



《大数据基础编程、实验和案例教程（第2版）》

教材官网：

<http://dmlab.xmu.edu.cn/post/bigdatapRACTICE2/>

温馨提示：编辑幻灯片母版，可以修改每页PPT的厦大校徽和底部文字

第2章 Linux系统的安装和使用

（PPT版本号：2020年12月版本）



扫一扫访问教材官网

林子雨

厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn ▶▶

主页: <http://dmlab.xmu.edu.cn/linziyu>





教材简介

本书是与《大数据技术原理与应用（第3版）》教材配套的唯一指定实验指导书

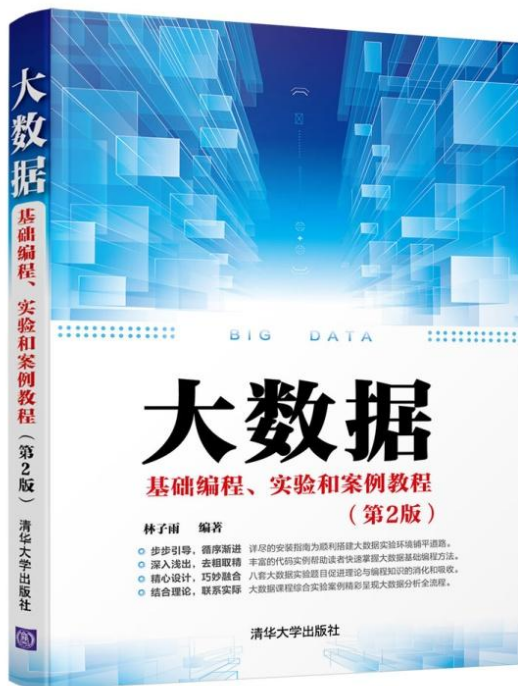
林子雨编著《大数据基础编程、实验和案例教程（第2版）》

清华大学出版社 ISBN:978-7-302-55977-1 定价：69元，2020年10月第2版

教材官网：<http://dbllab.xmu.edu.cn/post/bigdatapRACTICE2/>



扫一扫访问
教材官网



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，八套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程



提纲

- 2.1 Linux系统简介
- 2.2 Linux系统安装
- 2.3 Linux系统及相关软件的基本使用方法



高校大数据课程

公共服务平台

百度搜索厦门大学数据库实验室网站访问平台





2.1 Linux系统简介

- Linux是一套免费使用和自由传播的类Unix操作系统，是一个基于POSIX和UNIX的多用户、多任务、支持多线程和多CPU的操作系统。
- Linux有许多服务于不同目的的发行版，包括对不同计算机结构的支持、对一个具体区域或语言的本地化、实时应用和嵌入式系统等，已经有超过三百个发行版，但是，目前在全球范围内只有十个左右发行版被普遍使用，比如Fedora、Debian、Ubuntu、RedHat、SuSE、CentOS等。
- Linux的发行版本可以大体分为两类，一类是商业公司维护的发行版本，一类是社区组织维护的发行版本，前者以著名的Redhat为代表，后者以Debian为代表。Debian是社区类Linux的典范，是迄今为止最遵循GNU规范的Linux系统。Ubuntu严格来说不能算一个独立的发行版本，Ubuntu是基于Debian的unstable版本加强而来的。Ubuntu就是一个拥有Debian所有的优点以及自己所加强的优点的近乎完美的Linux桌面系统，在服务端和桌面端使用占比最高，网络上资料最齐全，因此，本教程采用Ubuntu。



2.2.Linux系统安装

- 2.2.1 下载安装文件
- 2.2.2 Linux系统的安装方式
- 2.2.3 安装Linux虚拟机
- 2.2.4 生成Linux虚拟机镜像文件



2.2.1 下载安装文件

访问Ubuntu官网下载（<https://www.ubuntu.org.cn/download/ubuntu-kylin>）。进入Ubuntu官网的下载页面后（如图2-1所示），会提供两种不同版本的下载地址，即32位版本和64位版本。如果计算机硬件配置较低，内存小于2GB，建议选择32位版本；如果内存大于4GB，建议选择64位版本。

Download Ubuntu Kylin

Ubuntu Kylin 16.04.1 LTS

Download the long-term support edition of Ubuntu Kylin 16.04.1 LTS ISO image file.
To install Ubuntu Kylin, burn the image file on a DVD or create a bootable USB disk.

Download (64-bit)

Download (32-bit)

If you have an older PC with less than 2GB of memory, choose the 32-bit download.



2.2.2 Linux系统的安装方式

Linux系统的安装主要有两种方式：虚拟机安装方式和双系统安装方式。

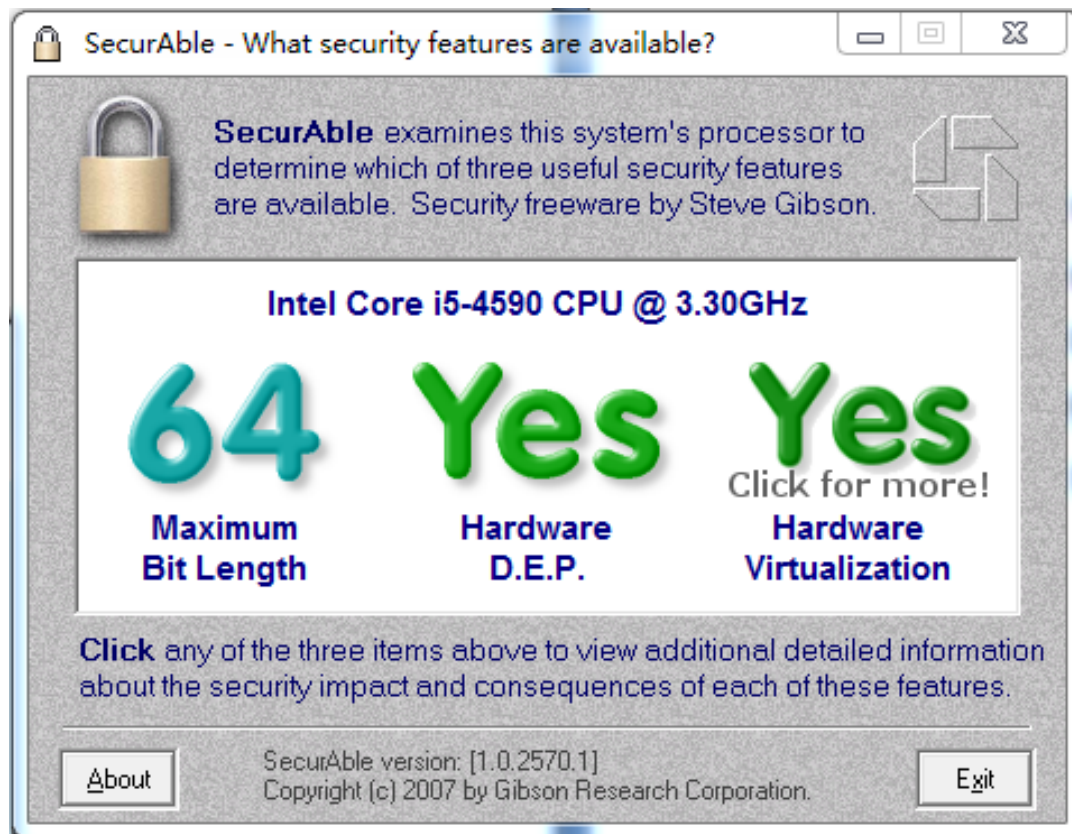
- (1) **虚拟机安装方式**：在Windows操作系统上安装虚拟机软件（比如VirtualBox或VMware），然后，在虚拟机软件之上安装Linux系统。采用这种安装方式时，Linux系统就相当于运行在Windows上的一个软件，如果要使用Linux系统，需要在电脑开机后，首先启动进入Windows操作系统，然后，在Windows操作系统中打开虚拟机软件（比如VirtualBox），最后，在虚拟机软件中启动Linux，之后才能使用Linux系统。
- (2) **双系统安装方式**：直接把Linux系统安装在电脑“裸机”上，而不是安装在Windows系统之上。采用这种安装方式时，Linux系统和Windows系统的地位是平等的，当电脑开机时，屏幕上会显示提示信息，让用户选择要启动的系统，如果用户选择Windows系统，电脑就继续启动进入Windows系统，如果用户选择Linux系统，电脑就继续启动进入Linux系统。



2.2.2 安装Linux虚拟机

1.检测CPU是否支持虚拟化

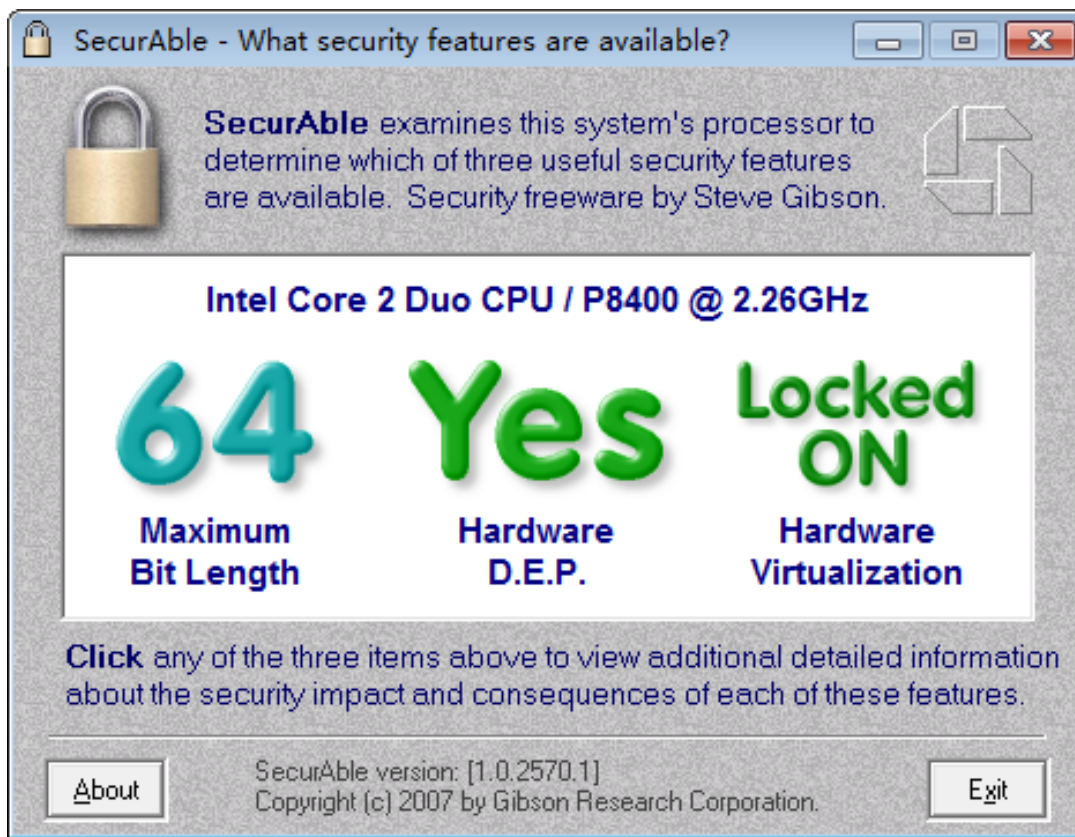
(1) CPU支持虚拟化





2.2.2 安装Linux虚拟机

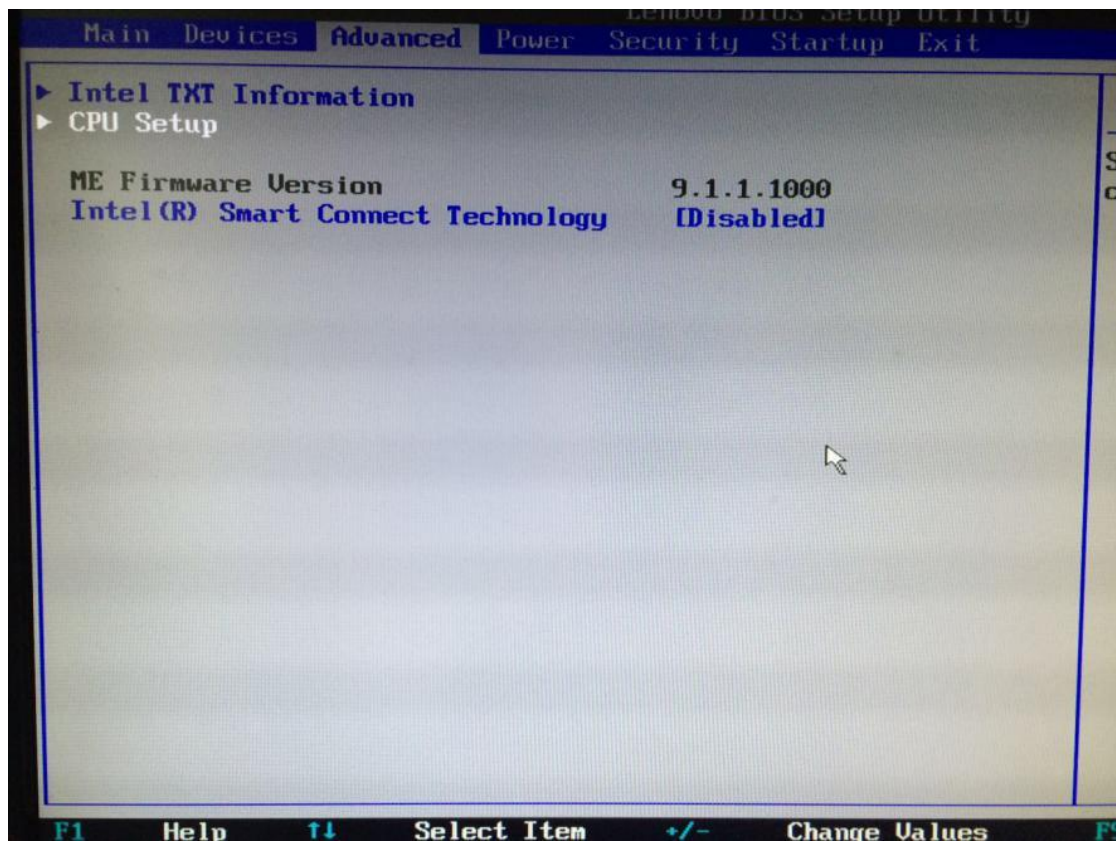
(2) 虚拟化在BIOS中没有开启





2.2.2 安装Linux虚拟机

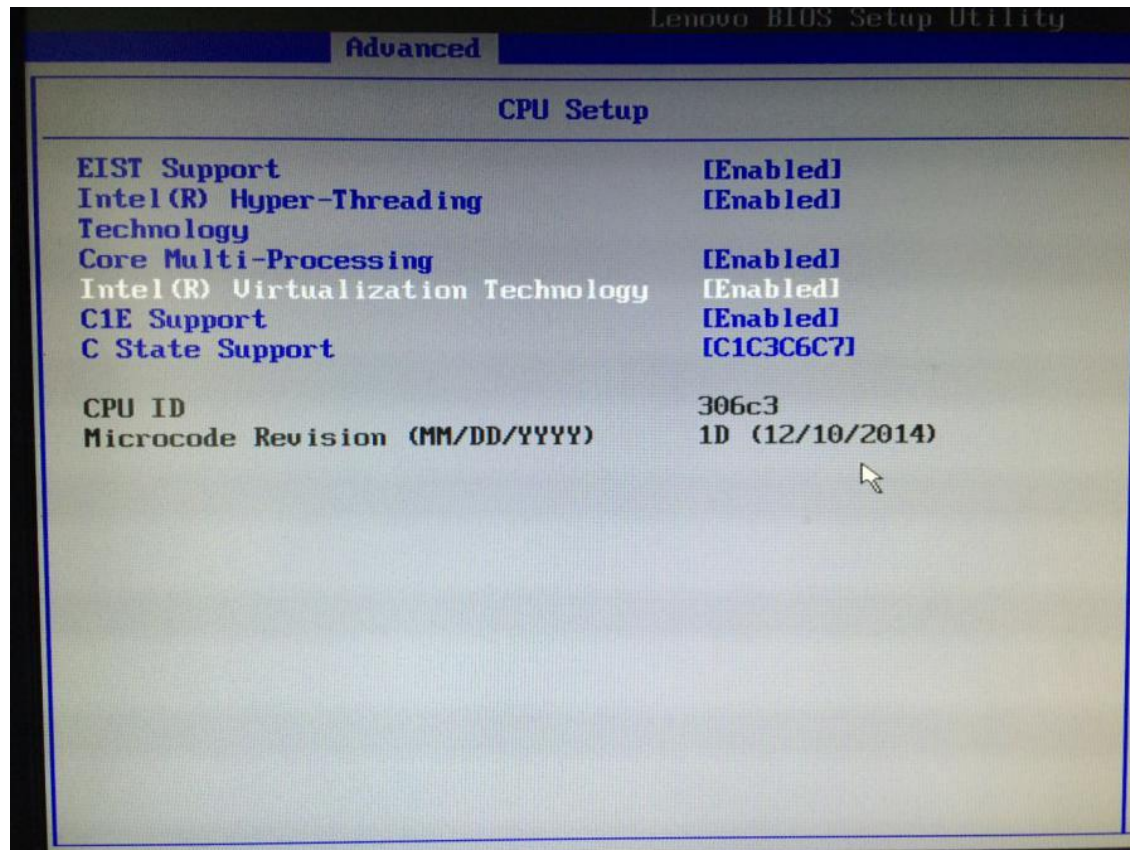
需要在BIOS中开启相关选项。关于如何修改BIOS以支持虚拟化，不同的电脑具有不同的方法。对于大部分电脑而言，只需要在开机的第一时间按下键盘上Del键即可进入BIOS系统。





2.2.2 安装Linux虚拟机

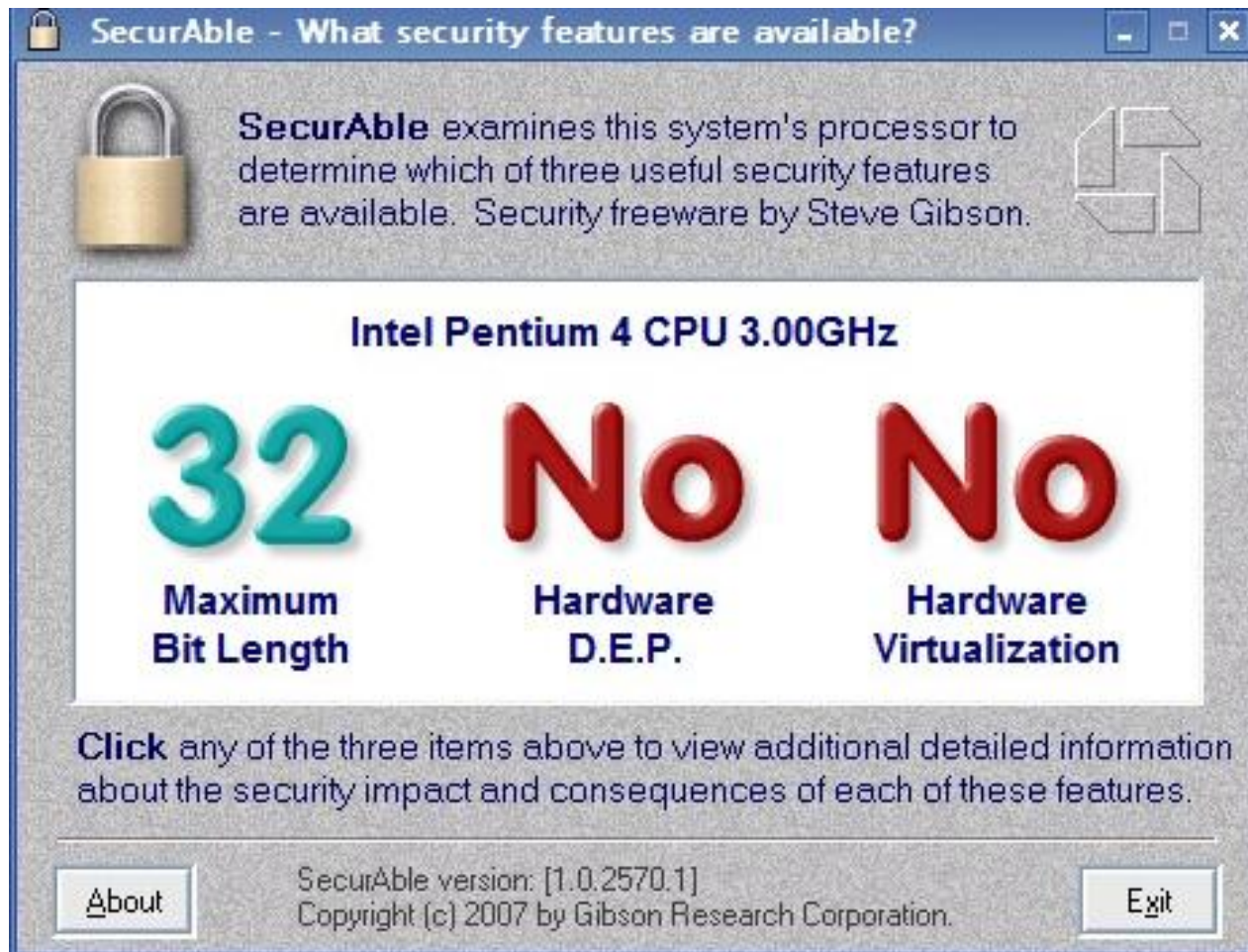
在该BIOS界面中，选择“CPU Setup”就可以进入到CPU的设置界面（如图2-5所示），需要把“Intel(R) Virtualization Technology”这个选项设置为“Enabled”，这样就开启了虚拟化功能。





2.2.2 安装Linux虚拟机

(3) CPU不支持虚拟化





2.2.2 安装Linux虚拟机

2.安装虚拟机软件

- 可以访问VirtualBox官网（<https://www.virtualbox.org/wiki/Downloads>）下载安装文件
- 有部分电脑无法成功安装VirtualBox，这时，请尝试获得商业软件VMware进行安装



2.2.2 安装Linux虚拟机

3. 安装Linux操作系统

(1) 创建虚拟机

(a) 步骤1: 新建一个虚拟机





2.2.2 安装Linux虚拟机

(b) 步骤2: 设置虚拟机名称和系统类型


新建虚拟机

虚拟机名称和系统类型

请选择新虚拟机的描述名称及要安装的操作系统类型。此名称将用于标识此虚拟机。

名称:

文件夹:

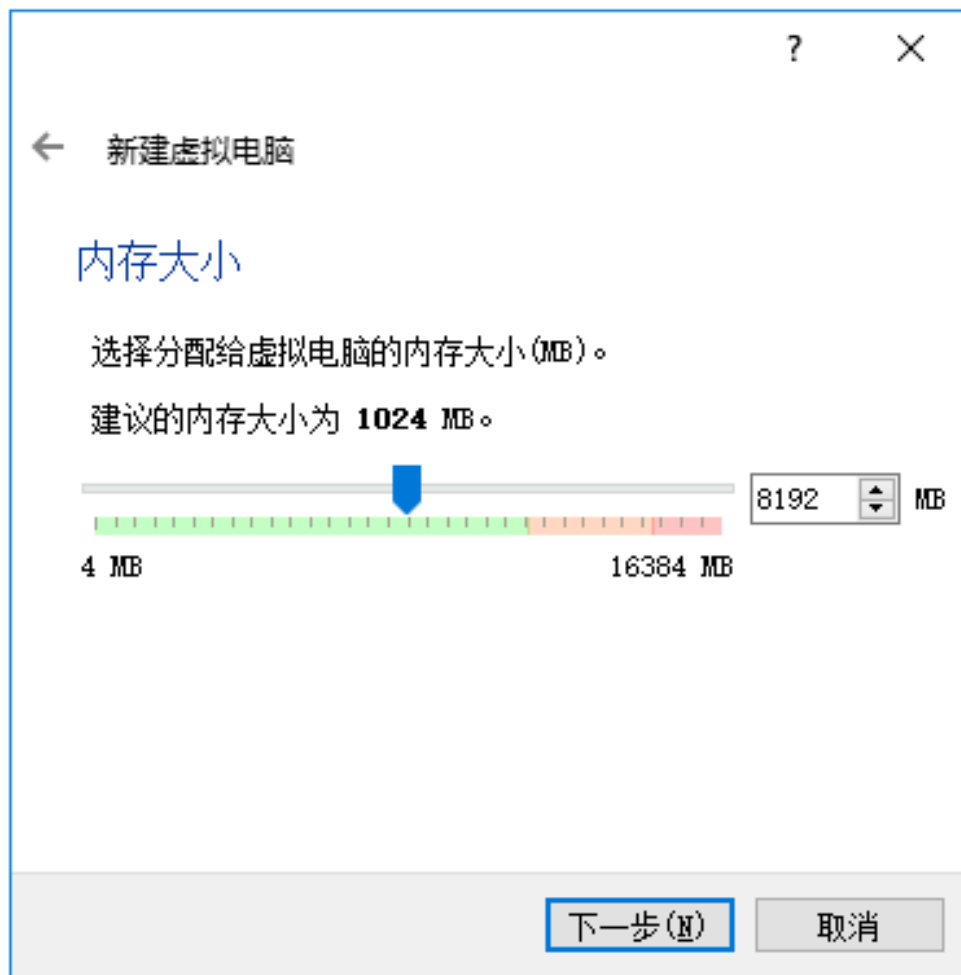
类型(T): 

版本(V):



2.2.2 安装Linux虚拟机

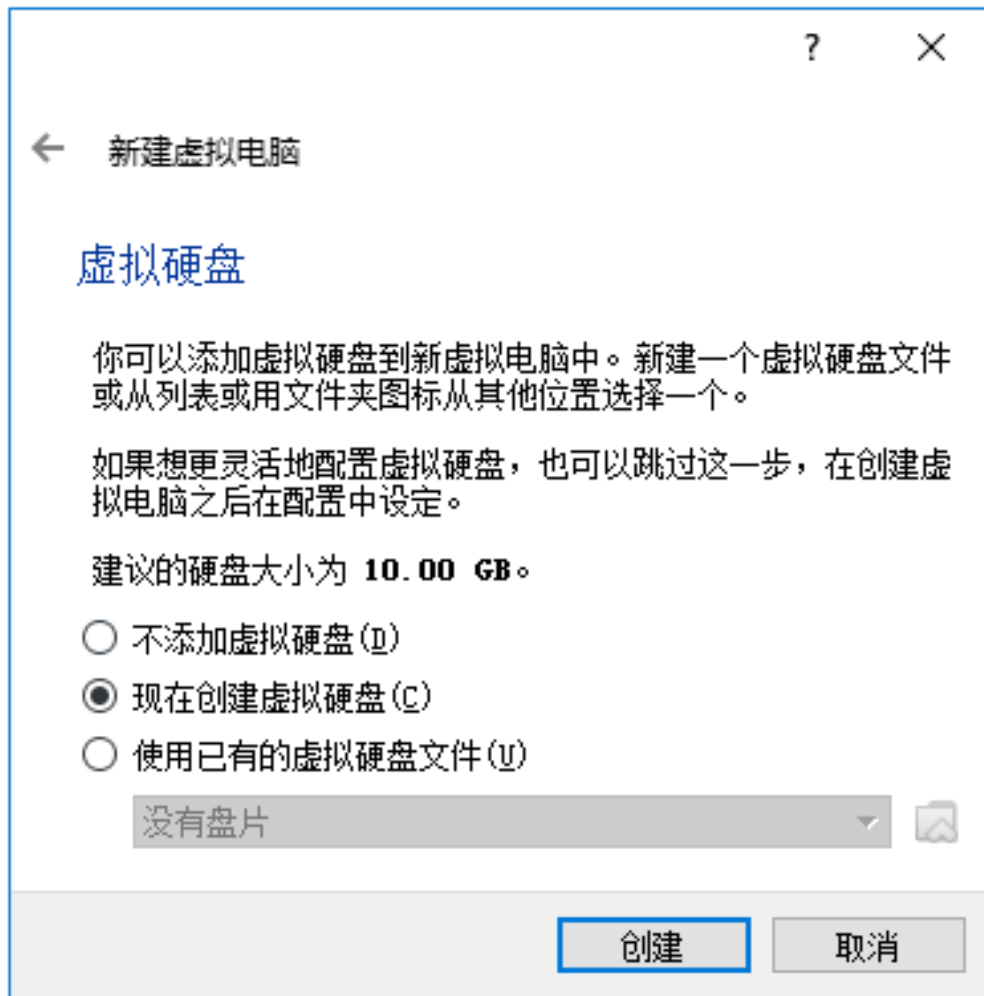
(c) 步骤3: 设置虚拟机内存大小





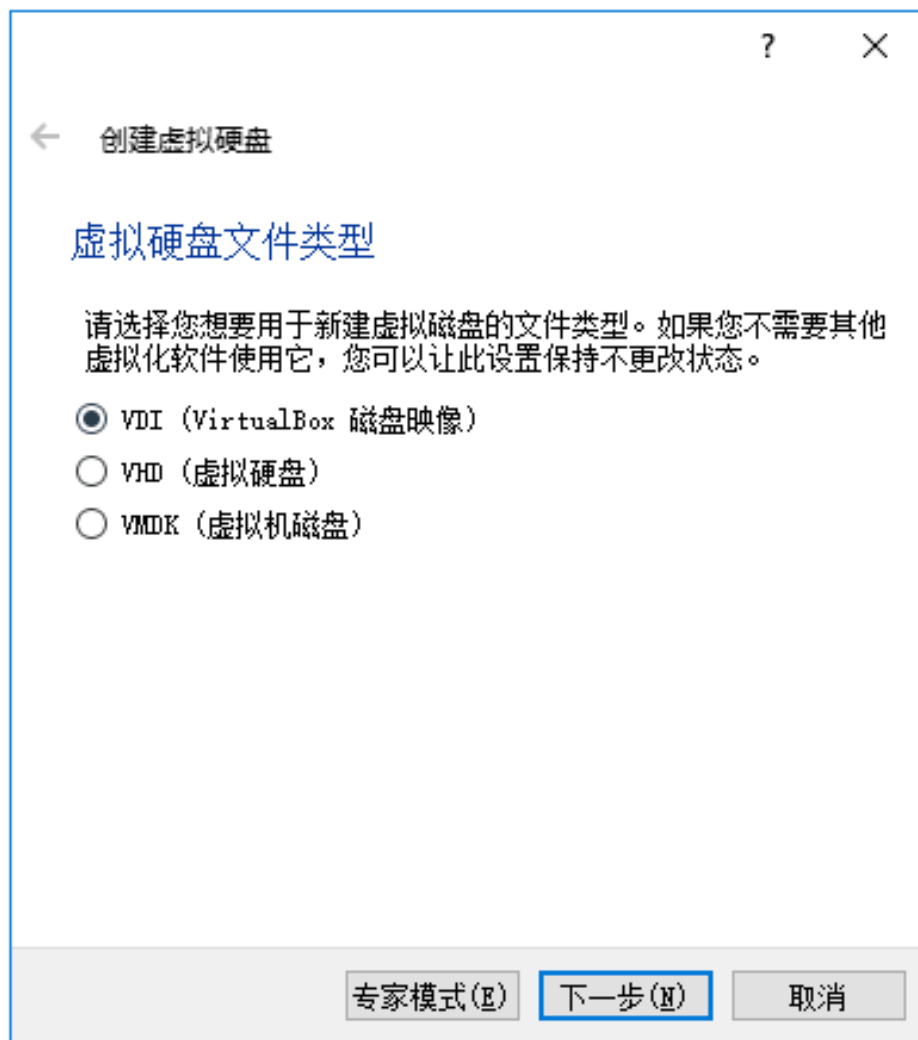
2.2.2 安装Linux虚拟机

(d) 步骤4: 创建虚拟硬盘



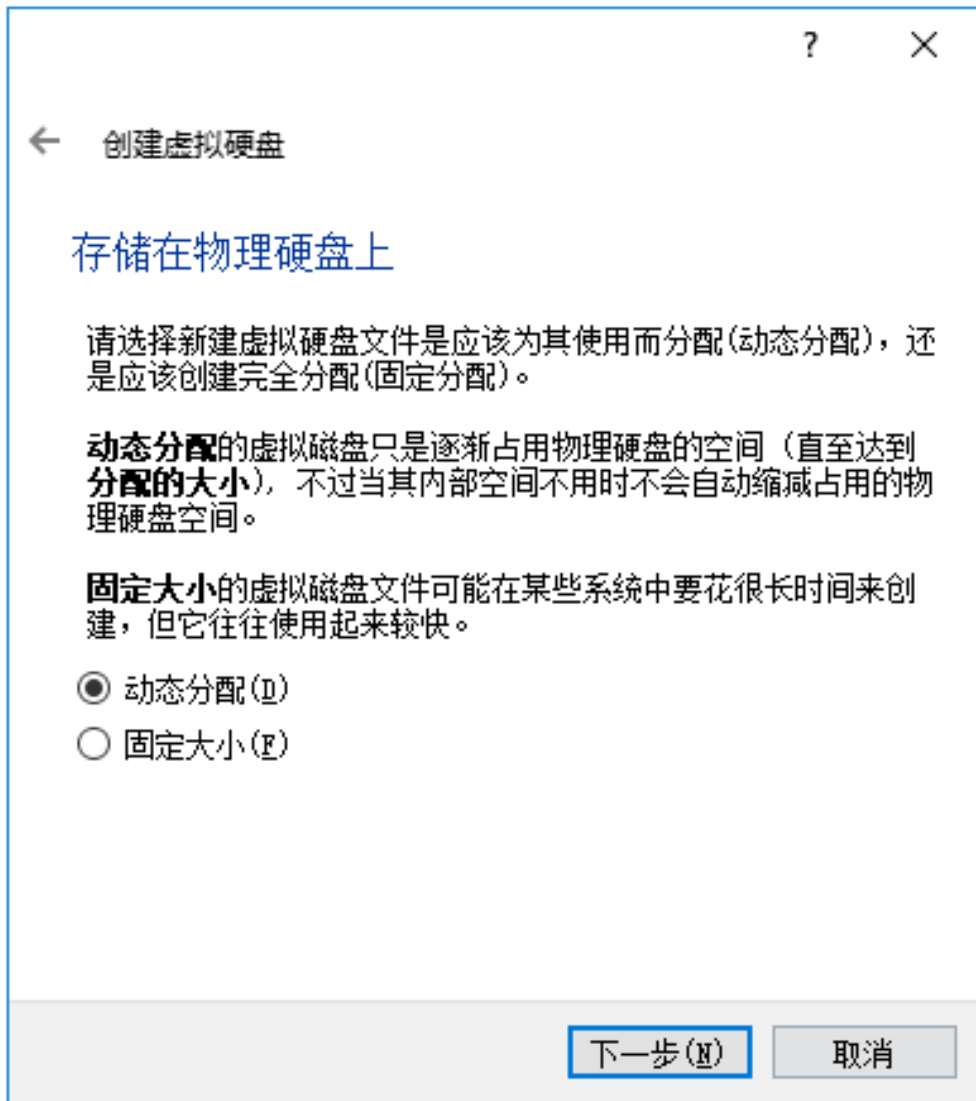


2.2.2 安装Linux虚拟机



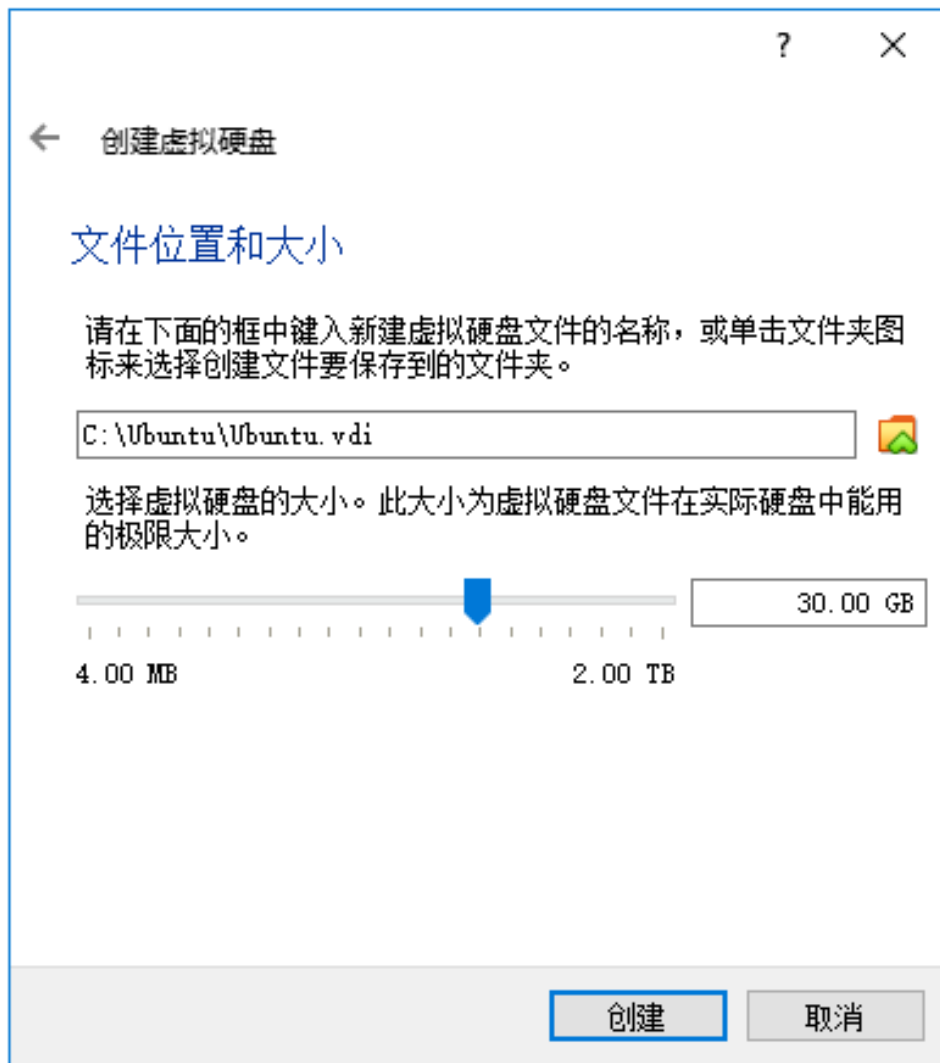


2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机





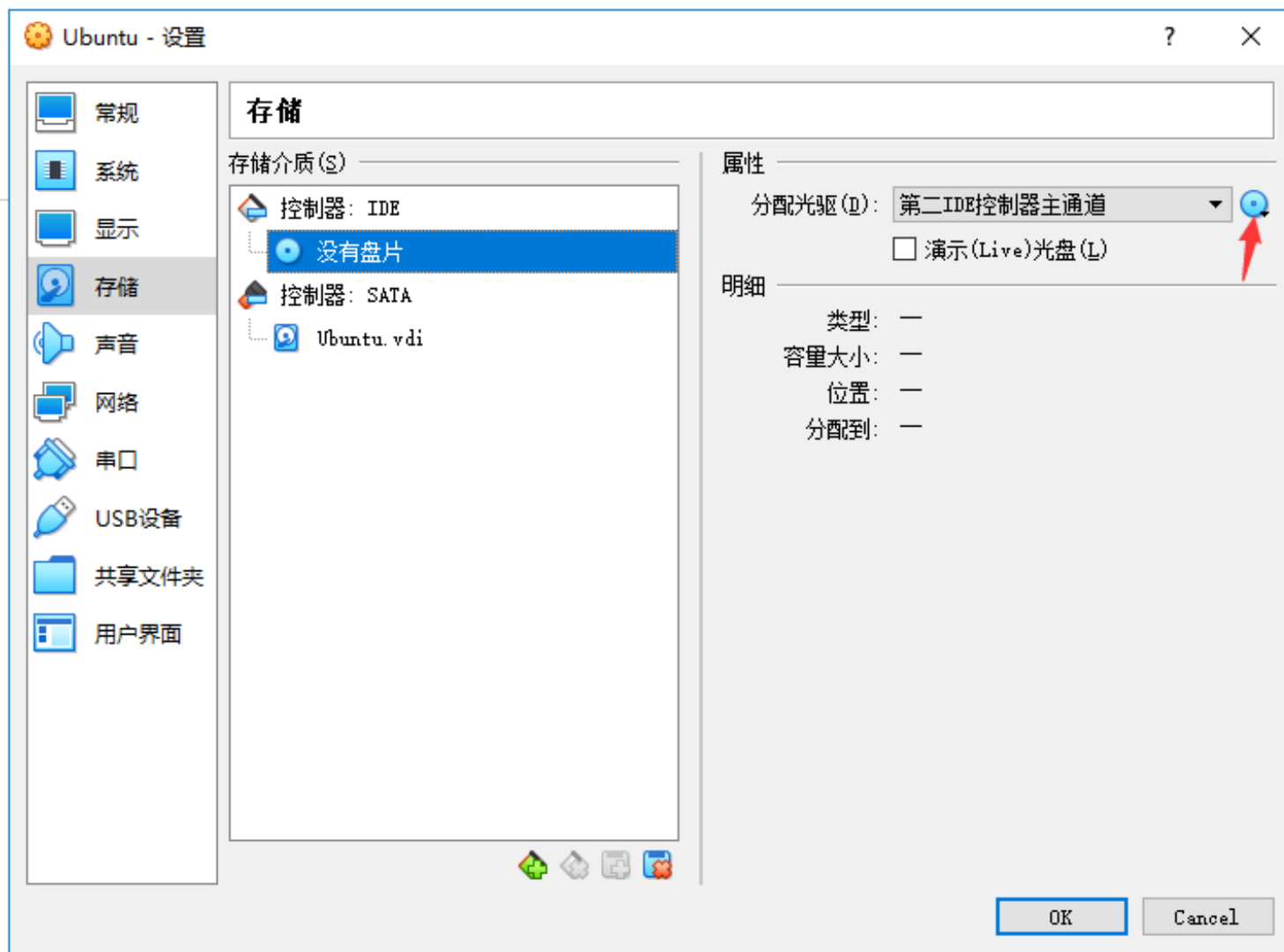
2.2.2 安装Linux虚拟机

(2) 在虚拟机上安装Linux系统





2.2.2 安装Linux虚拟机



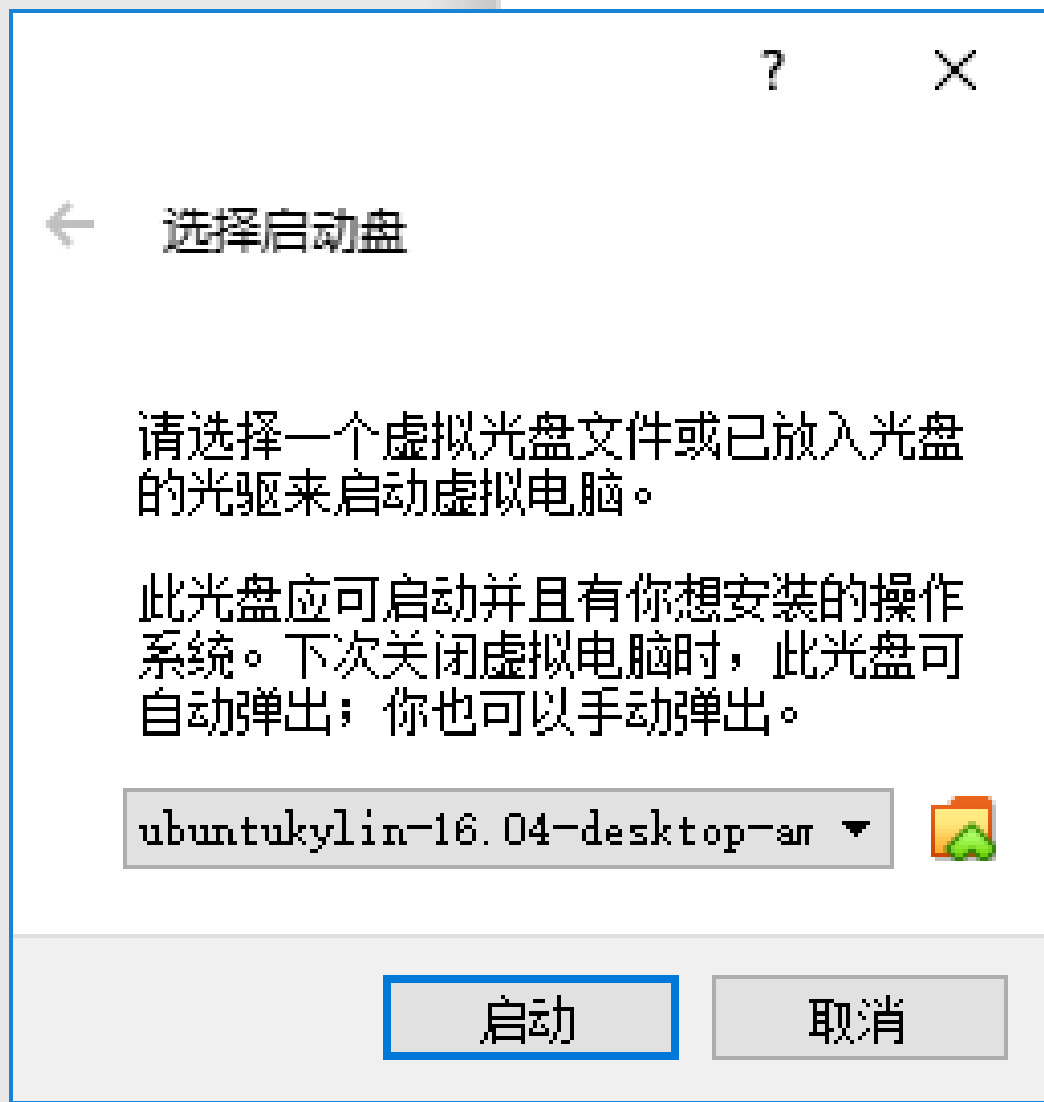


2.2.2 安装Linux虚拟机



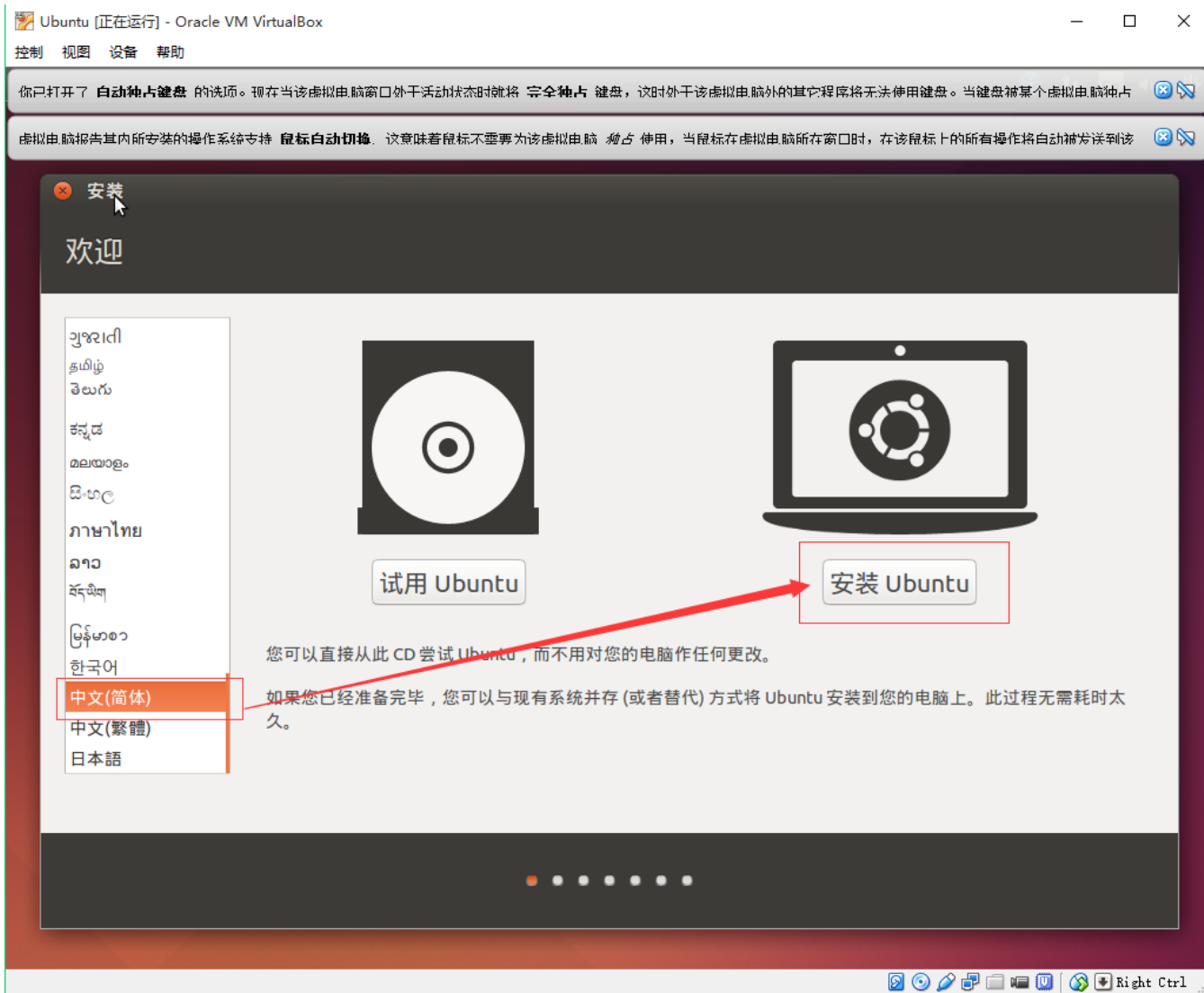


2.2.2 安装Linux虚拟机



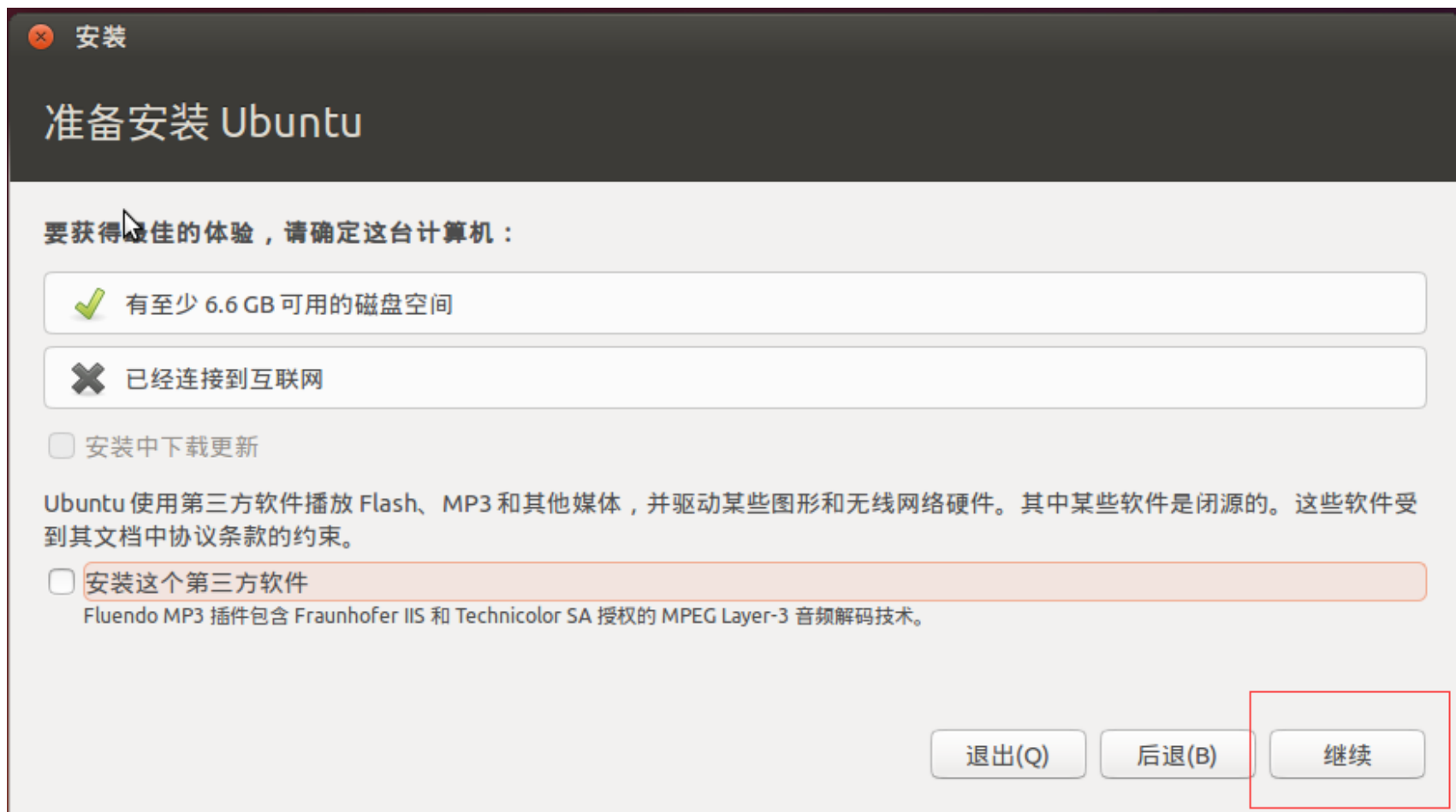


2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机



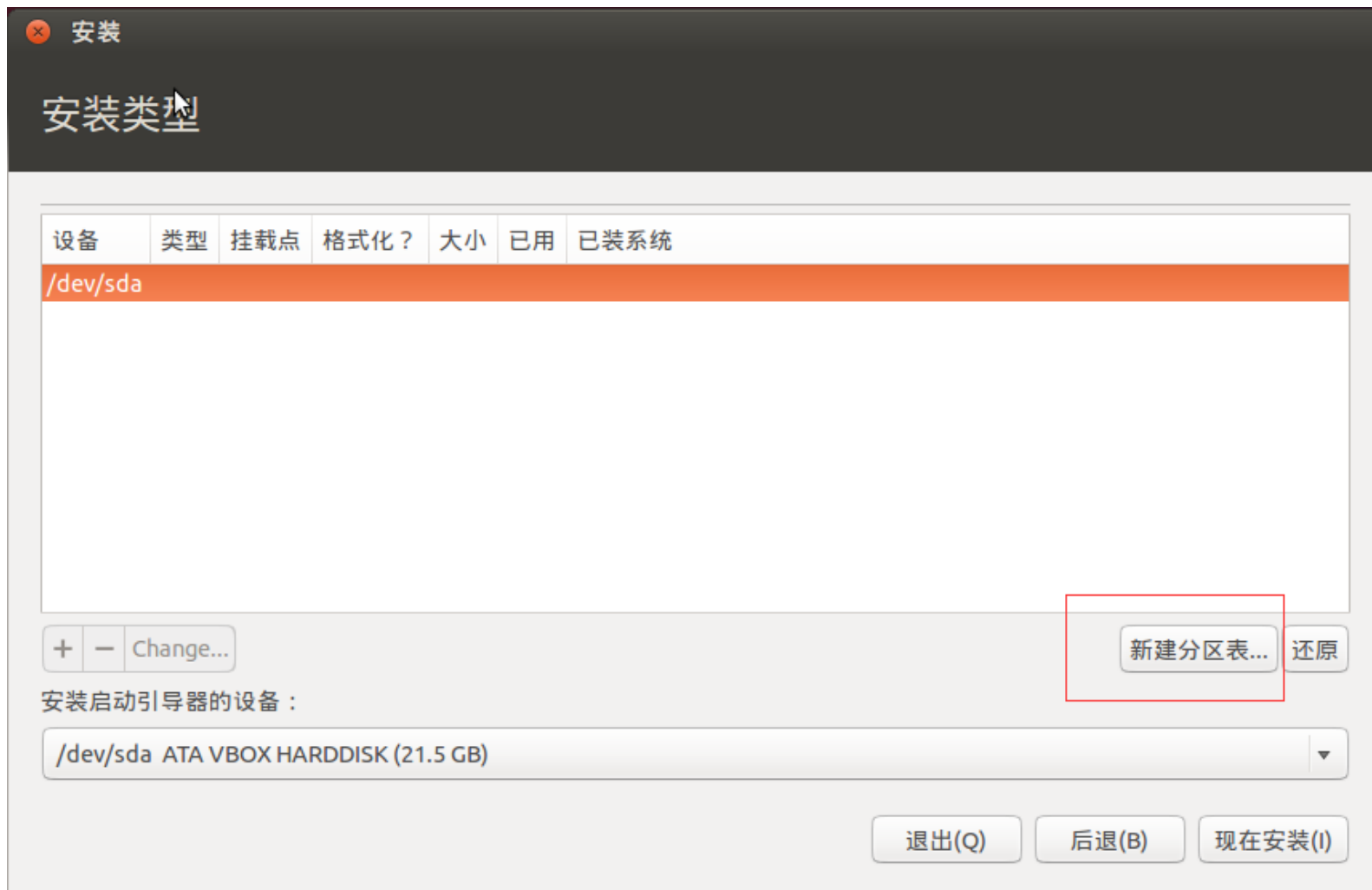


2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

✘ 要在此设备上创建新的空分区表吗？

您选择了对整个设备进行分区。如果您要在此设备上创建的分区表，那么现有的所有分区都将被删除。

注意：如果您希望的话，稍后依然可以撤消此动作。

后退

继续



2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

创建分区

大小： MB

新分区的类型： 主分区
 逻辑分区

新分区的位置： 空间起始位置
 空间结束位置

用于：



2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

创建分区

大小： 20963 - + MB

新分区的类型：
 主分区
 逻辑分区

新分区的位置：
 空间起始位置
 空间结束位置

用于： Ext4 日志文件系统 ▼

挂载点： / ▼

取消(C) 确定(O)



2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机



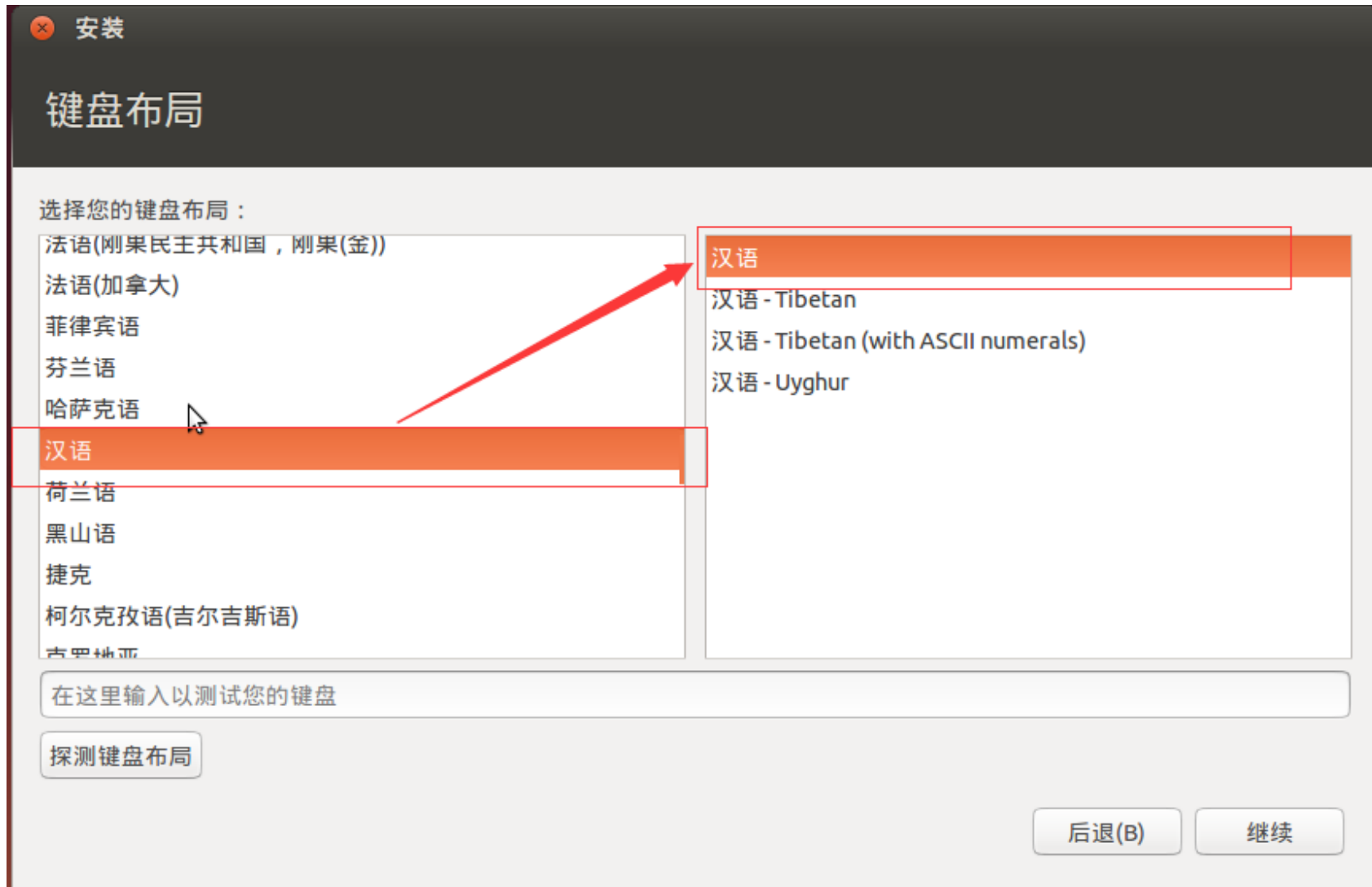


2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

安装

您是谁？

您的姓名： ✓

您的计算机名： ✓
与其他计算机联络时使用的名称。

选择一个用户名： ✓

选择一个密码： 密码强度：过短

确认您的密码： ✓

自动登录

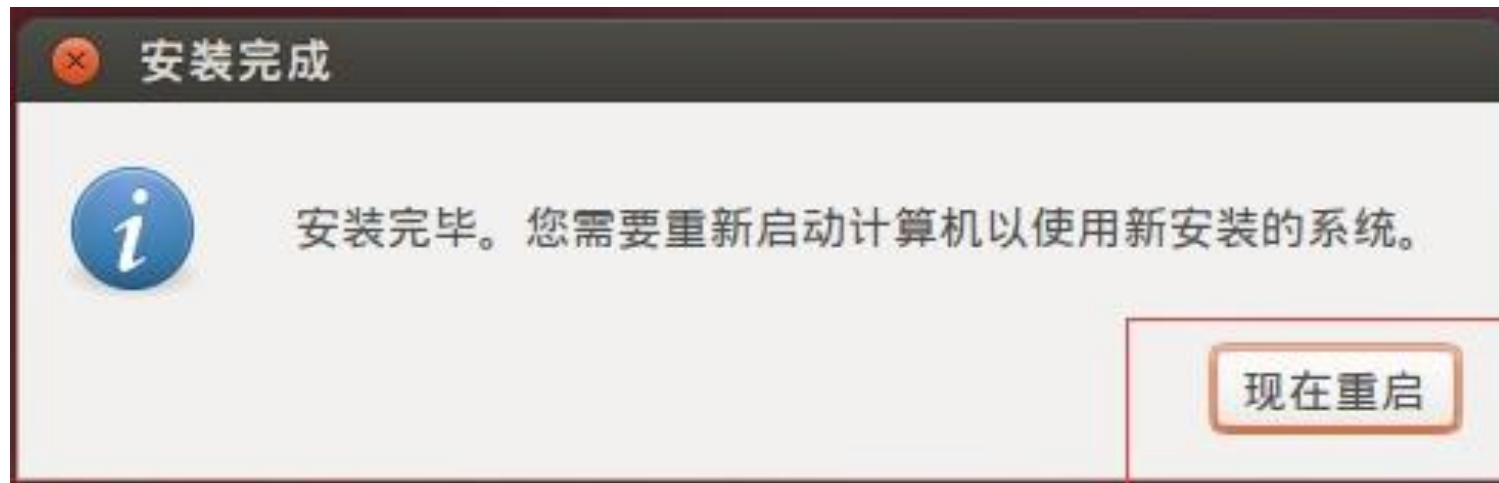
登录时需要密码

加密我的主目录

后退(B) 继续



2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

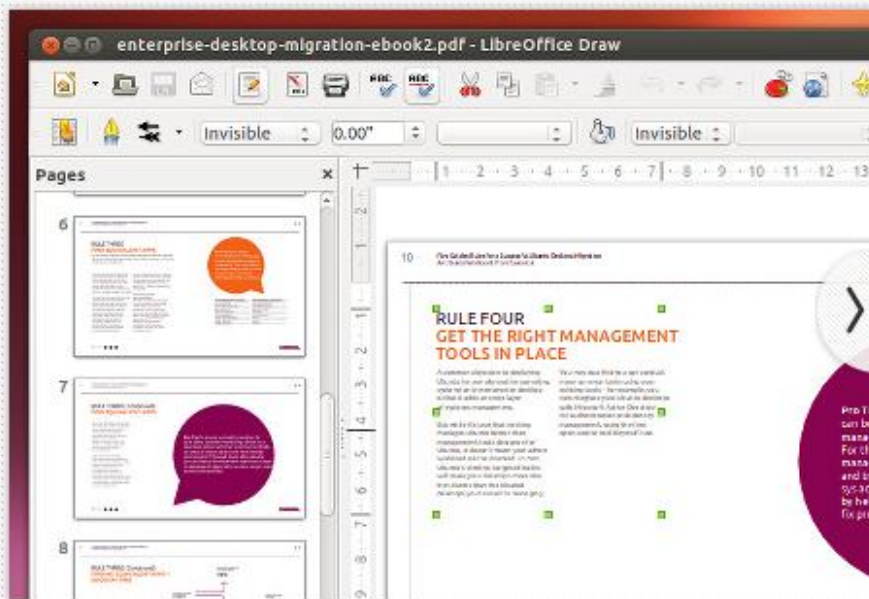
安装

办公所需 一应俱全

LibreOffice 是一个包括了您创建文档、电子表格与演示文稿一切所需的自由的办公套装软件。它与 Microsoft Office 文件格式兼容，并带给您需要的所有功能，而不需要一分钱。

预装软件

-  LibreOffice 文书处理
-  LibreOffice 电子表格
-  LibreOffice 演示文稿



▶ 正在下载语言包 (剩余 0:33)...

Skip



2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

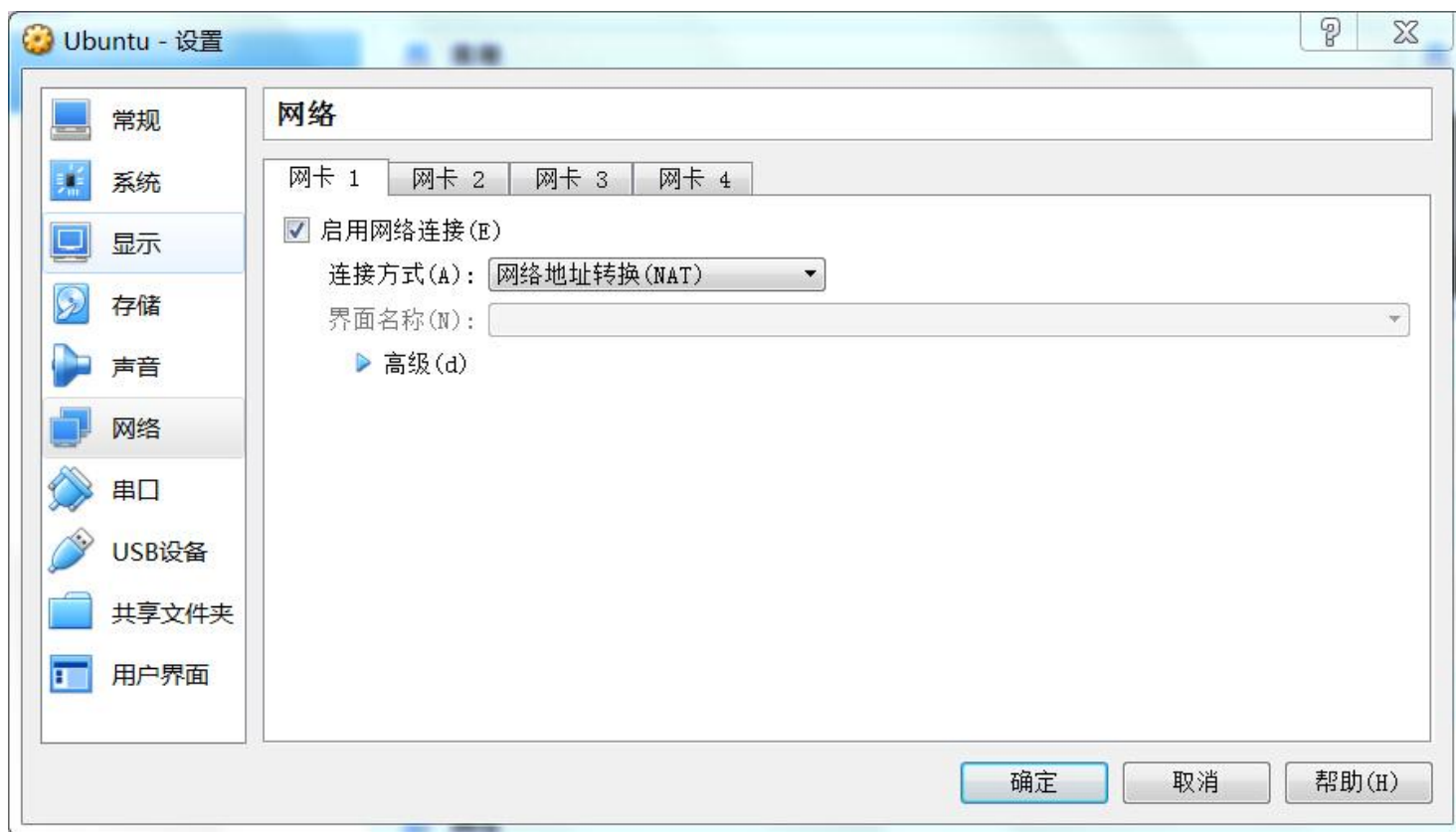




2.2.2 安装Linux虚拟机

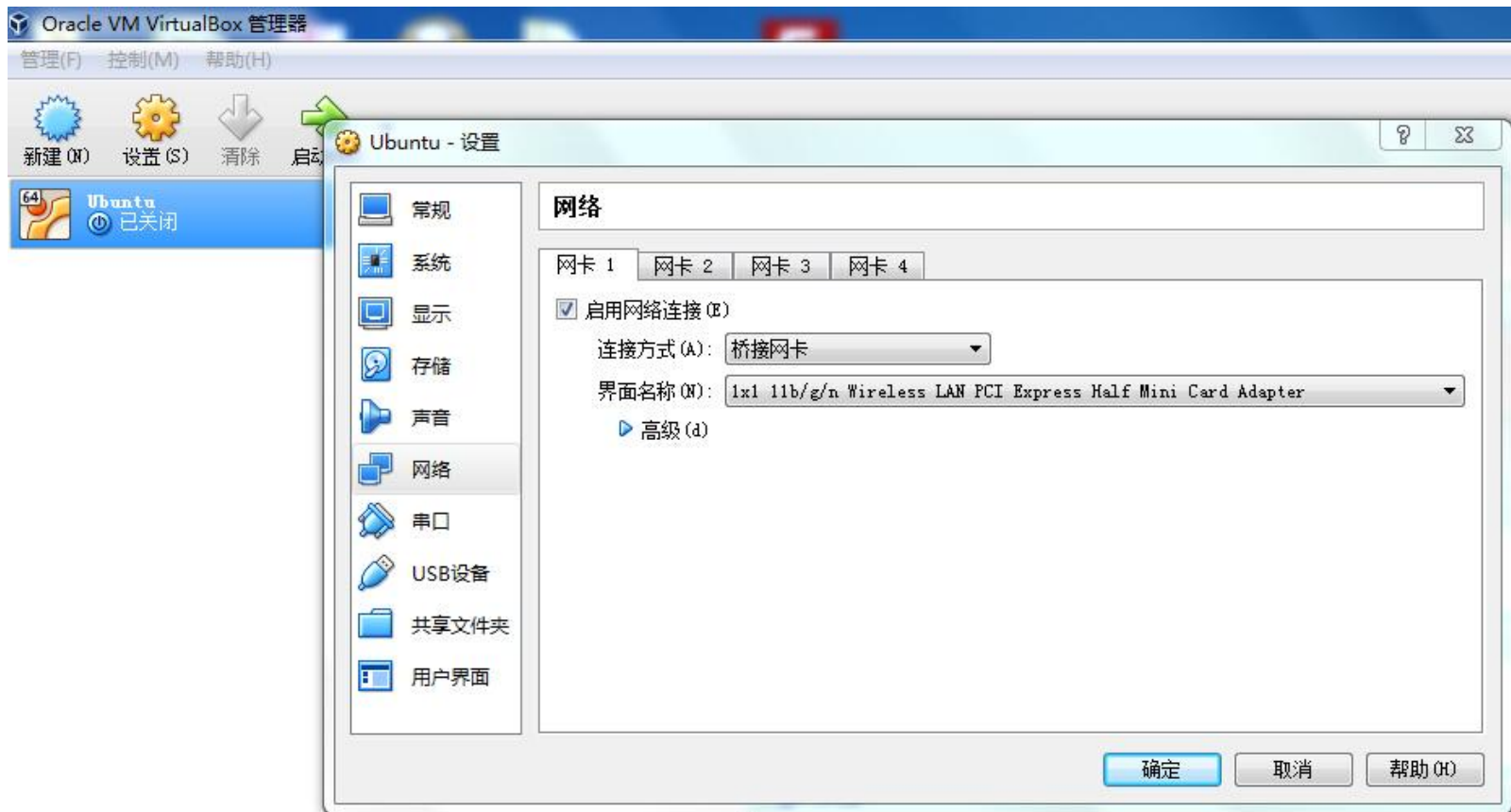
4.利用FTP软件向Ubuntu传输文件

(1) 设置网络连接方式





2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机

(2) 安装SSH

SSH是Secure Shell的缩写，它是建立在应用层基础上的安全协议，专为远程登录会话和其他网络服务提供安全性。利用SSH协议可以有效防止远程管理过程中的信息泄露问题。需要在Ubuntu上安装SSH服务端，才能让FTP软件通过SFTP（SSH File Transfer Protocol）方式连接Ubuntu系统。Ubuntu默认已安装了SSH客户端，因此，还需要安装SSH服务端，安装方法是，在Ubuntu系统中打开一个命令行终端，执行如下命令：

```
$sudo apt-get install openssh-server
```



2.2.2 安装Linux虚拟机

(3) 设置FTP软件

进入Shell命令提示符状态，输入命令“ifconfig”

```
终端 文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
hadoop@dblab-VirtualBox:~$ ifconfig
enp0s3  Link encap:以太网 硬件地址 08:00:27:de:ba:51
        inet 地址:192.168.0.104 广播:192.168.0.255 掩码:255.255.255.0
        inet6 地址: fe80::bd3f:14fe:f3f4:9725/64 Scope:Link
        UP BROADCAST RUNNING MULTICAST MTU:1500 跃点数:1
        接收数据包:43 错误:0 丢弃:0 过载:0 帧数:0
        发送数据包:116 错误:0 丢弃:0 过载:0 载波:0
        碰撞:0 发送队列长度:1000
        接收字节:7782 (7.7 KB) 发送字节:12293 (12.2 KB)

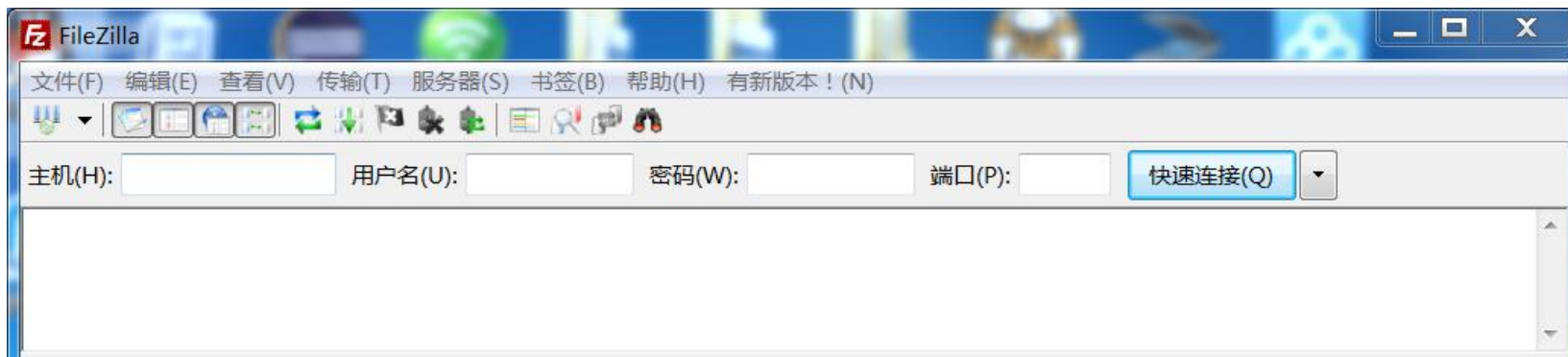
lo      Link encap:本地环回
        inet 地址:127.0.0.1 掩码:255.0.0.0
        inet6 地址: ::1/128 Scope:Host
        UP LOOPBACK RUNNING MTU:65536 跃点数:1
        接收数据包:234 错误:0 丢弃:0 过载:0 帧数:0
        发送数据包:234 错误:0 丢弃:0 过载:0 载波:0
        碰撞:0 发送队列长度:1
        接收字节:19651 (19.6 KB) 发送字节:19651 (19.6 KB)

hadoop@dblab-VirtualBox:~$
```



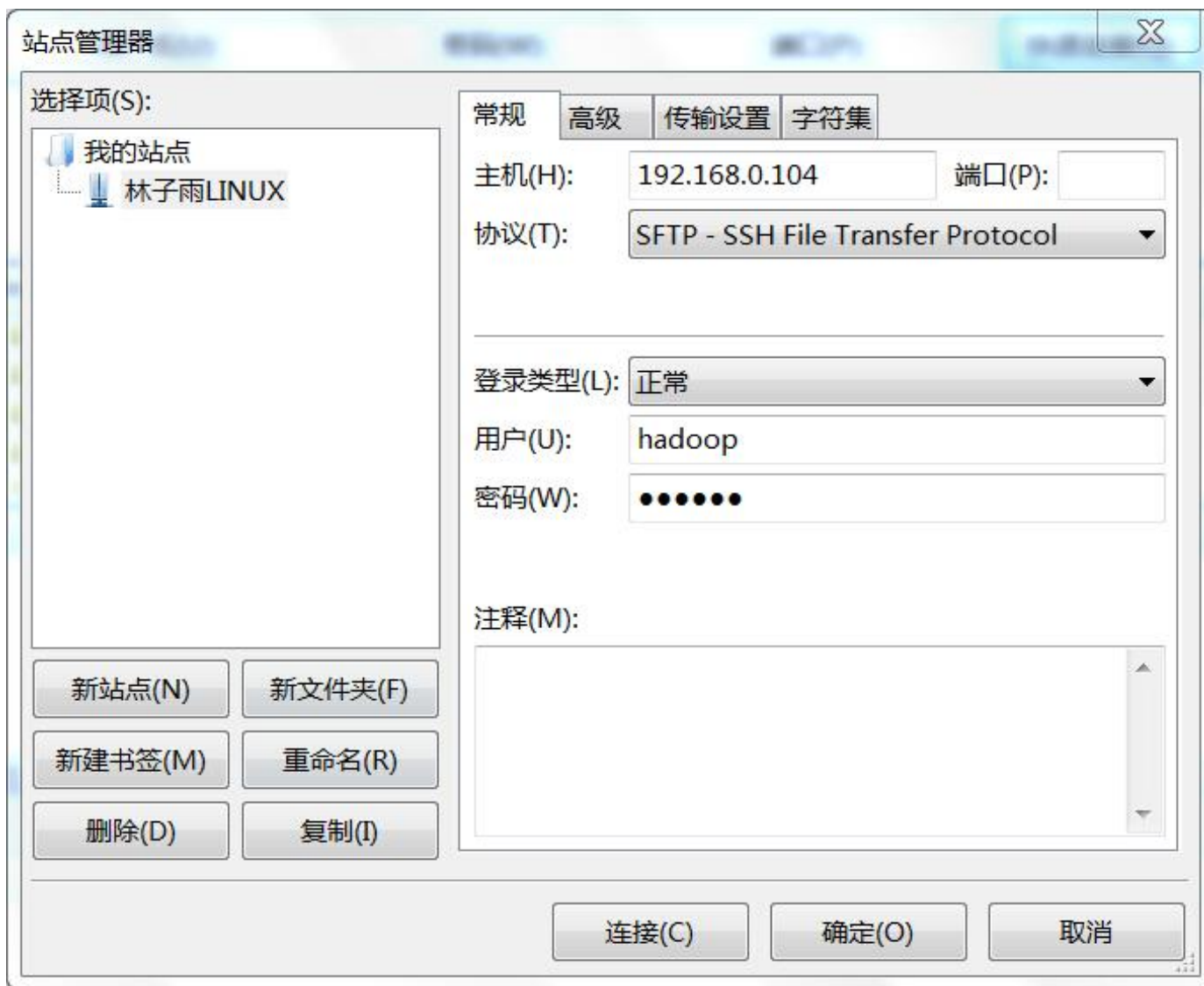
2.2.2 安装Linux虚拟机

获得Linux虚拟机IP地址信息以后，就可以使用FTP软件FileZilla连接Linux虚拟机了



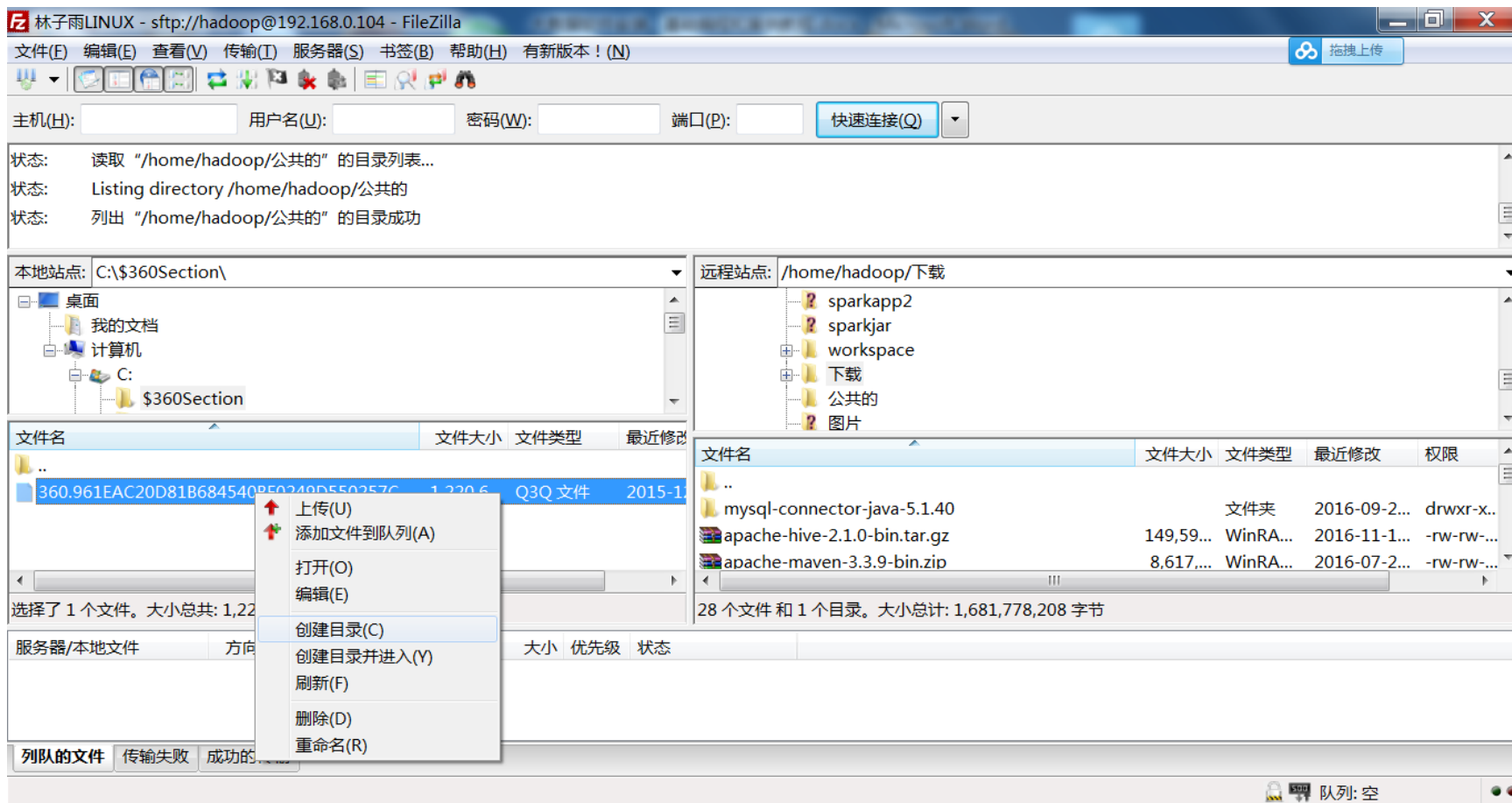


2.2.2 安装Linux虚拟机





2.2.2 安装Linux虚拟机





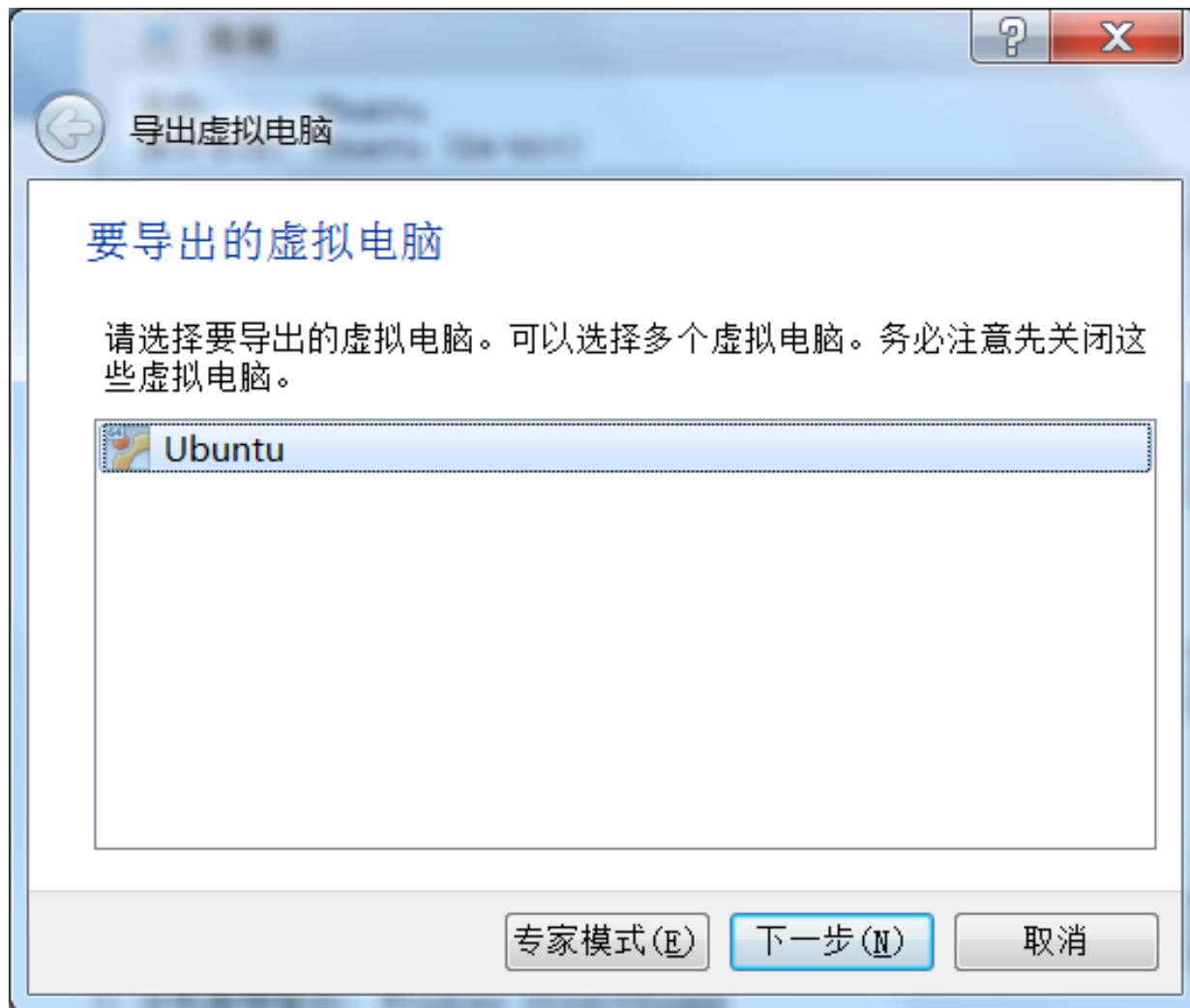
2.2.3 生成Linux虚拟机镜像文件

1. 导出生成镜像文件





2.2.3 生成Linux虚拟机镜像文件





2.2.3 生成Linux虚拟机镜像文件

虚拟电脑导出设置

请选择导出格式。

开放式虚拟化格式 仅支持 `ovf` 或 `ova` 扩展名。
若采用 `ovf` 格式，将分别存成多个独立文件。
若采用 `ova` 格式，这些文件将合并成一个开放式虚拟化格式包。

Oracle 公共云格式 仅支持 `tar.gz` 扩展名。
选中的虚拟机的主虚拟磁盘文件将保存到服务器上。

格式(O): 开放式虚拟化格式 1.0

请选择导出文件名。 另外可对导出文件大小和内容做一些设置。

文件(F): C:\Ubuntu.ova

MAC 地址设定(P): 仅包含 NAT 网卡的 MAC 地址

其它选项: 写入 Manifest 文件(W)

加入 ISO 映像文件(I)










2.2.3 生成Linux虚拟机镜像文件

虚拟系统设置

这是导出的虚拟电脑的描述信息。双击每行可以修改。

虚拟系统 1

 名称	Ubuntu16.04虚拟机镜像
 产品	高校大数据课程公共服务平台
 产品 URL	http://dblab.xmu.edu.cn/post/bigdatapRACTICE2/
 供应商	厦门大学数据库实验室
 供应商 URL	http://dblab.xmu.edu.cn/
 版本	2020年1月29日
 描述	
 许可	



2.2.3 生成Linux虚拟机镜像文件

2. 导入Linux虚拟机镜像文件

导入Linux虚拟机镜像文件的操作比较简单，打开虚拟机软件VirtualBox，点击界面左上角的“管理”菜单，在弹出的子菜单中选择“导入虚拟电脑”，然后，在弹出的界面中找到虚拟机镜像文件，点击“下一步”，在弹出的界面中点击“导入”按钮，就可以顺利完成导入，生成一个可用的Linux虚拟机。然后，就可以启动这个虚拟机，登录Linux系统。



2.3.Linux系统及相关软件基本使用方法

2.3.1 Shell

2.3.2 root用户

2.3.3 创建普通用户

2.3.4 sudo命令

2.3.5 常用的Linux系统命令

2.3.6 文件解压缩

2.3.7 常用的目录

2.3.8 目录的权限

2.3.9更新APT

2.3.10 切换中英文输入法

2.3.11 vim编辑器使用方法

2.3.12 在Windows系统中使用SSH方式登录Linux系统

2.3.13 在Linux中安装Eclipse

2.3.14 其他使用技巧



2.3.1 Shell

Shell是Linux系统的用户界面，提供了用户与Linux系统内核进行交互操作的一种接口，它接收用户输入的命令并把它送入内核去执行。当我们在Linux系统中打开一个终端（可以采用快捷键Ctrl+Alt+T）时，就进入了Shell命令提示符状态，在里面输入的用户命令，都会被送入Linux内核去执行。每个Linux系统的用户可以拥有自己的Shell，用以满足他们自己专门的Shell需要。

Shell也有多种不同的版本，主要有下列版本的Shell：

- Bourne Shell：由贝尔实验室开发；
- BASH：是GNU的Bourne Again Shell，是GNU操作系统上默认的Shell；
- Korn Shell：是对Bourne Shell的发展，在大部分内容上与Bourne Shell兼容；
- C Shell：是SUN公司Shell的BSD版本；
- Z Shell：Z是最后一个字母，也就是终极Shell，集成了BASH、ksh的重要特性，同时又增加了自己独有的特性。



2.3.2 root用户

- 对于Linux系统而言，超级用户一般命名为root，相当于Windows系统中的Administrator用户。
- root是系统中唯一的超级用户，具有系统中所有的权限，如启动或停止一个进程、删除或增加用户、增加或者禁用硬件等等。
- Linux中的root用户比Windows的Administrator用户的能力更大，足以把整个系统的大部分文件删掉，导致系统完全毁坏，不能再次使用。所以，用root进行不当的操作是相当危险的，轻微的可以造成死机，严重的甚至不能开机。因此，在实际使用中，除非确实需要，一般情况下都不推荐使用root用户登录Linux系统进行日常的操作。建议单独建立一个普通的用户，来学习大数据软件安装和开展编程实践。比如，本教程全部采用单独建立的hadoop用户来开展实验。



2.3.3 创建普通用户

打开一个终端（可以使用快捷键 **Ctrl+Alt+T**），使用如下命令创建一个用户 `hadoop`：

```
$ sudo useradd -m hadoop -s /bin/bash
```

使用如下命令为 `hadoop` 用户设置密码：

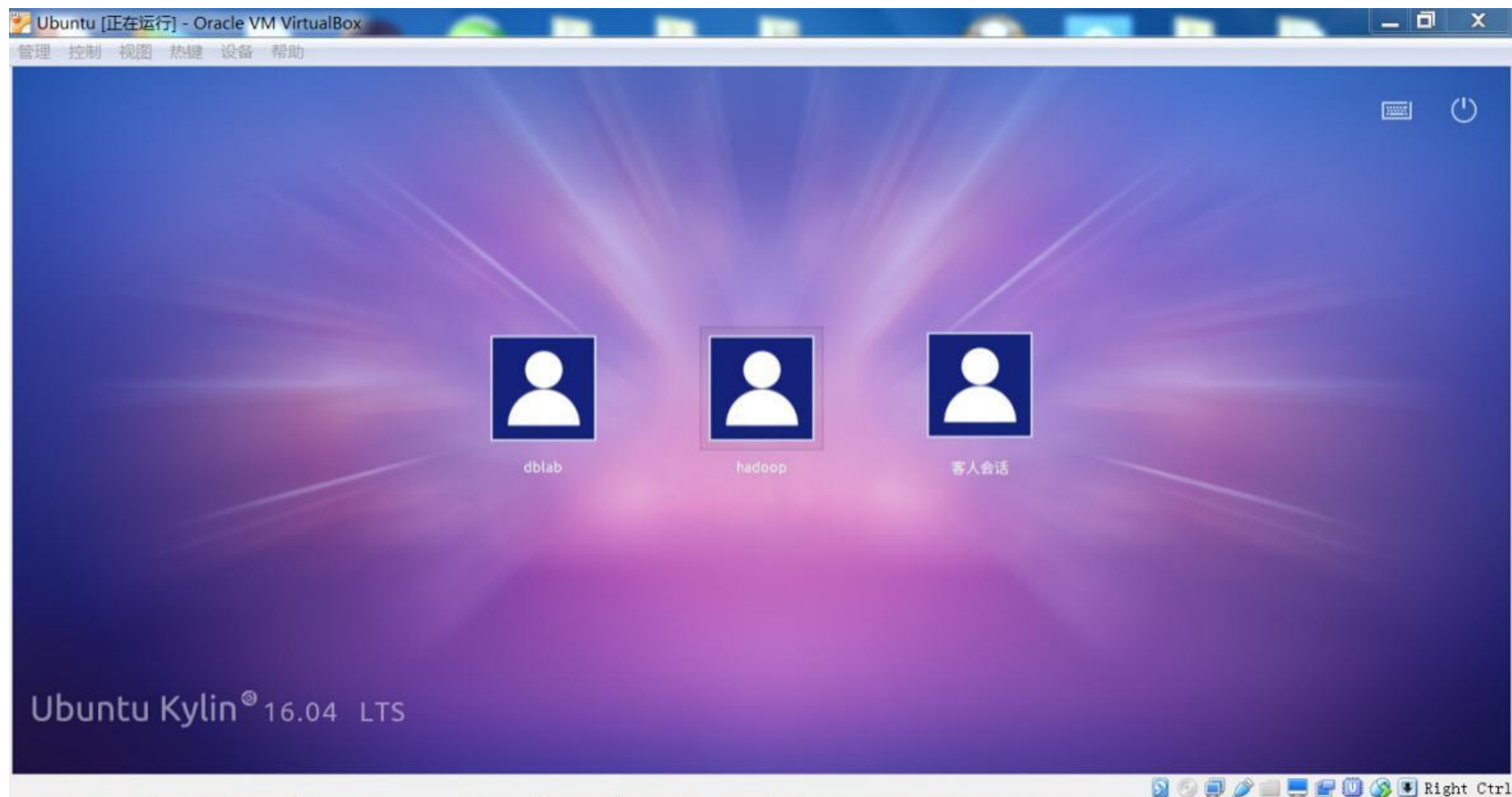
```
$ sudo passwd hadoop
```

为 `hadoop` 用户增加管理员权限：

```
$ sudo adduser hadoop sudo
```



2.3.3 创建普通用户





2.3.4 sudo命令

- **sudo**是Linux系统管理指令，管理员可以授权给一些普通用户去执行一些需要root权限执行的操作。这样不仅减少了root用户的登录和管理时间，同时也提高了安全性。本教程后面在执行软件安装时，都是采用hadoop用户登录，而不是root用户登录，因此，使用hadoop用户登录Linux系统以后，当需要执行只有root用户有权限执行的命令时，都要在命令前面加上**sudo**，才能够顺利执行，如果不加上**sudo**，就会被拒绝执行。
- 当使用**sudo**命令时，会要求输入当前用户的密码。需要注意的是，Windows系统中，用户输入密码，都会在屏幕上显示“*”号作为反馈，但是，在Linux系统中，当输入密码时，不会在屏幕上显示“*”号作为反馈，这时，不要误以为系统死机或者键盘出了问题，只要输完密码后回车即可。



2.3.5 常用的Linux系统命令

命令	含义
<code>cd /home/hadoop</code>	把/home/hadoop设置为当前目录
<code>cd ..</code>	返回上一级目录
<code>cd ~</code>	进入到当前Linux系统登录用户的主目录（或主文件夹）。在Linux系统中，~代表的是用户的主文件夹，即“/home/用户名”这个目录，如果当前登录用户名为hadoop，则~就代表“/home/hadoop/”这个目录
<code>ls</code>	查看当前目录中的文件
<code>ls -l</code>	查看文件和目录的权限信息
<code>cat /proc/version</code>	查看Linux系统内核版本信息
<code>cat /home/hadoop/word.txt</code>	把/home/hadoop/word.txt这个文件全部内容显示到屏幕上
<code>cat file1 file2 > file3</code>	把当前目录下的file1和file2两个文件进行合并生成文件file3



2.3.5 常用的Linux系统命令

<code>head -5 word.txt</code>	把当前目录下的word.txt文件中的前5行内容显示到屏幕上
<code>cp /home/hadoop/word.txt /usr/local/</code>	把/home/hadoop/word.txt文件复制到“/usr/local”目录下
<code>rm ./word.txt</code>	删除当前目录下的word.txt文件
<code>rm -r ./test</code>	删除当前目录下的test目录及其下面的所有文件
<code>rm -r test*</code>	删除当前目录下所有以test开头的目录和文件
<code>ifconfig</code>	查看本机IP地址信息
<code>exit</code>	退出并关闭Linux终端



2.3.6 文件解压缩

在Linux系统中，可以使用tar命令对后缀名为.tar.gz（或.tgz）的压缩文件进行解压。通常可以采用如下形式的命令：

```
$ tar -zxf /home/hadoop/下载/hbase-2.2.2-bin.tar.gz -C /usr/local
```



2.3.7 常用的目录

- Linux系统的根目录“/”下，存在很多个目录，其中有两个目录，是本教程学习过程中经常用到的，一个是“/home”目录，一个是“/usr”目录。“/home”目录包含了各个用户的用户目录，每当在Linux系统中新建一个普通用户时，系统就会自动为这个用户创建用户主目录（主文件夹）
- “/usr”目录是“Unix Software Resource”的简写，表示这里是各种软件安装的目录。对于“/usr”目录而言，只需要关注它下面的子目录“/usr/local”，一般由用户安装的软件都建议安装到该目录下



2.3.8 目录的权限

Linux系统对文件权限有着严格的规定，如果一个用户不具备权限，将无法访问目录及其下面的文件。比如，使用hadoop用户登录Linux系统以后，从网络上下载了HBase安装包文件，把文件解压缩到“/usr/local/”目录下，会得到一个类似“/usr/local/hbase”这样的目录，这时，hadoop用户并不是“/usr/local/hbase”这个目录的所有者，无法对该目录进行相关操作，从而无法正常使用HBase。这时，就必须采用chown命令进行授权，让hadoop用户拥有对该目录的权限，具体命令如下：

```
$ sudo chown -R hadoop /usr/local/hbase
```



2.3.9更新APT

APT是一个非常优秀的软件管理工具，Linux系统采用APT来安装和管理各种软件。安装成功Linux系统以后，需要及时更新APT软件，否则，后续一些软件可能无法正常安装。请登录Linux系统，打开一个终端（可以使用快捷键Ctrl+Alt+T），进入Shell命令提示符状态，然后输入下面命令：

```
$ sudo apt-get update
```







下载自：

<http://mirrors.aliyun.com/ubuntu>

可从光驱安装

Ubuntu 14.04 'Trusty Tahr' 光盘

官方支持
版权受限

还原(V)


关闭(C)



可用软件的列表信息已过时

要通过新加入或已变更的软件源来安装或更新软件包，您必须重新载入可用软件包列表。

您需要一个有效的互联网连接才能继续。

 重新载入(R)

关闭(C)



2.3.10 切换中英文输入法

Ubuntu中终端输入的命令一般都是使用英文输入。但是，有时候也会需要输入中文。Linux系统的中英文输入法的切换方式是使用键盘“**shift**”键来切换，或者，也可以点击系统桌面顶部菜单的输入法按钮进行切换。Ubuntu自带的Sunpinyin中文输入法已经可以很好满足中文输入要求，不需要额外安装中文输入法。



2.3.11 vim编辑器使用方法

执行下面命令完成vim编辑器的安装:

```
$ sudo apt-get install vim
```

可以执行下面命令创建一个word.txt文件:

```
$ cd ~  
$ vim word.txt
```



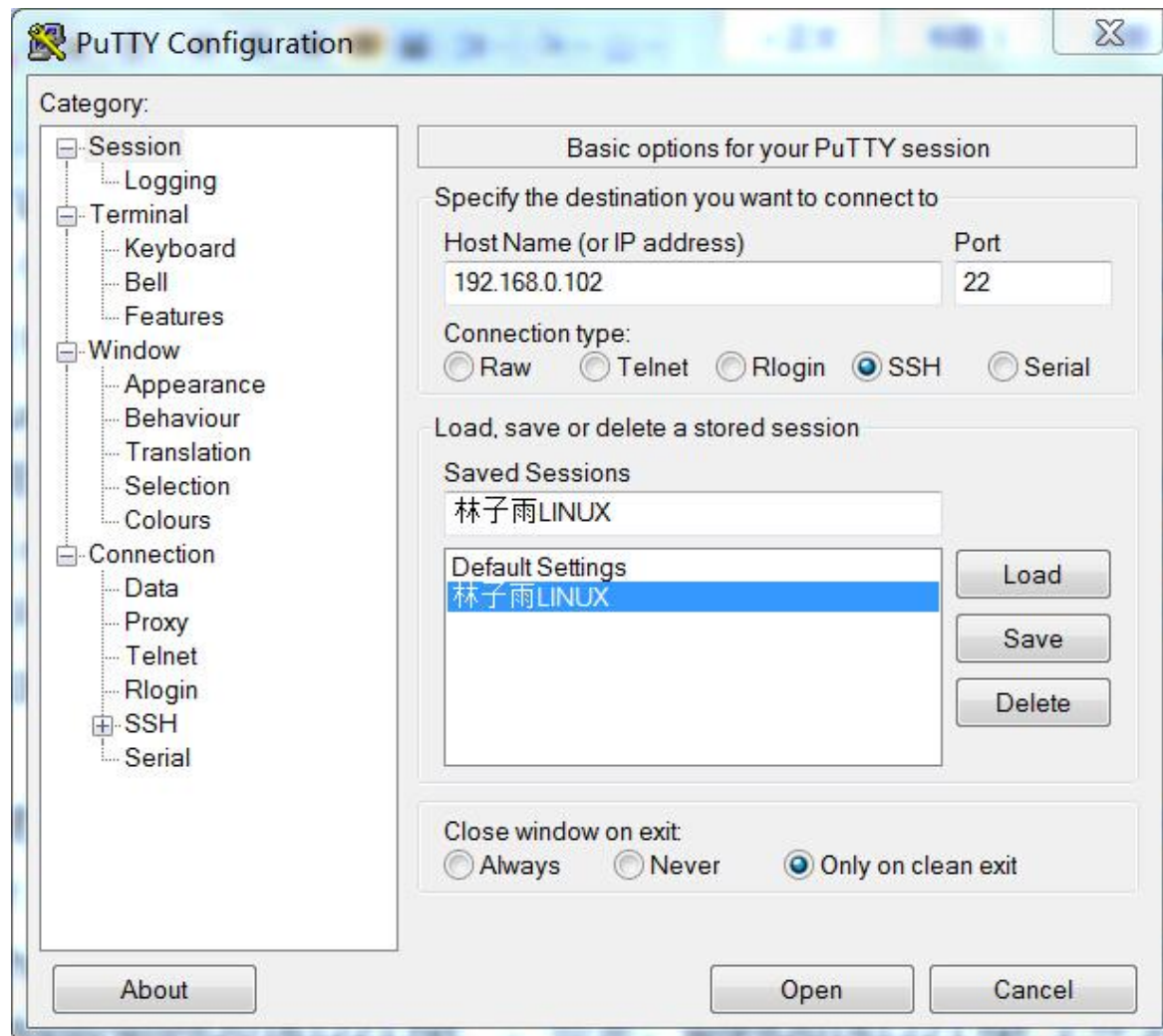

通过上面命令就打开了vim编辑器，然后，需要键盘敲击输入一个英文字母i，进入编辑状态以后才能修改内容，这时，就可以向word.txt文件中输入一些单词。修改后，需要按键盘上的Esc键退出vim的编辑状态，之后有以下几种选择：

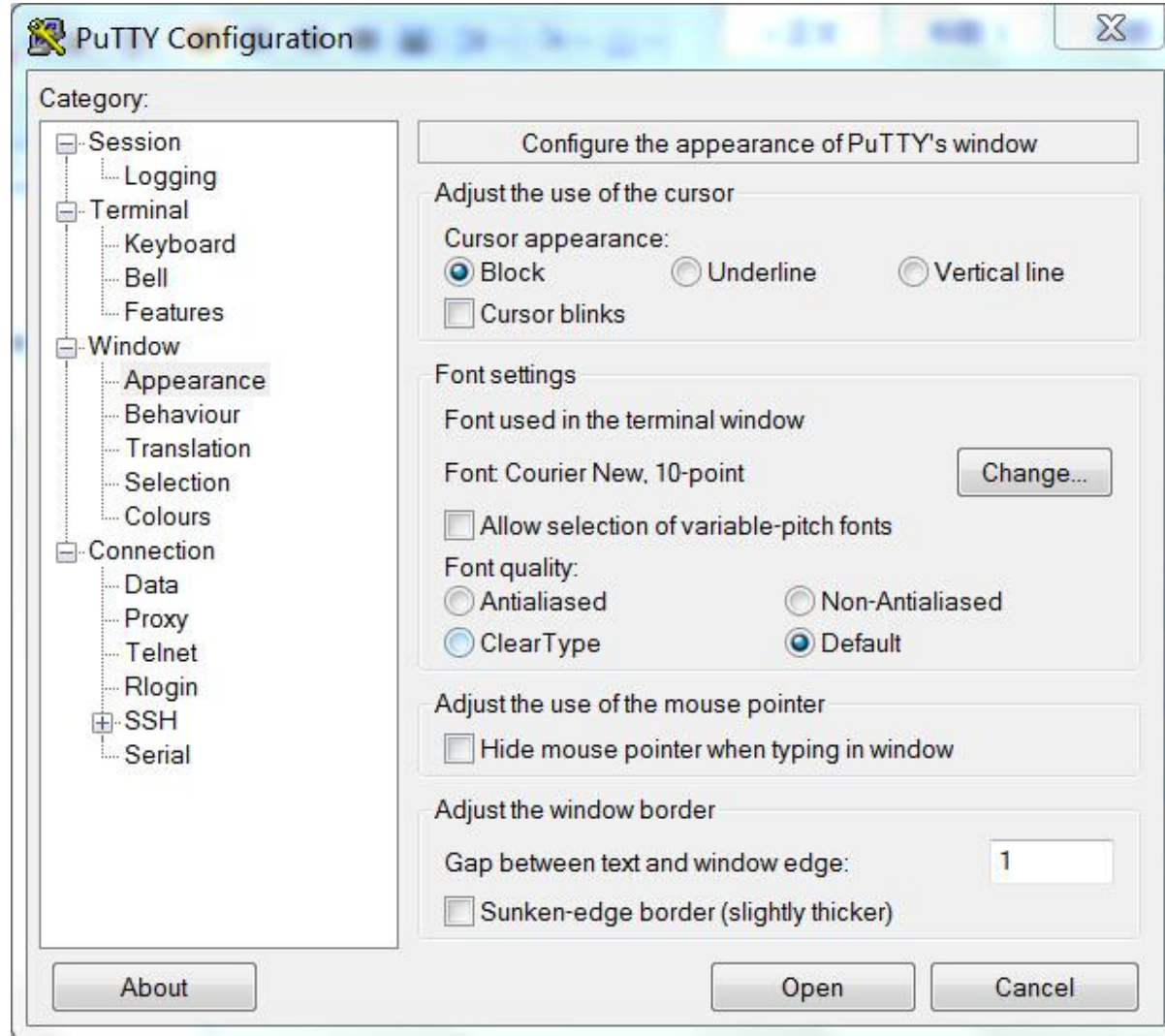
- 从键盘输入“:wq”三个英文字母，然后回车，表示保存文件并退出。
- 从键盘输入“:q”两个英文字母，然后回车，表示不保存并退出。如果本次编辑过程只是查看了文件内容，没有对文件做任何修改，则可以顺利退出。但是，如果已经修改了文件内容，则vim编辑器不允许就这样退出，会给出提示信息，这时，要想不保存就退出vim编辑器，就要采用下面一种方式，即输入“:q!”。
- 从键盘输入“:q!”三个英文字母，然后回车，表示不保存并强制退出。



2.3.12 在Windows系统中使用SSH方式登录Linux系统









```
hadoop@dmlab-VirtualBox ~  
login as: hadoop  
hadoop@192.168.0.102's password:  
Welcome to Ubuntu 16.04 LTS (GNU/Linux 4.4.0-57-generic x86_64)  
  
* Documentation: https://help.ubuntu.com/  
  
325 个可升级软件包。  
4 个安全更新。  
  
Last login: Thu Dec 22 08:28:13 2016 from 192.168.0.101  
hadoop@dmlab-VirtualBox:~$
```



2.3.13 在Linux中安装Eclipse

Eclipse是常用的程序开发工具，本教程很多程序代码都是使用Eclipse开发调试，因此，需要在Linux系统中安装Eclipse。

可以到Eclipse官网（<https://www.eclipse.org/downloads/>）下载安装包，或者直接访问教程官网下载该软件，位于“下载专区”的“软件”目录下，文件名是eclipse-4.7.0-linux.gtk.x86_64.tar.gz。假设安装文件下载后保存在了Linux系统的目录“~/Downloads”下，下面执行如下命令对文件进行解压缩：

```
$cd ~/Downloads  
$sudo tar -zxvf ./eclipse-4.7.0-linux.gtk.x86_64.tar.gz -C /usr/local
```

然后，执行如下命令启动Eclipse：

```
$cd /usr/local/eclipse  
$./eclipse
```



2.3.14 其他使用技巧

1. 使用**Tab**键自动补全命令
2. 隐藏文件
3. 重现历史命令

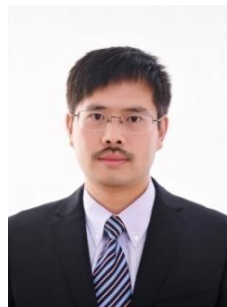


2.5 本章小结

- 本教程几乎所有软件（除了可视化工具Tableau以外），都是安装和运行在Linux操作系统之上，因此，顺利安装好Linux系统并且掌握Linux系统的基本使用方法，是开展后续章节内容学习的前提和基础。
- Linux系统可以采用双系统方式安装，也可以采用虚拟机方式安装，建议在大数据初学阶段采用虚拟机方式安装。本章详细介绍了如何安装Linux虚拟机，并给出了生成Linux虚拟机镜像文件的方法，可以实现Linux虚拟机的快速复制安装。另外，为了帮助读者更好开展Linux系统下的相关操作和实践，本章简要介绍了Linux系统及其相关软件的基本使用方法。最后，本章给出了关于本教程的一些约定，了解这些约定，有助于准确理解后续章节的内容。



附录A：主讲教师林子雨简介



主讲教师：林子雨

单位：厦门大学计算机科学系

E-mail: ziyulin@xmu.edu.cn

个人网页: <http://dblab.xmu.edu.cn/post/linziyu>

数据库实验室网站: <http://dblab.xmu.edu.cn>



扫一扫访问个人主页

林子雨，男，1978年出生，博士（毕业于北京大学），全国高校知名大数据教师，现为厦门大学计算机科学系副教授，曾任厦门大学信息科学与技术学院院长助理、晋江市发展和改革局副局长。中国计算机学会数据库专业委员会委员，中国计算机学会信息系统专业委员会委员。国内高校首个“数字教师”提出者和建设者，厦门大学数据库实验室负责人，厦门大学云计算与大数据研究中心主要建设者和骨干成员，2013年度、2017年度和2020年度厦门大学教学类奖教金获得者，荣获2019年福建省精品在线开放课程、2018年厦门大学高等教育成果特等奖、2018年福建省高等教育教学成果二等奖、2018年国家精品在线开放课程。主要研究方向为数据库、数据仓库、数据挖掘、大数据、云计算和物联网，并以第一作者身份在《软件学报》《计算机学报》和《计算机研究与发展》等国家重点期刊以及国际学术会议上发表多篇学术论文。作为项目负责人主持的科研项目包括1项国家自然科学基金青年基金项目(No.61303004)、1项福建省自然科学基金青年基金项目(No.2013J05099)和1项中央高校基本科研业务费项目(No.2011121049)，主持的教改课题包括1项2016年福建省教改课题和1项2016年教育部产学协作育人项目，同时，作为课题负责人完成了国家发改委城市信息化重大课题、国家物联网重大应用示范工程区域试点泉州市工作方案、2015泉州市互联网经济调研等课题。中国高校首个“数字教师”提出者和建设者，2009年至今，“数字教师”大平台累计向网络免费发布超过1000万字高价值的研究和教学资料，累计网络访问量超过1000万次。打造了中国高校大数据教学知名品牌，编著出版了中国高校第一本系统介绍大数据知识的专业教材《大数据技术原理与应用》，并成为京东、当当网等网店畅销书籍；建设了国内高校首个大数据课程公共服务平台，为教师教学和学生学习大数据课程提供全方位、一站式服务，年访问量超过200万次，累计访问量超过1000万次。



附录B：大数据学习路线图



大数据学习路线图访问地址：<http://dblab.xmu.edu.cn/post/10164/>



附录C：林子雨大数据系列教材



林子雨大数据系列教材

用于导论课、专业课、实训课、公共课

了解全部教材信息：<http://dbllab.xmu.edu.cn/post/bigdatabook/>



附录D：《大数据导论（通识课版）》教材

开设全校公共选修课的优质教材



本课程旨在实现以下几个培养目标：

- 引导学生步入大数据时代，积极投身大数据的变革浪潮之中
- 了解大数据概念，培养大数据思维，养成数据安全意识
- 认识大数据伦理，努力使自己的行为符合大数据伦理规范要求
- 熟悉大数据应用，探寻大数据与自己专业的应用结合点
- 激发学生基于大数据的创新创业热情

高等教育出版社 ISBN:978-7-04-053577-8 定价：32元

教材官网：<http://dbl原因.xmu.edu.cn/post/bigdataintroduction/>



附录E：《大数据导论》教材

- 林子雨 编著 《大数据导论》
- 人民邮电出版社，2020年9月第1版
- ISBN:978-7-115-54446-9 定价：49.80元

教材官网：<http://dbl原因.xmu.edu.cn/post/bigdata-introduction/>



开设大数据专业导论课的优质教材



扫一扫访问教材官网



附录F：《大数据技术原理与应用》教材

《大数据技术原理与应用——概念、存储、处理、分析与应用（第2版）》，由厦门大学计算机科学系林子雨博士编著，是国内高校第一本系统介绍大数据知识的专业教材。人民邮电出版社 ISBN:978-7-115-44330-4 定价：49.80元

全书共有15章，系统地论述了大数据的基本概念、大数据处理架构Hadoop、分布式文件系统HDFS、分布式数据库HBase、NoSQL数据库、云数据库、分布式并行编程模型MapReduce、Spark、流计算、图计算、数据可视化以及大数据在互联网、生物医学和物流等各个领域的应用。在Hadoop、HDFS、HBase和MapReduce等重要章节，安排了入门级的实践操作，让读者更好地学习和掌握大数据关键技术。

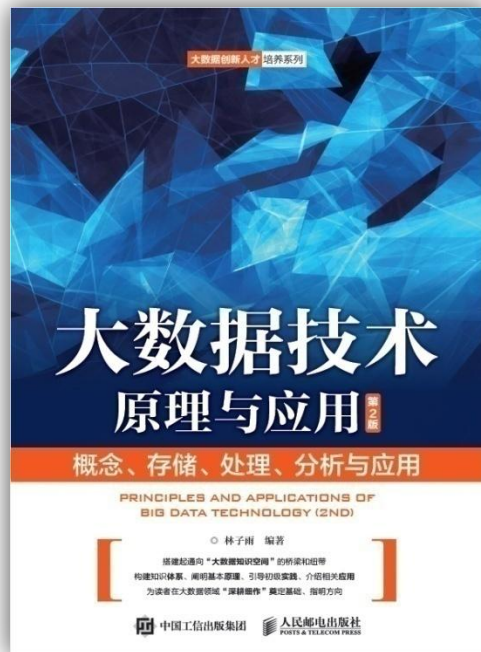
本书可以作为高等院校计算机专业、信息管理等相关专业的大数据课程教材，也可供相关技术人员参考、学习、培训之用。

欢迎访问《大数据技术原理与应用——概念、存储、处理、分析与应用》教材官方网站：

<http://dbl原因.xmu.edu.cn/post/bigdata>



扫一扫访问教材官网





附录G：《大数据基础编程、实验和案例教程（第2版）》

本书是与《大数据技术原理与应用（第3版）》教材配套的唯一指定实验指导书

大数据教材



1+1黄金组合
厦门大学林子雨编著

配套实验指导书



- 步步引导，循序渐进，详尽的安装指南为顺利搭建大数据实验环境铺平道路
- 深入浅出，去粗取精，丰富的代码实例帮助快速掌握大数据基础编程方法
- 精心设计，巧妙融合，八套大数据实验题目促进理论与编程知识的消化和吸收
- 结合理论，联系实际，大数据课程综合实验案例精彩呈现大数据分析全流程

林子雨编著《大数据基础编程、实验和案例教程（第2版）》

清华大学出版社 ISBN:978-7-302-55977-1 定价：69元 2020年10月第2版



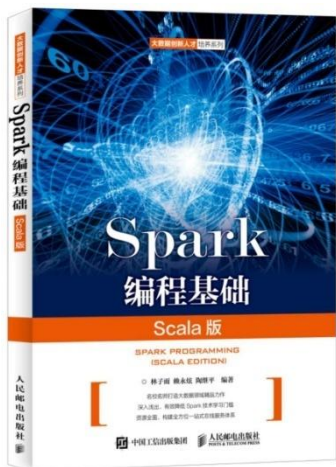
附录H: 《Spark编程基础 (Scala版)》

《Spark编程基础 (Scala版)》

厦门大学 林子雨, 赖永炫, 陶继平 编著

披荆斩棘, 在大数据丛林中开辟学习捷径
填沟削坎, 为快速学习Spark技术铺平道路
深入浅出, 有效降低Spark技术学习门槛
资源全面, 构建全方位一站式在线服务体系

人民邮电出版社出版发行, ISBN:978-7-115-48816-9
教材官网: <http://dblalab.xmu.edu.cn/post/spark/>

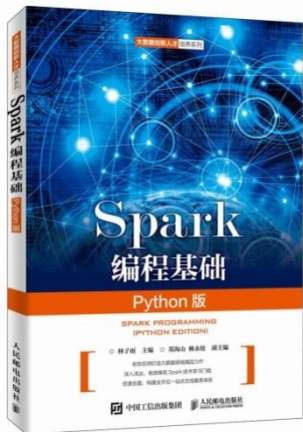


本书以Scala作为开发Spark应用程序的编程语言, 系统介绍了Spark编程的基础知识。全书共8章, 内容包括大数据技术概述、Scala语言基础、Spark的设计与运行原理、Spark环境搭建和使用方法、RDD编程、Spark SQL、Spark Streaming、Spark MLlib等。本书每个章节都安排了入门级的编程实践操作, 以便读者更好地学习和掌握Spark编程方法。本书官网免费提供了全套的在线教学资源, 包括讲义PPT、习题、源代码、软件、数据集、授课视频、上机实验指南等。



附录I: 《Spark编程基础 (Python版)》

《Spark编程基础 (Python版)》



厦门大学 林子雨, 郑海山, 赖永炫 编著

披荆斩棘, 在大数据丛林中开辟学习捷径
填沟削坎, 为快速学习Spark技术铺平道路
深入浅出, 有效降低Spark技术学习门槛
资源全面, 构建全方位一站式在线服务体系

人民邮电出版社出版发行, ISBN:978-7-115-52439-3

教材官网: <http://dblab.xmu.edu.cn/post/spark-python/>



本书以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、上机实验指南等。



附录J：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

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



扫一扫访问平台主页



扫一扫观看3分钟FLASH动画宣传片



附录K：高校大数据实训课程系列案例教材

为了更好地满足高校开设大数据实训课程的教材需求，厦门大学数据库实验室林子雨老师团队联合企业共同开发了《高校大数据实训课程系列案例》，目前已经完成开发的系列案例包括：

《电影推荐系统》（已经于2019年5月出版）

《电信用户行为分析》（已经于2019年5月出版）

《实时日志流处理分析》

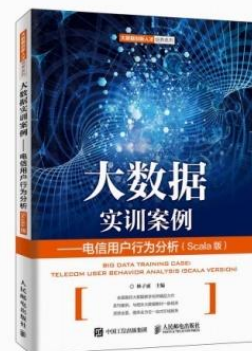
《微博用户情感分析》

《互联网广告预测分析》

《网站日志处理分析》

系列案例教材将于2019年陆续出版发行，教材相关信息，敬请关注网页后续更新！

<http://dblab.xmu.edu.cn/post/shixunkecheng/>



扫一扫访问大数据实训课程系列案例教材主页

The background of the slide features several faint, light-blue silhouettes of people. At the top, there are two groups of people standing and holding hands. On the right side, a person is shown in profile, looking towards the center. On the left side, two people are shown in profile, one appearing to be speaking or gesturing towards the other. The overall scene suggests a group of people in a meeting or presentation setting.

Thank You!

Department of Computer Science, Xiamen University, 2020