

开篇词 | 四纵四横，带你透彻理解分布式技术

2019-09-23 聂鹏程 来自北京

《分布式技术原理与算法解析》



你好，我是聂鹏程，智载云帆的 CTO。在接下来的三个月时间里，我将带你打卡分布式技术。

在众多计算机技术当中，分布式技术无疑是最璀璨的明珠之一。毫不夸张地说，没有分布式技术就没有互联网，也就没有现在的阿里巴巴、腾讯、亚马逊、Facebook、谷歌等科技巨头，更不会有以信息技术为核心的、对人类历史产生巨大变革的第三次工业革命。万维网、Email、DNS 等，都是分布式系统的典型代表。

随着分布式技术的不断发展，它也早已不再局限于传统的互联网等应用场景。在今年的两会期间，全国政协委员、360 董事长周鸿祎更是大唱 I' M ABCDE 字母歌。**IMABCDE 这 7 个字母所代表的 IoT 物联网、Mobile 移动计算、AI 人工智能、Blockchain 区块链、Cloud 云计算、Data 大数据、Edge 边缘计算，无不都是以分布式技术为基石。**

无疑，**谁更好地掌握了分布式技术，谁就更容易在新一轮技术浪潮中获得主动。**在全球经济增速趋缓的大背景下，与许多应用业务大量裁人形成鲜明对比的是，各大巨头的中间件团队、实验室等基础部门，依然在大规模地招兵买马。随着业务扩展，以及 IMABCDE 等新兴技术领域的布局，分布式技术人才已然成为巨头们争夺的焦点。

一方面是各大厂商的求贤若渴，一方面是分布式专业技术人才的一将难求。在多年的面试中，我经常能体会到，有些面试者确实非常积极主动，但他们表现出来的水平却无法通过面试。分布式技术人才市场的供应与需求，俨然一首冰与火之歌。

2007 年，我在西安电子科技大学攻读博士期间，就开始研究并行与分布式计算；毕业后，在 IBM 做过 HPC 大规模负载管理系统 LSF 相关的设计和研发工作，在华为负责过分布式 IoT 相关项目的架构设计，以及电信级业务微服务框架、函数服务框架的设计工作，也从事过区块链相关的研究工作。现在，我在智载云帆负责技术体系的构建，专注于无服务器 Serverless 的架构实践。

从我深入研究分布式技术这十多年的经验来看，分布式技术概念繁多、知识庞杂、新兴技术层出不穷，令许多新手望而却步，而许多有一定年限工作经验的老手，虽然也能对一些概念滔滔不绝，但追问到实质性问题就变得磕磕巴巴，顾左右而言它。比如：

各种分布式概念、名词学了一大堆，但经常张冠李戴，傻傻分不清楚；

做了多年技术，也参与了很多分布式技术实践，却无法回答工作中各种分布式技术、组件、框架选型背后的根源；

在一个分布式技术配套的典型场景往往能驾轻就熟，但一旦稍微变更考察业务场景、业务目标后，就变得毫无头绪。

究其原因，主要是**知识碎片化、不成体系、见树不见林**。如果再追究更深层次的原因，那无外乎就是两点：

在计算机学科课程设置中，分布式技术尴尬如同中小学中的性教育，重要但不受重视。鲜有高校在计算机本科专业中设置分布式课程，即便是有些高校在研究生教育中设置了相关课程，也是如同高手过招点到为止。

信息碎片化与信息孤立。在信息泛滥的信息时代，各种经典教材、最新文章自然是唾手可得。但，教材固然经典但严谨有余浅出不够，最重要的是没有与时下最新的场景相结合，一方面是因为教材“年久失修”，另一方面确实是因为分布式领域新技术推出的速度令人叹为观止；而网上的各种技术文章虽然多，却鲜有体系化的说明，一个个概念如同一座座信息孤岛。**如果不能体系化地理解这些概念，何谈掌握，更谈不上真正地去综合运用这些知识了。**

因为工作太忙，这些年我整体而系统的输出比较少。偶然一次听到任正非的讲话，他提到了基础教育、孩子是一个社会的未来，这让我感触良多。我想，如果说一个社会的未来，离不开朝气蓬勃的孩子们，那么**一个行业的兴盛，同样也离不开一个广泛的从业者基础。**而我如果能做好分布式通识课这样的一个专栏，既可以对自己这些年的经验做一次系统梳理，温故而知新，又能帮助更多的人系统理解并掌握分布式核心技术，为整个行业的兴盛略尽绵薄之力，何乐而不为呢？

其实，在工作、面试、演讲等多种场合，也经常会有人问我：“聂博士，分布式领域的新概念繁多、各种框架五花八门、各种组件层出不穷，应该如何应对啊？”我回答说：“其实你已经有了答案了。”

看着他们满脸疑惑，我笑着问：“RISC 芯片，程序设计中的封装、继承，还有现在提倡的中台战略，它们都在做一件什么事情呢？”他们答道：“莫非是重用？”

我说：“是的，既然指令可以重用，代码可以重用，技术、业务、数据等都可以重用，为什么知识体系不可以呢？学好分布式通识课，掌握了分布式的核心技术、体系，你就会发现很多新技术、新框架、新组件只不过是‘新瓶装旧酒’，将分布式核心技术进行了再包装、再组合，至多也就是做了一点延伸而已。”

那么，分布式通识课究竟该如何学呢？在接下来的三个月时间里，我会遵循以下 4 个思路带着你一起学习。

第一，分布式技术错综复杂，各种技术相互耦合，确实无法简单地像网络等技术一样划分层次，所以我会结合自己多年的积累和思考，首先为你梳理出一个脉络清晰、**四纵四横**的分布式核心技术知识体系，然后从这个纵横的技术体系中抽取最核心、最普适的技术思想以及概念，结合各种适用场景一一解析。这样的设计，旨在帮助你找到核心知识点，并将这些知识点联系

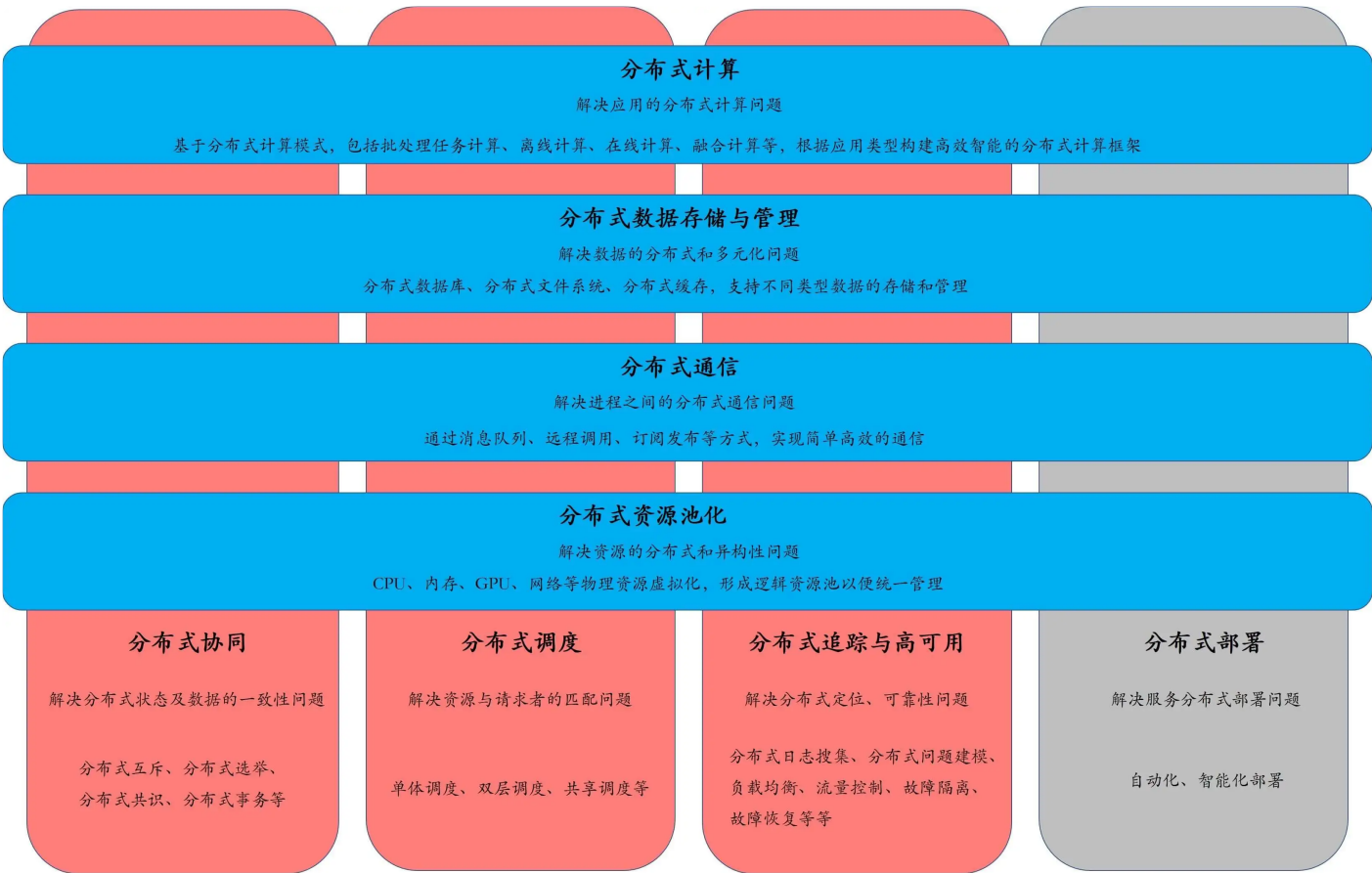
起来，快速形成分布式核心技术的知识网络，从而形成自己的技术判断力，进而规划出自己的技术路线。

第二，从一个熟知的事物出发，进行浅出的讲解，帮助你从已有知识体系扩展到新的知识体系，从而迅速、牢固地掌握分布式技术的核心知识点。

第三，透过表象深入讲解技术本质，而不是 case by case 地讲解，帮助你知其然并知其所以然，真正做到理解与运用时的举一反三。

第四，针对同一分布式问题的不同方法，从多维度、多角度进行对比、分析，方便你在工作中灵活选型，避免重复“造轮子”。你甚至可以综合权衡各种方法的优缺点，提炼发明出新的方法，最终做到活学活用。

讲到这里，你是不是也有点摩拳擦掌、跃跃欲试了呢？“分布式世界这么大，我要去看看！”别慌，请先看完这份技术地图。



分布式四纵四横知识体系图

首先，按照业务的架构层次栈，我自底向上按照资源、通信、数据与计算的维度，梳理出了 4 个技术层次：分布式资源池化、分布式通信、分布式数据存储与管理、分布式计算。这样的划分符合业务架构设计的一般规律，即**“在一定资源上，进行一定通信，通过一定计算，完成一定数据的加工和处理，从而对外提供特定的服务”**。另一方面，这样的划分也整合了零散的知识点，具有完备性。

既然横向的 4 个层次都已经完备了，那为什么又多出了 4 个纵向的技术呢？如果我们把横向的 4 个层次比作派生类的话，那么纵向的 4 条技术线应该是它们的基类。因为，在分布式环境下，无论是资源、通信、数据还是计算，都需要去解决协同、调度、追踪高可用，还有部署的问题。因此，我从横向的技术层次中，提炼出分布式协同、分布式调度、分布式追踪与高可用、分布式部署 4 个纵向技术线。由于分布式追踪、分布式部署虽属于支撑技术，但并不会影响业务的构成，因此我不会在本专栏中进行讲解。

最后，如果说现在分布式领域里各种包装出来的、五花八门的新技术，像是令人高不可攀的女神、男神的话，那么这个分布式通识课程中所提炼出来的体系和核心知识点无疑就是女神、男神素颜的样子。我想说，等你**看尽素颜，无论是女神、男神也好，还是各种高大上的技术也好，也就不会觉得那么高不可攀了。**

既然你已经看到了这里，相信你也看到了学习分布式技术知识的迫切需求，那么不妨请你在留言区做个自我介绍，给我说说你的困惑，也说说你想通过这个专栏收获些什么，这样我后续也可以根据你的情况进行有针对性的讲解。

我是聂鹏程，今天的内容就到这里了，我们下一讲再会。

© 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。

精选留言 (98)



而立斋

2019-09-23

想达到的第一个目标就是:在跟别人聊分布式的时候，有话可说，有深度的话可说，有体系深度的话可说。第二个目标就是:在实际的工作中能够运用自己对分布式的理解做出更好的选

型，更好的架构，同时让自己更好的进步。今年32岁，我渴望知识，渴望突破，更渴望成长！

作者回复：我发现现在技术圈存在一种怪相，很多从业者都觉得做技术就是吃青春饭，甚至一些所谓的大V也在社交媒体上助长这种观念。

商品市场上从来不缺普通的产品和服务，缺的是有竞争力的、有差异化的产品和服务。我觉得人才和商品是一样的，人才市场上也从不缺普通的求职者，缺的是有竞争力的、有差异化的人才。

很高兴你没有存在这种技术是吃青春饭的误区！相信只要你肯学习、积极进取，把自己的积累做广做深，没有人能阻止得了你前进和上升的步伐。It's never too late to learn. 加油！

共 14 条评论 >

👍 134



静水流深

2019-09-28

老师，我觉得我很笨，学习啥都很慢，比如有些算法问题，我都仔仔细细的想好多为什么，有时候觉得高手就是太聪明，写得代码那么优秀！别人为什么学得那么快？内心着急，但是我也不甘放弃。

作者回复：我们在战场上可以看到非常典型的两种兵种，机枪兵还有狙击兵，一个快一个慢。其实很多时候快是一方面，准又是另一方面！

积跬步，而终至千里！加油！

共 5 条评论 >

👍 28



贤蛋蛋

2019-09-24

即将毕业，在阿里实习，遇到了各种各样的中间件，希望能好好学习这门课，把工作学习中的知识构建到自己的体系结构中

作者回复：相对同龄人，你无论是履历还是学习意识，都已经赢在起跑线上了！加油！



👍 28



颖颖

2019-09-24

之前学过mit的.6.824分布式课，里面主要就是通过读论文，然后做实验来帮助理解分布式系统中的各种理论，但是一个比较大的问题就是缺乏应用场景，理论联系不到实践中，希望这门课能做到第二点的

共 3 条评论 >

👍 24



djfhchdh

2019-09-24

这个知识体系总结的很好啊

作者回复: 谢谢！宝剑赠英雄，红粉赠佳人！

共 2 条评论 >

👍 17



钱

2020-02-12

阅过留痕

我认为带着问题学习，采用淘金式的学习方式，先搞清楚自己的技能树是什么样子，新的技能枝叶应该怎么融合上去是比较好的学习方式。

那问题来了：

- 1：为什么说没有分布式技术就没有BATJTMD？
- 2：没有分布式技术之前的网络世界是怎样的？
- 3：分布式技术是怎么催生的，又是怎么发展的？
- 4：分布式技术解决了什么痛点问题？
- 5：分布式解决的痛点问题又是怎么产生的？
- 6：分布式技术的基础是啥？
- 7：分布式技术的核心原理是啥？
- 8：分布式技术解决问题的同时有带来了那些痛点？
- 9：分布式技术带来的问题又是怎么解决的？

OK，学完本专栏，如果这些问题都清楚了，那就不白花钱。

我现在的理解，分布式本质意思就是多个的意思，多个啥呢？直观感受就是多个电脑，至少有两台吧！为啥需要多个电脑？因为一个电脑不行，不能解决问题，不能解决啥问题？不能解决存储的问题，它空间有限，假如一台电脑的容量是无限的还有问题吗？还有的，还有就是性能问题，假如一台电脑容量无限计算数据和移动数据不需要时间立马就能完成呢？如果假设成立，我真不知道世界上还需要两台电脑的原因是什么？可能是防止这台电脑坏了吧！

OK，我的观点是

- 1：分布式解决的痛点问题就是单机容量和性能所形成的瓶颈问题。
- 2：分布式的核心原理就是一台电脑不行就来两台，其他的一切都是围绕怎么把两台电脑像一台电脑一样来使用未出发点来展开了。

3: 分布式的系统把N台电脑幻化成了一台超级电脑, 他的容量无限性能极佳斗转星移永不宕机。

共 3 条评论 >

👍 16



老男孩

2019-09-24

浮生有梦三千场, 穷尽千里诗酒荒, 徒把理想倾倒, 才见此专栏。相见恨晚啊!

作者回复: 人生最难的是等待, 而最美的是有值得等待的东西!



👍 12



oush

2019-09-23

目前工作是做大数据方面的开发, 各种技术框架层出不穷, 如: spark、flink、hadoop等, 想了解其底层的基本知识, 希望本课程能学习到这些东西。

作者回复: 你说的就属于我在文中提到的IMABCDE中的D (Data大数据) 的范畴, 大数据当然也是以分布式技术为基石, 常用的分布式核心技术, 比方说分布式数据存储、分布式计算、分布式协同等都会用到。与之相关的技术应用, 我也会在后续文章中进行展开讲解。

共 2 条评论 >

👍 12



刘福洋

2019-09-28

会Rust和Go的也可以试试来跟着PingCAP这个撸代码课程来做做看哦

<https://github.com/pingcap/talent-plan>

不知咱们这个专栏有没有什么代码作业玩呢?



👍 11



AKA三皮

2019-09-23

工作中使用eureka作为服务注册中心, 然后就接触到CAP理论, 进而想去理解CP、AP分布式项目的差异, 比如zookeeper, etcd, 然后又发现必须去了解各种一致性协议, zk的zab, etcd的raft。

作者回复: 你所关心的这些内容都是本专栏会讲到的, 敬请关注后续更新。只要你一直怀揣这种永不止步、积极探索的精神, 想不进步都是不可能的! 加油!



👍 10



allean

2019-09-23

希望通过这个专栏可以入门分布式

作者回复: 我做这个专栏的初衷就是: 通过划重点、寻路线、建体系, 帮助更多的人打通分布式技术任督二脉, 并提升职业竞争力



👍 9



Louris

2019-09-24

研究生在读, 现在课题组是分布式机器学习系统, 希望多了解分布式知识, 有个整体概念, 方便后面深入学习。

作者回复: 研究生阶段就有这样的体系化意识实属难能可贵! 加油!



👍 8



你猜我是不是隋晓宇

2019-09-23

哈哈, 分布式的大门向我打开了

作者回复: God helps those who help themselves. 加油!



👍 8



杜do度

2019-09-25

我想说, 等你看尽素颜, 无论是女神、男神也好, 还是各种高大上的技术也好, 也就不会觉得那么高不可攀了。

——聂老师是一个有故事的男人!!!

作者回复: 我觉得你就是传说中的子期:)



👍 7



leslie

2019-09-23

现在的系统已经不再是过去/传统的C/S或B/S: 面对越来越多的高并发、CAP理论, 分布式确实是解决问题的不错方式。

这就像运维已经发展到AIOPS或SRE了: 希望学习中能够补充这块缺陷, 把相应的知识梳理起来并适当补充-跟上时代的步伐。



👍 6



50包邮解君愁

2019-09-23

小白弱弱问一下: 数据挖掘需要吗? 比如hadoop,hive,spark

作者回复: 你说的就属于我在文中提到的IMABCDE中的D (Data大数据) 的范畴, 大数据当然也是以分布式技术为基石, 常用的分布式核心技术, 比方说分布式数据存储、分布式计算、分布式协同等都会用到。与之相关的技术应用, 我也会在后续文章中进行展开讲解。



👍 5



每天晒白牙

2019-09-23

老师代码用啥语言呀

作者回复: 我的《分布式技术原理与算法解析》专栏并没有严格限定语言。其实分布式计算当中, 包括微服务在内的很多技术努力在做的事情就是允许语言的多元化。正所谓尺有所短寸有所长, 十八般兵器, 不一定要样样精通, 但如果能代表性的掌握几种, 那么当我们面临问题时, 肯定会有更好的选择。

共 3 条评论 >

👍 5



jeri

2020-03-20

四横四纵, 精辟。

四横: 在一定的分布式资源上进行分布式通信, 然后通过分布式计算的处理结果, 落到分布式

存储上；

四纵：在构建分布式系统时，总要关注分布式协同、分布式调度、分布式高可用追踪和分布式部署。

作者回复：👍👍



👍 4



张理查

2019-12-19

分布式技术原理与算法#Day1

入行大数据快五年了，围绕toB的项目也做了不少，大数据从蓝海变成红海，apache为了满足不同场景孵化了一个又一个项目，但我常问自己，那么究竟什么是大数据？

大数据真的是数据大吗？见仁见智，GB TB 的项目都接触过，但以前数据就没上GB吗？

大数据真的是原来无法处理吗？不见得，无非受限于硬件发展，十年前上大2，那时候普通PC硬盘500GB/4core/8G已经挺大了，现在来说你看看苹果刚出的Mac pro，这不服务器吗？原来无非就是慢点而已。

大数据是什么呢？我认为就是利用分布式解决一些数据问题而已。至少1台机器你还真不好意思说大数据吧？

分布式包含哪些点呢？就是聂老师说的四横四纵，横着就是大数据的日常，资源，通信，存储，计算，竖着就是运维和架构关心的协同，调度，高可用，可扩展(或者说部署)

希望通过上课和打卡再研究新老组件时，既见树木又见森林，加油！

作者回复：通过已有知识体系延伸到新的知识体系是一种非常高效的学习方法，而且更有利于灵活运用。你做得很好，加油！

共 2 条评论 >

👍 4



QQ怪

2019-09-23

好期待老师讲的分布式精髓

作者回复：谢谢！我也非常期待能帮助到你！一起加油！



👍 4

