



Globus Online for APS Data Movement

Raj Kettimuthu

5/3/2011

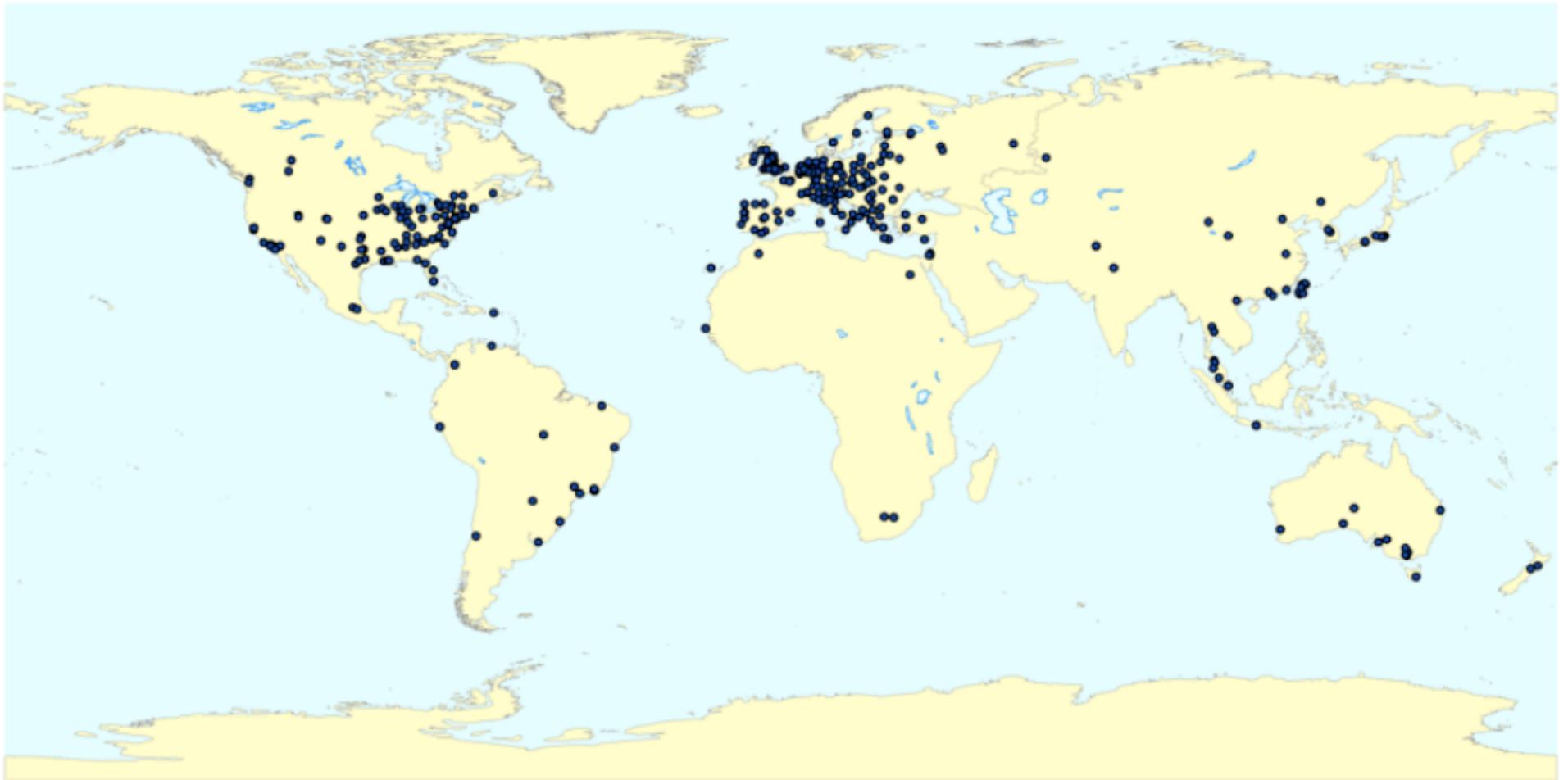


GridFTP

- **High-performance, reliable data transfer protocol optimized for high-bandwidth wide-area networks**
- **Globus GridFTP**
 - High Performance
 - Parallel TCP streams, Non TCP protocol such as UDT, Cluster-to-cluster data movement
 - Secure
 - Synchronize data sets
 - Checkpointing



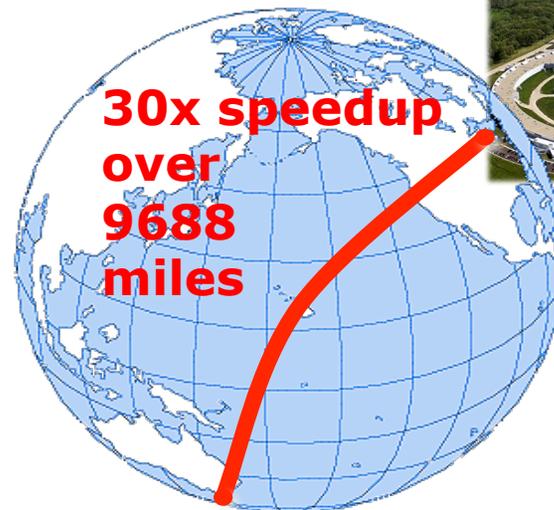
GridFTP



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



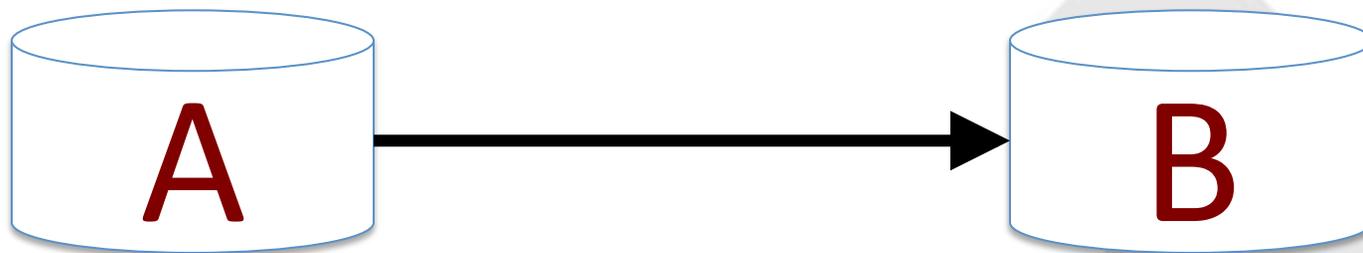
GridFTP data movement APS to Australia



One terabyte moved from an
Advanced Photon Source
to Australia, at a rate 30x
faster than SCP



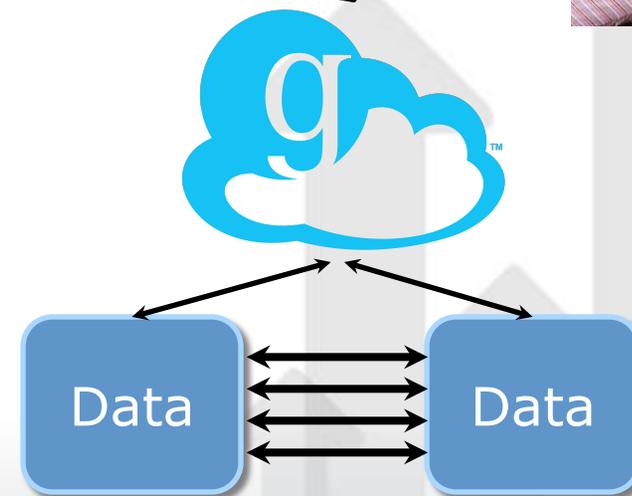
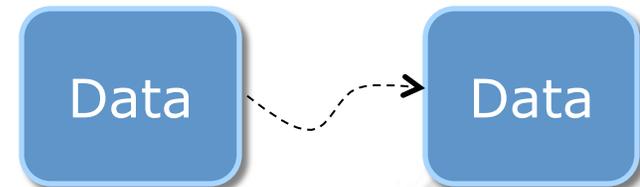
Discover endpoints, determine available protocols, negotiate firewalls, configure software, manage space, determine required credentials, configure protocols, detect and respond to failures, determine expected performance, determines actual performance, identify diagnose and correct network misconfigurations,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





Data Movement Scenarios

- **Transfer between two GridFTP server**
 - Error recovery
 - Auto tuning
- **Upload/download data to/from a GridFTP server to/from a user's machine**
 - Typically no GridFTP server on user's machine
 - GridFTP installation and security configuration
 - Globus Connect
- **Transfer between endpoints – neither has GridFTP server running**
 - Globus Connect for departments



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 Installation

Globus Connect allows you to use Globus Online to transfer files to and from your computer. [Need Help? Click Here](#)

Step One: Choose Your Download

Globus Connect For Mac OS X Globus Connect For Linux Globus Connect For Windows

Step Two: Get Your Globus Connect Setup Key

Endpoint Name:

Description:

[Generate Setup Key](#)

Step Three: Finish Globus Connect Setup

Copy the setup key displayed above. Run Globus Connect and paste the key into the Initial Setup window when prompted. This setup key can only be used once.

[Close](#)

Globus Connect

1 of 2 selected, 90.4 MB available

globus online

Reliable File Transfer. No IT Required.

[Globus Connect](#) Applications

Setup

Initial Setup

Please type or paste your Globus Connect setup key into the field below and click 'OK' when finished.

Setup Key:

▶ Advanced

[Ok](#)

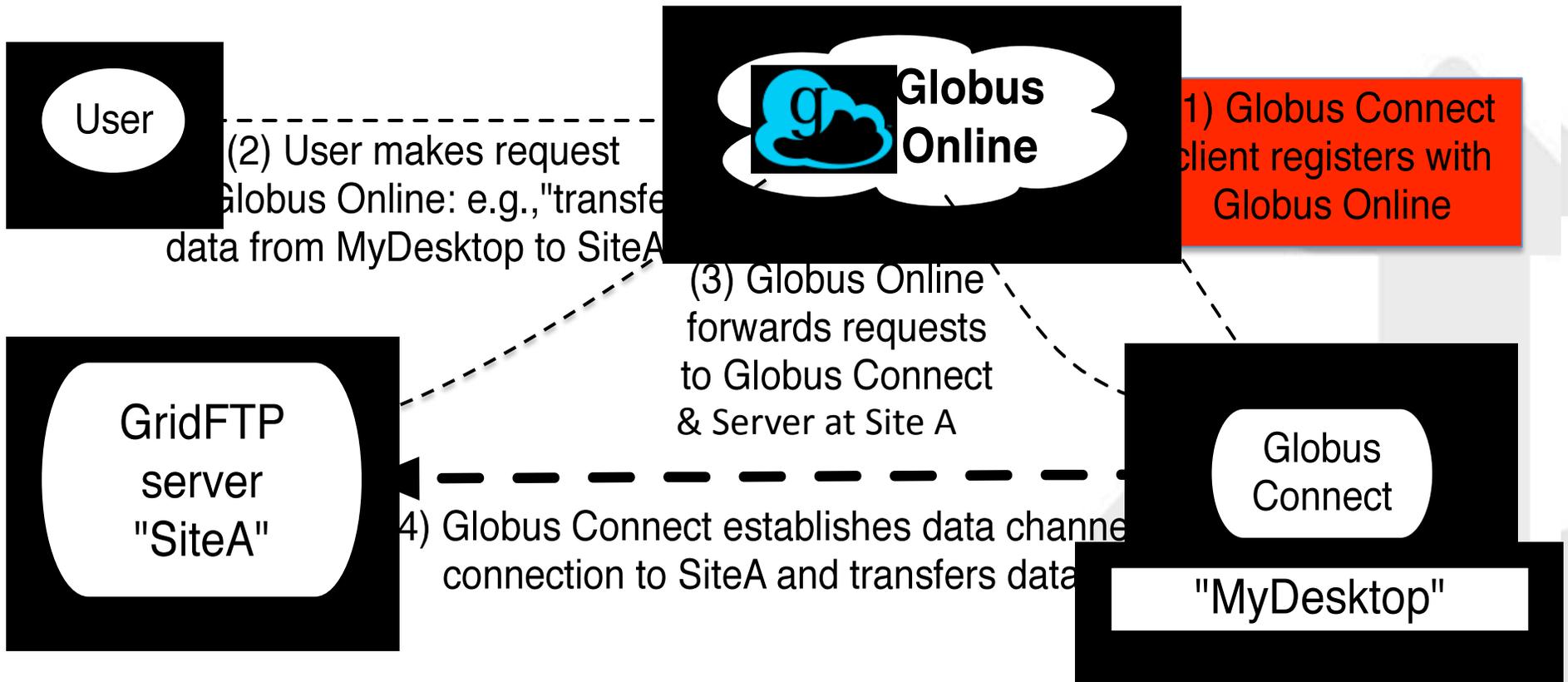


Globus Connect

- <http://confluence.globus.org/display/GOPUB/Globus+Connect>

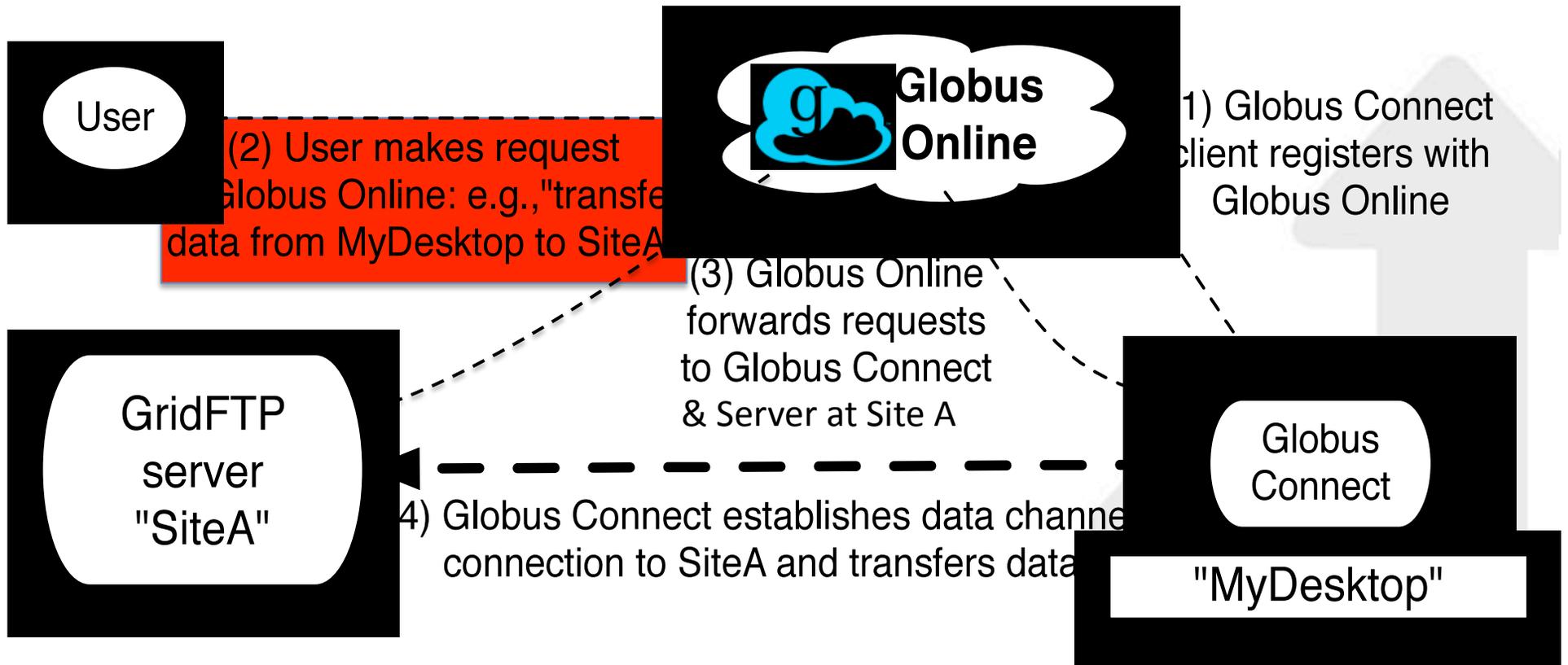


Globus Connect in Action



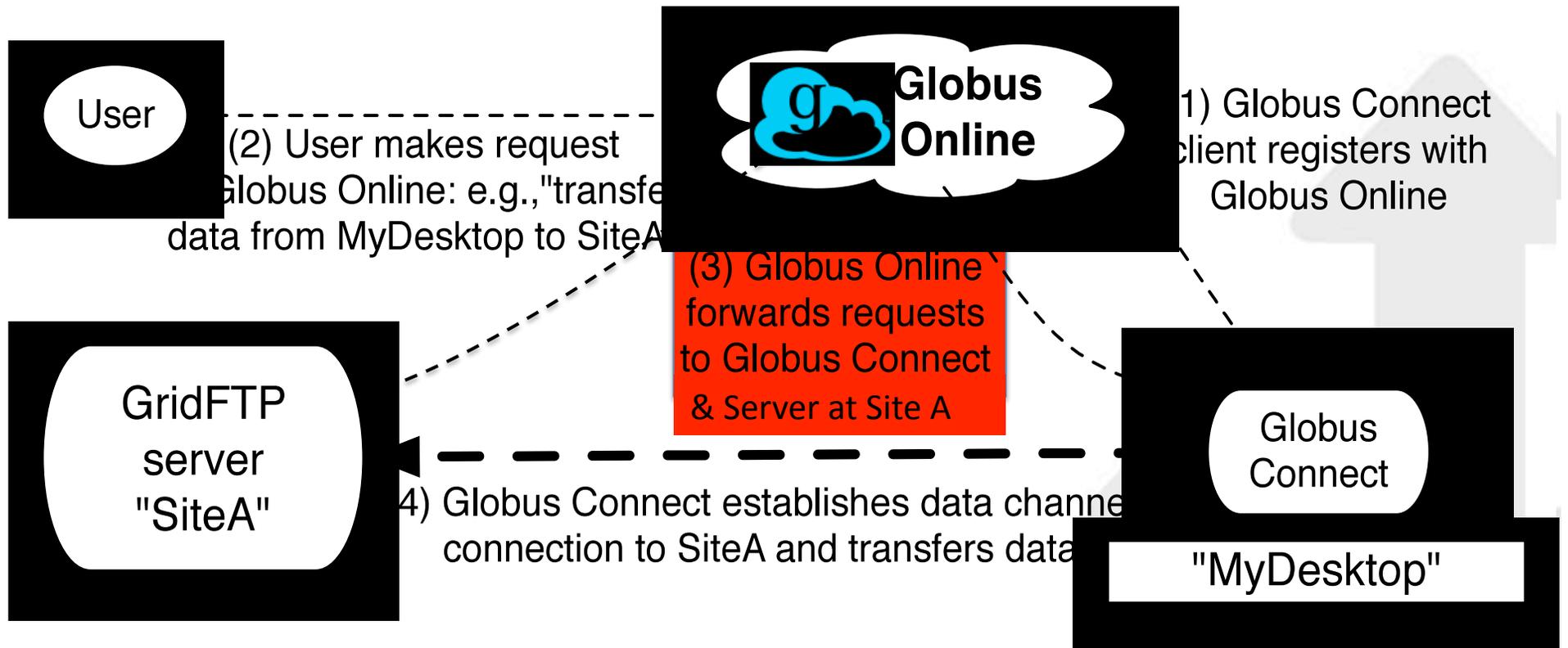


Globus Connect in Action



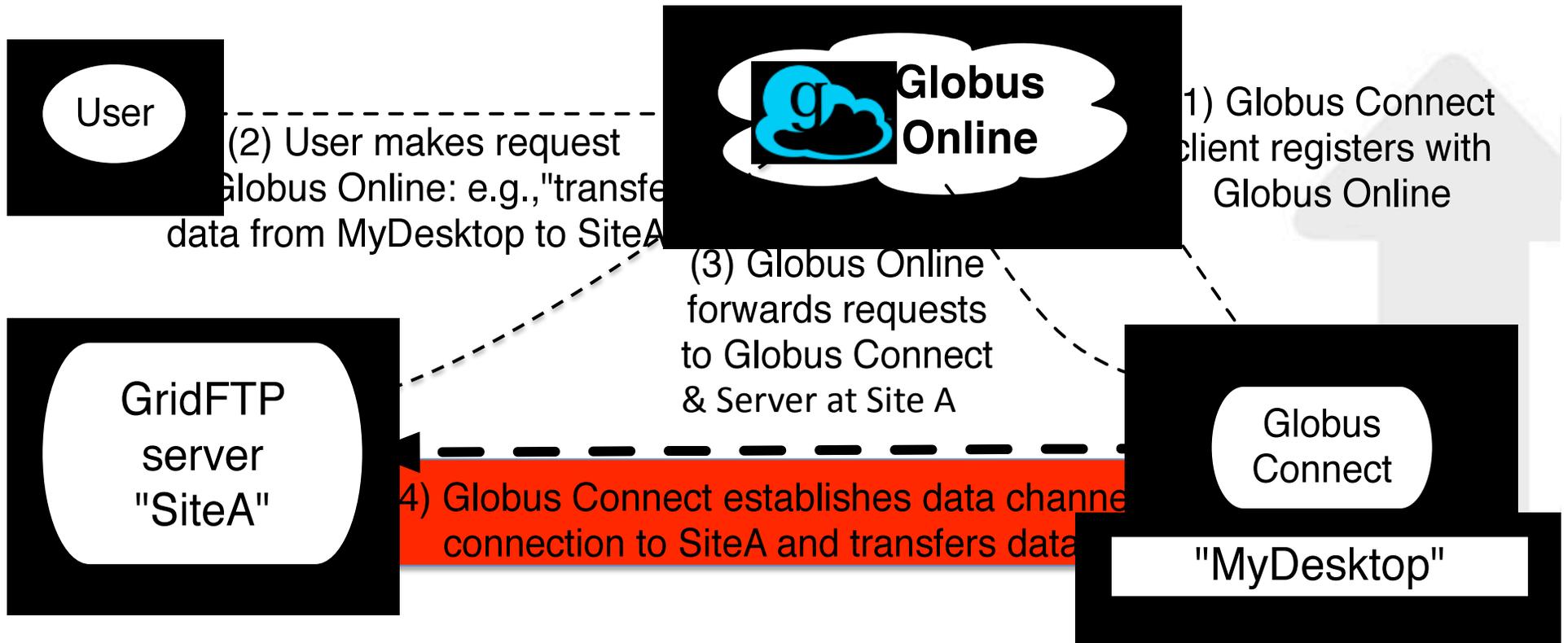


Globus Connect in Action





Globus Connect in Action



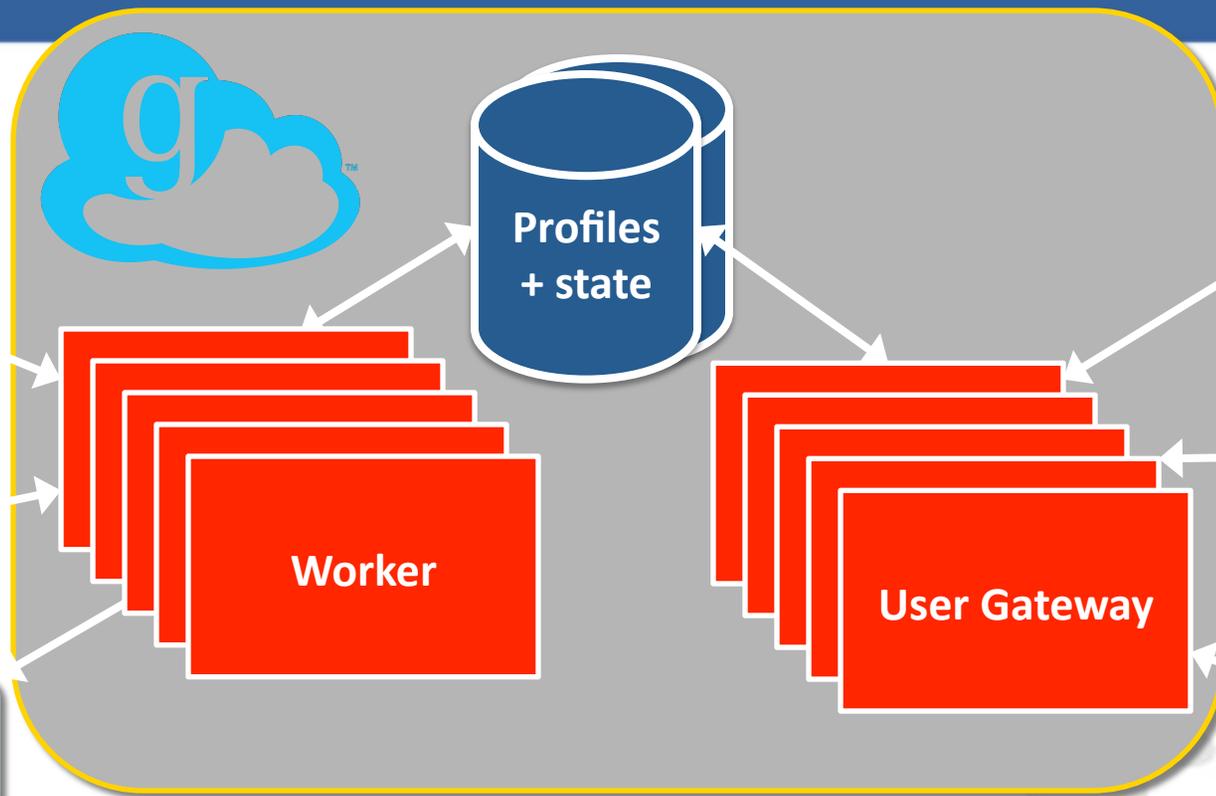


Globus Online

GridFTP

GridFTP

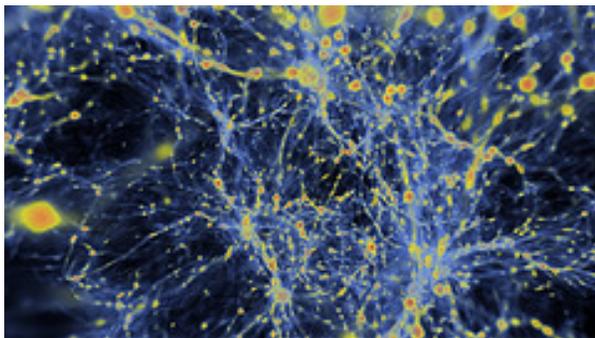
Notification target



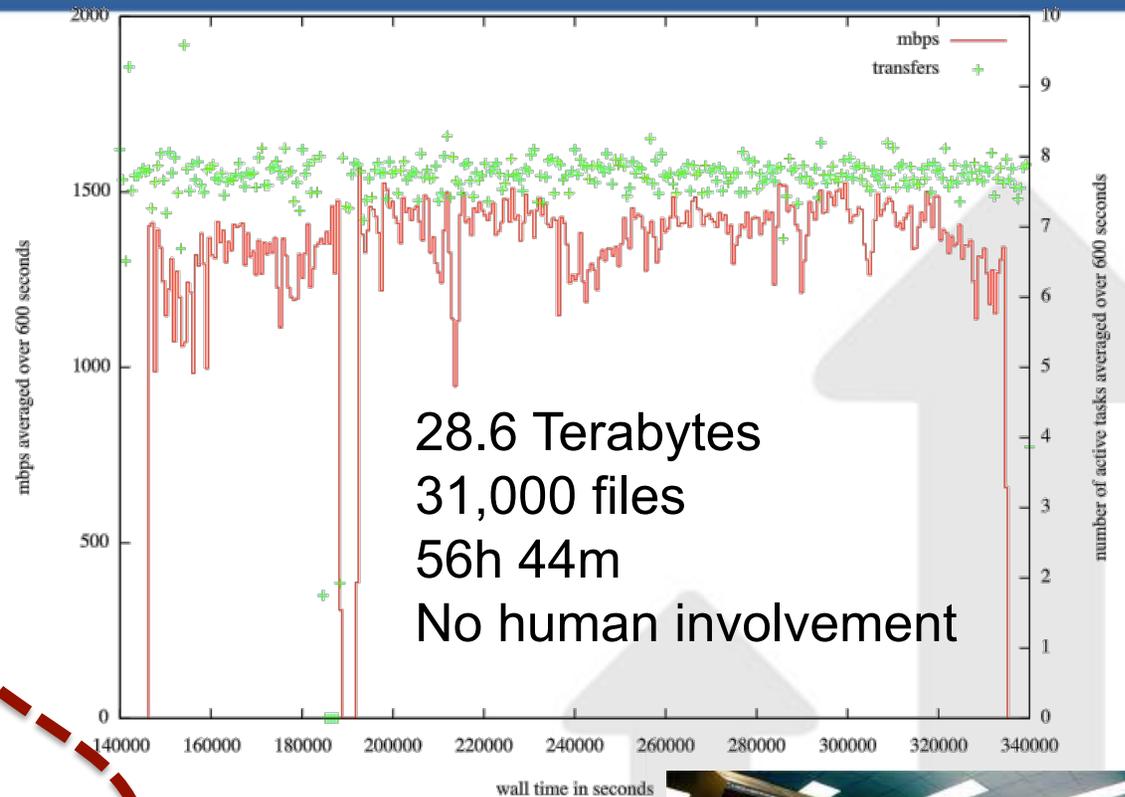
amazon.com[®]
web services



Globus Online in Action



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



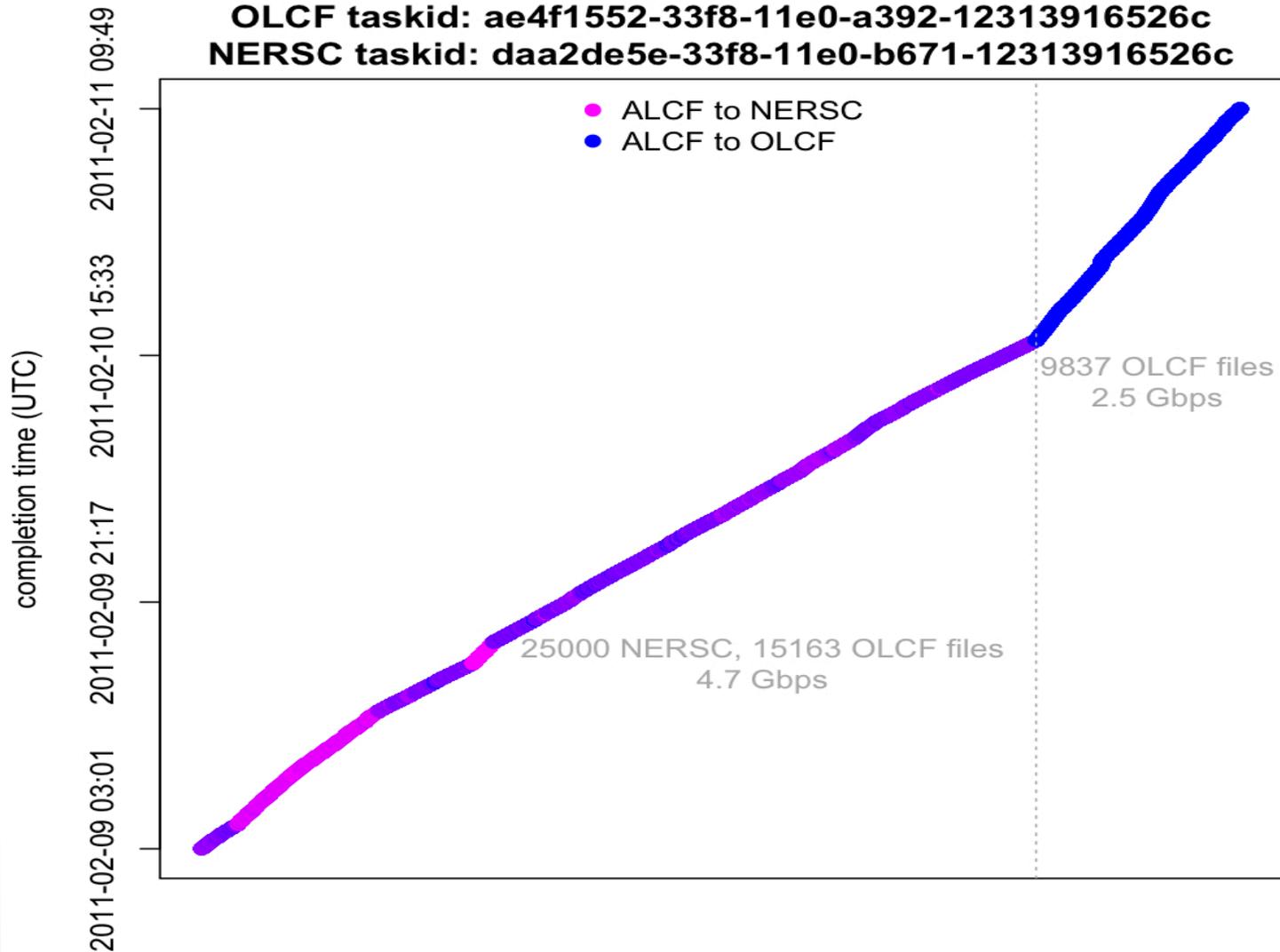


Performance Diagnosis

98TB Transfer (2x25000 2GB Files)

OLCF taskid: ae4f1552-33f8-11e0-a392-12313916526c

NERSC taskid: daa2de5e-33f8-11e0-b671-12313916526c





Auto tuning

- **Tuning GridFTP to maximize performance is tricky**
- **Depends on the characteristics of the data sets**
 - Large files
 - Lots of small files
- **Depends on the characteristics of the network**
 - Dedicated network
 - Shared network
- **Globus Online does auto tuning**
 - Groups files into different buckets
 - Provides manual control for advanced users



Sync feature

- **Feature to synchronize datasets**
- **Only transfer files where the destination does not exist or differs (size, timestamp, checksum) from the source**
- **'-sync-level'**
 - 0 - transfer if the destination does not exist
 - 1 - transfer if the size does not match
 - 2 - transfer if timestamp of dst is older
 - 3 - transfer if the checksums do not match.
 - The default sync level is 2.



Globus Online for APS Users



Globus Online - hosted service for high-speed, reliable, secure data movement

Advanced Photon Source



GridFTP Server



APS Beamline
(Data Acquisition)



HPC Cluster



GridFTP Server

External Data Analysis / Visualization



GridFTP Server

User's Home Institution

Globus Connect

User's Desktop/
Laptop



Globus Online for APS Users



Globus Online - hosted service for high-speed, reliable, secure data movement

Advanced Photon Source



GridFTP Server



APS Beamline
(Data Acquisition)



HPC Cluster

GridFTP Server

External Data Analysis / Visualization



GridFTP Server

User's Home Institution

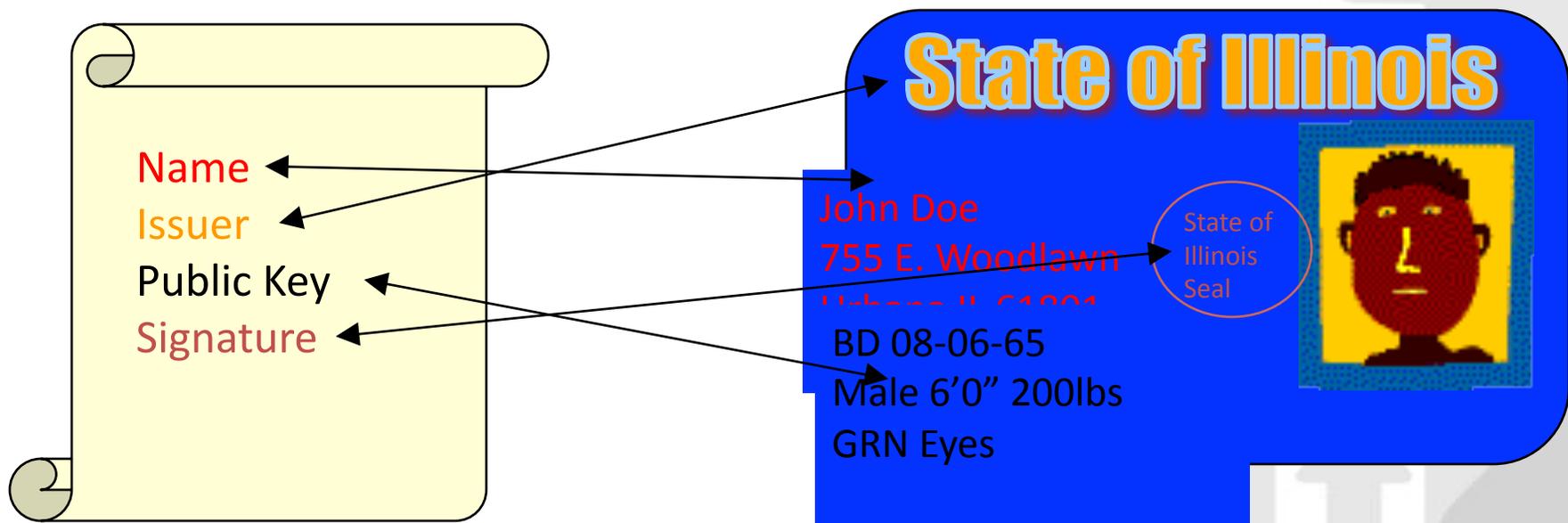
Globus Connect

User's Desktop/
Laptop



Certificates

- Digital document that certifies a public key is owned by a particular user
- Signed by 3rd party – the Certificate Authority (CA)
- X509 standard



To know if you should trust the certificate, you have to trust the CA

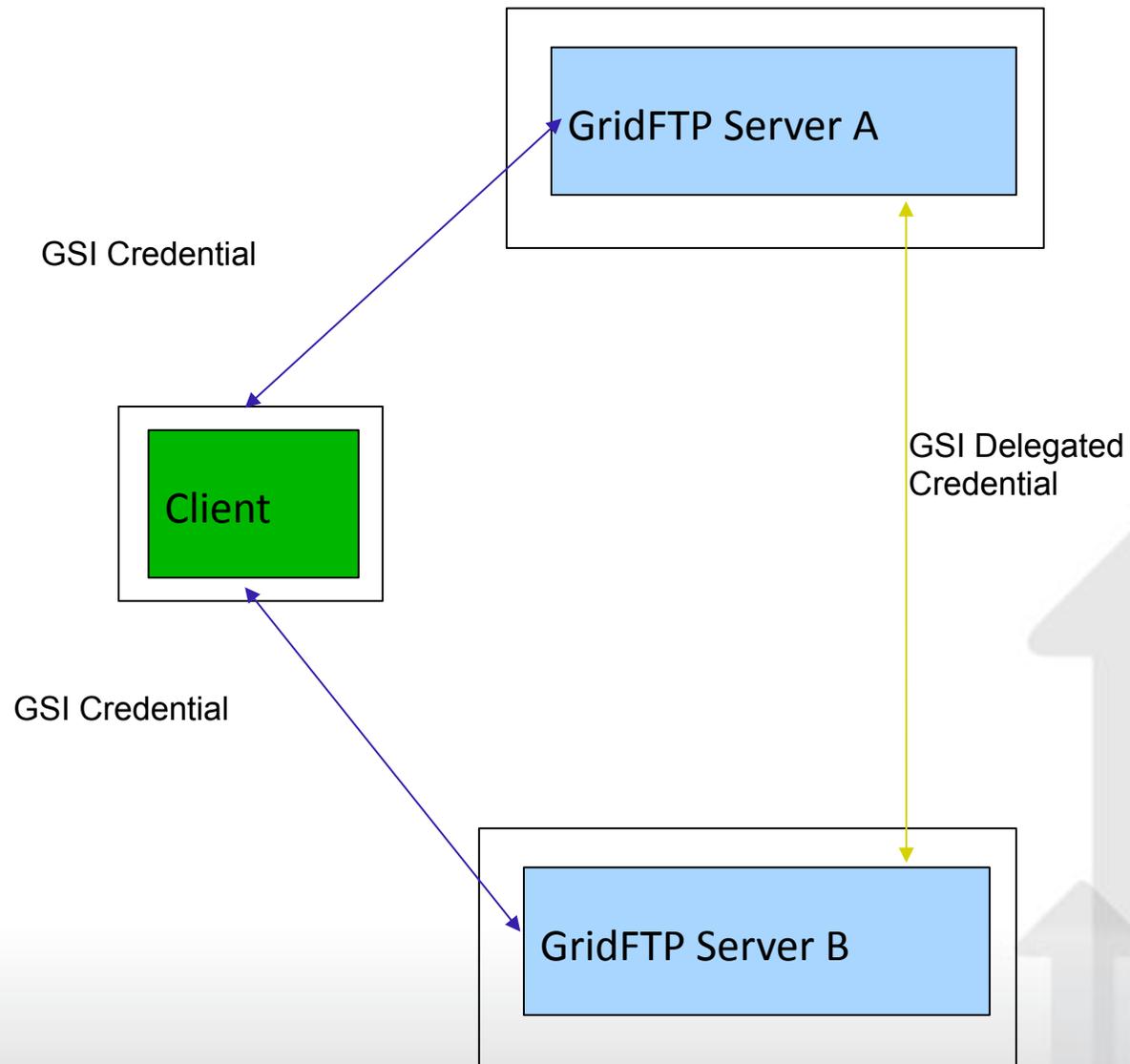


GSI Authentication

- **Public key based security**
 - X509 certificates
 - Short lived proxy credentials
- **Delegation**
 - Authenticates DC on clients behalf
 - Flexibility for automated workflows
 - Agents can authenticate to GridFTP servers on users behalf
 - Enables encryption, integrity on data channel

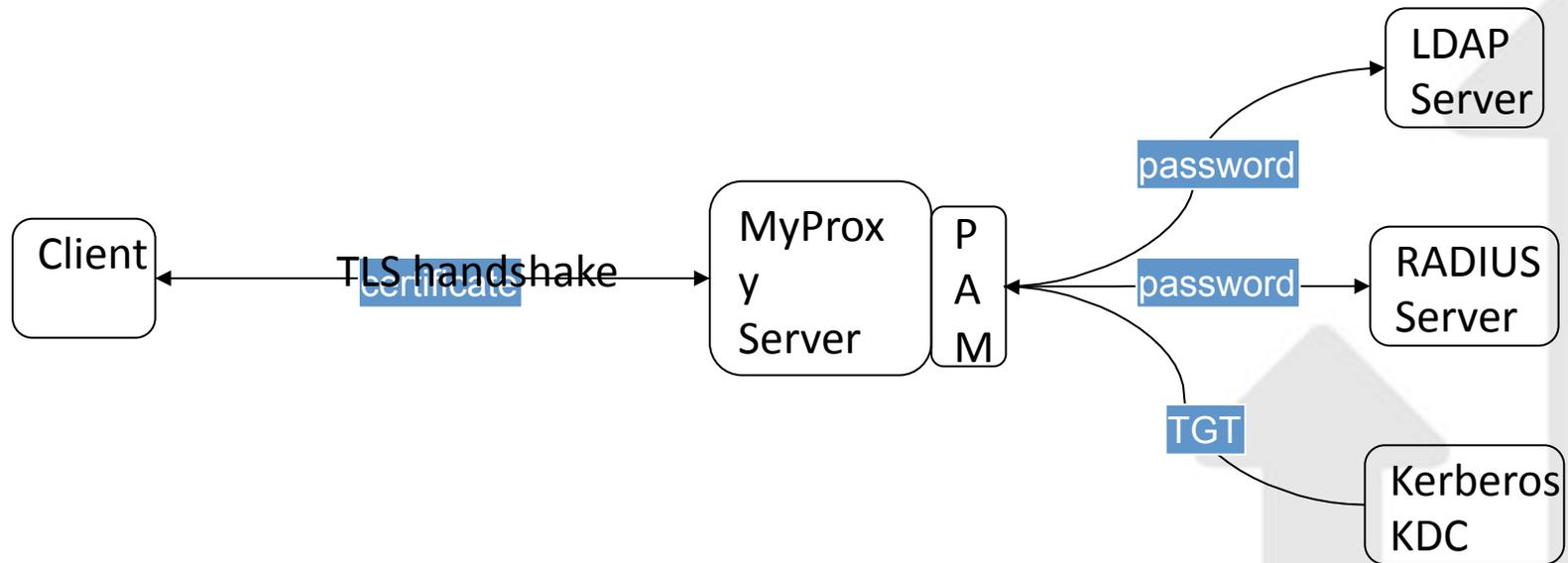


GSI Authentication





Myproxy CA with PAM





Globus Connect for Departments

- **Current Globus Connect is for individuals**
- **Many small lab and departments need Globus Connect like solution to run GridFTP with ease**
 - Single click GridFTP and Myproxy installation and setup
 - Used by all its users
 - Transfer between endpoints – neither has GridFTP server running - trivial



Questions

