

Expanse

Data Storage and Transfer

San Diego Supercomputer Center
October 2020

Storage Overview

- Home file system
- **Global Parallel Lustre filesystem**
 - /expanse/lustre/scratch
 - /expanse/lustre/projects
- Automated Purge Policy
- **NO BACKUPS!!!!**

Data Mover(DM) Nodes

- **Avoid using login nodes** to move large amounts of data
- Data Mover nodes enable efficient transfer of larger amounts of data
- Data movers can be used in conjunction with Globus Online and GridFTP, all available transfer methods
- Access:
 `>ssh oasis-dm.sdsc.edu`



Tools for Transferring Data

- scp(Secure Copy Protocol)
- rsync (CLI)
- Globus CLI – gsiscp, gsiftp
- Globus Connect -GUI
- globus-url-copy
- IBM Aspera (ascp)
- AWS CLI (sftp)



SCP & SFTP & RSYNC

- **Small files (< 2 GB)**
- **SCP**(Secure Copy Protocol) utility for secure file transfer
- **SFTP** (Secure File Transfer Protocol) utility for secure file access, transfer and management
- **RSYNC** Linux file transfer program to sync files or directories
- Syntax

`scp [-options] SourceFile username@host:/directory/TargetFile`

`rsync -h, --help`

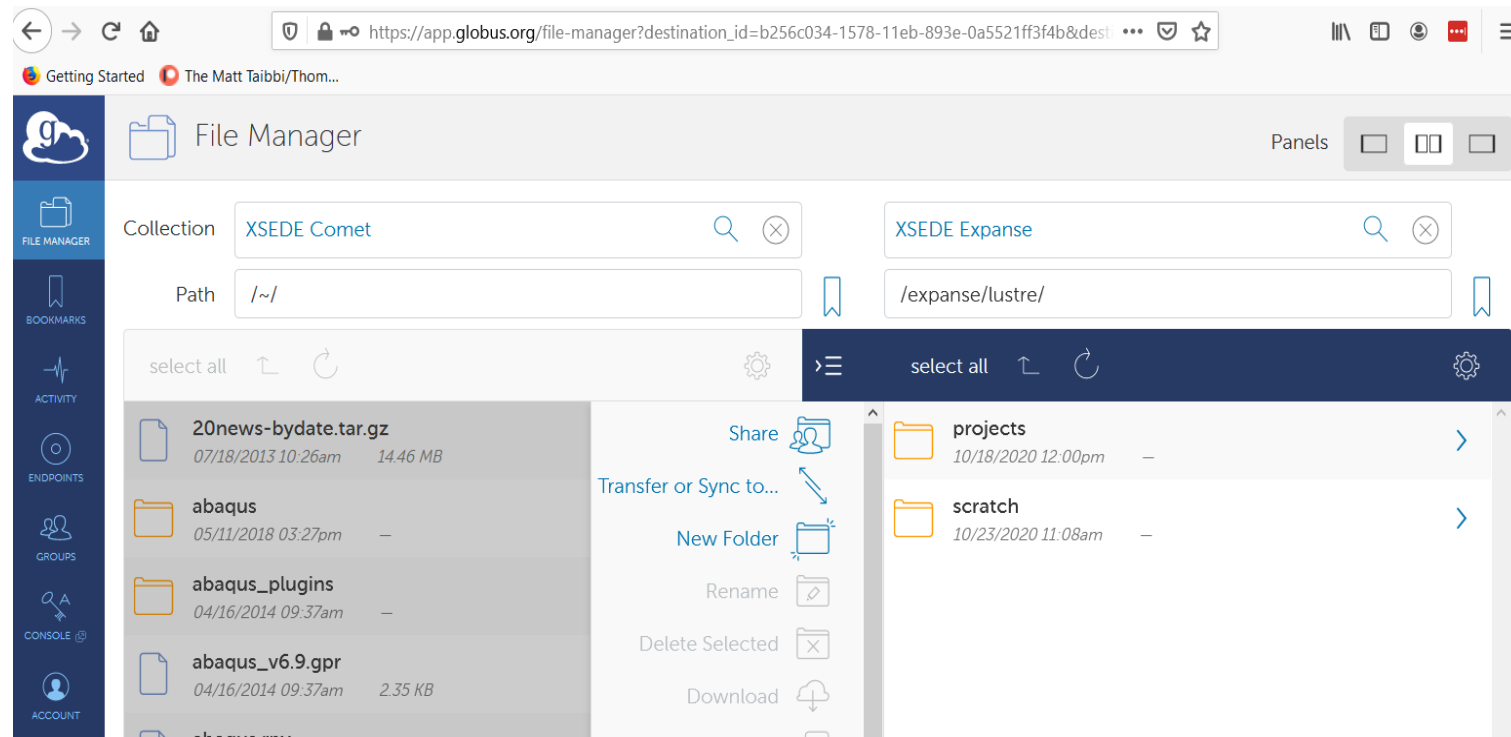


Globus GSI-OpenSSH, Globus-CLI

- The Globus GSI-OpenSSH implementation of OpenSSH
- Command-line transfer utilities supporting standard ssh and grid authentication
- Use in SSO hub
- Use gsissh to interface with Globus
- Transfer tools:
 - gsiscp and gsiftp commands = OpenSSH scp and sftp commands
- **Requires an active XSEDE allocation**
 - Contact help@xsede.org, consult@sdsc.edu if you do not have an XSEDE allocation

Globus GUI

- Globus GUI for high performance file transfer and sharing
- Uses XSEDE credentials
- Uses gridFTP protocol
- GUI (www.globus.org)
 - Endpoints
 - xsede#comet
 - xsede#exppanse
 - Paths
 - /exppanse/home/\$USER
 - /exppanse/lustre/projects
 - /exppanse/lustre/scratch



Globus-url-copy

- CLI for high speed data transfers
- Uses GridFTP protocol
- **Recommended for large files, and large number of files**
- Load globus module
- Active User Proxy certificate
- globus-url-copy --help

```
[nickel@comet-dm3-15-3 ~]$ globus-url-copy -vb -stripe -tcp-bs 8m -p 4 gsiftp://oasis-dm.sdsc.xsede.org:2811/home/nickel/linear_regression.tar.gz gsiftp://oasis-dm.sdsc.xsede.org:2811/exppanse/lustre/scratch/nickel/linear_regression2.tar.gz
Source: gsiftp://oasis-dm.sdsc.xsede.org:2811/home/nickel/
Dest:   gsiftp://oasis-dm.sdsc.xsede.org:2811/exppanse/lustre/scratch/nickel/
linear_regression.tar.gz -> linear_regression2.tar.gz

    37251072 bytes      315.78 MB/sec avg      315.78 MB/sec inst
[nickel@comet-dm3-15-3 ~]$
```


IBM ASPERA (ascp)

- Uses FASP technology to transfer and share files
- Available on data mover nodes
- Example

```
>ascp options [[user@]srcHost:]source_file1[,source_file2,...]  
[[user@]destHost:]target_path
```
- More information at
<https://download.asperasoft.com/download/docs/ascp/2.7/html/index.html>

AWS CLI

- Open source tool to interact with AWS services using commands in your command-line shell
- Uses Linux shells, Windows Command line
- AWS Transfer Family enables file transfers using File Transfer Protocol (FTP), File Transfer Protocol over SSL (FTPS), or Secure Shell (SSH) File Transfer Protocol (SFTP)
- More information on setting up AWS CLI

Install: <https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-install.html>

Options: <https://awscli.amazonaws.com/v2/documentation/api/latest/reference/s3/cp.html>

Review

- Lots of tools available on Expanse for data transfer
- Consider Size of files, number of files, workflow
- Avoid moving lots of small files
- Automated Purge on Lustre file system
- Comet Lustre file systems will go away Mid April.

