Globus for Researchers

Vas Vasiliadis - vas@uchicago.edu
May 23, 2023
ocellated
(Os-uh-lay-tid)
adjective: eyelike; having eyelike spots

From: A.Word.A.Day (wordsmith.org)
Globus is …
a non-profit service developed and operated by

THE UNIVERSITY OF CHICAGO
Our mission is to...

increase the efficiency and effectiveness of researchers engaged in data-driven science and scholarship through sustainable software.
Reimagining research IT with Globus

Managed transfer & sync

Collaborative data sharing

Unified data access

Publication & discovery

Managed remote execution

Reliable automation

Platform-as-a-Service

Software-as-a-Service
Fast, reliable file transfer ... from any to any system

- Fire-and-forget transfer, sync
- Optimized speed
- Assured reliability
- Unified view of storage
- HTTP/S access to data

Globally accessible multi-tenant service

User-initiated, or automated transfer request

Optional notifications

Instrument, Lab server

Globus transfers files reliably, securely

Compute Facility

Google Cloud
Mean = 52.1 Gbps; Max = 65.4 Gbps (n = 221)
Globus is reliable, secure ...and fast!

**Activity List**

- **RDA to ALCF noverify**
  - Transfer completed

**Overview**

- **Task Label**: RDA to ALCF noverify
- **Source**: NCAR RDA Dataset Archive
- **Destination**: DME PerfTest - Argonne
- **Task ID**: 20ebf766-a46d-11eb-8a95-d70d98a40c8d
- **Owner**: Vas Vasilakis (vas@globusid.org)
- **Condition**: SUCCEEDED
- **Requested**: 2021-04-23 02:50 pm
- **Completed**: 2021-04-23 02:53 pm
- **Duration**: 2 minutes 47 seconds
- **Transfer Settings**:
  - transfer is not encrypted
  - overwriting all files on destination

**Event Log**

- **6151** Files
- **2** Directories
- **1.51 TB** Bytes Transferred
- **9.10 GB/s** Effective Speed
- **0** Skipped files on sync
- **0** Skipped files on error

View debug data
Secure data sharing ...from any storage

- Fine-grained access control
- Storage system “overlay”
- Share with identity/email/group
- No data staging required

Globally accessible multi-tenant service

Select files to share, select user or group, and set access permissions

Collaborator logs into Globus and accesses shared files; no local account required; download via Globus

On-prem or public cloud storage

Laptop, server, compute facility

Globus controls access to shared files on existing storage
Globus supports diverse storage systems
Managed compute ...on any system

- Fire and forget function execution
- Federated authentication, and local access control
- Uniform interface to diverse compute resources
- Supports Python functions
- ...and everything else by wrapping as Python :-)

User submits a function to be run on compute endpoints

Globally accessible multi-tenant service

Globus manages the function execution on any endpoint

Results returned to the user

Laptop, server, compute facility
The **Globus Compute** model

- **Compute service** — Highly available cloud-hosted service for managed, fire-and-forget function execution
- **Compute endpoint** — Abstracts access to compute resources, from edge device to supercomputer
- **Compute SDK** — Python interface for interacting with the service, using the familiar (to many) Globus look and feel
- **Security** — Leverages Globus Auth; Compute endpoints are resource servers, authN and access via tokens
ML-based drug screening

National Virtual Biotechnology Laboratory, arXiv:2006.02431
Scalable data discovery … for any domain

- Scalable metadata store
- Fine grained visibility controls
- Schema agnostic
- Federated auth integration
- Query API with facets

User publishes metadata into search index

Globally accessible multi-tenant service

Globus manages metadata access

Users query and discover data of interest
Repository data distribution

Globally accessible multi-tenant service

Example: acdc.alcf.anl.gov
Reliable automation …spanning all resources

- Managed reliable task orchestration
- Declarative language for flow definition
- Event driven execution model
- Extensible for integrating external services
Automating cryoEM flows

Globus Flows

Transfer
- Transfer raw files

Compute
- Launch analysis job

Carbon!
- Correct, classify, ...

Share
- Set access controls

Transfer
- Move final files to repo

Search
- Ingest to index

Compute
- Extract metadata
One service, many interfaces

Globus service

CLI

Web

Flows

Globus service

GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json

Rest

API
Plugging into the Globus ecosystem

- **Globus Connect Server**
  - Data access on multi-user systems
  - [docs.globus.org/globus-connect-server](https://docs.globus.org/globus-connect-server)

- **Globus Connect Personal**
  - Data access on personal systems
  - [globus.org/globus-connect-personal](https://globus.org/globus-connect-personal)

- **Globus Compute Agent**
  - Compute access in single-user environments
  - [globus.org/compute](https://globus.org/compute)
Hand-on time!

Follow along on laptop or virtual machine

Watch demonstration (and do it yourself)
To participate...

Recent version of web browser
SSH client for later sessions
Let’s see how to...

Authenticate
Search collections and transfer files
View transfer activity and logs
Up/download files via browser
Install Globus Connect Personal on your laptop
Some cool stuff...

Sharing data
Linking identities
Creating bookmarks
Globus core security features

- **Access Control**
  - Identities provided and managed by institution
  - Institution controls all access policies
  - Globus is identity broker; no access to/storage of user credentials

- **Data remain at institutions, no storage/routing via Globus**

- **Integrity checks of transferred data**

- **Enforced encryption of Globus control data**

- **Institution-configured encryption of user data in transit**
Globus High Assurance features

- **Additional authentication assurance**
  - Reauthenticate after specified time period
  - Authenticate with a *specific identity* within session

- **Per application, per session auth context isolation**

- **Enforced encryption of all user data in transit**

- **Local audit logging**

- **Option to require MFA**

- **Blocked anonymous or open sharing**
Globus High Assurance for managing protected data

Security controls
→ NIST 800-53
→ NIST 800-171

Restricted data handling
→ PHI, PII, CUI
→ Compliant data sharing

BAA w/Uchicago
→ UChicago BAA with Amazon
Move without (worrying about) limits

- API request rates
- File size
- Data volume
- Third-party tools cannot circumvent...
- ...but Globus lets you “fire-and-forget”
- → it will (eventually) be done
More cool stuff...

Accessing protected data
Accessing cloud and archival storage
Managing access using Globus groups
On sustainability…
...Operations are funded by subscribers
Our sustainability model

• **File transfer for non-profit research is free to all**
  – Subscription required if transferring to/from a commercial entity

• **Subscriptions enable multiple enhanced features**
  – Researchers: Data sharing, transfer/sharing to/from personal endpoints, HTTPS access, Globus Flows for automation
  – Sysadmins: Management console, usage reporting
  – Developers: Globus Search, app integration support
  – Additional security/logging for protected data management
  – Priority support

• **Subscriptions are required for access to connectors**
Resources

- Documentation: docs.globus.org
- Community discussion: globus.org/mailing-lists
- Support: support@globus.org