MYSQL-性能优化篇

# 为什么要进行数据库优化？

1. 避免网站页面出现访问错误

由于数据库连接timeout产生页面5xx错误

由于慢查询造成页面无法加载

由于阻塞造成数据无法提交

1. 增加数据库的稳定性

很多数据库问题都是由于低效的查询引起的

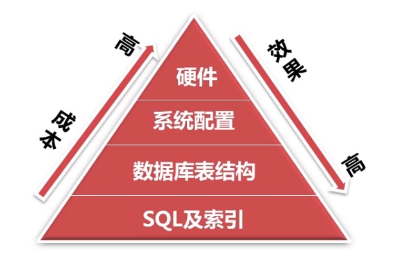
1. 优化用户体验

流畅页面的访问速度

良好的网站功能体验

# 2、mysql数据库优化

可以从哪几个方面进行数据库的优化？如下图所示：



1. SQL及索引优化

根据需求写出良好的SQL，并创建有效的索引，实现某一种需求可以多种写法，这时候我们就要选择一种效率最高的写法。这个时候就要了解sql优化

1. 数据库表结构优化

根据数据库的范式，设计表结构，表结构设计的好直接关系到写SQL语句。

1. 系统配置优化

大多数运行在Linux机器上，如tcp连接数的限制、打开文件数的限制、安全性的限制，因此我们要对这些配置进行相应的优化。

1. 硬件配置优化

选择适合数据库服务的cpu，更快的IO，更高的内存；cpu并不是越多越好，某些数据库版本有最大的限制，ＩＯ操作并不是减少阻塞。

注：通过上图可以看出，该金字塔中，优化的成本从下而上逐渐增高，而优化的效果会逐渐降低。

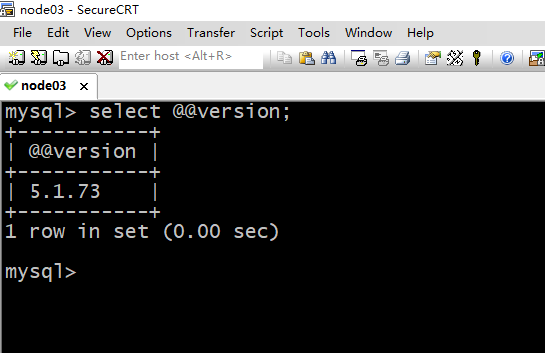
# ３、SQL及索引优化

## 1、mysql安装与卸载（linux在线安装与卸载）

## 2、数据库版本选择

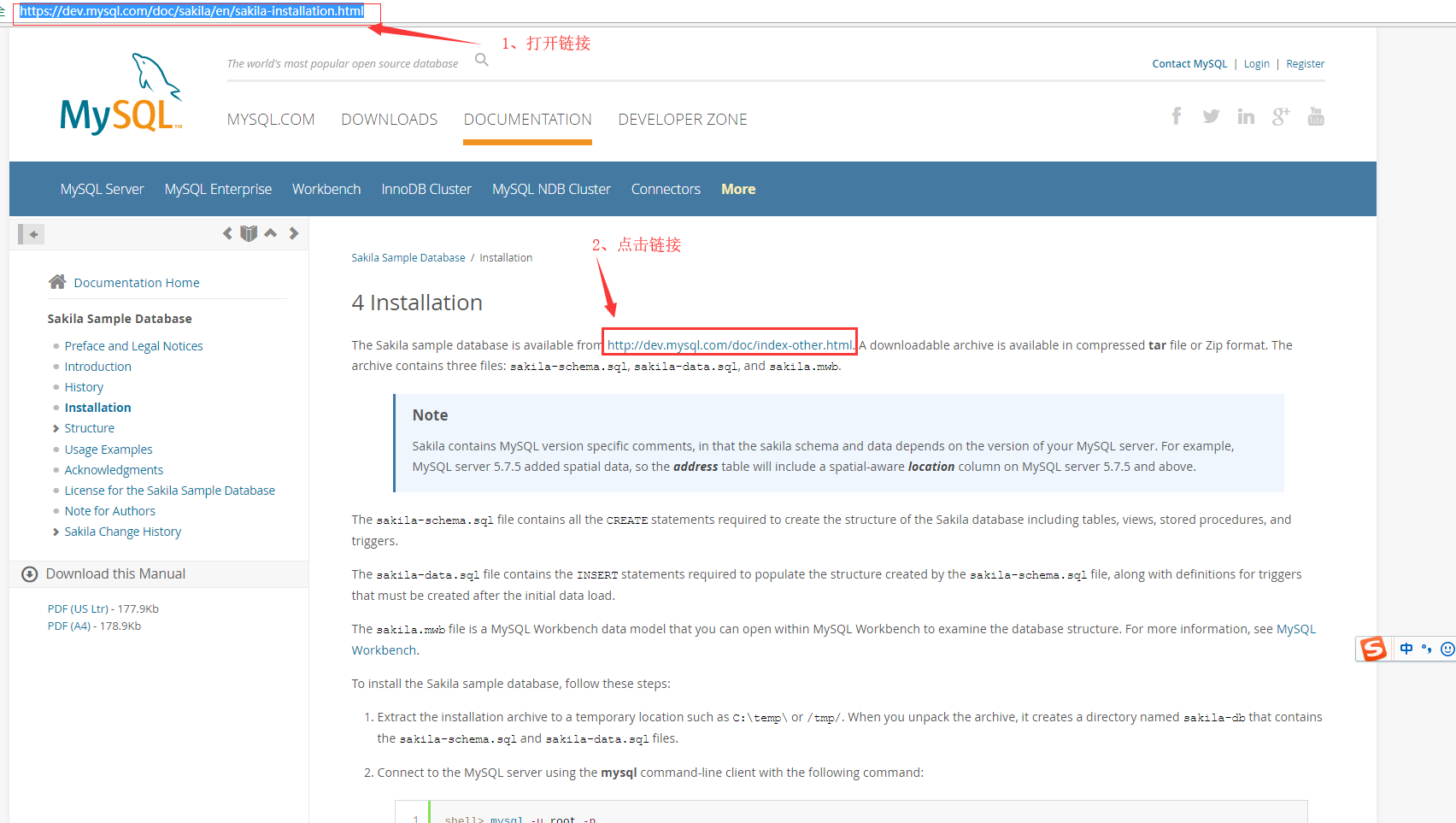
1、查看数据库的版本

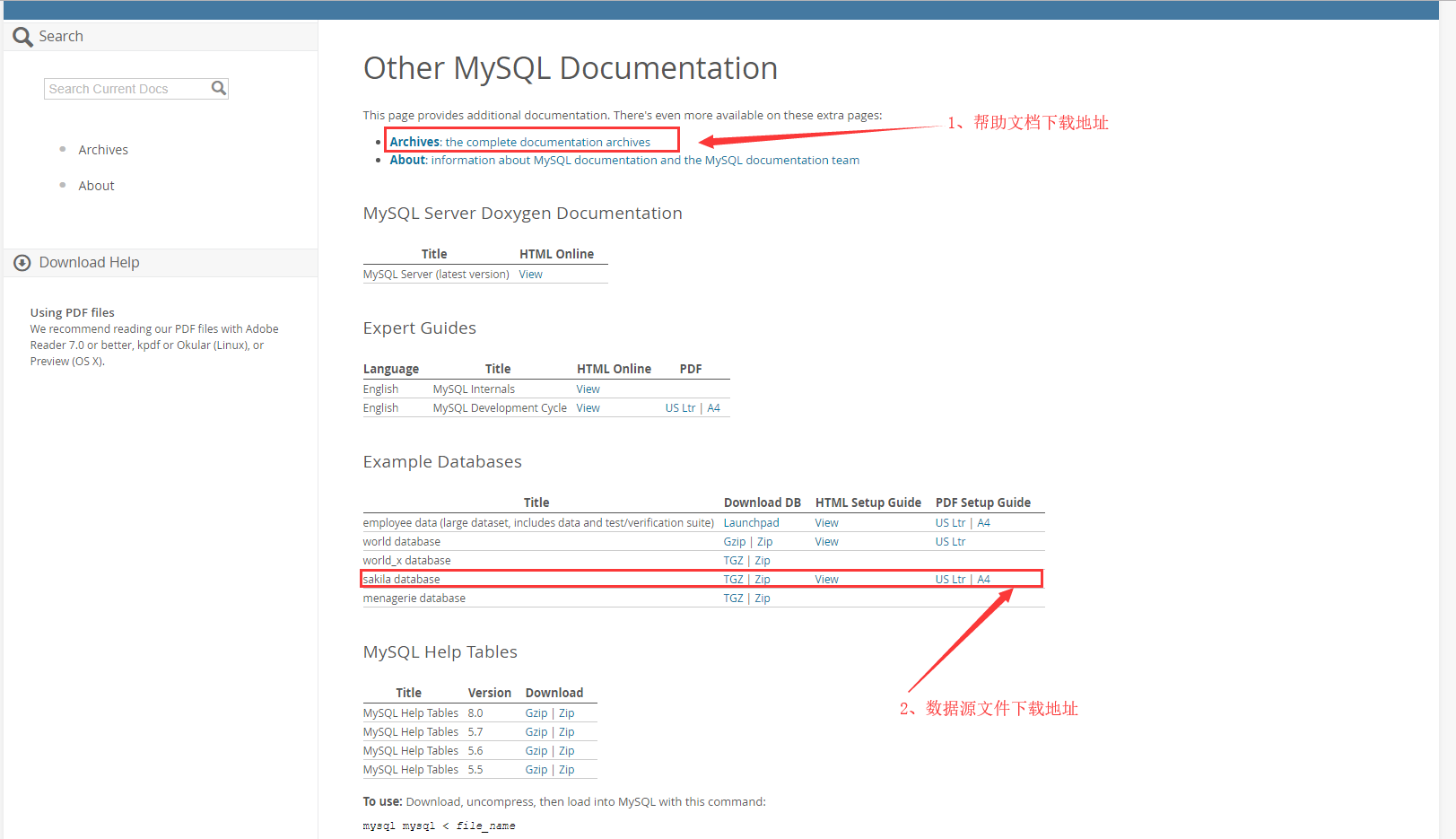
|  |
| --- |
| select @@version; |

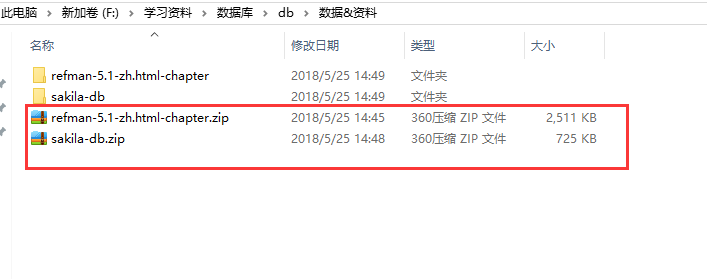


## 2、准备数据

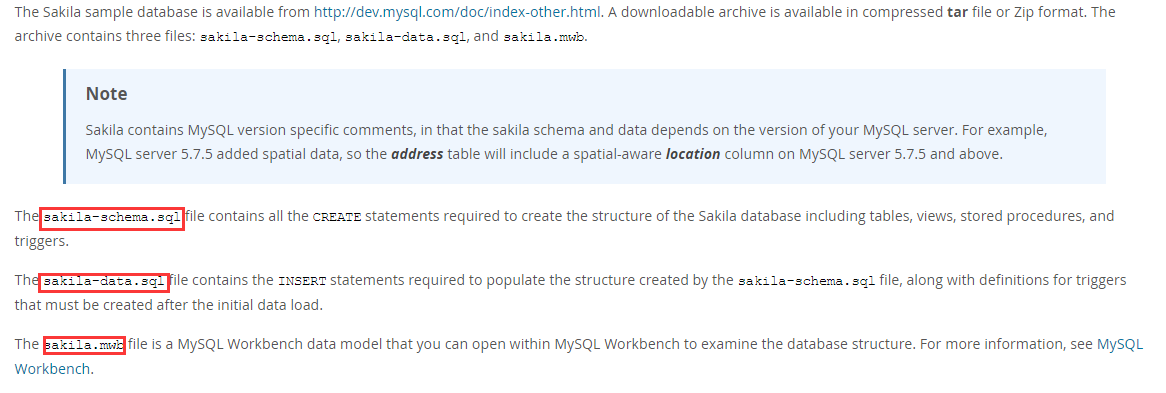
**网址：https://dev.mysql.com/doc/sakila/en/sakila-installation.html**





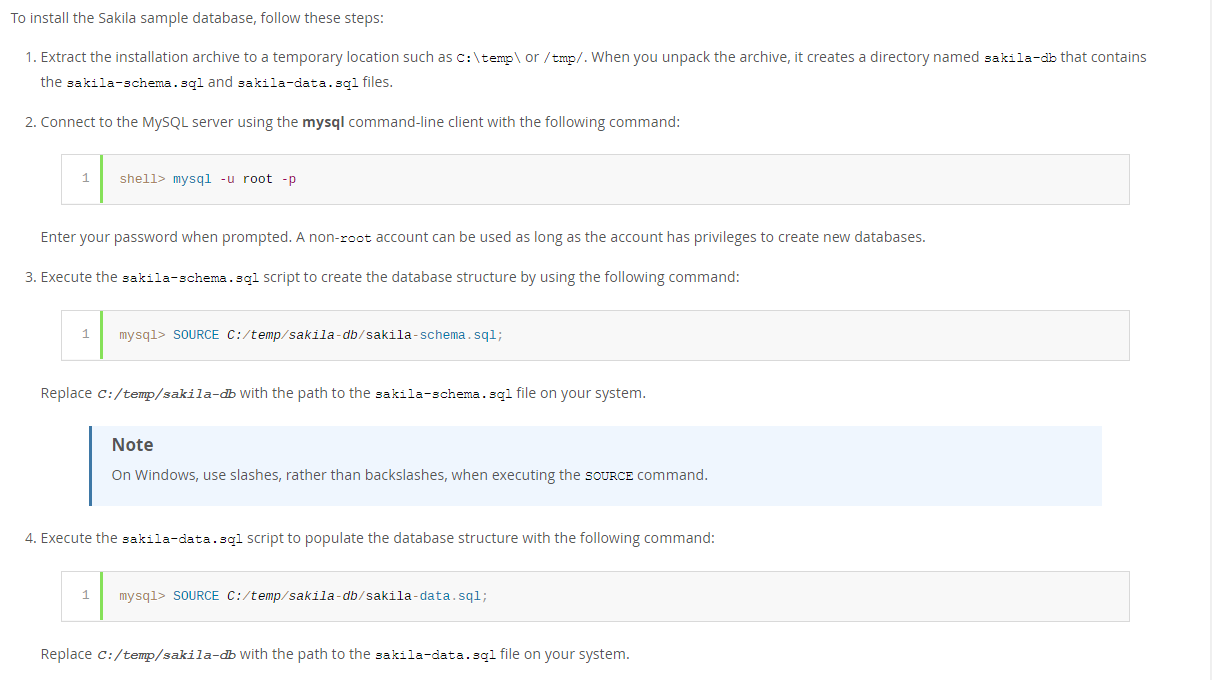


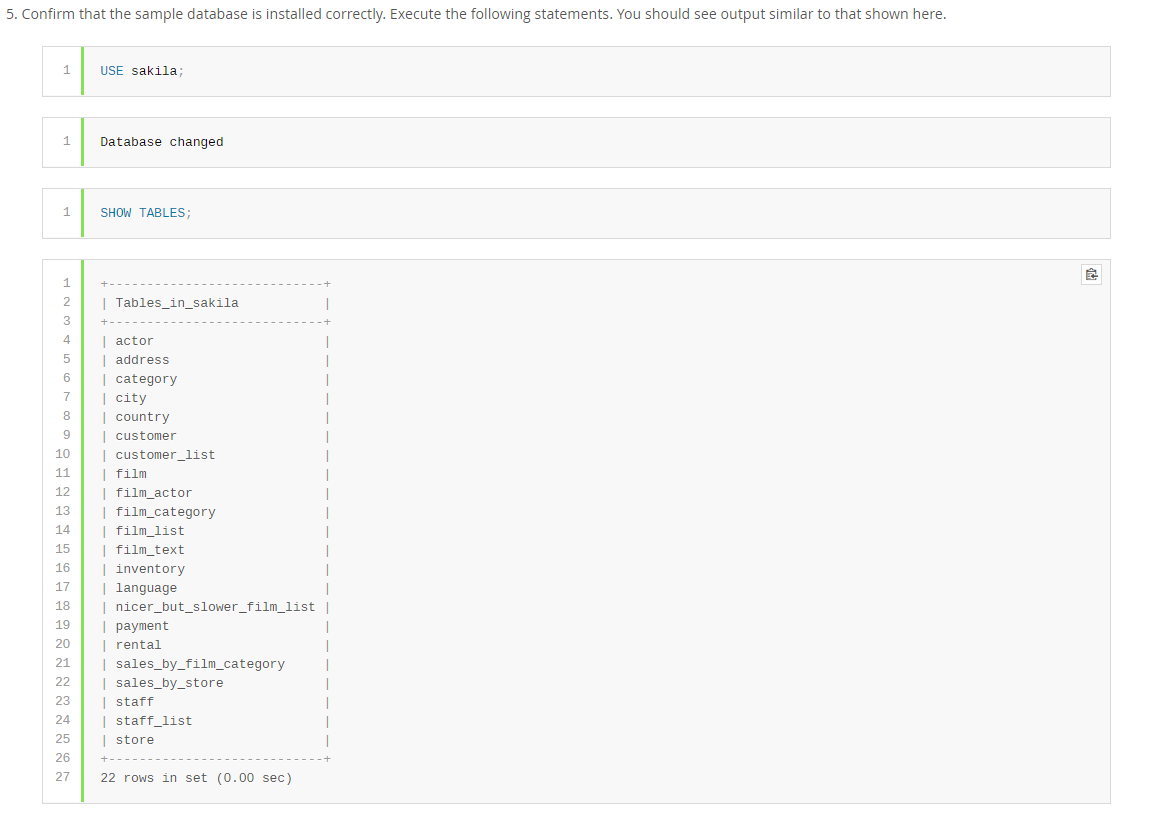
sakila-db.zip压缩包所包含的文件如下解释



加载数据

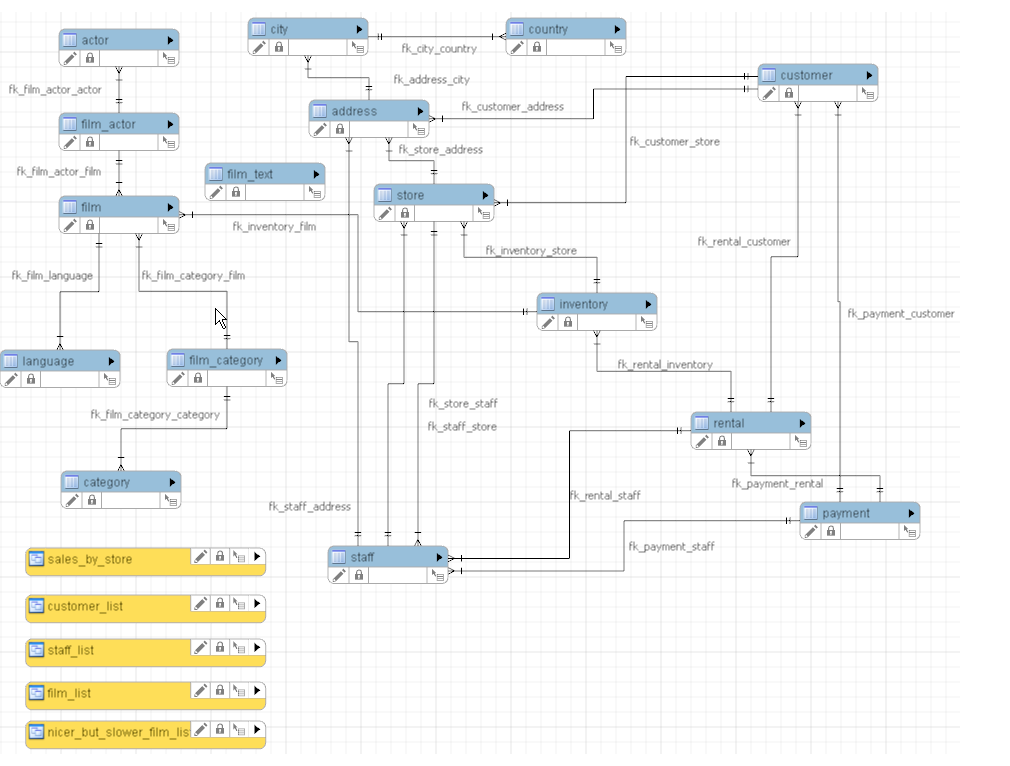
步骤如下图所示







## 3、表结构关系



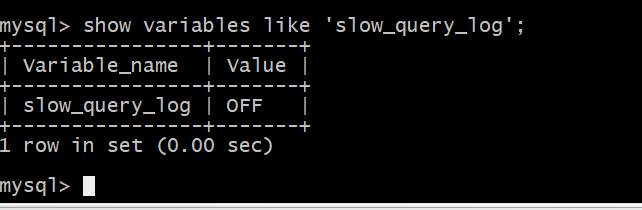
注：该表结构关系是用工具生成的。

## 如何发现有问题的SQL

MySQL慢查日志的开启方式和存储格式

### 1、检查慢查日志是否开启：

|  |
| --- |
| show variables like 'slow\_query\_log' |



|  |
| --- |
| show variables like 'slow\_query\_log'  //查看是否开启慢查询日志  set global slow\_query\_log\_file=' /usr/share/mysql/sql\_log/mysql-slow.log'  //慢查询日志的位置  set global log\_queries\_not\_using\_indexes=on;  //开启慢查询日志  set global long\_query\_time=1;  //大于1秒钟的数据记录到慢日志中，如果设置为默认0，则会有大量的信息存储在磁盘中，磁盘很容易满掉 |

### 2、查看所有日志的变量信息

|  |
| --- |
| show variables like '%log%' |

mysql> show variables like '%log%';

+-----------------------------------------+------------------------------------+

| Variable\_name | Value |

+-----------------------------------------+------------------------------------+

| back\_log | 80 |

| binlog\_cache\_size | 32768 |

| binlog\_checksum | CRC32 |

| binlog\_direct\_non\_transactional\_updates | OFF |

| binlog\_error\_action | IGNORE\_ERROR |

| binlog\_format | STATEMENT |

| binlog\_gtid\_simple\_recovery | OFF |

| binlog\_max\_flush\_queue\_time | 0 |

| binlog\_order\_commits | ON |

| binlog\_row\_image | FULL |

| binlog\_rows\_query\_log\_events | OFF |

| binlog\_stmt\_cache\_size | 32768 |

| binlogging\_impossible\_mode | IGNORE\_ERROR |

| expire\_logs\_days | 0 |

| general\_log | OFF |

| general\_log\_file | /var/lib/mysql/mysql-host.log |

| innodb\_api\_enable\_binlog | OFF |

| innodb\_flush\_log\_at\_timeout | 1 |

| innodb\_flush\_log\_at\_trx\_commit | 1 |

| innodb\_locks\_unsafe\_for\_binlog | OFF |

| innodb\_log\_buffer\_size | 8388608 |

| innodb\_log\_compressed\_pages | ON |

| innodb\_log\_file\_size | 50331648 |

| innodb\_log\_files\_in\_group | 2 |

| innodb\_log\_group\_home\_dir | ./ |

| innodb\_mirrored\_log\_groups | 1 |

| innodb\_online\_alter\_log\_max\_size | 134217728 |

| innodb\_undo\_logs | 128 |

| log\_bin | OFF |

| log\_bin\_basename | |

| log\_bin\_index | |

| log\_bin\_trust\_function\_creators | OFF |

| log\_bin\_use\_v1\_row\_events | OFF |

| log\_error | /var/log/mysqld.log |

| log\_output | FILE |

| log\_queries\_not\_using\_indexes | ON |

| log\_slave\_updates | OFF |

| log\_slow\_admin\_statements | OFF |

| log\_slow\_slave\_statements | OFF |

| log\_throttle\_queries\_not\_using\_indexes | 0 |

| log\_warnings | 1 |

| max\_binlog\_cache\_size | 18446744073709547520 |

| max\_binlog\_size | 1073741824 |

| max\_binlog\_stmt\_cache\_size | 18446744073709547520 |

| max\_relay\_log\_size | 0 |

| relay\_log | |

| relay\_log\_basename | |

| relay\_log\_index | |

| relay\_log\_info\_file | relay-log.info |

| relay\_log\_info\_repository | FILE |

| relay\_log\_purge | ON |

| relay\_log\_recovery | OFF |

| relay\_log\_space\_limit | 0 |

| simplified\_binlog\_gtid\_recovery | OFF |

| slow\_query\_log | OFF |

| slow\_query\_log\_file | /var/lib/mysql/mysql-host-slow.log |

| sql\_log\_bin | ON |

| sql\_log\_off | OFF |

| sync\_binlog | 0 |

| sync\_relay\_log | 10000 |

| sync\_relay\_log\_info | 10000 |

+-----------------------------------------+------------------------------------+

61 rows in set (0.01 sec)

开启慢查日志：

|  |
| --- |
| show variables like 'slow\_query\_log'  //查看是否开启慢查询日志  set global slow\_query\_log\_file=' /var/lib/mysql/mysql-host-slow.log '  //慢查询日志的位置  set global log\_queries\_not\_using\_indexes=on;  //开启慢查询日志  set global long\_query\_time=1;  //大于1秒钟的数据记录到慢日志中，如果设置为默认0，则会有大量的信息存储在磁盘中，磁盘很容易满掉 |

验证慢查询日志是否开启：

在mysql操作中，

Show databases;

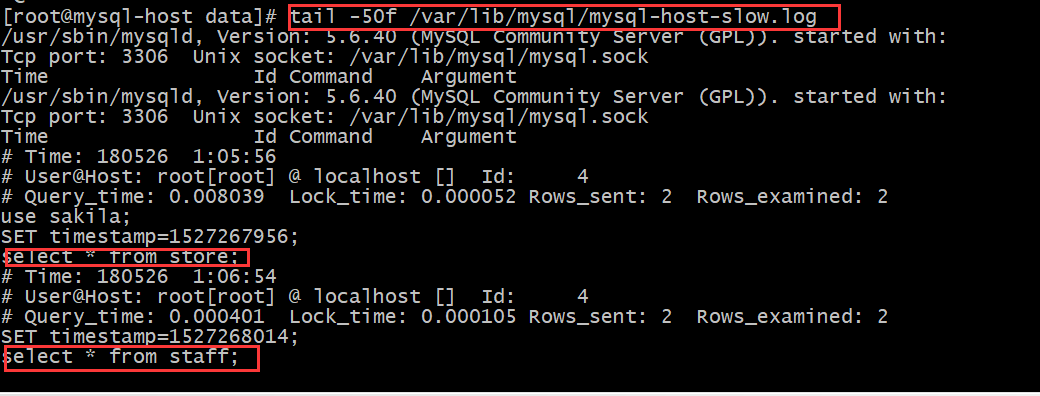
Use sakila;

select \* from store;

select \* from staff;

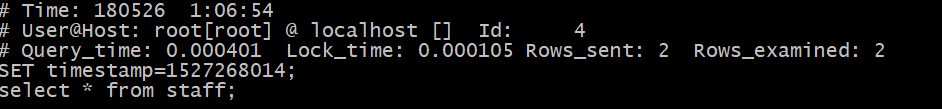
监听日志文件，看是否写入

tail -50f /var/lib/mysql/mysql-host-slow.log



### 3、MySQL慢查日志的存储格式

如下图所示：



说明：

1、# Time: 180526 1:06:54 -------🡪查询的执行时间

2、# User@Host: root[root] @ localhost [] Id: 4 -------🡪执行sql的主机信息

3、# Query\_time: 0.000401 Lock\_time: 0.000105 Rows\_sent: 2 Rows\_examined: 2-------🡪SQL的执行信息：

Query\_time：SQL的查询时间

Lock\_time：锁定时间

Rows\_sent：所发送的行数

Rows\_examined：锁扫描的行数

4、SET timestamp=1527268014; -------🡪SQL执行时间

5、select \* from staff; -------🡪SQL的执行内容

# 4、MySQL慢查日志分析工具（mysqldumpslow）

## 1、介绍

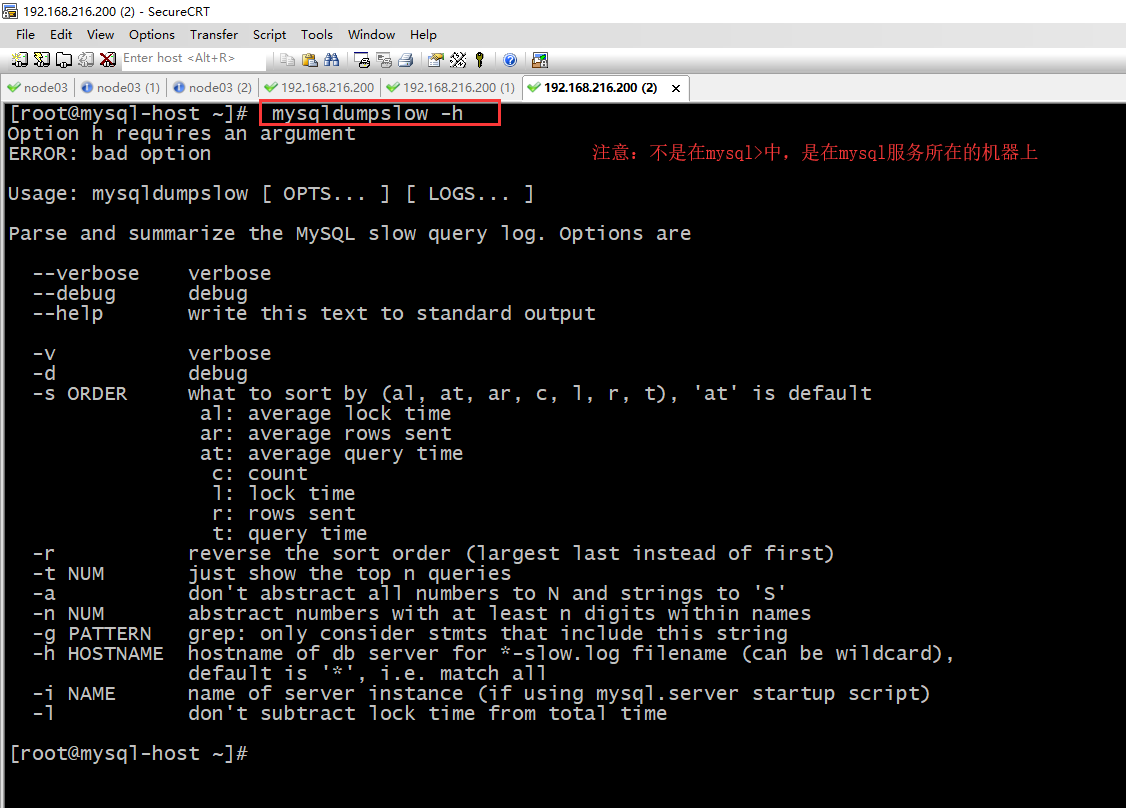
如何进行查看慢查询日志，如果开启了慢查询日志，就会生成很多的数据，然后我们就可以通过对日志的分析，生成分析报表，然后通过报表进行优化。

## 2、用法

接下来我们查看一下这个工具的用法：

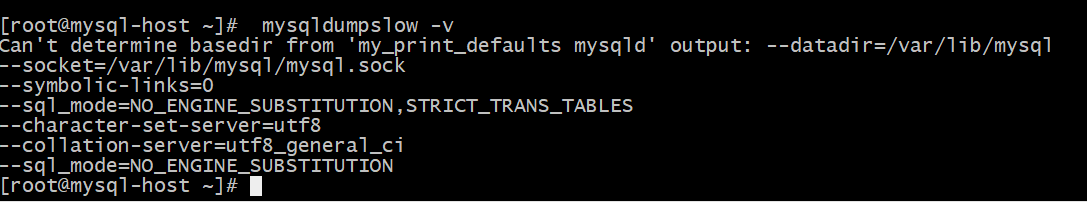
注意：在mysql数据库所在的服务器上，而不是在mysql>命令行中

该工具如何使用：**mysqldumpslow -h**

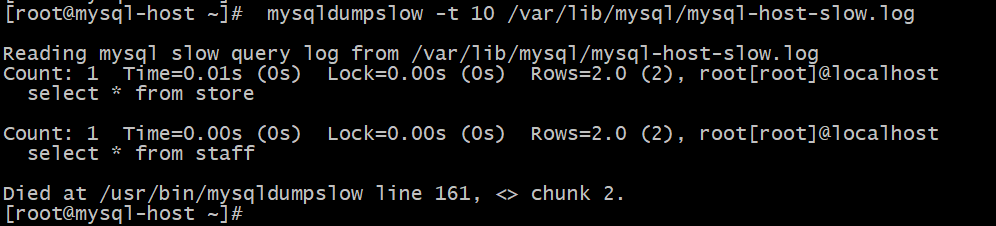


**查看verbose信息**

**Mysqldumpslow -v**



查看慢查询日志的前10个，mysqldumpslow 分析的结果如下

**mysqldumpslow -t 10 /var/lib/mysql/mysql-slow.log**

如上图两条就是分析的结果，每条结果都显示是执行时间，锁定时间，发送的行数，扫描的行数

这个工具是最常用的工具，通过安装mysql进行附带安装，但是该工具统计的结果比较少，对我们的优化锁表现的数据还是比较少。

# 5、MySQL慢查日志分析工具(pt-query-digest)

## 1、介绍及作用

作为一名优秀的mysql dba也需要有掌握几个好用的mysql管理工具，所以我也一直在整理和查找一些能够便于管理mysql的利器。以后的一段时间内，将会花一大部分的精力去搜索这些工具。

性 能的管理一直都是摆在第一位的，dba的很多工作管理层都看不到也没有办法衡量价值，但是如果一个系统慢的跟蜗牛一样，dba通过监控调优把系统从崩溃边 缘重新拉回到高铁时代。这种价值和触动应该是巨大的。（很多企业的领导认为系统跑不动了就需要换更快的CPU、更大的内存、更快的存储，而且这还不是少 数，所以DBA的价值也一直体现不出来，薪水自然也就不会很高）

mysql 的日志是跟踪mysql性能瓶颈的最快和最直接的方式了，系统性能出现瓶颈的时候，

①首先要打开慢查询日志，进行跟踪；

这段时间关于慢查询日志的管理和查看 已经整理过两篇文章了，不经意间又发现了一个查看慢查询日志的工具：mk-query-digest，这个工具网上号称mysql dba必须掌握的十大工具之首。

## 2、安装pt-query-digest工具

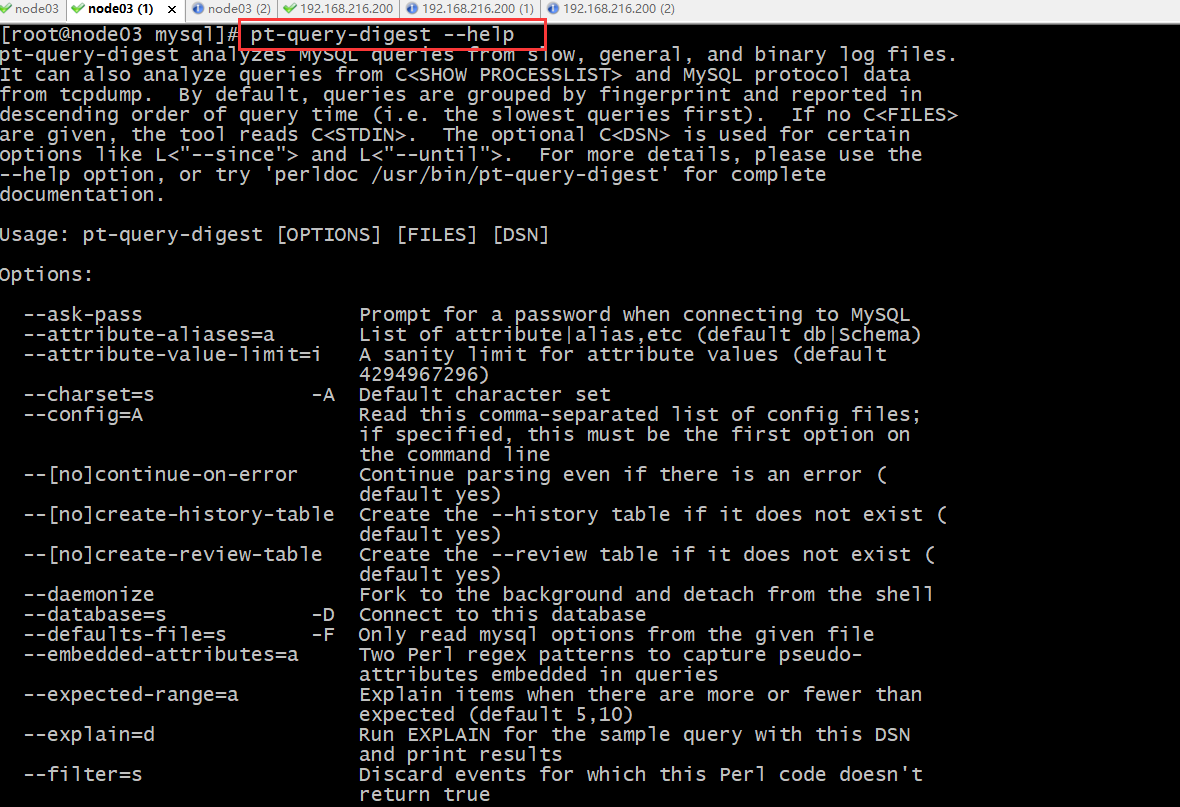
### 1.1、快速安装（注：必须先要安装wget）

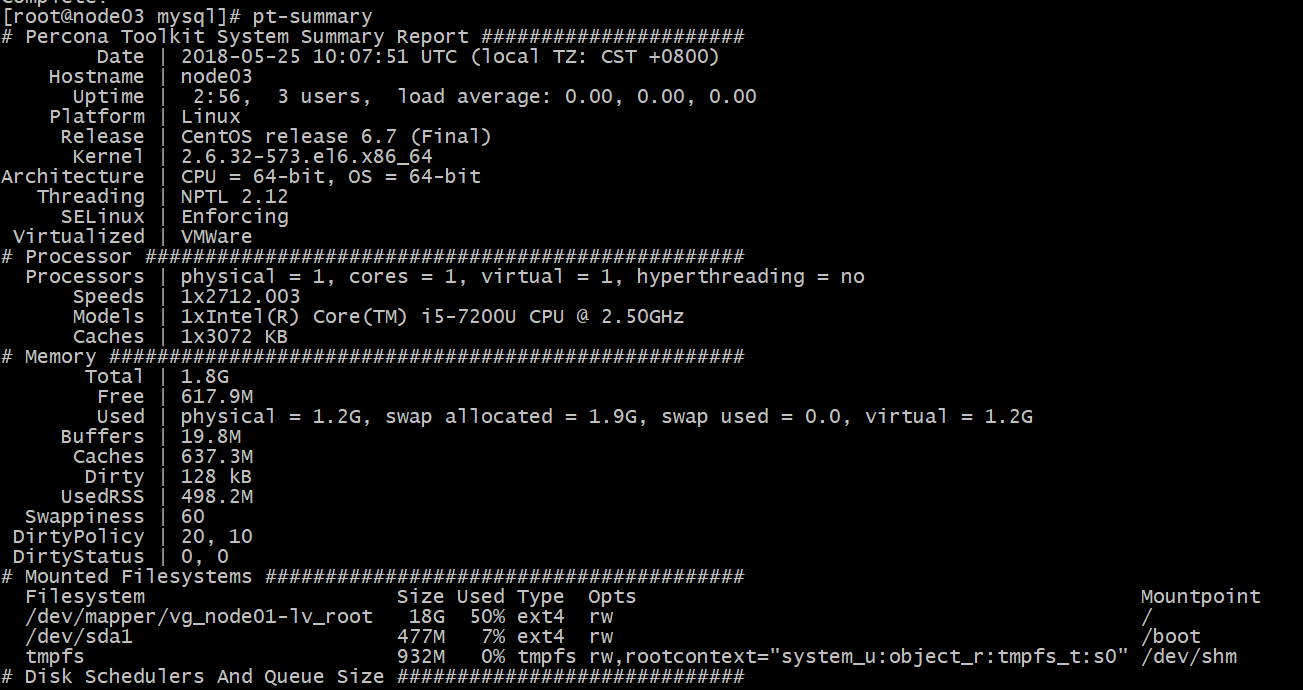
|  |
| --- |
| wget https://www.percona.com/downloads/percona-toolkit/2.2.16/RPM/percona-toolkit-2.2.16-1.noarch.rpm && yum localinstall -y  percona-toolkit-2.2.16-1.noarch.rpm |

### 1.2、检查是否安装完成：

命令行中输入：pt-summary

显示如下图所示：说明安装成功！输入【[root@node03 mysql]# pt-query-digest --help】





### 1.3、工具使用简介：

pt-summary –help

wget http://percona.com/get/pt-summary

#### 1、查看服务器信息

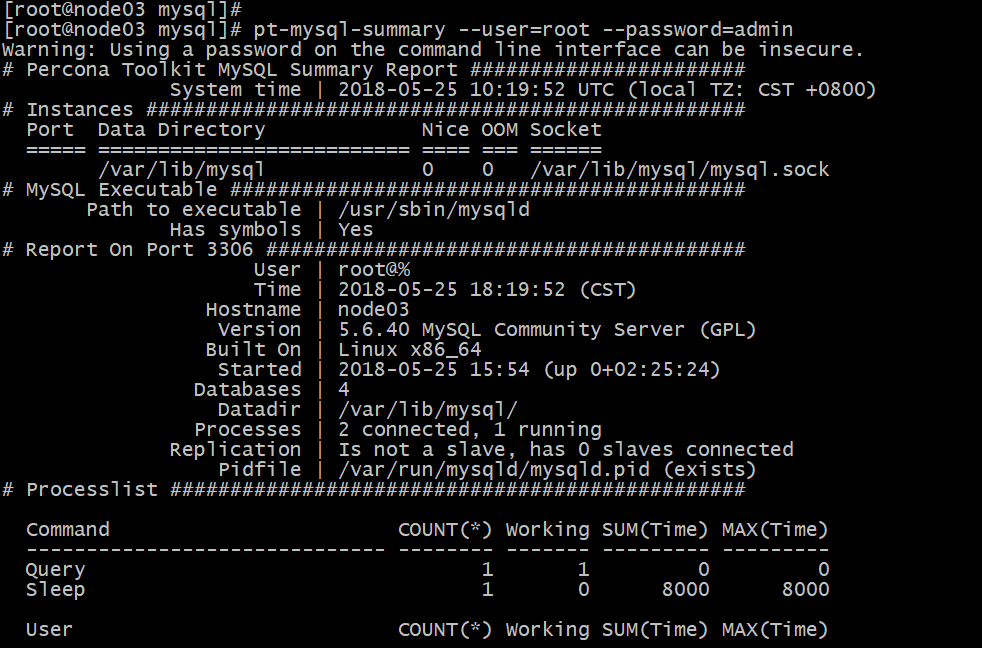
|  |
| --- |
| 命令：pt-summary |

#### 2、查看磁盘开销使用信息

|  |
| --- |
| 命令：pt-diskstats |

#### 3、查看mysql数据库信息

|  |
| --- |
| 命令：pt-mysql-summary --user=root --password=123456 |



#### 4、分析慢查询日志

|  |
| --- |
| 命令：pt-query-digest /data/mysql/data/db-3-12-slow.log |

#### 5、查找mysql的从库和同步状态

|  |
| --- |
| 命令：pt-slave-find --host=localhost --user=root --password=123456 |

#### 6、查看mysql的死锁信息

|  |
| --- |
| pt-deadlock-logger --user=root --password=123456 localhost |

#### 7、从慢查询日志中分析索引使用情况

|  |
| --- |
| pt-index-usage slow\_20131009.log |

#### 8、查找数据库表中重复的索引

|  |
| --- |
| pt-duplicate-key-checker --host=localhost --user=root --password=123456 |

#### 9、查看mysql表和文件的当前活动IO开销

|  |
| --- |
| pt-ioprofile |

#### 10、查看不同mysql配置文件的差异

|  |
| --- |
| pt-config-diff /etc/my.cnf /etc/my\_master.cnf |

#### 11、pt-find查找mysql表和执行命令，示例如下

查找数据库里大于2G的表：

|  |
| --- |
| pt-find --user=root --password=123456 --tablesize +2G |

查找10天前创建，MyISAM引擎的表：

|  |
| --- |
| pt-find --user=root --password=123456 --ctime +10 --engine MyISAM |

查看表和索引大小并排序

|  |
| --- |
| pt-find --user=root --password=123456 --printf "%T\t%D.%N\n" | sort -rn |

#### 12、pt-kill 杀掉符合标准的mysql进程

显示查询时间大于60秒的查询

|  |
| --- |
| pt-kill --user=root --password=123456 --busy-time 60 --print |

kill掉大于60秒的查询

|  |
| --- |
| pt-kill --user=root --password=123456 --busy-time 60 --kill |

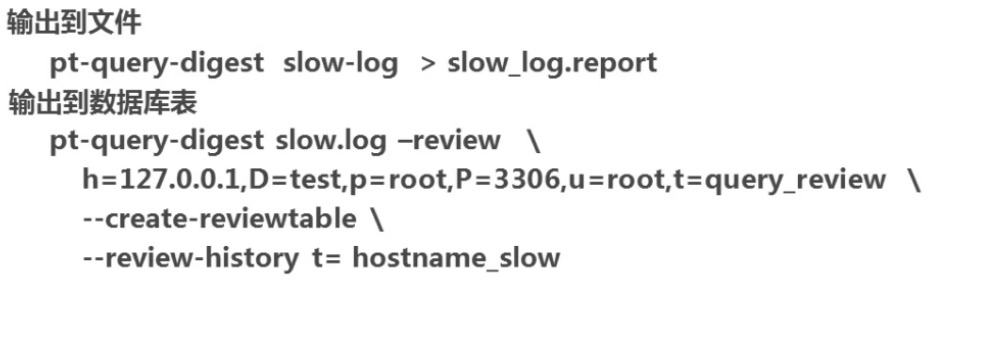
#### 13、查看mysql授权

|  |
| --- |
| 1、pt-show-grants --user=root --password=123456  2、pt-show-grants --user=root --password=123456 --separate –revoke |

#### 14、验证数据库复制的完整性

|  |
| --- |
| pt-table-checksum --user=root --password=123456 |

#### 15、附录：



# 6、如何通过慢查日志发现有问题的SQL

## 1、查询次数多且每次查询占用时间长的sql

通常为pt-query-digest分析的前几个查询；

该工具可以很清楚的看出每个SQL执行的次数及百分比等信息，执行的次数多，占比比较大的SQL

## 2、IO大的sql

注意pt-query-digest分析中的Rows examine项。扫描的行数越多，IO越大。

## 3、未命中的索引的SQL

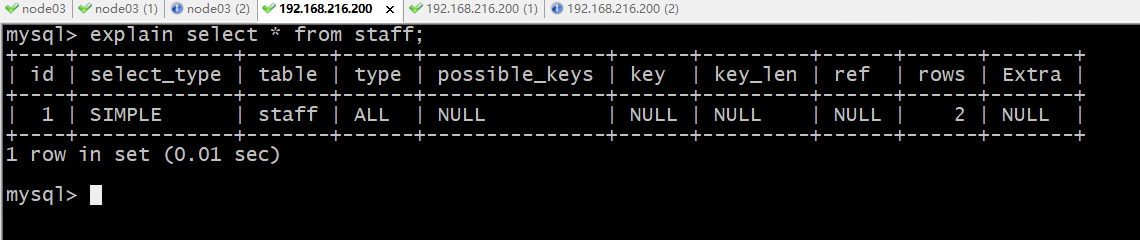
注意pt-query-digest分析中的Rows examine 和Rows Send的对比。说明该SQL的索引命中率不高，对于这种SQL，我们要重点进行关注。

# 7、通过explain查询分析SQL的执行计划

## 1、使用explain查询SQL的执行计划

SQL的执行计划侧面反映出了SQL的执行效率，具体执行方式如下所示：

在执行的SQL前面加上explain关键词即可；



## 2、每个字段的说明：

1）、id列数字越大越先执行，如果说数字一样大，那么就从上往下依次执行，id列为null的就表是这是一个结果集，不需要使用它来进行查询。

2）、select\_type列常见的有：

A：simple：表示不需要union操作或者不包含子查询的简单select查询。有连接查询时，外层的查询为simple，且只有一个

B：primary：一个需要union操作或者含有子查询的select，位于最外层的单位查询的select\_type即为primary。且只有一个

C：union：union连接的两个select查询，第一个查询是dervied派生表，除了第一个表外，第二个以后的表select\_type都是union

D：dependent union：与union一样，出现在union 或union all语句中，但是这个查询要受到外部查询的影响

E：union result：包含union的结果集，在union和union all语句中,因为它不需要参与查询，所以id字段为null

F：subquery：除了from子句中包含的子查询外，其他地方出现的子查询都可能是subquery

G：dependent subquery：与dependent union类似，表示这个subquery的查询要受到外部表查询的影响

H：derived：from字句中出现的子查询，也叫做派生表，其他数据库中可能叫做内联视图或嵌套select

3）、table

显示的查询表名，如果查询使用了别名，那么这里显示的是别名，如果不涉及对数据表的操作，那么这显示为null，如果显示为尖括号括起来的<derived N>就表示这个是临时表，后边的N就是执行计划中的id，表示结果来自于这个查询产生。如果是尖括号括起来的<union M,N>，与<derived N>类似，也是一个临时表，表示这个结果来自于union查询的id为M,N的结果集。

4）、type

依次从好到差：system，const，eq\_ref，ref，fulltext，ref\_or\_null，unique\_subquery，index\_subquery，range，index\_merge，index，ALL，除了all之外，其他的type都可以使用到索引，除了index\_merge之外，其他的type只可以用到一个索引

A：system：表中只有一行数据或者是空表，且只能用于myisam和memory表。如果是Innodb引擎表，type列在这个情况通常都是all或者index

B：const：使用唯一索引或者主键，返回记录一定是1行记录的等值where条件时，通常type是const。其他数据库也叫做唯一索引扫描

C：eq\_ref：出现在要连接过个表的查询计划中，驱动表只返回一行数据，且这行数据是第二个表的主键或者唯一索引，且必须为not null，唯一索引和主键是多列时，只有所有的列都用作比较时才会出现eq\_ref

D：ref：不像eq\_ref那样要求连接顺序，也没有主键和唯一索引的要求，只要使用相等条件检索时就可能出现，常见与辅助索引的等值查找。或者多列主键、唯一索引中，使用第一个列之外的列作为等值查找也会出现，总之，返回数据不唯一的等值查找就可能出现。

E：fulltext：全文索引检索，要注意，全文索引的优先级很高，若全文索引和普通索引同时存在时，mysql不管代价，优先选择使用全文索引

F：ref\_or\_null：与ref方法类似，只是增加了null值的比较。实际用的不多。

G：unique\_subquery：用于where中的in形式子查询，子查询返回不重复值唯一值

H：index\_subquery：用于in形式子查询使用到了辅助索引或者in常数列表，子查询可能返回重复值，可以使用索引将子查询去重。

I：range：索引范围扫描，常见于使用>,<,is null,between ,in ,like等运算符的查询中。

J：index\_merge：表示查询使用了两个以上的索引，最后取交集或者并集，常见and ，or的条件使用了不同的索引，官方排序这个在ref\_or\_null之后，但是实际上由于要读取所个索引，性能可能大部分时间都不如range

K：index：索引全表扫描，把索引从头到尾扫一遍，常见于使用索引列就可以处理不需要读取数据文件的查询、可以使用索引排序或者分组的查询。

L：all：这个就是全表扫描数据文件，然后再在server层进行过滤返回符合要求的记录。

5）、possible\_keys

查询可能使用到的索引都会在这里列出来

6）、key

查询真正使用到的索引，select\_type为index\_merge时，这里可能出现两个以上的索引，其他的select\_type这里只会出现一个。

7）、key\_len

用于处理查询的索引长度，如果是单列索引，那就整个索引长度算进去，如果是多列索引，那么查询不一定都能使用到所有的列，具体使用到了多少个列的索引，这里就会计算进去，没有使用到的列，这里不会计算进去。留意下这个列的值，算一下你的多列索引总长度就知道有没有使用到所有的列了。要注意，mysql的ICP特性使用到的索引不会计入其中。另外，key\_len只计算where条件用到的索引长度，而排序和分组就算用到了索引，也不会计算到key\_len中。

8）、ref

如果是使用的常数等值查询，这里会显示const，如果是连接查询，被驱动表的执行计划这里会显示驱动表的关联字段，如果是条件使用了表达式或者函数，或者条件列发生了内部隐式转换，这里可能显示为func

9）、rows

这里是执行计划中估算的扫描行数，不是精确值

10）、extra

这个列可以显示的信息非常多，有几十种，常用的有

A：distinct：在select部分使用了distinc关键字

B：no tables used：不带from字句的查询或者From dual查询

C：使用not in()形式子查询或not exists运算符的连接查询，这种叫做反连接。即，一般连接查询是先查询内表，再查询外表，反连接就是先查询外表，再查询内表。

D：using filesort：排序时无法使用到索引时，就会出现这个。常见于order by和group by语句中

E：using index：查询时不需要回表查询，直接通过索引就可以获取查询的数据。

F：using join buffer（block nested loop），using join buffer（batched key accss）：5.6.x之后的版本优化关联查询的BNL，BKA特性。主要是减少内表的循环数量以及比较顺序地扫描查询。

G：using sort\_union，using\_union，using intersect，using sort\_intersection：

using intersect：表示使用and的各个索引的条件时，该信息表示是从处理结果获取交集

using union：表示使用or连接各个使用索引的条件时，该信息表示从处理结果获取并集

using sort\_union和using sort\_intersection：与前面两个对应的类似，只是他们是出现在用and和or查询信息量大时，先查询主键，然后进行排序合并后，才能读取记录并返回。

H：using temporary：表示使用了临时表存储中间结果。临时表可以是内存临时表和磁盘临时表，执行计划中看不出来，需要查看status变量，used\_tmp\_table，used\_tmp\_disk\_table才能看出来。

I：using where：表示存储引擎返回的记录并不是所有的都满足查询条件，需要在server层进行过滤。查询条件中分为限制条件和检查条件，5.6之前，存储引擎只能根据限制条件扫描数据并返回，然后server层根据检查条件进行过滤再返回真正符合查询的数据。5.6.x之后支持ICP特性，可以把检查条件也下推到存储引擎层，不符合检查条件和限制条件的数据，直接不读取，这样就大大减少了存储引擎扫描的记录数量。extra列显示using index condition

J：firstmatch(tb\_name)：5.6.x开始引入的优化子查询的新特性之一，常见于where字句含有in()类型的子查询。如果内表的数据量比较大，就可能出现这个

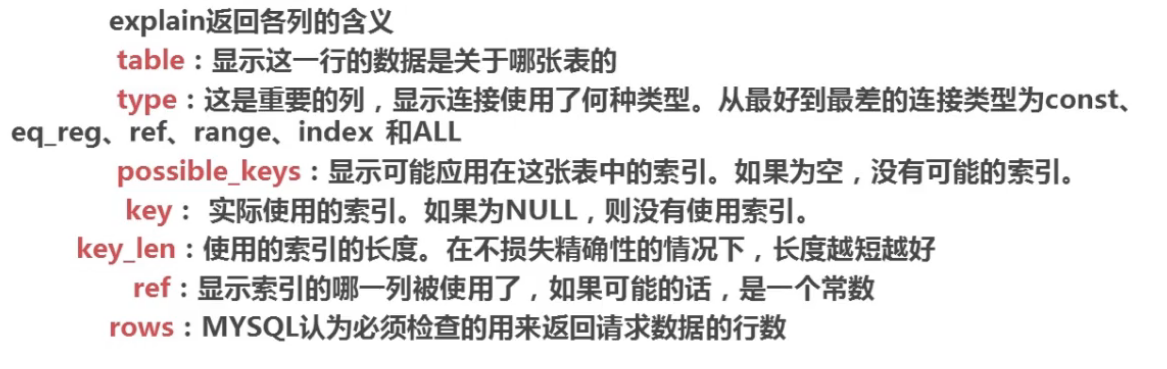
K：loosescan(m..n)：5.6.x之后引入的优化子查询的新特性之一，在in()类型的子查询中，子查询返回的可能有重复记录时，就可能出现这个

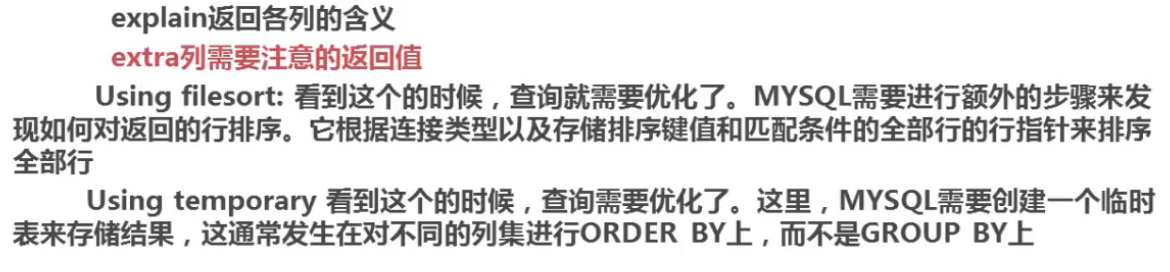
除了这些之外，还有很多查询数据字典库，执行计划过程中就发现不可能存在结果的一些提示信息

11）、filtered

使用explain extended时会出现这个列，5.7之后的版本默认就有这个字段，不需要使用explain extended了。这个字段表示存储引擎返回的数据在server层过滤后，剩下多少满足查询的记录数量的比例，注意是百分比，不是具体记录数。

附图：





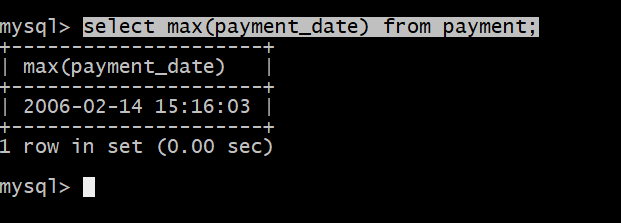
## 3、具体慢查询的优化案例

### 1、函数Max()的优化

用途：查询最后支付时间-优化max（）函数

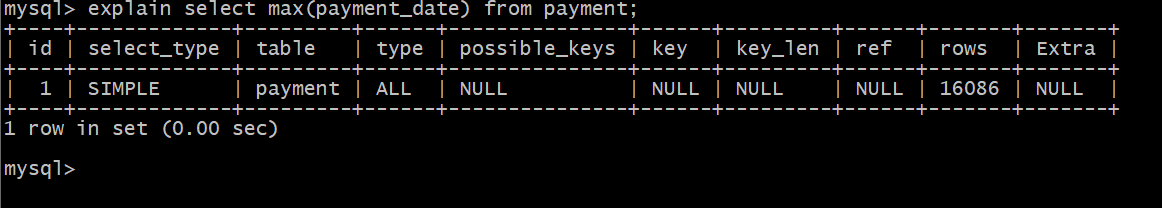
语句：

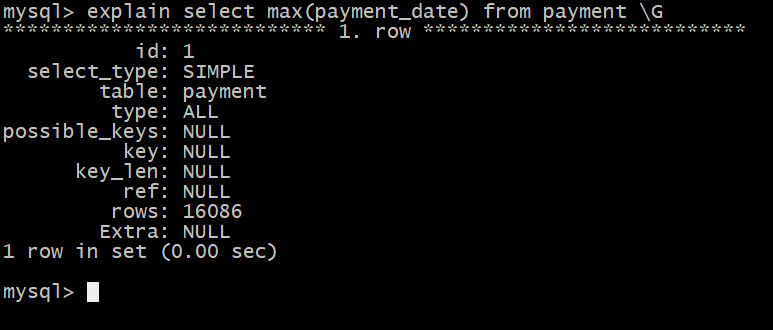
|  |
| --- |
| select max(payment\_date) from payment; |



执行计划：

|  |
| --- |
| explain select max(payment\_date) from payment; |

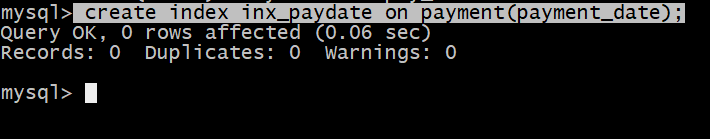


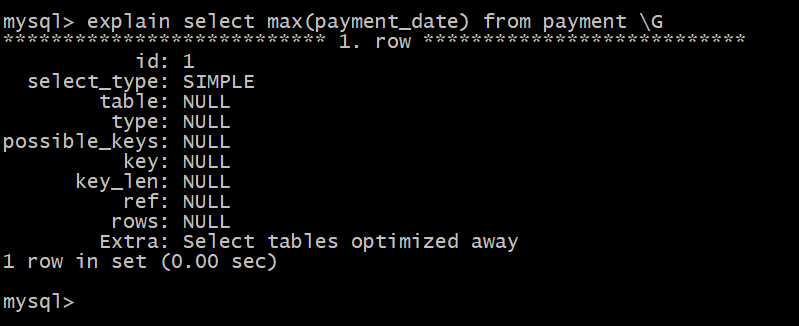


可以看到显示的执行计划，并不是很高效，可以拖慢服务器的效率，如何优化了？

创建索引

|  |
| --- |
| create index inx\_paydate on payment(payment\_date); |





索引是顺序操作的，不需要扫描表，执行效率就会比较恒定，

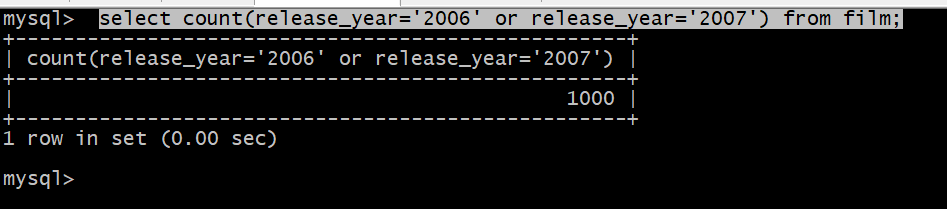
### 2、函数Count()的优化

需求：在一条SQL中同时查处2006年和2007年电影的数量

错误的方式：

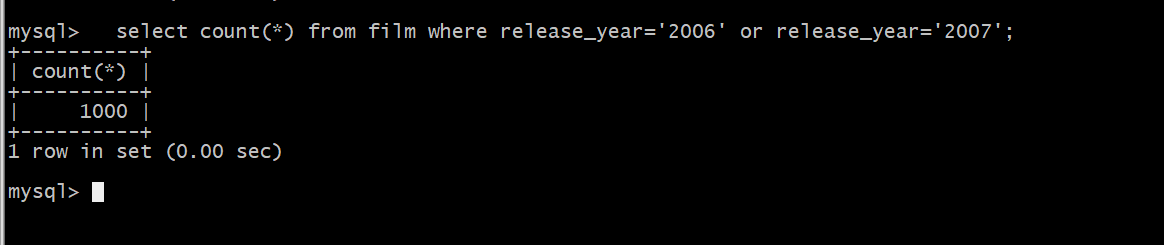
语句：

|  |
| --- |
| select count(release\_year='2006' or release\_year='2007') from film; |



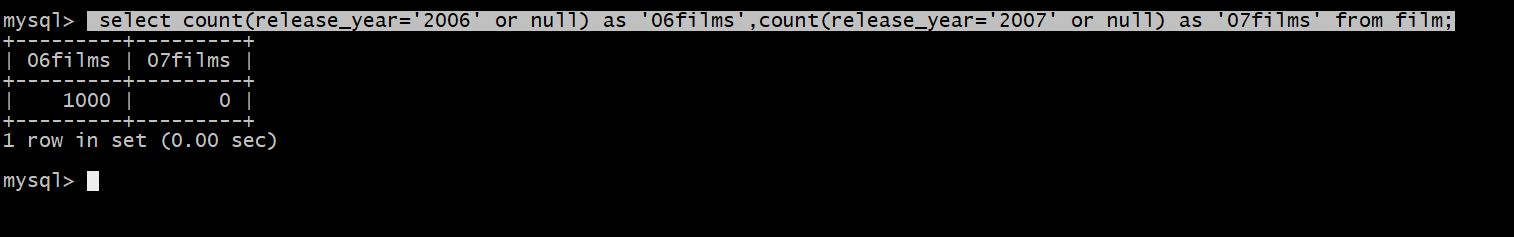
2006和2007年分别是多少，判断不出来

|  |
| --- |
| select count(\*) from film where release\_year='2006' or release\_year='2007'; |



正确的编写方式：

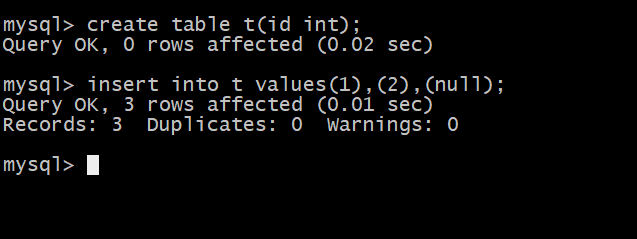
|  |
| --- |
| select count(release\_year='2006' or null) as '06films',count(release\_year='2007' or null) as '07films' from film; |



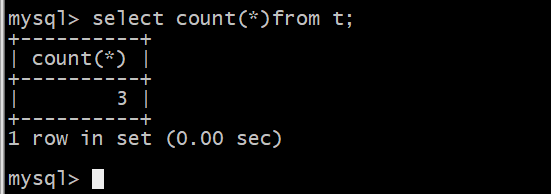
区别：count（\*）和count（id）

创建表并插入语句

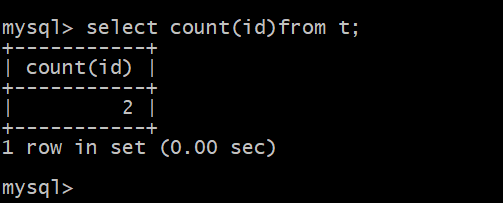
|  |
| --- |
| create table t(id int);    insert into t values(1),(2),(null); |



Count（\*）：select count(\*)from t;



Count（id）：select count(id)from t;



说明：

Count（id）是不包含null的值

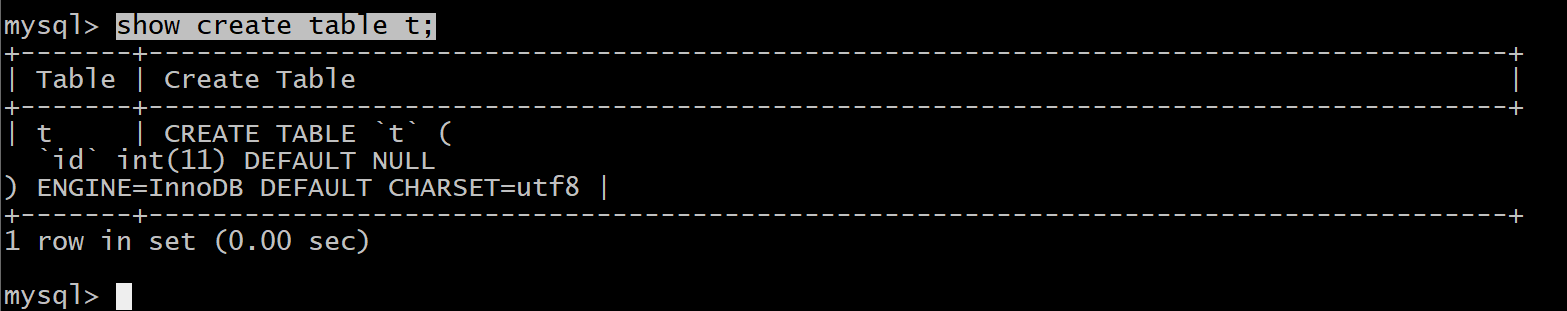
Count（\*）是包含null的值

### 3、子查询的优化

子查询是我们在开发过程中经常使用的一种方式，在通常情况下，需要把子查询优化为join查询但在优化是需要注意关联键是否有一对多的关系，要注意重复数据。

查看我们所创建的t表

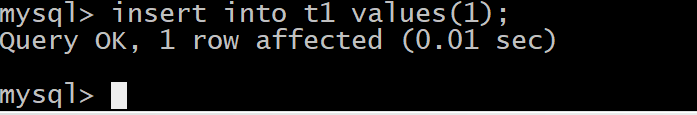
|  |
| --- |
| show create table t; |



接下来我们创建一个t1表

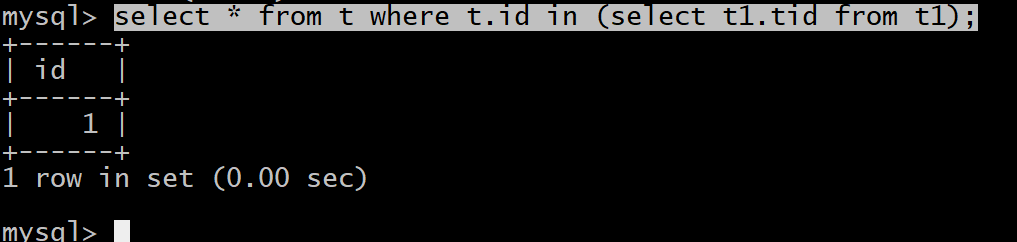
|  |
| --- |
| create table t1(tid int); |

并插入一条数据



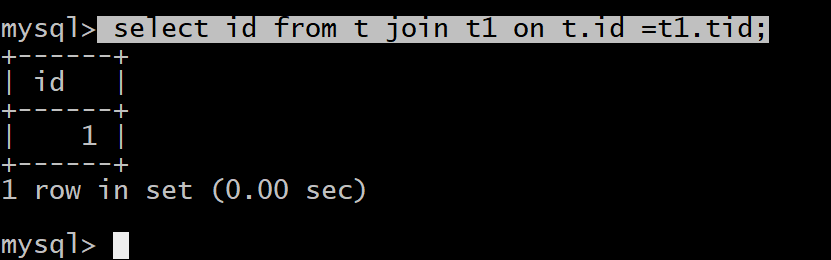
我们要进行一个子查询，需求：查询t表中id在t1表中tid的所有数据；

|  |
| --- |
| select \* from t where t.id in (select t1.tid from t1); |



接下来我们用join的操作来进行操作

|  |
| --- |
| select id from t join t1 on t.id =t1.tid; |

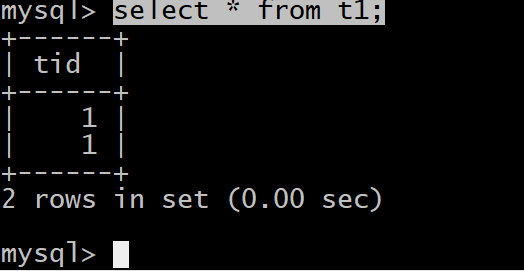


通过上面结果来看，查询的结果是一致的，我们就将子查询的方式优化为join操作。

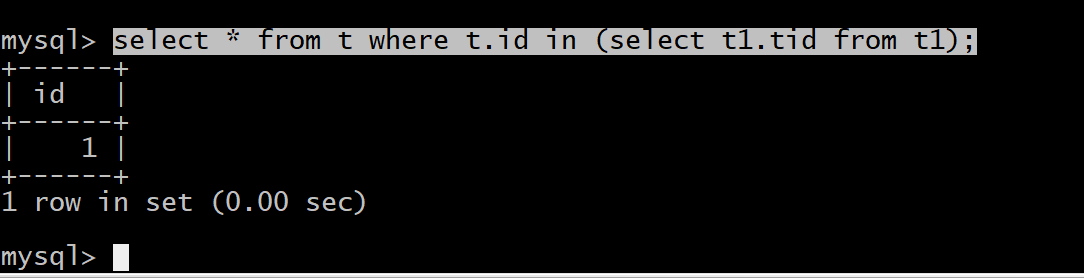
接下来，我们在t1表中再插入一条数据

|  |
| --- |
| insert into t1 values (1); |

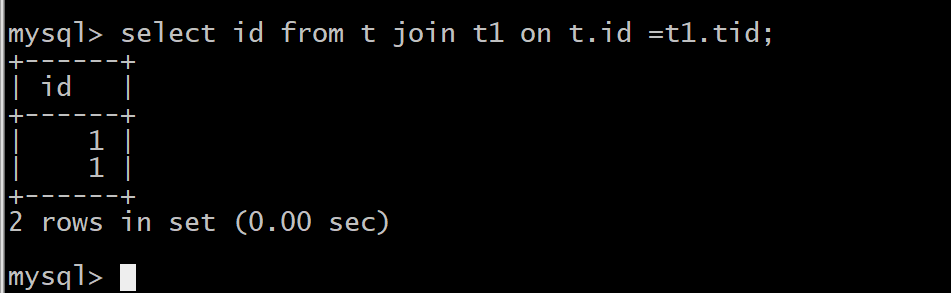
|  |
| --- |
| select \* from t1; |



在这种情况下，如果我们使用子查询方式进行查询，返回的结果就是如下图所示：

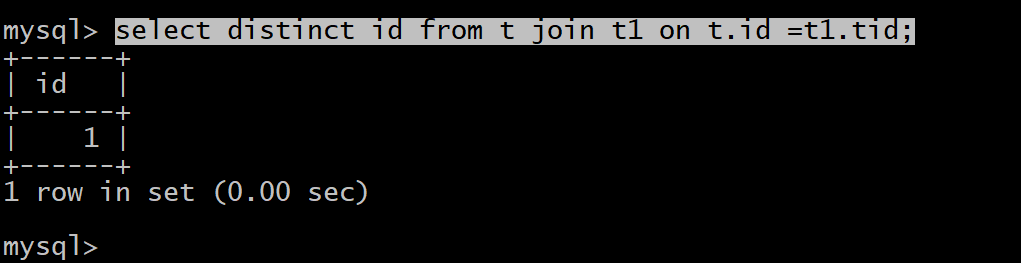


如果使用join方式进行查找，如下图所示：



在这种情况下出现了一对多的关系，会出现数据的重复，我们为了方式数据重复，不得不使用distinct关键词进行去重操作

|  |
| --- |
| select distinct id from t join t1 on t.id =t1.tid; |



注意：这个一对多的关系是我们开发过程中遇到的一个坑，出现数据重复，需要大家注意一下。

例子：查询sandra出演的所有影片：

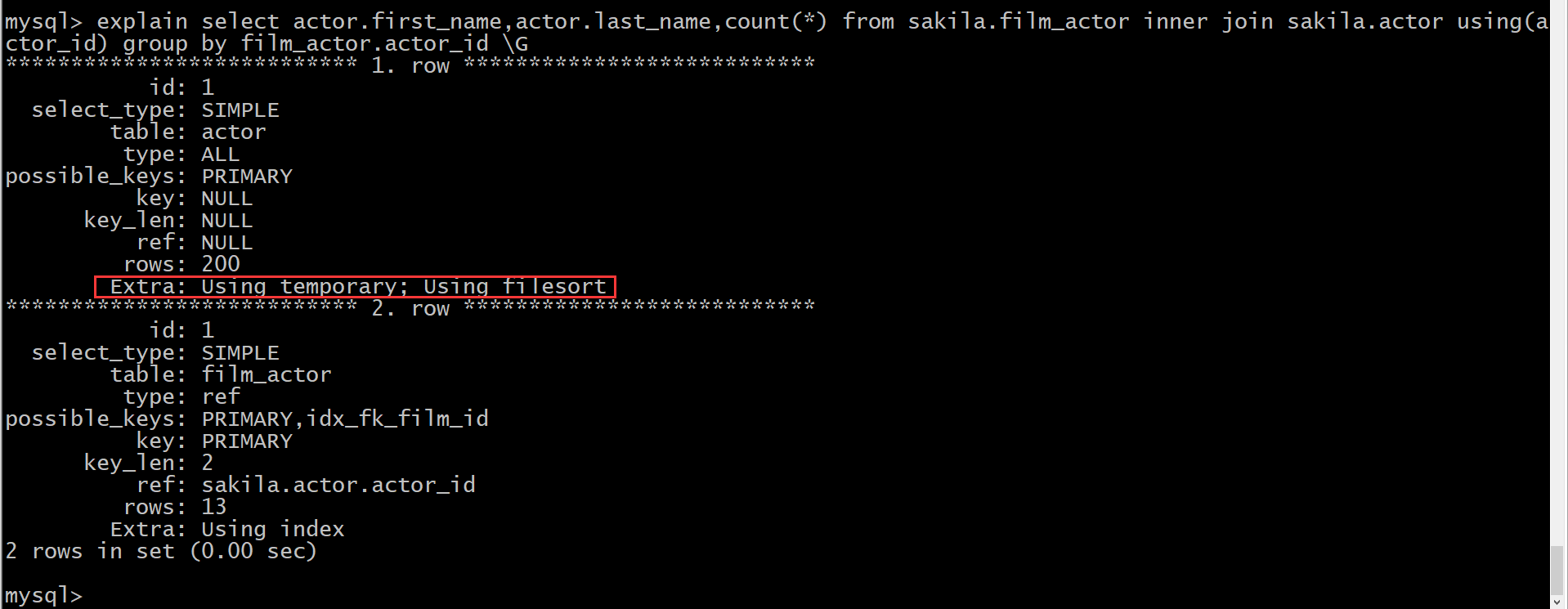
|  |
| --- |
| explain select title,release\_year,length  from film  where film\_id in (  select film\_id from film\_actor where actor\_id in (  select actor\_id from actor where first\_name='sandra')); |

### 4、group by的优化

最好使用同一表中的列，

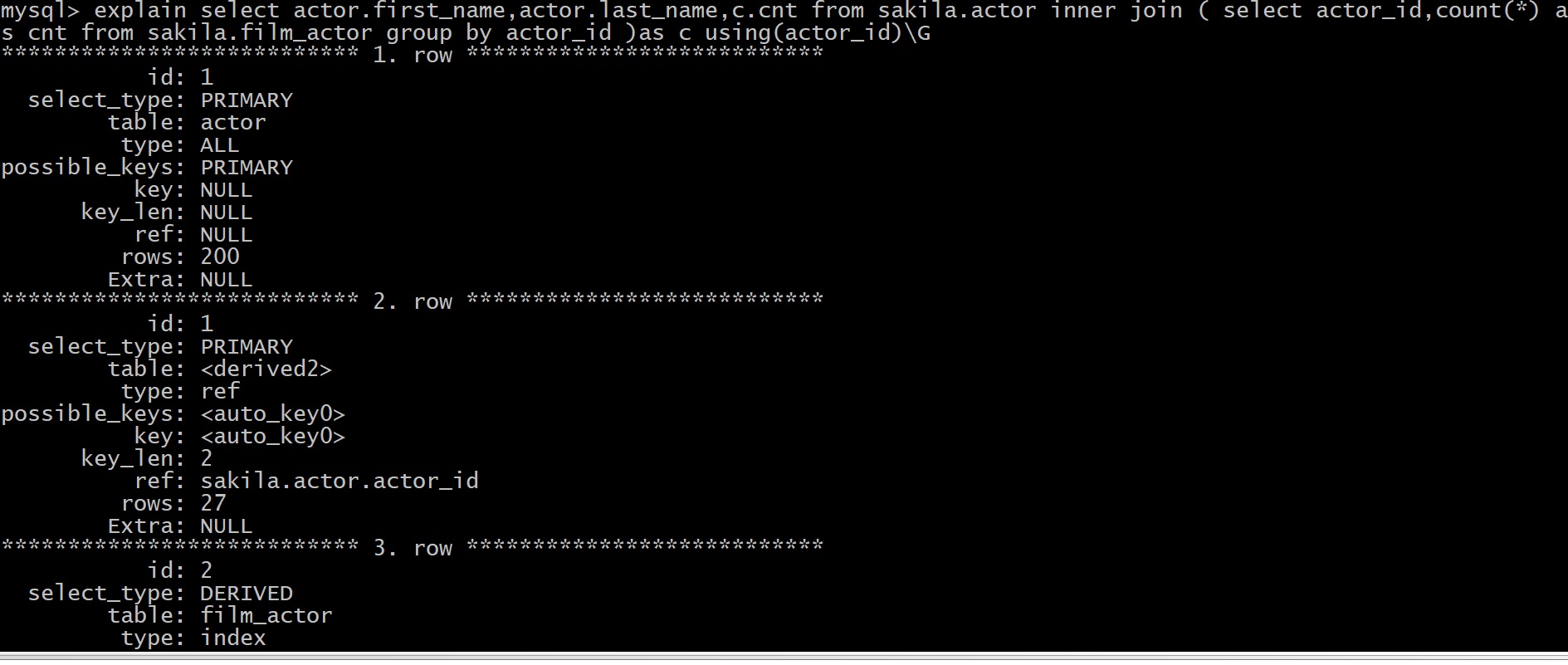
需求：每个演员所参演影片的数量-（影片表和演员表）

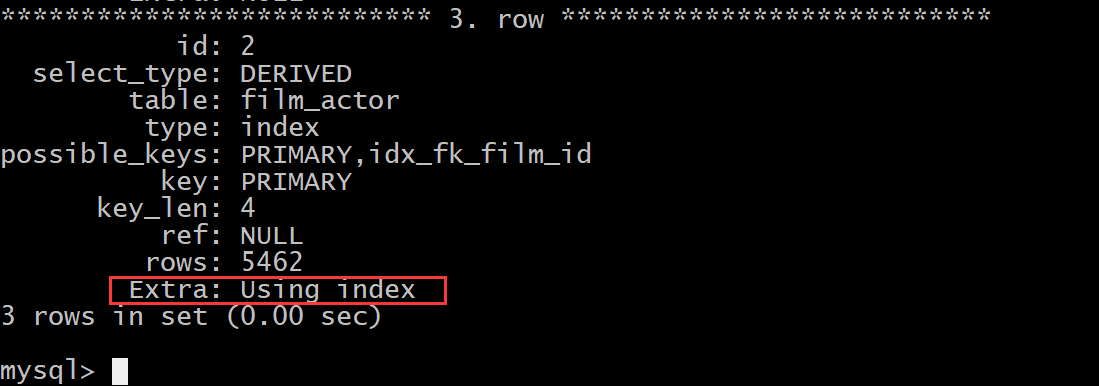
|  |
| --- |
| explain select actor.first\_name,actor.last\_name,count(\*)  from sakila.film\_actor  inner join sakila.actor using(actor\_id)  group by film\_actor.actor\_id; |



优化后的SQL：

|  |
| --- |
| explain select actor.first\_name,actor.last\_name,c.cnt  from sakila.actor inner join (  select actor\_id,count(\*) as cnt from sakila.film\_actor group by actor\_id  )as c using(actor\_id); |





说明：从上面的执行计划来看，这种优化后的方式没有使用临时文件和文件排序的方式了，取而代之的是使用了索引。查询效率老高了。

这个时候我们表中的数据比较大，会大量的占用IO操作，优化了sql执行的效率，节省了服务器的资源，因此我们就需要优化。

注意：

1、mysql 中using关键词的作用：也就是说要使用using,那么表a和表b必须要有相同的列。

2、在用Join进行多表联合查询时，我们通常使用On来建立两个表的关系。其实还有一个更方便的关键字，那就是Using。

3、如果两个表的关联字段名是一样的，就可以使用Using来建立关系，简洁明了。

### 5、Limit查询的优化

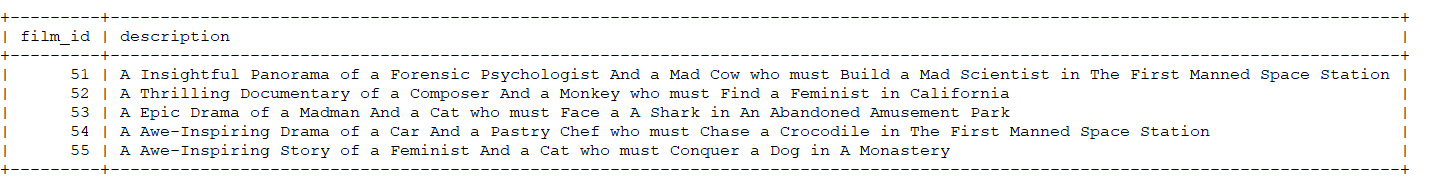
Limit常用于分页处理，时长会伴随order by从句使用，因此大多时候回使用Filesorts这样会造成大量的IO问题。

例子：

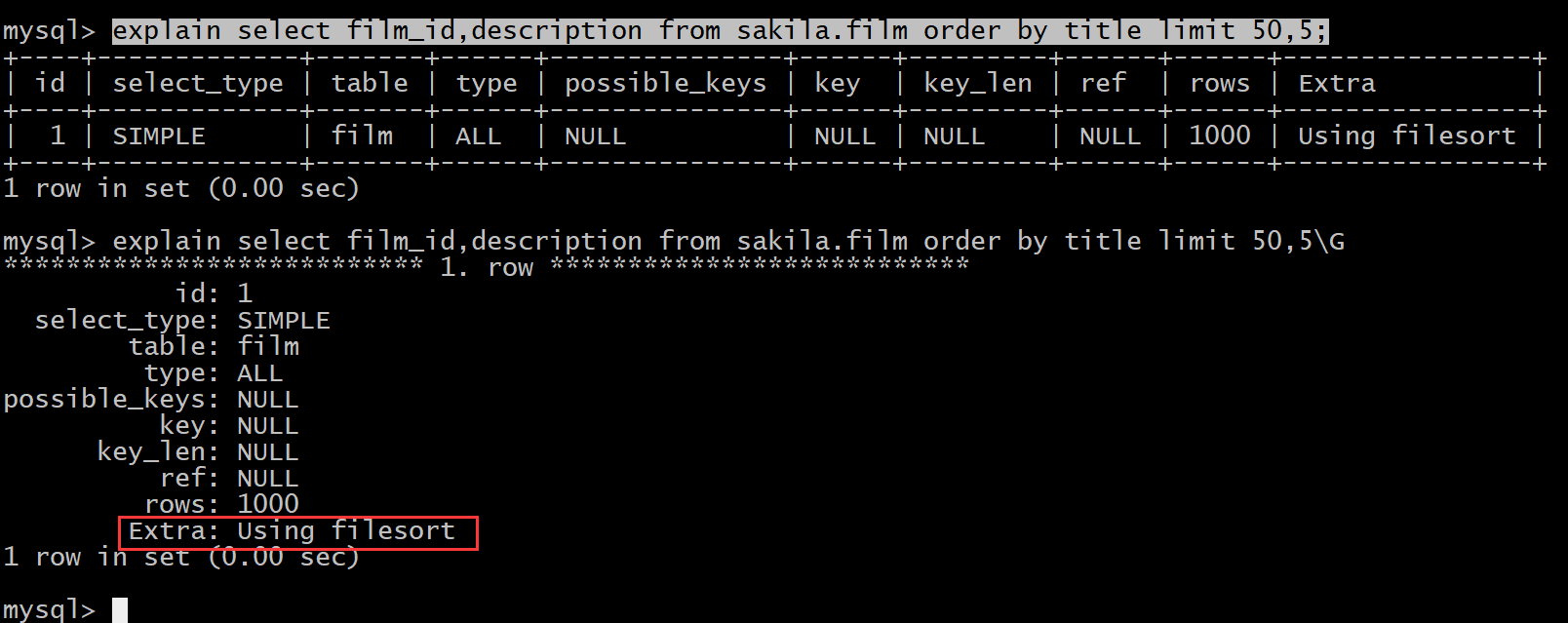
需求：查询影片id和描述信息，并根据主题进行排序，取出从序号50条开始的5条数据。

|  |
| --- |
| select film\_id,description from sakila.film order by title limit 50,5; |

执行的结果：



在查看一下它的执行计划：

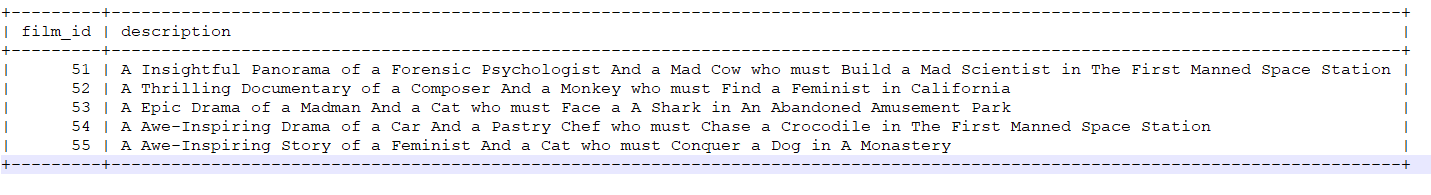


对于这种操作，我们该用什么样的优化方式了？

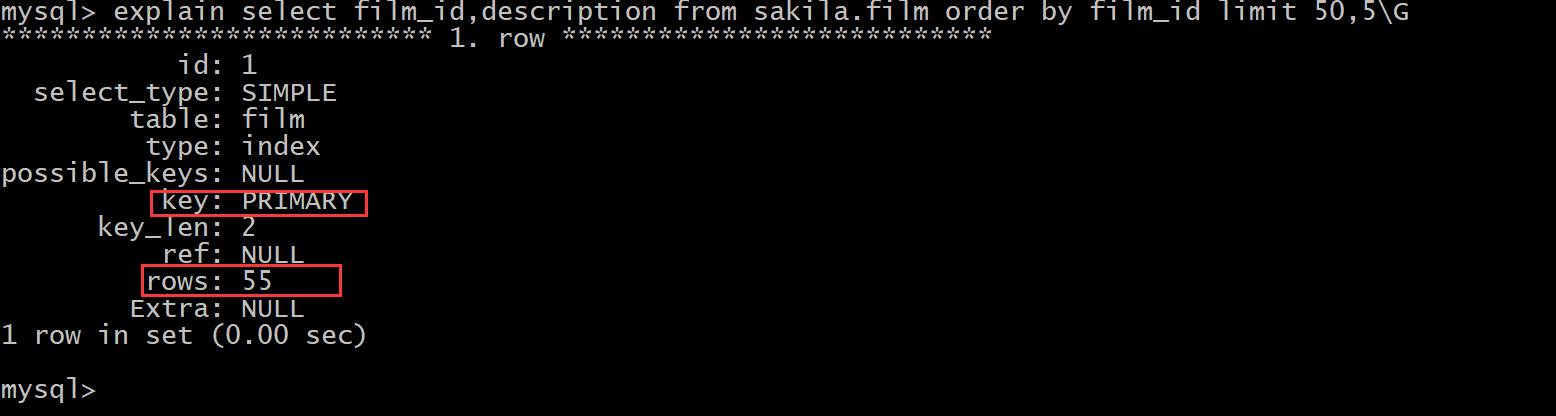
优化步骤1：

使用有索引的列或主键进行order by操作，因为大家知道，innodb是按照主键的逻辑顺序进行排序的。可以避免很多的IO操作。

|  |
| --- |
| select film\_id,description from sakila.film order by film\_id limit 50,5; |

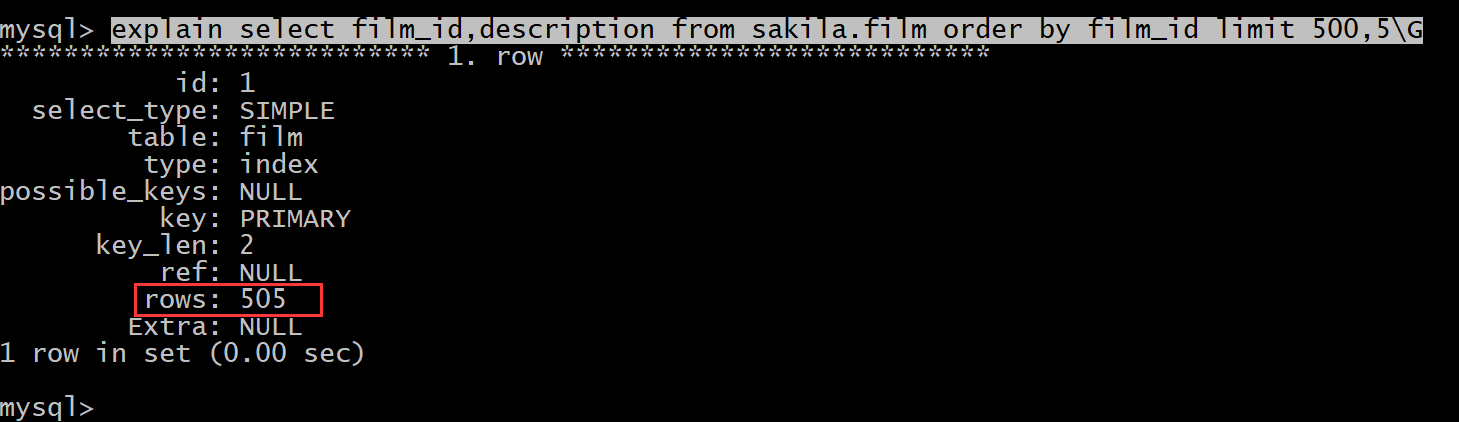


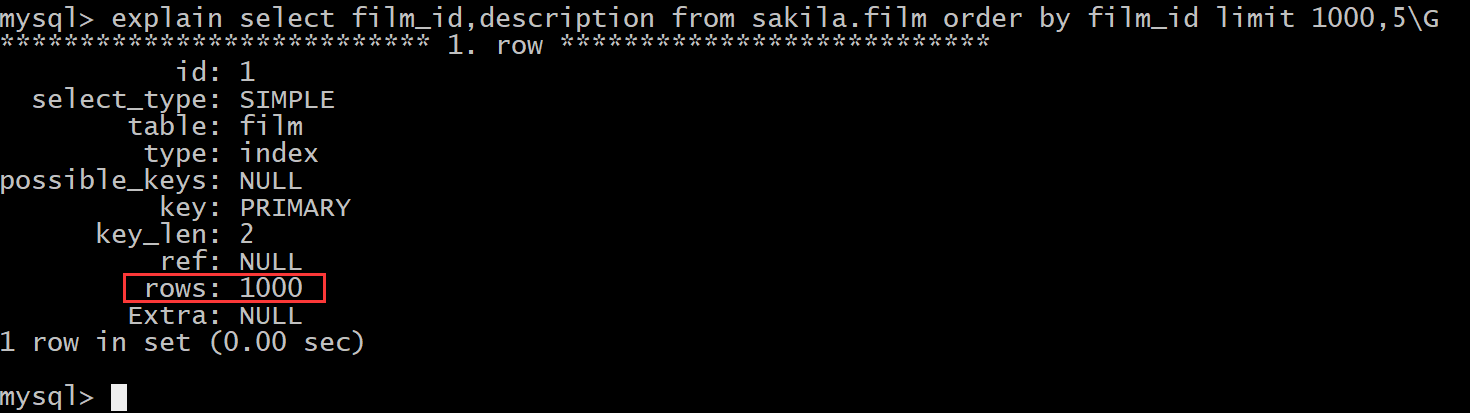
查看一下执行计划



那如果我们获取从500行开始的5条记录，执行计划又是什么样的了？

|  |
| --- |
| explain select film\_id,description from sakila.film order by film\_id limit 500,5\G |





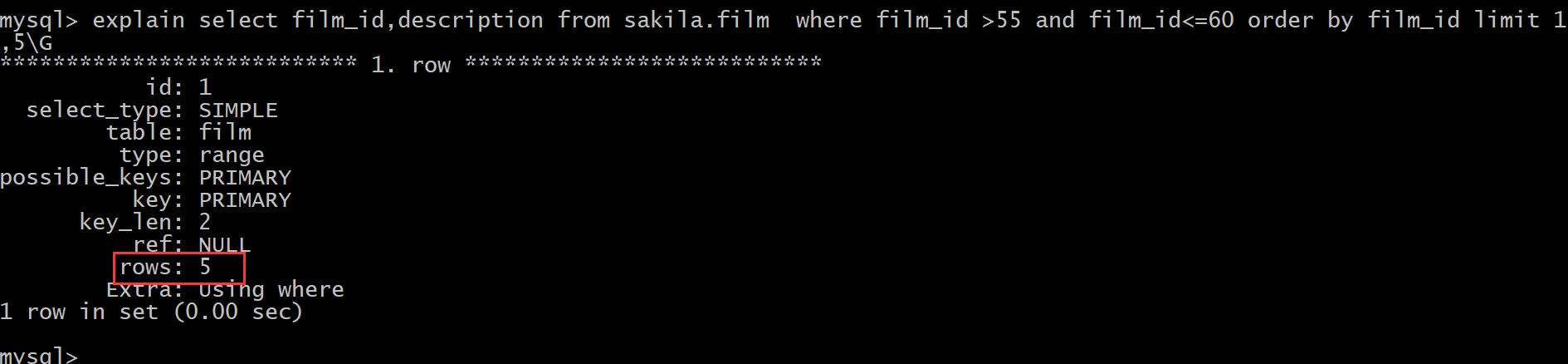
随着我们翻页越往后，IO操作会越来越大的，如果一个表有几千万行数据，翻页越后面，会越来越慢，因此我们要进一步的来优化。

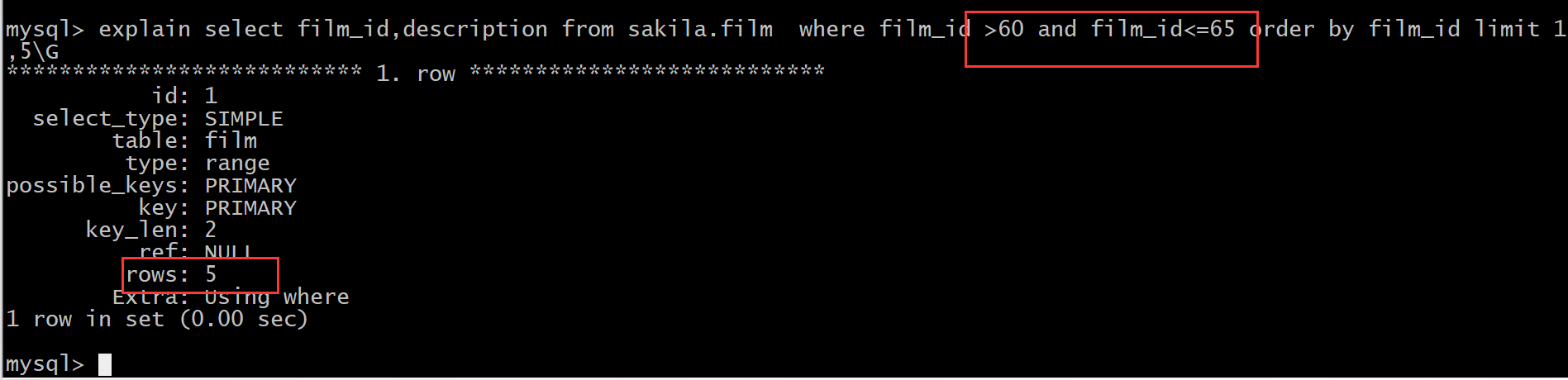
优化步骤2、记录上次返回的主键， 在下次查询时使用主键过滤。（说明：避免了数据量大时扫描过多的记录）

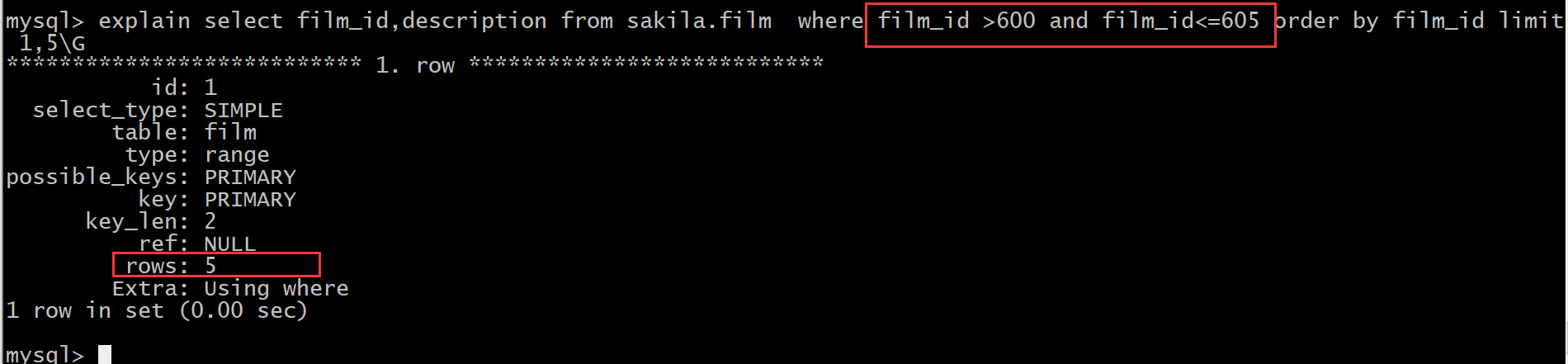
上次limit是50,5的操作，因此我们在这次优化过程需要使用上次的索引记录值，

|  |
| --- |
| select film\_id,description from sakila.film where film\_id >55 and film\_id<=60 order by film\_id limit 1,5; |

查看执行计划：







结论：**扫描行数不变，执行计划是很固定，效率也是很固定的**

注意事项：

主键要顺序排序并连续的，如果主键中间空缺了某一列，或者某几列，会出现列出数据不足5行的数据；如果不连续的情况，建立一个附加的列index\_id列，保证这一列数据要自增的，并添加索引即可。

### 6、索引的优化

#### 1、什么是索引？

索引的作用相当于图书的目录，可以根据目录中的页码快速找到所需的内容。

数据库使用索引以找到特定值，然后顺指针找到包含该值的行。在表中建立索引，然后在索引中找到符合查询条件的索引值，最后通过保存在索引中的ROWID（相当于页码）快速找到表中对应的记录。

索引的建立是表中比较有指向性的字段，相当于目录，

比如说行政区域代码，同一个地域的行政区域代码都是相同的，那么给这一列加上索引，避免让它重复扫描，从而达到优化的目的！

#### 2、如何创建索引

在执行CREATE TABLE语句时可以创建索引，也可以单独用CREATE INDEX或ALTER TABLE来为表增加索引。

1、ALTER TABLE

ALTER TABLE用来创建普通索引、UNIQUE索引或PRIMARY KEY索引。

ALTER TABLE table\_name ADD INDEX index\_name (column\_list)

ALTER TABLE table\_name ADD UNIQUE (column\_list)

ALTER TABLE table\_name ADD PRIMARY KEY (column\_list)

说明：其中table\_name是要增加索引的表名，column\_list指出对哪些列进行索引，多列时各列之间用逗号分隔。索引名index\_name可选，缺省时，MySQL将根据第一个索引列赋一个名称。另外，ALTER TABLE允许在单个语句中更改多个表，因此可以在同时创建多个索引。

2、CREATE INDEX

CREATE INDEX可对表增加普通索引或UNIQUE索引。

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

CREATE UNIQUE INDEX index\_name ON table\_name (column\_list)

说明：table\_name、index\_name和column\_list具有与ALTER TABLE语句中相同的含义，索引名不可选。另外，不能用CREATE INDEX语句创建PRIMARY KEY索引。

3、索引类型

在创建索引时，可以规定索引能否包含重复值。如果不包含，则索引应该创建为PRIMARY KEY或UNIQUE索引。对于单列惟一性索引，这保证单列不包含重复的值。对于多列惟一性索引，保证多个值的组合不重复。

PRIMARY KEY索引和UNIQUE索引非常类似。

事实上，PRIMARY KEY索引仅是一个具有名称PRIMARY的UNIQUE索引。这表示一个表只能包含一个PRIMARY KEY，因为一个表中不可能具有两个同名的索引。

下面的SQL语句对students表在sid上添加PRIMARY KEY索引。

ALTER TABLE students ADD PRIMARY KEY (sid)

4、删除索引

可利用ALTER TABLE或DROP INDEX语句来删除索引。类似于CREATE INDEX语句，DROP INDEX可以在ALTER TABLE内部作为一条语句处理，语法如下。

DROP INDEX index\_name ON talbe\_name

ALTER TABLE table\_name DROP INDEX index\_name

ALTER TABLE table\_name DROP PRIMARY KEY

其中，前两条语句是等价的，删除掉table\_name中的索引index\_name。

第3条语句只在删除PRIMARY KEY索引时使用，因为一个表只可能有一个PRIMARY KEY索引，因此不需要指定索引名。如果没有创建PRIMARY KEY索引，但表具有一个或多个UNIQUE索引，则MySQL将删除第一个UNIQUE索引。

如果从表中删除了某列，则索引会受到影响。对于多列组合的索引，如果删除其中的某列，则该列也会从索引中删除。如果删除组成索引的所有列，则整个索引将被删除。

5、查看索引

|  |
| --- |
| mysql> show index from tblname;  mysql> show keys from tblname; |

1. 什么情况下，使用索引了？

**适合索引**

1、表的主关键字

2、自动建立唯一索引

3、表的字段唯一约束

4、直接条件查询的字段（在SQL中用于条件约束的字段） where

5、查询中与其它表关联的字段 on

6、查询中排序的字段（排序的字段如果通过索引去访问那将大大提高排序速度）order by

7、查询中统计或分组统计的字段 group by

**不适合索引**

8、表记录太少（如果一个表只有5条记录，采用索引去访问记录的话，那首先需访问索引表，再通过索引表访问数据表，一般索引表与数据表不在同一个数据块）

9、经常插入、删除、修改的表（对一些经常处理的业务表应在查询允许的情况下尽量减少索引）-索引主要针对查询

10、数据重复且分布平均的表字段（假如一个表有10万行记录，有一个字段A只有T和F两种值，且每个值的分布概率大约为50%，那么对这种表A字段建索引一般不会提高数据库的查询速度。）

**不太理解**

11、经常和主字段一块查询但主字段索引值比较多的表字段

12、对千万级MySQL数据库建立索引的事项及提高性能的手段

#### 3、如何选择合适的列建立索引

1、在where从句，group by从句，order by从句，on从句中的列添加索引

2、索引字段越小越好（因为数据库数据存储单位是以“页”为单位的，数据存储的越多，IO也会越大）

3、离散度大的列放到联合索引的前面

例子：

|  |
| --- |
| select \* from payment where staff\_id =2 and customer\_id =584; |

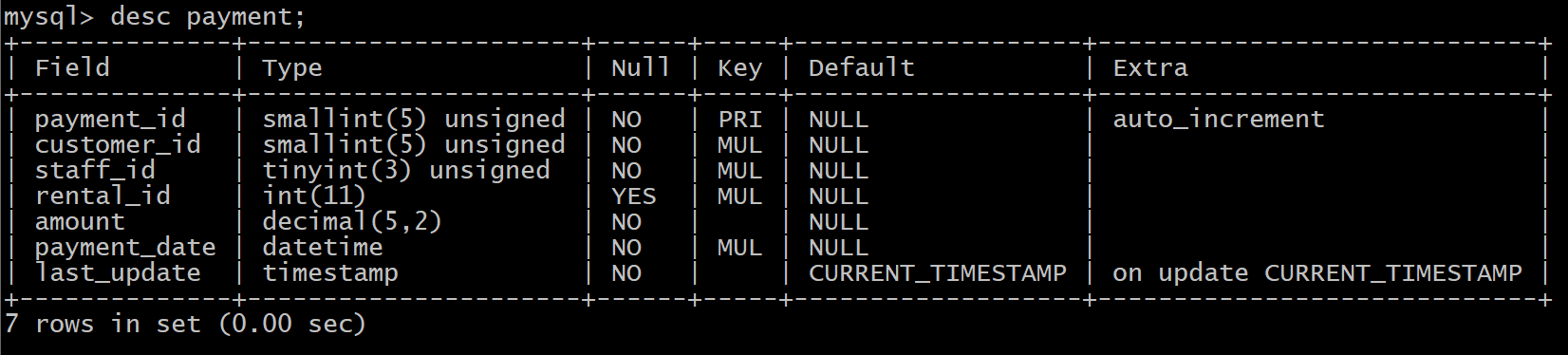
注意:

是index（staff\_id，customer\_id）好，还是index（customer\_id，staff\_id）好

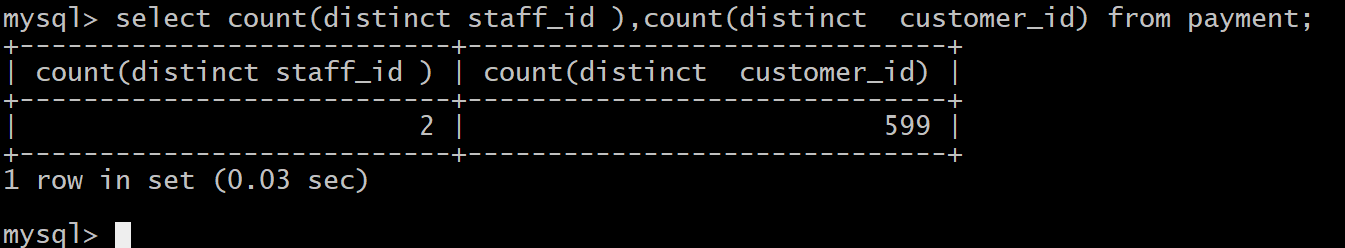
那我们怎么进行验证离散度好了？

A、我们先查看一下表结构

|  |
| --- |
| desc payment; |



B、分别查看这两个字段中不同的id的数量，数量越多，则表明离散程度越大：因此可以通过下图看出：customer\_id 离散程度大。



结论：由于customer\_id 离散程度大，使用index（customer\_id，staff\_id）好

C、mysql联合索引

①命名规则 ：表名\_字段名

1、需要加索引的字段，要在where条件中

2、数据量少的字段不需要加索引

3、如果where条件中是OR关系，加索引不起作用

4、符合最左原则

②什么是联合索引

1. 两个或更多个列上的索引被称作联合索引，又被称为是复合索引。
2. 利用索引中的附加列，您可以缩小搜索的范围，但使用一个具有两列的索引 不同于使用两个单独的索引。复合索引的结构与电话簿类似，人名由姓和名构成，电话簿首先按姓氏对进行排序，然后按名字对有相同姓氏的人进行排序。如果您知 道姓，电话簿将非常有用；如果您知道姓和名，电话簿则更为有用，但如果您只知道名不姓，电话簿将没有用处。

所以说创建复合索引时，应该仔细考虑列的顺序。对索引中的所有列执行搜索或仅对前几列执行搜索时，复合索引非常有用；仅对后面的任意列执行搜索时，复合索引则没有用处。

#### 4、索引优化SQL的方法

1、索引的维护及优化（重复及冗余索引）

增加索引会有利于查询效率，但会降低insert，update，delete的效率，但实际上往往不是这样的，过多的索引会不但会影响使用效率，同时会影响查询效率，这是由于数据库进行查询分析时，首先要选择使用哪一个索引进行查询，如果索引过多，分析过程就会越慢，这样同样的减少查询的效率，因此我们要知道如何增加，有时候要知道维护和删除不需要的索引

2、如何找到重复和冗余的索引

重复索引：

重复索引是指相同的列以相同的顺序建立的同类型的索引，如下表中的 primary key和ID列上的索引就是重复索引

|  |
| --- |
| create table test(  id int not null primary key,  name varchar(10) not null,  title varchar(50) not null,  unique(id)  )engine=innodb; |

冗余索引：

冗余索引是指多个索引的前缀列相同，或是在联合索引中包含了主键的索引，下面这个例子中key（name，id）就是一个冗余索引。

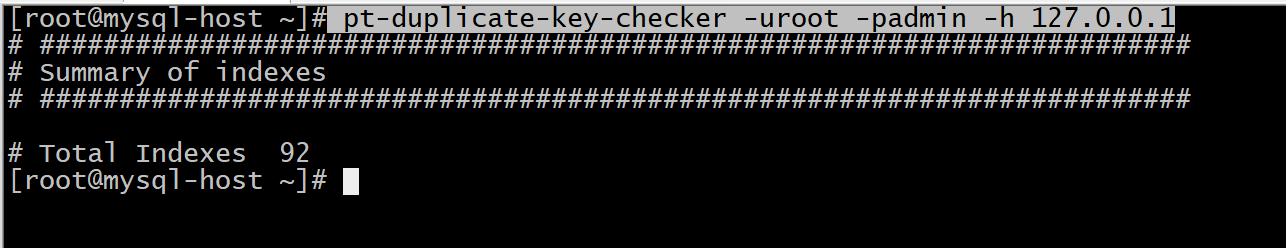
|  |
| --- |
| create table test(  id int not null primary key,  name varchar(10) not null,  title varchar(50) not null,  key(name,id)  )engine=innodb; |

说明：对于innodb来说，每一个索引后面，实际上都会包含主键，这时候我们建立的联合索引，又人为的把主键包含进去，那么这个时候就是一个冗余索引。

3、如何查找重复索引

工具：使用pt-duplicate-key-checker工具检查重复及冗余索引

|  |
| --- |
| pt-duplicate-key-checker -uroot -padmin -h 127.0.0.1 |

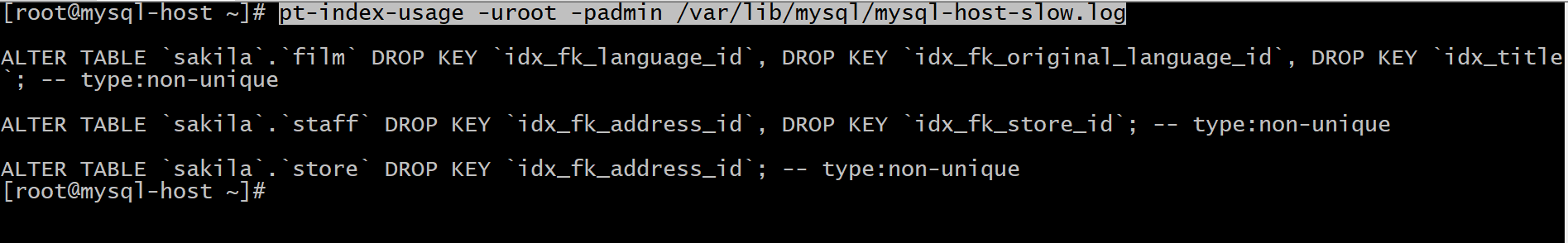


4、索引维护的方法

由于业务变更，某些索引是后续不需要使用的，就要进行删除。

在mysql中，目前只能通过慢查询日志配合pt-index-usage工具来进行索引使用情况的分析；

|  |
| --- |
| pt-index-usage -uroot -padmin /var/lib/mysql/mysql-host-slow.log |



附：https://www.percona.com/downloads/

#### 5、注意事项

设计好MySql的索引可以让你的数据库飞起来，大大的提高数据库效率。设计MySql索引的时候有一下几点注意：

1，创建索引

对于查询占主要的应用来说，索引显得尤为重要。很多时候性能问题很简单的就是因为我们忘了添加索引而造成的，或者说没有添加更为有效的索引导致。如果不加索引的话，那么查找任何哪怕只是一条特定的数据都会进行一次全表扫描，如果一张表的数据量很大而符合条件的结果又很少，那么不加索引会引起致命的性能下降。

但是也不是什么情况都非得建索引不可，比如性别可能就只有两个值，建索引不仅没什么优势，还会影响到更新速度，这被称为过度索引。

2，复合索引

比如有一条语句是这样的：select \* from users where area=’beijing’ and age=22;

如果我们是在area和age上分别创建单个索引的话，由于mysql查询每次只能使用一个索引，所以虽然这样已经相对不做索引时全表扫描提高了很多效率，但是如果在area、age两列上创建复合索引的话将带来更高的效率。如果我们创建了(area, age,salary)的复合索引，那么其实相当于创建了(area,age,salary)、(area,age)、(area)三个索引，这被称为最佳左前缀特性。

因此我们在创建复合索引时应该将最常用作限制条件的列放在最左边，依次递减。

3，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。

4，使用短索引

对字符串列进行索引，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。

5，排序的索引问题

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。

6，like语句操作

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。

7，不要在列上进行运算

select \* from users where

YEAR(adddate)

8，不使用NOT IN操作

NOT IN操作都不会使用索引将进行全表扫描。NOT IN可以NOT EXISTS代替