Data Storage and Transfer

Data Storage and Transfer

No. We NFS mount storage across all compute nodes so that data is available independent of which compute nodes are used. See this section for how to transfer data.

After creating your HPC Account, your home directory will not be created until you log in for the first time. Without your home directory, you will not be able to transfer your data to HPC. If you are struggling and receiving errors, sign into your account either using the CLI through the bastion or logging into OnDemand and then try again.

If you are using something like SCP and are receiving errors, make sure your hostname is set to filexfer.hpc.arizona.edu (not hpc.arizona.edu).

Unfortunately, no. Backups are not made and anything deleted is permanently erased. It is impossible for us to recover it. To ensure your data are safe, we recommend:

- Make frequent backups, ideally in three places and two formats. Helpful information on making backups can be found on our page Transferring Data.
- Use rm and rm -r with caution as these commands cannot be undone! Consider using rm -i when removing files/directories. The -i flag will prompt you to manually confirm file removals to make really sure they can be deleted.
- You can open a support ticket to request assistance. Files that are deleted may not have been removed from the storage array immediately (though this is not guaranteed), don’t wait more than a few days.

Endpoint too busy: This is most commonly seen when users are transferring directories to Google Drive. This is because Google has user limits restricting the number of files that can be transferred per unit time. When many files are being transferred at once, that limit may be exceeded. Globus will automatically hold the transfer until the limit is reset at which point it will continue. One way to avoid this is to archive your work prior to the transfer (e.g., in .tar.gz form). Additionally, archiving will also speed up your transfers considerably, sometimes by orders of magnitude.

Fatal FTP Response, PATH_EXISTS: Globus is finicky about the destination endpoint. If you get this error, check to see whether duplicate files/directories exist at the destination. This can happen frequently with Google Drive as multiple files/directories can exist in the same location with the same name. If duplicates exist, try moving, removing, or renaming them and reinitiate the transfer.

In our last maintenance update on July 20th, one of the changes was to ensure HIPAA compliance on the Data Transfer Nodes (DTNs). This change included the insertion of required text:

**Authorized uses only. All activity may be monitored and reported.**

This change breaks SCP activity. Not in all cases but frequently with WinSCP, Filezilla and from a terminal. Terminal activity will likely still work from Linux or MacOS.

The solution is to not use SCP (SCP is considered outdated, inflexible, and not readily fixed) and to use a more modern protocol like SFTP and rsync. Info on using SFTP:

- Putty supports SFTP with the “PSFTP” command
- For FileZilla, in the Toolbar, click on Edit and Settings, then click on SFTP
- For Cyberduck, choose SFTP in the dropdown for protocols.

Choose SFTP (SSH File Transfer Protocol), and use filexfer.hpc.arizona.edu as the Server. May enter Username and Password, and unselect “Add to KeyChain”.

For Mac, Cyberduck can be used. For version 8.4.2, select Go then Disconnect before clicking on Quit Cyberduck.

The recent Cyberduck patch 8.4.4 has created problems with two-factor authentication which is required to access filexfer.hpc.arizona.edu. Adding a SSH Key is suggested.

For Windows, WinSCP and FileZilla are recommended.
If your home directory is full and you can't find what is taking up all the space, it's possible the culprit is a hidden file or directory. Hidden objects are used for storing libraries, cached singularity/apptainer objects, saved R session, anaconda environments, configuration files, and more. It's important to be careful with hidden, or “dot”, files since they often control your environment and modifying them can lead to unintended consequences.

To view the sizes of all the objects (including hidden) in your home, one quick command is `du -hs $(ls -A ~)` , for example:

```
[netid@junonia ~]$ du -hs $(ls -A ~)
32K    Archives
192M   bin
4.7G   Software
46M    .anaconda
1.9M   .ansys
4.0K   .apptainer
16K    .bash_history
4.0K   .bash_logout
4.0K   .bash_profile
12K    .bashrc
20M    ondemand
```

Clearing out unwanted objects, moving data to a location with more space (e.g. /groups or /xdisk), and setting different defaults for data storage (e.g., resetting your apptainer cache directory or setting new working directories for Python/R) can help manage your home's space.

Unfortunately, without active university credentials it is not possible to access HPC compute or storage resources. External collaborators who need ongoing access may apply for Designated Campus Colleague, or DCC, status. This is a process done through HR and will give the applicant active university credentials allowing them to receive HPC sponsorship.

Otherwise, data will need to be moved off HPC and made available on a mutually-accessible platform. This may include (but is not limited to): Google Drive, AWS S3, Box, and CyVerse's Data Store.