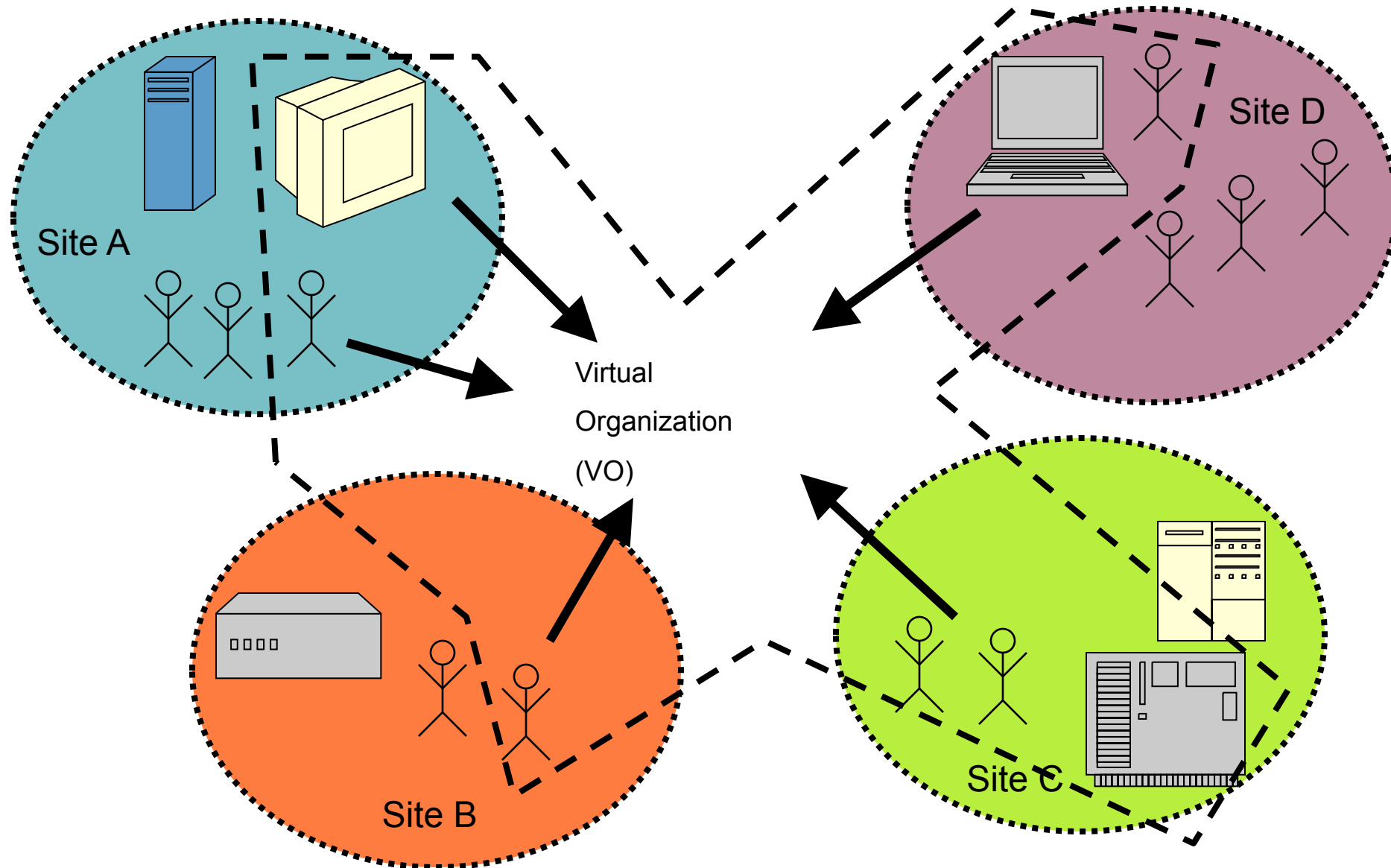# Globus Online: Hosted Service for High-speed, Reliable and Secure Data Movement

Raj Kettimuthu

Argonne National Laboratory and

The University of Chicago
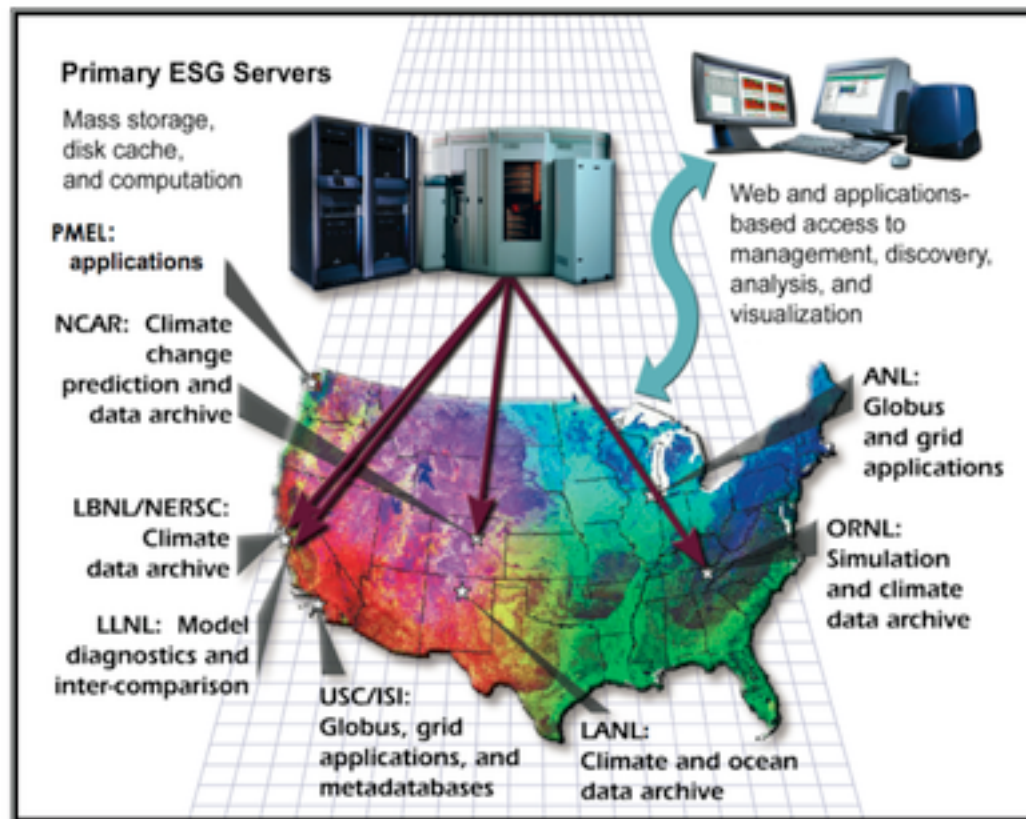
Grid

the globus alliance
www.globus.org

Site A

Site D

Virtual

Organization

(VO)

Site B

Site C

03/31/2011

TIFR, India

Distributed Science

- Distributed community of users to access and analyze large amounts of data

Large Hadron Collider

1800 Physicists, 150 Institutes, 32 Countries

100 PB of data by 20XX; 50,000 CPUs?
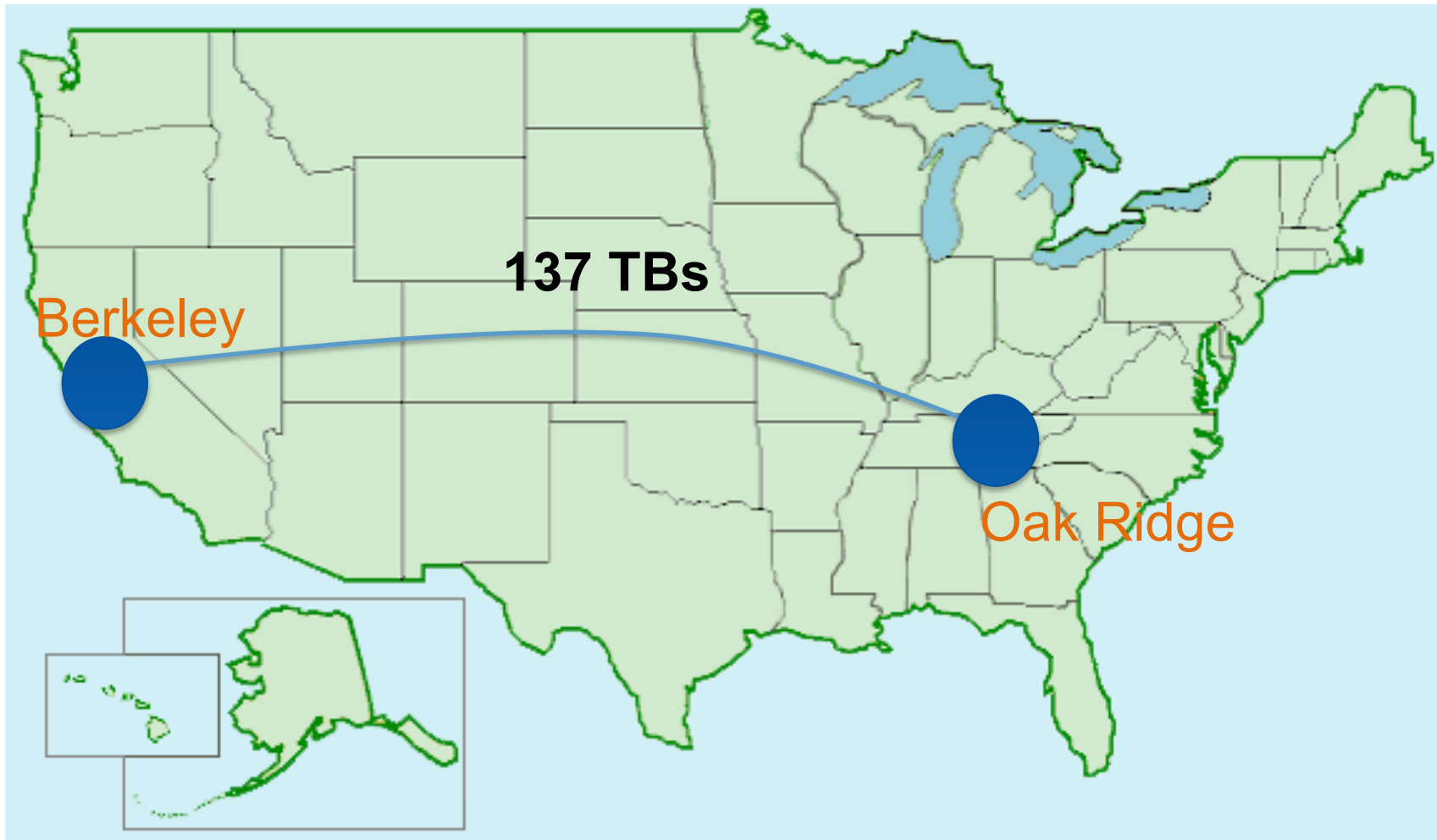
Data Movement

- 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

TIFR, India

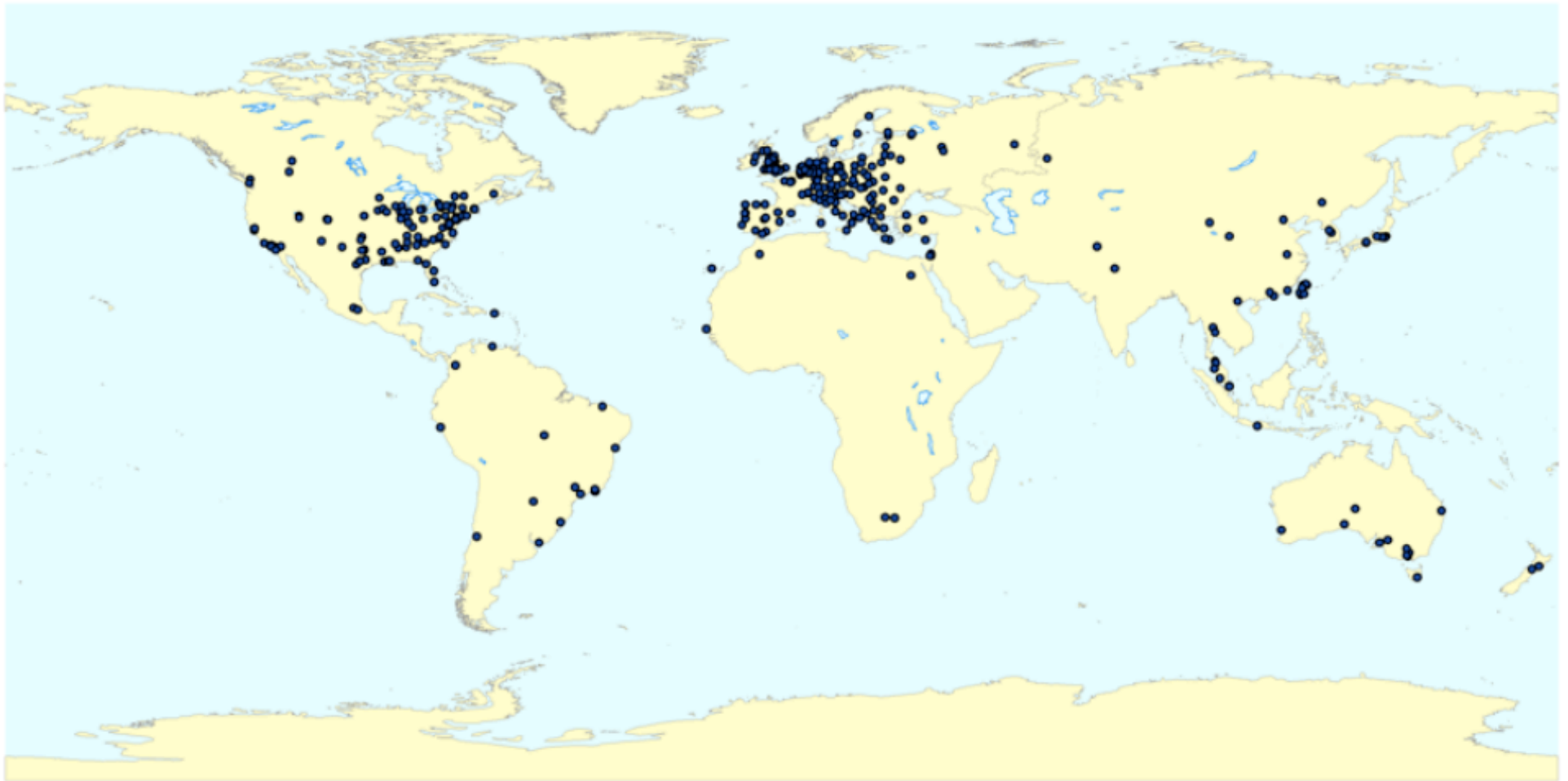# Data Movement Challenges

137 TBs

Berkeley

Oak Ridge

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

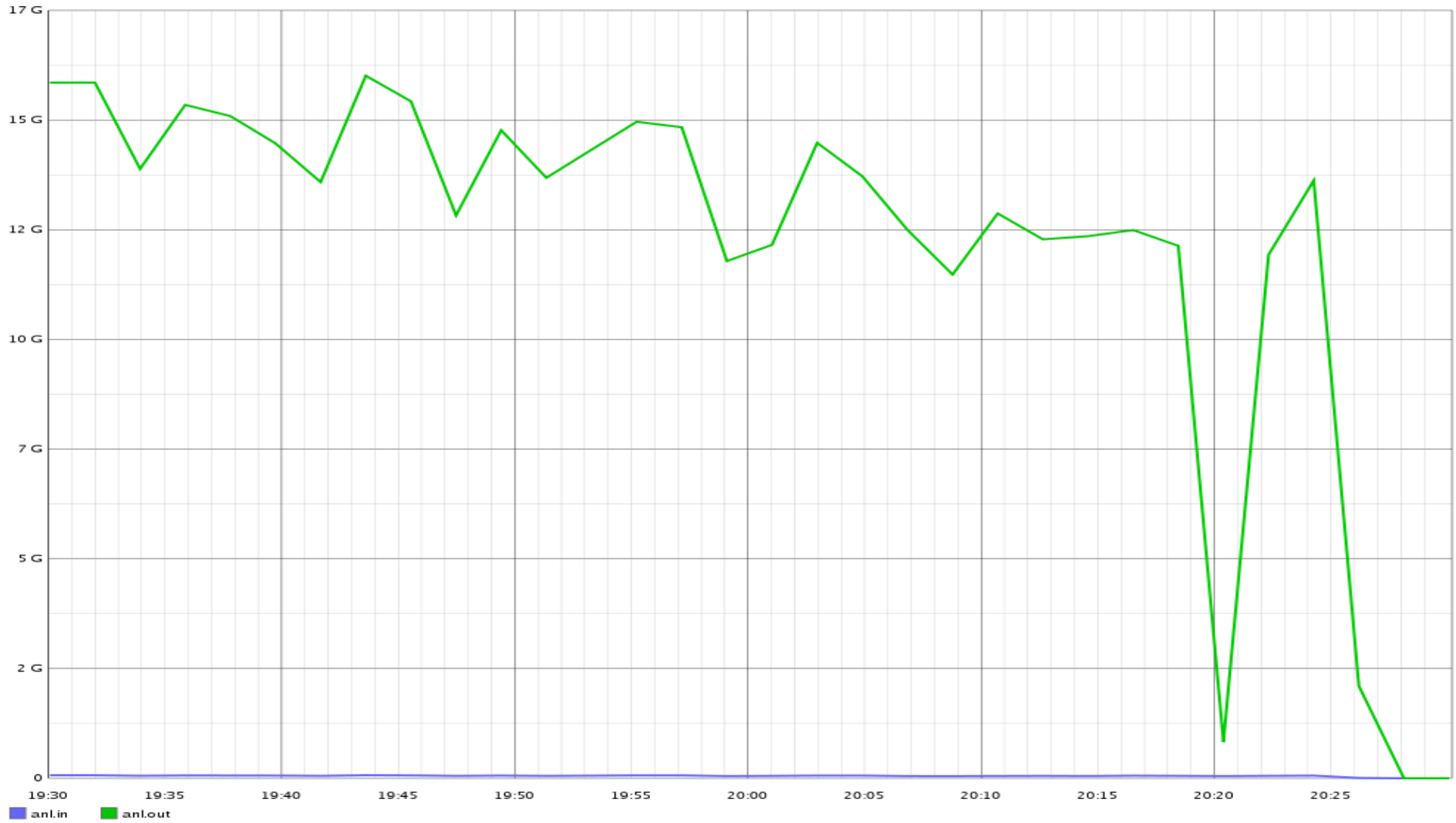# GridFTP Servers Around the World

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

Performance

TIFR, India

GridFTP Usage

Monthly Totals* of GridFTP File Transfers
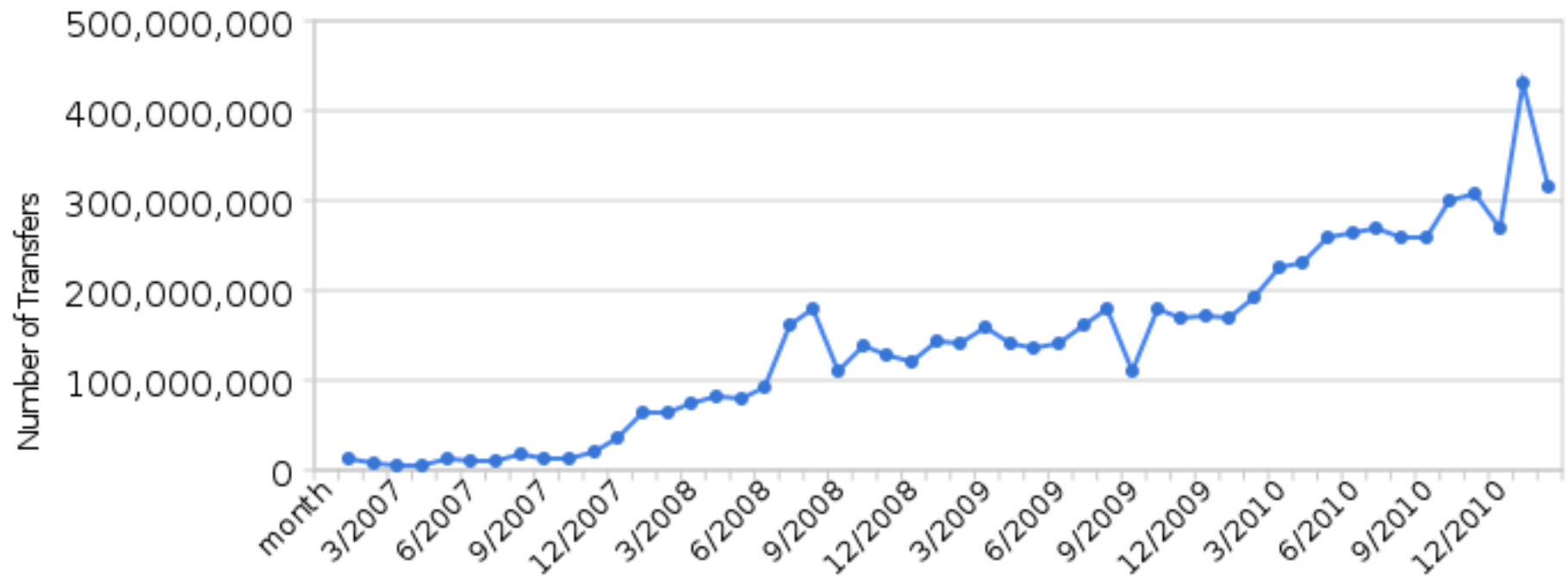
*for those "reporting"

Data Movement Challenges
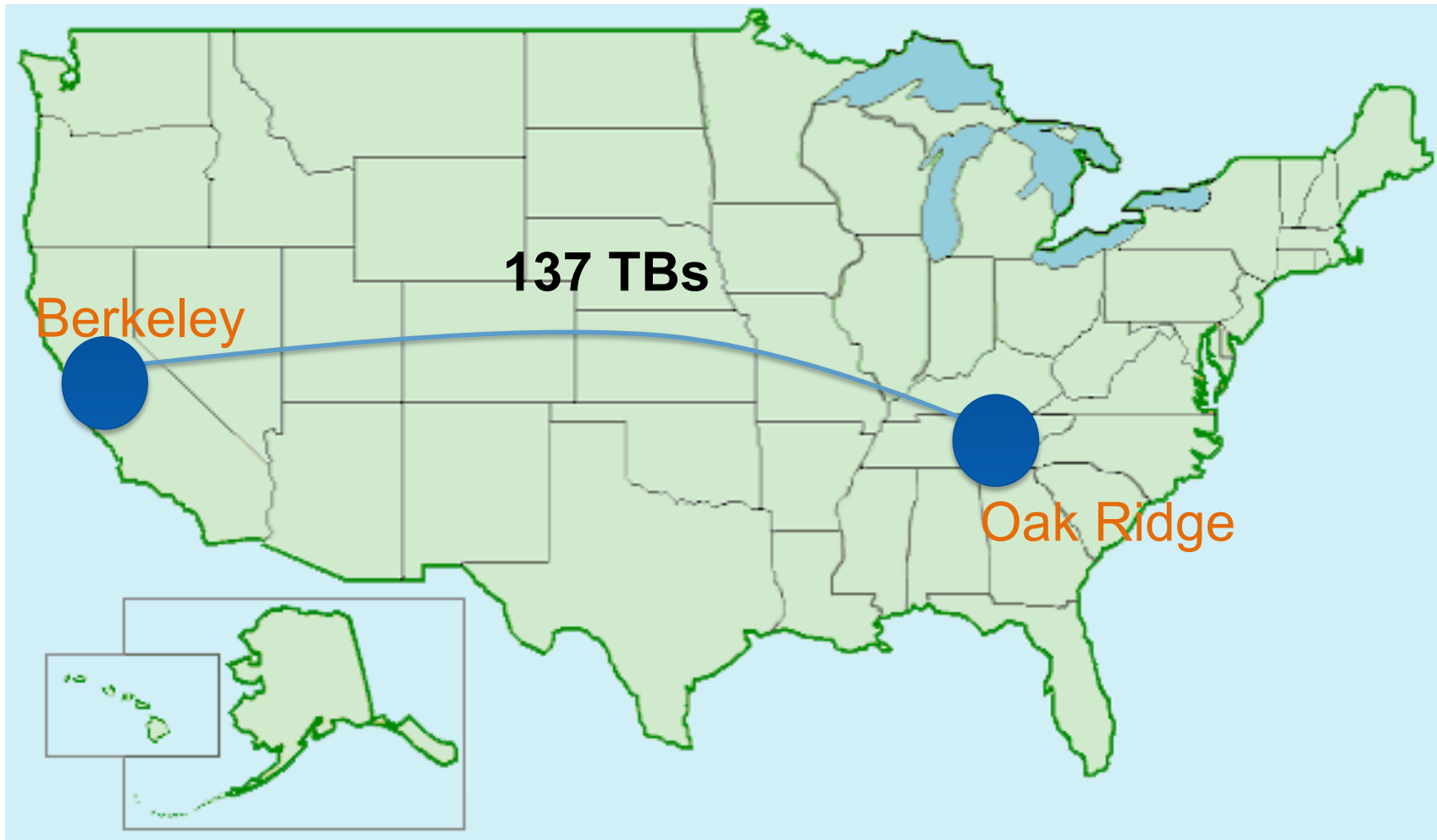
137 TBs

Berkeley

Oak Ridge

TIFR, India

Advanced Photon Source

TIFR, India

Data Movement Challenges

...he found himself crushed by the shear amount of data produced!

TIFR, India

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, …
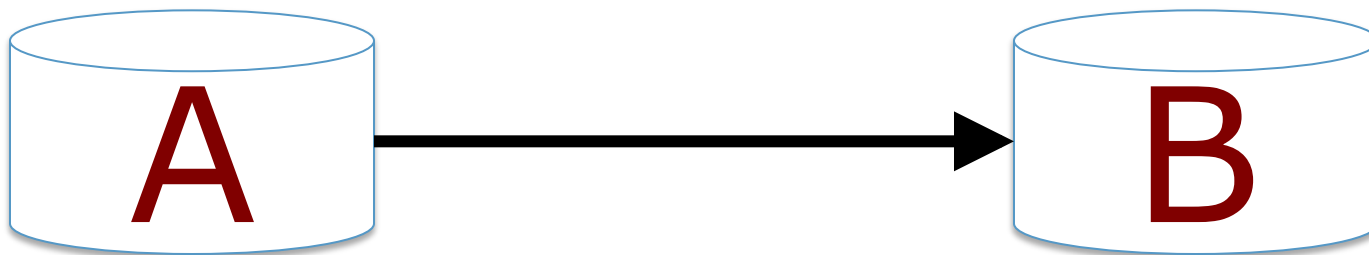
Cloud Computing
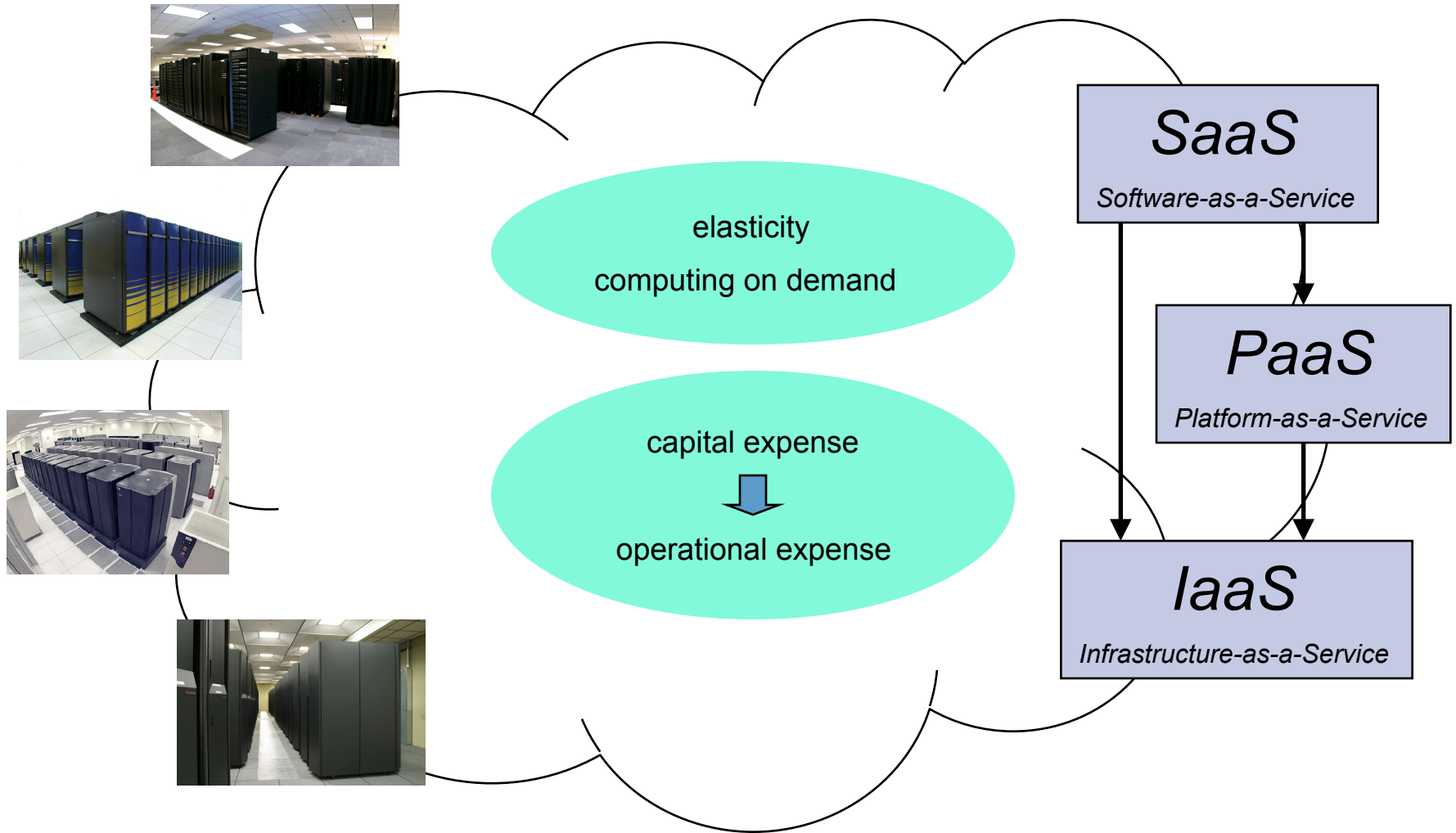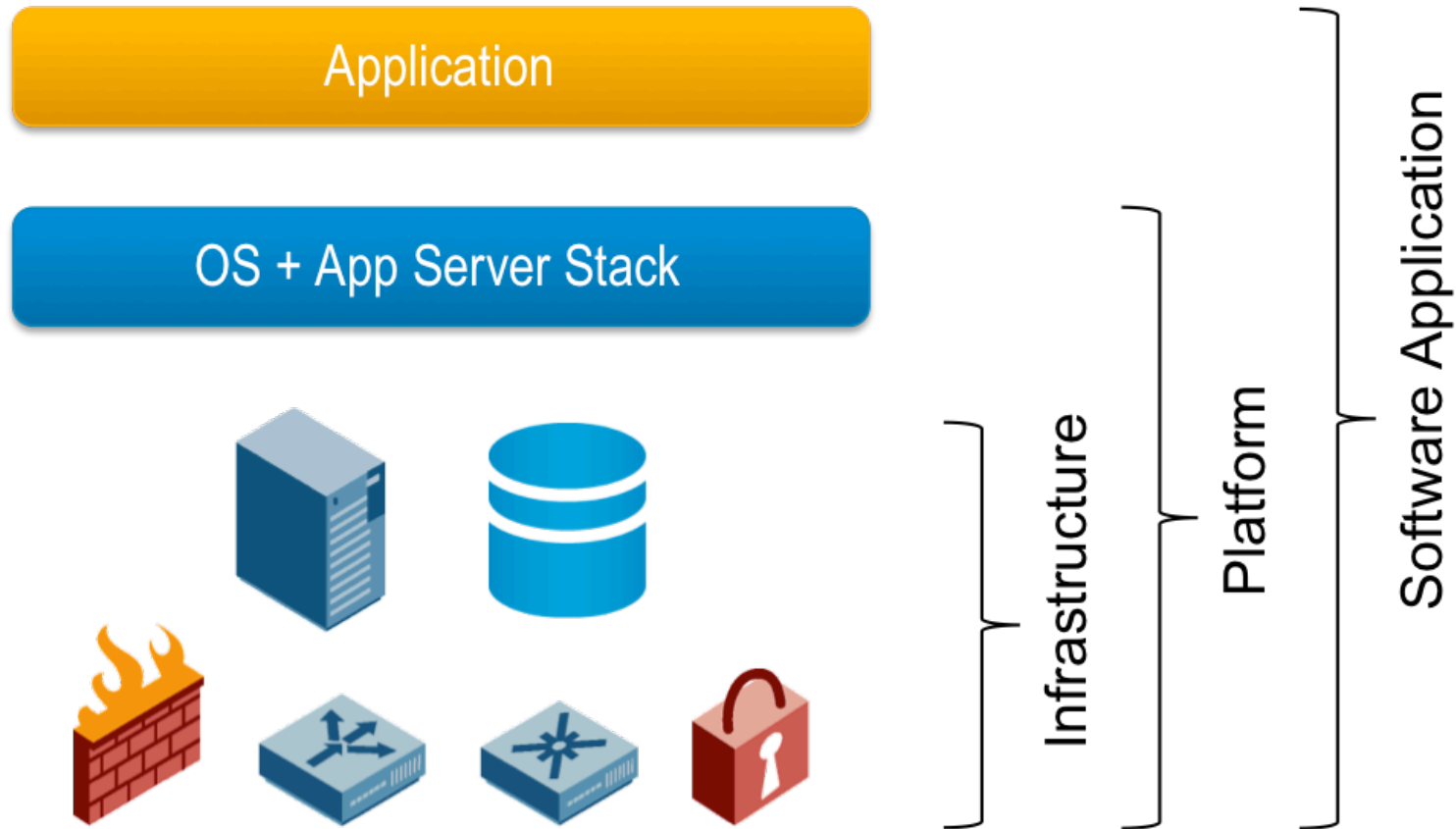
elasticity

computing on demand

capital expense

operational expense

SaaS
Software-as-a-Service

PaaS
Platform-as-a-Service

IaaS
Infrastructure-as-a-Service

Cloud Computing

TIFR, 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)

Time-consuming tasks in business

the globus alliance

Web presence

Email (hosted Exchange)

Calendar

Telephony (hosted VOIP)

Human resources and payroll

Accounting

Customer relationship mgmt

SaaS

Data analytics

Content distribution

...

IaaS

Cloud

the globus alliance
www.globus.org

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
- …

Globus Online

- *Globus Toolkit*

  Build the Grid

- *Globus Online*

  Use the Grid

Components for building
custom grid solutions

Cloud-hosted
file transfer service

globustoolkit.org

globusonline.org

TIFR, India

- 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
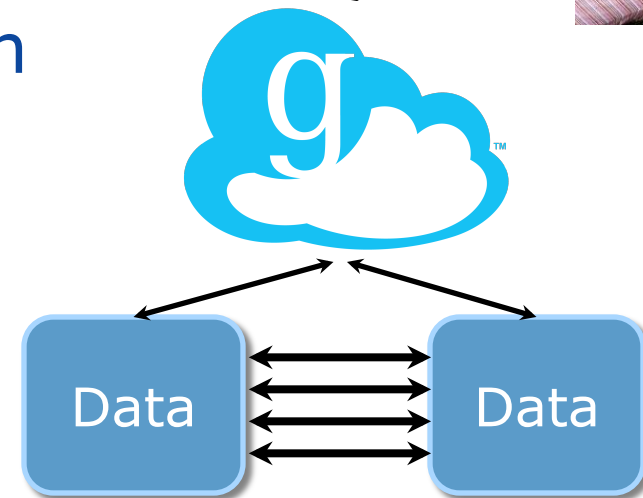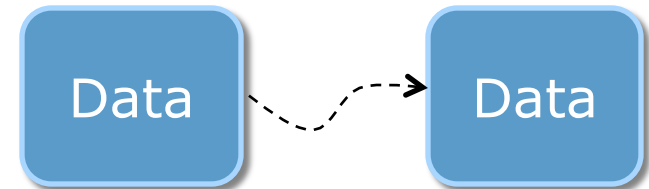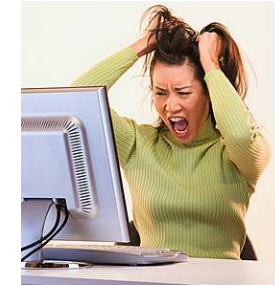
globus online beta

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

Raj Kettimuthu

Username

rajk

Email address

kettimut@mcs.anl.gov

Confirm email

kettimut@mcs.anl.gov

Save changes

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

03/31/2011   TIFR, India

Initiate Transfers

Transfers In Progress: 0   **View All Transfers**

Endpoint [_____] ▼   **Endpoint** [_____] ▼

(Go)   (Go)

Directory [_____]   **Mode:**   **Directory** [_____]

                                         [Transfer ▾]

(Go)   (Go)

📁 Up Directory   🔄 Refresh Directory   Select All   📁 Up Directory   🔄 Refresh Directory   Select All

```
┌──────────────────────────┐        ┌──────────────────────────┐
│                          │        │                          │
│                          │        │                          │
│                          │   >    │                          │
│                          │        │                          │
│                          │        │                          │
│                          │   <    │                          │
│                          │        │                          │
│                          │        │                          │
└──────────────────────────┘        └──────────────────────────┘
```

Clear Selections   Clear Selections

Globus Online	**Support**	**Community**
Overview	Documentation	News and Events
Sponsors	Video Tutorials	Globus Toolkit

03/31/2011   TIFR, India

Application: Initiate transfer ▼

Hi, rajk! Edit profile Sign-out

Initiate Transfers

Transfers In Progress: 0   View All Transfers

Endpoint [   ] ▼   Endpoint [   ] ▼

(Go)   (Go)

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

Directory [   ]   Mode: Directory [   ]

(Go)   Transfer ▼   (Go)

Up Dir   Select All   Up Directory   Refresh Directory   Select All

>

<

Clear Selections   Clear Selections

Globus Online   **Support**   **Community**
Overview   Documentation   News and Events
Sponsors   Video Tutorials   Globus Toolkit

03/31/2011   TIFR, India

rajk#alcf

rajk#alcf-personal

rajk#kraken

rajk#rajlaptop

rajk#stomp

rajk#tg-queenbee

rajk#tg-steele

alcf#dtn

ci#pads

go#ep1

globus online beta

Application: Monitor transfers ▾   **Hi, rajk!** Edit profile Sign-out

Monitor Transfers

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

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
[XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX] 1/1 0.00
mbps
ssh lcc@cli.globusonline.org ls go#ep2/~/
myfile.txt
```

CLI commands

cancel

details

endpoint-activate

endpoint-add

endpoint-deactivate

endpoint-list

endpoint-modify

endpoint-remove

endpoint-rename
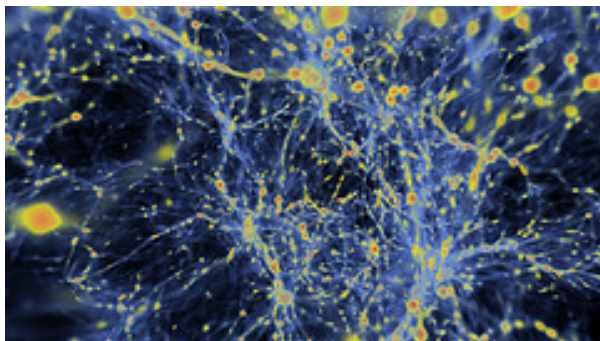
events

ls

profile
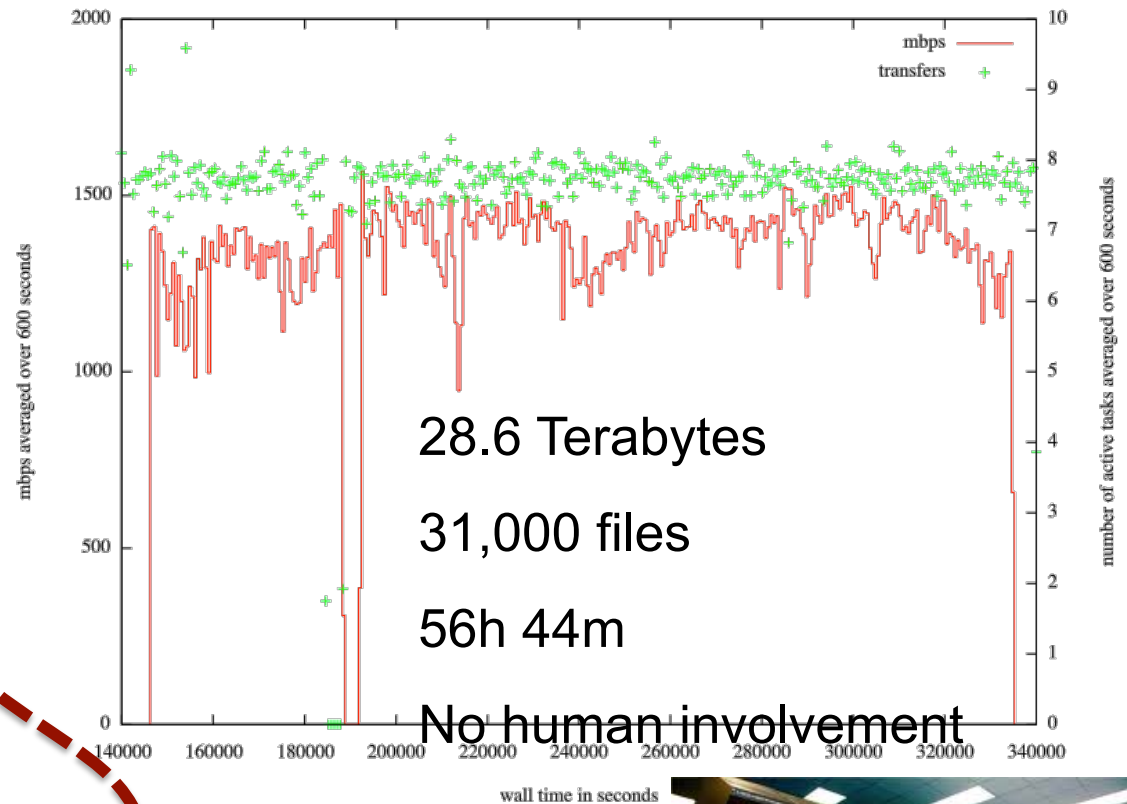
scp

status

transfer

versions

wait

the globus alliance
www.globus.org

Globus Online in Action

28.6 Terabytes
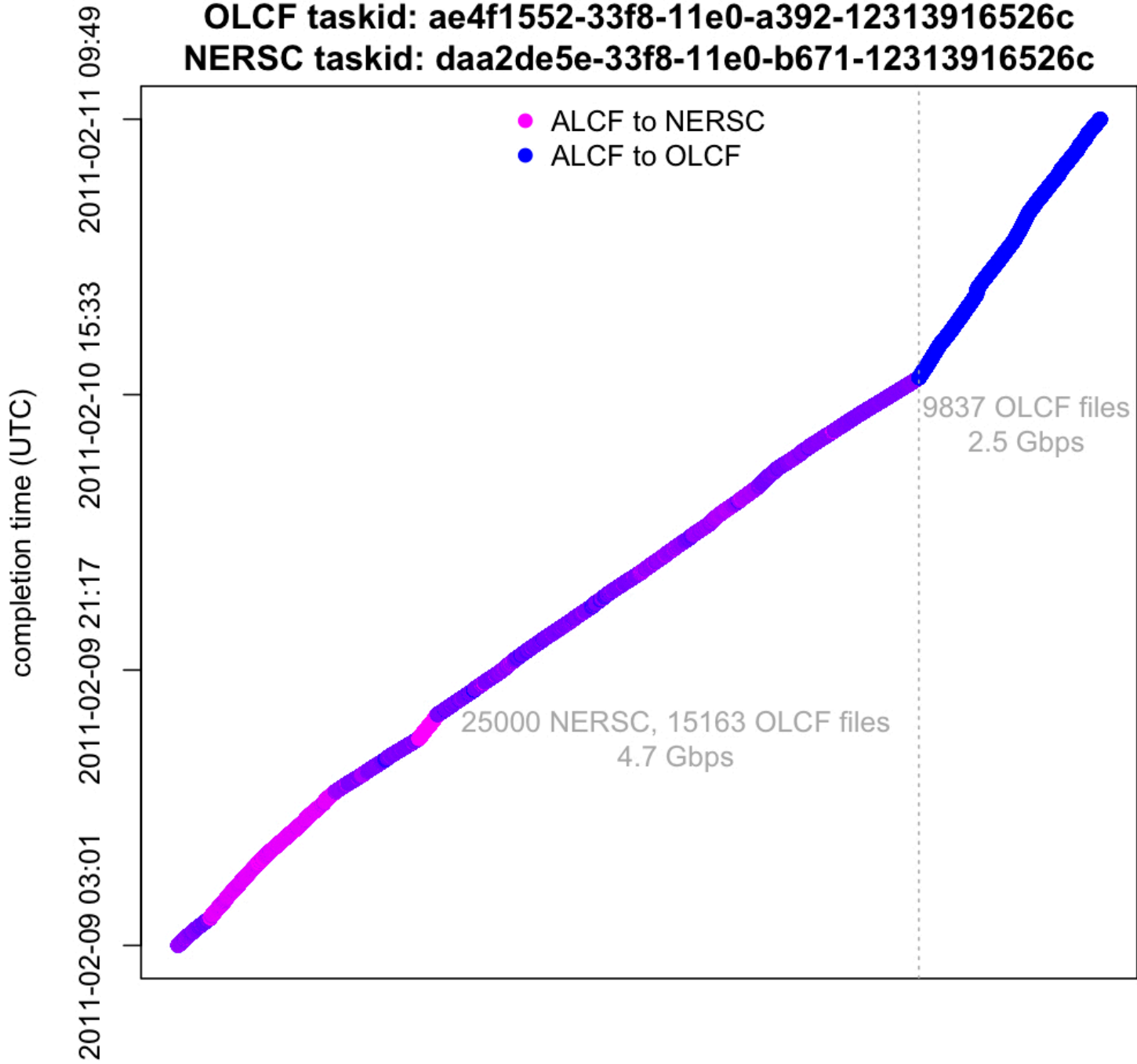
31,000 files

56h 44m

No human involvement

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

● ALCF to NERSC
● ALCF to OLCF

9837 OLCF files
2.5 Gbps

25000 NERSC, 15163 OLCF files
4.7 Gbps

completion time (UTC)

2011-02-11 09:49
2011-02-10 15:33
2011-02-09 21:17
2011-02-09 03:01

03/31/2

Globus Online

GridFTP

GridFTP

Notification target

Profiles + state

Worker

Request collector

amazon.com® web services

03/31/2011   TIFR, India

Managed data transfer service

Configure
Manage
Monitor
Retry
Diagnose

A → B

03/31/2011 TIFR, India
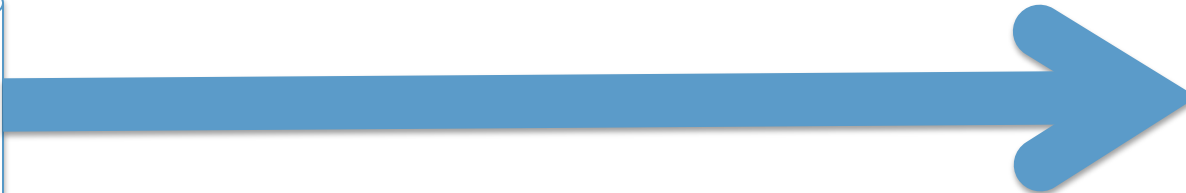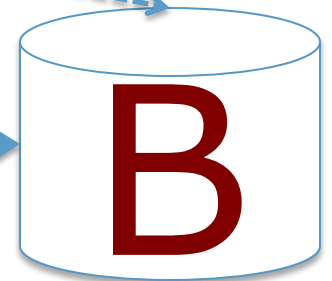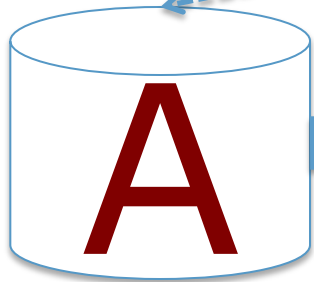
Summary

- 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

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

03/31/2011   TIFR, India

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

Questions