HPRC Primer
Data Management Practices on HPRC Resources
February 2, 2024
High Performance Research Computing
DIVISION OF RESEARCH
HPRC Resources

Knowledge Base
- https://hprc.tamu.edu/kb/

HPRC Website
- https://hprc.tamu.edu/

Youtube Channel
- https://www.youtube.com/texasamhprc
Computing Resources

The HPRC group currently administers four HPC clusters:

- ACES
- FASTER
- Grace
- Terra (retiring soon!)

You’ll need one of two options to use them:

<table>
<thead>
<tr>
<th>Credentials</th>
<th>Clusters</th>
<th>Who</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPRC Account</td>
<td>FASTER, Grace, Terra</td>
<td>Mostly Texas A&amp;M students/staff</td>
</tr>
<tr>
<td>ACCESS ID</td>
<td>FASTER and ACES</td>
<td>Anybody</td>
</tr>
</tbody>
</table>

Link to our Knowledge Base: [https://hprc.tamu.edu/kb/](https://hprc.tamu.edu/kb/)
Accessing Clusters via Portal

Access through (most) web browsers
– portal.hprc.tamu.edu
– Or dropdown menu on HPRC homepage:
Using the Portal - Shell Access

OnDemand provides an integrated, single access point for all of your HPC resources.

Message of the Day

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: https://hprc.tamu.edu/policies

!! WARNING: THERE ARE ONLY NIGHTLY BACKUPS OF USER HOME DIRECTORIES. !!
Log into the Grace cluster using the online Portal (If you’re not on TAMU Wifi, then you should use TAMU VPN before trying to connect to our clusters)
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 Backup
**File Systems and User Directories**

<table>
<thead>
<tr>
<th>Directory</th>
<th>Environment Variable</th>
<th>Space Limit</th>
<th>File Limit</th>
<th>Intended Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>/home/$USER</td>
<td>$HOME</td>
<td>10 GB</td>
<td>10,000</td>
<td>Small to modest amounts of processing.</td>
</tr>
<tr>
<td>/scratch/user/$USER</td>
<td>$SCRATCH</td>
<td>1 TB</td>
<td>250,000</td>
<td>Temporary storage of large files for on-going computations. Not intended to be a long-term storage area.</td>
</tr>
</tbody>
</table>

$SCRATCH is shared between FASTER and Grace clusters.

View file usage and quota limits using the command: `showquota`

Do NOT share your home or scratch directories. Request a group directory for sharing files.

[https://hprc.tamu.edu/kb/User-Guides/FASTER/Filesystems_and_Files/](https://hprc.tamu.edu/kb/User-Guides/FASTER/Filesystems_and_Files/)
OOD Portal Quota Increase Request

Portal Homepage → [Cluster name] Dashboard

Request quota increases directly from the dashboard with a guided form.

Contact help@hprc.tamu.edu if you need more help.
1. Navigate to the Dashboard to see your quotas
2. Run the `showquota` command in a shell to see your quotas
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

To move data between machines, use `scp` or `sftp` (which we’ll cover in a moment)
Command Line Tools: `rm`

Remove

Deletes a file:

```
rm some_file
```

WARNING: *There is no “trash bin” on the command line!* Once you `rm` an object, you cannot get it back!

Add the `-i` flag to be prompted prior to file deletion:

```
rm -i some_file
```
On your command line:
- Create a test directory and make a file
- Copy the file
- Remove one copy
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.gz
```

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>
On your command line:

- Create more copies of the test file you made before
- Condense them all into one object
- Extract them back out and check that their contents are the same
Data Transfer

Login nodes
- medium-bandwidth internet connection
- 60 minute process time limit

File transfer nodes
- high-bandwidth internet connection
- no process time limit

Command line tools \texttt{scp}, \texttt{sftp}, and \texttt{rsync} are all available
Data Transfer: Grace DTN

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

SSH (from a login node) to either DTN:

```
ssh NetID@grace-dtn1.hprc.tamu.edu
ssh NetID@grace-dtn2.hprc.tamu.edu
```

Large transfers should use the Data Transfer Nodes

Both nodes have **40 gigabit capability** (vs 10 gigabit on login node)

No programming environment installed → these are for transfers only!

These nodes have access to all of Grace’s filesystem (/home and /scratch)
Data Transfer: Terra FTN

Terra has 1 node dedicated to data transfer → Fast Transfer Node

SSH (from a login node) 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 and /scratch)
Data Transfer: FASTER and ACES

The FASTER and ACES cluster each have two Data Transfer Nodes—one of which is dedicated to ACCESS users.

*ssh to Data Transfer Node not available on ACES.*

These can be accessed through Globus (more about Globus later).

https://hprc.tamu.edu/kb/User-Guides/FASTER/#data-transfer-nodes
https://hprc.tamu.edu/kb/User-Guides/ACES/#data-transfer-nodes
Command Line Tools: sftp

Secure file transfer protocol

interactive file transfer program

```
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

Not Recommended for large file transfers.
Command Line Tools: scp

Secure copy

Copies files between hosts on a network

scp source_file <NetId>@<node>.hprc.tamu.edu:/home/<NetId>

Can be used
- local to remote (as above)
- remote to local
- remote to remote
Command Line Tools: rsync

Remote Synchronize

Copies files between hosts on a network

```
rsync source_file <NetId>@<node>.hprc.tamu.edu:/home/<NetId>
```

Can be used
- local to remote (as above)
- remote to local
- remote to remote

→ rsync is better than scp
→ rsync supports intermittent transfer
• ssh to grace-dtn2 (from login node)
• use rsync to copy a .tar file to terra-ftn
• verify the file arrived (on Terra login node)
Graphical User Interface (GUI) Clients

There are many GUI solutions for file transfer:

- Open OnDemand Portal
- MobaXterm
- WinSCP
- Cyberduck
- Globus Connect
GUI Clients: HPRC Portal

Access your files through (almost) any web browser

View, Edit, Upload, Download, and 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

https://hprc.tamu.edu/kb/Helpful-Pages/#mobaxterm-recommended
GUI Clients: WinSCP

Available on Windows machines

Connects to host directly with SFTP

Allows for transfers through the GUI

https://hprc.tamu.edu/kb/Helpful-Pages/File-Transfer/#tutorial-videos
GUI Clients: CyberDuck

Available on Windows & MacOS

Connects to host directly with SFTP

Allows for transfers through the GUI

https://hprc.tamu.edu/kb/Helpful-Pages/File-Transfer/#tutorial-videos
GUI Clients: Globus

Web-based, with application you can download

Grace endpoints:
- grace-dtn1
- grace-dtn2

Terra endpoint
- terra-ftn

FASTER endpoints:
- TAMU FASTER DTN1
- ACCESS faster.tamu.access-ci.org (ACCESS users)

ACES endpoint:
- ACCESS TAMU ACES DTN

https://www.globus.org/
https://hprc.tamu.edu/kb/Software/Globus/
Transfer files between FASTER and Grace:

- Create some test file in your home directory on FASTER or Grace
- Log in to https://app.globus.org/
- Search for “TAMU” in the Collection fields and select “TAMU grace-dtn” and “TAMU FASTER DTN1”
  
  Globus should show you the contents of your home directories
- Transfer your test file from one cluster to the other
Data Classification Tool

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

[Image of a user interface with options for data classification]

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

Intro to HPRC Video Tutorial Series

HPRC’s Knowledge Base
Need Help?

First check the FAQ: [https://hprc.tamu.edu/kb/FAQ/Accounts/](https://hprc.tamu.edu/kb/FAQ/Accounts/)
- Knowledge Base: [https://hprc.tamu.edu/kb/User-Guides/Portal/](https://hprc.tamu.edu/kb/User-Guides/Portal/)
- Email further questions to help@hprc.tamu.edu

Help us help you -- provide the following info:
- Which cluster you’re using
- Your 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
Thank you.

Any questions?