

# Introduction to Globus.org

Lisa Childers

OSG Summer School

22 July 2010

# Outline

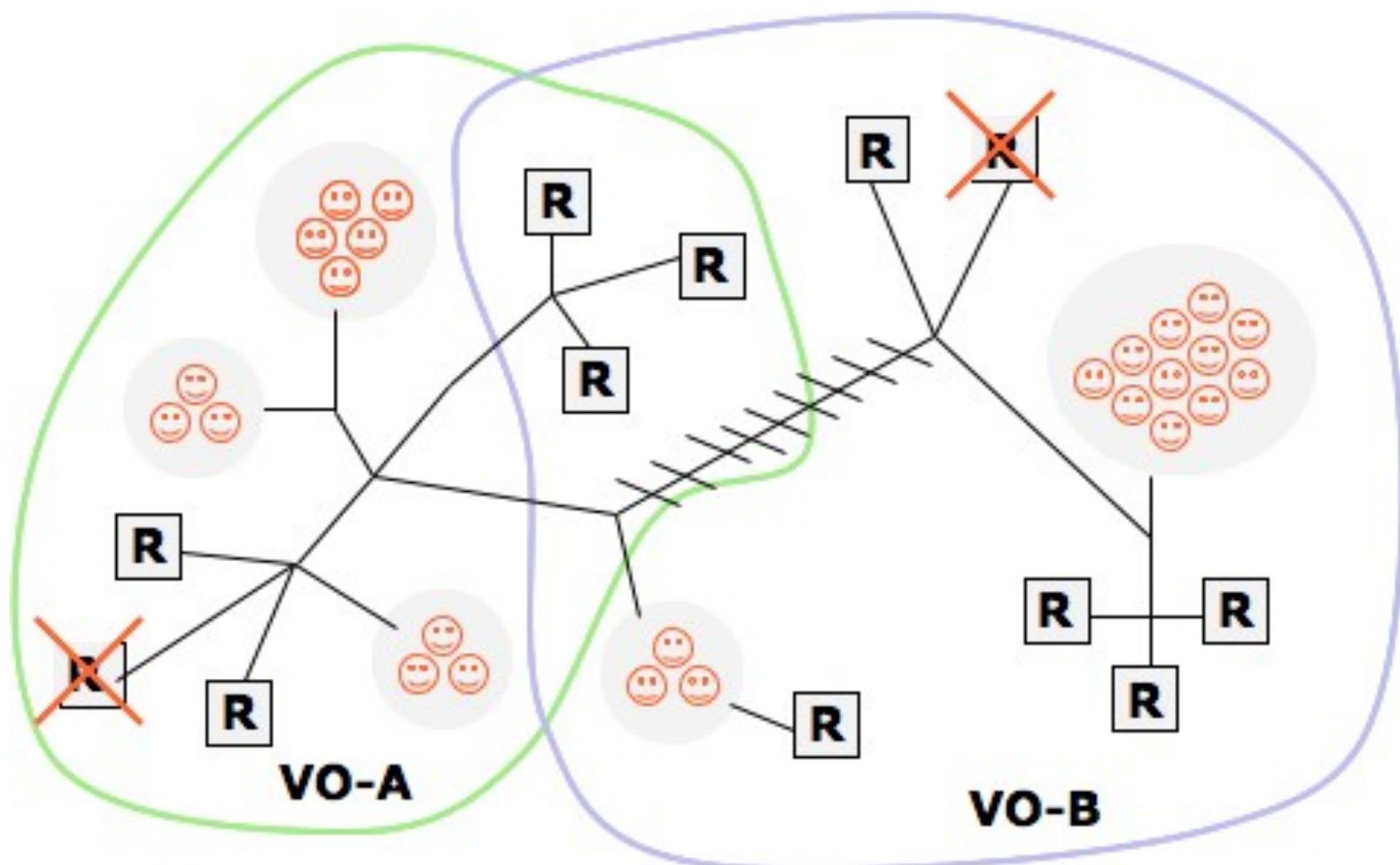
- Key problems and requirements
- The Globus.org approach
- For more information

# Problems & Requirements Topics

- Problem explanation
- Requirements summary
- Real world examples

Problem #1:  
Facilitate cross-administrative domain  
interactions while at the same time  
protecting local autonomy

# Facilitate The Work Of Virtual Organizations



# Support Heterogeneity and Local Control

- Local sites have their own
  - User policies
  - Authorization mechanisms
  - Data privacy policies
  - Hardware
  - Software stacks
  - Service and network configurations
- The sites should be able to share their resources without losing control over them

# Key Requirements

- Globus.org should strive to be compatible with the resource owner's preferred software stack
  - Support existing security mechanisms
  - Avoid imposing new software requirements

# Real World Examples

- Site-specific security requirements
  - Many resource owners, such as OLCF, require the use of a one-time password security token to access their resources
- Site-specific software stack controls
  - Many resource owners, such as GFDL, do not allow end-users to run 3<sup>rd</sup>-party applications on their transfer nodes

Problem #2:  
Most users lack the time and  
inclination to become experts in  
distributed computing technology

# Overview of Reported User Goals

*Perspectives on Distributed Computing User Interviews*



# Key Requirements

- Implement familiar user interfaces
  - Technology interactions should require no special expertise
- Minimize end-user software installation requirements
- Ease the infrastructure providers' support burden

# Real World Examples: Science Goals

- To provide theoretical underpinnings for observations from instruments such as the Hubble space telescope and the James Webb space telescope
- To understand the interactions of quarks and gluons and apply that understanding to the discovery of new, fundamental parameters of elementary particles
- To identify contamination sources in water distribution environments

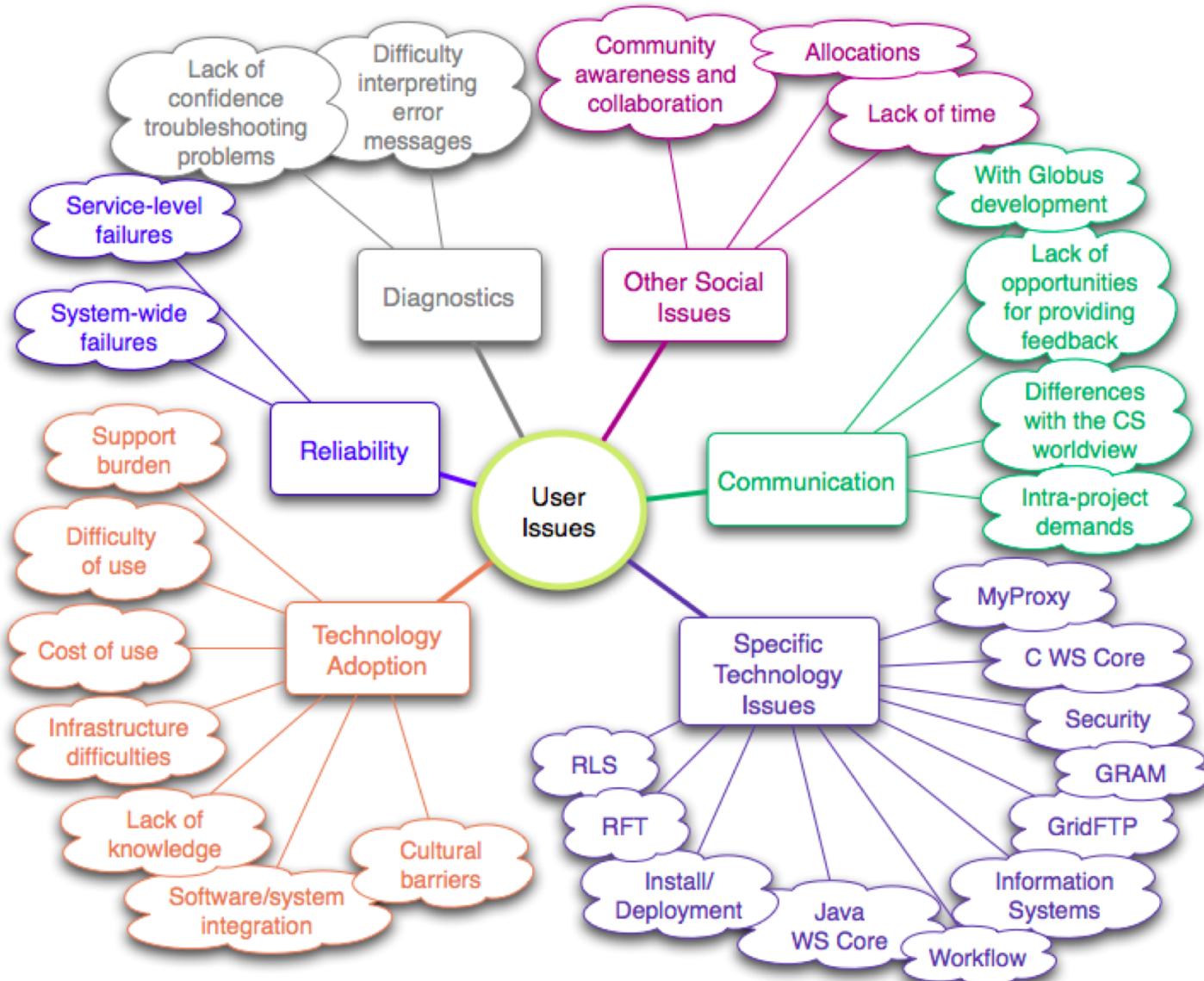
# Real World Examples: Infrastructure Provider Goals

- To provide stable computing facilities for people to do cutting-edge science
- To bring new users to the system
- To add redundancy so the system has no single point of failure
- To share resources that would otherwise not be available to ordinary researchers. They don't have to buy their own equipment; they can get their research done even if they don't have lots of money

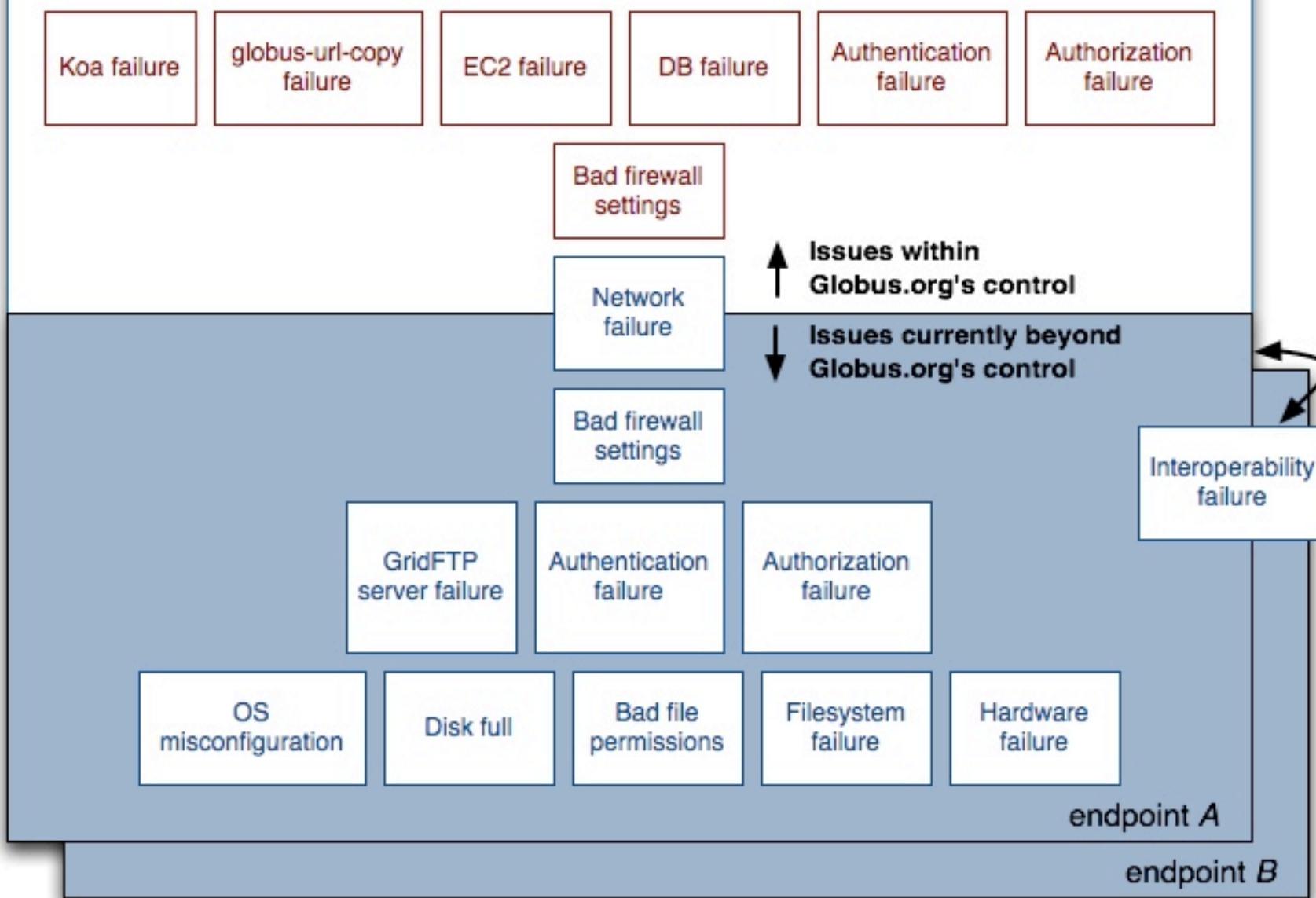
Problem #3:  
Both end-users and infrastructure providers struggle mightily with wide-area technology failures

# Overview of Reported User Issues

Perspectives on Distributed Computing User Interviews



## Potential Sources of File Movement Problems



# Key Requirements

- Manage an increasing number of technology failures on behalf of the user
- Provide users and resource owners with enough information (in words they can understand) to efficiently resolve problems
- Send notifications of interesting events
  - Now: send an email when a transfer completes
  - Some day: Give end-users and resource providers a heads-up about potential problems

# Real World Examples

Errors recorded by Globus.org during the  
3<sup>rd</sup> CEDPS data challenge:

**Unable to open destination file: permission denied**

**Failure to establish a secure connection**

**Stale NFS file handle on the source filesystem**

**Disk quota exceeded**

**Expired host certificate on the destination endpoint**

**SSLv3 handshake problems**

**GSSAPI authentication error**

**The transfer timed out**

# Problem #4: More Data Is Coming



# Anticipated ALCF Bandwidth Requirements \*

- 0-2 years: 10s of TB/day
- 2-5 years: 100s of TB/day
- 5+ years: PBs/day

\* Office of Advanced Scientific Computing Research Network  
Requirements Workshop, April 15-16, 2009

# Near-Term Requirements

- Meet the upcoming CEDPS challenges
  - Just recently met a 100k file, 200MB challenge
  - Start moving 40TB/day of GFDL data next year, eventually supporting 80TB/day

# Real World Examples



# Outline

- Key problems and requirements
- **The Globus.org approach**
- For more information

# The Globus.org Approach

## Topics

- Design overview
- Feature highlights
  - Command-line interface
  - Deadline model
  - Attempts
  - Fair use

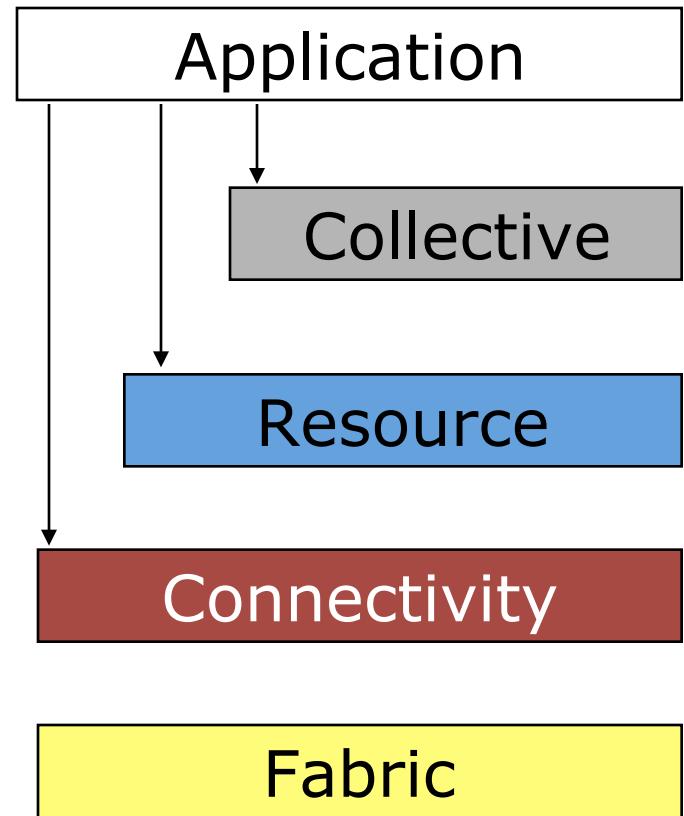
# Grid Architecture

“Coordinating multiple resources”: ubiquitous infrastructure services, app-specific distributed services

“Sharing single resources”: negotiating access, controlling use

“Talking to things”: communication (Internet protocols) and security

“Controlling things locally”: Access to, and control of resources



“The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, Foster, Kesselman, Tuecke, Intl Journal of High Performance Computing Applications, 15(3), 2001.

# Distributed Data Nodes At The Resource Layer

GridFTP

GridFTP

GridFTP

GridFTP

GridFTP

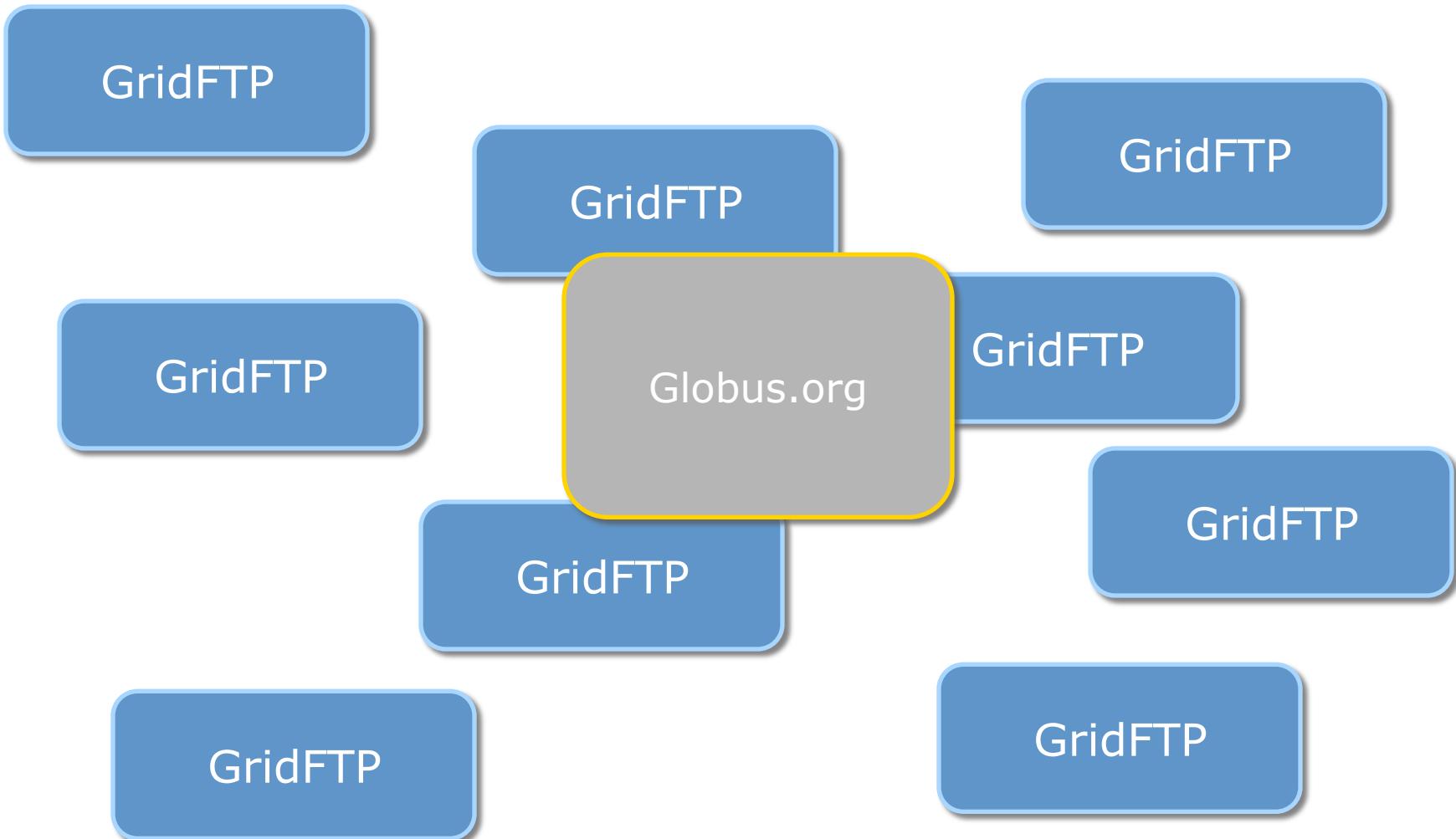
GridFTP

GridFTP

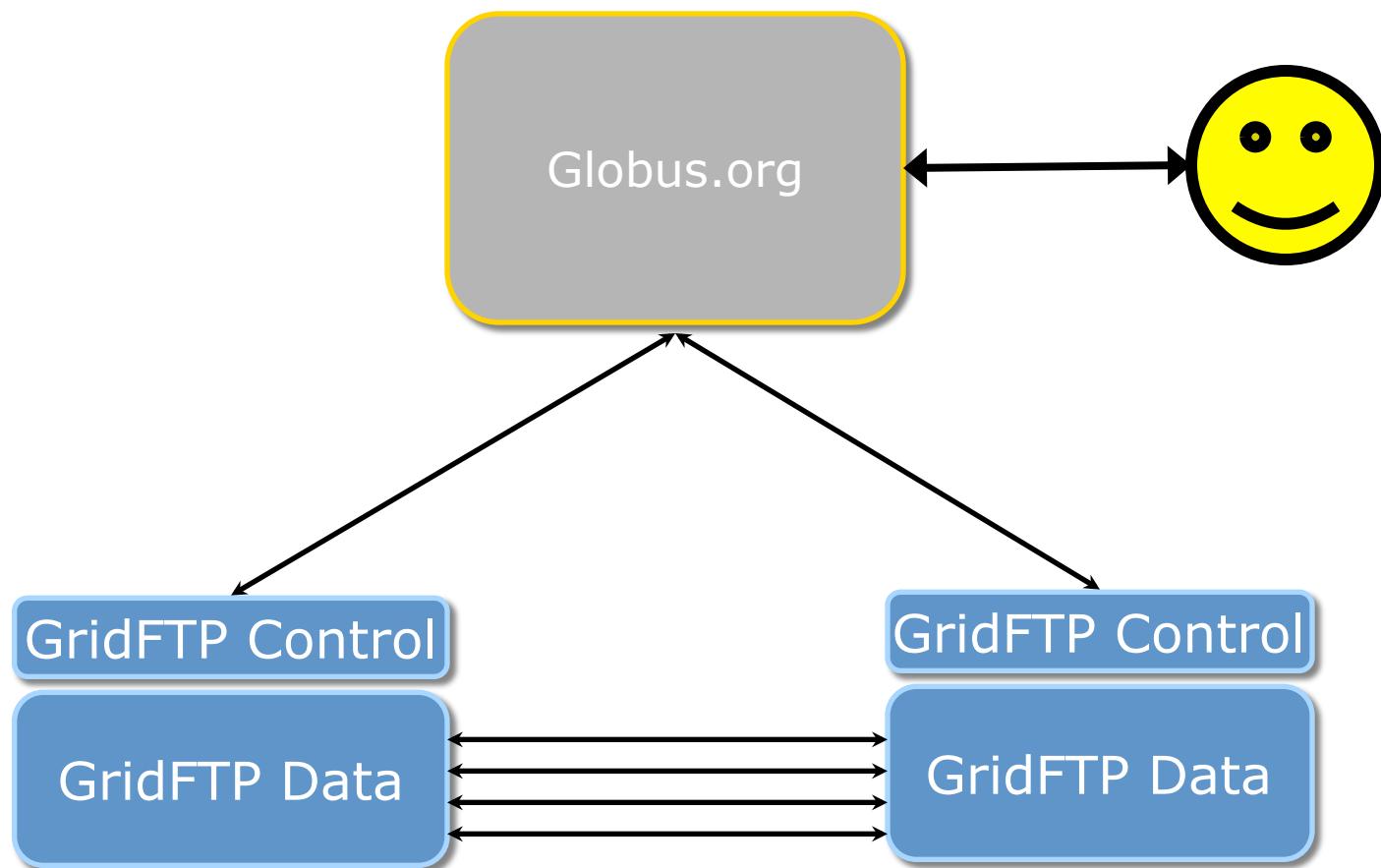
GridFTP

GridFTP

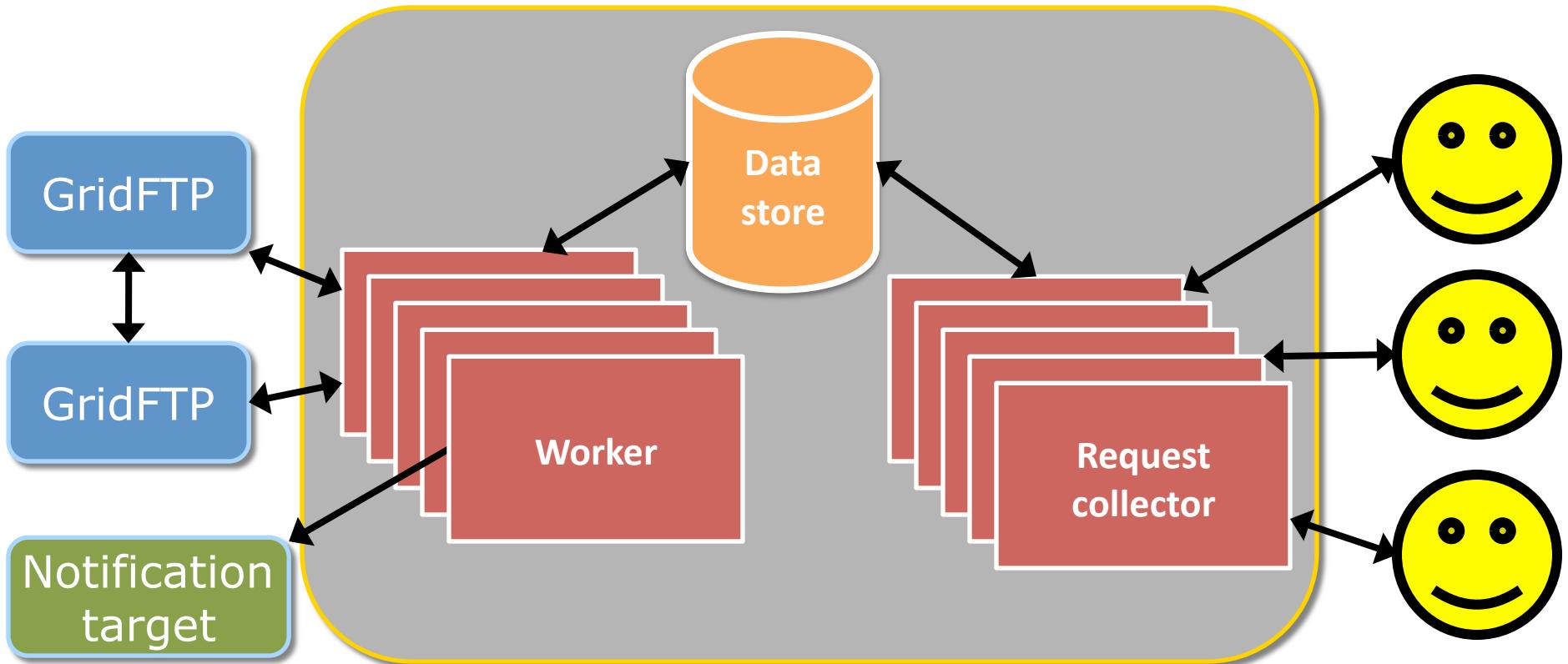
# Globus.org Operates At The Collective Layer



# Globus.org Manages 3<sup>rd</sup>-Party Transfers



# A Peek Inside Globus.org



# Command-Line Interface

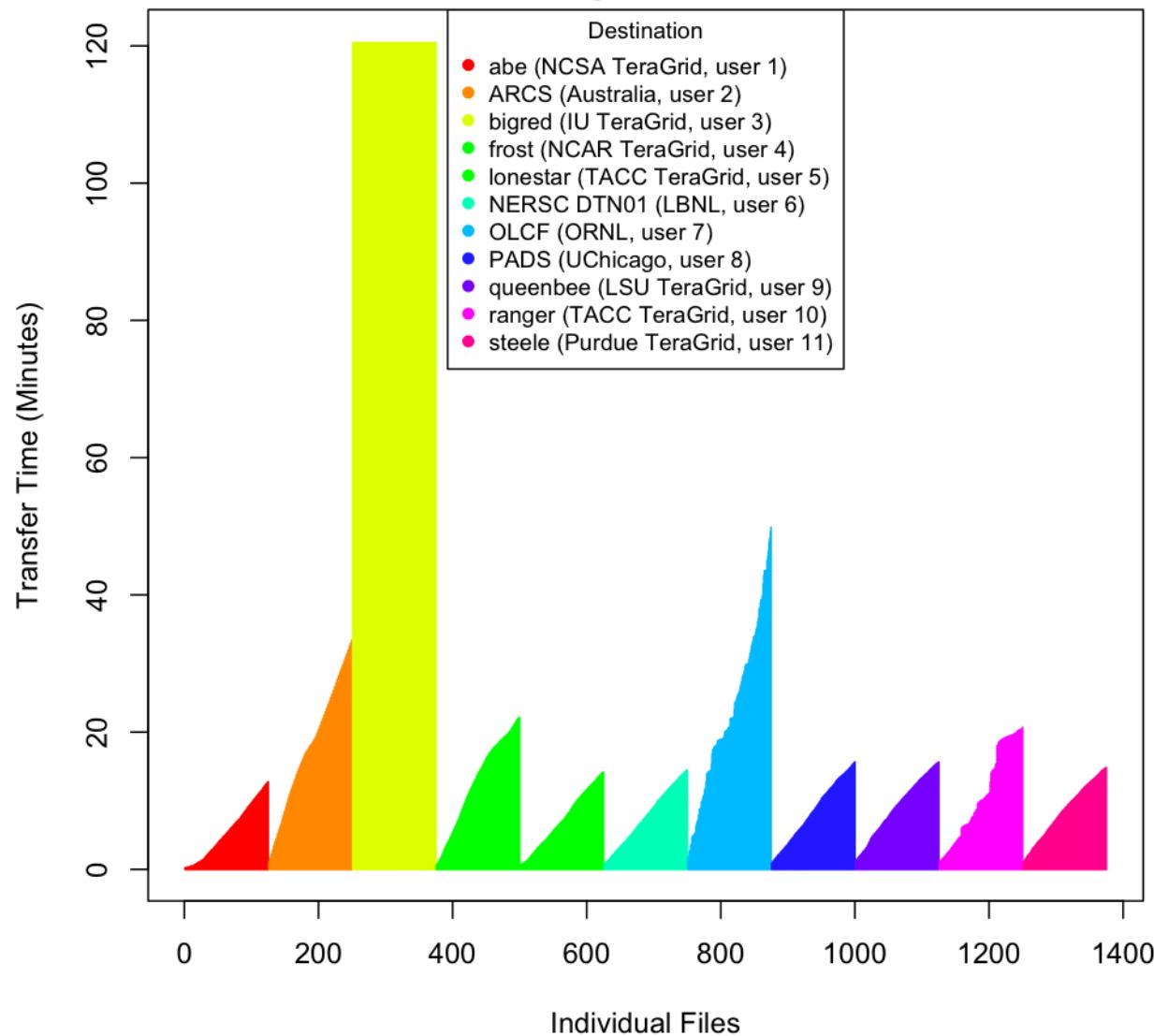
Welcome to globus.org. You may run any of the following commands:

passwd  
manage-endpoint  
list-endpoints  
activate  
xfer  
cancel  
summary  
status  
event-logs  
help

### CEDPS Data Challenge #3: Time (ordered by Destination, Time)

11 users each transferring 125 200M files from ALCF

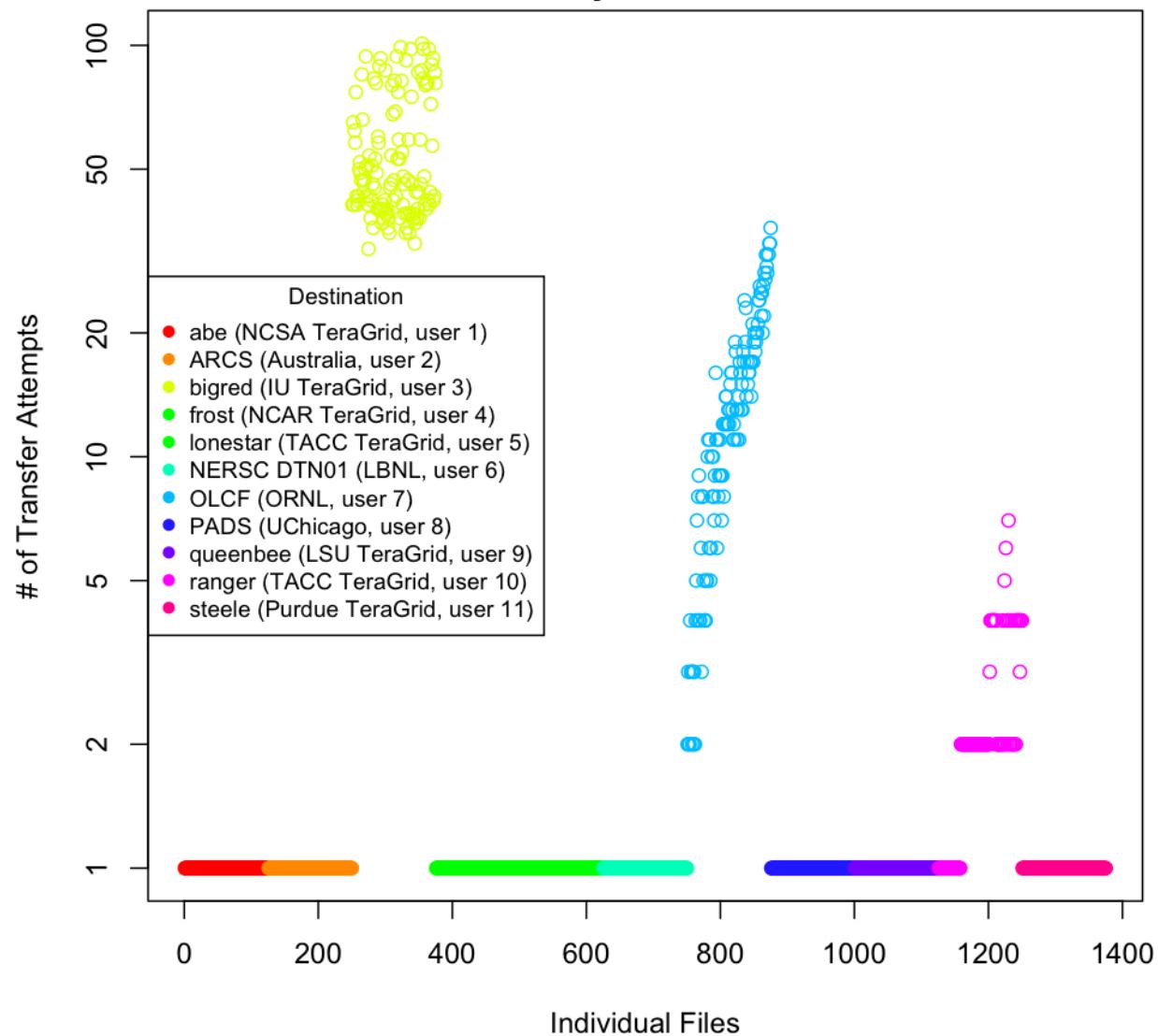
May 11 2010



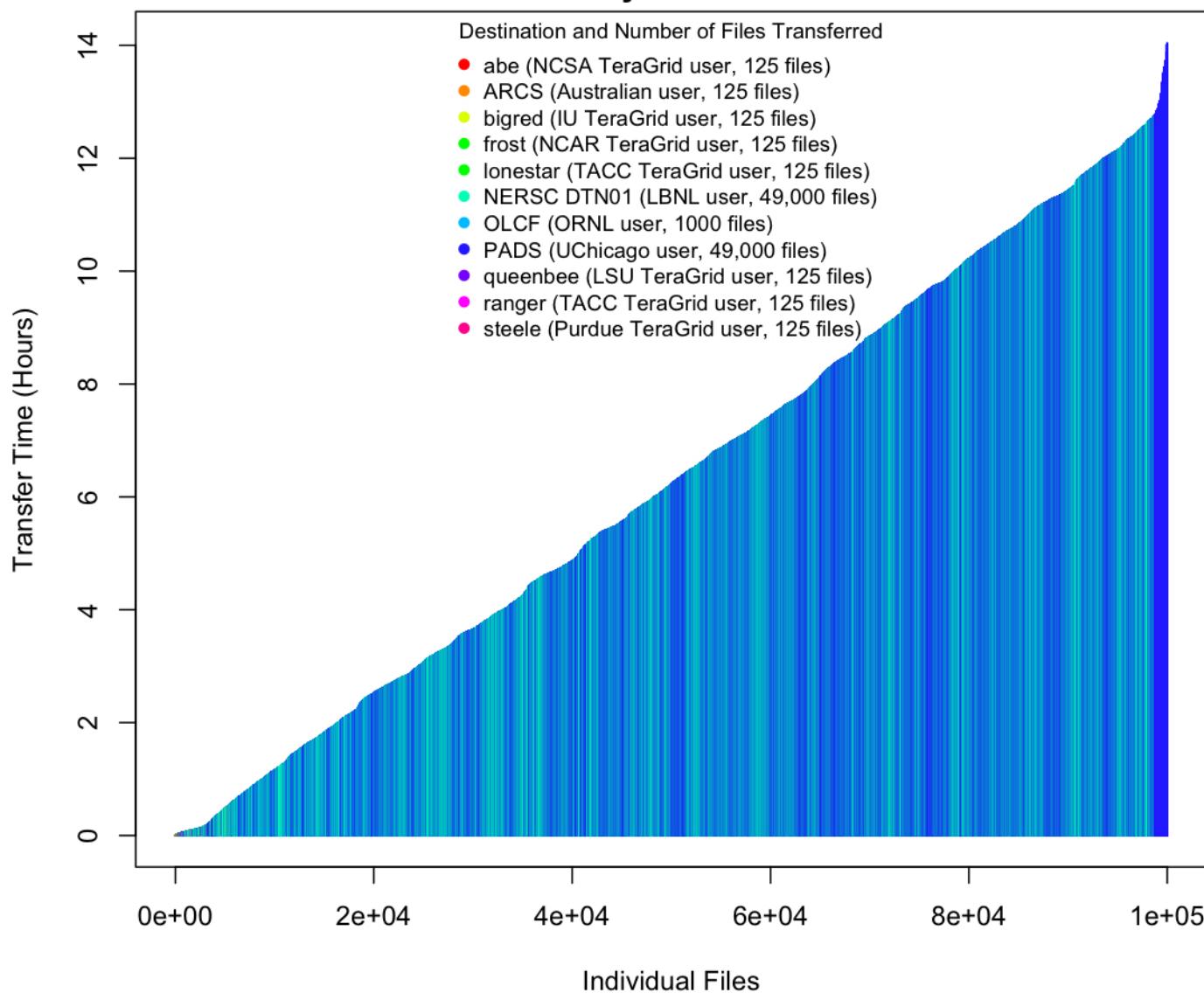
### CEDPS Data Challenge #3: Attempts (ordered by Destination, Time)

11 users each transferring 125 200M files from ALCF

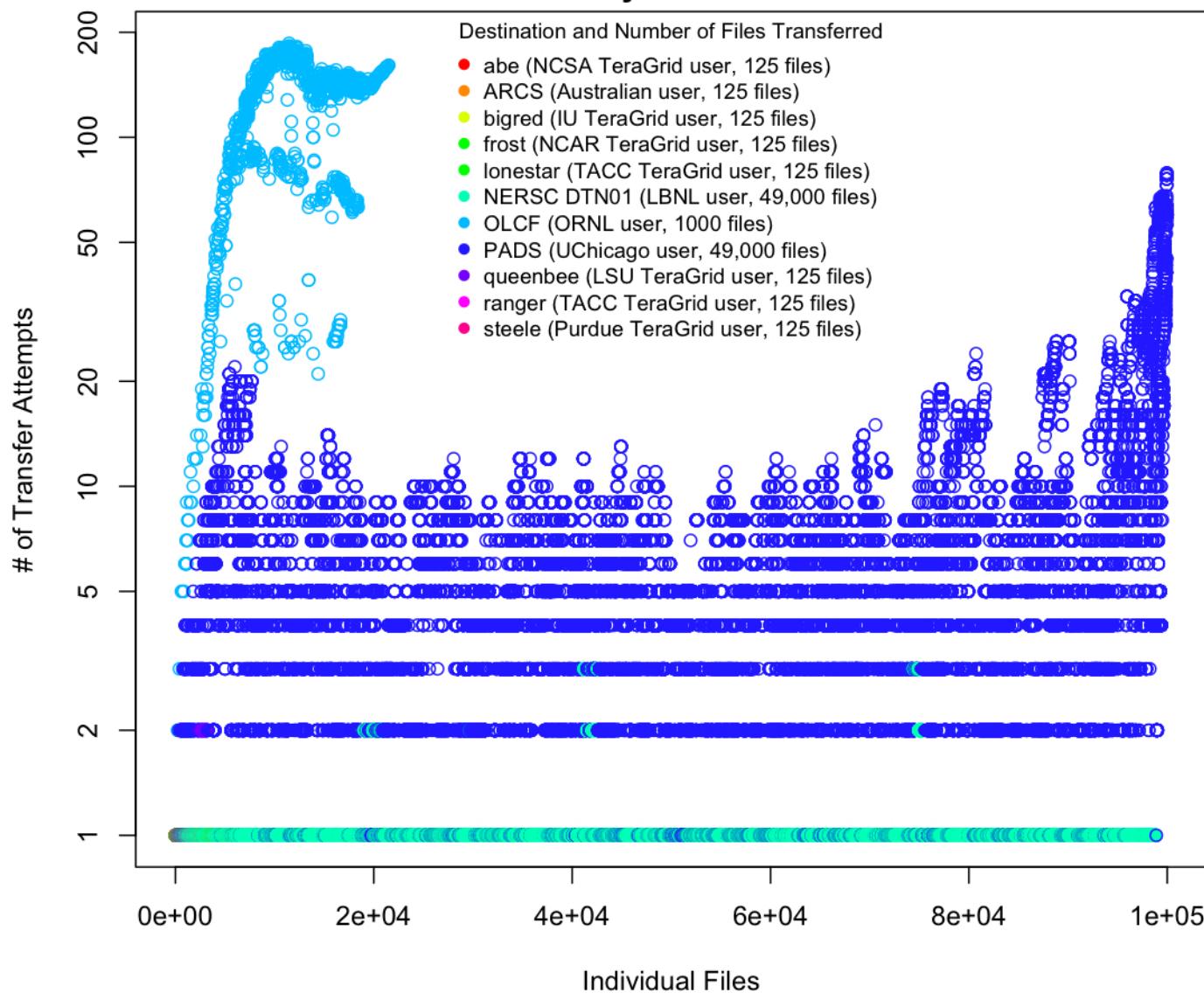
May 11 2010



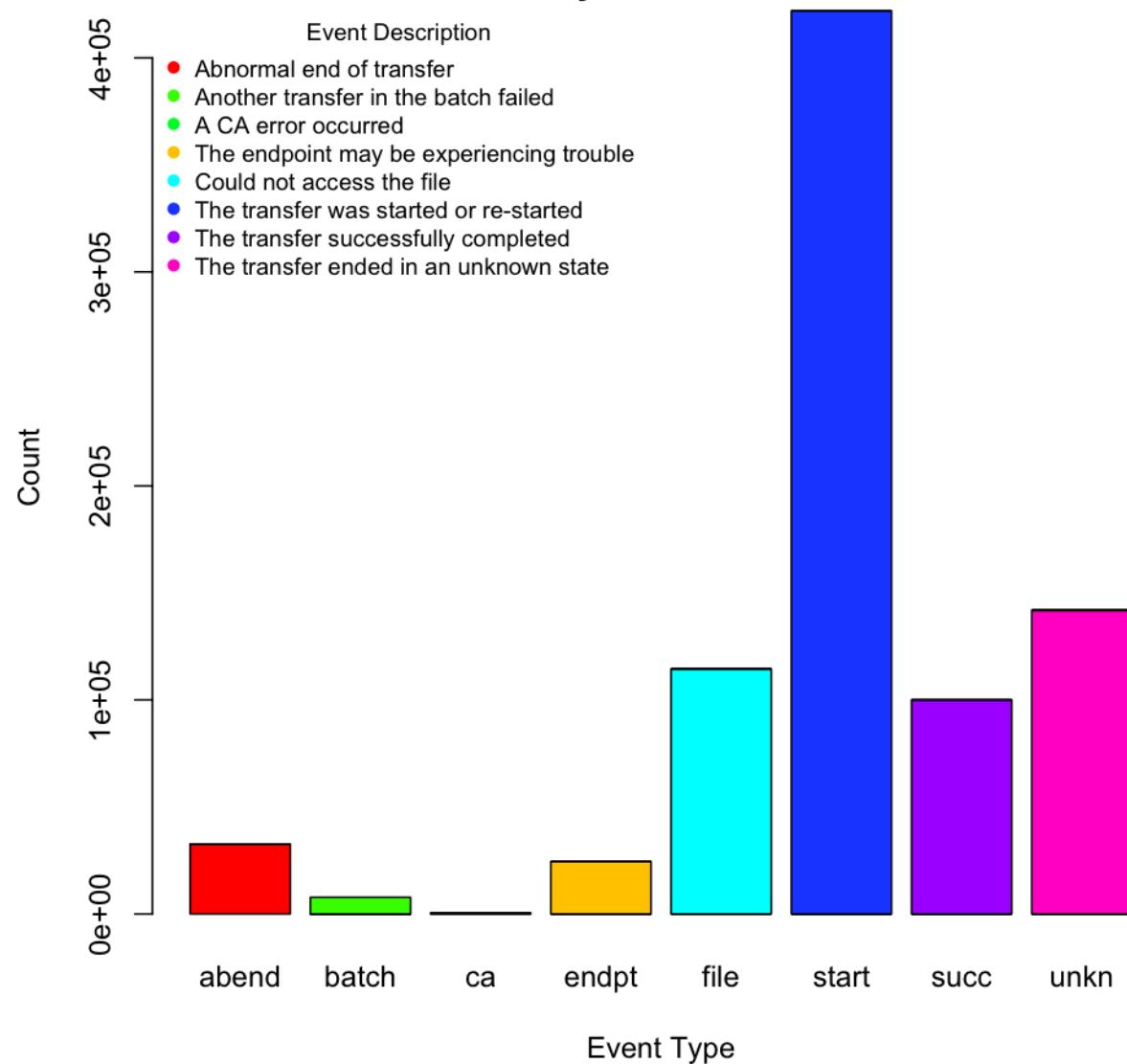
**CEDPS Data Challenge #3: Time (ordered by Transfer Time, Destination)**  
**11 users transferring a total of 100,000 200MB files from ALCF**  
**May 28 2010**



**CEDPS Data Challenge #3: Attempts (ordered by Transfer Time, Destination)**  
**11 users transferring a total of 100,000 200MB files from ALCF**  
**May 28 2010**



**CEDPS Data Challenge #3: Event Summary**  
**11 users transferring 100,000 200MB files from ALCF**  
**28 May 2010, Run #12**



### CEDPS Data Challenge #3: Throughput Comparison

#### Average Throughput Experienced by 125-File Users

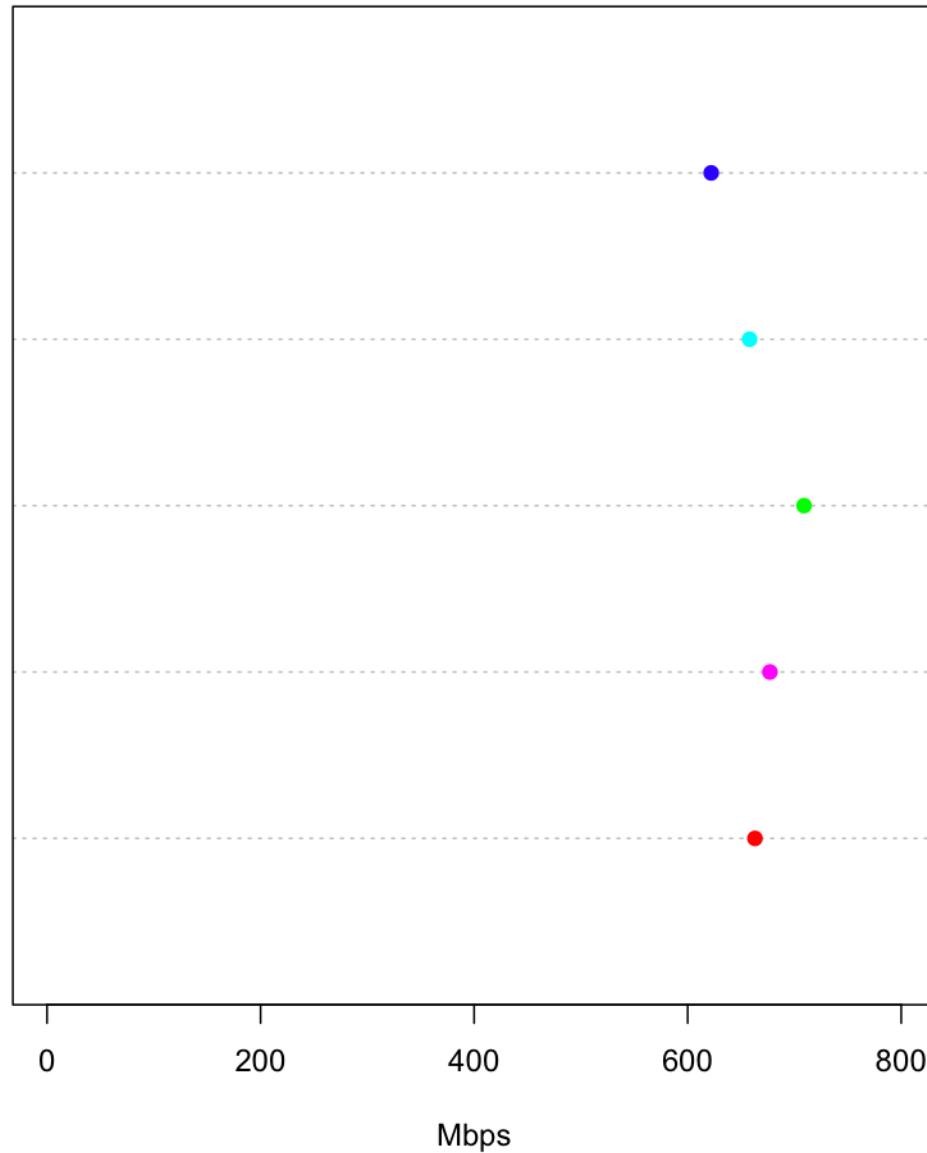
Run 8: 1,375 total transfer requests

Run 9: 1,375 total transfer requests

Run 10: 1,375 total transfer requests

Run 11: 100,000 total transfer requests

Run 12: 100,000 total transfer requests



# What is Globus.org?

- The latest iteration of Globus software
  - The same Globus vision, but an updated approach
- Hosted services
  - Data movement initially
  - Execution , information, and VO services to follow
- The Globus Toolkit isn't going away
  - Contains tools and services for resource owners
  - Compatible with Globus.org hosted services

# Key Goals

- Provide scientists with easy access to advanced computing resources
  - Familiar user interfaces
  - Technology interactions requiring no special expertise
  - No software to install
  - Support for well-known community and international resources
  - Ability to customize working environment
- Enable users to focus on domain-specific work
  - Manage technology failures
  - Notifications of interesting events
  - Provide users with enough information to resolve problems

# For More Information

- My email address
  - [childers@mcs.anl.gov](mailto:childers@mcs.anl.gov)
- Globus.org quickstart guide
  - [www.mcs.anl.gov/~childers/quickstart/](http://www.mcs.anl.gov/~childers/quickstart/)
- Globus.org marketing overview
  - [www.globus.org/service/](http://www.globus.org/service/)
- Hub-and-spoke data movement study
  - [www.mcs.anl.gov/~childers/CDC3/](http://www.mcs.anl.gov/~childers/CDC3/)
- User study
  - [www.mcs.anl.gov/~childers/perspectives/](http://www.mcs.anl.gov/~childers/perspectives/)

Thank You