《基于Hadoop的大数据分析和处理》

书籍信息

版 次：1
页 数：
字 数：
印刷时间：2017年06月01日
开 本：16开
纸 张：胶版纸
包 装：平装-胶订
是否套装：否
国际标准书号ISBN：9787121317392

内容简介

本书基于云计算和大数据，介绍大数据处理和分析的技术，分为两部分。*部分介绍Hadoop基础知识，内容包括：Hadoop的介绍和集群构建、Hadoop的分部式系统架构、MapReduce及其应用、Hadoop的版本特征及进化。第二部分以云计算为主题，详细论述利用Hadoop的大数据分析和处理工具，以及NoSQL技术，内容包括：云计算和Hadoop、Amazon服务中的MapReduce应用、Hadoop应用下的大数据分析、NoSQL、HBase。本书不单纯地讲述理论和概念，而是基于目具体的工具和技术(Hadoop和NoSQL)，利用大量实际案例，通过实际的操作和应用来组织大数据处理和分析技术，有利于读者从工程应用的角度进行实际掌握和利用。适合相关专业的本科生、研究生和软件工程师学习。

作者简介

魏祖宽，男，电子科技大学教授，博士，韩国科技协会、中国计算机学会、日本电子电器协会会员。承担计算机以及软件学院的本科和研究生的数据库课程教学和实验教学，及云计算和大数据方面的新课。主持数据库应用、GIS应用等方面的应用课题10多项(国家自然基金委，省/市级科技局等科研项目，以及企业横向项目)，现专注于云存储方面的应用科研项目。

目 录

目 录

前 言

前 言
欧盟的“ INFO2000计划 ”中对内容产业的定义是：那些制造、开发、包装和销售信息产品及其服务的企业，其中包括在各种媒介上的印刷品（报纸、书籍、杂志等）；电子出版物（联机数据库、音像制品服务，以传真及光盘为基础的服务以及电子游戏等）；音像传播（电视、录像、广播和影院），还有一些定义把部分软件业（包括课程软件）也放进去了。
“ 在不久未来，信息服务内容的质量高低将取决于如何加工大数据 ”。
很久以前就已经感觉到，内容（contents），在大部分的服务和产品中，已经成为最重要的决定要素。最初由谷歌出世、最近各家厂商纷纷推出的互联网电视，就是这样一个例子，虽然产品硬件各有特色地优异，但其中最核心的内容提供才是吸引顾客的关键。
问题是，随着互联网技术的急速发展，构建信息内容的数据量也在急速增加。这类量级巨大、急速增加的数据信息我们称为“ 大数据 ”。一般来讲，当我们说“ 信息内容的质量高低取决于如何加工信息大数据 ”的时候，就意味着优质高效地加工这些信息大数据所对应的软件技术是必需的。

欧盟的“ INFO2000计划 ”中对内容产业的定义是：那些制造、开发、包装和销售信息产品及其服务的企业，其中包括在各种媒介上的印刷品（报纸、书籍、杂志等）；电子出

版物（联机数据库、音像制品服务，以传真及光盘为基础的服务以及电子游戏等）；音像传播（电视、录像、广播和影院），还有一些定义把部分软件业（包括课程软件）也放进去了。

很久以前就已经感觉到，内容（contents），在大部分的服务和产品中，已经成为最重要的决定要素。最初由谷歌出世、最近各家厂商纷纷推出的互联网电视，就是这样一个例子，虽然产品硬件各有特色地优异，但其中最核心的内容提供才是吸引顾客的关键。

我们通过本书试图和读者们分享和思考“ 如何存储和处理这类信息大数据 ”。我们看到的YouTube或别的视频网站已经在多年前就在思考这些问题：适应不同的服务平台，从成千上万个视频中，根据顾客的兴趣，精心地经过推荐和过滤等环节，向顾客提供高质量的内容视频。本书中，正是要介绍可以简单地完成这些数据加工任务的开源软件Hadoop及其关联工具。特别的，对和Hadoop一起用于实际大数据分析的专用工具进行了有深度的探讨，并基于图表和案例进行了形象的说明。通过本书，比起对Hadoop的相关开源代码的理解来说，作者更着重于读者在实战中对实际大数据分析平台的理解和见识。特别是，在数据分析处理、平台架构构建时针对大数据处理所遇见的共通性必需技术进行了详细的介绍。

第二部分包括第5章 云计算和Hadoop、第6章 Amazon Elastic MapReduce的倍增利用、第7章 Hadoop应用下的大数据分析、第8章 数据中的DBMS、NoSQL和第9章 HBase：Hadoop中的NoSQL。该部分从云计算的基本概念讲起，通过介绍Amazon的主要服务内容，详细了解将云计算和大数据有效结合的典型云服务——Amazon Hadoop服务，对Hive、Pig、EC2 等可供应用的技术进行了说明；通过了解Mahout、R RHive和Giraph Framework等工具的设置方法和应用实例，进一步了解大数据分析的具体方法；最后介绍了高度综合大数据存储、实时查询及分析功能为一体的NoSQL技术，并详细讲解了Hadoop生态界中的NoSQL——HBase技术。

2017年春 于西南

显示全部信息