HPRC Primers
Data Management Practices

January 27, 2023

High Performance
Research Computing
DIVISION OF RESEARCH

HPRC's Newest Cluster

FASTER (Fostering Accelerated Sciences Transformation Education and Research) is a 180-node Intel cluster from Dell with an InfiniBand HDR-100 interconnect and Liqid PCIe Gen4 for composing the GPUs. Nvidia A100, A10, A30, A40 and T4 GPUs are available. The 180 compute nodes are based on the Intel Ice Lake processor.

	Quantity
Login nodes	4
Compute nodes	180
Cores/Node	64
Memory	256 GB
Local Disk Space	3.84 TB NVMe (/tmp)
1TB Large Memory Nodes	2

For more information:
https://hprc.tamu.edu/wiki/FASTER:Intro

Logging in to the system

SSH (secure shell)
–freely available for Linux/Unix and Mac OS X hosts

For Microsoft Windows PCs, use *MobaXterm*
– or *Putty*

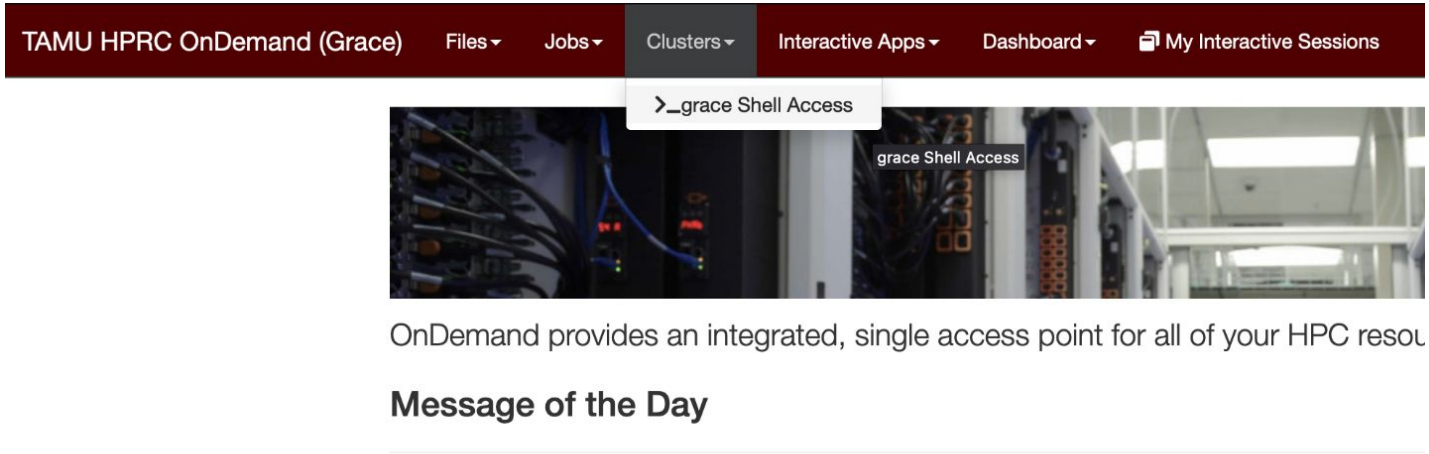
VPN needed for off campus access
– https://u.tamu.edu/VPN_help

Logging in via the OOD Portal

Access through (most) web browsers

–portal.hprc.tamu.edu

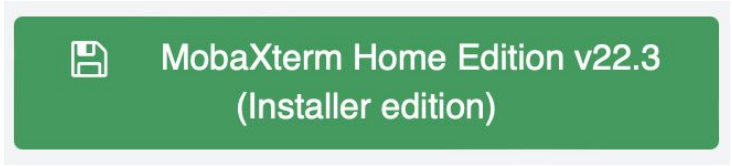
–Top Banner Menu “Clusters” -> “Shell Access”



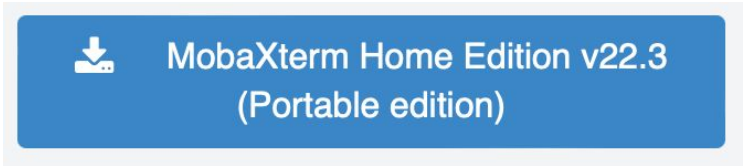
Windows Users:

To follow along, please download Mobaxterm
(a free SSH client)
Google Search “Mobaxterm” or navigate to:
https://mobaxterm.mobatek.net/download.html

Download the “Installer edition” (green button)



or "portable edition" if you do not have admin privileges to install



Using SSH (on a Linux/Unix Client)

https://hprc.tamu.edu/wiki/Grace:Access

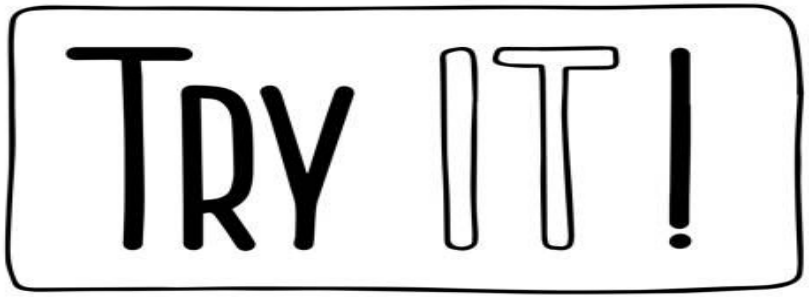
```
ssh NetID@grace.hprc.tamu.edu
```

You may see something like the following the first time you connect to the remote machine from your local machine:

```
Host key not found from the list of known hosts.
Are you sure you want to continue connecting (yes/no)?
```

Type yes, hit enter and you will then see the following:

```
Host 'grace.tamu.edu' added to the list of known hosts.
user_NetID@grace.tamu.edu's password:
```


TRY IT!

Good Data Practice

Rule of thumb:

1 is none

2 is one

Keep multiple copies of importar

Having just one copy is not enough

Backup Backup Backup

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
```

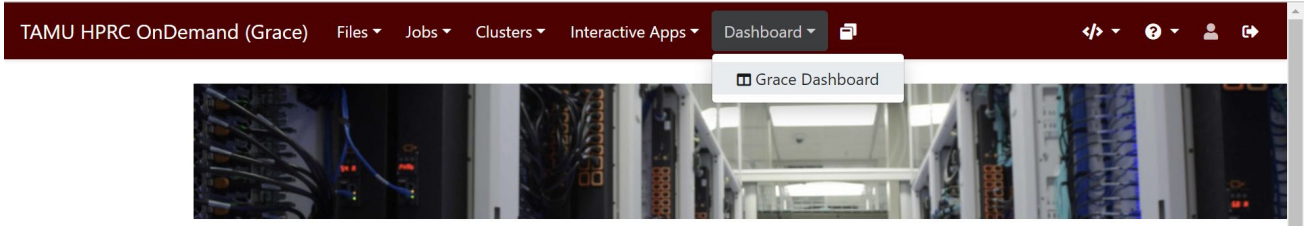
Need more space?

Submit a ***Quota Increase Request***

Contact help@hprc.tamu.edu

OOD Portal Quota Increase Request

Grace Portal Homepage → Grace Dashboard

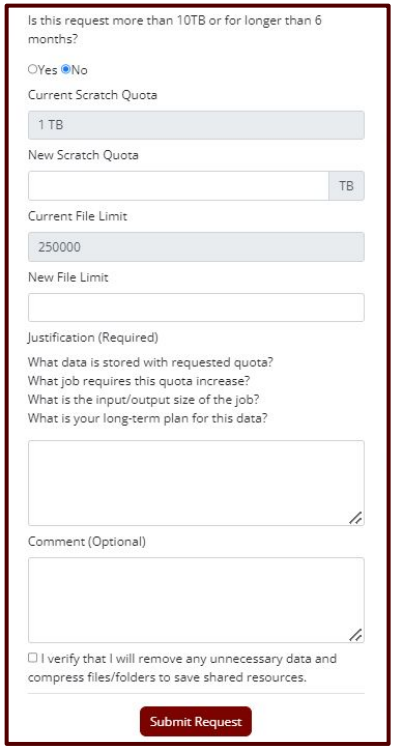


Request quota increases directly from the dashboard with a guided form

Disk Quotas

Disk	Disk Usage	Limit	File Usage	Limit
/home	160 KB (0.00 %)	10 GB	35 (0.35 %)	10000
/scratch	837.84 MB (0.08 %)	1 TB	11795 (4.72 %)	250000

Request Quota Increase



Data on Our Clusters: Grace and Terra

Default Limits

```
/home            10G   /   10,000 files
/scratch         1T    / 250,000 files
```

Data on Our Clusters: Grace and Terra

What's the difference between these filesystems?

/home	/scratch
high performance storage will not be expanded **backed up**	high performance storage can be expanded **not backed up**

Need more space?
Submit a *Quota Increase Request*

Contact help@hprc.tamu.edu

Data Transfer: Grace and Terra

Grace's login nodes have 10 Gigabit Ethernet to the TAMU Network

scp - sftp - *rsync* are all available

Login nodes have a 60 minute process limit

rsync is preferred → supports intermittent transfer

```
rsync source_file destination
```

GUI transfer programs are easiest for new users

WinSCP, MobaXterm, HPRC Portal

Data Transfer: Grace DTN

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

SSH to either DTN:

```
ssh NetID@grace-dtn1.tamu.edu
```

```
ssh NetID@grace-dtn2.tamu.edu
```

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
- /scratch

Data Transfer: Terra FTN

Terra has 1 node dedicated to data transfer → Fast Transfer Node
No process time limit

SSH to Terra's FTN:

```
ssh NetID@terra-ftn.hprc.tamu.edu
```

Large Transfers should use the Data Transfer Nodes
The node has 10 gigabit capability
No programming environment installed → these are for transfers only!
These nodes have access to all of Terra's filesystem

- /home
- /scratch

Command Line Tools

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

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*

How about moving data between machines?

Command Line Tools: rm

Remove

Deletes a file

```
rm some_file
```

Completely deletes a file

There is no “trash bin” on the command line
add the -i flag to be prompted prior to file deletion

```
rm -i some_file
```

Command Line Tools: scp

Secure copy

Copies files between hosts on a network -- uses ssh for data transfer (hence “Secure”)

```
scp source_file NetId@grace.tamu.edu:/home/NetID
```

Can be used to copy files:

from local to remote

from remote to local

between 2 remote systems from local system

Command Line Tools: sftp

Secure file transfer protocol

interactive file transfer program -- uses ssh (again so hence “secure”)

```
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
```

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
```

Useful for consolidating (and compressing) files prior to transfer

Important flags

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

Graphical User Interface (GUI) Clients

There are many GUI solutions for file transfer:

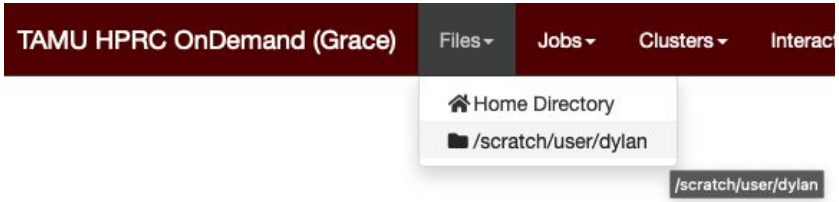
- MobaXterm
- Open OnDemand Portal
- Globus Connect
- WinSCP
- Cyberduck

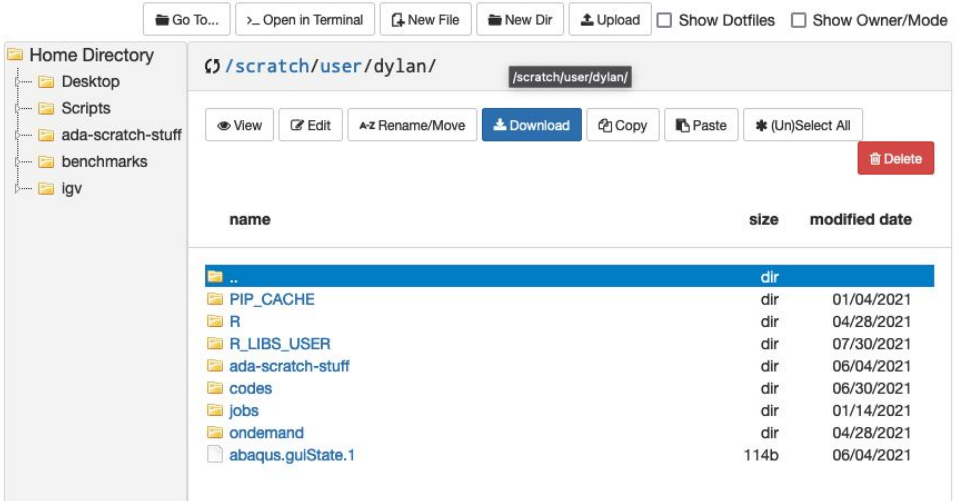
GUI Clients: HPRC Portal

Access your files through (almost) any web browser

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

https://portal.hprc.tamu.edu



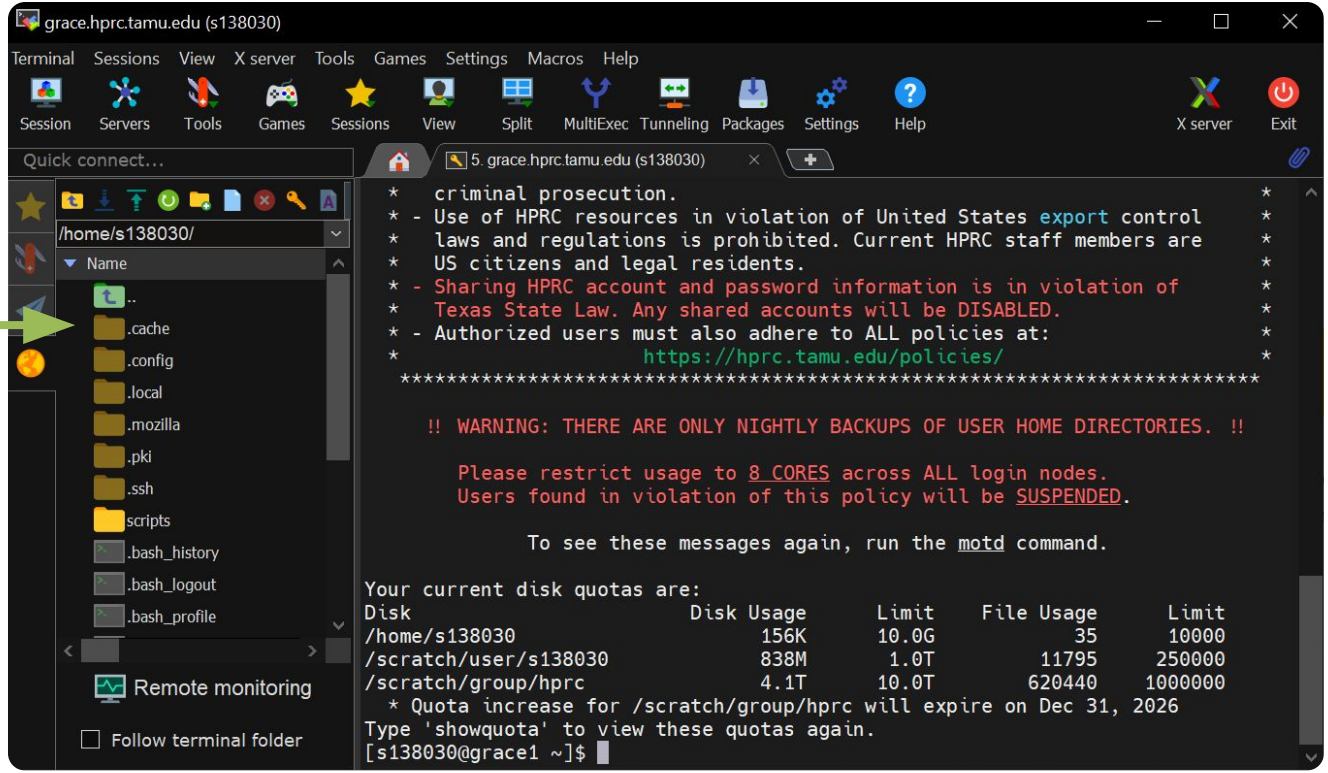


GUI Clients: MobaxTerm

Available on Windows machines

SFTP side panel in MobaxTerm

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

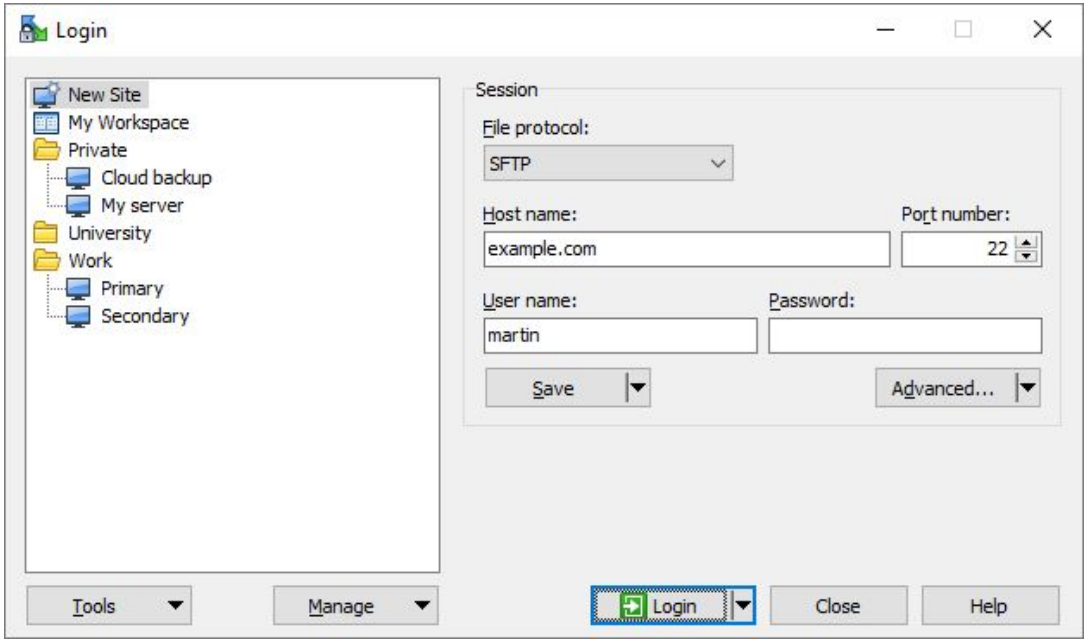


GUI Clients: WinSCP

Available on Windows machines

Connects to host directly with SFTP

Allows for transfers through the GUI

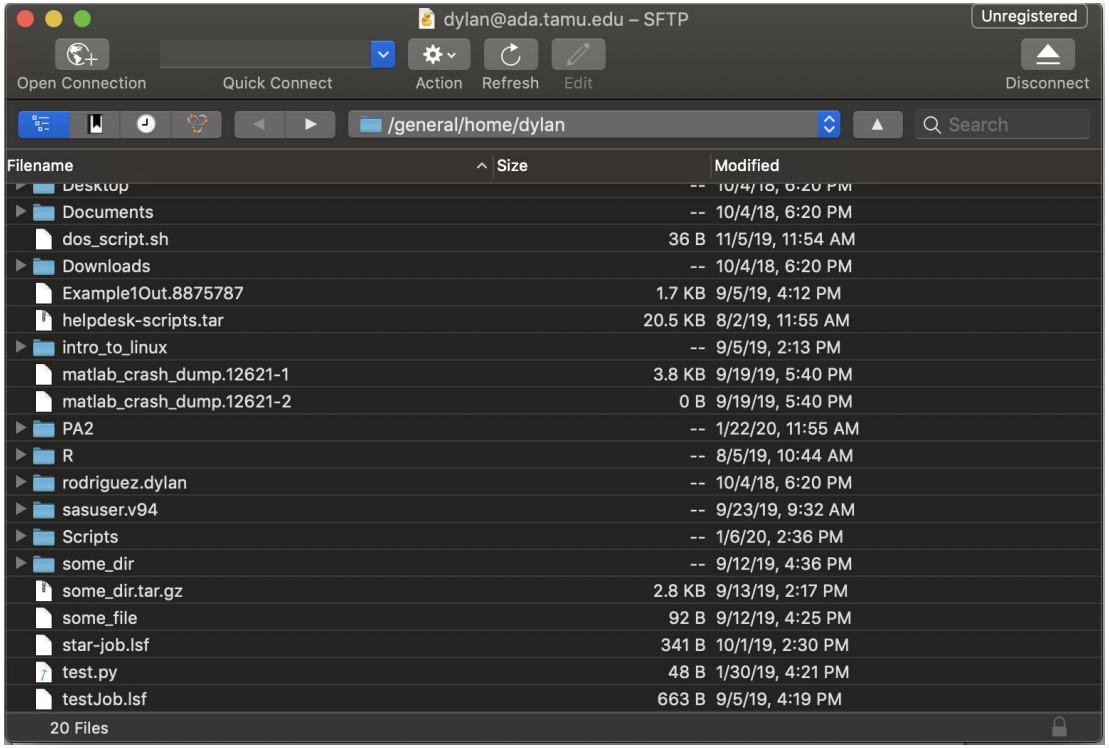


GUI Clients: CyberDuck

Available on Windows & MacOS

Connects to host directly with SFTP

Allows for transfers through the GUI

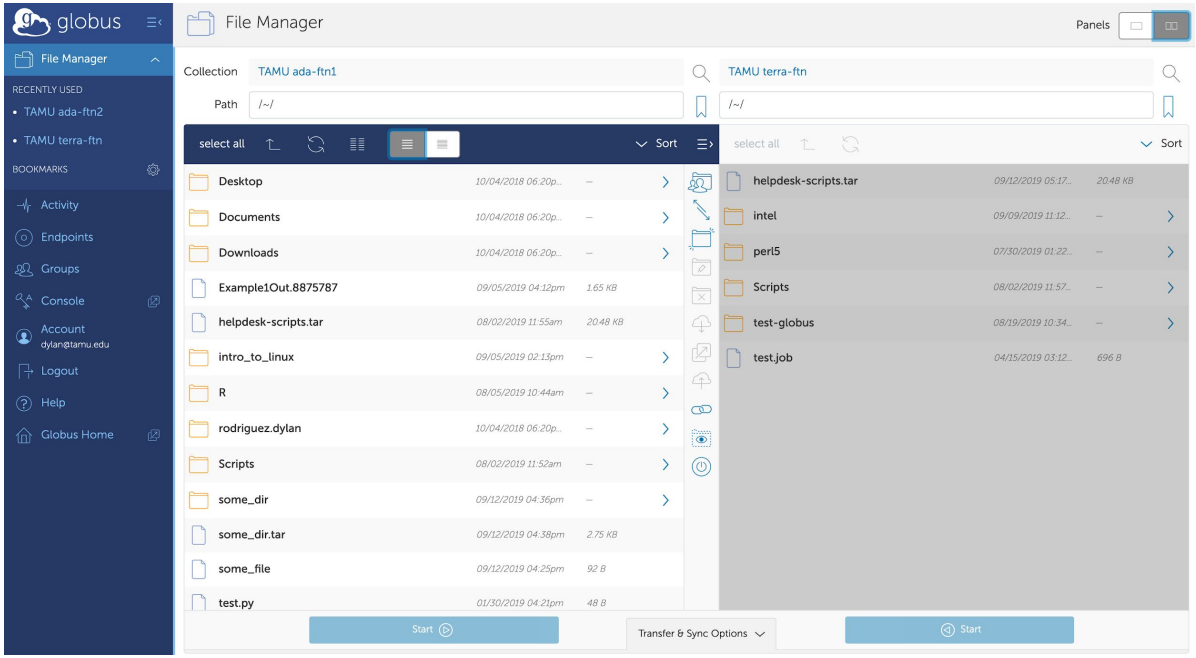


GUI Clients: Globus

Web based, with application you can download

Grace endpoints:
grace-dtn1
grace-dtn2

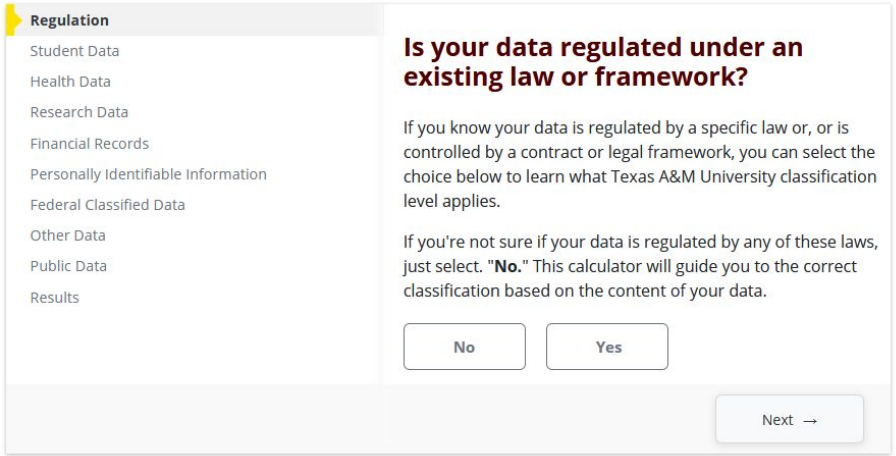
Terra endpoint
terra-ftn



https://www.globus.org/

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 Wiki Page

Need Help?

First check the FAQ, wiki, and YouTube channel

- Email your questions to help@hprc.tamu.edu

Help us, help you -- we need more info

- Which Cluster
- 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