

# **Command Line and Accessing Cypress**

**A workshop at Tulane  
University**

May 2024

# Goals of this workshop

- Promote open reproducible science
- Basic working understanding what command line is
- Demonstrate using remote hpc resources - in advance of more detailed Cypress and LONI workshops

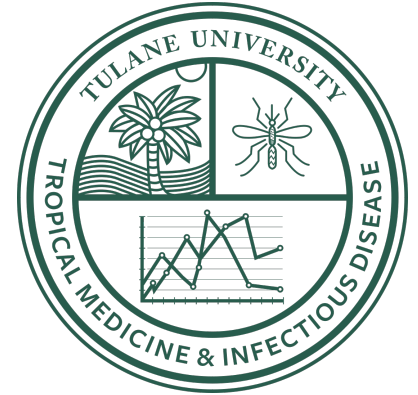
# Limitations of this workshop

- Time - 2h!
  - We are aiming to get folks past the potential hurdle of simply not having tried these things yet.
  - We are trying this out for the first time here, so if it is something you want more or less of, give that to us in the feedback!
- Diverse computer setups -> Diverse instructions -> Diverse errors

# Instructors and sources

Hosted by dept. of **Tropical Medicine and Infectious Disease**

With compute support from **the Cypress Team**



- Jessica Blanton
- Paula Castillo
- Arley Calle Tobon
- Samuel Jameson
- Bilen Mekonnen

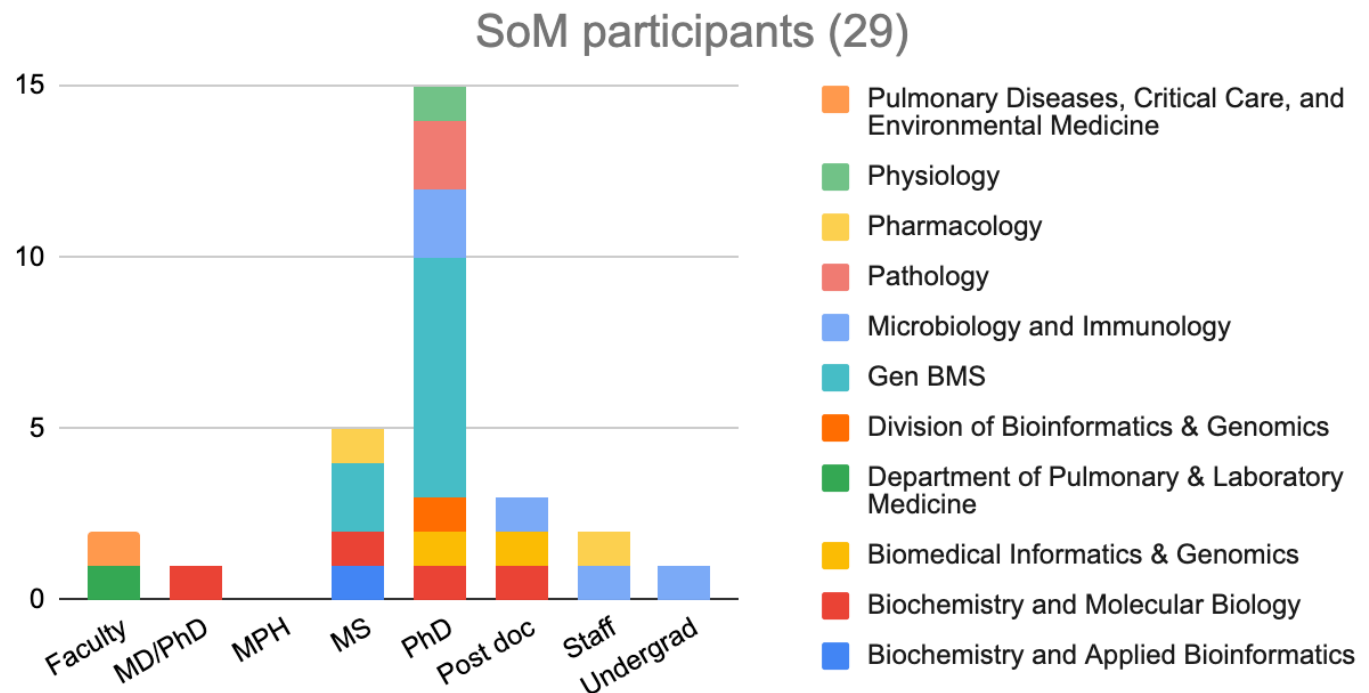
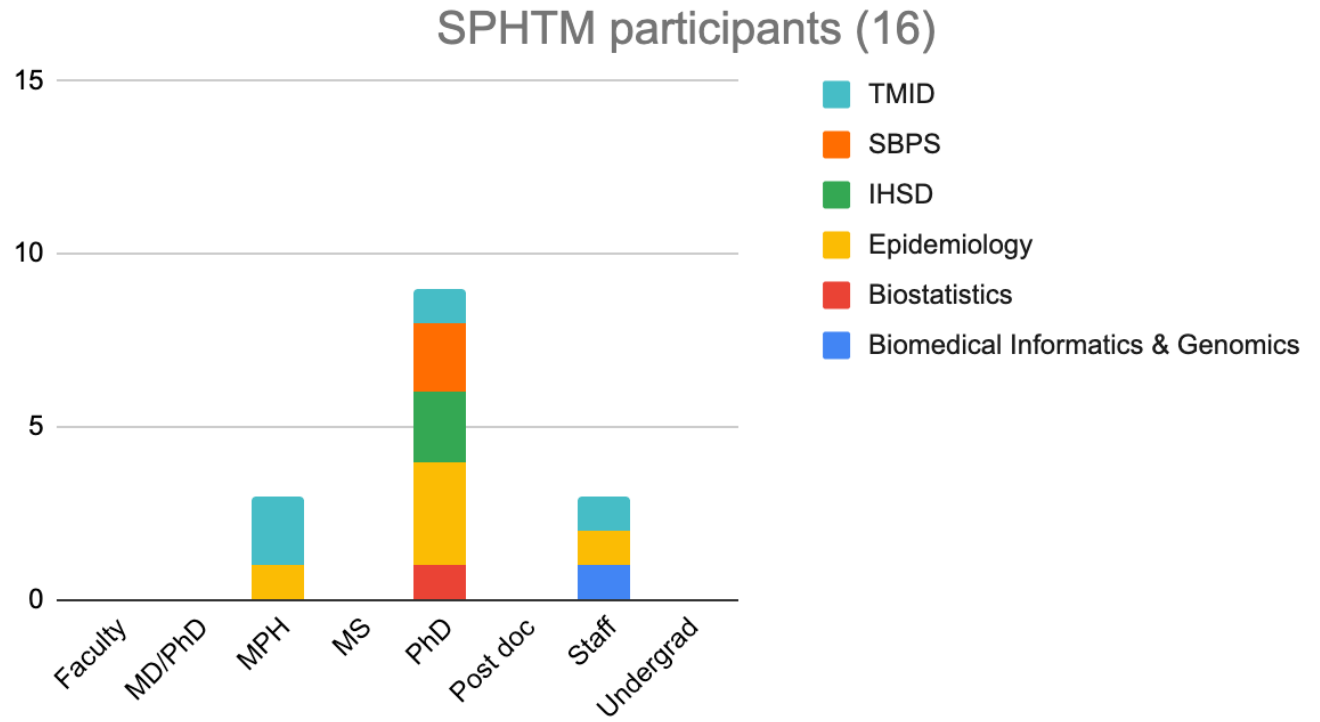
Elements of workshop materials may draw from lessons by

<https://wiki.hpc.tulane.edu/trac/wiki/cypress>

<https://software-carpentry.org/>

<https://www.hpc.lsu.edu/>

# Who are we in this room?

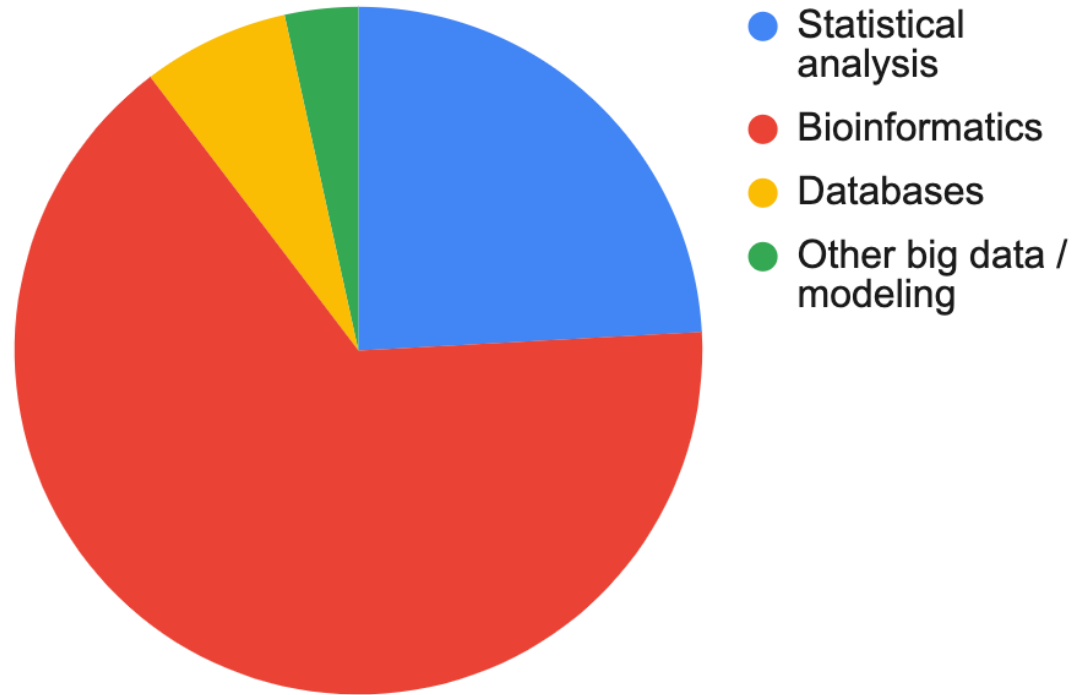


# What do **we** use scientific computing for?

SPHTM (20)

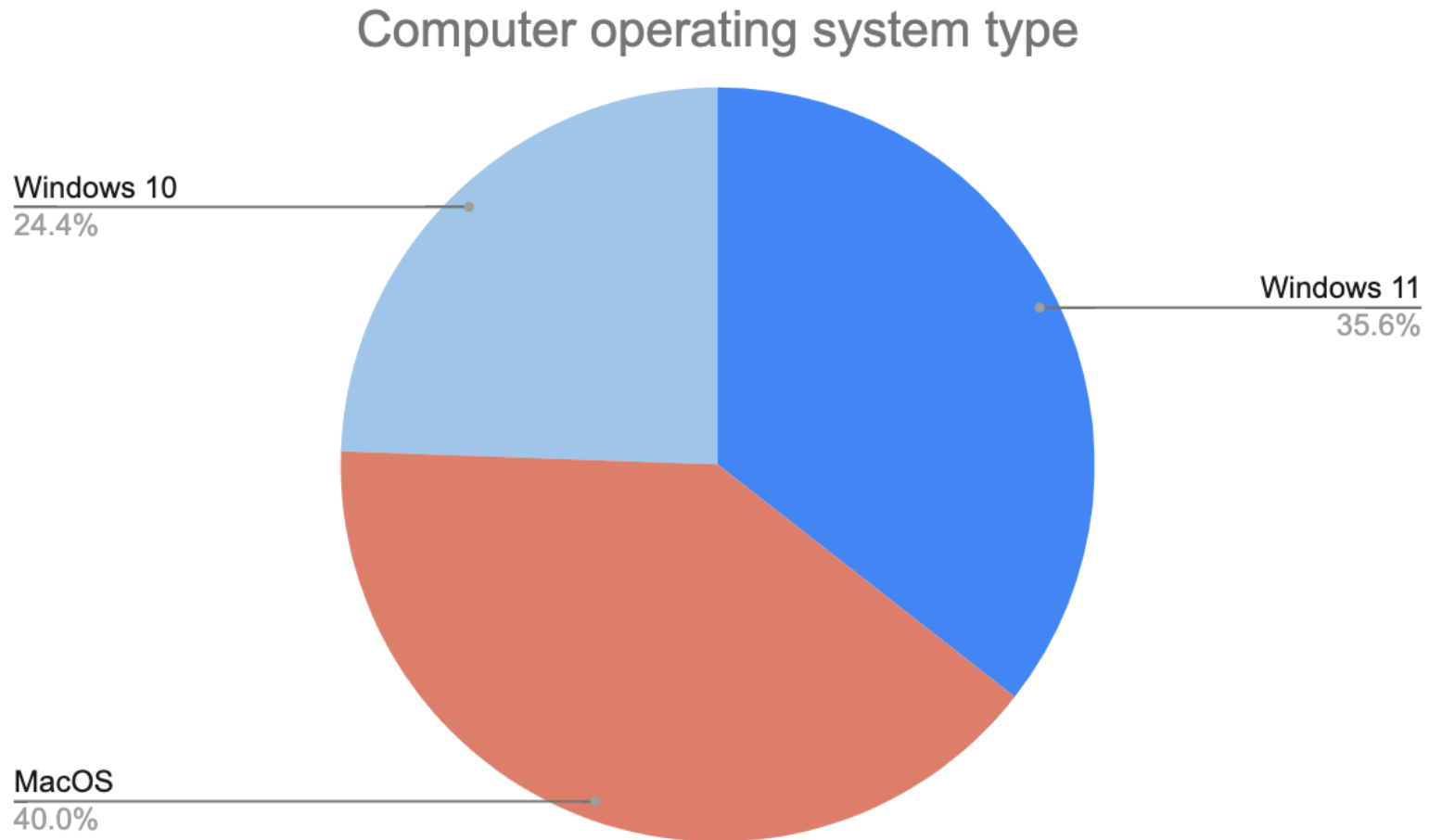


School of Medicine (29)



- Statistical analysis
- Bioinformatics
- Databases
- Other big data / modeling

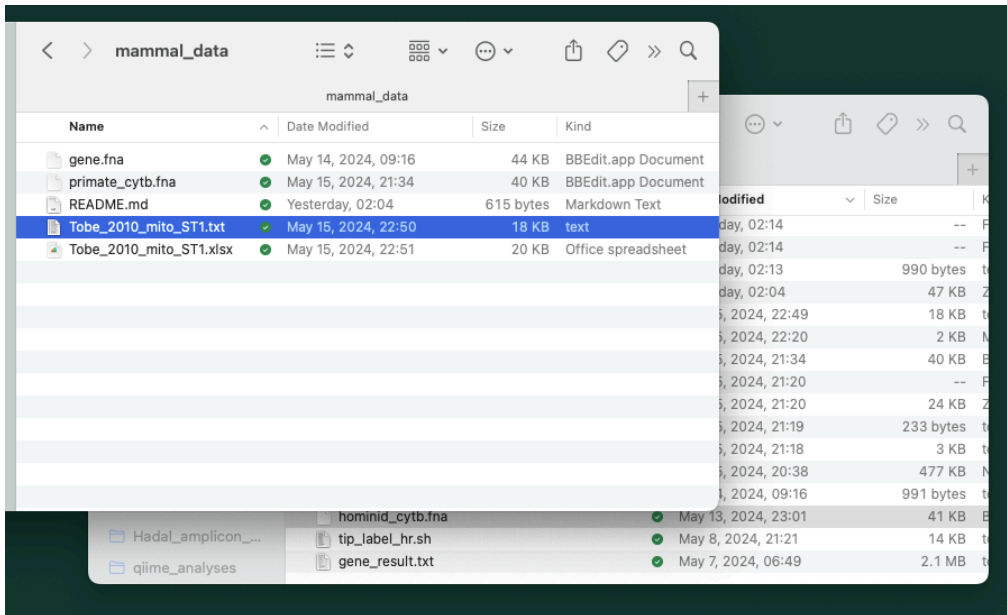
# What computers are **we** working from?



# How users communicate with the system (kernel)

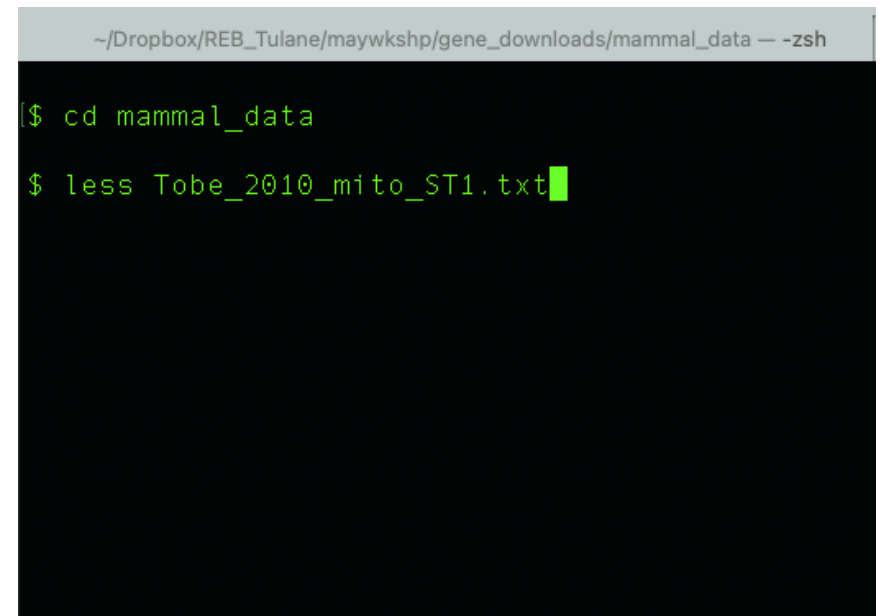
## Graphical User Interface (GUI)

Interaction using graphical representation-point and click



## Command Line Interface (CLI)

Interaction by typing commands





# Note taking

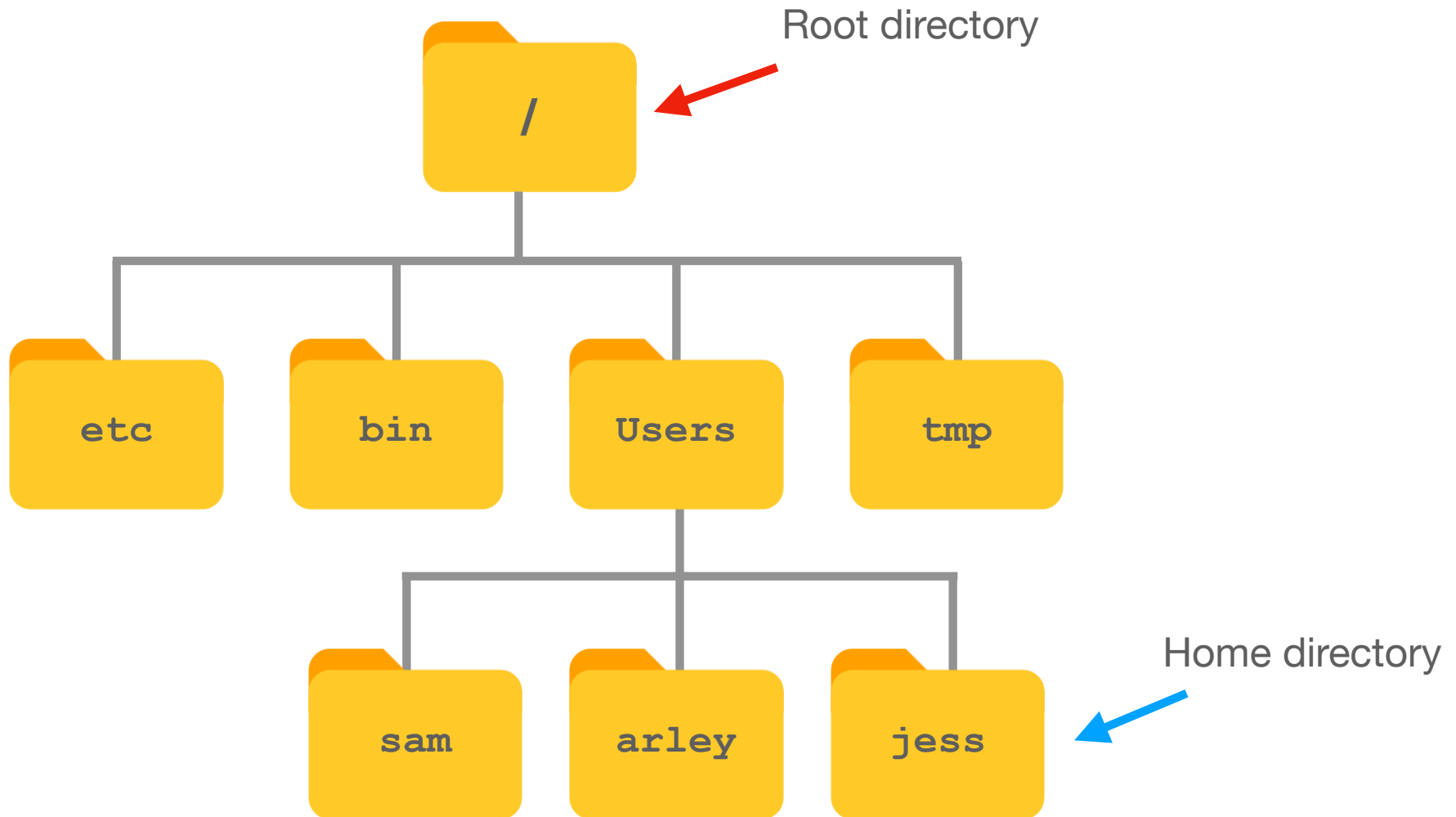
- Use a plain text editor such as TextEdit or Notepad, not Microsoft word
- Record both your notes & your commands!

There are many other nice programs to record notes and scripts in a simple commented fashion, some of which also support markdown.

A few easily come to mind:

- BBedit
- Sublime Text
- Atom (no longer updated but still great)

# Directory structure



# Paths

**Absolute** - The complete path to a directory/file on a given computer, starting from the home directory

```
/home/Users/jess/somefile.txt
```

**Relative** - The path that is relative to the current working directory

```
jess/somefile.txt
```

# Cypress supercomputer

A high performance computing (hpc) resource hosted at Tulane

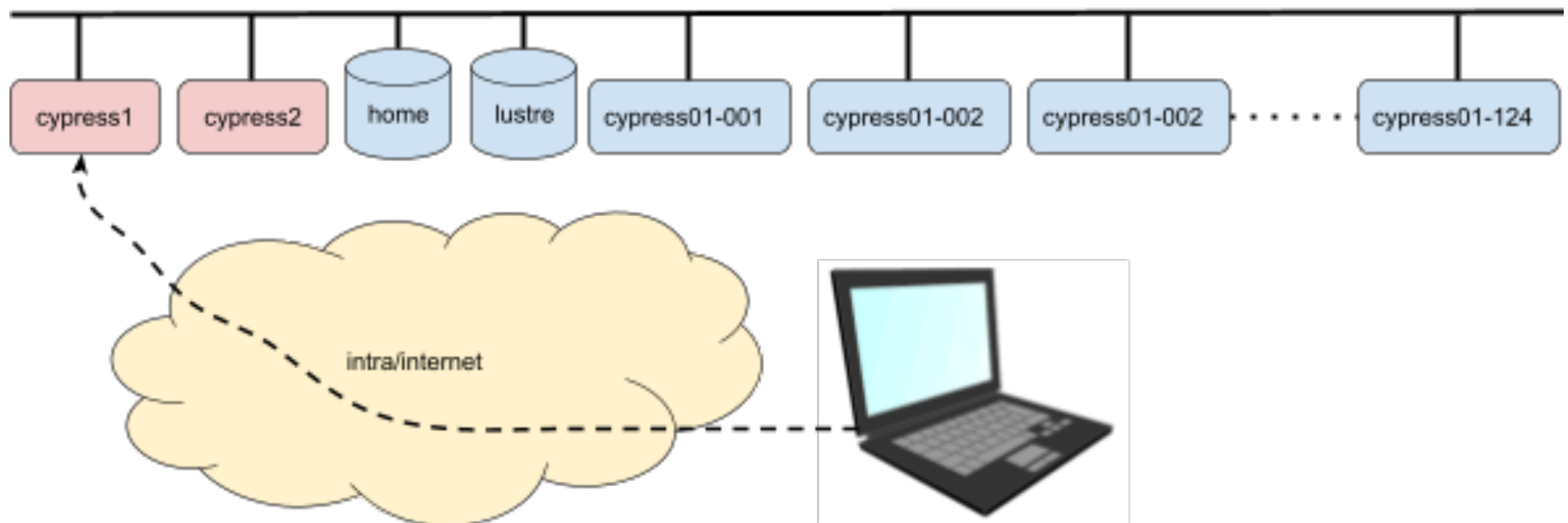


image source: Tulane hpc workshop materials

# Cypress supercomputer - A compute cluster

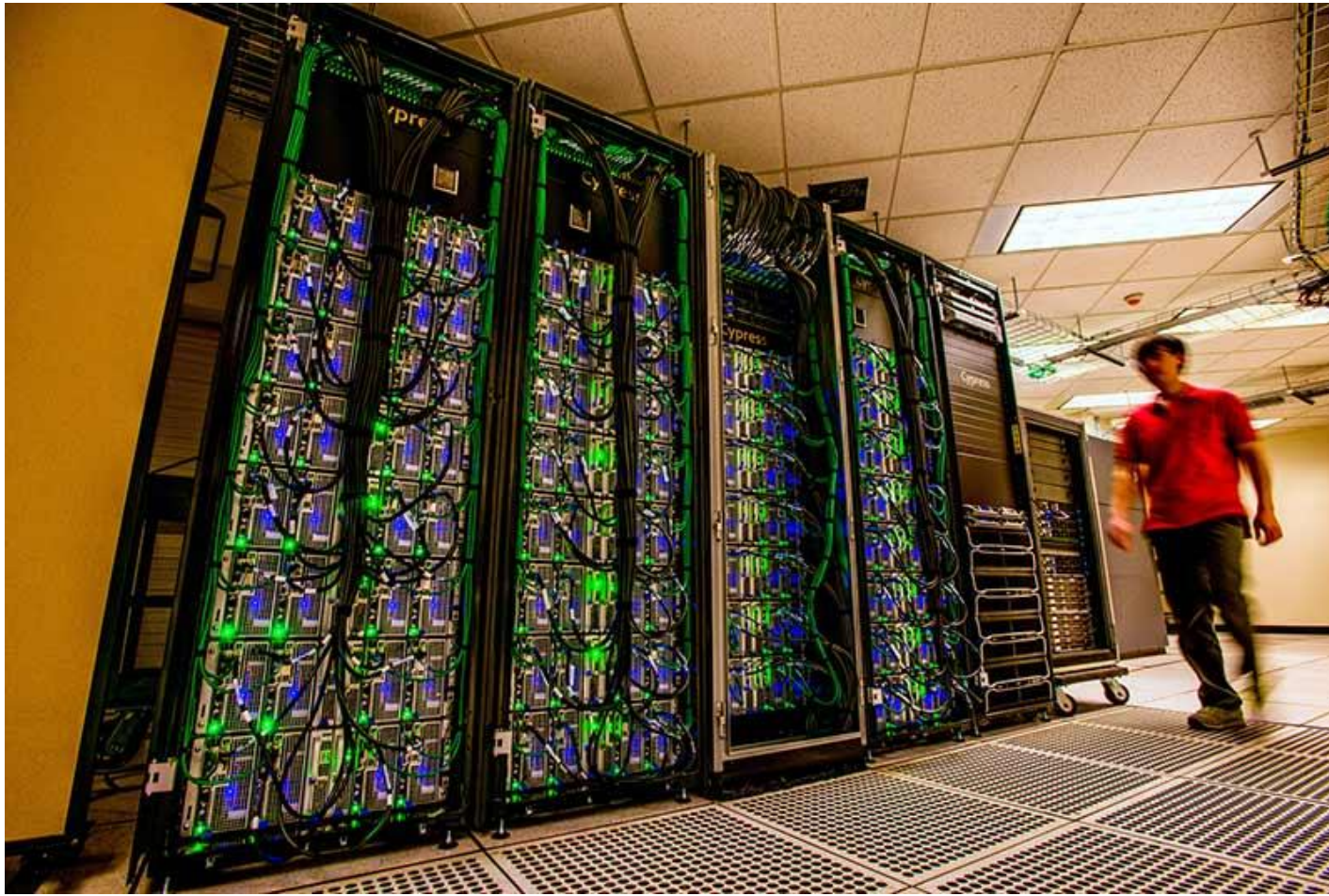


image source: Ryan Rivet

# Advantages of adding Cypress to your toolchest

- Speed & space for your computations
- Programs already installed or installable
- Expert and responsive SysAdmins!



# Cypress is accessed remotely

Located somewhere in downtown New Orleans...

But all we really need to know is that the cluster is on planet earth!



---

Not just connecting over wifi... we use "secure protocols":

Secure Shell (**SSH**)

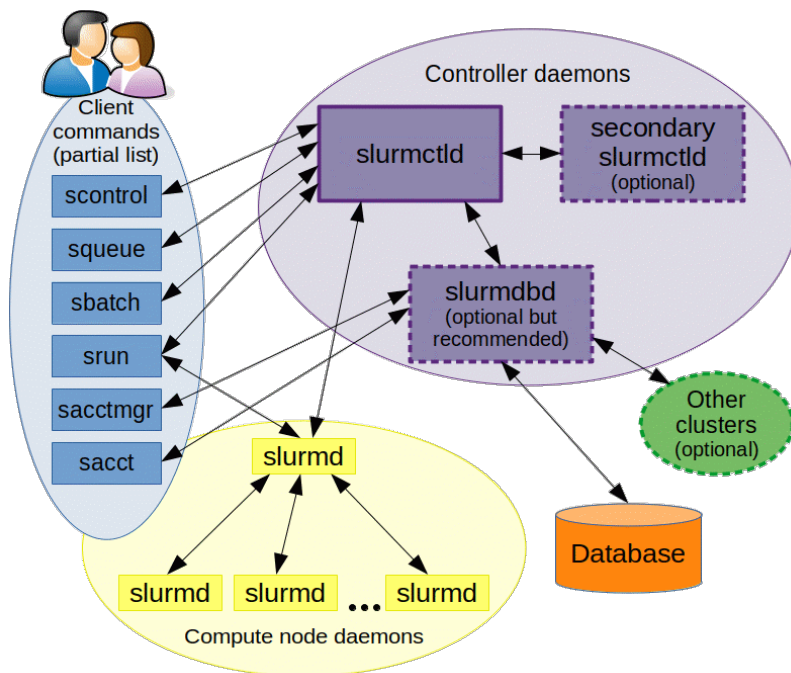
```
ssh location@remote.server
```

Secure Copy Protocol (**SCP**)

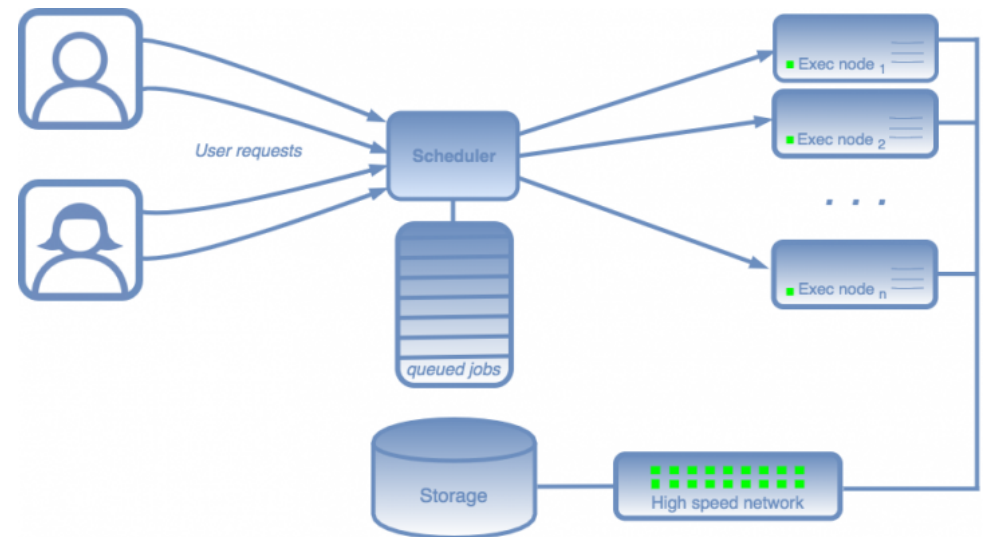
```
scp location@remote.server:filename location/local.server
```

# Slurm workload and scheduling manager

Cypress is a shared server. Slurm is used to coordinate all user requests!



<https://slurm.schedmd.com/>



<https://rdlab.cs.upc.edu/>



# Anatomy of a Slurm Batch Submission Script

```
#!/bin/bash
#SBATCH --job-name=p_tree      ### Job Name
#SBATCH --output=p_tree.out    ### File in which to store job output
#SBATCH --error=p_tree.err     ### File in which to store job error messages
#SBATCH --partition=workshop   ### partition for Cypress- default "defq"
#SBATCH --qos=workshop        ### Quality of service parameter- default "normal"
#SBATCH --time=0-00:10:00     ### Wall clock time limit in Days-HH:MM:SS
#SBATCH --nodes=1             ### Number of nodes to use
#SBATCH --ntasks-per-node=1   ### Number of tasks to run per node
#SBATCH --cpus-per-task=20    ### Number of cpus available for task

# Make sure muscle 5.1.linux64 is available
module load muscle/5.1

# Make sure FastTree version 2.1.10 is available
module load qiime2/2018.2

# Set up directory for output files
mkdir primate_tree

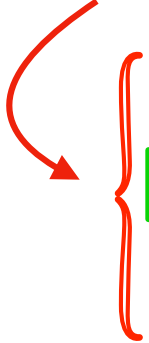
# Align sequences
muscle -align mammal_data/primate_cytb.fna -output primate_tree/aln.fna \
-threads 1 -log primate_tree/muscle_aln.log

# Build tree
FastTree -log primate_tree/fasttree.log -nt primate_tree/aln.fna > primate_tree/p_tree.nwk
```

# Anatomy of a Slurm Batch Submission Script

Directives for Slurm:  
management and job  
scheduling

Partition and resource  
limits, here specific to  
this workshop



```
#!/bin/bash
#SBATCH --job-name=p_tree      ### Job Name
#SBATCH --output=p_tree.out    ### File in which to store job output
#SBATCH --error=p_tree.err     ### File in which to store job error messages
#SBATCH --partition=workshop   ### partition for Cypress- default "defq"
#SBATCH --qos=workshop        ### Quality of service parameter- default "normal"
#SBATCH --time=0-00:10:00     ### Wall clock time limit in Days-HH:MM:SS
#SBATCH --nodes=1             ### Number of nodes to use
#SBATCH --ntasks-per-node=1    ### Number of tasks to run per node
#SBATCH --cpus-per-task=20     ### Number of cpus available for task
```

```
# Make sure muscle 5.1.linux64 is available
module load muscle/5.1

# Make sure FastTree version 2.1.10 is available
module load qiime2/2018.2
```

Make programs already  
installed on Cypress  
available

```
# Set up directory for output files
mkdir primate_tree
```

```
# Align sequences
muscle -align mammal_data/primate_cytb.fna -output primate_tree/aln.fna \
-threads 1 -log primate_tree/muscle_aln.log

# Build tree
FastTree -log primate_tree/fasttree.log -nt primate_tree/aln.fna > primate_tree/p_tree.nwk
```

Run main programs

# Other compute resources available to Tulanians

## The LA Optical Network (LONI)



Three clusters- QB2, QB3, QB4

More info:

- LSU <https://www.hpc.lsu.edu/docs/guides.php>
- Cypress workshop guides