

《大数据浪潮》

图书基本信息

书名：《大数据浪潮》

13位ISBN编号：9787568012700

出版时间：2015-11

作者：陈敏,张东,张引,亓开元

页数：251

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《大数据浪潮》

内容概要

本书站在大数据学科发展前沿和大数据产业发展方向上，结合学术界和工业界的研究成果和实践经验，传播大数据领域的研究动态和先进技术。

全书通过“基础、进阶、实战、应用”四个篇章，既深入浅出地介绍大数据的基本概念，又面面俱到地剖析了大数据整体解决方案所涉及的具体技术细节。

本书既适合初学者作为基础学习资料，又适合科研人员作为理论研究教程，更适合从业人员作为技术参考书目。

《大数据浪潮》

作者简介

陈敏，男，1980年12月出生，教授，博士生导师，华中科技大学嵌入与普适计算实验室主任，23岁获华南理工大学通信与信息系博士学位。发表学术论文180余篇，其中包括SCI论文90余篇，IEEE/ACM期刊论文40余篇；相关论文在美国科技信息索引SCI数据库中他引总数500次，Google Scholar中引用总数超过4250次，单篇论文引用超高520次，其中十篇第一作者论著引用次数超过1800次，H-index = 29；曾获第五届国际异构网服务质量大会QShine 2008最佳论文奖和IEEE ICC 2012最佳论文奖。2012年入选国家第二批“青年千人计划”，长期从事普适计算、移动通信、多媒体通信、传感网、云计算、远程医疗等领域的研究工作，担任Information Science等7个国际SCI学术期刊的副编辑或编委，以及IEEE Wireless Communications及IEEE Network等多个著名国际杂志及期刊的特邀编辑。任IEEE CIT 2012，TRIDENTCOM 2014，Mobimedia 2015等国际学术大会副总主席，任Cloudcomp 2014/2015大会指导委员会成员。陈敏教授具有多年海外工作经验，曾在韩国国立汉城大学、加拿大不列颠哥伦比亚大学从事博士后研究，曾任韩国首尔大学助理教授。1999年入选国际电子电气工程师学会（IEEE）高级会员，2014年担任IEEE计算机协会大数据技术委员会主席。入选爱思唯尔2014年中国高被引学者。

书籍目录

目录

- 1 大数据背景(8)
 - 1.1 大数据时代的到来(8)
 - 1.2 大数据定义及特征(4)
 - 1.3 大数据价值(6)
 - 1.4 大数据备受关注(7)
 - 1.5 大数据带来的挑战(8)
- 2 大数据基础(10)
 - 2.1 云计算(10)
 - 2.1.1 云计算概述(10)
 - 2.1.2 云计算与大数据的联系(12)
 - 2.2 物联网(13)
 - 2.2.1 物联网概述(13)
 - 2.2.2 物联网与大数据(14)
 - 2.3 数据中心(14)
 - 2.3.1 数据中心概述(14)
 - 2.3.2 数据中心与大数据(15)
- 3 大数据的生成和采集(19)
 - 3.1 大数据生成(19)
 - 3.1.1 企业内部数据(19)
 - 3.1.2 物联网数据(20)
 - 3.1.3 互联网数据(21)
 - 3.1.4 生物医疗数据(21)
 - 3.1.5 其他科学数据(22)
 - 3.2 大数据采集(22)
 - 3.2.1 数据收集(23)
 - 3.2.2 数据传输(25)
 - 3.2.3 数据预处理(26)
- 4 大数据存储(28)
 - 4.1 海量存储系统(28)
 - 4.2 分布式存储系统(29)
 - 4.3 大数据存储机制(31)
 - 4.3.1 数据库技术(32)
 - 4.3.2 数据库编程模型(36)
- 5 大数据分析(39)
 - 5.1 传统数据分析方法(39)
 - 5.2 大数据分析方法(40)
 - 5.3 大数据分析架构(42)
 - 5.4 大数据挖掘和分析软件(43)
- 大数据浪潮——大数据整体解决方案及关键技术探索目录6 大数据整体解决方案(47)
 - 6.1 大数据解决方案方法论(47)
 - 6.1.1 大数据解决方案参考模型(48)
 - 6.1.2 大数据解决方案分类(49)
 - 6.2 大数据硬件平台(50)
 - 6.2.1 可扩展性设计(51)
 - 6.2.2 可定制性设计(52)
 - 6.3 大数据软件系统(55)

- 6.3.1 大数据处理系统核心模块(55)
- 6.3.2 发行版增强功能、企业应用优化和增值服务(56)
- 6.3.3 基于内存计算的大数据处理系统(60)
- 6.4 大数据典型处理流程(63)
- 6.5 大数据一体化解决方案比较(67)
- 7 分布式文件系统HDFS(72)
 - 7.1 Hadoop I/O操作(72)
 - 7.1.1 I/O操作中的数据检查(73)
 - 7.1.2 数据的压缩(76)
 - 7.1.3 数据的I/O中序列化操作(78)
 - 7.2 Hadoop文件系统(87)
 - 7.3 HDFS体系结构(89)
 - 7.3.1 HDFS的特点和局限(89)
 - 7.3.2 HDFS相关概念(90)
 - 7.3.3 HDFS架构(91)
 - 7.4 HDFS文件结构(94)
 - 7.4.1 NameNode的文件结构(94)
 - 7.4.2 编辑日志(edit log)及文件系统映像(filesystem image)(95)
 - 7.4.3 Secondary NameNode的目录结构(96)
 - 7.4.4 DataNode的目录结构(97)
 - 7.5 HDFS读/写数据流(98)
 - 7.5.1 文件的读取(98)
 - 7.5.2 文件的写入(99)
 - 7.5.3 一致性模型(101)
 - 7.6 HDFS命令详解(102)
 - 7.6.1 通过distcp进行并行复制(102)
 - 7.6.2 HDFS平衡(103)
 - 7.6.3 其他命令(103)
- 8 并行编程模型MapReduce(108)
 - 8.1 MapReduce体系结构(108)
 - 8.1.1 MapReduce基本模型(108)
 - 8.1.2 MapReduce作业执行流程(108)
 - 8.2 MapReduce关键流程详解(110)
 - 8.2.1 partiton过程(110)
 - 8.2.2 combine过程(111)
 - 8.2.3 shuffle过程(112)
 - 8.3 MapReduce高级应用(114)
 - 8.3.1 二次排序(114)
 - 8.3.2 全排序(119)
 - 8.3.3 分布式缓存(121)
 - 8.3.4 MapReduce 小文件处理与文件压缩(123)
 - 8.3.5 MapReduce负载均衡(125)
- 9 NoSQL数据库HBase(128)
 - 9.1 HBase体系结构(128)
 - 9.2 RowKey的设计与数据访问(132)
 - 9.3 过滤器(135)
 - 9.3.1 比较过滤器(137)
 - 9.3.2 专用过滤器(137)
 - 9.3.3 附加过滤器(139)

- 9.3.4 FilterList(139)
- 9.3.5 自定义过滤器(140)
- 9.4 HBase多维数据访问(142)
 - 9.4.1 通过Filter实现(142)
 - 9.4.2 通过设计RowKey实现(143)
- 9.5 协处理器Coprocessor(143)
 - 9.5.1 Coprocessor类(144)
 - 9.5.2 协处理器的加载(144)
 - 9.5.3 观察者(146)
 - 9.5.4 终端(150)
- 9.6 二级索引(154)
 - 9.6.1 全局索引 (global index) (154)
 - 9.6.2 本地索引 (local index) (155)
- 10 交互式查询语言Hive(157)
 - 10.1 Hive体系结构(157)
 - 10.1.1 Hive客户端(159)
 - 10.1.2 Metastore(159)
 - 10.2 Hive数据类型(160)
 - 10.2.1 基本类型(161)
 - 10.2.2 复杂类型(161)
 - 10.3 Hive存储方式和压缩类型(162)
 - 10.3.1 托管表和外部表(162)
 - 10.3.2 存储方式(163)
 - 10.4 Hive关键技术(164)
 - 10.4.1 HiveQL简介(164)
 - 10.4.2 Hive表的创建(165)
 - 10.4.3 Hive表的数据加载(166)
 - 10.4.4 Hive表的查询(167)
 - 10.4.5 Hive表的更改(170)
 - 10.4.6 Hive表的删除(171)
 - 10.4.7 Hive表的分区(171)
 - 10.4.8 Hive表的分桶(173)
 - 10.4.9 用户定义函数(174)
 - 10.5 Hive优化技术(175)
 - 10.5.1 Join优化(175)
 - 10.5.2 数据倾斜优化(176)
 - 10.5.3 Map和Reduce个数控制(177)
- 11 资源管理和调度框架——YARN(180)
 - 11.1 MRv1架构面临的问题(180)
 - 11.2 YARN架构(181)
 - 11.2.1 YARN整体架构(181)
 - 11.2.2 RM组件的作用(183)
 - 11.2.3 AM组件的作用(184)
 - 11.2.4 NM组件的作用(185)
 - 11.2.5 运行在YARN上的计算框架(186)
 - 11.2.6 在YARN上定制计算框架(187)
 - 11.3 YARN管理后台简介(188)
 - 11.4 YARN资源调度(192)
- 12 内存计算引擎Spark(197)

- 12.1 Spark简介(197)
- 12.2 Spark整体架构(198)
- 12.3 Spark核心概念(199)
 - 12.3.1 弹性分布式数据集(199)
 - 12.3.2 RDD模型的优点(200)
 - 12.3.3 Spark DAG(201)
- 12.4 Spark编程模型(202)
 - 12.4.1 Spark初始化(203)
 - 12.4.2 RDDs(203)
 - 12.4.3 Shared Variables(205)
- 12.5 Spark相关组件(207)
- 12.6 Spark应用实例(208)
 - 12.6.1 InMemory Analytics(208)
 - 12.6.2 Traffic Modeling(209)
 - 12.6.3 Twitter Spam Classification (209)
- 13 大数据应用(213)
 - 13.1 大数据应用演化(213)
 - 13.2 大数据分析的关键领域(214)
 - 13.2.1 结构化数据分析(214)
 - 13.2.2 文本分析(215)
 - 13.2.3 Web分析(216)
 - 13.2.4 多媒体分析(217)
 - 13.2.5 网络分析(218)
 - 13.2.6 移动分析(219)
- 14 大数据案例分析(221)
 - 14.1 物联网大数据(221)
 - 14.1.1 物联网大数据的表示(222)
 - 14.1.2 物联网大数据的预处理(223)
 - 14.1.3 物联网大数据的快速处理(224)
 - 14.1.4 物联网大数据的并行分析(226)
 - 14.1.5 物联网大数据处理平台的搭建(227)
 - 14.2 其他大数据的典型应用(231)
 - 14.2.1 企业级应用(231)
 - 14.2.2 社交网络大数据(232)
 - 14.2.3 医疗健康(234)
 - 14.2.4 群智感知(235)
 - 14.2.5 智能电网(235)
- 15 总结(237)
 - 15.1 大数据的研究热点及研究方向(237)
 - 15.1.1 基础理论研究(237)
 - 15.1.2 关键技术研究(238)
 - 15.1.3 应用实践研究(238)
 - 15.1.4 数据安全研究(239)
 - 15.2 展望(240)
- 参考文献(243)

《大数据浪潮》

精彩短评

1、如今人人都在讲大数据，那么什么是大数据，大数据用来干什么，怎么利用大数据，这些问题你都能在这本书里找到答案，不仅如此，这本书还介绍了许多相关拓展知识，如分布式文件系统、交互式查询语言、物联网大数据等等，不管是对于想要深入了解大数据还是初窥大数据的人，都很合适。感觉这本书包罗万象，我刚买来用了一个月的时间看完，还觉得意犹未尽，没有完全地吸收这些知识，有空的时候还会翻一翻。

《大数据浪潮》

精彩书评

- 1、这几年大数据发展迅速，但优秀的书籍始终是从国外引进的，经过翻译再出版后，技术上早已落后当前的大数据发展了。这本书是少有的国人自己出的讲大数据的好书，前半部分讲理论，后半部分讲实践，全书语言精练，值得细细阅读。
- 2、这本书介绍了最近比较火的大数据的背景 and 基础，对大数据的生成与采集，存储与分析等方面进行了详细的介绍，并且介绍了常用的两个工具Hadoop和Spark。最后给出一些实例让堵住更清楚的认识大数据，内容非常详实，讲解很通俗易懂，是一本值得推荐的好书。
- 3、这是一本终于将大数据各个环节连起来的书，除了基础的理论，还有具体的实践和代码，系统的介绍了大数据方面需要掌握的各项知识。不仅包括了理论部分，还有实践部分（实践篇），对大数据最新的编程技术都进行了介绍。作者陈敏十几年前就曾经写过《OPNET网络仿真》，是个学术大牛！

《大数据浪潮》

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:www.tushu000.com