

# 设计数据密集型应用（影印版）



[设计数据密集型应用（影印版）\\_下载链接1](#)

著者:Martin Kleppmann

出版者:东南大学出版社

出版时间:2017-10-1

装帧:平装

isbn:9787564173852

书中包含以下内容:

深入分析你已经在使用的系统，并学习如何更高效地使用和运维这些系统

通过识别不同工具的优缺点，作出更明智的决策

了解一致性、可伸缩性、容错性和复杂度之间的权衡

理解分布式系统研究，这些研究是现代数据库构建的基石

走到一些主流在线服务的幕后，学习它们的架构

作者介绍:

Martin Kleppmann，是英国剑桥大学的一名分布式系统研究员。在此之前他曾是软件工程师和企业家，在 LinkedIn 和 Rapportive 工作过，从事大规模数据基础设施相关的工作。Martin 经常在大会做演讲，写博客，也是开源贡献者。

目录:

[设计数据密集型应用（影印版）\\_下载链接1](#)

标签

分布式

计算机

软件架构

数据库

大数据

计算机科学

数据分析

Programming

## 评论

梳理全面，见解独到

---

非常好的综述，把数据库、分布式系统、消息中间件、存储系统等都放在一起，比较说明各自的特点和相通之处，非常有启发。

---

条理、逻辑、语言都极度舒服，尤其是每一章都引用了非常多的参考文献，这些参考文献大多数都值得一读。算是分布式系统工程概述性方面写得最好的书籍！

---

真是一本好书，分析问题，解决问题，理论联系实际。有时间要二刷。PS：最后一章最后一节比一般技术书籍真是不知道要高到哪里去。 2019.12.2  
二刷完毕，感觉依然那么棒！

---

仅有一些存储引擎背景，粗略了解分布式，工作经验也没有点CRUD，大数据一套即没用过，也没做过。不过依然觉得书是好书，回答是什么，为什么，能做什么，可以怎么做的书很多，本书更像是综述，讲述历史脉络，且语言朴实生动，浅显易懂，可以做一个很好的开始，然后用参考文献继续深入，书里最后的数据驱动为监视驱动，也是发人深思的原则问题，机缘到时，再来深入阅读

---

没看懂

---

屠龙之技

---

很体系化的一本技术书籍，后面的论文可深入研究，五星不准掉，略微比预期差些。

---

-----  
为了复习看的，有种高屋建瓴的感觉！

-----  
经典需要多读几遍

-----  
主要啃英文了

-----  
高屋建瓴，常读常新

-----  
神书。看得好慢。。。最近几年唯一完整看完的书，还得刷第二遍。。。。

-----  
书看的差不多了，已近尾声。如果非要评价下这本书，只想用四个字“相见恨晚”，一本足以让你建立领域框架梗概，了解系统方法论的书籍，如果在计算机科学领域推荐另一本能够匹配的书，我会推荐性能之巅。

-----  
经典！

-----  
断断续续总算读完了，很赞；这本书就像一本地图，把数据相关的知识全都串起来，让人能站在全局的角度看问题；比起数据库开发者，更适合数据库使用者来读，对正确的设计系统架构很有帮助

-----  
可以读得停不下来的技术书，可谓精彩纷呈，引人入胜

-----

影响我计算机最大的书

-----  
(2018-9-4左右)辛苦读完，但感觉很多细节都不懂。读的时候为了图快，很多细节/小问题，我也没有课后自己去搜索Stack Overflow/去问别人，就这样过了。但还是收获很多。 - 工作中要是遇到相关的问题/展开，我可能也会记得对应书中哪一个章节，然后过来翻阅(拿来当说明书/字典用)。 - 估计一年内会读第二次，会读得更细，把这些遗憾补上

-----  
[设计数据密集型应用（影印版）\\_下载链接1](#)

书评

大概十一放假开始读这本书，中间经历了加班，双十一值班，自己病假，老婆生病请病假，娃生病请病假，断断续续到现在终于算是从头到尾看了一遍，实在是觉得不容易。这本书的作者是少有的从工业界干到学术界的牛人，知识面广得惊人，也善于举一反三，知识之间互相关联，比如有...

-----  
每一个数据工程师和产品经理都应该把最后一节，Doing the right thing，读十遍。为了说服大家为什么值得读十遍，我摘抄几句： automated systems can systematically and arbitrarily exclude a person from participating in society without any proof of guilt, and with l...

-----  
<https://github.com/Vonng/db>  
=====

-----  
  
-----  
  
-----

-----

-----

Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relati...

-----

以前看过一些分布式系统相关的资料，比如《分布系统原理---刘杰》，还有一些关于HDFS、MapReduce、Google  
三大经典论文之类的，也去官方看过一些开源项目(Hadoop、Zookeeper)的官方文档，总觉得分布式是很高深的理论知识。看到这本书后，发现它真的是深入浅出，对分布式原理的...

-----

1.事务及隔离级别 1.1.Read Committed 定义  
一个事务只能看到其它事务已经提交的修改，不能看到其它事务进行中产生的修改。  
实现方法  
对任一事务修改的数据，在事务提交前均同时记录新值和旧值。其它事务读到此数据时，使用旧值;本事务读取时，使用新值。 一致性缺陷：不可重复...

-----

1.数据模型  
从数据库的数据模型上，可以将数据库分为关系型、文档型以及图型数据库等。这三类数据库在形式和使用的场景上各有不同。 1.1.关系型  
关系型模型基于集合论中的关系理论，将数据视为Tuple的集合。关系型在1970年被提出，人们一开始质疑这类技术是否能高效地实现，但...

-----

订正Replication一章的错误，包括consistency prefix reads和“happens-before”的定义。以下均为个人意见，可能有错。 1. Consistency Prefix Reads The book definition of consistent prefix reads is not clear. In one place the book defines consistent prefix reads as...

-----

-----  
从高层视角系统的介绍了数据在存储和读取中的方方面面，非常有助于建立自己的知识体系。书中很多内容都是平时会接触到的一些问题，但是在实际工作中未必能够把这些零碎的东西串起来，形成体系框架。这本书可以让你整理你自己的知识，让他们连接在一起，变成一个树结构，后面你...

-----  
本书开头提到“当今很多新型应用都属于数据密集型（data-intensive）而不是计算密集型（compute-intensive）”  
当今机器学习越来越普及的情况下其实用户应用后面基础件层的compute-intensive应用越来越多了。“很可惜，让鄙人日常头秃都是compute-intensive的，啥时候有一本De...

-----  
这本书从数据的产生、传输、存储、检索、更新等方方面面，介绍了一个后台系统使用到的种种存储技术，包括了数据库、缓存、检索引擎、批处理、流处理。对于每一种技术，作者介绍了它们的基本原理、使用场景、优势和不足，并附带了大量的扩展阅读书目。对于后台开发、运维人员，...

-----  
Martin Kleppmann  
不仅是个牛逼的程序员，更是一个极富社会责任和人文关怀的牛逼程序员。而这是更难能可贵的。习武之人讲究“习武先修德”。Martin Kleppmann亦是如此。他用前十一章教会我们如何处理海量数据，用最后一章告诉我们如何正确使用数据。要保护用户隐私、要对自己...

-----  
作者的态度与高度于书的结尾再见一斑。“As software and data are having such a large impact on the world, we engineers must remember that we carry a responsibility to work toward the kind of world that we want to live in: a world that treats people with huma...

-----  
版权归作者所有，任何形式转载请联系作者。作者：荒城梦（来自豆瓣）  
来源：<https://www.douban.com/note/725242700/>  
陆陆续续有几个月过去了，终于把这本“鸿篇巨制”读完了。本书在计算机类著作里并不算特别厚，说鸿篇巨制是因为随着越往后读越发觉得此书内容之广度与深度已...

-----  
摘要，后面再读补总结。字数还说不能少于140字[捂脸哭]简直了。  
书中比较详细描述了数据库，从数据库衍生派生数据，批处理，流处理等衍生。主要是问题-思路-问题的方式，但其问题角度考虑极端，分布式也正是这些极端场景的发源地

。对自己是一个比较全面的分布式认知，对于很多...

-----  
[设计数据密集型应用（影印版）\\_下载链接1](#)