

# Kafka10 消费者

## 消费方式

- consumer 采用 pull(拉)模式从 broker 中读取数据

原因: **push**(推)模式很难适应消费者速率不同的消费者, 因为消息发送速率是由**broker**决定的。它的目标是尽可能以最快速度传递消息, 但是这样很容易造成**consumer**来不及处理消息, 典型的表现就是拒绝服务以及网络拥塞, 而**pull**模式则可以根据**consumer**的消费能力以适应当前的速率消费信息。

不足: 如果Kafka没有数据, 消费者可能会陷入循环中, 一直返回空数据。针对这一点, **kafka**的消费者在消费数据时会传入一个时长参数**timeout**, 如果当前没有数据可供消费, **consumer**会等待一段时间之后再返回, 这段时间即为**timeout**;

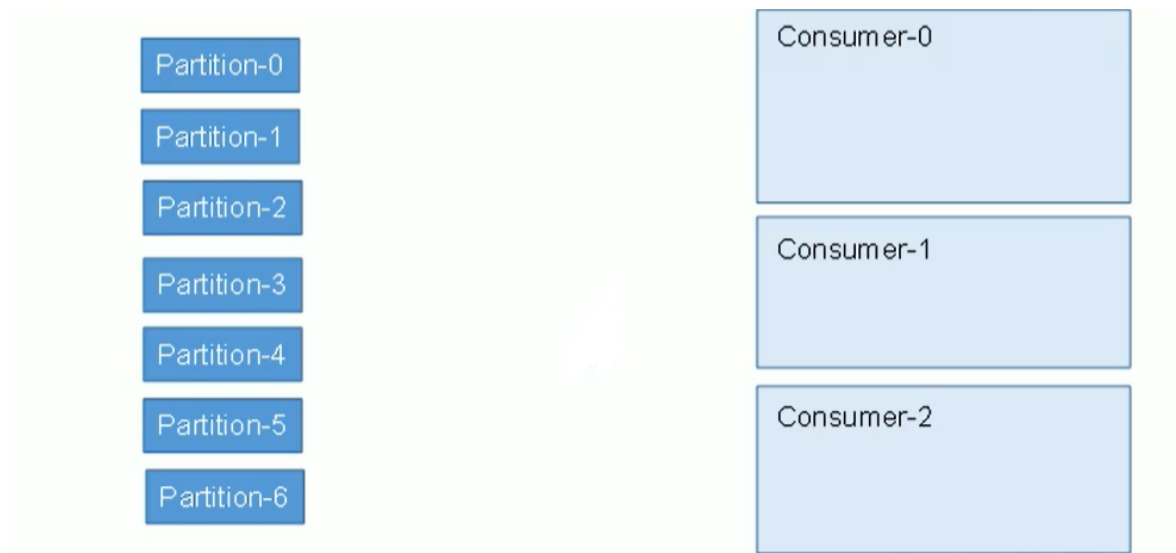
## 分区分配策略

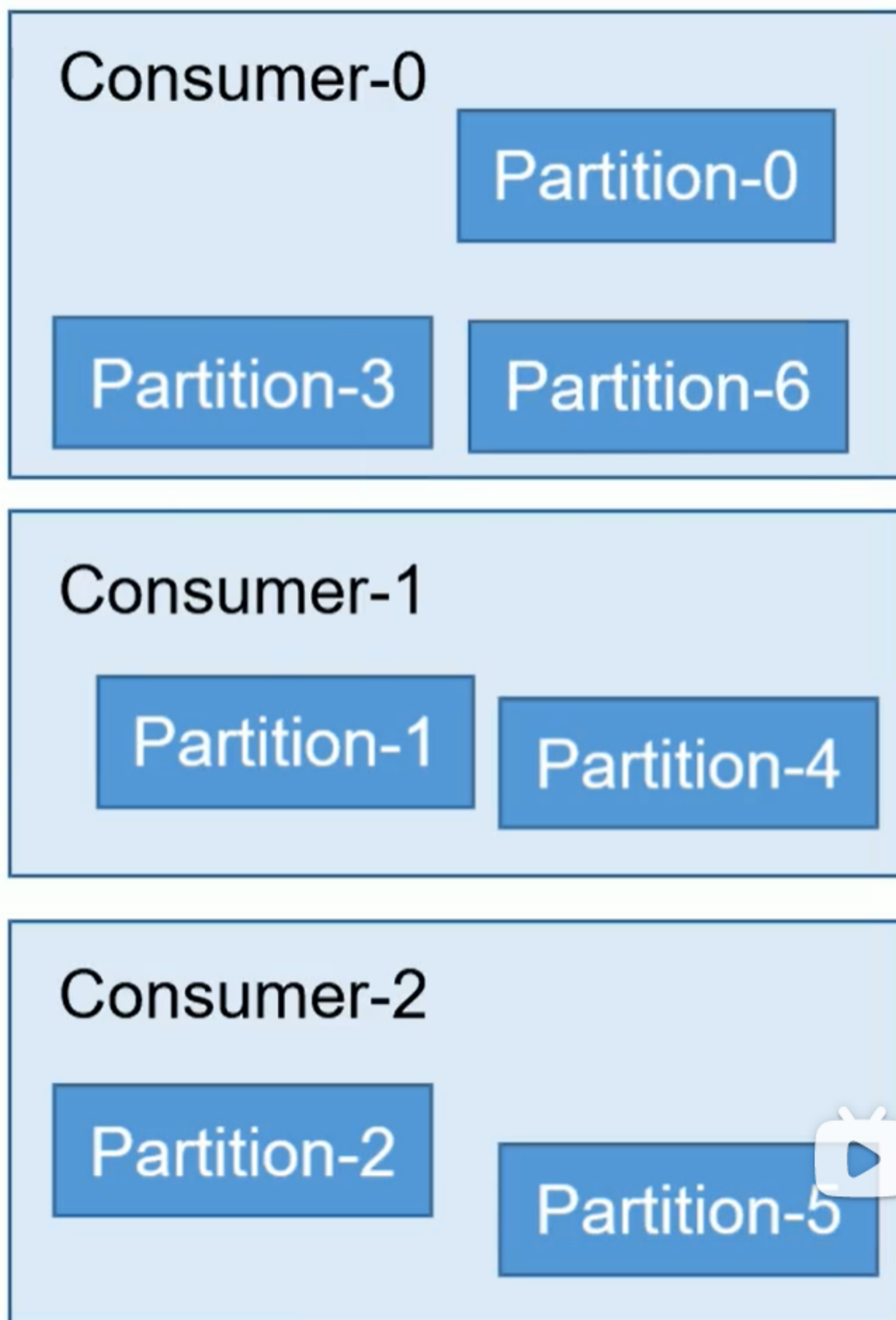
一个 **consumer group** 中有多个 **consumer**, 一个 **topic** 有多个 **partition**, 所以必然会涉及到 **partition** 的分配问题, 即确定哪个 **partition** 由哪个 **consumer** 来消费。

Kafka 有两种分配策略: 一个是 **RoundRobin**, 一个是 **Range**

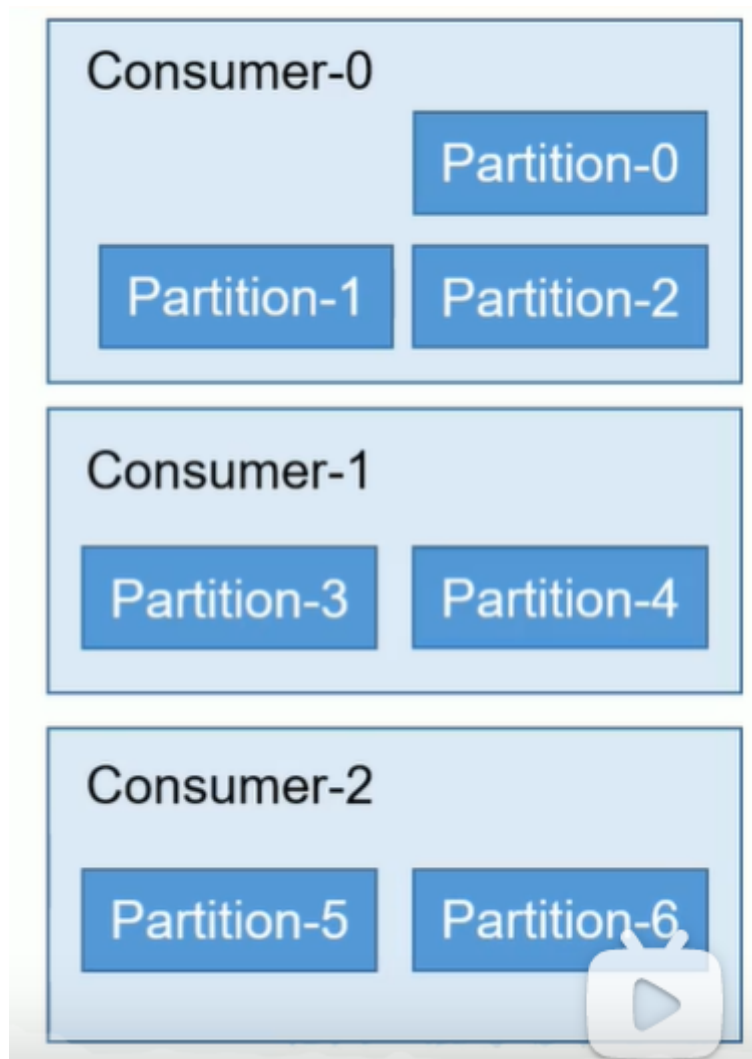
**RoundRobin** (面向组)

轮询, 一个一个分配。





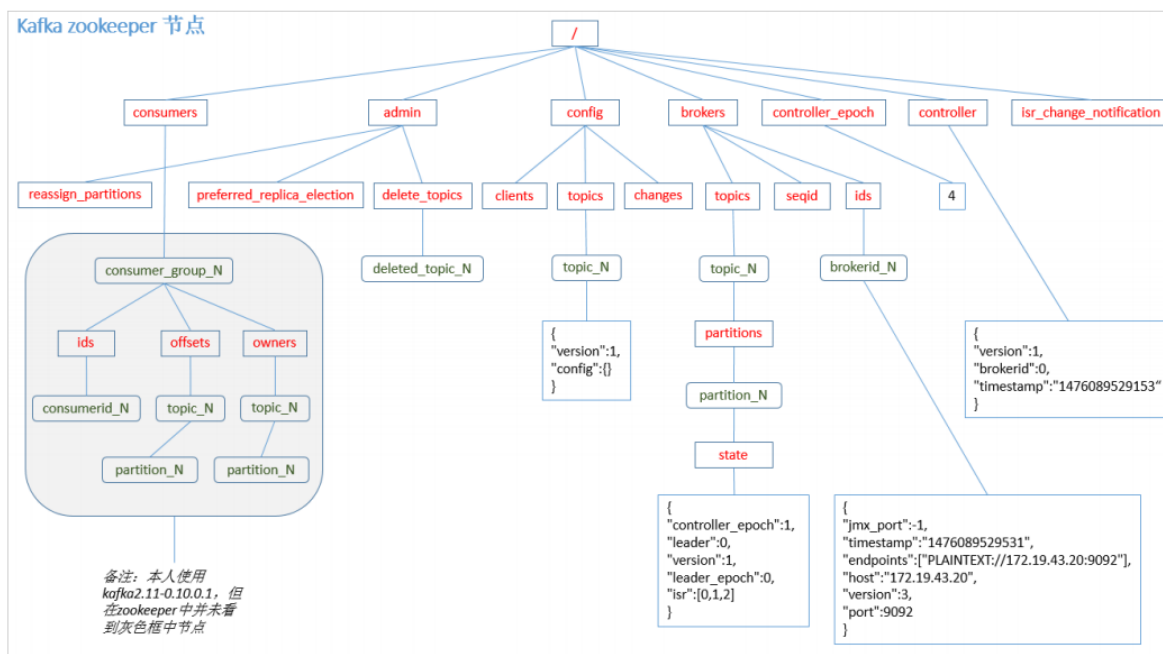
**Range** (面向主题)



## offset的维护

按照 G T P 进行存储

由于 consumer 在消费过程中可能会出现断电宕机等故障，consumer 恢复后，需要从故障前的位置的继续消费，所以 consumer 需要实时记录自己消费到了哪个 offset，以便故障恢复后继续消费。



注意：producer 不在 zk 中注册，消费者在 zk 中注册。

Kafka 0.9 版本之前，consumer 默认将 offset 保存在 Zookeeper 中，从 0.9 版本开始，consumer 默认将 offset 保存在 Kafka 一个内置的 topic 中，该 topic 为 \_\_consumer\_offsets。

1) 修改配置文件 cconsumer.properties

```
exclude.internal.topics=false
```

2) 读取 offset

0.11.0.0 之后版本 (含)

```
bin/kafka-console-consumer.sh --topic __consumer_offsets --zookeeper
hadoop102:2181 --formatter
"kafka.coordinator.group.GroupMetadataManager$OffsetsMessageFormatter" --
consumer.config config/consumer.properties --from-beginning
```

## 消费者组案例

- 需求：测试同一个消费者组中的消费者，同一时刻只能有一个消费者消费。
- 案例实操
  - (1) 在 hadoop102、hadoop103 上修改 /opt/module/kafka/config/consumer.properties 配置文件中的 group.id 属性为任意组名。

```
[atguigu@hadoop103 config] vi consumer.properties group.id=atguigu
```

(2) 在 hadoop102、hadoop103 上分别启动消费者

```
[atguigu@hadoop102 kafka]• bin/kafka-console-consumer.sh \--zookeeper  
hadoop102:2181 --topic first --consumer.config config/consumer.properties
```

```
[atguigu@hadoop103 kafka]• bin/kafka-console-consumer.sh --zookeeper  
hadoop102:2181 --topic first --consumer.config config/consumer.properties
```

(3) 在 hadoop104 上启动生产者

```
[atguigu@hadoop104 kafka]$ bin/kafka-console-producer.sh \--broker-list  
hadoop102:9092 --topic first  
>hello world
```

(4) 查看 hadoop102 和 hadoop103 的接收者。同一时刻只有一个消费者接收到消息