



globus online



Big data transfer and sharing ...as a service

Steve Tuecke

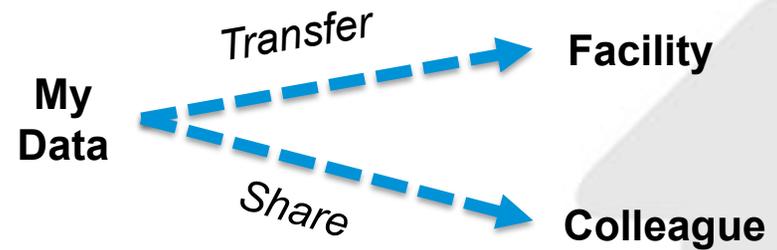
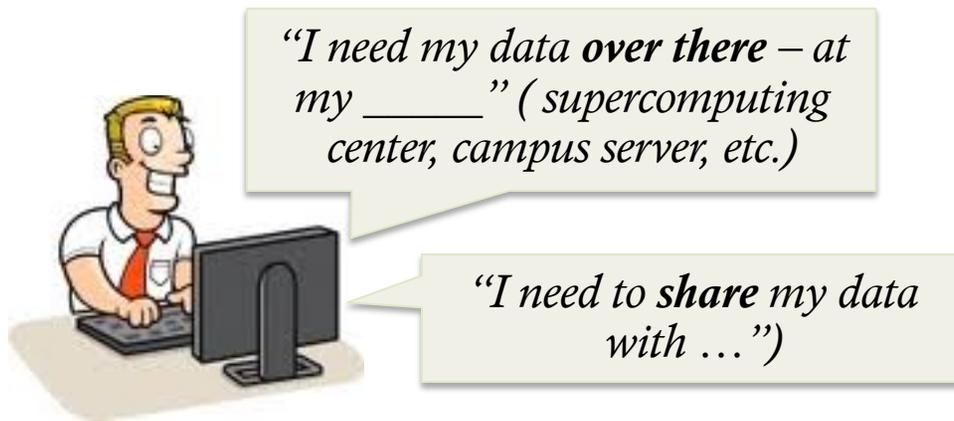
Computation Institute

University of Chicago and Argonne National Laboratory



The Challenge: Moving and Sharing Big Data *Easily*

- What should be trivial...



- ... can be tedious, time-consuming and expensive





Bandwidth Requirements

Bandwidth Requirements to move Y Bytes of data in Time X

Bits per Second Requirements

10PB	25,020.0 Gbps	3,127.5 Gbps	1,042.5 Gbps	148.9 Gbps	34.7 Gbps
1PB	2,502.0 Gbps	312.7 Gbps	104.2 Gbps	14.9 Gbps	3.5 Gbps
100TB	244.3 Gbps	30.5 Gbps	10.2 Gbps	1.5 Gbps	339.4 Mbps
10TB	24.4 Gbps	3.1 Gbps	1.0 Gbps	145.4 Mbps	33.9 Mbps
1TB	2.4 Gbps	305.4 Mbps	101.8 Mbps	14.5 Mbps	3.4 Mbps
100GB	238.6 Mbps	29.8 Mbps	9.9 Mbps	1.4 Mbps	331.4 Kbps
10GB	23.9 Mbps	3.0 Mbps	994.2 Kbps	142.0 Kbps	33.1 Kbps
1GB	2.4 Mbps	298.3 Kbps	99.4 Kbps	14.2 Kbps	3.3 Kbps
100MB	233.0 Kbps	29.1 Kbps	9.7 Kbps	1.4 Kbps	0.3 Kbps
	1H	8H	24H	7Days	30Days

<http://fasterdata.es.net>

Use a Science DMZ to optimize

<http://fasterdata.es.net/science-dmz/>



www.globusonline.org



What is Globus Online?

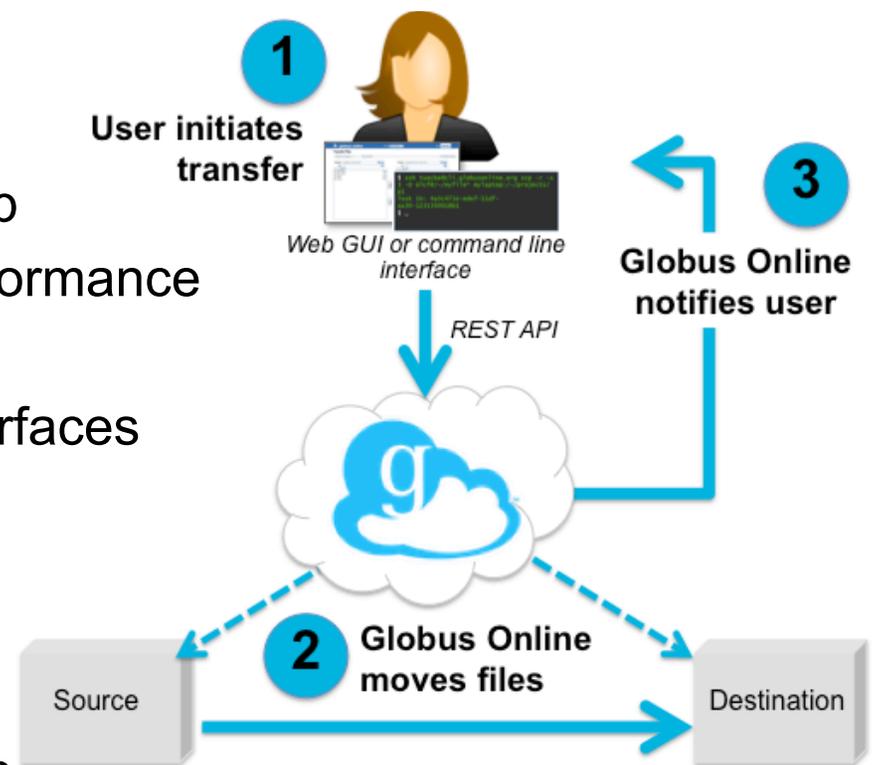
- **Move, sync, share files**

- Easy “fire-and-forget” transfers
- Share with any Globus user or group
- Automatic fault recovery & high performance
- Across multiple security domains
- Web, command line, and REST interfaces

- **Minimize IT costs**

- Software as a Service (SaaS)
 - No client software installation
 - New features automatically available
- Consolidated support & troubleshooting
- Simple endpoint installation with Globus Connect and GridFTP

- **Recommended by ALCF, ESnet, NERSC, XSEDE, Blue Waters, NCAR, many Universities**





Sample Endpoints

Argonne

- alcf#dtn
- alcf#dtn_intrepid
- alcf#dtn_mira
- alcf#dtn_surveyor
- mcs#home
- lcrc#fusion
- aps#clutch

Northwestern

- northwestern#quest

LBL

- nersc#dtn
- nersc#carver
- nersc#dtn_jgi
- nersc#hopper

UTexas, Austin

- tacc#spur
- xsede#stampede
- xsede#lonestar
- xsede#ranch

Fermi

- usqcd#fermi

Rice

- rice#dtn

Harvard

- sbgrid#hmseq

UChicago

- ucrc#midway
- ci#beagle
- uc3#uc3-data





globus online



Globus Online Demo



globus online



Enabling your server / cluster as a Globus Online Endpoint



What is a Science DMZ?



Consists of 3 key components, all required:

- **“friction free” network path**
 - Highly capable network devices (wire-speed, deep queues)
 - Virtual circuit connectivity option
 - Security policy and enforcement specific to science workflows
 - Located at or near site perimeter if possible
- **Dedicated, high-performance data movers**
 - a.k.a.: Data Transfer Node (DTN)
 - Optimized bulk data transfer tools such as GlobusOnline/GridFTP
- **Performance measurement/test node**
 - perfSONAR

Details at: <http://fasterdata.es.net/science-dmz/>



Globus Online in three steps...

To turn your resource into a GO endpoint, you will need to:

1) Provide your users a way of authenticating with your resource that can be used by GO.



2) Set up a GridFTP server that will be in charge of actually serving the files.



3) Create your GO endpoint.



Globus Online in three steps...

To turn your resource into a GO endpoint, you will need to:

- 1) Provide your users a way of authenticating with your resource that can be used by GO.

Install MyProxy CA, or use external CA



- 2) Set up a GridFTP server that will be in charge of actually serving the files.

Install Globus Toolkit GridFTP server



- 3) Create your GO endpoint.

Use Globus Online Web or CLI



... or just one step

To turn your resource into a GO endpoint,
you will need to:

- 1) Install **Globus Connect Multi-User**





Adding your server to Globus Online Globus Connect Multi-User

- **What is GCMU?**
 - Globus Connect version for easily creating (sharable) endpoints on multi-user storage servers
 - Packages a GridFTP server and MyProxy CA authentication server, pre-configured for use with Globus Online
- **Why GCMU?**
 - Create transfer endpoints in minutes
 - Avoid complex GridFTP install
- **To download:** www.globusonline.org/gcmu



"We used GCMU to form a campus-wide GSI authentication service spanning multiple servers. Now my users have a fast, easy way to get their data wherever it needs to go, and the setup process was trivial."

--University of Michigan



"As a resource admin, I've found GCMU an exceedingly useful tool.... With GCMU, setting up a GridFTP server and handling authentication for multiple users is easy."

--Oak Ridge National Lab

www.globusonline.org



Globus Connect Multi User Steps

Download from <http://www.globusonline.org/gcmu>

```
tar xzf globusconnect-multiuser-latest.tgz
```

```
cd gcmu-1.1.3/
```

```
./install
```

Follow the prompts...

Coming soon: Native RPM and Debian packages, sharing support, OAuth, no SSH key required, update-able, ...



Globus Online Platform-as-a-Service

- **Create custom (web) applications that leverage Globus Online services**
 - User identities, profiles, groups (Nexus)
 - File transfer, synchronization, sharing
- **Integrate via**
 - REST APIs
 - Scriptable command line interface
 - Branded web pages





End-User Pricing

- **Why charge?**
 - To sustain Globus Online as a *non-profit service* to the non-profit research community
- **Basic: Free**
 - File transfer and synchronization to/from servers
 - Server endpoints with Globus Connect Multi-User
 - Can host shared endpoints for Plus subscribers
 - Personal endpoints with Globus Connect
 - Access to shared endpoints created by others
- **Plus: \$7/month (or \$70/year)**
 - Create and manage shared endpoints (from any sharable or personal endpoint)
 - Peer-to-peer (Globus Connect to Globus Connect)
 - Support for web and command line interfaces



Provider Pricing

- **Basic: Free**
 - Server endpoints with Globus Connect Multi-User
 - Can host shared endpoints for Plus subscribers
- **Provider Starter: \$20k/year**
 - Branded web site with choice of default identity provider
 - Alternate identity provider
 - 1 provider endpoint: local users can create and manage shared endpoints hosted by this server endpoint
 - Support for branded, web, command line, and REST interfaces
- **Higher tiers with additional capabilities**
 - More branded sites, more provider endpoints, more local users, more GridFTP servers per endpoint, MSS optimizations, usage reports, operations console, advisory board seat, roadmap visibility, code escrow, ...



For More Information

- Visit www.globusonline.org/signup to:
 - Get a free account to start moving and sharing files
- Visit www.globusonline.org for:
 - Tutorials, FAQs, Pro Tips, Troubleshooting
 - Papers, Case Studies
- Visit support.globusonline.org or contact support@globusonline.org for:
 - Help
 - Forums
- Follow us at [@globusonline](https://twitter.com/globusonline) on Twitter and [Globus Online](https://www.facebook.com/globusonline) on Facebook