HPRC Short Course

Introduction to UNIX/Linux
For Assistance…

Website: hprc.tamu.edu
Email: help@hprc.tamu.edu
Telephone: (979) 845-0219
Visit us in person: Henderson Hall, Room 114A

Help us, help you -- we need more info

• Which Cluster
• UserID/NetID
• Job ID(s) if any
• Location of your job files, input/output files
• Application used & module(s) loaded if any
• Error messages
• Steps you have taken, so we can reproduce the problem
Logging in to the system

• SSH (secure shell)
  – The only program allowed for remote access; encrypted communication; freely available for Linux/Unix and Mac OS X hosts;

• For Microsoft Windows PCs, use MobaXterm
  • [https://hprc.tamu.edu/wiki/HPRC:MobaXterm](https://hprc.tamu.edu/wiki/HPRC:MobaXterm)
    – You are able to view images and use GUI applications with MobaXterm
  – or Putty
    – You can not view images or use GUI applications with PuTTY
Your Login Password

- Both State of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts;
- Don’t write down passwords;
- Don’t choose easy to guess/crack passwords;
- Change passwords frequently
Common Operating Systems (OS)

- **Linux**
  - RedHat
  - Debian
  - CentOS
  - Ubuntu

- **Mac**

- **Windows**
  - MS-DOS

**UNIX**
Using SSH - MobaXterm (on Windows)
https://hprc.tamu.edu/wiki/HPRC:MobaXterm

Use `titan.tamu.edu` as Remote host name.
Using SSH (on a Linux/Unix Client)

You may see something like this the first time you connect to the remote machine from your local machine:

```
ssh -X NetID@titan.tamu.edu
```

Host key not found from the list of known hosts.
Are you sure you want to continue connecting (yes/no)?

Type yes, hit enter and you will then see the following:

```
Host 'titan.tamu.edu' added to the list of known hosts.
NetID@titan.tamu.edu's password:
```

**Mac** users may need to use `ssh -Y` to enable X11 so you can view images and use GUI software

```
ssh -Y NetID@titan.tamu.edu
```

The host `titan.tamu.edu` is for this Linux class only and is not an HPRC cluster. To use the HPRC clusters, you must apply for an HPRC account.
UNIX Terminal Attributes

File and directory names are colored based on their attributes such as permissions and extension.

- Symbolic link: TURQUOISE
- Executable file: GREEN
- Compressed files: RED
- Image files: PURPLE
- Directories: BLUE
- Text files: WHITE
Finding your way around the Linux directory structure

```
/  
  /root
  /tmp
  /etc
  /home
     /sarah
     /chris
        /docs
        /scripts
  /var
     /log
     /www
```

```
/  
  /root
  /tmp
  /etc
  /home
     /home/sarah
     /home/chris
        /home/chris/docs
        /home/chris/scripts
  /var
     /var/log
     /var/www
```
Where Am I?

**pwd** command (print working directory)

Linux commands in green for you to type

```
pwd
```

command output in blue

```
/home/user_NetID
```

list contents of your pwd

```
ls
```
Linux Commands Have Options

Leave a space between the command and the options

double dash means there is a single option which is usually a descriptive word

```
ls --all
```

---all  show all files including

hidden files which begin with .

single dash means each character is an option

```
ls -a -l
```

-a  show all files including hidden

-1  show file details

most options can be combined behind one dash

```
ls -a -l
```

-a  show all files including hidden

-1  show file details

.    current working directory

..   parent directory
## Search for Linux Commands Options

Search the manual page for the Linux command `ls`

<table>
<thead>
<tr>
<th>Command</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>man ls</td>
<td>Move up one page</td>
</tr>
<tr>
<td></td>
<td>Move down one page</td>
</tr>
<tr>
<td></td>
<td>Move down one page</td>
</tr>
<tr>
<td></td>
<td>Move up and down</td>
</tr>
<tr>
<td></td>
<td>Search the man page for the text 'all'</td>
</tr>
<tr>
<td>/all</td>
<td>Search forward for next found match</td>
</tr>
<tr>
<td>n</td>
<td>Search backwards for next found match</td>
</tr>
<tr>
<td>N</td>
<td>Go to first line</td>
</tr>
<tr>
<td>g</td>
<td>Go to last line</td>
</tr>
<tr>
<td>G</td>
<td>Quit</td>
</tr>
<tr>
<td>q</td>
<td></td>
</tr>
</tbody>
</table>
File and directory names

Commonly used:
A-Z
a-z
0-9
.
- dash
_ underscore

Avoid using:
spaces
() parenthesis
" ' quotes
? Question mark
$ Dollar sign
* Asterisk
\ back slash
/ forward slash
: colon

- Avoid spaces.
- File and directory names are case sensitive.
- Avoid spaces in the file name ("my data file.txt" vs "my_data_file.txt").
- Avoid creating files on your Windows computer and copying to UNIX especially with spaces in the file name.
Common Directory Commands

**mkdir** command to make a new directory:
```
mkdir my_dir
```

**cd** to change to another directory:
```
cd my_dir
```

**rmdir** to remove an empty directory:
```
rmdir my_dir
```
Changing Directories: the `cd` cmd

- Return to your home directory
  
  ```
  cd
  cd ~
  cd ~/
  ```

- To switch to the parent directory of the current directory:
  
  ```
  cd ..
  ```

- Return to previous pwd
  
  ```
  cd -
  ```

- See the directory structure of your pwd
  
  ```
  tree
  ```
Clear Contents on Screen

Type **clear** command to clear screen contents.

You can still scroll up in your terminal to see past contents

You can also use Ctrl+l to clear contents
Absolute vs. Relative Path

If you are in the `project` directory:

```
pwd
```

```
/home/chris/project
```

The relative path to the README file is `../docs/README`

```
ls ../docs/README
```

The absolute path to the README file `/home/chris/docs/README`

```
ls /home/chris/docs/README
```
The **Gedit** Text Editor

From the Linux terminal command line enter this command to start gedit and edit a file called `my_favorite_foods.txt`

```bash
gedit my_favorite_foods.txt &
```

The `&` will detach gedit from the terminal so you can continue to use the terminal.

Recommendations for naming files:
- Use all lowercase characters
- Separate words with an underscore
- Make the filename concise and very descriptive of the file contents even if the name seems long

Some common file extensions are:
- `.pl` a Perl script
- `.py` a Python script
- `.gz` a file that has been compressed (zipped) to reduce file size (.zip)
- `.txt` a generic text file
- `.tsv` tab separated values (columns are separated by a tab)
- `.csv` comma separated values
- `.jar` a Java Archive file
Use Tab to Complete a File Name

Type the first few characters of the file name

```
ls my
```

then hit the tab key to autocomplete the file name

```
ls my_favorite_foods.txt
```

then hit enter to see the command results

If the tab key did not complete the file name then either the file does not exist or there are two or more files that begin with the same characters in which case you need to hit tab twice then type a few more characters and hit tab again to complete.
Count the Lines in a File

```
wc my_favorite_foods.txt
```

29 109 876 my_favorite_foods.txt

What does the output mean?

Use the man page for `wc` to find out.

How do you just print the newline counts?
Download a File from the Web to Your pwd

Click this link

Index of /goldenpath/hg19/chromosomes

hgdownload.cse.ucsc.edu/.../hg19/ch... University of California, Santa Cruz.

Files included in this directory: - chr*.fa.gz: compressed FASTA sequence of ... we recommend that you use ftp rather than downloading the files via our website.
Don't Left Click and download a file to your desktop

go to the bottom of the web page
Right Click and Copy the URL so you can download it directly to your Linux pwd
Copying a File Directly to a LINUX Directory

Use the `wget` command to get a file from a URL

Type `wget` then a space then right click and paste the URL

```
wget http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/chrY.fa.gz
```

```
wget http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/md5sum.txt
```

List the directory contents to see the file with details (-l) and human readable file sizes (-h)

```
ls -lh
```
Copying and Renaming Files

Use the tab key to help prevent typos when typing filenames and directories.

1a. \texttt{cp} \texttt{ch} (then hit tab)

UNIX will complete the file name for you.

1b. \texttt{cp} \texttt{chrY.fa.gz}

Make a copy of the \texttt{chrY.fa.gz} file called \texttt{chrY_copy.fa.gz}.

1c. \texttt{cp} \texttt{chrY.fa.gz} \texttt{chrY_copy.fa.gz}

Rename the \texttt{chrY_copy.fa.gz} file to \texttt{chrY_hg19.fa.gz}.

\texttt{mv} \texttt{chrY_copy.fa.gz} \texttt{chrY_hg19.fa.gz}
Deleting Files: The `rm` Command

```
rm [options] [file_name]
```

- Commonly used options with the `rm` command:
  - `-i` prompt user before any deletion

- Use the wildcard `*` to list all files ending with `gz`

Exercise:
```
ls *gz
rm -i chrY.fa.gz
ls
```
Displaying File Contents

- Check the file size before attempting to open with a text editor

```bash
ls -lh md5sum.txt
```
```bash
  cat  md5sum.txt
```

- `cat` prints all the contents of a file(s) to the screen.
- The `more` command, and its improved version `less`, display a text file one page (screen) at a time.
  - Hit space bar for next page; `less` search function is similar to man pages
  - Type `q` to quit
  - Use `zmore` or `zless` for compressed files (.gz)

```bash
  more  md5sum.txt
  less  md5sum.txt
  zmore chrY_hg19.fa.gz
```
Displaying File Contents

- Use `head` and `tail` commands to see first and last 10 lines of a file respectively

```
head md5sum.txt
tail md5sum.txt
```

- `head` and `tail` are not for compressed files (.gz)
- There is not a zhead or ztail command
- Use `zcat` together with the `head` or `tail` command
- Use the pipe `|` to send output of first command to the second command

```
zcat chrY_hg19.fa.gz | head
zcat chrY_hg19.fa.gz | tail
```
**grep – Searching for Pattern(s) in Files**

```bash
grep [options] PATTERN [files ...]
```

- `grep chrX md5sum.txt`

  the `-i` option is to ignore case

- `grep -i chrX md5sum.txt`

  Count the number of lines that match pattern

- `grep -c random md5sum.txt`
grep – Searching For Pattern(s) in Files

Search multiple matches

```
grep -e chrX -e chrY md5sum.txt
```

Exclude a pattern; show non-matching lines

```
grep -v random md5sum.txt
```

Use `zgrep` for compressed files (.gz)

```
zgrep chr chrY_hg19.fa.gz
```
Passing the Output of Linux Commands

Use the pipe character | to send results to another command

Search for the string 'chrX' in the output of md5sum.txt

```
cat md5sum.txt | grep chrX
```

Use the standard output redirect operator > to **create** a new file

```
grep chrX md5sum.txt > chr_xy.txt
```

Use the standard output redirect operator >> to **append** to a file

```
grep chrY md5sum.txt >> chr_xy.txt
```
History of Your Commands

- Your commands are saved to a file in your home directory (`.bash_history`)
- You can use the up/down arrows to scroll through your previous commands
- Type `history` to see your previously entered commands

See the last 10 commands

```
history
history | tail
```

Search your history commands using `|` and `grep`

```
history | grep wget
```
Types of File: the `file` cmd

`file [name]`

Displays a brief description of the contents or other information for a file or related object.

```
file md5sum.txt
```

`md5sum.txt`: ASCII English text

```
file chrY_hg19.fa.gz
```

`chrY_hg19.fa.gz`: gzip compressed data, was “chrY.fa”, last modified ...
CRLF Line Terminators

Windows editors such as Notepad will add hidden Carriage Return Line Feed (CRLF) characters that will cause problems with many applications.

```
cd ~/intro_to_linux/scripts/

file DOS_script.sh

DOS_script.sh: ASCII English text, with CRLF line terminators

dos2unix DOS_script.sh
file DOS_script.sh

DOS_script.sh: ASCII English text
```
Editing an ASCII file

- There are many editors available under Linux.
- Text mode
  - nano (simple)
  - vi or vim (more advanced)
  - emacs (more advanced)
- Graphic mode (require X11)
  - gedit
  - xemacs / gvim
- Be aware that a text file edited under Windows editors will most likely add CRLF characters. Use `dos2unix` to convert a DOS/Windows edited text file to UNIX format.
The Linux Filesystem
File Attributes: A look with `ls`

```
[user_NetID@titan ~]$ ls -l
```

<table>
<thead>
<tr>
<th>File Permission</th>
<th>File Type</th>
<th>Owner</th>
<th>Group</th>
<th>Hard Link Count</th>
<th>File Name</th>
<th>File Size in Bytes</th>
<th>File Modification Time</th>
<th>File Modification Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>drwx-</td>
<td>directory</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>7</td>
<td>abaqus_files</td>
<td>121 Sep 9 10:41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rw-</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>fluent-unique.txt</td>
<td>2252 Aug 24 10:47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rw-</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>fluent-use1.txt</td>
<td>13393007 Aug 24 10:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rw-</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>fluent.users</td>
<td>533 Aug 24 11:23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>drwxr-xr-x</td>
<td>directory</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>3</td>
<td>man</td>
<td>17 May 7 16:56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rw-</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>myHomeDir.tar</td>
<td>24627200 Sep 9 10:49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lrwxrwxrwx</td>
<td>link</td>
<td>root</td>
<td>root</td>
<td>1</td>
<td>README -&gt; /usr/local/etc/README</td>
<td>21 May 28 16:11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rwx-</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>spiros-ex1.bash</td>
<td>162 Sep 7 12:20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-rwx-x-x-x-x</td>
<td>file</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>1</td>
<td>split.pl</td>
<td>82 Aug 24 10:51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>drwxr-xr-x</td>
<td>directory</td>
<td>user_NetID</td>
<td>user_NetID</td>
<td>2</td>
<td>verifyOLD</td>
<td>6 May 5 11:32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Directory Permissions

- The meanings of the permission bits for a directory are slightly different than for regular files:
  - `r` permission means the user can list the directory’s contents
  - `w` permission means the user can add or delete files from the directory
  - `x` permission means the user can cd into the directory; it also means the user can execute programs stored in it
- Notice that if the file is a directory, the leading bit before the permissions is set to `d`, indicating directory.
There are 3 sets of permissions for each file:
- 1st set - user (the owner)
- 2nd set - group (to which file owner belongs)
- 3rd set - other (all other users)

The r indicates read permission
The w indicates write permission
The x indicates execute permission
Changing Attributes: The `chmod` cmd

`chmod [options] [permission mode] [target_file]`

```bash
cd ~/temp/hg19

chmod 755 chr_xy.txt   (the permissions will be set to -rwxr-xr-x)
chmod o-x chr_xy.txt   (the permissions will change to -rwxr-xr--)
chmod ug-x chr_xy.txt  (the permissions will change to -rw-r--r--)
chmod g+w chr_xy.txt   (the permissions will change to -rw-rw-r--)
```

- `u` = user, `r` = read
- `g` = group, `w` = write
- `o` = other, `x` = execute
- `-x` = remove executable permissions
- `+x` = enable executable permissions
Transfer Data From Windows Host to Linux Host

On a Windows system, you can use MobaXterm to transfer files to/from an HPRC cluster or other remote machine

https://hprc.tamu.edu/wiki/HPRC:MobaXterm
File Transfers Using FileZilla

The FileZilla Client:
- Available on Windows, OS X, and UNIX/Linux
- Allows permissions to be preserved or implied
- Easy to use without previous experience
- Can drag and drop files

Download from:
https://filezilla-project.org

Connect with your TAMU NetID login using Host: sftp://titan.tamu.edu
Transfering Files Using FileZilla

Local files

Remote files
Transfer Data between Hosts: the `scp` cmd

```
scp [[user@]host1:]filename1 [[user@]host2:]filename2
```

Copy a file **from** your Linux desktop **to** your titan home directory

```
scp myfile1 NetID@titan.tamu.edu:
```

Copy and rename file **from** your Linux desktop **to** dir1 in your $HOME

```
scp myfile1 NetID@titan.tamu.edu:dir1/remote_myfile1
```

Copy a file **to** your Linux desktop **from** your titan home directory

```
scp NetID@titan.tamu.edu:myfile2 ./
scp NetID@titan.tamu.edu:myfile2 local_myfile2
scp -r NetID@titan.tamu.edu:dir3 local_dir/
```
Bash Environment Variables

Use all caps for Bash Environment variable. A-Z 0-9 _
Use lowercase for the variables that you create. a-z 0-9 _

HOME Pathname of current user’s home directory
PATH The search path for commands.

Use the `echo` command to see the contents of a variable

```bash
echo $HOME

/home/user_NetID
```
list contents of your $HOME directory

```bash
ls $HOME
```
The Search Path

- The shell uses the PATH environment variable to locate commands typed at the command line.
- The value of PATH is a colon separated list of full directory names.
- The PATH is searched from left to right. If the command is not found in any of the listed directories, the shell returns an error message.
- If multiple commands with the same name exist in more than one location, the first instance found according to the PATH variable will be executed.

```bash
export PATH=$PATH:/home/user_NetID/bin
```

```
PATH=/opt/TurboVNC/bin:/software/tamusc/local/bin:/software/lsf/9.1/linux2.6-glibc2.3-x86_64/bin:
/usr/local/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/sbin:/usr/lpp/mmfs/bin:/opt/ibutils/bin:/home/user_NetID/bin
```

Add a directory to the PATH for the current Linux session.
Viewing image files with Eye of Gnome image viewer

eog ~/intro_to_linux/images/

Click to see the next image in the current directory, or use the arrow keys.
The `find` Command

```
find [target dir] [expression]
```

cd  
go to your home directory

find ./ -name "*.txt"

find $HOME -mtime -2 -type f

find $HOME -mtime +5

find $HOME -iname temp -type d

. is the same as ./ which means current directory

Anything in ./ ending in .txt

Files Modified within last 2 days

Modified more than 5 days ago

case insensitive search for a directory named ‘Temp’
Command Aliases

- Set a shortcut command or alias for the `grep` command
  ```bash
  alias grep='grep --color=auto'
  ```

- Try the new alias
  ```bash
  grep chrX ~/temp/hg19/md5sum.txt
  ```

- To save alias for each login session, add the alias command to your `~/.bashrc` file
  - After updating the `~/.bashrc` file, you will need to run the source command for the current session
  ```bash
  gedit ~/.bashrc
  ```
  ```bash
  # .bashrc
  # Source global definitions
  if [ -f /etc/bashrc ]; then
    . /etc/bashrc
  fi
  # User specific aliases and functions
  alias grep='grep --color=auto'
  ```
  ```bash
  source ~/.bashrc
  ```
The ‘tar’ Command

```
tar [options] [tar file] [file or dir name]
```

- Used to “package” multiple files (along with directories if any) into one file suffixed with a `.tar` suffix by convention.
- Commonly used options:
  - `x` extract files from a tar
  - `c` create a new tar
  - `t` list the contents of a tar
  - `v` verbosely list files processed
  - `f` use the specified tar file
  - `z` the tar file is compressed
The `tar` Command - Exercise

- **cd**
  
  Go to your home directory

- **tar** `-cvf` `my_hg19.tar temp`
  
  Package the temp directory into a file called `my_hg19.tar`

- **tar** `-cvzf` `my_hg19.tar.gz temp`
  
  Package the temp directory into a compressed file called `my_hg19.tar.gz`

- **tar** `-tzf` `my_hg19.tar.gz`
  
  Show the contents of the compressed tar file

- **mv** `temp temp_orig`
  
  Change the name of your original temp directory so you don’t overwrite it

- **tar** `-xvzf` `my_hg19.tar.gz`
  
  Extract all contents from the compressed tar file

---

### Options

- `-c` = create
- `-f` = file
- `-t` = list contents
- `-v` = verbose
- `-z` = zipped
- `-x` = extract
The `tar` Command

- Be careful when extracting files
  - `tar` will overwrite existing files.
- Where files are extracted depends on how they were packaged.
- Always a good idea to check Table of Contents (`-t` option) before extraction.
Shell Scripts

- A shell script is a file that contains one or more UNIX commands that can be run as a single batch of commands.
- Shell script error messages are sent to `/dev/stderr` (stderr).
- Shell script non-error output is sent to `/dev/stdout` (stdout).
- Add comments to your script using `#`.
- You can redirect the stderr and stdout content to a file using redirection operators.
Redirection Operators

<  redirects input

>  redirects output

>> appends output

<< input from here document (search the web for examples)

2> redirects error

&> redirects output and error

>>& redirects output and error

2>>&1 redirects error to where output is going

1>>&2 redirects output output to where error is going

bsub < job_script.sh

cmd > out.txt

cmd >> out.txt

cmd >& out-error.txt

cmd &> out-error.txt

cmd 2> error.txt

cmd &> out-error.txt

cmd >& out-error.txt
Shell Script Exercise

create a shell script

```bash
#!/bin/bash
# HPRC shell script exercise
my_city='College Station'
echo "Howdy $my_city" > my_city.txt
echo 'Howdy $my_city' >> my_city.txt
mkdir outdir
mv my_city.txt outdir/
cd outdir
```

make your shell script executable

```bash
chmod 755 ./my_script.sh
```

run your shell script

```bash
./my_script.sh
```

run your shell script send stdout to file

```bash
./my_script.sh > output.txt
```

send stdout and stderr to files

```bash
./my_script.sh > output.txt 2> err.txt
```

```bash
```
exit

# exit the terminal session
# can also use Ctrl+d to detach session
References

Here are some slides from TACC on the similar subject.

- Linux/Unix Basics for HPC: October 9, 2014 (with video) [TACC]
  - [https://portal.tacc.utexas.edu/-/linux-unix-basics-for-hpc](https://portal.tacc.utexas.edu/-/linux-unix-basics-for-hpc)
- Express Linux Tutorial: Learn Basic Commands in an Hour [TACC]
  - [https://portal.tacc.utexas.edu/c/document_library/get_file?uuid=ed6c16e9-bc9c-4b70-9311-5273b09508b8&groupId=13601](https://portal.tacc.utexas.edu/c/document_library/get_file?uuid=ed6c16e9-bc9c-4b70-9311-5273b09508b8&groupId=13601)
Additional Slides
Using SSH - Putty (on Windows)


Use ada.tamu.edu as host name.