

Baseline Skills and Setup



Objectives

- Basic understanding of Virtual Machines and the terminal
- Review basic Linux commands
- Become familiar with Atmosphere and the Data Store at CyVerse



Outline

- VMs & terminals
- File system navigation
- Creating, copying, moving, deleting files
- Networking commands
- Software installation & compression commands



What is a Virtual Machine?

- A VM is an operating system which has been installed inside a simulated environment.



What is a Terminal?

- A terminal is a textual interface for interacting with a computer.
- Using the terminal, one can issue powerful and concise command-line instructions for the computer to follow.



Why use the command-line (terminal)?

- Most software for biological data analysis is used through UNIX command-line operations.
- Most of the servers for biological data analysis use Linux/Unix as their operating system.
- Data analysis on calculation servers are much faster since we can use more CPUs and RAM than in a PC or laptop (e.g. BTI's "Boyce" server has 64 cores and 1TB RAM).
- Large NGS data files can not be opened or loaded in most graphical software and web sites.



We will use the CyVerse platform today

About CyVerse

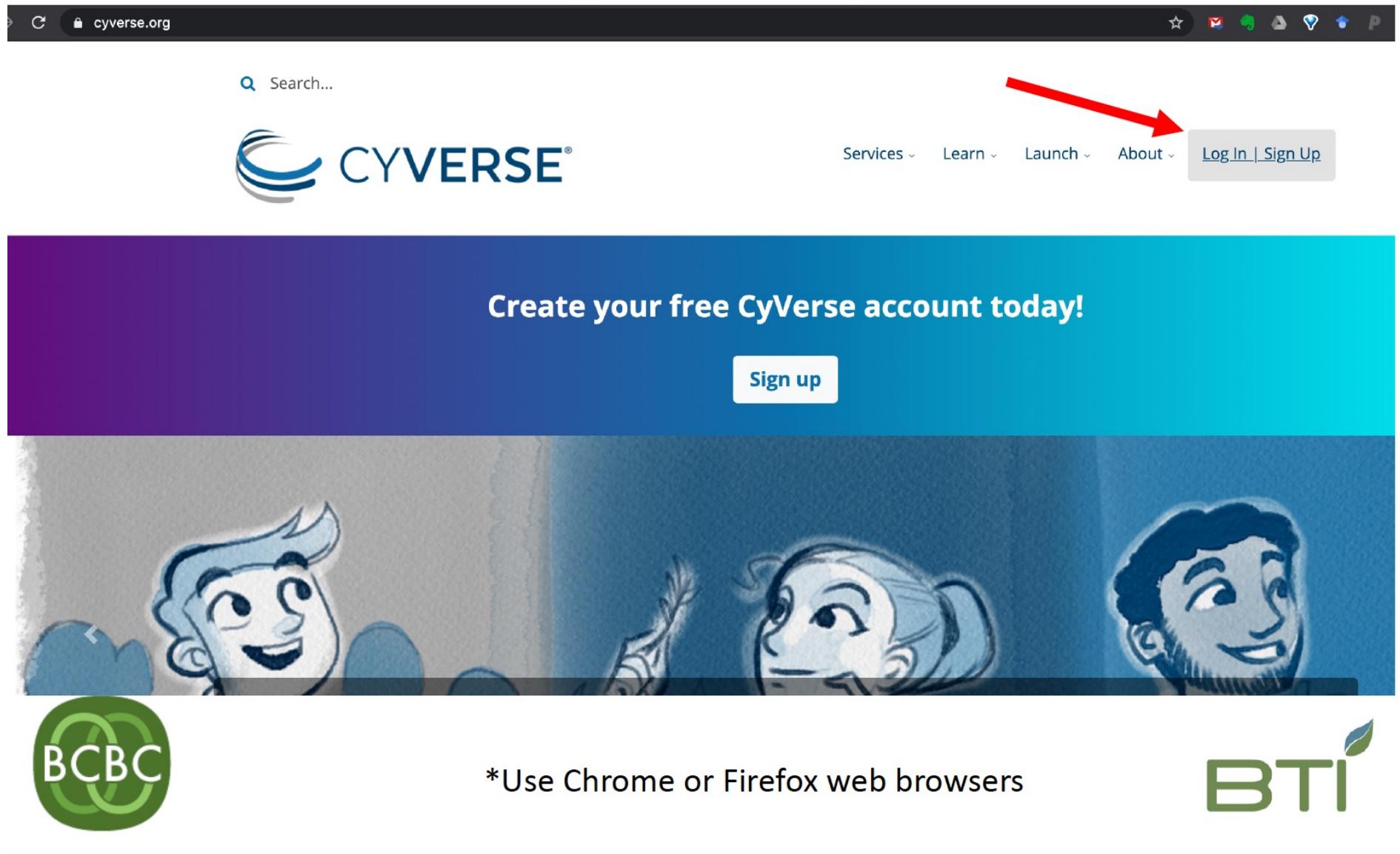
Vision: Transforming Science through Data-Driven Discovery

Mission: To design, deploy, and expand a national Cyberinfrastructure for Life Sciences research, and to train scientists in its use.

CyVerse provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. Our extensible platforms provide data storage, bioinformatics tools, data visualization, interactive analyses, cloud services, APIs, and more.

CyVerse is funded by the National Science Foundation's Directorate for Biological Sciences. We are a dynamic virtual organization, led by the University of Arizona in partnership with the Texas Advanced Computing Center and Cold Spring Harbor Laboratory, fulfilling a broad mission to enable data-driven, collaborative research.

Let's log in to CyVerse!



A screenshot of the CyVerse website login page. The page has a dark header with the URL "cyverse.org". Below the header is a search bar and the CyVerse logo. The main content area features a gradient background from purple to teal. Centered text reads "Create your free CyVerse account today!" with a "Sign up" button below it. A red arrow points to the "Log In | Sign Up" link in the top right corner. The bottom of the page shows a stylized illustration of three cartoon characters (two girls and one boy) looking up. Logos for BCBC and BTI are at the bottom.

cyverse.org

Search...

CYVERSE®

Services ▾ Learn ▾ Launch ▾ About ▾

[Log In | Sign Up](#)

Create your free CyVerse account today!

Sign up

BCBC

*Use Chrome or Firefox web browsers

BTI

The CyVerse User Portal interface displays a sidebar with navigation icons (grid, list, calendar, help) and a main content area titled "My Services".

Services:

- CoGe**: Online system for quick and easy retrieval and comparison of genomic information and sequences. [LAUNCH](#)
- Discovery Environment**: Use hundreds of bioinformatics apps and manage data in the CyVerse Data Store from a simple web... [LAUNCH](#)
- Atmosphere**: Cloud computing platform for CyVerse. [LAUNCH](#) (highlighted with a red arrow)
- Data Commons**: A unified system for managing and sharing your data across CyVerse's tools and services. [LAUNCH](#)

Top Right: HELP, ACCOUNT

Bottom Center: CLOSE AND DON'T SHOW AGAIN



Let's launch Atmosphere!



The screenshot shows the CYVERSE dashboard interface. At the top, there is a blue header bar with the CYVERSE logo, a navigation menu (Dashboard, Projects, Images, Help), and a user account dropdown (apowell). Below the header is a section titled "Getting Started" with three cards:

- Launch New Instance**: An icon of a rocket ship. Description: "Browse Atmosphere's list of available images and select one to launch a new instance."
- Browse Help Resources**: An icon of a cloud with a question mark. Description: "View a video tutorial, read the how-to guides, or email the Atmosphere support team."
- Change Your Settings**: An icon of a gear with a wrench. Description: "Modify your account settings, view your resource quota, or request more resources."

Below the "Getting Started" section is a "Resources Used" section with a "NEED MORE ?" button. It includes an "Allocation Source" chart showing allocation used (0%) and a "2 Instances" chart showing active networking and active instances.

Allocation Source

Allocation 0%
Percent of Allocation Used: 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100

2 Instances

active - networking active



[Resources](#) > NMW2021

NMW2021

Allocation Source

apowell

Allocation Used

0% of 336 AUs from apowell

Instance Details

Status ● Active

Activity N/A

Size large3 (CPU: 8, Mem: 64 GB, Disk: 480 GB root)

IP Address 128.196.142.17 [Copy](#)

Launched May 10, 2021 (35 minutes ago)

Based on [NMW2021 v1.0](#)

Actions

Report

Image

Suspend

Shelve

Stop

Reboot

Redeploy

Delete

Links

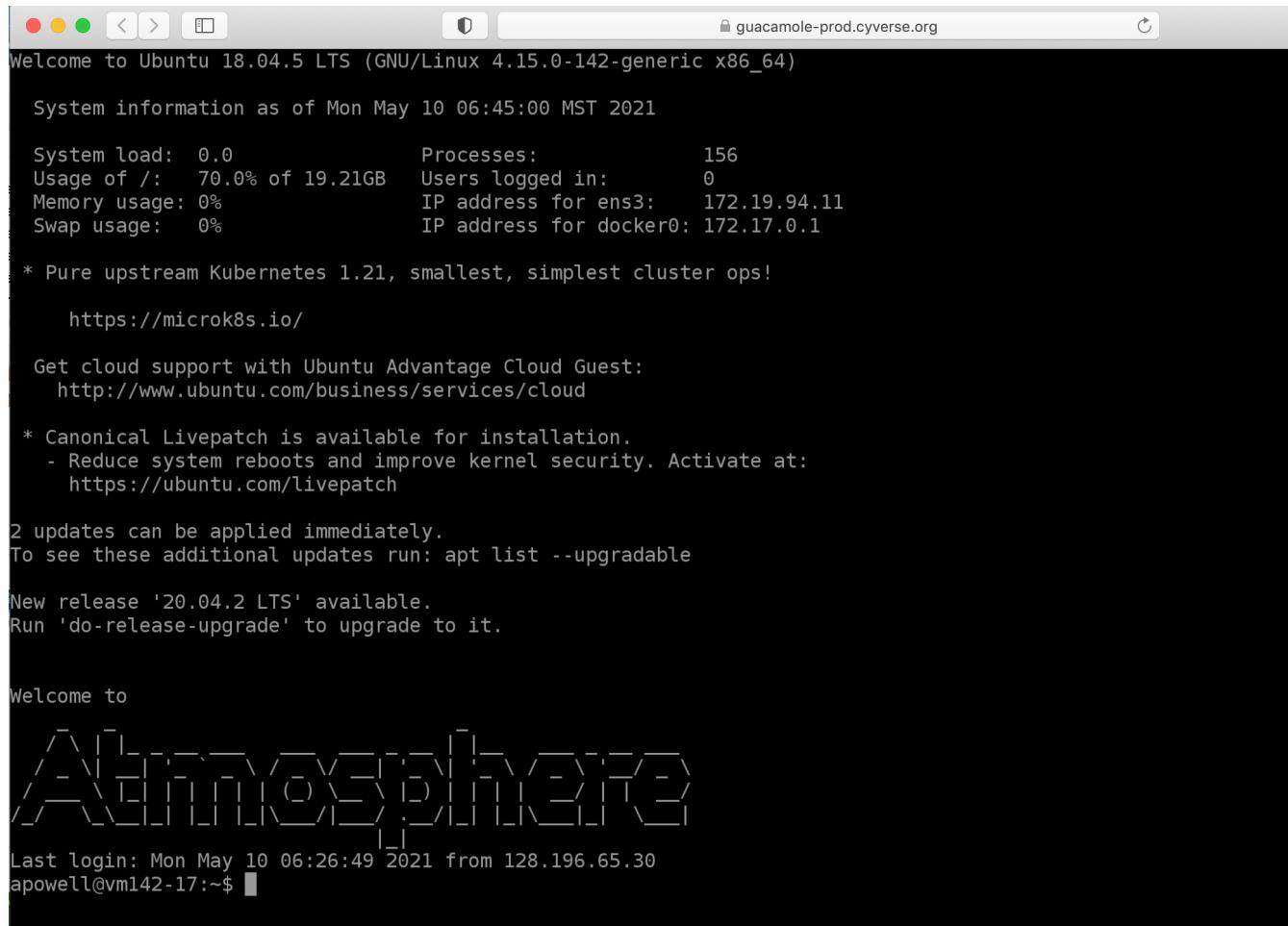
Open Web Shell ↗



Exercise 1:
Open the Atmosphere
terminal



Atmosphere Terminal



A screenshot of a macOS-style terminal window titled "guacamole-prod.cyverse.org". The window displays a terminal session on an Ubuntu 18.04.5 LTS system. The session starts with a welcome message and system information for May 10, 2021. It then lists various system metrics and resources. Following this, it provides links to Kubernetes documentation and Canonical Livepatch. It also mentions available updates, specifically a '20.04.2 LTS' release. The terminal concludes with a standard "Last login" message and a prompt for user "apowell".

```
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 4.15.0-142-generic x86_64)

System information as of Mon May 10 06:45:00 MST 2021

System load: 0.0          Processes:           156
Usage of /: 70.0% of 19.21GB  Users logged in:      0
Memory usage: 0%
Swap usage:  0%          IP address for ens3:   172.19.94.11
                           IP address for docker0: 172.17.0.1

* Pure upstream Kubernetes 1.21, smallest, simplest cluster ops!

https://microk8s.io/

Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

* Canonical Livepatch is available for installation.
- Reduce system reboots and improve kernel security. Activate at:
https://ubuntu.com/livepatch

2 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

New release '20.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Welcome to

A stylized logo composed of various line segments forming a circular pattern, resembling the Ubuntu logo.

Last login: Mon May 10 06:26:49 2021 from 128.196.65.30
apowell@vm142-17:~$ █
```



Command-line File System Navigation

- The cheat sheet you have been provided contains a list of common commands for navigating the file system.
- But what *is* the file system?

<https://github.com/bcbc-group/NMGWorkshop2021>



The image shows a 'UNIX Command-Line Cheat Sheet' from the BTI-SGN Bioinformatics Course 2016. The sheet is a grid of command descriptions. A red box highlights the 'File system Commands' section, which includes common commands like ls, cd, cp, mv, rm, and rm -r. The entire sheet is set against a background featuring a stylized sun icon in the top right corner.

File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info
apt-get install	installs programs in linux

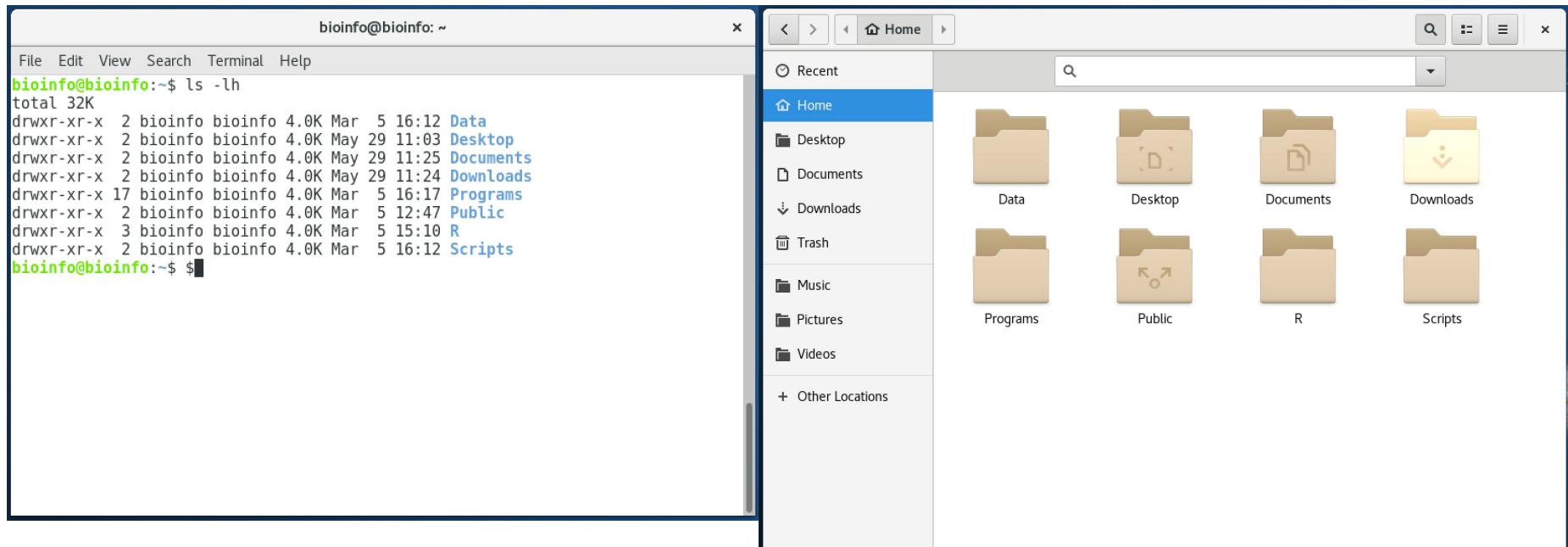
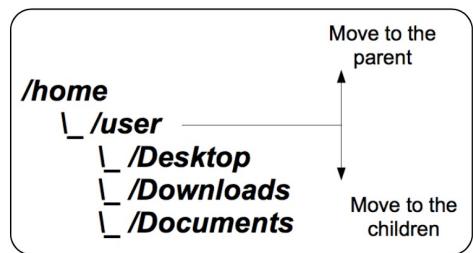
Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file content
less -N file	includes line numbers
less -S file	does not wrap long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ":"	concatenates the lines of input files by commas
sed	transforms text

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers



The File System



Home and Root Directories

```
Noes-MacBook-Pro:~ Noe$ ls -lht
total 0
drwx-----+ 29 Noe  staff   986B May 31 11:24 Desktop
drwx-----@  8 Noe  staff   272B May 31 08:26 Dropbox
drwx-----+ 54 Noe  staff   1.8K May 30 16:01 Downloads
drwx-----+  8 Noe  staff   272B May 28 21:06 Pictures
drwxr-xr-x  18 Noe  staff   612B May 17 11:12 BTI
drwxr-xr-x   5 Noe  staff   170B May  8 11:44 programs
drwx-----+ 15 Noe  staff   510B Apr 10 08:33 Documents
drwxr-xr-x   6 Noe  staff   204B Mar 18 09:22 VirtualBox VMs
drwxr-xr-x   8 Noe  staff   272B Mar 14 19:26 py_devel
drwx-----@ 51 Noe  staff   1.7K Mar 11 15:08 Library
```

Home directory

```
/home/bioinfo
/home/noe
/home/noe/Desktop
```

```
noe@debian-virtualbox:~$ ls -l /
total 108
drwxr-xr-x  2 root root  4096 Sep 26  2012 bin
drwxr-xr-x  3 root root  4096 Nov  9  2012 boot
drwxr-xr-x 15 root root 3140 May 31 12:46 dev
drwxr-xr-x 130 root root 12288 May 31 12:45 etc
drwxr-xr-x  5 root root  4096 Feb 28 13:54 export
drwxr-xr-x  4 root root  4096 Nov  7  2012 home
lrwxrwxrwx  1 root root   30 Sep 26 2012 initrd.img
drwxr-xr-x 12 root root 12288 Nov  9  2012 lib
drwxr-xr-x  2 root root 12288 Nov  9  2012 lib32
lrwxrwxrwx  1 root root   4 Sep 26 2012 lib64 -> /lib
drwxr-xr-x  2 root root 16384 Sep 26 2012 lost+found
drwxr-xr-x  3 root root  4096 Sep 26 2012 media
drwxr-xr-x  2 root root  4096 May  1  2012 mnt
drwxr-xr-x  2 root root  4096 Sep 26 2012 opt
dr-xr-xr-x 134 root root    0 May 31 12:45 proc
drwxr-xr-x 10 root root  4096 Nov 15  2012 root
drwxr-xr-x  2 root root  4096 Nov  9  2012 sbin
drwxr-xr-x  2 root root  4096 Jul 21  2010 selinux
drwxr-xr-x  2 root root  4096 Sep 26 2012 srv
drwxr-xr-x 13 root root    0 May 31 12:45 sys
drwxrwxrwt 11 root root  4096 May 31 19:56 tmp
drwxr-xr-x 11 root root  4096 Sep 26 2012 usr
drwxr-xr-x 14 root root  4096 Sep 26 2012 var
```

Root directory

/bin, /lib, /usr	code and code libraries
/var	logs and other data
/home	user directories
/tmp	temporary files
/etc	configuration information
/proc	special file system in Linux



pwd, ls, and cd

- **pwd** — where am I?
- **ls** — what is inside the current directory?
- **cd** — change my current directory.

Exercise 2:

Change into the “/opt/” directory and list the files found there.



Wildcards, Shortcuts, and Command History

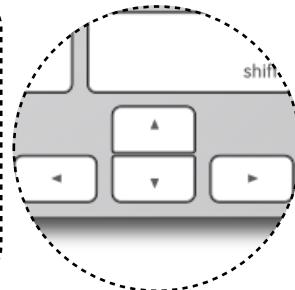
```
ls *txt
```

list all txt files in current directory

```
ls P*s
```

list files starting with P and ending with s,
e.g.: Pictures, Photos, Programs ...

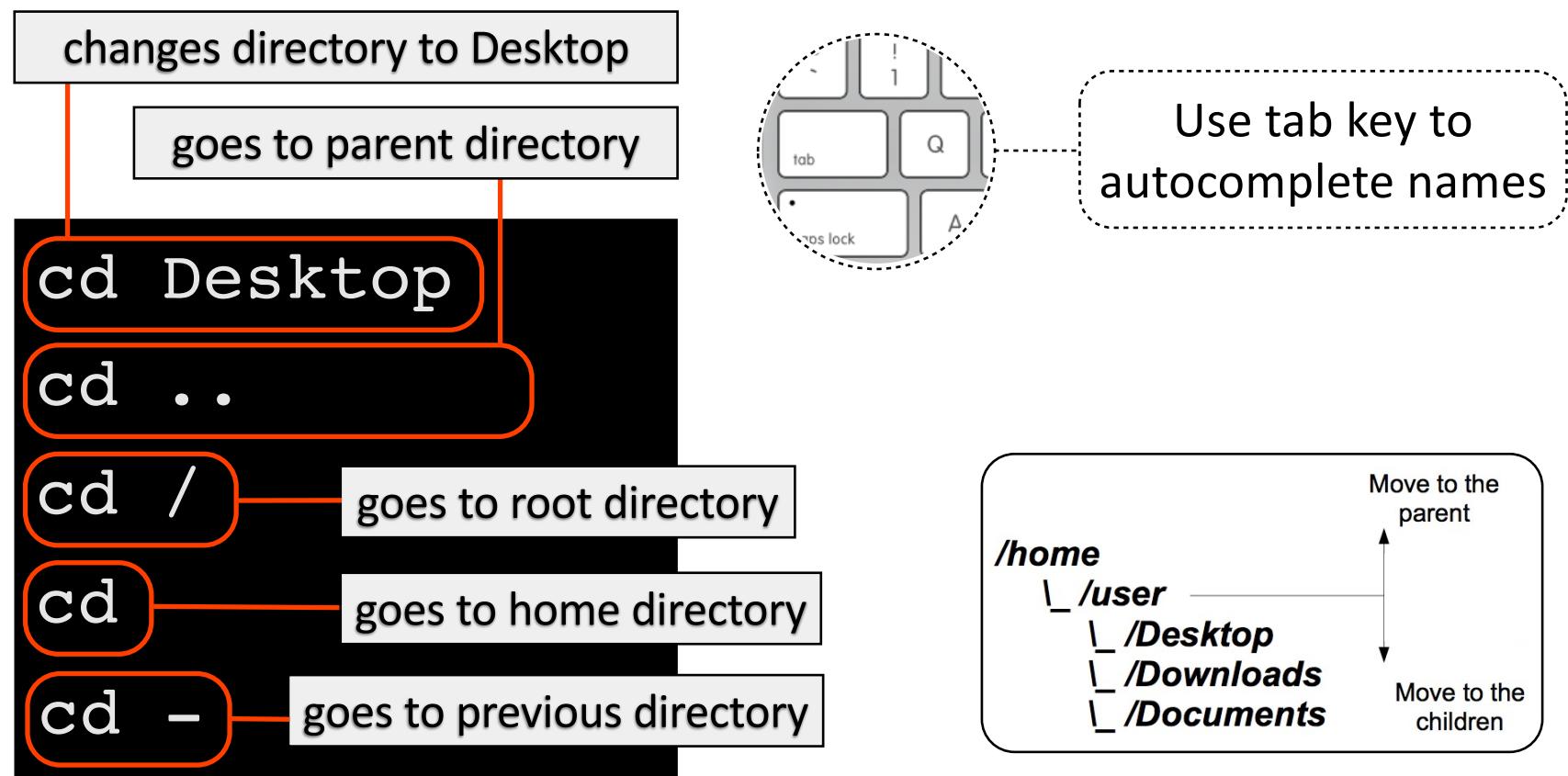
ctrl-c stop process
ctrl-a go to begin of line
ctrl-e go to end of line
ctrl-r search in command history



Use up and down arrows to navigate the command history



`cd` changes directory



Absolute and relative paths

list files in Desktop using an absolute path

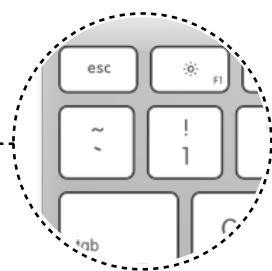
```
ls /home/user/Desktop
```

```
ls Desktop/
```

```
ls ~/Desktop
```

list files in Desktop using your home as a reference

list files in Desktop using a relative path (from your home: /home/bioinfo)



Absolute and relative paths

Absolute paths do not depend on where you are

```
ls /home/bioinfo/Desktop
```

```
ls ~/Desktop
```

~/ is equivalent to /home/bioinfo/



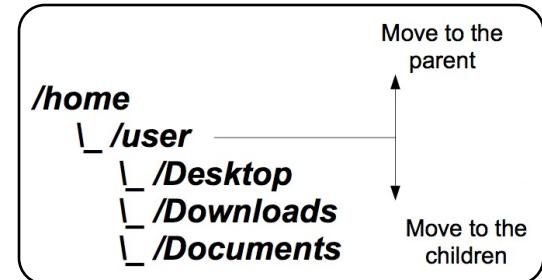
Absolute and relative paths

goes to *Desktop* from when you are in your home (/home/bioinfo)

```
cd Desktop/
```

```
ls .. /Documents
```

list files from *Documents* when you are in *Desktop*



Create, copy, move and delete files

creates an empty file called tmp_file.txt

copies tmp_file.txt in file_copy.txt

`touch tmp_file.txt`

`cp tmp_file.txt file_copy.txt`

`mv file1.txt file2.txt`

`rm file.txt`

Tip: name files in lower case and with underscores instead of using spaces



Create, copy and delete directories

creates an empty directory called *dir_name*

deletes *dir_name* directory if it is empty

```
mkdir dir_name
```

```
rmdir dir_name
```

```
rm -r dir_name
```

```
cp -r dir_name dir_copy
```

copy *dir_name* and its files in a new folder



Setting up our data files

cd

```
wget http://mirrors.kernel.org/ubuntu/pool/main/g/glibc/multiarch-support_2.27-3ubuntu1.4_amd64.deb  
http://ftp.se.debian.org/debian/pool/main/o/openssl/libssl1.0.0_1.0.1t-1+deb8u8_amd64.deb  
https://files.renci.org/pub/irods/releases/4.1.10/ubuntu14/irods-icommands-4.1.10-ubuntu14-x86_64.deb
```

```
sudo dpkg --install multiarch-support_2.27-3ubuntu1.4_amd64.deb  
libssl1.0.0_1.0.1t-1+deb8u8_amd64.deb irods-icommands-4.1.10-ubuntu14-x86_64.deb
```

```
sudo apt --fix-broken install  
sudo apt install nfs-common  
sudo mkdir /mnt  
cd /mnt  
sudo mkdir NMW2021Data  
sudo mount -t nfs storage.cyverse.org:/storage NMW2021Data/  
cp -r /mnt/NMW2021Data/Botany2020NMGWorkshop/ /scratch/
```



Networking Commands

● Networking commands

UNIX Command-Line Cheat Sheet	
BTI-SGN Bioinformatics Course 2014	
File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file1 file2	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows command's manual
top	shows process activity
df -h	shows disk space info
Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files
Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file
less -N file	includes line numbers
less -S file	wraps long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ","	concatenates the lines of input files by commas
sed	transforms text
Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux



Networking Commands

connects your terminal to your account in a server

Downloads the BCBC logo!

```
ssh user_name@server_address
```

```
wget https://btiscience.org/wp-content/uploads/BCBClogo.png
```

```
scp afp@boyce.sgn.cornell.edu:/home/afp/file.txt .
```

copy *file.txt* from your home in the server to the current directory in your computer

Tip: use the command `pwd` to get the path for `cp` and `scp`



Using terminal and ssh to access the Atmosphere instance

The screenshot shows the CYVERSE web interface for managing cloud instances. On the left, there's a sidebar with 'CYVERSE' logo, 'Dashboard', 'Projects', and 'Images' tabs. Below that, it says 'Resources > NMW2021'. In the center, there's a green circular icon labeled 'NMW2021'. To its right, under 'Allocation Source', it shows 'apowell'. Under 'Allocation Used', it says '0% of 336 AU's from apowell'. Below that is a section titled 'Instance Details' with the following information:

Status	● Active
Activity	N/A
Size	large3 (CPU: 8, Mem: 64 GB, Disk: 180 GB root)
IP Address	128.196.142.17 Copy
Launched	May 10, 2021 (35 minutes ago)
Based on	NMW2021 v1.0

On the right side of the interface, there's a terminal window showing a SSH session:

```
adrianpowell@Adrian-Powells-MacBook-Pro:~$ ssh apowell@128.196.142.17
The authenticity of host '128.196.142.17 (128.196.142.17)' can't be established.
ECDSA key fingerprint is SHA256:5mR5D+Z8M/DM7t79TYWLB4t9mf+zhnRvZIYgr1v3Wg.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '128.196.142.17' (ECDSA) to the list of known hosts.
apowell@128.196.142.17's password:
Welcome to Ubuntu 18.04.5 LTS (GNU/Linux 4.15.0-142-generic x86_64)

System information as of Mon May 10 07:58:43 MST 2021

System load: 0.0          Processes:           163
Usage of /: 77.0% of 18.21GB  Users logged in:      1
Memory usage: 0%           IP address for ens3:   172.19.94.11
Swap usage:  0%           IP address for docker0: 172.17.0.1

* Pure upstream Kubernetes 1.21, smallest, simplest cluster ops!
  http://microk8s.io/

Get cloud support with Ubuntu Advantage Cloud Guest:
  http://www.ubuntu.com/business/services/cloud

* Canonical Livepatch is available for installation.
  - Reduce system reboots and improve kernel security. Activate at:
    https://ubuntu.com/livepatch

2 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

New release '20.04.2 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Welcome to

[REDACTED]

Last login: Mon May 10 06:45:02 2021 from 128.196.65.30
apowell@vm142-17:~$
```





LEARNING CENTER

GETTING STARTED WEBINAR

CLOSE AND DON'T SHOW AGAIN

My Services



CoGe

Online system for quick and easy retrieval and comparison of genomic information and sequences.

[LAUNCH](#) 

Discovery Environment

Use hundreds of bioinformatics apps and manage data in the CyVerse Data Store from a simple web...

[LAUNCH](#) 

Atmosphere

Cloud computing platform for CyVerse

[LAUNCH](#) 

Data Commons

A unified system for managing and sharing your data across CyVerse's tools and services

[LAUNCH](#) 

CyVerse Discovery Environment

The screenshot displays the CyVerse Discovery Environment interface. On the left, a vertical sidebar contains icons for navigation: three horizontal lines (Menu), a house (Home), a grid (Community Data), a bar chart (Data Analysis), a person (User Profile), a gear (Settings), and a question mark (Help). The main content area shows a list of "Community Data" projects. Each project entry includes a checkbox, a folder icon, the project name, and a three-dot ellipsis icon. The projects listed are: Astrolabe, BioViz, Botany2020NMGWorkshop, Brassica_dwarf, CAP_TCN, CHEESEHEAD, EuPathDB-Data, G-OnRamp_hubs, GoreLab, KBase_staging, Legume_Federation, and NAM. At the top of the main area is a blue header bar with the "DISCOVERY ENVIRONMENT" logo, a search bar, a "Data" dropdown menu, and user profile icons for help, shopping cart, notifications, and account settings. A "Refresh" button is located in the top right corner of the main content area.

	Name	Actions
<input type="checkbox"/>	Astrolabe	⋮
<input type="checkbox"/>	BioViz	⋮
<input type="checkbox"/>	Botany2020NMGWorkshop	⋮
<input type="checkbox"/>	Brassica_dwarf	⋮
<input type="checkbox"/>	CAP_TCN	⋮
<input type="checkbox"/>	CHEESEHEAD	⋮
<input type="checkbox"/>	EuPathDB-Data	⋮
<input type="checkbox"/>	G-OnRamp_hubs	⋮
<input type="checkbox"/>	GoreLab	⋮
<input type="checkbox"/>	KBase_staging	⋮
<input type="checkbox"/>	Legume_Federation	⋮
<input type="checkbox"/>	NAM	⋮



CyVerse Discovery Environment

The screenshot shows the CyVerse Discovery Environment interface. On the left is a vertical sidebar with icons for navigation: three horizontal lines (Menu), Home, Community Data (selected), Databases (highlighted in blue), Grid, Bar chart, Line chart, Settings, and Help. The main area has a header with the 'DISCOVERY ENVIRONMENT' logo, a search bar, and buttons for 'Data', '?', '!', 'Bell', and a green circular profile icon with a white letter 'A'. Below the header, a sub-header says 'Community Data' with a dropdown arrow. To the right are buttons for 'Refresh', 'Details', 'Add To Bag', and three vertical dots. A 'Customize Columns' button is also present. The main content area displays a list of projects under 'Community Data': Astrolabe (checked), BioViz, Botany2020NMGWorkshop, Brassica_dwarf, CAP_TCN, CHEESEHEAD, EuPathDB-Data, G-OnRamp_hubs, GoreLab, KBase_staging, Legume_Federation, and NAM. Each project entry has a checkbox, a folder icon, and three vertical dots. A context menu is open over the 'Astrolabe' project, listing options: Details (with an info icon), Metadata (with a list icon), Copy Link (with a clipboard icon), Copy Path (with a folder icon, highlighted in grey), and Move to Trash (with a trash bin icon). At the bottom of the list are navigation buttons (<, 1, >) and a page number '100'.



iCommands

https://learning.cyverse.org/projects/data_store_guide/en/latest/step2.html

```
$ iinit
One or more fields in your iRODS environment file (.irodsEnv) are
missing; please enter them.
Enter the host name (DNS) of the server to connect to: data.cyverse.org
Enter the port number: 1247
Enter your irods user name: #your_cyverse_username
Enter your irods zone: iplant
Those values will be added to your environment file (for use by
other i-commands) if the login succeeds.

Enter your current iRODS password: #your_cyverse_password
```

```
cd /scratch
```

```
iget -rPT /iplant/home/shared/Botany2020NMGWorkshop
```



Commands to install software

```
aptitude search blast
```

```
sudo aptitude install blast2
```

```
sudo apt-get install pbzip2
```

ubuntu®

debian
GNU/Linux

call the command with super user permissions

installs *pbzip2* in your computer



Compression commands

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

BTI UNIX Command-Line Cheat Sheet
BTI-SGN Bioinformatics Course 2014



File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info

Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file
less -N file	include line numbers
less -S file	wraps long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ","	concatenates the lines of input files by commas
sed	transforms text

Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux



Compression commands

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

group and compress files

`tar -zcvf file.tar.gz f1 f2`

`tar -zxvf file.tar.gz`

decompress and ungroup a tar.gz file

BTI UNIX Command-Line Cheat Sheet
BTI-SGN Bioinformatics Course 2014

File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info
Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files
Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux

files, directories or wildcards



Installing MaSuRCA

```
cd /scratch  
git clone https://github.com/alekseyzimin/masurca.git  
cd masurca  
mv MaSuRCA-3.4.2.tar.gz /scratch  
cd ..  
tar -xvf MaSuRCA-3.4.2.tar.gz  
cd MaSuRCA-3.4.2  
../install.sh
```



Outline

- VMs & terminals
- File system navigation
- Creating, copying, moving, deleting files
- Networking commands
- Software installation & compression commands

