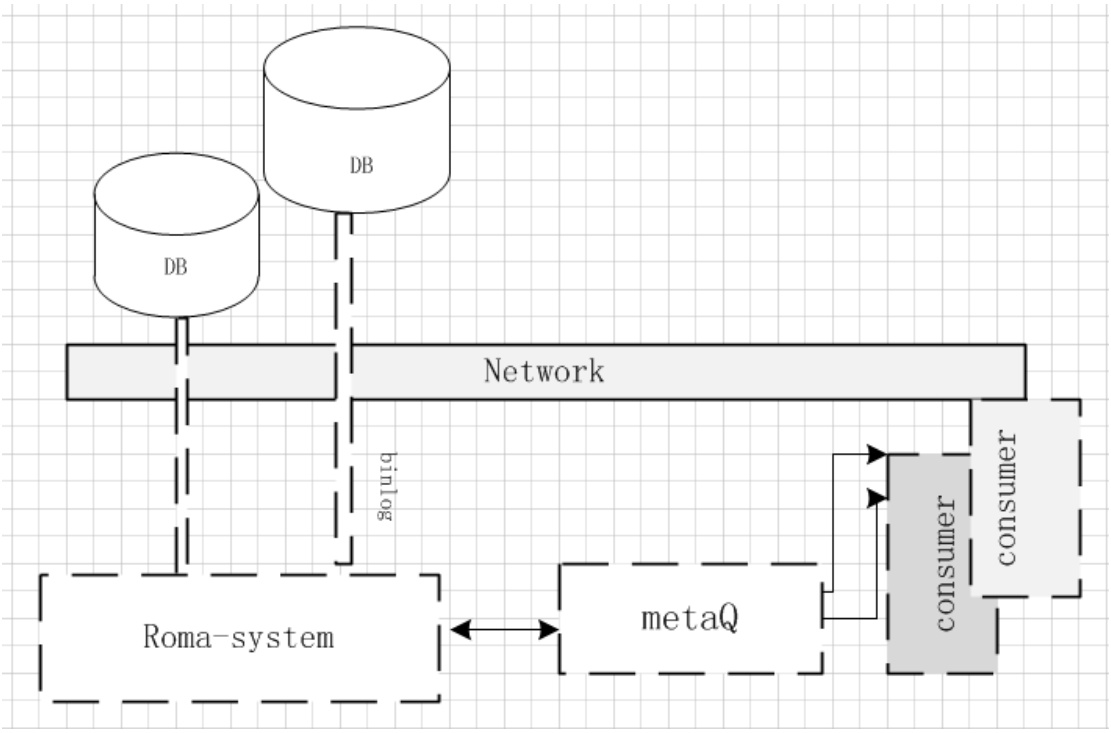
MYSQL 分布式消息的处理

在很多 MYSQL 环境中，对于 MYSQL 的分布式事物处理一直是个难题，在当前互联网环境中，大多数应用系统是基于 SOA 的很多复杂接口之间的调用，并且事物之间的处理优先级也是有先后的，所以对于实际入库的数据而言，不同的系统，对于当前入库的处理方式是不一样的，这样就衍生出了对于订阅 MYSQL 消息的需求。

在公司内部，这套分布式消息系统负责了各个子接口之间数据的衔接，同时肩负后端 DW 数据仓库的实时消息计算，多数的 RDBMS 数据，被分解成各种子消息队列，通过不同的 topic 被各种消费者订阅。

1. 如何分解消息

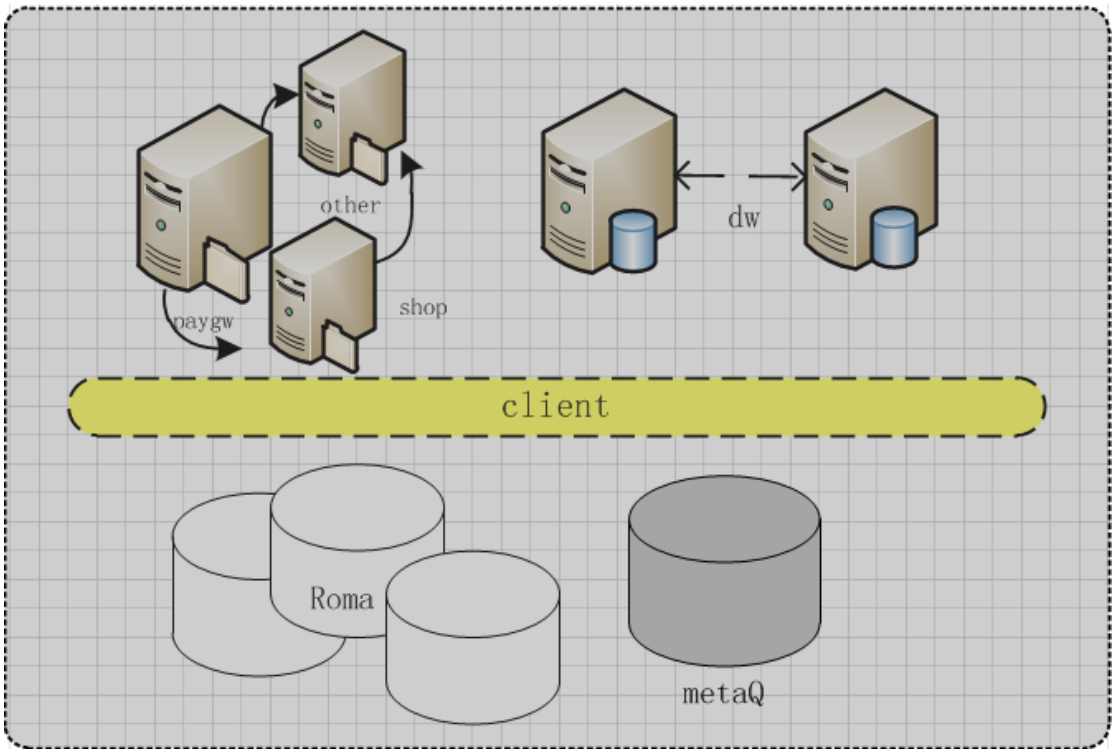
后端订阅程序（基于阿里巴巴的 canal）通过解析不同应用的 binlog (mysql 线上产生的二进制日志) 通过模拟 slave 的行为，将 binlog 顺序的订阅到本地，通过内部解析程序，将 binlog events 解析成对应的消息，通过 MetaQ 固化解析完成的消息，自定义存放时间，从而让 consumer 自行订阅到对应的系统，进行相关处理。



具体 roma 文档可以参考我的 blog:
http://www.vmcd.org/docs/roma_system.pdf

2. 何时订阅

通常当支付系统需要做异步分布式事务调用的时候，可以采用 roma 消息。采用水平拆分 DB 而需要一些统计类的需求的时候（合表） 可以订阅合并的 topics。当需要一个汇总的数据仓库，执行跨库 join 查询的时候 可以订阅 roma 消息。

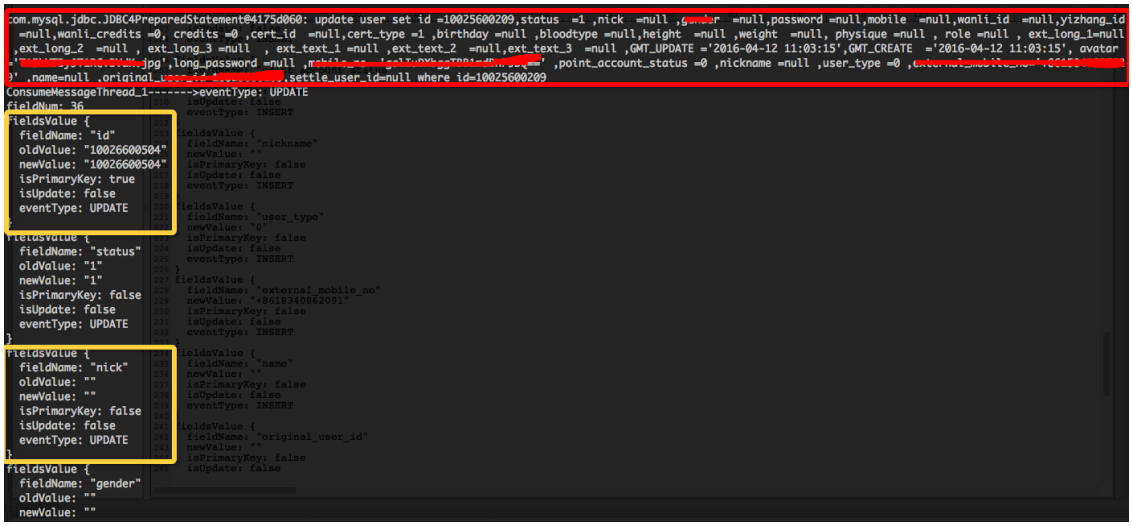


上图中，各类系统通过 RPC 框架进行异步调用，同时将订阅到的消息（roma 异步消息）进行相处理，将操作类型，操作细节发送给对应子系统，从而实现了操作的异步化（而 roma 对于前端数据库日志的实时解析保证了事物消息的实时性）。

3. 对于数据仓库

在我们的系统中，很多核心表被水平拆分成了 N 份，对于后端实时数据仓库来说，希望通过合并所有的拆分表，进行多维度的查询工作 （对 job 来说，可以通过定期任务抽取水平拆分的表，但是实时性是滞后的）。

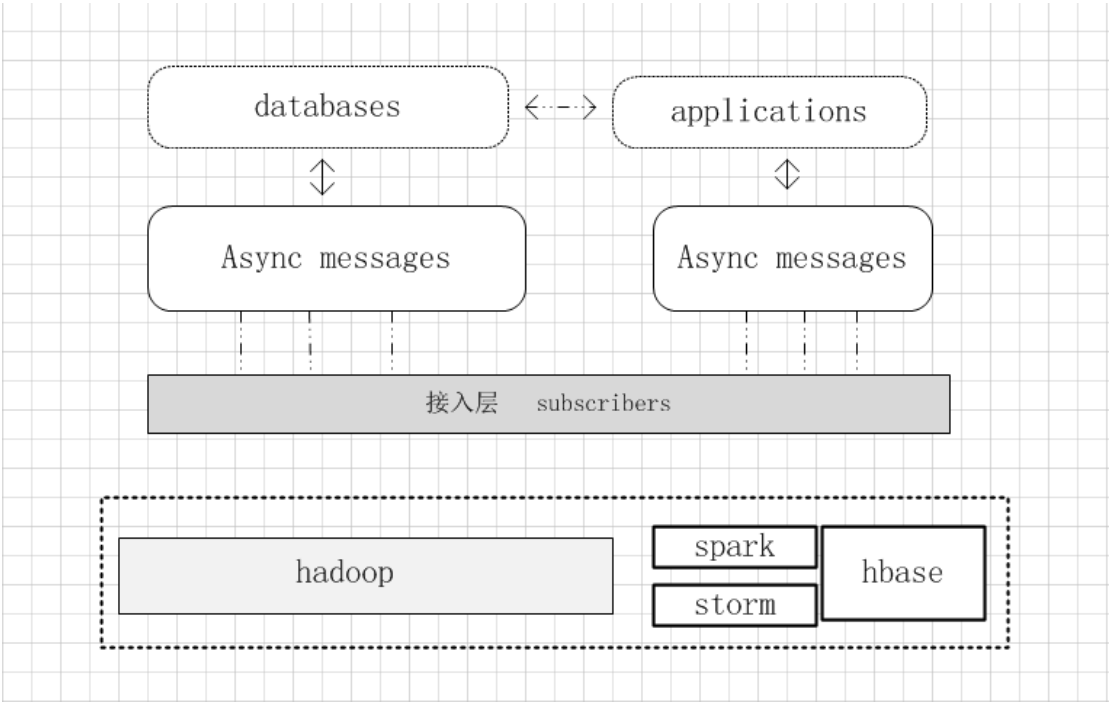
在中转服务器上，使用 java 程序直接订阅 roma 的消息，拼接成相应的 SQL 在后端 DW 上直接执行。



通过订阅同步消息，将前端更新实时同步到后端的数据仓库，从而达到实时分析的需求。后期结合 binlog server 的改进还可以进行所有系统的 binlog 集中化分层订阅。
具体可以参考：
https://www.mariadb.com/blog/binlog-server

4. 对于实时分析平台

同样可以订阅前端 RDBMS 操作到后端大数据平台，通过流式计算实现秒级的分析。



后期需要改进的：

roma 的订阅能力，对于前端 log 并发解析的粒度
智能的存储策略 动态调整没有被订阅消息的保存时间

Louis liu (www.vmcd.org)

平安健康互联网-数据库架构