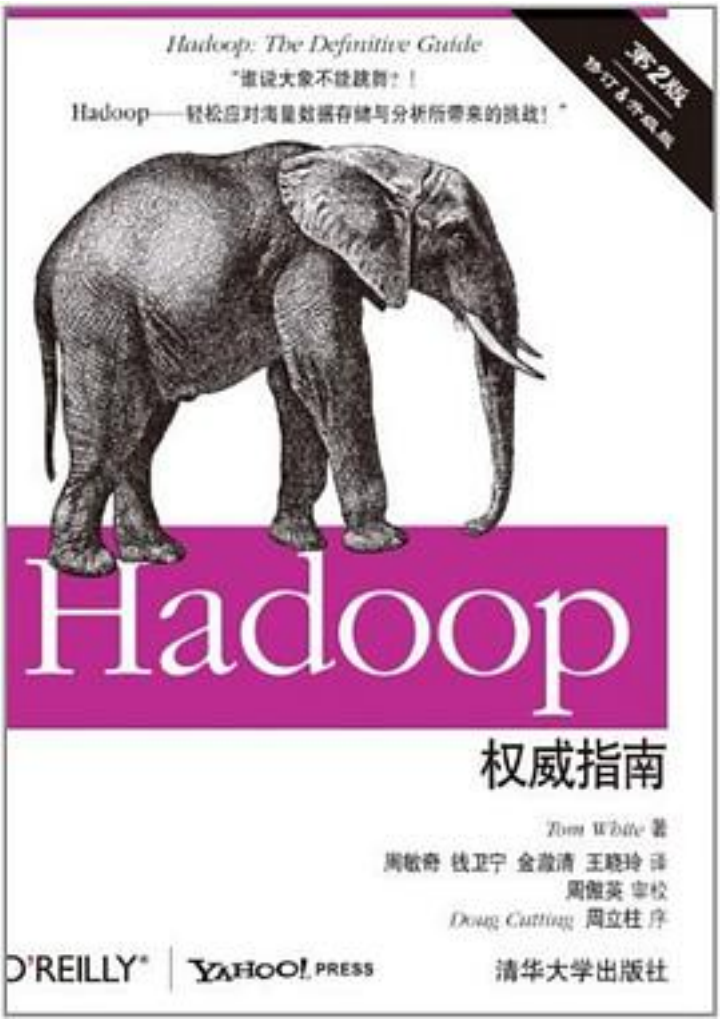
Hadoop权威指南（第2版）



Hadoop权威指南（第2版）_下载链接1_

著者:Tom White

出版者:清华大学出版社

出版时间:2011-7

装帧:平装

isbn:9787302257585

《Hadoop权威指南(第2版)(修订・升级版)》从Hadoop的缘起开始，由浅入深，结合理

论和实践，全方位地介绍Hadoop这一高性能处理海量数据集的理想工具。全书共16章，3个附录，涉及的主题包括：Haddoop简介；MapReduce简介；Hadoop分布式文件系统；Hadoop的I/O、MapReduce应用程序开发；MapReduce的工作机制；MapReduce的类型和格式；MapReduce的特性；如何构建Hadoop集群，如何管理Hadoop；Pig简介；Hbase简介；Hive简介；ZooKeeper简介；开源工具Sqoop，最后还提供了丰富的案例分析。

《Hadoop权威指南(第2版)(修订·升级版)》是Hadoop权威参考，程序员可从中探索如何分析海量数据集，管理员可以从中了解如何安装与运行Hadoop集群。

作者介绍:

Tom White从2007年以来，一直担任Apache Hadoop项目负责人。他是Apache软件基金会的成员之一，同时也是Cloudera的一名工程师。Tom为oreully.com、java.net和IBM的developerWorks写过大量文章，并经常在很多行业大会上发表演讲。

目录: 第1章 初识Hadoop
数据！数据！
数据存储与分析
与其他系统相比
关系型数据库管理系统
网格计算
志愿计算
1.3.4 Hadoop 发展简史
Apache Hadoop和Hadoop生态圈
第2章 关于MapReduce
一个气象数据集
数据的格式
使用Unix工具进行数据分析
使用Hadoop分析数据
map阶段和reduce阶段
横向扩展
合并函数
运行一个分布式的MapReduce作业
Hadoop的Streaming
Ruby版本
Python版本
Hadoop Pipes
编译运行
第3章 Hadoop分布式文件系统
HDFS的设计
HDFS的概念
数据块
namenode和datanode
命令行接口
基本文件系统操作
Hadoop文件系统
接口
Java接口
从Hadoop URL中读取数据
通过FileSystem API读取数据

Hadoop权威指南（第2版）_下载链接1_

标签

hadoop

分布式

MapReduce

云计算

大数据

计算机

O'Reilly

编程

评论

hadoop真牛逼

很快地读完了，没特别了解具体的语法，对于一个宏观上的掌握以及技术选型来说足够了~没有学过Hadoop的人值得一读。

如果一本书是由好几个人合伙翻译的，那么译者通常会漏掉Google，请务必做好精分的准备。

好书烂翻译

感觉这本书字体比较小，印刷的间距比较大，内容上比第1版没什么大的更新——那个SQL导入到Hadoop的工具倒是很有意思

简单读过

中规中矩吧，理论科普和实操手册，仅此

暂时不深究了。

科普，并非深入

翻译质量实在是不敢恭维…

太晦涩了，需要看多遍。。

只有看完之后以后才可能知道有用的部分，第一遍不要求看得多深入，只求知道有用的特性，在以后需要时能回想起来，这也许是阅读这样的工具书的有效方法；除去中文版的翻译问题，整本书还不错；鉴于内容过于啰嗦的问题，建议新手先看《Hadoop实战》，等有了一定的经验之后再看这本书，这样可以保证不陷入繁琐的细节又能增加涉猎。

知道翻译的不好，也知道可能会看起来很痛苦，还是买了，看英文实在太慢。
没看完。

比第一版增加的HIVE和Sqoop

工具书，粗读一遍，不求甚解

暂时还用不上，细看了前10章以及第14章的Zookeeper。

当手册用还OK，不适合入门

书很权威，这本翻译也很好！译本的页数与原本的一直，这对于对照阅读很有好处！囫囵吞枣的看了一遍！大概算是了解了hadoop以及相关项目是干嘛的吧！
完全分布式模式还没搭建起来！还要继续努力啊！

书的内容不错，学习Hadoop的经典，但是翻译的太粗糙了！

O'REILLy的书还是很不错的

Hadoop权威指南（第2版）_下载链接1_

书评

买了第一版，时间太紧，没来得及看，后来出了个号称修订升级的第二版，毫不犹豫又买了，后来听说第二版比第一版翻译得好，心中窃喜，再后来看了第二版，我震惊了，我TM就是一傻子，放着好好的英文版不看，赶什么时髦买中文版呢。在这个神奇的国度，牛奶里放的是三聚氰胺，火腿...

其实也不算全部读完了，读它主要是为了技术选型，考虑升级持久层架构、提高系统可

扩展性，仔细研读了前几章，对Hadoop、MapReduce、HDFS的模型、机制、使用场景有了一定了解。后面几章及其生态圈内的其他项目抱着了解的心态简单浏览了一下。整体感觉还行，至少从我看过的章节来...

中文版412页：
所以理论上，任何东西都可以表示成二进制形式，然后转化成为长整型的字符串或直接对数据结构进行序列化，来作为键值。 原文460页： ..., so theoretically anything can serve as row key, from strings to binary representations of long or even serialized ...

-- china-pub 赠书活动 -- http://www.douban.com/group/topic/20965935/
一直比较忙，整本书还没读完，只是粗略翻了个大概，其中有两三章细读了一遍。先做个大体评价吧，有时间全部细读后再评论。
从书的内容上来讲，大致上与网上该书的内容介绍一致。简单点概括：这本书对...

参加豆瓣China-pub抽奖，比较幸运的得到这本Hadoop权威指南中文第二版，拿来与第一版相比，发现新加入了Hive和Sqoop章节，译文质量也提高了不少，并且保留了英文索引。
这本书对Hadoop的介绍还算全面，有实践冲动的朋友基本可以拿着书、配合Google百度马上实现梦想。个人感觉 “...

看了几章中文版的，各种错误，太低级，实在是看不下去了。 建议还是看原版吧。
译者们的脸皮可真厚，英文译不明白也就罢了，中文都组织的不通顺，好意思吗！！
什么叫 “但是，.....，但是”啊，“但是体”啊。

很好的Hadoop教程，比Apache和Yahoo
!网页版guide详细很多，很多想不明白的Hadoop实现细节都可以在这本书里找到。

详见：http://www.cnblogs.com/aprilrain/archive/2013/03/07/2947664.html

---

很多地方翻译的不行，需要对照英文看才能明白。。。不过对于快速学习，仍然是不错的选择。建议译者看看每部分内容的重要性，不重要的瞎翻翻就算了，重要的部分还是好好花点功夫，不要本末倒置了。比如第三章的数据流部分，这么经典的地方居然被翻译烂的一塌糊涂。不知道译者会...

书中没有透露太多实现架构方面的细节，更多的是从使用者的角度上介绍了Hadoop的各种知识，包括MapReduce, HDFS, Hive, Pig, HBase, ZooKeeper。几乎涉及了Hadoop的所有关于使用方面的知识，包括安装和使用。
你甚至可以直接在自己的电脑上装上一个Hadoop，对着书中的例子实际演...

是我遇到过的翻译最烂的一本书，在译者的“妙语连珠”里折腾了半个钟头就再也没兴趣了。略举几例如下： P.6 任然 -> 仍然 P.21
输入键（为什么不像后面那样有个“的”？），输入的值，输出的键…… P. 27
“计数器”(Counter)，译文附原文；"Context Object"(上下文对象)，原...

专门登录来评论的，翻译也太烂了吧，真的真的建议强烈英语阅读能力好的人去读原版书，不要花冤枉钱在这上面，除了文字错误外，里边的图居然也有错，就比如260页的图最后两个年份应该是1901结果这里竟然是1900，我是真滴服了，一本神书被翻译成这样，作者得气死。zsbd zsbd zsbd...

你的履历添了一笔<hadoop权威指南>译者,但是你不配 这是我见过的最不用心的翻译, 字里行间行文不通顺, 请别勉强自己,map reduce shuffle机制都没翻译的好
虽然原作者写作功底也实在是一般 第 1 2 5 6 7 这几章 翻译的实在是太烂了
请不要呐Google翻译糊弄人阿 误人子弟 ...

首先，翻译太差，很多句子就是瞎翻，根本不通顺，很多时候你要停下来断句，慢慢去理解。
然后，这本书是很多人去翻译的，很多人连代码都不懂，曾经一段代码看到我蒙圈，去看了一下源代码，好家伙，四行有五个错误。另外，从代码瞎缩进也可以看出这是群没写过代码的人翻的，而且...

Cobub Razor APP数据统计分析工具官网上有篇文章是讲Hadoop Yarn调度器的选择和使用的，我觉得写的挺好的，推荐http://www.cobub.com/the-selection-and-use-of-hadoop-yarn-scheduler/

Hadoop权威指南（第2版）_下载链接1_