

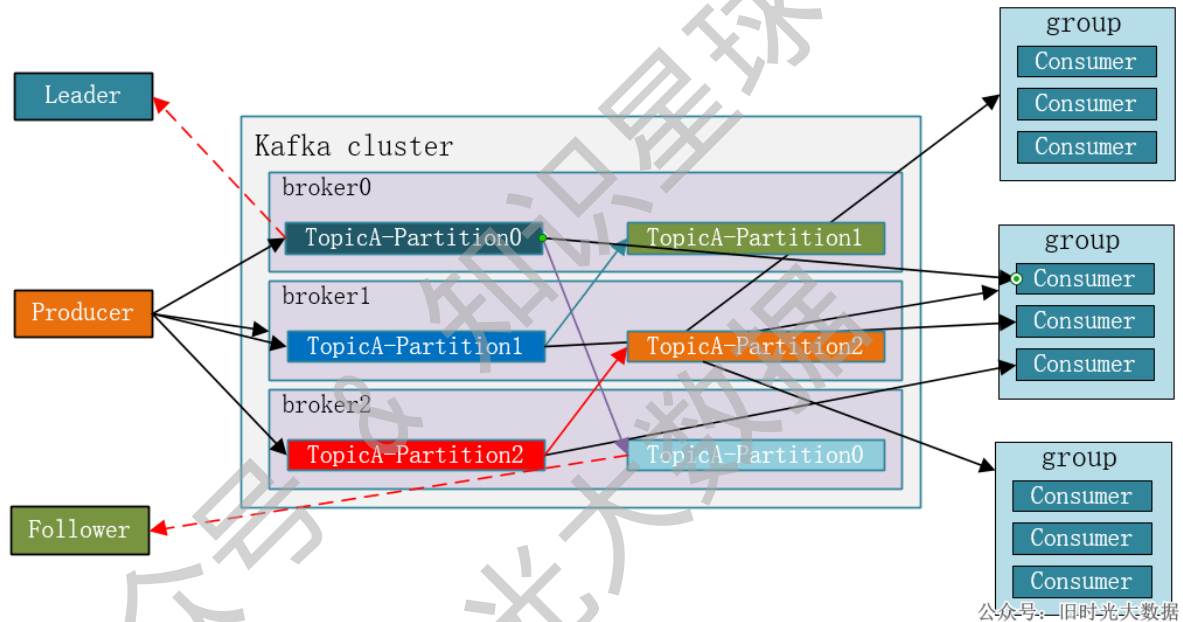
说下Kafka架构

可回答：1) Kafka的相关结构；2) Kafka的组成部分

参考答案：

Kafka基础架构

- 为方便扩展，并提高吞吐量，一个topic分为多个partition
- 配合分区的设计，提出消费者组的概念，组内每个消费者并行消费
- 为提高可用性，为每个partition增加若干副本，类似NameNode HA



- 1) **Producer**：消息生产者，负责向Broker发送消息
- 2) **Consumer**：消息消费者，从负责Broker读取并消费消息
- 3) **Consumer Group (CG)**：消费者组，由多个consumer组成。消费者组内每个消费者负责消费不同分区的数据，一个分区只能由一个消费者消费；消费者组之间互不影响。所有的消费者都属于某个消费者组，即消费者组是逻辑上的一个订阅者。
- 4) **Broker**：一台kafka服务器就是一个broker。一个集群由多个broker组成。一个broker可以容纳多个topic。
- 5) **Topic**：主题，与RocketMQ的Topic类似，使用Topic对消息进行分类，Kafka接收到的每条消息都会放入到一个Topic中。
 - Topic代表发布和消费记录的端点。生产者向主题发布消息，消费者订阅主题进行消费消息；
 - 每条记录有一个键，一个值，一个时间戳和一些元数据组成；
 - 在未指定分区的情况下发布消息时，将使用键的散列选择分区。
- 6) **Partition**：为了实现扩展性，一个非常大的topic可以分布到多个broker（即服务器）上，一个topic可以分为多个partition，对Topic的数据进行分布式存储的最小单位。；
- 7) **Replica**：副本，为保证集群中的某个节点发生故障时，该节点上的partition数据不丢失，且kafka仍然能够继续工作，kafka提供了副本机制，一个topic的每个分区都有若干个副本，一个leader和若干个Follower。

8) Leader: 每个分区多个副本的“主”，生产者发送数据的对象，以及消费者消费数据的对象都是leader。

9) Follower: 每个分区多个副本中的“从”，实时从leader中同步数据，保持和leader数据的同步。leader发生故障时，某个Follower会成为新的leader。

