1. 什么是kafka,(what),一般什么场景使用它?（when）

官网上是这么介绍的“kafka是一个分布式的流式处理平台”，那么这样一句话怎么去理解呢？什么有事流式处理平台呢？

一个流式处理平台是有这么三个核心能力的

* 1. 允许发布和订阅记录流。在这方面类似消息队列和企业级的消息系统
  2. 允许以容错的方式存储记录流。
  3. 允许以流的形式处理记录。

基于以上的几个特点，kafka常用语以下场景，

* 消息系统
* 流处理
* 应用监控
* 网站用户行为追踪
* 持久性日志

除了前两者，后面的功能大多也是基于前两者的功能，然后结合其他的相关技术实现的，

1. Kafka中的几个基本概念

首先看几张官网上的图：



从这个图中我们知道kafka中又这么几个点，

1. Topic: topic是kafka中比较核心的一个概念，它相当于是发布消息的类别，每个消息都有一个topic,每个topic都会有多个partition（记住这里的partition 是针对topic的）,每个分区中的数据是有序的，是不可变的记录顺序追加的commit log（log 文件），分区中的每个记录都有一个连续的ID,称之为offset,唯一表示分区内的记录
2. Producers: producer是生产者，往kafka中存放数据的，producer往kafka存放数据的方式是推送的方式
3. Broker

