g globus online

Integrating Galaxy and Globus Online

Ravi K Madduri

The Globus Project

May 24, 2011

globus online

Reliable File Transfer. No IT Required.

Reliable, high-performance, secure file transfer

Move files fast. No IT required.

+ HOW IT WORKS

Learn more about the service

> GET STARTED

Setup profile in 3 easy steps

Globus Online makes robust file transfer capabilities, traditionally available only on expensive, special-purpose software systems, accessible to everyone.

Learn more

Why Use Globus Online?
See how easy file transfer can

For HPC Resource Owners
Enable Globus Online

For Developers
Integrate with Globus Online

Globus products

Globus Toolkit
Use the Grid

Reliable file transfer
Software-as-a-Service
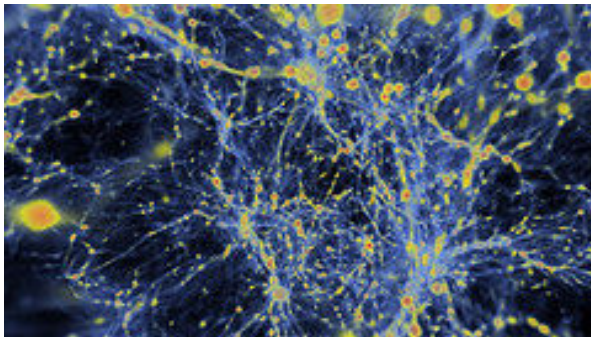
globusonline.org

Globus Online
Build the Grid

Components for building
custom grid solutions
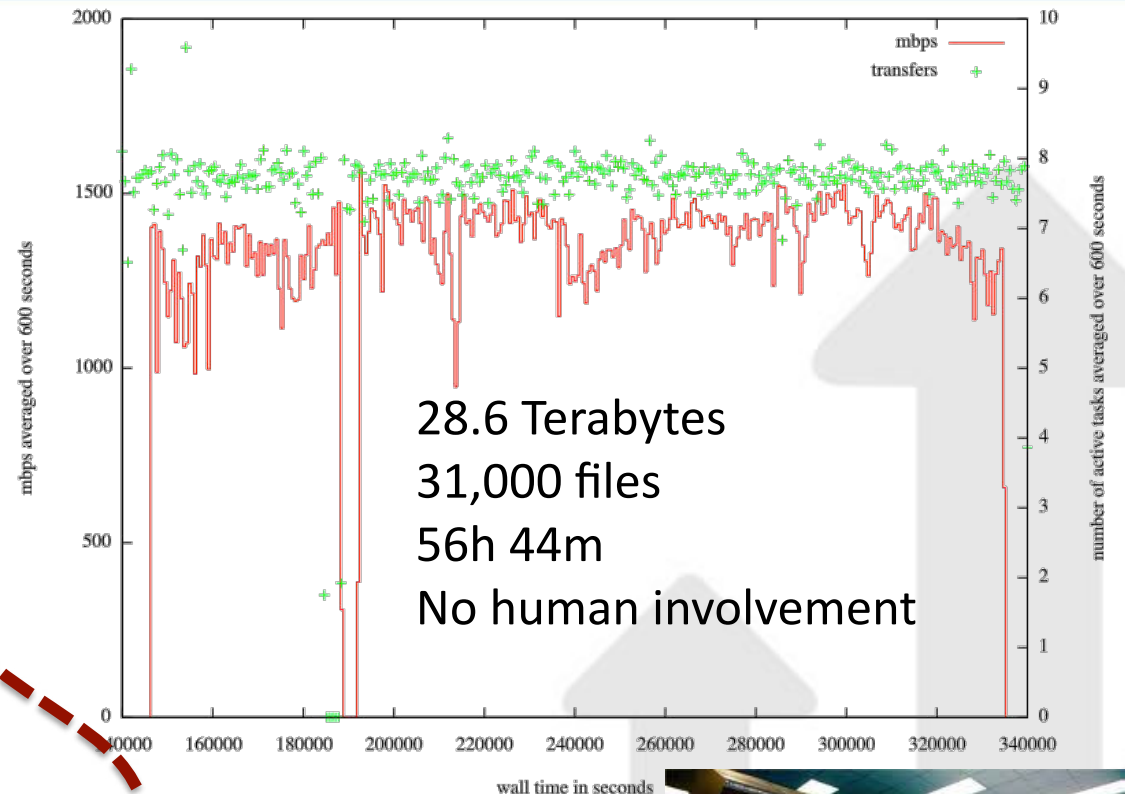
globustoolkit.org

www.globusonline.org

Globus Online In Action

Astrophysics simulation data
generated in Tennessee,
moved to Illinois for visualization
(Enzo, UCSD; Futures Lab, Argonne)

28.6 Terabytes
31,000 files
56h 44m
No human involvement

www.globusonline.org

Benefits of Globus Online

- **Reliable file transfer.**
  - Easy "fire and forget" file transfers
  - Automatic fault recovery
  - High performance
  - Across multiple security domains

- **No IT required.**
  - No client software installation
  - New features automatically available
  - Consolidated support and troubleshooting
  - Works with existing GridFTP servers
  - Globus Connect solves "last mile problem"

www.globusonline.org
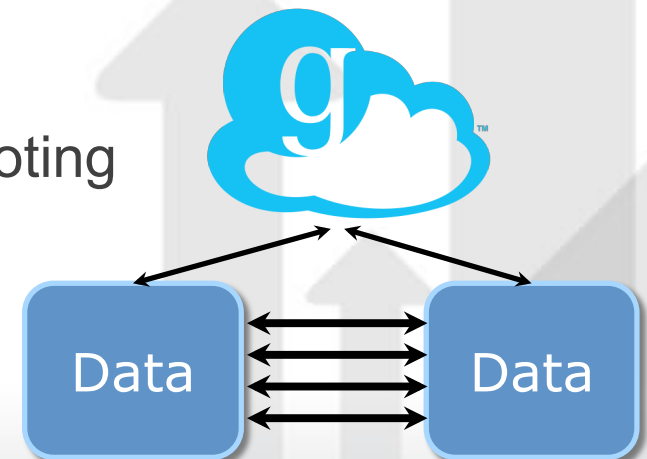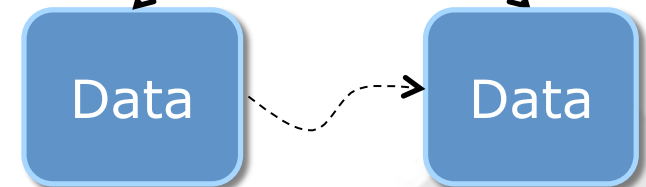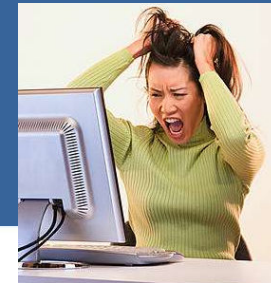
Who can benefit from Globus Online

- **Ad-hoc:** Non-programmers who need to move many files can use Web GUI

- **Scripted:** Users who want to create automated workflows can use Command Line Interface (CLI)

- **System builders:** Programmer who don't want to re-engineer file transfer solutions can use REST API

6

GO Galaxy Capabilities

- **Use GO transfer service to transfer data in and out of Galaxy reliably using high performance data transfer protocol, GridFTP**

- **Provision a EC2 cluster with Galaxy, Condor, GridFTP server and a set of users in <15mins**

- **Upload data sets of size more than 2GB reliably using Globus Connect**

- **Solve the last mile data transfer problem by making it easy to download data sets after analysis in Galaxy using Globus Connect**

Integration of Globus Online with Galaxy

www.globusonline.org

Globus Online Tools in Galaxy

www.globusonline.org

Sample workflow (Serial and Parallel Blast with GridFTP transfers using GO)

www.globusonline.org

Workflow with all the steps

www.globusonline.org

Workflow with all the steps

Workflow in Progress

Final Results

Final Results on a EC2 endpoint

Email Notifications of Transfers

www.globusonline.org

Transfer Monitoring

www.globusonline.org

Sequence Vs Parallel Blast Using Condor

- **Using m1.small instances:**
  - Sequential: 15m19s (BLAST only: 13m57s)
  - Parallel: 06m22s (BLAST only: 03m45s)

- **Using c1.medium:**
  - Sequential: 6m06s (BLAST only: 4m59s)
  - Parallel: 3m20s (BLAST only: 1m50s)

GO Galaxy Future Capabilities

- **Integrate flexible Globus Online Identity management, group management with Galaxy (BIRN project wrt InCommon, campus identity etc)**

- **Ability to easily share data with collaborators**

- **Ability to create a VO with the above capabilities (User/Groups management, File system, Transfer, HTC, Parallel execution, EC2 Cluster, data sharing) with click of a button.**

globus online

Demo