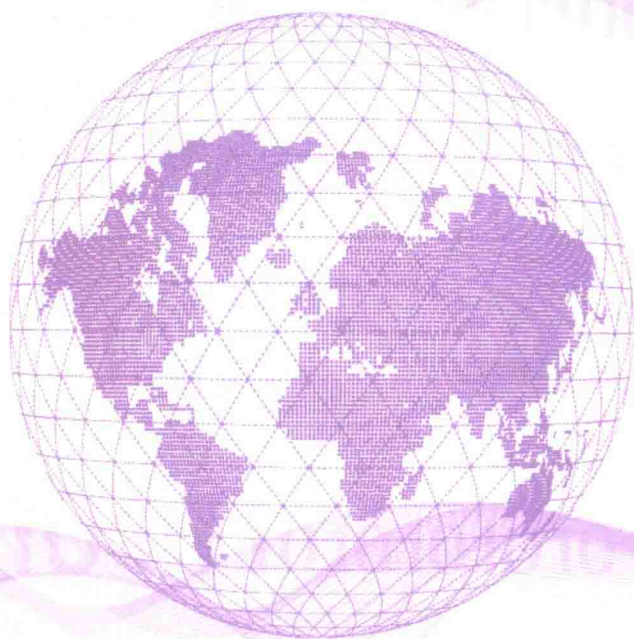


NITE 国家软件与集成电路公共服务平台信息技术紧缺人才培养工程指定教材



大数据技术与应用丛书

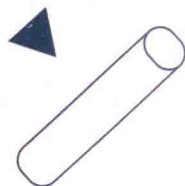
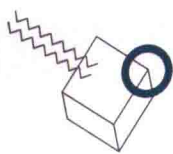
有问题，就找问答精灵！



# Spark

## 大数据分析实战

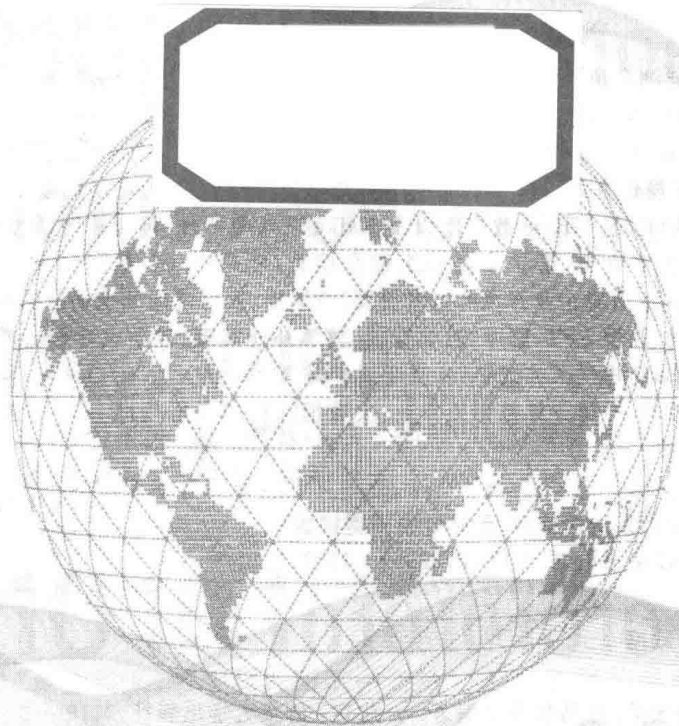
黑马程序员 / 编著



清华大学出版社



NITE 国家软件与集成电路公共服务平台信息技术紧缺人才培养工程指定教材

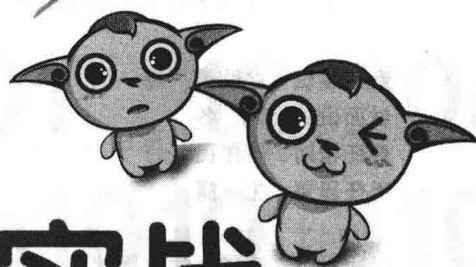


大数据技术与应用丛书

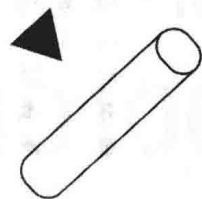
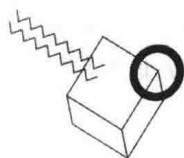
有问题，就找问答精灵！

# Spark

## 大数据分析实战



黑马程序员 / 编著



清华大学出版社  
北京

## 内 容 简 介

本书从初学者角度详细介绍了 Spark 应用程序体系架构的核心技术,全书共 9 章。第 1 章详细介绍开发 Spark 框架的 Scala 编程语言;第 2~4、7~8 章主要讲解 Spark 核心基础、SparkRDD 弹性分布式数据集、Spark SQL 处理结构化数据、Spark Streaming 实时计算框架、Spark MLlib 机器学习库,并包含了搭建 Spark 集群、Spark 集群的操作方式、利用 Spark 解决大数据工作中遇到的基本问题。第 5~6 章主要讲解大数据环境中常见的辅助系统, HBase 数据库以及 Kafka 流处理平台,包含辅助系统的搭建方式、使用方法以及相关底层实现的基本原理;第 9 章是一个综合项目,利用 Spark 框架开发流式计算系统。掌握 Spark 相关技术,能够很好地适应企业开发的技术需要,为离线、实时数据处理平台的开发奠定基础。

本书附有配套源代码、教学 PPT、题库、教学视频、教学补充案例、教学设计等资源。为了帮助初学者更好地学习本书中的内容,还提供了在线答疑,欢迎读者关注。

本书可作为高等院校本、专科计算机相关专业,大数据课程的专用教材,是一本适合广大计算机编程爱好者的优秀读物。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

### 图书在版编目(CIP)数据

Spark 大数据分析实战/黑马程序员编著. —北京:清华大学出版社,2019  
(大数据技术与应用丛书)  
ISBN 978-7-302-53432-7

I. ①S… II. ①黑… III. ①数据处理软件 IV. ①TP274

中国版本图书馆 CIP 数据核字(2019)第 166166 号

责任编辑:袁勤勇 杨 枫

封面设计:韩 冬

责任校对:胡伟民

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, [c-service@tup.tsinghua.edu.cn](mailto:c-service@tup.tsinghua.edu.cn)

质量反馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

课件下载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:北京富博印刷有限公司

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:15.25 字 数:378 千字

版 次:2019 年 9 月第 1 版 印 次:2019 年 9 月第 1 次印刷

定 价:49.00 元

产品编号:083855-01

# 序 言

江苏传智播客教育科技有限公司(简称传智播客)是一家致力于培养高素质软件开发人才的科技公司,“黑马程序员”是传智播客旗下高端 IT 教育品牌。

“黑马程序员”的学员多为大学毕业后,想从事 IT 行业,但各方面条件还不成熟的年轻人。“黑马程序员”的学员筛选制度非常严格,包括了严格的技术测试、自学能力测试,还包括性格测试、压力测试、品德测试等。百里挑一的残酷筛选制度确保学员质量,并降低企业的用人风险。

自“黑马程序员”成立以来,教学研发团队一直致力于打造精品课程资源,不断在产、学、研三个层面创新自己的执教理念与教学方针,并集中“黑马程序员”的优势力量,有针对性地出版了计算机系列教材 80 多种,制作教学视频数十套,发表各类技术文章数百篇。

“黑马程序员”不仅斥资研发 IT 系列教材,还为高校师生提供以下配套学习资源与服务。

## 为大学生提供的配套服务

1. 请登录“高校学习平台”<http://yx.ityxb.com>,免费获取海量学习资源,帮助高校学生解决学习问题。

2. 针对高校学生在学习过程中存在的压力等问题,我们还面向大学生量身打造了“IT 技术女神”——“播妞学姐”,可提供教材配套源码和习题答案,以及更多 IT 学习资源,同学们快来关注“播妞学姐”的微信公众号:boniu1024。



“播妞学姐”微信公众号

## 为教师提供的配套服务

针对高校教学,“黑马程序员”为 IT 系列教材精心设计了“教案+授课资源+考试系统+

题库+教学辅助案例”的系列教学资源,高校老师请登录“高校教辅平台”<http://yx.ityxb.com> 或关注码大牛老师微信/QQ:2011168841,获取教材配套资源,也可以扫描下方二维码,加入专为 IT 教师打造的师资服务平台——“教学好助手”,获取最新教师教学辅助资源的相关动态。



# 前言

## 为什么要学习本书

Apache Spark 是用于大规模数据处理的统一分析引擎,具有高效性、易用性、通用性、兼容性四大特性,并且在 Spark 生态体系中,包含了离线数据、流式数据、图计算、机器学习、NoSQL 查询等多个方面的解决方案,深受广大大数据工程师及算法工程师的喜爱。对于想从事大数据行业的开发人员来说,学好 Spark 尤为重要。

Spark 技术功能强大,涉及知识面较广,零基础的同学很难踏入 Spark 体系架构之中,因此本书采用理论和案例相结合的编写方式,用通俗易懂的语言讲解复杂、难以理解的原理,每章都包含多个案例,让读者学以致用。

## 关于本书

作为大数据技术 Spark 的入门教程,最重要且最难的一件事情就是将一些复杂、难以理解的问题简单化,让初学者能够轻松理解并快速掌握。本教材对每个知识点都进行了深入分析,并针对每个知识点精心设计了相关案例,然后模拟这些知识点在实际工作中的运用,真正做到了知识的由浅入深、由易到难。

本书共分为 9 章,接下来分别对每个章节进行简单的介绍,具体如下。

- 第 1 章主要讲解什么是 Scala 以及 Scala 编程相关知识。通过本章学习,读者应掌握 Scala 环境的安装配置,熟悉 Scala 语法规则,并实现使用 Scala 语言编写自己的第一个程序。
- 第 2 章主要介绍什么是 Spark,以及搭建 Spark 集群的方式,并通过 Spark Shell 学习 Spark 的基本操作方法。通过本章学习,读者应能独立搭建 Spark 集群,同时对 Spark 系统的基础操作和基本原理有初步了解。
- 第 3 章主要介绍什么是 Spark RDD、RDD 的处理过程以及操作 RDD 的方式。通过本章的学习,读者可以了解 RDD 处理数据核心思想,并且能够使用 RDD 编程解决实际问题。
- 第 4 章主要介绍 Spark SQL 的数据模型 DataFrame 和 Dataset,它是一个由多个列组成的结构化的分布式数据集合,类似于关系数据库中的表概念。通过本章的学习,读者应能够掌握利用 Spark SQL 操作 MySQL 和 Hive 两种常见数据源。
- 第 5 章主要介绍 HBase 分布式数据库的数据模型以及操作方式。通过本章学习,读者能够掌握部署 HBase 集群的方法,了解 HBase 存储数据的架构原理,并且能够使用 HBase 分布式数据库解决实际业务问题。

- 第 6 章主要介绍 Kafka 流处理平台, Kafka 是流式计算系统中常见的辅助工具, 通过 Kafka 工作原理的学习, 读者能够了解 Kafka 集群整体架构中各个组件的功能, 以及 Kafka 写入数据和消费数据的底层原理及操作方式。通过本章学习, 读者能够掌握部署 Kafka 集群的方法, 并能够通过执行命令和 API 方式操作 Kafka。
- 第 7 章主要介绍 Spark Streaming 的相关知识, Spark Streaming 是 Spark 生态系统中实现流式计算功能的重要组件。通过本章 Spark Streaming 案例式讲解, 读者能够掌握 Spark Streaming 程序的开发步骤, 及与 Kafka 整合使用的方法。
- 第 8 章主要介绍 Spark MLlib, 它是 Spark 提供的机器学习库, 其中整合了统计、分类、回归、过滤等主流的机器学习算法和丰富的 API, 降低用户使用机器学习的难度。通过本章学习, 读者能够了解利用 Spark 完成机器学习的方式, 即获取数据集, 调用训练模型算法得出模型, 通过模型分析当前数据。
- 第 9 章主要介绍利用 Spark 构建实时交易数据统计案例的开发流程。通过本章学习, 读者能够了解实时计算项目的基本架构模型, 以及本项目统计商品成交额的需求实现方式。

## 致谢

本书的编写和整理工作由传智播客教育科技有限公司完成, 主要参与人员有吕春林、高美云、翟振方、文燕等, 全体人员在这近一年的编写过程中付出了许多辛勤的汗水。除此之外, 还有传智播客 600 多名学员参与到了教材的试读工作中, 他们站在初学者的角度对教材提出了许多宝贵的修改意见, 在此一并表示衷心的感谢。

## 意见反馈

尽管我们付出了最大的努力, 但书中难免会有欠妥之处, 欢迎各界专家和读者朋友提出宝贵意见, 我们将不胜感激。您在阅读本书时, 如果发现任何问题或有不认同之处可以通过电子邮件与我们取得联系。

请发送电子邮件至 [itcast\\_book@vip.sina.com](mailto:itcast_book@vip.sina.com)。

黑马程序员

2019 年 7 月

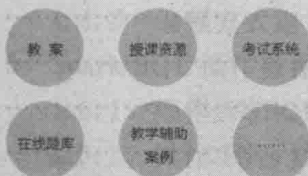
# 目 录

第 1 章 Scala 语言基础 .....	1
1.1 初识 Scala .....	1
1.1.1 Scala 概述 .....	1
1.1.2 Scala 的下载安装 .....	2
1.1.3 在 IDEA 开发工具中下载安装 Scala 插件 .....	4
1.1.4 开发第一个 Scala 程序 .....	6
1.2 Scala 的基础语法 .....	9
1.2.1 声明值和变量 .....	9
1.2.2 数据类型 .....	9
1.2.3 算术和操作符重载 .....	10
1.2.4 控制结构语句 .....	10
1.2.5 方法和函数 .....	13
1.3 Scala 的数据结构 .....	15
1.3.1 数组 .....	15
1.3.2 元组 .....	18
1.3.3 集合 .....	19
1.4 Scala 面向对象的特性 .....	24
1.4.1 类与对象 .....	24
1.4.2 继承 .....	25
1.4.3 单例对象和伴生对象 .....	26
1.4.4 特质 .....	28
1.5 Scala 的模式匹配与样例类 .....	29
1.5.1 模式匹配 .....	30
1.5.2 样例类 .....	31
1.6 本章小结 .....	31
1.7 课后习题 .....	32
第 2 章 Spark 基础 .....	33
2.1 初识 Spark .....	33

专属于教师及学生的在线教育平台  
<http://yx.ityxb.com/>

让 IT 教学更简单

教师获取教材配套资源



添加微信/QQ  
2011168841

让 IT 学习更有效

学生获取配套源码

关注微信公众号“播妞学姐”  
获取教材配套源码



专属大学生的圈子

2.1.1	Spark 概述	33
2.1.2	Spark 的特点	34
2.1.3	Spark 应用场景	35
2.1.4	Spark 与 Hadoop 对比	36
2.2	搭建 Spark 开发环境	37
2.2.1	环境准备	37
2.2.2	Spark 的部署方式	37
2.2.3	Spark 集群安装部署	38
2.2.4	Spark HA 集群部署	41
2.3	Spark 运行架构与原理	45
2.3.1	基本概念	45
2.3.2	Spark 集群运行架构	45
2.3.3	Spark 运行基本流程	46
2.4	体验第一个 Spark 程序	47
2.5	启动 Spark-Shell	49
2.5.1	运行 Spark-Shell 命令	49
2.5.2	运行 Spark-Shell 读取 HDFS 文件	50
2.6	IDEA 开发 WordCount 程序	52
2.6.1	以本地模式执行 Spark 程序	52
2.6.2	集群模式执行 Spark 程序	54
2.7	本章小结	58
2.8	课后习题	59
<b>第 3 章</b>	<b>Spark RDD 弹性分布式数据集</b>	<b>60</b>
3.1	RDD 简介	60
3.2	RDD 的创建方式	61
3.2.1	从文件系统加载数据创建 RDD	61
3.2.2	通过并行集合创建 RDD	62
3.3	RDD 的处理过程	63
3.3.1	转换算子	63
3.3.2	行动算子	67
3.3.3	编写 WordCount 词频统计案例	70
3.4	RDD 的分区	71
3.5	RDD 的依赖关系	71
3.6	RDD 机制	73
3.6.1	持久化机制	73
3.6.2	容错机制	75
3.7	Spark 的任务调度	76
3.7.1	DAG 的概念	76

3.7.2	RDD 在 Spark 中的运行流程 .....	76
3.8	本章小结 .....	78
3.9	课后习题 .....	78
<b>第 4 章</b>	<b>Spark SQL 结构化数据文件处理 .....</b>	<b>80</b>
4.1	Spark SQL 的基础知识 .....	80
4.1.1	Spark SQL 的简介 .....	80
4.1.2	Spark SQL 架构 .....	81
4.2	DataFrame 的基础知识 .....	82
4.2.1	DataFrame 简介 .....	82
4.2.2	DataFrame 的创建 .....	83
4.2.3	DataFrame 的常用操作 .....	86
4.3	Dataset 的基础知识 .....	89
4.3.1	Dataset 简介 .....	89
4.3.2	Dataset 对象的创建 .....	89
4.4	RDD 转换为 DataFrame .....	90
4.4.1	反射机制推断 Schema .....	90
4.4.2	编程方式定义 Schema .....	92
4.5	Spark SQL 操作数据源 .....	94
4.5.1	操作 MySQL .....	94
4.5.2	操作 Hive 数据集 .....	96
4.6	本章小结 .....	99
4.7	课后习题 .....	99
<b>第 5 章</b>	<b>HBase 分布式数据库 .....</b>	<b>101</b>
5.1	HBase 的基础知识 .....	101
5.1.1	HBase 的简介 .....	101
5.1.2	HBase 的数据模型 .....	102
5.2	HBase 的集群部署 .....	103
5.3	HBase 的基本操作 .....	107
5.3.1	HBase 的 Shell 操作 .....	107
5.3.2	HBase 的 Java API 操作 .....	112
5.4	深入学习 HBase 原理 .....	120
5.4.1	HBase 架构 .....	121
5.4.2	物理存储 .....	122
5.4.3	寻址机制 .....	123
5.4.4	HBase 读写数据流程 .....	124
5.5	HBase 和 Hive 的整合 .....	125
5.6	本章小结 .....	130

5.7	课后习题 .....	130
<b>第 6 章</b>	<b>Kafka 分布式发布订阅消息系统 .....</b>	<b>132</b>
6.1	Kafka 的基础知识 .....	132
6.1.1	消息传递模式简介 .....	132
6.1.2	Kafka 简介 .....	133
6.2	Kafka 工作原理 .....	134
6.2.1	Kafka 核心组件介绍 .....	134
6.2.2	Kafka 工作流程分析 .....	136
6.3	Kafka 集群部署与测试 .....	138
6.3.1	安装 Kafka .....	138
6.3.2	启动 Kafka 服务 .....	140
6.4	Kafka 生产者 and 消费者实例 .....	141
6.4.1	基于命令行方式使用 Kafka .....	141
6.4.2	基于 Java API 方式使用 Kafka .....	143
6.5	Kafka Streams .....	148
6.5.1	Kafka Streams 概述 .....	149
6.5.2	Kafka Streams 开发单词计数应用 .....	149
6.6	本章小结 .....	153
6.7	课后习题 .....	153
<b>第 7 章</b>	<b>Spark Streaming 实时计算框架 .....</b>	<b>155</b>
7.1	实时计算的基础知识 .....	155
7.1.1	什么是实时计算 .....	155
7.1.2	常用的实时计算框架 .....	155
7.2	Spark Streaming 的基础知识 .....	156
7.2.1	Spark Streaming 简介 .....	156
7.2.2	Spark Streaming 工作原理 .....	157
7.3	Spark 的 DStream .....	157
7.3.1	DStream 简介 .....	157
7.3.2	DStream 编程模型 .....	158
7.3.3	DStream 转换操作 .....	158
7.3.4	DStream 窗口操作 .....	164
7.3.5	DStream 输出操作 .....	168
7.3.6	DStream 实例——实现网站热词排序 .....	170
7.4	Spark Streaming 整合 Kafka 实战 .....	173
7.4.1	KafkaUtils.createDstream 方式 .....	173
7.4.2	KafkaUtils.createDirectStream 方式 .....	177
7.5	本章小结 .....	180

7.6	课后习题 .....	180
<b>第8章</b>	<b>Spark MLlib 机器学习算法库 .....</b>	<b>182</b>
8.1	初识机器学习 .....	182
8.1.1	什么是机器学习 .....	182
8.1.2	机器学习的应用 .....	183
8.2	Spark 机器学习库 MLlib 的概述 .....	184
8.2.1	MLlib 的简介 .....	184
8.2.2	Spark 机器学习工作流程 .....	185
8.3	数据类型 .....	186
8.3.1	本地向量 .....	186
8.3.2	标注点 .....	186
8.3.3	本地矩阵 .....	187
8.4	Spark MLlib 基本统计 .....	188
8.4.1	摘要统计 .....	188
8.4.2	相关统计 .....	189
8.4.3	分层抽样 .....	190
8.5	分类 .....	191
8.5.1	线性支持向量机 .....	191
8.5.2	逻辑回归 .....	192
8.6	案例——构建推荐系统 .....	193
8.6.1	推荐模型分类 .....	194
8.6.2	利用 MLlib 实现电影推荐 .....	194
8.7	本章小结 .....	200
8.8	课后习题 .....	200
<b>第9章</b>	<b>综合案例——Spark 实时交易数据统计 .....</b>	<b>202</b>
9.1	系统概述 .....	202
9.1.1	系统背景介绍 .....	202
9.1.2	系统架构设计 .....	202
9.1.3	系统预览 .....	203
9.2	Redis 数据库 .....	203
9.2.1	Redis 介绍 .....	204
9.2.2	Redis 部署与启动 .....	204
9.2.3	Redis 操作及命令 .....	205
9.3	模块开发——构建工程结构 .....	206
9.4	模块开发——构建订单系统 .....	211
9.4.1	模拟订单数据 .....	211
9.4.2	向 Kafka 集群发送订单数据 .....	212

9.5	模块开发——分析订单数据 .....	215
9.6	模块开发——数据展示 .....	219
9.6.1	搭建 Web 开发环境 .....	219
9.6.2	实现数据展示功能 .....	221
9.6.3	可视化平台展示 .....	227
9.7	本章小结 .....	228

# 第 1 章

## Scala 语言基础

### 学习目标

- 了解 Scala 的特点。
- 掌握 Scala 和 IDEA 的下载安装。
- 掌握 Scala 的基础语法。
- 掌握 Scala 的数据结构。
- 熟悉 Scala 面向对象的特性。
- 掌握 Scala 的模式匹配与样例类。

Spark 是专为大规模数据处理而设计的快速通用的计算引擎,它是用 Scala 语言开发实现的。大数据技术本身就是数据计算的技术,而 Scala 既有面向对象组织项目工程的能力,又具备计算数据的功能,同时与 Spark 紧密集成,本书将采用 Scala 语言开发 Spark 程序,所以学好 Scala 将有助于读者更好地掌握 Spark 框架。接下来,本章将讲解 Scala 语言的基础知识。

## 1.1 初识 Scala

### 1.1.1 Scala 概述

Scala 于 2001 年由瑞士洛桑联邦理工学院(EPFL)编程方法实验室研发,它由 Martin Odersky(马丁·奥德斯基)创建。目前,许多公司依靠 Java 进行的关键性业务应用已转向或正在转向 Scala,以提高应用程序的可扩展性和整体的可靠性,从而提高开发效率。

Scala 是 Scalable Language 的简称,它是一门多范式的编程语言,其设计初衷是实现一种可扩展的语言,并集成面向对象编程和函数式编程的各种特性。基于这个目标与设计,Scala 具有以下显著的特性。

(1) Scala 是面向对象的语言。

Scala 是一种纯粹的面向对象语言,每一个值都是对象。对象的数据类型以及行为由类和特征来描述,类抽象机制的扩展通过两种途径实现:一种是子类继承,另一种是混入机制,这两种途径都能够避免多重继承的问题。

(2) Scala 是函数式编程的语言。

Scala 也是一种函数式语言,其函数可以作为值来使用。Scala 提供了轻量级的语法用于定义匿名函数,支持高阶函数,允许嵌套多层函数,并支持柯里化。

(3) Scala 是静态类型的。

Scala 具备类型系统,通过编译时的类型检查来保证代码的安全性和一致性。类型系统支持的特性包括泛型类、注释、类型上下限约束、类别和抽象类型作为对象成员、复合类型、引用自己时显示指定类型、视图、多态方法等。

(4) Scala 是可扩展的。

在实际开发中,某个特定领域的应用程序开发往往需要特定领域的语言扩展。Scala 提供了许多独特的语言机制,它能够很容易地以库的方式无缝添加新的语言结构。

(5) Scala 是可以交互操作的。

Scala 可以与流行的 Java Runtime Environment(JRE)进行良好的交互操作。Scala 用 scalac 编译器把源文件编译成 Java 的 class 文件(即可以在 JVM 上运行的字节码)。我们可以从 Scala 中调用所有的 Java 类库,同样也可以从 Java 应用程序中调用 Scala 代码。

## 1.1.2 Scala 的下载安装

Scala 语言可以在 Windows、Linux、Mac OS 等系统上编译运行。由于 Scala 是运行在 JVM 平台上的,所以安装 Scala 之前必须配置好 JDK 环境(JDK 版本要求不低于 1.5)。本书使用的 JDK 版本是 jdk1.8,关于 JDK 的安装和配置这里不作详解。

在不同操作系统上安装 Scala 环境的相关介绍如下。

### 1. 在 Windows 下安装 Scala

访问 Scala 官网 <https://www.scala-lang.org/>,单击【DOWNLOAD】按钮进入下载页面,在该页面可以下载最新版本的 Scala。考虑到 Scala 的稳定性以及和 Spark 的兼容性,这里选择下载 Scala 2.11.8(下载地址为 <https://www.scala-lang.org/download/2.11.8.html>),具体如图 1-1 所示。

Archive	System	Size
scala-2.11.8.tgz	Mac OSX, Unix, Cygwin	27.35M
scala-2.11.8.msi	Windows (msi installer)	109.35M
scala-2.11.8.zip	Windows	27.40M
scala-2.11.8.deb	Debian	76.02M
scala-2.11.8.rpm	RPM package	108.16M
scala-docs-2.11.8.tgz	API docs	46.00M
scala-docs-2.11.8.zip	API docs	84.21M
scala-sources-2.11.8.tar.gz	Sources	

图 1-1 下载 Window 系统支持的 Scala 安装包

下载成功后,解压 Scala 的安装包 scala-2.11.8.zip,并配置 Windows 系统的环境变量,效果如图 1-2 和图 1-3 所示。

测试 Scala 环境是否安装成功。进入 Windows 的命令行,输入 scala 命令,按 Enter 键,效果如图 1-4 所示。

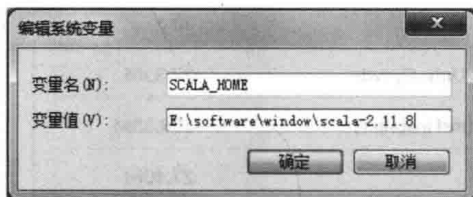


图 1-2 Scala 系统变量的配置

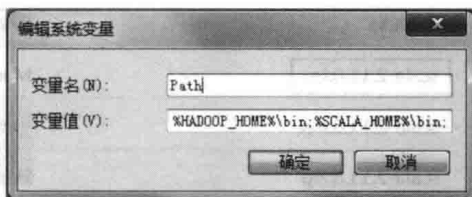


图 1-3 将 Scala 系统变量引入到环境中



图 1-4 测试 Scala 环境的安装

从图 1-4 可以看出,控制台输出了 Scala 的版本号 2.11.8,证明 Scala 环境已经安装成功。

## 2. 在 Linux 下安装 Scala

通过 Scala 官网下载 Linux 系统下的 Scala-2.11.8 的安装包 scala-2.11.8.tgz(下载地址为 <https://www.scala-lang.org/download/2.11.8.html>)。将安装包上传到 Linux 系统的 /export/software 目录下,进行解压安装,解压命令如下:

```
$tar -zxvf scala-2.11.8.tgz -C /export/servers/
```

执行 vi /etc/profile 命令,进入 Linux 环境变量的配置文件中,添加 Scala 环境变量,具体内容如下:

```
export SCALA_HOME=/export/servers/scala-2.11.8
export PATH=$PATH:$SCALA_HOME/bin
```

添加完上述的内容后,执行 source /etc/profile 命令,使配置的环境变量生效,Scala 在 Linux 系统下的环境安装完成。

## 3. 在 Mac 下安装 Scala

首先,通过 Scala 官网下载支持 Mac OS 的 Scala 安装包 scala-2.11.8.tgz,下载地址为 <https://www.scala-lang.org/download/2.11.8.html>,具体如图 1-5 所示。

下载成功后,解压安装包,并将其移动到主目录下(如果找不到主目录,可以回到桌面,按快捷键 Shift-Command-H 进入计算机主目录)。

然后,修改环境变量。将 bin 目录添加到路径中,路径通常存储在计算机主目录下的

Archive	System	Size
scala-2.11.8.tgz	Mac OSX, Unix, Cygwin	27.35M
scala-2.11.8.msi	Windows (msi installer)	109.35M
scala-2.11.8.zip	Windows	27.40M
scala-2.11.8.deb	Debian	76.02M
scala-2.11.8.rpm	RPM package	108.16M
scala-docs-2.11.8.bxz	API docs	46.00M
scala-docs-2.11.8.zip	API docs	84.21M
scala-sources-2.11.8.tar.gz	Sources	

图 1-5 下载 Mac 系统支持的 Scala 安装包

.profile 或 .bash\_profile 的文件中。假设此刻要编辑 .bash\_profile 文件,可以使用下列命令打开文件,具体如下:

```
$ touch ~/.bash_profile      #如果 bash_profile 不存在,可以使用此命令创建
$ open ~/.bash_profile      #打开 bash_profile 文件
```

接着,打开 bash\_profile 文件后,将下列内容添加到所有 PATH 语句之后。

```
export PATH="主路径/Scala/bin:$PATH"
```

最后,保存并关闭 bash\_profile 文件,重启计算机,并输入下列命令查看 Scala 版本号,测试 Scala 的安装情况,具体如下:

```
$ scala -version
```

如果 Scala 安装成功,计算机同样会输出 Scala 的版本号,说明 Scala 安装成功。

### 1.1.3 在 IDEA 开发工具中下载安装 Scala 插件

目前 Scala 的主流开发工具主要有两种: Eclipse 工具和 IDEA 工具,在这两个开发工具中可以安装对应的 Scala 插件进行 Scala 开发。由于 IDEA 工具可以自动识别代码错误并进行简单的修复,而且 IDEA 工具内置了很多优秀的插件,所以现在大多数 Scala 开发程序员都会选择 IDEA 作为开发 Scala 的工具。接下来,本书将以 Windows 操作系统为例,分步骤讲解如何在 IDEA 工具上下载安装 Scala 插件,具体步骤如下。

(1) 访问 <http://www.jetbrains.com/idea/download/previous.html> 下载 IDEA 工具,本书选择的版本是 2018. 2. 5 (IDEA 只是编程工具,读者可以任意选择);然后,打开 IDEA 安装包,单击【Next】按钮进行安装,直到安装结束。最终显示的效果如图 1-6 所示。

(2) 下载 Scala 插件(地址为 <https://plugins.jetbrains.com/plugin/1347-scala>),本书选择的版本是 2018. 2. 4 (scala-intellij-bin-2018. 2. 4. zip)。在 IDEA 工具上安装 Scala 插件,单击主界面右下角的【Configure】下拉按钮,然后选择【Plugins】命令,效果如图 1-7 所示。