

<<Hadoop实战>>

图书基本信息

书名：<<Hadoop实战>>

13位ISBN编号：9787111359449

10位ISBN编号：7111359445

出版时间：2011-10

出版时间：机械工业出版社华章公司

作者：陆嘉恒

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>>

内容概要

本书是一本系统且极具实践指导意义的hadoop工具书和参考书。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和mapreduce这两大核心内容，而且还包括hive、hbase、mahout、pig、zookeeper、avro、chukwa等与hadoop相关的子项目的内容。实战性强，为各个知识点精心设计了大量经典的小案例，易于理解，可操作性强。

全书一共18章：第1章全面介绍了hadoop的概念、优势、项目结构、体系结构，以及它与分布式计算的关系；第2章详细讲解了hadoop集群的安装和配置，以及常用的日志分析技巧；第3章分析了hadoop在yahoo！、ebay、facebook和百度的应用案例，以及hadoop平台上海量数据的排序；第4-7章深入地讲解了mapreduce计算模型、mapreduce应用的开发方法、mapreduce的工作机制，同时还列出了多个mapreduce的应用案例，涉及单词计数、数据去重、排序、单表关联和多表关联等内容；第8-11章全面地阐述了hadoop的i/o操作、hdfs的原理与基本操作，以及hadoop的各种管理操作，如集群的维护等；第12-17章详细而系统地讲解了hive、hbase、mahout、pig、zookeeper、avro、chukwa等所有与hadoop相关的子项目的原理及使用，以及这些子项目与hadoop的整合使用；第18章以实例的方式讲解了常用hadoop插件的使用和hadoop插件的开发。

本书既适合没有hadoop基础的初学者系统地学习，又适合有一定hadoop基础但是缺乏实践经验的读者实践和参考。

<<Hadoop实战>>

作者简介

陆嘉恒，中国人民大学副教授，新加坡国立大学博士，美国加利福尼亚大学尔湾分校(University of California, Irvine)

博士后。

专注于云计算及其相关技术的研究，对Hadoop有较深入的研究，积累了丰富的实践经验。

对分布式计算和海量数据处理有深刻的认识，主持并完成了多个国家863和自然科学基金项目的研究与实施。

2009年入选新世纪优秀人才，2010年入选北京科技新星。

主持《云计算概论》课程获教育部-IBM精品课程称号。

<<Hadoop实战>>

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的应用现状和发展趋势

1.2 Hadoop项目及其结构

1.3 Hadoop的体系结构

1.3.1 HDFS的体系结构

1.3.2 MapReduce的体系结构

1.4 Hadoop与分布式开发

1.5 Hadoop计算模型——MaDRcduce

1.6 Hadoop的数据管理

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 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Windows上安装与配置Hadoop

2.2.1 安装Cygwi~

2.2.2 配置环境变量

2.2.3 安装和启动sshd服务

2.2.4 配置SSH免密码登录

2.3 安装和配置Hadoop集群

2.3.1 网络拓扑

2.3.2 定义集群拓扑

2.3.3 建立和安装Cluster

2.4 日志分析及几个小技巧

2.5 小结

第3章 Hadoop应用案例分析

3.1 Hadoop在Yahoo!的应用

3.2 Hadoop在eBay的应用

3.3 Hadoop在百度的应用

3.4 Hadoop在Facebook的应用

3.5 Hadoop平台上的海量数据排序

3.6 小结

第4章 MapReduce计算模型

第5章 开发MapReduce应用程序

第6章 MapReduce应用案例

<<Hadoop实战>>

- 第7章 MapReduce工作机制
- 第8章 HadoopI/O
- 第9章 HDFS详解
- 第10章 Hadoop的管理
- 第11章 Hive详解
- 第12章 HBase详解
- 第13章 Mahout详解
- 第14章 Pigi详解
- 第15章 ZooKeepeti详解
- 第16章 Avro详解
- 第17章 Chukwai详解
- 第18章 Hadoop的常用插件与开发
- 附录A 云计算在线检测平台

章节摘录

版权页：插图：2.3 安装和配置Hadoop集群2.3.1 网络拓扑通常来说，一个Hadoop的集群体系结构由两层网络拓扑组成，如图2-1所示。

结合实际的应用来看，每个机架中会有30~40台机器，这些机器共享一个1GB带宽的网络交换机。

在所有的机架之上还有一个核心交换机或路由器，通常来说其网络交换能力为1GB或更高。

可以很明显地看出，同一个机架中机器节点之间的带宽资源肯定要比不同机架中机器节点间丰富。

这也是Hadoop随后设计数据读写分发策略要考虑的一个重要因素。

2.3.2 定义集群拓扑在实际应用中，为了使Hadoop集群获得更高的性能，读者需要配置集群使Hadoop能够感知其所在的网络拓扑结构。

当然如果集群中机器数量很少，而且它们存在于一个机架中，那么就不用做太多额外的工作，而当集群中存在多个机架时，就要使Hadoop清晰地知道每台机器所在的机架。

随后，在处理Map Reduce任务时，Hadoop会优先选择在机架内部进行数据传输，而不是在机架间，这样就可以更充分地使用网络带宽资源。

同时，HDFS可以更加智能地部署数据副本，并在性能和可靠性间寻找到最优的平衡。

<<Hadoop实战>>

媒体关注与评论

将网格计算、并行计算和虚拟化等技术融为一体的云计算技术已成为我们当下存储和处理海量数据的最佳选择之一。

Hadoop的开源、高性能、高容错、跨平台等特点又使其成为架构云计算平台的首选。

本书以实践为主，理论与实践相结合，全面阐述了整个Hadoop技术体系，适合读者系统地学习。

强烈推荐！

——Hadoop中文网 目前，国内的很多互联网企业都在使用或正准备使用Hadoop技术，这些企业都面临着一个共同的难题：Hadoop方面的人才难求。

Hadoop方面的人才之所以难找，一方面是因为Hadoop在国内应用的时间不长，从业人员并不多；另一方面是因为Hadoop技术本身较难以掌握，而且与涉及海量数据处理的实际生产环境密切相关。

本书很好地把握住了当前Hadoop从业者的核心需求，不仅理论知识全面，更重要的是包含大量与实际生产环境相结合的案例，极具指导意义。

——Hadoop用户社区 Hadoop可谓炙手可热，在全球范围内，已经有数量庞大的大中型互联网公司开始使用Hadoop，国外的Amazon、Facebook、Yahoo！

，国内的腾讯、百度、淘宝、阿里巴巴等都是成功应用Hadoop的典范。

然而，Hadoop技术本身却极为复杂，而且涉及众多其他的技术，学习门槛比较高。

本书从中初级读者的需求出发，以实践为导向，全面而系统地讲解了Hadoop技术本身，以及与之相关的其他各种技术。

对于想系统学习Hadoop和想增加实战经验的读者来说，本书不可多得！

——51CTO

<<Hadoop实战>>

编辑推荐

《Hadoop实战》内容全面，涵盖Hadoop技术本身和Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等所有与Hadoop相关的子项目，实战性强，为各个知识点精心设计了大量经典的小案例，易于理解，可操作性强。

当越来越多的人涌向互联网时，不仅互联网应用的种类和形态在不断发生变化，而且互联网上的信息量也在以我们无法想象的速度迅猛增长。

如今，无论是传统企业，还是互联网企业都越来越重视对所有与企业相关的各种数据的分析和利用。对海量数据的处理能力已成为现代企业的核心竞争力之一。

云计算技术的出现为海量数据的处理提供了良好的解决方案。

Hadoop是云计算技术的开源实现，具有高容错、跨平台等优势，用户可以利用Hadoop轻松地组织计算机资源来搭建自己的分布式云计算平台，从而充分地利用集群的计算和存储能力，完成对海量数据的处理。

《Hadoop实战》对Hadoop的整个技术体系进行了全面的讲解，不仅包括HDFS和MapReduce这两大核心内容，而且还包括Hive、HBase、Mahout、Pig、ZooKeeper、Avro、Chukwa等与Hadoop相关的子目的内容。

《Hadoop实战》“以理论为基础、以实践为重点”，希望能为大家奉上一场Hadoop的盛宴。

如果你想使用Hadoop，或正在使用Hadoop，《Hadoop实战》将会给你绝佳的指导。

<<Hadoop实战>>

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>