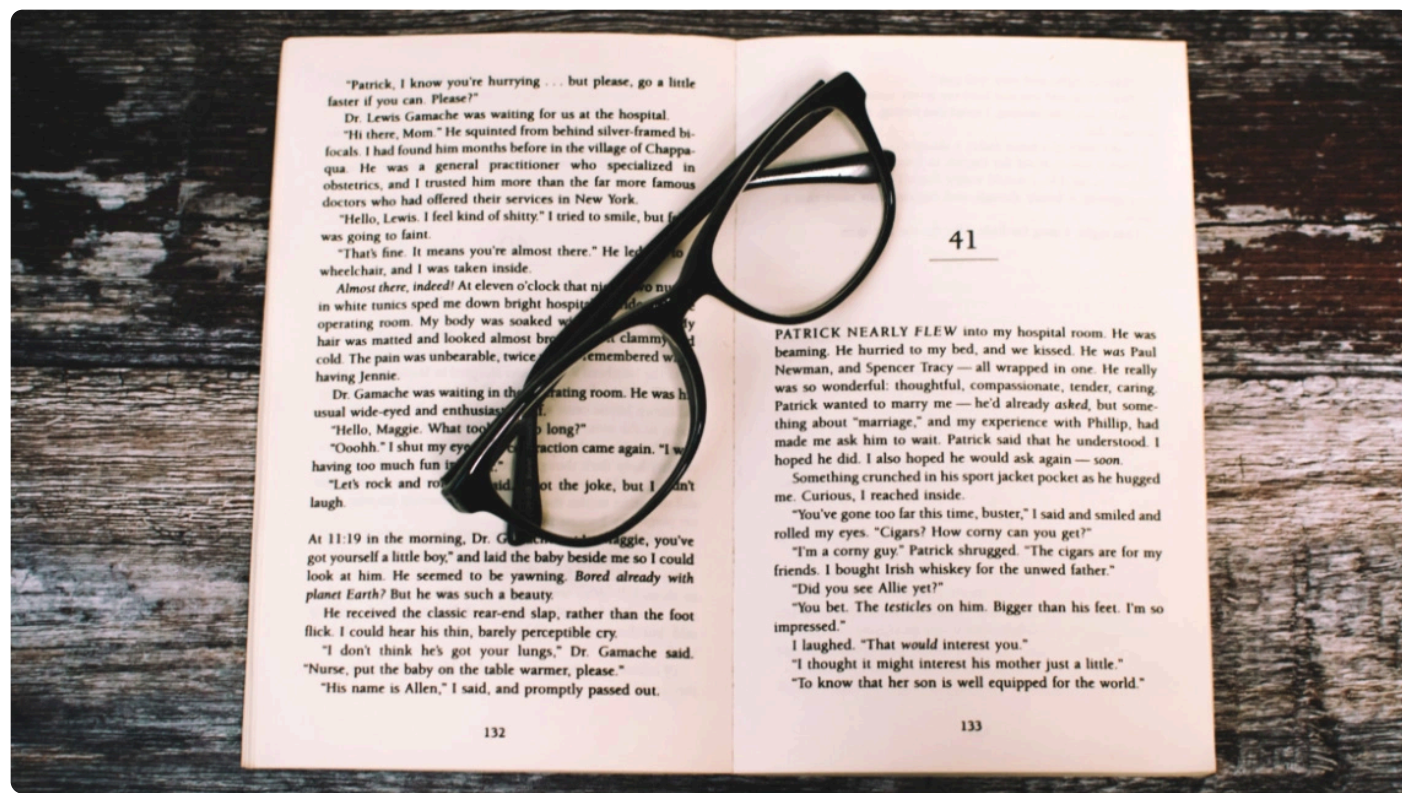


02 | 一篇文章带你快速搞定Kafka术语

胡夕 · Kafka核心技术与实战



你好，我是胡夕。今天我们正式开启 Apache Kafka 学习之旅。

在 Kafka 的世界中有很多概念和术语是需要你提前理解并熟练掌握的，这对于后面你深入学习 Kafka 各种功能和特性将大有裨益。下面我来盘点一下 Kafka 的各种术语。

在专栏的第一期我说过 Kafka 属于分布式的消息引擎系统，它的主要功能是提供一套完备的消息发布与订阅解决方案。在 Kafka 中，发布订阅的对象是主题（Topic），你可以为每个业务、每个应用甚至是每类数据都创建专属的主题。

向主题发布消息的客户端应用程序称为生产者（Producer），生产者程序通常持续不断地向一个或多个主题发送消息，而订阅这些主题消息的客户端应用程序就被称为消费者（Consumer）。和生产者类似，消费者也能够同时订阅多个主题的消息。我们把生产者和消费者统称为客户端（Clients）。你可以同时运行多个生产者和消费者实例，这些实例会不断地向 Kafka 集群中的多个主题生产和消费消息。

有客户端自然也就有服务器端。Kafka 的服务器端由被称为 Broker 的服务进程构成，即一个 Kafka 集群由多个 Broker 组成，Broker 负责接收和处理客户端发送过来的请求，以及对消息进行持久化。虽然多个 Broker 进程能够运行在同一台机器上，但更常见的做法是将不同的 Broker 分散运行在不同的机器上，这样如果集群中某一台机器宕机，即使在它上面运行的所有 Broker 进程都挂掉了，其他机器上的 Broker 也依然能够对外提供服务。这其实就是 Kafka 提供高可用的手段之一。

实现高可用的另一个手段就是备份机制（Replication）。备份的思想很简单，就是把相同的数据拷贝到多台机器上，而这些相同的数据拷贝在 Kafka 中被称为副本（Replica）。好吧，其实在整个分布式系统里好像都叫这个名字。副本的数量是可以配置的，这些副本保存着相同的数据，但却有不同的角色和作用。Kafka 定义了两类副本：领导者副本（Leader Replica）和追随者副本（Follower Replica）。前者对外提供服务，这里的对外指的是与客户端程序进行交互；而后者只是被动地追随领导者副本而已，不能与外界进行交互。当然了，你可能知道在很多其他系统中追随者副本是可以对外提供服务的，比如 MySQL 的从库是可以处理读操作的，但是在 Kafka 中追随者副本不会对外提供服务。对了，一个有意思的事情是现在已经不提倡使用 Master-Slave 来指代这种主从关系了，毕竟 Slave 有奴隶的意思，在美国这种严禁种族歧视的国度，这种表述有点政治不正确了，所以目前大部分的系统都改成 Leader-Follower 了。

副本的工作机制也很简单：生产者总是向领导者副本写消息；而消费者总是从领导者副本读消息。至于追随者副本，它只做一件事：向领导者副本发送请求，请求领导者把最新生产的消息发给它，这样它能保持与领导者的同步。

虽然有了副本机制可以保证数据的持久化或消息不丢失，但没有解决伸缩性的问题。伸缩性即所谓的 Scalability，是分布式系统中非常重要且必须要谨慎对待的问题。什么是伸缩性呢？我们拿副本来说，虽然现在有了领导者副本和追随者副本，但倘若领导者副本积累了太多的数据以至于单台 Broker 机器都无法容纳了，此时应该怎么办呢？一个很自然的想法就是，能否把数据分割成多份保存在不同的 Broker 上？如果你就是这么想的，那么恭喜你，Kafka 就是这么设计的。

这种机制就是所谓的分区（Partitioning）。如果你了解其他分布式系统，你可能听说过分片、分区域等提法，比如 MongoDB 和 Elasticsearch 中的 Sharding、HBase 中的

Region，其实它们都是相同的原理，只是 Partitioning 是最标准的名称。

Kafka 中的分区机制指的是将每个主题划分成多个分区（Partition），每个分区是一组有序的消息日志。生产者生产的每条消息只会被发送到一个分区中，也就是说如果向一个双分区的主题发送一条消息，这条消息要么在分区 0 中，要么在分区 1 中。如你所见，Kafka 的分区编号是从 0 开始的，如果 Topic 有 100 个分区，那么它们的分区号就是从 0 到 99。

讲到这里，你可能有这样的疑问：刚才提到的副本如何与这里的分区联系在一起呢？实际上，副本是在分区这个层级定义的。每个分区下可以配置若干个副本，其中只能有 1 个领导者副本和 $N-1$ 个追随者副本。生产者向分区写入消息，每条消息在分区中的位置信息由一个叫位移（Offset）的数据来表征。分区位移总是从 0 开始，假设一个生产者向一个空分区写入了 10 条消息，那么这 10 条消息的位移依次是 0、1、2、.....、9。

至此我们能够完整地串联起 Kafka 的三层消息架构：

第一层是主题层，每个主题可以配置 M 个分区，而每个分区又可以配置 N 个副本。

第二层是分区层，每个分区的 N 个副本中只能有一个充当领导者角色，对外提供服务；其他 $N-1$ 个副本是追随者副本，只是提供数据冗余之用。

第三层是消息层，分区中包含若干条消息，每条消息的位移从 0 开始，依次递增。

最后，客户端程序只能与分区的领导者副本进行交互。

讲完了消息层次，我们来说说 Kafka Broker 是如何持久化数据的。总的来说，Kafka 使用消息日志（Log）来保存数据，一个日志就是磁盘上一个只能追加写（Append-only）消息的物理文件。因为只能追加写入，故避免了缓慢的随机 I/O 操作，改为性能较好的顺序 I/O 写操作，这也是实现 Kafka 高吞吐量特性的一个重要手段。不过如果你不停地向一个日志写入消息，最终也会耗尽所有的磁盘空间，因此 Kafka 必然要定期地删除消息以回收磁盘。怎么删除呢？简单来说就是通过日志段（Log Segment）机制。在 Kafka 底层，一个日志又进一步细分成多个日志段，消息被追加写到当前最新的日志段中，当写满了一个日志段后，Kafka 会自动切分出一个新的日志段，并将老的日志段封存起来。Kafka 在后台还有定时任务会定期地检查老的日志段是否能够被删除，从而实现回收磁盘空间的目的。

这里再重点说说消费者。在专栏的第一期中我提到过两种消息模型，即点对点模型（Peer to Peer, P2P）和发布订阅模型。这里面的点对点指的是同一条消息只能被下游的一个消费者消费，其他消费者则不能染指。在 Kafka 中实现这种 P2P 模型的方法就是引入了消费者组（Consumer Group）。所谓的消费者组，指的是多个消费者实例共同组成一个组来消费一组主题。这组主题中的每个分区都只会被组内的一个消费者实例消费，其他消费者实例不能消费它。为什么要引入消费者组呢？主要是为了提升消费者端的吞吐量。多个消费者实例同时消费，加速整个消费端的吞吐量（TPS）。我会在专栏的后面详细介绍消费者组机制，所以现在你只需要了解消费者组是做什么的即可。另外这里的消费者实例可以是运行消费者应用的进程，也可以是一个线程，它们都称为一个消费者实例（Consumer Instance）。

消费者组里面的所有消费者实例不仅“瓜分”订阅主题的数据，而且更酷的是它们还能彼此协助。假设组内某个实例挂掉了，Kafka 能够自动检测到，然后把这个 Failed 实例之前负责的分区转移给其他活着的消费者。这个过程就是 Kafka 中大名鼎鼎的“重平衡”（Rebalance）。嗯，其实既是大名鼎鼎，也是臭名昭著，因为由重平衡引发的消费者问题比比皆是。事实上，目前很多重平衡的 Bug 社区都无力解决。

每个消费者在消费消息的过程中必然需要有个字段记录它当前消费到了分区的哪个位置上，这个字段就是消费者位移（Consumer Offset）。注意，这和上面所说的位移完全不是一个概念。上面的“位移”表征的是分区内的消息位置，它是不变的，即一旦消息被成功写入到一个分区上，它的位移值就是固定的了。而消费者位移则不同，它可能是随时变化的，毕竟它是消费者消费进度的指示器嘛。另外每个消费者有着自己的消费者位移，因此一定要区分这两类位移的区别。我个人把消息在分区中的位移称为分区位移，而把消费者端的位移称为消费者位移。

小结

我来总结一下今天提到的所有名词术语：

消息：Record。Kafka 是消息引擎嘛，这里的消息就是指 Kafka 处理的主要对象。

主题：Topic。主题是承载消息的逻辑容器，在实际使用中多用来区分具体的业务。

分区：Partition。一个有序不变的消息序列。每个主题下可以有多个分区。

消息位移：Offset。表示分区中每条消息的位置信息，是一个单调递增且不变的值。

副本：Replica。Kafka 中同一条消息能够被拷贝到多个地方以提供数据冗余，这些地方就是所谓的副本。副本还分为领导者副本和追随者副本，各自有不同的角色划分。副本是在分区层级下的，即每个分区可配置多个副本实现高可用。

生产者：Producer。向主题发布新消息的应用程序。

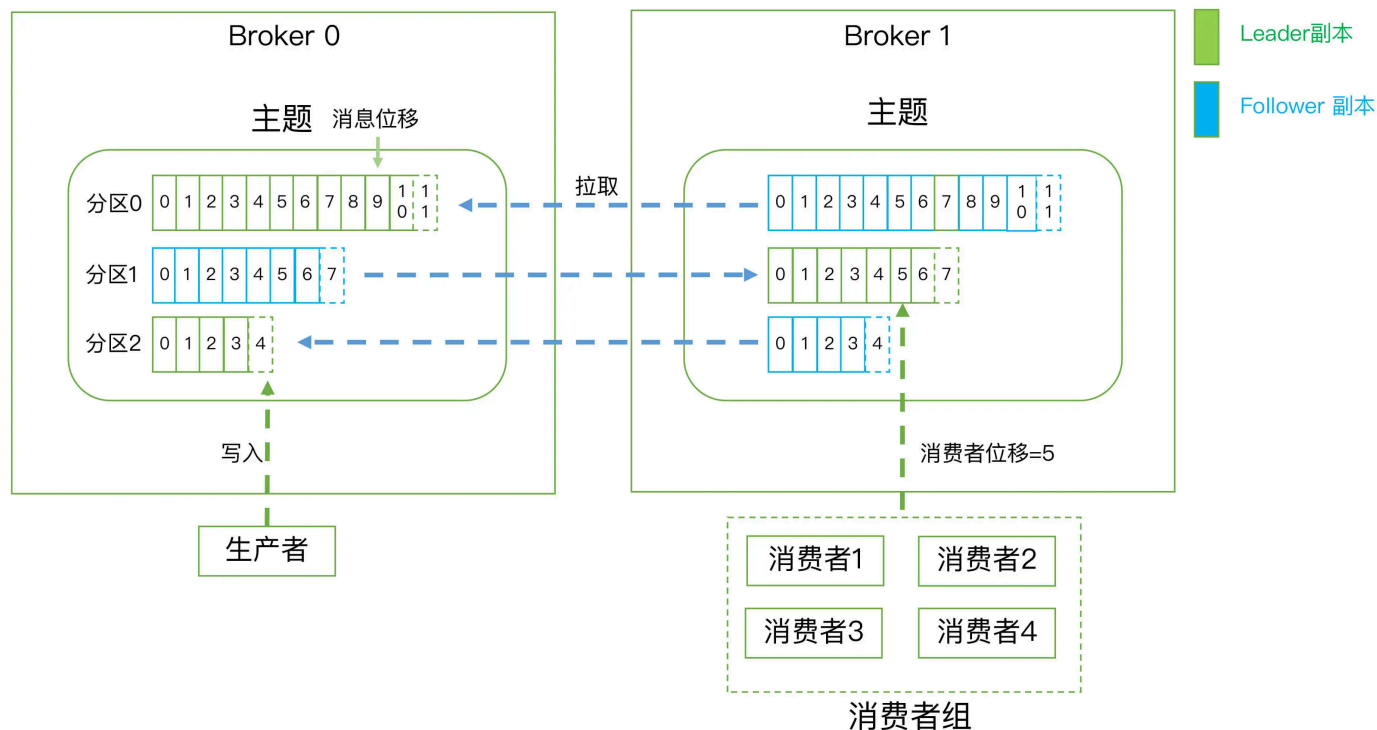
消费者：Consumer。从主题订阅新消息的应用程序。

消费者位移：Consumer Offset。表征消费者消费进度，每个消费者都有自己的消费者位移。

消费者组：Consumer Group。多个消费者实例共同组成的一个组，同时消费多个分区以实现高吞吐。

重平衡：Rebalance。消费者组内某个消费者实例挂掉后，其他消费者实例自动重新分配订阅主题分区的过程。Rebalance 是 Kafka 消费者端实现高可用的重要手段。

最后我用一张图来展示上面提到的这些概念，希望这张图能够帮助你形象化地理解所有这些概念：



开放讨论

请思考一下为什么 Kafka 不像 MySQL 那样允许追随者副本对外提供读服务？

欢迎写下你的思考和答案，我们一起讨论。如果你觉得有所收获，也欢迎把文章分享给你的朋友。

AI智能总结

Apache Kafka是一个分布式的消息引擎系统，为消息发布与订阅提供了完备的解决方案。本文详细介绍了Kafka的核心概念和术语，帮助读者快速了解其特点。主题、生产者、消费者、Broker、副本、分区、消费者组等术语被详细解释，展现了Kafka的功能和架构。Kafka的服务器端由多个Broker组成，负责接收和处理客户端请求，以及对消息进行持久化。副本机制和分区机制保证了数据的持久化和伸缩性。消费者组提升了消费者端的吞吐量，而重平衡机制能够自动调整消费者组内的分区分配。此外，文章还介绍了Kafka Broker持久化数据的机制和消费者位移的概念。总之，本文为读者提供了全面的Kafka术语解释，帮助读者快速了解Kafka的核心概念和特点。

Kafka术语解释包括消息、主题、分区、消息位移、副本、生产者、消费者、消费者位移、消费者组、重平衡等概念。这些术语构成了Kafka的核心架构和功能特点，帮助用户理解Kafka的工作原理和优势。文章还提出了一个开放讨论问题，引发读者思考Kafka与MySQL在追随者副本对外提供读服务方面的差异。通过这些内容，读者可以快速了解Kafka的关键概念和特点，为进一步深入学习和应用Kafka奠定了基础。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

全部留言 (219)

最新 精选



胡夕 置顶
2019-06-11

结尾处增加了一张图，提炼了02中讲到的Kafka概念和术语，希望能够帮助到你：)

共 6 条评论 >

👍 155



时光剪影
2019-06-16

整理一遍个人的理解：

Kafka体系架构=M个producer +N个broker +K个consumer+ZK集群

producer:生产者

Broker：服务代理节点，Kafka服务实例。

n个组成一个Kafka集群，通常一台机器部署一个Kafka实例，一个实例挂了其他实例仍可以使用，体现了高可用

consumer：消费者

消费topic 的消息，一个topic 可以让若干个consumer消费，若干个consumer组成一个 consumer group，一条消息只能被consumer group 中一个consumer消费，若干个partition 被若干个consumer 同时消费，达到消费者高吞吐量

topic：主题

partition：一个topic 可以拥有若干个partition（从 0 开始标识partition），分布在不同的broker 上，实现发布与订阅时负载均衡。producer 通过自定义的规则将消息发送到对应topic 下某个partition，以offset标识一条消息在一个partition的唯一性。

一个partition拥有多个replica，提高容灾能力。

replica 包含两种类型：leader 副本、follower副本，

leader副本负责读写请求，follower 副本负责同步leader副本消息，通过副本选举实现故障转移。

partition在机器磁盘上以log 体现，采用顺序追加日志的方式添加新消息、实现高吞吐量

作者回复: 厉害！感觉比我写的简洁：)

共 13 条评论 >

👍 173



邈邈的流浪剑客

2019-06-06

如果允许follower副本对外提供读服务（主写从读），首先会存在数据一致性的问题，消息从主节点同步到从节点需要时间，可能造成主从节点的数据不一致。主写从读无非就是为了减轻leader节点的压力，将读请求的负载均衡到follower节点，如果Kafka的分区相对均匀地分散到各个broker上，同样可以达到负载均衡的效果，没必要刻意实现主写从读增加代码实现的复杂程度

作者回复: 是的。前些天在知乎上就这个问题也解答了一下，有兴趣可以看看：<https://www.zhihu.com/question/327925275/answer/705690755>

共 13 条评论 >

👍 96



卡特

2020-04-17

加入a主题有4个分区，消费者组有2个实例，发布应用的时候，会先新启动一个服务节点，加入消费组，通过重平衡分配到至少1个最多2个分区，消费者的偏移量是

1，重新从0开始

2，拿到分配分区的上一个消费者偏移量？

如果按照文章说的，即偏移量为0，消息应该会重复消费；

如果拿到上一个消费者的偏移量则不会消息重复消费，具体过程又是怎样的？

求解惑，

作者回复: 整个故事是这样的：假设C1消费P0,P1, C2消费P2,P3。如果C1从未提交，C1挂掉，C2开始消费P0,P1，发现没有对应提交位移，那么按照C2的auto.offset.reset值决定从哪里消费，如果是earliest，从P0，P1的最小位移值（可能不是0）开始消费，如果是latest，从P0，P1的最新位移值（分区高水位值）开始消费。但如果C1之前提交了位移，那么C1挂掉之后C2从C1最新一次提交的位移值开始消费。

所谓的重复消费是指，C1消费了一部分数据，还没来得及提交这部分数据的位移就挂了。C2承接过来之后会重新消费这部分数据。

我说清楚了吗：)

共 3 条评论 >

👍 58



永光

2019-06-10

为什么 Kafka 不像 MySQL 那样允许追随者副本对外提供读服务？

答：因为mysql一般部署在不同的机器上一台机器读写会遇到瓶颈，Kafka中的领导者副本一般均匀分布在不同的broker中，已经起到了负载的作用。即：同一个topic的已经通过分区的形式负载到不同的broker上了，读写的时候针对的领导者副本，但是量相比mysql一个还实例少太多，个人觉得没有必要在提供度读服务了。（如果量大还可以使用更多的副本，让每一个副本本身都不太大）不知道这样理解对不对？

作者回复: 我个人觉得是很不错的答案，自己也学到了一些：)

共 6 条评论 >

👍 44



jacke

2019-06-22

胡老师：

还想问个分区的问题，比如一个topic分为0, 1, 2 个分区
写入0到9条消息，按照轮训分布：

0分区：0, 1, 2, 9

1分区：3, 4, 5,

2分区：6, 7, 8

那对于消费端来说，不管是p2p点对点模式，还是push/sub模式来说，
如何保证消费端的读取顺序也是从0到9？因为0到9条消息是分布在3个
分区上的，同时消费者是主动轮训模式去读分区数据的，
有没有可能读到后面写的的数据呢？比如先读到5在读到4？
ps:刚开始学习，问题比较多，谅解

作者回复：目前Kafka的设计中多个分区的话无法保证全局的消息顺序。如果一定要实现全局的消息顺序，只能单分区

共 7 条评论 >

👍 36



莫道不销魂

2019-06-11

老师，我想问下

- 1、kafka是按照什么规则将消息划分到各个分区的？
- 2、既然同一个topic下的消息分布在不同的分区，那是什么机制将topic、partition、record 关联或者说管理起来的？

作者回复：1. 如果producer指定了要发送的目标分区，消息自然是去到那个分区；否则就按照producer参数partitioner.class指定的分区策略来定；如果你没有指定过partitioner.class，那么默认的规则是：看消息是否有key，如果有则计算key的murmur2哈希值%topic分区数；如果没有key，按照轮询的方式确定分区。

2. 这个层级其实是逻辑概念。在物理上还是以日志段（log segment）文件的方式保存，日志段文件在内存中有对应的Java对象，里面关联了你说的这些。

共 2 条评论 >

👍 27



Happy

2019-06-20

主写从读无非就是为了减轻leader节点的压力，而kafka中数据分布相对比较均匀，所说的Follower从节点,实际上也是其他topic partition的Leader节点，所以在Follower可以读数据，那么会影响Follower节点上的做为Leader的partition的读性能，所以整体性能并没有提升，但是带来了主从数据同步延迟导致的数据不一致的问题

作者回复: 同意:)

共 3 条评论 >

👍 22



(`田ω田`)

2019-06-06

- 1、主题中的每个分区都只会被组内的一个消费者实例消费，其他消费者实例不能消费它。
- 2、假设组内某个实例挂掉了，Kafka 能够自动检测到，然后把这个 Failed 实例之前负责的分区分转移给其他活着的消费者。

意思是1个分区只能同时被1个消费者消费，但是1个消费者能同时消费多个分区是吗？那1个消费者里面就会有多个消费者位移变量？

如果1个主题有2个分区，消费者组有3个消费者，那至少有1个消费者闲置？

作者回复: 在一个消费者组下，一个分区只能被一个消费者消费，但一个消费者可能被分配多个分区，因而在提交位移时也能提交多个分区的位移。

针对你说的第二种情况，答案是：是的。有一个消费者将无法分配到任何分区，处于idle状态。

共 3 条评论 >

👍 22



快跑

2019-07-02

老师，你好

假如只有一个Producer进程，Kafka只有一分区。Producer按照1，2，3，4，5的顺序发送消息，Kafka这个唯一分区收到消息一定是1，2，3，4，5么？ Producer端，网络，数据格式等因素，会不会导致Kafka只有一个分区接收到数据顺序跟Producer发送数据顺序不一致

作者回复: 如果retries>0并且max.in.flight.requests.per.connection>1有可能出现消息乱序的情况

共 2 条评论 >

👍 15

