Simplify Big Data
Sharing and Transfer
www.globusonline.org

Steve Tuecke
tuecke@ci.uchicago.edu
Transferring and sharing data should be easy...
... but it’s hard and frustrating!

- Expired credentials
- NATs Firewalls
- Permission denied
- Failed Server / Net
## Bandwidth Requirements

### Bandwidth Requirements to move Y Bytes of data in Time X

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

![Diagram](image)

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
Some Life Sciences Uses

• Transfer/sync data with remote analysis resources
• Share data with collaborators
• Get data from sequencer, imager, etc.
• Get data from community collections
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
• Workflows can be easily defined and automated with integrated Galaxy Platform capabilities

• Data movement is streamlined with integrated Globus Online file-transfer functionality

• Resources can be provisioned on-demand with Amazon Web Services cloud based infrastructure

• Authenticate using your Globus Online account
Future Plans

• Dataset and metadata management
• HIPAA
We are a non-profit service provider to the non-profit research community
Our challenge:
Sustainability

We are a non-profit service provider to the non-profit research community.
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)
Our research is supported by:

U.S. DEPARTMENT OF ENERGY

NSF

THE UNIVERSITY OF CHICAGO

Argonne

powered by Amazon Web Services

• Signup: globusonline.org/signup
• Connect your system: globusonline.org/globus_connect
• Learn: support.globusonline.org/forums
• Need help? support.globusonline.org
• Follow us: @globusonline

www.globusonline.org