Kafka实际案例问题

<https://www.jianshu.com/p/3d8a415e673f>

# 1 kafka面试题

## 基本题

**1 kafka和传统消息中间件的区别**

1 架构模型方面

RabbitMQ遵循AMQP协议，RabbitMQ由Exchange,Binding,queue组成，consumer从queue获取消息进行消费；rabbitMQ以broker为中心，有消息确认机制。

kafka遵从一般MQ结构，producer，broker，consumer；以consumer为中心，无消息确认机制。

2 吞吐量

kafka具有高的吞吐量，内部采用消息的批量处理，效率高

rabbitMQ在吞吐量方面稍逊于kafka，rabbitMQ支持对消息的可靠的传递，支持事务，不支持批量的操作；

3 可用性

rabbitMQ支持miror的queue，主queue失效，miror queue接管。   
 kafka的broker支持主备模式。

4 集群负载均衡

kafka采用zookeeper对集群中的broker、consumer进行管理，可以注册topic到zookeeper上；通过zookeeper的协调机制，producer保存对应topic的broker信息，可以随机或者轮询发送到broker上；

**2 kafka丢失和重复消费数据**

<https://www.cnblogs.com/kaleidoscope/p/9763053.html>

**1、丢包问题**：消息推送服务，每天早上，手机上各终端都会给用户推送消息，这时候流量剧增，可能会出现kafka发送数据过快，导致服务器网卡爆满，或者磁盘处于繁忙状态，可能会出现丢包现象。

解决方案：首先对kafka进行限速， 其次启用重试机制，重试间隔时间设置长一些，最后Kafka设置acks=all，即需要相应的所有处于ISR的分区都确认收到该消息后，才算发送成功。

检测方法：使用重放机制，查看问题所在。

**3 Apache Kafka介绍**

Apache Kafka是由Apache开发的一种发布订阅消息系统，它提供分布式的、分区的和可重复的日志服务

**4 传统的消息传递方法?**

排队：在队列中，一组用户可以从服务器中读取消息，每条消息都发送给其中一个人。

发布-订阅：在这个模型中，消息被广播给所有的用户。

**5 请说明Kafka相对传统技术有什么优势?**

快速:单一的Kafka代理可以处理成千上万的客户端，每秒处理数兆字节的读写操作。

可伸缩:在一组机器上对数据进行分区和简化，以支持更大的数据

持久:消息是持久性的，并在集群中进行复制，以防止数据丢失。

设计:它提供了容错保证和持久性

6 kafka中broker的意义

Kafka集群包含多台服务器，一台Kafka服务器就是一个Broker，一个集群由多个broker组成，一个broker可以有多个topic。broker承担着中间缓存和分发的作用，broker将producer发送的数据分发注册到consumer中。

7 zookeeper的作用

1）注册：Broker注册、Topic注册、消费者注册

2）负载均衡：生产者负载均衡、消费者负载均衡

3）分区与消费者的关系：kafka中规定了每个消息分区只能被同组的一个消费者进行消费，因此，需要在 Zookeeper 上记录消息分区与Consumer 之间的关系。

4）消息、消费进度Offset 记录

kafka中需要定时地将分区消息的消费进度Offset记录到Zookeeper上。