第3

Linux 基础知识及相关软件的基本使用

在学习本教材之前,应当具备基本的Linux 知识和操作技能。为了确保后面的学习顺利进行,有必要再了解一下相关知识和基本使用方法。

3.1 Linux 基础知识

3.1.1 Linux 目录结构

本教材中的目录、文件夹、路径均为同一个概念,可混用。

Linux 文件系统与 Windows 文件系统有很大的不同, Linux 文件系统是一个树形的分层组织结构,根目录(/) 是整个文件系统的唯一起点,其他目录都从该点出发。Linux 系统的全部 文件按照一定的用途归类,合理地挂载在这棵"大树"的"树枝"或"树叶"上,如图 3-1 所示,完全不用考虑文件的实际存储位置。



在树形目录结构中,文件和目录都通过路径来表示。路径有两种表示方法:一种是从根目录开始,称为绝对路径,如/usr/local/bin/;另一种是从当前目录开始,称为相对路径,如./hello.txt。目录标记的规则将在后面详细介绍。下面将对与我们的大数据实验相关的常用目录做简要介绍。



1. / (根目录)

一般,在根目录下只存放目录,不存放文件。

2. /bin 目录和/sbin 目录

Linux 系统中的大部分命令都在/bin 和/sbin 这两个目录下。

/bin 目录通常用来存放用户最常用的基本命令,包括文件操作实用程序命令、系统实用程 序命令、压缩工具命令等程序,如常用的 ls、tar、mv、cat 等可执行文件均在此目录下。

/sbin 目录通常存放系统维护程序命令,放置系统管理员使用的可执行命令,如 fdisk、 shutdown、mount 等。注意,与/bin 目录不同的是,这个目录是给系统管理员 root 用户使用的 命令,一般用户只能查看,不能设置和使用。

3. /etc 目录

/etc 目录是存放系统配置文件的目录,比如系统环境配置文件 profile 就在此目录下。

4. /root 目录

/root 目录是系统管理员超级用户的主目录(主文件夹),普通用户无权进入。

5. /home 目录

/home 目录是系统默认的用户目录,又称用户主目录或用户主文件夹,新增一般用户时,用户的目录都存放在此目录下,如登录普通用户 hadoop 时,其主目录为/home/hadoop/, hadoop 用户的桌面路径就是/home/hadoop/桌面。

6. /tmp 目录和/var 目录

/tmp 目录是存放临时文件的目录;/var 目录包含系统产生的经常变化的文件,例如打印机、 邮件、新闻等假脱机目录、日志文件、格式化后的手册页,以及一些应用程序的数据文件等。 其中,/var/log/目录下记载了各种程序的日志文件。

7. /dev 目录

dev 是设备(device)的英文缩写。/dev 目录包含了 Linux 系统使用的所有外部设备。

8. /usr 目录

/usr 目录是应用程序存放目录,存放与用户相关的程序和文件。/usr/bin 目录存放应用程序; /usr/share 目录存放共享数据; /usr/lib 目录存放不能直接运行,但许多用户程序运行所必需的一 些函数库文件; /usr/local 目录存放软件升级包; /usr/share/doc 目录是系统说明文件的存放目录; /usr/share/man 目录则是程序说明文件的存放目录。

9. /opt 目录

它是给主机额外安装的软件所设置的目录。本教材大数据安装在此目录下。

10. /lib 目录

/lib 目录下存放系统本身需要用到的库文件,包括/bin 和/sbin 目录下的程序所需的库文件。

3.1.2 Shell 介绍

Shell 的英文含义是"壳",它是相对于 Linux 内核来说的,因为它是建立在内核的基础上的面向用户的一种表现形式。Shell 就是 Linux 内核的外壳,是 Linux 对外交互的窗口。

1. Shell 在 Linux 中的存在形式

Linux 中的 Shell 是指一个面向用户的命令接口,表现形式就是一个用户可以录入的界面,



这个界面也可以反馈运行信息。

由于 Linux 系统不同于 Windows 系统,它是内核与界面分离的,可以脱离图形界面单独运行,也可以在内核的基础上运行图形化的桌面。因此,在 Linux 系统中出现了两种 Shell 表现 形式,一种是无图形界面终端运行环境下的 Shell,如第 14 章中使用的 CentOS;另一种是在桌面上运行的类似 Windows 的 MS-DOS 运行窗口,如本教材使用的 Ubuntu 系统。前者一般被简称为"终端",后者一般直接被称为"Shell"。大部分时候,我们是不区分这两种表现形式的, 它们都是交互界面而已。

2. Shell 如何执行用户的指令

Shell 有两种执行指令的方式。

第一种方式是用户先编写一个 sh 脚本文件,内含 Shell 脚本,然后使用 Shell 程序执行该脚本。我们习惯称这种方式为 Shell 编程。

第二种方式则是用户直接在 Shell 界面执行 Shell 命令。由于 Shell 界面的特点,大家都习惯一行行地写入指令,很少写出成套的程序一起执行,所以也称命令行。

可以说,Shell 只是在用户与机器之间搭建的一个桥梁,让我们能够通过它对计算机进行 操作和交互,从而达到让计算机为我们服务的目的。

3.1.3 超级用户——root 用户

在 Linux 系统中,超级用户一般被命名为 root 用户,相当于 Windows 系统中的 System 用 户 Administrator。root 用户是系统中唯一的超级管理员,具有系统中所有的权限,例如启动或 停止一个进程;删除或增加用户;增加或禁用硬件等。

使用 root 用户进行不当的操作是相当危险的,轻则死机,重则不能开机。所以,在 Linux 系统中,除非确实需要,一般情况下不推荐使用 root 用户,最好单独建立一个普通用户。本教 材全部采用单独建立的 hadoop 用户开展实验,需要执行超级用户的指令和权限时,使用 sudo 命令实现。

【实例】为普通用户赋权。为安装时创建的 hadoop 用户增加管理员权限,以方便部署任务, 避免在普通用户下使用 Linux 命令时有是否添加 sudo 命令的困惑。打开 hadoop 用户的终端, 命令与执行过程如图 3-2 所示。



图 3-2 为 hadoop 用户赋予管理员权限

3.1.4 sudo 命令和 su 命令

sudo 命令是一种权限管理机制,管理员可以授权一些普通用户执行一些 root 用户执行的 操作,也就是说,sudo 命令允许一个已授权用户以超级用户的角色运行一个命令。sudo 命令又 称为超级命令。

这样不仅减少了 root 用户的登录和管理时间,也提高了安全性。sudo 命令不是对 Shell 的一个代替,它是面向每个命令的。sudo 的具体使用在第2章中已经出现多次,不再具体举例说明。

su 命令是用来改变当前用户的。比如,由普通用户转为超级用户,或者由超级用户转为普

41

通用户,或者平台用户之间相互转换。如果没有给 root 用户设置密码,由普通用户转为 root 用 户时会提示输入密码,但输入任何密码都提示"认证失败"。这时应给 root 用户设置密码:

```
hadoop@hadoop-ubuntu:~$ sudopasswd root[sudo] hadoop 的密码:#输入 hadoop 用户的密码输入新的 UNIX 密码:#设置 root 用户的密码重新输入新的 UNIX 密码:#输入两次 root 用户的密码passwd: 已成功更新密码
```

由普通用户 hadoop 转换为 root 用户:

hadoop@hadoop-ubuntu:~\$ su root 密码: #输入刚才设置的 root 用户的密码 root@hadoop-ubuntu:/home/hadoop#

在 root 用户下,就不必再使用 sudo 命令了,可以在终端中使用 root 权限。

由 root 用户转为普通用户:

\$ su hadoop #无须密码

3.1.5 创建与删除普通用户

(1) 创建新用户。

启动 Ubuntu 系统,进入安装时创建的 hadoop 用户,打开终端(可以使用快捷键 "Ctrl+Alt+T"),输入以下命令创建用户 spark:

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

这条命令创建了一个普通用户 spark, "/bin/bash" 说明该 Shell 类型为 bash, bash Shell 是 最常用的一种 Shell, 是大多数 Linux 发行版默认的 Shell。

(2) 使用以下命令为 spark 用户设置密码:

```
$ sudo passwd spark
```

按回车键后会提示为用户 spark 连续输入两次密码。

(3) 返回 Ubuntu 系统的登录界面,查看新建用户。

在 hadoop 用户的终端建立了 spark 用户后,可单击屏幕右上角的"齿轮"图标,选择"注 销"选项,注销当前登录的 hadoop 用户,返回 Ubuntu 系统的登录界面(见图 3-3),就会发现 多了一个新建用户 spark。



图 3-3 Ubuntu 系统的登录界面

```
(4) 删除新建用户 spark。
```

再次登录 hadoop 用户, 输入以下命令:

\$ sudo userdel -r spark

又所作

返回 Ubuntu 系统的登录界面,可以查看用户情况。 如果用户没有被删除,可进入 hadoop 用户再次进行删除操作。

3.1.6 目录标记规则

1. 主目录(用户目录)标记(~)

在 Linux 系统中,超级用户或普通用户打开一个终端时,都会进入一个特殊的目录,我们称之为主目录。超级用户登录系统后默认进入/root 目录,/root 目录就是超级用户的主目录。普通用户登录系统后进入普通用户设置的默认目录,这个默认目录就是普通用户的主目录。主目录标记为 "~",比如 hadoop 用户登录后,其用户目录为:/home/hadoop/,如图 3-4 所示。



图 3-4 查看用户目录

2. 当前目录标记(. 或 ./)

当前目录标记为"."或"./",实验测试过程和步骤如图 3-5 所示。

| nadoop@hadoop-ubuntu:~\$ | pwd | |
|---|-------|--|
| /home/hadoop | | |
| <pre>nadoop@hadoop-ubuntu:~\$</pre> | cd . | |
| nadoop@hadoop-ubuntu:~\$ | pwd | |
| /home/hadoop | | |
| <pre>nadoop@hadoop-ubuntu:~\$</pre> | cd ./ | |
| nadoop@hadoop-ubuntu:~\$ | pwd . | |
| /home/hadoop | | |
| nadoop@hadoop-ubuntu:~\$ | | |
| A REAL PROPERTY OF THE OWNER OF T | | |

图 3-5 当前目录标记验证

3. 上一级目录标记(.. 或 ../)

当前目录的上一级目录使用".."或"../"表示,实验测试过程和步骤如图 3-6 所示。

| hadoop@hadoop-ubuntu:~\$ pwd |
|----------------------------------|
| /home/hadoop |
| hadoop@hadoop-ubuntu:~\$ cd |
| hadoop@hadoop-ubuntu:/home\$ pwd |
| /home |
| hadoop@hadoop-ubuntu:/home\$ cd/ |
| hadoop@hadoop-ubuntu:/\$ pwd |
| / |
| hadoop@hadoop-ubuntu:/S |
| |

图 3-6 上一级目录标记验证

3.1.7 目录权限

Linux 系统对文件或目录权限有着严格的规定,如果用户不具备权限,将无法访问目录及 其下面的文件,特别是不能对该目录进行写操作。例如,hadoop 用户登录 Linux 系统后,把文 件解压到/opt 目录下,但这时用户 hadoop 并不是/opt 目录的拥有者,就无法对该目录进行操 作,不能创建子目录,也不能写入任何文件。这时,就要使用 chown 命令进行相关授权,让 hadoop 用户拥有对该目录的操作权限。具体命令如下(注意,"-R"中的字母是大写的):

\$ sudo chown -R hadoop /opt #用户目录授权,改变/opt 的拥有者为用户 hadoop

\$ sudo chown -R hadoop:hadoop /opt #用户目录和用户组授权,将/opt的拥有者和群组(group) 都设成用户 hadoop

"chown -R hadoop /opt"代表将"/opt"文件夹下所有递归目录和文件的拥有者(owner) 改为用户 hadoop。chown 命令用于修改目录的权限;-R 代表递归,就是从当前目录到其最下 面的子目录。

Linux 是多用户多任务操作系统,所有的目录或文件皆有拥有者。利用 chown 命令可以 改变文件的拥有者,一般该命令只能由系统管理者使用。普通用户没有权限改变别人的文件 的拥有者,也没有权限将自己的文件的拥有者改为别人。所以,在普通用户下要使用 sudo 命 令执行操作。

3.1.8 安装 SSH

SSH 类似于远程登录,可以从一台 Linux 主机登录到另外一台 Linux 主机,并且在其上运行命令。SSH 的详细介绍将在集群免密登录中说明。

在安装 hadoop 之前需要配置 SSH,因为集群和单节点模式都需要用到 SSH 登录。Ubuntu 系统默认已安装了 SSH Client (SSH 客户端),对于单节点的伪分布式模式,还需要安装 SSH Server (SSH 服务器)。如果不安装 SSH Server,但已配置好单节点 SSH 服务器和客户机的授 权登录,则无法启动将要安装的伪分布式 hadoop。

安装 SSH Server 的命令如下:

```
$ sudo apt-get install openssh-server #在线安装
```

安装完毕后,可以使用以下命令登录本机:

\$ ssh localhost

首次登录 SSH 时会有提示,输入"yes",按提示输入密码后就登录到本机了。

这样设置后每次登录都要输入密码,设置成无密码登录比较方便。

先退出刚才的 SSH,回到原先的终端窗口,再利用 ssh-keygen 生成密钥,并将密钥加入授 权中,执行下列一系列命令。

| \$exit | # 退出刚才的 ssh localhost |
|--|-------------------------------|
| \$cd ~/.ssh/ | # 若没有该目录, 先执行一次 ssh localhost |
| \$ssh-keygen -t rsa | # 全部按回车键就可以 |
| <pre>\$cat ./id_rsa.pub >> ./authorized_keys</pre> | # 加入授权 |

使用命令 "ssh-keygen-trsa" 后会生成公钥和私钥。默认在该目录 (~/.ssh/) 会生成 id_rsa (私钥) 和 id_rsa.pub (公钥) 两个文件。

使用命令"cat ~/.ssh/id_rsa.pub ≫ ~/.ssh/authorized_keys"将公钥密码导入认证文件 authorized_keys 中。

SSH 授权完成,此时再使用 ssh localhost 命令,无须输入密码就可以直接登录,也不会影响伪分布式 hadoop 的启动了。



3.2.1 修改主机名

修改主机名要对以下两个文件进行修改。

将安装时定义的主机名"hadoop-virtual-machine"修改为"hadoop-ubuntu"。执行以下命令:

hadoop@hadoop-virtual-machine:~\$ sudo vim /etc/hostname

将原来的内容修改为: hadoop-ubuntu,保存并退出编辑。

同时修改 hosts 的配置:

hadoop@hadoop-virtual-machine:~\$ sudo vim /etc/hosts

将原来的内容"127.0.1.1 hadoop-virtual-machine"修改为新的内容"127.0.1.1 hadoop-ubuntu"。 在 hosts 设置中,我们可能会见到"127.0.0.1"和"127.0.1.1"两个 IP 地址,"127.0.0.1" 是指本机地址,主要用于测试,用汉语表示就是"我自己"。"127.0.1.1"是一个本地的回环地 址,出现在/etc/hosts 文件中,可以用来解析自己的主机名。另外,localhost 常用于本地 DNS 解 析的 127.0.0.1 的域名。

保存并退出编辑。重启系统:

```
hadoop@hadoop-virtual-machine:~$ reboot
```

系统重启后,打开终端查看主机名,显示为:

hadoop@hadoop-ubuntu:~\$ hostname

hadoop-ubuntu

▲ 注 意 修改主机名要完整地修改两个文件: hostname 和 hosts。否则可能影响到后 面 Ubuntu 系统下 HDFS 的一些连接操作,会出现"拒绝连接"的错误。

3.2.2 目录操作

1. cd 命令

改变目录路径。

2. mkdir 命令

语句格式: mkdir [-p] dirName

在指定位置创建新的文件夹或目录。要创建文件夹或目录的用户必须对所创建的文件夹的 父文件夹具有写权限。

例如,在当前工作目录下,建立一个名为"AAA"的子目录: mkdir AAA。

参数-p确保目录名称存在,如果不存在就创建一个。例如,在/mnt/udisk/下创建名为"AAA"的子目录:mkdir-p/mnt/udisk/AAA,表示无论原来有无/mnt/udisk/目录,都可以在此目录下新建 AAA 子目录。创建多级目录必须使用参数-p。

3. pwd 命令

显示用户当前的工作路径,即当前目录。



3.2.3 查看文件内容和清屏

1. cat 命令: 查看文件内容

cat 主要有以下三大功能。

- 一次显示整个文件,如\$ cat filename。
- 利用键盘创建一个文件,如\$ cat > filename。只能创建新文件,不能编辑已有文件。
- 将几个文件合并为一个文件: \$ cat file1 file2 > file。
- 2. clear 命令

清屏。

3.2.4 文件操作

1. cp: 复制文件或目录命令

命令格式: cp [-参数] 源 目的

命令功能:将源文件复制到目标文件,或者将多个源文件复制到目标目录。 下面仅对常用的参数做介绍。

- a: 将源文件或目录状态、权限等信息按照原状态复制。
- r: 递归复制。若源文件中含有子目录,则将子目录下的文件依次复制至目的地。所有文 件夹复制都要使用该参数。
- f: 强制(force)复制。若目的地已经有相同文件名存在,则先将其删除再进行复制。

例如,-rf 表示复制所有文件,包括子目录,若目的地已存在相同的文件,则不予提示,直 接复制。

请试验将一个文件复制到另一个目录时,有无参数-a对复制结果的影响。实证如下:

```
$ mkdir /test
                                #在根目录下创建文件夹 test
$ cp file1.txt /test
                                 #如果没有 file1.txt 文件,则先创建
 $ cp -a file1.txt /test/file2.txt
                               #复制文件
 $ ll /test
```

检查两个文件的创建时间是否相同。思考为什么?

2. rm:删除文件或目录命令

命令格式: rm [-参数] 文件或目录 常用参数如下。

- f: 强制删除, 删除时无信息提示。
- r: 递归删除, 删除目录时必须有该参数。

3. mv: 移动文件或将文件改名

mv 是 move 的缩写, mv 命令可以用来移动文件或将文件改名, 是 Linux 系统中常用的命 令,经常用来备份文件或目录。

命令格式: mv [-参数] 源文件或目录 目标文件或目录

命令功能:视 mv 命令中第二个参数类型的不同(文件或目录), mv 命令将文件重命名或 将文件移至一个新的目录中。当第二个参数类型是目标文件时, mv 命令将文件重命名, 此时, 源文件只能有一个(也可以是源目录名),它将所给的源文件或目录重命名为给定的目标文件

46

成权所有

名。当第二个参数是已存在的目录名称时,源文件或目录参数可以有多个, mv 命令将各参数 指定的源文件均移至目标目录中。

命令参数如下。

- b: 若需覆盖文件,则先进行备份。
- f: 如果目标文件已经存在, 直接将其覆盖。
- i: 若目标文件已经存在, 询问是否将其覆盖。
- u: 若目标文件已经存在,且 source(源文件)比较新,则对目标文件更新(update)。 【实例 1】将文件改名。
- 命令: mv test.log test1.txt
- 说明:将文件 test.log 重命名为 test1.txt。
- 【实例 2】移动文件(假设 test3 目录已经存在)。
- 命令: mv test1.txt test3
- 说明:将文件 test1.txt 移到目录 test3 中。

3.2.5 文件解压

将文件打包、压缩与解压缩的命令为 tar,参数如下

- c: 建立一个压缩文件。
- x: 解开一个压缩文件。
- t: 查看压缩文件中的文件。

这三个参数不能同时使用, 仅能选择一个。因为不可能同时压缩与解压缩文件。

- z: 用 gzip 解压缩文件。
- j: 用 bzip2 压缩文件。
- v: 显示详细信息。
- f: 压缩后的文件名, 这个参数必须是最后一个参数。

本教材仅使用 z 解压方式。大数据软件安装包通常都是一个压缩文件,并且文件名以.tar.gz 为后缀(或者简写为.tgz),这样的压缩格式一般简称为 gz 格式,这种压缩文件需要通过 tar 解 压才能安装。通常解压命令格式如下:

\$ tar -zxvf gz 格式压缩文件-C 解压后转到的路径名

这里采用 z 解压方式,参数 v 表示可以看到解压的过程; -C 表示解压后转到指定的路径。

【实例】将 hbase-1.1.5-bin.tar.gz 文件下载到本地计算机,打开虚拟机,登录 hadoop 用户,

并且将 hbase-1.1.5-bin.tar.gz 文件复制到虚拟机 Ubuntu 的"下载"目录下。操作过程如下。

(1) 复制文件, 过程如图 3-7~图 3-9 所示。



图 3-7 打开桌面的"主文件夹"

| ← → 合主文件夹 | | |
|--------------|-------|----|
| ● 最近使用的 | | |
| ☆ Home | | |
| 📮 桌面 | 公共的 | 模板 |
| ₩ 视频 | | 4 |
| 🗖 图片 | 下裁 | 辛斤 |
| ▶ 文档 | 1. #2 | 日小 |
| ◆ 下载 | | |
| ▶ 音乐 | | |
| 前 回收站 | | |
| 器 网络 | | |
| ふ 计算机 | | |
| □ 软盘磁盘 | | |
| ■ 连接到服务器 | | |

图 3-8 打开"下载"目录



图 3-9 将 hbase-1.1.5-bin.tar.gz 文件复制到"下载"目录下

(2) 打开终端,执行以下命令:

\$ sudo tar -zxvf /home/hadoop/下载/hbase-1.1.5-bin.tar.gz -C /opt/

(3) 查看解压结果。

\$ ls /opt/hbase-1.1.5/

(4)修改目录路径名称。

\$ sudo mv /opt/hbase-1.1.5 /opt/hbase

3.2.6 进程与端口查看命令

1. ps: 监控后台进程命令

进程是 Linux 系统中一个非常重要的概念。Linux 是一个多任务操作系统,经常同时运行 多个进程。

48

ps 命令是基本且非常强大的进程查看命令,常用于监控后台进程的工作情况,显示目前系统进程的状况,命令如下:

| \$ ps | |
|------------|---------------|
| PID TTY | TIME CMD |
| 2438 pts/1 | 00:00:00 bash |
| 5819 pts/1 | 00:00:00 ps |

第一列 PID 表示进程的 ID,是进程唯一的标识符。最后一列是进程的名称。

2. jps: 查看本地运行的 java 程序进程

与 ps 命令类似, jps 命令用来显示本地有权限的 Java 进程,可以查看本地运行着几个 Java 程序,并显示它们的进程号。jps 命令是 Java 的一个命令文件,存放在 Java 安装路径的 bin 目录下。也就是说,只有安装 Java 并设置好它的环境变量之后,jps 命令才能正常使用。jps 命令 是查看 hadoop 进程的关键命令,hadoop 进程均是用 Java 编写的。

3. netstat: 查看端口被占用情况

查看所有的服务端口,显示 PID。

\$ netstat -anp

这样会显示很多端口,如果仅查看某个已知端口的占用情况,则可以通过管道("|")和 grep (搜索查找)命令查看。

比如, 查看 20893 端口的占用情况, 执行以下命令:

```
$ sudo netstat -ap|grep 20893
```

结果如图 3-10 所示。

| hadoop@hadoop-ubuntu:~\$; [sudo] hadoop 的密码: | sudo netstat | -ap grep | 20893 | | |
|--|--------------|----------|-------|---------------------|-----|
| unix 2 [ACC] /user/1000/keyring/contro | 流 Li ol | ISTENING | 20893 | 1632/gnome-keyring- | /ru |
| | | 10 | | | |

```
图 3-10 查看端口占用情况
```

可以发现,处于监控状态为 LISTENING 的端口已经被占用,端口 20893 被 PID 为 1632 的进程占用。

4. 杀死进程: kill PID

kill 命令用来中止一个进程。中止一个进程也被称为杀死进程,例如输入以下命令:

\$ kill 1632

该命令结合了上一条命令的查询结果,杀死了进程号为1632的进程,解除了对端口20893的占用。

3.2.7 系统管理命令

1. find: 查找文件命令

find 命令是在目录结构中搜索文件,并执行指定的操作。此命令提供了相当多的查找条件,功能很强大。

命令格式: find [起始目录] -name 搜索文件名 [操作]

命令功能:从指定的起始目录开始,递归搜索各个子目录,查找满足条件的文件,并且 进行相关操作。

49

如在/sbin 目录中寻找以"if"开始的文件:

```
hadoop@hadoop:~$ find /sbin -name if*
/sbin/ifdown
/sbin/ifup
/sbin/ifquery
/sbin/ifconfig
```

2. df: 显示磁盘剩余空间命令

命令如下,

| hadoop@hadoop | :~\$ df | -h | | | | |
|-----------------|---------|------|-------|------|----------------|---------|
| 文件系统 | 容量 E | 2月 可 | 「用 已月 | 用% 挂 | 载点 | |
| udev | 972M | 0 | 972M | 0% | /dev | |
| tmpfs | 200M | 6.4M | 193M | 4% | /run | |
| /dev/sda5 | 20G | 12G | 6.7G | 64% | / | |
| tmpfs | 996M | 244K | 996M | 1% | /dev/shm | |
| tmpfs | 5.0M | 4.0K | 5.0M | 1% | /run/lock | |
| tmpfs | 996M | 0 | 996M | 08 | /sys/fs/cgroup | |
| tmpfs | 200M | 48K | 200M | 1% | /run/user/1000 | 1 GK FJ |
| .8 数据流重定向和管道的使用 | | | | | | |
| 1. 数据流重知 | 定向 | | | | ニスナトル | |

3.2.8 数据流重定向和管道的使用

1. 数据流重定向

Linux 默认的输入设备是键盘,输出设备是显示器。可以使用重定向改变这些设置,数据 流重定向就是将输入或输出数据流从默认的位置定向到其他位置。比如,将某个命令执行后应 该显示在屏幕上的数据传输到其他地方 (文件或设备上)。

数据流重定向操作符功能描述如下。

• >: 输出导向,如果目标存在数据,则原始数据被替换掉。

• >>: 输出导向,如果目标存在数据,则在原始数据后面追加。

<: 输入导向,可以从一个文件中读入数据,如果原文件有数据会被替换掉。

【实例】

(1) 把/etc 目录中的文件信息存储到/tmp/etcfile 文件中。

```
hadoop@hadoop:~$ ll /etc/>/tmp/etcfile
hadoop@hadoop:~$ cat /tmp/etcfile
```

(2)回显信息到文件中,屏幕不输出信息,而是将信息传送到文件中。

```
hadoop@hadoop:~$ echo "hello world" > filetxt
hadoop@hadoop:~$ cat filetxt
```

追加数据(>>)的练习可自己进行,不赘述。

(3) 输入导向。

```
hadoop@hadoop:~$ cat > catfile < filetxt
hadoop@hadoop:~$ cat catfile
hello world
```



2. 管道的使用

管道就是将第一个命令的输出结果作为第二个命令的输入,并继续处理。例如,用户可以 用以下命令查看/etc 目录中的内容:

\$ ls -al /etc

演示输出结果,我们发现结果很快就显示完毕,几乎一闪而过,我们可以用管道的形式将 ls -al /etc 显示的结果输入到第二个命令 less 程序。less 程序是一个分页显示文件的工具,这 样就允许我们较悠闲地用"↑"键向上翻页,或者用"↓"键向下翻页,一页一页地查看/etc 目 录下的文件信息。

ls -al /etc | less

管道的操作符为英文状态下的" | ",可以连接多个命令,表示前一个命令的输出是后一 个命令的输入。

3.2.9 其他使用技巧介绍

1. 使用"Tab"键自动补全命令

在 Shell 中输入命令时,可以使用快捷键 "Tab"自动补全命令,以节省时间。例如输入: \$ cd /opt/hbase-1.1.5/,但 "hbase-1.1.5",特别是后面的版本号不好记,手动输入易出错,就可以采用自动补全命令,只需先输入 "cd /opt/hbase",再按 "Tab"键即可自动补全。当有两条以上命令可供选择时,连续按两次 "Tab"键,就会将所有选择列出来。

2. 隐藏文件

在 Ubuntu 中,给文件或文件名前加一个"."表示隐藏文件。隐藏文件通过 ls 命令是查看 不到的,但使用命令 ls –a 则可以显示出来。后面出现的环境变量配置文件.bashrc 就是一个隐 藏文件,用来保存系统的各种环境变量。

3. 重现历史命令

在 Ubuntu 终端命令窗口中,系统会自动保存输入的命令,通过键盘的"↑"和"↓"键 可翻看历史命令,找到某条历史命令后,可直接按回车键执行该命令,这对需要重复输入命令 的操作很有帮助。

3.3 在 Ubuntu 中安装 Eclipse



3.3.1 通过软件中心下载安装 Eclipse

Eclipse 是常用的程序开发工具,是 Java 开发的主要工具之一。可以利用 Ubuntu 自带的软件中心(桌面上的 UK 图标)安装 Eclipse 等软件。"Ubuntu Kylin 软件中心"快捷图标如图 3-11 所示。



图 3-11 "Ubuntu Kylin 软件中心"快捷图标

如果有一段时间没有进行系统更新,那么安装 Eclipse 之前应确保 Ubuntu 更新至最新版本。打开 Linux 终端,先做以下操作:

\$ sudo apt-get update

事实上,在本教材的学习过程中,只要是在线安装新软件,都应确保 Ubuntu 更新至最新版本,最妥帖的办法就是下载安装软件之前先通过 apt-get 命令做一次 update 更新操作。

然后打开桌面的"Ubuntu Kylin 软件中心",在搜索框输入"eclipse",会自动出现 Eclipse 软件的快捷安装按钮,如图 3-12 所示。单击"eclipse platform"选项安装即可。



图 3-12 搜索 Eclipse 的相关软件

安装时要保持计算机处于联网状态,若要输入管理员权限,输入当前用户的登录密码即可。 如果长时间不能正常安装,请检查网络;如果网络正常,也可参照"apt更新与更新源更换"一 节的内容,重新更换更新源,如选择阿里云源"mirrors.aliyun.com",重新安装 Eclipse。

Eclipse 安装完毕后,可以单击桌面快捷启动图标栏的"资源搜索"图标(图 3-11 最左侧的图标),搜索本地和在线资源,在弹出的"资源搜索"界面的搜索框中输入"eclipse",就可以找到已经安装好的 Eclipse 软件(见图 3-13),单击 Eclipse 图标,启动运行该软件。



图 3-13 资源搜索界面

3.3.2 在桌面创建 Eclipse 快捷方式

安装成功后,打开文件系统,在目录/usr/share/applications/下找到 Eclipse 图标,如图 3-14 所示。右击 Eclipse 图标,在弹出的快捷菜单中选择"复制"命令,回到桌面并在空白处右击,



在弹出的快捷菜单中选择"粘贴"命令,此时,Eclipse 的桌面快捷方式就创建完毕了。



图 3-14 /usr/share/applications/目录文件系统

本章主要回顾了 Linux 的一些基本操作,为下面的学习预热,还通过 Eclipse 的软件安装 与桌面快捷方式的创建,熟悉了 Linux 一般软件安装的方法。特别需要提醒的是,务必安装 SSH, 因为它是安装运行 hadoop 的必要环境。

