

Kafka 面试题

1、是什么确保了 Kafka 中服务器的负载平衡？

由于领导者的主要角色是执行分区的所有读写请求的任务，而追随者被动地复制领导者。因此，在领导者失败时，其中一个追随者接管了领导者的角色。基本上，整个过程可确保服务器的负载平衡。

2、消费者 API 的作用是什么？

允许应用程序订阅一个或多个主题并处理生成给它们的记录流的 API，我们称之为消费者 API。

3、解释流 API 的作用？

一种允许应用程序充当流处理器的 API，它还使用一个或多个主题的输入流，并生成一个输出流到一个或多个输出主题，此外，有效地将输入流转换为输出流，我们称之为流 API。

4、Kafka 为什么那么快？

- Cache Filesystem Cache PageCache 缓存。
- 顺序写：由于现代的操作系统提供了预读和写技术，磁盘的顺序写大多数情况下比随机写内存还要快。

- Zero-copy 零拷技术减少拷贝次数。
- Batching of Messages 批量处理。合并小的请求，然后以流的方式进行交互，直顶网络上限。
- Pull 拉模式 使用拉模式进行消息的获取消费，与消费端处理能力相符。

5、Kafka 系统工具有哪些类型？

- Kafka 迁移工具：它有助于将代理从一个版本迁移到另一个版本。
- Mirror Maker : Mirror Maker 工具有助于将一个 Kafka 集群的镜像提供给另一个。
- 消费者检查:对于指定的主题集和消费者组，它显示主题，分区，所有者。

6、partition 的数据如何保存到硬盘

topic 中的多个 partition 以文件夹的形式保存到 broker，每个分区序号从 0 递增，且消息有序。Partition 文件下有多个 segment (xxx.index , xxx.log) , segment 文件里的大小和配置文件大小一致可以根据要求修改默认为 1g。如果大小大于 1g 时，会滚动一个新的 segment 并且以上一个 segment 最后一条消息的偏移量命名。

7、Zookeeper 对于 Kafka 的作用是什么？

- 1、Zookeeper 是一个开放源码的、高性能的协调服务，它用于 Kafka 的分布式应用。
- 2、Zookeeper 主要用于在集群中不同节点之间进行通信。

3、在 Kafka 中，它被用于提交偏移量，因此如果节点在任何情况下都失败了，它都可以从之前提交的偏移量中获取。

4、除此之外，它还执行其他活动，如：leader 检测、分布式同步、配置管理、识别新节点何时离开或连接、集群、节点实时状态等等。

8、流 API 的作用是什么？

一种允许应用程序充当流处理器的 API，它还使用一个或多个主题的输入流，并生成一个输出流到一个或多个输出主题，此外，有效地将输入流转换为输出流，我们称之为流 API。

9、Kafka 的流处理是什么意思？

连续、实时、并发和以逐记录方式处理数据的类型，我们称之为 Kafka 流处理。

10、Kafka 集群中保留期的目的是什么？

保留期限保留了 Kafka 群集中的所有已记录。它不会检查它们是否已被消耗。此外，可以通过使用保留期的配置设置来丢弃记录，而且，它可以释放一些空间。