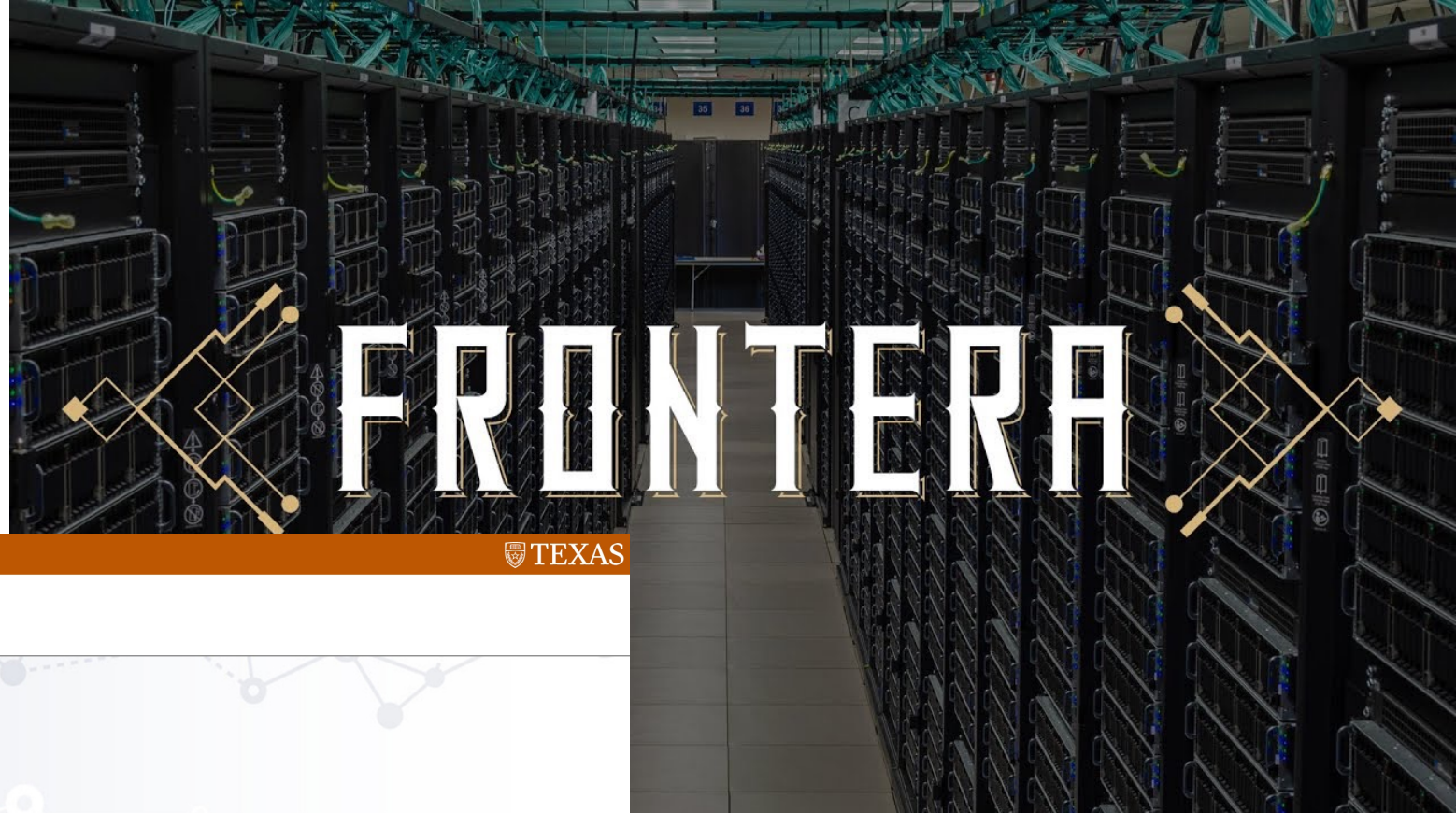




Working with the cluster on TACC

What is TACC?



College of Natural Sciences

COVID-19 Modeling Consortium

DASHBOARDS ▾ PUBLICATIONS PEOPLE COLLOQUIA

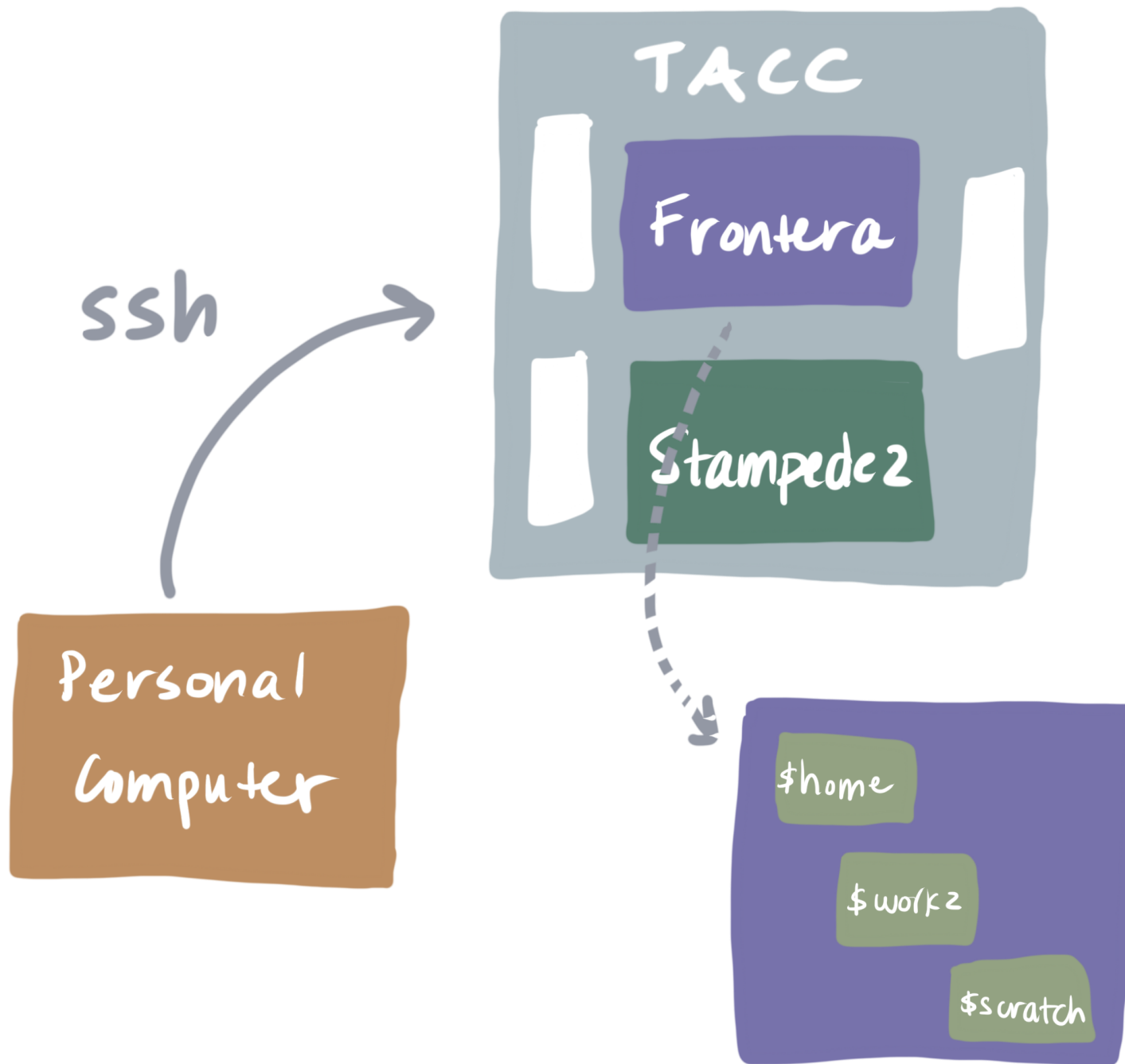
UT Austin COVID-19 Modeling Consortium

An interdisciplinary network of researchers and health professionals building models to detect, project, and combat COVID-19

<https://covid-19.tacc.utexas.edu/>

[2EnEN4EBQTs&ab_channel=TACCutexas](#)

Frontera is the largest and most powerful academic supercomputer in the world

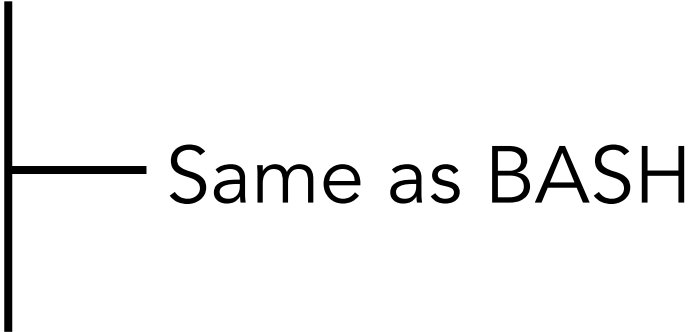



How do we work with TACC?

- **Access:**
 - Multi-factor authentication (TACC token code)
 - Secure shell on your computer (ssh)
- **Allocations/projects**
- **Computers**
 - Stampede2, Frontera, etc.
- **Directory structure**
- **Submitting jobs**

This is translatable to other supercomputer/cluster computing systems at other institutions (which is why I'm teaching it)

Useful starting commands for TACC

- pwd
 - cd [directory] (cdw, cd dir, cd ..)
 - ls or ll
 - cat
- 
- Same as BASH

- ssh
 - scp
 - sbatch
 - squeue
 - scancel
 - ctrl+c
- 
- TACC-specific commands

TACC: access

```
[annes-mbp:~ eac$ ssh eac3496@frontera.tacc.utexas.edu
```

secure shell

your TACC username
(I use my EID)

computer

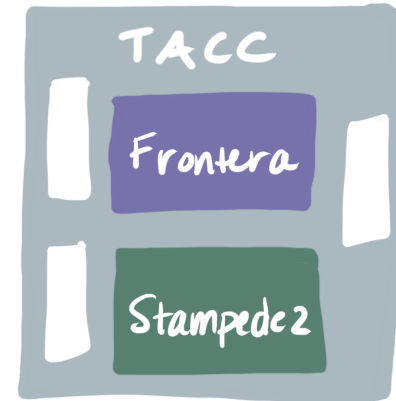
standard for TACC

```
[annes-mbp:~ eac$ ssh eac3496@stampede2.tacc.utexas.edu
```

LIVE CODE

ssh

Personal
Computer



```
Annex-MacBook-Pro:Day1_advanced_files eac$ ssh eac3496@frontera.tacc.utexas.edu
To access the system:
```

- 1) If not using ssh-keys, please enter your TACC password at the password prompt
- 2) At the TACC Token prompt, enter your 6-digit code followed by <return>.

```
Password:
TACC Token Code:
Last login: Sat Oct 23 11:47:06 2021 from 128.62.58.6
```

```
-----
                Welcome to the Frontera Supercomputer
        Texas Advanced Computing Center, The University of Texas at Austin
-----
```

```
** Unauthorized use/access is prohibited. **
```

If you log on to this computer system, you acknowledge your awareness of and concurrence with the UT Austin Acceptable Use Policy. The University will prosecute violators to the full extent of the law.

```
TACC Usage Policies:
http://www.tacc.utexas.edu/user-services/usage-policies/
```

```
-----
Welcome to Frontera, *please* read these important system notes:
```

```
--> Frontera user documentation is available at:
    https://portal.tacc.utexas.edu/user-guides/frontera
```

```
----- Project balances for user eac3496 -----
| Name          Avail SUs   Expires | Name          Avail SUs   Expires |
| tagmap         1995  2021-12-31 | Phylogenomics    882  2021-12-31 |
----- Disk quotas for user eac3496 -----
| Disk          Usage (GB)   Limit   %Used   File Usage      Limit   %Used |
| /home1         1.8        25.0    7.20    44533      200000  22.27 |
| /work2        813.2       1024.0  79.42    2364      3000000  0.08 |
| /scratch1     861.3         0.0    0.00    17577       0     0.00 |
| /scratch2       0.0         0.0    0.00       1       0     0.00 |
| /scratch3       0.0         0.0    0.00       1       0     0.00 |
-----
```

```
Tip 180  (See "module help tacc_tips" for features or how to disable)
```

You can use the "Alt" key on a Linux keyboard in X-windows to be a "Meta" key. See <https://www.emacswiki.org/emacs/MetaKeyProblems> for details. This can make editing on the command line easier by making a single keystroke Meta-b rather than Escape then b.

ssh command to get into TACC

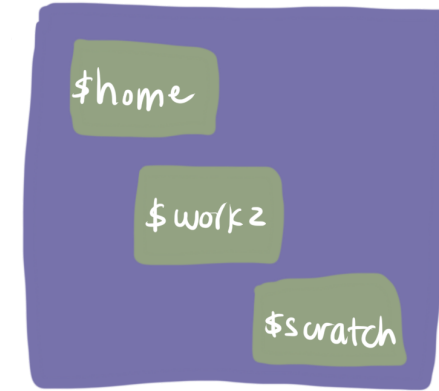
multi-factor authentication

allocations/projects I'm on

how much storage I have left

TACC directory structure

- scratch / home / work directories



File System	Quota	Key Features
\$HOME	1 FOR STORING SOFTWARE	Not intended for parallel or high-intensity file operations. Backed up regularly. Overall capacity ~1PB. Two Meta-Data Servers (MDS), four Object Storage Targets (OSTs). Defaults: 1 stripe, 1MB stripe size. Not purged.
\$WORK	1 STORING PERM. FILES ms, regardless of where on the file system the files reside.	Not intended for high-intensity file operations or jobs involving very large files. On the Global Shared File System that is mounted on most TACC systems. See Stockyard system description for more information. Defaults: 1 stripe, 1MB stripe size Not backed up. Not purged.
\$SCRATCH	n IN PROGRESS/TEMP. FILES	Overall capacity ~30PB. Four MDSs, 66 OSTs. Defaults: 1 stripe, 1MB stripe size. Not backed up. Files are <u>subject to purge</u> if access time* is more than 10 days old.

TACC directory structure: navigating

- scratch / home / work directories

```
[(base) login1.frontera(1000)$ pwd  
/home1/03123/eac3496  
[(base) login1.frontera(1001)$ cds  
[(base) login1.frontera(1002)$ pwd  
/scratch1/03123/eac3496
```

Submitting “jobs” on TACC: two options

- **Development nodes**

- max time allowed is 2 hours
- often get better error messages or can check that things are running before submitting a longer job

- **Submitting jobs**

- max time allowed is 48 hours; very high memory limit

Development node

`idev -t 2:00:00 -A Phylogenomics`

```
((base) login1.frontera(173)$ idev -t 2:00:00 -A Phylogenomics

NOTE: "->" are idev statements. "-->" are TACC/SLURM filter statements.

-> Checking on the status of development queue. OK

-> Defaults file      : ~/.idevrc
-> System             : frontera
-> Nodes              : 1              (idev default   )
-> Tasks per Node     : 56             (Queue default )
-> Time (hh:mm:ss)    : 2:00:00        (cmd line: -t   )
-> Project            : Phylogenomics  (cmd line: -A   )

-----
Welcome to the Frontera Supercomputer
-----

No reservation for this job
--> Verifying valid submit host (login1)...OK
--> Verifying valid jobname...OK
--> Verifying valid ssh keys...OK
--> Verifying access to desired queue (development)...OK
--> Checking available allocation (Phylogenomics)...OK
--> Verifying that quota for filesystem /home1/03123/eac3496 is at 6.79% allocated...OK
--> Verifying that quota for filesystem /work2/03123/eac3496/frontera is at 35.02% allocated...OK
Submitted batch job 3251180

-> After your idev job begins to run, a command prompt will appear,
-> and you can begin your interactive development session.
-> We will report the job status every 4 seconds: (PD=pending, R=running).

-> job status: R

-> Job is now running on masternode= c202-016...OK
-> Sleeping for 7 seconds...OK
-> Checking to make sure your job has initialized an env for you...OK
-> Creating interactive terminal session (login) on master node c202-016.

ssh -Y -o StrictHostKeyChecking no c202-016
TACC Frontera System
Provisioned on 21-Dec-2020 at 05:30

(base) c202-016[clx](99)$ █
```

`exit`

Submitting jobs: the slurm file

These are flags!



```
#!/bin/bash
#SBATCH -J merge_june
#SBATCH -o merge_june.o%j
#SBATCH -N 6
#SBATCH -n 64
#SBATCH -p large
#SBATCH -t 10:00:00
#SBATCH --mail-user=eachambers@utexas.edu
#SBATCH -A Phylogenomics
```

```
[your lines of code here]
```

Submitting jobs: the slurm file

```
(base) login4.stamped2(1004)$ cat allsites_raxmlng.slurm
#!/bin/bash
#SBATCH -J allsites_transect
#SBATCH -o allsites_transect.o%j
#SBATCH -N 4
#SBATCH -n 64
#SBATCH -p normal
#SBATCH -t 48:00:00
#SBATCH --mail-user=eachambers@utexas.edu
#SBATCH -A Phylogenomics

./raxml-ng --msa /scratch/03123/eac3496/transect_allsites_called.min4_RENAMED.phylip --model GTR+G --all --tree rand --bs-trees 200 --threads 2
```


Submitting jobs: sbatch [slurm file]

```
[(base) login3.stampede2(1028)$ sbatch lampro_s1.slurm
```

```
-----  
Welcome to the Stampede2 Supercomputer  
-----
```

```
No reservation for this job
```

```
--> Verifying valid submit host (login3)...OK
```

```
--> Verifying valid jobname...OK
```

```
--> Enforcing max jobs per user...OK
```

```
--> Verifying availability of your home dir (/home1/03123/eac3496)...OK
```

```
--> Verifying availability of your work2 dir (/work2/03123/eac3496/stampede2)...OK
```

```
--> Verifying availability of your scratch dir (/scratch/03123/eac3496)...OK
```

```
--> Verifying valid ssh keys...OK
```

```
--> Verifying access to desired queue (normal)...OK
```

```
--> Verifying job request is within current queue limits...OK
```

```
--> Checking available allocation (Phylogenomics)...OK
```

```
--> Warning: filesystem /home1/03123/eac3496 is over 90% allocated.
```

```
-->           Please, keep /home1/03123/eac3496 under 90% of quota allocation
```

```
--> Verifying that quota for filesystem /work2/03123/eac3496/stampede2 is at 35.02% allocated...OK
```

```
Submitted batch job 8058474
```

```
(base) login3.stampede2(1029)$ █
```

Checking on submitted jobs: `squeue -u [userID]`

```
((base) login1.frontera(123)$ squeue -u eac3496
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
3250710	normal	lampro_s	eac3496	PD	0:00	6	(Priority)

```
(base) login3.stampedede2(1001)$ squeue -u eac3496
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
8055975	normal	merge_ju	eac3496	R	18:31	6	c403-091,c419-104,c420-063,c422-114,c424-102,c427-112
8056020	normal	lampro_s	eac3496	R	1:35	4	c422-[002,124],c424-032,c425-024

```
(base) login3.stampedede2(1002)$
```

`scancel [jobID]`

Copying files to/from TACC: scp

LIVE CODE AN EXAMPLE

- scp stands for secure copy
- syntax is the same as for **cp** or **mv** commands in bash

```
scp [path to file to copy] [path to copied location]
```

- Easiest if you're in your personal computer (regardless of whether you're copying from or to TACC)

Copy file from your personal computer to TACC

```
scp . eac3496@frontera.tacc.utexas.edu:/scratch/file_to_copy.txt
```

Copy file from TACC to your personal computer

```
scp ./file_to_copy.txt eac3496@frontera.tacc.utexas.edu:/scratch/directory_copy_to
```

Running programs / software on TACC

- TACC doesn't have programs (i.e., iPyrad) pre-installed in its system, so we need to install and compile (assemble) the programs ourselves
- Always helpful information online (or through Google!)

Good housekeeping / reproducibility

- *Always* maintain a text file to document what you've done!!!

```
#####  
RUNNING IPYRAD  
  
## Because our data are in different pools with the same barcodes, we'll have to first  
## demultiplex separately and then merge the files together.  
  
JA19241 has 6 pools (degraded samples)  
JA19242 has 13 pools (good samples)  
JA20247 has 1 pool  
JA20248 has 7 pools  
  
# params file for project (modeled after 2b/ddrad project); make sure you're getting the  
# most current params file by first entering  
$ ipyrad -n JA19242pool1  
# This will create a new params file in your current directory; it's easiest to just edit  
# this directly to ensure everything is correctly formatted.  
  
PARAMS FILE  
----- ipyrad params file (v.0.9.13)-----  
JA19241_pool2          ## [0] [assembly_name]: Assembly name. Used to name output directories for assembl  
/scratch1/03123/eac3496/mxrana_ddrad_data/JA19241/pool2/      ## [1] [project_dir]: Project dir (made in curdir if n  
/scratch1/03123/eac3496/mxrana_ddrad_data/JA19241/pool2/*.fastq.gz      ## [2] [raw_fastq_path]:  
/scratch1/03123/eac3496/mxrana_ddrad_data/JA19241/JA19241_barcode_pool2.txt      ## [3] [barcod  
denovo                  ## [4] [sorted_fastq_path]: Location of demultiplexed/sorted fastq files  
                        ## [5] [assembly_method]: Assembly method (denovo, reference)  
                        ## [6] [reference_sequence]: Location of reference sequence file  
pairedddrad            ## [7] [datatype]: Datatype (see docs): rad, gbs, ddrad, etc.  
CATGC,                  ## [8] [restriction_overhang]: Restriction overhang (cut1,) or (cut1, cut2)
```