULL值时强制加引号。

CSV格式没有标准方式区分NULL值和空字符串。PostgreSQL的COPY通过引号来处理这一区别。NULL输出为NULL 字符串并且不加引号，而与NULL 字符串匹配的数据值则会加引号。因此，使用默认设置时，NULL 会写成一个未加引号的空字符串，而空字符串则写成双引号包围的形式（""）。读取值时遵循类似规则。你可以使用 FORCE NOT NULL来阻止对特定列进行NULL输入比较。

因为反斜线在 CSV 格式中不是特殊字符，数据结束标记 \. 也可能作为数据值出现。为避免误解，当 \. 数据值作为一行中的唯一字段出现时，输出时会自动给它加上引号；输入时，如果它带有引号，就不会被解释为数据结束标记。如果正在装载由其他应用程序创建的文件，其中只有一个不带引号的列，而且该列可能包含值 \.，那么可能需要在输入文件中给该值加上引号。

注意

在CSV格式中，所有字符都有意义。被空白字符或 DELIMITER之外其他字符包围的带引号值，会把这些字符也包含进值中。如果你导入的数据来自某个会用空白把CSV 行填充到固定宽度的系统，这可能导致错误。出现这种情况时，你可能需要在将数据导入PostgreSQL之前，先预处理CSV文件以移除尾随空白。

注意

CSV 格式既能识别也能生成这样的 CSV 文件：其中带引号的值包含内嵌的回车和换行。因此，这类文件不像文本格式文件那样严格地一行对应表中的一行。

注意

很多程序会生成奇怪、甚至近乎反常的CSV文件，因此这种文件格式更像一种约定而非标准。因而你可能会遇到无法用这种机制导入的文件，而COPY也可能生成其他程序无法处理的文件。

二进制格式

COPY BINARY所使用的文件格式在 PostgreSQL 7.4 中发生了变化。新格式由一个文件头、零个或多个包含行数据的元组以及一个文件尾组成。头部和数据现在都采用网络字节序。

文件头

文件头由 15 字节的固定字段组成，后面跟着一个可变长度的头部扩展区域。固定字段如下：

签名

11 字节序列PGCOPY\n\377\r\n\0 — 注意，零字节是签名中必不可少的一部分。（该签名的设计目的是便于识别那些在不具备 8 位透明性的传输过程中遭到破坏的文件。行尾转换过滤器、零字节丢失、高位丢失或奇偶校验变化等情况都会改变该签名。）

标志域

这是一个 32 位整数位掩码，用于表示文件格式的重要属性。位编号从 0（LSB）到 31（MSB）。注意，此字段与文件格式中的所有整数字段一样，采用网络字节序存储（最高有效字节在前）。位 16-31 保留用于表示文件格式的关键问题；如果读取程序发现这个范围内有非预期的位被置位，就应中止。位 0-15 保留用于表示向后兼容的格式问题；读取程序应直接忽略这个范围内非预期的置位。目前仅定义了一个标志位，其余位必须为零：

位 16

如果为 1，则数据中包含 OID；如果为 0，则不包含。

头部扩展区长度

32 位整数，表示头部剩余部分的长度（以字节计），不包括该字段本身。当前该值为零，因此其后紧接着第一个元组。未来对这种格式的更改可能允许在头部中包含额外数据。
如果读取程序不知道如何处理头部扩展区数据，应静默跳过它。

头部扩展区被设想为包含一系列可自我标识的块。
标志域并不用于告诉读取程序扩展区中包含哪些内容。
头部扩展内容的具体设计留待后续版本决定。

这种设计既允许向后兼容的头部新增（增加头部扩展块，或设置低位标志位），
也允许不向后兼容的更改（设置高位标志位来表明这类更改，
并在需要时向扩展区增加支持数据）。

元组

每个元组都以一个 16 位整数计数开头，用于表示该元组中的字段数。（目前，一个表中的所有元组都应有相同的计数，但这未必永远如此。）随后，对元组中的每个字段，都会有一个 32 位长度字，后跟该字段数据的相应字节数。（长度字不包括其本身，且可以为零。）特殊情况下，-1 表示一个 NULL 字段值；在 NULL 情况下，后面不会跟随任何值字节。

字段之间没有对齐填充或任何其他额外数据。

当前，`COPY BINARY`文件中的所有数据值都假定为二进制格式（格式代码一）。可以预见，未来的扩展可能会增加一个允许为各列分别指定格式代码的头部字段。

要确定实际元组数据应采用的二进制格式，你应该参考 PostgreSQL源码，特别是各列数据类型对应的`*send`和`*recv`函数（这些函数通常可以在源码分发包的`src/backend/utils/adt/`目录中找到）。

如果文件中包含 OID，则 OID 字段紧跟在字段计数值之后。它是一个普通字段，只是不计入字段数。特别是，它带有一个长度字段 – 这样就能较容易地处理 4 字节或 8 字节的 OID，也允许在将来确有需要时把 OID 表示为空值。

文件尾

文件尾由一个值为 -1 的 16 位整数构成。这很容易与元组的字段计数字区分开来。

如果字段计数字既不是 -1 也不是预期的列数，读取程序应报告错误。这提供了一项额外检查，以防与数据失去同步。

示例

下面的示例使用竖线（|）作为字段分隔符将一个表复制到客户端：

```
COPY country TO STDOUT WITH DELIMITER '|';
```

要将文件中的数据复制到country表中：

```
COPY country FROM '/usr1/proj/bray/sql/country_data';
```

要使用一个会自动删除的临时表，只把名称以 'A' 开头的国家复制到一个文件中：

```
BEGIN;
CREATE TEMP TABLE a_list_countries AS
    SELECT * FROM country WHERE country_name LIKE 'A%';
COPY a_list_countries TO '/usr1/proj/bray/sql/a_list_countries.copy';
ROLLBACK;
```

下面给出适合从STDIN复制到表中的示例数据：

```
AF      AFGHANISTAN
AL      ALBANIA
DZ      ALGERIA
ZM      ZAMBIA
ZW      ZIMBABWE
```

注意每一行中的空白实际上是一个制表符。

下面是用二进制格式输出的相同数据。该数据是用 Unix 工具 `od -c`过滤后显示的。该表具有三列，第一列类型是`char(2)`，第二列类型是`text`，第三列类型是`integer`。
所有行在第三列都是空值。

```
0000000   P   G   C   O   P   Y  \n 377  \r  \n  \0  \0  \0  \0  \0
 \0
0000020  \0  \0  \0  \0 003  \0  \0  \0 002   A   F  \0  \0  \0 013
 A
0000040   F   G   H   A   N   I   S   T   A   N 377 377 377 377  \0
 003
0000060  \0  \0  \0 002   A   L  \0  \0  \0 007   A   L   B   A   N
 I
0000100   A 377 377 377 377  \0 003  \0  \0  \0 002   D   Z  \0  \0
 \0
0000120 007   A   L   G   E   R   I   A 377 377 377 377  \0 003  \0
 \0
0000140  \0 002   Z   M  \0  \0  \0 006   Z   A   M   B   I   A 377
 377
0000160 377 377  \0 003  \0  \0  \0 002   Z   W  \0  \0  \0  \b   Z
 I
0000200   M   B   A   B   W   E 377 377 377 377 377 377
```

兼容性

SQL 标准中没有COPY语句。

以下语法曾在PostgreSQL7.3 之前的版本中使用，目前仍受支持：

```
COPY [ BINARY ] tablename [ WITH OIDS ]
    FROM { 'filename' | STDIN }
    [ [USING] DELIMITERS 'delimiter' ]
```

```
    [ WITH NULL AS 'null string' ]

COPY [ BINARY ] tablename [ WITH OIDS ]
    TO { 'filename' | STDOUT }
    [ [USING] DELIMITERS 'delimiter' ]
    [ WITH NULL AS 'null string' ]
```

CREATE AGGREGATE

CREATE AGGREGATE — 定义一个新的聚合函数

大纲

```
CREATE AGGREGATE name (
    BASETYPE = input_data_type,
    SFUNC = sfunc,
    STYPE = state_data_type
    [ , FINALFUNC = ffunc ]
    [ , INITCOND = initial_condition ]
    [ , SORTOP = sort_operator ]
)
```

描述

`CREATE AGGREGATE` 定义一个新的聚合函数。系统自带了一些基本且常用的聚合函数；它们在第 9.15 节中有文档说明。如果需要定义新类型或需要一个尚未提供的聚合函数，则可以使用 `CREATE AGGREGATE` 来实现所需功能。

如果给出了一个模式名（例如`CREATE AGGREGATE myschema.myagg ...`），则该聚合函数会在指定模式中创建。否则会在当前模式中创建。

聚合函数由其名称和输入数据类型标识。同一模式中的两个聚合如果操作于不同的输入类型，则可以有相同的名称。
聚合的名称和输入数据类型还必须与同一模式中每个普通函数的名称和输入数据类型都不同。

一个聚合函数由一个或两个普通函数构成：一个状态转移函数*sfunc*，以及一个可选的最终计算函数*ffunc*。它们的用法如下：

```
sfunc( internal-state, next-data-item ) ---> next-internal-state
ffunc( internal-state ) ---> aggregate-value
```

PostgreSQL会创建一个数据类型为 *stype*的临时变量，用来保存聚合的当前内部状态。对于每个输入数据项，都会调用状态转移函数来计算新的内部状态值。等到所有数据处理完毕后，再调用一次最终函数来计算聚合的返回值。如果没有最终函数，则原样返回结束时的状态值。

聚合函数可以提供一个初始条件，即内部状态值的初始值。它在数据库中以 `text`类型的列指定和存储，但它必须是该状态值数据类型常量的合法外部表示。如果未提供，则状态值初始为空值。

如果状态转移函数被声明为“strict”的，那么它不能以空值输入被调用。使用这种转移函数时，聚合的执行行为如下：空值输入会被忽略（不调用函数并保留先前的状态值）。
如果初始状态值为空，那么第一个非空输入值会取代状态值，
而转移函数会从第二个非空输入值开始被调用。这对于实现`max`这类聚合很方便。注意，只有当*state_data_type*与第一个*input_data_type*相同时，此行为才可用。
当这两种类型不同时，必须提供非空的初始条件或者使用非严格的转移函数。

如果状态转移函数不是严格的，那么它会在每个输入值上被无条件调用，
并且必须自行处理空值输入和空值转移值。这使聚合作者能完全控制聚合对空值的处理。

如果最终函数被声明为“strict”，那么当结束状态值为空时不会调用它；而是自动返回空结果。（这当然只是 strict 函数的正常行为。）无论如何，最终函数都可以选择返回空值。例如，`avg`的最终函数在发现输入行数为零时会返回空值。

行为类似于MIN或MAX的聚合，有时可以通过查阅索引而不是扫描每个输入行来优化。
如果该聚合可以这样优化，请通过指定一个排序操作符来表明。基本要求是，
该聚合必须返回该操作符所诱导的排序顺序中的第一个元素；换句话说：

```
SELECT agg(col) FROM tab;
```

必须等价于

```
SELECT col FROM tab ORDER BY col USING sortop LIMIT 1;
```

进一步的假设是，该聚合忽略空输入，并且当且仅当不存在非空输入时返回空结果。通常，
某种数据类型的<操作符是 MIN的合适排序操作符，而>是 MAX的合适排序操作符。注意，除非指定的操作符是 B-树索引操作符类中“小于”或“大于”策略成员，否则这种优化实际上永远不会生效。

参数

name

要创建的聚合函数的名称（可带模式限定）。

input_data_type

此聚合函数所操作的输入数据类型。对于不检查其输入值的聚合（例如count(*)），
可以将其指定为"ANY"。

sfunc

要为每个输入数据值调用的状态转移函数的名称。它通常是一个两参数函数，
第一个参数的类型为*state_data_type*，第二个的类型为*input_data_type*。或者，
对于不检查其输入值的聚合，该函数只接受一个类型为*state_data_type*的参数。
无论哪种情况，该函数都必须返回*state_data_type*类型的值。
此函数接受当前状态值和当前输入数据项，并返回下一个状态值。

state_data_type

聚合状态值的数据类型。

ffunc

在遍历完所有输入数据之后调用的、用于计算聚合结果的最终函数的名称。
该函数必须接受单个类型为*state_data_type*的参数。
聚合的返回数据类型被定义为该函数的返回类型。如果没有指定*ffunc*，
则结束状态值被用作聚合的结果，返回类型为*state_data_type*。

initial_condition

状态值的初始设置。这必须是 *state_data_type* 数据类型所接受形式的字符串常量。如果未指定，状态值将从 null 开始。

sort_operator

类MIN或MAX聚合的关联排序操作符。这只是一个操作符名称（可以用模式限定）。
假定该操作符具有与聚合相同的输入数据类型。

CREATE AGGREGATE的参数可以按任意顺序书写，而不只是按上面展示的顺序。

示例

见第 32.10 节。

兼容性

`CREATE AGGREGATE`是 PostgreSQL的语言扩展。SQL 标准不提供用户定义聚合函数。

另见

ALTER AGGREGATE, DROP AGGREGATE

CREATE CAST

CREATE CAST — 定义一种新的类型转换

大纲

```
CREATE CAST (sourcetype AS targettype)
    WITH FUNCTION funcname (argtypes)
    [ AS ASSIGNMENT | AS IMPLICIT ]

CREATE CAST (sourcetype AS targettype)
    WITHOUT FUNCTION
    [ AS ASSIGNMENT | AS IMPLICIT ]
```

描述

`CREATE CAST`定义一种新的类型转换。类型转换规定如何在两种数据类型之间执行转换。例如，

```
SELECT CAST(42 AS text);
```

通过调用一个预先指定的函数（此处是 `text(int4)`）把整型常量 42 转换成 `text`类型。（如果没有定义合适的类型转换，转换就会失败。）

两种类型可以是二进制兼容的，这意味着无需调用任何函数，就可以把它们“免费”地互相转换。这要求对应的值使用相同的内部表示。例如，`text`和`varchar`这两种类型就是二进制兼容的。

默认情况下，只有显式请求类型转换时才会调用一种类型转换，也就是显式使用`CAST(x AS typename)`或 `x::typename` 这种构造。

如果一种类型转换被标记为`AS ASSIGNMENT`，那么在把值赋给目标数据类型的列时就可以隐式调用它。例如，假设`foo.f1` 是一个`text`类型的列，那么如果从`integer`到 `text`的类型转换被标记为`AS ASSIGNMENT`，则：

```
INSERT INTO foo (f1) VALUES (42);
```

就会被允许，否则不会。（我们通常用术语赋值类型转换来描述这种类型转换。）

如果一种类型转换被标记为`AS IMPLICIT`，那么无论是在赋值上下文中还是在表达式内部，都可以在任何上下文中隐式调用它。例如，由于`||`接受`text`操作数，

```
SELECT 'The time is ' || now();
```

只有在从`timestamp`到`text`的类型转换被标记为 `AS IMPLICIT`时才会被允许。否则就必须显式写出该类型转换，例如

```
SELECT 'The time is ' || CAST(now() AS text);
```

（我们通常用术语隐式类型转换来描述这种类型转换。）

把类型转换标记为隐式时应当保持保守。过多的隐式类型转换路径可能导致 PostgreSQL对命令作出令人意外的解释，或者因为存在多种可能的解释而根本无法解析命令。
一个好的经验法则是，只有对同一一般类型分类中且能保留信息的类型间转换，
才让它可以被隐式调用。例如，从`int2`到`int4`的类型转换可以合理地设为隐式，但从`float8`

到`int4`的类型转换大概应仅限赋值使用。跨类型分类的类型转换，例如从`text`到`int4`，最好只允许显式调用。

要能够创建一个类型转换，你必须拥有源数据类型或目标数据类型之一。
要创建一个二进制兼容的类型转换，你必须是超级用户。（之所以有这一限制，是因为错误的二进制兼容类型转换很容易使服务器崩溃。）

参数

sourcetype

该类型转换的源数据类型的名称。

targettype

该类型转换的目标数据类型的名称。

funcname(*argtypes*)

用于执行该类型转换的函数。函数名可以用模式限定；如果没有，
则会在模式搜索路径中查找该函数。
该函数的结果数据类型必须与类型转换的目标类型一致。其参数见下文。

`WITHOUT FUNCTION`

表示源类型和目标类型是二进制兼容的，因此执行该类型转换不需要函数。

`AS ASSIGNMENT`

表示该类型转换可以在赋值上下文中隐式调用。

`AS IMPLICIT`

表示该类型转换可以在任何上下文中隐式调用。

类型转换实现函数可以有一到三个参数。第一个参数类型必须与类型转换的源类型相同。
第二个参数（如果有）必须是 `integer`类型；它接收与目标类型关联的类型修饰符，如果没有则为`-1`。第三个参数（如果有）必须是`boolean` 类型；如果该类型转换是显式类型转换，它接收`true`，否则接收`false`。（奇怪的是，SQL 标准在某些情况下要求显式类型转换和隐式类型转换具有不同的行为。这个参数是为必须实现这类类型转换的函数提供的。
不建议你把自己的数据类型设计成需要关心这一点。）

通常，类型转换的源数据类型和目标数据类型必须不同。不过，
如果它具有一个接受多个参数的类型转换实现函数，
则允许声明源类型和目标类型相同的类型转换。
这用于在系统目录中表示特定类型的长度强制函数。
所命名的函数用于将该类型的值强制为其第二个参数给定的类型修饰符值。

当一种类型转换的源类型和目标类型不同，且其函数接受多个参数时，
它支持在一个步骤中同时完成从一种类型到另一种类型的转换并应用长度强制。
如果没有这样的条目，
对使用类型修饰符的类型进行强制就需要两个类型转换步骤：先在数据类型之间进行转换，
再应用该修饰符。

注解

使用DROP CAST删除用户定义的转换。

请记住，如果你希望能够双向转换类型，就需要在两个方向上分别显式声明类型转换。

在 PostgreSQL 7.3 之前，每个与某种数据类型同名、返回该数据类型且接受一个不同类型参数的函数都会自动成为类型转换函数。随着模式的引入，
以及为了能在系统目录中表示二进制兼容的类型转换，这一惯例已被放弃。
内建的类型转换函数仍然遵循这一命名方案，但它们也必须显示为系统目录 `pg_cast` 中的转换。

虽然这不是强制要求，但仍建议你继续遵循这种老惯例，
即按目标数据类型为类型转换实现函数命名。
许多用户已经习惯于用函数风格的记法进行类型转换，也就是*typename*(*x*)。
这种记法实际上无非就是调用类型转换实现函数；它不会被特别当作类型转换处理。
如果你的转换函数没有按这种惯例命名，那么用户会感到意外。由于 PostgreSQL允许同一函数名按不同参数类型重载，
因此让来自不同类型的多个转换函数都使用目标类型的名称并不存在困难。

> **注意**
>
> 前一段有一个小小的谎言：仍然存在一种情况，`pg_cast` 会被用来解析一个表面上是函数调用的含义。如果函数调用 *name*(*x*)没有匹配到任何实际存在的函数，但*name*是一个数据类型名，并且 `pg_cast`显示存在从*x*的类型到该类型的二进制兼容类型转换，那么该调用会被解释为一个显式类型转换。作出这一例外，
> 是为了让二进制兼容的类型转换即使没有任何函数，也能使用函数语法调用。

示例

要使用函数`int4(text)`创建一种从类型 `text`到类型`int4`的类型转换：

```
CREATE CAST (text AS int4) WITH FUNCTION int4(text);
```

兼容性

`CREATE CAST`命令符合SQL标准，不过 SQL 没有对二进制兼容的类型转换或实现函数的额外参数作出规定。`AS IMPLICIT`也是 PostgreSQL的扩展。

另见

CREATE FUNCTION、CREATE TYPE、DROP CAST

CREATE CONSTRAINT TRIGGER

CREATE CONSTRAINT TRIGGER — 定义一个新的约束触发器

大纲

```
CREATE CONSTRAINT TRIGGER name
    AFTER events ON
    tablename constraint attributes
    FOR EACH ROW EXECUTE PROCEDURE funcname ( args )
```

描述

`CREATE CONSTRAINT TRIGGER`被`CREATE TABLE`/`ALTER TABLE`以及 pg_dump用来创建用于引用完整性的特殊触发器。它并不是为一般用途而设计的。

参数

name

约束触发器的名称。

events

应引发该触发器的事件类别。

tablename

触发事件发生的表（可能是模式限定的）的名称。

constraint

实际的约束说明。

attributes

约束的属性。

funcname(*args*)

作为触发器处理的一部分要调用的函数。

CREATE CONVERSION

CREATE CONVERSION — 定义一个新的编码转换

大纲

```
CREATE [DEFAULT] CONVERSION name
    FOR source_encoding TO dest_encoding FROM funcname
```

描述

`CREATE CONVERSION` 定义字符集编码之间的新转换。转换名可以在`convert`函数中用来指定一个特定的编码转换。此外，被标记为 `DEFAULT` 的转换可用于客户端与服务器之间的自动编码转换。为此目的，必须定义两个转换：从编码 A 到 B 以及从编码 B 到 A。

要创建一个转换，你必须对该函数具有`EXECUTE`权限，并在目标模式上具有`CREATE`权限。

参数

`DEFAULT`

`DEFAULT`子句表示该转换是这一特定源编码到目标编码的默认转换。对于某个编码对，在一个模式中应该只有一个默认转换。

`name`

转换的名称。转换名可以是模式限定的；如果不是，则该转换定义在当前模式中。转换名在一个模式中必须唯一。

`source_encoding`

源编码名称。

`dest_encoding`

目标编码名称。

`funcname`

用于执行该转换的函数。函数名可以是模式限定的；如果不是，则会在搜索路径中查找该函数。

该函数必须具有以下签名：

```
conv_proc(
    integer,  -- source encoding ID
    integer,  -- destination encoding ID
    cstring,  -- source string (null terminated C string)
    internal, -- destination (fill with a null terminated C string)
    integer   -- source string length
) RETURNS void;
```

注解

使用`DROP CONVERSION`可以删除用户定义的转换。

创建转换所要求的权限在未来版本中可能会更改。

示例

使用`myfunc`创建从编码`UTF8`到 `LATIN1`的转换：

```
CREATE CONVERSION myconv FOR 'UTF8' TO 'LATIN1' FROM myfunc;
```

兼容性

`CREATE CONVERSION`是 PostgreSQL的一个扩展。SQL 标准中没有`CREATE CONVERSION`语句。

另见

ALTER CONVERSION, CREATE FUNCTION, DROP CONVERSION

CREATE DATABASE

CREATE DATABASE — 创建一个新数据库

大纲

```
CREATE DATABASE name
    [ [ WITH ] [ OWNER [=] dbowner ]
           [ TEMPLATE [=] template ]
           [ ENCODING [=] encoding ]
           [ TABLESPACE [=] tablespace ]
           [ CONNECTION LIMIT [=] connlimit ] ]
```

描述

`CREATE DATABASE` 创建一个新的 PostgreSQL 数据库。

要创建数据库，你必须是超级用户，或者拥有特殊的 `CREATEDB` 权限。见CREATE USER。

通常，创建者会成为新数据库的所有者。超级用户可以使用 `OWNER` 子句创建由其他用户拥有的数据库，甚至可以创建由不具有特殊权限的用户拥有的数据库。拥有 `CREATEDB` 权限的非超级用户只能创建由自己拥有的数据库。

默认情况下，新数据库通过克隆标准系统数据库 `template1` 来创建。可以通过写成 `TEMPLATE` *`name`* 指定其他模板。特别是，写成 `TEMPLATE template0` 时，可以创建一个全新的数据库，它只包含你的 PostgreSQL 版本预定义的标准对象。如果你希望避免复制任何可能已添加到 `template1` 中的站点本地附加对象，这会很有用。

参数

`name`

要创建的数据库名称。

`dbowner`

将拥有新数据库的数据库用户的名称，或使用 `DEFAULT` 以使用默认值（即执行命令的用户）。

`template`

用于创建新数据库的模板名称，或者指定 `DEFAULT` 以使用默认模板（`template1`）。

`encoding`

新数据库要使用的字符集编码。可指定字符串常量（例如 `'SQL_ASCII'`）、整数编码编号，或者指定 `DEFAULT` 以使用默认编码。PostgreSQL 服务器支持的字符集见第 21.2.1 节。

`tablespace`

将与新数据库关联的表空间名称，或者指定 `DEFAULT` 以使用模板数据库的表空间。该表空间将成为在此数据库中创建对象时使用的默认表空间。详见CREATE TABLESPACE。

connlimit

可对该数据库建立的并发连接数。-1（默认值）表示不受限制。

可选参数可以按任意顺序书写，不一定要按照上面的顺序。

注解

`CREATE DATABASE` 不能在事务块内执行。

形如“could not initialize database directory”的错误，多半与数据目录权限不足、磁盘已满或其他文件系统问题有关。

使用DROP DATABASE删除数据库。

程序createdb是这个命令的一个包装器程序，为方便使用而提供。

尽管可以通过把某个数据库名指定为模板，从而复制它而不是复制 `template1`，但这（至少目前）并不打算作为一种通用的“`COPY DATABASE`”功能。我们建议将用作模板的数据库视为只读。详见第 19.3 节。

`CONNECTION LIMIT` 选项只是近似地被强制执行；如果两个新会话几乎同时启动，而该数据库只剩下一个连接“槽”，则两者都可能失败。此外，该限制对超级用户无效。

示例

要创建一个新数据库：

```
CREATE DATABASE lusiadas;
```

要创建一个由用户 `salesapp` 拥有、默认表空间为 `salesspace` 的数据库 `sales`：

```
CREATE DATABASE sales OWNER salesapp TABLESPACE salesspace;
```

要创建支持 ISO-8859-1 字符集的数据库`music`：

```
CREATE DATABASE music ENCODING 'LATIN1';
```

兼容性

SQL 标准中没有 `CREATE DATABASE` 语句。数据库相当于目录，而目录的创建由实现定义。

CREATE DOMAIN

CREATE DOMAIN — 定义一个新域

大纲

```
CREATE DOMAIN name [AS] data_type
    [ DEFAULT expression ]
    [ constraint [ ... ] ]

where constraint is:

[ CONSTRAINT constraint_name ]
{ NOT NULL | NULL | CHECK (expression) }
```

描述

`CREATE DOMAIN`创建一个新的数据域。定义域的用户将成为其拥有者。

如果给定了模式名（例如`CREATE DOMAIN myschema.mydomain ...`），则该域会在指定模式中创建。否则它会在当前模式中创建。域名在其所在模式中的现有类型和域之间必须唯一。

域适合把多个表之间的公共字段抽象到单一位置进行维护。例如，
一个电子邮件地址列可能在多个表中使用，且都具有相同的属性。定义一个域并使用它，
比为每个表分别设置约束更合适。

> **小心**
>
> 目前，将函数结果值声明为域是相当危险的，
> 因为没有任何过程语言会对函数结果强制执行域约束。
> 你需要确保函数代码本身遵守这些约束。在PL/pgSQL中，
> 一种可行的变通办法是在返回结果值时将其显式转换为域类型。
> PL/pgSQL同样不会对函数内的局部变量强制执行域约束。

参数

name

要创建的域名（可选地带模式限定）。

data_type

该域的底层数据类型。它可以包含数组说明符。

`DEFAULT` *expression*

`DEFAULT`子句为该域数据类型的列指定默认值。
该值可以是任意不含变量的表达式（但不允许子查询）。
默认表达式的数据类型必须与该域的数据类型匹配。如果未指定默认值，则默认值为空值。

默认表达式会在任何未为该列指定值的插入操作中使用。如果为某个特定列定义了默认值，它就会覆盖与该域关联的任何默认值。反过来，域默认值又会覆盖与底层数据类型关联的任何默认值。

`CONSTRAINT constraint_name`

约束的可选名称。如果未指定，系统会生成一个名称。

`NOT NULL`

该域的值不允许为空值。

`NULL`

该域的值允许为空值。这是默认行为。

该子句仅用于与非标准 SQL 数据库兼容。不鼓励在新应用中使用它。

`CHECK (expression)`

`CHECK`子句指定该域的值必须满足的完整性约束或测试。每个约束都必须是一个产生布尔结果的表达式。它应使用关键字`VALUE`来引用被测试的值。

当前，`CHECK`表达式不能包含子查询，也不能引用 `VALUE`之外的其他变量。

示例

这个示例创建`us_postal_code`数据类型，然后在一个表定义中使用该类型。这里使用正则表达式测试来验证该值看起来是否为一个合法的美国邮政编码：

```
CREATE DOMAIN us_postal_code AS TEXT
CHECK(
   VALUE ~ '^\d{5}$'
OR VALUE ~ '^\d{5}-\d{4}$'
);

CREATE TABLE us_snail_addy (
  address_id SERIAL PRIMARY KEY,
  street1 TEXT NOT NULL,
  street2 TEXT,
  street3 TEXT,
  city TEXT NOT NULL,
  postal us_postal_code NOT NULL
);
```

兼容性

命令`CREATE DOMAIN`符合 SQL 标准。

另见

ALTER DOMAIN, DROP DOMAIN

CREATE FUNCTION

CREATE FUNCTION — 定义一个新函数

大纲

```
CREATE [ OR REPLACE ] FUNCTION
    name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    [ RETURNS rettype ]
  { LANGUAGE langname
    | IMMUTABLE | STABLE | VOLATILE
    | CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT
    | [ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER
    | AS 'definition'
    | AS 'obj_file', 'link_symbol'
  } ...
    [ WITH ( attribute [, ...] ) ]
```

描述

`CREATE FUNCTION`定义一个新函数。`CREATE OR REPLACE FUNCTION`将创建一个新函数，或者替换现有定义。

如果包含模式名，则函数将在指定的模式中创建。否则将在当前模式中创建。
新函数的名称不能与同一模式中具有相同输入参数类型的任何现有函数匹配。然而，
不同参数类型的函数可以共享一个名称（这称为重载）。

要更新一个现有函数的定义，可以使用`CREATE OR REPLACE FUNCTION`。但不能用这种方式更改函数的名称或者参数类型（如果尝试这样做，实际上就会创建一个新的不同函数）。此外，`CREATE OR REPLACE FUNCTION`也不允许更改现有函数的返回类型。要做到这一点，必须删除该函数并重新创建。（使用`OUT`参数时，这意味着除非删除该函数，否则不能更改任何`OUT`参数的名称或类型。）

如果删除函数后再重新创建，新函数就不再是旧函数的同一实体；
你将必须删除引用旧函数的现有规则、视图、触发器等。使用`CREATE OR REPLACE FUNCTION`可以在不破坏引用该函数的对象的情况下更改函数定义。

创建该函数的用户将成为该函数的拥有者。

参数

name

要创建的函数名称（可以被模式限定）。

argmode

参数的模式：`IN`、`OUT`或`INOUT`。如果省略，默认为`IN`。

argname

参数的名称。某些语言（目前只有 PL/pgSQL）允许在函数体中使用该名称。对于其他语言，就函数本身而言，输入参数的名称只是额外文档。但输出参数的名称很重要，

因为它定义了结果行类型中的列名。（如果省略输出参数的名称，
系统将选择一个默认列名。）

argtype

该函数参数（如果有）的数据类型（可以是模式限定的）。参数类型可以是基础类型、
复合类型或者域类型，也可以引用一个表列的类型。

根据实现语言的不同，也可能允许指定诸如cstring这样的“伪类型”。
伪类型表示实际参数类型要么没有被完整指定，要么不属于普通 SQL 数据类型集合。

写成*tablename*.*columnname*%TYPE即可引用一个列的类型。
使用这种特性有时有助于让函数独立于表定义的变化。

rettype

该函数的返回数据类型（可以是模式限定的）。返回类型可以是基础类型、
复合类型或者域类型，也可以引用一个表列的类型。根据实现语言的不同，
也可能允许指定诸如cstring这样的“伪类型”。

当存在OUT或INOUT参数时，可以省略RETURNS子句。如果写出该子句，
它必须与输出参数所隐含的结果类型一致：如果有多个输出参数，则为RECORD；
如果只有一个输出参数，则为该输出参数的类型。

SETOF修饰符表示该函数将返回一组项，而不是单个项。

写成*tablename*.*columnname*%TYPE即可引用一个列的类型。

langname

实现该函数所用语言的名称。它可以是SQL、C、internal，
也可以是用户定义的过程语言的名称。为了向后兼容，名称可以用单引号括起。

IMMUTABLE
STABLE
VOLATILE

这些属性告知系统，为了运行时优化，是否可以安全地用一次求值替代对函数的多次求值。
最多只能指定其中一个。如果这些属性都没有出现，则默认假定为VOLATILE。

IMMUTABLE表示该函数在给定相同参数值时总是返回相同结果；也就是说，
它不会执行数据库查找，也不会以其他方式使用未直接出现在其参数列表中的信息。
如果给出此选项，任何使用全常量参数对该函数的调用都可以立即替换为该函数值。

STABLE表示在一次表扫描内，该函数对于相同参数值会一致地返回相同结果，
但其结果可能在不同 SQL 语句之间发生变化。这适用于结果依赖于数据库查找、参数变量
（例如当前时区）等的函数。另请注意，current_timestamp函数族也属于稳定函数，
因为它们的值在一个事务内不会变化。

VOLATILE表示该函数的值即使在一次表扫描内也可能发生变化，因此无法进行任何优化。
从这个意义上说，真正不稳定的数据库函数相对较少；一些例子是random()、currval(
)、timeofday()。但请注意，任何有副作用的函数都必须归类为不稳定，
即使其结果相当可预测，也必须如此，以防其调用被优化掉；例如setval()。

更多细节见第 32.6 节。

CALLED ON NULL INPUT
RETURNS NULL ON NULL INPUT
STRICT

CALLED ON NULL INPUT（默认）表示当某些参数为空值时，仍会正常调用该函数。如果有需要，则由函数作者负责检查空值并作出适当响应。

RETURNS NULL ON NULL INPUT或STRICT表示只要任一参数为空值，该函数总是返回空值。如果指定了这个选项，那么在参数中出现空值时不会执行该函数，而是自动假定结果为空值。

[EXTERNAL] SECURITY INVOKER
[EXTERNAL] SECURITY DEFINER

SECURITY INVOKER表示该函数将以调用它的用户的权限执行。这是默认设置。SECURITY DEFINER指定该函数将以创建它的用户的权限执行。

为了符合 SQL，允许使用关键字EXTERNAL。但它是可选的，因为与 SQL 不同，这个特性适用于所有函数，而不仅仅是外部函数。

definition

一个定义该函数的字符串常量，其含义取决于所用语言。它可以是一个内部函数名称、一个对象文件的路径、一个 SQL 命令，或者用一种过程语言编写的文本。

obj_file, *link_symbol*

当 C 语言源代码中的函数名称与 SQL 函数名称不同时，这种形式的 AS 子句用于可动态加载的 C 语言函数。字符串 *obj_file* 是包含该动态可加载对象的文件名，而 *link_symbol* 是函数的链接符号，即 C 语言源代码中函数的名称。如果省略链接符号，则假定它与正在定义的 SQL 函数名称相同。

attribute

用于指定函数可选信息的历史方式。以下属性可以出现在此处：

isStrict

等同于STRICT或RETURNS NULL ON NULL INPUT。

isCachable

isCachable是IMMUTABLE的已废弃等价写法；出于向后兼容的原因，目前仍接受它。

属性名称不区分大小写。

注解

有关编写函数的详细信息，请参阅第 32.3 节。

允许使用完整的SQL类型语法来声明输入参数和返回值。不过，类型说明的某些细节（例如numeric类型的精度字段）由底层函数实现负责，CREATE FUNCTION命令会悄悄地忽略它们（即不识别也不强制执行）。

PostgreSQL 允许函数重载；也就是说，只要输入参数类型不同，同一个名称就可以用于多个不同的函数。但是，所有函数的 C 名称都必须不同，因此必须为重载的 C 函数指定不同的 C 名称（例如，把参数类型用作 C 名称的一部分）。

如果两个函数具有相同的名称和输入参数类型，它们被认为相同（不考虑任何OUT参数）。
因此这些声明会冲突：

```
CREATE FUNCTION foo(int) ...
CREATE FUNCTION foo(int, out text) ...
```

当重复的CREATE FUNCTION调用引用同一个对象文件时，该文件只装载一次。要卸载并重新装载该文件（例如在开发期间），可使用LOAD命令。

使用DROP FUNCTION可以移除用户定义的函数。

使用美元引用（见第 4.1.2.2 节）来书写函数定义字符串通常会更有帮助，而不是使用普通的单引号语法。如果没有美元引用，函数定义中的任何单引号或者反斜线都必须用双写来转义。

要定义函数，用户必须具有该语言上的USAGE权限。

当CREATE OR REPLACE FUNCTION被用来替换一个现有函数时，该函数的拥有权和权限不会改变。所有其他的函数属性会按照该命令中指定的或者隐含的值赋值。
必须拥有（包括成为拥有角色的成员）该函数才能替换它。

示例

下面给出一些简单示例，帮助你开始使用。有关更多信息和示例，请参见第 32.3 节。

```
CREATE FUNCTION add(integer, integer) RETURNS integer
    AS 'select $1 + $2;'
    LANGUAGE SQL
    IMMUTABLE
    RETURNS NULL ON NULL INPUT;
```

在PL/pgSQL中，使用参数名把一个整数加 1：

```
CREATE OR REPLACE FUNCTION increment(i integer) RETURNS integer AS $$
        BEGIN
                RETURN i + 1;
        END;
$$ LANGUAGE plpgsql;
```

返回一个包含多个输出参数的记录：

```
CREATE FUNCTION dup(in int, out f1 int, out f2 text)
    AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
    LANGUAGE SQL;

SELECT * FROM dup(42);
```

你也可以用一个显式命名的复合类型，更详细地表达同样的意思：

```
CREATE TYPE dup_result AS (f1 int, f2 text);

CREATE FUNCTION dup(int) RETURNS dup_result
    AS $$ SELECT $1, CAST($1 AS text) || ' is text' $$
    LANGUAGE SQL;

SELECT * FROM dup(42);
```

安全地编写 SECURITY DEFINER 函数

因为 SECURITY DEFINER 函数要以创建它的用户的权限执行，所以必须小心确保该函数不会被滥用。出于安全考虑，应将search_path设置为排除任何可被不受信任用户写入的模式。
这可以防止恶意用户创建对象来遮蔽该函数使用的对象。在这方面尤其重要的是临时表模式；
默认情况下它最先被搜索，而且通常任何人都可写。
一个安全的安排是强制把临时模式放到搜索顺序的最后。要做到这一点，应把 pg_temp 写成 search_path 中的最后一项。下面这个函数展示了安全用法：

```
CREATE FUNCTION check_password(uname TEXT, pass TEXT)
RETURNS BOOLEAN AS $$
DECLARE passed BOOLEAN;
        old_path TEXT;
BEGIN
        -- Save old search_path; notice we must qualify current_
setting
        -- to ensure we invoke the right function
        old_path := pg_catalog.current_setting('search_path');

        -- Set a secure search_path: trusted schemas, then 'pg_temp'.
        -- We set is_local = true so that the old value will be
 restored
        -- in event of an error before we reach the function end.
        PERFORM pg_catalog.set_config('search_path', 'admin, pg_temp',
 true);

        -- Do whatever secure work we came for.
        SELECT  (pwd = $2) INTO passed
        FROM    pwds
        WHERE   username = $1;

        -- Restore caller's search_path
        PERFORM pg_catalog.set_config('search_path', old_path, true);

        RETURN passed;
END;
$$ LANGUAGE plpgsql SECURITY DEFINER;
```

兼容性

CREATE FUNCTION命令在 SQL:1999 及更高版本中定义。PostgreSQL版本与其相似，但并不完全兼容。这些属性和可用的不同语言都不具备可移植性。

为与某些其他数据库系统兼容，*argmode*可以写在*argname*之前或之后，
但只有前一种写法符合标准。

另见

ALTER FUNCTION, DROP FUNCTION, GRANT, LOAD, REVOKE, createlang

CREATE GROUP

CREATE GROUP — 定义一个新的数据库角色

大纲

```
CREATE GROUP name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE rolename [, ...]
    | IN GROUP rolename [, ...]
    | ROLE rolename [, ...]
    | ADMIN rolename [, ...]
    | USER rolename [, ...]
    | SYSID uid
```

描述

`CREATE GROUP` 现在是CREATE ROLE的别名。

兼容性

SQL 标准中没有 `CREATE GROUP` 语句。

另见

CREATE ROLE

CREATE INDEX

CREATE INDEX — 定义一个新索引

大纲

```
CREATE [ UNIQUE ] INDEX name ON table [ USING method ]
    ( { column | ( expression ) } [ opclass ] [, ...] )
    [ TABLESPACE tablespace ]
    [ WHERE predicate ]
```

描述

`CREATE INDEX` 在指定表上构造一个索引 *index_name*。索引主要用于提升数据库性能（但使用不当也可能导致性能下降）。

索引的键字段指定为列名，或者指定为写在圆括号中的表达式。如果索引方法支持多列索引，则可以指定多个字段。

索引字段可以是根据表行中一个或多个列值计算得到的表达式。
该特性可用于根据基础数据的某种变换来快速访问数据。例如，在`upper(col)`
上计算的索引可让子句 `WHERE upper(col) = 'JIM'`使用索引。

PostgreSQL提供了四种索引方法：B-树、R-树、hash 和 GiST。B-树索引方法是 Lehman-Yao 高并发 B-树的一种实现。R-树索引方法使用 Guttman 的二次分割算法实现标准 R-树。hash 索引方法是 Litwin 线性散列的一种实现。用户也可以定义自己的索引方法，但这相当复杂。

当`WHERE`子句存在时，会创建一个部分索引。部分索引只包含表中一部分行的索引项，
通常这一部分比表的其余部分更适合建立索引。例如，
如果一个表同时包含已开票和未开票订单，而未开票订单只占整个表的一小部分，
但这一部分又经常被访问，就可以只对这部分创建索引来提升性能。另一种可能的应用是将
`WHERE`与`UNIQUE`结合使用，以便在表的一个子集上强制唯一性。更多讨论请见第 11.7 节。

`WHERE`子句中使用的表达式只能引用底层表的列，但它可以使用所有列，
而不仅仅是被索引的列。当前，`WHERE`中也禁止使用子查询和聚合表达式。
同样的限制也适用于作为表达式的索引字段。

所有在索引定义中使用的函数和操作符都必须是“不可变的”，也就是说，
它们的结果只能依赖其参数，而不能受任何外部因素影响（例如另一个表的内容或当前时间）。
这种限制确保索引的行为定义明确。要在索引表达式或`WHERE`子句中使用用户定义的函数，
记得在创建该函数时将其标记为不可变。

参数

`UNIQUE`

使系统在创建索引时（如果数据已经存在）以及每次添加数据时，检查表中的重复值。
任何会导致重复项的插入或更新操作都会报错。

name

要创建的索引名称。此处不能包含模式名称；索引始终创建在其父表所在的同一模式中。

table

要建立索引的表名（可以是模式限定名）。

method

要使用的索引方法的名称。选择包括 `btree`、`hash`、`rtree` 和 `gist`。默认方法是 `btree`。

column

一个表列的名称。

expression

一个基于表中一个或多个列的表达式。通常必须像语法中所示那样写在外围圆括号中。不过，如果该表达式是函数调用形式，则可以省略圆括号。

opclass

一个操作符类的名称。详见下文。

tablespace

在其中创建索引的表空间。如果未指定，将查阅default_tablespace；如果 `default_tablespace`是空字符串，则使用数据库的默认表空间。

predicate

部分索引的约束表达式。

注解

关于索引何时能被使用、何时不被使用以及什么情况下它们有用的信息请见第 11 章。

目前，只有 B-树和 GiST 索引方法支持多列索引。默认最多可指定 32 个字段。（构建PostgreSQL时可以更改此限制。）目前只有 B-树支持唯一索引。

索引的每一列都可以指定一个操作符类。操作符类标识索引用于该列的操作符。例如，四字节整数上的 B-树索引会使用`int4_ops`类；此操作符类包含四字节整数的比较函数。实际上，列数据类型的默认操作符类通常已足够。操作符类存在的主要原因是，某些数据类型可能有多个有意义的排序方式。例如，我们可能希望按绝对值或实部对复数数据类型排序。我们可以为该数据类型定义两个操作符类，并在创建索引时选择合适的类。有关操作符类的更多信息，请参见第 11.8 节和第 32.14 节。

使用DROP INDEX删除索引。

默认情况下，索引不会被用于`IS NULL`子句。在这种情况下使用索引的最佳办法，是创建一个使用`IS NULL`谓词的部分索引。

示例

创建 B-树索引，列为`title`，所在表为`films`：

```
CREATE UNIQUE INDEX title_idx ON films (title);
```

要在表`films`的列`code`上创建一个索引，并让该索引驻留在表空间`indexspace`中：

```
CREATE INDEX code_idx ON films (code) TABLESPACE indexspace;
```

兼容性

`CREATE INDEX`是 PostgreSQL的语言扩展。SQL 标准中没有关于索引的规定。

另见

ALTER INDEX, DROP INDEX

CREATE LANGUAGE

CREATE LANGUAGE — 定义一种新的过程语言

大纲

```
CREATE [ PROCEDURAL ] LANGUAGE name
CREATE [ TRUSTED ] [ PROCEDURAL ] LANGUAGE name
    HANDLER call_handler [ VALIDATOR valfunction ]
```

描述

使用`CREATE LANGUAGE`，PostgreSQL用户可以在PostgreSQL数据库中注册一种新的过程语言。随后，可以用这种新语言定义函数和触发器过程。用户必须具有PostgreSQL超级用户权限才能注册新语言。

`CREATE LANGUAGE`实际上是将该语言名称与负责执行用这种语言编写的函数的处理器函数关联起来。关于语言处理器的更多信息，参见第 35 章。

`CREATE LANGUAGE`命令有两种形式。第一种形式中，用户只提供所需语言的名称，PostgreSQL服务器会查询`pg_pltemplate`系统目录，以确定正确的参数。第二种形式中，用户在提供语言名称的同时也提供语言参数。第二种形式可用来创建未在`pg_pltemplate`中定义的语言，但这种做法被认为已经过时。

当服务器在`pg_pltemplate`目录中找到给定语言名称的条目时，即使命令包含语言参数，也会使用目录中的数据。这种行为简化了旧转储文件的装载，因为它们很可能包含关于语言支持函数的过时信息。

参数

`TRUSTED`

`TRUSTED`指定该语言的调用处理器是安全的，也就是说，它不会让非特权用户获得任何绕过访问限制的功能。如果在注册该语言时省略这个关键字，只有拥有PostgreSQL超级用户权限的用户才能使用该语言创建新函数。

`PROCEDURAL`

这是一个噪声词。

`name`

新过程语言的名称。语言名称不区分大小写。该名称必须在与数据库中的其他语言名称不同。

为了向后兼容，名称可以用单引号括起来。

`HANDLER` *`call_handler`*

`call_handler` 是一个先前注册的函数名称，它会被调用来执行该过程语言中的函数。过程语言的调用处理器必须使用某种编译型语言（例如 C）编写，并采用版本 1 调用约定。它必须

在PostgreSQL 中注册为一个不带参数且返回language_handler类型的函数。该返回类型是一种占位符类型，仅用于把该函数标识为调用处理器。

VALIDATOR *valfunction*

valfunction 是一个先前注册的函数名称，在创建该语言中的新函数时会调用它来验证该新函数。如果没有指定验证器函数，那么新函数在创建时不会被检查。
验证器函数必须接受一个oid类型的参数，它将是待创建函数的 OID，并且它通常返回void。

验证器函数通常会检查函数体的语法正确性，但它也可以检查函数的其他属性，
例如该语言是否支持某些参数类型。要报告错误，验证器函数应使用ereport()函数。
函数的返回值会被忽略。

TRUSTED选项和支持函数名称会被忽略，前提是服务器在pg_pltemplate
中有指定语言名称的条目。

注解

createlang程序是对CREATE LANGUAGE命令的一个简单包装。它使从 shell 命令行安装过程语言变得更加容易。

使用DROP LANGUAGE删除过程语言，或者更好的方法是使用droplang程序。

系统目录pg_language（见第 42.20 节）记录当前已安装语言的信息。此外，createlang有一个可以列出已安装语言的选项。

要用某种过程语言创建函数，用户必须拥有该语言的 USAGE权限。默认情况下，对于受信任的语言，USAGE会授予PUBLIC（即所有人）。如果需要，也可以撤销这项授权。

过程语言是各个数据库本地的。不过，可以将某种语言安装到 template1数据库中，这样它就会在之后创建的所有数据库中自动可用。

如果服务器在pg_pltemplate中没有该语言的条目，
则调用处理器函数和验证器函数（如果有）必须已经存在。但如果有条目，
则这些函数不必事先存在；如果数据库中不存在它们，就会自动定义它们。
（如果安装中没有实现该语言的共享库，这可能导致CREATE LANGUAGE失败。）

在 7.3 之前的PostgreSQL版本中，必须将处理器函数声明为返回占位符类型opaque，而不是language_handler。为了支持装载旧转储文件，CREATE LANGUAGE会接受声明为返回opaque的函数，但会发出提示，并将函数声明的返回类型改为language_handler。

示例

创建任何标准过程语言的首选方式很简单：

```
CREATE LANGUAGE plperl;
```

对于 pg_pltemplate 目录中未知的语言，需要类似以下的命令序列：

```
CREATE FUNCTION plsample_call_handler() RETURNS language_handler
    AS '$libdir/plsample'
    LANGUAGE C;
CREATE LANGUAGE plsample
    HANDLER plsample_call_handler;
```

兼容性

`CREATE LANGUAGE`是一种 PostgreSQL扩展。

另见

ALTER LANGUAGE, CREATE FUNCTION, DROP LANGUAGE, GRANT, REVOKE, createlang, droplang

CREATE OPERATOR

CREATE OPERATOR — 定义一个新的操作符

大纲

```
CREATE OPERATOR name (
    PROCEDURE = funcname
    [, LEFTARG = lefttype ] [, RIGHTARG = righttype ]
    [, COMMUTATOR = com_op ] [, NEGATOR = neg_op ]
    [, RESTRICT = res_proc ] [, JOIN = join_proc ]
    [, HASHES ] [, MERGES ]
    [, SORT1 = left_sort_op ] [, SORT2 = right_sort_op ]
    [, LTCMP = less_than_op ] [, GTCMP = greater_than_op ]
)
```

描述

`CREATE OPERATOR`定义一个新的操作符 *name*。定义该操作符的用户将成为其拥有者。如果给出了模式名，该操作符将在指定的模式中创建；否则将在当前模式中创建。

操作符名称是一个字符序列，最多包含 `NAMEDATALEN`-1 个字符（默认为 63 个），字符取自以下列表：

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

名称的选择有以下几项限制：

- `--`和`/*`不能出现在操作符名称的任何位置，因为它们会被视为注释的开始。
- 多字符操作符名称不能以`+`或`-` 结尾，除非该名称中还至少包含下列字符之一：

  ~ ! @ # % ^ & | ` ?

  例如，`@-`是允许的操作符名称，而 `*-`则不允许。这一限制使 PostgreSQL能够在不要求记号之间必须有空格的情况下解析符合 SQL 规范的命令。

输入`!=`时会被映射为`<>`，因此这两个名称始终等效。

必须至少定义`LEFTARG`和`RIGHTARG`中的一个。对于二元操作符，必须同时定义两者。
对于右一元操作符，只应定义`LEFTARG`；对于左一元操作符，只应定义`RIGHTARG`。

*funcname*过程必须事先已用`CREATE FUNCTION`定义，并且其定义必须接受正确数量（一个或两个）的指定类型参数。

其他子句用于指定可选的操作符优化属性。其含义详见第 32.13 节。

参数

name

要定义的操作符名称。允许使用的字符请见上文。该名称可以带模式限定，例如`CREATE OPERATOR myschema.+ (...)`。如果不带模式限定，该操作符将在当前模式中创建。同一

模式中的两个操作符如果针对不同的数据类型进行操作，可以具有相同的名称。这被称为重载。

funcname

用于实现该操作符的函数。

lefttype

该操作符左操作数的数据类型（如果有）。对于左一元操作符，应省略此选项。

righttype

该操作符右操作数的数据类型（如果有）。对于右一元操作符，应省略此选项。

com_op

该操作符的交换子。

neg_op

该操作符的求反器。

res_proc

该操作符的限制选择度估计函数。

join_proc

该操作符的连接选择度估算函数。

HASHES

表示该操作符可以支持哈希连接。

MERGES

表示该操作符可以支持归并连接。

left_sort_op

如果该操作符可以支持归并连接，则为对该操作符左手数据类型排序的小于操作符。

right_sort_op

如果该操作符可以支持归并连接，则为对该操作符右手数据类型排序的小于操作符。

less_than_op

如果该操作符可以支持归并连接，则为比较该操作符输入数据类型的小于操作符。

greater_than_op

如果该操作符可以支持归并连接，则为比较该操作符输入数据类型的大于操作符。

要在*com_op*或其他可选参数中给出带模式限定的操作符名称，请使用OPERATOR()语法，例如：

```
COMMUTATOR = OPERATOR(myschema.===) ,
```

注解

更多信息请参见第 32.12 节。

使用DROP OPERATOR从数据库中删除用户定义的操作符。使用ALTER OPERATOR修改数据库中的操作符。

示例

以下命令定义一个新的面积相等操作符，适用的数据类型为 `box`：

```
CREATE OPERATOR === (
    LEFTARG = box,
    RIGHTARG = box,
    PROCEDURE = area_equal_procedure,
    COMMUTATOR = ===,
    NEGATOR = !==,
    RESTRICT = area_restriction_procedure,
    JOIN = area_join_procedure,
    HASHES, MERGES
);
```

兼容性

`CREATE OPERATOR`是 PostgreSQL扩展。在 SQL 标准中没有用户定义操作符的规定。

另见

ALTER OPERATOR, CREATE OPERATOR CLASS, DROP OPERATOR

CREATE OPERATOR CLASS

CREATE OPERATOR CLASS — 定义一个新的操作符类

大纲

```
CREATE OPERATOR CLASS name [ DEFAULT ] FOR TYPE data_type USING index_
method AS
  {  OPERATOR strategy_number operator_name [ ( op_type, op_type ) ] [
 RECHECK ]
   | FUNCTION support_number funcname ( argument_type [, ...] )
   | STORAGE storage_type
  } [, ... ]
```

描述

CREATE OPERATOR CLASS创建一个新的操作符类。操作符类定义特定数据类型如何用于索引。它指定某些操作符为该数据类型和索引方法承担特定角色或“策略”。
它还指定为索引列选择该操作符类时，索引方法要使用的支持函数。在创建操作符类之前，必须先定义它使用的所有操作符和函数。

如果给出了一个模式名称，那么该操作符类会被创建在指定模式中。否则，它会被创建在当前模式中。
同一模式中的两个操作符类只有在被用于不同的索引方法时才可以具有相同的名称。

定义操作符类的用户将成为其拥有者。目前，创建者必须是超级用户。（之所以有此限制，是因为错误的操作符类定义可能会使服务器混乱，甚至崩溃。）

CREATE OPERATOR CLASS当前不会检查操作符类定义是否包括该索引方法所要求的所有操作符和函数，也不会检查这些操作符和函数是否构成一个自洽的集合。
定义一个合法的操作符类是用户的责任。

进一步的信息可参考第 32.14 节。

参数

name

要创建的操作符类的名称。该名称可以被模式限定。

DEFAULT

如果出现，该操作符类将成为其数据类型的默认操作符类。
对于某个特定的数据类型和索引方法，最多只能有一个默认操作符类。

data_type

该操作符类对应的列数据类型。

index_method

该操作符类对应的索引方法名称。

strategy_number

与该操作符类关联的操作符在此索引方法中的策略号。

operator_name

与该操作符类关联的操作符名称（可以是模式限定的）。

op_type

操作符的操作数数据类型，或者用NONE表示一个左一元或右一元操作符。在通常情况下，如果操作数数据类型与操作符类的数据类型相同，则可以省略操作数数据类型。

RECHECK

如果存在，则表示索引对该操作符是“有损”的，因此使用索引检索到的行必须重新检查，以验证它们确实满足涉及该操作符的限定条件。

support_number

与操作符类关联的函数在索引方法中的支持函数编号。

funcname

作为该操作符类的索引方法支持函数的函数名称（可带模式限定）。

argument_types

该函数的参数数据类型。

storage_type

实际存储在索引中的数据类型。通常它与列数据类型相同，但是有些索引方法（目前是 GiST 和 GIN）允许它们不同。除非索引方法允许使用不同类型，否则必须省略STORAGE子句。

OPERATOR、FUNCTION和STORAGE 子句可以以任何顺序出现。

注解

操作符不应该由 SQL 函数定义。SQL 函数很有可能会被内联到调用查询中，这会妨碍优化器识别该查询匹配一个索引。

示例

下面的示例命令为数据类型_int4（int4数组）定义了一个 GiST 索引操作符类。完整示例见contrib/intarray/。

```
CREATE OPERATOR CLASS gist__int_ops
    DEFAULT FOR TYPE _int4 USING gist AS
        OPERATOR        3       &&,
        OPERATOR        6       =       RECHECK,
        OPERATOR        7       @>,
        OPERATOR        8       <@,
        OPERATOR        20      @@ (_int4, query_int),
        FUNCTION        1       g_int_consistent (internal, _int4,
 int, oid, internal),
```

```
        FUNCTION        2       g_int_union (internal, internal),
        FUNCTION        3       g_int_compress (internal),
        FUNCTION        4       g_int_decompress (internal),
        FUNCTION        5       g_int_penalty (internal, internal,
 internal),
        FUNCTION        6       g_int_picksplit (internal, internal),
        FUNCTION        7       g_int_same (_int4, _int4, internal);
```

兼容性

`CREATE OPERATOR CLASS`是一种 PostgreSQL扩展。在 SQL 标准中没有 `CREATE OPERATOR CLASS`语句。

另见

ALTER OPERATOR CLASS, DROP OPERATOR CLASS

CREATE ROLE

CREATE ROLE — 定义一个新的数据库角色

大纲

```
CREATE ROLE name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE rolename [, ...]
    | IN GROUP rolename [, ...]
    | ROLE rolename [, ...]
    | ADMIN rolename [, ...]
    | USER rolename [, ...]
    | SYSID uid
```

描述

`CREATE ROLE`向 PostgreSQL数据库集簇中新增一个角色。角色是一种可以拥有数据库对象并拥有数据库权限的实体；根据其使用方式，它既可以视为“用户”，也可以视为“组”，或者兼具这两种身份。有关用户管理和认证的信息，请参见第 18 章和第 20 章。要使用该命令，你必须拥有 `CREATEROLE`权限，或者是数据库超级用户。

注意，角色是在数据库集簇级别定义的，因此在该集簇中的所有数据库里都有效。

参数

name

新角色的名称。

`SUPERUSER`
`NOSUPERUSER`

这些子句决定新角色是否为“超级用户”；超级用户可以绕过数据库中的所有访问限制。超级用户身份具有危险性，只有在确有需要时才应使用。要创建新的超级用户，你自己必须先是超级用户。若未指定，默认值是 `NOSUPERUSER`。

CREATEDB
NOCREATEDB

这些子句定义角色创建数据库的能力。如果指定CREATEDB，
则允许正在定义的角色创建新数据库。指定NOCREATEDB会禁止角色创建数据库。
如果未指定，默认值为NOCREATEDB。

CREATEROLE
NOCREATEROLE

这些子句决定是否允许角色创建新角色（即执行CREATE ROLE）。具有CREATEROLE权限的角色还可以更改和删除其他角色。如果未指定，默认值为NOCREATEROLE。

CREATEUSER
NOCREATEUSER

这些子句是 SUPERUSER 和 NOSUPERUSER 的一种已过时但仍被接受的写法。注意，它们不等价于 CREATEROLE，尽管可能有人会天真地这么认为！

INHERIT
NOINHERIT

这些子句决定角色是否“继承”其所属角色的权限。具有INHERIT
属性的角色可以自动使用直接或间接所属的所有角色获得的任何数据库权限。没有INHERIT时，成为另一个角色的成员只授予对该角色执行SET ROLE的能力；另一个角色的权限只有在执行该操作后才可用。如果未指定，默认值为INHERIT。

LOGIN
NOLOGIN

这些子句决定是否允许角色登录；也就是说，在客户端连接期间，
是否可以将该角色作为初始会话授权名称。具有 LOGIN 属性的角色可以视为用户。没有该属性的角色可用于管理数据库权限，但通常意义上不属于用户。如果未指定，默认值为NOLOGIN，但通过其另一种拼写 CREATE USER 调用 CREATE ROLE 时除外。

CONNECTION LIMIT *connlimit*

如果该角色可以登录，此项指定该角色可建立多少并发连接。-1（默认值）表示不限制。
注意，只有普通连接会计入此限制。

PASSWORD *password*

设置角色的密码。（密码只对具有LOGIN属性的角色有用，但即使没有该属性，
你仍然可以为角色定义密码。）如果你不打算使用密码认证，可以省略此选项。

ENCRYPTED
UNENCRYPTED

这些关键字控制密码是否以加密形式存储在系统目录中。（如果两者都未指定，
默认行为由配置参数password_encryption决定。）如果给出的密码字符串已经是 MD5 加密格式，那么无论指定的是ENCRYPTED还是 UNENCRYPTED，它都会按原样以加密形式存储（因为系统无法解密指定的已加密密码字符串）。这样便可在转储/
恢复期间重新装载已加密的密码。

注意，较老的客户端可能缺少对 MD5 认证机制的支持，而处理加密存储的密码需要这种机制。

VALID UNTIL '*timestamp*'

VALID UNTIL子句设置一个日期和时间，在此之后角色的密码将不再有效。如果省略此子句，密码将始终有效。

IN ROLE *rolename*

IN ROLE子句列出一个或多个现有角色，新角色会立即作为新成员加入这些角色。（请注意，无法通过此选项将新角色添加为管理员；要实现这一点，请使用单独的GRANT命令。）

IN GROUP *rolename*

IN GROUP是IN ROLE的已废弃写法。

ROLE *rolename*

ROLE子句列出一个或多个现有角色，这些角色会自动作为新角色的成员加入。（实际上，这会使新角色成为一个“组”。）

ADMIN *rolename*

ADMIN子句类似于ROLE，但指定的角色会以WITH ADMIN OPTION加入新角色，从而获得向其他人授予该角色成员资格的权利。

USER *rolename*

USER子句是ROLE子句的已废弃写法。

SYSID *uid*

SYSID子句会被忽略，但为向后兼容而接受。

注解

使用ALTER ROLE更改角色属性，并使用DROP ROLE删除角色。CREATE ROLE指定的所有属性都可以由后续的ALTER ROLE命令修改。

向作为组使用的角色添加和移除成员，推荐使用GRANT和REVOKE。

VALID UNTIL子句只为密码定义过期时间，而不是为角色本身定义过期时间。特别是，使用非基于密码的认证方式登录时，不会强制执行该过期时间。

INHERIT属性控制可授予权限的继承（即数据库对象的访问权限和角色成员资格）。它不适用于CREATE ROLE和ALTER ROLE设置的特殊角色属性。例如，即使设置了INHERIT，成为拥有CREATEDB权限的角色成员也不会立即获得创建数据库的能力；在创建数据库之前，必须先通过SET ROLE成为该角色。

INHERIT属性出于向后兼容的原因而成为默认值：在PostgreSQL的早期版本中，用户总能访问其所属组的所有权限。不过，NOINHERIT更接近 SQL 标准规定的语义。

请谨慎对待CREATEROLE权限。CREATEROLE角色的权限不存在继承这一概念。这意味着，即使某个角色没有特定权限但被允许创建其他角色，它仍可轻易创建一个权限不同于自身的角色（创建具有超级用户权限的角色除外）。例如，如果角色“user”具有CREATEROLE权限但没有CREATEDB权限，它仍然可以创建一个具有CREATEDB权限的新角色。因此，应将拥有CREATEROLE权限的角色视为几乎等同超级用户的角色。

PostgreSQL包含一个程序createuser，其功能与`CREATE ROLE`相同（实际上它就是调用此命令），但它可以从命令 shell 中运行。

`CONNECTION LIMIT`选项只是近似地被强制执行；如果两个新会话在几乎同一时刻启动，而此时该角色只剩下一个连接“槽”，则两者都可能失败。此外，该限制从不对超级用户强制执行。

示例

创建一个可登录但没有密码的角色：

```
CREATE ROLE jonathan LOGIN;
```

创建一个带密码的角色：

```
CREATE USER davide WITH PASSWORD 'jw8s0F4';
```

（`CREATE USER`与`CREATE ROLE`相同，只是它隐含指定了`LOGIN`。）

创建一个密码有效期截止到 2004 年底的角色。当 2005 年的第一秒过去后，该密码就不再有效。

```
CREATE ROLE miriam WITH LOGIN PASSWORD 'jw8s0F4' VALID UNTIL '2005-01
-01';
```

创建一个可以创建数据库并管理角色的角色：

```
CREATE ROLE admin WITH CREATEDB CREATEROLE;
```

兼容性

该`CREATE ROLE`语句在 SQL 标准中有定义，但该标准只要求以下语法

```
CREATE ROLE name [ WITH ADMIN rolename ]
```

多个初始管理员，以及`CREATE ROLE`的所有其他选项，都是PostgreSQL扩展。

SQL 标准定义了用户和角色的概念，但将它们视为不同的概念，并将定义用户的所有命令留给各数据库实现自行规定。在PostgreSQL中，我们选择将用户和角色统一为一种实体。因此，与标准相比，角色具有更多可选属性。

将用户赋予 `NOINHERIT` 属性，而将角色赋予 `INHERIT` 属性，最接近 SQL 标准规定的行为。

参见

SET ROLE, ALTER ROLE, DROP ROLE, GRANT, REVOKE, createuser

CREATE RULE

CREATE RULE — 定义一条新的重写规则

大纲

```
CREATE [ OR REPLACE ] RULE name AS ON event
    TO table [ WHERE condition ]
    DO [ ALSO | INSTEAD ] { NOTHING | command | ( command ; command .
.. ) }
```

描述

CREATE RULE定义一条应用于指定表或视图的新规则。CREATE OR REPLACE RULE将创建一条新规则，或者替换同一个表上同名的现有规则。

PostgreSQL规则系统允许定义在数据库表上执行插入、更新或删除时的替代动作。粗略地说，当在某个给定表上执行某个命令时，一条规则会导致额外的命令被执行。另一种情况是，INSTEAD 规则可以用另一个命令替换给定命令，或者让某个命令完全不执行。规则也用于实现SQL 视图。必须认识到，规则实际上是一种命令转换机制，或者说是命令宏。这种转换发生在命令开始执行之前。如果你真正需要的是对每个物理行独立触发的操作，那么你很可能应该使用触发器，而不是规则。有关规则系统的更多信息见第 34 章。

目前，ON SELECT规则必须是无条件的INSTEAD规则，其动作必须由单个SELECT命令组成。因此，ON SELECT规则实际上会把表变为视图，其可见内容是该规则的SELECT命令返回的行，而不是表中原来存储的内容（如果有）。直接编写CREATE VIEW命令，被认为比创建真实表并为其定义ON SELECT规则的风格更好。

可以通过定义ON INSERT、ON UPDATE 和ON DELETE规则（或其中足以满足你需求的任意子集），来营造出可更新视图的效果，从而将视图上的更新动作替换为对其他表的适当更新。

如果你尝试对复杂视图更新使用有条件规则，这里有一个陷阱：对于你希望允许在该视图上执行的每一种动作，必须都有一条无条件的INSTEAD规则。如果该规则是有条件的，或者不是 INSTEAD规则，那么系统仍会拒绝执行该更新动作，因为它认为在某些情况下最终可能还是会尝试在该视图的虚拟表上执行该动作。如果你想用有条件规则处理所有有用的情况，可以加上一条无条件的 DO INSTEAD NOTHING规则，以确保系统明白它永远不会被要求去更新这个虚拟表。然后，把这些有条件规则改成非INSTEAD规则；在它们适用的情况下，它们会附加在默认的 INSTEAD NOTHING动作之上。

参数

name

要创建的规则的名称。它必须与同一个表上任何其他规则的名称相区分。同一个表上同一种事件类型的多条规则会按名称的字母顺序应用。

event

事件是SELECT、INSERT、UPDATE或者 DELETE之一。

table

规则适用的表或者视图的名称（可以是模式限定的）。

condition

任意的SQL条件表达式（返回 boolean）。条件表达式不能引用NEW和 OLD之外的任何表，并且不能包含聚合函数。

INSTEAD

INSTEAD表示这些命令应当替代原始命令执行。

ALSO

ALSO表示这些命令应当在原始命令之外执行。

如果既未指定ALSO，也未指定 INSTEAD，则默认值为ALSO。

command

组成规则动作的命令。可用的命令有SELECT、INSERT、UPDATE、DELETE或者NOTIFY。

在*condition*和 *command*中，可以使用特殊表名 NEW和OLD来引用目标表中的值。在ON INSERT和ON UPDATE规则中，NEW可用于引用待插入或待更新的新行。在 ON UPDATE和ON DELETE规则中，OLD可用于引用将被更新或删除的现有行。

注解

必须在表上具有RULE权限，才允许在表上定义规则。

避免循环规则非常重要。例如，尽管下面的两条规则定义都被 PostgreSQL所接受，SELECT命令将导致 PostgreSQL报告错误，因为规则会发生递归展开：

```
CREATE RULE "_RETURN" AS
    ON SELECT TO t1
    DO INSTEAD
        SELECT * FROM t2;

CREATE RULE "_RETURN" AS
    ON SELECT TO t2
    DO INSTEAD
        SELECT * FROM t1;

SELECT * FROM t1;
```

当前，如果一个规则动作包含一个NOTIFY命令，该NOTIFY命令将被无条件执行，也就是说，即使没有任何应当应用该规则的行，也会发出NOTIFY。例如：

```
CREATE RULE notify_me AS ON UPDATE TO mytable DO ALSO NOTIFY mytable;

UPDATE mytable SET name = 'foo' WHERE id = 42;
```

在UPDATE期间会发送一个NOTIFY事件，无论是否存在匹配条件id = 42的行。这是一个实现限制，未来版本中可能会修复。

兼容性

CREATE RULE是一种 PostgreSQL语言扩展，整个查询重写系统也是如此。

CREATE SCHEMA

CREATE SCHEMA — 定义一个新模式

大纲

```
CREATE SCHEMA schemaname [ AUTHORIZATION username ] [ schema_element [
 ... ] ]
CREATE SCHEMA AUTHORIZATION username [ schema_element [ ... ] ]
```

描述

CREATE SCHEMA在当前数据库中创建一个新模式。该模式名必须与当前数据库中任何现有模式的名称不同。

模式本质上是一个命名空间：它包含具名对象（表、数据类型、函数和操作符），这些对象的名称可以与其他模式中的对象重名。访问具名对象时，可以在其名称前加上模式名作为前缀来“限定”名称，或者设置一个包含所需模式的搜索路径。指定非限定对象名的CREATE命令会在当前模式中创建该对象（即搜索路径最前面的模式，可通过函数current_schema确定）。

CREATE SCHEMA还可以选择包含子命令，以便在新模式中创建对象。这些子命令基本上会被当作在创建模式之后单独发出的命令来处理。不过，如果使用了AUTHORIZATION子句，则所有创建的对象都将归该用户所有。

参数

schemaname

要创建的模式名称。如果省略，用户名将被用作模式名。该名称不能以 pg_开头，因为这类名称保留给系统模式。

username

将拥有新模式的用户的角色名。如果省略，则默认为执行该命令的用户。
要创建由另一个角色拥有的模式，你必须是该角色的直接或间接成员，或者是超级用户。

schema_element

定义要在该模式中创建的对象的 SQL 语句。当前，只有CREATE TABLE、CREATE VIEW、CREATE INDEX、CREATE SEQUENCE、CREATE TRIGGER和GRANT可作为 CREATE SCHEMA中的子句使用。其他类型的对象可以在模式创建之后通过单独的命令创建。

注解

要创建一个模式，执行该命令的用户必须拥有当前数据库的CREATE 权限（当然，超级用户可以绕过这项检查）。

示例

创建一个模式：

```
CREATE SCHEMA myschema;
```

为用户joe创建一个模式，该模式也将被命名为 joe：

```
CREATE SCHEMA AUTHORIZATION joe;
```

创建一个模式并且在其中创建一个表和视图：

```
CREATE SCHEMA hollywood
    CREATE TABLE films (title text, release date, awards text[])
    CREATE VIEW winners AS
        SELECT title, release FROM films WHERE awards IS NOT NULL;
```

请注意，各个子命令末尾都不带分号。

下面是实现相同结果的一种等效写法：

```
CREATE SCHEMA hollywood;
CREATE TABLE hollywood.films (title text, release date, awards text[]
);
CREATE VIEW hollywood.winners AS
    SELECT title, release FROM hollywood.films WHERE awards IS NOT
 NULL;
```

兼容性

SQL 标准允许在CREATE SCHEMA中使用 DEFAULT CHARACTER SET子句，并允许使用比PostgreSQL当前所接受的更多子命令类型。

SQL 标准规定，CREATE SCHEMA中的子命令可以按任意顺序出现。当前的PostgreSQL实现并不能处理子命令中所有前向引用的情况；有时可能需要重新排列子命令的顺序，以避免前向引用。

根据 SQL 标准，模式的拥有者总是拥有其中的所有对象。PostgreSQL允许模式包含由模式拥有者之外的用户所拥有的对象。只有当模式拥有者把其模式上的CREATE权限授予其他人，或者超级用户选择在该模式中创建对象时，才会发生这种情况。

另见

ALTER SCHEMA, DROP SCHEMA

CREATE SEQUENCE

CREATE SEQUENCE — 定义一个新的序列发生器

大纲

```
CREATE [ TEMPORARY | TEMP ] SEQUENCE name [ INCREMENT [ BY ] increment
 ]
    [ MINVALUE minvalue | NO MINVALUE ] [ MAXVALUE maxvalue | NO
 MAXVALUE ]
    [ START [ WITH ] start ] [ CACHE cache ] [ [ NO ] CYCLE ]
```

描述

`CREATE SEQUENCE`创建一个新的序列号发生器。这会创建并初始化一个名为*name*的特殊单行表。该发生器归发出该命令的用户所有。

如果给定了模式名，则序列会在指定模式中创建。否则，它会在当前模式中创建。
临时序列存在于一个特殊模式中，因此创建临时序列时不能指定模式名。
序列名必须不同于同一模式中任何其他关系（表、序列、索引、视图、物化视图或外部表）的名称。

创建序列后，可以使用函数 `nextval`、`currval`以及 `setval`来操作该序列。这些函数的文档见第 9.12 节。

虽然不能直接更新序列，但可以使用如下查询：

```
SELECT * FROM name;
```

来查看序列的参数和当前状态。特别是，序列的 `last_value`字段会显示任一会话最近分配的值。（当然，如果其他会话正在调用`nextval`，那么到该值被打印出来时，它可能已经过时。）

参数

`TEMPORARY` or `TEMP`

如果指定该项，则该序列对象只为当前会话创建，并会在会话退出时自动删除。
临时序列存在期间，已有的同名永久序列在本会话中不可见，
除非使用模式限定名来引用它们。

name

要创建的序列的名称（可以是模式限定的）。

increment

可选子句`INCREMENT BY` *increment*指定加到当前序列值上以生成新值的增量。正值生成升序序列，负值生成降序序列。默认值为 1。

minvalue
`NO MINVALUE`

可选子句`MINVALUE` *minvalue*确定序列可生成的最小值。如果未提供此子句，或者指定了`NO MINVALUE`，则使用默认值。升序序列和降序序列的默认值分别为 1 和 $-2^{63}-1$。

maxvalue
NO MAXVALUE

可选子句MAXVALUE *maxvalue*确定序列的最大值。如果未提供此子句，或者指定了 NO MAXVALUE，则使用默认值。升序序列和降序序列的默认值分别为 2^{63}-1 和 -1。

start

可选子句START WITH *start* 允许序列从任意值开始。升序序列的默认起始值为*minvalue*，降序序列则为*maxvalue*。

cache

可选子句CACHE *cache*指定要预分配并存储在内存中以加快访问的序列号数量。最小值为 1（一次只能生成一个值，即不缓存），这也是默认值。

CYCLE
NO CYCLE

CYCLE选项允许序列在升序序列达到*maxvalue*、或降序序列达到*minvalue*时分别回绕。如果达到该限制，下一个生成的数字将分别是*minvalue*或*maxvalue*。

如果指定NO CYCLE，则当序列达到其限值后，任何对 nextval的调用都会返回错误。如果既未指定 CYCLE也未指定NO CYCLE，则默认使用NO CYCLE。

注解

使用DROP SEQUENCE删除序列。

序列基于bigint算术，因此其范围不能超过八字节整数的范围（-9223372036854775808 到 9223372036854775807）。在一些较老的平台上，编译器可能不支持八字节整数，此时序列将使用常规的integer算术（范围为 -2147483648 到 +2147483647）。

如果对一个将由多个会话并发使用的序列对象设置了大于 1 的*cache*，则可能得到意想不到的结果。每个会话会在一次访问该序列对象时分配并缓存连续的序列值，并相应增加该序列对象的 last_value。随后，该会话中接下来的 *cache*-1 次nextval 调用会直接返回预分配的值，而不接触序列对象。因此，任何在会话中已分配但未使用的数字都会在会话结束时丢失，从而在序列中造成“空洞”。

此外，虽然多个会话保证会分配到互不相同的序列值，但从所有会话整体来看，这些值的生成顺序可能会乱序。例如，当*cache*设置为 10 时，会话 A 可能保留值 1..10 并返回nextval=1；然后会话 B 可能保留值 11..20 并返回nextval=11，而此时会话 A 尚未生成 nextval=2。因此，当*cache*设置为 1 时，可以安全地假定 nextval值按顺序生成；当*cache*设置大于 1 时，只应假定 nextval值彼此不同，而不应假定它们严格按顺序生成。此外，last_value会反映任一会话最近保留的值，不管该值是否已经由nextval返回。

另一个需要注意的问题是，在这种序列上执行的setval 不会立刻被其他会话察觉，直到它们用完自己缓存的所有预分配值为止。

示例

创建一个名为serial的升序序列，从 101 开始：

```
CREATE SEQUENCE serial START 101;
```

从该序列中取出下一个数字：

```
SELECT nextval('serial');

 nextval
---------
     114
```

在INSERT命令中使用该序列：

```
INSERT INTO distributors VALUES (nextval('serial'), 'nothing');
```

在一次COPY FROM之后更新序列值：

```
BEGIN;
COPY distributors FROM 'input_file';
SELECT setval('serial', max(id)) FROM distributors;
END;
```

兼容性

CREATE SEQUENCE符合SQL 标准，但下列情况除外：

- 不支持标准的AS <data type>表达式。
- 获取下一个值是使用nextval()函数，而不是标准的 NEXT VALUE FOR表达式完成的。

CREATE TABLE

CREATE TABLE — 定义一个新表

大纲

```
CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } ] TABLE table_name (
 [
  { column_name data_type [ DEFAULT default_expr ] [ column_constraint
 [ ... ] ]
    | table_constraint
    | LIKE parent_table [ { INCLUDING | EXCLUDING } DEFAULTS ] }
    [, ... ]
] )
[ INHERITS ( parent_table [, ... ] ) ]
[ WITH OIDS | WITHOUT OIDS ]
[ ON COMMIT { PRESERVE ROWS | DELETE ROWS | DROP } ]
[ TABLESPACE tablespace ]

where column_constraint is:

[ CONSTRAINT constraint_name ]
{ NOT NULL |
  NULL |
  UNIQUE [ USING INDEX TABLESPACE tablespace ] |
  PRIMARY KEY [ USING INDEX TABLESPACE tablespace ] |
  CHECK (expression) |
  REFERENCES reftable [ ( refcolumn ) ] [ MATCH FULL | MATCH PARTIAL |
 MATCH SIMPLE ]
    [ ON DELETE action ] [ ON UPDATE action ] }
[ DEFERRABLE | NOT DEFERRABLE ] [ INITIALLY DEFERRED | INITIALLY
 IMMEDIATE ]

and table_constraint is:

[ CONSTRAINT constraint_name ]
{ UNIQUE ( column_name [, ... ] ) [ USING INDEX TABLESPACE tablespace
 ] |
  PRIMARY KEY ( column_name [, ... ] ) [ USING INDEX
 TABLESPACE tablespace ] |
  CHECK (expression) |
  FOREIGN KEY ( column_name [, ... ] ) REFERENCES reftable [ (
 refcolumn [, ... ] ) ]
    [ MATCH FULL | MATCH PARTIAL | MATCH SIMPLE ] [ ON DELETE action ]
 [ ON UPDATE action ] }
[ DEFERRABLE | NOT DEFERRABLE ] [ INITIALLY DEFERRED | INITIALLY
 IMMEDIATE ]
```

描述

`CREATE TABLE` 将在当前数据库中创建一个新的、初始为空的表。该表归发出该命令的用户所有。

如果给出了模式名（例如 `CREATE TABLE myschema.mytable ...`），则表将在指定模式中创建。否则，它将在当前模式中创建。临时表存在于一个特殊模式中，因此创建临时表时不能给出模式名。表名必须与同一模式中任何其他表、序列、索引、视图或外部表的名称不同。

`CREATE TABLE` 还会自动创建一种数据类型，用以表示与该表一行对应的复合类型。因此，表名不能与同一模式中任何已有数据类型同名。

可选的约束子句指定插入或更新要成功时，新行或更新后的行必须满足的约束（测试）。约束是一种 SQL 对象，可用多种方式帮助定义表中允许的值集合。

定义约束有两种方式：表约束和列约束。列约束作为列定义的一部分定义。表约束则不绑定到特定列，并且可以涵盖多个列。每个列约束也都可以写成表约束；当约束只影响一列时，列约束只是一种书写上的方便。

参数

`TEMPORARY` or `TEMP`

如果指定该选项，表将创建为临时表。临时表会在会话结束时自动删除，或者也可在当前事务结束时删除（见下文 `ON COMMIT`）。在临时表存在期间，同名的现有永久表对当前会话不可见，除非使用带模式限定的名称引用它们。在临时表上创建的任何索引也都会自动成为临时索引。

可以选择在 `TEMPORARY` 或 `TEMP` 前写上 `GLOBAL` 或 `LOCAL`。这在 PostgreSQL 中没有任何区别，但参见兼容性。

`table_name`

要创建的表名（可选地带模式限定）。

`column_name`

要在新表中创建的列名。

`data_type`

列的数据类型。这可以包括数组说明符。有关 PostgreSQL 支持的数据类型的更多信息，请参见第 8 章。

`DEFAULT` *`default_expr`*

`DEFAULT`子句为其所在列定义的列指定默认数据值。该值可以是任何不含变量的表达式（不允许子查询，也不允许交叉引用当前表中的其他列）。默认表达式的数据类型必须与该列的数据类型匹配。

默认值表达式会用于任何未为该列指定值的插入操作。如果一列没有默认值，则默认值为 null。

`INHERITS ( parent_table [, ... ] )`

可选的 `INHERITS` 子句指定一组表，新表将自动从中继承所有列。

使用 `INHERITS` 会在新子表与其父表之间建立持久关系。对父表的模式修改通常也会传播到子表，且默认情况下，对父表的扫描会包含子表的数据。

如果同一列名出现在多个父表中，除非这些父表中该列的数据类型全部匹配，否则会报错。如果没有冲突，这些重复列会合并为新表中的单个列。
如果新表的列名列表中包含一个同样来自继承的列名，其数据类型也必须与继承列匹配，并且列定义会合并为一个。不过，
同名的继承列声明和新列声明不必指定完全相同的约束：来自任何声明的所有约束都会合并到一起，并且全部应用于新表。如果新表显式为该列指定了默认值，
该默认值会覆盖继承声明中的任何默认值。否则，
任何为该列指定默认值的父表都必须指定相同的默认值，否则会报错。

`LIKE` *`parent_table`* `[ { INCLUDING | EXCLUDING } DEFAULTS ]`

`LIKE` 子句指定一个表，新表会自动从中复制所有列名、数据类型及其非空约束。

与 `INHERITS` 不同，新表和原表在创建完成后就完全脱钩了。对原表的修改不会应用到新表，也不可能在扫描原表时包含新表的数据。

只有指定`INCLUDING DEFAULTS`时，才会复制所复制列定义的默认表达式。默认行为是不包含默认表达式，因此新表的所有列将具有空默认值。

`WITH OIDS`
`WITHOUT OIDS`

这个可选子句指定新表的行是否应被分配 OID（对象标识符）。如果既没有指定`WITH OIDS`也没有指定`WITHOUT OIDS`，默认值取决于default_with_oids配置参数。（如果新表继承自任何具有 OID 的表，则会强制使用`WITH OIDS`，即使命令指定了`WITHOUT OIDS`也是如此。）

如果显式或隐式指定了`WITHOUT OIDS`，新表将不存储 OID，也不会为插入其中的行分配 OID。通常认为这样做是值得的，因为它会减少 OID 的消耗，从而推迟 32 位 OID 计数器回卷。一旦计数器回卷，就不能再假定 OID 是唯一的，这会大大降低它们的用途。此外，不在表中包含 OID 可以减少在磁盘上存储该表所需的空间，在大多数机器上每行可减少 4 字节，从而略微提高性能。

要在表创建之后删除 OID，可以使用ALTER TABLE。

`CONSTRAINT` *`constraint_name`*

列约束或表约束的可选名称。如果没有指定，系统会生成一个。

`NOT NULL`

该列不允许包含空值。

`NULL`

该列允许包含空值。这是默认情况。

该子句仅为兼容非标准 SQL 数据库而提供，不建议在新应用中使用。

`UNIQUE` (column constraint)
`UNIQUE ( column_name [, ... ] )` (table constraint)

`UNIQUE` 约束指定表中一列或多列组成的一组只能包含唯一值。表级唯一约束的行为与列级唯一约束相同，只是它还能跨越多列。因此，
该约束要求任意两行在这些列中至少有一列不同。

对于唯一约束，空值不被视为相等。

每个唯一表约束必须命名一组列，这组列不能与该表主键约束所命名的列集合相同。（否则它就只是把同一个约束列出了两次。）

`PRIMARY KEY` (column constraint)
`PRIMARY KEY ( column_name [, ... ] )` (table constraint)

主键约束指定表的一个或多个列只能包含唯一（不重复）的非空值。从技术上讲，`PRIMARY KEY`只是`UNIQUE`和`NOT NULL`的组合，但把一组列标识为主键还为模式设计提供了元数据，因为主键意味着其他表可以依赖这组列作为行的唯一标识符。

一个表只能指定一个主键，无论是作为列约束还是表约束。

主键约束所命名的列集合应当不同于为同一表定义的任何唯一约束所命名的其他列集合。

`CHECK (expression)`

`CHECK` 子句指定一个产生布尔结果的表达式。要使插入或更新成功，新行或更新后的行必须满足该表达式。计算结果为 TRUE 或 UNKNOWN 的表达式视为成功。如果插入或更新操作中的任何一行得到 FALSE 结果，就会抛出错误异常，并且插入或更新不会修改数据库。作为列约束指定的检查约束只应引用该列的值，而出现在表约束中的表达式可以引用多个列。

当前，`CHECK` 表达式不能包含子查询，也不能引用当前行的列之外的变量。

`REFERENCES reftable [ ( refcolumn ) ] [ MATCH matchtype ] [ ON DELETE action ] [ ON UPDATE action ]` (column constraint)
`FOREIGN KEY ( column [, ... ] ) REFERENCES reftable [ ( refcolumn [, ... ] ) ] [ MATCH matchtype ] [ ON DELETE action ] [ ON UPDATE action ]` (table constraint)

这些子句指定外键约束，
要求新表的一个或多个列组成的列组只能包含与被引用表某一行的被引用列值匹配的值。
如果省略*refcolumn*列表，则使用*reftable*的主键。
被引用列必须是被引用表中不可延迟的唯一约束或主键约束的列。请注意，
不能在临时表和永久表之间定义外键约束。

插入到引用列中的值会按照给定的匹配类型，与被引用表及其被引用列中的值进行匹配。
共有三种匹配类型：`MATCH FULL`、`MATCH PARTIAL` 和 `MATCH SIMPLE`（默认值）。
`MATCH FULL` 不允许多列外键中的某一列为空，除非所有外键列都为空。`MATCH SIMPLE` 允许某些外键列为空而外键的其他部分不为空。`MATCH PARTIAL` 目前尚未实现。

此外，当被引用列中的数据发生变化时，会对本表列中的数据执行某些操作。`ON DELETE`子句指定删除被引用表中的被引用行时要执行的操作。同样，`ON UPDATE`子句指定将被引用表中的被引用列更新为新值时要执行的操作。如果行被更新，但被引用列实际上没有变化，
则不执行任何操作。除`NO ACTION`检查以外的引用操作都不能延迟，即使该约束声明为可延迟也是如此。每个子句可以指定以下操作：

`NO ACTION`

产生错误，指出删除或更新会违反外键约束。如果该约束被延迟，
则会在约束检查时仍存在引用行的情况下产生这个错误。这是默认操作。

`RESTRICT`

产生错误，指出删除或更新会违反外键约束。这与`NO ACTION`相同，但检查不能延迟。

CASCADE

分别删除任何引用已删除行的行，或将引用列的值更新为被引用列的新值。

SET NULL

将引用列设置为空值。

SET DEFAULT

将引用列设置为其默认值。（如果默认值不为空，
则被引用表中必须存在与这些默认值匹配的行，否则操作会失败。）

如果被引用列经常变化，可以考虑在外键列上添加索引，
使与该外键列关联的引用动作能够更高效地执行。

DEFERRABLE
NOT DEFERRABLE

这控制约束是否可以延迟。不可延迟的约束会在每条命令之后立即检查。
可延迟约束的检查可以推迟到事务结束（使用SET CONSTRAINTS命令）。NOT DEFERRABLE
是默认值。目前，只有外键约束接受此子句。所有其他类型的约束都不可延迟。

INITIALLY IMMEDIATE
INITIALLY DEFERRED

如果约束可延迟，则此子句指定检查约束的默认时间。如果约束为INITIALLY
IMMEDIATE，则在每条语句之后检查。这是默认值。如果约束为INITIALLY DEFERRED，
则仅在事务结束时检查。可以使用SET CONSTRAINTS命令更改约束检查时间。

ON COMMIT

可以使用 ON COMMIT 控制临时表在事务块结束时的行为。三种选项如下：

PRESERVE ROWS

在事务结束时不执行任何特殊操作。这是默认行为。

DELETE ROWS

临时表中的所有行都会在每个事务块结束时删除。实际上，
每次提交时都会自动执行一次TRUNCATE。

DROP

在当前事务块结束时删除临时表。

TABLESPACE *tablespace*

*tablespace*是要创建新表的表空间名称。如果未指定，则使用default_tablespace；
如果default_tablespace为空字符串，则使用数据库的默认表空间。

USING INDEX TABLESPACE *tablespace*

该子句允许选择与 UNIQUE 或 PRIMARY KEY 约束相关联的索引要创建在哪个表空间中。若
未指定，则使用default_tablespace；如果 default_tablespace为空字符串，则使用数
据库的默认表空间。

注解

不建议在新应用中使用 OID：在可能的情况下，优先使用SERIAL或其他序列生成器作为表的主键。不过，如果应用确实使用 OID 来标识表中的特定行，建议在该表的oid列上创建唯一约束，以确保即使计数器回卷，表中的 OID 也确实能唯一标识行。不要假定 OID 在不同表之间唯一；如果需要数据库范围的唯一标识符，请组合使用tableoid和行 OID。

> **提示**
>
> 对于没有主键的表，不建议使用WITHOUT OIDS，因为既没有 OID，也没有唯一数据键时，很难标识特定行。

PostgreSQL为每一个唯一约束和主键约束自动创建一个索引来强制唯一性。因此，没有必要显式地为主键列创建一个索引（详见CREATE INDEX）。

在当前的实现中，唯一约束和主键不会被继承。这使得继承与唯一约束的组合相当不实用。

一个表不能有超过 1600 列（实际上，由于元组长度限制，有效的限制通常更低）。

示例

创建表 films 和表 distributors：

```
CREATE TABLE films (
    code        char(5) CONSTRAINT firstkey PRIMARY KEY,
    title       varchar(40) NOT NULL,
    did         integer NOT NULL,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute
);

CREATE TABLE distributors (
     did    integer PRIMARY KEY DEFAULT nextval('serial'),
     name   varchar(40) NOT NULL CHECK (name <> '')
);
```

创建一个带二维数组列的表：

```
CREATE TABLE array_int (
    vector  int[][]
);
```

为表films定义一个唯一表约束。唯一表约束可以定义在表的一列或多列上：

```
CREATE TABLE films (
    code        char(5),
    title       varchar(40),
    did         integer,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute,
```

```
    CONSTRAINT production UNIQUE(date_prod)
);
```

定义一个列检查约束：

```
CREATE TABLE distributors (
    did     integer CHECK (did > 100),
    name    varchar(40)
);
```

定义一个表检查约束：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    CONSTRAINT con1 CHECK (did > 100 AND name <> '')
);
```

为表`films`定义一个主键表约束。主键表约束可以定义在表的一列或多列上：

```
CREATE TABLE films (
    code        char(5),
    title       varchar(40),
    did         integer,
    date_prod   date,
    kind        varchar(10),
    len         interval hour to minute,
    CONSTRAINT code_title PRIMARY KEY(code,title)
);
```

为表`distributors`定义一个主键约束。下面两个示例是等价的，第一个使用表约束语法，第二个使用列约束语法。

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    PRIMARY KEY(did)
);

CREATE TABLE distributors (
    did     integer PRIMARY KEY,
    name    varchar(40)
);
```

这里为列`name`指定一个字面常量默认值，将列`did`的默认值安排为从某个序列对象中取下一个值，并让`modtime`的默认值为插入该行的时间。

```
CREATE TABLE distributors (
    name      varchar(40) DEFAULT 'Luso Films',
    did       integer DEFAULT nextval('distributors_serial'),
    modtime   timestamp DEFAULT current_timestamp
);
```

在表`distributors`上定义两个`NOT NULL`列约束，其中一个显式指定了名称：

```
CREATE TABLE distributors (
```

```
    did     integer CONSTRAINT no_null NOT NULL,
    name    varchar(40) NOT NULL
);
```

为name列定义一个唯一约束：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40) UNIQUE
);
```

上面的写法等价于下面指定为表约束的形式：

```
CREATE TABLE distributors (
    did     integer,
    name    varchar(40),
    UNIQUE(name)
);
```

在表空间diskvol1中创建表cinemas：

```
CREATE TABLE cinemas (
        id serial,
        name text,
        location text
) TABLESPACE diskvol1;
```

兼容性

CREATE TABLE 命令符合 SQL 标准，但有下列例外。

临时表

尽管 CREATE TEMPORARY TABLE 的语法看起来类似于 SQL 标准，但其效果并不相同。按标准，临时表只需定义一次，并会自动存在于每个需要它的会话中（内容初始为空）。而 PostgreSQL 要求每个会话都为每个要使用的临时表发出自己的 CREATE TEMPORARY TABLE 命令。这使不同会话可以出于不同目的使用相同的临时表名；而标准做法则要求给定临时表名的所有实例都必须具有相同的表结构。

标准对临时表行为的定义在实践中被广泛忽略。PostgreSQL 在这一点上的行为与多种其他 SQL 数据库相似。

标准中全局临时表和局部临时表的区分在 PostgreSQL 中并不存在，因为这一区分依赖于模块的概念，而 PostgreSQL 没有这一概念。出于兼容性考虑，PostgreSQL 会接受在临时表声明中使用 GLOBAL 和 LOCAL 关键字，但它们没有效果。

临时表的 ON COMMIT 子句也与 SQL 标准相似，但存在一些差异。如果省略 ON COMMIT 子句，SQL 规定默认行为是 ON COMMIT DELETE ROWS。然而，PostgreSQL 中的默认行为是 ON COMMIT PRESERVE ROWS。SQL 中不存在 ON COMMIT DROP 选项。

列检查约束

SQL 标准规定，CHECK 列约束只能引用其所作用的列；只有 CHECK 表约束才能引用多列。PostgreSQL 并不强制这一限制；它对列检查约束和表检查约束一视同仁。

NULL“约束”

NULL“约束”（实际上并不是约束）是 PostgreSQL 对 SQL 标准的扩展；提供它是为了与其他一些数据库系统兼容（以及与 NOT NULL 约束保持对称）。由于它本来就是任意列的默认情况，所以它的存在只是噪声。

继承

通过 INHERITS 子句实现的多重继承是 PostgreSQL 的语言扩展。SQL:1999 及后续标准使用不同的语法和语义定义了单继承。PostgreSQL 尚不支持 SQL:1999 风格的继承。

零列表

PostgreSQL 允许创建没有列的表（例如 CREATE TABLE foo();）。这是对 SQL 标准的扩展，标准不允许零列的表。零列的表本身并不十分有用，但若禁止它们，就会让 ALTER TABLE DROP COLUMN 出现奇怪的特殊情况，因此忽略这一规范限制看起来更整洁。

对象 ID

PostgreSQL 的 OID 概念不是标准的概念。

表空间

PostgreSQL 的表空间概念不是标准的一部分。因此，TABLESPACE 和 USING INDEX TABLESPACE 子句都是扩展。

另见

ALTER TABLE, DROP TABLE, CREATE TABLESPACE

CREATE TABLE AS

CREATE TABLE AS — 根据查询结果定义一个新表

大纲

```
CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } ] TABLE table_name
    [ (column_name [, ...] ) ] [ [ WITH | WITHOUT ] OIDS ]
    AS query
```

描述

CREATE TABLE AS创建一个表，并用 SELECT命令或运行已预备的SELECT 命令的EXECUTE计算得到的数据填充该表。该表的列具有与SELECT输出列对应的名称和数据类型（但可以通过显式给出新的列名列表来覆盖列名）。

CREATE TABLE AS与创建视图有些相似，但实际上差异很大：它会创建一个新表，并且只执行一次该查询来完成新表的初始填充。这个新表不会跟踪该查询所引用源表的后续变化。
相比之下，视图每次被查询时都会重新执行其定义中的 SELECT语句。

参数

GLOBAL or LOCAL

出于兼容性考虑会被忽略。详见CREATE TABLE。

TEMPORARY or TEMP

如果指定，将把该表创建为临时表。详见CREATE TABLE。

table_name

要创建的表名（可选地带模式限定）。

column_name

新表中某一列的名称。如果未提供列名，则采用查询输出列的列名。如果表是从 EXECUTE 命令创建的，则不能指定列名列表。

WITH OIDS
WITHOUT OIDS

这个可选子句指定CREATE TABLE AS创建的表是否应包含 OID。如果未指定该子句的任何一种形式，则使用default_with_oids配置参数的值。

query

一条查询语句（也就是一个SELECT命令，或一个运行已预备的SELECT命令的EXECUTE 命令）。所允许语法的描述请分别参阅SELECT或EXECUTE。

注解

这个命令在功能上类似于SELECT INTO，但更推荐使用它，因为它更不容易与SELECT INTO语法的其他用法相混淆。此外，CREATE TABLE AS提供的功能是SELECT INTO所提供功能的超集。

在 PostgreSQL 8.0 之前，`CREATE TABLE AS`创建的表总是包含 OID。从 PostgreSQL 8.0 起，`CREATE TABLE AS`命令允许用户显式指定是否包含 OID。如果未显式指定是否存在 OID，则使用default_with_oids配置变量。而从 PostgreSQL 8.1 起，该变量默认为 false，因此默认行为与 8.0 之前的发行版不再相同。需要在`CREATE TABLE AS`所创建表中包含 OID 的应用，应显式指定`WITH OIDS`以确保所需行为。

示例

创建一个新表`films_recent`，其中只包含表 `films`中的近期条目：

```
CREATE TABLE films_recent AS
  SELECT * FROM films WHERE date_prod >= '2002-01-01';
```

兼容性

`CREATE TABLE AS`符合SQL标准，但有以下例外：

- 标准要求在子查询子句外围加括号；在 PostgreSQL中，这些括号是可选的。
- 标准定义了`ON COMMIT`子句；目前 PostgreSQL尚未实现它。
- 标准定义了`WITH [ NO ] DATA`子句；目前 PostgreSQL尚未实现它。PostgreSQL提供的行为等同于标准中的 `WITH DATA`情形。
- `WITH/WITHOUT OIDS`是 PostgreSQL的扩展。
- PostgreSQL处理临时表的方式与标准有相当大的不同；详见CREATE TABLE。

另见

CREATE TABLE, EXECUTE, SELECT, SELECT INTO

CREATE TABLESPACE

CREATE TABLESPACE — 定义一个新表空间

大纲

```
CREATE TABLESPACE tablespacename [ OWNER username ] LOCATION
 'directory'
```

描述

`CREATE TABLESPACE`注册一个新的集簇级别的表空间。该表空间的名称必须不同于数据库集簇中任何现有表空间的名称。

表空间允许超级用户在文件系统中定义一个替代位置，使数据库对象（如表和索引）的数据文件可以存放在那里。

具有适当权限的用户可以将 *tablespacename*传递给 `CREATE DATABASE`、`CREATE TABLE`、`CREATE INDEX`或`ADD CONSTRAINT`，从而将这些对象的数据文件存储在指定的表空间中。

参数

tablespacename

要创建的表空间名称。名称不能以`pg_`开头，因为这类名称保留给系统表空间使用。

username

将成为该表空间所有者的用户名。若省略，则默认为执行该命令的用户。只有超级用户可以创建表空间，但他们可以将表空间的所有权赋予非超级用户。

directory

将用于表空间的目录。该目录应为空，并且必须归PostgreSQL系统用户所有。必须使用绝对路径名指定该目录。

注解

表空间仅在支持符号链接的系统上受支持。

示例

创建一个表空间`dbspace`，位置为`/data/dbs`：

```
CREATE TABLESPACE dbspace LOCATION '/data/dbs';
```

创建一个表空间`indexspace`，位置为`/data/indexes`，所有者为用户`genevieve`：

```
CREATE TABLESPACE indexspace OWNER genevieve LOCATION '/data/indexes';
```

兼容性

`CREATE TABLESPACE`是一种PostgreSQL扩展。

另见

CREATE DATABASE, CREATE TABLE, CREATE INDEX, DROP TABLESPACE, ALTER TABLESPACE

CREATE TRIGGER

CREATE TRIGGER — 定义一个新触发器

大纲

```
CREATE TRIGGER name { BEFORE | AFTER } { event [ OR ... ] }
    ON table [ FOR [ EACH ] { ROW | STATEMENT } ]
    EXECUTE PROCEDURE funcname ( arguments )
```

描述

CREATE TRIGGER创建一个新触发器。触发器会关联到指定的表、视图或外部表，并在发生某些事件时执行指定的函数*funcname*。

可以指定触发器在尝试对某一行执行该操作之前引发（即在检查约束以及尝试执行INSERT、UPDATE或DELETE之前）；也可以在该操作完成之后引发（即在检查约束以及完成 INSERT、UPDATE或 DELETE之后）；或者取代该操作执行（用于视图上的插入、更新或删除）。如果触发器在事件之前引发，或者改为取代该事件执行，则它可以跳过对当前行的操作，或者修改待插入的行（仅适用于INSERT和UPDATE操作）。如果触发器在事件之后引发，则所有更改（包括其他触发器的影响）对该触发器都是“可见”的。

被标记为FOR EACH ROW的触发器会针对该操作修改的每一行调用一次。例如，一个影响 10 行的DELETE会导致目标关系上的 ON DELETE触发器分别被调用 10 次，即每个被删除的行调用一次。相比之下，被标记为FOR EACH STATEMENT的触发器对于任意给定操作只执行一次，不管该操作修改了多少行（特别是，修改零行的操作仍会导致所有适用的 FOR EACH STATEMENT触发器执行）。

如果针对同一事件定义了多个同类触发器，它们将按名称的字母顺序引发。

SELECT 不会修改任何行，因此无法创建 SELECT 触发器。对于这类情形，规则和视图更为合适。

参见第 33 章以了解有关触发器的更多信息。

参数

name

要赋予新触发器的名称。它必须不同于同一表上任何其他触发器的名称。

BEFORE
AFTER

决定该函数是在事件之前还是之后被调用。

event

INSERT、UPDATE或DELETE之一；它指定将引发该触发器的事件。可以使用OR指定多个事件。

table

该触发器所作用的表、视图或外部表的名称（可以是模式限定的）。

FOR EACH ROW
FOR EACH STATEMENT

这指定触发器过程是针对受触发事件影响的每一行引发一次，还是针对每条 SQL 语句只引发一次。如果两者都未指定，则默认为FOR EACH STATEMENT。

funcname

一个由用户提供的函数，声明为不接受参数并返回 trigger 类型；当触发器被触发时，就会执行该函数。

arguments

一个可选的、以逗号分隔的参数列表，在执行触发器时会传给该函数。
这些参数是字符串字面常量。这里也可以写简单名称和数字常量，
但它们都会被转换成字符串。请查阅该触发器函数所用实现语言的说明，
以了解如何在函数内部访问这些参数；这可能与普通函数参数不同。

注解

要在表上创建触发器，用户必须拥有该表的TRIGGER权限。

使用DROP TRIGGER删除触发器。

在PostgreSQL 7.3 之前的版本中，必须将触发器函数声明为返回占位类型opaque，而不是trigger。为了支持加载旧转储文件，CREATE TRIGGER将接受声明为返回opaque的函数，但会发出通知，并将该函数声明的返回类型更改为trigger。

示例

第 33.4 节包含一个完整示例。

兼容性

PostgreSQL 中的 CREATE TRIGGER 语句实现了 SQL 标准的一个子集。当前仍缺少以下功能：

- SQL 允许触发器在特定列被更新时引发（例如 AFTER UPDATE OF col1, col2）。
- SQL 允许为在触发动作定义中使用的“旧”行或“新”行定义别名（例如 CREATE TRIGGER ... ON tablename REFERENCING OLD ROW AS somename NEW ROW AS othername ...）。由于 PostgreSQL 允许用任意多种用户定义的语言编写触发器过程，对数据的访问是以语言相关的方式处理的。
- PostgreSQL 只允许通过执行用户定义函数来完成触发动作。标准则允许把许多其他 SQL 命令（例如 CREATE TABLE）作为触发动作执行。通过创建一个执行所需命令的用户定义函数，这一限制并不难绕过。

SQL 规定多个触发器应按创建时间顺序引发。PostgreSQL 则按名称顺序引发，这被认为更方便。

SQL 规定，级联删除上的 BEFORE DELETE 触发器应在级联 DELETE 完成之后引发。PostgreSQL 的行为则是 BEFORE DELETE 总会在删除动作之前引发，即使是级联删除也一样。这被认为更一致。如果 BEFORE 触发器在由引用动作引起的更新过程中修改行或阻止更新，也会出现非标准行为。这可能导致约束违反，或者使存储的数据不符合引用约束。

使用 OR 为单个触发器指定多个动作的能力，是 PostgreSQL 对 SQL 标准的扩展。

参见

CREATE FUNCTION, ALTER TRIGGER, DROP TRIGGER

CREATE TYPE

CREATE TYPE — 定义一种新的数据类型

大纲

```
CREATE TYPE name AS
    ( attribute_name data_type [, ... ] )

CREATE TYPE name (
    INPUT = input_function,
    OUTPUT = output_function
    [ , RECEIVE = receive_function ]
    [ , SEND = send_function ]
    [ , ANALYZE = analyze_function ]
    [ , INTERNALLENGTH = { internallength | VARIABLE } ]
    [ , PASSEDBYVALUE ]
    [ , ALIGNMENT = alignment ]
    [ , STORAGE = storage ]
    [ , DEFAULT = default ]
    [ , ELEMENT = element ]
    [ , DELIMITER = delimiter ]
)
```

描述

`CREATE TYPE`在当前数据库中注册一种新的数据类型。定义该类型的用户将成为其拥有者。

如果给出了模式名，则该类型会在指定模式中创建；否则会在当前模式中创建。
类型名必须不同于同一模式中任何现有类型或域的名称。（由于表也具有关联的数据类型，
类型名还必须不同于同一模式中任何现有表的名称。）

复合类型

第一种形式的`CREATE TYPE`创建一个复合类型。复合类型由属性名和数据类型的列表指定。它本质上与表的行类型相同，但如果想要的只是定义一种类型，使用`CREATE TYPE`可以避免创建一个实际的表。独立的复合类型可用作函数的参数类型或返回类型。

基础类型

第二种形式的`CREATE TYPE`创建一种新的基础类型（标量类型）。这些参数可以按任意顺序出现，不必局限于上面展示的顺序，而且大多数都是可选的。在定义该类型之前，
必须先注册两个或更多函数（使用 `CREATE FUNCTION`）。支持函数 *input_function* 和 *output_function*是必需的；*receive_function*、*send_function*和 *analyze_function*则是可选的。通常这些函数必须用 C 或其他低级语言编写。

*input_function*将类型的外部文本表示转换为该类型的操作符和函数所使用的内部表示。
output_function 执行相反的转换。输入函数可以声明为接受一个`cstring` 参数，或者接受三个参数，类型分别为`cstring`、`oid`、`integer`。第一个参数是以 C 字符串表示的输入文本，第二个参数是该类型自身的 OID（数组类型例外，此时传入的是其元素类型的 OID），第三个参

数是在已知情况下目标列的 typmod（未知则传入 -1）。输入函数必须返回该数据类型本身的值。输出函数必须声明为接受一个新数据类型参数，并且必须返回cstring类型。

可选的*receive_function* 把类型的外部二进制表示转换为内部表示。如果未提供此函数，该类型就不能参与二进制输入。外部二进制表示应选择为既能低成本转换为内部形式，又具有合理可移植性。（例如，标准整数数据类型把网络字节序用作外部二进制表示，而内部表示则使用机器的本地字节序。）接收函数应进行充分检查以确保值有效。它可以声明为接受一个 internal参数，或者接受三个参数，类型分别为 internal、oid、integer。第一个参数是指向保存已接收字节串的StringInfo缓冲区的指针；其余可选参数与文本输入函数相同。接收函数必须返回该数据类型本身的值。类似地，可选的*send_function*把内部表示转换为外部二进制表示。如果未提供此函数，该类型就不能参与二进制输出。发送函数必须声明为接受一个新数据类型参数，并且必须返回bytea类型。

读到这里，你可能会问：既然输入和输出函数必须在新类型创建之前创建，它们怎么能声明为返回或接受这个新类型呢？答案是，必须先创建输入函数，然后创建输出函数（以及二进制 I/O 函数，如果需要的话），最后才创建数据类型。PostgreSQL最初会把新数据类型的名称看作输入函数的返回类型。它会创建一个 “shell” 类型，这实际上只是系统目录中的一个占位条目，并把输入函数的定义链接到该 shell 类型。类似地，其他函数也会链接到（此时已经存在的）shell 类型。最后，CREATE TYPE用完整的类型定义替换 shell 条目，之后新类型就可以使用了。

可选的*analyze_function* 为该数据类型的列执行类型专用的统计信息收集。默认情况下，如果该类型有默认的 B-树操作符类，ANALYZE将尝试使用该类型的 “equals” 和 “less-than” 操作符收集统计信息。对于非标量类型，这种行为很可能不合适，因此可以通过指定自定义分析函数来覆盖。分析函数必须声明为接受一个 internal参数并返回boolean结果。分析函数的详细 API 见src/include/commands/vacuum.h。

虽然新类型内部表示的细节只有 I/O 函数以及你为该类型编写的其他函数才知道，但仍有若干内部表示属性必须向 PostgreSQL声明。其中最重要的是 *internallength*。基础类型可以是定长的，此时*internallength*为正整数；也可以是变长的，此时将*internallength*设为VARIABLE。（在内部，这通过把 typlen设为 -1 表示。）所有变长类型的内部表示都必须以一个 4 字节整数开头，用来给出该类型该值的总长度。

可选标志PASSEDBYVALUE表示该数据类型的值按值传递，而不是按引用传递。内部表示大于Datum类型大小（多数机器上为 4 字节，少数为 8 字节）的类型不能按值传递。

*alignment*参数指定该数据类型所需的存储对齐方式。允许的值分别对应按 1、2、4 或 8 字节边界对齐。注意，变长类型的对齐至少必须为 4，因为它们的第一个组成部分必然是一个int4。

*storage*参数允许为变长数据类型选择存储策略。（定长类型只允许 plain。）plain表示该类型数据始终内联存储且不压缩。extended表示系统会先尝试压缩较长的数据值，如果仍然过长，就把该值移出主表行。external允许把值移出主表，但系统不会尝试压缩它。main允许压缩，但不鼓励把值移出主表。（采用这种存储策略的数据项在没有其他办法让一行适配时仍可能被移出主表，但与extended和external 数据项相比，它们会被优先保留在主表中。）

如果用户希望该数据类型列的默认值不是空值，可以指定默认值。用 DEFAULT关键字指定默认值。（该默认值可以被附加到具体列上的显式DEFAULT子句覆盖。）

若要表明一种类型是数组类型，可用ELEMENT 关键字指定数组元素的类型。例如，要定义由 4 字节整数（int4）构成的数组，可指定 ELEMENT = int4。更多关于数组类型的细节见下文。

若要指定该类型数组在外部表示中用于分隔各值的分隔符，可把 *delimiter*设为特定字符。默认分隔符是逗号（,）。注意，这个分隔符关联的是数组元素类型，而不是数组类型本身。

数组类型

每当创建用户定义类型时，PostgreSQL 都会自动创建一个关联的数组类型，其名称由元素类型名前加一个下划线组成。解析器理解这一命名约定，会把对 `foo[]` 类型列的请求转换为对 `_foo` 类型的请求。这种隐式创建的数组类型是变长的，并使用内置输入/输出函数 `array_in` 和`array_out`。

如果系统会自动创建正确的数组类型，你可能会合理地问，为什么还需要`ELEMENT`选项。`ELEMENT`唯一有用的情况是：你正在创建一种定长类型，而它在内部恰好是若干相同元素组成的数组，并且除了为整个类型提供的操作之外，你还希望允许通过下标直接访问这些元素。例如，类型`name`允许按这种方式访问其组成的`char`元素。一个二维`point`类型可以允许其两个分量数字像`point[0]`和 `point[1]`那样被访问。注意，这种机制只适用于内部形式恰好是一串相同定长字段的定长类型。可用下标访问的变长类型必须采用`array_in`和`array_out`使用的通用内部表示。由于历史原因（也就是说这显然不对，但现在改已经太晚了），定长数组类型的下标从零开始，而变长数组则从一开始。

参数

name

要创建的类型名称（可选地带有模式限定）。

attribute_name

复合类型的一个属性（列）的名称。

data_type

将成为复合类型一列的现有数据类型名称。

input_function

将数据从类型的外部文本形式转换为内部形式的函数名。

output_function

将数据从类型的内部形式转换为外部文本形式的函数名。

receive_function

将数据从类型的外部二进制形式转换成内部形式的函数名。

send_function

将数据从类型的内部形式转换为外部二进制形式的函数名。

analyze_function

为该数据类型执行统计分析的函数名。

internallength

一个数字常量，用于指定新类型内部表示的字节长度。默认假定它是变长的。

alignment

该数据类型的存储对齐需求。如果被指定，它必须是 `char`、`int2`、`int4`或者`double`。默认是 `int4`。

storage

该数据类型的存储策略。如果被指定，必须是 `plain`、`external`、`extended`或者`main`。默认是`plain`。

default

该数据类型的默认值。若省略，默认值为 null。

element

正在创建的类型是数组；该参数指定数组元素类型。

delimiter

在由该类型构成的数组中各值之间使用的分隔符字符。

注解

用户定义类型的名称不能以下划线字符（`_`）开头，而且最长只能是 62 个字符（一般来说是 `NAMEDATALEN - 2`，而不是其他名称所允许的 `NAMEDATALEN - 1` 个字符）。以下划线开头的类型名保留给内部创建的数组类型名使用。

在PostgreSQL 7.3 以前的版本中，通常完全不创建 shell 类型，而是将函数中对类型名的前向引用替换为占位伪类型 `opaque`。在 7.3 以前，`cstring`参数和结果也必须声明为`opaque`。为了支持载入旧转储文件，`CREATE TYPE`会接受使用`opaque`声明的函数，但会发出通知，并修改函数的声明以使用正确的类型。

示例

这个示例创建一种复合类型，并在函数定义中使用它：

```
CREATE TYPE compfoo AS (f1 int, f2 text);

CREATE FUNCTION getfoo() RETURNS SETOF compfoo AS $$
    SELECT fooid, fooname FROM foo
$$ LANGUAGE SQL;
```

这个示例创建基础类型`box`，然后在表定义中使用它：

```
CREATE TYPE box (
    INTERNALLENGTH = 16,
    INPUT = my_box_in_function,
    OUTPUT = my_box_out_function
);

CREATE TABLE myboxes (
    id integer,
    description box
);
```

如果box的内部结构是由四个 float4元素构成的数组，则也可以改为这样写：

```
CREATE TYPE box (
    INTERNALLENGTH = 16,
    INPUT = my_box_in_function,
    OUTPUT = my_box_out_function,
    ELEMENT = float4
);
```

这样就能通过下标访问 box 值的各个分量。除此之外，该类型的行为与前例相同。

这个示例创建一种大对象类型，并在表定义中使用它：

```
CREATE TYPE bigobj (
    INPUT = lo_filein, OUTPUT = lo_fileout,
    INTERNALLENGTH = VARIABLE
);
CREATE TABLE big_objs (
    id integer,
    obj bigobj
);
```

更多示例（包括配套的输入和输出函数）请见第 32.11 节。

兼容性

这里的CREATE TYPE命令是一个PostgreSQL扩展。SQL标准中也有一个CREATE TYPE语句，但细节上相当不同。

另见

CREATE FUNCTION, DROP TYPE, ALTER TYPE

CREATE USER

CREATE USER — 定义一个新的数据库角色

大纲

```
CREATE USER name [ [ WITH ] option [ ... ] ]

where option can be:

      SUPERUSER | NOSUPERUSER
    | CREATEDB | NOCREATEDB
    | CREATEROLE | NOCREATEROLE
    | CREATEUSER | NOCREATEUSER
    | INHERIT | NOINHERIT
    | LOGIN | NOLOGIN
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | VALID UNTIL 'timestamp'
    | IN ROLE rolename [, ...]
    | IN GROUP rolename [, ...]
    | ROLE rolename [, ...]
    | ADMIN rolename [, ...]
    | USER rolename [, ...]
    | SYSID uid
```

描述

`CREATE USER`现为CREATE ROLE的别名。唯一的区别在于：当命令写为`CREATE USER`时，默认假定为`LOGIN`；而当命令写为 `CREATE ROLE`时，默认假定为`NOLOGIN`。

兼容性

`CREATE USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义留给具体实现决定。

另见

CREATE ROLE

CREATE VIEW

CREATE VIEW — 定义一个新视图

大纲

```
CREATE [ OR REPLACE ] [ TEMP | TEMPORARY ] VIEW name [ ( column_name [
, ...] ) ]
    AS query
```

描述

`CREATE VIEW`定义一个视图，其内容来自一个查询。该视图不会被实际物化。相反，每次在查询中引用该视图时，都会执行定义它的查询。

`CREATE OR REPLACE VIEW`与之类似，但如果已经存在同名视图，则会用新定义替换它。只能用一条生成完全相同的列集合（即相同的列名和数据类型）的新查询来替换视图。

如果给出了模式名称（例如，`CREATE VIEW myschema.myview ...`），则视图将在指定的模式中创建。否则，它将在当前模式中创建。临时视图存在于一个特殊模式中，
因此在创建临时视图时不能给出模式名称。视图的名称必须与同一模式中的任何其他关系（表、序列、索引、视图、物化视图或外部表）的名称不同。

参数

`TEMPORARY` or `TEMP`

如果指定该选项，视图将被创建为临时视图。临时视图会在当前会话结束时自动删除。
当临时视图存在时，除非使用模式限定名称引用，
否则具有相同名称的现有永久关系对当前会话不可见。

如果视图引用的任何表是临时表，则该视图也会被创建为临时视图（无论是否指定了`TEMPORARY`）。

`name`

要创建的视图名称（可以带模式限定）。

`column_name`

视图各列使用的名称列表，可选。若未给出，则从查询中推断列名。

`query`

一个将为该视图提供列和行的查询（即一条`SELECT`语句）。

有关有效查询的更多信息，请参见SELECT。

注解

当前，视图是只读的：系统不允许在视图上执行插入、更新或删除。你可以通过创建规则，
把对视图的插入等操作重写为对其他表的适当动作，来获得可更新视图的效果。
更多信息见CREATE RULE。

使用DROP VIEW语句删除视图。

应留意视图列的名称和类型是否按预期确定。例如：

```
CREATE VIEW vista AS SELECT 'Hello World';
```

这种写法在两方面都不好：列名默认是`?column?`，而且列数据类型默认为`unknown`。
如果想要在视图结果中使用字符串字面量，可使用类似下面的写法：

```
CREATE VIEW vista AS SELECT text 'Hello World' AS hello;
```

对视图中引用的表的访问由视图所有者的权限决定。在某些情况下，
这可以用于为底层表提供安全但受限的访问。不过，并非所有视图都能防止篡改；
详情请参见第 34.4 节。视图调用的函数与直接从使用该视图的查询中调用时的处理方式相同。因此，视图的用户必须拥有调用视图所用全部函数的权限。

对现有视图使用`CREATE OR REPLACE VIEW`时，只会更改视图的定义性 SELECT 规则。其他视图属性（包括所有权、权限和非 SELECT 规则）保持不变。必须拥有该视图才能替换它（包括作为所有者角色的成员）。

示例

创建一个由所有喜剧电影组成的视图：

```
CREATE VIEW comedies AS
    SELECT *
    FROM films
    WHERE kind = 'Comedy';
```

兼容性

SQL 标准为`CREATE VIEW`语句指定了一些额外的能力：

```
CREATE VIEW name [ ( column_name [, ...] ) ] AS query [ WITH [
 CASCADED | LOCAL ] CHECK OPTION ]
```

完整 SQL 命令的可选子句有：

`CHECK OPTION`

该选项与可更新视图有关。对视图的所有 `INSERT` 和 `UPDATE` 命令都会被检查，以确保数据满足视图定义条件（也就是说，新数据可以通过视图看到）。如果不满足，更新将被拒绝。

`LOCAL`

在该视图上检查完整性。

`CASCADED`

在该视图以及任何依赖视图上检查完整性。如果既未指定`CASCADED`也未指定`LOCAL`，则假定使用`CASCADED`。

`CREATE OR REPLACE VIEW` 是 PostgreSQL 的语言扩展。临时视图的概念也是如此。

另见

DROP VIEW

DEALLOCATE

DEALLOCATE — 释放一个预备语句

大纲

`DEALLOCATE [ PREPARE ] plan_name`

描述

`DEALLOCATE`用于释放一个先前准备好的 SQL 语句。如果不显式释放预备语句，则它会在会话结束时被释放。

更多关于预备语句的信息，请参见PREPARE。

参数

`PREPARE`

此关键字会被忽略。

`plan_name`

要释放的预备语句的名称。

兼容性

SQL 标准中包含一个`DEALLOCATE`语句，但它只用于嵌入式 SQL。

另见

EXECUTE, PREPARE

DECLARE

DECLARE — 定义一个游标

大纲

```
DECLARE name [ BINARY ] [ INSENSITIVE ] [ [ NO ] SCROLL ]
    CURSOR [ { WITH | WITHOUT } HOLD ] FOR query
    [ FOR { READ ONLY | UPDATE [ OF column [, ...] ] } ]
```

描述

DECLARE允许用户创建游标，游标可用于从较大的查询中一次取出少量行。游标可以使用FETCH以文本或二进制格式返回数据。

普通游标以文本格式返回数据，就像SELECT所产生的结果一样。
由于数据在内部是以二进制格式存储的，系统必须做一次转换才能产生文本格式。
而信息以文本形式返回后，客户端应用又可能需要把它转换回二进制格式才能处理。此外，
文本格式的数据通常比二进制格式占用的空间更大。二进制游标以二进制表示返回数据，
处理起来可能更方便。不过，如果你本来就打算把数据显示为文本，
那么以文本形式检索会省去客户端的一些工作量。

举例来说，如果某个查询从一个整型列返回值 1，那么默认游标会返回字符串 1，而二进制游标则会返回一个 4 字节字段，其中包含该值的内部表示（采用大端字节序）。

应谨慎使用二进制游标。许多应用程序（包括 psql）都没有准备好处理二进制游标，而是期望返回的数据为文本格式。

> **注意**
>
> 当客户端应用使用“扩展查询”协议发出FETCH命令时，Bind 协议消息会指定数据应以文本格式还是二进制格式提取。这一选择会覆盖定义游标时所指定的方式。因此，
> 在使用扩展查询协议时，二进制游标这一概念实际上已经过时了 – 任何游标都可以按文本或二进制方式处理。

参数

name

要创建的游标名称。它必须与该会话中任何其他活动游标的名称不同。

BINARY

使游标以二进制格式而不是文本格式返回数据。

INSENSITIVE

表示从游标取出的数据不应受到游标存在期间对其底层表所做更新的影响。
在PostgreSQL中，所有游标都是不敏感的；该关键字目前没有任何效果，只是为兼容 SQL 标准而存在。

SCROLL
NO SCROLL

SCROLL指定游标可用于以非顺序方式（例如向后）取出行。根据查询执行计划的复杂程度，指定SCROLL可能会给查询执行带来性能开销。NO SCROLL指定游标不能以非顺序方式取出行。默认情况下只在某些情形下允许滚动，这与显式指定 SCROLL并不相同。详见下文注解。

WITH HOLD
WITHOUT HOLD

WITH HOLD指定在创建该游标的事务成功提交后，仍可继续使用该游标。WITHOUT HOLD指定该游标不能在创建它的事务之外使用。如果既未指定WITHOUT HOLD也未指定WITH HOLD，默认值是WITHOUT HOLD。

query

一个将为游标提供要返回的行的SELECT命令。有关有效查询的更多信息，请参阅SELECT。

FOR READ ONLY
FOR UPDATE

FOR READ ONLY表示该游标将以只读方式使用。FOR UPDATE表示该游标将用于更新表。由于 PostgreSQL目前不支持游标更新，指定 FOR UPDATE会导致一条错误消息，而指定 FOR READ ONLY则没有任何效果。

column

将被游标更新的列。由于PostgreSQL目前不支持游标更新，FOR UPDATE子句会引发一条错误消息。

关键字BINARY、INSENSITIVE和SCROLL可以按任意顺序出现。

注解

除非指定了WITH HOLD，否则该命令创建的游标只能在当前事务中使用。因此，DECLARE若不带WITH HOLD，在事务块之外就毫无用处：游标只能存活到该语句执行完成。所以，如果在事务块之外使用这种命令，PostgreSQL会报错。可使用BEGIN、COMMIT和ROLLBACK来定义事务块。

如果指定了WITH HOLD，且创建游标的事务成功提交，那么在同一会话中的后续事务里仍可继续访问该游标。（但如果创建事务被中止，游标会被移除。）使用WITH HOLD创建的游标会在对其发出显式CLOSE命令时关闭，或在会话结束时关闭。在当前实现中，这种游标所表示的行会被复制到临时文件或内存区域中，以便它们在后续事务中仍然可用。

在定义将用于向后取出行的游标时，应指定SCROLL选项。这是 SQL 标准所要求的。不过，为了兼容早期版本，如果游标的查询计划足够简单，以致支持向后取出不需要额外开销，PostgreSQL也允许在未指定SCROLL的情况下向后取出行。不过，建议应用开发者不要依赖于从未用SCROLL创建的游标中向后取出行。如果指定了NO SCROLL，那么无论如何都不允许向后取出行。

SQL 标准只为嵌入式SQL中的游标作出规定。PostgreSQL服务器没有为游标实现OPEN语句；游标在声明时即被视为打开。不过，ECPG是PostgreSQL的嵌入式 SQL 预处理器，它支持标准 SQL 的游标约定，包括涉及DECLARE和OPEN语句的那些约定。

示例

声明一个游标：

```
DECLARE liahona CURSOR FOR SELECT * FROM films;
```

更多游标用法示例见FETCH。

兼容性

SQL 标准只允许在嵌入式SQL和模块中使用游标。PostgreSQL允许以交互方式使用游标。

SQL 标准允许游标更新表数据。所有 PostgreSQL游标都是只读的。

二进制游标是PostgreSQL的一种扩展。

另见

CLOSE, FETCH, MOVE

DELETE

DELETE — 删除表中的行

大纲

```
DELETE FROM [ ONLY ] table
    [ USING usinglist ]
    [ WHERE condition ]
```

描述

DELETE从指定表中删除满足 WHERE子句的行。如果省略WHERE 子句，效果就是删除表中的所有行。结果是一个合法但为空的表。

默认情况下，DELETE将删除指定表及其所有子表中的行。如果只想从提及的特定表中删除，必须使用ONLY子句。

> **提示**
>
> TRUNCATE是一种PostgreSQL扩展，它提供了一种更快的机制来移除表中的所有行。

有两种方法可以利用数据库中其他表所包含的信息来删除一个表中的行：使用子选择，或者在USING子句中指定附加表。哪种技术更合适取决于具体情况。

要从一个表中删除行，必须具有该表上的DELETE权限，以及USING子句中出现的任何表或者其值会在 *condition*中读取的任何表上的SELECT权限。

参数

ONLY

如果指定，则只从所提及的表中删除行。未指定时，继承自该表的任何表也会被处理。

table

现有表的名称（可以是模式限定的）。

usinglist

一个表表达式列表，允许其他表的列出现在WHERE条件中。这类似于SELECT语句的FROM Clause中可以指定的表列表；例如，可以为表名指定别名。不要把目标表重复写在 *usinglist*中，除非打算进行自连接。

condition

一个返回boolean类型值的表达式，它决定要删除哪些行。

输出

成功完成后，DELETE 命令返回如下形式的命令标签：

`DELETE` *`count`*

其中 *`count`* 是被删除的行数。如果 *`count`* 为 0，说明没有行匹配 *`condition`*（这不算错误）。

注解

PostgreSQL允许通过在 `USING`子句中指定其他表，在 `WHERE`条件中引用这些表的列。例如，要删除由给定制片人制作的所有电影，可以这样做：

```
DELETE FROM films USING producers
  WHERE producer_id = producers.id AND producers.name = 'foo';
```

本质上，这里发生的是在`films`和 `producers`之间进行连接，并将所有成功连接到的`films`行标记为删除。这种语法不是标准的。更标准的写法是

```
DELETE FROM films
  WHERE producer_id IN (SELECT id FROM producers WHERE name = 'foo');
```

在某些情况下，连接形式比子选择形式更容易书写或者执行更快。

示例

删除所有电影，但音乐剧除外：

```
DELETE FROM films WHERE kind <> 'Musical';
```

清空表`films`：

```
DELETE FROM films;
```

兼容性

这个命令符合 SQL 标准，不过`USING`子句以及在 `WHERE`子句中引用其他表的能力是PostgreSQL扩展。

DROP AGGREGATE

DROP AGGREGATE — 移除一个聚合函数

大纲

```
DROP AGGREGATE name ( type [ , ... ] ) [ CASCADE | RESTRICT ]
```

描述

`DROP AGGREGATE`移除一个现有聚合函数。要执行此命令，当前用户必须是该聚合函数的拥有者。

参数

name

现有聚合函数的名称（可以被模式限定）。

type

聚合函数作用于其上的输入数据类型。要引用一个零参数聚合函数，请在参数说明列表的位置写`*`。

`CASCADE`

自动删除依赖于该聚合函数的对象。

`RESTRICT`

如果有任何对象依赖于该聚合函数，则拒绝删除它。这是默认值。

示例

要移除用于类型`integer`的聚合函数`myavg`：

```
DROP AGGREGATE myavg(integer);
```

兼容性

SQL 标准中没有`DROP AGGREGATE`语句。

另见

ALTER AGGREGATE, CREATE AGGREGATE

DROP CAST

DROP CAST — 移除一个类型转换

大纲

```
DROP CAST (sourcetype AS targettype) [ CASCADE | RESTRICT ]
```

描述

`DROP CAST`移除一个先前定义的类型转换。

要删除一种类型转换，你必须拥有其源数据类型或目标数据类型。
这与创建该类型转换所需的权限相同。

参数

sourcetype

该类型转换的源数据类型的名称。

targettype

该类型转换的目标数据类型的名称。

`CASCADE`
`RESTRICT`

这些关键字没有任何作用，因为没有对象依赖于类型转换。

示例

要删除从类型`text`到类型`int`的类型转换：

```
DROP CAST (text AS int);
```

兼容性

`DROP CAST`命令符合 SQL 标准。

另见

CREATE CAST

DROP CONVERSION

DROP CONVERSION — 移除一个转换

大纲

```
DROP CONVERSION name [ CASCADE | RESTRICT ]
```

描述

`DROP CONVERSION`移除一个先前定义的转换。要能够删除一个转换，你必须拥有该转换。

参数

name

转换的名称。转换名可以是模式限定的。

`CASCADE`
`RESTRICT`

这些关键字没有任何作用，因为没有对象依赖于转换。

示例

移除名为`myname`的转换：

```
DROP CONVERSION myname;
```

兼容性

SQL 标准中没有`DROP CONVERSION`语句。

另见

ALTER CONVERSION, CREATE CONVERSION

DROP DATABASE

DROP DATABASE — 移除一个数据库

大纲

`DROP DATABASE name`

描述

`DROP DATABASE`删除一个数据库。它会移除数据库的系统目录条目，并删除包含数据的目录。只有数据库拥有者才能执行此命令。此外，当你或任何其他人连接到目标数据库时，都不能执行它。（请连接到`postgres`或任何其他数据库来发出此命令。）

`DROP DATABASE`无法撤销。请谨慎使用！

参数

name

要移除的数据库名称。

注解

`DROP DATABASE`不能在事务块内执行。

由于该命令不能在连接到目标数据库时执行，因此改用程序dropdb可能更方便，它是此命令的一个包装器。

兼容性

SQL 标准中没有`DROP DATABASE`语句。

另见

CREATE DATABASE

DROP DOMAIN

DROP DOMAIN — 移除一个域

大纲

```
DROP DOMAIN name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP DOMAIN`移除一个域。只有域的拥有者才能移除它。

参数

name

要移除的现有域的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该域的对象（例如表列）。

`RESTRICT`

如果有任何对象依赖于该域，则拒绝删除它。这是默认值。

示例

要移除域`box`：

```
DROP DOMAIN box;
```

兼容性

此命令符合 SQL 标准。

另见

CREATE DOMAIN, ALTER DOMAIN

DROP FUNCTION

DROP FUNCTION — 移除一个函数

大纲

```
DROP FUNCTION name ( [ [ argmode ] [ argname ] argtype [, ...] ] )
    [ CASCADE | RESTRICT ]
```

描述

`DROP FUNCTION`移除一个现有函数的定义。要执行此命令，用户必须是该函数的拥有者。必须指定该函数的参数类型，因为可能存在多个名称相同但参数列表不同的函数。

参数

name

现有函数的名称（可以被模式限定）。

argmode

参数的模式：`IN`、`OUT`或 `INOUT`。如果省略，默认值为`IN`。注意，`DROP FUNCTION`实际上并不关注`OUT`参数，因为确定函数标识只需要输入参数。因此，只列出 `IN`和`INOUT` 参数就足够了。

argname

一个参数的名称。注意，`DROP FUNCTION` 实际上并不关注参数名，因为确定函数标识只需要参数数据类型。

argtype

函数参数的数据类型（如果有，可以被模式限定）。

`CASCADE`

自动删除依赖于该函数的对象（例如操作符或触发器）。

`RESTRICT`

如果有任何对象依赖于该函数，则拒绝删除它。这是默认值。

示例

此命令删除平方根函数：

```
DROP FUNCTION sqrt(integer);
```

兼容性

`DROP FUNCTION`语句在 SQL 标准中有定义，但与该命令不兼容。

另见

CREATE FUNCTION, ALTER FUNCTION

DROP GROUP

DROP GROUP — 移除一个数据库角色

大纲

```
DROP GROUP name [, ...]
```

描述

`DROP GROUP` 现在是DROP ROLE的别名。

兼容性

SQL 标准中没有 `DROP GROUP` 语句。

另见

DROP ROLE

DROP INDEX

DROP INDEX — 移除一个索引

大纲

```
DROP INDEX name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP INDEX`从数据库系统中删除一个现有索引。要执行此命令，你必须是该索引的拥有者。

参数

name

要移除的索引名称（可以被模式限定）。

`CASCADE`

自动删除依赖于该索引的对象。

`RESTRICT`

如果有任何对象依赖于该索引，则拒绝删除它。这是默认值。

示例

此命令会删除索引`title_idx`：

```
DROP INDEX title_idx;
```

兼容性

`DROP INDEX`是PostgreSQL 的一种语言扩展。SQL 标准中没有关于索引的规定。

另见

CREATE INDEX

DROP LANGUAGE

DROP LANGUAGE — 移除一种过程语言

大纲

```
DROP [ PROCEDURAL ] LANGUAGE name [ CASCADE | RESTRICT ]
```

描述

`DROP LANGUAGE`将移除先前注册的名为 *name*的过程语言的定义。

参数

name

现有过程语言的名称。为向后兼容，名称可以用单引号括起。

`CASCADE`

自动删除依赖于该语言的对象（例如以该语言编写的函数）。

`RESTRICT`

如果有任何对象依赖于该语言，则拒绝删除它。这是默认值。

示例

此命令移除过程语言`plsample`：

```
DROP LANGUAGE plsample;
```

兼容性

SQL 标准中没有`DROP LANGUAGE`语句。

另见

ALTER LANGUAGE, CREATE LANGUAGE, droplang

DROP OPERATOR

DROP OPERATOR — 移除一个操作符

大纲

```
DROP OPERATOR name ( { lefttype | NONE } , { righttype | NONE } ) [
 CASCADE | RESTRICT ]
```

描述

DROP OPERATOR从数据库系统中删除一个现有操作符。要执行此命令，你必须是该操作符的拥有者。

参数

name

一个现有操作符的名称（可以是模式限定的）。

lefttype

操作符左操作数的数据类型；如果该操作符没有左操作数，请写 NONE。

righttype

操作符右操作数的数据类型；如果操作符没有右操作数，请写NONE。

CASCADE

自动删除依赖于该操作符的对象。

RESTRICT

如果有任何对象依赖于该操作符，则拒绝删除它。这是默认值。

示例

删除幂操作符a^b，其类型为 integer：

```
DROP OPERATOR ^ (integer, integer);
```

删除前缀一元按位取反操作符 ~b，其类型为 bit：

```
DROP OPERATOR ~ (none, bit);
```

删除后缀一元阶乘操作符 x!，其类型为 bigint：

```
DROP OPERATOR ! (bigint, none);
```

兼容性

SQL 标准中没有DROP OPERATOR语句。

另见

CREATE OPERATOR, ALTER OPERATOR

DROP OPERATOR CLASS

DROP OPERATOR CLASS — 移除一个操作符类

大纲

```
DROP OPERATOR CLASS name USING index_method [ CASCADE | RESTRICT ]
```

描述

`DROP OPERATOR CLASS`删除一个现有操作符类。要执行此命令，你必须是该操作符类的拥有者。

参数

name

一个现有操作符类的名称（可以被模式限定）。

index_method

该操作符类所属索引访问方法的名称。

`CASCADE`

自动删除依赖于该操作符类的对象。

`RESTRICT`

如果有任何对象依赖于该操作符类，则拒绝删除它。这是默认值。

示例

移除 B-树操作符类`widget_ops`：

```
DROP OPERATOR CLASS widget_ops USING btree;
```

如果仍有索引使用该操作符类，此命令将不会成功。加上`CASCADE`
可以在删除操作符类的同时删除这类索引。

兼容性

SQL 标准中没有`DROP OPERATOR CLASS`语句。

另见

ALTER OPERATOR CLASS, CREATE OPERATOR CLASS

DROP ROLE

DROP ROLE — 移除一个数据库角色

大纲

```
DROP ROLE name [, ...]
```

描述

`DROP ROLE`移除指定的角色。要删除超级用户角色，你自己也必须是超级用户；要删除非超级用户角色，必须拥有`CREATEROLE`权限。

如果某角色仍在该集簇的任何数据库中被引用，就不能移除它，否则会报错。在删除角色之前，必须删除它拥有的所有对象（或重新分配其拥有关系），
并撤销该角色在其他对象上被授予的任何权限。

不过，不必手工移除与该角色相关的角色成员资格；`DROP ROLE`会自动撤销目标角色在其他角色中的任何成员资格，以及其他角色在目标角色中的任何成员资格。其他角色不会被删除，
也不会受到影响。

参数

name

要移除的角色名称。

注解

PostgreSQL包含一个程序dropuser，它与此命令具有相同功能（实际上它就是调用此命令），但可以从命令 shell 中运行。

示例

删除一个角色：

```
DROP ROLE jonathan;
```

兼容性

SQL 标准定义了`DROP ROLE`，但它一次只允许删除一个角色，并且规定的权限要求也不同于PostgreSQL。

另见

CREATE ROLE, ALTER ROLE, SET ROLE

DROP RULE

DROP RULE — 移除一条重写规则

大纲

```
DROP RULE name ON relation [ CASCADE | RESTRICT ]
```

描述

`DROP RULE`移除一条重写规则。

参数

name

要移除的规则名称。

relation

该规则适用的表或视图名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该规则的对象。

`RESTRICT`

如果有任何对象依赖于该规则，则拒绝删除它。这是默认值。

示例

要移除重写规则`newrule`：

```
DROP RULE newrule ON mytable;
```

兼容性

SQL 标准中没有 `DROP RULE` 语句。

另见

CREATE RULE

DROP SCHEMA

DROP SCHEMA — 移除一个模式

大纲

```
DROP SCHEMA name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP SCHEMA`从数据库中移除一个或多个模式。

一个模式只能由其拥有者或超级用户删除。注意，
即使该模式的拥有者并不拥有模式内的某些对象，
也仍然可以删除该模式（从而删除其中的全部对象）。

参数

name

模式名称。

`CASCADE`

自动删除模式中包含的对象（表、函数等）。

`RESTRICT`

如果该模式中含有任何对象，则拒绝删除它。这是默认值。

示例

要从数据库中移除模式`mystuff`及其包含的全部内容：

```
DROP SCHEMA mystuff CASCADE;
```

兼容性

`DROP SCHEMA`完全符合 SQL 标准，但标准每条命令只允许删除一个模式。

另见

ALTER SCHEMA, CREATE SCHEMA

DROP SEQUENCE

DROP SEQUENCE — 移除一个序列

大纲

```
DROP SEQUENCE name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP SEQUENCE`移除序列号生成器。

参数

name

要删除的序列名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该序列的对象。

`RESTRICT`

如果有任何对象依赖于该序列，则拒绝删除它。这是默认值。

示例

要移除序列`serial`：

```
DROP SEQUENCE serial;
```

兼容性

`DROP SEQUENCE`符合SQL标准，但标准每条命令只允许删除一个序列。

另见

CREATE SEQUENCE

DROP TABLE

DROP TABLE — 移除一个表

大纲

```
DROP TABLE name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP TABLE` 从数据库中删除表。只有表的拥有者才能删除表。要在不破坏表的情况下清空表中的行，可使用`DELETE`。

`DROP TABLE`总会移除目标表上存在的所有索引、规则、触发器和约束。但是，要删除一个被视图或另一张表的外键约束引用的表，必须指定`CASCADE`。（`CASCADE`
会将依赖于该表的视图整个删除，但在外键场景下，它只会删除外键约束，
不会把另一张表整体删除。）

参数

name

要删除的表名称（可选地使用模式限定）。

`CASCADE`

自动删除依赖于该表的对象（例如视图）。

`RESTRICT`

如果有任何对象依赖于该表，则拒绝删除它。这是默认值。

示例

要删除两个表`films`和`distributors`：

```
DROP TABLE films, distributors;
```

兼容性

此命令符合 SQL 标准，只是标准每条命令只允许删除一个表。

另见

ALTER TABLE, CREATE TABLE

DROP TABLESPACE

DROP TABLESPACE — 移除一个表空间

大纲

```
DROP TABLESPACE tablespacename
```

描述

`DROP TABLESPACE`从系统中移除一个表空间。

一个表空间只能由其拥有者或超级用户删除。删除前，该表空间中必须不含任何数据库对象。
即使当前数据库中没有对象正在使用该表空间，
其他数据库中的对象也可能仍然驻留在该表空间中。

参数

tablespacename

表空间的名称。

示例

要从系统中移除表空间`mystuff`：

```
DROP TABLESPACE mystuff;
```

兼容性

`DROP TABLESPACE`是一种PostgreSQL扩展。

另见

CREATE TABLESPACE, ALTER TABLESPACE

DROP TRIGGER

DROP TRIGGER — 移除一个触发器

大纲

```
DROP TRIGGER name ON table [ CASCADE | RESTRICT ]
```

描述

`DROP TRIGGER`移除一个现有触发器的定义。要执行此命令，当前用户必须是定义该触发器的表的所有者。

参数

name

要移除的触发器的名称。

table

定义该触发器的表的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该触发器的对象。

`RESTRICT`

如果有任何对象依赖于该触发器，则拒绝删除它。这是默认值。

示例

删除表`films`上的触发器`if_dist_exists`：

```
DROP TRIGGER if_dist_exists ON films;
```

兼容性

PostgreSQL中的`DROP TRIGGER`语句与 SQL 标准不兼容。在 SQL 标准中，触发器名称不是表局部的，因此该命令仅为 `DROP TRIGGER name`。

参见

CREATE TRIGGER

DROP TYPE

DROP TYPE — 移除一个数据类型

大纲

```
DROP TYPE name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP TYPE`移除一个用户定义的数据类型。只有该类型的拥有者才能移除它。

参数

name

要移除的数据类型的名称（可以是模式限定的）。

`CASCADE`

自动删除依赖于该类型的对象（例如表列、函数和操作符）。

`RESTRICT`

如果有任何对象依赖于该类型，则拒绝删除它。这是默认值。

示例

要移除数据类型`box`：

```
DROP TYPE box;
```

兼容性

此命令与 SQL 标准中的对应命令相似，但请注意，`CREATE TYPE`命令以及 PostgreSQL中的数据类型扩展机制都不同于 SQL 标准。

另见

CREATE TYPE, ALTER TYPE

DROP USER

DROP USER — 移除一个数据库角色

大纲

```
DROP USER name [, ...]
```

描述

`DROP USER`现在是DROP ROLE的一个别名。

兼容性

`DROP USER`语句是 PostgreSQL的一种扩展。SQL 标准将用户的定义留给具体实现决定。

另见

DROP ROLE

DROP VIEW

DROP VIEW — 移除一个视图

大纲

```
DROP VIEW name [, ...] [ CASCADE | RESTRICT ]
```

描述

`DROP VIEW`删除一个现有视图。要执行此命令，你必须是该视图的拥有者。

参数

name

要移除的视图名称（可选地使用模式限定）。

`CASCADE`

自动删除依赖于该视图的对象（例如其他视图）。

`RESTRICT`

如果有任何对象依赖于该视图，则拒绝删除它。这是默认值。

示例

这个命令将移除名为`kinds`的视图：

```
DROP VIEW kinds;
```

兼容性

此命令符合 SQL 标准，只是标准每条命令只允许删除一个视图。

另见

CREATE VIEW

END

END — 提交当前事务

大纲

```
END [ WORK | TRANSACTION ]
```

描述

`END`提交当前事务。该事务所做的所有更改都会对其他会话可见，并且如果发生崩溃，也能保证其持久性。该命令是PostgreSQL扩展，等效于COMMIT。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用ROLLBACK中止事务。

在事务块之外发出`END`不会造成任何影响，但会产生一条警告消息。

示例

提交当前事务，并使所有更改永久生效：

```
END;
```

兼容性

命令`END`是PostgreSQL扩展，提供了与 SQL 标准规定的COMMIT等效的功能。

另见

BEGIN, COMMIT, ROLLBACK

EXECUTE

EXECUTE — 执行一个预备语句

大纲

```
EXECUTE plan_name [ (parameter [, ...] ) ]
```

描述

EXECUTE用于执行一个先前创建的预备语句。由于预备语句只在一个会话期间内存在，因此该预备语句必须由当前会话中较早执行的PREPARE语句创建。

如果创建该语句的PREPARE语句指定了某些参数，那么传递给EXECUTE语句的参数集合就必须与之兼容，否则将引发错误。注意，预备语句（不同于函数）不会根据其参数的类型或个数进行重载；预备语句的名称在一个数据库会话内必须唯一。

有关预备语句的创建和用法的更多信息，请参见PREPARE。

参数

plan_name

要执行的预备语句名称。

parameter

预备语句中某个参数的实际值。它必须是一个表达式，其求值结果的类型需与创建该预备语句的PREPARE命令中为此参数位置指定的数据类型兼容。

输出

EXECUTE返回的命令标签是该预备语句的命令标签，而不是EXECUTE。

Examples

示例位于示例一节，详见PREPARE文档。

Compatibility

SQL 标准包含EXECUTE语句，但它只用于嵌入式 SQL。这里的EXECUTE语句还使用了稍有不同的语法。

参见

DEALLOCATE, PREPARE

EXPLAIN

EXPLAIN — 显示一个语句的执行计划

大纲

```
EXPLAIN [ ANALYZE ] [ VERBOSE ] statement
```

描述

这个命令显示PostgreSQL规划器为给定语句生成的执行计划。
执行计划会显示将如何扫描该语句引用的表 – 例如普通顺序扫描、索引扫描等 –，如果引用了多个表，还会显示将使用哪些连接算法来汇集每个输入表中的所需行。

显示结果中最关键的部分是语句执行代价的估计值，
它是规划器对运行该语句需要多长时间的猜测（以任意代价单位衡量，
但按惯例表示磁盘页面抓取次数）。实际上会显示两个数字：返回第一行之前的启动代价，
以及返回全部行的总代价。对大多数查询来说，总代价才是关键；但在某些场景中，例如EXISTS中的子查询，
规划器会选择启动代价最小而不是总代价最小的计划（因为无论如何执行器都会在得到一行后停止）。此外，如果你用LIMIT子句限制返回的行数，规划器会在端点代价之间作适当插值，
以估计哪个计划实际上最便宜。

ANALYZE 选项导致语句被真正执行，而不仅仅是规划。每个计划节点内部耗费的总时间（以毫秒计）以及它实际返回的总行数会被加入显示。这有助于查看规划器的估计是否接近现实。

> **重要**
>
> 请记住，使用 ANALYZE 选项时，语句会被实际执行。虽然 EXPLAIN 会丢弃 SELECT 返回的任何输出，但语句的其他副作用仍会照常发生。如果希望使用 EXPLAIN ANALYZE 分析 INSERT，UPDATE，DELETE 或 EXECUTE 语句而不让命令影响数据，可以采用以下方法：
>
> ```
> BEGIN;
> EXPLAIN ANALYZE ...;
> ROLLBACK;
> ```

参数

ANALYZE

执行命令并显示实际的运行时间。

VERBOSE

显示计划树的完整内部表示，而不只是摘要。通常此选项只对专门的调试目的有用。
VERBOSE输出是否美观打印取决于explain_pretty_print配置参数的设置。

statement

任何你希望查看其执行计划的SELECT、INSERT、UPDATE、DELETE、EXECUTE
或DECLARE语句。

注解

关于 PostgreSQL 中优化器对代价信息的使用，文档很少。更多信息参见第 13.1 节。

为了让 PostgreSQL 查询规划器在优化查询时能做出合理知情的决定，应当运行ANALYZE语句来记录表中数据分布的统计信息。如果你没有这样做（或者自上次运行 ANALYZE 以来表中数据的统计分布发生了显著变化），估计的代价不太可能符合查询的真实属性，
因此可能选出一个较差的查询计划。

遗传查询优化（GEQO）随机测试执行计划。因此，当连接关系数超过geqo_threshold
从而启用遗传查询优化时，每次执行该语句时执行计划都很可能发生变化。

在PostgreSQL 7.3 之前，执行计划是以NOTICE消息的形式发出的。现在它作为查询结果出现（格式化为一个只有单个文本列的表）。

示例

要显示一个只有单个integer列且包含 10000 行的表上的简单查询计划：

```
EXPLAIN SELECT * FROM foo;

                       QUERY PLAN
-----------------------------------------------------------
 Seq Scan on foo  (cost=0.00..155.00 rows=10000 width=4)
(1 row)
```

如果存在索引，并且我们使用了带有可索引WHERE条件的查询，EXPLAIN
可能会显示不同的计划：

```
EXPLAIN SELECT * FROM foo WHERE i = 4;

                         QUERY PLAN
--------------------------------------------------------------
 Index Scan using fi on foo  (cost=0.00..5.98 rows=1 width=4)
   Index Cond: (i = 4)
(2 rows)
```

下面是使用聚合函数的查询的查询计划示例：

```
EXPLAIN SELECT sum(i) FROM foo WHERE i < 10;

                             QUERY PLAN
---------------------------------------------------------------------
 Aggregate  (cost=23.93..23.93 rows=1 width=4)
   ->  Index Scan using fi on foo  (cost=0.00..23.92 rows=6 width=4)
         Index Cond: (i < 10)
(3 rows)
```

下面是使用 EXPLAIN EXECUTE 显示预备查询执行计划的示例：

```
PREPARE query(int, int) AS SELECT sum(bar) FROM test
    WHERE id > $1 AND id < $2
    GROUP BY foo;
```

```
EXPLAIN ANALYZE EXECUTE query(100, 200);

                                                        QUERY PLAN

-----------------------------------------------------------------------
--------------------------------------------------
 HashAggregate  (cost=39.53..39.53 rows=1 width=8) (actual time=0.661.
.0.672 rows=7 loops=1)
   ->  Index Scan using test_pkey on test  (cost=0.00..32.97 rows=1311
 width=8) (actual time=0.050..0.395 rows=99 loops=1)
         Index Cond: ((id > $1) AND (id < $2))
 Total runtime: 0.851 ms
(4 rows)
```

当然，这里显示的具体数字取决于所涉及表的实际内容。还要注意，由于规划器的改进，这些数字甚至所选查询策略在PostgreSQL的不同版本之间都可能发生变化。此外，`ANALYZE`命令使用随机采样来估计数据统计信息；因此，即使表中数据的实际分布没有变化，在重新运行一次`ANALYZE`之后，代价估计也可能发生变化。

兼容性

SQL 标准中没有定义`EXPLAIN`语句。

另见

ANALYZE

FETCH

FETCH — 使用游标从查询中检索行

大纲

```
FETCH [ direction [ FROM | IN ] ] cursorname

where direction can be empty or one of:

    NEXT
    PRIOR
    FIRST
    LAST
    ABSOLUTE count
    RELATIVE count
    count
    ALL
    FORWARD
    FORWARD count
    FORWARD ALL
    BACKWARD
    BACKWARD count
    BACKWARD ALL
```

描述

FETCH使用先前创建的游标检索行。

游标带有一个关联位置，FETCH会使用该位置。游标位置可以位于查询结果的第一行之前、结果中的任意一行上，或者位于结果的最后一行之后。游标创建时位于第一行之前。提取若干行后，游标会定位在最近提取的那一行上。如果FETCH 越过了可用行的末尾，那么游标会留在最后一行之后；如果是向后提取，则会留在第一行之前。FETCH ALL或者 FETCH BACKWARD ALL总是会让游标位于最后一行之后或第一行之前。

NEXT、PRIOR、FIRST、LAST、ABSOLUTE、RELATIVE 这些形式会在适当地移动游标后提取一行。如果不存在这样的行，则返回空结果，并根据情况将游标定位在第一行之前或最后一行之后。

使用FORWARD和BACKWARD的形式会在向前或向后的方向上提取指定数量的行，并将游标定位在最后返回的那一行上（如果*count*超过可用行数，则定位在所有行之后或之前）。

RELATIVE 0、FORWARD 0以及 BACKWARD 0都会请求提取当前行而不移动游标，也就是重新提取最近一次提取的行。除非游标位于第一行之前或最后一行之后，否则该操作都会成功；在这两种情况下，不会返回任何行。

参数

direction

direction 定义抓取方向和要抓取的行数。它可以是以下值之一：

NEXT

提取下一行。如果省略*direction*，这将是默认值。

PRIOR

提取前一行。

FIRST

提取查询的第一行（与ABSOLUTE 1相同）。

LAST

提取查询的最后一行（与ABSOLUTE -1相同）。

ABSOLUTE *count*

提取查询的第*count*行；如果*count*为负，则提取从末尾算起的第abs(*count*)行。如果 *count*超出范围，则定位在第一行之前或最后一行之后。特别地，ABSOLUTE 0会定位在第一行之前。

RELATIVE *count*

抓取后面的第*count*行，或前面的第abs(*count*)行（当*count*为负时）。RELATIVE 0会重新抓取当前行（如果存在）。

count

提取接下来的*count*行（与FORWARD *count*相同）。

ALL

提取所有剩余行（与FORWARD ALL相同）。

FORWARD

提取下一行（与NEXT相同）。

FORWARD *count*

提取接下来的*count*行。FORWARD 0会重新提取当前行。

FORWARD ALL

提取所有剩余行。

BACKWARD

提取前一行（与PRIOR相同）。

BACKWARD *count*

提取前面的*count*行（反向扫描）。BACKWARD 0会重新提取当前行。

BACKWARD ALL

提取之前的所有行（反向扫描）。

count

*count*是一个可以带符号的整数常量，决定抓取位置或抓取行数。对于FORWARD和BACKWARD，指定负的*count*等同于交换FORWARD和BACKWARD的方向。

cursorname

一个已打开游标的名称。

输出

成功完成时，FETCH命令会返回如下形式的命令标签：

FETCH *count*

其中*count*是提取到的行数（可能为零）。注意，在psql中，命令标签实际上不会显示，因为psql会改为显示提取到的行。

注解

如果打算使用除FETCH NEXT或带正数计数的 FETCH FORWARD之外的任何FETCH 变体，则应当在声明游标时使用SCROLL选项。对于简单查询，PostgreSQL允许从未使用SCROLL 声明的游标向后提取，但最好不要依赖这种行为。如果游标声明为 NO SCROLL，则不允许向后提取。

ABSOLUTE提取并不比通过相对移动到达目标行更快：无论如何，
底层实现都必须遍历所有中间行。
负值的绝对提取甚至更糟：必须先把查询读到末尾以找到最后一行，然后再从那里反向遍历。
不过，回卷到查询起始处（如FETCH ABSOLUTE 0）是很快的。

目前，PostgreSQL还不支持通过游标更新数据。

DECLARE用于定义游标。使用MOVE可以在不检索数据的情况下改变游标位置。

示例

下面的示例展示了如何使用游标遍历一个表：

```
BEGIN WORK;

-- 建立一个游标:
DECLARE liahona SCROLL CURSOR FOR SELECT * FROM films;

-- 从游标 liahona 中提取前 5 行:
FETCH FORWARD 5 FROM liahona;

 code  |          title          | did | date_prod  |   kind   |  len
-------+-------------------------+-----+------------+----------+-----
--
 BL101 | The Third Man           | 101 | 1949-12-23 | Drama    | 01:44
 BL102 | The African Queen       | 101 | 1951-08-11 | Romantic | 01:43
 JL201 | Une Femme est une Femme | 102 | 1961-03-12 | Romantic | 01:25
 P_301 | Vertigo                 | 103 | 1958-11-14 | Action   | 02:08
 P_302 | Becket                  | 103 | 1964-02-03 | Drama    | 02:28
```

```
-- 提取前一行:
FETCH PRIOR FROM liahona;

 code  |  title  | did | date_prod  |  kind  |  len
-------+---------+-----+------------+--------+-------
 P_301 | Vertigo | 103 | 1958-11-14 | Action | 02:08

-- 关闭游标并结束事务:
CLOSE liahona;
COMMIT WORK;
```

兼容性

SQL 标准只为嵌入式 SQL 定义了FETCH。这里描述的FETCH变体会像 SELECT结果那样返回数据，而不是把数据放入主变量中。除此之外，FETCH与 SQL 标准完全向上兼容。

涉及FORWARD和BACKWARD的 FETCH形式，以及形式FETCH *count*和FETCH ALL（其中隐含了FORWARD）都是 PostgreSQL扩展。

SQL 标准只允许在游标名之前使用FROM；允许使用 IN，或者将两者都完全省略，都是扩展。

另见

CLOSE, DECLARE, MOVE

GRANT

GRANT — 定义访问权限

大纲

```
GRANT { { SELECT | INSERT | UPDATE | DELETE | RULE | REFERENCES |
 TRIGGER }
    [,...] | ALL [ PRIVILEGES ] }
    ON [ TABLE ] tablename [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT { { CREATE | TEMPORARY | TEMP } [,...] | ALL [ PRIVILEGES ] }
    ON DATABASE dbname [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT { EXECUTE | ALL [ PRIVILEGES ] }
    ON FUNCTION funcname ( [ [ argmode ] [ argname ] argtype [, ...] ]
 ) [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT { USAGE | ALL [ PRIVILEGES ] }
    ON LANGUAGE langname [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT { { CREATE | USAGE } [,...] | ALL [ PRIVILEGES ] }
    ON SCHEMA schemaname [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT { CREATE | ALL [ PRIVILEGES ] }
    ON TABLESPACE tablespacename [, ...]
    TO { username | GROUP groupname | PUBLIC } [, ...] [ WITH GRANT
 OPTION ]

GRANT role [, ...] TO username [, ...] [ WITH ADMIN OPTION ]
```

描述

`GRANT`命令有两种基本变体：一种用于授予数据库对象（表、视图、序列、数据库、函数、过程语言、模式或表空间）上的权限，另一种用于授予角色的成员资格。
这两种变体在许多方面很相似，但差别也足够大，因此分别介绍。

从 PostgreSQL 8.1 起，用户和组的概念已统一为一种称为角色的单一实体。因此，不再需要使用 `GROUP` 关键字来区分权限接收者是用户还是组。命令中仍然允许 `GROUP`，但它只是一个无实际意义的词。

在数据库对象上 GRANT

这种 GRANT 命令变体将数据库对象上的特定权限授予一个或多个角色。如果此前已经授予过某些权限，新授予的权限会加到现有权限之上。

关键字 PUBLIC 表示要把权限授予所有角色，包括以后可能创建的角色。PUBLIC 可以视为一个隐式定义的组，并且始终包含所有角色。任何特定角色实际拥有的权限，
是直接授予给它的权限、授予给它当前所属任一角色的权限，以及授予给 PUBLIC 的权限之和。

如果指定了 WITH GRANT OPTION，权限接收者随后可以再把该权限授予其他人。没有授予选项时，接收者不能这样做。授予选项不能授予给 PUBLIC。

没有必要向对象拥有者（通常是创建它的用户）授予权限，因为拥有者默认拥有全部权限。
（不过，出于安全考虑，拥有者也可以选择撤销自己的某些权限。）
删除对象或以任何方式更改其定义的权利，不被视为一种可授予的权限；它是拥有者固有的，
不能被授予或撤销。拥有者还隐式拥有该对象上的全部授予选项。

根据对象类型的不同，初始的默认权限可能包括向 PUBLIC 授予某些权限。默认情况下，表、模式和表空间没有公共访问权限；数据库有创建 TEMP 表的权限；函数有 EXECUTE 权限；语言有 USAGE 权限。当然，对象拥有者可以撤销这些权限。（为了最大程度的安全，应在创建对象的同一个事务中发出 REVOKE；这样其他用户就没有可以使用该对象的时间窗口。）

可用权限如下：

SELECT

允许使用SELECT读取指定表、视图或序列的任何列。还允许使用COPY TO。在UPDATE或DELETE中引用现有列值时也需要此权限。对于序列，此权限还允许使用currval函数。对于大对象，此权限允许读取该对象。

INSERT

允许使用INSERT向指定表插入新行。还允许使用COPY FROM。

UPDATE

允许使用UPDATE更新指定表的任何列。（实际上，任何非简单的UPDATE命令还需要SELECT权限，因为它必须引用表列，以确定要更新的行和/或计算列的新值。）SELECT ... FOR UPDATE和SELECT ... FOR SHARE也要求具有此权限，此外还需要SELECT权限。对于序列，此权限允许使用nextval和setval函数。对于大对象，此权限允许写入或截断该对象。

DELETE

允许使用DELETE从指定表删除一行。（实际上，任何非简单的DELETE命令还需要SELECT权限，因为它必须引用表列，以确定要删除的行。）

RULE

允许在表/视图上创建规则。（参见CREATE RULE语句。）

REFERENCES

要创建外键约束，必须在引用表和被引用表上都拥有此权限。

TRIGGER

允许在指定表上创建触发器。（参见CREATE TRIGGER语句。）

CREATE

对于数据库，允许在数据库中创建新的模式。

对于模式，允许在模式中创建新对象。要重命名现有对象，必须拥有该对象，并且在包含它的模式上具有此权限。

对于表空间，允许在其中创建表和索引，也允许创建以该表空间为默认表空间的数据库。（注意，撤销此权限不会改变现有对象的存放位置。）

TEMPORARY
TEMP

允许在使用指定数据库时创建临时表。

EXECUTE

允许使用指定的函数，以及基于该函数实现的任何操作符。这是唯一适用于函数的权限类型。（此语法也适用于聚合函数。）

USAGE

对于过程语言，允许使用指定语言创建以该语言编写的函数。这是唯一适用于过程语言的权限类型。

对于模式，允许访问指定模式中包含的对象（假定也满足对象自身的权限要求）。本质上，这允许被授权者“查找”模式内的对象。

ALL PRIVILEGES

一次授予所有可用权限。`PRIVILEGES`关键字在PostgreSQL中是可选的，但在严格 SQL 中是必需的。

其他命令所需的权限列在相应命令的参考页上。

角色上的 GRANT

这种`GRANT`命令变体把一个角色的成员资格授予一个或多个其他角色。角色成员资格之所以重要，是因为它会把授予该角色的权限传递给其每个成员。

如果指定了`WITH ADMIN OPTION`，成员就可以将该角色的成员资格继续授予其他人，也可以撤销该角色的成员资格。没有管理选项时，普通用户不能这样做。不过，数据库超级用户可以向任何人授予或撤销任何角色的成员资格。具有`CREATEROLE`权限的角色可以授予或撤销任何非超级用户角色的成员资格。

与权限不同，角色成员资格不能授予`PUBLIC`。还要注意，这种形式的命令不允许使用无实际作用的`GROUP`一词。

注解

REVOKE命令用于撤销访问权限。

目前，PostgreSQL不支持授予或撤销表中单个列上的权限。一种可能的变通办法是创建一个只包含所需列的视图，然后把权限授予该视图。

当对象的非拥有者试图在该对象上执行 `GRANT` 时，如果该用户在该对象上完全没有任何权限，命令会立即失败。只要有某项权限可用，命令就会继续执行，

但只会授予那些该用户持有授予选项的权限。如果未持有任何授予选项，GRANT ALL PRIVILEGES 形式会发出警告；而其他形式如果命令中特别列出的任一权限未持有其授予选项，也会发出警告。（原则上，这些说明也适用于对象拥有者；
但由于拥有者总是被视为持有全部授予选项，这种情况实际上不会发生。）

需要注意，数据库超级用户可以访问所有对象，而不受对象权限设置的影响。这可类比于 Unix 系统中的 root 权限。和 root 一样，除非绝对必要，否则不宜以超级用户身份操作。

如果超级用户选择执行 GRANT 或 REVOKE 命令，该命令会像由受影响对象的拥有者发出一样执行。特别是，通过这种命令授予的权限看起来会像是由对象拥有者授予的。
（对于角色成员资格，则看起来像是由被授予成员资格的角色本身授予的。）

GRANT 和 REVOKE 也可以由并非受影响对象拥有者的角色执行，只要该角色是拥有该对象之角色的成员，或者是持有该对象上 WITH GRANT OPTION 权限之角色的成员。在这种情况下，权限会记录为由实际拥有该对象的角色，或者由持有 WITH GRANT OPTION 权限的角色授予。例如，如果表 t1 由角色 g1 拥有，而角色 u1 是它的成员，那么 u1 可以把 t1 上的权限授予给 u2，但这些权限看起来会像是直接由 g1 授予的。角色 g1 的任何其他成员之后都可以撤销这些权限。

如果执行 GRANT 的角色通过多条角色成员资格路径间接持有所需权限，则系统不会指明会被记录为执行该授权的是哪一个上层角色。在这种情况下，最佳做法是使用 SET ROLE 切换成你希望作为其身份执行 GRANT 的那个具体角色。

使用psql的 \z 命令可以获取现有权限的信息。例如：

```
=> \z mytable

                        Access privileges for database "lusitania"
 Schema |  Name   | Type  |                     Access privileges
--------+---------+-------+-----------------------------------------
-----------------
 public | mytable | table | {miriam=arwdRxt/miriam,=r/miriam,"group
 todos=arw/miriam"}
(1 row)
```

由 \z 显示的条目解释如下：

```
         uname=xxxx -- privileges granted to a user
   group gname=xxxx -- privileges granted to a group
              =xxxx -- privileges granted to PUBLIC

                  r -- SELECT ("read")
                  w -- UPDATE ("write")
                  a -- INSERT ("append")
                  d -- DELETE
                  R -- RULE
                  x -- REFERENCES
                  t -- TRIGGER
                  X -- EXECUTE
                  U -- USAGE
                  C -- CREATE
                  T -- TEMPORARY
            arwdRxt -- ALL PRIVILEGES (for tables)
                  * -- grant option for preceding privilege

              /yyyy -- user who granted this privilege
```

上例中的显示结果会出现在用户 `miriam` 创建表 `mytable` 并执行以下命令之后：

```
GRANT SELECT ON mytable TO PUBLIC;
GRANT SELECT, UPDATE, INSERT ON mytable TO GROUP todos;
```

如果某个对象的 “Access privileges” 列为空，表示该对象具有默认权限（即其权限列为 null）。默认权限始终包括所有者的全部权限，并且可能根据对象类型包含授予`PUBLIC`的某些权限，如上所述。在对象上首次执行`GRANT`或`REVOKE`时，会先实例化默认权限（例如生成`{miriam=arwdRxt/miriam}`），然后根据指定请求修改它们。

注意，访问权限显示中不会标记所有者隐含的授权选项。只有在显式向某人授予授权选项时，才会出现`*`。

示例

将表 `films` 上的插入权限授予所有用户：

```
GRANT INSERT ON films TO PUBLIC;
```

将视图 `kinds` 上的所有可用权限授予用户 `manuel`：

```
GRANT ALL PRIVILEGES ON kinds TO manuel;
```

请注意，如果上述命令由超级用户或 `kinds` 的拥有者执行，确实会授予所有权限；但如果由其他人执行，则只会授予该执行者持有授予选项的那些权限。

将角色 `admins` 的成员资格授予用户 `joe`：

```
GRANT admins TO joe;
```

兼容性

根据 SQL 标准，`ALL PRIVILEGES` 中的 `PRIVILEGES` 关键字是必需的。SQL 标准也不支持每条命令对多个对象设置权限。

PostgreSQL 允许对象拥有者撤销自己的普通权限：例如，表拥有者可以通过撤销自己的 INSERT、UPDATE 和 DELETE 权限，使该表对自己变成只读。这在 SQL 标准中是不可能的。原因是 PostgreSQL 把拥有者的权限视为拥有者授予给自己的；因此他们也可以撤销这些权限。在 SQL 标准中，拥有者的权限由一个假定实体 “_SYSTEM” 授予。由于拥有者并不是 “_SYSTEM”，因此不能撤销这些权利。

SQL 标准允许为表中的单个列设置权限：

```
GRANT privileges
    ON table [ ( column [, ...] ) ] [, ...]
    TO { PUBLIC | username [, ...] } [ WITH GRANT OPTION ]
```

SQL 标准还为其他种类的对象提供 `USAGE` 权限：字符集、排序规则、翻译、域。

`RULE`权限，以及数据库、表空间、模式、语言和序列上的权限都是 PostgreSQL扩展。

参见

REVOKE

INSERT

INSERT — 在表中插入新行

大纲

```
INSERT INTO table [ ( column [, ...] ) ]
    { DEFAULT VALUES | VALUES ( { expression | DEFAULT } [, ...] ) |
 query }
```

描述

INSERT将新行插入表中。可以插入由值表达式指定的单个行，也可以插入作为查询结果的多行。

目标列名可以按任意顺序列出。如果根本未给出列名列表，
则默认使用按声明顺序排列的全部列；或者如果VALUES子句或 *query*只提供了*N*列，则默认使用按声明顺序排列的前*N*个列名。VALUES子句或*query* 提供的值，会按从左到右的顺序与显式或隐式列列表对应起来。

任何未出现在显式或隐式列列表中的列都会填入默认值；如果没有声明默认值，则填入空值。

如果任何列的表达式数据类型不正确，将尝试自动进行类型转换。

要向表中插入行，必须具有该表上的INSERT权限。如果使用*query*子句从查询中插入行，
还需要在查询使用的任何表上具有SELECT权限。

参数

table

现有表的名称（可选地使用模式限定）。

column

table 中某一列的名称。如有需要，列名可以用子字段名或数组下标限定。（只向组合列的某些字段插入时，其余字段保持为空。）

DEFAULT VALUES

所有列都会填入各自的默认值。

expression

要赋予对应 *column* 的表达式或值。

DEFAULT

对应的 *column* 将用它的默认值填充。

query

提供要插入行的查询（SELECT语句）。其语法说明请参见SELECT命令。

输出

成功完成时，`INSERT` 命令返回以下形式的命令标签：

```
INSERT oid count
```

其中，*count* 是插入的行数。如果 *count* 恰好为一，且目标表有 OID，则 *oid* 表示分配给插入行的 OID。该单行必须是插入的，而不是更新的。否则 *oid* 为零。

示例

向表`films`插入一行：

```
INSERT INTO films VALUES
    ('UA502', 'Bananas', 105, '1971-07-13', 'Comedy', '82 minutes');
```

在这个示例中，省略了`len`列，因此它将取默认值：

```
INSERT INTO films (code, title, did, date_prod, kind)
    VALUES ('T_601', 'Yojimbo', 106, '1961-06-16', 'Drama');
```

这个示例对日期列使用`DEFAULT`子句，而不是显式指定值：

```
INSERT INTO films VALUES
    ('UA502', 'Bananas', 105, DEFAULT, 'Comedy', '82 minutes');
INSERT INTO films (code, title, did, date_prod, kind)
    VALUES ('T_601', 'Yojimbo', 106, DEFAULT, 'Drama');
```

要插入一行，其所有列都由默认值构成：

```
INSERT INTO films DEFAULT VALUES;
```

这个示例从与`films`列布局相同的 `tmp_films`表中选出一些行并插入到 `films`表：

```
INSERT INTO films SELECT * FROM tmp_films WHERE date_prod < '2004-05
-07';
```

这个示例向数组列插入值：

```
-- Create an empty 3x3 gameboard for noughts-and-crosses
-- (these commands create the same board)
INSERT INTO tictactoe (game, board[1:3][1:3])
    VALUES (1,'{{"","",""},{"","",""},{"","",""}}');
INSERT INTO tictactoe (game, board)
    VALUES (2,'{{,,},{,,},{,,}}');
```

兼容性

`INSERT`符合 SQL 标准。标准不允许省略列名列表、而又不是所有列都由`VALUES`子句或*query*填充的情况。

*query*子句可能存在的限制见SELECT。

LISTEN

LISTEN — 监听通知

大纲

```
LISTEN name
```

描述

LISTEN将当前会话注册为名为*name*的通知条件的监听者。
如果当前会话已经在该通知条件上注册为监听者，则不会执行任何操作。

每当命令NOTIFY *name*被执行时，无论是由当前会话还是由另一个连接到同一数据库的会话执行，所有当前正在监听该通知条件的会话都会收到通知，
而每个会话随后都会通知其所连接的客户端应用。有关更多信息，请参见对NOTIFY的讨论。

会话可以使用UNLISTEN命令取消在给定通知条件上的监听注册。会话结束时，
其监听注册会被自动清除。

客户端应用必须采用何种方法来检测通知事件，取决于它所使用的 PostgreSQL应用程序编程接口。使用 libpq库时，应用程序将LISTEN 作为普通 SQL 命令发出，然后必须定期调用 PQnotifies函数，以判断是否已经收到任何通知事件。其他接口，例如libpgtcl，则提供了处理通知事件的更高层方法；事实上，在libpgtcl中，应用程序员甚至不应直接发出LISTEN
或UNLISTEN。详见所使用接口的相关文档。

NOTIFY中对LISTEN和NOTIFY的使用有更详细的讨论。

参数

name

通知条件的名称（任意标识符）。

示例

在psql中配置并执行一组 listen/notify 操作：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
```

兼容性

SQL 标准中没有LISTEN语句。

另见

NOTIFY, UNLISTEN

LOAD

LOAD — 装载或重新装载一个共享库文件

大纲

```
LOAD 'filename'
```

描述

该命令将一个共享库文件装载到PostgreSQL服务器的地址空间中。如果该文件先前已被装载，则会先卸载它。
这条命令主要用来卸载并重新装载自服务器首次装载之后发生了变化的共享库文件。
要使用该共享库，需要用CREATE FUNCTION命令声明其中的函数。

文件名的指定方式与CREATE FUNCTION中共享库名的指定方式相同；特别地，可以依赖搜索路径以及系统标准共享库文件名扩展名的自动添加。关于此主题的更多信息，见第 32.9 节。

兼容性

`LOAD`是PostgreSQL的一种扩展。

另见

CREATE FUNCTION

LOCK

LOCK — 锁定表

大纲

```
LOCK [ TABLE ] name [, ...] [ IN lockmode MODE ] [ NOWAIT ]

where lockmode is one of:

    ACCESS SHARE | ROW SHARE | ROW EXCLUSIVE | SHARE UPDATE EXCLUSIVE
    | SHARE | SHARE ROW EXCLUSIVE | EXCLUSIVE | ACCESS EXCLUSIVE
```

描述

`LOCK TABLE`获取一个表级锁；如有必要，会等待任何冲突锁被释放。如果指定了`NOWAIT`，`LOCK TABLE`就不会等待获取所需的锁：如果无法立即获得，命令将被中止并报错。锁一旦获得，就会一直持有到当前事务结束。（没有`UNLOCK TABLE`命令；锁总是在事务结束时释放。）

为引用表的命令自动获取锁时，PostgreSQL始终使用限制最少的可用锁模式。`LOCK TABLE`用于可能需要更严格锁定的情况。例如，假设一个应用在`READ COMMITTED`隔离级别运行事务，并且需要确保表中的数据在事务期间保持稳定。为此，可以在查询之前对表获取`SHARE`锁。
这会阻止并发数据更改，确保后续对表的读取能看到已提交数据的稳定视图，因为`SHARE`锁模式与写入者获取的`ROW EXCLUSIVE`锁冲突，而你的`LOCK TABLE name IN SHARE MODE`语句会等待，直到所有并发持有ROW EXCLUSIVE模式锁的事务提交或回滚。因此，一旦获得该锁，就不存在尚未提交的写入；而且在释放该锁之前，也不会开始任何写入。

要在 Serializable（可串行化）隔离级别下运行事务时获得类似的效果，你必须先执行`LOCK TABLE`语句，然后才能执行任何`SELECT`或数据修改语句。
一个可串行化事务的数据视图将在它的第一条`SELECT`或数据修改语句开始时被冻结。
事务中较晚执行的`LOCK TABLE`仍然可以阻止并发写–但它无法确保该事务读取到的是最新已提交的值。

如果这类事务还要修改表中的数据，那么它应使用`SHARE ROW EXCLUSIVE`锁模式，而不是`SHARE`模式。这样可以确保同一时间只有一个这类事务在运行。
否则就可能发生死锁：两个事务都可能先获得`SHARE`模式，
然后都无法再获得实际执行更新所需的`ROW EXCLUSIVE`模式。（注意，事务自己的锁永远不会互相冲突，因此事务在持有`SHARE`模式时仍可获得 `ROW EXCLUSIVE`模式，但前提是没有其他人持有`SHARE`模式。）为避免死锁，要确保所有事务都按相同顺序对相同对象获取锁；
如果同一对象需要多种锁模式，则事务应始终先获取限制最严格的模式。

关于锁模式和锁策略的更多信息，请参见第 12.3 节。

参数

name

要锁定的现有表（可选模式限定）的名称。

命令`LOCK TABLE a, b;`等效于 `LOCK TABLE a; LOCK TABLE b;`。这些表会按 `LOCK TABLE`命令中指定的顺序逐个锁定。

lockmode

锁模式指定该锁会与哪些锁冲突。锁模式见第 12.3 节。

如果未指定锁模式，则使用限制最严格的ACCESS EXCLUSIVE模式。

NOWAIT

指定LOCK TABLE不等待任何冲突锁被释放：如果指定的锁无法在不等待的情况下立即获得，事务就会中止。

注解

LOCK TABLE ... IN ACCESS SHARE MODE要求对目标表具有SELECT权限。所有其他形式的LOCK 都要求具有表级UPDATE和/或DELETE权限。

LOCK TABLE只有在事务块（BEGIN/COMMIT对）内才有用，因为锁在事务结束时就会被释放。出现在任何事务块之外的 LOCK TABLE命令会构成一个自包含的事务，因此锁在获取后立即就会被释放。

LOCK TABLE只处理表级锁，因此名称中带有ROW的模式其实都不准确。这些模式名称通常应理解为：用户打算在被锁定的表中获取行级锁。此外，ROW EXCLUSIVE模式本身也是一种可共享的表锁。请记住，就LOCK TABLE而言，所有锁模式的语义完全相同，差别只在于哪些模式彼此冲突。关于如何获取真正的行级锁，请参阅第 12.3.2 节以及FOR UPDATE/FOR SHARE 子句一节（位于SELECT参考文档中）。

示例

在准备向外键表执行插入时，在主键表上获取一个SHARE锁：

```
BEGIN WORK;
LOCK TABLE films IN SHARE MODE;
SELECT id FROM films
    WHERE name = 'Star Wars: Episode I - The Phantom Menace';
-- 如果未返回记录则执行 ROLLBACK
INSERT INTO films_user_comments VALUES
    (_id_, 'GREAT! I was waiting for it for so long!');
COMMIT WORK;
```

在准备执行删除操作时，在主键表上获取一个SHARE ROW EXCLUSIVE锁：

```
BEGIN WORK;
LOCK TABLE films IN SHARE ROW EXCLUSIVE MODE;
DELETE FROM films_user_comments WHERE id IN
    (SELECT id FROM films WHERE rating < 5);
DELETE FROM films WHERE rating < 5;
COMMIT WORK;
```

兼容性

SQL 标准中没有LOCK TABLE，而是使用SET TRANSACTION 来指定事务的并发级别。PostgreSQL也支持这一点；详见SET TRANSACTION。

除ACCESS SHARE、ACCESS EXCLUSIVE和 SHARE UPDATE EXCLUSIVE锁模式外，PostgreSQL的锁模式和LOCK TABLE语法与Oracle中的对应语法兼容。

MOVE

MOVE — 定位游标

大纲

```
MOVE [ direction { FROM | IN } ] cursorname
```

描述

MOVE在不检索任何数据的情况下重新定位游标。MOVE与FETCH命令的工作方式完全相同，只是它只定位游标而不返回行。

有关语法和用法的细节，请参阅FETCH。

输出

成功完成时，MOVE命令会返回如下形式的命令标签：

```
MOVE count
```

其中*count*是用相同参数执行 FETCH命令时本应返回的行数（可能为零）。

示例

```
BEGIN WORK;
DECLARE liahona CURSOR FOR SELECT * FROM films;

-- Skip the first 5 rows:
MOVE FORWARD 5 IN liahona;
MOVE 5

-- Fetch the 6th row from the cursor liahona:
FETCH 1 FROM liahona;
 code  | title  | did | date_prod  |  kind  |  len
-------+--------+-----+------------+--------+-------
 P_303 | 48 Hrs | 103 | 1982-10-22 | Action | 01:37
(1 row)

-- Close the cursor liahona and end the transaction:
CLOSE liahona;
COMMIT WORK;
```

兼容性

SQL 标准中没有MOVE语句。

另见

CLOSE, DECLARE, FETCH

NOTIFY

NOTIFY — 发出一个通知

大纲

NOTIFY *name*

描述

NOTIFY 命令向当前数据库中曾为指定通知名执行过 LISTEN *name* 的每个客户端应用发送一个通知事件。

NOTIFY为访问同一个PostgreSQL数据库的一组进程提供了一种简单的进程间通信机制。如果需要传递结构化数据，也可以借助数据库中的表，把附加数据（而不只是通知名）从通知发送者传递给一个或多个监听者，从而构建更高层的机制。

传递给客户端的通知事件信息包括：通知名以及发出通知的会话对应的服务器进程PID。在某个数据库中使用哪些通知名以及各自含义，由数据库设计者自行决定。

常见的做法是让通知名与数据库中的某个表同名，而通知事件基本上意味着：“我改动了这张表，去看看有什么新变化”。不过，NOTIFY和LISTEN命令并不会强制这种关联。例如，数据库设计者可以使用多个不同的通道名，来标识同一张表上的不同类型变更；

当NOTIFY用于表明某张特定表发生变化时，一种很有用的编程技巧是把NOTIFY放进由表更新触发的规则中。这样一来，每当表发生变化时就会自动发出通知，应用程序员也不容易忘记这么做。

NOTIFY会以几种重要方式与 SQL 事务交互。首先，如果在事务内部执行NOTIFY，那么只有在事务提交之后，通知事件才会被递送；如果事务被中止，则其中所有命令都不会生效，NOTIFY也不例外。这种行为是合理的，但如果你期望通知立即送达，可能会感到意外。其次，如果某个正在监听的会话在事务内部收到了通知信号，那么在该事务结束（提交或中止）之前，通知事件都不会递送给它所连接的客户端。原因同样在于：如果通知在事务内部就已递送，而该事务后来又被中止，我们会希望通知也能被撤销，但服务器一旦把通知发送给客户端，就无法再把它“收回”。因此，通知事件只会在事务之间递送。由此得出的结论是，使用NOTIFY做实时信号的应用，应尽量让事务保持短小。

NOTIFY在一个重要方面的行为类似 Unix 信号：如果在短时间内多次发出同一通知名的信号，接收者可能对多次执行的NOTIFY只收到一个通知事件。因此，依赖收到通知的数量并不是一个好主意。正确的做法是用NOTIFY唤醒需要关注某事的应用，而用一个数据库对象（例如序列）来跟踪发生了什么以及发生了多少次。

执行NOTIFY的客户端自己同时也在监听同一通知条件，这是很常见的情形。在这种情况下，它会像其他监听会话一样收到一个通知事件。根据应用逻辑，这可能导致无用功，例如再次读取自己刚刚写入更新的数据库表。要避免这种额外工作，可以检查通知事件消息中提供的发出通知的服务器进程PID，是否与当前会话自身的PID（可通过libpq获得）相同。如果二者相同，就说明这是当前会话自己发出的通知回送给自己，可以直接忽略。（尽管前一段有那样的说法，但这是一种安全的技术：PostgreSQL会把自身通知与来自其他会话的通知分开保存，因此忽略自己的通知并不会让你错过来自外部的通知。）

参数

name

要发信号的通知名称（任意标识符）。

示例

在psql中配置并执行一组 listen/notify 操作：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
```

兼容性

SQL 标准中没有`NOTIFY`语句。

另见

LISTEN, UNLISTEN

PREPARE

PREPARE — 为执行准备一个语句

大纲

```
PREPARE plan_name [ (datatype [, ...] ) ] AS statement
```

描述

PREPARE 创建一个预备语句。预备语句是一种可用于优化性能的服务器端对象。执行 PREPARE 语句时，指定的语句会被解析、重写和规划。随后对预备语句发出 EXECUTE 命令时，只需执行该预备语句即可。这样，解析、重写和规划阶段只执行一次，而不是每次执行语句时都进行。

预备语句可以带参数，也就是在执行语句时被代入语句中的值。要在预备语句中使用参数，需在PREPARE语句中提供一个数据类型列表，并在要预备的语句本身中，使用$1、$2等按位置引用这些参数。执行语句时，在 EXECUTE 语句中为这些参数指定实际值。有关详情，参见EXECUTE。

预备语句只在当前数据库会话期间存在。会话结束时，预备语句就会被遗忘，因此再次使用前必须重新创建。这也意味着单个预备语句不能由多个并发的数据库客户端共用；不过，每个客户端都可以创建自己要使用的预备语句。预备语句也可以用DEALLOCATE命令手工释放。

当单个会话要执行大量相似语句时，预备语句能带来最大的性能优势。如果语句在规划或重写时比较复杂，例如查询涉及许多表的连接，或者需要应用多条规则，那么这种性能差异会特别明显。如果语句相对容易规划和重写，但执行本身代价较高，则预备语句带来的性能优势就不那么明显。

参数

plan_name

为这个特定预备语句指定的任意名称。它在单个会话内必须唯一，随后用于执行或释放先前已准备的语句。

datatype

预备语句中某个参数的数据类型。要在预备语句本身中引用参数，可使用 $1、$2 等。

statement

任何SELECT、INSERT、UPDATE 或DELETE语句。

注解

在某些情况下，为预备语句生成的查询计划可能不如语句被正常提交并执行时所选出的查询计划。这是因为在对语句做规划、规划器尝试确定最优查询计划时，语句中指定的任何参数的实际值都不可用。PostgreSQL收集表中数据分布的统计信息，并且可以利用语句中的常量值来推测执行该语句的可能结果。

由于在规划带参数的预备语句时这些数据不可用，所选出的计划可能是次优的。要检查PostgreSQL为预备语句选择的查询计划，可以使用EXPLAIN。

有关查询规划以及 PostgreSQL 为此收集的统计信息的更多内容，请参见ANALYZE文档。

示例

为一个INSERT 语句创建预备语句，然后执行它：

```
PREPARE fooplan (int, text, bool, numeric) AS
    INSERT INTO foo VALUES($1, $2, $3, $4);
EXECUTE fooplan(1, 'Hunter Valley', 't', 200.00);
```

为一个 SELECT 语句创建预备语句，然后执行它：

```
PREPARE usrrptplan (int, date) AS
    SELECT * FROM users u, logs l WHERE u.usrid=$1 AND u.usrid=l.usrid
    AND l.date = $2;
EXECUTE usrrptplan(1, current_date);
```

兼容性

SQL 标准包含PREPARE语句，但它只用于嵌入式 SQL。这里的PREPARE语句在语法上也略有不同。

另见

DEALLOCATE, EXECUTE

PREPARE TRANSACTION

PREPARE TRANSACTION — 为两阶段提交准备当前事务

大纲

PREPARE TRANSACTION *transaction_id*

描述

PREPARE TRANSACTION为两阶段提交准备当前事务。执行此命令后，该事务将不再与当前会话相关联；相反，它的状态会被完整地存储到磁盘上，因此即使在请求提交之前数据库发生崩溃，它也极有可能成功提交。

事务进入预备状态后，稍后可以分别使用COMMIT PREPARED或ROLLBACK PREPARED来提交或回滚。发出这些命令的不必是执行原始事务的那个会话，任何会话都可以这样做。

从发出该命令的会话来看，PREPARE TRANSACTION颇似ROLLBACK：执行之后，将不再有活动的当前事务，并且该预备事务的效果也不再可见。（如果该事务随后被提交，这些效果会再次可见。）

如果PREPARE TRANSACTION因任何原因失败，它就等同于执行了一次ROLLBACK：当前事务会被取消。

参数

transaction_id

一个任意标识符，后续可通过它在COMMIT PREPARED或ROLLBACK PREPARED中标识该事务。该标识符必须写成字符串字面值，长度必须小于 200 字节，并且不能与当前任何已处于预备状态的事务标识符相同。

注解

该命令必须在事务块内部使用。请用BEGIN启动事务块。

目前不允许 PREPARE 一个执行过任何涉及临时表操作、创建过 WITH HOLD 游标，或执行过 LISTEN 或 UNLISTEN 的事务。这些特性与当前会话绑定得太紧密，在一个要预备的事务中没有用处。

如果该事务曾用SET修改过任何运行时参数，那么这些效果在执行PREPARE TRANSACTION之后仍会保留，并且不会受到后续任何COMMIT PREPARED或ROLLBACK PREPARED的影响。因此，仅就这一点而言，PREPARE TRANSACTION的行为更像COMMIT而不是ROLLBACK。

当前所有处于预备状态的事务都列在 pg_prepared_xacts系统视图中。

从性能角度看，让事务长时间停留在预备状态是不明智的：例如这会妨碍 VACUUM回收存储空间的能力。还要记住，该事务会继续持有它原本持有的锁。
此功能的预期用法是：一旦外部事务管理器确认其他数据库也已准备好提交，预备事务通常就会被提交或回滚。

如果你会大量使用预备事务，可能需要增大max_prepared_transactions的值，因为默认设置相当小（以避免为不使用该功能的用户浪费资源）。建议将其至少设置为与max_connections相等，这样每个会话都可以有一个待处理的预备事务。

示例

为两阶段提交准备当前事务，并使用`foobar`作为事务标识符：

```
PREPARE TRANSACTION 'foobar';
```

另见

COMMIT PREPARED, ROLLBACK PREPARED

REINDEX

REINDEX — 重建索引

大纲

```
REINDEX { INDEX | TABLE | DATABASE | SYSTEM } name [ FORCE ]
```

描述

REINDEX使用索引所属表中存储的数据重建索引，并替换索引的旧副本。使用REINDEX主要有两个原因：

- 一个索引已经损坏，不再包含有效数据。尽管理论上这不该发生，但在实践中索引可能因软件缺陷或硬件故障而损坏。REINDEX提供了一种恢复方法。
- 该索引包含大量未被回收的死索引页。在 PostgreSQL 中，B-树索引在某些访问模式下可能出现这种情况。REINDEX提供了一种通过写入一个不含死页的新版本索引来减少索引空间消耗的方法。详见第 22.2 节。

参数

INDEX

重新创建指定的索引。

TABLE

重新创建指定表的所有索引。如果该表有一个辅助 “TOAST” 表，也会对其重新索引。

DATABASE

重新创建当前数据库中的所有索引。共享系统目录上的索引会被跳过（独立模式除外，见下文）。

SYSTEM

重新创建当前数据库内系统目录上的所有索引。用户表上的索引不会被处理。另外，共享系统目录上的索引会被跳过（独立模式除外，见下文）。

name

要重新索引的特定索引、表或数据库的名称。索引名和表名可以带模式限定。目前，REINDEX DATABASE和 REINDEX SYSTEM只能对当前数据库重新索引，所以其参数必须与当前数据库名匹配。

FORCE

在每个索引被重建时打印进度报告。

注解

如果怀疑某个用户表上的索引已经损坏，可以使用 REINDEX INDEX或REINDEX TABLE 直接重建该索引，或者重建该表上的所有索引。

如果需要从系统表上的索引损坏中恢复，情况就更复杂了。在这种情况下，重要的是系统本身没有使用任何可疑索引。（事实上，在这种场景下，你可能会发现服务器进程在启动时立即崩溃，因为它依赖损坏的索引。）要安全恢复，必须用-P选项启动服务器，该选项会阻止服务器在查找系统目录时使用索引。

一种做法是关闭服务器，并在命令行中包含-P选项来启动单用户 PostgreSQL 服务器。然后可以根据希望重建的范围，执行REINDEX DATABASE、REINDEX SYSTEM、REINDEX TABLE 或REINDEX INDEX。如果拿不准，就使用REINDEX SYSTEM来重建该数据库中的所有系统索引。然后退出单用户服务器会话并重新启动常规服务器。关于如何与单用户服务器接口交互的更多信息，参见postgres参考页。

另一种方法是启动一个常规服务器会话，并在其命令行选项中包含 -P。具体做法因客户端而异，但对于所有基于 libpq的客户端，都可以在启动客户端之前将环境变量PGOPTIONS设置为-P。注意，尽管这种方法不需要阻止其他客户端，但在修复完成之前，阻止其他用户连接到受损数据库可能仍然更稳妥。

如果怀疑任何共享系统目录（即pg_authid、pg_auth_members、pg_database、pg_pltemplate、pg_shdepend和pg_tablespace）上的索引已经损坏，则必须使用独立服务器来修复。REINDEX在多用户模式下不会处理共享目录。

对于除共享系统目录之外的所有索引，REINDEX都是崩溃安全且事务安全的。REINDEX对共享索引并不崩溃安全，这正是正常操作期间不允许这种情况的原因。如果在独立模式下重建这类目录之一时发生故障，在问题得到纠正之前，将无法重新启动常规服务器。（部分重建的共享索引的典型症状是“index is not a btree”错误。）

REINDEX类似于删除并重新创建索引，因为索引内容都是从头重建的。不过，两者在锁方面的考量相当不同。REINDEX会阻止该索引所属表上的写入，但不阻止读取。它还会对正在处理的特定索引获取独占锁，从而阻塞试图使用该索引的读取。相比之下，DROP INDEX会短暂地对父表获取独占锁，同时阻塞写入和读取。随后的 CREATE INDEX会阻止写入但不阻止读取；由于索引不存在，读取不会尝试使用它，因此不会发生阻塞，但读取可能被迫使用代价高昂的顺序扫描。另一个要点是，删除/重建方式会使所有使用该索引的缓存查询计划失效，而REINDEX不会。

对单个索引或表重新索引，要求用户是该索引或表的拥有者。对数据库重新索引，则要求用户是该数据库的拥有者（因此，拥有者可以重建其他用户所拥有表的索引）。当然，超级用户始终可以重新索引任何对象。

在PostgreSQL 8.1 之前，REINDEX DATABASE只处理系统索引，而不是像其名字所暗示的那样处理所有索引。这一行为已被更改，以减少出人意料的因素。旧的行为可以通过REINDEX SYSTEM使用。

重建表my_table上的所有索引：REINDEX TABLE my_table;

示例

重建单个索引：

```
REINDEX INDEX my_index;
```

重建表my_table上的所有索引：

```
REINDEX TABLE my_table;
```

在不假定系统索引已经有效的情况下，重建某个数据库中的所有索引：

```
$ export PGOPTIONS="-P"
$ psql broken_db
...
broken_db=> REINDEX DATABASE broken_db;
broken_db=> \q
```

兼容性

在 SQL 标准中没有`REINDEX`命令。

RELEASE SAVEPOINT

RELEASE SAVEPOINT — 销毁一个先前定义的保存点

大纲

```
RELEASE [ SAVEPOINT ] savepoint_name
```

描述

`RELEASE SAVEPOINT` 销毁先前在当前事务中定义的保存点。

销毁保存点后，就不能再将它用作回滚点，但除此之外没有其他用户可见的行为。它不会撤销保存点建立之后执行的命令的效果。（要这样做，请参见ROLLBACK TO SAVEPOINT。）在不再需要保存点时将其销毁，可以让系统在事务结束之前回收一些资源。

`RELEASE SAVEPOINT` 还会销毁在指定保存点建立之后建立的所有保存点。

参数

savepoint_name

要销毁的保存点名称。

注解

指定一个此前未定义的保存点名称是一种错误。

当事务处于已中止状态时，不能释放保存点。

如果有多个保存点使用相同名称，则只会释放最近定义的那个。

示例

要建立保存点，并在稍后销毁它：

```
BEGIN;
    INSERT INTO table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

上述事务将同时插入 3 和 4。

兼容性

该命令符合SQL标准。标准规定关键字`SAVEPOINT`是强制的，但PostgreSQL允许省略它。

另见

BEGIN, COMMIT, ROLLBACK, ROLLBACK TO SAVEPOINT, SAVEPOINT

RESET

RESET — 将一个运行时参数的值恢复为默认值

大纲

```
RESET name
RESET ALL
```

描述

RESET将运行时参数恢复为其默认值。RESET是

```
SET parameter TO DEFAULT
```

的另一种拼写。详见SET。

默认值指的是：在当前会话中，如果从未对该参数执行过SET，它本应具有的值。
该值的实际来源可能是编译时内置的默认值、配置文件、命令行选项，
或者针对特定数据库或特定用户的默认设置。详见第 17 章。

关于RESET的事务行为的细节，请参见 SET参考页。

参数

name

运行时参数的名称。列表请参见SET。

ALL

将所有可设置的运行时参数重置为默认值。

示例

将timezone配置参数设置为其默认值：

```
RESET timezone;
```

兼容性

RESET是一种 PostgreSQL扩展。

REVOKE

REVOKE — 撤销访问权限

大纲

```
REVOKE [ GRANT OPTION FOR ]
    { { SELECT | INSERT | UPDATE | DELETE | RULE | REFERENCES |
 TRIGGER }
    [,...] | ALL [ PRIVILEGES ] }
    ON [ TABLE ] tablename [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { CREATE | TEMPORARY | TEMP } [,...] | ALL [ PRIVILEGES ] }
    ON DATABASE dbname [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { EXECUTE | ALL [ PRIVILEGES ] }
    ON FUNCTION funcname ( [ [ argmode ] [ argname ] argtype [, ...] ]
 ) [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { USAGE | ALL [ PRIVILEGES ] }
    ON LANGUAGE langname [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { { CREATE | USAGE } [,...] | ALL [ PRIVILEGES ] }
    ON SCHEMA schemaname [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ GRANT OPTION FOR ]
    { CREATE | ALL [ PRIVILEGES ] }
    ON TABLESPACE tablespacename [, ...]
    FROM { username | GROUP groupname | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ]

REVOKE [ ADMIN OPTION FOR ]
    role [, ...] FROM username [, ...]
    [ CASCADE | RESTRICT ]
```

描述

REVOKE 命令从一个或多个角色那里撤销先前授予的权限。关键字 PUBLIC 指的是由所有角色组成的隐式定义组。

有关各种权限类型的含义，请参见GRANT命令的描述。

请注意，任何特定角色实际拥有的权限，是直接授予给它的权限、授予给它当前所属任一角色的权限，以及授予给 PUBLIC 的权限之和。因此，例如，从 PUBLIC 撤销 SELECT 权限，并不一定意味着所有角色都失去了该对象上的 SELECT 权限：那些被直接授予该权限或通过其他角色获得该权限的角色仍然拥有它。

如果指定了 GRANT OPTION FOR，则只撤销该权限的授予选项，而不撤销权限本身。否则，权限和授予选项都会被撤销。

如果某个用户持有带授予选项的权限，并且已经将其授予其他用户，那么其他用户持有的这些权限称为依赖权限。若正在撤销第一个用户持有的该权限或其授予选项，并且存在依赖权限，则在指定 CASCADE 时这些依赖权限也会被一并撤销；否则，该撤销操作会失败。此递归撤销只影响那些通过某条可追溯到本 REVOKE 命令目标用户的用户链授予的权限。因此，如果受影响用户还通过其他用户获得了同一权限，那么他们实际上可能仍保有该权限。

在撤销角色成员资格时，GRANT OPTION改称为ADMIN OPTION，但行为类似。另请注意，这种形式的命令不接受噪声词GROUP。

注解

使用psql的 \z 命令可以显示现有对象上授予的权限。有关格式的信息，参见GRANT。

用户只能撤销由自己直接授予的权限。例如，如果用户 A 已将某项带授予选项的权限授予用户 B，而用户 B 又将其授予用户 C，则用户 A 不能直接从 C 撤销该权限。相反，用户 A 可以从 B 撤销授予选项，并使用 CASCADE 选项，这样该权限就会进一步从 C 撤销。再例如，如果 A 和 B 都将同一权限授予了 C，A 可以撤销他自己授予的那份，但不能撤销 B 授予的那份，因此 C 实际上仍会拥有该权限。

当对象的非拥有者尝试对该对象执行 REVOKE 时，如果该用户在该对象上完全没有任何权限，命令将立即失败。只要该用户至少拥有某项权限，命令就会继续执行，但只会撤销那些该用户持有授予选项的权限。如果未持有任何授予选项，REVOKE ALL PRIVILEGES 形式会发出警告；而其他形式如果命令中特别列出的任一权限未持有其授予选项，也会发出警告。（原则上，这些说明也适用于对象拥有者；但由于拥有者始终被视为持有全部授予选项，这种情况实际上不会发生。）

如果超级用户选择执行 GRANT 或 REVOKE 命令，则该命令会像由受影响对象的拥有者发出那样执行。由于所有权限最终都来自对象拥有者（可能经由授予选项链间接传递），超级用户可以撤销所有权限，但如上所述，这可能需要使用 CASCADE。

REVOKE 也可以由并非受影响对象拥有者的角色执行，只要该角色是拥有该对象之角色的成员，或者是持有该对象上 WITH GRANT OPTION 权限之角色的成员。在这种情况下，该命令会视同由实际拥有该对象的角色，或持有该对象上 WITH GRANT OPTION 权限的角色发出。例如，如果表 t1 由角色 g1 拥有，而角色 u1 是其成员，那么 u1 可以撤销 t1 上那些记录为由 g1 授出的权限。这包括由 u1 以及角色 g1 的其他成员作出的授权。

如果执行 REVOKE 的角色通过多条角色成员资格路径间接持有权限，则系统不会指明将使用哪一个上层角色来执行该命令。在这种情况下，最佳做法是使用 SET ROLE 切换成你希望作为其身份

执行 REVOKE 的那个具体角色。否则，可能会撤销掉并非你本意要撤销的权限，或者根本没有撤销任何权限。

示例

撤销表 films 上授予所有用户的插入权限：

```
REVOKE INSERT ON films FROM PUBLIC;
```

撤销用户 manuel 在视图 kinds 上的所有权限：

```
REVOKE ALL PRIVILEGES ON kinds FROM manuel;
```

请注意，这实际上意味着“撤销所有由我授予的权限”。

撤销用户 joe 在角色 admins 中的成员资格：

```
REVOKE admins FROM joe;
```

兼容性

GRANT命令的兼容性注解同样适用于 REVOKE。语法摘要如下：

```
REVOKE [ GRANT OPTION FOR ] privileges
    ON object [ ( column [, ...] ) ]
    FROM { PUBLIC | username [, ...] }
    { RESTRICT | CASCADE }
```

按照标准，关键词 RESTRICT 或 CASCADE 是必需的，但 PostgreSQL 默认假定为 RESTRICT。

另见

GRANT

ROLLBACK

ROLLBACK — 中止当前事务

大纲

```
ROLLBACK [ WORK | TRANSACTION ]
```

描述

`ROLLBACK`回滚当前事务，并丢弃该事务所做的全部更新。

参数

`WORK`
`TRANSACTION`

可选关键字，没有任何作用。

注解

使用COMMIT成功结束事务。

在事务块外执行 `ROLLBACK` 会发出警告，除此之外没有任何效果。

示例

要中止所有更改：

```
ROLLBACK;
```

兼容性

SQL 标准只规定了 `ROLLBACK` 和 `ROLLBACK WORK` 两种形式。除此之外，该命令完全符合标准。

另见

BEGIN, COMMIT, ROLLBACK TO SAVEPOINT

ROLLBACK PREPARED

ROLLBACK PREPARED — 回滚一个先前为两阶段提交而预备的事务

大纲

`ROLLBACK PREPARED` *`transaction_id`*

描述

`ROLLBACK PREPARED`回滚一个处于预备状态的事务。

参数

`transaction_id`

要回滚的事务的事务标识符。

注解

要回滚预备事务，执行者必须是最初执行该事务的同一用户，或者是超级用户；但不必处在执行该事务的同一会话中。

这个命令不能在事务块内执行。该预备事务会被立即回滚。

当前所有处于预备状态的事务都列在 `pg_prepared_xacts` 系统视图中。

示例

回滚事务标识符为`foobar`的事务：

```
ROLLBACK PREPARED 'foobar';
```

另见

PREPARE TRANSACTION, COMMIT PREPARED

ROLLBACK TO SAVEPOINT

ROLLBACK TO SAVEPOINT — 回滚到一个保存点

大纲

```
ROLLBACK [ WORK | TRANSACTION ] TO [ SAVEPOINT ] savepoint_name
```

描述

回滚该保存点建立后执行的所有命令。该保存点仍然有效，如有需要，
之后还可以再次回滚到它。

`ROLLBACK TO SAVEPOINT`会隐式销毁在指定保存点之后建立的所有保存点。

参数

savepoint_name

要回滚到的保存点名称。

注解

使用RELEASE SAVEPOINT可以销毁一个保存点，而不丢弃在它建立之后执行的命令所产生的效果。

指定一个尚未建立的保存点名称是一种错误。

就保存点而言，游标带有一些非事务性的行为。凡是在某个保存点内打开的游标，
在回滚该保存点时都会被关闭。如果先前打开的游标在某个随后又被回滚的保存点内受到了`FETCH`命令的影响，那么该游标会保留在`FETCH`使其指向的位置上（也就是说，`FETCH`
不会被回滚）。关闭游标同样不会因回滚而撤销。如果某个游标的执行导致事务中止，
该游标会进入不可执行状态，因此即使事务可以通过`ROLLBACK TO SAVEPOINT`恢复，该游标也不能再使用。

示例

要撤销在`my_savepoint`建立后执行的命令的效果：

```
ROLLBACK TO SAVEPOINT my_savepoint;
```

游标位置不受保存点回滚的影响：

```
BEGIN;

DECLARE foo CURSOR FOR SELECT 1 UNION SELECT 2;

SAVEPOINT foo;

FETCH 1 FROM foo;
 ?column?
```

```
----------
        1

ROLLBACK TO SAVEPOINT foo;

FETCH 1 FROM foo;
 ?column?
----------
        2

COMMIT;
```

兼容性

SQL标准规定关键字`SAVEPOINT`是强制的，但 PostgreSQL和Oracle允许省略它。SQL 只允许在`ROLLBACK`之后使用`WORK`作为噪声词，不允许使用`TRANSACTION`。此外，SQL 还有一个可选子句 `AND [ NO ] CHAIN`，而PostgreSQL当前尚不支持。除此之外，该命令符合 SQL 标准。

另见

BEGIN, COMMIT, RELEASE SAVEPOINT, ROLLBACK, SAVEPOINT

SAVEPOINT

SAVEPOINT — 在当前事务中定义一个新的保存点

大纲

```
SAVEPOINT savepoint_name
```

描述

`SAVEPOINT`在当前事务中建立一个新的保存点。

保存点是事务中的一种特殊标记，它允许回滚在其建立之后执行的所有命令，
从而将事务状态恢复到建立保存点时的状态。

参数

savepoint_name

赋予新保存点的名称。

注解

使用ROLLBACK TO SAVEPOINT回滚到保存点。使用RELEASE SAVEPOINT销毁保存点，同时保留其建立后执行的命令所产生的效果。

保存点只能在事务块内部建立。一个事务中可以定义多个保存点。

示例

要建立一个保存点，并在稍后撤销其建立后执行的所有命令的效果：

```
BEGIN;
    INSERT INTO table1 VALUES (1);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (2);
    ROLLBACK TO SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (3);
COMMIT;
```

上述事务将插入值 1 和 3，而不会插入 2。

要建立保存点，并在稍后销毁它：

```
BEGIN;
    INSERT INTO table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

上述事务将同时插入 3 和 4。

兼容性

SQL 要求在建立另一个同名保存点时自动销毁原有保存点。在 PostgreSQL中，旧保存点会被保留，不过在回滚或释放时只会使用最近建立的那个。（如果使用释放较新的保存点后，较旧的保存点将再次可供`ROLLBACK TO SAVEPOINT`和 `RELEASE SAVEPOINT`使用。）除此之外，`SAVEPOINT` 完全符合 SQL。

另见

BEGIN, COMMIT, RELEASE SAVEPOINT, ROLLBACK, ROLLBACK TO SAVEPOINT

SELECT

SELECT — 从表或视图中检索行

大纲

```
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
    * | expression [ AS output_name ] [, ...]
    [ FROM from_item [, ...] ]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING condition [, ...] ]
    [ { UNION | INTERSECT | EXCEPT } [ ALL ] select ]
    [ ORDER BY expression [ ASC | DESC | USING operator ] [ NULLS {
 FIRST | LAST } ] [, ...] ]
    [ LIMIT { count | ALL } ]
    [ OFFSET start ]
    [ FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ] ]

where from_item can be one of:

    [ ONLY ] table_name [ * ] [ [ AS ] alias [ ( column_alias [, ...]
 ) ] ]
    ( select ) [ AS ] alias [ ( column_alias [, ...] ) ]
    function_name ( [ argument [, ...] ] ) [ AS ] alias [ ( column_
alias [, ...] | column_definition [, ...] ) ]
    function_name ( [ argument [, ...] ] ) AS ( column_definition [, .
..] )
    from_item [ NATURAL ] join_type from_item [ ON join_condition |
 USING ( join_column [, ...] ) ]
```

描述

SELECT从零个或更多表中检索行。SELECT的总体处理如下：

1. 所有FROM列表中的元素都会被计算。（FROM列表中的每个元素都是一个真实或虚拟表。）如果在FROM列表中指定了多个元素，则它们会被交叉连接在一起。（参见下面的FROM Clause。）

2. 如果指定了WHERE子句，则不满足条件的所有行将从输出中删除。（请参见下面的WHERE Clause。）

3. 如果指定了GROUP BY子句，或者存在聚合函数调用，输出将被组合成在一个或多个值上匹配的行组，并计算聚合函数的结果。如果存在HAVING子句，它将消除不满足给定条件的组。（参见GROUP BY Clause和HAVING Clause。）

4. 实际输出行是使用每个选定行或行组的SELECT输出表达式计算的。（参见下面的SELECT List。）

5. 使用操作符UNION、INTERSECT和EXCEPT，可以把多条SELECT语句的输出合并成一个结果集。UNION操作符返回在一个或两个结果集中的所有行。INTERSECT操作符返回严格同时在两个结果集中的所有行。EXCEPT

操作符返回在第一个结果集中但不在第二个结果集中的行。在所有这三种情况下，除非指定了ALL，否则重复行都会被消除。（参见下面的UNION Clause、INTERSECT Clause和EXCEPT Clause。）

6. 如果指定了ORDER BY子句，则返回的行按指定顺序排序。如果没有给出ORDER BY，则按系统认为最快的顺序返回行。（参见下面的ORDER BY Clause。）

7. DISTINCT消除结果中的重复行。DISTINCT ON会消除在所有指定表达式上匹配的行，只保留每组中的第一行。ALL（默认）将返回所有候选行，包括重复行。（参见下面的DISTINCT Clause。）

8. 如果指定了LIMIT或OFFSET子句，SELECT语句只返回结果行的子集。（参见下面的LIMIT Clause。）

9. 如果指定了 FOR UPDATE 或 FOR SHARE 子句，SELECT 语句将选定的行锁定，防止并发更新。(See FOR UPDATE/FOR SHARE 子句 below.)

你必须在一个表上拥有SELECT权限才能读取它的值。使用FOR UPDATE或FOR SHARE还要求UPDATE权限。

参数

FROM Clause

FROM 子句为 SELECT 指定一个或更多源表。如果指定了多个源表，结果将是所有源表的笛卡尔积（交叉连接）。但通常会增加限定条件，把返回的行限制为该笛卡尔积的一个小子集。

该 FROM 子句可以包含以下元素：

table_name

要扫描的现有表或视图的名称（可选模式限定符）。如果指定了ONLY，则只扫描该表。如果未指定ONLY，则扫描该表及其所有后代表（如果有）。可以在表名后附加*来指示要扫描后代表，但在当前版本中，这是默认行为。（在 7.1 之前的版本中，ONLY是默认行为。）可以通过更改sql_inheritance配置选项来修改默认行为。

alias

包含该别名的FROM项的替代名称。别名可用于简写，或者消除自连接（同一张表被扫描多次）中的歧义。提供别名后，它会完全隐藏表或函数的实际名称；例如给定FROM foo AS f，SELECT的其余部分必须把这个FROM 项写成f而不是foo。如果写了别名，还可以写列别名列表，为该表的一个或多个列提供替代名称。

select

子SELECT可以出现在FROM子句中，它的作用就像在这个SELECT命令的执行期间创建了一个临时表。注意，子SELECT必须用圆括号括起来，并且必须为其提供别名。

function_name

函数调用可以出现在FROM子句中。（这对于返回结果集的函数尤其有用，但任何函数都可以使用。）其效果就像在这条SELECT命令执行期间，将函数的输出创建成了一张临时表。也可以使用别名。如果写了别名，

还可以写一个列别名列表，为函数复合返回类型中的一个或多个属性提供替代名称。
如果函数被定义为返回record数据类型，则必须给出别名或关键字AS，
后面跟一个如下形式的列定义列表：(*column_name data_type* [, ...])。列定义列表必须与函数实际返回的列数和类型相匹配。

join_type

以下之一：

- [INNER] JOIN
- LEFT [OUTER] JOIN
- RIGHT [OUTER] JOIN
- FULL [OUTER] JOIN
- CROSS JOIN

对于INNER和OUTER连接类型，必须指定一个连接条件，即NATURAL、ON *join_condition*或USING (*join_column* [, ...])三者中恰好之一。其含义见下文。对于CROSS JOIN，这些子句都不能出现。

JOIN 子句组合两个 FROM 项。如有必要可用圆括号确定嵌套的顺序。在没有圆括号时，JOIN 从左到右嵌套。在任何情况下，JOIN 的结合都比分隔 FROM 项的逗号更紧。

CROSS JOIN和INNER JOIN产生简单的笛卡尔积，与在FROM顶层列出这两个表得到的结果相同，但会受到连接条件（如果有）的限制。CROSS JOIN等价于INNER JOIN ON (TRUE)，也就是说，没有行会被条件过滤掉。这些连接类型只是提供了一种方便的记法，因为它们所做的一切都可以用普通的FROM和WHERE完成。

LEFT OUTER JOIN返回已限定笛卡尔积中的所有行（即，通过其连接条件的所有组合行），外加左侧表中每一行的一个副本；对于这些行，不存在通过连接条件的右侧行。
这样的左侧行会通过在右侧列中插入空值而扩展到连接表的完整宽度。注意，
在决定哪些行有匹配项时，只考虑JOIN子句自身的条件；外层条件是在之后应用的。

相反，RIGHT OUTER JOIN返回所有连接后的行，再加上每个未匹配右侧行对应的一行（左侧用空值扩展）。这只是一种记法上的便利，因为你可以通过交换左右表将其改写成LEFT OUTER JOIN。

FULL OUTER JOIN返回所有连接后的行，再加上每个未匹配的左侧行（右侧用空值扩展），以及每个未匹配的右侧行（左侧用空值扩展）。

ON *join_condition*

join_condition 是一个表达式，其结果类型为 boolean（类似于 WHERE 子句），用于指定连接中哪些行被视为匹配。

USING (*join_column* [, ...])

一个形如USING (a, b, ...)的子句是ON left_table.a = right_table.a AND left_table.b = right_table.b ...的简写。此外，USING意味着只有每对等价列中的一个会包含在连接输出中，而不是两者都包含。

NATURAL

NATURAL 是一个简写，表示一个列出两个表中所有同名列的 USING 列表。

WHERE Clause

可选的WHERE子句的形式

WHERE *condition*

其中*condition* 是任一计算得到boolean类型结果的表达式。任何不满足这个条件的行都会从输出中被消除。如果用一行的实际值替换其中的变量引用后，该表达式返回真，则该行符合条件。

GROUP BY Clause

可选的 GROUP BY 子句的一般形式为

GROUP BY *expression* [, ...]

GROUP BY会把所有在分组表达式上具有相同值的已选中行压缩成单独一行。*expression*可以是输入列名、输出列（SELECT列表项）的名称或序号或者由输入列值构成的任意表达式。在出现歧义时，GROUP BY名称将被解释为输入列名而不是输出列名。

聚合函数（如果使用了的话）会在组成每个组的所有行上进行计算，为每个组产生一个单独的值（而没有GROUP BY时，聚合会产生一个在所有被选中的行上计算出的单一值）。当存在GROUP BY时，SELECT列表表达式引用未分组的列是无效的，除非是在聚合函数内部，因为对一个未分组的列可能会有多个可能的返回值。

HAVING Clause

可选的HAVING子句的形式

HAVING *condition*

其中*condition*与 WHERE子句中指定的条件相同。

HAVING消除不满足该条件的分组行。HAVING与WHERE不同：WHERE会在应用GROUP BY之前过滤个体行，而HAVING过滤由 GROUP BY创建的分组行。*condition*中引用的每一列都必须无歧义地引用某个分组列，除非该引用出现在聚合函数中，或者该非分组列函数依赖于分组列。

即使没有GROUP BY子句，HAVING 的存在也会把一个查询转变成一个分组查询。这和查询中包含聚合函数但没有 GROUP BY子句时的情况相同。所有被选择的行都被认为是一个单一分组，并且SELECT列表和 HAVING子句只能从聚合函数内部引用表列。如果该 HAVING条件为真，这样的查询将输出单独一行；否则不返回行。

SELECT List

SELECT列表（位于关键词 SELECT和FROM之间）指定构成 SELECT语句输出行的表达式。这些表达式可以（并且通常确实会）引用FROM子句中计算得到的列。使用AS *output_name*子句可以为输出列指定另一个名称。这个名称主要用于给要显示的列做标记。它也可以用来在ORDER BY和GROUP BY子句中引用该列的值，但不能用在WHERE或HAVING子句中；在那里必须写出表达式。

可以在输出列表中写*来取代表达式，它是被选中行的所有列的一种简写方式。还可以写 *table_name*.*，它是只来自那个表的所有列的简写形式。

UNION Clause

UNION 子句的一般形式如下：

select_statement UNION [ALL] *select_statement*

select_statement 是不带 ORDER BY、LIMIT、FOR UPDATE 或 FOR SHARE 子句的任意 SELECT 语句。（如果ORDER BY和LIMIT所附属的子表达式被圆括号括起，则它们可以附加到该子表达式上。不带圆括号时，这些子句会被认为适用于UNION的结果，而不是它的右手输入表达式。）

UNION操作符计算相关 SELECT语句所返回的行的并集。如果一行至少出现在两个结果集中的一个内，它就会在并集中。作为 UNION两个操作数的 SELECT语句必须产生相同数量的列并且对应位置上的列必须具有兼容的数据类型。

UNION的结果不包含任何重复行，除非指定了ALL选项。ALL阻止消除重复行。（因此，UNION ALL通常明显快于UNION；能用ALL时就用它。）

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 UNION操作符会从左至右计算。

当前，FOR UPDATE 和 FOR SHARE 不能为 UNION 的结果或 UNION 的任何输入指定。

INTERSECT Clause

INTERSECT 子句的一般形式如下：

select_statement INTERSECT [ALL] *select_statement*

select_statement 是不带 ORDER BY、LIMIT、FOR UPDATE 或 FOR SHARE 子句的任意 SELECT 语句。

INTERSECT操作符计算相关 SELECT语句返回的行的交集。如果一行同时出现在两个结果集中，它就在交集中。

INTERSECT的结果不包含任何重复行，除非指定了ALL选项。带ALL时，在左表中有m个重复、在右表中有n个重复的行将在结果集中出现 min(m,n) 次。

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 INTERSECT操作符会从左至右计算。INTERSECT的优先级比 UNION更高。也就是说，A UNION B INTERSECT C将被读成A UNION (B INTERSECT C)。

当前，FOR UPDATE 和 FOR SHARE 不能为 INTERSECT 的结果或 INTERSECT 的任何输入指定。

EXCEPT Clause

EXCEPT 子句的一般形式如下：

select_statement EXCEPT [ALL] *select_statement*

select_statement 是不带 ORDER BY、LIMIT、FOR UPDATE 或 FOR SHARE 子句的任意 SELECT 语句。

EXCEPT操作符计算位于左侧 SELECT语句的结果中但不在右侧语句结果中的行集合。

EXCEPT的结果不包含任何重复行，除非指定了ALL选项。带ALL时，在左表中有m个重复、在右表中有n个重复的行将在结果集中出现 max(m-n,0) 次。

除非用圆括号指定计算顺序，同一个SELECT语句中的多个 EXCEPT操作符会从左至右计算。EXCEPT的优先级与 UNION相同。

当前，FOR UPDATE 和 FOR SHARE 不能为 EXCEPT 的结果或 EXCEPT 的任何输入指定。

ORDER BY Clause

可选的ORDER BY子句的形式如下：

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

ORDER BY子句导致结果行被按照指定的表达式排序。如果两行按照最左边的表达式是相等的，则会根据下一个表达式比较它们，依次类推。如果按照所有指定的表达式它们都是相等的，则它们被返回的顺序取决于实现。

每一个*expression* 可以是输出列（SELECT列表项）的名称或者序号，它也可以是由输入列值构成的任意表达式。

序号指的是输出列的顺序（从左至右）位置。
这种特性可以为不具有唯一名称的列定义一个顺序。这不是绝对必要的，因为总是可以使用 AS子句为输出列赋予一个名称。

也可以在ORDER BY子句中使用任意表达式，包括没有出现在SELECT输出列表中的列。因此，下面的语句是合法的：

```
SELECT name FROM distributors ORDER BY code;
```

这种特性的一个限制是一个应用在UNION、INTERSECT或EXCEPT子句结果上的 ORDER BY只能指定输出列名称或序号，但不能指定表达式。

如果一个ORDER BY表达式是一个既匹配输出列名称又匹配输入列名称的简单名称，ORDER BY将把它解读成输出列名称。这与在同样情况下GROUP BY会做出的选择相反。这种不一致是为了与SQL 标准兼容。

可以在ORDER BY子句中任一表达式之后附加关键字 ASC（升序）或DESC（降序）。如果没有指定，ASC被假定为默认值。或者，可以在USING 子句中指定一个特定的排序操作符名称。一个排序操作符必须是某个 B-树操作符族的小于或者大于成员。ASC通常等价于 USING <而DESC通常等价于 USING >（但是一种用户定义数据类型的创建者可以准确地定义默认排序顺序是什么，并且它可能会对应于其他名称的操作符）。

空值的排序高于任何其他值。换句话说，在升序排序时，空值排在最后；在降序排序时，空值排在最前。

字符串数据按照数据库集簇初始化时建立的区域相关排序顺序来排序。

DISTINCT Clause

如果指定了DISTINCT，所有重复的行会被从结果集中移除（为每一组重复的行保留一行）。ALL则指定相反的行为：所有行都会被保留，这也是默认情况。

DISTINCT ON (*expression* [, ...]) 只保留在给定表达式上计算相等的行集合中的第一行。DISTINCT ON表达式使用和 ORDER BY相同的规则（见上文）解释。注意，除非用 ORDER BY来确保所期望的行出现在第一位，每一个集合的“第一行”是不可预测的。例如：

```
SELECT DISTINCT ON (location) location, time, report
    FROM weather_reports
    ORDER BY location, time DESC;
```

为每个地点检索最近的天气报告。但是如果我们不使用 ORDER BY来强制对每个地点的时间值进行降序排序，我们为每个地点得到的报告的时间可能是无法预测的。

DISTINCT ON表达式必须匹配最左边的 ORDER BY表达式。ORDER BY子句通常将包含额外的表达式，这些额外的表达式用于决定在每一个 DISTINCT ON分组内行的优先级。

LIMIT Clause

该 LIMIT 子句由两个独立的子句组成：

```
LIMIT { count | ALL }
OFFSET start
```

count 指定最多返回多少行，而 *start* 指定开始返回行之前要跳过的行数。如果两者都指定，则先跳过 *start* 行，然后开始计数并返回 *count* 行。

在使用LIMIT时，用一个ORDER BY子句把结果行约束到一个唯一顺序是个好办法。否则你将得到该查询结果行的一个不可预测的子集 – 你可能要求从第 10 到第 20 行，但是在什么顺序下的第 10 到第 20 呢？除非指定ORDER BY，你是不知道顺序的。

查询规划器在生成一个查询计划时会考虑LIMIT，因此根据你使用的LIMIT和OFFSET，你很可能得到不同的计划（得到不同的行序）。所以，使用不同的 LIMIT/OFFSET值来选择一个查询结果的不同子集将会给出不一致的结果，除非你用ORDER BY强制一种可预测的结果顺序。这不是一个缺陷，它是 SQL 不承诺以任何特定顺序（除非使用 ORDER BY来约束顺序）给出一个查询结果这一事实造成的必然后果。

FOR UPDATE/FOR SHARE 子句

FOR UPDATE 子句的形式如下：

```
FOR UPDATE [ OF table_name [, ...] ] [ NOWAIT ]
```

与它密切相关的 FOR SHARE 子句的形式如下：

```
FOR SHARE [ OF table_name [, ...] ] [ NOWAIT ]
```

FOR UPDATE使SELECT语句检索到的行像被更新一样被锁定。这防止其他事务在当前事务结束之前修改或删除它们。也就是说，其他事务对这些行尝试UPDATE、DELETE或SELECT FOR UPDATE时将被阻塞，直到当前事务结束。此外，如果来自另一个事务的UPDATE、DELETE或SELECT FOR UPDATE已经锁定了被选中的一行或多行，SELECT FOR UPDATE将等待那个事务完成，然后锁定并返回更新后的行（如果该行已被删除，则不返回行）。
进一步的讨论见第 12 章。

为了防止该操作等待其他事务提交，可使用NOWAIT 选项。SELECT FOR UPDATE NOWAIT会在选中的行无法立即锁定时报错，而不是等待。注意，NOWAIT只适用于行级锁；所需的ROW SHARE表级锁仍会按常规方式取得（见第 12 章）。如果需要不等待就取得表级锁，可以使用LOCK的NOWAIT选项。

FOR SHARE 的行为类似，但它在每个被检索的行上获取共享锁而非排他锁。共享锁阻塞其他事务对这些行执行 UPDATE、DELETE 或 SELECT FOR UPDATE，但不会阻止它们执行 SELECT FOR SHARE。

当前，不允许单条 SELECT 语句同时包含 FOR UPDATE 和 FOR SHARE，也不允许语句的不同部分同时使用 NOWAIT 和普通的等待模式。

如果在 FOR UPDATE 或 FOR SHARE 中提到了特定的表，则只有来自那些表的行会被锁定；SELECT 中用到的任何其他表还是被简单地照常读取。

在返回的行无法清楚地与各个表行对应起来的上下文中，不能使用 FOR UPDATE 和 FOR SHARE；例如它们不能与聚合一起使用。

> **小心**
>
> 避免锁定一行然后又在后面的保存点或 PL/pgSQL 异常块中修改它。后续的回滚会导致锁丢失。例如：
>
> ```
> BEGIN;
> SELECT * FROM mytable WHERE key = 1 FOR UPDATE;
> SAVEPOINT s;
> UPDATE mytable SET ... WHERE key = 1;
> ROLLBACK TO s;
> ```
>
> 在执行 ROLLBACK 之后，该行实际上已被解锁，而不是回到保存点之前已锁定但未修改的状态。如果当前事务中锁定的行被更新或删除，或者共享锁被升级为排他锁，就会出现这种风险：在所有这些情况下，之前的锁状态都会被遗忘。如果事务随后回滚到原始锁定命令和后续更改之间的某个状态，该行将看起来完全没有被锁定。这是一个实现上的缺陷，将在 PostgreSQL 的未来版本中解决。

> **小心**
>
> 一条同时使用LIMIT和FOR UPDATE/SHARE子句的SELECT命令，有可能返回少于LIMIT所指定数量的行。这是因为LIMIT先被应用。该命令先选出指定数量的行，但随后可能会在试图获取其中一行或多行的锁时阻塞。一旦SELECT解除阻塞，该行可能已被删除或更新，不再满足查询的WHERE条件，这种情况下它就不会被返回。

示例

连接表 films 和表 distributors：

```
SELECT f.title, f.did, d.name, f.date_prod, f.kind
    FROM distributors d, films f
    WHERE f.did = d.did

       title       | did |     name     | date_prod  |   kind
-------------------+-----+--------------+------------+----------
 The Third Man     | 101 | British Lion | 1949-12-23 | Drama
 The African Queen | 101 | British Lion | 1951-08-11 | Romantic
 ...
```

要对所有电影的len列求和并且用 kind对结果分组：

```
SELECT kind, sum(len) AS total FROM films GROUP BY kind;

   kind   | total
----------+-------
```

```
 Action   | 07:34
 Comedy   | 02:58
 Drama    | 14:28
 Musical  | 06:42
 Romantic | 04:38
```

要对所有电影的len列求和、对结果按照 kind分组并且显示总长小于 5 小时的分组：

```
SELECT kind, sum(len) AS total
    FROM films
    GROUP BY kind
    HAVING sum(len) < interval '5 hours';

   kind   | total
----------+-------
 Comedy   | 02:58
 Romantic | 04:38
```

下面两个示例都是根据第二列（name）的内容来排序结果：

```
SELECT * FROM distributors ORDER BY name;
SELECT * FROM distributors ORDER BY 2;

 did |       name
-----+------------------
 109 | 20th Century Fox
 110 | Bavaria Atelier
 101 | British Lion
 107 | Columbia
 102 | Jean Luc Godard
 113 | Luso films
 104 | Mosfilm
 103 | Paramount
 106 | Toho
 105 | United Artists
 111 | Walt Disney
 112 | Warner Bros.
 108 | Westward
```

接下来的示例展示了如何得到表distributors和 actors的并集，把结果限制为那些在每个表中以字母 W 开始的行。这里只需要不重复的行，因此省略了关键词 ALL。

```
distributors:               actors:
 did |     name              id |     name
-----+--------------        ----+----------------
 108 | Westward                1 | Woody Allen
 111 | Walt Disney             2 | Warren Beatty
 112 | Warner Bros.            3 | Walter Matthau
 ...                          ...

SELECT distributors.name
    FROM distributors
    WHERE distributors.name LIKE 'W%'
UNION
SELECT actors.name
```

```
    FROM actors
    WHERE actors.name LIKE 'W%';

      name
----------------
 Walt Disney
 Walter Matthau
 Warner Bros.
 Warren Beatty
 Westward
 Woody Allen
```

这个示例展示了如何在FROM子句中使用函数，分别使用和不使用列定义列表：

```
CREATE FUNCTION distributors(int) RETURNS SETOF distributors AS $$
    SELECT * FROM distributors WHERE did = $1;
$$ LANGUAGE SQL;

SELECT * FROM distributors(111);
 did |    name
-----+-------------
 111 | Walt Disney

CREATE FUNCTION distributors_2(int) RETURNS SETOF record AS $$
    SELECT * FROM distributors WHERE did = $1;
$$ LANGUAGE SQL;

SELECT * FROM distributors_2(111) AS (f1 int, f2 text);
 f1  |     f2
-----+-------------
 111 | Walt Disney
```

兼容性

当然，SELECT语句与 SQL 标准兼容。但它也有一些扩展和缺失的特性。

省略 FROM 子句

PostgreSQL 允许省略 FROM 子句。它可以直接用于计算简单表达式的结果：

```
SELECT 2+2;

 ?column?
----------
        4
```

其他一些 SQL 数据库必须引入一个只有一行的虚拟表，才能从该表执行 SELECT。

请注意，如果未指定 FROM 子句，查询就不能引用任何数据库表。例如，以下查询是无效的：

```
SELECT distributors.* WHERE distributors.name = 'Westward';
```

PostgreSQL 在 8.1 之前的版本会接受这种形式的查询，并为查询引用的每个表在该查询的 FROM 子句中添加一个隐式项。这种做法已不再是默认行为，因为它不符合 SQL 标准，而且被许多人认为容易出错。为了兼容依赖这一行为的应用，可以启用add_missing_from配置变量。

AS 关键字

在 SQL 标准中，可选关键词AS只是一个噪音词，可以省略它而不影响含义。PostgreSQL解析器在重命名输出列时要求这个关键词，因为类型可扩展性特性使得没有它就会出现解析歧义。
不过，AS在FROM项中是可选的。

GROUP BY和ORDER BY可用的名字空间

在 SQL-92 标准中，一个ORDER BY子句只能使用输出列名或者序号，而一个GROUP BY子句只能使用基于输入列名的表达式。PostgreSQL扩展了这两种子句以允许它们使用其他的选择（但如果有歧义时还是使用标准的解释）。
PostgreSQL也允许两种子句指定任意表达式。
注意出现在一个表达式中的名称将总是被当做输入列名而不是输出列名。

SQL:1999 及其后的标准使用了一种略微不同的定义，它并不完全向后兼容 SQL-92。不过，在大部分的情况下，PostgreSQL会以与 SQL:1999 相同的方式解释ORDER BY或GROUP BY表达式。

非标准子句

DISTINCT ON、LIMIT和OFFSET子句未在 SQL 标准中定义。

SELECT INTO

SELECT INTO — 根据查询结果定义一个新表

大纲

```
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
    * | expression [ AS output_name ] [, ...]
    INTO [ TEMPORARY | TEMP ] [ TABLE ] new_table
    [ FROM from_item [, ...] ]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING condition [, ...] ]
    [ { UNION | INTERSECT | EXCEPT } [ ALL ] select ]
    [ ORDER BY expression [ ASC | DESC | USING operator ] [, ...] ]
    [ LIMIT { count | ALL } ]
    [ OFFSET start ]
    [ FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ] ]
```

描述

`SELECT INTO`创建一个新表，并用查询计算得到的数据填充该表。与普通的`SELECT`不同，这些数据不会返回给客户端。新表的列具有与`SELECT`输出列对应的名称和数据类型。

参数

`TEMPORARY` or `TEMP`

如果指定，该表将创建为临时表。详见CREATE TABLE。

`new_table`

要创建的表的名称（可选地带模式限定）。

所有其他参数都在SELECT中有详细说明。

注解

CREATE TABLE AS在功能上与 `SELECT INTO`类似。`CREATE TABLE AS` 是推荐使用的语法，因为这种形式的`SELECT INTO`在ECPG 或PL/pgSQL中不可用，因为它们对 `INTO`子句有不同的解释。此外，`CREATE TABLE AS`提供的功能是 `SELECT INTO`所提供功能的超集。

在 PostgreSQL 8.1 之前，`SELECT INTO` 创建的表默认包含 OID。而在 PostgreSQL 8.1 中不再如此——要在新表中包含 OID，必须启用default_with_oids配置变量。或者，也可以使用 `CREATE TABLE AS` 加上 `WITH OIDS` 子句。

示例

创建一个新表`films_recent`，它只包含表 `films`中的最近条目：

```
SELECT * INTO films_recent FROM films WHERE date_prod >= '2002-01-01';
```

兼容性

SQL 标准使用`SELECT INTO`表示将值选入宿主程序的标量变量，而不是创建一个新表。这确实是 ECPG（见第 30 章）和 PL/pgSQL（见第 36 章）中的用法。PostgreSQL使用`SELECT INTO`表示创建表则是历史遗留用法。对于新代码，最好为此目的使用`CREATE TABLE AS`。

另见

CREATE TABLE AS

SET

SET — 更改运行时参数

大纲

```
SET [ SESSION | LOCAL ] name { TO | = } { value | 'value' | DEFAULT }
SET [ SESSION | LOCAL ] TIME ZONE { timezone | LOCAL | DEFAULT }
```

描述

SET命令用于更改运行时配置参数。许多在第 17 章中列出的运行时参数都可以通过 SET即时更改。（有些参数需要超级用户权限才能更改，
还有一些参数在服务器启动或会话开始之后便无法再更改。）SET只影响当前会话所使用的值。

如果在一个随后被中止的事务中发出SET 或SET SESSION，那么在事务回滚时，SET命令的效果就会消失。（这一行为与 PostgreSQL 7.3 之前的版本不同，那时 SET的效果在之后的错误发生后不会回滚。）一旦该事务提交，这些效果就会持续到会话结束，除非被另一个SET覆盖。

SET LOCAL的效果只持续到当前事务结束，无论事务是否提交。一种特殊情况是：在同一个事务中先执行 SET，再执行SET LOCAL。这样SET LOCAL设置的值会一直生效到事务结束，但在此之后（如果该事务被提交），SET设置的值将生效。

参数

SESSION

指定该命令对当前会话生效。（如果既没有出现SESSION，也没有出现LOCAL，
这就是默认情况。）

LOCAL

指定该命令只对当前事务生效。在COMMIT或 ROLLBACK之后，会话级设置会再次生效。注意，如果在 BEGIN 块之外执行，SET LOCAL 看起来不会有任何效果。

name

可设置的运行时参数名称。可用参数见第 17 章以及下文。

value

参数的新值。根据具体参数的不同，值可以指定为字符串常量、标识符、数字，
或由这些构成的逗号分隔列表。可以使用DEFAULT 来指定把该参数重置为其默认值。

除了第 17 章中记载的配置参数外，还有几个参数只能通过SET命令调整，或者具有特殊语法：

NAMES

SET NAMES *value*是 SET client_encoding TO *value*的别名。

SEED

设置随机数生成器（即函数 random）的内部种子。允许的值是 -1 到 1 之间的浮点数，这些值随后会乘以 2^{31}-1。

种子也可以通过调用函数`setseed`来设置：

```
SELECT setseed(value);
```

`TIME ZONE`

`SET TIME ZONE value`是以下写法的别名：`SET timezone TO value`。语法`SET TIME ZONE`允许使用特殊语法指定时区。下面是一些有效值的示例：

`'PST8PDT'`

加利福尼亚州伯克利所使用的时区。

`'Europe/Rome'`

意大利所使用的时区。

`-7`

比 UTC 向西 7 小时的时区（等同于 PDT）。正值表示位于 UTC 以东。

`INTERVAL '-08:00' HOUR TO MINUTE`

比 UTC 向西 8 小时的时区（等同于 PST）。

`LOCAL`
`DEFAULT`

把时区设置为你的本地时区（即服务器操作系统默认的时区）。

有关时区的更多信息，请参见第 8.5 节。此外，附录 B中列出了可识别的时区名称。

注解

函数`set_config`提供了等效功能；见第 9.20 节。

示例

设置模式搜索路径：

```
SET search_path TO my_schema, public;
```

将日期风格设置为传统POSTGRES风格，并采用“日在月之前”的输入惯例：

```
SET datestyle TO postgres, dmy;
```

设置加利福尼亚州伯克利的时区：

```
SET TIME ZONE 'PST8PDT';
```

将时区设置为意大利所使用的时区：

```
SET TIME ZONE 'Europe/Rome';
```

兼容性

`SET TIME ZONE`扩展了 SQL 标准定义的语法。标准只允许使用数值形式的时区偏移量，而 PostgreSQL 允许更灵活的时区指定方式。其他所有`SET`特性都是 PostgreSQL扩展。

另见

RESET, SHOW

SET CONSTRAINTS

SET CONSTRAINTS — 为当前事务设置约束检查模式

大纲

```
SET CONSTRAINTS { ALL | name [, ...] } { DEFERRED | IMMEDIATE }
```

描述

SET CONSTRAINTS设置当前事务中约束检查的行为。IMMEDIATE约束会在每条语句结束时检查。DEFERRED约束要等到事务提交时才检查。每个约束都有各自的IMMEDIATE或DEFERRED模式。

创建约束时，它会被赋予以下三种特性之一：DEFERRABLE INITIALLY DEFERRED、DEFERRABLE INITIALLY IMMEDIATE或者 NOT DEFERRABLE。第三类始终是 IMMEDIATE，不会受到 SET CONSTRAINTS命令的影响。前两类在每个事务开始时都处于其指定的模式，但其行为可以在事务中通过 SET CONSTRAINTS修改。

带约束名称列表的SET CONSTRAINTS只会修改这些约束的模式（它们都必须是可延迟的）。如果任何给定名称匹配多个约束，则所有匹配的约束都会受到影响。SET CONSTRAINTS ALL会修改所有可延迟约束的模式。

当SET CONSTRAINTS将某个约束的模式从 DEFERRED改为IMMEDIATE时，新模式具有追溯效力：任何原本会在事务结束时才检查的未决数据修改，都会改为在执行SET CONSTRAINTS命令期间检查。如果违反了任何此类约束，SET CONSTRAINTS就会失败（并且不会改变该约束的模式）。因此，可以利用SET CONSTRAINTS强制在事务中的特定点执行约束检查。

当前，只有外键约束受到这个设置的影响。检查约束和唯一约束实际上总是不可延迟的。

注解

这个命令只会改变当前事务中约束的行为。因此，如果你在事务块（BEGIN/COMMIT对）之外执行此命令，它将不会产生任何效果。

兼容性

这个命令符合 SQL 标准定义的行为，但有一个限制：在 PostgreSQL中，它只应用于外键约束。

SQL 标准规定SET CONSTRAINTS中出现的约束名可以带模式限定。PostgreSQL尚不支持这一点：这些名称必须是不限定的，而且无论约束位于哪个模式中，所有与命令匹配的约束都会受到影响。

SET ROLE

SET ROLE — 设置当前会话的当前用户标识符

大纲

```
SET [ SESSION | LOCAL ] ROLE rolename
SET [ SESSION | LOCAL ] ROLE NONE
RESET ROLE
```

描述

该命令把当前 SQL 会话的当前用户标识符设置为 *rolename*。角色名可以写成标识符或字符串字面量。在 SET ROLE 之后，SQL 命令的权限检查会视同指定角色是最初登录的角色。

当前会话用户必须是指定的 *rolename* 角色的成员。（如果会话用户是超级用户，则可以选择任意角色。）

SESSION 和 LOCAL 修饰符的作用与常规的SET命令相同。

NONE 和 RESET 形式会把当前用户标识符重置为当前会话用户标识符。这些形式可以由任何用户执行。

注解

使用这个命令，既可以增加权限，也可以限制自己的权限。如果会话用户角色具有 INHERIT 属性，它会自动拥有每个可通过 SET ROLE 切换到的角色的全部权限；在这种情况下，SET ROLE 实际上会去掉直接分配给会话用户及其所属其他角色的全部权限，只保留指定角色可用的权限。另一方面，如果会话用户角色具有 NOINHERIT 属性，SET ROLE 会放弃直接分配给会话用户的权限，转而获得指定角色可用的权限。

特别是，当超级用户选择通过 SET ROLE 切换到非超级用户角色时，会失去其超级用户权限。

SET ROLE的效果与SET SESSION AUTHORIZATION相近，但其中涉及的权限检查完全不同。此外，SET SESSION AUTHORIZATION会决定后续 SET ROLE命令允许使用哪些角色，而通过 SET ROLE更改角色并不会改变后续 SET ROLE允许使用的角色集合。

SET ROLE不能在 SECURITY DEFINER函数中使用。

示例

```
SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 peter        | peter

SET ROLE 'paul';

SELECT SESSION_USER, CURRENT_USER;
```

```
 session_user | current_user
--------------+--------------
 peter        | paul
```

兼容性

PostgreSQL允许使用标识符语法（"rolename"），而 SQL 标准要求将角色名写成字符串字面量。SQL 不允许在事务中执行这个命令；PostgreSQL不施加这一限制，因为没有理由这样做。SESSION和LOCAL修饰符，以及 RESET语法，都是PostgreSQL扩展。

另见

SET SESSION AUTHORIZATION

SET SESSION AUTHORIZATION

SET SESSION AUTHORIZATION — 设置当前会话的会话用户标识符和当前用户标识符

大纲

```
SET [ SESSION | LOCAL ] SESSION AUTHORIZATION username
SET [ SESSION | LOCAL ] SESSION AUTHORIZATION DEFAULT
RESET SESSION AUTHORIZATION
```

描述

这个命令将当前 SQL 会话的会话用户标识符和当前用户标识符设置为 *username*。用户名可以写成一个标识符或者一个字符串字面量。例如，可以使用这个命令临时变为一个非特权用户，之后再切换回超级用户。

会话用户标识符初始时被设置为客户端提供的（可能已认证的）用户名。
当前用户标识符通常与会话用户标识符相同，但可能在 `SECURITY DEFINER`函数和类似机制的上下文中临时更改；它也可以通过SET ROLE更改。当前用户标识符用于权限检查。

只有当初始会话用户（即已认证用户）最初具有超级用户权限时，才能更改会话用户标识符。
否则，只有当该命令指定的是已认证用户名时，才会被接受。

`SESSION` 和 `LOCAL` 修饰符的作用与常规的SET命令相同。

`DEFAULT`和`RESET`形式会将会话用户标识符和当前用户标识符重置为最初已认证的用户名。
这些形式可以由任何用户执行。

注解

`SET SESSION AUTHORIZATION`不能在一个 `SECURITY DEFINER`函数中使用。

示例

```
SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 peter        | peter

SET SESSION AUTHORIZATION 'paul';

SELECT SESSION_USER, CURRENT_USER;

 session_user | current_user
--------------+--------------
 paul         | paul
```

兼容性

SQL 标准允许在字面值*username*的位置上使用某些其他表达式，但这些选项在实践中并不重要。PostgreSQL允许使用标识符语法（`"username"`），而 SQL 标准不允许。SQL 不允许在事

务中使用这个命令；PostgreSQL并不做此限制，因为没有理由这样做。`SESSION`和`LOCAL`修饰符以及`RESET`语法都是PostgreSQL扩展。

标准把执行这个命令所需的权限留给实现定义。

另见

SET ROLE

SET TRANSACTION

SET TRANSACTION — 设置当前事务的特性

大纲

```
SET TRANSACTION transaction_mode [, ...]
SET SESSION CHARACTERISTICS AS TRANSACTION transaction_mode [, ...]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
```

描述

SET TRANSACTION命令设置当前事务的特性。它对任何后续事务都没有影响。SET SESSION CHARACTERISTICS设置一个会话中后续事务的默认事务特性。对于单个事务，这些默认值可以用 SET TRANSACTION覆盖。

可用的事务特性是事务隔离级别和事务访问模式（读/写或只读）。

事务的隔离级别决定当其他事务并发运行时该事务能看到什么数据：

READ COMMITTED

一条语句只能看到在它开始之前已提交的行。这是默认值。

SERIALIZABLE

当前事务中的所有语句都只能看到在该事务中执行第一条查询或数据修改语句之前已提交的行。

SQL 标准还定义了另外两个级别：READ UNCOMMITTED和REPEATABLE READ。在PostgreSQL中，READ UNCOMMITTED被当作READ COMMITTED对待，而REPEATABLE READ被当作SERIALIZABLE对待。

在事务执行第一条查询或数据修改语句（SELECT，INSERT，DELETE，UPDATE，FETCH，或COPY）之后，事务隔离级别就不能再更改。有关事务隔离和并发控制的更多信息，请参见第 12 章。

事务访问模式决定事务是读/写还是只读。读/写是默认值。当事务为只读时，下列 SQL 命令会被禁止：INSERT、UPDATE、DELETE 以及 COPY FROM，前提是它们要写入的表不是临时表；所有 CREATE、ALTER 和 DROP 命令；COMMENT、GRANT、REVOKE、TRUNCATE；以及 EXPLAIN ANALYZE 和 EXECUTE，前提是它们要执行的命令属于上述列表。这是一个较高层面的只读概念，并不会阻止所有写入磁盘的行为。

注解

如果执行SET TRANSACTION之前没有 START TRANSACTION或者 BEGIN，它会发出一条警告，除此之外没有任何效果。

也可以不使用 `SET TRANSACTION`，而是在 `BEGIN` 或 `START TRANSACTION` 中直接指定所需的 *`transaction_modes`*。

会话默认的事务模式也可以通过设置配置参数default_transaction_isolation和default_transaction_read_only来设置。（事实上，`SET SESSION CHARACTERISTICS`只是用`SET`设置这些变量的一种冗长的等价形式。）这意味着这些默认值可以在配置文件中设置，也可以通过`ALTER DATABASE`等方式设置。更多信息请参考第 17 章。

兼容性

两个命令都在SQL标准中定义。标准中默认的事务隔离级别是`SERIALIZABLE`。在PostgreSQL中默认通常是`READ COMMITTED`，但你可以像上面提到的那样更改它。由于缺少谓词锁，`SERIALIZABLE`级别并不是真正可串行化的。详见第 12 章。

在 SQL 标准中，还可以用这些命令设置另一项事务特性：诊断区域的大小。这个概念特定于嵌入式 SQL，因此没有在 PostgreSQL服务器中实现。

SQL 标准要求在连续的*`transaction_modes`*之间有逗号，但出于历史原因，PostgreSQL允许省略逗号。

SHOW

SHOW — 显示一个运行时参数的值

大纲

```
SHOW name
SHOW ALL
```

描述

SHOW将显示运行时参数的当前设置。这些变量可以使用SET语句来设置，也可以通过编辑postgresql.conf配置文件、使用 PGOPTIONS环境变量（在使用 libpq或基于libpq的应用时），或者在启动postmaster时通过命令行标志来设置。详见第 17 章。

参数

name

运行时参数的名称。可用参数的说明见第 17 章以及SET参考页。此外，还有几个可以显示但不能设置的参数：

SERVER_VERSION

显示服务器的版本号。

SERVER_ENCODING

显示服务器端字符集编码。目前，该参数可以显示但不能设置，因为编码在数据库创建时就已确定。

LC_COLLATE

显示数据库用于排序规则（文本排序）的区域设置。目前，该参数可以显示但不能设置，因为该设置在initdb时就已确定。

LC_CTYPE

显示数据库用于字符分类的区域设置。目前，该参数可以显示但不能设置，因为该设置在initdb时就已确定。

IS_SUPERUSER

如果当前角色具有超级用户权限，则为真。

ALL

显示所有配置参数的值及其描述。

注解

函数current_setting也会产生等效的输出，见第 9.20 节。

示例

显示参数`DateStyle`的当前设置：

```
SHOW DateStyle;
 DateStyle
-----------
 ISO, MDY
(1 row)
```

显示参数`geqo`的当前设置：

```
SHOW geqo;
 geqo
------
 on
(1 row)
```

显示所有设置：

```
SHOW ALL;
            name          | setting |                 description
--------------------------+---------+-----------------------------------
---------------
 allow_system_table_mods | off     | Allows modifications of the
 structure of ...
    .
    .
    .
 xmloption               | content | Sets whether XML data in implicit
 parsing ...
 zero_damaged_pages      | off     | Continues processing past damaged
 page headers.
(196 rows)
```

兼容性

`SHOW`命令是一种 PostgreSQL扩展。

另见

SET, RESET

START TRANSACTION

START TRANSACTION — 开始一个事务块

大纲

```
START TRANSACTION [ transaction_mode [, ...] ]

where transaction_mode is one of:

    ISOLATION LEVEL { SERIALIZABLE | REPEATABLE READ | READ COMMITTED
 | READ UNCOMMITTED }
    READ WRITE | READ ONLY
```

描述

该命令开始一个新的事务块。如果指定了隔离级别、读写模式或可延迟模式，
新事务就会具有这些特性，就像执行了SET TRANSACTION一样。它与BEGIN命令具有相同功能。

参数

关于本语句参数含义的信息，参见SET TRANSACTION。

兼容性

在 SQL 标准中，并不需要发出`START TRANSACTION`来开始一个事务块：任何 SQL 命令都会隐式开始一个事务块。PostgreSQL的行为可以看作是：对每条不跟在`START TRANSACTION`（或`BEGIN`）之后的命令，都会在其后隐式发出一个`COMMIT`，因此这种行为通常被称为“自动提交”。其他关系型数据库系统也可能为了方便而提供自动提交特性。

SQL 标准要求在连续的*transaction_modes*之间有逗号，但出于历史原因，PostgreSQL允许省略逗号。

另见SET TRANSACTION的兼容性小节。

另见

BEGIN, COMMIT, ROLLBACK, SAVEPOINT, SET TRANSACTION

TRUNCATE

TRUNCATE — 清空一个表或一组表

大纲

```
TRUNCATE [ TABLE ] name [, ...]
```

描述

TRUNCATE可以快速移除一组表中的所有行。它的效果与对每个表执行不带条件的DELETE相同，但由于它并不实际扫描这些表，因此速度更快。这一点对于大表尤其有用。

参数

name

要截断的表名（可以是模式限定的）。

注解

只有表的所有者才能对它执行TRUNCATE。

TRUNCATE不能用于被其他表通过外键引用的表，除非所有这类表也在同一条命令中被截断。因为在这种情况下检查有效性将需要扫描表，而该命令的意义恰恰在于避免这样做。

TRUNCATE不会运行这些表上可能存在的任何用户定义的 ON DELETE触发器。

示例

截断表bigtable和fattable：

```
TRUNCATE TABLE bigtable, fattable;
```

兼容性

SQL 标准中没有TRUNCATE命令。

UNLISTEN

UNLISTEN — 停止监听通知

大纲

```
UNLISTEN { name | * }
```

描述

UNLISTEN用于移除现有的 NOTIFY事件注册。UNLISTEN会取消当前 PostgreSQL会话作为名为*name*的通知的任何现有监听注册。特殊通配符*会取消当前会话的所有监听注册。

NOTIFY中对LISTEN和 NOTIFY的使用有更详细的讨论。

参数

name

通知的名称（任意标识符）。

*

该会话当前的所有监听注册都会被清除。

注解

可以取消监听自己并未监听的通道，不会出现警告或错误。

每个会话结束时，都会自动执行UNLISTEN *。

示例

建立一个监听注册：

```
LISTEN virtual;
NOTIFY virtual;
Asynchronous notification "virtual" received from server process with
 PID 8448.
```

执行UNLISTEN之后，后续的NOTIFY 命令将被忽略：

```
UNLISTEN virtual;
NOTIFY virtual;
-- no NOTIFY event is received
```

兼容性

SQL 标准中没有UNLISTEN语句。

另见

LISTEN, NOTIFY

UPDATE

UPDATE — 更新表中的行

大纲

```
UPDATE [ ONLY ] table SET column = { expression | DEFAULT } [, ...]
    [ FROM fromlist ]
    [ WHERE condition ]
```

描述

UPDATE会更改所有满足条件的行中指定列的值。只需在SET子句中提及需要修改的列；未被显式修改的列会保留其原有值。

默认情况下，UPDATE将更新指定表及其所有子表中的行。如果只想更新提及的特定表，必须使用ONLY子句。

有两种方法可以利用数据库中其他表所包含的信息来修改一个表：使用子选择，或者在FROM子句中指定附加表。哪种技术更合适取决于具体情况。

必须拥有该表上的UPDATE权限才能更新它，同时还必须拥有在*expression*或*condition*中读取其值的任何表上的SELECT权限。

参数

table

要更新的现有表的名称（可以是模式限定的）。

column

table 所指定表中的列名。如有需要，列名可以使用子字段名或数组下标进行限定。

expression

要赋给该列的表达式。该表达式可以使用该表中这一列和其他列的旧值。

DEFAULT

把该列设置为其默认值（如果没有为它指定特定的默认表达式，则该值为 NULL）。

fromlist

一个表表达式列表，允许其他表的列出现在WHERE条件以及更新表达式中。这类似于SELECT语句的FROM Clause中可以指定的表列表。注意，除非打算进行自连接（此时它必须在*fromlist*中带有别名出现），否则目标表不能出现在*fromlist*中。

condition

一个返回boolean类型值的表达式。只有使这个表达式返回true的行才会被更新。

输出

成功完成后，UPDATE 命令返回如下形式的命令标签：

```
UPDATE count
```

其中*count*是被更新的行数。如果*count*为 0，说明没有行匹配*condition*（这不算错误）。

注解

当存在FROM子句时，本质上会将目标表与 *fromlist*中提到的表连接起来，而连接的每一条输出行都代表对目标表的一次更新操作。使用 FROM时，应确保对每一个要修改的目标行，连接至多生成一条输出行。换言之，一条目标行不应与其他表中的多于一行成功连接。如果发生这种情况，则只会使用其中某一条连接行来更新目标行，但具体使用哪一条并不容易预测。

因为存在这种不确定性，所以仅在子选择中引用其他表会更安全，
尽管这种写法通常比使用连接更难阅读、也更慢。

示例

把表films的kind 列中的单词Drama改为Dramatic：

```
UPDATE films SET kind = 'Dramatic' WHERE kind = 'Drama';
```

在表weather的一行中调整温度项，并将降水量重置为其默认值：

```
UPDATE weather SET temp_lo = temp_lo+1, temp_hi = temp_lo+15, prcp =
 DEFAULT
  WHERE city = 'San Francisco' AND date = '2003-07-03';
```

将负责 Acme Corporation 账户的销售人员的销量计数加一，使用 FROM 子句语法：

```
UPDATE employees SET sales_count = sales_count + 1 FROM accounts
  WHERE accounts.name = 'Acme Corporation'
  AND employees.id = accounts.sales_person;
```

执行同样的操作，在 WHERE 子句中使用子选择：

```
UPDATE employees SET sales_count = sales_count + 1 WHERE id =
  (SELECT sales_person FROM accounts WHERE name = 'Acme Corporation');
```

尝试插入一个新库存项及其库存量。如果该项已存在，则改为更新现有项的库存量。
要在不致使整个事务失败的情况下做到这一点，请使用保存点：

```
BEGIN;
-- 其他操作
SAVEPOINT sp1;
INSERT INTO wines VALUES('Chateau Lafite 2003', '24');
-- 假设上面的语句因唯一键冲突而失败，
-- 那么现在发出这些命令：
ROLLBACK TO sp1;
UPDATE wines SET stock = stock + 24 WHERE winename = 'Chateau Lafite
 2003';
-- 继续执行其他操作，最后
```

```
COMMIT;
```

兼容性

这个命令符合SQL标准，不过FROM子句是 PostgreSQL扩展。

某些其他数据库系统提供一种 FROM 选项，其中目标表被认为要在 FROM 中再次列出。PostgreSQL 不是这样解释 FROM 的。移植使用这一扩展的应用时要小心。

VACUUM

VACUUM — 垃圾收集并按需分析数据库

大纲

```
VACUUM [ FULL | FREEZE ] [ VERBOSE ] [ table ]
VACUUM [ FULL | FREEZE ] [ VERBOSE ] ANALYZE [ table [ (column [, ...
] ) ] ]
```

描述

VACUUM回收被删除元组占用的存储空间。在正常的 PostgreSQL运行中，被删除或因更新而过时的元组并不会从其表中物理移除；它们会一直保留，直到执行 VACUUM。因此有必要定期执行VACUUM，尤其是对频繁更新的表。

不带参数时，VACUUM会处理当前数据库中的每个表。带参数时，VACUUM则只处理该表。

VACUUM ANALYZE会对每个选定表先执行 VACUUM，再执行ANALYZE。这种便捷的组合形式很适合例行维护脚本使用。关于其处理细节，参见ANALYZE。

普通的VACUUM（不带FULL）只是回收空间并使其可被重用。
这种形式的命令可以与表的正常读写并行运行，因为它不会获得独占锁。VACUUM FULL会做更彻底的处理，包括跨块移动元组，以尽量把表压缩到最少数量的磁盘块。这种形式要慢得多，
并且在处理每个表时都需要独占锁。

FREEZE是一种特殊用途的选项，它使元组尽可能早地被标记为“冻结”，
而不是等到它们相当旧时才冻结。如果在同一数据库中没有其他打开的事务时执行此操作，
那么可以保证数据库中的所有元组都已被“冻结”，无论数据库多长时间不清理，都不会受到事务ID 回卷问题的影响。不建议在日常场景中使用FREEZE。它唯一的预期用途是在准备用户定义模板数据库时，或者用于其他完全只读、不会接受例行VACUUM维护操作的数据库。参见第 22 章了解更多细节。

参数

FULL

选择“完全”清理，它可以回收更多空间，但耗时更长，并且会独占锁定该表。
这种方法还需要额外的磁盘空间，因为它会写出该表的一个新副本，
并且在操作完成之前不会释放旧副本。通常，
只有当需要从表内回收大量空间时才应使用这种方法。

FREEZE

选择激进地“冻结”元组。

VERBOSE

为每个表输出详细的清理活动报告。

ANALYZE

更新规划器用来确定查询最高效执行方式的统计信息。

table

要清理的特定表的名称（可选地带模式限定）。默认为当前数据库中的所有表。

column

要分析的特定列名。默认会分析所有列。

输出

指定VERBOSE时，VACUUM会输出进度消息，指示当前正在处理哪个表，
同时还会打印这些表的各种统计信息。

注解

我们建议经常（至少每晚）清理活跃的生产数据库，以移除过期行。在添加或删除大量行之后，对受影响的表执行VACUUM ANALYZE命令可能是个好主意。这会用所有近期变更的结果更新系统目录，使PostgreSQL查询规划器在规划查询时能够做出更好的选择。

FULL选项不建议在日常场景中使用，但在某些特殊情况下可能很有用。例如，
当你删除了表中的绝大多数行，并希望该表在物理上收缩以占用更少磁盘空间时，
这个选项就比较合适。VACUUM FULL通常会比普通 VACUUM更大幅度地收缩表。

VACUUM会显著增加 I/O 流量，这可能导致其他活动会话性能变差。因此，有时建议使用基于代价的清理延迟特性。详见第 17.4.4 节。

示例

下面是对回归数据库中的一个表运行 VACUUM 的例子：

```
regression=# VACUUM (VERBOSE, ANALYZE) onek;
INFO:  vacuuming "public.onek"
INFO:  index "onek_unique1" now contains 1000 tuples in 14 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.08u sec elapsed 0.18 sec.
INFO:  index "onek_unique2" now contains 1000 tuples in 16 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.00s/0.07u sec elapsed 0.23 sec.
INFO:  index "onek_hundred" now contains 1000 tuples in 13 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.08u sec elapsed 0.17 sec.
INFO:  index "onek_stringu1" now contains 1000 tuples in 48 pages
DETAIL:  3000 index tuples were removed.
0 index pages have been deleted, 0 are currently reusable.
CPU 0.01s/0.09u sec elapsed 0.59 sec.
INFO:  "onek": removed 3000 tuples in 108 pages
DETAIL:  CPU 0.01s/0.06u sec elapsed 0.07 sec.
INFO:  "onek": found 3000 removable, 1000 nonremovable tuples in 143
 pages
DETAIL:  0 dead tuples cannot be removed yet.
There were 0 unused item pointers.
```

```
0 pages are entirely empty.
CPU 0.07s/0.39u sec elapsed 1.56 sec.
INFO:  analyzing "public.onek"
INFO:  "onek": 36 pages, 1000 rows sampled, 1000 estimated total rows
VACUUM
```

兼容性

在SQL标准中没有`VACUUM`语句。

另见

vacuumdb, 基于代价的清理延迟

PostgreSQL 客户端应用

本部分包含PostgreSQL客户端应用和工具的参考信息。并非所有这些命令都适用于一般用途；有些可能需要特殊权限。这些应用的共同点是，它们都可以在任意主机上运行，而不受数据库服务器所在位置的限制。

目录

clusterdb 873
createdb 875
createlang 878
createuser 880
dropdb 884
droplang 886
dropuser 888
ecpg 890
pg_config 892
pg_dump 895
pg_dumpall 901
pg_restore 905
psql 911
reindexdb 935
vacuumdb 938

clusterdb

clusterdb — 对一个PostgreSQL数据库进行聚簇

大纲

`clusterdb` [*connection-option*...] [--table | -t *table*] [*dbname*]
`clusterdb` [*connection-option*...] [[--all] | [-a]]

描述

clusterdb是一个用于重新聚簇PostgreSQL 数据库中各表的工具。它会找出此前已经聚簇过的表，并按照上次使用的同一索引再次对其进行聚簇。从未聚簇过的表不会受到影响。

clusterdb是 SQL 命令CLUSTER的一个包装器。使用该工具聚簇数据库，与使用其他访问服务器的方法聚簇数据库，并没有实质区别。

选项

clusterdb接受以下命令行参数：

`-a`
`--all`

聚簇所有数据库。

`[-d]` *dbname*
`[--dbname=]`*dbname*

指定要聚簇的数据库名。如果未指定并且没有使用 `-a`（或`--all`），则从环境变量 `PGDATABASE` 读取数据库名。如果该环境变量也未设置，则使用为该连接指定的用户名作为数据库名。

`-e`
`--echo`

回显clusterdb生成并发送到服务器的命令。

`-q`
`--quiet`

不显示响应。

`-t` *table*
`--table` *table*

只对 *table* 执行聚簇。

clusterdb还接受以下用于连接参数的命令行参数：

`-h` *host*
`--host` *host*

指定运行服务器的机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p port`
`--port port`

指定服务器监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。

`-U username`
`--username username`

用于连接的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

如果遇到问题，请参见CLUSTER和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须在目标主机上运行。此外，
libpq前端库所使用的任何默认连接设置和环境变量也都会生效。

示例

要聚簇数据库`test`：

```
$ clusterdb test
```

要聚簇名为`xyzzy`的数据库中的单个表`foo`：

```
$ clusterdb --table foo xyzzy
```

参见

CLUSTER, 环境变量（第 28.11 节）

createdb

createdb — 创建一个新的PostgreSQL数据库

大纲

`createdb` [*option*...] [*dbname*] [*description*]

描述

createdb创建一个新的PostgreSQL数据库。

通常，执行此命令的数据库用户会成为新数据库的所有者。不过，
如果执行用户具有相应的权限，也可以通过`-O`选项指定不同的所有者。

createdb是SQL命令CREATE DATABASE的包装器。通过此工具创建数据库与通过其他访问服务器的方法创建数据库，在效果上没有区别。

选项

createdb accepts the following command-line arguments:

dbname

指定要创建的数据库名称。该名称必须在此数据库集簇中的所有PostgreSQL数据库中唯一。默认情况下，会创建一个与当前系统用户同名的数据库。

description

指定要与新创建的数据库关联的注释。

`-D` *tablespace*
`--tablespace` *tablespace*

指定数据库的默认表空间。

`-e`
`--echo`

回显createdb生成并发送到服务器的命令。

`-E` *encoding*
`--encoding` *encoding*

指定该数据库使用的字符集编码。PostgreSQL服务器支持的字符集见第 21.2.1 节。

`-O` *owner*
`--owner` *owner*

指定将拥有新数据库的数据库用户。

`-q`
`--quiet`

不显示响应。

`-T` *`template`*
`--template` *`template`*

指定用于构建该数据库的模板数据库。

`-D`、`-E`、`-O`和`-T`选项对应底层 SQL 命令CREATE DATABASE的选项；有关这些选项的更多信息请参见该命令。

createdb还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host` *`host`*

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p` *`port`*
`--port` *`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

以该用户名进行连接。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`

如果设置了该变量，其值就是要创建的数据库名称，除非被命令行覆盖。

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数。如果命令行和`PGDATABASE`都没有指定要创建的数据库名称，则`PGUSER`也会决定该名称。

诊断

如果遇到问题，请参阅CREATE DATABASE和psql中关于潜在问题和错误消息的讨论。数据库服务器必须在目标主机上运行。此外，由
libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要使用默认数据库服务器创建`demo`数据库：

```
$ createdb demoCREATE DATABASE
```

这个响应与你运行`CREATE DATABASE` SQL命令所得到的响应相同。

要使用主机`eden`上、端口为 5000 的服务器，并以 `LATIN1`编码方案创建`demo`数据库，可以看看下面的底层命令：

```
$ createdb -p 5000 -h eden -E LATIN1 -e demoCREATE DATABASE "demo"
 WITH ENCODING = 'LATIN1'CREATE DATABASE
```

参见

dropdb, CREATE DATABASE, 环境变量（第 28.11 节）

createlang

createlang — 定义一种新的PostgreSQL过程语言

大纲

```
createlang [connection-option...] langname [dbname]
createlang [connection-option...] [--list] | [-l] dbname
```

描述

createlang是一个向PostgreSQL数据库添加新编程语言的工具。createlang只是CREATE LANGUAGE命令的一个包装，更多信息请参见该命令。

选项

createlang accepts the following command-line arguments:

langname

指定要安装的过程语言的名称。

[-d] *dbname*
[--dbname=]*dbname*

指定要向哪个数据库添加该语言。默认使用与当前系统用户同名的数据库。

-e
--echo

在执行 SQL 命令时把它们显示出来。

-l
--list

显示目标数据库中已安装语言的列表。

createlang 还接受下列用于连接参数的命令行参数：

-h *host*
--host *host*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

-p *port*
--port *port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

-U *username*
--username *username*

要用来连接的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

大多数错误消息都是不言自明的。如果不然，可以带 `--echo` 选项运行 createlang，然后查看相应的 SQL 命令了解详情。

注解

使用droplang可移除一个语言。

示例

要将语言 `pltcl` 安装到数据库 `template1` 中：

```
$ createlang pltcl template1
```

注意，将该语言安装到 `template1` 中后，之后创建的数据库中也会自动安装它。

参见

droplang, CREATE LANGUAGE, 环境变量（第 28.11 节）

createuser

createuser — 定义一个新的PostgreSQL用户账户

大纲

`createuser` [*option*...] [*username*]

描述

createuser创建一个新的PostgreSQL用户（更准确地说，是一个角色）。只有超级用户以及具有`CREATEROLE`权限的用户才能创建新用户，因此调用
createuser的用户必须能够以超级用户或具有`CREATEROLE`权限的用户身份连接。

如果希望创建新的超级用户，必须以超级用户身份连接，而不能仅凭`CREATEROLE`权限。
超级用户意味着能够绕过数据库内所有访问权限检查，因此不应轻易授予超级用户身份。

createuser是SQL命令CREATE　ROLE的包装器。通过此工具创建用户与通过其他访问服务器的方法创建用户，在效果上没有区别。

选项

createuser接受以下命令行参数：

username

指定要创建的PostgreSQL用户的名称。
该名称必须不同于此PostgreSQL安装中所有现有角色的名称。

`-s`
`--superuser`

新用户将成为超级用户。

`-S`
`--no-superuser`

新用户将成为超级用户。

`-d`
`--createdb`

允许新用户创建数据库。

`-D`
`--no-createdb`

允许新用户创建数据库。

`-r`
`--createrole`

允许新用户创建新角色（即该用户将具有`CREATEROLE`权限）。

`-R`
`--no-createrole`

不允许新用户创建新角色。

`-l`
`--login`

允许新用户登录（也就是说，该用户名可用作初始会话用户标识符）。这是默认设置。

`-L`
`--no-login`

不允许新用户登录。（不具有登录权限的角色仍可用于管理数据库权限。）

`-i`
`--inherit`

新角色将自动继承其所属角色的权限。这是默认设置。

`-I`
`--no-inherit`

新角色将不会自动继承其所属角色的权限。

`-c` *`number`*
`--connection-limit` *`number`*

为新用户设置最大连接数。默认不设限制。

`-P`
`--pwprompt`

如果给出此选项，createuser会提示输入新用户的密码。如果不打算使用密码认证，则不需要这样做。

`-E`
`--encrypted`

加密存储在数据库中的用户密码。如果未指定，则使用默认的密码行为。

`-N`
`--unencrypted`

不加密存储在数据库中的用户密码。如果未指定，则使用默认的密码行为。

`-e`
`--echo`

回显createuser生成并发送给服务器的命令。

`-q`
`--quiet`

不显示响应。

如果命令行上没有指定名称和其他缺少的信息，将会提示你输入。

createuser还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host` *`host`*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

`-p` *`port`*
`--port` *`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

指定连接时使用的用户名（不是要创建的用户名）。

`-W`
`--password`

强制提示输入密码（用于连接服务器，而不是新用户的密码）。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

若遇到困难，请参见CREATE ROLE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须在目标主机上运行。此外，libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要在默认数据库服务器上创建用户 `joe`：

```
$ createuser joeShall the new role be a superuser? (y/n) nShall the
 new role be allowed to create databases? (y/n) nShall the new role be
 allowed to create more new roles? (y/n) nCREATE USER
```

要使用主机 `eden`、端口 5000 上的服务器创建同一个用户 `joe`，避开提示并查看底层命令：

```
$ createuser -h eden -p 5000 -S -D -R -e joeCREATE ROLE joe
 NOSUPERUSER NOCREATEDB NOCREATEROLE INHERIT LOGIN;CREATE ROLE
```

创建用户`joe`作为超级用户，并立即分配密码：

```
$ createuser -P -s -e joeEnter password for new role: xyzzyEnter
 it again: xyzzyCREATE ROLE joe PASSWORD 'xyzzy' SUPERUSER CREATEDB
 CREATEROLE INHERIT LOGIN;CREATE ROLE
```

上面的示例中，输入密码时实际上不会回显新密码，但为清晰起见，我们显示了所输入的内容。不过，如示例所示，密码会出现在回显的命令中——因此，如果还有其他人能看到你的屏幕，在分配密码时就不要使用`-e`。

参见

dropuser, CREATE ROLE, 环境变量（第 28.11 节）

dropdb

dropdb — 移除一个PostgreSQL数据库

大纲

`dropdb` [*option*...] *dbname*

描述

dropdb删除一个现有的 PostgreSQL数据库。执行此命令的用户必须是数据库超级用户或该数据库的拥有者。

dropdb是SQL命令DROP DATABASE的包装器。通过此工具删除数据库与通过其他访问服务器的方法删除数据库，在效果上没有区别。

选项

dropdb接受以下命令行参数：

dbname

指定要移除的数据库名称。

`-e`
`--echo`

回显dropdb生成并发送给服务器的命令。

`-i`
`--interactive`

在执行任何破坏性操作之前发出确认提示。

`-q`
`--quiet`

不显示响应。

dropdb也接受下列用于连接参数的命令行参数：

`-h` *host*
`--host` *host*

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p` *port*
`--port` *port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *username*
`--username` *username*

指定用于连接的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

若遇到困难，请参见DROP DATABASE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须运行在目标主机上。此外，libpq前端库使用的任何默认连接设置和环境变量也都会生效。

示例

要在默认数据库服务器上删除数据库`demo`：

```
$ dropdb demoDROP DATABASE
```

要使用主机`eden`、端口 5000 上的服务器删除数据库`demo`，并带确认提示以及查看底层命令：

```
$ dropdb -p 5000 -h eden -i -e demoDatabase "demo" will be permanently
 deleted.
Are you sure? (y/n) yDROP DATABASE "demo"
DROP DATABASE
```

另见

createdb, DROP DATABASE, 环境变量（第 28.11 节）

droplang

droplang — 移除一种PostgreSQL过程语言

大纲

```
droplang [connection-option...] langname [dbname]
droplang [connection-option...] [--list] | [-l] dbname
```

描述

droplang是一个从PostgreSQL数据库移除现有编程语言的工具。
droplang可以删除任何过程语言，甚至是那些不由PostgreSQL发行版提供的过程语言。

虽然后端编程语言可以直接用几条SQL命令移除，但建议使用droplang，
因为它会执行一些检查，而且用起来容易得多。更多信息见DROP LANGUAGE。

选项

droplang 接受下列命令行参数：

langname

指定要移除的过程语言的名称。

[-d] *dbname*
[--dbname=]*dbname*

指定要从哪个数据库中移除该语言。默认使用与当前系统用户同名的数据库。

-e
--echo

在执行 SQL 命令时把它们显示出来。

-l
--list

显示目标数据库中已安装语言的列表。

droplang 还接受下列用于连接参数的命令行参数：

-h *host*
--host *host*

指定服务器所在主机的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字所在目录。

-p *port*
--port *port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

要用来连接的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

大多数错误消息都是不言自明的。如果不然，可以带 `--echo` 选项运行 droplang，然后查看相应的 SQL 命令了解详情。

注解

使用createlang可添加一个语言。

示例

要移除语言 `pltcl`：

```
$ droplang pltcl dbname
```

参见

createlang, DROP LANGUAGE, 环境变量（第 28.11 节）

dropuser

dropuser — 移除一个PostgreSQL用户账户

大纲

`dropuser` [*option*...] [*username*]

描述

dropuser移除现有的PostgreSQL用户。只有超级用户和拥有`CREATEROLE`权限的用户才能移除PostgreSQL用户。（要移除超级用户，你自己也必须是超级用户。）

dropuser是SQL命令DROP ROLE的包装器。通过此工具删除用户与通过其他访问服务器的方法删除用户，在效果上没有区别。

选项

dropuser 接受下列命令行参数：

username

指定要移除的 PostgreSQL 用户的名称。如果命令行上没有指定名称，将会提示你输入。

`-e`
`--echo`

回显dropuser生成并发送给服务器的命令。

`-i`
`--interactive`

在实际移除用户之前提示确认。

`-q`
`--quiet`

不显示响应。

dropuser也接受下列用于连接参数的命令行参数：

`-h` *host*
`--host` *host*

指定服务器运行所在主机的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字所在目录。

`-p` *port*
`--port` *port*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

用于连接的用户名（不是要删除的用户名）。

`-W`
`--password`

强制提示输入密码。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

若遇到困难，请参见DROP ROLE和psql，其中讨论了潜在问题和错误消息。数据库服务器必须在目标主机上运行。此外，libpq前端库所使用的任何默认连接设置和环境变量也都会生效。

示例

要从默认数据库服务器上移除用户`joe`：

```
$ dropuser joeDROP ROLE
```

要使用主机`eden`、端口 5000 上的服务器移除用户`joe`，并带确认提示以及查看底层命令：

```
$ dropuser -p 5000 -h eden -i -e joeRole "joe" will be permanently
 removed.
Are you sure? (y/n) yDROP ROLE "joe"
DROP ROLE
```

另见

createuser, DROP ROLE, 环境变量（第 28.11 节）

ecpg

ecpg — 嵌入式 SQL C 预处理器

大纲

ecpg [*option*...] *file*...

描述

ecpg是供 C 程序使用的嵌入式 SQL 预处理器。它会把嵌入了 SQL 语句的 C 程序转换成普通 C 代码，方法是将 SQL 调用替换为特殊的函数调用。生成的输出文件随后可以由任何 C 编译器工具链处理。

ecpg 会把命令行上给出的每个输入文件转换为对应的 C 输出文件。输入文件最好带有扩展名 .pgc，这种情况下会用 .c 替换该扩展名来确定输出文件名。如果输入文件的扩展名不是 .pgc，则通过在整个文件名后追加 .c 来计算输出文件名。也可以使用 -o 选项覆盖输出文件名。

本参考页不描述嵌入式 SQL 语言。有关该主题的更多信息，见第 30 章。

选项

ecpg accepts the following command-line arguments:

-c

自动从 SQL 代码生成某些 C 代码。目前，这适用于EXEC SQL TYPE。

-C *mode*

设置兼容模式。*mode*可以是INFORMIX或INFORMIX_SE。

-D *symbol*

定义一个 C 预处理器符号。

-i

也解析系统 include 文件。

-I *directory*

指定一个附加的 include 路径，用于查找通过EXEC SQL INCLUDE包含的文件。默认路径依次为 .（当前目录）、/usr/local/include、在编译时定义的PostgreSQL include 目录（默认：/usr/local/pgsql/include），以及 /usr/include。

-o *filename*

指定ecpg应将全部输出写入给定的*filename*。

-r *option*

选择一种运行时行为。当前，*option* 只能是 no_indicator。

`-t`

打开事务自动提交。在此模式下，除非 SQL 命令位于显式事务块中，否则每条 SQL 命令都会自动提交。在默认模式下，只有发出`EXEC SQL COMMIT`时才会提交命令。

`-v`

打印附加信息，包括版本和 include 路径。

`--help`

显示命令用法的简要摘要，然后退出。

`--version`

输出版本信息，然后退出。

注解

在编译预处理后的 C 代码文件时，编译器必须能够在PostgreSQL的 include 目录中找到ECPG头文件。因此，调用编译器时你可能需要使用`-I`选项（例如`-I/usr/local/pgsql/include`）。

使用嵌入式 SQL 的 C 程序必须与`libecpg`库链接，例如可使用链接器选项`-L/usr/local/pgsql/lib -lecpg`。

适合该安装的这两个目录的值可以通过pg_config查出。

示例

如果你有一个名为`prog1.pgc`的嵌入式 SQL C 源文件，就可以使用下列命令序列创建一个可执行程序：

```
ecpg prog1.pgc
cc -I/usr/local/pgsql/include -c prog1.c
cc -o prog1 prog1.o -L/usr/local/pgsql/lib -lecpg
```

pg_config

pg_config — 获取已安装的PostgreSQL版本的信息

大纲

`pg_config` [*option*...]

Description

pg_config工具打印当前已安装的PostgreSQL版本的配置参数。例如，希望与PostgreSQL进行接口的软件包可以使用它，以便找到所需的头文件和库。

Options

要使用 pg_config，请提供下列一个或多个选项：

`--bindir`

打印用户可执行文件的位置。例如，可以用它来寻找 `psql` 程序。通常这也是 `pg_config` 程序所在的位置。

`--docdir`

打印文档文件所在的位置。（如果构建PostgreSQL时指定了 `--without-docdir`，这里将是一个空字符串。）

`--includedir`

打印 HTML 文档文件所在的位置。

`--pkgincludedir`

打印客户端接口的 C 头文件所在的位置。

`--includedir-server`

打印其他 C 头文件所在的位置。

`--libdir`

打印服务器编程所用 C 头文件所在的位置。

`--pkglibdir`

打印目标代码库所在的位置。

`--localedir`

打印区域支持文件的位置。（如果在构建 PostgreSQL 时没有配置区域支持，这将是一个空字符串。）

`--mandir`

打印手册页的位置。

`--sharedir`

打印手册页所在的位置。

`--sysconfdir`

打印与体系结构无关的支持文件所在的位置。

`--pgxs`

打印系统范围的配置文件所在的位置。

`--configure`

打印为构建 PostgreSQL 而配置时传递给 `configure` 脚本的选项。这可用于复现完全相同的配置，或者查明某个二进制包是用什么选项构建的。（不过注意，
二进制包通常包含厂商特有的自定义补丁。）另请参见下面的例子。

`--cc`

打印构建PostgreSQL时使用的CC宏的值。

`--cppflags`

打印构建 PostgreSQL 时使用的 CPPFLAGS 宏的值。它显示预处理时需要的 C 编译器开关（通常是 `-I` 开关）。

`--cflags`

打印构建 PostgreSQL 时使用的 CFLAGS 宏的值。它显示 C 编译器开关。

`--cflags_sl`

打印构建PostgreSQL时所用CFLAGS_SL宏的值。这显示构建共享库时使用的额外 C 编译器开关。

`--ldflags`

打印构建PostgreSQL时所用LDFLAGS宏的值。这显示链接器开关。

`--ldflags_sl`

打印构建PostgreSQL时所用LDFLAGS_SL宏的值。这显示构建共享库时使用的链接器开关。

`--libs`

打印构建 PostgreSQL 时使用的 LIBS 宏的值。它通常包含链接进 PostgreSQL 的外部库的 `-l` 开关。

`--version`

打印 PostgreSQL 的版本。

如果给出了多个选项，信息会按该顺序打印，每行一项。如果没有给出选项，
则打印所有可用信息，并带有标签。

注意

`--includedir-server` 选项是在 PostgreSQL 7.2 中加入的。在之前的版本中，服务器头文件与客户端头文件安装在同一位置，可以用 `--includedir` 选项查询。要让你的软件包同时处理这两种情况，请先尝试较新的选项，并测试退出状态以确定它是否成功。

选项`--docdir`、`--pkgincludedir`、`--localedir`、`--mandir`、`--sharedir`、`--sysconfdir`、`--cc`、`--cppflags`、`--cflags`、`--cflags_sl`、`--ldflags`、`--ldflags_sl`以及`--libs`是在PostgreSQL 8.1 中加入的。

示例

要重现当前 PostgreSQL 安装的构建配置，请运行下列命令：

```
eval ./configure `pg_config --configure`
```

`pg_config --configure`的输出包含 shell 引号，因此带空格的参数能够被正确表示。因此，要得到正确的结果，需要使用`eval`。

History

`pg_config`工具最早出现在 PostgreSQL 7.1 中。

在 PostgreSQL 7.1 之前的发行版中，还没有 `pg_config`，当时不存在获取等效配置信息的方法。

pg_dump

pg_dump — 将 PostgreSQL 数据库导出为 SQL 脚本或其他格式

大纲

pg_dump [*option*...] [*dbname*]

描述

pg_dump是一个用于备份PostgreSQL数据库的工具。即使数据库正在被并发使用，它也能生成一致的备份。pg_dump不会阻塞其他用户访问数据库（无论读还是写）。

转储可以输出为脚本格式或归档文件格式。脚本转储是纯文本文件，包含把数据库重建到保存时状态所需的 SQL 命令。要从这样的脚本恢复，只需将其交给psql。脚本文件甚至可以在其他机器和其他体系结构上用于重建数据库；经过一些修改后，甚至也可以用于其他 SQL 数据库产品。

另一类归档文件格式必须结合pg_restore来重建数据库。它们允许 pg_restore 有选择地恢复某些内容，甚至在恢复之前重新排列条目。归档文件格式被设计为可跨体系结构移植。

当使用一种归档文件格式并与pg_restore结合使用时，pg_dump提供了一种灵活的归档和传输机制。pg_dump可以用来备份整个数据库，然后pg_restore可以用来检查归档和/或选择要恢复数据库的哪些部分。最灵活的输出文件格式是“custom”格式（`-Fc`）。它允许选择和重新排序所有归档条目，并且默认会被压缩。tar格式（`-Ft`）不压缩，而且在装载时无法对数据重新排序，但在其他方面相当灵活；此外，它可以用标准的 Unix 工具（如`tar`）来处理。

运行 pg_dump 时，应检查输出中是否有任何警告（打印到标准错误），尤其要结合下面列出的限制来查看。

选项

下列命令行选项控制输出的内容和格式。

dbname

指定要转储的数据库名称。如果未指定，则使用环境变量`PGDATABASE`。如果未设置该变量，则使用连接指定的用户名。

`-a`
`--data-only`

只转储数据，不转储模式（数据定义）。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-c`
`--clean`

输出在创建数据库对象之前清除（删除）它们的命令。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-C`
`--create`

让输出以创建数据库本身并重新连接到新建数据库的命令开始。（使用这种形式的脚本时，在运行脚本之前连接到哪个数据库无关紧要。）

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-d`
`--inserts`

将数据转储为`INSERT`命令（而不是`COPY`）。这会使恢复非常缓慢；它主要用于生成可以装入非PostgreSQL数据库的转储。注意，如果重新安排了列顺序，恢复可能会彻底失败。`-D`选项更安全一些，但速度也更慢。另外，虽然这个选项在遇到无效数据时会生成错误，但它允许其他`INSERT`继续向表中装载数据。

`-D`
`--column-inserts`
`--attribute-inserts`

将数据转储为带显式列名的`INSERT`命令（`INSERT INTO` *`table`* `(`*`column`*`, ...) VALUES ...`）。这会使恢复非常缓慢；它主要用于生成可以装入非PostgreSQL数据库的转储。另外，虽然这个选项在遇到无效数据时会生成错误，但它允许其他`INSERT`继续向表中装载数据。

`-E` *`encoding`*
`--encoding=`*`encoding`*

在指定的字符集编码中创建转储。默认情况下，转储将以数据库编码创建。（获得相同结果的另一种方法是将`PGCLIENTENCODING`环境变量设置为所需的转储编码。）

`-f` *`file`*
`--file=`*`file`*

把输出发送到指定的文件。如果省略，则使用标准输出。

`-F` *`format`*
`--format=`*`format`*

选择输出的格式。*`format`*可以是下列之一：

`plain`

输出纯文本 SQL 脚本文件（默认）

`t`

输出一个适合输入到pg_restore的`tar`归档。使用这种归档格式，可以在恢复数据库时重新排序和/或排除数据库对象。还可以限制恢复时重新装载哪些数据。

`c`

输出适合供 pg_restore 使用的 custom 归档。这是最灵活的输出格式，因为它既允许对装载的数据进行重新排序，也允许对对象定义进行重新排序。这种格式默认也会压缩。

`-i`
`--ignore-version`

忽略pg_dump与数据库服务器之间的版本不匹配。

pg_dump可以处理以前版本PostgreSQL的数据库，但非常老的版本不再受支持（目前是 7.0 之前的版本）。如果需要跳过版本检查，可使用此选项（如果pg_dump随后失败，别怪我们没有提醒你）。

`-n` *`schema`*
`--schema=`*`schema`*

只转储*`schema`*的内容。未指定此选项时，将转储目标数据库中的所有非系统模式。

> **注意**
>
> 在这种模式下，pg_dump不会尝试转储所选模式中的对象可能依赖的任何其他数据库对象。因此，不能保证单模式转储的结果能够单独成功恢复到一个干净的数据库中。

`-o`
`--oids`

将对象标识符（OID）作为每个表的数据的一部分进行转储。如果应用程序以某种方式引用OID列（例如在外键约束中），请使用此选项。否则，不应使用此选项。

`-O`
`--no-owner`

不输出用于把对象所有权设置成与原始数据库一致的命令。默认情况下，pg_dump 会发出 `ALTER OWNER` 或 `SET SESSION AUTHORIZATION` 语句来设置新建数据库对象的所有权。除非脚本由超级用户（或拥有脚本中所有对象的同一用户）启动，否则这些语句会在运行脚本时报错。若要创建一个可由任意用户恢复、并让该用户拥有所有对象的脚本，请指定 `-O`。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-R`
`--no-reconnect`

这个选项已经过时，但仍然被接受以保持向后兼容性。

`-s`
`--schema-only`

只转储对象定义（模式），不转储数据或统计信息。

`-S` *`username`*
`--superuser=`*`username`*

指定在禁用触发器时要使用的超级用户名。这只在使用 `--disable-triggers` 时相关。（通常最好省略它，而是以超级用户身份运行生成的脚本。）

`-t` *`table`*
`--table=`*`table`*

只转储*`table`*的数据。不同模式中可能存在多个同名的表；如果是这种情况，所有匹配的表都会被转储。要只选择一个表，可以同时指定`--schema`和`--table`。

> **注意**
>
> 在这种模式下，pg_dump不会尝试转储所选表可能依赖的任何其他数据库对象。因此，不能保证单表转储的结果能够单独成功恢复到一个干净的数据库中。

`-v`
`--verbose`

指定详细模式。这会使 pg_dump 将详细的对象注释、开始/停止时间写入转储文件，并把进度消息写到标准错误。

`-x`
`--no-privileges`
`--no-acl`

不转储访问权限（grant/revoke 命令）。

`-X disable-dollar-quoting`
`--disable-dollar-quoting`

此选项禁用函数体中的 dollar quoting，并强制改用 SQL 标准字符串语法对它们进行引用。

`-X disable-triggers`
`--disable-triggers`

此选项只在创建包含数据但不包含模式的转储时才相关。它指示 pg_dump 在输出中包含一些命令，以便在恢复数据时临时禁用目标表上的触发器。如果这些表上存在不希望在数据恢复期间触发的引用完整性检查或其他触发器，请使用此选项。

目前，为 `--disable-triggers` 输出的这些命令必须由超级用户执行。因此，还应通过 `-S` 指定一个超级用户名，或者更好的做法是谨慎地以超级用户身份运行生成的脚本。

此选项只对纯文本格式有意义。对于归档格式，可以在调用 `pg_restore` 时指定该选项。

`-X use-set-session-authorization`
`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是 `ALTER OWNER` 命令来确定对象所有权。这会让转储更符合标准，但根据转储中对象的历史，可能无法正确恢复。另外，使用 `SET SESSION AUTHORIZATION` 的转储肯定需要超级用户权限才能正确恢复，而 `ALTER OWNER` 只需较低权限。

`-Z` *`0..9`*
`--compress=`*`0..9`*

指定所用的压缩级别。零表示不压缩。对于自定义格式归档，该选项指定对各个表数据段的压缩，默认以适中的级别压缩。对于纯文本输出，

设置非零压缩级别会压缩整个输出文件，就像将其传给gzip处理一样；但默认不压缩。tar 归档格式目前完全不支持压缩。

下列命令行选项控制数据库连接参数。

`-h host`
`--host=host`

指定运行服务器的机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。默认值取自 `PGHOST` 环境变量（如果已设置）；否则会尝试使用 Unix 域套接字连接。

`-p port`
`--port=port`

指定服务器正在监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。默认值取自 `PGPORT` 环境变量（如果已设置），否则使用编译时默认值。

`-U username`

连接时使用的用户名。

`-W`

强制提示输入密码。如果服务器要求密码认证，这应该会自动发生。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数。

诊断

pg_dump 在内部执行 `SELECT` 语句。如果运行 pg_dump 时遇到问题，请确认能够例如使用psql从数据库中查询信息。

注意

如果数据库集簇在 `template1` 数据库中有任何本地添加的内容，要小心把 pg_dump 的输出恢复到一个真正空的数据库中；否则很可能因为这些新增对象的重复定义而报错。要创建一个不含任何本地添加的空数据库，应从 `template0` 而不是 `template1` 复制，例如：

```
CREATE DATABASE foo WITH TEMPLATE template0;
```

pg_dump有一些限制：

- 当选择仅数据转储并使用`--disable-triggers`选项时，pg_dump会在插入数据前发出命令禁用用户表上的触发器，并在数据插入完成后发出命令重新启用它们。如果恢复在中途停止，系统目录可能会保持在错误状态。

tar 归档的成员被限制为小于 8 GB 的尺寸。（这是 tar 文件格式固有的限制。）因此，如果任何一个表的文本表示超过了这个尺寸，就不能使用这种格式。tar 归档和任何其他输出格式的总大小没有限制，除非操作系统有限制。

pg_dump生成的转储文件不包含优化器用于决定查询计划的统计信息。因此，明智的做法是在从转储文件恢复后运行`ANALYZE`，以保证良好的性能。

由于pg_dump常被用来把数据转移到更新版本的PostgreSQL，可以期望pg_dump的输出能够装载到比pg_dump自身版本更新的PostgreSQL服务器版本中。pg_dump也可以从比它自身版本更旧的PostgreSQL服务器上转储。（目前支持回到 7.0 版本的服务器。）然而，pg_dump不能从比它自身主版本更新的PostgreSQL服务器上转储；它甚至会拒绝尝试，而不冒生成无效转储的风险。此外，不能保证pg_dump的输出能够装载到更旧主版本的服务器中–即使转储正是从该版本的服务器上取得的。把转储文件装载到较旧的服务器时，可能需要手工编辑转储文件，以移除较旧服务器不理解的语法。

示例

要转储一个数据库：

```
$pg_dump mydb > db.out
```

要重新装载这个数据库：

```
$psql -d database -f db.out
```

要把一个名为 `mydb` 的数据库转储到 `tar` 文件中：

```
$pg_dump -Ft mydb > db.tar
```

要把这个转储重新装载到一个名为 `newdb` 的现有数据库中：

```
$pg_restore -d newdb db.tar
```

历史

pg_dump 工具最早出现在 Postgres95 的 0.02 版中。非纯文本输出格式是在 PostgreSQL 7.1 版中引入的。

参见

pg_dumpall, pg_restore, psql, 环境变量（第 28.11 节）

pg_dumpall

pg_dumpall — 将一个PostgreSQL数据库集簇转储为脚本文件

大纲

`pg_dumpall` [*option*...]

描述

pg_dumpall是一个实用程序，用于写出（“转储”）
一个集簇中的所有PostgreSQL数据库到一个脚本文件中。该脚本文件包含SQL命令，可作为psql的输入来恢复这些数据库。它通过对集簇中的每个数据库调用pg_dump来实现这一点。pg_dumpall还会转储所有数据库共有的全局对象。（pg_dump不会保存这些对象。）
目前包括数据库用户和组、表空间，以及适用于整个数据库的访问权限等属性的信息。

由于pg_dumpall会读取所有数据库中的表，因此通常必须以数据库超级用户身份连接，
才能生成完整的转储。另外，要执行保存的脚本，也需要具备超级用户权限，
这样才能添加用户和组并创建数据库。

SQL 脚本将写入标准输出。应使用 shell 操作符将其重定向到文件。

pg_dumpall需要多次连接到PostgreSQL服务器（每个数据库一次）。如果使用密码认证，
它每次都会提示输入密码。在这种情况下准备一个`~/.pgpass`文件会比较方便。详见第 28. 12 节。

选项

下列命令行选项控制输出的内容和格式。

`-a`
`--data-only`

只转储数据，不转储模式（数据定义）。

`-c`
`--clean`

包含用于在重新创建数据库之前清理（删除）数据库的 SQL 命令。也会添加用于角色和表空间的`DROP`命令。

`-d`
`--inserts`

将数据转储为 `INSERT` 命令（而不是 `COPY`）。这会使恢复非常缓慢；它主要用于生成可以装入非 PostgreSQL 数据库的转储。注意，如果重新安排了列顺序，恢复可能会彻底失败。`-D` 选项更安全，但速度更慢。

`-D`
`--column-inserts`
`--attribute-inserts`

将数据转储为带有显式列名的 `INSERT` 命令（`INSERT INTO table (column, ...) VALUES ...`）。这会使恢复变得非常缓慢；它主要用于生成可以装入非 PostgreSQL 数据库的转储。

`-g`
`--globals-only`

只转储全局对象，不转储数据库。

`-i`
`--ignore-version`

忽略pg_dumpall与数据库服务器之间的版本不匹配。

pg_dumpall可以处理 PostgreSQL 以前版本的数据库，但非常老的版本（目前是 7.0 之前）已不再受支持。如果你需要跳过版本检查，可以使用此选项（如果 pg_dumpall 随后失败了，可别说我没警告过你）。

`-o`
`--oids`

将对象标识符（OID）作为每个表的数据的一部分进行转储。如果应用程序以某种方式引用 OID 列（例如在外键约束中），请使用此选项。否则，不应使用此选项。

`-O`
`--no-owner`

不要输出用于将对象所有权设置为与原始数据库一致的命令。默认情况下，pg_dumpall 会发出 `ALTER OWNER` 或 `SET SESSION AUTHORIZATION` 语句，以设置已创建模式元素的所有权。除非该脚本由超级用户（或拥有脚本中全部对象的同一用户）启动，否则这些语句在运行时会失败。若要创建一个可由任意用户恢复、并让该用户获得所有对象所有权的脚本，请指定 `-O`。

`-s`
`--schema-only`

只转储对象定义（模式），不转储数据。

`-S` *`username`*
`--superuser=`*`username`*

指定在禁用触发器时要使用的超级用户名。只有在使用 `-X disable-triggers` 时才相关。（通常更好的做法是省略此选项，而以超级用户身份运行生成的脚本。）

`-v`
`--verbose`

指定详细模式。这会让 pg_dumpall 把开始/停止时间写入转储文件，并将进度消息输出到标准错误。它还会启用 pg_dump 的详细输出。

`-x`
`--no-privileges`
`--no-acl`

不转储访问权限（grant/revoke 命令）。

`-X disable-dollar-quoting`
`--disable-dollar-quoting`

此选项禁用函数体中的 dollar quoting，并强制改用 SQL 标准字符串语法对它们进行引用。

`-X disable-triggers`
`--disable-triggers`

此选项只在创建只包含数据而不包含模式的转储时才相关。它指示 pg_dumpall 在输出中包含一些命令，以便在重新载入数据时临时禁用目标表上的触发器。
如果这些表上存在不希望在数据重新载入期间触发的引用完整性检查或其他触发器，
请使用此选项。

目前，为 `-X disable-triggers` 输出的这些命令必须由超级用户执行。因此，还应通过 `-S` 指定一个超级用户名，或者更好的做法是确保以超级用户身份运行生成的脚本。

`-X use-set-session-authorization`
`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是用 `ALTER OWNER` 命令来确定对象所有权。这会让转储更符合标准，但根据转储中对象的历史，可能无法正确恢复。

下列命令行选项控制数据库连接参数。

-h *host*

指定运行数据库服务器的机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。默认值取自 `PGHOST` 环境变量（如果已设置）；否则会尝试使用 Unix 域套接字连接。

-p *port*

指定服务器监听连接的 TCP 端口，或本地 Unix 域套接字文件扩展名。默认值取自`PGPORT`环境变量（如果已设置），否则使用编译时默认值。

-U *username*

以给定用户身份连接。

-W

强制提示输入密码。如果服务器要求密码认证，这应会自动发生。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

注解

由于pg_dumpall在内部调用pg_dump，因此某些诊断消息会提到pg_dump。

恢复完成后，最好对每个数据库运行`ANALYZE`，以便优化器获得有用的统计信息。也可以运行`vacuumdb -a -z`来分析所有数据库。

pg_dumpall要求在恢复之前，所有必需的表空间目录都已经存在；否则，
位于非默认位置的数据库在创建时将会失败。

示例

要转储所有数据库：

```
$pg_dumpall > db.out
```

要从此文件重新装载数据库，可以使用：

```
$psql -f db.out postgres
```

（这里连接哪个数据库并不重要，因为 pg_dumpall 创建的脚本文件会包含适当的命令，用于创建并连接到已保存的数据库。）

另见

pg_dump。请在那里查阅可能的错误条件的详细信息。另请参见受支持的环境变量（第 28.11 节）。

pg_restore

pg_restore — 从由 pg_dump 创建的归档文件恢复一个PostgreSQL数据库

大纲

`pg_restore` [*option*...] [*filename*]

描述

pg_restore 是一个用于从pg_dump以非纯文本格式创建的归档中恢复 PostgreSQL 数据库的工具。它会发出必要的命令，将数据库重建为保存时的状态。归档文件还允许 pg_restore 有选择地恢复其中的内容，甚至在恢复前重新排列各项的顺序。归档文件被设计为可跨体系结构移植。

pg_restore 可以以两种模式运行。如果指定了数据库名，pg_restore 就会连接到该数据库，并将归档内容直接恢复到数据库中。否则，它会创建一个脚本，其中包含重建数据库所需的 SQL 命令，并将其写入文件或标准输出。这种脚本输出等价于 pg_dump 的纯文本输出格式。因此，一些控制输出的选项与 pg_dump 的选项相对应。

显然，pg_restore 无法恢复归档文件中不存在的信息。例如，如果归档是使用“将数据转储为 `INSERT` 命令”选项生成的，pg_restore 就不能使用 `COPY` 语句装载数据。

选项

pg_restore accepts the following command line arguments.

filename

指定要恢复的归档文件的位置。如果没有指定，则使用标准输入。

`-a`
`--data-only`

只恢复数据，不恢复模式（数据定义）。

`-c`
`--clean`

在重新创建数据库对象之前清理（删除）它们。

`-C`
`--create`

在恢复之前先创建数据库。（使用此选项时，由 `-d` 指定的数据库仅用于发出初始的 `CREATE DATABASE` 命令。所有数据都会恢复到归档中出现的数据库名中。）

`-d` *dbname*
`--dbname=`*dbname*

连接到数据库 *dbname*，并直接恢复到该数据库中。

`-e`
`--exit-on-error`

一个已弃用、现在会被忽略的选项。

`-f` *filename*
`--file=`*filename*

指定生成脚本的输出文件，或者在与 `-l` 一起使用时指定列表输出文件。默认为标准输出。

`-F` *format*
`--format=`*format*

指定归档的格式。没有必要指定格式，因为pg_restore会自动确定格式。如果指定，它可以是下列之一：

`t`

归档是 `tar` 归档。

`c`

归档采用 pg_dump 的自定义格式。

`-i`
`--ignore-version`

忽略数据库版本检查。

`-I` *index*
`--index=`*index*

仅恢复指定名称的索引的定义。

`-l`
`--list`

列出归档的内容。此操作的输出可用作 `-L` 选项的输入。请注意，如果在 `-l` 中使用 `-n` 或 `-t` 等过滤开关，它们会限制被列出的项目。

`-L` *list-file*
`--use-list=`*list-file*

仅恢复在 *list-file* 中列出的归档元素，并按它们在文件中出现的顺序进行恢复。请注意，如果在 `-L` 中使用 `-n` 或 `-t` 等过滤开关，它们还会进一步限制被恢复的项目。

list-file 通常是通过编辑先前 `-l` 操作的输出来创建的。可以移动或删除行，也可以通过在行首放置分号（`;`）将其注释掉。示例见下文。

`-n` *namespace*
`--schema=`*schema*

仅恢复指定模式中的对象。这可以与 `-t` 选项结合使用，只恢复特定表。

`-O`
`--no-owner`

不输出用于设置对象所有权以匹配原始数据库的命令。默认情况下，pg_restore 会发出 `ALTER OWNER` 或 `SET SESSION AUTHORIZATION` 语句，为所创建的模式元素设置所有权。除非对数据库的初始连接由超级用户发起（或者由脚本中所有对象的同一所有者发起），否则这些语句会失败。使用 `-O` 时，初始连接可以使用任意用户名，并且该用户将拥有所有创建的对象。

`-P` *`function-name(argtype [, ...])`*
`--function=`*`function-name(argtype [, ...])`*

仅恢复指定函数。请务必按转储文件目录中显示的形式准确拼写函数名和参数。

`-R`
`--no-reconnect`

此选项已过时，但为了向后兼容仍然被接受。

`-s`
`--schema-only`

只恢复模式（数据定义），不恢复数据（表内容）。当前的序列值也不会被恢复。
（不要将它与 `--schema` 选项混淆，那里“schema”一词的含义不同。）

`-S` *`username`*
`--superuser=`*`username`*

指定在禁用触发器时要使用的超级用户用户名。仅当使用 `--disable-triggers` 时才相关。

`-t` *`table`*
`--table=`*`table`*

只恢复指定表的定义和/或数据。

`-T` *`trigger`*
`--trigger=`*`trigger`*

仅恢复指定名称的触发器。

`-v`
`--verbose`

指定详细模式。

`-x`
`--no-privileges`
`--no-acl`

阻止恢复访问权限（grant/revoke 命令）。

`-X use-set-session-authorization`
`--use-set-session-authorization`

输出符合 SQL 标准的 `SET SESSION AUTHORIZATION` 命令，而不是使用 `ALTER OWNER` 命令来确定对象所有权。这样会使转储结果更符合标准，但根据转储中对象的历史，可能无法正确恢复。

`-X disable-triggers`
`--disable-triggers`

此选项仅在执行仅数据恢复时才相关。它会指示 pg_restore 在恢复数据期间发出命令，临时禁用目标表上的触发器。
如果你不希望在恢复数据期间触发表上的引用完整性检查或其他触发器，请使用此选项。

目前，为 `--disable-triggers` 发出的命令必须以超级用户身份执行。因此，你还应通过 `-S` 指定超级用户用户名，或者更好的是，直接以 PostgreSQL 超级用户身份运行 pg_restore。

pg_restore还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host=`*`host`*

指定服务器运行所在机器的主机名。如果该值以斜杠开头，则它被用作 Unix 域套接字目录。默认值取自 `PGHOST` 环境变量（如果已设置），否则会尝试使用 Unix 域套接字连接。

`-p` *`port`*
`--port=`*`port`*

指定服务器监听连接所使用的 TCP 端口或本地 Unix 域套接字文件扩展名。默认值取自 `PGPORT` 环境变量（如果已设置），否则使用编译时的默认值。

`-U` *`username`*

以给定用户身份连接

`-W`

强制提示输入密码。如果服务器要求密码认证，这应会自动发生。

环境

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

当使用 `-d` 选项指定直接数据库连接时，pg_restore 会在内部执行 SQL 语句。如果运行 pg_restore 时遇到问题，请确保你能够使用例如psql从该数据库中查询信息。

注解

如果你的安装在 `template1` 数据库中有任何本地添加内容，请务必将 pg_restore 的输出装载到一个真正空的数据库中；否则很可能会因为这些附加对象的重复定义而报错。
要创建一个不带任何本地添加内容的空数据库，应从 `template0` 而不是 `template1` 复制，例如：

```
CREATE DATABASE foo WITH TEMPLATE template0;
```

pg_restore 的局限性详述如下。

- 当把数据恢复到一个预先存在的表中，并且使用了 `--disable-triggers` 选项时，pg_restore 会在插入数据前发出命令，禁用用户表上的触发器，然后在数据插入后再发出命令重新启用它们。如果恢复在中途停止，系统目录可能会处于错误状态。
- pg_restore 不会为单个表恢复大对象。如果归档中包含大对象，那么所有大对象都将被恢复。

关于 pg_dump 的局限性的细节也可参见pg_dump文档。

恢复完成后，最好对每个恢复的表运行ANALYZE，以便优化器获得有用的统计信息。

示例

要把一个名为 mydb 的数据库转储到一个 tar 文件中：

```
$pg_dump -Ft mydb > db.tar
```

要把这个转储重新装载到一个已存在、名为 newdb 的数据库中：

```
$pg_restore -d newdb db.tar
```

要重新排列数据库项的顺序，首先需要转储归档的目录：

```
$pg_restore -l archive.file > archive.list
```

列表文件由一个头部和每个项各占一行的内容组成，例如：

```
;
; Archive created at Fri Jul 28 22:28:36 2000
;     dbname: birds
;     TOC Entries: 74
;     Compression: 0
;     Dump Version: 1.4-0
;     Format: CUSTOM
;
;
; Selected TOC Entries:
;
2; 145344 TABLE species postgres
3; 145344 ACL species
4; 145359 TABLE nt_header postgres
5; 145359 ACL nt_header
6; 145402 TABLE species_records postgres
7; 145402 ACL species_records
8; 145416 TABLE ss_old postgres
9; 145416 ACL ss_old
10; 145433 TABLE map_resolutions postgres
11; 145433 ACL map_resolutions
12; 145443 TABLE hs_old postgres
13; 145443 ACL hs_old
```

分号表示一条注释的开始，而行首的数字表示分配给每个项的内部归档 ID。

文件中的行可以被注释掉、删除并重新排序。例如：

```
10; 145433 TABLE map_resolutions postgres
;2; 145344 TABLE species postgres
;4; 145359 TABLE nt_header postgres
6; 145402 TABLE species_records postgres
;8; 145416 TABLE ss_old postgres
```

可以将这样的文件作为 pg_restore 的输入，这样它就只会按该顺序恢复项 10 和 6：

```
$pg_restore -L archive.list archive.file
```

历史

pg_restore 工具最早出现于 PostgreSQL 7.1。

参见

pg_dump, pg_dumpall, psql, 环境变量（第 28.11 节）

psql

psql — PostgreSQL交互式终端

大纲

`psql` [*option*...] [*dbname* [*username*]]

描述

psql是PostgreSQL的一个基于终端的前端。它使你能够交互式地输入查询，将其发送给PostgreSQL，并查看查询结果。也可以从文件提供输入。此外，
它还提供了若干元命令和多种类似 shell 的特性，以便于编写脚本和自动化执行各种任务。

选项

`-a`
`--echo-all`

在读入时将所有非空输入行打印到标准输出（不适用于交互式行读取）。
这等效于把变量`ECHO`设置为 `all`。

`-A`
`--no-align`

切换到非对齐输出模式（默认输出模式是对齐的）。

`-c` *command*
`--command` *command*

指定psql执行一个命令字符串*command*，然后退出。这在 shell 脚本中很有用。

*command*必须是一个服务器完全可解析的命令字符串（即不包含psql专有的特性）
或者单个反斜线命令。因此不能用这个选项混合SQL和psql元命令。要那样做，
可以把字符串用管道输送到psql中，例如`echo "\x \\ select * from foo;" | psql`。

如果命令字符串包含多条 SQL 命令，它们会在单个事务中处理，除非字符串中包含显式的BEGIN/COMMIT 命令将其分成多个事务。这与把同一字符串送入psql标准输入时的行为不同。

`-d` *dbname*
`--dbname` *dbname*

指定要连接的数据库的名称。这等效于指定*dbname*为命令行上的第一个非选项参数。

`-e`
`--echo-queries`

也把发送到服务器的所有 SQL 命令复制到标准输出。这等效于把变量`ECHO`设置为`queries`。

`-E`
`--echo-hidden`

回显`\d`以及其他反斜线命令生成的实际查询。可以用它来学习psql的内部操作。这等效于在psql中设置变量`ECHO_HIDDEN`。

`-f` *`filename`*
`--file` *`filename`*

使用文件*`filename`*作为命令来源，而不是交互式地读取命令。文件处理完毕后，psql终止。这在很多方面等价于元命令`\i`。

如果*`filename`*是`-`（连字符），则会读取标准输入。

使用这个选项与写成`psql <` *`filename`*有细微差别。通常两种形式都会得到你期望的结果，但使用`-f`可以启用一些有用的特性，例如带行号的错误消息。
使用这个选项也还有一点机会降低启动开销。另一方面，使用 shell 输入重定向的形式在理论上能保证得到与你手工逐行输入时完全相同的输出。

`-F` *`separator`*
`--field-separator` *`separator`*

使用*`separator`*作为非对齐输出的字段分隔符。这等效于`\pset fieldsep`或者`\f`。

`-h` *`hostname`*
`--host` *`hostname`*

指定运行服务器的机器的主机名。如果该值以斜线开头，则它会被用作 Unix 域套接字所在的目录。

`-H`
`--html`

切换到HTML表格输出模式。这等效于`\pset format html`或者`\H`命令。

`-l`
`--list`

列出所有可用的数据库，然后退出。其他非连接选项会被忽略。这与元命令`\list`类似。

`-L` *`filename`*
`--log-file` *`filename`*

除了把所有查询输出写到普通输出目标之外，还写到文件*`filename`*中。

`-n`
`--no-readline`

不要使用 readline 进行行编辑，也不要使用命令历史记录。这有助于在剪切和粘贴时关闭 TAB 补全。

`-o` *`filename`*
`--output` *`filename`*

把所有查询输出放到文件*`filename`*中。这等效于命令`\o`。

`-p` *`port`*
`--port` *`port`*

指定服务器用于监听连接的 TCP 端口或者本地 Unix 域套接字文件扩展名。默认是`PGPORT`环境变量的值，如果没有设置，则默认为编译时指定的端口号（通常是5432）。

`-P` *`assignment`*
`--pset` *`assignment`*

以 `\pset` 的形式指定打印选项。注意，这里必须用一个等号而不是空格来分隔名称和值。例如，要把输出格式设置为 LaTeX，可以写 `-P format=latex`。

`-q`
`--quiet`

指定psql应该安静地工作。默认情况下，它会打印出欢迎消息和各种提示信息。
如果使用了这个选项，以上那些就都不会输出。在使用`-c`选项时，配合这个选项很有用。在psql中，你还可以设置`QUIET`变量来达到同样的效果。

`-R` *`separator`*
`--record-separator` *`separator`*

把*`separator`*用作非对齐输出的记录分隔符。这等效于`\pset recordsep`命令。

`-s`
`--single-step`

运行在单步模式中。这意味着在每个命令被发送给服务器之前都会提示用户，
并允许取消执行。使用这个选项可以调试脚本。

`-S`
`--single-line`

运行在单行模式中，其中换行符会终止一个 SQL 命令，就像分号的作用一样。

> **注意**
>
> 这种模式是为坚持使用它的用户提供的，但并不一定值得推荐。特别是，
> 如果在一行中混合了SQL和元命令，对于没有经验的用户来说，
> 它们的执行顺序未必总是清楚的。

`-t`
`--tuples-only`

关闭打印列名和结果行计数页脚等。这等效于`\t`命令。

`-T` *`table_options`*
`--table-attr` *`table_options`*

指定要放在HTML `table`标签内的选项。详见`\pset`。

`-u`

强制psql在连接数据库之前提示输入用户名和密码。

这个选项已被弃用，因为它在概念上有缺陷。（提示输入非默认用户名，
与因为服务器要求而提示输入密码，实际上是两件不同的事情。）建议你改用`-U`和`-W`选项。

`-U` *`username`*
`--username` *`username`*

作为用户*`username`*而不是默认用户连接到数据库（当然，你必须具有这样做的权限）。

`-v` *`assignment`*
`--set` *`assignment`*
`--variable` *`assignment`*

执行一次变量赋值，和`\set`元命令相似。
注意你必须在命令行上用等号分隔名字和值（如果有）。要取消变量的设置，
去掉等号就行。要把一个变量设为空字符串，使用等号但是去掉值。
这些赋值在启动的非常早期阶段完成，因此为内部目的保留的变量可能会在稍后被覆盖。

`-V`
`--version`

打印psql版本并且退出。

`-W`
`--password`

强制psql在连接数据库之前提示输入密码。

只要服务器请求密码认证，psql就应当自动提示输入密码。然而，
目前的密码请求检测并不完全可靠，因此提供了这个选项来强制提示。
如果没有发出密码提示而服务器又要求密码认证，连接尝试将失败。

即使你后来用元命令`\connect`改变了数据库连接，
这个选项在整个会话期间也将保持有效。

`-x`
`--expanded`

打开扩展表格式模式。这等效于`\x`命令。

`-X`,
`--no-psqlrc`

不读取启动文件（既不读取系统范围的`psqlrc`文件，也不读取用户的`~/.psqlrc`文件）。

`-?`
`--help`

显示有关psql命令行参数的帮助并且退出。

退出状态

如果psql正常结束，它会向 shell 返回 0；如果它自身发生致命错误（例如内存耗尽、找不到文件），则返回 1；如果到服务器的连接发生故障且该会话不是交互式的，则返回 2；如果脚本中发生错误且变量`ON_ERROR_STOP`已设置，则返回 3。

用法

连接到数据库

psql 是常规的 PostgreSQL 客户端应用程序。要连接到数据库，你需要知道目标数据库名称、服务器的主机名和端口号，以及要以哪个数据库用户名连接。psql 可以通过命令行选项 `-d`、`-h`、`-p` 和 `-U` 分别指定这些参数。如果遇到一个不属于任何选项的参数，它将被解释为数据库名（如果数据库名已经给出，则解释为数据库用户名）。并非所有这些选项都是必需的；
它们都有有用的默认值。如果省略主机名，psql 将通过 Unix 域套接字连接到本地主机上的服务器，而在没有 Unix 域套接字的机器上则通过 TCP/IP 连接到 `localhost`。默认端口号在编译时确定。由于数据库服务器使用相同的默认值，因此在大多数情况下不必指定端口。
默认用户名是你的操作系统用户名，默认数据库名也是如此。请注意，
你不能随意以任意数据库用户名连接到任意数据库。
数据库管理员应当已经告知你拥有的访问权限。

当默认值不完全合适时，可以通过把环境变量`PGDATABASE`、`PGHOST`、`PGPORT`和`PGUSER`设置为适当的值来少敲一些键盘（额外的环境变量见第 28.11 节）。另外，准备一个`~/.pgpass`文件也很方便，这样就不必经常手工输入密码。详见第 28.12 节。

如果由于任何原因（例如权限不足、服务器没有在目标主机上运行等）导致连接无法建立，
psql将返回一个错误并且终止。

输入 SQL 命令

在正常操作时，psql会提供一个提示符，该提示符是
psql当前连接到的数据库名称后面跟上字符串`=>`。例如：

```
$ psql testdb
Welcome to psql 8.1.23, the PostgreSQL interactive terminal.

Type:  \copyright for distribution terms
       \h for help with SQL commands
       \? for help with psql commands
       \g or terminate with semicolon to execute query
       \q to quit

testdb=>
```

在提示符下，用户可以输入SQL命令。通常，当遇到表示命令结束的分号时，
输入的内容就会被发送给服务器。行结束并不会终止一条命令，因此为了提高清晰度，
命令可以分布在多行上。如果命令被发送并成功执行，其结果就会显示在屏幕上。

每当执行命令时，psql也会轮询由LISTEN和NOTIFY生成的异步通知事件。

元命令

你输入到psql中的任何以未加引号的反斜线开始的东西都是一个psql元命令，它们由
psql自行处理。这些命令让psql对管理和编写脚本更有用。
元命令常常被称作斜线或者反斜线命令。

psql命令的格式是用反斜线后面直接跟上一个命令动词，然后是一些参数。
参数与命令动词和其他参数之间用任意多个空白字符分隔开。

要在参数中包含空白，可以用单引号将它括起来。要在这样的参数中包含一个单引号，
可以在它前面加一个反斜线。单引号中的内容还会接受类似 C 语言的替换：`\n`（换行）、`\t`（制表符）、`\digits`（八进制）和 `\xdigits`（十六进制）。

如果未加引号的参数以冒号（`:`）开头，它会被当作一个 psql 变量，该变量的值将被用作参数。

用反引号（`` ` ``）包围的参数会被当作传给 shell 的命令行。该命令的输出（去掉末尾的换行符）将被用作参数值。上述转义序列在反引号中同样适用。

有些命令把SQL标识符（例如表名）作为参数。
这些参数遵循SQL的语法规则：未加引号的字母会被强制转换为小写，而双引号（`"`）
可以保护字母不发生大小写转换，并允许在标识符中包含空白。在双引号内，
成对的双引号会在结果名称中折叠成一个双引号。例如，`FOO"BAR"BAZ`会被解释为`foo BARbaz`，而`"A weird"" name"`会变成`A weird" name`。

参数解析会在行尾或遇到另一个未加引号的反斜线时停止。
未加引号的反斜线会被视为新元命令的开始。特殊序列`\`（两个反斜线）表示参数结束，
并继续解析SQL命令（如果还有）。通过这种方式，
SQL命令和psql命令可以自由地混合在同一行中。但无论如何，
元命令的参数都不能延续到下一行。

定义了下列元命令：

`\a`

如果当前表格输出格式是非对齐，则切换为对齐；否则切换为非对齐。
保留此命令是为了向后兼容。更通用的解决方案请参见`\pset`。

`\cd [ directory ]`

将当前工作目录更改为*directory*。如果没有参数，则切换到当前用户的主目录。

> **提示**
>
> 要打印当前工作目录，请使用`\! pwd`。

`\C [ title ]`

设置作为查询结果打印的表的标题，或取消此类标题。该命令等价于`\pset title title`。（此命令的名称源自“caption”，因为它过去只用于设置HTML表的标题。）

`\connect` (or `\c`) `[ dbname [ username ] ]`

建立到一个新数据库和/或以另一个用户名的连接。先前的连接会被关闭。如果*dbname*为`-`，
则假定使用当前的数据库名。

如果省略*username*，则假定使用当前的用户名。

作为一条特殊规则，不带任何参数的`\connect`
将以默认用户连接到默认数据库（就像你启动psql时不带任何参数一样）。

如果连接尝试失败（用户名错误、访问被拒绝等），当且仅当psql处于交互模式时，
才会保留先前的连接。在执行非交互式脚本时，处理会立即停止并报错。

选择这种区别对待，一方面是为了让用户方便地应对输入错误，
另一方面是作为一种安全机制，防止脚本意外地作用于错误的数据库。

`\copy table [ ( column_list ) ] { from | to } { filename | stdin | stdout | pstdin | pstdout } [ with ] [ oids ] [ delimiter [ as ] 'character' ] [ null [ as ] 'string' ] [ csv [ quote [ as ] 'character' ] [ escape [ as ] 'character' ] [ force quote column_list ] [ force not null column_list ] ]`

执行前端（客户端）复制。该操作会运行一个SQL COPY命令，但并不是由服务器读取或写入指定文件，而是由psql读取或写入文件，并在服务器与本地文件系统之间转送数据。
这意味着文件可访问性和权限取决于本地用户，而不是服务器，也不需要 SQL 超级用户权限。

该命令的语法与 SQL 的COPY命令类似。注意，正因为如此，`\copy` 命令适用特殊的解析规则。特别是，变量替换规则和反斜线转义在这里不适用。

`\copy table from stdin | stdout`分别基于命令的输入和输出进行读取/写入。所有行都从发出该命令的同一输入源读取，直到读到仅包含 `\.` 的一行或流到达 EOF 为止。输出会发送到与命令输出相同的位置。若要读写psql的标准输入或输出，请使用 `pstdin` 或 `pstdout`。这个选项适合在 SQL 脚本文件中以内联方式填充表。

> **提示**
>
> 这个操作不像SQL的`COPY`命令那样高效，因为所有数据都必须经过客户端/服务器连接。对于大量数据，SQL命令可能更可取。

`\copyright`

显示PostgreSQL的版权和分发条款。

`\d [ pattern ]`
`\d+ [ pattern ]`

对于每个匹配*pattern*的关系（表、视图、索引或序列），显示其所有列、列的类型、表空间（如果不是默认表空间），以及任何特殊属性，例如`NOT NULL`或默认值（如果有）。关联的索引、约束、规则和触发器也会显示；如果该关系是视图，还会显示视图定义。（“匹配模式”的定义见下文。）

命令形式`\d+`与之相同，只是显示的信息更多：会显示与表各列关联的任何注释，以及表中是否存在 OID。

> **注意**
>
> 如果使用`\d`时没有*pattern*参数，它等同于`\dtvs`，将显示所有表、视图和序列的列表。这纯粹是一种便利措施。

`\da [ pattern ]`

列出所有可用的聚合函数及它们所操作的数据类型。如果指定了*pattern*，则只显示名称匹配该模式的聚合函数。

\db [*pattern*]
\db+ [*pattern*]

列出所有可用的表空间。如果指定了*pattern*，则只显示名称匹配该模式的表空间。
如果在命令名后附加+，每个对象还会与它关联的权限一起列出。

\dc [*pattern*]

列出所有可用的字符集编码之间的转换。如果指定了*pattern*，
则只列出名称匹配该模式的转换。

\dC

列出所有可用的类型转换。

\dd [*pattern*]

显示匹配*pattern*的对象的描述，如果没有给出参数则显示所有可见对象的描述。
但无论哪种情况，都只列出有描述的对象。（“对象”涵盖聚合、函数、操作符、类型、
关系（表、视图、索引、序列、大对象）、规则和触发器。）例如：

```
=> \dd version
                      Object descriptions
   Schema   |  Name   |  Object  |        Description
------------+---------+----------+---------------------------
 pg_catalog | version | function | PostgreSQL version string
(1 row)
```

可以用COMMENT SQL命令为对象创建描述。

\dD [*pattern*]

列出所有可用的域。如果指定了*pattern*，则只显示匹配该模式的域。

\df [*pattern*]
\df+ [*pattern*]

列出可用的函数及其参数和返回类型。如果指定了*pattern*，
则只显示名称匹配该模式的函数。如果使用\df+形式，还会显示每个函数的附加信息，
包括语言和描述。

> **注意**
>
> 要查找接受特定类型的参数或返回特定类型值的函数，请使用分页器的搜索功能浏览\df输出。
>
> 为了减少杂乱，\df不显示数据类型的 I/O 函数。其实现方式是忽略接受或返回cstring类型的函数。

\dg [*pattern*]

列出所有数据库角色。如果指定了*pattern*，则只列出名称匹配该模式的角色。
（此命令现在实际上等价于\du。）

\distvS [*pattern*]

这不是实际的命令名：字母i、s、t、v、S分别代表索引、序列、表、视图和系统表。你可以按任意顺序指定这些字母中的任意个或全部，来获得所有匹配对象的列表。字母 S 把列表限定为系统对象；不带S时，只显示非系统对象。如果在命令名后面追加+，每个对象还会连同它关联的描述（如果有）一起列出。

如果指定了*pattern*，则只列出名称匹配该模式的对象。

\dl

这是\lo_list的别名，用于显示大对象列表。

\dn [*pattern*]
\dn+ [*pattern*]

列出所有可用的模式（命名空间）。如果指定了*pattern*（一个正则表达式），则只列出名称匹配该模式的模式。非本地的临时模式会被隐藏。如果在命令名后面追加+，每个对象还会连同它关联的权限和描述（如果有）一起列出。

\do [*pattern*]

列出可用的操作符及其操作数和返回类型。如果指定了*pattern*，则只列出名称匹配该模式的操作符。

\dp [*pattern*]

列出所有可用的表、视图和序列及其关联的访问权限。如果指定了*pattern*，则只列出名称匹配该模式的表、视图和序列。

GRANT和 REVOKE命令用来设置访问权限。更多信息见GRANT。

\dT [*pattern*]
\dT+ [*pattern*]

列出所有数据类型，或者只列出匹配*pattern*的类型。命令形式\dT+显示额外的信息。

\du [*pattern*]

列出所有数据库角色，或者只列出匹配*pattern*的角色。

\edit (or \e) [*filename*]

如果指定了*filename*，则编辑该文件；编辑器退出后，其内容会被复制回查询缓冲区。如果没有给出参数，则把当前查询缓冲区复制到一个临时文件中，然后以相同的方式编辑。

然后，新的查询缓冲区会按照psql的正常规则被重新解析，整个缓冲区被当作单独一行。（因此你不能用这种方式编写脚本。请使用\i。）这也意味着，如果查询以分号结尾（或者更确切地说包含分号），它会立即被执行。在其他情况下，它只会在查询缓冲区中等待。

提示

psql按顺序搜索环境变量PSQL_EDITOR、EDITOR和VISUAL来寻找要使用的编辑器。如果它们都未设置，Unix 系统上使用vi，Windows 系统上使用notepad.exe。

`\echo` *text* [...]

将参数打印到标准输出，参数之间用一个空格分隔，末尾跟随换行符。这可用于在脚本输出中插入信息。例如：

```
=> \echo `date`
Tue Oct 26 21:40:57 CEST 1999
```

如果第一个参数是未加引号的`-n`，则不输出末尾的换行符。

> **提示**
>
> 如果使用`\o`命令重定向查询输出，可能会想用`\qecho`代替这个命令。

`\encoding` [*encoding*]

设置客户端字符集编码。没有参数时，此命令显示当前编码。

`\f` [*string*]

设置非对齐查询输出的字段分隔符。默认值是竖线（`|`）。另请参见`\pset`，那里介绍了设置输出选项的通用方法。

`\g` [{ *filename* | | *command* }]

把当前查询输入缓冲区发送给服务器，并可选地把查询输出存储到*filename*中，或者把输出通过管道送给执行*command*的单独 Unix shell。不带参数的`\g`实际上等效于分号。带参数的`\g`是`\o`命令的一种“一次性”替代。

`\help` (or `\h`) [*command*]

给出指定SQL命令的语法帮助。如果未指定*command*，则psql将列出所有可用语法帮助的命令。如果*command*是星号（`*`），则显示所有SQL命令的语法帮助。

> **注意**
>
> 为了简化输入，由多个单词组成的命令不需要加引号。因此，可以直接输入**`\help alter table`**。

`\H`

打开HTML查询输出格式。如果HTML格式已经打开，则切换回默认的对齐文本格式。此命令是为兼容性和便利性而保留的；设置其他输出选项的方法见`\pset`。

`\i` *filename*

从文件*filename*中读取输入，并像在键盘上输入一样执行它。

> **注意**
>
> 如果想在屏幕上看到被读入的各行，请将变量`ECHO`设置为`all`。

`\l` (or `\list`)
`\l+` (or `\list+`)

列出服务器中所有数据库的名称、拥有者和字符集编码。如果在命令名后附加 `+`，还会显示数据库的描述。

`\lo_export` *`loid filename`*

从数据库中读取具有OID*`loid`*的大对象，并将其写入*`filename`*。请注意，这与服务器函数`lo_export`略有不同，后者使用运行数据库服务器的用户的权限，
并在服务器的文件系统上操作。

> **提示**
>
> 使用`\lo_list`命令来查找大对象的OID。

`\lo_import` *`filename`* `[` *`comment`* `]`

将文件存储到一个PostgreSQL大对象中。可选地，它将给定的注释与对象关联起来。例如：

```
foo=> \lo_import '/home/peter/pictures/photo.xcf' 'a picture of me'
lo_import 152801
```

响应表明大对象获得了对象 ID 152801，如果将来还想访问该对象，就应当记住这个 ID。因此，建议始终为每个对象关联一条便于人阅读的注释。随后可以用`\lo_list`命令查看这些注释。

请注意，此命令与服务器端的`lo_import`略有不同，
因为它作为本地用户在本地文件系统上操作，而不是服务器的用户和文件系统。

`\lo_list`

列出当前存储在数据库中的所有PostgreSQL大对象，以及为它们提供的注释。

`\lo_unlink` *`loid`*

从数据库中删除OID 为 *`loid`*的大对象。

> **提示**
>
> 使用`\lo_list`命令来查找大对象的OID。

`\o` `[` {*`filename`* | `|` *`command`*} `]`

把以后的查询结果保存到文件*`filename`*中，或通过管道把以后的结果送给单独的 Unix shell 去执行*`command`*。如果未指定参数，查询输出将重置为标准输出。

“查询结果”包括从数据库服务器获得的所有表、命令响应和提示，
以及各种查询数据库的反斜线命令（例如`\d`）的输出，但不包括错误消息。

> **提示**
>
> 要在查询结果之间穿插文本输出，可使用`\qecho`。

`\p`

将当前查询缓冲区打印到标准输出。

`\pset` *`parameter`* `[` *`value`* `]`

这个命令设置影响查询结果表输出的选项。*`parameter`*描述要设置哪个选项。*`value`*的含义取决于该选项。

可调整的打印选项有：

`format`

将输出格式设置为 `unaligned`、`aligned`、`html`、`latex` 或 `troff-ms` 之一。允许使用无歧义的缩写。（这意味着一个字母就足够了。）

“Unaligned” 格式把一行的所有列写在一行上，用当前生效的字段分隔符分隔。这适合创建可能要由其他程序读入的输出（例如制表符分隔、逗号分隔格式）。“Aligned” 模式是默认的标准、人类可读、格式美观的文本输出。“HTML” 和 “LaTeX” 模式生成的表格旨在包含在使用相应标记语言的文档中。它们不是完整的文档！（这在 HTML 中可能不是必需的，但在 LaTeX 中，必须有一个完整文档的外层结构。）

`border`

第二个参数必须是数字。一般来说，数字越大，表格的边框和分隔线就越多，但这取决于具体的格式。在 HTML 模式中，它会直接转换为 `border=...` 属性；在其他模式中，只有值 0（无边框）、1（内部分隔线）和 2（表格外框）有意义。

`expanded` (or `x`)

在常规格式和扩展格式之间切换。启用扩展格式时，查询结果以两列显示，左侧为列名，右侧为数据。如果数据在通常的 “横向” 模式下无法在屏幕上完整显示，这种模式就很有用。

所有四种输出格式都支持扩展模式。

`null`

设置用于代替空值打印的字符串。默认不打印任何内容，这很容易被误认为空字符串。例如，你可能更喜欢使用 `\pset null '(null)'`。

`fieldsep`

指定非对齐输出格式使用的字段分隔符。这样就可以生成例如制表符分隔或逗号分隔的输出，其他程序可能更喜欢这种格式。要把制表符设置为字段分隔符，可以键入 `\pset fieldsep '\t'`。默认的字段分隔符是 `'|'`（竖线）。

`footer`

切换默认页脚 `(x rows)` 的显示。

`numericlocale`

切换是否显示区域感知的字符来分隔小数点左侧的数字组。
它还会启用区域感知的小数点标记。

`recordsep`

指定非对齐输出格式使用的记录（行）分隔符。默认是换行符。

`tuples_only` (or `t`)

在仅元组显示和完整显示之间切换。完整显示包含列标题、
表格标题和各种页脚等附加信息。在仅元组模式下，只显示实际的表格数据。

`title [ text ]`

设置随后打印的任何表格的标题。这可以为输出提供描述性标签。如果没有给出参数，
则取消标题的设置。

`tableattr` (or `T`) `[ text ]`

指定要放在 HTML `table` 标签内部的属性。例如可以是`cellpadding`或`bgcolor`。注意，你可能不想在这里指定`border`，因为这已经由`\pset border`处理。

`pager`

控制查询和 psql 帮助输出是否使用分页器程序。如果设置了环境变量 `PAGER`，输出会通过管道传递给指定的程序。否则，使用与平台有关的默认程序（如 `more`）。

当 pager 为 off 时，不使用分页器。当 pager 为 on 时，只在适当的时候使用分页器，即输出目标是终端且无法在屏幕上完整显示时。（psql 对何时使用分页器的估计并不完美。）`\pset pager` 打开和关闭分页器。分页器也可以设置为`always`，这使分页器总是被使用。

这些不同格式的外观示例可参见示例一节。

提示

有各种用于 `\pset` 的快捷命令。请参见 `\a`、`\C`、`\H`、`\t`、`\T` 和 `\x`。

注意

不带任何参数调用 `\pset` 是一个错误。在将来的版本中，这种情况可能会显示所有打印选项的当前状态。

`\q`

退出psql程序。

`\qecho text [ ... ]`

这个命令与`\echo`相同，只是输出会写入由`\o`设置的查询输出通道。

`\r`

重置（清空）查询缓冲区。

`\s [ filename ]`

打印或保存命令行历史记录到*filename*。如果省略*filename*，
历史记录将被写入标准输出。只有当psql被配置为使用GNU Readline库时，此选项才可用。

`\set [ name [ value [ ... ] ] ]`

把内部变量 *name* 设置为 *value*；如果给出多个值，则设置为所有值的串接。如果没有给出第二个参数，变量只是被设置为没有值。要取消变量的设置，使用 `\unset` 命令。

合法的变量名可以包含字母、数字和下划线。详情见下面的变量。变量名区分大小写。

尽管你可以随意将任何变量设置为任何值，但psql将其中若干变量视为特殊变量。
这些变量在下面有关变量的章节中介绍。

> **注意**
>
> 这个命令与 SQL 命令SET完全无关。

`\t`

切换输出中的列名标题和行数页脚的显示状态。这个命令等价于`\pset tuples_only`，提供它是为了使用方便。

`\T table_options`

指定在HTML输出格式中放在`table`标签内的属性。这个命令等价于`\pset tableattr table_options`。

`\timing`

切换以毫秒为单位显示每条 SQL 语句执行耗时的开关状态。

`\w {filename | |command}`

把当前查询缓冲区输出到文件*filename*，或通过管道送给 Unix 命令*command*。

`\x`

设置或切换扩展表格格式模式。它等价于`\pset expanded`。

`\z [ pattern ]`

产生一份所有表、视图和序列及其关联访问权限的列表。如果指定了*pattern*，
则只列出名称匹配该模式的表、视图和序列。

`GRANT`和`REVOKE`命令用于设置访问权限。更多信息见GRANT。

这是`\dp`的别名（“显示权限”）。

`\! [ command ]`

进入一个单独的 shell，或执行 shell 命令 *command*。参数不会被进一步解释；shell 会原样看到它们。

`\?`

显示有关反斜线命令的帮助信息。

很多`\d`命令都可以用一个*pattern*参数来指定要被显示的对象名称。`*`表示“任意字符序列”，`?`表示“任意单个字符”（这种记号方法与 Unix shell 的文件名模式相当）。高级用户还可以使用字符类等正则表达式记法，例如用`[0-9]`匹配“任意数字”。
要让这些模式匹配字符中的任何一个按字面意思解释，可以用双引号把它括起来。

包含（未加引号的）点号的模式会被解释为模式名称的匹配模式，后接对象名称的匹配模式。
例如，`\dt foo*.bar*`显示模式名以`foo`开头且表名以`bar`开始的所有表。如果没有点号，则该匹配模式只匹配当前模式搜索路径中可见的对象。

每当完全省略*pattern*参数时，`\d`命令会显示当前模式搜索路径中可见的所有对象。
要查看数据库中的所有对象，可使用模式`*.*`。

高级特性

变量

psql 提供了与常见 Unix 命令 shell 相似的变量替换特性。变量就是简单的名称/值对，其中值可以是任意长度的任意字符串。要设置变量，使用 psql 元命令 `\set`：

```
testdb=> \set foo bar
```

把变量 `foo` 设置为值 `bar`。要获取变量的内容，在名称前加上冒号，并把它用作任何斜线命令的参数：

```
testdb=> \echo :foo
bar
```

> **注意**
>
> `\set`的参数服从与其他命令相同的替换规则。因此可以构造有趣的引用，例如`\set :foo 'something'`以及分别得到Perl或者PHP的“软链接”或者“可变变量”。不幸的是（或者幸运的是？），这些构造出来的东西并没有什么用处。在另一方面，`\set bar :foo`是一种很好的拷贝变量的方法。

如果调用 `\set` 时没有第二个参数，该变量会被设置，其值为空字符串。要取消设置（即删除）一个变量，使用命令 `\unset`。

psql 的内部变量名可以由字母、数字和下划线按任意顺序、任意数量组成。其中一些变量会被 psql 特殊对待。它们表示特定的选项设置（运行时可以通过更改该变量的值来改变），或者表示应用的某种状态。尽管你可以把这些变量用于其他目的，但不建议这样做，
因为程序的行为可能会很快变得非常奇怪。按照惯例，
所有被特殊对待的变量都由全大写字母（以及可能的数字和下划线）组成。
为了确保将来最大的兼容性，请避免把这类变量名用于自己的目的。
下面是所有被特殊对待的变量的列表。

AUTOCOMMIT

在被设置为on（默认）时，每一个 SQL 命令在成功完成时会被自动提交。在这种模式中要推迟提交，必须输入一个BEGIN或者START TRANSACTION SQL 命令。当被设置为off或者被取消设置时，在显式发出COMMIT或者END之前，SQL 命令不会被提交。自动提交关闭模式会为你发出一个隐式的BEGIN，这会发生在任何不在一个事务块中且本身既不是BEGIN及其他事务控制命令且不是无法在事务块中执行的命令（例如VACUUM）之前。

注意

在自动提交关闭模式中，必须通过ABORT或者ROLLBACK显式地放弃任何失败的事务。还要记住，如果退出会话时没有提交，则所有的工作都会丢失。

注意

自动提交打开模式是PostgreSQL的传统行为，但是自动提交关闭模式更接近于 SQL 的规范。如果更喜欢自动提交关闭模式，可以在系统级的psqlrc文件或者个人的~/.psqlrc文件中设置它。

DBNAME

当前已连接的数据库名称。每次连接到一个数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

ECHO

如果被设置为all，所有非空输入行会在读入时打印到标准输出（不适用于交互式读取的行）。要在程序开始时选择这种行为，可以使用开关-a。如果被设置为queries，psql会在发送每个查询给服务器时将它们打印到标准输出。选择这种行为的开关是-e。如果未设置，或被设置为上述值以外的其他值，则不会显示任何查询。

ECHO_HIDDEN

当设置了该变量且一个反斜线命令查询数据库时，相应的查询会被先显示。这种特性可以帮助我们学习PostgreSQL的内部并且在自己的程序中提供类似的功能（要在程序开始时选择这种行为，可以使用开关-E）。如果把这个变量设置为值noexec，则对应的查询只会被显示而并不真正被发送给服务器执行。

ENCODING

当前的客户端字符集编码。

HISTCONTROL

如果这个变量被设置为ignorespace，则以一个空格开始的行不会被放入到历史列表中。如果被设置为值ignoredups，则与上一条历史记录相同的行不会被放入。值ignoreboth组合了上述两种值。如果未设置，或被设置为上述值以外的其他值，所有在交互模式中被读入的行都会保存在历史列表中。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`HISTFILE`

用于存储历史记录列表的文件名。默认值是`~/.psql_history`。例如，将以下内容：

```
\set HISTFILE ~/.psql_history- :DBNAME
```

放入 `~/.psqlrc` 会使 psql 为每个数据库维护单独的历史记录。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`HISTSIZE`

存储在命令历史中的命令数量。默认值是 500。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`HOST`

当前连接到的数据库服务器主机。
每次连接到数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

`IGNOREEOF`

如果未设置，向一个psql的交互式会话发送一个EOF字符（通常是**Control+D**）
将会终止应用。如果设置为一个数字值，则会有该数量的EOF字符被忽略，
然后应用才会终止。如果该变量被设置但没有数字值，则默认为 10。

> **注意**
>
> 这个特性是可耻地从Bash抄袭过来的。

`LASTOID`

最后被影响的 OID 的值，这可能会由`INSERT`或者`\lo_import`命令返回。这个变量只保证在下一个SQL命令的结果被显示完之前有效。

`ON_ERROR_ROLLBACK`

当被设置为`on`时，如果事务块中的一个语句产生一个错误，
该错误会被忽略并且该事务会继续。当被设置为`interactive`时，

只在交互式会话中忽略这类错误，而读取脚本文件时则不会忽略错误。当未设置或被设置为off时，事务块中产生错误的一个语句会中止整个事务。
错误回滚模式的工作原理是在事务块的每个命令之前都为你发出一个隐式的SAVEPOINT，然后在该命令失败时回滚到该保存点。

ON_ERROR_STOP

默认情况下，如果非交互式脚本遇到错误（例如一个格式错误的SQL命令或内部元命令），处理会继续进行。这是psql的传统行为，但有时并不理想。如果设置了该变量，脚本处理将在出错后立即终止。如果该脚本是从另一个脚本调用的，那个脚本也会以同样的方式终止。如果最外层的脚本不是从交互式psql会话中调用、而是使用-f选项调用的，psql将返回错误码 3，以区分这种情况与致命错误情况（错误码1）。

PORT

当前连接到的数据库服务器端口。
每次连接到数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

PROMPT1
PROMPT2
PROMPT3

这些变量指定psql发出的提示符的模样。见下文的提示符。

QUIET

这个变量等效于命令行选项-q。在交互模式下可能用处不大。

SINGLELINE

这个变量等效于命令行选项-S。

SINGLESTEP

这个变量等效于命令行选项-s。

USER

当前连接的数据库用户。每次连接到一个数据库时都会设置该变量（包括程序启动时），但可以被取消设置。

VERBOSITY

这个变量可以被设置为值default、verbose或者terse来控制错误报告的详细程度。

SQL 插值

psql 变量的另一个有用特性是可以把它们替换（“插值”）到常规 SQL 语句中。这种操作的语法同样是在变量名前加上冒号（:）：

```
testdb=> \set foo 'my_table'
testdb=> SELECT * FROM :foo;
```

将查询表my_table。变量的值会被按字面拷贝，
因此它甚至可能包含不平衡的引号或反斜线命令。你必须确保把它放在那里是有意义的。
不会对加了引号的SQL实体执行变量插值。

这种机制的一个流行应用是在后续语句中引用最后插入的OID来构建外键场景。
这种机制的另一个可能用途是把一个文件的内容拷贝到一个表列中。
首先把该文件载入到一个变量，然后按上面的做法进行：

```
testdb=> \set content '\'' `cat my_file.txt` '\''
testdb=> INSERT INTO my_table VALUES (:content);
```

这种做法的一个可能问题是`my_file.txt`可能包含单引号。这些单引号需要被转义，
以免在处理第二行时导致语法错误。这可以用`sed`程序来完成：

```
testdb=> \set content '\'' `sed -e "s/'/\\\\\\'/g" < my_file.txt` '\''
```

注意反斜线的正确个数（6 个）！它的工作方式是这样的：psql解析这一行后，会把`sed -e "s/'/\\\'/g" < my_file.txt`传给 shell。shell 会在双引号内做自己的处理，然后带着参数`-e`和`s/'/\\'/g`执行`sed`。`sed`解析它时会把两个反斜线替换成一个，然后做替换。也许在某个时候你会觉得所有 Unix 命令使用相同的转义字符是一件很棒的事。）这里还没有考虑这样的事实：由于SQL文本常量也要经受某些解释，
你可能还必须转义所有的反斜线。在那种情况下，也许更好的做法是在外部先准备好这个文件。

由于冒号可以合法地出现在 SQL 命令中，适用如下规则：字符序列 “:name” 不会被改变，除非 “name” 是当前已设置的某个变量的名称。无论哪种情况，都可以用反斜线对冒号转义，
以避免它被替换。（变量的冒号语法是嵌入式查询语言（例如ECPG）的标准SQL语法。
数组切片和类型转换的冒号语法是PostgreSQL的扩展，因此存在冲突。）

提示符

psql 发出的提示符可以按你的喜好进行定制。`PROMPT1`、`PROMPT2` 和 `PROMPT3` 这三个变量包含描述提示符外观的字符串和特殊转义序列。提示符 1 是 psql 请求新命令时发出的常规提示符。提示符 2 会在录入命令期间还需要更多输入时发出，例如命令尚未以分号结束，或者引号尚未闭合时。在执行 SQL `COPY FROM STDIN` 命令并需要在终端中输入一行值时，会发出提示符 3。

被选中的提示符变量的值会按字面打印，除非遇到百分号（`%`）。根据下一个字符的不同，
会替换为某些其他文本。已定义的替换有：

`%M`

数据库服务器的完整主机名（含域名）；如果通过 Unix 域套接字连接，则为`[local]`；如果 Unix 域套接字不在编译时指定的默认位置，则为`[local:/dir/name]`。

`%m`

数据库服务器的主机名，在第一个点号处截断；如果通过 Unix 域套接字连接，则为`[local]`。

`%>`

数据库服务器监听的端口号。

`%n`

数据库会话用户名。（在数据库会话期间，`SET SESSION AUTHORIZATION`命令可能改变该值的扩展结果。）

`%/`

当前数据库的名称。

`%~`

类似 `%/`，但如果该数据库是你的默认数据库，则输出 `~`（波浪号）。

`%#`

如果会话用户是数据库超级用户，则为`#`，否则为`>`。（在数据库会话期间，`SET SESSION AUTHORIZATION`命令可能改变该值的扩展结果。）

`%R`

在提示符 1 中通常是`=`，但如果处于单行模式则是`^`；如果会话已与数据库断开连接（例如`\connect`失败时可能发生），则是`!`。在提示符 2 中，该序列分别被替换为`-`、`*`、单引号、双引号或美元符，具体取决于psql期待更多输入是因为命令尚未结束、
因为你正处于一个`/* ... */`注释中，还是因为你正处于一个带引号的或用美元符转义的字符串中。在提示符 3 中，该序列不产生任何内容。

`%x`

事务状态：如果当前不在事务块中，则为空字符串；如果处于事务块中，则为 `*`；如果处于失败的事务块中，则为 `!`；如果事务状态不确定（例如因为当前没有连接），则为 `?`。

`%`*`digits`*

替换为指定八进制代码对应的字符。

`%:`*`name`*`:`

psql 变量 *`name`* 的值。详见变量。

`%` ` ` *`command`* ` ` `

`command` 的输出，类似普通的“反引号”替换。

`%[ ... %]`

提示符中可以包含终端控制字符，例如改变提示文本的颜色、背景或样式，
或者改变终端窗口标题。为了让 Readline 的行编辑功能正常工作，这些不可打印的控制字符必须用 `%[` 和 `%]` 包围起来，以标记为不可见字符。提示符中可以出现多组这样的标记。例如：

```
testdb=> \set PROMPT1 '%[%033[1;33;40m%]%n@%/%R%[%033[0m%#%] '
```

其效果是在兼容 VT100 且支持颜色的终端上，生成一个粗体（`1;`）的黑底黄字提示符（`33;40`）。

要在提示符中插入一个百分号，可以写`%%`。默认的提示符是提示符 1 和 2 的`'%/%R%# '`，以及提示符 3 的`'>> '`。

注意

这个特性是可耻地从tcsh抄袭过来的。

命令行编辑

psql 支持 Readline 库，便于编辑和检索输入行。命令历史记录会在 psql 退出时自动保存，并在 psql 启动时重新载入。也支持 Tab 补全，不过其补全逻辑并不声称自己是 SQL 解析器。如果出于某种原因你不喜欢 Tab 补全，可以将以下内容放入主目录下名为 `.inputrc` 的文件中，将其关闭：

```
$if psql
set disable-completion on
$endif
```

（这不是 psql 的功能，而是 Readline 的功能。更多细节请阅读其文档。）

环境

`PAGER`

如果查询结果无法在屏幕上完整显示，则会通过管道传递给这个命令。典型值为 `more` 或 `less`。默认值取决于平台。可以使用 `\pset` 命令禁用分页器的使用。

`PGDATABASE`

默认连接数据库。

`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数。

`PSQL_EDITOR`
`EDITOR`
`VISUAL`

`\e`命令使用的编辑器。这些变量按列出的顺序检查；使用第一个被设置的变量。

`SHELL`

被`\!`命令执行的命令。

`TMPDIR`

存储临时文件的目录。默认是`/tmp`。

文件

- psql 在启动之前会尝试从系统级的 `psqlrc` 文件和用户的 `~/.psqlrc` 文件中读取并执行命令。（在 Windows 上，用户的启动文件名为 `%APPDATA%\postgresql\psqlrc.conf`。）有关设置系统级文件的信息，参见 *PREFIX*`/share/psqlrc.sample`。它可用来按个人喜好设置客户端或服务器（使用 `\set` 和 `SET` 命令）。
- 系统范围的 `psqlrc` 文件和用户的 `~/.psqlrc` 文件都可以通过在文件名后附加连字符和 PostgreSQL 的版本号来使其与版本相关，例如 `~/.psqlrc-8.1.23`。匹配的版本特定文件将优先于非版本特定的文件被读取。

- 命令行历史被存储在文件`~/.psql_history`中，或者是 Windows 的文件`%APPDATA%\postgresql\psql_history`中。

注解

- 在早期版本中，psql 允许一个单字母反斜线命令的第一个参数直接写在该命令后面，中间不需要空白。出于兼容性考虑，这种写法在一定程度上仍然得到支持，
  但我们不打算在这里解释细节，因为这种用法是不鼓励的。如果你收到了奇怪的消息，
  请记住这一点。例如：

  ```
  testdb=> \foo
  Field separator is "oo".
  ```

  这可能不是你所期望的。

- psql 只能与相同版本的服务器顺利配合工作。这并不意味着其他组合会完全失败，但可能出现或微妙或不那么微妙的问题。如果服务器的版本不同，反斜线命令特别容易失败。

Windows 用户的注意事项

psql是一个“控制台应用”。由于 Windows 的控制台窗口使用的是一种和系统中其他应用不同的编码，在psql中使用 8 位字符时要特别注意。如果psql检测到一个有问题的控制台代码页，它将会在启动时警告你。要更改控制台代码页，有两件事是必要的：

- 输入**`cmd.exe /c chcp 1252`**可以设置代码页（1252 是适用于德语的一个代码页，请在这里替换成你的值）。如果正在使用 Cygwin，可以把这个命令放在`/etc/profile`中。

- 把控制台字体设置为“Lucida Console”，因为栅格字体无法与 ANSI 代码页一起使用。

示例

第一个示例展示如何将一个命令分散在多行输入中。请注意提示符的变化：

```
testdb=> CREATE TABLE my_table (
testdb(>  first integer not null default 0,
testdb(>  second text)
testdb-> ;
CREATE TABLE
```

现在再看看表定义：

```
testdb=> \d my_table
             Table "my_table"
 Attribute |  Type   |      Modifier
-----------+---------+--------------------
 first     | integer | not null default 0
 second    | text    |
```

现在我们把提示符改得更有趣一些：

```
testdb=> \set PROMPT1 '%n@%m %~%R%# '
peter@localhost testdb=>
```

假设你已经在表中填入数据，并想查看一下：

```
peter@localhost testdb=> SELECT * FROM my_table;
 first | second
-------+--------
     1 | one
     2 | two
     3 | three
     4 | four
(4 rows)
```

要以不同方式显示表格，可以使用\pset命令：

```
peter@localhost testdb=> \pset border 2
Border style is 2.
peter@localhost testdb=> SELECT * FROM my_table;
+-------+--------+
| first | second |
+-------+--------+
|     1 | one    |
|     2 | two    |
|     3 | three  |
|     4 | four   |
+-------+--------+
(4 rows)

peter@localhost testdb=> \pset border 0
Border style is 0.
peter@localhost testdb=> SELECT * FROM my_table;
first second
----- ------
    1 one
    2 two
    3 three
    4 four
(4 rows)

peter@localhost testdb=> \pset border 1
Border style is 1.
peter@localhost testdb=> \pset format unaligned
Output format is unaligned.
peter@localhost testdb=> \pset fieldsep ","
Field separator is ",".
peter@localhost testdb=> \pset tuples_only
Showing only tuples.
peter@localhost testdb=> SELECT second, first FROM my_table;
one,1
two,2
three,3
four,4
```

也可以使用简短命令：

```
peter@localhost testdb=> \a \t \x
Output format is aligned.
Tuples only is off.
```

```
Expanded display is on.
peter@localhost testdb=> SELECT * FROM my_table;
-[ RECORD 1 ]-
first  | 1
second | one
-[ RECORD 2 ]-
first  | 2
second | two
-[ RECORD 3 ]-
first  | 3
second | three
-[ RECORD 4 ]-
first  | 4
second | four
```

另见

环境变量（第 28.11 节）

reindexdb

reindexdb — 重建一个PostgreSQL数据库的索引

大纲

```
reindexdb [connection-option...] [--table | -t table ] [--index | -i index ] [dbname]
reindexdb [connection-option...] [--all | -a]
reindexdb [connection-option...] [--system | -s] [dbname]
```

描述

reindexdb是用于重建PostgreSQL数据库中索引的工具。

reindexdb 是 SQL 命令REINDEX的一个包装器。通过该工具重建数据库索引与通过其他方法访问服务器来重建索引，在效果上没有区别。

选项

reindexdb接受以下命令行参数：

`-a`
`--all`

重建所有数据库的索引。

`-s`
`--system`

重建数据库的系统目录上的索引。

`-t table`
`--table table`

只对 *table* 重新建立索引。

`-i index`
`--index index`

只重建 *index*。

`[-d] dbname`
`[--dbname] dbname`

指定要重建索引的数据库名称。如果未指定并且没有使用 `-a`（或 `--all`），则从环境变量`PGDATABASE`中读取数据库名称。如果该变量未设置，则使用为连接指定的用户名。

`-e`
`--echo`

回显reindexdb生成并发送给服务器的命令。

`-q`
`--quiet`

不显示响应。

reindexdb还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host` *`host`*

指定服务器所在机器的主机名。如果该值以斜杠开头，则它将被用作 Unix 域套接字的目录。

`-p` *`port`*
`--port` *`port`*

指定服务器监听连接所用的 TCP 端口或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

要作为其身份连接的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

如果遇到困难，请参阅REINDEX和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须运行在目标主机上。此外，
libpq前端库所使用的任何默认连接设置和环境变量也都会生效。

注解

reindexdb 可能需要多次连接到 PostgreSQL 服务器，每次都要求输入密码。在这种情况下，使用 `~/.pgpass` 文件会很方便。更多信息见第 28.12 节。

示例

要重建数据库`test`中的索引：

```
$ reindexdb test
```

要重建名为`abcd`的数据库中表`foo`的索引和索引`bar`：

```
$ reindexdb --table foo --index bar abcd
```

另见

REINDEX, 环境变量（第 28.11 节）

vacuumdb

vacuumdb — 对一个PostgreSQL数据库进行垃圾收集和分析

大纲

`vacuumdb` [*connection-option*...] [[--full] | [-f]] [[--analyze] | [-z]] [[--freeze] | [-F]] [--table | -t *table* [(*column* [,...])]] [*dbname*]
`vacuumdb` [*connection-options*...] [[--all] | [-a]] [[--full] | [-f]] [[--analyze] | [-z]] [[--freeze] | [-F]]

描述

vacuumdb 是用于清理 PostgreSQL 数据库的工具。vacuumdb 还会生成供 PostgreSQL 查询优化器使用的内部统计信息。

vacuumdb是 SQL 命令VACUUM的包装器。通过该工具或通过访问服务器的其他方式来清理和分析数据库，在效果上没有区别。

选项

vacuumdb接受下列命令行参数：

`-a`
`--all`

清理所有数据库。

`[-d] dbname`
`[--dbname] dbname`

指定要清理或分析的数据库名称。如果未指定并且没有使用 `-a`（或`--all`），则从环境变量 `PGDATABASE` 中读取数据库名。如果该变量未设置，则使用连接时指定的用户名。

`-e`
`--echo`

回显 vacuumdb 生成并发送给服务器的命令。

`-f`
`--full`

执行“完全”清理。

`-q`
`--quiet`

不显示响应。

`-t table [ (column [,...]) ]`
`--table table [ (column [,...]) ]`

只清理或分析 *table*。列名只能与 `--analyze` 选项一起指定。

> **提示**
>
> 如果指定了列，你很可能需要在 shell 中转义括号。（见下面的示例。）

`-v`
`--verbose`

在处理过程中输出详细信息。

`-z`
`--analyze`

计算供优化器使用的统计信息。

vacuumdb还接受以下用于连接参数的命令行参数：

`-h` *`host`*
`--host` *`host`*

指定服务器运行所在机器的主机名。如果该值以斜杠开头，则将其用作 Unix 域套接字的目录。

`-p` *`port`*
`--port` *`port`*

指定服务器监听连接所用的 TCP 端口，或本地 Unix 域套接字文件扩展名。

`-U` *`username`*
`--username` *`username`*

连接时使用的用户名。

`-W`
`--password`

强制提示输入密码。

环境

`PGDATABASE`
`PGHOST`
`PGPORT`
`PGUSER`

默认连接参数

诊断

如遇困难，请参见VACUUM和psql中关于潜在问题和错误消息的讨论。
数据库服务器必须在目标主机上运行。此外，libpq 前端库所使用的任何默认连接设置和环境变量也都会生效。

注解

vacuumdb可能需要多次连接到PostgreSQL服务器，每次都要询问密码。在这种情况下，使用`~/.pgpass`文件会比较方便。更多信息见第 28.12 节。

示例

清理数据库 `test`：

```
$ vacuumdb test
```

清理并为优化器分析数据库 `bigdb`：

```
$ vacuumdb --analyze bigdb
```

清理单个表 `foo`，它位于数据库 `xyzzy` 中，并为优化器分析该表的单列 `bar`：

```
$ vacuumdb --analyze --verbose --table 'foo(bar)' xyzzy
```

另见

VACUUM, 环境变量（第 28.11 节）

PostgreSQL 服务器端应用

本部分包含PostgreSQL服务器端应用和支持工具的参考信息。
这些命令只有在数据库服务器所在的主机上运行时才有意义。其他工具程序列在PostgreSQL 客户端应用中。

目录

initdb 942
ipcclean 945
pg_controldata 946
pg_ctl 947
pg_resetxlog 951
postgres 953
postmaster 957

initdb

initdb — 创建一个新的PostgreSQL数据库集簇

大纲

initdb [*option*...] [--pgdata] | [-D]*directory*

描述

initdb创建一个新的PostgreSQL数据库集簇。
数据库集簇是由单个服务器实例管理的一组数据库。

创建数据库集簇包括创建用于存放数据库数据的目录，
生成共享系统目录表（属于整个集簇而不是某个特定数据库的表），以及创建template1
和postgres数据库。以后创建新数据库时，会复制template1数据库中的所有内容。（因此，
安装在template1中的任何东西都会自动复制到以后创建的每个数据库中。）postgres
数据库是一个默认数据库，供用户、工具程序和第三方应用程序使用。

虽然initdb会尝试创建指定的数据目录，但如果所需数据目录的父目录归 root 所有，它可能没有足够的权限。要在这种环境中初始化，可先由 root 创建一个空的数据目录，然后用chown将该目录的所有权赋予数据库用户账户，再用 su切换为该数据库用户来运行initdb。

initdb必须以将拥有服务器进程的用户身份运行，因为服务器需要访问 initdb创建的文件和目录。由于服务器不能以 root 身份运行，因此也绝不能以 root 身份运行initdb。（实际上它会拒绝这样做。）

initdb会初始化数据库集簇的默认区域设置和字符集编码。排序顺序（LC_COLLATE）
和字符集分类（LC_CTYPE，例如大写、小写、数字）对所有数据库固定，不能更改。使用C
或POSIX之外的排序顺序也会带来性能损失。因此，在运行initdb
时选择正确的区域设置很重要。其余区域设置类别可以在以后启动服务器时更改。
所有服务器区域设置值（lc_*）都可以通过SHOW ALL显示。更多细节见第 21.1 节。

字符集编码可以在创建数据库时单独设置。initdb为template1数据库确定编码，
它将作为所有其他数据库的默认值。要更改默认编码，请使用--encoding选项。
更多细节见第 21.2 节。

选项

-A *authmethod*
--auth=*authmethod*

此选项指定 pg_hba.conf 中用于本地用户的认证方法。除非你信任系统上的所有本地用户，否则不要使用 trust。为了便于安装，默认使用 trust。

-D *directory*
--pgdata=*directory*

该选项指定数据库集簇应存放的目录。这是initdb所需的唯一信息，但也可以通过设置PGDATA环境变量来省去显式写出它；这通常更方便，因为数据库服务器（postmaster）之后也可以通过同一变量找到数据目录。

`-E` *encoding*
`--encoding=`*encoding*

选择模板数据库的编码。这也会成为以后创建的任何数据库的默认编码，除非在创建时覆盖它。默认值由区域设置推导而来，如果无法推导，则使用`SQL_ASCII`。PostgreSQL服务器支持的字符集在第 21.2.1 节中介绍。

`--locale=`*locale*

设置数据库集簇的默认区域设置。如果未指定该选项，区域设置将继承自 `initdb`运行时所在的环境。区域设置支持见第 21.1 节。

`--lc-collate=`*locale*
`--lc-ctype=`*locale*
`--lc-messages=`*locale*
`--lc-monetary=`*locale*
`--lc-numeric=`*locale*
`--lc-time=`*locale*

类似于`--locale`，但只在指定的类别中设置区域设置。

`-U` *username*
`--username=`*username*

选择数据库超级用户的用户名。默认值是运行`initdb`的有效用户的名称。超级用户的名称本身并不重要，不过即使操作系统用户名称不同，也可以选择沿用惯常的名称 postgres。

`-W`
`--pwprompt`

使`initdb`提示输入要赋给数据库超级用户的密码。如果不打算使用密码认证，这一点并不重要。否则，在设置密码之前将无法使用密码认证。

`--pwfile=`*filename*

使`initdb`从文件中读取数据库超级用户的密码。文件的第一行会被当作密码。

还可以使用以下较少用到的选项：

`-d`
`--debug`

打印引导后端的调试输出，以及少量普通用户通常不感兴趣的其他消息。引导后端是 `initdb`用来创建系统目录表的程序。该选项会产生大量极其乏味的输出。

`-L` *directory*

指定`initdb`初始化数据库集簇时应到哪里查找其输入文件。通常不需要这样做。若需要显式指定其位置，系统会提示。

`-n`
`--noclean`

默认情况下，如果`initdb`发现某个错误使其无法完整创建数据库集簇，就会删除它在发现无法完成任务之前可能已创建的所有文件。该选项会禁止这种清理，因此对调试有用。

环境

`PGDATA`

指定数据库集簇应存放的目录；可使用`-D`选项覆盖。

参见

postgres, postmaster

ipcclean

ipcclean — 从失败退出的PostgreSQL服务器中移除共享内存和信号量

大纲

```
ipcclean
```

描述

`ipcclean`移除当前用户拥有的所有共享内存段和信号量集。
它用于在PostgreSQL服务器（postmaster）崩溃之后进行清理。注意，
立即重启服务器同样会清理共享内存和信号量，因此这条命令实际上没有什么用处。

只有数据库管理员才应当执行这个程序，因为如果在多用户运行期间执行它，
可能导致古怪的行为（即崩溃）。如果在服务器运行期间执行这条命令，
该服务器分配的共享内存和信号量将被删除，这会给该服务器带来相当严重的后果。

注解

这个脚本只是一个权宜之计，但自它写成以来的许多年里，
还没有人提出过同样有效且可移植的替代方案。由于`postmaster`现在能够自行清理，
`ipcclean`今后不太可能再得到改进。

这个脚本对`ipcs` 工具的输出格式做了一些假设，而这些假设在不同的操作系统上未必成立。因此，它可能无法在你的特定操作系统上工作。使用前先查看一下脚本内容是明智的。

pg_controldata

pg_controldata — 显示 PostgreSQL 数据库集簇的控制信息

大纲

`pg_controldata` [*datadir*]

描述

`pg_controldata`会打印在`initdb`期间初始化的信息，例如系统目录版本。它还会显示有关预写式日志和检查点处理的信息。这些信息是整个集簇范围内的，不针对任何单个数据库。

该工具只能由初始化该集簇的用户运行，因为它需要对数据目录具有读取权限。可以在命令行中指定数据目录，也可以使用环境变量`PGDATA`。

环境

`PGDATA`

默认数据目录位置

pg_ctl

pg_ctl — 启动、停止或重启PostgreSQL服务器

大纲

`pg_ctl` start [-w] [-s] [-D *datadir*] [-l *filename*] [-o *options*] [-p *path*]
`pg_ctl` stop [-W] [-s] [-D *datadir*] [-m [s[mart]] | [f[ast]] | [i[mmediate]]]
`pg_ctl` restart [-w] [-s] [-D *datadir*] [-m [s[mart]] | [f[ast]] | [i[mmediate]]] [-o *options*]
`pg_ctl` reload [-s] [-D *datadir*]
`pg_ctl` status [-D *datadir*]
`pg_ctl` kill [*signal_name*] [*process_id*]
`pg_ctl` register [-N *servicename*] [-U *username*] [-P *password*] [-D *datadir*] [-w] [-o *options*]
`pg_ctl` unregister [-N *servicename*]

描述

pg_ctl是一个实用工具，用于启动、停止或重启PostgreSQL数据库服务器（postmaster），或者显示正在运行服务器的状态。虽然服务器也可以手工启动，但pg_ctl将重定向日志输出、正确地与终端和进程组脱离等任务封装了起来。它还提供了便于实施受控关闭的选项。

在`start`模式中，会启动一个新的服务器。服务器在后台启动，标准输入被连接到`/dev/null`。标准输出和标准错误要么被追加到一个日志文件（如果使用了`-l`选项），要么被重定向到pg_ctl的标准输出（而不是标准错误）。如果没有选择日志文件，建议把pg_ctl的标准输出重定向到一个文件，或者通过管道传给另一个进程（例如像rotatelogs这样的日志轮转程序），否则`postmaster`会把它的输出写到控制终端（从后台），并且不会离开 shell 的进程组。

在`stop`模式中，运行在指定数据目录中的服务器被关闭。可以用`-m`选项选择三种不同的关闭方法：“Smart” 模式等待所有客户端断开连接。这是默认值。“Fast” 模式不等待客户端断开连接。所有活动事务都会被回滚，客户端被强制断开，然后服务器关闭。“Immediate” 模式会不做干净关闭直接中止所有服务器进程。这会导致重启时进行一次恢复运行。

`restart` 模式实际上是先执行停止再执行启动。这允许更改 `postmaster` 的命令行选项。

`reload`模式只是向`postmaster`服务器进程发送一个SIGHUP信号，使其重新读取配置文件（`postgresql.conf`、`pg_hba.conf`等）。这样就可以修改那些无需完全重启服务器即可生效的配置文件选项。

`status` 模式检查在指定数据目录中是否有一个服务器正在运行。如果有，将显示它的 PID 以及调用它时使用的命令行选项。

`kill`模式向指定进程发送信号。这在没有内置kill命令的Microsoft Windows上尤其有用。使用`--help`可查看受支持的信号名称列表。

`register`模式允许你在Microsoft Windows上注册一个系统服务。

`unregister`模式允许你在Microsoft Windows上注销一个先前用`register`命令注册过的系统服务。

选项

`-D` *`datadir`*

指定数据库配置文件所在的文件系统位置。若省略此选项，则使用环境变量`PGDATA`。

`-l` *`filename`*

将服务器日志输出追加到*`filename`*。如果该文件不存在，就会创建它。umask被设置为077，因此默认情况下其他用户无法访问该日志文件。

`-m` *`mode`*

指定关闭模式。*`mode`*可以是`smart`、`fast`或`immediate`，或者是这三者之一的首字母。

`-o` *`options`*

指定要直接传递给`postmaster`命令的选项。

通常应将选项放在单引号或双引号中，以确保它们作为一个整体被传递。

`-p` *`path`*

指定`postmaster`可执行程序的位置。默认情况下，`postmaster`可执行程序取自与`pg_ctl`相同的目录；如果那里没有，则取自硬编码的安装目录。除非采用了某些非常规方式，并收到找不到`postmaster`可执行程序的错误，否则通常不需要使用此选项。

`-s`

只打印错误，不打印信息性消息。

`-w`

等待启动或关闭完成。60 秒后超时。这是关闭操作的默认选项。一次成功的关闭由PID文件的移除来指示。对于启动，一次成功的`psql -l`表示成功。`pg_ctl`会尝试为psql使用正确的端口。如果环境变量`PGPORT`存在，则使用它。否则，它会查看`postgresql.conf`文件中是否设置了一个端口。如果这两者都未被使用，它将使用PostgreSQL编译时的默认端口（默认是 5432）。等待时，`pg_ctl`会根据启动或关闭的成功与否返回准确的退出码。

`-W`

不等待启动或关闭完成。对 start 和 restart 模式来说，这是默认行为。

Windows 选项

`-N` *`servicename`*

要注册的系统服务名称。该名称会同时用作服务名和显示名。

`-P` *`password`*

用于启动该服务的用户密码。

`-U` *`username`*

用于启动该服务的用户名。对于域用户，请使用`DOMAIN\username`格式。

环境

`PGDATA`

默认数据目录位置。

`PGPORT`

Default port for psql（由 -w 选项使用）。

其他的服务器变量见postmaster。

文件

`postmaster.pid`

数据目录中该文件的存在与否，用于帮助pg_ctl确定服务器当前是否正在运行。

`postmaster.opts.default`

如果该文件存在于数据目录中，pg_ctl（在`start`模式下）
会将该文件的内容作为选项传递给`postmaster`命令，除非被`-o`选项覆盖。

`postmaster.opts`

如果该文件存在于数据目录中，pg_ctl（在`restart`模式下）
会将该文件的内容作为选项传递给postmaster，除非被`-o`选项覆盖。
该文件的内容也会在`status`模式下显示出来。

`postgresql.conf`

数据目录中的这个文件会被解析，以便在`start`模式中给出`-w`时，找到与
psql一起使用的正确端口。

注意

等待完全启动并不是一个定义良好的操作，
如果访问控制的设置使得本地客户端不经过手工交互（例如密码认证）就无法连接，
它可能会失败。

示例

启动服务器

要启动服务器：

```
$pg_ctl start
```

一个启动服务器并阻塞到服务器启动完成的例子：

```
$pg_ctl -w start
```

要使用端口 5433 启动服务器，并在未启用`fsync`的情况下运行，可使用：

```
$pg_ctl -o "-F -p 5433" start
```

停止服务器

```
$pg_ctl stop
```

停止服务器。使用 -m 开关可以控制后端如何关闭。

重启服务器

重启服务器几乎等同于先停止服务器再重新启动，只不过pg_ctl
会保存并重用传给前一个运行实例的命令行选项。以最简单的形式重启服务器，可使用：

```
$pg_ctl restart
```

要重启服务器并等待它关闭和重新启动：

```
$pg_ctl -w restart
```

要使用端口 5433 重启服务器，并在重启时禁用fsync，可使用：

```
$pg_ctl -o "-F -p 5433" restart
```

显示服务器状态

下面是pg_ctl状态输出的一个示例：

```
$pg_ctl statuspg_ctl: postmaster is running (pid: 13718)
Command line was:
/usr/local/pgsql/bin/postmaster '-D' '/usr/local/pgsql/data' '-p'
 '5433' '-B' '128'
```

这是在重启模式中会被调用的命令行。

参见

postmaster

pg_resetxlog

pg_resetxlog — 重置一个 PostgreSQL 数据库集簇的事务日志及其他控制信息

大纲

pg_resetxlog [-f] [-n] [-o *oid*] [-x *xid*] [-m *mxid*] [-O *mxoff*] [-l *timelineid*,*fileid*,*seg*] *datadir*

描述

pg_resetxlog 清空事务日志（WAL），并有选择地重置 pg_control 文件中存储的一些其他控制信息。当这些文件损坏时，有时就需要这项功能。它应当只作为最后的手段使用，即当服务器因这种损坏而无法启动时。

运行这条命令之后，应当就能启动服务器了，但请记住，由于部分提交的事务，数据库中可能包含不一致的数据。你应当立即转储数据、运行 initdb 并重新装载。重新装载之后，检查不一致之处并按需修复。

这个工具只能由安装服务器的用户运行，因为它需要对数据目录的读写访问权限。出于安全原因，你必须在命令行上指定数据目录。pg_resetxlog 不使用环境变量 PGDATA。

如果 pg_resetxlog 抱怨说无法确定 pg_control 的有效数据，你可以通过指定 -f（强制）选项迫使它继续执行。在这种情况下，缺失的数据会用貌似合理的值代替。大多数字段都可以期望是匹配的，但下一个 OID、下一个事务 ID、下一个多事务 ID 及偏移量、WAL 起始地址以及数据库区域字段可能需要人工协助。前五项可以用下面讨论的选项来设置。pg_resetxlog自身的环境是它对区域字段的猜测来源；请确保LANG等环境变量与运行 initdb 时的环境相匹配。如果你无法为所有这些字段确定正确的值，仍可以使用 -f，但恢复出的数据库必须受到比平常更多的怀疑：必须立即转储并重新装载。在转储之前不要在数据库中执行任何修改数据的操作，因为任何此类操作都很可能使损坏加剧。

-o、-x、-m、-O 和 -l 选项允许手动设置下一个 OID、下一个事务 ID、下一个多事务 ID、下一个多事务偏移量以及 WAL 起始地址。只有当 pg_resetxlog 无法通过读取 pg_control 确定合适的值时，才需要这些选项。安全值可以按如下方法确定：

- 下一个事务 ID（-x）的安全值可以通过在数据目录下的 pg_clog 目录中查找数值最大的文件名、加一、再乘以 1048576 来确定。注意文件名为十六进制。通常最简单的做法也是用十六进制指定选项值。例如，如果 0011 是 pg_clog 中最大的项，则 -x 0x1200000 可行（五个尾随零提供正确的乘数）。

- 下一个多事务 ID（-m）的安全值可以这样确定：在数据目录下的 pg_multixact/offsets 目录中找出数值最大的文件名，加一，再乘以 65536。如上所述，文件名是十六进制的，所以最简单的做法是用十六进制指定开关值并加上四个零。

- 下一个多事务偏移（-O）的安全值可以这样确定：在数据目录下的 pg_multixact/members 目录中找出数值最大的文件名，加一，再乘以 65536。如上所述，文件名是十六进制的，所以最简单的做法是用十六进制指定开关值并加上四个零。

- WAL 起始地址（-l）应当大于数据目录下 pg_xlog 目录中现存的任何 WAL 段文件名。这些名称也是十六进制的，并且由三部分组成。第一部分是“时间线 ID”，通常应保持不变。第三部分不要选择大于 255（0xFF）的值；应当递增第二部分并把第三部分重置为 0。例如，如果 000000010000000320000004A 是 pg_xlog 中最大的项，-l 0x1,0x32,0x4B 即可；但如果最大的项是 000000010000003A000000FF，请选择 -l 0x1,0x3B,0x0 或更大。

- 没有同样简单的方法来确定一个超出数据库中最大 OID 的下一个 OID，但幸运的是，下一个 OID 设置得是否正确并不关键。

`-n`（无操作）开关指示 `pg_resetxlog` 打印从 `pg_control` 重构出的值，然后不做任何修改直接退出。这主要是一个调试工具，但在真正允许`pg_resetxlog`继续执行之前，
也可用作一次健全性检查。

注解

服务器正在运行时不得使用这条命令。如果 `pg_resetxlog` 在数据目录中发现服务器锁文件，它将拒绝启动。如果服务器崩溃过，可能会留下锁文件；这种情况下你可以删除锁文件以允许 `pg_resetxlog` 运行。但在这样做之前，请务必再三确认没有仍然存活的`postmaster`或任何后端服务器进程。

postgres

postgres — 以单用户模式运行 PostgreSQL 服务器

大纲

postgres [-A [0] | [1]] [-B *nbuffers*] [-c *name*=*value*] [-d *debug-level*] [--describe-config] [-D *datadir*] [-e] [-E] [-f [s] | [i] | [t] | [n] | [m] | [h]] [-F] [-N] [-o *filename*] [-O] [-P] [[-s] | [-t [pa] | [pl] | [ex]]] [-S *work-mem*] [-W *seconds*] [--*name*=*value*] *database*
postgres [-A [0] | [1]] [-B *nbuffers*] [-c *name*=*value*] [-d *debug-level*] [-D *datadir*] [-e] [-f [s] | [i] | [t] | [n] | [m] | [h]] [-F] [-o *filename*] [-O] [-p *database*] [-P] [[-s] | [-t [pa] | [pl] | [ex]]] [-S *work-mem*] [-v *protocol*] [-W *seconds*] [--*name*=*value*]

描述

postgres可执行文件是实际处理查询的 PostgreSQL服务器进程。它通常不直接调用；而是启动一个postmaster多用户服务器。

上面的第二种形式是 postgres被postmaster调用的方式（这只是概念上的，因为postmaster和 postgres实际上是同一个程序）；不应直接以这种方式调用它。第一种形式以交互式单用户模式直接调用服务器。这种模式的主要用途是在initdb执行引导期间。
有时它也用于调试或灾难恢复。

在 shell 中以交互模式调用时，用户可以输入查询，结果将打印到屏幕上，但其形式对开发者比终端用户更有用。但请注意，运行单用户服务器其实并不真正适合调试服务器，
因为不会发生真实的进程间通信和锁定。

运行独立服务器时，会话用户将被设置为 ID 为 1 的用户。该用户实际上不必存在，因此独立服务器可用于手工恢复系统目录遭受某些意外损坏的情况。在独立模式下，会向 ID 为 1 的用户隐式授予超级用户权限。

选项

当postgres由一个postmaster启动时，它会继承后者设置的所有选项。此外，postgres特有的选项可以通过 postmaster的-o开关传递。

通过设置配置文件，你可以免于键入这些选项。详见第 17 章。某些（安全的）选项还可以由连接的客户端以应用相关的方式设置。例如，如果设置了环境变量PGOPTIONS，那么基于libpq的客户端就会将该字符串传给服务器，服务器会将其解释为 postgres命令行选项。

通用选项

选项-A、-B、-c、-d、-D、-F和--*name*的含义与postmaster中的相同，只是 -d 0会阻止postmaster的服务器日志级别传播到postgres。

-e

将默认日期风格设置为“European”，即输入日期字段采用 DMY顺序。这也会在某些日期输出格式中使日期显示为日在月前。更多信息见第 8.5 节。

-o *filename*

将所有服务器日志输出发送到 *filename*。如果postgres运行在 postmaster之下，则忽略此选项，使用从 postmaster继承的stderr。

`-P`

读取系统表时忽略系统索引，但在修改表时仍然更新这些索引。
这在从损坏的系统索引中恢复时很有用。

`-s`

在每条命令结束时打印时间信息和其他统计信息。
这对基准测试或者调整缓冲区数量很有用。

`-S` *`work-mem`*

指定内部排序和散列操作在转而使用临时磁盘文件之前可使用的内存量。有关 `work_mem`配置参数的说明，请参阅第 17.4.1 节。

独立模式的选项

`database`

指定要访问的数据库名称。如果省略，则默认为用户名。

`-E`

回显所有命令。

`-N`

禁用换行符作为语句分隔符。

半内部选项

还有一些其他选项可以指定，主要供调试之用。这里列出这些选项，仅供PostgreSQL系统开发者使用。强烈不建议使用这些选项中的任何一个。此外，这些选项中的任何一个都可能在未来的版本中不经通知而消失或改变。

`-f { s | i | m | n | h }`

禁止使用特定的扫描和连接方法：`s`和`i` 分别禁用顺序扫描和索引扫描，而 `n`、`m`和`h` 分别禁用嵌套循环、归并和哈希连接。

> **注意**
>
> 顺序扫描和嵌套循环连接都无法被完全禁用；`-fs`和 `-fn`选项只是在优化器有其他选择时，尽量不使用这些计划类型。

`-O`

允许修改系统表的结构。该选项由 `initdb`使用。

`-p` *`database`*

表明此进程是由一个 `postmaster`启动的，并指定要使用的数据库。

`-t pa[rser] | pl[anner] | e[xecutor]`

打印与各个主要系统模块相关的每个查询的时间统计信息。该选项不能与`-s`选项一起使用。

-v *protocol*

指定特定会话要使用的前端/后端协议版本号。

-W *seconds*

一旦遇到此选项，进程就会休眠指定的秒数。
这为开发者把调试器附着到服务器进程提供了时间。

--describe-config

该选项会以制表符分隔的COPY格式导出服务器内部配置变量、说明以及默认值。
它主要是为管理工具设计的。

环境

PGDATA

默认数据目录位置

至于在单用户模式下影响不大的其他环境变量，请参见postmaster。

注解

要取消一个正在运行的查询，可向运行该命令的postgres进程发送 SIGINT 信号。

要让postgres重新加载配置文件，可发送SIGHUP信号。通常，最好对postmaster
发送SIGHUP；postmaster会依次对其每个子进程发送SIGHUP。但在某些情况下，
可能只需要一个postgres进程重新加载配置文件。

postmaster使用SIGTERM通知一个postgres进程正常退出，使用 SIGQUIT在不做正常清理
的情况下终止。这些信号不应由用户使用。向一个postgres进程发送SIGKILL
同样是不明智的–postmaster会把这解释为 postgres的一次崩溃，并作为其标准崩溃恢复过
程的一部分，强制所有兄弟 postgres进程退出。

用法

使用类似下面的命令启动独立服务器：

postgres -D /usr/local/pgsql/data *other-options* my_database

使用-D提供正确的数据库目录路径，或者确保已经设置了环境变量PGDATA。
还要指定你想要操作的那个数据库的名称。

通常，独立服务器将换行视为命令输入终止符；它不像psql那样会对分号进行智能处理。
要让一个命令跨越多行，必须在除最后一个换行之外的每个换行前输入反斜线。

但是，如果你使用-N命令行开关，换行并不会终止命令输入。在这种情况下，
服务器会读取标准输入直到文件末尾（EOF）标记，然后把输入作为单个命令字符串处理。
在这种情况下，反斜线-换行不会被特殊处理。

要退出会话，输入EOF（通常是**Control+D**）。如果你使用了-N，则需要连续输入两次
EOF才能退出。

注意，独立服务器不提供复杂的行编辑功能（例如没有命令历史）。

另见

initdb, ipcclean, postmaster

postmaster

postmaster — PostgreSQL多用户数据库服务器

大纲

postmaster [-A [0] | [1]] [-B *nbuffers*] [-c *name*=*value*] [-d *debug-level*] [-D *datadir*] [-F] [-h *hostname*] [-i] [-k *directory*] [-l] [-N *max-connections*] [-o *extra-options*] [-p *port*] [-S] [--*name*=*value*] [[-n] | [-s]]

描述

postmaster是 PostgreSQL多用户数据库服务器。客户端应用要访问数据库时，会（通过网络或本地）连接到一个正在运行的 postmaster。postmaster随后会启动一个独立的服务器进程（“postgres”）来处理该连接。postmaster还管理各服务器进程之间的通信。

默认情况下，postmaster在前台启动，并将日志消息打印到标准错误流。在实际应用中，postmaster 应作为后台进程启动，例如在系统启动时启动。

一个postmaster始终只管理一个数据库集簇的数据。
数据库集簇是一组存储在同一文件系统位置（“数据区域”）中的数据库。
只要使用不同的数据区域和不同的通信端口（见下文），一个系统上就可以同时运行多个postmaster进程。数据区域是用initdb创建的。

当postmaster启动时，它需要知道数据区域的位置。该位置必须通过-D选项或PGDATA环境变量指定；没有默认值。通常，-D或PGDATA会直接指向由initdb创建的数据区域目录。其他可能的文件布局见第 17.2 节。

选项

postmaster接受以下命令行参数。有关这些选项的详细讨论，请参阅第 17 章。通过设置配置文件，你可以省去键入其中大多数选项的麻烦。

-A 0|1

启用运行时断言检查，这是一种帮助检测编程错误的调试辅助手段。只有在编译 PostgreSQL 时启用了断言，此选项才可用；如果是这样，默认值为开。

-B *nbuffers*

设置供服务器进程使用的共享缓冲区数量。该参数的默认值由initdb自动选择；
更多信息参见第 17.4.1 节。

-c *name=value*

设置一个命名的运行时参数。PostgreSQL支持的配置参数见第 17 章。大多数其他命令行选项实际上都是这种参数赋值的简写形式。-c可以出现多次以设置多个参数。

-d *debug-level*

设置调试级别。该值设得越高，写入服务器日志的调试输出就越多。取值范围是 1 到 5。

-D *datadir*

指定数据目录或配置文件的文件系统位置。详见第 17.2 节。

`-F`

禁用`fsync`调用以提高性能，但在系统崩溃时会有数据损坏的风险。指定此选项等效于禁用fsync配置参数。使用前请先阅读详细文档！

`--fsync=true`选项与该选项的效果相反。

`-h` *`hostname`*

指定`postmaster`用来侦听来自客户端应用的 TCP/IP 连接的 IP 主机名或地址。该值也可以是逗号分隔的地址列表，或者用`*`指定侦听所有可用接口。空值表示不侦听任何 IP 地址，在这种情况下只能使用 Unix 域套接字连接到 `postmaster`。默认只侦听 localhost。指定此选项等效于设置listen_addresses配置参数。

`-i`

允许远程客户端通过 TCP/IP（Internet 域）连接。没有此选项时，只接受本地连接。该选项等效于在 `postgresql.conf`中，或通过`-h`将 `listen_addresses`设置为`*`。

该选项已被弃用，因为它无法访问listen_addresses的全部功能。
通常最好直接设置`listen_addresses`。

`-k` *`directory`*

指定 `postmaster` 监听来自客户端应用连接的 Unix 域套接字的目录。默认通常是 `/tmp`，但可以在构建时更改。

`-l`

启用使用SSL的安全连接。若要使用此选项，PostgreSQL必须在编译时启用 SSL支持。关于使用SSL的更多信息，请参阅第 16.7 节。

`-N` *`max-connections`*

设置该`postmaster`将接受的最大客户端连接数。默认值为 32，但可以设置为你系统所支持的最高值。（注意`-B`必须至少是`-N`的两倍。关于大量客户端连接的系统资源需求的讨论，参见第 16.4 节。）指定此选项等效于设置max_connections配置参数。

`-o` *`extra-options`*

在*`extra-options`*中指定的命令行形式选项，会传给由这个 `postmaster`启动的所有服务器进程。相关可能性参见postgres。如果选项字符串中包含空格，整个字符串必须加引号。

`-p` *`port`*

指定`postmaster` 用来侦听客户端应用连接的 TCP/IP 端口或本地 Unix 域套接字文件扩展名。默认值取自`PGPORT`环境变量；如果`PGPORT`未设置，
则默认值就是在编译期间确定的值（通常为 5432）。如果你指定了非默认端口，那么所有客户端应用都必须通过命令行选项或`PGPORT`指定相同的端口。

`-S`

指定`postmaster` 进程应以静默模式启动。也就是说，它将与用户的（控制）终端脱离关联，启动自己的进程组，并把标准输出和标准错误重定向到`/dev/null`。

使用这个开关会丢弃所有日志输出，这很可能不是你想要的效果，
因为它会使问题的排查变得非常困难。在后台启动 `postmaster`的更好方法见下文。

`--silent-mode=false`与该选项的效果相反。

`--name=value`

设置一个命名的运行时参数；这是`-c`的较短形式。

还有两个额外的命令行选项可用于调试导致服务器进程异常终止的问题。在这种情况下，通常的策略是通知所有其他服务器进程必须终止，然后重新初始化共享内存和信号量。这是因为出错的服务器进程可能在终止前破坏了某些共享状态。这些选项指定 `postmaster`在这种情况下采取的替代行为。这两个选项都不用于普通操作。

这些特殊情况选项是：

`-n`

`postmaster` 将不重新初始化共享数据结构。经验丰富的系统程序员随后可以使用调试器检查共享内存和信号量的状态。

`-s`

`postmaster` 将通过发送`SIGSTOP`信号停止所有其他服务器进程，但不使它们终止。这允许系统程序员手工收集所有服务器进程的核心转储。

环境

`PGCLIENTENCODING`

客户端使用的默认字符编码。（客户端可以单独覆盖它。）该值也可以在配置文件中设置。

`PGDATA`

默认数据目录位置

`PGDATESTYLE`

DateStyle运行时参数的默认值。（该环境变量的用法已被弃用。）

`PGPORT`

默认端口（最好在配置文件中设置）

`TZ`

服务器时区

诊断

提到`semget`或 `shmget`的失败消息，很可能表明你需要配置内核，以提供足够的共享内存和信号量。更多讨论见第 16.4 节。

> **提示**
>
> 你也许可以通过降低shared_buffers来减少PostgreSQL的共享内存消耗，以及/或通过降低max_connections来减少信号量消耗，从而推迟重新配置内核。

表明另一个 postmaster 已在运行的失败消息应仔细检查，例如可根据你的系统使用命令

```
$ps ax | grep postmaster
```

或

```
$ps -ef | grep postmaster
```

如果你确信没有冲突的 postmaster 在运行，可以删除消息中提到的锁文件，然后重试。

表明无法绑定端口的失败消息，可能意味着该端口已经被某个非PostgreSQL进程占用。你在终止postmaster 后如果立即使用同一端口重新启动它，也可能会得到这个错误；在这种情况下，你只需等待几秒钟，直到操作系统关闭该端口后再重试。最后，如果你指定了一个操作系统认为是保留的端口号，也可能会收到这个错误。例如，许多版本的 Unix 认为低于 1024 的端口号是“受信任的”，并且只允许 Unix 超级用户访问它们。

注解

只要有可能，就不要使用 SIGKILL杀死 postmaster。这样会阻止 postmaster在终止前释放其持有的系统资源（例如共享内存和信号量）。这可能会导致新的postmaster 运行实例在启动时出现问题。

要正常终止postmaster，可以使用 SIGTERM、SIGINT或 SIGQUIT信号。第一个会在退出前等待所有客户端终止，第二个会强制断开所有客户端，而第三个会在不进行正确关闭的情况下立即退出，从而在重启期间导致一次恢复运行。SIGHUP 信号会重新加载服务器配置文件。

实用命令pg_ctl可用于安全、便捷地启动和关闭postmaster。

--选项在FreeBSD或OpenBSD上不起作用。请改用-c。这是受影响操作系统中的一个缺陷；如果这个问题未被修复，未来版本的PostgreSQL 将提供变通方案。

示例

要使用默认值在后台启动postmaster，输入：

```
$nohup postmaster >logfile 2>&1 </dev/null &
```

要使用特定端口启动postmaster：

```
$postmaster -p 1234
```

这条命令将启动postmaster，通过端口 1234 通信。为了使用psql连接到这个 postmaster，你需要这样运行它：

```
$psql -p 1234
```

或者设置环境变量PGPORT：

```
$export PGPORT=1234$psql
```

命名的运行时参数可以采用以下任一种形式设置：

```
$postmaster -c work_mem=1234$postmaster --work-mem=1234
```

两种形式都会覆盖postgresql.conf中可能存在的 work_mem设置。请注意，参数名中的下划线在命令行中既可以写成下划线，也可以写成连字符。

> **提示**
>
> 除了短期实验之外，与其依赖命令行开关来设置参数，通常更好的做法是直接编辑`postgresql.conf`中的设置。

另见

initdb, pg_ctl

部分 VII. 内部

本部分包含一些可能对PostgreSQL开发人员有用的各类信息。

目录

41. PostgreSQL 内部概述 966
41.1. 查询的路径 966
41.2. 连接是如何建立的 966
41.3. 解析器阶段 967
41.3.1. 解析器 967
41.3.2. 转换过程 967
41.4. PostgreSQL 规则系统 968
41.5. 规划器/优化器 968
41.5.1. 生成可能的计划 968
41.6. 执行器 969
42. 系统目录 971
42.1. 概述 971
42.2. pg_aggregate 972
42.3. pg_am 972
42.4. pg_amop 973
42.5. pg_amproc 974
42.6. pg_attrdef 974
42.7. pg_attribute 974
42.8. pg_authid 976
42.9. pg_auth_members 977
42.10. pg_autovacuum 977
42.11. pg_cast 978
42.12. pg_class 979
42.13. pg_constraint 980
42.14. pg_conversion 981
42.15. pg_database 982
42.16. pg_depend 983
42.17. pg_description 984
42.18. pg_index 984
42.19. pg_inherits 985
42.20. pg_language 986
42.21. pg_largeobject 986
42.22. pg_listener 987
42.23. pg_namespace 987
42.24. pg_opclass 987
42.25. pg_operator 988
42.26. pg_pltemplate 989
42.27. pg_proc 989
42.28. pg_rewrite 991
42.29. pg_shdepend 992
42.30. pg_statistic 993
42.31. pg_tablespace 994
42.32. pg_trigger 994
42.33. pg_type 995
42.34. 系统视图 999
42.35. pg_group 999
42.36. pg_indexes 1000
42.37. pg_locks 1000
42.38. pg_prepared_xacts 1002
42.39. pg_roles 1002

42.40. pg_rules 1003
42.41. pg_settings 1003
42.42. pg_shadow 1004
42.43. pg_stats 1004
42.44. pg_tables 1006
42.45. pg_user 1006
42.46. pg_views 1007
43. 前端/后端协议 1008
43.1. 概述 1008
43.1.1. 消息概述 1008
43.1.2. 扩展查询概述 1009
43.1.3. 格式与格式代码 1009
43.2. 消息流 1009
43.2.1. 启动 1009
43.2.2. 简单查询 1011
43.2.3. 扩展查询 1012
43.2.4. 函数调用 1015
43.2.5. COPY 操作 1016
43.2.6. 异步操作 1016
43.2.7. 取消正在进行的请求 1017
43.2.8. 终止 1018
43.2.9. SSL 会话加密 1018
43.3. 消息数据类型 1018
43.4. 消息格式 1019
43.5. 错误和通知消息的字段 1033
43.6. 协议 2.0 以来的变更总结 1034
44. PostgreSQL 编码约定 1036
44.1. 格式 1036
44.2. 在服务器内部报告错误 1036
44.3. 错误消息风格指南 1038
45. 本地语言支持 1043
45.1. 给翻译者 1043
45.1.1. 要求 1043
45.1.2. 概念 1043
45.1.3. 创建和维护消息目录 1044
45.1.4. 编辑 PO 文件 1045
45.2. 给程序员 1045
45.2.1. 实现机制 1045
45.2.2. 消息编写指南 1046
46. 编写过程语言调用处理器 1048
47. 遗传查询优化器 1050
47.1. 将查询处理看成是一个复杂的优化问题 1050
47.2. 遗传算法 1050
47.3. PostgreSQL 中的遗传查询优化（GEQO） 1051
47.3.1. PostgreSQL GEQO的未来实现任务 1052
47.4. 进一步阅读 1052
48. 索引访问方法接口定义 1053
48.1. 索引的目录条目 1053
48.2. 索引访问方法函数 1054
48.3. 索引扫描 1056
48.4. 索引锁定注意事项 1057
48.5. 索引唯一性检查 1058
48.6. 索引代价估算函数 1058

49. GiST 索引 1061
49.1. 简介 1061
49.2. 可扩展性 1061
49.3. 实现 1061
49.4. 示例 1062
49.5. 崩溃恢复 1062
50. 数据库物理存储 1064
50.1. 数据库文件布局 1064
50.2. TOAST 1065
50.3. 数据库页面布局 1066
51. BKI 后端接口 1069
51.1. BKI 文件格式 1069
51.2. BKI 命令 1069
51.3. 引导 BKI 文件的结构 1070
51.4. 示例 1070
52. 规划器如何使用统计信息 1072
52.1. 行估计示例 1072

第 41 章 PostgreSQL 内部概述

作者

本章最初源自[sim98] Stefan Simkovics 在维也纳技术大学完成的硕士论文，该论文由 O. Univ.Prof.Dr. Georg Gottlob 和 Univ.Ass. Mag. Katrin Seyr 指导。

本章概述 PostgreSQL 后端的内部结构。阅读以下各节之后，你应该能够大致了解一个查询是如何被处理的。本章并不试图详细描述 PostgreSQL 的内部运作，因为那样的文档会非常庞大。相反，本章旨在帮助读者理解，从后端收到一个查询开始，到将结果返回给客户端为止，内部通常有哪些操作步骤。

41.1. 查询的路径

这里将简要概述一个查询为得到结果必须经过的各个阶段。

1. 必须先建立从应用程序到 PostgreSQL 服务器的连接。应用程序将查询发送给服务器，并等待接收服务器返回的结果。

2. 解析器阶段检查应用程序传来的查询是否具有正确的语法，并创建一棵查询树。

3. 重写系统接收由解析器阶段创建的查询树，并查找任何可应用于该查询树的规则（存储在系统目录中）。它执行这些规则体中给出的转换。

   重写系统的一个应用是实现视图。每当对一个视图（即虚拟表）发出查询时，重写系统都会将用户的查询重写成一个改为访问视图定义中给出的基表的查询。

4. 规划器/优化器接收（重写后的）查询树，并创建一个查询计划，将作为执行器的输入。

   它首先生成所有能得到同一结果的可能路径。例如，如果待扫描的某个关系上有一个索引，那么该扫描就有两条路径：一种是简单的顺序扫描，另一种是使用该索引。接着会估算执行每条路径的代价，并选择代价最低的路径。代价最低的路径会被展开成一个完整的计划，供执行器使用。

5. 执行器会递归地遍历计划树，并以计划所表示的方式提取行。执行器在扫描关系时会使用存储系统，执行排序和连接，计算限定条件，最后返回得到的行。

在后续各节中，我们将更详细地介绍上述各项内容，以便更好地理解 PostgreSQL 的内部控制和数据结构。

41.2. 连接是如何建立的

PostgreSQL 使用一种简单的“每用户一个进程”客户端/服务器模型。在这种模型中，一个客户端进程恰好连接到一个服务器进程。由于事先不知道会建立多少个连接，因此必须使用一个主进程，在每次收到连接请求时派生一个新的服务器进程。这个主进程称为 `postmaster`，在指定的 TCP/IP 端口上监听传入连接。每当检测到连接请求时，`postmaster` 进程就会派生一个新的服务器进程，称为`postgres`。服务器任务（`postgres` 进程）之间使用信号量和共享内存通信，以确保并发数据访问期间的数据完整性。

客户端进程可以是任何理解 PostgreSQL 协议（见第 43 章）的程序。很多客户端基于 C 语言库 libpq，但该协议也有若干独立实现，例如 Java 的 JDBC 驱动。

一旦连接建立起来，客户端进程就可以向后端（服务器）发送查询。查询以纯文本方式传输，也就是说，前端（客户端）不进行解析。服务器解析查询，创建一个执行计划，执行该计划，并通过已建立的连接把检索到的行返回给客户端。

41.3. 解析器阶段

解析器阶段由两部分组成：

- 解析器定义在 `gram.y` 和 `scan.l` 中，它使用 Unix 工具 yacc 和 lex 构建。
- 转换过程，它会对解析器返回的数据结构做修改和补充。

41.3.1. 解析器

解析器必须检查查询串（以普通 ASCII 文本形式到达）的语法是否有效。如果语法正确，就构建一棵解析树并返回；否则返回错误。解析器和词法分析器是用著名的 Unix 工具yacc和lex实现的。

词法分析器定义在文件 `scan.l` 中，负责识别标识符、SQL 关键字等。每找到一个关键字或标识符，就会生成一个词元并交给解析器。

解析器定义在文件 `gram.y` 中，由一组语法规则和动作组成，每当某条规则被触发时，对应动作就会执行。动作中的代码（实际上是 C 代码）用于构造语法解析树。

文件 `scan.l` 会被转换成 C 源文件 `scan.c`，所用程序是 lex；而 `gram.y` 则会被转换成 `gram.c`，所用程序是 yacc。这些转换完成之后，就可以使用普通的 C 编译器来构建解析器。绝不要修改这些生成出来的 C 文件，因为下一次调用 lex 或 yacc 时，它们都会被覆盖。

> **注意**
>
> 上述转换和编译通常都是借助 makefiles 自动完成的，这些文件随 PostgreSQL 源代码发行版一同提供。

对 yacc 的详细描述，或者对 `gram.y` 中给出的语法规则的说明，都超出了本文的范围。有许多书籍和文档专门讨论 lex 和 yacc。在开始研究之前，你应该先熟悉 yacc，然后再去看 `gram.y` 中给出的语法，否则你将无法理解其中发生了什么。

41.3.2. 转换过程

解析器阶段仅依据 SQL 语法结构的固定规则创建一棵语法解析树。它不会在系统目录中做任何查找，因此不可能理解所请求操作的详细语义。解析器完成之后，
转换过程会接收解析器返回的树，并进行必要的语义解释，以理解查询引用了哪些表、函数和操作符。为表示这些信息而构建的数据结构称为查询树。

之所以将原始解析与语义分析分开，是因为系统目录查找只能在事务内部进行，
而我们不希望在刚收到查询字符串时就立刻启动事务。
原始解析阶段已经足以识别事务控制命令（`BEGIN`、`ROLLBACK` 等），因此这些命令可以在无需进一步分析的情况下被正确执行。一旦我们知道当前处理的是实际查询（例如 `SELECT` 或 `UPDATE`），如果尚未处于事务中，就可以启动事务。只有在那之后才能调用转换过程。

转换过程创建的查询树在大多数地方的结构都与原始语法解析树相似，但在细节上有许多不同。例如，语法解析树中的一个 `FuncCall` 节点表示某个在语法上看起来像函数调用的东西。它可能

被转换成 `FuncExpr` 节点，也可能被转换成 `Aggref` 节点，这取决于被引用的名称最终被判定为普通函数还是聚合函数。此外，
关于列和表达式结果的实际数据类型的信息也会被加入到查询树中。

41.4. PostgreSQL 规则系统

PostgreSQL 提供了一个强大的规则系统，用于定义视图以及处理有歧义的视图更新。最初，PostgreSQL 的规则系统由两种实现组成：

- 第一种实现使用行级处理，并且深度实现于执行器内部。每当访问到单独一行时，
  规则系统就会被调用。这种实现在 1995 年被移除，当时 Berkeley Postgres 项目的最后一个官方发行版被转换成了 Postgres95。
- 规则系统的第二种实现是一种称为查询重写的技术。
  重写系统是位于解析器阶段与规划器/优化器之间的一个模块。这种技术至今仍在使用。

关于查询重写器，第 34 章中已有相当详细的讨论，因此这里没有必要再展开。我们只指出一点：重写器的输入和输出都是查询树，也就是说，树的表示形式以及语义细节层次都不会发生变化。重写可以被看作某种形式的宏展开。

41.5. 规划器/优化器

规划器/优化器的任务是创建一个最优的执行计划。给定的一个 SQL 查询（因此也就是一棵查询树）实际上可以用多种不同方式执行，而这些方式都会产生同样的结果集。如果在计算上可行，查询优化器将考察这些可能执行计划中的每一种，并最终选择预计运行最快的那个执行计划。

> **注意**
>
> 在某些情况下，考察查询的每一种可能执行方式会耗费过多时间和内存。
> 尤其是在执行涉及大量连接操作的查询时更是如此。
> 为了在合理时间内确定一个合理的（不一定最优的）查询计划，PostgreSQL 会使用遗传查询优化器。

规划器的搜索过程实际上使用一种称为路径的数据结构，它只是计划的简化表示，
仅包含规划器做出决策所需的信息。在确定出代价最低的路径之后，会构建一棵完整的
计划树传递给执行器。这棵树以足够细致的方式表示了期望的执行计划，使执行器能够运行它。
在本节余下部分，我们将忽略路径与计划之间的区别。

41.5.1. 生成可能的计划

规划器/优化器首先为查询中使用的每个单独关系（表）生成扫描计划。
可能的计划由每个关系上可用的索引决定。对一个关系总是可以执行顺序扫描，
因此一定会生成顺序扫描计划。假设某个关系上定义了一个索引（例如一个 B-树索引），并且查询中包含限制条件 `relation.attribute OPR constant`。如果 `relation.attribute` 恰好匹配该 B-树索引的键，而且 `OPR` 是该索引操作符类中列出的操作符之一，那么就会生成另一个使用该 B-树索引扫描该关系的计划。如果还存在其他索引，并且查询中的限制条件恰好匹配某个索引的键，则还会考虑更多计划。

如果查询需要连接两个或更多关系，那么会在找出所有可行的单关系扫描计划之后，
再考虑关系连接计划。规划器/优化器会优先考虑那些在`WHERE`
限定条件中存在相应连接子句的关系对之间的连接（即存在类似`where rel1.attr1=rel2.`

attr2这样的限制）。没有连接子句的连接对只有在别无选择时才会被考虑，也就是说，某个关系与其他任何关系之间都没有可用的连接子句。规划器/优化器会为它所考虑的每一对连接生成所有可能的计划。有以下三种可用的连接策略：

- 嵌套循环连接：左关系中找到的每一行，都会使右关系被扫描一次。这种策略实现起来很容易，但可能非常耗时。（不过，如果右关系可以通过索引扫描，那么这也可能是一种不错的策略。可以把左关系当前行中的值作为右关系索引扫描的键。）
- 归并连接：在连接开始之前，每个关系都会先按照连接属性排序。然后两个关系并行扫描，将匹配的行组合成连接结果行。这种连接方式很有吸引力，因为每个关系只需扫描一次。所需的排序既可以通过显式排序步骤完成，也可以利用连接键上的索引按适当顺序扫描关系来完成。
- 哈希连接：首先扫描右关系，并使用其连接属性作为哈希键将其装入一个哈希表。接着扫描左关系，并将找到的每一行中的相应值作为哈希键，用来在该哈希表中定位匹配的行。

当查询涉及两个以上的关系时，最终结果必须由一棵连接步骤树构建出来，每个连接步骤都有两个输入。规划器会考察不同的可能连接顺序，以找出代价最低的那个。

最终完成的计划树由基表的顺序扫描或索引扫描，再加上所需的嵌套循环、归并或哈希连接节点，以及任何需要的辅助步骤（例如排序节点或聚合函数计算节点）组成。这些计划节点类型中的大多数还具有执行选择（丢弃不满足指定布尔条件的行）和投影（根据给定列值计算派生列集，也就是在需要时计算标量表达式）的额外能力。规划器的职责之一，就是把来自 WHERE 子句的选择条件以及所需输出表达式的计算附加到计划树中最合适的节点上。

41.6. 执行器

执行器接收由规划器/优化器创建的计划，并递归地处理它，以提取所需的行集合。这本质上是一种按需拉取的流水线机制。每当某个计划节点被调用时，它必须再交付一行，或者报告自己已经完成了行的交付。

为了给出一个具体示例，假设顶层节点是一个 MergeJoin 节点。在执行归并之前，必须先取到两行（每个子计划各一行）。因此执行器会递归调用自身去处理这些子计划（从挂接在 lefttree 上的子计划开始）。新的顶层节点，也就是左子计划的顶层节点，假设是一个 Sort 节点，那么又需要通过递归来取得一个输入行。Sort 的子节点可能是一个 SeqScan 节点，它表示真正去读取一个表。执行该节点会使执行器从表中取出一行，并把它返回给调用者节点。Sort 节点会反复调用它的子节点，以取得所有待排序的行。当输入耗尽时（子节点返回的是 NULL 而不是一行），Sort 代码就执行排序，并最终能够返回它的第一条输出行，也就是排序后的第一行。它会把其余行保存起来，以便在后续请求中按排序后的顺序交付这些行。

MergeJoin 节点同样会向它的右子计划请求第一行。然后它比较这两行，看它们是否可以连接；如果可以，就向调用者返回一条连接结果行。在下一次调用时，或者在当前这一对输入无法连接时立即，它会前进到某一个表的下一行或另一个表的下一行（取决于比较结果），并再次检查是否匹配。最终，其中一个子计划会耗尽，MergeJoin 节点就返回 NULL，以表明无法再形成更多连接结果行。

复杂查询可能涉及许多层计划节点，但总体方法是相同的：每个节点在每次被调用时，都会计算并返回它的下一条输出行。每个节点还负责应用规划器分配给它的任何选择或投影表达式。

执行器机制用于求值全部四种基本 SQL 查询类型：SELECT、INSERT、UPDATE 和 DELETE。对于 SELECT，顶层执行器代码只需要把查询计划树返回的每一行发送给客户端。对于 INSERT，每

个返回的行会被插入到为该 INSERT 指定的目标表中。（一个简单的 INSERT ... VALUES 命令会创建一棵极为简单的计划树，它只包含一个计算单条结果行的 Result 节点。但 INSERT ... SELECT 可能需要执行器机制的全部能力。）对于 UPDATE，规划器会安排每个计算得到的行包含所有被更新列的值，再加上原始目标行的 TID（tuple ID，即元组 ID 或行 ID）；执行器顶层会利用这些信息创建一个新的更新后行，并将旧行标记为已删除。对于 DELETE，计划实际返回的唯一列就是 TID，而执行器顶层只需利用该 TID 访问每个目标行并将其标记为已删除。

第 42 章 系统目录

系统目录是关系型数据库管理系统存放模式元数据的地方，
例如表和列的信息以及内部记账信息。PostgreSQL 的系统目录是普通表。你可以删除并重建这些表、增加列、插入和更新数值，并借此把系统严重搞坏。通常，不应手工修改系统目录，
总是有 SQL 命令可用于完成这些操作。（例如，`CREATE DATABASE` 会向 `pg_database` 目录中插入一行 – 并且实际上会在磁盘上创建该数据库。）对于某些特别深奥的操作存在一些例外，
例如增加索引访问方法。

42.1. 概述

表 42.1列出了系统目录。每个目录更详细的文档见后文。

大多数系统目录在创建数据库时都会从模板数据库复制过来，因此之后是数据库专有的。
少数目录在物理上由一个集簇中的所有数据库共享；这些目录会在各自的说明中注明。

表 42.1. 系统目录

目录名	用途
pg_aggregate	聚合函数
pg_am	索引访问方法
pg_amop	访问方法操作符
pg_amproc	访问方法支持函数
pg_attrdef	列默认值
pg_attribute	表列(“属性”)
pg_authid	授权标识符（角色）
pg_auth_members	授权标识符成员关系
pg_autovacuum	逐关系的自动清理配置参数
pg_cast	转换（数据类型转换）
pg_class	表、索引、序列、视图（“关系”）
pg_constraint	检查约束、唯一约束、主键约束、外键约束
pg_conversion	编码转换信息
pg_database	本数据库集簇中的数据库
pg_depend	数据库对象间的依赖
pg_description	数据库对象上的描述或注释
pg_index	索引信息
pg_inherits	表继承层次
pg_language	编写函数的语言
pg_largeobject	大对象的数据页
pg_listener	异步通知
pg_namespace	模式
pg_opclass	访问方法操作符类
pg_operator	操作符
pg_pltemplate	过程语言的模板数据

目录名	用途
pg_proc	函数和过程
pg_rewrite	查询重写规则
pg_shdepend	共享对象上的依赖
pg_statistic	规划器统计
pg_tablespace	本数据库集簇内的表空间
pg_trigger	触发器
pg_type	数据类型

42.2. pg_aggregate

目录pg_aggregate存储关于聚合函数的信息。
聚合函数是对一组值（典型的是每个匹配查询条件的行中的同一个列的值）进行操作的函数，它返回从这些值中计算出的单个值。典型的聚合函数是 sum、count和max。pg_aggregate里的每个项都是一个pg_proc项的扩展。pg_proc项记载该聚合的名字、输入和输出数据类型，以及其他一些和普通函数类似的信息。

表 42.2. pg_aggregate 列

名称	类型	引用	描述
aggfnoid	regproc	pg_proc.oid	聚合函数的pg_proc OID
aggtransfn	regproc	pg_proc.oid	转移函数
aggfinalfn	regproc	pg_proc.oid	最终函数（如果没有就为零）
aggsortop	oid	pg_operator.oid	相关联的排序操作符（如果没有则为0）
aggtranstype	oid	pg_type.oid	聚合函数的内部转移（状态）数据的数据类型
agginitval	text		转移状态的初始值。这是一个文本字段，包含初始值的外部字符串表示形式。如果此字段为空，则转移状态值从空值开始。

新的聚集函数通过CREATE AGGREGATE命令注册。更多关于编写聚合函数以及转移函数的含义等信息请参见第 32.10 节。

42.3. pg_am

目录pg_am存储关于索引访问方法的信息。
系统支持的每种索引访问方法在这个目录中都有一行。此目录的内容在第 48 章中详细讨论。

表 42.3. pg_am 列

名称	类型	引用	描述
amname	name		访问方法的名字
amstrategies	int2		此访问方法的操作符策略数，如果访问方法没有固

名称	类型	引用	描述
			定的操作符策略集合则为零
amsupport	int2		此访问方法的支持例程数
amorderstrategy	int2		如果索引不提供排序顺序则为零，否则为描述该排序顺序的策略操作符的策略编号
amcanunique	bool		访问方法是否支持唯一索引？
amcanmulticol	bool		访问方法是否支持多列索引？
amoptionalkey	bool		访问方法是否支持对第一个索引列没有任何约束的扫描？
amindexnulls	bool		访问方法是否支持多列索引？
amconcurrent	bool		该访问方法是否支持并发更新？
aminsert	regproc	pg_proc.oid	“插入这个元组”函数
ambeginscan	regproc	pg_proc.oid	“（重新）启动索引扫描”函数
amgettuple	regproc	pg_proc.oid	“下一个有效元组”函数，如果没有则为零
amgetmulti	regproc	pg_proc.oid	“取出所有有效元组”函数，如果没有则为零
amrescan	regproc	pg_proc.oid	“（重新）启动索引扫描”函数
amendscan	regproc	pg_proc.oid	“（重新）启动索引扫描”函数
ammarkpos	regproc	pg_proc.oid	“标记当前扫描位置”函数
amrestrpos	regproc	pg_proc.oid	“恢复被标记的扫描位置”函数
ambuild	regproc	pg_proc.oid	“构建新索引”函数
ambulkdelete	regproc	pg_proc.oid	批量删除函数
amvacuumcleanup	regproc	pg_proc.oid	VACUUM之后的清理函数
amcostestimate	regproc	pg_proc.oid	估计索引扫描开销的函数

42.4. pg_amop

目录pg_amop存储与索引访问方法操作符类相关的操作符信息。
每个作为操作符类成员的操作符对应一行。

表 42.4. pg_amop列

名称	类型	引用	描述
amopclaid	oid	pg_opclass.oid	此条目所属的索引操作符类

名称	类型	引用	描述
amopsubtype	oid	pg_type.oid	用于区分一个策略的多个条目的子类型；默认值为零
amopstrategy	int2		操作符策略编号
amopreqcheck	bool		索引命中必须重新检查
amopopr	oid	pg_operator.oid	操作符的 OID

42.5. pg_amproc

目录pg_amproc存储与索引访问方法操作符类相关的支持过程信息。
每个属于某个操作符类的支持过程对应一行。

表 42.5. pg_amproc列

名称	类型	引用	描述
amopclaid	oid	pg_opclass.oid	此条目所属的索引操作符类
amprocsubtype	oid	pg_type.oid	如果是跨类型例程则为子类型，否则为零
amprocnum	int2		支持过程编号
amproc	regproc	pg_proc.oid	过程的 OID

42.6. pg_attrdef

目录pg_attrdef存储列默认值。列的主要信息存储在pg_attribute中（见下文）。
只有在创建表或添加列时显式指定默认值的列，才会在这里有一项。

表 42.6. pg_attrdef 列

名称	类型	引用	描述
adrelid	oid	pg_class.oid	该列所属的表
adnum	int2	pg_attribute.attnum	列的编号
adbin	text		列默认值的内部表示
adsrc	text		适合人阅读的默认值表示

adsrc字段是历史遗留，最好不要使用，因为它不会跟踪可能影响默认值表示的外部变化。
显示默认值时，更好的方法是反向编译adbin字段（例如使用pg_get_expr）。

42.7. pg_attribute

目录pg_attribute存储关于表列的信息。数据库中每个表的每一列，在pg_attribute
中都恰好有一行。（其中也包括索引的属性项，实际上，凡是在pg_class中有项的对象，
都有对应的属性项。）

术语属性等同于列，这里使用它只是出于历史原因。

表 42.7. pg_attribute 列

名称	类型	引用	描述
attrelid	oid	pg_class.oid	此列所属的表
attname	name		列名称
atttypid	oid	pg_type.oid	此列的数据类型
attstattarget	int4		attstattarget 控制ANALYZE为该列收集统计信息时的细节级别。零值表示不应收集统计信息。负值表示使用系统默认统计目标。正值的确切含义依赖于数据类型。对于标量数据类型，attstattarget 既是要收集的“高频值”目标个数，也是要创建的直方图桶目标个数。
attlen	int2		该列数据类型的pg_type.typlen副本
attnum	int2		列的编号。一般列从1开始向上编号。系统列，如oid，则拥有（任意）负值编号。
attndims	int4		如果该列是数组类型，则为维数；否则为 0。（目前并不会强制检查数组维数，因此任何非零值实际上都只意味着“这是一个数组”。）
attcacheoff	int4		在存储中总是 -1，但加载到内存中的行描述符后，此值可能被更新，用于缓存该属性在行内的偏移量
atttypmod	int4		atttypmod记录了在表创建时提供的类型相关数据（例如一个varchar列的最大长度）。它会被传递给类型相关的输入函数和长度强制转换函数。对于那些不需要atttypmod的类型，这个值通常为 -1。
attbyval	bool		该列类型的pg_type.typbyval的一个拷贝
attstorage	char		通常是该列类型的pg_type.typstorage的一个拷贝。对于可TOAST的数据类型，这可以在列创建后被修改以控制存储策略。
attalign	char		该列类型的 pg_type.typalign 的一个拷贝

名称	类型	引用	描述
attnotnull	bool		这代表一个非空约束
atthasdef	bool		该列有一个默认值，在此情况下pg_attrdef目录中会有一个对应项来真正定义该值。
attisdropped	bool		该列被删除且不再有效。一个删除的列仍然物理存在于表中，但是会被解析器忽略并因此无法通过SQL访问。
attislocal	bool		该列是由关系本地定义的。注意一个列可以同时是本地定义和继承的。
attinhcount	int4		该列直接祖先的数量。祖先数量非零的列不能被删除，也不能被重命名。

在被删除列的 pg_attribute 条目中，atttypid 被重置为零，但 attlen 以及其他从 pg_type 复制的字段仍然有效。这种安排用于应对被删除列的数据类型后来也被删除、因而不再有相应 pg_type 行的情况。attlen 和其他字段可用于解释表中一行的内容。

42.8. pg_authid

目录pg_authid包含关于数据库授权标识符（角色）的信息。角色涵盖了“用户”和“组”这两个概念。用户本质上只是设置了rolcanlogin标志的角色。
任何角色（无论是否设置了rolcanlogin）都可以拥有其他角色作为成员；参见pg_auth_members。

由于这个目录包含密码，它不应公开可读。pg_roles是建立在pg_authid之上的公开可读视图，它会隐藏密码字段。

第 18 章包含关于用户和权限管理的详细信息。

由于用户标识符是集簇范围的，pg_authid
在一个集簇的所有数据库之间共享：在一个集簇中只有一份pg_authid拷贝，
而不是每个数据库一份。

表 42.8. pg_authid 列

名称	类型	描述	引用
rolname	name		角色名
rolsuper	bool		角色有超级用户权限
rolinherit	bool		该角色是否会自动继承其所属其他角色的权限
rolcreaterole	bool		角色能创建更多角色
rolcreatedb	bool		角色能创建数据库
rolcatupdate	bool		角色能否直接更新系统目录。（即使是超级用户，也只有该列为真时才能这样做）

名称	类型	描述	引用
rolcanlogin	bool		角色是否能登录。即该角色是否能够作为初始会话授权标识符
rolconnlimit	int4		对于可以登录的角色，本列设置该角色能够建立的最大并发连接数。-1 表示无限制。
rolpassword	text		密码（可能被加密）；如果没有则为空
rolvaliduntil	timestamptz		密码过期时间（仅用于密码认证）；不设过期时间时为空值
rolconfig	text[]		运行时配置变量的会话默认值

42.9. pg_auth_members

目录pg_auth_members展示了角色之间的成员关系。允许任何无环的关系集合。

由于用户标识符是集簇范围的，pg_auth_members
在一个集簇的所有数据库之间共享：在一个集簇中只有一份pg_auth_members拷贝，
而不是每个数据库一份。

表 42.9. pg_auth_members 列

名称	类型	引用	描述
roleid	oid	pg_authid.oid	拥有成员的角色的ID
member	oid	pg_authid.oid	roleid的成员角色的ID
grantor	oid	pg_authid.oid	授权此成员关系的角色的ID
admin_option	bool		如果member能够将roleid的成员资格授予其他角色，则为真

42.10. pg_autovacuum

目录pg_autovacuum存储自动清理守护进程可选的逐表配置参数。如果某个表在这里有条目，
自动清理该表时将使用给定的参数。如果没有条目，则使用系统级默认值。

表 42.10. pg_autovacuum列

名称	类型	引用	描述
vacrelid	oid	pg_class.oid	此条目所属的表
enabled	bool		如果为假，此表永远不会被自动清理
vac_base_thresh	integer		触发清理前已修改元组的最小数量
vac_scale_factor	float4		与 reltuples 相乘后加到vac_base_thresh上的乘数

名称	类型	引用	描述
anl_base_thresh	integer		触发分析前已修改元组的最小数量
anl_scale_factor	float4		与 reltuples 相乘后加到anl_base_thresh上的乘数
vac_cost_delay	integer		自定义的vacuum_cost_delay参数
vac_cost_limit	integer		自定义的vacuum_cost_limit参数

当某个表已更新或已删除的元组数超过vac_base_thresh加上vac_scale_factor乘以该表当前估计的活元组数时，autovacuum 守护进程将对这个表发起一次VACUUM操作。类似地，当已插入、已更新或已删除的元组数超过anl_base_thresh加上anl_scale_factor乘以该表当前估计的活元组数时，它将发起一次ANALYZE操作。

任何数值字段都可以填-1（实际上任何负值都行），表示这个特定值使用系统级默认值。注意，vac_cost_delay变量的默认值继承自配置参数autovacuum_vacuum_cost_delay；如果前者设为负值，则继承自vacuum_cost_delay。vac_cost_limit同理。

42.11. pg_cast

目录pg_cast存储数据类型转换路径，包括内置路径以及使用CREATE CAST定义的路径。

表 42.11. pg_cast列

名称	类型	引用	描述
castsource	oid	pg_type.oid	源数据类型的OID
casttarget	oid	pg_type.oid	目标数据类型的OID
castfunc	oid	pg_proc.oid	执行该转换的函数的OID。如果该转换方法不需要一个函数则存储0。
castcontext	char		指示该转换能被调用的环境。e表示仅能作为一个显式转换（使用CAST或::语法）。a表示在赋值给目标列时隐式调用，和显式调用一样。i表示在表达式中隐式调用，和其他转换一样。

在pg_cast里列出的类型转换函数必须总是以转换的源类型作为它的第一个参数类型，并且返回转换的目标类型作为它的结果类型。一个类型转换函数最多有三个参数。如果出现了第二个参数，必须是integer类型；它接受与目标类型关联的修饰词，如果没有，就是-1。如果出现了第三个参数，那么必须是boolean类型；如果该类型转换是一种明确的转换，那么它接受true，否则接受false。

在pg_cast里创建一条源类型和目标类型相同的记录是合理的，只要相关联的函数接受多过一个参数。这样的记录代表“长度转换函数”，它们把该类型的值转换为对特定的类型合法的值。

如果一个pg_cast的项有着不同的原类型和目标类型，并且有一个接收多于一个参数的函数，那么它会在一个步骤中完成从一种类型到另外一种类型的转换并应用一个长度转换。

如果没有这样的项，使用一个类型修改器的转换涉及两个步骤，一个是在数据类型之间转换，另外一个是应用修改器。

42.12. pg_class

目录pg_class记录表，以及几乎所有其他具有列或在其他方面与表类似的对象。这包括索引（但请参见pg_index）、序列、视图、复合类型和TOAST表；请参见relkind。在下面，当我们指的是所有这些类型的对象时，我们称之为“关系（relations）”。并非所有列对所有关系类型都有意义。

表 42.12. pg_class 列

名称	类型	引用	描述
relname	name		表、索引、视图等的名字
relnamespace	oid	pg_namespace.oid	包含该关系的名字空间的OID
reltype	oid	pg_type.oid	与此表的行类型对应的数据类型的 OID（如有）；索引没有pg_type项，此值为零
relowner	oid	pg_authid.oid	关系的拥有者
relam	oid	pg_am.oid	如果这是索引，则为使用的访问方法（B-树、hash 等）
relfilenode	oid		该关系的磁盘文件名；如果没有则为 0
reltablespace	oid	pg_tablespace.oid	存储此关系的表空间。如果为零，则暗示使用数据库的默认表空间。（如果关系没有磁盘文件，则没有意义。）
relpages	int4		该表磁盘表示的尺寸，以页面计（页面尺寸为BLCKSZ）。这只是一个由规划器使用的估计值。它被VACUUM、ANALYZE以及一些DDL命令（如CREATE INDEX）所更新。
reltuples	float4		表中的行数。这只是一个由规划器使用的估计值。它被VACUUM、ANALYZE以及一些DDL命令（如CREATE INDEX）所更新。
reltoastrelid	oid	pg_class.oid	与该表相关联的TOAST表的OID，如果没有则为零。TOAST表将大属性“线外”存储在一个二级表中。
reltoastidxid	oid	pg_class.oid	该约束所在的表，如果不是表约束则为零

名称	类型	引用	描述
relhasindex	bool		如果这是一个表并且其上建有（或最近建有）索引则为真。它由CREATE INDEX设置，但不会立即被DROP INDEX清除。VACUUM会在发现表没有索引时清除relhasindex
relisshared	bool		如果该表在集簇中的所有数据库间共享则为真。只有某些系统目录（如pg_database）是共享的。
relkind	char		r = 普通表，i = 索引，S = 序列，v = 视图，c = 复合类型，s = 特殊，t = TOAST 表
relnatts	int2		关系中用户列的数量（不计系统列）。在pg_attribute中必须有这么多对应项。另请参阅pg_attribute.attnum。
relchecks	int2		表上检查约束的数目，参见pg_constraint目录
reltriggers	int2		表上的触发器数量；见pg_trigger目录
relukeys	int2		未使用（并非唯一键的数量）
relfkeys	int2		未使用（并非表上外键的数量）
relrefs	int2		未使用
relhasoids	bool		如果为此关系的每一行生成 OID，则为真
relhaspkey	bool		如果表有（或曾经有过）主键，则为真
relhasrules	bool		如果表有（或曾有）规则则为真，参见pg_rewrite目录
relhassubclass	bool		如果表有（或曾经有过）任何继承子类，则为真
relacl	aclitem[]		访问权限，详见GRANT和REVOKE

42.13. pg_constraint

目录pg_constraint存储表上的检查约束、主键约束、唯一约束和外键约束。（列约束不会被特殊对待。每一个列约束都将被视为一个表约束。）非空约束由pg_attribute目录表示。

域上的检查约束也存储在这里。

表 42.13. pg_constraint 列

名称	类型	引用	描述
conname	name		约束名字（不需要唯一！）
connamespace	oid	pg_namespace.oid	包含此约束的名字空间的OID
contype	char		c = 检查约束，f = 外键约束，p = 主键约束，u = 唯一约束
condeferrable	bool		该约束是否能被延迟？
condeferred	bool		该约束是否默认被延迟？
conrelid	oid	pg_class.oid	该约束所在的表，如果不是表约束则为零
contypid	oid	pg_type.oid	该约束所在的域，如果不是域约束则为0
confrelid	oid	pg_class.oid	如果此约束是一个外键约束，此列为被引用的表，否则为零
confupdtype	char		外键更新动作代码
confdeltype	char		外键删除动作代码
confmatchtype	char		外键匹配类型
conkey	int2[]	pg_attribute.attnum	如果是一个表约束（包括外键但不包括约束触发器），此列是被约束列的列表
confkey	int2[]	pg_attribute.attnum	如果是一个外键，此列是被引用列的列表
conbin	text		如果是一个检查约束，此列是表达式的一个内部表示
consrc	text		如果是检查约束，则为适合人阅读的表达式表示

注意

被引用对象发生变化时，consrc不会更新；例如，它不会跟踪列重命名。最好使用pg_get_constraintdef()提取检查约束的定义，而不是依赖这个字段。

注意

pg_class.relchecks需要与该表中为每个关系找到的检查约束条目数保持一致。

42.14. pg_conversion

目录pg_conversion描述编码转换函数。更多信息参见CREATE CONVERSION。

表 42.14. pg_conversion 列

名称	类型	引用	描述
conname	name		转换的名字（在一个名字空间内唯一）
connamespace	oid	pg_namespace.oid	包含此转换的名字空间的OID
conowner	oid	pg_authid.oid	转换的拥有者
conforencoding	int4		源编码ID
contoencoding	int4		目标编码ID
conproc	regproc	pg_proc.oid	转换函数
condefault	bool		如果这是默认转换则为真

42.15. pg_database

目录pg_database存储有关可用数据库的信息。数据库通过CREATE DATABASE命令创建。更多关于其参数的信息请查阅第 19 章。

和大部分系统目录不同，pg_database
是在集簇的所有数据库之间共享的：在一个集簇中只有一份pg_database拷贝，
而不是每个数据库一份。

表 42.15. pg_database 列

名称	类型	引用	描述
datname	name		数据库名称
datdba	oid	pg_authid.oid	数据库的拥有者，通常是创建它的用户
encoding	int4		数据库的字符编码
datistemplate	bool		如果为真，则这个数据库可以被用在CREATE DATABASE的TEMPLATE子句中，来创建一个作为该数据库克隆的新数据库
datallowconn	bool		如果为假则没有人能连接到这个数据库。这可以用来保护template0数据库不被修改。
datconnlimit	int4		设置可以连接到此数据库的最大并发连接数。-1表示没有限制。
datlastsysoid	oid		数据库中最后一个系统OID；对 pg_dump 尤其有用
datvacuumxid	xid		由该事务 ID 之前的事务 ID 插入或删除的所有行，都已在此数据库中被标记为已知已提交或已知已中止。这用于确定何时可以回收提交日志空间。

名称	类型	引用	描述
datfrozenxid	xid		在此事务 ID 之前的所有事务 ID 插入的行，都已经在该数据库中被重新标记为永久的（“冻结的”）事务 ID。这可用于检查是否需要尽快清理数据库，以避免事务 ID 回卷问题。
dattablespace	oid	pg_tablespace.oid	此数据库的默认表空间。在此数据库中，所有pg_class.reltablespace为0的表都将被存储在这个表空间中；尤其是非共享系统目录都会在其中。
datconfig	text[]		运行时配置变量的会话默认值
datacl	aclitem[]		访问权限，详见GRANT和REVOKE

42.16. pg_depend

目录pg_depend记录数据库对象之间的依赖关系。这些信息允许DROP命令查找必须被DROP CASCADE删除的其他对象，或者在DROP RESTRICT情况下阻止删除。

另请参阅pg_shdepend，它对在一个数据库集簇中共享的对象之间的依赖提供了相似的功能。

表 42.16. pg_depend 列

名称	类型	引用	描述
classid	oid	pg_class.oid	依赖对象所在系统目录的OID
objid	oid	任意 OID 列	特定依赖对象的OID
objsubid	int4		对于一个表列，这里是列号（objid和classid指表本身）。对于所有其他对象类型，此列为0。
refclassid	oid	pg_class.oid	被引用对象所在的系统目录的OID
refobjid	oid	任意 OID 列	指定被引用对象的OID
refobjsubid	int4		对于一个表列，这里是列号（refobjid和refclassid指表本身）。对于所有其他对象类型，此列为0。
deptype	char		定义此依赖关系语义的一个代码，见文本

在所有情况下，一条pg_depend记录都表示：如果不同时删除依赖对象，就不能删除被引用对象。不过，还存在若干由deptype标识的子类型：

DEPENDENCY_NORMAL (n)

两个独立创建的对象之间的正常关系。依赖对象可以被删除而不影响被引用对象。被引用对象只能通过指定CASCADE来删除，这样依赖对象也会被删除。例如：表列对其数据类型有一个正常依赖关系。

DEPENDENCY_AUTO (a)

依赖对象可以与被引用对象分开删除，并且在被引用对象删除时应自动删除（无论使用RESTRICT还是CASCADE模式）。例如，表上的命名约束会自动依赖于该表，因此删除表时该约束也会消失。

DEPENDENCY_INTERNAL (i)

依赖对象作为被引用对象创建过程的一部分而创建，实际上只是其内部实现的一部分。对依赖对象执行DROP会被直接禁止（我们会告诉用户改为对被引用对象执行DROP）。对被引用对象执行DROP会传播到依赖对象，使其被删除，无论是否指定CASCADE。例如，为实施外键约束而创建的触发器会在内部依赖于该约束的pg_constraint项。

DEPENDENCY_PIN (p)

不存在依赖对象；这种类型的项表示系统本身依赖于被引用对象，因此该对象绝不能删除。这种类型的项只由initdb创建。用于表示依赖对象的各列包含零。

未来可能需要其他依赖关系的变种。

42.17. pg_description

目录pg_description存储对每一个数据库对象可选的描述（注释）。描述可以通过COMMENT命令操作，并可使用psql的\d命令查看。在pg_description的初始内容中提供了很多内置系统对象的描述。

表 42.17. pg_description 列

名称	类型	引用	描述
objoid	oid	任意 OID 列	描述所属对象的OID
classoid	oid	pg_class.oid	该对象所在的系统目录的OID
objsubid	int4		对于一个表列上的一个注释，这里是列号（objoid和classoid指表本身）。对所有其他对象类型，此列为0。
description	text		作为该对象描述的任意文本

42.18. pg_index

目录pg_index包含关于索引的部分信息。其他信息大部分在pg_class中。

表 42.18. pg_index 列

名称	类型	引用	描述
indexrelid	oid	pg_class.oid	此索引的pg_class项的OID

名称	类型	引用	描述
indrelid	oid	pg_class.oid	此索引所对应的基表的pg_class项的OID
indnatts	int2		索引中的总列数（与pg_class.relnatts重复）
indisunique	bool		如为真，这是唯一索引
indisprimary	bool		如为真，表示索引为表的主键（如果此列为真，indisunique也总是为真）
indisclustered	bool		如果为真，表示表最后以此索引进行了聚簇
indkey	int2vector	pg_attribute.attnum	这是一个包含indnatts值的数组，用于指示该索引为哪些表列建立索引。例如，1 3的值表示第一个和第三个表列组成索引键。此数组中的零表示相应的索引属性是表列上的表达式，而不是简单的列引用。
indclass	oidvector	pg_opclass.oid	对于索引键中的每一列，这里包含了要使用的操作符类的OID。详见pg_opclass。
indexprs	text		非简单列引用索引属性的表达式树（以nodeToString()形式）。对于indkey中每一个为0的项，这个列表中都有一个元素。如果所有的索引属性都是简单引用，此列为空。
indpred	text		部分索引谓词的表达式树（以nodeToString()形式）。如果不是部分索引，此列为空。

42.19. pg_inherits

目录pg_inherits记录有关表继承层次的信息。数据库中每个直接子表在这里都有一项。（非直接继承可以通过顺着项构成的链来确定。）

表 42.19. pg_inherits 列

名称	类型	引用	描述
inhrelid	oid	pg_class.oid	子表的OID
inhparent	oid	pg_class.oid	父表的OID
inhseqno	int4		如果一个孩子表有多于一个直接父表（多继承），这个数字说明了继承列被

名称	类型	引用	描述
			排列的顺序。计数从1开始。

42.20. pg_language

目录pg_language注册了可用于编写函数或存储过程的语言。
更多关于语言处理器的信息请参阅CREATE LANGUAGE和第 35 章。

表 42.20. pg_language 列

名称	类型	引用	描述
lanname	name		语言的名字
lanispl	bool		内部语言为假（如SQL），用户定义语言为真。当前，pg_dump仍然使用这个列来决定要转储哪些语言，但在未来这可能会被一种不同的机制所取代。
lanpltrusted	bool		为真表示这是一种可信的语言，即它被相信不会向普通SQL执行环境之外的任何东西授予权限。只有超级用户可以在非可信语言中创建函数。
lanplcallfoid	oid	pg_proc.oid	对于非内部语言，此列引用语言处理器，它是一个特殊函数，负责执行所有用这种语言编写的函数。
lanvalidator	oid	pg_proc.oid	此列引用一个负责在函数创建时对其进行语法和可用性检查的语言验证函数。如果没有提供验证器则为0。
lanacl	aclitem[]		访问权限，详见GRANT和REVOKE

42.21. pg_largeobject

目录pg_largeobject保存构成“大对象”的数据。一个大对象在被创建时会被分配一个OID。
每个大对象被分解成足够小的段或“页”，以便方便地作为行存储在pg_largeobject中。
每页中的数据量被定义为LOBLKSIZE（目前是BLCKSZ/4，通常为 2 kB）。

表 42.21. pg_largeobject 列

名称	类型	描述	引用
loid	oid		包含此页的大对象的标识符
pageno	int4		此页在它所属大对象中的页号（从0开始计）

名称	类型	描述	引用
data	bytea		实际存储在大对象中的数据。它从不会超过LOBLKSIZE字节，也可能更少。

pg_largeobject的每一行保存一个大对象的一个页的数据，从对象内部的字节偏移量（pageno * LOBLKSIZE）开始。现在的实现允许稀疏存储：页面可能丢失，并且可能比LOBLKSIZE字节短（即便不是最后一页）。一个大对象中丢失的区域会被读出为0。

42.22. pg_listener

目录 pg_listener 支持LISTEN和NOTIFY命令。监听者为自己正在监听的每个通知名称在 pg_listener 中创建一个条目。通知者会扫描 pg_listener 并更新每个匹配的条目，以表明发生了一个通知。通知者还会（使用表中记录的 PID）发送信号，把监听者从睡眠中唤醒。

表 42.22. pg_listener 列

名称	类型	引用	描述
relname	name		通知条件名。（该名称不必匹配数据库中的任何实际关系；字段名 relname 是历史遗留。）
listenerpid	int4		创建此条目的服务器进程的 PID
notification	int4		如果该监听者没有待处理的事件则为零。如果有事件待处理，则为发送通知的服务器进程的 PID

42.23. pg_namespace

目录pg_namespace存储名字空间。
名字空间是SQL模式之下的结构：每个名字空间拥有一个独立的表、类型等的集合，
且其中没有名字冲突。

表 42.23. pg_namespace 列

名称	类型	引用	描述
nspname	name		名字空间的名字
nspowner	oid	pg_authid.oid	名字空间的拥有者
nspacl	aclitem[]		访问权限，详见GRANT和REVOKE

42.24. pg_opclass

目录pg_opclass定义索引访问方法的操作符类。
每一个操作符类定义了一种特定数据类型和一种特定索引访问方法的索引列的语义。注意，
一个给定的数据类型/访问方法组合可以有多个操作符类，从而支持多种行为。

操作符类详见第 32.14 节。

定义一个操作符类的信息大部分实际上并不在它的pg_opclass行中，而在关联的pg_amop和pg_amproc行中。这些行被视为操作符类定义的一部分--这与一个关系由单条pg_class行加上pg_attribute及其他表中的关联行来定义的方式并无不同。

表 42.24. pg_opclass 列

名称	类型	引用	描述
opcamid	oid	pg_am.oid	操作符类所属的索引访问方法
opcname	name		操作符类的名称
opcnamespace	oid	pg_namespace.oid	操作符类所属的名字空间
opcowner	oid	pg_authid.oid	操作符类的拥有者
opcintype	oid	pg_type.oid	操作符类索引的数据类型
opcdefault	bool		如果此操作符类为opcintype的默认值则为真
opckeytype	oid	pg_type.oid	存储在索引中的数据的类型，如果值为0表示与opcintype相同

42.25. pg_operator

目录pg_operator存储关于操作符的信息。详见CREATE OPERATOR和第 32.12 节。

表 42.25. pg_operator 列

名称	类型	引用	描述
oprname	name		操作符的名称
oprnamespace	oid	pg_namespace.oid	操作符所属的名字空间的OID
oprowner	oid	pg_authid.oid	操作符的拥有者
oprkind	char		b = 中缀（“双目”），l = 前缀（“左目”），r = 后缀（“右目”）
oprcanhash	bool		该操作符支持哈希连接
oprleft	oid	pg_type.oid	左操作数类型
oprright	oid	pg_type.oid	右操作数类型
oprresult	oid	pg_type.oid	结果类型
oprcom	oid	pg_operator.oid	该操作符的交换子（如有）
oprnegate	oid	pg_operator.oid	该操作符的否定（如有）
oprlsortop	oid	pg_operator.oid	如果此操作符支持合并连接，则为对左操作数类型排序的操作符（L<L）
oprrsortop	oid	pg_operator.oid	如果此操作符支持合并连接，则为对右操作数类型排序的操作符（R<R）

名称	类型	引用	描述
oprltcmpop	oid	pg_operator.oid	如果此操作符支持合并连接，则为比较左右操作数类型的小于操作符（L<R）
oprgtcmpop	oid	pg_operator.oid	如果此操作符支持合并连接，则为比较左右操作数类型的大于操作符（L>R）
oprcode	regproc	pg_proc.oid	实现该操作符的函数
oprrest	regproc	pg_proc.oid	该操作符的限制选择率估算函数
oprjoin	regproc	pg_proc.oid	该操作符的连接选择率估算函数

未使用的列包含零。例如，前缀操作符的oprleft为零。

42.26. pg_pltemplate

目录pg_pltemplate存储过程语言的“模板”信息。有了语言模板，就可以用简单的CREATE LANGUAGE命令在特定数据库中创建该语言，无需指定实现细节。

与大多数系统目录不同，pg_pltemplate
由一个集簇中的所有数据库共享：每个集簇只有一份pg_pltemplate，
而不是每个数据库各有一份。这样，每个数据库都可以在需要时访问这些信息。

表 42.26. pg_pltemplate 列

名称	类型	描述	引用
tmplname	name		此模板所对应的语言名称
tmpltrusted	boolean		如果该语言被视为可信语言，则为真
tmplhandler	text		调用处理器函数的名称
tmplvalidator	text		验证器函数的名称；如果没有则为空
tmpllibrary	text		实现该语言的共享库的路径
tmplacl	aclitem[]		模板的访问权限（实际上未使用）

目前没有任何命令可以操纵过程语言模板；要更改内置信息，超级用户必须使用普通的 INSERT、DELETE 或 UPDATE 命令修改此表。PostgreSQL的将来版本很可能提供以更干净的方式更改这些条目的命令。

实现之后，tmplacl字段将为模板本身（即使用它创建语言的权限）提供访问控制，
而不是为由模板创建的语言提供访问控制。

42.27. pg_proc

目录pg_proc存放有关函数（或过程）的信息。更多信息请参见CREATE FUNCTION和第 32.3 节。

该表同时包含聚合函数和普通函数的数据。如果proisagg为真，在pg_aggregate中应该有一个相匹配的行。

表 42.27. pg_proc 列

名称	类型	引用	描述
proname	name		函数的名字
pronamespace	oid	pg_namespace.oid	函数所属的名字空间的OID
proowner	oid	pg_authid.oid	函数的拥有者
prolang	oid	pg_language.oid	实现语言或该函数的调用接口
proisagg	bool		该函数是聚合函数
prosecdef	bool		函数是一个安全性定义者（即，一个“setuid”函数）
proisstrict	bool		当任一调用参数为 NULL 时，函数是否返回空值。在这种情况下，函数实际上根本不会被调用。非“strict”函数必须准备好处理空值输入。
proretset	bool		函数是否返回一个集合（即，指定数据类型的多个值）
provolatile	char		provolatile说明函数是仅仅只依赖于它的输入参数，还是会被外部因素影响。值i表示“不变的”函数，它对于相同的输入总是输出相同的结果。值s表示“稳定的”函数，它的结果（对于固定输入）在一次扫描内不会变化。值v表示“不稳定的”函数，它的结果在任何时候都可能变化（使用v也表示函数具有副作用，因此对它们的调用无法得到优化）
pronargs	int2		输入参数的个数
prorettype	oid	pg_type.oid	返回值的数据类型
proargtypes	oidvector	pg_type.oid	一个函数参数的数据类型的数组。这只包括输入参数（含 INOUT 参数），因此也表现了函数的调用签名。
proallargtypes	oid[]	pg_type.oid	一个函数参数的数据类型的数组。这包括所有参数（含 OUT 和 INOUT 参数）。但是，如果所有参数都是 IN 参数，这个域将为空。注意下标是从1开始，然而由于历史原因

名称	类型	引用	描述
			proargtypes的下标是从0开始。
proargmodes	char[]		一个函数参数的模式的数组，编码为 i表示 IN 参数，o表示 OUT 参数，b表示 INOUT 参数。如果所有的参数都是 IN 参数，这个域为空。注意这里的下标对应着proallargtypes而不是proargtypes中的位置。
proargnames	text[]		一个函数参数的名字的数组。没有名字的参数在数组中设置为空字符串。如果没有一个参数有名字，这个域为空。注意这里的下标对应着proallargtypes而不是proargtypes中的位置。
prosrc	text		这个字段告诉函数处理器如何调用该函数。它可能是针对解释型语言的真实源码、一个链接符号、一个文件名，或任何其他取决于实现语言/调用约定的内容。
probin	bytea		关于如何调用函数的附加信息。其解释是与语言相关的。
proacl	aclitem[]		访问权限，详见GRANT和REVOKE

对于编译好的函数，包括内置函数和动态装载函数，prosrc包含函数的 C 语言名字（链接符号）。所有其他已知的语言类型，prosrc包含函数的源码文本。
除了对于动态载入的C函数之外，probin不使用，对于动态载入的C函数，它给定了包含该函数的共享库文件的名称。

42.28. pg_rewrite

目录pg_rewrite存储对于表和视图的重写规则。

表 42.28. pg_rewrite 列

名称	类型	引用	描述
rulename	name		规则名称
ev_class	oid	pg_class.oid	使用该规则的表
ev_attr	int2		该规则所针对的列（目前总是 -1，表示整个表）
ev_type	char		使用该规则的事件类型：1 = SELECT, 2 =

名称	类型	引用	描述
			UPDATE, 3 = INSERT, 4 = DELETE
is_instead	bool		为真表示是一个INSTEAD规则
ev_qual	text		规则条件的表达式树（以nodeToString()表示）
ev_action	text		规则动作的查询树（以nodeToString()表示）

注意

如果一个表在这个目录中有任何规则，pg_class.relhasrules必须为真。

42.29. pg_shdepend

目录pg_shdepend记录数据库对象和共享对象之间的依赖关系，例如角色。这些信息使得PostgreSQL可以确保对象在被删除时没有被其他对象引用。

另请参阅pg_depend，它对单个数据库中对象之间的依赖提供了相似的功能。

与大部分其他系统目录不同，pg_shdepend在整个集簇的所有数据库之间共享：在每一个集簇中只有一个pg_shdepend的拷贝，而不是每个数据库一份。

表 42.29. pg_shdepend 列

名称	类型	引用	描述
dbid	oid	pg_database.oid	依赖对象所在数据库的OID，对于共享对象为零
classid	oid	pg_class.oid	依赖对象所在系统目录的OID
objid	oid	任意 OID 列	特定依赖对象的OID
refclassid	oid	pg_class.oid	被引用对象所在的系统目录的OID（必须是一个共享的目录）
refobjid	oid	任意 OID 列	指定被引用对象的OID
deptype	char		定义此依赖关系语义的一个代码，见文本

在所有情况下，一条pg_shdepend记录都表示：如果不同时删除依赖对象，就不能删除被引用对象。不过，还存在若干由deptype标识的子类型：

SHARED_DEPENDENCY_OWNER (o)

被引用对象（必须是一个角色）是依赖对象的拥有者。

SHARED_DEPENDENCY_ACL (a)

被引用对象（必须是角色）出现在依赖对象的 ACL（访问控制列表，即权限列表）中。（不会为对象拥有者创建 SHARED_DEPENDENCY_ACL 项，因为对象拥有者无论如何都会有一个 SHARED_DEPENDENCY_OWNER 项。）

SHARED_DEPENDENCY_PIN (p)

不存在依赖对象；这种项表示系统本身依赖于被引用对象，因此该对象绝不能删除。这种项只由initdb创建。用于表示依赖对象的各列包含零。

将来可能还需要其他依赖类型。特别要注意的是，当前定义只支持把角色作为被引用对象。

42.30. pg_statistic

目录pg_statistic存储有关数据库内容的统计数据。其中的项由ANALYZE创建，查询规划器会使用这些数据来进行查询规划。注意所有的统计数据天然就是近似的，即使它刚刚被更新。

pg_statistic也存储关于索引表达式值的统计数据，就好像它们是真正的数据列，但在这种情况中starelid指索引。对一个普通非表达式索引列不会创建项，因为它将是底层表列的项的冗余。

由于不同种类的数据可能适合不同种类的统计信息，pg_statistic 在设计上尽量不对所存储的统计信息种类作出假定。只有极为通用的统计信息（例如空值情况）才在 pg_statistic 中有专用的列。其余统计信息都存储在“槽位”中。每个槽位都是一组相关的列，其内容由其中一列的代码编号来标识。更多信息见 src/include/catalog/pg_statistic.h。

pg_statistic不应公开可读，因为即使是表内容的统计信息也可能被认为是敏感的（例如，一个工资列的最大值和最小值就可能非常引人关注）。pg_stats是建立在pg_statistic之上的公开可读视图，它只会显示当前用户可读取的表的信息。

表 42.30. pg_statistic 列

名称	类型	引用	描述
starelid	oid	pg_class.oid	被描述列所属的表或索引
staattnum	int2	pg_attribute.attnum	被描述列的编号
stanullfrac	float4		列中空值所占的比例
stawidth	int4		非空项的平均存储宽度，以字节计
stadistinct	float4		列中不同非空数据值的数量。一个大于零的值是不同值的真正数目。一个小于零的值是表中行数的乘数的负值；例如，对于一个约 80% 的值为非空且每个非空值平均出现两次的列，可以表示为 stadistinct = -0.4。一个0值表示不同值的数目未知。
stakind*N*	int2		一个代码，它表示存储在该pg_statistic行中第*N*个“槽位”的统计类型。

名称	类型	引用	描述
staop*N*	oid	pg_operator.oid	一个用于生成这些存储在第*N*个“槽位”的统计信息的操作符。例如，直方图槽位会使用<操作符，该操作符定义了这些数据的排序顺序。
stanumbers*N*	float4[]		第*N*个“槽位”中相应种类的数值统计信息；如果该槽位种类不涉及数值，则为 NULL
stavalues*N*	anyarray		第*N*个“槽位”的类型的列值，如果该槽位类型不存储任何数据值则为 NULL。每个数组的元素值实际上都是指定列的数据类型或者是一个相关类型（如数组元素类型），因此，无法把这些列的类型定义得比anyarray更具体。

42.31. pg_tablespace

目录pg_tablespace存储关于可用表空间的信息。
表可以被放置在特定表空间中以实现磁盘布局的管理。

与大部分其他系统目录不同，pg_tablespace
在整个集簇的所有数据库之间共享：在每一个集簇中只有一个pg_tablespace的拷贝，
而不是每个数据库一份。

表 42.31. pg_tablespace 列

名称	类型	引用	描述
spcname	name		表空间名
spcowner	oid	pg_authid.oid	表空间的拥有者，通常是创建它的用户
spclocation	text		表空间的位置（目录路径）
spcacl	aclitem[]		访问权限，详见GRANT和REVOKE

42.32. pg_trigger

目录pg_trigger存储表和视图上的触发器。详见CREATE TRIGGER。

表 42.32. pg_trigger 列

名称	类型	引用	描述
tgrelid	oid	pg_class.oid	触发器所在的表
tgname	name		触发器名（在同一个表的触发器中必须唯一）

名称	类型	引用	描述
tgfoid	oid	pg_proc.oid	要被触发器调用的函数
tgtype	int2		标识触发器触发条件的位掩码
tgenabled	bool		如果触发器被启用则为真
tgisconstraint	bool		如果触发器实现一个引用完整性约束则为真
tgconstrname	name		引用完整性约束名
tgconstrrelid	oid	pg_class.oid	被一个引用完整性约束引用的表
tgdeferrable	bool		如果可延迟则为真
tginitdeferred	bool		如果初始为延迟则为真
tgnargs	int2		传递给触发器函数的参数字符串个数
tgattr	int2vector		目前未使用
tgargs	bytea		传递给触发器的参数字符串，每一个都以NULL结尾

注意

pg_class.reltriggers必须与给定关系在本目录中找到的触发器数量一致。

42.33. pg_type

目录pg_type存储有关数据类型的信息。基础类型和枚举类型（标量类型）使用CREATE TYPE创建，而域使用CREATE DOMAIN创建。数据库中的每一个表都会有一个自动创建的复合类型，用于表示表的行结构。也可以使用CREATE TYPE AS创建复合类型。

表 42.33. pg_type 列

名称	类型	引用	描述
typname	name		数据类型的名字
typnamespace	oid	pg_namespace.oid	包含此类型的名字空间的OID
typowner	oid	pg_authid.oid	类型的拥有者
typlen	int2		对于一个固定尺寸的类型，typlen是该类型内部表示的字节数。 对于一个变长类型，typlen为负值。-1表示一个“varlena”类型（具有长度字），-2表示一个以空值结尾的C字符串。
typbyval	bool		typbyval判断内部例程传递这个类型的数值时是通过传值还是传引用。 如果typlen不是1、2或4（或者在Datum为8字

名称	类型	引用	描述
			节的机器上为8），因此typbyval最好是假。变长类型总是传引用。注意即使长度允许传值，typbyval也可以为假；例如float4类型目前就是这样。
typtype	char		typtype为b表示基本类型，c表示复合类型（例如表的行类型），d表示域，p表示伪类型。另见typrelid和typbasetype。
typisdefined	bool		如果此类型已被定义则为真，如果此类型只是一个表示还未定义类型的占位符则为假。当typisdefined为假，除了类型名字、名字空间和OID之外什么都不能被依赖。
typdelim	char		在分析数组输入时，分隔两个此类型值的字符。注意该分隔符是与数组元素数据类型相关联的，而不是和数组的数据类型关联。
typrelid	oid	pg_class.oid	如果这是一个复合类型（见typtype），那么此列指向pg_class中定义对应表的项（对于自由存在的复合类型，pg_class项并不表示一个表，但不管怎样该类型的pg_attribute项需要链接到它）。对非复合类型此列为零。
typelem	oid	pg_type.oid	如果typelem不为 0，则它标识pg_type中的另一行。此时可以像数组一样对当前类型使用下标，得到typelem类型的值。“真正的”数组类型是变长的（typlen = -1），但某些定长类型（typlen > 0）的typelem也非零，例如name和point。如果定长类型具有typelem，则其内部表示必须是若干个typelem数据类型的值，不含其他数据。变长数组类型的首部由数组子例程定义。
typinput	regproc	pg_proc.oid	输入转换函数（文本格式）

名称	类型	引用	描述
typoutput	regproc	pg_proc.oid	输出转换函数（文本格式）
typreceive	regproc	pg_proc.oid	输入转换函数（二进制格式），如果没有则为零
typsend	regproc	pg_proc.oid	输出转换函数（二进制格式），如果没有则为零
typanalyze	regproc	pg_proc.oid	自定义 ANALYZE 函数，零表示使用标准函数
typalign	char		typalign表示存储此类型值时所要求的对齐方式。它适用于磁盘存储，以及该值在 PostgreSQL 内部的大多数表示形式。当多个值连续存放时，例如在磁盘上完整行的表示中，这种类型的数据前会插入填充，使其从指定边界开始。对齐参考点是该序列中第一个数据的起始位置。 可能的值为： • c = char对齐，即不需要对齐。 • s = short对齐（在大部分机器上为2字节）。 • i = int对齐（在大部分机器上为4字节）。 • d = double对齐（在很多机器上为8字节，但绝不是全部）。 **注意** 对于系统表中使用的类型，关键的是在pg_type定义的大小和对齐与编译器在表示表行的结构体中布局列的方式一致。
typstorage	char		typstorage用于指示 varlena 类型（即typlen = -1）是否支持 TOAST，以及这种类型的属性应采用什么默认策略。可能的值为：

名称	类型	引用	描述
			• p：值必须始终以普通方式存储。 • e：值可以存储在一个“二级”关系中（如果关系有这样的关系，参见pg_class.reltoastrelid）。 • m：值可以被压缩并内联存储。 • x：值可以被压缩并内联存储，或存储在“二级”存储中。 注意，m列也可以被移动到二级存储，但只能作为最后手段（e和x列会先被移动）。
typnotnull	bool		typnotnull表示类型上的一个非空约束。只用于域。
typbasetype	oid	pg_type.oid	如果这是一个域（见typtype），则typbasetype标识该域所基于的类型。如果此类型不是域，则为 0。
typtypmod	int4		域使用typtypmod来记录被应用于它们基础类型的typmod（如果基础类型不使用typmod，则为-1）。如果此类型不是一个域则为-1。
typndims	int4		对于作为数组的域，typndims是数组维度数（即，typbasetype是一个数组类型；该域的typelem将匹配基类型的typelem）。对于不是基于数组类型的域的其他类型，此值为零。
typdefaultbin	text		如果typdefaultbin非空，那么它就是该类型默认表达式的nodeToString()表示形式。此列只用于域。
typdefault	text		如果某类型没有相关默认值，那么typdefault为空。如果typdefaultbin不为空，那么typdefault必须包含由typdefaultbin表示的默认表达式的人类可读版本。如果typdefaultbin为空而

名称	类型	引用	描述
			typdefault不为空，则typdefault是该类型默认值的外部表示形式，它可以被交给该类型的输入转换器来产生一个常量。

42.34. 系统视图

除了系统目录之外，PostgreSQL还提供了一些内置视图。
一些系统视图提供了对系统目录中一些常用查询的便捷访问。
其他视图则提供了对内部服务器状态的访问。

信息模式（第 31 章）提供了一组与系统视图功能重叠的替代视图。由于信息模式是SQL标准，而这里描述的视图是PostgreSQL特有的，所以如果信息模式提供了所需的全部信息，
通常最好使用信息模式。

表 42.34列出了这里描述的系统视图。每个视图的更详细文档见下文。还有一些额外的视图提供了对统计信息收集器结果的访问；它们在表 24.1中描述。

除非另有说明，否则这里描述的所有视图都是只读的。

表 42.34. 系统视图

视图名称	用途
pg_group	数据库用户组
pg_indexes	索引
pg_locks	当前持有的锁
pg_prepared_xacts	预备事务
pg_roles	数据库角色
pg_rules	规则
pg_settings	参数设置
pg_shadow	数据库用户
pg_stats	规划器统计信息
pg_tables	表
pg_user	数据库用户
pg_views	视图

42.35. pg_group

视图pg_group存在是为了向后兼容性：它模拟了在PostgreSQL版本8.1之前存在的目录。
它显示了所有被标记为非rolcanlogin的角色的名称和成员，
这是对被用作组的角色集合的近似。

表 42.35. pg_group 列

名称	类型	引用	描述
groname	name	pg_authid.rolname	组的名称

名称	类型	引用	描述
grosysid	oid	pg_authid.oid	组ID
grolist	oid[]	pg_authid.oid	包含此组中角色ID的一个数组

42.36. pg_indexes

视图pg_indexes提供了有关数据库中每个索引的有用信息。

表 42.36. pg_indexes 列

名称	类型	引用	描述
schemaname	name	pg_namespace.nspname	包含表和索引的模式名
tablename	name	pg_class.relname	此索引的基表名称
indexname	name	pg_class.relname	索引名
tablespace	name	pg_tablespace.spcname	包含索引的表空间名（如果是数据库的默认值则为空）
indexdef	text		索引定义（重建出的 CREATE INDEX 命令）

42.37. pg_locks

视图pg_locks提供了对数据库服务器中打开的事务持有的锁的信息的访问。
有关锁定的更多讨论，请参见第 12 章。

pg_locks包含每个活动的可锁定对象、请求的锁模式和相关事务的一行。因此，
如果多个事务正在持有或等待锁定它，同一可锁定对象可能会出现多次。但是，
当前没有任何锁定的对象将不会出现。

有几种不同类型的可锁定对象：整个关系（例如，表）、关系的单个页面、关系的单个元组、事务 ID（虚拟和永久 ID 均包括），以及一般的数据库对象（由类 OID 和对象 OID 标识，与pg_description或pg_depend中的方式相同）。此外，扩展一个关系的权限被表示为一个单独的可锁定对象。

表 42.37. pg_locks 列

名称	类型	引用	描述
locktype	text		可锁定对象的类型：relation、extend、page、tuple、transactionid、object或userlock
database	oid	pg_database.oid	锁目标存在的数据库的OID，如果目标是一个共享对象则为0，如果目标是一个事务ID则为 NULL
relation	oid	pg_class.oid	作为锁目标的关系的OID，如果目标既不是关

名称	类型	引用	描述
			系，也不是关系的一部分则此列为 NULL
page	integer		作为锁目标的页在关系中的页号，如果目标不是一个关系页或元组则此列为 NULL
tuple	smallint		作为锁目标的元组在页中的元组号，如果目标不是一个元组则此列为 NULL
transactionid	xid		锁定目标事务的标识，如果目标不是事务 ID，则为 NULL
classid	oid	pg_class.oid	包含锁目标的系统目录的 OID，如果目标不是一个普通数据库对象则此列为 NULL
objid	oid	任意 OID 列	锁目标在它的系统目录中的OID，如果目标不是一个普通数据库对象则为 NULL
objsubid	smallint		对于一个表列，这里是列号（objid和classid指表本身）。对于所有其他对象类型，此列为0。如果对象不是普通数据库对象则为 NULL。
transaction	xid		持有或等待此锁的事务的 ID。
pid	integer		保持这个锁或者正在等待这个锁的服务器进程的 PID，如果此锁被一个预备事务所持有则此列为 NULL
mode	text		此进程已持有或者希望持有的锁模式的名称（参见第 12.3.1 节）
granted	boolean		如果锁已授予则为真，如果锁被等待则为假

在表示由指定事务持有的锁的行中，granted 为真。为假表示该事务当前正在等待获取此锁，这意味着某个其他事务正在同一可锁定对象上持有冲突的锁模式。等待的事务会休眠，直到另一个锁被释放（或检测到死锁情况）。单个事务一次最多只能等待获取一个锁。

每个事务在其整个持续期内都对其事务 ID 持有独占锁。当一个事务发现有必要专门等待另一个事务时，它会尝试获取另一个事务 ID 上的共享锁。只有当另一个事务终止并释放其锁时，这才会成功。

虽然元组是一种可锁定的对象类型，但关于行级锁的信息存储在磁盘上，而不是内存中，因此行级锁通常不会出现在此视图中。如果一个事务正在等待行级锁，它通常会出现在视图中，等待当前持有该行锁的事务 ID。

如果使用了用户定义锁，它们会使用普通数据库对象的列来显示。不过，在这种情况下锁字段的实际含义由用户决定。

当访问pg_locks视图时，内部锁管理器数据结构会被短暂锁定，
并为该视图制作一份要显示的拷贝。这确保该视图产生一致的结果集，
同时不会在不必要的时间内阻塞正常的锁管理器操作。尽管如此，如果频繁访问这个视图，
可能会对数据库性能产生一些影响。

pg_locks提供了集簇中所有锁的全局视图，不仅包括与当前数据库相关的锁。虽然它的relation列可以与pg_class.oid 连接来识别被锁定的关系，但这仅对当前数据库中的关系有效（即 database 列为当前数据库的 OID 或零的那些关系）。

如果你启用了统计收集器，pid列可以与pg_stat_activity视图的procpid列连接，
以获取持有或等待该锁的会话的更多信息。另外，如果你正在使用预备事务，transaction
列可以与pg_prepared_xacts视图的transaction列连接，
以获取持有锁的预备事务的更多信息。（预备事务永远不会等待锁，
但会继续持有它运行期间获得的锁。）

42.38. pg_prepared_xacts

视图pg_prepared_xacts显示当前准备进行两阶段提交的事务的信息 (详见PREPARE TRANSACTION)。

pg_prepared_xacts 为每个预备事务包含一行。事务提交或回滚时，对应条目会被移除。

表 42.38. pg_prepared_xacts 列

名称	类型	引用	描述
transaction	xid		预备事务的数字事务 ID
gid	text		分配给事务的全局标识符
prepared	timestamp with time zone		此事务为提交准备好的时间
owner	name	pg_authid.rolname	执行此事务的用户名
database	name	pg_database.datname	执行此事务所在数据库的名称

当访问pg_prepared_xacts视图时，内部事务管理器数据结构会被暂时锁定，
并为视图制作一份副本用以显示。这确保视图生成一致的结果集，
同时不会不必要地阻塞正常操作。尽管如此，如果频繁访问该视图，
可能会对数据库性能产生一些影响。

42.39. pg_roles

视图pg_roles提供数据库角色的信息。它本质上是 pg_authid 的公开可读视图，并隐藏密码字段。

此视图显式显示底层表的 OID 列，因为与其他目录连接时需要此列。

表 42.39. pg_roles 列

名称	类型	引用	描述
rolname	name		角色名
rolsuper	bool		角色有超级用户权限
rolinherit	bool		该角色是否会自动继承其所属其他角色的权限

名称	类型	引用	描述
rolcreaterole	bool		角色能创建更多角色
rolcreatedb	bool		角色能创建数据库
rolcatupdate	bool		角色能否直接更新系统目录。（即使是超级用户，也只有该列为真时才能这样做）
rolcanlogin	bool		角色是否能登录。即该角色是否能够作为初始会话授权标识符
rolconnlimit	int4		对于可以登录的角色，本列设置该角色能够建立的最大并发连接数。-1 表示无限制。
rolpassword	text		不是密码（始终显示为 ********）
rolvaliduntil	timestamptz		密码过期时间（仅用于密码认证）；不设过期时间时为空值
rolconfig	text[]		运行时配置变量的角色特定默认值
oid	oid	pg_authid.oid	角色的ID

42.40. pg_rules

视图pg_rules提供了有关查询重写规则的有用信息。

表 42.40. pg_rules 列

名称	类型	引用	描述
schemaname	name	pg_namespace.nspname	包含表的模式名称
tablename	name	pg_class.relname	规则适用的表名
rulename	name	pg_rewrite.rulename	规则名
definition	text		规则定义（重建的创建命令）

pg_rules视图排除了视图的 ON SELECT 规则；那些规则可以在pg_views中看到。

42.41. pg_settings

视图pg_settings提供对服务器运行时参数的访问。它提供SHOW和SET命令的一个接口。它还提供了一些关于每个参数的事实，这些事实无法直接从SHOW中获取，例如最小值和最大值。

表 42.41. pg_settings 列

名称	类型	描述	引用
name	text		运行时配置参数名

名称	类型	描述	引用
setting	text		参数的当前值
category	text		参数的逻辑组
short_desc	text		参数的简短描述
extra_desc	text		附加的参数的详细描述
context	text		要求设置此参数值的上下文 (参见下方)
vartype	text		参数类型（bool、integer、real或string）
source	text		当前参数值的来源
min_val	text		参数的最小允许值（对非数值参数为空）
max_val	text		参数的最大允许值（对非数值参数为空）

pg_settings视图不能插入或删除，但可以更新。对 pg_settings 的某一行执行 UPDATE，等同于对该命名参数执行SET命令。更改只影响当前会话所用的值。
如果在之后被中止的事务中发出了 UPDATE，那么事务回滚时 UPDATE 命令的效果也会消失。一旦外围事务提交，这些效果会持续到会话结束，除非被另一个 UPDATE 或 SET 覆盖。

42.42. pg_shadow

视图pg_shadow存在是为了向后兼容性：它模拟了在PostgreSQL 8.1版本之前存在的目录。它显示了所有被标记为rolcanlogin的角色的属性。

这个名称源于该表不应向公众开放读取，因为它包含密码。pg_user 是 pg_shadow 的公开可读视图，并隐藏密码字段。

表 42.42. pg_shadow 列

名称	类型	引用	描述
usename	name	pg_authid.rolname	用户名
usesysid	oid	pg_authid.oid	用户的ID
usecreatedb	bool		用户可创建数据库
usesuper	bool		用户为一个超级用户
usecatupd	bool		用户能否更新系统目录。（即使是超级用户，也只有该列为真时才能这样做。）
passwd	text		密码（可能已加密）
valuntil	abstime		密码过期时间（仅用于密码认证）
useconfig	text[]		运行时配置变量的会话默认值

42.43. pg_stats

视图pg_stats提供对存储在pg_statistic目录中信息的访问。
此视图仅允许访问用户具有读取权限的表对应的pg_statistic行，
因此可以安全地允许对此视图进行公共读取访问。

pg_stats也旨在以比底层目录更易读的格式呈现信息— 但其模式必须在为pg_statistic定义新的槽类型时进行扩展。

表 42.43. pg_stats 列

名称	类型	引用	描述
schemaname	name	pg_namespace.nspname	包含表的模式名称
tablename	name	pg_class.relname	表的名称
attname	name	pg_attribute.attname	被此行描述的列名
null_frac	real		列项中为空的比例
avg_width	integer		列的条目的平均字节宽度
n_distinct	real		如果大于零，表示列中可区分值的估计个数。如果小于零，是可区分值个数除以行数的负值（当ANALYZE认为可区分值的数量会随着表增长而增加时采用负值的形式，而如果认为列具有固定数量的可选值时采用正值的形式）。例如，-1表示一个唯一列，即其中可区分值的个数等于行数。
most_common_vals	anyarray		列中高频值的一个列表（如果没有任何一个值看起来比其他值更常用，此列为空）
most_common_freqs	real[]		高频值的频率列表，即每一个高频值的出现次数除以总行数（如果most_common_vals为空，则此列为空）
histogram_bounds	anyarray		将列值划分成大小接近的组的值列表。如果存在most_common_vals，其中的值会被直方图计算所忽略（如果列类型没有一个<操作符或者most_common_vals等于整个值集合，则此列为空）
correlation	real		物理行顺序和列值逻辑顺序之间的统计关联。其范围从-1到+1。当值接近-1或+1时，在列上的一个索引扫描被认为比值接近0时的代价更低，因为这种情况减少了对磁盘的随机访问（如果列数据类型不

名称	类型	引用	描述
			具有一个<操作符，则此列为空）

most_common_vals和histogram_bounds数组中的条目最大数量可以逐列设置，使用ALTER TABLE SET STATISTICS命令，或者通过设置default_statistics_target运行时参数来全局设置。

42.44. pg_tables

视图pg_tables提供了对数据库中每个表的有用信息的访问。

表 42.44. pg_tables 列

名称	类型	引用	描述
schemaname	name	pg_namespace.nspname	包含表的模式名称
tablename	name	pg_class.relname	表的名称
tableowner	name	pg_authid.rolname	表拥有者名称
tablespace	name	pg_tablespace.spcname	包含该表的表空间名称（如果使用数据库的默认表空间，此列为空）
hasindexes	boolean	pg_class.relhasindex	如果表有（或最近有过）任何索引，此列为真
hasrules	boolean	pg_class.relhasrules	如果表有（或曾经有过）规则，此列为真
hastriggers	boolean	pg_class.reltriggers	如果表有（或者曾经有过）触发器，此列为真

42.45. pg_user

视图pg_user提供数据库用户的信息。它本质上是 pg_shadow 的公开可读视图，并隐藏密码字段。

表 42.45. pg_user 列

名称	类型	描述	引用
usename	name		用户名
usesysid	int4		用户 ID（用来引用该用户的任意数字）
usecreatedb	bool		用户可创建数据库
usesuper	bool		用户为一个超级用户
usecatupd	bool		用户能否更新系统目录。（即使是超级用户，也只有该列为真时才能这样做。）
passwd	text		不是密码（始终显示为 ********）

名称	类型	描述	引用
valuntil	abstime		密码过期时间（仅用于密码认证）
useconfig	text[]		运行时配置变量的会话默认值

42.46. pg_views

视图pg_views提供了对数据库中每个视图的有用信息的访问。

表 42.46. pg_views 列

名称	类型	引用	描述
schemaname	name	pg_namespace.nspname	包含视图的模式名
viewname	name	pg_class.relname	视图名称
viewowner	name	pg_authid.rolname	视图拥有者名称
definition	text		视图定义（一个重建的SELECT 查询）

第 43 章 前端/后端协议

PostgreSQL使用一种基于消息的协议在前端和后端（客户端与服务器）之间进行通信。
该协议既支持TCP/IP，也支持 Unix 域套接字。端口号 5432 已在 IANA 注册为支持该协议的服务器的惯用 TCP 端口号，但实际上任何非特权端口号都可以使用。

本文档描述协议 3.0 版，它在 PostgreSQL 7.4 及以后版本中实现。早期协议版本的描述参见 PostgreSQL 文档的以前发行版。单个服务器可以支持多个协议版本。最初的启动请求消息告诉服务器客户端试图使用哪个协议版本，然后服务器在有能力的情况下遵循该协议。

建立在该协议之上的更高层特性（例如 libpq 如何在建立连接时传递某些环境变量）将在别处介绍。

为了高效地为多个客户端提供服务，服务器会为每个客户端启动一个新的“后端”进程。
在当前实现中，一旦检测到传入连接，就会立刻创建新的子进程。不过，
这一点对协议而言是透明的。就协议而言，术语“后端”和“服务器”可以互换；同样，“前端”和“客户端”也可以互换。

43.1. 概述

协议分为启动和正常操作两个阶段。在启动阶段，前端打开到服务器的连接，
并完成服务器所要求的认证。（这可能只涉及一条消息，
也可能因所用认证方法不同而需要多条消息。）如果一切顺利，
服务器随后会向前端发送状态信息，并最终进入正常操作。除最初的启动请求消息外，
协议的这一部分由服务器驱动。

在正常操作中，前端向后端发送查询及其他命令，后端则返回查询结果和其他响应。
少数情况下（例如 `NOTIFY`），后端会发送未请求的消息，但会话中的绝大多数交互仍由前端请求驱动。

会话通常由前端选择终止，但在某些情况下也可能由后端强制终止。无论哪种情况，
后端关闭连接时，都会在退出前回滚所有打开的（未完成的）事务。

在正常操作中，SQL 命令可以通过两种子协议之一执行。在“简单查询”协议中，前端只需发送文本形式的查询字符串，后端会立即对其进行解析并执行。在“扩展查询”协议中，
查询处理被拆分为多个步骤：解析、参数值绑定以及执行。这带来了更高的灵活性和性能收益，
但代价是额外的复杂性。

正常操作还包含用于`COPY`等特殊操作的额外子协议。

43.1.1. 消息概述

所有通信都通过消息流进行。消息的第一个字节标识消息类型，
接下来的四个字节给出消息其余部分的长度（该长度计数包含自身，但不包括消息类型字节）。
消息剩余内容由消息类型决定。由于历史原因，客户端发送的第一条消息（启动消息）
没有开头的消息类型字节。

为了避免与消息流失去同步，服务器和客户端通常都会先根据字节计数把整条消息读入缓冲区，
然后再处理其内容。这样一来，如果在处理内容时检测到错误，就比较容易恢复。
在极端情况下（例如没有足够内存缓冲整条消息），
接收方也可以利用字节计数判断在恢复读取消息之前需要跳过多少输入。

反过来，服务器和客户端也必须注意绝不能发送不完整的消息。通常的做法是在开始发送之前，先在缓冲区中整理好整条消息。如果在发送或接收消息的途中发生通信故障，唯一合理的做法就是放弃连接，因为几乎不可能重新恢复消息边界的同步。

43.1.2. 扩展查询概述

在扩展查询协议中，SQL 命令的执行被划分为多个步骤。各步骤之间保留的状态由两类对象表示：预备语句和 portal。预备语句表示对文本查询字符串完成解析和语义分析（可选地还有规划）后的结果。预备语句本身还不一定能直接执行，因为它可能缺少特定参数的值。portal 表示一条已经可以执行、或已经部分执行过的语句，其中所有缺失的参数值都已补齐。（对于 `SELECT` 语句，portal 等价于一个打开的游标；但由于游标不能处理非 `SELECT` 语句，这里采用了不同的措辞。）

整个执行周期包括一个解析步骤，它从文本查询字符串创建预备语句；一个绑定步骤，它根据预备语句和所需参数值创建 portal；以及一个执行步骤，用于执行 portal 中的查询。对于返回行的查询（`SELECT`、`SHOW`等），可以要求执行步骤只取回有限数量的行，因此可能需要多次执行步骤才能完成整个操作。

后端可以跟踪多个预备语句和 portal（但请注意，它们只存在于单个会话内，绝不会在会话之间共享）。已有的预备语句和 portal 都通过创建时赋予的名称来引用。此外，还存在一个“未命名”的预备语句和 portal。虽然它们的行为与有名对象大体相同，但对未命名对象的操作是为“只执行一次然后丢弃”的场景优化的，而对有名对象的操作则是基于会被多次使用的预期进行优化的。

43.1.3. 格式与格式代码

某一特定数据类型的数据可以使用多种不同的格式之一进行传输。自 PostgreSQL 7.4 起，当前只支持“文本”和“二进制”两种格式，但协议为未来扩展留出了空间。任意值所需的格式由格式代码指定。客户端可以为每个传输的参数值以及查询结果的每一列指定格式代码。文本格式的代码为零，二进制格式的代码为一，其他格式代码则保留供将来定义。

值的文本表示是相应数据类型的输入/输出转换函数生成和接受的字符串。在传输形式中，值的末尾没有空字符；前端若要将收到的值作为 C 字符串处理，必须自行添加一个。（文本格式也不允许内嵌空字符。）

整数的二进制表示采用网络字节序（最高有效字节在前）。至于其他数据类型，请查阅文档或源代码了解其二进制表示形式。要注意，复杂数据类型的二进制表示可能会在不同服务器版本之间发生变化；文本格式通常是可移植性更好的选择。

43.2. 消息流

本节描述消息流以及各种消息类型的语义（每种消息的精确格式见第 43.4 节）。根据连接所处的状态不同，存在若干不同的子协议：启动、查询、函数调用、`COPY`以及终止。对于异步操作（包括通知响应和命令取消）还有专门规定，它们可能在启动阶段结束后的任何时刻发生。

43.2.1. 启动

要开始一个会话，前端会打开到服务器的连接并发送一条启动消息。该消息包含用户名以及用户希望连接的数据库名；它还指明要使用的协议版本。（启动消息也可以选择性地包含运行时参数的附加设置。）

随后服务器会结合这些信息以及配置文件（例如 `pg_hba.conf`）的内容，初步判断是否接受该连接，以及需要何种额外认证（如果需要）。

接着服务器会发送适当的认证请求消息，
前端必须以适当的认证响应消息进行回应（例如密码）。原则上，认证请求/
响应周期可能需要多次往返，但目前的认证方法中没有一种会使用多于一次的请求和响应。
在某些方法中，前端根本不需要发送响应，因此也不会出现认证请求。

认证周期要么以服务器拒绝连接（ErrorResponse）结束，要么以 AuthenticationOk 结束。

该阶段服务器可能发送的消息有：

ErrorResponse

连接尝试被拒绝。随后服务器立即关闭连接。

AuthenticationOk

认证交换成功完成。

AuthenticationKerberosV5

前端现在必须参与与服务器之间的 Kerberos V5 认证对话（此处不再描述，它属于 Kerberos 规范的一部分）。如果对话成功，服务器会响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationCleartextPassword

前端现在必须发送一个以明文形式包含密码的 PasswordMessage。如果密码正确，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationCryptPassword

前端现在必须发送一个 PasswordMessage，其中包含用 crypt(3) 加密的密码，加密时使用 AuthenticationCryptPassword 消息中指定的 2 字符盐。如果密码正确，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationMD5Password

前端现在必须发送一个 PasswordMessage，其中包含使用 AuthenticationMD5Password 消息中指定的 4 字符盐经过 MD5 加密的密码。如果密码正确，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

AuthenticationSCMCredential

此响应仅适用于支持 SCM 凭据消息的平台上的本地 Unix 域连接。前端必须发出一条 SCM 凭据消息，然后发送一个单字节数据。（该数据字节的内容无关紧要；它仅用于确保服务器等待足够长的时间来接收凭据消息。）如果凭据可接受，服务器响应 AuthenticationOk；否则响应 ErrorResponse。

如果前端不支持服务器要求的认证方式，那么它应该马上关闭连接。

在收到 AuthenticationOk 消息之后，前端必须继续等待来自服务器的后续消息。在这个阶段，
一个后端进程正在启动，而前端只是旁观者。启动尝试仍可能失败（ErrorResponse），
但通常情况下，后端会发送一些 ParameterStatus 消息、BackendKeyData，以及最后的 ReadyForQuery。

在这个阶段，后端会尝试应用启动消息中给出的任何额外运行时参数设置。如果成功，
这些值就会成为会话的默认值。发生错误则会导致 ErrorResponse 并退出。

这个阶段来自后端的可能消息是：

BackendKeyData

该消息提供密钥数据。如果前端希望稍后发送取消请求，就必须保存这些数据。
前端不应响应该消息，而应继续等待 ReadyForQuery 消息。

ParameterStatus

该消息告知前端后端参数的当前（初始）设置，例如client_encoding或DateStyle。
前端可以忽略这些信息，也可以记录下来供后续使用；详见第 43.2.6 节。前端不应响应该消息，而应继续等待 ReadyForQuery 消息。

ReadyForQuery

启动成功，前端现在可以发出命令。

ErrorResponse

启动失败，在发送完这个消息之后连接被关闭。

NoticeResponse

已发出一条警告消息。前端应显示该消息，但仍应继续等待 ReadyForQuery 或 ErrorResponse。

ReadyForQuery 消息与后端在每个命令周期结束后发出的消息是同一个。前端可根据自身编码需要，将 ReadyForQuery 视为一个命令周期的开始，或视为启动阶段以及后续每个命令周期的结束。

43.2.2. 简单查询

一个简单查询周期由前端向后端发送一条 Query 消息来启动。该消息包含一条或多条以文本字符串表示的 SQL 命令。后端随后会根据查询命令串的内容向前端发送一条或多条响应消息，最后再发送一条 ReadyForQuery 响应消息。ReadyForQuery 通知前端它现在可以安全地发送新的命令。（实际上，前端并不一定非要等到收到 ReadyForQuery 才发送下一条命令，但这样一来，前端就必须自己处理“先前命令失败而后续已发出的命令成功”这种情况。）

后端可能返回的响应消息有：

CommandComplete

一条 SQL 命令正常完成。

CopyInResponse

后端已准备好把数据从前端复制到表中；参见第 43.2.5 节。

CopyOutResponse

后端已准备好把数据从表中复制到前端；参见第 43.2.5 节。

RowDescription

表示即将返回行作为对`SELECT`、`FETCH`等查询的响应。此消息的内容描述了行的列布局。
该消息之后，每个返回给前端的行都对应一条 DataRow 消息。

DataRow

由SELECT、FETCH等查询返回的一组行中的一个。

EmptyQueryResponse

识别出了一条空查询字符串。

ErrorResponse

发生了一个错误。

ReadyForQuery

查询字符串的处理已完成。发送一个单独的消息来指示这一点，
因为查询字符串可能包含多个SQL命令。（Command
Complete标记了一个SQL命令的处理结束，而不是整个字符串的结束。）
无论处理是成功还是出现错误，都将始终发送ReadyForQuery。

NoticeResponse

与查询相关的警告消息已发出。通知是其他响应的补充，即后端将继续处理命令。

SELECT 查询（或其他返回行集的查询，如 EXPLAIN 或 SHOW）的响应通常包含 Row Description、零条或多条 DataRow 消息，以及最后的 CommandComplete。在前端与服务器之间执行 COPY 输入或输出时，会使用第 43.2.5 节所述的特殊协议。所有其他类型的查询通常只产生一条 CommandComplete 消息。

由于查询字符串可能包含若干条查询（以分号分隔），
因此在后端完成整个查询字符串的处理之前，可能会出现多个这样的响应序列。
只有在整个字符串处理完毕且后端已准备好接受新的查询字符串时，才会发出 ReadyForQuery 消息。

如果收到的是一条完全为空的查询字符串（除了空白字符之外没有任何内容），响应就是一条 EmptyQueryResponse，后面跟着 ReadyForQuery。

一旦发生错误，就会发出一条 ErrorResponse 消息，后面跟着 ReadyForQuery。Error Response 会中止该查询字符串中后续所有处理（即使其中还包含其他查询）。请注意，这种情况可能发生在处理单条查询所生成的消息序列中途。

在简单 Query 模式下，取回值的格式总是文本，除非给定命令是从一个用 BINARY 选项声明的游标进行的 FETCH。那种情况下，取回的值为二进制格式。RowDescription 消息中给出的格式代码指明使用的是哪种格式。

当前端正在等待其他类型的消息时，也必须准备好接收 ErrorResponse 和 NoticeResponse 消息。参见第 43.2.6 节，了解后端因外部事件而可能生成的消息。

建议以状态机的方式编写前端，使其能够在任何合理的时机接收相应类型的消息，
而不把消息确切顺序的假设写死在代码中。

43.2.3. 扩展查询

扩展查询协议把上文描述的简单查询协议拆分成多个步骤。准备步骤的结果可以重复使用，
从而提高效率。此外，它还提供了额外特性，例如可以把数据值作为独立参数提供，
而不必直接插入查询字符串中。

在扩展协议中，前端首先发送一条 Parse 消息，其中包含文本查询字符串、可选的参数占位符数据类型信息，以及目标预备语句对象的名称（空字符串表示未命名预备语句）。响应要么是

ParseComplete，要么是 ErrorResponse。参数数据类型可以用 OID 指定；如果未给出，解析器会像处理无类型字面字符串常量那样尝试推断其数据类型。

> **注意**
>
> 一个参数的数据类型可以通过设为零，或让参数类型 OID 数组短于查询字符串中参数符号（`$n`）的数量来保持未指定。另一个特例是，参数类型可以指定为 `void`（即伪类型 `void` 的 OID）。这样做是为了允许参数符号用于那些实际上是 OUT 参数的函数参数。通常并不存在可使用 `void` 参数的上下文，但如果这样的参数符号出现在函数参数列表中，它实际上会被忽略。例如，像 `foo($1,$2,$3,$4)` 这样的函数调用，如果 `$3` 和 `$4` 被指定为 `void` 类型，就可能匹配一个带有两个 IN 参数和两个 OUT 参数的函数。

> **注意**
>
> Parse 消息中的查询字符串不能包含多于一条 SQL 语句；否则会报告语法错误。这个限制在简单查询协议中并不存在，但在扩展协议中必须如此，因为若允许预备语句或 portal 包含多条命令，会使协议变得过于复杂。

如果成功创建，命名预备语句对象会一直持续到当前会话结束，除非被显式销毁。
未命名预备语句只会持续到下一条把未命名语句作为目标的 Parse 语句发出为止。（注意，简单 Query 消息也会销毁未命名语句。）有名预备语句在被另一条 Parse 消息重新定义之前必须显式关闭，但未命名语句则无此要求。有名预备语句也可以在 SQL command level, using `PREPARE` and `EXECUTE`.

一旦预备语句存在，就可以用 Bind 消息把它准备为可执行状态。Bind 消息给出源预备语句的名称（空字符串表示未命名预备语句）、目标 portal 的名称（空字符串表示未命名 portal），以及预备语句中所有参数占位符应使用的值。所提供的参数集必须与预备语句所需参数相匹配。（如果你在 Parse 消息中声明了任何 `void` 参数，那么在 Bind 消息中应为它们传递 NULL 值。）Bind 还会指定查询返回数据所使用的格式；格式既可以统一指定，也可以按列指定。响应要么是 BindComplete，要么是 ErrorResponse。

> **注意**
>
> 输出采用文本还是二进制格式，由 Bind 中给出的格式代码决定，而与涉及的 SQL 命令无关。在使用扩展查询协议时，游标声明中的 `BINARY` 属性并不起作用。

对命名预备语句对象的查询规划发生在处理 Parse 消息时。如果一条查询将用不同参数反复执行，那么发送一条包含参数化查询的 Parse 消息，随后再发送多条 Bind 和 Execute 消息可能是有益的。这样可以避免在每次执行时重新规划查询。

对于未命名预备语句，如果 Parse 消息没有指定参数类型，查询规划会延迟到收到该语句的第一条 Bind 消息时才进行。规划器在规划查询时会考虑 Bind 消息中提供的参数实际值。

> **注意**
>
> 从参数化查询生成的查询计划，
> 可能不如用实际参数值替换后的等价查询生成的计划高效。
> 在为分配给命名预备语句对象的参数化查询做规划时，

> 查询规划器无法基于实际参数值（例如索引选择性）做决策。
> 使用未命名语句可以避免这种可能的性能损失，
> 因为直到有了实际参数值之后才会进行规划。
>
> 如果在没有插入 Parse 的情况下，又收到了引用未命名预备语句对象的第二个或后续 Bind，查询不会被重新规划。第一次 Bind 消息中使用的参数值所产生的查询计划，可能只对可能的参数值中的一部分是高效的。要强制为新一组参数重新规划查询，
> 可以再发送一条 Parse 消息来替换未命名预备语句对象。

如果成功创建，命名 portal 对象会一直持续到当前事务结束，除非被显式销毁。未命名 portal 会在事务结束时销毁，或者在下一条把未命名 portal 作为目标的 Bind 语句发出时立即销毁。（注意，简单 Query 消息也会销毁未命名 portal。）有名 portal 在被另一条 Bind 消息重新定义之前必须显式关闭，而未命名 portal 则无此要求。有名 portal 也可以在 SQL command level, using `DECLARE CURSOR` and `FETCH`.

一旦 portal 存在，就可以使用 Execute 消息执行它。Execute 消息指定 portal 的名称（空字符串表示未命名 portal）以及一个最大的结果行计数（零表示“取回全部行”）。结果行计数只对包含返回行集命令的 portal 有意义；在其他情况下，命令总会执行到完成，而行计数会被忽略。Execute 的可能响应与通过简单查询协议发出的查询相同，只是 Execute 不会导致后端发送 ReadyForQuery 或 RowDescription。

如果 Execute 在 portal 执行完成之前终止（因为达到了非零的结果行计数），它会发送一条 PortalSuspended 消息；该消息表明前端应当针对同一个 portal 再发出一条 Execute 消息，以完成此次操作。在 portal 执行完成之前，不会发送表示源 SQL 命令结束的 Command Complete 消息。因此，一个 Execute 阶段总会以下列消息中的恰好一条结束：Command Complete、EmptyQueryResponse（如果 portal 是从空查询字符串创建的）、ErrorResponse 或 PortalSuspended。

每一组扩展查询消息完成后，前端都应发送一条 Sync 消息。这条无参数消息会让后端关闭当前事务，如果当前事务并不处在 `BEGIN/COMMIT` 事务块内（这里的“关闭”指的是：无错误则提交，有错误则回滚）。随后会发送一条 ReadyForQuery 响应。Sync 的目的是为错误恢复提供一个重新同步点。如果在处理任何扩展查询消息时检测到错误，后端会发出 ErrorResponse，然后持续读取并丢弃消息，直到遇到 Sync，再发送 ReadyForQuery 并回到正常的消息处理流程。（但要注意，如果是在处理 Sync 期间检测到错误，则不会跳过任何消息，这样就能保证每条 Sync 恰好对应一条 ReadyForQuery。）

> **注意**
>
> Sync 不会关闭由`BEGIN`打开的事务块。之所以能够检测出这种情况，是因为 ReadyForQuery 消息中包含事务状态信息。

除了这些基本的、必需的操作之外，在扩展查询协议里还有几种可选的操作可以使用。

Describe 消息（portal 变体）指定一个现有 portal 的名称（或者用空字符串表示未命名 portal）。响应要么是一条 RowDescription 消息，用于描述执行该 portal 时将返回的行；要么是一条 NoData 消息，如果该 portal 不包含会返回行的查询；要么是 ErrorResponse，如果不存在这样的 portal。

Describe 消息（语句变体）指定一个现有预备语句的名称（或者用空字符串表示未命名预备语句）。响应是一条 ParameterDescription 消息，用于描述该语句所需的参数，随后是一条 RowDescription 消息，用于描述该语句最终执行时将返回的行（如果该语句不返回行，则为 NoData 消息）。如果不存在这样的预备语句，则返回 ErrorResponse。请注意，由于还没有发出 Bind，

后端尚不知道返回列将使用什么格式；因此在这种情况下，RowDescription 消息中的格式代码字段将为零。

> **提示**
>
> 在大多数场景下，前端都应在发出 Execute 之前先发送某一种 Describe 变体，以确保它知道应如何解释接收到的结果。

Close 消息会关闭一个现有的预备语句或 portal，并释放相关资源。针对不存在的语句或 portal 发出 Close 并不算错误。响应通常是 CloseComplete，但如果在释放资源时遇到困难，也可能返回 ErrorResponse。请注意，关闭预备语句会隐式关闭所有基于该语句构造出的打开 portal。

Flush 消息本身不会产生任何特定输出，但会强制后端发送其输出缓冲区中所有尚待发送的数据。如果前端希望在发出更多命令之前检查某条扩展查询命令的结果，那么除了 Sync 之外，任何扩展查询命令之后都必须发送 Flush。如果没有 Flush，后端返回的消息会尽量合并成最少的数据包，以降低网络开销。

> **注意**
>
> 简单 Query 消息大致等价于一串 Parse、Bind、portal Describe、Execute、Close、Sync 操作，它们使用未命名预备语句和未命名 portal 对象，并且不带参数。不同之处在于，简单 Query 会接受包含多条 SQL 语句的查询字符串，并依次自动为每条语句执行绑定、描述和执行序列；另一个区别是它不会返回 ParseComplete、BindComplete、CloseComplete 或 NoData 消息。

43.2.4. 函数调用

函数调用子协议允许客户端请求直接调用数据库`pg_proc`系统目录中的任意函数。
客户端必须具有该函数的执行权限。

> **注意**
>
> 函数调用子协议是一个遗留的特性，在新代码里可能最好避免用它。
> 类似的结果可以通过设置一个执行`SELECT function($1, ...)`的预备语句得到。这样函数调用周期就可以用 Bind/Execute 代替。

函数调用周期由前端向后端发送一条 FunctionCall 消息来启动。后端随后根据函数调用的结果发送一条或多条响应消息，最后发送一条 ReadyForQuery 响应消息。ReadyForQuery 告知前端，可以安全地发送新的查询或函数调用。

来自后端的可能的响应消息是：

ErrorResponse

发生了一个错误。

FunctionCallResponse

函数调用完成并且在消息中返回一个结果（请注意函数调用协议只能处理单个标量结果，不能处理行类型或者结果集合）。

ReadyForQuery

函数调用处理完成。ReadyForQuery将总是被发送，
不管是成功完成处理还是发生一个错误。

NoticeResponse

发出了一条有关该函数调用的警告信息。通知是附加在其他响应上的，也就是说，
后端将继续处理该命令。

43.2.5. COPY 操作

COPY命令允许在服务器和客户端之间进行高速大批量数据传输。
拷贝入和拷贝出操作每个都把连接切换到一个独立的子协议中，并且持续到操作结束。

拷贝入模式（向服务器传输数据）在后端执行COPY FROM STDIN SQL 语句时启动。后端会向前端发送一条 CopyInResponse 消息。随后前端应发送零条或多条 CopyData 消息，构成一条输入数据流。（消息边界与行边界之间没有任何对应关系要求，尽管让它们对齐通常是合理的选择。）前端可以通过发送 CopyDone 消息来结束拷贝入模式（允许成功结束），也可以发送 CopyFail 消息（这会使COPY语句以错误失败）。然后后端会恢复到COPY开始之前的命令处理模式，也就是简单查询协议或扩展查询协议。接下来它会发送 CommandComplete（成功时）或 ErrorResponse（失败时）。

如果在拷贝入模式期间后端检测到错误（包括收到 CopyFail 消息），后端会发出一条 ErrorResponse 消息。如果COPY命令是通过扩展查询消息发出的，那么后端会从此开始丢弃前端消息，直到收到一条 Sync 消息，然后发出 ReadyForQuery 并恢复正常处理。如果COPY命令是在简单 Query 消息中发出的，那么该消息的剩余部分会被丢弃，并发送 ReadyForQuery。无论哪种情况，前端随后发出的任何 CopyData、CopyDone 或 CopyFail 消息都会被直接丢弃。

后端会忽略在拷贝入模式期间收到的 Flush 和 Sync 消息。收到任何其他非拷贝类型的消息都会构成错误，并按上述方式中止拷贝入状态。（Flush 和 Sync 的例外是为了方便那些总是在 Execute 消息之后发送 Flush 或 Sync、而不检查待执行命令是否为COPY FROM STDIN的客户端库。）

拷贝出模式（数据从服务器发出）是在后端执行一个COPY TO STDOUT语句时启动的。后端发出一个CopyOutResponse消息给前端，后面跟着零或者多个CopyData消息（总是每行一个），
然后跟着CopyDone。然后后端回退到它在COPY开始之前的命令处理模式，然后发送CommandComplete。前端不能中止传输（除非是关闭连接或者发出一个Cancel请求），
但是它可以抛弃不需要的CopyData和CopyDone消息。

在拷贝出模式中，如果后端检测到错误，那么它将发出一个Error
Response消息并且回到正常的处理。前端应该把收到Error
Response当作终止拷贝出模式的标志。

NoticeResponse 消息有可能穿插在 CopyData 消息之间；前端必须处理这种情况，并应准备好处理其他异步消息类型（参见第 43.2.6 节）。除此之外，可以将任何除 CopyData 或 CopyDone 以外的消息类型视为拷贝出模式的终止标志。

CopyInResponse 和 CopyOutResponse 消息包含一些字段，告知前端每行的列数以及每列所使用的格式代码。（就当前实现而言，给定 COPY 操作中的所有列都使用相同的格式，但消息的设计并不假定这一点。）

43.2.6. 异步操作

在若干情况下，后端会发送并非由前端命令流直接触发的消息。
前端必须随时准备处理这些消息，即使当前并未处于查询过程中。至少，
在开始读取查询响应之前应检查这些情况。

NoticeResponse 消息可能因外部活动而产生；例如，如果数据库管理员发起一次“快速”数据库关闭，后端会在关闭连接之前发送一条 NoticeResponse 说明这一事实。因此，前端应始终准备好接收并显示 NoticeResponse 消息，即使连接表面上处于空闲状态。

只要后端认为前端应当知晓的某个参数的当前有效值发生变化，就会生成 ParameterStatus 消息。最常见的情况是响应前端执行的`SET`命令，这种情况实际上是同步的；
但也可能是管理员修改了配置文件，然后向服务器发送SIGHUP信号，
从而导致参数状态发生变化。同样，如果某条`SET`命令被回滚，也会生成适当的 ParameterStatus 消息，用于报告当前生效的值。

目前有一组硬编码的参数会生成 ParameterStatus，它们是 `server_version`、`server_encoding`、`client_encoding`、`is_superuser`、`session_authorization`、`DateStyle`、`TimeZone`、`integer_datetimes` 和 `standard_conforming_strings`。（`server_encoding`、`TimeZone` 和 `integer_datetimes` 在 8.0 之前的发行版中不报告；`standard_conforming_strings` 在 8.1 之前的发行版中不报告。）注意，`server_version`、`server_encoding` 和 `integer_datetimes` 是启动后无法更改的伪参数。这个集合将来可能会改变，甚至可能变为可配置。因此，前端应直接忽略它不理解或不关心的参数的 ParameterStatus。

如果前端发出`LISTEN`命令，那么每当针对同一通知名执行`NOTIFY`命令时，后端都会发送一条 NotificationResponse 消息（不要与 NoticeResponse 混淆）。

> **注意**
>
> 目前，NotificationResponse只能在一个事务外面发送，
> 因此它将不会在一个命令响应序列中间出现，但是它可能正好在ReadyForQuery之前出现。不过，在前端逻辑中做上述假设是不明智的。
> 好的做法是在协议的任何点上都可以接受NotificationResponse。

43.2.7. 取消正在进行的请求

在一条查询正在处理的时候，前端可以请求取消该查询。
这种取消请求不是直接通过打开的连接发送给后端的，
这么做是因为实现的效率：我们不希望后端在处理查询的过程中不停地检查前端来的输入。
取消请求应该相对而言比较少见，所以我们把取消做得稍微笨拙一些，
以便不影响正常状况的性能。

要发出取消请求，前端会新建到服务器的连接，并发送 CancelRequest 消息，而不是新连接通常发送的 StartupMessage 消息。服务器处理该请求后便会关闭连接。出于安全原因，服务器不会直接回复取消请求消息。

除非CancelRequest消息包含在连接启动过程中传递给前端的相同的密钥数据（PID 和密钥），
否则它将被忽略。如果该请求匹配当前运行着的后端的PID和密钥，
则中止当前查询的处理（目前的实现里采用的方法是向正在处理该查询的后端进程发送一个特殊的信号）。

取消信号可能有效，也可能无效；例如，如果它在后端已经处理完查询之后才到达，
就不会起作用。如果取消生效，当前命令就会以一条错误消息提前终止。

这么做是对安全性和效率通盘考虑的结果，前端没有直接的方法获知一个取消请求是否成功。
它必须继续等待后端对查询的响应。发出一个取消仅仅是增加了当前查询快些结束的可能性，
同时也增加了当前查询会伴随着一条错误消息失败而不是成功执行的可能性。

由于取消请求是通过一条新的连接发送给服务器，而不是通过常规的前端/后端通信链路发送，因此发出取消请求的可以是任意进程，而不一定非要是要取消查询的那个前端。这为构建多进程应用提供了额外的灵活性，同时也带来了安全风险，因为未授权用户可能会尝试取消查询。通过要求在取消请求中提供动态生成的密钥，可以缓解这一安全风险。

43.2.8. 终止

通常优雅的终止过程是前端发送一条Terminate消息并且立刻关闭连接。一旦收到消息，后端马上关闭连接并且终止。

在少数情况下（比如一个管理员命令数据库关闭），后端可能在没有任何前端请求的情况下断开连接。在这种情况下，后端将在它断开连接之前尝试发送一个错误或者通知消息给出断开的原因。

其他终止场景来自各种故障，例如任一端发生 core dump、通信链路中断、消息边界同步丢失等。如果前端或后端看到连接意外关闭，就应清理并终止。若前端不想自行终止，也可以重新联系服务器以启动一个新的后端。如果收到无法识别的消息类型，同样建议关闭连接，因为这通常意味着消息边界同步已经丢失。

不管是正常还是不正常的终止，任何打开的事务都会回滚而不是提交。不过，我们应该注意的是如果一个前端在一个非`SELECT`查询正在处理的时候断开，那么后端很可能在发现断开之前先完成查询的处理。如果查询处于任何事务块之外（`BEGIN` ... `COMMIT`序列），那么其结果可能在发现连接断开之前被提交。

43.2.9. SSL 会话加密

如果PostgreSQL构建时包含了SSL支持，前端/后端通信就可以用SSL加密。这在攻击者可能能够捕获会话流量的环境中提供通信安全。关于用SSL加密PostgreSQL会话的更多信息，参见第 16.7 节。

要发起 SSL 加密连接，前端首先发送 SSLRequest 消息，而不是 StartupMessage。服务器随后返回包含 `S` 或 `N` 的单个字节，分别表示愿意或不愿意进行 SSL 通信。如果前端不满意此响应，可以在此时关闭连接。要在收到 `S` 后继续，应与服务器执行 SSL 启动握手（这里不作说明，它属于 SSL 规范的一部分）。如果握手成功，则继续发送通常的 StartupMessage。此时，Startup Message 及其后的所有数据都将经过 SSL 加密。要在收到 `N` 后继续，则发送通常的 Startup Message，并以不加密的方式继续通信。

前端还应准备好处理服务器对 SSLRequest 的 ErrorMessage 响应。这种情况只会发生在服务器版本早于 PostgreSQL 加入 SSL 支持的时候。此时必须关闭连接，但前端可以选择打开一个新连接并继续，而不请求 SSL。

如果建立连接是为了发送 CancelRequest 消息，也可以先发送 SSLRequest。

虽然协议本身没有提供让服务器强制使用SSL加密的方法，但管理员可以配置服务器，使其在认证检查中拒绝未加密的会话。

43.3. 消息数据类型

本节描述消息中使用的基础数据类型。

Int*n*(*i*)

以网络字节序（最高有效字节在前）表示的 *n* 位整数。如果指定了 *i*，则表示其精确值；否则表示该值可变。例如：Int16、Int32(42)。

Int*n*[*k*]

由 *k* 个 *n* 位整数组成的数组，每个整数都按网络字节序排列。数组长度 *k* 总是由消息中更早的某个字段确定。例如：Int16[M]。

String(*s*)

一个以空字符结尾的字符串（C 风格字符串）。字符串没有特定的长度限制。如果指定了 *s*，则表示其精确值；否则表示该值可变。例如：String、String("user")。

> **注意**
>
> 后端返回的字符串长度没有预定义的上限。前端的好做法是使用可扩展的缓冲区，这样凡是能装进内存的内容都能接受。如果做不到这一点，就读取整个字符串，然后丢弃固定大小缓冲区装不下的尾部字符。

Byte*n*(*c*)

精确 *n* 字节。如果字段宽度 *n* 不是常数，它总是可以由消息中更早的字段确定。如果指定了 *c*，则表示其精确值。例如：Byte2、Byte1('\n')。

43.4. 消息格式

本节描述每条消息的详细格式。每条消息都标记了可由前端（F）、后端（B）或双方（F&B）发送。请注意，虽然每条消息开头都带有字节计数，
但大多数消息格式都定义为无需参考该计数也能确定消息边界。
这一设计最初是出于历史原因（早期已废弃的协议 v2 没有显式长度字段），同时也有助于有效性校验。

AuthenticationOk (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(0)

表示认证成功。

AuthenticationKerberosV5 (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(2)

表示需要 Kerberos V5 认证。

AuthenticationCleartextPassword (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(3)

表示需要明文密码。

AuthenticationCryptPassword (B)

Byte1('R')

标识该消息为一个认证请求。

Int32(10)

消息内容的长度（以字节计，包括自身）。

Int32(4)

指定要求一个用 crypt() 加密的密码。

Byte2

加密密码时使用的盐值。

AuthenticationMD5Password (B)

Byte1('R')

将该消息标识为认证请求。

Int32(12)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(5)

表示需要经过 MD5 加密的密码。

Byte4

加密密码时使用的盐。

AuthenticationSCMCredential (B)

Byte1('R')

将该消息标识为认证请求。

Int32(8)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32(6)

表示需要 SCM 凭证消息。

BackendKeyData (B)

Byte1('K')

将此消息标识为取消请求密钥数据。如果前端希望以后能够发送 CancelRequest 消息，就必须保存这些值。

Int32(12)

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

此后端的进程 ID。

Int32

此后端的密钥。

Bind (F)

Byte1('B')

将该消息标识为 Bind 命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

目标 portal 的名称（空字符串选择未命名的 portal）。

String

源预备语句的名称（空字符串选择未命名的预备语句）。

Int16

后续参数格式码的数量（下文以 *C* 表示）。可以为零，表示没有参数，或者所有参数都使用默认格式（文本）；也可以为一，此时指定的格式码应用于所有参数；
还可以等于实际参数数量。

Int16[*C*]

参数格式码。目前每个格式码必须为零（文本）或一（二进制）。

Int16

后续参数值的数量（可以为零）。必须与查询所需的参数数量一致。

接下来，每个参数都有以下一对字段：

Int32

参数值的长度，以字节为单位（不包括此长度字段本身）。可以为零。特殊值 -1 表示 NULL 参数值。为 NULL 时，后面不再有值的字节。

Byte*n*

参数值，格式由对应的格式码指明。*n* 为上述长度。

最后一个参数之后是以下字段：

Int16

后续结果列格式码的数量（下文以 *R* 表示）。可以为零，表示没有结果列，或者所有结果列都应使用默认格式（文本）；也可以为一，
此时指定的格式码应用于所有结果列（如果有）；还可以等于查询实际的结果列数量。

Int16[*R*]

结果列格式码。目前每个格式码必须为零（文本）或一（二进制）。

BindComplete (B)

Byte1('2')

将该消息标识为 Bind 完成指示。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CancelRequest (F)

Int32(16)

消息内容以字节计的长度，包括自身。

Int32(80877102)

取消请求代码。此值的最高 16 位为 `1234`，最低 16 位为 `5678`。（为避免混淆，此代码不能与任何协议版本号相同。）

Int32

目标后端的进程 ID。

Int32

目标后端的密钥。

Close (F)

Byte1('C')

将该消息标识为 Close 命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

“`S`”表示关闭预备语句；“`P`”表示关闭 portal。

String

要关闭的预备语句或 portal 的名称（空字符串选择未命名的预备语句或 portal）。

CloseComplete (B)

Byte1('3')

将该消息标识为 Close 完成指示。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CommandComplete (B)

Byte1('C')

将该消息标识为命令完成响应。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

命令标签。通常是一个单词，用来标识已完成的 SQL 命令。

对于 `INSERT` 命令，标签是 `INSERT` *`oid rows`*，其中 *`rows`* 是插入的行数。如果 *`rows`* 为 1 且目标表具有 OID，则 *`oid`* 是插入行的对象 ID；否则 *`oid`* 为 0。

对于`DELETE`命令，标签是`DELETE` *`rows`*，其中*`rows`*表示删除的行数。

对于`UPDATE`命令，标签是`UPDATE` *`rows`*，其中*`rows`*是更新的行数。

对于`MOVE`命令，标签是`MOVE` *`rows`*，其中*`rows`*表示游标位置改变的行数。

对于`FETCH`命令，标签是`FETCH` *`rows`*，其中*`rows`*是从游标中检索出的行数。

CopyData (F & B)

Byte1('d')

标识消息为`COPY`数据。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte*n*

数据是`COPY`数据流的一部分。来自后端的消息始终对应单个数据行，
但来自前端的消息可能会任意划分数据流。

CopyDone (F & B)

Byte1('c')

将消息标识为`COPY`完成指示符。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

CopyFail (F)

Byte1('f')

将消息标识为`COPY`失败指示器。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

作为失败原因报告的错误消息。

CopyInResponse (B)

Byte1('G')

标识消息为开始复制输入的响应。前端此时必须发送复制输入数据（如果尚未准备好，应发送 CopyFail 消息）。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int8

0表示整体`COPY`格式是文本的（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式是二进制的（类似于DataRow格式）。更多信息请参见COPY。

Int16

要复制的数据中的列数（以下用*N*表示）。

Int16[*N*]

各列使用的格式代码。目前每个代码必须为零（文本）或一（二进制）。
如果整体复制格式为文本，则所有代码都必须为零。

CopyOutResponse (B)

Byte1('H')

标识消息为开始复制输出的响应。该消息之后会发送复制输出数据。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int8

0表示整体`COPY`格式是文本的（行由换行符分隔，列由分隔符分隔等）。
1表示整体复制格式是二进制的（类似于DataRow格式）。更多信息请参见COPY。

Int16

要复制的数据中的列数（以下用*N*表示）。

Int16[*N*]

各列使用的格式代码。目前每个代码必须为零（文本）或一（二进制）。
如果整体复制格式为文本，则所有代码都必须为零。

DataRow (B)

Byte1('D')

标识消息为数据行。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

后面跟着的列值的数量（可能为零）。

接下来，每列都有以下两个字段：

Int32

列值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示列值为 NULL，此时后面没有值字节。

Byte*n*

列的值，格式由相关的格式代码指示。*n*是上述长度。

Describe (F)

Byte1('D')

标识消息为描述命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

'S' 表示描述一个预备语句；或者 'P' 表示描述一个 portal。

String

要描述的预备语句或 portal 的名称（空字符串选择未命名的预备语句或 portal）。

EmptyQueryResponse (B)

Byte1('I')

标识消息为对空查询字符串的响应。（此消息替代 CommandComplete。）

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

ErrorResponse (B)

Byte1('E')

将消息标识为错误。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

消息体由一个或多个带标识的字段组成，最后以一个零字节终止。
字段可以按任意顺序出现。每个字段包含以下内容：

Byte1

一个用于标识字段类型的代码；如果为零，则这是消息终止符，后面没有字符串。
目前定义的字段类型列在第 43.5 节中。由于将来可能会添加更多的字段类型，前端应该静默地忽略未识别类型的字段。

String

字段值。

Execute (F)

Byte1('E')

标识消息为一个执行命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

要执行的 portal 的名称（空字符串选择未命名的 portal）。

Int32

如果 portal 包含返回行的查询，则这是最多返回的行数（否则忽略此值）。零表示“无限制”。

Flush (F)

Byte1('H')

将消息标识为Flush命令。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

FunctionCall (F)

Byte1('F')

标识消息为函数调用。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

指定要调用的函数的对象 ID。

Int16

后续参数格式代码的数量（以下用 C 表示）。可以为零，表示没有参数，或所有参数都采用默认格式（文本）；也可以为一，表示将指定的格式代码用于所有参数；
还可以等于实际参数数量。

Int16[C]

参数格式代码。每个目前必须是零（文本）或一（二进制）。

Int16

指定传递给函数的参数数量。

接下来，每个参数都有以下两个字段：

Int32

参数值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示参数值为 NULL，此时后面没有值字节。

Byte$n$

参数的值，以相关格式代码指示的格式表示。n是上述长度。

最后一个参数之后还有以下字段：

Int16

函数结果的格式代码。目前必须为零（文本）或一（二进制）。

FunctionCallResponse (B)

Byte1('V')

标识消息为函数调用结果。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

函数结果值的长度，以字节为单位（不包括本长度字段自身）。可以为零。特殊值 -1 表示函数结果为 NULL，此时后面没有值字节。

Byte$n$

函数结果的值，格式由相关的格式代码指示。n是上述长度。

NoData (B)

Byte1('n')

将消息标识为无数据指示器。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

NoticeResponse (B)

Byte1('N')

将消息标识为通知。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

消息体由一个或多个带标识的字段组成，最后以一个零字节终止。
字段可以按任意顺序出现。每个字段包含以下内容：

Byte1

一个用于标识字段类型的代码；如果为零，则这是消息终止符，后面没有字符串。
目前定义的字段类型列在第 43.5 节中。由于将来可能会添加更多的字段类型，前端应该静默地忽略未识别类型的字段。

String

字段值。

NotificationResponse (B)

Byte1('A')

标识消息为通知响应。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int32

发出通知的后端进程的进程 ID。

String

发出该通知的通知名。

String

通知进程传来的附加信息。

ParameterDescription (B)

Byte1('t')

标识消息为参数描述。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

语句使用的参数数量（可以为零）。

接下来，每个参数都有以下内容：

Int32

指定参数数据类型的对象 ID。

ParameterStatus (B)

Byte1('S')

标识消息为运行时参数状态报告。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

所报告的运行时参数的名称。

String

参数的当前值。

Parse (F)

Byte1('P')

将消息标识为解析命令。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

目标预备语句的名称（空字符串选择未命名的预备语句）。

String

要解析的查询字符串。

Int16

指定的参数数据类型的数量（可以为零）。请注意，
这不是查询字符串中可能出现的参数数量的指示，
而是前端希望为其预先指定类型的参数数量。

接下来，每个参数都有以下内容：

Int32

指定参数数据类型的对象 ID。此处填零等同于不指定类型。

ParseComplete (B)

Byte1('1')

将消息标识为解析完成指示器。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

PasswordMessage (F)

Byte1('p')

标识消息为密码响应。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

密码（如果需要，已加密）。

PortalSuspended (B)

Byte1('s')

标识消息为 portal 挂起指示器。请注意，仅当执行消息的行数限制达到时才会出现此消息。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

Query (F)

Byte1('Q')

标识消息为简单查询。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

String

查询字符串本身。

ReadyForQuery (B)

Byte1('Z')

标识消息类型。ReadyForQuery在后端准备好进行新的查询周期时发送。

Int32(5)

消息内容的长度，以字节为单位，包括此长度字段本身。

Byte1

当前后端事务状态指示器。可能的值为'I'，如果空闲（不在事务块中）；'T'，如果在事务块中；或'E'，如果在失败的事务块中（查询将被拒绝，直到块结束）。

RowDescription (B)

Byte1('T')

标识消息为行描述。

Int32

消息内容的长度，以字节为单位，包括此长度字段本身。

Int16

指定一行中的字段数量（可以为零）。

接下来，每个字段都有以下内容：

String

字段名称。

Int32

如果能够确定该字段是某个特定表的列，则为该表的对象 ID；否则为零。

Int16

如果能够确定该字段是某个特定表的列，则为该列的属性编号；否则为零。

Int32

字段数据类型的对象 ID。

Int16

数据类型大小（参见`pg_type.typlen`）。注意，负值表示可变宽度类型。

Int32

类型修饰符（参见`pg_attribute.atttypmod`）。修饰符的含义是特定于类型的。

Int16

字段所使用的格式代码。目前为零（文本）或一（二进制）。对于 Describe 的语句变体所返回的 RowDescription，格式代码尚未确定，始终为零。

SSLRequest (F)

Int32(8)

消息内容以字节计的长度，包括自身。

Int32(80877103)

SSL 请求代码。此值的最高 16 位为 `1234`，最低 16 位为 `5679`。（为避免混淆，此代码不能与任何协议版本号相同。）

StartupMessage (F)

Int32

消息内容以字节计的长度，包括自身。

Int32(196608)

协议版本号。最高 16 位是协议主版本号（本文档描述的协议为 3）。最低 16 位是协议次版本号（本文档描述的协议为 0）。

协议版本号后面跟着一对或多对参数名和参数值字符串。
最后一对名称/值对之后必须有一个零字节作为终止符。参数可以按任意顺序出现。`user` 是必需的，其他参数可选。每个参数指定为：

String

参数名。目前可识别的名称有：

`user`

要连接的数据库用户名。必需；没有默认值。

`database`

要连接的数据库。默认为用户名。

`options`

后端的命令行参数。（已弃用，建议设置单独的运行时参数。）

除上述参数外，还可以列出任何可在后端启动时设置的运行时参数。
这些设置将在后端启动期间应用（如果有的话，在解析命令行选项之后）。
这些值将作为会话默认值。

String

参数的值。

Sync (F)

Byte1('S')

将消息标识为同步命令。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

Terminate (F)

Byte1('X')

标识消息为终止。

Int32(4)

消息内容的长度，以字节为单位，包括此长度字段本身。

43.5. 错误和通知消息的字段

本节描述了ErrorResponse和NoticeResponse消息中可能出现的字段。
每个字段类型都有一个单字节的标识符。请注意，
任何给定的字段类型在消息中最多只能出现一次。

`S`

严重性：字段内容为 `ERROR`、`FATAL` 或 `PANIC`（在错误消息中），或 `WARNING`、`NOTICE`、`DEBUG`、`INFO` 或 `LOG`（在通知消息中），或这些内容的本地化翻译。始终存在。

`C`

代码：错误的 SQLSTATE 代码（见附录 A）。不可本地化。始终存在。

`M`

消息：主要的可读错误消息。应准确而简洁（通常为一行）。始终存在。

`D`

详细信息：可选的次要错误消息，提供有关该问题的更多细节。可能占多行。

`H`

提示：可选的建议，说明如何处理该问题。它与详细信息的区别在于，
它提供建议（可能并不合适），而非确凿事实。可能占多行。

`P`

位置：字段值是一个十进制ASCII整数，表示错误光标位置，作为原始查询字符串的索引。
第一个字符的索引为1，位置以字符而非字节计量。

`p`

内部位置：这与`P`字段的定义相同，
但在光标位置指向内部生成的命令而不是客户端提交的命令时使用。当此字段出现时，`q`
字段将始终出现。

`q`

内部查询：一个失败的内部生成命令的文本。例如，这可能是由PL/pg
SQL函数发出的SQL查询。

`W`

上下文：说明错误发生时的上下文。
目前包括活跃的过程语言函数和内部生成的查询的调用栈回溯。回溯每行一个条目，
最近的调用在前。

`F`

文件：报告错误的源代码位置所在的文件名。

L

行号：报告错误的源代码位置所在的行号。

R

例程：报告错误的源代码例程的名称。

客户端负责按自身需要格式化显示的信息；尤其应根据需要将长行断开。
错误消息字段中的换行符应视为段落分隔符，而非行分隔符。

43.6. 协议 2.0 以来的变更总结

本节提供一份简要的变更清单，供准备将现有客户端库更新到协议 3.0 的开发者参考。

初始启动包采用灵活的字符串列表格式，取代了固定格式。注意，
运行时参数的会话默认值现在可以直接在启动包中指定。（实际上，以前也能通过 options 字段实现，但由于 options 的宽度有限，且无法用引号保护值中的空白字符，这种方法并不稳妥。）

现在，所有消息都在消息类型字节之后紧跟一个长度计数（启动包除外，它没有类型字节）。
另请注意，PasswordMessage 现在也有类型字节。

ErrorResponse 和 NoticeResponse（'E' 和 'N'）消息现在包含多个字段，客户端代码可以利用这些字段组合出所需详细程度的错误消息。注意，各字段通常不会以换行符结束，
而旧协议发送的单个字符串总是以换行符结束。

ReadyForQuery（'Z'）消息包含一个事务状态指示器。

BinaryRow 与 DataRow 消息类型不再有区别；单一的 DataRow 消息类型用于返回所有格式的数据。注意，DataRow 的布局已经改变，使其更容易解析。另外，二进制值的表示方式也已改变，不再直接取决于服务器内部的表示方式。

新增了“扩展查询”子协议，其中增加了前端消息类型 Parse、Bind、Execute、Describe、Close、Flush 和 Sync，以及后端消息类型 ParseComplete、BindComplete、PortalSuspended、ParameterDescription、NoData 和 CloseComplete。现有客户端不必关注这个子协议，但使用它可能有助于改善性能或功能。

COPY 数据现在封装在 CopyData 和 CopyDone 消息中。COPY 期间的错误恢复已有明确定义的方法。不再需要特殊的最后一行“\.”，COPY OUT 期间也不再发送它。（在 COPY IN 期间仍会将它识别为终止符，但这种用法已弃用，最终将被移除。）现已支持二进制 COPY。CopyInResponse 和 CopyOutResponse 消息包含了表示列数及各列格式的字段。

FunctionCall 和 FunctionCallResponse 消息的布局已经改变。FunctionCall 现在支持向函数传递 NULL 参数，也可以使用文本或二进制格式传递参数和取得结果。由于它不再提供对服务器内部数据表示的直接访问，因此也不再有理由将 FunctionCall 视为潜在的安全漏洞。

在连接启动期间，后端会为其认为客户端库关注的所有参数发送 ParameterStatus（'S'）消息。之后，只要这些参数中任何一个的当前值发生变化，就会发送一条 ParameterStatus 消息。

RowDescription（'T'）消息针对所描述行中的每一列，增加了表 OID 和列编号字段，并显示各列的格式代码。

后端不再生成 CursorResponse（'P'）消息。

NotificationResponse（'A'）消息增加了一个额外的字符串字段，该字段目前为空，
但将来可能携带从 NOTIFY 事件发送方传来的附加数据。

EmptyQueryResponse（'I'）消息以前包含一个空字符串参数，现在已将其移除。

第 44 章 PostgreSQL 编码约定

44.1. 格式

源代码格式使用 4 列制表宽度，并保留制表符（即，不会把制表符展开为空格）。每个逻辑缩进层级都对应再增加一个制表位。布局规则（花括号位置等）遵循 BSD 约定。

尽管提交的补丁并不绝对必须遵循这些格式规则，但这样做是个好主意。你的代码会经过 pgindent 处理，因此按另一套格式约定把它写得再漂亮也没有意义。

对于 Emacs，请把以下内容（或类似内容）加入你的 `~/.emacs` 初始化文件：

```
;; check for files with a path containing "postgres" or "pgsql"
(setq auto-mode-alist
  (cons '("\\(postgres\\|pgsql\\).*\\.[ch]\\'" . pgsql-c-mode)
        auto-mode-alist))
(setq auto-mode-alist
  (cons '("\\(postgres\\|pgsql\\).*\\.cc\\'" . pgsql-c-mode)
        auto-mode-alist))

(defun pgsql-c-mode ()
  ;; sets up formatting for PostgreSQL C code
  (interactive)
  (c-mode)
  (setq-default tab-width 4)
  (c-set-style "bsd")             ; set c-basic-offset to 4, plus
 other stuff
  (c-set-offset 'case-label '+)   ; tweak case indent to match PG
 custom
  (setq indent-tabs-mode t))      ; make sure we keep tabs when
 indenting
```

对于 vi，你的 `~/.vimrc` 或等价的文件应包含以下内容：

```
set tabstop=4
```

或者在 vi 中等效地使用

```
:set ts=4
```

文本浏览工具 more 和 less 可以这样调用：

```
more -x4
less -x4
```

以便正确显示制表符。

44.2. 在服务器内部报告错误

服务器代码内生成的错误、警告和日志消息应使用 `ereport`，或其更老的近亲 `elog` 来创建。这一函数的用法相当复杂，因此需要一些解释。

每条消息都必须包含两个元素：严重性级别（范围从 `DEBUG` 到 `PANIC`）以及主消息文本。此外还可以有可选元素，其中最常见的是遵循 SQL 规范 SQLSTATE 约定的错误标识符代码。`ereport`

本身只是一个包装函数，主要为了语法上的便利，使消息生成在 C 源代码中看起来像一次函数调用。ereport 唯一直接接受的参数是严重性级别。主消息文本以及任何可选消息元素都是通过在 ereport 调用中调用辅助函数（例如 errmsg）来生成的。

对ereport的一次典型调用可能如下：

```
ereport(ERROR,
        (errcode(ERRCODE_DIVISION_BY_ZERO),
         errmsg("division by zero")));
```

这指定了错误严重性级别ERROR（一种普通错误）。errcode调用使用定义在src/include/utils/errcodes.h中的一个宏指定 SQLSTATE 错误代码。errmsg调用提供主消息文本。注意辅助函数调用外面多了一层圆括号 – 它们虽然烦人，但在语法上是必需的。

这里有一个更复杂的示例：

```
ereport(ERROR,
        (errcode(ERRCODE_AMBIGUOUS_FUNCTION),
         errmsg("function %s is not unique",
                func_signature_string(funcname, nargs,
                                      actual_arg_types)),
         errhint("Unable to choose a best candidate function. "
                 "You may need to add explicit typecasts.")));
```

这展示了如何用格式代码把运行时值嵌入消息文本中。此外还提供了一条可选的“提示”消息。

适用于 ereport 的辅助例程有：

- errcode(sqlerrcode) 为该条件指定 SQLSTATE 错误标识符代码。如果不调用这个例程，则默认错误标识符在错误严重性级别为 ERROR 或更高时为 ERRCODE_INTERNAL_ERROR，在错误级别为 WARNING 时为 ERRCODE_WARNING，否则（对于 NOTICE 及以下）为 ERRCODE_SUCCESSFUL_COMPLETION。虽然这些默认值常常很方便，但在省略 errcode() 调用之前，始终要先想想它们是否合适。

- errmsg(const char *msg, ...) 指定主错误消息文本，以及可能要插入其中的运行时值。插入项通过 sprintf 风格的格式代码指定。除了 sprintf 接受的标准格式代码外，还可以使用格式代码 %m 插入 strerror 为 errno 当前值返回的错误消息。[1] %m 不需要在 errmsg 的参数列表中有任何对应项。注意，在处理格式代码之前，消息字符串会先经过 gettext 以便可能进行本地化。

- errmsg_internal(const char *msg, ...) 与 errmsg 相同，只是消息字符串不会被翻译，也不会被收入国际化消息字典。这应当用于那些“不可能发生”、大概不值得为之投入翻译精力的情况。

- errdetail(const char *msg, ...) 提供一条可选的“详情”消息；当有额外信息但似乎不适合放在主消息中时，可使用它。消息字符串的处理方式与 errmsg 完全相同。

- errhint(const char *msg, ...) 提供一条可选的“提示”消息；当要就如何修复问题给出建议（而非关于出了什么问题的事实性详情）时，可使用它。消息字符串的处理方式与 errmsg 完全相同。

- errcontext(const char *msg, ...) 通常不会直接在 ereport 消息处调用；它用于 error_context_stack 回调函数中，提供错误发生时的上下文信息，例如 PL 函数中的当前

[1] 也就是说，是到达 ereport 调用点时的那个值；辅助报告例程内部对 errno 的更改不会影响它。如果显式写出 strerror(errno) 作为 errmsg 的参数列表内容，就不是这样了；因此不要这么做。

位置。消息字符串的处理方式与 `errmsg` 完全相同。与其他辅助函数不同，它在每次 `ereport` 调用中可以被调用多次；这样提供的 successive 字符串会用换行符分隔拼接起来。

- `errposition(int cursorpos)` 指定错误在查询字符串中的文本位置。目前，它只对在查询处理的词法和语法分析阶段检测到的错误有用。
- `errcode_for_file_access()` 是一个便捷函数，用于为与文件访问相关的系统调用失败选择合适的 SQLSTATE 错误标识符。它使用保存下来的 `errno` 来确定要生成哪种错误代码。通常应将它与主错误消息文本中的 `%m` 结合使用。
- `errcode_for_socket_access()` 是一个便捷函数，用于为与套接字相关的系统调用失败选择合适的 SQLSTATE 错误标识符。

还有一个较旧的函数`elog`，至今仍被大量使用。一个`elog`调用：

```
elog(level, "format string", ...);
```

完全等价于：

```
ereport(level, (errmsg_internal("format string", ...)));
```

注意，SQLSTATE 错误代码总会取默认值，而且消息字符串不会被翻译。因此，`elog`只应用于内部错误和低层调试日志。凡是普通用户可能感兴趣的消息，都应通过`ereport`。尽管如此，系统中仍有足够多的内部“不可能发生”错误检查，因此`elog`依然被广泛使用；对这类消息来说，由于记法更简洁，它更受青睐。

关于如何编写良好的错误消息，可参见第 44.3 节。

44.3. 错误消息风格指南

提供这份风格指南，是希望在 PostgreSQL 生成的所有消息中保持一致、对用户友好的风格。

什么放在哪里

主消息应简短、陈述事实，并避免提及具体函数名等实现细节。“简短”的意思是“在正常情况下应能放在一行里”。如果为了保持主消息简短，或者你觉得有必要提及特定失败的系统调用之类的实现细节，可使用详情消息。主消息和详情消息都应当陈述事实。提示消息则用于给出如何修复问题的建议，尤其是在该建议未必总是适用时。

例如，不要写：

```
IpcMemoryCreate: shmget(key=%d, size=%u, 0%o) failed: %m
(plus a long addendum that is basically a hint)
```

而要写：

```
Primary:    could not create shared memory segment: %m
Detail:     Failed syscall was shmget(key=%d, size=%u, 0%o).
Hint:       the addendum
```

原理：保持主消息简短，有助于使其切中要点，也便于客户端假定一行足以容纳错误消息，从而安排屏幕空间。详情消息和提示消息可以归入详细模式，或者放到弹出的错误详情窗口中。此外，为节省空间，详情消息和提示消息通常会从服务器日志中省略。最好避免提及实现细节，因为用户反正也不知道这些细节。

格式

不要在消息文本中写入任何特定的格式假设。应预期客户端和服务器日志会按照各自需要折行。
对于较长的消息，可以使用换行字符（\n）来表示建议的分段。不要让消息以换行结束。
不要使用制表符或其他格式控制字符。（在错误上下文显示中，
会自动添加换行以分隔函数调用等上下文层次。）

原理：消息不一定显示在终端类设备上。在 GUI 显示或浏览器中，这些格式指令充其量也只是被忽略。

引号

英文文本在适合引用时应使用双引号。其他语言的文本应一致地使用一种引号，
这种引号应符合该语言的出版习惯以及其他程序的计算机输出习惯。

原理：选择双引号而不是单引号多少有些武断，但通常更符合首选用法。有人建议按照 SQL 约定，根据对象类型来选择引号的种类（即字符串用单引号，标识符用双引号）。
但这是语言内部的技术问题，很多用户甚至并不熟悉；它也无法扩展到其他类型的被引术语，
更无法翻译到其他语言中，而且本身也没什么意义。

引号的使用

对文件名、用户提供的标识符以及其他可能包含单词的变量，始终使用引号界定。
对于不会包含单词的变量（例如操作符名），则不要加引号。

后端中有些函数会在需要时自行给其输出加上双引号（例如 `format_type_be()`）。不要再对这类函数的输出额外加引号。

原理：对象的名称在嵌入消息时可能造成歧义。对于插入的名称从哪里开始、到哪里结束，
要保持一致。但也不要用不必要或重复的引号把消息弄得杂乱无章。

语法和标点

主错误消息与详情/提示消息的规则不同：

主错误消息：首字母不要大写。消息末尾不要加句号。更不要想着在消息末尾加感叹号。

详情消息和提示消息：使用完整句子，并且每条都用句号结束。句子的第一个词要首字母大写。
如果后面还有一句，在句号后放两个空格（针对英文文本；对其他语言可能并不合适）。
错误上下文字符串：首字母不要大写，字符串末尾不要加句号。
上下文字符串通常不应是完整句子。

原理：避免使用标点，能让客户端应用更容易把消息嵌入各种语法上下文中。
主消息往往本来也不是语法完整的句子。（而如果它们长到不止一句，
就应当拆分成主消息和详情消息。）不过，详情消息和提示消息更长，
也可能需要包含多个句子。为了保持一致，即便只有一个句子，它们也应遵循完整句子的风格。

大写与小写

消息措辞应使用小写，包括主错误消息的首字母。如果消息中出现 SQL 命令或关键字，则使用大写形式。

原理：这样更容易让所有消息看起来一致，因为有些消息是完整句子，而有些则不是。

避免被动语态

使用主动语态。有施动者时，用完整句子（“A could not do B”）。如果施动者就是程序本身，则使用无主语的电报式写法，但不要把程序写成“I”。

原理：程序不是人。不要假装它是。

现在时与过去时

如果一次尝试做某事失败了，但下次仍可能成功（也许在修复某个问题之后），就使用过去时。如果失败显然是永久性的，就使用现在时。

下面两种句式在语义上有明显差别：

```
could not open file "%s": %m
```

和：

```
cannot open file "%s"
```

第一种表示尝试打开文件失败了。消息应给出原因，例如“磁盘已满”或“文件不存在”。过去时更合适，因为下次磁盘可能就不满了，或者所请求的文件可能已经存在。

第二种形式表示，程序中根本不存在打开该命名文件的功能，或者从概念上就不可能。现在时更合适，因为这种情况会无限期持续下去。

原理：诚然，普通用户未必能仅凭消息的时态得出什么重要结论，但既然语言提供了语法，我们就应当正确使用它。

对象的类型

在引用对象名称时，要说明它是什么类型的对象。

原理：否则没人会知道“foo.bar.baz”指的是什么。

括号

方括号只应用于：（1）命令概要中表示可选参数，或（2）表示数组下标。

原理：其他任何用法都不符合广为人知的习惯用法，而且会让人困惑。

组装错误消息

当消息中包含其他地方生成的文本时，应按以下方式嵌入：

```
could not open file %s: %m
```

原理：要把这种文本拼进一个流畅的单句中，同时照顾所有可能的错误代码会很困难，因此需要某种标点。也有人建议把嵌入的文本放在圆括号里，但如果嵌入的文本很可能是消息中最重要的部分（而这往往确实如此），那样就显得不自然了。

错误的原因

消息总应说明错误发生的原因。例如：

```
BAD:    could not open file %s
BETTER: could not open file %s (I/O failure)
```

如果不知道原因，最好去修复代码。

函数名

不要在错误文本中包含报告该错误的例程名。需要时，我们有其他机制可以找出它，而且对大多数用户来说，这项信息没有帮助。如果去掉函数名后错误文本就不够清楚，请重新措辞。

```
BAD:    pg_atoi: error in "z": can't parse "z"
BETTER: invalid input syntax for integer: "z"
```

也要避免提及被调用函数的名称；应改为说明代码想要做什么：

```
BAD:    open() failed: %m
BETTER: could not open file %s: %m
```

如果确实有必要，可在详情消息中提及系统调用。（在某些情况下，在详情消息中提供传给该系统调用的实际值也许是合适的信息。）

原理：用户并不知道那些函数都做了什么。

要避免的棘手词汇

Unable. “Unable” 几乎就是被动语态。更好的做法是视情况使用 “cannot” 或 “could not” 。

Bad. 像 “bad result” 这样的错误消息，实在很难被合理理解。最好写出结果为什么是 “bad” ，例如 “invalid format” 。

Illegal. “Illegal” 指的是违法，其余情况应使用 “invalid” 。更好的是，直接说明为什么无效。

Unknown. 尽量避免 “unknown” 。考虑 “error: unknown response” 。如果你都不知道响应是什么，又如何知道它有错？ “Unrecognized” 往往是更好的选择。此外，一定要包含被抱怨的值。

```
BAD:    unknown node type
BETTER: unrecognized node type: 42
```

Find vs. Exists. 如果程序为了定位某个资源使用了非平凡算法（例如路径搜索），而该算法失败了，那么说程序没能 “find” 该资源是公平的。另一方面，如果资源的预期位置已知，但程序无法在那里访问它，那么就应说该资源并不 “exist” 。在这种情况下使用 “find” 显得含糊，也会混淆问题。

正确拼写

单词应完整拼写。例如，避免使用：

- spec
- stats
- parens

- auth
- xact

原理：这会提升一致性。

本地化

请记住，错误消息文本需要被翻译成其他语言。请遵循第 45.2.2 节中的指导，避免给翻译者制造麻烦。

第 45 章 本地语言支持

Peter Eisentraut

45.1. 给翻译者

PostgreSQL 程序（服务器和客户端）可以用你所偏好的语言发出消息 – 前提是这些消息已经被翻译。创建和维护翻译后的消息集，需要那些精通自己语言并愿意为 PostgreSQL 项目作出贡献的人的帮助。做这件事完全不必是程序员。本节说明如何提供帮助。

45.1.1. 要求

我们不会评判你的语言能力 – 本节讨论的是软件工具。理论上，你只需要一个文本编辑器。但这只有在你不打算试用自己翻译的消息时才成立，而这种情况不太可能发生。
配置你的源代码树时，请务必使用 `--enable-nls` 选项。这还会检查 libintl 库和 `msgfmt` 程序，而终端用户无论如何也需要它们。要试用你的工作成果，请遵循安装说明中适用的部分。

如果你想开始一项新的翻译工作，或者想执行一次消息目录合并（后文会描述），则分别需要 GNU 兼容实现的 `xgettext` 和 `msgmerge` 程序。今后我们会尽量做到：如果你使用打包的源码发布版，就不需要 `xgettext` 了。（如果从 CVS 工作，你仍然需要它。）目前推荐使用 GNU Gettext 0.10.36 或更高版本。

你本地的 gettext 实现应该自带文档。其中有些内容可能与下文重复，但要了解更多细节，你应当去查阅那些文档。

45.1.2. 概念

原始（英语）消息及其（可能存在的）译文对应项，以成对形式保存在消息目录中；
每个程序以及每种目标语言各有一份消息目录（尽管相关程序可以共享同一个消息目录）。
消息目录有两种文件格式：第一种是“PO”文件（Portable Object，可移植对象），它是带有特殊语法的纯文本文件，由翻译者编辑。第二种是“MO”文件（Machine Object，机器对象），它是根据相应的 PO 文件生成的二进制文件，在国际化后的程序运行时使用。翻译者不需要处理 MO 文件；事实上几乎没有人需要。

消息目录文件的扩展名，毫不意外地是 `.po` 或 `.mo`。基本名则视情况而定，要么是其所对应的程序名，要么是该文件所对应的语言。这多少有些令人困惑。例子包括 `psql.po`（psql 的 PO 文件）或 `fr.mo`（法语的 MO 文件）。

PO 文件的格式示例如下：

```
# comment

msgid "original string"
msgstr "translated string"

msgid "more original"
msgstr "another translated"
"string can be broken up like this"

...
```

msgid 行是从程序源代码中提取出来的。（并非一定如此，但这是最常见的做法。）msgstr 行起初为空，由翻译者填入有意义的字符串。字符串可以包含 C 风格的转义字符，也可以像上例那样跨多行延续。（下一行必须从行首开始。）

字符引入注释。如果 # 字符后面紧跟空白，那么这就是由翻译者维护的注释。也可能存在自动注释，即 # 后紧跟非空白字符的注释。这些注释由各种处理 PO 文件的工具维护，目的是帮助翻译者。

```
#. automatic comment
#: filename.c:1023
#, flags, flags
```

#. 风格的注释是从使用该消息的源文件中提取出来的。程序员可能已经为翻译者插入了信息，例如期望的对齐方式。#: 注释指出该消息在源代码中使用的精确位置。翻译者不必查看程序源代码，但如果对正确译法有疑问，也可以去看。#, 注释包含以某种方式描述该消息的标志。目前有两种标志：如果某条消息可能因程序源代码的更改而过期，则会设置 `fuzzy`。翻译者随后可以核实这一点，并在可能时移除 fuzzy 标志。注意，fuzzy 消息不会提供给终端用户。另一个标志是 `c-format`，它表示该消息是一个 `printf` 风格的格式模板。这意味着译文也应该是一个格式字符串，并具有相同数量和类型的占位符。有一些工具会据此验证格式字符串，它们就是根据 c-format 标志来检查的。

45.1.3. 创建和维护消息目录

好，那么该如何创建一个“空白”消息目录呢？首先，进入包含你想翻译其消息的程序的目录。如果那里有一个 `nls.mk` 文件，就说明这个程序已经为翻译做好准备了。

如果已经有一些 `.po` 文件，那么就说明已经有人做过部分翻译工作。这些文件命名为 *`language`*`.po`，其中 *`language`* 是 ISO 639-1 两字母语言代码（小写）[1]，例如法语对应 `fr.po`。如果某种语言确实需要进行多套翻译工作，这些文件也可以命名为 *`language_region`*`.po`，其中 *`region`* 是 ISO 3166-1 两字母国家代码（大写）[2]，例如巴西葡萄牙语对应 `pt_BR.po`。如果你找到了自己想要的语言，就可以直接开始编辑那个文件。

如果需要启动一项新的翻译工作，请先运行命令：

```
gmake init-po
```

这会创建文件*`progname`*`.pot`。（用`.pot`来与“正在使用中”的 PO 文件区分，`T`代表“模板”。）把这个文件复制为*`language`*`.po`并编辑它。为了让别人知道新语言已可用，还要编辑文件`nls.mk`，把语言（或语言和国家）代码加到形如下面的行中：

```
AVAIL_LANGUAGES := de fr
```

（当然也可以是其他语言。）

随着底层程序或库的变化，程序员可能会更改或添加消息。这种情况下不必从头开始，只需运行命令：

```
gmake update-po
```

该命令会创建一个新的空白消息目录文件（即你最初使用的 pot 文件），并把它与现有的 PO 文件合并。如果合并算法对某个消息没有把握，就会如上所述把它标记为“模糊”。新的 PO 文件以`.po.new`扩展名保存。

[1] http://lcweb.loc.gov/standards/iso639-2/englangn.html
[2] http://www.iso.org/iso/country_names_and_code_elements

45.1.4. 编辑 PO 文件

PO 文件可以用普通文本编辑器编辑。翻译者只应修改 msgstr 指令后引号之间的内容、添加注释，以及调整 fuzzy 标志。Emacs 里（不出意外地）有一个 PO 模式，相当有用。

PO 文件不必完全填满。如果没有可用译文（或译文为空），软件会自动回退到原始字符串。把不完整的翻译提交到源码树中也没有问题；这样别人就有机会接手你的工作。不过，完成一次合并之后，建议优先清除 fuzzy 条目。记住，fuzzy 条目不会被安装；它们只用来作为可能正确译文的参考。

编辑译文时要记住以下几点：

- 务必确保如果原文以换行结束，译文也同样以换行结束。制表符等也是如此。
- 如果原文是一个 `printf` 格式字符串，译文也必须如此。译文还需要按相同顺序保留相同的格式说明符。有时语言自身的规则会让这一点变得不可能，或者至少很别扭。在这种情况下，你可以像下面这样修改格式说明符：

  ```
  msgstr "Die Datei %2$s hat %1$u Zeichen."
  ```

  这样一来，第一个占位符实际上会使用参数列表中的第二个参数。*digits*`$` 必须紧跟在 `%` 后面，位于其他任何格式修饰符之前。（这个特性确实存在于 `printf` 函数族中。你以前可能没听说过它，因为在消息国际化之外，它几乎没有用武之地。）
- 如果原始字符串包含语言错误，就报告它（或者你自己在程序源代码中修正它），然后照常翻译。等程序源代码更新之后，修正过的字符串就可以被合并进来。如果原始字符串包含事实性错误，就报告它（或者你自己修正它），并且不要翻译它。相反，你可以在 PO 文件里为该字符串加上一条注释。
- 保持原始字符串的风格和语气。特别是，那些不是完整句子的消息（`cannot open file %s`）通常不应以大写字母开头（如果你的语言区分大小写），也不应以句号结尾（如果你的语言使用标点符号）。阅读第 44.3 节可能会有帮助。
- 如果你不知道某条消息是什么意思，或者它有歧义，就到开发者邮件列表里提问。很可能说英语的终端用户也同样看不懂它，或者会觉得它有歧义，因此最好直接改进那条消息。

45.2. 给程序员

45.2.1. 实现机制

本节描述如何在 PostgreSQL 发行版中的程序或库里实现本地语言支持。目前它只适用于 C 程序。

为程序添加 NLS 支持

1. 把下面的代码插入到该程序的启动序列中：

   ```
   #ifdef ENABLE_NLS
   #include <locale.h>
   #endif

   ...
   ```

```
#ifdef ENABLE_NLS
setlocale(LC_ALL, "");
bindtextdomain("progname", LOCALEDIR);
textdomain("progname");
#endif
```

（其中 *progname* 实际上可以自由选择。）

2. 凡是遇到可能需要翻译的消息，都需要插入对 `gettext()` 的调用。例如：

   ```
   fprintf(stderr, "panic level %d\n", lvl);
   ```

   将改成：

   ```
   fprintf(stderr, gettext("panic level %d\n"), lvl);
   ```

   （如果没有配置 NLS 支持，`gettext` 会被定义成一个空操作。）

   这样往往会增加很多杂乱内容。一个常用的简写是：

   ```
   #define _(x) gettext(x)
   ```

   如果该程序的大部分通信都是通过一个或少数几个函数完成的，例如后端中的 `ereport()`，另一种可行的解决办法是让这个函数在内部对所有输入字符串调用 `gettext`。

3. 在程序源代码所在目录中添加一个`nls.mk`文件。这个文件会作为 makefile 读取。这里需要设置以下变量：

   `CATALOG_NAME`

   程序名，即 `textdomain()` 调用中提供的名称。

   `AVAIL_LANGUAGES`

   已提供的翻译列表 – 最初为空。

   `GETTEXT_FILES`

   包含可翻译字符串的文件列表，也就是那些使用 `gettext` 或其他替代方案标记过的文件。最终，这会包含该程序几乎所有的源文件。如果这个列表太长，
   可以让第一个“文件”是一个 `+`，第二个词则是一个文件名，该文件每行包含一个文件名。

   `GETTEXT_TRIGGERS`

   为翻译者生成消息目录的工具需要知道哪些函数调用包含可翻译字符串。默认只认识 `gettext()` 调用。如果你使用了 `_` 或其他标识符，就需要在这里列出它们。如果可翻译字符串不是第一个参数，条目就需要写成 `func:2` 这样的形式（表示第二个参数）。

构建系统会自动处理消息目录的构建和安装。

45.2.2. 消息编写指南

下面列出一些便于翻译的消息编写指南。

- 不要像下面这样在运行时拼接句子：

```
printf("Files were %s.\n", flag ? "copied" : "removed");
```

句子中的词序在其他语言里可能完全不同。另外，即使你记得对每个片段都调用 gettext()，这些片段分开来看也可能难以妥善翻译。最好稍微重复一点代码，
让每条待翻译消息都成为一个语义完整的整体。只有数字、
文件名之类的运行时变量才应该在运行时插入消息文本中。

- 出于类似的原因，下面这样也行不通：

```
printf("copied %d file%s", n, n!=1 ? "s" : "");
```

因为它假定了复数形式的构造方式。如果你以为可以这样解决：

```
if (n==1)
    printf("copied 1 file");
else
    printf("copied %d files", n):
```

那就会失望了。有些语言的复数形式不止两种，而且规则相当特殊。
将来我们也许会针对这个问题提供解决方案，但目前最好从设计上彻底避开它。
你可以这样写：

```
printf("number of copied files: %d", n);
```

- 如果你想向翻译者传达某些信息，例如一条消息应如何与其他输出对齐，
就在该字符串出现之前放置一个以 `translator` 开头的注释，例如：

```
/* translator: This message is not what it seems to be. */
```

这些注释会被复制到消息目录文件中，这样翻译者就能看到它们。

第 46 章 编写过程语言调用处理器

凡是不是使用当前针对编译型语言的“版本 1”接口编写的函数，在被调用时都会经过该语言专用的调用处理器函数。这包括用户定义过程语言中的函数、用 SQL 编写的函数，以及使用版本 0 编译语言接口的函数。调用处理器负责以恰当的方式执行该函数，例如解释所提供的源文本。本章概述如何编写新的过程语言调用处理器。

过程语言的调用处理器是一个“普通”函数，必须使用诸如 C 这样的编译型语言、按照版本 1 接口编写，并在 PostgreSQL 中注册为不接受参数且返回 `language_handler` 类型。这个特殊伪类型会将该函数标识为调用处理器，并阻止它在 SQL 命令中被直接调用。

调用处理器与其他函数的调用方式相同：它接收一个指向 `FunctionCallInfoData struct` 的指针，其中包含参数值和被调用函数的信息，并应返回一个 `Datum` 结果（如果要返回 SQL 空值，还需要设置 `FunctionCallInfoData` 结构体的 `isnull` 字段）。调用处理器与普通被调用函数的区别在于，`FunctionCallInfoData` 结构体的 `flinfo->fn_oid` 字段包含的是实际要调用的函数的 OID，而不是调用处理器自身的 OID。调用处理器必须使用该字段确定应执行哪个函数。此外，传入的参数列表也是按照目标函数的声明设置的，而不是按照调用处理器的声明。

调用处理器必须从系统目录 `pg_proc` 中取出该函数的条目，并分析被调用函数的参数类型和返回类型。该函数的 `CREATE FUNCTION` 命令中的 `AS` 子句内容，保存在 `pg_proc` 对应行的 `prosrc` 列中。这里通常是一段过程语言源文本，但理论上也可以是其他内容，例如某个文件的路径名，或者任何能详细告诉调用处理器该做什么的信息。

同一个函数在执行一条 SQL 语句期间往往会被调用很多次。调用处理器可以利用 `flinfo->fn_extra` 字段，避免重复查找被调用函数的信息。该字段起初为 `NULL`，但调用处理器可以把它设置为指向与被调用函数有关的信息。在后续调用中，如果 `flinfo->fn_extra` 已经不是 `NULL`，就可以直接使用它并跳过信息查找步骤。调用处理器必须确保 `flinfo->fn_extra` 指向的内存至少能存活到当前查询结束，因为 `FmgrInfo` 数据结构可能会保留这么久。一种做法是在 `flinfo->fn_mcxt` 指定的内存上下文中分配这些额外数据；这类数据通常会与 `FmgrInfo` 本身具有相同的生命周期。不过，处理器也可以选择使用生命周期更长的内存上下文，以便跨查询缓存函数定义信息。

过程语言函数作为触发器调用时，不会按通常方式传入参数，但 `FunctionCallInfoData` 的 `context` 字段会指向一个 `TriggerData` 结构体，而不是像普通函数调用那样为 `NULL`。语言调用处理器应提供让过程语言函数获取触发器信息的机制。

下面是用 C 编写的过程语言处理器模板：

```
#include "postgres.h"
#include "executor/spi.h"
#include "commands/trigger.h"
#include "fmgr.h"
#include "access/heapam.h"
#include "utils/syscache.h"
#include "catalog/pg_proc.h"
#include "catalog/pg_type.h"

#ifdef PG_MODULE_MAGIC
PG_MODULE_MAGIC;
#endif

PG_FUNCTION_INFO_V1(plsample_call_handler);
```

```
Datum
plsample_call_handler(PG_FUNCTION_ARGS)
{
    Datum          retval;

    if (CALLED_AS_TRIGGER(fcinfo))
    {
        /*
         * 作为触发器过程调用
         */
        TriggerData    *trigdata = (TriggerData *) fcinfo->context;

        retval = ...
    }
    else
    {
        /*
         * 作为函数调用
         */

        retval = ...
    }

    return retval;
}
```

只需在省略号处添加几千行代码，就能完成这个调用处理器。

将处理器函数编译为可加载模块后（参见第 32.9.6 节），可用以下命令注册示例过程语言：

```
CREATE FUNCTION plsample_call_handler() RETURNS language_handler
    AS 'filename'
    LANGUAGE C;
CREATE LANGUAGE plsample
    HANDLER plsample_call_handler;
```

标准发行版中附带的各过程语言，在尝试编写自己的语言调用处理器时都是很好的参考资料。请查看源码树中的 `src/pl` 子目录。

第 47 章 遗传查询优化器

Martin Utesch, University of Mining and Technology

作者

由 Martin Utesch <utesch@aut.tu-freiberg.de> 为德国弗赖贝格矿业和技术大学自动控制研究所编写。

47.1. 将查询处理看成是一个复杂的优化问题

在所有关系操作符中，最难处理和优化的是连接。随着查询中连接数目的增加，可能的查询计划数量会呈指数增长。为了处理单个连接而支持多种连接方法（例如 PostgreSQL 中的嵌套循环、哈希连接和归并连接），以及作为关系访问路径的多种索引（例如 PostgreSQL 中的 R-树、B-树、哈希），也进一步增加了优化工作量。

当前的PostgreSQL优化器实现会在可选策略空间中执行近似穷举搜索。该算法最早在 “System R” 数据库中引入，能够产生近乎最优的连接顺序；但当查询中的连接数变得很大时，它可能耗费极大的时间和内存空间。因此，普通的PostgreSQL查询优化器并不适合需要连接大量表的查询。

德国弗赖贝格矿业和技术大学自动控制研究所遇到了上述问题，因为他们希望将 PostgreSQL DBMS 用作一个用于电网维护的基于知识的决策支持系统的后端。该 DBMS 需要为该基于知识系统的推理机处理大型连接查询。

在探索可能查询计划空间方面的性能困难，催生了开发一种新优化技术的需求。

下文将介绍一种遗传算法的实现，它以适合涉及大量连接的查询的高效方式来解决连接顺序问题。

47.2. 遗传算法

遗传算法（GA）是一种通过非确定性的随机化搜索进行工作的启发式优化方法。优化问题的可能解集合被视为由若干个体组成的种群。个体对其环境的适应程度由其适应度表示。

一个个体在搜索空间中的坐标由染色体表示，本质上是一组字符串。基因是染色体的一个片段，它编码某个待优化参数的值。基因的典型编码可以是二进制或整数。

通过模拟重组、变异和选择这些进化操作，可以找到平均适应度高于前代的新一代搜索点。

根据comp.ai.genetic FAQ 中的说法，再怎么强调也不过分：GA并不是为了求解问题而进行的纯粹随机搜索。GA会使用随机过程，但其结果显然并非随机的（优于随机）。

图 47.1. 遗传算法的结构图

P(t)	时刻 t 的祖先代
P''(t)	时刻 t 的后代代

```
+=========================================+
|>>>>>>>>>>>  Algorithm GA  <<<<<<<<<<<<<<|
+=========================================+
| INITIALIZE t := 0                       |
+=========================================+
| INITIALIZE P(t)                         |
+=========================================+
| evaluate FITNESS of P(t)                |
+=========================================+
| while not STOPPING CRITERION do         |
|   +-------------------------------------+
|   | P'(t)  := RECOMBINATION{P(t)}       |
|   +-------------------------------------+
|   | P''(t) := MUTATION{P'(t)}           |
|   +-------------------------------------+
|   | P(t+1) := SELECTION{P''(t) + P(t)}  |
|   +-------------------------------------+
|   | evaluate FITNESS of P''(t)          |
|   +-------------------------------------+
|   | t := t + 1                          |
+===+=====================================+
```

47.3. PostgreSQL 中的遗传查询优化（GEQO）

GEQO模块把查询优化问题视为著名的旅行商问题（TSP）来处理。
可能的查询计划被编码为整数串。每个串表示查询中从一个关系到下一个关系的连接顺序。
例如，连接树

```
   /\
  /\ 2
 /\ 3
4  1
```

会被编码为整数串 '4-1-3-2'，这意味着先连接关系 '4' 和 '1'，再连接 '3'，最后连接 '2'；其中 1、2、3、4 是PostgreSQL优化器内部的关系 ID。

GEQO模块的部分内容改编自 D. Whitley 的 Genitor 算法。

PostgreSQL 中 GEQO 实现的具体特征包括：

- 采用稳态 GA（替换种群中适应度最低的个体，而不是整代替换），能够快速收敛到更优的查询计划。这对于在合理时间内处理查询至关重要；
- 采用边重组交叉，它特别适合在利用 GA 求解 TSP 时将边损失保持在较低水平；
- 不使用变异作为遗传操作符，因此无须借助修复机制来生成合法的 TSP 回路。

GEQO模块使PostgreSQL查询优化器能够通过非穷举搜索有效支持大型连接查询。

47.3.1. PostgreSQL GEQO的未来实现任务

为了改进遗传算法的参数设置，仍有一些工作要做。在文件`src/backend/optimizer/geqo/geqo_main.c`中的例程`gimme_pool_size`和`gimme_number_generations`里，我们必须为参数设置找到一种折中，以满足两个相互竞争的需求：

- 查询计划的最优性
- 计算时间

从更基础的层面看，用一个为 TSP 设计的 GA 算法来解决查询优化问题是否合适，也并不明确。在 TSP 情况下，与任何子串（部分巡回）相关的代价都独立于巡回的其余部分，但对于查询优化显然并非如此。因此，边重组交叉是否是最有效的变异过程，仍然值得怀疑。

47.4. 进一步阅读

下列资源包含关于遗传算法的更多信息：

- 《Hitch-Hiker 演化计算指南》[1]（comp.ai.genetic[2] FAQ）
- 《演化计算及其在艺术和设计中的应用》[3]，Craig Reynolds 著
- [elma04]
- [fong]

[1] http://www.cs.bham.ac.uk/Mirrors/ftp.de.uu.net/EC/clife/www/location.htm
[2] news://comp.ai.genetic
[3] http://www.red3d.com/cwr/evolve.html

第 48 章 索引访问方法接口定义

本章定义 PostgreSQL 核心系统与管理各个索引类型的索引访问方法之间的接口。除这里规定的内容之外，核心系统对索引一无所知，因此可以通过编写附加代码来开发全新的索引类型。

PostgreSQL 中所有索引在技术上都称为二级索引；也就是说，索引与它所描述的表文件在物理上是分离的。每个索引都存储为独立的物理关系，因此在 pg_class 目录中都有一个条目。索引的内容完全由其索引访问方法控制。实践中，
所有索引访问方法都会把索引划分为标准大小的页面，
以便使用常规的存储管理器和缓冲区管理器访问索引内容。（现有的所有索引访问方法还都使用第 50.3 节中描述的标准页面布局，而且它们都对索引元组头使用相同的格式；但这些决定并不是访问方法必须遵循的。）

索引本质上是从一些数据键值到行版本（元组）在索引父表中的元组标识符（TIDs）的映射。
一个 TID 由块号以及该块中的项号组成（见第 50.3 节）。这些信息足以从表中取出某个特定的行版本。索引并不直接知道在 MVCC 之下同一逻辑行可能会存在多个现存版本；对索引来说，每个元组都是一个独立对象，需要它自己的索引条目。因此，
对一行的更新总会为该行创建全新的索引条目，即使键值并未改变也是如此。
当死元组自身被回收时（通过清理），它们对应的索引条目也会被回收。

48.1. 索引的目录条目

每种索引访问方法都由系统目录 pg_am 中的一行描述（参见第 42.3 节）。pg_am 行的主要内容是一些指向 pg_proc 条目的引用，它们标识了该访问方法所提供的索引访问函数。这些函数的 API 将在本章后面定义。此外，pg_am 行还指定了访问方法的一些固定属性，例如它是否支持多列索引。目前没有任何专门用于创建或删除 pg_am 条目的支持；能够编写新访问方法的人，理应有能力自己插入合适的行。

要使一种索引访问方法真正可用，还必须在 pg_opclass、pg_amop 和 pg_amproc 中定义一个或多个操作符类。这些条目使规划器能够确定该访问方法的索引可以使用哪些查询限定条件。关于操作符类的说明见第 32.14 节；阅读本章前应先掌握这些内容。

单个索引由两个系统目录条目共同定义：一个 pg_class 条目把它描述为一个物理关系，另一个 pg_index 条目给出索引的逻辑内容，也就是它包含哪些索引列，以及由相关操作符类定义的这些列的语义。索引列（键值）既可以是底层表的简单列，也可以是基于表行的表达式。
索引访问方法通常并不关心索引键值来自何处（它拿到的总是预先计算好的键值），
但会非常关心 pg_index 中的操作符类信息。这两个目录条目都可以作为传递给索引上所有操作的 Relation 数据结构的一部分来访问。

pg_am 的某些标志列还有一些不那么直观的含义。amcanunique 的要求见第 48.5 节，amconcurrent 的要求见第 48.4 节。amcanmulticol 标志表明该访问方法支持多列索引，而 amoptionalkey 表明它允许在第一索引列上没有给出可索引限制子句时进行扫描。当 amcanmulticol 为假时，amoptionalkey 实质上表示该访问方法是否支持没有任何限制子句的全索引扫描。
支持多个索引列的访问方法必须支持在第一列之后省略任意一个或全部列限制条件的扫描；
但它们可以要求第一索引列必须出现某种限制条件，这通过把 amoptionalkey 设为假来表示。
amindexnulls 表明会为 NULL 键值创建索引条目。由于大多数可索引操作符都是严格的（strict），因此对 NULL 输入不可能返回 TRUE，所以乍看之下不存储 NULL 值的索引条目似乎很有吸引力：无论如何，这些条目似乎都不可能被索引扫描返回。然而，
当某个索引扫描对给定索引列没有限制子句时，这个论证就不成立了。实践中这意味着，
若索引把 amoptionalkey 设为真，就必须索引 NULL 值，因为规划器可能决定在完全没有扫描键的情况下使用这种索引。与此相关的另一个限制是，支持多个索引列的索引访问方法
必须支持对第一列之后各列中的 NULL 值建立索引，因为规划器会假定该索引可用于不限制这些

列的查询。例如，考虑一个在 (a,b) 上的索引以及查询 `WHERE a = 4`。系统会假定该索引可用于扫描满足 `a = 4` 的行；如果索引省略了 `b` 为 NULL 的行，这个假定就是错误的。不过，省略第一索引列为 NULL 的行是可以的。（GiST 目前就是这样做的。）因此，只有当索引访问方法对所有行（包括 NULL 值的任意组合）都建立索引时，才应把 `amindexnulls` 设为真。

48.2. 索引访问方法函数

索引访问方法必须提供如下索引构建和维护函数：

```
void
ambuild (Relation heapRelation,
         Relation indexRelation,
         IndexInfo *indexInfo);
```

构建新索引。索引关系已在物理上创建，但内容为空。必须填入访问方法所需的固定数据，以及表中所有已有元组对应的条目。通常，`ambuild`函数会调用`IndexBuildHeapScan()`扫描表中现有的元组，并计算需要插入索引的键。

```
bool
aminsert (Relation indexRelation,
          Datum *values,
          bool *isnull,
          ItemPointer heap_tid,
          Relation heapRelation,
          bool check_uniqueness);
```

将一个新元组插入已有索引。`values`和`isnull`数组给出要索引的键值，`heap_tid`则是要索引的 TID。如果访问方法支持唯一索引（其`pg_am.amcanunique`标志为 true），则`check_uniqueness`可能为真，此时访问方法必须验证不存在冲突的行；这也是访问方法通常唯一需要`heapRelation`参数的情形。详情参见第 48.5 节。如果插入了索引条目则结果为 TRUE，否则为 FALSE。（FALSE 结果并不表示错误条件，而是用于诸如索引方法拒绝索引 NULL 值之类的情况。）

```
IndexBulkDeleteResult *
ambulkdelete (Relation indexRelation,
              IndexBulkDeleteCallback callback,
              void *callback_state);
```

从索引中删除元组。这是一个“批量删除”操作，旨在通过扫描整个索引并检查每个条目是否应被删除来实现。可以调用传入的 `callback` 函数，其调用形式为 `callback(`*`TID`*`, callback_state) returns bool`，以确定由其引用的 TID 所标识的某个索引条目是否应删除。该函数必须返回 NULL，或者返回一个通过 palloc 分配的结构体，其中包含此次删除操作影响的统计信息。

```
IndexBulkDeleteResult *
amvacuumcleanup (Relation indexRelation,
                 IndexVacuumCleanupInfo *info,
                 IndexBulkDeleteResult *stats);
```

在一次`VACUUM`操作（一次或多次`ambulkdelete`调用）之后执行清理。索引访问方法不必提供此函数（如果不提供，`pg_am` 中的相应条目必须为零）。如果提供了此函数，它通常用于批量清理，例如回收空索引页面。`info` 提供一些额外的参数，例如统计报告的消息级别；`stats` 是最后一次 `ambulkdelete` 调用返回的结果。`amvacuumcleanup` 可以在返回之前替换或修改这个结构体。如果结果不是 NULL，它必须是一个通过 palloc 分配的结构体。其中的统计信息将在指定了 `VERBOSE` 时由 `VACUUM` 报告。

索引的目的当然是支持扫描那些匹配可索引 WHERE 条件的元组，这种条件常被称为限定词或扫描键。关于索引扫描的语义，将在下面的第 48.3 节中更详细地说明。索引访问方法必须提供的扫描相关函数如下：

```
IndexScanDesc
ambeginscan (Relation indexRelation,
             int nkeys,
             ScanKey key);
```

开始一次新扫描。key 数组（长度为 nkeys）描述该索引扫描的扫描键。结果必须是一个通过 palloc 分配的结构体。出于实现上的原因，索引访问方法必须通过调用 RelationGetIndexScan() 来创建这个结构体。大多数情况下，ambeginscan 除了做这次调用之外不会做太多工作；索引扫描启动中真正有意思的部分在 amrescan 中。

```
boolean
amgettuple (IndexScanDesc scan,
            ScanDirection direction);
```

沿给定方向（在索引中向前或向后）取出给定扫描中的下一个元组。若取得元组则返回 TRUE，若没有剩余匹配元组则返回 FALSE。返回 TRUE 时，元组 TID 存入scan结构体中。注意，“成功”仅表示索引中包含与扫描键匹配的条目，并不表示该元组一定仍存在于堆中，
或者能够通过调用方的快照测试。

```
boolean
amgetmulti (IndexScanDesc scan,
            ItemPointer tids,
            int32 max_tids,
            int32 *returned_tids);
```

在给定扫描中取出多个元组。如果扫描还应继续则返回 TRUE，如果没有剩余匹配元组则返回 FALSE。tids指向一个由调用方提供的、包含max_tids个ItemPointerData记录的数组，
本调用会把匹配元组的 TID 填入其中。*returned_tids被设为实际返回的 TID 数。即使返回值为 TRUE，这个数也可以小于max_tids，甚至为零。（这一规定允许访问方法在扫描中选择最高效的停止点，例如索引页边界。）amgetmulti和amgettuple
不能在同一次索引扫描中同时使用；使用amgetmulti时还有其他限制，详见第 48.3 节。

```
void
amrescan (IndexScanDesc scan,
          ScanKey key);
```

重新开始给定的扫描，并且可以使用新的扫描键（若要继续使用旧键，则为 key 传入 NULL）。注意，键的数量不允许改变。实践中，
这种重启特性通常用于嵌套循环连接选中了新的外层元组、因此需要新的键比较值，
但扫描键结构体保持不变的场景。RelationGetIndexScan() 也会调用此函数，因此它既用于索引扫描的初始设置，也用于重新扫描。

```
void
amendscan (IndexScanDesc scan);
```

结束扫描并释放资源。scan 结构体本身不应被释放，但访问方法内部获取的任何锁或引用（pin）都必须被释放。

```
void
ammarkpos (IndexScanDesc scan);
```

标记当前扫描位置。访问方法只需要为每次扫描支持一个被记住的扫描位置。

```
void
amrestrpos (IndexScanDesc scan);
```

将扫描恢复到最近一次标记的位置。

```
void
amcostestimate (PlannerInfo *root,
                IndexOptInfo *index,
                List *indexQuals,
                Cost *indexStartupCost,
                Cost *indexTotalCost,
                Selectivity *indexSelectivity,
                double *indexCorrelation);
```

估计索引扫描的代价。该函数在下文第 48.6 节中有完整描述。

按照惯例，索引访问方法函数的 `pg_proc` 条目应显示正确的参数个数，但把所有参数都声明为 `internal` 类型（因为大多数参数的类型对 SQL 而言是未知的，而且我们也不希望用户直接调用这些函数）。返回类型根据实际情况声明为 `void`、`internal` 或 `boolean`。

48.3. 索引扫描

在索引扫描中，索引访问方法负责返回它所知道的、匹配扫描键的所有元组的 TID。访问方法不负责从索引父表中实际取出这些元组，
也不负责判断它们是否能通过扫描的可见性测试或其他条件。

扫描键是如下形式的 `WHERE` 子句的内部表示：*`index_key operator constant`*。其中，索引键是该索引的一列，而操作符是与该索引列关联的操作符族成员之一。
一次索引扫描可以有零个或多个扫描键，这些扫描键会被隐式地以 AND 连接，也就是说，返回的元组预期应满足所有给定条件。

操作符族可以表明索引对于某个特定操作符是有损的；
这意味着索引扫描会返回所有通过扫描键的条目，并可能额外返回一些未通过扫描键的条目。
随后，核心系统的索引扫描机制会再次把该操作符应用到堆元组上，
以验证它是否真的应被选中。对于非有损操作符，索引扫描必须精确返回匹配条目的集合，
因为没有重检。

请注意，确保正确找出通过全部给定扫描键的所有且仅有的条目，
这项工作完全由访问方法负责。此外，核心系统只是简单地把所有与索引键和操作符族匹配的 `WHERE` 子句交给访问方法，而不会做任何语义分析来判断它们是否冗余或互相矛盾。例如，给定 `WHERE x > 4 AND x > 14`，其中 `x` 是一个 B-树索引列，那么由 B-树的 `amrescan` 函数负责识别第一个扫描键是冗余的并可被丢弃。`amrescan` 期间需要做多少预处理，取决于索引访问方法需要在多大程度上把扫描键归约成一种“正规化”形式。

`amgettuple` 函数有一个 `direction` 参数，它可以是 `ForwardScanDirection`（通常情况）或 `BackwardScanDirection`。如果 `amrescan` 之后的第一次调用指定了 `BackwardScanDirection`，那么匹配条件的索引条目集合就要按从后到前的方向扫描，而不是通常的从前到后，因此 `amgettuple` 必须返回索引中最后一个匹配元组，而不是通常情况下的第一个。（这只会发生在通过把 `pg_am.amorderstrategy`设为非零而声明支持有序扫描的访问方法上。）第一次调用之后，`amgettuple` 必须准备好从最近一次返回条目的位置继续按任一方向推进扫描。

访问方法必须支持在扫描中“标记”一个位置，并在之后返回到该标记位置。
同一个位置可能会被恢复多次。不过，每次扫描只需要记住一个位置；新的 `ammarkpos` 调用会覆盖先前标记的位置。

无论是扫描位置还是标记位置（如果有），在面对索引中的并发插入或删除时都必须保持一致。如果某个新插入的条目没有被一次扫描返回，
而假如该条目在扫描开始前就已存在则本应被返回，这也是可以接受的；同样，
即使某个条目在第一次扫描通过时未被返回，
后续重新扫描或反向扫描时返回它也是可以接受的。类似地，
并发删除也可能会或者不会反映在扫描结果中。重要的是，插入或删除本身不能导致扫描漏掉，
或者重复返回那些本身并未被插入或删除的条目。（对于没有设置 pg_am.amconcurrent 的索引类型，只需处理由执行扫描的同一后端所执行的插入或删除就足够了。但当 amconcurrent 为真时，还必须处理来自其他后端的插入或删除。）

除了使用 amgettuple 之外，索引扫描也可以用 amgetmulti 每次调用取出多个元组。这可以比 amgettuple 明显更高效，因为它可以避免在访问方法内部反复加锁/解锁。原则上，amgetmulti 的效果应与反复调用 amgettuple 相同，但我们施加了一些限制以简化问题。首先，amgetmulti不接受 direction 参数，因此它不支持反向扫描，也不支持扫描中途改变方向。其次，访问方法也不需要在 amgetmulti 扫描期间支持标记或恢复扫描位置。（这些限制代价不大，因为在 amgetmulti 扫描中使用这些特性本来就很困难：调整调用方缓冲的 TID 列表会非常复杂。）最后，amgetmulti 不保证对返回的元组加任何锁，其影响在第 48.4 节中有详细说明。

48.4. 索引锁定注意事项

索引访问方法可以自行选择是否支持多个进程并发更新同一个索引。如果该方法pg_am.amconcurrent标志为真，那么PostgreSQL 核心系统会在索引扫描期间对索引获取 AccessShareLock，并在更新索引时获取 RowExclusiveLock。由于这些锁类型彼此不冲突，所以访问方法必须自行处理它可能需要的任何细粒度锁定。对整个索引的排他锁只会在索引创建、
销毁或 REINDEX 期间获取。当 amconcurrent 为假时，PostgreSQL 仍会在索引扫描期间获取 AccessShareLock，但在任何更新期间获取 AccessExclusiveLock。这确保了更新者独占使用该索引。注意，这里隐含地假定了索引扫描是只读的；
在扫描期间可能修改索引的访问方法，仍须自行处理锁定以应对并发扫描的情况。

请记住，后端自身的锁永远不会冲突；因此，即使是并发型索引，
也必须准备好处理某个后端正在它自己扫描的索引中插入或删除条目的情况。
（这当然是支持使用索引查找要更新行的 UPDATE 所必需的。）

构建一种支持并发更新的索引类型，通常需要对所需行为进行大量而细致的分析。对于 B-树和 hash 索引类型，可以阅读 src/backend/access/nbtree/README 和 src/backend/access/hash/README 中涉及的设计决策。

除了索引自身内部一致性的要求外，并发更新还会带来父表（即堆）与索引之间一致性的问题。由于 PostgreSQL 把对堆的访问和更新与对索引的访问和更新分离开来，因此存在一些窗口期，在这些窗口期内索引可能与堆不一致。我们通过以下规则处理这个问题：

- 先创建新的堆条目，再创建它对应的索引条目。
  （因此并发索引扫描很可能看不到这个堆条目。这是可以接受的，
  因为索引读取者本来也不会关心未提交的行。但也请见第 48.5 节。）

- 当某个堆条目将要被删除（由 VACUUM 执行）时，必须先删除它的所有索引条目。

- 对于并发型索引，索引扫描必须在保存 amgettuple 最近一次返回条目的索引页面上保持钉住状态（pin），而 ambulkdelete 不能从被其他后端钉住的页面中删除条目。下面会解释为什么需要这条规则。

如果索引是并发型的，那么索引读取者就有可能在某个索引条目被 VACUUM 删除之前先看到它，然后在对应的堆条目已被 VACUUM 删除之后才到达那里。如果读取者到达时，该项号

仍未被重新使用，就不会造成严重问题，因为空项槽位会被 `heap_fetch()` 忽略。但如果第三个后端已经把这个项槽位重新用作别的东西呢？在使用 MVCC 兼容快照时不会有问题，因为该槽位的新占用者一定太新，无法通过快照测试。然而，对于非 MVCC 兼容的快照（例如 `SnapshotNow`），就可能错误地接受并返回一行实际上并不匹配扫描键的数据。我们可以要求在所有情况下都对堆行重新检查扫描键，以防范这种情形，但那样代价太高。于是，
我们改为把索引页面的钉住状态作为一种代理，
表示读取者从索引条目到匹配堆条目的访问过程可能仍“在进行中”。让 `ambulkdelete` 因这种钉住状态而阻塞，能够保证 `VACUUM` 不会在读取者处理完之前删除对应的堆条目。（对于非并发型索引，这种情况是不可能发生的，因为将要获取的索引级锁是相互冲突的。）
这种方案运行时开销很小，只有在真正发生冲突的少数情况下才会增加阻塞开销。

这种解决方案要求索引扫描是“同步的”：我们必须在扫描到相应索引条目之后立刻取出对应的堆元组。由于多种原因，这会比较昂贵。相反，“异步”扫描可以先从索引中收集许多 TID，再在稍后某个时刻访问堆元组，这样索引锁定的开销要小得多，也能实现更高效的堆访问模式。
根据前面的分析，对于非 MVCC 兼容快照，我们必须采用同步方式；而对于使用 MVCC 快照的查询，异步扫描则是可行的。

在 `amgetmultti` 索引扫描中，访问方法不会为任何返回的元组保持索引页面的钉住状态。因此，只有把这种扫描与 MVCC 兼容的快照一起使用才是安全的。

48.5. 索引唯一性检查

PostgreSQL 使用唯一索引来强制执行 SQL 唯一性约束；所谓唯一索引，就是不允许存在多个具有相同键值条目的索引。支持这一特性的访问方法会把 `pg_am.amcanunique` 设为真。（目前只有 B-树支持这一点。）

由于 MVCC 的存在，索引中在物理上总是必须允许存在重复条目：这些条目可能指向同一个逻辑行的连续版本。我们真正想强制的行为是，任何 MVCC 快照都不能同时包含两行拥有相同索引键的记录。在向唯一索引插入一行新数据时，必须检查以下几种情况：

- 如果某一条冲突的有效行已被当前事务删除，那么这是允许的。（特别是，因为一次 UPDATE 总是在插入新版本前删除旧版本，这就允许对某一行执行不改变键值的 UPDATE。）
- 如果某条冲突的行是由一个尚未提交的事务插入的，那么当前准备插入的事务必须等待，
  看看那个事务是否提交。如果它回滚，就不存在冲突；
  如果它提交并且没有再次删除那条冲突行，就发生了唯一性违背。（实际上，
  我们只是等待另一个事务结束，然后把可见性检查整个重新做一遍。）
- 类似地，如果某条冲突的有效行是由一个尚未提交的事务删除的，
  那么当前准备插入的事务必须等待该事务提交或中止，然后重新执行测试。

我们要求索引访问方法自行应用这些测试，
这意味着它必须深入堆中检查那些根据索引内容显示为具有重复键的行的提交状态。毫无疑问，
这样做既丑陋又不够模块化，但它避免了重复工作：如果我们单独再做一次探测，
那么在寻找新行索引条目插入位置时，查找冲突行的索引搜索实际上就会被重复执行。更何况，
除非把冲突检查作为插入新索引条目动作的一个组成部分，
否则也没有明显的方法可以避免竞争条件。

该方案的主要局限在于，它没有支持延迟唯一性检查的便捷方法。

48.6. 索引代价估算函数

amcostestimate 函数会得到一个已被确定可用于该索引的 WHERE 子句列表。它必须返回访问该索引的代价估计，以及 WHERE 子句的选择率估计（即在索引扫描期间将被检索的父表行所占的

比例）。对于简单情况，代价估算器几乎全部工作都可以通过调用优化器中的标准例程来完成；之所以要提供 amcostestimate 函数，是为了让索引访问方法能够提供与索引类型相关的知识，以便在可能的情况下改进标准估计。

每个 amcostestimate 函数都必须具有下面的签名：

```
void
amcostestimate (PlannerInfo *root,
                IndexOptInfo *index,
                List *indexQuals,
                Cost *indexStartupCost,
                Cost *indexTotalCost,
                Selectivity *indexSelectivity,
                double *indexCorrelation);
```

前三个参数是输入：

root

规划器关于当前正在处理查询的信息。

index

正在被考虑的索引。

indexQuals

索引限定子句的列表（隐含地做 AND 连接）；一个 NIL 列表表示没有可用的限定条件。注意，该列表包含的是表达式树，而不是 ScanKey。

最后四个参数是按引用传递的输出参数：

*indexStartupCost

设为索引启动处理的代价。

*indexTotalCost

设为索引处理的总代价。

*indexSelectivity

设为索引选择率。

*indexCorrelation

设为索引扫描顺序与底层表顺序之间的相关系数。

请注意，代价估算函数必须用 C 编写，而不能用 SQL 或任何可用的过程语言，因为它们必须访问规划器/优化器的内部数据结构。

索引访问代价应使用 `src/backend/optimizer/path/costsize.c` 所用的参数来计算：顺序磁盘块访问的代价为 1.0，非顺序访问的代价为 `random_page_cost`，而处理一个索引行的代价通常取为 `cpu_index_tuple_cost`。此外，对于索引处理（尤其是 indexQuals 本身的求值）期间调用的每个比较操作符，都应计入 `cpu_operator_cost` 的适当倍数。

访问代价应包括与扫描索引本身有关的全部磁盘和 CPU 代价，但不包括取出或处理由索引标识出的父表行的代价。

“启动代价”是整个扫描总代价中必须在开始取第一行之前先付出的那一部分。
对大多数索引来说，这可以视为零；但启动代价较高的索引类型可能希望把它设为非零。

indexSelectivity 应设为在索引扫描期间将从父表中检索出的行的估计比例。对于有损查询，这个值通常会高于实际通过给定限定条件的行比例。

indexCorrelation 应设为索引顺序与表顺序之间的相关性（范围从 -1.0 到 1.0）。该值用于调整从父表取行代价的估计。

代价估算

一个典型的代价估算器会按如下步骤进行：

1. 根据给定的限定条件，估计并返回将被访问的父表行所占的比例。
   在没有索引类型相关知识的情况下，使用标准优化器函数 `clauselist_selectivity()`：

   ```
   *indexSelectivity = clauselist_selectivity(root, indexQuals,
                                              index->rel->relid,
                                              JOIN_INNER, NULL);
   ```

2. 估计扫描期间将被访问的索引行数。对于许多索引类型，它等于 indexSelectivity 乘以索引中的行数，但也可能更多。（注意，索引的大小（页数和行数）可以从 IndexOptInfo 结构体中获得。）

3. 估计扫描期间将读取的索引页面数。它可能仅仅是 indexSelectivity 乘以索引总页面数。

4. 计算索引访问代价。一个通用估计器可能会这样做：

   ```
   /*
    * Our generic assumption is that the index pages will be read
    * sequentially, so they have cost 1.0 each, not random_page_cost.
    * Also, we charge for evaluation of the indexquals at each index
    row.
    * All the costs are assumed to be paid incrementally during the
    scan.
    */
   cost_qual_eval(&index_qual_cost, indexQuals, root);
   *indexStartupCost = index_qual_cost.startup;
   *indexTotalCost = numIndexPages +
       (cpu_index_tuple_cost + index_qual_cost.per_tuple) * numIndex
   Tuples;
   ```

5. 估计索引的相关性。对于单列上的简单有序索引，这个值可以从 pg_statistic 中取得。如果相关性未知，保守估计应为零（即无相关性）。

代价估算器函数的示例可在 `src/backend/utils/adt/selfuncs.c` 中找到。

第 49 章 GiST 索引

49.1. 简介

GiST是 Generalized Search Tree（通用搜索树）的缩写。它是一种平衡的树形访问方法，可作为实现任意索引方案的基本模板。B+-树、R 树以及许多其他索引方案都可以在GiST中实现。

GiST的一个优点是，它使数据类型所属领域的专家而不是数据库专家，能够开发带有适当访问方法的自定义数据类型。

这里的一些内容来自加州大学伯克利分校的 GiST 索引项目网站[1]以及 Marcel Kornacker 的论文 Access Methods for Next-Generation Database Systems[2]。PostgreSQL中的GiST 实现主要由 Teodor Sigaev 和 Oleg Bartunov 维护，他们的网站[3]上还有更多信息。

49.2. 可扩展性

传统上，实现一种新的索引访问方法意味着大量艰难的工作。必须理解数据库的内部机制，例如锁管理器和预写式日志。GiST接口具有很高的抽象层次，只要求访问方法实现者实现被访问数据类型的语义。GiST层本身会处理并发、日志记录以及树结构的搜索。

这种可扩展性不应与其他标准搜索树在可处理数据方面的可扩展性相混淆。例如，PostgreSQL支持可扩展的 B+-树和 R 树。这意味着你可以用PostgreSQL在任意数据类型上构建 B+-树或 R 树。但 B+-树只支持范围谓词（<、=、>），而 R 树只支持 n-D 范围查询（包含、被包含、相等）。

因此，如果你用PostgreSQL的 B+-树为一个图像集合建立索引，你只能发出诸如“imagex 是否等于 imagey”、“imagex 是否小于 imagey”以及“imagex 是否大于 imagey”之类的查询。取决于你如何在这种上下文中定义“等于”、“小于”和“大于”，这可能仍然有用。不过，使用基于GiST的索引，你就可以构造出能够提出特定领域问题的查询方式，例如“找出所有马的图片”或者“找出所有曝光过度的图片”。

要让一个 GiST 访问方法运行起来，只需实现七个用户定义的方法，这些方法定义了树中键的行为。当然，要支持复杂查询，这些方法本身也必须足够巧妙；但对于所有标准查询（B+-树、R 树等），它们都相对直接。简而言之，GiST 把可扩展性与通用性、代码复用以及清晰的接口结合了起来。

49.3. 实现

一个GiST索引操作符类必须提供七个方法：

consistent

给定树页面上的一个谓词`p`和一个用户查询`q`，如果可以确定对于给定的数据项`p`和`q`不可能同时为真，该方法将返回假。

union

此方法合并树中的信息。给定一组项，该函数生成一个对所有这些项都为真的新谓词。

[1] http://gist.cs.berkeley.edu/
[2] http://www.sai.msu.su/~megera/postgres/gist/papers/concurrency/access-methods-for-next-generation.pdf.gz
[3] http://www.sai.msu.su/~megera/postgres/gist/

compress

把数据项转换为适合在索引页面中物理存储的格式。

decompress

`compress`方法的逆操作。把数据项的索引表示形式转换为数据库可以操纵的格式。

penalty

返回一个值，指示把新项插入树的特定分支的“代价”。项将被沿着树中`penalty`最小的路径插入。

picksplit

当必须进行页面分裂时，该函数决定页面上哪些项留在旧页面，哪些项移动到新页面。

same

如果两个项完全相同则返回真，否则返回假。

49.4. 示例

PostgreSQL源代码发行包包含了若干使用GiST实现的索引方法示例。
核心系统目前为某些内置几何数据类型提供了与 R 树等价的功能（见`src/backend/access/gist/gistproc.c`）。下列`contrib`模块中也包含GiST操作符类：

btree_gist

为多种数据类型提供与 B-Tree 等价的功能

cube

多维立方体的索引

intarray

一维 int4 值数组的 RD 树

ltree

树状结构的索引

pg_trgm

基于三字符组匹配的文本相似度

seg

“float 范围” 的索引

tsearch2

全文索引

49.5. 崩溃恢复

通常，重放 WAL 日志足以在数据库崩溃之后恢复 GiST 索引的完整性。但也存在一些边角情况，索引状态在其中不会被完全重建。索引在功能上仍然正确，但可能出现一些性能下降。

发生这种情况时，可以对该索引所在的表执行 VACUUM 来修复索引，或者用 REINDEX 重建索引。在某些情况下，普通 VACUUM 并不够，需要 VACUUM FULL 或 REINDEX。如果崩溃恢复期间出现这条日志消息，就表明需要执行上述操作之一：

```
LOG:  index NNN/NNN/NNN needs VACUUM or REINDEX to finish crash
 recovery
```

或者，在例行的索引插入过程中出现这条日志消息也表明这一点：

```
LOG:  index "FOO" needs VACUUM or REINDEX to finish crash recovery
```

如果普通 VACUUM 发现自己无法完全完成恢复，它会返回一个提示：

```
NOTICE:  index "FOO" needs VACUUM FULL or REINDEX to finish crash
 recovery
```

第 50 章 数据库物理存储

本章概述 PostgreSQL 数据库所使用的物理存储格式。

50.1. 数据库文件布局

本节从文件和目录的层次描述存储格式。

数据库集簇所需的全部数据都存放在集簇的数据目录中，该目录通常称为 PGDATA（名字来自可用于定义它的环境变量）。PGDATA 的一个常见位置是 /var/lib/pgsql/data。同一台机器上可以存在多个由不同服务器实例管理的集簇。

PGDATA 目录包含若干子目录和控制文件，如表 50.1所示。除这些必需项之外，集簇配置文件 postgresql.conf、pg_hba.conf 和 pg_ident.conf 传统上也存放在 PGDATA 中（不过在 PostgreSQL 8.0 及以后版本中，可以把它们放在其他地方）。

表 50.1. PGDATA 的内容

Item	描述
PG_VERSION	包含 PostgreSQL 主版本号的文件
base	包含各数据库子目录的子目录
global	包含集簇范围的表（例如 pg_database）的子目录
pg_clog	包含事务提交状态数据的子目录
pg_multixact	包含多事务状态数据的子目录（用于共享行锁）
pg_subtrans	包含子事务状态数据的子目录
pg_tblspc	包含指向表空间的符号链接的子目录
pg_twophase	包含预备事务状态文件的子目录
pg_xlog	包含 WAL（预写式日志）文件的子目录
postmaster.opts	记录 postmaster 上次启动时所用命令行选项的文件
postmaster.pid	记录当前 postmaster PID 和共享内存段 ID 的锁文件（postmaster 关闭后不存在）

对于集簇中的每个数据库，PGDATA/base 中都有一个子目录，其名称是该数据库在 pg_database 中的 OID。这个子目录是该数据库文件的默认位置；特别是，它的系统目录就存放在这里。

每个表和索引都存储在单独的文件中，文件以表或索引的 filenode 编号命名，该编号可在 pg_class.relfilenode 中找到。

> **小心**
>
> 请注意，虽然表的 filenode 往往与其 OID 相同，但这并非必然如此；有些操作，例如 TRUNCATE、REINDEX、CLUSTER 以及某些形式的 ALTER TABLE，会在保留 OID 的同时改变 filenode。不要假定 filenode 和表 OID 一定相同。

当表或索引超过 1 GB 时，会被划分成若干个 1 GB 大小的段。第一段的文件名与 filenode 相同，后续各段则命名为 filenode.1、filenode.2，依此类推。这样的安排避免了在文件大小受限的平台上出现问题。表和索引的内容将在第 50.3 节中进一步讨论。

如果一个表包含可能具有很大条目的列，它就会有关联的 TOAST 表，用于对那些大到无法直接保存在表行中的字段值进行线外存储。pg_class.reltoastrelid 会把该表关联到其 TOAST 表（如果存在的话）。更多信息见第 50.2 节。

表空间使情况变得更复杂。每个用户定义的表空间都在 PGDATA/pg_tblspc 目录中有一个符号链接，指向其物理表空间目录（即该表空间的 CREATE TABLESPACE 命令中指定的位置）。该符号链接以表空间的 OID 命名。在物理表空间目录中，对于每个在该表空间中拥有对象的数据库，都有一个以该数据库 OID 命名的子目录。表和索引就使用 filenode 命名方案存储在该目录中。pg_default 表空间不通过 pg_tblspc 访问，而是对应于 PGDATA/base。同样，pg_global 表空间也不通过 pg_tblspc 访问，而是对应于 PGDATA/global。

50.2. TOAST

本节概述TOAST（超长属性存储技术）。

由于 PostgreSQL 使用固定页面大小（通常为 8Kb），并且不允许元组跨越多个页面，因此无法直接存储非常大的字段值。在 PostgreSQL 7.1 之前，一个表行可以容纳的数据总量有一个略小于一页的硬性限制。在 7.1 及以后的版本中，这一限制通过允许压缩大字段值并且/或者把它们拆分成多个物理行来克服。这个过程对用户是透明的，而且对大部分后端代码的影响很小。
这项技术被亲切地称为 TOAST（或者“自切片面包问世以来最棒的东西”）。

只有某些数据类型支持 TOAST，因为没有必要让那些不可能产生大字段值的数据类型承担这部分额外开销。要支持 TOAST，数据类型必须具有变长（varlena）表示形式，其中任何已存储值的第一个 32 位字都包含该值以字节计的总长度（包括其自身）。TOAST 不约束表示形式的其余部分。所有支持可 TOAST 数据类型的 C 级函数都必须小心处理已 TOAST 化的输入值。（通常是在对输入值做任何处理之前调用 PG_DETOAST_DATUM，但在某些情况下也可以采用更高效的方法。）

TOAST 占用 varlena 长度字的高阶两位，从而把任何可 TOAST 数据类型值的逻辑大小限制为 1Gb（2^{30} - 1 字节）。当这两个位都为零时，该值就是这种数据类型的普通、未经 TOAST 处理的值。其中一位如果被置位，就表示该值已经被压缩，使用前必须先解压。另一位如果被置位，则表示该值已采用线外存储。在这种情况下，值的其余部分实际上只是一个指针，正确的数据必须到别处查找。当两个位都被置位时，线外数据也已被压缩。在每种情况下，varlena 字低位中的长度指示的都是 datum 的实际大小，而不是经过解压或取回线外数据后所提取出的逻辑值的大小。

如果一个表的任意列是可 TOAST 的，该表就会有一个相关联的 TOAST 表，其 OID 存储在该表的 pg_class.reltoastrelid 项中。线外 TOAST 化值保存在 TOAST 表中，下文会更详细地描述。

所用的压缩技术是 LZ 压缩技术族中一种相当简单且非常快速的方法。详情见 src/backend/utils/adt/pg_lzcompress.c。

线外值会被划分成（如果启用了压缩，则在压缩之后划分）最多 TOAST_MAX_CHUNK_SIZE 字节的块（该值略小于 BLCKSZ/4，默认大约是 2000 字节）。每个块都作为独立的一行，存储在所属表的 TOAST 表中。每个 TOAST 表都有 chunk_id 列（标识某个特定 TOAST 化值的 OID）、chunk_seq 列（该块在其值内的序号）以及 chunk_data 列（该块的实际数据）。在 chunk_id 和 chunk_seq 上建立的唯一索引可提供快速检索。因此，一个表示线外 TOAST 化值的指针 datum，需要存储要查找的 TOAST 表的 OID，以及该特定值的 OID（即其 chunk_id）。为方便起见，指针 datum 还会存储逻辑 datum 大小（原始未压缩数据长度）和物理存储大小（如果应用了压缩，这两者会不同）。再加上 varlena 头部字，TOAST 指针 datum 的总大小因此恒为 20 字节，与所表示值的实际大小无关。

TOAST 代码只有在要存入表中的行值宽于 BLCKSZ/4 字节（通常为 2Kb）时才会被触发。TOAST 代码会压缩并且/或者把字段值移到线外，直到该行值短于 BLCKSZ/4 字节，或者已经无法再获

得更多收益。在 UPDATE 操作中，未更改字段的值通常会原样保留；因此，如果更新一行时其线外值都没有变化，便不会产生 TOAST 成本。

TOAST 代码为存储可 TOAST 的列识别出四种不同的策略：

- `PLAIN` 既不允许压缩，也不允许线外存储。这是不可 TOAST 数据类型列唯一可能的策略。
- `EXTENDED` 同时允许压缩和线外存储。这是大多数可 TOAST 数据类型的默认策略。系统会先尝试压缩，如果行仍然过大，再使用线外存储。
- `EXTERNAL` 允许线外存储，但不允许压缩。使用 `EXTERNAL` 会让宽 `text` 和 `bytea` 列上的子串操作更快（代价是占用更多存储空间），因为在值未压缩时，这些操作经过优化，只需提取线外值中所需的部分。
- `MAIN` 允许压缩，但不允许线外存储。（实际上，对于这类列，仍然可能进行线外存储，但只有在别无他法、必须这样做才能让行足够小以放入页面时，才会把它作为最后手段。）

每种可 TOAST 的数据类型都会为该数据类型的列指定默认策略，但可以通过以下命令更改某个表列的策略：`ALTER TABLE SET STORAGE`。

与允许行值跨页之类更直接的方法相比，这种方案有不少优点。
假定查询通常是通过与相对较小的键值比较来限定的，
执行器的大部分工作都会只使用主行项完成。TOAST 化属性的大值只有在结果集发送给客户端时才会被取出（前提是它们确实被选中了）。因此，主表会小得多，
其更多的行能够装入共享缓冲区缓存，这是没有线外存储时做不到的。排序集也会缩小，
因此排序更常能够完全在内存中完成。一个小测试表明，一个包含典型 HTML 页面及其 URL 的表，其总存储量（包括 TOAST 表）大约只有原始数据大小的一半，而主表只包含全部数据的大约 10%（URL 以及一些较小的 HTML 页面）。与未进行 TOAST 处理的对照表相比，运行时没有差异；在那个对照表中，所有 HTML 页面都被裁剪到 7 kB 以内以便放得下。

50.3. 数据库页面布局

本节概述 PostgreSQL 表和索引内部所使用的页面格式。[1]序列和 TOAST 表的格式与普通表相同。

在下面的说明中，假定一个字节包含 8 个位。另外，术语项指的是存储在页面上的单个数据值。在表中，项是一行；在索引中，项是一条索引条目。

每个表和索引都存储为固定大小的页面数组（通常为 8 kB，不过在编译服务器时可以选择不同的页面大小）。在表中，所有页面在逻辑上都是等价的，因此某个特定项（行）
可以存放在任意页面中。在索引中，第一页通常保留为元页，用于保存控制信息；
并且根据索引访问方法的不同，索引中还可能存在不同类型的页面。

表 50.2展示了页面的总体布局。每个页面包含五个部分。

表 50.2. 总体页面布局

Item	描述
PageHeaderData	长度为 20 字节。包含页面的一般信息，包括空闲空间指针。

[1] 实际上，索引访问方法不要求必须使用这种页面格式。现有的所有索引方法都使用这种基本格式，但索引元页中保存的数据通常并不遵循项布局规则。

Item	描述
ItemIdData	指向实际项的（偏移量、长度）对数组。每项 4 字节。
空闲空间	尚未分配的空间。新的项指针从该区域起始处分配，新的项从末尾处分配。
项	项本身。
特殊空间	索引访问方法特有的数据。不同方法存储不同数据。普通表中为空。

每个页面的前 24 个字节由页头（PageHeaderData）组成。它的格式详见表 50.3。前两个字段跟踪与该页面有关的最新 WAL 记录。接下来是一个 2 字节字段，包含标志位。随后是三个 2 字节整数字段（`pd_lower`、`pd_upper` 和 `pd_special`）。它们分别保存从页面起始位置到未分配空间起始位置、未分配空间结束位置以及特殊空间起始位置的字节偏移量。页头接下来的 2 个字节 `pd_pagesize_version` 同时存储页面大小和版本指示器。从 PostgreSQL 8.1 起版本号为 3；PostgreSQL 8.0 使用版本号 2；PostgreSQL 7.3 和 7.4 使用版本号 1；更早的版本使用版本号 0。（在这些版本中，基本页面布局和页头格式并未改变，但堆行头部的布局发生过变化。）
页面大小字段基本上只是用于交叉检查；一个安装中并不支持同时存在多种页面大小。
最后一个字段是一个提示，用于显示对页面进行剪枝是否可能有利：它跟踪页面上最老的、尚未剪枝的 XMAX。

表 50.3. PageHeaderData 布局

Field	Type	Length	描述
pd_lsn	XLogRecPtr	8 字节	LSN：上次修改此页面的 xlog 记录最后一个字节之后的下一个字节
pd_tli	TimeLineID	4 字节	TLI of last change
pd_lower	LocationIndex	2 字节	到空闲空间起始位置的偏移量
pd_upper	LocationIndex	2 字节	到空闲空间结束位置的偏移量
pd_special	LocationIndex	2 字节	到特殊空间起始位置的偏移量
pd_pagesize_version	uint16	2 字节	页面大小和布局版本号信息

所有细节都可以在 `src/include/storage/bufpage.h` 中找到。

页头之后是项标识符（`ItemIdData`），每个需要四个字节。
一个项标识符包含项起始位置的字节偏移量、其字节长度，以及若干影响解释方式的属性位。
新的项标识符会根据需要从未分配空间的起始处分配。当前已有多少个项标识符，可以通过查看 `pd_lower` 得知；分配新标识符时它会增加。因为一个项标识符在被释放之前永远不会移动，所以即使项本身为了压缩空闲空间而在页面内移动，其索引仍然可以被长期用来引用该项。
事实上，每个指向项的指针（`ItemPointer`，也称 `CTID`）在 PostgreSQL 中都由页号和项标识符索引构成。

项本身存储在从未分配空间末尾开始、向后分配的区域中。
其确切结构取决于表要包含什么内容。表和序列都使用一种名为 `HeapTupleHeaderData` 的结构体，如下所述。

最后一部分是“特殊部分”，其中可以存放访问方法希望保存的任何内容。例如，B-树索引会在这里保存指向页面左、右兄弟的链接，以及其他一些与索引结构相关的数据。
普通表完全不使用特殊部分（通过将 `pd_special` 设为页面大小来表示）。

所有表行的结构都相同。它们都有一个固定大小的头部（在大多数机器上占 27 字节），后面跟着可选的空值位图、可选的对象 ID 字段以及用户数据。头部的详细格式见表 50.4。实际用户数据（行的各列）从 `t_hoff` 指示的偏移位置开始，它必须始终是该平台 MAXALIGN 对齐单位的整数倍。只有当 HEAP_HASNULL 位在 `t_infomask` 中被置位时，空值位图才存在。若存在，它紧随固定头部之后，并占用足够多的字节，以便为每个数据列提供一位（也就是说，总共为 `t_natts` 位）。在这个位图中，1 表示非空，0 表示空值。当位图不存在时，假定所有列都非空。只有当 HEAP_HASOID 位在 `t_infomask` 中被置位时，对象 ID 才存在。若存在，它位于 `t_hoff` 边界之前。为了让 `t_hoff` 成为 MAXALIGN 的整数倍所需的任何填充，都会出现在空值位图与对象 ID 之间。（这又反过来保证了对象 ID 的对齐是合适的。）

表 50.4. HeapTupleHeaderData 布局

Field	Type	Length	描述
t_xmin	TransactionId	4 字节	insert XID stamp
t_cmin	CommandId	4 字节	insert CID stamp
t_xmax	TransactionId	4 字节	delete XID stamp
t_cmax	CommandId	4 字节	delete CID stamp (overlays with t_xvac)
t_xvac	TransactionId	4 字节	VACUUM 操作移动某个行版本时的 XID
t_ctid	ItemPointerData	6 字节	本行版本或更新行版本的当前 TID
t_natts	int16	2 字节	number of attributes
t_infomask	uint16	2 字节	various flag bits
t_hoff	uint8	1 byte	offset to user data

所有细节都可以在 `src/include/access/htup.h` 中找到。

要解释实际数据，必须借助从其他表中获得的信息，其中大部分来自 `pg_attribute`。识别字段位置所需的关键值是 `attlen` 和 `attalign`。除非所有字段都是定宽且没有空值，否则没有办法直接取得某个特定属性。所有这些技巧都封装在 heap_getattr、fastgetattr 和 heap_getsysattr 这些函数中。

读取数据时，需要依次检查每个属性。首先根据空值位图判断该字段是否为 NULL。如果是，就继续下一个。然后确认对齐是否正确。如果字段是定宽字段，那么它的所有字节都是直接摆放的；如果它是变长字段（attlen = -1），情况就会稍复杂一些。所有变长数据类型都共享一个通用头部结构 `varattrib`，其中包含已存储值的总长度以及一些标志位。根据这些标志，数据可能是线内存储的，也可能位于 TOAST 表中；它也可能是经过压缩的（见第 50.2 节）。

第 51 章 BKI 后端接口

后端接口（BKI）文件是使用一种特殊语言编写的脚本，PostgreSQL 后端在“引导”模式下运行时能够理解这种语言。引导模式允许从零开始创建并填充系统目录，而普通 SQL 命令要求目录已经存在。因此，BKI 文件可以用于最初创建数据库系统。（它们可能也没有其他用途。）

initdb 在创建新的数据库集簇时，使用 BKI 文件完成部分工作。initdb 使用的输入文件作为构建和安装 PostgreSQL 的一部分，由名为 `genbki.pl` 的程序创建。该程序读取源码树 `src/include/catalog/` 目录中采用特殊格式的 C 头文件。创建的 BKI 文件名为 `postgres.bki`，通常安装在安装树的 `share` 子目录中。

相关信息可以在 initdb 的文档中找到。

51.1. BKI 文件格式

本节描述 PostgreSQL 后端如何解释 BKI 文件。如果手边有一份 `postgres.bki` 文件作为示例，这段说明会更容易理解。

BKI 输入由一系列命令组成。根据命令语法的不同，每条命令由若干词元组成。词元通常由空白分隔，但如果不会产生歧义，也可以不分隔。没有专门的命令分隔符；
语法上不可能属于前一条命令的下一个词元，就会开始一条新命令。（通常为了清晰，
你会把新命令写在新的一行。）词元可以是某些关键字、特殊字符（圆括号、逗号等）、数字，
或者双引号字符串。所有内容都区分大小写。

以 `#` 开头的行会被忽略。

51.2. BKI 命令

`create` [`bootstrap`] [`shared_relation`] [`without_oids`] *`tablename`* *`tableoid`* (*`name1`* = *`type1`* [, *`name2`* = *`type2`*, ...])

创建一个名为 *`tablename`*、OID 为 *`tableoid`* 的表，其列定义在括号中给出。

`bootstrap.c` 直接支持以下列类型：`bool`、`bytea`、`char`（1 字节）、`name`、`int2`、`int4`、`regproc`、`regclass`、`regtype`、`text`、`oid`、`tid`、`xid`、`cid`、`int2vector`、`oidvector`、`_int4`（数组）、`_text`（数组）、`_oid`（数组）、`_char`（数组）以及 `_aclitem`（数组）。虽然也可以创建包含其他类型列的表，但这必须等到 `pg_type` 创建完成并填入合适条目之后才能做到。（实际上，这意味着引导目录中只能使用这些列类型，而非引导目录则可以包含任何内置类型。）

指定 `bootstrap` 时，表只会在磁盘上创建；不会在 `pg_class`、`pg_attribute` 等目录中为它写入任何信息。因此，在通过 `insert` 命令以较为原始的方式补入这些条目之前，普通 SQL 操作无法访问该表。这个选项用于创建 `pg_class` 等目录自身。

如果指定了 `shared_relation`，该表就会作为共享表创建。除非指定 `without_oids`，否则该表将具有 OID。

`open` *`tablename`*

打开名为 *`tablename`* 的表，以便插入数据。当前任何已打开的表都会被关闭。

`close` [*`tablename`*]

关闭当前打开的表。可以给出表名进行交叉校验，但并非必须。

`insert` [`OID =` *`oid_value`*] (*`value1 value2`* ...)

向当前打开的表插入一行新数据，使用 *`value1`*、*`value2`* 等作为各列的值，使用 *`oid_value`* 作为其 OID。如果 *`oid_value`* 为零（0）或省略该子句，并且该表具有 OID，则分配下一个可用 OID。

NULL 值可以用特殊关键字 `_null_` 表示。包含空格的值必须用双引号括起来。

`declare` [`unique`] `index` *`indexname indexoid`* `on` *`tablename`* `using` *`amname`* (*`opclass1 name1`* [, ...])

在名为 *`tablename`* 的表上，使用 *`amname`* 访问方法，创建一个名为 *`indexname`*、OID 为 *`indexoid`* 的索引。要建立索引的字段分别称为 *`name1`*、*`name2`* 等；使用的操作符类分别为 *`opclass1`*、*`opclass2`* 等。该命令会创建索引文件并写入相应的目录条目，但不会初始化索引内容。

`build indices`

填充之前已声明的索引。

51.3. 引导 BKI 文件的结构

在 `open` 命令可以使用之前，它所依赖的那些表必须已经存在，并且其中已经有了将要被打开之表对应的条目。（这些最基本的表包括 `pg_class`、`pg_attribute`、`pg_proc` 和 `pg_type`。）为了让这些表本身也能够被填充，带 `bootstrap` 选项的 `create` 命令会隐式打开刚创建的表，以便插入数据。

因此，`postgres.bki` 文件的结构必须如下：

1. 对某个关键表执行 `create bootstrap`
2. 用 `insert` 插入至少足以描述这些关键表的数据
3. `close`
4. 对其他关键表重复上述步骤。
5. 对某个非关键表执行 `create`（不带 `bootstrap`）
6. `open`
7. 用 `insert` 插入所需数据
8. `close`
9. 对其他非关键表重复上述步骤。
10. 定义索引和 TOAST 表。
11. `build indices`

无疑还存在其他未文档化的顺序依赖。

51.4. 示例

以下命令序列会创建表`test_table`，其 OID 为 420，包含两列`cola`和`colb`，类型分别为`int4`和`text`，并向表中插入两行：

```
create test_table 420 (cola = int4, colb = text)
open test_table
insert OID=421 ( 1 "value1" )
insert OID=422 ( 2 _null_ )
close test_table
```

第 52 章 规划器如何使用统计信息

本章建立在第 13.1 节和第 13.2 节所介绍的内容之上，说明规划器如何利用系统统计信息来估计查询各阶段可能返回的行数。这是规划/优化过程的重要组成部分，
为代价计算提供了大量原始材料。

本章的目的不是为代码编写文档 – 那更适合在代码本身中完成，而是概述其工作方式。这或许能让随后想要阅读代码的人更容易入门。因此，我们选择的方式是分析一系列复杂度递增的示例。

下面展示的输出和算法取自版本 8.0。更早（或更晚）的版本行为可能有所不同。

52.1. 行估计示例

使用取自回归测试数据库的示例，让我们从一个很简单的查询开始：

```
EXPLAIN SELECT * FROM tenk1;

                         QUERY PLAN
-------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..445.00 rows=10000 width=244)
```

关于规划器如何确定 tenk1 的基数，第 13.1 节中已经介绍过；这里为了完整起见再重复一次。行数是从 pg_class 中查得的：

```
SELECT reltuples, relpages FROM pg_class WHERE relname = 'tenk1';

 reltuples | relpages
----------+-----------
     10000 |       345
```

规划器会检查relpages估计（这是一项开销很小的操作），如果它不正确，就可能对reltuples进行缩放以获得行数估计。本例中没有这样做，因此：

```
rows = 10000
```

让我们接着看一个在WHERE子句中包含范围条件的示例：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 1000;

                         QUERY PLAN
--------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..470.00 rows=1031 width=244)
   Filter: (unique1 < 1000)
```

规划器检查WHERE子句中的条件：

```
unique1 < 1000
```

并在pg_operator中查找操作符<的限制函数。它保存在oprrest列中，本例的结果为scalarltsel。scalarltsel函数从pg_statistics中取得unique1的直方图 – 我们可以通过更简单的pg_stats视图来查看它：

```
SELECT histogram_bounds FROM pg_stats
WHERE tablename='tenk1' AND attname='unique1';
```

```
                    histogram_bounds
-------------------------------------------------------
 {1,970,1943,2958,3971,5069,6028,7007,7919,8982,9995}
```

接下来，计算“< 1000”在直方图中所占的比例，这就是选择率。直方图把范围划分为等频率的桶，因此只需找到目标值所在的桶，计入该桶的部分和前面所有桶的全部。值 1000 显然在第二个桶（970-1943）中。假设每个桶内部的值呈线性分布，可以按如下方式计算选择率：

```
selectivity = (1 + (1000 - bucket[2].min)/(bucket[2].max - bucket[2].
min))/num_buckets
            = (1 + (1000 - 970)/(1943 - 970))/10
            = 0.1031
```

即一个完整的桶加上第二个桶中的线性比例，再除以桶数。用选择率乘以`tenk1`的基数，即可计算估计行数：

```
rows = rel_cardinality * selectivity
     = 10000 * 0.1031
     = 1031
```

接下来考虑一个在`WHERE`子句中包含等值条件的示例：

```
EXPLAIN SELECT * FROM tenk1 WHERE stringu1 = 'ATAAAA';

                        QUERY PLAN
-----------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..470.00 rows=31 width=244)
   Filter: (stringu1 = 'ATAAAA'::name)
```

规划器同样检查`WHERE`子句中的条件：

```
stringu1 = 'ATAAAA'
```

并查找`=`的限制函数，即`eqsel`。这个例子有些不同，因为要使用最常见值 — MCV 来确定选择率。查看这些值，同时带上一些后面会有用的额外列：

```
SELECT null_frac, n_distinct, most_common_vals, most_common_freqs FROM
 pg_stats
WHERE tablename='tenk1' AND attname='stringu1';

null_frac         | 0
n_distinct        | 672
most_common_vals  | {FDAAAA,NHAAAA,ATAAAA,BGAAAA,EBAAAA,MOAAAA,NDAAAA,
OWAAAA,BHAAAA,BJAAAA}
most_common_freqs | {0.00333333,0.00333333,0.003,0.003,0.003,0.003,0.
003,0.003,0.00266667,0.00266667}
```

选择率就是第三个MCV – 'ATAAAA' 对应的最常见值频率（MCF）：

```
selectivity = mcf[3]
            = 0.003
```

估计行数与之前一样，是这个值与`tenk1`基数的乘积：

```
rows = 10000 * 0.003
```

```
     = 30
```

由于一些估计后的检查，EXPLAIN显示的数字比它多 1。

现在考虑同一个查询，但使用一个不在MCV列表中的常量：

```
EXPLAIN SELECT * FROM tenk1 WHERE stringu1 = 'xxx';

                        QUERY PLAN
-----------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..470.00 rows=15 width=244)
   Filter: (stringu1 = 'xxx'::name)
```

这是一个完全不同的问题：当值不在MCV 列表中时，如何估计选择率。方法是利用该值不在列表中的事实，再结合所有 MCV的频率信息：

```
selectivity = (1 - sum(mvf))/(num_distinct - num_mcv)
            = (1 - (0.00333333 + 0.00333333 + 0.003 + 0.003 + 0.003
            + 0.003 + 0.003 + 0.003 + 0.00266667 + 0.00266667))/(672 -
 10)
            = 0.001465
```

也就是说，把所有MCV的频率加起来，再用一减去这个总和 – 因为该值不是其中之一 – 然后除以其余的非重复值数量。注意，这里没有空值，所以不必考虑它们。估计行数照常计算：

```
rows = 10000 * 0.001465
     = 15
```

现在考虑在WHERE子句中有多个条件的情形：

```
EXPLAIN SELECT * FROM tenk1 WHERE unique1 < 1000 AND stringu1 = 'xxx';

                         QUERY PLAN
------------------------------------------------------------
 Seq Scan on tenk1  (cost=0.00..495.00 rows=2 width=244)
   Filter: ((unique1 < 1000) AND (stringu1 = 'xxx'::name))
```

规划器假定这两个条件相互独立，因此可以将各个子句的选择率相乘：

```
selectivity = selectivity(unique1 < 1000) * selectivity(stringu1 =
 'xxx')
            = 0.1031 * 0.001465
            = 0.00015104
```

估计行数照常计算：

```
rows = 10000 * 0.00015104
     = 2
```

最后，我们来看一个同时包含JOIN和 WHERE子句的查询：

```
EXPLAIN SELECT *  FROM tenk1 t1, tenk2 t2
WHERE t1.unique1 < 50 AND t1.unique2 = t2.unique2;

                                       QUERY PLAN
-----------------------------------------------------------------------
-------------------
```

```
 Nested Loop  (cost=0.00..346.90 rows=51 width=488)
   ->  Index Scan using tenk1_unique1 on tenk1 t1  (cost=0.00..192.57
 rows=51 width=244)
         Index Cond: (unique1 < 50)
   ->  Index Scan using tenk2_unique2 on tenk2 t2  (cost=0.00..3.01
 rows=1 width=244)
         Index Cond: ("outer".unique2 = t2.unique2)
```

对tenk1的限制条件“unique1 < 50”在嵌套循环连接之前求值。它的处理方式与前面的范围示例类似。<的限制函数仍是scalarlteqsel，但这次值 50 位于 unique1直方图的第一个桶中：

```
selectivity = (0 + (50 - bucket[1].min)/(bucket[1].max - bucket[1].
min))/num_buckets
            = (0 + (50 - 1)/(970 - 1))/10
            = 0.005057

rows        = 10000 * 0.005057
            = 51
```

连接的限制条件是：

```
t2.unique2 = t1.unique2
```

这是因为连接方法是嵌套循环，而tenk1位于外层循环。操作符仍是我们熟悉的=，
但限制函数是从 pg_operator的oprjoin列获得的，它是eqjoinsel。此外，我们要同时使用 tenk2和tenk1的统计信息：

```
SELECT tablename, null_frac,n_distinct, most_common_vals FROM pg_stats
WHERE tablename IN ('tenk1', 'tenk2') AND attname='unique2';

tablename  | null_frac | n_distinct | most_common_vals
-----------+-----------+------------+------------------
 tenk1     |         0 |         -1 |
 tenk2     |         0 |         -1 |
```

这里unique2没有MCV信息，因为所有值看起来都是唯一的，
所以我们可以使用一种只依赖两个关系的不同值数量以及它们空值比例的算法：

```
selectivity = (1 - null_frac1) * (1 - null_frac2) * min(1/num_
distinct1, 1/num_distinct2)
            = (1 - 0) * (1 - 0) * min(1/10000, 1/1000)
            = 0.0001
```

也就是说，对每个关系用一减去其空值比例，再除以两个不同值数量中较大的那个。
连接可能产生的行数按嵌套循环中两个节点的笛卡尔积基数乘以选择率计算：

```
rows = (outer_cardinality * inner_cardinality) * selectivity
     = (51 * 10000) * 0.0001
     = 51
```

如果有兴趣了解更多细节，关系中行数的估计位于 src/backend/optimizer/util/plancat.c。子句选择率的计算逻辑位于src/backend/optimizer/path/clausesel.c。操作符与连接限制函数的实际实现可在 src/backend/utils/adt/selfuncs.c中找到。

部分 VIII. 附录

目录

A. PostgreSQL错误代码 1086
B. 日期/时间支持 1093
B.1. 日期/时间输入解释 1093
B.2. 日期/时间关键字 1094
B.3. 历法的历史 1108
C. SQL 关键字 1110
D. SQL 符合性 1129
D.1. 支持的特性 1130
D.2. 不支持的特性 1140
E. 发布说明 1147
E.1. 版本 8.1.23 1147
E.1.1. 迁移到版本 8.1.23 1147
E.1.2. 变更 1147
E.2. 版本 8.1.22 1148
E.2.1. 迁移到版本 8.1.22 1149
E.2.2. 变更 1149
E.3. 版本 8.1.21 1150
E.3.1. 迁移到版本 8.1.21 1150
E.3.2. 变更 1150
E.4. 版本 8.1.20 1151
E.4.1. 迁移到版本 8.1.20 1151
E.4.2. 变更 1151
E.5. 版本 8.1.19 1152
E.5.1. 迁移到版本 8.1.19 1152
E.5.2. 变更 1152
E.6. 版本 8.1.18 1153
E.6.1. 迁移到版本 8.1.18 1153
E.6.2. 变更 1154
E.7. 版本 8.1.17 1154
E.7.1. 迁移到版本 8.1.17 1155
E.7.2. 变更 1155
E.8. 版本 8.1.16 1155
E.8.1. 迁移到版本 8.1.16 1155
E.8.2. 变更 1156
E.9. 版本 8.1.15 1156
E.9.1. 迁移到版本 8.1.15 1156
E.9.2. 变更 1156
E.10. 版本 8.1.14 1157
E.10.1. 迁移到版本 8.1.14 1157
E.10.2. 变更 1157
E.11. 版本 8.1.13 1158
E.11.1. 迁移到版本 8.1.13 1159
E.11.2. 变更 1159
E.12. 版本 8.1.12 1159
E.12.1. 迁移到版本 8.1.12 1159
E.12.2. 变更 1159
E.13. 版本 8.1.11 1161
E.13.1. 迁移到版本 8.1.11 1161
E.13.2. 变更 1161
E.14. 版本 8.1.10 1162
E.14.1. 迁移到版本 8.1.10 1163

E.14.2. 变更 1163
E.15. 版本 8.1.9 1163
E.15.1. 迁移到版本 8.1.9 1163
E.15.2. 变更 1163
E.16. 版本 8.1.8 1164
E.16.1. 迁移到版本 8.1.8 1164
E.16.2. 变更 1164
E.17. 版本 8.1.7 1164
E.17.1. 迁移到版本 8.1.7 1164
E.17.2. 变更 1165
E.18. 版本 8.1.6 1165
E.18.1. 迁移到版本 8.1.6 1165
E.18.2. 变更 1165
E.19. 版本 8.1.5 1166
E.19.1. 迁移到版本 8.1.5 1166
E.19.2. 变更 1166
E.20. 版本 8.1.4 1167
E.20.1. 迁移到版本 8.1.4 1167
E.20.2. 变更 1168
E.21. 版本 8.1.3 1169
E.21.1. 迁移到版本 8.1.3 1169
E.21.2. 变更 1169
E.22. 版本 8.1.2 1170
E.22.1. 迁移到版本 8.1.2 1171
E.22.2. 变更 1171
E.23. 版本 8.1.1 1172
E.23.1. 迁移到版本 8.1.1 1172
E.23.2. 变更 1172
E.24. 版本 8.1 1173
E.24.1. 概述 1173
E.24.2. 迁移到版本 8.1 1174
E.24.3. 附加变更 1176
E.25. 版本 8.0.26 1186
E.25.1. 迁移到版本 8.0.26 1186
E.25.2. 变更 1186
E.26. 版本 8.0.25 1188
E.26.1. 迁移到版本 8.0.25 1188
E.26.2. 变更 1188
E.27. 版本 8.0.24 1189
E.27.1. 迁移到版本 8.0.24 1189
E.27.2. 变更 1189
E.28. 版本 8.0.23 1190
E.28.1. 迁移到版本 8.0.23 1190
E.28.2. 变更 1190
E.29. 版本 8.0.22 1191
E.29.1. 迁移到版本 8.0.22 1191
E.29.2. 变更 1191
E.30. 版本 8.0.21 1192
E.30.1. 迁移到版本 8.0.21 1192
E.30.2. 变更 1192
E.31. 版本 8.0.20 1193
E.31.1. 迁移到版本 8.0.20 1193
E.31.2. 变更 1193
E.32. 版本 8.0.19 1193

E.32.1. 迁移到版本 8.0.19 1194
E.32.2. 变更 1194
E.33. 版本 8.0.18 1194
E.33.1. 迁移到版本 8.0.18 1194
E.33.2. 变更 1194
E.34. 版本 8.0.17 1195
E.34.1. 迁移到版本 8.0.17 1195
E.34.2. 变更 1195
E.35. 版本 8.0.16 1196
E.35.1. 迁移到版本 8.0.16 1196
E.35.2. 变更 1196
E.36. 版本 8.0.15 1197
E.36.1. 迁移到版本 8.0.15 1197
E.36.2. 变更 1198
E.37. 版本 8.0.14 1199
E.37.1. 迁移到版本 8.0.14 1199
E.37.2. 变更 1199
E.38. 版本 8.0.13 1199
E.38.1. 迁移到版本 8.0.13 1200
E.38.2. 变更 1200
E.39. 版本 8.0.12 1200
E.39.1. 迁移到版本 8.0.12 1200
E.39.2. 变更 1200
E.40. 版本 8.0.11 1200
E.40.1. 迁移到版本 8.0.11 1201
E.40.2. 变更 1201
E.41. 版本 8.0.10 1201
E.41.1. 迁移到版本 8.0.10 1201
E.41.2. 变更 1201
E.42. 版本 8.0.9 1202
E.42.1. 迁移到版本 8.0.9 1202
E.42.2. 变更 1202
E.43. 版本 8.0.8 1203
E.43.1. 迁移到版本 8.0.8 1203
E.43.2. 变更 1203
E.44. 版本 8.0.7 1204
E.44.1. 迁移到版本 8.0.7 1204
E.44.2. 变更 1204
E.45. 版本 8.0.6 1205
E.45.1. 迁移到版本 8.0.6 1205
E.45.2. 变更 1205
E.46. 版本 8.0.5 1206
E.46.1. 迁移到版本 8.0.5 1206
E.46.2. 变更 1206
E.47. 版本 8.0.4 1207
E.47.1. 迁移到版本 8.0.4 1207
E.47.2. 变更 1207
E.48. 版本 8.0.3 1208
E.48.1. 迁移到版本 8.0.3 1209
E.48.2. 变更 1209
E.49. 版本 8.0.2 1210
E.49.1. 迁移到版本 8.0.2 1210
E.49.2. 变更 1210
E.50. 版本 8.0.1 1212

E.50.1. 迁移到版本 8.0.1 1212
E.50.2. 变更 1212
E.51. 版本 8.0 1213
E.51.1. 概述 1213
E.51.2. 迁移到版本 8.0 1214
E.51.3. 废弃特性 1215
E.51.4. 变更 1215
E.52. 版本 7.4.30 1227
E.52.1. 迁移到版本 7.4.30 1227
E.52.2. 变更 1227
E.53. 版本 7.4.29 1228
E.53.1. 迁移到版本 7.4.29 1228
E.53.2. 变更 1228
E.54. 版本 7.4.28 1229
E.54.1. 迁移到版本 7.4.28 1229
E.54.2. 变更 1229
E.55. 版本 7.4.27 1230
E.55.1. 迁移到版本 7.4.27 1230
E.55.2. 变更 1230
E.56. 版本 7.4.26 1231
E.56.1. 迁移到版本 7.4.26 1231
E.56.2. 变更 1231
E.57. 版本 7.4.25 1232
E.57.1. 迁移到版本 7.4.25 1232
E.57.2. 变更 1232
E.58. 版本 7.4.24 1232
E.58.1. 迁移到版本 7.4.24 1232
E.58.2. 变更 1233
E.59. 版本 7.4.23 1233
E.59.1. 迁移到版本 7.4.23 1233
E.59.2. 变更 1233
E.60. 版本 7.4.22 1234
E.60.1. 迁移到版本 7.4.22 1234
E.60.2. 变更 1234
E.61. 版本 7.4.21 1234
E.61.1. 迁移到版本 7.4.21 1234
E.61.2. 变更 1234
E.62. 版本 7.4.20 1235
E.62.1. 迁移到版本 7.4.20 1235
E.62.2. 变更 1235
E.63. 版本 7.4.19 1236
E.63.1. 迁移到版本 7.4.19 1236
E.63.2. 变更 1236
E.64. 版本 7.4.18 1237
E.64.1. 迁移到版本 7.4.18 1237
E.64.2. 变更 1237
E.65. 版本 7.4.17 1237
E.65.1. 迁移到版本 7.4.17 1238
E.65.2. 变更 1238
E.66. 版本 7.4.16 1238
E.66.1. 迁移到版本 7.4.16 1238
E.66.2. 变更 1238
E.67. 版本 7.4.15 1239
E.67.1. 迁移到版本 7.4.15 1239

E.67.2. 变更 1239
E.68. 版本 7.4.14 1239
E.68.1. 迁移到版本 7.4.14 1239
E.68.2. 变更 1239
E.69. 版本 7.4.13 1240
E.69.1. 迁移到版本 7.4.13 1240
E.69.2. 变更 1240
E.70. 版本 7.4.12 1241
E.70.1. 迁移到版本 7.4.12 1241
E.70.2. 变更 1241
E.71. 版本 7.4.11 1242
E.71.1. 迁移到版本 7.4.11 1242
E.71.2. 变更 1242
E.72. 版本 7.4.10 1242
E.72.1. 迁移到版本 7.4.10 1242
E.72.2. 变更 1243
E.73. 版本 7.4.9 1243
E.73.1. 迁移到版本 7.4.9 1243
E.73.2. 变更 1243
E.74. 版本 7.4.8 1244
E.74.1. 迁移到版本 7.4.8 1244
E.74.2. 变更 1245
E.75. 版本 7.4.7 1246
E.75.1. 迁移到版本 7.4.7 1246
E.75.2. 变更 1246
E.76. 版本 7.4.6 1247
E.76.1. 迁移到版本 7.4.6 1247
E.76.2. 变更 1247
E.77. 版本 7.4.5 1248
E.77.1. 迁移到版本 7.4.5 1248
E.77.2. 变更 1248
E.78. 版本 7.4.4 1248
E.78.1. 迁移到版本 7.4.4 1249
E.78.2. 变更 1249
E.79. 版本 7.4.3 1249
E.79.1. 迁移到版本 7.4.3 1249
E.79.2. 变更 1249
E.80. 版本 7.4.2 1250
E.80.1. 迁移到版本 7.4.2 1250
E.80.2. 变更 1251
E.81. 版本 7.4.1 1252
E.81.1. 迁移到版本 7.4.1 1252
E.81.2. 变更 1253
E.82. 版本 7.4 1254
E.82.1. 概述 1254
E.82.2. 迁移到版本 7.4 1255
E.82.3. 变更 1256
E.83. 版本 7.3.21 1269
E.83.1. 迁移到版本 7.3.21 1269
E.83.2. 变更 1269
E.84. 版本 7.3.20 1270
E.84.1. 迁移到版本 7.3.20 1270
E.84.2. 变更 1270
E.85. 版本 7.3.19 1270

E.85.1. 迁移到版本 7.3.19 1270
E.85.2. 变更 1271
E.86. 版本 7.3.18 1271
E.86.1. 迁移到版本 7.3.18 1271
E.86.2. 变更 1271
E.87. 版本 7.3.17 1271
E.87.1. 迁移到版本 7.3.17 1271
E.87.2. 变更 1272
E.88. 版本 7.3.16 1272
E.88.1. 迁移到版本 7.3.16 1272
E.88.2. 变更 1272
E.89. 版本 7.3.15 1272
E.89.1. 迁移到版本 7.3.15 1272
E.89.2. 变更 1273
E.90. 版本 7.3.14 1273
E.90.1. 迁移到版本 7.3.14 1273
E.90.2. 变更 1274
E.91. 版本 7.3.13 1274
E.91.1. 迁移到版本 7.3.13 1274
E.91.2. 变更 1274
E.92. 版本 7.3.12 1275
E.92.1. 迁移到版本 7.3.12 1275
E.92.2. 变更 1275
E.93. 版本 7.3.11 1275
E.93.1. 迁移到版本 7.3.11 1275
E.93.2. 变更 1275
E.94. 版本 7.3.10 1276
E.94.1. 迁移到版本 7.3.10 1276
E.94.2. 变更 1277
E.95. 版本 7.3.9 1277
E.95.1. 迁移到版本 7.3.9 1278
E.95.2. 变更 1278
E.96. 版本 7.3.8 1278
E.96.1. 迁移到版本 7.3.8 1278
E.96.2. 变更 1278
E.97. 版本 7.3.7 1279
E.97.1. 迁移到版本 7.3.7 1279
E.97.2. 变更 1279
E.98. 版本 7.3.6 1279
E.98.1. 迁移到版本 7.3.6 1279
E.98.2. 变更 1279
E.99. 版本 7.3.5 1280
E.99.1. 迁移到版本 7.3.5 1280
E.99.2. 变更 1280
E.100. 版本 7.3.4 1281
E.100.1. 迁移到版本 7.3.4 1281
E.100.2. 变更 1281
E.101. 版本 7.3.3 1282
E.101.1. 迁移到版本 7.3.3 1282
E.101.2. 变更 1282
E.102. 版本 7.3.2 1284
E.102.1. 迁移到版本 7.3.2 1285
E.102.2. 变更 1285
E.103. 版本 7.3.1 1286

E.103.1. 迁移到版本 7.3.1 1286
E.103.2. 变更 1286
E.104. 版本 7.3 1287
E.104.1. 概述 1287
E.104.2. 迁移到版本 7.3 1288
E.104.3. 变更 1288
E.105. 版本 7.2.8 1299
E.105.1. 迁移到版本 7.2.8 1299
E.105.2. 变更 1299
E.106. 版本 7.2.7 1300
E.106.1. 迁移到版本 7.2.7 1300
E.106.2. 变更 1300
E.107. 版本 7.2.6 1300
E.107.1. 迁移到版本 7.2.6 1301
E.107.2. 变更 1301
E.108. 版本 7.2.5 1301
E.108.1. 迁移到版本 7.2.5 1301
E.108.2. 变更 1301
E.109. 版本 7.2.4 1302
E.109.1. 迁移到版本 7.2.4 1302
E.109.2. 变更 1302
E.110. 版本 7.2.3 1302
E.110.1. 迁移到版本 7.2.3 1302
E.110.2. 变更 1303
E.111. 版本 7.2.2 1303
E.111.1. 迁移到版本 7.2.2 1303
E.111.2. 变更 1303
E.112. 版本 7.2.1 1304
E.112.1. 迁移到版本 7.2.1 1304
E.112.2. 变更 1304
E.113. 版本 7.2 1304
E.113.1. 概述 1305
E.113.2. 迁移到版本 7.2 1305
E.113.3. 变更 1306
E.114. 版本 7.1.3 1315
E.114.1. 迁移到版本 7.1.3 1315
E.114.2. 变更 1315
E.115. 版本 7.1.2 1315
E.115.1. 迁移到版本 7.1.2 1315
E.115.2. 变更 1316
E.116. 版本 7.1.1 1316
E.116.1. 迁移到版本 7.1.1 1316
E.116.2. 变更 1316
E.117. 版本 7.1 1316
E.117.1. 迁移到版本 7.1 1317
E.117.2. 变更 1317
E.118. 版本 7.0.3 1321
E.118.1. 迁移到版本 7.0.3 1321
E.118.2. 变更 1321
E.119. 版本 7.0.2 1322
E.119.1. 迁移到版本 7.0.2 1322
E.119.2. 变更 1322
E.120. 版本 7.0.1 1322
E.120.1. 迁移到版本 7.0.1 1322

E.120.2. 变更 1322
E.121. 版本 7.0 1323
E.121.1. 迁移到版本 7.0 1323
E.121.2. 变更 1324
E.122. 版本 6.5.3 1330
E.122.1. 迁移到版本 6.5.3 1330
E.122.2. 变更 1330
E.123. 版本 6.5.2 1330
E.123.1. 迁移到版本 6.5.2 1330
E.123.2. 变更 1330
E.124. 版本 6.5.1 1331
E.124.1. 迁移到版本 6.5.1 1331
E.124.2. 变更 1331
E.125. 版本 6.5 1332
E.125.1. 迁移到版本 6.5 1333
E.125.2. 变更 1333
E.126. 版本 6.4.2 1336
E.126.1. 迁移到版本 6.4.2 1337
E.126.2. 变更 1337
E.127. 版本 6.4.1 1337
E.127.1. 迁移到版本 6.4.1 1337
E.127.2. 变更 1337
E.128. 版本 6.4 1338
E.128.1. 迁移到版本 6.4 1338
E.128.2. 变更 1338
E.129. 版本 6.3.2 1342
E.129.1. 变更 1342
E.130. 版本 6.3.1 1343
E.130.1. 变更 1343
E.131. 版本 6.3 1344
E.131.1. 迁移到版本 6.3 1345
E.131.2. 变更 1345
E.132. 版本 6.2.1 1348
E.132.1. 从版本 6.2 迁移到版本 6.2.1 1349
E.132.2. 变更 1349
E.133. 版本 6.2 1349
E.133.1. 从版本 6.1 迁移到版本 6.2 1349
E.133.2. 从版本 1.x 迁移到版本 6.2 1350
E.133.3. 变更 1350
E.134. 版本 6.1.1 1352
E.134.1. 从版本 6.1 迁移到版本 6.1.1 1352
E.134.2. 变更 1352
E.135. 版本 6.1 1352
E.135.1. 迁移到版本 6.1 1353
E.135.2. 变更 1353
E.136. 版本 6.0 1355
E.136.1. 从版本 1.09 迁移到版本 6.0 1355
E.136.2. 从 1.09 之前版本迁移到版本 6.0 1355
E.136.3. 变更 1355
E.137. 版本 1.09 1357
E.138. 版本 1.02 1357
E.138.1. 从版本 1.02 迁移到版本 1.02.1 1357
E.138.2. 转储/重载程序 1358

E.138.3. 变更 1358
E.139. 版本 1.01 1359
E.139.1. 从版本 1.0 迁移到版本 1.01 1359
E.139.2. 变更 1361
E.140. 版本 1.0 1361
E.140.1. 变更 1361
E.141. Postgres95 版本 0.03 1362
E.141.1. 变更 1362
E.142. Postgres95 版本 0.02 1364
E.142.1. 变更 1364
E.143. Postgres95 版本 0.01 1364
F. 源代码仓库 1365
F.1. 通过 Git 获取源代码 1365
G. 文档 1366
G.1. DocBook 1366
G.2. 工具集 1366
G.2.1. Linux RPM 安装 1367
G.2.2. FreeBSD 安装 1367
G.2.3. Debian 软件包 1368
G.2.4. 从源码手动安装 1368
G.2.5. 通过 `configure` 检测 1370
G.3. 构建文档 1370
G.3.1. HTML 1370
G.3.2. 手册页 1370
G.3.3. 通过 JadeTeX 生成打印输出 1371
G.3.4. 通过 RTF 生成打印输出 1371
G.3.5. 纯文本文件 1372
G.3.6. 语法检查 1372
G.4. 文档编写 1373
G.4.1. Emacs/PSGML 1373
G.4.2. 其他 Emacs 模式 1374
G.5. 风格指南 1374
G.5.1. 参考页 1374
H. 外部项目 1376
H.1. 外部开发的接口 1376
H.2. 扩展 1377

附录 A. PostgreSQL错误代码

PostgreSQL 服务器发出的所有消息都被分配了五字符错误代码，遵循 SQL 标准对“SQLSTATE”代码的约定。需要知道发生了哪一种错误条件的应用，通常应测试错误代码，而不是查看错误消息的文本内容。这些错误代码在不同 PostgreSQL 版本之间更不容易发生变化，也不会因错误消息的本地化而改变。请注意，PostgreSQL 产生的错误代码中，只有一部分由 SQL 标准定义；还有一些用于标准未定义情况的附加错误代码，是新设的或从其他数据库借用而来的。

根据标准，错误代码的前两个字符表示错误类别，后三个字符表示该类别中的特定条件。因此，即使应用无法识别某个具体错误代码，仍可能从错误类别推断出应采取的操作。

表 A.1列出了 PostgreSQL 8.1.23 中定义的所有错误代码。（其中有些当前实际上并未使用，不过它们由 SQL 标准定义。）其中也显示了错误类别。对于每个错误类别，都有一个后三个字符为 `000` 的“标准”错误代码。该代码只用于那些属于该类别但没有分配更具体代码的错误情况。

每个错误代码的 PL/pgSQL 条件名与表中所示的短语相同，只是把空格替换成下划线。例如，代码 `22012`（DIVISION BY ZERO）的条件名是 `DIVISION_BY_ZERO`。条件名可以大写或小写。（注意，PL/pgSQL 不识别警告（而非错误）的条件名；它们是类别 00、01 和 02。）

表 A.1. PostgreSQL错误代码

Error Code	Meaning	Constant
Class 00 — Successful Completion		
`00000`	SUCCESSFUL COMPLETION	successful_completion
Class 01 — Warning		
`01000`	WARNING	warning
`0100C`	DYNAMIC RESULT SETS RETURNED	dynamic_result_sets_returned
`01008`	IMPLICIT ZERO BIT PADDING	implicit_zero_bit_padding
`01003`	NULL VALUE ELIMINATED IN SET FUNCTION	null_value_eliminated_in_set_function
`01007`	PRIVILEGE NOT GRANTED	privilege_not_granted
`01006`	PRIVILEGE NOT REVOKED	privilege_not_revoked
`01004`	STRING DATA RIGHT TRUNCATION	string_data_right_truncation
`01P01`	DEPRECATED FEATURE	deprecated_feature
Class 02 — No Data (this is also a warning class per the SQL standard)		
`02000`	NO DATA	no_data
`02001`	NO ADDITIONAL DYNAMIC RESULT SETS RETURNED	no_additional_dynamic_result_sets_returned
Class 03 — SQL Statement Not Yet Complete		
`03000`	SQL STATEMENT NOT YET COMPLETE	sql_statement_not_yet_complete
Class 08 — Connection Exception		
`08000`	CONNECTION EXCEPTION	connection_exception
`08003`	CONNECTION DOES NOT EXIST	connection_does_not_exist
`08006`	CONNECTION FAILURE	connection_failure

Error Code	Meaning	Constant
08001	SQLCLIENT UNABLE TO ESTABLISH SQLCONNECTION	sqlclient_unable_to_establish_sqlconnection
08004	SQLSERVER REJECTED ESTABLISHMENT OF SQLCONNECTION	sqlserver_rejected_establishment_of_sqlconnection
08007	TRANSACTION RESOLUTION UNKNOWN	transaction_resolution_unknown
08P01	PROTOCOL VIOLATION	protocol_violation
Class 09 – Triggered Action Exception		
09000	TRIGGERED ACTION EXCEPTION	triggered_action_exception
Class 0A – Feature Not Supported		
0A000	FEATURE NOT SUPPORTED	feature_not_supported
Class 0B – Invalid Transaction Initiation		
0B000	INVALID TRANSACTION INITIATION	invalid_transaction_initiation
Class 0F – Locator Exception		
0F000	LOCATOR EXCEPTION	locator_exception
0F001	INVALID LOCATOR SPECIFICATION	invalid_locator_specification
Class 0L – Invalid Grantor		
0L000	INVALID GRANTOR	invalid_grantor
0LP01	INVALID GRANT OPERATION	invalid_grant_operation
Class 0P – Invalid Role Specification		
0P000	INVALID ROLE SPECIFICATION	invalid_role_specification
Class 21 – Cardinality Violation		
21000	CARDINALITY VIOLATION	cardinality_violation
Class 22 – Data Exception		
22000	DATA EXCEPTION	data_exception
2202E	ARRAY SUBSCRIPT ERROR	array_subscript_error
22021	CHARACTER NOT IN REPERTOIRE	character_not_in_repertoire
22008	DATETIME FIELD OVERFLOW	datetime_field_overflow
22012	DIVISION BY ZERO	division_by_zero
22005	ERROR IN ASSIGNMENT	error_in_assignment
2200B	ESCAPE CHARACTER CONFLICT	escape_character_conflict
22022	INDICATOR OVERFLOW	indicator_overflow
22015	INTERVAL FIELD OVERFLOW	interval_field_overflow
2201E	INVALID ARGUMENT FOR LOGARITHM	invalid_argument_for_logarithm
2201F	INVALID ARGUMENT FOR POWER FUNCTION	invalid_argument_for_power_function
2201G	INVALID ARGUMENT FOR WIDTH BUCKET FUNCTION	invalid_argument_for_width_bucket_function
22018	INVALID CHARACTER VALUE FOR CAST	invalid_character_value_for_cast

Error Code	Meaning	Constant
22007	INVALID DATETIME FORMAT	invalid_datetime_format
22019	INVALID ESCAPE CHARACTER	invalid_escape_character
2200D	INVALID ESCAPE OCTET	invalid_escape_octet
22025	INVALID ESCAPE SEQUENCE	invalid_escape_sequence
22P06	NONSTANDARD USE OF ESCAPE CHARACTER	nonstandard_use_of_escape_character
22010	INVALID INDICATOR PARAMETER VALUE	invalid_indicator_parameter_value
22020	INVALID LIMIT VALUE	invalid_limit_value
22023	INVALID PARAMETER VALUE	invalid_parameter_value
2201B	INVALID REGULAR EXPRESSION	invalid_regular_expression
22009	INVALID TIME ZONE DISPLA CEMENT VALUE	invalid_time_zone_displa cement_value
2200C	INVALID USE OF ESCAPE CHARACTER	invalid_use_of_escape_character
2200G	MOST SPECIFIC TYPE MISMATCH	most_specific_type_mismatch
22004	NULL VALUE NOT ALLOWED	null_value_not_allowed
22002	NULL VALUE NO INDICATOR PARAMETER	null_value_no_indicator_parame ter
22003	NUMERIC VALUE OUT OF RANGE	numeric_value_out_of_range
22026	STRING DATA LENGTH MISMAT CH	string_data_length_mismatch
22001	STRING DATA RIGHT TRUNCA TION	string_data_right_truncation
22011	SUBSTRING ERROR	substring_error
22027	TRIM ERROR	trim_error
22024	UNTERMINATED C STRING	unterminated_c_string
2200F	ZERO LENGTH CHARACTER STRING	zero_length_character_string
22P01	FLOATING POINT EXCEPTION	floating_point_exception
22P02	INVALID TEXT REPRESENTATION	invalid_text_representation
22P03	INVALID BINARY REPRESENTATI ON	invalid_binary_representation
22P04	BAD COPY FILE FORMAT	bad_copy_file_format
22P05	UNTRANSLATABLE CHARACTER	untranslatable_character
Class 23 – Integrity Constraint Violation		
23000	INTEGRITY CONSTRAINT VIOLAT ION	integrity_constraint_violation
23001	RESTRICT VIOLATION	restrict_violation
23502	NOT NULL VIOLATION	not_null_violation
23503	FOREIGN KEY VIOLATION	foreign_key_violation
23505	UNIQUE VIOLATION	unique_violation
23514	CHECK VIOLATION	check_violation

Error Code	Meaning	Constant
Class 24 – Invalid Cursor State		
24000	INVALID CURSOR STATE	invalid_cursor_state
Class 25 – Invalid Transaction State		
25000	INVALID TRANSACTION STATE	invalid_transaction_state
25001	ACTIVE SQL TRANSACTION	active_sql_transaction
25002	BRANCH TRANSACTION ALREADY ACTIVE	branch_transaction_already_active
25008	HELD CURSOR REQUIRES SAME ISOLATION LEVEL	held_cursor_requires_same_isolation_level
25003	INAPPROPRIATE ACCESS MODE FOR BRANCH TRANSACTION	inappropriate_access_mode_for_branch_transaction
25004	INAPPROPRIATE ISOLATION LEVEL FOR BRANCH TRANSACTION	inappropriate_isolation_level_for_branch_transaction
25005	NO ACTIVE SQL TRANSACTION FOR BRANCH TRANSACTION	no_active_sql_transaction_for_branch_transaction
25006	READ ONLY SQL TRANSACTION	read_only_sql_transaction
25007	SCHEMA AND DATA STATEMENT MIXING NOT SUPPORTED	schema_and_data_statement_mixing_not_supported
25P01	NO ACTIVE SQL TRANSACTION	no_active_sql_transaction
25P02	IN FAILED SQL TRANSACTION	in_failed_sql_transaction
Class 26 – Invalid SQL Statement Name		
26000	INVALID SQL STATEMENT NAME	invalid_sql_statement_name
Class 27 – Triggered Data Change Violation		
27000	TRIGGERED DATA CHANGE VIOLATION	triggered_data_change_violation
Class 28 – Invalid Authorization Specification		
28000	INVALID AUTHORIZATION SPECIFICATION	invalid_authorization_specification
Class 2B – Dependent Privilege Descriptors Still Exist		
2B000	DEPENDENT PRIVILEGE DESCRIPTORS STILL EXIST	dependent_privilege_descriptors_still_exist
2BP01	DEPENDENT OBJECTS STILL EXIST	dependent_objects_still_exist
Class 2D – Invalid Transaction Termination		
2D000	INVALID TRANSACTION TERMINATION	invalid_transaction_termination
Class 2F – SQL Routine Exception		
2F000	SQL ROUTINE EXCEPTION	sql_routine_exception
2F005	FUNCTION EXECUTED NO RETURN STATEMENT	function_executed_no_return_statement
2F002	MODIFYING SQL DATA NOT PERMITTED	modifying_sql_data_not_permitted
2F003	PROHIBITED SQL STATEMENT ATTEMPTED	prohibited_sql_statement_attempted

Error Code	Meaning	Constant
2F004	READING SQL DATA NOT PERMITTED	reading_sql_data_not_permitted
Class 34 – Invalid Cursor Name		
34000	INVALID CURSOR NAME	invalid_cursor_name
Class 38 – External Routine Exception		
38000	EXTERNAL ROUTINE EXCEPTION	external_routine_exception
38001	CONTAINING SQL NOT PERMITTED	containing_sql_not_permitted
38002	MODIFYING SQL DATA NOT PERMITTED	modifying_sql_data_not_permitted
38003	PROHIBITED SQL STATEMENT ATTEMPTED	prohibited_sql_statement_attempted
38004	READING SQL DATA NOT PERMITTED	reading_sql_data_not_permitted
Class 39 – External Routine Invocation Exception		
39000	EXTERNAL ROUTINE INVOCATION EXCEPTION	external_routine_invocation_exception
39001	INVALID SQLSTATE RETURNED	invalid_sqlstate_returned
39004	NULL VALUE NOT ALLOWED	null_value_not_allowed
39P01	TRIGGER PROTOCOL VIOLATED	trigger_protocol_violated
39P02	SRF PROTOCOL VIOLATED	srf_protocol_violated
Class 3B – Savepoint Exception		
3B000	SAVEPOINT EXCEPTION	savepoint_exception
3B001	INVALID SAVEPOINT SPECIFICATION	invalid_savepoint_specification
Class 3D – Invalid Catalog Name		
3D000	INVALID CATALOG NAME	invalid_catalog_name
Class 3F – Invalid Schema Name		
3F000	INVALID SCHEMA NAME	invalid_schema_name
Class 40 – Transaction Rollback		
40000	TRANSACTION ROLLBACK	transaction_rollback
40002	TRANSACTION INTEGRITY CONSTRAINT VIOLATION	transaction_integrity_constraint_violation
40001	SERIALIZATION FAILURE	serialization_failure
40003	STATEMENT COMPLETION UNKNOWN	statement_completion_unknown
40P01	DEADLOCK DETECTED	deadlock_detected
Class 42 – Syntax Error or Access Rule Violation		
42000	SYNTAX ERROR OR ACCESS RULE VIOLATION	syntax_error_or_access_rule_violation
42601	SYNTAX ERROR	syntax_error
42501	INSUFFICIENT PRIVILEGE	insufficient_privilege
42846	CANNOT COERCE	cannot_coerce

Error Code	Meaning	Constant
42803	GROUPING ERROR	grouping_error
42830	INVALID FOREIGN KEY	invalid_foreign_key
42602	INVALID NAME	invalid_name
42622	NAME TOO LONG	name_too_long
42939	RESERVED NAME	reserved_name
42804	DATATYPE MISMATCH	datatype_mismatch
42P18	INDETERMINATE DATATYPE	indeterminate_datatype
42809	WRONG OBJECT TYPE	wrong_object_type
42703	UNDEFINED COLUMN	undefined_column
42883	UNDEFINED FUNCTION	undefined_function
42P01	UNDEFINED TABLE	undefined_table
42P02	UNDEFINED PARAMETER	undefined_parameter
42704	UNDEFINED OBJECT	undefined_object
42701	DUPLICATE COLUMN	duplicate_column
42P03	DUPLICATE CURSOR	duplicate_cursor
42P04	DUPLICATE DATABASE	duplicate_database
42723	DUPLICATE FUNCTION	duplicate_function
42P05	DUPLICATE PREPARED STATEM ENT	duplicate_prepared_statement
42P06	DUPLICATE SCHEMA	duplicate_schema
42P07	DUPLICATE TABLE	duplicate_table
42712	DUPLICATE ALIAS	duplicate_alias
42710	DUPLICATE OBJECT	duplicate_object
42702	AMBIGUOUS COLUMN	ambiguous_column
42725	AMBIGUOUS FUNCTION	ambiguous_function
42P08	AMBIGUOUS PARAMETER	ambiguous_parameter
42P09	AMBIGUOUS ALIAS	ambiguous_alias
42P10	INVALID COLUMN REFERENCE	invalid_column_reference
42611	INVALID COLUMN DEFINITION	invalid_column_definition
42P11	INVALID CURSOR DEFINITION	invalid_cursor_definition
42P12	INVALID DATABASE DEFINITION	invalid_database_definition
42P13	INVALID FUNCTION DEFINITION	invalid_function_definition
42P14	INVALID PREPARED STATEMENT DEFINITION	invalid_prepared_statement_ definition
42P15	INVALID SCHEMA DEFINITION	invalid_schema_definition
42P16	INVALID TABLE DEFINITION	invalid_table_definition
42P17	INVALID OBJECT DEFINITION	invalid_object_definition
Class 44 – WITH CHECK OPTION Violation		
44000	WITH CHECK OPTION VIOLATION	with_check_option_violation
Class 53 – Insufficient Resources		
53000	INSUFFICIENT RESOURCES	insufficient_resources

Error Code	Meaning	Constant
53100	DISK FULL	disk_full
53200	OUT OF MEMORY	out_of_memory
53300	TOO MANY CONNECTIONS	too_many_connections
Class 54 – Program Limit Exceeded		
54000	PROGRAM LIMIT EXCEEDED	program_limit_exceeded
54001	STATEMENT TOO COMPLEX	statement_too_complex
54011	TOO MANY COLUMNS	too_many_columns
54023	TOO MANY ARGUMENTS	too_many_arguments
Class 55 – Object Not In Prerequisite State		
55000	OBJECT NOT IN PREREQUISITE STATE	object_not_in_prerequisite_state
55006	OBJECT IN USE	object_in_use
55P02	CANT CHANGE RUNTIME PARAM	cant_change_runtime_param
55P03	LOCK NOT AVAILABLE	lock_not_available
Class 57 – Operator Intervention		
57000	OPERATOR INTERVENTION	operator_intervention
57014	QUERY CANCELED	query_canceled
57P01	ADMIN SHUTDOWN	admin_shutdown
57P02	CRASH SHUTDOWN	crash_shutdown
57P03	CANNOT CONNECT NOW	cannot_connect_now
Class 58 – System Error (errors external to PostgreSQL itself)		
58030	IO ERROR	io_error
58P01	UNDEFINED FILE	undefined_file
58P02	DUPLICATE FILE	duplicate_file
Class F0 – Configuration File Error		
F0000	CONFIG FILE ERROR	config_file_error
F0001	LOCK FILE EXISTS	lock_file_exists
Class P0 – PL/pgSQL Error		
P0000	PLPGSQL ERROR	plpgsql_error
P0001	RAISE EXCEPTION	raise_exception
Class XX – Internal Error		
XX000	INTERNAL ERROR	internal_error
XX001	DATA CORRUPTED	data_corrupted
XX002	INDEX CORRUPTED	index_corrupted

附录 B. 日期/时间支持

PostgreSQL 对所有日期/时间输入都使用一个内部的启发式解析器。日期和时间以字符串形式输入，并被拆分为不同的字段，同时初步判断每个字段中可能包含哪类信息。
每个字段都会被解释，并被赋予数值、忽略或拒绝。
解析器内部为所有文本字段都维护了查找表，包括月份、星期几以及时区。

本附录包含关于这些查找表内容的信息，并描述了解析器解码日期和时间时所使用的步骤。

B.1. 日期/时间输入解释

日期/时间输入字符串按照下面的过程进行解码。

1. 将输入字符串拆分为词元，并把每个词元归类为字符串、时间、时区或数字。

   a. 如果数字词元包含冒号（`:`），则它是一个时间字符串。
      把其后的所有数字和冒号都包含进来。

   b. 如果数字词元包含连字符（`-`）、斜杠（`/`）或两个以上的点（`.`），
      则它是一个日期字符串，并且可能带有文本形式的月份。

   c. 如果词元仅由数字组成，那么它要么是单个字段，要么是 ISO 8601 的拼接式日期（例如，`19990113` 表示 1999 年 1 月 13 日）或时间（例如，`141516` 表示 14:15:16）。

   d. 如果词元以加号（`+`）或减号（`-`）开头，那么它要么是时区，要么是特殊字段。

2. 如果词元是一个纯字母字符串，则将其与可能的字符串匹配。

   a. 对该词元做二分查找表查询，将它识别为特殊字符串（例如 `today`）、星期名（例如 `Thursday`）、月份名（例如 `January`）或噪声词（例如 `at`、`on`）。

      设置字段值和字段位掩码。例如，为`today`设置年、月、日，为`now`还额外设置时、分、秒。

   b. 如果没有找到，再做一次类似的二分查找表查询，将该词元与时区匹配。

   c. 如果仍然没有找到，则抛出错误。

3. 当词元是一个数字或数字字段时：

   a. 如果有八位或六位数字，并且之前还没有读取到其他日期字段，
      则将其解释为“拼接式日期”（例如 `19990118` 或 `990118`）。其解释方式分别为 `YYYYMMDD` 或 `YYMMDD`。

   b. 如果词元是三位数字，并且已经读取到了年份，则将其解释为一年中的第几天。

   c. 如果是四位或六位数字，并且已经读取到了年份，则将其解释为时间（`HHMM` 或 `HHMMSS`）。

   d. 如果有三位或更多位数字，并且尚未找到任何日期字段，
      则将其解释为年份（这会强制其余日期字段按 yy-mm-dd 的顺序解释）。

   e. 否则，假定日期字段顺序遵循 `DateStyle` 设置：mm-dd-yy、dd-mm-yy 或 yy-mm-dd。如果发现月份字段或日字段超出范围，则抛出错误。

4. 如果指定了 BC，则将年份取反并加一后再用于内部存储。（格里高利历中没有 0 年，因此从数值上说，公元前 1 年会变成年 0。）

5. 如果没有指定 BC，并且年份字段的长度为两位，则将年份调整为四位。如果该字段小于 70，则加上 2000，否则加上 1900。

> **提示**
>
> 格里高利历公元 1-99 年可以使用带前导零的四位数字输入（例如，`0099` 表示公元 99 年）。以前版本的PostgreSQL接受三位数和一位数的年份，但从版本 7.0 开始，规则被收紧以减少歧义的可能性。

B.2. 日期/时间关键字

表 B.1展示了哪些词元会被识别为月份名称。

表 B.1. 月份名称

Month	Abbreviations
January	Jan
February	Feb
March	Mar
April	Apr
May	
June	Jun
July	Jul
August	Aug
September	Sep, Sept
October	Oct
November	Nov
December	Dec

表 B.2展示了哪些词元会被识别为星期名称。

表 B.2. 星期名称

Day	Abbreviations
Sunday	Sun
Monday	Mon
Tuesday	Tue, Tues
Wednesday	Wed, Weds
Thursday	Thu, Thur, Thurs
Friday	Fri
Saturday	Sat

表 B.3展示了用于各种修饰用途的词元。

表 B.3. 日期/时间字段修饰符

Identifier	Description
ABSTIME	Ignored
AM	Time is before 12:00
AT	忽略
JULIAN, JD, J	下一个字段是儒略日
ON	忽略
PM	时间为 12:00 或之后
T	下一个字段是时间

关键字 ABSTIME 出于历史原因而被忽略：在非常老的 PostgreSQL 版本中，abstime 类型的无效值会被输出为 Invalid Abstime。不过现在已不再如此，而且这个关键字很可能在未来版本中被删除。

表 B.4列出了 PostgreSQL在日期/时间输入值中识别的时区缩写。注意，这些名字不一定会用于日期/时间输出--输出由与当前选定的timezone参数设置相关联的官方时区缩写决定。（未来的发行版很可能也会在输入中使用timezone。）

该表按时区与UTC的偏移组织，而不是按字母顺序。
这是为了便于在本地用法与已识别缩写可能不一致的情况下进行匹配。

表 B.4. 输入用的时区缩写

时区	与 UTC 的偏移	描述
NZDT	+13:00	新西兰夏令时
IDLE	+12:00	国际日期变更线以东
NZST	+12:00	新西兰标准时间
NZT	+12:00	新西兰时间
AESST	+11:00	澳大利亚东部夏令标准时间
ACSST	+10:30	澳大利亚中部夏令标准时间
CADT	+10:30	澳大利亚中部夏令时
SADT	+10:30	南澳大利亚夏令时
AEST	+10:00	澳大利亚东部标准时间
EAST	+10:00	东澳大利亚标准时间
GST	+10:00	关岛标准时间，俄罗斯 9 区
LIGT	+10:00	澳大利亚墨尔本
SAST	+09:30	南澳大利亚标准时间
CAST	+09:30	澳大利亚中部标准时间
AWSST	+09:00	澳大利亚西部夏令标准时间
JST	+09:00	日本标准时间，俄罗斯 8 区
KST	+09:00	韩国标准时间
MHT	+09:00	夸贾林时间
WDT	+09:00	西澳大利亚夏令时
MT	+08:30	马鲁古时间
AWST	+08:00	澳大利亚西部标准时间

时区	与 UTC 的偏移	描述
CCT	+08:00	中国沿海时间
WADT	+08:00	西澳大利亚夏令时
WST	+08:00	西澳大利亚标准时间
JT	+07:30	爪哇时间
ALMST	+07:00	阿拉木图夏令时间
WAST	+07:00	西澳大利亚标准时间
CXT	+07:00	圣诞岛时间
MMT	+06:30	缅甸时间
ALMT	+06:00	阿拉木图时间
MAWT	+06:00	莫森（南极洲）时间
IOT	+05:00	印度查戈斯时间
MVT	+05:00	马尔代夫时间
TFT	+05:00	凯尔盖朗时间
AFT	+04:30	阿富汗时间
EAST	+04:00	塔那那利佛夏令时间
MUT	+04:00	毛里求斯时间
RET	+04:00	留尼汪时间
SCT	+04:00	马埃岛时间
IRT, IT	+03:30	伊朗时间
EAT	+03:00	塔那那利佛、科摩罗时间
BT	+03:00	巴格达时间
EETDST	+03:00	东欧夏令时
HMT	+03:00	希腊地中海时间（?）
BDST	+02:00	英国双夏令时
CEST	+02:00	中欧夏令时间
CETDST	+02:00	中欧夏令时
EET	+02:00	东欧时间，俄罗斯 1 区
FWT	+02:00	法国冬令时间
IST	+02:00	以色列标准时间
MEST	+02:00	中欧夏令时间
METDST	+02:00	中欧夏令时
SST	+02:00	瑞典夏令时间
BST	+01:00	英国夏令时
CET	+01:00	中欧时间
DNT	+01:00	Dansk Normal Tid
FST	+01:00	法国夏令时间
MET	+01:00	中欧时间
MEWT	+01:00	中欧冬令时间
MEZ	+01:00	Mitteleuropäische Zeit
NOR	+01:00	挪威标准时间
SET	+01:00	塞舌尔时间

时区	与 UTC 的偏移	描述
SWT	+01:00	瑞典冬令时间
WETDST	+01:00	西欧夏令时
GMT	00:00	格林尼治标准时间
UT	00:00	世界时间
UTC	00:00	世界协调时间
Z	00:00	与 UTC 相同
ZULU	00:00	与 UTC 相同
WET	00:00	西欧时间
WAT	-01:00	西非时间
FNST	-01:00	费尔南多·迪诺罗尼亚夏令时间
FNT	-02:00	费尔南多·迪诺罗尼亚时间
BRST	-02:00	巴西利亚夏令时间
NDT	-02:30	纽芬兰夏令时
ADT	-03:00	大西洋夏令时
AWT	-03:00	（未知）
BRT	-03:00	巴西利亚时间
NFT	-03:30	纽芬兰标准时间
NST	-03:30	纽芬兰标准时间
AST	-04:00	大西洋标准时间（加拿大）
ACST	-04:00	大西洋/阿克里港夏令时间
EDT	-04:00	东部夏令时
ACT	-05:00	大西洋/阿克里港标准时间
CDT	-05:00	中部夏令时
EST	-05:00	东部标准时间
CST	-06:00	中部标准时间
MDT	-06:00	山地夏令时
MST	-07:00	山地标准时间
PDT	-07:00	太平洋夏令时
AKDT	-08:00	阿拉斯加夏令时
PST	-08:00	太平洋标准时间
YDT	-08:00	育空夏令时
AKST	-09:00	阿拉斯加标准时间
HDT	-09:00	夏威夷/阿拉斯加夏令时
YST	-09:00	育空标准时间
MART	-09:30	马克萨斯时间
AHST	-10:00	阿拉斯加/夏威夷标准时间
HST	-10:00	夏威夷标准时间
CAT	-10:00	阿拉斯加中部时间
NT	-11:00	诺姆时间
IDLW	-12:00	国际日期变更线以西

澳大利亚时区. 澳大利亚时区名与南北美洲常用的时区名之间存在三个命名冲突：ACST、CST和 EST。如果运行时选项 australian_timezones设置为真，那么 ACST、CST、EST和SAT将被解释为澳大利亚时区名，如表 B.5所示。如果它为假（默认值），那么 ACST、CST和 EST将被当作美国时区名，而SAT被解释为表示星期六的噪声词。

表 B.5. 输入用的澳大利亚时区缩写

时区	与 UTC 的偏移	描述
ACST	+09:30	澳大利亚中部标准时间
CST	+10:30	澳大利亚中部标准时间
EST	+10:00	澳大利亚东部标准时间
SAT	+09:30	南澳大利亚标准时间

表 B.6列出了 PostgreSQL识别为timezone参数有效设置的时区名。注意，这些名字在概念上以及实际中都与表 B.4中所示的名字不同：这些名字大多隐含一条本地夏令时规则，而前者中的每个名字只代表一个固定的与 UTC 的偏移。

在许多情况下，同一个时区有几个等价的名字。它们列在同一行中。
该表主要按时区主要城市的名字排序。

表 B.6. 用于设置 timezone 的时区名

时区
Africa/Abidjan
Africa/Accra
Africa/Addis_Ababa
Africa/Algiers
Africa/Asmera
Africa/Bamako
Africa/Bangui
Africa/Banjul
Africa/Bissau
Africa/Blantyre
Africa/Brazzaville
Africa/Bujumbura
Africa/Cairo Egypt
Africa/Casablanca
Africa/Ceuta
Africa/Conakry
Africa/Dakar
Africa/Dar_es_Salaam
Africa/Djibouti
Africa/Douala
Africa/El_Aaiun
Africa/Freetown
Africa/Gaborone
Africa/Harare

时区
Africa/Johannesburg
Africa/Kampala
Africa/Khartoum
Africa/Kigali
Africa/Kinshasa
Africa/Lagos
Africa/Libreville
Africa/Lome
Africa/Luanda
Africa/Lubumbashi
Africa/Lusaka
Africa/Malabo
Africa/Maputo
Africa/Maseru
Africa/Mbabane
Africa/Mogadishu
Africa/Monrovia
Africa/Nairobi
Africa/Ndjamena
Africa/Niamey
Africa/Nouakchott
Africa/Ouagadougou
Africa/Porto-Novo
Africa/Sao_Tome
Africa/Timbuktu
Africa/Tripoli Libya
Africa/Tunis
Africa/Windhoek
America/Adak America/Atka US/Aleutian
America/Anchorage SystemV/YST9YDT US/Alaska
America/Anguilla
America/Antigua
America/Araguaina
America/Aruba
America/Asuncion
America/Bahia
America/Barbados
America/Belem
America/Belize
America/Boa_Vista
America/Bogota

时区
America/Boise
America/Buenos_Aires
America/Cambridge_Bay
America/Campo_Grande
America/Cancun
America/Caracas
America/Catamarca
America/Cayenne
America/Cayman
America/Chicago CST6CDT SystemV/CST6CDT US/Central
America/Chihuahua
America/Cordoba America/Rosario
America/Costa_Rica
America/Cuiaba
America/Curacao
America/Danmarkshavn
America/Dawson
America/Dawson_Creek
America/Denver MST7MDT SystemV/MST7MDT US/Mountain America/Shiprock Navajo
America/Detroit US/Michigan
America/Dominica
America/Edmonton Canada/Mountain
America/Eirunepe
America/El_Salvador
America/Ensenada America/Tijuana Mexico/BajaNorte
America/Fortaleza
America/Glace_Bay
America/Godthab
America/Goose_Bay
America/Grand_Turk
America/Grenada
America/Guadeloupe
America/Guatemala
America/Guayaquil
America/Guyana
America/Halifax Canada/Atlantic SystemV/AST4ADT
America/Havana Cuba
America/Hermosillo
America/Indiana/Indianapolis America/Indianapolis America/Fort_Wayne EST SystemV/EST5 US/East-Indiana
America/Indiana/Knox America/Knox_IN US/Indiana-Starke

时区
America/Indiana/Marengo
America/Indiana/Vevay
America/Inuvik
America/Iqaluit
America/Jamaica Jamaica
America/Jujuy
America/Juneau
America/Kentucky/Louisville America/Louisville
America/Kentucky/Monticello
America/La_Paz
America/Lima
America/Los_Angeles PST8PDT SystemV/PST8PDT US/Pacific US/Pacific-New
America/Maceio
America/Managua
America/Manaus Brazil/West
America/Martinique
America/Mazatlan Mexico/BajaSur
America/Mendoza
America/Menominee
America/Merida
America/Mexico_City Mexico/General
America/Miquelon
America/Monterrey
America/Montevideo
America/Montreal
America/Montserrat
America/Nassau
America/New_York EST5EDT SystemV/EST5EDT US/Eastern
America/Nipigon
America/Nome
America/Noronha Brazil/DeNoronha
America/North_Dakota/Center
America/Panama
America/Pangnirtung
America/Paramaribo
America/Phoenix MST SystemV/MST7 US/Arizona
America/Port-au-Prince
America/Port_of_Spain
America/Porto_Acre America/Rio_Branco Brazil/Acre
America/Porto_Velho
America/Puerto_Rico SystemV/AST4

时区
America/Rainy_River
America/Rankin_Inlet
America/Recife
America/Regina Canada/East-Saskatchewan Canada/Saskatchewan SystemV/CST6
America/Santiago Chile/Continental
America/Santo_Domingo
America/Sao_Paulo Brazil/East
America/Scoresbysund
America/St_Johns Canada/Newfoundland
America/St_Kitts
America/St_Lucia
America/St_Thomas America/Virgin
America/St_Vincent
America/Swift_Current
America/Tegucigalpa
America/Thule
America/Thunder_Bay
America/Toronto Canada/Eastern
America/Tortola
America/Vancouver Canada/Pacific
America/Whitehorse Canada/Yukon
America/Winnipeg Canada/Central
America/Yakutat
America/Yellowknife
Antarctica/Casey
Antarctica/Davis
Antarctica/DumontDUrville
Antarctica/Mawson
Antarctica/McMurdo Antarctica/South_Pole
Antarctica/Palmer
Antarctica/Rothera
Antarctica/Syowa
Antarctica/Vostok
Asia/Aden
Asia/Almaty
Asia/Amman
Asia/Anadyr
Asia/Aqtau
Asia/Aqtobe
Asia/Ashgabat Asia/Ashkhabad
Asia/Baghdad

时区
Asia/Bahrain
Asia/Baku
Asia/Bangkok
Asia/Beirut
Asia/Bishkek
Asia/Brunei
Asia/Calcutta
Asia/Choibalsan
Asia/Chongqing Asia/Chungking
Asia/Colombo
Asia/Dacca Asia/Dhaka
Asia/Damascus
Asia/Dili
Asia/Dubai
Asia/Dushanbe
Asia/Gaza
Asia/Harbin
Asia/Hong_Kong Hongkong
Asia/Hovd
Asia/Irkutsk
Asia/Jakarta
Asia/Jayapura
Asia/Jerusalem Asia/Tel_Aviv Israel
Asia/Kabul
Asia/Kamchatka
Asia/Karachi
Asia/Kashgar
Asia/Katmandu
Asia/Krasnoyarsk
Asia/Kuala_Lumpur
Asia/Kuching
Asia/Kuwait
Asia/Macao Asia/Macau
Asia/Magadan
Asia/Makassar Asia/Ujung_Pandang
Asia/Manila
Asia/Muscat
Asia/Nicosia Europe/Nicosia
Asia/Novosibirsk
Asia/Omsk
Asia/Oral

时区
Asia/Phnom_Penh
Asia/Pontianak
Asia/Pyongyang
Asia/Qatar
Asia/Qyzylorda
Asia/Rangoon
Asia/Riyadh
Asia/Riyadh87 Mideast/Riyadh87
Asia/Riyadh88 Mideast/Riyadh88
Asia/Riyadh89 Mideast/Riyadh89
Asia/Saigon
Asia/Sakhalin
Asia/Samarkand
Asia/Seoul ROK
Asia/Shanghai PRC
Asia/Singapore Singapore
Asia/Taipei ROC
Asia/Tashkent
Asia/Tbilisi
Asia/Tehran Iran
Asia/Thimbu Asia/Thimphu
Asia/Tokyo Japan
Asia/Ulaanbaatar Asia/Ulan_Bator
Asia/Urumqi
Asia/Vientiane
Asia/Vladivostok
Asia/Yakutsk
Asia/Yekaterinburg
Asia/Yerevan
Atlantic/Azores
Atlantic/Bermuda
Atlantic/Canary
Atlantic/Cape_Verde
Atlantic/Faeroe
Atlantic/Madeira
Atlantic/Reykjavik Iceland
Atlantic/South_Georgia
Atlantic/St_Helena
Atlantic/Stanley
Australia/ACT Australia/Canberra Australia/NSW Australia/Sydney
Australia/Adelaide Australia/South

时区
Australia/Brisbane Australia/Queensland
Australia/Broken_Hill Australia/Yancowinna
Australia/Darwin Australia/North
Australia/Hobart Australia/Tasmania
Australia/LHI Australia/Lord_Howe
Australia/Lindeman
Australia/Melbourne Australia/Victoria
Australia/Perth Australia/West
CET
EET
Etc/GMT+1
Etc/GMT+2
Etc/GMT+3
Etc/GMT+4
Etc/GMT+5
Etc/GMT+6
Etc/GMT+7
Etc/GMT+8
Etc/GMT+9
Etc/GMT+10
Etc/GMT+11
Etc/GMT+12
Etc/GMT-1
Etc/GMT-2
Etc/GMT-3
Etc/GMT-4
Etc/GMT-5
Etc/GMT-6
Etc/GMT-7
Etc/GMT-8
Etc/GMT-9
Etc/GMT-10
Etc/GMT-11
Etc/GMT-12
Etc/GMT-13
Etc/GMT-14
Europe/Amsterdam
Europe/Andorra
Europe/Athens
Europe/Belfast
Europe/Belgrade Europe/Ljubljana Europe/Sarajevo Europe/Skopje Europe/Zagreb

时区
Europe/Berlin
Europe/Brussels
Europe/Bucharest
Europe/Budapest
Europe/Chisinau Europe/Tiraspol
Europe/Copenhagen
Europe/Dublin Eire
Europe/Gibraltar
Europe/Helsinki
Europe/Istanbul Asia/Istanbul Turkey
Europe/Kaliningrad
Europe/Kiev
Europe/Lisbon Portugal
Europe/London GB GB-Eire
Europe/Luxembourg
Europe/Madrid
Europe/Malta
Europe/Minsk
Europe/Monaco
Europe/Moscow W-SU
Europe/Oslo Arctic/Longyearbyen Atlantic/Jan_Mayen
Europe/Paris
Europe/Prague Europe/Bratislava
Europe/Riga
Europe/Rome Europe/San_Marino Europe/Vatican
Europe/Samara
Europe/Simferopol
Europe/Sofia
Europe/Stockholm
Europe/Tallinn
Europe/Tirane
Europe/Uzhgorod
Europe/Vaduz
Europe/Vienna
Europe/Vilnius
Europe/Warsaw Poland
Europe/Zaporozhye
Europe/Zurich
Factory
GMT GMT+0 GMT-0 GMT0 Greenwich Etc/GMT Etc/GMT+0 Etc/GMT-0 Etc/GMT0 Etc/Greenwich
Indian/Antananarivo

时区
Indian/Chagos
Indian/Christmas
Indian/Cocos
Indian/Comoro
Indian/Kerguelen
Indian/Mahe
Indian/Maldives
Indian/Mauritius
Indian/Mayotte
Indian/Reunion
MET
Pacific/Apia
Pacific/Auckland NZ
Pacific/Chatham NZ-CHAT
Pacific/Easter Chile/EasterIsland
Pacific/Efate
Pacific/Enderbury
Pacific/Fakaofo
Pacific/Fiji
Pacific/Funafuti
Pacific/Galapagos
Pacific/Gambier SystemV/YST9
Pacific/Guadalcanal
Pacific/Guam
Pacific/Honolulu HST SystemV/HST10 US/Hawaii
Pacific/Johnston
Pacific/Kiritimati
Pacific/Kosrae
Pacific/Kwajalein Kwajalein
Pacific/Majuro
Pacific/Marquesas
Pacific/Midway
Pacific/Nauru
Pacific/Niue
Pacific/Norfolk
Pacific/Noumea
Pacific/Pago_Pago Pacific/Samoa US/Samoa
Pacific/Palau
Pacific/Pitcairn SystemV/PST8
Pacific/Ponape
Pacific/Port_Moresby

时区
Pacific/Rarotonga
Pacific/Saipan
Pacific/Tahiti
Pacific/Tarawa
Pacific/Tongatapu
Pacific/Truk
Pacific/Wake
Pacific/Wallis
Pacific/Yap
UCT Etc/UCT
UTC Universal Zulu Etc/UTC Etc/Universal Etc/Zulu
WET

除了表中列出的名字之外，PostgreSQL还接受 POSIX 风格的时区规范，形式为*STDoffset*或*STDoffsetDST*，其中 *STD*是一个时区缩写，*offset*是相对 UTC 向西的数字偏移（小时），而*DST*是一个可选的夏令时时区缩写，假定它表示比给定偏移提前一小时。例如，如果`EST5EDT`还不是一个已识别的时区名，它也会被接受，并且在功能上等同于美国东海岸时间。当出现夏令时时区名时，假定它按照`zic`时区数据库的`posixrules`条目所用的相同夏令时转换规则使用。在标准的PostgreSQL安装中，`posixrules`与`US/Eastern`相同，因此 POSIX 风格的时区规范遵循美国的夏令时规则。如果需要，你可以通过替换`posixrules`文件来调整这一行为。

应当警惕 POSIX 风格的时区特性可能导致悄悄接受无效输入，因为没有对时区缩写的合理性检查。例如，`SET TIMEZONE TO FOOBAR0`会生效，使系统实际上使用一个相当古怪的 UTC 缩写。

B.3. 历法的历史

儒略日由法国学者 Joseph Justus Scaliger（1540-1609）发明，其名称大概取自他的父亲、意大利学者 Julius Caesar Scaliger（1484-1558）。天文学家使用儒略周期为自公元前 4713 年 1 月 1 日以来的每一天分配一个唯一的编号。这就是所谓的儒略日（JD）。JD 0 表示从 UTC 公元前 4713 年 1 月 1 日正午到 UTC 公元前 4713 年 1 月 2 日正午的 24 小时。

“儒略日”与“儒略历”不同。儒略历由 Julius Caesar 于公元前 45 年引入。在 1582 年各国开始改用格里高利历之前，它在西方世界一直被广泛使用。在儒略历中，回归年近似为 365 1/4 天 = 365.25 天。这会在大约 128 年里累计 1 天的误差。

不断累积的历法误差促使教皇格里高利十三世按照特伦托会议的指示改革历法。
在格里高利历中，回归年近似为 365 + 97 / 400 天 = 365.2425 天。因此，大约要经过 3300 年，回归年与格里高利历之间才会累积出 1 天的偏差。

365+97/400 这一近似值是通过每 400 年设置 97 个闰年来实现的，具体规则如下：

每个能被 4 整除的年份都是闰年。
不过，每个能被 100 整除的年份都不是闰年。
但是，每个能被 400 整除的年份仍然是闰年。

因此，1700、1800、1900、2100 和 2200 都不是闰年，但 1600、2000 和 2400 是闰年。相比之下，较早的儒略历规定所有能被 4 整除的年份都是闰年。

1582 年 2 月发布的教皇诏书规定，应从 1582 年 10 月删去 10 天，使 10 月 15 日紧接在 10 月 4 日之后。这一改革在意大利、波兰、葡萄牙和西班牙得到执行。其他天主教国家不久后也

跟进，但新教国家不愿改变，而希腊正教国家直到 20 世纪初才改用格里高利历。大不列颠及其属地（包括今天的美国）在 1752 年实施了这项改革。因此，1752 年 9 月 2 日之后紧接着就是 1752 年 9 月 14 日。这就是为什么提供 `cal` 程序的 Unix 系统会产生如下输出：

```
$ cal 9 1752
   September 1752
 S  M Tu  W Th  F  S
       1  2 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
```

> **注意**
>
> SQL 标准指出：“在‘日期时间字面量’的定义中，‘日期时间值’受格里高利历日期和时间的自然规则约束”。1752-09-03 到 1752-09-13 之间的日期虽然在某些国家被教皇敕令取消，但它们符合“自然规则”，因此是有效日期。

世界上不同地区曾发展出各种历法，其中许多早于格里高利历系统。例如，
中国历法的起源可以追溯到公元前 14 世纪。传说黄帝在公元前 2637 年发明了历法。中华人民共和国在民事上使用格里高利历，中国传统历法则用于确定节日。

附录 C. SQL 关键字

表 C.1列出了在 SQL 标准以及 PostgreSQL 8.1.23 中被视为关键字的所有词元。背景信息可参见第 4.1.1 节。

SQL 区分保留关键字和非保留关键字。根据标准，保留关键字才是真正的关键字，它们绝不允许作为标识符使用。非保留关键字只在特定上下文中具有特殊含义，
在其他上下文中则可以作为标识符使用。大多数非保留关键字实际上是 SQL 规定的内置表和内置函数名称。非保留关键字这一概念，
本质上只是为了声明某个词在某些上下文中附带了预定义含义。

在PostgreSQL解析器中，情况要更复杂一些。这里的词元分属若干不同类别，
从绝不能用作标识符的词元，到在解析器中完全没有特殊地位、
只是被当作普通标识符处理的词元都有。（后一类通常就是 SQL 规定的函数名。）即使是保留关键字，在PostgreSQL中也不是绝对保留的，它们仍然可以用作列标签（例如`SELECT 55 AS CHECK`，尽管`CHECK`是一个保留关键字）。

在表 C.1中针对PostgreSQL的那一列里，我们把解析器明确识别、但又允许作为列名或表名的那些关键字归类为“非保留”。某些原本属于非保留的关键字不能用作函数名或数据类型名，
因此会另外加以标记。（这类词大多表示具有特殊语法的内置函数或数据类型。
函数或类型本身仍然可用，但用户不能重新定义它们。）
不允许作为列名或表名的词元则标为“保留”。有些保留关键字仍允许作为函数名或数据类型名，这一点也会在表中标出。如果没有这样的标记，保留关键字就只允许作为“AS”列标签名使用。

一般来说，如果你在某条命令中把这里列出的关键字当作标识符使用，
并因此遇到意外的解析器错误，应尝试为该标识符加上引号，看看问题是否消失。

在研读表 C.1之前，有一点非常重要：某个关键字在PostgreSQL中不被保留，并不意味着与该词相关的特性尚未实现；反过来，某个关键字的存在也不意味着相应特性一定存在。

表 C.1. SQL 关键字

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
A		非保留		
ABORT	非保留			
ABS		保留	非保留	
ABSOLUTE	非保留	非保留	保留	保留
ACCESS	非保留			
ACTION	非保留	非保留	保留	保留
ADA		非保留	非保留	非保留
ADD	非保留	非保留	保留	保留
ADMIN	非保留	非保留	保留	
AFTER	非保留	非保留	保留	
AGGREGATE	非保留		保留	
ALIAS			保留	
ALL	保留	保留	保留	保留
ALLOCATE		保留	保留	保留
ALSO	非保留			
ALTER	非保留	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
ALWAYS		非保留		
ANALYSE	保留			
ANALYZE	保留			
AND	保留	保留	保留	保留
ANY	保留	保留	保留	保留
ARE		保留	保留	保留
ARRAY	保留	保留	保留	
AS	保留	保留	保留	保留
ASC	保留	非保留	保留	保留
ASENSITIVE		保留	非保留	
ASSERTION	非保留	非保留	保留	保留
ASSIGNMENT	非保留	非保留	非保留	
ASYMMETRIC	保留	保留	非保留	
AT	非保留	保留	保留	保留
ATOMIC		保留	非保留	
ATTRIBUTE		非保留		
ATTRIBUTES		非保留		
AUTHORIZATION	保留（可用作函数名）	保留	保留	保留
AVG		保留	非保留	保留
BACKWARD	非保留			
BEFORE	非保留	非保留	保留	
BEGIN	非保留	保留	保留	保留
BERNOULLI		非保留		
BETWEEN	保留（可用作函数名）	保留	非保留	保留
BIGINT	非保留（不可用作函数或类型名）	保留		
BINARY	保留（可用作函数名）	保留	保留	
BIT	非保留（不可用作函数或类型名）		保留	保留
BITVAR			非保留	
BIT_LENGTH			非保留	保留
BLOB		保留	保留	
BOOLEAN	非保留（不可用作函数或类型名）	保留	保留	
BOTH	保留	保留	保留	保留
BREADTH		非保留	保留	
BY	非保留	保留	保留	保留
C		非保留	非保留	非保留
CACHE	非保留			

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
CALL		保留	保留	
CALLED	非保留	保留	非保留	
CARDINALITY		保留	非保留	
CASCADE	非保留	非保留	保留	保留
CASCADED		保留	保留	保留
CASE	保留	保留	保留	保留
CAST	保留	保留	保留	保留
CATALOG		非保留	保留	保留
CATALOG_NAME		非保留	非保留	非保留
CEIL		保留		
CEILING		保留		
CHAIN	非保留	非保留	非保留	
CHAR	非保留（不可用作函数或类型名）	保留	保留	保留
CHARACTER	非保留（不可用作函数或类型名）	保留	保留	保留
CHARACTERISTICS	非保留	非保留		
CHARACTERS		非保留		
CHARACTER_LENGTH		保留	非保留	保留
CHARACTER_SET_CATALOG		非保留	非保留	非保留
CHARACTER_SET_NAME		非保留	非保留	非保留
CHARACTER_SET_SCHEMA		非保留	非保留	非保留
CHAR_LENGTH		保留	非保留	保留
CHECK	保留	保留	保留	保留
CHECKED			非保留	
CHECKPOINT	非保留			
CLASS	非保留		保留	
CLASS_ORIGIN		非保留	非保留	非保留
CLOB		保留	保留	
CLOSE	非保留	保留	保留	保留
CLUSTER	非保留			
COALESCE	非保留（不可用作函数或类型名）	保留	非保留	保留
COBOL		非保留	非保留	非保留
COLLATE	保留	保留	保留	保留
COLLATION		非保留	保留	保留
COLLATION_CATALOG		非保留	非保留	非保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
COLLATION_NAME		非保留	非保留	非保留
COLLATION_SCHEMA		非保留	非保留	非保留
COLLECT		保留		
COLUMN	保留	保留	保留	保留
COLUMN_NAME		非保留	非保留	非保留
COMMAND_FUNCTION		非保留	非保留	非保留
COMMAND_FUNCTION_CODE		非保留	非保留	
COMMENT	非保留			
COMMIT	非保留	保留	保留	保留
COMMITTED	非保留	非保留	非保留	非保留
COMPLETION			保留	
CONDITION		保留		
CONDITION_NUMBER		非保留	非保留	非保留
CONNECT		保留	保留	保留
CONNECTION	非保留	非保留	保留	保留
CONNECTION_NAME		非保留	非保留	非保留
CONSTRAINT	保留	保留	保留	保留
CONSTRAINTS	非保留	非保留	保留	保留
CONSTRAINT_CATALOG		非保留	非保留	非保留
CONSTRAINT_NAME		非保留	非保留	非保留
CONSTRAINT_SCHEMA		非保留	非保留	非保留
CONSTRUCTOR		非保留	保留	
CONTAINS		非保留	非保留	
CONTINUE		非保留	保留	保留
CONVERSION	非保留			
CONVERT	非保留（不可用作函数或类型名）	保留	非保留	保留
COPY	非保留			
CORR		保留		
CORRESPONDING		保留	保留	保留
COUNT		保留	非保留	保留
COVAR_POP		保留		
COVAR_SAMP		保留		
CREATE	保留	保留	保留	保留
CREATEDB	非保留			

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
CREATEROLE	非保留			
CREATEUSER	非保留			
CROSS	保留（可用作函数名）	保留	保留	保留
CSV	非保留			
CUBE		保留	保留	
CUME_DIST		保留		
CURRENT		保留	保留	保留
CURRENT_DATE	保留	保留	保留	保留
CURRENT_DEFAULT_TRANSFORM_GROUP		保留		
CURRENT_PATH		保留	保留	
CURRENT_ROLE	保留	保留	保留	
CURRENT_TIME	保留	保留	保留	保留
CURRENT_TIMESTAMP	保留	保留	保留	保留
CURRENT_TRANSFORM_GROUP_FOR_TYPE		保留		
CURRENT_USER	保留	保留	保留	保留
CURSOR	非保留	保留	保留	保留
CURSOR_NAME		非保留	非保留	非保留
CYCLE	非保留	保留	保留	
DATA		非保留	保留	非保留
DATABASE	非保留			
DATE		保留	保留	保留
DATETIME_INTERVAL_CODE		非保留	非保留	非保留
DATETIME_INTERVAL_PRECISION		非保留	非保留	非保留
DAY	非保留	保留	保留	保留
DEALLOCATE	非保留	保留	保留	保留
DEC	非保留（不可用作函数或类型名）	保留	保留	保留
DECIMAL	非保留（不可用作函数或类型名）	保留	保留	保留
DECLARE	非保留	保留	保留	保留
DEFAULT	保留	保留	保留	保留
DEFAULTS	非保留	非保留		
DEFERRABLE	保留	非保留	保留	保留
DEFERRED	非保留	非保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
DEFINED		非保留	非保留	
DEFINER	非保留	非保留	非保留	
DEGREE		非保留		
DELETE	非保留	保留	保留	保留
DELIMITER	非保留			
DELIMITERS	非保留			
DENSE_RANK		保留		
DEPTH		非保留	保留	
DEREF		保留	保留	
DERIVED		非保留		
DESC	保留	非保留	保留	保留
DESCRIBE		保留	保留	保留
DESCRIPTOR		非保留	保留	保留
DESTROY			保留	
DESTRUCTOR			保留	
DETERMINISTIC		保留	保留	
DIAGNOSTICS		非保留	保留	保留
DICTIONARY			保留	
DISABLE	非保留			
DISCONNECT		保留	保留	保留
DISPATCH		非保留	非保留	
DISTINCT	保留	保留	保留	保留
DO	保留			
DOMAIN	非保留	非保留	保留	保留
DOUBLE	非保留	保留	保留	保留
DROP	非保留	保留	保留	保留
DYNAMIC		保留	保留	
DYNAMIC_FUNCTION		非保留	非保留	非保留
DYNAMIC_FUNCTION_CODE		非保留	非保留	
EACH	非保留	保留	保留	
ELEMENT		保留		
ELSE	保留	保留	保留	保留
ENABLE	非保留			
ENCODING	非保留			
ENCRYPTED	非保留			
END	保留	保留	保留	保留
END-EXEC		保留	保留	保留
EQUALS		非保留	保留	
ESCAPE	非保留	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
EVERY		保留	保留	
EXCEPT	保留	保留	保留	保留
EXCEPTION		非保留	保留	保留
EXCLUDE		非保留		
EXCLUDING	非保留	非保留		
EXCLUSIVE	非保留			
EXEC		保留	保留	保留
EXECUTE	非保留	保留	保留	保留
EXISTING			非保留	
EXISTS	非保留（不可用作函数或类型名）	保留	非保留	保留
EXP		保留		
EXPLAIN	非保留			
EXTERNAL	非保留	保留	保留	保留
EXTRACT	非保留（不可用作函数或类型名）	保留	非保留	保留
FALSE	保留	保留	保留	保留
FETCH	非保留	保留	保留	保留
FILTER		保留		
FINAL		非保留	非保留	
FIRST	非保留	非保留	保留	保留
FLOAT	非保留（不可用作函数或类型名）	保留	保留	保留
FLOOR		保留		
FOLLOWING		非保留		
FOR	保留	保留	保留	保留
FORCE	非保留			
FOREIGN	保留	保留	保留	保留
FORTRAN		非保留	非保留	非保留
FORWARD	非保留			
FOUND		非保留	保留	保留
FREE		保留	保留	
FREEZE	保留（可用作函数名）			
FROM	保留	保留	保留	保留
FULL	保留（可用作函数名）	保留	保留	保留
FUNCTION	非保留	保留	保留	
FUSION		保留		
G		非保留	非保留	
GENERAL		非保留	保留	
GENERATED		非保留	非保留	

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
GET		保留	保留	保留
GLOBAL	非保留	保留	保留	保留
GO		非保留	保留	保留
GOTO		非保留	保留	保留
GRANT	保留	保留	保留	保留
GRANTED	非保留	非保留	非保留	
GREATEST	非保留（不可用作函数或类型名）			
GROUP	保留	保留	保留	保留
GROUPING		保留	保留	
HANDLER	非保留			
HAVING	保留	保留	保留	保留
HEADER	非保留			
HIERARCHY		非保留	非保留	
HOLD	非保留	保留	非保留	
HOST			保留	
HOUR	非保留	保留	保留	保留
IDENTITY		保留	保留	保留
IGNORE			保留	
ILIKE	保留（可用作函数名）			
IMMEDIATE	非保留	非保留	保留	保留
IMMUTABLE	非保留			
IMPLEMENTATION		非保留	非保留	
IMPLICIT	非保留			
IN	保留	保留	保留	保留
INCLUDING	非保留	非保留		
INCREMENT	非保留	非保留		
INDEX	非保留			
INDICATOR		保留	保留	保留
INFIX			非保留	
INHERIT	非保留			
INHERITS	非保留			
INITIALIZE			保留	
INITIALLY	保留	非保留	保留	保留
INNER	保留（可用作函数名）	保留	保留	保留
INOUT	非保留（不可用作函数或类型名）	保留	保留	
INPUT	非保留	非保留	保留	保留
INSENSITIVE	非保留	保留	非保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
INSERT	非保留	保留	保留	保留
INSTANCE		非保留	非保留	
INSTANTIABLE		非保留	非保留	
INSTEAD	非保留			
INT	非保留（不可用作函数或类型名）	保留	保留	保留
INTEGER	非保留（不可用作函数或类型名）	保留	保留	保留
INTERSECT	保留	保留	保留	保留
INTERSECTION		保留		
INTERVAL	非保留（不可用作函数或类型名）	保留	保留	保留
INTO	保留	保留	保留	保留
INVOKER	非保留	非保留	非保留	
IS	保留（可用作函数名）	保留	保留	保留
ISNULL	保留（可用作函数名）			
ISOLATION	非保留	非保留	保留	保留
ITERATE			保留	
JOIN	保留（可用作函数名）	保留	保留	保留
K		非保留	非保留	
KEY	非保留	非保留	保留	保留
KEY_MEMBER		非保留	非保留	
KEY_TYPE		非保留	非保留	
LANCOMPILER	非保留			
LANGUAGE	非保留	保留	保留	保留
LARGE	非保留	保留	保留	
LAST	非保留	非保留	保留	保留
LATERAL		保留	保留	
LEADING	保留	保留	保留	保留
LEAST	非保留（不可用作函数或类型名）			
LEFT	保留（可用作函数名）	保留	保留	保留
LENGTH		非保留	非保留	非保留
LESS			保留	
LEVEL	非保留	非保留	保留	保留
LIKE	保留（可用作函数名）	保留	保留	保留
LIMIT	保留		保留	
LISTEN	非保留			

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
LN		保留		
LOAD	非保留			
LOCAL	非保留	保留	保留	保留
LOCALTIME	保留	保留	保留	
LOCALTIMESTAMP	保留	保留	保留	
LOCATION	非保留			
LOCATOR		非保留	保留	
LOCK	非保留			
LOGIN	非保留			
LOWER		保留	非保留	保留
M		非保留	非保留	
MAP		非保留	保留	
MATCH	非保留	保留	保留	保留
MATCHED		非保留		
MAX		保留	非保留	保留
MAXVALUE	非保留	非保留		
MEMBER		保留		
MERGE		保留		
MESSAGE_LENGTH		非保留	非保留	非保留
MESSAGE_OCTET_LENGTH		非保留	非保留	非保留
MESSAGE_TEXT		非保留	非保留	非保留
METHOD		保留	非保留	
MIN		保留	非保留	保留
MINUTE	非保留	保留	保留	保留
MINVALUE	非保留	非保留		
MOD		保留	非保留	
MODE	非保留			
MODIFIES		保留	保留	
MODIFY			保留	
MODULE		保留	保留	保留
MONTH	非保留	保留	保留	保留
MORE		非保留	非保留	非保留
MOVE	非保留			
MULTISET		保留		
MUMPS		非保留	非保留	非保留
NAME		非保留	非保留	非保留
NAMES	非保留	非保留	保留	保留
NATIONAL	非保留（不可用作函数或类型名）	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
NATURAL	保留（可用作函数名）	保留	保留	保留
NCHAR	非保留（不可用作函数或类型名）	保留	保留	保留
NCLOB		保留	保留	
NESTING		非保留		
NEW	保留	保留	保留	
NEXT	非保留	非保留	保留	保留
NO	非保留	保留	保留	保留
NOCREATEDB	非保留			
NOCREATEROLE	非保留			
NOCREATEUSER	非保留			
NOINHERIT	非保留			
NOLOGIN	非保留			
NONE	非保留（不可用作函数或类型名）	保留	保留	
NORMALIZE		保留		
NORMALIZED		非保留		
NOSUPERUSER	非保留			
NOT	保留	保留	保留	保留
NOTHING	非保留			
NOTIFY	非保留			
NOTNULL	保留（可用作函数名）			
NOWAIT	非保留			
NULL	保留	保留	保留	保留
NULLABLE		非保留	非保留	非保留
NULLIF	非保留（不可用作函数或类型名）	保留	非保留	保留
NULLS		非保留		
NUMBER		非保留	非保留	非保留
NUMERIC	非保留（不可用作函数或类型名）	保留	保留	保留
OBJECT	非保留	非保留	保留	
OCTETS		非保留		
OCTET_LENGTH		保留	非保留	保留
OF	非保留	保留	保留	保留
OFF	保留		保留	
OFFSET	保留			
OIDS	非保留			
OLD	保留	保留	保留	
ON	保留	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
ONLY	保留	保留	保留	保留
OPEN		保留	保留	保留
OPERATION			保留	
OPERATOR	非保留			
OPTION	非保留	非保留	保留	保留
OPTIONS		非保留	非保留	
OR	保留	保留	保留	保留
ORDER	保留	保留	保留	保留
ORDERING		非保留		
ORDINALITY		非保留	保留	
OTHERS		非保留		
OUT	非保留（不可用作函数或类型名）	保留	保留	
OUTER	保留（可用作函数名）	保留	保留	保留
OUTPUT		非保留	保留	保留
OVER		保留		
OVERLAPS	保留（可用作函数名）	保留	非保留	保留
OVERLAY	非保留（不可用作函数或类型名）	保留	非保留	
OVERRIDING		非保留	非保留	
OWNER	非保留			
PAD		非保留	保留	保留
PARAMETER		保留	保留	
PARAMETERS			保留	
PARAMETER_MODE		非保留	非保留	
PARAMETER_NAME		非保留	非保留	
PARAMETER_ ORDINAL_POSITI ON		非保留	非保留	
PARAMETER_ SPECIFIC_ CATALOG		非保留	非保留	
PARAMETER_ SPECIFIC_NAME		非保留	非保留	
PARAMETER_ SPECIFIC_ SCHEMA		非保留	非保留	
PARTIAL	非保留	非保留	保留	保留
PARTITION		保留		
PASCAL		非保留	非保留	非保留
PASSWORD	非保留			

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
PATH		非保留	保留	
PERCENTILE_CONT		保留		
PERCENTILE_DISC		保留		
PERCENT_RANK		保留		
PLACING	保留	非保留		
PLI		非保留	非保留	非保留
POSITION	非保留（不可用作函数或类型名）	保留	非保留	保留
POSTFIX			保留	
POWER		保留		
PRECEDING		非保留		
PRECISION	非保留（不可用作函数或类型名）	保留	保留	保留
PREFIX			保留	
PREORDER			保留	
PREPARE	非保留	保留	保留	保留
PREPARED	非保留			
PRESERVE	非保留	非保留	保留	保留
PRIMARY	保留	保留	保留	保留
PRIOR	非保留	非保留	保留	保留
PRIVILEGES	非保留	非保留	保留	保留
PROCEDURAL	非保留			
PROCEDURE	非保留	保留	保留	保留
PUBLIC		非保留	保留	保留
QUOTE	非保留			
RANGE		保留		
RANK		保留		
READ	非保留	非保留	保留	保留
READS		保留	保留	
REAL	非保留（不可用作函数或类型名）	保留	保留	保留
RECHECK	非保留			
RECURSIVE		保留	保留	
REF		保留	保留	
REFERENCES	保留	保留	保留	保留
REFERENCING		保留	保留	
REGR_AVGX		保留		
REGR_AVGY		保留		
REGR_COUNT		保留		
REGR_INTERCEPT		保留		

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
REGR_R2		保留		
REGR_SLOPE		保留		
REGR_SXX		保留		
REGR_SXY		保留		
REGR_SYY		保留		
REINDEX	非保留			
RELATIVE	非保留	非保留	保留	保留
RELEASE	非保留	保留		
RENAME	非保留			
REPEATABLE	非保留	非保留	非保留	非保留
REPLACE	非保留			
RESET	非保留			
RESTART	非保留	非保留		
RESTRICT	非保留	非保留	保留	保留
RESULT		保留	保留	
RETURN		保留	保留	
RETURNED_CARDINALITY		非保留		
RETURNED_LENGTH		非保留	非保留	非保留
RETURNED_OCTET_LENGTH		非保留	非保留	非保留
RETURNED_SQLSTATE		非保留	非保留	非保留
RETURNS	非保留	保留	保留	
REVOKE	非保留	保留	保留	保留
RIGHT	保留（可用作函数名）	保留	保留	保留
ROLE	非保留	非保留	保留	
ROLLBACK	非保留	保留	保留	保留
ROLLUP		保留	保留	
ROUTINE		非保留	保留	
ROUTINE_CATALOG		非保留	非保留	
ROUTINE_NAME		非保留	非保留	
ROUTINE_SCHEMA		非保留	非保留	
ROW	非保留（不可用作函数或类型名）	保留	保留	
ROWS	非保留	保留	保留	保留
ROW_COUNT		非保留	非保留	非保留
ROW_NUMBER		保留		
RULE	非保留			

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
SAVEPOINT	非保留	保留	保留	
SCALE		非保留	非保留	非保留
SCHEMA	非保留	非保留	保留	保留
SCHEMA_NAME		非保留	非保留	非保留
SCOPE		保留	保留	
SCOPE_CATALOG		非保留		
SCOPE_NAME		非保留		
SCOPE_SCHEMA		非保留		
SCROLL	非保留	保留	保留	保留
SEARCH		保留	保留	
SECOND	非保留	保留	保留	保留
SECTION		非保留	保留	保留
SECURITY	非保留	非保留	非保留	
SELECT	保留	保留	保留	保留
SELF		非保留	非保留	
SENSITIVE		保留	非保留	
SEQUENCE	非保留	非保留	保留	
SERIALIZABLE	非保留	非保留	非保留	非保留
SERVER_NAME		非保留	非保留	非保留
SESSION	非保留	非保留	保留	保留
SESSION_USER	保留	保留	保留	保留
SET	非保留	保留	保留	保留
SETOF	非保留（不可用作函数或类型名）			
SETS		非保留	保留	
SHARE	非保留			
SHOW	非保留			
SIMILAR	保留（可用作函数名）	保留	非保留	
SIMPLE	非保留	非保留	非保留	
SIZE		非保留	保留	保留
SMALLINT	非保留（不可用作函数或类型名）	保留	保留	保留
SOME	保留	保留	保留	保留
SOURCE		非保留	非保留	
SPACE		非保留	保留	保留
SPECIFIC		保留	保留	
SPECIFICTYPE		保留	保留	
SPECIFIC_NAME		非保留	非保留	
SQL		保留	保留	保留
SQLCODE				保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
SQLERROR				保留
SQLEXCEPTION		保留	保留	
SQLSTATE		保留	保留	保留
SQLWARNING		保留	保留	
SQRT		保留		
STABLE	非保留			
START	非保留	保留	保留	
STATE		非保留	保留	
STATEMENT	非保留	非保留	保留	
STATIC		保留	保留	
STATISTICS	非保留			
STDDEV_POP		保留		
STDDEV_SAMP		保留		
STDIN	非保留			
STDOUT	非保留			
STORAGE	非保留			
STRICT	非保留			
STRUCTURE		非保留	保留	
STYLE		非保留	非保留	
SUBCLASS_ ORIGIN		非保留	非保留	非保留
SUBLIST			非保留	
SUBMULTISET		保留		
SUBSTRING	非保留（不可用作函数或类型名）	保留	非保留	保留
SUM		保留	非保留	保留
SUPERUSER	非保留			
SYMMETRIC	保留	保留	非保留	
SYSID	非保留			
SYSTEM	非保留	保留	非保留	
SYSTEM_USER		保留	保留	保留
TABLE	保留	保留	保留	保留
TABLESAMPLE		保留		
TABLESPACE	非保留			
TABLE_NAME		非保留	非保留	非保留
TEMP	非保留			
TEMPLATE	非保留			
TEMPORARY	非保留	非保留	保留	保留
TERMINATE			保留	
THAN			保留	
THEN	保留	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
TIES		非保留		
TIME	非保留（不可用作函数或类型名）	保留	保留	保留
TIMESTAMP	非保留（不可用作函数或类型名）	保留	保留	保留
TIMEZONE_HOUR		保留	保留	保留
TIMEZONE_MINUTE		保留	保留	保留
TO	保留	保留	保留	保留
TOAST	非保留			
TOP_LEVEL_COUNT		非保留		
TRAILING	保留	保留	保留	保留
TRANSACTION	非保留	非保留	保留	保留
TRANSACTIONS_COMMITTED		非保留	非保留	
TRANSACTIONS_ROLLED_BACK		非保留	非保留	
TRANSACTION_ACTIVE		非保留	非保留	
TRANSFORM		非保留	非保留	
TRANSFORMS		非保留	非保留	
TRANSLATE		保留	非保留	保留
TRANSLATION		保留	保留	保留
TREAT	非保留（不可用作函数或类型名）	保留	保留	
TRIGGER	非保留	保留	保留	
TRIGGER_CATALOG		非保留	非保留	
TRIGGER_NAME		非保留	非保留	
TRIGGER_SCHEMA		非保留	非保留	
TRIM	非保留（不可用作函数或类型名）	保留	非保留	保留
TRUE	保留	保留	保留	保留
TRUNCATE	非保留			
TRUSTED	非保留			
TYPE	非保留	非保留	非保留	非保留
UESCAPE		保留		
UNBOUNDED		非保留		
UNCOMMITTED	非保留	非保留	非保留	非保留
UNDER		非保留	保留	
UNENCRYPTED	非保留			
UNION	保留	保留	保留	保留

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
UNIQUE	保留	保留	保留	保留
UNKNOWN	非保留	保留	保留	保留
UNLISTEN	非保留			
UNNAMED		非保留	非保留	非保留
UNNEST		保留	保留	
UNTIL	非保留			
UPDATE	非保留	保留	保留	保留
UPPER		保留	非保留	保留
USAGE		非保留	保留	保留
USER	保留	保留	保留	保留
USER_DEFINED_TYPE_CATALOG		非保留	非保留	
USER_DEFINED_TYPE_CODE		非保留		
USER_DEFINED_TYPE_NAME		非保留	非保留	
USER_DEFINED_TYPE_SCHEMA		非保留	非保留	
USING	保留	保留	保留	保留
VACUUM	非保留			
VALID	非保留			
VALIDATOR	非保留			
VALUE		保留	保留	保留
VALUES	非保留	保留	保留	保留
VARCHAR	非保留（不可用作函数或类型名）	保留	保留	保留
VARIABLE			保留	
VARYING	非保留	保留	保留	保留
VAR_POP		保留		
VAR_SAMP		保留		
VERBOSE	保留（可用作函数名）			
VIEW	非保留	非保留	保留	保留
VOLATILE	非保留			
WHEN	保留	保留	保留	保留
WHENEVER		保留	保留	保留
WHERE	保留	保留	保留	保留
WIDTH_BUCKET		保留		
WINDOW		保留		
WITH	非保留	保留	保留	保留
WITHIN		保留		
WITHOUT	非保留	保留	保留	

Key Word	PostgreSQL	SQL:2003	SQL:1999	SQL-92
WORK	非保留	非保留	保留	保留
WRITE	非保留	非保留	保留	保留
YEAR	非保留		保留	保留
ZONE	非保留	非保留	保留	保留

附录 D. SQL 符合性

本节尝试概述PostgreSQL在多大程度上符合当前的 SQL 标准。下文并不是一份完整的符合性声明，但会在合理且对用户有用的范围内，尽可能详细地介绍主要主题。

SQL 标准的正式名称是 ISO/IEC 9075“数据库语言 SQL”。该标准不时发布修订版本；最近一次修订发布于 2003 年末。该版本称为 ISO/IEC 9075:2003，或简称 SQL:2003。在此之前的版本是 SQL:1999 和 SQL-92。每个版本都取代前一个版本，因此声称符合早期版本并没有官方意义。PostgreSQL的开发目标是在不与传统特性或常识相矛盾的前提下，符合该标准的最新官方版本。PostgreSQL 项目在 SQL:2003 制定期间并未参与 ISO/IEC 9075 工作组，但即便如此，SQL:2003 要求的许多特性都得到了支持，尽管有时语法或功能略有差异。今后可以预期会进一步向标准靠拢。

SQL-92为符合性定义了三个特性集：入门（Entry）、中级（Intermediate）和完整（Full）。多数宣称符合SQL标准的数据库管理系统，实际上只符合入门（Entry）级别，因为中级（Intermediate）和完整（Full）级别中的全部特性要么过于庞杂，要么与既有行为相冲突。

从SQL:1999开始，SQL 标准不再沿用SQL-92中那种效果不佳且过于宽泛的三级分类，而是定义了大量独立特性。其中很大一个子集构成了核心（“Core”）特性，也就是每个符合 SQL 标准的实现都必须提供的特性。其余特性则完全是可选的。某些可选特性被组合在一起，形成“包”，SQL 实现可以声明符合这些包，从而表明符合特定的特性组。

SQL:2003标准也分为若干部分，每部分都有一个简称。注意这些部分的编号并不连续。

- ISO/IEC 9075-1 框架（SQL/Framework）
- ISO/IEC 9075-2 基础（SQL/Foundation）
- ISO/IEC 9075-3 调用级接口（SQL/CLI）
- ISO/IEC 9075-4 持久存储模块（SQL/PSM）
- ISO/IEC 9075-9 外部数据管理（SQL/MED）
- ISO/IEC 9075-10 对象语言绑定（SQL/OLB）
- ISO/IEC 9075-11 信息和定义模式（SQL/Schemata）
- ISO/IEC 9075-13 使用 Java 语言的例程和类型（SQL/JRT）
- ISO/IEC 9075-14 XML 相关规范（SQL/XML）

PostgreSQL涵盖第 1、2 和 11 部分。第 3 部分类似于 ODBC 接口，第 4 部分类似于PL/pgSQL编程语言，但在这两种情况下，都未专门打算或核验确切的符合性。

PostgreSQL 支持 SQL:2003 的大部分主要特性。在完全符合 Core 所要求的 164 个强制特性中，PostgreSQL 至少符合 150 个。此外，还有一长串受支持的可选特性。值得注意的是，截至撰写时，尚无任何数据库管理系统的当前版本声称完全符合 Core SQL:2003。

在下面两节中，我们先列出PostgreSQL支持的特性，然后列出SQL:2003中定义但PostgreSQL尚不支持的特性。这两个列表都是近似的：被列为受支持的特性可能存在个别不符合标准的细节，而不受支持的特性也可能实际上已实现了很大一部分。关于什么可用什么不可用的最准确信息，始终以文档正文为准。

注意

带连字符的特性代码表示子特性。因此，如果某个特定子特性不受支持，则即使该主特性的其他子特性受支持，该主特性也会被列为不受支持。

D.1. 支持的特性

标识符	包	描述	注释
B012		Embedded C	
B021		Direct SQL	
E011	Core	Numeric data types	
E011-01	Core	INTEGER and SMALLINT data types	
E011-02	Core	REAL, DOUBLE PRECIS ION, and FLOAT data types	
E011-03	Core	DECIMAL and NUMERIC data types	
E011-04	Core	Arithmetic operators	
E011-05	Core	Numeric comparison	
E011-06	Core	Implicit casting among the numeric data types	
E021	Core	Character data types	
E021-01	Core	CHARACTER data type	
E021-02	Core	CHARACTER VARYING data type	
E021-03	Core	Character literals	
E021-04	Core	CHARACTER_LENGTH function	trims trailing spaces from CHARACTER values before counting
E021-05	Core	OCTET_LENGTH functi on	
E021-06	Core	SUBSTRING function	
E021-07	Core	Character concatenatio n	
E021-08	Core	UPPER and LOWER functions	
E021-09	Core	TRIM function	
E021-10	Core	Implicit casting among the character string types	
E021-11	Core	POSITION function	
E021-12	Core	Character comparison	
E031	Core	Identifiers	
E031-01	Core	Delimited identifiers	

标识符	包	描述	注释
E031-02	Core	Lower case identifiers	
E031-03	Core	Trailing underscore	
E051	Core	Basic query specificatio n	
E051-01	Core	SELECT DISTINCT	
E051-02	Core	GROUP BY clause	
E051-04	Core	GROUP BY can contain columns not in <select list>	
E051-05	Core	Select list items can be renamed	AS is required
E051-06	Core	HAVING clause	
E051-07	Core	Qualified * in select list	
E051-08	Core	Correlation names in the FROM clause	
E051-09	Core	Rename columns in the FROM clause	
E061	Core	Basic predicates and search conditions	
E061-01	Core	Comparison predicate	
E061-02	Core	BETWEEN predicate	
E061-03	Core	IN predicate with list of values	
E061-04	Core	LIKE predicate	
E061-05	Core	LIKE predicate ESCAPE clause	
E061-06	Core	NULL predicate	
E061-07	Core	Quantified comparison predicate	
E061-08	Core	EXISTS predicate	
E061-09	Core	Subqueries in compar ison predicate	
E061-11	Core	Subqueries in IN predic ate	
E061-12	Core	Subqueries in quanti fied comparison predic ate	
E061-13	Core	Correlated subqueries	
E061-14	Core	Search condition	
E071	Core	Basic query expres sions	
E071-01	Core	UNION DISTINCT table operator	
E071-02	Core	UNION ALL table operator	

标识符	包	描述	注释
E071-03	Core	EXCEPT DISTINCT table operator	
E071-05	Core	Columns combined via table operators need not have exactly the same data type	
E071-06	Core	Table operators in subqueries	
E081-01	Core	SELECT privilege	
E081-02	Core	DELETE privilege	
E081-03	Core	INSERT privilege at the table level	
E081-04	Core	UPDATE privilege at the table level	
E081-06	Core	REFERENCES privilege at the table level	
E081-08	Core	WITH GRANT OPTION	
E081-10	Core	EXECUTE privilege	
E091	Core	Set functions	
E091-01	Core	AVG	
E091-02	Core	COUNT	
E091-03	Core	MAX	
E091-04	Core	MIN	
E091-05	Core	SUM	
E091-06	Core	ALL quantifier	
E091-07	Core	DISTINCT quantifier	
E101	Core	Basic data manipu lation	
E101-01	Core	INSERT statement	
E101-03	Core	Searched UPDATE statement	
E101-04	Core	Searched DELETE statement	
E111	Core	Single row SELECT statement	
E121-01	Core	DECLARE CURSOR	
E121-02	Core	ORDER BY columns need not be in select list	
E121-03	Core	Value expressions in ORDER BY clause	
E121-04	Core	OPEN statement	
E121-08	Core	CLOSE statement	
E121-10	Core	FETCH statement implic it NEXT	
E121-17	Core	WITH HOLD cursors	

标识符	包	描述	注释
E131	Core	Null value support (nulls in lieu of values)	
E141	Core	Basic integrity constr aints	
E141-01	Core	NOT NULL constraints	
E141-02	Core	UNIQUE constraints of NOT NULL columns	
E141-03	Core	PRIMARY KEY constr aints	
E141-04	Core	Basic FOREIGN KEY constraint with the NO ACTION default for both referential delete action and referential update action	
E141-06	Core	CHECK constraints	
E141-07	Core	Column defaults	
E141-08	Core	NOT NULL inferred on PRIMARY KEY	
E141-10	Core	Names in a foreign key can be specified in any order	
E151	Core	Transaction support	
E151-01	Core	COMMIT statement	
E151-02	Core	ROLLBACK statement	
E152	Core	Basic SET TRANSA CTION statement	
E152-01	Core	SET TRANSACTION statement: ISOLATION LEVEL SERIALIZABLE clause	
E152-02	Core	SET TRANSACTION statement: READ ONLY and READ WRITE clause s	
E161	Core	SQL comments using leading double minus	
E171	Core	SQLSTATE support	
F021	Core	Basic information schema	
F021-01	Core	COLUMNS view	
F021-02	Core	TABLES view	
F021-03	Core	VIEWS view	
F021-04	Core	TABLE_CONSTRAINTS view	
F021-05	Core	REFERENTIAL_CONSTR AINTS view	

标识符	包	描述	注释
F021-06	Core	CHECK_CONSTRAINTS view	
F031	Core	Basic schema manipu lation	
F031-01	Core	CREATE TABLE statem ent to create persistent base tables	
F031-02	Core	CREATE VIEW statement	
F031-03	Core	GRANT statement	
F031-04	Core	ALTER TABLE statem ent: ADD COLUMN clause	
F031-13	Core	DROP TABLE statement: RESTRICT clause	
F031-16	Core	DROP VIEW statement: RESTRICT clause	
F031-19	Core	REVOKE statement: RESTRICT clause	
F032		CASCADE drop behavi or	
F033		ALTER TABLE statem ent: DROP COLUMN clause	
F034		Extended REVOKE statement	
F034-01		REVOKE statement performed by other than the owner of a schema object	
F034-02		REVOKE statement: GRANT OPTION FOR clause	
F034-03		REVOKE statement to revoke a privilege that the grantee has WITH GRANT OPTION	
F041	Core	Basic joined table	
F041-01	Core	Inner join (but not necessarily the INNER keyword)	
F041-02	Core	INNER keyword	
F041-03	Core	LEFT OUTER JOIN	
F041-04	Core	RIGHT OUTER JOIN	
F041-05	Core	Outer joins can be nested	
F041-07	Core	The inner table in a left or right outer join can also be used in an inner join	

标识符	包	描述	注释
F041-08	Core	All comparison operat ors are supported (rather than just =)	
F051	Core	Basic date and time	
F051-01	Core	DATE data type (includ ing support of DATE literal)	
F051-02	Core	TIME data type (includ ing support of TIME literal) with fractional seconds precision of at least 0	
F051-03	Core	TIMESTAMP data type (including support of TIMESTAMP literal) with fractional second s precision of at least 0 and 6	
F051-04	Core	Comparison predicate on DATE, TIME, and TIMESTAMP data types	
F051-05	Core	Explicit CAST betwee n datetime types and character string types	
F051-06	Core	CURRENT_DATE	
F051-07	Core	LOCALTIME	
F051-08	Core	LOCALTIMESTAMP	
F052	Enhanced datetime facilities	Intervals and datetime arithmetic	
F053		OVERLAPS predicate	
F081	Core	UNION and EXCEPT in views	
F111		Isolation levels other than SERIALIZABLE	
F111-01		READ UNCOMMITTED isolation level	
F111-02		READ COMMITTED isolation level	
F111-03		REPEATABLE READ isolation level	
F131	Core	Grouped operations	
F131-01	Core	WHERE, GROUP BY, and HAVING clauses supported in queries with grouped views	
F131-02	Core	Multiple tables suppor ted in queries with grouped views	

标识符	包	描述	注释
F131-03	Core	Set functions suppor ted in queries with grouped views	
F131-04	Core	Subqueries with GROUP BY and HAVING clauses and grouped views	
F131-05	Core	Single row SELECT with GROUP BY and HAVING clauses and grouped views	
F171		Multiple schemas per user	
F191	Enhanced integrity management	Referential delete actions	
F201	Core	CAST function	
F221	Core	Explicit defaults	
F222		INSERT statement: DEFAULT VALUES clause	
F231		Privilege tables	
F231-01		TABLE_PRIVILEGES view	
F231-02		COLUMN_PRIVILEGES view	
F231-03		USAGE_PRIVILEGES view	
F251		Domain support	
F261	Core	CASE expression	
F261-01	Core	Simple CASE	
F261-02	Core	Searched CASE	
F261-03	Core	NULLIF	
F261-04	Core	COALESCE	
F271		Compound character literals	
F281		LIKE enhancements	
F302		INTERSECT table operat or	
F302-01		INTERSECT DISTINCT table operator	
F302-02		INTERSECT ALL table operator	
F304		EXCEPT ALL table operator	
F311-01	Core	CREATE SCHEMA	
F311-02	Core	CREATE TABLE for persis tent base tables	
F311-03	Core	CREATE VIEW	

标识符	包	描述	注释
F311-05	Core	GRANT statement	
F321		User authorization	
F361		Subprogram support	
F381		Extended schema manipulation	
F381-01		ALTER TABLE statem ent: ALTER COLUMN clause	
F381-02		ALTER TABLE statem ent: ADD CONSTRAINT clause	
F381-03		ALTER TABLE statem ent: DROP CONSTRAINT clause	
F391		Long identifiers	
F401		Extended joined table	
F401-01		NATURAL JOIN	
F401-02		FULL OUTER JOIN	
F401-04		CROSS JOIN	
F411	Enhanced datetime facilities	Time zone specification	differences regarding literal interpretation
F421		National character	
F431		Read-only scrollable cursors	
F431-01		FETCH with explicit NEXT	
F431-02		FETCH FIRST	
F431-03		FETCH LAST	
F431-04		FETCH PRIOR	
F431-05		FETCH ABSOLUTE	
F431-06		FETCH RELATIVE	
F441		Extended set function support	
F471	Core	Scalar subquery values	
F481	Core	Expanded NULL predic ate	
F491	Enhanced integrity management	Constraint manage ment	
F501	Core	Features and confor mance views	
F501-01	Core	SQL_FEATURES view	
F501-02	Core	SQL_SIZING view	
F501-03	Core	SQL_LANGUAGES view	
F502		Enhanced docume ntation tables	

标识符	包	描述	注释
F502-01		SQL_SIZING_PROFILES view	
F502-02		SQL_IMPLEMENTATION_INFO view	
F502-03		SQL_PACKAGES view	
F531		Temporary tables	
F555	Enhanced datetime facilities	Enhanced seconds precision	
F561		Full value expressions	
F571		Truth value tests	
F591		Derived tables	
F611		Indicator data types	
F651		Catalog name qualif iers	
F672		Retrospective check constraints	
F701	Enhanced integrity management	Referential update actions	
F711		ALTER domain	
F761		Session management	
F771		Connection manage ment	
F781		Self-referencing operat ions	
F791		Insensitive cursors	
F801		Full set function	
S071	Enhanced object support	SQL paths in function and type name resolu tion	
S111	Enhanced object support	ONLY in query expres sions	
S211	Enhanced object support	User-defined cast functions	
T031		BOOLEAN data type	
T071		BIGINT data type	
T141		SIMILAR predicate	
T151		DISTINCT predicate	
T171		LIKE clause in table definition	
T191	Enhanced integrity management	Referential action RESTRICT	
T201	Enhanced integrity management	Comparable data types for referential constr aints	

标识符	包	描述	注释
T211-01	Active database, Enhanced integrity management	Triggers activated on UPDATE, INSERT, or DELETE of one base table	
T211-02	Active database, Enhanced integrity management	BEFORE triggers	
T211-03	Active database, Enhanced integrity management	AFTER triggers	
T211-04	Active database, Enhanced integrity management	FOR EACH ROW triggers	
T211-07	Active database, Enhanced integrity management	TRIGGER privilege	
T212	Enhanced integrity management	Enhanced trigger capability	
T231		Sensitive cursors	
T241		START TRANSACTION statement	
T271		Savepoints	
T312		OVERLAY function	
T321-01	Core	User-defined functions with no overloading	
T321-03	Core	Function invocation	
T321-06	Core	ROUTINES view	
T321-07	Core	PARAMETERS view	
T322	PSM	Overloading of SQL-invoked functions and procedures	
T323		Explicit security for external routines	
T351		Bracketed SQL comments (/*...*/ comments)	
T441		ABS and MOD functions	
T461		Symmetric BETWEEN predicate	
T501		Enhanced EXISTS predicate	
T551		Optional key words for default syntax	
T581		Regular expression substring function	
T591		UNIQUE constraints of possibly null columns	

D.2. 不支持的特性

PostgreSQL的本版本未实现下列SQL:2008中定义的特性。在少数情况下，有等价的功能可用。

标识符	包	描述	注释
B011		Embedded Ada	
B013		Embedded COBOL	
B014		Embedded Fortran	
B015		Embedded MUMPS	
B016		Embedded Pascal	
B017		Embedded PL/I	
B031		Basic dynamic SQL	
B032		Extended dynamic SQL	
B032-01		<describe input statement>	
B033		Untyped SQL-invoked function arguments	
B034		Dynamic specification of cursor attributes	
B041		Extensions to embedded SQL exception declarations	
B051		Enhanced execution rights	
B111		Module language Ada	
B112		Module language C	
B113		Module language COBOL	
B114		Module language Fortran	
B115		Module language MUMPS	
B116		Module language Pascal	
B117		Module language PL/I	
B121		Routine language Ada	
B122		Routine language C	
B123		Routine language COBOL	
B124		Routine language Fortran	
B125		Routine language MUMPS	
B126		Routine language Pascal	
B127		Routine language PL/I	
B128		Routine language SQL	

标识符	包	描述	注释
C011	Core	Call-Level Interface	
E081	Core	Basic Privileges	
E081-05	Core	UPDATE privilege at the column level	
E081-07	Core	REFERENCES privilege at the column level	
E081-09	Core	USAGE privilege	
E121	Core	Basic cursor support	
E121-06	Core	Positioned UPDATE statement	
E121-07	Core	Positioned DELETE statement	
E153	Core	Updatable queries with subqueries	
E182	Core	Module language	
F121		Basic diagnostics management	
F121-01		GET DIAGNOSTICS statement	
F121-02		SET TRANSACTION statement: DIAGNO STICS SIZE clause	
F181	Core	Multiple module support	
F262		Extended CASE expres sion	
F263		Comma-separated predicates in simple CASE expression	
F291		UNIQUE predicate	
F301		CORRESPONDING in query expressions	
F311	Core	Schema definition statement	
F311-04	Core	CREATE VIEW: WITH CHECK OPTION	
F312		MERGE statement	
F341		Usage tables	
F392		Unicode escapes in identifiers	
F393		Unicode escapes in literals	
F402		Named column joins for LOBs, arrays, and multisets	
F442		Mixed column refere nces in set functions	

标识符	包	描述	注释
F451		Character set definition	
F461		Named character sets	
F521	Enhanced integrity management	Assertions	
F641		Row and table constr uctors	
F661		Simple tables	
F671	Enhanced integrity management	Subqueries in CHECK	intentionally omitted
F691		Collation and transl ation	
F692		Enhanced collation support	
F693		SQL-session and client module collations	
F695		Translation support	
F696		Additional translation documentation	
F721		Deferrable constraints	foreign keys only
F731		INSERT column privil eges	
F741		Referential MATCH types	no partial match yet
F751		View CHECK enhanc ements	
F811		Extended flagging	
F812	Core	Basic flagging	
F813		Extended flagging	
F821		Local table references	
F831		Full cursor update	
F831-01		Updatable scrollable cursors	
F831-02		Updatable ordered cursors	
S011	Core	Distinct data types	
S011-01	Core	USER_DEFINED_TYPES view	
S023	Basic object support	Basic structured types	
S024	Enhanced object support	Enhanced structured types	
S025		Final structured types	
S026		Self-referencing struct ured types	
S027		Create method by specific method name	

标识符	包	描述	注释
S028		Permutable UDT options list	
S041	Basic object support	Basic reference types	
S043	Enhanced object support	Enhanced reference types	
S051	Basic object support	Create table of type	
S081	Enhanced object support	Subtables	
S091		Basic array support	
S091-01		Arrays of built-in data types	
S091-02		Arrays of distinct types	
S091-03		Array expressions	
S092		Arrays of user-defined types	
S094		Arrays of reference types	
S095		Array constructors by query	
S096		Optional array bounds	
S097		Array element assignment	
S151	Basic object support	Type predicate	
S161	Enhanced object support	Subtype treatment	
S162		Subtype treatment for references	
S201		SQL-invoked routines on arrays	
S201-01		Array parameters	
S201-02		Array as result type of functions	
S202		SQL-invoked routines on multisets	
S231	Enhanced object support	Structured type locators	
S232		Array locators	
S233		Multiset locators	
S241		Transform functions	
S242		Alter transform statement	
S251		User-defined orderings	
S261		Specific type method	
S271		Basic multiset support	
S272		Multisets of user-defined types	

标识符	包	描述	注释
S274		Multisets of reference types	
S275		Advanced multiset support	
S281		Nested collection types	
S291		Unique constraint on entire row	
T011		Timestamp in Inform ation Schema	
T041	Basic object support	Basic LOB data type support	
T041-01	Basic object support	BLOB data type	
T041-02	Basic object support	CLOB data type	
T041-03	Basic object support	POSITION, LENGTH, LOWER, TRIM, UPPER, and SUBSTRING functi ons for LOB data types	
T041-04	Basic object support	Concatenation of LOB data types	
T041-05	Basic object support	LOB locator: non-holdable	
T042		Extended LOB data type support	
T051		Row types	
T052		MAX and MIN for row types	
T053		Explicit aliases for all-fields reference	
T061		UCS support	
T111		Updatable joins, unions, and columns	
T121		WITH (excluding RECURSIVE) in query expression	
T122		WITH (excluding RECURSIVE) in subque ry	
T131		Recursive query	
T132		Recursive query in subquery	
T152		DISTINCT predicate with negation	
T172		AS subquery clause in table definition	
T173		Extended LIKE clause in table definition	
T174		Identity columns	

标识符	包	描述	注释
T175		Generated columns	
T176		Sequence generator support	
T211	Active database, Enhanced integrity management	Basic trigger capability	
T211-05	Active database, Enhanced integrity management	Ability to specify a search condition that must be true before the trigger is invoked	
T211-06	Active database, Enhanced integrity management	Support for run-time rules for the interaction of triggers and constr aints	
T211-08	Active database, Enhanced integrity management	Multiple triggers for the same event are executed in the order in which they were created in the catalog	intentionally omitted
T251		SET TRANSACTION statement: LOCAL option	
T261		Chained transactions	
T272		Enhanced savepoint management	
T281		SELECT privilege with column granularity	
T301		Functional depend encies	
T321	Core	Basic SQL-invoked routines	
T321-02	Core	User-defined stored procedures with no overloading	
T321-04	Core	CALL statement	
T321-05	Core	RETURN statement	
T324		Explicit security for SQL routines	
T325		Qualified SQL parame ter references	
T326		Table functions	
T331		Basic roles	
T332		Extended roles	
T401		INSERT into a cursor	
T411		UPDATE statement: SET ROW option	
T431	OLAP	Extended grouping capabilities	

标识符	包	描述	注释
T432		Nested and concat enated GROUPING SETS	
T433		Multiargument GROUPI NG function	
T434		GROUP BY DISINCT	
T471		Result sets return value	
T491		LATERAL derived table	
T511		Transaction counts	
T541		Updatable table refere nces	
T561		Holdable locators	
T571		Array-returning extern al SQL-invoked functi ons	
T572		Multiset-returning external SQL-invoked functions	
T601		Local cursor references	
T611	OLAP	Elementary OLAP operations	
T612		Advanced OLAP operat ions	
T613		Sampling	
T621		Enhanced numeric functions	
T631	Core	IN predicate with one list element	
T641		Multiple column assign ment	
T651		SQL-schema statem ents in SQL routines	
T652		SQL-dynamic statem ents in SQL routines	
T653		SQL-schema statem ents in external routin es	
T654		SQL-dynamic statem ents in external routin es	
T655		Cyclically dependent routines	

附录 E. 发布说明

发布说明包含每个PostgreSQL发布中的重要变更，并将主要特性和迁移问题列在最前面。
发布说明不包含只影响少数用户的变更，也不包含仅限内部实现、因而对用户不可见的变更。
例如，优化器几乎在每次发布中都会得到改进，
但用户通常只会把这些改进感知为查询速度更快。

每个版本的一份完整变更列表可以通过查看该版本的 Git 日志获得。pgsql-committers 邮件列表[1]也记录了所有源代码变更。另有一个Web 界面[2]可显示特定文件的变更。

每个条目旁边标出的姓名表示该条目的主要开发者。当然，
所有变更都经历了社区讨论和补丁审查，因此每个条目实际上都是社区共同努力的成果。

E.1. 版本 8.1.23

发行日期
2010-12-16

本次发布包含对 8.1.22 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

这预计是 PostgreSQL 8.1.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.1.1. 迁移到版本 8.1.23

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅 8.1.18 的发布说明。

E.1.2. 变更

- 强制 Linux 上的默认 `wal_sync_method` 为 `fdatasync`（Tom Lane、Marti Raudsepp）

  Linux 上的默认值实际上多年来一直是 `fdatasync`，但最近的内核变更导致 PostgreSQL 改选 `open_datasync`。这一选择没有带来任何性能改进，反而在某些文件系统（特别是使用 `data=journal` 挂载选项的 `ext4`）上导致彻底失败。

- 修复当起始检查点 WAL 记录与其重做点不在同一 WAL 段时从基础备份恢复的问题（Jeff Davis）

- 新增在 `IA64` 上检测寄存器栈溢出的支持（Tom Lane）

  `IA64` 架构有两个硬件栈。要完全防止栈溢出失败需要同时检查两者。

- 在 `copyObject()` 中添加栈溢出检查（Tom Lane）

  给定足够复杂的查询，某些代码路径可能因栈溢出而崩溃。

- 修复临时 GiST 索引中页面分裂的检测（Heikki Linnakangas）

[1] http://archives.postgresql.org/pgsql-committers/
[2] http://git.postgresql.org/gitweb?p=postgresql.git;a=summary

在临时索引中可能出现“并发”页面分裂，例如当插入时有打开的游标正在扫描该索引。GiST 未能检测到这种情况，因此当游标继续执行时可能给出错误结果。

- 避免在 `ANALYZE` 复杂索引表达式时内存泄漏（Tom Lane）
- 确保使用整行 Var 的索引仍依赖于其表（Tom Lane）

  像 `create index i on t (foo(t.*))` 这样声明的索引在删除其表时不会自动被删除。

- 不要“内联”带多个 `OUT` 参数的 SQL 函数（Tom Lane）

  这避免了因丢失预期结果行类型信息而可能出现的崩溃。

- 修复 `COALESCE()` 表达式的常量折叠（Tom Lane）

  规划器有时会尝试求值实际上永远不可达的子表达式，可能导致意外错误。

- 为 `InhRelation` 节点添加打印功能（Tom Lane）

  这避免了启用 `debug_print_parse` 且执行某些类型的查询时的失败。

- 修复点到水平线段距离的错误计算（Tom Lane）

  该缺陷影响多个不同的几何距离测量操作符。

- 修复 PL/pgSQL 对“简单”表达式的处理，使其在递归或错误恢复情况下不失败（Tom Lane）
- 修复 `contrib/cube` 的 GiST picksplit 算法中的缺陷（Alexander Korotkov）

  这可能导致 `cube` 列上的 GiST 索引相当低效，尽管答案并非真正错误。如果你有这样的索引，考虑在安装此更新后对其执行 `REINDEX`。

- `contrib/dblink` 除了在创建新连接时外不再发出“identifier will be truncated”通知（Itagaki Takahiro）
- 修复 `contrib/pgcrypto` 在缺少公钥时可能的核心转储（Marti Raudsepp）
- 修复 `contrib/xml2` 的 XPath 查询函数中的内存泄漏（Tom Lane）
- 把时区数据文件更新为 tzdata 版本 2010o for DST law changes in Fiji and Samoa; also historical corrections for Hong Kong。

E.2. 版本 8.1.22

发行日期
2010-10-04

本次发布包含对 8.1.21 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

PostgreSQL 社区将于 2010 年 11 月停止对 8.1.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.2.1. 迁移到版本 8.1.22

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅 8.1.18 的发布说明。

E.2.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

  感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）

- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。

- 防止 show_session_authorization() 在自动清理进程中崩溃（Tom Lane）

- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）

- 修复对按引用传递的函数结果计算哈希时可能的失败（Tao Ma、Tom Lane）

- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）

- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）

- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）

- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）

- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）

- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

E.3. 版本 8.1.21

发行日期
2010-05-17

本次发布包含对 8.1.20 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.3.1. 迁移到版本 8.1.21

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅 8.1.18 的发布说明。

E.3.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）
- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）
- 防止 psql 在展开引用自身的变量时无限递归（Tom）
- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）
- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。
- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

E.4. 版本 8.1.20

发行日期
2010-03-15

本次发布包含对 8.1.19 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.4.1. 迁移到版本 8.1.20

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅 8.1.18 的发布说明。

E.4.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。
- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）
- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）
- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。
- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）
- 修复若干病态缓慢的正则表达式匹配情形（Tom）
- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）

- 修复更多临时文件泄漏的情形（Heikki）

  这纠正了上一个小版本中引入的问题。失败的一种情形是返回集合的 plpgsql 函数在另一个函数的异常处理器中被调用。

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）

- 修复 psql 的 `numericlocale` 选项，使其不在 latex 和 troff 输出格式中格式化不应格式化的字符串（Heikki）

- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）

- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）

- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。

- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）

- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）

- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.5. 版本 8.1.19

发行日期
2009-12-14

本次发布包含对 8.1.18 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.5.1. 迁移到版本 8.1.19

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.18 的版本升级，请参阅 8.1.18 的发布说明。

E.5.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。

- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。

- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。
- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 确保预备事务所持有的共享元组级锁不被忽略（Heikki）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。
- 修复 `CREATE OR REPLACE FUNCTION` 期间所有权依赖的处理（Tom）
- 确保 Perl 数组在由返回集合的 PL/Perl 函数返回时被正确转换为 PostgreSQL 数组（Andrew Dunstan、Abhijit Menon-Sen）

  对非返回集合的函数这早已正确工作。
- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。
- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.6. 版本 8.1.18

发行日期
2009-09-09

本次发布包含对 8.1.17 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.6.1. 迁移到版本 8.1.18

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.1.18 后对它们执行 `REINDEX`。另外，如果你是从早于 8.1.15 的版本升级，请参阅 8.1.15 的发布说明。

E.6.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

  它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）

- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）

- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.7. 版本 8.1.17

发行日期
2009-03-16

本次发布包含对 8.1.16 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.7.1. 迁移到版本 8.1.17

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.15 的版本升级，请参阅 8.1.15 的发布说明。

E.7.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，
  解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）

- 修复带隐式强制转换的 `CASE WHEN` 的反编译（Tom）

  在试图检查或转储视图时，这一错误在启用断言的构建中可能导致断言失败，
  在其他情况下可能导致“意外的 CASE WHEN 子句”错误消息。

- 修复 TOAST 表行类型所有者可能的错误分配（Tom）

  如果 `CLUSTER` 或重写形式的 `ALTER TABLE` 由表所有者之外的人执行，表的 TOAST 表的 `pg_type` 条目最终会被标记为那个人所拥有。这不会立即引起问题，因为普通数据库操作不检查 TOAST 行类型的权限。但是，如果后来试图删除发出该命令的角色（在 8.1 或 8.2 中），或者在那之后 pg_dump 发出“数据类型的所有者似乎无效”警告（在 8.3 中），则可能导致意外失败。

- 在块退出时完全清理 PL/pgSQL 错误状态变量（Ashesh Vashi 和 Dave Page）

  这对 PL/pgSQL 本身不是问题，但这一疏漏可能导致 PL/pgSQL 调试器在检查函数状态时崩溃。

- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.8. 版本 8.1.16

发行日期
2009-02-02

本次发布包含对 8.1.15 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.8.1. 迁移到版本 8.1.16

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.15 的版本升级，请参阅 8.1.15 的发布说明。

E.8.2. 变更

- 修复 autovacuum 中的崩溃（Alvaro）

  该崩溃只在为防止事务回卷而清理整个数据库之后发生，这意味着它很少发生且难以追查。

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 确保可保持游标的内容不依赖于 TOAST 表的内容（Tom）

  此前，游标结果中的大字段值可能被表示为 TOAST 指针，如果被引用的表在游标被读取之前被删除，或者大值被删除随后被清理掉，这就会失败。普通游标不会发生这种情况，但对于保持到创建事务之后的游标则可能。

- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 修复 configure 脚本，使其在无法获得 PL/Perl 的链接信息时正确报告失败（Andrew）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.9. 版本 8.1.15

发行日期
2008-11-03

本次发布包含对 8.1.14 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.9.1. 迁移到版本 8.1.15

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。另外，如果你运行的是之前的 8.1.X 版本，建议在升级后对所有 GiST 索引执行 `REINDEX`。

E.9.2. 变更

- 修复 GiST 索引因删除后把错误的索引条目标记为“死”而损坏的问题（Teodor）

  这会导致索引搜索找不到本应找到的行。损坏的索引可以用 `REINDEX` 修复。

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）
- 修复当子 `SELECT` 出现在 `FROM` 中的函数调用、多行 `VALUES` 列表或 `RETURNING` 列表中时规则查询的错误展开（Tom）

  此问题通常的症状是“无法识别的节点类型”错误。
- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）
- 防止用 `ALTER SET TABLESPACE` 把表移到另一个表空间时 `relfilenode` 号可能的冲突（Heikki）

  该命令试图重用现有文件名，而不是选取目标目录中已知未使用的文件名。
- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。
- 修复 ecpg 对 `CREATE ROLE` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.10. 版本 8.1.14

发行日期
2008-09-22

本次发布包含对 8.1.13 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.10.1. 迁移到版本 8.1.14

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.10.2. 变更

- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。

- 修复 GiST 索引扫描期间可能的元组重复输出（Teodor）
- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  `ALTER COLUMN TYPE` 之后重用先前缓存的计划可能产生这种情形。该检查防范随之而来的数据损坏和/或崩溃。

- 修复 `AT TIME ZONE`：先尝试把时区参数解释为时区缩写，仅在失败时才尝试完整时区名，而不是像以前那样相反（Tom）

  时间戳输入函数一直按此顺序解析有歧义的时区名。让 `AT TIME ZONE` 也这样做改进了一致性，并修复了 8.1 中引入的兼容性缺陷：在有歧义的情形下，我们现在的行为与 8.0 及更早版本相同，因为在旧版本中 `AT TIME ZONE` 只接受缩写。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复嵌套子查询表达式的规划器缺陷（Tom）

  如果外层子查询不直接依赖父查询而内层依赖，外层的值可能不会为新的父查询行重新计算。

- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/pgSQL 在 `FOR` 循环的目标变量是包含组合类型字段的记录时不失败（Tom）
- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）
- 修复 PL/Python 使其与 Python 2.5 一起工作

  这是 8.2 开发周期中所做修复的向后移植。

- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）
- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）
- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.11. 版本 8.1.13

发行日期
2008-06-12

此发行版在 8.1.12 的基础上包含一个严重和一个次要的修复。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.11.1. 迁移到版本 8.1.13

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.11.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

- 让 `ALTER AGGREGATE ... OWNER TO` 更新 `pg_shdepend`（Tom）

  如果聚合后来涉及 `DROP OWNED` 或 `REASSIGN OWNED` 操作，这一疏漏可能导致问题。

E.12. 版本 8.1.12

发行日期
从未发布

本次发布包含对 8.1.11 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.12.1. 迁移到版本 8.1.12

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.12.2. 变更

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。

- 修复正则表达式子串匹配（`substring(string from pattern)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。

- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴、阿根廷/圣路易斯和智利的 DST 法规变更）

- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）

- 修复 `contrib/xml2` 的 `xpath_table()` 函数在输入查询返回 NULL 值时的核心转储（Tom）

- 修复 `contrib/xml2` 的 makefile，使其不覆盖 `CFLAGS`（Tom）

- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 不允许在预备事务内使用 `LISTEN` 和 `UNLISTEN`（Tom）

  这以前是允许的，但尝试这样做会有各种不愉快的后果，特别是只要 `UNLISTEN` 未提交，发起的后端就无法退出。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）

- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。

- 修复 `ALTER OWNER` 某些变体中的“无法识别的节点类型”错误（Tom）

- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 `pg_ctl start -w` 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。

- 使用 `-fwrapv` 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。

- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.13. 版本 8.1.11

发行日期
2008-01-07

此发行版包含 8.1.10 之后的多种修复，包括对重要安全问题的修复。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

这是 PostgreSQL 社区为 Windows 制作二进制包的最后一个 8.1.X 版本。建议 Windows 用户迁移到 8.2.X 或更高版本，因为 8.2.X 中有一些不便于向后移植的 Windows 特定修复。8.1.X 在其他平台上将继续受支持。

E.13.1. 迁移到版本 8.1.11

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.13.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  8.1.10 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 更新时区数据文件到 tzdata 发行版 2007k（特别是阿根廷最近的变更）（Tom）
- 改进规划器在非 C 区域设置下对 LIKE/正则表达式估计的处理（Tom）
- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）
- 保留被 `ALTER TABLE ... ALTER COLUMN TYPE` 重建的索引的表空间（Tom）
- 让归档恢复总是开始新的 WAL 时间线，而不是只在使用恢复停止时间时（Simon）

  这避免了试图覆盖最后一个 WAL 段现有归档副本的极端情形风险，而且看起来比原来的定义更简单干净。
- 当表太小而不适用时，让 `VACUUM` 不使用全部 `maintenance_work_mem`（Alvaro）
- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）
- 修复超过 68 年的 interval 在 `extract(epoch from interval)` 中的溢出（Tom）
- 修复 PL/Perl 在可信函数中使用 UTF-8 正则表达式时不失败（Andrew）
- 修复 PL/Perl 在平台 Perl 把 `bool` 类型定义为 `int` 而非 `char` 时的问题（Tom）

  虽然理论上这可能在任何地方发生，但没有标准 Perl 构建这样做.....直到 Mac OS X 10.5。
- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）
- 修复 pg_dump，使其正确处理默认表达式与父表不同的继承子表（Tom）
- 修复 `PGPASSFILE` 指向非普通文件时 libpq 的崩溃（Martin Pitt）
- ecpg 解析器修复（Michael）
- 让 `contrib/pgcrypto` 防御在超过 128 位密钥上失败的 OpenSSL 库——至少某些 Solaris 版本是这种情形（Marko Kreen）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）
- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.14. 版本 8.1.10

发行日期
2007-09-17

本次发布包含对 8.1.9 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.14.1. 迁移到版本 8.1.10

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.14.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 允许 `interval` 数据类型接受仅由毫秒或微秒组成的输入（Neil）
- 加快 rtree 索引插入（Teodor）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复日志，使使用 syslogger 进程时日志消息永不交错（Andrew）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 修复某些外键极端情形的错误处理（Tom）
- 防止 `REINDEX` 和 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 更新时区数据库规则，特别是新西兰即将到来的变更（Tom）
- Windows 套接字改进（Magnus）
- 因可能的编码不匹配，在 Windows 上抑制日志时间戳中的时区名（`%Z`）（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.15. 版本 8.1.9

发行日期
2007-04-23

此发行版包含 8.1.8 之后的多种修复，包括一个安全修复。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.15.1. 迁移到版本 8.1.9

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.15.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）
- 要求 `COMMIT PREPARED` 在事务预备时所在的同一数据库中执行（Heikki）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 规划器修复，包括改进外连接和位图扫描选择逻辑（Tom）
- 修复哈希索引扩容期间的 PANIC（8.1.6 中引入的缺陷）（Tom）
- 修复 POSIX 风格时区规范以遵循新的美国 DST 规则（Tom）

E.16. 版本 8.1.8

发行日期
2007-02-07

此发行版包含 8.1.7 之后的一个修复。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.16.1. 迁移到版本 8.1.8

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.16.2. 变更

- 移除约束和函数索引中对类型长度的过度严格检查（Tom）

E.17. 版本 8.1.7

发行日期
2007-02-05

此发行版包含 8.1.6 之后的多种修复，包括一个安全修复。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.17.1. 迁移到版本 8.1.7

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.17.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  这些漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，以及更改表列的数据类型（CVE-2007-0555、CVE-2007-0556）。这些错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 改进有大量表的数据库的 `VACUUM` 性能（Tom）

- 修复 autovacuum，避免在不可连接的数据库中留下非持久事务 ID（Alvaro）

  此缺陷只影响 8.1 分支。

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

- 修复 Windows 上因试图 fsync 已删除文件而发生的虚假“权限被拒绝”失败（Magnus、Tom）

- 修复已在使用中的 PL/pgSQL 函数被更新时可能的崩溃（Tom）

E.18. 版本 8.1.6

发行日期
2007-01-08

本次发布包含对 8.1.5 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.18.1. 迁移到版本 8.1.6

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.18.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 pg_restore 处理包含带注释的大对象（blob）的 tar 格式备份（Tom）

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）

- 在服务器重启期间清除 `pg_internal.init` 缓存文件（Simon）

这避免了缓存文件在 PITR 恢复后可能包含陈旧数据的风险。

- 修复 `VACUUM` 截断跨越 GB 边界的大关系时的竞态条件（Tom）
- 修复导致行级锁上不必要死锁错误的缺陷（Tom）
- 修复影响多 GB 哈希索引的缺陷（Tom）
- 修复 Windows 信号处理中可能的死锁（Teodor）
- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）
- 修复 ecpg 连接期间的内存泄漏（Michael）
- 修复 Darwin（OS X）编译（Tom）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。

- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。

- 更新时区数据库

  这特别影响澳大利亚和加拿大的夏令时规则。

E.19. 版本 8.1.5

发行日期
2006-10-16

本次发布包含对 8.1.4 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.19.1. 迁移到版本 8.1.5

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.19.2. 变更

- 不允许在 `UPDATE` 命令中使用聚合函数，子 SELECT 中除外（Tom）

  这种聚合的行为不可预测，而且在 8.1.X 中可能导致崩溃，因此已被禁用。SQL 标准也不允许这样做。

- 修复把无类型字面值当作 ANYARRAY 时的核心转储
- 修复扩展查询协议中执行 `COMMIT` 或 `ROLLBACK` 时持续时间日志记录的核心转储

- 修复查询包含返回多行的 SQL 函数时 AFTER 触发器的错误处理（Tom）
- 修复 `ALTER TABLE ... TYPE`，对 `USING` 子句重新检查 `NOT NULL`（Tom）
- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。
- 修复 `AM/PM` 格式的 `to_timestamp()`（Bruce）
- 修复 autovacuum 决定是否需要 `ANALYZE` 的计算（Alvaro）
- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- ecpg 的大量健壮性修复（Joachim Wieland）
- 修复 /contrib/dbmirror 中的反斜杠转义
- /contrib/dblink 和 /contrib/tsearch2 的小修复
- 哈希表和位图索引扫描的效率改进（Tom）
- 修复 Windows 上统计收集的不稳定（Tom、Andrew）
- 修复 `statement_timeout` 在 Win32 上使用正确的单位（Bruce）

  在以前的 Win32 8.1.X 版本中，延迟偏差了 100 倍。
- MSVC 和 Borland C++ 编译器的修复（Hiroshi Saito）
- AIX 和 Intel 编译器的修复（Tom）
- 修复连续归档中的罕见缺陷（Tom）

E.20. 版本 8.1.4

发行日期
2006-05-23

此发行版包含 8.1.3 之后的多种修复，包括针对极其严重安全问题的补丁。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.20.1. 迁移到版本 8.1.4

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐

的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescape StringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.20.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复 pgcrypto 中的弱密钥选择（Marko Kreen）

  fortuna PRNG 重播种逻辑中的错误可能在某些情况下导致 `pgp_sym_encrypt()` 选择可预测的会话密钥。这只影响不使用 OpenSSL 的构建。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`win866_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 让 autovacuum 在 `pg_stat_activity` 中可见（Alvaro）

- 禁用 `full_page_writes`（Tom）

  某些情况下，关闭 `full_page_writes` 会导致崩溃恢复失败。适当的修复将出现在 8.2 中；目前只是禁用它。

- 各种规划器修复，特别是位图索引扫描和 MIN/MAX 优化（Tom）

- 修复归并连接中的错误优化（Tom）

  外连接有时可能发出未匹配行的多个副本。

- 修复在同一事务中使用并修改 plpgsql 函数导致的崩溃

- 修复 B-Tree 索引被截断情形的 WAL 重放

- 修复涉及 `|` 的模式的 `SIMILAR TO`（Tom）

- 修复 `SELECT INTO` 和 `CREATE TABLE AS`，在默认表空间而不是基目录中创建表（Kris Jurka）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 改进 qsort 性能（Dann Corbit）

  目前此代码只在 Solaris 上使用。
- 修复 x86 系统上的 OS/X Bonjour（Ashley Clark）
- 修复各种小的内存泄漏
- 修复某些 Win32 系统上的口令提示问题（Robert Kinberg）
- 改进 pg_dump 对域默认值的处理
- 修复 pg_dumpall 合理处理同名的用户和组（只在从 8.1 之前的服务器转储时可能）（Tom）

  用户和组将被合并为带 `LOGIN` 权限的单个角色。此前合并后的角色没有 `LOGIN` 权限，使其不能作为用户使用。
- 修复 pg_restore 的 `-n`，使其按文档工作（Tom）

E.21. 版本 8.1.3

发行日期
2006-02-14

此发行版包含 8.1.2 之后的多种修复，包括一个非常严重的安全问题。有关 8.1 主版本新特性的信息，请参阅第 E.24 节。

E.21.1. 迁移到版本 8.1.3

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.1.2 的版本升级，请参阅 8.1.2 的发布说明。

E.21.2. 变更

- 修复允许任何已登录用户把 `SET ROLE` 设为任何其他数据库用户 id 的缺陷（CVE-2006-0553）

  由于有效性检查不充分，用户可以利用 `SET ROLE` 通常在错误后用来恢复先前角色设置的特殊情形。这例如允许普通用户获得超级用户状态。权限提升风险只存在于 8.1.0-8.1.2。但在直到 7.3 的所有版本中，`SET SESSION AUTHORIZATION` 都有一个相关缺陷，如果服务器编译时启用了断言（这不是默认），非特权用户可使服务器崩溃。感谢 Akio Ishida 报告此问题。
- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 8.0.4、7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 和 pg_subtrans 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复缓存失效消息恰好在错误时刻到达时可能导致崩溃的情形（Tom）
- 为预备语句中的 `UNKNOWN` 参数正确检查 `DOMAIN` 约束（Neil）
- 确保 `ALTER COLUMN TYPE` 按正确顺序处理 `FOREIGN KEY`、`UNIQUE` 和 `PRIMARY KEY` 约束（Nakano Yoshihisa）
- 修复以允许恢复有对自定义操作符或操作符类的跨模式引用的转储（Tom）
- 允许 pg_restore 在 `COPY` 失败后正确继续；此前它试图把剩余的 `COPY` 数据当作 SQL 命令（Stephen Frost）
- 修复未指定数据目录时 pg_ctl 的 `unregister` 崩溃（Magnus）
- 修复 libpq 的 `PQprint` HTML 标签（Christoph Zwerschke）
- 修复 AMD64 和 PPC 上 ecpg 的崩溃（Neil）
- 允许在函数结果类型声明中同时使用 `SETOF` 和 `%TYPE`
- 修复 PL/python 参数传递期间发生错误时的正确恢复（Neil）
- 修复 `plperl_return_next` 的内存泄漏（Neil）
- 修复 PL/perl 在 Win32 上对区域设置的处理以与后端匹配（Andrew）
- 各种优化器修复（Tom）
- 修复 Win32 上 `postgresql.conf` 中 `log_min_messages` 设为 `DEBUG3` 或更高时的崩溃（Bruce）
- 修复 Win32、Cygwin、OS X、AIX 的 pgxs `-L` 库路径指定（Bruce）
- 在检查 Win32 管理员权限时检查 SID 是否启用（Magnus）
- 正确拒绝超出范围的日期输入（Kris Jurka）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）
- 通过避免每数据行一次内核调用，改进经 libpq 的 `COPY IN` 速度（Alon Goldshuv）
- 改进 `/contrib/tsearch2` 索引创建的速度（Tom）

E.22. 版本 8.1.2

发行日期
2006-01-09

本次发布包含对 8.1.1 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.22.1. 迁移到版本 8.1.2

对于运行 8.1.X 的用户，不需要进行转储/恢复。不过，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.22.2. 变更

- 修复 Windows 代码，使 ShmemBackendArray 没有更多空间时 postmaster 继续而不是退出（Magnus）

  如果过多连接请求几乎同时到达，以前的行为可能导致拒绝服务情形。这只适用于 Windows 移植。

- 修复 8.0 中引入的缺陷，它可能使 ReadBuffer 返回已使用过的页当作新页，可能导致最近提交的数据丢失（Tom）

- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）

- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 允许更灵活地重新定位安装目录（Tom）

  以前的版本只有在所有安装目录路径除最后一个组件外都相同时才支持重新定位。

- 防止因使用 `ISO-8859-5` 和 `ISO-8859-9` 编码导致的崩溃（Tatsuo）

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）

- 修复 COPY CSV 模式把任何 `\.` 视为复制数据结束的缺陷

  新代码要求 `\.` 单独出现在一行上，如文档所述。

- 让 COPY CSV 模式给字面数据值 `\.` 加引号，以确保它不会被解释为数据结束标记（Bruce）

- 对返回 `RECORD` 的函数的各种修复（Tom）

- 修复 `postgresql.conf` 的处理，使没有换行符的最后一行被正确处理（Tom）

- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复处理表达式索引时 autovacuum 的崩溃

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.23. 版本 8.1.1

发行日期
2005-12-12

本次发布包含对 8.1.0 的多方面修复。有关 8.1 大版本中新特性的信息，请参见第 E.24 节。.

E.23.1. 迁移到版本 8.1.1

对于运行 8.1.X 的用户，不需要进行转储/恢复。

E.23.2. 变更

- 修复外连接条件的错误优化（Tom）
- 修复涉及被优化器平展的子查询的情形下报告列名错误的问题（Tom）
- 修复涉及 CHECK 约束、TOAST 列和索引的场景中的更新失败（Tom）
- 修复从错误恢复后 bgwriter 的问题（Tom）

  发现后台写入器在写错误后泄漏缓冲区固定。虽然本身不致命，但这可能导致后续 VACUUM 命令的神秘阻塞。
- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败
- `/contrib/tsearch2` 和 `/contrib/ltree` 修复（Teodor）
- 修复需要单词重排的语言（如土耳其语）中翻译错误消息的问题；还有输出字符串意外截断和 bigint 最小可能值显示错误的问题（Andrew、Tom）

  这些问题只出现在使用我们的 `port/snprintf.c` 代码的平台上，包括指定了 `--enable-nls` 的 BSD 变体，可能还有其他平台。此外，根据所使用的 `libintl` 版本，翻译错误消息问题的另一种形式可能出现在 Windows 上。
- 重新允许 `to_char(time)` 和 `to_char(interval)` 使用 `AM`/`PM`、`HH`、`HH12` 和 `D` 格式说明符。（`to_char(interval)` 或许应该用 `HH24`。）（Bruce）
- AIX、HPUX 和 MSVC 编译修复（Tom、Hiroshi Saito）
- 优化器改进（Tom）
- 在 Windows NO_SYSTEM_RESOURCES 错误后重试文件读写（Qingqing Zhou）
- 防止 autovacuum 在分析表达式索引期间崩溃（Alvaro）
- 修复 ON COMMIT DELETE ROWS 临时表的问题
- 修复触发器更改 SELECT DISTINCT 查询输出时的问题
- 新增关于如何迁移无效 `UTF-8` 字节序列的 8.1.0 发行说明条目（Paul Lindner）

E.24. 版本 8.1

发行日期
2005-11-08

E.24.1. 概述

此发行版的主要变更：

改进对共享缓冲区缓存的并发访问（Tom）

对共享缓冲区缓存的访问被认定为显著的可伸缩性问题，特别是在多 CPU 系统上。在此版本中，缓冲区管理器中加锁的方式被彻底改造，以减少锁争用并改进可伸缩性。
缓冲区管理器还被改为使用“时钟扫描”替换策略。

允许索引扫描使用中间内存位图（Tom）

在以前的版本中，对表做查找只能使用单个索引。有了此特性，如果查询有 `WHERE tab.col1 = 4 and tab.col2 = 9`，而 `col1` 和 `col2` 上没有多列索引，但 `col1` 上有一个索引、`col2` 上有另一个索引，则可以搜索两个索引并在内存中合并结果，然后只对同时匹配 `col1` 和 `col2` 限制的行做堆获取。这在有大量非结构化查询、不可能创建匹配所有可能访问条件的索引的环境中非常有用。位图扫描即使对单个索引也有用，因为它减少了所需的随机访问量；
位图索引扫描对于检索整表中相当大的部分是高效的，而普通索引扫描则不是。

Add two-phase commit (Heikki Linnakangas, Alvaro, Tom)

两阶段提交允许在多台计算机上"预备"事务，
一旦所有计算机都成功预备了各自的事务（无一失败），所有事务就可以提交。
即使机器在预备后崩溃，预备的事务也可以在机器重启后提交。新语法包括 `PREPARE TRANSACTION` 和 `COMMIT/ROLLBACK PREPARED`。还新增了系统视图 `pg_prepared_xacts`。

创建取代用户和组的新角色系统（Stephen Frost）

角色是用户和组的组合。像用户一样，它们可以有登录能力；像组一样，
角色可以拥有其他角色作为成员。角色基本上消除了用户和组的区别。例如，角色可以：

- （可选地）拥有登录能力
- 拥有对象
- 持有数据库对象的访问权限
- 从它是成员的其他角色继承权限

用户登录到角色后，就获得该登录角色加上任何继承角色的能力，并可以用 `SET ROLE` 切换到她是成员的其他角色。此特性是 SQL 标准角色概念的推广。此变更还用支持角色的新目录 `pg_authid` 和 `pg_auth_members` 取代了 `pg_shadow` 和 `pg_group`。旧表被重新定义为新角色表上的只读视图。

Automatically use indexes for `MIN()` and `MAX()` (Tom)

在以前的版本中，让 `MIN()` 或 `MAX()` 使用索引的唯一方法是把查询改写为 `SELECT col FROM tab ORDER BY col LIMIT 1`。现在索引的使用自动发生。

Move `/contrib/pg_autovacuum` into the main server (Alvaro)

把 autovacuum 集成到服务器中使其可以随数据库服务器同步自动启动和停止，并允许从 `postgresql.conf` 配置 autovacuum。

使用 `SELECT ... FOR SHARE` 增加共享行级锁（Alvaro）

虽然 PostgreSQL 的 MVCC 锁定使 `SELECT` 从不被写入者阻塞，因此典型操作不需要共享行锁，但共享锁对需要共享行锁定的应用程序很有用。
特别是这降低了参照完整性检查施加的锁定要求。

增加对共享对象（特别是角色）的依赖（Alvaro）

依赖机制的这一扩展防止角色在其仍拥有数据库对象时被删除。
此前可能通过删除所有者而意外地使对象“孤立”。虽然这可以恢复，但既混乱又令人不快。

改进分区表的性能（Simon）

新的 `constraint_exclusion` 配置参数避免在约束表明子表中不存在匹配行时查找子表。

这支持一种基本的表分区。如果子表存储独立的键范围并用适当的 `CHECK` 约束强制执行，当约束保证子表中不存在匹配行时，优化器将跳过对子表的访问。

E.24.2. 迁移到版本 8.1

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

8.0 发行版宣布 interval 的 `to_char()` 函数将在 8.1 中移除。但由于没有提出更好的 API，`to_char(interval)` 在 8.1 中得到增强并将保留在服务器中。

请注意以下不兼容之处：

- `add_missing_from` 现在默认为 false（Neil）

  默认情况下，如果查询中使用的表没有 `FROM` 引用，我们现在会产生错误。旧行为仍可用，但必须把参数设为 'true' 才能获得。

  如果转储包含用隐式 `FROM` 语法创建的视图或规则，可能需要把 `add_missing_from` 设为 true 才能装载现有转储文件。这应当只是一次性的麻烦，因为 PostgreSQL 8.1 会把这样的视图和规则转换为标准的显式 `FROM` 语法。因此后续转储不会有此问题。

- 使 `float4`/`float8`/`oid` 输入零长度字符串（`''`）抛出错误，而不是把它当作零（Neil）

  此变更与当前整数的零长度字符串处理一致。此变更的时间表在 8.0 中已宣布。

- `default_with_oids` 现在默认为 false（Neil）

  此选项设为 false 后，用户创建的表不再有 OID 列，除非在 `CREATE TABLE` 中指定 `WITH OIDS`。虽然 OID 在 PostgreSQL 的所有版本中都存在，但它们的用途有限，因为它们只有四字节长，而且计数器在所有已安装数据库之间共享。唯一标识行的首选方式是通过序列和 `SERIAL` 类型，它们自 PostgreSQL 6.4 起就受支持。

- 新增 E'' 语法，使最终普通字符串能按字面处理反斜杠（Bruce）

  目前 PostgreSQL 把字符串字面量中的反斜杠当作引入特殊转义序列（如 \n 或 \010）处理。虽然这便于输入特殊值，但它非标准且使从其他数据库移植应用程序更困难。因此，PostgreSQL 项目计划移除字符串中反斜杠的特殊含义。为了向后兼容和满足想要特殊反斜杠处理的用户，创建了新的字符串语法。
  新字符串语法的写法是在启动字符串的单引号前紧挨着写 E，例如 E'hi\n'。虽然此版本不改变字符串中反斜杠的处理，但它确实新增了帮助用户为未来版本迁移应用程序的配置参数：

  - standard_conforming_strings——此版本是否按字面处理普通字符串中的反斜杠?

  - escape_string_warning——对普通（非 E）字符串中的反斜杠发出警告

  standard_conforming_strings 的值是只读的。应用程序可以检索该值以了解反斜杠如何被处理。（参数的存在也可被视为支持 E'' 字符串语法的标志。）在未来版本中，standard_conforming_strings 将为 true，意味着非 E 字符串中的反斜杠将按字面处理。为此变更做准备：在需要特殊反斜杠处理的地方使用 E'' 字符串，并打开 escape_string_warning 以找出其他需要转换为使用 E'' 的字符串。此外，用两个单引号（''）在字符串中嵌入字面单引号，而不要用 PostgreSQL 支持的反斜杠单引号（\'）语法。前者符合标准且不需要使用 E'' 字符串语法。也可以使用 $$ 字符串语法，它不特殊处理反斜杠。

- 让 REINDEX DATABASE 重新索引数据库中的所有索引（Tom）

  此前 REINDEX DATABASE 只重新索引系统表。新行为看起来更直观。新命令 REINDEX SYSTEM 提供只重新索引系统表的旧功能。

- 只读大对象描述符现在遵循 MVCC 快照语义

  当大对象以 INV_READ（而非 INV_WRITE）打开时，从描述符读取的数据现在将反映调用 lo_open() 的查询所用事务快照时刻大对象状态的“快照”。要获得总是返回最新已提交数据的旧行为，请在 lo_open() 的模式标志中包含 INV_WRITE。

- 为序列函数的参数添加适当的依赖（Tom）

  在以前的版本中，传递给 nextval()、currval() 和 setval() 的序列名存储为简单文本字符串，这意味着重命名或删除 DEFAULT 子句中使用的序列会使子句无效。此版本把所有新创建的序列函数参数存储为内部 OID，使它们能跟踪序列重命名，并添加防止不当删除序列的依赖信息。它还使这样的 DEFAULT 子句不受模式重命名和搜索路径更改的影响。

  有些应用程序可能依赖旧的运行时查找序列名的行为。仍然可以通过把参数显式转换为 text 来做到，例如 nextval('myseq'::text)。

  装载到 8.1 的 8.1 之前数据库转储将使用旧的基于文本的表示，因此不会有 OID 存储参数的特性。但可以更新包含基于文本的 DEFAULT 子句的数据库。首先，把下面这个查询保存到文件中，例如 fixseq.sql：

```
SELECT  'ALTER TABLE ' ||
   pg_catalog.quote_ident(n.nspname) || '.' ||
   pg_catalog.quote_ident(c.relname) ||
   ' ALTER COLUMN ' || pg_catalog.quote_ident(a.attname) ||
   ' SET DEFAULT ' ||
   regexp_replace(d.adsrc,
                  $$val\(\(('[^']*')::text\)::regclass$$,
                  $$val(\1$$,
                  'g') ||
```

```
   ';'
FROM    pg_namespace n, pg_class c, pg_attribute a, pg_attrdef d
WHERE   n.oid = c.relnamespace AND
   c.oid = a.attrelid AND
   a.attrelid = d.adrelid AND
   a.attnum = d.adnum AND
   d.adsrc ~ $$val\(\('[^']*'::text\)::regclass$$;
```

然后，对数据库运行该查询以找出需要哪些调整，例如对数据库 `db1`：

```
psql -t -f fixseq.sql db1
```

这会显示把数据库转换为较新的基于 OID 的表示所需的 `ALTER TABLE` 命令。如果这些命令看起来合理，运行以下命令更新数据库：

```
psql -t -f fixseq.sql db1 | psql -e db1
```

此过程必须在每个要更新的数据库中重复进行。

- 在 psql 中，把未加引号的 `\{digit}+` 序列当作八进制（Bruce）

  在以前的版本中，`\{digit}+` 序列被当作十进制，只有 `\0{digit}+` 被当作八进制。此变更是为了保持一致。

- 移除前缀和后缀 `%` 和 `^` 操作符的语法产生式（Tom）

  它们从未被记录，而且使取模操作符（`%`）与负数一起使用变得复杂。

- 让多边形的 `&<` 和 `&>` 与 box 的"over"操作符一致（Tom）

- `CREATE LANGUAGE` 可以忽略提供的参数而采用 `pg_pltemplate` 中的信息（Tom）

  定义了新系统目录 `pg_pltemplate`，携带过程语言首选定义的信息（例如是否有验证器函数）。当该目录中存在正要创建的语言的条目时，`CREATE LANGUAGE` 将忽略除语言名之外的所有参数而改用目录信息。采取此措施是因为旧转储文件装载过时语言定义的问题日益增多。从 8.1 起，pg_dump 将把过程语言定义只转储为 `CREATE LANGUAGE` *`name`*，依赖装载时存在模板条目。我们预期这将是更面向未来的表示。

- 让 `pg_cancel_backend(int)` 返回 `boolean` 而不是 `integer`（Neil）

- 一些用户在向 8.1.X 装载 UTF-8 数据时遇到问题。这是因为以前的版本允许把无效 UTF-8 字节序列输入数据库，而此版本只正确接受有效的 UTF-8 序列。纠正转储文件的一种方法是运行命令 `iconv -c -f UTF-8 -t UTF-8 -o cleanfile.sql dumpfile.sql`。`-c` 选项移除无效字符序列。两个文件的 diff 会显示无效的序列。`iconv` 把整个输入文件读入内存，因此可能需要用 split 把转储拆成多个较小的文件处理。

E.24.3. 附加变更

下面你将找到 PostgreSQL 8.1 与上一个主版本之间额外变更的详细说明。

E.24.3.1. 性能改进

- 改进 GiST 和 R-tree 索引性能（Neil）

- 改进优化器，包括哈希连接的自动调整大小（Tom）

- 彻底改造若干领域的内部 API
- 把 WAL 记录的 CRC 从 64 位改为 32 位（Tom）

  我们认定计算 64 位 CRC 的额外代价显著，而可靠性收益太小不值得。
- 防止在 WAL 页中写入大的空隙（Tom）
- 改进 SMP 机器（特别是 Opteron）上的自旋锁行为（Tom）
- 允许多列索引使用非连续的索引列（Tom）

  例如，这允许 a,b,c 列上的索引用于 `WHERE a = 4 and c = 10` 的查询。
- 跳过 `CREATE TABLE AS` / `SELECT INTO` 的 WAL 日志（Simon）

  由于 `CREATE TABLE AS` 期间的崩溃会导致表在恢复期间被删除，没有理由在装载表时写 WAL 日志。（但如果启用了 WAL 归档，仍会记录。）
- 允许并发 GiST 索引访问（Teodor、Oleg）
- 新增配置参数 `full_page_writes`，控制向 WAL 写整页（Bruce）

  为防止部分磁盘写损坏数据库，PostgreSQL 在检查点后首次修改每个数据库磁盘页时把它的完整副本写入 WAL。此选项为提高速度关闭该功能。在使用电池供电的磁盘缓存（不会发生部分页写）时可以安全使用。
- 当 `wal_sync_method` 使用 `O_SYNC` 时，如果可用则使用 `O_DIRECT`（Itagaki Takahiro）

  `O_DIRECT` 使磁盘写绕过内核缓存，对 WAL 写入而言这改进了性能。
- 改进 `COPY FROM` 的性能（Alon Goldshuv）

  这是通过以更大的块读取 `COPY` 输入而不是逐字符读取实现的。
- 改进 `COUNT()`、`SUM`、`AVG()`、`STDDEV()` 和 `VARIANCE()` 的性能（Neil、Tom）

E.24.3.2. 服务器变更

- 防止事务 ID（XID）回卷导致的问题（Tom）

  当事务计数器接近回卷点时，服务器现在会发出警告。如果计数器过于接近回卷，服务器将停止接受查询。这确保在执行所需清理之前数据不会丢失。
- 修复 OID 计数器回卷后对象 ID（OID）与现有系统对象冲突的问题（Tom）
- 新增关于 `VACUUM` 期间需要增大 `max_fsm_relations` 和 `max_fsm_pages` 的警告（Ron Mayer）
- 新增 `temp_buffers` 配置参数，允许用户确定临时表访问的本地缓冲区大小（Tom）
- 为 `pg_stat_activity` 新增会话开始时间和客户端 IP 地址（Magnus）
- 为位图扫描调整 `pg_stat` 视图（Tom）

  某些字段的含义略有变化。
- 增强 `pg_locks` 视图（Tom）

- 记录客户端 PREPARE 和 EXECUTE 的查询（Simon）
- 允许在 postgresql.conf 中指定 Kerberos 名和用户名的大小写敏感性（Magnus）
- 新增配置参数 krb_server_hostname，使服务器主机名可以被指定为服务主体的一部分（Todd Kover）

  如果未设置，可以使用与 keytab 中条目匹配的任何服务主体。这是此版本新的 Kerberos 匹配行为。
- 为 log_line_prefix 新增毫秒时间戳（%m）和远程主机（%h）选项（Ed L.）
- 为 GiST 索引新增 WAL 日志（Teodor、Oleg）

  GiST 索引现在对崩溃和时间点恢复是安全的。
- 执行 pg_stop_backup() 时移除旧的 *.backup 文件（Bruce）

  这防止 pg_xlog/ 中存在大量 *.backup 文件。
- 新增控制 TCP/IP keep-alive 空闲时间、间隔和次数的配置参数（Oliver Jowett）

  可以更改这些值以更快检测丢失的客户端连接。
- 新增逐用户和逐数据库的连接限制（Petr Jelinek）

  使用 ALTER USER 和 ALTER DATABASE，现在可以对以特定用户身份或到特定数据库并发连接的最大会话数施加限制。把限制设为零禁用用户或数据库连接。
- 允许 64 位机器上超过 2GB 的共享内存和逐后端工作内存（Koichi Suzuki）
- 新系统目录 pg_pltemplate 允许覆盖转储文件中过时的过程语言定义（Tom）

E.24.3.3. 查询变更

- 新增临时视图（Koju Iijima、Neil）
- 修复没有任何聚合函数或 GROUP BY 的 HAVING，使查询返回单个组（Tom）

  此前这种情形会把 HAVING 子句当作 WHERE 子句。这不符合规范。
- 新增 USING 子句，允许为 DELETE 指定额外的表（Euler Taveira de Oliveira、Neil）

  在以前的版本中，没有明确的方法为 DELETE 语句指定用于连接的额外表。UPDATE 已有用于此目的的 FROM 子句。
- 新增对后端和 ecpg 字符串中 \x 十六进制转义的支持（Bruce）

  这就像标准的 C \x 转义语法。八进制转义此前已支持。
- 新增 BETWEEN SYMMETRIC 查询语法（Pavel Stehule）

  此特性允许 BETWEEN 比较而不要求第一个值小于第二个值。例如 2 BETWEEN [ASYMMETRIC] 3 AND 1 返回假，而 2 BETWEEN SYMMETRIC 3 AND 1 返回真。BETWEEN ASYMMETRIC 此前已支持。
- 为 SELECT ... FOR UPDATE/SHARE 新增 NOWAIT 选项（Hans-Juergen Schoenig）

虽然 statement_timeout 配置参数允许取消耗时超过一定时间的查询，但 NOWAIT 选项允许在 SELECT ... FOR UPDATE/SHARE 命令不能立即获得行锁时立刻取消查询。

E.24.3.4. 对象操作变更

- 跟踪共享对象的依赖（Alvaro）

  PostgreSQL 允许全局表（用户、数据库、表空间）引用多个数据库中的信息。此新增为全局表添加了依赖信息，因此例如用户所有权可以跨数据库跟踪，
  在任何数据库中拥有某些东西的用户不能再被删除。数据库本地对象的依赖跟踪此前已存在。

- 允许对象所有者执行有限的 ALTER OWNER 命令（Stephen Frost）

  以前的版本只允许超级用户更改对象所有者。现在，
  如果执行命令的用户拥有该对象并能以新所有者身份创建它（即该用户是新所有角色的成员且该角色具有重新创建对象所需的 CREATE 权限），就可以转移所有权。

- 为某些对象类型（表、函数、类型）新增 ALTER 对象 SET SCHEMA 能力（Bernd Helmle）

  这允许把对象移到不同的模式。

- 新增 ALTER TABLE ENABLE/DISABLE TRIGGER 来禁用触发器（Satoshi Nagayasu）

E.24.3.5. 工具命令变更

- 允许 TRUNCATE 用单条命令截断多个表（Alvaro）

  由于参照完整性检查，不允许截断作为参照完整性约束一部分的表。使用此新功能，
  如果参照完整性约束涉及的两个表在同一条 TRUNCATE 命令中被截断，就可以用 TRUNCATE 截断这样的表。

- 在 COPY CSV 模式中正确处理回车和换行（Andrew）

  在 8.0 版本中，CSV COPY TO 中的回车和换行的处理方式不一致。（这记录在 TODO 列表中。）

- 新增 COPY WITH CSV HEADER，允许 COPY 的第一行为头行（Andrew）

  这允许处理把列名放在数据文件首行的常见 CSV 用法。对 COPY TO，首行包含列名；对 COPY FROM，首行被忽略。

- 在 Windows 上，EXPLAIN ANALYZE 显示更好的亚秒精度（Magnus）

- 为 EXPLAIN ANALYZE 新增触发器时长显示（Tom）

  以前的版本把触发器执行时间计入总执行时间，但没有单独显示。
  现在可以看到每个触发器花费多少时间。

- 新增对 COPY 中 \x 十六进制转义的支持（Sergey Ten）

  以前的版本只支持八进制转义。

- 让 SHOW ALL 包含变量描述（Matthias Schmidt）

  SHOW varname 仍然只显示变量的值，不包含描述。

- 让 initdb 创建名为 postgres 的新标准数据库，并把工具转换为用 postgres 而不是 template1 做标准查找（Dave）

  在以前的版本中，template1 既被用作 createuser 等工具的默认连接，又被用作新数据库的模板。这导致 CREATE DATABASE 有时失败，因为如果其他人在模板数据库中就无法创建新数据库。有了此变更，默认连接数据库现在是 postgres，意味着 CREATE DATABASE 期间有人正在使用 template1 的可能性小得多。

- 通过把 /contrib/reindexdb 移入服务器来创建新的 reindexdb 命令行工具（Euler Taveira de Oliveira）

E.24.3.6. 数据类型和函数变更

- 为数组类型新增 MAX() 和 MIN() 聚合（Koju Iijima）

- 修复 to_date() 和 to_timestamp() 在同时使用 CC 和 YY 字段时合理行事（Karel Zak）

  如果格式说明包含 CC 而年规格为 YYY 或更长，忽略 CC。如果年规格为 YY 或更短，把 CC 解释为上一个世纪。

- 新增 md5(bytea)（Abhijit Menon-Sen）

  md5(text) 此前已存在。

- 新增对 numeric ^ numeric 的支持，基于 power(numeric, numeric)

  该函数此前已存在，但没有为它分配操作符。

- 修复 NUMERIC 取模，在计算期间正确截断商（Bruce）

  在以前的版本中，大值的取模有时因商的四舍五入而返回负结果。

- 新增函数 lastval()（Dennis Björklund）

  lastval() 是 currval() 的简化版本。它根据当前会话最近执行的 nextval() 或 setval() 调用自动确定正确的序列名。

- 新增 to_timestamp(DOUBLE PRECISION)（Michael Glaesemann）

  把自 1970 年以来的 Unix 秒数转换为 TIMESTAMP WITH TIMEZONE。

- 新增 pg_postmaster_start_time() 函数（Euler Taveira de Oliveira、Matthias Schmidt）

- 允许在 AT TIME ZONE 中完整使用时区名，而不只是以前可用的短列表（Magnus）

  此前 AT TIME ZONE 只支持预定义的时区名列表。现在可以使用任何受支持的时区名，例如：

  ```
  SELECT CURRENT_TIMESTAMP AT TIME ZONE 'Europe/London';
  ```

  在上面的查询中，使用的时区根据所提供日期当时有效的夏令时规则调整。

- 新增 GREATEST() 和 LEAST() 可变参数函数（Pavel Stehule）

  这些函数接受可变数量的参数，返回参数中的最大或最小值。

- 新增 pg_column_size()（Mark Kirkwood）

它返回列的存储大小，该列可能被压缩过。

- 新增 `regexp_replace()`（Atsushi Ogawa）

  这允许像 sed 那样做正则表达式替换。可选的标志参数允许选择全局（全部替换）和不区分大小写模式。

- 修复 interval 的除法和乘法（Bruce）

  以前的版本有时返回没有道理的结果，例如 `'4 months'::interval / 5` 返回 `'1 mon -6 days'`。

- 修复 timestamp、time 和 interval 输出的舍入行为（Tom）

  这修复了秒字段会显示为 `60` 而不是进位到更高字段的一些情形。

- 为 `interval` 类型新增单独的天字段，使一天间隔可与 24 小时间隔区分（Michael Glaesemann）

  包含夏令时调整的天不是 24 小时长，而通常是 23 或 25 小时。此变更创建了“多少天”间隔与“多少小时”间隔之间的概念区别。向时间戳加 `1 day` 现在即使中间发生夏令时调整，也在第二天给出相同的本地时间，而加 `24 hours` 在这种情况下会给出不同的本地时间。例如，按美国 DST 规则：

  ```
  '2005-04-03 00:00:00-05' + '1 day' = '2005-04-04 00:00:00-04'
  '2005-04-03 00:00:00-05' + '24 hours' = '2005-04-04 01:00:00-04'
  ```

- 新增 `justify_days()` 和 `justify_hours()`（Michael Glaesemann）

  这两个函数分别把天调整为适当的整月数加天数，把小时调整为适当的整天数加小时数。

- 把 `/contrib/dbsize` 移入后端，并重命名其中一些函数（Dave Page、Andreas Pflug）

  - `pg_tablespace_size()`
  - `pg_database_size()`
  - `pg_relation_size()`
  - `pg_total_relation_size()`
  - `pg_size_pretty()`

  `pg_total_relation_size()` 包括索引和 TOAST 表。

- 新增对集群目录只读文件访问的函数（Dave Page、Andreas Pflug）

  - `pg_stat_file()`
  - `pg_read_file()`
  - `pg_ls_dir()`

- 新增 `pg_reload_conf()`，强制重新装载配置文件（Dave Page、Andreas Pflug）

- 新增 `pg_rotate_logfile()`，强制轮转服务器日志文件（Dave Page、Andreas Pflug）

- 更改 `pg_stat_*` 视图以包含 TOAST 表（Tom）

E.24.3.7. 编码和区域设置变更

- 重命名一些编码使其更一致并遵循国际标准（Bruce）
  - `UNICODE` 现在是 `UTF8`
  - `ALT` 现在是 `WIN866`
  - `WIN` 现在是 `WIN1251`
  - `TCVN` 现在是 `WIN1258`

  原名仍然有效。
- 新增对 `WIN1252` 编码的支持（Roland Volkmann）
- 新增对四字节 `UTF8` 字符的支持（John Hansen）

  此前只支持一、二、三字节的 `UTF8` 字符。这对支持某些中文字符集特别重要。
- 允许 `EUC_JP` 和 `SJIS` 之间的直接转换以改进性能（Atsushi Ogawa）
- 允许 UTF8 编码在 Windows 上工作（Magnus）

  这是通过把 UTF8 映射到 Windows 本机 UTF16 实现完成的。

E.24.3.8. 常规服务器端语言变更

- 修复 `ALTER LANGUAGE RENAME`（Sergey Yatskevich）
- 允许通过 `ALTER FUNCTION` 修改函数特性，如严格性和易变性（Neil）
- 把函数参数的最大数量提高到 100（Tom）
- 允许 SQL 和 PL/pgSQL 函数使用 `OUT` 和 `INOUT` 参数（Tom）

  `OUT` 是函数返回值的另一种方式。除了使用 `RETURN`，还可以通过给声明为 `OUT` 或 `INOUT` 的参数赋值来返回值。这在某些情形下书写更简单，特别是需要返回多个值时。虽然以前的版本也可以从函数返回多个值，但这大幅简化了该过程。
  （该特性将在未来版本中扩展到其他服务器端语言。）
- 把语言处理函数移入 `pg_catalog` 模式

  这使按需删除 public 模式变得更容易。
- 为 SPI 新增 `SPI_getnspname()`（Neil）

E.24.3.9. PL/pgSQL 服务器端语言变更

- 彻底改造 PL/pgSQL 函数的内存管理（Neil）

  每个函数的分析树现在存储在单独的内存上下文中。这使该内存在不再需要时可以轻松回收。
- 在 `CREATE FUNCTION` 时检查函数语法，而不是在运行时（Neil）

此前，大多数语法错误只在函数执行时才被报告。

- 允许 OPEN 打开非 SELECT 查询，如 EXPLAIN 和 SHOW（Tom）
- 不再要求函数发出 RETURN 语句（Tom）

  这是新增的 OUT 和 INOUT 功能的副产品。当不需要提供函数返回值时可以省略 RETURN。
- 为 PL/pgSQL 的 EXECUTE 语句新增可选的 INTO 子句（Pavel Stehule、Neil）
- 让 CREATE TABLE AS 设置 ROW_COUNT（Tom）
- 定义 SQLSTATE 和 SQLERRM，返回当前异常的 SQLSTATE 和错误消息（Pavel Stehule、Neil）

  这些变量只在异常块内定义。
- 允许 RAISE 语句的参数为表达式（Pavel Stehule、Neil）
- 新增循环 CONTINUE 语句（Pavel Stehule、Neil）
- 允许块和循环标签（Pavel Stehule）

E.24.3.10. PL/Perl 服务器端语言变更

- 允许高效返回大结果集（Abhijit Menon-Sen）

  这允许函数使用 return_next() 来避免在内存中构建整个结果集。
- 允许逐行获取查询结果（Abhijit Menon-Sen）

  这允许函数使用 spi_query() 和 spi_fetchrow() 来避免在内存中累积整个结果集。
- 如果服务器编码为 UTF8，强制 PL/Perl 把字符串当作 UTF8 处理（David Kamholz）
- 为 PL/Perl 新增验证器函数（Andrew）

  这使语法错误可以在定义时而不是执行时报告。
- 当函数返回数组类型时允许 PL/Perl 返回 Perl 数组（Andrew）

  这基本上把 PostgreSQL 数组映射到 Perl 数组。
- 允许 Perl 非致命警告生成 NOTICE 消息（Andrew）
- 允许启用 Perl 的 strict 模式（Andrew）

E.24.3.11. psql 变更

- 新增 \set ON_ERROR_ROLLBACK，允许事务中的语句出错而不影响事务的其余部分（Greg Sabino Mullane）

  这基本上是通过把每条语句包装在子事务中实现的。
- 新增对 psql 变量中 \x 十六进制字符串的支持（Bruce）

  八进制转义此前已支持。

- 新增对 `troff -ms` 输出格式的支持（Roger Leigh）
- 允许历史文件的位置由 `HISTFILE` 控制（Andreas Seltenreich）

  这允许配置逐数据库的历史存储。
- 防止 `\x`（扩展模式）影响 `\d tablename` 的输出（Neil）
- 为 psql 新增 `-L` 选项以记录会话（Lorne Sunley）

  添加此选项是因为某些操作系统没有简单的命令行活动记录功能。
- 让 `\d` 显示索引的表空间（Qingqing Zhou）
- 允许 psql 帮助（`\h`）对正确的帮助信息做最佳猜测（Greg Sabino Mullane）

  这使用户只需在语法错误的查询前加上 `\h` 就能获得支持语法的帮助。此前必须删除命令名之外的任何附加查询文本才能使用 `\h`。
- 新增 `\pset numericlocale`，允许以区域感知的格式输出数字（Eugen Nedelcu）

  例如，使用 `C` 区域设置时 `100000` 会输出为 `100,000.0`，而欧洲区域设置可能输出为 `100.000,0`。
- 让启动横幅在服务器版本号与 psql 版本号不同时同时显示两者（Bruce）

  此外，如果服务器和 psql 来自不同的主版本，会显示警告。

E.24.3.12. pg_dump 变更

- 为 pg_restore 新增 `-n` / `--schema` 开关（Richard van den Berg）

  这允许只恢复指定模式中的对象。
- 允许 pg_dump 即使在文本模式也转储大对象（Tom）

  有了此变更，大对象现在总是被转储；以前的 `-b` 开关是无操作。
- 允许 pg_dump 转储大对象的一致快照（Tom）
- 转储大对象的注释（Tom）
- 为 pg_dump 新增 `--encoding`（Magnus Hagander）

  这允许以不同于服务器编码的编码转储数据库。
  当把转储传送到不同编码的机器时这很有价值。
- 过程语言依靠 `pg_pltemplate`（Tom）

  如果过程语言的调用处理器在 `pg_catalog` 模式中，pg_dump 不转储该处理器。而是只用 `CREATE LANGUAGE` *`name`* 转储该语言，依赖 `pg_pltemplate` 目录在装载时提供该语言的创建参数。

E.24.3.13. libpq 变更

- 新增 `PGPASSFILE` 环境变量，指定口令文件的文件名（Andrew）

- 新增 `lo_create()`，类似于 `lo_creat()` 但允许指定大对象的 OID（Tom）
- 让 libpq 在 `malloc()` 失败时一致地向客户端应用程序返回错误（Neil）

E.24.3.14. 源代码变更

- 修复 pgxs 以支持针对重新定位的安装构建
- 新增使用 Intel 编译器的 Itanium 处理器自旋锁支持（Vikram Kalsi）
- 新增 Windows 的 Kerberos 5 支持（Magnus）
- 新增中文 FAQ（laser@pgsqldb.com）
- 把 Rendezvous 改名为 Bonjour 以与 OS/X 特性改名一致（Bruce）
- 新增 Darwin 上对 `fsync_writethrough` 的支持（Chris Campbell）
- 精简服务器、优化器和锁系统内部的信息传递（Tom）
- 允许用 MSVC 编译 pg_config（Andrew）

  这是用 MSVC 构建 DBD::Pg 所必需的。
- 移除对 Kerberos V4 的支持（Magnus）

  Kerberos 4 有安全漏洞且不再维护。
- 代码清理（由 EnterpriseDB 执行 Coverity 静态分析）
- 修改 `postgresql.conf`，使用文档默认的 `on/off` 而不是 `true/false`（Bruce）
- 增强 pg_config，使其能报告更多构建期值（Tom）
- 允许在 Windows 上把 libpq 构建为线程安全（Dave Page）
- 允许在 Windows 上使用 IPv6 连接（Andrew）
- 新增关于 I/O 子系统可靠性的服务器管理文档（Bruce）
- 把私有声明从 `gist.h` 移到 `gist_private.h`（Neil）

  在以前的版本中，`gist.h` 既包含公共 GiST API（供 GiST 索引实现的作者使用），也包含 GiST 自身实现使用的一些私有声明。后者已被移到单独的文件 `gist_private.h`。大多数 GiST 索引实现应不受影响。
- 彻底改造 GiST 内存管理（Neil）

  GiST 方法现在总是在短生命周期的内存上下文中调用。因此，通过 `palloc()` 分配的内存将被自动回收，GiST 索引实现不需要通过 `pfree()` 手动释放分配的内存。

E.24.3.15. 附加模块变更

- 新增 `/contrib/pg_buffercache` contrib 模块（Mark Kirkwood）

  它显示缓冲区缓存的内容，用于调试和性能调优。

- 移除 `/contrib/array`，因为它已过时（Tom）
- 清理 `/contrib/lo` 模块（Tom）
- 把 `/contrib/findoidjoins` 移到 `/src/tools`（Tom）
- 从 `/contrib/cube` 移除 `<<`、`>>`、`&<` 和 `&>` 操作符

  这些操作符没有用处。
- 改进 `/contrib/btree_gist`（Janko Richter）
- 改进 `/contrib/pgbench`（Tomoaki Sato、Tatsuo）

  现在有一套用用户提供的 SQL 命令脚本进行测试的设施，而不是只有硬编码的命令序列。
- 改进 `/contrib/pgcrypto`（Marko Kreen）
  - 实现 OpenPGP 对称密钥和公钥加密

    支持 RSA 和 Elgamal 两种公钥算法。
  - 独立构建：包含 SHA256/384/512 哈希、Fortuna PRNG
  - OpenSSL 构建：支持 3DES，OpenSSL < 0.9.7 时使用内建 AES
  - 从 `configure` 结果获取构建参数（OpenSSL、zlib）

    不再需要编辑 `Makefile`。
  - 移除对 `libmhash` 和 `libmcrypt` 的支持

E.25. 版本 8.0.26

发行日期
2010-10-04

本次发布包含对 8.0.25 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

这预计是 PostgreSQL 8.0.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.25.1. 迁移到版本 8.0.26

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅 8.0.22 的发布说明。

E.25.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一

点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，
  会出现此错误。

- 防御返回 setof record 但并非所有返回行实际为同一行类型的函数（Tom Lane）
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。

- 避免在为深层嵌套的子事务分配 XID 时递归（Andres Freund、Robert Haas）

  在栈空间有限时，原来的编码可能导致崩溃。

- 修复 `log_line_prefix` 的 `%i` 转义，它可能在后端启动早期产生垃圾（Tom Lane）
- 修复启用归档时 `ALTER TABLE ... SET TABLESPACE` 中可能的数据损坏（Jeff Davis）
- 允许 `CREATE DATABASE` 和 `ALTER DATABASE ... SET TABLESPACE` 被查询取消中断（Guillaume Lelarge）
- 在 PL/Python 中，防御 `PyCObject_AsVoidPtr` 和 `PyCObject_FromVoidPtr` 的空指针结果（Peter Eisentraut）
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 修复 `contrib/dblink` 以正确处理超过 62 字节的连接名（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）
- 把时区数据文件更新为 tzdata 版本 2010l for DST law changes in Egypt and Palestine; also historical corrections for Finland。

  此变更还为两个密克罗尼西亚时区新增了名称：Pacific/Chuuk 现在优先于 Pacific/Truk（首选缩写为 CHUT 而非 TRUT），Pacific/Pohnpei 优先于 Pacific/Ponape。

E.26. 版本 8.0.25

发行日期
2010-05-17

本次发布包含对 8.0.24 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 8.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.26.1. 迁移到版本 8.0.25

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅 8.0.22 的发布说明。

E.26.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）

- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，
  因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 防止 psql 在展开引用自身的变量时无限递归（Tom）

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

- 把时区数据文件更新为 tzdata 版本 2010j for DST law changes in Argentina, Australian Antarctic, Bangladesh, Mexico, Morocco, Pakistan, Palestine, Russia, Syria, Tunisia; also historical corrections for Taiwan。

E.27. 版本 8.0.24

发行日期
2010-03-15

本次发布包含对 8.0.23 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 8.0.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.27.1. 迁移到版本 8.0.24

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅 8.0.22 的发布说明。

E.27.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 修复试图从子事务启动失败中恢复时可能的崩溃（Tom）

- 修复与使用保存点以及客户端编码不同于服务器编码相关的服务器内存泄漏（Tom）

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复整数到位串的转换：当输出位宽比给定整数宽出非 8 位倍数时，正确处理第一个小数字节（Tom Lane）

- 修复若干病态缓慢的正则表达式匹配情形（Tom）

- 修复备份历史文件中的 `STOP WAL LOCATION` 条目：当结束位置恰好位于段边界时报告下一个 WAL 段的名称（Itagaki Takahiro）

- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。

- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 修复组合列被设为 NULL 的一种情形下 plpgsql 的失败（Tom）
- 在 PL/Python 中添加 `volatile` 标记，以避免可能的编译器特定错误行为（Zdenek Kotala）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）
- 修复 `contrib/xml2` 中因草率内存管理导致的若干崩溃（Tom）
- 把时区数据文件更新为 tzdata 版本 2010e for DST law changes in Bangladesh, Chile, Fiji, Mexico, Paraguay, Samoa。

E.28. 版本 8.0.23

发行日期
2009-12-14

本次发布包含对 8.0.22 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.28.1. 迁移到版本 8.0.23

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.22 的版本升级，请参阅 8.0.22 的发布说明。

E.28.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

这可能发生在规划器对哈希连接结果大小的估计极大时。

- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）
- 修复在子事务内访问的游标所用临时文件被过早删除的问题（Heikki）
- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。
- 修复 PL/Python 异常处理中的罕见崩溃（Peter）
- 确保 psql 的 flex 模块用正确的系统头文件定义编译（Tom）

  这修复了在 `--enable-largefile` 导致生成代码不兼容变更的平台上的构建失败。
- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）
- 把时区数据文件更新为 tzdata 版本 2009s for DST law changes in Antarctica, Argentina, Bangladesh, Fiji, Novokuznetsk, Pakistan, Palestine, Samoa, Syria; also historical corrections for Hong Kong。

E.29. 版本 8.0.22

发行日期
2009-09-09

本次发布包含对 8.0.21 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.29.1. 迁移到版本 8.0.22

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 8.0.22 后对它们执行 `REINDEX`。另外，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.29.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）
- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）
- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 把 `to_char(..., 'TH')` 在 `'HH'/'HH12'` 下当作大写序数后缀（Heikki）

  它以前被当作 `'th'`（小写）处理。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 修复 pg_ctl，使其在 `postgresql.conf` 为空时不进入无限循环（Jeff Davis）

- 修复 `contrib/xml2` 的 `xslt_process()`，使其正确处理最大参数数量（二十个）（Tom）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）

- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

- 把时区数据文件更新为 tzdata 版本 2009l for DST law changes in Bangladesh, Egypt, Jordan, Pakistan, Argentina/San_Luis, Cuba, Jordan (historical correction only), Mauritius, Morocco, Palestine, Syria, Tunisia。

E.30. 版本 8.0.21

发行日期
2009-03-16

本次发布包含对 8.0.20 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.30.1. 迁移到版本 8.0.21

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.30.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，

解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）
- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.31. 版本 8.0.20

发行日期
2009-02-02

本次发布包含对 8.0.19 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.31.1. 迁移到版本 8.0.20

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.31.2. 变更

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）
- 更新时区数据文件到 tzdata 发行版 2009a（针对加德满都以及瑞士、古巴的历史 DST 修正）

E.32. 版本 8.0.19

发行日期
2008-11-03

本次发布包含对 8.0.18 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.32.1. 迁移到版本 8.0.19

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.32.2. 变更

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
  修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复从触发器调用深层嵌套函数时可能的崩溃（Tom）
- 确保新定义的 PL/pgSQL 触发器函数被当作普通函数调用时报错（Tom）
- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。

- 修复 ecpg 对 `CREATE USER` 的解析（Michael）
- 修复 `pg_ctl restart` 最近损坏的问题（Tom）
- 更新时区数据文件到 tzdata 发行版 2008i（针对阿根廷、巴西、毛里求斯、叙利亚的 DST 法规变更）

E.33. 版本 8.0.18

发行日期
2008-09-22

本次发布包含对 8.0.17 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.33.1. 迁移到版本 8.0.18

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.33.2. 变更

- 把本地锁计数器从 32 位拓宽到 64 位（Tom）

  这是针对计数器在足够长的事务中可能溢出的报告，该溢出会导致意外的“锁已被持有”错误。

- 在执行器启动时添加检查，确保 `INSERT` 或 `UPDATE` 产生的元组与目标表的当前行类型匹配（Tom）

  `ALTER COLUMN TYPE` 之后重用先前缓存的计划可能产生这种情形。该检查防范随之而来的数据损坏和/或崩溃。

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。

- 修复 PL/Tcl 在 Tcl 8.5 下的正确行为，并更小心地处理发送到 Tcl 或从 Tcl 发出的数据的编码（Tom）
- 修复 PL/Python 使其与 Python 2.5 一起工作

  这是 8.2 开发周期中所做修复的向后移植。

- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）
- 修复 pg_ctl，使其在 `restart` 期间正确保留 postmaster 命令行参数（Bruce）
- 更新时区数据文件到 tzdata 发行版 2008f（针对阿根廷、巴哈马、巴西、毛里求斯、摩洛哥、巴基斯坦、巴勒斯坦和巴拉圭的 DST 法规变更）

E.34. 版本 8.0.17

发行日期
2008-06-12

此发行版在 8.0.16 的基础上包含一个严重的缺陷修复。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.34.1. 迁移到版本 8.0.17

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.34.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

E.35. 版本 8.0.16

发行日期

从未发布

本次发布包含对 8.0.15 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.35.1. 迁移到版本 8.0.16

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.35.2. 变更

- 修复 `ALTER TABLE ADD COLUMN ... PRIMARY KEY`，使新列被正确检查是否已全部初始化为非空（Brendan Jurd）

  以前的版本完全忽略了对这一要求的检查。

- 修复从多个都继承自共同祖先的父关系继承“同一”约束时 `CREATE TABLE` 可能失败的问题（Tom）

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。

- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。

- 更新时区数据文件到 tzdata 发行版 2008c（针对摩洛哥、伊拉克、乔巴山、巴基斯坦、叙利亚、古巴、阿根廷/圣路易斯和智利的 DST 法规变更）

- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）

- 修复 `contrib/xml2` 的 `xpath_table()` 函数在输入查询返回 NULL 值时的核心转储（Tom）

- 修复 `contrib/xml2` 的 makefile，使其不覆盖 `CFLAGS`（Tom）

- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 修复使用哈希索引的查询期间发生错误时的罕见崩溃（Heikki）

- 修复公元前年份 2 月 29 日日期时间值的输入（Tom）

  以前的编码弄错了哪些年份是闰年。

- 修复 `ALTER OWNER` 某些变体中的“无法识别的节点类型”错误（Tom）

- 修复 pg_ctl，使其从命令行选项正确提取 postmaster 的端口号（Itagaki Takahiro、Tom）

  此前 `pg_ctl start -w` 可能试图在错误端口上联系 postmaster，导致虚假的启动失败报告。

- 使用 `-fwrapv` 防御最新 gcc 版本可能的错误优化（Tom）

  已知用 gcc 4.3 或更高版本构建 PostgreSQL 时这是必需的。

- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

  显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.36. 版本 8.0.15

发行日期
2008-01-07

此发行版包含 8.0.14 之后的多种修复，包括对重要安全问题的修复。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

这是 PostgreSQL 社区为 Windows 制作二进制包的最后一个 8.0.X 版本。建议 Windows 用户迁移到 8.2.X 或更高版本，因为 8.2.X 中有一些不便于向后移植的 Windows 特定修复。8.0.X 在其他平台上将继续受支持。

E.36.1. 迁移到版本 8.0.15

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.36.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  8.0.14 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 更新时区数据文件到 tzdata 发行版 2007k（特别是阿根廷最近的变更）（Tom）

- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）

- 保留被 `ALTER TABLE ... ALTER COLUMN TYPE` 重建的索引的表空间（Tom）

- 让归档恢复总是开始新的 WAL 时间线，而不是只在使用恢复停止时间时（Simon）

  这避免了试图覆盖最后一个 WAL 段现有归档副本的极端情形风险，而且看起来比原来的定义更简单干净。

- 当表太小而不适用时，让 `VACUUM` 不使用全部 `maintenance_work_mem`（Alvaro）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）

- 修复 PL/Perl 在平台 Perl 把 `bool` 类型定义为 `int` 而非 `char` 时的问题（Tom）

  虽然理论上这可能在任何地方发生，但没有标准 Perl 构建这样做.....直到 Mac OS X 10.5。

- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）

- 修复 pg_dump，使其正确处理默认表达式与父表不同的继承子表（Tom）

- ecpg 解析器修复（Michael）

- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）

- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）

- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）

- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.37. 版本 8.0.14

发行日期
2007-09-17

本次发布包含对 8.0.13 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.37.1. 迁移到版本 8.0.14

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.37.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复日志，使使用 syslogger 进程时日志消息永不交错（Andrew）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 修复某些外键极端情形的错误处理（Tom）
- 防止 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 更新时区数据库规则，特别是新西兰即将到来的变更（Tom）
- Windows 套接字改进（Magnus）
- 因可能的编码不匹配，在 Windows 上抑制日志时间戳中的时区名（`%Z`）（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.38. 版本 8.0.13

发行日期
2007-04-23

此发行版包含 8.0.12 之后的多种修复，包括一个安全修复。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.38.1. 迁移到版本 8.0.13

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.38.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）
- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）
- 修复哈希索引扩容期间的 PANIC（8.0.10 中引入的缺陷）（Tom）
- 修复 POSIX 风格时区规范以遵循新的美国 DST 规则（Tom）

E.39. 版本 8.0.12

发行日期
2007-02-07

此发行版包含 8.0.11 之后的一个修复。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.39.1. 迁移到版本 8.0.12

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.39.2. 变更

- 移除约束和函数索引中对类型长度的过度严格检查（Tom）

E.40. 版本 8.0.11

发行日期
2007-02-05

此发行版包含 8.0.10 之后的多种修复，包括一个安全修复。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.40.1. 迁移到版本 8.0.11

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.40.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  这些漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，以及更改表列的数据类型（CVE-2007-0555、CVE-2007-0556）。这些错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.41. 版本 8.0.10

发行日期
2007-01-08

本次发布包含对 8.0.9 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.41.1. 迁移到版本 8.0.10

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.41.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）

- 修复 `VACUUM` 截断跨越 GB 边界的大关系时的竞态条件（Tom）

- 修复影响多 GB 哈希索引的缺陷（Tom）

- 修复 Windows 信号处理中可能的死锁（Teodor）

- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）

- 修复 ecpg 连接期间的内存泄漏（Michael）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。
- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。
- 更新时区数据库

  这特别影响澳大利亚和加拿大的夏令时规则。

E.42. 版本 8.0.9

发行日期
2006-10-16

本次发布包含对 8.0.8 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.42.1. 迁移到版本 8.0.9

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.42.2. 变更

- 修复在规则 WHERE 表达式中引用 `NEW` 行值时的崩溃（Tom）
- 修复把无类型字面值当作 ANYARRAY 时的核心转储
- 修复查询包含返回多行的 SQL 函数时 AFTER 触发器的错误处理（Tom）
- 修复 `ALTER TABLE ... TYPE`，对 `USING` 子句重新检查 `NOT NULL`（Tom）
- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。
- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- ecpg 的大量健壮性修复（Joachim Wieland）
- 修复 /contrib/dbmirror 中的反斜杠转义
- 修复 Win32 上统计收集的不稳定（Tom、Andrew）
- AIX 和 Intel 编译器的修复（Tom）

E.43. 版本 8.0.8

发行日期
2006-05-23

此发行版包含 8.0.7 之后的多种修复，包括针对极其严重安全问题的补丁。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.43.1. 迁移到版本 8.0.8

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.43.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 修复有时导致 OR 索引扫描漏掉本应返回的行的缺陷

- 修复 btree 索引被截断情形的 WAL 重放
- 修复涉及 | 的模式的 SIMILAR TO（Tom）
- 修复 SELECT INTO 和 CREATE TABLE AS，在默认表空间而不是基目录中创建表（Kris Jurka）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 修复 Intel Mac 上的 Bonjour（Ashley Clark）
- 修复各种小的内存泄漏
- 修复某些 Win32 系统上的口令提示问题（Robert Kinberg）

E.44. 版本 8.0.7

发行日期
2006-02-14

本次发布包含对 8.0.6 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.44.1. 迁移到版本 8.0.7

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.6 的版本升级，请参阅 8.0.6 的发布说明。

E.44.2. 变更

- 修复 SET SESSION AUTHORIZATION 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 8.0.4、7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 和 pg_subtrans 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复缓存失效消息恰好在错误时刻到达时可能导致崩溃的情形（Tom）
- 为预备语句中的 UNKNOWN 参数正确检查 DOMAIN 约束（Neil）
- 确保 ALTER COLUMN TYPE 按正确顺序处理 FOREIGN KEY、UNIQUE 和 PRIMARY KEY 约束（Nakano Yoshihisa）
- 修复以允许恢复有对自定义操作符或操作符类的跨模式引用的转储（Tom）
- 允许 pg_restore 在 COPY 失败后正确继续；此前它试图把剩余的 COPY 数据当作 SQL 命令（Stephen Frost）

- 修复未指定数据目录时 pg_ctl 的 `unregister` 崩溃（Magnus）
- 修复 AMD64 和 PPC 上 ecpg 的崩溃（Neil）
- 修复 PL/python 参数传递期间发生错误时的正确恢复（Neil）
- 修复 PL/perl 在 Win32 上对区域设置的处理以与后端匹配（Andrew）
- 修复 Win32 上 `postgresql.conf` 中 `log_min_messages` 设为 `DEBUG3` 或更高时的崩溃（Bruce）
- 修复 Win32、Cygwin、OS X、AIX 的 pgxs `-L` 库路径指定（Bruce）
- 在检查 Win32 管理员权限时检查 SID 是否启用（Magnus）
- 正确拒绝超出范围的日期输入（Kris Jurka）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.45. 版本 8.0.6

发行日期
2006-01-09

本次发布包含对 8.0.5 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.45.1. 迁移到版本 8.0.6

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅 8.0.3 的发布说明。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.45.2. 变更

- 修复 Windows 代码，使 ShmemBackendArray 没有更多空间时 postmaster 继续而不是退出（Magnus）

  如果过多连接请求几乎同时到达，以前的行为可能导致拒绝服务情形。这只适用于 Windows 移植。

- 修复 8.0 中引入的缺陷，它可能使 ReadBuffer 返回已使用过的页当作新页，可能导致最近提交的数据丢失（Tom）
- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）
- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 允许更灵活地重新定位安装目录（Tom）

  以前的版本只有在所有安装目录路径除最后一个组件外都相同时才支持重新定位。

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）
- 对返回 `RECORD` 的函数的各种修复（Tom）
- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.46. 版本 8.0.5

发行日期
2005-12-12

本次发布包含对 8.0.4 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.46.1. 迁移到版本 8.0.5

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅 8.0.3 的发布说明。

E.46.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- 修复从错误恢复后 bgwriter 的问题（Tom）

  发现后台写入器在写错误后泄漏缓冲区固定。虽然本身不致命，但这可能导致后续 VACUUM 命令的神秘阻塞。

- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败
- `/contrib/ltree` 修复（Teodor）
- AIX 和 HPUX 编译修复（Tom）
- 在 Windows NO_SYSTEM_RESOURCES 错误后重试文件读写（Qingqing Zhou）

- 修复 `log_line_prefix` 包含 `%i` 时的间歇性失败
- 修复 Windows 上长脚本的 psql 性能问题（Merlin Moncure）
- 修复 `pg_group` 平面文件更新的缺失
- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。
- 把时区初始化推迟到 `postmaster.pid` 创建之后

  这避免让期待 pid 文件快速出现的启动脚本困惑。
- 防止表被删除后 pg_autovacuum 的核心转储
- 修复对子查询结果的整行引用（`foo.*`）的问题

E.47. 版本 8.0.4

发行日期
2005-10-04

本次发布包含对 8.0.3 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.47.1. 迁移到版本 8.0.4

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 8.0.3 的版本升级，请参阅 8.0.3 的发布说明。

E.47.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。
- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

  在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。
- 在提交 `CREATE DATABASE` 前强制检查点

  这应能修复最近报告的 `CREATE DATABASE` 之后不久发生崩溃时“索引不是 btree”失败的报告。
- 修正 `COPY` 中只读事务测试的方向

  以前的代码禁止 `COPY TO`，而它应禁止的是 `COPY FROM`。
- 处理 `COPY` CSV 模式输入中连续的嵌入换行符

- 修复年末附近日期的 `date_trunc(week)`
- 修复只引用内侧关系的外连接 ON 子句的规划问题
- 对 `x FULL JOIN y ON true` 极端情形的进一步修复
- 修复对 `x IN (SELECT DISTINCT ...)` 及相关情形的过度优化
- 修复因考虑不周的“模糊”代价比较导致小 `LIMIT` 值查询的错误规划
- 让 `array_in` 和 `array_recv` 更严格地验证其 OID 参数
- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题
- 改进日期时间解析的健壮性
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- 改进 MIPS 和 M68K 自旋锁代码
- postmaster 启动期间不要试图打开超过 `max_files_per_process` 个文件
- 各种内存泄漏修复
- 各种可移植性改进
- 更新时区数据文件
- 改进 Windows 上 DLL 装载失败的处理
- 改进 Windows 上的随机数生成
- 让 `psql -f filename` 在打开文件失败时返回非零退出码
- 更改 pg_dump 以更可靠地处理继承的检查约束
- 修复 Windows 上 pg_restore 的口令提示
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`
- 修复 PL/Perl 的 `%_SHARED`，使其真正共享
- 修复 `contrib/pg_autovacuum`，允许超过 2000 秒的休眠间隔
- 更新 `contrib/tsearch2` 以使用当前的 Snowball 代码

E.48. 版本 8.0.3

发行日期
2005-05-09

此发行版包含 8.0.2 之后的多种修复，包括几个与安全相关的问题。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.48.1. 迁移到版本 8.0.3

对于运行 8.0.X 的用户，不需要进行转储/恢复。不过，它是处理在 8.0.X 系统目录初始内容中发现的两个重要安全问题的可能方法之一。用 8.0.3 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

较大的安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）

较小的问题是 `contrib/tsearch2` 模块创建了几个被错误声明为返回 `internal` 的函数，而它们并不接受 `internal` 参数。这破坏了所有使用 `internal` 参数的函数的类型安全。

强烈建议所有安装通过 initdb 或按照下述手工修复程序修复这些错误。这些错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行 7.4.8 发行说明中所示的相同手工修复程序。

E.48.2. 变更

- 更改编码函数签名以防止误用
- 更改 `contrib/tsearch2` 以避免不安全地使用 `INTERNAL` 函数结果
- 防范 `record_out` 的错误第二参数
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

  此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化

- 防止 `COALESCE(NULL,NULL)` 的崩溃
- 修复 libpq 的 Borland makefile
- 修复 `contrib/btree_gist` 的 `timetz` 类型（Teodor）
- 让 `pg_ctl` 检查在 `postmaster.pid` 中找到的 PID，看它是否仍是活进程
- 修复因添加转储时间戳导致的 `pg_dump`/`pg_restore` 问题
- 修复事务提交期间物化可保持游标与触发延迟触发器之间的相互作用
- 修复返回按引用传递数据类型的 SQL 函数中的内存泄漏

E.49. 版本 8.0.2

发行日期
2005-04-07

本次发布包含对 8.0.1 的多方面修复。有关 8.0 大版本中新特性的信息，请参见第 E.51 节。.

E.49.1. 迁移到版本 8.0.2

对于运行 8.0.* 的用户，不需要进行转储/恢复。此发行版更新了 PostgreSQL 库的主版本号，因此如果某些用户应用程序找不到正确编号的共享库，可能需要重新链接它们。

E.49.2. 变更

- 递增所有接口库的主版本号（Bruce）

  这本应在 8.0.0 中完成。需要它才能使 7.4.X 版本的 PostgreSQL 客户端应用程序（如 psql）能与 8.0.X 应用程序在同一台机器上使用。这可能需要重新链接使用这些库的用户应用程序。

- 新增仅 Windows 的 `wal_sync_method` 设置 `fsync_writethrough`（Magnus、Bruce）

  此设置使 PostgreSQL 在写 WAL 时穿透磁盘驱动器的任何写缓存。此行为以前称为 `fsync`，但因它在其他平台上的行为与 `fsync` 大不相同而被改名。

- 在 Windows 上启用 `wal_sync_method` 的 `open_datasync` 设置，并使其成为该平台的默认（Magnus、Bruce）

  由于默认不再是 `fsync_writethrough`，如果磁盘驱动器启用了写缓存，断电期间可能丢失数据。要在 Windows 上关闭写缓存，请在设备管理器中选择驱动器属性，然后选择策略。

- 新缓存管理算法 2Q 取代 ARC（Tom）

  这是为了避免 ARC 的一项待决美国专利。对某些负载，2Q 代码可能比 ARC 慢几个百分点。更好的缓存管理算法将出现在 8.1 中。

- 规划器调整以改进新建表上的行为（Tom）
- 允许 plpgsql 给初始为 `NULL` 的数组的元素赋值（Tom）

此前数组会保持 NULL，但现在它变成单元素数组。主 SQL 引擎在 8.0 中已改为以这种方式处理空数组值的 UPDATE，但 plpgsql 中的类似情形被遗漏了。

- 把 plpython 函数体中的 \r\n 和 \r 转换为 \n（Michael Fuhr）

  这防止在 Windows 或 Mac 客户端上编写 plpython 代码时出现语法错误。

- 允许 SPI 游标处理返回行的工具命令，如 EXPLAIN（Tom）
- 修复 ALTER TABLE SET WITHOUT OIDS 之后 CLUSTER 的失败（Tom）
- 降低 ALTER TABLE ADD COLUMN 的内存使用（Neil）
- 修复 ALTER LANGUAGE RENAME（Tom）
- 记录 pg_ctl 仅 Windows 的 register 和 unregister 选项（Magnus）
- 确保后端关闭期间完成的操作被统计收集器计数

  这有望解决关于 pg_autovacuum 不够频繁地清理系统目录的报告——它没有被通知后端退出期间临时表删除导致的目录删除。

- 把 Windows 上配置参数 log_destination 的默认改为 eventlog（Magnus）

  默认情况下，在 Windows 上运行的服务器现在将把日志输出发送到 Windows 事件记录器而不是标准错误。

- 使 Kerberos 认证在 Windows 上工作（Magnus）
- 允许没有 CREATEDB 权限标志的超级用户执行 ALTER DATABASE RENAME（Tom）
- 修改 CREATE 和 DROP DATABASE 的 WAL 日志条目，不指定绝对路径（Tom）

  这允许在数据库位置可能不同的另一台机器上做时间点恢复。注意 CREATE TABLESPACE 在这种情形下仍有风险。

- 修复后端带着创建过表并对其打开游标的未决事务退出时的崩溃（Tom）
- 修复 array_map()，使其能调用 PL 函数（Tom）
- 若干 contrib/tsearch2 和 contrib/btree_gist 修复（Teodor）
- 修复某些平台上一些 contrib/pgcrypto 函数的崩溃（Marko Kreen）
- 修复 contrib/intagg 在 64 位平台上的问题（Tom）
- 修复 ecpg 解析 CREATE 语句的缺陷（Michael）
- 绕过 powerpc 和 amd64 上的 gcc 缺陷，它导致 ecpg 中的问题（Christof Petig）
- 当区域设置为 C 时不使用区域感知版本的 upper()、lower() 和 initcap()（Bruce）

  这使这些函数能在区域设置为 C 时对非 7 位数据报错的平台上工作。

- 修复 quote_ident()，给匹配关键字的名称加引号（Tom）
- 修复 to_date() 在同时使用 CC 和 YY 字段时合理行事（Karel）

- 防止 `to_char(interval)` 在给定零月 interval 时失败（Tom）
- 修复 `date_trunc('week')` 返回错误星期的问题（Bruce）

  `date_trunc('week')` 在某些年份对一月的前几天返回错误的年份。
- 为 `INET` 数据类型的 `D` 类地址使用正确的默认掩码长度（Tom）

E.50. 版本 8.0.1

发行日期
2005-01-31

此发行版包含 8.0.0 之后的多种修复，包括几个与安全相关的问题。有关 8.0 主版本新特性的信息，请参阅第 E.51 节。

E.50.1. 迁移到版本 8.0.1

对于运行 8.0.0 的用户，不需要进行转储/恢复。

E.50.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。
- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。
- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 让 `ALTER TABLE ADD COLUMN` 在所有情形下强制执行域约束
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。
- 改进分组聚合查询的规划
- `ROLLBACK TO` *`savepoint`* 关闭自保存点以来创建的游标
- 修复 Windows 上后端栈大小不足的问题
- 在 Windows 上避免 SHGetSpecialFolderPath()（Magnus）

- 修复作为 Windows 服务运行 pg_autovacuum 的一些问题（Dave Page）
- pg_dump/pg_restore 的多个小缺陷修复
- 修复 ecpg 在 typedef 中使用命名结构时的段错误（Michael）

E.51. 版本 8.0

发行日期
2005-01-19

E.51.1. 概述

此发行版的主要变更：

Microsoft Windows Native Server

这是第一个作为服务器原生运行于 Microsoft Windows® 上的 PostgreSQL 版本。它可以作为 Windows 服务运行。此版本支持基于 NT 的 Windows 版本，如 Windows 2000 SP4、Windows XP 和 Windows 2003。更老的版本如 Windows 95、Windows 98 和 Windows ME 不受支持，因为这些操作系统没有支持 PostgreSQL 的基础设施。已创建单独的安装程序项目以简化在 Windows 上的安装——参见http://www.postgresql.org/ftp/win32/。

虽然在整个发布周期中都做了测试，但 Windows 移植没有 PostgreSQL 在 Unix 平台上多年生产环境使用的积累。因此应像对待新产品一样谨慎对待它。

以前的版本需要 Unix 仿真工具包 Cygwin 才能在 Windows 操作系统上运行服务器。PostgreSQL 在 Windows 上的本机客户端支持已有多年。

Savepoints

保存点允许中止事务的特定部分而不影响事务的其余部分。以前的版本没有这种能力；事务内的语句失败除了中止整个事务外无法恢复。
此特性对需要在复杂事务内进行错误恢复的应用程序作者很有价值。

Point-In-Time Recovery

在以前的版本中，除了从以前的备份恢复或使用备用复制服务器外，无法从磁盘驱动器故障中恢复。时间点恢复允许对服务器做连续备份。
你可以恢复到故障点或过去的某个事务。

Tablespaces

表空间允许管理员为各个表、索引和数据库的存储选择不同的文件系统。
这改进了性能和对磁盘空间使用的控制。以前的版本使用 initlocation 和手工符号链接管理来完成此类任务。

Improved Buffer Management, `CHECKPOINT`, `VACUUM`

此版本有更智能的缓冲区替换策略，能更好地利用可用共享缓冲区并改进性能。
清理和检查点的性能影响也减轻了。

Change Column Types

现在可以用 `ALTER TABLE` 更改列的数据类型。

New Perl Server-Side Language

新版本的 plperl 服务器端语言现在支持持久共享存储区、触发器、返回记录和记录数组以及访问数据库的 SPI 调用。

Comma-separated-value (CSV) support in `COPY`

`COPY` 现在可以读写逗号分隔值文件。它还能灵活解释非标准的引号和分隔字符。

E.51.2. 迁移到版本 8.0

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

- 在 `READ COMMITTED` 序列化模式下，易变函数现在能看到并发事务提交到函数内每条语句开始处为止的结果，而不是到调用函数的交互命令开始处为止。
- 声明为 `STABLE` 或 `IMMUTABLE` 的函数总是使用调用查询的快照，因此看不到调用查询开始之后采取的动作的效果，无论在其自身事务还是其他事务中。这样的函数也必须是只读的，即它不能使用 `SELECT` 之外的任何 SQL 命令。
- 非延迟的 `AFTER` 触发器现在在触发查询完成后立即触发，而不是在当前交互命令完成时。当触发查询发生在函数内时这就有区别：触发器在函数继续其下一个操作之前被调用。
- 服务器配置参数 `virtual_host` 和 `tcpip_socket` 已被更通用的参数 `listen_addresses` 取代。此外，服务器现在默认监听 `localhost`，这在许多场景中消除了对 `-i` postmaster 开关的需要。
- 服务器配置参数 `SortMem` 和 `VacuumMem` 已改名为 `work_mem` 和 `maintenance_work_mem`，以更好地反映其用途。原名在 `SET` 和 `SHOW` 中仍受支持。
- 服务器配置参数 `log_pid`、`log_timestamp` 和 `log_source_port` 已被更通用的参数 `log_line_prefix` 取代。
- 服务器配置参数 `syslog` 已被更合逻辑的 `log_destination` 变量取代，以控制日志输出目的地。
- 服务器配置参数 `log_statement` 已更改，可以有选择地只记录数据库修改或数据定义语句。服务器配置参数 `log_duration` 现在只在 `log_statement` 打印查询时打印。
- 服务器配置参数 `max_expr_depth` 已被 `max_stack_depth` 取代，后者测量物理栈大小而不是表达式嵌套深度。这有助于防止因递归函数导致的栈溢出造成的会话终止。
- `length()` 函数不再统计 `CHAR(n)` 值中的尾随空格。
- 把整数转换为 `BIT(N)` 现在选取整数最右边的 N 位，而不是像以前那样最左边 N 位。
- 更新 NULL 数组值的元素或切片现在产生非空数组结果，即只包含被赋值位置的数组。
- 数组输入值的语法检查已大幅收紧。
  以前在奇怪位置被允许并产生奇怪结果的垃圾现在会导致错误。空字符串元素值现在必须写成

""，而不是什么也不写。还更改了数组元素周围空白的行为：尾随空白现在被忽略，与前导空白（一直被忽略）对称。

- 整数算术运算中的溢出现在被检测并报告为错误。
- 与单字节 "char" 数据类型相关的算术操作符已被移除。
- extract() 函数（也称 date_part）现在为 BC 日期返回正确的年份。此前它返回比正确年份小一的值。该函数现在还为千年和世纪返回正确的值。
- CIDR 值现在必须使其未掩码的位为零。例如，我们不再允许 204.248.199.1/31 作为 CIDR 值。这类值本就不应被 PostgreSQL 接受，现在将被拒绝。
- EXECUTE 现在返回与被执行语句匹配的完成标签。
- psql 的 \copy 命令现在读写查询的 stdin/stdout，而不是 psql 的 stdin/stdout。以前的行为可通过新的 pstdin/pstdout 参数访问。
- JDBC 客户端接口已从核心发行版中移除，现托管于http://jdbc.postgresql.org。
- Tcl 客户端接口也被移除。现在有几个 Tcl 接口托管于http://gborg.postgresql.org。
- 服务器现在使用自己的时区数据库，而不是操作系统提供的。这将在所有平台上提供一致的行为。在大多数情况下，时区行为应没有明显差异，只是 SET/SHOW TimeZone 使用的时区名可能与你的平台提供的不同。
- Configure 的线程选项不再要求用户运行测试或编辑配置文件；线程选项现在自动检测。'
- 既然表空间已经实现，initlocation 已被移除。
- 用户定义 GiST 索引的 API 已更改。Union 和 PickSplit 方法现在传递指向特殊 GistEntryVector 结构的指针，而不是 bytea。

E.51.3. 废弃特性

PostgreSQL 行为的某些方面已被确定为次优。为了向后兼容，它们在 8.0 中没有被移除，但被认为已废弃，将在下一个主版本中移除。

- 8.1 版本将移除 interval 的 to_char() 函数。
- 服务器现在对传递给 oid/float4/float8 数据类型的空字符串发出警告，但仍像以前一样把它们解释为零。在下一个主版本中，空字符串将被视为这些数据类型的无效输入。
- 默认情况下，PostgreSQL 8.0 及更早版本中的表创建时带 OID。在下一个版本中，情况将不再如此：要创建包含 OID 的表，必须指定 WITH OIDS 子句或设置 default_with_oids 配置参数。鼓励用户在表需要 OID 时显式指定 WITH OIDS，以与 PostgreSQL 的未来版本兼容。

E.51.4. 变更

下面你将找到 8.0 版本与上一个主版本之间变更的详细说明。

E.51.4.1. 性能改进

- 支持跨数据类型的索引使用（Tom）

  在此变更之前，如果数据类型不完全匹配，许多查询不会使用索引。此改进使索引使用更直观一致。

- 改进缓存的新缓冲区替换策略（Jan）

  以前的版本使用最近最少使用（LRU）缓存把最近引用的页保留在内存中。LRU 算法不考虑特定缓存条目被访问的次数，因此大表扫描可能驱逐有用的缓存页。
  新缓存算法使用四个独立的列表来跟踪最近使用和最常使用的缓存页，
  并根据负载动态优化其替换。这应能带来对共享缓冲区缓存高效得多的利用。
  过去测试过共享缓冲区大小的管理员应使用此新缓存替换策略重新测试。

- 新增周期性写脏缓冲区的子进程以减少检查点写入（Jan）

  在以前的版本中，
  每隔几分钟运行一次的检查点进程会把所有脏缓冲区写入操作系统的缓冲区缓存，
  然后把所有脏的操作系统缓冲区刷到磁盘。这导致周期性的磁盘使用尖峰，常常损害性能。
  新代码使用后台写入器以稳定的节奏细流式写磁盘，因此检查点要写到磁盘的脏页少得多。
  此外，新代码不发出全局 `sync()` 调用，而是只对自上次检查点以来写入的文件执行 `fsync()`。这应能改进性能并使检查点期间的退化最小化。

- 新增延长清理时间以降低性能影响的能力（Jan）

  在繁忙系统上，`VACUUM` 执行许多 I/O 请求，可能损害其他用户的性能。此版本允许放慢 `VACUUM` 以降低对其他用户的影响，尽管这会增加 `VACUUM` 的总时长。

- 改进重复键的 B-tree 索引性能（Dmitry Tkach、Tom）

  这改进了索引中存在许多重复值时扫描索引的方式。

- 规划时使用动态生成的表大小估计（Tom）

  此前规划器使用最后一次 `VACUUM` 或 `ANALYZE` 看到的值来估计表大小，包括物理表大小（页数）和行数。现在，当前物理表大小从内核获得，行数通过把表大小乘以最后一次 `VACUUM` 或 `ANALYZE` 看到的行密度（每页行数）来估计。在表大小自上次维护命令以来变化显著的情形下，这应能产生更可靠的估计。

- 改进 `OR` 子句的索引使用（Tom）

  这允许优化器在有许多 OR 子句的语句中使用索引，这些语句过去不会被索引。它还可以使用多列索引，其中第一列被指定而第二列是 `OR` 子句的一部分。

- 改进部分索引子句的匹配（Tom）

  服务器现在更善于在涉及复杂 `WHERE` 子句的查询中使用部分索引。

- 改进 GEQO 优化器的性能（Tom）

  GEQO 优化器用于规划涉及许多表的查询（默认十二个或更多）。此版本加快了查询分析方式，以减少花在优化上的时间。

- 其他优化器改进

  这里没有篇幅列出所做的所有小改进，但许多特殊情形比以前的版本工作得更好。

- 改进 C 函数的查找速度（Tom）

  此版本使用哈希表查找动态装载的 C 函数的信息。这改进了它们的速度，使其几乎与内建在服务器可执行文件中的函数一样快。

- 新增类型特定的 `ANALYZE` 统计能力（Mark Cave-Ayland）

此特性允许为非标准数据类型生成统计信息时有更多灵活性。

- ANALYZE 现在收集表达式索引的统计信息（Tom）

  表达式索引（也称函数索引）允许用户不仅索引列，还索引表达式和函数调用的结果。
  有了此版本，优化器可以收集和使用有关表达式索引内容的统计信息。
  这将大幅改进表达式索引相关查询的规划质量。

- ANALYZE 的新两阶段采样方法（Manfred Koizar）

  当表中不同区域有效行的密度差异很大时，这给出更好的统计信息。

- 加快 TRUNCATE（Tom）

  这挽回了一些 7.4 中观察到的性能损失，同时仍保持 TRUNCATE 事务安全。

E.51.4.2. 服务器变更

- 新增 WAL 文件归档和时间点恢复（Simon Riggs）
- 新增表空间，使管理员能控制磁盘布局（Gavin）
- 新增内建日志轮转程序（Andreas Pflug）

  现在可以方便地记录服务器消息，而不依赖 syslog 或外部日志轮转程序。

- 新增只读服务器配置参数以显示服务器编译期设置：block_size、integer_datetimes、max_function_args、max_identifier_length、max_index_keys（Joe）
- 使 sameuser、samegroup 和 all 的加引号能移除这些词在 pg_hba.conf 中的特殊含义（Andrew）
- 在默认 pg_hba.conf 中为 localhost 使用更清晰的 IPv6 名称 ::1/128（Andrew）
- 在 pg_hba.conf 示例中使用 CIDR 格式（Andrew）
- 把服务器配置参数 SortMem 和 VacuumMem 改名为 work_mem 和 maintenance_work_mem（旧名仍受支持）（Tom）

  此变更是为了澄清索引和外键创建等批量操作使用 maintenance_work_mem，而 work_mem 用于查询执行期间使用的工作区。

- 允许用服务器配置 log_disconnections 记录会话断开（Andrew）
- 新增服务器配置参数 log_line_prefix，允许控制每行日志中发出的信息（Andrew）

  可用信息包括用户名、数据库名、远程 IP 地址和会话开始时间。

- 移除服务器配置参数 log_pid、log_timestamp、log_source_port；功能被 log_line_prefix 取代（Andrew）
- 用统一的 listen_addresses 参数取代 virtual_host 和 tcpip_socket 参数（Andrew、Tom）

  virtual_host 只能指定监听单个 IP 地址。listen_addresses 允许指定多个地址。

- 默认监听 localhost，这在许多场景中消除了对 `-i` postmaster 开关的需要（Andrew）

  监听 localhost（`127.0.0.1`）不会打开新的安全漏洞，但允许 Windows 和 JDBC 等不支持本地套接字的配置无需特殊调整即可工作。

- 移除 `syslog` 服务器配置参数，并新增更合逻辑的 `log_destination` 变量来控制日志输出位置（Magnus）

- 更改服务器配置参数 `log_statement`，取值 `all`、`mod`、`ddl` 或 `none`，以选择记录哪些查询（Bruce）

  这允许管理员只记录数据定义更改或只记录数据修改语句。

- 一些与日志相关的配置参数以前可由普通用户调整，但只能向“更详细”方向。
  现在对它们的处理更严格：只有超级用户能设置。但超级用户可以用 `ALTER USER` 为非超级用户提供这些值的逐用户设置。此外，超级用户现在还可以通过 `PGOPTIONS` 设置仅超级用户配置参数的值。

- 允许把配置文件放在数据目录之外（mlw）

  默认情况下，配置文件保存在集群的顶级目录。有了此新增，
  配置文件可以放在数据目录之外，便于管理。

- 只在首次执行时规划预备查询，使常数可用于统计（Oliver Jowett）

  预备语句一次规划查询并多次执行。虽然预备查询避免了每次使用时重新规划的开销，
  但计划质量因不知道查询中将使用的确切参数而受损。在此版本中，
  未命名预备语句的规划推迟到首次执行，并把该执行的实际参数值用作优化提示。
  这允许使用外线参数传递而不带来性能损失。

- 允许 `DECLARE CURSOR` 接受参数（Oliver Jowett）

  现在在带参数的 `Parse` 消息中发出 `DECLARE CURSOR` 是有用的。`Bind` 时发送的参数值将被代入游标查询的执行。

- 修复 `inet` 和 `cidr` 数据类型的哈希连接和聚合（Tom）

  7.4 版本对混合 `inet` 和 `cidr` 值的哈希处理不正确。（此缺陷在以前的版本中不存在，因为它们不会尝试哈希这两种数据类型。）

- 让 `log_duration` 只在 `log_statement` 打印查询时打印（Ed L.）

E.51.4.3. 查询变更

- 新增保存点（嵌套事务）（Alvaro）

- 不支持的隔离级别现在被接受并提升到最接近的受支持级别（Peter）

  SQL 规范规定，如果数据库不支持特定隔离级别，应使用下一个更严格的级别。此变更遵循该建议。

- 允许 `BEGIN WORK` 像 `START TRANSACTION` 那样指定事务隔离级别（Bruce）

- 修复规则生成的查询类型与最初提交的查询不同情形下的表权限检查（Tom）

- 实现美元引用以简化单引号的使用（Andrew、Tom、David Fetter）

在以前的版本中，由于必须用单引号引用函数体，
函数文本中单引号的使用需要用两个单引号或其他易错的记法。有了此版本，
我们新增了用"美元引用"引用文本块的能力。
在不同嵌套级别使用不同引用定界符的能力大幅简化了正确引用的任务，
特别是在复杂函数中。美元引用可用于任何需要引用文本的地方。

- 让 `CASE val WHEN compval1 THEN ...` 只求值 `val` 一次（Tom）

  `CASE` 不再多次求值被测试的表达式。当表达式复杂或易变时这有好处。

- 在计算聚合查询的目标列表之前先测试 `HAVING`（Tom）

  修复诸如 `SELECT SUM(win)/SUM(lose) ... GROUP BY ... HAVING SUM(lose) > 0` 之类情形的不当失败。这本应工作，但此前可能因除以零而失败。

- 用 `max_stack_depth` 参数（以栈大小千字节计）取代 `max_expr_depth` 参数（Tom）

  这为防止因失控递归函数而崩溃提供了相当可靠的防御。我们现在直接测量执行栈的大小，
  而不是测量表达式嵌套的深度。

- 允许任意行表达式（Tom）

  此版本允许 SQL 表达式包含任意组合类型，即行值。它还允许函数更容易地接受行作为参数并返回行值。

- 允许 `LIKE`/`ILIKE` 用作行和子查询比较中的操作符（Fabien Coelho）

- 避免标识符和关键字中的基本 ASCII 字母的区域特定大小写转换（Tom）

  这解决了包含 `I` 和 `i` 的词被破坏的“土耳其问题”。7 位 ASCII 集之外字符的折叠仍是区域感知的。

- 改进语法错误报告（Fabien、Tom）

  语法错误报告比以前更有用。

- 更改 `EXECUTE`，使其返回与被执行语句匹配的完成标签（Kris Jurka）

  以前的版本对任何 `EXECUTE` 调用都返回 `EXECUTE` 标签。在此版本中，返回的标签将反映被执行的命令。

- 避免在规则列表中发出 `NATURAL CROSS JOIN`（Tom）

  这样的子句在逻辑上没有意义，但在某些情形下规则反编译器以前会产生这种语法。

E.51.4.4. 对象操作变更

- 为转换、编码转换、语言、操作符类和大对象新增 `COMMENT ON`（Christopher）

- 新增服务器配置参数 `default_with_oids`，控制表默认是否带 `OID` 创建（Neil）

  这允许管理员控制 `CREATE TABLE` 命令默认创建带或不带 `OID` 列的表。（注意：`default_with_oids` 当前的出厂默认设置为 `TRUE`，但未来版本中默认将变为 `FALSE`。）

- 为 `CREATE TABLE AS` 新增 `WITH` / `WITHOUT OIDS` 子句（Neil）

- 允许 `ALTER TABLE DROP COLUMN` 删除 `OID` 列（`ALTER TABLE SET WITHOUT OIDS` 仍有效）（Tom）
- 允许组合类型作为表列（Tom）
- 允许带默认值和 `NOT NULL` 约束的 `ALTER ... ADD COLUMN`；按 SQL 规范工作（Rod）

  现在 `ADD COLUMN` 可以创建初始不用 NULL 填充而是用指定默认值填充的列。
- 新增 `ALTER COLUMN TYPE` 来更改列的类型（Rod）

  现在可以在不删除并重新添加列的情况下更改列的数据类型。
- 允许在单条 `ALTER TABLE` 命令中有多个 `ALTER` 动作（Rod）

  这对重写表的 `ALTER` 命令（包括 `ALTER COLUMN TYPE` 和带默认值的 `ADD COLUMN`）特别有用。通过把 `ALTER` 命令分组，表只需重写一次。
- 允许 `ALTER TABLE` 添加 `SERIAL` 列（Tom）

  这是新列可指定默认值这一新能力的自然结果。
- 允许更改聚合、编码转换、数据库、函数、操作符、操作符类、模式、类型和表空间的所有者（Christopher、Euler Taveira de Oliveira）

  此前这需要直接修改系统表。
- 允许把临时对象创建限制为 `SECURITY DEFINER` 函数（Sean Chittenden）
- 新增 `ALTER TABLE ... SET WITHOUT CLUSTER`（Christopher）

  在此版本之前，除了修改系统表外没有办法清除自动聚簇规格。
- 约束/索引/`SERIAL` 名称现在为 *`table_column_type`* 并附加数字以保证在模式内唯一（Tom）

  SQL 规范规定此类名称应在模式内唯一。
- 新增 `pg_get_serial_sequence()`，返回 `SERIAL` 列的序列名（Christopher）

  这使自动脚本能可靠地找到 `SERIAL` 序列名。
- 当主/外键数据类型不匹配需要昂贵查找时发出警告
- 新 `ALTER INDEX` 命令，允许在表空间之间移动索引（Gavin）
- 让 `ALTER TABLE OWNER` 也更改依赖序列的所有权（Alvaro）

E.51.4.5. 工具命令变更

- 允许 `CREATE SCHEMA` 创建触发器、索引和序列（Neil）
- 为 `CREATE RULE` 新增 `ALSO` 关键字（Fabien Coelho）

  这允许在创建规则时添加 `ALSO`，与 `INSTEAD` 规则相对。
- 为 `LOCK` 新增 `NOWAIT` 选项（Tatsuo）

这允许 `LOCK` 命令在必须等待所请求的锁时失败。

- 允许 `COPY` 读写逗号分隔值（CSV）文件（Andrew、Bruce）
- 当 `COPY` 分隔符与 NULL 字符串冲突时报错（Bruce）
- `GRANT/REVOKE` 行为更贴近 SQL 规范
- 避免 `CREATE INDEX` 与 `CHECKPOINT` 之间的锁定冲突（Tom）

  在 7.3 和 7.4 中，长时间运行的 B-tree 索引构建可能阻塞并发 `CHECKPOINT` 完成，从而因 WAL 日志无法回收导致 WAL 膨胀。

- 数据库范围的 `ANALYZE` 不再跨表持有锁（Tom）

  这降低了与其他想独占锁定表的后端发生死锁的可能性。要获得此变更的好处，不要在事务块（`BEGIN` 块）内执行数据库范围的 `ANALYZE`；它必须能为每个表提交并开始新事务。

- `REINDEX` 不再独占锁定索引的父表

  索引本身仍被独占锁定，但如果表的读者不使用正被重建的特定索引，他们可以继续。

- 用户被重命名时擦除 MD5 用户口令（Bruce）

  PostgreSQL 用 MD5 加密口令时把用户名用作盐。当用户名被更改时，盐将不再与存储的 MD5 口令匹配，因此存储的口令变得无用。在此版本中会生成通知并清除口令。之后如果用户要能用口令登录，必须分配新口令。

- pg_ctl 新增用于 Windows 的 `kill` 选项（Andrew）

  Windows 没有向后端发送信号的 `kill` 命令，因此把此能力添加到了 pg_ctl。

- 信息模式改进
- 为 initdb 新增 `--pwfile` 选项，使初始口令可由 GUI 工具设置（Magnus）
- 在 initdb 中检测区域设置/编码不匹配（Peter）
- 为 pg_ctl 新增 `register` 命令来注册 Windows 操作系统服务（Dave Page）

E.51.4.6. 数据类型和函数变更

- 更完整的组合类型（行类型）支持（Tom）

  组合值可以用在许多以前只有标量值可用的地方。

- 把非矩形数组值作为错误拒绝（Joe）

  此前 `array_in` 会静默构建令人惊讶的结果。

- 整数算术运算中的溢出现在被检测（Tom）
- 与单字节 `"char"` 数据类型相关的算术操作符已被移除。

  此前，解析器会在许多情形下选择这些操作符，而“无法选择操作符”错误会更合适，例如 `null * null`。如果确实想对 `"char"` 列做算术，可以显式把它转换为整数。

- 数组输入值的语法检查大幅收紧（Joe）

  以前在奇怪位置被允许并产生奇怪结果的垃圾现在导致 `ERROR`，例如右花括号之后的非空白。

- 空字符串数组元素值现在必须写成 `""`，而不是什么也不写（Joe）

  此前两种写空字符串元素值的方式都被允许，但现在要求带引号的空字符串。
  完全什么都不出现的情形在未来某个版本中可能会被视为 NULL 元素值。

- 数组元素的尾随空白现在被忽略（Joe）

  此前前导空白被忽略，但元素值与分隔符或右花括号之间的尾随空白是有意义的。
  现在尾随空白也被忽略。

- 当下界不为一时以显式数组界限输出数组值（Joe）
- 接受 `YYYY-monthname-DD` 作为日期字符串（Tom）
- 让 `netmask` 和 `hostmask` 函数返回最大长度的掩码长度（Tom）
- 更改阶乘函数使其返回 `numeric`（Gavin）

  返回 `numeric` 使阶乘函数能用于更大范围的输入值。

- `to_char`/`to_date()` 日期转换改进（Kurt Roeckx、Fabien Coelho）
- 让 `length()` 无视 `CHAR(n)` 中的尾随空格（Gavin）

  此变更是为了改进一致性：尾随空格在 `CHAR(n)` 数据中语义上无关紧要，因此不应被 `length()` 统计。

- 对传递给 `OID`/`float4`/`float8` 数据类型的空字符串发出警告（Neil）

  8.1 将改为抛出错误。

- 允许 `int2`/`int4`/`int8`/`float4`/`float8` 输入例程中的前导或尾随空白（Neil）
- 更好地支持 `float4`/`float8` 中的 IEEE `Infinity` 和 `NaN` 值（Neil）

  它们现在应在所有支持 IEEE 兼容浮点运算的平台上工作。

- 为 `date_trunc()` 新增 `week` 选项（Robert Creager）
- 修复 `to_char` 对 `1 BC` 的处理（此前它返回 `1 AD`）（Bruce）
- 修复 BC 日期的 `date_part(year)`（此前它返回比正确年份小一的值）（Bruce）
- 修复 `date_part()`，使其返回正确的千年和世纪（Fabien Coelho）

  与这类事物的标准推算相比，以前的版本中世纪和千年的结果数字错误且起始年份错误。

- 为符合标准，新增 `ceiling()` 作为 `ceil()` 的别名，`power()` 作为 `pow()` 的别名（Neil）
- 更改 `ln()`、`log()`、`power()` 和 `sqrt()`，使其按 SQL:2003 的规定在某些错误条件下发出正确的 `SQLSTATE` 错误码（Neil）
- 新增 SQL:2003 定义的 `width_bucket()` 函数（Neil）
- 新增 `generate_series()` 函数以简化数字集合的处理（Joe）

- 修复 `upper/lower/initcap()` 函数使其支持多字节编码（Tom）
- 新增布尔和按位整数 `AND/OR` 聚合（Fabien Coelho）
- 新增返回客户端和服务器网络地址的会话信息函数（Sean Chittenden）
- 新增确定闭合路径面积的函数（Sean Chittenden）
- 新增向其他后端发送取消请求的函数（Magnus）
- 新增 `interval` 加 `datetime` 操作符（Tom）

  相反的顺序（`datetime` 加 `interval`）此前已支持，但两者都是 SQL 标准要求的。
- 把整数转换为 `BIT(N)` 选取整数最右边的 N 位（Tom）

  在以前的版本中选取的是最左边 N 位，但这被认为没有用处，更不用说与从 bit 到 int 的转换不一致了。
- 要求 `CIDR` 值的所有未掩码位为零（Kevin Brintnall）

E.51.4.7. 服务器端语言变更

- 在 `READ COMMITTED` 序列化模式下，易变函数现在能看到并发事务提交到函数内每条语句开始处为止的结果，而不是到调用函数的交互命令开始处为止。
- 声明为 `STABLE` 或 `IMMUTABLE` 的函数总是使用调用查询的快照，因此看不到调用查询开始之后采取动作的效果，无论在其自身事务还是其他事务中。这样的函数也必须是只读的，即不能使用 `SELECT` 之外的任何 SQL 命令。把函数声明为 `STABLE` 或 `IMMUTABLE` 而不是 `VOLATILE` 有可观的性能收益。
- 非延迟的 `AFTER` 触发器现在在触发查询完成后立即触发，而不是在当前交互命令完成时。当触发查询发生在函数内时这就有区别：触发器在函数继续其下一个操作之前被调用。例如，如果函数向表插入新行，任何非延迟外键检查都在函数继续之前进行。
- 允许函数参数带名称声明（Dennis Björklund）

  这允许对函数做更好的文档记录。名称是否真起作用取决于所用的具体函数语言。
- 允许在函数中引用 PL/pgSQL 参数名（Dennis Björklund）

  这基本上为每个命名参数创建了自动别名。
- 在创建时对 PL/pgSQL 函数做最简语法检查（Tom）

  这使我们能更早捕获简单的语法错误。
- PL/pgSQL 中对组合类型（行和记录变量）的更多支持

  例如，现在可以把行类型变量作为单个变量传递给另一个函数。
- PL/pgSQL 变量的默认值现在可以引用先前声明的变量
- 改进 PL/pgSQL FOR 循环的解析（Tom）

  解析现在由 `".."` 的存在驱动，而不是 `FOR` 变量的数据类型。这对正确的函数没有区别，但在出错时应产生更易理解的错误消息。

- PL/Perl 服务器端语言的重大改造（Command Prompt、Andrew Dunstan）
- 在 PL/Tcl 中，SPI 命令现在在子事务中运行。如果发生错误，子事务被清理，错误被报告为普通 Tcl 错误，可以用 `catch` 捕获。此前无法捕获此类错误。
- PL/pgSQL 接受 `ELSEIF`（Neil）

  此前 PL/pgSQL 只允许 `ELSIF`，但许多人习惯把此关键字拼写为 `ELSEIF`。

E.51.4.8. psql 变更

- 改进 psql 对数据库对象的信息显示（Christopher）
- 允许 psql 在 `\du` 和 `\dg` 中显示组成员身份（Markus Bertheau）
- 防止 psql 的 `\dn` 显示临时模式（Bruce）
- 允许 psql 处理文件名的波浪号用户展开（Zach Irmen）
- 允许 psql 通过 readline 显示精致提示符，包括颜色（Reece Hart、Chet Ramey）
- 让 psql 的 `\copy` 完全匹配 `COPY` 命令语法（Tom）
- 显示语法错误的位置（Fabien Coelho、Tom）
- 向 psql 的 `\d` 显示中添加 `CLUSTER` 信息（Bruce）
- 更改 psql 的 `\copy stdin/stdout`，使其从命令输入/输出读取（Bruce）
- 新增 `pstdin`/`pstdout`，从 psql 的 `stdin`/`stdout` 读取（Mark Feit）
- 新增全局 psql 配置文件 `psqlrc.sample`（Bruce）

  这允许用一个中心文件存储全局 psql 启动命令。
- 让 psql 的 `\d+` 指示表是否有 `OID` 列（Neil）
- 在 Windows 上，psql 读取文件时使用二进制模式，使 control-Z 不被视为文件结束
- 让 `\dn+` 显示模式的权限和描述（Dennis Björklund）
- 改进 Tab 补全支持（Stefan Kaltenbrunn、Greg Sabino Mullane）
- 允许用大写或小写设置布尔设置（Michael Paesold）

E.51.4.9. pg_dump 变更

- 使用依赖信息改进 pg_dump 的可靠性（Tom）

  这应能解决相关对象有时以错误顺序转储的长期问题。
- 让 pg_dump 尽可能按字母顺序输出对象（Tom）

  这应使识别转储文件之间的更改更容易。
- 允许 pg_restore 忽略某些 SQL 错误（Fabien Coelho）

这使 pg_restore 的行为类似于把 pg_dump 输出脚本送入 psql 的结果。在大多数情况下，忽略错误继续前进是最有用的做法。还添加了 pg_restore 选项以提供出错即退出的旧行为。

- pg_restore 的 `-l` 显示现在包含对象的模式名
- pg_dump 文本输出中新增开始/结束标记（Bruce）
- 为详细模式下的 pg_dump/pg_dumpall 新增开始/停止时间（Bruce）
- 允许在 pg_dumpall 中使用大多数 pg_dump 选项（Christopher）
- 让 pg_dump 默认使用 `ALTER OWNER` 而不是 `SET SESSION AUTHORIZATION`（Christopher）

E.51.4.10. libpq 变更

- 使 libpq 的 `SIGPIPE` 处理线程安全（Bruce）
- 新增 `PQmbdsplen()`，返回字符的显示长度（Tatsuo）
- 为 SSL 和 Kerberos 连接添加线程锁定（Manfred Spraul）
- 允许 `PQoidValue()`、`PQcmdTuples()` 和 `PQoidStatus()` 用于 `EXECUTE` 命令（Neil）
- 新增 `PQserverVersion()`，提供对服务器版本号更便捷的访问（Greg Sabino Mullane）
- 新增 `PQprepare/PQsendPrepared()` 函数，支持预备语句而不必指定其参数的数据类型（Abhijit Menon-Sen）
- 许多 ECPG 改进，包括 `SET DESCRIPTOR`（Michael）

E.51.4.11. 源代码变更

- 允许数据库服务器原生运行于 Windows（Claudio、Magnus、Andrew）
- 把 shell 脚本命令转换为 C 版本以支持 Windows（Andrew）
- 创建扩展 makefile 框架（Fabien Coelho、Peter）

  这简化了在原始源代码树之外构建扩展的任务。

- 支持可重新定位的安装（Bruce）

  已安装文件（如 `/share` 目录）的目录路径现在相对于可执行文件的实际位置计算，因此安装树可以移动到其他地方而无需重新配置和重新构建。

- 用 `--with-docdir` 选择文档的安装位置；还允许 `--infodir`（Peter）
- 新增 `--without-docdir` 以阻止安装文档（Peter）
- 升级到 DocBook V4.2 SGML（Peter）
- 新的 `PostgreSQL` CVS 标签（Marc）

  这是为了使组织更容易管理自己的 PostgreSQL CVS 仓库副本。来自主仓库的文件版本戳不会因向复制的仓库检入或从中检出而被破坏。

- 澄清锁定代码（Manfred Koizar）
- 缓冲区管理器清理（Neil）
- 把平台测试与 CPU 自旋锁代码解耦（Bruce、Tom）
- 为 gcc 在 PA-RISC 上新增内联 test-and-set 代码（ViSolve、Tom）
- 改进 i386 自旋锁代码（Manfred Spraul）
- 清理自旋锁汇编代码以避免较新 gcc 版本的警告（Tom）
- 从源代码树移除 JDBC；现为独立项目
- 移除 libpgtcl 客户端接口；现为独立项目
- 更准确地估计内存和文件描述符的使用（Tom）
- Mac OS X 启动脚本的改进（Ray A.）
- 新的 `fsync()` 测试程序（Bruce）
- 重大文档改进（Neil、Peter）
- 移除 pg_encoding；不再需要
- 移除 pg_id；不再需要
- 移除 initlocation；不再需要
- 自动检测线程标志（不再需要手工测试）（Bruce）
- 使用 Olson 的公有领域 timezone 库（Magnus）
- 启用线程时，Unixware 上的后端可执行文件也使用线程标志（Bruce）

  Unixware 不能在同一可执行文件中混用线程化和非线程化目标文件，因此所有内容都必须编译为线程化。
- psql 现在使用 flex 生成的词法分析器处理命令字符串
- 重新实现后端各处使用的链表数据结构（Neil）

  这通过使列表追加和长度操作更高效来改进性能。
- 允许动态装载的模块创建自己的服务器配置参数（Thomas Hallgren）
- 新的巴西葡萄牙语版 FAQ（Euler Taveira de Oliveira）
- 新增法语 FAQ（Guillaume Lelarge）
- 新的用于 Windows 日志的 pgevent
- 使 libpq 和 ECPG 在 OS X 上构建为正确的共享库（Tom）

E.51.4.12. 附加模块变更

- `contrib/dblink` 的彻底改造（Joe）
- `contrib/dbmirror` 改进（Steven Singer）

- 新增 `contrib/xml2`（John Gray、Torchbox）
- 更新了 `contrib/mysql`
- 新版本的 `contrib/btree_gist`（Teodor）
- 新的 `contrib/trgm`，PostgreSQL 的三字母组匹配（Teodor）
- 许多 `contrib/tsearch2` 改进（Teodor）
- 为 `contrib/fuzzystrmatch` 添加双变音位（Andrew）
- 允许 `contrib/pg_autovacuum` 作为 Windows 服务运行（Dave Page）
- 为 `contrib/dbsize` 新增函数（Andreas Pflug）
- 移除 `contrib/pg_logger`：被集成日志子进程取代
- 移除 `contrib/rserv`：被多个独立项目取代

E.52. 版本 7.4.30

发行日期
2010-10-04

本次发布包含对 7.4.29 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

这预计是 PostgreSQL 7.4.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.52.1. 迁移到版本 7.4.30

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅 7.4.26 的发布说明。

E.52.2. 变更

- 在 PL/Perl 和 PL/Tcl 中为每个调用方 SQL 用户 ID 使用单独的解释器（Tom Lane）

  此变更防止了可通过破坏稍后将在同一会话中以另一 SQL 用户身份执行的 Perl 或 Tcl 代码（例如在 `SECURITY DEFINER` 函数内）而造成的安全问题。大多数脚本语言都提供许多实现这一点的手段，例如重新定义目标函数调用的标准函数或操作符。如果没有此变更，任何拥有 Perl 或 Tcl 语言使用权的 SQL 用户基本上都能以目标函数所有者的 SQL 权限为所欲为。

  此变更的代价是 Perl 和 Tcl 函数之间的有意通信变得更困难。为提供出路，PL/PerlU 和 PL/TclU 函数继续每会话只使用一个解释器。这不被视为安全问题，因为所有这类函数本来就以数据库超级用户的信任级别执行。

  声称提供可信执行的第三方过程语言很可能存在类似的安全问题。
  我们建议就安全关键用途联系你所依赖的任何 PL 的作者。

  感谢 Tim Bunce 指出此问题（CVE-2010-3433）。

- 禁止以并非它预期配合使用的系统目录列的参数调用 `pg_get_expr()`，以防止可能的崩溃（Heikki Linnakangas、Tom Lane）
- 修复“无法处理未规划的子查询”错误（Tom Lane）

  当子查询包含一个连接别名引用，而该引用展开成包含另一个子查询的表达式时，会出现此错误。
- 写入锁文件（`postmaster.pid` 和套接字锁文件两者）内容时注意对其执行 fsync（Tom Lane）

  如果机器在 postmaster 启动后不久崩溃，这一疏漏可能导致锁文件内容损坏。这又可能使后续启动 postmaster 的尝试失败，直到手工删除锁文件。
- 改进 `contrib/dblink` 对包含已删除列的表的处理（Tom Lane）
- 修复 `contrib/dblink` 中“重复连接名”错误后的连接泄漏（Itagaki Takahiro）
- 更新构建基础设施和文档，反映源代码仓库从 CVS 迁移到 Git 的变化（Magnus Hagander 等）

E.53. 版本 7.4.29

发行日期
2010-05-17

本次发布包含对 7.4.28 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 7.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.53.1. 迁移到版本 7.4.29

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅 7.4.26 的发布说明。

E.53.2. 变更

- 用应用于整个解释器的操作码掩码来强制 `plperl` 中的限制，而不是使用 `Safe.pm`（Tim Bunce、Andrew Dunstan）

  近期的事态发展使我们确信 `Safe.pm` 太不安全，不能依靠它使 `plperl` 可信。此变更完全移除了对 `Safe.pm` 的使用，改为使用始终应用操作码掩码的独立解释器。此变更的良好副作用包括：现在可以在 `plperl` 中以自然的方式使用 Perl 的 `strict` 编译指令，Perl 的 `$a` 和 `$b` 变量在排序例程中按预期工作，以及函数编译显著加快。（CVE-2010-1169）
- 防止 PL/Tcl 执行来自 `pltcl_modules` 的不可信代码（Tom）

  PL/Tcl 从数据库表自动装载 Tcl 代码的特性可能被利用于特洛伊木马攻击，因为对谁能创建该表或向其中插入没有限制。此变更禁用该特性，除非 `pltcl_modules` 由超级用户拥有。（但是，不检查该表上的权限，

因此确实需要较不安全的模块表的安装仍可以向受信任的非超级用户授予适当权限。）此外，除非确实要执行 `pltclu` 函数，否则防止把代码装载到无限制的“普通”Tcl 解释器中。（CVE-2010-1170）

- 不允许非特权用户重置仅超级用户可用的参数设置（Alvaro）

  此前，如果非特权用户为自己运行 `ALTER USER ... RESET ALL`，或为他拥有的数据库运行 `ALTER DATABASE ... RESET ALL`，这将移除该用户或数据库的所有特殊参数设置，包括本应只可由超级用户更改的那些。现在，`ALTER` 只移除用户有权更改的参数。

- 避免后端关闭发生在需要向日志条目添加 `CONTEXT` 内容时导致可能的崩溃（Tom）

  某些情况下上下文打印函数会失败，因为到打印日志消息时当前事务已被回滚。

- 为现代 Perl 版本更新 pl/perl 的 `ppport.h`（Andrew）

- 修复 pl/python 中的若干内存泄漏（Andreas Freund、Tom）

- 确保 `contrib/pgstattuple` 函数及时响应取消中断（Tatsuhito Kasahara）

- 让服务器启动正确处理 `shmget()` 对已存在的共享内存段返回 `EINVAL` 的情形（Tom）

  此行为曾在包括 OS X 在内的 BSD 衍生内核上观察到。它导致完全误导性的启动失败，抱怨共享内存请求大小过大。

E.54. 版本 7.4.28

发行日期
2010-03-15

本次发布包含对 7.4.27 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

PostgreSQL 社区将于 2010 年 7 月停止对 7.4.X 版本系列的更新。建议用户尽快更新到较新的版本分支。

E.54.1. 迁移到版本 7.4.28

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅 7.4.26 的发布说明。

E.54.2. 变更

- 新增配置参数 `ssl_renegotiation_limit`，控制对 SSL 连接进行会话密钥重协商的频率（Magnus）

  可以把它设为零以完全禁用重协商，这在使用有缺陷的 SSL 库时可能是必需的。特别是，某些厂商发布的 CVE-2009-3555 临时补丁会导致重协商尝试失败。

- 让 `bit` 类型的 `substring()` 把任何负长度都视为“字符串的其余全部”（Tom）

  以前的编码只对 `-1` 这样处理，对其他负值会产生无效的结果值，可能导致崩溃（CVE-2010-0442）。

- 修复若干病态缓慢的正则表达式匹配情形（Tom）
- 读取 `pg_hba.conf` 及相关文件时，如果 `@` 出现在引号内，不要把 `@something` 当作文件包含请求；另外，永远不要把单独的 `@` 当作文件包含请求（Tom）

  这防止了角色名或数据库名以 `@` 开头时的古怪行为。如果需要包含路径名包含空格的文件，仍然可以做到，但必须写 `@"/path to/file"` 而不是把引号放在整个构造之外。
- 防止在某些平台上把目录命名为 `pg_hba.conf` 及相关文件中的包含目标时出现无限循环（Tom）
- 确保 PL/Tcl 完全初始化 Tcl 解释器（Tom）

  此疏漏唯一已知的症状是使用 Tcl 8.5 或更高版本时 Tcl 的 `clock` 命令行为异常。
- 防止在向 `dblink_build_sql_*` 函数指定过多键列时 `contrib/dblink` 崩溃（Rushabh Lathia、Joe Conway）

E.55. 版本 7.4.27

发行日期
2009-12-14

本次发布包含对 7.4.26 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.55.1. 迁移到版本 7.4.27

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.26 的版本升级，请参阅 7.4.26 的发布说明。

E.55.2. 变更

- 防范索引函数更改会话局部状态导致的间接安全威胁（Gurjeet Singh、Tom）

  此变更防止据称不可变的索引函数可能破坏超级用户的会话（CVE-2009-4136）。
- 拒绝在公共名（CN）字段中包含嵌入空字节的 SSL 证书（Magnus）

  这防止在 SSL 验证期间证书与服务器或客户端名称的非预期匹配（CVE-2009-4034）。
- 修复后端启动时缓存初始化期间可能的崩溃（Tom）
- 防止信号在不安全的时刻中断 `VACUUM`（Alvaro）

  此修复防止了 `VACUUM FULL` 在已提交其元组移动后被取消时的 PANIC，以及普通 `VACUUM` 在截断表之后被中断时的瞬态错误。
- 修复哈希表大小计算中因整数溢出而可能的崩溃（Tom）

  这可能发生在规划器对哈希连接结果大小的估计极大时。
- 修复 `inet/cidr` 比较中非常罕见的崩溃（Chris Mikkelson）

- 修复 PAM 口令处理使其更健壮（Tom）

  已知以前的代码在 Linux `pam_krb5` PAM 模块与 Microsoft Active Directory 域控制器组合下失败。它在其他地方可能也有问题，因为它对 PAM 栈会传递给它什么参数做了无根据的假设。

- 让 postmaster 忽略连接请求包中的任何 `application_name` 参数，以改进与未来 libpq 版本的兼容性（Tom）

E.56. 版本 7.4.26

发行日期
2009-09-09

本次发布包含对 7.4.25 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.56.1. 迁移到版本 7.4.26

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你在 `interval` 列上有任何哈希索引，必须在更新到 7.4.26 后对它们执行 `REINDEX`。另外，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.56.2. 变更

- 不允许在安全定义者函数内使用 `RESET ROLE` 和 `RESET SESSION AUTHORIZATION`（Tom、Heikki）

  这覆盖了先前禁止在安全定义者函数内使用 `SET ROLE` 和 `SET SESSION AUTHORIZATION` 的补丁中遗漏的一种情形。（参见 CVE-2007-6600）

- 修复出现在外层聚合函数参数中的子 SELECT 的处理（Tom）

- 修复 `interval` 数据类型的哈希计算（Tom）

  这纠正了 interval 值上哈希连接的错误结果。它还更改了 interval 列上哈希索引的内容。如果你有任何这样的索引，必须在更新后对它们执行 `REINDEX`。

- 修复 *x* 超过 200 万且使用整数日期时间时 `INTERVAL 'x ms'` 的溢出（Alex Hunsaker）

- 修复点与线段之间距离的计算（Tom）

  这导致多个几何操作符的结果不正确。

- 修复 `money` 数据类型，使其在货币金额没有小数位的区域设置（例如日本）中工作（Itagaki Takahiro）

- 正确地对像 `00:12:57.999999999999999999999999999` 这样的日期时间输入四舍五入（Tom）

- 修复 GiST R-tree 操作符类中页分裂点选择不佳的问题（Teodor）

- 修复 plperl 初始化中的可移植性问题（Andrew Dunstan）

- 改进 libpq 从 `COPY FROM STDIN` 期间错误中恢复的代码的健壮性（Tom）
- 避免在两个库都安装时包含相互冲突的 readline 和 editline 头文件（Zdenek Kotala）

E.57. 版本 7.4.25

发行日期
2009-03-16

本次发布包含对 7.4.24 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.57.1. 迁移到版本 7.4.25

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.57.2. 变更

- 防止编码转换失败时的错误递归崩溃（Tom）

  此变更扩展了上两个小版本中针对相关失败情形所做的修复。
  以前的修复是针对最初的问题报告量身定制的，但我们现在认识到，编码转换函数抛出的
  任何错误都可能在试图报告该错误时导致无限递归。因此，
  解决方案是：如果发现已处于递归错误报告情形，就禁用翻译和编码转换并报告错误消息的纯 ASCII 形式。（CVE-2009-0922）

- 不允许 `CREATE CONVERSION` 为指定的转换函数使用错误的编码（Heikki）

  这防止了编码转换失败的一种可能情形。前一个变更是兜底，防范同一领域其他类型的失败。

- 修复 `to_char()` 被给定与数据参数类型不适当的格式代码时的核心转储（Tom）
- 把 `MUST`（毛里求斯岛夏令时）加入已知时区缩写的默认列表（Xavier Bugaud）

E.58. 版本 7.4.24

发行日期
2009-02-02

本次发布包含对 7.4.23 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.58.1. 迁移到版本 7.4.24

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.58.2. 变更

- 改进 `headline()` 函数中对 URL 的处理（Teodor）
- 改进 `headline()` 函数对超长标题的处理（Teodor）
- 防止在为指定的编码对创建了错误转换函数的编码转换时可能的断言失败或错误转换（Tom、Heikki）
- 避免 `VACUUM` 中对小表的不必要锁定（Heikki）
- 修复 `contrib/tsearch2` 的 `get_covers()` 函数中的未初始化变量（Teodor）
- 修复 `to_char()` 处理 `TH` 格式代码中的缺陷（Andreas Scherbaum）
- 让所有文档在适当之处引用 `pgsql-bugs` 和/或 `pgsql-hackers`，而不是现已停用的 `pgsql-ports` 和 `pgsql-patches` 邮件列表（Tom）

E.59. 版本 7.4.23

发行日期
2008-11-03

本次发布包含对 7.4.22 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.59.1. 迁移到版本 7.4.23

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.59.2. 变更

- 修复客户端编码无法表示本地化错误消息时的后端崩溃（Tom）

  我们此前处理过类似问题，但如果“字符没有等价物”消息本身无法转换，仍会失败。
  修复方法是在检测到这种情形时禁用本地化并发送纯 ASCII 错误消息。

- 修复单个查询项匹配文本首词时 tsearch2 标题生成不正确的问题（Sushant Sinha）
- 修复 `--enable-integer-datetimes` 构建中使用非 ISO datestyle 时 interval 值小数秒显示不当的问题（Ron Mayer）
- 确保 `SPI_getvalue` 和 `SPI_getbinval` 在传入的元组和元组描述符列数不同时行为正确（Tom）

  当表添加或删除过列时这种情况是正常的，但这两个函数没有正确处理它。
  唯一可能的后果是错误的错误指示。

- 修复 ecpg 对 `CREATE USER` 的解析（Michael）

E.60. 版本 7.4.22

发行日期
2008-09-22

本次发布包含对 7.4.21 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.60.1. 迁移到版本 7.4.22

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.60.2. 变更

- 修复日期时间输入函数在 64 位平台上正确检测整数溢出（Tom）
- 改进向 syslog 写很长日志消息的性能（Tom）
- 修复 `SELECT DISTINCT ON` 查询上游标反向扫描的缺陷（Tom）
- 修复规划器，使其估计产生布尔结果的 `GROUP BY` 表达式总是产生两个组，而不考虑表达式的内容（Tom）

  对于某些布尔测试（如 *col* `IS NULL`），这比常规的 `GROUP BY` 估计准确得多。
- 改进 pg_dump 和 pg_restore 在发送 SQL 命令失败后的错误报告（Tom）

E.61. 版本 7.4.21

发行日期
2008-06-12

此发行版在 7.4.20 的基础上包含一个严重的缺陷修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.61.1. 迁移到版本 7.4.21

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.61.2. 变更

- 让 `pg_get_ruledef()` 给负常数加括号（Tom）

  在此修复之前，视图或规则中的负常数可能被转储成如 `-42::integer`，这有微妙的错误：由于操作符优先级规则，它应该是 `(-42)::integer`。通常这影响不大，但它可能与另一个最

近的补丁相互作用，导致 PostgreSQL 拒绝原本有效的 `SELECT DISTINCT` 视图查询。由于这可能导致 pg_dump 输出无法重载，它被当作高优先级修复。转储输出确实不正确的已发布版本只有 8.3.1 和 8.2.7。

E.62. 版本 7.4.20

发行日期

从未发布

本次发布包含对 7.4.19 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.62.1. 迁移到版本 7.4.20

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.62.2. 变更

- 修复 ISO-8859-5 与其他编码之间的转换对西里尔“Yo”字符（带两个点的 `e` 和 `E`）的处理（Sergey Burladyan）

- 修复几个数据类型输入函数，
  它们允许结果中未使用的字节包含未初始化的不可预测值（Tom）

  这可能导致两个看起来相同的字面值不被视为相等的失败，导致解析器抱怨 `ORDER BY` 和 `DISTINCT` 表达式不匹配。

- 修复正则表达式子串匹配（`substring(`*`string`*` from `*`pattern`*`)`）中的一个极端情形（Tom）

  问题发生在模式整体有匹配但用户指定了带括号的子表达式而该子表达式没有匹配时。例子是 `substring('foo' from 'foo(bar)?')`。由于 `(bar)` 未匹配，它应返回 NULL，但它错误地返回了整个模式的匹配（即 `foo`）。

- 修复 ecpg 的 `PGTYPEStimestamp_sub()` 函数结果不正确的问题（Michael）

- 修复 `DatumGetBool` 宏，使其在 gcc 4.3 下不失败（Tom）

  此问题影响返回布尔的“旧式”（V0）C 函数。修复已在 8.3 中，但当时没有意识到需要向后移植。

- 修复长期存在的 `LISTEN/NOTIFY` 竞态条件（Tom）

  极少数情况下，刚执行过 `LISTEN` 的会话可能收不到通知，即使按理应当收到，因为观察到执行 `NOTIFY` 的并发事务稍后提交了。

  此修复的一个副作用是：执行了尚未提交的 `LISTEN` 命令的事务如果去查看，将看不到该 `LISTEN` 在 `pg_listener` 中的行；以前能看到。这一行为从未被明确记录，但有些应用程序可能依赖旧行为。

- 修复 `ORDER BY` 和 `GROUP BY` 中常量表达式的显示（Tom）

显式转换的常量会被错误显示。这例如可能导致视图定义在转储和重载期间损坏。

- 修复 libpq，使其在 COPY OUT 期间正确处理 NOTICE 消息（Tom）

  只观察到此失败在用户定义数据类型的输出例程发出 NOTICE 时发生，但不能保证它不会因其他原因发生。

E.63. 版本 7.4.19

发行日期
2008-01-07

此发行版包含 7.4.18 之后的多种修复，包括对重要安全问题的修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.63.1. 迁移到版本 7.4.19

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.63.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 修复正则表达式包中的若干缺陷（Tom、Will Drewry）

  精心构造的正则表达式模式可能导致崩溃、无限或接近无限的循环和/或大量内存消耗，这些都对接受来自不可信来源的正则搜索模式的应用程序构成拒绝服务风险。（CVE-2007-4769、CVE-2007-4772、CVE-2007-6067）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  7.4.18 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 修复某些 `WHERE false AND var IN (SELECT ...)` 情形下的规划器失败（Tom）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）

- 修复 PL/Python 在长异常消息上不崩溃（Alvaro）
- ecpg 解析器修复（Michael）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 修复 `tsvector` 和 `tsquery` 输出例程以正确转义反斜杠（Teodor、Bruce）
- 修复 `to_tsvector()` 在超长输入字符串上的崩溃（Teodor）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.64. 版本 7.4.18

发行日期
2007-09-17

此发行版包含 7.4.17 之后的修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.64.1. 迁移到版本 7.4.18

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.64.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 SSL 错误消息的过度记录（Tom）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 防止 `CLUSTER` 因试图处理其他会话的临时表而失败（Alvaro）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.65. 版本 7.4.17

发行日期
2007-04-23

此发行版包含 7.4.16 之后的修复，包括一个安全修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.65.1. 迁移到版本 7.4.17

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.65.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- `/contrib/tsearch2` 崩溃修复（Teodor）

- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）

- 修复哈希索引扩容期间的 PANIC（7.4.15 中引入的缺陷）（Tom）

E.66. 版本 7.4.16

发行日期
2007-02-05

此发行版包含 7.4.15 之后的多种修复，包括一个安全修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.66.1. 迁移到版本 7.4.16

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.66.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  该漏洞涉及抑制 SQL 函数返回其声明数据类型的常规检查，或更改 SQL 函数中使用的表列的数据类型（CVE-2007-0555）。此错误很容易被利用来导致后端崩溃，
  而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 修复由 `UNION` 触发的罕见 Assert() 崩溃（Tom）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.67. 版本 7.4.15

发行日期
2007-01-08

本次发布包含对 7.4.14 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.67.1. 迁移到版本 7.4.15

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.67.2. 变更

- 改进 AIX 上 `getaddrinfo()` 的处理（Tom）

  这修复了启动统计收集器等问题。

- 修复 `VACUUM` 中的“未能重新找到父键”错误（Tom）
- 修复影响多 GB 哈希索引的缺陷（Tom）
- 修复构造由多个空元素组成的 `ARRAY[]` 时的错误（Tom）
- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。

- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。

E.68. 版本 7.4.14

发行日期
2006-10-16

本次发布包含对 7.4.13 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.68.1. 迁移到版本 7.4.14

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.68.2. 变更

- 修复把无类型字面值当作 ANYARRAY 时的核心转储

- 修复 `string_to_array()`，处理分隔符字符串的重叠匹配

  例如 `string_to_array('123xx456xxx789', 'xx')`。

- 修复 psql 的 `\d` 命令模式匹配的极端情形

- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）

- 修复 /contrib/dbmirror 中的反斜杠转义

- 针对美国 DST 法规的最近变更调整回归测试

E.69. 版本 7.4.13

发行日期
2006-05-23

此发行版包含 7.4.12 之后的多种修复，包括针对极其严重安全问题的补丁。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.69.1. 迁移到版本 7.4.13

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.69.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题和 `standard_conforming_strings`

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题，并使它们为将来转向 SQL 标准字符串字面量语法做好准备。并发使用多个 PostgreSQL 连接的应用

程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）
- 修复有时导致 OR 索引扫描漏掉本应返回的行的缺陷
- 修复 btree 索引被截断情形的 WAL 重放
- 修复涉及 `|` 的模式的 `SIMILAR TO`（Tom）
- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）
- 修复 Intel Mac 上的 Bonjour（Ashley Clark）
- 修复各种小的内存泄漏

E.70. 版本 7.4.12

发行日期
2006-02-14

本次发布包含对 7.4.11 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.70.1. 迁移到版本 7.4.12

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.11 的版本升级，请参阅 7.4.11 的发布说明。

E.70.2. 变更

- 修复 `SET SESSION AUTHORIZATION` 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 7.4.9 和 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 为预备语句中的 `UNKNOWN` 参数正确检查 `DOMAIN` 约束（Neil）
- 修复以允许恢复有对自定义操作符跨模式引用的转储（Tom）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.71. 版本 7.4.11

发行日期
2006-01-09

本次发布包含对 7.4.10 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.71.1. 迁移到版本 7.4.11

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅 7.4.8 的发布说明。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.71.2. 变更

- 修复在事务之外或失败事务中发出协议级 Describe 消息的问题（Tom）
- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。
- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。
- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）
- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。
- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.72. 版本 7.4.10

发行日期
2005-12-12

本次发布包含对 7.4.9 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.72.1. 迁移到版本 7.4.10

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅 7.4.8 的发布说明。

E.72.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- 防止客户端在当前事务已中止时发送 Bind 协议消息导致的失败
- `/contrib/ltree` 修复（Teodor）
- AIX 和 HPUX 编译修复（Tom）
- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。

- 防止表被删除后 pg_autovacuum 的核心转储

E.73. 版本 7.4.9

发行日期
2005-10-04

本次发布包含对 7.4.8 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.73.1. 迁移到版本 7.4.9

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.4.8 的版本升级，请参阅 7.4.8 的发布说明。

E.73.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。

- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

  在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。

- 修正 `COPY` 中只读事务测试的方向

  以前的代码禁止 `COPY TO`，而它应禁止的是 `COPY FROM`。

- 修复只引用内侧关系的外连接 ON 子句的规划问题
- 对 `x FULL JOIN y ON true` 极端情形的进一步修复
- 让 `array_in` 和 `array_recv` 更严格地验证其 OID 参数
- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题

- 改进日期时间解析的健壮性
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- postmaster 启动期间不要试图打开超过 `max_files_per_process` 个文件
- 各种内存泄漏修复
- 各种可移植性改进
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`
- 更新 `contrib/tsearch2` 以使用当前的 Snowball 代码

E.74. 版本 7.4.8

发行日期
2005-05-09

此发行版包含 7.4.7 之后的多种修复，包括几个与安全相关的问题。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.74.1. 迁移到版本 7.4.8

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，它是处理在 7.4.X 系统目录初始内容中发现的两个重要安全问题的可能方法之一。用 7.4.8 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

较大的安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）

较小的问题是 `contrib/tsearch2` 模块创建了几个被错误声明为返回 `internal` 的函数，而它们并不接受 `internal` 参数。这破坏了所有使用 `internal` 参数的函数的类型安全。

强烈建议所有安装通过 initdb 或按照下述手工修复程序修复这些错误。这些错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行以下程序。作为数据库超级用户，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[3] = 'internal'::regtype
WHERE pronamespace = 11 AND pronargs = 5
     AND proargtypes[2] = 'cstring'::regtype;
-- The command should report having updated 90 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

然后，如果你安装了 `contrib/tsearch2`，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[0] = 'internal'::regtype
WHERE oid IN (
   'dex_init(text)'::regprocedure,
   'snb_en_init(text)'::regprocedure,
   'snb_ru_init(text)'::regprocedure,
   'spell_init(text)'::regprocedure,
   'syn_init(text)'::regprocedure
);
-- The command should report having updated 5 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

如果此命令失败并显示类似 “function "dex_init(text)" does not exist” 的消息，则要么该数据库中未安装 `tsearch2`，要么你已经做过此更新。

上述程序必须在安装的每个数据库中执行，包括 `template1`，理想情况下也包括 `template0`。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。`template1` 可以像任何其他数据库一样修复，但修复 `template0` 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 `template0` 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.74.2. 变更

- 更改编码函数签名以防止误用
- 更改 `contrib/tsearch2` 以避免不安全地使用 `INTERNAL` 函数结果
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。

- 确保后端关闭期间完成的操作被统计收集器计数

  这有望解决关于 pg_autovacuum 不够频繁地清理系统目录的报告——它没有被通知后端退出期间临时表删除导致的目录删除。

- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 防止 `COALESCE(NULL,NULL)` 的崩溃
- 修复 `array_map`，使其正确调用 PL 函数
- 修复 `ALTER DATABASE RENAME` 中的权限检查
- 修复 `ALTER LANGUAGE RENAME`
- 让 `RemoveFromWaitQueue` 清理自身

  这修复了一个锁管理错误，它只在事务被踢出锁等待（通常由查询取消）且锁持有者随后在极窄的窗口内释放锁时可见。

- 修复 `INSERT ... SELECT` 中出现无类型参数的问题
- 修复 `ALTER TABLE SET WITHOUT OIDS` 之后 `CLUSTER` 的失败

E.75. 版本 7.4.7

发行日期
2005-01-31

此发行版包含 7.4.6 之后的多种修复，包括几个与安全相关的问题。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.75.1. 迁移到版本 7.4.7

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.75.2. 变更

- 不允许非超级用户使用 `LOAD`

在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。

- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 plperl 处理元组字段中引号的问题
- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示
- 让 age(timestamptz) 用本地时区而不是 GMT 计算

E.76. 版本 7.4.6

发行日期
2004-10-22

本次发布包含对 7.4.5 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.76.1. 迁移到版本 7.4.6

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.76.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 `make_oidjoins_check` 在 `/tmp` 中使用临时文件

这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

- 防止强制后端关闭重新发出先前的命令结果

  极少数情况下，客户端可能认为其最后一条命令成功了，
  而它实际上已被强制数据库关闭中止。

- 修复 `pg_stat_get_backend_idset` 中的缺陷

  这可能导致某些系统统计视图行为异常。

- 修复 postmaster 中的小内存泄漏

- 修复“预期两个交换的表都有 TOAST 表”缺陷

  这可能发生在如 ALTER TABLE DROP COLUMN 之后再 CLUSTER 的情形。

- 防止 `pg_ctl restart` 多次添加 `-D`

- 修复 GiST 索引中 NULL 值的问题

- `::` 不再被解释为 ECPG prepare 语句中的变量

E.77. 版本 7.4.5

发行日期
2004-08-18

此发行版在 7.4.4 的基础上包含一个严重的缺陷修复。有关 7.4 主版本新特性的信息，请参阅第 E.82 节。

E.77.1. 迁移到版本 7.4.5

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.77.2. 变更

- 修复并发 B-tree 索引插入期间可能的崩溃

  此补丁修复了并发插入 B-tree 索引可能导致服务器 panic 的罕见情形。不会造成永久损坏，但仍值得重新发布。该缺陷在 7.4 之前的版本中不存在。

E.78. 版本 7.4.4

发行日期
2004-08-16

本次发布包含对 7.4.3 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.78.1. 迁移到版本 7.4.4

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.78.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 在求值聚合计划的结果列表之前检查 HAVING 限制
- 避免会话的当前用户 ID 被删除时的崩溃
- 修复零行情形的哈希 crosstab（Joe）
- 在外键中重命名列后强制更新缓存
- 正确美化 UNION 查询的打印
- 让 psql 在 COPY IN 中正确处理 `\r\n` 换行
- pg_dump 对带 grant 选项的 ACL 处理不正确
- 修复 OS X 和 Solaris 的线程支持
- 更新了 JDBC 驱动（build 215），包含各种修复
- ECPG 修复
- 翻译更新（多位贡献者）

E.79. 版本 7.4.3

发行日期
2004-06-14

本次发布包含对 7.4.2 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.79.1. 迁移到版本 7.4.3

对于运行 7.4.X 的用户，不需要进行转储/恢复。

E.79.2. 变更

- 修复使用非哈希聚合时的临时内存泄漏（Tom）

- ECPG 修复，包括一些针对 Informix 兼容性的（Michael）
- 修复线程安全编译的问题，特别是 Solaris（Bruce）
- 修复使用旧网络协议时 COPY IN 结束时的错误（ljb）
- pg_autovacuum 的几个重要修复，包括大表、无符号 oid、稳定性、临时表和调试模式的修复（Matthew T. O'Connor）
- 修复在 NetBSD 和 BSD/OS 上读取 tar 格式转储的问题（Bruce）
- 若干 JDBC 修复
- 修复 last_value 等于重启值时的 ALTER SEQUENCE RESTART（Tom）
- 修复嵌套子查询未重新计算的问题（Tom）
- 修复 LIMIT/OFFSET 中非常量表达式的问题
- 支持无连接子句的 FULL JOIN，如 X FULL JOIN Y ON TRUE（Tom）
- 修复另一个零列表缺陷（Tom）
- 改进子查询中 GROUP BY 子句里非限定标识符的处理（Tom）

  子查询内的选择列表别名现在优先于来自外层查询级别的名称。
- 反编译规则时不生成“NATURAL CROSS JOIN”（Tom）
- 为二进制 COPY 中的无效字段长度添加检查（Tom）

  这修复了一个难以利用的安全漏洞。
- 避免 `ANALYZE` 与 `LISTEN/NOTIFY` 之间的锁定冲突
- 大量翻译更新（多位贡献者）

E.80. 版本 7.4.2

发行日期
2004-03-08

本次发布包含对 7.4.1 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.80.1. 迁移到版本 7.4.2

对于运行 7.4.X 的用户，不需要进行转储/恢复。不过，作为合并 7.4.X 系统目录初始内容中发现的两个错误的修复的最简单方法，它可能是可取的。用 7.4.2 的 initdb 做转储/initdb/重载序列会自动纠正这些问题。

两个错误中较严重的是数据类型 `anyarray` 的对齐标签错误；这是一个问题，因为 `pg_statistic` 系统目录使用 `anyarray` 列。当规划涉及双对齐列（如 `float8` 和 `timestamp`）

上 WHERE 子句的查询时，错误标记可能导致规划器误估甚至崩溃。强烈建议所有安装通过 initdb 或按照下述手工修复程序修复此错误。

较小的错误是系统视图 pg_settings 应被标记为具有公共更新访问权，以允许用 UPDATE pg_settings 替代 SET。这也可以通过 initdb 或手工修复，但除非你想用 UPDATE pg_settings，否则不必修复。

如果你不想做 initdb，以下程序可以修复 pg_statistic。作为数据库超级用户，执行：

```
-- clear out old data in pg_statistic:
DELETE FROM pg_statistic;
VACUUM pg_statistic;
-- this should update 1 row:
UPDATE pg_type SET typalign = 'd' WHERE oid = 2277;
-- this should update 6 rows:
UPDATE pg_attribute SET attalign = 'd' WHERE atttypid = 2277;
--
-- At this point you MUST start a fresh backend to avoid a crash!
--
-- repopulate pg_statistic:
ANALYZE;
```

这可以在运行中的数据库上完成，但要注意在重新填充 pg_statistic 之前，被更改数据库中运行的所有后端都必须重启才是安全的。

要修复 pg_settings 错误，只需执行：

```
GRANT SELECT, UPDATE ON pg_settings TO PUBLIC;
```

上述程序必须在安装的每个数据库中执行，包括 template1，理想情况下也包括 template0。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。template1 可以像任何其他数据库一样修复，但修复 template0 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 template0 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.80.2. 变更

7.4.2 版本包含 7.3.6 版本中的所有修复，外加以下修复：

- 修复可能导致优化器崩溃的 pg_statistics 对齐缺陷

  有关此问题的详情见上文。

- 允许非超级用户更新 pg_settings

- 修复几个优化器缺陷，其中大多数导致 “在子计划目标列表中未找到变量” 错误

- 避免大型多索引扫描启动期间的内存不足失败
- 修复可能导致 `COPY IN` 期间“内存不足”错误的多字节问题
- 修复从无 OID 的表 `SELECT INTO` / `CREATE TABLE AS` 的问题
- 修复并行测试期间 `alter_table` 回归测试的问题
- 修复达到打开文件限制的问题，特别是在 OS X 上（Tom）
- 对土耳其语区域设置问题的部分修复

  initdb 现在能在土耳其语区域设置下成功，但仍有一些与 `i/I` 问题相关的不便。
- 让 pg_dump 在恢复时设置客户端编码
- 其他次要的 pg_dump 修复
- 允许 ecpg 再次使用 C 关键字作为列名（Michael）
- 为 ecpg 的 `SELECT/INSERT/UPDATE/DELETE` 添加 `WHENEVER NOT_FOUND`（Michael）
- 修复调用返回集合函数的查询导致的 ecpg 崩溃（Michael）
- 各种其他 ecpg 修复（Michael）
- Borland 编译器的修复
- 线程构建改进（Bruce）
- 各种其他构建修复
- 各种 JDBC 修复

E.81. 版本 7.4.1

发行日期
2003-12-22

本次发布包含对 7.4 的多方面修复。有关 7.4 大版本中新特性的信息，请参见第 E.82 节。.

E.81.1. 迁移到版本 7.4.1

对于运行 7.4 的用户，不需要进行转储/恢复。

如果你想在信息模式中安装这些修复，需要把它重新装载到数据库中。这可以通过运行 `initdb` 初始化新集群来完成，或者在安装新版本后，以超级用户身份在 psql 中对每个数据库（理想情况下包括 `template1`）运行以下 SQL 命令序列：

```
DROP SCHEMA information_schema CASCADE;
\i /usr/local/pgsql/share/information_schema.sql
```

在第二条命令中替换你的安装路径。

E.81.2. 变更

- 修复 ECPG 中 `CREATE SCHEMA` 解析的缺陷（Michael）
- 修复 `--enable-thread-safety` 与 `--with-perl` 一起使用时的编译错误（Peter）
- 修复使用哈希连接的子查询（Tom）

  某些使用哈希连接的子查询会因不当共享的结构而崩溃。
- 修复空闲空间映射压缩缺陷（Tom）

  这修复了空闲空间映射的压缩可能导致数据库服务器关闭的缺陷。
- 修复 libpq 的 Borland 编译器构建（Bruce）
- 修复 `netmask()` 和 `hostmask()`，使其返回最大长度的掩码长度（Tom）

  修复这些函数使其返回与 7.4 之前版本一致的值。
- 若干 `contrib/pg_autovacuum` 修复

  修复包括不当的变量初始化、`TRUNCATE` 后缺失的清理以及长清理的时长计算溢出。
- 允许 `contrib/cube` 在 Cygwin 下编译（Jason Tishler）
- 修复 Solaris 在未定义口令时使用口令文件的问题（Tom）

  修复 Solaris 上因在未定义口令时使用任何类型的口令认证导致的崩溃。
- JDBC 线程问题修复及其他修复
- 修复 `bytea` 索引查找（Joe）
- 修复信息模式对位数据类型的支持（Peter）
- 强制在从 WAL 恢复期间开启 zero_damaged_pages
- 防止“变量不在子计划目标列表中”的一些晦涩情形
- 使 `PQescapeBytea` 与 `byteaout` 相互一致（Joe）
- 对大于 0x7e 的字节转义 `bytea` 输出（Joe）

  如果 `bytea` 输出和输入使用不同的客户端编码，`bytea` 值可能被不同的编码破坏。此修复转义了所有可能受影响的字节。
- 为 dblink 的 `get_tuple_of_interest()` 添加缺失的 `SPI_finish()` 调用（Joe）
- 新的捷克语 FAQ
- 修复信息模式视图 `constraint_column_usage` 对外键的支持（Peter）
- ECPG 修复（Michael）
- 修复多个 `IN` 子查询及子查询内连接的缺陷（Tom）
- 允许 `COUNT('x')` 工作（Tom）

- 把 Informix 兼容的 ECPG 头文件安装到单独的目录（Peter）

  一些 Informix 兼容的 ECPG 头文件名与操作系统头文件冲突。通过把它们安装在自己的目录中，减少了名称冲突。

- 修复 SSL 内存泄漏（Neil）

  此发行版修复了 7.4 中 SSL 未释放其分配的全部内存的缺陷。

- 防止 `pg_service.conf` 把服务名用作默认数据库名（Bruce）
- 修复 FreeBSD 上的本地 ident 认证（Tom）

E.82. 版本 7.4

发行日期
2003-11-17

E.82.1. 概述

此发行版的主要变更：

`IN` / `NOT IN` 子查询现在的效率高得多

在以前的版本中，`IN`/`NOT IN` 子查询通过顺序扫描子查询寻找匹配来与上层查询连接。7.4 的代码使用与普通连接相同的复杂技术，因此快得多。`IN` 现在通常与等价的 `EXISTS` 子查询一样快或更快；这颠倒了适用于以前版本的传统观念。

通过使用哈希桶改进了 `GROUP BY` 处理

在以前的版本中，要分组的行必须先排序。7.4 的代码可以不做排序就完成 `GROUP BY`，方法是把结果累积到每组一个条目的哈希表中。但如果估计哈希表太大放不进 `sort_mem`，它仍会使用排序技术。

新的多键哈希连接能力

在以前的版本中，哈希连接只能在单键上发生。此版本允许多列哈希连接。

使用显式 `JOIN` 语法的查询现在得到了更好的优化

以前的版本只按语法隐含的顺序求值使用显式 `JOIN` 语法的查询。7.4 允许对这些查询做完全优化，即优化器考虑所有可能的连接顺序并选择最高效的。
但外连接仍必须遵循声明的顺序。

更快、更强大的正则表达式代码

整个正则表达式模块已被 Henry Spencer 的新版本（最初为 Tcl 编写）取代。该代码大幅改进了性能并支持多种风格的正则表达式。

简单 SQL 函数的函数内联

简单 SQL 函数现在可以通过把其 SQL 包含在主查询中来内联。这通过消除每次调用的开销改进了性能。这意味着简单 SQL 函数现在的行为像宏。

完整支持 IPv6 连接和 IPv6 地址数据类型

以前的版本只允许 IPv4 连接，IP 数据类型也只支持 IPv4 地址。此版本在这两方面都增加了完整的 IPv6 支持。

SSL 性能与可靠性的重大改进

几位非常熟悉 SSL API 的人彻底改造了我们的 SSL 代码，以改进 SSL 密钥协商和错误恢复。

让空闲空间映射高效地重用空索引页面，以及其他空闲空间管理改进

在以前的版本中，因行被删除而变空的 B-tree 索引页只能被索引值与该页上最初索引的行相似的行重用。在 7.4 中，`VACUUM` 记录空索引页并允许它们被未来的任何索引行重用。

SQL 标准信息模式

信息模式提供了访问数据库中定义的模式对象信息的标准化且稳定的方式。

游标更贴近 SQL 标准

`FETCH` 和 `MOVE` 命令已被彻底改造，以更贴近 SQL 标准。

游标可以存在于事务之外

这些游标也称为可保持游标。

新的客户端-服务器协议

新协议增加了错误码、更多状态信息、更快的启动、对二进制数据传输的更好支持、与 SQL 命令分离的参数值、协议级的预备语句以及从 `COPY` 失败更干净的恢复。服务器和客户端仍支持较旧的协议。

libpq 和 ECPG 应用程序现在完全线程安全

虽然以前的 libpq 版本已支持线程，但此版本通过修复数据库连接启动期间使用的一些非线程安全代码改进了线程安全。必须使用 `configure` 选项 `--enable-thread-safety` 才能启用此特性。

新版本的全文索引

`contrib/tsearch2` 中提供了新的全文索引套件。

新的 autovacuum 工具

`contrib/autovacuum` 中的新 autovacuum 工具监控数据库统计表中的 `INSERT`/`UPDATE`/`DELETE` 活动，并在需要时自动清理表。

数组处理得到了改进并被移入服务器核心

数组的许多限制已被移除，数组的行为更像完全受支持的数据类型。

E.82.2. 迁移到版本 7.4

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

请注意以下不兼容之处：

- 服务器端的自动提交设置被移除，并在客户端应用程序和语言中重新实现。
  服务器端自动提交给想要控制自己自动提交行为的语言和应用程序带来了太多问题，
  因此自动提交从服务器中移除并酌情添加到各个客户端 API。
- 此版本中错误消息的措辞发生了实质变化。投入了大量努力使消息更一致、更面向用户。
  如果你的应用程序试图通过解析错误消息来检测不同的错误条件，
  强烈建议改用新的错误码机制。
- 使用显式 `JOIN` 语法的内连接可能行为不同，因为它们现在得到了更好的优化。
- 为清晰起见，许多服务器配置参数已被重命名，主要是与日志相关的那些。
- `FETCH 0` 或 `MOVE 0` 现在不做任何事情。在以前的版本中，`FETCH 0` 会获取所有剩余行，`MOVE 0` 会移到游标末尾。
- `FETCH` 和 `MOVE` 现在返回实际获取/移动的行数，如果位于游标开头/末尾则返回零。以前的版本会返回传给命令的行数，而不是实际获取或移动的行数。
- `COPY` 现在能处理使用回车或回车/换行行尾序列的文件。数据值中不再接受字面回车和换行；请改用 `\r` 和 `\n`。
- 从 `char(n)` 转换为 `varchar(n)` 或 `text` 时现在会修剪尾随空格。这本来就是大多数人一直期望的。
- 数据类型 `float(p)` 现在用二进制位而不是十进制位度量 *p*。新行为遵循 SQL 标准。
- 有歧义的日期值现在必须匹配 `datestyle` 设置指定的顺序。在以前的版本中，即使 `datestyle` 指定日在前，`10/20/03` 的日期规格也会被解释为十月的日期。7.4 将在日期规格对当前 `datestyle` 设置无效时抛出错误。
- 函数 `oidrand`、`oidsrand` 和 `userfntest` 已被移除。这些函数被认定不再有用。
- 指定随时间变化的日期/时间值的字符串字面量（如 `'now'` 或 `'today'`）在列默认表达式中不再按预期工作；它们现在使表创建时刻成为默认值，而不是插入时刻。应改用 `now()`、`current_timestamp` 或 `current_date` 等函数。

  在以前的版本中，有特殊代码使 `'now'` 之类的字符串在 `INSERT` 时而不是表创建时被解释，但这一变通没有覆盖所有情形。7.4 版本现在要求用 `now()` 或 `current_timestamp()` 等函数正确定义默认值。这些在所有情形下都能工作。
- 操作符名中不再允许美元符号（`$`）。它可以用作标识符中的非首字符。
  这是为了改进与其他数据库系统的兼容性，并避免参数占位符（`$n`）
  紧邻操作符书写时的语法问题。

E.82.3. 变更

下面你将找到 7.4 版本与上一个主版本之间变更的详细说明。

E.82.3.1. 服务器操作变更

- 允许 IPv6 服务器连接（Nigel Kukard、Johan Jordaan、Bruce、Tom、Kurt Roeckx、Andrew Dunstan）
- 修复 SSL 以干净地处理错误（Nathan Mueller）

  在以前的版本中，某些 SSL API 错误报告未被正确处理。此版本修复了这些问题。

- SSL 协议安全和性能改进（Sean Chittenden）

  SSL 密钥重协商过于频繁，导致 SSL 性能不佳。此外，初始密钥处理也得到了改进。

- 检测到死锁时打印锁信息（Tom）

  这使调试死锁情形更容易。

- 定期更新 `/tmp` 套接字的修改时间以避免它们被删除（Tom）

  这应有助于防止清理 `/tmp` 目录的管理脚本删除服务器套接字文件。

- 为 Mac OS X 启用 PAM（Aaron Hillegass）

- 使 B-tree 索引完全 WAL 安全（Tom）

  在以前的版本中，某些罕见情形下服务器崩溃可能导致 B-tree 索引损坏。此版本消除了最后几个罕见情形。

- 允许 B-tree 索引压缩和空页重用（Tom）

- 修复首个根页分裂期间索引查找不一致的问题（Tom）

  在以前的版本中，当单页索引分裂为两页时，
  有一小段时间另一个数据库会话可能看不到某个索引条目。此版本修复了这一罕见失败情形。

- 改进空闲空间映射的分配逻辑（Tom）

- 在服务器重启之间保留空闲空间信息（Tom）

  在以前的版本中，postmaster 停止时空闲空间映射不被保存，因此新启动的服务器没有空闲空间信息。此版本保存空闲空间映射，并在服务器重启时重新装载它。

- 为 `pg_stat_activity` 添加开始时间（Neil）

- 检测损坏磁盘页的新代码；用 `zero_damaged_pages` 清除（Tom）

- 新的客户端/服务器协议：更快、无用户名长度限制、允许从 `COPY` 干净退出（Tom）

- 为客户端/服务器协议添加事务状态、表 ID、列 ID（Tom）

- 为客户端/服务器协议添加二进制 I/O（Tom）

- 移除服务器端自动提交设置；移至客户端应用程序（Tom）

- 新的错误消息措辞、错误码和三级错误细节（Tom、Joe、Peter）

E.82.3.2. 性能改进

- 为 `GROUP BY` 聚合添加哈希（Tom）

- 使嵌套循环连接更聪明地使用多列索引（Tom）

- 允许多键哈希连接（Tom）

- 改进常量折叠（Tom）

- 新增内联简单 SQL 函数的能力（Tom）

- 降低使用复杂函数的查询的内存使用（Tom）

  在以前的版本中，返回已分配内存的函数直到查询完成才释放它。
  此版本允许在函数调用完成时释放函数分配的内存，减少了函数使用的总内存。

- 改进 GEQO 优化器性能（Tom）

  此版本修复了 GEQO 优化器管理潜在查询路径方式中的几个低效之处。

- 允许 `IN`/`NOT IN` 通过哈希表处理（Tom）
- 改进 `NOT IN (subquery)` 的性能（Tom）
- 允许大多数 `IN` 子查询作为连接处理（Tom）
- 模式匹配操作可以使用索引而不受区域设置影响（Peter）

  非 ASCII 区域设置无法使用标准索引做 `LIKE` 比较。此版本增加了一种为 `LIKE` 创建特殊索引的方法。

- 允许 postmaster 用 `preload_libraries` 预装载库（Joe）

  对于需要很长时间装载的共享库，此选项使库可以在 postmaster 中预装载并被所有数据库会话继承。

- 改进优化器代价计算，特别是对子查询（Tom）
- 当子查询 `ORDER BY` 与上层查询匹配时避免排序（Tom）
- 从 `WHERE a.x = b.y AND b.y = 42` 推出 `a.x = 42`（Tom）
- 允许复杂连接上的哈希/归并连接（Tom）
- 允许更多数据类型的哈希连接（Tom）
- 允许显式内连接的连接优化，可用 `join_collapse_limit` 禁用（Tom）
- 新增参数 `from_collapse_limit`，控制子查询到连接的转换（Tom）
- 使用来自 Tcl 的更快更强大的正则表达式代码（Henry Spencer、Tom）
- 在优化器中使用位图关系集（Tom）
- 改进连接启动时间（Tom）

  新的客户端/服务器协议启动数据库会话需要更少的网络包。

- 改进触发器/约束的性能（Stephan）
- 改进 `col IN (const, const, const, ...)` 的速度（Tom）
- 修复罕见情形下损坏的哈希索引（Tom）
- 改进哈希索引的并发和速度（Tom）

  以前的版本哈希索引性能不佳，特别是在高并发情形下。此版本修复了这一点，
  开发组有兴趣了解比较 B-tree 与哈希索引性能的报告。

- 把共享缓冲区对齐到 32 字节边界以改进复制速度（Manfred Spraul）

某些 CPU 在地址按 32 字节对齐时执行更快的数据复制。

- 为更好的性能重新实现了 `numeric` 数据类型（Tom）

  `numeric` 以前以 100 为基数存储。新代码使用 10000 为基数，性能显著提升。

E.82.3.3. 服务器配置变更

- 把服务器参数 `server_min_messages` 改名为 `log_min_messages`（Bruce）

  这是为了使大多数控制服务器日志的参数以 `log_` 开头。

- 把 `show_*_stats` 改名为 `log_*_stats`（Bruce）
- 把 `show_source_port` 改名为 `log_source_port`（Bruce）
- 把 `hostname_lookup` 改名为 `log_hostname`（Bruce）
- 新增 `checkpoint_warning`，对过度检查点发出警告（Bruce）

  在以前的版本中，很难确定检查点是否发生得过于频繁。
  此特性在过度检查点发生时向服务器日志添加警告。

- 新的本地化只读服务器参数（Tom）
- 把调试服务器日志消息改为以 `DEBUG` 而不是 `LOG` 输出（Bruce）
- 防止服务器日志变量被非超级用户关闭（Bruce）

  这是一个安全特性，使非超级用户不能禁用管理员启用的日志。

- `log_min_messages`/`client_min_messages` 现在控制 `debug_*` 输出（Bruce）

  这集中了客户端调试信息，使所有调试输出都可以发送到客户端或服务器日志。

- 新增 Mac OS X Rendezvous 服务器支持（Chris Campbell）

  这允许 Mac OS X 主机向网络查询可用的 PostgreSQL 服务器。

- 新增用 `log_min_duration_statement` 只打印慢语句的能力（Christopher）

  这是常被要求的调试特性，允许管理员在服务器日志中只看到慢查询。

- 允许 `pg_hba.conf` 接受 CIDR 格式的网络掩码（Andrew Dunstan）

  这允许管理员把主机 IP 地址和网络掩码字段合并为 `pg_hba.conf` 中的单个 CIDR 字段。

- 新的只读参数 `is_superuser`（Tom）
- 新参数 `log_error_verbosity`，控制错误细节（Tom）

  它与新的错误报告特性一起工作，提供提示、文件名和行号等额外错误信息。

- `postgres --describe-config` 现在转储服务器配置变量（Aizaz Ahmed、Peter）

  此选项对需要知道配置变量名及其最小值、最大值、默认值和描述的管理工具很有用。

- 为 pg_settings 新增列：context、type、source、min_val、max_val（Joe）
- 如果可能，把默认 shared_buffers 设为 1000、max_connections 设为 100（Tom）

  以前的版本默认 64 个共享缓冲区，以便 PostgreSQL 甚至能在很老的系统上启动。此版本测试平台允许的共享内存量并在可能时选择更合理的默认值。当然，
  仍鼓励用户评估其资源负载并相应地设定 shared_buffers 的大小。
- 新的 pg_hba.conf 记录类型 hostnossl，阻止 SSL 连接（Jon Jensen）

  在以前的版本中，如果客户端和服务器都支持 SSL，就没有办法阻止 SSL 连接。此选项提供了该能力。
- 移除参数 geqo_random_seed（Tom）
- 新增服务器参数 regex_flavor，控制正则表达式处理（Tom）
- 让 pg_ctl 更好地处理非标准端口（Greg）

E.82.3.4. 查询变更

- 新的 SQL 标准信息模式（Peter）
- 新增只读事务（Peter）
- 在外键违反消息中打印键名和值（Dmitry Tkach）
- 允许用户在 pg_stat_activity 中看到自己的查询（Kevin Brown）

  在以前的版本中，只有超级用户能用 pg_stat_activity 看到查询字符串。现在普通用户可以看到自己的查询字符串。
- 修复子查询中的聚合以符合 SQL 标准（Tom）

  SQL 标准规定，如果嵌套子查询中出现的聚合函数的参数只包含外层查询变量，则它属于外层查询。以前的 PostgreSQL 版本没有正确处理这个细节。
- 新增阻止自动添加查询中引用的表的选项（Nigel J. Andrews）

  默认情况下，查询中提到的表如果不在 FROM 子句中会被自动添加到其中。这与历史 POSTGRES 行为兼容但违背 SQL 标准。此选项允许选择标准兼容的行为。
- 允许 UPDATE ... SET col = DEFAULT（Rod）

  这允许 UPDATE 把列设为其声明的默认值。
- 允许在 LIMIT/OFFSET 中使用表达式（Tom）

  在以前的版本中，LIMIT/OFFSET 只能使用常数，不能用表达式。
- 实现 CREATE TABLE AS EXECUTE（Neil、Peter）

E.82.3.5. 对象操作变更

- 使 CREATE SEQUENCE 语法更符合 SQL:2003（Neil）
- 新增语句级触发器（Neil）

虽然这允许触发器在语句结束时触发，但不允许触发器访问该语句修改的所有行。此能力计划在未来版本中提供。

- 为域新增检查约束（Rod）

  这通过允许域使用检查约束大幅提高了域的用处。

- 新增 `ALTER DOMAIN`（Rod）

  这允许操纵现有域。

- 修复几个零列表缺陷（Tom）

  PostgreSQL 支持零列表。这修复了使用此类表时发生的各种缺陷。

- 让 `ALTER TABLE ... ADD PRIMARY KEY` 添加非空约束（Rod）

  在以前的版本中，`ALTER TABLE ... ADD PRIMARY` 会添加唯一索引但不添加非空约束。此版本修复了这一点。

- 新增 `ALTER TABLE ... WITHOUT OIDS`（Rod）

  这允许控制新行和更新的行是否有 OID 列。这对节省存储空间最有用。

- 新增 `ALTER SEQUENCE`，修改最小值、最大值、增量、缓存、循环值（Rod）

- 新增 `ALTER TABLE ... CLUSTER ON`（Alvaro Herrera）

  此命令被 `pg_dump` 用来记录每个先前聚簇过的表的聚簇列。该信息被数据库范围的 cluster 用来聚簇所有先前聚簇过的表。

- 改进域的自动类型转换（Rod、Tom）

- 允许标识符中的美元符号，但不能作为首字符（Tom）

- 不允许操作符名中的美元符号，使 `x=$1` 可行（Tom）

- 允许用 `LIKE` *`subtable`* 复制表模式，还有 SQL:2003 特性 `INCLUDING DEFAULTS`（Rod）

- 为 `GRANT` 新增 `WITH GRANT OPTION` 子句（Peter）

  这使 `GRANT` 能授予其他用户在对象上授予权限的能力。

E.82.3.6. 工具命令变更

- 为 `CREATE TABLE` 的临时表新增 `ON COMMIT` 子句（Gavin）

  这增加了在事务提交时删除表或删除所有行的能力。

- 允许用 `WITH HOLD` 在事务之外使用游标（Neil）

  在以前的版本中，游标在创建它的事务结束时被删除。现在可以用 `WITH HOLD` 选项创建游标，这允许在创建事务提交后继续访问它们。

- `FETCH 0` 和 `MOVE 0` 现在不做任何事情（Bruce）

  在以前的版本中，`FETCH 0` 获取所有剩余行，`MOVE 0` 移到游标末尾。

- 按照 SQL 标准，让 `FETCH` 和 `MOVE` 返回获取/移动的行数，如果位于游标开头/末尾则返回零（Bruce）

  在以前的版本中，`FETCH` 和 `MOVE` 返回的行数没有准确反映处理的行数。

- 正确处理游标的 `SCROLL`，否则报告错误（Neil）

  对某些类型的查询允许随机访问（前向和后向滚动）需要一些额外的工作才能做到。
  如果创建游标时指定了 `SCROLL`，将执行这些额外工作。此外，如果游标以 `NO SCROLL` 创建，则不允许随机访问。

- 为 `FETCH` 和 `MOVE` 实现 SQL 兼容的选项 `FIRST`、`LAST`、`ABSOLUTE` *n*、`RELATIVE` *n*（Tom）

- 允许对 `DECLARE CURSOR` 使用 `EXPLAIN`（Tom）

- 允许 `CLUSTER` 默认使用标记为预聚簇的索引（Alvaro Herrera）

- 允许 `CLUSTER` 聚簇所有表（Alvaro Herrera）

  这允许用单条命令重新聚簇数据库中所有先前聚簇过的表。

- 禁止对部分索引使用 `CLUSTER`（Tom）

- 允许 `COPY` 文件中的 DOS 和 Mac 行尾（Bruce）

- 不允许字面回车作为数据值，反斜杠-回车和 `\r` 仍被允许（Bruce）

- `COPY` 变更（二进制、`\.`）（Tom）

- 从 `COPY` 失败中干净恢复（Tom）

- 防止 `COPY` 中可能的内存泄漏（Tom）

- 使 `TRUNCATE` 事务安全（Rod）

  `TRUNCATE` 现在可以在事务内使用。如果事务中止，`TRUNCATE` 所做的更改会自动回滚。

- 允许预备/绑定 `FETCH` 和 `EXPLAIN` 等工具命令（Tom）

- 新增 `EXPLAIN EXECUTE`（Neil）

- 通过减少 WAL 流量改进索引上的 `VACUUM` 性能（Tom）

- 函数索引已被推广为表达式上的索引（Tom）

  在以前的版本中，函数索引只支持应用于一个或多个列名的简单函数。
  此版本允许任何类型的标量表达式。

- 让 `SHOW TRANSACTION ISOLATION` 与 `SET TRANSACTION ISOLATION` 的输入匹配（Tom）

- 让对非本地数据库的 `COMMENT ON DATABASE` 生成警告而不是错误（Rod）

  数据库注释存储在数据库本地表中，因此对数据库的注释必须存储在每个数据库中。

- 改进 `LISTEN`/`NOTIFY` 的可靠性（Tom）

- 允许 REINDEX 可靠地重新索引非共享系统目录索引（Tom）

  这允许重新索引系统表而不需要独立会话，这在以前的版本中是必需的。
  现在唯一需要独立会话来重新索引的是全局系统表 pg_database、pg_shadow 和 pg_group。

E.82.3.7. 数据类型和函数变更

- 新服务器参数 extra_float_digits，控制浮点数的精度显示（Pedro Ferreira、Tom）

  这控制导致回归测试问题的输出精度。

- 允许 +1300 作为数字时区说明符，用于 FJST（Tom）

- 移除很少使用的函数 oidrand、oidsrand 和 userfntest（Neil）

- 把 md5() 函数加入主服务器，此前已在 contrib/pgcrypto 中（Joe）

  MD5 函数是常被要求的。需要更复杂的加密能力时，请使用 contrib/pgcrypto。

- 扩大 timestamp 的日期范围（John Cochran）

- 更改 EXTRACT(EPOCH FROM timestamp)，使 timestamp without time zone 被假定处于本地时间而不是 GMT（Tom）

- 以防操作系统不防止除以零，捕获它（Tom）

- 把 numeric 数据类型内部改为以 10000 为基数（Tom）

- 新函数 hostmask()（Greg Wickham）

- to_char() 和 to_timestamp() 的修复（Karel）

- 允许接受任意参数数据类型并返回任意数据类型的函数，使用 anyelement 和 anyarray（Joe）

  这允许创建能用于任何数据类型的函数。

- 数组现在可以指定为 ARRAY[1,2,3]、ARRAY[['a','b'],['c','d']] 或 ARRAY[ARRAY[ARRAY[2]]]（Joe）

- 允许数组的正确比较，包括 ORDER BY 和 DISTINCT 支持（Joe）

- 允数组列上的索引（Joe）

- 允许用 || 连接数组（Joe）

- 允许 WHERE 限定 *exprop* ANY/SOME/ALL (*array_expr*)（Joe）

  这允许数组表现得像值列表，用于诸如 SELECT * FROM tab WHERE col IN (array_val) 之类的用途。

- 新数组函数 array_append、array_cat、array_lower、array_prepend、array_to_string、array_upper、string_to_array（Joe）

- 允许用户定义的聚合使用多态函数（Joe）

- 允许给空数组赋值（Joe）
- 允许 `time`、`timestamp` 和 `interval` 输入值的秒字段为 60（Tom）

  闰秒需要六十秒的值。
- 允许 `cidr` 数据类型转换为 `text`（Tom）
- 不允许 SET TIMEZONE 中的无效时区名
- 当 `char` 转换为 `varchar` 或 `text` 时修剪尾随空格（Tom）
- 让 `float(p)` 用二进制位而不是十进制位度量精度 *p*（Tom）
- 为 `inet` 和 `cidr` 数据类型添加 IPv6 支持（Michael Graff）
- 新增 `family()` 函数，报告地址是 IPv4 还是 IPv6（Michael Graff）
- 让 `SHOW datestyle` 生成与 `SET datestyle` 所用类似的输出（Tom）
- 让 `EXTRACT(TIMEZONE)` 和 `SET/SHOW TIME ZONE` 遵循 SQL 对时区偏移符号的约定，即正值为 UTC 以东（Tom）
- 修复 `date_trunc('quarter', ...)`（Böjthe Zoltán）

  以前的版本对此函数调用返回不正确的值。
- 使 `initcap()` 更兼容 Oracle（Mike Nolan）

  `initcap()` 现在把任何非字母数字字符之后出现的字母大写，而不只是空白之后。
- 对非 ISO-8601 格式的日期值只允许 `datestyle` 字段顺序（Greg）
- 新增 `datestyle` 值 `MDY`、`DMY` 和 `YMD` 来设置输入字段顺序；为向后兼容保留 `US` 和 `European`（Tom）
- 像 `'now'` 或 `'today'` 这样的字符串字面量将不再可用作列默认值。请改用 `now()`、`current_timestamp` 等函数。（预备语句需要此变更）（Tom）
- 在 `min()`/`max()` 中把 NaN 当作大于任何其他值（Tom）

  在大多数用途中 NaN 已排在普通数值之后，但 `min()` 和 `max()` 没有正确处理。
- 防止 interval 抑制 `:00` 秒的显示
- 新函数 `pg_get_triggerdef(prettyprint)` 和 `pg_conversion_is_visible()`（Christopher）
- 允许时间指定为 `040506` 或 `0405`（Tom）
- 输入日期顺序现在必须是 `YYYY-MM-DD`（4 位年份）或匹配 `datestyle`
- 让 `pg_get_constraintdef` 支持唯一、主键和检查约束（Christopher）

E.82.3.8. 服务器端语言变更

- 防止对零行记录变量使用 `RETURN NEXT` 时 PL/pgSQL 崩溃（Tom）

- 让 PL/Python 的 `spi_execute` 接口正确处理空值（Andrew Bosma）
- 允许 PL/pgSQL 不用 `%ROWTYPE` 声明组合类型的变量（Tom）
- 修复 PL/Python 的 `_quote()` 函数以处理大整数
- 把 PL/Python 改为不可信语言，现名 `plpythonu`（Kevin Jacobs、Tom）

  Python 语言不再支持受限执行环境，因此删除了可信版 PL/Python。如果这种情况改变，将重新添加可供非超级用户使用的 PL/Python 版本。
- 允许多态 PL/pgSQL 函数（Joe、Tom）
- 允许多态 SQL 函数（Joe）
- 改进了 PL/pgSQL 的编译函数缓存机制，完全支持多态（Joe）
- 在 PL/pgSQL 中新增表示函数实际返回类型的新参数 `$0`（Joe）
- 允许 PL/Tcl 和 PL/Python 在多个表上使用同一触发器（Tom）
- 修复 PL/Tcl 的 `spi_prepare`，使其接受参数类型列表中完全限定的类型名（Jan）

E.82.3.9. psql 变更

- 新增 `\pset pager always`，总是使用分页器（Greg）

  这强制使用分页器，即使行数少于屏幕高度。这对跨多个屏幕行的行很有价值。
- 改进 Tab 补全（Rod、Ross Reedstrom、Ian Barwick）
- 把 `\?` 帮助重排为分组（Harald Armin Massa、Bruce）
- 新增列出模式、转换和编码转换的反斜杠命令（Christopher）
- `\encoding` 现在根据服务器参数 `client_encoding` 变化（Tom）

  在以前的版本中，`\encoding` 不知道用 `SET client_encoding` 所做的编码更改。
- 把编辑器缓冲区保存到 readline 历史（Ross）

  当用 `\e` 编辑查询时，结果被保存到 readline 历史中，可用上箭头检索。
- 改进 `\d` 显示（Christopher）
- 增强 HTML 模式使其更符合标准（Greg）
- 新的 `\set AUTOCOMMIT off` 能力（Tom）

  这取代了被移除的服务器参数 `autocommit`。
- 新的 `\set VERBOSITY`，控制错误细节（Tom）

  这控制新的错误报告细节。
- 新的提示符转义序列 `%x`，显示事务状态（Tom）
- psql 的长选项现在在所有平台上可用

E.82.3.10. pg_dump 变更

- 多个 pg_dump 修复，包括 tar 格式和大对象
- 允许 pg_dump 转储特定的模式（Neil）
- 让 pg_dump 保留列存储特性（Christopher）

  这保留 `ALTER TABLE ... SET STORAGE` 信息。
- 让 pg_dump 保留 `CLUSTER` 特性（Christopher）
- 让 pg_dumpall 用 `GRANT/REVOKE` 转储数据库级权限（Tom）
- 允许 pg_dumpall 支持 pg_dump 的选项 `-a`、`-s`、`-x`（Tom）
- 防止 pg_dump 把命令行上指定的标识符转为小写（Tom）
- pg_dump 选项 `--use-set-session-authorization` 和 `--no-reconnect` 现在无作用，所有转储都使用 `SET SESSION AUTHORIZATION`

  pg_dump 不再重新连接来切换用户，而是始终使用 `SET SESSION AUTHORIZATION`。这将减少恢复期间的口令提示。
- pg_dump 的长选项现在在所有平台上可用

  PostgreSQL 现在包含自己的长选项处理例程。

E.82.3.11. libpq 变更

- 新增用于在 Windows 上释放内存的函数 `PQfreemem`，建议用于 `NOTIFY`（Bruce）

  Windows 要求库中分配的内存由同一库中的函数释放，因此 `free()` 不能释放 libpq 分配的内存。`PQfreemem` 是释放 libpq 内存的正确方式，特别是在 Windows 上，也建议在其他平台上使用。
- 记录服务能力并添加示例文件（Bruce）

  这允许客户端在客户端机器上的中心文件中查找连接信息。
- 让 `PQsetdbLogin` 与 `PQconnectdb` 有相同的默认值（Tom）
- 允许 libpq 在结果集过大时干净地失败（Tom）
- 改进函数 `PQunescapeBytea` 的性能（Ben Lamb）
- 用 `configure` 选项 `--enable-thread-safety` 允许线程安全的 libpq（Lee Kindness、Philip Yarra）
- 允许函数 `pqInternalNotice` 接受格式串和参数而不只是预先格式化的消息（Tom、Sean Chittenden）
- 用 `sslmode` 值 `disable`、`allow`、`prefer` 和 `require` 控制 SSL 协商（Jon Jensen）
- 允许新的错误码和文本级别（Tom）
- 允许访问查询结果的基础表和列（Tom）

这对想知道与特定结果集关联的基础表和列名的查询构建器应用程序很有帮助。

- 允许访问当前事务状态（Tom）
- 新增把二进制数据直接传递给服务器的能力（Tom）
- 新增函数 `PQexecPrepared` 和 `PQsendQueryPrepared`，执行先前预备语句的绑定/执行（Tom）

E.82.3.12. JDBC 变更

- 允许可更新结果集上的 `setNull`
- 允许预备语句上的 `executeBatch`（Barry）
- 支持 SSL 连接（Barry）
- 处理结果集中的模式名（Paul Sorenson）
- 新增 refcursor 支持（Nic Ferrier）

E.82.3.13. 其他接口变更

- 防止 libpgtcl 关闭期间可能的内存泄漏或核心转储（Tom）
- 为 ECPG 新增 Informix 兼容性（Michael）

  这允许 ECPG 处理使用某些 Informix 扩展编写的嵌入式 C 程序。
- 为 ECPG 新增定长的 `decimal` 类型，用于 Informix（Michael）
- 用 `configure` 选项 `--enable-thread-safety` 允许线程安全的嵌入式 SQL 程序（Lee Kindness、Bruce）

  这允许多个线程同时访问数据库。
- 把 Python 客户端 PyGreSQL 移到http://www.pygresql.org（Marc）

E.82.3.14. 源代码变更

- 消除对单独的平台几何回归结果文件的需要（Tom）
- 改进的 PPC 锁定原语（Reinhard Max）
- 新函数 `palloc0`，分配并清零内存（Bruce）
- 修复 s390x CPU（64 位）的锁定代码（Tom）
- 允许 OpenBSD 使用本地 ident 凭据（William Ahern）
- 使查询计划树对执行器只读（Tom）
- 新增 Darwin 启动脚本（David Wheeler）
- 允许 libpq 用 Borland C++ 编译器编译（Lester Godwin、Karl Waclawek）

- 需要时使用我们自己的 `getopt_long()` 版本（Peter）
- 把管理脚本转换为 C（Peter）
- 如果从 CVS 构建，现在需要 Bison >= 1.85 来构建 PostgreSQL 语法
- 把文档合并为一本书（Peter）
- 新增 Windows 兼容函数（Bruce）
- 允许客户端接口在 MinGW 下编译（Bruce）
- 新的用于错误报告的 `ereport()` 函数（Tom）
- 支持 Linux 上的 Intel 编译器（Peter）
- 改进 Linux 启动脚本（Slawomir Sudnik、Darko Prenosil）
- 新增对 AMD Opteron 和 Itanium 的支持（Jeffrey W. Baker、Bruce）
- 从 `configure` 移除 `--enable-recode` 选项

  既然有了 `CREATE CONVERSION`，这不再需要。
- 如果未找到自旋锁代码则产生编译错误（Bruce）

  没有自旋锁代码的平台现在将编译失败，而不是静默使用信号量。此失败可以用新的 `configure` 选项禁用。

E.82.3.15. 附加模块变更

- 把 dbmirror 许可证改为 BSD
- 改进 earthdistance（Bruno Wolff III）
- pgcrypto 的可移植性改进（Marko Kreen）
- 防止 xml 中的崩溃（John Gray、Michael Richards）
- 更新 oracle
- 更新 mysql
- 更新 cube（Bruno Wolff III）
- 更新 earthdistance 以使用 cube（Bruno Wolff III）
- 更新 btree_gist（Oleg）
- 新的 tsearch2 全文搜索模块（Oleg、Teodor）
- 为 tablefuncs 新增基于哈希的 crosstab 函数（Joe）
- 为 tablefuncs 新增用来对 `connectby()` 兄弟节点排序的 serial 列（Nabil Sayegh、Joe）
- 为 dblink 新增命名持久连接（Shridhar Daithanka）

- 新的 pg_autovacuum 允许自动 `VACUUM`（Matthew T. O'Connor）
- 让 pgbench 遵循环境变量 `PGHOST`、`PGPORT`、`PGUSER`（Tatsuo）
- 改进 intarray（Teodor Sigaev）
- 改进 pgstattuple（Rod）
- 修复 fuzzystrmatch 中 `metaphone()` 的缺陷
- 改进 adddepend（Rod）
- 更新 spi/timetravel（Böjthe Zoltán）
- 修复 dbase 的 `-s` 选项并改进非 ASCII 处理（Thomas Behr、Márcio Smiderle）
- 移除 array 模块，因为其功能现已默认包含（Joe）

E.83. 版本 7.3.21

发行日期
2008-01-07

此发行版包含 7.3.20 之后的多种修复，包括对重要安全问题的修复。

这预计是 PostgreSQL 7.3.X 系列的最后一个发行版。建议用户尽快更新到较新的发行分支。

E.83.1. 迁移到版本 7.3.21

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.83.2. 变更

- 防止索引中的函数以运行 `VACUUM`、`ANALYZE` 等的用户的权限执行（Tom）

  每当产生新的表条目时，都会求值索引表达式和部分索引谓词中使用的函数。人们早就明白，如果修改不可信用户拥有的表，这会带来特洛伊木马代码执行的风险。（注意触发器、默认值、检查约束等也构成同类风险。）但索引中的函数带来额外危险，因为它们会由例程维护操作（如 `VACUUM FULL`）执行，而这些操作通常在超级用户账户下自动执行。例如，恶意用户可以通过设置特洛伊木马索引定义并等待下一次例程清理来以超级用户权限执行代码。修复安排标准维护操作（包括 `VACUUM`、`ANALYZE`、`REINDEX` 和 `CLUSTER`）以表所有者而非调用用户的身份执行，使用已用于 `SECURITY DEFINER` 函数的同一权限切换机制。为防止绕过此安全措施，现在禁止在 `SECURITY DEFINER` 上下文中执行 `SET SESSION AUTHORIZATION` 和 `SET ROLE`。（CVE-2007-6600）

- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

  7.3.20 中出现的修复并不完整，它只堵住了部分 `dblink` 函数的漏洞。（CVE-2007-6601、CVE-2007-3278）

- 修复使用多字节数据库编码时 `translate()` 可能的崩溃（Tom）
- 让 `contrib/tablefunc` 的 `crosstab()` 把 NULL rowid 本身作为一个类别处理，而不是崩溃（Joe）
- 要求在重新生成 `configure` 脚本时使用特定版本的 Autoconf（Peter）

  这只影响开发者和打包者。此变更是为了防止意外使用未经测试的 Autoconf 与 PostgreSQL 版本组合。如果确实想使用不同的 Autoconf 版本，可以移除版本检查，但结果是否可用由你自己负责。

E.84. 版本 7.3.20

发行日期
2007-09-17

此发行版包含 7.3.19 之后的修复。

E.84.1. 迁移到版本 7.3.20

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.84.2. 变更

- 防止事务插入行后在同一表上并发 `VACUUM` 接近结束时中止导致的索引损坏（Tom）
- 让 `CREATE DOMAIN ... DEFAULT NULL` 正确工作（Tom）
- 修复 `log_min_error_statement` 日志记录内存耗尽时的崩溃（Tom）
- 作为安全措施，要求使用 `/contrib/dblink` 的非超级用户只使用口令认证（Joe）

E.85. 版本 7.3.19

发行日期
2007-04-23

此发行版包含 7.3.18 之后的修复，包括一个安全修复。

E.85.1. 迁移到版本 7.3.19

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.85.2. 变更

- 支持在 `search_path` 中显式放置临时表模式，并禁用为函数和操作符搜索它（Tom）

  这是允许安全定义者函数设置真正安全的 `search_path` 值所必需的。没有它，非特权 SQL 用户可以利用临时对象以安全定义者函数的权限执行代码（CVE-2007-2138）。更多信息请参阅 `CREATE FUNCTION`。

- 修复 `VACUUM FULL` 处理 `UPDATE` 链方式中的潜在数据损坏缺陷（Tom、Pavan Deolasee）

E.86. 版本 7.3.18

发行日期
2007-02-05

此发行版包含 7.3.17 之后的多种修复，包括一个安全修复。

E.86.1. 迁移到版本 7.3.18

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.86.2. 变更

- 移除允许已连接用户读取后端内存的安全漏洞（Tom）

  该漏洞涉及更改 SQL 函数中使用的表列的数据类型（CVE-2007-0555）。此错误很容易被利用来导致后端崩溃，而且原则上可能被用来读取用户不应能访问的数据库内容。

- 修复罕见的缺陷：btree 索引页分裂可能因选择不可行的分裂点而失败（Heikki Linnakangas）

- 收紧对超过三字节的 UTF8 序列的多字节字符处理的安全性（Tom）

E.87. 版本 7.3.17

发行日期
2007-01-08

此发行版包含 7.3.16 之后的多种修复。

E.87.1. 迁移到版本 7.3.17

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.87.2. 变更

- 对于新的 initdb 安装，`to_number()` 和 `to_char(numeric)` 现在是 `STABLE` 而不是 `IMMUTABLE`（Tom）

  这是因为 `lc_numeric` 可能改变这些函数的输出。

- 改进使用括号的正则表达式的索引利用（Tom）

  这同时也改进了 psql 的 `\d` 性能。

E.88. 版本 7.3.16

发行日期
2006-10-16

此发行版包含 7.3.15 之后的多种修复。

E.88.1. 迁移到版本 7.3.16

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.88.2. 变更

- 修复 psql 的 `\d` 命令模式匹配的极端情形
- 修复 /contrib/ltree 中损坏索引的缺陷（Teodor）
- 向后移植 7.4 自旋锁代码，以改进性能并更好地支持 64 位架构
- 修复 libpq 中与 SSL 相关的内存泄漏
- 修复 /contrib/dbmirror 中的反斜杠转义
- 针对美国 DST 法规的最近变更调整回归测试

E.89. 版本 7.3.15

发行日期
2006-05-23

此发行版包含 7.3.14 之后的多种修复，包括针对极其严重安全问题的补丁。

E.89.1. 迁移到版本 7.3.15

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

要完全防范 CVE-2006-2313 和 CVE-2006-2314 所述的 SQL 注入攻击，可能需要更改应用程序代码。如果你有把不可信字符串嵌入 SQL 命令的应用程序，应尽快检查它们，确保使用了推荐的转义技术。在大多数情况下，应用程序应使用库或驱动提供的子例程（如 libpq 的 `PQescapeStringConn()`）来执行字符串转义，而不是依靠临时代码来做。

E.89.2. 变更

- 更改服务器，在所有情形下拒绝无效编码的多字节字符（Tatsuo、Tom）

  虽然 PostgreSQL 向这个方向发展已有一段时间，但现在检查统一应用于所有编码和所有文本输入，而且总是错误而不只是警告。此变更可防范 CVE-2006-2313 所述类型的 SQL 注入攻击。

- 拒绝字符串字面量中对 `\'` 的不安全使用

  作为针对 CVE-2006-2314 所述类型 SQL 注入攻击的服务器端防御，服务器现在只接受 `''` 而不接受 `\'` 作为 SQL 字符串字面量中 ASCII 单引号的表示。默认情况下，只有当 `client_encoding` 设为仅客户端编码（SJIS、BIG5、GBK、GB18030 或 UHC）时才拒绝 `\'`，这正是可能发生 SQL 注入的场景。需要时可用新的配置参数 `backslash_quote` 调整此行为。注意，要完全防范 CVE-2006-2314 可能需要客户端更改；`backslash_quote` 的目的之一是让不安全的客户端明显地不安全。

- 修改 libpq 的字符串转义例程，使其感知编码问题

  这为使用 libpq 的应用程序修复了 CVE-2006-2313 和 CVE-2006-2314 所述的安全问题。并发使用多个 PostgreSQL 连接的应用程序应迁移到 `PQescapeStringConn()` 和 `PQescapeByteaConn()`，以确保为每个数据库连接中使用的设置正确转义。以“手工”方式做字符串转义的应用程序应改为依靠库例程。

- 修复一些不正确的编码转换函数

  `win1251_to_iso`、`alt_to_iso`、`euc_tw_to_big5`、`euc_tw_to_mic`、`mic_to_euc_tw` 都在不同程度上损坏。

- 清理字符串中偶然残留的 `\'` 使用（Bruce、Jan）

- 修复服务器正确使用自定义 DH SSL 参数（Michael Fuhr）

- 修复各种小的内存泄漏

E.90. 版本 7.3.14

发行日期
2006-02-14

此发行版包含 7.3.13 之后的多种修复。

E.90.1. 迁移到版本 7.3.14

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.13 的版本升级，请参阅 7.3.13 的发布说明。

E.90.2. 变更

- 修复 `SET SESSION AUTHORIZATION` 中可能的崩溃（CVE-2006-0553）

  如果服务器编译时启用了断言（这不是默认），非特权用户可以使服务器进程崩溃，导致对其他用户的瞬时拒绝服务。感谢 Akio Ishida 报告此问题。

- 修复自插入行的行可见性逻辑缺陷（Tom）

  极少数情况下，当前命令插入的行可能被看作已经有效，而不应该如此。修复 7.3.11 版本中引入的缺陷。

- 修复 pg_clog 文件创建期间可能导致“文件已存在”错误的竞态条件（Tom）
- 修复以允许恢复有对自定义操作符跨模式引用的转储（Tom）
- 修复 configure 期间测试 `finite` 和 `isinf` 存在性的可移植性问题（Tom）

E.91. 版本 7.3.13

发行日期
2006-01-09

此发行版包含 7.3.12 之后的多种修复。

E.91.1. 迁移到版本 7.3.13

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅 7.3.10 的发布说明。另外，如果你受到下述区域设置或 plperl 问题的影响，可能需要在更新后对文本列上的索引执行 `REINDEX`。

E.91.2. 变更

- 修复把不同字符组合视为相等的区域设置（如匈牙利语）下的字符串比较（Tom）

  这可能需要 `REINDEX` 来修复文本列上的现有索引。

- 在 postmaster 启动期间设置区域设置环境变量，以确保 plperl 之后不会更改区域设置

  这修复了 postmaster 启动时环境变量指定的区域设置与 initdb 所被告知的不同时发生的问题。在这些条件下，任何 plperl 的使用都可能导致索引损坏。如果你遇到过这种情况，可能需要 `REINDEX` 来修复文本列上的现有索引。

- 修复某些很少使用的亚洲多字节字符集中 strpos() 和正则表达式处理的长期缺陷（Tatsuo）
- 修复 `/contrib/pgcrypto` gen_salt 中的缺陷，该缺陷使 MD5 和 XDES 算法不能使用全部可用盐空间（Marko Kreen、Solar Designer）

  Blowfish 和标准 DES 的盐不受影响。

- 修复 `/contrib/dblink`，当指定的列数与查询实际返回的列数不同时抛出错误而不是崩溃（Joe）

E.92. 版本 7.3.12

发行日期
2005-12-12

此发行版包含 7.3.11 之后的多种修复。

E.92.1. 迁移到版本 7.3.12

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅 7.3.10 的发布说明。

E.92.2. 变更

- 修复事务日志管理中的竞态条件

  存在一个窄窗口，其中可能为错误的页发起 I/O 操作，导致断言失败或数据损坏。

- `/contrib/ltree` 修复（Teodor）

- 修复外连接的长期规划错误

  此缺陷有时导致虚假错误“RIGHT JOIN 只支持可归并连接的连接条件”。

- 防止表被删除后 pg_autovacuum 的核心转储

E.93. 版本 7.3.11

发行日期
2005-10-04

此发行版包含 7.3.10 之后的多种修复。

E.93.1. 迁移到版本 7.3.11

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，如果你是从早于 7.3.10 的版本升级，请参阅 7.3.10 的发布说明。

E.93.2. 变更

- 修复允许 `VACUUM` 过早移除 `ctid` 链的错误，并在跟随 `ctid` 链接的代码中添加更多检查

  这修复了一个在极罕见情况下可能导致崩溃的长期问题。

- 修复 `CHAR()` 在使用多字节字符集时正确填充空格到指定长度（Yoshiyuki Asaba）

在以前的版本中，`CHAR()` 的填充是不正确的，因为它只填充到指定的字节数而不考虑存储了多少字符。

- 修复列 `a` 上有 GiST 索引时 `UPDATE a=... WHERE a...` 之类查询丢行的问题
- 改进对部分写入 WAL 页的检查
- 改进启用 SSL 时信号处理的健壮性
- 各种内存泄漏修复
- 各种可移植性改进
- 修复 PL/pgSQL 在变量为按引用传递类型时正确处理 `var := var`

E.94. 版本 7.3.10

发行日期
2005-05-09

此发行版包含 7.3.9 之后的多种修复，包括几个与安全相关的问题。

E.94.1. 迁移到版本 7.3.10

对于运行 7.3.X 的用户，不需要进行转储/恢复。不过，它是处理在 7.3.X 系统目录初始内容中发现的严重安全问题的可能方法之一。用 7.3.10 的 initdb 做转储/initdb/重载序列会自动纠正此问题。

安全问题是内建字符集编码转换函数可以由非特权用户从 SQL 命令调用，但这些函数并非为此设计，对恶意选择的参数不安全。修复包括更改这些函数的声明参数列表，使其不能再从 SQL 命令调用。（这不影响编码转换机制对它们的正常使用。）强烈建议所有安装通过 initdb 或按照下述手工修复程序修复此错误。该错误至少允许非特权数据库用户使其服务器进程崩溃，而且可能允许非特权用户获得数据库超级用户的权限。

如果你不想做 initdb，请执行以下程序。作为数据库超级用户，执行：

```
BEGIN;
UPDATE pg_proc SET proargtypes[3] = 'internal'::regtype
WHERE pronamespace = 11 AND pronargs = 5
     AND proargtypes[2] = 'cstring'::regtype;
-- The command should report having updated 90 rows;
-- if not, rollback and investigate instead of committing!
COMMIT;
```

上述程序必须在安装的每个数据库中执行，包括 `template1`，理想情况下也包括 `template0`。如果不修复模板数据库，之后创建的任何数据库都会包含相同的错误。`template1` 可以像任何其他数据库一样修复，但修复 `template0` 需要额外的步骤。首先，从任意数据库发出：

```
UPDATE pg_database SET datallowconn = true WHERE datname =
 'template0';
```

接下来连接到 `template0` 并执行上述修复程序。最后执行：

```
-- re-freeze template0:
VACUUM FREEZE;
-- and protect it against future alterations:
UPDATE pg_database SET datallowconn = false WHERE datname =
 'template0';
```

E.94.2. 变更

- 更改编码函数签名以防止误用
- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。
- 修复关系扩展与 VACUUM 之间的竞态条件

  这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
  没有已知案例造成超过断言失败的后果。
- 修复 `TIME WITH TIME ZONE` 值的比较

  在使用了 `--enable-integer-datetimes` 配置开关的情形下比较代码是错误的。注意：如果你在 `TIME WITH TIME ZONE` 列上有索引，安装此更新后需要对它执行 `REINDEX`，因为该修复纠正了列值的排序顺序。
- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- 修复 `INTERVAL` 值负小数秒的错误显示

  此错误只在使用了 `--enable-integer-datetimes` 配置开关时发生。
- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump，正确转储包含 `%` 的触发器名（Neil）
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）
- `contrib/intagg` 的更多 64 位修复
- 防止对返回 `RECORD` 的函数的错误优化

E.95. 版本 7.3.9

发行日期
2005-01-31

此发行版包含 7.3.8 之后的多种修复，包括几个与安全相关的问题。

E.95.1. 迁移到版本 7.3.9

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.95.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 检查聚合函数的创建者是否有权执行指定的转换函数

  这一疏漏使得绕过函数 EXECUTE 权限的拒绝成为可能。

- 修复 contrib/intagg 的安全和 64 位问题
- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 plperl 处理元组字段中引号的问题
- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示

E.96. 版本 7.3.8

发行日期
2004-10-22

此发行版包含 7.3.7 之后的多种修复。

E.96.1. 迁移到版本 7.3.8

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.96.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 make_oidjoins_check 在 /tmp 中使用临时文件

  这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

E.97. 版本 7.3.7

发行日期
2004-08-16

此发行版在 7.3.6 的基础上包含一个关键修复和一些次要项。

E.97.1. 迁移到版本 7.3.7

对于运行 7.3.X 的用户，不需要进行转储/恢复。

E.97.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 移除 tsearch 中不对称的词处理（Teodor）

- 在 pg_dump 转储 CAST 时正确地对函数名加模式限定

E.98. 版本 7.3.6

发行日期
2004-03-02

此发行版包含 7.3.5 之后的多种修复。

E.98.1. 迁移到版本 7.3.6

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.98.2. 变更

- 回退规则权限检查中的错误变更

7.3.3 中为修复规则权限检查极端情形而应用的补丁结果在许多并非极端的情形中禁用了与规则相关的权限检查。这例如会允许用户向他们本无权插入的视图插入。因此我们回退了 7.3.3 的补丁。原始缺陷将在 8.0 中修复。

- 修复 GetNewTransactionId() 中错误的操作顺序

  此缺陷可能导致磁盘空间耗尽条件下失败，包括释放磁盘空间后也无法重启。

- 确保即使提供了外部 CFLAGS 值，configure 也选择 -fno-strict-aliasing

  在某些平台上，用 -fstrict-aliasing 构建会导致缺陷。

- 让 pg_restore 正确处理 64 位 off_t

  此缺陷妨碍了从超过 4 GB 的归档文件正确恢复。

- 让 contrib/dblink 不假定本地和远程类型 OID 匹配（Joe）
- 正确引用 connectby() 的 start_with 参数（Joe）
- 当 plpgsql 函数的行类型参数为 NULL 时不崩溃
- 避免在规划 LIKE 操作时的极端情形生成无效字符编码序列
- 确保 text_position() 在多字节情形不能扫描越过源字符串末尾（Korea PostgreSQL Users' Group）
- 修复 bytea 列上 LIKE 操作的索引优化和选择率估计（Joe）

E.99. 版本 7.3.5

发行日期
2003-12-03

这包含 7.3.4 之后的多种修复。

E.99.1. 迁移到版本 7.3.5

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.99.2. 变更

- 强制在从 WAL 恢复期间开启 zero_damaged_pages
- 防止“变量不在子计划目标列表中”的一些晦涩情形
- 强制统计进程从共享内存脱离，确保更干净的关闭
- 使 PQescapeBytea 与 byteaout 相互一致（Joe）
- 为 dblink 的 get_tuple_of_interest() 添加缺失的 SPI_finish() 调用（Joe）

- 修复规则重写 INSERT 时可能的外键违反（Jan）
- 支持 PL/Tcl 的 spi_prepare 命令中限定类型名（Jan）
- 让 pg_dump 处理位于 pg_catalog 中的过程语言处理器
- 让 pg_dump 处理自定义操作符类在另一个模式中的情形
- 让 pg_dump 正确转储二进制兼容的转换（Jan）
- 修复把包含子查询的表达式插入规则体的问题
- 修复 clusterdb 脚本中不正确的参数处理（Anand Ranganathan）
- 修复 plpython 触发器中已删除列的问题
- 修复 to_char() 读取越过输入字符串末尾的问题（Karel）
- 修复 GB18030 映射错误（Tatsuo）
- 修复 SSL 错误处理和异步 SSL I/O 的几个问题
- 移除 JDBC 中把值列表绑定到单个参数的能力（防止可能的 SQL 注入攻击）
- 修复 HAVE_INT64_TIMESTAMP 代码路径中的一些错误
- 修复与首个根页分裂并行的 btree 搜索的极端情形

E.100. 版本 7.3.4

发行日期
2003-07-24

这包含 7.3.3 之后的多种修复。

E.100.1. 迁移到版本 7.3.4

对于运行 7.3.* 的用户，不需要进行转储/恢复。

E.100.2. 变更

- 修复 2000 年之前日期的时间戳到日期转换的损坏
- 防止服务器启动失败的罕见可能（Tom）
- 修复 interval 到 time 转换中的缺陷（Tom）
- 在 pg_dump 的几处添加约束名（Rod）
- 改进多参数函数的性能（Tom）
- 修复 to_ascii() 缓冲区溢出（Tom）

- 防止恢复数据库注释时抛出错误（Tom）
- 绕过某些 Solaris 版本中存在缺陷的 strxfrm()（Tom）
- 正确转义 jdbc setObject() 字符串以改进安全（Barry）

E.101. 版本 7.3.3

发行日期
2003-05-22

此发行版包含 7.3.2 版本的多种修复。

E.101.1. 迁移到版本 7.3.3

对于运行 7.3.* 版本的用户，不需要进行转储/恢复。

E.101.2. 变更

- 修复崩溃后 StartUpID 有时不正确的计算
- 避免一个事务中大量延迟触发器的缓慢（Stephan）
- 当 `UPDATE` 未更改外键值时不锁定被引用行（Jan）
- 在 Sparc 上使用 `-fPIC` 而不是 `-fpic`（Tom Callaway）
- 修复 contrib/reindexdb 缺乏模式感知的问题
- 修复 contrib/intarray 对零元素结果数组的错误（Teodor）
- 确保 createuser 脚本在 control-C 时退出（Oliver）
- 修复已删除列的类型本身也被删除时的错误
- `CHECKPOINT` 在非关键步骤失败时不再导致数据库 panic
- 接受 timestamp、time、interval 输入值秒字段中的 60
- 当 `TIMESTAMP`、`TIME` 或 `INTERVAL` 精度过大时发出通知而不是错误
- 修复 `abstime-to-time` 转换函数（除非 initdb，否则不应用修复）
- 修复 `timestampt_izone` 的 pg_proc 条目（除非 initdb，否则不应用修复）
- 让 `EXTRACT(EPOCH FROM timestamp without time zone)` 把输入视为本地时间
- 如果事务早些时候更改了时区，`'now'::timestamptz` 给出错误答案
- 带时区 time 的 `HAVE_INT64_TIMESTAMP` 代码覆写了其输入
- 接受 `GLOBAL TEMP/TEMPORARY` 作为 `TEMPORARY` 的同义词

- 避免外键触发器中不当的模式权限检查失败
- 修复 `SET DEFAULT` 动作的外键触发器中的缺陷
- 修复 `UPDATE` 和 `DELETE` 触发器行获取中不正确的 time-qual 检查
- `ALTER TABLE ADD COLUMN` 中外键子句被解析但被忽略
- 修复处理函数已存在时 createlang 脚本的损坏
- 修复 pg_dump、COPY、ANALYZE 等处对零列表的行为异常
- 修复 `func_error()` 对包含 '%' 的类型名的行为异常
- 修复 `replace()` 对包含 '%' 的字符串的行为异常
- 包含某些多字节字符的正则表达式模式失败
- 在连接大小估计中在更多情形正确考虑 `NULL`
- 避免与 `isblank()` 函数或宏的系统定义冲突
- 修复 EUC_TW 转换中未能转换大码点值的问题（Tatsuo）
- 修复 `SSL_read`/`SSL_write` 调用的错误恢复
- 不提前做类型强制表达式的常量折叠
- 读入任何页后立即验证页头字段
- 修复未命名连接中对未分组变量的不正确检查
- 修复 `to_ascii` 中的缓冲区溢出（Guido Notari）
- contrib/ltree 修复（Teodor）
- 修复 char 为无符号的机器上死锁检测的核心转储
- 避免多路索引扫描耗尽缓冲区（7.3 中引入的缺陷）
- 修复规划器的选择率估计函数以正确处理域
- 修复 dbmirror 内存分配缺陷（Steven Singer）
- 防止 `ln(numeric)` 因舍入误差无限循环
- 如果存在多个相等的 GROUP BY 项，`GROUP BY` 会混乱
- 修复继承的 `UPDATE`/`DELETE` 引用另一个继承表时的坏计划
- 防止在不完整（部分或不存储 NULL）的索引上聚簇
- 如果关闭请求在启动期间到达，在正确时刻处理它
- 修复临时索引中的左链接（可能使反向扫描漏掉条目）
- 修复 postgresql.conf 中 client_encoding 设置的不正确处理（Tatsuo）

- 修复 Async_NotifyHandler 运行后不响应 `pg_ctl stop -m fast` 的问题
- 修复规则包含多个同类型语句情形的 SPI
- 修复规则查询中检查错误类型访问权限的问题
- 修复 `CREATE RULE` 中 `EXCEPT` 的问题
- 防止删除带 serial 列的临时表的问题
- 修复复杂视图中 replace_vars_with_subplan_refs 的失败
- 修复单字节编码中正则表达式的缓慢（Tatsuo）
- 允许 `CREATE CAST` 和 `DROP CAST` 中的限定类型名
- 接受 `SETOF type[]`，以前必须写成 `SETOF _type`
- 修复某些带过程语言情形下 pg_dump 的核心转储
- 为可移植性在 pg_dump 输出中强制 ISO datestyle（Oliver）
- pg_dump 未能处理 `lo_read` 的错误返回（Oleg Drokin）
- pg_dumpall 对没有成员的组失败（Nick Eskelinen）
- pg_dumpall 未能识别 --globals-only 开关
- 如果指定了 -X disable-triggers，pg_restore 未能恢复 blob
- 修复 plpgsql 中的函数内内存泄漏
- 如果给定错误参数，pltcl 的 `elog` 命令转储核心（Ian Harding）
- plpython 使用了错误的 `atttypmod` 值（Brad McLean）
- 修复 Python 接口中布尔值的不当引用（D'Arcy）
- 为 JDBC 的 PGConnection 接口添加 `addDataType()` 方法
- 修复 JDBC 可更新 ResultSet 的各种问题（Shawn Green）
- 修复 JDBC DatabaseMetaData 的各种问题（Kris Jurka、Peter Royal）
- 修复 JDBC 中解析表 ACL 的问题
- JDBC 中字符集转换问题的更好错误消息

E.102. 版本 7.3.2

发行日期
2003-02-04

此发行版包含 7.3.1 版本的多种修复。

E.102.1. 迁移到版本 7.3.2

对于运行 7.3.* 版本的用户，不需要进行转储/恢复。

E.102.2. 变更

- 恢复 CREATE TABLE AS / SELECT INTO 中 OID 列的创建
- 修复转储带注释视图时 pg_dump 的核心转储
- 正确转储 DEFERRABLE/INITIALLY DEFERRED 约束
- 修复子表列编号与父表不同时的 UPDATE
- 提高 max_fsm_relations 的默认值
- 修复单行查询游标反向获取时的问题
- 让 SELECT DISTINCT 查询游标的反向获取正确工作
- 修复装载包含 contrib/lo 用法的 pg_dump 文件的问题
- 修复全数字用户名的问题
- 修复 libpgtcl 断开连接期间可能的内存泄漏和核心转储
- 让 plpython 的 spi_execute 命令正确处理空值（Andrew Bosma）
- 调整 plpython 错误报告使其回归测试再次通过
- 与 bison 1.875 一起工作
- 正确处理 plpgsql %type 中的大小写混合名（Neil）
- 修复执行被规则重写的查询时 pltcl 的核心转储
- 修复数组下标越界（据 Yichen Xie 报告）
- 把浮点情形的 MAX_TIME_PRECISION 从 13 减到 10
- 在逐数据库和逐用户设置中正确折叠变量名的大小写
- 修复 SELECT INTO 记录不返回行时 plpgsql RETURN NEXT 的核心转储
- 修复 python 客户端接口中对 pg_type.typprtlen 的过时使用
- 正确处理 JDBC 驱动中时间戳的小数秒
- 改进 JDBC 中 getImportedKeys() 的性能
- 使共享库符号链接在 HPUX 上按标准工作（Giles）
- 修复 timestamp、time、interval 不一致的舍入行为
- SSL 协商修复（Nathan Mueller）

- 使 libpq 的 ~/.pgpass 特性在用 PQconnectDB 连接时工作
- 更新 my2pg、ora2pg
- 翻译更新
- 在 contrib/lo 中添加 lo 与 oid 类型之间的转换
- fastpath 代码现在检查调用函数的权限

E.103. 版本 7.3.1

发行日期
2002-12-18

此发行版包含 7.3 版本的多种修复。

E.103.1. 迁移到版本 7.3.1

对于运行 7.3 版本的用户，不需要进行转储/恢复。但应注意，主 PostgreSQL 接口库 libpq 在此版本有了新的主版本号，某些情况下可能需要重新编译客户端代码。

E.103.2. 变更

- 修复客户端/服务器编码不匹配时 COPY TO 的核心转储（Tom）
- 允许 pg_dump 与 7.2 之前的服务器一起工作（Philip）
- contrib/adddepend 修复（Tom）
- 修复删除逐用户/逐数据库配置设置的问题（Tom）
- contrib/vacuumlo 修复（Tom）
- 即使 pg_shadow 包含 MD5 口令也允许 'password' 加密（Bruce）
- contrib/dbmirror 修复（Steven Singer）
- 优化器修复（Tom）
- contrib/tsearch 修复（Teodor Sigaev、Magnus）
- 允许区域设置名为大小写混合（Nicolai Tufar）
- 递增 libpq 库的主版本号（Bruce）
- pg_hba.conf 错误报告修复（Bruce、Neil）
- 新增 SCO Openserver 5.0.4 为受支持平台（Bruce）
- 防止 EXPLAIN 使服务器崩溃（Tom）

- SSL 修复（Nathan Mueller）
- 防止通过 ALTER TABLE 创建组合列（Tom）

E.104. 版本 7.3

发行日期
2002-11-27

E.104.1. 概述

此发行版的主要变更：

Schemas

模式允许用户在独立的命名空间中创建对象，因此两个人或应用程序可以拥有同名的表。还有用于共享表的 public 模式。可以通过移除 public 模式上的权限来限制表/索引的创建。

Drop Column

PostgreSQL 现在支持 `ALTER TABLE ... DROP COLUMN` 功能。

Table Functions

返回多行和/或多列的函数现在比以前好用得多。可以在 `SELECT` 的 `FROM` 子句中调用这样的“表函数”，把其输出当作表。此外，PL/pgSQL 函数现在可以返回集合。

Prepared Queries

PostgreSQL 现在支持预备查询，以改进性能。

Dependency Tracking

PostgreSQL 现在记录对象依赖，这在许多方面带来了改进。`DROP` 语句现在接受 `CASCADE` 或 `RESTRICT` 来控制是否也删除依赖对象。

Privileges

函数和过程语言现在有权限，函数可以定义为以其创建者的权限运行。

Internationalization

多字节和区域设置支持现在总是启用。

Logging

多种日志选项得到了增强。

Interfaces

大量接口已被移到 http://gborg.postgresql.org，在那里可以独立开发和发布。

Functions/Identifiers

默认情况下，函数现在最多可接受 32 个参数，标识符最长可达 63 字节。此外，`OPAQUE` 现已废弃：有特定的“伪数据类型”来表示函数参数和结果类型中 `OPAQUE` 以前的每种含义。

E.104.2. 迁移到版本 7.3

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。如果你的应用程序检查系统目录，由于 7.3 中引入了模式，还需要额外的更改；更多信息参见http://developer.postgresql.org/~momjian/upgrade_tips_7.3。

请注意以下不兼容之处：

- 不再支持 6.3 之前的客户端。
- `pg_hba.conf` 现在有用户名列和额外特性。现有文件需要调整。
- 几个 `postgresql.conf` 日志参数已被重命名。
- `LIMIT #,#` 已被禁用；请使用 `LIMIT # OFFSET #`。
- 带列列表的 `INSERT` 语句必须为每个指定的列提供值。例如 `INSERT INTO tab (col1, col2) VALUES ('val1')` 现在无效。如果 `INSERT` 没有列列表，仍允许提供比预期少的列。
- `serial` 列不再自动 `UNIQUE`；因此不会自动创建索引。
- 已中止事务内的 `SET` 命令现在会被回滚。
- `COPY` 不再把缺失的尾随列视为空。需要指定所有列。（不过，可以通过在 `COPY` 命令中指定列列表达到类似效果。）
- 数据类型 `timestamp` 现在等价于 `timestamp without time zone`，而不是 `timestamp with time zone`。
- 装载到 7.3 的 7.3 之前数据库不会有 `serial` 列、唯一约束和外键的新对象依赖。参见 `contrib/adddepend/` 目录中的详细说明和添加此类依赖的脚本。
- 不再允许空字符串（`''`）作为整数字段的输入。以前它被静默解释为 0。

E.104.3. 变更

E.104.3.1. 服务器操作

- 新增显示锁的 pg_locks 视图（Neil）
- 口令协商内存分配的安全修复（Neil）
- 移除对版本 0 FE/BE 协议的支持（PostgreSQL 6.2 及更早）（Tom）
- 为超级用户保留最后几个后端槽，新增参数 superuser_reserved_connections 来控制（Nigel J. Andrews）

E.104.3.2. 性能

- 通过只调用一次 localtime() 改进启动（Tom）

- 把系统目录信息缓存在平面文件中以加快启动（Tom）
- 改进索引信息的缓存（Tom）
- 优化器改进（Tom、Fernando Nasser）
- 目录缓存现在存储失败的查找（Tom）
- 哈希函数改进（Neil）
- 改进查询词法化和网络处理的性能（Peter）
- 大对象恢复的速度改进（Mario Weilguni）
- 首次查找时标记过期索引条目，节省后续堆获取（Tom）
- 避免过多的 NULL 位图填充（Manfred Koizar）
- 为 Solaris 添加 BSD 许可的 qsort()，以提升性能（Bruce）
- 把每行开销减少四字节（Manfred Koizar）
- 修复 GEQO 优化器缺陷（Neil Conway）
- 使 WITHOUT OID 实际每行节省四字节（Manfred Koizar）
- 新增 default_statistics_target 变量，指定 ANALYZE 桶数（Neil）
- 临时表使用本地缓冲区缓存，因此没有 WAL 开销（Tom）
- 改进大表上空闲空间映射的性能（Stephen Marshall、Tom）
- 改进的 WAL 写并发（Tom）

E.104.3.3. 权限

- 新增函数和过程语言上的权限（Peter）
- 为 CREATE DATABASE 新增 OWNER，使超级用户能代表非特权用户创建数据库（Gavin Sherry、Tom）
- 新增对象权限位 EXECUTE 和 USAGE（Tom）
- 新增 SET SESSION AUTHORIZATION DEFAULT 和 RESET SESSION AUTHORIZATION（Tom）
- 允许函数以其所有者的权限执行（Peter）

E.104.3.4. 服务器配置

- 服务器日志消息现在标记为 LOG 而不是 DEBUG（Bruce）
- 为 pg_hba.conf 添加用户列（Bruce）
- 让 log_connections 在日志文件中输出两行（Tom）
- 从 postgresql.conf 移除 debug_level，现为 server_min_messages（Bruce）

- 新的 ALTER DATABASE/USER ... SET 命令，用于逐用户/数据库初始化（Peter）
- 新参数 server_min_messages 和 client_min_messages，控制哪些消息发送到服务器日志或客户端应用程序（Bruce）
- 允许 pg_hba.conf 指定以逗号分隔的用户/数据库列表、以 + 为前缀的组名和以 @ 为前缀的文件名（Bruce）
- 移除辅助口令文件能力和 pg_password 工具（Bruce）
- 新增变量 db_user_namespace，用于数据库本地用户名（Bruce）
- SSL 改进（Bear Giles）
- 使存储口令的加密成为默认（Bruce）
- 允许通过调用 pg_stat_reset() 重置 pg_statistics（Christopher）
- 新增 log_duration 参数（Bruce）
- 把 debug_print_query 改名为 log_statement（Bruce）
- 把 show_query_stats 改名为 show_statement_stats（Bruce）
- 新增参数 log_min_error_statement，在出错时把命令打印到日志（Gavin）

E.104.3.5. 查询

- 使游标不敏感，即其内容不变（Tom）
- 禁用 LIMIT #,# 语法；现在只支持 LIMIT # OFFSET #（Bruce）
- 把标识符长度增加到 63（Neil、Bruce）
- UNION 修复：合并不同长度的 3 列以上（Tom）
- 为 INSERT 新增 DEFAULT 关键字，例如 INSERT ... (..., DEFAULT, ...)（Rod）
- 允许视图用 ALTER COLUMN ... SET DEFAULT 拥有默认值（Neil）
- 对列列表未提供所有列值的 INSERT 失败，例如 INSERT INTO tab (col1, col2) VALUES ('val1')；（Rod）
- 连接别名的修复（Tom）
- FULL OUTER JOIN 的修复（Tom）
- 改进无效标识符和位置的报告（Tom、Gavin）
- 修复 OPEN cursor(args)（Tom）
- 允许在视图中使用 'ctid' 以及 currtid(viewname)（Hiroshi）
- 带 UNION 的 CREATE TABLE AS 的修复（Tom）
- SQL99 语法改进（Thomas）

- 新增 statement_timeout 变量来取消查询（Bruce）
- 允许用 PREPARE/EXECUTE 预备查询（Neil）
- 允许 FOR UPDATE 出现在 LIMIT/OFFSET 之后（Bruce）
- 新增变量 autocommit（Tom、David Van Wie）

E.104.3.6. 对象操作

- 使 CREATE DATABASE 中的等号可选（Gavin Sherry）
- 让 ALTER TABLE OWNER 也更改索引所有权（Neil）
- 新 ALTER TABLE tabname ALTER COLUMN colname SET STORAGE 控制 TOAST 存储和压缩（John Gray）
- 新增模式支持，CREATE/DROP SCHEMA（Tom）
- 为临时表创建模式（Tom）
- 新增用于模式搜索的变量 search_path（Tom）
- 新增 ALTER TABLE SET/DROP NOT NULL（Christopher）
- 新 CREATE FUNCTION 易变性级别（Tom）
- 使规则名只在每个表内唯一（Tom）
- 为 DROP RULE 和 COMMENT ON RULE 新增 'ON tablename' 子句（Tom）
- 新增 ALTER TRIGGER RENAME（Joe）
- 新 current_schema() 和 current_schemas() 查询函数（Tom）
- 允许函数返回多行（表函数）（Joe）
- 为一致性使 CREATE DATABASE 中的 WITH 可选（Bruce）
- 新增对象依赖跟踪（Rod、Tom）
- 为 DROP 命令新增 RESTRICT/CASCADE（Rod）
- 为非 CHECK 约束新增 ALTER TABLE DROP（Rod）
- 删除带 SERIAL 的表时自动销毁序列（Rod）
- 如果列被外键使用则阻止删除列（Rod）
- 对象被删除时自动删除约束/函数（Rod）
- 新增 CREATE/DROP OPERATOR CLASS（Bill Studenmund、Tom）
- 新增 ALTER TABLE DROP COLUMN（Christopher、Tom、Hiroshi）
- 防止继承列被删除或重命名（Alvaro Herrera）
- 修复外键约束在中间数据库状态不报错（Stephan）

- 把列或表重命名传播到外键约束
- 新增 CREATE OR REPLACE VIEW（Gavin、Neil、Tom）
- 新增 CREATE OR REPLACE RULE（Gavin、Neil、Tom）
- 让规则按字母顺序执行，返回更可预测的值（Tom）
- 触发器现在按字母顺序触发（Tom）
- 新增 /contrib/adddepend 来处理 7.3 之前的对象依赖（Rod）
- 插入/更新值时允许更好的转换（Tom）

E.104.3.7. 工具命令

- 让 COPY TO 把嵌入的回车和换行输出为 \r 和 \n（Tom）
- 允许 COPY FROM 中的 DELIMITER 为 8 位干净（Tatsuo）
- 让 pg_dump 使用 ALTER TABLE ADD PRIMARY KEY 以提升性能（Neil）
- 禁用多语句规则中的方括号（Bruce）
- 禁止在函数内调用 VACUUM（Bruce）
- 允许 dropdb 和其他脚本使用带空格的标识符（Bruce）
- 把数据库注释更改限制在当前数据库
- 允许对操作符添加注释，独立于底层函数（Rod）
- 回退已中止事务中的 SET 命令（Tom）
- EXPLAIN 现在作为查询输出（Tom）
- 在 EXPLAIN 中显示条件表达式和排序键（Tom）
- 新增 'SET LOCAL var = value'，为单个事务设置配置变量（Tom）
- 允许 ANALYZE 在事务中运行（Bruce）
- 用新 WITH 子句改进 COPY 语法，保持向后兼容（Bruce）
- 修复 pg_dump 在非 ASCII 转储中一致输出标签（Bruce）
- 使转储文件中的外键约束更清晰（Rod）
- 新增 COMMENT ON CONSTRAINT（Rod）
- 允许 COPY TO/FROM 指定列名（Brent Verner）
- 把 UNIQUE 和 PRIMARY KEY 约束转储为 ALTER TABLE（Rod）
- 让 SHOW 输出查询结果（Joe）
- 对短 COPY 行产生失败而不是填充 NULL（Neil）

- 修复 CLUSTER 以保留所有表属性（Alvaro Herrera）
- 新 pg_settings 表，用于查看/修改 GUC 设置（Joe）
- 为 pg_dump 输出添加智能引用和可移植性改进（Peter）
- 把 serial 列转储为 SERIAL（Tom）
- 为 pg_dump 启用大文件支持（>2G）（Peter、Philip Warner、Bruce）
- 不允许对涉及参照约束的表使用 TRUNCATE（Rod）
- 让 TRUNCATE 也自动截断关系的 toast 表（Tom）
- 新增 clusterdb 工具，基于先前的 CLUSTER 操作自动聚簇整个数据库（Alvaro Herrera）
- 彻底改造 pg_dumpall（Peter）
- 允许对 TOAST 表执行 REINDEX（Tom）
- 按 SQL99 实现 START TRANSACTION（Neil）
- 修复页分裂影响批量删除时的罕见索引损坏（Tom）
- 修复继承的 ALTER TABLE ... ADD COLUMN（Alvaro Herrera）

E.104.3.8. 数据类型和函数

- 修复 factorial(0) 使其返回 1（Bruce）
- 日期/时间/时区改进（Thomas）
- 数组切片提取的修复（Tom）
- 修复 extract/date_part 为 timestamp 报告正确的微秒（Tatsuo）
- 允许 text_substr() 和 bytea_substr() 更高效地读取 TOAST 值（John Gray）
- 新增域支持（Rod）
- 使 WITHOUT TIME ZONE 成为 TIMESTAMP 和 TIME 数据类型的默认（Thomas）
- 允许用 configure 的 --enable-integer-datetimes 为日期/时间类型使用 64 位整数的替代存储方案（Thomas）
- 让 timezone(timestamptz) 返回 timestamp 而不是字符串（Thomas）
- 允许 1BC 之前日期的日期/时间类型带小数秒（Thomas）
- 把 timestamp 数据类型的精度限制为 6 位小数（Thomas）
- 把时区转换函数从 timetz() 改为 timezone()（Thomas）
- 新增配置变量 datestyle 和 timezone（Tom）
- 新增 OVERLAY()，允许替换字符串中的子串（Thomas）
- 新增 SIMILAR TO（Thomas、Tom）

- 新增正则表达式 SUBSTRING(string FROM pat FOR escape)（Thomas）
- 新增 LOCALTIME 和 LOCALTIMESTAMP 函数（Thomas）
- 用 CREATE TYPE typename AS (column) 新增命名组合类型（Joe）
- 允许在表别名子句中定义组合类型（Joe）
- 新增 API 以简化 C 语言表函数的创建（Joe）
- 从与标准不匹配的 SQL99 函数调用中移除 ODBC 兼容的空括号（Thomas）
- 允许 macaddr 数据类型接受无分隔符的 12 个十六进制数字（Mike Wyer）
- 新增 CREATE/DROP CAST（Peter）
- 新增 IS DISTINCT FROM 操作符（Thomas）
- 新增 SQL99 TREAT() 函数，CAST() 的同义词（Thomas）
- 新增 pg_backend_pid()，输出后端 pid（Bruce）
- 新增 IS OF / IS NOT OF 类型谓词（Thomas）
- 允许不完全指定长度的位串常量（Thomas）
- 允许 8 字节整数与位串之间的转换（Thomas）
- 实现十六进制字面量到位串字面量的转换（Thomas）
- 允许表函数出现在 FROM 子句中（Joe）
- 把函数参数的最大数量增加到 32（Bruce）
- 不再自动为 SERIAL 列创建索引（Tom）
- 新增 current_database()（Rod）
- 修复 cash_words() 使其不溢出缓冲区（Tom）
- 新增函数 replace()、split_part()、to_hex()（Joe）
- 修复 bytea 作为右侧参数的 LIKE（Joe）
- 防止 SELECT cash_out(2) 导致的崩溃（Tom）
- 修复 to_char(1,'FM999.99') 使其返回句点（Karel）
- 修复返回 OPAQUE 的触发器/类型/语言函数，使其返回正确的类型（Tom）

E.104.3.9. 国际化

- 新增编码：韩语（JOHAB）、泰语（WIN874）、越南语（TCVN）、阿拉伯语（WIN1256）、简体中文（GBK）、韩语（UHC）（Eiji Tokuya）
- 默认启用区域设置支持（Peter）
- 新增区域设置变量（Peter）

- 在 PQescapeBytea/PQunescapeBytea 中为多字节转义 >= 0x7f 的字节（Tatsuo）
- 为正则表达式字符类添加区域感知
- 默认启用多字节支持（Tatsuo）
- 新增 GB18030 多字节支持（Bill Huang）
- 新增 CREATE/DROP CONVERSION，允许可装载的编码（Tatsuo、Kaori）
- 新增 pg_conversion 表（Tatsuo）
- 新增 SQL99 CONVERT() 函数（Tatsuo）
- pg_dumpall、pg_controldata 和 pg_resetxlog 现在支持国家语言（Peter）
- 新增和更新的翻译

E.104.3.10. 服务器端语言

- 允许递归 SQL 函数（Peter）
- 更改 PL/Tcl 构建以使用配置的编译器和 Makefile.shlib（Peter）
- 彻底改造 PL/pgSQL 的 FOUND 变量以更兼容 Oracle（Neil、Tom）
- 允许 PL/pgSQL 处理带引号的标识符（Tom）
- 允许返回集合的 PL/pgSQL 函数（Neil）
- 使 PL/pgSQL 感知模式（Joe）
- 移除一些内存泄漏（Nigel J. Andrews、Tom）

E.104.3.11. psql

- 为兼容 7.2.0，psql \connect 不再把数据库名转为小写（Tom）
- 新增 psql \timing 来为用户查询计时（Greg Sabino Mullane）
- 让 psql \d 显示索引信息（Greg Sabino Mullane）
- 新 psql \dD 显示域（Jonathan Eisler）
- 允许 psql 显示视图上的规则（Paul ?）
- psql 变量替换的修复（Tom）
- 允许 psql \d 显示临时表结构（Tom）
- 允许 psql \d 显示外键（Rod）
- 修复 \? 以遵循 \pset pager（Bruce）
- 让 psql 在启动时报告其版本号（Tom）
- 允许 \copy 指定列名（Tom）

E.104.3.12. libpq

- 新增 ~/.pgpass 来存储主机/用户口令组合（Alvaro Herrera）
- 为 libpq 新增 PQunescapeBytea() 函数（Patrick Welche）
- 修复通过非阻塞连接发送大查询的问题（Bernhard Herzog）
- 修复 libpq 在 Win9X 上使用定时器的问题（David Ford）
- 允许 libpq notify 处理标识符长度不同的服务器（Tom）
- 把 libpq 的 PQescapeString() 和 PQescapeBytea() 添加到 Windows（Bruce）
- 修复非阻塞连接的 SSL 问题（Jack Bates）
- 新增 libpq 连接超时参数（Denis A Ustimenko）

E.104.3.13. JDBC

- 允许 JDBC 用 JDK 1.4 编译（Dave）
- 新增 JDBC 3 支持（Barry）
- 允许 JDBC 通过在连接 URL 中添加 ?loglevel=X 来设置 loglevel（Barry）
- 新增打印版本号的 Driver.info() 消息（Barry）
- 新增可更新结果集（Raghu Nidagal、Dave）
- 新增对可调用语句的支持（Paul Bethe）
- 新增查询取消能力
- 新增 refresh row（Dave）
- 修复多字节服务器的 MD5 加密处理（Jun Kawai）
- 新增对预备语句的支持（Barry）

E.104.3.14. 其他接口

- 修复 ECPG 关于单引号中八进制数的缺陷（Michael）
- 把 src/interfaces/libpgeasy 移到 http://gborg.postgresql.org（Marc、Bruce）
- 改进 Python 接口（Elliot Lee、Andrew Johnson、Greg Copeland）
- 新增 libpgtcl 连接关闭事件（Gerhard Hintermayer）
- 把 src/interfaces/libpq++ 移到 http://gborg.postgresql.org（Marc、Bruce）
- 把 src/interfaces/odbc 移到 http://gborg.postgresql.org（Marc）
- 把 src/interfaces/libpgeasy 移到 http://gborg.postgresql.org（Marc、Bruce）
- 把 src/interfaces/perl5 移到 http://gborg.postgresql.org（Marc、Bruce）

- 从主树移除 src/bin/pgaccess，现位于 http://www.pgaccess.org（Bruce）
- 为 libpgtcl 新增 pg_on_connection_loss 命令（Gerhard Hintermayer、Tom）

E.104.3.15. 源代码

- 并行 make 的修复（Peter）
- 链接 Tcl 的 AIX 修复（Andreas Zeugswetter）
- 允许 PL/Perl 在 Cygwin 下构建（Jason Tishler）
- 改进 MIPS 编译（Peter、Oliver Elphick）
- 要求 Autoconf 2.53 版本（Peter）
- configure 默认要求 readline 和 zlib（Peter）
- 允许 Solaris 使用紧密共享内存（ISM）以提升性能（Scott Brunza、P.J. Josh Rovero）
- 编译时总是启用 syslog，移除 --enable-syslog 选项（Tatsuo）
- 编译时总是启用多字节，移除 --enable-multibyte 选项（Tatsuo）
- 编译时总是启用区域设置，移除 --enable-locale 选项（Peter）
- Win9x DLL 创建的修复（Magnus Naeslund）
- 修复 WAL 代码在 Windows、BeOS 上使用 link() 的问题（Jason Tishler）
- 把 sys/types.h 添加到 c.h，从主文件移除（Peter、Bruce）
- 修复 SMP 机器上的 AIX 挂起（Tomoyuki Niijima）
- AIX SMP 挂起修复（Tomoyuki Niijima）
- 修复较新 glibc 库上 1970 年之前的日期处理（Tom）
- 修复 PowerPC SMP 锁定（Tom）
- 防止使用 gcc -ffast-math（Peter、Tom）
- 开发者构建现在要求 Bison >= 1.50
- Kerberos 5 支持现在能用 Heimdal 构建（Peter）
- 在用户指南中新增列出 SQL 特性的附录（Thomas）
- 改进可装载模块链接以使用 RTLD_NOW（Tom）
- 新错误级别 WARNING、INFO、LOG、DEBUG[1-5]（Bruce）
- 新 src/port 目录存放替代的 libc 函数（Peter、Bruce）
- 用于模式的新 pg_namespace 系统目录（Tom）
- 为模式新增 pg_class.relnamespace（Tom）

- 为模式新增 pg_type.typnamespace（Tom）
- 为模式新增 pg_proc.pronamespace（Tom）
- 重构聚合使其有 pg_proc 条目（Tom）
- 系统关系现在有自己的命名空间，不需要 pg_* 测试（Fernando Nasser）
- 把 TOAST 索引名改为 *_index 而不是 *_idx（Neil）
- 为操作符、操作符类新增命名空间（Tom）
- 为服务器控制文件新增额外检查（Thomas）
- 新波兰语 FAQ（Marcin Mazurek）
- 新增 Posix 信号量支持（Tom）
- 记录重新索引的需要（Bruce）
- 重命名一些内部标识符以简化 Windows 编译（Jan、Katherine Ward）
- 新增关于计算磁盘空间的文档（Bruce）
- 从 GUC 移除 KSQO（Bruce）
- 修复 rtree 中的内存泄漏（Kenneth Been）
- 为一致性修改一些错误消息（Bruce）
- 移除未使用的系统表列（Peter）
- 在适当处使系统列为 NOT NULL（Tom）
- 清理 sprintf 的使用，改用 snprintf()（Neil、Jukka Holappa）
- 移除 OPAQUE 并创建特定子类型（Tom）
- 数组内部处理的清理（Joe、Tom）
- 不允许 pg_atoi('')（Bruce）
- 移除参数 wal_files，因为 WAL 文件现在被回收（Bruce）
- 为堆页新增版本号（Tom）

E.104.3.16. 附加模块

- 允许 /contrib/array 中的 inet 数组（Neil）
- GiST 修复（Teodor Sigaev、Neil）
- 升级 /contrib/mysql
- 新增 /contrib/dbsize，无需清理即可显示表大小（Peter）
- 新增 /contrib/intagg，整数聚合例程（mlw）

- 改进 /contrib/oid2name（Neil、Bruce）
- 改进 /contrib/tsearch（Oleg、Teodor Sigaev）
- /contrib/rserver 的清理（Alexey V. Borzov）
- 更新 /contrib/oracle 转换工具（Gilles Darold）
- 更新 /contrib/dblink（Joe）
- 改进 /contrib/vacuumlo 支持的选项（Mario Weilguni）
- /contrib/intarray 的改进（Oleg、Teodor Sigaev、Andrey Oktyabrski）
- 新增 /contrib/reindexdb 工具（Shaun Thomas）
- 为 /contrib/isbn_issn 新增索引（Dan Weston）
- 新增 /contrib/dbmirror（Steven Singer）
- 改进 /contrib/pgbench（Neil）
- 新增 /contrib/tablefunc 表函数示例（Joe）
- 新增用于树结构的 /contrib/ltree 数据类型（Teodor Sigaev、Oleg Bartunov）
- 把 /contrib/pg_controldata、pg_resetxlog 移入主树（Bruce）
- /contrib/cube 的修复（Bruno Wolff）
- 改进 /contrib/fulltextindex（Christopher）

E.105. 版本 7.2.8

发行日期
2005-05-09

此发行版包含 7.2.7 之后的多种修复，包括一个与安全相关的问题。

E.105.1. 迁移到版本 7.2.8

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.105.2. 变更

- 修复古老的竞态条件，它允许事务在某些用途（如 SELECT FOR UPDATE）上比其他用途稍早被视为已提交

  这是一个极其严重的缺陷，因为它可能导致应用程序短暂看到表面的数据不一致。

- 修复关系扩展与 VACUUM 之间的竞态条件

这在理论上可能导致丢失一页刚插入的数据，尽管该情形概率似乎很低。
没有已知案例造成超过断言失败的后果。

- 修复 `TIME WITH TIME ZONE` 值的 `EXTRACT(EPOCH)`
- plpgsql 中的额外缓冲区溢出检查（Neil）
- 修复 pg_dump 正确转储包含 `%` 的索引名和触发器名（Neil）
- 防止 `to_char(interval)` 在月份相关格式上转储核心
- 修复 `contrib/pgcrypto` 在较新 OpenSSL 构建上的问题（Marko Kreen）

E.106. 版本 7.2.7

发行日期
2005-01-31

此发行版包含 7.2.6 之后的多种修复，包括几个与安全相关的问题。

E.106.1. 迁移到版本 7.2.7

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.106.2. 变更

- 不允许非超级用户使用 `LOAD`

  在会自动执行共享库初始化函数的平台（至少包括 Windows 和基于 ELF 的 Unix）上，`LOAD` 可被用来让服务器执行任意代码。感谢 NGS Software 报告此问题。

- 为一些 contrib 函数添加所需的 STRICT 标记（Kris Jurka）
- 避免 plpgsql 游标声明参数过多时的缓冲区溢出（Neil）
- 修复 FULL 和 RIGHT 外连接的规划错误

  连接的结果被错误地认为与左输入的排序相同。这不仅会向用户交付排序错误的输出，
  而且在嵌套归并连接的情形下可能给出完全错误的答案。

- 修复 SQL 和 GERMAN datestyle 下负 interval 的显示

E.107. 版本 7.2.6

发行日期
2004-10-22

此发行版包含 7.2.5 之后的多种修复。

E.107.1. 迁移到版本 7.2.6

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.107.2. 变更

- 修复磁盘上提示位可能未更新的问题

  极少数情况下，这一疏漏可能导致“无法访问事务状态”失败，
  这使它成为一个潜在的数据丢失缺陷。

- 确保哈希外连接不漏掉元组

  在恰当的数据分布下，使用哈希连接计划的超大型左连接可能无法输出未匹配的左侧行。

- 不允许以 root 运行 pg_ctl

  这是为了防范任何可能的安全问题。

- 避免 make_oidjoins_check 在 /tmp 中使用临时文件

  这被报告为安全问题，不过几乎不值得担心，因为非开发者本来就没有理由使用此脚本。

- 更新到更新版本的 Bison

E.108. 版本 7.2.5

发行日期
2004-08-16

此发行版包含 7.2.4 之后的多种修复。

E.108.1. 迁移到版本 7.2.5

对于运行 7.2.X 的用户，不需要进行转储/恢复。

E.108.2. 变更

- 防止崩溃期间已提交事务的可能丢失

  由于事务提交与检查点之间的互锁不足，
  最近一次检查点之前刚提交的事务可能在数据库崩溃和重启后全部或部分丢失。这是自 PostgreSQL 7.1 以来就存在的严重缺陷。

- 修复与首个根页分裂并行的 btree 搜索的极端情形

- 修复 `to_ascii` 中的缓冲区溢出（Guido Notari）

- 修复 char 为无符号的机器上死锁检测的核心转储

- 修复 Async_NotifyHandler 运行后不响应 `pg_ctl stop -m fast` 的问题
- 修复 pg_dump 中的内存泄漏
- 避免与 `isblank()` 函数或宏的系统定义冲突

E.109. 版本 7.2.4

发行日期
2003-01-30

此发行版包含 7.2.3 版本的多种修复，包括防止可能数据丢失的修复。

E.109.1. 迁移到版本 7.2.4

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.109.2. 变更

- 修复 VACUUM "No one parent tuple was found" 错误的一些额外情形
- 禁止在函数内调用 VACUUM（Bruce）
- 确保在标记检查点完成之前 pg_clog 更新已同步到磁盘
- 避免大型哈希连接期间的整数溢出
- 使 GROUP 命令在 pg_group.grolist 大到需要 TOAST 时也能工作
- 修复日期时间表中的错误；一些时区名未被识别
- 修复 circle_poly()、path_encode()、path_add() 中的整数溢出（Neil）
- 修复 lseg_eq()、lseg_ne()、lseg_center() 中长期存在的逻辑错误

E.110. 版本 7.2.3

发行日期
2002-10-01

此发行版包含 7.2.2 版本的多种修复，包括防止可能数据丢失的修复。

E.110.1. 迁移到版本 7.2.3

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.110.2. 变更

- 防止压缩事务日志的可能丢失（Tom）
- 防止非超级用户推进最近的清理信息（Tom）
- 处理较新版本 glibc 中 1970 年之前的日期值（Tom）
- 修复服务器关闭期间可能的挂起
- 防止 SMP PPC 机器上的自旋锁挂起（Tomoyuki Niijima）
- 修复 pg_dump 正确转储 FULL JOIN USING（Tom）

E.111. 版本 7.2.2

发行日期
2002-08-23

此发行版包含 7.2.1 版本的多种修复。

E.111.1. 迁移到版本 7.2.2

对于运行 7.2.* 版本的用户，不需要进行转储/恢复。

E.111.2. 变更

- 允许在 PL/pgSQL 中 EXECUTE "CREATE TABLE AS ... SELECT"（Tom）
- 修复压缩事务日志 id 回卷的问题（Tom）
- 修复 PQescapeBytea/PQunescapeBytea 使其处理 > 0x7f 的字节（Tatsuo）
- 修复 psql 和 pg_dump 以不存在的长选项调用时的崩溃（Tatsuo）
- 修复调用几何操作符时的崩溃（Tom）
- 允许 OPEN cursor(args)（Tom）
- rtree_gist 索引构建的修复（Teodor）
- 转储用户定义聚合的修复（Tom）
- contrib/intarray 修复（Oleg）
- 修复使用括号的复杂 UNION/EXCEPT/INTERSECT 查询（Tom）
- pg_convert 的修复（Tatsuo）
- 长 DATA 字符串崩溃的修复（Thomas、Neil）
- repeat()、lpad()、rpad() 与长字符串的修复（Neil）

E.112. 版本 7.2.1

发行日期
2002-03-21

此发行版包含 7.2 版本的多种修复。

E.112.1. 迁移到版本 7.2.1

对于运行 7.2 版本的用户，不需要进行转储/恢复。

E.112.2. 变更

- 确保序列计数器在崩溃后不会倒退（Tom）
- 修复 pgaccess 汉字转换键绑定（Tatsuo）
- 优化器改进（Tom）
- Cash I/O 改进（Tom）
- 新俄语 FAQ
- 缺失 AuthBlockSig 的编译修复（Heiko）
- 附加时区和时区修复（Thomas）
- 允许 psql \connect 处理大小写混合的数据库和用户名（Tom）
- 即使有 ON INSERT 规则在命令完成时也返回正确的 OID（Tom）
- 允许 COPY FROM 使用 8 位 DELIMITERS（Tatsuo）
- 修复 extract/date_part 对毫秒/微秒的缺陷（Tatsuo）
- 改进多个不同长度 UNION 的处理（Tom）
- contrib/btree_gist 改进（Teodor Sigaev）
- contrib/tsearch 词典改进，附加安装步骤见 README.tsearch（Thomas T. Thai、Teodor Sigaev）
- 数组下标处理的修复（Tom）
- 允许在 PL/pgSQL 中 EXECUTE "CREATE TABLE AS ... SELECT"（Tom）

E.113. 版本 7.2

发行日期
2002-02-04

E.113.1. 概述

此发行版改进了 PostgreSQL 在高流量应用程序中的使用。

此发行版的主要变更：

VACUUM

清理不再锁定表，因此允许在清理期间正常的用户访问。新 `VACUUM FULL` 命令通过锁定表并收缩表的磁盘副本执行旧式清理。

Transactions

超过四十亿事务的安装不再有问题。

OIDs

OID 现在是可选的。用户现在可以在 OID 使用过多的情形下创建不带 OID 的表。

Optimizer

系统现在在 `ANALYZE` 期间计算直方图列统计，使优化器的选择好得多。

Security

新的 MD5 加密选项允许更安全地存储和传输口令。Linux 和 BSD 系统上有新的 Unix 域套接字认证选项。

Statistics

管理员可以使用新的表访问统计模块获取表和索引使用的细粒度信息。

Internationalization

程序和库消息现在可以用多种语言显示。

E.113.2. 迁移到版本 7.2

希望从任何以前的版本迁移数据的用户，需要使用 `pg_dump` 进行转储/恢复。

请注意以下不兼容之处：

- `VACUUM` 命令的语义在此版本中已更改。你可能需要相应地更新维护程序。
- 在此版本中，使用 `= NULL` 的比较总是返回假（更准确地说是 NULL）。以前的版本自动把此语法转换为 `IS NULL`。旧行为可以用 `postgresql.conf` 参数重新启用。
- `pg_hba.conf` 和 `pg_ident.conf` 配置现在只在收到 SIGHUP 信号后重新装载，而不是每次连接时。
- 函数 `octet_length()` 现在返回未压缩的数据长度。
- 日期/时间值 `'current'` 不再可用。你需要重写你的应用程序。
- `timestamp()`、`time()` 和 `interval()` 函数不再可用。不要用 `timestamp()`，请改用 `timestamp 'string'` 或 `CAST`。

`SELECT ... LIMIT #,#` 语法将在下一个版本中移除。你应把查询改为使用独立的 LIMIT 和 OFFSET 子句，例如 `LIMIT 10 OFFSET 20`。

E.113.3. 变更

E.113.3.1. 服务器操作

- 在单独的目录中创建临时文件（Bruce）
- postmaster 启动时删除孤立的临时文件（Bruce）
- 为一些系统表添加唯一索引（Tom）
- 系统表操作符重组（Oleg Bartunov、Teodor Sigaev、Tom）
- 把 pg_log 改名为 pg_clog（Tom）
- 启用 SIGTERM、SIGQUIT 来终止后端（Jan）
- 移除后端数量的编译期限制（Tom）
- 信号量资源失败时更好的清理（Tatsuo、Tom）
- 允许安全的事务 ID 回卷（Tom）
- 从一些系统表移除 OID（Tom）
- 移除"触发数据更改违反"错误检查（Tom）
- 预备/保存计划的 SPI 门户创建（Jan）
- 允许 SPI 列函数用于系统列（Tom）
- 长值压缩改进（Tom）
- 表、索引访问的统计收集器（Jan）
- 把超长序列名截断为合理的值（Tom）
- 以毫秒测量事务时间（Thomas）
- 修复 TID 顺序扫描（Hiroshi）
- 超级用户 ID 现在固定为 1（Peter E）
- pg_ctl 新 "reload" 选项（Tom）

E.113.3.2. 性能

- 优化器改进（Tom）
- 优化器的新直方图列统计（Tom）
- 重用预写日志文件而不是丢弃它们（Tom）
- 缓存改进（Tom）

- IS NULL、IS NOT NULL 优化器改进（Tom）
- 改进锁管理器以减少锁争用（Tom）
- 为索引访问支持函数保留 relcache 条目（Tom）
- 使 NUMERIC 中的 NaN 和无穷大有更好的选择率（Tom）
- R-tree 性能改进（Kenneth Been）
- B-tree 分裂更高效（Tom）

E.113.3.3. 权限

- 把 UPDATE、DELETE 权限改为彼此独立（Peter E）
- 新 REFERENCES、TRIGGER 权限（Peter E）
- 允许一次对多个用户 GRANT/REVOKE（Peter E）
- 新 has_table_privilege() 函数（Joe Conway）
- 允许非超级用户清理数据库（Tom）
- 新 SET SESSION AUTHORIZATION 命令（Peter E）
- 修复新建表上权限修改的缺陷（Tom）
- 禁止非超级用户访问 pg_statistic，新增用户可访问的视图（Tom）

E.113.3.4. 客户端认证

- 在认证前 fork postmaster 以防止挂起（Peter E）
- 在 Linux、*BSD 上通过 Unix 域套接字新增 ident 认证（Helge Bahmann、Oliver Elphick、Teodor Sigaev、Bruce）
- 新增使用 MD5 加密的口令认证方法（Bruce）
- 允许用 MD5 加密存储的口令（Bruce）
- PAM 认证（Dominic J. Eidson）
- 只在启动和 SIGHUP 时装载 pg_hba.conf 和 pg_ident.conf（Bruce）

E.113.3.5. 服务器配置

- 一些时区缩写解释为澳大利亚而不是北美，现在可在运行时设置（Bruce）
- 设置默认事务隔离级别的新参数（Peter E）
- 启用把 "expr = NULL" 转换为 "expr IS NULL" 的新参数，默认关闭（Peter E）
- 控制 VACUUM 内存使用的新参数（Tom）
- 设置客户端认证超时的新参数（Tom）
- 设置最大打开文件数的新参数（Tom）

E.113.3.6. 查询

- INSERT 规则添加的语句现在在 INSERT 之后执行（Jan）
- 阻止目标列表中不加修饰的关系名（Bruce）
- ORDER BY 中 NULL 现在排在所有正常值之后（Tom）
- 新 IS UNKNOWN、IS NOT UNKNOWN 布尔测试（Tom）
- 新 SHARE UPDATE EXCLUSIVE 锁模式（Tom）
- 新 EXPLAIN ANALYZE 命令，显示运行时间和行数（Martijn van Oosterhout）
- 修复 LIMIT 与子查询的问题（Tom）
- 修复推入子查询的 LIMIT、DISTINCT ON（Tom）
- 修复嵌套的 EXCEPT/INTERSECT（Tom）

E.113.3.7. 模式操作

- 修复临时表中的 SERIAL（Bruce）
- 允许临时序列（Bruce）
- 序列现在内部使用 int8（Tom）
- 新 SERIAL8 创建带序列的 int8 列，默认仍为 SERIAL4（Tom）
- 用 WITHOUT OIDS 使 OID 可选（Tom）
- 为 CREATE TYPE 新增 %TYPE 语法（Ian Lance Taylor）
- 为 CHECK 约束新增 ALTER TABLE / DROP CONSTRAINT（Christopher Kings-Lynne）
- 新 CREATE OR REPLACE FUNCTION 更改现有函数（保留函数 OID）（Gavin Sherry）
- 新增 ALTER TABLE / ADD [UNIQUE | PRIMARY]（Christopher Kings-Lynne）
- 允许视图中重命名列
- 让 ALTER TABLE / RENAME COLUMN 更新索引的列名（Brent Verner）
- 修复继承表的 ALTER TABLE / ADD CONSTRAINT ... CHECK（Stephan Szabo）
- ALTER TABLE RENAME 正确更新外键触发器参数（Brent Verner）
- DROP AGGREGATE 和 COMMENT ON AGGREGATE 现在接受 aggtype（Tom）
- 为 SQL 函数新增自动返回类型数据转换（Tom）
- 允许 GiST 索引处理 NULL 和多键索引（Oleg Bartunov、Teodor Sigaev、Tom）
- 启用部分索引（Martijn van Oosterhout）

E.113.3.8. 工具命令

- 新增 RESET ALL、SHOW ALL（Marko Kreen）

- CREATE/ALTER USER/GROUP 现在允许选项任意顺序（Vince）
- 新增 LOCK A, B, C 功能（Neil Padgett）
- CREATE/ALTER USER 的新 ENCRYPTED/UNENCRYPTED 选项（Bruce）
- 新的轻量 VACUUM 不锁定表；旧语义以 VACUUM FULL 提供（Tom）
- 禁用视图上的 COPY TO/FROM（Bruce）
- COPY DELIMITERS 字符串必须恰好一个字符（Tom）
- 关于索引元组少于堆的 VACUUM 警告现在只在适当时出现（Martijn van Oosterhout）
- 修复 CREATE INDEX 的权限检查（Tom）
- 禁止不当使用 CREATE/DROP INDEX/TRIGGER/VIEW（Tom）

E.113.3.9. 数据类型和函数

- SUM()、AVG()、COUNT() 现在内部使用 int8 以提升速度（Tom）
- 新增 convert()、convert2()（Tatsuo）
- 新函数 bit_length()（Peter E）
- 使 CHAR(n)/VARCHAR(n) 中的 "n" 表示字符数而不是字节数（Tatsuo）
- CHAR()、VARCHAR() 现在拒绝过长的字符串（Peter E）
- BIT VARYING 现在拒绝过长的位串（Peter E）
- BIT 现在拒绝与声明大小不匹配的位串（Peter E）
- INET、CIDR 文本转换函数（Alex Pilosov）
- INET、CIDR 操作符 << 和 <<= 可索引（Alex Pilosov）
- Bytea \### 现在要求有效的三位八进制数
- Bytea 比较改进，现在支持 =、<>、>、>=、< 和 <=
- Bytea 现在支持 B-tree 索引
- Bytea 现在支持 LIKE、LIKE...ESCAPE、NOT LIKE、NOT LIKE...ESCAPE
- Bytea 现在支持连接
- 新 bytea 函数：position、substring、trim、btrim 和 length
- 新 encode() 函数模式 "escaped"，在最小转义的 bytea 与 text 之间转换
- 新增 pg_database_encoding_max_length()（Tatsuo）
- 新增 pg_client_encoding() 函数（Tatsuo）
- now() 返回毫秒精度的时间（Thomas）
- 新 TIMESTAMP WITHOUT TIMEZONE 数据类型（Thomas）

- 新增带 "T" 的 ISO 日期/时间规格，yyyy-mm-ddThh:mm:ss（Thomas）
- 新 xid/int 比较函数（Hiroshi）
- 为 TIME、TIMESTAMP 和 INTERVAL 数据类型新增精度（Thomas）
- 修改类型强制逻辑，先尝试二进制兼容函数（Tom）
- 新 encode() 函数默认安装（Marko Kreen）
- 改进的 to_*() 转换函数（Karel Zak）
- 优化单字节编码下的 LIKE/ILIKE（Tatsuo）
- contrib/pgcrypto 中的新函数：crypt()、hmac()、encrypt()、gen_salt()（Marko Kreen）
- 修正 translate() 函数的描述（Bruce）
- 为 SET TIME ZONE 新增 INTERVAL 参数（Thomas）
- 新增 INTERVAL YEAR TO MONTH（等）语法（Thomas）
- 优化单字节编码下的长度函数（Tatsuo）
- 修复 path_inter、path_distance、path_length、dist_ppath 以处理闭合路径（Curtis Barrett、Tom）
- octet_length(text) 现在返回未压缩的长度（Tatsuo、Bruce）
- 处理日期/时间字面量中的 "July" 全名（Greg Sabino Mullane）
- 一些 datatype() 函数调用现在的求值方式不同
- 新增对儒略和 ISO 时间规格的支持（Thomas）

E.113.3.10. 国际化

- psql、pg_dump、libpq 和服务器的国家语言支持（Peter E）
- 中文（简体、繁体）、捷克语、法语、德语、匈牙利语、俄语、瑞典语的消息翻译（Peter E、Serguei A. Mokhov、Karel Zak、Weiping He、Zhenbang Wei、Kovacs Zoltan）
- 使 trim、ltrim、rtrim、btrim、lpad、rpad、translate 感知多字节（Tatsuo）
- 新增 LATIN5,6,7,8,9,10 支持（Tatsuo）
- 新增 ISO 8859-5,6,7,8 支持（Tatsuo）
- 更正 LATIN5 为 ISO-8859-9 而不是 ISO-8859-5（Tatsuo）
- 使 mic2ascii() 不感知非 ASCII（Tatsuo）
- 拒绝无效的多字节字符序列（Tatsuo）

E.113.3.11. PL/pgSQL

- 现在 SELECT 循环使用门户，允许超大结果集（Jan）

- CURSOR 和 REFCURSOR 支持（Jan）
- 现在可以返回打开的游标（Jan）
- 新增 ELSEIF（Klaus Reger）
- 改进 PL/pgSQL 错误报告，包括错误位置（Tom）
- 为兼容允许游标声明中的 IS 或 FOR 关键字（Bruce）
- SELECT ... FOR UPDATE 的修复（Tom）
- 返回多行的 PERFORM 的修复（Tom）
- 让 PL/pgSQL 使用服务器的类型强制代码（Tom）
- 内存泄漏修复（Jan、Tom）
- 使尾部分号可选（Tom）

E.113.3.12. PL/Perl

- 新不可信 PL/Perl（Alex Pilosov）
- 即使 libperl 不是共享的，PL/Perl 现在也能在一些平台上构建（Peter E）

E.113.3.13. PL/Tcl

- 现在报告 errorInfo（Vsevolod Lobko）
- 新增 spi_lastoid 函数（bob@redivi.com）

E.113.3.14. PL/Python

- ……是新的（Andrew Bosma）

E.113.3.15. psql

- \d 按唯一、主键分组显示索引（Christopher Kings-Lynne）
- 允许反斜杠命令中的尾部分号（Greg Sabino Mullane）
- 如果可能从 /dev/tty 读取口令
- 更改用户和数据库时强制新的口令提示（Tatsuo、Tom）
- 为 Unicode 格式化正确的列数（Patrice）

E.113.3.16. libpq

- 新函数 PQescapeString()，转义命令字符串中的引号（Florian Weimer）
- 新函数 PQescapeBytea()，转义二进制字符串以用作 SQL 字符串字面量

E.113.3.17. JDBC

- 返回 INSERT 的 OID（Ken K）

- 处理更多数据类型（Ken K）
- 处理字符串中的单引号和换行（Ken K）
- 处理 NULL 变量（Ken K）
- 时区处理的修复（Barry Lind）
- 改进的 Druid 支持
- 允许非多字节服务器使用 8 位字符（Barry Lind）
- 支持 BIT、BINARY 类型（Ned Wolpert）
- 降低内存使用（Michael Stephens、Dave Cramer）
- 更新 DatabaseMetaData（Peter E）
- 新增 DatabaseMetaData.getCatalogs()（Peter E）
- 编码修复（Anders Bengtsson）
- Get/setCatalog 方法（Jason Davies）
- DatabaseMetaData.getColumns() 现在返回列默认值（Jason Davies）
- DatabaseMetaData.getColumns() 性能改进（Jeroen van Vianen）
- 一些 JDBC1 和 JDBC2 的合并（Anders Bengtsson）
- 事务性能改进（Barry Lind）
- 数组修复（Greg Zoller）
- 序列化添加
- 批处理修复（Rene Pijlman）
- ExecSQL 方法重组（Anders Bengtsson）
- GetColumn() 修复（Jeroen van Vianen）
- 修复 isWriteable() 函数（Rene Pijlman）
- 更好地通过 JDBC2 一致性测试（Rene Pijlman）
- 新增 bytea 类型能力（Barry Lind）
- 新增 isNullable()（Rene Pijlman）
- JDBC 日期/时间测试套件修复（Liam Stewart）
- SELECT 'id' AS xxx FROM table 的修复（Dave Cramer）
- 修复 DatabaseMetaData 以正确显示精度（Mark Lillywhite）
- 新 getImported/getExported 键（Jason Davies）
- MD5 口令加密支持（Jeremy Wohl）

- 修复为真正使用类型缓存（Ned Wolpert）

E.113.3.18. ODBC

- 移除查询大小限制（Hiroshi）
- 移除文本字段大小限制（Hiroshi）
- 多字节模式下 SQLPrimaryKeys 的修复（Hiroshi）
- 允许 ODBC 过程调用（Hiroshi）
- 改进布尔处理（Aidan Mountford）
- 大多数配置选项现在可通过 DSN 设置（Hiroshi）
- 多字节、性能修复（Hiroshi）
- 允许驱动与 iODBC 或 unixODBC 一起使用（Peter E）
- MD5 口令加密支持（Bruce）
- 为 odbc.sql 添加更多兼容函数（Peter E）

E.113.3.19. ECPG

- 实现 EXECUTE ... INTO（Christof Petig）
- 多行描述符支持（如 CARDINALITY）（Christof Petig）
- GRANT 参数的修复（Lee Kindness）
- 修复 INITIALLY DEFERRED 缺陷
- 各种缺陷修复（Michael、Christof Petig）
- 指示符变量数组的自动分配（int *ind_p=NULL）
- 字符串数组的自动分配（char **foo_pp=NULL）
- 修复 ECPGfree_auto_mem
- 所有有外部链接的函数名现在都以 ECPG 为前缀
- 结构数组的修复（Michael）

E.113.3.20. 其他接口

- Python 修复 fetchone()（Gerhard Haring）
- 在适当处在 Tcl 中使用 UTF、Unicode（Vsevolod Lobko、Reinhard Max）
- 新增 Tcl COPY TO/FROM（ljb）
- 防止 pg_dump 输出默认索引操作符类（Tom）
- 修复 libpgeasy 内存泄漏（Bruce）

E.113.3.21. 构建和安装

- Configure、动态装载器和共享库修复（Peter E）
- QNX 4 移植的修复（Bernd Tegge）
- Cygwin 和 Windows 移植的修复（Jason Tishler、Gerhard Haring、Dmitry Yurtaev、Darko Prenosil、Mikhail Terekhov）
- Windows 套接字通信失败的修复（Magnus、Mikhail Terekhov）
- Hurd 编译修复（Oliver Elphick）
- BeOS 修复（Cyril Velter）
- 移除 configure --enable-unicode-conversion，现由多字节启用（Tatsuo）
- AIX 修复（Tatsuo、Andreas）
- 修复并行 make（Peter E）
- 把 SQL 语言手册页安装到操作系统特定的目录（Peter E）
- 把 config.h 改名为 pg_config.h（Peter E）
- 重新组织头文件的安装布局（Peter E）

E.113.3.22. 源代码

- 移除 SEP_CHAR（Bruce）
- 新 GUC 钩子（Tom）
- 合并 GUC 和命令行处理（Marko Kreen）
- 移除 EXTEND INDEX（Martijn van Oosterhout、Tom）
- 新的缩进 java 代码的 pgjindent 工具（Bruce）
- 在 C++ 下编译时移除 true/false 的定义（Leandro Fanzone、Tom）
- pgindent 修复（Bruce、Tom）
- 在适当处用 strcmp() 取代 strcasecmp()（Peter E）
- Dynahash 可移植性改进（Tom）
- 在自旋锁结构中添加 'volatile' 用法
- 改进信号处理逻辑（Tom）

E.113.3.23. 附加模块

- 新 contrib/rtree_gist（Oleg Bartunov、Teodor Sigaev）
- 新 contrib/tsearch 全文索引（Oleg、Teodor Sigaev）
- 新增用于远程数据库访问的 contrib/dblink（Joe Conway）

- contrib/ora2pg Oracle 转换工具（Gilles Darold）
- contrib/xml XML 转换工具（John Gray）
- contrib/fulltextindex 修复（Christopher Kings-Lynne）
- 新 contrib/fuzzystrmatch，含 levenshtein 和 metaphone，合并了 soundex（Joe Conway）
- 为 contrib/intarray 新增布尔查询、二分搜索和修复（Oleg Bartunov）
- 新 pg_upgrade 工具（Bruce）
- 新增 pg_resetxlog 选项（Bruce、Tom）

E.114. 版本 7.1.3

发行日期
2001-08-15

E.114.1. 迁移到版本 7.1.3

对于运行 7.1.X 的用户，不需要进行转储/恢复。

E.114.2. 变更

移除大事务未使用的 WAL 段（Tom）
多动作规则修复（Tom）
PL/pgSQL 内存分配修复（Jan）
VACUUM 缓冲区修复（Tom）
回归测试修复（Tom）
pg_dump 对视图上 GRANT/REVOKE/注释及用户定义类型的修复（Tom）
修复带 DISTINCT ON 或 LIMIT 的子查询（Tom）
BeOS 修复
禁用视图上的 COPY TO/FROM（Tom）
Cygwin 构建（Jason Tishler）

E.115. 版本 7.1.2

发行日期
2001-05-11

这包含 7.1.1 之后的一个修复。

E.115.1. 迁移到版本 7.1.2

对于运行 7.1.X 的用户，不需要进行转储/恢复。

E.115.2. 变更

修复返回零行时的 PL/pgSQL SELECT
修复 psql 反斜杠核心转储
参照完整性权限修复
优化器修复
pg_dump 清理

E.116. 版本 7.1.1

发行日期
2001-05-05

这包含 7.1 之后的多种修复。

E.116.1. 迁移到版本 7.1.1

对于运行 7.1 的用户，不需要进行转储/恢复。

E.116.2. 变更

修复 numeric MODULO 操作符 (Tom)
pg_dump 修复 (Philip)
pg_dump 可转储 7.0 数据库 (Philip)
readline 4.2 修复 (Peter E)
JOIN 修复 (Tom)
AIX、MSWIN、VAX、N32K 修复 (Tom)
多字节修复 (Tom)
Unicode 修复 (Tatsuo)
优化器改进 (Tom)
修复函数中的整行 (Tom)
修复 pg_ctl 及带空格的选项字符串 (Peter E)
ODBC 修复 (Hiroshi)
EXPLAIN 现在可接受字符串参数 (Thomas)
Python 修复 (Darcy)

E.117. 版本 7.1

发行日期
2001-04-13

此发行版专注于移除 PostgreSQL 代码中存在多年的限制。

此发行版的主要变更：

Write-ahead Log (WAL)

为了在操作系统崩溃时保持数据库一致性，以前的 PostgreSQL 版本强制在每次事务提交前把所有数据修改写到磁盘。有了 WAL，只需把一个日志文件刷到磁盘，大幅改进了性能。如果你在以前的版本中一直用 -F 禁用磁盘刷写，可能要考虑停止使用它。

TOAST

TOAST——以前的版本有编译期行长度限制，通常为 8k - 32k。这一限制使长文本字段的存储很困难。有了 TOAST，任意长度的长行都能以良好的性能存储。

Outer Joins

我们现在支持外连接。不再需要用 UNION/NOT IN 变通实现外连接。我们使用 SQL92 外连接语法。

Function Manager

以前的 C 函数管理器不能正确处理空值，也不支持 64 位 CPU（Alpha）。新的函数管理器可以。你可以继续使用旧的自定义函数，
但将来可能想重写它们以使用新的函数管理器调用接口。

Complex Queries

以前版本不支持的大量复杂查询现在可以工作。视图、聚合、UNION、LIMIT、游标、
子查询和继承表的许多组合现在都能正常工作。继承表现在默认被访问。FROM 中的子查询现在受支持。

E.117.1. 迁移到版本 7.1

希望从任何以前的版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。

E.117.2. 变更

```
缺陷修复
---------
大量多字节/Unicode/区域设置修复 (Tatsuo 等)
更可靠的 ALTER TABLE RENAME (Tom)
Kerberos V 修复 (David Wragg)
修复目标列表带子查询的 INSERT INTO...SELECT (Tom)
在标准错误上提示用户名/口令 (Bruce)
大对象 inv_read/inv_write 修复 (Tom)
to_char()、to_date()、to_ascii() 和 to_timestamp() 的修复 (Karel、
   Daniel Baldoni)
防止查询表达式泄漏内存 (Tom)
允许更新数组元素 (Tom)
取消时唤醒锁等待者 (Hiroshi)
修复使用哈希连接时游标的罕见崩溃 (Tom)
修复回滚事务中的 DROP TABLE/INDEX (Hiroshi)
修复启用 MULTIBYTE 时 \l+ 导致 psql 崩溃的问题 (Peter E)
修复 CREATE VIEW 期间规则名截断的问题 (Ross Reedstrom)
修复 PL/perl (Alex Kapranoff)
禁止视图上的 LOCK (Mark Hollomon)
禁止视图上的 INSERT/UPDATE/DELETE (Mark Hollomon)
禁止视图上的 DROP RULE、CREATE INDEX、TRUNCATE (Mark Hollomon)
允许 PL/pgSQL 接受非 ASCII 标识符 (Tatsuo)
```

允许视图正确处理 GROUP BY、聚合、DISTINCT (Tom)
修复 TRUNCATE 命令的罕见失败 (Tom)
允许 UNION/INTERSECT/EXCEPT 与 ALL、子查询、视图、
DISTINCT、ORDER BY、SELECT...INTO 一起使用 (Tom)
修复中止事务期间的解析器失败 (Tom)
允许临时关系正确清理索引 (Bruce)
修复 VACUUM 在同页移动行的问题 (Tom)
修改 pg_dump 以更好地处理 template1 中的用户定义项 (Philip)
允许 VIEW 中使用 LIMIT (Tom)
要求游标 FETCH 遵循 LIMIT (Tom)
允许继承列上的 PRIMARY/FOREIGN KEY 定义 (Stephan)
允许子查询中的 ORDER BY、LIMIT (Tom)
允许 CREATE RULE 中的 UNION (Tom)
使 ALTER/DROP TABLE 可回滚 (Vadim、Tom)
把 initdb 的排序规则存储在 pg_control 中使排序规则不可更改 (Tom)
修复带规则的 INSERT...SELECT (Tom)
修复视图和子查询内的 FOR UPDATE (Tom)
修复 OVERLAPS 操作符在 NULL 方面符合 SQL92 规范 (Tom)
修复 lpad() 和 rpad() 以处理长度小于输入字符串的情形 (Tom)
修复某些规则中 NOTIFY 的使用 (Tom)
彻底改造 btree 代码 (Tom)
修复 PL/pgSQL 变量中 NOT NULL 的使用 (Tom)
彻底改造 GIST 代码 (Oleg)
修复 CLUSTER 以保留约束和列默认值 (Tom)
改进的死锁检测处理 (Tom)
允许表中有多个 SERIAL 列 (Tom)
防止偶发索引损坏 (Vadim)

增强

添加 OUTER JOIN (Tom)
函数管理器彻底改造 (Tom)
允许对索引 ALTER TABLE RENAME (Tom)
改进 CLUSTER (Tom)
改进更多平台的 ps 状态显示 (Peter E、Marc)
改进 CREATE FUNCTION 失败消息 (Ross)
JDBC 改进 (Peter、Travis Bauer、Christopher Cain、William Webber、
Gunnar)
大一统配置方案/GUC。现在许多选项可以在
data/postgresql.conf、postmaster/postgres 标志或 SET 命令中设置 (Peter
E)
改进文件描述符缓存的处理 (Tom)
关于自动创建表别名条目的新警告代码 (Bruce)
彻底改造 initdb 流程 (Tom、Peter E)
继承表的彻底改造；继承表现在默认被访问；
新 ONLY 关键字可阻止 (Chris Bitmead、Tom)
ODBC 清理/改进 (Nick Gorham、Stephan Szabo、Zoltan Kovacs、
Michael Fork)
允许重命名临时表 (Tom)
彻底改造内存管理器上下文 (Tom)
pg_dumpall 使用 CREATE USER 或 CREATE GROUP 而不是 COPY (Peter E)
彻底改造 pg_dump (Philip Warner)
允许 pg_hba.conf 辅助口令文件只指定用户名 (Peter E)
创建临时表时允许 TEMPORARY 或 TEMP 关键字 (Bruce)

新内存泄漏检查器 (Karel)
新 SET SESSION CHARACTERISTICS (Thomas)
允许嵌套块注释 (Thomas)
添加 WITHOUT TIME ZONE 类型限定符 (Thomas)
新 ALTER TABLE ADD CONSTRAINT (Stephan)
INTEGER 聚合使用 NUMERIC 累加器 (Tom)
彻底改造聚合代码 (Tom)
新 VARIANCE 和 STDDEV() 聚合
改进 pg_dump 的依赖排序 (Philip)
新 pg_restore 命令 (Philip)
新 pg_dump tar 输出选项 (Philip)
新 pg_dump 大对象支持 (Philip)
LIKE 的新 ESCAPE 选项 (Thomas)
新不区分大小写的 LIKE—ILIKE (Thomas)
允许函数索引使用二进制兼容类型 (Tom)
允许 SQL 函数用于更多上下文 (Tom)
新 pg_config 工具 (Peter E)
新 PL/pgSQL EXECUTE 命令，允许动态 SQL 和工具语句
(Jan)
新 PL/pgSQL GET DIAGNOSTICS 语句用于 SPI 值访问 (Jan)
新 quote_identifiers() 和 quote_literal() 函数 (Jan)
新 ALTER TABLE table OWNER TO user 命令 (Mark Hollomon)
允许 FROM 中的子查询，即 FROM (SELECT ...) [AS] alias (Tom)
把 PyGreSQL 更新到 3.1 版 (D'Arcy)
以 OID 命名的文件存储表 (Vadim)
新 SQL 函数 setval(seq,val,bool) 供 pg_dump 使用 (Philip)
要求用 DROP VIEW 删除视图，不再用 DROP TABLE (Mark)
允许 DROP VIEW view1, view2 (Mark)
允许 DROP INDEX、DROP RULE 和 DROP TYPE 中的多个对象 (Tom)
允许与 Unicode 的自动转换 (Tatsuo、Eiji)
新 /contrib/pgcrypto 哈希函数 (Marko Kreen)
新 pg_dumpall --globals-only 选项 (Peter E)
WAL 的新 CHECKPOINT 命令，创建新 WAL 日志文件 (Vadim)
新 AT TIME ZONE 语法 (Thomas)
允许配置 Unix 域套接字的位置 (David J. MacKenzie)
允许 postmaster 监听特定 IP 地址 (David J. MacKenzie)
允许通过前导斜杠在主机名中指定套接字路径名
(David J. MacKenzie)
允许 CREATE DATABASE 指定模板数据库 (Tom)
新工具把 MySQL 模式转储转换为 SQL92 和 PostgreSQL (Thomas)
新 /contrib/rserv 复制工具包 (Vadim)
COPY BINARY 的新文件格式 (Tom)
新 /contrib/oid2name 把数字文件映射到表名 (B Palmer)
新 "idle in transaction" ps 状态消息 (Marc)
更新到 pgaccess 0.98.7 (Constantin Teodorescu)
pg_ctl 关闭时现在默认 -w (等待)，新 -l (日志) 选项
向 pg_dump 添加初步的依赖检查 (Philip)

类型

修复 INET/CIDR 类型排序并添加新函数 (Tom)
使 OID 表现为无符号类型 (Tom)
允许 BIGINT 作为 INT8 的同义词 (Peter E)
新 int2 和 int8 比较操作符 (Tom)

新 BIT 和 BIT VARYING 类型 (Adriaan Joubert、Tom、Peter E)
因 TOAST, CHAR() 不再比 VARCHAR() 快 (Tom)
新 GIST seg/cube 示例 (Gene Selkov)
改进的 round(numeric) 处理 (Tom)
修复 CIDR 输出格式化 (Tom)
新 CIDR abbrev() 函数 (Tom)

性能

预写日志 (WAL) 以更少性能开销提供崩溃恢复
    (Vadim)
VACUUM 的 ANALYZE 阶段不再独占锁定表 (Bruce)
减少文件寻道 (Denis Perchine)
改进重复键的 BTREE 代码 (Tom)
把所有大对象存储在单个表中 (Denis Perchine、Tom)
改进内存分配性能 (Karel、Tom)

源代码

新函数管理器调用约定 (Tom)
SGI 可移植性修复 (David Kaelbling)
新 configure --enable-syslog 选项 (Peter E)
新 BSDI README (Bruce)
configure 脚本移到顶级而不是 /src (Peter E)
Makefile/配置/编译彻底改造 (Peter E)
新 configure --with-python 选项 (Peter E)
Solaris 清理 (Peter E)
彻底改造 /contrib Makefile (Karel)
新 OpenSSL 配置选项 (Magnus、Peter E)
AIX 修复 (Andreas)
QNX 修复 (Maurizio)
新 heap_open()、heap_openr() API (Tom)
移除冒号和分号操作符 (Thomas)
视图的新 pg_class.relkind 值 (Mark Hollomon)
把 ichar() 改名为 chr() (Karel)
btrim()、ascii()、chr()、repeat() 的新文档 (Karel)
NT/Cygwin 修复 (Pete Forman)
AIX 移植修复 (Andreas)
新 BeOS 移植 (David Reid、Cyril Velter)
把校对者的更改添加到文档 (Addison-Wesley、Bruce)
新 Alpha 自旋锁代码 (Adriaan Joubert、Compaq)
UnixWare 移植彻底改造 (Peter E)
新 Darwin/Mac OS X 移植 (Peter Bierman、Bruce Hartzler)
新 FreeBSD Alpha 移植 (Alfred)
彻底改造共享内存段 (Tom)
添加 IBM S/390 支持 (Neale Ferguson)
把 macmanuf 移到 /contrib (Larry Rosenman)
Syslog 改进 (Larry Rosenman)
不含用户添加内容的新 template0 数据库 (Tom)
新 /contrib/cube 和 /contrib/seg GIST 示例代码 (Gene Selkov)
允许用 NetBSD 的 libedit 代替 readline (Peter)
改进的汇编语言源代码格式 (Bruce)
新 contrib/pg_logger
createdb 的新 --template 选项

新 contrib/pg_control 工具 (Oliver)
新 FreeBSD 工具 ipc_check、start-scripts/freebsd

E.118. 版本 7.0.3

发行日期
2000-11-11

这包含 7.0.2 之后的多种修复。

E.118.1. 迁移到版本 7.0.3

对于运行 7.0.* 的用户，不需要进行转储/恢复。

E.118.2. 变更

Jdbc 修复 (Peter)
大对象修复 (Tom)
修复 COPY WITH OIDS 泄漏的倾向 (Tom)
修复反向索引扫描 (Tom)
修复 SELECT ... FOR UPDATE 使其检查重复键 (Hiroshi)
为 configure 添加 --enable-syslog (Marc)
修复罕见情形下后端退出时中止事务的问题 (Tom)
启用多字节时 psql \l+ 的修复 (Tatsuo)
允许 PL/pgSQL 接受非 ASCII 标识符 (Tatsuo)
使 vacuum 总是刷新缓冲区 (Tom)
修复以允许在等待锁时取消 (Hiroshi)
用户认证代码中内存分配问题的修复 (Tom)
移除对 int4out() 的虚假使用 (Tom)
COALESCE 或 BETWEEN 中多个子查询的修复 (Tom)
某些情形下堆打开时触发器失败的修复 (Jeroen van Vianen)
修复不相等的选择率错误 (Tom)
修复 strcmp() 的错误使用 (Tom)
修复存储管理器访问文件末尾之外条目的缺陷 (Tom)
修复以在所有 smgr elog 消息中包含内核 errno 消息 (Tom)
构建时 '.' 不在 PATH 中的修复 (SL Baur)
文件描述符耗尽错误的修复 (Tom)
修复以使 pg_dump 转储函数的 'iscachable' 标志 (Tom)
Append 节点目标列表中子查询的修复 (Tom)
归并连接计划的修复 (Tom)
修复带索引关系上的 TRUNCATE 失败 (Tom)
避免写错误时数据库范围重启 (Hiroshi)
修复 nodeMaterial 通过重算输出以遵循 chgParam (Tom)
修复当行版本的源和目的地位于同一页时 VACUUM 移动更新行版本链的问题 (Tom)
修复 user.c CommandCounterIncrement (Tom)
to_char() 中 AM/PM 边界问题的修复 (Karel Zak)
修复 TIME 聚合处理 (Tom)
修复 to_char() 以避免对 NULL 输入核心转储 (Tom)
缓冲区修复 (Tom)
修复向 char() 数据类型插入/复制较长多字节字符串的问题 (Tatsuo)

修复中止时后端崩溃的问题（Tom）

E.119. 版本 7.0.2

发行日期
2000-06-05

这是 7.0.1 的重新打包，附加了文档。

E.119.1. 迁移到版本 7.0.2

对于运行 7.* 的用户，不需要进行转储/恢复。

E.119.2. 变更

已把文档添加到压缩包。

E.120. 版本 7.0.1

发行日期
2000-06-01

这是 7.0 的清理版本。

E.120.1. 迁移到版本 7.0.1

对于运行 7.0 的用户，不需要进行转储/恢复。

E.120.2. 变更

修复许多 CLUSTER 失败（Tom）
允许 ALTER TABLE RENAME 用于索引（Tom）
修复 plpgsql 处理 datetime->timestamp 和 timespan->interval（Bruce）
新的 configure --with-setproctitle 开关以使用 setproctitle()（Marc、Bruce）
修复 6.5.3 起 ResultSet 的差一错误等
jdbc ResultSet 修复（Joseph Shraibman）
优化器调优（Tom）
为 pgaccess 修复 create user
修复 UNLISTEN 失败
IRIX 修复（David Kaelbling）
QNX 修复（Andreas Kardos）
降低 COPY IN 锁级别（Tom）
把 libpqeasy 改为使用 PQconnectdb() 风格参数（Bruce）
修复 pg_dump 处理 OID 索引（Tom）
修复小内存泄漏（Tom）

Solaris 上 createdb/dropdb 的修复 (Tatsuo)
非阻塞连接修复 (Alfred Perlstein)
修复 RENAME TABLE 失败后的不当恢复 (Tom)
把 pg_ident.conf.sample 复制到安装的 /lib 目录 (Bruce)
新增 SJIS UDC (NEC selection IBM 汉字) 支持 (Eiji Tokuya)
修复过长的 syslog 消息 (Tatsuo)
修复带引号索引名过长的问题 (Tom)
JDBC ResultSet.getTimestamp() 修复 (Gregory Krasnow & Floyd Marinescu)
ecpg 变更 (Michael)

E.121. 版本 7.0

发行日期
2000-05-08

此发行版包含许多领域的改进，展示了 PostgreSQL 的持续成长。7.0 的改进和修复比以往任何版本都多。开发者确信这是迄今最好的版本；我们尽力只发布扎实的版本，这个版本也不例外。

此发行版的主要变更：

Foreign Keys

外键现在已实现，PARTIAL MATCH 外键除外。许多用户一直要求此特性，我们很高兴提供它。

Optimizer Overhaul

延续一年前开始的工作，优化器得到改进，允许更好的查询计划选择，性能更快且内存使用更少。

Updated psql

我们的交互式终端监视器 psql 已更新，新增多种特性。详情见 psql 手册页。

Join Syntax

SQL92 连接语法现在受支持，但此版本只支持 INNER JOIN。JOIN、NATURAL JOIN、JOIN/USING 和 JOIN/ON 都可用，列相关名也可用。

E.121.1. 迁移到版本 7.0

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。对于从 6.5.* 升级的用户，可以改用 pg_upgrade 升级到此版本；但完整的转储/重装安装始终是最可靠的升级方法。

新版本需要考虑的接口和兼容性问题包括：

- 日期/时间类型 datetime 和 timespan 已被 SQL92 定义的 timestamp 和 interval 取代。虽然我们做了一些努力来简化过渡，允许 PostgreSQL 识别废弃的类型名并转换为新类型名，但这一机制对你的现有应用程序不可能是完全透明的。
- 优化器在查询代价估计方面得到显著改进。在某些情况下，由于优化器为首选计划做出更好的选择，查询时间会缩短。

但在少数情形下（通常涉及病态的数据分布），查询时间可能上升。如果你处理大量数据，可能想检查查询以验证性能。

- JDBC 和 ODBC 接口已升级和扩展。
- 字符串函数 CHAR_LENGTH 现在是原生函数。以前的版本把它转换为对 LENGTH 的调用，这可能与实现 LENGTH 的其他类型（如几何类型）产生歧义。

E.121.2. 变更

```
缺陷修复
---------
防止函数调用超过最大参数数量 (Tom)
改进 CASE 构造 (Tom)
修复 SELECT coalesce(f1,0) FROM int4_tbl GROUP BY f1 (Tom)
修复 SELECT sentence.words[0] FROM sentence GROUP BY sentence.words[0]
 (Tom)
修复 GROUP BY 扫描缺陷 (Tom)
SQL 语法处理的改进 (Tom)
修复涉及 INSERT ... SELECT ... 的视图的问题 (Tom)
修复 SELECT a/2, a/2 FROM test_missing_target GROUP BY a/2 (Tom)
修复 INSERT ... SELECT 中的子查询 (Tom)
防止 INSERT ... SELECT ... ORDER BY (Tom)
大于 2GB 关系的修复，包括 vacuum
改进系统表变更向其他后端的传播 (Tom)
改进用户表变更向其他后端的传播 (Tom)
修复复杂情形下临时表的处理 (Bruce、Tom)
允许在表打开时锁定表，改进并发可靠性 (Tom)
在 pg_dump 中正确引用序列名 (Ross J. Reedstrom)
防止他人访问时 DROP DATABASE
若无行被处理则防止 GROUP BY 返回任何行 (Tom)
修复无行匹配 WHERE 时 SELECT COUNT(1) FROM table WHERE ... (Tom)
修复 pg_upgrade 使其适用于 MVCC (Tom)
修复 SELECT ... WHERE x IN (SELECT ... HAVING SUM(x) > 1) (Tom)
修复 "f1 datetime DEFAULT 'now'" (Tom)
修复 DEFAULT 中使用 CURRENT_DATE 的问题 (Tom)
允许纯注释行，也允许 ;;; 行。 (Tom)
改进磁盘写入失败、磁盘满后的恢复 (Hiroshi)
修复 FROM 中提到表但未连接的情形 (Tom)
允许无聚合函数的 HAVING 子句 (Tom)
修复 "--" 注释且无尾随换行的问题，如 perl 接口中所见
改进 pg_dump 失败错误报告 (Bruce)
允许排序和哈希超过 2GB 文件大小 (Tom)
修复 pg_dump 转储继承规则的问题 (Tom)
修复 NULL 处理比较的问题 (Tom)
修复失败 CREATE/DROP 命令导致的不一致状态 (Hiroshi)
修复带连字符的数据库名的问题
防止 DROP INDEX 干扰其他后端 (Tom)
修复 verify_password() 中的文件描述符泄漏
修复 "Unable to identify an operator =$" 问题
修复 ODBC 在启用 CommLog 和 Debug 时不再段错误 (Dirk Niggemann)
修复递归 exit 调用 (Massimo)
修复超长时区的问题 (Jeroen van Vianen)
使 pg_dump 保留主键信息 (Peter E)
```

防止带单引号的数据库名 (Peter E)
防止事务内 DROP DATABASE (Peter E)
ecpg 内存泄漏修复 (Stephen Birch)
修复 SELECT null::text、SELECT int4fac(null) 和 SELECT 2 + (null) (Tom)
Y2K 时间戳修复 (Massimo)
修复 VACUUM 'HEAP_MOVED_IN was not expected' 错误 (Tom)
修复含空格表/列名的视图的问题 (Tom)
防止索引上的权限 (Peter E)
修复生成错误时自旋锁卡住的问题 (Hiroshi)
修复 Linux 上的 ipcclean
修复 NULL 约束条件的处理 (Tom)
修复 odbc 驱动中的内存泄漏 (Nick Gorham)
修复 UNION 表上的权限检查 (Tom)
修复以允许 SELECT 'a' LIKE 'a' (Tom)
修复 SELECT 1 + NULL (Tom)
CHAR 的修复
修复 numeric 类型上的 log() (Tom)
废弃 ':' 和 ';' 操作符
允许清理临时表
不允许与新列同名的继承列
磁盘空间耗尽时恢复或强制失败 (Hiroshi)
修复 AS 列与结果列匹配的 INSERT INTO ... SELECT
修复 INSERT ... SELECT ... GROUP BY 按目标列而非源列分组的问题 (Tom)
修复带 INSERT 的 CREATE TABLE test (a char(5) DEFAULT text '', b int4) (Tom)
修复带 LIMIT 的 UNION
修复 CREATE TABLE x AS SELECT 1 UNION SELECT 2
修复 CREATE TABLE test(col char(2) DEFAULT user)
修复 CREATE TABLE ... DEFAULT 中的类型不匹配
修复 SELECT * FROM pg_class where oid in (0,-1)
修复 SELECT COUNT('asdf') FROM pg_class WHERE oid=12
防止能创建数据库的用户修改 pg_database 表 (Peter E)
修复 btree 在键 > 1/2 (页 - 开销) 时给出有用的 elog (Tom)
修复向 DECIMAL(4,4) 字段插入 0.0 的问题 (Tom)

增强

基于 SQL3/SQL98 的新 CLI 接口头文件 sqlcli.h
移除查询长度的所有限制，行长度限制仍存在 (Tom)
把 jdbc 协议更新到 2.0 (Jens Glaser <jens@jens.de>)
添加 TRUNCATE 命令以快速截断关系 (Mike Mascari)
修复以给超级用户和 createdb 用户适当的更新目录权限 (Peter E)
允许 ecpg bool 变量具有 NULL 值 (Christof)
对无 NULL 指示符变量的 NULL 值发出 ecpg 错误 (Christof)
允许 ^C 取消 COPY 命令 (Massimo)
添加 SET FSYNC 和 SHOW PG_OPTIONS 命令 (Massimo)
动态装载 C 函数的函数名重载 (Frankpitt)
向 libpq++ 添加 CmdTuples() (Vince)
新 CREATE CONSTRAINT TRIGGER 和 SET CONSTRAINTS 命令 (Jan)
允许 CREATE FUNCTION/WITH 子句用于所有语言类型
configure --enable-debug 添加 -g (Peter E)
configure --disable-debug 移除 -g (Peter E)
允许更复杂的默认表达式 (Tom)
第一个真正可用的 FOREIGN KEY 约束触发器功能 (Jan)

添加 FOREIGN KEY ... MATCH FULL ... ON DELETE CASCADE (Jan)
添加 FOREIGN KEY ... MATCH <unspecified> 参照动作 (Don Baccus)
允许对 ctid (物理堆位置) 的 WHERE 限制 (Hiroshi)
把 pginterface 从 contrib 移到 interface 目录并改名为 pgeasy (Bruce)
更改 pgeasy connectdb() 参数顺序 (Bruce)
要求 SELECT DISTINCT 目标列表包含所有 ORDER BY 列 (Tom)
添加 Oracle 的 COMMENT ON 命令 (Mike Mascari <mascarim@yahoo.com>)
libpq 的 PQsetNoticeProcessor 函数现在返回先前的钩子 (Peter E)
防止 PQsetNoticeProcessor 被设为 NULL (Peter E)
使 COPY 中的 USING 可选 (Bruce)
允许目标列表中的子查询 (Tom)
允许比较操作符左侧的子查询 (Tom)
新并行回归测试 (Jan)
更改后端侧 COPY 使写文件的权限为 644 而不是 666 (Tom)
强制 PGDATA 目录权限安全，即使它已存在 (Tom)
添加 psql LASTOID 变量以返回最后插入的 oid (Peter E)
允许并发 vacuum 并移除 pg_vlock vacuum 锁文件 (Tom)
为 vacuum 添加权限检查 (Peter E)
新的允许异步连接的 libpq 函数：PQconnectStart()、PQconnectPoll()、PQresetStart()、PQresetPoll()、PQsetenvStart()、PQsetenvPoll()、PQsetenvAbort (Ewan Mellor)
新 libpq PQsetenv() 函数 (Ewan Mellor)
create/alter user 扩展 (Peter E)
$PGDATA 下新的 postmaster.pid 和 postmaster.opts (Tatsuo)
新 create/drop user/db 脚本 (Peter E)
psql 重大改造 (Peter E)
向 libpq 接口添加 const (Peter E)
新 libpq 函数 PQoidValue (Peter E)
显示导致 GROUP BY 问题的特定非聚合 (Tom)
使 pg_shadow 的更改重建 pg_pwd 文件 (Peter E)
添加 aggregate(DISTINCT ...) (Tom)
允许用标志控制 NULL 的 COPY 输入/输出 (Peter E)
使 postgres 用户默认有口令 (Peter E)
添加 CREATE/ALTER/DROP GROUP (Peter E)
所有管理脚本现在支持 --long 选项 (Peter E、Karel)
Vacuumdb 脚本现在支持 --all 选项 (Peter E)
ecpg 新的可移植 FETCH 语法
添加 ecpg 的 EXEC SQL IFDEF、EXEC SQL IFNDEF、EXEC SQL ELSE、EXEC SQL ELIF
和 EXEC SQL ENDIF 指令
添加 pg_ctl 脚本以控制后端启动 (Tatsuo)
添加 postmaster.opts.default 文件以存储启动标志 (Tatsuo)
允许 --with-mb=SQL_ASCII
把索引键的最大数量增加到 16 (Bruce)
把函数参数的最大数量增加到 16 (Bruce)
允许配置索引键和参数的最大数量 (Bruce)
允许非特权用户更改自己的口令 (Peter E)
启用口令认证；新用户必须使用 (Peter E)
不允许删除拥有数据库的用户 (Peter E)
把 initdb 选项 --with-mb 改为 --enable-multibyte
添加 initdb 提示超级用户口令的选项 (Peter E)
允许复杂类型转换如 col::numeric(9,2) 和 col::int2::float8 (Tom)
更新 initdb、initlocation、pg_dump、ipcclean 的用户界面 (Peter E)
新 pg_char_to_encoding() 和 pg_encoding_to_char() 函数 (Tatsuo)

libpq 非阻塞模式 (Alfred Perlstein)
改进未指定长度转换中类型的转换
新 plperl 内部编程语言 (Mark Hollomon)
允许 COPY IN 读取不以换行结尾的文件 (Tom)
指示长标识符何时被截断 (Tom)
允许聚合使用类型等价 (Peter E)
添加 Oracle 的 to_char()、to_date()、to_datetime()、to_timestamp()、to_number()
转换函数 (Karel Zak <zakkr@zf.jcu.cz>)
添加 SELECT DISTINCT ON (expr [, expr ...]) targetlist ... (Tom)
检查以确保 ORDER BY 与 DISTINCT 操作兼容 (Tom)
向 ODBC 添加 NUMERIC 和 int8 类型
改进 Append、Group、Agg、Unique 的 EXPLAIN 结果 (Tom)
添加 ALTER TABLE ... ADD FOREIGN KEY (Stephan Szabo)
允许 PL/pgSQL 中的 SELECT .. FOR UPDATE (Hiroshi)
即使到达 EOF 也启用反向顺序扫描 (Hiroshi)
添加布尔值及 >= 和 <= 的 btree 索引 (Don Baccus)
COPY FROM 失败时打印当前行号 (Massimo)
识别 POSIX 时区，如 "PST+8" 和 "GMT-8" (Thomas)
添加 DEC 作为 DECIMAL 的同义词 (Thomas)
添加 SESSION_USER 作为 SQL92 关键字，同 CURRENT_USER (Thomas)
实现 SQL92 列别名（也称相关名）(Thomas)
实现 SQL92 连接语法 (Thomas)
使 INTERVAL 保留字可用作列标识符 (Thomas)
实现 REINDEX 命令 (Hiroshi)
接受聚合函数 SUM(ALL col) 中的 ALL (Tom)
防止 GROUP BY 使用列别名 (Tom)
新 psql \encoding 选项 (Tatsuo)
允许 PQrequestCancel() 在等待锁状态下终止 (Hiroshi)
允许在所有情形下对负数取负
添加 ecpg 描述符 (Christof、Michael)
允许 CREATE VIEW v AS SELECT f1::char(8) FROM tbl
允许带长度的转换，如 foo::char(8)
新 libpq 函数 PQsetClientEncoding()、PQclientEncoding() (Tatsuo)
添加对 SJIS 用户定义字符的支持 (Tatsuo)
支持更大的视图/规则
使 libpq 的 PQconndefaults() 线程安全 (Tom)
禁用 // 作为注释以符合 ANSI，应使用 -- (Tom)
允许视图上的列别名 CREATE VIEW name (collist)
修复带子查询的视图的问题 (Tom)
允许 UPDATE table SET fld = (SELECT ...) (Tom)
SET 命令选项不再需要引号
把 pgaccess 更新到 0.98.6
新 SET SEED 命令
新 pg_options.sample 文件
新 SET FSYNC 命令 (Massimo)
允许创建表时 pg_descriptions
允许创建类型、列和函数时 pg_descriptions
允许 psql \copy 使用分隔符 (Peter E)
允许 psql 把 null 打印为与 "" 不同的 [null] (Peter E)

类型

大量数组修复 (Tom)

允许裸列名作为数组下标 (Tom)
改进 int 和 float 常量的类型转换 (Tom)
清理 int8 输入、范围检查和类型转换 (Tom)
修复 SELECT timespan('21:11:26'::time) (Tom)
netmask('x.x.x.x/0') 为 255.255.255.255 而不是 0.0.0.0 (Oleg Sharoiko)
在 NUMERIC 上添加 btree 索引 (Jan)
含 NUL 字符大对象的 Perl 修复 (Douglas Thomson)
大对象的 ODBC 修复 (free)
修复 cidr 数据类型的索引
修复以太网 MAC 地址 (macaddr 类型) 比较的问题
修复计算中发生溢出时日期/时间类型的问题 (Tom)
允许 int8 上的数组 (Peter E)
修复 NUMERIC 类型 (如 NUMERIC(4,4)) 的舍入/溢出问题 (Tom)
允许 NUMERIC 数组
修复 NUMERIC ceil() 和 floor() 函数中的缺陷 (Tom)
使 char_length()/octet_length 包含尾随空白 (Tom)
使 abstime/reltime 使用 int4 而不是 time_t (Peter E)
用于压缩文本字段的新 lztext 数据类型
修改处理 int 和 float 常量强制的代码 (Tom)
开始编写实现 BIT 和 BIT VARYING 类型的新代码 (Adriaan Joubert)
NUMERIC 现在接受科学记数法 (Tom)
NUMERIC 到 int4 舍入 (Tom)
正确地把 float4/8 转换为 NUMERIC (Tom)
允许与 NUMERIC 的类型转换 (Thomas)
使 ISO 日期风格 (2000-02-16 09:33) 成为默认 (Thomas)
添加 NATIONAL CHAR [VARYING] (Thomas)
允许 NUMERIC round 和 trunc 接受负标度 (Tom)
新 TIME WITH TIME ZONE 类型 (Thomas)
在 time 类型上添加 MAX()/MIN() (Thomas)
为 int8 添加 abs()、mod()、fac() (Thomas)
把函数改名为 float8 的 round()、sqrt()、cbrt()、pow() (Thomas)
为 float8 添加超越数学函数 (如 sin()、acos()) (Thomas)
为 NUMERIC 类型添加 exp() 和 ln()
把 NUMERIC power() 改名为 pow() (Thomas)
改进的 TRANSLATE() 函数 (Edwin Ramirez、Tom)
允许 X=-Y 操作符 (Tom)
允许 SELECT float8(COUNT(*))/(SELECT COUNT(*) FROM t) FROM t GROUP BY f1; (Tom)
允许 LOCALE 在正则表达式搜索中使用索引 (Tom)
允许创建函数索引时使用默认类型

性能

防止大量 AND 和 OR 的指数级空间消耗 (Tom)
收集系统列的属性选择率值 (Tom)
降低聚合的内存使用 (Tom)
修复 LIKE 优化以在多字节编码下使用索引 (Tom)
修复 r-tree 索引优化器选择率 (Thomas)
改进优化器选择率计算和函数 (Tom)
优化存在许多相等键情形的 btree 搜索 (Tom)
只在有索引时启用快速 LIKE 索引处理 (Tom)
重用带重复项索引页上的空闲空间 (Tom)
改进哈希连接处理 (Tom)
结果已排序时防止降序排序 (Hiroshi)

允许索引扫描查询限定的交换 (Tom)
在需要 ORDER BY/GROUP BY 的情形优先索引扫描 (Tom)
为性能把大内存请求按固定大小块分配 (Tom)
通过减少内存分配请求修复 vacuum 的性能 (Tom)
实现常量表达式化简 (Bernard Frankpitt、Tom)
用辅助列确定索引扫描的起点 (Hiroshi)
防止内部排序时四倍的磁盘空间使用 (Tom)
通过调用更少函数加快排序 (Tom)
创建与所有系统缓存匹配的系统索引 (Bruce、Hiroshi)
使系统缓存使用系统索引 (Bruce)
使所有系统索引唯一 (Bruce)
改进 pg_statistics 管理以提升 VACUUM 速度 (Tom)
降低后端缓存刷新频率 (Tom、Hiroshi)
COPY 现在重用先前的内存分配，改进性能 (Tom)
改进优化代价估计 (Tom)
改进范围查询 x > lowbound AND x < highbound 的优化器估计 (Tom)
在适当处使用 DNF 而不是 CNF (Tom、Taral)
进一步清理 OR-of-AND WHERE 子句 (Tom)
使用 OR 子句中的索引 (x = 1 AND y = 2) OR (x = 2 AND y = 4) (Tom)
随机索引页访问的更智能优化器计算 (Tom)
控制优化器代价的新 SET 变量 (Tom)
基于 LIMIT、OFFSET 和 EXISTS 限定优化查询 (Tom)
减少优化器连接路径的内部管理以提速 (Tom)
子查询大幅提速 (Tom)
未禁用 fsync 时减少 fsync 写 (Tom)
改进的 LIKE 优化器估计 (Tom)
防止 SELECT-only 查询中的 fsync (Vadim)
使索引创建使用 psort 代码，因为它现在更快 (Tom)
允许创建 > 1GB 的排序临时表

源代码树变更

修复 linux PPC 编译
新通用表达式树遍历子例程 (Tom)
把 form() 改为 varargform() 以防止可移植性问题
改进 Alpha 上大整数的范围检查
清理 /include 目录中的 #include (Bruce)
添加检查 include 的脚本 (Bruce)
从 *.c 文件移除不需要的 #include (Bruce)
更改 #include 以酌情使用 <> 和 "" (Bruce)
启用 libpq 的 Windows 编译
来自 Uncle George <gatgul@voicenet.com> 的 Alpha 自旋锁修复
优化器数据结构的彻底改造 (Tom)
修复 cygipc 库 (Yutaka Tanida)
允许 pgsql 在更新的 Cygwin 快照上工作 (Dan)
新目录版本号 (Tom)
添加 Linux ARM
把 heap_replace 改名为 heap_update
为 QNX 更新 (Dr. Andreas Kardos)
新平台特定的回归处理 (Tom)
把 oid8 改名为 oidvector、int28 改名为 int2vector (Bruce)
把所有 yacc 和 lex 文件包含进发行版 (Peter E.)
移除 lextest，不再需要 (Peter E)
修复 Windows 上的 libpq 和 psql (Magnus)

内部把 datetime 和 timespan 改为 timestamp 和 interval (Thomas)
修复 BSD/OS 上的 plpgsql
向回归测试添加 SQL_ASCII 测试用例 (Tatsuo)
configure --with-mb 现已废弃 (Tatsuo)
NT 修复
NetBSD 修复 (Johnny C. Lam <lamj@stat.cmu.edu>)
Alpha 编译修复
新多字节编码

E.122. 版本 6.5.3

发行日期
1999-10-13

这基本上是 6.5.2 的清理版本。我们添加了 6.5.2 中缺失的新 PgAccess，并安装了 NT 特定的修复。

E.122.1. 迁移到版本 6.5.3

对于运行 6.5.* 的用户，不需要进行转储/恢复。

E.122.2. 变更

更新了 pgaccess 0.98 版本
NT 特定补丁
修复继承表上规则的转储

E.123. 版本 6.5.2

发行日期
1999-09-15

这基本上是 6.5.1 的清理版本。我们修复了 6.5.1 用户报告的多种问题。

E.123.1. 迁移到版本 6.5.2

对于运行 6.5.* 的用户，不需要进行转储/恢复。

E.123.2. 变更

子查询+CASE 修复 (Tom)
为 solaris_i386 和 solaris_sparc 移植添加 SHLIB_LINK 设置 (Daren Sefcik)
WHERE 连接子句中 CASE 的修复 (Tom)
修复 BTScan 中止 (Tom)
修复对冗余 UNIQUE 和 PRIMARY KEY 索引的检查 (Thomas)
改进为可检查多列约束 (Thomas)

修复启用 MB 时 Windows 的制作问题 (Hiroki Kataoka)
允许 BSD yacc 和 bison 编译 pl 代码 (Bruce)
修复 SET NAMES 的功能
int8 修复 (Thomas)
修复 vacuum 的内存消耗 (Hiroshi、Tatsuo)
降低 vacuum 的总内存消耗 (Tom)
修复 timestamp(datetime)
规则反编译缺陷修复 (Tom)
修复 mkMakefile.tcldefs.sh.in 和 mkMakefile.tkdefs.sh.in 中的引号问题 (Tom)
这是为重用 vacuum 释放的索引页空间 (Vadim)
为 pg_dump 记录 -x (Bruce)
修复规则反编译器中的一元操作符 (Tom)
在 mdtruncate() 期间注释掉多余段的 FileUnlink (Tom)
来自 Yu Cao 的 IRIX 链接修复
修复 LIKE 中的逻辑错误：到达模式末尾而文本未结束时不应返回 LIKE_ABORT (Tom)
修复事务中止期间堆内存分配的不当清理 (Tom)
更新 pgaccess 0.98 版本

E.124. 版本 6.5.1

发行日期
1999-07-15

这基本上是 6.5 的清理版本。我们修复了 6.5 用户报告的多种问题。

E.124.1. 迁移到版本 6.5.1

对于运行 6.5 的用户，不需要进行转储/恢复。

E.124.2. 变更

添加 NT README 文件
linux_ppc、IRIX、linux_alpha、OpenBSD、alpha 的可移植性修复
移除 QUERY_LIMIT，使用 SELECT...LIMIT
修复继承上的 EXPLAIN (Tom)
补丁允许对多段表进行清理 (Hiroshi)
R-Tree 优化器选择率修复 (Tom)
ACL 文件描述符泄漏修复 (Atsushi Ogawa)
新表达式子树代码 (Tom)
只读事务避免磁盘写 (Vadim)
修复最后事务中止时临时表的移除 (Bruce)
防止创建过大的行 (Bruce)
plpgsql 修复
允许 32k - 64k 的端口号 (Bruce)
添加 ^ 优先级 (Bruce)
把名为 pg_temp 的排序文件改名为 pg_sorttemp (Bruce)
修复时间值中的微秒 (Tom)
教程源码清理
新 linux_m68k 移植

修复某些情形下 NULL 的排序（Tom）
修复共享库依赖（Tom）
修复影响子查询中 GROUP BY 的故障（Tom）
修复一些编译器警告（Tomoaki Nishiyama）
新增 Win1250（捷克语）支持（Pavel Behal）

E.125. 版本 6.5

发行日期
1999-06-09

此版本标志着开发团队对从 Berkeley 继承的源代码的掌握迈出了一大步。你会看到我们现在能轻松添加重大特性，这要归功于我们全球开发团队不断扩大的规模和日益丰富的经验。

以下是较显著变更的简要总结：

多版本并发控制（MVCC）

这移除了我们旧的表级锁定，代之以优于大多数商业数据库系统的锁定系统。
在传统系统中，每个被修改的行都被锁定直到提交，阻止其他用户读取。MVCC 利用 PostgreSQL 天生的多版本特性，允许读者在写入者活动期间继续读取一致的数据。写入者继续使用紧凑的 pg_log 事务系统。这一切都不需要像传统数据库系统那样为每行分配锁。因此，基本上我们不再受简单表级锁定的限制；我们有了比行级锁定更好的东西。

Hot backups from pg_dump

pg_dump 利用新的 MVCC 特性，在数据库保持在线并可用于查询的同时给出一致的数据库转储/备份。

Numeric data type

我们现在有了真正的数字数据类型，精度由用户指定。

Temporary tables

临时表保证在数据库会话内名称唯一，并在会话退出时销毁。

New SQL features

我们现在支持 CASE、INTERSECT 和 EXCEPT 语句。新增了 LIMIT/OFFSET、SET TRANSACTION ISOLATION LEVEL、SELECT ... FOR UPDATE 以及改进的 LOCK TABLE 命令。

Speedups

得益于团队内的多种才能，我们继续加快 PostgreSQL 的速度。我们加速了内存分配、优化、表连接和行传输例程。

Ports

我们继续扩大移植列表，这次包括 Windows NT/ix86 和 NetBSD/arm32。

Interfaces

大多数接口有了新版本，现有功能得到改进。

Documentation

文档中到处都有新的和更新的材料。SGI 和 AIX 平台有了新的 FAQ。教程有 Stefan Simkovics 编写的 SQL 入门信息。用户指南有涵盖 postmaster 和更多工具程序的参考页，新附录包含日期/时间行为的细节。管理员指南有 Tom Lane 编写的故障排除新章节。程序员指南有 Stefan 编写的查询处理描述，以及通过匿名 CVS 和 CVSup 获取 PostgreSQL 源代码树的细节。

E.125.1. 迁移到版本 6.5

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 进行转储/恢复。pg_upgrade 不能用于升级到此版本，因为表的磁盘结构相对于以前的版本已更改。

新的多版本并发控制（MVCC）特性在多用户环境中可能给出略为不同的行为。
请阅读并理解以下章节，以确保你的现有应用程序能给出你需要的行为。

E.125.1.1. 多版本并发控制

由于 6.5 中的读者不锁定数据（无论事务隔离级别如何），一个事务读取的数据可能被另一个事务覆写。换句话说，如果 `SELECT` 返回了一行，并不意味着该行在被返回时（即语句或事务开始之后的某个时刻）确实存在，也不意味着该行在当前事务提交或回滚之前受到保护，
不被并发事务删除或更新。

为确保某行确实存在并保护它不被并发更新，必须使用 `SELECT FOR UPDATE` 或适当的 `LOCK TABLE` 语句。从以前的 PostgreSQL 版本和其他环境移植应用程序时应考虑到这一点。

如果你在使用 `contrib/refint.*` 触发器做参照完整性，请记住上述内容。现在需要额外的技术。一种方法是：如果事务要更新/删除主键，使用 `LOCK parent_table IN SHARE ROW EXCLUSIVE MODE` 命令；如果事务要更新/插入外键，使用 `LOCK parent_table IN SHARE MODE` 命令。

> **注意**
>
> 注意，如果你以 SERIALIZABLE 模式运行事务，则必须在执行事务中的任何 DML 语句（`SELECT/INSERT/DELETE/UPDATE/FETCH/COPY_TO`）之前执行上述 `LOCK` 命令。

当实现读取脏（未提交）数据的能力（无论隔离级别）和真正的参照完整性后，
这些不便将消失。

E.125.2. 变更

```
缺陷修复
---------
修复 text<->float8 和 text<->float4 转换函数 (Thomas)
修复创建带大小写混合约束表的问题 (Billy)
更改 exp()/pow() 行为以在下溢/上溢时报错 (Jan)
修复 pg_dump -z 中的缺陷
内存越界清理 (Tatsuo)
修复 lo_import 崩溃 (Tatsuo)
调整数据类型名的处理以抑制双引号 (Thomas)
```

用类型强制匹配列和 DEFAULT (Thomas)
修复死锁，使其在一秒睡眠后只检查一次 (Bruce)
聚合和 PL/pgsql 的修复 (Hiroshi)
修复子查询崩溃 (Vadim)
修复 libpq 函数 PQfnumber 和大小写不敏感名称的问题 (Bahman Rafatjoo)
修复大对象中间写入、无额外块和内存消耗的问题 (Tatsuo)
修复 pg_dump -d 或 -D 以及 INSERT 中特殊字符的引号问题
修复 dynahash 的严重问题 (Tom)
修复 INET/CIDR 可移植性问题
修复 ALTER TABLE ADD COLUMN 中选择率错误的问题 (Bruce)
修复执行器使不同列类型的归并连接可用 (Tom)
修复 Alpha OR 选择率缺陷
修复 OR 索引选择率问题 (Bruce)
修复 \d 为 char()/varchar() 显示正确长度的问题 (Ryan)
修复教程代码 (Clark)
改进 destroyuser 检查 (Oliver)
修复 Kerberos 的问题 (Rodney McDuff)
修复存在脏缓冲区时删除数据库的问题 (Bruce)
修复使 sequence nextval() 可区分大小写的问题 (Bruce)
修复 !!= 操作符
销毁数据库文件前丢弃缓冲区 (Bruce)
修复执行器两次求值函数的情形 (Tatsuo)
允许 sequence nextval 动作区分大小写 (Bruce)
修复优化器对负数索引不工作的问题 (Bruce)
修复执行器中 fjIsNull 的内存泄漏
修复聚合内存泄漏 (Erik Riedel)
允许含连字符的用户名授予权限
清理 inet 类型中的 NULL
清理系统表缺陷 (Tom)
修复 PAGER 和 \? 命令的问题 (Masaaki Sakaida)
把默认多段文件大小限制降为 1GB (Peter)
修复 CREATE OPERATOR 的转储 (Tom)
修复游标反向扫描 (Hiroshi Inoue)
修复使用 \i 时的 COPY FROM STDIN (Tom)
修复子查询在表达式内比较的问题 (Jan)
修复返回行时的错误报告处理 (Tom)
修复数组类型引用的问题 (Tom、Jan)
防止 UPDATE SET oid (Jan)
修复 pg_dump 使 -t 选项可处理大小写敏感的表名
特殊情形 GROUP BY 的修复 (Tom、Jan)
修复失败查询的内存泄漏 (Tom)
DEFAULT 现在支持大小写混合的标识符 (Tom)
修复表和索引 DROP/RENAME 的多段使用 (Ole Gjerde)
禁用 pg_dump 同时使用 -o 和 -d 选项 (Bruce)
允许 pg_dump 正确转储组权限 (Bruce)
修复 INSERT INTO table SELECT * FROM table2 中的 GROUP BY (Jan)
修复视图中的计算 (Jan)
修复数组索引上的聚合 (Tom)
修复 DEFAULT 处理值中需要过多引号的单引号的问题
修复非超级用户导入/导出大对象的安全问题 (Tom)
创建表的事务回滚现在正确清理 (Tom)
修复以允许长表名和列名生成正确的序列名 (Tom)

增强

```
------------
添加 "vacuumdb" 工具
通过更好的内存分配加速 libpq (Tom)
EXPLAIN 显示所有使用的索引 (Tom)
实现 CASE、COALESCE、NULLIF 表达式 (Thomas)
新 pg_dump 表输出格式 (Constantin)
添加字符串 min()/max() 函数 (Thomas)
把新类型强制技术扩展到聚合 (Thomas)
新 moddatetime contrib (Terry)
更新到 pgaccess 0.96 (Constantin)
为单字节 "char" 类型添加例程 (Thomas)
改进的 substr() 函数 (Thomas)
改进的多字节处理 (Tatsuo)
多版本并发控制/MVCC (Vadim)
新 Serialized 模式 (Vadim)
修复超过 2GB 表的问题 (Peter)
新 SET TRANSACTION ISOLATION LEVEL (Vadim)
新 LOCK TABLE IN ... MODE (Vadim)
更新 ODBC 驱动 (Byron)
新 NUMERIC 数据类型 (Jan)
新 SELECT FOR UPDATE (Vadim)
处理输入值的 "NaN" 和 "Infinity" (Jan)
改进的日期/年份处理 (Thomas)
改进的后端连接处理 (Magnus)
日志文件的新选项 ELOG_TIMESTAMPS 和 USE_SYSLOG (Massimo)
新 TCL_ARRAYS 选项 (Massimo)
新 INTERSECT 和 EXCEPT (Stefan)
用于主键跟踪的新 pg_index.indisprimary (D'Arcy)
允许在创建前删除表的新 pg_dump 选项 (Brook)
行输出例程加速 (Tom)
新 READ COMMITTED 隔离级别 (Vadim)
新 TEMP 表/索引 (Bruce)
结果已排序时防止排序 (Jan)
新内存分配优化 (Jan)
允许 psql 执行 \p\g (Bruce)
允许多个规则动作 (Jan)
添加 LIMIT/OFFSET 功能 (Jan)
改进连接大量表时的优化器 (Bruce)
来自 S. Simkovics 硕士论文的新 SQL 介绍 (Stefan、Thomas)
来自 S. Simkovics 硕士论文的后端处理新介绍 (Stefan)
改进的 int8 支持 (Ryan Bradetich、Thomas、Tom)
int8 与 text/varchar 类型之间转换的新例程 (Thomas)
新浓密计划，其中连接元表 (Bruce)
默认启用右手查询 (Bruce)
允许在 configure 时设置可靠的最大后端数
      (--with-maxbackends 和 postmaster 开关 (-N backends) ) (Tom)
因优化器提速，GEQO 默认现为 10 表 (Tom)
为 MS-SQL 可移植性允许 NULL=Var (Michael、Bruce)
修改 contrib check_primary_key() 使其可为 "automatic" 或
 "dependent" (Anand)
允许 psql 对视图的 \d 显示查询 (Ryan)
LIKE 提速 (Bruce)
Ecpg 修复/特性，参见 src/interfaces/ecpg/ChangeLog 文件 (Michael)
JDBC 修复/特性，参见 src/interfaces/jdbc/CHANGELOG (Peter)
```

使 % 操作符具有与 / 相同的优先级 (Bruce)
添加新 postgres -O 选项以允许系统表结构更改 (Bruce)
更新 contrib/pginterface/findoidjoins 脚本 (Tom)
带索引已删行 vacuum 的大幅提速 (Vadim)
允许非 SQL 函数根据参数运行不同版本 (Tom)
添加 -E 选项显示 \dt 等发送的实际查询 (Masaaki Sakaida)
在 psql 启动横幅中添加版本号 (Masaaki Sakaida)
新 contrib/vacuumlo 移除未被引用的大对象 (Peter)
新表大小初始化使未清理的表性能更好 (Tom)
改进连接被拒绝时的错误消息 (Tom)
支持 char() 和 varchar() 字段数组 (Massimo)
哈希代码的彻底改造以提高可靠性和性能 (Tom)
更新到 PyGreSQL 2.4 (D'Arcy)
更改调试选项使 -d4 和 -d5 产生不同的节点显示 (Jan)
新 pg_options: pretty_plan、pretty_parse、pretty_rewritten (Jan)
更好的系统表访问优化统计 (Tom)
更好地处理非默认块大小 (Massimo)
改进 GEQO 优化器内存消耗 (Tom)
UNION 现在支持对不在目标列表中的列 ORDER BY (Jan)
libpq++ 的重大改进 (Vince Vielhaber)
pg_dump 现在默认使用 -z (ACL) (Bruce)
后端缓存和内存提速 (Tom)
让 pg_dump 在一个快照事务中完成所有工作 (Vadim)
修复大对象内存泄漏和 pg_dump 问题 (Tom)
INET 类型现在比较时遵循网络掩码
使 VACUUM ANALYZE 只使用读锁 (Vadim)
允许 UNION 上的视图 (Jan)
pg_dump 现在能在活动数据库上生成一致快照 (Vadim)

源代码树变更

改进移植匹配 (Tom)
SunOS 可移植性修复
添加 Windows NT 后端移植并启用动态装载 (Magnus 和 Daniel Horak)
新移植到运行 Linux 的 Cobalt Qube (Mips) (Tatsuo)
移植到 NetBSD/m68k (Mr. Mutsuki Nakajima)
移植到 NetBSD/sun3 (Mr. Mutsuki Nakajima)
移植到 NetBSD/macppc (Toshimi Aoki)
修复 tcl/tk 配置 (Vince)
移除规则查询的 CURRENT 关键字 (Jan)
NT 动态装载现在可用 (Daniel Horak)
添加 ARM32 支持 (Andrew McMurry)
更好地支持 HP-UX 11 和 UnixWare
改进文件处理使其更统一，防止文件描述符泄漏 (Tom)
plpgsql 的新安装命令 (Jan)

E.126. 版本 6.4.2

发行日期
1998-12-20

6.4.1 版本打包不当。此版本还包含一个额外的缺陷修复。

E.126.1. 迁移到版本 6.4.2

对于运行 6.4.* 的用户，不需要进行转储/恢复。

E.126.2. 变更

修复某些平台上日期时间常量的问题（Thomas）

E.127. 版本 6.4.1

发行日期
1998-12-18

这基本上是 6.4 的清理版本。我们修复了 6.4 用户报告的多种问题。

E.127.1. 迁移到版本 6.4.1

对于运行 6.4 的用户，不需要进行转储/恢复。

E.127.2. 变更

新增 pg_dump -N 标志以强制标识符周围加双引号。这是默认（Thomas）
修复 where 子句中 NOT 导致崩溃的问题（Bruce）
EXPLAIN VERBOSE 核心转储修复（Vadim）
修复 Linux 上的共享库问题
修复表存在性测试以允许表名中的大小写混合和空白（Thomas）
修复几个 pg_dump 缺陷
Configure 更好地匹配 template/.similar 条目（Tom）
把内建函数名从 SPI_* 改为 spi_*
OR WHERE 子句修复（Vadim）
大小写混合表名的修复（Billy）
contrib/linux/postgres.init.csh/sh 修复（Thomas）
libpq 内存越界修复
SunOS 修复（Tom）
更改 exp() 行为在下溢时报错（Thomas）
pg_dump 对内存泄漏、继承约束、布局变更的修复
把 pgaccess 更新到 0.93
64 位平台的原型修复
多字节修复（Tatsuo）
新 ecpg 手册页
内存越界修复（Tatsuo）
lo_import() 崩溃修复（Bruce）
更好地搜索安装程序（Tom）
时区修复（Tom）
HP-UX 修复（Tom）
用隐式类型强制匹配 DEFAULT 值（Thomas）

添加帮助处理单字节（内部）字符类型的例程（Thomas）
Windows 下 libpq 编译修复（Magnus）
升级到 PyGreSQL 2.2（D'Arcy）

E.128. 版本 6.4

发行日期
1998-10-30

此版本有许多新特性和改进。感谢我们的开发者和维护者，
自上一个版本以来系统的几乎每个方面都得到了关注。以下是简要而不完整的总结：

- 得益于 Jan Wieck 在重写规则系统中的大量新代码，视图和规则现在可以工作了。他还为程序员指南写了相关章节。
- Jan 还贡献了第二个过程语言 PL/pgSQL，与他在上个版本贡献的最初的 PL/pgTCL 过程语言相伴。
- 我们有 Tatsuo Ishii 提供的可选多字节字符集支持，以补充现有的区域设置支持。
- 客户端/服务器通信得到了清理，感谢 Tom Lane，对异步消息和中断的支持更好了。
- 解析器现在执行自动类型强制，以把参数匹配到可用的操作符和函数，
  并把列和表达式与目标列匹配。这使用了支持 PostgreSQL 类型可扩展特性的通用机制。用户指南有涵盖此主题的新章节。
- 新增了三种数据类型。其中两种，inet 和 cidr，支持各种形式的 IP 网络、子网和机器寻址。某些平台上现在有 8 字节整数类型可用。详情见用户指南中的数据类型章节。第四种类型 serial 现在被解析器支持，作为 int4 类型、序列和唯一索引的混合体。
- 新增了更多 SQL92 兼容的语法特性，包括 INSERT DEFAULT VALUES
- 自动配置和安装系统得到了一些关注，应比以往在更多平台上更健壮。

E.128.1. 迁移到版本 6.4

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 或 pg_dumpall 进行转储/恢复。

E.128.2. 变更

缺陷修复

修复 PQsetdb/PQfinish 中的小内存泄漏（Bryan）
移除 char2-16 数据类型，使用 char/varchar（Darren）
Pqfn 现在处理 NOTICE 消息（Anders）
降低多后端时自旋锁的忙等开销（dg）
卡死自旋锁检测（dg）
修复 "ISO 风格" timespan 的解码和编码（Thomas）

修复事务回滚后删除表的问题 (Vadim)
更改错误消息并移除无效的更新消息 (Vadim)
修复 COPY 数组检查的问题
修复 SELECT 1 UNION SELECT NULL
修复大对象调用中的缓冲区泄漏 (Pascal)
把 owner 从 oid 类型改为 int4 (Bruce)
修复 Oracle 兼容函数 btrim()、ltrim() 和 rtrim() 中的缺陷
修复共享失效缓存溢出 (Massimo)
防止失败的 COPY 中的文件描述符泄漏 (Bruce)
修复 libpgtcl 的 pg_select 中的内存泄漏 (Constantin)
修复超过 8 字符的用户名/口令问题 (Tom)
修复后端处理异步 NOTIFY 的问题 (Tom)
修复许多错误的系统表条目 (Tom)

增强

升级 ecpg 和 ecpglib，参见 src/interfaces/ecpc/ChangeLog (Michael)
在 EXPLAIN 中显示使用的索引 (Zeugswetter)
EXPLAIN 调用规则系统并显示被重写查询的计划 (Jan)
通过 configure 使许多数据类型和函数感知多字节 (Tatsuo)
新 configure --with-mb 选项 (Tatsuo)
新 initdb --pgencoding 选项 (Tatsuo)
新 createdb -E 多字节选项 (Tatsuo)
Select version(); 现在返回 PostgreSQL 版本 (Jeroen)
libpq 现在允许异步客户端 (Tom)
允许从客户端取消后端查询 (Tom)
psql 现在可用 Control-C 取消查询 (Tom)
libpq 用户不必发送哑查询即可获取 NOTIFY 消息 (Tom)
NOTIFY 现在发送发送者的 PID，可以判断是否是自己的 (Tom)
PGresult 结构现在包含相关的错误消息（如果有） (Tom)
定义 date_part() 的 "tz_hour" 和 "tz_minute" 参数 (Thomas)
添加 varchar 和 bpchar 之间转换的例程 (Thomas)
添加允许把 varchar 和 bpchar 调整到目标列大小的例程 (Thomas)
添加位标志以支持数据检索中的时区小时和分钟 (Thomas)
允许有效浮点数的更多变体（如 ".1"、"1e6"） (Thomas)
修复带前导空格的一元负号解析 (Thomas)
按 SQL92 规范实现 TIMEZONE_HOUR、TIMEZONE_MINUTE (Thomas)
检查并正确忽略 FOREIGN KEY 列约束 (Thomas)
按 SQL92 规范定义 USER 为 CURRENT_USER 的同义词 (Thomas)
启用 HAVING 子句但其他地方尚无修复。
使 "char" 类型成为 "char(1)" 的同义词（实际实现为 bpchar） (Thomas)
为 DEFAULT 子句处理保存指定的字符串类型 (Thomas)
强制涉及不同数据类型的操作 (Thomas)
允许对不同类型列的某些索引使用 (Thomas)
添加自动类型转换能力 (Thomas)
清理大对象，使文件在打开时被截断 (Peter)
Readline 清理 (Tom)
允许 psql \f \ 用空格作为分隔符 (Bruce)
把 pg_attribute.atttypmod 传递给前端以获取列字段长度 (Tom、Bruce)
/contrib 中的 Msql 兼容库 (Aldrin)
移除 ORDER/GROUP BY 子句标识符必须
包含在目标列表中的要求 (David)
转换列以匹配 UNION 子句中的列 (Thomas)
移除 fork()/exec() 只做 fork() (Bruce)

Jdbc 清理 (Peter)
在 ps 命令行上显示后端状态 (只在某些平台上有效) (Bruce)
Pg_hba.conf 的 database 字段现在有 sameuser 选项
使 lo_unlink 接受 oid 参数而不是 int4
为无法处理我们宏的编译器新增 DISABLE_COMPLEX_MACRO (Bruce)
Libpgtcl 现在把 NOTIFY 作为 Tcl 事件处理，不必发送哑查询 (Tom)
libpgtcl 清理 (Tom)
向 libpgtcl 的 pg_result 命令添加 -error 选项 (Tom)
新区域设置补丁，参见 docs/README/locale (Oleg)
修复 pg_dump 使 CONSTRAINT 和 CHECK 语法正确 (ccb)
新 contrib/lo 代码用于移除孤立大对象 (Peter)
多字节特性的新 psql 命令 "SET CLIENT_ENCODING TO 'encoding'",
参见 /doc/README.mb (Tatsuo)
contrib/noupdate 代码用于撤销列上的更新权限
libpq 现在可以在 Windows 上编译 (Magnus)
在 libpq 中添加 PQsetdbLogin()
新 8 字节整数类型，由 configure 检查操作系统支持 (Thomas)
更好地支持带引号的表/列名 (Thomas)
在 pg_dump 中用双引号包围表和列名 (Thomas)
PQreset() 现在可用于口令 (Tom)
处理 GROUP BY 目标列表列号越界的情形 (David)
允许子查询中的 UNION
向 \d? 命令添加屏幕自动调整大小 (Bruce)
用 UNION 在一个查询中显示所有 \d? 结果 (Bruce)
添加 \d? 字段搜索特性 (Bruce)
Pg_dump 发出更少的 \connect 请求 (Tom)
使 pg_dump -z 标志更好地工作并在手册页中记录 (Tom)
添加对子查询和 union 完全支持的 HAVING 子句 (Stephan)
contrib/fulltextindex 中的全文索引例程 (Maarten)
事务 id 现在存储在共享内存中 (Vadim)
发出 COPY 命令时的新 PGCLIENTENCODING (Tatsuo)
支持 SQL92 语法 "SET NAMES" (Tatsuo)
支持 LATIN2-5 (Tatsuo)
添加 UNICODE 回归测试用例 (Tatsuo)
锁管理器清理，LLL 的新锁定模式 (Vadim)
允许 OR 子句使用索引 (Bruce)
允许 "SELECT NULL ORDER BY 1;"
Explain VERBOSE 打印计划，现在还把计划美化打印到
postmaster 日志文件 (Bruce)
向 \d 命令添加索引显示 (Bruce)
允许对函数 GROUP BY (David)
大对象的新 pg_class.relkind (Bruce)
把 libpq NOTICE 消息发送到其他位置的新方法 (Tom)
psql 的新 \w 写入命令 (Bruce)
新 /contrib/findoidjoins 扫描 oid 列以发现连接关系 (Bruce)
检查含常量的限制子句的有效索引时
允许考虑二进制兼容的索引 (Thomas)
/contrib/isbn_issn 中的新 ISBN/ISSN 代码
允许 NOT LIKE、IN、NOT IN、BETWEEN 和 NOT BETWEEN 约束 (Thomas)
新重写系统修复了规则和视图的许多问题 (Jan)

* 关系上的规则可用
* insert/update/delete 上的事件限定可用
* 新的 OLD 变量引用 CURRENT，CURRENT 未来将被移除
* 更新规则可在规则限定/动作中引用 NEW 和 OLD

```
        * 视图上的 insert/update/delete 规则可用
        * 现在支持多个规则动作，用括号包围
        * 普通用户可以在有 RULE 权限的表上创建视图/规则
        * 规则和视图继承创建者的权限
        * 没有列级别的规则
        * 没有 UPDATE NEW/OLD 规则
        * 新 pg_tables、pg_indexes、pg_rules 和 pg_views 系统视图
        * SELECT 规则只有单个动作
        * 彻底的重写改造，也许在 6.5
        * 处理子查询
        * 处理视图上的聚合
        * 处理 insert into select from view 可用
系统索引现在是多键的 (Bruce)
移除 Oidint2、oidint4 和 oidname 类型 (Bruce)
更多系统表查找使用系统缓存 (Bruce)
backend/pl 中的新后端编程语言 PL/pgSQL (Jan)
新 SERIAL 数据类型，自动创建序列/索引 (Thomas)
无需重新编译即可启用断言检查 (Massimo)
用户锁增强 (Massimo)
设置序列值的新 setval() 命令 (Massimo)
启动时若无 postmaster 运行则自动移除 unix 套接字文件 (Massimo)
条件跟踪包 (Massimo)
新 UNLISTEN 命令 (Massimo)
psql 和 libpq 现在可用 win32.mak 在 Windows 下编译 (Magnus)
Lo_read 不再存储尾随 NULL (Bruce)
标识符现在内部截断为 31 字符 (Bruce)
Createuser 选项现在可在命令行上使用
添加 64 位整数支持代码，configure 已测试，int8 类型 (Thomas)
防止失败 COPY 的文件描述符泄漏 (Bruce)
新 pg_upgrade 命令 (Bruce)
更新的 /contrib 目录 (Massimo)
新 CREATE TABLE DEFAULT VALUES 语句可用 (Thomas)
新 INSERT INTO TABLE DEFAULT VALUES 语句可用 (Thomas)
新 DECLARE 和 FETCH 特性 (Thomas)
libpq 的内部结构现在不导出 (Tom)
允许多达 8 键的索引 (Bruce)
移除 ARCHIVE 关键字，它不再使用 (Thomas)
pg_dump -n 标志以抑制标识符周围的引号
禁用视图的系统列 (Jan)
用于网络地址的新 INET 和 CIDR 类型 (TomH、Paul)
psql 输出中不再有双引号
pg_dump 现在转储视图 (Terry)
新 SET QUERY_LIMIT (Tatsuo、Jan)

源代码树变更
--------------------
/contrib 清理 (Jun)
内联一些每行都调用的小函数 (Bruce)
Alpha/linux 修复
HP-UX 清理 (Tom)
多字节回归测试 (Soonmyung.)
从 configure 移除 --disabled 选项
定义 PGDOC 默认使用 POSTGRESDIR
使回归可选
```

移除 pgindent 的额外花括号代码 (Bruce)
添加 bsdi 共享库支持 (Bruce)
新 --without-CXX 支持 configure 选项 (Brook)
新 FAQ_CVS
更新 tools/backend 中的后端流程图 (Bruce)
把 atttypmod 从 int16 改为 int32 (Bruce、Tom)
为没有 Getrusage() 的平台提供修复 (Tom)
把 PQconnectdb、PGUSER、PGPASSWORD 添加到 libpq 手册页
NS32K 平台修复 (Phil Nelson、John Buller)
SCO 7/UnixWare 2.x 修复 (Billy 等)
Sparc/Solaris 2.5 修复 (Ryan)
Pgbuiltin.3 已过时，移到 doc 文件 (Thomas)
更多的文档 (Thomas)
Nextstep 支持 (Jacek)
Aix 支持 (David)
pginterface 手册页 (Bruce)
所有共享库都有版本号
把所有操作系统特定的共享库定义合并到一个文件
更智能的 TCL/TK 配置检查 (Billy)
更智能的 perl 配置 (Brook)
未找到 install 脚本时 configure 使用随附的 install-sh (Tom)
用于共享库配置的新 Makefile.shlib (Tom)

E.129. 版本 6.3.2

发行日期
1998-04-07

这是 6.3.x 的缺陷修复版本。有关 6.3 版本系列新特性更完整的摘要，请参阅 6.3 版本的发行说明。

摘要：

- 修复某些平台（包括 Linux）的自动配置支持。
- 正确处理 BETWEEN 和 LIKE 子句左侧的函数调用。

对于运行 6.3 或 6.3.1 的用户，不需要进行转储/恢复。只需做一次 make distclean、make 和 make install。最后一步应在 postmaster 未运行时执行。你应重新链接任何使用 PostgreSQL 库的自定义应用程序。

对于从 6.3 之前安装的升级，请参阅 6.3 版本的安装和迁移说明。

E.129.1. 变更

对 tcl/tk 的配置检测改进 (Brook Milligan、Alvin)
手册页改进 (Bruce)
BETWEEN 和 LIKE 修复 (Thomas)
修复 pg_dump 所用 psql \connect 的问题 (Oliver Elphick)
新 odbc 驱动

pgaccess 0.86 版
移除 qsort，现使用 libc 版本并清理 (Jeroen)
修复检测到的缓冲区越界 (Maurice Gittens)
修复 libpgtcl 中的缓冲区越界 (Randy Kunkee)
修复带 DISTINCT 或 ORDER BY 的 UNION (Bruce)
gettimeofday configure 检查 (Doug Winterburn)
修复 "indexes not used" 缺陷 (Vadim)
文档补充 (Thomas)
修复后端内存泄漏 (Bruce)
libreadline 清理 (Erwan MAS)
移除 DISTDIR (Bruce)
Makefile 依赖清理 (Jeroen van Vianen)
ASSERT 修复 (Bruce)

E.130. 版本 6.3.1

发行日期
1998-03-23

摘要：

- 对多字节字符集的额外支持。
- 修复混合端序客户端和服务器的字节顺序。
- 对允许的 SQL 语法做小的更新。
- 改进安装时配置的自动检测。

对于运行 6.3 的用户，不需要进行转储/恢复。只需做一次 make distclean、make 和 make install。最后一步应在 postmaster 未运行时执行。你应重新链接任何使用 PostgreSQL 库的自定义应用程序。

对于从 6.3 之前安装的升级，请参阅 6.3 版本的安装和迁移说明。

E.130.1. 变更

ecpg 清理/修复，现为 1.1 版 (Michael Meskes)
pg_user 清理 (Bruce)
pg_dump 和 tclsh 的大对象修复 (alvin)
修复多个相邻下划线的 LIKE
修复内建函数的重定义 (Thomas)
ultrix4 清理
升级到 pg_access 0.83
更新 CLUSTER 手册页
多字节字符集支持，参见 doc/README.mb (Tatsuo)
configure --with-pgport 修复
pg_ident 修复
大对象位图修复 (Massimo)
修复所有 postgres 用户的 Postgres 大小写问题

修复 CREATE TRIGGER 中的对齐错误
pg_dump 的国家语言支持 (Fabien.Mabille)
关于把 max vector 消息长度乘以 3 的查询
为 libpq 添加自定义标题
修复 libpq 中断时的内存泄漏 (D. Sears)
修复 pg_proc 定义中的问题
pg_dump 修复 (Bruce)
修复 isnull 缺陷 (José Soares)
透明的多字节转换
上下文和位置错误消息 (Ian Burrell)
修复小内存泄漏
拆除 libpq 中的所有代码重复
乱序对象修复
新索引估计器
启用单个动作规则重写 (Jan)
外连接语法现在生成查询计划 (Thomas)
>= 与 <= 与 = 的 btree 索引
服务器对客户端编码转换指令 (Tatsuo)
对 EUC_TW 转换的 BIG5 修复 (Tatsuo)

E.131. 版本 6.3

发行日期

1998-03-01

此版本有许多新特性和改进。以下是简要而不完整的总结：

- 许多新的 SQL 特性，包括完整的 SQL92 子查询能力（只缺目标列表子查询，其余一切都有了）。
- 支持用客户端环境变量指定时区和日期样式。
- 客户端/服务器连接的套接字接口。现在这是默认，所以你可能需要用 -i 标志启动 postmaster。
- 更好的口令授权机制。默认表权限已更改。
- 旧式时间旅行已被移除。性能得到改进。

注意

Bruce Momjian 写了以下说明来介绍新版本。

我想提一些 6.3 的普遍问题。这里只是一句话说不完的大项。关于全部特性列表，请参阅发行说明末尾的详细变更列表。

首先，我们现在有了子查询。既然有了它们，我想指出：没有子查询的 SQL 是一种非常受限的语言。子查询是绝大多数严肃的商业应用所必需的，我们终于在特性列表中勾掉了它。

我们对此特别自豪，因为这一代码里程碑意味着 PostgreSQL 的查询处理器现在完全准备就绪。注意，外部连接现在可以使用子查询语法实现。我们准备了一个脚本在 contrib 目录中来做这件事，它会把 UNION 连接重写为子查询。

其次，6.3 默认使用 Unix 域套接字而不是 TCP/IP。要启用来自其他机器的连接，必须使用新的 postmaster -i 选项，当然还要编辑 `pg_hba.conf`。此外，由于这个原因，`pg_hba.conf` 的格式已经更改。

第三，`char()` 字段现在允许比 `varchar()` 或 `text` 更快的访问。具体来说，`text` 和 `varchar()` 在访问该类型第一列之后的任何列时有性能损失。`char()` 过去也有这种访问损失，但现在没有了。这可能提示你重新设计一些表，特别是如果你有定义为 `varchar()` 或 `text` 的短字符列。这项和其他变更使 6.3 比以前的版本更快。

我们现在可以独立于任何 Unix 文件定义口令。有新的 SQL USER 命令。更多信息见管理员指南。还有一个新表 pg_shadow，用于存储用户信息和用户口令，默认只有 postgres 超级用户可以 SELECT。pg_user 现在是 pg_shadow 的视图，可被 PUBLIC SELECT。你的应用程序应继续使用 pg_user 而无需更改。

用户创建的表现在默认不再对 PUBLIC 授予 SELECT 权限。这样做是因为 ANSI 标准要求这样做。如果需要，你可以通过 GRANT 语句更改这些权限。

我们还有真正的死锁检测代码。不再有六十秒超时。新的锁定代码实现了 SQL2 事务隔离级别。

以前的版本中有很多关于文档不足的抱怨。Thomas 在许多新文档上投入了大量精力。包括新的数据库管理员指南、入门指南、程序员指南和教程。文档树的 v6.3 版本有 8.6MB，而这个版本的文档源有 7.2MB。

出于性能原因，时间旅行已被移除，但可以用触发器实现（参见 `pgsql/contrib/spi/README`）。请查看新的 \d 命令以了解类型、操作符等。此外，视图现在有自己的权限，不再基于底层表，因此必须单独设置视图的权限。查看 `/pgsql/interfaces` 了解与 PostgreSQL 交互的一些新方式。

这是第一个真正需要为现有用户做解释的版本。在许多方面，这是必要的，因为版本 6.3 为很多问题提供了最终的解决方案，这些问题以前只能通过低效的变通方法处理。
这是强烈建议的升级；它包含许多重要的新特性和对现有特性的修复。

E.131.1. 迁移到版本 6.3

希望从任何以前的 PostgreSQL 版本迁移数据的用户，需要使用 pg_dump 或 pg_dumpall 进行转储/恢复。

E.131.2. 变更

```
缺陷修复
---------
修复被 MOVE 实现破坏的二进制游标 (Vadim)
修复 tcl 库崩溃 (Jan)
修复数组处理的问题，来自 Gerhard Hintermayer
修复 acl 错误并移除重复的 pqtrace (Bruce)
修复 psql \e 对空文件的问题 (Bruce)
修复 varchar() 字段上的 textcat (Bruce)
修复 DBT Sendproc (Zeugswetter Andres)
修复 vacuum analyze 语法问题 (Bruce)
```

修复国际标识符的问题 (Tatsuo)
修复继承表上的聚合 (Bruce)
修复 substr() 的越界数据问题
修复 select 1=1 or 2=2、select 1=1 and 2=2 和 select sum(2+2) (Bruce)
修复无 tty 输出以显示状态结果。-q 选项仍会关闭它 (Bruce)
修复 count(*)、带视图和多表的聚合以及 sum(3) (Bruce)
修复 cluster (Bruce)
修复 PQtrace 多次启动/停止的问题 (Bruce)
修复各种锁定问题，如较新的锁等待者先于较旧者
获得锁、写者等待锁时读者不共享锁、等待的写者
没有优先于等待的读者 (Bruce)
修复从外部文件执行查询时 psql 的崩溃 (James)
修复多列 order by 且第一列含 NULL 值的问题 (Jeroen)
为 float8 和 int4 使用正确的哈希表支持函数 (Thomas)
重新启用 CREATE OPERATOR 语句中的 JOIN= 选项 (Thomas)
更改布尔操作符的优先级以符合预期行为 (Thomas)
对过大整数生成 elog(ERROR) (Bruce)
允许约束子句中的多参数函数 (Thomas)
检查布尔输入字面量 'true'、'false'、'yes'、'no'、'1'、'0',
无法识别时抛出 elog(ERROR) (Thomas)
重要的大对象修复
修复 GROUP BY 显示重复项的问题 (Vadim)
修复 MergeJoin 中的索引扫描 (Vadim)

增强

带 EXISTS、IN、ALL、ANY 关键字的子查询 (Vadim、Bruce、Thomas)
新用户手册 (Thomas 等)
通过内联一些频繁调用的函数提速
真正的死锁检测，不再超时 (Bruce)
添加 SQL92 "常量" CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP、
CURRENT_USER (Thomas)
修改约束语法以符合 SQL92 (Thomas)
用索引实现 SQL92 PRIMARY KEY 和 UNIQUE 子句 (Thomas)
识别 FOREIGN KEY 的 SQL92 语法。抛出 elog 通知 (Thomas)
允许 NOT NULL UNIQUE 约束子句（以前各自单独允许） (Thomas)
允许非常量的 PostgreSQL 风格转换 ("::") (Thomas)
添加对 SQL3 TRUE 和 FALSE 布尔常量的支持 (Thomas)
支持 IS TRUE/IS FALSE/IS NOT TRUE/IS NOT FALSE 的 SQL92 语法 (Thomas)
允许更短的布尔字面量字符串（如 "t"、"tr"、"tru") (Thomas)
允许 SQL92 分隔标识符 (Thomas)
实现 SQL92 二进制和十六进制字符串解码 (b'10' 和 x'1F') (Thomas)
支持字面量字符串类型强制的 SQL92 语法
(如 "DATETIME 'now'") (Thomas)
添加 int2、int4 和 OID 类型与 text 之间的转换 (Thomas)
构建索引时使用共享锁 (Vadim)
事务块内为用户查询分配的内存在该查询完成后释放，
此功能在 <= 6.2.1 中被关闭 (Vadim)
新 SQL 语句 CREATE PROCEDURAL LANGUAGE (Jan)
新 PostgreSQL 过程语言 (PL) 后端接口 (Jan)
把 pg_dump -H 选项改名为 -h (Bruce)
添加 Java 对口令和欧洲日期的支持 (Peter)
为 LIKE 和 ~、!~ 操作使用索引 (Bruce)
添加 datetime 和 timespan 的哈希函数 (Thomas)

移除时间旅行 (Vadim、Bruce)
为 \d 和 \z 添加分页并修复 \i (Bruce)
向后端和前端库添加 Unix 域套接字支持 (Goran)
实现 CREATE DATABASE/WITH LOCATION 和 initlocation 工具 (Thomas)
允许更多 SQL92 和/或 PostgreSQL 保留字作为列标识符 (Thomas)
增强对 SQL92 SET TIME ZONE... 的支持 (Thomas)
SET/SHOW/RESET TIME ZONE 使用 TZ 后端环境变量 (Thomas)
实现 SET keyword = DEFAULT 和 SET TIME ZONE DEFAULT (Thomas)
启用使用 TZ 环境变量的 SET TIME ZONE (Thomas)
把 PGDATESTYLE 环境变量添加到前端和后端初始化 (Thomas)
添加 PGTZ、PGCOSTHEAP、PGCOSTINDEX、PGRPLANS、PGGEQO
前端库初始化环境变量 (Thomas)
回归测试时区用 "setenv PGTZ PST8PDT" 自动设置 (Thomas)
添加 pg_description 表以保存表、列、操作符、类型和
聚合的信息 (Bruce)
把系统表/索引名 16 字符的限制增加到 32 字符 (Bruce)
重命名系统索引 (Bruce)
向 SET DATESTYLE 添加 'GERMAN' 选项 (Thomas)
定义带 "hh:mm:ss" 字段的 "ISO 风格" timespan 输出格式 (Thomas)
允许 delta time 的小数值 (如 '2.5 days') (Thomas)
更仔细地验证 delta time 的数字输入 (Thomas)
实现把年积日作为 date_part() 的可能输入 (Thomas)
定义 timespan_finite() 和 text_timespan() 函数 (Thomas)
移除归档内容 (Bruce)
允许有独立于系统口令文件的 pg_password 认证数据库 (Todd)
转储 ACL、GRANT、REVOKE 权限 (Matt)
定义 text、varchar 和 bpchar 字符串长度函数 (Thomas)
修复继承的查询处理和代价计算 (Bruce)
实现 CREATE TABLE/AS SELECT (SELECT/INTO 的替代) (Thomas)
允许约束中的 NOT、IS NULL、IS NOT NULL (Thomas)
实现 SELECT 的 UNION (Bruce)
向 INSERT 添加 UNION、GROUP、DISTINCT (Bruce)
varchar() 在磁盘上只存储必要的字节 (Bruce)
修复 BLOB 的问题 (Peter)
JDBC 超大补丁……变更列表见 README_6.3 (Peter)
从 PQconnectdb() 移除未使用的 "option"
新 LOCK 命令和描述死锁的锁手册页 (Bruce)
添加新 psql \da、\dd、\df、\do、\dS 和 \dT 命令 (Bruce)
增强 psql \z 以显示序列 (Bruce)
在 psql \d 表中显示 NOT NULL 和 DEFAULT (Bruce)
新 psql .psqlrc 文件启动 (Andrew)
修改 contrib/linux 中的示例启动脚本以显示 syslog (Thomas)
contrib/ip_and_mac 中 IP 和 MAC 地址的新类型 (TomH)
contrib/unixdate 中与日期/时间类型的 Unix 系统时间转换 (Thomas)
contrib 内容的更新 (Massimo)
向 DBD::Pg 添加 Unix 套接字支持 (Goran)
新 python 接口 (PyGreSQL 2.0) (D'Arcy)
新前端/后端协议有版本号和网络字节序 (Phil)
pg_hba.conf 的安全特性得到增强和记录，大量清理 (Phil)
CHAR() 现在比 VARCHAR() 或 TEXT 访问更快
ecpg 嵌入式 SQL 预处理器
降低系统列开销 (Vadmin)
移除 pg_time 表 (Vadim)
添加 pg_type 属性以标识需要长度的类型 (bpchar、varchar)

COPY 命令失败时添加对问题行的报告
允许视图权限与底层表分开设置。
为安全起见，酌情对视图使用 GRANT/REVOKE (Jan)
表现在没有默认的 GRANT SELECT TO PUBLIC。你必须
显式授予此类权限。
清理教程示例 (Darren)

源代码树变更

添加新的 html 开发工具和 /tools/backend 中的流程图
修复 SCO 编译
Stratus 计算机移植 (Robert Gillies)
为 BSD44_derived & i386_solaris 添加 shlib 支持
使 configure 更加自动化 (Brook)
添加检查回归测试结果的脚本
把解析器函数拆分为更小的文件并分组 (Bruce)
把 heap_create 改名为 heap_create_and_catalog，把 heap_creatr
改名为 heap_create() (Bruce)
Sparc/Linux 的锁定补丁 (TomS)
移除 PORTNAME 并重组移植特定内容 (Marc)
添加优化器 README 文件 (Bruce)
移除优化器中的一些递归并清理其中一些代码 (Bruce)
修复 NetBSD 锁定 (Henry)
修复 libptcl 的制作 (Tatsuo)
AIX 补丁 (Darren)
把 IS TRUE、IS FALSE、... 更改为使用 "=" 的表达式而不是
对 istrue() 或 isfalse() 的函数调用以允许优化 (Thomas)
各种 NetBSD/Sparc 相关修复 (TomH)
Alpha linux 锁定 (Travis、Ryan)
把 elog(WARN) 改为 elog(ERROR) (Bruce)
FreeBSD 的 FAQ (Marc)
把 PostODBC 源代码树纳入我们的标准发行版 (Marc)
HP/UX 10 与 9 的小补丁 (Stan)
新 pg_attribute.atttypmod 用于 varchar 长度等类型特定信息 (Bruce)
UnixWare 补丁 (Billy)
自旋锁 asm 的新 i386 'lock' (Billy)
移除对多路复用后端的支持
开始 OpenBSD 移植
开始 AUX 移植
开始 Cygnus 移植
向回归套件添加字符串函数 (Thomas)
扩展一些以前截断为 16 字符的函数名 (Thomas)
移除不需要的 malloc() 调用并用 palloc() 替换 (Bruce)

E.132. 版本 6.2.1

发行日期
1997-10-17

6.2.1 是 6.2 之上的缺陷修复和易用性版本。

摘要：

- 按照 SQL92，允许字符串跨行。
- 包含在表更新时插入用户名的示例触发器函数。

这是 6.2 之上的次要缺陷修复版本。从 6.2 之前的系统升级需要完整的转储/重载。请参阅 6.2 版本的发行说明了解操作说明。

E.132.1. 从版本 6.2 迁移到版本 6.2.1

这是次要缺陷修复版本。从 6.2 版本不需要转储/重载，但从 6.2 之前的任何版本需要。

从 6.2 版本升级时，如果你选择转储/重载，你会发现 avg(money) 现在计算正确。所有其他缺陷修复在更新可执行文件后生效。

避免转储/重载的另一种方法是从 `psql` 使用以下 SQL 命令更新现有的系统表：

```
update pg_aggregate set aggfinalfn = 'cash_div_flt8'
 where aggname = 'avg' and aggbasetype = 790;
```

这需要对每个现有数据库执行，包括 template1。

E.132.2. 变更

允许 `TIME` 和 `TYPE` 作为列名 (`Thomas`)
允许更大范围的 `true/false` 布尔值 (`Thomas`)
支持 `"now"` 和 `"current"` 的输出 (`Thomas`)
正确处理 `INSERT NULL` 与 `DEFAULT` (`Vadim`)
修复缓冲区管理器中关系引用计数的问题 (`Vadim`)
像 `ANSI` 一样允许字符串跨行 (`Thomas`)
修复带 `ORDER BY` 的反向游标 (`Vadim`)
修复 `avg(cash)` 计算 (`Thomas`)
修复 `ORDER/GROUP BY` 中重复指定列的问题 (`Vadim`)
记录返回受影响行数的新 `libpq` 函数 `PQcmdTuples` (`Bruce`)
用于在 `INSERT/UPDATE` 时插入用户名的触发器函数 (`Brook Milligan`)

E.133. 版本 6.2

发行日期
1997-10-02

希望从以前的 PostgreSQL 版本迁移数据的用户，需要进行转储/恢复。

E.133.1. 从版本 6.1 迁移到版本 6.2

此迁移需要完整转储 6.1 数据库并在 6.2 中恢复。

注意，应使用 6.2 的 `pg_dump` 和 `pg_dumpall` 工具来转储 6.1 数据库。如果使用 6.2 之前版本的这些工具，把 template0 复制到 template1 数据库的步骤会失败。

E.133.2. 从版本 1.*x* 迁移到版本 6.2

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.133.3. 变更

```
缺陷修复
---------
修复 pg_dump 对继承、序列、归档表的问题 (Bruce)
修复 Solaris 上因移位、无符号和坏原型
        导致的溢出编译错误 (Diab Jerius)
修复几何线段算术中的缺陷（错误的交点计算） (Thomas)
在端点处检查几何相交以避免舍入难看 (Thomas)
捕获无效的删除尝试 (Vadim)
把时间函数名改得更一致 (Michael Reifenberg)
检查除零 (Michael Reifenberg)
修复使命令更改/插入的行对命令本身可见的很久以来的缺陷
        （因此曾出现已更新行的多次更新等） (Vadim)
修复 SELECT null, 'fail' FROM pg_am (Patrick)
现在允许 SELECT NULL as EMPTY_FIELD (Patrick)
从 contrib/pginterface 移除不需要的信号代码
修复 OR (where x != 1 or x isnull 不返回 x 为 NULL 的行) (Vadim)
修复 time_cmp 函数 (Vadim)
修复 WHERE 子句中第一个参数非属性的
       函数的处理 (Vadim)
修复条目顺序与目标列表顺序不同时的
       GROUP BY (Vadim)
修复 pg_dump 对无 sfunc1 聚合的处理 (Vadim)

增强
------------
默认遗传优化器 GEQO 参数现在是 8 (Bruce)
允许在目标列表中使用函数内带聚合的参数 (Vadim)
添加 JDBC 驱动作为接口 (Adrian & Peter)
pg_password 工具
返回 INSERT/UPDATE/DELETE 等插入/影响的行数 (Vadim)
用 CREATE TRIGGER 实现触发器 (SQL3) (Vadim)
SPI (服务器编程接口) 允许在 C 函数内执行查询 (Vadim)
实现 NOT NULL (SQL92) (Robson Paniago de Miranda)
包含用于字符串处理、外连接和并集的保留字 (Thomas)
用独占状态实现扩展注释 ("/* ... */") (Thomas)
添加 "//" 单行注释 (Bruce)
移除操作符名中字符的一些限制 (Thomas)
实现表的 DEFAULT 和 CONSTRAINT (SQL92) (Vadim & Thomas)
添加文本连接操作符和函数 (SQL92) (Thomas)
支持 WITH TIME ZONE 语法 (SQL92) (Thomas)
支持 INTERVAL unit TO unit 语法 (SQL92) (Thomas)
定义类型 DOUBLE PRECISION、INTERVAL、CHARACTER
       和 CHARACTER VARYING (SQL92) (Thomas)
定义类型 FLOAT(p) 和基础的 DECIMAL(p,s)、NUMERIC(p,s) (SQL92) (Thomas)
定义 EXTRACT()、POSITION()、SUBSTRING() 和 TRIM() (SQL92) (Thomas)
定义 CURRENT_DATE、CURRENT_TIME、CURRENT_TIMESTAMP (SQL92) (Thomas)
为 UNION、HAVING、INNER 和 OUTER JOIN 添加语法和警告 (SQL92) (Thomas)
```

添加更多保留字，主要为符合 SQL92 (Thomas)
允许 timespan/reltime 类型的 hh:mm:ss 时间输入 (Thomas)
为 lseg、path、polygon 添加 center() 例程 (Thomas)
为 circle-polygon、polygon-polygon 添加 distance() 例程 (Thomas)
用轴交叉算法显式检查多边形内包含的点和多边形 (Thomas)
添加 circle-box 转换例程 (Thomas)
合并不同几何数据类型的冲突操作符 (Thomas)
把距离操作符 "<===>" 替换为 "<->" (Thomas)
把"上方"操作符 "!^" 替换为 ">^", "下方"操作符 "!|" 替换为 "<^" (Thomas)
添加两端文本修剪、子串和字符串位置例程 (Thomas)
添加转换例程 circle(box) 和 poly(circle) (Thomas)
允许内部排序存储在内存中而不是文件中 (Bruce & Vadim)
允许内部相同类型上的函数和操作符成功 (Bruce)
通过性能分析加快后端启动 (Bruce)
内联频繁调用的函数以提升性能 (Bruce)
减少 open() 调用 (Bruce)
psql: 为 \h 和 \? 添加 PAGER, \C 修复
修复无 tty 时 psql 分页器的问题 (Bruce)
新 entab 工具 (Bruce)
用于参照完整性的一般触发器函数 (Vadim)
用于时间旅行的一般触发器函数 (Vadim)
用于 AUTOINCREMENT/IDENTITY 特性的一般触发器函数 (Vadim)
MOVE 实现 (Vadim)

源代码树变更

HP-UX 10 补丁 (Vladimir Turin)
添加 SCO 支持 (Daniel Harris)
MkLinux 补丁 (Tatsuo Ishii)
把几何 box 术语从 "length" 改为 "width" (Thomas)
废弃几何代码中的临时未存储斜率字段 (Thomas)
从 INSTALL 移除重启说明 (Bruce)
先在 /usr/ucb 中查找 install (Bruce)
修复 c++ 复制示例代码 (Thomas)
在 psql 手册页添加 -o (Bruce)
防止 relname 未分配字符串长度被复制进数据库 (Bruce)
清理 NAMEDATALEN 的使用 (Bruce)
修复输出中超过 15 字符的 pg_proc 名 (Bruce)
添加 strNcpy() 函数 (Bruce)
移除一些不必要的 (void) 转换 (Bruce)
新的 interfaces 目录 (Marc)
用对 fd.c 函数的调用替换 fopen() 调用 (Bruce)
尽可能把函数设为 static (Bruce)
把未使用的函数包在 #ifdef NOT_USED 中 (Bruce)
移除时间戳支持中对 difftime() 的调用以修复 SunOS (Bruce & Thomas)
针对 Digital Unix 的变更
pg_dumpall 的可移植性修复 (Bruce)
把 pg_attribute.attnvals 改名为 attdispersion (Bruce)
"intro/unix" 手册页现为 "pgintro" (Bruce)
"built-in" 手册页现为 "pgbuiltin" (Bruce)
"drop" 手册页现为 "drop_table" (Bruce)
添加 "create_trigger"、"drop_trigger" 手册页 (Thomas)
添加约束回归测试 (Vadim & Thomas)
添加注释语法回归测试 (Thomas)

添加 PGINDENT 和支持程序 (Bruce)
对所有 *.c 和 *.h 文件运行 PGINDENT 的大规模提交 (Bruce)
文件移到 /src/tools 目录 (Bruce)
SPI 和触发器编程指南 (Vadim & D'Arcy)

E.134. 版本 6.1.1

发行日期
1997-07-22

E.134.1. 从版本 6.1 迁移到版本 6.1.1

这是次要缺陷修复版本。从 6.1 版本不需要转储/重载，但从 6.1 之前的任何版本需要。更多细节请参阅 6.1 的发行说明。

E.134.2. 变更

修复带选项的 SET (Thomas)
允许 pg_dump/pg_dumpall 保留所有表/对象的所有权 (Bruce)
新 psql \connect 选项允许只更改用户名而不更改数据库
修复 initdb --debug 选项 (Yoshihiko Ichikawa))
lextest 清理 (Bruce)
哈希修复 (Vadim)
修复日期/时间的月份边界算术 (Thomas)
修复某些移植的时区夏令时处理 (Thomas、Bruce、Tatsuo)
timestamp 彻底改造为使用标准函数 (Thomas)
日期/时间例程的其他代码清理 (Thomas)
psql 的 \d 现在不区分大小写 (Bruce)
psql 的反斜杠命令现在可以带尾部分号 (Bruce)
修复使用 \g 时 psql 的内存泄漏 (Bruce)
与服务器通信的字节序处理重大修复 (Thomas、Tatsuo)
Solaris 汇编器和头文件的修复 (Yoshihiko Ichikawa)
允许用户名中使用下划线 (Bruce)
pg_dumpall 现在返回正确状态，可移植性修复 (Bruce)

E.135. 版本 6.1

发行日期
1997-06-08

回归测试已经为 PostgreSQL 6.1 版本做了适配和大量修改。

新增了三种数据类型（datetime、timespan 和 circle），已加入到 PostgreSQL 本地类型集中。点、框、路径和多边形在各数据类型间的输出格式已统一。misc.out 中的多边形输出只相对原始回归输出做了抽查。

PostgreSQL 6.1 引入了使用遗传算法的新替代优化器。当查询包含多个限定条件或多个表时（优化器可选择求值顺序），这些算法会给查询结果的顺序引入随机行为。
已有几个回归测试被修改为显式排序结果，因此对优化器的选择不敏感。
少数回归测试针对本质上无序的数据类型（如点和时间间隔），涉及这些类型的测试显式地用 `set geqo to 'off'` 和 `reset geqo` 括起来。

数组说明符（原子值周围的花括号）
的解释似乎在最初的回归测试生成之后的某个时候发生了变化。当前的 `./expected/*.out` 文件反映了这一新解释，它可能并不正确！

float8 回归测试至少在某些平台上失败。这是由于 `pow()` 和 `exp()` 实现的差异以及上溢/下溢条件所用的信号机制不同，不是 PostgreSQL 的缺陷。

random 测试中的“random”结果会使“random”测试被标记为“failed”，因为回归测试是用简单的 diff 评估的。不过，“random”在我的测试机（Linux/gcc/i686）上似乎不会产生随机结果。仔细检查输出会发现输出是有效的。

E.135.1. 迁移到版本 6.1

此迁移需要完整转储 6.0 数据库并在 6.1 中恢复。

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.135.2. 变更

```
缺陷修复
---------
库例程中的包长度检查
锁管理器优先级补丁
检查 float8 的下溢/上溢 (Bruce)
多表连接修复 (Vadim)
SIGPIPE 崩溃修复 (Darren)
大对象修复 (Sven)
允许 btree 索引处理 NULL (Vadim)
时区修复 (D'Arcy)
无行时 select SUM(x) 可返回 NULL (Thomas)
内部优化器、执行器缺陷修复 (Vadim)
修复 < 或 <= 内层循环无行的问题 (Vadim)
防止连接索引子句重新交换 (Vadim)
修复多表的连接子句 (Vadim)
修复数组的哈希和哈希连接 (Vadim)
修复 abstime 类型的 btree (Vadim)
大对象修复 (Raymond)
修复哈希索引中的缓冲区泄漏 (Vadim)
修复 rtree 以用于内层扫描 (Vadim)
修复 gist 以用于内层扫描并清理 (Vadim、Andrea)
避免不必要的本地缓冲区分配 (Vadim、Massimo)
修复事务中止时本地缓冲区泄漏 (Vadim)
修复文件管理器内存泄漏并清理 (Vadim、Massimo)
修复存储管理器内存泄漏 (Vadim)
修复 btree 重复项处理 (Vadim)
修复 vacuum 导致已删行复活的问题 (Vadim)
修复 SELECT varchar()/char() INTO TABLE 生成零长度字段的问题 (Bruce)
用 Purify 修复许多 psql、pg_dump 和 libpq 内存泄漏 (Igor)
```

增强

属性优化统计 (Bruce)
快得多的新 btree 批量装载代码 (Paul)
批量装载代码加入 BTREE UNIQUE (Vadim)
新锁调试代码 (Massimo)
libpg++ 的大规模变更 (Leo)
新 GEQO 优化器加速多表优化 (Martin)
对唯一键非唯一插入的新 WARN 消息 (Marc)
update x=-3 (无空格) 现在有效 (Bruce)
移除标识符的大小写敏感处理 (Bruce、Thomas、Dan)
调试后端现在美化打印树 (Darren)
新 Oracle 字符函数 (Edmund)
新明文口令函数 (Dan)
"no such class or insufficient privilege" 拆分为不同的消息 (Dan)
新 ANSI timestamp 函数 (Dan)
新 ANSI Time 和 Date 类型 (Thomas)
在后端中移动大数据块 (Martin)
多列 btree 索引 (Vadim)
新 SET var TO value 命令 (Martin)
读取时更新事务状态 (Dan)
字符类型的新区域设置 (Oleg)
新 SEQUENCE 序列号生成器 (Vadim)
现在可以使用 GROUP BY 函数 (Vadim)
重新组织回归测试 (Thomas、Marc)
新优化器操作权重 (Vadim)
新 psql \z 授予/许可选项 (Marc)
新 MONEY 数据类型 (D'Arcy、Thomas)
tcp 套接字通信速度改进 (Vadim)
VACUUM 的属性统计及特定列新选项 (Vadim)
许多几何类型改进 (Thomas、Keith)
附加回归测试 (Thomas)
新 datestyle 变量 (Thomas、Vadim、Martin)
更多用于排序的比较操作符 (Thomas)
新转换函数 (Thomas)
新更紧凑的 btree 格式 (Vadim)
允许 pg_dumpall 保留数据库所有权 (Bruce)
新 SET GEQO=# 和 R_PLANS 变量 (Vadim)
旧 (!GEQO) 优化器可使用右侧计划 (Vadim)
SQL 解析器类型检查改进 (Bruce)
新 SET、SHOW、RESET 命令 (Thomas、Vadim)
新 \connect database USER 选项
新 destroydb -i 选项 (Igor)
新 \dt 和 \di psql 命令 (Darren)
SELECT "\n" 现在转义换行 (A. Duursma)
从旧格式转换的新几何转换函数 (Thomas)

源代码树变更

新配置脚本 (Marc)
添加 readline 配置选项 (Marc)
移除操作系统特定的配置选项 (Marc)
新的操作系统特定模板文件 (Marc)

不再需要编辑 Makefile.global (Marc)
重新安排头文件 (Marc)
nextstep 补丁 (Gregor Hoffleit)
移除 Windows 特定代码 (Bruce)
移除 postmaster -e 选项，现在只有 postgres -e 选项 (Bruce)
合并前后端中重复的库代码 (Martin)
现在可与 eBones、国际 Kerberos 一起工作 (Jun)
更多共享库支持
c++ 头文件清理 (Bruce)
对有缺陷的 flex 发出警告 (Bruce)
DG/UX、Ultrix、IRIX、AIX 可移植性修复

E.136. 版本 6.0

发行日期
1997-01-29

希望从以前的 PostgreSQL 版本迁移数据的用户，需要进行转储/恢复。

E.136.1. 从版本 1.09 迁移到版本 6.0

此迁移需要完整转储 1.09 数据库并在 6.0 中恢复。

E.136.2. 从 1.09 之前版本迁移到版本 6.0

从更早的 1.* 版本迁移的用户应先升级到 1.09，因为 COPY 输出格式从 1.09 开始得到了改进。

E.136.3. 变更

缺陷修复

ALTER TABLE 缺陷—运行中的 postgres 进程需要重新读取表定义
允许对一个表或整个数据库运行 vacuum (Bruce)
数组修复
修复数组内存写的越界 (Kurt)
修复难以捉摸的 btree 范围/非范围缺陷 (Dan)
修复某些类型（如 time 和 date）上的哈希索引
修复 pg_log 大小爆炸的问题
修复 lo_export() 的权限 (Bruce)
修复未初始化的内存读取 (Kurt)
修复 ALTER TABLE ... char(3) 缺陷 (Bruce)
修复了几个小的内存泄漏
修复 EXPLAIN 的选项处理并更改了 full_path 选项名
修复组 acl 权限的输出
内存泄漏（用 Purify 等工具查找并消灭）(Kurt)
规则系统的小改进
NOTIFY 修复
新的运行检查断言
彻底改造 parser/analyze 代码以正确报告错误并提高速度

Pg_dump -d 现在正确处理 NULL (Bruce)
防止 SELECT NULL 使服务器崩溃 (Bruce)
INSERT ... SELECT 列不匹配时正确报告错误
插入列名不正确时正确报告错误
psql \g filename 现在可用 (Bruce)
修复 psql 一行多语句多输出的问题
移除重复的系统 OID
SELECT * INTO TABLE . GROUP/ORDER BY 在表已存在时给出 unlink 错误 (Bruce)
几项对使后端崩溃查询的修复
插入字符串中起始引号的错误 (Bruce)
提交空查询现在返回空状态，而不只是 " " 查询 (Bruce)

增强

添加 EXPLAIN 手册页 (Bruce)
添加 UNIQUE 索引能力 (Dan)
添加主机名/用户级访问控制，而不只是主机名和用户
为 <> 添加 != 同义词 (Bruce)
允许 "select oid,* from table"
允许 BY、ORDER BY 按编号或按非别名的 table.column 指定列 (Bruce)
允许从前端 COPY (Bryan)
允许 GROUP BY 使用别名列名 (Bruce)
允许真正的压缩，而不只是同一页上的重用 (Vadim)
允许安装配置选项自动添加所有本地用户 (Bryan)
允许 libpq 区分文本值 '' 和 null (Bruce)
允许有 createdb 权限的非 postgres 用户 destroydb
允许限制谁可以创建 C 函数 (Bryan)
允许限制谁可以做后端 COPY (Bryan)
可以收缩表、pg_time 和 pg_log (Vadim & Erich)
更改调试级别 2 为只打印查询，更改调试标题布局 (Bruce)
把十进制常量的默认表示从 float4 改为 float8 (Bruce)
postmaster 启动时现在设置欧洲日期格式
找不到精确大小写时执行小写函数名
聚合/GROUP 处理的修复，允许 'select sum(func(x),sum(x+y) from z'
Gist 现在包含在发行版中 (Marc)
本地用户的 Ident 认证 (Bryan)
实现 BETWEEN 限定符 (Bruce)
实现 IN 限定符 (Bruce)
libpq 有 PQgetisnull() (Bruce)
libpq++ 改进
initdb 的新选项 (Bryan)
Pg_dump 允许转储 OID (Bruce)
Pg_dump 为提高速度在表装载后创建索引 (Bruce)
Pg_dumpall 转储所有数据库和用户表
Pginterface 对 NULL 值的补充 (Bruce)
防止以 root 运行 postmaster
psql \h 和 \? 现在可读 (Bruce)
psql 允许行内任意位置的反斜杠和分号 (Bruce)
psql 更改查询中或引号中行的命令提示符 (Bruce)
psql char(3) 现在在 \d 输出中显示为 (bp)char (Bruce)
psql 返回码现在更准确 (Bryan?)
psql 更新的帮助语法 (Bruce)
重新检查并修复 vacuum (Vadim)
减小回归 diff 的大小，移除时区名差异 (Bruce)

移除编译期参数以支持二进制发行 (Bryan)
反转 HBA 掩码的含义 (Bryan)
本地用户的安全认证 (Bryan)
加快 vacuum (Vadim)
Vacuum 现在有 VERBOSE 选项 (Bruce)

源代码树变更

所有函数现在都有与调用比较的原型
允许从 Makefile.global 轻松禁用断言 (Bruce)
把代码中使用的 oid 常量改为 #define 名称
解耦 sparc 和 solaris 定义 (Kurt)
Gcc -Wall 干净编译，警告只来自无法修复的构造
主要的头文件重组/精简 (Marc)
编译失败时 make 现在停止 (Bryan)
Makefile 重构 (Bryan、Marc)
把 bsdi_2_1 合并到 bsdi (Bruce)
移除 Monitor 程序
从 Postgres95 改名为 PostgreSQL
新 config.h 文件 (Marc、Bryan)
PG_VERSION 现在设为 6.0 并被 postmaster 使用
可移植性补充，包括 Ultrix、DG/UX、AIX 和 Solaris
减少 #define 的数量，集中化 #define
移除系统表中的重复 OID (Dan)
移除重复的系统目录信息或报告不匹配 (Dan)
移除了许多操作系统特定的 #define
重构目标文件的生成/位置 (Bryan、Marc)
重构移植特定文件的位置 (Bryan、Marc)
纠正未使用/未初始化的变量

E.137. 版本 1.09

发行日期
1996-11-04

抱歉，我们没有记录 1.02 到 1.09 的变更。6.0 中列出的一些变更实际上包含在 1.09 版本中。

E.138. 版本 1.02

发行日期
1996-08-01

E.138.1. 从版本 1.02 迁移到版本 1.02.1

这是 1.02.1 的新迁移文件。它包括 'copy' 变更和一个把旧 ASCII 文件转换为新格式的脚本。

> **注意**
>
> 以下说明是为了帮助希望把数据库从 Postgres95 1.01 和 1.02 迁移到 Postgres95 1.02.1 的用户。
>
> 如果你全新开始使用 Postgres95 1.02.1 而不需要迁移旧数据库，则无需执行任何操作。

为了把较旧的 Postgres95 1.01 或 1.02 版本数据库升级到 1.02.1 版本，需要以下步骤：

1. 启动新的 1.02.1 postmaster

2. 把 1.02.1 的新内建函数和操作符添加到 1.01 或 1.02 数据库。方法是用新的 1.02.1 服务器运行你自己的 1.01 或 1.02 数据库，并应用文件末尾随附的查询。这可以通过 `psql` 轻松完成。如果你的 1.01 或 1.02 数据库命名为 `testdb`，并且你已从本文件末尾剪下命令并保存到 `addfunc.sql`：

   ```
   % psql testdb -f addfunc.sql
   ```

   升级 1.02 数据库的用户在执行文件中的最后两条语句时会得到一个警告，因为它们已存在于 1.02 中。这不必担心。

E.138.2. 转储/重载程序

如果你试图重新装载用以前版本生成的 pg_dump 或文本模式 `copy tablename to stdout` 输出，需要先对 ASCII 文件运行随附的 `sed` 脚本再装入数据库。旧格式用 '.' 作为数据结束，而 '\.' 现在是数据结束标记。此外，空字符串现在装填为 '' 而不是 NULL。完整细节参见 copy 手册页。

```
sed 's/^\.$/\\./g' <in_file >out_file
```

如果你在装载较旧的二进制备份或非 stdout 副本，则没有数据结束字符，因此无需转换。

```
-- following lines added by agc to reflect the case-insensitive
-- regexp searching for varchar (in 1.02), and bpchar (in 1.02.1)
create operator ~* (leftarg = bpchar, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = bpchar, rightarg = text, procedure =
 texticregexne);
create operator ~* (leftarg = varchar, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = varchar, rightarg = text, procedure =
 texticregexne);
```

E.138.3. 变更

```
源代码维护和开发
  * 全球志愿者团队
  * 源代码树现在在 ftp.ki.net 的 CVS 中

增强
  * psql（及底层 libpq 库）现在有更多输出格式化选项，
    包括 HTML
```

```
 * pg_dump 现在输出模式和/或数据，并有许多提升完整性的修复。
 * 管理 shell 脚本中用 psql 代替 monitor。
   monitor 将在下个版本中废弃。
 * 日期/时间函数增强
 * NULL 插入/更新/比较修复/增强
 * TCL/TK 库和 shell 修复为同时支持 tck7.4/tk4.0 和 tcl7.5/tk4.1

缺陷修复（几乎多到无法列举）
 * 索引
 * 存储管理
 * 解引用前检查 NULL 指针
 * Makefile 修复

新移植
 * 新增 SolarisX86 移植
 * 新增 BSD/OS 2.1 移植
 * 新增 DG/UX 移植
```

E.139. 版本 1.01

发行日期
1996-02-23

E.139.1. 从版本 1.0 迁移到版本 1.01

以下说明是为了帮助希望把数据库从 Postgres95 1.0 迁移到 Postgres95 1.01 的用户。

如果你全新开始使用 Postgres95 1.01 而不需要迁移旧数据库，则无需执行任何操作。

为了在用 Postgres95 1.0 版本创建的数据库上运行 Postgres95 1.01 版本，需要以下步骤：

1. 把 `src/Makefile.global` 中 `NAMEDATALEN` 的定义改为 16，把 `OIDNAMELEN` 的定义改为 20

2. 决定是否要使用基于主机的认证。

   a. 如果要用，你必须在顶级数据目录（通常是 `$PGDATA` 的值）中创建名为 `pg_hba` 的文件。`src/libpq/pg_hba` 给出了示例语法。

   b. 如果不想使用基于主机的认证，可以注释掉这一行：

      ```
      HBA = 1
      ```

      （位于 `src/Makefile.global`）

      注意，基于主机的认证默认开启，如果你不执行上述 A 或 B 步骤，默认的 pg_hba 文件将拒绝所有到 Postgres 的连接，因此请执行上面的 A 或 B，或删除 pg_hba 文件。

3. 编译并安装 1.01，但不要执行 `initdb` 步骤。

4. 在做任何其他事情之前，终止你的 1.0 postmaster，并备份现有的 `$PGDATA` 目录。

5. 把 PGDATA 环境变量指向你的 1.0 数据库，但把路径设置为使用 1.01 二进制文件。
6. 把文件 $PGDATA/PG_VERSION 从 5.0 改为 5.1
7. 启动新的 1.01 postmaster
8. 把 1.01 的新内建函数和操作符添加到 1.0 数据库。方法是用新的 1.01 服务器运行你自己的 1.0 数据库，并应用随附保存在 1.0_to_1.01.sql 文件中的查询。这可以通过 psql 轻松完成。如果你的 1.0 数据库名为 testdb：

   ```
   % psql testdb -f 1.0_to_1.01.sql
   ```

   然后执行以下命令（从此处剪切粘贴）：

   ```
   -- add builtin functions that are new to 1.01

   create function int4eqoid (int4, oid) returns bool as 'foo'
   language 'internal';
   create function oideqint4 (oid, int4) returns bool as 'foo'
   language 'internal';
   create function char2icregexeq (char2, text) returns bool as 'foo'
   language 'internal';
   create function char2icregexne (char2, text) returns bool as 'foo'
   language 'internal';
   create function char4icregexeq (char4, text) returns bool as 'foo'
   language 'internal';
   create function char4icregexne (char4, text) returns bool as 'foo'
   language 'internal';
   create function char8icregexeq (char8, text) returns bool as 'foo'
   language 'internal';
   create function char8icregexne (char8, text) returns bool as 'foo'
   language 'internal';
   create function char16icregexeq (char16, text) returns bool as
    'foo'
   language 'internal';
   create function char16icregexne (char16, text) returns bool as
    'foo'
   language 'internal';
   create function texticregexeq (text, text) returns bool as 'foo'
   language 'internal';
   create function texticregexne (text, text) returns bool as 'foo'
   language 'internal';

   -- add builtin functions that are new to 1.01

   create operator = (leftarg = int4, rightarg = oid, procedure =
    int4eqoid);
   create operator = (leftarg = oid, rightarg = int4, procedure =
    oideqint4);
   create operator ~* (leftarg = char2, rightarg = text, procedure =
    char2icregexeq);
   create operator !~* (leftarg = char2, rightarg = text, procedure =
    char2icregexne);
   create operator ~* (leftarg = char4, rightarg = text, procedure =
    char4icregexeq);
   ```

```
create operator !~* (leftarg = char4, rightarg = text, procedure =
 char4icregexne);
create operator ~* (leftarg = char8, rightarg = text, procedure =
 char8icregexeq);
create operator !~* (leftarg = char8, rightarg = text, procedure =
 char8icregexne);
create operator ~* (leftarg = char16, rightarg = text, procedure =
 char16icregexeq);
create operator !~* (leftarg = char16, rightarg = text, procedure
 = char16icregexne);
create operator ~* (leftarg = text, rightarg = text, procedure =
 texticregexeq);
create operator !~* (leftarg = text, rightarg = text, procedure =
 texticregexne);
```

E.139.2. 变更

不兼容变更:
* 只要用户按照 MIGRATION_from_1.0_to_1.01 文件中概述的步骤操作，1.01 与 1.0 数据库向后兼容。
  如果未采取这些步骤，1.01 与 1.0 数据库不兼容。

增强:
* 为 libpq 添加 PQdisplayTuples() 并更改 monitor 和 psql 使用它
* 新增 NeXT 移植（需要 SysVIPC 实现）
* 新增 CAST .. AS 语法
* 为 libpq 添加 PQtrace()/PQuntrace()
* 新增为表字段设置默认值的能力
* 新增了出色的 QNX 移植
* 新增了许多新的内置函数
* 在 tclsh 中支持字符串数组

E.140. 版本 1.0

发行日期
1995-09-05

E.140.1. 变更

版权变更:
* Postgres 1.0 的版权已放宽为可自由修改并可用于任何目的。
  请阅读 COPYRIGHT 文件。感谢 Michael Stonebraker 教授使之成为可能。

不兼容变更:
* 日期格式必须为 MM-DD-YYYY（或使用欧式风格时的 DD-MM-YYYY）。这遵循 SQL-92 规范。
* "double precision"（浮点数）而不是 "float8" 字面量
  （float8 现在自动提升为 double precision）
* 直接调用内部日期/时间例程（例如 abstimein）的输入现在产生错误

（内部例程继续存在，但从 SQL 解释器调用它们现在给出错误）
* 在 "create" 语句中，现在允许（且仅允许）为列加上标签

新特性：
* 为 SELECT 结果提供终端监控器 psql（附带 readline 支持）
* 为列增加唯一的行标识符，使用新的类型名 "oid"
* 服务器同时支持 SQL 和 POSTQUEL 语法
* 无重复字段名限制；可以有任意数量的同名列
* DELETE 语句可在 RSELECT 语句中使用子句
* 若启用了访问控制，则所有用户需要认证访问

E.141. Postgres95 版本 0.03

发行日期
1995-07-21

E.141.1. 变更

不兼容变更：
* BETA-0.3 与以前版本创建的数据库不兼容
（由于系统目录和索引结构的变更）。
* 双引号（"）作为字符串字面量的引号字符已废弃；
需要把它们转换为单引号（'）。
* 聚合名（如 int4sum）已按照 SQL 标准改名（如 sum）。
* CHANGE ACL 语法被 GRANT/REVOKE 语法取代。
* 浮点字面量（如 3.14）现在为 float4 类型（以前版本中是 float8）；
如果你依赖它是 float8 类型，可能需要做类型转换。如果忽略类型转换
而把浮点字面量赋给 float8 类型的字段，可能会存储不正确的值！
* LIBPQ 已彻底改造，使前端应用程序
可以连接多个后端
* pg_user 中的 usesysid 字段从 int2 改为 int4，
以允许更大范围的 Unix 用户 id。
* netbsd/freebsd/bsd 操作系统移植已合并为
单一的 BSD44_derived 移植。（感谢 Alistair Crooks）

SQL 标准符合性（以下详述使 postgres95
更符合 SQL-92 标准的变更）：
* 以下 SQL 类型现在是内建的：smallint、int(eger)、float、real、
char(N)、varchar(N)、date 和 time。

以下是对现有 postgres 类型的别名：
```
smallint -> int2
integer, int -> int4
float, real  -> float4
```
char(N) 和 varchar(N) 实现为截断的文本类型。此外，
char(N) 做空白填充。
* 单引号（'）用于引用字符串字面量；''（除
\' 外）也支持作为在字符串中插入单引号的手段
* 使用 SQL 标准聚合名（MAX、MIN、AVG、SUM、COUNT）
（此外，聚合现在可以重载，即你可以定义自己的 MAX 聚合来接受用户定义的类型。）

```
 * 移除 CHANGE ACL。添加 GRANT/REVOKE 语法。
   - 可以使用 "GROUP" 关键字向组授予权限。
       例如:
                GRANT SELECT ON foobar TO GROUP my_group;
       也支持关键字 'PUBLIC' 表示所有用户。

       一次只能向一个用户或组授予或撤销权限。

       不支持 "WITH GRANT OPTION"。只有类所有者可以更改
       访问控制
   - 默认访问控制是给用户只读访问。
     必须显式向用户授予插入/更新权限。要更改这一点，请修改
                src/backend/utils/acl.h
     中定义 ACL_WORLD_DEFAULT 的那一行

缺陷修复:
 * 修复了空表聚合不运行的缺陷。现在，
   对空表运行的聚合将返回聚合的初始条件。因此，空表的 COUNT 现在正确返回 0。
   空表的 MAX/MIN 将返回值为 NULL 的一行。
 * 允许在 monitor 中使用 \;
 * LISTEN/NOTIFY 异步通知机制现在可用
 * 规则动作体中的 NOTIFY 现在可用
 * 哈希索引可用，访问方法总体上性能应更好。
   大 btree 索引的创建应快得多。(感谢 Paul
   Aoki)

其他变更和增强:
 * 新增 EXPLAIN 语句用于解释查询执行计划
   (例如 "EXPLAIN SELECT * FROM EMP" 打印该查询的执行计划)。
 * WARN 和 NOTICE 消息不再带时间戳。要打开错误消息的时间戳，请取消注释
   src/backend/utils/elog.h 中的这一行:
       /* define ELOG_TIMESTAMPS */
 * 发生访问控制违规时，将给出消息
       "Either no such class or insufficient privilege"。
   这与找不到类时返回的消息相同。这可以打消非特权用户
   猜测特权类是否存在的念头。
 * 做了一些用户不可见的额外系统目录变更。

libpgtcl 变更:
 * 向 "pg_result" tcl 命令添加了 -oid 选项。
   pg_result -oid 返回最后插入行的 oid。如果
   最后一条命令不是 INSERT，则 pg_result -oid 返回 ""。
 * 大对象接口以 pg_lo* tcl 命令提供:
   pg_lo_open、pg_lo_close、pg_lo_creat 等。

可移植性增强和新移植:
 * flex/lex 的问题已澄清。现在，你应该可以在任何平台上使用
   flex 代替 lex。我们不再根据你使用的平台假设
   你使用什么词法分析器。
 * 现在支持 Linux-ELF 移植。已测试各种配置：已知以下配置可用:
       kernel 1.2.10, gcc 2.6.3, libc 4.7.2, flex 2.5.2, bison 1.24
   全部为 ELF 格式，

新工具:
```

```
 * ipcclean 添加到发行版中
   通常不需要运行 ipcclean，但如果你的后端崩溃
   并留下悬空的共享内存段，ipcclean 会
   帮你清理。

新文档：
 * 用户手册已修订并添加了 libpq 文档。
```

E.142. Postgres95 版本 0.02

发行日期
1995-05-25

E.142.1. 变更

```
不兼容变更：
* 创建数据库的 SQL 语句是 'CREATE DATABASE' 而不是 'CREATEDB'。
  类似地，删除数据库是 'DROP DATABASE' 而不是 'DESTROYDB'。
  但可执行文件名 'createdb' 和 'destroydb' 保持不变。

新工具：
* pgperl - Postgres95 的 Perl (4.036) 接口
* pg_dump - 转储 postgres 数据库的工具
* 启动 Postgres95 服务器的 initdb 工具

高级特性：
* 打开数据库的异步 Postgres 信号量 I/O
* 支持基于主机的访问控制文件 postmaster/pg_hba.conf
```

E.143. Postgres95 版本 0.01

发行日期
1995-05-01

初始版本。

附录 F. 源代码仓库

PostgreSQL的源代码使用Git版本控制系统进行存储和管理。官方提供了主仓库的公开镜像；主仓库发生任何变更后，该镜像都会在一分钟内更新。

我们的 wiki http://wiki.postgresql.org/wiki/Working_with_Git上有一些关于使用 Git 的讨论。

注意，从源代码仓库构建PostgreSQL需要相当新的bison、flex和Perl版本。从发行版压缩包构建时不需要这些工具，因为它们所构建的文件已包含在压缩包中。其他工具要求与第 14 章中所示相同。

F.1. 通过 Git 获取源代码

使用Git时，你将在本地机器上创建整个代码仓库的一个副本，因此即使离线也能访问全部历史和分支。这是开发或测试补丁最快且最灵活的方式。

Git

1. 你需要安装Git，可以从http://git-scm.com获取。许多系统默认已安装了较新版本的Git，或者可以在其软件包分发系统中找到。

2. 要开始使用 Git 仓库，请克隆官方镜像：

   ```
   git clone git://git.postgresql.org/git/postgresql.git
   ```

   这会把整个仓库复制到你的本地机器，因此可能需要一段时间才能完成，尤其是在网络连接较慢的情况下。文件将被放置在当前目录的一个新子目录`postgresql`中。

   如果防火墙阻止了对 Git 协议的访问，也可以通过 HTTP 协议访问 Git 镜像。只需把 URL 前缀改为`http`，例如：

   ```
   git clone http://git.postgresql.org/git/postgresql.git
   ```

   HTTP 协议的效率低于 Git 协议，因此使用起来会更慢。

3. 每当想要获取系统的最新更新时，`cd`进入该仓库目录并运行：

   ```
   git fetch
   ```

Git除了获取源代码外还能做更多事情。欲了解更多信息，请查阅Git的手册页，或访问http://git-scm.com网站。

附录 G. 文档

PostgreSQL有四种主要的文档格式：

- 纯文本，用于安装前信息
- HTML，用于在线浏览和参考
- PDF 或 PostScript，用于打印
- 手册页，用于快速参考。

另外，若干纯文本`README`文件可以在 PostgreSQL源代码树各处找到，用于说明各种实现方面的问题。

HTML文档和手册页属于标准发行包的一部分，并会默认安装。PDF 和 PostScript 格式的文档可单独下载。

G.1. DocBook

文档源码使用DocBook编写，这是一种表面上类似于HTML的标记语言。
这两种语言都是标准通用标记语言（SGML）的应用，后者本质上是一种描述其他语言的语言。
下文会同时使用 DocBook 和SGML这两个术语，但从技术上讲，它们不能互换。

DocBook允许作者指定技术文档的结构和内容，而无需操心展示细节。
文档样式定义这些内容如何被渲染成若干最终形式之一。DocBook 由 OASIS group[1]维护。官方 DocBook 站点[2]提供了很好的入门和参考文档，以及一本可供在线阅读的完整 O'Reilly 图书。NewbieDoc Docbook Guide[3]对初学者非常有帮助。FreeBSD Documentation Project[4]也使用 DocBook，并提供了一些很有价值的信息，其中包括若干值得参考的风格指南。

G.2. 工具集

以下工具用于处理文档。其中一些可能是可选的，具体见说明。

DocBook DTD[5]

这是 DocBook 自身的定义。我们当前使用 4.2 版；不能使用更高或更低的版本。注意 DocBook 还有一个 XML 版本 – 不要使用它。

ISO 8879 character entities[6]

DocBook 需要这些实体，但由于它们由 ISO 维护，因此单独分发。

OpenJade[7]

这是处理 SGML 的基础软件包。它包含一个 SGML 解析器、一个 DSSSL 处理器（即使用 DSSSL 样式表将 SGML 转换为其他格式的程序），以及若干相关工具。Jade 现在由 Open Jade 小组维护，而不再由 James Clark 维护。

[1] http://www.oasis-open.org
[2] http://www.oasis-open.org/docbook/
[3] http://newbiedoc.sourceforge.net/metadoc/docbook-guide.html
[4] http://www.freebsd.org/docproj/docproj.html
[5] http://www.oasis-open.org/docbook/sgml/
[6] http://www.oasis-open.org/cover/ISOEnts.zip
[7] http://openjade.sourceforge.net

DocBook DSSSL Stylesheets[8]

它们包含将 DocBook 源文件转换成其他格式（例如 HTML）的处理指令。

DocBook2X tools[9]

这个可选软件包用于创建手册页。它自身还有许多先决软件包。请查看网站。

JadeTeX[10]

如果你愿意，还可以安装 JadeTeX，用 TeX 作为 Jade 的格式化后端。JadeTeX 可以创建 PostScript 或 PDF 文件（后者带书签）。

不过，JadeTeX 的输出质量不如 RTF 后端。问题特别多的地方是表格，以及各种垂直和水平间距的瑕疵。另外，也没有机会对结果进行手工润色。

我们已经记录了多种安装处理文档所需工具的方法，下面将加以介绍。
这些工具也可能还有其他打包发行形式。请将软件包状态报告到文档邮件列表，
我们会把这些信息补充到这里。

G.2.1. Linux RPM 安装

大多数发行版厂商都在其发行版中提供了用于 DocBook 处理的完整 RPM 集合。安装时查找“SGML”选项，或者查找下列软件包：`sgml-common`、`docbook`、`stylesheets`、`openjade`（或 `jade`）。可能还需要 `sgml-tools`。如果你的发行版厂商没有提供这些软件包，那么应该可以使用其他足够兼容的厂商提供的软件包。

G.2.2. FreeBSD 安装

FreeBSD Documentation Project 本身就大量使用 DocBook，因此 FreeBSD 提供了完整的文档工具“ports”也就不足为奇了。要在 FreeBSD 上构建文档，需要安装以下 ports。

- `textproc/sp`
- `textproc/openjade`
- `textproc/iso8879`
- `textproc/dsssl-docbook-modular`

目前似乎没有可用于 DocBook V4.2 SGML DTD 的 port，需要手动安装它。

`/usr/ports/print` 中的一些内容（`tex`、`jadetex`）可能也值得安装。

ports 可能没有更新 `/usr/local/share/sgml/catalog` 中的主 catalog 文件。请确保其中包含下面这一行：

```
CATALOG "/usr/local/share/sgml/docbook/4.2/docbook.cat"
```

如果不想编辑该文件，也可以把环境变量 `SGML_CATALOG_FILES` 设置为以冒号分隔的 catalog 文件列表（例如上面那些）。

[8] http://docbook.sourceforge.net/projects/dsssl/index.html
[9] http://docbook2x.sourceforge.net
[10] http://jadetex.sourceforge.net

更多关于 FreeBSD 文档工具的信息，请参阅FreeBSD Documentation Project 的说明[11]。

G.2.3. Debian 软件包

Debian GNU/Linux 也提供了完整的文档工具软件包集。只需使用以下命令即可安装：

```
apt-get install jade
apt-get install docbook
apt-get install docbook-stylesheets
```

G.2.4. 从源码手动安装

手动安装 DocBook 工具的过程有些复杂，因此如果有预构建的软件包，就使用它们。这里只描述一种标准配置，使用合理且标准的安装路径，不使用任何“花哨”功能。有关细节，应查阅各软件包的文档，并阅读SGML入门资料。

G.2.4.1. 安装 OpenJade

1. OpenJade 的安装采用 GNU 风格的`./configure; make; make install`构建过程。详细信息可在 OpenJade 源码发行包中找到。简要来说：

   ```
   ./configure --enable-default-catalog=/usr/local/share/sgml/catalog
   make
   make install
   ```

   务必记住“默认目录文件”的存放位置，后面会用到它。也可以不指定，但这样以后每次使用jade时，都必须设置环境变量`SGML_CATALOG_FILES`，使其指向该文件。（如果 OpenJade 已经安装，而你希望在本地安装工具链的其余部分，也可以采用这种方法。）

2. 此外，还应该把`dsssl`目录中的文件 `dsssl.dtd`、`fot.dtd`、`style-sheet.dtd`和`catalog` 安装到某个位置，例如 `/usr/local/share/sgml/dsssl`。最简单的做法可能是复制整个目录：

   ```
   cp -R dsssl /usr/local/share/sgml
   ```

3. 最后，创建文件 `/usr/local/share/sgml/catalog`，并向其中加入这一行：

   ```
   CATALOG "dsssl/catalog"
   ```

   （这是对安装到步骤 2中的文件的相对路径引用。如果你选择了不同的安装布局，务必作相应调整。）

G.2.4.2. 安装 DocBook DTD 工具包

1. 获取DocBook V4.2 发行包[12]。

2. 创建目录`/usr/local/share/sgml/docbook-4.2`并切换到该目录。
   （具体位置无关紧要，但这个位置在本文所采用的布局中是合理的。）

   ```
   $ mkdir /usr/local/share/sgml/docbook-4.2$ cd /usr/local/share/
   sgml/docbook-4.2
   ```

[11]http://www.freebsd.org/doc/en_US.ISO8859-1/books/fdp-primer/tools.html
[12]http://www.docbook.org/sgml/4.2/docbook-4.2.zip

3. 解压归档包：

   ```
   $ unzip -a ...../docbook-4.2.zip
   ```

   （归档包会将文件解压到当前目录。）

4. 编辑文件/usr/local/share/sgml/catalog（或者安装时告诉 jade 的其他位置），并加入类似下面的一行：

   ```
   CATALOG "docbook-4.2/docbook.cat"
   ```

5. 下载ISO 8879 字符实体归档包[13]，将其解压，然后把文件放入存放 DocBook 文件的同一目录：

   ```
   $ cd /usr/local/share/sgml/docbook-4.2$ unzip ...../ISOEnts.zip
   ```

6. 在包含 DocBook 和 ISO 文件的目录中运行以下命令：

   ```
   perl -pi -e 's/iso-(.*).gml/ISO\1/g' docbook.cat
   ```

   （这会修正 DocBook 目录文件中使用的名称与 ISO 字符实体文件的实际名称之间的混淆。）

G.2.4.3. 安装 DocBook DSSSL 样式表

要安装样式表，先解压发行包，再把它移到合适的位置，例如 /usr/local/share/sgml。（归档会自动创建一个子目录。）

```
$gunzip docbook-dsssl-1.xx.tar.gz$tar -C /usr/local/share/sgml -xf
 docbook-dsssl-1.xx.tar
```

/usr/local/share/sgml/catalog 中通常的目录项也可以加上：

```
CATALOG "docbook-dsssl-1.xx/catalog"
```

由于样式表变化相当频繁，而且有时尝试其他版本是有益的，PostgreSQL不使用这个目录项。关于如何改为选择样式表，参见第 G.2.5 节。

G.2.4.4. 安装 JadeTeX

要安装和使用JadeTeX，你需要可用的 TeX和LaTeX2e安装，包括受支持的tools和 graphics宏包、Babel、AMS 字体和 AMS-LaTeX、PSNFSS扩展及“35 种字体”配套包、用于生成 PostScript的dvips程序，以及宏包fancyhdr、hyperref、minitoc、url和ot2enc。所有这些都可以在邻近的 CTAN 站点[14]上找到。TeX基础系统的安装远远超出了本介绍的范围。任何能运行TeX的系统都应该有可用的二进制软件包。

在将JadeTeX用于 PostgreSQL文档源之前，你需要增大 TeX内部数据结构的尺寸。相关细节可以在 JadeTeX的安装说明中找到。

完成之后，就可以安装JadeTeX：

```
$gunzip jadetex-xxx.tar.gz$tar xf jadetex-xxx.tar$cd jadetex$make
 install$mktexlsr
```

最后两条需要以root身份执行。

[13]http://www.oasis-open.org/cover/ISOEnts.zip
[14]http://www.ctan.org

G.2.5. 通过 configure 检测

在构建文档之前，需要运行configure脚本，就像构建PostgreSQL程序本身时一样。
检查运行即将结束时的输出，它应该类似于：

```
checking for onsgmls... onsgmls
checking for openjade... openjade
checking for DocBook V4.2... yes
checking for DocBook stylesheets... /usr/lib/sgml/stylesheets/nwalsh-
modular
checking for sgmlspl... sgmlspl
```

如果onsgmls和nsgmls都没有找到，你就不会看到最后这 4 行。nsgmls是 Jade 软件包的一部分。如果没有找到“DocBook V4.2”，则说明 DocBook DTD 工具包没有安装到 Jade 能找到的位置，或者目录文件没有正确设置。请参阅上面的安装提示。DocBook 样式表会在若干个较为标准的位置中查找，但如果你把它们放在其他地方，则应该设置环境变量DOCBOOKSTYLE指向该位置，然后重新运行configure。

G.3. 构建文档

一切设置妥当后，切换到doc/src/sgml目录，
并运行后续各小节中介绍的某个命令来构建文档。（记得使用 GNU make。）

G.3.1. HTML

要构建文档的 HTML 版本：

```
doc/src/sgml$ gmake html
```

这也是默认目标。

构建 HTML 文档时，该过程还会为索引条目生成链接信息。因此，如果你希望文档末尾带有概念索引，就需要先构建一次 HTML 文档，然后按你喜欢的任何格式再次构建文档。

为了便于在最终发行版中处理，组成 HTML 文档的各个文件可以被打包成一个 tar 归档，并在安装时解开。要创建 HTML 文档包，使用命令：

```
cd doc/src
gmake postgres.tar.gz
```

在发行版中，这些归档位于 doc 目录，并且默认随 gmake install 一起安装。

G.3.2. 手册页

我们使用 docbook2man 工具将 DocBook refentry 页面转换为适合手册页的 *roff 输出。手册页也以 tar 归档形式分发，与 HTML 版本类似。要创建手册页包，使用命令：

```
cd doc/src
gmake man.tar.gz
```

这会在 doc/src 目录中生成一个 tar 文件。

要生成高质量的手册页，可能需要使用打过补丁的转换工具版本，或者做一些手工后处理。
所有手册页在分发前都应经过人工检查。

G.3.3. 通过 JadeTeX 生成打印输出

如果你想使用 JadeTex 生成文档的可打印版本，可以使用下列命令之一：

- 要生成 DVI 版本：

  ```
  doc/src/sgml$ gmake postgres.dvi
  ```

- 要从 DVI 生成 Postscript：

  ```
  doc/src/sgml$ gmake postgres.ps
  ```

- 要生成 PDF：

  ```
  doc/src/sgml$ gmake postgres.pdf
  ```

  （当然，你也可以用 Postscript 生成 PDF 版本，但如果直接生成 PDF，它将带有超链接和其他增强特性。）

G.3.4. 通过 RTF 生成打印输出

你也可以把PostgreSQL文档转换为 RTF并用办公套件做一些小的格式修正，从而生成可打印版本。视具体办公套件的能力而定，随后可以把文档转换为 PostScript 或 PDF。下面的过程以 Applixware 为例说明。

> **注意**
>
> 当前版本的PostgreSQL文档似乎会触发 OpenJade 的某个错误，或者超出其大小限制。如果RTF版本的构建过程长时间挂起且输出文件大小仍为 0，你可能是遇到了这个问题。（不过请记住，正常构建也需要 5 到 10 分钟，所以不要太早中止。）

Applixware RTF 清理

OpenJade没有为正文文本指定默认样式。过去，
这个未确诊的问题导致目录生成过程极其漫长。不过，在 Applixware方面的大力帮助下，症状已得到诊断，并且有了可用的变通方法。

1. 输入以下命令来生成 RTF 版本：

   ```
   doc/src/sgml$ gmake postgres.rtf
   ```

2. 修复 RTF 文件，使其正确指定所有样式，特别是默认样式。如果文档包含 `refentry`小节，还必须替换把前一段落与当前段落绑定的格式提示，改为把当前段落与后一段落绑定。`doc/src/sgml` 中提供了一个实用程序`fixrtf`来完成这些修复：

   ```
   doc/src/sgml$ ./fixrtf --refentry postgres.rtf
   ```

   该脚本会添加`{\s0 Normal;}`作为文档的第 0 号样式。按照Applixware的说法，RTF 标准不允许添加隐式的第 0 号样式，不过 Microsoft Word 恰好能处理这种情况。对于修复`refentry`小节，脚本会把`\keepn` 标记替换为`\keep`。

3. 在Applixware Words中打开一个新文档，然后导入RTF文件。

4. 使用Applixware生成新目录（ToC）。
   a. 从第一行第一个字符开始到最后一行最后一个字符为止，选中现有的 ToC 行。
   b. 使用Tools → Book Building → Create Table of Contents构建新的 ToC。选择让 ToC 包含前三级标题。这会用Applixware原生的 ToC 替换从 RTF 导入的现有行。
   c. 使用Format → Style 调整 ToC 格式，依次选择三种 ToC 样式，并调整`First` 和`Left`的缩进。使用以下数值：

样式	首行缩进（英寸）	左缩进（英寸）
`TOC-Heading 1`	`0.4`	`0.4`
`TOC-Heading 2`	`0.8`	`0.8`
`TOC-Heading 3`	`1.2`	`1.2`

5. 在整个文档中完成以下工作：
   - 调整分页。
   - 调整表格列宽。
6. 用正确的值替换 ToC 中 Examples 和 Figures 部分右对齐的页码。这只需要几分钟。
7. (可选的) 如果索引节为空，从文档中删除它。
8. 重新生成并调整目录。
   a. 选中 ToC 域。
   b. 选择Tools → Book Building → Create Table of Contents。
   c. 通过选择Tools → Field Editing → Unprotect 解除 ToC 的保护。
   d. 删除 ToC 中的第一行，那是 ToC 自身的条目。
9. 将文档保存为Applixware Words原生格式，以便日后更容易地进行最后时刻的编辑。
10. 把文档 “打印” 到 PostScript 格式的文件。

G.3.5. 纯文本文件

若干文件以纯文本形式分发，供安装过程中阅读。`INSTALL` 文件对应第 14 章，并针对不同语境作了少量调整。要重新生成该文件，请切换到 `doc/src/sgml` 目录，然后输入 **`gmake INSTALL`**。这会创建一个 `INSTALL.html` 文件，可以用 Netscape Navigator 把它保存为文本，并放到现有文件的位置。在 HTML 到文本的转换方面，Netscape 似乎能提供最好的质量（优于 lynx 和 w3m）。

`HISTORY` 文件可以用命令 **`gmake HISTORY`** 类似地创建。对于文件 `src/test/regress/README`，命令是 **`gmake regress_README`**。

G.3.6. 语法检查

构建文档可能非常耗时。但有一种方法可以只检查文档文件的语法是否正确，这只需要几秒钟：

```
doc/src/sgml$ gmake check
```

G.4. 文档编写

SGML和DocBook的开源编写工具并不多。
最常用的工具组合是Emacs/XEmacs编辑器加上适当的编辑模式。在某些系统上，
典型的完整安装会包含这些工具。

G.4.1. Emacs/PSGML

PSGML是编辑SGML文档最常用、也最强大的模式。正确配置后，可以使用
Emacs插入标签并检查标记一致性。它也可以用于HTML。下载、安装说明和详细文档请参阅
PSGML 网站[15]。

使用PSGML时有一点需要注意：它的作者假定主SGML DTD目录是`/usr/local/lib/sgml`。如果像本章示例一样使用`/usr/local/share/sgml`，就必须作相应调整，可以设置`SGML_CATALOG_FILES`环境变量，也可以定制PSGML的安装配置（其手册介绍了具体方法）。

将以下内容放入 `~/.emacs` 环境文件中（将路径名调整为适合你系统的值）：

```
; ********** for SGML mode (psgml)

(setq sgml-omittag t)
(setq sgml-shorttag t)
(setq sgml-minimize-attributes nil)
(setq sgml-always-quote-attributes t)
(setq sgml-indent-step 1)
(setq sgml-indent-data t)
(setq sgml-parent-document nil)
(setq sgml-default-dtd-file "./reference.ced")
(setq sgml-exposed-tags nil)
(setq sgml-catalog-files '("/usr/local/share/sgml/catalog"))
(setq sgml-ecat-files nil)

(autoload 'sgml-mode "psgml" "Major mode to edit SGML files." t )
```

并在同一文件中为 SGML 加入一项，添加到现有的变量定义中。该变量是 `auto-mode-alist`：

```
(setq
  auto-mode-alist
  '(("\\.sgml$" . sgml-mode)
   ))
```

目前，每个 SGML 源文件的末尾都有下面这样的块：

```
<!-- Keep this comment at the end of the file
Local variables:
mode: sgml
sgml-omittag:t
sgml-shorttag:t
sgml-minimize-attributes:nil
sgml-always-quote-attributes:t
sgml-indent-step:1
sgml-indent-data:t
```

[15] http://www.lysator.liu.se/projects/about_psgml.html

```
sgml-parent-document:nil
sgml-default-dtd-file:"./reference.ced"
sgml-exposed-tags:nil
sgml-local-catalogs:("/usr/lib/sgml/catalog")
sgml-local-ecat-files:nil
End:
-->
```

即使你不设置 `~/.emacs` 文件，这也会配置好许多编辑模式参数；但有点不幸的是，如果你遵循了上面的安装说明，那么 catalog 路径将与你的安装位置不匹配。因此，你可能需要关闭局部变量：

```
(setq inhibit-local-variables t)
```

PostgreSQL 发布中包含一个已解析的 DTD 定义文件 `reference.ced`。使用 PSGML 时，你可能会发现，处理这些分别保存书籍各部分的文件，有一种方便的方式：在编辑时插入适当的 `DOCTYPE` 声明。例如，当前这个源码文件是一章附录，因此可以将它指定为 DocBook 文档的 “appendix” 实例，把第一行写成这样：

```
<!DOCTYPE appendix PUBLIC "-//OASIS//DTD DocBook V4.2//EN">
```

这样，所有读取 SGML 的工具都能正确处理该文档，并且可以用 `nsgmls -s docguide.sgml` 验证文档。（但在构建整套文档之前，需要移除这一行。）

G.4.2. 其他 Emacs 模式

GNU Emacs 自带另一种 SGML 模式，功能不如 PSGML 强大，但更容易理解，也更轻量。它还提供语法高亮（font lock），这可能很有帮助。

Norm Walsh 提供了一种专用于 DocBook 的主模式[16]，也带有 font-lock 和若干减少输入量的功能。

G.5. 风格指南

G.5.1. 参考页

参考页应遵循标准布局。这能让用户更快找到所需信息，
也能促使作者记录命令的所有相关方面。不仅希望PostgreSQL参考页彼此保持一致，
也希望它们与操作系统及其他软件包提供的参考页保持一致。因此制定了下列准则。
它们在很大程度上与多个操作系统所建立的类似准则保持一致。

描述可执行命令的参考页应按以下顺序包含这些小节。不适用的小节可以省略。
只有在特殊情况下才应增加顶层小节；这些信息通常应放在 “用法” 小节中。

名称

本节自动生成。其中包含命令名以及一句简短的功能摘要。

概要

本节包含命令的语法图。概要通常不应列出每一个命令行选项；这些内容会在下文说明。
相反，它应列出命令行的主要组成部分，例如输入文件和输出文件的位置。

[16] http://nwalsh.com/emacs/docbookide/index.html

描述

用若干段解释该命令的作用。

选项

用列表描述每个命令行选项。如果选项很多，可以使用小节。

退出状态

如果程序以 0 表示成功、以非零表示失败，那么你不需要记录这一点。如果不同的非零退出代码有各自的含义，就在这里列出它们。

用法

描述程序的任何子语言或运行时接口。如果程序不是交互式的，通常可以省略本节。否则，本节就是用于描述运行时特性的总括性小节。适当时可以使用小节。

环境

列出程序可能使用的所有环境变量。尽量完整；即便像 `SHELL` 这样看似微不足道的变量，也可能让用户感兴趣。

文件

列出程序可能会隐式访问的任何文件。也就是说，
不要列出在命令行上指定的输入和输出文件，而应列出配置文件等内容。

诊断

解释程序可能产生的任何不寻常输出。不要去列出所有可能的错误消息。这项工作量很大，而在实践中用处不大。不过，例如如果错误消息有一种用户可以解析的标准格式，
那么这里就是解释它的地方。

注解

放不进其他地方的任何内容，尤其包括错误、实现缺陷、安全性考虑和兼容性问题。

示例

示例

历史

如果该程序的历史上有一些重要里程碑，可以在这里列出。通常本节可以省略。

另见

交叉引用按以下顺序列出：其他 PostgreSQL 命令参考页、PostgreSQL SQL 命令参考页、对 PostgreSQL 手册的引用、其他参考页（例如操作系统或其他软件包）、其他文档。同一组中的条目按字母顺序排列。

描述 SQL 命令的参考页应包含下列小节：名称、概要、描述、参数、输出、注解、示例、兼容性、历史、另见。参数小节类似于选项小节，但在决定列出命令的哪些子句时有更大的自由度。只有当命令返回的内容不是默认的命令完成标签时，才需要输出小节。
兼容性小节应解释该命令在多大程度上符合 SQL 标准，或者它与哪种其他数据库系统兼容。SQL 命令的另见小节应先列出 SQL 命令，再列出对程序的交叉引用。

附录 H. 外部项目

PostgreSQL是一个复杂的软件项目，管理它并不容易。我们发现，对 PostgreSQL的许多增强功能，如果与核心项目分开开发会更有效率。独立的项目可以有各自的开发团队、邮件列表、错误跟踪和发布时间表。虽然这种独立性使开发更容易，
但它也让用户的工作变得更难：他们不得不四处寻找满足自己需要的数据库增强功能。
本节介绍一些较受欢迎的外部开发的增强功能，并指导你如何找到它们。

许多与PostgreSQL相关的项目托管在 GBorg[1] 或 pgFoundry[2]。还有一些与PostgreSQL相关的项目托管在其他地方，但你需要通过 Internet 搜索才能找到它们。

H.1. 外部开发的接口

PostgreSQL的基础发行版中只包含很少的接口。libpq被打包在内，因为它是主要的C接口，
而且许多其他接口都构建在它之上。ecpg被打包在内，因为它与服务器端文法紧密相关，
对数据库版本有很强的依赖性。所有其他接口都是独立的项目，必须单独安装。

一些较受欢迎的接口包括：

psqlODBC

这是Windows应用程序最常用的接口。

pgjdbc

一个 JDBC接口。

Npgsql

面向较新Windows应用程序的 .Net接口。

libpqxx

一个较新的 C++接口。

libpq++

一个较老的 C++接口。

pgperl

一个 Perl接口，其API类似于 libpq。

DBD-Pg

一个使用 DBD标准API的 Perl接口。

pgtclng

Tcl接口的较新版本。

pgtcl

Tcl接口的原始版本。

[1] http://gborg.postgresql.org
[2] http://pgfoundry.org

PyGreSQL

一个 Python接口库。

所有这些都可以在 GBorg[3] 或 pgFoundry[4] 找到。

H.2. 扩展

PostgreSQL从一开始就被设计为可扩展的。因此，
装载到数据库中的扩展可以像随数据库打包的功能一样工作。源码发行版中的 `contrib/`目录包含大量扩展。该目录中的 `README`文件包含一个摘要。它们包括转换工具、全文索引、XML工具以及更多的数据类型和索引方法。其他扩展是独立开发的，例如 PostGIS。甚至PostgreSQL的复制解决方案也是在外部开发的。例如，Slony-I是一种流行的主/从复制解决方案，
它独立于核心项目开发。

PostgreSQL有多种可用的管理工具。其中最受欢迎的是 pgAdmin，另外还有一些商业化的工具。

[3]http://gborg.postgresql.org
[4]http://pgfoundry.org

参考书目

这里列出了一些关于 SQL 和 PostgreSQL 的参考文献与读物。

来自最初的 POSTGRES 开发团队的一些白皮书和技术报告，可在加州大学伯克利分校计算机系的网站[1]上获取。

SQL 参考书

Reference texts for SQL features.

[bowman01] The Practical SQL Handbook. Bowman et al, 2001. Using SQL Variants. Fourth Edition. Judith Bowman、Sandra Emerson和Marcy Darnovsky. ISBN 0-201-70309-2. 2001. Addison-Wesley Professional. 版权 © 2001.

[date97] A Guide to the SQL Standard. Date and Darwen, 1997. A user's guide to the standard database language SQL. Fourth Edition. C. J. Date和Hugh Darwen. ISBN 0-201-96426-0. 1997. Addison-Wesley. 版权 © 1997 Addison-Wesley Longman, Inc..

[date04] An Introduction to Database Systems. Date, 2004. Eighth Edition. C. J. Date. ISBN 0-321-19784-4. 2003. Addison-Wesley. 版权 © 2004 Pearson Education, Inc..

[elma04] Fundamentals of Database Systems. Fourth Edition. Ramez Elmasri和Shamkant Navathe. ISBN 0-321-12226-7. 2003. Addison-Wesley. 版权 © 2004.

[melt93] Understanding the New SQL. Melton and Simon, 1993. A complete guide. Jim Melton和Alan R. Simon. ISBN 1-55860-245-3. 1993. Morgan Kaufmann. 版权 © 1993 Morgan Kaufmann Publishers, Inc..

[ull88] Principles of Database and Knowledge. Base Systems. Ullman, 1988. Jeffrey D. Ullman. Volume 1. Computer Science Press. 1988.

PostgreSQL 专题文档

本节收录相关文档。

[sim98] Enhancement of the ANSI SQL Implementation of PostgreSQL. Simkovics, 1998. Stefan Simkovics.

该文讨论了 SQL 的历史和语法，并描述了将 `INTERSECT` 和 `EXCEPT` 构造加入 PostgreSQL 的过程。它是作者在维也纳技术大学、在 O. Univ. Prof. Dr. Georg Gottlob 和 Univ. Ass. Mag. Katrin Seyr 支持下完成的硕士学位论文。

November 29, 1998. Department of Information Systems, Vienna University of Technology. Vienna, Austria.

[yu95] The Postgres95. User Manual. Yu and Chen, 1995. A. Yu和J. Chen. . Sept. 5, 1995. University of California. Berkeley, California.

[fong] The design and implementation of the POSTGRES query optimizer [2]. Zelaine Fong. University of California, Berkeley, Computer Science Department.

[1] http://db.cs.berkeley.edu/papers/
[2] http://db.cs.berkeley.edu/papers/UCB-MS-zfong.pdf

会议论文和期刊文章

本节收录文章和简报。

[olson93] Partial indexing in POSTGRES: research project. Olson, 1993. Nels Olson. 1993. UCB Engin T7.49.1993 O676. University of California. Berkeley, California.

[ong90] "A Unified Framework for Version Modeling Using Production Rules in a Database System". Ong and Goh, 1990. L. Ong和J. Goh. ERL Technical Memorandum M90/33. April, 1990. University of California. Berkely, California.

[rowe87] " The POSTGRES data model [3]". Rowe and Stonebraker, 1987. L. Rowe和M. Stonebraker. VLDB Conference, Sept. 1987.

[seshadri95] " Generalized Partial Indexes [4] ". Seshardri, 1995. P. Seshadri和A. Swami. Eleventh International Conference on Data Engineering, 6-10 March 1995. 1995. Cat. No.95CH35724. IEEE Computer Society Press. Los Alamitos, California. 420-7.

[ston86] " The design of POSTGRES [5]". Stonebraker and Rowe, 1986. M. Stonebraker和L. Rowe. ACM-SIGMOD Conference on Management of Data, May 1986.

[ston87a] "The design of the POSTGRES. rules system". Stonebraker, Hanson, Hong, 1987. M. Stonebraker、E. Hanson和C. H. Hong. IEEE Conference on Data Engineering, Feb. 1987.

[ston87b] " The design of the POSTGRES storage system [6]". Stonebraker, 1987. M. Stonebraker. VLDB Conference, Sept. 1987.

[ston89] " A commentary on the POSTGRES rules system [7]". Stonebraker et al, 1989. M. Stonebraker、M. Hearst和S. Potamianos. SIGMOD Record 18(3). Sept. 1989.

[ston89b] " The case for partial indexes [8]". Stonebraker, M, 1989b. M. Stonebraker. SIGMOD Record 18(4). 4-11. Dec. 1989.

[ston90a] " The implementation of POSTGRES [9]". Stonebraker, Rowe, Hirohama, 1990. M. Stonebraker、L. A. Rowe和M. Hirohama. Transactions on Knowledge and Data Engineering 2(1). IEEE. March 1990.

[ston90b] " On Rules, Procedures, Caching and Views in Database Systems [10]". Stonebraker et al, ACM, 1990. M. Stonebraker、A. Jhingran、J. Goh和S. Potamianos. ACM-SIGMOD Conference on Management of Data, June 1990.

[3] http://db.cs.berkeley.edu/papers/ERL-M87-13.pdf
[4] http://simon.cs.cornell.edu/home/praveen/papers/partindex.de95.ps.Z
[5] http://db.cs.berkeley.edu/papers/ERL-M85-95.pdf
[6] http://db.cs.berkeley.edu/papers/ERL-M87-06.pdf
[7] http://db.cs.berkeley.edu/papers/ERL-M89-82.pdf
[8] http://db.cs.berkeley.edu/papers/ERL-M89-17.pdf
[9] http://db.cs.berkeley.edu/papers/ERL-M90-34.pdf
[10] http://db.cs.berkeley.edu/papers/ERL-M90-36.pdf

索引

符号

$，32
$libdir，465
*，78
.pgpass，383
主机名，350
事务隔离级别，204
位图扫描，197
共享库，473
函数
  调用中的类型解析，190
  输出参数，458
删除，67
别名
  查询中的表名，12
区域设置，244
历史
  PostgreSQL 的，xxv
右连接，71
合取，114
否定，114
大小写敏感性
  SQL 命令，25
字符串
  串接，121，130
密码，287
左连接，71
引号
  与标识符，26
  转义，26
引导
  期间启动服务器，244
插入，66
操作符
  调用中的类型解析，187
操作符类，201，496
数据库
  创建权限，287
数据类型
  分类，187
  转换，186
更新，67
析取，114
标识符
  长度，25
登录权限，287
磁盘使用情况，336
磁盘驱动器，338
空值
  在 PL/Perl 中，586
  比较，115
端口，351
类型转换，28
索引，195
  B-树，196
  基于表达式，198
  hash，196
  R-树，196
  唯一，198
  多列，197
  检查使用情况，202
  组合多个索引，197
  部分，199
规则
  与触发器比较，531
角色
  创建权限，287
触发器
  与规则比较，531
超级用户，287
连接，10
  右，71
  外连接，11
  左，71
  自连接，12
通知处理器，381
通知接收器，381
重载
  操作符，491
错误消息，358
非阻塞连接，352

A

ABORT，646
abs，117
acos，120
add_missing_from configuration parameter，

age，149
aggregate function，12
  built-in，170
  invocation，34
  user-defined，486
AIX
  IPC configuration，
alias
  in the FROM clause，73
  in the select list，78
ALL，172，175
ALTER AGGREGATE，647
ALTER CONVERSION，649
ALTER DATABASE，650
ALTER DOMAIN，652

ALTER FUNCTION，655
ALTER GROUP，658
ALTER INDEX，659
ALTER LANGUAGE，661
ALTER OPERATOR，662
ALTER OPERATOR CLASS，663
ALTER ROLE，287，664
ALTER SCHEMA，667
ALTER SEQUENCE，668
ALTER TABLE，670
ALTER TABLESPACE，677
ALTER TRIGGER，678
ALTER TYPE，679
ALTER USER，680
ANALYZE，314，681
AND（操作符），114
any，113
ANY，171，172，175
anyarray，113
anyelement，113
archive_command configuration parameter，

area，160
array，101
  constant，101
  constructor，35
  of user-defined type，491
ARRAY，35
  结果类型的确定，193
ascii，123
asin，120
AT TIME ZONE，157
atan，120
atan2，120
australian_timezones configuration parameter，
authentication_timeout configuration parameter，
auto-increment (见 serial)
autocommit
  bulk-loading data，219
  psql，
autovacuum
  configuration parameters，276
  general information，316
  table-specific configuration，977
autovacuum configuration parameter，
autovacuum_analyze_scale_factor configuration parameter，
autovacuum_analyze_threshold configuration parameter，
autovacuum_naptime configuration parameter，
autovacuum_vacuum_cost_delay configuration parameter，
autovacuum_vacuum_cost_limit configuration parameter，
autovacuum_vacuum_scale_factor configuration parameter，
autovacuum_vacuum_threshold configuration parameter，
average，12，170

B

B-树 (见索引)
backslash_quote configuration parameter，

backup，183，319
base type，452
BEGIN，683
BETWEEN，115
BETWEEN SYMMETRIC，115
bgwriter_all_maxpages configuration parameter，
bgwriter_all_percent configuration parameter，

bgwriter_delay configuration parameter，
bgwriter_lru_maxpages configuration parameter，
bgwriter_lru_percent configuration parameter，

bigint，28，83
bigserial，85
binary data，88
  functions，130
binary string
  length，131
bison，228
bit string
  constant，28
  data type，100
bit strings
  functions，132
bitmap scan，
bit_and，171
bit_length，121
bit_or，171
BLOB (见 large object)
block_size configuration parameter，
bonjour_name configuration parameter，
Boolean
  data type，95
  operators (见 operators, logical)
bool_and，171
bool_or，171
box (data type)，97

BSD/OS
  IPC configuration，
  shared library，
btrim，123
bytea，88
  in libpq，369

C

C，350，405
canceling
  SQL command，373
CASCADE
  with DROP，64
  foreign key action，46
CASE，166
  结果类型的确定，193
cbrt，117
ceiling，117
center，160
char，86
character，86
character set，，，307
character string
  constant，26
  data types，86
  length，121
character varying，86
char_length，121
check constraint，41
checkpoint，339
CHECKPOINT，685
checkpoint_segments configuration parameter，
checkpoint_timeout configuration parameter，
checkpoint_warning configuration parameter，
check_function_bodies configuration parameter，
chr，123
cid，112
cidr，99
circle，98
client authentication，296
  timeout during，
client_encoding configuration parameter，
client_min_messages configuration parameter，
CLOSE，686
cluster
  of databases (见 database cluster)
CLUSTER，687
clusterdb，873
cmax，48
cmin，47
COALESCE，168
column，7，39
  adding，48
  removing，49
  renaming，50
  system column，47
column data type
  changing，50
column reference，32
col_description，178
comment
  about database objects，178
  in SQL，30
COMMENT，689
COMMIT，692
COMMIT PREPARED，693
commit_delay configuration parameter，
commit_siblings configuration parameter，
comparison
  operators，114
  row-wise，175
  subquery result row，172
compiling
  libpq applications，384
composite type，108，452
  constant，109
  constructor，36
computed field，458
concurrency，204
conditional expression，166
configuration
  of the server，257
  of the server
    functions，183
configure，229
config_file configuration parameter，
connection service file，383
constant，26
constraint，41
  adding，49
  check，41
  foreign key，45
  name，41
  NOT NULL，43
  primary key，44
  removing，49
  unique，43
constraint exclusion，62，
constraint_exclusion configuration parameter，
CONTINUE

in PL/pgSQL，554
convert，121
COPY，9，694
with libpq，376
cos，120
cot，120
count，12
cpu_index_tuple_cost configuration parameter，
cpu_operator_cost configuration parameter，
cpu_tuple_cost configuration parameter，
CREATE AGGREGATE，702
CREATE CAST，705
CREATE CONSTRAINT，708
CREATE CONVERSION，709
CREATE DATABASE，291，711
CREATE DOMAIN，713
CREATE FUNCTION，715
CREATE GROUP，720
CREATE INDEX，721
CREATE LANGUAGE，724
CREATE OPERATOR，727
CREATE OPERATOR CLASS，730
CREATE ROLE，286，733
CREATE RULE，737
CREATE SCHEMA，739
CREATE SEQUENCE，741
CREATE TABLE，7，744
CREATE TABLE AS，753
CREATE TABLESPACE，294，755
CREATE TRIGGER，757
CREATE TYPE，760
CREATE USER，765
CREATE VIEW，766
createdb，3，292，875
createlang，878
createuser，286，880
cross join，
crypt，301
cstring，113
ctid，48
CTID，519
current_date，149
current_time，149
current_timestamp，149
currval，164
cursor
CLOSE，686
DECLARE，769
FETCH，802
in PL/pgSQL，558
MOVE，817
showing the query plan，799
custom_variable_classes configuration parameter，

D

data area (见 database cluster)
data type，82
base，452
composite，452
constant，28
internal organization，465
numeric，83
type cast，34
user-defined，488
database，291
creating，3
database activity
monitoring，330
database cluster，7，243
data_directory configuration parameter，
date，89，91
constants，93
current，157
output format，94
(参见 formatting)
DateStyle configuration parameter，
date_part，149，153
date_trunc，149，156
DBI，588
db_user_namespace configuration parameter，
deadlock，209
timeout during，
deadlock_timeout configuration parameter，
DEALLOCATE，768
debug_assertions configuration parameter，
debug_pretty_print configuration parameter，
debug_print_parse configuration parameter，
debug_print_plan configuration parameter，
debug_print_rewritten configuration parameter，
decimal (见 numeric)
DECLARE，769
decode，123
default value，40
changing，50
default_statistics_target configuration parameter，

default_tablespace configuration parameter，
default_transaction_isolation configuration parameter，
default_transaction_read_only configuration parameter，
default_with_oids configuration parameter，
degrees，117
DELETE，14，67，772
diameter，160
Digital UNIX (见 Tru64 UNIX)
dirty read，204
disk space，313
DISTINCT，10，79
dollar quoting，27
double precision，84
DROP AGGREGATE，774
DROP CAST，775
DROP CONVERSION，776
DROP DATABASE，293，777
DROP DOMAIN，778
DROP FUNCTION，779
DROP GROUP，781
DROP INDEX，782
DROP LANGUAGE，783
DROP OPERATOR，784
DROP OPERATOR CLASS，786
DROP ROLE，286，787
DROP RULE，788
DROP SCHEMA，789
DROP SEQUENCE，790
DROP TABLE，8，791
DROP TABLESPACE，792
DROP TRIGGER，793
DROP TYPE，794
DROP USER，795
DROP VIEW，796
dropdb，294，884
droplang，886
dropuser，286，888
duplicate，10
duplicates，79
dynamic loading， ，464
dynamic_library_path，465
dynamic_library_path configuration parameter，

E

ECPG，405
ecpg，890
effective_cache_size configuration parameter，
elog，1036
  in PL/Perl，
  in PL/Python，595
  in PL/Tcl，
embedded SQL
  in C，405
enable_bitmapscan configuration parameter，
enable_hashagg configuration parameter，
enable_hashjoin configuration parameter，
enable_indexscan configuration parameter，
enable_mergejoin configuration parameter，
enable_nestloop configuration parameter，
enable_seqscan configuration parameter，
enable_sort configuration parameter，
enable_tidscan configuration parameter，
encode，123
encryption，253
END，797
environment variable，382
ereport，1036
error codes
  libpq，
  list of，1086
escape_string_warning configuration parameter，
escaping strings，368
every，171
EXCEPT，79
exceptions
  in PL/PgSQL，556
EXECUTE，798
EXISTS，172
EXIT
  in PL/pgSQL，553
exp，117
EXPLAIN，212，799
explain_pretty_print configuration parameter，
expression
  order of evaluation，37
  syntax，31
extending SQL，452
extensions，1377
external_pid_file configuration parameter，
extract，149，153

extra_float_digits configuration parameter，

F

false，95
FAQ，xxvi
fast path，374
FETCH，802
field
  computed，458
field selection，33
flex，228
float4 (见 real)
float8 (见 double precision)
floating point，84
floating-point
  display，
floor，117
foreign key，15，45
formatting，143
format_type，178
free space map，262
FreeBSD
  IPC configuration，
  shared library，
  启动脚本，245
FROM
  missing，
from_collapse_limit configuration parameter，

fsync configuration parameter，
full_page_writes configuration parameter，

function，114
  in the FROM clause，74
  internal，464
  invocation，33
  polymorphic，453
  user-defined，453
    in C，464
    in SQL，454

G

generate_series，176
genetic query optimization，
GEQO (见 genetic query optimization)
geqo configuration parameter，
geqo_effort configuration parameter，
geqo_generations configuration parameter，

geqo_pool_size configuration parameter，
geqo_selection_bias configuration parameter，
geqo_threshold configuration parameter，
get_bit，131
get_byte，131
GiST (见 index)
global data
  in PL/Python，594
  in PL/Tcl，579
GRANT，288，806
GREATEST，168
  结果类型的确定，193
GROUP BY，13，76
grouping，76

H

hash (见索引)
has_database_privilege，178
has_function_privilege，178
has_language_privilege，178
has_schema_privilege，178
has_tablespace_privilege，178
has_table_privilege，178
HAVING，13，77
hba_file configuration parameter，
height，160
hierarchical database，7
HP-UX
  IPC configuration，
  shared library，

I

ident，302
identifier
  syntax of，25
ident_file configuration parameter，
IFNULL，168
IMMUTABLE，462
IN，172，175
index
  for user-defined data type，496
  GiST，1061
  locks，210
index scan，
inet (data type)，99
inet_client_addr，178
inet_client_port，178
inet_server_addr，178
inet_server_port，178
information schema，422
inheritance，18，55，
initcap，123
initdb，243，942
input function，488
  of a data type，488

INSERT，8，66，811
installation，226
  on Windows，227，241
instr，570
int2 (见 smallint)
int4 (见 integer)
int8 (见 bigint)
integer，28，83
integer_datetimes configuration parameter，

interfaces，1376
internal，113
INTERSECT，79
interval，89，93
ipcclean，945
IRIX
  shared library，
IS DISTINCT FROM，115，176，176
IS FALSE，116
IS NOT FALSE，116
IS NOT NULL，115，175
IS NOT TRUE，116
IS NOT UNKNOWN，116
IS NULL，115，175，
IS TRUE，116
IS UNKNOWN，116
isclosed，160
isfinite，149
ISNULL，115
isopen，160

J

join，70
  controlling the order，218
  cross，
  natural，70
  outer，
join_collapse_limit configuration parameter，

justify_days，149
justify_hours，149

K

Kerberos，301
key word
  list of，1110
  syntax of，25
krb_caseins_users configuration parameter，

krb_server_hostname configuration parameter，
krb_server_keyfile configuration parameter，
krb_srvname configuration parameter，

L

label (见 alias)
language_handler，113
large object，396
lastval，164
lc_collate configuration parameter，
lc_ctype configuration parameter，
lc_messages configuration parameter，
lc_monetary configuration parameter，
lc_numeric configuration parameter，
lc_time configuration parameter，
ldconfig，235
LEAST，168
  结果类型的确定，193
length，160
  of a binary string (见 binary strings, length)
  of a character string (见 character string, length)
libperl，227
libpq，350
libpq-fe.h，350，355
libpq-int.h，355，385
libpython，227
LIKE，133
  与区域，306
LIMIT，81
line segment，97
Linux
  IPC configuration，
  shared library，
  启动脚本，245
LISTEN，813
listen_addresses configuration parameter，

ln，117
LOAD，814
locale，305
localtime，149
localtimestamp，149
lock，207
  monitoring，335
LOCK，207，815
log，117
log_connections configuration parameter，

log_destination configuration parameter，
log_directory configuration parameter，
log_disconnections configuration parameter，

log_duration configuration parameter，

log_error_verbosity configuration parameter，
log_executor_stats configuration parameter，
log_filename configuration parameter，
log_hostname configuration parameter，
log_line_prefix configuration parameter，
log_min_duration_statement configuration parameter，
log_min_error_statement configuration parameter，
log_min_messages configuration parameter，
log_parser_stats configuration parameter，
log_planner_stats configuration parameter，
log_rotation_age configuration parameter，
log_rotation_size configuration parameter，
log_statement configuration parameter，
log_statement_stats configuration parameter，
log_truncate_on_rotation configuration parameter，
loop
  in PL/pgSQL，553
lower，121
  与区域，306
lo_close，398
lo_creat，396，399
lo_create，397，399
lo_export，397，399
lo_import，397，399
lo_lseek，398
lo_open，397
lo_read，398
lo_tell，398
lo_unlink，399，399
lo_write，398
lpad，123
lseg，97
ltrim，123

M

MAC address (见 macaddr)
macaddr (data type)，100
MacOS X
  IPC configuration，
  shared library，
maintenance，313
maintenance_work_mem configuration parameter，
make，226
MANPATH，235
max，12
max_connections configuration parameter，
max_files_per_process configuration parameter，
max_fsm_pages configuration parameter，
max_fsm_relations configuration parameter，
max_function_args configuration parameter，
max_identifier_length configuration parameter，
max_index_keys configuration parameter，
max_locks_per_transaction configuration parameter，
max_prepared_transactions configuration parameter，
max_stack_depth configuration parameter，
md5，123
MD5，301
memory context
  in SPI，626
min，12
mod，117
monitoring
  database activity，330
MOVE，817
MVCC，204

N

name
  qualified，52
  syntax of，25
  unqualified，53
natural join，70
NetBSD
  IPC configuration，
  shared library，
  启动脚本，245
network
  data types，98
nextval，164
nonblocking connection，370
nonrepeatable read，204
NOT IN，172，175
not-null constraint，42

notice processing
  in libpq，381
NOTIFY，818
  in libpq，375
NOTNULL，115
NOT（操作符），114
now，149
npoints，160
null value
  with check constraints，42
  default value，40
  in DISTINCT，79
  in libpq，366
  in PL/Python，594
  with unique constraints，44
NULLIF，168
number
  constant，28
numeric，28
numeric (data type)，84
NVL，168

O

object identifier
  data type，112
object-oriented database，7
obj_description，178
octet_length，121
OFFSET，81
OID
  libpq 中的，368
  列，47
oid，112
on-line backup，319
ONLY，70
opaque，113
OpenBSD
  IPC configuration，
  shared library，
  启动脚本，245
OpenSSL，
  (参见 SSL)
operator，114
  invocation，33
  logical，114
  precedence，30
  syntax，29
  user-defined，491
Oracle
  porting from PL/SQL to PL/pgSQL，568
ORDER BY，10，80
  与区域，306
ordering operator，502
OR（操作符），114
outer join，
output function，488
  of a data type，488
overlay，121
overloading
  functions，462
owner，288

P

palloc，473
PAM，232，304
parameter
  syntax，32
parenthesis，32
partitioning，58
password
  authentication，301
  超级用户的，244
password file，383
password_encryption configuration parameter，
PATH，235
path
  for schemas，
path (data type)，97
pattern matching，133
pclose，160
performance，212
Perl，585
permission (见 privilege)
pfree，473
PGcancel，374
PGCLIENTENCODING，383
PGconn，350
PGCONNECT_TIMEOUT，382
PGDATA，243
PGDATABASE，382
PGDATESTYLE，383
PGGEQO，383
PGHOST，382
PGHOSTADDR，382
PGKRBSRVNAME，382
PGLOCALEDIR，383
PGOPTIONS，382
PGPASSFILE，382
PGPASSWORD，382
PGPORT，382
PGREALM，382
PGREQUIRESSL，382
PGresult，361
PGSERVICE，382
PGSSLMODE，382

PGSYSCONFDIR，383
PGTZ，383
PGUSER，382
pgxs，475
pg_aggregate，972
pg_am，972
pg_amop，973
pg_amproc，974
pg_attrdef，974
pg_attribute，974
pg_authid，976
pg_auth_members，977
pg_autovacuum，977
pg_cancel_backend，183
pg_cast，978
pg_class，979
pg_client_encoding，123
pg_column_size，183
pg_config，892
  用于 libpq，385
  用于用户定义的 C 函数，472
pg_constraint，980
pg_controldata，946
pg_conversion，981
pg_conversion_is_visible，178
pg_ctl，244，947
pg_database，293，982
pg_database_size，183
pg_depend，983
pg_description，984
pg_dump，895
pg_dumpall，901
  use during upgrade，228
pg_function_is_visible，178
pg_get_constraintdef，178
pg_get_expr，178
pg_get_indexdef，178
pg_get_ruledef，178
pg_get_serial_sequence，178
pg_get_triggerdef，178
pg_get_userbyid，178
pg_get_viewdef，178
pg_group，999
pg_has_role，178
pg_hba.conf，296
pg_ident.conf，303
pg_index，984
pg_indexes，1000
pg_inherits，985
pg_language，986
pg_largeobject，986
pg_listener，987
pg_locks，1000
pg_ls_dir，183
pg_namespace，987
pg_opclass，987
pg_opclass_is_visible，178
pg_operator，988
pg_operator_is_visible，178
pg_pltemplate，989
pg_postmaster_start_time，178
pg_prepared_xacts，1002
pg_proc，989
pg_read_file，183
pg_relation_size，183
pg_reload_conf，183
pg_restore，905
pg_rewrite，991
pg_roles，1002
pg_rotate_logfile，183
pg_rules，1003
pg_service.conf，383
pg_settings，1003
pg_shadow，1004
pg_shdepend，992
pg_size_pretty，183
pg_start_backup，183
pg_statistic，217，993
pg_stats，217，1004
pg_stat_file，183
pg_stop_backup，183
pg_tables，1006
pg_tablespace，994
pg_tablespace_databases，178
pg_tablespace_size，183
pg_table_is_visible，178
pg_total_relation_size，183
pg_trigger，994
pg_type，995
pg_type_is_visible，178
pg_user，1006
pg_views，1007
phantom read，204
pi，117
PIC，473
PID
  确定服务器进程的 PID
    在 libpq 中，358
PITR，319
PL/Perl，585
PL/PerlU，591
PL/pgSQL，536
PL/Python，594
PL/SQL (Oracle)
  porting to PL/pgSQL，568
PL/Tcl，578
point，97
point-in-time recovery，319

polygon，98
polymorphic function，453
polymorphic type，453
popen，160
port configuration parameter，
position，121
POSTGRES，xxv
postgres，291
postgres user，243
Postgres95，xxv
postgresql.conf，257
postgres（程序），953
postmaster，3，244，957
power，117
PQbackendPID，358
PQbinaryTuples，366
  with COPY，377
PQcancel，374
PQclear，363
PQcmdStatus，367
PQcmdTuples，367
PQconndefaults，354
PQconnectdb，350
PQconnectPoll，352
PQconnectStart，352
PQconsumeInput，372
PQdb，356
PQendcopy，379
PQerrorMessage，358
PQescapeBytea，369
PQescapeByteaConn，369
PQescapeString，368
PQescapeStringConn，368
PQexec，359
PQexecParams，359
PQexecPrepared，361
PQfformat，365
  with COPY，377
PQfinish，355
PQflush，373
PQfmod，365
PQfn，374
PQfname，364
PQfnumber，364
PQfreeCancel，374
PQfreemem，370
PQfsize，366
PQftable，365
PQftablecol，365
PQftype，365
PQgetCancel，373
PQgetCopyData，378
PQgetisnull，366
PQgetlength，367
PQgetline，378
PQgetlineAsync，378
PQgetResult，372
PQgetssl，358
PQgetvalue，366
PQhost，356
PQisBusy，372
PQisnonblocking，373
PQmakeEmptyPGresult，364
PQnfields，364
  用于 COPY，376
PQnotifies，375
PQntuples，364
PQoidStatus，368
PQoidValue，368
PQoptions，356
PQparameterStatus，357
PQpass，356
PQport，356
PQprepare，360
PQprint，367
PQprotocolVersion，357
PQputCopyData，377
PQputCopyEnd，377
PQputline，379
PQputnbytes，379
PQrequestCancel，374
PQreset，355
PQresetPoll，355
PQresetStart，355
PQresStatus，362
PQresultErrorField，362
PQresultErrorMessage，362
PQresultStatus，361
PQsendPrepare，371
PQsendQuery，371
PQsendQueryParams，371
PQsendQueryPrepared，371
PQserverVersion，358
PQsetdb，352
PQsetdbLogin，352
PQsetErrorVerbosity，380
PQsetnonblocking，373
PQsetNoticeProcessor，381
PQsetNoticeReceiver，381
PQsocket，358
PQstatus，356
PQtrace，380
PQtransactionStatus，357
PQtty，356
PQunescapeBytea，370
PQuntrace，380
PQuser，356
predicate locking，206

preload_libraries configuration parameter，

PREPARE，820
PREPARE TRANSACTION，822
prepared statements
  creating，820
  executing，798
  removing，768
  showing the query plan，799
preparing a query
  in PL/pgSQL，536
  in PL/Python，595
  in PL/Tcl，581
pre_auth_delay configuration parameter，
primary key，44
privilege，50，288
  querying，179
  with rules，529
  for schemas，54
  with views，529
procedural language，534
  handler for，1048
ps
  to monitor activity，330
psql，5，911
Python，594

Q

qualified name，52
query，9，69
query plan，212
query tree，512
quote_ident，123
  use in PL/PgSQL，548
quote_literal，123
  use in PL/PgSQL，548

R

R-树 (见索引)
radians，117
radius，160
RAISE，562
random，117
random_page_cost configuration parameter，

range table，
read-only transaction，
readline，226
real，84
record，113
rectangle，97
redirect_stderr configuration parameter，
referential integrity，15，45
regclass，112
regexp_replace，134
regex_flavor configuration parameter，
regoper，112
regoperator，112
regproc，112
regprocedure，112
regression test，233
regression tests，341
regtype，112
regular expression，134，135
  (参见 pattern matching)
regular expressions，
reindex，317
REINDEX，824
reindexdb，935
relation，7
relational database，7
RELEASE SAVEPOINT，827
repeat，123
replace，123
reporting errors
  in PL/PgSQL，562
RESET，828
RESTRICT
  with DROP，64
  foreign key action，46
REVOKE，288，829
role，286
  membership in，288
ROLLBACK，832
rollback
  psql，
ROLLBACK PREPARED，833
ROLLBACK TO SAVEPOINT，834
round，117
row，7，39
ROW，36
row estimation
  planner，1072
row type，108
  constructor，36
rpad，123
rtrim，123
rule，512
  and views，513
  for DELETE，520
  for INSERT，520
  for SELECT，514
  for UPDATE，520

S

SAVEPOINT，836

savepoints
  defining，836
  releasing，827
  rolling back，834
scalar (见 expression)
schema，51，291
  creating，52
  current，53，178
  public，53
  removing，52
SCO OpenServer
  IPC configuration，
search path，53
  current，178
search_path，53
search_path configuration parameter，
SELECT，9，69，838
  select list，78
SELECT INTO，849
  in PL/pgSQL，546
semaphores，247
sequence，164
  and serial type，85
sequential scan，
serial，85
serial4，85
serial8，85
serializability，206
server log，271
  log file maintenance，318
server_encoding configuration parameter，
server_version configuration parameter，
SET，183，851
SET CONSTRAINTS，854
set difference，79
set intersection，79
set operation，79
set returning functions
  functions，176
SET ROLE，855
SET SESSION AUTHORIZATION，857
SET TRANSACTION，859
set union，79
SETOF，454
  (参见 function)
setseed，117
setval，164
set_bit，131
set_byte，131
shared library，234
shared memory，247
shared_buffers configuration parameter，
SHMMAX，247
SHOW，183，861
shutdown，253
SIGHUP，257，299，303
SIGINT，253
sign，117
signal
  backend processes，183
significant digits，
SIGQUIT，253
SIGTERM，253
silent_mode configuration parameter，
SIMILAR TO，134
sin，120
sliced bread (见 TOAST)
smallint，83
Solaris
  IPC configuration，
  shared library，
  启动脚本，245
SOME，171，172，175
sorting，80
SPI，597
SPI_connect，598
SPI_copytuple，630
SPI_cursor_close，617
SPI_cursor_fetch，615
SPI_cursor_find，614
SPI_cursor_move，616
SPI_cursor_open，613
SPI_exec，605
SPI_execp，612
SPI_execute，602
SPI_execute_plan，610
spi_exec_query
  in PL/Perl，
SPI_finish，599
SPI_fname，619
SPI_fnumber，620
SPI_freeplan，635
SPI_freetuple，633
SPI_freetuptable，634
SPI_getargcount，607
SPI_getargtypeid，608
SPI_getbinval，622
SPI_getnspname，626
SPI_getrelname，625
SPI_gettype，623
SPI_gettypeid，624
SPI_getvalue，621
SPI_is_cursor_plan，609
spi_lastoid，
SPI_modifytuple，632
SPI_palloc，627
SPI_pfree，629

SPI_pop，601
SPI_prepare，606
SPI_push，600
SPI_repalloc，628
SPI_returntuple，631
SPI_saveplan，618
split_part，123
sql_inheritance configuration parameter，
sqrt，117
ssh，255
SSL，254，383
  with libpq，351
  在libpq中，358
ssl configuration parameter，
ssl_renegotiation_limit configuration parameter，
STABLE，462
standard deviation，171
standard_conforming_strings configuration parameter，
START TRANSACTION，863
statement_timeout configuration parameter，

statistics，331
  of the planner，216，314
stats_block_level configuration parameter，

stats_command_string configuration parameter，
stats_reset_on_server_start configuration parameter，
stats_row_level configuration parameter，
stats_start_collector configuration parameter，

string (见 character string)
strings
  backslash quotes，
  escape warning，
  standard conforming，
strpos，123
subquery，12，35，74，172
subscript，32
substr，123
substring，121，130，134
sum，12
superuser，5
superuser_reserved_connections configuration parameter，
syntax
  SQL，25
syslog_facility configuration parameter，
syslog_ident configuration parameter，
system catalog
  schema，54

T

table，7，39
  creating，39
  inheritance，55
  modifying，48
  partitioning，58
  removing，40
  renaming，50
table expression，69
table function，74
tableoid，47
tablespace，294
  default，
tan，120
target list，
Tcl，578
tcp_keepalives_count configuration parameter，
tcp_keepalives_idle configuration parameter，

tcp_keepalives_interval configuration parameter，
template0，292
template1，292，292
temp_buffers configuration parameter，
test，341
text，86
threads
  with libpq，384
tid，112
time，89，91
  constants，93
  current，157
  output format，94
    (参见 formatting)
time span，89
time with time zone，89，91
time without time zone，89，91
time zone，95，
  Australian，
  configuration names，1098
  conversion，157
  input abbreviations，1095
timelines，319
timeofday，149
timeout
  client authentication，
  deadlock，
timestamp，89，92
timestamp with time zone，89，92
timestamp without time zone，89，92
timezone configuration parameter，
TOAST，1065

and user-defined types，491
per-column storage settings，
versus large objects，396
token，25
to_ascii，123
to_char，143
与区域，306
to_date，143
to_hex，123
to_number，143
to_timestamp，143
trace_notify configuration parameter，
trace_sort configuration parameter，
transaction，16
transaction ID
wraparound，315
transaction isolation，204
transaction isolation level，
read committed，205
serializable，205
transaction log (见 WAL)
transform_null_equals configuration parameter，
translate，123
trigger，113，504
arguments for trigger functions，505
in C，505
in PL/pgSQL，562
in PL/Python，595
in PL/Tcl，582
trim，121
Tru64 UNIX
shared library，
true，95
trunc，117
TRUNCATE，864
trusted
PL/Perl，591
type (见 data type)
polymorphic，453
type cast，34

U

UNION，79
结果类型的确定，193
unique constraint，43
Unix 域套接字，350
UnixWare
IPC configuration，
shared library，
unix_socket_directory configuration parameter，
unix_socket_group configuration parameter，
unix_socket_permissions configuration parameter，
UNLISTEN，865
unqualified name，53
UPDATE，14，67，866
upgrading，228，328
upper，121
与区域，306
user，286
current，178

V

vacuum，313
VACUUM，869
vacuumdb，938
vacuum_cost_delay configuration parameter，
vacuum_cost_limit configuration parameter，
vacuum_cost_page_dirty configuration parameter，
vacuum_cost_page_hit configuration parameter，
vacuum_cost_page_miss configuration parameter，
value expression，31
varchar，86
variance，171
version，6，178
compatibility，328
view，15
implementation through rules，513
updating，524
void，113
VOLATILE，462
volatility
functions，462

W

WAL，338
wal_buffers configuration parameter，
wal_debug configuration parameter，
wal_sync_method configuration parameter，
WHERE，75
where to log，271
WHILE
in PL/pgSQL，555
width，160
width_bucket，117
work_mem configuration parameter，

X

xid，112
xmax，47
xmin，47

Y

yacc，228

Z

zero_damaged_pages configuration parameter，
zlib，226，232