

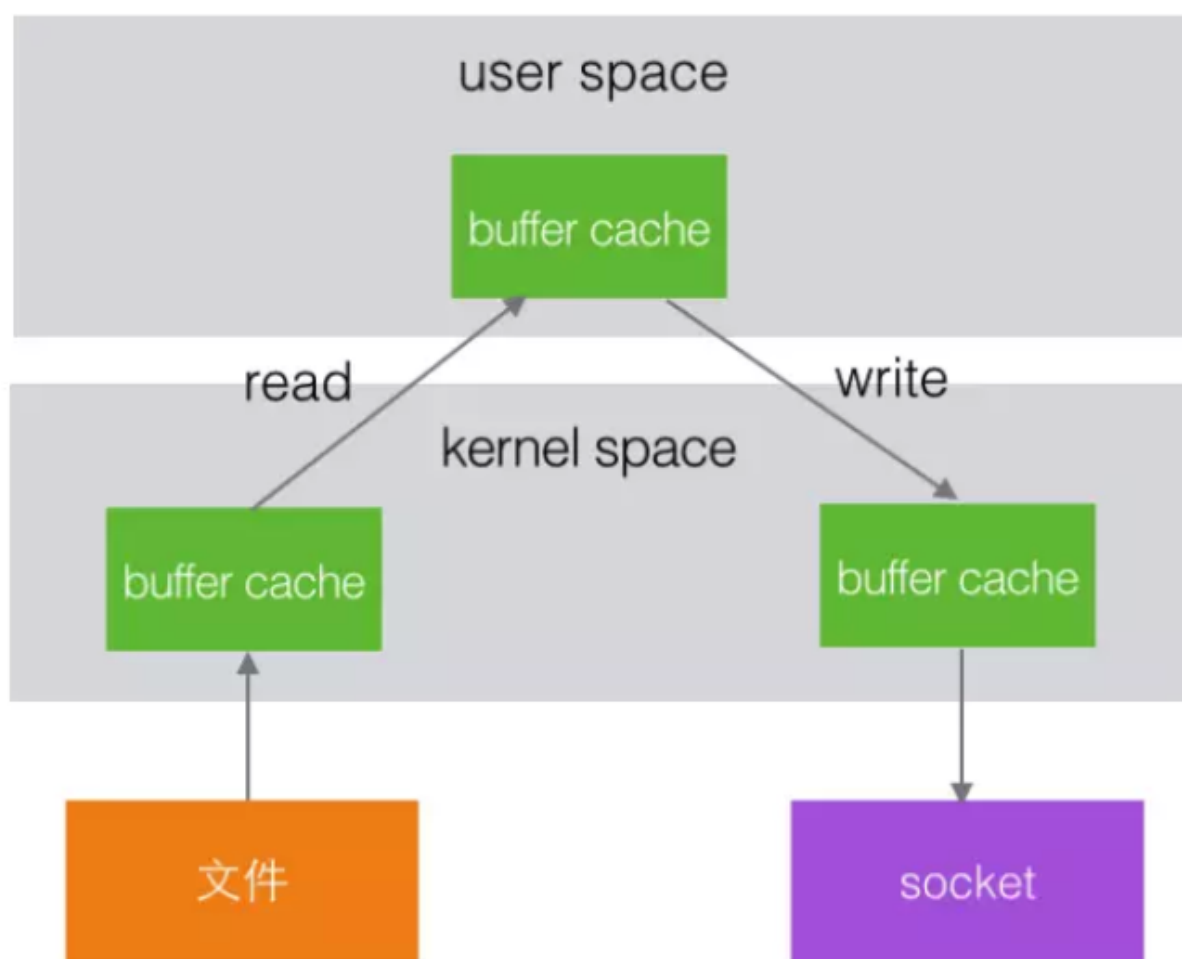
kafka吞吐量高的原因

1. 顺序读写

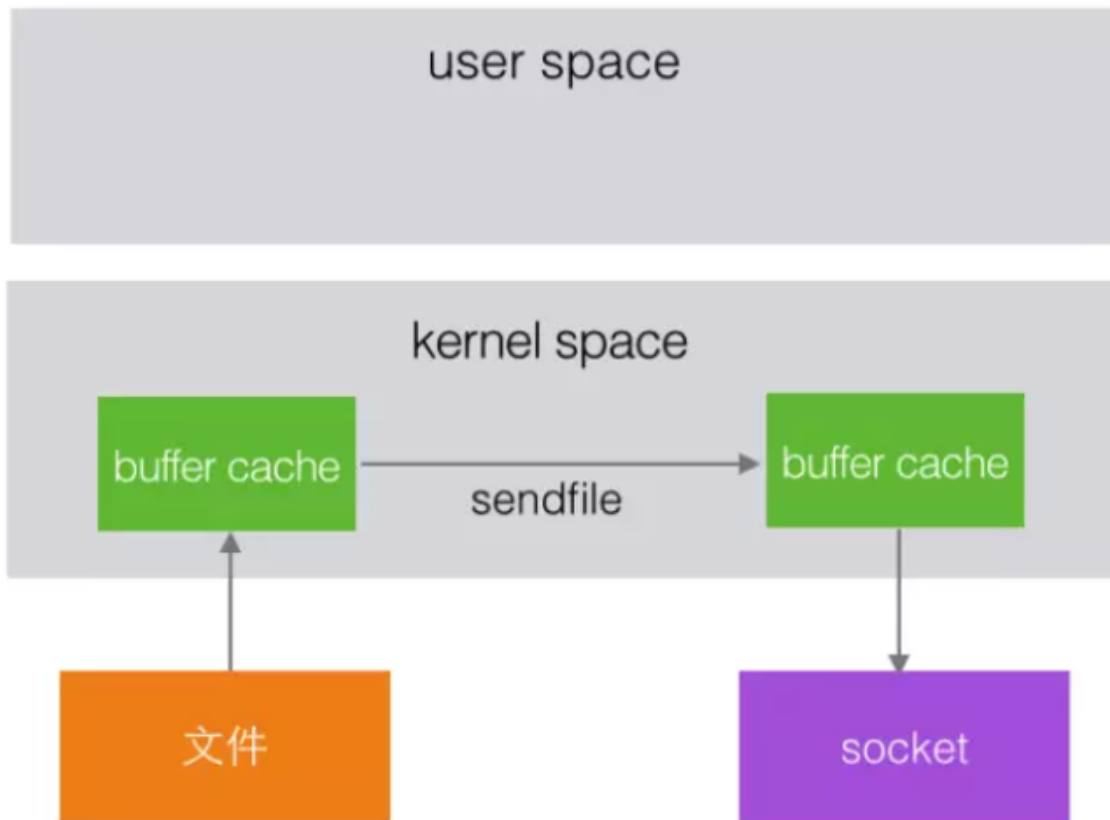
kafka的消息是不断追加到文件中的，这个特性使kafka可以充分利用磁盘的顺序读写性能。顺序读写不需要硬盘磁头的寻道时间，只需很少的扇区旋转时间，所以速度远快于随机读写。

2. 零拷贝

在linux kernel 2.2 之后出现了一种叫做“零拷贝”系统调用机制，就是跳过“用户缓冲区”的拷贝，建立一个磁盘空间和内存的直接映射，数据不再复制到“用户态缓冲区”

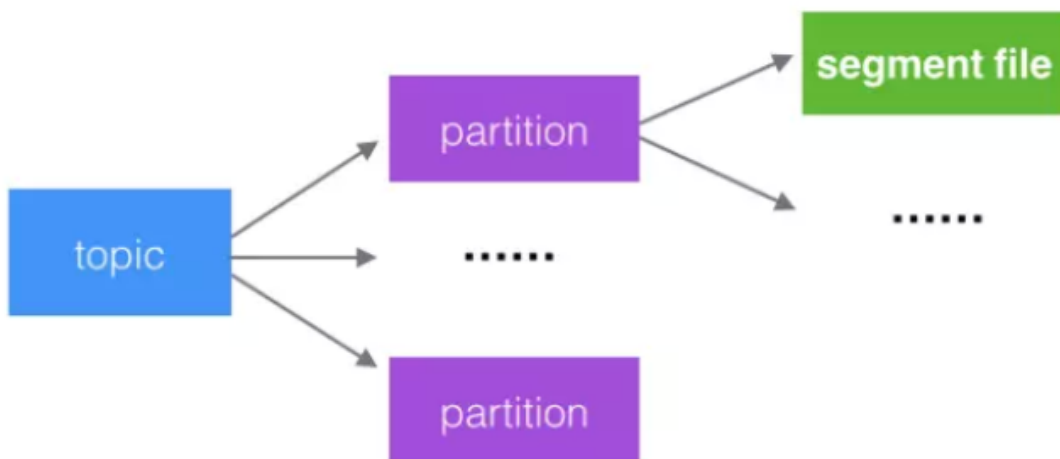


173206xip766ntnpziapk2.png



3. 分区

kafka中的topic中的内容可以被分为多个partition存在，每个partition又分为多个段segment，所以每次操作都是针对一小部分做操作，很轻便，并且增加并行操作的能力。



4. 批量发送

kafka允许进行批量发送消息，producer发送消息的时候，可以将消息缓存在本地，等到了固定条件发送到kafka

1. 等消息条数到了固定条数
2. 一段时间发送一次

5. 数据压缩

kafka还支持对消息集合进行压缩，producer可以通过gzip或snappy格式对消息集合进行压缩

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。

批量发送和数据压缩一起使用，单条做数据压缩的话，效果不明显。