

厦门大学林子雨，赖永炫，陶继平 编著

# 《Spark 编程基础（Scala 版）》

## 教材配套 机房上机实验指南

### 实验 2 Scala 编程初级实践

（版本号：2018 年 7 月 19 日版本）

（答案）



主讲教师：林子雨  
厦门大学数据库实验室  
二零一八年七月

## 目录

一、实验目的 .....	1
二、实验平台 .....	1
三、实验内容和要求 .....	1
1. 计算级数.....	1
2. 模拟图形绘制.....	2
3. 统计学生成绩.....	4
四、实验报告 .....	7
附录 1:任课教师介绍 .....	8
附录 2: 课程教材介绍.....	8
附录 3: 高校大数据课程公共服务平台介绍.....	9

厦门大学林子雨,赖永炫,陶继平 编著

## 《Spark 编程基础 (Scala 版)》

### 第 2 章 Scala 语言基础

#### 教材配套机房上机实验指南

#### 实验 2 Scala 编程初级实践

#### (答案)

主讲教师: 林子雨

E-mail: ziyulin@xmu.edu.cn 个人主页: <http://www.cs.xmu.edu.cn/linziyu>

### 一、实验目的

1. 掌握 Scala 语言的基本语法、数据结构和控制结构;
2. 掌握面向对象编程的基础知识, 能够编写自定义类和特质;
3. 掌握函数式编程的基础知识, 能够熟练定义匿名函数。熟悉 Scala 的容器类库的基本层次结构, 熟练使用常用的容器类进行数据;
4. 熟练掌握 Scala 的 REPL 运行模式和编译运行方法。

### 二、实验平台

已经配置完成的 Scala 开发环境。Scala 版本为 2.11.8.

### 三、实验内容和要求

#### 1. 计算级数

请用脚本的方式编程计算并输出下列级数的前  $n$  项之和  $S_n$ , 直到  $S_n$  刚好大于或等于  $q$  为止, 其中  $q$  为大于 0 的整数, 其值通过键盘输入。

$$S_n = \frac{2}{1} + \frac{3}{2} + \frac{4}{3} + \cdots + \frac{n+1}{n}$$

例如, 若  $q$  的值为 50.0, 则输出应为:  $S_n=50.416695$ 。请将源文件保存为 exercise2-1.scala, 在 REPL 模式下测试运行, 测试样例:  $q=1$  时,  $S_n=2$ ;  $q=30$  时,  $S_n=30.891459$ ;  $q=50$  时,  $S_n=50.416695$ 。

#### 【参考答案】

```
import io.StdIn._
var Sn:Float = 0
var n:Float=1
println("please input q:")
val q = readInt()
```

```
while (Sn<q) {  
  Sn+=(n+1)/n  
  n+=1  
}  
println(s"$Sn")
```

## 2. 模拟图形绘制

对于一个图形绘制程序,用下面的层次对各种实体进行抽象。定义一个 Drawable 的特质,其包括一个 draw 方法,默认实现为输出对象的字符串表示。定义一个 Point 类表示点,其混入了 Drawable 特质,并包含一个 shift 方法,用于移动点。所有图形实体的抽象类为 Shape,其构造函数包括一个 Point 类型,表示图形的具体位置(具体意义对不同的具体图形不一样)。Shape 类有一个具体方法 moveTo 和一个抽象方法 zoom,其中 moveTo 将图形从当前位置移动到新的位置,各种具体图形的 moveTo 可能会有不一样的地方。zoom 方法实现对图形的放缩,接受一个浮点型的放缩倍数参数,不同具体图形放缩实现不一样。继承 Shape 类的具体图形类型包括直线类 Line 和圆类 Circle。Line 类的第一个参数表示其位置,第二个参数表示另一个端点,Line 放缩的时候,其中点位置不变,长度按倍数放缩(注意,缩放时,其两个端点信息也改变了),另外,Line 的 move 行为影响了另一个端点,需要对 move 方法进行重载。Circle 类第一个参数表示其圆心,也是其位置,另一个参数表示其半径,Circle 缩放的时候,位置参数不变,半径按倍数缩放。另外直线类 Line 和圆类 Circle 都混入了 Drawable 特质,要求对 draw 进行重载实现,其中类 Line 的 draw 输出的信息样式为“Line:第一个端点的坐标—第二个端点的坐标”,类 Circle 的 draw 输出的信息样式为“Circle center:圆心坐标,R=半径”。如下的代码已经给出了 Drawable 和 Point 的定义,同时也给出了程序入口 main 函数的实现,请完成 Shape 类、Line 类和 Circle 类的定义。

```
case class Point(var x:Double,var y:Double) extends Drawable{  
  def shift(deltaX:Double,deltaY:Double){x+=deltaX;y+=deltaY}  
}  
trait Drawable{  
  def draw(){println(this.toString)}  
}
```

// 请完成 Shape 类、Line 类和 Circle 类的定义。

```
object MyDraw{  
  def main(args: Array[String]) {  
    val p=new Point(10,30)  
    p.draw;  
    val line1 = new Line(Point(0,0),Point(20,20))  
    line1.draw  
    line1.moveTo(Point(5,5)) //移动到一个新的点  
    line1.draw  
    line1.zoom(2) //放大两倍  
    line1.draw  
    val cir= new Circle(Point(10,10),5)
```

```
        cir.draw
        cir.moveTo(Point(30,20))
        cir.draw
        cir.zoom(0.5)
        cir.draw
    }
}
```

编译运行程序,期望的输出结果如下:

```
Point(10.0,30.0)
Line:(0.0,0.0)--(20.0,20.0)
Line:(5.0,5.0)--(25.0,25.0)
Line:(-5.0,-5.0)--(35.0,35.0)
Circle center:(10.0,10.0),R=5.0
Circle center:(30.0,20.0),R=5.0
Circle center:(30.0,20.0),R=2.5
```

### 【参考答案】

```
case class Point(var x:Double,var y:Double) extends Drawable{
    def shift(deltaX:Double,deltaY:Double){x+=deltaX;y+=deltaY}
}
trait Drawable{
    def draw(){println(this.toString)}
}
abstract class Shape(var location:Point){//location是Shape的一个可变字段
    def moveTo(newLocation:Point){ //默认实现,只是修改位置
        location = newLocation
    }
    def zoom(scale:Double)
}
class Line(beginPoint:Point,var endPoint:Point) extends
Shape(beginPoint) with Drawable{
    override def draw(){
        println(s"Line: (${location.x},${location.y})--(${endPoint.x},${endPoint.y})") //按指定格式重载click
    }
    override def moveTo(newLocation:Point){
        endPoint.shift(newLocation.x - location.x,newLocation.y - location.y) //直线移动时,先移动另外一个端点
        location = newLocation //移动位置
    }
    override def zoom(scale:Double){
        val midPoint = Point((endPoint.x + location.x)/2,(endPoint.y + location.y)/2) //求出中点,并按中点进行缩放
        location.x = midPoint.x + scale * (location.x - midPoint.x)
    }
}
```

```
        location.y = midPoint.y + scale * (location.y - midPoint.y)
        endPoint.x = midPoint.x + scale * (endPoint.x - midPoint.x)
        endPoint.y = midPoint.y + scale * (endPoint.y - midPoint.y)
    }
}

class Circle(center:Point,var radius:Double) extends Shape(center)
with Drawable{
    override def draw(){//按指定格式重载click
        println(s"Circle
center:(${location.x},${location.y}),R=$radius")
    }
    override def zoom(scale:Double){
        radius = radius*scale //对圆的缩放只用修改半径
    }
}

object MyDraw{
    def main(args: Array[String]) {
        val p=new Point(10,30)
        p.draw;
        val line1 = new Line(Point(0,0),Point(20,20))
        line1.draw
        line1.moveTo(Point(5,5))
        line1.draw
        line1.zoom(2)
        line1.draw
        val cir= new Circle(Point(10,10),5)
        cir.draw
        cir.moveTo(Point(30,20))
        cir.draw
        cir.zoom(0.5)
        cir.draw
    }
}
```

### 3. 统计学生成绩

学生的成绩清单格式如下所示,第一行为表头,各字段意思分别为学号、性别、课程名 1、课程名 2 等,后面每一行代表一个学生的信息,各字段之间用空白符隔开

Id	gender	Math	English	Physics
301610	male	80	64	78
301611	female	65	87	58
...				

给定任何一个如上格式的清单 (不同清单里课程数量可能不一样), 要求尽可能采用函数式编程, 统计出各门课程的平均成绩, 最低成绩, 和最高成绩; 另外还需按男女同学分开, 分别统计各门课程的平均成绩, 最低成绩, 和最高成绩。

测试样例 1 如下:

Id	gender	Math	English	Physics
301610	male	80	64	78
301611	female	65	87	58
301612	female	44	71	77
301613	female	66	71	91
301614	female	70	71	100
301615	male	72	77	72
301616	female	73	81	75
301617	female	69	77	75
301618	male	73	61	65
301619	male	74	69	68
301620	male	76	62	76
301621	male	73	69	91
301622	male	55	69	61
301623	male	50	58	75
301624	female	63	83	93
301625	male	72	54	100
301626	male	76	66	73
301627	male	82	87	79
301628	female	62	80	54
301629	male	89	77	72

样例 1 的统计结果输出为:

course	average	min	max
Math:	69.20	44.00	89.00
English:	71.70	54.00	87.00
Physics:	76.65	54.00	100.00
course	average	min	max (males)
Math:	72.67	50.00	89.00
English:	67.75	54.00	87.00
Physics:	75.83	61.00	100.00
course	average	min	max (females)
Math:	64.00	44.00	73.00
English:	77.63	71.00	87.00
Physics:	77.88	54.00	100.00

测试样例 2

Id	gender	Math	English	Physics	Science
301610	male	72	39	74	93
301611	male	75	85	93	26
301612	female	85	79	91	57
301613	female	63	89	61	62
301614	male	72	63	58	64

301615	male	99	82	70	31
301616	female	100	81	63	72
301617	male	74	100	81	59
301618	female	68	72	63	100
301619	male	63	39	59	87
301620	female	84	88	48	48
301621	male	71	88	92	46
301622	male	82	49	66	78
301623	male	63	80	83	88
301624	female	86	80	56	69
301625	male	76	69	86	49
301626	male	91	59	93	51
301627	female	92	76	79	100
301628	male	79	89	78	57
301629	male	85	74	78	80

样例 2 的统计结果为:

course	average	min	max
Math:	79.00	63.00	100.00
English:	74.05	39.00	100.00
Physics:	73.60	48.00	93.00
Science:	65.85	26.00	100.00
course	average	min	max
Math:	77.08	63.00	99.00
English:	70.46	39.00	100.00
Physics:	77.77	58.00	93.00
Science:	62.23	26.00	93.00
course	average	min	max
Math:	82.57	63.00	100.00
English:	80.71	72.00	89.00
Physics:	65.86	48.00	91.00
Science:	72.57	48.00	100.00

### 【参考答案】

```
object scoreReport{
  def main(args: Array[String]) {
    // 假设数据文件在当前目录下
    val inputFile = scala.io.Source.fromFile("test.txt")

    //"\s+"是字符串正则表达式,将每行按空白字符(包括空格/制表符)分开
    // 由于可能涉及多次遍历,同 toList 将 Iterator 装为 List
    // originalData 的类型为 List[Array[String]]
    val originalData =
inputFile.getLines.map{_.split("\\s+")} .toList

    val courseNames = originalData.head.drop(2) //获取第一行中的课程名
    val allStudents = originalData.tail // 去除第一行剩下的数据
    val courseNum = courseNames.length

    // 统计函数,参数为需要常用统计的行
    //用到了外部变量 courseNum,属于闭包函数
```



```
def statistic(lines:List[Array[String]])= {  
// for 推导式, 对每门课程生成一个三元组, 分别表示总分, 最低分和最高分  
  (for(i<- 2 to courseNum+1) yield {  
    // 取出需要统计的列  
    val temp = lines map {elem=>elem(i).toDouble}  
    (temp.sum,temp.min,temp.max)  
  }) map {case (total,min,max) => (total/lines.length,min,max)}  
  } // 最后一个map 对 for 的结果进行修改, 将总分转为平均分  
}  
  
// 输出结果函数  
def printResult(theresult:Seq[(Double,Double,Double)]) {  
  // 遍历前调用 zip 方法将课程名容器和结果容器合并, 合并结果为二元组容器  
  (courseNames zip theresult) foreach {  
    case (course,result)=>  
      println(f"${course+":"}%-10s${result._1}%5.2f${result._2}%8.2f${result._3}%8.2f")  
  }  
}  
  
// 分别调用两个函数统计全体学生并输出结果  
val allResult = statistic(allStudents)  
println("course    average    min    max")  
printResult(allResult)  
  
//按性别划分为两个容器  
val (maleLines,femaleLines) = allStudents partition  
{_(1)=="male"}  
  
// 分别调用两个函数统计男学生并输出结果  
val maleResult = statistic(maleLines)  
println("course    average    min    max")  
printResult(maleResult)  
  
// 分别调用两个函数统计女学生并输出结果  
val femaleResult = statistic(femaleLines)  
println("course    average    min    max")  
printResult(femaleResult)  
}  
}
```

## 四、实验报告

《Spark 编程基础》实验报告		
题目:	姓名:	日期:
实验环境:		
解决问题的思路:		
实验内容与完成情况:		
出现的问题:		
解决方案 (列出遇到的问题和解决办法, 列出没有解决的问题):		

## 附录 1: 任课教师介绍



林子雨(1978—),男,博士,厦门大学计算机科学系助理教授,主要研究领域为数据库、数据仓库、数据挖掘、大数据

主讲课程:《大数据处理技术》

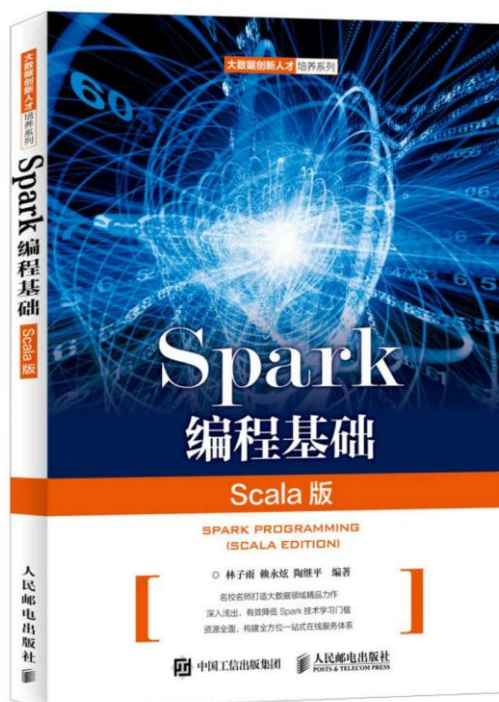
办公地点:厦门大学海韵园科研 2 号楼

E-mail: ziyulin@xmu.edu.cn

个人主页: <http://www.cs.xmu.edu.cn/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>

## 附录 2: 课程教材介绍



林子雨、赖永炫、陶继平编著《Spark 编程基础（Scala 版）》

人民邮电出版社 ISBN:978-7-115-48816-9 定价: 49.80 元

厦门大学林子雨、赖永炫和陶继平老师编著《Spark 编程基础》，以 Scala 作为开发 Spark 应用程序的编程语言，系统介绍了 Spark 编程的基础知识。全书共 8 章，内容包括大数据技术概述、Scala 语言基础、Spark 的设计与运行原理、Spark 环境搭建和使用方法、RDD 编程、Spark SQL、Spark Streaming、Spark MLlib 等。本书每个章节都安排了入门级的编程实践操作，以便读者更好地学习和掌握 Spark 编程方法。本书官网免费提供了全套的在线教学资源，包括讲义 PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。

本书可以作为高等院校计算机、软件工程、数据科学与大数据技术等专业的进阶级大数据课程教材，用于指导 Spark 编程实践，也可供相关技术人员参考。

欢迎访问《Spark 编程基础》教材官方网站: <http://dblab.xmu.edu.cn/post/spark/>



扫一扫访问教材官网

## 附录 3：高校大数据课程公共服务平台介绍



**高校大数据课程**

公 共 服 务 平 台

高校大数据课程公共服务平台，由中国高校首个“数字教师”的提出者和建设者——林子雨老师发起，由厦门大学数据库实验室全力打造，由厦门大学云计算与大数据研究中心、海峡云计算与大数据应用研究中心携手共建。这是国内第一个服务于高校大数据课程建设的公共服务平台，旨在促进国内高校大数据课程体系建设，提高大数据课程教学水平，降低大数据课程学习门槛，提升学生课程学习效果。平台服务对象涵盖高校、教师和学生。平台为高校开设大数据课程提供全流程辅助，为教师开展教学工作提供一站式服务，为学生学习大数据课程提供全方位辅导。平台重点打造“11 个 1 工程”，即 1 本教材（含官网）、1 个教师服务站、1 个学生服务站、1 个公益项目、1 堂巡讲公开课、1 个示范班级、1 门在线课程、1 个交流群（QQ 群、微信群）、1 个保障团队、1 个培训基地和 1 个实验平台。目前平台每年访问量已经超过 100 万次，成为国内高校大数据教学知名品牌。

平台主页：<http://dblab.xmu.edu.cn/post/bigdata-teaching-platform/>



扫一扫访问平台主页