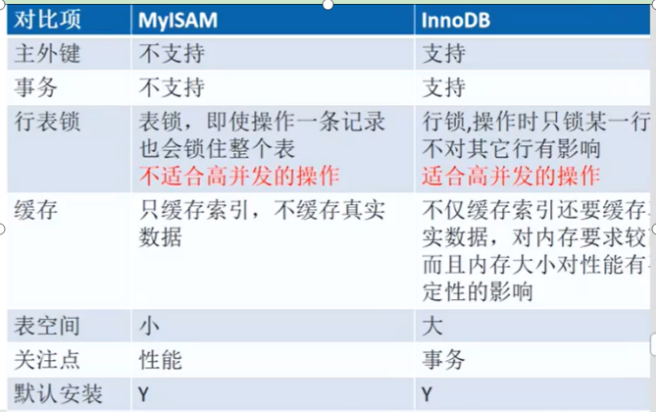
**MySQL**

# MyISAM和InnoDB的区别

1. MyISAM只支持表锁，InfoDB支持表锁和行锁。
2. MyISAM不支持事务，InfoDB支持事务。



# MyISAM

## 共享读锁

1. 创建一张表

|  |
| --- |
| CREATE TABLE `testisam` (  `id` int(11) DEFAULT NULL  ) ENGINE=MyISAM DEFAULT CHARSET=utf8mb4; |

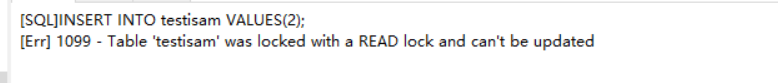
1. 插入数据

|  |
| --- |
| INSERT INTO testisam VALUES(2); |

1. 开启行锁

|  |
| --- |
| LOCK TABLE testisam READ; |

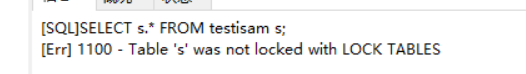
1. 再次执行第二步中的插入语句，结果如下：



1. 加上别名

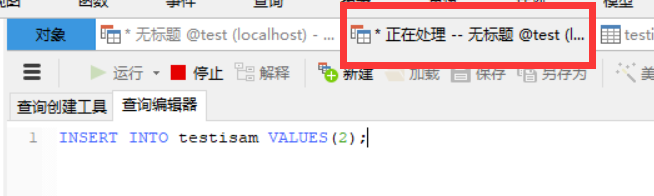
|  |
| --- |
| SELECT s.\* FROM testisam s; |

查询报错:

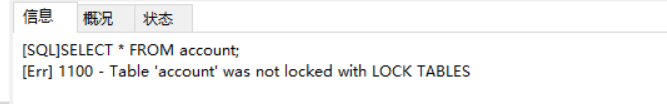


1. 打开另一个会话，继续进行插入语句，结果如下:

处于等待状态



1. 在开启行锁的相同的会话中，查询其他的表，结果如下:



1. 解锁

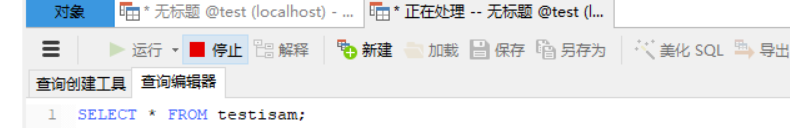


## 写锁

1. 开启读锁

|  |
| --- |
| LOCK TABLE testisam WRITE; |

1. 执行插入、更新、删除、查询语句都是可以的。
2. 查询其他表也是不能成功的（和开启读锁同一个会话）。
3. 在另一个会话中，查询testisam 中的数据，结果：**等待中**



## 总结锁



# InnoDB

## innodb\_page\_size

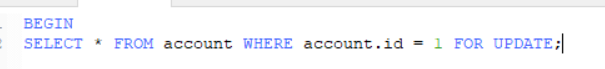
show variables like 'innodb\_page\_size';



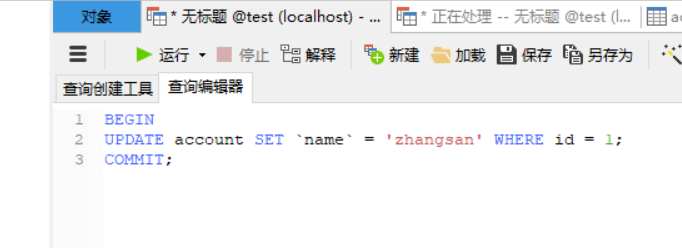
只支持B-tree索引

## 行锁

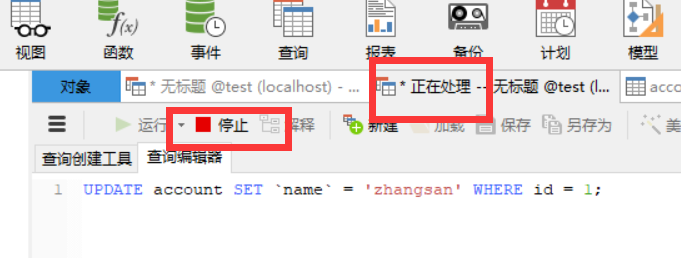
1. 开启行锁



**Update，删除语句会自动添加行锁**



1. 在另一个会话中修改数据，id一样，结果**:在等待，因为开启了行锁，在将上面的事务提交之后，该条数据自动修改成功。**



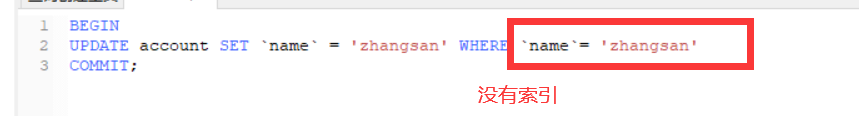
**超时之后，自动取消**



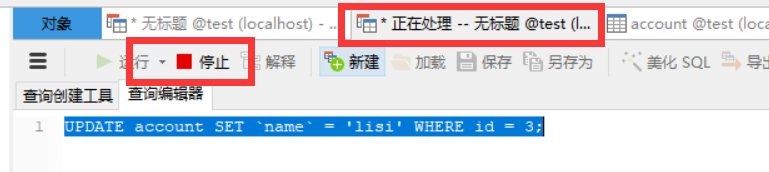
1. 在另一个会话中修改数据，但是id不一样，结果是可以修改成功，因为和锁住的数据的id不一样。

---------------------------------------------------------------------------------------------------------------------------------

1. 开启事务，执行一条修改语句，条件涉及的字段是没有索引的

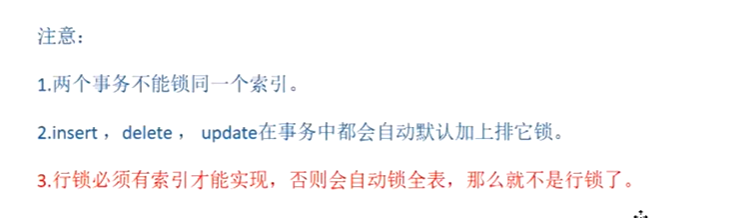


1. 在另一个会话中修改数据，与上个语句的条件不一样，结果:数据不一样，为啥还修改不成功呢？答案：在上一个会话中SQL语句涉及的字段是没有索引的，那么修改的时候会锁住所有的数据，所以这里修改的时候会等待上一个事务提交。

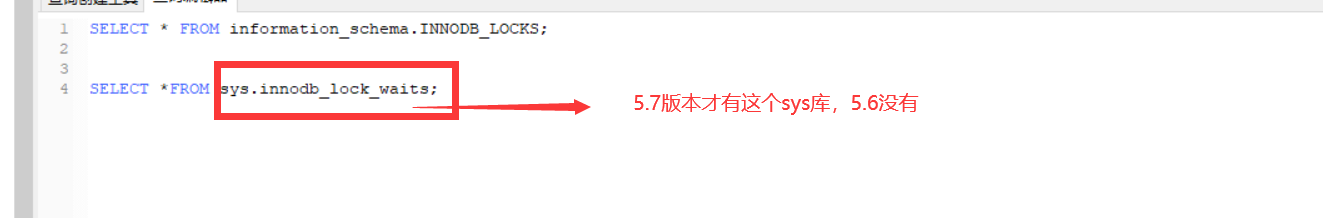


## 总结（行锁）

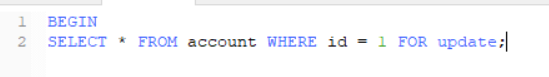
两条SQL语句，id一样，但是只能有一条才能修改成功，如果加上版本号，一条数据修改成功之后，版本号+1，另一条SQL语句再执行不会成功，因为版本号已经改变了，这就是乐观锁。

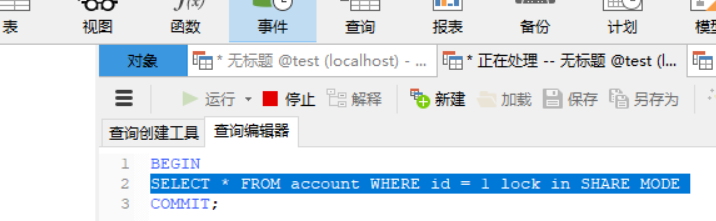


## 查询锁的信息



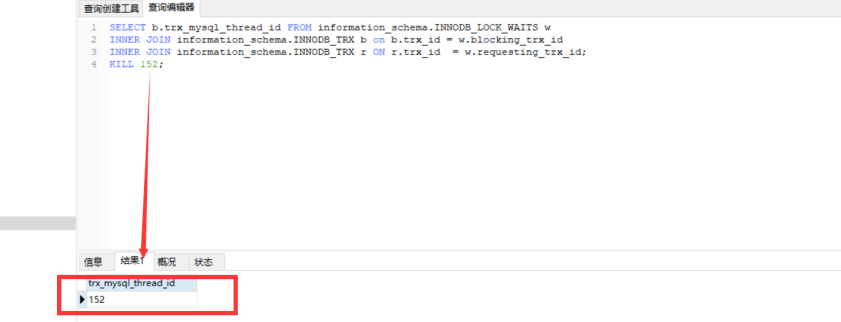
## 解决锁的问题，在开发过程中可能会遇到



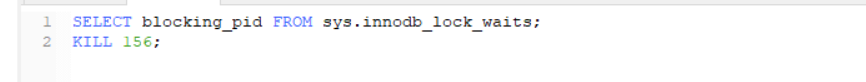


执行这个第二提交SQL语句，在等待中，解决方法如下(5.6版本):

|  |
| --- |
| SELECT b.trx\_mysql\_thread\_id FROM information\_schema.INNODB\_LOCK\_WAITS w  INNER JOIN information\_schema.INNODB\_TRX b on b.trx\_id = w.blocking\_trx\_id  INNER JOIN information\_schema.INNODB\_TRX r ON r.trx\_id = w.requesting\_trx\_id; |



### 5.7版本解决锁等待的问题



# 事务

## 什么是事务

事务是指是程序中一系列严密的逻辑操作，而且所有操作必须全部成功完成，否则在每个操作中所作的所有更改都会被撤消。可以通俗理解为：就是把多件事情当做一件事情来处理，好比大家同在一条船上，要活一起活，要完一起完 。

## 事务的特性

* **原子性**（Atomicity）**：**操作这些指令时，要么全部执行成功，要么全部不执行。只要其中一个指令执行失败，所有的指令都执行失败，数据进行回滚，回到执行指令前的数据状态。
* **一致性**（Consistency）**：**事务的执行使数据从一个状态转换为另一个状态，但是对于整个数据的完整性保持稳定。
* **隔离性**（Isolation）**：**隔离性是当多个用户并发访问数据库时，比如操作同一张表时，数据库为每一个用户开启的事务，不能被其他事务的操作所干扰，多个并发事务之间要相互隔离。
* **持久性**（Durability）**：**当事务正确完成后，它对于数据的改变是永久性的。

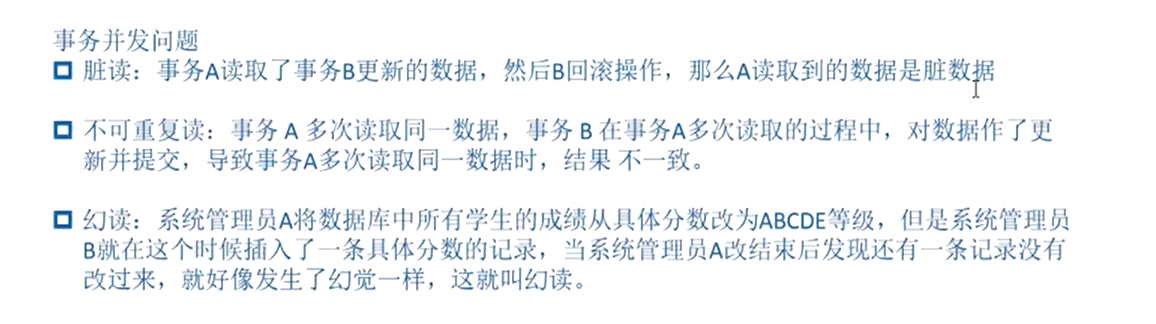
## 隔离性

未提交读(脏读)

已提交读(不可重复读)

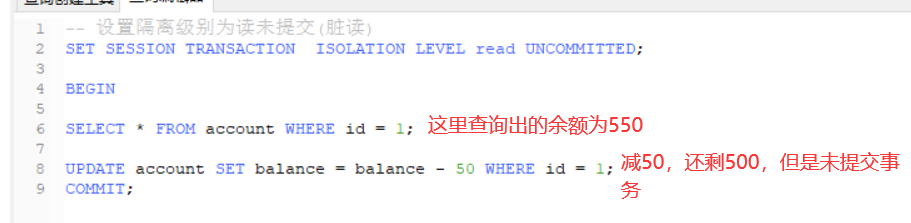
可重复读

序列化



## 通过案例来演示事务的隔离性

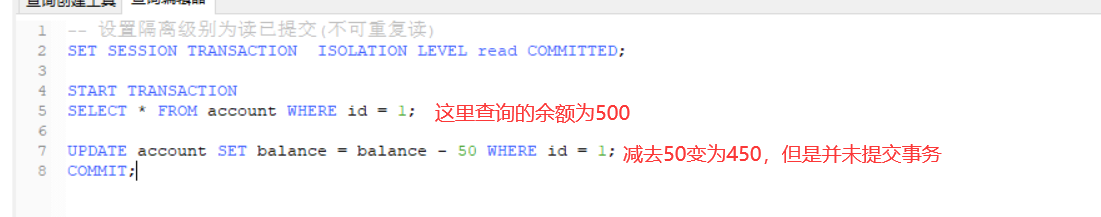
### 读未提交

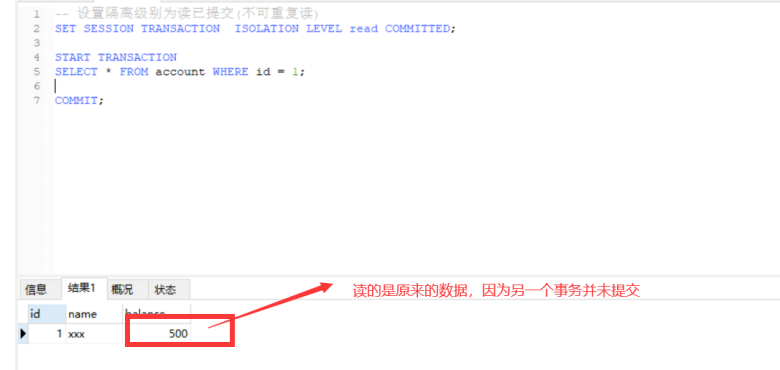




### 读已提交（不可重复读）

|  |
| --- |
| -- 设置隔离级别为读已提交(不可重复读)  SET SESSION TRANSACTION ISOLATION LEVEL read COMMITTED;  START TRANSACTION  SELECT \* FROM account WHERE id = 1;  UPDATE account SET balance = balance - 50 WHERE id = 1;  COMMIT; |



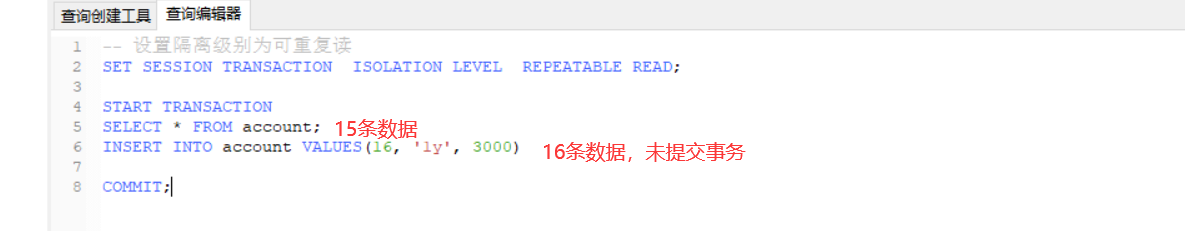


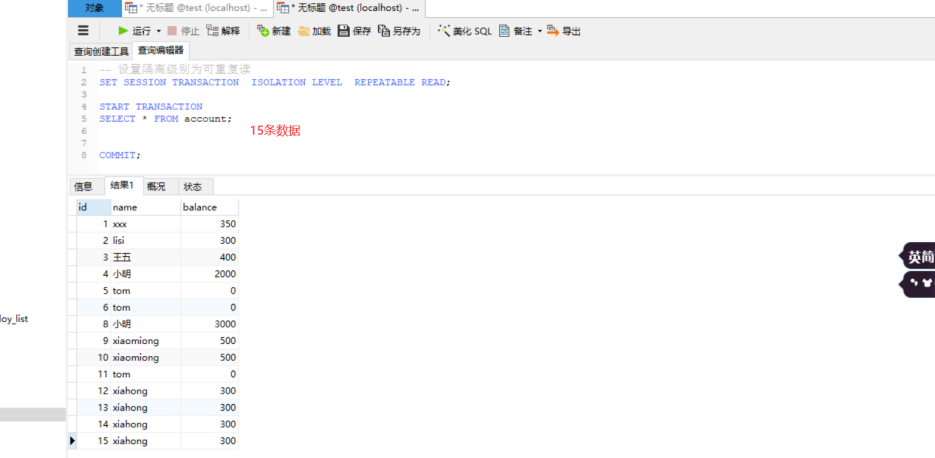
将第一个事务提交，余额变为450，第二个会话再去查询，结果也变为了450，第二个会话查询的两次的结果并不一样，这种情况叫做不可重复读。

### 可重复读

解决了不可重复读的问题

**存在的问题:幻读**



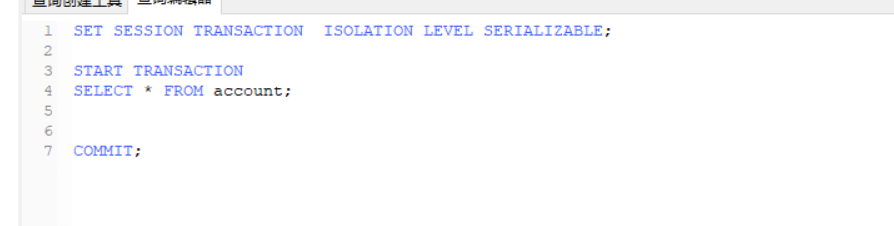


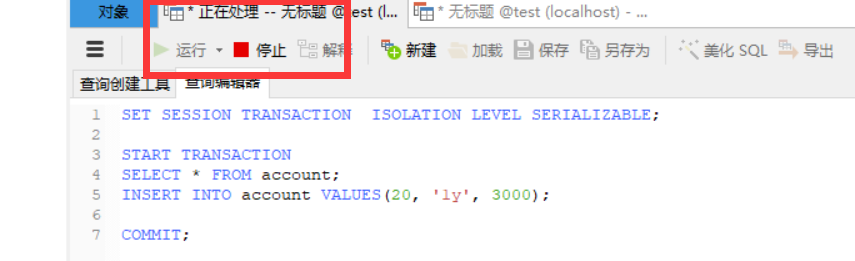
将第一个会话中的事务提交

再在第二会话中查询，结果如下:结果还是15条数据，然后再插入一条id为16的数据，却插入不进去，因为id=16的数据已经存在了，这种情况叫做幻读。

### 可序列化

事务串行执行，一个一个执行，a事务在执行，b执行必须等到a事务提交之后才能执行。

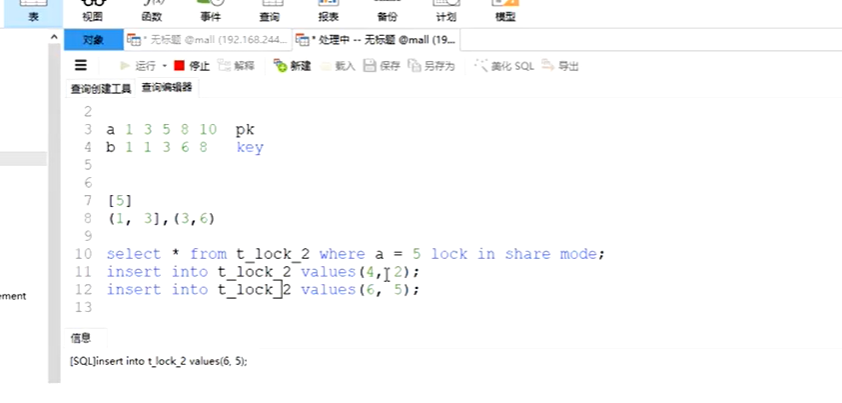




# 间隙锁

<https://www.cnblogs.com/crazylqy/p/7821481.html>

a为主键，b为普通的索引

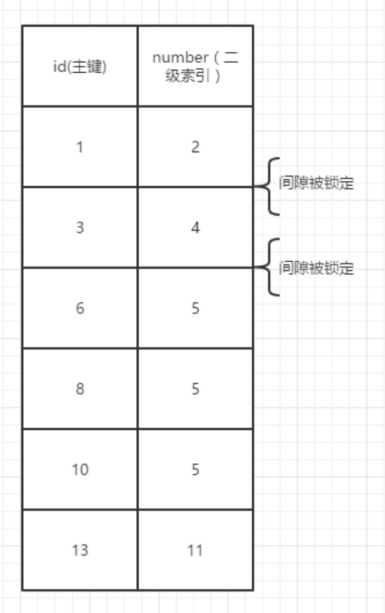


|  |
| --- |
| CREATE TABLE `grap` (  `id` int(11) NOT NULL,  `number` int(11) DEFAULT NULL,  PRIMARY KEY (`id`),  KEY `number` (`number`)  ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4; |

## 案例一:

|  |
| --- |
| session **1**:  start transaction ;select \* from news where number=**4** for update ;  session **2**:  start transaction ;  insert into news value(**2**,**4**);#（阻塞）  insert into news value(**2**,**2**);#（阻塞）  insert into news value(**4**,**4**);#（阻塞）  insert into news value(**4**,**5**);#（阻塞）  insert into news value(**7**,**5**);#（执行成功）  insert into news value(**9**,**5**);#（执行成功）  insert into news value(**11**,**5**);#（执行成功） |

**检索条件number=4,向左取得最靠近的值2作为左区间，向右取得最靠近的5作为右区间，因此，session 1的间隙锁的范围（2，4），（4，5），如下图所示：**



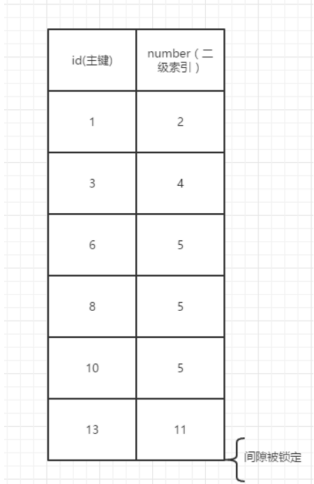
间隙锁锁定的区间为（2，4），（4，5），即记录（id=1,number=2）和记录（id=3,number=4）之间间隙会被锁定，记录（id=3,number=4）和记录（id=6,number=5）之间间隙被锁定。

因此记录（id=2,number=4），（id=2,number=2），（id=4,number=4），（id=4,number=5）正好处在（id=3,number=4）和（id=6,number=5）之间，所以插入不了，需要等待锁的释放，而记录(id=7,number=5)，（id=9,number=5），（id=11,number=5）不在上述锁定的范围内，因此都会插入成功。

## 案例二:

|  |
| --- |
| SESSION 1;  start transaction ;  select \* from grap where number=13 for update ;  SESSION 2;  start transaction ;  insert into grap value(11,5);#(执行成功)  insert into grap value(12,11);#(执行成功)  insert into grap value(14,11);#(阻塞)  insert into grap value(15,12);#(阻塞)  update grap set id=14 where number=11;#(阻塞)  update grap set id=11 where number=11;#(执行成功) |

检索条件number=13,向左取得最靠近的值11作为左区间，向右由于没有记录因此取得无穷大作为右区间，因此，session 1的间隙锁的范围（11，无穷大），如下图所示：



此表中没有number=13的记录的，innodb依然会为该记录左右两侧加间隙锁，间隙锁的范围（11，无穷大）。

有人会问，为啥update news set id=14 where number=11会阻塞，但是update news set id=11 where number=11却执行成功呢？

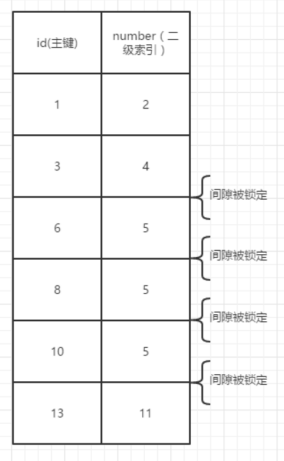
间隙锁采用在指定记录的前面和后面以及中间的间隙上加间隙锁的方式避免数据被插入，此图间隙锁锁定的区域是（11，无穷大），也就是记录（id=13,number=11）之后不能再插入记录，update news set id=14 where number=11这条语句如果执行的话，将会被插入到（id=13,number=11）的后面，也就是在区间（11，无穷大）之间，由于该区间被间隙锁锁定，所以只能阻塞等待，而update news set id=11 where number=11执行后是会被插入到（id=13,number=11）的记录前面，也就不在（11，无穷大）的范围内，所以无需等待，执行成功。

## 案例三

|  |
| --- |
| session 1:  start transaction ;  select \* from grap where number=5 for update;  session 2:  start transaction ;  insert into grap value(4,4);#(阻塞)  insert into grap value(4,5);#(阻塞)  insert into grap value(5,5);#(阻塞)  insert into grap value(7,11);#(阻塞)  insert into grap value(9,12);#(执行成功)  insert into grap value(12,11);#(阻塞)  update grap set number=5 where id=1;#(阻塞)  update grap set id=11 where number=11;#(阻塞)  update grap set id=2 where number=4 ;#（执行成功）  update grap set id=4 where number=4 ;#（阻塞）  ROLLBACK |

插入的时候先按照主键排序，再按照number排序

检索条件number=5,向左取得最靠近的值4作为左区间，向右取得11为右区间，因此，session 1的间隙锁的范围（4，5），（5，11），如下图所示：



有人会问，为啥insert into news value(9,12)会执行成功？间隙锁采用在指定记录的前面和后面以及中间的间隙上加间隙锁的方式避免数据被插入，（id=9,number=12）很明显在记录（13,11）的后面，因此不再锁定的间隙范围内。

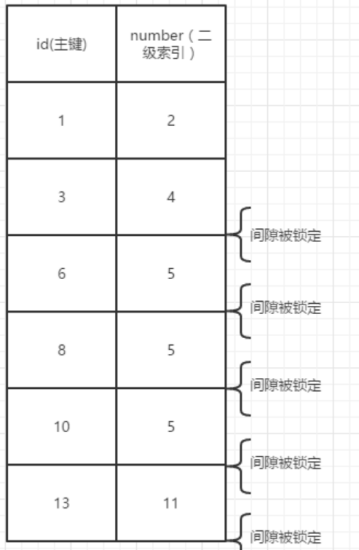
为啥update news set number=5 where id=1会阻塞？  
number=5的记录的前面，后面包括中间都被封锁了，你这个update news set number=5 where id=1根本没法执行，因为innodb已经把你可以存放的位置都锁定了，因为只能等待。

同理，update news set id=11 where number=11由于记录（id=10,number=5）与记录（id=13,number=11）中间的间隙被封锁了，你这句sql也没法执行，必须等待，因为存放的位置被封锁了。

## 案例四

|  |
| --- |
| session 1:  start transaction;  select \* from grap where number>4 for update;  session 2:  start transaction;  update grap set id=2 where number=4 ;#(执行成功)  update grap set id=4 where number=4 ;#(阻塞)  update grap set id=5 where number=5 ;#(阻塞)  insert into grap value(2,3);#(执行成功)  insert into grap value(null,13);#(阻塞)  ROLLBACK; |

检索条件number>4,向左取得最靠近的值4作为左区间，向右取无穷大，因此，session 1的间隙锁的范围（4，无穷大），如下图所示：

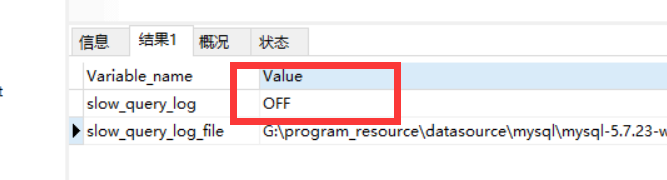


session2中之所以有些阻塞，有些执行成功，其实就是因为插入的区域被锁定，从而阻塞。

# 慢查询

show VARIABLES LIKE '%slow\_query\_log%'

默认是关闭的

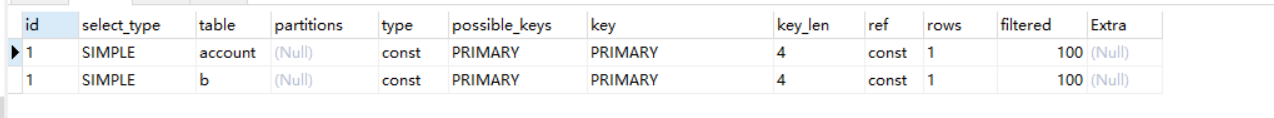


SET GLOBAL slow\_query\_log = 1;// 开启慢查询

SET GLOBAL long\_query\_time=0;// 查过0秒的查询都会记录慢查询日志文件中

可能修改查询时间并没有生效，只要将会话关闭再重新打开就生效了。

# 执行计划



## Id

1. Id相同，执行顺序由上到下
2. Id不同，id越大，优先级越高
3. Id相同不同

# Hash和B+tree

## 扇出

<https://blog.csdn.net/shenchaohao12321/article/details/83243314>

<https://blog.csdn.net/hao65103940/article/details/89032538>

索引的分裂和合并

https://blog.csdn.net/shenchaohao12321/article/details/83243314

## Hash

Hash 索引结构的特殊性，其检索效率非常高，索引的检索可以一次定位，不像B-Tree 索引需要从根节点到枝节点，最后才能访问到页节点这样多次的IO访问，所以 Hash 索引的查询效率要远高于 B-Tree 索引。

可能很多人又有疑问了，既然 Hash 索引的效率要比 B-Tree 高很多，为什么大家不都用 Hash 索引而还要使用 B-Tree 索引呢？任何事物都是有两面性的，Hash 索引也一样，虽然 Hash 索引效率高，但是 Hash 索引本身由于其特殊性也带来了很多限制和弊端，主要有以下这些。

（1）Hash 索引仅仅能满足"=","IN"和"<=>"查询，不能使用范围查询。

由于 Hash 索引比较的是进行 Hash 运算之后的 Hash 值，所以它只能用于等值的过滤，不能用于基于范围的过滤，因为经过相应的 Hash 算法处理之后的 Hash 值的大小关系，并不能保证和Hash运算前完全一样。

（2）Hash 索引无法被用来避免数据的排序操作。

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；

（3）Hash 索引不能利用部分索引键查询。

对于组合索引，Hash 索引在计算 Hash 值的时候是组合索引键合并后再一起计算 Hash 值，而不是单独计算 Hash 值，所以通过组合索引的前面一个或几个索引键进行查询的时候，Hash 索引也无法被利用。

（4）Hash 索引在任何时候都不能避免表扫描。

前面已经知道，Hash 索引是将索引键通过 Hash 运算之后，将 Hash运算结果的 Hash 值和所对应的行指针信息存放于一个 Hash 表中，由于不同索引键存在相同 Hash 值，所以即使取满足某个 Hash 键值的数据的记录条数，也无法从 Hash 索引中直接完成查询，还是要通过访问表中的实际数据进行相应的比较，并得到相应的结果。

1. 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。

## 设计索引的规则

1. 最适合索引的列是出现在where子句中的列，或连接子句中指定的列，而不是出现在select关键字后的选择列表中的列。
2. 索引的列的基数越大，索引的效果越好。
3. 尽量使用短索引。能够节省大量索引空间，也可能使查询更快。
4. 不要过度索引。索引都要占用额外的磁盘空间，并降低写操作的性能。在修改表的内容时，索引必须进行更新，有时可能需要重构。

## BTREE索引和HASH索引的比较

1、B-Tree索引

索引存储的值按索引列中的顺序排列。可以用B-Tree索引进行全关键字、关键字范围和关键字前缀查询。如果使用索引，必须保证按索引最左边前缀进行查询。由于B树中节点是顺序存储的，可以对查询结果进行order by。

限制：

1)查询必须从索引的最左边的列开始

2)不能跳过某一索引列。

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，所以dob是不能使用索引的。

1. Hash索引

MySQL中只有Memory存储引擎显示支持hash索引，是默认索引类型，它也支持B-Tree索引。

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。

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，

int列的索引和长字符串列的索引一样大。

1. 总结
2. HASH索引适合等式比较的操作，不能用来加速order by操作，也不能确定在两个值之间大约有多少行，会影响一些查询的执行效率。而且只能使用整个关键字来搜索一行。
3. BTREE索引，使用大于，小于，BETWEEN，不等于，LIKE等操作符的时候都可以用。对索引字段进行范围查询的时候，只有BTREE索引可以通过索引访问。HASH索引实际上是全表扫描的。

**如果对多列进行索引，列的顺序非常重要，MySQL仅能对索引最左边的前缀进行有效查找。**

假设存在组合索引it1c1c2(c1,c2)，查询语句select \* from t1 where c1=1 and c2=2能够

使用该索引。查询语句select \* from t1 where c1=1也能够使用该索引。

但是，查询语句select \* from t1 where c2=2不能够使用该索引，

因为没有组合索引的引导列，即，要想使用c2列进行查找，必需出现c1等于某值。

索引是在存储引擎中实现的，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。

## 高性能索引策略

### 聚簇索引

https://baijiahao.baidu.com/s?id=1645514817836645220&wfr=spider&for=pc&qq-pf-to=pcqq.group

<https://www.cnblogs.com/jiawen010/p/11805241.html>

将数据存储与索引放到了一块，找到索引也就找到了数据

非聚簇索引:将数据存储于索引分开结构，索引结构的叶子节点指向了数据的对应行，myisam通过key\_buffer把索引先缓存到内存中，当需要访问数据时（通过索引访问数据），在内存中直接搜索索引，然后通过索引找到磁盘相应数据，这也就是为什么索引不在key buffer命中时，速度慢的原因.

聚簇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簇索引，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簇索引。

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簇索引。

目前，只有solidDB和InnoDB支持。

### 覆盖索引

1）索引不会包含有null值得列

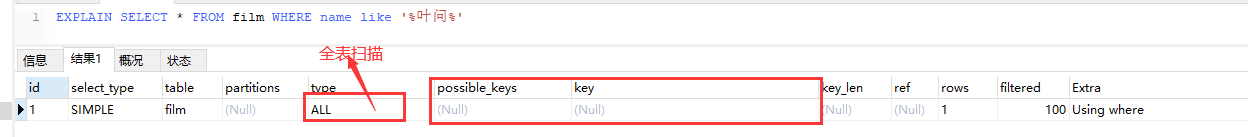
2）尽量使用短索引，对串列进行索引，如果可能应该指定一个前缀长度。短索引不仅可以提高查询速度而且可以节省磁盘空间和IO操作

3）like语句操作：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

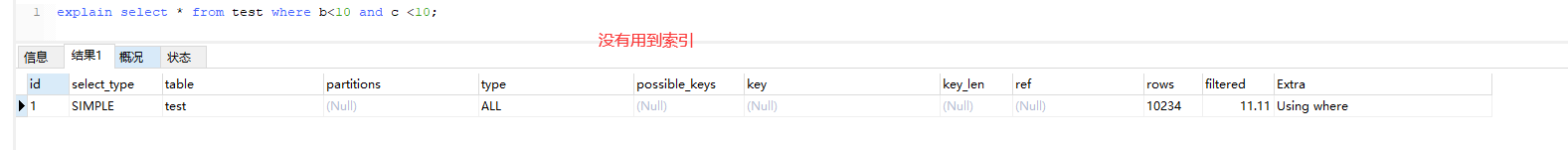
# 什么情况下索引会失效

1. like ‘%.....’，将%放到最前面，索引会失效.

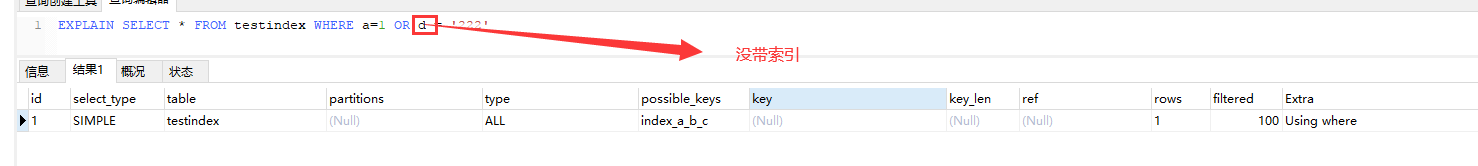


1. 不遵循最左原则

|  |
| --- |
| create table test( a int , b int, c int, d int, key index\_abc(a,b,c) )engine=InnoDB default charset=utf8; |
| DROP PROCEDURE IF EXISTS proc\_initData;  DELIMITER $  CREATE PROCEDURE proc\_initData()  BEGIN  DECLARE i INT DEFAULT 1;  WHILE i<=10000 DO  INSERT INTO test(a,b,c,d) VALUES(i,i,i,i);  SET i = i+1;  END WHILE;  END $  CALL proc\_initData(); |



1. 如果条件中有 or， 有条件没有使用索引，即使其中有条件带索引也不会使用，换言之， 就是要求使用的所有字段，都必须单独使用时能使用索引。



(4)如果列类型是字符串，那么一定要在条件中使用引号引用起来，否则不使用索引。



1. 如果mysql认为全表扫面要比使用索引快，则不使用索引。

# 十大优化策略

1. 尽量全值匹配
2. 最左前缀法则
3. 不再索引列上做任何操作
4. 范围条件放最后
5. 覆盖索引尽量用
6. 不等于要慎用
7. Null/not有影响
8. Like查询要当心
9. 字符类型要加引号
10. OR该Union效率更高

# 一个6亿的表a，一个3亿的表b，通过外间tid关联，你如何最快的查询出满足条件的第50000到第50200中的这200条数据记录。

1、如果A表TID是自增长,并且是连续的,B表的ID为索引

select \* from a,b where a.tid = b.id and a.tid>500000 limit 200;

2、如果A表的TID不是连续的,那么就需要使用覆盖索引.TID要么是主键,要么是辅助索引,B表ID也需要有索引。

select \* from b , (select tid from a limit 50000,200) a where b.id = a .tid;

# 左连接，右连接，内连接，笛卡尔积

内连接查出相同的数据

## 左外连接

左外连接是A表的所有行匹配上B表得出的结果集。

## 右外连接

右外连接是B表的所有行匹配上A表得出的结果集。

## 一个表的主键为自增主键，现在共有10条数据，最大的id为10，删掉一条，在增加一条，此时的id为多少

如果索引是存到内存中，那么就是10，如果是存到磁盘中的，那么就是11。