链滴

kafka 的详解

作者：wgl530

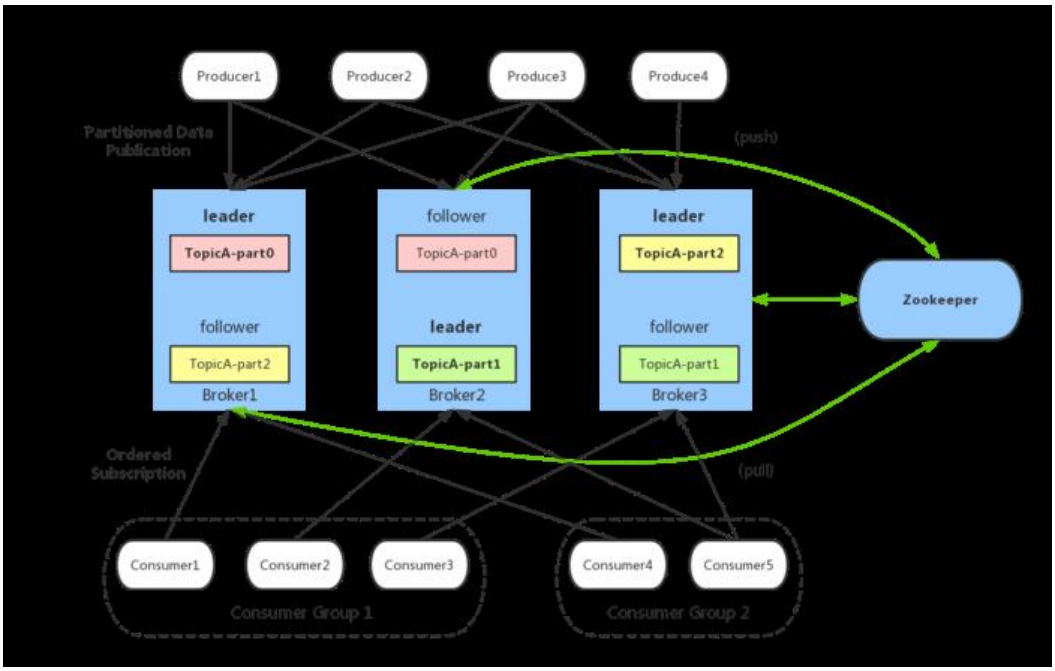
原文链接：https://ld246.com/article/1577444179357

来源网站：链滴

Kafka 概念

Kafka 是一种高吞吐量、分布式、基于发布/订阅的消息系统，最初由 LinkedIn 公司开发，使用 Scala 语言编写，目前是 Apache 的开源项目。

1. broker：Kafka 服务器，负责消息存储和转发
2. topic：消息类别，Kafka 按照 topic 来分类消息
3. partition：topic 的分区，一个 topic 可以包含多个 partition，topic 消息保存在各个 partition 上
4. offset：消息在日志中的位置，可以理解是消息在 partition 上的偏移量，也是代表该消息的唯一序号
5. Producer：消息生产者
6. Consumer：消息消费者
7. Consumer Group：消费者分组，每个 Consumer 必须属于一个 group
8. Zookeeper：保存着集群 broker、topic、partition 等 meta 数据；另外，还负责 broker 故障发现，partition leader 选举，负载均衡等功能



Kafka 数据存储设计

partition 的数据文件（ offset，MessageSize，data ）

partition中的每条Message包含了以下三个属性：offset，MessageSize，data，其中offset表示 Message 在这个 partition 中的偏移量，offset 不是该 Message 在 partition 数据文件中的实际储位置，而是逻辑上一个值，它唯一确定了partition中的一条Message，可以认为offset是 partition 中 Message 的 id；MessageSize 表示消息内容 data 的大小；data 为 Message 的具体内容。

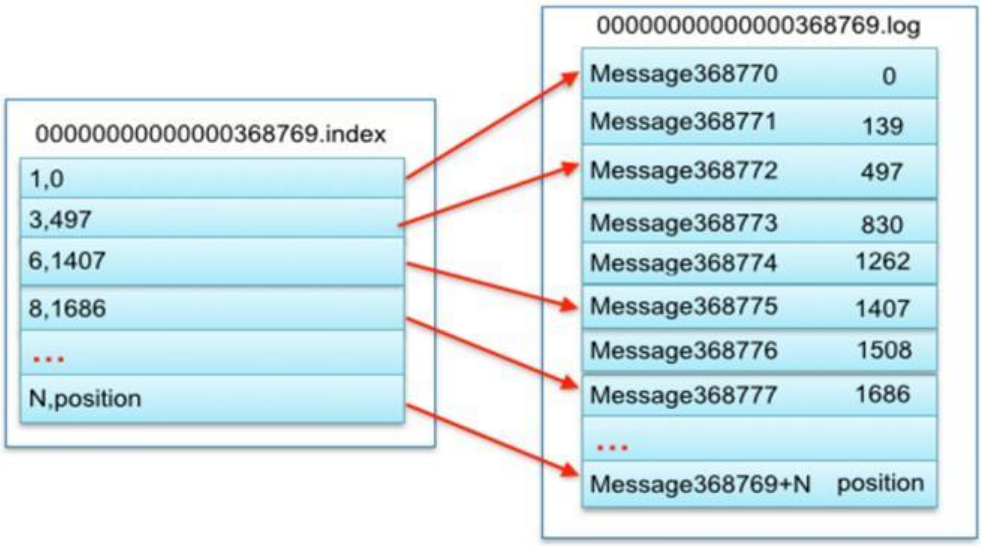
原文链接：kafka 的详解

数据文件分段 segment（顺序读写、分段命令、二分查找）

partition 物理上由多个 segment 文件组成，每个 segment 大小相等，顺序读写。每个 segment 数据文件以该段中最小的 offset 命名，文件扩展名为.log。这样在查找指定 offset 的 Message 的时候，用二分查找就可以定位到该 Message 在哪个 segment 数据文件中。

数据文件索引（分段索引、 稀疏存储）

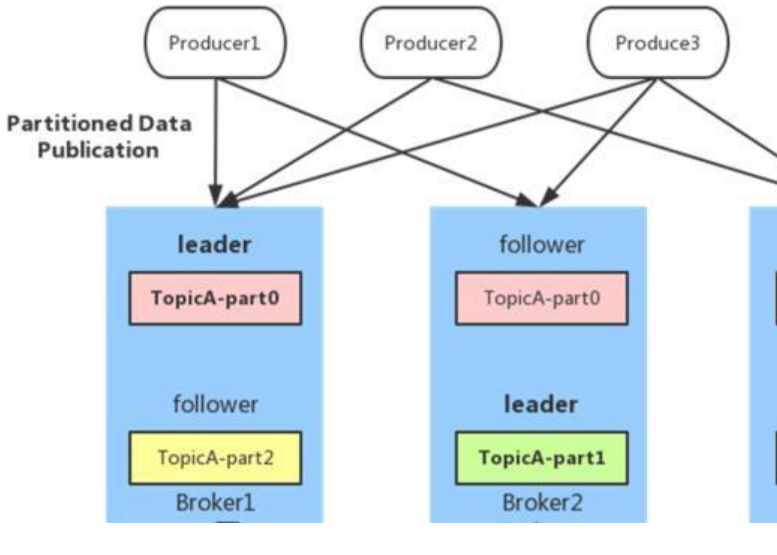
Kafka 为每个分段后的数据文件建立了索引文件，文件名与数据文件的名字是一样的，只是文件扩展名为.index。index 文件中并没有为数据文件中的每条 Message 建立索引，而是采用了稀疏存储的方式，每隔一定字节的数据建立一条索引。这样避免了索引文件占用过多的空间，从而可以将索引文件保留在内存中。



生产者设计

负载均衡（partition 会均衡分布到不同 broker 上）

由于消息 topic 由多个 partition 组成，且 partition 会均衡分布到不同 broker 上，因此，为了有效利用 broker 集群的性能，提高消息的吞吐量，producer 可以通过随机或者 hash 等方式，将消息平均发送到多个 partition 上，以实现负载均衡。



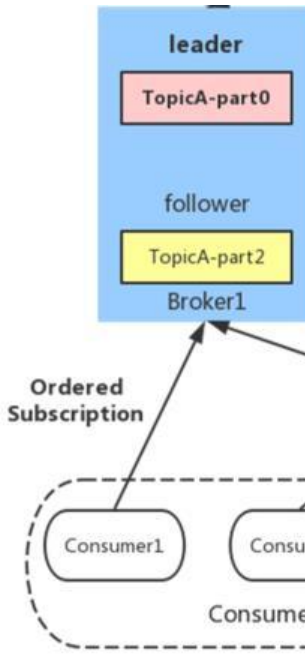
原文链接：kafka 的详解

批量发送

是提高消息吞吐量重要的方式，Producer 端可以在内存中合并多条消息后，以一次请求的方式发送了批量的消息给 broker，从而大大减少 broker 存储消息的 IO 操作次数。但也一定程度上影响了消息的实时性，相当于以时延代价，换取更好的吞吐量。

压缩（ GZIP 或 Snappy ）

Producer 端可以通过 GZIP 或 Snappy 格式对消息集合进行压缩。Producer 端进行压缩之后，在 Consumer 端需进行解压。压缩的好处就是减少传输的数据量，减轻对网络传输的压力，在对大数据处理上，瓶颈往往体现在网络上而不是 CPU（压缩和解压会耗掉部分 CPU 资源）。



Consumer Group

同一 Consumer Group 中的多个 Consumer 实例，不同时消费同一个 partition，等效于队列模式。partition 内消息是有序的，Consumer 通过 pull 方式消费消息。Kafka 不删除已消费的消息对于 partition，顺序读写磁盘数据，以时间复杂度 O(1)方式提供消息持久化能力。