

40 | insert语句的锁为什么这么多？

2019-02-13 林晓斌



讲述：林晓斌

时长 12:16 大小 11.25M



在上一篇文章中，我提到 MySQL 对自增主键锁做了优化，尽量在申请到自增 id 以后，就释放自增锁。

因此，**insert 语句是一个很轻量的操作**。不过，这个结论对于“普通的 insert 语句”才有效。也就是说，还有些 insert 语句是属于“特殊情况”的，在执行过程中需要给其他资源加锁，或者无法在申请到自增 id 以后就立马释放自增锁。


那么，今天这篇文章，我们就一起来聊聊这个话题。

insert ... select 语句

我们先从昨天的问题说起吧。表 t 和 t2 的表结构、初始化数据语句如下，今天的例子我们还是针对这两个表展开。

```
1 CREATE TABLE `t` (  
2   `id` int(11) NOT NULL AUTO_INCREMENT,  
3   `c` int(11) DEFAULT NULL,  
4   `d` int(11) DEFAULT NULL,  
5   PRIMARY KEY (`id`),  
6   UNIQUE KEY `c` (`c`)  
7 ) ENGINE=InnoDB;  
8  
9 insert into t values(null, 1,1);  
10 insert into t values(null, 2,2);  
11 insert into t values(null, 3,3);  
12 insert into t values(null, 4,4);  
13  
14 create table t2 like t
```

现在，我们一起来看看为什么在可重复读隔离级别下，binlog_format=statement 时执行：

 复制代码

```
1 insert into t2(c,d) select c,d from t;
```

这个语句时，需要对表 t 的所有行和间隙加锁呢？


其实，这个问题我们需要考虑的还是日志和数据的一致性。我们看下这个执行序列：

session A	session B
insert into t values(-1, -1,-1);	insert into t2(c,d) select c,d from t;

图 1 并发 insert 场景

实际的执行效果是，如果 session B 先执行，由于这个语句对表 t 主键索引加了 $(-\infty, 1]$ 这个 next-key lock，**会在语句执行完成后，才允许 session A 的 insert 语句执行。**

但如果没有锁的话，就可能出现 session B 的 insert 语句先执行，但是后写入 binlog 的情况。于是，在 binlog_format=statement 的情况下，binlog 里面就记录了这样的语句序列：

 复制代码

```
1 insert into t values(-1,-1,-1);
2 insert into t2(c,d) select c,d from t;
```


这个语句到了备库执行，就会把 id=-1 这一行也写到表 t2 中，出现主备不一致。

insert 循环写入

当然了，执行 insert ... select 的时候，对目标表也不是锁全表，而是只锁住需要访问的资源。

如果现在有这么一个需求：要往表 t2 中插入一行数据，这一行的 c 值是表 t 中 c 值的最大值加 1。

此时，我们可以这么写这条 SQL 语句：

 复制代码

```
1 insert into t2(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
```

这个语句的加锁范围，就是表 t 索引 c 上的 (3,4] 和 (4,supremum] 这两个 next-key lock，以及主键索引上 id=4 这一行。

它的执行流程也比较简单，从表 t 中按照索引 c 倒序，扫描第一行，拿到结果写入到表 t2 中。

因此整条语句的扫描行数是 1。


这个语句执行的慢查询日志（slow log），如下图所示：

```
# Query_time: 0.000732 Lock_time: 0.000356 Rows_sent: 0 Rows_examined: 1
SET timestamp=1548852517;
insert into t2(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
```

图 2 慢查询日志 -- 将数据插入表 t2

通过这个慢查询日志，我们看到 Rows_examined=1，正好验证了执行这条语句的扫描行数为 1。

那么，如果我们是想要把这样的一行数据插入到表 t 中的话：

 复制代码

```
1 insert into t(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
```

语句的执行流程是怎样的？扫描行数又是多少呢？

这时候，我们再看慢查询日志就会发现不对了。

```
# Query_time: 0.000478 Lock_time: 0.000128 Rows_sent: 0 Rows_examined: 5
SET timestamp=1548852287;
insert into t(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
```

图 3 慢查询日志 -- 将数据插入表 t

可以看到，这时候的 Rows_examined 的值是 5。

我在前面的文章中提到过，希望你都能够学会用 explain 的结果来“脑补”整条语句的执行过程。今天，我们就一起来试试。

如图 4 所示就是这条语句的 explain 结果。

```
mysql> explain insert into t(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| id | select_type | table | partitions | type | possible_keys | key | key_len | ref | rows | filtered | Extra |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 1 | INSERT      | t     | NULL       | ALL  | NULL          | NULL | NULL    | NULL | NULL | NULL     | NULL |
| 1 | SIMPLE      | t     | NULL       | index | NULL          | c   | 5       | NULL | 1    | 100.00   | Using temporary |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

图 4 explain 结果

从 Extra 字段可以看到“Using temporary”字样，表示这个语句用到了临时表。也就是说，执行过程中，需要把表 t 的内容读出来，写入临时表。

图中 rows 显示的是 1，我们不妨先对这个语句的执行流程做一个猜测：如果说是把子查询的结果读出来（扫描 1 行），写入临时表，然后再从临时表读出来（扫描 1 行），写入表 t 中。那么，这个语句的扫描行数就应该是 2，而不是 5。

所以，这个猜测不对。实际上，Explain 结果里的 rows=1 是因为受到了 limit 1 的影响。

从另一个角度考虑的话，我们可以看看 InnoDB 扫描了多少行。如图 5 所示，是在执行这个语句前后查看 Innodb_rows_read 的结果。

```
mysql> show status like '%Innodb_rows_read%';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| Innodb_rows_read | 2242 |
+-----+-----+
1 row in set (0.00 sec)

mysql> insert into t(c,d) (select c+1, d from t force index(c) order by c desc limit 1);
Query OK, 1 row affected (0.00 sec)
Records: 1 Duplicates: 0 Warnings: 0

mysql> show status like '%Innodb_rows_read%';
+-----+-----+
| Variable_name | Value |
+-----+-----+
| Innodb_rows_read | 2246 |
+-----+-----+
1 row in set (0.00 sec)
```

图 5 查看 Innodb_rows_read 变化

可以看到，这个语句执行前后，Innodb_rows_read 的值增加了 4。因为默认临时表是使用 Memory 引擎的，所以这 4 行查的都是表 t，也就是说对表 t 做了全表扫描。

这样，我们就把整个执行过程理清楚了：

1. 创建临时表，表里有两个字段 c 和 d。
2. 按照索引 c 扫描表 t，依次取 c=4、3、2、1，然后回表，读到 c 和 d 的值写入临时表。这时，Rows_examined=4。
3. 由于语义里面有 limit 1，所以只取了临时表的第一行，再插入到表 t 中。这时，Rows_examined 的值加 1，变成了 5。


也就是说，这个语句会导致在表 t 上做全表扫描，并且会给索引 c 上的所有间隙都加上共享的 next-key lock。所以，这个语句执行期间，其他事务不能在这个表上插入数据。

至于这个语句的执行为什么需要临时表，原因是这类一边遍历数据，一边更新数据的情况，如果读出来的数据直接写回原表，就可能在遍历过程中，读到刚刚插入的记录，新插入的记录如果参与计算逻辑，就跟语义不符。

由于实现上这个语句没有直接在子查询中就直接使用 limit 1，从而导致了这个语句的执行需要遍历整个表 t。它的优化方法也比较简单，就是用前面介绍的方法，先 insert into 到临时

表 temp_t，这样就只需要扫描一行；然后再从表 temp_t 里面取出这行数据插入表 t1。

当然，由于这个语句涉及的数据量很小，你可以考虑使用内存临时表来做这个优化。使用内存临时表优化时，语句序列的写法如下：

 复制代码

```
1 create temporary table temp_t(c int,d int) engine=memory;
2 insert into temp_t (select c+1, d from t force index(c) order by c desc limit 1);
3 insert into t select * from temp_t;
4 drop table temp_t;
```

insert 唯一键冲突

前面的两个例子是使用 insert ... select 的情况，接下来我要介绍的这个例子就是最常见的 insert 语句出现唯一键冲突的情况。

对于有唯一键的表，插入数据时出现唯一键冲突也是常见的情况了。我先给你举一个简单的唯一键冲突的例子。

session A	session B
insert into t values(10,10,10);	
begin; insert into t values(11,10,10); (Duplicate entry '10' for key 'c')	
	insert into t values(12,9,9); (blocked)

图 6 唯一键冲突加锁

这个例子也是在可重复读（repeatable read）隔离级别下执行的。可以看到，session B 要执行的 insert 语句进入了锁等待状态。

也就是说，session A 执行的 insert 语句，发生唯一键冲突的时候，并不只是简单地报错返回，还在冲突的索引上加了锁。我们前面说过，一个 next-key lock 就是由它右边界的值定义的。这时候，session A 持有索引 c 上的 (5,10] 共享 next-key lock（读锁）。

至于为什么要加这个读锁，其实我也没有找到合理的解释。从作用上来看，这样做可以避免这一行被别的事务删掉。

这里[官方文档](#)有一个描述错误，认为如果冲突的是主键索引，就加记录锁，唯一索引才加 next-key lock。但实际上，这两类索引冲突加的都是 next-key lock。

备注：这个 bug，是我在写这篇文章查阅文档时发现的，已经[发给官方](#)并被 verified 了。

有同学在前面文章的评论区问到，在有多个唯一索引的表中并发插入数据时，会出现死锁。但是，由于他没有提供复现方法或者现场，我也无法做分析。所以，我建议你评论区发问题的时候，尽量同时附上复现方法，或者现场信息，这样我才好和你一起分析问题。

这里，我就先和你分享一个经典的死锁场景，如果你还遇到过其他唯一键冲突导致的死锁场景，也欢迎给我留言。

	session A	session B	session C
T1	begin; insert into t values(null, 5,5);		
T2		insert into t values(null, 5,5);	insert into t values(null, 5,5);
T3	rollback;		(Deadlock found)

图 7 唯一键冲突 -- 死锁

在 session A 执行 rollback 语句回滚的时候，session C 几乎同时发现死锁并返回。

这个死锁产生的逻辑是这样的：

1. 在 T1 时刻，启动 session A，并执行 insert 语句，此时在索引 c 的 c=5 上加了记录锁。注意，这个索引是唯一索引，因此退化为记录锁（如果你的印象模糊了，可以回顾下[第 21 篇文章](#)介绍的加锁规则）。
2. 在 T2 时刻，session B 要执行相同的 insert 语句，发现了唯一键冲突，加上读锁；同样地，session C 也在索引 c 上，c=5 这一个记录上，加了读锁。

3. T3 时刻，session A 回滚。这时候，session B 和 session C 都试图继续执行插入操作，都要加上写锁。两个 session 都要等待对方的行锁，所以就出现了死锁。

这个流程的状态变化图如下所示。

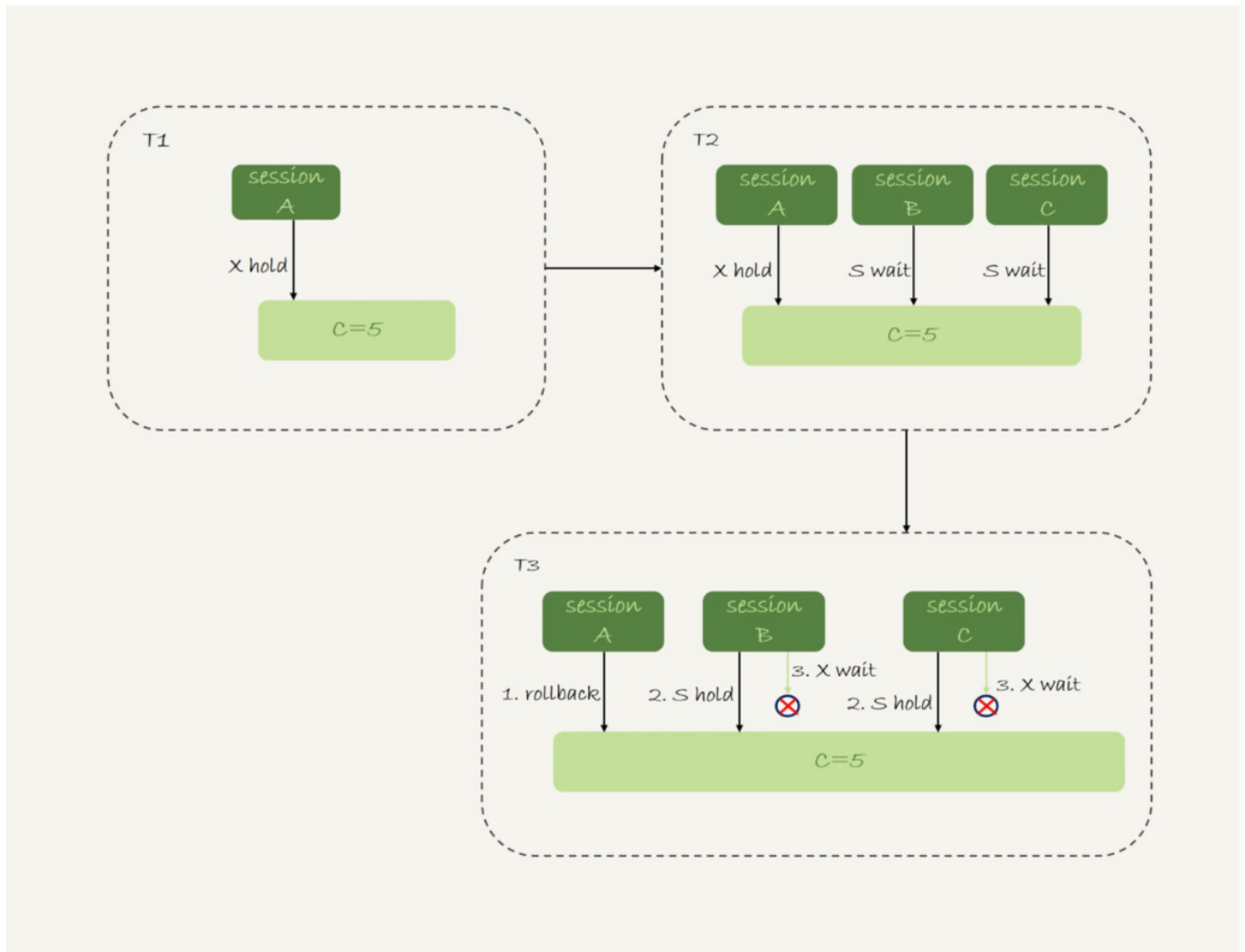


图 8 状态变化图 -- 死锁

insert into ... on duplicate key update

上面这个例子是主键冲突后直接报错，如果是改写成

[复制代码](#)

```
1 insert into t values(11,10,10) on duplicate key update d=100;
```

的话，就会给索引 c 上 (5,10] 加一个排他的 next-key lock (写锁)。

insert into ... on duplicate key update 这个语义的逻辑是，插入一行数据，如果碰到唯一键约束，就执行后面的更新语句。

注意，如果有多个列违反了唯一性约束，就会按照索引的顺序，修改跟第一个索引冲突的行。

现在表 t 里面已经有了 (1,1,1) 和 (2,2,2) 这两行，我们再来看看下面这个语句执行的效果：

```
mysql> insert into t values(2,1, 100) on duplicate key update d=100;
Query OK, 2 rows affected (0.00 sec)
mysql> select * from t where id<=2;
+----+-----+-----+
| id | c     | d     |
+----+-----+-----+
| 1  | 1     | 1     |
| 2  | 2     | 100    |
+----+-----+-----+
2 rows in set (0.00 sec)
```

图 9 两个唯一键同时冲突

可以看到，主键 id 是先判断的，MySQL 认为这个语句跟 id=2 这一行冲突，所以修改的是 id=2 的行。

需要注意的是，执行这条语句的 affected rows 返回的是 2，很容易造成误解。实际上，真正更新的只有一行，只是在代码实现上，insert 和 update 都认为自己成功了，update 计数加了 1，insert 计数也加了 1。

小结

今天这篇文章，我和你介绍了几种特殊情况下的 insert 语句。

insert ... select 是很常见的在两个表之间拷贝数据的方法。你需要注意，在可重复读隔离级别下，这个语句会给 select 的表里扫描到的记录和间隙加读锁。

而如果 insert 和 select 的对象是同一个表，则有可能会造成循环写入。这种情况下，我们需要引入用户临时表来做优化。

insert 语句如果出现唯一键冲突，会在冲突的唯一值上加共享的 next-key lock(S 锁)。因此，碰到由于唯一键约束导致报错后，要尽快提交或回滚事务，避免加锁时间过长。

最后，我给你留一个问题吧。

你平时在两个表之间拷贝数据用的是什么方法，有什么注意事项吗？在你的应用场景里，这个方法，相较于其他方法的优势是什么呢？

你可以把你的经验和分析写在评论区，我会在下一篇文章的末尾选取有趣的评论来和你一起分析。感谢你的收听，也欢迎你把这篇文章分享给更多的朋友一起阅读。

上期问题时间

我们已经在文章中回答了上期问题。

有同学提到，如果在 insert ... select 执行期间有其他线程操作原表，会导致逻辑错误。其实，这是不会的，如果不加锁，就是快照读。

一条语句执行期间，它的一致性视图是不会修改的，所以即使有其他事务修改了原表的数据，也不会影响这条语句看到的数据。

评论区留言点赞板：

@长杰 同学回答得非常准确。



MySQL 实战 45 讲

从原理到实战，丁奇带你搞懂 MySQL

林晓斌

网名丁奇
前阿里资深技术专家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得转载

上一篇 39 | 自增主键为什么不是连续的？

下一篇 41 | 怎么最快地复制一张表？

精选留言 (28)

写留言



huolang

2019-02-13

5

老师，死锁的例子，关于sessionA拿到的c=5的记录锁，sessionB和sessionC发现唯一键冲突会加上读锁我有几个疑惑：

1. sessionA拿到的c=5的记录锁是写锁吗？
2. 为什么sessionB和sessionC发现唯一键冲突会加上读锁？
3. 如果sessionA拿到c=5的记录所是写锁，那为什么sessionB和sessionC还能加c=5的...

展开 ∨

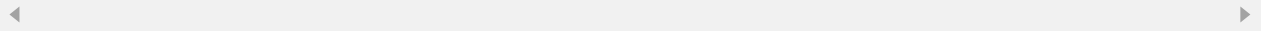
作者回复: 1. 是的

2. 这个我觉得是为了防止这个记录再被删除（不过这个理由不是很硬，我还没有找到其他解释

3. 互斥的，所以这两个语句都在等待。注意next-key lock是由间隙锁和记录锁组成的哦，间隙

锁加成功了。好问题。

4. 还没有提交，但是这个记录已经作为最新记录写进去了，复习一下08篇哈



老杨同志

2019-02-13

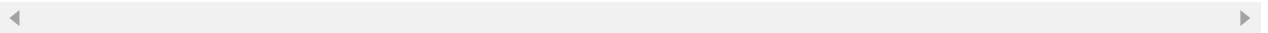
👍 4

课后问题：

我用的最多还是insert into select。如果数量比较大，会加上limit 100,000这种。并且看看后面的select条件是否走索引。缺点是会锁select的表。方法二：导出成excel，然后拼sql 成 insert into values(),(),()的形式。方法3，写类似淘宝调动的定时任务，任务的逻辑是查询100条记录，然后多个线程分到几个任务执行，比如是个线程，每个线程10...

展开 ▾

作者回复: 🍌



sonic

2019-02-14

👍 3

你好，

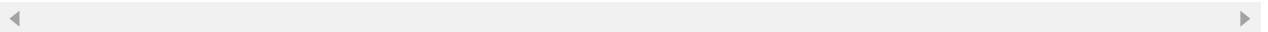
我想问下文章中关于为什么需要创建临时表有这一句话：

如果读出来的数据直接写回原表，就可能在遍历过程中，读到刚刚插入的记录，新插入的记录如果参与计算逻辑，就跟语义不符。

...

展开 ▾

作者回复: 可重复读隔离级别下，事务是可以看到自己刚刚修改的数据的，好问题



滔滔

2019-02-13

👍 2

老师，之前提到的一个有趣的问题"A、B两个用户，如果互相喜欢，则成为好友。设计上是有两张表，一个是like表，一个是friend表，like表有user_id、liker_id两个字段，我设置为复合唯一索引即uk_user_id_liker_id。语句执行顺序是这样的：

以A喜欢B为例：

1、先查询对方有没有喜欢自己（B有没有喜欢A）...

展开 ▾

作者回复: 好问题

这种情况下一般是造成锁等待，不会造成死锁吧 😊



夹心面包

2019-02-13

👍 2

1 关于insert造成死锁的情况,我之前做过测试,事务1并非只有insert,delete和update都可能造成死锁问题,核心还是插入唯一值冲突导致的.我们线上的处理办法是 1 去掉唯一值检测 2减少重复值的插入 3降低并发线程数量
2 关于数据拷贝大表我建议采用pt-archiver,这个工具能自动控制频率和速度,效果很不错...
展开 ▾

作者回复: 🐼，这两点都是很有用的建议



王伯轩

2019-02-19

👍 1

老师你好,去年双11碰到了dbcrash掉的情况.至今没有找到答案,心里渗得慌.老师帮忙分析一下.
我是一个开发,关于db的知识更多是在应用和基本原理上面,实在是找不到原因. 我也搜了一些资料 感觉像是mysql的bug,不过在其buglist中没有找到完全一致的,当然也可能是我们业务也许导致库的压力大的原因. ...
展开 ▾

作者回复: 这就是压力太大了。。一般伴随着ioutil很大，语句执行特别慢，别的语句就被堵着等锁，等超时就自己crash



phpzheng...

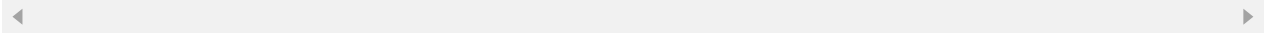
2019-02-15

👍 1

循环插入数据，然后拿着刚刚插入的主键id，更新数据。请问怎么提高这个情况的效率
展开 ▾

作者回复: insert以后
select last_insert_id,
再update ,
只能这么做啦

如果要快一些，可能可以考虑减少交互，比如写成存储过程



张永志

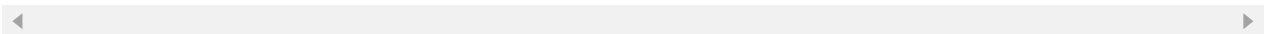
2019-02-28



对主键插入加读锁的个人理解，两个会话insert同样记录，在没有提交情况下，insert主键加读锁是为了避免第一个会话回滚后，第二个会话可以正常执行；第一个会话提交后，第二个会话再报错。

展开 ∨

作者回复: 是为了实现这个目的，是吧□



Mr.Strive...

2019-02-27



老师您好：

关于文中的锁描述有所疑惑。

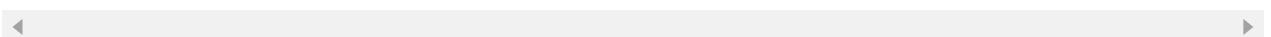
文中出现过 共享的next-key锁 和 排他的next-key锁。

...

展开 ∨

作者回复: 是这样的，gap lock是无所谓S还是X的。
但是record lock 有。

Gap lock + 排他的record 就称作 排他的next-key lock 吧☺



滔滔



2019-02-24

老师，select c+1, d from t force index(c) order by c desc limit 1;这条语句如果单独执行，是会对表t进行全表加锁，还是只加(3,4],[4,sup]这两个next key锁。还有一个问题，这里为什么要加force index(c)，不加会是怎样的效果呢？□

展开▽

作者回复: (4,sup]

以防优化器不走索引，影响我们结论（比如数据量比较小的时候）



发条橙子 ...

2019-02-23

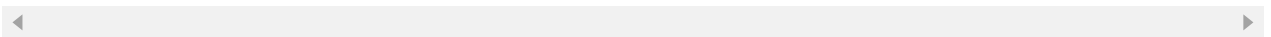


老师，年后过来狂补课程了哈哈，看到老师的bug留言已经被fix掉准备在最新版本发布了呢。

这里我有一个疑问，我之前以为只有更新的时候才会加锁，参考前面的文章，innodb要先扫描表中数据，被扫描到的行要加锁。...

展开▽

作者回复: 不是都会，是在要写入的时候，发现有主键冲突，才会加上这个next-key lock的锁



滔滔

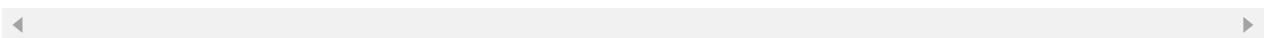
2019-02-21



老师，有个问题insert into ... on duplicate key update语句在发生冲突的时候是先加next key读锁，然后在执行后面的update语句时再给冲突记录加上写锁，从而把之前加的next key读锁变成了写锁，是这样的吗？

展开▽

作者回复: 不是，发现冲突直接加的就是写锁



王伯轩

2019-02-18



内存锁 大大计划讲下么,实际中碰到内存锁被持有后一直不释放导致db直接crash掉

展开 ▾

作者回复: 这个系列里没讲到了

这种我碰到比较多的是io压力特别大，导致有的事务执行不下去，但是占着锁

然后其他事务就拿不到锁，有一个600计时，超过就crash了



信信

2019-02-17

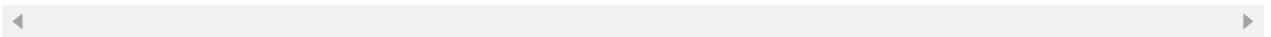


老师好，文中提到：`insert into t2(c,d) (select c+1, d from t force index(c) order by c desc limit 1)`的加锁范围是表 t 索引 c 上的 $(4, \text{supremum}]$ 这个 next-key lock 和主键索引上 $\text{id}=4$ 这一行。

可是如果我把表t的id为3这行先删除，再执行这个`insert...select`，那么别的会话执行`insert into t values(3,3,3)`会被阻塞，这说明4之前也是有间隙锁的？...

展开 ▾

作者回复: 你说的对，这里其实是“向左扫描”，加锁范围应该是 $(3,4]$ 和 $(4, \text{supremum}]$ 。



Justin

2019-02-15



插入意向锁的gal lock和next key lock中的 gaplock互斥吗？

展开 ▾

作者回复: 额，

这里我们要澄清一下哈

只有一个gap lock，就是 $\text{next key lock} = \text{gap lock} + \text{record lock}$ ；

我们说一个insert语句如果要插入一个间隙，而这个间隙上有gap lock的话，insert语句会被堵住，这个被堵住的效果，实现机制上是用插入意向锁和gap lock相互作用来实现的。

gap lock并不属于插入意向锁的一部分，就没有“插入意向锁的gal lock”这个概念哈





Justin

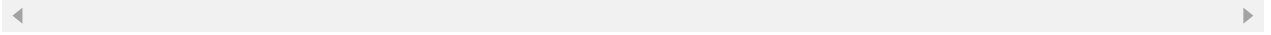
2019-02-15



那如果是冲突时 会加next key lock的话 间隙锁和插入意向锁不会冲突吗？

展开 ∨

作者回复: 如果发生冲突，就不是插入在间隙里面，这里就还没到“插入意向锁”的逻辑哈



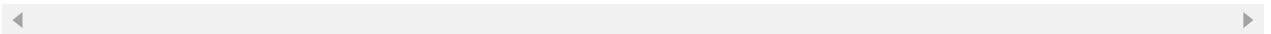
Justin

2019-02-15



为什么insert 还会使用到next key lock 呢，我记得我原来看的资料写的是插入使用的是插入意向锁啊

作者回复: 是说插入碰到唯一键冲突的时候才会哈



夹心面包

2019-02-15



我来补充应用表空间迁移的场景

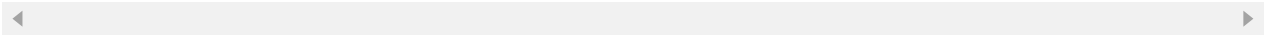
1 冷数据表的复制和迁移

2 大表数据的恢复,线上DDL操作失误,需要恢复时,利用备份+binlog进行恢复后,表空间迁移进行导入

对于热表数据的复制建议还是采用pt-archiver慢慢搞

展开 ∨

作者回复: 🍷



信信

2019-02-14



老师好，

图6下方“发生主键冲突的时候”是不是应该改为“发生唯一键冲突的时候”？因为c不是主键。

还有，图7下方：T2时刻session b 发现“唯一键冲突”，这里为啥不是锁冲突？因为如果没有锁冲突，仅有唯一键冲突，就对应图6的情况，这时加的是next-key lock，而不仅...

展开 ∨

作者回复: 1. 你说得对，👉细致，发起勘误了哈

2. 图7 这里说的唯一键冲突，就是发现“已经有一个c=5的行存在”，所以转为加next-key lock，没有单独的加行锁的逻辑哈



cyberbit

2019-02-14



在表件迁移拷贝数据，用的pt-archiver工具做的，但是它不支持utf8mb4字符集，这个问题一直困扰我，不知道怎么解决。