

《零基础学大数据算法》

图书基本信息

书名：《零基础学大数据算法》

13位ISBN编号：9787121289377

出版时间：2016-7

作者：王宏志,林可

页数：268

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.tushu111.com

《零基础学大数据算法》

内容概要

《零基础学大数据算法》是通俗易懂的大数据算法教程。通篇采用师生对话的形式，旨在用通俗的语言、轻松的气氛，帮助读者理解大数据计算领域中的基础算法和思想。

《零基础学大数据算法》由背景篇、理论篇、应用篇和实践篇四部分组成。背景篇介绍大数据、算法、大数据算法等基本概念和背景；理论篇介绍解决大数据问题的亚线性算法、磁盘算法、并行算法、众包算法的基本思想和理论知识；应用篇介绍与大数据问题息息相关的数据挖掘和推荐系统的相关知识；实践篇从实际应用出发，引导读者动手操作，帮助读者通过实际程序和实验验证磁盘算法、并行算法和众包算法。

在讲解每一个大数据问题之前，《零基础学大数据算法》都会介绍大量的经典算法和基础数据结构知识，不仅可以帮助学习过数据结构与算法、算法设计与分析等课程的同学复习，同时能够让入门的“菜鸟”们，不会因为学习过经典算法而对《零基础学大数据算法》望而却步，轻松地掌握大数据算法！

书籍目录

第1篇 背景篇

第1章 何谓大数据 4

1.1 身边的大数据 4

1.2 大数据的特点和应用 6

第2章 何谓算法 8

2.1 算法的定义 8

2.2 算法的分析 .. 14

2.3 基础数据结构——线性表 .. 24

2.4 递归——以阶乘为例 28

第3章 何谓大数据算法 31

第2篇 理论篇

第4章 窥一斑而见全豹——亚线性算法 34

4.1 亚线性算法的定义 34

4.2 空间亚线性算法 35

4.2.1 水库抽样 35

4.2.2 数据流中的频繁元素 37

4.3 时间亚线性计算算法 40

4.3.1 图论基础回顾 40

4.3.2 平面图直径 45

4.3.3 最小生成树 46

4.4 时间亚线性判定算法 53

4.4.1 全0 数组的判定 53

4.4.2 数组有序的判定 55

第5章 价钱与性能的平衡——磁盘算法 58

5.1 磁盘算法概述 58

5.2 外排序 62

5.3 外存数据结构——磁盘查找树 71

5.3.1 二叉搜索树回顾 71

5.3.2 外存数据结构——B 树 78

5.3.3 高维外存查找结构——KD 树 80

5.4 表排序 83

5.5 表排序的应用 86

5.5.1 欧拉回路技术 86

5.5.2 父子关系判定 87

5.5.3 前序计数 88

5.6 时间前向处理技术 90

5.7 缩图法 98

第6章 $1+1>2$ ——并行算法 103

6.1 MapReduce 初探 103

6.2 MapReduce 算法实例 106

6.2.1 字数统计 106

6.2.2 平均数计算 108

6.2.3 单词共现矩阵计算 .111

6.3 MapReduce 进阶算法 115

6.3.1 join 操作 115

6.3.2 MapReduce 图算法概述 122

6.3.3 基于路径的图算法 125

第7章超越MapReduce的并行计算	131
7.1 MapReduce 平台的局限	131
7.2 基于图处理平台的并行算法	136
7.2.1 概述	136
7.2.2 BSP 模型下的单源最短路径	137
7.2.3 计算子图同构	141
第8章众人拾柴火焰高——众包算法	144
8.1 众包概述	144
8.1.1 众包的定义	144
8.1.2 众包应用举例	146
8.1.3 众包的特点	149
8.2 众包算法例析	152
第3篇 应用篇	
第9章大数据中有黄金——数据挖掘	158
9.1 数据挖掘概述	158
9.2 数据挖掘的分类	159
9.3 聚类算法——k-means	160
9.4 分类算法——Naive Bayes	166
第10章推荐系统	170
10.1 推荐系统概述	170
10.2 基于内容的推荐方法	173
10.3 协同过滤模型	176
第4篇 实践篇	
第11章磁盘算法实践	186
第12章并行算法实践	194
12.1 Hadoop MapReduce 实践	194
12.1.1 环境搭建	194
12.1.2 配置Hadoop	201
12.1.3 “Hello World” 程序——WordCount	203
12.1.4 Hadoop 实践案例——记录去重	213
12.1.5 Hadoop 实践案例——等值连接	216
12.1.6 多机配置	221
12.2 适于迭代并行计算的平台——Spark	224
12.2.1 Spark 初探	224
12.2.2 单词出现行计数	230
12.2.3 在Spark 上实现WordCount	236
12.2.4 在HDFS 上使用Spark	241
12.2.5 Spark 的核心操作——Transformation 和Action	244
12.2.6 Spark 实践案例——PageRank	247
第13章众包算法实践	251
13.1 认识AMT	251
13.2 成为众包工人	252

《零基础学大数据算法》

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.tushu111.com