

开篇词 | 从这里开始，带你走上硅谷一线系统架构师之路

蔡元楠 2019-04-15



00:00

讲述：蔡元楠 大小：7.51M

08:11

你好，我是蔡元楠。目前是 Google Brain 的软件工程师。

在接下来的 4 个月时间里，我会与你一起探索大规模数据处理的世界。

在开始我们的系统性学习之前，我想先和你分享两个我亲历的故事，借此告诉你，我为什么要开这个专栏。

2014 年，我刚开始在美国找工作，在一次面试中，面试官让我解释一下 C++ 的 smart pointer 和 string view。我完全回答不上来。

当时我面露难色怀疑人生，难道我的前半生学的是“假的”C++ 吗？

回想学习经历，虽然我在一个还可以的高校里学习了 C++ 课程，考试成绩也是 90 分以上。但我的学习资料只有按 C++ 98 标准编写的教材，和当时流行的 CSDN 论坛文章。

而当时的教材里根本没有提到过 smart pointer 和 string view。

没想到工业界早就进入了 C++ 0x11 标准甚至是 C++ 17（当时是试验标准）。我后悔没有学习紧跟时代的最新技术，而信息的不对称就会造成巨大的认知偏差。

那学习“最新技术”一定就是好事情吗？我想通过第二个故事来回答这个问题。

2017 年，我帮 Google Ventures（Google 的风险投资基金，会寻找并帮助优秀初创公司）在投的初创公司做导师，那时候经常参加一些对方公司的技术架构评审。

一次评审中，对方的技术 VP 眉飞色舞地介绍他们的技术框架和××大厂一样，罗列了 Kylin 和 Tornado 等一些时髦的技术名词。

因为我并不了解提到的几个技术，就好奇地问为什么 Kylin 适合他们团队，Kylin 为他们解决了哪些独特的问题？

他当时的回答并没能说服我和别的同事：“因为××大厂也在用，这就是未来的技术方向”。

看得出来，这位技术 VP 还没有真的搞清楚使用一个技术的原因。

学会用一个技术只是第一步，最重要的是要追问自己：

- 这个技术解决了哪些痛点？
- 别的技术为什么不能解决？
- 这个技术用怎样的方法解决问题？
- 采用这个技术真的是最好的方法吗？
- 如果不用这个技术，你会怎样独立解决这类问题？

如果没有这些深层次的思考，你就永远只是在赶技术的时髦而已，不会拥有影响他人的技术领导力。

事实上在 Google，类似这样的“灵魂追问”每天都在发生。

这里敢于打破任何权威，所有的技术设计都是从问题出发。每一个工程师都会独立思考究竟什么是最佳方案，而不是照搬现有结论。

正如乔布斯所说，过去的点最终会连成线。这两个经历让我深感有使命去帮助更多技术同行，比如第一个故事里的我，或者是第二个故事里的初创公司 VP。我想设计一个专栏，去解决故事里的问题。

第一，我想要介绍硅谷最前沿技术和真实的案例。

比如，在大规模数据处理领域，MapReduce 或者 Apache Storm 的不少设计理念已经无法胜任最新的挑战。

所以，我会介绍最新的知识，例如框架层面的前后端分离理念，和批处理流处理统一的思想。

第二，我不想只停留在照本宣科的层面。

正如上文所说，学会用一个或者两个技术是不够的。

更重要的是，我会剖析技术框架产生的原因和它们解决的问题。这样当下一次你再碰到相似的问题时，就不用照搬别人的方法。

为什么写大规模数据处理？

为什么会选择大规模数据处理这个主题呢？并不是因为我觉得这个主题受众多、销量好，相反，我认为大部分人都还没有正确理解数据处理技术，常常见到的误区有如下几种。

第一，低估了数据处理的重要性。

因为我在 Google Brain 的 AI 应用领域工作，切身感受到过，没有高质量的数据处理的话，人工智能是只有人工没有智能的。

Google 也曾在很长的一段时间里低估过数据处理。

例如，在语义理解上，Google 认为自己有最多的搜索文本数据，最好的算法，那就一定能把语义理解做的最好。

可是到 2016 年左右，一个名不见经传的德国小公司却一举超过了 Google，大家都很惊讶。后来发现原来他们凭借的是高质量的数据标注和处理。

第二，低估了数据处理工程师在组织架构上的重要性。

许多工程师都喜欢自嘲自己的工作“搬砖”，事实也正是如此。

包括我在内，很多人的工作内容都避不开数据的搬运和处理。把数据从这个格式处理成那个格式，把数据从这个数据库搬到那个数据库，这个服务器搬到那个服务器，这个客户端搬到那个客户端。

可能连你自己都还没有意识到，即使是一个写前端的工程师，他的很多工作还是数据处理。

大数据领域泰斗级人物 Jesse Anderson 曾做过一项研究，一个人工智能团队的合理组织架构，需要 4/5 的数据处理工程师。很不幸，很多团队没有认识到这一点。

第三，低估了数据处理规模变大带来的复杂度。

我把这个专栏定位在“大规模”数据处理，因为我想着重在数据规模变大时需要的技术思想。

很多人可能还没有遇到过“大规模”数据的问题，容易把问题想简单了。

我在 Google 面试过很多优秀的候选人，应对普通的编程问题，他们能够用算法和数据结构解决得很好。可是当我追问数据规模变大时怎么设计系统，他们的回答却并不让人满意。

当你的产品从 1 万用户到 1 亿用户，技术团队从 10 个人到 1000 个人，之前的方法还能奏效吗？

第四，高估了上手数据处理的难度。

一方面我们需要认识到大规模的数据处理是有复杂的因素的。但另一方面，我想在这个专栏里教会你，有了正确的工具和技术理念，现在上手数据处理并不困难。

在 Google，我见到很多应届生来了半年后也能轻松应对上亿的数据量。

我给开篇词起名为《从这里开始，带你走上硅谷一线系统架构师之路》，就是为了给你设计切实可操作的学习路径，让你比别人更准确深入地掌握实用的大规模数据处理技术，最终通往硅谷一线系统架构师的水平。

因此，我们的学习路径会是这样的。

第一部分，先会用原汁原味**最实际的硅谷一线大厂的案例**，向你解释 MapReduce 为什么不能应对最新的技术挑战。然后我会从实际的问题出发，从头开始引导你怎样从顶层设计一个数据处理框架。

第二部分，同样是结合实战案例，来讲解在数据处理框架的使用和设计中必需的一些基础知识。这些案例紧贴应用，可能就是**你的团队明天会碰到的问题**。

第三、第四部分深入拆解了 Apache Spark 和 Apache Beam。不仅会用实际的案例教会你如何使用，还要教会你为什么它们这么设计。你会发现它们的设计其实大致和第一部分的顶层设计是一致的。**下一次，即使这个世界一无所有，你也能构建类似的框架解决一系列问题。**

第五部分按 Google T6 级别设计，是带着代码的真枪实弹的架构设计。毫不夸张地说，能完整掌握第五部分的思想精髓，你就能比肩硅谷一线大规模数据处理架构师。

第六部分着重培养你的技术远见。因为**是否能现在就开始准备应对 10 年后人类社会的技术挑战，是你拉开与别人差距的重要一站。**

《大规模数据处理实战》课程目录

- ① 从这里开始，带你走上硅谷一线系统架构师之路

模块一：直通硅谷大规模数据处理技术

- ① 为什么 MapReduce 会被硅谷一线公司淘汰？
- ② MapReduce 后谁主沉浮：怎样设计下一代数据处理技术？
- ③ 大规模数据处理初体验：怎样实现大型电商热销榜？

模块二：实战学习大规模数据处理基本功

- ④ 分布式系统（上）：如何用服务等级协议 SLA 来评估你的系统？
- ⑤ 分布式系统（下）：架构师不得不知的三大指标
- ⑥ 如何区分批处理还是流处理？
- ⑦ Workflow 设计模式：让你在大规模数据世界中君临天下
- ⑧ 发布 / 订阅模式：流处理架构中的瑞士军刀
- ⑨ CAP 定理：三选二，架构师必须学会的取舍
- ⑩ Lambda 架构：Twitter 千亿级实时数据分析架构背后的倚天剑
- ⑪ Kappa 架构：使用 Kafka 锻造的屠龙刀

模块三：抽丝剥茧剖析 Apache Spark 设计精髓

- 12 Spark 的独有优势：为什么世界需要 Spark ?
- 13 弹性分布式数据集：Spark 大厦的地基为何如此设计（上）
- 14 弹性分布式数据集：Spark 大厦的地基为何如此设计（下）
- 15 Spark SQL：Spark 数据查询的利器
- 16 如何用 Spark DataFrame API 进行实时数据分析？
- 17 Spark Streaming：Spark 的实时流计算 API
- 18 WordCount：从零开始运行你的第一个 Spark 应用
- 19 综合案例实战：加州房屋信息的线性回归模型 & 成年人收入的预测模型
- 20 更多、更快、更好、更省：实例解析 Spark 应用程序性能优化
- 21 深入对比 Flink 与 Spark: 帮你的系统设计两开花

模块四：Apache Beam 为何能一统江湖

- 22 Apache Beam 的前世今生
- 23 站在 Google 的肩膀上学习 Beam 编程模型
- 24 PCollection：为何 Beam 要如此抽象封装数据？
- 25 Beam 数据转换操作的抽象方法
- 26 Pipeline：Beam 如何抽象多步骤的数据流水线？
- 27 Pipeline I/O：Beam 数据中转的设计模式
- 28 如何设计好一个 Beam Pipeline ?
- 29 如何测试 Beam Pipeline ?

模块五：决战 Apache Beam 真实硅谷案例

- 30 Apache Beam 实战冲刺：Beam 如何 run everywhere?
- 31 WordCount Beam Pipeline 实战
- 32 Beam Window：打通流处理的任督二脉
- 33 横看成岭侧成峰：再战 Streaming WordCount
- 34 Amazon 热销榜 Beam Pipeline 实战
- 35 Facebook 游戏实时流处理 Beam Pipeline 实战（上）
- 36 Facebook 游戏实时流处理 Beam Pipeline 实战（下）

模块六：大规模数据处理的挑战与未来

- 37 5G 时代，如何处理实时超大规模物联网数据？
- 38 大规模数据处理在深度学习中如何应用？
- 39 从 SQL 到 Streaming SQL：突破静态数据查询的次元
- 40 总结与答疑

在刚开始动笔写专栏的时候，我就在设想，什么样的人会是这个专栏的目标读者呢？

直到专栏快上线，我写下这篇开篇词，我才真正定义读者的标签——应该就是跟我一样渴望成长的人。是的，我和你一样，都渴望成长——渴望知识的成长，渴望经验的成长，渴望财富的成长。

所以我想把这个专栏设计成一份共同的成长规划，而不是一本死板的教材。

正如我在开头的小故事里所说的，这个世界没有谁是绝对的权威。

我希望你每一期都能在留言栏里质疑、提问和讨论。这些互动能帮助我和别的同学一起提高。

最后，我期待和你一起开始学习，共同成长！



©



由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。

Ctrl + Enter 发表

0/2000字

提交留言

精选留言(49)



As Sunshine 置顶

这个课程对学员的要求有没有什么限制

👍 3 2019-04-15

作者回复: 很好的问题，另一个同学也提到了类似问题。我们在内容设计时并没有对读者对知识背景作任何假设，所以即使一些基础的技术概念都会举例解释一下（如果你会了可能会觉得啰嗦）。有一些任何语言的编程经验会看起来快一点，因为有一些示例代码是Python的。但是设

计类型的案例，我不觉得有特别的技术要求。希望你后面继续跟踪一下吧，如果有哪些讲的不清楚，或者解释的过多，后面可以调整内容。谢谢提问！



你家隔壁宋叔叔 置顶

灵魂追问需要通过阅读文档和源码，并加上自己实践和思考才能够回答。

👍 3 2019-04-15

作者回复: 的确是互为补充，专栏讲解的案例是有限的，但我希望在有限的案例里把思考方式讲清楚。另外相比文档，这里设计的案例会更实际一点。也欢迎你把自己专栏外的学习收获在这里分享。



韩程 置顶

老师你好，你上文提到的AI落地的基础是大规模的数据和高质量的标注，目前能满足的这个条件是否只有一些超大规模的一线互联网公司。那是否意味着大数据处理也只有在这些公司中才能发挥真正的价值，那对于在小型互联网公司工作的程序员，学习大数据处理的意义在哪里呢？

👍 2 2019-04-16

作者回复: 我觉得这个问题很好啊。我一部分同意大规模的互联网公司天生数据量大一点。另一方面，1. 对于公司来讲小型互联网公司甚至是传统企业，并不是不需要数据处理技能，而是他们还没有从数据中挖掘business insight的意识，没有数据驱动决策的意识，甚至没有收集数据的意识。举个我工作中见到的例子，比如有奶牛的农户几十年来根本不知道什么是数据，但是我们帮他们细致的搜集牛的每天的数据，比如饮食，运动，作息，产奶，他们能从中找到最经济最优的饲料投放。2. 对于个人来讲一定要看长期的职业发展，公司会从小变大，职位会从低变高，当你能更多影响决策当你数据量变多，当你跳槽之后，数据的处理能力都是至关重要的。我们可以继续就这个问题探讨！



流殇忘情 置顶

有个问题想问一下老师，为什么Google开源的Apache Beam不用Go写，而是使用Java来写呢？理论上无论从性能还是代码规范来说Go是个更好的选择，也能推动Go社区的发展，为什么最终还是用了Java实现呢？

👍 1 2019-04-15

作者回复: 我第二篇会提到，beam采用前后端分离的设计。也就是说作为普通开发者可以用go/python/java描述数据处理流程。但是后端计算引擎的实现是C++等实现的。一般在google大型项目的实现都是混合语言的，具体需要根据哪一块场景最适合什么，另外还有很多自己的语言。所以比较的其实不是单纯的java和go。



卜 Sapphm

高效的数据处理和高质量的标注是数据分析的前提，在AI战场厮杀的不仅仅是复杂的算法，还要依托于背后的大数据处理能力。看了目录，内容很干货。

话说，这又是一个亲身上阵自己录音的老师，作者本人读出来的文章是有灵魂的~

👍 13 2019-04-15

作者回复: 谢谢鼓励



风之伤

学习这专栏需要什么基础知识

👍 8 2019-04-15

作者回复: 很好的问题。设计时并没有对读者基础作任何假设, 所以碰到任何技术概念, 都会举例解释一下。可能需要些编程基础会学的快一点, 专栏里一些示例代码是Python。如果有哪里觉得不清楚的后面可以再提出来, 我们可以再调整。谢谢提问!



李

深度拷问灵魂中技术的本质, 做一个有技术远见的工程师

👍 4 2019-04-15

作者回复: 是, 李同学, 你理解的很对



听水的湖

又是一个Google大佬, 大佬是南方人吧。带着耳机听的, 专栏用作者本人的声音真是很良心了, 更有代入感。不过感觉有点难度, 希望文章内容有深度的同事能兼顾一下宽度……学渣倒地不起……

👍 4 2019-04-15

作者回复: 哈哈, 确实是南方口音。是会兼顾各方面同学的需求, 不过难度和宽度并不冲突。比如在第二篇里面, 我们分析一个案例, 会看数据量100的时候怎么解决, 1亿又是怎么解决。我希望展现一个问题解决的立体全景。



大王叫我来巡山

很多时候公司淘汰一个人的原因不是因为他年龄大了, 而是他的技术没有随着年龄增长

👍 3 2019-04-16

作者回复: 很多时候是因为那个公司傻逼, 没有意识到程序员的价值在于经验, 下次解决相似的问题知道哪些路可以哪些路不行。年轻干的动只是一小部分。



李海明

听起来老师很安静, 喜欢您的分享。鼓励自己多留言

👍 3 2019-04-15

作者回复: 是的海明同学, 希望多看到你的留言



Mark Lee

什么时候开课呀! 我是小厂的大数据架构师, 但我已经迫不及待听你的灵魂追问了

👍 3 2019-04-15

作者回复: Mark你好, 应该是这周就会上正文。希望后面继续交流



William.Sheng

最喜欢这种既有理论高度又有项目实战的课程了。看完这篇开篇词, 让我这个Java老兵心潮澎湃, 期待老师的课程~

👍 2 2019-04-16

作者回复: 谢谢, 希望让你有收获, 后面继续交流!



高立明

看到老师的专栏很激动，觉得这应该是我后续职业发现的方向，因此果断订阅了。老师您好，我现在是一名dba，想转型大数据架构，您觉得需要掌握哪些知识，大数据知识和应用的体系结构是什么样的，能给出一张思维导图吗？希望和老师一起成长，收获满满！

👍 2 2019-04-16

作者回复: 大数据架构可以先把大去掉，先是数据架构，包含了数据的storage processing serving index learning 存储处理服务索引等等。所以你的dba的经验是非常宝贵的。然后再考虑大规模的问题。从小规模到大规模的思考方式会有很大不同，这些我们专栏里都会提到。



老米屋

老师能不能提供课程里测试数据啊？

👍 2 2019-04-15

作者回复: 可以



JensonYao

看了一半觉得很让人激动，果断订阅了，虽然对相关的知识不是很了解，但是期待和老师一起成长。

👍 2 2019-04-15

作者回复: 是的Jenson，我同意你，知识基础如何并不会阻碍成长。开篇词里也提到14年的时候我也不太会C++。只要每次都提高10%，坚持下来回头看就是很大的进步。



json

请问老师Google T6是什么概念？

👍 2 2019-04-15

作者回复: 相当于阿里P9吧



爱吃彩虹糖的猫~

我想跟着老师的脚步，去拷问我自己那愚钝的灵魂

👍 2 2019-04-15



yiwu

声音好听

👍 2 2019-04-15



hua168

老师，学习这个需要什么知识为提前？

👍 2 2019-04-15

作者回复: 很好的问题，另一个同学也提到了类似问题。我们在内容设计时并没有对读者对知识背景作任何假设，所以即使一些基础的技术概念都会举例解释一下（如果你会了可能会觉得啰嗦）。有一些任何语言的编程经验会看起来快一点，因为有一些示例代码是Python的。但是设计类型的案例，我不觉得有特别的技术要求。希望你后面继续跟踪一下吧，如果有哪些讲的不清

楚，或者解释的过多，后面可以调整内容。谢谢提问！



jiji

作为数据行业从业者表示深深同意开篇词，非常实用的提纲挈领。现在的人工智能如果还是只关注算法那说明应用者太浮躁了，喜欢这份专栏

👍 1 2019-04-16

作者回复: 谢谢你的肯定！让我们一起成长



hua168

老师，问个问题外话:

- 1.现在运维发展方向是什么？
- 2.我现在是小公司运维，三十多岁了，会
网络:自学CCNP
运维:常用软件环境搭建及配置，shell
编程:去年开始学了点spring boot，django，
前端:js，jquery，css，bootstrap，layui
现学编程和大数据还来得及吗？☹

中年危机了，找工作感觉比之前难，如果40岁后的路怎么走？迷茫……

👍 1 2019-04-16

作者回复: 运维不变的方向就是自动化吧，把更多的任务交给软件完成。我觉得编程什么时候开始学都来得及，我在google看到很多白发苍苍的人还在coding，一方面是喜欢，另一方面经验的增长是相比年轻时候拼体力更大的优势。



doze_worm

人工智能，我们就嘲笑说：先人工，后智能还是智障说不定。数据的预处理、清洗、标注，这个成本太夸张了，处处都是人工。。。什么时候能从大数据小功能，过度到数据功能量性对等？有点遥远了，我还是先学学怎么减少人工吧。期待老师的课程。

👍 1 2019-04-16

作者回复: 嗯嗯



专注

这个跟云厂商的大数据有联系么

👍 1 2019-04-16

作者回复: “大数据”作为广告词包含了多重意思，数据存储，数据处理，数据服务，等等。我们这个专栏是侧重于数据的处理。



移动的130☹

其实我有些迷糊，对于大数据这个概念的界定是什么？是数据量太大？还是服务承载能力不足？或是什么？有人说他的数据每天流入以G计算！大数据技术帮他解决了计算问题！那，不用大数据技术就没法解决？还是只能用大数据技术解决？

👍 1 2019-04-16

作者回复: 这些都是大规模的例子。大规模和小规模没有明确界限的，有时候即使小的规模100个数据也会用大规模的方式解决。我们其实在第三篇文章有具体的案例讲解怎样从小规模升级为

大规模。



Geek_d6f9c0

喜欢老师的风格，好的思想跟技术功底是相辅相成的，难分孰轻孰重，互为支撑，互相成就

👍 1 2019-04-16



InfoQ_Albert

“看到5G时代，如何处理实时超大规模物联网数据”时果断订阅，这是5G时代来领真正的挑战，期待老师的讲解。

👍 1 2019-04-16



sunxboy

渴望成长，共同成长

👍 1 2019-04-16



Zach_

老师，大数据是用python多一点 还是jav多一点哇？

👍 1 2019-04-16

作者回复: 一句话回答的话是：语言不重要，任何语言都可以解决大量问题。

长的回答的话：大部分框架都有python 和java的binding比如apache beam, spark。重要的是要掌握数据系统背后的设计思想，然后就会发现换个语言都是一样的。



SpanningWings

我觉得老师的五个问题问得好：这个技术解决哪些痛点，为什么别的技术不能解决，它是如何解决的，是最好的方法吗，不用它我如何来独立解决。回答类似问题在我司别名叫讲清楚，实在很不容易。谢谢。

👍 1 2019-04-16

作者回复: 很高兴看到你也是类似追问



Li221

读了开篇词，觉得作者非常平易近人，鼓励读者互动，希望可以学习到一些硅谷的最新知识，充实自己的工作！

👍 1 2019-04-16



jone

看到是大数据处理的 果断订阅 弱弱的问一句 大佬 学习这个的基础是什么？虽然之前自学过一些大数据的一些组件 不过都是肤浅的，不知道能不能听懂您的技术分享！不过还是决定要学习！

👍 1 2019-04-16

作者回复: 好几个同学都问到了这个问题。我们没有对知识基础作假设，所以即使一些基本技术概念也会有案例解释。后面如果你觉得哪里不清楚，也可以及时提出来，我们调整，谢谢！



渡码

您觉得研究Mapreduce或者hdfs框架有必要吗

👍 1 2019-04-16

作者回复: Mapreduce和hdfs都解决了很多问题。但是后面一篇会提到mapreduce本身的局限性。我看一个技术不会拘泥于“现在还有必要学吗”。任何技术产生都是有原因的,肯定能解决一些问题,肯定也有被更好技术取代的一天,但更重要的是明白技术怎么产生怎么设计的。



Jackjin

我需要一点肯定,老师你认为学习是专一好,还是各个领域都要涉略?

👍 1 2019-04-15

作者回复: 这个没有标准答案,取决于你的目标。比如你想成为一个技术型leader那不得不对系统各组成的领域都有所涉及;比如你想成为一个领域专家,就不得不在一个领域花更多时间。但是深度和广度也并不一定矛盾,有时候能融会贯通,比如Chris Lattner(Swift发明人LLVM发明人)在编程语言和编译器原理上有深厚的积累,到了人工智能领域,他最新在google的工作MLIR能够更方便把同样的模型在不同硬件上计算。所以在我们专栏里,更多强调技术背后的思维,为什么这个技术是这样的。思考了这些,下次看到别的技术,只是另一个相似的思路罢了。正如畅销书principle中文名《原则》里面所说,看多了就会发现一个新技术只是“another one of those”。



小勒.MVP

跟上前沿技术的思维,争取能在工作场景中落地

👍 1 2019-04-15

作者回复: 嗯嗯,后面也可以继续交流



Yang

师者,传道师行业解惑。

👍 1 2019-04-15



茅坑是渣渣

老师,大规模数据处理解决方案上会有银弹嘛?

👍 1 2019-04-15

作者回复: 怎么定义“银弹”?从技术框架上看从mapreduce到beam是巨大的简化和提高。另外对于具体问题的设计方案上,一个方案比一个方案高效很多也是可能的。



零

打卡day01

👍 1 2019-04-15

作者回复: 哈哈,期待看到你的day40



jimmy

这个是目前听到最舒服，最有水平的开篇辞

👍 1 2019-04-15

作者回复: 谢谢认可。包括后面的文章也是，改了好多版，几乎重写了几次。



韩亚强

目前Java程序员，目标大数据处理，需要学的还很多……

👍 1 2019-04-15

作者回复: 这个目标很好啊，你可以参考下我们这里的学习路径。如果后面有问题或困惑，也可以进一步留言讨论。



陈晓

看到spark就立刻订阅了！

👍 1 2019-04-15

作者回复: 对的，陈晓，spark是这里面想要深入展开的一块。希望到时候多交流



leben krieg

我最想要学的就是数据开发成长之路

👍 1 2019-04-15

作者回复: 兴趣很重要！那后面继续多讨论吧



小度

渴望知识的成长，渴望经验的成长，渴望财富的成长。

👍 1 2019-04-15



Hurt

渴望成长👍👍

👍 1 2019-04-15



Jeeno

渴望知识的成长，渴望经验的成长，渴望财富的成长

👍 1 2019-04-15



darrykinger.com

技术跟着需求走，观望中……

👍 1 2019-04-15



任大鹏

是否能现在就开始准备应对 10 年后人类社会的技术挑战，是你拉开与别人差距的重要一站

毫不夸张地讲，读过开篇词之后热血沸腾，果断订阅！

👍 1 2019-04-15

作者回复: 好的大鹏，希望后面继续留言讨论！

 Geek_32772e

实战的时候用啥语言？Scala吗？

👍 2019-04-16

作者回复: 主要是Python。语言不是很重要，关键是把系统的设计思路，和问题的解决方法掌握。

 台风骆骆

正如老师所说的，来到这里学习的人都是渴望技术成长的人，我也是，希望成为坚持学习，坚持地对技术进行灵魂深度提问来成为一个合格的大规模数据处理架构师

👍 2019-04-16

作者回复: 是，希望后面继续交流

 Destroy_

打卡打卡。。刚看完数据分析专栏，继续进修！

👍 2019-04-16

作者回复: 加油！希望与你共同成长！