Globus Endpoint Administration

Vas Vasiliadis
vas@uchicago.edu

NCAR – September 5, 2018
Get your server: bit.ly/ec2ip

1. Select an empty row in the spreadsheet
2. Enter your name and email address
3. Make a note of the IP address displayed

Slides and useful links: globusworld.org/tutorials
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, ...)

Local system users

Globus Connect Server

MyProxy CA

OAuth Server

GridFTP Server

DTN

globus
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 CLI to flag as managed endpoint (also configurable via the web app)**

* 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)
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>
High Assurance Endpoints

• App instance isolation
• Additional authentication assurance (IdP locking)
• Comprehensive audit logging
• Require Globus Connect Server v5.2+
  – New installation method (using client ID, secret)
  – New architecture/terminology

docs.globus.org/high-assurance/
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
Illustrative performance

Petascale DTN Project

November 2017
L380 Data Set

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

34.1/46.8/48.4 Gbps

41.0/42.2/43.9 Gbps

43.0/50.0/56.3 Gbps

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

35.9/39.0/40.7 Gbps

34.6/47.5/56.8 Gbps

23.1/33.7/39.7 Gbps

29.9/33.1/35.5 Gbps

33.2/43.4/50.3 Gbps

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

55.4/56.7/57.4 Gbps

21.2/22.6/24.5 Gbps

26.7/34.7/39.9 Gbps

Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.47 bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313894248 bytes (11G 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 - 1000G bytes: 3 files

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

24.5 Gbps
Disk-to-Disk Throughput: ESnet Testing

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

Disk-to-Disk Throughput (Mbps)

- 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 network configuration

![Diagram of network configuration]

- **Source**: Security filters
- **Destination**: Security filters
- **Destination**: Science DMZ
- **Source**: Science DMZ
- **Source**: Border Router
- **Destination**: Border Router
- **Source Router**: Data Transfer Node (DTN)
- **Destination Router**: Data Transfer Node (DTN)
- **User Organization**: Physical Data Path

**DATA**
- *Ports 50000-51000*

**CONTROL**
- *Ports 443, 2811, 7512*

*Please see TCP ports reference: https://docs.globus.org/resource-provider-guide/#open-tcp-ports_section*
Common endpoint configuration

- Data Transfer Node
- POSIX Connector
- ~/scratch
- ext*, XFS, ZFS
Common endpoint configuration

Data Transfer Node

POIX Connector

~/projects

ext* XFS ZFS

~/scratch

GPFS Lustre
Multi-endpoint configuration

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

- ~/projects
- ~/archive
- ~/scratch
- ext*, XFS, ZFS
- GPFS Lustre
Multi-endpoint configuration

Data Transfer Node

~/.vault

POSIX Connector

~/.projects

GPFS Lustre

Western Digital ActiveScale Connector

~/.archive

Amazon S3 Bucket

Amazon S3 Connector

~/.vault

ext*, XFS, ZFS

~/.scratch

~/.projects

~/.archive
Deploying a premium connector gateway

Object Store / Archive

Globus endpoint
(e.g. WD ActiveScale gateway)

Data

Other Globus endpoints

Control
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

- Science DMZ
- Data Transfer Node
- GridFTP Server
- Control Channel
- Data Channel
- Internet2 path
- GridFTP Server

Diagram showing a network setup with dual-homed DTN, high-speed data path, and integration with GridFTP and Science DMZ.
Dual-homed DTN – high speed data path

Science DMZ

Data Transfer Node

GridFTP Server

Data Channel

LAN/Intranet path

Control 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

- GridFTP Server
- MyProxy CA
- OAuth Server
- ext* XFS ZFS

Data Transfer Node

Firewall
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
Globus Network Manager
For environments with super duper special network constraints…
(a.k.a. "for the very brave")
Globus Network Manager

• Information from GridFTP to facilitate dynamic network changes
• Callbacks during GridFTP execution on local DTN
• Supplements information available via Globus transfer API
Globus Network Manager Callbacks

- Pre-listen (binding of socket)
- Post-listen
- Pre-accept/Pre-connect (no Data yet)
- Post-accept/Post-connect (data in flight)
- Pre-close
- Post-close
Network manager use cases

• **Science DMZ Traffic Engineering**
  – Use SDN to dynamically route data path
  – Control path uses traditional route

• **Automated WAN bandwidth reservation**
  – OSCARS, AL2S

• **Note: All this requires custom code**
Open Discussion