Globus for System Administrators

Vas Vasiliadis
vas@uchicago.edu

May 2, 2019
Globus Connect Server

- Makes your storage accessible via Globus
- Multi-user server, installed and managed by sysadmin
- Default access for all local accounts
- Native packaging
  Linux: DEB, RPM

docs.globus.org/globus-connect-server-installation-guide/
Globus Connect Server

Non-POSIX Connectors

POSIX-compliant Connector

Local Storage System (HPC cluster, NAS, ...)

MyProxy CA
OAuth Server
GridFTP Server

DTN

Local system users
Storage Connectors - globus.org/connectors

Current
- IBM Spectrum Scale
- SPECTRA
- ceph
- Amazon web services™ S3
- Google Drive
- Lustre
- box
- HPSS
- Western Digital.™ ACTIVE SCALE™
- hadoop

Planned
- Google Cloud
- Microsoft Azure
- wasabi

hot cloud storage
Globus Connect Server v5 Milestones

- **v5.0:** Google Drive
- **v5.1:** POSIX guest collections, HTTPS
- **v5.2:** High assurance
- **v5.3:** Version Unification
- **v5.x:** ...
- **v5.x+:** v4 feature parity+

Other features:
- Multi DTN support
- Additional connectors
- Custom IdPs
- …
Which version of Globus Connect Server do I use?

By default, assume you should use GCS v4.
Which version of Globus Connect Server do I use?

Are you a Globus subscriber?

- Yes → Do you have a High Assurance or BAA subscription (and plan to set up a high assurance endpoint)?
  - Yes → Use GCS v5.3
  - No → Do you plan to use the S3, Box, Google Drive or Ceph connectors?
    - Yes → Would you like to support multiple storage system types on one endpoint?
      - Yes → Do you want to enable HTTPS access to your storage?
        - Yes → Use GCS v4
        - No → Use GCS v4
      - No → Use GCS v4
    - No → Use GCS v4
  - No → Use GCS v4
Creating a Globus endpoint on your server

• In this example, Server = Amazon EC2 instance
• Installation and configuration of Globus Connect Server requires a Globus ID
• Go to globusid.org
• Click “create a Globus ID”
  – Optional: associate it with your Globus account
What we are going to do:

1. Install Globus Connect Server
   - Access server as user "campusadmin"
   - Update repo
   - Install package
   - Setup Globus Connect Server

2. Log into Globus

3. Access the newly created endpoint (as user 'researcher')

4. Transfer a file
Access your server

• Get the IP address for your EC2 server ([bit.ly/ec2ip](bit.ly/ec2ip))
• Log in as user ‘campusadmin’
  
  `ssh campusadmin@<EC2_instance_IP_address>`
• Please `sudo su` before continuing
  – User ‘campusadmin’ has passwordless sudo privileges
Install Globus Connect Server

$ sudo su
$ curl -LOs
http://downloads.globus.org/toolkit/globus-connect-server/globus-connect-server-repo_latest_all.deb
$ dpkg -i globus-connect-server-repo_latest_all.deb
$ apt-get update
$ apt-get -y install globus-connect-server
$ globus-connect-server-setup

You have a working Globus endpoint!

Use your Globus ID username and password when prompted
Access the Globus endpoint

• Go to Manage Data → Transfer Files

• Access the endpoint you just created
  – Search for your EC2 host name in the Endpoint field
  – Log in as “researcher”; you will see the user’s home directory

• Transfer files between a test endpoint (e.g. ESnet read-only) and your EC2 endpoint
Globus accounts and endpoint access

• **Globus account:** Primary identity (+ Linked Identities)
• **Endpoint initially accessible by creator**
• **Endpoint not visible?**
  – Primary identity is your institutional ID?
  – Link your Globus ID!
Configuring Globus Connect Server
Endpoint configuration

• On the Globus service: app.globus.org/endpoints

• On your DTN: /etc/globus-connect-server.conf
  – Standard .ini format: [Section] Option = Value
  – To enable changes run globus-connect-server-setup
  – “Rinse and repeat”
Common configuration options

• **Endpoints page**
  – Display Name
  – Visibility
  – Encryption

• **DTN configuration file**
  – RestrictPaths
  – Sharing
  – IdentityMethod (CILogon, Oauth)
  – SharingRestrictPaths
Exercise: Make your endpoint visible

• **Edit endpoint attributes**
  – Change the name to something useful, e.g. `<your_name> EC2 Endpoint`
  – For the “Visible To” attribute select “Public - Visible to all users”

• **Find your neighbor’s endpoint**
  – Thanks to our superb security …you can access it too 😊
Path Restriction

- **Default configuration:**
  - All paths allowed, access control handled by the OS

- **Use RestrictPaths to customize**
  - Specifies a comma separated list of full paths that clients may access
  - Each path may be prefixed by R (read) and/or W (write), or N (none) to explicitly deny access to a path
  - ‘~’ for authenticated user’s home directory, and * may be used for simple wildcard matching.

- **e.g. Full access to home directory, read access to /data:**
  - `RestrictPaths = RW~,R/data`

- **e.g. Full access to home directory, deny hidden files:**
  - `RestrictPaths = RW~,N~/.*`
Exercise: Restrict access

- **Set** `RestrictPaths=RW~,N~/archive`
- Run `globus-connect-server-setup`
- Access your endpoint as ‘researcher’
- What’s changed?
Enabling sharing on an endpoint

- **In config file, set** `Sharing=True`
- Run `globus-connect-server-setup`
- Flag endpoint as “managed” (in web app or via CLI)

* Note: Creation of shared endpoints requires a Globus subscription for the managed endpoint
Limit sharing to specific accounts

- `SharingUsersAllow =`
- `SharingGroupsAllow =`
- `SharingUsersDeny =`
- `SharingGroupsDeny =`
Sharing Path Restriction

• Restrict paths where users can create shared endpoints
• Use `SharingRestrictPaths` to customize
  – Same syntax as `RestrictPaths`
• e.g. Full access to home directory, deny hidden files:
  – `SharingRestrictPaths = RW~,N~/.*`
• e.g. Full access to public folder under home directory:
  – `SharingRestrictPaths = RW~/public`
• e.g. Full access to `/proj`, read access to `/scratch`:
  – `SharingRestrictPaths = RW/proj,R/scratch`
Endpoint Access Control/Activation
Ports needed for Globus

- Inbound: 2811 (control channel)
- Inbound: 7512 (MyProxy), 443 (OAuth)
- Inbound: 50000-51000 (data channel)
- If restricting outbound connections, allow connections on:
  - 80, 2223 (used during install/config)
  - 50000-51000 (GridFTP data channel)
Endpoint activation using MyProxy

Default configuration (avoid if at all possible)
Endpoint activation using MyProxy OAuth

Best practice configuration
Single Sign-On with InCommon/CILogon

• Your Shibboleth server must release R&S attributes to CILogon—especially the ePPN attribute

• Local account must match institutional ID (InCommon ID)
  – Test by creating a local user with same name

• In /etc/globus-connect-server.conf set:
  
  AuthorizationMethod = CILogon
  
  CILogonIdentityProvider = 
  <institution_listed_in_CILogon_IdP_list>
Managed endpoints and subscriptions
Subscription configuration

• **Subscription manager**
  – Create/upgrade managed endpoints
  – Requires Globus ID linked to Globus account

• **Management console permissions**
  – Independent of subscription manager
  – Map managed endpoint to Globus ID

• **Globus Plus group**
  – Subscription Manager is admin
  – Can grant admin rights to other members
Creating managed endpoints

• **Required** for sharing, management console, reporting, …
• Convert existing endpoint to managed via CLI (or web):
  
globus endpoint update --managed <endpt_uuid>
• Must be run by subscription manager
• **Important:** Re-run endpoint update after deleting/re-creating endpoint
Monitoring and managing Globus endpoint activity
Management console

• Monitor all transfers
• Pause/resume specific transfers
• Add pause conditions with various options
• Resume specific tasks overriding pause conditions
• Cancel tasks
• View sharing ACLs
Endpoint Roles

• **Administrator**: define endpoint and roles
• **Access Manager**: manage permissions
• **Activity Manager**: perform control tasks
• **Activity Monitor**: view activity
Demonstration:
Management console
Endpoint Roles
Usage Reporting
...on performance
Balance: performance - reliability

- Network use parameters: concurrency, parallelism
- Maximum, Preferred values for each
- Transfer considers source and destination endpoint settings
  \[
  \min(\max(\text{preferred src}, \text{preferred dest}), \max \text{ src}, \max \text{ dest})
  \]
- Service limits, e.g. concurrent requests
Deployment Scenarios
Common endpoint configuration (GCSv4)
Common endpoint configuration (GCSv4)
Multi-endpoint configuration (GCSv4)

- Data Transfer Node
  - POSIX Connector
  - Western Digital ActiveScale Connector

- File Systems
  - ext* XFS ZFS
  - GPFS Lustre

- Directories
  - ~/projects
  - ~/archive
  - ~/scratch
Multi-endpoint configuration (GCSv4)

- Data Transfer Node
- POSIX Connector
- Western Digital ActiveScale Connector
- Amazon S3 Connector

- ~/vault
- ~/archive
- ~/projects
- ~/scratch

- ext* XFS ZFS
- GPFS Lustre
- Amazon S3 Bucket
- Amazon S3 Connector
Network paths

- Separate control and data interfaces
- "DataInterface =" option in globus-connect-server-conf
- Common scenario: route data flows over Science DMZ link
Dual-homed DTN – high speed data path

Science DMZ

Data Transfer Node

GridFTP Server

Internet2 path

Data Channel

GridFTP Server

Control Channel

Control Channel
Dual-homed DTN – internal data path

- Science DMZ
- Data Transfer Node
- GridFTP Server
- Control Channel
- LAN/Intranet path
- Data Channel
- Firewall
- Data Transfer Node
- GridFTP Server
- Control Channel
Other Deployment Options
Encryption

• Requiring encryption on an endpoint
  – User cannot override
  – Useful for “sensitive” data

• Globus uses OpenSSL cipher stack as currently configured on your DTN

• FIPS 140-2 compliance: ensure use of FIPS capable OpenSSL libraries on DTN

Distributing Globus Connect Server components

- **Globus Connect Server components**
  - `globus-connect-server-io, -id, -web`

- **Default**: `-io, –id` and `-web` on single server

- **Common options**
  - Multiple `-io` servers for load balancing, failover, and performance
  - No `-id` server, e.g. third-party IdP
  - `-id` on separate server, e.g. non-DTN nodes
  - `-web` on either `-id` server or separate server for OAuth interface
Distributing Globus Connect Server components

Science DMZ (ACL limited)
Port 2811 accepts inbound connections from Globus

Data Transfer Node

GridFTP Server

MyProxy CA

ext* XFS ZFS

OAuth Server

Firewall

Microsoft AD

OpenID Connect
Setting up multiple –io servers

• **Guidelines**
  – Use the same .conf file on all servers
  – First install on the server running the –id component, then all others

• **Install Globus Connect Server on all servers**

• **Edit .conf file on one of the servers and set [MyProxy] Server to the hostname of the server you want the –id component installed on**

• **Copy Globus Connect Server configuration file to all servers**

• **Run globus-connect-server-setup on the server running the –id component**

• **Run globus-connect-server-setup on all other servers**

• **Repeat steps 2-5 as necessary to update configurations**
Example: Two-node DTN

On “primary” DTN node (34.20.29.57):
/etc/globus-connect-server.conf
[Endpoint] Name = globus_dtn
[MyProxy] Server = 34.20.29.57

On other DTN nodes:
/etc/globus-connect-server.conf
[Endpoint] Name = globus_dtn
[MyProxy] Server = 34.20.29.57
Open Discussion