Introduction to Data Literacy and Data Management
Use the MaroonBase Student App to earn points for attending events and using resources listed in the app. Earn points, win prizes!

Monthly drawings – 3 students win $500 each
Semester grand prizes - top 3 students earn up to $2,000

Instructions to get started at maroonbase.tamu.edu

INNOVATION | CREATIVITY | LEADERSHIP | ENTREPRENEURSHIP
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](https://hprc.tamu.edu/wiki/HPRC:MobaXterm)
    – You are able to view images and use GUI applications with MobaXterm
  – or **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)

- IMPORTANT POLICY INFORMATION -
  - Unauthorized use of HPRC resources is prohibited and subject to criminal prosecution.
  - Use of HPRC resources in violation of United States export control laws and regulations is prohibited. Current HPRC staff members are US citizens and legal residents.
  - Sharing HPRC account and password information is in violation of State Law. Any shared accounts will be DISABLED.
  - Authorized users must also adhere to all policies at:

** WARNING: There are NO active backups of user data. **

Please restrict usage to 8 CORES across ALL Ada login nodes. Users found in violation of this policy will be SUSPENDED.

**** Ada Scheduled Maintenance Completed ****

The maintenance for Ada has been completed. Batch job scheduling has resumed.

Your current disk quotas are:

- Disk
  - Usage: 117.2M
  - Limit: 10G
- scratch
  - Usage: 6.684G
  - Limit: 30G
- tiered
  - Usage: 7.017
  - Limit: 467

Type "showquota" to view these quotas again.

[whomp@ajax5 ~]$
Using SSH (on a Linux/Unix Client)

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

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 → AKA quota
The limits are on Disk Space & File Usage

```
showquota
```

View your current quota with this command

Your current disk quotas are:

<table>
<thead>
<tr>
<th>Disk</th>
<th>Disk Usage</th>
<th>Limit</th>
<th>File Usage</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home</td>
<td>416.1M</td>
<td>10G</td>
<td>4489</td>
<td>10000</td>
</tr>
<tr>
<td>/scratch</td>
<td>18.64G</td>
<td>1T</td>
<td>122616</td>
<td>250000</td>
</tr>
<tr>
<td>/tiered</td>
<td>0</td>
<td>10T</td>
<td>4</td>
<td>250000</td>
</tr>
</tbody>
</table>
### Data on Our Clusters: Ada

#### Default Limits

<table>
<thead>
<tr>
<th>Directory</th>
<th>Space</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home</td>
<td>10G</td>
<td>10,000</td>
</tr>
<tr>
<td>/scratch</td>
<td>1T</td>
<td>250,000</td>
</tr>
<tr>
<td>/tiered</td>
<td>10T</td>
<td>250,000</td>
</tr>
</tbody>
</table>

Need more space?  
Contact help@hprc.tamu.edu
Data on Our Clusters: Ada

What’s the difference between these filesystems?

/home
- backed up
- will not be expanded

/scratch
- ‘high performance storage’
- can be expanded
- not backed up

/tiered
- archival file system
- can be expanded
- not backed up

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
There are limits on data on our clusters → AKA quota
The limits are on Disk Space & File Usage

```markdown
showquota
```
View your current quota with this command

<table>
<thead>
<tr>
<th>Disk</th>
<th>Disk Usage</th>
<th>Limit</th>
<th>File Usage</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home</td>
<td>416.1M</td>
<td>10G</td>
<td>4489</td>
<td>10000</td>
</tr>
<tr>
<td>/scratch</td>
<td>18.64G</td>
<td>1T</td>
<td>122616</td>
<td>250000</td>
</tr>
</tbody>
</table>
## Data on Our Clusters: Terra

### Default Limits

<table>
<thead>
<tr>
<th>Directory</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home</td>
<td>10G / 10,000 files</td>
</tr>
<tr>
<td>/scratch</td>
<td>1TB / 250,000</td>
</tr>
</tbody>
</table>

Need more space?  
Contact help@hprc.tamu.edu
Data on Our Clusters: Terra

What’s the difference between these filesystems?

- /home
  - backed up nightly
  - will not be expanded

- /scratch
  - ‘high performance storage’
  - can be expanded
  - not backed up

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

Delete 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**

<table>
<thead>
<tr>
<th>Flag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-cf</td>
<td>create archive</td>
</tr>
<tr>
<td>-xf</td>
<td>extract archive</td>
</tr>
<tr>
<td>-v</td>
<td>verbose</td>
</tr>
<tr>
<td>-z</td>
<td>compress with gzip</td>
</tr>
</tbody>
</table>
GUI Clients

There are many GUI solutions for file transfer:

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

Globus Connect
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: 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
GUI Clients: CyberDuck

Available on Windows & MacOS

2 Factor Authentication makes FileZilla hard to use

Connects to host directly with SFTP

Allows for transfers through the GUI
GUI Clients: Globus

Web based, with application you can download

Ada endpoints:
ada-ftn1
ada-ftn2

Terra endpoint
terra-ftn

https://www.globus.org/
Questions?