

# PhxQueue

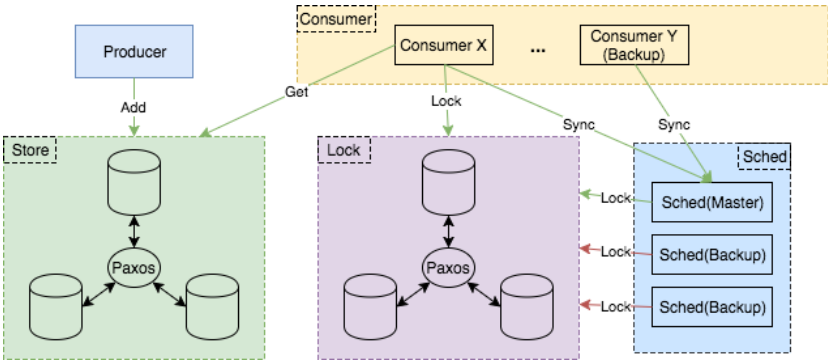
ryanyycao

## 1. PhxQueue

PhxQueue Paxos At-Least-Once Delivery / /

## 2. PhxQueue

•



• Store -

Store PhxPaxos Paxos 3  
paxos group paxos group master master Store master  
Store , plog  
Store paxos group Group Commit Paxos  
paxos group

• Producer -

Producer key key

• Consumer -

Consumer Store Consumer Topic Handler

• Scheduler -

Scheduler , Consumer Consumer Consumer  
 Scheduler Consumer  
 master , master master master failover

• Lock -

Lock Paxos PhxQueue Scheduler leader Con-  
 sumer

### 3. PhxQueue VS Kafka

#### 3.1.

特性	Kafka	PhxQueue	备注
刷盘方式	异步为主，支持同步（但是会出现写放大降低吞吐量）	同步刷盘	
水平扩展最小粒度	Partition	Queue	
物理文件存储粒度	Partition	Paxos group	PhxQueue 有意区分 queue 和 paxos group 的概念。一个 queue 只属于一个 paxos group，一个 paxos group 内可包含多个 queue，从而避免逻辑水平扩展影响写盘并发
存储层选举	Broker 依赖 Zookeeper 选举出 Controller，再由 Controller 选举出各分区的 leader	Store 自身依赖 Paxos 选举 master	Kafka Broker 引入 Controller 解决了 Zookeeper 压力大的问题；PhxQueue Store 不依赖外部选举，每一组均能独立进行 Paxos 决议，分散了选举压力
批量生产能力	仅 Producer 有 batch 逻辑	Producer、Store 均有独立的 batch 逻辑	PhxQueue 为了应对高输入场景下 Producer 端 batch 效果不好的情况，在 Store 中加入了 batch 逻辑
同步延迟	全同步协议，所有 ISR 返回 ack 后完成同步，延迟取决于最慢节点	Paxos 协议，多数派 accept 即完成同步，最慢节点不影响整体吞吐	PhxQueue 只需多数派应答即可完成同步，Kafka 同步需要等待最慢节点
存储层的组间容灾隔离	无	有	对于 PhxQueue，Paxos 只要求组内多数节点存活即可完成写入，所以单节点离线造成的失败，可以在组内换节点重试成功；对于 Kafka，单节点离线会造成整组暂时不可写，重试逻辑需要换组进行
存储层的服务发现	通过 Metadata RPC 获取存储层信息	通过本地配置文件获取存储层信息	PhxQueue 通过配置文件做服务发现，使用者需维护各机器配置的一致性；Kafka 以 Zookeeper 作为配置管理中心
消费隔离	以消费分组（Consumer Group）为单位	以订阅（Sub）为单位	
消费管理	由 Coordinator 对每个 Consumer Group 选举出一个 Consumer 作为 leader，决定该 Consumer Group 内的队列分配	各 Consumer 与 Scheduler 维持心跳并上报负载，Scheduler 根据负载调整各 Consumer 的消费权重，Consumer 再根据消费权重决定要处理的队列	PhxQueue 新增了负载均衡功能，当某 Consumer 负载过高时，可自动调整分配；Kafka 并无该功能

#### 3.2.

• b70\*3

cpu	48 * Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz
memory	128G
network	10 Gigabit Ethernet
cluster nodes	3

- kafka qps 13 / 230ms PhxQueue kafka qps IOPS  
 kafka Memory Mapped Files PhxQueue wal s
- sata kafka PhxQueue PhxQueue ssd kafka ssd PhxQueue  
 cpu 70% sata PhxQueue Kafka

### 3.3. failover

- **Kafka** Failover 0% ~ 33% Failover 10s
- **PhxQueue** Failover 66% Failover 5s Failover  
90+%

### 3.4. API

- **Kafka** API  
**Producer API** topics  
**Consumer API** topics  
**Streams API** input topics output topics  
**Connect API** kafka kafka  
**AdminClient API** topics brokers kafka
- **Phxqueue** bg API  
 svrkit evcpubid Api, Public Commit/RollBack  
 http httpsvr logicsvr svrkit  
 http httpsvr logicsvr svrkit

## 4.

PhxQueue Paxos Store batch Producer batch  
Kafka