# SQL执行很慢的原因

**一、开始装逼：分类讨论**

一条 SQL 语句执行的很慢，那是每次执行都很慢呢？还是大多数情况下是正常的，偶尔出现很慢呢？所以我觉得，我们还得分以下两种情况来讨论。

**1、大多数情况是正常的，只是偶尔会出现很慢的情况。**

**2、在数据量不变的情况下，这条SQL语句一直以来都执行的很慢。**

针对这两种情况，我们来分析下可能是哪些原因导致的。

**二、针对偶尔很慢的情况**

一条 SQL 大多数情况正常，偶尔才能出现很慢的情况，针对这种情况，我觉得这条SQL语句的书写本身是没什么问题的，而是其他原因导致的，那会是什么原因呢？

**1、数据库在刷新脏页（flush）我也无奈啊**

当我们要往数据库插入一条数据、或者要更新一条数据的时候，我们知道数据库会在**内存**中把对应字段的数据更新了，但是更新之后，这些更新的字段并不会马上同步持久化到**磁盘**中去，而是把这些更新的记录写入到 redo log 日记中去，等到空闲的时候，在通过 redo log 里的日记把最新的数据同步到**磁盘**中去。

当内存数据页跟磁盘数据页内容不一致的时候，我们称这个内存页为“脏页”。内存数据写入到磁盘后，内存和磁盘上的数据页的内容就一致了，称为“干净页”。

**刷脏页有下面4种场景（后两种不用太关注“性能”问题）：**

* **redolog写满了：**redo log 里的容量是有限的，如果数据库一直很忙，更新又很频繁，这个时候 redo log 很快就会被写满了，这个时候就没办法等到空闲的时候再把数据同步到磁盘的，只能暂停其他操作，全身心来把数据同步到磁盘中去的，而这个时候，**就会导致我们平时正常的SQL语句突然执行的很慢**，所以说，数据库在在同步数据到磁盘的时候，就有可能导致我们的SQL语句执行的很慢了。
* **内存不够用了：**如果一次查询较多的数据，恰好碰到所查数据页不在内存中时，需要申请内存，而此时恰好内存不足的时候就需要淘汰一部分内存数据页，如果是干净页，就直接释放，如果恰好是脏页就需要刷脏页。
* **MySQL 认为系统“空闲”的时候：**这时系统没什么压力。
* **MySQL 正常关闭的时候：**这时候，MySQL 会把内存的脏页都 flush 到磁盘上，这样下次 MySQL 启动的时候，就可以直接从磁盘上读数据，启动速度会很快。

**2、拿不到锁我能怎么办**

这个就比较容易想到了，我们要执行的这条语句，刚好这条语句涉及到的**表**，别人在用，并且加锁了，我们拿不到锁，只能慢慢等待别人释放锁了。或者，表没有加锁，但要使用到的某个一行被加锁了，这个时候，我也没办法啊。

如果要判断是否真的在等待锁，我们可以用 **show processlist**这个命令来查看当前的状态哦，这里我要提醒一下，有些命令最好记录一下，反正，我被问了好几个命令，都不知道怎么写，呵呵。

下来我们来访分析下第二种情况，我觉得第二种情况的分析才是最重要的

**三、针对一直都这么慢的情况**

如果在数据量一样大的情况下，这条 SQL 语句每次都执行的这么慢，那就就要好好考虑下你的 SQL 书写了，下面我们来分析下哪些原因会导致我们的 SQL 语句执行的很不理想。

我们先来假设我们有一个表，表里有下面两个字段,分别是主键 id，和两个普通字段 c 和 d。

mysql> CREATE TABLE `t` (  
  `id` int(11) NOT NULL,  
  `c` int(11) DEFAULT NULL,  
  `d` int(11) DEFAULT NULL,  
  PRIMARY KEY (`id`)  
) ENGINE=InnoDB;

**1、扎心了，没用到索引**

没有用上索引，我觉得这个原因是很多人都能想到的，例如你要查询这条语句

select \* from t where 100 <c and c < 100000;

**（1）、字段没有索引**

刚好你的 c 字段上没有索引，那么抱歉，只能走全表扫描了，你就体验不会索引带来的乐趣了，所以，这回导致这条查询语句很慢。

**（2）、字段有索引，但却没有用索引**

好吧，这个时候你给 c 这个字段加上了索引，然后又查询了一条语句

select \* from t where c - 1 = 1000;

我想问大家一个问题，这样子在查询的时候会用索引查询吗？

答是不会，如果我们在字段的左边做了运算，那么很抱歉，在查询的时候，就不会用上索引了，所以呢，大家要注意这种**字段上有索引，但由于自己的疏忽，导致系统没有使用索引**的情况了。

正确的查询应该如下

select \* from t where c = 1000 + 1;

有人可能会说，右边有运算就能用上索引？难道数据库就不会自动帮我们优化一下，自动把 c - 1=1000 自动转换为 c = 1000+1。

不好意思，确实不会帮你，所以，你要注意了。

**（3）、函数操作导致没有用上索引**

如果我们在查询的时候，对字段进行了函数操作，也是会导致没有用上索引的，例如

select \* from t where pow(c,2) = 1000;

这里我只是做一个例子，假设函数 pow 是求 c 的 n 次方，实际上可能并没有 pow(c,2)这个函数。其实这个和上面在左边做运算也是很类似的。

所以呢，一条语句执行都很慢的时候，可能是该语句没有用上索引了，不过具体是啥原因导致没有用上索引的呢，你就要会分析了，我上面列举的三个原因，应该是出现的比较多的吧。

**2、呵呵，数据库自己选错索引了**

我们在进行查询操作的时候，例如

select \* from t where 100 < c and c < 100000;

我们知道，主键索引和非主键索引是有区别的，主键索引存放的值是**整行字段的数据**，而非主键索引上存放的值不是整行字段的数据，而且存放**主键字段的值**。不大懂的可以看这篇文章：  [【思维导图-索引篇】搞定数据库索引就是这么简单](https://mp.weixin.qq.com/s?__biz=Mzg2OTA0Njk0OA==&mid=2247484848&idx=1&sn=77a0e6e82944ec385f5df17e91ce3bf2&chksm=cea24a7bf9d5c36d4b289cccb017292f9f36da9f3c887fd2b93ecd6af021fcf30121ba09799f&token=1082669959&lang=zh_CN&scene=21#wechat_redirect)  里面有说到主键索引和非主键索引的区别

也就是说，我们如果走 c 这个字段的索引的话，最后会查询到对应主键的值，然后，再根据主键的值走主键索引，查询到整行数据返回。

好吧扯了这么多，其实我就是想告诉你，就算你在 c 字段上有索引，系统也并不一定会走 c 这个字段上的索引，而是有可能会直接扫描扫描全表，找出所有符合 100 < c and c < 100000 的数据。

**为什么会这样呢？**

其实是这样的，系统在执行这条语句的时候，会进行预测：究竟是走 c 索引扫描的行数少，还是直接扫描全表扫描的行数少呢？显然，扫描行数越少当然越好了，因为扫描行数越少，意味着I/O操作的次数越少。

如果是扫描全表的话，那么扫描的次数就是这个表的总行数了，假设为 n；而如果走索引 c 的话，我们通过索引 c 找到主键之后，还得再通过主键索引来找我们整行的数据，也就是说，需要走两次索引。而且，我们也不知道符合 100 c < and c < 10000 这个条件的数据有多少行，万一这个表是全部数据都符合呢？这个时候意味着，走 c 索引不仅扫描的行数是 n，同时还得每行数据走两次索引。

**所以呢，系统是有可能走全表扫描而不走索引的。那系统是怎么判断呢？**

判断来源于系统的预测，也就是说，如果要走 c 字段索引的话，系统会预测走 c 字段索引大概需要扫描多少行。如果预测到要扫描的行数很多，它可能就不走索引而直接扫描全表了。

那么问题来了，**系统是怎么预测判断的呢？**这里我给你讲下系统是怎么判断的吧，虽然这个时候我已经写到脖子有点酸了。

系统是通过**索引的区分度**来判断的，一个索引上不同的值越多，意味着出现相同数值的索引越少，意味着索引的区分度越高。我们也把区分度称之为**基数**，即区分度越高，基数越大。所以呢，基数越大，意味着符合 100 < c and c < 10000 这个条件的行数越少。

所以呢，一个索引的基数越大，意味着走索引查询越有优势。

**那么问题来了，怎么知道这个索引的基数呢？**

系统当然是不会遍历全部来获得一个索引的基数的，代价太大了，索引系统是通过遍历部分数据，也就是通过**采样**的方式，来预测索引的基数的。

**扯了这么多，重点的来了**，居然是采样，那就有可能出现**失误**的情况，也就是说，c 这个索引的基数实际上是很大的，但是采样的时候，却很不幸，把这个索引的基数预测成很小。例如你采样的那一部分数据刚好基数很小，然后就误以为索引的基数很小。**然后就呵呵，系统就不走 c 索引了，直接走全部扫描了**。

所以呢，说了这么多，得出结论：**由于统计的失误，导致系统没有走索引，而是走了全表扫描**，而这，也是导致我们 SQL 语句执行的很慢的原因。

这里我声明一下，系统判断是否走索引，扫描行数的预测其实只是原因之一，这条查询语句是否需要使用使用临时表、是否需要排序等也是会影响系统的选择的。

不过呢，我们有时候也可以通过强制走索引的方式来查询，例如

select \* from t force index(a) where c < 100 and c < 100000;

我们也可以通过

show index from t;

来查询索引的基数和实际是否符合，如果和实际很不符合的话，我们可以重新来统计索引的基数，可以用这条命令

analyze table t;

来重新统计分析。

**既然会预测错索引的基数，这也意味着，当我们的查询语句有多个索引的时候，系统有可能也会选错索引哦**，这也可能是 SQL 执行的很慢的一个原因。

好吧，就先扯这么多了，你到时候能扯出这么多，我觉得已经很棒了，下面做一个总结。

**四、总结**

以上是我的总结与理解，最后一个部分，我怕很多人不大懂**数据库居然会选错索引**，所以我详细解释了一下，下面我对以上做一个总结。

一个 SQL 执行的很慢，我们要分两种情况讨论：

1、大多数情况下很正常，偶尔很慢，则有如下原因

(1)、数据库在刷新脏页，例如 redo log 写满了需要同步到磁盘。

(2)、执行的时候，遇到锁，如表锁、行锁。

2、这条 SQL 语句一直执行的很慢，则有如下原因。

(1)、没有用上索引：例如该字段没有索引；由于对字段进行运算、函数操作导致无法用索引。

(2)、数据库选错了索引。