

10 | 生产者压缩算法面面观

胡夕 · Kafka核心技术与实战



你好，我是胡夕。今天我要和你分享的内容是：生产者压缩算法面面观。

说起压缩（compression），我相信你一定不会感到陌生。它秉承了用时间去换空间的经典 trade-off 思想，具体来说就是用 CPU 时间去换磁盘空间或网络 I/O 传输量，希望以较小的 CPU 开销带来更少的磁盘占用或更少的网络 I/O 传输。在 Kafka 中，压缩也是用来做这件事的。今天我就来跟你分享一下 Kafka 中压缩的那些事儿。

怎么压缩？

Kafka 是如何压缩消息的呢？要弄清楚这个问题，就要从 Kafka 的消息格式说起了。目前 Kafka 共有两大类消息格式，社区分别称之为 V1 版本和 V2 版本。V2 版本是 Kafka 0.11.0.0 中正式引入的。

不论是哪个版本，Kafka 的消息层次都分为两层：消息集合（message set）以及消息（message）。一个消息集合中包含若干条日志项（record item），而日志项才是真正封装

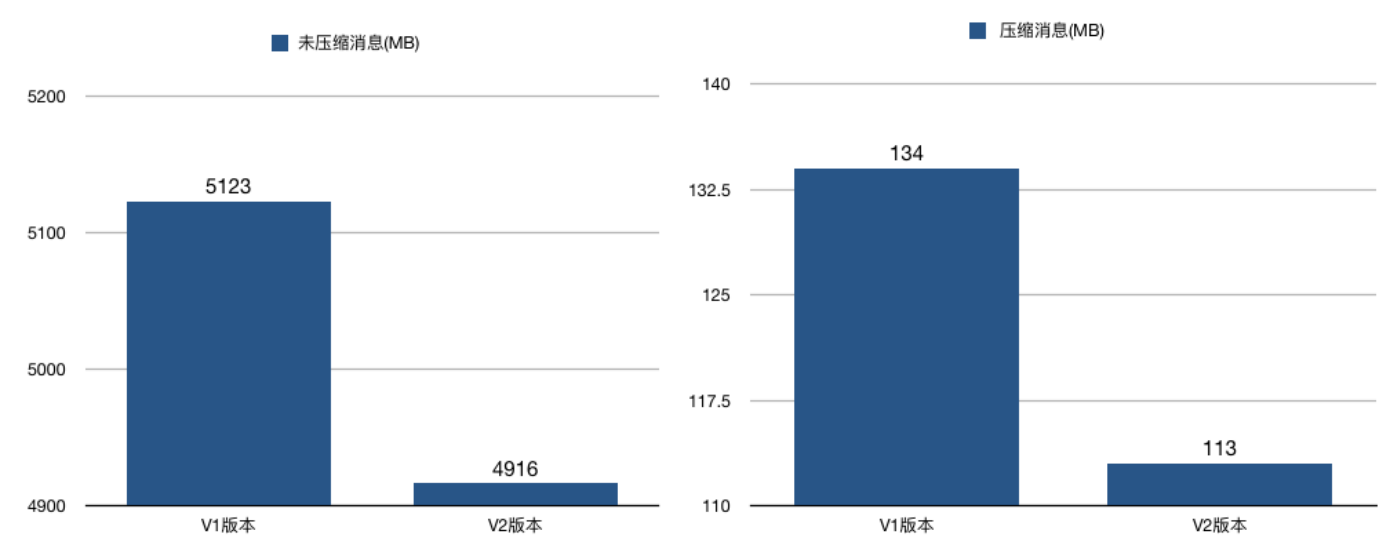
消息的地方。Kafka 底层的消息日志由一系列消息集合日志项组成。Kafka 通常不会直接操作具体的一条条消息，它总是在消息集合这个层面上进行写入操作。

那么社区引入 V2 版本的目的是什么呢？V2 版本主要是针对 V1 版本的一些弊端做了修正，和我们今天讨论的主题相关的修正有哪些呢？先介绍一个，就是把消息的公共部分抽取出来放到外层消息集合里面，这样就不用每条消息都保存这些信息了。

我来举个例子。原来在 V1 版本中，每条消息都需要执行 CRC 校验，但有些情况下消息的 CRC 值是会发生变化的。比如在 Broker 端可能会对消息时间戳字段进行更新，那么重新计算之后的 CRC 值也会相应更新；再比如 Broker 端在执行消息格式转换时（主要是为了兼容老版本客户端程序），也会带来 CRC 值的变化。鉴于这些情况，再对每条消息都执行 CRC 校验就有点没必要了，不仅浪费空间还耽误 CPU 时间，因此在 V2 版本中，消息的 CRC 校验工作就被移到了消息集合这一层。

V2 版本还有一个和压缩息息相关的改进，就是保存压缩消息的方法发生了变化。之前 V1 版本中保存压缩消息的方法是把多条消息进行压缩然后保存到外层消息的消息体字段中；而 V2 版本的做法是对整个消息集合进行压缩。显然后者应该比前者有更好的压缩效果。


我对两个版本分别做了一个简单的测试，结果显示，在相同条件下，不论是否启用压缩，V2 版本都比 V1 版本节省磁盘空间。当启用压缩时，这种节省空间的效果更加明显，就像下面这两张图展示的那样：



何时压缩？

在 Kafka 中，压缩可能发生在两个地方：生产者端和 Broker 端。

生产者程序中配置 `compression.type` 参数即表示启用指定类型的压缩算法。比如下面这段程序代码展示了如何构建一个开启 GZIP 的 Producer 对象：

 复制代码

```
1 Properties props = new Properties();
2 props.put("bootstrap.servers", "localhost:9092");
3 props.put("acks", "all");
4 props.put("key.serializer", "org.apache.kafka.common.serialization.StringSeriali
5 props.put("value.serializer", "org.apache.kafka.common.serialization.StringSeria
6 // 开启GZIP压缩
7 props.put("compression.type", "gzip");
8
9 Producer<String, String> producer = new KafkaProducer<>(props);
```

这里比较关键的代码行是 `props.put("compression.type", "gzip")`，它表明该 Producer 的压缩算法使用的是 GZIP。这样 Producer 启动后生产的每个消息集合都是经 GZIP 压缩过的，故而能很好地节省网络传输带宽以及 Kafka Broker 端的磁盘占用。

在生产者端启用压缩是很自然的想法，那为什么我说在 Broker 端也可能进行压缩呢？其实大部分情况下 Broker 从 Producer 端接收到消息后仅仅是原封不动地保存而不会对其进行任何修改，但这里的“大部分情况”也是要满足一定条件的。有两种例外情况就可能让 Broker 重新压缩消息。

情况一：Broker 端指定了和 Producer 端不同的压缩算法。

先看一个例子。想象这样一个对话。

Producer 说：“我要使用 GZIP 进行压缩。”

Broker 说：“不好意思，我这边接收的消息必须使用 Snappy 算法进行压缩。”

你看，这种情况下 Broker 接收到 GZIP 压缩消息后，只能解压缩然后使用 Snappy 重新压缩一遍。如果你翻开 Kafka 官网，你会发现 Broker 端也有一个参数叫 `compression.type`，和上面那个例子中的同名。但是这个参数的默认值是 `producer`，这表示 Broker 端会“尊重”Producer 端使用的压缩算法。可一旦你在 Broker 端设置了不同的 `compression.type` 值，就一定要小心了，因为可能会发生预料之外的压缩 / 解压缩操作，通常表现为 Broker 端 CPU 使用率飙升。

情况二：Broker 端发生了消息格式转换。

所谓的消息格式转换主要是为了兼容老版本的消费者程序。还记得之前说过的 V1、V2 版本吧？在一个生产环境中，Kafka 集群中同时保存多种版本的消息格式非常常见。为了兼容老版本的格式，Broker 端会对新版本消息执行向老版本格式的转换。这个过程中会涉及消息的解压缩和重新压缩。一般情况下这种消息格式转换对性能是有很影响的，除了这里的压缩之外，它还让 Kafka 丧失了引以为豪的 Zero Copy 特性。

所谓“Zero Copy”就是“零拷贝”，我在专栏 [第 6 期](#) 提到过，说的是当数据在磁盘和网络进行传输时避免昂贵的内核态数据拷贝，从而实现快速的数据传输。因此如果 Kafka 享受不到这个特性的话，性能必然有所损失，所以尽量保证消息格式的统一吧，这样不仅可以避免不必要的解压缩 / 重新压缩，对提升其他方面的性能也大有裨益。如果有兴趣你可以深入地了解下 Zero Copy 的原理。

何时解压缩？

有压缩必有解压缩！通常来说解压缩发生在消费者程序中，也就是说 Producer 发送压缩消息到 Broker 后，Broker 照单全收并原样保存起来。当 Consumer 程序请求这部分消息时，Broker 依然原样发送出去，当消息到达 Consumer 端后，由 Consumer 自行解压缩还原成之前的消息。

那么现在问题来了，Consumer 怎么知道这些消息是用何种压缩算法压缩的呢？其实答案就在消息中。Kafka 会将启用了哪种压缩算法封装进消息集合中，这样当 Consumer 读取到消息集合时，它自然就知道了这些消息使用的是哪种压缩算法。如果用一句话总结一下压缩和解压缩，那么我希望你记住这句话：**Producer 端压缩、Broker 端保持、Consumer 端解压缩。**

除了在 Consumer 端解压缩，Broker 端也会进行解压缩。注意了，这和前面提到消息格式转换时发生的解压缩是不同的场景。每个压缩过的消息集合在 Broker 端写入时都要发生解压缩操作，目的就是为了对消息执行各种验证。我们必须承认这种解压缩对 Broker 端性能是有一定影响的，特别是对 CPU 的使用率而言。

事实上，最近国内京东的小伙伴们刚刚向社区提出了一个 bugfix，建议去掉因为做消息校验而引入的解压缩。据他们称，去了解压缩之后，Broker 端的 CPU 使用率至少降低了 50%。不过有些遗憾的是，目前社区并未采纳这个建议，原因就是这种消息校验是非常重要的，不可盲目去之。毕竟先把事情做对是最重要的，在做对的基础上，再考虑把事情做好做快。针对这个使用场景，你也可以思考一下，是否有一个两全其美的方案，既能避免消息解压缩也能对消息执行校验。

各种压缩算法对比

那么我们来谈谈压缩算法。这可是重头戏！之前说了这么多，我们还是要比较一下各个压缩算法的优劣，这样我们才能有针对性地配置适合我们业务的压缩策略。

在 Kafka 2.1.0 版本之前，Kafka 支持 3 种压缩算法：GZIP、Snappy 和 LZ4。从 2.1.0 开始，Kafka 正式支持 Zstandard 算法（简称为 zstd）。它是 Facebook 开源的一个压缩算法，能够提供超高的压缩比（compression ratio）。

对了，看一个压缩算法的优劣，有两个重要的指标：一个指标是压缩比，原先占 100 份空间的东西经压缩之后变成了占 20 份空间，那么压缩比就是 5，显然压缩比越高越好；另一个指标就是压缩 / 解压缩吞吐量，比如每秒能压缩或解压缩多少 MB 的数据。同样地，吞吐量也是越高越好。

下面这张表是 Facebook Zstandard 官网提供的一份压缩算法 benchmark 比较结果：

| Compressor name | Ratio | Compression | Decompress. |
|----------------------|-------|-------------|-------------|
| zstd 1.3.4 -1 | 2.877 | 470 MB/s | 1380 MB/s |
| zlib 1.2.11 -1 | 2.743 | 110 MB/s | 400 MB/s |
| brrotli 1.0.2 -0 | 2.701 | 410 MB/s | 430 MB/s |
| quicklz 1.5.0 -1 | 2.238 | 550 MB/s | 710 MB/s |
| lzo1x 2.09 -1 | 2.108 | 650 MB/s | 830 MB/s |
| lz4 1.8.1 | 2.101 | 750 MB/s | 3700 MB/s |
| snappy 1.1.4 | 2.091 | 530 MB/s | 1800 MB/s |
| lzf 3.6 -1 | 2.077 | 400 MB/s | 860 MB/s |

从表中我们可以发现 zstd 算法有着最高的压缩比，而在吞吐量上的表现只能说中规中矩。反观 LZ4 算法，它在吞吐量方面则是毫无疑问的执牛耳者。当然对于表格中数据的权威性我不做过多解读，只想用它来说明一下当前各种压缩算法的大致表现。

在实际使用中，GZIP、Snappy、LZ4 甚至是 zstd 的表现各有千秋。但对于 Kafka 而言，它们的性能测试结果却出奇得一致，即在吞吐量方面：LZ4 > Snappy > zstd 和 GZIP；而在压缩比方面，zstd > LZ4 > GZIP > Snappy。具体到物理资源，使用 Snappy 算法占用的网络带宽最多，zstd 最少，这是合理的，毕竟 zstd 就是要提供超高的压缩比；在 CPU 使用率方面，各个算法表现得差不多，只是在压缩时 Snappy 算法使用的 CPU 较多一些，而在解压缩时 GZIP 算法则可能使用更多的 CPU。

最佳实践

了解了这些算法对比，我们就能根据自身的实际情况有针对性地启用合适的压缩算法。

首先来说压缩。何时启用压缩是比较合适的时机呢？

你已经知道 Producer 端完成的压缩，那么启用压缩的一个条件就是 Producer 程序运行机器上的 CPU 资源要很充足。如果 Producer 运行机器本身 CPU 已经消耗殆尽了，那么启用消息压缩无疑是雪上加霜，只会适得其反。

除了 CPU 资源充足这一条件，如果你的环境中带宽资源有限，那么我也建议你开启压缩。事实上我见过的很多 Kafka 生产环境都遭遇过带宽被打满的情况。这年头，带宽可是比 CPU 和内存还要珍贵的稀缺资源，毕竟万兆网络还不是普通公司的标配，因此千兆网络中 Kafka 集群带宽资源耗尽这件事情就特别容易出现。如果你的客户端机器 CPU 资源有很多富余，我强烈建议你开启 zstd 压缩，这样能极大地节省网络资源消耗。

其次说说解压缩。其实也没什么可说的。一旦启用压缩，解压缩是不可避免的事情。这里只想强调一点：我们对不可抗拒的解压缩无能为力，但至少能规避掉那些意料之外的解压缩。就像我前面说的，因为要兼容老版本而引入的解压缩操作就属于这类。有条件的话尽量保证不要出现消息格式转换的情况。

小结

总结一下今天分享的内容：我们主要讨论了 Kafka 压缩的各个方面，包括 Kafka 是如何对消息进行压缩的、何时进行压缩及解压缩，还对比了目前 Kafka 支持的几个压缩算法，最后我给出了工程化的最佳实践。分享这么多内容，我就只有一个目的：就是希望你能根据自身的实际情况恰当地选择合适的 Kafka 压缩算法，以求实现最大的资源利用率。

生产者压缩算法面面观

- 在Kafka中，压缩可能发生在两个地方：生产者端和Broker端。
- 让Broker重新压缩消息的2种例外情况：Broker端指定了和Producer端不同的压缩算法；Broker端发生了消息格式转换。
- 如果用一句话总结一下压缩和解压缩，那么我希望你记住这句话：Producer端压缩、Broker端保持、Consumer端解压缩。
- 对Kafka而言，这4种压缩算法的性能测试结果很一致：在吞吐量方面，LZ4 > Snappy > zstd和GZIP；在压缩比方面，zstd > LZ4 > GZIP > Snappy。



极客时间

开放讨论

最后给出一道作业题，请花时间思考一下：前面我们提到了 Broker 要对压缩消息集合执行解压缩操作，然后逐条对消息进行校验，有人提出了一个方案：把这种消息校验移到 Producer 端来做，Broker 直接读取校验结果即可，这样就可以避免在 Broker 端执行解压缩操作。你认同这种方案吗？

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

Kafka消息压缩是通过在生产者端和Broker端进行压缩和解压缩来实现的。V2版本相比V1版本在消息压缩方面有了一些改进，如将整个消息集合进行压缩，节省了磁盘空间。在生产者端，可以通过配置compression.type参数来启用指定类型的压缩算法，而在Broker端也可能进行重新压缩消息，例如在消息格式转换或使用不同压缩算法时。解压缩通常发生在消费者程序中，但在Broker端写入时也会进行解压缩操作以执行各种验证。文章强调了消息校验的重要性，尽管去掉解压缩可以降低Broker端的CPU使用率，但在做对的基础上考虑如何做好做快更为重要。Kafka的消息压缩机制涉及生产者端压缩、Broker端保持和消费者端解压缩，而不同压缩算法和消息格式转换可能会对压缩和解压缩过程产生影响。

Kafka支持多种压缩算法，包括GZIP、Snappy、LZ4和Zstandard（zstd）。这些算法在压缩比和吞吐量上各有优劣，因此在实际使用中需要根据实际情况选择合适的压缩算法。对于Kafka而言，在吞吐量方面，LZ4 > Snappy > zstd和GZIP；而在压缩比方面，zstd > LZ4 > GZIP > Snappy。根据环境的CPU和带宽资源情况，可以有针对性地选择合适的压缩算法，以实现最大的资源利用率。

最佳实践方面，建议在Producer端CPU资源充足且带宽资源有限时启用压缩，而在解压缩方面则需要规避意外情况，尽量避免消息格式转换。总的来说，文章分享了Kafka压缩的各个方面，包括压缩实现、压缩算法对比和最佳实践，旨在帮助读者根据实际情况选择合适的Kafka压缩算法，以实现最大的资源利用率。

© 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 (97)

最新 精选



胡夕 置顶
2019-06-26

刚刚看到4天前京东提的那个jira已经修复了，看来规避了broker端为执行校验而做的解压缩操作，代码也merge进了2.4版本。有兴趣的同学可以看一下：<https://issues.apache.org/jira/browse/KAFKA-8106>

共 5 条评论 >

👍 115



Geek_8441fd

2019-06-25

broker端校验可以分两步走。

第1步，message set 层面，增加一个 crc，这样可以不用解压缩，直接校验压缩后的数据。如果校验不成功，说明message set 中有损坏的message；这时，再做解压操作，挨个校验message，找出损坏的那一个。

这样的话，绝大部分情况下，是不用做解压操作的；只有在确实发生错误时，才需要解压。请指正。

作者回复: 嗯嗯，挺好的。我自己也学到了一些。另外校验不仅仅是CRC校验，还有消息级别的检查。

共 6 条评论 >

👍 57



Li Shunduo

2019-06-25

假如一个消息集合里有10条消息，并且被压缩，但是消费端配置每次只poll 5条消息。这种情况下，消费端怎么解压缩？矛盾点是 如果只取5条消息，需要broker帮助解压缩；如果取整个消息集合10条消息，会有贷款等资源的浪费？

作者回复: 目前java consumer的设计是一次取出一批，缓存在客户端内存中，然后再过滤出max.poll.records条消息返给你，也不算太浪费吧，毕竟下次可以直接从缓存中取，不用再发请求了。

共 5 条评论 >

👍 27



衣申人

2019-12-11

老师，我看源码，broker在接收producer消息并落盘这块貌似没有用零拷贝啊！只有传输给consumer时用了，求解答

作者回复: 是的，就是你理解的那样。Kafka使用Zero Copy优化将页缓存中的数据直接传输到Socket——这的确是发生在broker到consumer的链路上。这种优化能成行的前提是consumer端能够识别磁盘上的消息格式。

共 3 条评论 >

👍 24



风中花

2019-06-26

胡老师您好！ 我们已经学历10多节课了！ 针对我们得留言和反馈，不知道您有没有给我们一些后续得课程得学习建议和方法？ 我目前得学习就是您告诉我们得，我必须学会记住。但是看同学们得评论和反馈，我觉得貌似还有很多很多知识啊且不知也不懂，故有此一问！ 希望老师能给与一点一点学习建议？ 感谢老师

作者回复: 个人觉得学一个东西最重要的还是要用，如果只是参加一些培训课程很难全面的理解。您这么多的留言我一直坚持回复。我也一直是这个观点：用起来，自然问题就来了。

我学机器学习的经历和您现在学Kafka很像。没有实际使用场景怎么学都觉得深入不了。

我给您的建议是：把Kafka官网通读几遍然后再实现一个实时日志收集系统（比如把服务器日志实时放入Kafka）。事实上，能把官网全面理解的话已经比很多Kafka使用者要强了。

共 2 条评论 >

👍 21



星期八

2019-07-12

老师那再问一下，如果多条消息组成消息集合发送，那是什么条件控制消息发送，如果是一条又是什么条件控制触发发送的呢

作者回复: 主要是这两个参数：batch.size和linger.ms。如果是生产了一条消息且linger.ms=0，通常producer就会立即发送者一条消息了。

共 2 条评论 >

👍 16



南辕北辙

2019-06-27

老师有一点不是很明白，在正常情况下broker端会原样保存起来，但是为了检验需要解压缩。该怎么去理解这个过程呢，broker端解压缩以后还会压缩还原吗？
这个过程是在用户态执行的吗，总感觉怪怪的

作者回复: 它只是解压缩读取而已，不会将解压缩之后的数据回写到磁盘。另外就像我置顶的留言那样，目前社区已经接纳了京东小伙伴的修改，貌似可以绕过这部分解压缩了，。

共 3 条评论 >

👍 15



dream

2019-06-25

老师，我对消息层次、消息集合、消息、日志项这些概念与它们之间的关系感觉很懵，消息层次都分消息集合以及消息，消息集合中包含日志项，日志项中封装消息，那么日志项中封装的是producer发送的消息吗？

一个日志项中会包含多条消息吗？

消息集合中消息项封装的消息与消息层次包含的消息有什么关系呢？

这两个消息与producer发送的消息有什么关系呢？

一个消息集合对应是producer发送的一条消息还是多条消息呢？

最后，老师能不能详细说一下CRC校验，谢谢！

作者回复: 消息批次RecordBatch里面包含若干条消息（record）。

你可以认为消息批次和消息集合是等价的，消息和日志项是等价的。

这样消息层次有两层：外层是消息批次（或消息集合）；里层是消息（或日志项）。

Producer以recordbatch为单位发送消息，对于V2版本一个batch中通常包含多条消息。

在V2版本中，在batch层面计算CRC值；在V1版本中，每条消息都要计算CRC值。

共 6 条评论 >

👍 13



曾轶麟

2019-06-25

不认同，因为网络传输也会造成丢失，但是我建议可以在消息里面使用一种消耗较小的签名方式sign，比如多使用位移等方式，broke端也这么操纵，如果签名不一致证明有数据丢失，同时签名的方式可以避免CPU大量消耗

作者回复: 有一定道理。有机会我向社区反馈下：)

共 4 条评论 >

👍 11



南辕北辙

2019-07-01

老师有一点点迷惑，broker为了多版本消息兼容，意思是一份消息有多个版本存在吗，是这个意思吗？

作者回复: 同一台broker上可能存在多个版本的消息，但每条消息只会以1个版本的形式保存。

共 3 条评论 >

👍 9

