High Performance Research Computing A Resource for Research and Discovery



# Introduction to Data Literacy and Data Management



# Contact the HPRC Helpdesk

Website: hprc.tamu.edu
Email: help@hprc.tamu.edu
Telephone: (979) 845-0219
Visit us in person: Henderson Hall, Room 114A

## Help us, help you -- we need more info

- Which Cluster
- UserID/NetID
- •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

# Logging in to the system

- SSH (secure shell)
  - The only program allowed for remote access; encrypted communication; freely available for Linux/Unix and Mac OS X hosts;
- For Microsoft Windows PCs, use MobaXterm
  - https://hprc.tamu.edu/wiki/HPRC:MobaXterm
    - You are able to view images and use GUI applications with MobaXterm
  - or *Putty* 
    - https://hprc.tamu.edu/wiki/HPRC:Access#Using\_PuTTY
      - You can not view images or use GUI applications with PuTTY

# Your Login Password

- Both state of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts
- Don't write down passwords
- Don't choose easy to guess/crack passwords
- Change passwords frequently

# **Course Topics**

- 1. Data on our Clusters
- 2. Good data practices
- 3. Data management tools (command line)
- 4. Data management tools (GUI clients)

# Using SSH - MobaXterm (on Windows)



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

Texas A&M University

High Performance Research Computing – https://hprc.tamu.edu

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

### ssh user NetID@ada.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 'ada.tamu.edu' added to the list of known hosts. user\_NetID@ada.tamu.edu's password:

## Data on Our Clusters: Ada

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

sho	wquota	View yo	ur current quota w	vith this command	
Your curre Disk /home /scratch	ent disk quotas a Disk Usage 416.1M 18.64G	rre: Limit 10G 1T	File Usage 4489 122616	Limit 10000 250000	
/tiered	0	10T	4	250000	

## Data on Our Clusters: Ada

#### **Default Limits**

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

Need more space? Contact help@hprc.tamu.edu

A M

## Data on Our Clusters: Ada

What's the difference between these filesystems?



Need more space? Contact help@hprc.tamu.edu

# Data on Our Clusters: Ada's Fast Transfer Nodes

Ada has 2 nodes dedicated to fast file transfers:

ada-ftn1@ada.tamu.edu ada-ftn2@ada.tamu.edu

Access the node via ssh:

ssh netID@ada-ftn1@ada.tamu.edu

Both nodes have 40 gigabit capability No programming environment installed → these are for transfers only! These nodes have access to all of Ada's filesystem /home /scratch /tiered

## 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	vith this command	
Your curre Disk /home /scratch	ent disk quotas a Disk Usage 416.1M 18.64G	re: Limit 10G 1T	File Usage 4489 122616	Limit 10000 250000	

## Data on Our Clusters: Terra

#### **Default Limits**

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

Need more space? Contact help@hprc.tamu.edu

A M

## Data on Our Clusters: Terra

What's the difference between these filesystems?



Need more space? Contact help@hprc.tamu.edu

## **Good Data Practices**

## 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

## **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@ada.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@ada.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

Globus



## **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 https://hprc.tamu.edu/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

New Site     My Workspace     Private     Cloud backup     My server     University     Work     Primary     Secondary	Session File protocol: SFTP V Host name: example.com User name: Pass martin	Port number:		
	Save V	A <u>d</u> vanced <b> </b> ▼		

## **GUI Clients: FileZilla**

Open source, available on all platforms

Connects to host directly with SFTP

Allows for transfers through the GUI

Ă M

ost: sftp://ada.tamu.ed tatus: Connecting to ada: Connected to ada.tatus: Retrieving directory tatus: Listing directory/gy	u Username: dylan amu.edu amu.edu listing eneral/home/dylan	Password:		Port: Qui	kconnect		
latus. Directory insting of	/general/nome/uyian success	101					
ocal site: /Users/dylan/			<u> </u>	Remote site: /general/h	ome/dylan		
Guest     Guest     Gyin     Gyin     Volumes     bin     cores     dev			I	▼ ? / ▼ ? general ▼ ? home ► b dylan			
h nto		Citation Citations	Li set si O	Filenese A	Filesies Filetons	I ant medified Descriptions	0
ename 23		Filesize Filetype	Last II	Filename A	Filesize Filetype	Last modified Permissions	Owner/Group
Trash		Directory	09/01	.cache	Directory	12/14/2018 1 drwxrwxr-x	dvlan dvlan
bash sessions		Directory	09/12	.config	Directory	11/01/2018 1 drwxrwxr-x	dylan dylan
cache		Directory	02/27	.dbus	Directory	09/12/2019 1 drwx	dylan dylan
confia		Directory	05/20	.emacs.d	Directory	01/23/2019 1 drwx	dvlan dvlan
cups		Directory	04/03	felix	Directory	09/02/2019 1 drwxrwxr-x	dylan dylan
ipython		Directory	06/26	fitk	Directory	01/09/2019 1 drwx	dylan dylan
itmstransporter		Directory	05/22	fluxbox	Directory	04/03/2019 1 drwx	dylan dylan
iupyter		Directory	06/26	fontconfig	Directory	04/03/2019 1 drwxr-xr-x	dylan dylan
keras		Directory	06/26	aconf	Directory	09/12/2019 1 drwx	dylan dylan
local		Directory	02/27	anome?	Directory	11/01/2018 1 drwxr-xr-x	dylan dylan
matnlotlib		Directory	06/26	anuna	Directory	01/24/2019 1 drwx	dylan dylan
narok2		Directory	03/21	grupg	Directory	11/01/2018 1 drwx	dylan dylan
ngi ukz		Directory	02/27	invthon	Directory	10/19/2018 1 drwyr-yr-y	dylan dylan
noue-gyp		Directory	02/27	iunvter	Directory	10/19/2018 1 drwyr-yr-y	dylan dylan
oracle ire usage		Directory	02/27,	koras	Directory	07/20/2010 1 drawsers v	dylan dylan
oracio_Jre_usage		Directory	03/15/		Directory	12/02/2019 1 druwtrwxr-x	dylan dylan
putty		Directory	02/20		Directory	00/02/2010 1 drwxfWXf-X	dylan dylan
runente		Directory	03/15,	Iocal	Directory	09/02/2019 1 drwx	dulan dular
5511		Directory	03/10,	motion	Directory	09/03/2019 1 drwx	dulan dular
		Directory	03/21/	natiap	Directory	11/01/2019 1 drwxfWXf-X	dylan dylan
files and 37 directories. To	tal size: 71,759 bytes			<ul> <li>19 files and 37 directories.</li> </ul>	otal size: 57,345 bytes	11/01/2018 1 drwxF-XF-X	uyian uyian
ver/l ocal file	Direction Remote file		Size Driorit	Statue	, , , , , , , , , , , , , , , , , , , ,		
1 files and 37 directories. To erver/Local file	tal size: 71,759 bytes Direction Remote file	1	Size Priorit	19 files and 37 directories.	fotal size: 57,345 bytes	11/01/2018 I drwxt-xr-x	dylan dylan

## **GUI Clients: Globus**

Web based, with application you can download

Ada endpoints: ada-ftn1 ada-ftn2

Terra endpoint terra-ftn

Ă M

https://www.globus.org/

🔊 globus	≡∢	🗂 File Manage	er							ţ	Panels	
File Manager		Collection TAMU ada	a-ftn1				Q	TAMU terra-ftn				Q
• TAMU ada-ftn2		Path /~/						/~/				
TAMU terra-ftn		select all 🔶 🕻	3 II 🔳 🗉			∽ Sort		select all 🔶				🗸 Sort
BOOKMARKS		Desktop		10/04/2018 06:20p	-	>	Ð	helpdesk-scripts	s.tar 09/12/201	9 05:17	20.48 KB	
/ <sub>r</sub> Activity		Documents		10/04/2018 06:20p	-	>	1	intel	09/09/20.	9 11:12		>
O Endpoints		Downloads		10/04/2018 06:20p	-	>		perl5	07/30/201	9 01:22		>
SA Console		Example1Out.887	75787	09/05/2019 04:12pm	1.65 KB			Scripts	08/02/201	9 11:57		>
Account		helpdesk-scripts.	tar	08/02/2019 11:55am	20.48 KB		4	test-globus	08/19/201	9 10:34		>
<ul> <li>dylanøtamu.edu</li> <li>□→ Logout</li> </ul>		intro_to_linux		09/05/2019 02:13pm	-	>		test.job	04/15/201	9 03:12	696 B	
? Help		R		08/05/2019 10:44am	-	>	4					
Globus Home		rodriguez.dylan		10/04/2018 06:20p	-	>	۲					
		Scripts		08/02/2019 11:52am	-	>	0					
		some_dir		09/12/2019 04:36pm	-	>						
		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			

## **Questions?**