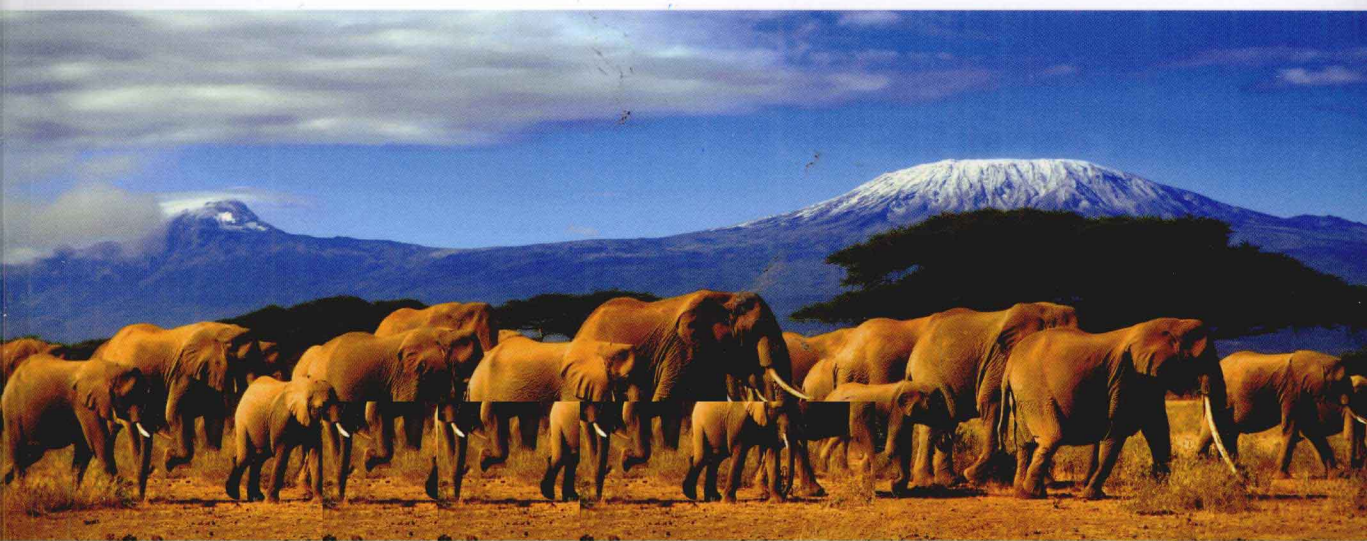


PostgreSQL

数据库内核分析



彭智勇 彭煜玮 ◎编著



机械工业出版社
China Machine Press

数据库技术

PostgreSQL

数据库内核分析

彭智勇 彭煜玮 ◎编著



机械工业出版社
China Machine Press

PostgreSQL 是目前广泛应用的开源数据库管理系统。本书从 PostgreSQL 数据库的源代码入手, 深入分析了该数据库管理系统的底层实现细节, 揭示了数据库运行的基本原理。本书的主要内容包括: PostgreSQL 数据库的体系结构、存储管理、索引机制、查询编译、查询执行、并发控制机制以及安全等。每个主题都引用了大量的数据结构、图表等进行说明, 使读者对实现过程和机理一目了然。

本书适合从事数据库领域相关研究的技术人员、高等院校相关专业高年级本科生或研究生阅读, 有助于读者进行数据库的高级开发、基于 PostgreSQL 定制满足需求的数据库系统, 或者在数据库内核中实现并验证数据库新技术。

封底无防伪标均为盗版

版权所有, 侵权必究

本书法律顾问 北京市展达律师事务所

图书在版编目 (CIP) 数据

PostgreSQL 数据库内核分析/彭智勇, 彭煜玮编著. —北京: 机械工业出版社, 2011.9
ISBN 978-7-111-35905-0

I. P… II. ①彭… ②彭… III. 关系数据库—数据库管理系统, PostgreSQL IV. TP311.138

中国版本图书馆 CIP 数据核字 (2011) 第 191369 号

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 朱 劼

北京京师印务有限公司印刷

2012 年 1 月第 1 版第 1 次印刷

186mm × 240mm · 28.75 印张

标准书号: ISBN 978-7-111-35905-0

定价: 79.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

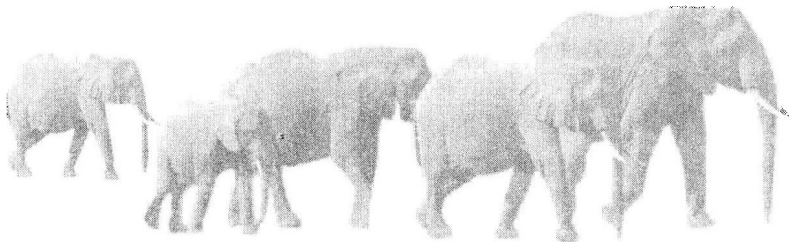
客服热线: (010) 88378991; 88361066

购书热线: (010) 68326294; 88379649; 68995259

投稿热线: (010) 88379604

读者信箱: hzjsj@hzbook.com

前 言



随着计算机与互联网技术的发展和普及，数据库技术已经渗透到人们日常生活的每个角落，无论是信息检索、电子商务、网上交友，还是日常工作时使用的各种信息管理系统，无一例外地都用到了数据库管理系统，它已经成为最常用的软件之一。

目前，市场上有多种数据库管理系统产品，商业版数据库管理系统产品有 Oracle、DB2、SQL Server 等，开源版数据库管理系统产品有 PostgreSQL、MySQL 等。商业版数据库管理系统虽占据着较大的市场份额，但随着新的应用领域的出现以及成本方面的考虑，有很多用户开始采用开源版的数据库管理系统或者在开源数据库基础之上定制自己的数据库管理系统。

开源数据库管理系统的广泛使用也吸引了大量的开发者、技术人员，与此同时也出现了大量与数据库管理系统相关的技术性书籍和教材。但这些书籍要么是介绍数据库管理系统的管理和使用技巧，要么是介绍如何在这些数据库管理系统上开发应用，鲜见从数据库底层实现的角度来分析数据库系统内部实现机制的书籍。

本书尝试从 PostgreSQL 数据库源代码入手，介绍典型开源数据库管理系统的内部实现细节，揭开一个真正的数据库管理系统的内部奥秘。写作本书的重要动机是向数据库内核开发人员以及科研人员介绍 PostgreSQL 的实现机制，帮助他们基于 PostgreSQL 定制满足特定需求的数据库系统，或者在数据库内核中实现并验证数据库的新技术。我们希望本书有助于不同背景和经验的人交换对 PostgreSQL 源代码的理解，为进一步推进 PostgreSQL 的发展及使用做出贡献。

读者定位

对本书感兴趣的读者包括：数据库内核开发人员、研究人员、DBA、高校学生、教师以及其他希望了解数据库内核实现的人们。

对于那些希望在 PostgreSQL 内核中进行改进或裁剪形成自己组织内部定制数据库的开发人员来说，本书将有助于他们了解 PostgreSQL 内部的实现机制，从而可以按照需要对数据库进行修改。

本书也可以作为计算机专业高年级本科生或者研究生的教材，尤其对于那些日后希望从事数据库领域研究的学生，我们希望本书能够激发学生对数据库内部实现机制的兴趣。

为更好地理解本书的内容，我们希望读者能具备以下基础：

- 了解数据库系统的基本概念和常用术语。
- 有使用 PostgreSQL 开发应用的实际经验。
- 阅读过介绍数据库系统实现基本原理的书籍。

本书的组织

本书的组织如下：

第 1 章介绍 PostgreSQL 的发展历史、特点、在商业领域的应用以及源代码的结构、安装、使用等内容。

第 2 章介绍 PostgreSQL 的体系结构，包括系统表的结构和用途、数据集簇、进程结构，并重点分析了守护进程 Postmaster、服务进程 Postgres 以及各个辅助进程的实现机制。

第 3 章介绍 PostgreSQL 的存储管理，对 PostgreSQL 中管理内存和外存的方式进行了详细分析。

第 4 章介绍 PostgreSQL 的索引机制，分析了 B-Tree、Hash、GiST、GIN 等索引的组织结构、操作方式，并介绍了全文搜索 TSearch 2 模块的实现与使用。

第 5 章介绍 PostgreSQL 的查询编译器，给出了如何将一个 SQL 命令进行编译、优化并最终形成执行计划的过程。

第 6 章介绍 PostgreSQL 的查询执行器，描述了执行计划如何被 PostgreSQL 进行解释执行，对执行计划中每一种节点的执行算法给出了详细分析。

第 7 章介绍 PostgreSQL 的并发控制机制，给出了事务系统和日志系统的具体实现方法。

第 8 章介绍 PostgreSQL 的安全措施，描述了用户标识与认证、权限管理和对象访问控制等有关数据库安全措施方面的实现细节。

源代码版本

本书对 PostgreSQL 的分析工作基于 PostgreSQL 8.4.1，该版本的源代码可以从以下地址下载：
<ftp://ftp-archives.postgresql.org/pub/source/v8.4.1/>。

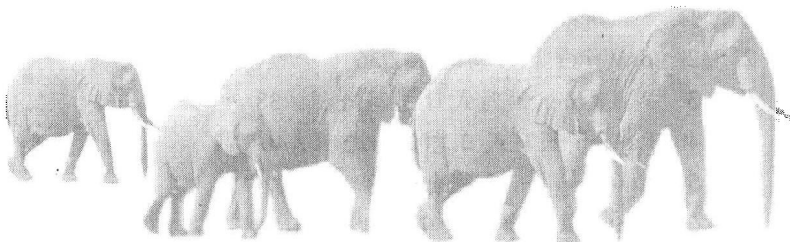
错误

由于作者的水平有限，本书可能包含笔误、差错或遗漏，我们非常希望您能够告诉我们所发现的错误以及对于本书的建议或批评。您可以发送电子邮件到 pgbook@whu.edu.cn，很抱歉我们可能无法一一回答所有的电子邮件，但是我们会将发现的错误和收到的建议整理后公布在网址 <http://totem-db.whu.edu.cn/pgbook/> 以及华章网站 <http://www.hzbook.com> 上。

致谢

本书的内容是武汉大学珞珈图腾数据库实验室对于 PostgreSQL 十年分析工作的结晶，在这十年中，实验室的各届研究生进行了不懈的努力，对此做出了巨大的贡献，在此两位作者对他们在 PostgreSQL 代码分析过程中付出的辛勤劳动表示由衷的感谢！在编写本书的过程中，作者参阅了相关的书籍、网上资料、技术文档等，在此也向这些资料的作者表示感谢！

目 录



前 言

第 1 章 PostgreSQL 系统概述 / 1

- 1.1 PostgreSQL 简介及发展历程 / 1
- 1.2 PostgreSQL 的特性 / 2
- 1.3 PostgreSQL 的应用 / 3
- 1.4 PostgreSQL 代码结构 / 3
- 1.5 安装 PostgreSQL / 4
- 1.6 PostgreSQL 数据库命令 / 5

第 2 章 PostgreSQL 的体系结构 / 7

- 2.1 系统表 / 8
 - 2.1.1 主要系统表功能及依赖关系 / 8
 - 2.1.2 系统视图 / 14
- 2.2 数据集簇 / 15
 - 2.2.1 initdb 的使用 / 17
 - 2.2.2 postgres.bki / 17
 - 2.2.3 initdb 的执行过程 / 19
 - 2.2.4 系统数据库 / 20
- 2.3 PostgreSQL 进程结构 / 21
- 2.4 守护进程 Postmaster / 22
 - 2.4.1 初始化内存上下文 / 23
 - 2.4.2 配置参数 / 23

- 2.4.3 创建监听套接字 / 28
- 2.4.4 注册信号处理函数 / 29
- 2.4.5 辅助进程启动 / 31
- 2.4.6 装载客户端认证文件 / 32
- 2.4.7 循环等待客户连接请求 / 33
- 2.5 辅助进程 / 34
 - 2.5.1 SysLogger 系统日志进程 / 34
 - 2.5.2 BgWriter 后台写进程 / 36
 - 2.5.3 WalWriter 预写式日志写进程 / 38
 - 2.5.4 PgArch 预写式日志归档进程 / 40
 - 2.5.5 AutoVacuum 系统自动清理进程 / 41
 - 2.5.6 PgStat 统计数据收集进程 / 44
- 2.6 服务进程 Postgres / 47
 - 2.6.1 初始化内存环境 / 48
 - 2.6.2 配置运行参数和处理客户端传递的 GUC 参数 / 48
 - 2.6.3 设置信号处理和信号屏蔽 / 48
 - 2.6.4 初始化 Postgres 的运行环境 / 49
 - 2.6.5 创建内存上下文并设置查询取消跳跃点 / 50
 - 2.6.6 循环等待处理查询 / 50
 - 2.6.7 简单查询的执行流程 / 52
- 2.7 小结 / 53

第3章 存储管理 / 54

- 3.1 存储管理器的体系结构 / 54
- 3.2 外存管理 / 57
 - 3.2.1 表和元组的组织方式 / 58
 - 3.2.2 磁盘管理器 / 60
 - 3.2.3 VFD 机制 / 61
 - 3.2.4 空闲空间映射表 / 64
 - 3.2.5 可见性映射表 / 69
 - 3.2.6 大数据存储 / 69
- 3.3 内存管理 / 76
 - 3.3.1 内存上下文概述 / 77
 - 3.3.2 高速缓存 / 84
 - 3.3.3 缓冲池管理 / 94
 - 3.3.4 IPC / 102

- 3.4 表操作与元组操作 / 106
 - 3.4.1 表操作 / 106
 - 3.4.2 元组操作 / 112
- 3.5 VACUUM 机制 / 115
 - 3.5.1 VACUUM 操作 / 115
 - 3.5.2 Lazy VACUUM / 116
 - 3.5.3 Full VACUUM / 118
- 3.6 ResourceOwner 资源跟踪 / 120
- 3.7 小结 / 122

第4章 索引 / 123

- 4.1 概述 / 123
 - 4.1.1 索引方式 / 124
 - 4.1.2 索引类型 / 125
 - 4.1.3 索引相关系统表 / 125
 - 4.1.4 索引的操作函数 / 129
- 4.2 B-Tree 索引 / 131
 - 4.2.1 B-Tree 索引的组织结构 / 132
 - 4.2.2 B-Tree 索引的操作 / 135
- 4.3 Hash 索引 / 146
 - 4.3.1 Hash 索引的组织结构 / 147
 - 4.3.2 Hash 索引的实现 / 153
- 4.4 GiST 索引 / 161
 - 4.4.1 GiST 的扩展性 / 161
 - 4.4.2 GiST 索引的组织结构 / 162
 - 4.4.3 GiST 索引的实现 / 163
 - 4.4.4 GiST 索引实例 / 166
- 4.5 GIN 索引 / 168
 - 4.5.1 GIN 索引的扩展性 / 168
 - 4.5.2 GIN 索引的组织结构 / 170
 - 4.5.3 GIN 索引的操作 / 172
- 4.6 TSearch2 全文搜索 / 179
 - 4.6.1 全文索引的创建 / 180
 - 4.6.2 全文索引的查询 / 184
 - 4.6.3 查询结果处理 / 185
- 4.7 小结 / 186

第5章 查询编译 / 187

- 5.1 概述 / 187
- 5.2 查询分析 / 188
 - 5.2.1 Lex 和 Yacc 简介 / 189
 - 5.2.2 词法和语法分析 / 193
 - 5.2.3 语义分析 / 208
- 5.3 查询重写 / 218
 - 5.3.1 规则系统 / 219
 - 5.3.2 查询重写的处理操作 / 224
- 5.4 查询规划 / 228
 - 5.4.1 总体处理流程 / 229
 - 5.4.2 预处理 / 237
 - 5.4.3 生成路径 / 240
 - 5.4.4 生成可优化的 MIN/MAX 聚集计划 / 253
 - 5.4.5 生成普通计划 / 254
 - 5.4.6 生成完整计划 / 257
 - 5.4.7 整理计划树 / 259
 - 5.4.8 实例分析 / 260
- 5.5 代价估计 / 266
 - 5.5.1 代价估算公式 / 267
 - 5.5.2 选择度 / 268
 - 5.5.3 单个表的扫描代价 / 268
 - 5.5.4 两个表的连接代价 / 268
- 5.6 PostgreSQL 中的遗传算法 / 268
 - 5.6.1 个体编码方式及种群初始化 / 269
 - 5.6.2 适应值 / 270
 - 5.6.3 父体选择策略 / 270
 - 5.6.4 杂交算子 / 270
 - 5.6.5 变异算子 / 275
 - 5.6.6 终止条件 / 275
 - 5.6.7 基于排列生成路径 / 276
 - 5.6.8 实例分析 / 276
- 5.7 小结 / 280

第6章 查询执行 / 282

- 6.1 查询执行策略 / 283

- 6.1.1 可优化语句和数据定义语句 / 284
- 6.1.2 四种执行策略 / 284
- 6.1.3 策略选择的实现 / 286
- 6.1.4 Portal 执行的过程 / 288
- 6.2 数据定义语句执行 / 289
 - 6.2.1 数据定义语句执行流程 / 289
 - 6.2.2 执行实例 / 290
 - 6.2.3 主要的功能处理器函数 / 293
- 6.3 可优化语句执行 / 293
 - 6.3.1 物理代数与处理模型 / 294
 - 6.3.2 物理操作符的数据结构 / 295
 - 6.3.3 执行器的运行 / 297
 - 6.3.4 执行实例 / 301
- 6.4 计划节点 / 303
 - 6.4.1 控制节点 / 304
 - 6.4.2 扫描节点 / 310
 - 6.4.3 物化节点 / 319
 - 6.4.4 连接节点 / 328
- 6.5 其他子功能介绍 / 334
 - 6.5.1 元组操作 / 334
 - 6.5.2 表达式计算 / 337
 - 6.5.3 投影操作 / 340
- 6.6 小结 / 342

第7章 事务处理与并发控制 / 343

- 7.1 事务系统简介 / 343
- 7.2 事务系统的上层 / 344
 - 7.2.1 事务块状态 / 345
 - 7.2.2 事务块操作 / 346
- 7.3 事务系统的底层 / 348
 - 7.3.1 事务状态 / 348
 - 7.3.2 事务操作函数 / 350
 - 7.3.3 简单查询事务执行过程实例 / 352
- 7.4 事务保存点和子事务 / 353
 - 7.4.1 保存点实现原理 / 353
 - 7.4.2 子事务 / 353
- 7.5 两阶段提交 / 354

- 7.5.1 预提交阶段 / 354
- 7.5.2 全局提交阶段 / 355
- 7.6 PostgreSQL 的并发控制 / 355
- 7.7 PostgreSQL 中的三种锁 / 357
 - 7.7.1 SpinLock / 357
 - 7.7.2 LWLock / 358
 - 7.7.3 RegularLock / 360
- 7.8 锁管理机制 / 368
 - 7.8.1 表粒度的锁操作 / 368
 - 7.8.2 页粒度的锁操作 / 369
 - 7.8.3 元组粒度的锁操作 / 369
 - 7.8.4 事务粒度的锁操作 / 369
 - 7.8.5 一般对象的锁操作 / 369
- 7.9 死锁处理机制 / 370
 - 7.9.1 死锁处理相关数据结构 / 371
 - 7.9.2 死锁处理相关操作 / 372
- 7.10 多版本并发控制 / 373
 - 7.10.1 MVCC 相关数据结构 / 373
 - 7.10.2 MVCC 相关操作 / 375
 - 7.10.3 MVCC 与快照 / 379
- 7.11 日志管理 / 380
 - 7.11.1 SLRU 缓冲池 / 381
 - 7.11.2 CLOG 日志管理器 / 386
 - 7.11.3 SUBTRANS 日志管理器 / 390
 - 7.11.4 MULTIXACT 日志管理器 / 392
 - 7.11.5 XLOG 日志管理器 / 394
 - 7.11.6 日志管理器总结 / 402
- 7.12 小结 / 403

第8章 数据库安全 / 404

- 8.1 PostgreSQL 安全简介 / 404
- 8.2 用户标识和认证 / 405
 - 8.2.1 客户端配置文件 / 406
 - 8.2.2 认证方法 / 408
 - 8.2.3 客户端认证 / 410
- 8.3 基于角色的权限管理 / 413
 - 8.3.1 用户和角色 / 413

8.3.2 角色相关的系统表 / 415

8.3.3 角色管理 / 416

8.4 对象访问控制 / 426

8.4.1 访问控制列表 / 426

8.4.2 对象权限管理 / 428

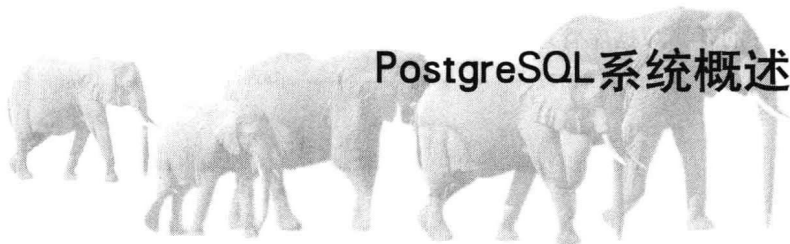
8.4.3 对象权限检查 / 432

8.5 小结 / 433

附录 A 用 Eclipse 开发和调试 PostgreSQL / 434

第 1 章

PostgreSQL 系统概述



1.1 PostgreSQL 简介及发展历程

PostgreSQL 是一种先进的对象 - 关系数据库管理系统 (ORDBMS)，它不仅支持关系数据库的各种功能，而且还具备类、继承等对象数据库的特征。它是目前功能最强大、特性最丰富和结构最复杂的开源数据库管理系统，其中有些特性甚至连商业数据库都不具备。这个起源于加州大学伯克利分校 (UCB) 的数据库研究计划，现在已经衍生成一项国际开发项目，并且拥有广泛的用户群。PostgreSQL 主要运行在 Unix 和 Linux 操作系统上 (从 8.0 版本开始推出了 Windows 平台上的版本)，并且免费开放源代码，用户可以在其官方网站 www.postgresql.org 上下载各种安装程序和文档。

PostgreSQL 的发展历程见证了数据库理论和技术的发展历程，PostgreSQL 由 UCB 计算机科学教授 Michael Stonebraker 于 1986 年创建。在此之前，Stonebraker 教授领导了关系数据库 Ingres 研究项目 (Ingres 项目的源代码可以从 www.ingres.com 免费获取)，1982 年他离开 UCB 并将 Ingres 商业化使之成为 Relational Technologies (RT) 公司的一个产品。而后该公司被 Computer Associates (CA) 公司收购。2004 年，CA 在开源许可下发布了 Ingres release 3，并继续开发销售 Ingres。而 Stonebraker 教授在返回 UCB 后开始了一项 post-Ingres 计划，该计划致力于解决基于关系模型的数据库管理系统产品的局限性，这即是 Postgres (但还不是现在的 PostgreSQL) 的开端。

从 1986 年开始，Stonebraker 教授发表了一系列论文，引入对象关系理念，探讨了新的数据库的结构设计和扩展设计。1988 年，他提出了 Postgres 的第一个原型设计，1989 年 6 月发布了版本 1，1990 年 6 月发布了带有重写后的规则系统的版本 2。1991 年发布了版本 3，在版本 3 里改进了规则系统，增加了对多种存储系统支持的能力，并且改进了查询引擎。1993 年，Postgres 用户开始剧增，并且特性需求急剧增加。在做了一些代码清理后发布了版本 4，之后 Postgres 项目正式终止。随后，Stonebraker 再次创业，成立 Illustra 公司提供对 Postgres 的商业支持，Illustra 在 1997 年被 Informix 收购，而 Stonebraker 成为 Informix 的 CTO，Informix 由于财务问题在 2001 年被 IBM 收购。

尽管 Postgres 计划终止了，但 BSD 许可证 (UCB 在其下发行的 Postgres) 却使开放源代码开发者获得副本并进一步开发系统。1994 年，两个 UCB 的研究生，Andrew Yu 和 Jolly Chen，增加了一

个 SQL 语言解释器来替代早先的基于 Ingres 的 QUEL 系统，建立了 Postgres95。代码随后被发行到互联网上。1996 年，该计划被重新命名为 PostgreSQL，以反映数据库的新查询语言 SQL，来自世界各地的数据库开发者和志愿者通过互联网协作起来，发行了 PostgreSQL 的第一个版本 6.0，并且一直维护着这套软件。自此以后，PostgreSQL 开始持续稳定地发布新版本，在新版本中有很多改进。2005 年 1 月 19 日，版本 8.0 发行，从这个版本开始，PostgreSQL 以原生的方式（即不需要模拟中间层的支持）开始支持 Windows 操作系统。

从 Michael Stonebraker 教授开始，全世界无数著名的数据库专家和优秀的黑客为 PostgreSQL 的发展做出了杰出的贡献，使 PostgreSQL 项目充满活力，不断向前发展，并使得 PostgreSQL 成为目前最好的开源数据库管理系统之一。PostgreSQL 的发展历程如图 1-1 所示。

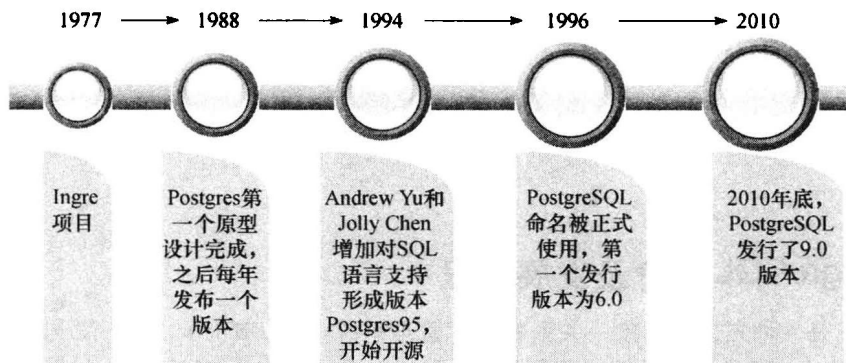


图 1-1 PostgreSQL 发展路线图

1.2 PostgreSQL 的特性

PostgreSQL 是一种几乎可以运行在各种平台上的免费的开放源码的对象关系数据库管理系统，拥有与企业级数据库相媲美的特性，如完善的 SQL 标准支持、多版本并发控制、时间点恢复、表空间机制、异步复制、嵌套事务、在线/热备份、一个复杂的查询优化器、预写日志容错技术。它支持国际字符集、多字节字符编码、Unicode，并且对格式化、排序、大小写敏感提供本地化支持。PostgreSQL 在管理大数据量方面有良好的可扩展性，对并发用户管理具有自适应性。现在已经出现具有管理超过 4 万亿字节数据能力的实用版本产品。

- **开放特性**：PostgreSQL 内置了丰富的数据类型，如任意精度的数值、无限制长度的文本、几何图元、IP 地址、数组等；同时还允许用户定义基于正规 SQL 类型的新类型，让数据库自身理解复杂数据，自定义类型中还可以包含继承关系。用户可以为数据库内几乎所有的对象定义新的类型，如索引、操作符（可重载现有操作符）、聚集函数、数据域、数据类型转换、会话（编码转换）等。
- **可编程性**：PostgreSQL 同样拥有大量的编程接口供用户开发使用，如 ODBC、JDBC（Java）、Libpq（C/C++）等。
- **可定制性**：PostgreSQL 拥有广泛的编程语言支持来实现函数功能，包括内置的 PL/pgSQL 过程语言，PL/Perl、plPHP、PL/Python、PL/Ruby、PL/Tcl 等脚本语言，以及 Java、C/C++ 等高级编程语言。函数的输出是一系列行类型的集合，可以在查询中当做表来使用，函数

也可以被定义成以创建者或者调用者的身份运行。在其他的数据库产品中，函数也会被称为“存储过程”。

- 索引手段：用户可以自定义索引方法或者使用内置 B-Tree 索引、Hash 表索引、GiST 索引、GIN 索引。GiST 索引不是某种特定的索引类型，而是一种通用索引基础结构，可以在这种结构上实现很多不同的索引策略。PostgreSQL 同时还支持如下功能：反向索引检索、表达式索引、部分索引、位图索引扫描。
- 多种身份认证方式：PostgreSQL 中可以使用数据库用户/角色、操作系统、PAM、Kerberos 等方式，根据主机配置文件（pg_hba.conf）中的设置执行对应的身份认证。

1.3 PostgreSQL 的应用

PostgreSQL 提供经济有效的、易于部署的复杂数据管理基础设施，在工业界得到广泛应用，在许多国际化大公司的应用中取得了良好的效果。

1. Sony 在线娱乐网站

Sony 在线娱乐网站（SOE）经营着很多知名游戏，因此其数据库需要快速处理大量的用户数据，其原有数据库系统为几十个 Oracle 9i RAC 集群。SOE 分析业务后发现，Oracle 数据库过于昂贵，且许可证限制过多，缺乏灵活性，最终 SOE 将其数据库系统转换为稳定、可扩展、高性能的使用 PostgreSQL 技术的 EnterpriseDB 数据库。

2. Hi5 社交网站

Hi5 从 2003 发展开始已经成为世界上最大的社交网站之一，有超过 200 个国家的八千万注册用户，每个月用户访问量超过 5600 万，是在 Alexa 全球排名前 20 的网站；Hi5 拥有一个运行在几百台服务器节点上的商业 PostgreSQL OLTP 集群，这也是世界上最大的 PostgreSQL 集群，运行着 Hi5 社交网络的所有服务，包括用户数据、配置信息、图片、评论等数据，并提供高效的检索。另外，FortiusOne、Florists' Transworld Delivery、NNT 等公司都选择廉价而高效的 PostgreSQL 作为其数据管理基础设施，取得了良好的应用效果。

PostgreSQL 经过多年的发展，得到了学术界和工业界的充分认可。获得 2008 Developer.com 编辑选择的数据库工具方向的年度产品，2000 年、2003 ~ 2006 年荣获 Linux Journal 杂志编辑评选的“最佳数据库”奖，2004 年获 ArsTechnica 最佳服务器应用奖，2002 年获 Linux New Media 杂志编辑评选的“最佳数据库”奖，1999 年获 Linux World 杂志评选的“最佳数据库”奖等多项荣誉。

1.4 PostgreSQL 代码结构

PostgreSQL 是遵照一个和 BSD 开源协议类似的协议 PostgreSQL License 发布。该许可证可以在 PostgreSQL 源代码的 COPYRIGHT 文件中找到。从该协议可以看到，开发人员或者商业组织只要遵循该协议，便可以自由地使用 PostgreSQL，可以完全控制这些第三方代码，在必要的时候可以加以修改或者二次开发。

PostgreSQL 源代码包含 3400 多个文件（截至 8.4.1 版本），主要程序由 C 语言编写，包括十几个大型模块，定义了几百个主要的数据结构和上万个函数。PostgreSQL 源代码结构清晰，每一个子目录都对应一个模块，其中主要目录（模块）及用途如下。

- Bootstrap: 用于支持 Bootstrap 运行模式, 该模式主要用来创建初始的模板数据库。
- Main: 主程序模块, 它负责将控制权转到 Postmaster 进程或 Postgres 进程。
- Postmaster: 监听用户请求的守护进程, 并控制 Postgres 进程的启动和终止。
- Libpq: C/C++ 的库函数, 处理与客户端间的通信, 几乎所有的模块都依赖它。
- Tcop: Postgres 服务进程的主要处理部分, 它调用 Parser、Optimizer、Executor、和 Commands 中的函数来执行客户端提交的查询。
- Parser: 编译器, 将 SQL 查询转化为内部查询树。
- Optimizer: 优化器, 根据查询树创建最优的查询路径和查询计划。
- Executor: 执行器, 执行来自 Optimizer 的查询计划。
- Commands: 执行非计划查询的 SQL 命令, 如创建表命令等。
- Catalog: 系统表操作, 包含用于操作系统表的专用函数。
- Storage: 管理各种类型的存储系统 (如磁盘、闪存等)。
- Access: 提供各种存取方法, 支持堆、索引等对数据的存取。
- Nodes: 定义系统内部所用到的节点、链表等结构, 以及处理这些结构的函数。
- Utils: 各种支持函数, 如错误报告、各种初始化操作等。
- Regex: 正则表达式库及相关函数, 用于支持正则表达式处理。
- Rewrite: 查询重写, 根据规则系统对查询进行重写。
- Initdb: 初始化数据库集簇。
- TSearch: 全文检索。
- Psql: 数据库交互工具。
- Port: 平台兼容性处理相关的函数。

PostgreSQL 内部模块之间的调用关系如

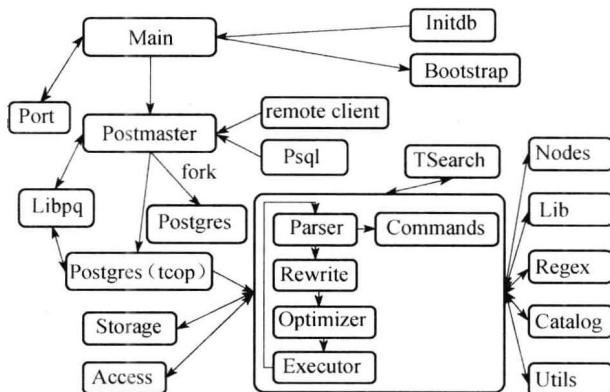


图 1-2 PostgreSQL 模块调用关系

1.5 安装 PostgreSQL

在开始使用 PostgreSQL 之前, 必须先安装它。在 Windows 下安装 PostgreSQL 比较简单, 按照安装程序的提示一步步进行即可, 这里不再赘述。在 Linux 下, 则需要从源代码进行安装, 过程如下:

- 1) 从 PostgreSQL 官方网站下载最新源代码, 通常是 .tar.gz 的压缩格式。
- 2) 通过 tar 或者图形化软件解压源码包, 假设解压到 /usr/local/src/pgsrc 目录下。
- 3) 从源代码编译安装 PostgreSQL 需要用到 C 语言开发库 (包括 gcc、ld 等), 默认配置的情况下还需要 readline、zlib 和 xml2 这几个开发库的支持, 在 Ubuntu 等 Linux 发行版下可以通过 apt-get 安装上述开发库。
- 4) 在源代码目录下运行 configure 脚本, 默认情况下 PostgreSQL 将被安装在 /usr/local/postgresql 目录下。
- 5) 在源代码目录下分别执行 make 和 make install 命令即可从源代码编译 PostgreSQL 并将相关文

件拷贝到目标目录下。

```
6) 初始化数据库集簇:  
cd/usr/local/pgsql  
mkdir data (创建数据库文件的存放目录)  
chown postgres:postgres data  
su postgres  
cd bin  
./initdb --no-locale -D ../data  
7) 启动数据库服务器:  
./pg_ctl start -D ../data  
8) 创建和访问数据库 test:  
./createdb test  
./psql test
```

1.6 PostgreSQL 数据库命令

数据库安装完成后，在 bin 目录下包含了 PostgreSQL 的工具命令。这些命令包括数据库初始化，创建、删除数据库，启动、停止数据库，连接数据库，备份、恢复数据库等丰富的工具。PostgreSQL 中各种命令的功能介绍如表 1-1 所示。

表 1-1 PostgreSQL 常用数据库命令

命令名	常用参数	功能介绍
initdb	-D 数据簇位置 -E 新建数据库默认编码 -X 事务日志位置	创建一个用于存储数据库的 PostgreSQL 数据目录，并创建预定义的模板数据库 template0 和 template1，生成共享目录表 catalog；此程序通常只在安装 PostgreSQL 时运行一次。如 initdb -D ../data
createuser	-c 指定用户的最大连接数 -d 新用户可以创建数据库 -l 新用户具有登录权限 -U 连接的用户名（不是要创建的用户）	创建一个新的 PostgreSQL 的用户（和 SQL 语句：CREATE USER 相同），如 createuser -d -l -U postgres pguser
dropuser	-U 连接数据库的用户名	删除用户，如 dropuser pguser
createdb	-E 数据库的编码 -l 数据库 locale 设置 -O 数据库所有者	创建一个新的 PostgreSQL 的数据库（和 SQL 语句 CREATE DATABASE 相同），如 createdb test
dropdb	-i 删除之前提示确认操作 -U 连接数据库的用户名	删除数据库，如 dropdb test
pg_dump	-f 备份文件名称 -F 备份文件格式，custom、tar、plain text -Z 压缩等级 -a 只备份数据、不含模式 -b 备份包含大对象 -C 包含创建数据库命令	将 PostgreSQL 数据库导出到一个备份文件，如 pg_dump test >/home/pgsql/backup/1. bak
pg_dumpall	-f 备份文件名称 --lock-wait-timeout 有表锁时若超时则备份失败	将所有的 PostgreSQL 数据库导出到一个备份文件