

<<Hadoop实战 (第2版) >>

图书基本信息

书名：<<Hadoop实战 (第2版) >>

13位ISBN编号：9787111395836

10位ISBN编号：7111395832

出版时间：2012-11

出版时间：机械工业出版社华章公司

作者：陆嘉恒

页数：498

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<Hadoop实战 (第2版)>>

前言

为什么写这本书计算技术已经改变了我们的工作、学习和生活。

分布式的云计算技术是当下IT领域最热门的话题之一，它通过整合资源，为降低成本和能源消耗提供了一种简化、集中的计算平台。

这种低成本、高扩展、高性能的特点促使其迅速发展，遍地开发，悄然改变着整个行业的面貌。

社会各界对云计算的广泛研究和应用无疑证明了这一点：在学术界，政府和很多高校十分重视对云计算技术的研究和投入；在产业界，各大IT公司也在研究和开发相关的云计算产品上投入了大量的资源。

这些研究和应用推动与云计算相关的新兴技术和产品不断涌现，传统的信息服务产品向云计算模式转型。

Hadoop作为Apache基金会的开源项目，是云计算研究和应用最具代表性的产品。

Hadoop分布式框架为开发者提供了一个分布式系统的基础架构，用户可以在不了解分布式系统底层细节的情况下开发分布式的应用，充分利用由Hadoop统一起来的集群存储资源、网络资源和计算资源，实现基于海量数据的高速运算和存储。

在编写本书第一版时，鉴于Hadoop技术本身和应用环境较为复杂，入门和实践难度较大，而关于Hadoop的参考资料又非常少，笔者根据自己的实际研究和经历，理论与实践并重，从基础出发，为读者全面呈现了Hadoop的相关知识，旨在为Hadoop学习者提供一本工具书。

但是时至今日，Hadoop的版本已从本书第一版介绍的0.20升级至正式版1.0，读者的需求也从入门发展到更加深入地了解Hadoop的实现细节，了解Hadoop的更新和发展的趋势，了解Hadoop在企业中的应用。

虽然本书第一版受到广大Hadoop学习者的欢迎，但是为了保持对最新版Hadoop的支持，进一步满足读者的需求，继续推动Hadoop技术在国内的普及和发展，笔者不惜时间和精力，搜集资料，亲自实践，编写了本书第二版。

第2版与第1版的区别基于Hadoop 1.0版本和相关项目的最新版，本书在第1版的基础上进行了更新和调整：每章都增加了新内容（如第1章增加了与Hadoop安全相关的知识，第2增加了在Mac OS X系统上安装Hadoop的介绍，第9章增加了WebHDFS等）；部分章节深入剖析了Hadoop源码；增加了对Hadoop接口及实践方面的介绍（附录C和附录D）；增加了对下一代MapReduce的介绍（第8章）；将企业应用介绍移到本书最后并更新了内容（第19章）；增加了对Hadoop安装和代码执行的集中介绍（附录B）。

本书面向的读者在编写本书时，笔者力图使不同背景、职业和层次的读者都能从这本书中获益。

如果你是专业技术人员，本书将带领你深入云计算的世界，全面掌握Hadoop及其相关技术细节，帮助你使用Hadoop技术解决当前面临的问题。

如果你是系统架构人员，本书将成为你搭建Hadoop集群、管理集群，并迅速定位和解决问题的工具书。

如果你是高等院校计算机及相关专业的学生，本书将为你在课堂之外了解最新的IT技术打开了一扇窗户，帮助你拓宽视野，完善知识结构，为迎接未来的挑战做好知识储备。

在学习本书之前，大家应该具有如下的基础：要有一定的分布式系统的基础知识，对文件系统的基本操作有一定的了解。

要有一定的Linux操作系统的基础知识。

有较好的编程基础和阅读代码的能力，尤其是要能够熟练使用Java语言。

对数据库、数据仓库、系统监控，以及网络爬虫等知识最好也能有一些了解。

如何阅读本书从整体内容上讲，本书包括19章和4个附录。

前10章、第18章、第19章和4个附录主要介绍了Hadoop背景知识、Hadoop集群安装和代码执行、MapReduce机制及编程知识、HDFS实现细节及管理知识、Hadoop应用。

第11章至第17章结合最新版本详细介绍了与Hadoop相关的其他项目，分别为Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa，以备读者扩展知识面之用。

<<Hadoop实战 (第2版) >>

在阅读本书时，笔者建议大家先系统地学习Hadoop部分的理论知识（第1章、第3章、第6章至第10章），这样可对Hadoop的核心内容和实现机制有一个很好的理解。

在此基础上，读者可进一步学习Hadoop部分的实践知识（第2章、第4章、第5章、第18章、第19章和4个附录），尝试搭建自己的Hadoop集群，编写并运行自己的MapReduce代码。

对于本书中关于Hadoop相关项目的介绍，大家可以有选择地学习。

在内容的编排上，各章的知识点是相对独立的，是并行的关系，因此大家可以有选择地进行学习。

当然，如果时间允许，还是建议大家系统地学习全书的内容，这样能够对Hadoop系统的机制有一个完整而系统的理解，为今后深入地研究和实践Hadoop及云计算技术打下坚实的基础。

另外，笔者希望大家在学习本书时能一边阅读，一边根据书中的指导动手实践，亲自实践本书中所给出的编程范例。

例如，先搭建一个自己的云平台，如果条件受限，可以选择伪分布的方式。

致谢在本书的编写过程中，很多Hadoop方面的实践者和研究者做了大量的工作，他们是冯博亮、程明、徐文韬、张林林、朱俊良、许翔、陈东伟、谭果、林春彬等，在此表示感谢。

陆嘉恒2012年6月于北京

<<Hadoop实战(第2版)>>

内容概要

《Hadoop实战(第2版)》能满足读者全面学习最新的Hadoop技术及其相关技术(Hive、HBase等)的需求，是一本系统且极具实践指导意义的Hadoop工具书和参考书。

第1版上市后广受好评，被誉为学习Hadoop技术的经典著作之一。

与第1版相比，第2版技术更新颖，所有技术都针对最新版进行了更新；内容更全面，几乎每一个章节都增加了新内容，而且增加了新的章节；实战性更强，案例更丰富；细节更完美，对第1版中存在的缺陷和不足进行了修正。

本书内容全面，对Hadoop整个技术体系进行了全面的讲解，不仅包括HDFS、MapReduce、YARN等核心内容，而且还包括Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等与Hadoop技术相关的重要内容。实战性强，不仅为各个知识点精心设计了大量经典的小案例，而且还包括Yahoo！等多个大公司的企业级案例，可操作系极强。

《Hadoop实战(第2版)》全书一共19章：第1~2章首先对Hadoop进行了全方位的宏观介绍，然后介绍了Hadoop在三大主流操作系统平台上的安装与配置方法；第3~6章分别详细讲解了MapReduce计算模型、MapReduce的工作机制、MapReduce应用的开发方法，以及多个精巧的MapReduce应用案例；第7章全面讲解了Hadoop的I/O操作；第8章对YARN进行了介绍；第9章对HDFS进行了详细讲解和分析；第10章细致地讲解了Hadoop的管理；第11~17章对Hadoop大生态系统中的Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等技术进行了详细的讲解；第18章讲解了Hadoop的各种常用插件，以及Hadoop插件的开发方法；第19章分析了Hadoop在Yahoo！、eBay、百度、Facebook等企业中的应用案例。本书由陆嘉恒著。

<<Hadoop实战 (第2版) >>

作者简介

陆嘉恒，资深数据库专家和云计算技术专家，对Hadoop及其相关技术有非常深入的研究，主持了多个分布式云计算项目的研究与实施，积累了丰富的实践经验。

获得新加坡国立大学博士学位，美国加利福尼亚大学尔湾分校(University of California, Irvine)

博士后，现为中国人民大学教授，博士生导师。

此外，他对数据挖掘和Web信息搜索等技术也有深刻的认识。

<<Hadoop实战 (第2版)>>

书籍目录

前言

第1章 Hadoop简介

1.1 什么是Hadoop

1.1.1 Hadoop概述

1.1.2 Hadoop的历史

1.1.3 Hadoop的功能与作用

1.1.4 Hadoop的优势

1.1.5 Hadoop应用现状和发展趋势

1.2 Hadoop项目及其结构

1.3 Hadoop体系结构

1.4 Hadoop与分布式开发

1.5 Hadoop计算模型—MapReduce

1.6 Hadoop数据管理

1.6.1 HDFS的数据管理

1.6.2 HBase的数据管理

1.6.3 Hive的数据管理

1.7 Hadoop集群安全策略

1.8 本章小结

第2章 Hadoop的安装与配置

2.1 在Linux上安装与配置Hadoop

2.1.1 安装JDK 1.6

2.1.2 配置SSH免密码登录

2.1.3 安装并运行Hadoop

2.2 在Mac OSX上安装与配置Hadoop

2.2.1 安装Homebrew

2.2.2 使用Homebrew安装Hadoop

2.2.3 配置SSH和使用Hadoop

2.3 在Windows上安装与配置Hadoop

2.3.1 安装JDK 1.6或更高版本

2.3.2 安装Cygwin

2.3.3 配置环境变量

2.3.4 安装sshd服务

2.3.5 启动sshd服务

2.3.6 配置SSH免密码登录

2.3.7 安装并运行Hadoop

2.4 安装和配置Hadoop集群

2.4.1 网络拓扑

2.4.2 定义集群拓扑

2.4.3 建立和安装Cluster

2.5 日志分析及几个小技巧

2.6 本章小结

第3章 MapReduce计算模型

3.1 为什么要用MapReduce

3.2 MapReduce计算模型

3.2.1 MapReduce Job

<<Hadoop实战 (第2版) >>

- 3.2.2 Hadoop中的Hello World程序
- 3.2.3 MapReduce的数据流和控制流
- 3.3 MapReduce任务的优化
- 3.4 Hadoop流
 - 3.4.1 Hadoop流的工作原理
 - 3.4.2 Hadoop流的命令
 - 3.4.3 两个例子
- 3.5 Hadoop Pipes
- 3.6 本章小结
- 第4章 开发MapReduce应用程序
 - 4.1 系统参数的配置
 - 4.2 配置开发环境
 - 4.3 编写MapReduce程序
 - 4.3.1 Map处理
 - 4.3.2 Reduce处理
 - 4.4 本地测试
 - 4.5 运行MapReduce程序
 - 4.5.1 打包
 - 4.5.2 在本地模式下运行
 - 4.5.3 在集群上运行
 - 4.6 网络用户界面
 - 4.6.1 JobTracker页面
 - 4.6.2 工作页面
 - 4.6.3 返回结果
 - 4.6.4 任务页面
 - 4.6.5 任务细节页面
 - 4.7 性能调优
 - 4.7.1 输入采用大文件
 - 4.7.2 压缩文件
 - 4.7.3 过滤数据
 - 4.7.4 修改作业属性
 - 4.8 MapReduce工作流
 - 4.8.1 复杂的Map和Reduce函数
 - 4.8.2 MapReduce Job中全局共享数据
 - 4.8.3 链接MapReduce Job
 - 4.9 本章小结
- 第5章 MapReduce应用案例
 - 5.1 单词计数
 - 5.1.1 实例描述
 - 5.1.2 设计思路
 - 5.1.3 程序代码
 - 5.1.4 代码解读
 - 5.1.5 程序执行
 - 5.1.6 代码结果
 - 5.1.7 代码数据流
 - 5.2 数据去重
 - 5.2.1 实例描述

<<Hadoop实战 (第2版) >>

5.2.2 设计思路

5.2.3 程序代码

5.3 排序

5.3.1 实例描述

5.3.2 设计思路

5.3.3 程序代码

5.4 单表关联

5.4.1 实例描述

5.4.2 设计思路

5.4.3 程序代码

5.5 多表关联

5.5.1 实例描述

5.5.2 设计思路

5.5.3 程序代码

5.6 本章小结

第6章 MapReduce工作机制

6.1 MapReduce作业的执行流程

6.1.1 MapReduce任务执行总流程

6.1.2 提交作业

6.1.3 初始化作业

6.1.4 分配任务

6.1.5 执行任务

6.1.6 更新任务执行进度和状态

6.1.7 完成作业

6.2 错误处理机制

6.2.1 硬件故障

6.2.2 任务失败

6.3 作业调度机制

6.4 Shuffle和排序

6.4.1 Map端

6.4.2 Reduce端

6.4.3 shuffle过程的优化

6.5 任务执行

6.5.1 推测式执行

6.5.2 任务JVM重用

6.5.3 跳过坏记录

6.5.4 任务执行环境

6.6 本章小结

第7章 Hadoop IO操作

7.1 IO操作中的数据检查

7.2 数据的压缩

7.2.1 Hadoop对压缩工具的选择

7.2.2 压缩分割和输入分割

7.2.3 在MapReduce程序中使用压缩

7.3 数据的IO中序列化操作

7.3.1 Writable类

7.3.2 实现自己的Hadoop数据类型

<<Hadoop实战 (第2版) >>

- 7.4 针对Mapreduce的文件类
 - 7.4.1 SequenceFile类
 - 7.4.2 MapFile类
 - 7.4.3 ArrayFile、SetFile和BloomMapFile
- 7.5 本章小结
- 第8章 下一代MapReduce : YARN
 - 8.1 MapReduce V2设计需求
 - 8.2 MapReduce V2主要思想和架构
 - 8.3 MapReduce V2设计细节
 - 8.4 MapReduce V2优势
 - 8.5 本章小结
- 第9章 HDFS详解
 - 9.1 Hadoop的文件系统
 - 9.2 HDFS简介
 - 9.3 HDFS体系结构
 - 9.3.1 HDFS的相关概念
 - 9.3.2 HDFS的体系结构
 - 9.4 HDFS的基本操作
 - 9.4.1 HDFS的命令行操作
 - 9.4.2 HDFS的Web界面
 - 9.5 HDFS常用Java API详解
 - 9.5.1 使用Hadoop URL读取数据
 - 9.5.2 使用FileSystem API读取数据
 - 9.5.3 创建目录
 - 9.5.4 写数据
 - 9.5.5 删除数据
 - 9.5.6 文件系统查询
 - 9.6 HDFS中的读写数据流
 - 9.6.1 文件的读取
 - 9.6.2 文件的写入
 - 9.6.3 一致性模型
 - 9.7 HDFS命令详解
 - 9.7.1 通过distcp进行并行复制
 - 9.7.2 HDFS的平衡
 - 9.7.3 使用Hadoop归档文件
 - 9.7.4 其他命令
 - 9.8 WebHDFS
 - 9.8.1 WebHDFS的配置
 - 9.8.2 WebHDFS命令
 - 9.9 本章小结
- 第10章 Hadoop的管理
 - 10.1 HDFS文件结构
 - 10.2 Hadoop的状态监视和管理工具
 - 10.2.1 审计日志
 - 10.2.2 监控日志
 - 10.2.3 Metrics
 - 10.2.4 Java管理扩展

<<Hadoop实战 (第2版) >>

- 10.2.5 Ganglia
- 10.2.6 Hadoop管理命令
- 10.3 Hadoop集群的维护
 - 10.3.1 安全模式
 - 10.3.2 Hadoop的备份
 - 10.3.3 Hadoop的节点管理
 - 10.3.4 系统升级
- 10.4 本章小结
- 第11章 Hive详解
 - 11.1 Hive简介
 - 11.1.1 Hive的数据存储
 - 11.1.2 Hive的元数据存储
 - 11.2 Hive的基本操作
 - 11.2.1 在集群上安装Hive
 - 11.2.2 配置MySQL存储Hive元数据
 - 11.2.3 配置Hive
 - 11.3 Hive QL详解
 - 11.3.1 数据定义(DDL)操作
 - 11.3.2 数据操作(DML)
 - 11.3.3 SQL操作
 - 11.3.4 Hive QL使用实例
 - 11.4 Hive网络(Web UI)接口
 - 11.4.1 Hive网络接口配置
 - 11.4.2 Hive网络接口操作实例
 - 11.5 Hive的JDBC接口
 - 11.5.1 Eclipse环境配置
 - 11.5.2 程序实例
 - 11.6 Hive的优化
 - 11.7 本章小结
- 第12章 HBase详解
 - 12.1 HBase简介
 - 12.2 HBase的基本操作
 - 12.2.1 HBase的安装
 - 12.2.2 运行HBase
 - 12.2.3 HBase Shell
 - 12.2.4 HBase配置
 - 12.3 HBase体系结构
 - 12.3.1 HRegion
 - 12.3.2 HRegion服务器
 - 12.3.3 HBase Master服务器
 - 12.3.4 ROOT表和META表
 - 12.3.5 ZooKeeper
 - 12.4 HBase数据模型
 - 12.4.1 数据模型
 - 12.4.2 概念视图
 - 12.4.3 物理视图
 - 12.5 HBase与RDBMS

<<Hadoop实战 (第2版) >>

- 12.6 HBase与HDFS
- 12.7 HBase客户端
- 12.8 Java API
- 12.9 HBase编程
 - 12.9.1 使用Eclipse开发HBase应用程序
 - 12.9.2 HBase编程
 - 12.9.3 HBase与MapReduce
- 12.10 模式设计
 - 12.10.1 模式设计应遵循的原则
 - 12.10.2 学生表
 - 12.10.3 事件表
- 12.11 本章小结
- 第13章 Mahout详解
 - 13.1 Mahout简介
 - 13.2 Mahout的安装和配置
 - 13.3 Mahout API简介
 - 13.4 Mahout中的频繁模式挖掘
 - 13.4.1 什么是频繁模式挖掘
 - 13.4.2 Mahout中的频繁模式挖掘
 - 13.5 Mahout中的聚类和分类
 - 13.5.1 什么是聚类和分类
 - 13.5.2 Mahout中的数据表示
 - 13.5.3 将文本转化成向量
 - 13.5.4 Mahout中的聚类、分类算法
 - 13.5.5 算法应用实例
 - 13.6 Mahout应用：建立一个推荐引擎
 - 13.6.1 推荐引擎简介
 - 13.6.2 使用Taste构建一个简单的推荐引擎
 - 13.6.3 简单分布式系统下基于产品的推荐系统简介
 - 13.7 本章小结
- 第14章 Pig详解
 - 14.1 Pig简介
 - 14.2 Pig的安装和配置
 - 14.2.1 Pig的安装条件
 - 14.2.2 Pig的下载、安装和配置
 - 14.2.3 Pig运行模式
 - 14.3 Pig Latin语言
 - 14.3.1 Pig Latin语言简介
 - 14.3.2 Pig Latin的使用
 - 14.3.3 Pig Latin的数据类型
 - 14.3.4 Pig Latin关键字
 - 14.4 用户定义函数
 - 14.4.1 编写用户定义函数
 - 14.4.2 使用用户定义函数
 - 14.5 Zebra简介
 - 14.5.1 Zebra的安装
 - 14.5.2 Zebra的使用简介

<<Hadoop实战 (第2版) >>

- 14.6 Pig实例
 - 14.6.1 Local模式
 - 14.6.2 MapReduce模式
- 14.7 Pig进阶
 - 14.7.1 数据实例
 - 14.7.2 Pig数据分析
- 14.8 本章小结
- 第15章 ZooKeeper详解
 - 15.1 ZooKeeper简介
 - 15.1.1 ZooKeeper的设计目标
 - 15.1.2 数据模型和层次命名空间
 - 15.1.3 ZooKeeper中的节点和临时节点
 - 15.1.4 ZooKeeper的应用
 - 15.2 ZooKeeper的安装和配置
 - 15.2.1 安装ZooKeeper
 - 15.2.2 配置ZooKeeper
 - 15.2.3 运行ZooKeeper
 - 15.3 ZooKeeper的简单操作
 - 15.3.1 使用ZooKeeper命令的简单操作步骤
 - 15.3.2 ZooKeeper API的简单使用
 - 15.4 ZooKeeper的特性
 - 15.4.1 ZooKeeper的数据模型
 - 15.4.2 ZooKeeper会话及状态
 - 15.4.3 ZooKeeper watches
 - 15.4.4 ZooKeeper ACL
 - 15.4.5 ZooKeeper的一致性保证
 - 15.5 使用ZooKeeper进行Leader选举
 - 15.6 ZooKeeper锁服务
 - 15.6.1 ZooKeeper中的锁机制
 - 15.6.2 ZooKeeper提供的一个写锁的实现
 - 15.7 使用ZooKeeper创建应用程序
 - 15.7.1 使用Eclipse开发ZooKeeper应用程序
 - 15.7.2 应用程序实例
 - 15.8 BooKeeper
 - 15.9 本章小结
- 第16章 Avro详解
 - 16.1 Avro介绍
 - 16.1.1 模式声明
 - 16.1.2 数据序列化
 - 16.1.3 数据排列顺序
 - 16.1.4 对象容器文件
 - 16.1.5 协议声明
 - 16.1.6 协议传输格式
 - 16.1.7 模式解析
 - 16.2 Avro的CC++实现
 - 16.3 Avro的Java实现
 - 16.4 GenAvro(Avro IDL)语言

<<Hadoop实战 (第2版) >>

- 16.5 Avro SASL概述
- 16.6 本章小结
- 第17章 Chukwa详解
 - 17.1 Chukwa简介
 - 17.2 Chukwa架构
 - 17.2.1 客户端及其数据模型
 - 17.2.2 收集器
 - 17.2.3 归档器和分离解析器
 - 17.2.4 HICC
 - 17.3 Chukwa的可靠性
 - 17.4 Chukwa集群搭建
 - 17.4.1 基本配置要求
 - 17.4.2 Chukwa的安装
 - 17.4.3 Chukwa的运行
 - 17.5 Chukwa数据流的处理
 - 17.6 Chukwa与其他监控系统比较
 - 17.7 本章小结
- 本章参考资料
- 第18章 Hadoop的常用插件与开发
 - 18.1 Hadoop Studio的介绍和使用
 - 18.1.1 Hadoop Studio的介绍
 - 18.1.2 Hadoop Studio的安装配置
 - 18.1.3 Hadoop Studio的使用举例
 - 18.2 Hadoop Eclipse的介绍和使用
 - 18.2.1 Hadoop Eclipse的介绍
 - 18.2.2 Hadoop Eclipse的安装配置
 - 18.2.3 Hadoop Eclipse的使用举例
 - 18.3 Hadoop Streaming的介绍和使用
 - 18.3.1 Hadoop Streaming的介绍
 - 18.3.2 Hadoop Streaming的使用举例
 - 18.3.3 使用Hadoop Streaming常见的问题
 - 18.4 Hadoop Libhdfs的介绍和使用
 - 18.4.1 Hadoop Libhdfs的介绍
 - 18.4.2 Hadoop Libhdfs的安装配置
 - 18.4.3 Hadoop Libhdfs API简介
 - 18.4.4 Hadoop Libhdfs的使用举例
 - 18.5 本章小结
- 第19章 企业应用实例
 - 19.1 Hadoop在Yahoo !
的应用
 - 19.2 Hadoop在eBay的应用
 - 19.3 Hadoop在百度的应用
 - 19.4 即刻搜索中的Hadoop
 - 19.4.1 即刻搜索简介
 - 19.4.2 即刻Hadoop应用架构
 - 19.4.3 即刻Hadoop应用分析
 - 19.5 Facebook中的Hadoop和HBase

<<Hadoop实战 (第2版) >>

19.5.1 Facebook中的任务特点

19.5.2 MySQL VS Hadoop+HBase

19.5.3 Hadoop和HBase的实现

19.6 本章小结

本章参考资料

附录A 云计算在线检测平台

附录B Hadoop安装、运行与使用说明

附录C 使用DistributedCache的MapReduce程序

附录D 使用ChainMapper和ChainReducer的MapReduce程序

章节摘录

第1章 Hadoop 简介 本章内容 什么是 Hadoop Hadoop 项目及其结构 Hadoop 体系结构 Hadoop 与分布式开发 Hadoop 计算模型—MapReduce Hadoop 数据管理 Hadoop 集群安全策略 本章小结 1.1 什么是 Hadoop 1.1.1 Hadoop 概述 Hadoop 是 Apache 软件基金会旗下的一个开源分布式计算平台。

以 Hadoop 分布式文件系统 (Hadoop Distributed File System, HDFS) 和 MapReduce (Google MapReduce 的开源实现) 为核心的 Hadoop 为用户提供了系统底层细节透明的分布式基础架构。

HDFS 的高容错性、高伸缩性等优点允许用户将 Hadoop 部署在低廉的硬件上, 形成分布式系统; MapReduce 分布式编程模型允许用户在不了解分布式系统底层细节的情况下开发并行应用程序。所以用户可以利用 Hadoop 轻松地组织计算机资源, 从而搭建自己的分布式计算平台, 并且可以充分利用集群的计算和存储能力, 完成海量数据的处理。

经过业界和学术界长达 10 年的锤炼, 目前的 Hadoop 1.0.1 已经趋于完善, 在实际的数据处理和分析任务中担当着不可替代的角色。

<<Hadoop实战 (第2版)>>

媒体关注与评论

经过学术界和业界近10年的努力，Hadoop技术已经趋于完善而且应用广泛，几乎已经成为Big Data领域的事实标准。

Hadoop技术本身比较复杂，而且还涉及Pig、ZooKeeper、Hive、HBase等一系列技术，学习门槛比较高，对于初学者和基础不太扎实的读者而言，有一本适合系统学习的Hadoop图书显得十分重要。

本书即是专门为这两类读者量身定做的：第一，它的内容非常全面和前沿，不仅讲解了最新的Hadoop技术和第二代MapReduce，还讲解了涉及的所有周边技术，能满足系统学习的需求；第二，实战性非常强，不仅很多知识点配有精心设计的小案例，而且有完整的企业级案例，能满足操作实践的需求；第三，这一版在上一版的基础上根据最新的技术做了更新和补充，能满足读者学习最新技术的需求。本书第1版不仅取得了好的销量，而且广受好评，第2版在内容上有很大的提升，相信能让更多的读者从中受益。

——EasyHadoop 国内专业的Hadoop社区，致力于让Hadoop大数据分析更简单

<<Hadoop实战 (第2版) >>

编辑推荐

《Hadoop实战(第2版)》编辑推荐：第1版广受好评，第2版基于Hadoop及其相关技术最新版本撰写，从多角度做了全面的修订和补充。

不仅详细讲解了新一代的Hadoop技术，而且全面介绍了Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等重要技术，是系统学习Hadoop技术的首选之作。

<<Hadoop实战 (第2版) >>

版权说明

本站所提供下载的PDF图书仅提供预览和简介, 请支持正版图书。

更多资源请访问:<http://www.tushu007.com>