# 1 Kafka基础概念

## 消息

消息是Kafka的数据单元，由字节数组组成（类似Java中的byte[]）。

### 键

键是消息的一个可选的元数据，也是由字节数组组成。

### 偏移量

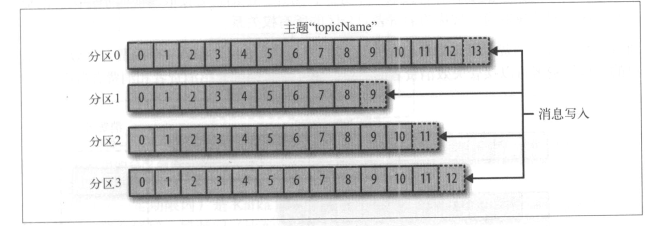
偏移量是消息的一个元数据，在消息创建时由Kafka添加到消息中，它是一个不断增加的整数值。

（注意：同个分区中每个消息的偏移量都是唯一的）

## 分区

一个主题被分为多个分区，同一个分区中的消息以先入先出的顺序读取。

图1



### 一个主题的多个分区横跨多个服务器

一个主题的不同分区可以部署在不同的服务器上，所以说一个主题可以横跨多个服务器。

### 一个分区同时只能被一个消费者读取

一个分区同时只能被一个消费者读取，但一个消费者可以读取多个分区。

### 分区的所有权关系

分区的所有权关系是指：消费者和分区之间的映射关系。

### 分区复制

如果把一个分区分配给多个broker则会发生分区复制。分区复制冗余了分区的消息，但当一个broker失效时，其他broker的分区可以使用。

## 主题

主题就是一个分类。消息类似于表中的一条记录，主题类似于表。

## 生产者

生产者默认情况下，将消息均匀分布在一个主题的所有分区上。

（注意：也可以指定写入某一个分区上）

## 消费者

消费者订阅一个或多个主题，并按照消息生成的顺序读取（因为同个分区中的消息是先入先出的）。

### 消费者读取消息偏移量

消费者会把每个分区的最后读取的消息偏移量保存到Zookeeper或Kafka上，如果消费者关闭，读取状态不会丢失。

## broker

一个独立的Kafka服务器称为broker。

### broker和集群的关系

每个集群中会有一个broker充当集群控制器的角色（从集群的活跃成员中选举出来）。

### broker和分区的关系

一个分区属于一个broker（该broker称为该分区的头领），如果把一个分区分配给多个broker则会发生分区复制。