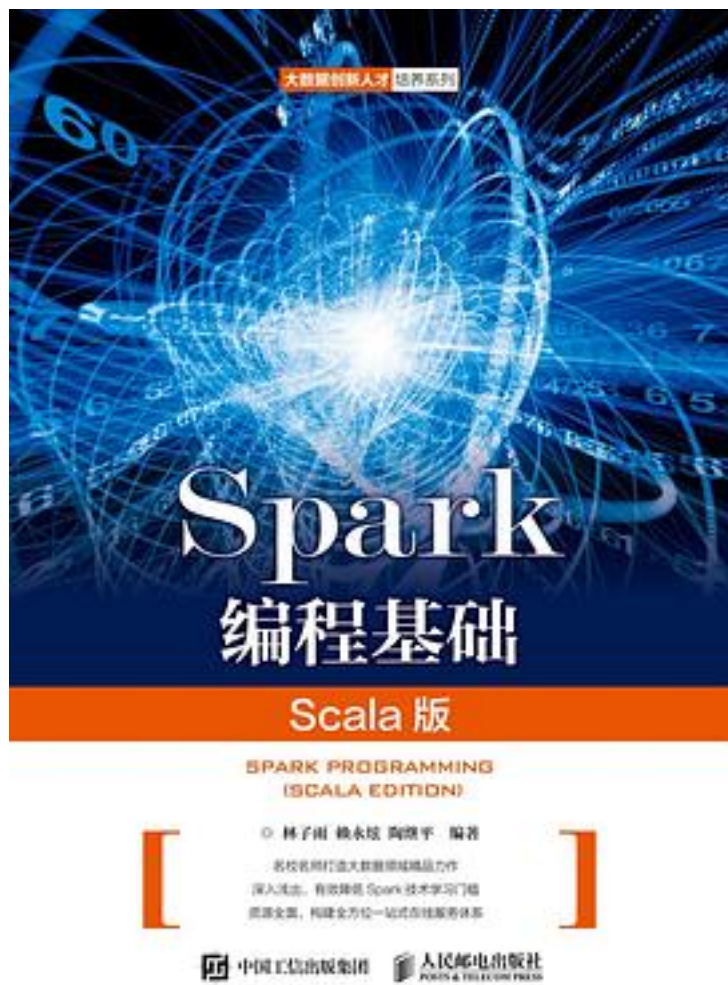


Spark编程基础（Scala版）



[Spark编程基础（Scala版）_下载链接1](#)

著者:

出版者:

出版时间:2018-7

装帧:平装

isbn:9787115488169

作者介绍:

目录: 第1章 大数据技术概述	1
1.1 大数据的概念与关键技术	2
1.1.1 大数据的概念	2
1.1.2 大数据关键技术	2
1.2 代表性大数据技术	4
1.2.1 Hadoop	4
1.2.2 Spark	8
1.2.3 Flink	10
1.2.4 Beam	11
1.3 编程语言的选择	12
1.4 在线资源	13
1.5 本章小结	14
1.6 习题	14
实验1 Linux系统的安装和常用命令	15
一、实验目的	15
二、实验平台	15
三、实验内容和要求	15
四、实验报告	16
第2章 Scala语言基础	17
2.1 Scala语言概述	18
2.1.1 计算机的缘起	18
2.1.2 编程范式	19
2.1.3 Scala简介	20
2.1.4 Scala的安装	21
2.1.5 HelloWorld	21
2.2 Scala基础知识	23
2.2.1 基本数据类型和变量	23
2.2.2 输入/输出	26
2.2.3 控制结构	28
2.2.4 数据结构	31
2.3 面向对象编程基础	37
2.3.1 类	37
2.3.2 对象	42
2.3.3 继承	47
2.3.4 参数化类型	50
2.3.5 特质	52
2.3.6 模式匹配	55
2.3.7 包	58
2.4 函数式编程基础	59
2.4.1 函数的定义与使用	60
2.4.2 高阶函数	61
2.4.3 闭包	62
2.4.4 偏应用函数和Curry化	62
2.4.5 针对容器的操作	64
2.4.6 函数式编程实例	69
2.5 本章小结	70
2.6 习题	70
实验2 Scala编程初级实践	71
一、实验目的	71
二、实验平台	71
三、实验内容和要求	72

四、实验报告	75
第3章 Spark的设计与运行原理	76
3.1 概述	77
3.2 Spark生态系统	78
3.3 Spark运行架构	79
3.3.1 基本概念	79
3.3.2 架构设计	80
3.3.3 Spark运行基本流程	81
3.3.4 RDD的设计与运行原理	82
3.4 Spark的部署方式	91
3.5 本章小结	92
3.6 习题	93
第4章 Spark环境搭建和使用方法	94
4.1 安装Spark	95
4.1.1 基础环境	95
4.1.2 下载安装文件	95
4.1.3 配置相关文件	96
4.1.4 Spark和Hadoop的交互	97
4.2 在spark-shell中运行代码	97
4.2.1 spark-shell命令	98
4.2.2 启动spark-shell	99
4.3 开发Spark独立应用程序	99
4.3.1 安装编译打包工具	100
4.3.2 编写Spark应用程序代码	101
4.3.3 编译打包	101
4.3.4 通过spark-submit运行程序	104
4.4 Spark集群环境搭建	104
4.4.1 集群概况	105
4.4.2 搭建Hadoop集群	105
4.4.3 在集群中安装Spark	106
4.4.4 配置环境变量	106
4.4.5 Spark的配置	106
4.4.6 启动Spark集群	107
4.4.7 关闭Spark集群	107
4.5 在集群上运行Spark应用程序	108
4.5.1 启动Spark集群	108
4.5.2 采用独立集群管理器	108
4.5.3 采用Hadoop YARN管理器	109
4.6 本章小结	110
4.7 习题	111
实验3 Spark和Hadoop的安装	111
一、实验目的	111
二、实验平台	111
三、实验内容和要求	111
四、实验报告	112
第5章 RDD编程	113
5.1 RDD编程基础	114
5.1.1 RDD创建	114
5.1.2 RDD操作	115
5.1.3 持久化	121
5.1.4 分区	122
5.1.5 一个综合实例	126
5.2 键值对RDD	128
5.2.1 键值对RDD的创建	128

5.2.2 常用的键值对转换操作	129
5.2.3 一个综合实例	133
5.3 数据读写	134
5.3.1 文件数据读写	135
5.3.2 读写HBase数据	137
5.4 综合实例	141
5.4.1 求TOP值	141
5.4.2 文件排序	143
5.4.3 二次排序	144
5.5 本章小结	146
实验4 RDD编程初级实践	146
一、实验目的	146
二、实验平台	146
三、实验内容和要求	146
四、实验报告	148
第6章 Spark SQL	149
6.1 Spark SQL简介	150
6.1.1 从Shark说起	150
6.1.2 Spark SQL架构	151
6.1.3 为什么推出Spark SQL	152
6.2 DataFrame概述	152
6.3 DataFrame的创建	153
6.4 DataFrame的保存	154
6.5 DataFrame的常用操作	155
6.6 从RDD转换得到DataFrame	156
6.6.1 利用反射机制推断RDD模式	157
6.6.2 使用编程方式定义RDD模式	158
6.7 使用Spark SQL读写数据库	160
6.7.1 通过JDBC连接数据库	160
6.7.2 连接Hive读写数据	162
6.8 本章小结	166
6.9 习题	166
实验5 Spark SQL编程初级实践	167
一、实验目的	167
二、实验平台	167
三、实验内容和要求	167
四、实验报告	168
第7章 Spark Streaming	169
7.1 流计算概述	170
7.1.1 静态数据和流数据	170
7.1.2 批量计算和实时计算	171
7.1.3 流计算概念	171
7.1.4 流计算框架	172
7.1.5 流计算处理流程	173
7.2 Spark Streaming	174
7.2.1 Spark Streaming设计	174
7.2.2 Spark Streaming与Storm的对比	175
7.2.3 从“Hadoop+Storm”架构转向Spark架构	176
7.3 DStream操作概述	177
7.3.1 Spark Streaming工作机制	177
7.3.2 编写Spark Streaming程序的基本步骤	178
7.3.3 创建StreamingContext对象	178
7.4 基本输入源	179
7.4.1 文件流	179

7.4.2 套接字流	181
7.4.3 RDD队列流	186
7.5 高级数据源	187
7.5.1 Kafka简介	188
7.5.2 Kafka准备工作	188
7.5.3 Spark准备工作	189
7.5.4 编写Spark Streaming程序使用Kafka数据源	190
7.6 转换操作	194
7.6.1 DStream无状态转换操作	194
7.6.2 DStream有状态转换操作	195
7.7 输出操作	199
7.7.1 把DStream输出到文本文件中	199
7.7.2 把DStream写入到关系数据库中	200
7.8 本章小结	202
7.9 习题	202
实验6 Spark Streaming编程初级实践	203
一、实验目的	203
二、实验平台	203
三、实验内容和要求	203
四、实验报告	204
第8章 Spark MLlib	205
8.1 基于大数据的机器学习	206
8.2 机器学习库MLlib概述	207
8.3 基本数据类型	208
8.3.1 本地向量	208
8.3.2 标注点	208
8.3.3 本地矩阵	209
8.4 机器学习流水线	210
8.4.1 流水线的概念	210
8.4.2 流水线工作过程	211
8.5 特征提取、转换和选择	212
8.5.1 特征提取	213
8.5.2 特征转换	215
8.5.3 特征选择	220
8.5.4 局部敏感哈希	221
8.6 分类算法	222
8.6.1 逻辑斯蒂回归分类器	222
8.6.2 决策树分类器	226
8.7 聚类算法	229
8.7.1 K-Means聚类算法	230
8.7.2 GMM聚类算法	232
8.8 协同过滤算法	234
8.8.1 推荐算法的原理	235
8.8.2 ALS算法	235
8.9 模型选择和超参数调整	239
8.9.1 模型选择工具	239
8.9.2 用交叉验证选择模型	240
8.10 本章小结	242
8.11 习题	242
实验7 Spark机器学习库MLlib编程实践	243
一、实验目的	243
二、实验平台	243
三、实验内容和要求	243
四、实验报告	244

参考文献 245

• • • • • ([收起](#))

[Spark编程基础（Scala版）_下载链接1](#)

标签

大数据

scala

Spark

spark

Hadoop

计算机科学

教材

评论

spark的入门，从理论到架构到实现都有很详细的讲解，还有网站博客可以参考，很建议用这一个系列来入门了。

作为入门书还是比较清晰明了的，终于区分了rdd编程，spark sql编程和streaming

rdd还不错，其他的有点简略

还是很不错的，深入全面易上手。适合像我这种动手能力稀烂又强迫症一样想了解知识全貌并懂得技术原理的数学系学生。

非常适合入门spark，对scala介绍短短一张介绍的很清晰，对spark的各个组件要求介绍的很全，我自己也明确了那一块能着重的看

[Spark编程基础（Scala版）_下载链接1](#)

书评

[Spark编程基础（Scala版）_下载链接1](#)