

<<实战Hadoop>>

图书基本信息

书名：<<实战Hadoop>>

13位ISBN编号：9787121144752

10位ISBN编号：7121144751

出版时间：2011-8

出版时间：电子工业出版社

作者：刘鹏

页数：456

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>

## <<实战Hadoop>>

### 内容概要

作为谷歌云计算基础架构的模仿实现，Hadoop堪称业界最经典的开源云计算平台软件。

《实战Hadoop--开启通向云计算的捷径》(作者刘鹏)是原著的Hadoop编程技术书籍，是云计算专家刘鹏教授继《云计算》教材取得成功后，再次组织团队精心编写的又一力作，其作者均来自拥有丰富实践经验的云计算技术研发和教学团队。

《实战Hadoop--开启通向云计算的捷径》强调动手、强调实战，以风趣幽默的语言和一系列生动的实战应用案例，系统地讲授了Hadoop的核心技术和扩展技术，包括：

HDFS、MapReduce、HBase、Hive、Pig、Cassandra、

Chukwa和ZooKeeper等，并给出了3个完整的Hadoop云计算综合应用实例，最后介绍了保障Hadoop平台可靠性的方法。

本书读者对象为各类云计算相关企业、高校和科研机构的研发人员，亦适合作为高校研究生和本科生教材。

## 作者简介

刘鹏，清华大学博士，解放军理工大学教授、学科带头人，中国云计算专家委员会委员。主要研究方向为信息网格和云计算，完成科研课题18项，发表论文70余篇，获部级科技进步奖6项。曾夺得国际计算机排序比赛冠军，并两次夺得全国高校科技比赛最高奖，获“全军十大学习成才标兵”、“南京十大杰出青年”和“清华大学学术新秀”等称号。2002年首倡的“网格计算池”和2003年研发的“反垃圾邮件网格”分别为云计算和云安全的前身。创办了知名的中国网格（chinagrid.net）和中国云计算（chinacloud.cn）网站。

## &lt;&lt;实战Hadoop&gt;&gt;

## 书籍目录

## 第1章 神奇的大象——Hadoop

## 1.1 初识神象

## 1.2 Hadoop初体验

## 1.2.1 了解Hadoop的构架

## 1.2.2 查看Hadoop活动

## 1.3 Hadoop族群

## 1.4 Hadoop安装

## 1.4.1 在Linux系统中安装Hadoop

## 1.4.2 在Windows系统中安装Hadoop

## 1.4.3 站在象背上说“hello”

## 1.4.4 Eclipse下的Hadoop应用开发

## 参考文献

## 第2章 HDFS——不怕故障的海量存储

## 第3章 分久必合——MapReduce

## 第4章 一张无限大的表——HBase

## 第5章 更上一层楼——MapReduce进阶

## 第6章 Hive——飞进数据仓库的小蜜蜂

## 第7章 Pig——一头什么都能吃的猪

## 第8章 Facebook的女神——Cassandra

## 第9章 Chukwa——收集数据的大乌龟

## 第10章 一统天下——ZooKeeper

## 第11章 综合实战1——打造一个搜索引擎

## 第12章 综合实战2——生物信息学应用

## 第13章 综合实战3——移动通信信令监测与查询

## 第14章 高枕无忧——Hadoop容错

## 章节摘录

版权页：插图：在对源文件进行功能性处理之前，有必要对11.2.2节生成的源文件进行一次预分析和过滤。

主要原因有以下几个。

(1) 在遇到故障并恢复爬行后，爬虫会从日志文件的最近一页（一页包含20个帖子）重新继续爬取工作。

从最近一页爬取保证了帖子不被遗漏，但是会导致有些帖子被重复爬取（一般不会超过2次）。

为保证索引时指定文件的唯一性，有必要过滤相同帖子的记录，保证帖子的唯一性。

(2) 爬取过程中，从遇到的帖子中抽取出的信息可能并不符合我们的要求，比如，源文件中可能会有之类的记录。

这样的空信息记录也必须过滤掉。

对源文件进行一次预分析和过滤以确保数据的完整性和正确性，可以避免后续步骤中由于数据不正确而引发的一些问题。

对元数据进行预分析是海量数据处理过程中很必要也是很自然的一步。

## 媒体关注与评论

随着计算模式从个人计算向云计算的发展，开源软件的比重和作用越来越大，现在多数云服务都基于开源平台Hadoop是国际著名的云计算开源平台，本书通过许多实例，深入系统地分析了Hadoop的技术、源码以及构建实用系统的方法。

为国内产学研各界开启了一条通向云计算的捷径。

——中国工程院院士、中国电子学会计算专家委员会副主任 倪光南作为云计算基础架构的开源软件，Hadoop堪称业界最成熟和经典的案例该书是云计算业界专家刘鹏教授带领科研团队，基于多年研究心得，做了大量技术开发和测试，以严谨的态度和诙谐幽默的语言编写的Hadoop实战宝典。

相信对推动中国云计算技术的发展和具有重要价值。

——中兴通讯首席构师、中国电子学会云计算专家委员会委员 罗圣美热烈祝贺刘鹏教授主编的实战Hadoop——开启通向云计算的捷径。

一书出版发行，它将大力推动我国云计算的应用。

——清华大学教授、中国电子学会云计算专家委员会委员 郑纬民

编辑推荐

《实战Hadoop:开启通向云计算的捷径》：云计算核心研发团队剖析Hadoop:怎么装？  
怎么编程？  
怎么解决实际问题？

#### 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>