Globus Endpoint Administration

Rachana Ananthakrishnan - rachana@globus.org
Greg Nawrocki - greg@globus.org

Johns Hopkins University
April 11, 2019
The DTN and Globus Connect Server

Current - Full feature set: GCS 4.x

Future – Some features currently available: GCS 5.x
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/](docs.globus.org/globus-connect-server-installation-guide/)
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)
- 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 to/from 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

- **Globus service “Manage Endpoints” page**
- **DTN (Globus Connect Server) config**
  
  `/etc/globus-connect-server.conf`
  - Standard .ini format: `[Section] Option = Value`
  - To enable changes you must run:
    
    `globus-connect-server-setup`
  - “Rinse and repeat”
Common configuration options

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

• **DTN configuration file**
  – RestrictPaths
  – IdentityMethod (CILogon, OAuth)
  – Sharing
  – 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`
- **Use the web app to flag as managed endpoint**

* 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
Accessing Endpoints
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)
Network Paths - Illustrative

* Please see TCP ports reference: https://docs.globus.org/resource-provider-guide/#open-tcp-ports_section
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: \(<\text{RandS}>1</\text{RandS}>\)
- Local account must match institutional ID (InCommon ID)
  - Test by creating a local user with same name
- `ln /etc/globus-connect-server.conf set:
  AuthorizationMethod = CILogon`
  
  `CILogonIdentityProvider = <institution_listed_in_CILogon_IdP_list>`
  
  `https://cilogon.org/include/idplist.xml`
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
Petascale DTN Project

November 2017
L380 Data Set

Illustrative performance

Gigabits per second (min/avg/max), three transfers

ALCF DTN cluster
Globus endpoint: alcf#dtn_mira
Filesystem: /projects

33.0/35.0/37.8 Gbps

44.1/46.8/48.4 Gbps

NERSC DTN cluster
Globus endpoint: nersc#dtn
Filesystem: /project

41.0/42.2/43.9 Gbps

43.0/50.0/56.3 Gbps

35.9/39.0/40.7 Gbps

34.6/47.5/56.8 Gbps

41.2/22.6/24.5 Gbps

55.4/56.7/57.4 Gbps

33.2/43.4/50.3 Gbps

OLCF DTN cluster
Globus endpoint: olcf#dtn_atlas
Filesystem: atlas2

29.9/33.1/35.5 Gbps

NCSA DTN cluster
Globus endpoint: ncsa#BlueWaters
Filesystem: /scratch

Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (110G bytes)
Size distribution:
    1 - 10 bytes: 7 files
    10 - 100 bytes: 1 files
    100 - 1K bytes: 59 files
    1K - 10K bytes: 3170 files
    10K - 100K bytes: 1560 files
    100K - 1M bytes: 2817 files
    1M - 10M bytes: 3901 files
    10M - 100M bytes: 3800 files
    100M - 1G bytes: 2295 files
    1G - 10G bytes: 1647 files
    10G - 100G bytes: 3 files
Disk-to-Disk Throughput: ESnet Testing

- **GridFTP (4 streams)**
- **GridFTP (1 stream)**
- **sftp**
- **scp (w/HPN)**
- **scp**

- Berkeley, CA to Argonne, IL (RTT: 53 ms, Capacity: 10Gbps)
- scp is 24x slower than GridFTP on this path
- >1 Gbps (125 MB/s) disk-to-disk requires RAID array

Source: ESnet (2016)
Deployment Scenarios
Best-practice deployment

- **WAN**
- **WAN**
- **10G**
- **10GE**
- **10GE**
- **10GE**
- **Border Router**
- **Enterprise Border Router/Firewall**
- **Clean, High-bandwidth WAN path**
- **Site / Campus access to Science DMZ resources**
- **Science DMZ Switch/Router**
- **High performance Data Transfer Node with high-speed storage**
- **Per-service security policy control points**
- **perfSONAR**
- **Details at: fasterdata.es.net**
The Data Transfer Node

On-premises Data Transfer Node (DTN)

- Data Storage Interface (DSI) for POSIX-compliant filesystems
- Non-POSIX DSI
  - Google Drive
  - Amazon S3 (native)
  - Spectra BlackPearl
  - Ceph S3 RadosGW
  - HPSS

Cloud-hosted DTN

- AWS EBS Volume
- AWS S3 Bucket
- Non-POSIX DSI
  - Google Drive
  - Amazon S3 (native)
  - Spectra BlackPearl
  - Ceph S3 RadosGW
  - HPSS

On-prem and cloud based endpoint hosting
Common endpoint configuration

Data Transfer Node

~/
~/scratch

ext*
XFS
ZFS

POSIX Connector
Common endpoint configuration

- Data Transfer Node
- POSIX Connector
- ~/projects
- ext* XFS ZFS
- ~/scratch
- GPFS Lustre
Multi-endpoint configuration

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

- File Systems:
  - ext*, XFS, ZFS
  - GPFS Lustre
  - ~/projects
  - ~/archive
  - ~/scratch

- ~/projects
- ~/archive
- ~/scratch
Multi-endpoint configuration

Data Transfer Node

- POSIX Connector
- Western Digital ActiveScale Connector
- Amazon S3 Connector

Endpoints:
- ~/scratch
- ~/projects
- ~/archive
- ~/vault

File Systems:
- ext*, XFS, ZFS
- GPFS Lustre
- Amazon S3 Bucket
Western Digital ActiveScale

- Turnkey on-premise object storage
- Globus connector using S3 API
- Low TCO: Manufactures own drives
- Erasure coding
- Auto data integrity checks with self-healing
- Cloud-based systems management tools
- Data Forever: automatic migration to new tech

docs.globus.org/premium-storage-connectors/wd-activescale/
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
Dual-homed DTN – high speed data path
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

- GridFTP Server
- MyProxy CA
- OAuth Server

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

Firewall

ext* XFS ZFS

Data Transfer Node

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

1. Install Globus Connect Server on all servers
2. 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
3. Copy Globus Connect Server configuration file to all servers
4. Run globus-connect-server-setup on the server running the –id component
5. 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
Join the Globus community

- Access the service: globus.org/login
- Create a personal endpoint: globus.org/app/endpoints/create-gcp
- Documentation: docs.globus.org
- Engage: globus.org/mailing-lists
- Subscribe: globus.org/subscriptions
- Need help? support@globus.org
- Follow us: @globusonline