

Linux基础培训

生信帮

云南农业大学

云南省生物大数据重点实验室

2023年7月15日

操作系统

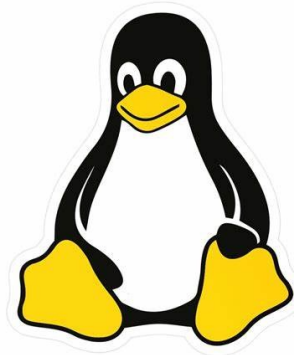


雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮



ANDROID

HarmonyOS

Linux OS



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

常见的Linux发行版，如Ubuntu、CentOS、Debian等。

```
[bioinformatics@localhost: 03:05:49 /home/MD120044T/bioinformatics/01.Linux]
$ll
total 75844
-rw-r--r--. 1 bioinformatics users      180 Jul 12 09:29 data.config
-rw-r--r--. 1 bioinformatics users  212788 Jul 12 09:29 gene.list
-rw-r--r--. 1 bioinformatics users   17596 Jul 12 09:36 L13.mt.vcf
-rw-r--r--. 1 bioinformatics users   1508 Jul 12 09:30 peak.xls
-rw-r--r--. 1 bioinformatics users   6174 Jul 12 09:35 pep.fa
-rw-r--r--. 1 bioinformatics users    227 Jul 12 09:29 R1_1.fq.gz
-rw-r--r--. 1 bioinformatics users    449 Jul 12 09:29 R1_2.fq
drwxr-xr-x. 6 bioinformatics users     54 Jul 12 09:31 scrNA-seq
-rw-r--r--. 1 bioinformatics users 12919810 Jul 12 09:36 test.bam
-rw-r--r--. 1 bioinformatics users    259 Jul 12 09:34 test.fa
-rw-r--r--. 1 bioinformatics users    8281 Jul 12 09:35 test.gff3
-rw-r--r--. 1 bioinformatics users 64449593 Jul 12 09:37 test.sam
-rw-r--r--. 1 bioinformatics users     46 Jul 13 03:10 x
-rw-r--r--. 1 bioinformatics users      9 Jul 13 02:43 xx
-rw-r--r--. 1 bioinformatics users    17 Jul 13 02:59 xx2
-rw-r--r--. 1 bioinformatics users    12 Jul 13 03:01 xx3
```

为什么要用Linux



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

使用Linux的原因有多个：

- 1. 开源性和安全性高：** Linux是一个开源操作系统，意味着它的源代码可以被任何人查看、修改和分发。这使得用户可以自由地定制和优化系统，以满足自己的需求。此外，开源性也意味着安全漏洞可以更快地被发现和修复。
- 2. 稳定性：** Linux系统以其稳定性而闻名。它可以在长时间运行的服务器上提供可靠的服务，而无需频繁重新启动或崩溃。这对于需要**持续运行的任务**和关键业务非常重要。
- 3. 支持命令行界面：** Linux系统提供了强大的命令行界面（CLI），允许用户通过命令输入指令来执行各种复杂任务。这使得用户可以更高效地进行系统管理、软件配置和**自动化**工作。
- 4. 丰富的生信软件库：** Linux系统通常拥有庞大而丰富的软件库，其中包含开源和免费的应用程序、工具和工作环境。用户可以轻松地通过软件包管理器安装所需的生信软件，满足各种不同的需求。
- 5. 支持多任务，多用户：** Linux操作系统是一个多任务操作系统，这意味着它可以同时运行多个程序。无论是在桌面环境还是服务器上，用户可以同时打开多个应用程序，可以同时执行多个计算任务等；Linux操作系统支持多用户环境。多个用户可以同时登录到同一台服务器上，并在各自的用户账户中进行独立的操作。每个用户都有自己的文件和设置，彼此之间相互独立。

如何登录Linux



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

- Linux远程登录的方式：TELNET, **SSH**, SFTP...

由于SSH在安全性，功能和性能上都比较好，推荐使用SSH。

- 常用登录终端：

- ◆ iTerm2 (macOS)

- ◆ Xshell, PuTTY, MobaXterm (Windows)

- ◆ XFTP, WinSCP, filezilla:用于从服务器下载数据到本地

- 远程登录Linux步骤：

- ◆ 打开远程登录软件，输入登录节点IP地址；

- ◆ 输入用户名，回车，密码，回车

- ◆ 登录界面

- ◆ 切换到计算节点

登陆界面



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

```
[bioinformatics@localhost: 03:09:36 ~]
```

~代表“用户家目录”的意思。

bioinformatics的家目录在/home/bioinformatics。



在Linux系统中，路径（path）用于指定文件或目录的位置。Linux使用一种类似于树形结构的文件系统，根目录为 /，其他目录都是从根目录开始的子目录。以下是一些常见的Linux目录：

根目录（root）：表示整个文件系统的起始点，表示为 /。

普通目录：普通目录是文件系统中包含文件和其他子目录的标准目录。例如 /home、/usr、/var 等。

家目录（home）：每个用户都有一个家目录，用于存储用户的个人文件和配置文件。通常位于 /home/<用户名>，例如 /home/john 表示用户 "john" 的家目录。

当前目录（current directory）：表示当前所在的工作目录。在命令行中，可以使用 . 表示当前目录。

父目录（parent directory）：表示当前目录的上一级目录。在命令行中，可以使用 .. 表示父目录。

绝对路径（absolute path）：从根目录开始的完整路径，例如 /home/soft/bedtools。

相对路径（relative path）：相对于当前目录的路径，例如 ../test/file.txt 表示当前目录的上一级目录中的 test 目录下的 file.txt 文件



在Linux环境中，环境变量是一些用于存储系统配置和运行时信息的动态值。它们为用户和系统进程提供了访问重要信息的方法。以下是一些常见的Linux环境变量：

- PATH：指示shell查找可执行文件的路径。当您输入命令时，系统会在这些路径中查找对应的可执行文件并执行。可以使用 `echo $PATH` 命令查看当前的PATH值。
- LD_LIBRARY_PATH：用于指定共享库文件（.so文件）的搜索路径。

```
export LD_LIBRARY_PATH=/share/app/LACHESIS/boost/lib/:$LD_LIBRARY_PATH
```

- PS1：表示shell提示符的格式。这些环境变量可以在系统的全局范围内设置，也可以在用户级别或特定的shell会话中设置。
- 临时设置：可以通过直接在命令行中使用 `export` 命令设置环境变量，例如：

```
export PATH=/usr/local/bin:$PATH
```

上述命令将 `/usr/local/bin` 添加到现有的PATH值中。

- 要永久设置环境变量，可以将其添加到用户的配置文件（如 `~/.bashrc`）中，这样每次登录时都会加载这些配置文件。



在Linux中，命令行是与操作系统进行交互的一种方式。用户可以通过输入命令来执行特定的操作。下面是Linux命令行的基本格式：

```
command [options] [arguments]
```

command：要执行的命令，比如 ls、cd、mkdir 等。

options：命令的选项，用来修改命令的行为。选项通常以短横线（-）或双短横线（--）开头，比如 -l、-a、--help 等。

arguments：命令的参数，用来指定命令需要操作的对象。参数可以是文件名、目录名、字符串等。

如：less -Ns test.txt

- man 命令用于查看 Linux 系统中的手册页（manual pages）。man 命令的格式：man [选项] <命令/手册页>，其中，选项 是可选的命令参数，用于控制 man 命令的行为和显示。
- help：--help 或 -h 是一种常见的命令行选项，用于显示命令的简要帮助信息。不同的示例用法：
- 显示命令的帮助信息：command --help；缩写形式：command -h



文件和目录操作：

- cd：切换到指定目录。
- ls：列出当前目录下的文件和目录。
- pwd：显示当前所在目录的路径。
- mkdir：创建一个新的目录。
- cp：复制文件或目录。
- rm：删除文件或目录。
- mv：移动或重命名文件或目录。
- ln:链接
- find：寻找某个文件或者目录
- paste：合并（按列或者转换为行）两个文件
- join：按照指定列合并

Linux重点掌握命令



雲南農業大學
Yunnan Agricultural University



生信帮

云南省生物大数据重点实验室

在Linux中，ls -l 是一个常用的命令，用于以详细的格式列出文件和目录的信息。下面是 ls -l 命令输出的列信息的解释：

```
-rw-r--r--. 1 bioinformatics users 1508 Jul 12 09:30 peak.xls
-rw-r--r--. 1 bioinformatics users 6174 Jul 12 09:35 pep.fa
-rw-r--r--. 1 bioinformatics users 227 Jul 12 09:29 R1_1.fq.gz
-rw-r--r--. 1 bioinformatics users 449 Jul 12 09:29 R1_2.fq
drwxr-xr-x. 6 bioinformatics users 54 Jul 12 09:31 scrNA-seq
-rw-r--r--. 1 bioinformatics users 12919810 Jul 12 09:36 test.bam
```

- 第一列:文件类型和权限：-：普通文件。d：目录。l：符号链接。c：字符设备文件。b：块设备文件。s：套接字文件。p：命名管道。
- 接下来的九个字符表示文件的访问权限，三个字符一组，每个组由三个字符组成，分别表示读（r）、写（w）和执行（x）权限。如果权限未设置，则会显示短横线（-）。第一组（左侧）是所有者的权限;第二组是与组相关的用户的权限;第三组是其他用户（不属于所有者或组）的权限。
- 硬链接数：第二列显示链接到该文件或目录的硬链接数。
- 所有者和组：第三列和第四列分别显示了文件或目录的所有者和所属组。
- 文件大小：第五列显示了文件的大小（以字节为单位）。
- 修改日期和时间：第六列和第七列分别显示了文件的最后修改日期和时间。
- 文件名：最后一列显示了文件或目录的名称。



文件内容展示类：查看大文件使用less，查看程序代码或者编辑小文件使用vi,less可以查看压缩文件。行数少的可以用cat。

- vi：打开，编辑文件内容
- cat：显示文件全部内容。
- less：逐页显示文件内容。
- head：显示文件的前几行。
- tail：显示文件的后几行。



文本处理类：

- sort：排序
- uniq：去重
- wc：统计行数
- grep：筛选
- sed：替换
- awk：条件判断
- |：管道符号。是管道符号，用于将一个命令的输出作为另一个命令的输入。它可以将多个命令串联起来，实现更复杂的数据处理和操作。| 本身并不是一个命令，而是一种连接多个命令的方式。
- >：重定向。是重定向操作符，用于将命令的输出重定向到文件中，或者创建一个新的文件并将命令的输出写入其中。它可以用于覆盖文件内容或者创建新文件。
- >>：追加符号



压缩和解压缩类：

➤ tar命令：

压缩文件或目录：tar -cf archive.tar file1 file2 directory1

解压缩文件或目录：tar -xf archive.tar

➤ gzip命令：

压缩文件：gzip file

解压缩文件：gunzip file.gz

➤ bzip2命令：

压缩文件：bzip2 file

解压缩文件：bunzip2 file.bz2

➤ zip命令（需要安装zip软件包）：

压缩文件或目录：zip archive.zip file1 file2 directory1

解压缩文件或目录：unzip archive.zip

➤ rar命令（需要安装rar软件包）：

压缩文件或目录：rar -a archive.rar file1 file2 directory1

解压缩文件或目录：rar -x archive.rar



系统操作和管理：

- &：后台符，将一个命令直接放到后台运行。
- nohup：nohup 是一个用于在后台运行命令并使其不受终端关闭影响的命令，所有输出重定向到 nohup.out 文件中。
- ps：显示当前正在运行的进程。
- top：实时显示系统资源占用情况。
- kill：挂起或者杀死程序
- su：以超级用户权限执行命令。
- chmod：修改目录和文件权限操作命令
- passwd：更改密码
- du：文件和目录大小占用情况
- df：磁盘的使用情况
- history：查看历史操作命令
- which：定位程序的安装路径



这只是正则表达式的一部分，Linux 中使用正则表达式的命令有很多，如 grep、sed、awk 等。在实际使用中，你可以根据具体需求来选择合适的工具和正则表达式语法来实现文本匹配和操作。

1. 简单字符匹配：

`\.`: 匹配任意一个字符。

`[abc]`: 匹配字符 a、b 或 c。

`[^abc]`: 匹配除了字符 a、b 和 c 之外的任意字符。

`[0-9]`: 匹配任何一个数字。

`\\`: 转义字符，用于匹配特殊字符本身，如 `\\.` 匹配句点。

2. 重复匹配：

`*`: 匹配前一个元素出现零次或多次。

`+`: 匹配前一个元素出现一次或多次。

`?`: 匹配前一个元素出现零次或一次。

`{n}`: 匹配前一个元素出现恰好 n 次。

`{n,}`: 匹配前一个元素出现至少 n 次。



3. 特殊字符:

`\^`: 匹配行的开头。

`\$`: 匹配行的结尾。

`\<`: 匹配单词的开头。

`\>`: 匹配单词的结尾。

4. 字符类:

`\s`: 匹配任何空白字符（空格、制表符等）。

`\S`: 匹配任何非空白字符。

`\w`: 匹配任何字母、数字或下划线字符。

`\W`: 匹配任何非字母、非数字和非下划线字符。

`\D`: 匹配任何非数字字符。

`\d`: 匹配任何数字。`\d+`: 匹配一个或多个连续的数字。`\d*`: 匹配零个或多个连续的数字。`-?\d+`: 匹配一个或多个连续的数字，可包含一个可选的负号。



Shell编程的基础知识

1. 脚本的创建：使用文本编辑器创建一个以`.sh`为扩展名的文件，例如 `script.sh`。
2. 声明解释器：在脚本的第一行添加解释器的路径，指定要使用的Shell解释器，例如 `#!/bin/bash`。
3. 变量：使用变量来存储和处理数据。可以使用**赋值运算符 (=)** 将值分配给变量，例如 `name="John"`。在使用变量时，需要在**变量名前加上 \$ 符号**，例如 `echo $name`。
4. 输入输出：脚本可以从用户输入读取数据，并向终端输出结果。**read 命令**用于读取用户输入，`echo 命令`用于输出文本到终端。
5. 注释：使用注释提供代码解释和说明。在Shell中，注释以 **# 开头**，后面的内容将被视为注释，不会被执行。
6. 条件判断与循环：可以使用条件语句（如 `if`、`elif` 和 `else`）来根据条件执行不同的命令块。循环结构（如 `for`、`while` 和 `until`）用于重复执行一系列命令。
7. 函数：可以定义函数以封装一系列命令，并在脚本中多次调用。函数通过 **`function_name() { ... }`** 的形式定义，然后可以使用 `function_name` 来调用函数。
8. 运行脚本：通过在终端中输入脚本的文件名来运行脚本，例如 `./script.sh`。在运行脚本之前，需要确保脚本文件具有执行权限，可以使用 **`chmod +x script.sh`** 命令添加执行权限。

```
#!/bin/bash
# 定义函数，用于打印欢迎消息
function welcome {
    echo "Welcome to Shell
Programming!"
}
# 调用函数
welcome
# 提示用户输入名字
echo "Please enter your name:"
read name
# 判断名字是否为空，并输出相应的
消息
if [ -z "$name" ]; then
    echo "Hello, stranger!"
else
    echo "Hello, $name!"
fi
# 提示用户输入数字
echo "Please enter a number:"
read number
# 循环打印数字从1到用户输入的数字
for ((i=1; i<=number; i++)); do
    echo "Number: $i"
done
# 结束脚本
exit 0
```

注意事项



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

在编写Linux命令时，有一些注意事项可以帮助您提高命令的效率和正确性。以下是一些常见的注意事项：

- 1. 命令格式：**大部分Linux命令都遵循特定的格式，通常以命令本身开始，后跟选项和参数。命令和选项之间使用空格分隔。确保按照正确的格式书写命令，以避免语法错误。**注意区分大小写！**
- 2. 选项和参数：**了解命令支持的选项和参数，并在书写命令时使用正确的标记和值。通常，选项以一个或两个短横线（-）或一个长横线（--）开始，然后是选项标记或名称，有时也需要提供选项的值。
- 3. 文件路径：**当涉及到文件和目录路径时，确保路径的准确性和正确性。使用相对路径或绝对路径，具体取决于您的需求。
- 4. 通配符：**在命令中使用通配符可以方便地匹配多个文件或目录。常见的通配符有星号（*）和问号（?）。确保在正确的位置使用通配符，并注意通配符可能带来的潜在风险。
- 5. 输入输出重定向：**了解如何将命令的输入和输出重定向到文件或其他命令。使用重定向符号（>、>>、<、|）来指示输入和输出的流向。

注意事项



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

6. **管道**：利用管道符号（|）将多个命令连接在一起，以实现数据流的传递和处理。熟悉如何使用管道来组合命令并实现更复杂的操作。
7. **引号和转义字符**：在命令中使用引号（单引号或双引号）可以保留特殊字符的原义，并将其视为文本而非命令元素。同时，了解如何使用转义字符（\）来转义特殊字符。
8. **权限和安全性**：特别是在需要进行系统级操作或管理任务时，确保您具有必要的权限并谨慎行事。避免以 root 用户身份长时间工作，以减少意外操作的风险。
9. **命令参数提示**：使用命令行自动完成功能（按下 Tab 键）可以帮助您快速输入命令和路径，并减少可能的拼写错误。
10. **学习和文档**：Linux 有很多命令和工具，不同命令的选项和用法也各不相同。阅读相关的文档、手册页或使用 `man` 命令查看命令的详细说明，以便了解命令的使用方法和最佳实践。

通过遵循这些注意事项，您可以更有效地编写 Linux 命令，并减少错误的发生。记住，熟能生巧，随着您对 Linux 操作系统和命令行的熟悉度的提高，编写命令将变得更加容易和自然。

超算使用注意点



雲南農業大學
Yunnan Agricultural University

云南省生物大数据重点实验室



生信帮

- 尽量使用软链接 `ln -s`
- 文件命名规范及快速匹配，目录及脚本命名尽量通俗易懂
- 每个目录下须有统一命名的*.sh文件（eg. work.sh）用于记录该目录下的所有操作
- 同级目录下文件数量<1000
- 程序输出文件较大，运行程序前请先用 `df` 查看空间使用情况
- 警惕超大的*.nohup,
- 项目结束时请清理无用数据，重要的数据及脚本打包(压缩之后的文件尽量不要超过50G)
- 工作空间的程序、脚本、数据链接进行版本管理，**定期备份**。



雲南農業大學
Yunnan Agricultural University

生信帮，答你所问

云南省生物大数据重点实验室