

Science Gateways on the TeraGrid

Jeffrey P. Gardner

Pittsburgh Supercomputing Center

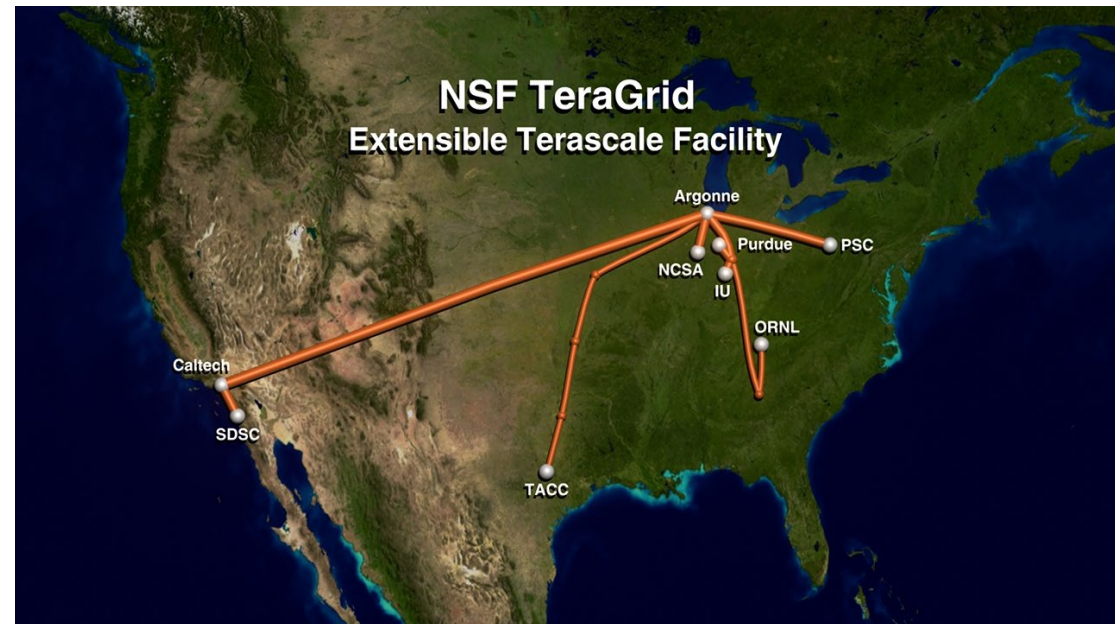
gardnerj@psc.edu

Outline

- Brief introduction to the TeraGrid
- Making the TeraGrid useful for Science Gateway communities

The TeraGrid

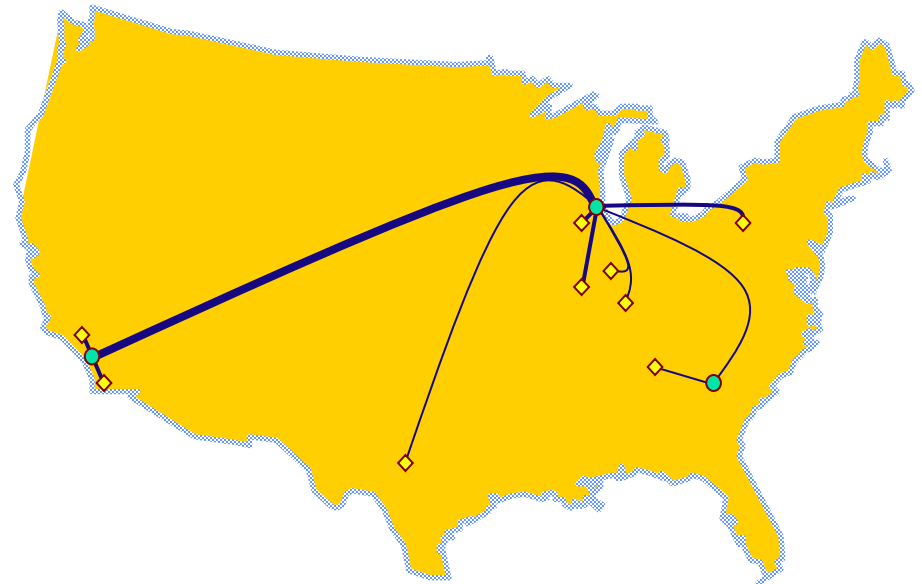
- A major paradigm shift for HPC resource providers
- Make NSF resources useful to a wider community



- *Strength through diversity!*

The TeraGrid Strategy

- Building a distributed system of unprecedented scale
 - 40+ teraflops compute
 - 1+ petabyte storage
 - 10-40Gb/s networking
- Creating a **unified user environment** across heterogeneous resources
 - Single user support resources.
 - Single authentication point
 - Common software functionality
 - Common job management infrastructure
 - Globally-accessible data storage
- Create a ***unified*** national HPC infrastructure that is both ***heterogeneous*** and ***extensible***



“Grid-Like” Usage Scenarios

Currently Enabled by the TeraGrid

TeraGrid is in its infancy!

- “Traditional” massively parallel jobs
 - Tightly-coupled interprocessor communication
 - storing vast amounts of data remotely
 - remote visualization
- Thousands of independent jobs
 - Automatically scheduled amongst many TeraGrid machines
 - Use data from a distributed data collection
- Multi-site parallel jobs
 - Compute upon many TeraGrid sites simultaneously

Science Gateways

A new initiative for the TeraGrid

- Science communities increasingly **building their own cyberinfrastructure: "Science Gateways"**.
 - Provides an easy-to-use solution for community computational needs.
 - Both software and hardware
- Science Gateways will
 - **increase overall science throughput by making computers easier to use for knowledge discovery.**

Science Gateways

A new initiative for the TeraGrid

- But many will often have computing needs that far **exceed their own capacity**.
- The **TeraGrid** can provide massive computing resources to these communities **on demand**.
- Science Gateways can use the TeraGrid to
 - **provide community-tailored access to massive compute resources at no additional cost**
- Goals:
 - *Integration* and *interoperability*.

Three Types of Science Gateways

■ Web-based Portals

- User interacts with community-deployed web interface.
- Runs community-deployed codes
- Service requests forwarded to TeraGrid resources

■ Bridges to Community Grids

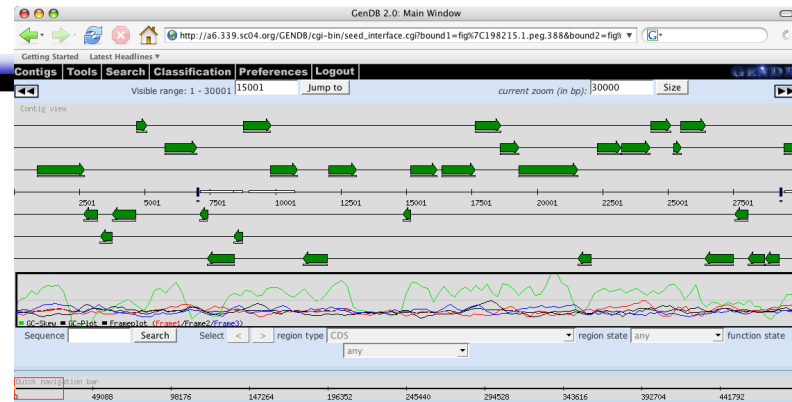
- Jobs forwarded seamlessly between science community grids and TeraGrid

■ Service Gateways

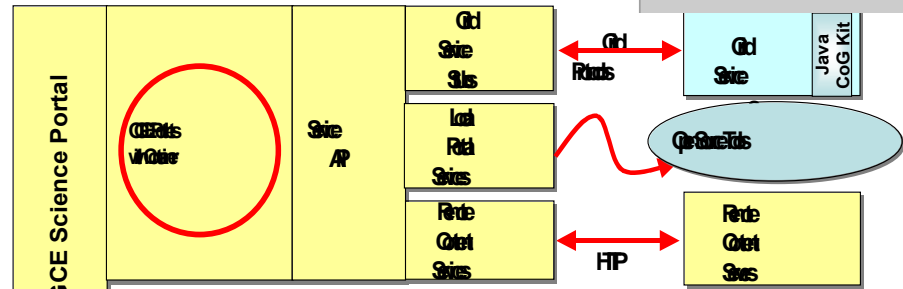
- Application programs running on users' machines but access services in TeraGrid (and elsewhere)

Types of Gateways: *Web-Based Portal*

- A portal is a gateway to a set of distributed services accessible from a Web browser or desktop tools
- Example: **GEON**
- Provides single point of access to:
 - Grid authentication
 - Community-deployed data
 - Community-deployed services (e.g. **Synseis**)
- Seamlessly forwards service requests (and data?) to TeraGrid resources



This block contains two screenshots. The top one is the 'LEAD Portal Home Page' with a navigation menu (LEAD Home, Education & Outreach, Research & Applications, News & Information), a login form (Name, Password, Login), and a central map area. The bottom screenshot is the 'Workflow Composer' interface, showing a 'Component Selector' with various services like 'Decoder and Data Mover Service' and 'THREDDS catalog Generator', and a visual workflow diagram connecting these components.



Three Types of Science Gateways

■ Web-based Portals

- User interacts with community-deployed web interface.
- Runs community-deployed codes
- Service requests forwarded to TeraGrid resources

■ Bridges to Community Grids

- Jobs forwarded seamlessly between science community grids and TeraGrid

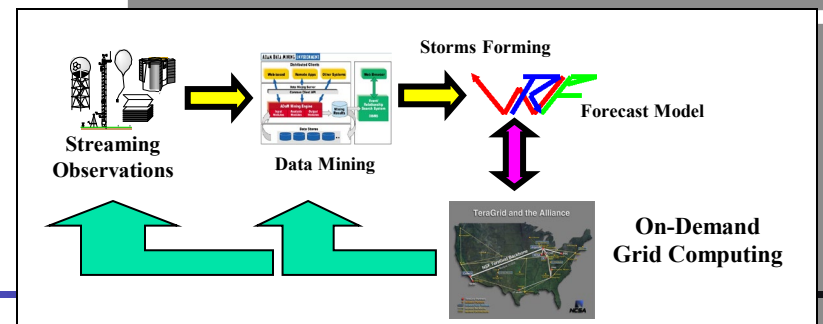
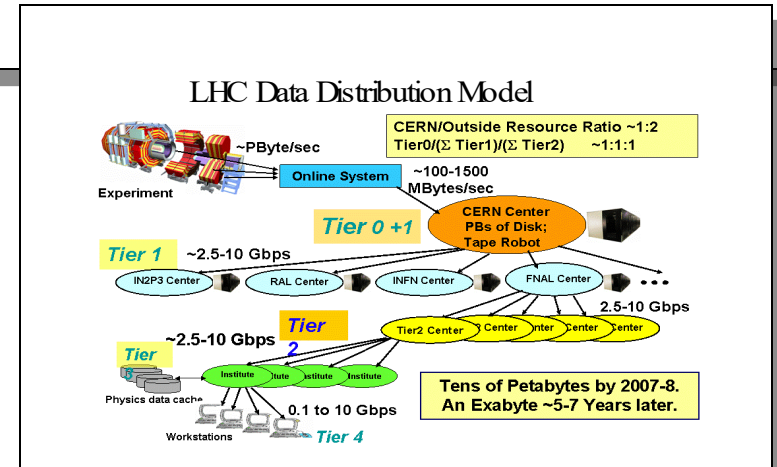
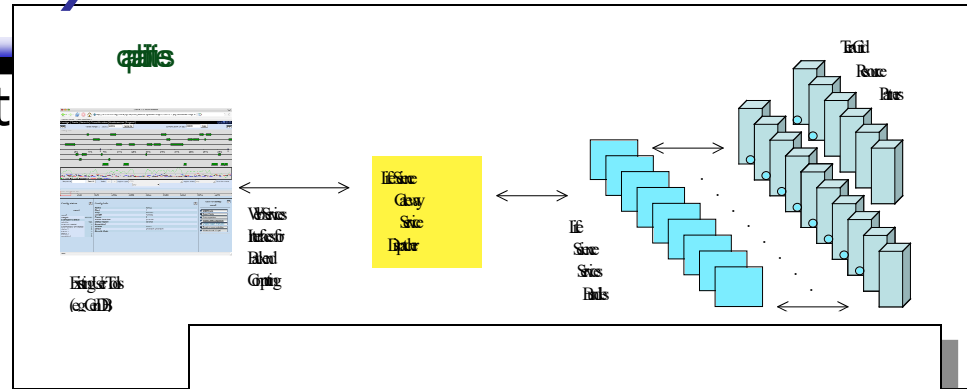
■ Service Gateways

- Application programs running on users' machines but access services in TeraGrid (and elsewhere)

Types of Gateways:

Bridges to Community Grids

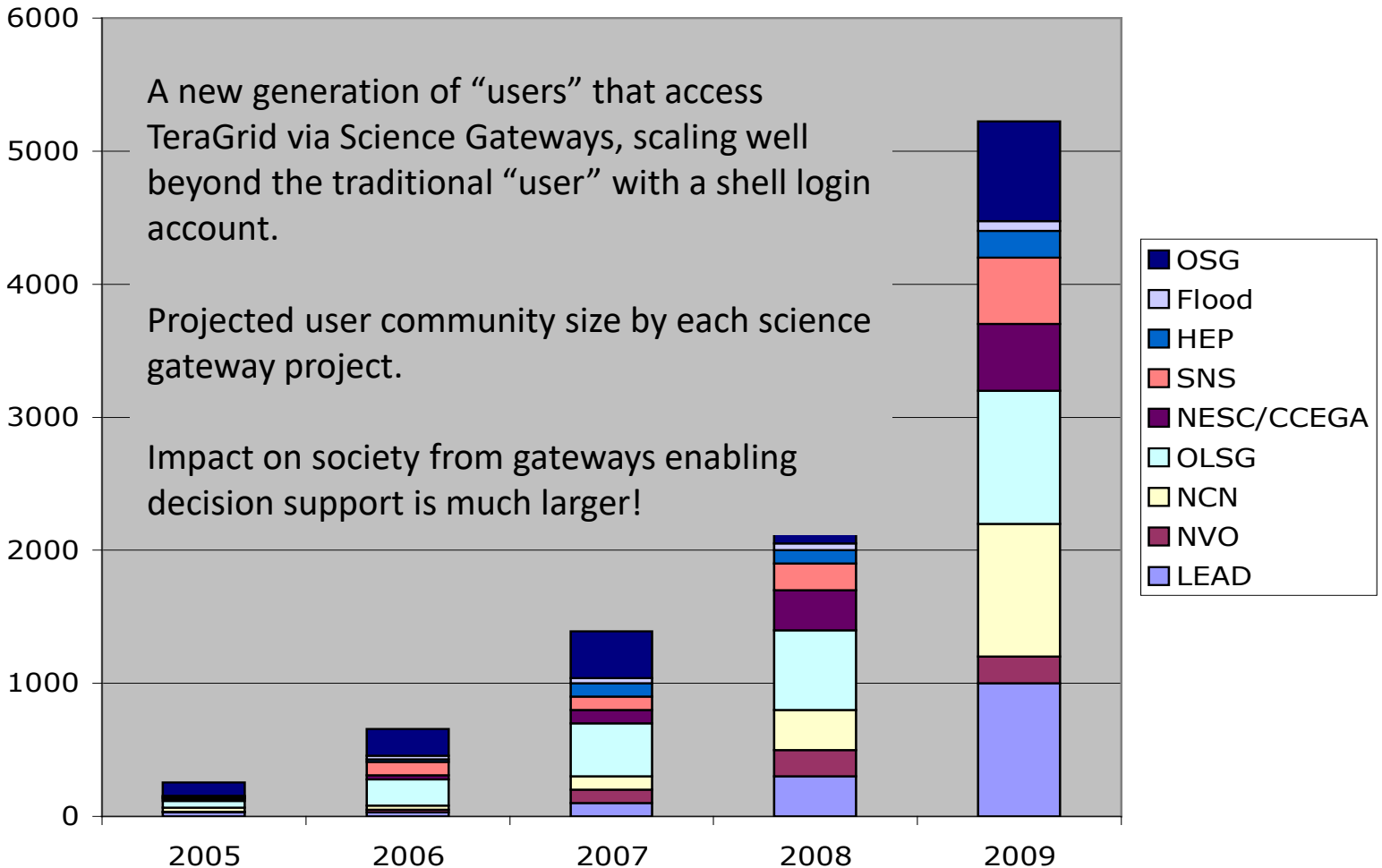
- Many Community Grids already exist or are being built
 - NEESGrid, LIGO, Earth Systems Grid, NVO, Open Science Grid, etc.
- TeraGrid will provide a service framework to enable access in ways that are transparent to their users.
 - The community maintains and controls the Gateway
- Different Communities have different requirements.
 - NEES and LEAD will use TeraGrid to provision compute services
 - LIGO and NVO have substantial data distribution problems.



Science Gateways

An expanding user base

Users



Science Gateways

A new ~~initiative~~ for the TeraGrid

Challenge

- **Heterogeneity**
 - Software stacks - every gateway is a "custom job".
 - Resources - different architectures at local, national and international levels
 - Users- from HPC developer to K-12 student
 - Security policies
- How can "centers/Institutions" like the TeraGrid provide and operate in this heterogeneous world?
- Working with Gateways, TeraGrid is starting to answer that question
 - Identifying important generic CyberInfrastructure services
 - Providing those services to communities.

Initial Focus on 10 Gateways

<u>Science Gateway Prototype</u>	<u>Discipline</u>	<u>Science Partner(s)</u>	<u>TeraGrid Liaison</u>
Linked Environments for Atmospheric Discovery (LEAD)	Atmospheric	Droegemeier (OU)	Gannon (IU), Pennington (NCSA)
National Virtual Observatory (NVO)	Astronomy	Szalay (Johns Hopkins)	Williams (Caltech)
Network for Computational Nanotechnology (NCN) and "nanoHUB"	Nanotechnology	Lundstrum (PU)	Goasguen (PU)
National Microbial Pathogen Data Resource Center (NMPDR)	Biomedicine and Biology	Schneewind (UC), Osterman (Burnham/UCSD), DeLong (MIT), Dusko (INRA)	Stevens (UC/Argonne)
NSF National Evolutionary Biology Center (NESC), NIH Carolina Center for Exploratory Genetic Analysis, State of North Carolina Bioinformatics Portal project	Biomedicine and Biology	Cunningham (Duke), Magnuson (UNC)	Reed (UNC), Blatecky (UNC)
Neutron Science Instrument Gateway	Physics	Dunning (ORNL)	Cobb (ORNL)
Grid Analysis Environment	High-Energy Physics	Newman (Caltech)	Bunn (Caltech)
Transportation System Decision Support	Homeland Security	Stephen Eubanks (LANL)	Beckman (Argonne)
Groundwater/Flood Modeling	Environmental	Wells (UT-Austin), Engel (ORNL)	Boisseau (TACC)
Science Grid [GrPhyN/ivDGL/Grid3]	Multiple	Pordes (FNAL), Huth (Harvard), Avery (Uflorida)	Foster (UC/Argonne), Kesselman (USC-ISI), Livny (UW)

Timelines - Fall, 2005

- Deploy 3 prototype portals
 - LEAD, Bioinformatics, Evolutionary Biology
- Define work plan and application characteristics
 - NVO, nanoHub, Neutron Science
- Port/install software
 - Homeland Security, Flood Analysis, OSG
- Re-Analyze Gateway needs
- Draft Gateways Primer document

Gateways Primer Outline:

Defining generic cyberinfrastructure services

- 1. Introduction
- 2. Science Gateway in Context
 - a. Science Gateway (SGW) Definition(s)
 - b. Science Gateway user modes
 - c. Distinction between SGW and other TeraGrid user modes
- 3. Components of a Science Gateway
 - a. User Model
 - b. Gateway targeted community
 - c. Gateway Services
 - d. Integration with TeraGrid external resources (data collections, services, ...)
 - e. Organizational and administrative structure
- 4. TeraGrid services and policies available for Science Gateways
 - a. Portal middleware tools (user portal and other portal tools)
 - b. Account Management (user models, community accounts,)
 - c. Security environment (security models)
 - d. Web Services
 - e. Scheduling services (and meta-scheduling)
 - f. Community accounts and allocations
 - g. Community Software Areas
 - h. All traditional TeraGrid services and resources
 - i. Ability to propose additional services and how that would interact with TeraGrid operations
- 5. Responsibilities and Requirements for Science Gateways
 - a. Interaction with and compatibility with TeraGrid communities
 - b. Control procedures
 - i. Community user identification and tracking (map TeraGrid usage to Portal user)
 - ii. Use monitoring and reporting
 - iii. Security and trust
 - iv. Appropriate use
- 6. How to get started
 - a. Existing resources
 - i. Publication references
 - ii. Web areas with more details
 - iii. Online tutorials
 - iv. Upcoming presentations and tutorials
 - b. Who to contact for initial discussions
 - c. How to propose a new Gateway
 - d. How to integrate with TeraGrid Gateways efforts.
 - e. How to obtain a resource allocation


Spring, 2006

- Integrate TeraGrid compute resources
 - LEAD, nanoHUB, Bioinf., Evo. Bio., HEP, OSG
- Prototypes
 - web/grid services (Bioinformatics)
 - Data archive hosting (Neutron Science)
 - Data federation models with compute support (Evolutionary Biology)
 - Application hosting services, initial compute resource brokering and data federation. Test for security, scalability.
- Explore authentication methods (NVO)

Summary

- TeraGrid offers Science Gateways the ability to extend their compute power on demand.
- Supporting Science Gateways is important to the TeraGrid.
- TeraGrid is starting to work closely with ~ 10 Science Gateway communities to define and implement common services.
- Prototype gateways deployed in ~ 1 year

Portal Example: The TeraGrid Portal!


My Workspace
Sep 12, 2004 01:55 pm

Home

Membership

Schedule

Resources

Proxy Manager

TeraGrid CVS

TeraGrid UserInfo

TeraGrid Monitor

TeraGrid INCA

TeraGrid Networks

TeraGrid Bug Tracking

Newsgroup (Read/Post)

Newsgroup (Request)

Systems View

Logout

Customize

Eric Roberts

Gpir Browser [Help](#)

Note: This portlet requires an external service (GPIR) to be installed and configured before it can be used. Please consult the help documentation for further details.

Systems View | [Grid View](#)

High Performance Computing Resources

Institution	Dept	Name	System	Peak CPUs	Memory GFlops	Memory GBytes	Disk GBytes	Motd	Status	Load	Jobs
The University of Texas at Austin	Texas Advanced Computing Center	Bandera	Dell Linux Cluster	8	25	8	365	Q	↑	0%	0R-0Q-00
The University of Texas at Austin	Texas Advanced Computing Center	Blanco	Dell Linux Cluster	8	25	8	365	Q	↑	0%	0R-0Q-00
The University of Texas at Austin	Texas Advanced Computing Center	Buda	Dell Linux Cluster	8	25	8	365	Q	↑	0%	0R-0Q-00

Visualization Resources

Institution	Dept	Name	System	Peak CPUs	Memory GFlops	Memory GBytes	Disk GBytes	Motd	Status
Test University	Test Computing Center	TestResource10000	Test System	5	55	55	555	Q	

Archival Systems

[Cache](#) [Storage](#)

Grid Networking Status

Sep 12, 2004 01:56 pm



My Workspace

Gpir Browser

[Help](#)

- Home
- Membership
- Schedule
- Resources
- Proxy Manager
- TeraGrid CVS
- TeraGrid UserInfo
- TeraGrid Monitor
- TeraGrid INCA
- TeraGrid Networks
- TeraGrid Bug Tracking
- Newsgroup (Read/Post)
- Newsgroup (Request)
- Systems View**
- Logout
- Customize
- Eric Roberts

[Systems View](#) | Grid View

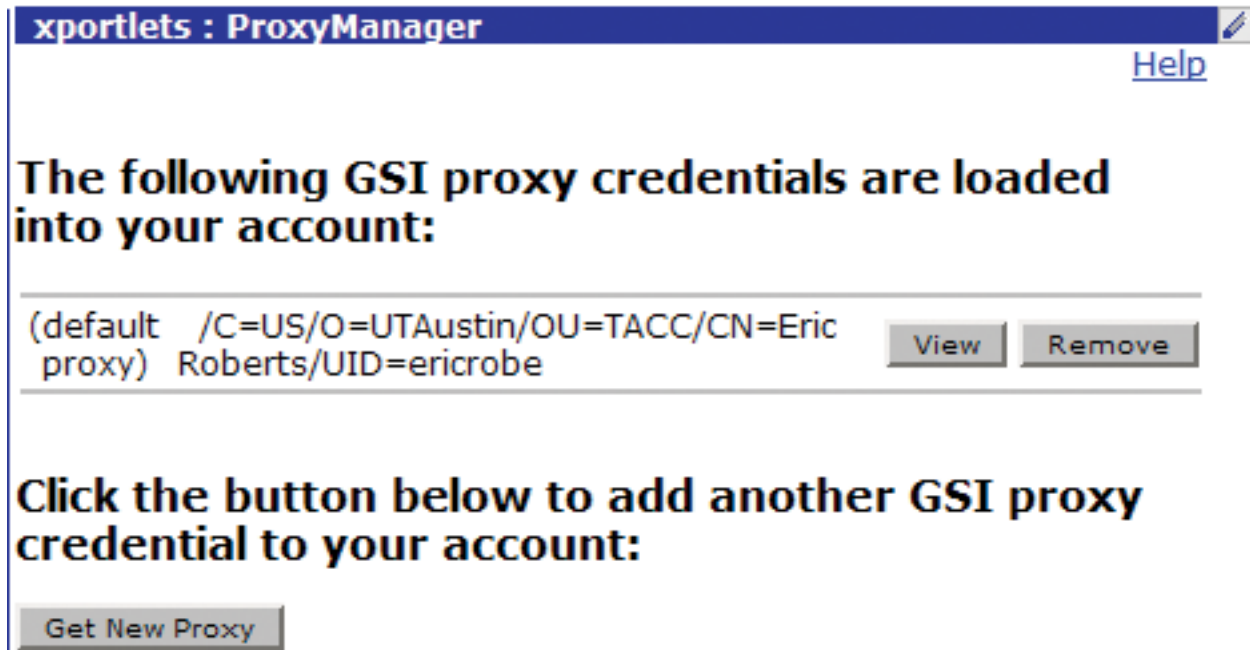
Total Statistics

Total Processors: 29
 Total Nodes: 16
 Total Peak Performance (GFlops): 130
 Total Memory (GBytes): 79
 Total Scratch Disk (GBytes): 1650
 Total Offline Disk Space (GBytes): 555
 Total Online Disk Space (TBytes): 555
 Total Peak Polygons (MPoly/sec): 55

Network Performance

	bandera	blanco	buda	test10000	test9000
bandera	N / A	931.56476 0.094	863.8969 0.745	-	-
blanco	912.65924 0.765	N / A	934.9154 0.091	-	-
buda	875.6196 0.76	871.46454 0.761	N / A	-	-
test10000	-	-	-	N / A	-
test9000	-	-	-	-	N / A

Manage Proxies



xportlets : ProxyManager [Help](#)

The following GSI proxy credentials are loaded into your account:

(default /C=US/O=UTAustin/OU=TACC/CN=Eric proxy) Roberts/UID=ericrobe

Click the button below to add another GSI proxy credential to your account:

Remote Command Execution

XPortlets: Gram Job Launcher [Help](#)

Launch a Gram Job

Executable name *

Parameters

Host name *

Executable Directory

Output file *

Job Submission

Grid Job Submission Portlet

[Monitor submitted jobs](#)

Job name

Host name

Port

Certificate to use ▼

Note: If you don't see all your certificates above, that means some of them have expired

Executable

Arguments

Directory:

Standard Output File

Standard Error File

File Management

xportlets : GridFTP Client

Grid FTP Client

One Host Only Two Hosts

Please specify the Grid FTP Host 1:
lonestar.tacc.utexas.edu Port: 2811

Please specify Grid FTP Host 2, if you choose "Two Hosts":
longhorn.tacc.utexas.edu Port: 2811

Enter

TeraGrid Resources

	ANL/UC	Caltech	IU	NCSA	ORNL	PSC	Purdue	SDSC	TACC
Compute Resources	Itanium2 (0.5 TF) IA-32 (0.5 TF)	Itanium2 (0.8 TF)	Itanium2 (0.2 TF) IA-32 (2.0 TF)	Itanium2 (10 TF) SGI SMP (6.5 TF)	IA-32 (0.3 TF)	XT3 (10 TF) TCS (6 TF) Marvel (0.3 TF)	Hetero (1.7 TF)	Itanium2 (4.4 TF) Power4 (1.1 TF)	IA-32 (6.3 TF) Sun (Vis)
Online Storage	20 TB	155 TB	32 TB	600 TB	1 TB	150 TB		540 TB	50 TB
Mass Storage			1.2 PB	3 PB		2.4 PB		6 PB	2 PB
Data Collections			Yes	Yes			Yes	Yes	Yes
Visualization	Yes		Yes			Yes	Yes		Yes
Instruments		Yes	Yes		Yes				
Network (Gb/s,Hub)	30 CHI	30 LA	10 CHI	30 CHI	10 ATL	30 CHI	10 CHI	30 LA	10 CHI

TeraGrid Resource Partners



OAK RIDGE NATIONAL LABORATORY

CCS The Center for
Computational Sciences

DOE High Performance Computing Research Center

