UT Southwestern Medical Center
Lyda Hill Department of Bioinformatics

BioHPC

Introduction to BioHPC

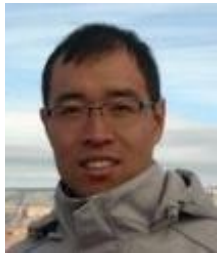
Training will begin at 10:35 AM – please mute your microphones.

[email] biohpc-help@utsouthwestern.edu
[register] portal.biohpc.swmed.edu/accounts/register
[portal] portal.biohpc.swmed.edu

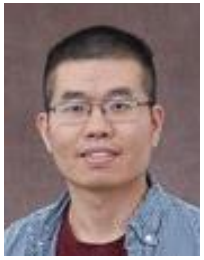
The BioHPC Team

Liqiang Wang
Director

Paniz Karbasi
Ph.D. Electrical and computer engineering



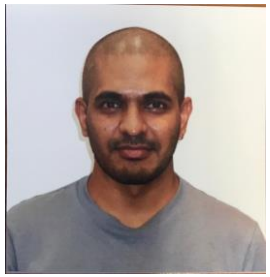
Peng Lian *Ph.D. Biomedical Engineering, Bioinformatics*



Xiaochu Lou -
Ph.D. Biochemistry

Devin O'Kelly
Ph.D., Biomedical and Molecular Imaging

Hung Le
Ph.D., Computational Chemistry



Suresh Pannerselvam
Ph.D., Medical Sciences

Kenny Huynh
B.S., Computer Science

Xueyan Li
M.S., Computational Geophysics

BioHPC institutional resources are available to member departments

- Center for Alzheimer's and Neurodegenerative Diseases
- Cecil H. and Ida Green Center Reproductive Biology Sciences
- Children's Research Institute
- Department of Cell Biology
- Department of Biophysics
- Department of Biochemistry
- Department of Immunology
- Department of Ophthalmology
- Department of Pathology
- Department of Radiology
- Department of Urology
- Department of Population and Data Sciences
- Department of Psychiatry
- Department of Internal Medicine
- Department of Anesthesiology and Pain Management
- Green Center for Systems Biology
- Hamon Center for Therapeutic Oncology Research
- Harold C. Simmons Comprehensive Cancer Center
- Lyda Hill Department of Bioinformatics
- McDermott Center for Human Growth and Development
- Peter O'Donnell Jr. Brain Institute
  - And sub-departments
- Texas Institute for Brain Injury and Repair
- Touchstone Diabetes Center

Regularly updated list available at:
https://portal.biohpc.swmed.edu/content/about/

If you don't see your department in our membership list, come talk with us biohpc-help@utsouthwestern.edu

BioHPC Business Model

- The UT Southwestern BioHPC team provides and maintains high-performance computing, storage and client systems for the UTSW research community.

- The business model can be summarized as "**By the users - For the users**"
  - BioHPC team works directly with Department chair or administrator on resource requirements
  - Department chair or administrator internally coordinates with PIs on the total expected amount of compute resources and storage.

- Questions? Please reach out to us: biohpc-help@utsouthwestern.edu

What is HPC, and why do we need it?

High-performance computing (HPC) is the use of aggregated computing power to solve large-scale or computationally intensive problems in science and engineering.

The power to solve problems that personal computers can't

CHALLENGES

Huge, Diverse, "Big Data" Datasets

Complex Algorithms

Difficult / Inefficient Software

Quick Start - Usage Scenario

- **You have a complex analysis that takes a long time to run on your desktop PC**, and you would like to use your desktop for other things like doing research, making figures and writing manuscripts.

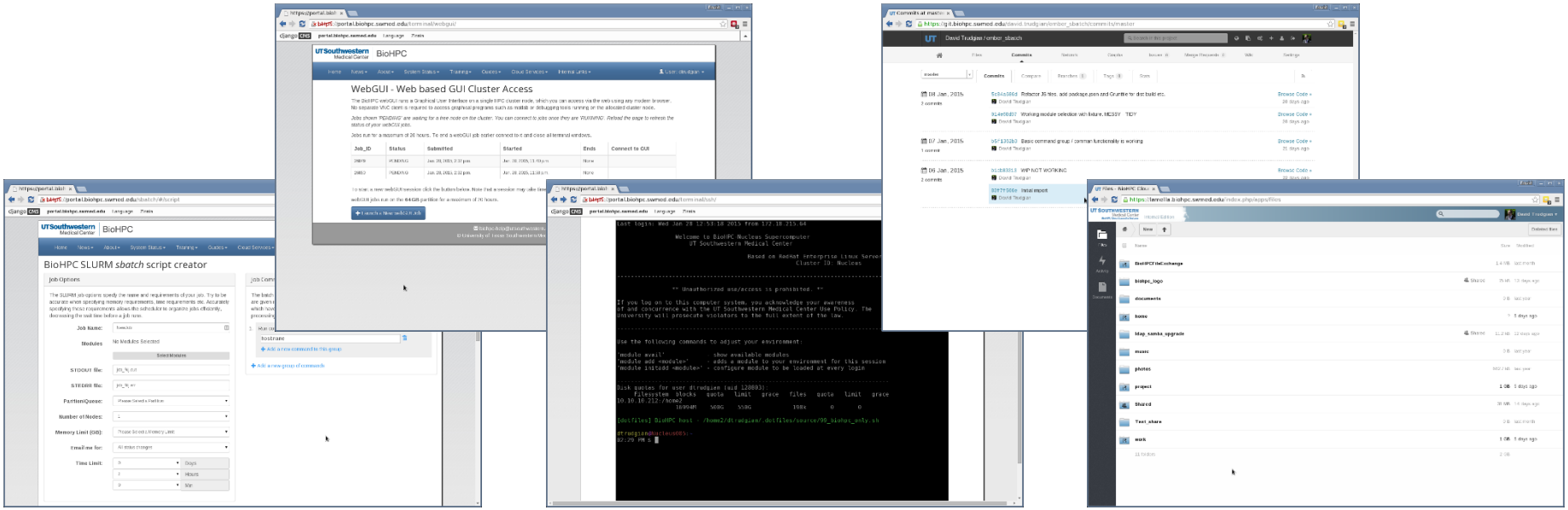
- Everything is on your desktop's hard drive: scripts, data, libraries for running code, etc.

- You would like to securely store your data and code on BioHPC and use its computational power to run your analysis.

What does BioHPC provide? – *Cloud Services*

A big focus at BioHPC is ***easy access*** to our systems.

Our ***cloud services*** provide web-based access to resources, with only a browser.



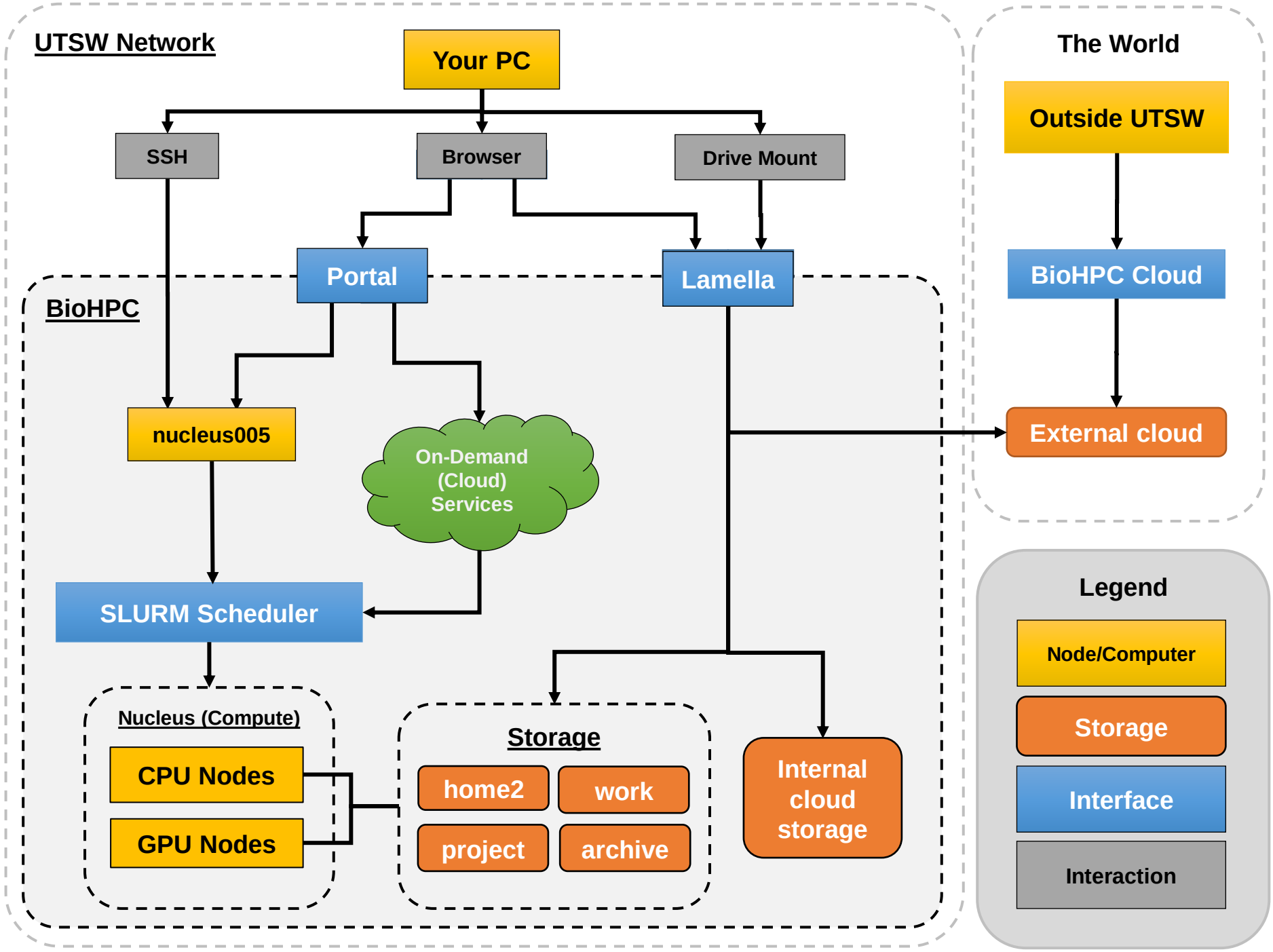
All cloud services accessible via ***portal.biohpc.swmed.edu***

What does BioHPC provide? – *Software access*

A wide range of software packages are available on Nucleus as *modules*.

Module	Usage	url	Keywords	Versions
3D modeling				
mricron	0	http://people.cas.sc.edu/rorden/mricron/index.html	imaging	10-2016
Alignment				
bustools	1	https://github.com/BUStools/bustools	RNA-Seq, Alignment, analysis	0.39.3
mauve	0	http://darlinglab.org/mauve/mauve.html	genome, alignment	2.4.0
star	4968	https://github.com/STAR-Fusion/STAR-Fusion	RNA-Seq, Alignment, analysis	2.7.2b 2.4.2a; 2.5.2b; 2
Allignment				
star	4968	https://github.com/STAR-Fusion/STAR-Fusion	RNA-Seq, Alignment, analysis	2.7.2b 2.4.2a; 2.5.2b; 2
Assembler				
spades	2180	http://cab.spbu.ru/files/release3.13.0/manual.html	genome, assemly	gcc/3.13.0
Bioinfomatics				
bfast	0	https://sourceforge.net/projects/bfast	search, sequencing	0.7.0a
bismark	104	http://www.bioinformatics.babraham.ac.uk/projects/bismark/	cytosine, methylation, mapping	0.21.0 0.14.5;
Biophysics				
nmrpipe	0	https://www.ibbr.umd.edu/nmrpipe/index.html	nmr, spectroscopy, pipeline	20181101
Biostastics				
mzmine	12	http://mzmine.github.io/	LC-MS, biostatistics, spectroscopy	2.40.1
Biostatistics				
R	8418	http://www.r-project.org	Biology, Biostatisticst	3.3.2-gccmkl(default) 2.15. 3.1.0-intel; 3.2.1-intel; 3.4.1-gccmkl; 3.4.1-gcc gccmkl; 3.5.1-gccmkl_2
C				

You can ask biohpc-help@utsouthwestern.edu for additions/upgrades etc.


UTSW Network
Your PC
SSH
Browser
Drive Mount
Portal
Lamella
BioHPC
nucleus005
On-Demand (Cloud) Services
SLURM Scheduler
Nucleus (Compute)
CPU Nodes
GPU Nodes
Storage
home2
work
project
archive
Internal cloud storage
The World
Outside UTSW
BioHPC Cloud
External cloud
Legend
Node/Computer
Storage
Interface
Interaction

Nucleus is our supercomputing cluster (our "compute")

CPU Nodes

Light Nodes:

32 GB	276 nodes

Heavy Nodes:

128 GB	24 nodes
256 GB	89 nodes
256 GBv1	54 nodes
384 GB	18 nodes

Nucleus (Compute)

CPU Nodes

GPU Nodes

GPU Nodes

Tesla K20/K40	8 nodes
Tesla P4	16 nodes
Tesla P40	16 nodes
2x Tesla P100	12 nodes
Tesla V100	32 nodes
4x Tesla V100	12 nodes
Ampere A100	16 nodes
4x Ampere A100	16 nodes

Coming soon

256 GB	~300 nodes

https://portal.biohpc.swmed.edu/content/about/systems/

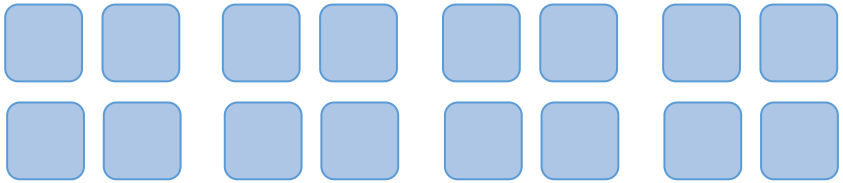
RHEL 7.7, GNOME 3, Bash

How many nodes can I use at once (if they are available)?

4 GPU nodes

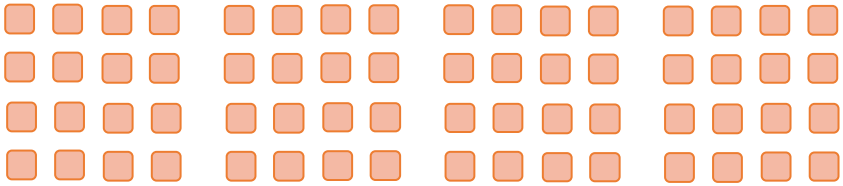
AND

16
heavy nodes
(> 32 GB ea.)



OR

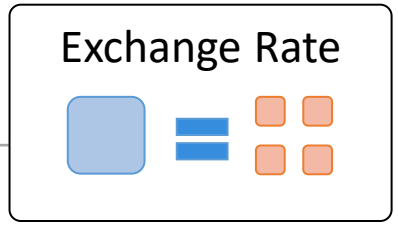
64
light nodes
(32 GB ea.)



Exchange Rate

One (1) Heavy = Four (4) Light

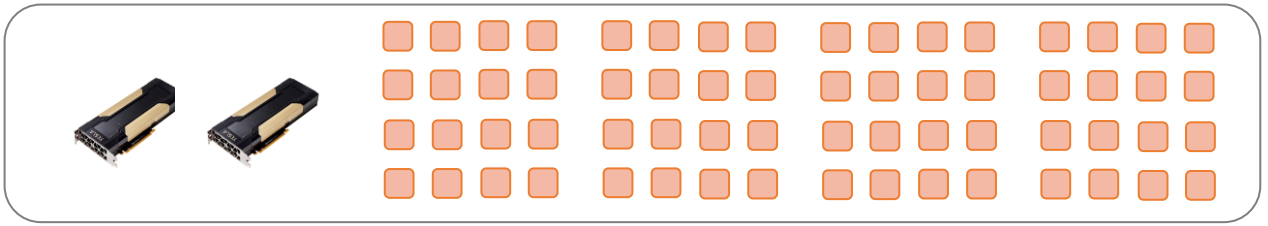
Possible total allocations



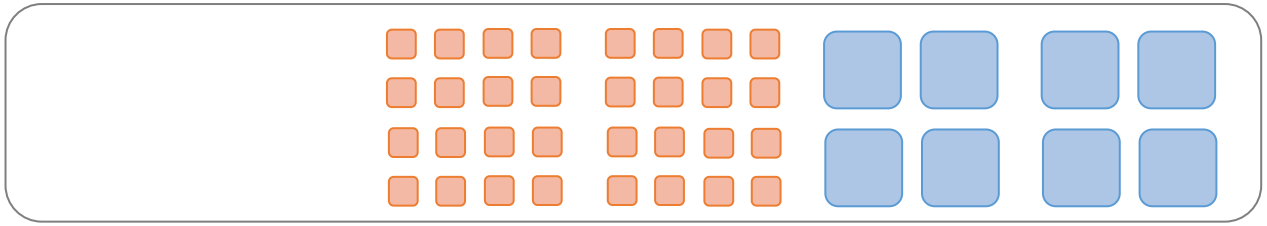
2 GPU + 64 light nodes

1024 CPU cores

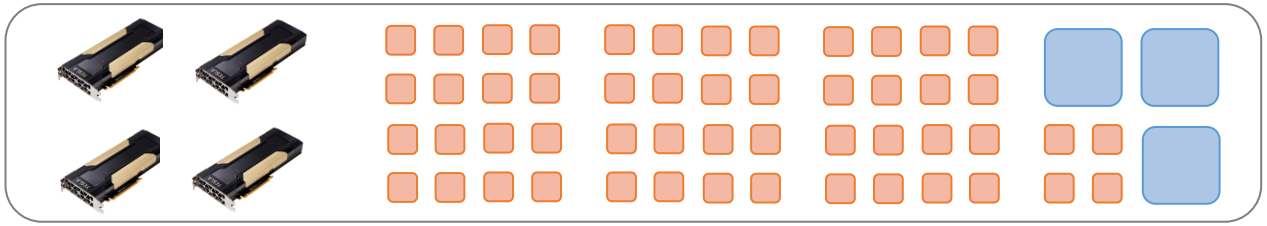
2048 GB RAM



0 GPU + 8 heavy + 32 light



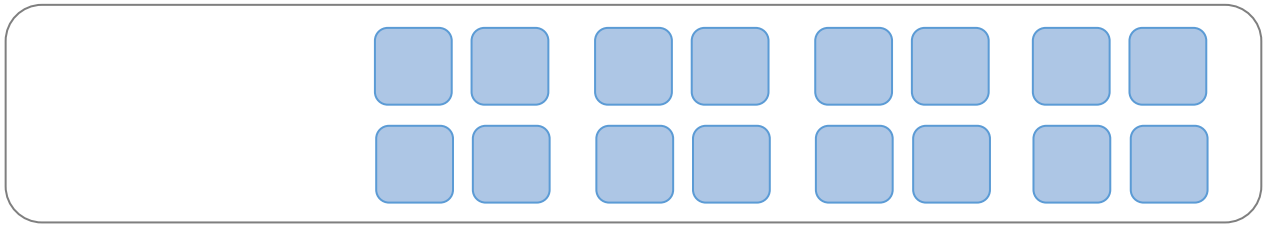
4 GPU + 3 heavy + 52 light

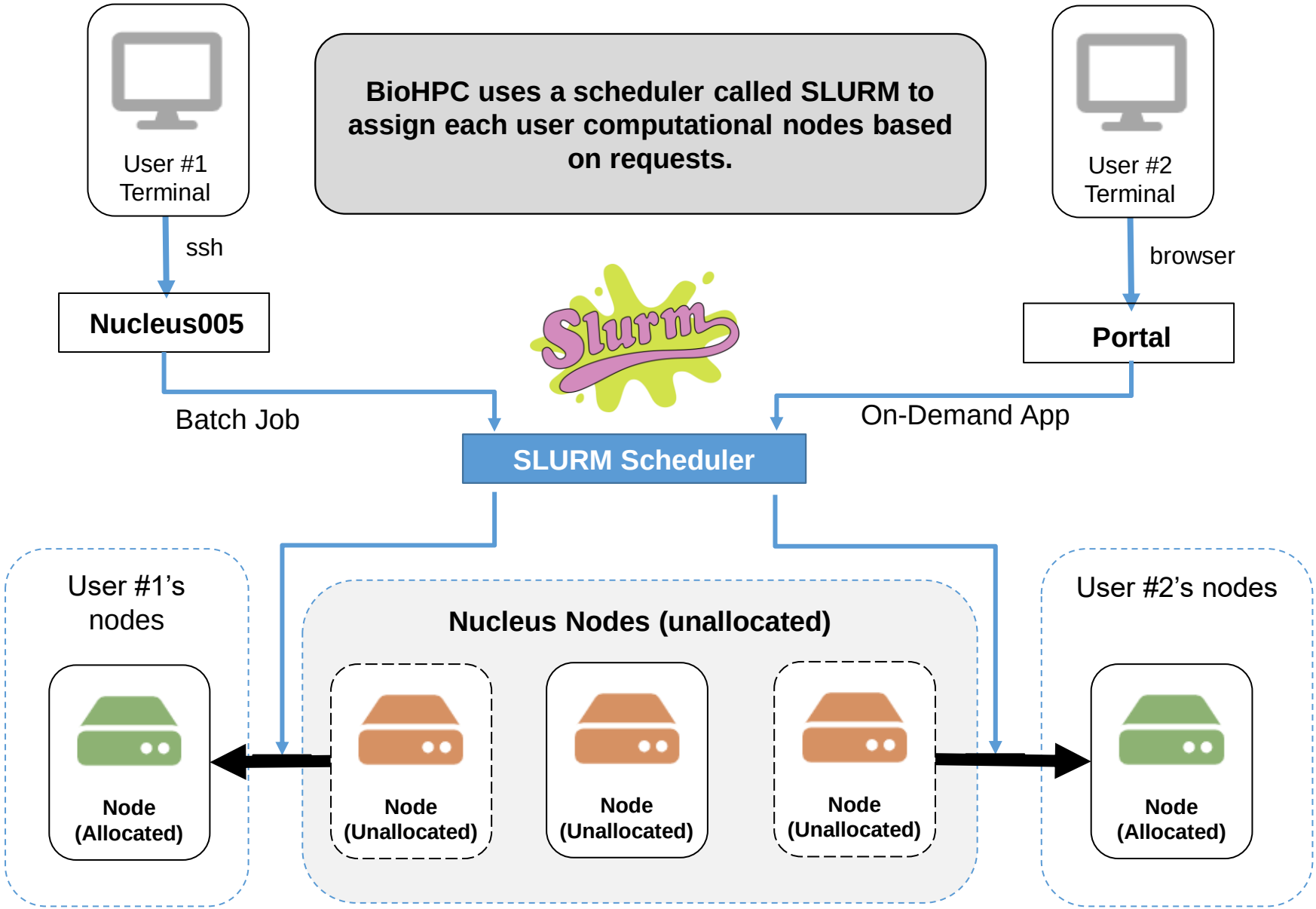


0 GPU + 16 heavy

448 CPU cores

4096 GB RAM




User #1
Terminal
BioHPC uses a scheduler called SLURM to assign each user computational nodes based on requests.
User #2
Terminal
ssh
browser
Nucleus005
Slurm
Portal
Batch Job
On-Demand App
SLURM Scheduler
User #1's nodes
Nucleus Nodes (unallocated)
User #2's nodes
Node (Allocated)
Node (Unallocated)
Node (Unallocated)
Node (Unallocated)
Node (Allocated)

BioHPC - Storage

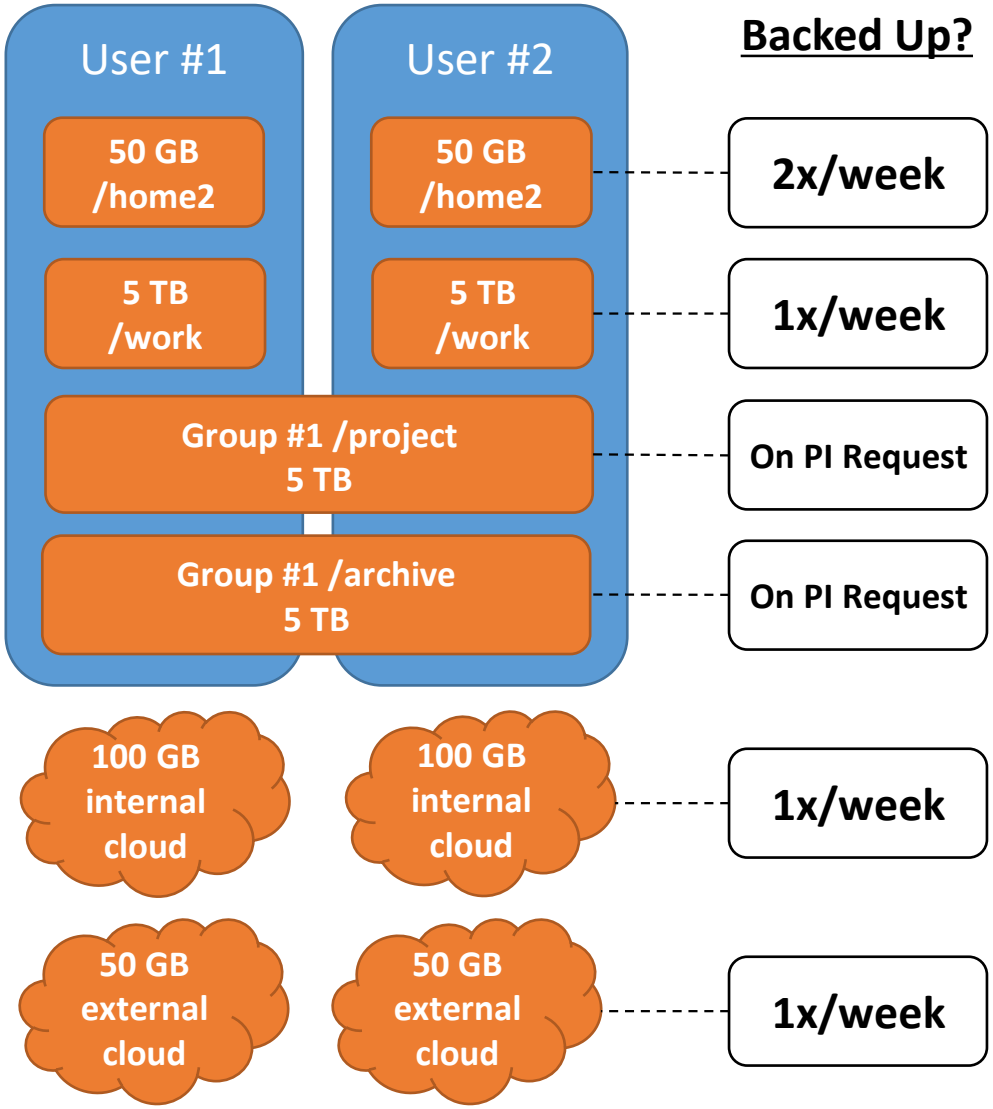
Every user has a **user quota**.
Every group has a **group quota.**
Group quotas are shared among members of that group.

Every user gets:

- **50 GB** in **/home2**
- **5 TB** in **/work**
- **100 GB** internal cloud storage
- **50 GB** external cloud storage

Every group gets:

- **5 TB** in **/project**
- **5 TB** in **/archive – 7.5 TB effective**
  - **2/3x "data discount"**



Info and backup policy: https://portal.biohpc.swmed.edu/content/guides/storage-cheat-sheet/

Scripts, Configurations

/home2

/home2/<username>
Small files which are frequently accessed

"Hot" data

/work

/work/<department>/<username>
Data which is read and written frequently; active analyses

"Warm" data

/project

/project/<department>/<group>/<username>
Data which is read and written occasionally; raw data

"Cold" data

/archive

/archive/<department>/<group>/<username>
Data which is rarely read; archival/retained data

Inside UTSW

100 GB internal cloud

Dropbox-like interface, for sharing **inside** UTSW network

Outside UTSW

50 GB external cloud

Dropbox-like interface, for sharing **outside** UTSW network

Exceptions for O'Donnell Brain Institute (OBI)

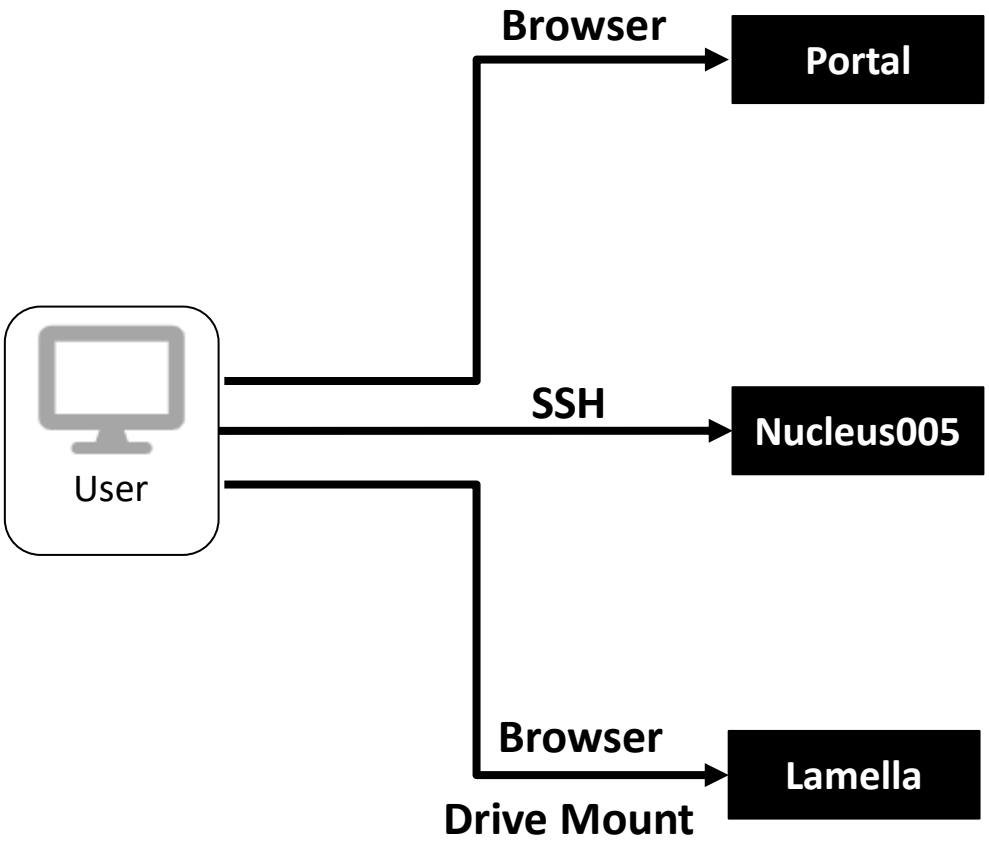
If you are a member of OBI's sub-departments, your paths are:

The same!

/home2	**/home2/<username>** Small files which are frequently accessed
/work	**/work/OBI/<department>/<username>** Data which is read and written frequently; active analyses
/project	**/project/OBI/<department>/<group>/<username>** Data which is read and written occasionally; raw data
/archive	**/archive/OBI/<department>/<group>/<username>** Data which is rarely read; archival/retained data

Whenever you see <department> in general instructions, replace it with OBI/<department>

Three basic ways to interact with BioHPC



portal.biohpc.swmed.edu

Web Services, Guides, Training, Information

nucleus.biohpc.swmed.edu

The login node. Schedule jobs, hop to another node

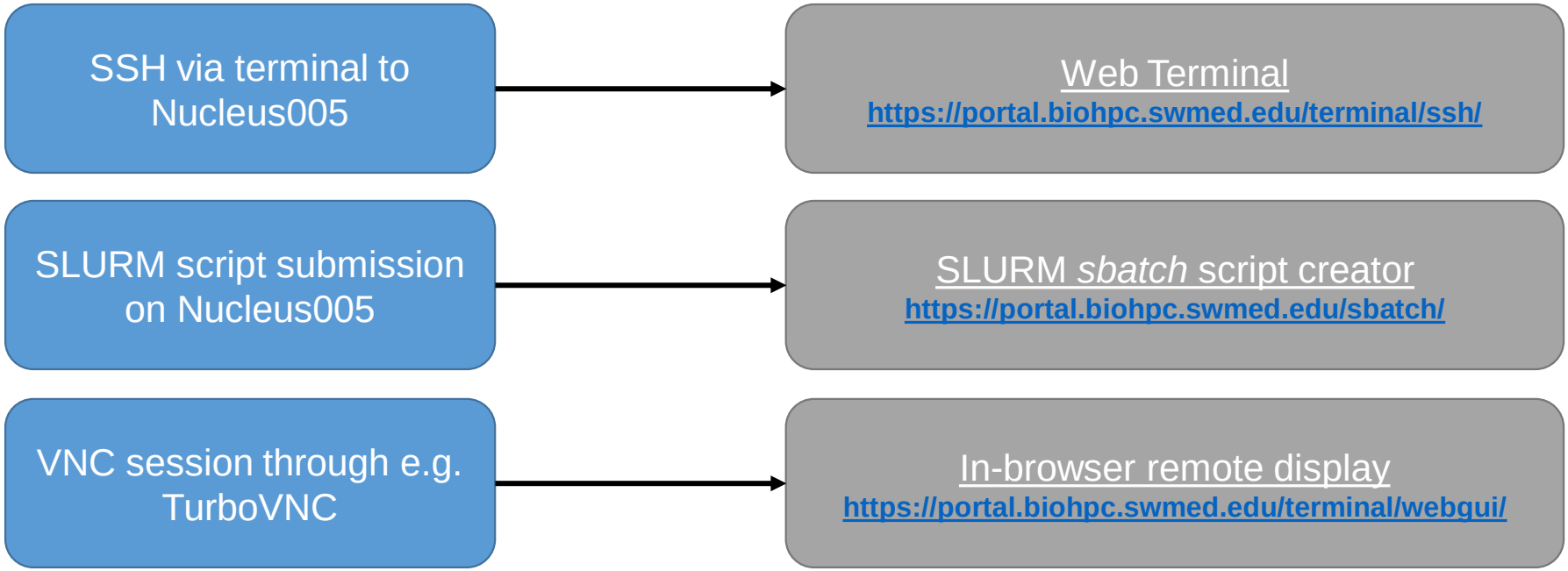
lamella.biohpc.swmed.edu

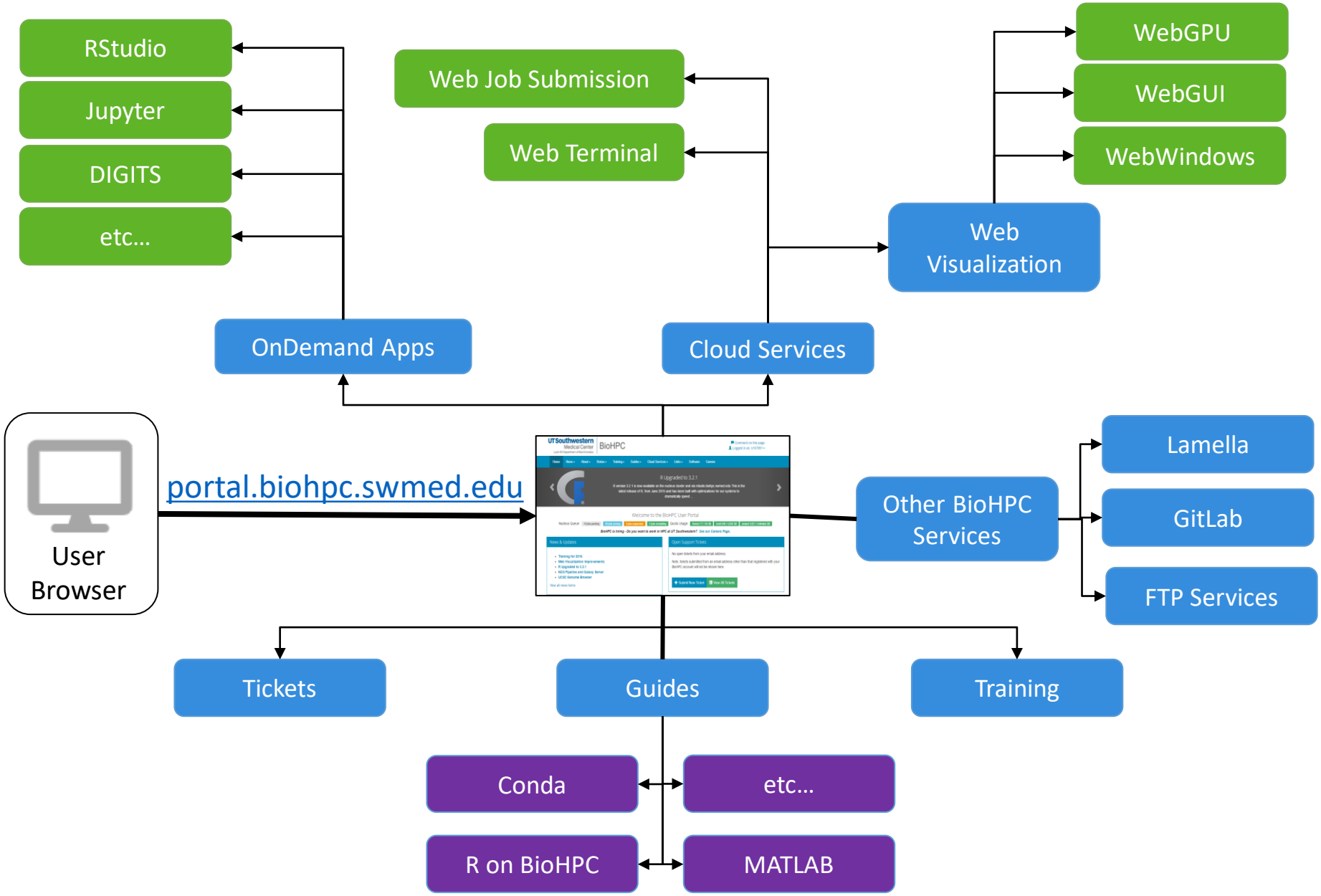
Storage

Accessing BioHPC – Portal

Our goal is for BioHPC to be available to users of all technical backgrounds, and to streamline the process of becoming more comfortable and familiar with HPC by lowering the barriers to access.

Many basic tasks have a Portal-based alternative.




RStudio
Jupyter
DIGITS
etc...
Web Job Submission
Web Terminal
WebGPU
WebGUI
WebWindows
Web Visualization
OnDemand Apps
Cloud Services
User Browser
portal.biohpc.swmed.edu
Other BioHPC Services
Lamella
GitLab
FTP Services
Tickets
Guides
Training
Conda
etc...
R on BioHPC
MATLAB

Accessing BioHPC – Secure Shell (SSH)

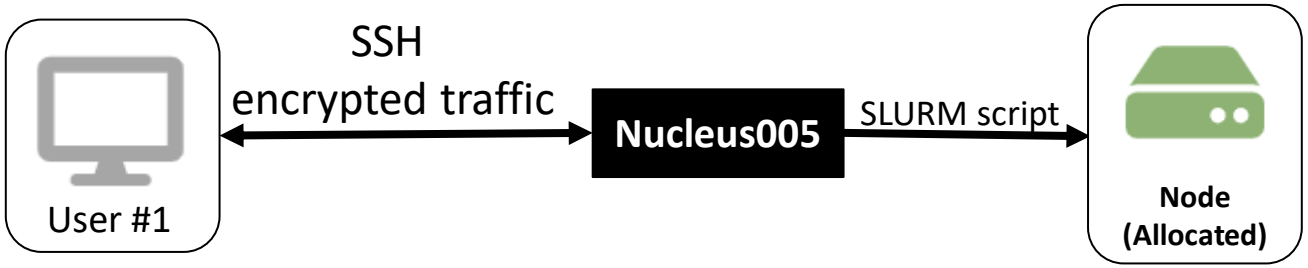
SSH provides a command line that you can use to perform tasks on a remote computer.

Users may connect to the Nucleus login node (**nucleus005)** via Secure Shell (SSH) sessions. This is preferred to other methods (e.g. WebGUI) as it does not use up a node.

```
> ssh <username>@nucleus.biohpc.swmed.edu
```

Nucleus005 is the main hub, and is for scheduling jobs to run on **other nodes**.

You should never run analyses or intensive computation on Nucleus005



Accessing BioHPC - SSH Clients

Linux

Most Linux distributions have OpenSSH installed by default.

Simply open a terminal window.

Mac

The Mac Terminal should already be able to SSH by default.

Windows

Either:
Install an SSH client

- PuTTY
- Windows Subsystem for Linux (WSL)
- Git Bash (provided by Git)
- ConEmu
- MobaXterms

Or use Command Prompt (cmd) or Powershell

BioHPC Portal

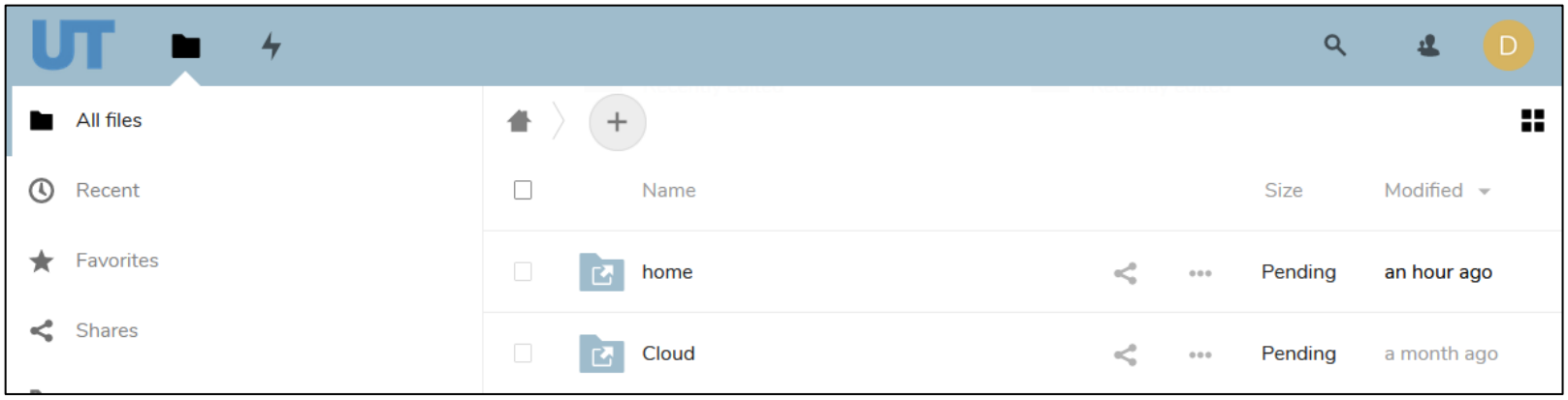
Navigate to Web Terminal.
Portal > Cloud Services > Web Terminal
https://portal.biohpc.swmed.edu/terminal/ssh/

Accessing BioHPC - Lamella

Lamella is BioHPC's storage gateway. By default, it is configured to provide cloud access to your **internal cloud**, **external cloud**, and **home2** storage.

Lamella translates data transfer protocols **so that you can access BioHPC storage from non-BioHPC computers.**

There is a browser-based interface, similar to Dropbox or other cloud storage, at http://lamella.biohpc.swmed.edu

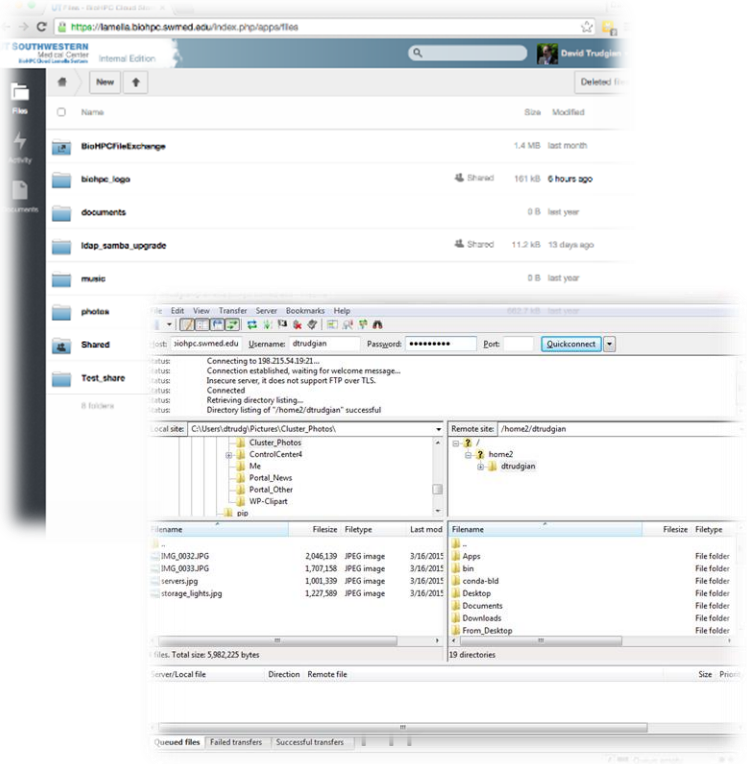


Cloud storage guide: https://portal.biohpc.swmed.edu/content/guides/biohpc-cloud-storage/

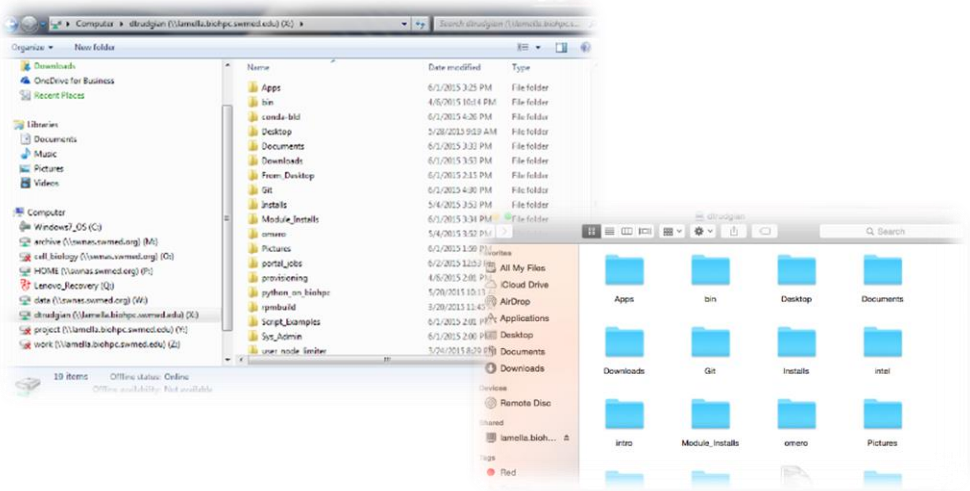
Accessing BioHPC - Lamella

Multiple ways to access your files:

Web Interface (Lamella)
FTP Clients (Thunder/Flash)



Windows / Mac network drive mounts
(SMB /WebDav)



lamella.biohpc.swmed.edu

Accessing BioHPC – Thin Clients and Workstations

- BioHPC has specially-configured desktop clients available for order.

- Can submit jobs directly to the cluster **(like Nucleus005**)
- Direct access to the cluster storage systems.
- Same software stack and modules as compute nodes.
- Provides you with a graphical desktop (like a web visualization session)
- **Customizable (to some extent)**, persistent, can run some software we disallow on compute nodes (e.g. Virtualbox)

- Thin clients are useful for light workloads.
- Workstations are designed for intensive development or local computation.

Okay, sounds great….

But how do I use any of this?

Quick Start - Usage Scenario

You have a complex analysis that takes a long time to run on your desktop PC, and you would like to use your desktop for other things like doing research, making figures and writing manuscripts.

We will walk through a sketch of the process via the following basic walkthrough:

1. Move data and code from your desktop to BioHPC, via **Lamella mounts**
2. Submitting a SLURM batch script for your code, via **Web Job Submission**
3. Checking on our script via the **Web Terminal**
4. Visualizing our results via a **Web Visualization** session and **loading modules**
5. Running OnDemand apps

Hands on BioHPC – 1. Manage Files with Lamella / Cloud Storage Gateway

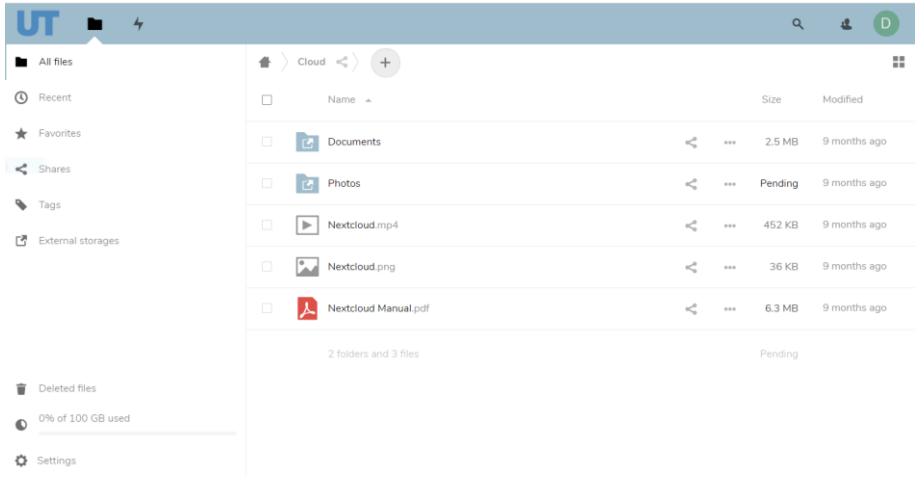
Cloud storage gateway – web-based.

https://lamella.biohpc.swmed.edu

- 100GB separate space +
- Mount /home /project /work
- Internal

https://cloud.biohpc.swmed.edu

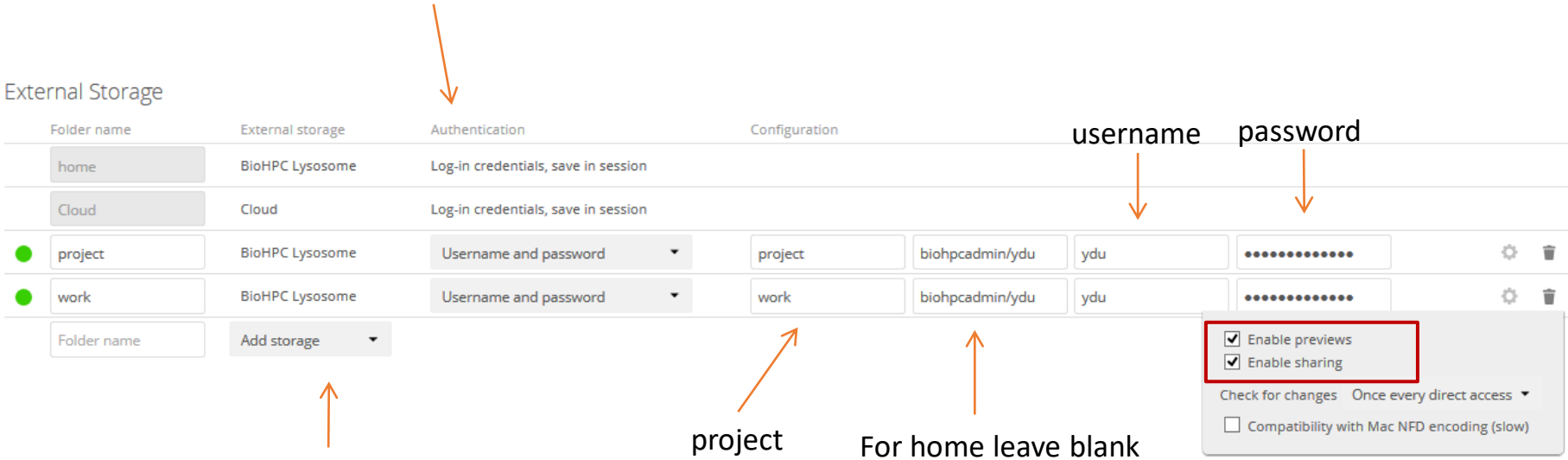
- 50GB space
- External file transfer
- Accessible from Internet



Hands on BioHPC – 1. Setting up Lamella to access project and workspace

https://lamella.biohpc.swmed.edu

Log-in credentials, save in session uses the BioHPC login credentials and only saved in the user session, giving increased security. The drawbacks are that sharing is disabled, as lamella has no access to the cluster storage credentials.
Username and password mechanism requires a manually-defined username and password.
Remember to click the gear icon and enable sharing.



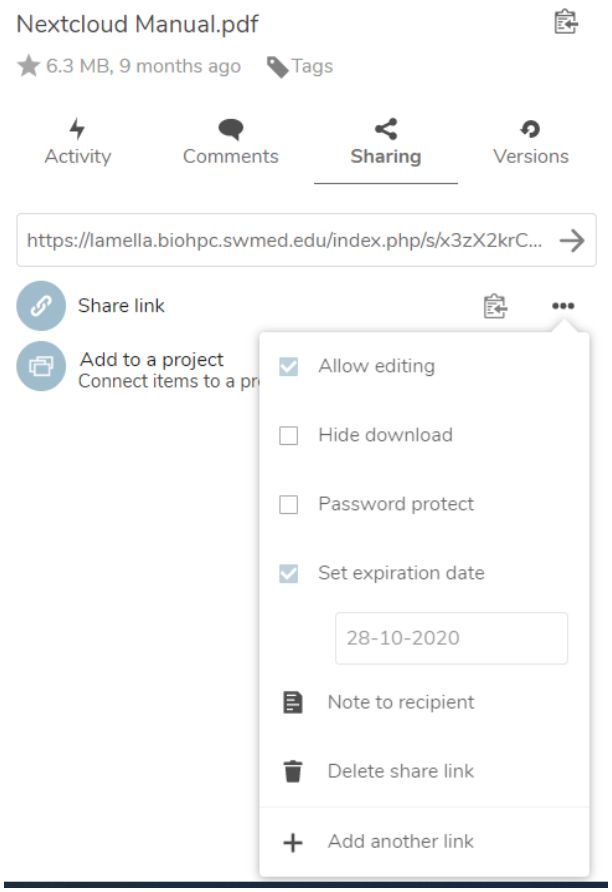
BioHPC Endosome/Lysosome

project
work
home

For home leave blank
For private project space:
department/lab/user
For lab shared project space:
department/lab/shared

Hands on BioHPC – 1. Manage Files with Lamella / Cloud Storage Gateway

File Sharing



Lamella cloud storage https://lamella.biohpc.swmed.edu : sharing with user **inside** UTSW
File Exchange https://cloud.biohpc.swmed.edu : sharing with user **outside** UTSW

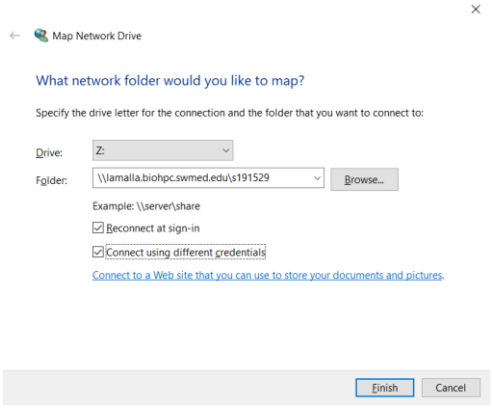
Hands on BioHPC – 1. Accessing BioHPC Storage Directly from Windows

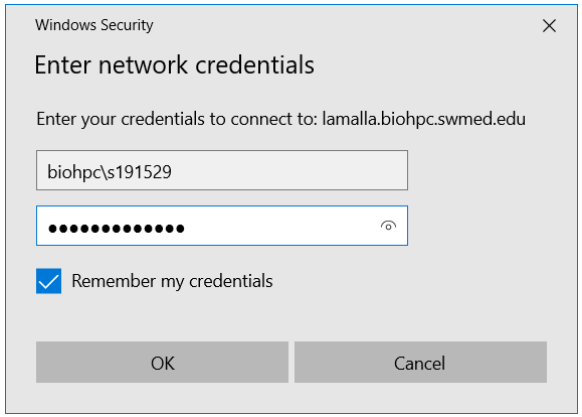
Computer -> Map Network Drive

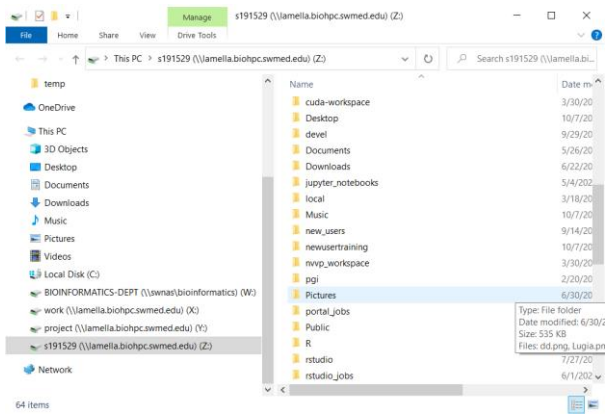
Folder is: \\lamella.biohpc.swmed.edu*username* *(home dir)*
\\lamella.biohpc.swmed.edu\project
\\lamella.biohpc.swmed.edu\work

Check ‘Connect using different credentials’

Enter your BioHPC username and password when prompted.





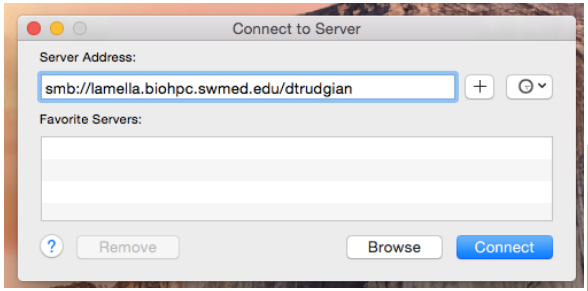


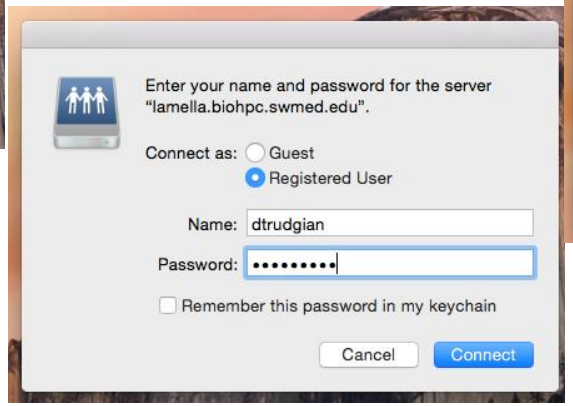
Hands on BioHPC – 1. Accessing BioHPC Storage Directly from Mac OSX

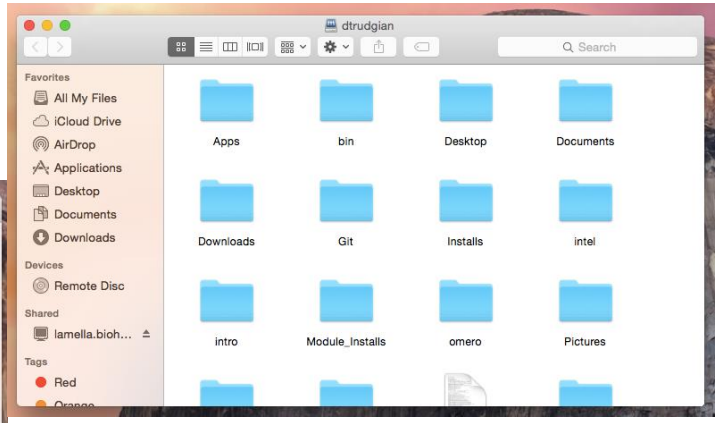
Finder -> Go -> Connect to Server

Folder is: smb://lamella.biohpc.swmed.edu/*username* *(home dir)*
smb://lamella.biohpc.swmed.edu/project
smb://lamella.biohpc.swmed.edu/work

Enter your BioHPC username and password when prompted.

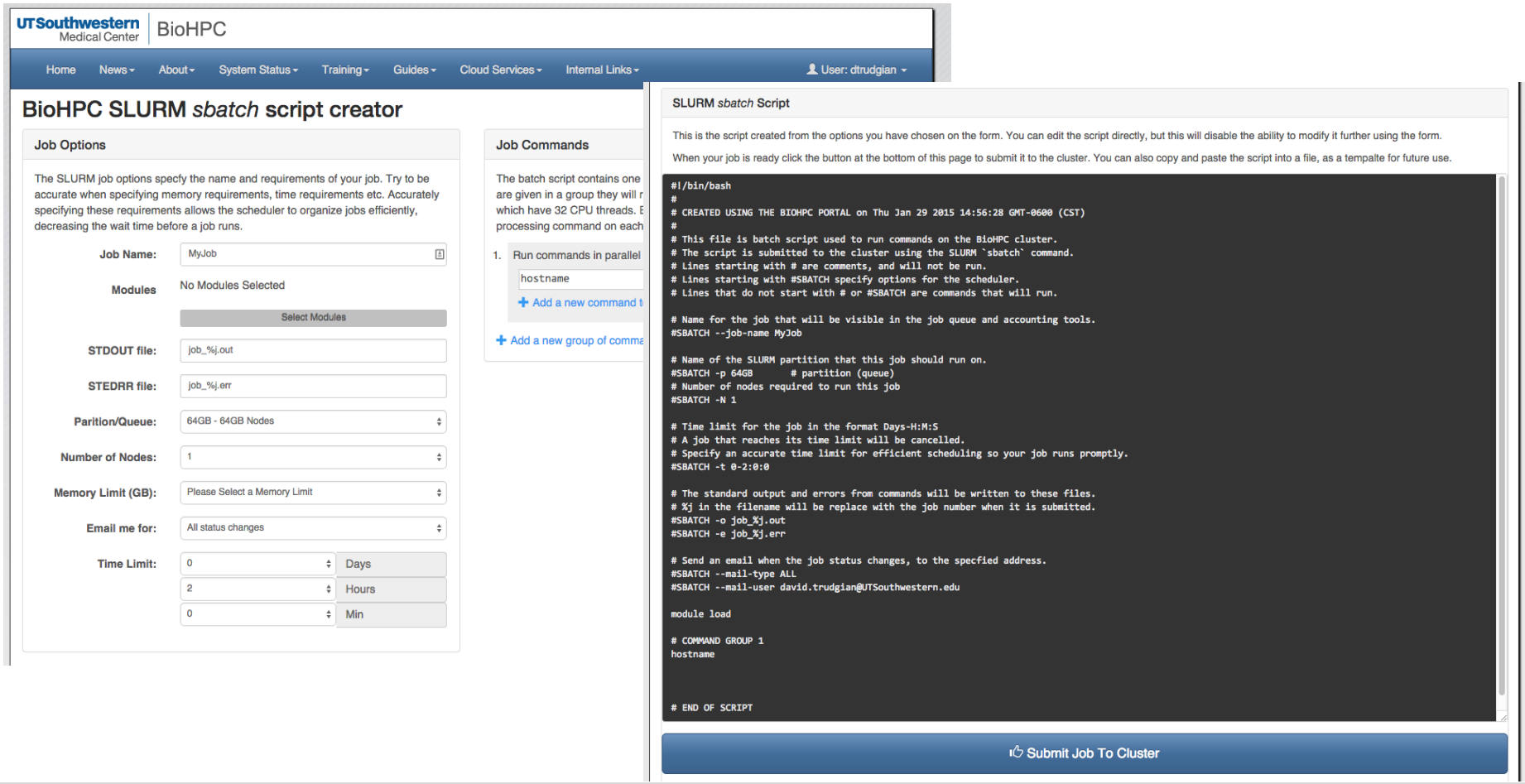






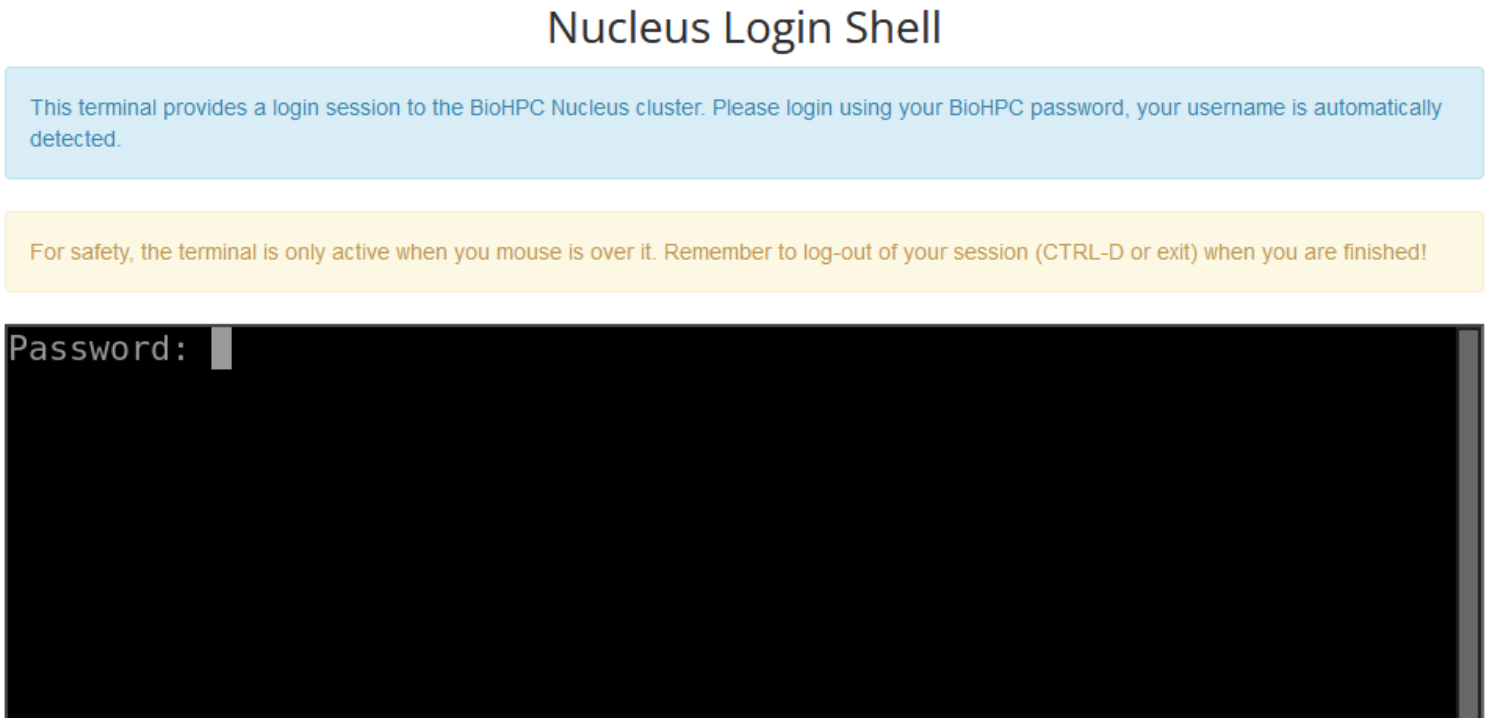
Hands on BioHPC – 2. Web Job Script Generator

https://portal.biohpc.swmed.edu -> Cloud Services -> Web Job Submission



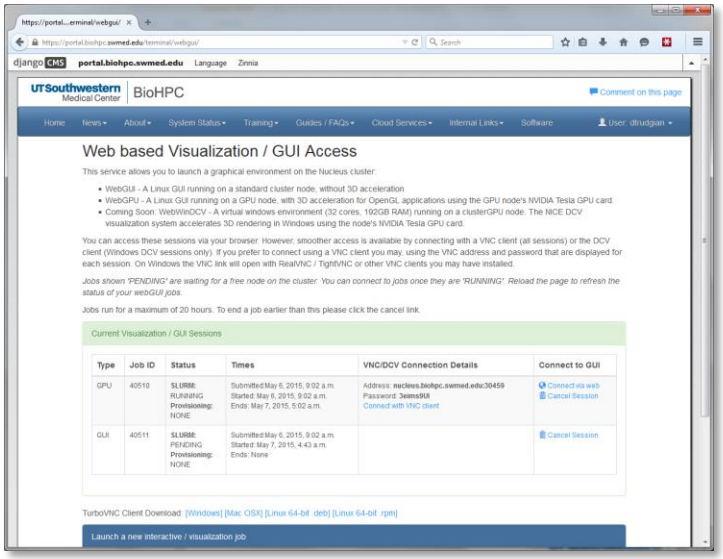
Hands on BioHPC – 3. Web Terminal

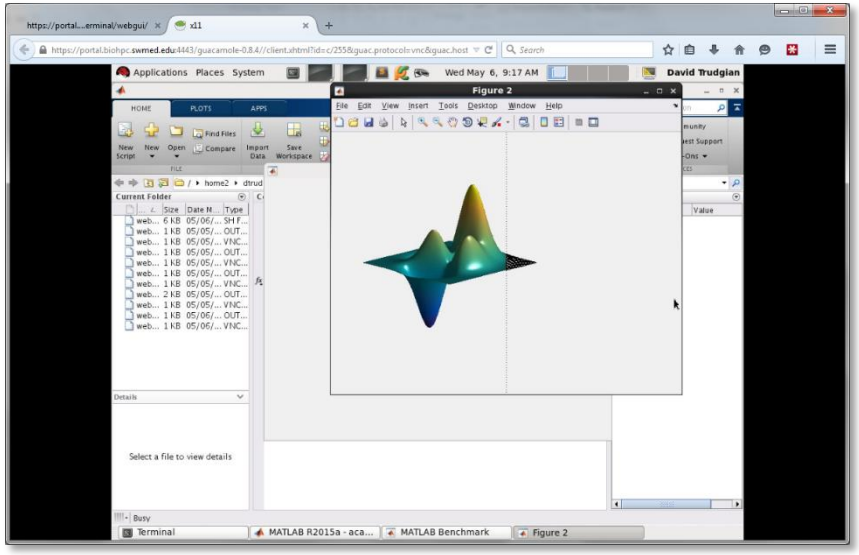
https://portal.biohpc.swmed.edu -> Cloud Services -> Web Terminal



Hands on BioHPC – 4. Web Visualization: Graphical Interactive Session via Web Portal/VNC

https://portal.biohpc.swmed.edu -> Cloud Services -> Web Visualization





Connects to GUI running on a cluster node. WebGPU sessions have access to GPU card for 3D rendering.

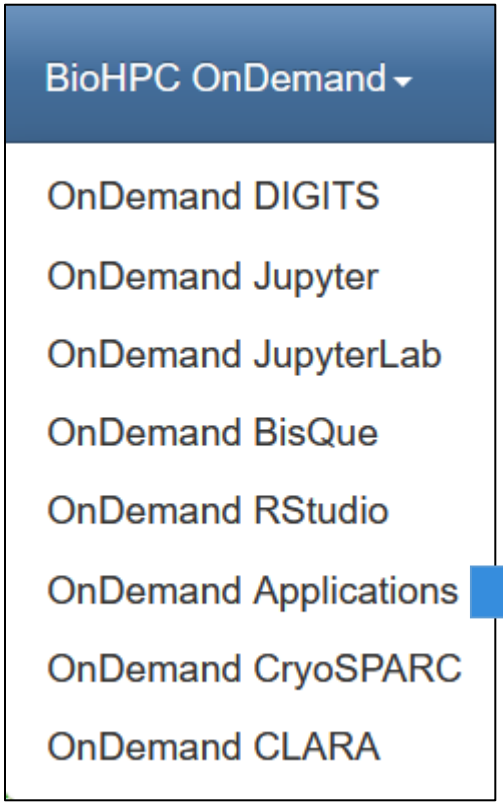
Hands on BioHPC – 4. (continued) Software Modules

```
dtrudgian@Nucleus005:~
03:16 PM $ module avail

------------------------------------------------------ /cm/local/modulefiles ------------------------------------------------------
cluster-tools/6.1             ipmitool/1.8.12               null                          use.own
cmd                           module-git                    openldap                      version
dot                           module-info                   openmpi/gcc/64/1.6.5-mlnx-ofed
freeipmi/1.2.6                mvapich2/gcc/64/1.9-mlnx-ofed shared

------------------------------------------------------ /cm/shared/modulefiles -----------------------------------------------------
abyss/1.3.6                        fftw2/openmpi/open64/64/double/2.1.5 NAMD/2.9/ibverbs-smp-CUDA
acml/gcc/64/5.3.1                  fftw2/openmpi/open64/64/float/2.1.5  NAMD/2.9/multicore
acml/gcc/fma4/5.3.1                fftw3/openmpi/gcc/64/3.3.3           NAMD/2.9/multicore-CUDA
acml/gcc/mp/64/5.3.1               fftw3/openmpi/intel/3.3.3            netcdf/gcc/64/4.3.0
acml/gcc/mp/fma4/5.3.1             fftw3/openmpi/open64/64/3.3.3        netcdf/intel/4.3.0
acml/gcc-int64/64/5.3.1            fftw3/shared/3.3.4                   netcdf/open64/64/4.3.0
acml/gcc-int64/fma4/5.3.1          gcc/4.8.1                            netperf/2.6.0
```

Command	Summary
module list	Show loaded modules
module avail	Show available modules
module load <module_name>	Load module into environment
module unload <module_name>	Unload module from environment
module help <module_name>	Help text for a specific module
module -h	General module command help

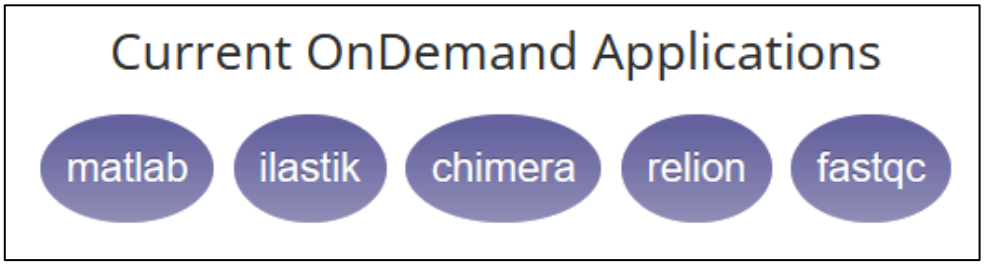
*BioHPC may install additional modules upon request.

Hands on BioHPC – 5. OnDemand apps

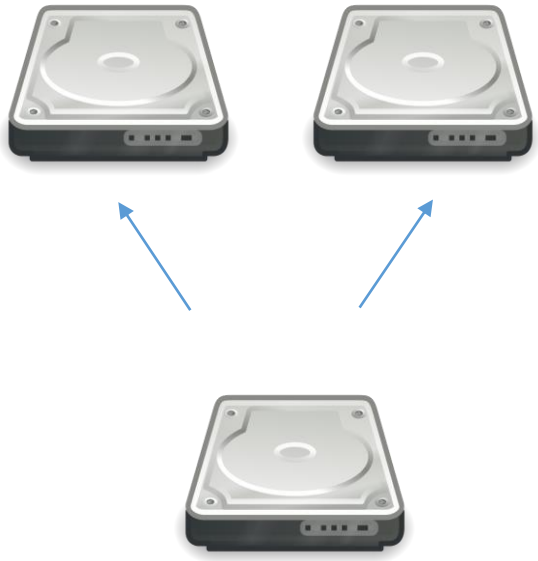


BioHPC has made convenient interfaces for accessing some commonly used software packages.

OnDemand apps consume 1 light (32 GB) node each.



What is BioHPC not intended for?



Rigorous file version backups

BioHPC is primarily a *computational resource*, not a *storage provider*.

Non-Research Computing

BioHPC is publicly funded; every user is expected to be responsible with taxpayer dollars.

Things to Remember

1. If you ever have any BioHPC-related questions or need assistance, contact us at ***biohpc-help@utsouthwestern.edu***

2. Be conservative when requesting resources – Do you really need a heavy node when a light node will do?

3. Make reasonable attempts to use the resources efficiently.
   - Use one node for multiple tasks if you can, rather than reserving several nodes.
   - Cancel or close any jobs/session you no longer need.
   - Try to optimize your code.

4. Keep notes in case you need our help troubleshooting.

5. **Never run any intensive code on Nucleus005, AKA login node, AKA web terminal AKA nucleus.biohpc.swmed.edu.**

Persistent, improper use of BioHPC resources is grounds for having your account frozen.

Getting Effective Help

Email the ticket system: biohpc-help@utsouthwestern.edu

What is the problem?

Provide any error messages, logs, and diagnostic output you may have.

When did it happen?

What time? Cluster node or client workstation? What job ID? What folders were you in and accessing?

How did you run it?

What did you run? Which parameters and settings did you use?

Any unusual circumstances?

Have you compiled your own software? Do you customize startup scripts?

Can we look at your scripts and data?

Depending on the problem and its complexity, we may need your permission to access your storage allocation. Please let us know if there is any data that is sensitive.

Special Consulting

BioHPC can also provide more detailed assistance to help *facilitate* research.

- **Collaborations**
  - Need help beyond a support ticket?
  - The BioHPC team can provide in-depth consultations to assist with particular research projects.
- **Liaison**
  - Lab with need for *dedicated*, domain-specific computational support
  - Co-hired by BioHPC and the lab

Next Steps

- Please register immediately if you have not done so at (UTSW VPN required):
  - https://portal.biohpc.swmed.edu/accounts/register/
- New user registrations will be manually reviewed and processed (**a week or two**)
  - Your PI should be prepared to approve your activation.
- You would receive an activation notice when registration is processed.
- Spend some time experimenting with our systems and browse our guides.
- Check the training schedule and attend relevant sessions on BioHPC Portal
  - *https://portal.biohpc.swmed.edu/content/training/calendar-2022/*
- QUESTIONS: biohpc-help@utsouthwestern.edu

Thank you!

Please acknowledge our contribution:

This research was supported in part by the computational resources provided by the BioHPC supercomputing facility located in the Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center.

If your PI wants to include info in a grant proposal on BioHPC as a resource, or request grant funds to contribute to their Department's HPC costs, please contact our Department Administrator, **Rebekah Craig (Rebekah.Craig@utsouthwestern.edu)** for advice on grant language.