

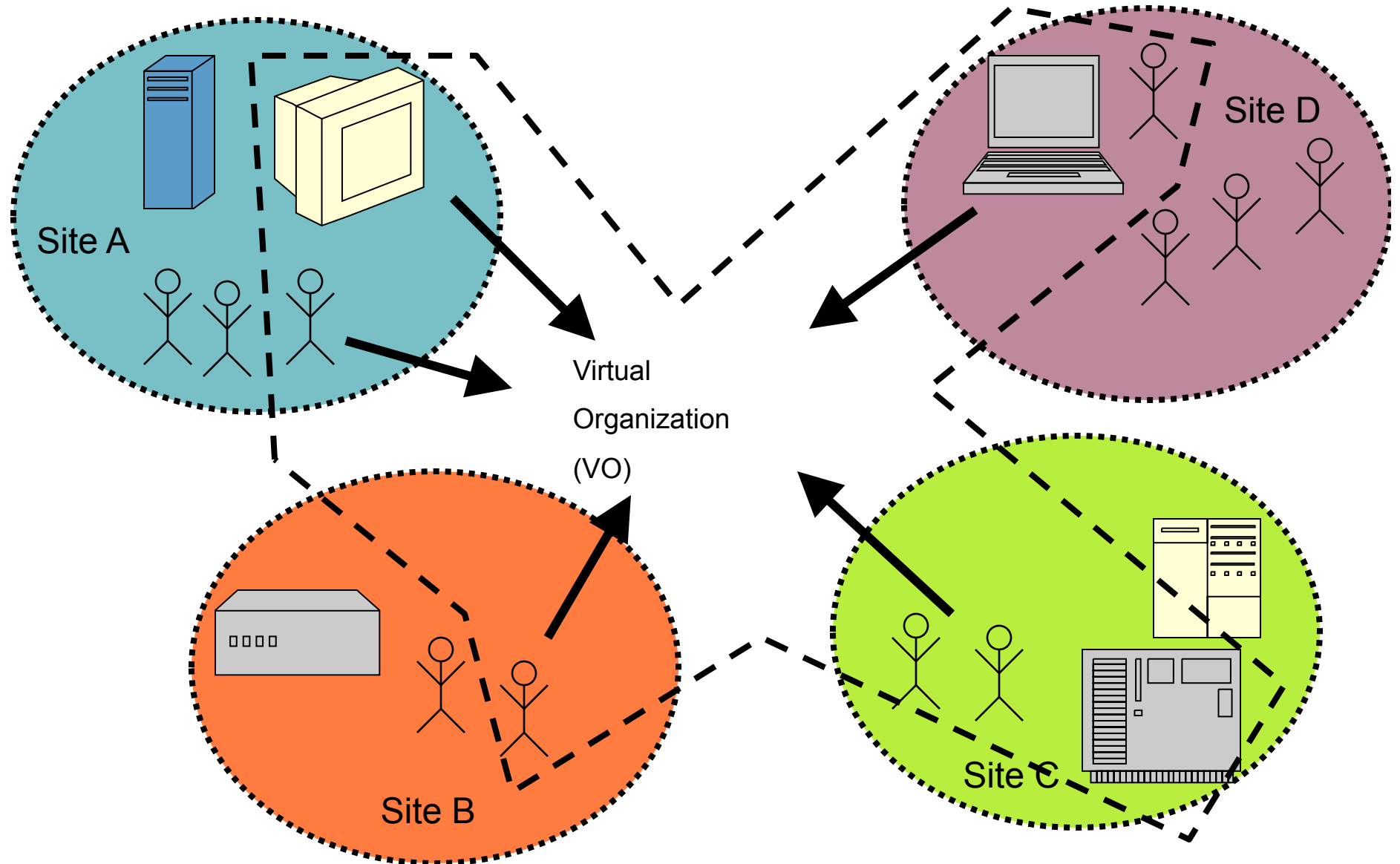


Grid-Cloud Intersection: Hosted Service for High- Speed, Secure and Reliable Data Movement

Raj Kettimuthu
Argonne National Laboratory and
The University of Chicago



Grid





Grid and other systems

Grid Vs. Web

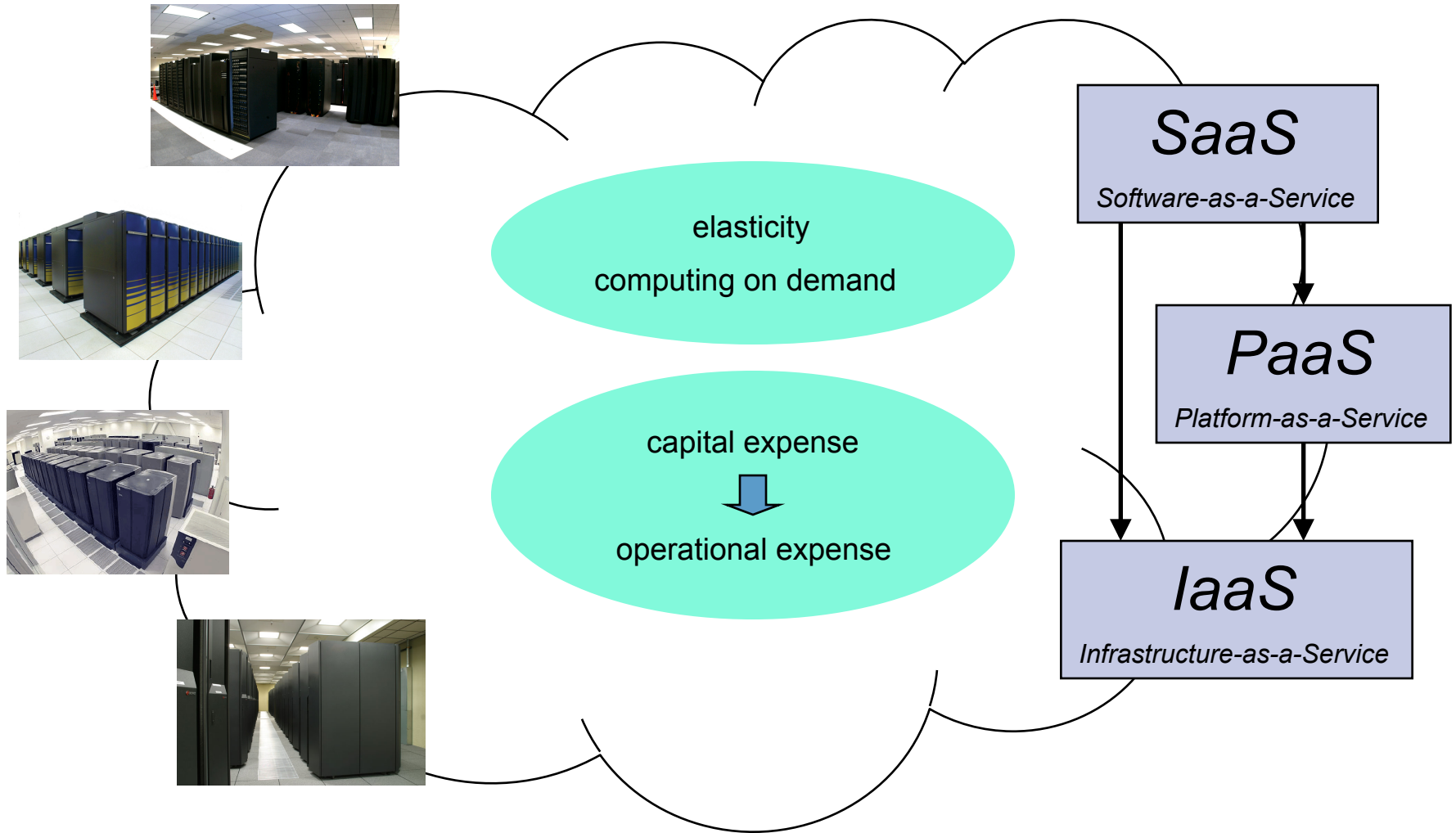
- Like the web Grid keeps complexity hidden: multiple users enjoy a single, unified experience.
- Unlike the Web which mainly enables communication, grid computing enables full collaboration toward common business or scientific goals.

Grid Vs. P2P

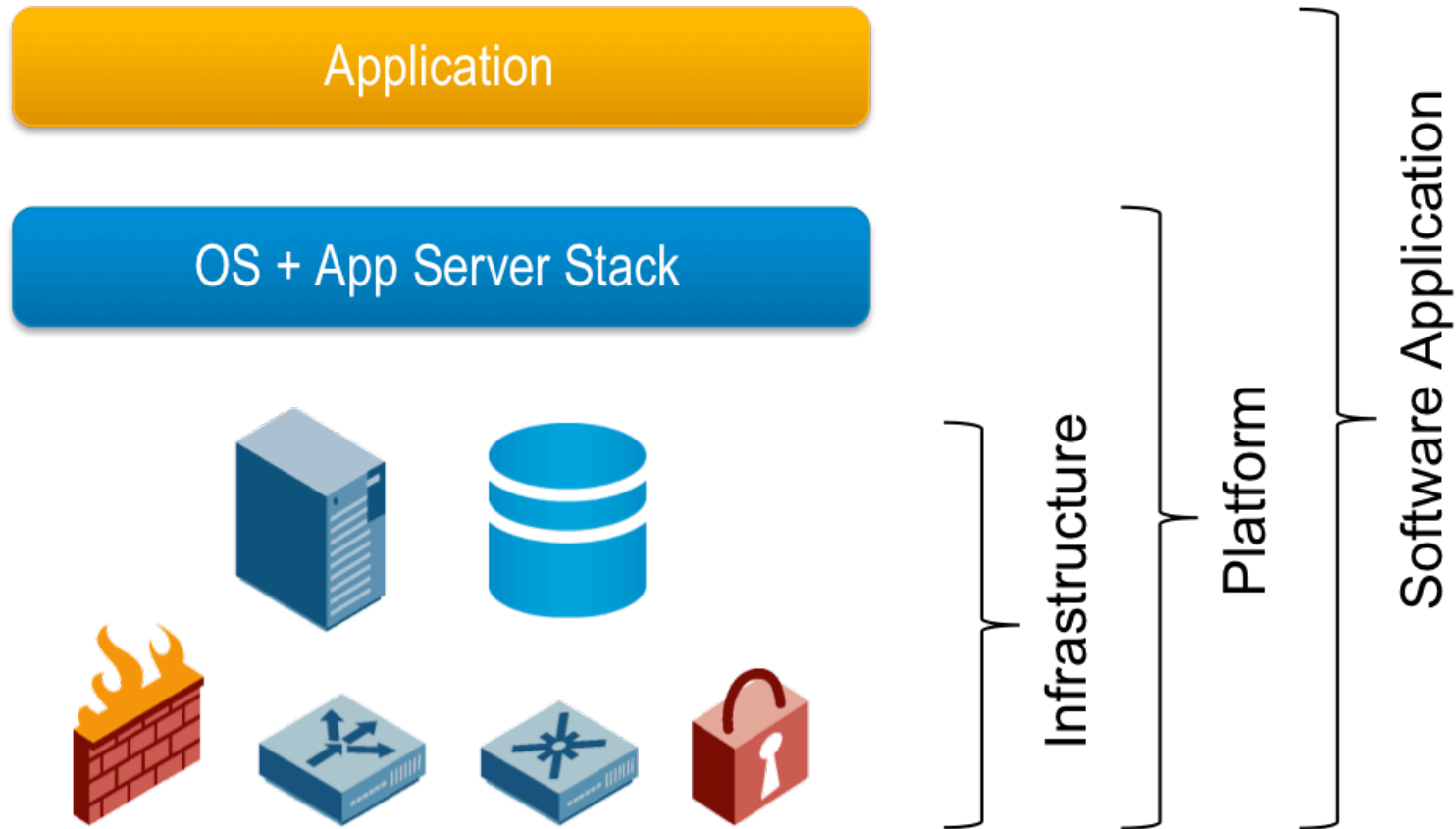
- Like peer-to-peer grid computing allows users to share files.
- Unlike peer-to-peer grid computing allows many-to-many sharing not only files but other resources as well.



Cloud Computing



Cloud Computing





Cloud versus Grid

- Eliminates costs and complexity of buying, configuring, and managing H/W and S/W needed to build and deploy applications
- H/W and S/W resources and applications are delivered as a service over the Internet
- Benefits small and medium-sized business
- Grid's vision was allow consumers to obtain computing power on demand
- Grid systems are designed for collaborative sharing of resources
- Only big science projects can afford to create and operate dedicated grid infrastructures

Grid = federation

Cloud = hosting



Time consuming tasks in Science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature
- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...



the globus alliance
www.globus.org

Grid



03/30/2011

KCT, India



Cloud

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)



Cloud

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)



Cloud

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)



the globus alliance

www.globus.org

Time-consuming tasks in business

Web presence

Email (hosted Exchange)

Calendar

Telephony (hosted VOIP)

Human resources and payroll

Accounting

Customer relationship mgmt

Data analytics

Content distribution

...

SaaS

IaaS



Cloud

Software-as-a-Service (SaaS)

Platform-as-a-Service (PaaS)

Infrastructure-as-a-Service (IaaS)



Time consuming tasks in Science

- Run experiments
- Collect data
- Manage data
- Move data
- Acquire computers
- Analyze data
- Run simulations
- Compare experiment with simulation
- Search the literature
- Communicate with colleagues
- Publish papers
- Find, configure, install relevant software
- Find, access, analyze relevant data
- Order supplies
- Write proposals
- Write reports
- ...



the globus alliance

www.globus.org

Data Movement

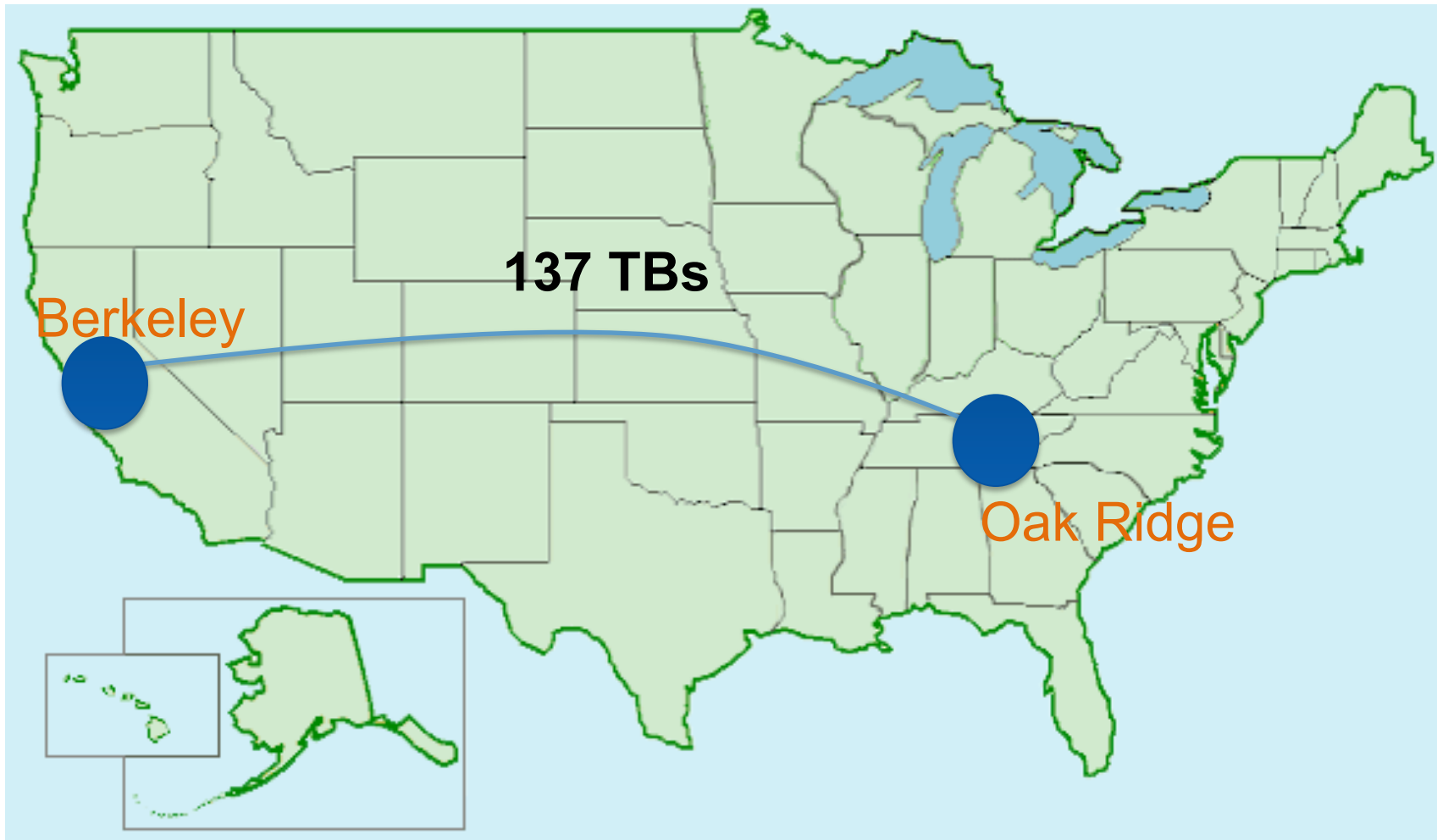
- Deceptively simple
- Stick it in email
- Too large
- 100,000 files totaling 10 Terabytes
- Move from a federal laboratory where they were generated to my home institution
- Can be very difficult





the globus alliance
www.globus.org

Data Movement Challenges



GridFTP

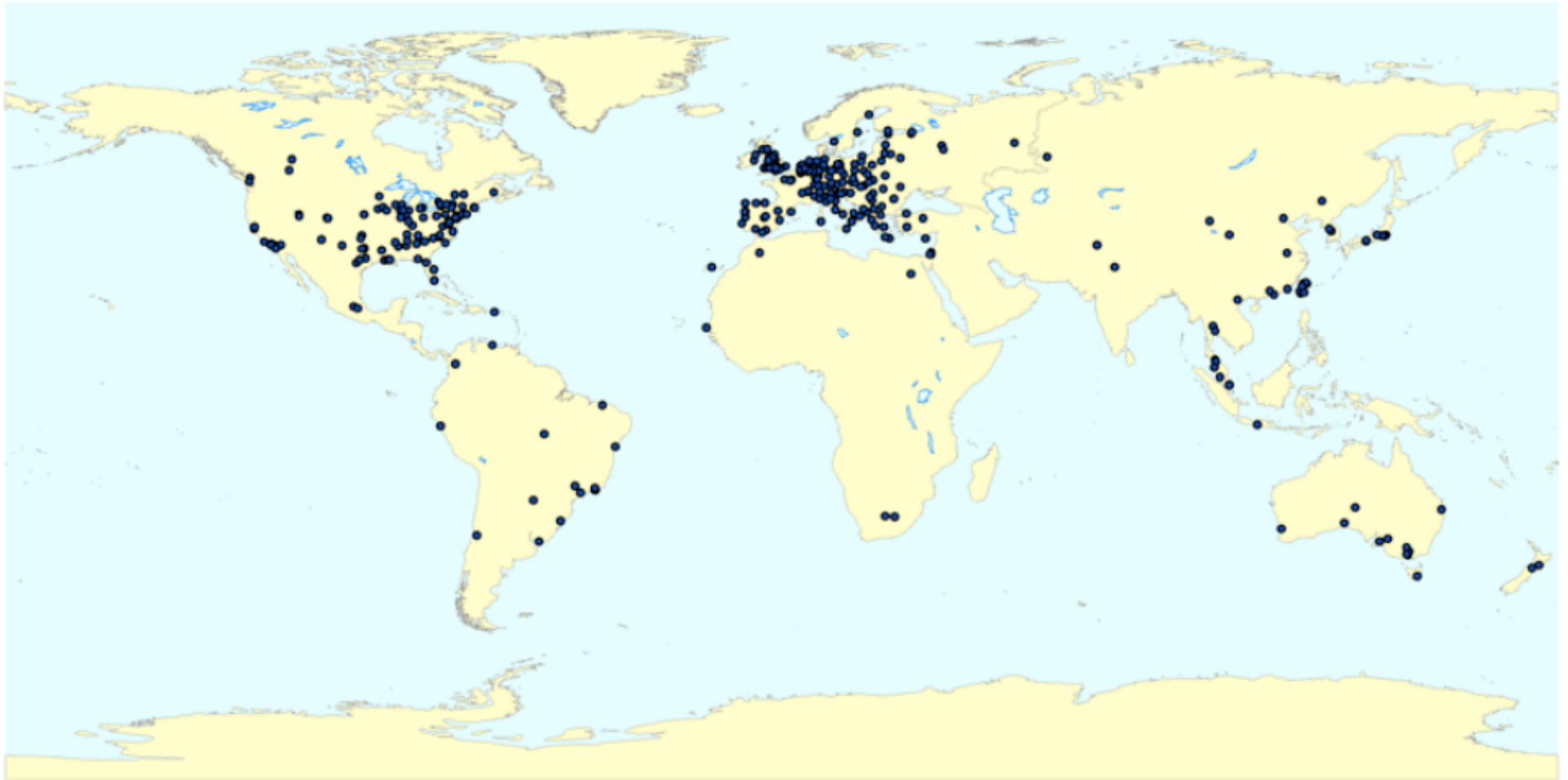
- High-performance, reliable data transfer protocol optimized for high-bandwidth wide-area networks
- Globus GridFTP
 - ◆ Parallel TCP streams, optimal TCP buffer
 - ◆ Non TCP protocol such as UDT
 - ◆ Cluster-to-cluster data movement
 - ◆ SSH, GSI
 - ◆ Restartable transfers



the globus alliance

www.globus.org

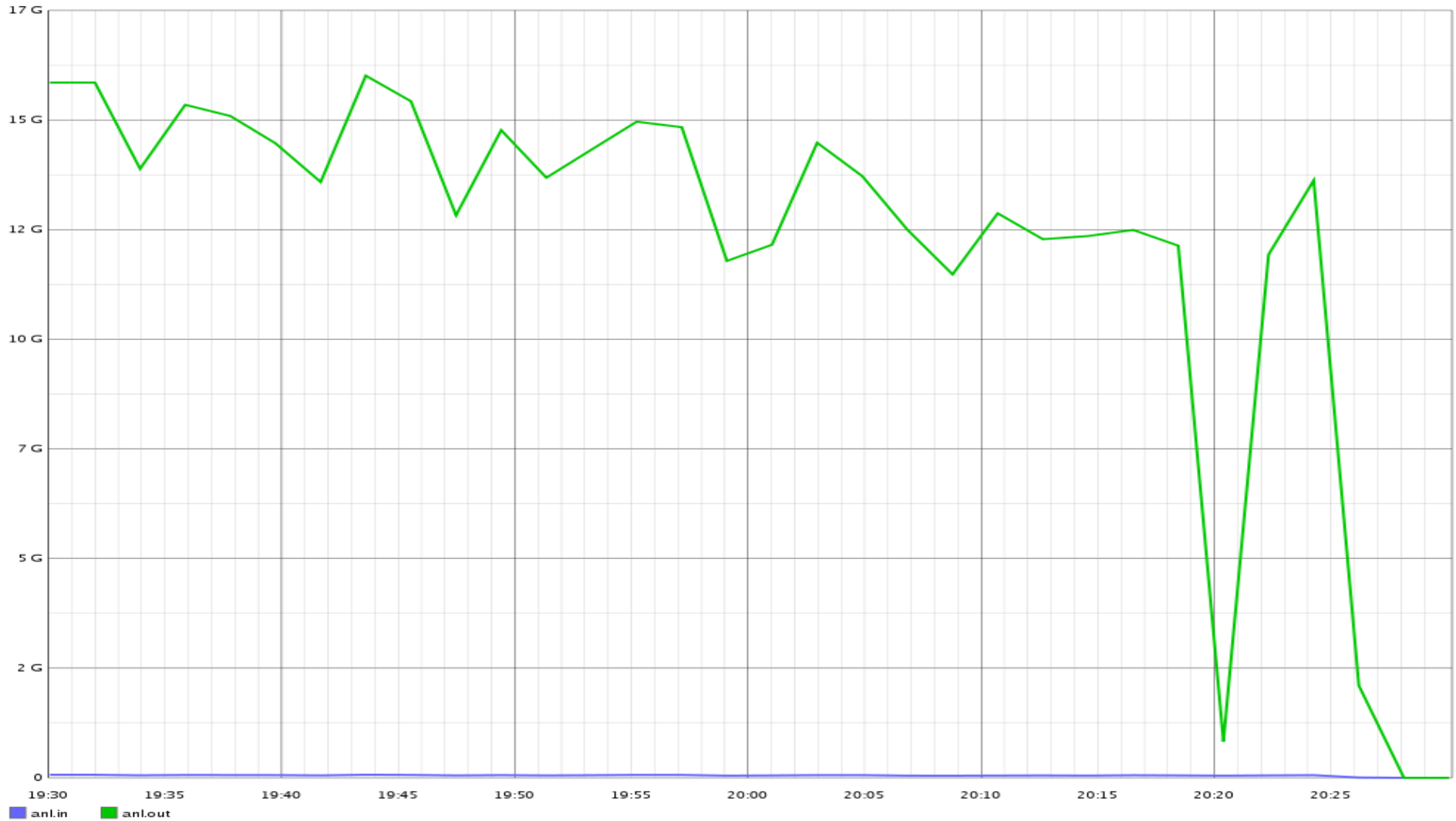
GridFTP Servers Around the World



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



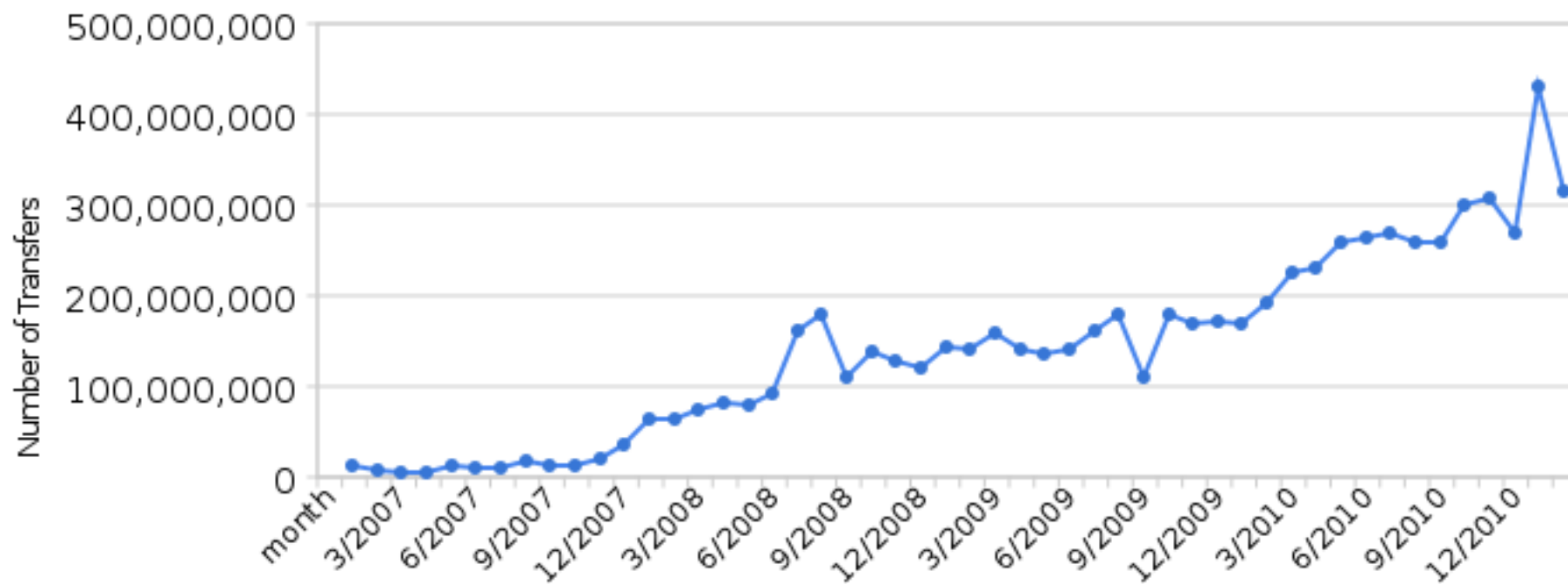
Performance





GridFTP Usage

Monthly Totals* of GridFTP File Transfers



*for those "reporting"



the globus alliance

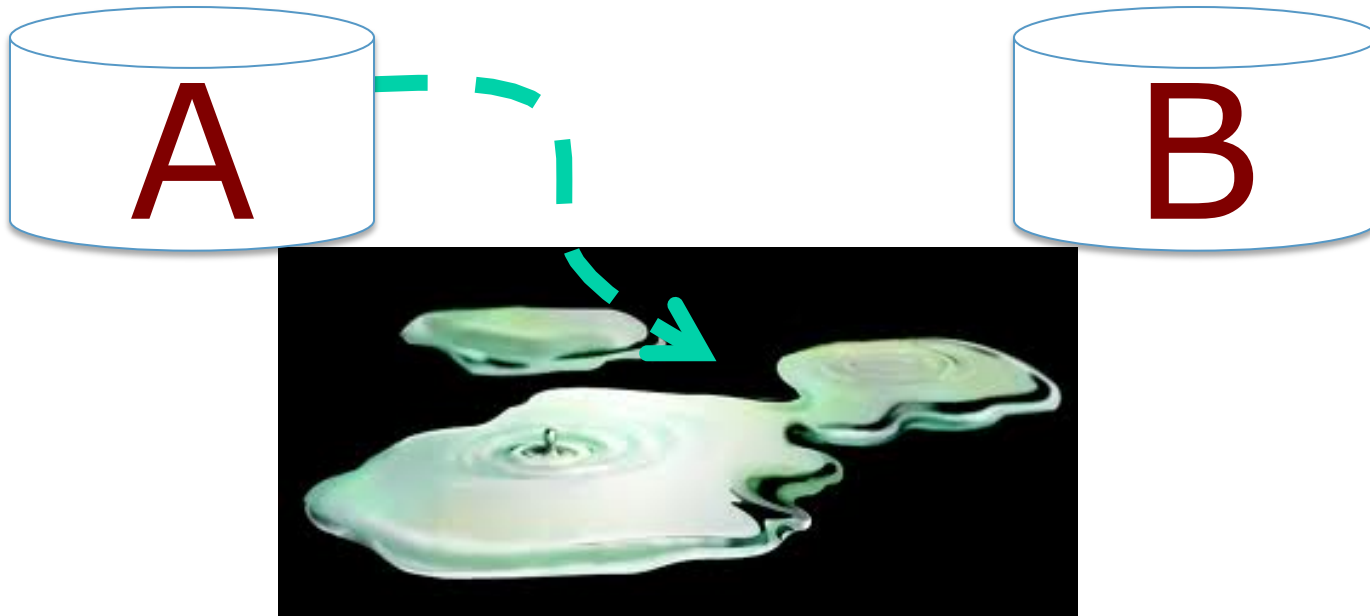
www.globus.org

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, ...





the globus alliance
www.globus.org

Globus Online

- *Globus Toolkit*

Build the Grid



Components for building
custom grid solutions

globustoolkit.org

- *Globus Online*

Use the Grid

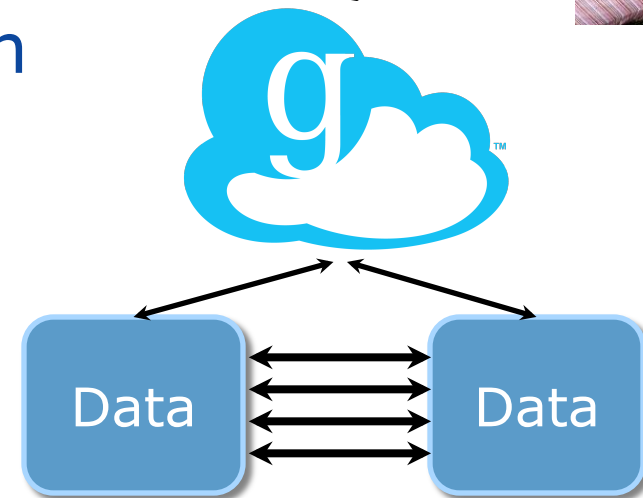
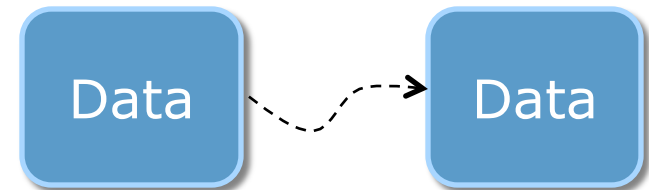
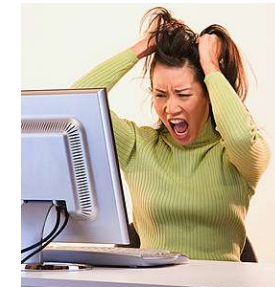


Cloud-hosted
file transfer service

globusonline.org



- 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



Login to globusonline.org

Sign-in with Google OpenID

Sign-in with MyProxy

Create a profile

Not yet a member? [Join today.](#)

Join

[Learn more about Globus Online](#)

Basic Profile

- [Basic information](#)
- [Login Accounts](#)
- [Command-line \(CLI\)](#)

Use this section of Edit Profile to specify your name and preferred email address for notification from Globus Online.

Full Name

Username

Email address

Confirm email

Globus Online

- [Overview](#)
- [Sponsors](#)
- [Contact Us](#)
- [Privacy Policy](#)
- [Terms of Use](#)

Support

- [Documentation](#)
- [Video Tutorials](#)
- [Quickstart](#)
- [Using the CLI](#)
- [Beyond Basics](#)

Community

- [News and Events](#)
- [Globus Toolkit](#)
- [Computation Institute](#)
- [University of Chicago](#)
- [Argonne](#)

Initiate Transfers

Transfers In Progress: 0

[View All Transfers](#)

Endpoint ▾

[Go](#)

Directory

[Go](#)

[Up Directory](#)

[Refresh Directory](#)

[Select All](#)

[Clear Selections](#)

Mode:

Transfer ▾

Endpoint ▾

[Go](#)

Directory

[Go](#)

[Up Directory](#)

[Refresh Directory](#)

[Select All](#)

[Clear Selections](#)



Globus Online

[Overview](#)
[Sponsors](#)

Support

[Documentation](#)
[Video Tutorials](#)

Community

[News and Events](#)
[Globus Toolkit](#)

Initiate Transfers

Transfers In Progress: 0

[View All Transfers](#)

Endpoint

Directory

Up Dir

- rajk#alcf
- rajk#alcf-personal
- rajk#kraken
- rajk#rajlaptop
- rajk#stomp
- rajk#tg-queenbee
- rajk#tg-steele
- alcf#dtn
- ci#pads
- go#ep1

[Clear Selections](#)

Mode: Transfer ▾



Endpoint

Directory

Up Directory

[Clear Selections](#)

Globus Online

- [Overview](#)
- [Sponsors](#)

Support

- [Documentation](#)
- [Video Tutorials](#)

Community

- [News and Events](#)
- [Globus Toolkit](#)



the globus alliance

www.globus.org

raj#alcf

raj#alcf-personal

raj#kraken

raj#rajlaptop

raj#stomp

raj#tg-queenbee

raj#tg-steele

alcf#dtn

ci#pads

go#ep1

Monitor Transfers

Remove Data Filter

Cancel

◀◀ 2 of 2 ▶▶

10 ▾

	Status	ID	Task Progress	Username	Completion Time	Request Time
<input type="checkbox"/>	❌	b0460...	0 / 1 / 1	rajk	12/01/2010 10:55 PM	11/30/2010 10:54 PM
<input type="checkbox"/>	❌	ca555...	0 / 1 / 1	rajk	12/01/2010 07:48 PM	11/30/2010 07:47 PM
<input type="checkbox"/>	✅	54282...	1 / 1	rajk	11/18/2010 05:36 PM	11/18/2010 05:36 PM
<input type="checkbox"/>	✅	95a4f...	1 / 1	rajk	11/17/2010 10:11 PM	11/17/2010 10:11 PM
<input type="checkbox"/>	✅	72382...	1 / 1	rajk	11/16/2010 04:54 AM	11/16/2010 04:53 AM
<input type="checkbox"/>	❌	27b4b...	0 / 1 / 1	rajk	10/12/2010 05:39 PM	10/12/2010 05:35 PM
<input type="checkbox"/>	❌	016b6...	0 / 1 / 1	rajk	10/12/2010 05:29 PM	10/12/2010 05:27 PM
<input type="checkbox"/>	❌	9e166...	0 / 1 / 1	rajk	10/01/2010 06:02 PM	10/01/2010 05:59 PM
<input type="checkbox"/>	❌	c878c...	31256 / 93 / 31349	rajk	10/03/2010 11:18 PM	10/01/2010 04:34 PM
<input type="checkbox"/>	❌	c121f...	0 / 1 / 1	rajk	09/29/2010 10:21 PM	09/29/2010 10:13 PM
<input type="checkbox"/>	✅	56914...	1 / 1	rajk	08/02/2010 07:59 PM	08/02/2010 07:58 PM
<input type="checkbox"/>	✅	eb3c1...	20 / 20	rajk	07/29/2010 04:56 PM	07/29/2010 04:54 PM
<input type="checkbox"/>	✅	ba769...	25 / 25	rajk	07/21/2010 07:09 PM	07/21/2010 06:56 PM
<input type="checkbox"/>	❌	8df71...	0 / 40 / 40	rajk	07/21/2010 06:28 PM	07/21/2010 04:18 PM

Cancel

◀◀ 2 of 2 ▶▶

10 ▾

Globus Online

- [Overview](#)
- [Sponsors](#)
- [Contact Us](#)
- [Privacy Policy](#)
- [Terms of Use](#)

Support

- [Documentation](#)
- [Video Tutorials](#)
- [Quickstart](#)
- [Using the CLI](#)
- [Beyond Basics](#)

Community

- [News and Events](#)
- [Globus Toolkit](#)
- [Computation Institute](#)
- [University of Chicago](#)
- [Argonne](#)

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>
```


Command

Endpoints

`scp go#ep1:/share/godata/file1.txt 'go#ep2:~/myfile.txt'`

```
ssh lcc@cli.globusonline.org scp go#ep1:/share/godata/file1.txt  
go#ep2:~/myfile.txt
```

```
Contacting 'myproxy.tutorial.globusonline.org'...
```

```
Activating 'ep2'
```

```
Activating 'ep1'
```

```
Task ID: 19029d64-ecec-11df-aa30-1231350018b1
```

```
[XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX] 1/1 0.00  
mbps
```

```
ssh lcc@cli.globusonline.org ls go#ep2/~/  
myfile.txt
```



the globus alliance

www.globus.org

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

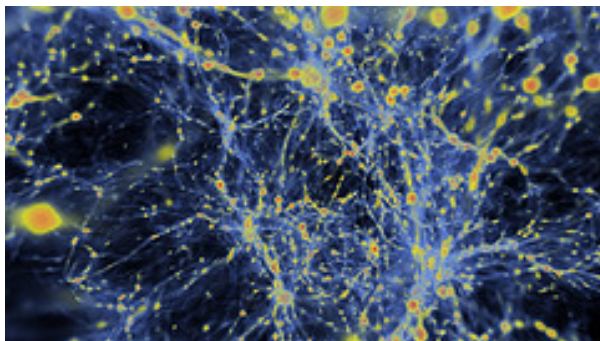
wait



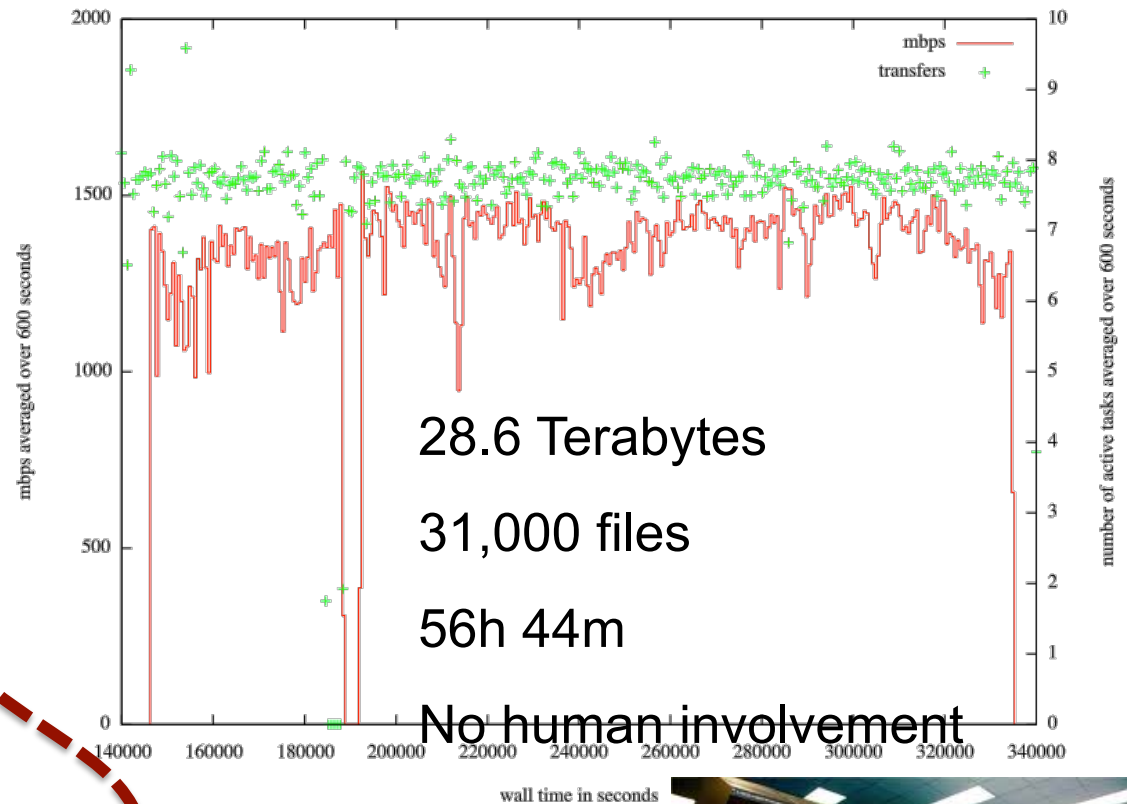
the globus alliance

www.globus.org

Globus Online in Action



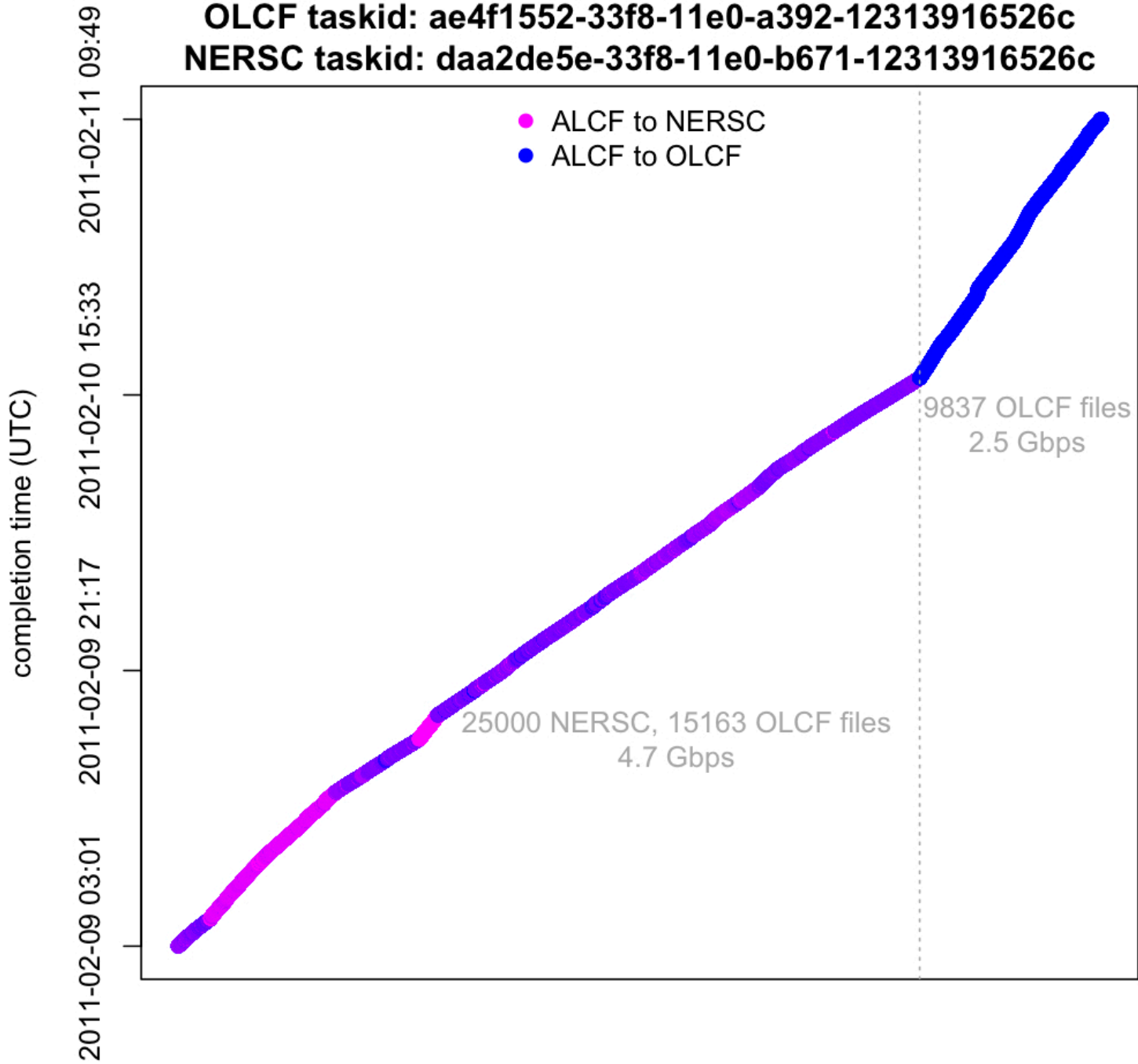
Astrophysics simulation data
generated in Tennessee,
moved to Illinois for visualization
(Enzo, UCSD; Futures Lab, Argonne)





98TB Transfer (2x25000 2GB Files)

OLCF taskid: ae4f1552-33f8-11e0-a392-12313916526c
NERSC taskid: daa2de5e-33f8-11e0-b671-12313916526c





the globus alliance

www.globus.org

Globus Online

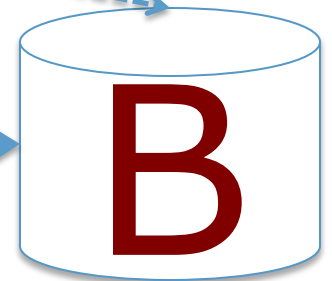
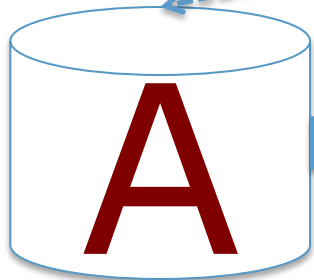




us.org

Managed
data transfer
service

Configure
Manage
Monitor
Retry
Diagnose





Summary

- Grid computing has revolutionized the way science is done
 - ◆ While big science projects can afford to operate dedicated grid infrastructures, smaller teams can't
- Commercial cloud providers deliver on-demand cycles, storage and S/W as a service
 - ◆ Benefit small and medium-sized business
- Globus Online is our first foray into what you might call a computational science cloud: hosted services that let you "use the grid" without installing software
 - ◆ Benefit big science projects as well as small and medium research groups



the globus alliance

www.globus.org

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 Kandaswamy, 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