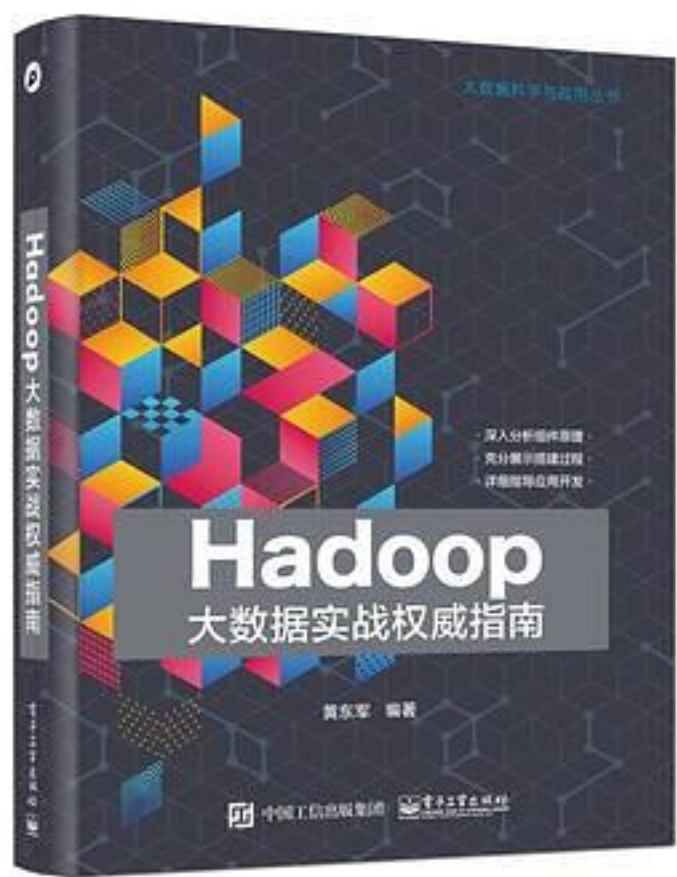


Hadoop大数据实战权威指南



[Hadoop大数据实战权威指南_下载链接1](#)

著者:黄东军

出版者:电子工业出版社

出版时间:2017-7

装帧:平装

isbn:9787121318214

大数据贵在落实！

本书是一本讲解大数据实战的图书，按照“深入分析组件原理、充分展示搭建过程、详细指导应用开发”编写。全书分为三篇，第一篇为大数据的基本概念和技术，主要介绍大数据的背景、发展及关键技术；第二篇为Hadoop大数据平台搭建与基本应用，内容

涉及Linux、HDFS、MapReduce、YARN、Hive、HBase、Sqoop、Kafka、Spark等；第三篇为大数据处理与项目开发，包括交互式数据处理、协同过滤推荐系统、销售数据分析系统，并就京东的部分销售数据应用大数据进行处理分析。

作者介绍:

目录: 第一篇 大数据的基本概念和技术

第1章 绪论 3

1.1 时代背景 3

1.1.1 全球大数据浪潮 3

1.1.2 我国的大数据国家战略 5

1.2 大数据的概念 7

1.2.1 概念 7

1.2.2 特征 8

1.3 技术支撑体系 9

1.3.1 概览 9

1.3.2 大数据采集层 9

1.3.3 大数据存储层 10

1.3.4 大数据分析（处理与服务）层 11

1.3.5 大数据应用层 11

1.3.6 垂直视图 13

1.4 大数据人才及其能力要求 14

1.4.1 首席数据官 14

1.4.2 数据科学家（数据分析师） 15

1.4.3 大数据开发工程师 16

1.4.4 大数据运维工程师 17

1.5 本章小结 17

第2章 Hadoop大数据关键技术 19

2.1 Hadoop生态系统 19

2.1.1 架构的基本理论 19

2.1.2 主要组件及其关系 21

2.2 数据采集 24

2.2.1 结构化数据采集工具 24

2.2.2 日志文件采集工具与技术 25

2.3 大数据存储技术 29

2.3.1 相关概念 29

2.3.2 分布式文件存储系统 34

2.3.3 数据库与数据仓库 38

2.4 分布式计算框架 43

2.4.1 离线计算框架 43

2.4.2 实时流计算平台 50

2.5 数据分析平台与工具 57

2.5.1 面向大数据的数据挖掘与分析工具 57

2.5.2 机器学习 61

2.6 本章小结 66

第二篇 Hadoop大数据平台搭建与基本应用

第3章 Linux操作系统与集群搭建 69

3.1 Linux操作系统 69

3.1.1 概述 69

3.1.2 特点 70

3.1.3 Linux的组成 72

3.2 Linux安装与集群搭建 75

3.2.1 安装VMware Workstation	75
3.2.2 在VMware上安装Linux (CentOS7)	79
3.3 集群的配置	91
3.3.1 设置主机名	91
3.3.2 网络设置	93
3.3.3 关闭防火墙	98
3.3.4 安装JDK	99
3.3.5 免密钥登录配置	102
3.4 Linux基本命令	105
3.5 本章小结	112
第4章 HDFS安装与基本应用	113
4.1 HDFS概述	113
4.1.1 特点	113
4.1.2 主要组件与架构	114
4.2 HDFS架构分析	114
4.2.1 数据块	114
4.2.2 NameNode	115
4.2.3 DataNode	116
4.2.4 SecondaryNameNode	117
4.2.5 数据备份	117
4.2.6 通信协议	118
4.2.7 可靠性保证	118
4.3 文件操作过程分析	119
4.3.1 读文件	119
4.3.2 写文件	120
4.3.3 删除文件	122
4.4 Hadoop HDFS安装与配置	122
4.4.1 解压Hadoop安装包	122
4.4.2 配置Hadoop环境变量	123
4.4.3 配置Yarn环境变量	124
4.4.4 配置核心组件文件	125
4.4.5 配置文件系统	125
4.4.6 配置yarn-site.xml文件	126
4.4.7 配置MapReduce计算框架文件	128
4.4.8 配置Master的slaves文件	129
4.4.9 复制Master上的Hadoop到Slave节点	129
4.5 Hadoop集群的启动	130
4.5.1 配置操作系统环境变量	130
4.5.2 创建Hadoop数据目录	131
4.5.3 格式化文件系统	132
4.5.4 启动和关闭Hadoop	133
4.5.5 验证Hadoop是否启动成功	133
4.6 Hadoop集群的基本应用	136
4.6.1 HDFS基本命令	136
4.6.2 在Hadoop集群中运行程序	139
4.7 本章小结	141
第5章 MapReduce与Yarn	143
5.1 MapReduce程序的概念	143
5.1.1 基本编程模型	143
5.1.2 计算过程分析	144
5.2 深入理解Yarn	147
5.2.1 Yarn的基本架构	147
5.2.2 Yarn的工作流程	151
5.3 在Linux平台安装Eclipse	152

- 5.3.1 Eclipse简介 153
- 5.3.2 安装并启动Eclipse 154
- 5.4 开发MapReduce程序的基本方法 155
 - 5.4.1 为Eclipse安装Hadoop插件 156
 - 5.4.2 WordCount：第一个MapReduce程序 160
- 5.5 本章小结 175
- 第6章 Hive和HBase的安装与应用 177
 - 6.1 在CentOS7下安装MySQL 177
 - 6.1.1 下载或复制MySQL安装包 177
 - 6.1.2 执行安装命令 178
 - 6.1.3 启动MySQL 179
 - 6.1.4 登录MySQL 179
 - 6.1.5 使用MySQL 181
 - 6.1.6 问题与解决办法 182
 - 6.2 Hive安装与应用 183
 - 6.2.1 下载并解压Hive安装包 183
 - 6.2.2 配置Hive 184
 - 6.2.3 启动并验证Hive 187
 - 6.2.4 Hive的基本应用 189
 - 6.3 ZooKeeper集群安装 190
 - 6.3.1 ZooKeeper简介 190
 - 6.3.2 安装ZooKeeper 191
 - 6.3.3 配置ZooKeeper 191
 - 6.3.4 启动和测试 193
 - 6.4 HBase的安装与应用 195
 - 6.4.1 解压并安装HBase 195
 - 6.4.2 配置HBase 196
 - 6.4.3 启动并验证HBase 199
 - 6.4.4 HBase的基本应用 200
 - 6.4.5 应用HBase中常见问题及其解决办法 203
 - 6.5 本章小结 204
- 第7章 Sqoop和Kafka的安装与应用 205
 - 7.1 安装部署Sqoop 205
 - 7.1.1 下载或复制Sqoop安装包 205
 - 7.1.2 解压并安装Sqoop 206
 - 7.1.3 配置Sqoop 206
 - 7.1.4 启动并验证Sqoop 208
 - 7.1.5 测试Sqoop与MySQL的连接 209
 - 7.2 安装部署Kafka集群 211
 - 7.2.1 下载或复制Kafka安装包 211
 - 7.2.2 解压缩Kafka安装包 211
 - 7.2.3 配置Kafka集群 211
 - 7.2.4 Kafka的初步应用 213
 - 7.3 本章小结 218
- 第8章 Spark集群安装与开发环境配置 219
 - 8.1 深入理解Spark 219
 - 8.1.1 Spark系统架构 219
 - 8.1.2 关键概念 221
 - 8.2 安装与配置Scala 224
 - 8.2.1 下载Scala安装包 225
 - 8.2.2 安装Scala 225
 - 8.2.3 启动并应用Scala 226
 - 8.3 Spark集群的安装与配置 226
 - 8.3.1 安装模式 226

8.3.2 Spark的安装	227
8.3.3 启动并验证Spark	230
8.3.4 几点说明	234
8.4 开发环境安装与配置	236
8.4.1 IDEA简介	236
8.4.2 IDEA的安装	236
8.4.3 IDEA的配置	238
8.5 本章小结	243
第9章 Spark应用基础	245
9.1 Spark程序的运行模式	245
9.1.1 Spark on Yarn-cluster	245
9.1.2 Spark on Yarn-client	246
9.2 Spark应用设计	247
9.2.1 分布式估算圆周率	248
9.2.2 基于Spark MLlib的贷款风险预测	265
9.3 本章小结	285
第三篇 数据处理与项目开发术	
第10章 交互式数据处理	289
10.1 数据预处理	289
10.1.1 查看数据	289
10.1.2 数据扩展	291
10.1.3 数据过滤	292
10.1.4 数据上传	293
10.2 创建数据仓库	294
10.2.1 创建Hive数据仓库的基本命令	294
10.2.2 创建Hive分区表	296
10.3 数据分析	299
10.3.1 基本统计	299
10.3.2 用户行为分析	301
10.3.3 实时数据	303
10.4 本章小结	304
第11章 协同过滤推荐系统	305
11.1 推荐算法概述	305
11.1.1 基于人口统计学的推荐	305
11.1.2 基于内容的推荐	306
11.1.3 协同过滤推荐	307
11.2 协同过滤推荐算法分析	308
11.2.1 基于用户的协同过滤推荐	308
11.2.2 基于物品的协同过滤推荐	310
11.3 Spark MLlib推荐算法应用	312
11.3.1 ALS算法原理	312
11.3.2 ALS的应用设计	315
11.4 本章小结	329
第12章 销售数据分析系统	331
12.1 数据采集	331
12.1.1 在Windows下安装JDK	331
12.1.2 在Windows下安装Eclipse	334
12.1.3 将WebCollector项目导入Eclipse	335
12.1.4 在Windows下安装MySQL	336
12.1.5 连接JDBC	339
12.1.6 运行爬虫程序	340
12.2 在HBase集群上准备数据	342
12.2.1 将数据导入到MySQL	342
12.2.2 将MySQL表中的数据导入到HBase表中	344

- 12.3 安装Phoenix中间件 347
 - 12.3.1 Phoenix架构 347
 - 12.3.2 解压安装Phoenix 348
 - 12.3.3 Phoenix环境配置 349
 - 12.3.4 使用Phoenix 350
- 12.4 基于Web的前端开发 353
 - 12.4.1 将Web前端项目导入Eclipse 353
 - 12.4.2 安装Tomcat 355
 - 12.4.3 在Eclipse中配置Tomcat 355
 - 12.4.4 在Web浏览器中查看执行结果 359
- 12.5 本章小结 361
- • • • • (收起)

[Hadoop大数据实战权威指南_下载链接1](#)

标签

大数据

hadoop

bd

评论

入门科普……

一本单纯指引如何搭建hadoop平台的说明书

[Hadoop大数据实战权威指南_下载链接1](#)

书评

[Hadoop大数据实战权威指南 下载链接1](#)