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

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/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:

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: Ada

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	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	1T / 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.tamu.edu

Both nodes have <u>40 gigabit capability</u> 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

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
- 4. Cyberduck

Globus Connect







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

- THVGOL				
Cloud backup My server	SFTP ~	Po <u>r</u> t number:		
Work	example.com	vord:		
Secondary	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

ĂЙ

cocal site: /Users/dylan/ Guest Shared Volumes bin cores dev Filesize Filetype		Remote site: /general/hor	ne/dylan		
Guest	1	▼ ? / ▼ ? general ▼ ? home ▶ ► dylan			
lename A Filesize Filetype					
	Last m	Filename 🔨	Filesize Filetype	Last modified Permissions	Owner/Group
.Trash Directory	09/01,	📒 .cache	Directory	12/14/2018 1 drwxrwxr-x	dylan dylan
.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
.config Directory	05/20	iemacs.d	Directory	01/23/2019 1 drwx	dylan dylan
.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	E .fluxbox	Directory	04/03/2019 1 drwx	dylan dylan
.jupyter Directory	06/26	.fontconfig	Directory	04/03/2019 1 drwxr-xr-x	dylan dylan
.keras Directory	06/26	.gconf	Directory	09/12/2019 1 drwx	dylan dylan
.local Directory	02/27,	gnome2	Directory	11/01/2018 1 drwxr-xr-x	dylan dylan
.matplotlib Directory	06/26	📒 .gnupg	Directory	01/24/2019 1 drwx	dylan dylan
.ngrok2 Directory	03/21,	gvfs	Directory	11/01/2018 1 drwx	dylan dylan
.node-gyp Directory	02/27,	ipython	Directory	10/19/2018 1 drwxr-xr-x	dylan dylan
.npm Directory	02/27	jupyter	Directory	10/19/2018 1 drwxr-xr-x	dylan dylan
.oracle_jre_usage Directory	03/15,	ikeras	Directory	07/30/2019 1 drwxrwxr-x	dylan dylan
.putty Directory	02/26	lmod.d	Directory	12/03/2018 1 drwxrwxr-x	dylan dylan
runelite Directory	03/15,	local	Directory	09/02/2019 1 drwx	dylan dylan
.ssh Directory	03/10,	Isbatch	Directory	09/05/2019 1 drwx	dylan dylan
.vim Directory	03/21,	imatlab.	Directory	09/02/2019 1 drwxrwxr-x	dylan dylan
		📒 .nautilus	Directory	11/01/2018 1 drwxr-xr-x	dylan dylan
files and 37 directories. Total size: 71,759 bytes		19 files and 37 directories. Tot	tal size: 57,345 bytes		
ver/Local file Direction Remote file	Size Priori	ty Status			

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	Unregistered
C+ Quick Connect	Action Refresh Edit	Disconnect
	general/home/dylan	🗘 🔺 Q Search
Filename	^ Size	Modified
Desktop		
		10/4/18, 6:20 PM
dos_script.sh	36 B	11/5/19, 11:54 AM
Downloads		10/4/18, 6:20 PM
Example1Out.8875787	1.7 KB	9/5/19, 4:12 PM
helpdesk-scripts.tar	20.5 KB	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
👔 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

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 🔶 🕻	S 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	
/ _r 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		>
 dylanøtamu.edu □→ Logout 		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?