

图文 81 如何基于RocketMQ的数据过滤机制，提升订单数据库同步的处理效率

383 人次阅读 2020-01-17 07:00:00

[详情](#) [评论](#)

狸猫技术

[如何基于RocketMQ的数据过滤机制，提升订单数据库同步的处理效率](#)

The banner features large blue text '从零开始' (From Zero to Hero) and 'MySQL实战优化高手' (MySQL Practical Optimization Expert). Below the main title, it says '带你成为' (Lead you to become) and '继爆款专栏《从零开始带你成为JVM实战高手》后，又一重磅力作' (After the blockbuster column 'From Zero to Hero Lead you to become JVM Practical Expert', another重量级 work). It also mentions '专栏作者：救火队队长 阿里资深技术专家' (Column author: Firefighting Team Leader, Senior Technical Expert at Alibaba). On the right side, there's a sidebar titled '相关频道' (Related Channels) with a thumbnail for '从0开始' (From Zero to Hero) which has been updated 9 times.

继《从零开始带你成为JVM实战高手》后，又一重磅力作：

(点击下方蓝字试听)

[《从零开始带你成为MySQL实战优化高手》](#)

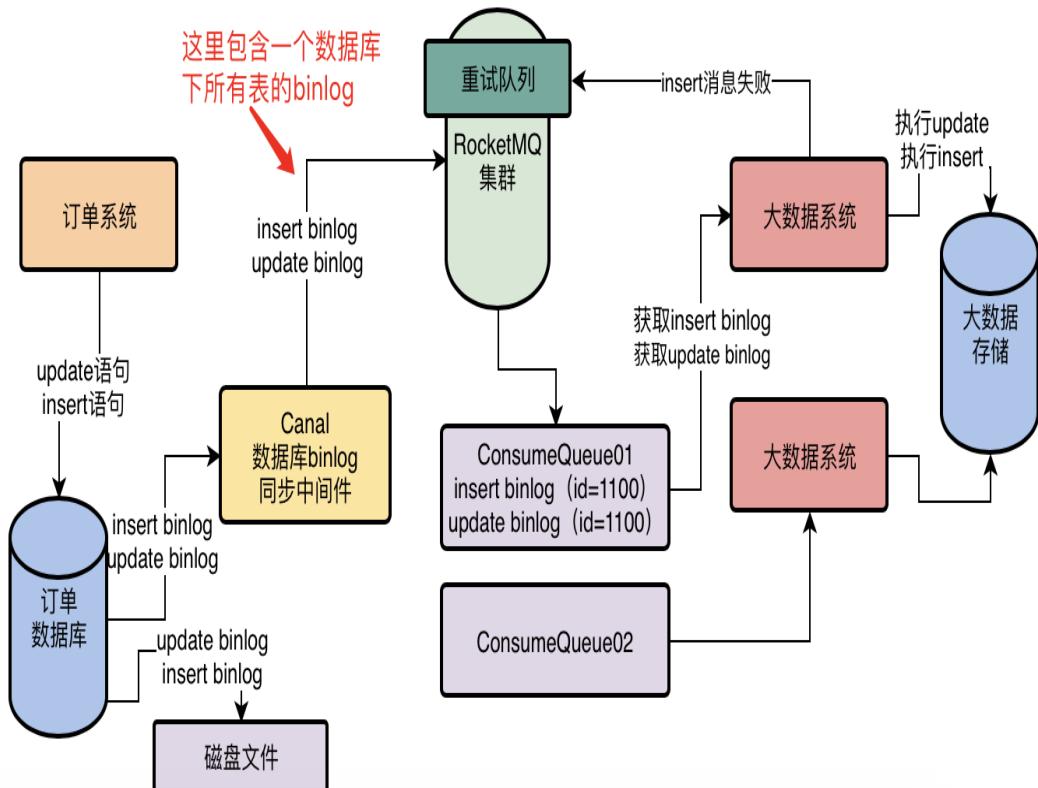
1、混杂在一起的订单数据库的binlog

我们已经花费了一些篇幅讲完了消息顺序方案了，那么我们现在就接着订单数据库同步的这个场景，来简单的看一下如何对混杂在一起的数据进行过滤的方案。

首先我们都应该知道，一个数据库中可能会包含很多表的数据，比如订单数据库，他里面除了订单信息表以外，可能还包含很多其他的表。

所以我们在进行数据库binlog同步的时候，很可能是把一个数据库里所有表的binlog都推送到MQ里去的！

我们看下面的图



所以在MQ的某个Topic中，可能是混杂了订单数据库里几个甚至十几个表的binlog数据的，不一定仅仅包含我们想要的表的binlog数据！

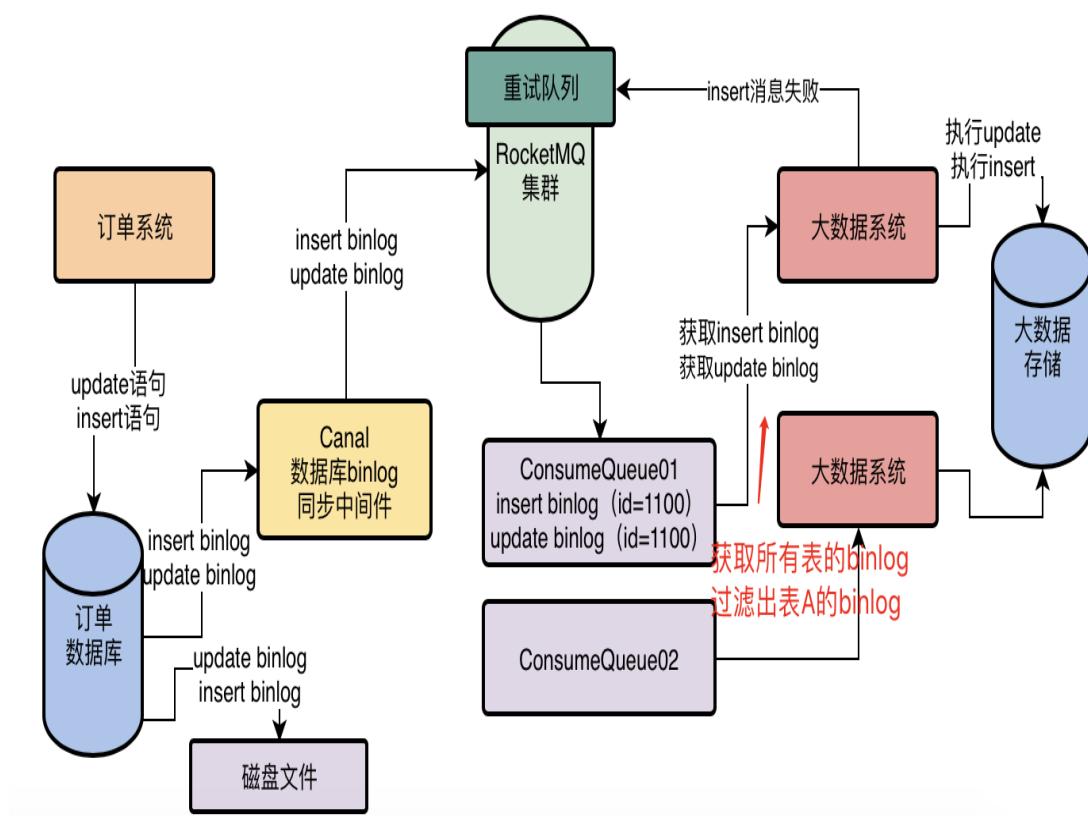
2、处理不关注的表的binlog，有多么浪费时间！

慕课都有联系微信642600657

那么此时假设我们的大数据系统仅仅关注订单数据库中的表A的binlog，并不关注其他表的binlog，那么大数据系统可能需要在获取到所有表的binlog之后，对每条binlog判断一下，是否是表A的binlog？

如果不是表A的binlog，那么就直接丢弃不要处理；如果是表A的binlog，才会去进行处理！

但是这样的话，必然会导致大数据系统处理很多不关注的表的binlog，也会很浪费时间，降低消息的效率，我们看下图



3、在发送消息的时候，给消息设置tag和属性

针对这个问题，我们可以采用RocketMQ支持的数据过滤机制，来让大数据系统仅仅关注他想要的表的binlog数据即可。

首先，我们在发送消息的时候，可以给消息设置tag和属性，我们看下面的代码。

```
Message msg = new Message()  
    "TopicOrderDbData", // 这是我们订单数据库写入的Topic  
    "TableA", // 这是这条数据的tag，可以是表的名字  
    ("binlog").getBytes(RemotingHelper.DEFAULT_CHARSET) // 这是一条binlog数据  
);  
// 我们可以给一条消息设置一些属性  
msg.putUserProperty("a", 10);  
msg.putUserProperty("b", "abc");
```

上面的代码清晰的展示了我们发送消息的时候，其实是可以给消息设置tag、属性等多个附加的信息的。

4、在消费数据的时候根据tag和属性进行过滤

接着我们可以在消费的时候根据tag和属性进行过滤，比如我们可以通过下面的代码去指定，我们只要tag=TableA和tag=TableB的数据。

```
consumer.subscribe("TopicOrderDbData", "TableA || TableB");
```

或者我们也可以通过下面的语法去指定，我们要根据每条消息的属性的值进行过滤，此时可以支持一些语法，比如：

```
consumer.subscribe(  
    "TopicOrderDbData",  
    MessageSelector.bySql("a > 5 AND b = 'abc'");  
    // 只要a大于5, b等于abc的数据
```

慕课都有联系微信642600657

RocketMQ还是支持比较丰富的数据过滤语法的，如下所示：

- (1) 数值比较，比如：>, >=, <, <=, BETWEEN, =;
- (2) 字符比较，比如：=, <>, IN;
- (3) IS NULL 或者 IS NOT NULL;
- (4) 逻辑符号 AND, OR, NOT;
- (5) 数值，比如：123, 3.1415;
- (6) 字符，比如：'abc'，必须用单引号包裹起来；
- (7) NULL，特殊的常量
- (8) 布尔值，TRUE 或 FALSE

5、基于数据过滤减轻Consumer负担

今天学习了这块知识后，我们以后就知道在使用MQ的时候，如果MQ里混杂了大量的数据，可能Consumer仅仅对其中一部分数据感兴趣，此时可以在Consumer端使用tag等数据过滤语法，过滤出自己感兴趣的数据来消费。

End

专栏版权归公众号**狸猫技术窝**所有

未经许可不得传播，如有侵权将追究法律责任

狸猫技术窝精品专栏及课程推荐：

[《从零开始带你成为JVM实战高手》](#)
[《21天互联网Java进阶面试训练营》（分布式篇）](#)
[《互联网Java工程师面试突击》（第1季）](#)
[《互联网Java工程师面试突击》（第3季）](#)

重要说明：

如何提问：每篇文章都有评论区，大家可以尽情留言提问，我会逐一答疑

如何加群：购买狸猫技术窝专栏的小伙伴都可以加入狸猫技术交流群，一个非常纯粹的技术交流的地方

具体加群方式，请参见目录菜单下的文档：《付费用户如何加群》（[购买后可见](#)）

Copyright © 2015-2020 深圳小鹅网络技术有限公司 All Rights Reserved. [粤ICP备15020529号](#)

 小鹅通提供技术支持

慕课都有联系微信642600657