Globus 2017: Beyond Big Data

Lee Liming
University of Chicago & Argonne National Laboratory
Researcher initiates transfer request; or requested automatically by script, science gateway.

Instrument

Globus transfers files reliably, securely.

Transfer

Compute Facility

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

Globus transfers files reliably, securely.

Share

Curator reviews and approves; data set published on campus or other system.

Publication Repository

• Only a Web browser required
• Use storage system of your choice
• Access using your campus credentials

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

Personal Computer

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

Researcher assembles data set; describes it using metadata (Dublin core and domain-specific).

Peer, collaborators search and discover datasets; transfer and share using Globus.
Globus by the numbers

- 48 most server endpoints on one campus
- 258 PB transferred
- 43 billion tasks processed
- 56,000 registered users
- 350 100TB+ users
- 10,000 active users
- 3 months longest running managed transfer
- 10,000 active endpoints
- 300+ identity providers
- 1 PB largest single transfer to date
- 5,119 active shared endpoints
- 99.5% uptime
Benefits of SaaS

• **Users do not need to deploy software**
  – Easy to access via Web browser
  – Command line, REST interfaces for flexible automation and integration

• **New features automatically available**

• **Reduced IT operational costs**
  – Small local footprint (Globus Connect)
  – Consolidated support and troubleshooting
Demo

- Login using existing identity
- Endpoint search
- Transfer
New features and enhancements
Storage connectors

• **Standard storage connectors (Posix)**
  – Linux, Windows, MacOS
  – Lustre, GPFS, OrangeFS, etc.

• **Premium storage connectors**
  
  - AWS S3
  - Ceph RadosGW (S3 API)
  - Spectra Logic BlackPearl
  - Google Drive
  - HPSS
  - HDFS (in progress)
  - iRODS (in progress)
  - HGST Active Archive (in progress)

[docs.globus.org/premium-storage-connectors](https://docs.globus.org/premium-storage-connectors)
Globus Auth

- ORCID

- OpenID Connect compliance

- Email aliases at InCommon domains
New Command Line Interface

- Transfer and Auth
- Replaces old SSH-based command line shell
- Uses Python SDK
- Open source

```
$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose                   Control level of output
  -h, --help                      Show this message and exit.
  -F, --format [json|text]       Output format for stdout. Defaults to text
  --jmespath, --jg TEXT          A JMESPath expression to apply to json output.
  --map-http-status TEXT         Map HTTP statuses to any of these exit codes:
                                0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
  delete        Submit a Delete Task
  endpoint      Manage Globus Endpoint definitions
  get-identities Lookup Globus Auth Identities
  list-commands List all CLI Commands
  login         Login to Globus to get credentials for the Globus CLI
  logout        Logout of the Globus CLI
  ls            List Endpoint directory contents
  mkdir         Make a directory on an Endpoint
  rename        Rename a file or directory on an Endpoint
  task          Manage asynchronous Tasks
  transfer      Submit a Transfer Task
  version       Show the version and exit
  whoami        Show the currently logged-in identity.
```

[github.com/globus/globus-cli] [docs.globus.org/cli]
Globus Python SDK

- Python SDK for Globus PaaS
- Transfer and Auth APIs
- Open source

github.com/globus/globus-sdk-python
Management Console

- **Updated user interface**
  - Shared endpoints
    - MC on shared endpoint
    - Host endpoint admin can see all shares & permissions
  - Search by local account

- **Self-help management console access via roles**
  - Activity manager
  - Activity monitor
HTTPS Access to Endpoints
(coming soon)

• **Enhanced use of research storage:**
  – asynchronous, bulk transfer via GridFTP
  – synchronous remote access via HTTPS

• **Leverages Science DMZ**

• **Enhanced Globus web app**
  – Browser-based upload/download
  – Inline file viewer

• **Integrate with your apps and services**
A busy year of training

- GlobusWorld Tour: 5 stops, over 200 people
- Lots of open source developer materials
- Contact us to host event

PennState
May 16, 2017

NEW YORK UNIVERSITY
May 31 – June 1, 2017
The Globus Platform: Automation

DMagic
ANL APS

dmagic.readthedocs.io

DMAGIC
a Globus implementation at the APS

 Courtesy of Francesco De Carlo, Argonne National Laboratory (2016)
The Globus Platform: Automation

Spot Suite Portal
LBNL ALS

Welcome to the SPOT Suite Portal for Light-Source Data and Simulations.

Log in with your NERSC account at the top right to gain access to data and NERSC compute resources.

This site is a work in progress. Please report any problems with the page to spot-help@lists.lbl.gov

spot.nersc.gov
The Globus Platform: Native App Support

github.com/globus/native-app-examples
19-day sample of Globus platform use (March/April 2017)

4665 Non-Globus IdP Logins
3829 Unique Identities
152 IdPs

2062 Non-Globus Webapp Logins

1. Jetstream
2. Blue Waters
3. Compute Canada
4. GPCR
5. CIRM
6. XSEDE Moodle
7. KBase

Top 10 IdPs

- XSEDE
- Google
- University of Michigan
- University of Chicago
- NIH
- Purdue University
- Johns Hopkins University
- NCSA
- Illinois
- UCLA
App registration

Developers.globus.org

Register your app with Globus

Developer documentation

developers.globus.org
Mailing lists

Globus Email Lists

If you use Globus, then participating in one or more of the public email lists is an excellent way to keep in touch with your peers in the Globus Community. There are three lists you can join — all are public forums:

User List
user-discuss@globus.org
For general questions about using the Globus service.

Admin List
admin-discuss@globus.org
For questions about managing your Globus deployment, e.g., installing software for a Globus endpoint, configuring your firewall, and integrating your institution's identity system.

Developer List
developer-discuss@globus.org
For questions about integrating Globus services into your science gateway, portal, or other web application via the Globus REST APIs.

globus.org/mailing-lists
Thank you to our sponsors

U.S. Department of Energy

National Institute of Standards and Technology
U.S. Department of Commerce

NIH

The University of Chicago

Argonne National Laboratory

powered by Amazon Web Services
Endpoint Roles

Manage Endpoints

Globus Project Share on BlueWaters Lustre

Host Endpoint: ncsfBlueWaters
Host Path: /projects/vendor/globus/share/

<table>
<thead>
<tr>
<th>user or group</th>
<th>role</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:SteveTuecke@globusid.org">SteveTuecke@globusid.org</a></td>
<td>administrator</td>
</tr>
</tbody>
</table>

Assign New Role to Globus Project Share on BlueWaters Lustre

Assign To: user
Identity/E-mail: ves@uchicago.edu
Role: Administrator
- modify endpoint definition, delete the endpoint, manage roles, perform file system operations and transfers, and all capabilities of the Access Manager role
- Access Manager
  - view, add, and delete all access rules on the endpoint; implicitly gives read/write access to the root of the endpoint
- Activity Manager
  - view and control tasks and other endpoint activity
- Activity Monitor
  - view tasks and other activity to or from the endpoint

Add
Spectra BlackPearl + Globus Solution

On-line/High performance storage system

Data Transfer Node

Black Pearl Gateway

Tape

Arctic Blue

docs.globus.org/premium-storage-connectors/black-pearl/
AWS S3, Ceph RadosGW

- Now generally available
- Admin sets up credentials
- User self-registration coming

[docs.globus.org/premium-storage-connectors/aws-s3/](docs.globus.org/premium-storage-connectors/aws-s3/)
Globus Auth: Identity and Access Management PaaS

- Federated identity login to apps
  - Web, mobile, desktop, command line

- Protect all REST API communications
Data Publication and Discovery
Data Publication

Globus Data Publication (Trial)

Welcome to the Globus data publication "trial sandbox". The User Guide explains how to publish data. Please contact us with questions or feedback. When you are ready to create a publication community for your institution, complete this form.

Communities
Choose a community to browse its collections.
- Compute Canada
- Globus Demo Community
- Kansas City Federal Reserve
- Open Trial
- Open to all Globus users.
- Pilot Communities
- Groups performing trials

Discover

Subject

Submit: Describe Dataset

All fields with an asterisk (*) are required.

Title *
- Jane's First Dataset to Create Trial Collection

Authors *
- Smith

Publication Year *
- 2015

Language
- English (United States)

Publisher *
- Trial Sandbox Holdings

© 2010-2015 Computation Institute, University of Chicago, Argonne National Laboratory
Other future enhancements

- **Symbolic link handling**
  - Keep, copy, or ignore symlinks
- **Data search**
- **Compliance**
  - HIPAA, FedRAMP gap analysis underway
New pricing model

Globus Subscriptions

Globus file transfer is free to use for non-profit research and educational purposes. Subscriptions offer researchers access to enhanced features such as file sharing and data publication, and provide system administrators with a comprehensive view into how storage systems are being utilized. Your subscription helps us make Globus self-sustaining so we can continue to offer high-quality data management infrastructure to all researchers.

Click here to start your 90-day free trial of Globus premium features

Subscriptions for Non-Profit Research and Education

If you wish to use Globus in a commercial setting, you must have a commercial subscription.

<table>
<thead>
<tr>
<th>Features (click for description)</th>
<th>Basic - Free</th>
<th>Starter Subscription</th>
<th>Standard Subscription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer level</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>User level</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Managed endpoints</td>
<td>None</td>
<td>1</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Management console</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Usage reports</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Support for Globus Connect, Web, CLI</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shared endpoints</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Globus Plus users</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
</tbody>
</table>

https://www.globus.org/subscriptions