

## Kafka的消费者和消费者组有什么区别？为什么需要消费者组？

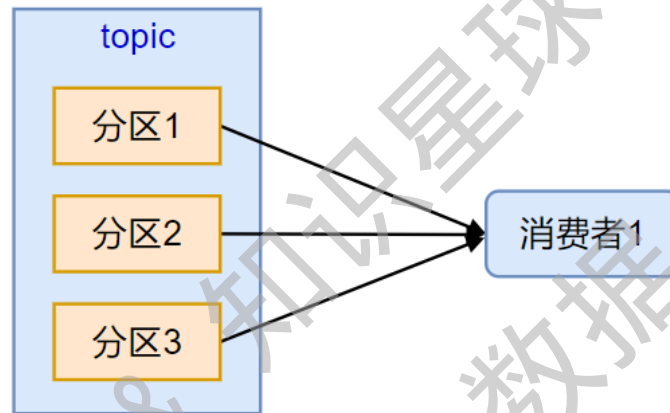
可回答：1) 说下Kafka的消费者和消费者组，以及它们的作用是什么？

问过的一些公司：字节提前批(2020.08)，京东(2020.08)，Bigo(2020.08)，拼多多学霸批(2019.08)

参考答案：

### 1、什么是消费者

顾名思义，消费者就是从kafka集群消费数据的客户端，如下图，展示了一个消费者从一个topic中消费数据的模型。



### 2、为什么需要消费者组

如果这个时候 kafka 上游生产的数据很快，超过了这个 消费者1 的消费速度，那么就会导致数据堆积，产生一些大家都知道的蛋疼事情了，那么我们只能加强 消费者 的消费能力，所以也就有了 消费者组。

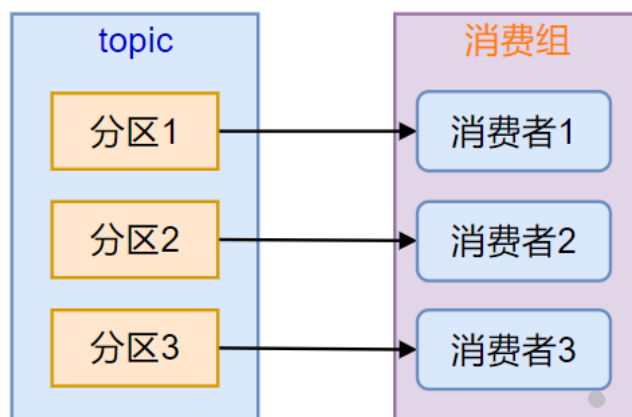
### 3、什么是消费者组

所谓 消费者组，其实就是一组 消费者 的集合，当我们看到下面这张图是不是就特别舒服了，我们采用了一个 消费组 来消费这个 topic，众人拾柴火焰高，其消费能力那是按倍数递增的，所以这里我们一般来说都是采用 消费者组 来消费数据，而不会是 单消费者 来消费数据的。

注意：

一个topic可以被多个消费者组消费，但是每个消费者组消费的数据是互不干扰的，也就是说，每个消费组消费的都是完整的数据。

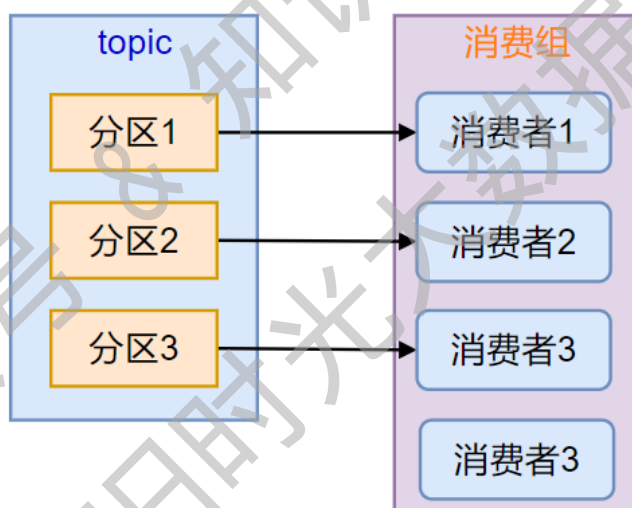
一个分区只能被同一个消费组内的一个消费者消费，而不能拆给多个消费者消费，也就是说如果你某个消费者组内的消费者数比该 Topic 的分区数还多，那么多余的消费者是不起作用的。



扩展一下：

1) 是不是一个消费组的消费者越多其消费能力就越强呢？

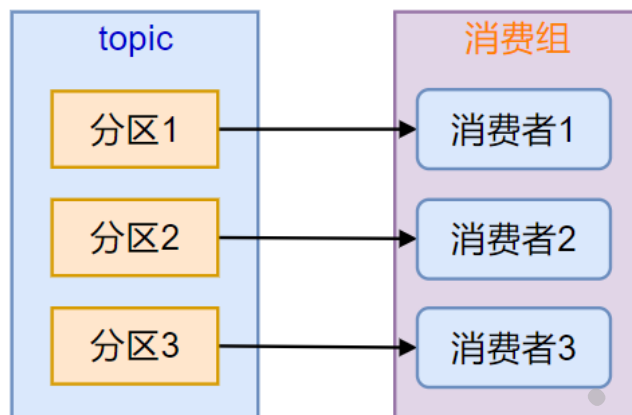
从下图我们就可以很好的可以回答这个问题了，我们可以看到消费者4是完全没有消费任何的数据的，所以如果你想要加强消费者组的能力，除了添加消费者，分区的数量也是需要跟着增加的，只有这样他们的并行度才能上的去，消费能力才会强。



2) 为了提高消费组的消费能力，我是不是可以随便添加分区和消费者呢？

答案当然是否定的。。。我们看到下图，一般来说我们建议消费者数量和分区数量是一致的，当我们的消费能力不够时，就必须通过调整分区的数量来提高并行度，但是，我们应该尽量来避免这种情况发生。

比如：现在我们需要在下图的基础上增加一个分区4，那么这个分区4该由谁来消费呢？这个时候kafka会进行分区再均衡，来为这个分区分配消费者，分区再均衡期间该Topic是不可用的，并且作为一个被消费者，分区数的改动将影响到每一个消费者组，所以在创建 topic 的时候，我们就应该考虑好分区数，来尽量避免这种情况发生。



欢迎加入知识星球，获取《大数据面试题 V4.0》以及更多大数据开发内容



知识星球

长按扫码领取优惠

