厦门大学林子雨，郑海山，赖永炫 编著

《Spark编程基础（Python版）》

教材配套

机房上机实验指南

实验1 Linux系统的安装和常用命令

（版本号：2020年4月版本）

（答案）



主讲教师：林子雨

厦门大学数据库实验室

二零二零年四月

目录

[一、实验目的 1](#_Toc495049316)

[二、实验平台 1](#_Toc495049317)

[三、实验内容和要求 1](#_Toc495049318)

[1．安装Linux虚拟机 1](#_Toc495049319)

[2．使用Linux系统的常用命令 2](#_Toc495049320)

[3. 在Windows系统和Linux系统之间互传文件 3](#_Toc495049321)

[四、实验报告 3](#_Toc495049322)

[**附录1:任课教师介绍** 4](#_Toc495049323)

[**附录2：课程教材介绍** 4](#_Toc495049324)

[**附录3：高校大数据课程公共服务平台介绍** 5](#_Toc495049325)

厦门大学林子雨,郑海山,赖永炫 编著《Spark编程基础（Python版）》

第1章 大数据技术概述

教材配套机房上机实验指南

实验1 Linux系统的安装和常用命令

（答案）

**主讲教师：林子雨**

E-mail: [ziyulin@xmu.edu.cn](mailto:ziyulin@xmu.edu.cn) 个人主页：<http://dblab.xmu.edu.cn/post/linziyu>

# 一、实验目的

（1）掌握Linux虚拟机的安装方法。Spark和Hadoop等大数据软件在Linux操作系统上运行可以发挥最佳性能，因此，本教程中，Spark都是在Linux系统中进行相关操作，同时，下一章的Scala语言也会在Linux系统中安装和操作。鉴于目前很多读者正在使用Windows操作系统，因此，为了顺利完成本教程的后续实验，这里有必要通过本实验，让读者掌握在Windows操作系统上搭建Linux虚拟机的方法。当然，安装Linux虚拟机只是安装Linux系统的其中一种方式，实际上，读者也可以不用虚拟机，而是采用双系统的方式安装Linux系统。本教程推荐使用虚拟机方式。

（2）熟悉Linux系统的基本使用方法。本教程全部在Linux环境下进行实验，因此，需要读者提前熟悉Linux系统的基本用法，尤其是一些常用命令的使用方法。

# 二、实验平台

操作系统：Windows系统和Ubuntu（推荐）。

虚拟机软件：推荐使用的开源虚拟机软件为VirtualBox。VirtualBox是一款功能强大的免费虚拟机软件，它不仅具有丰富的特色，性能也很优异，且简单易用，可虚拟的系统包括Windows、Mac OS X、Linux、OpenBSD、Solaris、IBM OS2甚至Android 4.0系统等操作系统。读者可以在Windows系统上安装VirtualBox软件，然后在VirtualBox上安装并且运行Linux操作系统。本次实验默认的Linux发行版为Ubuntu16.04。

# 三、实验内容和要求

## 1．安装Linux虚拟机

请登录Windows系统，下载VirtualBox软件和Ubuntu16.04镜像文件。

VirtualBox软件的下载地址：<https://www.virtualbox.org/wiki/Downloads>。

Ubuntu16.04的镜像文件下载地址：http://www.ubuntu.org.cn/download/desktop。

或者也可以直接到本教程官网的“下载专区”的“软件”中下载Ubuntu安装文件ubuntukylin-16.04-desktop-amd64.iso。

首先，在Windows系统上安装虚拟机软件VirtualBox软件，然后在虚拟机软件VirtualBox上安装Ubuntu16.04操作系统，具体请参考本教程官网的“实验指南”中的“在Windows中使用VirtualBox安装Ubuntu”。

【参考答案】

请参照下面网络教程完成Linux虚拟机的安装。

http://dblab.xmu.edu.cn/blog/337-2/

## 2．使用Linux系统的常用命令

【参考答案】

启动Linux虚拟机，进入Linux系统，通过查阅相关Linux书籍和网络资料，或者参考本教程官网的“实验指南”的“Linux系统常用命令”，完成如下操作：

（1）切换到目录 /usr/bin；

|  |
| --- |
| $ cd /usr/bin |

（2）查看目录/usr/local下所有的文件；

|  |
| --- |
| $cd /usr/local  $ls -al |

（3）进入/usr目录，创建一个名为test的目录，并查看有多少目录存在；

|  |
| --- |
| $ cd /usr  $ mkdir test  $ ls -al |

（4）在/usr下新建目录test1，再复制这个目录内容到/tmp；

|  |
| --- |
| $ cd /usr  $ mkdir test1  $ sudo cp -r /usr/test1 /tmp |

（5）将上面的/tmp/test1目录重命名为test2；

|  |
| --- |
| $ sudo mv /tmp/test1 /tmp/test2 |

（6）在/tmp/test2目录下新建word.txt文件并输入一些字符串保存退出；

|  |
| --- |
| $ cd /tmp/test2  $ vim word.txt  #按键盘Esc键退出vim编辑状态，输入“:wq”保存文件并退出vim编辑器 |

（7）查看word.txt文件内容；

|  |
| --- |
| $ cat /tmp/test2/word.txt |

（8）将word.txt文件所有者改为root帐号，并查看属性；

|  |
| --- |
| $ sudo chown root /tmp/test2/word.txt  $ ls -l /tmp/test2/word.txt |

（9）找出/tmp目录下文件名为test2的文件；

|  |
| --- |
| $ sudo find /tmp -name test2 |

（10）在/目录下新建文件夹test,然后在/目录下打包成test.tar.gz；

|  |
| --- |
| $ sudo mkdir /test  $ sudo tar -zcv -f /test.tar.gz test |

（11）将test.tar.gz解压缩到/tmp目录。

|  |
| --- |
| $ sudo tar -zxv -f /test.tar.gz -C /tmp |

## 3. 在Windows系统和Linux系统之间互传文件

本教程大量实验都是在Linux虚拟机上完成，因此，需要掌握如何把Windows系统中的文件上传到Linux系统，以及如何把Linux系统中的文件下载到Windows系统中。

首先，到本教程官网的“下载专区”中的“软件”目录中，下载FTP软件FileZilla的安装文件FileZilla\_3.17.0.0\_win64\_setup.exe，把FileZilla安装到Windows系统中；然后，请参考本教程官网“实验指南”栏目的“在Windows系统中利用FTP软件向Ubuntu系统上传文件”，完成以下操作：

（1）在Windows系统中新建一个文本文件test.txt，并通过FTP软件FileZilla，把test.txt上传到Linux系统中的“/home/hadoop/下载”目录下，把利用Linux命令把该文件名修改为test1.txt；

（2）通过FTP软件FileZilla，把Linux系统中的“/home/hadoop/下载”目录下的test1.txt文件下载到Windows系统的某个目录下。

【参考答案】

请参照下面网络教程完成该操作：

http://dblab.xmu.edu.cn/blog/1608-2/

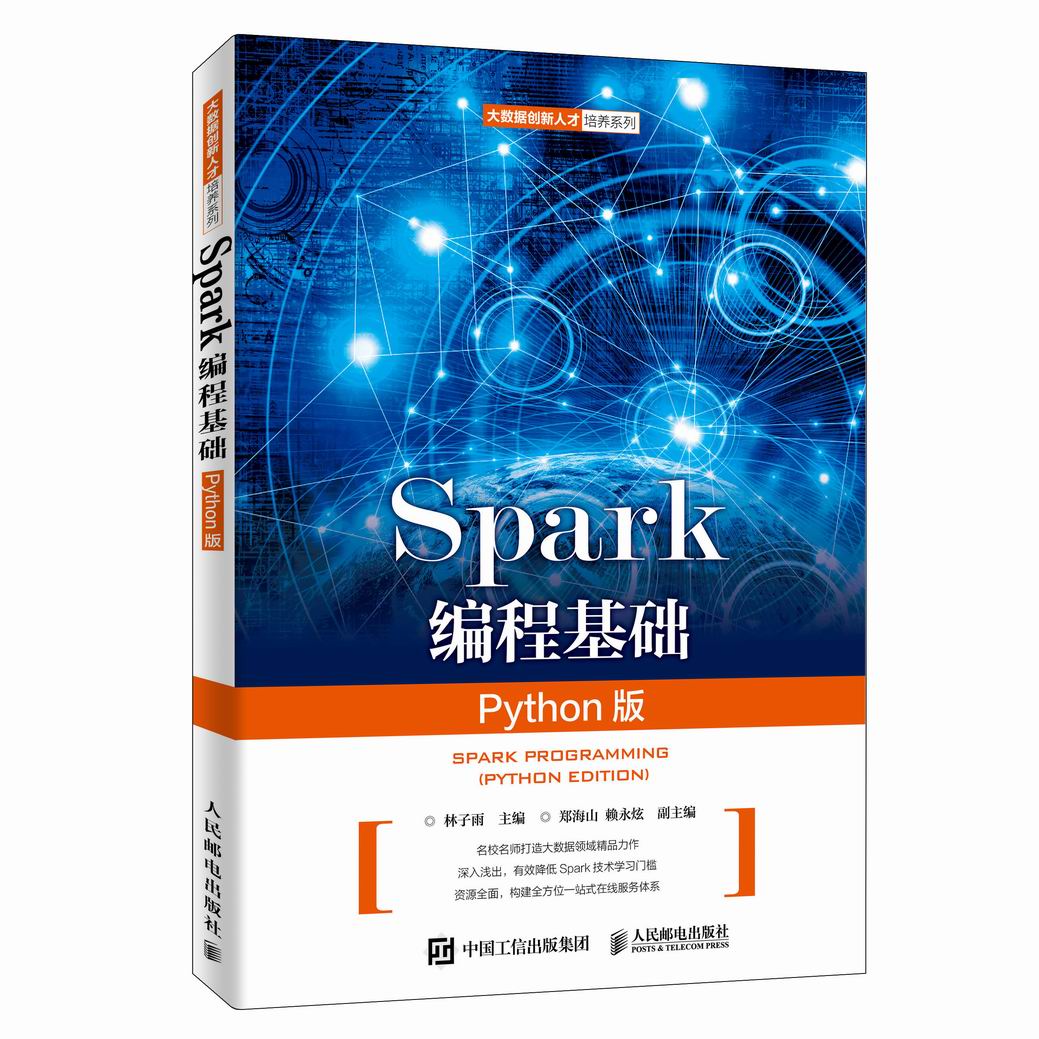
# 四、实验报告

|  |  |  |
| --- | --- | --- |
| 《Spark编程基础》实验报告 | | |
| 题目： | 姓名： | 日期： |
| 实验环境： | | |
| 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**

|  |  |
| --- | --- |
| [http://www.cs.xmu.edu.cn/database/linziyu/images/linziyu2016.jpg](http://dblab.xmu.edu.cn/post/linziyu/) | 林子雨（1978－），男，博士，国内高校知名大数据教师，厦门大学计算机科学系副教授，厦门大学云计算与大数据研究中心创始成员，厦门大学数据库实验室负责人，中国计算机学会数据库专委会委员，中国计算机学会信息系统专委会委员，中国高校首个“数字教师”提出者和建设者。2013年开始在厦门大学开设大数据课程，建设了国内高校首个大数据课程公共服务平台，平台累计网络访问量超过1000万次，成为全国高校大数据教学知名品牌，并荣获“2018年福建省教学成果二等奖”，主持的课程《大数据技术原理与应用》获评“2018年国家精品在线开放课程”。  E-mail: ziyulin@xmu.edu.cn  个人主页：http://dblab.xmu.edu.cn/post/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、郑海山、赖永炫编著《Spark编程基础（Python版）》

人民邮电出版社 ISBN:978-7-115-52439-3 定价：49.80元

厦门大学林子雨、郑海山、赖永炫老师编著《Spark编程基础（Python版）》，以Python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、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、Structured 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(Python版)》教材官方网站：http://dblab.xmu.edu.cn/post/spark-python/



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，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200万次，累计访问量超过1000万次，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