

开篇词-为什么要学习Kafka？

你好，我是胡夕，Apache Kafka的一名代码贡献者，目前在社区的Patch提交总数位列第22位，应该说算是国内比较活跃的贡献者了。

在过去5年中，我经历了Kafka从最初的0.8版本逐步演进到现在的2.3版本的完整过程，踩了很多坑也交了很多学费，慢慢地我梳理出了一个相对系统、完整的Kafka应用实战指南，最终以“Kafka核心技术与实战”专栏的形式呈现给你，希望分享我对Apache Kafka的理解和实战方面的经验，帮你透彻理解Kafka、更好地应用Kafka。

你可能会有这样的疑问，**我为什么要学习Kafka呢**？要回答这个问题，我们不妨从更大的视角来审视它，先聊聊我对这几年互联网技术发展的理解吧。

互联网蓬勃发展的这些年涌现出了很多令人眼花缭乱的新技术。以我个人的浅见，截止到2019年，当下互联网行业最火的技术当属ABC了，即所谓的AI人工智能、BigData大数据和Cloud云计算云平台。我个人对区块链技术发展前景存疑，毕竟目前没有看到特别好的落地应用场景，也许在未来几年它会更令人刮目相看吧。

在这ABC当中，坦率说A和C是有点曲高和寡的，不是所有玩家都能入场。反观B要显得平民得多，几乎所有公司都能参与进来。我曾经到过过一个理发厅，那里的人都宣称他们采用了大数据系统帮助客户设计造型，足见BigData是很“下里巴人”的。

作为工程师或架构师，你在实际工作过程中一定参与到了很多大数据业务系统的构建。由于这些系统都是为公司业务服务的，所以通常来说它们仅仅是执行一些常规的业务逻辑，因此它们不能算是计算密集型应用，相反更应该是数据密集型的。

对于数据密集型应用来说，如何应对数据量激增、数据复杂度增加以及数据变化速率变快，是彰显大数据工程师、架构师功力的最有效表征。我们欣喜地发现Kafka在帮助你应对这些问题方面能起到非常好的效果。就拿数据量激增来说，Kafka能够有效隔离上下游业务，将上游突增的流量缓存起来，以平滑的方式传导到下游子系统中，避免了流量的不规则冲击。由此可见，如果你是一名大数据从业人员，熟练掌握Kafka是非常必要的一项技能。

刚刚所举的例子仅仅是Kafka助力业务的一个场景罢了。事实上，Kafka有着非常广阔的应用场景。不谦虚地说，目前Apache Kafka被认为是整个消息引擎领域的执牛耳者，仅凭这一点就值得我们好好学习一下它。另外，从学习技术的角度而言，Kafka也是很有亮点的。我们仅需要学习一套框架就能在实际业务系统中实现消息引擎应用、应用程序集成、分布式存储构建，甚至是流处理应用的开发与部署，听起来还是很超值的吧。

不仅如此，再给你看一个数据。援引美国2019年Dice技术薪资报告中的数据，在10大薪资最高的技术技能中，掌握Kafka以平均每年12.8万美元排名第二！排名第一位的是13.2万美元/年的Go语言。好吧，希望你看到这个之后不会立即关闭我的专栏然后转头直奔隔壁的Go语言专栏。虽然这是美国人才市场的数据，但是我们有理由相信在国内Kafka的行情也是水涨船高。2019年两会上再一次提到了要深化**大数据**、人工智能等研发应用，而Kafka无论是作为消息引擎还是实时流处理平台，都能在大数据工程领域发挥重要的作用。

总之Kafka是个利器，值得一试！既然知道了为什么要学Kafka，那我们就要行动起来，把它学透，而学透Kafka有什么路径吗？

如果你是一名软件开发工程师的话，掌握Kafka的第一步就是要根据你掌握的编程语言去寻找对应的Kafka客户端。当前Kafka最重要的两大客户端是Java客户端和libkafka客户端，它们更新和维护的速度很快，非常适合你持续花时间投入。

一旦确定了要使用的客户端，马上去官网上学习一下代码示例，如果能够正确编译和运行这些样例，你就能轻松地驾驭客户端了。

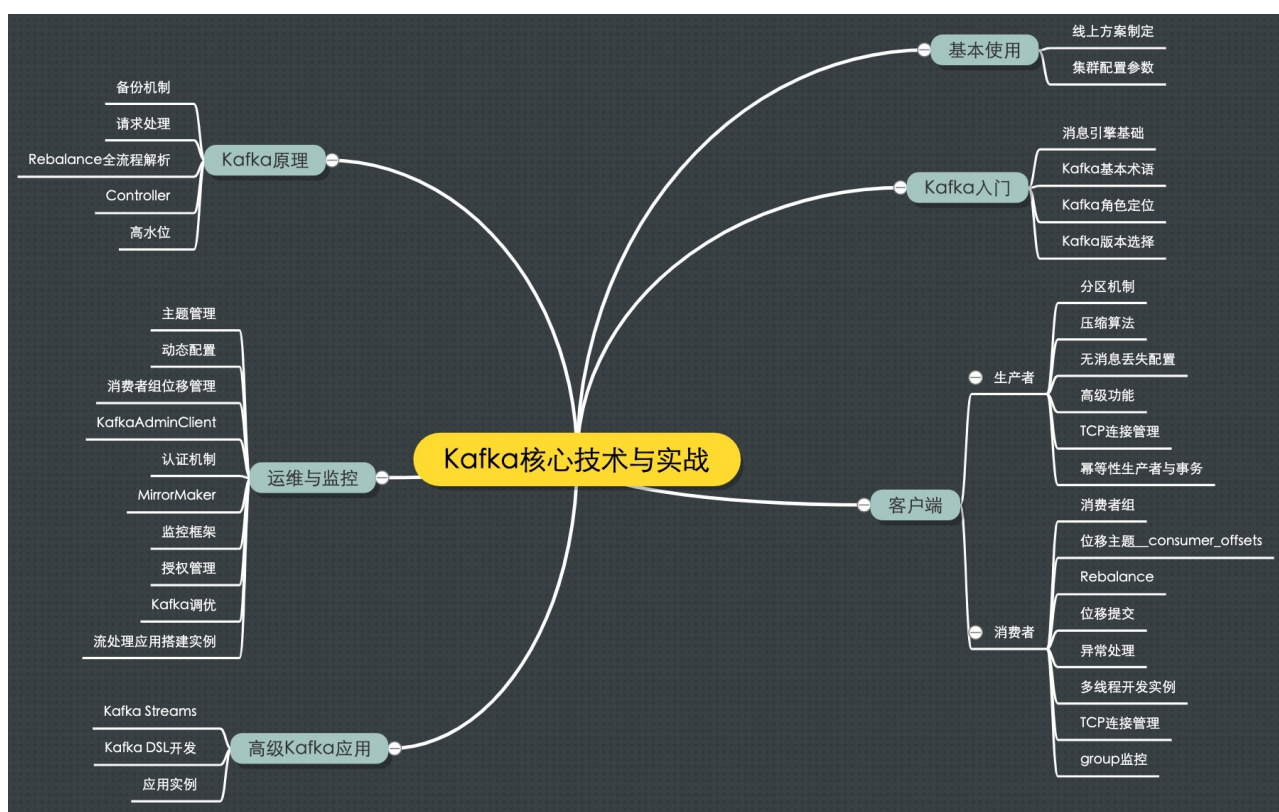
下一步你可以尝试修改样例代码尝试去理解并使用其他的API，之后观测你修改的结果。如果这些都没有难倒你，你可以自己编写一个小型项目来验证下学习成果，然后就是改善和提升客户端的可靠性和性能了。到了这一步，你可以熟读一遍Kafka官网文档，确保你理解了那些可能影响可靠性和性能的参数。

最后是学习Kafka的高级功能，比如流处理应用开发。流处理API不仅能够生产和消费消息，还能执行高级的流式处理操作，比如时间窗口聚合、流处理连接等。

如果你是系统管理员或运维工程师，那么相应的学习目标应该是学习搭建及管理Kafka线上环境。如何根据实际业务需求评估、搭建生产线上环境将是你主要的学习目标。另外对生产环境的监控也是重中之重的工作，Kafka提供了超多的JMX监控指标，你可以选择任意你熟知的框架进行监控。有了监控数据，作为系统运维管理员的你，势必要观测真实业务负载下的Kafka集群表现。之后如何利用已有的监控指标来找出系统瓶颈，然后提升整个系统的吞吐量，这也是最能体现你工作价值的地方。

在明确了自己要学什么以及怎么学之后，你现在会不会有一种感慨：原来我要学习这么多东西呀！不用担心，刚刚我提到的所有内容都会在专栏中被覆盖到。

下面是我特意为专栏画的一张思维导图，可以帮你迅速了解这个专栏的知识结构体系是什么样的。专栏大致从六个方面展开，包括Kafka入门、Kafka的基本使用、客户端详解、Kafka原理介绍、Kafka运维与监控以及高级Kafka应用。



• 专栏的第一部分我会介绍消息引擎这类系统大致的原理和用途，以及作为优秀消息引擎代表的Kafka在这

方面的表现。

- 第二部分则重点探讨Kafka如何用于生产环境，特别是线上环境方案的制定。
- 在第三部分中我会陪你一起学习Kafka客户端的方方面面，既有生产者的实操讲解也有消费者的原理剖析，你一定不要错过。
- 第四部分会着重介绍Kafka最核心的设计原理，包括Controller的设计机制、请求处理全流程解析等。
- 第五部分则涵盖Kafka运维与监控的内容，想获得高效运维Kafka集群以及有效监控Kafka的实战经验？我必当倾囊相助！
- 最后一个部分我会简单介绍一下Kafka流处理组件Kafka Streams的实战应用，希望能让你认识一个不太一样的Kafka。

这里不得不提的是，有熟悉我的读者可能知道我出版过的图书《Apache Kafka实战》。你可能有这样的疑问：既然有书了，那么这个专栏与书的区别又是什么呢？《Apache Kafka实战》这本书是基于Kafka 1.0版本撰写的，但目前Kafka已经演进到2.3版本了，我必须要承认书中的部分内容已经过时甚至是不准确了，而专栏的写作是基于Kafka的最新版。并且专栏作为一次全新的交付，我希望能用更轻松更容易理解的语言和形式，帮你获取到最新的Kafka实战经验。

我希望通过学习这个专栏，你不仅能够将Kafka熟练运用到实际工作当中去，而且还能培养出对于Kafka或是其他技术框架的浓厚学习兴趣。

最后我希望用一句话收尾与你共勉：Stay focused and work hard !



Kafka 核心技术与实战

全面提升你的 Kafka 实战能力

胡夕
人人贷计算平台部总监
Apache Kafka Contributor



新版升级：点击「 请朋友读」，20位好友免费读，邀请订阅更有**现金**奖励。

精选留言：

- 阿武 2019-06-03 23:51:11
老师 发量 保持的还不错！ [10赞]
- jeffery 2019-06-03 18:51:45
书+专栏更完美、期待蜕变 [5赞]

作者回复2019-06-03 19:00:43

感谢支持与鼓励：) 我们一起学习

- Fan 2019-06-03 18:12:03
Stay focused and work hard ! [5赞]

作者回复2019-06-03 18:53:15

特别喜欢这句话：)

- miofy 2019-06-03 18:50:28
相比人工智能和区块链，大数据和云计算成熟很多。云计算的成熟，又促成了大数据技术的大规模普及。不过真正有“大”数据的公司并不多。 [4赞]
- 大番茄 2019-06-04 10:11:35
 - 1.希望老师 讲下rabbitmq和kafka区别，传统企业如何选择?
 2. 课程更新周期是怎么样子的？
 3. 学习这个课，老师能推荐些辅助资料么？

感谢 [3赞]

- Geek_e5967d 2019-06-03 21:39:54
赞一个，最近开始学习kafka运维，我们使用的还是用的0.8.1版本🐼 [3赞]
- kaiux 2019-06-03 17:20:21
这个专栏等了一段时间了，前段时间正好在学习Kafka，后期会用到项目实践中，期待能有一个深入的理解。 [3赞]

作者回复2019-06-03 18:57:47

感谢支持与鼓励：) 期待你的反馈我们一起学习~~

- Fouy_飞虎 2019-06-03 18:40:02
赞一个，正准备深入学习一下。 [2赞]

作者回复2019-06-03 19:01:08

感谢支持与鼓励：)

- 开水 2019-06-05 07:41:53
老师，后续课程中会对比一下其他消息框架么。比如rocket，rabbit [1赞]

作者回复2019-06-06 08:05:31

坦率说不会有太多涉及，毕竟还是以介绍Apache Kafka为主

- roger 2019-06-05 07:41:28
有新出比kafka更灵活的消息队列吗？ [1赞]

作者回复2019-06-05 08:26:30

不确定百分之百理解了“灵活”的含义，不过可以关注下Apache Pulsar

- Alexdown 2019-06-04 14:40:01

期待，同时期望老师在专栏中穿插一些高频的面试题 ——作为消息中间件使用时，消息的可靠传输、顺序、重复消费等问题在kafka中是如何解决的。这不仅仅对面试有帮助，也对进入公司后快速上手熟悉系统、排查线上问题等有帮助，谢谢 [1赞]

作者回复2019-06-04 19:12:14

嗯嗯，专栏中有些文章的标题就是常见的面试题。比如消息无丢失配置，再比如消费者组rebalance过程等，可以关注下~~

- 绝影 2019-06-06 11:27:33

跟着老师，系统学习kafka

- 佑儿 2019-06-06 10:24:55

打卡

- 蹇小东 2019-06-06 00:16:42

打卡学习

- 风起青萍 2019-06-05 23:14:39

我也想成为大佬。学习

- 侯代烨 2019-06-05 20:27:31

我们的kafka集群，每个节点只能存储35G的数据量，超过这个量之后，kafka进程就会挂掉，启动会报内存溢出的错误，困扰很久了，不知道学习了这个课程，能不能解决这个问题。

作者回复2019-06-06 08:08:52

后面会有性能调优的部分。分析性能问题还是要结合具体的错误来看，比如内存溢出要明确是哪部分内存溢出，然后才有可能有针对性的调整

- Michael 2019-06-05 07:51:22

老师 这个专栏是从最初级的入门开始的嘛？怕后面听不懂，没接触过

作者回复2019-06-05 08:52:40

专栏前面有一些术语和基本概念的介绍。如果没接触过可以采用这样的学习方法：先把专栏中基本概念和术语搞懂，然后一点点实践专栏后面的内容，跟着专栏一起学：)

- 虹桥念苏 2019-06-04 23:09:36

一直在看Flink，感觉Kafka还是深入了解一下比较好，不知道胡老师会不会带一些pulsar比较，这个后起之秀不能忽视，谢谢！

作者回复2019-06-05 08:14:01

我和Pulsar的郭总、翟总都相识，都Pulsar不敢妄言比较，但我同意Apache Pulsar是一款非常优秀的消息引擎：)

- 小北 2019-06-04 22:21:19

希望后续会有php以及go客户端的讲解

- huaweichen 2019-06-04 13:51:34

老师的Kafka Stream是不是讲得有点少？

作者回复2019-06-04 18:11:46

嗯嗯，目前Kafka Streams还在积极演进中。我挑了一些已经比较稳定的功能来讲