Broadview®
www.broadview.com.cn

The Internals of PostgreSQL
for Database Administrators and System Developers

PostgreSQL指南

内幕探索

[日]铃木启修◎著
冯若航　刘阳明　张文升◎译

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

The Internals of PostgreSQL
for Database Administrators and System Developers

PostgreSQL指南

内幕探索

[日]铃木启修◎著
冯若航　刘阳明　张文升◎译

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

本书介绍了 PostgreSQL 内部的工作原理，包括数据库对象的逻辑组织与物理实现，进程与内存的架构，并依次剖析了几个重要的子系统：查询处理、外部数据包装器、并发控制、清理过程、缓冲区管理、WAL、备份及流复制。本书为 DBA 与系统开发者提供了一幅全景概念地图，有助于读者形成对数据库实现的整体认识，亦可作为深入学习 PostgreSQL 源代码的导读手册，对于理解数据库原理与 PostgreSQL 内部实现大有裨益。

本书适合数据库开发人员及相关领域的研究人员、数据库 DBA 及高等院校相关专业的学生阅读。

图书在版编目（CIP）数据

PostgreSQL 指南：内幕探索 /（日）铃木启修（Hironobu Suzuki）著；冯若航，刘阳明，张文升译.
北京：电子工业出版社，2019.6
书名原文：The Internals of PostgreSQL for Database Administrators and System Developers
ISBN 978-7-121-35709-1

Ⅰ. ①P… Ⅱ. ①铃… ②冯… ③刘… ④张… Ⅲ. ①关系数据库系统 Ⅳ. ①TP311.138

中国版本图书馆 CIP 数据核字（2018）第 281117 号

策划编辑：符隆美
责任编辑：张春雨
印　　刷：山东华立印务有限公司
装　　订：山东华立印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：787×980　1/16　印张：15.25　字数：350 千字
版　　次：2019 年 6 月第 1 版
印　　次：2019 年 6 月第 1 次印刷
定　　价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

· 作者简介 ·

Hironobu Suzuki（铃木启修）

毕业于北海道大学信息工程研究生院，获得信息工程硕士学位，曾在多家公司担任软件开发人员和技术经理/技术主管。在数据库和系统集成领域出版了7本书。

2010年—2016年担任日本PostgreSQL用户组的主任，连续7年组织了日本PostgreSQL技术研讨会，并担任日本2013年PostgreSQL大会的委员会主席。

推荐序

PostgreSQL 已获得 DB-Engines 排行榜 2017 年和 2018 年的“年度数据库”称号，发展速度迅猛，PostgreSQL 被广泛应用的主要原因在于 PostgreSQL 和 Oracle 一样，特别适合于非常复杂的企业应用场景，代替 Oracle 的位置。同时 PostgreSQL 的开源协议允许用户非常友好地应用在自用业务、对外商用业务中，没有法律风险，也不要求用户开源等。

很多大型企业应用 PostgreSQL 也有数年，例如阿里巴巴、邮储、平安、中兴、苏宁、亚信、探探等，长期的 PostgreSQL 数据库应用使得公司内部的数据库团队在 PostgreSQL 的管理、开发、内核等各方面的经验、人才都得到了大量的积累。

文升是我多年的好友，同时也是 PostgreSQL 核心组成员，为 PostgreSQL 中文技术社区做出了巨大的贡献。

2019 年，张文升及其技术团队又要出书了，这次带来的是《The Internals of PostgreSQL for database Administrators and System Developers》一书的中文翻译书籍，作者为日本 PostgreSQL 数据库专家 Suzuki，对数据库的集群、架构、SQL 处理、外部表接口、并发控制、垃圾回收、HOT、堆表、索引存储结构、WAL 日志、时间点恢复、流复制等原理进行了深入浅出的讲解，对期望了解 PostgreSQL 内部原理的开发者、管理员、架构师来说，无疑是一本非常好的入门书籍。

感谢译者对 PostgreSQL 社区的付出，期待这本图片的出版。

德哥

2019 年 3 月

译者序

相信选择这本书的读者，大多已经对 PostgreSQL 有所了解。本书从 PostgreSQL 的整体架构展开，依次介绍了各个功能模块的来龙去脉，方便 DBA（数据库管理员）与数据库系统开发人员了解数据库内部原理、阅读学习 PostgreSQL 源码。

数据库是信息系统的核心组件，关系型数据库则是数据库皇冠上的明珠，而 PostgreSQL 的头衔是“世界上最先进的开源关系型数据库”。PostgreSQL 在各行各业的各种场景下都有着广泛应用。但是会用只是“知其然”，知道背后的原理才能“知其所以然”。理解数据库原理及其具体实现，能让架构师以最小复杂度的代价实现所需的功能，让程序员以最小复杂度的代价写出更加高效可靠的代码，让 DBA 在遇到“疑难杂症”时拥有精准的直觉与深刻的洞察。

数据库是一个博大精深的领域，存储、I/O、计算，无所不包。PostgreSQL 可以视作关系型数据库实现的典范，用 100 万行不到的 C 代码实现了功能如此丰富的软件系统，非常凝练。它的每一个功能模块都值得用一本甚至几本书的篇幅去介绍。本书虽限于篇幅而无法一一深入所有细节，但它为读者进一步深入理解 PostgreSQL 提供了一幅全局的概念地图。读者完全可以顺着各个章节的线索，以点破面，深入挖掘源码背后的设计思路。

我们偶然发现了本书的英文版本，读完之后感觉受益匪浅。看到这么好的书没有中文译本，实在是遗憾，遂萌生了翻译的念头。译者不才，愿为 PostgreSQL 在中国的发展贡献一份力量，但鉴于水平有限，翻译如有疏漏，还望读者海涵。

作者序

中国的 PostgreSQL 用户们：

你们好！

本书详细解释了 PostgreSQL 的内部工作细节，目标读者为 DBA 与系统开发人员。理解数据库内部机制很有挑战，愿本书能在你们精通 PostgreSQL 的道路上有所助益。

本书能出中文版，我真的感到非常高兴，我这样认为是有原因的。

首先，这是我写的书中第一本被翻译的。当自己的书出版时，心情愉悦自不必说，而自己的作品能被翻译出版，更是一件非常令人激动的事情。

其次，我收到了来自世界各地的电子邮件，请求将这本书翻译成各种语言。实际上，至少有一半的邮件来自中国。许多邮件提及“这本书对中国 PostgreSQL 用户而言很有帮助”，因此，我很高兴终于能对他们的要求做出回应。

而最重要的原因与我的家族史有关。先父曾在中国的哈尔滨市住过几年，此后他回到日本，我出生了。他患心脏病很长时间，因此在我的印象中，他的心情总是不好，除了提到一件事时。

当我还是孩子的时候，每晚睡觉前父亲都会给我和妹妹讲他在中国的经历。在讲这些经历时，他总是显得非常高兴，经常说中国人民帮助了他。

他的故事给我留下了深刻的印象，当然也影响了我的人生。因为这些故事，我从小就对这个世界和未知的事物有强烈的好奇心，而我的好奇心也引领我走入软件工程领域。现在，我已经在中国出版了第一本书。简而言之，这是一个关于在日本家庭与中国之间传递爱的故事。

我很高兴终于可以回报中国人民了。当然，这本书只是一份很小的礼物，然而我想充满感激之情地将它献给您。

2017 年，我有幸访问中国，亲眼目睹了中国令人惊叹的发展。对我而言，这是一次非常难

忘的经历。我希望能再去中国，如果有机会，我还想在中国工作。

无论如何，让我们享受 PostgreSQL 吧！

铃木启修

2018 年 11 月

致谢

本书的出版离不开电子工业出版社工作人员与译者们共同的努力与付出，对此我深表感激。

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **提交勘误**：您对书中内容的修改意见可在 提交勘误 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 读者评论 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：http://www.broadview.com.cn/35709

· 译者简介 ·

冯若航

PostgreSQL DBA，全栈工程师。PostgreSQL与Golang布道者，活跃于PostgreSQL技术社区，致力于企业PostgreSQL培训与技术支持。

刘阳明

毕业于中国人民大学信息学院，对数据库内核有浓厚兴趣，有丰富的PostgreSQL管理与使用经验，坚信PostgreSQL是一个还未被大众了解的优秀数据库产品，坚持推动PostgreSQL的发展。

张文升

参与了《PostgreSQL实战》一书的写作，中国开源软件推进联盟PostgreSQL分会核心成员之一。常年活跃于PostgreSQL、MySQL、Redis等开源技术社区，坚持推动PostgreSQL在中国的发展，多次参与组织PostgreSQL全国用户大会。近年来致力于企业PostgreSQL培训与技术支持。

目录

第1章 数据库集簇、数据库和数据表

第1章和第2章会简单介绍一些PostgreSQL的基础知识，以帮助读者理解后续章节的内容。本章包括以下几个主题：

- 数据库集簇的逻辑结构
- 数据库集簇的物理结构
- 堆表文件的内部布局
- 从表中读写数据的方式

如果你已经熟悉这些内容，可以跳过本章。

> 译者注：本书中的database cluster与PostgreSQL中文文档统一，译为数据库集簇。与高可用数据库集群不同，这里的集簇表示多个逻辑的数据库在同一个数据库实例中。

1.1 数据库集簇的逻辑结构

数据库集簇（database cluster）是一组数据库（database）的集合，由一个PostgreSQL服务器管理。读者第一次听到这个定义也许会产生疑惑，其实PostgreSQL中的术语“数据库集簇”，并非意味着“一组数据库服务器”。一个PostgreSQL服务器只会在单机上运行并管理单个数据库集簇。

图1.1展示了一个数据库集簇的逻辑结构。数据库是数据库对象（database object）的集合。在关系型数据库理论中，数据库对象用于存储或引用数据的数据结构。（堆）表就是一个典型的

例子，还有更多对象，例如索引、序列、视图、函数等。在 PostgreSQL 中，数据库本身也是数据库对象，并在逻辑上彼此分离。所有其他的数据库对象（例如表、索引等）都归属于各自相应的数据库。

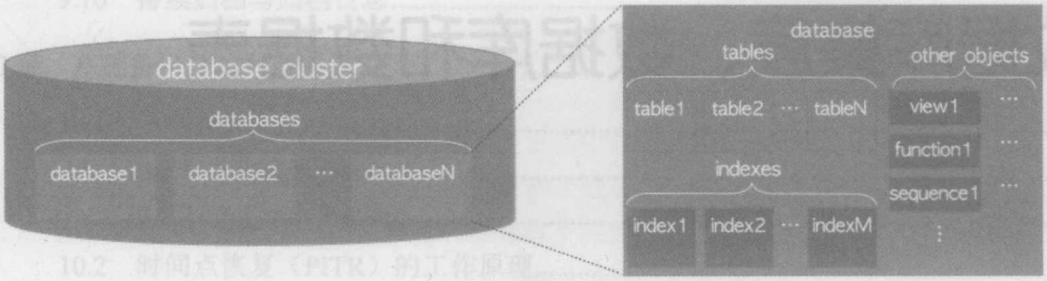

图 1.1 数据库集簇的逻辑结构

在 PostgreSQL 内部，所有的数据库对象都通过相应的对象标识符（object identifier，oid）进行管理，这些标识符是无符号的 4 字节整型。数据库对象与相应 oid 之间的关系存储在对应的系统目录中，依具体的对象类型而异。例如数据库和堆表对象的 oid 分别存储在 pg_database 和 pg_class 中，因此，当你希望找出 oid 时，可以执行以下查询：

```
sampledb=# SELECT datname, oid FROM pg_database WHERE datname = 'sampledb';
 datname  |  oid
----------+-------
 sampledb | 16384
(1 row)

sampledb=# SELECT relname, oid FROM pg_class WHERE relname = 'sampletbl';
  relname  |  oid
-----------+-------
 sampletbl | 18740
(1 row)
```

1.2 数据库集簇的物理结构

数据库集簇在本质上就是一个文件目录，即基础目录，包含着一系列子目录与文件。执行 initdb 命令会在指定目录下创建基础目录，从而初始化一个新的数据库集簇。通常基础目录的路径会被配置到环境变量 PGDATA 中，但这并不是必要的。

图 1.2 展示了一个 PostgreSQL 数据库集簇的例子。base 子目录中的每一个子目录都对应一