

当期文章

MySQL · 答疑解惑 · MySQL Sort 分页

背景

6.5号，小编在 Aliyun 的论坛中发现一位开发者提的一个问题，说 RDS 发现了一个超级大BUG，吓的小编一身冷汗 = =!!

赶紧来看看，背景是一个RDS用户创建了一张表，在一个都是NULL值的非索引字段上进行了排序并分页，用户发现第二页和第一页的数据有重复，然后以为是NULL值的问题，把这个字段都更新成相同的值，发现问题依旧。详细的信息可以登录阿里云的[官方论坛查看](#)。

小编进行了尝试，确实如此，并且5.5的版本和5.6的版本行为不一致，所以，必须要查明原因。

原因调查

在MySQL 5.6的版本上，优化器在遇到order by limit语句的时候，做了一个优化，即使用了priority queue。

参考伪代码：

```
while (get_next_sortkey())
{
    if (using priority queue)
        push sort key into queue
    else
    {
        if (no free space in sort_keys buffers)
        {
            sort sort_keys buffer;
            dump sorted sequence to 'tempfile';
            dump BUFFPEK describing sequence location into 'buffpek_pointers';
        }
        put sort key into 'sort_keys';
    }
}

if (sort_keys has some elements && dumped at least once)
    sort-dump-dump as above;
else
    don't sort, leave sort_keys array to be sorted by caller.
```

使用 priority queue 的目的，就是在不能使用索引有序性的时候，如果要排序，并且使用了limit n，那么只需要在排序的过程中，保留n条记录即可，这样虽然不能解决所有记录都需要排序的开销，但是只需要 sort buffer 少量的内存就可以完成排序。

之所以5.6出现了第二页数据重复的问题，是因为 priority queue 使用了堆排序的排序方法，而堆排序是一个不稳定的排序方法，也就是相同的值可能排序出来的结果和读出来的数据顺序不一致。

5.5 没有这个优化，所以也就不会出现这个问题。

解决方法

1. 索引排序字段

之前的月报中，我们讨论过三星索引的设计，其中第二条就是利用索引的有序性，如果用户在字段添加上索引，就直接按照索引的有序性进行读取并分页，从而可以规避遇到的这个问题。

2. 正确理解分页

还是要正确理解分页，分页是建立在排序的基础上，进行了数量范围分割。排序是数据库提供的功能，而分页却是衍生的出来的应用需求。在MySQL和Oracle的官方文档中提供了limit n和rownum < n的方法，但却没有明确的定义分页这个概念。还有重要的一点，虽然上面的解决方法可以缓解用户的这个问题，但按照用户的理解，依然还有问题：比如，这个表插入比较频繁，用户查询的时候，在read-committed的隔离级别下，第一页和第二页仍然会有重合。

分页一直都有这个问题，我们看分页常用的场景：1)早期的论坛 2)个人交易记录。这些场景都对数据分页都没有非常高的准确性要求。

究竟是不是BUG

究竟归于bug问题还是用户使用理解上的问题？

小编觉得应该分开看待这个问题，如果是排序的问题，那就算是BUG，如果是分页的这个问题，那它确实完成了order by的功能，也完成了limit n功能，那就不能说它是BUG，分页就纯粹变成了用户理解的问题了。

用户在使用数据库的时候常见的一些问题：

1. 不加order by的时候的排序问题

用户在使用Oracle或MySQL的时候，发现MySQL总是有序的，Oracle却很混乱，这个主要是因为Oracle是堆表，MySQL是索引聚簇表的原因。所以没有order by的时候，数据库并不保证记录返回的顺序性，并且不保证每次返回都一致的。

2. 分页问题

分页重复的问题，就如前面所描述的，分页是在数据库提供的排序功能的基础上，衍生出来的应用需求，数据库并不保证分页的重复问题。

3. NULL值和空串问题

不同的数据库对于NULL值和空串的理解和处理是不一样的，比如Oracle NULL和NULL值是无法比较的，既不是相等也不是不相等，是未知的。而对于空串，在插入的时候，MySQL是一个字符串长度为0的空串，而Oracle则直接进行NULL值处理。