第一图书网, tushu007.com

<<物理数据库设计>>

图书基本信息

书名：<<物理数据库设计>>

13位ISBN编号：9787302239314

10位ISBN编号：7302239312

出版时间：2010-12

出版时间：清华大学出版社

作者：（美）莱特斯通　等著

页数：283

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：http://www.tushu007.com

前言

IBM工程师E.F.Codd在1970年提出关系模型后，关系数据库成为管理和查询结构化数据的事实标准。
在过去的15年中，随着Internet的兴起，联机事务处理、在线银行业务、与不同系统连接的能力使数据量有大幅度的增长。
容量在TB级以上的数据已经很普遍。
这一数据增长的同时，遵守摩尔定律的CPU性能急剧增长，磁盘技术的进步使磁盘存储密度进一步提升。
现代数据库经常需要支持成千上万的用户同时访问。
数据库系统的性能和可维护性大大取决于其物理设计。
主要的数据库厂商开发了一系列很好的物理特性和功能设计技术。
当今的数据库可以在内存和磁盘上进行切片、切块、正移，还可以以各种方式旋转等。
但是到目前为止，并没有多少文章涉及“物理数据库设计”。
当然，也有过白皮书、一些个人观点和一些产品的文章涉及这个问题，但并没有把它作为一个整体来对待。
数据库设计人员在实际工作中遇到的困难也很少提及，“蠕变特性”的复杂性给行业带来了不小的麻烦。
因为这些原因，我们迫切需要关于物理数据库设计的资料。
我们编写的这本新书面向广大的用户，包括数据库系统的学生和数据库领域的专业人士。
在书中，我们引入物理数据库设计的主要概念，包括索引（B+、哈希、位图）、物化视图（延迟和实时）、范围分区、散列（哈希）分区、无共享设计、多维群集、服务器拓扑、数据分布和基础物理子系统（NUMA、SMP、MPP、SAN、NAS、RAID设备）等。
为符合编写这本书的目标人群——学生和数据库专业人员，我们把重点放在实际的问题和解决方案上。
在每一个细分市场和使用关系数据库系统的各个方面，有关物理数据库设计的关键问题似乎无处不在，如联机事务处理（OLTP）、数据挖掘（DM）、多维联机分析处理（MOLAP）、企业资源规划（ERP）、管理资源计划（MRP）、数据库管理员（DBA）团队设计和管理的所有内部企业系统以及开发过程中的独立软件提供商的应用程序（ISVA）。
我们希望对物理数据库设计、使用范例、产品的特定语法和最优方法的关注，能使本书成为数据库资料的一个很好的补充。
文章结构第1章是物理数据库设计的概述以及如何适应数据库的生命周期。
第2章介绍了业界目前用得较多的索引方法和B+树索引的基本原理，并介绍了简单索引和复合索引的变化，还对不同方法的性能测试加以比较。
第3章主要从数据库专家的角度介绍了查询优化的基础和查询执行计划选择等需要了解的数据库设计背景知识。
第4～8章讨论物理数据库设计所需要的个别重要的设计决策。
第4章介绍了如何完成索引选择的具体内容，以及在选择和连接操作中索引策略的替代方案。
第5章介绍了如何为独立关系数据库选择物化视图，以及如何在数据仓库中的数据库集合设置星型模型，用了几个例子说明物化视图选择中所涉及的折中方案。
第6章介绍了如何执行无共享分区来分步解决庞大而且需要复杂计算的数据库问题，并说明了无共享分区、物化视图复制以及索引之间的关系。
第7章专门讨论范围分区，将大表分成多个较小的表来保存一定范围内的数据和索引，方便寻址。
第8章讨论了群集数据的优点，以及在扩展到多维数据时这种技术是多么的有效。
这样可以使系统在同一时间内群集多维数据，而不用复制数据。
第9章讨论物理设计决策的整合问题。
物理设计决策由每个决策如何影响其他决策来决定，并指导设计人员使用这些组件来优化设计。
第10章的重点是计数和抽样数据，以协助改善独立查询设计的方法、物化视图的选择、群集和分区等

。
第11章通过对一些比较有用的工具的描述来讲述查询执行计划，这些工具允许用户了解查询执行计划，以及设计数据库时选择的方法，如索引选择和物化视图等。
第12章详细说明了自动化物理设计决策的重要性，主要的关系数据库如DB2、SQLServer和Oracle提供的自动化设计支持，讨论了如何使用这些工具更快速地设计高效率的数据库。
第13章提出数据库设计人员需要了解的一些系统问题，如多处理器服务器、磁盘系统、网络拓扑、灾难恢复技术和内存管理等。
第14章讨论了物理设计如何支持数据仓库和使用OLAP技术进行有效的信息检索。
第15章定义了什么是逆规范化，并且说明规范化程度和数据库性能之间的权衡。
第16章是分布式数据之间分配策略的基础知识，包括由于数据复制影响的快速查询响应时间和多个数据副本更新时间成本的权衡。
附录A简要介绍了一个简单的运算性能模型，用来评估和比较单一数据库的不同物理设计的策略。
该模型阐明了几个章节中所讲述的物理设计方法——权衡分析和设计决策。
附录B包含了两个商业化灾难恢复技术的比较，分别为IBM的高可用性灾难恢复和Oracle的Data Guard。
每一章都有数据库设计人员的提示和独特见解，对读者了解每一章的设计方法很有帮助。
接下来是参考书目汇总，使读者能够选择相应的物理设计主题进行进一步的了解。
使用范例逻辑设计和物理设计之间的主要区别之一，是物理设计的基本特点和数据库服务器的物理属性（软件和硬件）的关系越来越紧密。
虽然逻辑设计可以抽象地进行，有些产品和组件相对独立，可以用来实现所做的设计，但也不能说是物理设计。
因此，在本书中，我们仔细选择了一些例子，包括关于物理数据库设计的数据库服务器产品中的部分主要产品，如DB2forzOSV8.1、DB29（Linux、UNIX和Windows）、Oracle10g、SQLServer2005、InformixDataserver和NCRTeradata的产品。
我们认为，这包括了行业内流行的大部分数据库。
一些比较流行的数据库如MySQL和Sybase则没有加以介绍，只是为了更简单地说明问题。
文献综述和书目与之前的逻辑数据库设计书籍《数据库建模与设计：逻辑设计》（《DatabaseModelingandDesign:LogicalDesign,4th》）一样，本书在每章的最后都有文献综述。
章节当中所涉及的材料有论文和相关参考，具体有两种形式：与本章讨论的物理数据库设计概念有关、原创的具有突破性想法、对数据库设计发展有重大影响的论文。
关于最新研究和突破性想法的主要论文。
除每一章后面的文献综述外，本书的最后有个更全面的参考书目表。
反馈与勘误表如果您有任何意见或建议，我们希望能够收到您的来信，这对我们得到反馈和改进或者更正这本书的内容十分有帮助。
不知道有人会注意到database这几个字母都是用左手打出来的吗？
现在标准键盘的布局设计是为了使两只手的负担能够均等。
因此，按照这样设计数据库不仅仅是不合理的，并且比它看起来要困难得多。
——佚名虽然这是一句俏皮话，但您可以想象作者写这本书时，要用左手打上百次的database，如果您使用“设计数据库（designing databases）”来代替“编写数据库（writing about databases）”，那么对来自全球性社区的数据库设计人才会更有说服力。
致谢如同做其他事一样，作者周围会有许多人帮忙审阅、编辑、出版，最后才会出版这样一本书。
我们在这里特别感谢公司里的同事和为这本书做出贡献的合作伙伴，他们是Sanjay Agarwal、Eric Alton、Hermann Baer、Kevin Beck、Surajit Chaudhuri、Kitman Cheung、Leslie Cranston、Yuri Deigin、Chris Eaton、Scott Fadden、Lee Goddard、Peter Haas、Scott Hayes、Lilian Hobbs、John Hornibrook、Martin Hubel、John Kennedy、Eileen Lin、Guy Lohman、Wenbin Ma、Roman Melnyk、Mughees Minhas、Vivek Narasayya、Jack Raitto、Haider Rizvi、Peter Shum、Danny Zilio and Calisto Zuzarte。
谢谢Linda Peterson和Rebekah Smith对手稿所做的帮助。

我们还要感谢为本书提供了极有价值见解的校对人员，他们深入阅读并给出了新的方向，使我们能更好地完成本书。
感谢Mike Blaha、Philippe Bonnet、Philipe Carino和Patrick O ’ Neil。
谢谢概念评审Bob Muller、Dorian Pyle、James Bean、Jim Gray和Michael Blaha。
我们要感谢我们的妻子和孩子们，有了他们的支持，我们才有时间来做这个项目。
我们经常工作到凌晨。
向全球社区的学生和数据库设计人员致敬。
你们的工作比一般人所了解的更为艰巨和复杂。
现代关系数据库系统每种可能的设计属性都有其自身的复杂性，要解决这些问题，作为真正的数据库设计人员来说是个很大的挑战，大家都知道对人类来说这比较难。
事实上，对任何有一定复杂关系的系统，在数学上真正完全表达数据库优化设计是不可能的。
在一次分析中我们发现，一个普通的数据库可能的设计选择数目远远超过了宇宙中目前估计的原子数量（1081次方）几个数量级！
然而，尽管现代数据库系统有很大的复杂性，但是也要设法加以研究、掌握并继续进行设计。
毫不夸张地说，世界上的数据就在大家的手中。
我们希望这本书成为您的宝贵工具。
我们希望这本书能够引领学生和数据库设计人员提高，这是改进世界数据管理基础结构的重要方法。
工程师行业是一个伟大的职业。
他们将主观想象出来的事务通过科学手段在理论上进行实现，然后转化为对石头、金属或能源的认识。
随后，它给人们带来了家园，提高了人们的生活水平，并增加了生活的舒适性。
这是工程师的最高特权。
——赫伯特?胡佛（Herbert Hoover 1874—1964） 大部分专家都同意，摧毁世界最有可能的方式是事故。
这正是我们的用武之地，我们是计算机专家，能引发事故。
——纳森尼尔?伯伦斯坦（Nathaniel Borenstein 1957）

内容概要

本书全面讲述数据库物理设计方案，主要包括物理数据库设计概况，基本索引方法，查询优化和方案选择，选择索引，物化视图选择，无共享分区，范围分区，多维群集，相互依赖的问题，物理设计探索中的计数和数据抽样，查询执行计划和物理设计，自动化物理数据库设计，服务器资源和拓扑，决策支持、数据仓库和OLAP系统中的物理设计，逆规范化以及分布式数据分配等内容。

本书适合作为高等院校计算机及相关专业的教材和教学参考书，也可作为相关开发人员的自学教材和参考手册。

作者简介

作者：（美国）莱特斯通（Sam Lightstone） （美国）特里（Toby Teorey） （美国）纳多（Tom Nadeau） 译者：吴骅 王学昌 韩潼瑜Sam Lightstone，是IBM的DB2产品开发团队研发经理及高级技术人员，他的工作涉及 自主计算和关系数据库管理系统的许多方面。
Sam Lightstone是DB2自主计算研发工作的 领导者和创始人之一，是自治数据库系统方面IEEE数据工程组的主席，是自治和自主计 算方面IEEE计算机协会任务组的成员。
2003年，他当选为IBM技术研究院加拿大子公司 的杰出技术委员会委员。
Sam Lightstone是IBM的发明大师，拥有超过25项专利和未决专 利，他所发表的成果涉及许多关于关系数据库系统自主计算的内容，从1991年开始一直就 职于IBM。
Toby Teorey是密歇根大学安娜堡分校电气工程与计算机科学系名誉教授以及工程学 院学术项目主任。他先后获得亚利桑那大学图森分校电气工程的理学学士、理学硕士学位 以及威斯康星大学麦迪逊分校的计算机科学博士学位，也一直是各种数据库会议委员会的 项目委员会主席和成员。
Tom Nadeau是阿拉丁软件(aladdinsoftware．com)的创始人，研究方向是数据及文本 挖掘领域。
他先后获得密歇根大学安娜堡分校计算机科学理学学士、电气工程和计算机科 学硕士、博士学位，其研究的技术领域包括数据仓库、OLAP、数据挖掘和机器学习，并于 2001年获得IBM CASCON会议的最佳论文奖。

<<物理数据库设计>>

书籍目录

第1章　物理数据库设计概况　1.1　动机——数据增长和与日俱增的物理数据库设计　1.2　数据库生命周期　1.3　物理设计的元素：索引、分区和群集　　1.3.1　索引　　1.3.2　物化视图　　1.3.3　分区和多维群集　　1.3.4　物理数据库设计的其他方法　1.4　物理设计为什么这么困难　1.5　文献综述第2章　基本索引方法第3章　查询优化和方案选择第4章　选择索引第5章　物化视图选择第6章　无共享分区第7章　范围分区第8章　多维群集第9章　相互依赖的问题第10章　物理设计探索中的计数和数据抽样第11章　查询执行计划和物理设计第12章　自动化物理数据库设计第13章　实质探讨：服务器资源和拓扑第14章　决策支持、数据仓库和OLAP系统中的物理设计第15章　逆规范化第16章　分布式数据分配附录A　一个简单的性能模型数据库附录B　Oracle　Data　Guard与DB2　HADR的数据库灾难恢复技术比较参考书目

章节摘录

第1章 物理数据库设计概况我没有丢弃我的想法，它已经备份在了某一张磁盘上。
——佚名在1974年密歇根州安阿伯市，美国计算机协会文件说明与翻译组（ACM SIGFIDE，现更名为：国际数据管理协会，SIGMOD）举办的年会及其随后的数据系统语言会议报告中，关系数据模型之父特德考得与网络数据模型思想提出者查理贝克汉姆发生了激烈辩论。
到底什么样的逻辑模型才是最好的数据库逻辑模型？
将近30年各种学术期刊和商业杂志都充斥着这种辩论，直到2003年特德考得去世后，这种辩论才逐渐平息下来。
尽管最终关系数据库模型占据了上风，但由于最初的争论，许多数据库系统开始建立并支持这两种数据库模型，各种系统所使用的基础物理数据库结构也随之发展起来了。
最初物理设计主要是系统能够采用什么类型的索引，B+树索引几乎占据了所有系统场景。
后来，类似群集和分区的其他概念开始变得重要，但在20世纪70年代的争论之后，这些方法与逻辑结构的关系越来越少。
逻辑数据库设计，更确切地说是对于一个特定数据库系统的逻辑设计，大多数应用程序设计人员和程序员已经掌握了基本数据关系及概念定义。
设计人员的逻辑设计工作可以利用并借助ERWin数据模型工具、Rational Rose UML建模或使用纯手工方法实现。
物理数据库设计就是在您正在使用的计算机平台上，实现高效的数据存储、检索机制，通常是数据库管理员的工作，今天有一些厂家提供的工具可以帮助数据库管理员设计一个非常高效的数据库。
本书将主要致力于关系数据库的物理设计方法和流行工具介绍，书中例子也将采取常见的数据库系统——Oracle、DB2（IBM）和SQL Server（Microsoft）来描述这些物理数据库的设计概念。
1.1 动机——数据增长和与日俱增的物理数据库设计物理数据库设计真的有那么重要吗？
答案是肯定的。
一些计算机专家除了运行自己正在运行的商业外，很少去做其他的事情，更不要说帮助客户提升他们的表索引设计性能。
这也同样令人印象深刻，问题查询提升了50倍的效率。
真正推动物理数据库设计的是数据卷，毕竟，一个数据库只有几行数据，对于数据库物理设计并不是什么问题，执行一个应用程序存储一个小数据库不可能影响到基础的数据库物理设计。
在特定的领域内，实际的索引选择对一个有20多行数据的数据库并不是很困难。
然而，随着数据卷的增长，位于数据存储模式下的物理结构就显得非常重要了。
物理数据库设计实际上是被数据卷引发的。
毕竟，一个只有几行数据的数据库对于物理数据库设计实际上是没有任何争论的，访问小型数据的应用程序性能不可能被基础系统的物理设计深入影响。
从实践的观点来看，索引选择对于只有20行记录的数据库实际上并不重要。
然而，随着数据卷的增长，位于数据库访问模式之下的物理结构日渐重要。
一些因素正在激发数据按照3种形式呈戏剧化增长，这3种形式是结构化（如关系元组）、半结构化（如XML）和非结构化数据（如音频、视频等）。
大多数数据增长可以归结于快速扩展和无处不在的网络计算机和终端，几乎存在于每个家庭、商业机构之中，存储在工业世界中。
此外，随着个人快速使用移动电话和PDA等可以连接网络并用于共享数据的工具，数据卷开始进一步跳跃式发展。
在企业系统中，以10T测算的数据库也开始变得很普通。
随着人类基因组的3亿个碱基对的测绘工作，制药公司正在研究覆盖人类基因的以蛋白质为基础的基因工程网络，这将会在数据库中产生几拍（它）字节PB（1PB=1000T或1PB=1000000G）大小的分析数据。
表1.1数据显示的是伯克加州大学1999年展开的一项调查。

从这个调查研究报告中，可以看到存储在磁盘上的部门和企业数据正在以每年100%的速度快速增长。
事实上，没有人能够知道数据增长模式将在哪里停止或数据增长将何时停止。
表1.1 全球性产品的原始内容、存储数据（以TB为单位）存 储 介 质 类 型 TB年（高） TB年（低） 增长率%纸质 书 8 1 2 报纸 25 2 -2 期刊杂志 12 1 2 官方文件 195 19 2 小计 240 23 2胶片 照片 410000 41000 5 影视 16 16 3 X光 17200 17200 2 小计 427216 58216 4光学介质 音乐CD 58 6 3 数据CD 3 3 2 DVD 22 22 100 小计 83 31 70磁性介质 摄像机磁带 300000 300000 5 PC机磁盘驱动器 766000 7660 100 部门级服务器 460000 161000 100 企业级服务器 167000 108550 100 小计 1693000 577210 55总计 2120539 635480 50资料来源：美国加州大学伯克利分校的研究，1999。
另外，还有一些特殊事情的发生带动了数据量。
这些事情悄无声息地发生，没有任何人干扰并提起它们，但变化还是可以定量和令人深思的。
在2000年左右，存储的价格下降了1个点，在计算机磁盘上存储数据开始变得比在纸张上存储数据便宜（如图1.1所示）。
事实上，这大概是西方文明发展史上一次伟大的转折点。
羊皮纸、纸莎草和文件已经记录了过去的2000多年文明，突然间，这种文字记录模式开始出现衰败，现在电子文件不仅可以用来满足分享和分析的乐趣，而且也更为经济。
图1.1 存储价格（来源：IBM的研究）剧烈增长的数据模式已经改变了应用系统必须访问和处理的关系数据库的数据量大小，但是快速增长的数据模式并没有改变完成操作的速度。
事实上，在很大程度上，计算机定义的数据处理的执行目标要比人类定义的目标更加合理，例如，一个人愿意等待完成事务的时间可以是等待自动提款机的时间，也可以是每天晚上收盘和每天早晨开盘之间的几个小时非高峰可用时间。
这些限制因素主要取决于人类的期望，并且不依赖于所操作的数据卷。
虽然数据卷和分析的复杂度在快速增长，作为人类的期望变化速度却慢得多。
一些缓解手段已经被发现，因为尽管数据卷在不断增长，但随着现代数据服务器的计算能力的提升，数据卷的计算能力也在增长。
然而，固化服务器技术减少IT支出、增加数据处理能力的现象正在减轻，正是因为这个原因，随着服务器处理能力的增加，数据服务器往往需要承担更多的用途，而不是为了让单一的任务执行得更快。
尽管按照摩尔定律，20世纪70年代中期以来，CPU的处理能力每18个月增长一倍，但磁盘读写速度也在按照一个比较适中的速度快速增加（请参见第13章关于摩尔定律的更加深入的讨论）。
最终，数据正在越来越多地用于检测信息，而不仅仅是处理数据，与其同时，正在崛起的在线分析处理（OLAP）、数据挖掘和其他形式的商业智能计算，造成实际处理过程复杂性的剧增。
这些因素诱发了物理数据库设计复杂而精密的方法。
为什么呢？
通过利用设计技巧，设计人员能够在某些场景中减少几个数量级的业务处理时间。
当您等候在银行取款机准备提取您的现金，或等一个投资数百万美元的投资决策商业贸易分析结果时，提升计算效率上千倍是真实且有价值的，也是非常有必要的。
1.2 数据库生命周期数据库设计生命周期的基本步骤包括利用数据库管理系统（DBMS）的具体表定义方法从用户需求的概念模型设计概念模型和对物理数据库的索引、分区、群集和具体选择性进行实际性能调优。
对于分布式数据库，物理数据库设计还涉及分配计算机网络数据。
一旦设计完成，数据库生命周期将在实现和维护中继续延续。
数据库生命周期如图1.2所示。
物理数据库设计被定义在整个数据库生命周期的上下文中，用于显示物理数据库设计关系到其他设计步骤。

媒体关注与评论

我强烈推荐由Lightstone、Teorey以及Nadeau所撰写的《物理数据库设计》一书。
本书涵盖了诸多有关良好物理设计的内容：如在使用索引、物化视图的权衡以及物理数据布局等不同方法时的效果。
与其他类似书籍所不同的是，本书并不是局限在某种特定的产品之上，而是那些被产品表面所遮蔽的更深层次的内容。
本书还涉及到交易密集型应用程序(OLTP)以及数据仓库(OLAP)方面的内容。
总之，本书十分值得期待。
——Michael Blaha，OMT Associates，Inc。
这是一部物理数据库设计方面的优秀著作，她给予读者务实的模型以及建议。
为广大的从业者以及学习者提供了丰富的信息，并通过DB2、Oracle以及SQL Server中的实例介绍相关分析模型以及实用的提示信息。
——Jim Gray，Microsoft Research

编辑推荐

《物理数据库设计:索引、视图和存储技术》这是第一本全面讲述数据库物理设计的书籍。
关系数据库中所包含信息的迅速增加，使得数据库、性能和可维护性变得不再简单，DBA比以往任何时候都要在更大压力下去优化数据库结构的系统性能和管理。
《物理数据库设计——索引、视图和存储技术》一书讨论数据库的物理结构如何影响性能，包括具体的例子、向导、各种DBMS和配置的最好和最差的应用实例；一些简单的内容，如提高表的索引设计对性能有深远的影响；不同形式的数据库，如联机事务处理（OLTP）、企业资源管理（ERP）、数据挖掘（DM）和管理资源计划（MRP）等，这些数据库都可以使用这本书中提供的方法改善性能。

第一图书网, tushu007.com

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:http://www.tushu007.com