High Performance Research Computing

A Resource for Research and Discovery



Data Management Practices Primer

Windows Users:

To follow along, please download Mobaxterm (a free SSH client)

Google Search "Mobaxterm" or navigate to: <u>https://mobaxterm.mobatek.net/download.html</u>

Download the "Installer edition" (green button)

MobaXterm Home Edition v20.2 (Installer edition)



ĂМ

HPRC's Newest Cluster

Grace is a 925-node Intel cluster from Dell with an InfiniBand HDR-100 interconnect, A100 GPUs, RTX 6000 GPUs and T4 GPUs. There are 925 nodes based on the Intel Cascade Lake processor.

> Grace 3TB Large Memory-80 cores/nodes Other Login Nodes-48 cores/node



Login Nodes	5
384GB memory general compute nodes	800
GPU - A100 nodes with 384GB memory	100
GPU - RTX 6000 nodes with 384GB memory	9
GPU - T4 nodes with 384GB memory	8
3TB Large Memory	8

For more information: https://hprc.tamu.edu/wiki/Grace: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
 - <u>https://u.tamu.edu/VPN_help</u>

Using SSH (on a Linux/Unix Client) https://hprc.tamu.edu/wiki/Grace:Access

ssh user NetID@grace.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:

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

Ă Ň

Data on Our Clusters: Grace

There are limits on data on our clusters \rightarrow AKA quota The limits are on *Disk Space* & *File Usage*

showquota

View your current quota with this command

Your curre	ent disk quotas a	re:			
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

AN

Data on Our Clusters: Grace

Default Limits

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

Data on Our Clusters: Grace

What's the difference between these filesystems?



Need more space? Submit a *Quota Increase Request*

Contact help@hprc.tamu.edu

Data Transfer: Grace

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 \rightarrow supports intermittent transfer

rsync source file destination

GUI transfer programs are easiest for new users WIN SCP, MobaXterm, HPRC Web Portal

Data Transfer: Grace DTN

Grace has 2 nodes dedicated to data transfer → Data Transfer Nodes Accessible via Grace's login nodes

SSH to either DTN from any Grace login node:

ssh netID@grace-dtn1.tamu.edu ssh netID@grace-dtn2.tamu.edu

Large Transfers should use the Data Transfer Nodes Both nodes have <u>40 gigabit capability</u> No programming environment installed → these are for transfers only! These nodes have access to all of Grace's filesystem /home /scratch

Data on Our Clusters: Terra

There are limits on data on our clusters \rightarrow AKA quota The limits are on *Disk Space* & *File Usage*

sho	wquota	View you	ur current quota w	ith this command	
Your curre	ent disk quotas a	re:			
Disk	Disk Usage	Limit	File Usage	Limit	
/home	416.1M	10G	4489	10000	
/scratch	18.64G	1T	122616	250000	

Data on Our Clusters: Terra

Default Limits

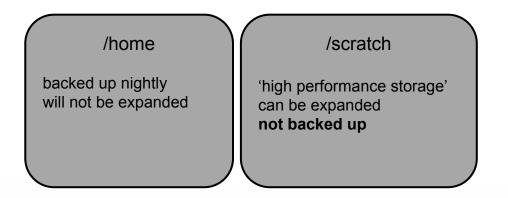
home	10G / 10,000 files
scratch	1TB / 250,000

Need more space? Submit a *Quota Increase Request*

Contact help@hprc.tamu.edu

Data on Our Clusters: Terra

What's the difference between these filesystems?



Data Transfer: Terra

Terra'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 \rightarrow supports intermittent transfer

rsync source file destination

GUI transfer programs are easiest for new users WIN SCP, MobaXterm, HPRC Web Portal

Data Transfer: Terra FTN

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

SSH to Terra's FTN directly:

ssh netID@terra-ftn.hprc.tamu.edu

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

Pop Quiz

What is the process limit for the login node?

- A. 45 minutes
- B. 75 minutes
- C. 60 minutes
- D. No process limit

Pop Quiz

What is the process limit for the login node?

- A. 45 minutes
- B. 75 minutes
- C. 60 minutes
- D. No process limit

Command Line Tools

- 1. cp -- copy
- 2. rm -- remove
- 3. scp -- secure copy (remote copy)
- 4. sftp -- secure file transfer
- 5. 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

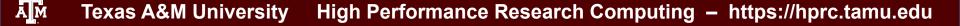
Deletes a file

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:

- from local to remote
- from remote to local
- between 2 remote systems from local system

Command Line Tools: sftp

Secure file transfer program

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

GUI Clients

There are many GUI solutions for file transfer:

- 1. MobaXterm
- 2. WinSCP
- 3. FileZilla
- 4. Cyberduck

Globus Connect





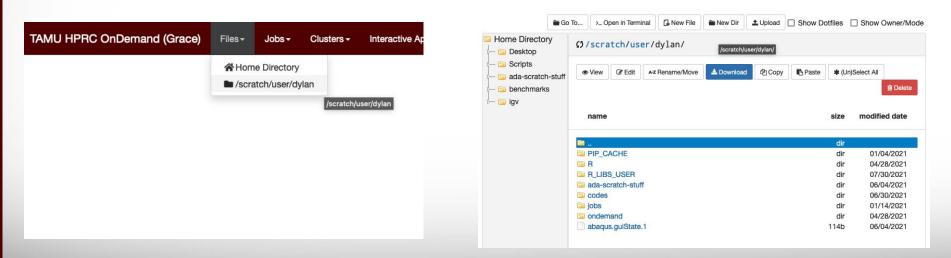


GUI Clients: HPRC Portal

Access your files through (almost) any web browser

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

https://portal.hprc.tamu.edu



GUI Clients: MobaxTerm

whomps@login5:~ X Available on Windows machines Terminal Sessions View X server Tools Games Settings Macros Help Ċ 2 X Solit MultiExec Tunneling Settings Help Exit Session Servers Tools Games Sessions View X server Ouick connect... / 📮 2. whomps@login5:~ 0 🚖 🖸 💕 🗋 💥 🥄 💻 🗛 🛅 Texas A&M University High Performance Research Computing /general/home/whomps/ s/ Website: http://hprc.tamu.edu Size (KB) Last 1 ^ Name Consulting: help@hprc.tamu.edu or (979) 845-0219 Ada Documentation: https://hprc.tamu.edu/wiki/index.php/Ada aienv fea2.015.0 cache 2015 SFTP side panel in aieny fea2.017.1 cache 2016 Altair 2015 === IMPORTANT POLICY INFORMATION === MobaxTerm .altair 2015 * -Unauthorized use of HPRC resources is prohibited and subject to .altair_licensing 2015 criminal prosecution. * -Use of HPRC resources in violation of United States export control laws 2016 .ansys and regulations is prohibited. Current HPRC staff members are US .cache 2016 US citizens and legal residents. .config 2016 * -Sharing HPRC account and password information is in violation of State .dbus 2015 Law. Any shared accounts will be DISABLED. 2017 .fontconfig * -Authorized users must also adhere to all policies at: Can download, upload 2017 .aconf /wiki/index.php/HPRC:Policies 2017 .aconfd files with a few clicks .anome2 2016 .gnome2 private 2015 !! WARNING: There are NO active backups of user data. !! .gvfs 2015 from the CLI Please restrict usage to <u>8 CORES</u> across ALL Ada login nodes. .intel 2015 Users found in violation of this policy will be SUSPENDED. .ipython 2016 📓 .java 2015 **** Ada Scheduled Maintenance Completed **** Imod.d 2016 he maintenance for Ada has been completed. Batch job scheduling has resumed. Iocal 2015 Isbatch 2017 Your current disk quotas are: matlab 2016 Disk Disk Usage Limit File Usage Limit 117.2M 10G 1419 10000 /home .mozilla 2015 /scratch 6 804G 303 250000 2016 \ .mw 10T 50000 /tiered ype 'showquota' to view these quotas again. Follow terminal folder [whomps@ada5 ~]\$

UNREGISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: http://mobaxterm.mobatek.net

GUI Clients: WinSCP

Available on Windows machines

Connects to host directly with SFTP

Allows for transfers through the GUI

Ă Ň

Yew Site My Workspace	Session Eile protocol:			
Private Cloud backup	SFTP V			
My server	Host name:	Port number:		
University	example.com	22 🚔		
Primary Secondary	User name: Pas	sword:		
	martin			
	<u>S</u> ave ▼	A <u>d</u> vanced		

GUI Clients: FileZilla

Open source - available on all platforms

2 Factor Authentication makes FileZilla hard to use

Connects to host directly with SFTP

Allows for transfers through the GUI

Ă Ň

Status: Connecting to ada.ta Status: Connected to ada.tar Status: Retrieving directory I Status: Listing directory /gen Status: Directory listing of "/gen	nu.edu isting						
	neral/home/dylan general/home/dylan" su	uccessful					
Local site: /Users/dylan/				Remote site: /general/l	nome/dylan		
Guest Guest Gyun Gyun Volumes bin cores dev			1	V ? general V ? home b dylan			
Filename 🔨		Filesize Filetype	Last m	Filename 🔨	Filesize Filetype	Last modified Permissions	Owner/Group
 Trash Jash_sessions Cache Config Cups Jpython Jimstransporter Jupyter Jupyter Jupyter Jocal matplotlib .ngrok2 .ngrok2 .ngrom .oracle_jre_usage .putty		Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	05/22 06/26 06/26 02/27, 06/26 03/21, 02/27, 02/27, 03/15, 02/26	 .cache .cache .cantes .cantes .cantes .cache .amacs.d .flux .flux .foutsox .foutsox .foutsox .gconf .gconf .grupg .guyfs .juyten .jupyten .keras .lmod.d	Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory Directory	01/23/2019 1 drvx 09/02/2019 1 drvx 04/03/2019 1 drvx 04/03/2019 1 drvx 04/03/2019 1 drvx 04/03/2019 1 drvx 01/01/2018 1 drvx 11/01/2018 1 drvx 01/04/2018 1 drvx 0/19/2018 1 drvx 0/19/2018 1 drvx 0/19/2018 1 drvx 0/19/2018 1 drvx-r-x-	dylan dylan dylan dylan
.runelite .ssh .vim		Directory Directory Directory	03/15, 03/10, 03/21,	local Isbatch .matlab	Directory Directory Directory	09/02/2019 1 drwx 09/05/2019 1 drwx 09/02/2019 1 drwxrwxr-x	
• • • •				.nautilus	Directory	11/01/2018 1 drwxr-xr-x	
4 files and 37 directories. Tota	Direction Remote		Size Priority	19 files and 37 directories. Status	Total size: 57,345 bytes		

GUI Clients: CyberDuck

Available on Windows & MacOS

Connects to host directly with SFTP

Allows for transfers through the GUI

Ă Ň

● ● ●	🤞 dylan@ada.tamu.edu – SFTP	
Open Connection Quick Connect	Action Refresh Edit	Disconnect
	🔲 /general/home/dylan	🗘 🔺 Q Search
Filename	^ Size	Modified
Desktop		10/4/ 10, 0·20 MM
Documents		10/4/18, 6:20 PM
dos_script.sh		11/5/19, 11:54 AM
▶ 💼 Downloads		10/4/18, 6:20 PM
Example1Out.8875787		9/5/19, 4:12 PM
helpdesk-scripts.tar		8/2/19, 11:55 AM
intro_to_linux		9/5/19, 2:13 PM
matlab_crash_dump.12621-1	3.8 KB	9/19/19, 5:40 PM
matlab_crash_dump.12621-2	0 B	9/19/19, 5:40 PM
▶ 🚞 PA2		1/22/20, 11:55 AM
🕨 🚞 R		8/5/19, 10:44 AM
🕨 🚞 rodriguez.dylan		10/4/18, 6:20 PM
sasuser.v94		9/23/19, 9:32 AM
Scripts		1/6/20, 2:36 PM
▶ 💼 some_dir		9/12/19, 4:36 PM
some_dir.tar.gz	2.8 KB	9/13/19, 2:17 PM
some_file	92 B	9/12/19, 4:25 PM
star-job.lsf	341 B	10/1/19, 2:30 PM
7 test.py	48 B	1/30/19, 4:21 PM
testJob.lsf	663 B	9/5/19, 4:19 PM
20 Files		

GUI Clients: Globus

Web based, with application you can download

Grace endpoints: grace-dtn1 grace-dtn2

Terra endpoint terra-ftn

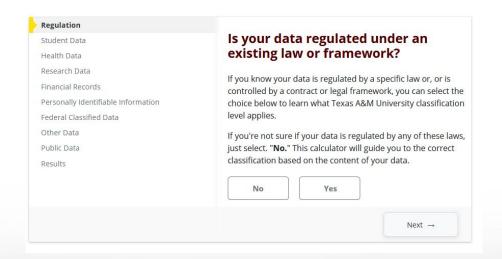
Ă M

https://www.globus.org/

🖭 globus	≡<	File Manager						p.	anels	00
File Manager		Collection TAMU ada-ftn1				Q	TAMU terra-ftn			Q
		Path /~/					/~/			<u></u> [
		select all 🔶 🙄 🏢			🗸 Sort	≡,	select all 👔		,	✓ Sort
		Desktop	10/04/2018 06:20p	-	>	Ð	helpdesk-scripts.tar	09/12/2019 05:17	20.48 KB	
		Documents	10/04/2018 06:20p	-	>	1	intel	09/09/2019 11:12		>
		Downloads	10/04/2018 06:20p	-	>	,C	perl5	07/30/2019 01:22		>
		Example1Out.8875787	09/05/2019 04:12pm	1.65 KB			Scripts	08/02/2019 11:57		>
		helpdesk-scripts.tar	08/02/2019 11:55am	20.48 KB		4	test-globus	08/19/2019 10:34		>
dylan@tamu.edu		intro_to_linux	09/05/2019 02:13pm	-	>		test.job	04/15/2019 03:12	696 B	
		R	08/05/2019 10:44am	-	>					
		rodriguez.dylan	10/04/2018 06:20p	-	>	8				
		Scripts	08/02/2019 11:52am	-	>					
		some_dir	09/12/2019 04:36pm	_	>	0				
		some_dir.tar	09/12/2019 04:38pm	2.75 KB						
		some_file	09/12/2019 04:25pm	92 B						
		test.py	01/30/2019 04:21pm	48 B						
			Start 🕞		Transfer &	Sync O	ptions 🗸	(d) Start		

Data Classification Tool

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



Can be found at:

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

Questions?

Continued Learning

Intro to HPRC Video Tutorial Series

HPRC's Wiki Page