



下载APP



## 结束语 | 长风破浪会有时，直挂云帆济沧海

2022-01-31 徐文浩

《大数据经典论文解读》

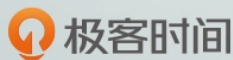
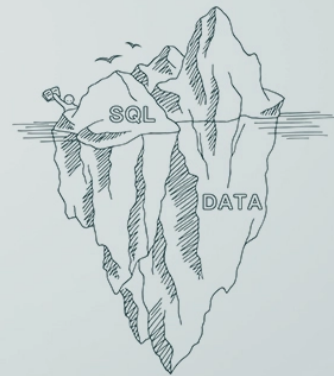
课程介绍 &gt;

**徐文浩**

bothub 创始人

你好，我是徐文浩。

在这门课里，我们一起看过了很多大数据论文。希望这门课的结束，对你来说是打开了一扇新的大门。在未来的日子里，你能够通过论文去追踪计算机领域最新的进展，让你自己迈上一个可以长久发展的职业生涯。

**讲述：徐文浩**

时长 06:41 大小 6.13M



你好，我是徐文浩。

今天正好是除夕夜，在这个特别的、萦绕着浓厚节日气息的日子里，咱们《大数据经典论文解读》这个课程，即将要跟你说声再见了。在此，先祝你新春快乐，新的一年，学习、工作都顺顺利利。

领资料

### 不够完美却也不留遗憾

相比于之前《深入浅出计算机组成原理》课程，这个课程对我来说是一个更大的挑战。☆ 竟，组成原理想要讲的内容相对来说是确定的，而在整个“大数据”领域，决定到底选取哪些论文就已经不容易了。并且我自己也对这个课程有一个小小的期望，希望它不要变成简单地翻译或者重新组织一下论文。

我希望这个课程，一方面能够给你解释清楚，这些论文本身涉及到的系统的基本设计思路，另一方面也能够让你理解，这些论文在当时出现的背景和实际在工程界的需求。之所以这样考虑，也是来自于我一直持有并且反复强调的一个观点——那就是在学习的过程中，**搞清楚“为什么”比“是什么”更重要。**

在课程已经完结的今天来看，这个课程比理想中的“完美形象”可能还要差上不少。其中，固然有不少论文，我觉得是解读得非常清晰，让自己也很满意的。也有一些论文，可能还留有一些遗憾，对于“为什么”可能讲得还不够透彻。我也相信，这些解读中，多少还会有一些错漏，也只能在之后根据同学们的留言和反馈去进行修订，让整个课程尽量做到清晰、准确。

更多的遗憾还是来自于有限的时间和内容，其实还不足以覆盖过去 20 年整个大数据领域的发展。我们讲解了 Spark 这个系统，不过却没有足够的篇幅再来具体讲一讲 Spark Streaming 了。我们讲解了 Megastore 这个在 Bigtable 之上，支持同实体组内事务的 KV 数据库，却也跳过了 Percolator 这个在两者发展过程中，起到了桥梁作用的增量数据处理系统。而在解读 Chubby 这样的系统的时候，也没有给 Paxos 算法留下足够的篇幅，只是给出了一个描述性的介绍。

不过，“**完成比完美更重要**”，更何况这个课程也覆盖了过去 20 年，绝大部分最重要的大数据相关的论文。在经过了半年左右“痛苦”的写作经历之后，它还是以一个完整的姿态和你见面了。这一点，要特别感谢极客时间的各位朋友，和正在坚持学习课程的你，如果没有你们的支持和鼓励，我想我大概早就已经“半途而废”了。相信很多人和我一样，需要一些来自外部的压力和最后期限，才会让我们能创造出一点点东西来。

## 新时代技术人的新机会

回到这个课程的主题，“大数据经典论文解读”其实是一个两年前我就已经有了想法的主题。“大数据”可谓是计算机工程界过去 20 年，发展最迅速的一门“显学”了，乃至今天很多对于互联网公司的“负面”新闻，往往也是关于“大数据杀熟”的话题。而且，整个“大数据”领域也在不断进步，并没有什么晦涩难懂的理论，或者复杂高深的数据公式。

相信你在整个课程的解读过程中，也已经发现了这是一个非常**工程化**的领域。所有的论文、系统设计都是基于实际需要解决的问题，所给出的工程解决方案。只不过，这些方案不是靠一两个版本、一两个月甚至是一两年的过程进化出来的。而是通过十几年时间，通

过不同公司、不同国家的成千上万个工程师，在一个松散的合作方式下逐步迭代出来的。而这十几篇论文，则是这十几年全世界顶尖工程师协作下的精华，可谓是字字珠玑了。

何况，整个大数据领域，其实相当于在十年的时间里，重写了一套分布式环境下的“操作系统”“文件系统”和“数据库”。这也是以另外一种形式，让我们可以重新对计算机科学的底层原理在应用层的作用，有一个更深刻的了解。

我希望这个课程，能够让更多的工程师**直接去学习论文这样最新的一手资料**，而不是能够在技术成熟之后，去读一些教程，调用一些 API。

20 年前，国内最优秀的工程师，往往也只有开发应用系统的机会。而到了现在，我周围就有大量朋友和同学所在的创业公司，在做图数据库、向量数据库、流式数据库这样更有技术挑战的产品了。而这些朋友，往往在过去十多年里，都深度参与了各个互联网公司的大数据系统的建设。代码没有少写，论文也没有少读。正是在这个过程中，他们从青年的学生变成了这个时代的技术中坚。

也很幸运，在这个时代，各类最先进的论文会直接公布在互联网上，最优秀的系统以开源的形式出现在互联网上，而各类视频教程和这样的课程唾手可得。**只要你有学习的意愿，绝对不会缺少学习的材料。**而这个课程，也是一个去学习和了解大数据系统的很好的起点。

当然，这个学习过程不会是一个“舒适”的过程。对你来说，也许挑战来自于论文都是英文的，你阅读起来没有那么轻松；或者是因为这些论文里，大都是抽象的文字讲解，并没有太多你可以直接运行调试的代码。那么，我希望我对于这些论文的详细解读，能够在这上面帮到你。

## 长风破浪会有时，直挂云帆济沧海

不知道此时此刻，在读这篇结束语的你，有没有把课程的每一讲都学习完了。有没有在学完课程之后，也去把每一篇论文都读完了。如果你或多或少去读了一些论文的原文，那么我想，我过去几个月“痛苦”的努力就没有白费。

事实上，在整个课程写作的过程中，对我自己来说也是一个“温故而知新”的过程。虽然这些论文在过去的十多年里，我大都仔细读过，但是之前的阅读也更多只是知道论文讲了

些什么、我们在实践中可以如何运用。而在这个写作的过程中，却也给我了一个机会，去重新思考这些论文、系统的由来和相互之间的联系。

对你来说，通过课程知道我给你讲解的内容是一回事儿，回头自己去研读论文，想明白里面的每一个点，是更有挑战的新的一步。相信你和我一样，会去迎接这个挑战的。

当然，课程到这里，无论对你还是我来说，**既是一个终点更是一个起点**。之后，我还有大量的课程留言需要花时间去回复，对你来说，未来还有更多的新论文、新系统等你去研究、学习，乃至发表。当然，这一定不是一个简单而是一个困难的过程。

最后，借用李白在《行路难》里的几句诗词，和你互相勉励。

行路难！行路难！多岐路，今安在？

长风破浪会有时，直挂云帆济沧海。

好了，最后的最后，我还给你准备了一份毕业问卷，希望你能花两三分钟填写一下，非常期待能听到你对这门课程的反馈。



徐文浩

bothub 创始人

感谢一起走过的这段时间，非常想听听你对我和这门课程的反馈与建议。在 2 月 20 日前提交问卷，将有机会获得



极客原创 | VS Code 快捷口令  
超大鼠标垫 价值 **¥99**

或



极客时间课程阅码  
价值 **¥99**

填写问卷 

我是徐文浩，感谢你一直以来的学习和陪伴，咱们下次再会啦。

分享给需要的人，Ta购买本课程，你将得 20 元

生成海报并分享

赞 2 提建议

© 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。

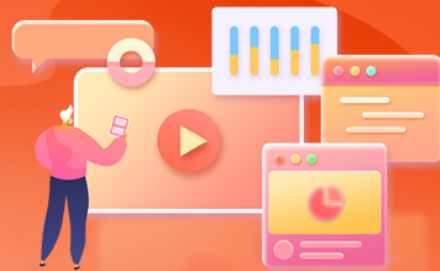
上一篇 40 | 十年一梦，一起来看Facebook的数据仓库变迁（二）

## 更多学习推荐

# 190 道大数据高频面试真题

涵盖 11 个核心技术栈 + 4 套大厂真题

免费领取



## 精选留言 (6)

写留言



陈冰清

2022-02-03

好喜欢徐老师的课哦。

让我这个不懂计算机的人 一直从去年听到现在，因为每篇的讲解逻辑很清楚，我能复述出来是怎么一回事（虽然里面的专业术语还有代码看不懂..但这并不影响）。

还期望徐老师出其他课程，比如 讲算法的、还有计算机入门的、以及时下流行的黑科技解

读等等。可能比较适合我这样的入门选手。或者其他像编译反编译原理及其应用、异构...  
展开 ∨



**王子凡kerry**

2022-01-31

老师辛苦了，很棒的专栏



**Alex**

2022-02-09

感谢老师的辛苦付出



**在路上**

2022-02-09

徐老师辛苦了，非常棒的专栏，不仅帮助我进入了大数据领域，也让我掌握了学习技术的新方法。



**CRT**

2022-02-05

辛苦老师，很好的专栏



**尚春**

2022-02-05

虽然跳过了一些对我一个入门的人来说比较晦涩的章节，不过这段时间的学习让我对整个大数据脉络有了更深的理解。老师辛苦了，我还会二刷三刷，逐步加深理解。

