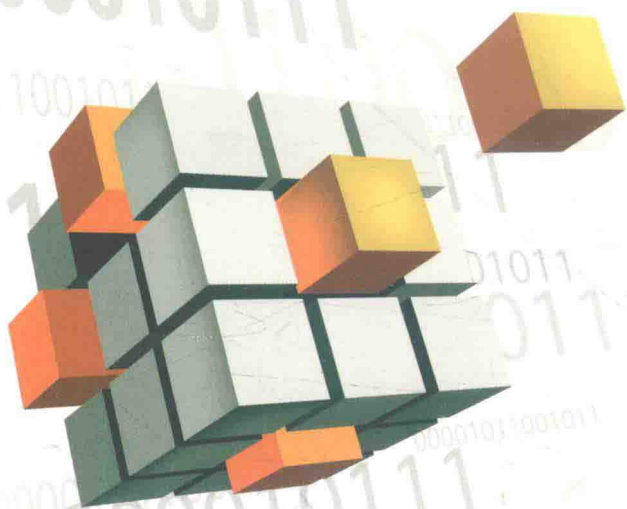




新世纪高等学校规划教材·大数据系列



陈明◎编著

# 数据科学与 大数据技术导论实验

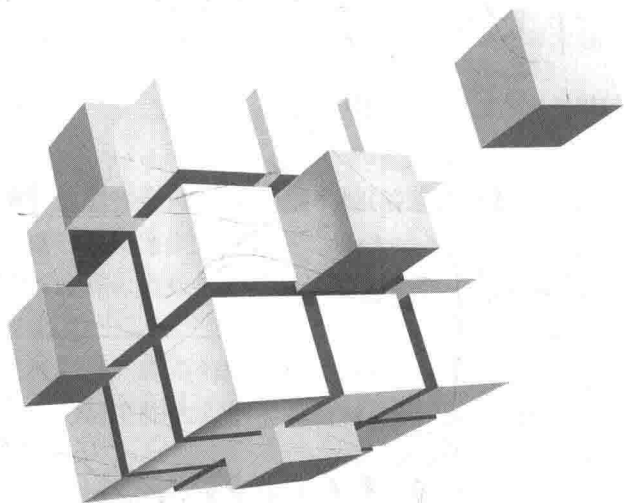
- ◆ 基于开源的大数据处理软件工具
- ◆ 呈现数据生命周期中的典型实验



北京师范大学出版集团  
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP  
北京师范大学出版社



新世纪高等学校规划教材·大数据系列



陈明◎编著

数据科学与



大数据技术导论实验

SHUJU KEXUE YU DASHUJU JISHU DAOLUN SHIYAN



北京师范大学出版集团  
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP  
北京师范大学出版社

---

图书在版编目 (CIP) 数据

数据科学与大数据技术导论实验/陈明编著. —北京: 北京师范大学出版社, 2018. 7

(新世纪高等学校规划教材·大数据系列)

ISBN 978-7-303-23450-9

I. ①数… II. ①陈… III. ①数据处理—实验—高等学校—教材

IV. ①TP274.33

中国版本图书馆 CIP 数据核字 (2018) 第 022637 号

---

营 销 中 心 电 话      010-62978190 62979006

北师大出版社科技与经管分社      www.jswsbook.com

电 子 信 箱      jswsbook@163.com

---

出版发行: 北京师范大学出版社      www.bnup.com

北京市海淀区新街口外大街 19 号

邮政编码: 100875

印 刷: 北京京师印务有限公司

经 销: 全国新华书店

开 本: 787mm×1092mm 1/16

印 张: 14.25

字 数: 315 千字

版 次: 2018 年 7 月第 1 版

印 次: 2018 年 7 月第 1 次印刷

定 价: 32.80 元

---

策划编辑: 赵洛育      责任编辑: 赵洛育

美术编辑: 刘 超      装帧设计: 刘 超

责任校对: 马子杰      责任印制: 赵非非

**版权所有 侵权必究**

反盗版、反侵权举报电话: 010-62978190

北京读者服务部电话: 010-62979006-8021

外埠邮购电话: 010-62978190

本书如有印装质量问题, 请与印制管理部联系调换。

印制管理部电话: 010-62979006-8006

# 内 容 简 介

本书是“数据科学与大数据技术导论实验”课程的实验教材，可以配合主教材使用。主要内容包括 Linux 操作系统部署、Hadoop 开发环境部署、网页数据获取、大数据去重、大数据最大值计算、大数据排序、大数据倒排索引、大数据平均值计算、大数据单表关联、大数据可视化。

本书选材先进、特点鲜明，注重应用。通过本书所提供的实验练习，可以提高大数据处理能力。本书可用作大数据实验教材，也可以作为科学技术人员学习和应用大数据的科技参考书。

# 前 言

在“数据科学与大数据技术导论”课程的教学过程中，为了使能够快速理解课程的内容，增强理论与实践的结合，尤其是培养应用能力，编写了《数据科学与大数据技术导论实验》教材，可以作为课程实验的内容，配合主教材使用。

本教材主要设置了 10 个实验，分别是 Linux 操作系统部署、Hadoop 开发环境部署、网页数据获取、大数据去重、大数据最大值计算、大数据排序、大数据倒排索引、大数据平均值计算、大数据单表关联和大数据可视化。

上述的 10 个实验，涉及了大数据处理周期中内容的基本实验，在大数据处理平台 Hadoop 开发环境下完成，主要应用了 MapReduce 分布编程模型完成基本的分析处理功能。各实验的内容简介如下。

## 实验 1 Linux 操作系统部署

Linux 操作系统是部署 Hadoop 开发环境的最佳选择，通过 Linux 操作系统部署的实验，可以掌握安装 Linux 操作系统的方法，进而为大数据 Hadoop 环境部署奠定基础。

## 实验 2 Hadoop 开发环境部署

通过 Hadoop 环境部署实验练习，可以在理解本实验所涉及的理论基础之上，掌握 Hadoop 系统安装方法、伪分布式 Hadoop 的安装方法和 Eclipse 开发环境的具体安装过程。

## 实验 3 网页数据获取

通过网页数据获取的实验，可以掌握利用网络爬虫获取网页数据的方法，进而解决网页数据获取的实际问题。

## 实验 4 大数据去重

通过大数据去重实验，可以掌握在 Hadoop 环境下大数据去重的方法。

## 实验 5 大数据最大值计算

通过大数据最大值计算实验，可以掌握在 Hadoop 环境下大数据最大值计算的方法。

## 实验 6 大数据排序

通过大数据排序实验，可以掌握在 Hadoop 环境下大数据排序的方法。

## 实验 7 大数据倒排索引

通过 Hadoop 倒排索引实验，可以掌握在 Hadoop 环境下大数据倒排索引的方法。

## 实验 8 大数据平均值计算

通过大数据平均值计算实验，可以掌握在 Hadoop 环境下大数据平均值计算的方法。

## 实验 9 大数据单表关联

通过大数据表关联实验，可以掌握在 Hadoop 环境下大数据表关联的方法。

## 实验 10 大数据可视化

通过大数据可视化实验，可以掌握 Superset（或 Echar）环境构建方法、直方图、时序线图、时序面积图、饼图和文字云图可视化方法，进而为大数据分析结果展现奠定基础。

在本教材中涉及 Hadoop 环境的实验中，使用了流行的基于 MapReduce 分布编程的典型程序（<http://www.cnblogs.com/xia52opi/archive/2012/06/04/2534533.html>），这些程序使用 Java 语言书写，如果具有 Java 程序设计与阅读能力，应尽可能理解程序的功能和内容。

在本书的编写过程中，得到了北京师范大学出版社赵洛育编辑和王锁柱教授的支持，李猛坤博士、张广骏博士、李勇博士参加了部分实验实践的工作，在此，一并表示谢意。



2017.11

# 目 录

实验 1 Linux 操作系统部署 .....	1	2.6 实验总结 .....	49
1.1 实验目的 .....	1	2.7 思考拓展 .....	49
1.2 实验要求 .....	1	实验 3 网页数据获取 .....	50
1.3 实验原理 .....	1	3.1 实验目的 .....	50
1.3.1 Linux 操作系统 .....	1	3.2 实验要求 .....	50
1.3.2 虚拟机 .....	2	3.3 实验原理 .....	50
1.4 Linux 安装方法 .....	3	3.3.1 网络爬虫的工作过程 .....	51
1.4.1 Ubuntu 安装方法 .....	3	3.3.2 通用网络爬虫 .....	51
1.4.2 RHEL 安装方法 .....	12	3.3.3 聚焦网络爬虫 .....	53
1.4.3 Linux 基本命令使用方法 .....	21	3.3.4 数据抓取目标的定义 .....	56
1.5 实验内容 .....	22	3.3.5 网页分析算法 .....	57
1.6 实验总结 .....	22	3.3.6 更新策略 .....	59
1.7 思考拓展 .....	22	3.3.7 分布式爬虫的系统结构 .....	60
实验 2 Hadoop 开发环境部署 .....	23	3.3.8 ForeSpider 数据采集系统 .....	62
2.1 实验目的 .....	23	3.3.9 GooSeeker 集搜客爬虫软件 .....	67
2.2 实验要求 .....	23	3.4 数据获取方法 .....	68
2.3 实验原理 .....	23	3.4.1 前嗅 ForeSpider 爬虫软件数据 采集过程 .....	68
2.3.1 OpenSSH .....	24	3.4.2 基于 GooSeeker 集搜客爬虫软 件的数据采集 .....	79
2.3.2 Hadoop 处理平台 .....	27	3.5 实验内容 .....	85
2.3.3 MapReduce 分布计算模型 .....	29	3.6 实验总结 .....	85
2.3.4 Eclipse 集成开发环境 .....	31	3.7 思考拓展 .....	86
2.4 Hadoop 开发环境部署方法 .....	31	实验 4 大数据去重 .....	87
2.4.1 安装 SSH 协议 .....	31	4.1 实验目的 .....	87
2.4.2 安装 OpenJDK 1.8 开发 环境 .....	32	4.2 实验要求 .....	87
2.4.3 安装 Hadoop 系统 .....	33	4.3 实验原理 .....	87
2.4.4 伪分布式 Hadoop 环境部署 .....	35	4.3.1 数据去重的步骤 .....	87
2.4.5 集成开发环境 Eclipse 部署 .....	39	4.3.2 按时间的维度去重 .....	88
2.4.6 Eclipse 开发环境的使用 方法 .....	42	4.3.3 按字段相似度的去重 .....	89
2.5 实验内容 .....	49		

4.4 Hadoop 数据去重方法 .....	89	6.4.2 准备数据 .....	124
4.4.1 进入/usr/local/hadoop 目录.....	89	6.4.3 修改/usr/local/hadoop/etc/hadoop/ 目录下的 Hadoop 配置文件 ...	125
4.4.2 准备数据 .....	89	6.4.4 NameNode 的格式化 .....	125
4.4.3 修改/usr/local/hadoop/etc/hadoop/ 目录下的 Hadoop 配置文件... 91		6.4.5 在集成开发环境 Eclipse 中实现 Hadoop .....	125
4.4.4 NameNode 的格式化 .....	92	6.5 实验内容 .....	137
4.4.5 在集成开发环境 Eclipse 中实现 Hadoop 数据去重 .....	92	6.6 实验总结 .....	137
4.5 实验内容 .....	104	6.7 思考拓展 .....	137
4.6 实验总结 .....	104	<b>实验 7 大数据倒排索引 .....</b>	<b>138</b>
4.7 思考拓展 .....	104	7.1 实验目的 .....	138
<b>实验 5 大数据最大值计算 .....</b>	<b>105</b>	7.2 实验要求 .....	138
5.1 实验目的 .....	105	7.3 实验原理 .....	138
5.2 实验要求 .....	105	7.3.1 倒排索引基本概念 .....	139
5.3 实验原理 .....	105	7.3.2 倒排索引文件实例分析 .....	140
5.4 Hadoop 最大值计算方法 .....	105	7.3.3 倒排索引处理过程 .....	142
5.4.1 进入/usr/local/hadoop 目录... 106		7.4 Hadoop 倒排索引方法.....	144
5.4.2 准备数据 .....	106	7.4.1 进入/usr/local/hadoop 目录... 144	
5.4.3 修改/usr/local/hadoop/etc/hadoop/ 目录下的 Hadoop 配置文件 ... 107		7.4.2 准备数据 .....	145
5.4.4 NameNode 的格式化 .....	107	7.4.3 修改/usr/local/hadoop/etc/hadoop/ 目录下的 Hadoop 配置文件 ... 145	
5.4.5 在集成开发环境 Eclipse 中 Hadoop 最大值计算 .....	107	7.4.4 NameNode 的格式化 .....	145
5.5 实验内容 .....	117	7.4.5 在集成开发环境 Eclipse 中实现 Hadoop 倒排索引.....	146
5.6 实验总结 .....	117	7.5 实验内容 .....	158
5.7 思考拓展 .....	118	7.6 实验总结 .....	158
<b>实验 6 大数据排序 .....</b>	<b>119</b>	7.7 思考拓展 .....	158
6.1 实验目的 .....	119	<b>实验 8 大数据平均值计算 .....</b>	<b>159</b>
6.2 实验要求 .....	119	8.1 实验目的 .....	159
6.3 实验原理 .....	119	8.2 实验要求 .....	159
6.3.1 排序方法分类 .....	119	8.3 实验原理 .....	159
6.3.2 排序方法 .....	120	8.3.1 算术平均数 .....	159
6.3.3 大数据排序算法 .....	121	8.3.2 平均值 .....	160
6.3.4 基于 MapReduce 排序过程... 122		8.4 大数据平均值计算方法 .....	161
6.4 Hadoop 数据排序方法 .....	123	8.4.1 进入/usr/local/hadoop 目录... 161	
6.4.1 进入/usr/local/hadoop 目录... 124		8.4.2 准备数据 .....	161



8.4.3	修改/usr/local/hadoop/etc/hadoop/ 目录下的Hadoop配置文件 ...	162	9.4.5	在集成开发环境 Eclipse 中实现 大数据平均值 .....	181
8.4.4	NameNode 的格式化 .....	162	9.5	实验内容 .....	193
8.4.5	在集成开发环境 Eclipse 中实现 Hadoop 平均值 .....	162	9.6	实验总结 .....	193
8.5	实验内容 .....	174	9.7	思考拓展 .....	194
8.6	实验总结 .....	174	实验 10	大数据可视化 .....	195
8.7	思考拓展 .....	174	10.1	实验目的 .....	195
实验 9	大数据单表关联 .....	175	10.2	实验要求 .....	195
9.1	实验目的 .....	175	10.3	实验原理 .....	195
9.2	实验要求 .....	175	10.3.1	数据的可视化展现 .....	195
9.3	实验原理 .....	175	10.3.2	Superset 处理平台 .....	198
9.3.1	单表关联的概念 .....	175	10.4	基于数据可视化的方法 .....	198
9.3.2	MapReduce 单表关联计算 过程 .....	176	10.4.1	基于 ECharts.js 可视化 方法 .....	198
9.4	Hadoop 单表关联方法 .....	180	10.4.2	基于 Superset 数据可视化的 方法 .....	201
9.4.1	进入/usr/local/hadoop 目录 ...	180	10.5	实验内容 .....	216
9.4.2	准备数据 .....	181	10.6	实验总结 .....	216
9.4.3	修改/usr/local/hadoop/etc/hadoop/ 目录下的Hadoop配置文件 ...	181	10.7	思考拓展 .....	216
9.4.4	NameNode 的格式化 .....	181	参考文献	.....	217

# 实验 1 Linux 操作系统部署

Linux 操作系统应用日益广泛，现已成为主流的网络操作系统。云计算、物联网、移动互联网和大数据等研究热点与应用领域的出现与发展，都应用了 Linux 操作系统。随着互联网的广泛应用，Linux 用户也迅速扩展，Linux 操作系统发挥出越来越大的作用。

## 1.1 实验目的

通过 Linux 操作系统部署的实验，可以掌握虚拟机平台 VirtualBox 及扩展包安装方法、创建 Linux 虚拟机方法、安装 Linux 操作系统方法，进而为大数据 Hadoop 环境部署奠定基础。

## 1.2 实验要求

在了解 Linux 操作系统安装的相关知识基础之上，通过实例完成下述任务。

- (1) 虚拟机平台 VirtualBox 及扩展包安装。
- (2) 创建虚拟机。
- (3) 安装 Ubuntu 操作系统。

## 1.3 实验原理

在当前最为普及的 Windows 操作系统中，利用甲骨文公司的 Oracle VM VirtualBox 虚拟化平台及扩展包搭建 Linux。

### 1.3.1 Linux 操作系统

Windows、Linux、UNIX 和 OSX 等操作系统是目前最常用的操作系统，Linux 操作系统的主要 4 个版本简介如下。

#### 1. RHEL

Red Hat Enterprise Linux (RHEL) 是 Red Hat Linux 的企业版。RHEL 主要的版本有 Advanced Server (AS)、ES Workstation (WS)、Desktop，它们之间差别只是功能方面的差别。Red Hat 分为两个系列：一个系列是由红帽公司提供收费技术支持和更新的 Red Hat Enterprise Linux，另一个系列是由社区开发的免费 Fedora Core 操作系统。

## 2. Fedora Core

Fedora Core (FC) 是红帽公司的新技术实验场地, 许多新技术都在 FC 企业中检验, 如果新技术稳定, 则红帽公司将其加入 Red Hat Enterprise Linux 中。FC 的定位是桌面用户, 提供了最新的软件包, 同时, 其版本更新周期短, 仅 6 个月。由于版本更新频繁, 影响了性能和稳定性, 因此, 一般在服务器上不采用 Fedora Core。其实 Fedora 就是 Red Hat 发行 Red Hat Enterprise Linux 的一个实验版本, 通过用户测试, 为 Red Hat Enterprise Linux 发布建立基础。

## 3. Ubuntu

Ubuntu 是一个以桌面应用为主的开源 GNU/Linux 操作系统, Ubuntu 是基于 Debian GNU/Linux, 支持 x86、amd64 (即 x64) 和 ppc 架构, 由全球化的专业开发团队 (Canonical Ltd) 开发。本实验所使用的操作系统为 Ubuntu-16.10 版本。

## 4. CentOS

CentOS (Community Enterprise Operating System) 是社区企业操作系统, 是 Red Hat 的 AS/ES/WS 的免费版本, 具有 RHEL 的基本功能和性能。CentOS 还提供免费的长期升级和更新服务。虽然 CentOS 是 RHEL 的克隆, 但并不完全相同, RHEL 软件在 CentOS 上不能保证 100% 正常工作。

### 1.3.2 虚拟机

在实体计算机上, 通过软件模拟出一台或者多台虚拟计算机简称虚拟机。虚拟机使用宿主机的硬件资源, 拥有真实计算机的绝大多数功能。可以在虚拟机中安装虚拟机软件所支持的操作系统, 例如, 可以在虚拟机中安装 Linux, 也可以在虚拟机中安装 Windows, 无论宿主机使用的是 UNIX 系统还是 Windows 系统。

#### 1. Oracle VM VirtualBox

VirtualBox 是一款开源虚拟机软件。VirtualBox 是由德国 Innotek 公司开发, 由 Sun Microsystems 公司出品的软件, 使用 Qt 编写, 在 Sun 被 Oracle 收购后正式更名成 Oracle VM VirtualBox。

VirtualBox 是免费虚拟机软件, 不仅具有丰富的特色, 而且性能优异, 简单易用, 可虚拟的系统包括 Windows、Mac OS X、Linux、OpenBSD、Solaris、IBM OS2 甚至 Android 等操作系统。使用者可以在 VirtualBox 上安装并且运行上述操作系统, 与同性质的 VMware 及 Virtual PC 比较, VirtualBox 的优势是远端桌面协定 (RDP)、iSCSI 及 USB 的支持, VirtualBox 在客户端操作系统上已可以支持 USB 2.0 的硬件装置, 但需要安装 VirtualBox Extension Pack。

#### 2. VMware Workstation

在 Windows 系统下有很多虚拟机软件, 如 VMware Workstation 和 Virtual PC 等。

VMware Workstation 是一款功能强大的桌面虚拟计算机软件, 可为用户提供在单一的桌面上同时运行不同的操作系统来进行开发、测试、部署新的应用程序的最佳解决方案。

VMware Workstation 可在一部实体机器上模拟完整的网络环境,以及便于携带的虚拟机器,其更好的灵活性与先进的技术超过了市面上其他的虚拟计算机软件。VMware 在虚拟网络、实时快照、拖曳共享文件夹、支持 PXE 等方面的特点使其成为必不可少的工具。

## 1.4 Linux 安装方法

本节介绍常用的 Ubuntu 操作系统和 RHEL 操作系统安装方法。

### 1.4.1 Ubuntu 安装方法

甲骨文公司的 VirtualBox 是 x86 系统的跨平台虚拟化软件,可以从官方网址:<http://www.oracle.com/technetwork/cn/server-storage/virtualbox/downloads/index.html> 下载适合 Windows 系统的 Oracle VM VirtualBox 扩展包软件。

#### 1. 支持 VT 虚拟化功能

为了保障虚拟机运行性能和对 64 位操作系统的支持,需要在安装虚拟机的宿主计算机的 BIOS 设置中打开 Intel CPU 的 VT 虚拟化功能。具体方法是,可通过在 Windows 下运行资源目录中的 SecurAble (VT 检测工具) 汉化版.exe 来检测是否已经支持了 VT 虚拟化功能。VT 虚拟化功能界面如图 1-1 所示,最右边的一项 Hardware Virtualization 必须为“是”,表示 VT 虚拟化功能设置成功;如果不为“是”,需要重启计算机调整 BIOS 设置。由于各种计算机设置步骤不统一,所以有关设置操作可以参考所用宿主计算机 BIOS 配置说明。



图 1-1 VT 虚拟化功能界面

#### 2. 虚拟机平台 VirtualBox 及扩展包的安装

甲骨文公司的 VirtualBox 是 x86 系统的跨平台虚拟化软件,可以从官方网址:

<http://www.oracle.com/technetwork/cn/server-storage/virtualbox/downloads/index.html> 下载适合 Windows 系统的 Oracle VM VirtualBox 扩展包软件。

虚拟机平台 VirtualBox 的安装可以按照 Windows 下的常用软件默认安装步骤，首先打开 VirtualBox 软件管理菜单项的“全局设定”配置对话框，在左栏选择“扩展”项后，单击右侧第一个按钮“添加新包”，选中资源目录的 Oracle\_VM\_VirtualBox\_Extension\_Pack-5.1.26-117224 .vbox-extpack 文件完成扩展安装。图 1-2 所示为已完成安装设定的扩展配置界面。

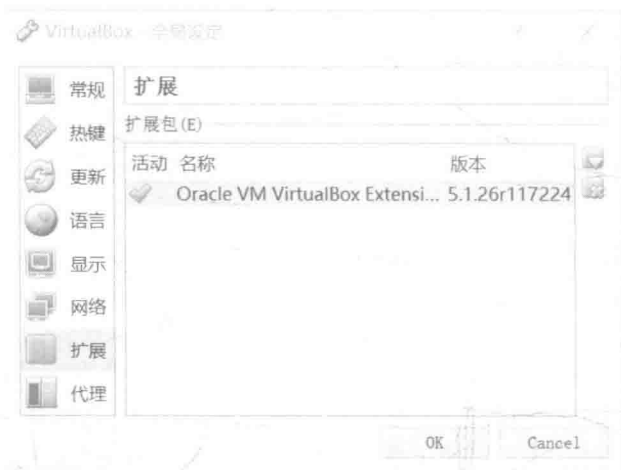


图 1-2 完成安装设定的扩展配置界面

### 3. 创建 Linux 虚拟机

(1) 新建虚拟机电脑。在 Oracle VM VirtualBox 界面工具栏中单击“新建”按钮，在弹出的“新建虚拟电脑”对话框中单击“专家模式”切换后，出现如图 1-3 所示的界面，在这个界面下进行设定：名称为 rootuser，类型为 Linux，版本为 Ubuntu (64-bit)，内存大小调整为 4096MB。



图 1-3 新建虚拟电脑

(2) 创建虚拟硬盘。在图 1-3 中,选中“现在创建虚拟硬盘”单选按钮,单击“创建”按钮继续,出现如图 1-4 所示的“创建虚拟硬盘”对话框。在文件位置选择具有足够硬盘空间的目录位置,文件大小调整为 16.00GB,其余选项默认,单击“创建”按钮。至此一个 Linux 虚拟机已在 Oracle VM VirtualBox 中创建。

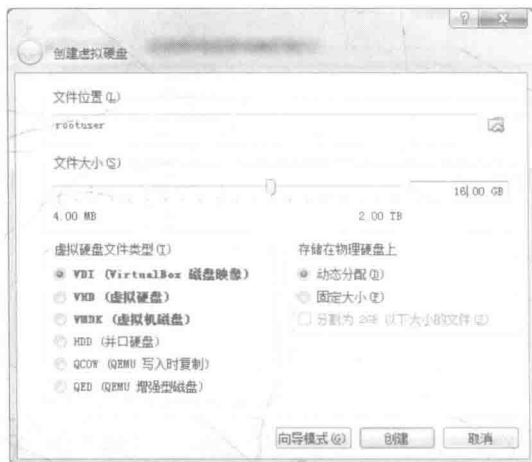


图 1-4 创建虚拟硬盘

#### 4. 安装 Ubuntu 操作系统

(1) 在 Oracle VM VirtualBox 中(见图 1-5)选中已经创建的 rootuser 虚拟机,单击工具栏中的“设置”按钮,出现如图 1-6 所示的设置对话框。



图 1-5 Oracle VM VirtualBox 界面

在左边栏中选择“存储”项后,单击右侧光盘状“选择一个虚拟光盘文件……”按钮,选中资源目录的 ubuntu-17.04-desktop-amd64.iso 文件(类似版本的 Ubuntu 安装光盘的 ISO

文件可以从 Ubuntu 的官网下载页面 <https://www.ubuntu.com/download/desktop> 免费下载得到)。图 1-6 所示的是在“控制器：IDE”已加载 ubuntu-16.10-desktop-amd64.iso 安装光盘镜像文件后的设置情况。



图 1-6 安装光盘镜像文件后的设置情况

(2) 返回到 Oracle VM VirtualBox 界面，选中 rootuser 虚拟机，单击工具栏中的“启动”按钮，此时弹出虚拟机启动运行窗口。图 1-7 所示是光盘启动后出现的安装欢迎界面，在左侧列表中选择“中文（简体）”切换为中文，单击右侧的“安装 Ubuntu”按钮继续。其他选项保持默认状态，然后单击“继续”按钮。



图 1-7 光盘启动后的安装欢迎界面

(3) 在虚拟机运行窗口继续 Ubuntu 安装，在时区确认处单击世界地图中国区域，图 1-8 所示的文本框显示 Shanghai 后单击“继续”按钮。



图 1-8 时区确认界面

(4) 安装的键盘布局设置完成后, 如图 1-9 所示的安装在的用户设置页面需要慎重配置, 例如, 姓名为 rootuser; 计算机名为 BigData; 用户名为 hadoop; 密码可自行设置但要牢记, 选中“自动登录”单选按钮后单击“继续”按钮。

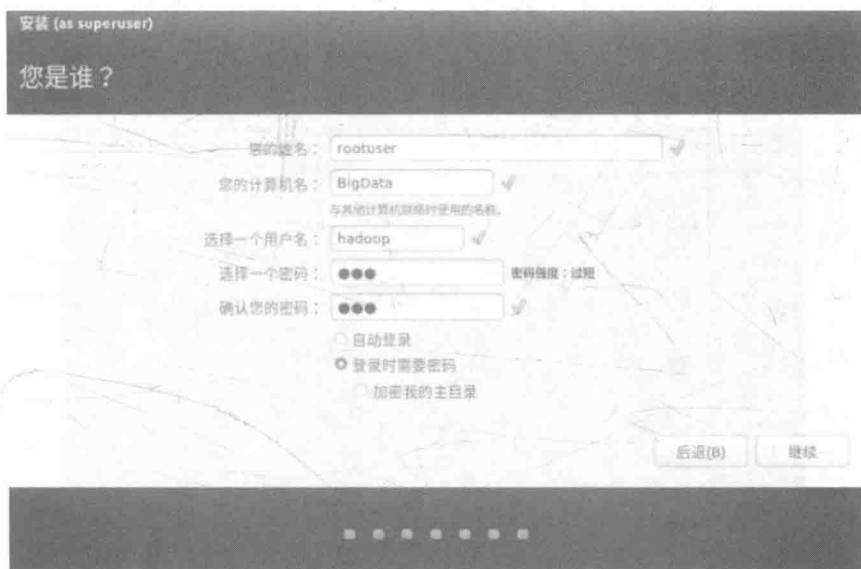


图 1-9 安装的用户设置

经过上述设置后, 将在虚拟机中自动执行安装过程, 完成后按提示按 Enter 键确认, 重启虚拟机。重启后在虚拟机运行窗口进入 Ubuntu 默认桌面显示。

(5) 终端程序的快捷打开方式。后续需要多次用到 Linux 的终端程序进行命令执行等操作。下面介绍终端程序的一种快捷打开方式。单击如图 1-10 所示的 Ubuntu 桌面左侧工具栏“终端”图标按钮 (即第一个按钮), 在弹出搜索框中输入 te 后找到的第一个应用程



序就是“终端”，单击后即可进入终端命令行模式，同时注意在搜索窗口右上角的小按钮可以像 Windows 一样快速关闭、最大化显示当前应用。



图 1-10 Ubuntu 桌面

(6) 快速进入终端应用的方法。打开终端应用后，如图 1-11 所示，可在左侧工具栏的终端图标处右击，在弹出的快捷菜单中选择“锁定到启动器”命令，这样每次可直接从工具栏终端图标处快速进入终端应用。

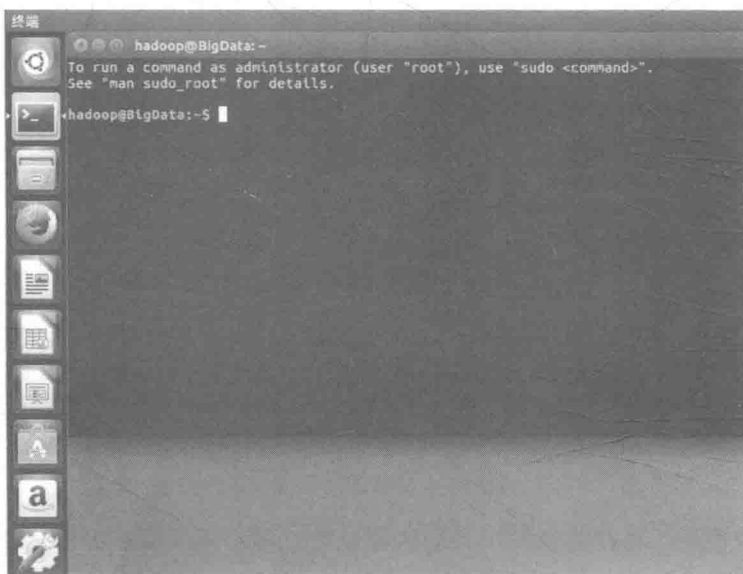


图 1-11 快速进入终端应用

(7) 在 Ubuntu 系统中还需要安装对应的 Oracle VM VirtualBox 虚拟机扩展包文件。如图 1-12 所示，在虚拟机运行完整大窗口的菜单栏，选择设备菜单列表的“安装增强