**Kafka**

**是Linkedin公司开发，scala语言编写，是一个分布式，支持多分区，多副本的，基于zookeeper协调的分布式消息系统。**

**最大特点：实时的处理大量数据已满足各种需求场景**

比如：基于hadoop的批处理系统，低延迟的实时系统，storm/spark流式处理引擎，

Web/nginx日志，访问日志，消息服务等

消息队列的性能好坏，其文件存储机制设计是衡量一个消息队列服务技术水平和最关键指标之一。

**主要功能：**

1. 发布订阅消息流
2. 以容错的方式记录消息流，以文件的方式来存储
3. 在消息发布的时候进行处理

角色：

Producer：生产者，向kafka集群发送消息，发送前会对消息进行分类

Topic：主题，通过对消息制定主题可以对消息进行分类

Consumer：消费者，与kafka集群建立长连接的方式，不断从集群中拉取消息，对这些消息进行处理

**Kafka服务器消息存储策略：**

**分区：partitions**

**创建一个topic时，同时可以指定分区的数目，**