

1.Kafka: 是由 Linked 开源提供的分布式消息队列,由 scala 语言写成的
消息队列的作用: 解耦 异步 并行

2.传统的消息队列给予 pub sub 发布和订阅

Kafka 消息队列基于:push pull 过程 推送和拉取

3.构架模型

Producer 生产者,生产消息的 主要接收一些外部的数据源 从外部获取数据

我们可以从 flume 中获取数据

kafkaAPI 生产数据,通过 push 的方式主动将数据推送到 kafka 的 topic 当中去

topic:主题,说白了里面就是一些各种数据

partition:消息的分区,解决横向扩展问题.为了解决 partition 丢失问题,引用了一个副本机制.

Broker:一个服务器叫做一个 broker;

Consumer:消费者,主要用来消费数据的,主动的 pull 到 topic 拉取数据

Zookeeper:为了解决消费者消费数据的时候,确定 topic 中到底有多少个 parttion,都在哪些机器上.

Kafka 消费模型:组的概念,同一时间,一个组中,只能有一个线程去消费 parttion 中的数据

Partition 里面包含了多个 segment,segment 里面两个文件.log 文件 .index

.log 记录了我们的数据,文件是顺序读写的

.index 文件,记录了.log 文件的索引

Offset:消息的偏移量,我们消费数据的时候,都要记录消息的 offset,下次再消费的时候,我们就可以确定数据该从哪里进行消费