

Hadoop大数据挖掘从入门到进阶实战（视频教学版）



[Hadoop大数据挖掘从入门到进阶实战（视频教学版）_下载链接1](#)

著者:邓杰

出版者:机械工业出版社

出版时间:2018-6

装帧:平装

isbn:9787111600107

本书采用“理论+实战”的形式编写，全面介绍了Hadoop大数据挖掘的相关知识。本书秉承循序渐进、易于理解、学以致用和便于查询的讲授理念，讲解时结合了大量实例和作者多年积累的一线开发经验。本书作者拥有丰富的视频制作与在线教学经验，曾经与极客学院合作开设过在线视频教学课程。为了帮助读者高效、直观地学习本书内容，作者特意为本书录制了配套教学视频，这些教学视频和本书配套源代码文件读者都可以免费获取。

本书共分为13章，涵盖的主要内容有：集群及开发环境搭建；快速构建一个Hadoop项目并线上运行；Hadoop套件实战；Hive编程——使用SQL提交MapReduce任务到Hadoop集群；游戏玩家的用户行为分析——特征提取；Hadoop平台管理与维护；Hadoop异常处理解决方案；初识Hadoop核心源码；Hadoop通信机制和内部协议；Hadoop分布式文件系统剖析；ELK实战案例——游戏应用实时日志分析平台；Kafka实战案例——实时处理游戏用户数据；Hadoop拓展——Kafka剖析。

本书通俗易懂，案例丰富，实用性强，不但适合初学者系统学习Hadoop的各种基础语法和开发技巧，而且也适合有开发经验的程序员进阶提高。另外，本书还适合社会培训

机构和相关院校作为教材或者教学参考书。

作者介绍:

邓杰

博客园资深博主，资深大数据全栈开发者，极客学院大数据讲师，开源爱好者。善于开发大数据监控系统辅助日常工作，提升工作效率。主导开发了大数据自助类平台系统。开发并在GitHub上发布了Kafka系统监控管理工具Kafka Eagle，深受业内开发者的赞誉。作为极客学院特邀讲师，制作了多个技术视频，讲授Hadoop和Kafka等相关技术课程，广受学员好评。

目录: 前言

第1章 集群及开发环境搭建 1

1.1 环境准备 1

1.1.1 基础软件下载 1

1.1.2 准备Linux操作系统 2

1.2 安装Hadoop 4

1.2.1 基础环境配置 4

1.2.2 Zookeeper部署 7

1.2.3 Hadoop部署 9

1.2.4 效果验证 21

1.2.5 集群架构详解 24

1.3 Hadoop版Hello World 25

1.3.1 Hadoop Shell介绍 25

1.3.2 WordCount初体验 27

1.4 开发环境 28

1.4.1 搭建本地开发环境 28

1.4.2 运行及调试预览 31

1.5 小结 34

第2章 实战：快速构建一个Hadoop项目并线上运行 35

2.1 构建一个简单的项目工程 35

2.1.1 构建Java Project结构工程 35

2.1.2 构建Maven结构工程 36

2.2 操作分布式文件系统（HDFS） 39

2.2.1 基本的应用接口操作 39

2.2.2 在高可用平台上的使用方法 42

2.3 利用IDE提交MapReduce作业 43

2.3.1 在单点上的操作 43

2.3.2 在高可用平台上的操作 46

2.4 编译应用程序并打包 51

2.4.1 编译Java Project工程并打包 51

2.4.2 编译Maven工程并打包 55

2.5 部署与调度 58

2.5.1 部署应用 58

2.5.2 调度任务 59

2.6 小结 60

第3章 Hadoop套件实战 61

3.1 Sqoop——数据传输工具 61

3.1.1 背景概述 61

3.1.2 安装及基本使用 62

3.1.3 实战：在关系型数据库与分布式文件系统之间传输数据 64

3.2 Flume——日志收集工具 66

3.2.1 背景概述	67
3.2.2 安装与基本使用	67
3.2.3 实战：收集系统日志并上传到分布式文件系统（HDFS）上	72
3.3 HBase——分布式数据库	74
3.3.1 背景概述	74
3.3.2 存储架构介绍	75
3.3.3 安装与基本使用	75
3.3.4 实战：对HBase业务表进行增、删、改、查操作	79
3.4 Zeppelin——数据集分析工具	85
3.4.1 背景概述	85
3.4.2 安装与基本使用	85
3.4.3 实战：使用解释器操作不同的数据处理引擎	88
3.5 Drill——低延时SQL查询引擎	92
3.5.1 背景概述	93
3.5.2 安装与基本使用	93
3.5.3 实战：对分布式文件系统（HDFS）使用SQL进行查询	95
3.5.4 实战：使用SQL查询HBase数据库	99
3.5.5 实战：对数据仓库（Hive）使用类实时统计、查询操作	101
3.6 Spark——实时流数据计算	104
3.6.1 背景概述	104
3.6.2 安装部署及使用	105
3.6.3 实战：对接Kafka消息数据，消费、计算及落地	108
3.7 小结	114
第4章 Hive编程——使用SQL提交MapReduce任务到Hadoop集群	115
4.1 环境准备与Hive初识	115
4.1.1 背景介绍	115
4.1.2 基础环境准备	116
4.1.3 Hive结构初识	116
4.1.4 Hive与关系型数据库（RDBMS）	118
4.2 安装与配置Hive	118
4.2.1 Hive集群基础架构	119
4.2.2 利用HAProxy实现Hive Server负载均衡	120
4.2.3 安装分布式Hive集群	123
4.3 可编程方式	126
4.3.1 数据类型	126
4.3.2 存储格式	128
4.3.3 基础命令	129
4.3.4 Java编程语言操作数据仓库（Hive）	131
4.3.5 实践Hive Streaming	134
4.4 运维和监控	138
4.4.1 基础命令	138
4.4.2 监控工具Hive Cube	140
4.5 小结	143
第5章 游戏玩家的用户行为分析——特征提取	144
5.1 项目应用概述	144
5.1.1 场景介绍	144
5.1.2 平台架构与数据采集	145
5.1.3 准备系统环境和软件	147
5.2 分析与设计	148
5.2.1 整体分析	148
5.2.2 指标与数据源分析	149
5.2.3 整体设计	151
5.3 技术选型	153
5.3.1 套件选取简述	154

5.3.2 套件使用简述 154

5.4 编码实践 157

5.4.1 实现代码 157

5.4.2 统计结果处理 163

5.4.3 应用调度 169

5.5 小结 174

第6章 Hadoop平台管理与维护 175

6.1 Hadoop分布式文件系统 (HDFS) 175

6.1.1 HDFS特性 175

6.1.2 基础命令详解 176

6.1.3 解读NameNode Standby 179

6.2 Hadoop平台监控 182

6.2.1 Hadoop日志 183

6.2.2 常用分布式监控工具 187

6.3 平台维护 196

6.3.1 安全模式 196

6.3.2 节点管理 198

6.3.3 HDFS快照 200

6.4 小结 203

第7章 Hadoop异常处理解决方案 204

7.1 定位异常 204

7.1.1 跟踪日志 204

7.1.2 分析异常信息 208

7.1.3 阅读开发业务代码 209

7.2 解决问题的方式 210

7.2.1 搜索关键字 211

7.2.2 查看Hadoop JIRA 212

7.2.3 阅读相关源码 213

7.3 实战案例分析 216

7.3.1 案例分析1：启动HBase失败 216

7.3.2 案例分析2：HBase表查询失败 219

7.3.3 案例分析3：Spark的临时数据不自动清理 222

7.4 小结 223

第8章 初识Hadoop核心源码 224

8.1 基础准备与源码编译 224

8.1.1 准备环境 224

8.1.2 加载源码 228

8.1.3 编译源码 230

8.2 初识Hadoop 2 233

8.2.1 Hadoop的起源 233

8.2.2 Hadoop 2源码结构图 234

8.2.3 Hadoop模块包 235

8.3 MapReduce框架剖析 236

8.3.1 第一代MapReduce框架 236

8.3.2 第二代MapReduce框架 238

8.3.3 两代MapReduce框架的区别 239

8.3.4 第二代MapReduce框架的重构思路 240

8.4 序列化 241

8.4.1 序列化的由来 242

8.4.2 Hadoop序列化 243

8.4.3 Writable实现类 245

8.5 小结 247

第9章 Hadoop通信机制和内部协议 248

9.1 Hadoop RPC概述 248

9.1.1 通信模型	248
9.1.2 Hadoop RPC特点	250
9.2 Hadoop RPC的分析与使用	251
9.2.1 基础结构	251
9.2.2 使用示例	257
9.2.3 其他开源RPC框架	264
9.3 通信协议	266
9.3.1 MapReduce通信协议	266
9.3.2 RPC协议的实现	273
9.4 小结	277
第10章 Hadoop分布式文件系统剖析	278
10.1 HDFS介绍	278
10.1.1 HDFS概述	278
10.1.2 其他分布式文件系统	282
10.2 HDFS架构剖析	283
10.2.1 设计特点	283
10.2.2 命令空间和节点	285
10.2.3 数据备份剖析	289
10.3 数据迁移实战	292
10.3.1 HDFS跨集群迁移	292
10.3.2 HBase集群跨集群数据迁移	297
10.4 小结	301
第11章 ELK实战案例——游戏应用实时日志分析平台	302
11.1 Logstash——实时日志采集、分析和传输	302
11.1.1 Logstash介绍	302
11.1.2 Logstash安装	306
11.1.3 实战操作	308
11.2 Elasticsearch——分布式存储及搜索引擎	309
11.2.1 应用场景	309
11.2.2 基本概念	310
11.2.3 集群部署	312
11.2.4 实战操作	317
11.3 Kibana——可视化管理系统	323
11.3.1 Kibana特性	324
11.3.2 Kibana安装	324
11.3.3 实战操作	328
11.4 实时日志分析平台案例	331
11.4.1 案例概述	331
11.4.2 平台体系架构与剖析	332
11.4.3 实战操作	334
11.5 小结	339
第12章 Kafka实战案例——实时处理游戏用户数据	340
12.1 应用概述	340
12.1.1 Kafka回顾	340
12.1.2 项目简述	347
12.1.3 Kafka工程准备	348
12.2 项目的分析与设计	349
12.2.1 项目背景和价值概述	349
12.2.2 生产模块	350
12.2.3 消费模块	352
12.2.4 体系架构	352
12.3 项目的编码实践	354
12.3.1 生产模块	354
12.3.2 消费模块	356

12.3.3 数据持久化 362
12.3.4 应用调度 364
12.4 小结 369
第13章 Hadoop拓展——Kafka剖析 370
13.1 Kafka开发与维护 370
13.1.1 接口 370
13.1.2 新旧API编写 372
13.1.3 Kafka常用命令 380
13.2 运维监控 383
13.2.1 监控指标 384
13.2.2 Kafka开源监控工具——Kafka Eagle 384
13.3 Kafka源码分析 391
13.3.1 源码工程环境构建 391
13.3.2 分布式选举算法剖析 394
13.3.3 Kafka Offset解读 398
13.3.4 存储机制和副本 398
13.4 小结 402
• • • • • (收起)

[Hadoop大数据挖掘从入门到进阶实战（视频教学版）](#) [下载链接1](#)

标签

hadoop

评论

[Hadoop大数据挖掘从入门到进阶实战（视频教学版）](#) [下载链接1](#)

书评

[Hadoop大数据挖掘从入门到进阶实战（视频教学版）](#) [下载链接1](#)