第一图书网, tushu007.com

<<海量数据库解决方案>>

图书基本信息

书名：<<海量数据库解决方案>>

13位ISBN编号：9787121118838

10位ISBN编号：7121118831

出版时间：2010-10

出版时间：电子工业出版社

作者：李华植

页数：440

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

这已经是第四次为本书写作者序言了，此时此刻过去20年的生活如同电影般在我的脑海里一一掠过。
当我最初决定步入IT领域时就为自己立下了誓言，时至今日回想起多年走过的历程，其间充满了艰辛，也正是这无数的艰辛让我最终体验了收获的愉悦。
回望这20多年的足迹，我一直努力用新的视角去观察他人所忽视的领域，尝试用崭新的思维和充满创意的双手去耕耘。
尽管如此，也仍然无法紧跟IT技术飞快的发展步伐，我为实现理想而终日不停前行的脚步，虽然忙碌但却无限满足。
众所周知，能够加工成宝石的原石比比皆是，一分耕耘，一分收获，每当我们初次接触某个新的东西时都会或多或少有些紧张。
因此从这一层面来看，数据库散发着无穷的魅力，它如同渊博精深的智者般质朴，总是以真实、坦诚的心去面对每一位学习和研究它的人。
在过去并不短暂的岁月里我一直深信数据库的骨骼就是“数据”，并为这一理论的发展不断努力，吸收同仁们分享的经验而持续奋斗。
为了打破始终在理论表面徘徊的固有模式而不断寻求新的尝试，并试图探求能够让IT工作者在实际工作中轻松应用并掌控的巧妙方法。
这种巧妙方法不能是只通过经验和试验才能获得的，它必须是利用日常常识就可以理解说明的方法。
有这么一句话“会者不难，难者不会”，如果能够把一些复杂的理论与通俗浅显的常识相结合，那么不仅有利于人们的理解，更有利于人们在合适的情况下加以灵活运用。
相反，有这么一句话“一知半解以为是”，意思是指那些只观其表不观其里就加以相信的人。
很多程序员只忠实地相信自己的经验，当问及为何如此时，大部分人的答案都是“因为我那样做过”或“那样比较好”。
10种类型的原理可以组合出10的阶乘（3 628 800）种现象，那么100种类型的原理所能够表现出来的现象数可以认为是一个天文数字。
如若仅凭经验去思考问题，无论怎么努力，最终也只能获得其中一部分的原理而已。
然而，事实上我们是完全有能力深刻地理解这100种原理的。
但如果不试图进行深刻钻研而只停留在表面，最终只能是一无所获。
宝石是不会被轻易发现的，只有凭借最大的努力去寻找方能找到。
在不知不觉中当我们遇到了从表面上看无法解决的复杂问题时，会出现两种人：其一，是坚持不懈、彻夜不休也要寻找到最佳解决办法的人，这种人通过不懈的努力最终能够获得什么呢？
事实上随着岁月的流逝，他们终将成为众人皆知的专家；其二，是认为过于烦琐，直接予以放弃的人，这种人只会让自己的血汗变成廉价的废弃物。
可以自豪地说“我付出了常人所无法想象的艰辛”，为了寻求完美的真理舍弃了很多常人的生活。
在没有钓到鱼时钓鱼人也许会为此而耿耿于怀，但在我看来问题的关键在于没有寻找到有效的钓鱼方法。
如果钓鱼人能够充分理解我的想法，并甘愿为了改变自己的固有观念而付出较大努力，尽管他也可能会为此而花费大量的时间和心血，但坚信他一定能够获得别人所无法获得的成果。
如果他研究出了别人所无法研究出的钓鱼方法，那么从此就再也不用为钓不到鱼而担心了。
各位读者在工作的同时究竟是否一直在使用一种平凡的方法呢?还是为了解决明天必须要完成的任务而临时抱佛脚呢？
现在该到结束这种恶性循环的时候了。
应用程序其实就是处理数据的手段而已，它需要紧跟流行的步伐，如不及时进行更新，在不经意之间就已经落伍了。
然而数据和数据库并非如此，不论岁月如何流失，我们积攒起来的“内功”是不会消失的。
如果能对其原理有一个深刻的理解，那么不论何时何地都能够随心所欲地钓到很多鱼。
随着数据库技术的发展进步，能够精确执行指令的DBMS与日俱增，随着对DBMS应用能力的不同所获

得的性能差异使我们从技术中获得满足感。

内容概要

本书将整体内容分为两部分，在第1部分中以影响数据读取效率的所有要素为类别，对其各自的概念、原理、 特征、应用准则，以及表的结构特征、多样化的索引类型、优化器的内部作用、优化器为各种结果制定的执行计划予以详细说明，并以对优化器的正确理解为基础，提出对执行计划和执行速度产生最大影响的索引构建战略方案；在第2部分中主要介绍提高数据读取效率的具体战略方案，在这部分中介绍与数据读取效率相关的局部范围扫描的原理和具体应用方法，以及对被认为是提高数据库使用效率基础的表连接的所有类型予以详细说明。

《海量数据库解决方案》系列丛书深受广大读者的喜爱已经长达10年之久，在被誉为“ 圣经 ”的同时，它已经变成了数据库用户不可或缺的必读书籍。

作者竭力探求能够让IT工作者在实际工作中轻松应用并掌控的巧妙方法，提供事半功倍的海量数据库解决之道。

第一图书网, tushu007.com

作者简介

李华植 代表韩国的数据库技术先驱
集基于EA的数据架构方法论之大成
在韩国最早提出了数据专家顾问的概念
现任EN-CORE CONSULTING总经理及代表顾问
曾任韩国Oracle公司担任200多家企业的技术顾问

书籍目录

第1部分 影响数据读取的因素第1章 数据的存储结构和特征 1 1.1 表和索引分离型 5 1.1.1 堆表的结构 5 1.1.2 聚簇因子（Cluster Factor） 10 1.1.3 影响读取的因素 13 1.1.3.1 大范围数据读取的处理方案 14 1.1.3.2 提高聚簇因子的手段 17 1.2 索引组织表（Index-Organized Table） 19 1.2.1 堆表和索引组织表的比较 19 1.2.2 索引组织表的结构和特征 20 1.2.3 逻辑ROWID和物理猜（Physical Guess） 22 1.2.4 溢出区（Overflow Area） 24 1.2.5 索引组织表的创建 25 1.3 聚簇表 26 1.3.1 聚簇表的概念 27 1.3.2 单表聚簇 29 1.3.3 复合表聚簇 31 1.3.4 聚簇表的代价 34 1.3.5 哈希聚簇 39第2章 索引的类型和特征 43 2.1 B-Tree 索引 44 2.1.1 B-Tree 索引的结构 44 2.1.2 B-Tree 索引的应用 47 2.1.3 反向键索引 52 2.2 位图索引 53 2.2.1 位图索引的形成背景 54 2.2.2 位图索引的结构和特征 55 2.2.3 位图索引的读取 57 2.3 基于自定义的函数索引 60 2.3.1 基于自定义的函数索引的概念和结构 60 2.3.2 基于自定义函数索引的约束 61 2.3.3 基于自定义函数索引的灵活运用 64第3章 SQL的执行计划（Explain Plan） 74 3.1 SQL和优化器 75 3.1.1 优化器的作用和人的作用 77 3.1.2 优化器的类型 80 3.1.2.1 基于规则的优化器 82 3.1.2.2 基于成本的优化器 86 3.1.2.3 优化器目标的选择 93 3.1.2.4 执行计划的固定化方案 97 3.1.2.5 优化器的局限 103 3.1.3 优化器的最优化步骤 106 3.1.4 查询语句的转换 112 3.1.4.1 传递性规则 113 3.1.4.2 视图合并（View Merging） 116 3.1.4.3 查看用户定义的绑定变量 122 3.1.5 开发者的作用 123 3.2 执行计划的类型 126 3.2.1 扫描的基本类型 126 3.2.1.1 全表扫描 127 3.2.1.2 ROWID扫描 132 3.2.1.3 索引扫描 133 3.2.1.4 B-Tree聚簇读取（Cluster Access） 138 3.2.1.5 哈希聚簇读取（Hash Cluster Access） 139 3.2.1.6 采样表扫描（Sample Table Scan） 140 3.2.2 表连接的执行计划 143 3.2.2.1 嵌套循环连接（Nested Loops Join） 143 3.2.2.2 排序合并连接（Sort Merge Join） 146 3.2.2.3 哈希连接（Hash Join） 148 3.2.2.4 半连接（Semi Join） 149 3.2.2.5 笛卡儿连接 151 3.2.2.6 外连接（Outer Join） 154 3.2.2.7 索引连接 159 3.2.3 其他运算方式的执行计划 161 3.2.3.1 IN-List迭代执行计划 162 3.2.3.2 连锁执行计划 163 3.2.3.3 远程执行计划 165 3.2.3.4 排序操作执行计划 168 3.2.3.5 集合操作执行计划 171 3.2.3.6 COUNT(STOPKEY)执行计划 174 3.2.4 位图（Bitmap）执行计划 175 3.2.4.1 各种条件运算符的位图执行计划 176 3.2.4.2 子查询执行计划 182 3.2.4.3 与B-Tree索引相结合的执行计划 184 3.2.5 其他特殊处理的执行计划 185 3.2.5.1 递归展开（Recursive Implosion）执行计划 186 3.2.5.2 修改子查询执行计划 191 3.2.5.3 特殊类型的执行计划 193 3.3 执行计划的控制 203 3.3.1 提示的活用准则 204 3.3.2 使用提示实现最优化目标 206 3.3.3 使用提示改变表连接顺序 207 3.3.4 表连接方式选择过程中提示的使用 208 3.3.5 并行操作中提示的使用 209 3.3.6 数据读取方法选择中提示的使用 211 3.3.7 查询转换（Query Transformation）过程中提示的使用 214 3.3.8 其他提示 216第4章 构建索引的战略方案 221 4.1 索引的选定准则 222 4.1.1 不同类型表的索引应用准则 223 4.1.2 离散度和损益分界点 227 4.1.3 索引合并和组合索引的比较 229 4.1.4 组合索引的特征 232 4.1.5 组合索引中列序的决定准则 239 4.1.6 索引选定步骤 242 4.2 决定聚簇类型的准则 263 4.2.1 全局性聚簇 263 4.2.2 局部性聚簇 265 4.2.3 单表聚簇 266 4.2.4 单位聚簇大小的决定 267 4.2.5 确保聚簇被使用的措施 270第2部分 最优化数据读取方案第5章 局部范围扫描（Partial range scan） 274 5.1 局部范围扫描的概念 276 5.2 局部范围扫描的应用原则 281 5.2.1 局部范围扫描的条件 281 5.2.2 不同优化器模式下的局部范围扫描 284 5.3 提高局部范围扫描执行速度的原理 285 5.4 向局部范围扫描引导的方法 289 5.4.1 利用访问路径实现对Sort的代替 289 5.4.2 只使用索引的局部范围扫描 292 5.4.3 MIN、MAX 的处理 293 5.4.4 FILTER型局部范围扫描 298 5.4.5 ROWNUM的灵活运用 300 5.4.6 利用嵌套视图的局部范围扫描 306 5.4.7 利用函数的局部范围扫描 308 5.4.8 利用查询语句二元化特性的局部范围扫描 316 5.4.9 Web留言板中的局部范围扫描 318第6章 表连接的最优化方案 336 6.1 JOIN和LOOP QUERY的比较 339 6.1.1 全部范围扫描方式下的比较 341 6.1.2 局部范围扫描方式下的比较 349 6.2 连接条件状态对表连接的影响 351 6.2.1 连接条件正常 353 6.2.2 连接条件一边异常 358 6.2.3 连接条件两边异常 361 6.3 各种表连接方式的特征及活用方案 365 6.3.1 嵌套循环连接 366 6.3.1.1 嵌套循环连接的基本概念 367 6.3.1.2 嵌套循环连接顺序的决定 370 6.3.2 排序合并连接 379 6.3.3 嵌套循环连接和排序合并连接的比较 383 6.3.4 哈希连接（Hash Join） 387 6.3.4.1 IN-MEMORY哈希连接 392 6.3.4.2 延迟哈希连接 395 6.3.5 半连接（Semi Join） 398 6.3.5.1 半连接

第一图书网, tushu007.com

的概念和特征 399　6.3.5.2 半连接的执行计划 401　6.3.6 星型（Star）连接 417　6.3.7 星变形（Star Transformation）连接 425　6.3.8 位图连接索引 436

章节摘录

插图：所谓的表和索引分离型的存储结构其实就是堆表，即用来存储数据的表和为了快速查找特定数据而使用的索引完全作为不同的对象来存储。
事实上，在以前的关系型数据库中，大部分的键（Key）和数据都有着非常密切的关系。
当我们从存储的数据中查找某个特定范围中的数据时必须要依赖于键，否则无法正常查找所需要的数据。
因此，键必然会对数据的存储产生一定的影响。
在以前的数据库中查找数据时，只需要查找到键就一定能够在相应的位置查到与其对应的数据，因此从某个角度来看，查找键就是查找数据。
也正因为如此，一度的经验是把键存储在数据的旁边。
其实即使我们把数据和键分别存储在不同的位置，也同样能够通过查找键来定位到数据所在的位置，所以完全没有必要把键和数据存储在一起。
从代价的角度来分析，由于需要执行两次查找，即键的查找和数据的查找，所以与以前的方法相比必然会存在额外的代价。
使用键和数据分离的方法在存储数据时不会受到索引的任何影响，从这一角度来看，又在很大程度上减少了代价的支出。
如果使用这种方法来存储数据，则意味着数据与存储方式之间将不再有任何的关联，而数据完全可以被无条件地存储。
这就好像往抽屉里放东西一样，如果并不需要我们去追究将要放入的物品是什么，而只需要将其挨着放入就可以了的话，应该再没有比这更为简单的方法了。

编辑推荐

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