University of Wyoming
Advanced Research Computing Center (ARCC)
arcc-info@uwyo.edu
(307) 766-7748

Presented Sept 28, 2017
Jeff Lang
System Architect
and Administrator

Jared Baker
Application Optimization
and Development

Mike Killean
Storage Architect

Jeff Lang
System Architect
and Administrator

Dylan Perkins
Database Analyst

Karen Milberger
Data Ombudsman

Nikhil Shetty
3D Visualization Specialist

Introduction to Globus
Topics

• Introduction:
  Features
  Endpoints
  Access (credentials, logging in)

• Data Transfer:
  Web interface
  Command line interface (CLI)
  Globus Plus

• Discussion
Introduction

Globus can transfer data between two machines – two servers, a server and a personal machine, or (using Globus Plus) two personal machines. It is ideal for large files and available for many institutional clusters and national labs.

When the amount of data to transfer exceeds around 100 GB, other methods like scp, sftp, rsync may be too slow, and Globus will be faster for transferring collections of files due to doing so in parallel.

Globus website: https://www.globus.org
References: https://docs.globus.org
https://docs.globus.org/how-to/get-started
https://docs.globus.org/cli
Advantages

- Secure, handles errors, verifies integrity of transferred files
- Automatically resumes after interruption
- Emails user when transfer is complete, or when error occurs
- Can check status at any time
- Accelerated transfer rates – transfers in parallel where available
- Web and command line interfaces for transfer
- Links to major HPC sites (ORNL, TACC, NCAR, etc.)

So, is a reliable method of transfer and is recommended for transferring to/from petaLibrary, Bighorn, and other research sites (between two servers, or server and desktop/laptop).
An endpoint is a location that data can be transferred from/to:

- A server endpoint is set up by an administrator to provide access to a system via Globus.

- A personal endpoint is set up by an individual on their computer to transfer files between it and other endpoints. To create a personal endpoint, the “Globus Connect Personal” software must be installed on the computer. It is available from globus.org and free to use at non-profit research and education institutions.
# Some Server Endpoints

## UW endpoints:

<table>
<thead>
<tr>
<th>Endpoint Description</th>
<th>Endpoint Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bighorn (Mount Moran's storage system)</td>
<td>ARCC Bighorn</td>
</tr>
<tr>
<td>petaLibrary</td>
<td>ARCC petaLibrary</td>
</tr>
</tbody>
</table>

## Other endpoints:

<table>
<thead>
<tr>
<th>Endpoint Description</th>
<th>Endpoint Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAR GLADE (XSEDE): (Yellowstone and Cheyenne)</td>
<td>XSEDE NCAR GLADE</td>
</tr>
<tr>
<td>University of Utah:</td>
<td>Search for “uofuchpc”, list will show up</td>
</tr>
<tr>
<td>CI-Water:</td>
<td>University of Utah – CIWater Dedicated Endpoint</td>
</tr>
</tbody>
</table>
Introduction to Globus

File Transfer

1. You submit a transfer request
2. Globus moves the data for you
3. Globus notifies you once the transfer is complete
Access

Logging in

1) Go to globus.org and click the “Log in” button.

2) In the drop-down list, select “University of Wyoming” and click Continue.

3) This redirects to a UWyo login page, login as normal there. (The first time logging in, follow the prompts to complete account set up, including accepting Terms of Service.)

4) After logging in, get to a Transfer Files page, from which various types of file transfers can be done.
Introduction to Globus

Make a Personal Endpoint

1) Download and install “Globus Connect Personal”:
   - Go to globus.org → “Products” → “Globus Connect” → “Get Globus Connect Personal”, and choose the version for Mac, Linux, or Windows.
   - Follow the instructions on the website to create an endpoint. This involves logging in, downloading and running the installer for your selected version, generating a “Setup Key” on the website, and entering it when prompted during installation. (Note that on Linux, a “Security Code” is requested – enter the “Setup Key” here.)

2) When installation is complete, Globus Connect Personal will start running – shows up as a window (Linux) or icon in running apps (other systems). This program will need to be running whenever the endpoint is used, and needs to show a status of “Connected”.

![Globus Connect Personal screen shot]
Make a Personal Endpoint

3) On Linux and Mac systems, the above process didn't have a way to select which directory/directories should be used for the endpoint – it just automatically used the home directory. This can be changed after installation by going to **File → Preferences** (Linux), or from the Globus icon, **Preferences... → Access** (Mac). The Windows installation process does allow this selection to be made during installation – if it needs to be changed later, that can be done through **Options... → Access** from the main menu.
File transfer – Web interface

1) Login to the Globus website as described above.

2) If you don't automatically get to the Transfer File page, then select it in the menu at the top of the page.

3) Select the two endpoints that data will be transferred between
   - If either one is a personal endpoint, make sure that Globus Connect Personal is running.
   - To use a server endpoint, it will be necessary to login to access it. For the ARCC servers, this uses two-factor authentication.

4) Once an endpoint is selected, double-click on the necessary folder/directory links to navigate to the locations between which the transfer will occur. Select the file(s)/directory(ies) to transfer with a single-click. This will make one of the two gray arrows at the top of the page change to blue. To initiate the transfer, click the blue arrow.
Introduction to Globus

File transfer – Web interface

Endpoint: personal-endpoint
Path: /xxxxxxxxxx/personal-endpoint

Endpoint: ARCC petaLibrary
Path: /petaLibrary/homes/xxxxxxxx/

Select all ○ up one folder ○ refresh list

- LoggerData
- junk.txt

Select all ○ up one folder ○ refresh list

- Desktop
- Documents
- Downloads
- Music

Label This Transfer

Transfer Settings
- sync - only transfer new or changed files
- delete files on destination that do not exist on source
- preserve source file modification times
- verify file integrity after transfer
- encrypt transfer

Get Globus Connect Personal
Turn your computer into an endpoint.
File transfer – Web interface

**Activity page:** shows recent transfers with their status (completed, in progress, canceled, interrupted). To see details, like effective speed or progress, click the link for that transfer.

**Email notification:**
- Sent when transfer is complete. If a transfer is interrupted, automatically retries for up to 3 days, and will resume rather than retransmitting data already copied.
- If an interrupted transfer can't be resumed for 3 days, an email notification of failure is sent.
- If transfer can't complete within authentication expiration period of an endpoint, an email notice to re-authenticate is sent.
File transfer – Command line interface

• The new Globus CLI is available. Installation instructions are at https://docs.globus.org/cli/installation/

• Globus (or “new”) CLI: Must install an application called `globus` which is used instead of `ssh`. Doesn’t restrict IDs to only Globus IDs, but does require logging in on website. For endpoints that require activation, can be done via website or CLI.

• Legacy (or “hosted”) CLI: Required using Globus ID instead of institutional ID, and setting up SSH keys. Data transfer was done via the `ssh` command, using an interactive `ssh` session or by running a command directly over `ssh`.

• Reference: https://docs.globus.org/cli/
Globus CLI

System requirements:

- Python 2.7+ or 3.3+
- "python pip" – a package manager for python (used similarly to yum or apt-get)
- "virtualenv" – a python package for creating isolated python environments

Installation:

1) Install "python pip" if needed (only needed for Python >= 2.7 and < 2.7.9, for versions 2.7.9+ is included with Python):

   sudo yum install python-pip

2) Install "virtualenv" if needed ("which virtualenv" is easy way to check):

   sudo pip install virtualenv

3) Use virtualenv to create an isolated python environment and install the CLI in it:

   ...
Globus CLI

Installation, cont:

1) Use virtualenv to create an isolated python environment and install the CLI in it:
   virtualenv "$HOME/.globus-cli-virtualenv"
   source "$HOME/.globus-cli-virtualenv/bin/activate"
   pip install globus-cli
   deactivate

2) Add location of newly installed CLI to path:
   export PATH="$PATH:$HOME/.globus-cli-virtualenv/bin"
   To end of $HOME/.bashrc (or other init file), add the line:
   export PATH="$PATH:$HOME/.globus-cli-virtualenv/bin"

3) Check if the CLI is installed:
   globus --help
Globus CLI

Transferring Data

To use the CLI to view data, transfer data, or look up info, start by logging in:
globus login

- Opens a web page which requires normal Globus login – follow instructions on page to accept the CLI login – when complete, above CLI command will complete (If browser isn't detected, will display link to copy-and-paste into browser.)
- Prompt looks like regular OS prompt, but can run globus commands
- The login applies to a machine+user, so after logging in can run globus commands from any window, and login persists through reboot
- Will remain logged in until:
  - logout with “globus logout” command, or
  - 6 months after last use of CLI, or
  - cancel CLI access while logged in to Globus account on website
Using the CLI

- Endpoints are always identified by ID (not name)
- For **server endpoints**: Commands that access contents of endpoint (like 'ls' and 'transfer') require endpoint to be activated first - can either login to it in the web interface, or use 'globus endpoint activate' command
- For **personal endpoints**: The "globusconnect" program must be running and the endpoint must have a "Connected" status.

Useful commands

- `globus whoami`: see login status/user
- `globus endpoint search 'ARCC'`: search for a known endpoint
- `globus endpoint search "Globus Tutorial Endpoint"`
- `globus endpoint activate --myproxy '<endpoint_id>'`
- `globus ls '<endpoint_id>'`: if no path given, shows listing of default path
- `globus ls '<endpoint_id>:/full/path/to/dir/'`: must be dir, not file
Globus CLI

Transfer a file

To transfer a file from a server endpoint to a personal endpoint, run 'globus endpoint search' commands for the endpoint names to find their IDs, then:

```
  globus transfer 'srv_endpt_id:/share/godata/file2.txt' 'prsnl_endpt_id:/full/path/to/file2.txt'
```

Notes:

- Must give destination file name even if it is the same as source file name
- The above command submits a transfer request and then returns – it doesn't wait for transfer to complete
- When a transfer request is successfully submitted, the Task ID for the transfer is displayed – can use to see task status and details
- Can use 'globus ls' command before transfer to make sure file/path names are correct
Globus CLI

Transfer a directory

To transfer a directory from a server endpoint to a personal endpoint, run 'globus endpoint search' commands for the endpoint names to find their IDs, then:

```
globus transfer -r 'srv_endpt_id:/share/godata' 'prsnl_endpt_id:/full/path/new_dir'
```

Notes:

- If "new_dir" already exists at destination, it will be replaced by "godata" (renamed to "new_dir"), rather than putting "godata" into "new_dir"
- All transfer options that can be selected in the web interface can also be used in the CLI – see 'globus transfer -h' for details
- There are additional globus commands for changing files/directories at endpoints (e.g. mkdir, rename, delete) – see 'globus -h' for details
Globus CLI

Viewing task details

When a task has been created, some useful commands are:

- `globus task list`: show all current and recent tasks, with status, of user
- `globus task show <task-ID>`: show details about a particular task
- `globus task event-list <task-ID>`: can see progress of transfer, and if there are any errors, details about them

If a task is found to have an unrecoverable error (like trying to use a path or endpoint that doesn't exist), can cancel it (rather than having it retry for 3 days before failing):

- `globus task cancel <task-ID>`
Introduction to Globus

Globus CLI

Using Globus CLI for automation

- Do "globus login" and complete login on website once as part of initial setup – login won't expire as long as used more often than every 6 months
- CLI activation of endpoints should normally not be done (and may not be possible) through a script:
  - putting a password in a script (or file read by a script) is insecure
  - for endpoints that use 2-factor authentication (such as “ARCC Bighorn” and “ARCC petaLibrary”), it isn't possible to get the 2nd factor into a script
- The recommended method of scripting a transfer involving a server endpoint is to make a shared endpoint on the required directory, and use the share instead of the actual endpoint. (Shared endpoints don't need to be activated.)
- If a script does an automatic transfer, it can be run at scheduled times through cron/crontab (on Linux/Unix), or other scheduling software on other systems. (Please contact ARCC if need help with this.)
Globus Plus

Enables:
• Sharing a personal endpoint to any Globus user or group
• Transfer between two personal endpoints

To upgrade account to Globus Plus:
• Login at globus.org and go to the Account page – it should contain an Identities section followed by a Globus Plus section. Under Globus Plus, the presence or absence of a message like “You have Globus Plus, because...” will show whether or not it is a Globus Plus account.
• If you need to request Globus Plus access, in the Globus Plus section, select an organization from the drop-down list. Note that “University of Wyoming” is NOT an available choice here, and “Advanced Research Computing Center” is the option that needs chosen. Follow the instructions on the page loaded by making this choice, or email arcc-help@uwyo.edu if you have any questions or problems with this.
Globus Plus

Sharing a personal endpoint:
- Set endpoint configuration to allow sharing: In the Globus Connect Personal menu, go to File → Preferences (Linux), Preferences… → Access (Mac), Options… → Access (Windows), and click to turn on Shared or Shareable.
- Ensure that Globus Connect Personal is running and has a status of Connected.
- Login at globus.org and go to the Transfer Files page. For an endpoint, select the personal endpoint to share. Navigate to the particular folder to be shared, and single-click the folder name to select it. Click the menu icon in the top right corner of the navigation box and choose the option “share”. Follow the prompts to select the user(s)/group(s) and permissions for sharing.

Transferring data between two personal endpoints:
- Make sure that the machines hosting both personal endpoints are on, and that Globus Connect is running and has “Connected” status on both of them.
- Using the web or command line interface, initiate a transfer between the endpoints as normal.
Discussion

What are your needs for data transfer, storage, and management?