High Performance Research Computing

A Resource for Research and Discovery



HPRC Short Course

Introduction to UNIX/Linux



For Assistance...

Website:hprc.tamu.eduEmail:help@hprc.tamu.eduTelephone:(979) 845-0219Visit 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
 - <u>https://hprc.tamu.edu/wiki/HPRC:MobaXterm</u>
 - You are able to view images and use GUI applications with MobaXterm
 - or *Putty*
 - <u>https://hprc.tamu.edu/wiki/HPRC:Access#Using_PuTTY</u>
 - 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)



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

	Session settings	.
MobaXterm Terminal Sessions	Image: SSH Image: SSH <th></th>	
Quick connect.	Basic SSH settings	
Saved session	Remote host * titan.tamu.edu 🖉 Specify username netid 🔊 Port 22 🐑	×
🖄 Macros 😿 Tools 🗱 Sessions	Advanced SSH settings Terminal settings Network settings Bookmark settings	Holp Holp UREvec Tunneling Settings Help X server Ext There T
UNREGISTERED VERSI	Solution Cancel	Welcome to MobaXterm ss < return > to start a new session or type a remote server na CygUtils plugin not found on your system.
titan	.tamu.edu as Remote host	name.

Texas A&M University High Performance Research Computing – https://hprc.tamu.edu

Use

A M

Using SSH (on a Linux/Unix Client)

ssh -X NetID@titan.tamu.edu

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

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

AAF -> AAF.py
AAF.py
aaf_tip.py
data.gz
image.jpg
phylip_src
phylokmer
README
run_aaf.sh



Finding your way around the Linux directory structure



/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

--all show all files including

hidden files which begin with .

single dash means each character is an option

-a show all files including hidden

-1 show file details

most options can be combined behind one dash

ls -al	 -a show all files including hidden -1 show file details
	. current working directory

Search for Linux Commands Options

Search the manual page for the Linux command 1s

man ls

Page up Page down Spacebar Mouse scroll wheel Move up one page Move down one page Move down one page Move up and down

/all
n
N
a
G G

search the man page for the text 'all' search forward for next found match search backwards next found match go to first line Go to last line guit

File and directory names

Avoid using:

" 'quotes

()

?

\$

*

spaces

Asterisk

colon

parenthesis

Dollar sign

back slash

Question mark

forward slash



• 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:

• Return to previous pwd

cd -

cd ..

• See the directory structure of your pwd

tree

cd mkdir temp **mkdir** temp/hq19 cd temp pwd cd hq19 pwd cd ../.. pwd cd pwd cd .. pwd cd 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

/root /tmp /etc /home /home/sarah /home/chris/project /home/chris/docs/README /var /var/log /var/www

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

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



Don't Left Click and download a file to your desktop

chrun qluuuuzzu.ta.qz chrUn gl000221.fa.gz chrUn gl000222.fa.gz chrUn gl000223.fa.gz chrUn gl000224.fa.gz chrUn gl000225.fa.gz chrUn gl000226.fa.gz chrUn gl000227.fa.gz chrUn gl000228.fa.gz chrUn gl000229.fa.gz chrUn gl000230.fa.gz chrUn gl000231.fa.gz chrUn gl000232.fa.gz chrUn gl000233.fa.gz chrUn gl000234.fa.gz chrUn gl000235.fa.gz chrUn gl000236.fa.gz chrUn gl000237.fa.gz chrUn gl000238.fa.gz chrUn gl000239.fa.gz chrUn gl000240.fa.gz chrUn_ql000241.fa.qz chrUn gl000242.fa.gz chrUn gl000243.fa.gz chrUn gl000244.fa.gz chrUn gl000245.fa.gz chrUn gl000246.fa.gz chrUn_ql000247.fa.qz chrUn gl000248.fa.gz chrUn gl000249.fa.gz chrX.fa.gz chrY.fa.gz md5sum.txt



go to the bottom of the web page

Right Click and Copy the URL so you can download it directly to your Linux pwd

cholin glooo	227 fo dz	20 Mar 2000 00	.02
chrun glooo	227.1d.92	20-Mar 2009 09	02 2
chrun globo	220.1d.yz	20-Mar - 2009 09	102 6
chron gto	Open Link in	New Tab	0.
chrun glo	open Link in	1100	0
chron gto	Open Link in	New Window	9
chron gto	open enix in	The Manager	
chrun gto	Open Link in New Private Window		ww 1
chrun glo	Open Link in New Phote Window		
chrun glo			
chrun glo	Bookmark This Link		1
chrun glo	500 6000 M		1
chrun glo	Save Lin <u>k</u> As		1
chrlin glo	A second second second		
chrlin glo	Copy Link Lo	oc <u>a</u> tion	
chrUn alo			1
chrUn alo	Search Yahoo for "chrY.fa.gz"		1
chrUn alo			- 5
chrUn alo	Inspect Flen	nent (Q)	
chrUn alo	inspect Lten		- 15
chrUn alo	Zotero		> 1
chrUn al0	Lotero		
chrUn al0]
chrX fa d	Inspect Element with Firebug		4
off Arrang			
chrY.fa.gz		20-Mar-2009 09	:02 7.

ĂМ

Copying a File Directly to a LINUX Directory

cd ~/temp/hg19

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 (-I) 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. **cp** ch



UNIX will complete the file name for you

Make a copy of the chrY.fa.gz file called chrY_copy.fa.gz

Rename the chrY_copy.fa.gz file to chrY_hg19.fa.gz

mv chrY_copy.fa.gz 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



Displaying File Contents

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

ls -lh md5sum.txt

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)

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

grep – Searching for Pattern(s) in Files

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



• 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 1s

[user NetID@titan ~]\$ ls -1 total_37216 drwx----7 user NetID user NetID 121 Sep 9 10:41 abaqus files 2252 Aug 24 10:47 fluent-unique.txt 1 user NetID user NetID -rw----1 user NetID user NetID 13393007 Aug 24 10:40 fluent-use1.txt -rw----1 user NetID user NetID 533 Aug 24 11:23 fluent.users -rw-----3 user NetID drwxr-xr-x user NetID 17 May 7 16:56 man 1 user NetID user NetID 24627200 Sep 9 10:49 myHomeDir.tar -rw----21 May 28 16:11 README -> /usr/local/etc/README lrwxrwxrwx 1 root root 162 Sep 7 12:20 spiros-ex1.bash 1 user NetID user NetID -rwx-----1 user NetID user NetID 82 Aug 24 10:51 split.pl -rwx - -x - x2 user NetID 6 May 5 11:32 verifyOLD drwxr-xr-x user NetID group name file name user name file modification time hard link count file modification date file size in bytes file permissions **Texas A&M University** High Performance Research Computing – https://hprc.tamu.edu

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.

File Ownership and Permissions



- 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]

cd ~/temp/hg19

chmod 755 chr xy.txt (the permissions will be set to -nwxr-xr-x)

chmod o-x chr_xy.txt (the permissions will change to -nwxr-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 -nw-nw-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





Transfering Files Using FileZilla

	🛃 sftp://whomps@ada.tamu.edu - FileZilla	– o x	
	File Edit View Transfer Server Bookmarks Help		
	1 - V E C - + W 1 & V E R 7 A		
	Host: sftp://ada.tamu.edi Username: whomps Password: •••••••• Port:	Quickconnect 💌	
	Status: Listing directory /general/home/whomps Status: Calculating timezone offset of server Command: mtime ".ssh" Response: 1478885224 Status: Timezone offsets: Server: -21600 seconds. Local: -21600 seconds. Difference: 0 seconds.		
	Status: Directory listing successful		
Local files	Local site: H\Downloads\ 	Remote site: //general/home/whomps	Remote files
	Custom Office Templates Downloads MATLAB Mobiliterm	- ? , altair - ? , Altair - ? , altair	v
	Filename Filesize Filetype Last modified	Filename Filesize Filetype Last modified Permissions Owner/Gro	^
	SRECYCLE.BIN File folder 1/13/2017 2:54:14	aienv_fe File folder 10/1/2015 drwxrwxr-x whomps w	
	blocker-screens File folder 1/13/2017 2:58:22	aienv_fe File folder 5/20/2016 drwxrwxr-x whomps w	
	Lab I File folder 9/30/2014 2:30:45 Moha Yterm v8 3 File folder 11/0/2015 9:07:25	altair File folder 7/30/2015 drwxrwxr-x whomps w	
	desktop.ini 282 Configuration 12/12/2016 2:13:52	altair_lic File folder 7/30/2015 drwxr-xr-x whomps w	
		ansys File folder 9/29/2016 1:37: drwxrwxr-x whomps w	
	1 file and 4 directories. Total size: 282 bytes	44 files and 42 directories. Total size: 93,307 bytes	
	Server/Local file Direction Remote file Size Priority	Status	
	Queued files Failed transfers Successful transfers		
		🔂 🚥 Queue: empty 🛛 👁	

Texas A&M University High Performance Research Computing – https://hprc.tamu.edu

ĀМ

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

HOMEPathname of current user's home directoryPATHThe search path for commands.

Use the echo command to see the contents of a variable

echo \$HOME

/home/user_NetID

list contents of your \$HOME directory

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.

echo \$PATH

PATH=/opt/TurboVNC/bin:/software/tamusc/local/bin: /software/lsf/9.1/linux2.6-glibc2.3-x86_64/bin: /usr/local/bin:/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

export PATH=\$PATH:/home/user_NetID/bin

Viewing image files with Eye of Gnome image viewer

eog ~/intro_to_linux/images/



The find Command

find [target dir] [expression]

cd	go to your home directory	
find ./ -n	ame "*.txt"	Anything in ./ ending in .txt
find \$HOME	-mtime -2 -type f	Files Modified within last 2 days
find \$HOME	-mtime +5	Modified more than 5 days ago
find \$HOME	-iname temp -type d	case insensitive search for a directory named 'Temp'

is the same as ./ which means current directory

Ă Ň

Command Aliases

• Set a shortcut command or alias for the grep command

alias grep='grep --color=auto'

• Try the new alias

grep chrX ~/temp/hg19/md5sum.txt

- To save alias for each login session, add the alias command to your
 - ~/.bashrc file gedit ~/.bashrc
 - After updating the ~/.bashrc file, you will need to run the source command for the current session

source ~/.bashrc

.bashrc

Source global definitions
if [-f /etc/bashrc]; then
 . /etc/bashrc

fi

User specific aliases and functions
alias grep='grep --color=auto'

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

go to your home directory

Package the temp directory into a file called my_hg19.tar

tar -cvf my_hg19.tar temp

Package the temp directory into a compressed file called my_hg19.tar.gz

tar -cvzf my_hg19.tar.gz temp

Show the contents of the compressed tar file

tar -tzf my_hg19.tar.gz

Change the name of your original temp directory so you don't overwrite it

mv temp temp_orig

cd

Extract all contents from the compressed tar file

tar -xvzf my hg19.tar.gz

-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

<

>

<<

bsub < job_script.sh</pre>

command > out.txt

command >> out.txt

- input from here document (search the web for examples)
- 2> redirects error
 &> redirects output and error
 >& out-error.txt
 >& command >> out-error.txt
 >& command >> out-error.txt
 >& redirects error to where output is going
 1>&2

Shell Script Exercise



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]
 <u>https://portal.tacc.utexas.edu/-/linux-unix-basics-for-hpc</u>
- Express Linux Tutorial: Learn Basic Commands in an Hour [TACC]
 - <u>https://portal.tacc.utexas.edu/c/document_library/get_file?uuid=ed6</u>
 <u>c16e9-bcbc-4b70-9311-5273b09508b8&groupId=13601</u>

Additional Slides

Using SSH - Putty (on Windows)

https://hprc.tamu.edu/wiki/HPRC:Access:Windows#Using_PuTTY

