

# 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



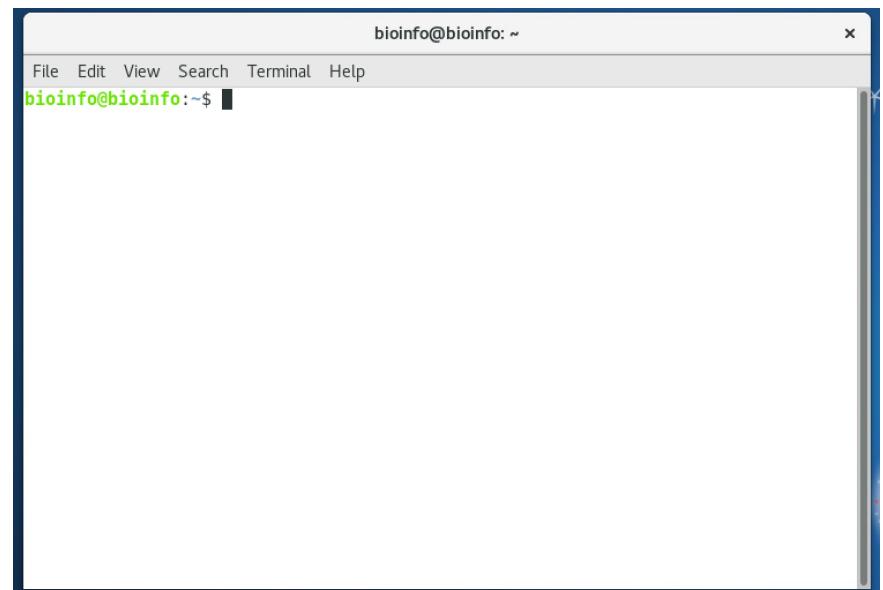
# What is a Virtual Machine?

- 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

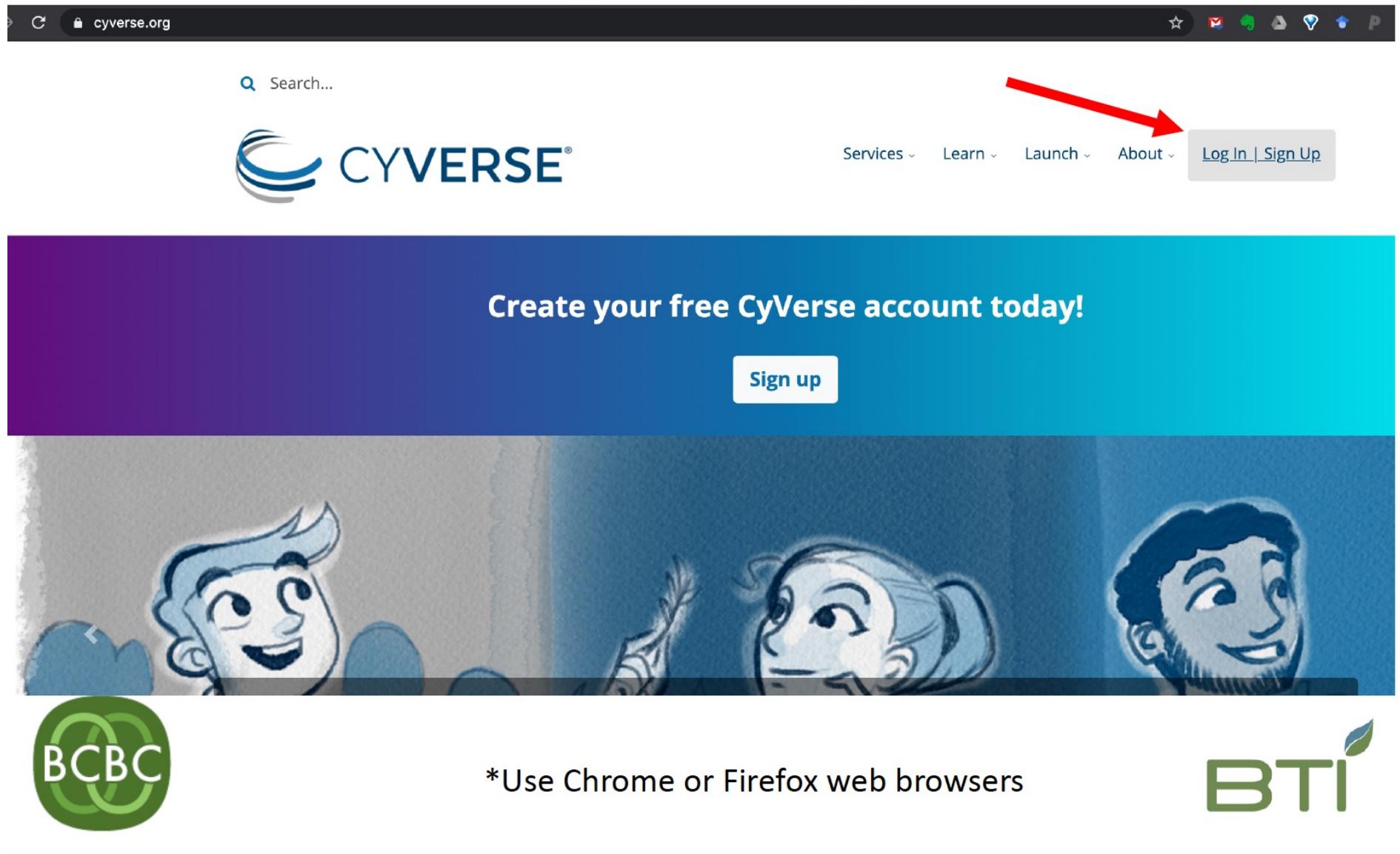
Vision: Transforming Science through Data-Driven Discovery

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 blue. 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 cartoon illustration of three people 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" containing three cards: "Launch New Instance" (with a rocket icon), "Browse Help Resources" (with a question mark icon), and "Change Your Settings" (with a gear icon). Further down, there are sections for "Resources Used" (with a "NEED MORE ?" button) and "Allocation Source" (a chart showing allocation usage for "apowell"). On the right, there is a summary of "2 Instances" with a donut chart indicating active vs. networking status.

**Dashboard**

**Getting Started**

**Launch New Instance**

Browse Atmosphere's list of available images and select one to launch a new instance.

**Browse Help Resources**

View a video tutorial, read the how-to guides, or email the Atmosphere support team.

**Change Your Settings**

Modify your account settings, view your resource quota, or request more resources.

**Resources Used** NEED MORE ?

**Allocation Source**

apowell

Allocation 0% 10 20 30 40 50 60 70 80 90 100

Percent of Allocation Used

**2 Instances**

active – networking active

Provider Resources



[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

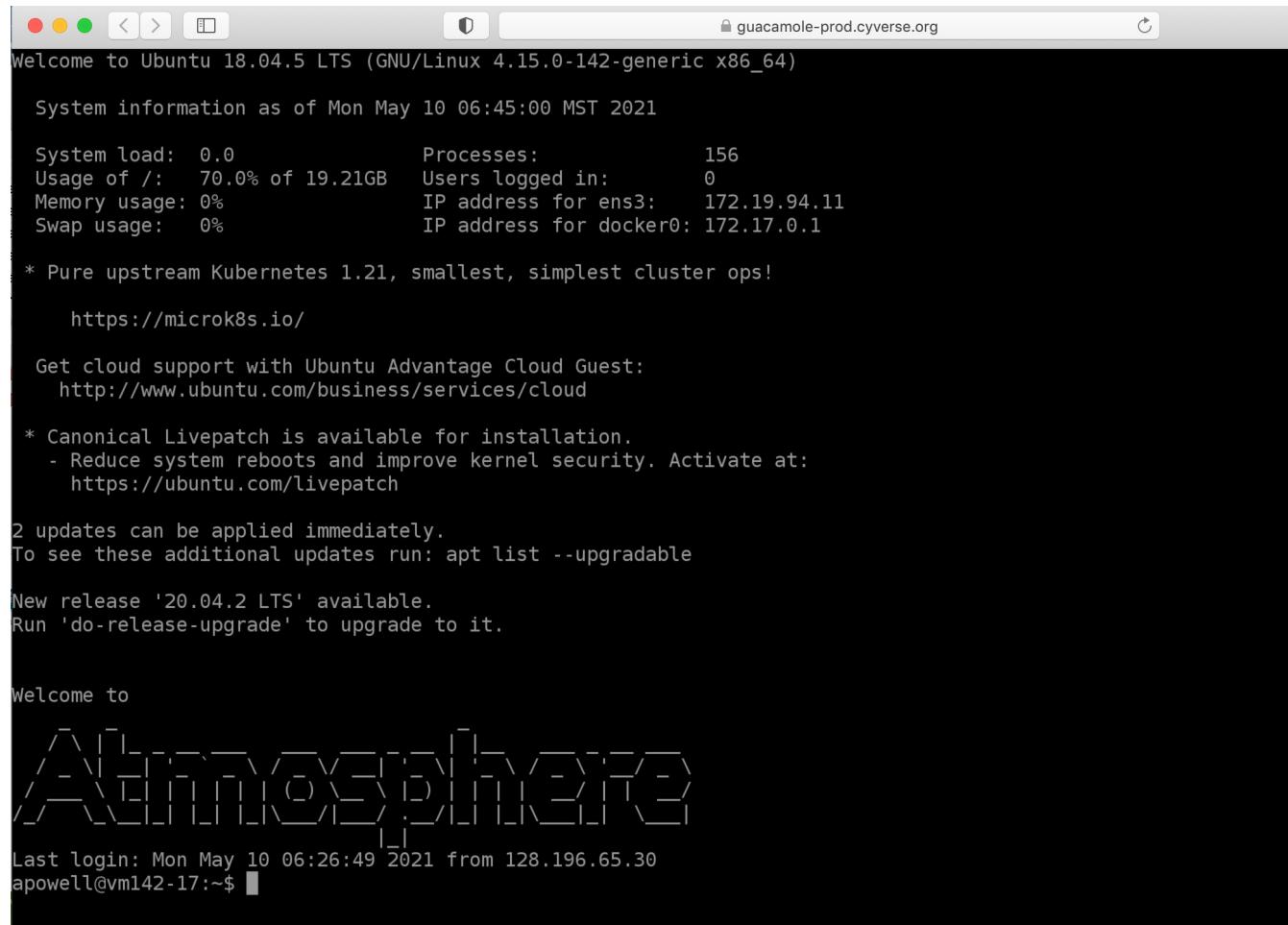
Open Web Shell ↗



**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 links to Kubernetes and Canonical Livepatch. Following this, it shows two available updates and provides instructions for upgrading. The terminal concludes with a welcome message from the user "apowell" and their last login details.

```
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

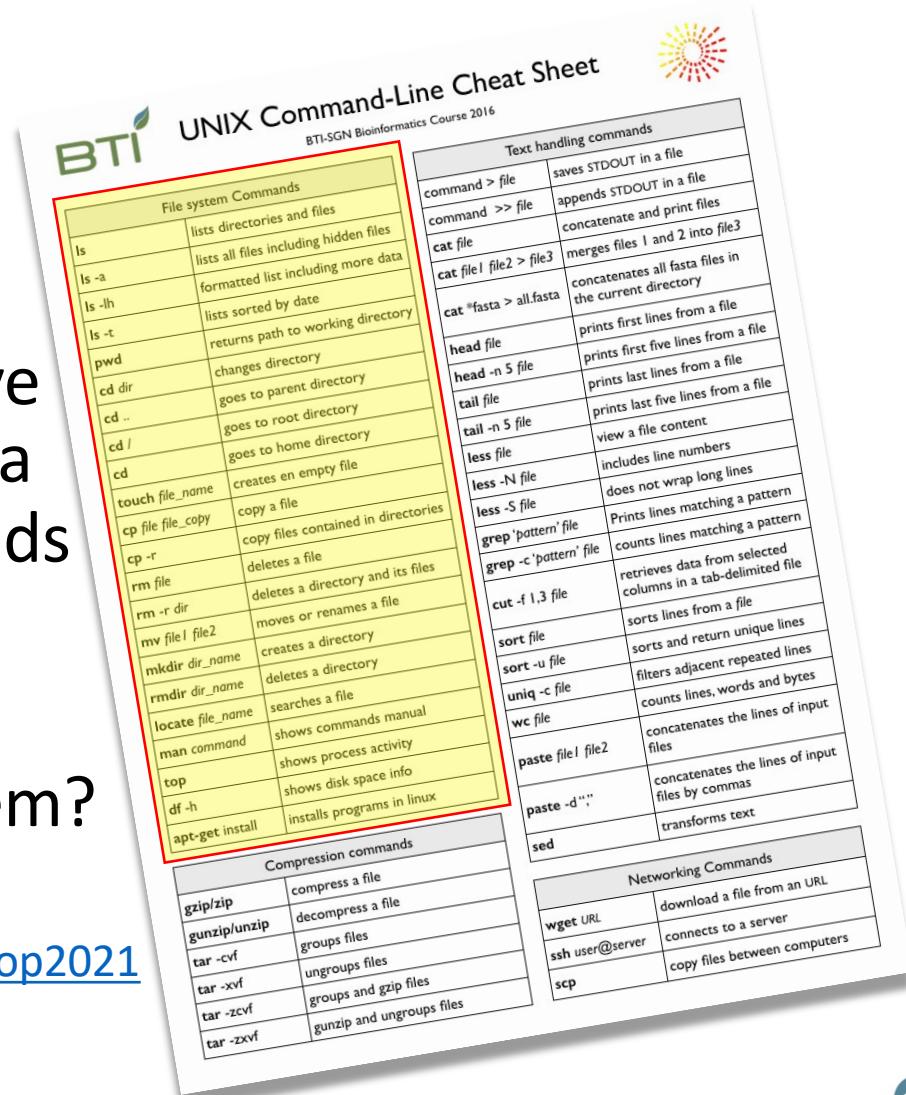
Last login: Mon May 10 06:26:49 2021 from 128.196.65.30
apowell@vm142-17:~$ █
```



# 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, rm, mv, mkdir, rmdir, locate, man, top, df, and apt-get. Another red box highlights the 'Text handling commands' section, which includes commands like cat, head, tail, less, grep, cut, sort, uniq, wc, paste, and sed. A third red box highlights the 'Compression commands' section, which includes gzip, gunzip, tar, and various options for compression and decompression. The 'Networking Commands' section at the bottom right includes wget, ssh, and scp. The BTI logo is in the top left corner, and a sun icon is 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

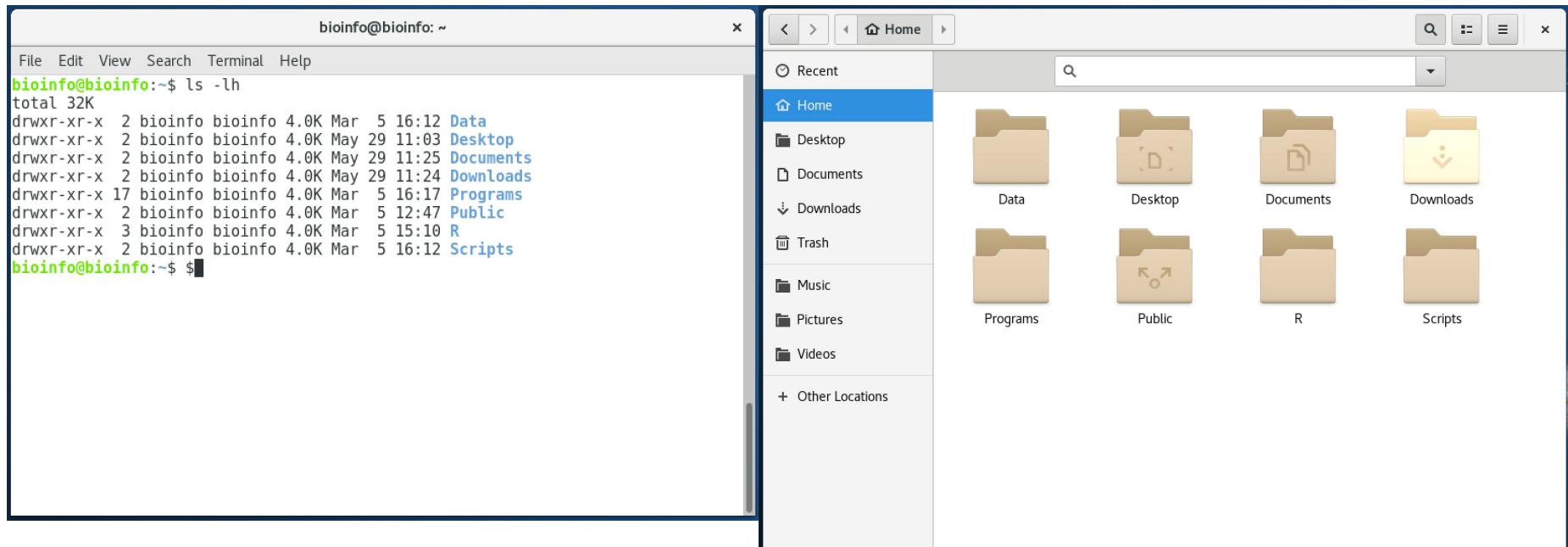
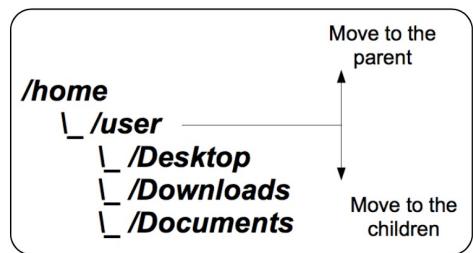
Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines 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

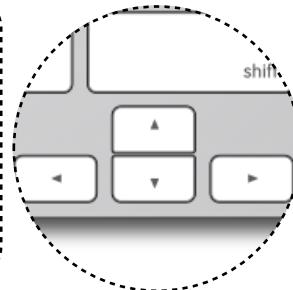
```
ls *txt
```

list all txt files in current directory

```
ls P*s
```

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

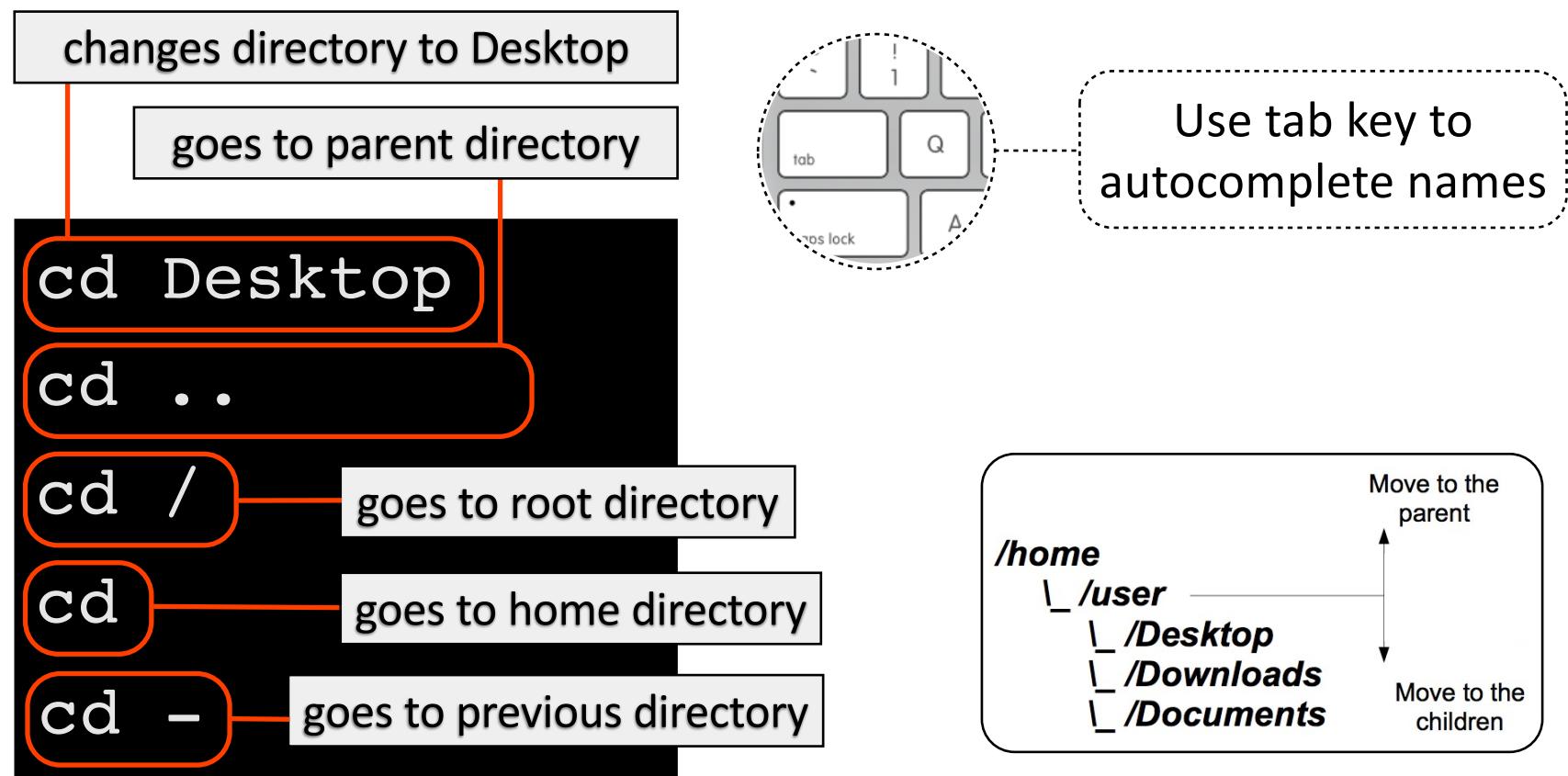
ctrl-c stop process  
ctrl-a go to begin of line  
ctrl-e go to end of line  
ctrl-r search in command history



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

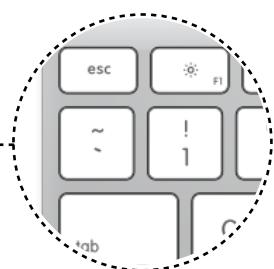
```
ls /home/user/Desktop
```

```
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/



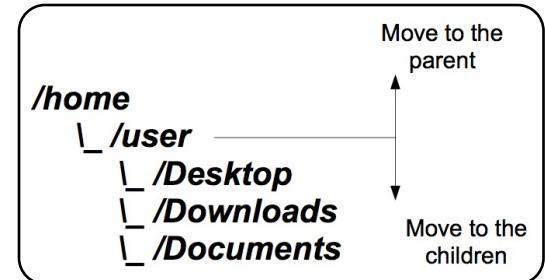
# Absolute and relative paths

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

```
sudo apt install nfs-common  
sudo mkdir /mnt  
sudo mkdir /mnt/NMW2021Data  
sudo mount -t nfs storage.cyverse.org:/storage /mnt/NMW2021Data/
```

```
sudo cp -r /mnt/NMW2021Data/Botany2020NMGWorkshop/ /scratch/  
sudo chown -R usrid /scratch/Botany2020NMGWorkshop
```



# Networking Commands

## ● Networking commands

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

CYVERSE Dashboard Projects Images

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: 180 GB root)
IP Address	128.196.142.17 <a href="#">Copy</a>
Launched	May 10, 2021 (35 minutes ago)
Based on	NMW2021 v1.0

```
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 180.21GB   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

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. The list includes:

- Astrolabe
- BioViz
- Botany2020NMGWorkshop
- Brassica\_dwarf
- CAP\_TCN
- CHEESEHEAD
- EuPathDB-Data
- G-OnRamp\_hubs
- GoreLab
- KBase\_staging
- Legume\_Federation
- NAM

Each item in the list has a checkbox icon to its left and a three-dot ellipsis icon to its right. A "Customize Columns" button is located at the top right of the list area. At the bottom of the list, there are navigation controls: a left arrow, a page number (1), a right arrow, and a dropdown menu set to "100". The top navigation bar features the "DISCOVERY ENVIRONMENT" logo, a search bar, and user account information.



# 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, grid, chart, video, settings, and help. The main area has a blue header with the 'DISCOVERY ENVIRONMENT' logo, a search bar, and a 'Data' dropdown. Below the header is a toolbar with 'Refresh', 'Details', 'Add To Bag', and a three-dot menu. A 'Community Data' dropdown is open, showing a list of projects. The list includes: Astrolabe (selected), BioViz, Botany2020NMGWorkshop, Brassica\_dwarf, CAP\_TCN, CHEESEHEAD, EuPathDB-Data, G-OnRamp\_hubs, GoreLab, KBase\_staging, Legume\_Federation, and NAM. To the right of the list is a context menu with options: Details, Metadata, Copy Link (highlighted), Copy Path, and Move to Trash. At the bottom of the main area are navigation arrows and a page number indicator (1/100).

Community Data	
	Name
<input checked="" 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



# iCommands

[https://learning.cyverse.org/projects/data\\_store\\_guide/en/latest/step2.html](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
```



# Compression commands

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

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



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

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

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



# Compression commands

Compression commands	
<b>gzip/zip</b>	compress a file
<b>gunzip/unzip</b>	decompress a file
<b>tar -cvf</b>	groups files
<b>tar -xvf</b>	ungroups files
<b>tar -zcvf</b>	groups and gzip files
<b>tar -zxvf</b>	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	
<b>ls</b>	lists directories and files
<b>ls -a</b>	lists all files including hidden files
<b>ls -lh</b>	formatted list including more data
<b>ls -t</b>	lists sorted by date
<b>pwd</b>	returns path to working directory
<b>cd dir</b>	changes directory
<b>cd ..</b>	goes to parent directory
<b>cd /</b>	goes to root directory
<b>cd</b>	goes to home directory
<b>touch file_name</b>	creates an empty file
<b>cp file file_copy</b>	copy a file
<b>cp -r</b>	copy files contained in directories
<b>rm file</b>	deletes a file
<b>rm -r dir</b>	deletes a directory and its files
<b>mv file1 file2</b>	moves or renames a file
<b>mkdir dir_name</b>	creates a directory
<b>rmdir dir_name</b>	deletes a directory
<b>locate file_name</b>	searches a file
<b>man command</b>	shows commands manual
<b>top</b>	shows process activity
<b>df -h</b>	shows disk space info

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

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

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

files, directories or wildcards



## 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



# 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
```



# Containers

## Developing with Docker

Developing apps today requires so much more than writing code. Multiple languages, frameworks, architectures, and discontinuous interfaces between tools for each lifecycle stage creates enormous complexity. Docker simplifies and accelerates your workflow, while giving developers the freedom to innovate with their choice of tools, application stacks, and deployment environments for each project.



# Changing docker file storage location

```
#stop docker
```

```
$ sudo service docker stop
```

```
#edit daemon.json (only if necessary)
```

```
$ sudo emacs /etc/docker/daemon.json
```

```
#and add (may already be done):
```

```
{
```

```
  "graph": "/scratch/docker"
```

```
}
```

```
#copy current dir to new one
```

```
$ sudo rsync -aP /var/lib/docker/ /scratch/docker
```

```
#rename old docker dir, do no delete until you test config works
```

```
$ sudo mv /var/lib/docker /var/lib/docker.old
```

```
$ sudo service docker start
```



# Install some programs with docker

```
$ docker pull maplesond/portcullis:stable #this is in 3_rnaseq_cleanup.sh  
$ docker pull cyverseuk/mikado           #this is in 3_rnaseq_cleanup.sh  
$ docker pull ezlabgva/busco:v5.1.2_cv1
```



# Outline

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

