

联系我们



请扫描二维码联系客服
webmaster@csdn.net
400-660-0108
QQ客服 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有
京ICP证09002463号

经营性网站备案信息
网络110报警服务
中国互联网举报中心
北京互联网违法和不良信息举报中心



文章搜索

- 文章分类
- 算法练习 (13)

数据结构 (12)

设计模式 (4)

数据库相关 (2)

Java相关 (3)

计算机网络 (2)

架构 (1)

- 文章存档
- 2017年11月 (1)

2017年5月 (2)

2017年3月 (3)

2016年10月 (4)

2016年8月 (14)
- 展开

- 阅读排行
- 数据库索引详解 (4530)

经典面试题之最长滑道 (2017... (2338)

数据结构之图的深度优先遍历... (1957)

设计模式系列之单例模式和观... (730)

Java跨平台的原理 (606)

2016华为暑期实习生编程题之... (591)

哈夫曼树和哈夫曼编码 (499)

腾讯2017实习生编程题之有趣... (493)

JVM内存分区和GC回收算法 (479)

经典面试题之字符串的全排列 (471)

数据库索引详解

2016年08月10日 15:46:27 4561人阅读 评论(0) 收藏 举报

分类： 数据库相关 (1)

版权声明：本文为博主原创文章，未经博主允许不得转载。 https://blog.csdn.net/sinat_30186009/article/details/52169057

目录(?) [+]

1.什么是索引？索引的常见类型有什么？

索引就是加快检索表中数据的方法。数据库的索引类似于书籍的索引。在书籍中，索引允许用户不必翻阅完整本书就能迅速地找到所需要的信息。在数据库中，索引也允许数据库程序迅速地找到表中的数据，而不必扫描整个数据库。

分类：这里我们主要介绍常见的聚集索引和非聚集索引

聚集索引：对正文内容按照一定规则排列的目录称为聚集索引

非聚集索引：目录自己按照一定规则排列，正文自己按照另一种规则排列，目录主要是保存对正文的一个映射关系，这种称为非聚集索引

看完上边的定义，你可能对聚集索引和非聚集索引还是一头雾水，没关系，看看下边的例子你应该就明白了

以我们的汉语字典为例，我们要查一个字“爱”，我们知道它的发音，我们会自然的翻开目录中的A目录下去找读音为“ai”的字，找到它并找到它具体对应到哪一页，这里A目录下所有字具体在哪一页都是连续的，假设A目录下“爱”对应18页，在A目录中“爱”的前一个字是“矮”，对应的是17页，也就是他们在目录中的相对顺序也是他们在具体页数中的相对顺序，这个就是聚集索引。

但是如果我们遇到一个字，并不知道它的读音，我们就会采用另一种查找方式，根据“偏旁部首”去查找，然后根据这个字后的页码直接翻到某页来找到您要找的字。但你结合“部首目录”和“检字表”而查到的字的排序并不是真正的正文的排序方法，比如你查“张”字，我们可以看到在查部首之后的检字表中“张”的页码是672页，检字表中“张”的上面是“驰”字，但页码却是63页，“张”的下面是“弩”字，页面是390页，这样目录中的排列方式并不是正文实际的排列方式，这就是非聚集索引。

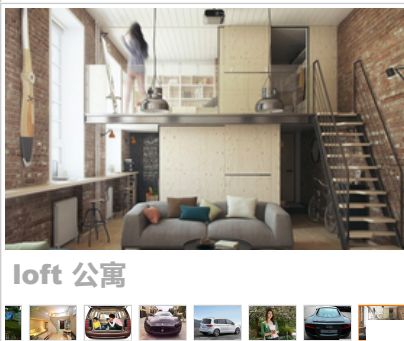
2.索引采取什么数据结构存储？为什么采取这样的数据结构？

大规模的数据不可能全部存储在内存中，故要存储到磁盘上，这样查找读取等操作时就涉及到磁盘IO，那么索引就要尽量减少磁盘IO次数，才能保证查找速度。如果采用普通的二叉查找树结构，会由于树的高度过深进行多次磁盘IO，导致查询效率低下，那么就要尽量减少树的高度，这就引出了B-Tree和B+-Tree,即B树和B+树。

总结：为什么使用B+树？

- 文件很大，不可能全部存储在内存中，故要存储到磁盘上
- 索引的结构组织要尽量减少查找过程中磁盘I/O的存取次数（为什么使用B-/+Tree，还跟磁盘存取原理有关，具体看下边分析）
- 局部性原理与磁盘预读，预读的长度一般为页（page）的整倍数，（在许多操作系统中，页得大小通常为4k）
- 数据库系统巧妙利用了磁盘预读原理，将一个节点的大小设为等于一个页，这样每个节点只需要一次I/O就可以完全载入，（由于节点中有两个数组，所以地址连续）。而红黑树这种结构，h明显要深的多。由于逻辑上很近的节点（父子）物理上可能很远，无法利用局部性

以下资料出自：http://blog.csdn.net/v_JULY_v。



loft 公寓



联系我们



请扫描二维码联系客服

✉ webmaster@csdn.net

☎ 400-660-0108

💬 QQ客服 💬 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

磁盘的构造

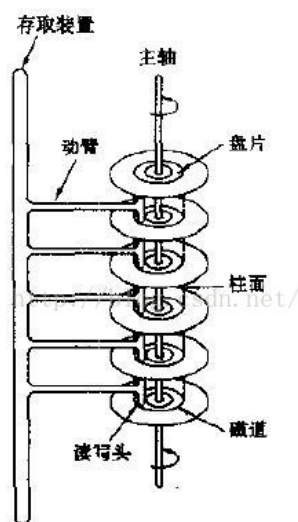


图 11.3 活动头盘示意图

磁盘是一个扁平的圆盘。盘面上有许多称为磁道的圆圈，数据就记录在这些磁道上。磁盘可以是单片的，也可以是由若干盘片组成的盘组，每一盘片上有两个面。如上图11.3中所示的6片盘组为例，除去最顶端和最底端的外侧面不存储数据之外，一共有10个面可以用来保存信息

当磁盘驱动器执行读/写功能时。盘片装在一个主轴上，并绕主轴高速旋转，当磁道在读/写头(又叫磁头)下通过时，就可以进行数据的读 / 写了。

一般磁盘分为固定头盘(磁头固定)和活动头盘。固定头盘的每一个磁道上都有独立的磁头，它是固定不动的，专门负责这一磁道上数据的读/写。

活动头盘(如上图)的磁头是可移动的。每一个盘面上只有一个磁头(磁头是双向的，因此正反盘面都能读写)。它可以从该面的一个磁道移动到另一个磁道。所有磁头都装在同一个动臂上，因此不同盘面上的所有磁头都是同时移动的(行动整齐划一)。当盘片绕主轴旋转的时候，磁头与旋转的盘片形成一个圆柱体。各个盘面上半径相同的磁道组成了一个圆柱面，我们称为柱面。因此，柱面的个数也就是盘面上的磁道数。

磁盘的读写原理及效率

磁盘上的数据需要使用一个三维地址来表示：柱面号、盘面号和块号

读/写磁盘的三个步骤：

- (1) 首先移动臂根据柱面号使磁头移动到所需要的柱面上，这一过程被称为定位或查找。
- (2) 如上图11.3中所示的6盘组示意图中，所有磁头都定位到了10个盘面的10条磁道上(磁头都是双向的)。这时根据盘面号来确定指定盘面上的磁道。
- (3) 盘面确定以后，盘片开始旋转，将指定块号的磁道段移动至磁头下。

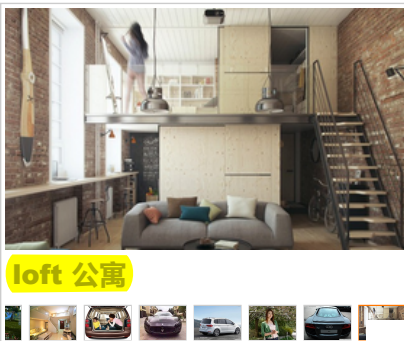
耗时间：

- 查找时间：即完成步骤(1)的时间，这部分耗时最多
- 等待时间：即完成步骤(3)的时间
- 传输时间：数据通过系统总线送到内存的时间

磁盘读取数据是以盘块(block)为基本单位的。位于同一盘块中的所有数据都能被一次性全部读取出来。而磁盘IO代价主要花费在查找时间 T_s 上。因此我们应该尽量将相关信息存放在同一盘块，同一磁道中。或者至少放在同一柱面或相邻柱面上，以求在读/写信息时尽量减少磁头来回移动的次數，避免过多的查找时间

B-Tree

这里不对B-Tree的基础概念的介绍，请先自行百度了解。这里主要介绍一个B树查找的例子。



loft 公寓



联系我们



请扫描二维码联系客服

webmaster@csdn.net

400-660-0108

QQ客服 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有

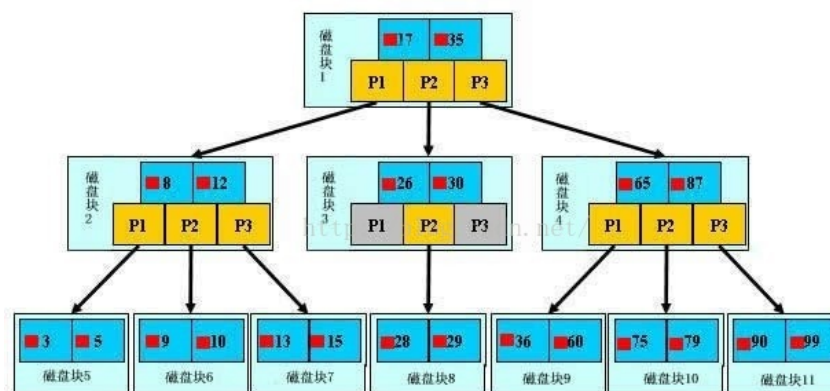
京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心



这里用少量数据构造一棵3叉树的形式，实际应用中的B树结点中关键字很多的。上面的图中比如根结点，其中17表示一个磁盘文件的文件名；小红方块表示这个17文件内容在硬盘中的存储位置；p1表示指向17左子树的指针。

我们假设一个盘块刚好只能存储一个结点，那么上图中一个结点就表示一个盘块，其子树指针就是指向另一个盘块的地 址。

现在我们来模拟查找文件29的过程：

1. 根据根结点指针找到文件目录的根磁盘块1，将其中的信息导入内存。【磁盘IO操作 1次】
2. 此时内存中有两个文件名17、35和三个存储其他磁盘页面地址的数据。根据算法我们发现： $17 < 29 < 35$ ，因此我们找到指针p2。
3. 根据p2指针，我们定位到磁盘块3，并将其中的信息导入内存。【磁盘IO操作 2次】
4. 此时内存中有两个文件名26、30和三个存储其他磁盘页面地址的数据。根据算法我们发现： $26 < 29 < 30$ ，因此我们找到指针p2。
5. 根据p2指针，我们定位到磁盘块8，并将其中的信息导入内存。【磁盘IO操作 3次】
6. 此时内存中有两个文件名28、29。根据算法我们查找到文件名29，并定位了该文件内存的磁盘地址。

分析上面的过程，发现需要3次磁盘IO操作和3次内存查找操作。关于内存中的文件名查找，由于是一个有序表结构，可以利用折半查找提高效率。至于IO操作是影响整个B树查找效率的决定因素。

当然，如果我们使用平衡二叉树的磁盘存储结构来进行查找，磁盘4次，最多5次，而且文件越多，B树比平衡二叉树所用的磁盘IO操作次数将越少，效率也越高。（这里需要重点解释一下，为什么使用平衡二叉树会有更多的IO操作，个人理解是假设内存一次加载的盘块大小固定，上图中我们已经假设了一个盘块刚好只能存储一个三叉树结点大小，如果换成是二叉树，则相比每次读取的关键字少，所以需要更多次的IO）

B+-Tree

B+树是B树的一种变形，这里要注意B+树的两个重要特点：

- 1.所有叶子节点包含了全部关键字的信息，以及指向这些关键字记录的指针，且叶子节点本身依关键字的大小自小而大的顺序链接
- 2.所有非终端节点可以看成是索引，节点中仅含有其子树根节点最大（或最小）的关键字，不包含查找的有效信息（B树则包含）



loft 公寓



联系我们



请扫描二维码联系客服

✉ webmaster@csdn.net

☎ 400-660-0108

💬 QQ客服 💬 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有

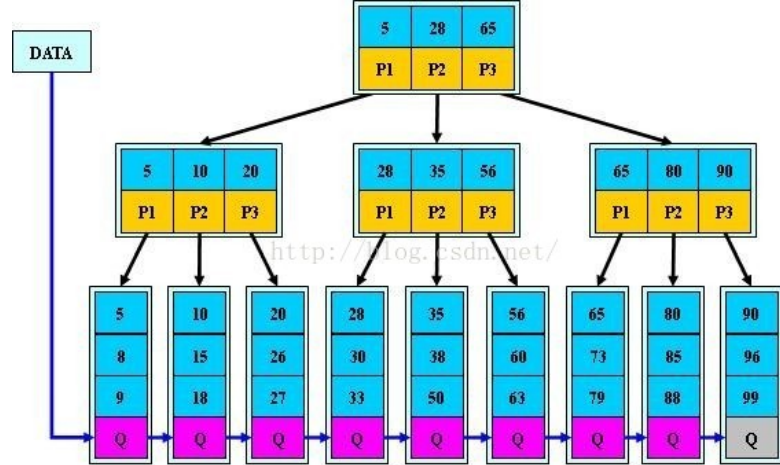
京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心



为什么B+树比B树更适合实际应用中操作系统的文件索引和数据库索引？

1) B⁺-tree的磁盘读写代价更低

B⁺的内部结点并没有指向关键字具体信息的指针。因此其内部结点相对B树更小。如果把所有同一内部结点的关键字存放在同一盘块中，那么盘块所能容纳的关键字数量也越多。一次性读入内存中的需要查找的关键字也就越多。相对来说IO读写次数也就降低了。

举个例子，假设磁盘中的一个盘块容纳16bytes，而一个关键字2bytes，一个关键字具体信息指针2bytes。一棵9阶B树(一个结点最多8个关键字)的内部结点需要2个盘快。而B⁺树内部结点只需要1个盘快。当需要把内部结点读入内存中的时候，B树就比B⁺树多一次盘块查找时间(在磁盘中就盘片旋转的时间)。

2) B⁺-tree的查询效率更加稳定

由于非终结点并不是最终指向文件内容的结点，而只是叶子结点中关键字的索引。所以任何关键字的查找必须走一条从根结点到叶子结点的路。所有关键字查询的路径长度相同，导致每一个数据的查询效率相当。

3.B+树索引和哈希索引区别

如果是等值查询，那么哈希索引有明显有绝对优势，因为只需要经过一次算法即可找到相应的键值；当然了，这个前提是，键值都是唯一的。如果键值不是唯一的，就需要先找到该键所在位置，然后再根据链表往后扫描，直到找到相应的数据从示意图中也能看到。

如果是范围查询检索，这时候哈希索引就毫无用武之地了，因为原先是有序的键值，经过哈希算法后，有可能变成不连续的了，就没办法再利用索引完成范围查询检索；

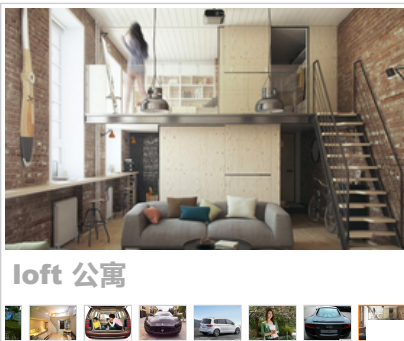
同理，哈希索引也没办法利用索引完成排序，以及like 'xxx%' 这样的部分模糊查询（这种部分模糊查询，其实本质上也是范围查询）；

哈希索引也不支持多列联合索引的最左匹配规则；

B+树索引的关键字检索效率比较平均，不像B树那样波动幅度大，在有大量重复键值情况下，哈希索引的效率也是极低的，因为存在所谓的哈希碰撞问题。

4.InnoDB和MyISAM存储引擎

InnoDB的**主索引**：主索引的key是数据表的主键，因此InnoDB表数据文件本身就是主索引，所以必须有主键，如果没有显示定义，自动为生成一个隐含字段作为主键



loft 公寓

联系我们



请扫描二维码联系客服

webmaster@csdn.net

400-660-0108

QQ客服 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有

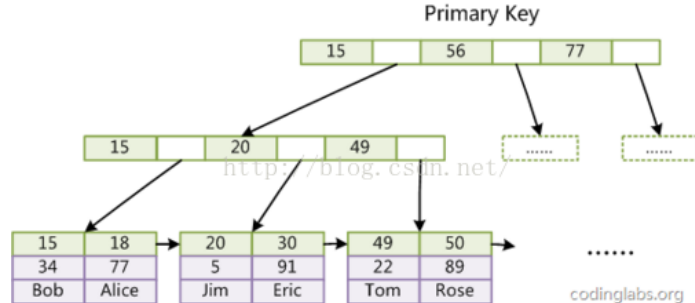
京ICP证09002463号

经营性网站备案信息

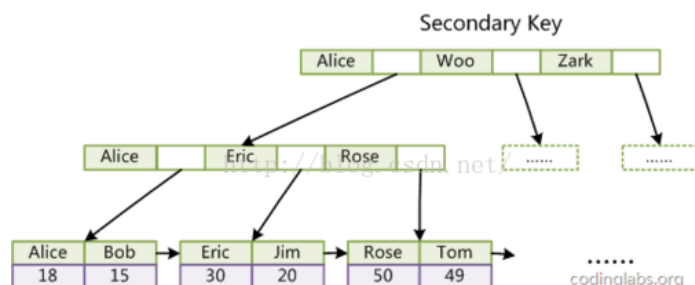
网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

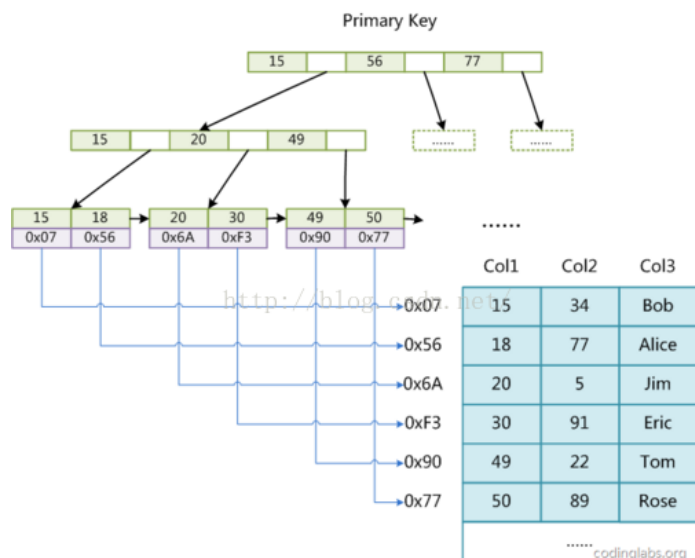


InnoDB的辅助索引：辅助索引也会包含主键列，比如名字建立索引，内部节点 会包含名字，叶子节点会包含该名字对应的主键的值（下图以名字做一个辅助索引）

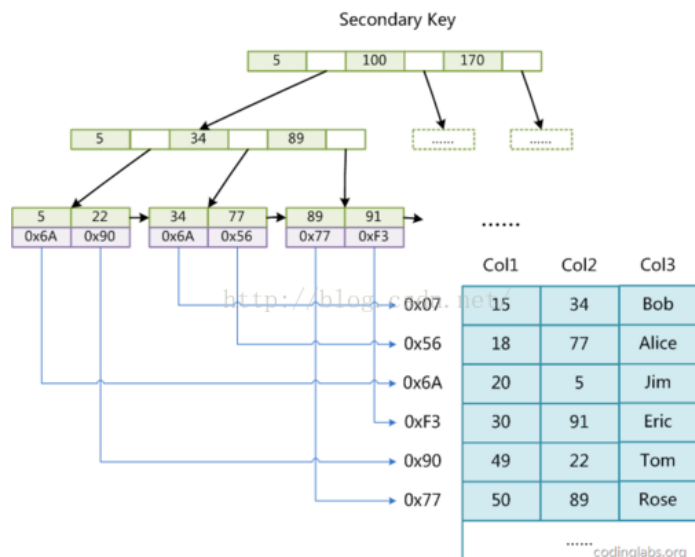


MyISAM主索引和辅助索引在结构上没有任何区别，只是主索引要求key是唯一的，辅助索引可以重复,其叶子节点存储都是数据记录的地址

主索引：



辅助索引：



InnoDB索引和 MyISAM索引 的区别：

一是主索引的区别，InnoDB的数据文件本身就是索引文件(聚集索引)，而MyISAM的索引和数据是分开的（非聚集索引）。

二是辅助索引的区别：InnoDB的辅助索引data域存储相应记录主键的值而不是地址。而MyISAM的辅助索引和主索引没有多大区别。

5.索引的优点和缺点

优点：

大大加快数据的检索速度（主要原因）

缺点：

创建索引和维护索引要耗费时间，这种时间随着数据量的增加而增加

索引需要占物理空间

- [上一篇](#) [经典面试题之求连续子数组最大和](#)
- [下一篇](#) [数据结构之最小生成树](#)

看 Python 如何诠释“薪”时代

Python全栈开发包含Python爬虫、前端、网站后台、Python机器学习与数据挖掘等,从0基础小白到Python 企业级web开发达人、自动化运维开发能手的进击,课程真实企业项目实战演练,全面系统学习python编程语言,从容应对企业中各式各样的.....

[查看更多>>](#)

29224

[查看更多>>](#)



四合院出售



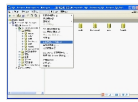
大腿抽脂



种植胡须多少钱



一克拉钻戒价格



sql数据库培训



洗牙有害



电视卫星接收

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

数据库索引概念和分类

luzhangting 2016年05月26日 21:41 2229

数据库索引概念和分类，几种常见索引。

数据库索引类型及实现方式

sdgihsdv 2017年07月12日 18:43 2995

数据库索引类型及实现方式 1、索引定义 数据库索引好比是一本书前面的目录，能加快数据库的查询速度。索引是对数据库表中一个或多个列（例如，employee 表的姓氏 (lname) ...

程序员不会英语怎么行？

老司机教你一个数学公式秒懂天下英语
广告



数据库调优教程（五）索引的作用和索引的种类

hzy38324 2015年04月07日 16:55 2414

这一章，我们将全方位讲解如何使用索引来优化我们的数据库

数据库中索引的优缺点和分类

dianzijinglin 2016年08月19日 16:42 1274

1、索引的概念 索引就是为了提高数据的检索速度。数据库的索引类似于书籍的索引。在书籍中，索引允许用户不必翻阅整本书就能迅速地找到所需要的信息。在数据库中，索引也允许数据库程序迅速地找到表中的数据，...



联系我们

请扫描二维码联系客服

✉ webmaster@csdn.net

☎ 400-660-0108

🗣 QQ客服 💬 客服论坛

关于 招聘 广告服务 百度

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

数据库索引类型

ducduc 2009年05月31日 09:59 9339

逻辑上：Single column 单行索引Concatenated 多行索引Unique 唯一索引NonUnique 非唯一索引Function-based函数索引Domain 域索引物理上：Par...

知网论文查重入口

知网论文检测查重系统

百度广告

数据库-索引（概念，优缺点，分类）

zdplife 2015年08月27日 20:08 3021

（1）索引概念：索引是由用户创建，能够被修改和删除的，实际存储在数据库中的物理存在，它是某一个表中一行或者若干列值的集合和相应的指向表中物理标志这些值的数据页的逻辑指针清单。（2）索引的优点：第...

mysql索引总结----mysql 索引类型以及创建

xluren 2014年06月20日 23:28 209142

关于MySQL索引的好处，如果正确合理设计并且使用索引的MySQL是一辆兰博基尼的话，那么没有设计和使用索引的MySQL就是一个人力三轮车。对于没有索引的表，单表查询可能几十万数据就是瓶颈，而通常大型...

数据库索引的优缺点及原理

MiracleWW 2016年11月26日 16:47 9426

1. 什么是索引：索引就像是书的目录，是与表或视图关联的磁盘上结构，可以加快从表或视图中检索行的速度。索引中包含由表或视图中的一行或多行生成的键。这些键存储在一个结构（BTree）中，使SQL可以...

Mysql几种索引类型的区别及适用情况

dyllove98 2013年07月30日 18:39 69533

如大家所知道的，Mysql目前主要有以下几种索引类型：FULLTEXT，HASH，BTREE，RTREE。那么，这几种索引有什么功能和性能上的不同呢？FULLTEXT 即为全文索引，目前只有...

数据库的几种索引

chinasoftosg 2013年03月09日 07:52 5036

①按照索引列值的唯一性，索引可分为唯一索引和非唯一索引 ②非唯一索引：B树索引 create index 索引名 on 表名（列名）tablespace 表空间名; ③唯一索引：建立主键或...

程序员不会英语怎么行？

老司机教你一个数学公式秒懂天下英语

数据库索引-扫盲

flyhawk_xjtu 2016年03月02日 14:59 536

位图索引主要针对大量相同值的列而创建。拿全国居民登录一第表来说，假设有四个字段：姓名、性别、年龄、和身份证号，年龄和性别两个字段会产生许多相同的值，性别只有男女两种值，年龄，1到120（假设最大年龄1...

MySQL数据库索引类型以及创建

yccowdy 2017年08月20日 16:07 119

相关概念优点：提高数据库从表中检索数据的速度。索引检索速度快的原因：由于数据存储在数据表中所以索引是创建在数据表对象上的，由表中一个或者多个键组成了索引，这些键存储在数据结构（b-tree或者ha...

数据库索引的类型

sanyaouxu_2 2018年01月08日 17:39 34

B-树索引 位图索引 HASH索引 索引编排表 反转键索引 基于函数的索引 分区索引 本地和全局索引 索引结构：B-tree：适合与大量的增、删、改（OL...

数据库索引到底是什么，是怎样工作的？

weiliangliang111 2016年05月19日 16:37 35257



联系我们



请扫描二维码联系客服

✉ webmaster@csdn.net

☎ 400-660-0108

🗨 QQ客服 🗨 客服论坛

关于 招聘 广告服务 

©1999-2018 CSDN版权所有
京ICP证09002463号

经营性网站备案信息
网络110报警服务
中国互联网举报中心
北京互联网违法和不良信息举报中心

我们通过一个简单的例子来开始教程，解释为什么我们需要数据库索引。假设我们有一个数据库表 Employee，这个表有三个字段（列）分别是 Employee_Name、Employee_Age 和Emp...

免费云主机试用一年

云服务器免费试用

百度广告



数据库索引的通俗理解

 sundacheng1989 2016年11月10日 16:40 4634

最近使用到Oracle数据库的索引比较多，所以就好好研究一下索引到底是什么。毕竟作为一个Application Developer，而不是DBA，所以这篇文字也是很通俗，特别浅显的描述了一下索引...

数据库索引的实现原理

 kennyrose 2012年05月03日 17:05 226575

强烈建议参阅链接：<http://www.linezing.com/blog/?p=798#nav-1> 说白了，索引问题就是一个查找问题。。。数据库索引，是数据库管理系统中一个排序...

数据库索引的使用

 jianxuanlu 2008年06月13日 11:19 7825

实际上，您可以把索引理解为一种特殊的目录。微软的SQL SERVER提供了两种索引：聚集索引（clustered index，也称聚类索引、簇集索引）和非聚集索引（nonclustered index...

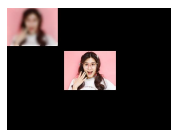
数据库索引原理及优化

 suifeng3051 2016年09月26日 14:30 18158

本文内容主要来源于互联网上主流文章，只是按照个人理解稍作整合，后面附有参考链接。一、摘要本文以MySQL数据库为研究对象，讨论与数据库索引相关的一些话题。特别需要说明的是，MySQL支持诸多存储引擎，...

码农不会英语怎么行？英语文档都看不懂！

软件工程出身的英语老师，教你用数学公式读懂天下英文→



深入浅出数据库索引原理

 weizhiai12 2017年04月03日 17:20 1014

深入浅出数据库索引原理 前段时间，公司一个新上线的网站出现页面响应速度缓慢的问题，一位负责这个项目的但并不是搞技术的妹子找到我，让我想办法提升网站的访问速度，因为已经有很多用户来投诉了。我第...

数据库索引的作用优点和缺点

 u013310119 2016年09月13日 17:10 8289

为什么要创建索引呢？这是因为，创建索引可以大大提高系统的性能。 第一，通过创建唯一性索引，可以保证数据库表中每一行数据的唯一性。 第二，可以大大加快 数据的检索速度，这也是创建索引的最主要的原因...