HIGH PERFORMANCE RESEARCH COMPUTING

HPRC Primer

Data Management Practices on HPRC Resources

August 30, 2024

High Performance
Research Computing
DIVISION OF RESEARCH

HPRC Resources

Knowledge Base

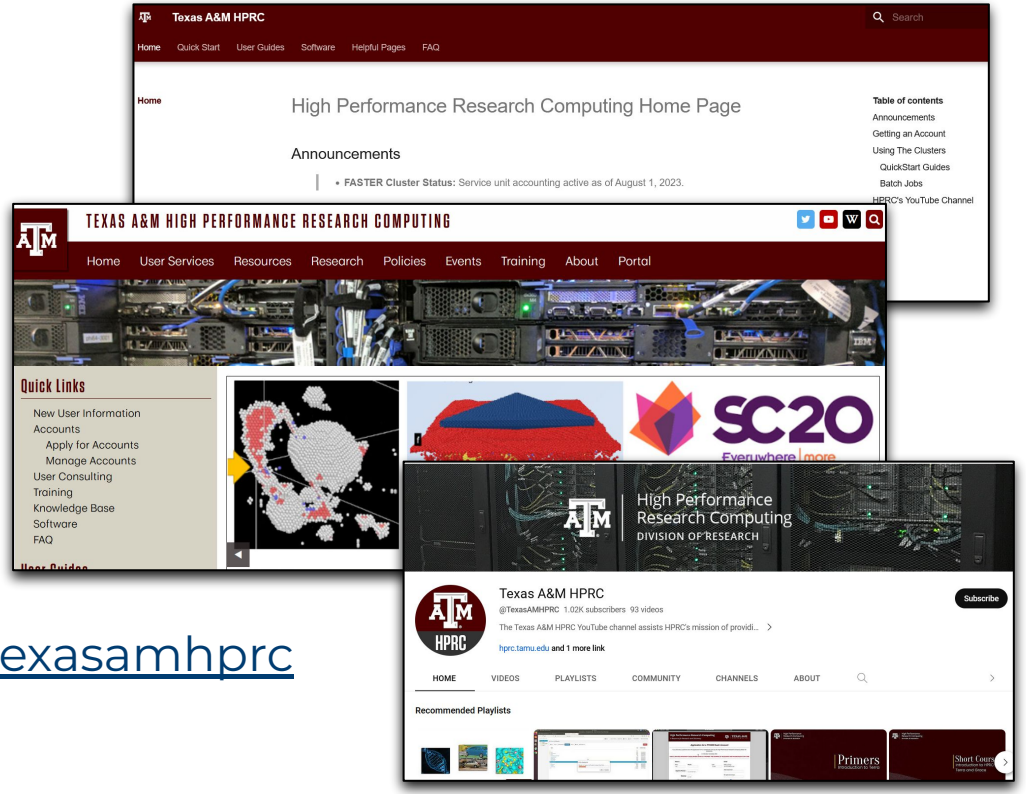
– https://hprc.tamu.edu/kb/

HPRC Website

– https://hprc.tamu.edu/

Youtube Channel

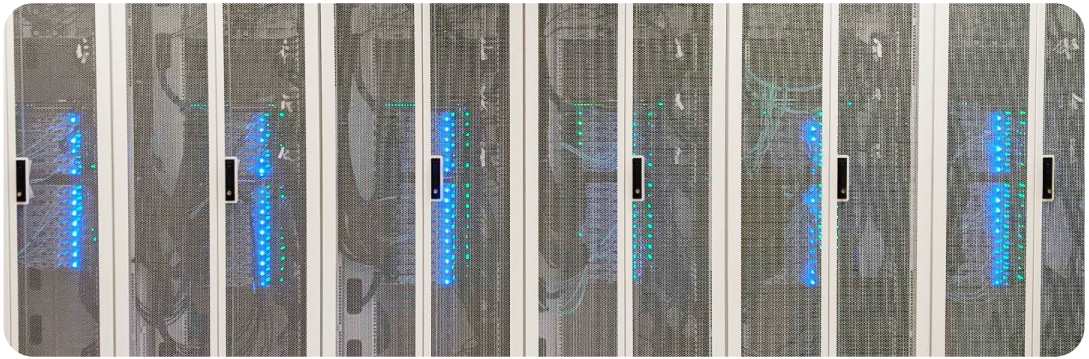
– https://www.youtube.com/texasamhprc



Computing Resources

The HPRC group currently administers four HPC clusters:

- ACES
- FASTER
- Grace
- Launch



You'll need one of two options to use them:

Credentials	Clusters	Who
HPRC Account	FASTER, Grace	Only Texas A&M students/staff
ACCESS ID	FASTER, ACES, Launch	Anybody

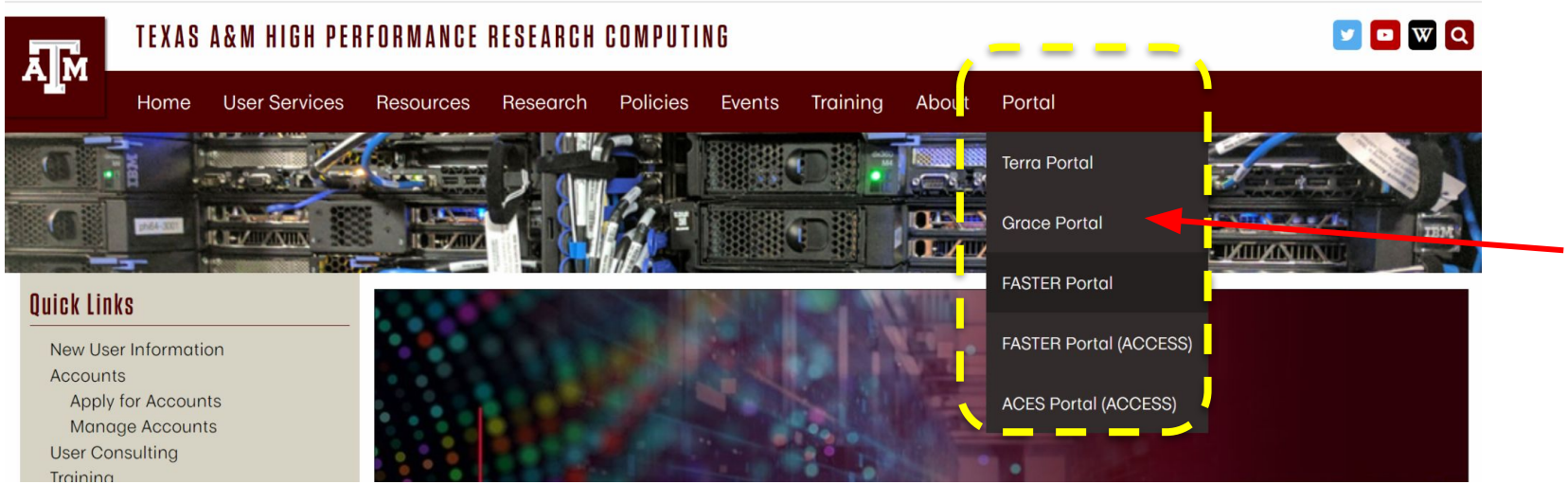
Link to our Knowledge Base: https://hprc.tamu.edu/kb/

Accessing Clusters via Portal

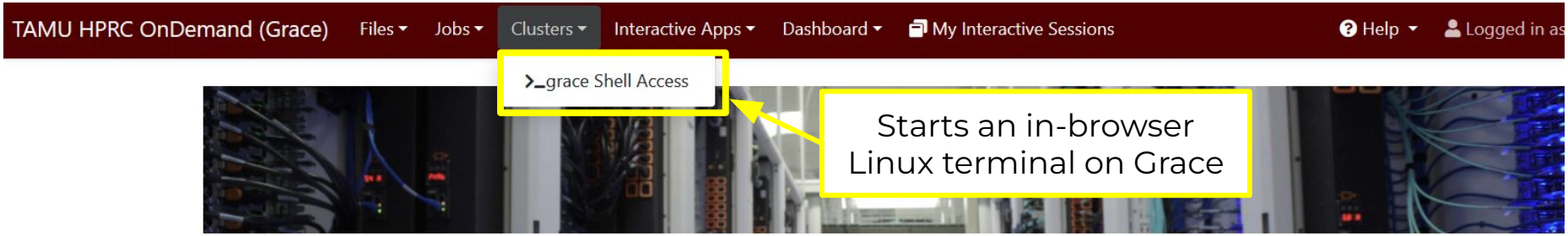
Access through web browser:

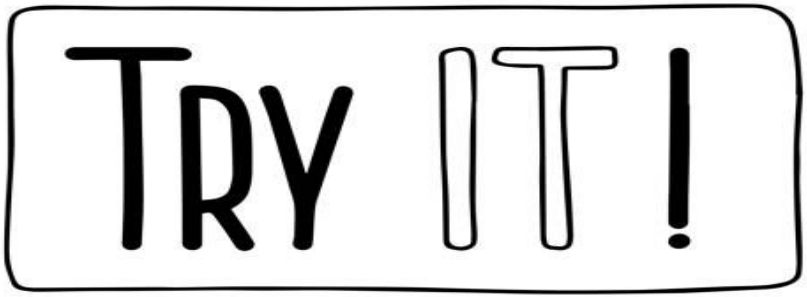
– portal.hprc.tamu.edu

– Or dropdown menu on HPRC homepage:



Using the Portal - Shell Access





Log into the Grace cluster using the online Portal

(If you're not on TAMU Wifi, then you should use TAMU VPN before trying to connect to our clusters)

Good Data Practice

Rule of thumb:

- **1 is none**
- **2 is one**

Keep multiple copies of important data!

Having just one copy is not enough

Backup Backup Backup

File Systems and User Directories

Directory	Environment Variable	Space Limit	File Limit	Intended Use
/home/$USER	$HOME	10 GB	10,000	Small to modest amounts of processing.
/scratch/user/$USER	$SCRATCH	1 TB	250,000	Temporary storage of large files for on-going computations. Not intended to be a long-term storage area.

`$SCRATCH` is shared between FASTER and Grace clusters.

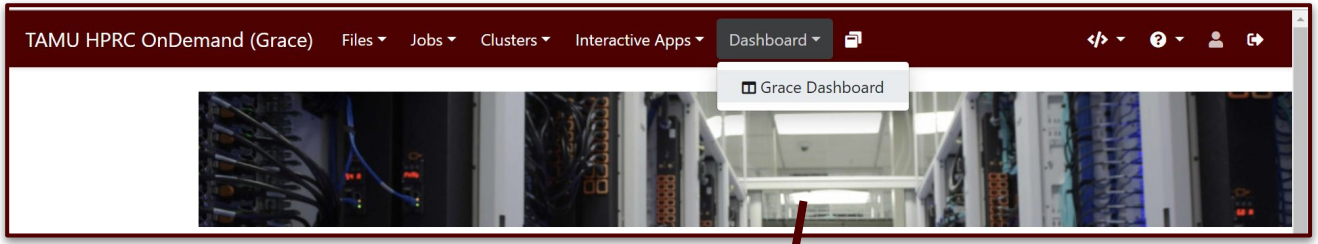
View file usage and quota limits using the command: **`showquota`**

Do NOT share your home or scratch directories. Request a group directory for sharing files.

https://hprc.tamu.edu/kb/User-Guides/FASTER/Filesystems_and_Files/

OOD Portal Quota Increase Request

Portal Homepage → [Cluster name] Dashboard

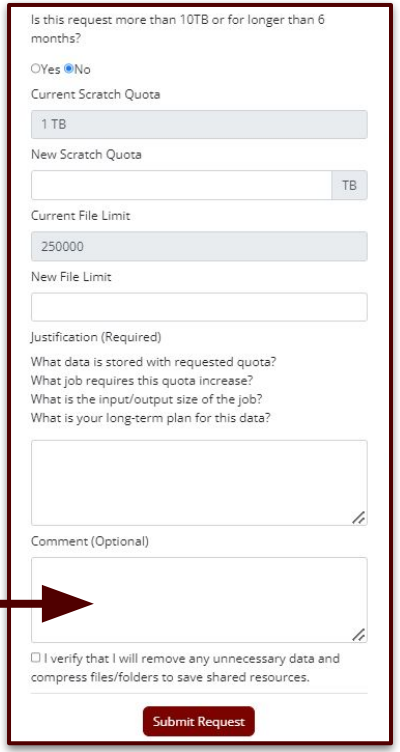


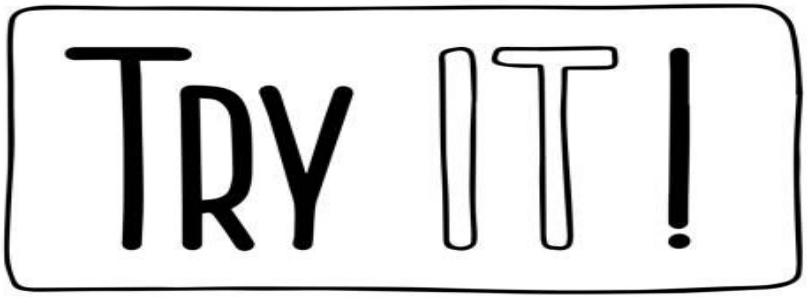
Request quota increases directly from the dashboard with a guided form

Contact help@hprc.tamu.edu if you need more help.

Disk Quotas

Disk	Disk Usage	Limit	File Usage	Limit	
home	318M (3.11 %)	10.0G	3,951 (39.51 %)	10,000	
scratch	679M (0.06 %)	1.0T	27,247 (10.90 %)	250,000	Request Quota Increase





1. Navigate to the Dashboard to see your quotas
2. Run the `showquota` command in a shell to see your quotas

Command Line Tools

```
cp    -- copy
rm    -- remove
tar   -- archiving
sftp  -- secure file transfer
scp   -- secure copy (remote copy)
rsync -- remote synchronization
```

Command Line Tools: cp

Copy

Makes a copy of a file:

```
cp source_file new_fileName
```

Easy solution for copying a file onto the *same machine*

To move data between machines, use `scp` or `sftp` (which we'll cover in a moment)

Command Line Tools: rm

Remove

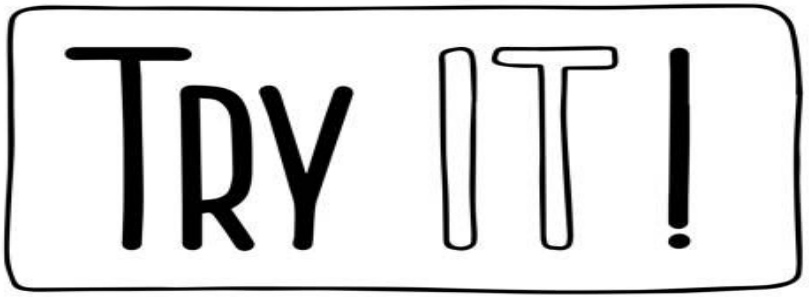
Deletes a file:

```
rm some_file
```

WARNING: *There is no "trash bin" on the command line!*
Once you `rm` an object, you cannot get it back!

Add the -i flag to be prompted prior to file deletion:

```
rm -i some_file
```



On your command line:

1. Create a test directory and make a file
2. Copy the file
3. Remove one copy

Command Line Tools: tar

Archiving files

saves many files together into a single file (archive):

```
tar -cvf  archive.tar source
```

create a compressed archive:

```
tar -czvf  archive.tar.gz source
```

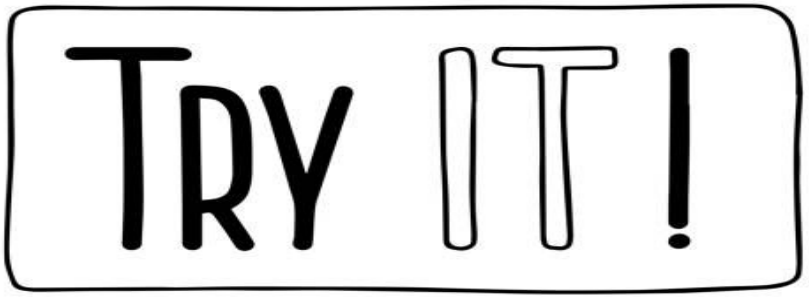
extract an archive:

```
tar -xvf  archive.tar.gz
```

Important flags

-cf	create archive
-xf	extract archive
-v	verbose
-z	compress with gzip

Useful for consolidating (and compressing) files prior to transfer



On your command line:

1. Create more copies of the test file you made before
2. Condense them all into one object
3. Extract them back out and check that their contents are the same

Data Transfer

Login nodes

- 60 minute process time limit

Data transfer nodes

- no process time limit

Command line tools **`scp`**, **`sftp`**, and **`rsync`** are all available

Globus is a preferred method for large data transfer (more about Globus later)

Command Line Tools: sftp

Secure file transfer protocol

interactive file transfer program

```
sftp NetID@grace.tamu.edu
```

Connects and logs into specified host, enters command mode

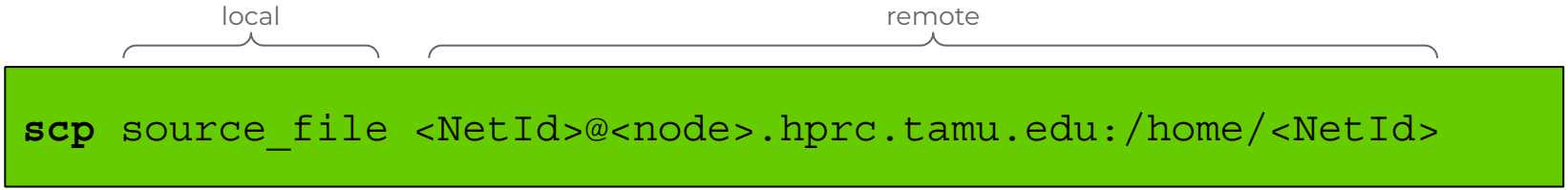
```
cd  - change directory
get - download file
put - upload file
bye - quit sftp
```

Not Recommended for large file transfers.

Command Line Tools: scp

Secure copy

Copies files between hosts on a network



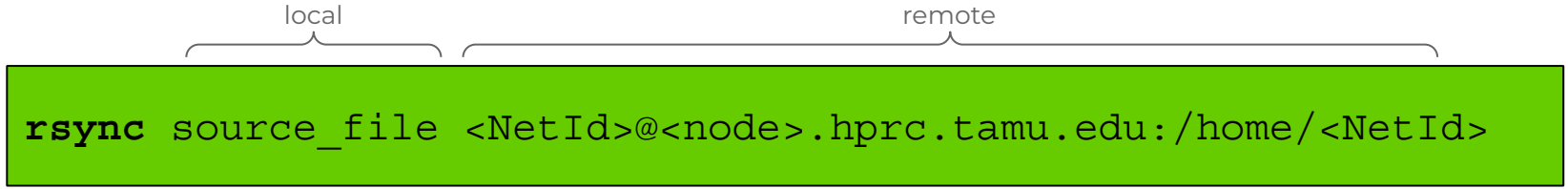
Can be used

- local to remote (as above)
- remote to local
- remote to remote

Command Line Tools: rsync

Remote Synchronize

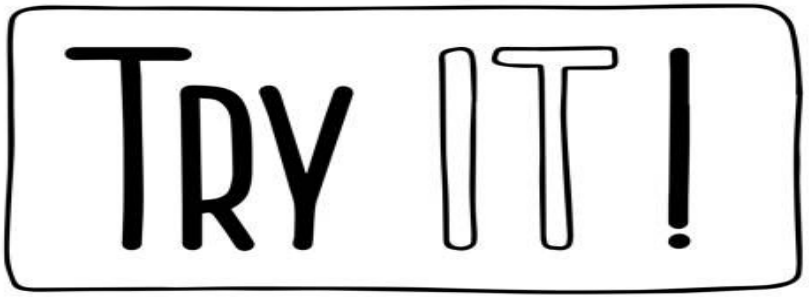
Copies files between hosts on a network



Can be used

- local to remote (as above)
- remote to local
- remote to remote

→ rsync is better than scp
→ rsync supports *intermittent* transfer



1. use `rsync` to copy a `.tar` file to `faster`
2. verify the file arrived (on FASTER login node)

Graphical User Interface (GUI) Clients

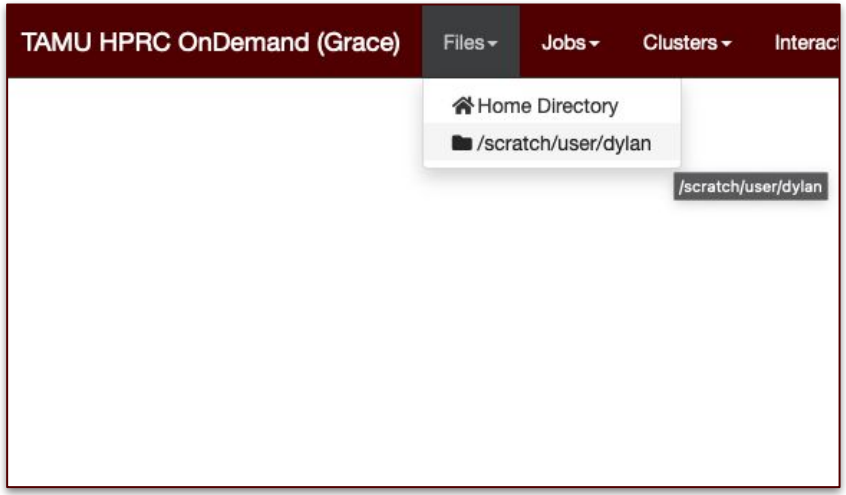
There are many GUI solutions for file transfer:

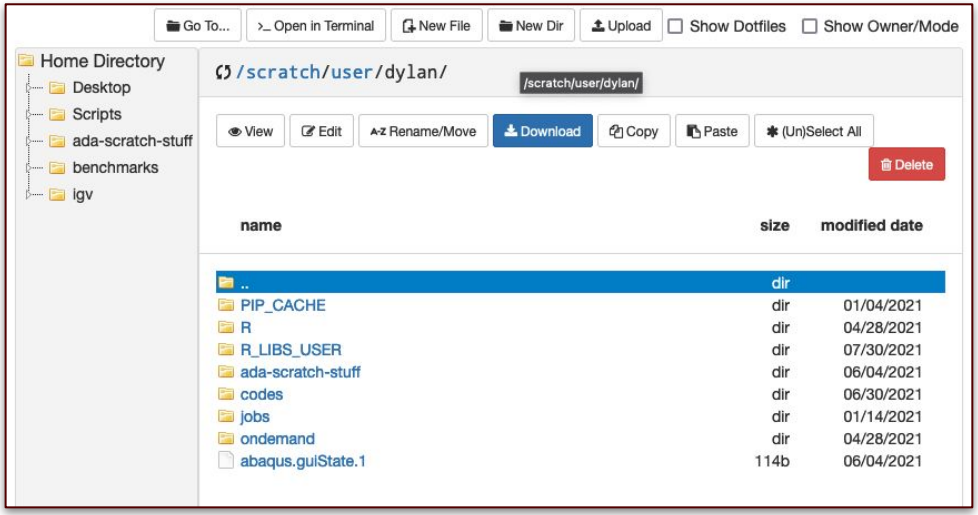
- Open OnDemand Portal
- MobaXterm
- WinSCP
- Cyberduck
- Globus Connect

GUI Clients: HPRC Portal

Access your files through a web browser

View, Edit, Upload, Download, and Remove through the Portal





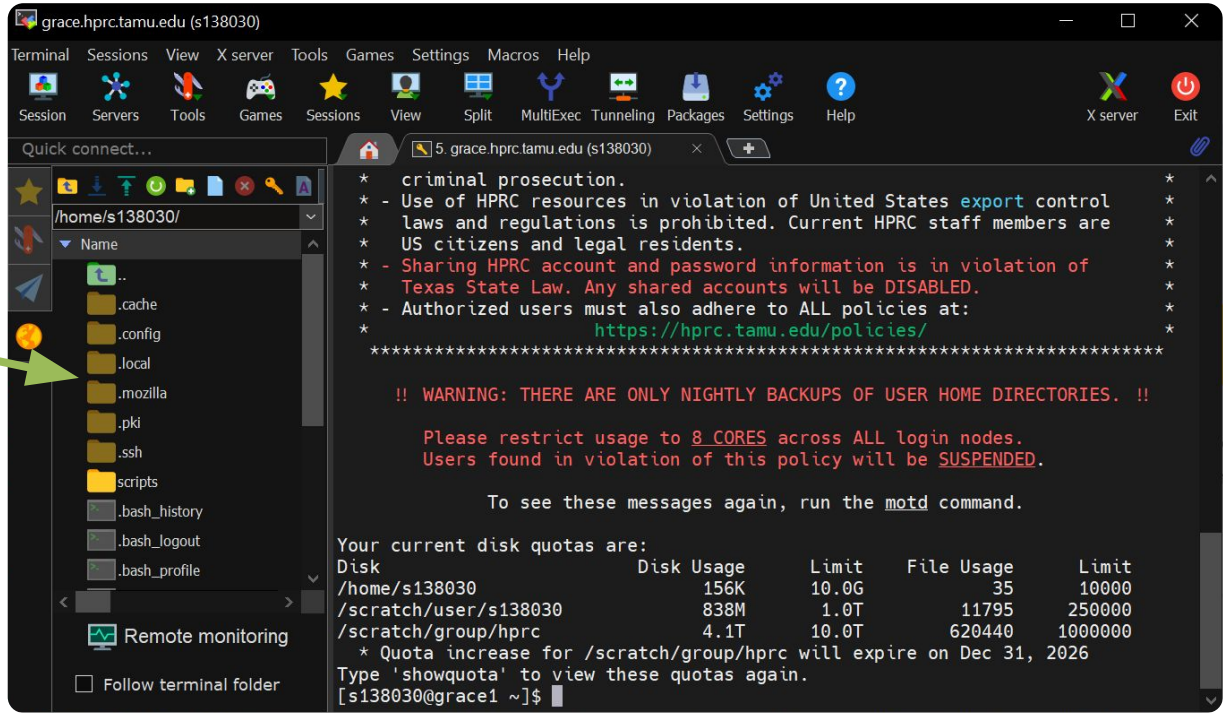
https://portal.hprc.tamu.edu

GUI Clients: MobaXterm

Available on Windows machines

SFTP side panel in MobaXterm

Can download, upload files with a few clicks from the CLI



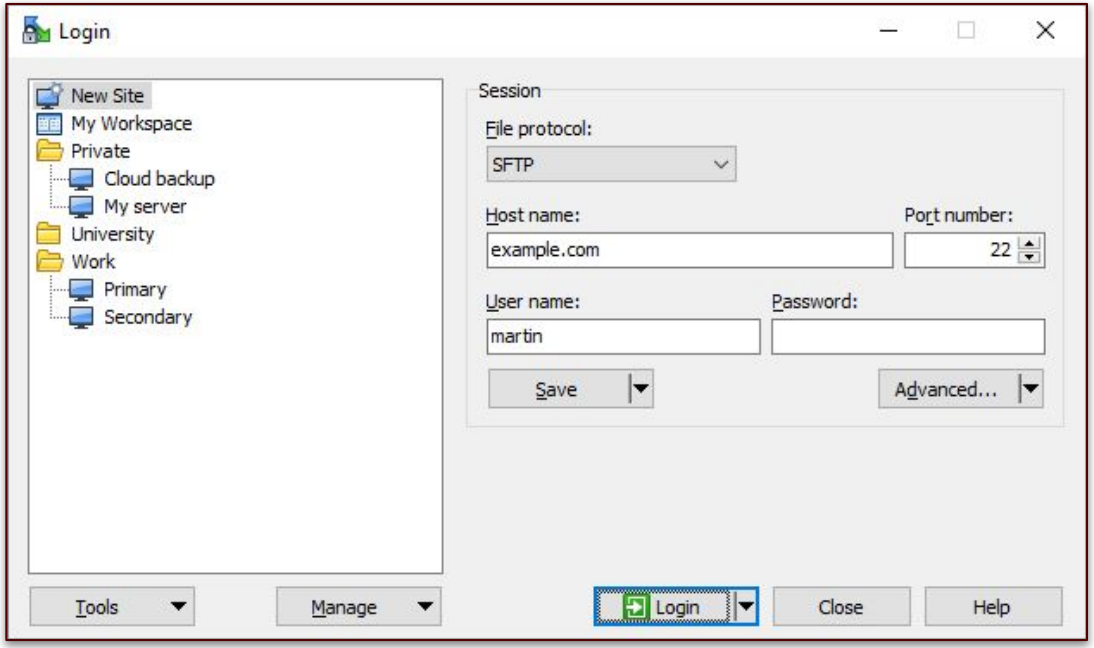
https://hprc.tamu.edu/kb/Helpful-Pages/#mobaxterm-recommended

GUI Clients: WinSCP

Available on Windows machines

Connects to host directly with SFTP

Allows for transfers through the GUI



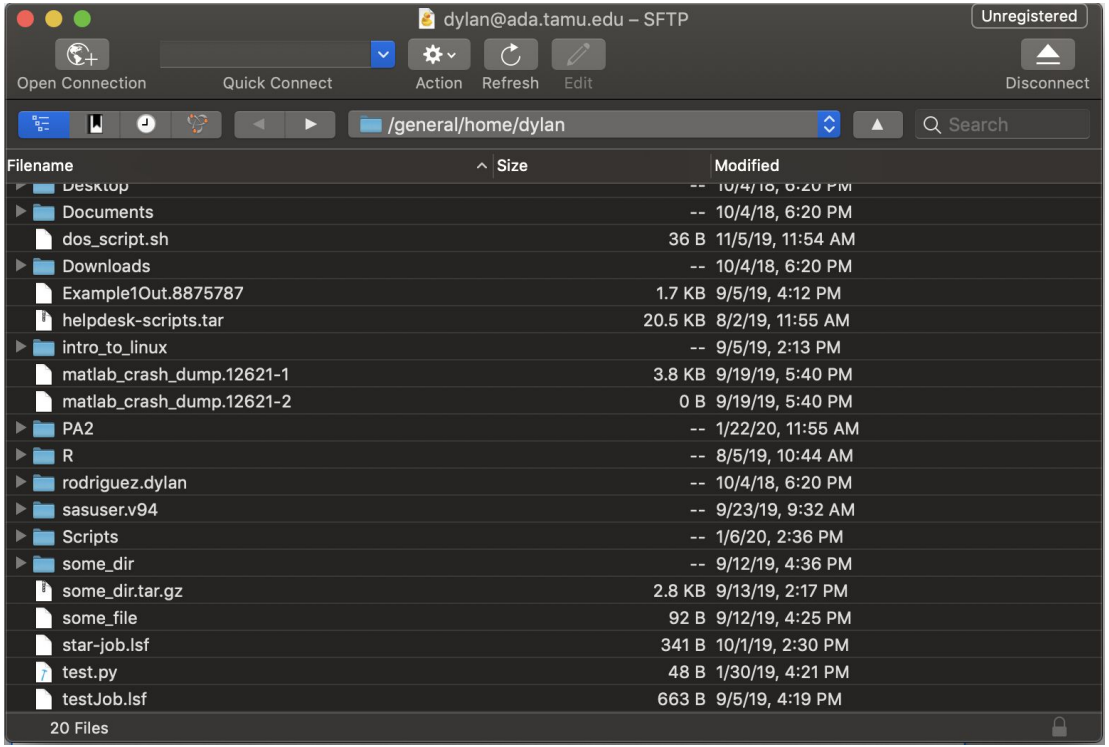
https://hprc.tamu.edu/kb/Helpful-Pages/File-Transfer/#tutorial-videos

GUI Clients: CyberDuck

Available on Windows & MacOS

Connects to host directly with SFTP

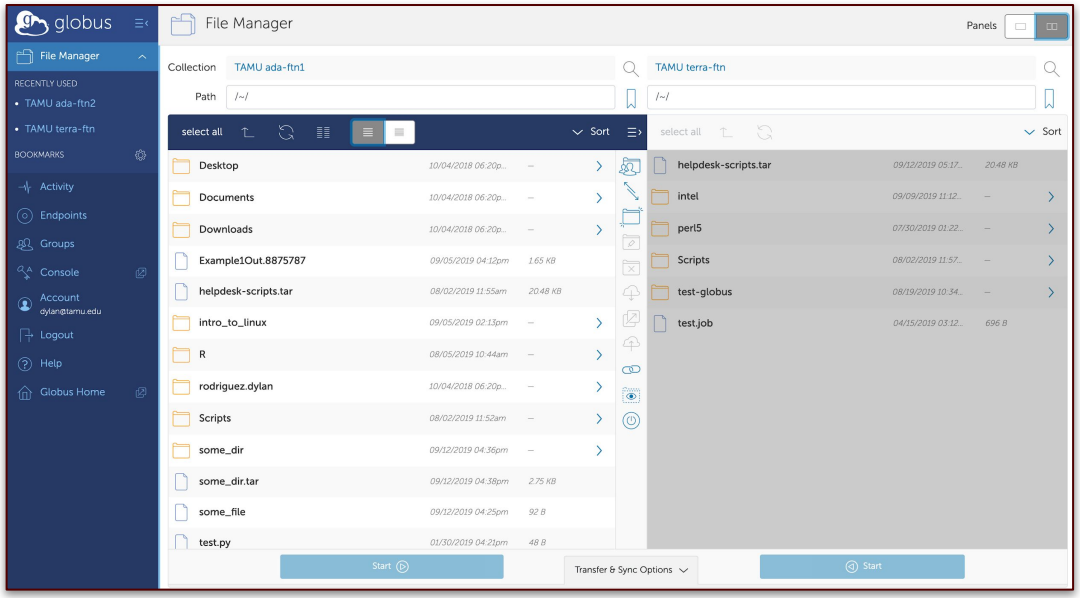
Allows for transfers through the GUI



https://hprc.tamu.edu/kb/Helpful-Pages/File-Transfer/#tutorial-videos

GUI Clients: Globus

Available on
Windows, MacOS, Linux &
through a web portal.



Grace endpoints:

- grace-dtn1
- grace-dtn2

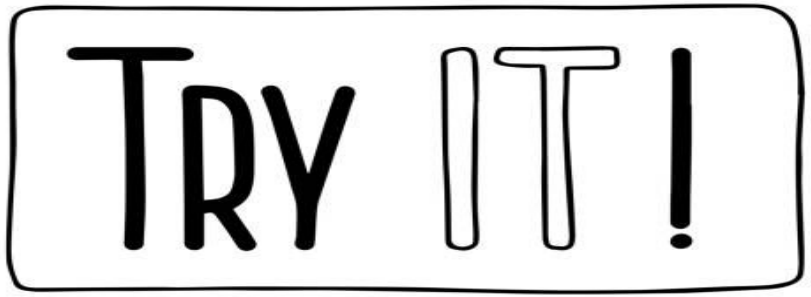
ACES endpoint:

- ACCESS TAMU ACES DTN

FASTER endpoints:

- TAMU FASTER DTN1
- ACCESS faster.tamu.access-ci.org (ACCESS users)

https://www.globus.org/
https://hprc.tamu.edu/kb/Software/Globus/

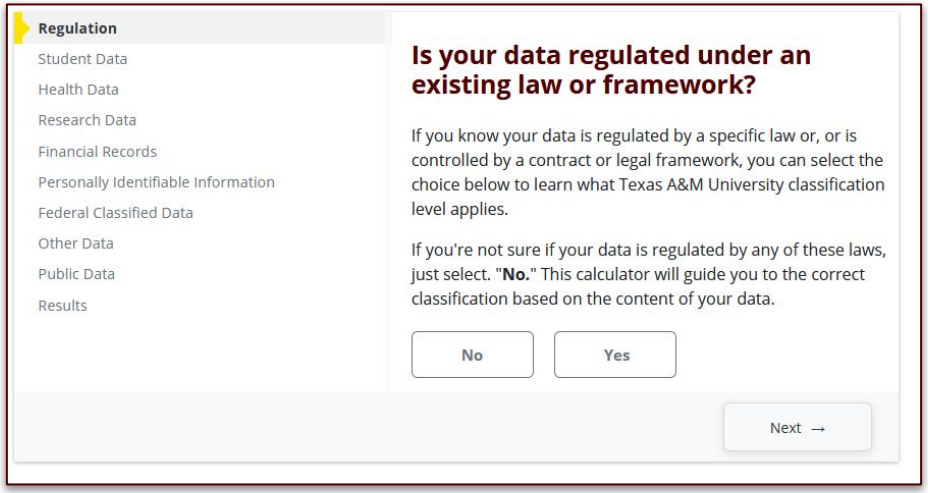


Transfer files between FASTER and Grace:

1. Create some test file in your home directory on FASTER or Grace
2. Log in to https://app.globus.org/
3. Search for “TAMU” in the *Collection* fields and select “**TAMU grace-dtn**” and “**TAMU FASTER DTN1**”
   Globus should show you the contents of your home directories
4. Transfer your test file from one cluster to the other

Data Classification Tool

The process of sorting and categorizing data based on the sensitivity of information and the impact of potential loss



https://it.tamu.edu/community/tools/data-classification.php

Continued Learning

Intro to HPRC Video Tutorial Series

HPRC's Knowledge Base

Need Help?

First check the FAQ

- Knowledge Base
- Send us a ticket using the dashboard tab on our web portal
- Email further questions to help@hprc.tamu.edu

Help us help you -- when you contact us, tell us:

- Which cluster you're using
- Your username
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used, if any
- Module(s) loaded, if any
- Error messages
- Steps you have taken, so we can reproduce the problem

High Performance
Research Computing
DIVISION OF RESEARCH

Thank you.

Any questions?

Please let us know what you thought of this course by filling out this survey:
https://u.tamu.edu/hprc_shortcourse_survey

Data Transfer: Grace DTN

Grace has 2 nodes dedicated to data transfer → Data Transfer Nodes

Large transfers should use the Data Transfer Nodes

Both nodes have **40 gigabit capability**

No programming environment installed → these are for transfers only!

These nodes have access to all of Grace's filesystem (/home and /scratch)

Data Transfer: FASTER and ACES

The FASTER and ACES cluster each have two Data Transfer Nodes–one of which is dedicated to ACCESS users.

ssh to Data Transfer Node not available on ACES.

These can be accessed through Globus.

https://hprc.tamu.edu/kb/User-Guides/FASTER/#data-transfer-nodes
https://hprc.tamu.edu/kb/User-Guides/ACES/#data-transfer-nodes

Data on Our Clusters: Grace and Terra

There are limits on data on our clusters → AKA quota

The limits are on *disk space* & *file usage*

`showquota` View your current quota with this command

```
Your current disk quotas are:
Disk    Disk Usage          Limit       File Usage  Limit
/home       416.1M          10G         4489        10000
/scratch    18.64G          1T          122616      250000
```

Need more space?

Submit a ***Quota Increase Request***

Data on Our Clusters: Grace and Terra

What's the difference between these filesystems?

/home

- high performance storage
- will not be expanded
- **backed up**

/scratch

- high performance storage
- can be expanded
- **not backed up**

Need more space?
Submit a ***Quota Increase Request***

Contact help@hprc.tamu.edu

Accessing Clusters via SSH

- SSH (Secure SHell) allows users to establish a connection between their local machine and the TAMU HPRC clusters.
- SSH Programs

Operating System	Windows	MacOS	Linux
Programs	MobaXTerm* PuTTY SSH Windows Subsystem for Linux (WSL) Windows Command Prompt	Terminal*	Terminal*
* Recommended			

- SSH Command

```
ssh [NetID]@faster.hprc.tamu.edu
```

- TAMU (on-campus: Duo 2FA; off-campus: VPN and Duo 2FA)
- ACCESS (ssh not available)
  https://hprc.tamu.edu/kb/User-Guides/FASTER/ACCESS-CI/