

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 1 Linux 系统的安装和常用命令

（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. 安装 Linux 虚拟机	1
2. 使用 Linux 系统的常用命令	2
3. 在 Windows 系统和 Linux 系统之间互传文件	3
四、实验报告.....	3
附录 1:任课教师介绍	4
附录 2: 课程教材介绍.....	4
附录 3: 高校大数据课程公共服务平台介绍.....	5

厦门大学林子雨,赖永炫,陶继平 编著《Spark 编程基础 (Scala 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 个人主页: <http://www.cs.xmu.edu.cn/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 `ubuntu16.04-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 Ubuntu14.04 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://dbllab.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



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》

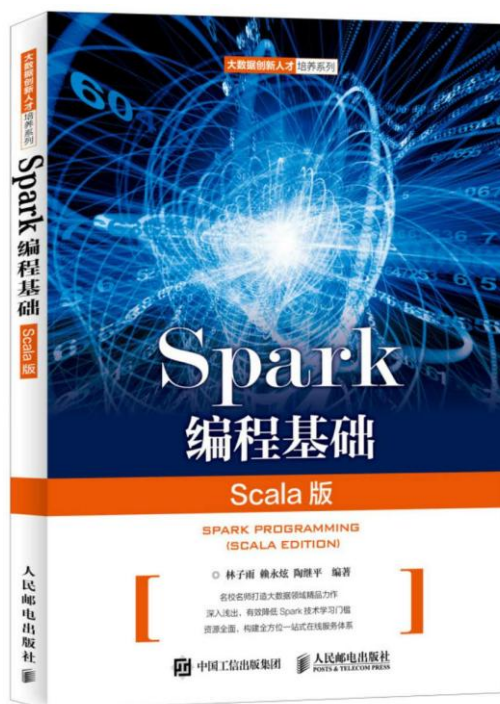
办公地点:厦门大学海韵园科研 2 号楼

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