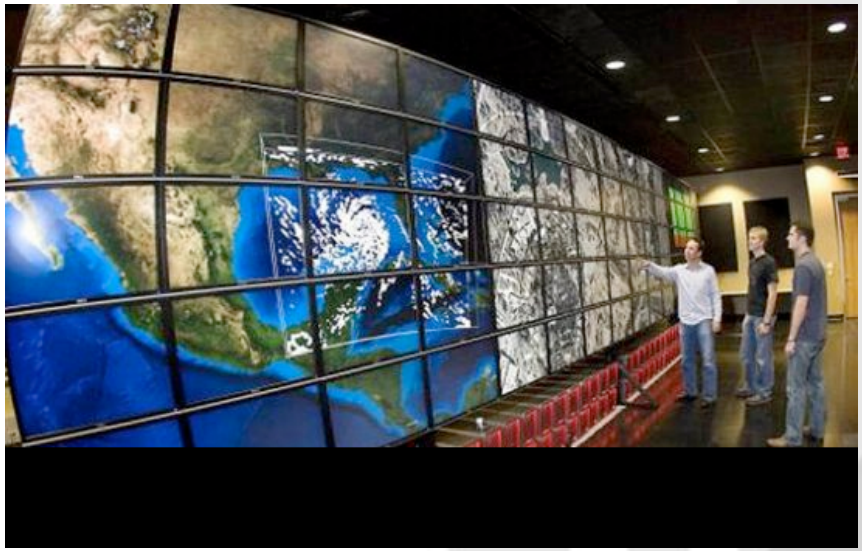
Globus Online: Simplifying Data Movement in HPC Environments

Raj Kettimuthu
Argonne National Laboratory and
The University of Chicago

Data Movement

NCSA



TACC

Data Movement

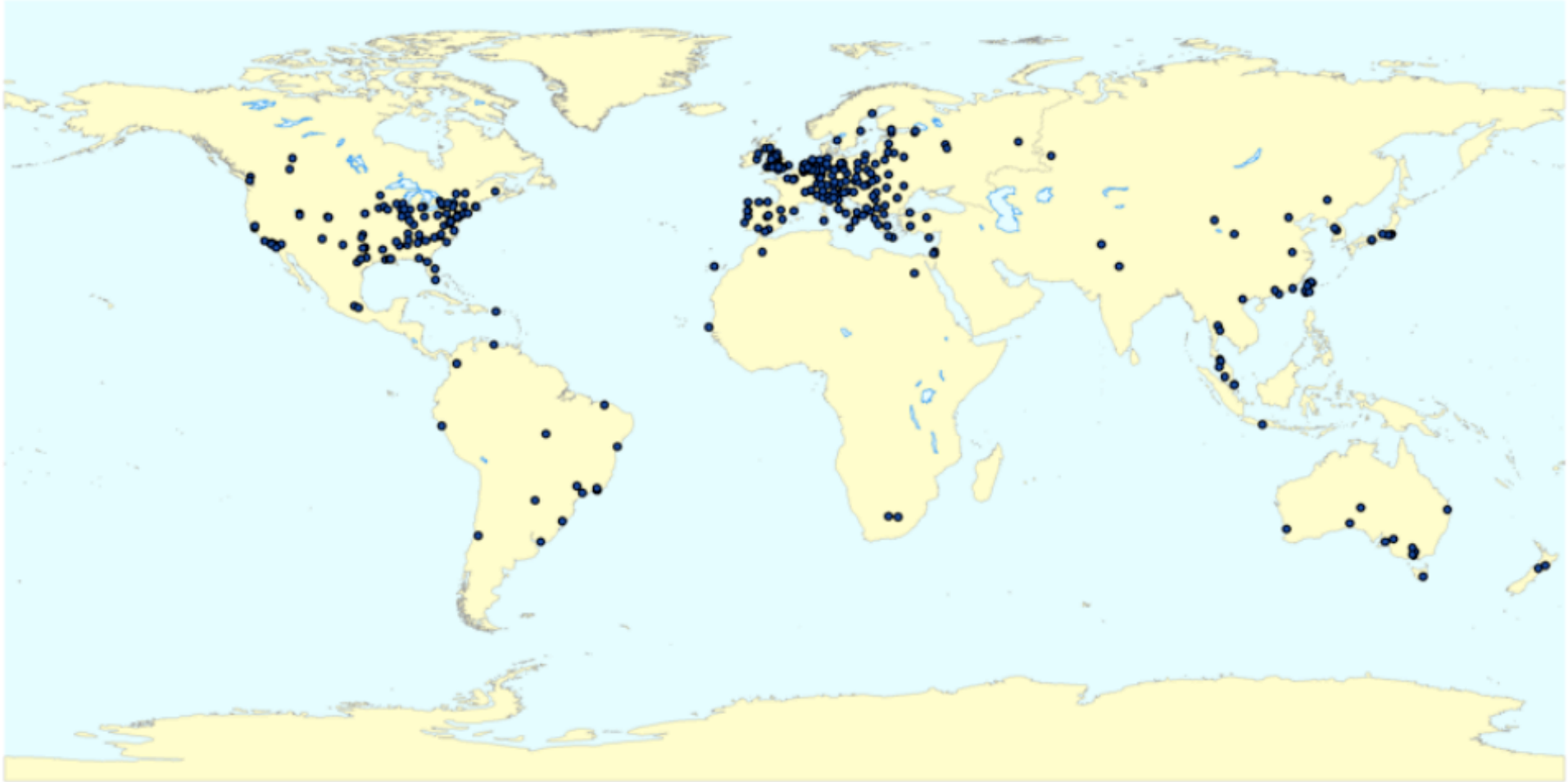
- **Deceptively simple**
- **Too large**
- **100,000 files totaling 10 Terabytes**
- **Move from a federal laboratory where they were generated to my home institution**
- **Can be very difficult**

Data Movement Challenges

GridFTP

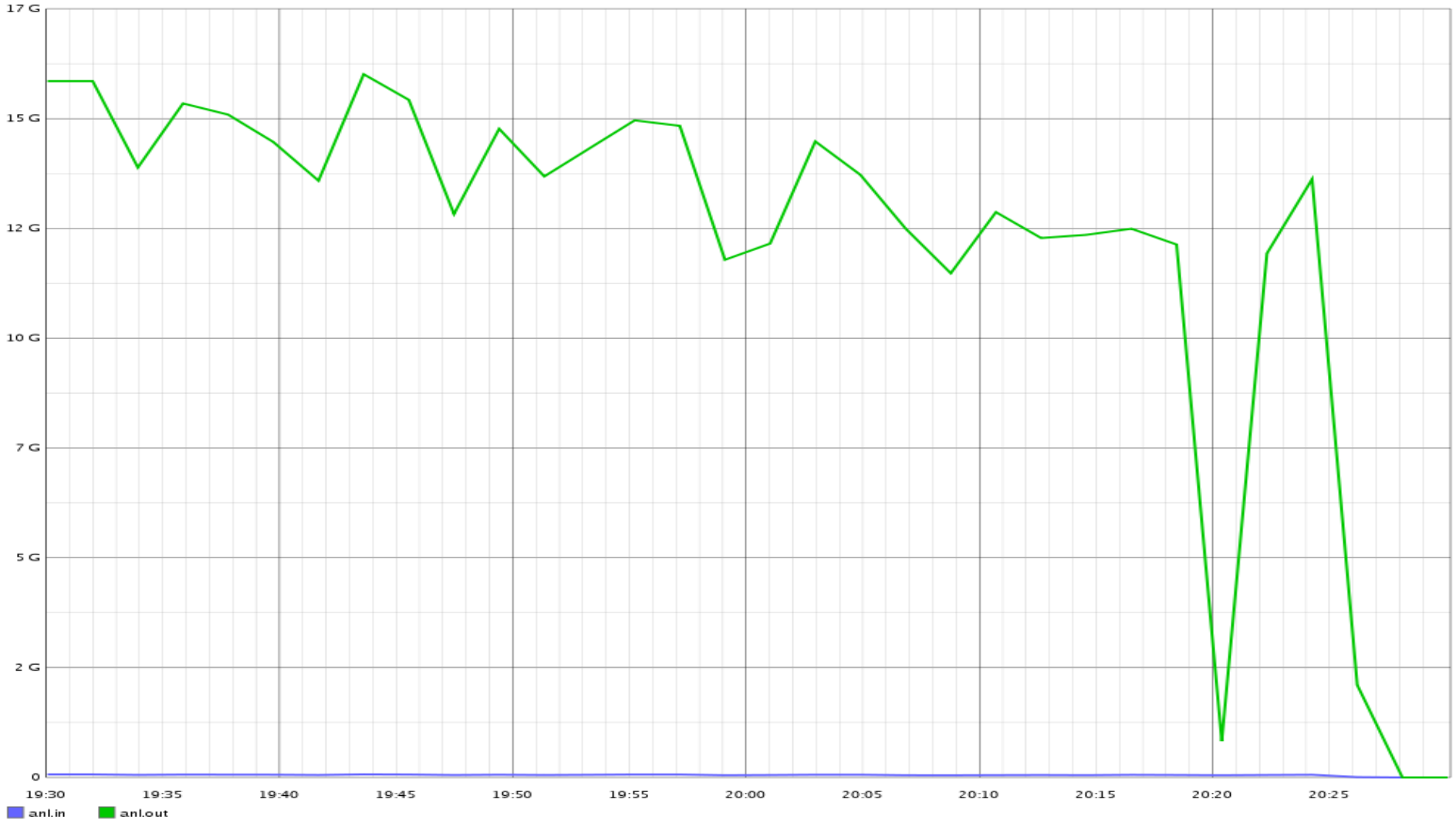
- **High-performance, reliable data transfer protocol optimized for high-bandwidth wide-area networks**
- **Globus GridFTP**
  - Parallel TCP streams, optimal TCP buffer
  - Non TCP protocol such as UDT
  - Cluster-to-cluster data movement
  - SSH, GSI
  - Restartable transfers

GridFTP Servers round the World

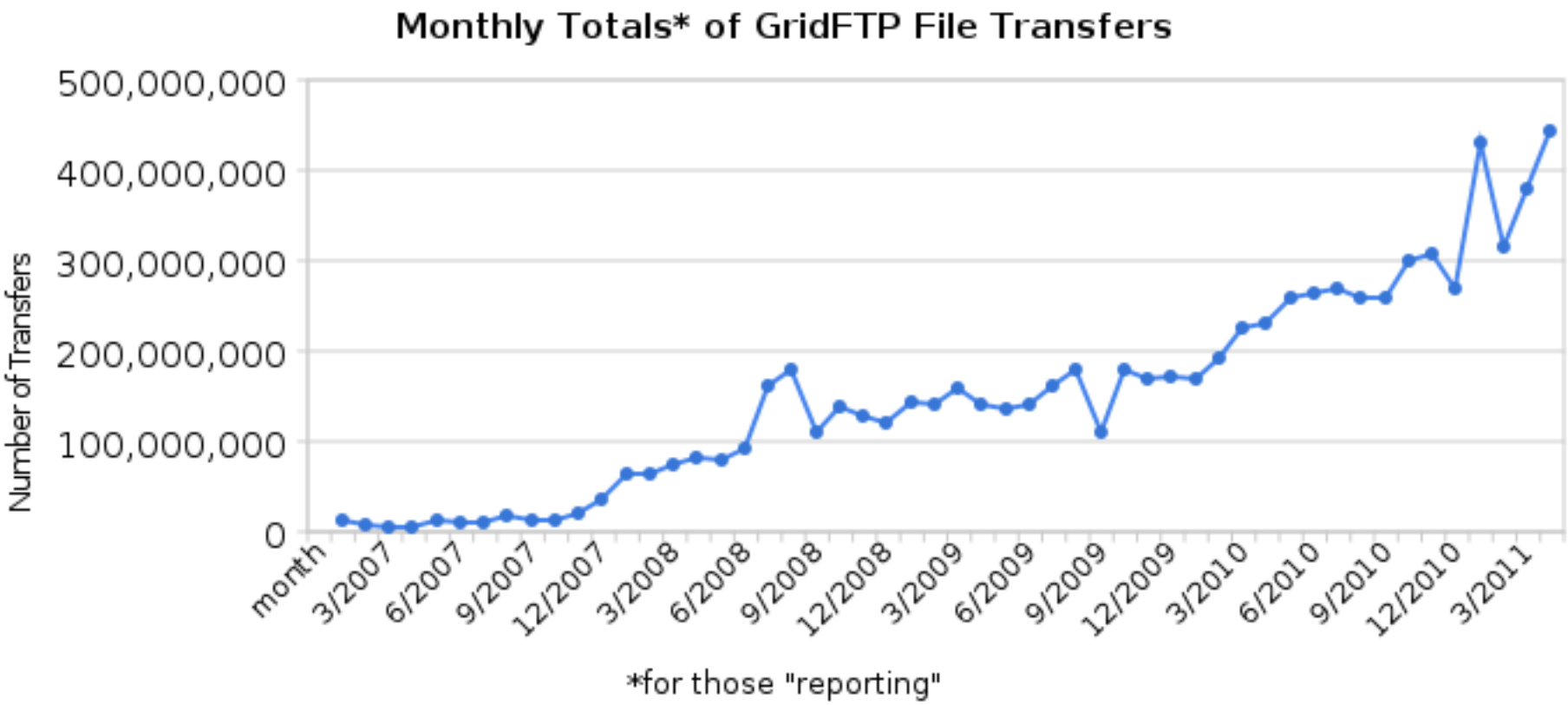


Created by Tim Pinkawa (Northern Illinois University) using MaxMind's GeoIP technology (http://www.maxmind.com/app/ip-locate).

Performance

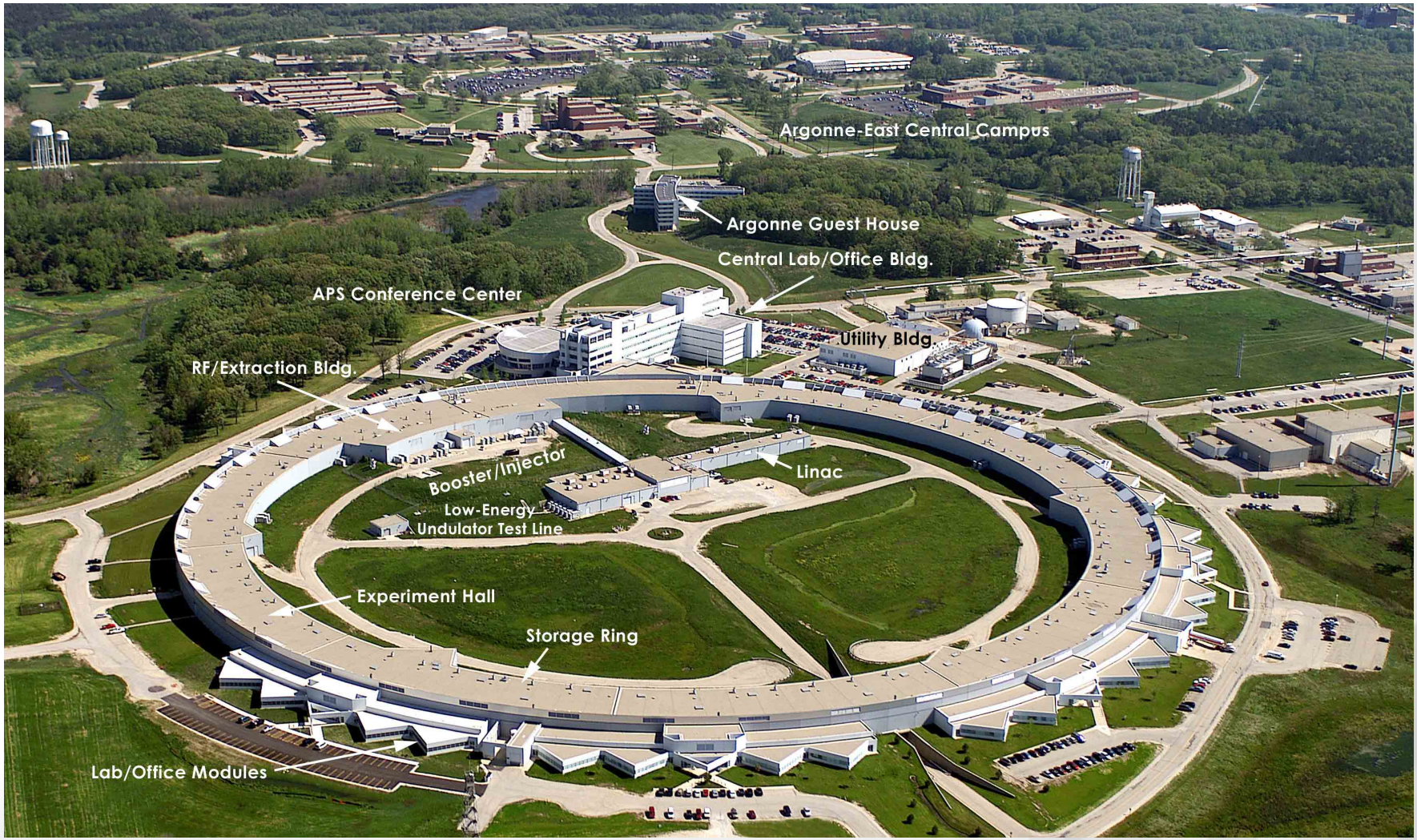


GridFTP Usage



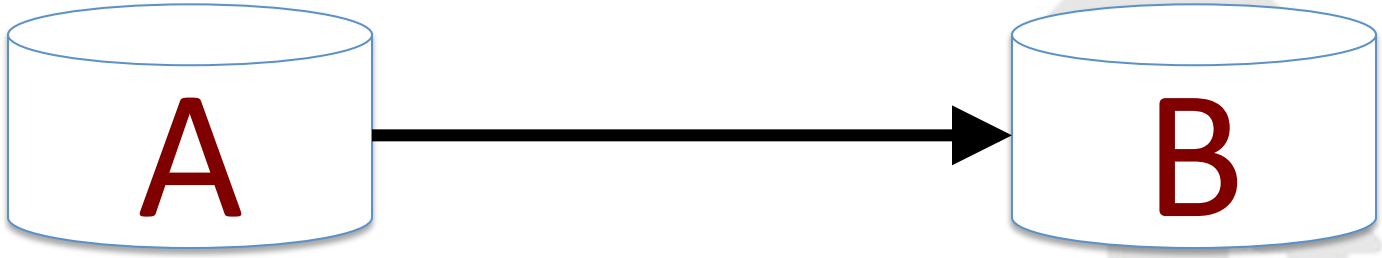
Data Movement Challenges

Advanced Photon Source



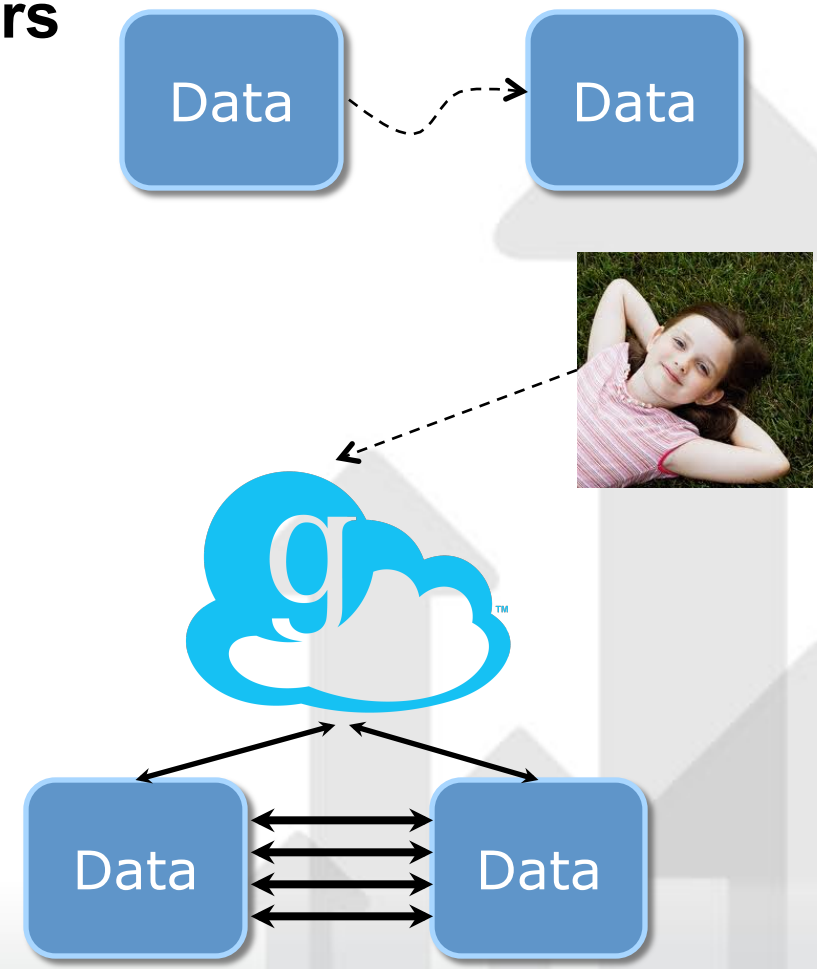
Data Movement Challenges

Discover endpoints, determine available protocols, negotiate firewalls, configure software, manage space, determine required credentials, configure protocols, detect and respond to failures, determine expected performance, determines actual performance, identify diagnose and correct network misconfigurations, integrate with file systems, …



Globus Online

- **Easy "fire and forget" file transfers**
- **Automatic fault recovery**
- **High performance**
- **Simplify use of multiple security domains**
- **No client software installation**
- **New features automatically available**
- **Consolidated support and troubleshooting**



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 Online

Command line interface
ls alcf#dtn:~
scp alcf#dtn:~/myfile \
nersc#dtn:~/myfile

HTTP REST interface
POST https://transfer.api.globusonline.org/ v0.10/transfer <transfer-doc>

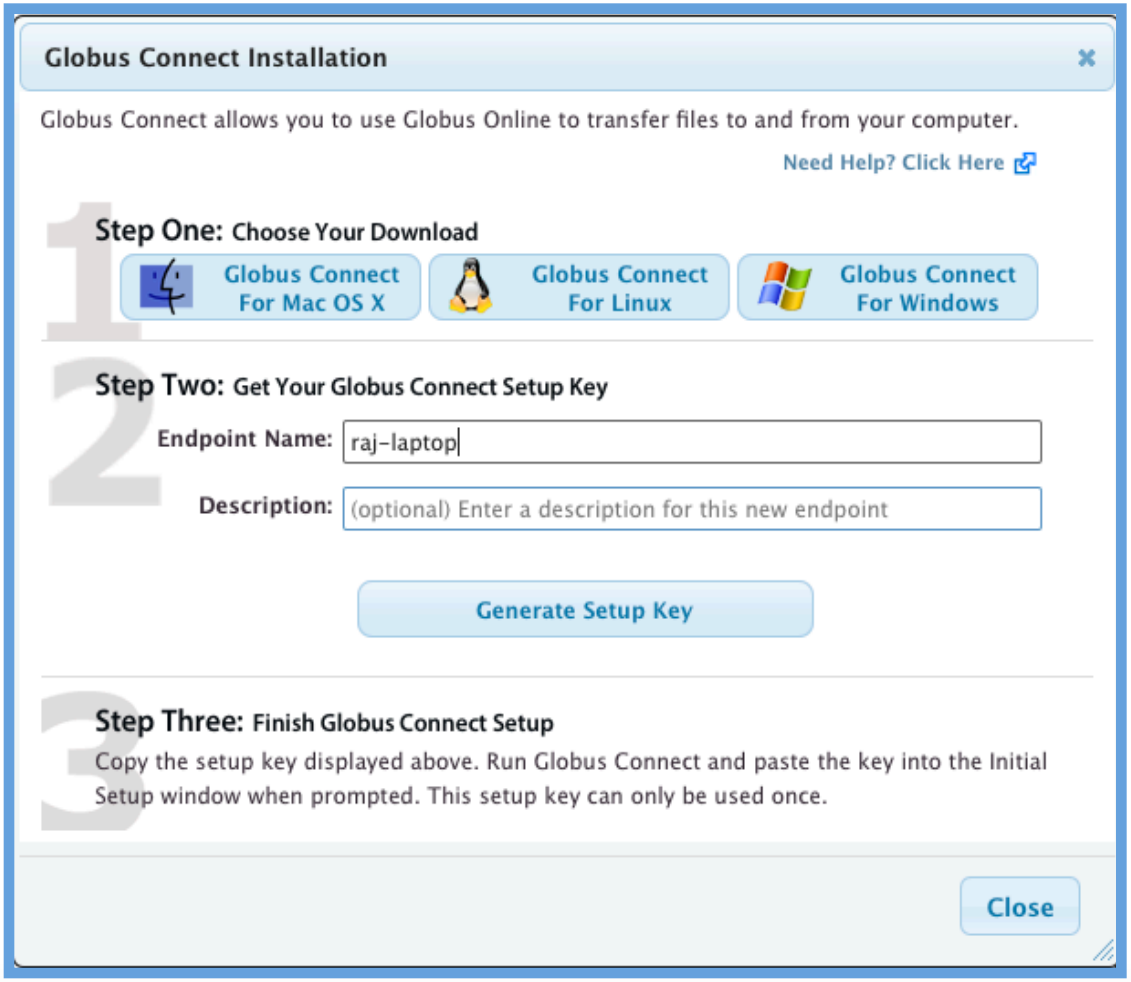
Fire-and-forget data movement
Many files and lots of data
Third-party transfers
Performance optimization
Expert operations and monitoring

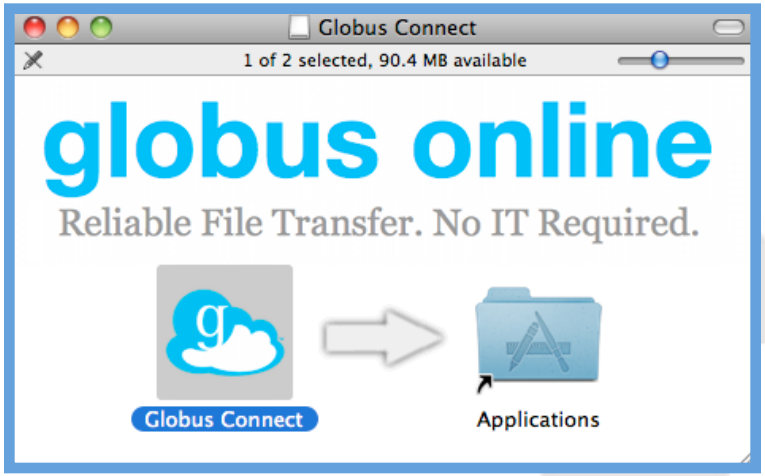
GridFTP servers
FTP servers

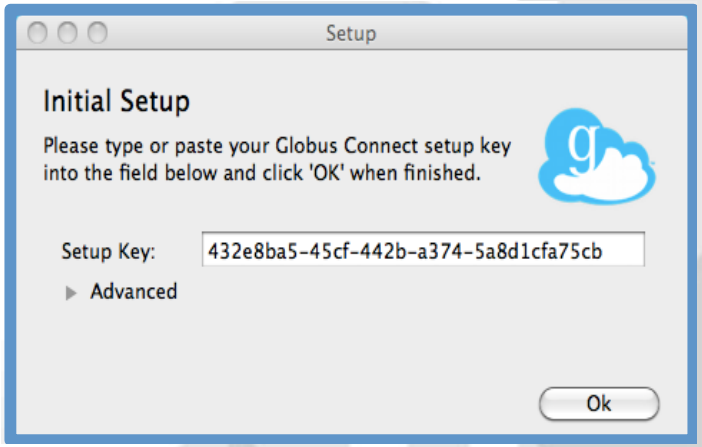
High-performance data transfer nodes

Globus Connect on local computers

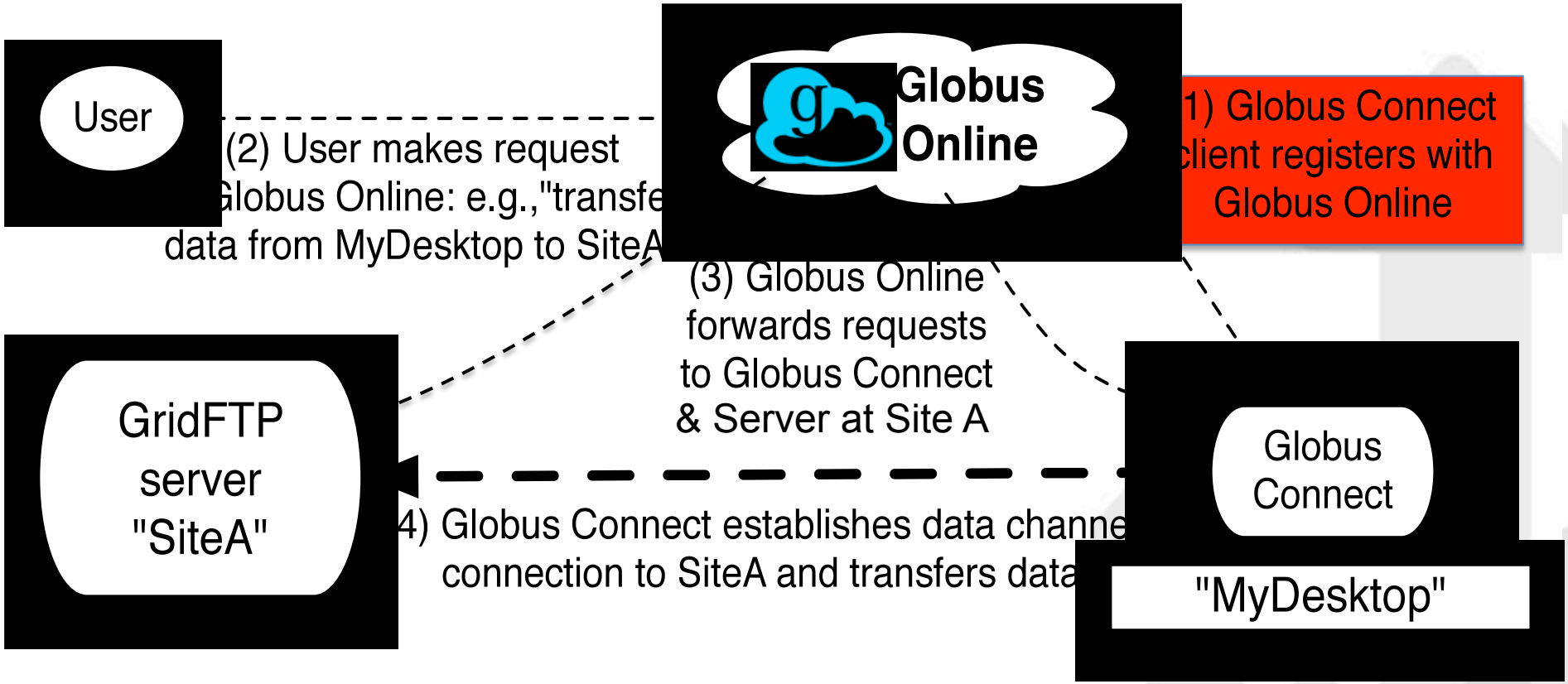
Globus Connect



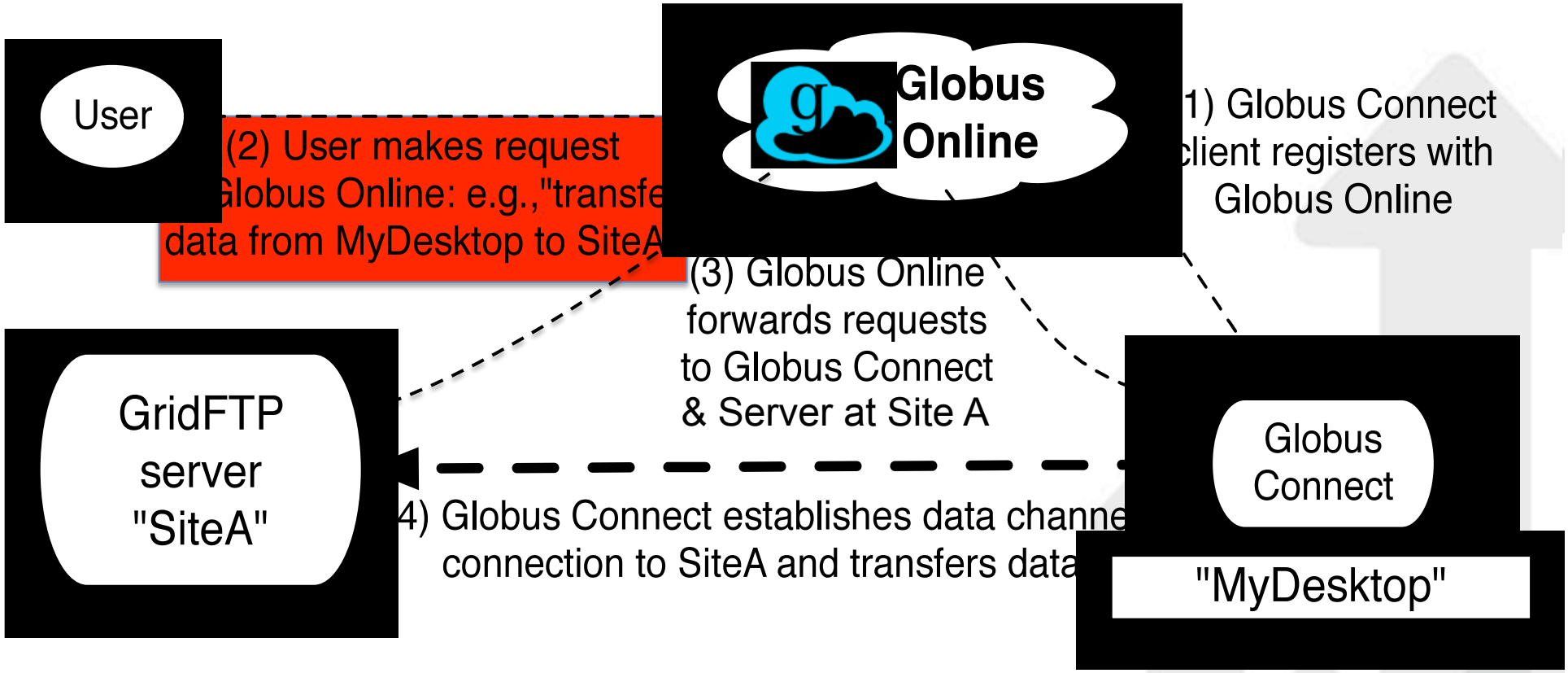




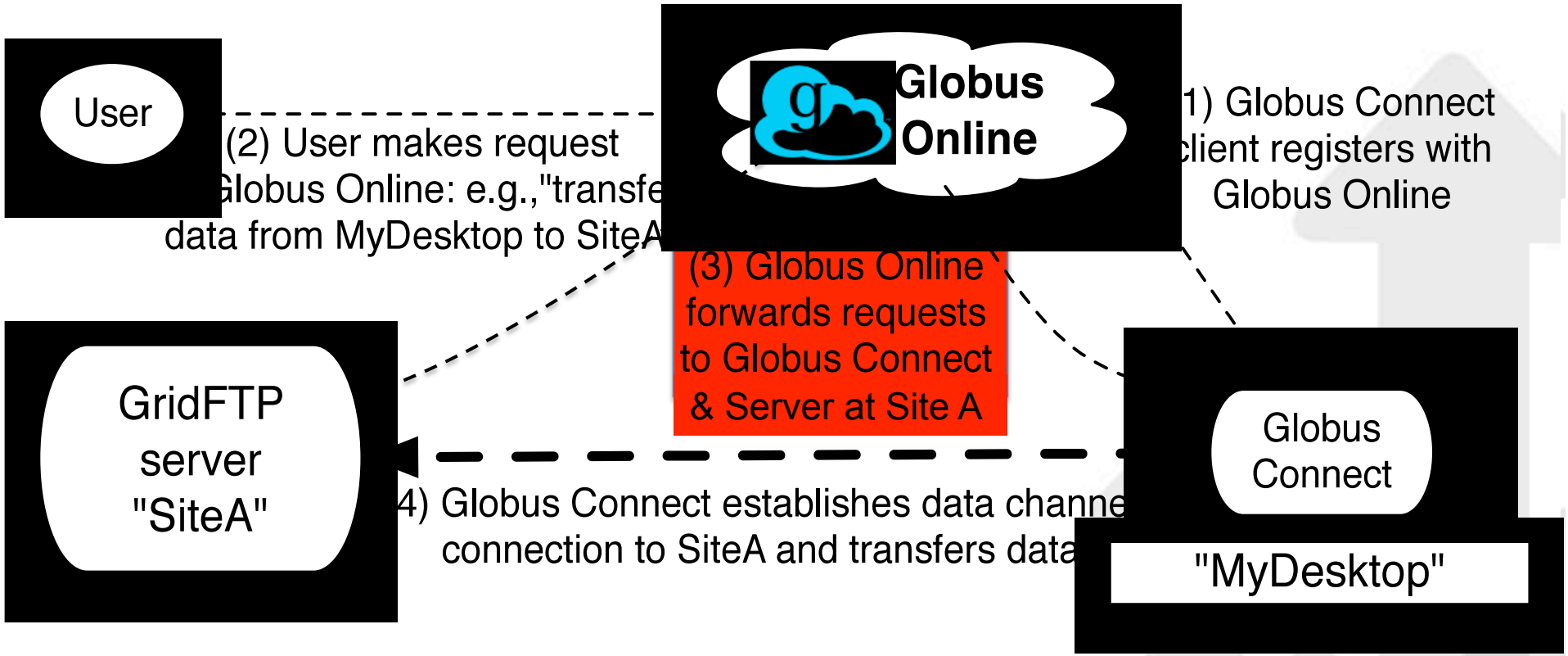
Globus Connect in Action



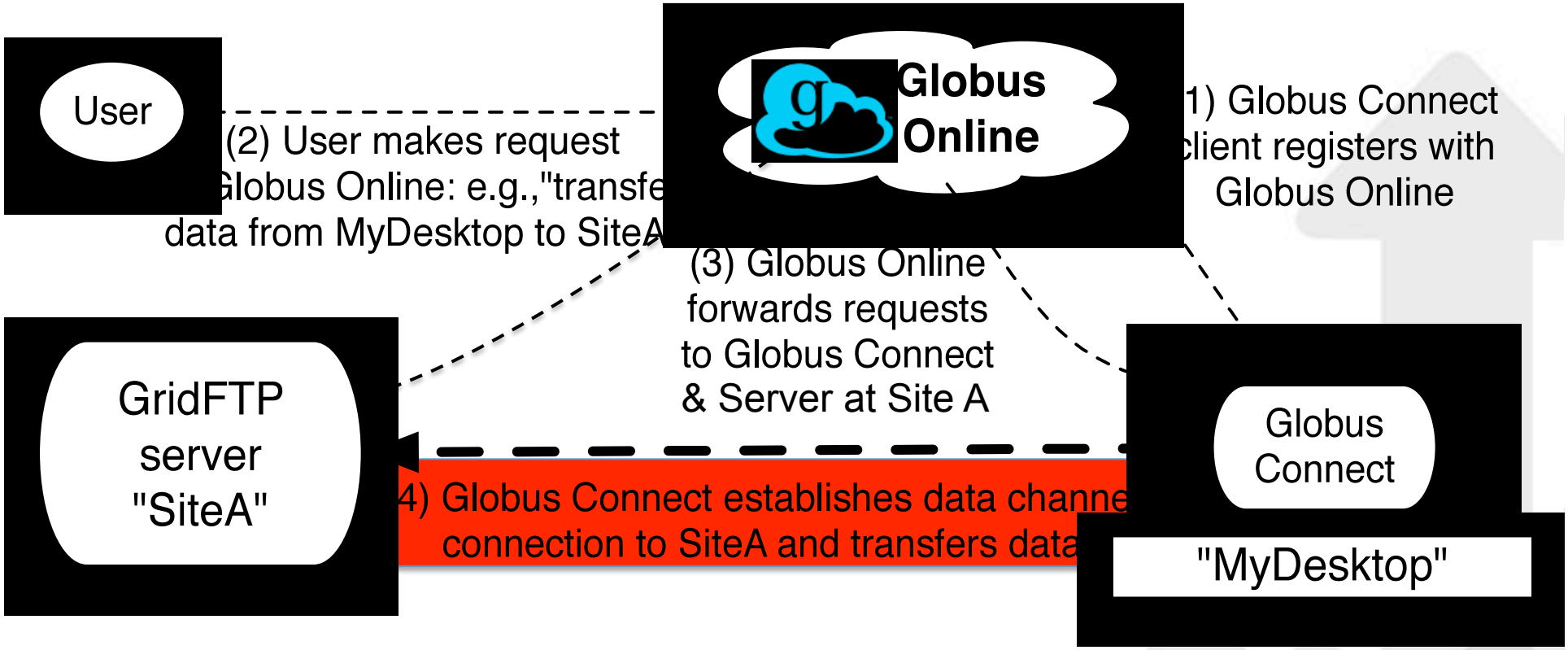
Globus Connect in Action



Globus Connect in Action



Globus Connect in Action



CLI 2.0

```
ssh -t <user>@cli.globusonline.org <command> <options> <params>
```

Globus Online user name

Globus Online command

Use as needed to hide password text

```
gsissh <user>@cli.globusonline.org <command> <options> <params>
```

CLI commands

cancel

details

endpoint-activate

endpoint-add

endpoint-deactivate

endpoint-list

endpoint-modify

endpoint-remove

endpoint-rename

events

ls

profile

scp

status

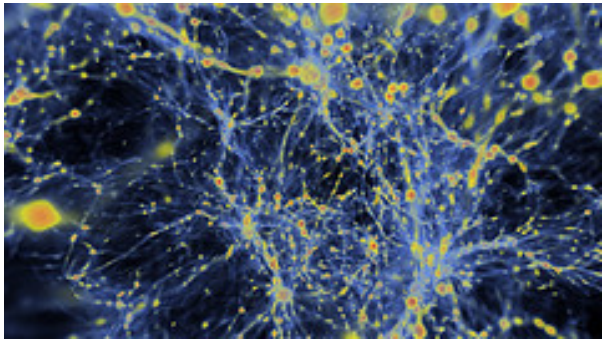
transfer

versions

wait

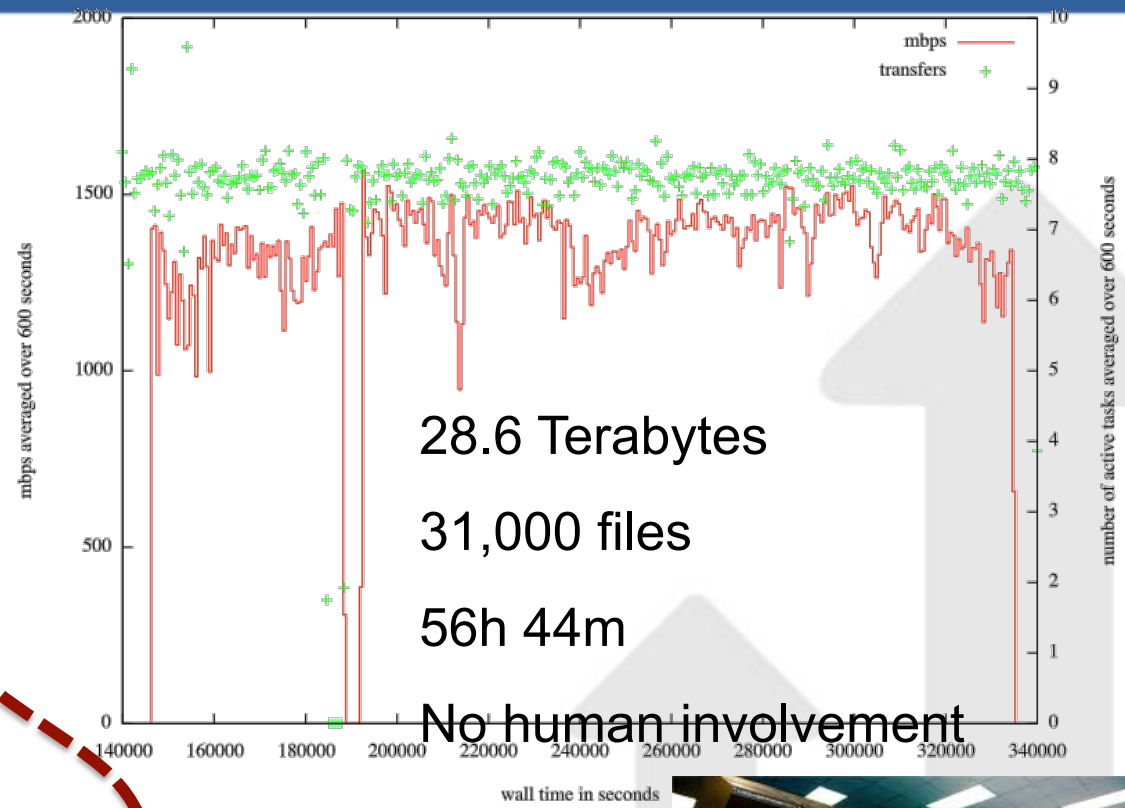
Globus Online

Globus Online in Action



Astrophysics simulation data generated in Tennessee, moved to Illinois for visualization

(Enzo, UCSD; Futures Lab, Argonne)

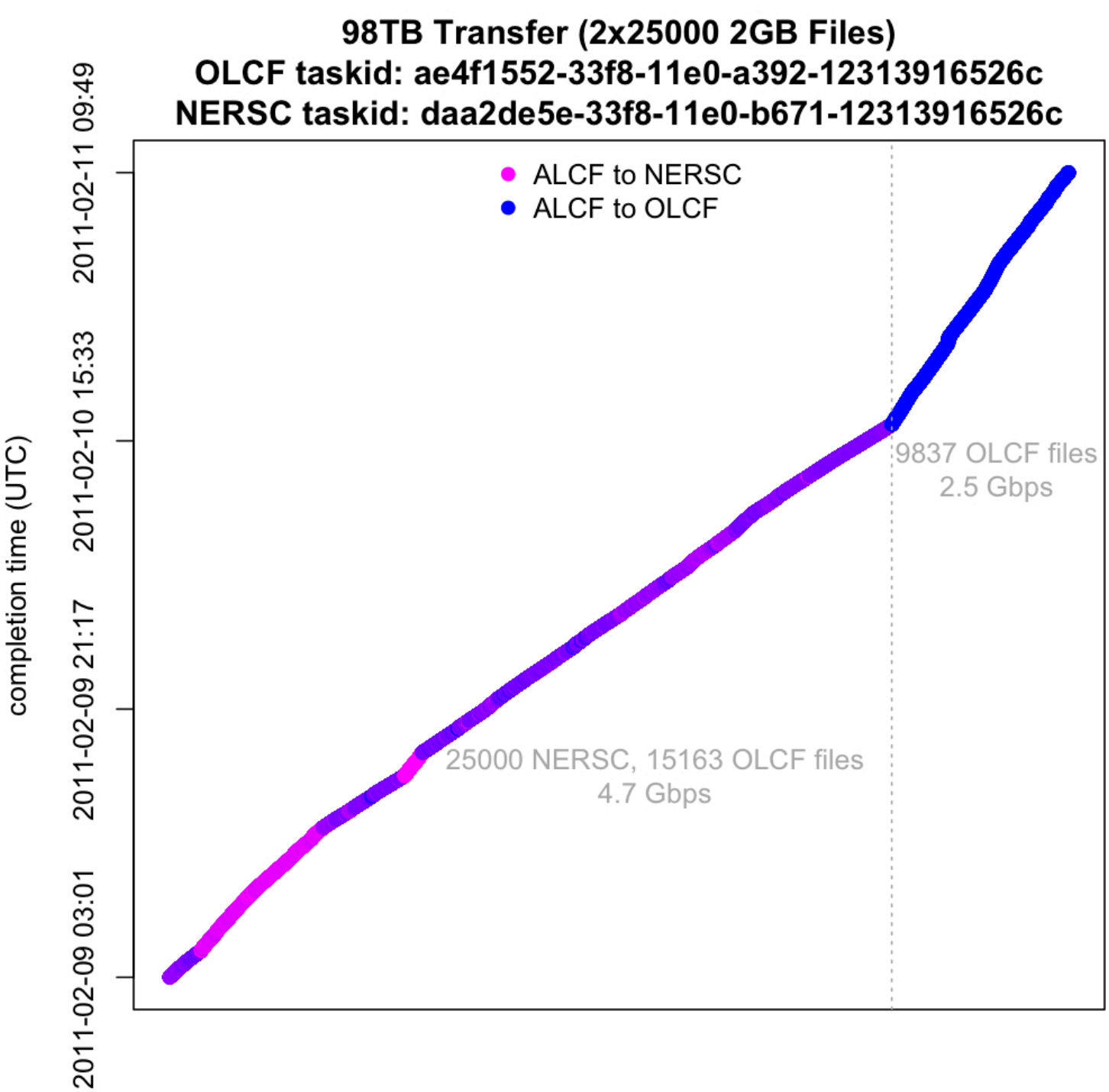


28.6 Terabytes

31,000 files

56h 44m

No human involvement


98TB Transfer (2x25000 2GB Files)
OLCF taskid: ae4f1552-33f8-11e0-a392-12313916526c
NERSC taskid: daa2de5e-33f8-11e0-b671-12313916526c
ALCF to NERSC
ALCF to OLCF
completion time (UTC)
2011-02-09 03:01
2011-02-09 21:17
2011-02-10 15:33
2011-02-11 09:49
25000 NERSC, 15163 OLCF files
4.7 Gbps
9837 OLCF files
2.5 Gbps

Acknowledgments

Numerous people have contributed to the Globus Online work, including:

Bryce Allen, Joshua Boverhof, John Bresnahan, Lisa Childers, Paul Dave', Fred Dech, Ian Foster, Dan Gunter, Gopi Kandaswany, Nick Karonis, Raj Kettimuthu, Jack Kordas, Lee Liming, Mike Link, Stu Martin, JP Navarro, Karl Pickett, Mei Hui Su, Steve Tuecke, Vas Vasiliadis

Many thanks to our funders: DOE, NSF, and the University of Chicago

More Information at
http://www.globusonline.org/

Questions