What’s New with Globus

Rachana Ananthakrishnan
rachana@globus.org

Greg Nawrocki
greg@globus.org
Research data management today

How do we...
...move?
...share?
...discover?
...reproduce?
Globus delivers...
Fast and reliable big data transfer, sharing, and platform services...
...directly from your own storage systems...
...via software-as-a-service using existing identities with the overarching goal of...
Unifying access to data across tiers

Research Computing HPC

National Resources

Personal Resources

Desktop Workstations

Mass Storage

Instruments

Public Cloud
Storage Connectors - globus.org/connectors

Current
- IBM Spectrum Scale
- SPECTRA
- ceph
- Amazon S3
- box
- Google Drive
- HDFS
- Lustre
- HPSS
- HGST ActiveScale

Planned
- Google Cloud
- wasabi
- Microsoft Azure
Share with collaborators/community

Project repositories, replication stores

Public repositories

External campus storage

Public/private cloud stores

XSEDE
Jetstream

EC2

Google Drive

Amazon Web Services

red cloud

S3

openstack

ceph
Manage data from instruments

- Next-Gen Sequencer
- Cryo-EM
- Advanced Light Source
- MRI
- Light Sheet Microscope

Analysis store
High-durability, low-cost store
Remote visualization
Personal system
Use(r)-appropriate interfaces

- Web
- CLI
- Rest API

GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
Globus SaaS / PaaS: Research data lifecycle

1. Researcher initiates transfer request; or requested automatically by script, science gateway.

2. Globus transfers files reliably, securely.

3. Researcher selects files to share, selects user or group, and sets access permissions.

4. Globus controls access to shared files on existing storage; no need to move files to cloud storage!

5. Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus.

6. The Globus Command Line Interface, API sets, and Python SDK provide a platform…

7. …for building science gateways, portals and publication services.

8. Automating research workflows and ensuring those that need access to the data have it.

- Use a Web browser or platform services
- Access any storage
- Use an existing identity
Conceptual architecture: Hybrid SaaS

 Subscriber Control Domain

 Subscriber owned and administered storage system

 Source Endpoint

 Globus "client" software

 Control Channel

 Destination Endpoint

 DATA Channel

 Single, globally accessible multi-tenant service

 No data relay or staging via Globus
Conceptual architecture: Sharing

- External User Control Domain
- Global Control Domain
- Managed Endpoint
- Subscriber Control Domain
- Shared Endpoint

- DATA Channel
- CONTROL Channel

- Subscriber managed filesystem permissions
- Globus managed "overlay" permissions

External User
Control Domain

GLOBAL CONNECT

Managed Endpoint

GLOBAL CONNECT

Global Control Domain
...makes your storage system a Globus endpoint
Globus Connect Personal

• Now your laptop / local workstation is a Globus endpoint
  – Installers do not require admin access
  – Zero configuration; auto updating
  – Handles NATs
  – Installs in seconds – easy to delete - I’ll prove it!
  – https://www.globus.org/globus-connect-personal
Almost demo time…

• **How do I get a Globus account?**
  – A Globus Account is
    o A Primary Identity
    o Possible Linked Identities
  – Your existing institutional identity may already work
  – Linking / Managing Identities
  – Consents

• **But I don’t have any Endpoints (Collections)!**
  – Globus Connect Personal
  – Globus Tutorial Endpoint 1
  – Globus Tutorial Endpoint 2
  – ESnet Test Endpoints
    o Contain file samples of various sizes
Demo time!

Identities and Accounts

The Activity Monitor

The Console

Transfer Details

Transfer

Sharing

Bookmarks

Responsive Interface

Groups

5.3 - Building a Collection on a Storage Gateway

The Hamburger Menu

5.3 - HTTPS

Roles
Globus sustainability model

- **Standard Subscription**
  - Shared endpoints
  - HTTPS support
  - Management console
  - Usage reporting
  - Priority support
- **Branded Web Site**
- **Premium Storage Connectors**
- **Alternate Identity Provider (InCommon is standard)**
The path to sustainability
Globus by the numbers...

- 1900+ most shared endpoints at a single institution
- 600+ PB moved
- 83 billion files processed
- 1800 active server endpoints
- 100+ subscribers
- 127,000 registered users
- 3 months longest running transfer
- 21,000 active personal endpoints
- 600+ identity providers
- 1 PB largest single transfer to date
- 8300 active shared endpoints
- 99.9% availability
Support resources

• Globus documentation: docs.globus.org
• Helpdesk and issue escalation: support@globus.org
• Mailing Lists
  – https://www.globus.org/mailing-lists
• Customer engagement team
• Globus professional services team
  – Assist with portal/gateway/app architecture and design
  – Develop custom applications that leverage the Globus platform
  – Advise on customized deployment and integration scenarios
Globus on your Campus

- Webinars
- Programs
  - Helping you evangelize Globus within your institution.
- Professional Services
- Globus World Tour
  - Taking the show on the road.