1. 索引
2. 索引的定义
3. 索引的数据类型
4. 索引的类型
5. MySQL索引的优化
6. 底层索引的实现
7. 高性能的索引策略
8. 索引与加锁

一、索引

1. 索引的定义
2. 索引是一种特殊的文件(InnoDB数据表上的索引是表空间的一个组成部分)，它们包含着对数据表里所有记录的引用指针。更通俗的说，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。
3. 索引仅仅提供创建、删除和修改的功能，索引不是关系数据库理论的产物，是为了性能和维护工作而出现的东西，主要用来加快数据操作速度和提高数据的存储利用。
4. 索引的数据类型

MySQL支持很多数据类型，选择合适的数据类型存储数据对性能有很大的影响。通常来说，可以遵循以下一些指导原则：

(1) 越小的数据类型通常更好：越小的数据类型通常在磁盘、内存和CPU缓存中都需要更少的空间，处理起来更快。

(2) 简单的数据类型更好：整型数据比起字符，处理开销更小，因为字符串的比较更复杂。在MySQL中，应该用内置的日期和时间数据类型，而不是用字符串来存储时间；以及用整型数据类型存储IP地址。

(3) 尽量避免NULL：应该指定列为NOT NULL，除非你想存储NULL。在MySQL中，含有空值的列很难进行查询优化，因为它们使得索引、索引的统计信息以及比较运算更加复杂。你应该用0、一个特殊的值或者一个空串代替空值。

2.1 选择合适的标识符

(1) 整型：通常是作为标识符的最好选择，因为可以更快的处理，而且可以设置为AUTO\_INCREMENT。

(2) 字符串：尽量避免使用字符串作为标识符，它们消耗更好的空间，处理起来也较慢。

而且，通常来说，字符串都是随机的，所以它们在索引中的位置也是随机的，这会导致页面分裂、随机访问磁盘，聚簇索引分裂（对于使用聚簇索引的存储引擎）。

1. 索引的类型

每种存储引擎的索引都不一定完全相同，并不是所有的存储引擎都支持所有的索引类型。

* 1. 普通索引（INDEX）

最基本的索引，没有任何限制，MyIASM中默认的BTREE类型的索引，也是大多数情况下用到的索引。

1. 实例代码：

–直接创建索引

CREATE INDEX index\_name ON table(column(length))

–修改表结构的方式添加索引

ALTER TABLE table\_name ADD INDEX index\_name ON (column(length))

–创建表的时候同时创建索引

CREATE TABLE `table` (

`id` int(11) NOT NULL AUTO\_INCREMENT ,

`title` char(255) CHARACTER SET utf8 COLLATE utf8\_general\_ci NOT NULL ,

`content` text CHARACTER SET utf8 COLLATE utf8\_general\_ci NULL ,

`time` int(10) NULL DEFAULT NULL ,

PRIMARY KEY (`id`),

INDEX index\_name (title(length)))

–删除索引

DROP INDEX index\_name ON table

* 1. 唯一索引（UNIQUE）

与普通索引类似，不同的就是：索引列的值必须唯一，但允许有空值（注意和主键不同）。如果是组合索引，则列值的组合必须唯一，创建方法和普通索引类似。

1. 实例代码：

–创建唯一索引

CREATE UNIQUE INDEX indexName ON table(column(length))

–修改表结构

ALTER TABLE table\_name ADD UNIQUE indexName ON (column(length))

–创建表的时候直接指定

CREATE TABLE `table` (

`id` int(11) NOT NULL AUTO\_INCREMENT ,

`title` char(255) CHARACTER SET utf8 COLLATE utf8\_general\_ci NOT NULL ,

`content` text CHARACTER SET utf8 COLLATE utf8\_general\_ci NULL ,

`time` int(10) NULL DEFAULT NULL ,

PRIMARY KEY (`id`),

UNIQUE indexName (title(length))

);

* 1. 全文索引（FULLTEXT）

FULLTEXT索引仅可用于 MyISAM 表；他们可以从CHAR、VARCHAR或TEXT列中作为CREATE TABLE语句的一部分被创建，或是随后使用ALTER TABLE 或CREATE INDEX被添加。对于较大的数据集，将你的资料输入一个没有FULLTEXT索引的表中，然后创建索引，其速度比把资料输入现有FULLTEXT索引的速度更为快。不过切记对于大容量的数据表，生成全文索引是一个非常消耗时间非常消耗硬盘空间的做法。

1. 实例代码：

–创建表的适合添加全文索引

CREATE TABLE `table` (

`id` int(11) NOT NULL AUTO\_INCREMENT ,

`title` char(255) CHARACTER SET utf8 COLLATE utf8\_general\_ci NOT NULL ,

`content` text CHARACTER SET utf8 COLLATE utf8\_general\_ci NULL ,

`time` int(10) NULL DEFAULT NULL ,

PRIMARY KEY (`id`),

FULLTEXT (content)

);

–修改表结构添加全文索引

ALTER TABLE article ADD FULLTEXT index\_content(content)

–直接创建索引

CREATE FULLTEXT INDEX index\_content ON article(content)

* 1. 单列索引和多列索引

多个单列索引与单个多列索引的查询效果不同，因为执行查询时，MySQL只能使用一个索引，会从多个索引中选择一个限制最为严格的索引。

* 1. 组合索引（最左前缀）

1. 平时用的SQL查询语句一般都有比较多的限制条件，所以为了进一步榨取MySQL的效率，就要考虑建立组合索引。例如上表中针对title和time建立一个组合索引：ALTER TABLE article ADD INDEX index\_titme\_time (title(50),time(10))。建立这样的组合索引，其实是相当于分别建立了下面两组组合索引：

–title,time

–title

1. 为什么没有time这样的组合索引呢？这是因为MySQL组合索引“最左前缀”的结果。简单的理解就是只从最左面的开始组合。并不是只要包含这两列的查询都会用到该组合索引，如下面的几个SQL所示

–使用到上面的索引

SELECT \* FROM article WHERE title='测试' AND time=1234567890;

SELECT \* FROM article WHERE title='测试';

–不使用上面的索引

SELECT \* FROM article WHREE time=1234567890;

1. MySQL索引的优化
2. 缺点：

虽然索引大大提高了查询速度，同时却会降低更新表的速度，如对表进行INSERT、UPDATE和DELETE。因为更新表时，MySQL不仅要保存数据，还要保存一下索引文件。建立索引会占用磁盘空间的索引文件。

1. 分类：

索引分为聚集索引和非聚集索引两种，聚集索引是按照数据存放的物理位置为顺序的，而非聚簇索引就不一样了。聚集索引能提高多行检索的速度，而非聚集索引对于单行的检索很快。

* 1. 聚集索引和非聚集索引的应用场景



* 1. 索引的一些注意事项

1. 索引不会包含有NULL值的列

只要列中包含有NULL值都将不会被包含在索引中，复合索引中只要有一列含有NULL值，那么这一列对于此复合索引就是无效的。所以我们在数据库设计时不要让字段的默认值为NULL。

1. 使用短索引

对串列进行索引，如果可能应该指定一个前缀长度。例如，如果有一个CHAR(255)的列，如果在前10个或20个字符内，多数值是惟一的，那么就不要对整个列进行索引。短索引不仅可以提高查询速度而且可以节省磁盘空间和I/O操作。

如：ALERT table\_name ADD INDEX index\_name ON id(id(4));

1. 索引列排序

MySQL查询只使用一个索引，因此如果where子句中已经使用了索引的话，那么order by中的列是不会使用索引的。因此数据库默认排序可以符合要求的情况下不要使用排序操作；尽量不要包含多个列的排序，如果需要最好给这些列创建复合索引。

1. like语句操作（不会使用索引）

一般情况下不鼓励使用like操作，如果非使用不可，如何使用也是一个问题。like “%aaa%” 不会使用索引而like “aaa%”可以使用索引。

1. 不要在列上进行运算

在列上进行运算会导致数据库不使用索引，而进行全盘扫描。

1. MySQL只对一下操作符才使用索引：<,<=,=,>,>=,between,in,以及某些时候的like(不以通配符%或\_开头的情形)，而理论上每张表里面最多可创建16个索引.
2. 底层索引的实现

索引是在存储引擎中实现的，而不是在服务器层中实现的。所以，每种存储引擎的索引都不一定完全相同，并不是所有的存储引擎都支持所有的索引类型。

* 1. B-Tree索引

创建一张表

CREATE TABLE People (

last\_name varchar(50) not null,

first\_name varchar(50) not null,

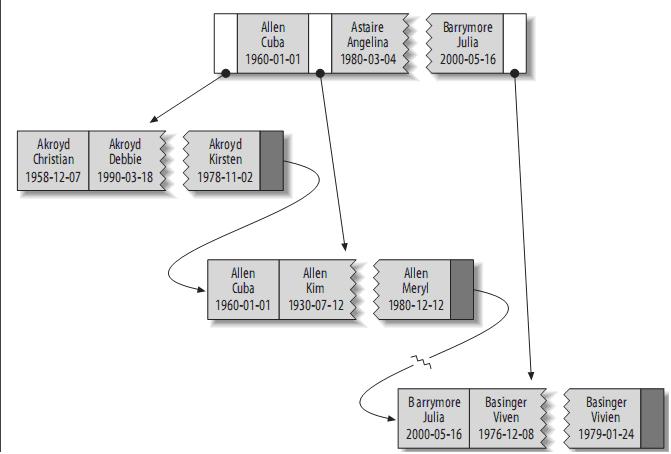
dob date not null,

gender enum('m', 'f') not null,

key(last\_name, first\_name, dob)

);

B树索引结构



索引存储的值按索引列中的顺序排列。可以利用B-Tree索引进行**全关键字、关键字范围和关键字前缀查询**，当然，如果想使用索引，你必须保证**按索引的最左边前缀**(leftmost prefix of the index)来进行查询。-- 叶子节点通过链表连接

(1) 匹配全值(Match the full value)：对索引中的所有列都指定具体的值。例如，上图中索引可以帮助你查找出生于1960-01-01的Cuba Allen。

(2) 匹配最左前缀(Match a leftmost prefix)：你可以利用索引查找last name为Allen的人，仅仅使用索引中的第1列。

(3) 匹配列前缀(Match a column prefix)：例如，你可以利用索引查找last name以J开始的人，这仅仅使用索引中的第1列。

(4) 匹配值的范围查询(Match a range of values)：可以利用索引查找last name在Allen和Barrymore之间的人，仅仅使用索引中第1列。

(5) 匹配部分精确而其它部分进行范围匹配(Match one part exactly and match a range on another part)：可以利用索引查找last name为Allen，而first name以字母K开始的人。

(6) 仅对索引进行查询(Index-only queries)：如果查询的列都位于索引中，则不需要读取元组的值。

由于B-树中的节点都是顺序存储的，所以可以利用索引进行查找(找某些值)，也可以对查询结果进行**ORDER BY**。当然，使用B-tree索引有以下一些限制：

(1) **查询必须从索引的最左边的列开始**。关于这点已经提了很多遍了。例如你不能利用索引查找在某一天出生的人。

(2) 不能跳过某一索引列。例如，你不能利用索引查找last name为Smith且出生于某一天的人。

(3) 存储引擎不能使用索引中范围条件右边的列。例如，如果你的查询语句为WHERE last\_name="Smith" AND first\_name LIKE 'J%' AND dob='1976-12-23'，则该查询只会使用索引中的前两列，因为LIKE是范围查询。

* 1. Hash索引

MySQL中，只有Memory存储引擎显示支持hash索引，是Memory表的默认索引类型，尽管Memory表也可以使用B-Tree索引。Memory存储引擎支持非唯一hash索引，这在数据库领域是罕见的，如果多个值有相同的hash code，索引把它们的行指针用链表保存到同一个hash表项中。

创建一张表：

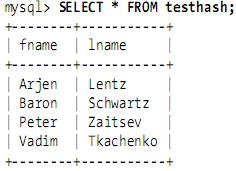
CREATE TABLE testhash (

fname VARCHAR(50) NOT NULL,

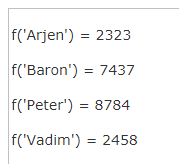
lname VARCHAR(50) NOT NULL,

KEY USING HASH(fname)

) ENGINE=MEMORY;



1. 表使用hash函数-f()



1. Hash索引的结构



1. Hash的执行顺序
2. mysql> SELECT lname FROM testhash WHERE fname='Peter';
3. MySQL会计算’Peter’的hash值，然后通过它来查询索引的行指针。因为f('Peter') = 8784，MySQL会在索引中查找8784，得到指向记录3的指针。
4. **先计算hash值，查询索引的行指针，查询到结果。**
5. 因为索引自己仅仅存储很短的值，所以，索引非常紧凑。Hash值不取决于列的数据类型，一个TINYINT列的索引与一个长字符串列的索引一样大。
6. Hash索引的一些限制

(a) 由于索引仅包含hash code和记录指针，所以，MySQL不能通过使用索引避免读取记录。但是访问内存中的记录是非常迅速的，不会对性能造成太大的影响。

(b) 不能使用Hash索引排序。由于 Hash 索引中存放的是经过 Hash 计算之后的 Hash值，而且Hash值的大小关系并不一定和 Hash运算前的键值完全一样，所以数据库无法利用索引的数据来避免任何排序运算；

(c) Hash索引不支持键的部分匹配，因为是通过整个索引值来计算hash值的。

(d) Hash索引只支持等值比较，例如使用=，IN( )和<=>。对于WHERE price>100并不能加速查询。

(e) Hash索引遇到大量Hash值相等的情况后性能并不一定就会比B-Tree索引高。

* 1. 空间(R-Tree)索引

MyISAM支持空间索引，主要用于地理空间数据类型，例如GEOMETRY。

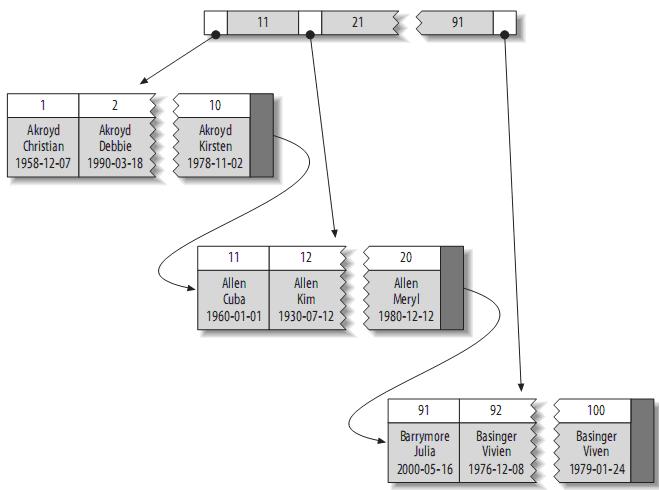
* 1. 全文(Full-text)索引

全文索引是MyISAM的一个特殊索引类型，主要用于全文检索。

1. 高性能的索引策略
   1. 聚集索引

聚集索引保证关键字的值相近的元组存储的物理位置也相同（所以字符串类型不宜建立聚集索引，特别是随机字符串，会使得系统进行大量的移动操作），且一个表只能有一个聚集索引。因为由存储引擎实现索引，所以，并不是所有的引擎都支持聚集索引。目前，只有solidDB和InnoDB支持**。**

1. 聚集索引的结构



注：**叶子页面包含完整的元组，而内节点页面仅包含索引的列(索引的列为整型)**。一些DBMS允许用户指定聚簇索引，但是MySQL的存储引擎到目前为止都不支持。**InnoDB对主键建立聚簇索引。**如果你不指定主键，InnoDB会用一个具有唯一且非空值的索引来代替。如果不存在这样的索引，InnoDB会定义一个隐藏的主键，然后对其建立聚簇索引。**一般来说，DBMS都会以聚簇索引的形式来存储实际的数据，它是其它二级索引的基础**。

* 1. InnoDB和MyISAM的数据布局的比较

为了更加理解聚簇索引和非聚簇索引，或者primary索引和second索引(MyISAM不支持聚簇索引)，来比较一下InnoDB和MyISAM的数据布局，对于如下表：

CREATE TABLE layout\_test (

col1 int NOT NULL,

col2 int NOT NULL,

PRIMARY KEY(col1),

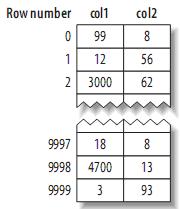
KEY(col2)

);

假设主键的值位于1---10,000之间，且按随机顺序插入，然后用OPTIMIZE TABLE进行优化。col2随机赋予1---100之间的值，所以会存在许多重复的值。

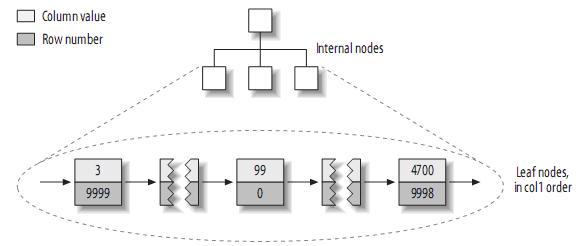
1. MyISAM的数据布局

其布局十分简单，MyISAM按照插入的顺序在磁盘上存储数据，如下：



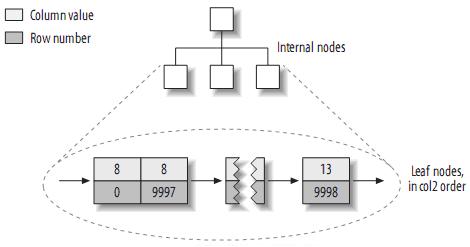
注：左边为行号(row number)，从0开始。因为元组的大小固定，所以MyISAM可以很容易的从表的开始位置找到某一字节的位置。

1. 据默认建立的primary key的索引结构大致如下：



注：MyISAM不支持聚簇索引，索引中每一个叶子节点仅仅包含行号(row number)，且叶子节点按照col1的顺序存储。

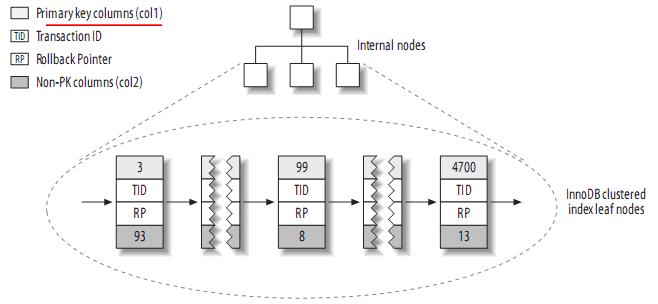
1. 列col2的索引结构：



实际上，在MyISAM中，primary key和其它索引没有什么区别。Primary key仅仅只是一个叫做PRIMARY的唯一，非空的索引而已。

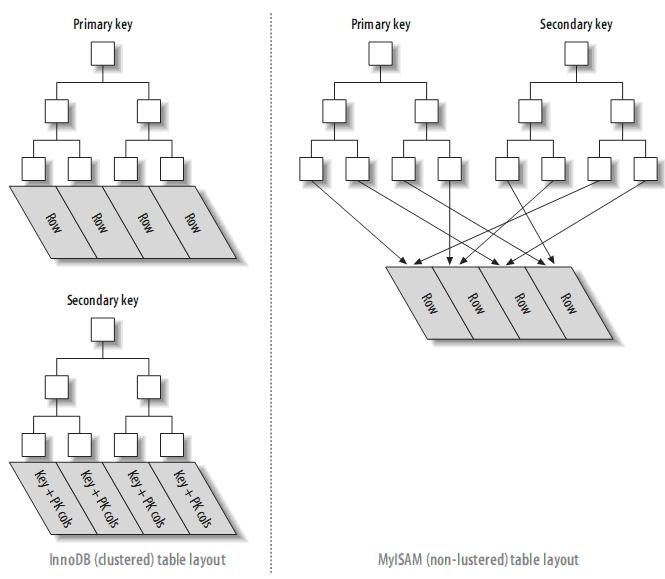
1. InnoDB的数据布局

InnoDB按聚簇索引的形式存储数据，所以它的数据布局有着很大的不同。它存储表的结构大致如下：



注：**聚簇索引中的每个叶子节点包含primary key的值，事务ID和回滚指针(rollback pointer)——用于事务和MVCC，和余下的列(如col2)。**

1. 聚簇索引和非聚簇索引表的对比：



1. 按primary key的顺序插入行(InnoDB)

如果你用InnoDB，而且不需要特殊的聚簇索引，一个好的做法就是使用代理主键(surrogate key)——独立于你的应用中的数据。最简单的做法就是使用一个AUTO\_INCREMENT的列，这会保证记录按照顺序插入，而且能提高使用primary key进行连接的查询的性能。**应该尽量避免随机的聚簇主键，例如，字符串主键就是一个不好的选择，它使得插入操作变得随机。**

* 1. 覆盖索引

**如果索引包含满足查询的所有数据，就称为覆盖索引。**覆盖索引是一种非常强大的工具，能大大提高查询性能。

只需要读取索引而不用读取数据有以下一些优点：

(1) 索引项通常比记录要小，所以MySQL访问更少的数据；

(2) 索引都按值的大小顺序存储，相对于随机访问记录，需要更少的I/O；

(3) 大多数据引擎能更好的缓存索引。比如MyISAM只缓存索引。

(4) 覆盖索引对于InnoDB表尤其有用，因为InnoDB使用聚集索引组织数据，如果二级索引中包含查询所需的数据，就不再需要在聚集索引中查找了。

* + 1. 覆盖索引的实现

1. 覆盖索引不能是任何索引，只有B-TREE索引存储相应的值。**而且不同的存储引擎实现覆盖索引的方式都不同，并不是所有存储引擎都支持覆盖索引(Memory和Falcon就不支持)。**
2. 对于索引覆盖查询(index-covered query)，使用EXPLAIN时，可以在Extra一列中看到“Using index”。例如，在sakila的inventory表中，有一个组合索引(store\_id,film\_id)，对于只需要访问这两列的查询，MySQL就可以使用索引，如下：

mysql> EXPLAIN SELECT store\_id, film\_id FROM sakila.inventory\G

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 1. row \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

id: 1

select\_type: SIMPLE

table: inventory

type: index

possible\_keys: NULL

key: idx\_store\_id\_film\_id

key\_len: 3

ref: NULL

rows: 5007

Extra: Using index

1 row in set (0.17 sec)

1. 在大多数引擎中，只有当查询语句所访问的列是索引的一部分时，索引才会覆盖。但是，InnoDB不限于此，InnoDB的二级索引在叶子节点中存储了primary key的值。因此，sakila.actor表使用InnoDB，而且对于是last\_name上有索引，所以，索引能覆盖那些访问actor\_id的查询，如：

mysql> EXPLAIN SELECT actor\_id, last\_name

-> FROM sakila.actor WHERE last\_name = 'HOPPER'\G

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 1. row \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

id: 1

select\_type: SIMPLE

table: actor

type: ref

possible\_keys: idx\_actor\_last\_name

key: idx\_actor\_last\_name

key\_len: 137

ref: const

rows: 2

Extra: Using where; Using index

* 1. 利用索引进行排序

MySQL中，有两种方式生成有序结果集：一是使用filesort，二是按索引顺序扫描。**利用索引进行排序操作是非常快的，而且可以利用同一索引同时进行查找和排序操作**。当索引的顺序与ORDER BY中的列顺序相同且所有的列是同一方向(全部升序或者全部降序)时，可以使用索引来排序。如果查询是连接多个表，仅当ORDER BY中的所有列都是第一个表的列时才会使用索引。其它情况都会使用filesort。

1. 创建表

create table actor(

actor\_id int unsigned NOT NULL AUTO\_INCREMENT,

name varchar(16) NOT NULL DEFAULT '',

password varchar(16) NOT NULL DEFAULT '',

PRIMARY KEY(actor\_id),

KEY (name)

) ENGINE=InnoDB

insert into actor(name,password) values('cat01','1234567');

insert into actor(name,password) values('cat02','1234567');

insert into actor(name,password) values('ddddd','1234567');

insert into actor(name,password) values('aaaaa','1234567');

1. 利用索引排序

mysql> explain select actor\_id from actor order by actor\_id \G

Extra: Using index

mysql> explain select actor\_id from actor order by password \G

Extra: Using filesort

mysql> explain select actor\_id from actor order by name \G

Extra: Using index

* + 1. 当不能使用索引排序

当MySQL不能使用索引进行排序时，就会利用自己的排序算法(快速排序算法)在内存(sort buffer)中对数据进行排序，如果内存装载不下，它会将磁盘上的数据进行分块，再对各个数据块进行排序，然后将各个块合并成有序的结果集（实际上就是外排序）。对于filesort，MySQL有两种排序算法。

1. 两遍扫描算法(Two passes)

实现方式是先将须要排序的字段和可以直接定位到相关行数据的指针信息取出，然后在设定的内存（通过参数sort\_buffer\_size设定）中进行排序，完成排序之后再次通过行指针信息取出所需的Columns。

注：该算法是4.1之前采用的算法，它需要两次访问数据，尤其是第二次读取操作会导致大量的随机I/O操作。另一方面，内存开销较小。

1. 一次扫描算法(single pass)

该算法一次性将所需的Columns全部取出，在内存中排序后直接将结果输出。

注：从 MySQL 4.1 版本开始使用该算法。它减少了I/O的次数，效率较高，但是内存开销也较大。如果我们将并不需要的Columns也取出来，就会极大地浪费排序过程所需要的内存。在 MySQL 4.1 之后的版本中，可以通过设置 max\_length\_for\_sort\_data 参数来控制 MySQL 选择第一种排序算法还是第二种。**当取出的所有大字段总大小大于**

**max\_length\_for\_sort\_data 的设置时，MySQL 就会选择使用第一种排序算法，反之，则会选择第二种。**为了尽可能地提高排序性能，我们自然更希望使用第二种排序算法，所以在 Query 中仅仅取出需要的 Columns 是非常有必要的。

1. 当对连接操作进行排序时，如果ORDER BY仅仅引用第一个表的列，MySQL对该表进行filesort操作，然后进行连接处理，此时，EXPLAIN输出“Using filesort”；否则，MySQL必须将查询的结果集生成一个临时表，在连接完成之后进行filesort操作，此时，EXPLAIN输出“Using temporary;Using filesort”。
2. 索引与加锁
3. 索引对于InnoDB非常重要，因为它可以让查询锁更少的元组。这点十分重要，因为MySQL 5.0中，InnoDB直到事务提交时才会解锁。有两个方面的原因：首先，即使InnoDB行级锁的开销非常高效，内存开销也较小，但不管怎么样，还是存在开销。其次，对不需要的元组的加锁，会增加锁的开销，降低并发性。
4. InnoDB仅对需要访问的元组加锁，而索引能够减少InnoDB访问的元组数。**但是，只有在存储引擎层过滤掉那些不需要的数据才能达到这种目的。**一旦索引不允许InnoDB那样做（即达不到过滤的目的），MySQL服务器只能对InnoDB返回的数据进行WHERE操作，此时，已经无法避免对那些元组加锁了：InnoDB已经锁住那些元组，服务器无法解锁了。-- 使服务器锁住更少的元组数

如：SET AUTOCOMMIT=0;

BEGIN;

SELECT actor\_id FROM actor WHERE actor\_id < 4

AND actor\_id <> 1 FOR UPDATE;

* InnoDB只对SELECT语句中WHERE子句actor\_id<4使用索引，没有对actor\_id<>1使用索引，所以MySQL服务器会锁住1,2,3三个元组。使用索引就是为了让MySQL服务器锁住更少的元组，更高的锁住元组。