

数据库 查询优化器 的艺术

原理解析与SQL性能优化

李海翔 著

The Art of Database Query Optimizer
Principle and SQL Performance Optimization

- Oracle公司MySQL全球开发团队资深专家撰写，拥有10余年数据库内核研究经验，数据库领域泰斗王珊教授亲自作序推荐
- PostgreSQL中国社区和中国用户会发起人，以及来自Oracle、新浪、网易、华为等企业的数位资深数据库专家联袂推荐
- 从原理角度深度解读和展示数据库查询优化器的技术细节和全貌；从源码实现角度全方位深入分析MySQL和PostgreSQL两大主流开源数据库查询优化器的实现原理；从工程实践的角度对比了两大数据库的查询优化器的功能异同和实现异同



机械工业出版社
China Machine Press

014013169

TP311.138SQ

542

数据库 查询优化器 的艺术

原理解析与SQL性能优化

The Art of Database Query Optimizer
Principle and SQL Performance Optimization

李海翔 著



北航

C1700445

TP311.138SQ
542



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

数据库查询优化器的艺术：原理解析与 SQL 性能优化 / 李海翔著. —北京：机械工业出版社，2014.1
(数据库技术丛书)

ISBN 978-7-111-44746-7

I. 数… II. 李… III. 关系数据库系统—系统最优化 IV. TP311.138

中国版本图书馆 CIP 数据核字 (2013) 第 269607 号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书是数据库查询优化领域的里程碑之作，由 Oracle 公司 MySQL 全球开发团队、资深专家撰写，作者有 10 余年数据库内核和查询优化器研究经验。数据库领域泰斗王珊教授亲自作序推荐，PostgreSQL 中国社区和中国用户会发起人以及来自 Oracle、新浪、网易、华为等企业的数位资深数据库专家联袂推荐。从原理角度深度解读和展示数据库查询优化器的技术细节和全貌；从源码实现角度全方位深入分析 MySQL 和 PostgreSQL 两大主流开源数据库查询优化器的实现原理；从工程实践的角度对比了两大数据库的查询优化器的功能异同和实现异同。它是所有数据开发工程师、内核工程师、DBA 以及其他数据库相关工作人员值得反复研读的一本书。

全书一共 19 章，分为四个部分：第一篇（第 1～4 章）对数据库查询优化技术的范围、逻辑查询优化、物理查询优化，以及查询优化器与其他模块的关系做了非常细致、深入的讲解；第二篇（第 5～10 章）首先从源码角度对 PostgreSQL 查询优化器的架构、层次、设计思想、相关数据结构和实现原理进行了深入、系统的分析，然后从功能角度对 PostgreSQL 的逻辑查询优化、物理查询优化、查询优化器的关键算法，以及 PostgreSQL 查询优化器与其他模块的关系做了深入的讲解；第三篇（第 11～16 章）首先从源码角度对 MySQL 查询优化器的架构、层次、设计思想、相关数据结构和实现原理进行了深入、系统的分析，然后从功能角度对 MySQL 的逻辑查询优化、物理查询优化、查询优化器的关键算法，以及 MySQL 查询优化器与其他模块的关系做了深入的讲解；第四篇（第 17～19 章）对 PostgreSQL 与 MySQL 的逻辑查询优化技术、物理查询优化技术、设计思想和编码规范等各方面进行了深度的比较。

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：孙海亮

北京市荣盛彩色印刷有限公司印刷

2014 年 1 月第 1 版第 1 次印刷

186mm × 240mm • 33.25 印张

标准书号：ISBN 978-7-111-44746-7

定 价：89.00 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

投稿热线：(010) 88379604

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzsj@hzbook.com

推荐序一

随着数据库系统应用的广泛和深入，特别是近年来实际应用中数据库数据量不断增长，形成了所谓的海量数据，进一步的，现在人人都在讲的大数据时代已经到来。数据库系统的性能提升在传统数据库应用中同样受到现实的挑战。

对于一个选定的数据库管理系统（DBMS）产品，数据库系统性能的表现可以有很大差异，它和 DBMS 参数的选择、数据库模式的设计、应用系统的设计、软硬件环境的配置等多个因素密切相关。这就对数据库管理员（即 DBA）、应用系统分析和设计人员提出了要求，要求他们能够根据实际的应用环境、应用需求结合 DBMS 查询优化技术找到提升系统性能的方法或解决方案。

数据库查询优化技术一直是 DBMS 实现技术中的精华，也是难点和重点。数据库领域广大的研究工作者和实际开发者几十年来一直对查询优化技术孜孜不倦地探索着，更快、更好、更有效，这种精神使得查询优化技术不断推陈出新，形成了较为完善的优化技术体系，包括基于语法的、基于语义的、基于规则的、基于代价的等多角度、多方面的优化技术。

对于广大的工程第一线的技术人员、对于 DBMS 开发设计人员、对于数据管理领域的学生来说，仅了解查询优化技术的原理是不够的，还必须从工程实践的角度入手，把 DBMS 查询优化基本原理和实现代码结合起来，才能真正掌握 DBMS 查询优化技术，才能运用掌握的技术解决实际系统中性能提升的问题。

本书以 PostgreSQL 和 MySQL 两大开源 DBMS 查询优化器的实现为蓝本，介绍和分析了它们的实现原理、实现过程，帮助读者深入理解和掌握数据库查询优化的核心技术，在 DBMS 调优过程中不仅掌握怎么调优而且掌握为什么可以调优。

本书作者努力把数据库查询优化原理、查询优化引擎源码实现和 SQL 优化工程实践相结合，这是本书的特色，期望本书能成为读者的良师益友。

王珊

中国人民大学信息学院教授 / 博士生导师

教育部数据工程与知识工程重点实验室学术委员会副主任

推荐序二

PostgreSQL、MySQL、MongoDB、HBase 等都是非常优秀的开源数据库，大数据的到来，使得数据管理正走向一个新的复兴时期，PostgreSQL 在这次复兴中是一支重要力量。我一直在国内从事 PostgreSQL 社区的培育和推广工作，在这个过程中，我的一个比较深的体会是，目前国内非常缺乏与之相关的书籍和资料，这给 PostgreSQL 的推广和普及工作带来了非常大的阻力，所以社区非常希望有更多的出版社推动 PostgreSQL 相关书籍的出版。机械工业出版社华章公司非常具有前瞻性，对 PostgreSQL 非常关注和支持，他们策划的《PostgreSQL 数据库内核分析》一书在业界引起了很大的反响，我身边的朋友人手一本，为整个社区带来了非常大的帮助。海翔的这本书，是他们努力下的又一个丰硕成果。

数据库除了性能，还是性能，性能无疑是数据库工作者最为关注的重要方面之一。DBA 需要深入理解数据库优化的原则，只有这样才能有效调整和驾驭数据库的性能；数据库内核开发人员更需要理解其中的具体原理和实现，这样才能开发出高性能的产品。这本书独辟蹊径，从内核中选出查询优化这个最为复杂和引人入胜的部分为主题，同时也涵盖了两大开源主流数据库 PostgreSQL 和 MySQL 的内部实现，辅以理论和具体实现相结合的手法，使得读者可以在充分理解原理的基础上，深入掌握在具体工作中用到的技术，这不能不说是本书的一个精巧的地方。不仅是 PostgreSQL 的相关工作者，MySQL、Oracle 的数据库工作者，也可以通过本书，从原理和代码层面真正理解数据库性能优化，从而使自己成为一名性能优化“大师”。本书会给大家真正深入理解查询优化带来莫大的帮助。

李元佳

PostgreSQL 中国用户会发起人

前言

为什么写这本书

数据库引擎是一项包罗万象的技术大全，纷繁复杂，不能轻易穷尽。

数据库查询优化器是数据库引擎的重中之重。掌握了查询优化器，就等于掌握了数据库的精髓，得查询优化器者则得数据库天下。

在数据库查询优化技术领域，尚没有一本将理论与实践结合的书籍，**能帮助查询优化技术的爱好者掌握其全貌、掌握其细节、掌握其精髓、掌握其奥妙。**

如果能有一本书可以包括数据库查询优化技术的完整面貌，相信一定会使探索者少些崎岖摸索，这自然是一件好事。于是笔者心怀忐忑，虽自知力不能及，但还是决定奋笔一试，如工匠砌砖、蝼蚁啃土，一点一滴，将自己多年总结的经验汇集于此，终于小成。

本书的主要内容

本书主要讨论以下问题：

□“查询优化”究竟包括什么样的优化技术？为什么能做那些优化？

□“查询优化器”是怎么实现的？“查询优化器”是怎么做优化工作的？

□PostgreSQL 和 MySQL 的查询优化器各自实现了什么样的查询优化技术？

□PostgreSQL 和 MySQL 的查询优化功能有什么异同？

第一个问题在第一篇中进行了讨论。第二、第三个问题在第二、第三篇中通过对 PostgreSQL 和 MySQL 查询优化器的源码分析进行了回答。第四个问题在第四篇中通过对 PostgreSQL 和 MySQL 查询优化的功能对比，对两者的差异进行了回答。

从全书来看，本书涉及查询优化器的原理、代码实现、工程实践 3 个部分。

原理部分，概括总结了查询优化器的两大优化策略，一是逻辑优化，二是物理优化。在每一个优化策略中，都结合原理、运用示例，努力展示了查询优化器的全貌。这部分内容将解答：“查询

优化”究竟包括什么样的优化技术？为什么能做那些优化？

代码实现部分，通过深入数据库内核的源码，对查询优化器的实现进行深度的、全方位的探索，力求为程序员和数据库爱好者展示数据库引擎查询优化器的真实面目。行文选取了开源数据库 PostgreSQL 和 MySQL 的查询优化器，抽丝剥茧，进行了详细、全面的分析。这部分内容将解答：“查询优化器”是怎么实现的？“查询优化器”是怎么做优化工作的？

工程实践部分，通过原理分析、代码分析和示例，进行了以下 3 个对比：

- 对比 PostgreSQL 和 MySQL 的查询优化器的实现异同。
- 对比 PostgreSQL 和 MySQL 的查询优化器的功能异同。
- 对比查询优化器和原理之间的异同。

工程实践部分还通过大量详细的示例比较，总结了 PostgreSQL、MySQL 支持的优化点和不支持的优化点，为 DBA 调优提供了许多新颖的、好的思路。揭秘内核，帮助读者掌握其实现方式，可使读者在面对 SQL 调优时不再是雾里看花，而是心中有数。系统地掌握这些，可以超越依靠日积月累的学习方式，能够快速进阶。

工程实践部分通过 3 个对比，完成了以下 3 个结合：

- 通过代码分析，首次将关系代数、代价估算等查询优化原理与代码实现、工程实践结合。
- 通过对 PostgreSQL 和 MySQL 的对比，首次揭秘了两大开源数据库的查询优化器的异同，使得数据库爱好者能够以本书为基础，更好地结合查询优化技术，进而掌握各种类型的数据库 SQL 调优。
- 通过原理、代码、分析、示例，完成了将数据库爱好者对查询优化器强烈渴望掌握的愿望与轻松掌握的现实的结合，使得查询优化器不再那么神秘莫测。

本书的主要特色

本书的主要特色如下：

- **全面揭秘数据库查询优化技术。**把数据库的各种查询优化技术以工程化的视角进行剖析，集原理、源码分析、示例实践为一体。
- **重点分析了 PostgreSQL 和 MySQL 的查询优化器的实现。**以两大开源数据库的查询优化器的实现为蓝本，分析了它们的实现过程、实现原理，帮助读者深入理解掌握数据库的核心技术，使他们在 SQL 调优过程中不仅掌握怎么调优而且知道为什么可以调优。
- **详细对比了 PostgreSQL 和 MySQL 的查询优化技术的支持能力。**通过大量示例，分析对比了两大开源数据库的查询优化功能的强弱，帮助读者掌握两个数据库查询优化功能相同和不同之处，为应用系统选型提供衡量查询优化功能的依据。

本书面向的主要读者

如果您是一名数据库内核的工作者，本书对数据库查询优化器的内核代码分析，一定会帮您全面掌握查询优化这一技术。再结合理论部分，将使您理解得更加深刻。

本书不仅讲述了 PostgreSQL 查询优化器的实现，还讲述了 MySQL 查询优化器的实现，这对于从事这两个数据库内核开发的人员来说尤其有帮助。如果您能把结合本书和源码（务必要读代码）两者相互印证进行掌握，将像弄潮儿涛头立，查询优化技术尽在您的掌中。

本书从理论角度出发构造了大量示例，通过源码和功能的对比分析，更是进一步揭示了这两个数据库查询优化器的功能异同。对于从事查询优化器性能提升的工作人员帮助更大——因为原理部分讲了为什么，代码分析部分讲了怎么做，示例与对比部分讲了有什么不同，可做的事情只剩下“没有实现的优化怎么做”了。路已铺垫，启示已明，发挥就靠您自己了。

如果您是一名数据库实践者、DBA，本书一定会帮到您。查询优化器能做什么样的优化？为什么查询优化器能做某种类型的优化？对于这两个问题，原理部分能帮您解惑。如果您是一名 DBA，您将更多地从代码分析、工程实践部分获益。源码级的内幕揭秘，无疑能帮您更好地做好 SQL 语句的日常优化工作；代码、功能、实践的比较，无疑能帮助您从初学到理解，进而过渡到得心应手地应用。

生活不是一成不变的，工作中您也许会在 PostgreSQL 与 MySQL 间切换，掌握原理又能帮您以不变应万变。本书所涉及的查询优化器的知识，适用于 Oracle 的 DBA、DB2 的 DBA、SQL Server 的 DBA、SYBASE 的 DBA。也许从此以后，您会成为一名万能的 DBA！

如果您是一名数据库研究者、开发者、教学者、学生、爱好者，那么本书对您来说是一本不可多得的参考资料。查询优化器作为数据库中最重要、最难学、最精彩的部分，探索起来道路漫长，艰苦而无趣，也许本书能有幸成为您的同行伙伴，一路走来，不离不弃。探索数据库内核，犹如面对一座大山，陡峭艰险，无路可寻。看似非常艰难的事情，当您一书在手，仿佛为您提供了一柄手杖，步行于山涧沟壑时，且健且速。

特别是对于学生，如果能凭着这本薄书摸索进入数据库内核的世界，当是笔者最大的心愿。学生是未来的栋梁，是将技术发扬光大的希望。

多年来，作者一直翘首以盼，希望能有这么一类书，与己成为好友。现在，有了这么一本薄书，唯愿读者开卷有益。

如何阅读本书

从模块的角度看，本书可分为 4 部分：原理、PostgreSQL 实现、MySQL 实现、三者对比总结。

所以建议读者以模块为单位分别阅读。如果是读 PostgreSQL 实现、MySQL 实现的章节，则应结合代码，前后联系沟通为好，不要掌握局部忽视整体。

从内容的分布上看，原理是综述部分，既讲述了原理对查询优化的指导，也对 PostgreSQL 实现、MySQL 实现进行了总结，所以要从实践个例上升到理论的高度才能看清全貌。把原理和 PostgreSQL 实现结合起来，反复互为印证，会对掌握 PostgreSQL 有帮助；把原理和 MySQL 实现结合起来，反复互为印证，会对掌握 MySQL 有帮助；把 4 部分完全结合起来一起掌握，不仅可以帮助 DBA 和爱好者理解 PostgreSQL、MySQL，还可以帮助他们理解其他数据库。因为本书中的两套源码的比较加上原理解析，有助于帮助理解其他数据库查询优化器的原理；示例则有助于在其他数据库上执行以检查数据库对查询优化的支持程度。

另外，为便于读者更好地掌握数据库查询优化技术及本书内容，本书的附录从方法学的角度对阅读本书提供了 4 个方面的建议。

为什么特别推荐学生阅读本书

对于爱好数据库技术的学生来说，选择数据处理技术，是一条明智之路。数据越来越多，数据源越来越多，大数据的时代不仅已经来临了，而且会一直发展、不断发展。数据库技术是处理大数据的核心技术，数据库技术会随着时代的需求，一直前进。而查询优化，在数据库领域，不仅有着重要的理论地位，而且在实践中也异常重要且用途广泛，SQL 调优占据着数据库性能调优的半壁江山。

现今，就业竞争异常激烈，一份好的工作往往会有数以千计的人同时去争取。这对于诸多刚刚毕业、没有什么工作经验的学生来说，无疑是非常严峻的考验，尤其对于想找到理想工作的学生更是难上加难。因为现在用人单位提供的工作岗位对应聘者的素质和技能的要求越来越高，这无疑是刚刚毕业的学生最大的短板。

对于用人单位来说，想招到一名合格的员工也不容易，招到一名技能纯熟、素质高的员工更难。

在这样的大背景下，如果学生在求学期间能够打下良好的专业基础，找工作时定会增加很多机会；如果学生能够在自己的专业领域深入探索，对数据结构、操作系统、数据库等重要课程熟练掌握，使自己具备良好的专业素质，必然能使自己在找工作时具备更强的竞争力；如果学生能在某些课程上把理论和实践结合起来，把理论和源码结合起来，熟练掌握重量级软件的内核代码，找到心仪的工作将更为容易。用人单位喜欢有理论、工作初始就能上手做事的人才。而学生们想成为这样的人才，就需要在学校里打下坚实的基础。

本书在数据库领域就能起到这样的作用。引领学生把数据库理论和数据库实践结合起来，把数据库理论和数据库引擎源码结合起来，有效帮助读者梳理数据库技术中最重要的部分——查询优化技术的全貌，这必定对掌握这部分技术有很大的帮助。

所以，笔者特别推荐学生阅读本书。建议在本科高年级阶段或研究生阶段，大家能与本书为伴，边读源码（PostgreSQL 和 MySQL 的源码）边读本书，互为印证。笔者相信通过短短数月的努力，大家得到的将是足够好的技能、足够高的起点、足够强的竞争力！

学生是未来、是希望，学生强，社会亦将受益。学生若能通过本书使自己变强，将是笔者最大之幸。

勘误及支持

由于笔者的水平有限，书中难免会有笔误、差错、格式错误或遗漏等问题，希望广大读者能把发现的错误告诉笔者，我将不胜感谢。本书的进步和完善，有您的帮助和爱护，定能再上一层楼。您可以发送电子邮件到：database_XX@163.com。由于时间有限，也许笔者不能一一答复所有的电子邮件，但是会定期整理并汇总信息到笔者的博客（http://bolg.163.com/li_hx）上。

书中分析、探索的数据库源码的下载地址如下：

PostgreSQL, V9.2.3, 源自 <http://www.postgresql.org/ftp/source/v9.2.3/>。

MySQL, V5.6.10, 源自 <http://dev.mysql.com/downloads/mysql/>。

致谢

在我的生命中，家人是最重要的。伏案疾书不舍昼夜，心中一直牵挂的是父母、姐妹、妻子和各地的亲人们。他们是我写作的动力源泉，他们是我得以坚持完稿的坚强后盾。感谢父母给予了我生命并育我成人，感谢妻子朝夕相伴给予我鼓励和慰藉，感谢我远方的亲人们让我心意暖暖。

感谢中国人民大学信息学院王珊教授多年的教导，并高屋建瓴地为本书提出修订意见，她严谨的工作作风影响了本书的风格。成书之际，王老师乐为作序，我心怀感激，念念在心。

感谢为开源社区无私奉献的人们，没有他们就没有本书。本书的内容源于对开源数据库内核代码的分析，故笔者认为本书也应该回归社区为众人服务。坚持写下自己的所得，与他人共享，是一件快事，也是一件幸事。

感谢编辑杨福川先生和孙海亮先生为本书付出的努力和耗费的心血，书名源于杨先生，书稿样式由孙先生设定。写了一本书，交了两个好朋友。谢谢他们。

感谢每一位读者，我们一起进步，你们将是本书继续完善的新动力。

我深知本书尚不能达到书名标识的高度，而今迈步从头越。

目 录

推荐序一	
推荐序二	
前 言	

第一篇 查询优化技术

第 1 章 数据管理系统的查询优化 2

1.1 数据库调优	3
1.2 查询优化技术	5
1.2.1 查询重用	5
1.2.2 查询重写规则	6
1.2.3 查询算法优化	6
1.2.4 并行查询优化	8
1.2.5 分布式查询优化	9
1.2.6 其他优化	9
1.3 本章小结	9

第 2 章 逻辑查询优化 10

2.1 查询优化技术的理论基础	10
2.1.1 关系代数	11
2.1.2 关系代数等价变换规则 对优化的指导意义	13
2.2 查询重写规则	17
2.2.1 子查询的优化	18
2.2.2 视图重写	28

2.2.3 等价谓词重写	29
2.2.4 条件化简	32
2.2.5 外连接消除	33
2.2.6 嵌套连接消除	37
2.2.7 连接消除	38
2.2.8 语义优化	40
2.2.9 针对非 SPJ 的优化	41

2.3 启发式规则在逻辑优化阶段 的应用	42
2.4 本章小结	43

第 3 章 物理查询优化 44

3.1 查询代价估算	44
3.1.1 代价模型	44
3.1.2 选择率计算的常用方法	45
3.2 单表扫描算法	45
3.2.1 常用的单表扫描算法	45
3.2.2 单表扫描代价计算	47
3.3 索引	47
3.3.1 如何利用索引	47
3.3.2 索引列的位置对使用 索引的影响	50
3.3.3 联合索引对索引使用的 影响	56
3.3.4 多个索引对索引使用的 影响	57

3.4 两表连接算法	59
3.4.1 基本的两表连接算法	59
3.4.2 进一步认识两表连接算法	61
3.4.3 连接操作代价计算	61
3.5 多表连接算法	62
3.5.1 多表连接顺序	62
3.5.2 常用的多表连接算法	63
3.5.3 多表连接算法的比较	68
3.6 本章小结	68

第4章 查询优化器与其他模块的关系

4.1 查询优化器整体介绍	70
4.2 查询优化器与其他模块的关系	73
4.3 本章小结	74

第二篇 PostgreSQL 查询优化器原理解析

第5章 PostgreSQL 查询优化器概述

5.1 PostgreSQL 查询执行过程	76
5.2 PostgreSQL 查询优化器的架构和设计思想	78
5.2.1 PostgreSQL 查询优化器架构	79
5.2.2 PostgreSQL 查询优化器的层次	81
5.2.3 PostgreSQL 查询优化器设计思想	81
5.3 主要概念	81
5.4 代码层次结构	85

5.5 本章小结	86
----------------	----

第6章 PostgreSQL 查询优化器相关数据结构

6.1 主要数据结构	88
6.1.1 基本数据结构	88
6.1.2 查询树	91
6.1.3 各种对象的结构	95
6.1.4 连接操作相关的结构	99
6.1.5 查询执行计划相关的结构	104
6.2 各个结构之间的关系	108
6.3 各个阶段间和主要结构体间的关系	109
6.4 本章小结	110

第7章 PostgreSQL 查询优化器实现原理解析

7.1 查询优化整体流程	111
7.2 查询优化器实现原理解析	115
7.2.1 planner——主入口函数	115
7.2.2 standard_planner——标准的查询优化器函数	116
7.2.3 subquery_planner——生成(子)查询执行计划函数	117
7.2.4 grouping_planner——生成查询执行计划并对非 SPJ 优化	139
7.2.5 build_minmax_path——聚集函数 MIN/MAX 的优化函数	141

7.2.6	query_planner——生成 最优的查询路径函数	142	8.1.1	视图重写	193
7.2.7	make_one_rel——构造多 表连接路径并选出最优 路径函数	148	8.1.2	子查询优化	197
7.2.8	make_rel_from_joinlist—— 生成多表连接路径函数	153	8.1.3	等价谓词重写	209
7.2.9	optimize_minmax_ aggregates——聚集操作 MIN/MAX 优化函数	163	8.1.4	条件化简	209
7.2.10	create_plan——创建 查询执行计划函数	164	8.1.5	外连接消除	210
7.2.11	非 SPJ 处理——grouping_ planner 的各个子模块	166	8.1.6	嵌套连接消除	217
7.2.12	其他重要的函数与操作	170	8.1.7	连接的消除	218
7.3	代价估算实现原理解析	174	8.1.8	语义优化	221
7.3.1	查询代价估算	174	8.1.9	选择操作下推	226
7.3.2	单表扫描方式的代价 估算	174	8.1.10	非 SPJ 优化	226
7.3.3	两表连接的代价估算	178	8.2	优化器之物理查询优化	229
7.3.4	其他代价估算函数	184	8.2.1	PostgreSQL 的物理优化 主要完成的工作	229
7.3.5	选择率的计算	185	8.2.2	启发式规则在物理查询 优化阶段的使用	230
7.4	从目录结构和文件功能角度 看查询优化器	186	8.2.3	两表连接	230
7.4.1	查询优化子模块与主要 文件的关系	187	8.2.4	代价估算	230
7.4.2	查询优化器代码结构	187	8.2.5	PostgreSQL 的索引与 查询优化	231
7.5	本章小结	190	8.3	其他	237
第 8 章	从功能的角度看 PostgreSQL 查询优化	192	8.3.1	grouping_planner 函数 主干再分析	238
8.1	优化器之逻辑查询优化	192	8.3.2	用户指定的连接语义与 PostgreSQL 实现两表连接 的函数及算法的关系	240
8.1.1	视图重写	193	8.3.3	集合操作优化	242
8.1.2	子查询优化	197	8.4	本章小结	245
8.1.3	等价谓词重写	209	第 9 章	PostgreSQL 查询优化 的关键算法	246
8.1.4	条件化简	209	9.1	动态规划算法	246
8.1.5	外连接消除	210	9.1.1	动态规划算法的处理 流程	247
8.1.6	嵌套连接消除	217			
8.1.7	连接的消除	218			
8.1.8	语义优化	221			
8.1.9	选择操作下推	226			
8.1.10	非 SPJ 优化	226			
8.2	优化器之物理查询优化	229			
8.2.1	PostgreSQL 的物理优化 主要完成的工作	229			
8.2.2	启发式规则在物理查询 优化阶段的使用	230			
8.2.3	两表连接	230			
8.2.4	代价估算	230			
8.2.5	PostgreSQL 的索引与 查询优化	231			
8.3	其他	237			
8.3.1	grouping_planner 函数 主干再分析	238			
8.3.2	用户指定的连接语义与 PostgreSQL 实现两表连接 的函数及算法的关系	240			
8.3.3	集合操作优化	242			
8.4	本章小结	245			

9.1.2	紧密树处理流程	248
9.2	遗传算法	248
9.2.1	PostgreSQL 遗传算法的 处理流程	248
9.2.2	主要的数据结构	250
9.2.3	主要的函数和变量	251
9.2.4	应用遗传算法实现表 连接的语义	253
9.2.5	应用遗传算法计算适 应度	254
9.2.6	进一步理解 PostgreSQL 的遗传算法	255
9.3	动态规划算法与遗传算法对比	256
9.4	本章小结	257

第 10 章 PostgreSQL 查询优化器与其他部分的关系 259

10.1	查询优化器与语法分析器	259
10.2	查询优化器与执行器	260
10.3	查询优化器与缓冲区 管理模块	261
10.4	查询优化器与对象访问模块	262
10.5	查询优化器与统计模块	262
10.6	查询优化器与索引模块	263
10.7	本章小结	263

第三篇 MySQL 查询优化器 原理解析

第 11 章 MySQL 查询优化器 概述 266

11.1	MySQL 查询执行过程	266
------	--------------	-----

11.2	MySQL 查询优化器的架构 和设计思想	267
11.2.1	MySQL 查询优化器 架构	268
11.2.2	MySQL 查询优化器的 层次	269
11.2.3	MySQL 查询优化器 设计思想	269
11.3	主要概念	270
11.3.1	常量表	270
11.3.2	表数据的访问方式	270
11.4	代码层次结构	272
11.5	本章小结	274

第 12 章 MySQL 查询优化器 相关数据结构 275

12.1	主要的类和数据结构	275
12.1.1	查询树	275
12.1.2	基本对象	276
12.1.3	连接对象与执行计划	278
12.1.4	代价估算类	281
12.2	各个阶段主要结构体间的 关系	282
12.3	本章小结	282

第 13 章 MySQL 查询优化器的 原理解析 283

13.1	查询优化器整体流程	283
13.2	优化器的代码详解	285
13.2.1	JOIN.prepare——优化 前的准备工作	286

13.2.2	JOIN.optimize——优化 器主入口方法	299
13.2.3	make_join_statistics—— 计算最优的查询优化执 行计划	315
13.2.4	choose_table_order—— 求解多表连接最优连接 路径	324
13.2.5	make_join_statistics 函数的其他子函数	339
13.2.6	make_join_select—— 对条件求值、下推连 接条件到表中	348
13.2.7	test_if_skip_sort_order——排 序操作的优化	350
13.2.8	make_join_readinfo—— 为连接的每个表构造 信息	351
13.2.9	JOIN.exec——执行查询 执行计划的函数	353
13.3	代价估算	354
13.3.1	查询代价估算模型	354
13.3.2	查询代价估算过程	355
13.3.3	其他的代价估算	358
13.3.4	对存储引擎的调用接口	362
13.3.5	统计信息	364
13.4	本章小结	365

第 14 章 从功能的角度看

MySQL 查询优化

14.1	优化器之逻辑查询优化	366
14.1.1	视图重写	367

14.1.2	子查询优化	371
14.1.3	等价谓词重写	387
14.1.4	条件化简	388
14.1.5	外连接消除	389
14.1.6	嵌套连接消除	396
14.1.7	连接的消除	398
14.1.8	语义优化	400
14.1.9	非 SPJ 优化	406
14.2	优化器之物理查询优化	412
14.2.1	MySQL 的物理优化 主要完成的工作	412
14.2.2	启发式规则在物理查询 优化阶段的使用	413
14.2.3	MySQL 的索引与查询 优化	413
14.2.4	用户指定的连接语义与 MySQL 实现两表连接 的算法	417
14.3	本章小结	418

第 15 章 MySQL 查询优化的

关键算法

15.1	深入理解 MySQL 的多表 连接算法	419
15.2	本章小结	424

第 16 章 MySQL 查询优化器与

其他部分的关系

16.1	查询优化器与语法分析器	425
16.2	查询优化器与执行器	426
16.3	查询优化器与缓冲区管理 模块	426

16.4	查询优化器与索引模块	426
16.5	本章小结	427

第四篇 PostgreSQL 查询优化器 VS MySQL 查询优化器

第 17 章 PostgreSQL 和 MySQL 的逻辑查询优化技术 430

17.1	查询重写	430
17.1.1	子查询优化	430
17.1.2	视图重写	443
17.1.3	等价谓词重写	446
17.1.4	条件化简	447
17.1.5	外连接消除	448
17.1.6	嵌套连接消除	449
17.1.7	连接消除	451
17.1.8	语义优化	452
17.2	非 SPJ 的优化	452
17.3	本章小结	456

第 18 章 PostgreSQL 和 MySQL 的物理查询优化技术 457

18.1	查询代价估算模型比较	457
------	------------------	-----

18.2	单表扫描算法	458
18.3	索引	458
18.4	两表连接算法	466
18.5	多表连接算法	467
18.6	本章小结	467

第 19 章 PostgreSQL 和 MySQL 的其他异同 468

19.1	启发式规则的使用比较	468
19.2	综合比较	469
19.2.1	基本概念的比较	469
19.2.2	数据结构的比较	469
19.2.3	设计思想的比较	469
19.2.4	编码规范的比较	470
19.3	本章小结	471

附录 A 如何掌握数据库内核 472

附录 B 如何阅读本书 496

附录 C 如何阅读查询执行计划 498

附录 D 如何跟踪查询执行计划 508

第一篇

查询优化技术

本篇介绍了数据库的查询优化技术，从数据库的理论出发界定查询优化技术的范围，讨论了包括逻辑查询优化、物理查询优化两个方面的查询优化技术。全篇立足于数据库的基本理论——关系代数，在第1章首先界定了本书讨论的查询优化的技术范围；在第2章运用关系代数理论和关系法则，对逻辑查询优化进行全面而深入的探讨，对各种逻辑查询优化技术进行全面介绍，指出关系代数对于查询优化技术的指导意义，通过示例巩固对查询优化技术的理解和认识；在第3章，通过对代价估算模型、索引和表扫描、表连接算法的介绍，对各种物理查询优化技术进行全面描绘，构造了清晰、完整的物理查询优化技术图谱；在第4章，对实现查询优化技术的数据库查询优化器的相关模块进行了介绍，意在使读者了解查询优化技术的相关上下文内容。