

File Systems and Data Transfer



Filesystems and Policies

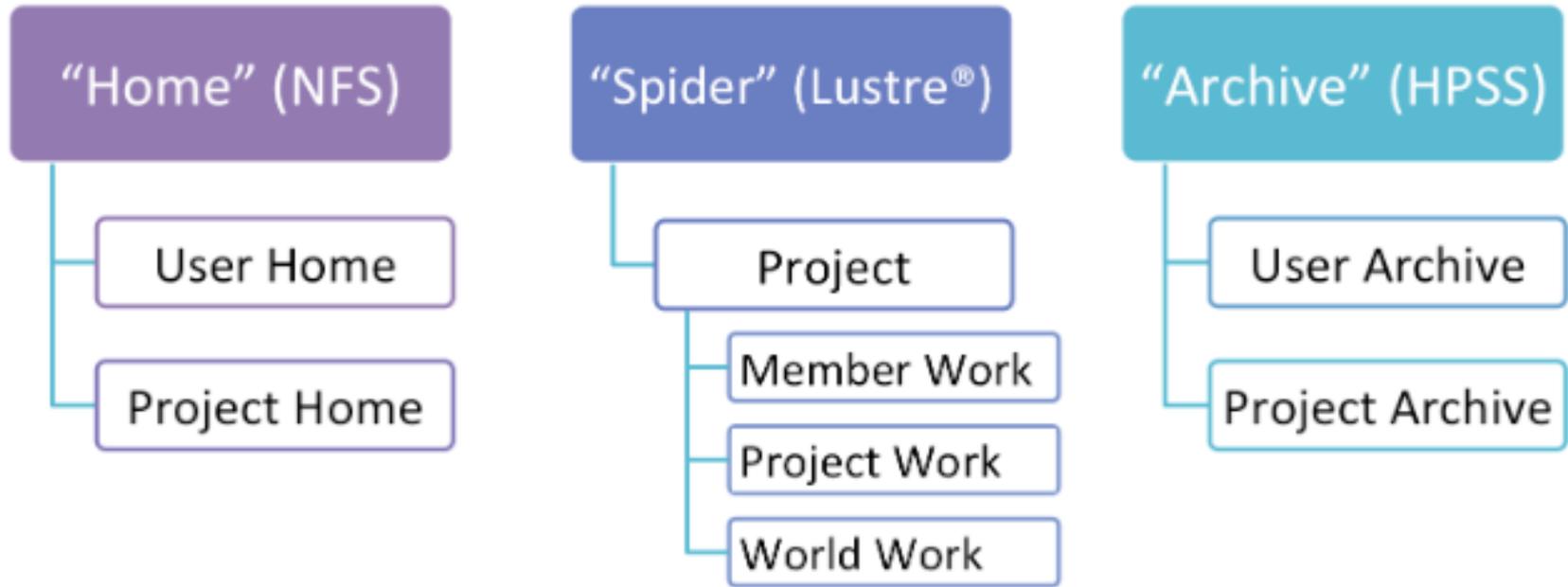
Suzanne Parete-Koon

Data transfer

Cover Scripts for Scheduled DTNS (Example HSI)

Cover GridFTP basics with Globusonline Example

A Storage Area for Every Activity



<https://www.olcf.ornl.gov/support/system-user-guides/titan-user-guide/#2695>

The OLCF machines all mount “Home” and “Spider”. Your files in User home on Rhea will be in User home on Titan.

“Home”, “Spider” and “Archive” are separate files systems. You must copy data between them.

Definitions

- **Quota** The limits placed on total number of bytes and/or files in the storage area.
- **Backups** States if the data is automatically duplicated for disaster recovery purposes.
- **Purged** Period of time, post-file-creation, after which a file will be marked as eligible for permanent deletion.
- **Retention** Period of time, post-account-deactivation or post-project-end, after which data will be marked as eligible for permanent deletion.
 - For sensitive projects only, all data related to the project must be purged from all OLCF computing resources within 30 days of the project's end or termination date.

A Storage Area for Every Activity

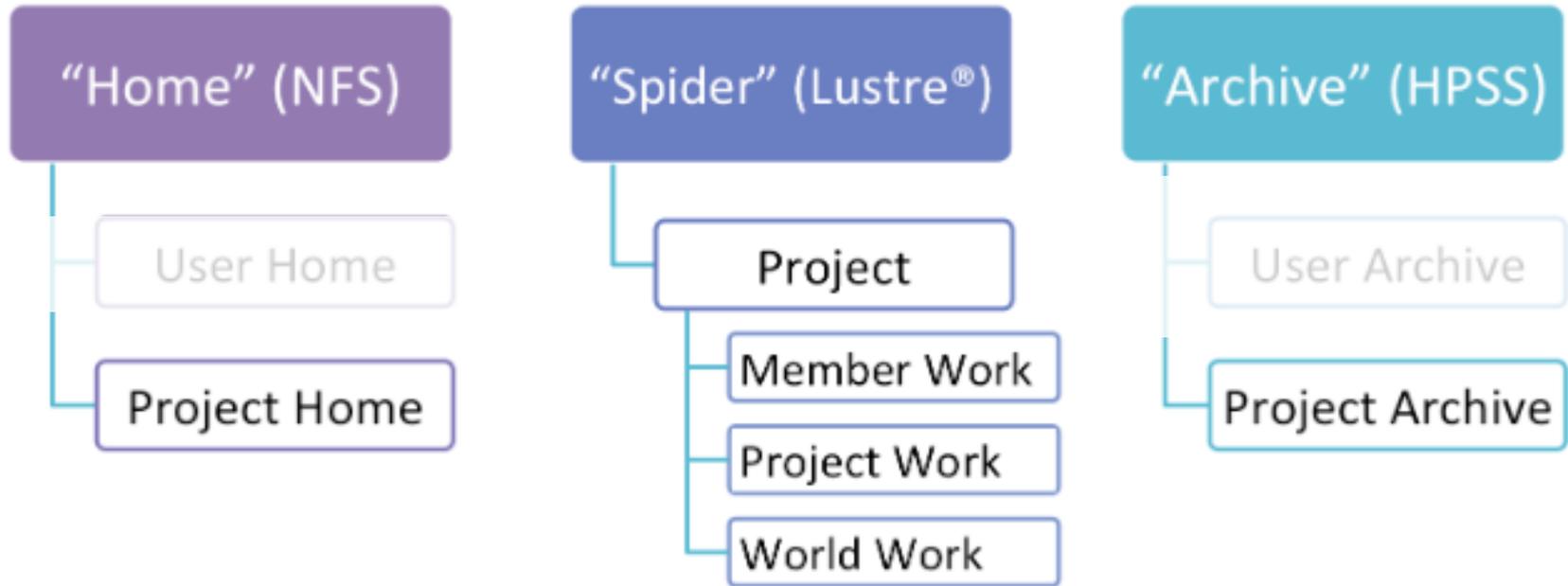


User-Centric Storage Areas : Long-term access not related to a project

Area	Path	Permissions	Quota	Backups	Purged	Retention
User Home	\$HOME	User-controlled	10 GB	Yes	No	90 days
User Archive	/home/\$USER	User-controlled	2 TB ^[1]	No	No	90 days

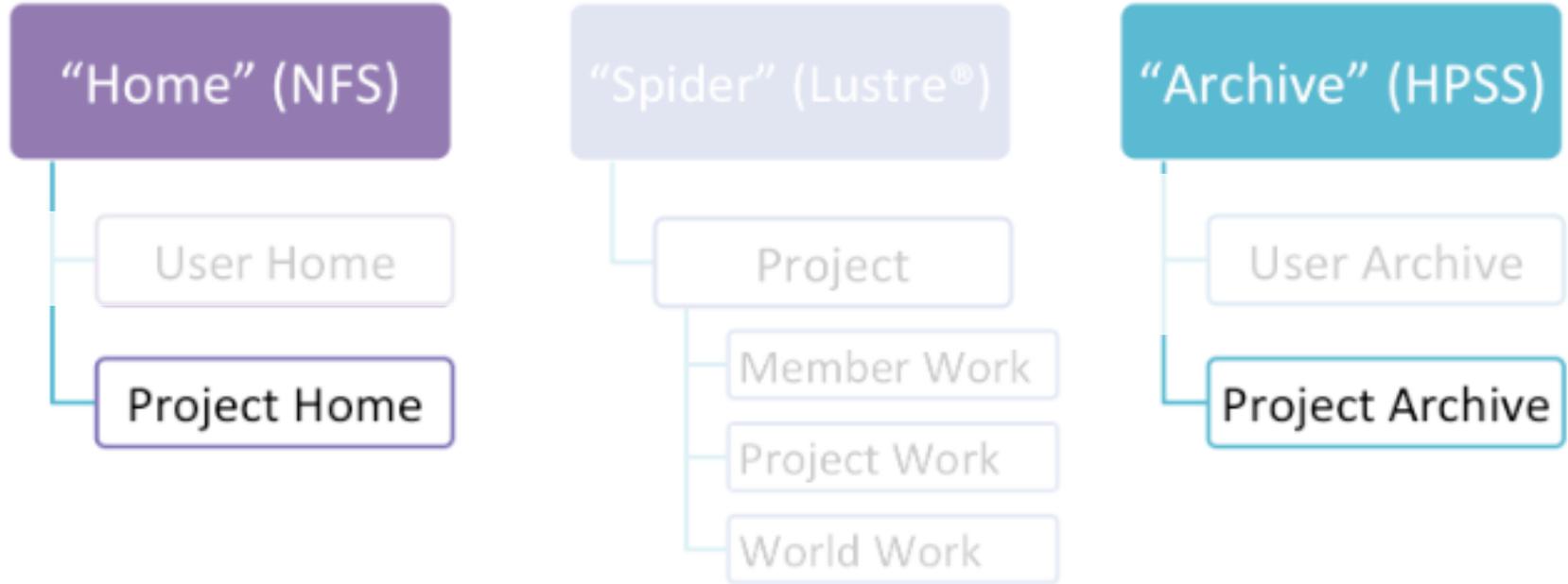
Tip: To see hpss /home/\$USER: hsi ls

A Storage Area for Every Activity



Project-Centric Areas : a common place to store code, data files, documentation, and other files related to your project.

A Storage Area for Every Activity

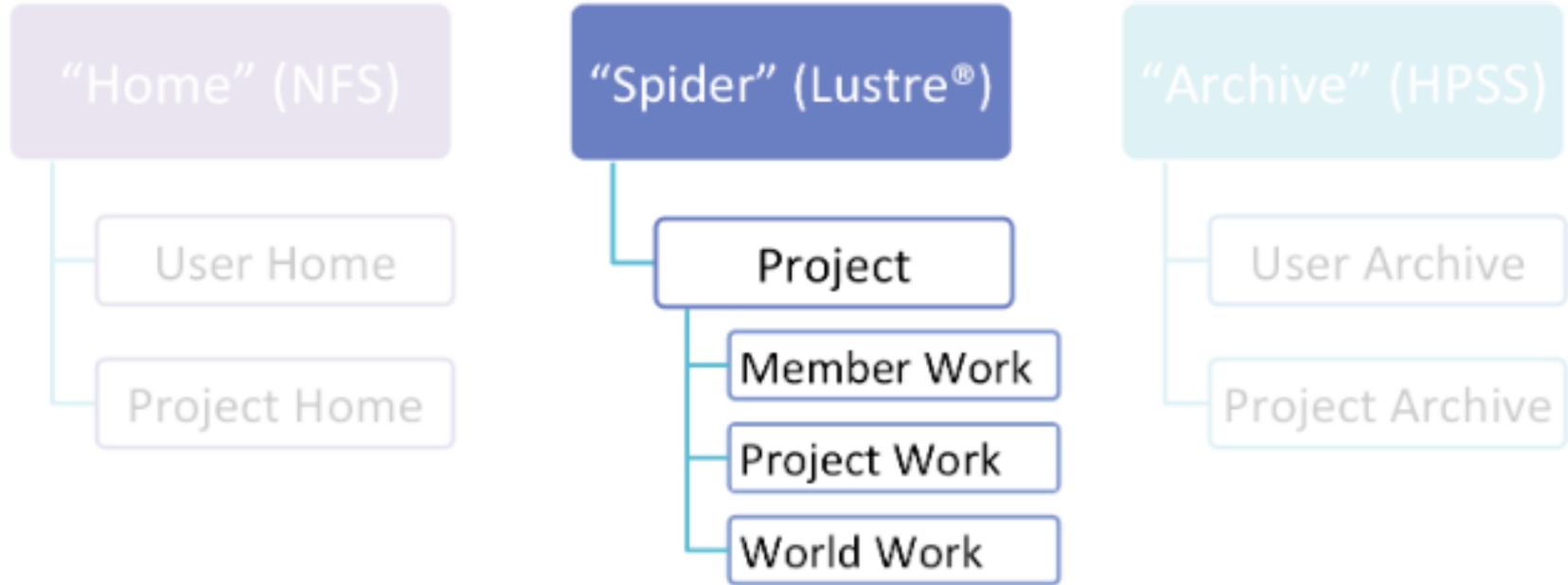


Long Term Project-Centric Areas: Not Purged

Area	Path	Permissions	Quota	Backups	Retention
Project Home	/ccs/proj/[projid]	770	50 GB	Yes	90 days
Project Archive	/proj/[projid]	770	100 TB	No	90 days

Tip: 770 = User rwx group rwx world ---

A Storage Area for Every Activity



Short Term Project-Centric Areas: Purged, No Backup

Area	Path	Permissions	Quota	Purged Days	Retention Days
Member Work	\$MEMBERWORK/[projid]	700	10 TB	14	14
Project Work	\$PROJWORK/[projid]	770	100 TB	90	90
World Work	\$WORLDWORK/[projid]	775	10 TB	14	14

Tip: 775 =user rwx group rwx world r-x

Prohibited Data

- The OLCF computer systems are operated as research systems and only contain data related to scientific research and do not contain personally identifiable information (data that falls under the Privacy Act of 1974 5U.S.C. 552a). Use of OLCF resources to store, manipulate, or remotely access any national security information is strictly prohibited. Authors/generators/owners of information are responsible for its correct categorization as sensitive or non-sensitive. For questions, contact help@olcf.ornl.gov.

- **Please read the full policy here :**

https://www.olcf.ornl.gov/kb_articles/data-management-policy/

Data Transfer

	GridFTP + GridCert	SFTP/SCP	BBCP
Data Security	insecure (default) / secure (w/configuration)	secure	insecure (unsuited for sensitive projects)
Authentication	GridCert	Passcode	Passcode
Transfer speed	Fast	Slow	Fast
Required Infrastructure	GridFTP server at remote site + user OSG Cert	Comes standard with SSH install	BBCP installed at remote site

Where to find Documentation

- **bbcp** : Multi-stream transfer tool
https://www.olcf.ornl.gov/kb_articles/transferring-data-with-bbcp/
- **GridFTP**: Two Flavors – globus (online) and a globus_url_copy -
Transfer based on Internet FTP protocol
https://www.olcf.ornl.gov/kb_articles/transferring-data-with-gridftp/
- **SCP**: Single Stream transfer tool
https://www.olcf.ornl.gov/kb_articles/transferring-data-with-sftp-and-scp/
- **hsi/htar** - hsi and htar transfer tools for User Archive and Project Archive spaces on the OLCF's HPSS-based archival storage system.
https://www.olcf.ornl.gov/kb_articles/transferring-data-with-hsi-and-htar/

Data Transfer Hardware

- dtn.ccs.ornl.gov

	Type	Remote	Local	Authentication	Transfer methods
dtn03	Interactive ssh dtn.ccs.ornl.gov	X	X	ssh, OSG cert.	Globus_url_copy , Globus, scp, bbcp
dtn04	Interactive ssh dtn.ccs.ornl.gov	X	X	ssh, OSG cert.	online, scp, bbcp
dtn- sch	Batch Job Scheduled	X	X	OSG, cert	Scp, bbcp, Globus)url_copy, GriFTP, DCP
dtnHSI	HPSS Only		X		hsi

Tip: dtn- Data Transfer Node

Scheduled (Batch) DTNs

- The scheduled DTNs are accessed by batch scripts from Titan.
- You can not enter a password.
- Options:
 - 1. Use as a part of your workflow for local transfers with hpss.
 - 2. Use with cp or dcp for local transfers between filesystem partitions See: https://www.olcf.ornl.gov/kb_articles/transferring-data-with-dcp/
 - 3. Use for transfer to remote places that have single password entry and allow SSH keys (for example NERSC).
 - 4. Use with GridFTP and certificate based authentication.
- Your allocation is not billed for DTN usage.

Local Data Transfer: Archive HPSS

Use his/htar to Transfer with the HPSS

To send a file to HPSS, you might use:

```
dtn04 $ hsi put a.out
```

Or

```
dtn03 $ htar -cvf myfiles.tar dir1/*
```

To retrieve a file to HPSS, you might use:

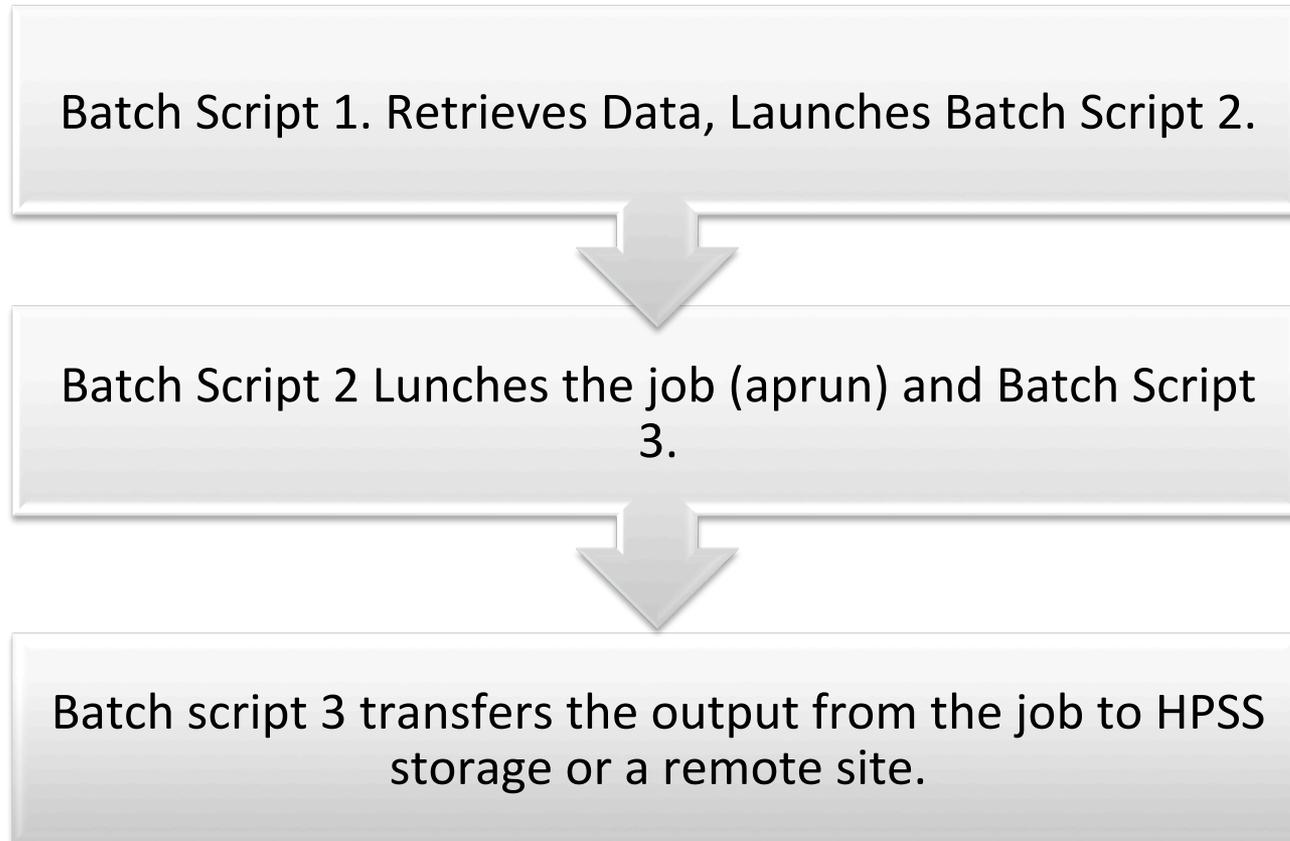
```
dtn03 $ hsi get a.out
```

Or

```
dtn04 $ htar -xvf myfiles.tar
```

Batch DTN Example

- You can script data transfers as part of your workflow.
- #PBS -l partition=dtm
- Launch script from Titan



Batch DTN Example

Full example scripts at: https://www.olcf.ornl.gov/kb_articles/employing-data-transfer-nodes/

```
$ cat backup.pbs

# Batch script to back-up resulting data

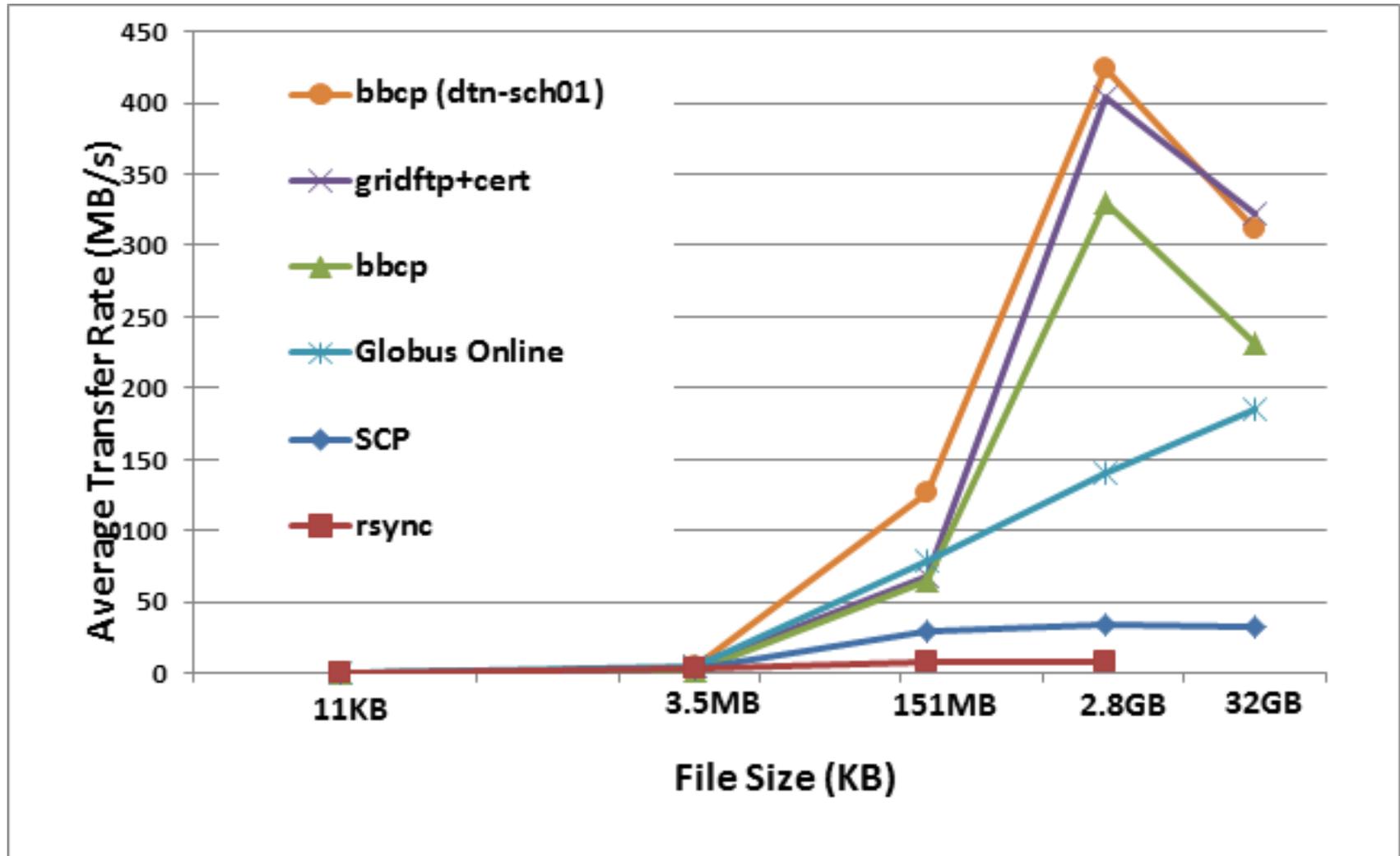
# PBS directives
#PBS -l partition=dtm
#PBS -A PROJ123
#PBS -l walltime=8:00:00

# Store resulting data
cd $MEMBERWORK/pjt000
hsi put largedatfileC
hsi put largedatafileD

$
```

Remote Transfers

Transfer Rates OLCF to NERSC

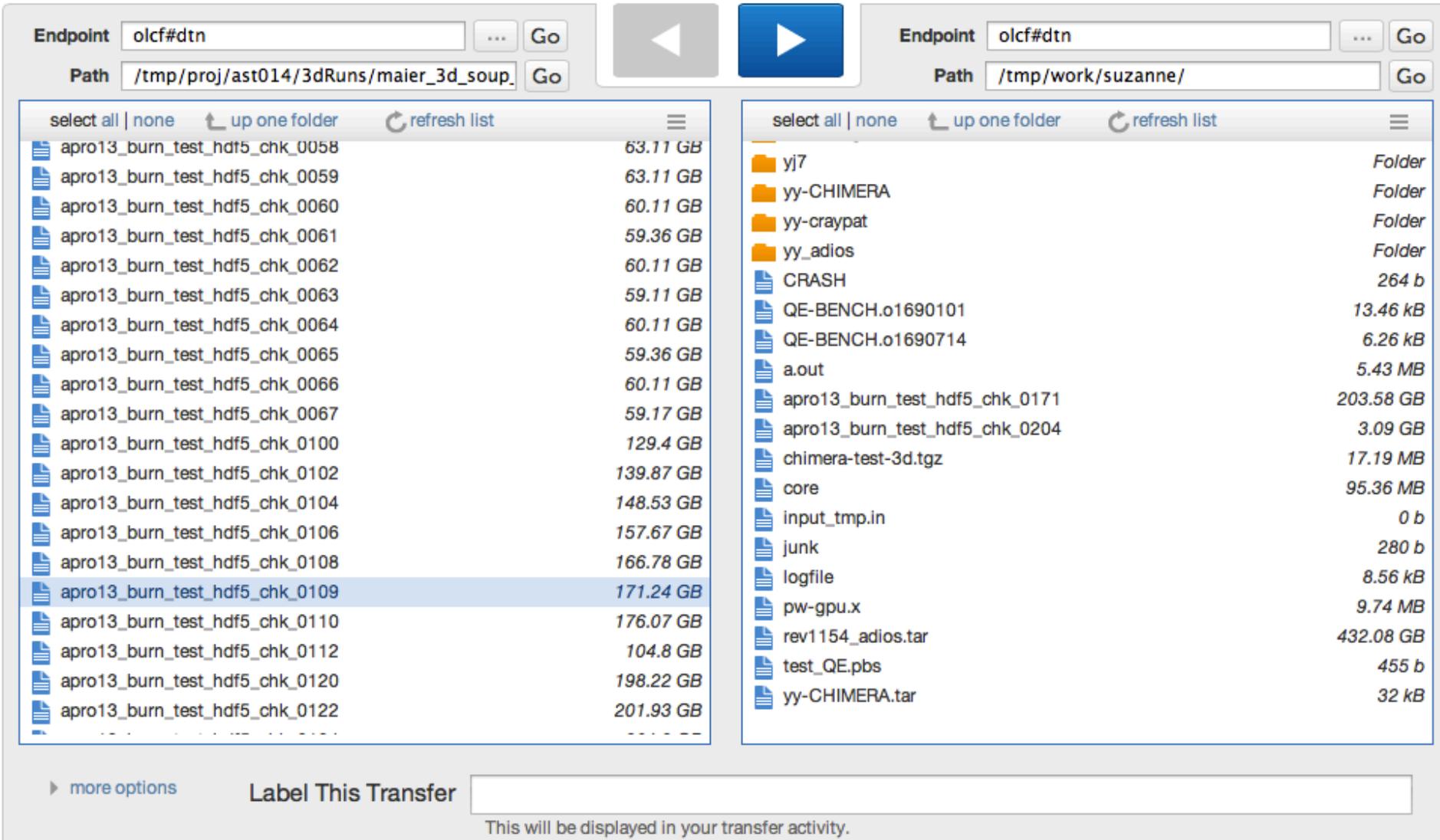


GridFTP

- GridFTP is a high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks that is based upon the Internet FTP protocol.
- Two flavors: globus(online), globus-url-copy
- Both ends of the transfer must have GridFTP set up.
- OLCF Require the Open Science Grid Certificates for authentication
- You must have a valid certificate setup at both ends of the transfer.

Globus (Online)

OLCF Globus Endpoint is olcf#dtn



Endpoint: ... Path:

Endpoint: ... Path:

select all | none [up one folder](#) [refresh list](#)

	apro13_burn_test_hdf5_chk_0058	63.11 GB
	apro13_burn_test_hdf5_chk_0059	63.11 GB
	apro13_burn_test_hdf5_chk_0060	60.11 GB
	apro13_burn_test_hdf5_chk_0061	59.36 GB
	apro13_burn_test_hdf5_chk_0062	60.11 GB
	apro13_burn_test_hdf5_chk_0063	59.11 GB
	apro13_burn_test_hdf5_chk_0064	60.11 GB
	apro13_burn_test_hdf5_chk_0065	59.36 GB
	apro13_burn_test_hdf5_chk_0066	60.11 GB
	apro13_burn_test_hdf5_chk_0067	59.17 GB
	apro13_burn_test_hdf5_chk_0100	129.4 GB
	apro13_burn_test_hdf5_chk_0102	139.87 GB
	apro13_burn_test_hdf5_chk_0104	148.53 GB
	apro13_burn_test_hdf5_chk_0106	157.67 GB
	apro13_burn_test_hdf5_chk_0108	166.78 GB
	apro13_burn_test_hdf5_chk_0109	171.24 GB
	apro13_burn_test_hdf5_chk_0110	176.07 GB
	apro13_burn_test_hdf5_chk_0112	104.8 GB
	apro13_burn_test_hdf5_chk_0120	198.22 GB
	apro13_burn_test_hdf5_chk_0122	201.93 GB

select all | none [up one folder](#) [refresh list](#)

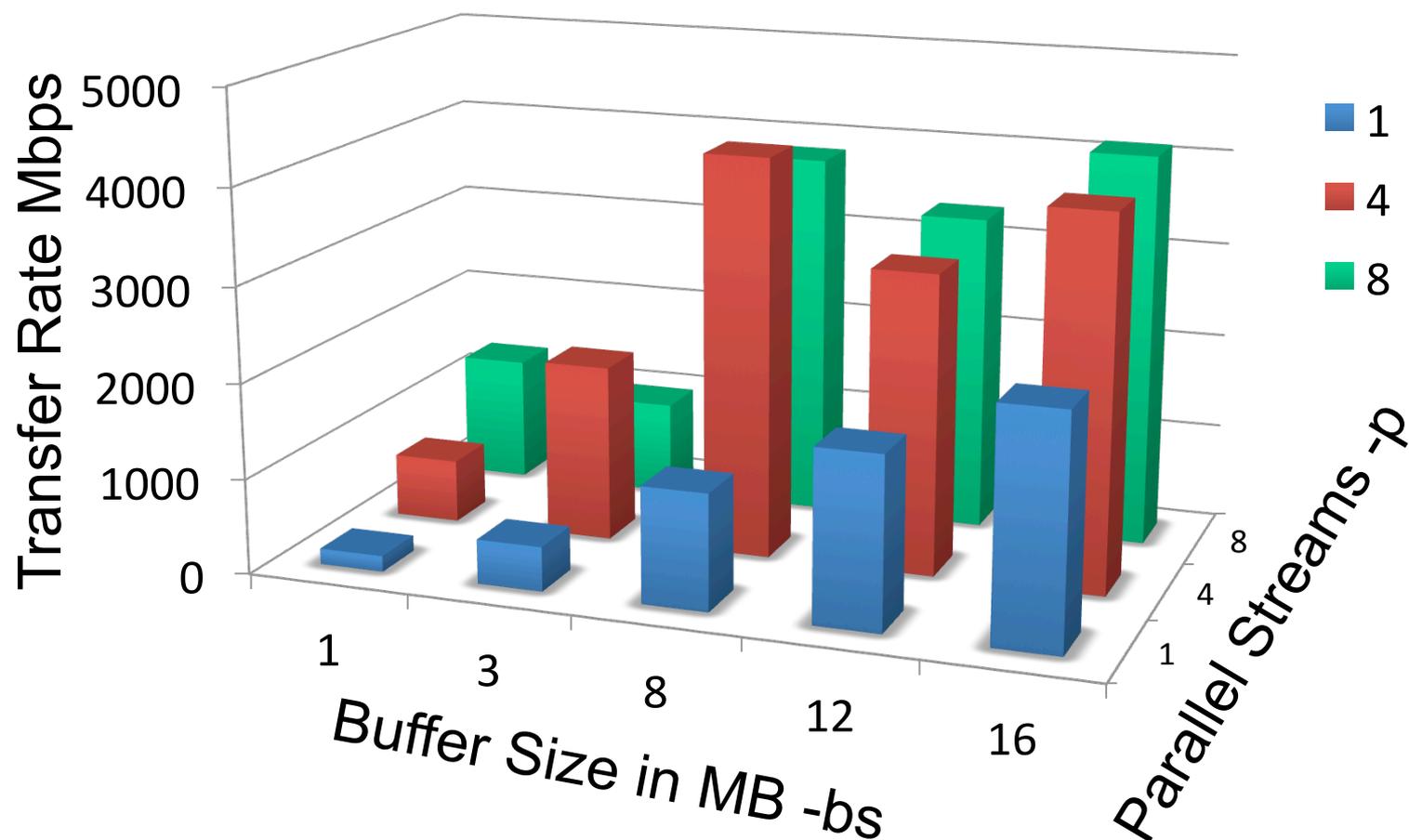
	yj7	Folder
	yy-CHIMERA	Folder
	yy-craypat	Folder
	yy_adios	Folder
	CRASH	264 b
	QE-BENCH.o1690101	13.46 kB
	QE-BENCH.o1690714	6.26 kB
	a.out	5.43 MB
	apro13_burn_test_hdf5_chk_0171	203.58 GB
	apro13_burn_test_hdf5_chk_0204	3.09 GB
	chimera-test-3d.tgz	17.19 MB
	core	95.36 MB
	input_tmp.in	0 b
	junk	280 b
	logfile	8.56 kB
	pw-gpu.x	9.74 MB
	rev1154_adios.tar	432.08 GB
	test_QE.pbs	455 b
	yy-CHIMERA.tar	32 kB

[more options](#) Label This Transfer

This will be displayed in your transfer activity.

globus_url_copy

```
globus-url-copy -tcp-bs 12M -bs 12M -p 4 gsiftp://  
dtn03.ccs.ornl.gov/$HOME/file1 gsiftp://remote.sys/tmp/shared/  
$USER/file2
```



Workflow for GridFTP

1. Command line (done once):
Set up and map OSG certificate

2. Command line (done every time*):
Setup proxy certificate on OLCF myproxy server from
dtn03 or dtn04

3. Transfer data

Globus: Web browser (done every time):
Go to Globusline.org, setup/active endpoints at both
ends of your transfer

Globus_url_copy: run from dtns or script from Titan

OSG Certificate Setup at OLCF

1. Go to <https://oim.grid.iu.edu/oim/certificaterequesthost> and fill out the certificate request form.

2. Retrieve grid certificate from your personal OSG Web link.

3. Move Yourcert.p12 to .globus directory in your home directory on dtn03 or dtn04.

4. Extract certificate and private key.

5. Register (map) your grid ID with OLCF and your remote transfer site.

Tip: Find a detailed walk-through in the extra slides

Workflow for GridFTP

1. Command line (done once):
Set up and map OSG certificate

2. Command line (done every time*):
Setup proxy certificate on OLCF myproxy server from
dtn03 or dtn04

3. Transfer data

Globus: Web browser (done every time):
Go to Globusline.org, setup/active endpoints at both
ends of your transfer

Globus_url_copy: run from dtns or script from Titan

Globus: Proxy Setup

The first time you use globus or globus online each day you must generate a proxy certificate from the command-line on dtn03 or dtn04 via two scripts.

- The first script generates the proxy certificate and requires your certificate pass phrase.

```
$ myproxy-init -n
```

```
Your identity: /DC=com/DC=DigiCert-Grid/O=Open Science Grid/  
OU=People/CN=test name 1234
```

```
Enter GRID pass phrase for this identity:
```

PASS PHRASE YOU SET UP

Globus: Proxy Setup

- The second script distributes the proxy certificate. It uses your RSA pin and token code as a the “MyProxy pass phrase”.

```
$ myproxy-logon
```

```
Enter MyProxy pass phrase: THIS IS YOUR PIN AND RSA TOKEN  
CODE
```

```
A credential has been received for username in /tmp/  
x123up_123456
```



Data Transfer

- Scriptable DTNs can be used as part of your data management workflow for methods that do not require interactive authentication.
 - HPSS transfers.
 - Transfers **to** places with single passwords (bbcp,scp).
 - Transfers with GridFTP and certificate authentication.
- Interactive DTNs are available for all OLCF transfer tools and should be user for local and remote transfers rather than Titan login nodes.
- For GridFTP: **yy-CHIMERA.tar**
 - Both ends of transfer must support GridFTP.
 - Proxy certificate only live 12 hours.
 - The grid certificate password is used to create the proxy and your pin and RSA token code is used to distribute the proxy and for most other things.
 - OLCF is forming a whitelist of remote globusonline endpoints.
 - OLCF has an Interactive DTN globusonline endpoint olcf#dtn.

Questions?

Extra Slides

Workflow for GridFTP

1. Command line (done once):
Set up and map OSG certificate

2. Command line (done every time*):
Setup proxy certificate on OLCF myproxy server from
dtn03 or dtn04

3. Transfer data

Globus: Web browser (done every time):
Go to Globusline.org, setup/active endpoints at both
ends of your transfer

Globus_url_copy: run from dtns or script from Titan

OSG Certificate Setup at OLCF

1. Go to <https://oim.grid.iu.edu/oim/certificaterequesthost> and fill out the certificate request form.

2. Retrieve grid certificate from your personal OSG Web link.

3. Move Yourcert.p12 to .globus directory in your home directory on dtn03 or dtn04.

4. Extract certificate and private key.

5. Register (map) your grid ID with OLCF and your remote transfer site.

OSG Web Form

- 🏠 Home
- USER CERTIFICATES**
- + Request New**
- 🔍 Search
- HOST CERTIFICATES**
- + Request New
- 📁 GridAdmins
- 🔍 Search

Please submit a [GOC Ticket](#) for assistance.

User Certificate Request

Contact Information

Following information will be used to issue your new user certificate, and also used to contact you during the process.

Full Name

 * Required

Phone

 * Required

Email

 * Required

Profile Information

Following information will be used to register you as a new OIM user.

City

 * Required

State

 * Required

Zipcode

 * Required

Country

 * Required

OSG Web Form

Profile - Introduce yourself to the rest of OSG community.

Please enter your role within OSG community, and maybe a small introduction of who you are and what you do.

* Required

List ORNL contact here: PI, Scientific Computing Liaison

Choose a password

Please choose a password to issue your certificate and encrypt your private key.

IMPORTANT: If you forget this password, you will not be able to issue your certificate and import it your browser after it is approved.

* Required

Re-enter password

* Required

OSG Certificate Setup at OLCF

1. Go to <https://oim.grid.iu.edu/oim/certificaterequesthost> and fill out the certificate request form.

2. Retrieve grid certificate from your personal OSG Web link.

3. Move Yourcert.p12 to .globus directory in your home directory on dtn03 or dtn04.

4. Extract certificate and private key.

5. Register (map) your grid ID with OLCF and your remote transfer site.

OSG Certificate Setup

OSG will send you an email containing a certificate retrieval link:

<https://oim.grid.iu.edu/oimcertificateuser?id=###>.

Requester to issue certificate & download

Please enter the password you chose during a request submission to retrieve your certificate & encrypt your private key

Password

 * Required

 Issue Certificate ...

- Login to dtn03.ccs.ornl.gov, load the globus module, and make a directory called .globus in you home directory. Copy the certificate from your workstation to dtn03.

```
$ scp YourCert.p12 username@dtn03.ccs.ornl.gov:/ccs/home/  
username/.globus
```

OSG Certificate Setup at OLCF

1. Go to <https://oim.grid.iu.edu/oim/certificaterequesthost> and fill out the certificate request form.

2. Retrieve grid certificate from your personal OSG Web link.

3. Move Yourcert.p12 to .globus directory in your home directory on dtn03 or dtn04.

4. Extract certificate and private key.

5. Register (map) your grid ID with OLCF and your remote transfer site.

OSG Certificate Extraction

You will need your OSG issue password for both extractions.

- To extract your certificate on dtn03 type:

```
$ openssl pkcs12 -in YourCert.p12 -clcerts -nokeys -out $HOME  
/.globus/usercert.pem
```

- To extract the encrypted private key type:

```
$ openssl pkcs12 -in YourCert.p12 -nocerts -out $HOME/.globus/  
userkey.pem
```

You will be asked to generate and confirm a “PEM” passphrase. Remember this passphrase. You will need both it and your RSA token code to use your certificate.

OSG Certificate Setup

- Ensure appropriate permissions are set on `usercert.pem` and `userkey.pem`

```
$ chmod 644 $HOME/.globus/usercert.pem  
$ chmod 600 $HOME/.globus/userkey.pem
```

- Ensure that your `.p12` file is securely stored or remove it if you have a backup on your local computer

OSG Certificate Setup at OLCF

1. Go to <https://oim.grid.iu.edu/oim/certificaterequesthost> and fill out the certificate request form.

2. Retrieve grid certificate from your personal OSG Web link.

3. Move Yourcert.p12 to .globus directory in your home directory on dtn03 or dtn04.

4. Extract certificate and private key.

5. Register (map) your grid ID with OLCF and your remote transfer site.

OSG Certificate Registration (Mapping)

- This is done from a script on dtn03 or dtn04:

```
$ module load globus  
$ register_globus_creds
```

- You will receive an email once your certificate has been successfully registered.
- **The whole certificate process should take less than a week.**

Workflow for GridFTP

1. Command line (done once):
Set up and map OSG certificate

2. Command line (done every time*):
Setup proxy certificate on OLCF myproxy server from
dtn03 or dtn04

3. Transfer data

Globus: Web browser (done every time):
Go to Globusonline.org, setup/active endpoints at both
ends of your transfer

Globus_url_copy: run from dtns or script from Titan

Globus: Proxy Setup

The first time you use globus or globus online each day you must generate a proxy certificate from the command-line on dtn03 or dtn04 via two scripts.

- The first script generates the proxy certificate and requires your certificate pass phrase.

```
$ myproxy-init -n
```

```
Your identity: /DC=com/DC=DigiCert-Grid/O=Open Science Grid/  
OU=People/CN=test name 1234
```

```
Enter GRID pass phrase for this identity:
```

PASS PHRASE YOU SET UP

Globus: Proxy Setup

- The second script distributes the proxy certificate. It uses your RSA pin and token code as a the “MyProxy pass phrase”.

```
$ myproxy-logon
```

```
Enter MyProxy pass phrase: THIS IS YOUR PIN AND RSA TOKEN  
CODE
```

```
A credential has been received for username in /tmp/  
x123up_123456
```



Workflow for GridFTP

1. Command line (done once):
Set up and map OSG certificate

2. Command line (done every time*):
Setup proxy certificate on OLCF myproxy server from
dtn03 or dtn04

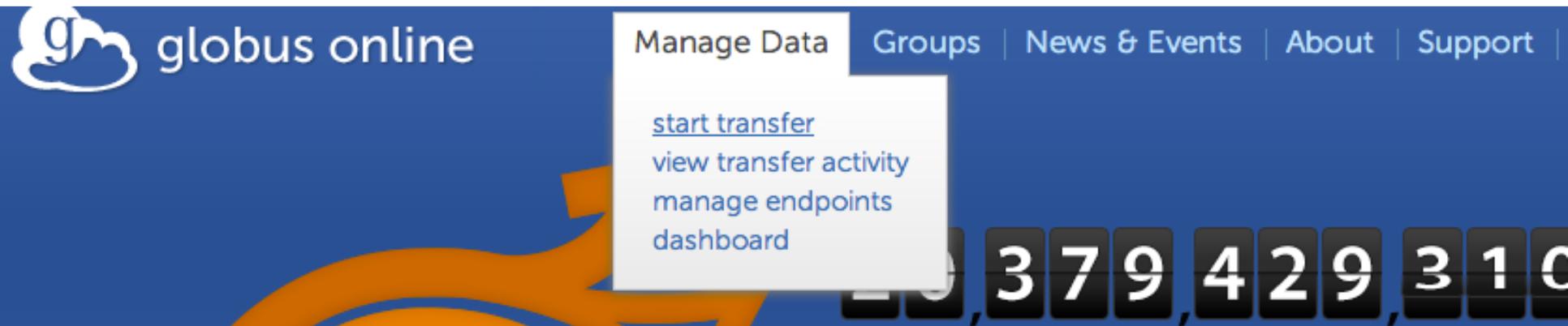
3. Transfer data

Globus: Web browser (done every time):
Go to Globusline.org, setup/active endpoints at both
ends of your transfer

Globus_url_copy: run from dtns or script from Titan

Steps for Globus (Online)

- Go to <https://www.globus.org> and login.



- Click "start transfer".
- Enter olcf#dtn in one endpoint window and the name of the remote endpoint in the other.

A screenshot of the Globus endpoint configuration form. It has two rows. The first row is labeled "Endpoint" and contains a text input field with "olcf#dtn" entered, a "..." button, and a "Go" button. The second row is labeled "Path" and contains an empty text input field and a "Go" button.

Steps for Globus Online

Activate Endpoint: olcf#dtn ✕

Please Enter Your Credentials

*MyProxy Server

*Username

*Passphrase

Server DN

Credential Lifetime (hours)

* Required



Using Globus_url_copy

- Your grid ID may need to be registered at both ends of the transfer.
- You must create a proxy certificate once a day from dtn03 or dtn04 as shown for Globus Online.
- The commands used are the globus-url-copy.
- globus-url-copy [options] Source_URL/file remote_URL/dir/file

-bs	Buffer size of transfer method
-tcp-bs	Tcp buffer size
-p	Number of parallel streams

- Example:

```
globus-url-copy -tcp-bs 12M -bs 12M -p 4 gsiftp://  
dtn03.ccs.ornl.gov/$HOME/file1 gsiftp://remote.sys/tmp/shared/  
$USER/file2
```