



下载APP

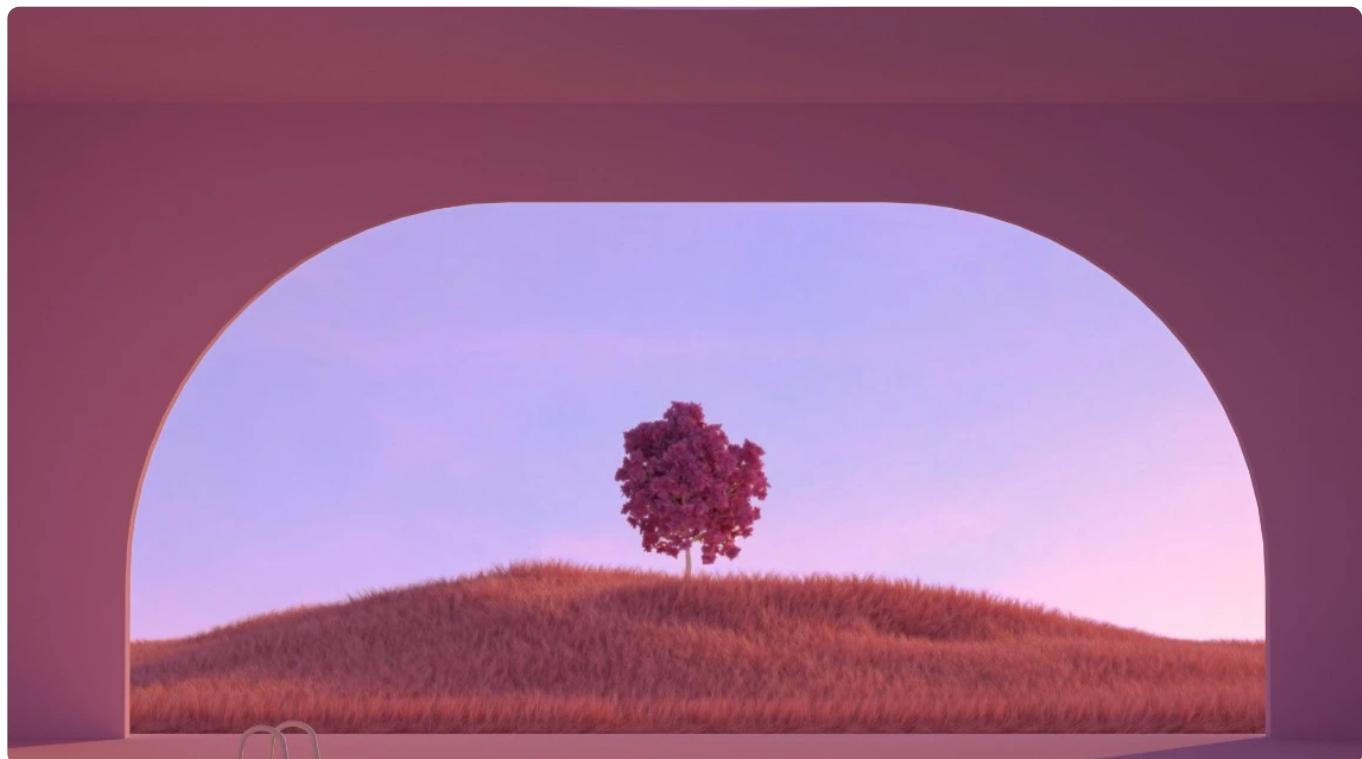


# 用户故事 | 陈煌：唯有自强不息，方能屹立不倒

2022-01-05 陈煌

《大数据经典论文解读》

[课程介绍 >](#)



讲述：王惠

时长 07:57 大小 7.30M



你好，我叫陈煌，目前在南京一家通信公司从事大数据开发工作，主要负责公司自研分布式数据库的开发与运维工作。

我从《大数据经典论文解读》这门课程刚推出的时候，就开始跟随徐老师脚步，学习各种大数据相关的知识。在这个过程中，我收获了很多新的认知，也对各种数据库的设计思路与框架有了一个较为深刻的理解。所以，非常荣幸能在这里，与你分享我学习这门课的心得与体会、学习思路与方法。



领资料



## 为什么学习大数据经典论文解读？



其实早在老师开设这门课程之前，我在攻读研究生的时候，就已经阅读了一些大数据相关的经典论文。

一开始我有想去阅读大数据开发相关论文的念头，应该是在 2018 年末，当时由于自己考上了计算机相关的专业硕士，又对 Google 的三驾马车充满了好奇，于是在好奇心的驱使下，开始尝试去读大数据相关的经典论文。但由于当时自己的功力不够深厚，很多的论文都只能浅尝辄止。

在 2019 年末的时候，我得到了目前所在公司的大数据开发实习岗位的工作机会，算是正式踏入了大数据开发的领域。当时我们小组正好在为公司自研一款分布式数据库，让我负责数据库的持久化文件合并功能的开发。对于我这样一个小白来说，这是一个很大的挑战，无论是合并功能一开始的底层架构、设计思路，还是之后针对功能优化所需要的新算法，几乎都需要我去认真研读论文，学习前人留下的宝贵财富。

所以，除了我个人对大数据论文的关注之外，在工作上的各种变化和挑战，也驱使我想要快速提升阅读大数据论文的能力，并能通过阅读论文中所涉及数据库的设计思路，帮助指导和解决我在优化合并功能上遇到的各种难题。

## 我是如何学习大数据经典论文解读的？

那么，我是怎么学习这门课程的呢？

其实在一开始阅读大数据论文的时候，我也走了一些弯路，无论是在论文的选择上，还是确定论文阅读的切入点，我都花费了大量的时间，但最后，却没达到我理想的效果。所以这里，徐文浩老师的《大数据经典论文解读》这门课，可以说对我产生了巨大的帮助。

### 记笔记

首先，我从买完徐老师的课程之后，每学习一节课，就会给这节课单独建一页笔记。我选择了有道云笔记来记录，如下图所示：

The screenshot shows the Geekbang Time application interface. On the left, there's a sidebar with navigation links like '升级' (Upgrade), '同步' (Sync), '+新建' (New Note), '最新' (Latest), '加星' (Star), '与我分享' (Share with me), '我的文件夹' (My Folders), 'Big Data', 'MIT-分布式...', '工作备忘', '极客时间-大... (selected)', '技术考核', 'Computer Tec...', 'Explore', 'Java', 'Machine Lear...', 'Master', 'Training', '来自手机', '我的资源', and '回收站'. Below this is a note list with titles such as '19 Megastore (一) : 全国各地都...', '18 Spark: 别忘了内存比硬盘快多少 过去几年里，无论是 Hive 這種基於 MapReduce 的系統，還是 Dremel 這樣...', '17 从Dremel到Parquet (二) : 他... 在上节课里，我们看到了 Dremel 这个系统的数据...', '16 从Dremel到Parquet (一) : 深... 在解读 Hive 论文的过程 中，我们看到 Hive 已经...', '15 Hive: 来来去去的DSL (Domai... 过去几篇论文的解读，相信现在你已经深入掌握好了大数据系统的基本知识。而在 Google 的这些论文发表之后，整个工业界也行动起来了。很快，我们就有了开源的 GFS 和 MapReduce 的实现 Hadoop，以及 Bigtable 的实现 HBase，这些系统，的确帮助我们解决了很多海量数据处理的问题。并且这些系统设计得还算易用，作为工程师，我们基本不太需要对分布式系统本身有深入地了解，就能够使用它们。不过，这些系统都还很原始和粗糙，随便干点什么都很麻烦。所以自然而然地，工程师们就会通过封装和抽象，来提供更好用的系统。这次的系统，不再是来自于 Google，而是 Facebook 了，它的名字叫做 Hive。Facebook 在 2009 年发表了 Hive 的论文《Hive: a warehousing solution over a map-reduce framework》，并把这个系统开源。而在 2010 年，Facebook 又把这篇论文丰富了一下，作为...', '14 分布式版Chubby (三) : 移... 过去的讲解里，我们都在尝试做一些事情，就是在...', '13 分布式版Chubby (二) : 众所... 总共 21 篇'.

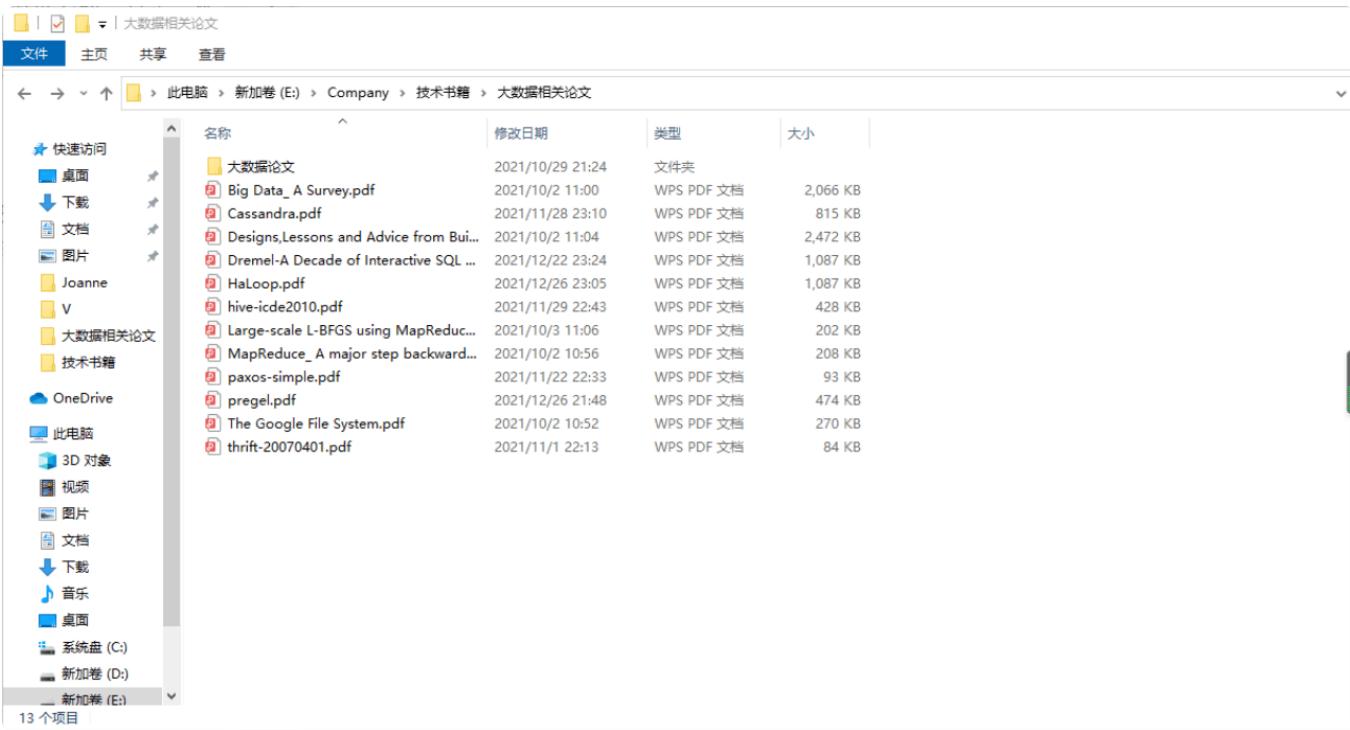
The main content area displays a detailed note titled '15 Hive: 来来去去的DSL (Domai...' with a PDF icon and a file size of '427.72KB'. The note content discusses the evolution of Hive from its initial design to its eventual implementation in Hadoop and its impact on the industry. It also highlights the challenges of distributed systems and the role of Facebook in advancing the technology.

我之所以这样做，是因为我发现，在边学习边做笔记的过程中，会进一步加深我对这节课的理解，而且我所做的笔记，可以让我在任何想要温习自己所学知识的时候，都可以拿出来回顾。最重要的一点是，我可以将徐老师的课程进行拆分，对我认为比较重要的地方进行记录，并做出重点标注解读。后来我发现，极客时间的做笔记功能也是十分强大，可以根据个人喜好选择做笔记的方式。

## 关注“推荐阅读”

徐文浩老师在每一节课的最后，都会给我们推荐几篇与课程相关的论文或者资料。我们一定要关注这些论文资料，徐老师并不会平白无故地给我们推荐不适合的论文，增加我们的负担。这些论文能够帮助我们对课程的核心思想有更加深刻的理解，并让学有余力的同学能够拓展思考。

所以，哪怕我们在眼下没有时间去阅读，也可以把这些论文下载下来，放在一个单独的文件夹中，供日后查阅。



## 关注“思考题”

人是懒惰的动物，哪怕是爱学习的我们也是一样的。我们习惯于“投喂式学习”，习惯于让老师帮我们把所需要的知识全部都总结好，我相信有很多人在学习课程的过程中，遇到思考题就直接跳过，懒于思考。但是我认为，徐老师设置思考题的真正意义，就是在于检验我们是否真的读懂了课程，是否真的勤于思考。如果只是囫囵吞枣般地学习课程，那效率将大大降低。

所以，我很鼓励你去思考，哪怕有很多思考题你并不能给出答案，但是这个思考的过程非常重要，一旦养成了思考的习惯，你会发现学习每节课的内容，都能收获到更多。

## 关注“留言”

留言区往往是大神云集的地方，刚刚我们提到，也许会有很多思考题你连思考的方向都没有，这时候留言区就派上了很大的用场。各路大师各显神通，从他们各自的角度出发，对思考题作出了解答，在留言区互动不仅可以对我们有所启发，而且可以让我们交到更多的朋友。徐文浩老师也对不少留言进行了回复，所以留言区绝对是每节课被大家“遗忘”的亮点。

## 我是如何运用这门课所教授的知识的？

读万卷书不如行万里路，对大数据论文的学习也是一样的，我们不能只阅读论文，还需要经常去做尝试。只有不断地试错，才能得到成长。就比如说：

对于小白同学来说，在学习这门课程的过程当中，首先要做的就是自己动手去搭建一个分布式数据库，从实践中获得的成就感，会不断促使你去学习新的理论知识。

对于像我这样有一些大数据开发经验的同学来说，在我工作的过程当中如果遇到了相关问题，我会有意识地让自己去回想课程里，是否有提供相关的解决思路。

比如，我公司自研的数据库目前暂不支持并发写索引，最近我写的一个入库工具在入库的过程中遇到了这个问题，但我的工具是单线程的。这时候，我便联想到了徐老师在《从 Dremel 到 Parquet（二）：他山之石的 MPP 数据库》中提到的 MapReduce 的推测执行功能，在一个任务运行比预期慢的时候，它会尽量检测，并启动另一个相同的任务作为备份。因此我立刻在配置项中关闭了该功能，之后便再也没有遇到并发写索引问题。

因此，我非常推荐去践行 “Talk is cheap, show me the code.” 的原则，不要只会纸上谈兵，一定要创造条件去实践。

## 写在最后

最后，很荣幸能和你一起阅读大数据经典论文，一起学习徐文浩老师的这门课程。

我个人认为，能够订阅徐文浩老师这门课程的同学，大部分都是和我一样有志于在大数据开发领域有所建树的程序员，能够和大家共同进步、共同见证彼此的成长，是一件无比快乐的事情。

当然，阅读论文的过程可以说是非常痛苦的，我时常会绕进自己思维的“死胡同”，但是每当我解决了困扰自己已久的难题时，都会有一种“山重水复疑无路，柳暗花明又一村”的快感。

世上无难事，只要肯登攀。祝我们都能跨过大数据领域的一座又一座大山，站在山顶一览众山小。也期待各位同学的留言，我们一起交流和讨论，一起进步！

分享给需要的人，Ta订阅后你可得 20 元现金奖励

[生成海报并分享](#)[赞 1](#)[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 加餐3 | 我该使用什么样的大数据系统？](#)

[下一篇 32 | Raft（一）：不会背叛的信使](#)

## 更多课程推荐

### 陈天 · Rust 编程第一课

实战驱动，快速上手 Rust

陈天

Tubi TV 研发副总裁



涨价倒计时



今日订阅 **¥89**，1月12日涨价至**¥199**

## 精选留言 (1)

[写留言](#)

在路上

2022-01-05

优秀的陈煌同学，感谢你的分享。学以致用，举一反三，向你学习。

