Simplify Big Data
Sharing and Transfer
www.globusonline.org

Raj Kettimuthu
kettimut@mcs.anl.gov
Argonne National Laboratory
and University of Chicago
We asked ourselves: What if the research work flow could be managed as easily as…

...our pictures

...our e-mail

...home entertainment
What makes these services great?

Great User Experience

+ Scalable (but invisible) infrastructure
We aspire (initially) to create a great user experience for research data management. What would a “dropbox for science” look like?
Managing data should be easy …
… but it’s hard and frustrating!
Network Requirements

Bandwidth Requirements to move Y Bytes of data in Time X

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

Use a Science DMZ to optimize
http://fasterdata.es.net/science-dmz/

http://fasterdata.es.net
What is Globus Online?

Big data transfer and sharing…

…with Dropbox-like simplicity…

…directly from your own storage systems
Reliable, secure, high-performance file transfer and synchronization

- “Fire-and-forget” transfers
- Automatic fault recovery
- Seamless security integration

1. User initiates transfer request
2. Globus Online moves and syncs files
3. Globus Online notifies user
Simple, secure sharing off existing storage systems

- Easily share large data with any user or group
- No cloud storage required

1. User A selects file(s) to share, selects user or group, and sets permissions

2. Globus Online tracks shared files; no need to move files to cloud storage!

3. User B logs in to Globus Online and accesses shared file
Globus Online is SaaS

- Web, command line, and REST interfaces
- Reduced IT operational costs
- New features automatically available
- Consolidated support & troubleshooting
- Easy to add your laptop, server, cluster, supercomputer, etc. with Globus Connect
Demonstration
Globus Connect Multiuser

- Create endpoint in minutes; no complex GridFTP install
- Enable all users with local accounts to transfer files
- Native packages: RPMs and DEBs
- Also available as part of the Globus Toolkit
Globus Connect Multiuser for providers

Deliver advanced data management services to researchers

Provide an integrated user experience

Reduce your support burden
Globus Connect Multiuser

• Create endpoint in minutes; no complex GridFTP install
• Enable all users with local accounts to transfer files
• Native packages: RPMs and DEBs
• Also available as part of the Globus Toolkit
GCMU Demonstration
Let’s get started…

On RPM distributions:

$ yum install globus-connect-multiuser

On Debian distributions:

$ curl -LOs http://www.globus.org/ftppub/gt5/5.2/stable/installers/repo/globus-repository-5.2-stable-precise_0.0.3_all.deb
$ sudo dpkg -I globus-repository-5.2-stable-precise_0.0.3_all.deb
$ sudo aptitude update
$ sudo aptitude -y install globus-connect-multiuser
GCMU Basic Configuration

• Edit the GCMU configuration file:
  /etc/globus-connect-multiuser.conf
  – Set [Endpoint] Host = "myendpointname"

• Run: globus-connect-multiuser-setup
  – Enter your Globus Online username and password when prompted

• That’s it! You now have an endpoint ready for file transfers

• More configurations at: support.globusonline.org/forums/22095911
Globus Online–GCMU interaction

Step 1: Access Endpoint

Step 2: username & password

Step 3: TLS handshake

Step 4: username & password

Step 5: certificate

Step 6: Transfer request

Step 7: certificate

Step 8: Authorization

Step 9: Authentication & Data Transfer

Campus Cluster

Local Authentication System (LDAP, RADIUS, Kerberos, …)

Globus Connect Multiuser

MyProxy Online CA

PAM

GridFTP Server

Globus Online (hosted service)

Remote Cluster / User’s PC

Local Storage

GridFTP Server

Access files
Firewall configuration

- **Allow inbound connections to port 2811** (GridFTP control channel), 7512 (MyProxy CA), 443 (OAuth)
- **Allow inbound connections to ports 50000-51000** (GridFTP data channel)
  - If transfers to/from this machine will happen only from/to a known set of endpoints (not common), you can restrict connections to this port range only from those machines
- **If your firewall restricts outbound connections**
  - Allow outbound connections if the source port is in the range 50000-51000
MyProxy OAuth server

- **Web-based endpoint activation**
  - Sites run a MyProxy OAuth server
    - MyProxy OAuth server in Globus Connect Multiuser
  - Users enter username/password only on site’s webpage to activate an endpoint
  - Globus Online gets short-term X.509 credential via OAuth protocol

- **MyProxy without OAuth**
  - Site passwords flow through Globus Online to site MyProxy server
  - Globus Online does not store passwords
  - Still a security concern for some sites
Globus Connect Multiuser

1. Access Endpoint
2. Username and password
3. Redirect
4. Username and password
5. Certificate
6. Certificate
7. Certificate
8. Transfer request
9. Certificate
10. Certificate
11. Certificate

Steps:
- Step 1: Access Endpoint
- Step 2: Username and password
- Step 3: Redirect
- Step 4: Username and password
- Step 5: Certificate
- Step 6: Certificate
- Step 7: Certificate
- Step 8: Transfer request
- Step 9: Certificate
- Step 10: Certificate
- Step 11: Certificate

Services:
- Globus Online (hosted service)
- OAuth Server
- Local Authentication System (LDAP, RADIUS, Kerberos, …)
- MyProxy Online CA
- GridFTP Server
- PAM
- GridFTP Server
- Local Storage

Processes:
- Access files
- Authentication & Data Transfer

Clusters:
- Campus Cluster
- Remote Cluster / User’s PC
Enable your resource. It’s easy.

• Signup: globusonline.org/signup
• Connect your system: globusonline.org/gcmu
• Learn: support.globusonline.org/forums/22095911
• Need help? support.globusonline.org
• Follow us: @globusonline
GCMU Advanced Configuration

- Customizing filesystem access
- Using host certificates
- Using CILogon certificates
- Configuring multiple GridFTP servers
- Setting up an anonymous endpoint
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~/*`
Sharing Path Restriction

- Define additional restrictions on which paths your users are allowed to create shared endpoint
- Use `SharingRestrictPaths` to customize
  - Same syntax as `RestrictPaths`
- E.g. Full access to home directory, deny hidden files:
  - `RestrictPaths = RW~,N~/.*`
- E.g. Full access to public folder under home directory:
  - `RestrictPaths = RW~/public`
- E.g. Full access to `/project`, read access to `/scratch`:
  - `RestrictPaths = RW/project,R/scratch`
Per user sharing control

- **Default: SharingFile = False**
  - Sharing is enabled for all users when Sharing = True

- **SharingFile = True**
  - Sharing is enabled only for users who have the file 
    ~/.globus_sharing

- **SharingFile can be set to an existing path in order for sharing to be enabled**
  - e.g. enable sharing for any user for which a file exists in /var/globusonline/sharing/:
  - SharingFile = "/var/globusonline/sharing/$USER"
Use CILogon issued certificates for authentication

• Your organization must allow CILogon to release ePPN attribute in the certificate

• Set AuthorizationMethod = CILogon in the globus connect multiuser configuration

• Set CILogonIdentityProvider = <your_institution_as_listed_in_CILogon_identity_provider_list>

• Add CILogon CA to your trustroots (/var/lib/globus-connect-multiuser/grid-security/certificates/)
Setting up additional GridFTP servers for your endpoint

- curl -LOs http://www.globus.org/ftppub/gt5/5.2/stable/installers/repo/globus-repository-5.2-stable-oneiric_0.0.3_all.deb
- sudo dpkg -i globus-repository-5.2-stable-oneiric_0.0.3_all.deb
- sudo aptitude update
- sudo aptitude -y install globus-connect-multiuser
- sudo vi /etc/globus-connect-multiuser.conf
- <-- comment ‘Server = %(HOSTNAME)s’ in ‘MyProxy Config’
- Copy contents of ‘/var/lib/globus-connect-multiuser/grid-security/certificates/’ from the first machine to same location on this machine
- sudo globus-connect-multiuser-setup <-- enter Globus Online username and password
Setting up additional GridFTP servers for your endpoint

$ curl -LOs http://www.globus.org/ftppub/gt5/5.2/stable/installers/repo/globus-repository-5.2-stable-precise_0.0.3_all.deb
$ dpkg -I globus-repository-5.2-stable-precise_0.0.3_all.deb
$ sudo aptitude update
$ sudo aptitude -y install globus-connect-multiuser
$ sudo vi /etc/globus-connect-multiuser.conf

- Comment Server = %(HOSTNAME)s in [MyProxy] section
- Copy contents of ‘/var/lib/globus-connect-multiuser/grid-security/certificates/’ from the first machine to same location on this machine

$ sudo globus-connect-multiuser-setup
- Enter Globus Online username and password when prompted
Setting up an anonymous endpoint

$ globus-gridftp-server -aa -anonymous-user <user>

• --anonymous-user <user> needed if run as root

$ endpoint-add <name> -p ftp://<host>:<port>
$ endpoint-modify --myproxy-server=myproxy.globusonline.org
We are a non-profit service provider to the non-profit research community.
We are a non-profit service provider to the non-profit research community

Our challenge: 
Sustainability
Globus Online End User Plans

• **Basic: Free**
  – File transfer and synchronization to/from servers
  – Create private and public endpoints
  – Access to shared endpoints created by others

• **Plus: $7/month (or $70/year)**
  – Create and manage shared endpoints
  – Peer-to-peer transfer and sharing
Globus Online Provider Plans

Bundle of Plus subscriptions
+ Features for providers

http://globusonline.org/provider-plans
(Working on NET+ offering)

- **Signup**: globusonline.org/signup
- **Connect your system**: globusonline.org/globus_connect
- **Learn**: support.globusonline.org/forums
- **Need help?** support.globusonline.org
- **Follow us**: @globusonline
Our research is supported by:
Questions

Contact: support@globusonline.org

Providers: globusonline.org/provider-plans

www.globusonline.org