

The Cyberinfrastructure Laboratory, NYS Grid, & NYSGrid.org

Russ Miller

Director, CI Laboratory

Dept of Comp Sci & Eng

Hauptman-Woodward Med Res Inst

Executive Director, NYSGrid.org

NSF, NIH, DOE, NIMA, NYS, HP

www.cse.buffalo.edu/faculty/miller/CI/



Academia in the 21st Century

- **Embrace digital data-driven society**
- **Empower students to compete in knowledge-based economy**
- **Support HPC infrastructure, research, and applications**
- **Support education, outreach, and training**
- **Deliver *high-end cyberinfrastructure* to enable efficient**
 - **Collection of data**
 - **Management/Organization of data**
 - **Distribution of data**
 - **Analysis of data**
 - **Visualization of data**



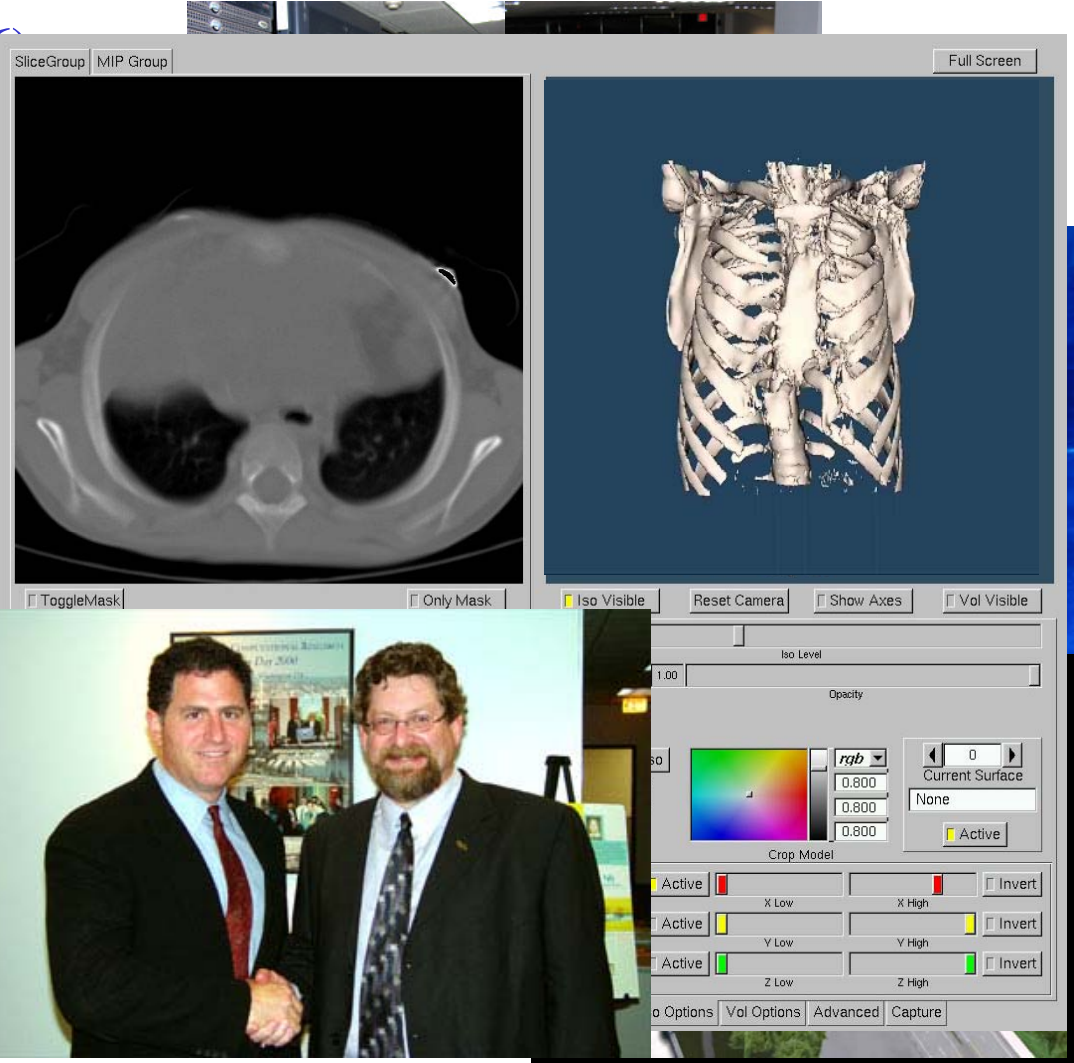
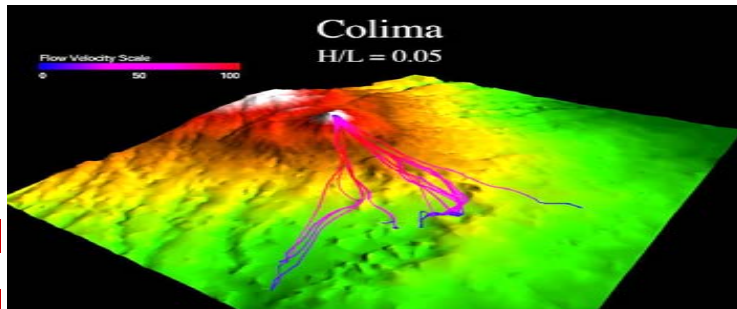
Center for Computational Research (CCR)

■ Founding Director (1998-2006)

■ Peak Achievements

- ❑ Top Academic HPC Center in the US
- ❑ ~25 TF of HPC
- ❑ ~600 TB of High-End Storage
- ❑ Significant Visualization
- ❑ Special-Purpose Systems
- ❑ ~30 FTEs Staff
- ❑ 140 Projects Annually

■ EOT



MTV

Song: I'm OK (I Promise)
Band: Chemical Romance

IBC Digital & CCR Gaming Environment: Death Jr.

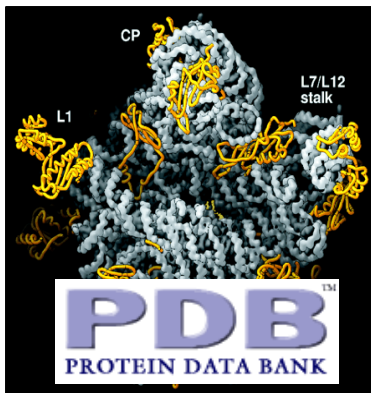
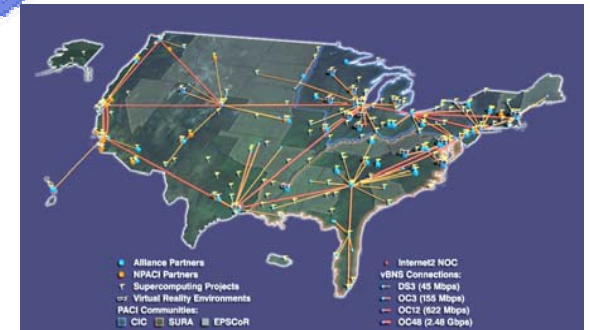
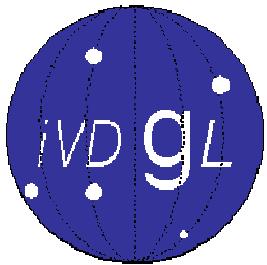


University at Buffalo *The State University of New York*

Cyberinfrastructure Laboratory

CI Lab

Grid Computing



Asia-Pacific Advanced Network



Advanced Center for Computational Research Data Center

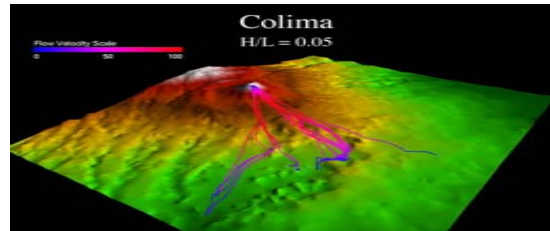


University at Buffalo The State University of New York

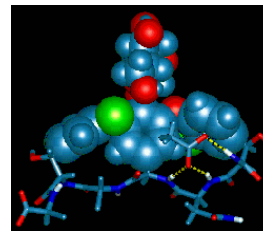
Cyberinfrastructure Laboratory

CI Lab

Grid Computing Overview



Data Acquisition



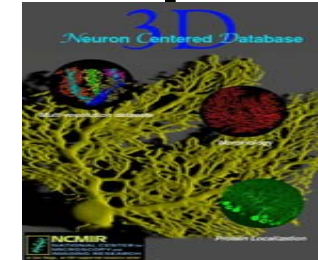
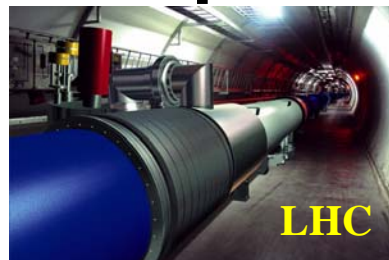
Advanced Visualization



Analysis



Imaging Instruments



Large-Scale Databases

- Coordinate Computing Resources, People, Instruments in Dynamic Geographically-Distributed Multi-Institutional Environment
- Treat Computing Resources like Commodities
 - ❑ Compute cycles, data storage, instruments
 - ❑ Human communication environments
- No Central Control; No Trust



"Middleware"

- **Intermediate Software Layer between Application Codes and Grid Resources**
- **Required for applications, users, and resource providers to operate effectively in a manner transparent to the user**
- **Security; Resource Management; Data Access; Policies; Accounting;**
- **Globus; Condor**
- **Checks availability of Resources**
 - **CPUs; Storage; Networking; Render Farms; etc.**
- **Scheduling / Workload Management System**
- **Resource Broker**
 - **Evaluates Job and Breaks Up/Submits**



Grid Issues

- **High-Throughput Computing**
- **Transparent Integration of Data, Computing, Sensors/Devices, Networking**
- **Heterogeneous Resources**
- **Standards (Grid, Data)**
- **Major User Communities**
 - **High-Energy Physics and Astrophysics**
 - **Medicine and Biological Sciences**
 - **Earth Sciences**
- **Public Funding Still Critical**
- **Grids are in their Infancy**



Major Grid Initiatives

- **EGEE: Enabling Grids for E-Science (European Commission)**
 - Initial Focus on CERN (5PB of Data/Year)
 - High-Energy Physics and Life Sciences
 - Expanded Focus Includes Virtually All Scientific Domains
 - 200 Institutions; 40 Countries
 - 20K+ CPUs; 5PB; 25,000 jobs per day!
- **OSG (DOE, NSF)**
 - High-Throughput Distributed Facility
 - Open & Heterogeneous
 - Biology, Computer Science, Astrophysics, LHC
 - 57 Compute Sites; 11 Storage Sites;
 - 10K CPUS; 6PB
- **TeraGrid (NSF)**
 - Integrates High-End Resources
 - High-Performance (Dedicated) Networks
 - 9 Sites; 100TF & 15PB
 - 100+ Databases Available

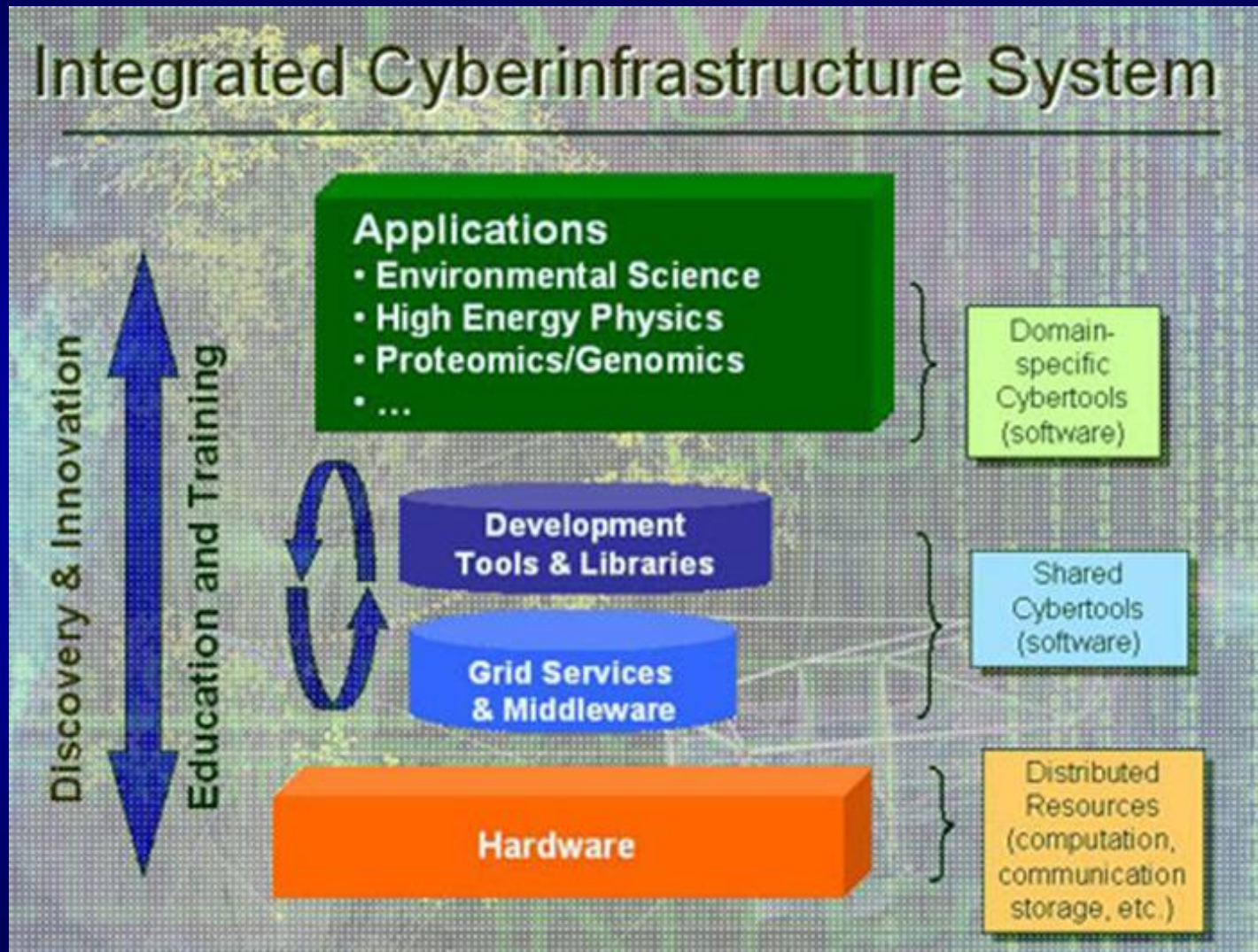


Cyberinfrastructure

- **Foster & Kesselman: “a domain-independent computational infrastructure designed to support science.”**
- **NSF: “comprehensive phenomenon that involves creation, dissemination, preservation, and application of knowledge”**
- **Generic: transparent and ubiquitous application of technologies central to contemporary engineering and science**
- **NSF Cyberinfrastructure (OCI)**
 - **HPC Hardware and Software**
 - **Data Collections**
 - **Science Gateways/Virtual Organizations**
 - **Support of Next Generation Observing Systems**



NSF Integrated Cyberinfrastructure



NSF Director Arden L. Bement: "leadership in cyberinfrastructure may determine America's continued ability to innovate – and thus our ability to compete successfully in the global arena."


Home Page - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.cse.buffalo.edu/faculty/miller/CI/

Mail Home My Netscape

New Tab Home Page



Cyberinfrastructure Laboratory

Ubiquitous High-End Computing, Data, Networking, & Visualization

Dr. Russ Miller
UB Distinguished Professor of Computer Science & Engineering

CI Lab

[Biography](#)

[Research](#)

- ◆ [Overview](#)
- ◆ [Papers](#)
- ◆ [Presentations](#)
- ◆ [CI Lab](#)
 - ◇ [Overview](#)
 - ◇ [Projects](#)
 - ◇ [News](#)
 - ◇ [Personnel](#)
 - ◇ [NYS Grid](#)
 - ◇ [Collaborations](#)
 - ◇ [Monitor](#)
 - ◇ [Portal](#)
- ◆ [SnB](#)
- ◆ [NYSGrid.org](#)
- ◆ [CCR](#)

[Teaching](#)

[Personal Info](#)

[Utilities](#)

[Contact Info](#)

[Home](#)

Introduction

The Cyberinstitute of the State of New York (CSNY) was approved in June of 2006 by the Center of Excellence in Bioinformatics and Life Sciences. Delays in the announcement of CSNY and its establishment have led to the creation of the CyberInfrastructure lab (CI). CI projects are currently supported by an NSF ITR grant, an NSF CRI grant, and the Center for Computational Research.

In the 21st century, leading academic institutions will embrace our digital data-driven society and empower students to compete in this knowledge-based economy. In order to support research, scholarship, education, and community outreach, CI has been established to integrate research in disciplinary domains, including science, engineering, and biomedicine, with research in enabling technologies and interfaces. This will allow students and scientists to transparently collect, manage, organize, analyze, and visualize data without having to worry about details such as where the data is stored, where the data is processed, where the data is rendered, and so forth. This ease of use and high availability of data and information processing tools will allow for revolutionary advances in all areas of science, engineering, and beyond.

Cyberinfrastructure sits at the core of modern simulation and modeling, which allows for entirely new methods of investigation that allow scholars to address previously unsolvable problems. Specifically, the development of necessary software, algorithms, portals, and interfaces that will enable research and

Cyberinfrastructure Lab

- **CI sits at core of modern simulation & modeling**
- **CI allows for new methods of investigation to address previously unsolvable problems**
- **Focus on development of**
 - algorithms
 - portals
 - interfaces
 - middleware
- **Free end-users to do disciplinary work**
- **Funding (2001-pres): NSF ITR, NSF CRI, NSF MRI, NYS, Fed**



Evolution of CI Lab Projects

■ ACDC-Grid

- ❑ Experimental Grid: Globus & Condor
- ❑ Integrate Data & Compute, Monitor, Portal, Node Swapping, Predictive Scheduling/Resource Management
- ❑ GRASE VO: Structural Biology, Groundwater Modeling, Earthquake Eng, Comp Chemistry, GIS/BioHazards
- ❑ Buffalo, Buffalo State, Canisius, Hauptman-Woodward

■ WNY Grid

- ❑ Heterogeneous System: Hardware, Networking, Utilization
- ❑ Buffalo, Geneseo, Hauptman-Woodward, Niagara

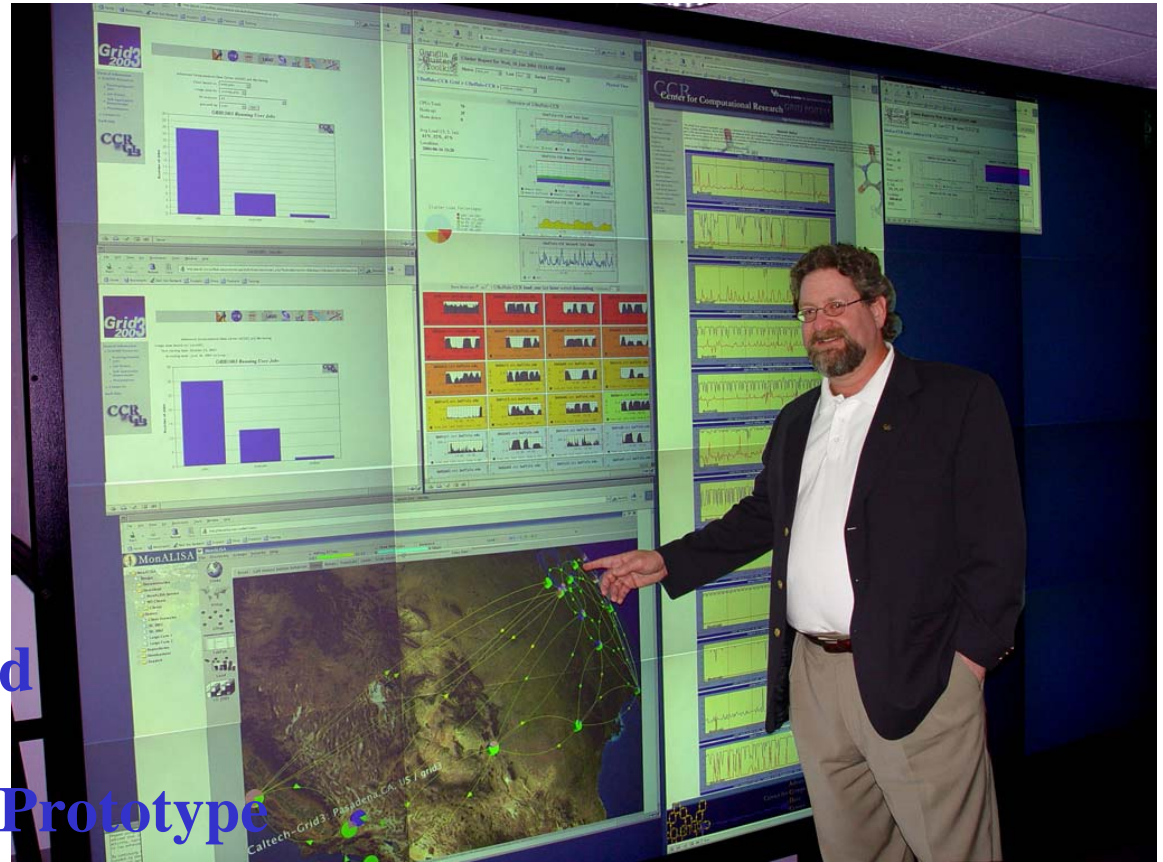
■ NYS Grid

- ❑ Extension to Hardened Production-Level System State-Wide
- ❑ Albany, Binghamton, Buffalo, Geneseo, Canisius, Columbia, HWI, Niagara, [Cornell, NYU, RIT, Rochester, Syracuse, Marist], {Stony Brook, RPI, Iona}



Grid Collaborations

- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
 - Only External Member
- Open Science Grid
 - GRASE VO
- NYSGrid.org
 - NYS CI Initiative
 - Executive Director
 - Various WGs
- Grid-Lite: Campus Grid
 - HP Labs Collaboration
- Innovative Laboratory Prototype
 - Dell Collaboration

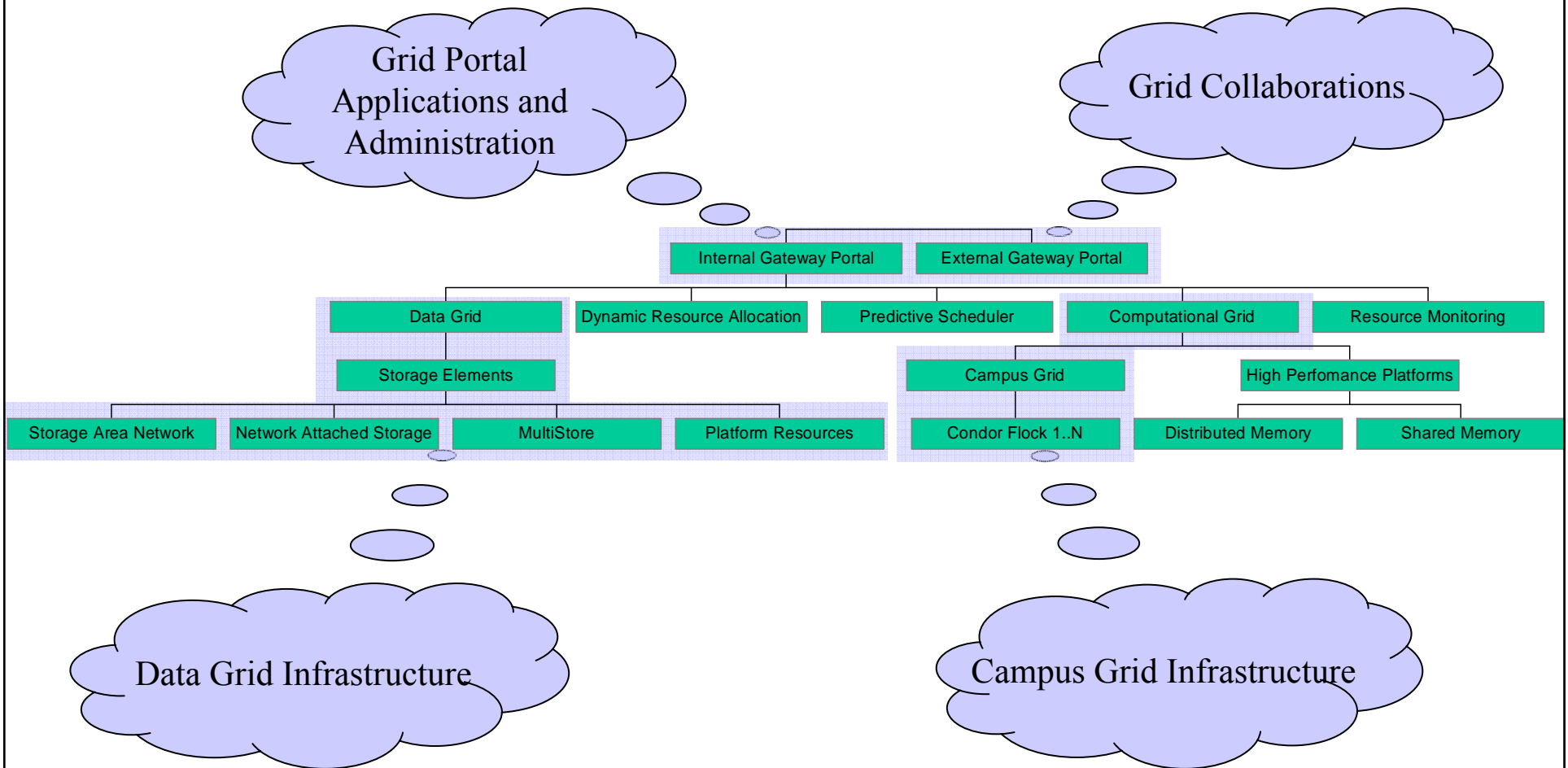


CI Lab Projects

- **Lightweight Grid Monitor (Dashboard)**
- **Predictive Scheduler**
 - Define quality of service estimates of job completion, by better estimating job runtimes by profiling users.
- **Dynamic Resource Allocation**
 - Develop automated procedures for dynamic computational resource allocation.
- **High-Performance Grid-Enabled Data Repositories**
 - Develop automated procedures for dynamic data repository creation and deletion.
- **Integrated Data Grid**
 - Automated Data File Migration based on profiling users.
- **Grid Portal**



ACDC-Grid System Architecture

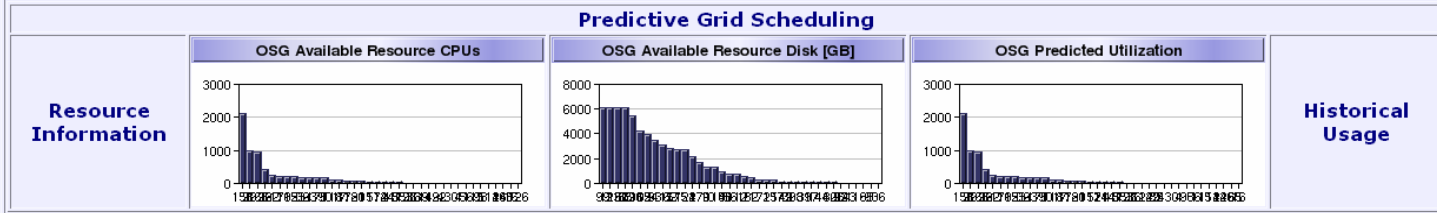
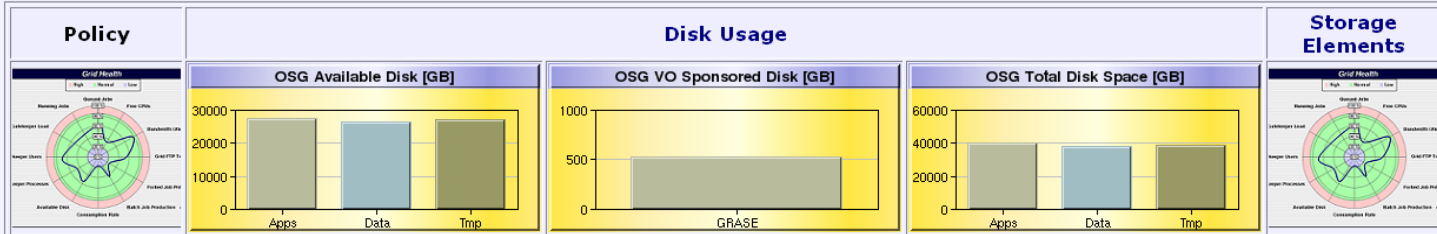
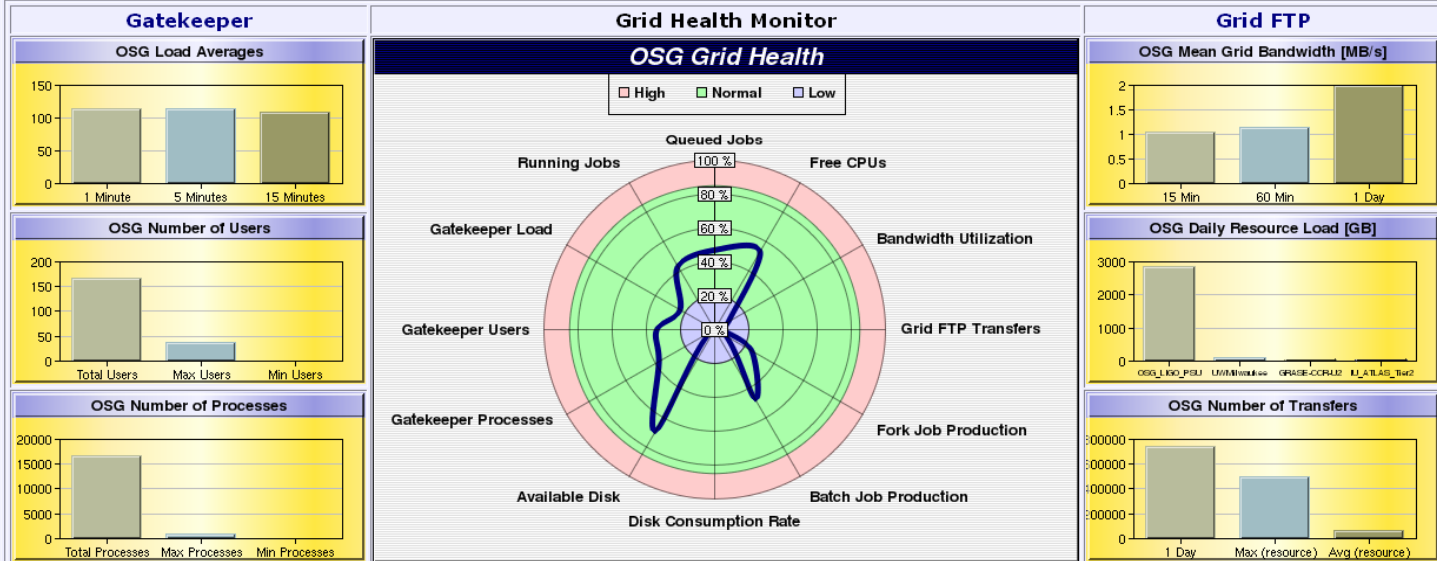
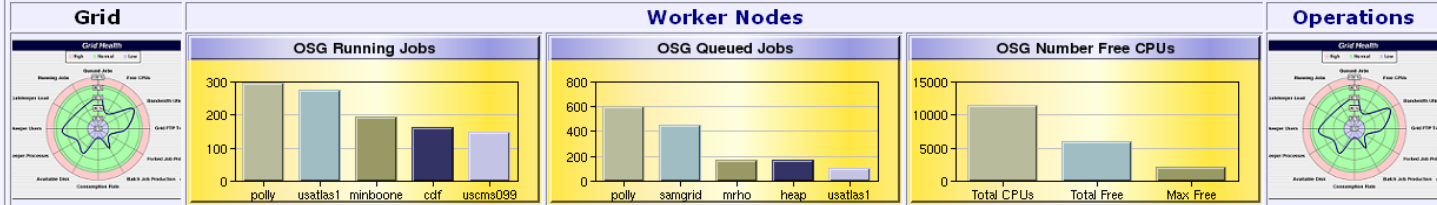


CI Lab

- CI Lab Grid Monitor
- Grid Dashboard
- Operations Dashboard
- Historical Dashboard
- Running/Queued Jobs
- Job History
- Detailed Job History
- VO Sponsor CPUs
- Free/Running/Queued CPUs
- VO Support Matrix
- Current Bandwidth Matrix
- Historical Bandwidth Matrix
- Current Latency Matrix
- Historical Latency Matrix
- Resource Queue Visualization
- Resource User Visualization
- SnB Application Demonstrator
- ACDC Grid Dashboard Site Status
- ACDC Grid Dashboard Tutorial
- GRASE VO
 - Overview
 - Request Membership
 - Request Help
- Staff Only
- Contact Us / Staff
- CI Lab

CI LAB GRID DASHBOARD

OSG-ITB OSG ACDC TeraGrid Ad-Hoc



Supported by the National Science Foundation and the Department of Energy

CI Lab Grid Monitor: <http://osg.ccr.buffalo.edu/>

Grid



CI LAB OPERATIONS DASHBOARD

OSG-ITB
 OSG
 ACDC
 TeraGrid
 Ad-Hoc

VIRTUAL ORGANIZATIONS

MIS
 GRASE
 OSG
 CDF
 Fermilab
 GADU
 GLOW
 USATLAS
 fMRI
 iVDGL
 nanoHUB

Version:
 All
 OSG-0.2.1
 OSG-0.3.0
 OSG-0.4.0
 OSG-0.4.1

Detailed Service Status

Operations



Site Resource - Service Matrix

Site Resource - Service Matrix

No Information
 Pass
 Error
 Fail
 Untested
 Excluded

Production Sites

| | Remote Host is Reachable | Running Gatekeeper | Authentication | Hello, World Application Check | Remote Host Uptime | Internet Network Services | Internet Servers Database | Certificate Expiration | Certificate Revocation | Gatekeeper Conf | Apps Directory Write | Data Directory Write | Temp Directory Write | CSIF-TP | CSIF-TP Local -> Remote | CSIF-TP Remote -> Local | Grid Tools | Grid Services | Scheduler | MDS | Grid / GLUE Attributes | Infrastructure Version | VDT Version | Grid3 Information | MonAJSA | Ganglia | VO-Specific Tests | |
|---------------------------|--------------------------|--------------------|----------------|--------------------------------|--------------------|---------------------------|---------------------------|------------------------|------------------------|-----------------|----------------------|----------------------|----------------------|---------|-------------------------|-------------------------|------------|---------------|-----------|------|------------------------|------------------------|-------------|-------------------|---------|---------|---------------------|--|
| athena.rit.albany.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:08 | |
| rommel.cs.binghamton.edu | Pass | Pass | Pass | Fail | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:45:52 | |
| gridgk01.racf.bnl.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:05 | |
| gridgk02.racf.bnl.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:44 | |
| idun.hwi.buffalo.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2007-01-22 14:08:32 | |
| u2-grid.ccr.buffalo.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:18 | |
| ctcnysgrid.tc.cornell.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:45:59 | |
| osgc01.grid.sinica.edu.tw | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:47:02 | |
| cms-xen2.fnal.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:50:12 | |
| cmsosgce.fnal.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:52:45 | |
| fngp-osg.fnal.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:55:13 | |
| tam01.fnal.gov | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:05 | |
| atlas.iu.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:50:19 | |
| nysgrid11.is.marist.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2007-01-22 15:28:04 | |
| bench.es.its.nyu.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:50 | |
| ouhep0.nhn.ou.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:47:20 | |
| grid3.aset.psu.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:50:39 | |
| grid.physics.purdue.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:14 | |
| osg.rcac.purdue.edu | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 14:07:04 | |
| stars.if.usp.br | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | Pass | 2006-12-13 13:46:33 | |

CI Lab Operations Dashboard

Predictive Scheduler

- **Build profiles based on statistical analysis of logs of past jobs**
 - Per User/Group
 - Per Resource
- **Use these profiles to predict runtimes of new jobs**
- **Make use of these predictions to determine**
 - Resources to be utilized
 - Availability of Backfill

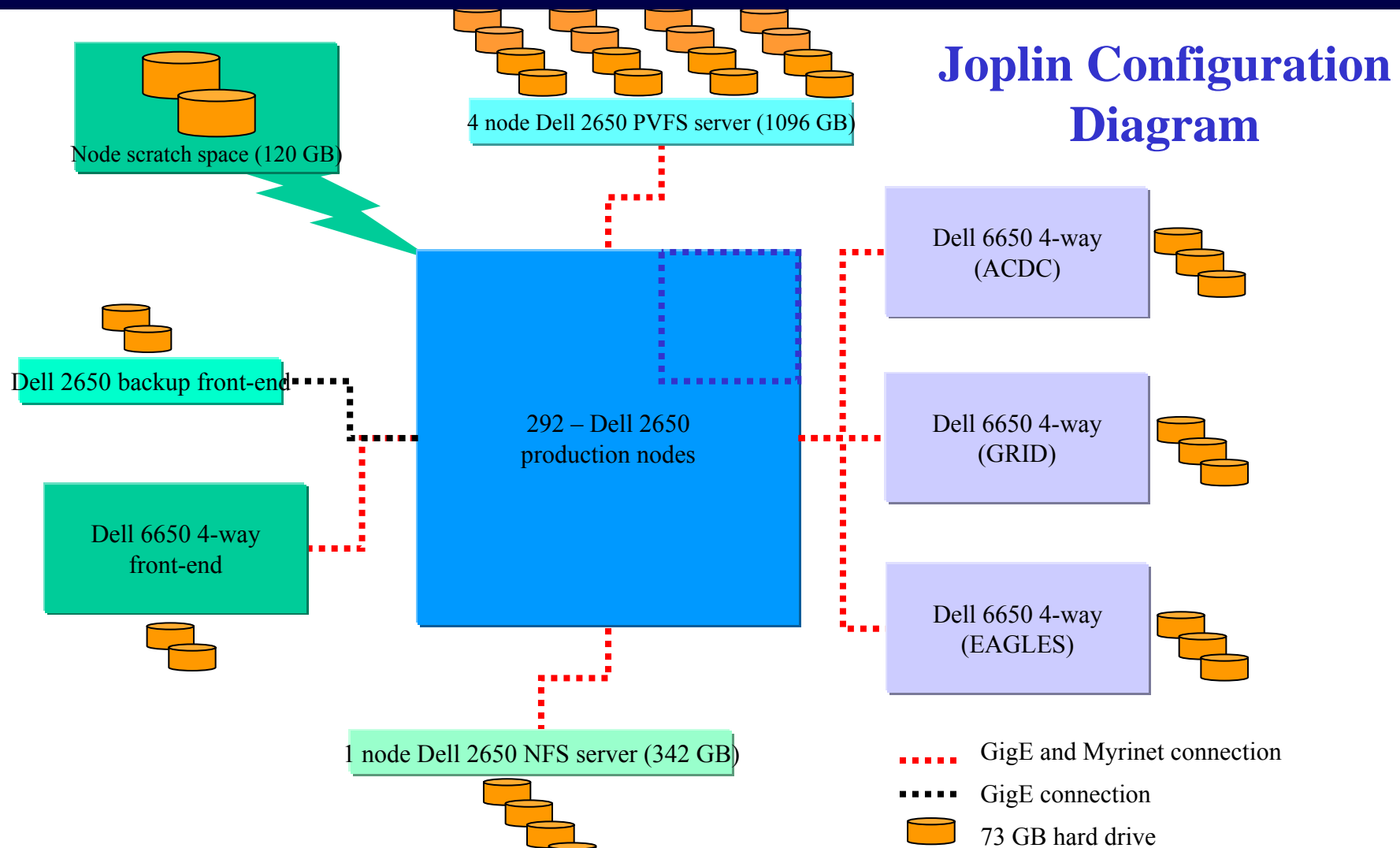


ACDC-Grid Dynamic Resource Allocation at SC03 with Grid3

- **Small number (40) of CPUs were dedicated at night**
- **An additional 400 CPUs were dynamically allocated during the day**
- **No human intervention was required**
- **Grid applications were able to utilize the resources and surpassed the Grid3 goals**



ACDC-Grid Dynamic Resource Allocation



Grid Administration

The image displays four overlapping screenshots of the CCR Grid Portal interface, each showing a different administrative function. The portal header for all screenshots includes the University at Buffalo logo and the text 'Center for Computational Research GRID PORTAL High Performance Grid Computing'.

Grid Site Administration (top-left screenshot): Shows a sidebar with 'PORTAL LOGOUT' and 'User Tools' (Manage Account, Grid General Info, Projects, Resources, Computational Grid, Job Submission, Job/Queue Status, Data Grid, Data Grid Statistics, Network Status, Running/Queued Jobs, PBS Job History, Grid Portal Statistics, Condor Flock Statistics, User Information, Education/Outreach, Staff Only, CCR HOME). The main content area is titled 'Grid Site Administration' and includes sections for 'Users' (Groups, Portal Event Log, Database Job List), 'Organizations' (add, edit, delete), 'Resources' (view, refresh, ping, delete, create host certificate), and 'Globus Administration' (Reports (machine usage, user access to machines, etc.)).

Generate Globus grid-mapfile (top-right screenshot): Shows a form for generating a Globus grid-mapfile. It includes fields for 'Optional include file:' (with value /home/griddev/www/grid-mapfile.inc) and 'Optional grid-mapfile path:'. A checkbox 'Do not stage the file to the grid nodes' is present. 'Generate' and 'Reset' buttons are at the bottom.

Create New Database Job (bottom-left screenshot): Shows a form for creating a new database job. It includes fields for 'Job Name:', 'Full Path To Script:', 'Accepts Arguments:' (set to No), 'Run Script:' (set to No), and 'Run As User:' (set to admin (Admin, Admin)). 'Create Job' and 'Reset' buttons are at the bottom.

MDS Resource Update Status (bottom-right screenshot): Shows a table of MDS Resource Update Status. The current time is 16-September-2003 10:58:12. The table has columns for Resource, Last Updated, Next Update, and Status.

| Resource | Last Updated | Next Update | Status |
|---------------------------|----------------------------|-------------|--------|
| crosby.ccr.buffalo.edu | 16-September-2003 09:15:30 | 2 minutes | OK |
| fogerty.ccr.buffalo.edu | 16-September-2003 10:45:30 | 2 minutes | OK |
| joplin.ccr.buffalo.edu | 16-September-2003 10:45:15 | 2 minutes | OK |
| mama.ccr.buffalo.edu | 16-September-2003 10:45:15 | 2 minutes | OK |
| nash.ccr.buffalo.edu | 16-September-2003 10:45:15 | 2 minutes | OK |
| nexus.hwi.buffalo.edu | 16-September-2003 10:45:20 | 2 minutes | OK |
| yardbirds.ccr.buffalo.edu | 16-September-2003 10:45:13 | 2 minutes | OK |
| young.ccr.buffalo.edu | 16-September-2003 10:45:27 | 2 minutes | OK |

Below the table, there are links: 'Return to the Grid Resource Admin menu.' and 'Return to the Grid Admin menu.'



ACDC-Grid Data Grid Functionality

- **Basic file management functions are accessible via a platform-independent web interface.**
- **User-friendly menus/interface.**
- **File Upload/Download to/from the Data Grid Portal.**
- **Simple Web-based file editor.**
- **Efficient search utility.**
- **Logical display of files (user/ group/ public).**
- **Ability to logically display files based on metadata (file name, size, modification date, etc.)**



Grid-Enabling Application Templates (GATs)

- **Structural Biology**
 - *SnB* and *BnP* for Molecular Structure Determination/Phasing
- **Groundwater Modeling**
 - *Ostrich*: Optimization and Parameter Estimation Tool
 - *POMGL*: Princeton Ocean Model Great Lakes for Hydrodynamic Circulation
 - *Split*: Modeling Groundwater Flow with Analytic Element Method
- **Earthquake Engineering**
 - *EADR*: Evolutionary Aseismic Design and Retrofit; Passive Energy Dissipation System for Designing Earthquake Resilient Structures
- **Computational Chemistry**
 - *Q-Chem*: Quantum Chemistry Package
- **Geographic Information Systems & BioHazards**
 - *Titan*: Computational Modeling of Hazardous Geophysical Mass Flows



Grid Enabled *SnB*

■ Required Layered Grid Services

□ Grid-enabled Application Layer

- *Shake – and – Bake* application
- Apache web server
- MySQL database

□ High-level Service Layer

- Globus, NWS, PHP, Fortran, and C

□ Core Service Layer

- Metacomputing Directory Service, Globus Security Interface, GRAM, GASS

□ Local Service Layer

- Condor, MPI, PBS, Maui, WINNT, IRIX, Solaris, RedHat Linux



Required Grid Services

■ Application Layer

- *Shake-and-Bake*
- Apache web server
- MySQL database

■ High-level Services

- Globus, PHP, Fortran, C

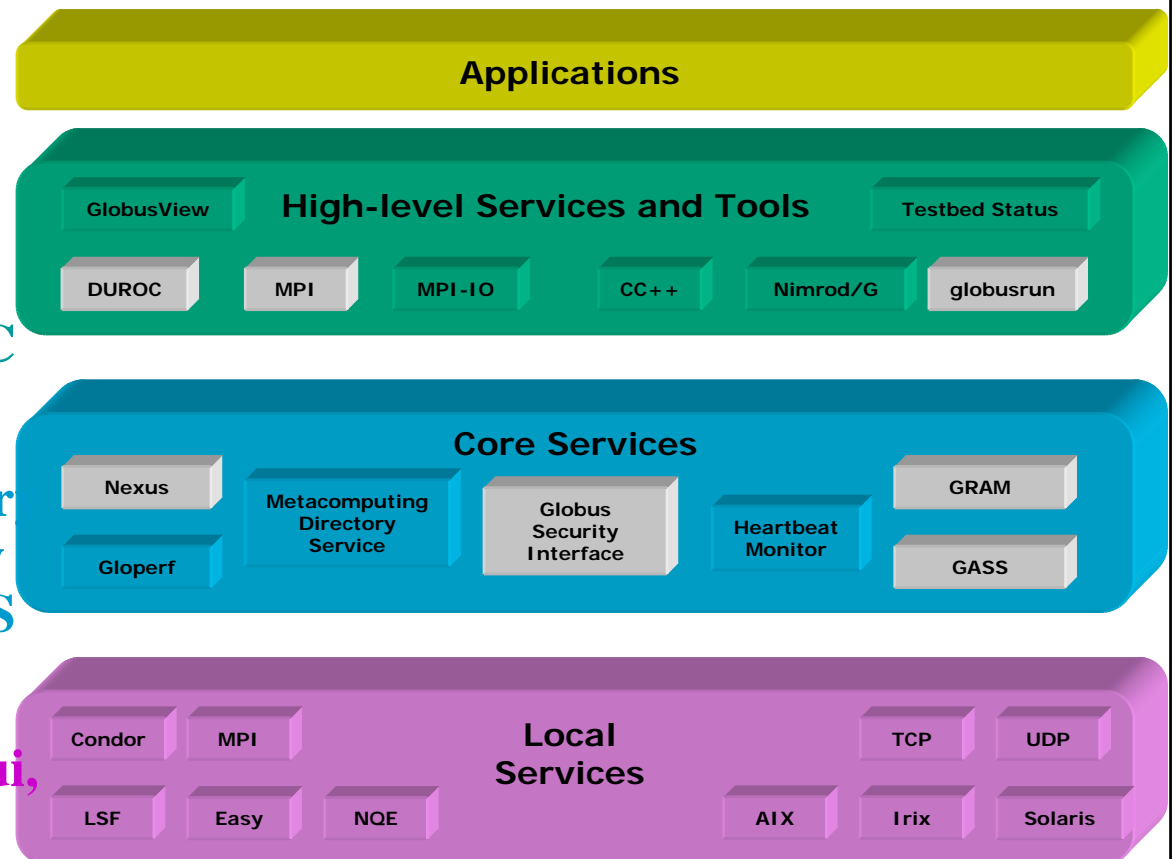
■ Core Services

- Metacomputing Director Service, Globus Security Interface, GRAM, GASS

■ Local Services

- Condor, MPI, PBS, Maui, WINNT, IRIX, Solaris, RedHat Linux


Grid Implementation as a Layered Set of Services



https://grid.ccr.buffalo.edu/

Mail Home My Netscape

New Tab CCR Grid Computing Services:



Cyberinfrastructure Laboratory

Grid Portal

Dr. Russ Miller
UB Distinguished Professor of Computer Science & Engineering

CI Lab

- Grid Portal Info
 - Overview
 - Portal Login
 - Grid Account Info
- Computational Grid
 - Job Submission
 - Job/Queue Status
 - MDS Information
 - Network Status
 - Running/Queued Jobs
 - PBS Job History
 - Condor Flock Statistics
 - GAT/Resource Matrix
- Data Grid
 - Data Grid Tree
 - Data Grid Upload
 - Data Grid Download
 - Data Grid File Manager
 - Data Grid Replica Manager
 - Data Grid Simulator
 - Data Grid Admin Tools
 - Data Grid Admin File Tools
- Contact Us / Staff
 - CI Lab
 - Staff Only

Welcome to the Cyberinfrastructure Laboratory Grid Portal

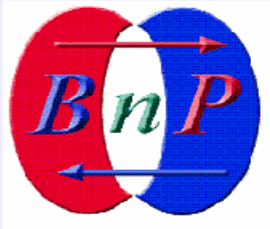
The **Cyberinfrastructure Laboratory**, in conjunction with the **Center for Computational Research**, has created an integrated Data and Computational Grid. This site is devoted to a Grid Portal that provides access to applications that can be run on a variety of grids. A related site contains a **Grid Monitoring System** designed by the Cyberinfrastructure Laboratory.

Applications may be run on the Cyberinfrastructure Laboratory's **ACDC Grid**, **Western New York Grid**, and **New York State Grid**, which includes computational and data storage systems from dozens of institutions throughout the State of New York.

The applications available to the users cover a variety of disciplines, including Bioinformatics, Computational Chemistry, Crystallography and Medical Imaging, to name a few.

The grids developed by the CI Lab support teaching and research activities, as well as providing infrastructure that includes high-end data, computing, imaging, grid-enabled software, all of which relies on the New York State Research Network (**NYSERNet**).

This work is funded by the National Science Foundation (ITR, MRI, CRI), three program projects from The National Institutes of Health, and the Department of Energy.



Software : BnP
Field : Protein crystal structure determination

Startup Screen for CI Lab Grid Job Submission

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Software → Template → **General Information** → Detailed Information → Job Definition → Review → Execution Scenario

Advanced Computational Data Center Grid Job Submission Instructions

The grid-enabling application templates used on the ACDC-Grid are created from the application developers grid user profiles that contain the users standard information uid, name, organization, address, etc., and more specific information such as group id and access level information for each of grid-enabled applications. This information is stored in a database for each of the grid-enabled applications and can be accessed through selected queries throughout the ACDC-Grid Web Portal.

Additionally, each grid-enabled scientific application profile contains information about specific execution parameters, required data files, optional data files, computational requirements, etc. and statistics on application historical ACDC-Grid jobs for predictive runtime estimates. MySQL provides the speed and reliability required for this task and it is currently being used as the ACDC-Grid Web Portal database provider.

The grid-enabled versions of many well-defined scientific and engineering applications have very similar general requirements and core functionality that are require for execution in the ACDC-Grid environment. We have identified that sequentially defining milestones for the grid user to complete intuitively guides them through the application workflow.

- Software Application:** Grid user chooses a grid-enabled software application.
- Template:** Grid user selects the required and/or optional data files from the ACDC Data Grid. User defined computational requirements are input or a template defined computational requirement runtime estimate is selected.
- Job Definition:** Grid user defines application specific runtime parameters or accepts default template parameter definitions.
- Review:** Grid user accepts the template complete job definition workflow or corrects any part of job definition.
- Execution Scenario:** The grid user has the ability to input an execution scenario or select a ACDC-Grid determined template defined execution scenario.
- Grid Job Status:** The grid user can view specific grid job completion status, grid job current state (COMPLETE, RUNNING, QUEUED, BLOCKED, FAILED, ETC.), detailed information on all running or queued grid jobs and grid-enabled application specific intermediate and post processing grid job graphics, plots and tables.

Each item of the job definition workflow is then stored in the ACDC-Grid Web Portal database so the grid user may use/modify any previously created workflow in creating new job definitions. The job definitions can also be accessed via batch script files for executing hundreds of similar workflows in an automated fashion. For example, a grid user would first define/save a relatively generic job workflow template for the grid-enabled application and then use the batch script capabilities to change the job definition workflow data files or application parameters and execute a series of new grid jobs.

Continue

Reset Sequence

Reset Current Stage

Cancel

Instructions and Description for Running a Job on ACDC-Grid

Expand All Collapse All
PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics
Data Grid
Education/Outreach
Staff Only
CCR HOME
Printer Friendly

Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

Select a GAT: BnP Auto Run
BnP Auto Run
EADR
Ostrich
POM
Q-Chem
SnB
SnB DREAR
Split
snb-dev

Continue Reset Current Stage Cancel

Return to the

Software Package Selection

Expand All Collapse All
PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics
Data Grid
Education/Outreach
Staff Only
CCR HOME
Printer Friendly

Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

Enter structure definition manually
 Select structure from Data Grid: Select Config File

Continue Reset Sequence Reset Current Stage Cancel

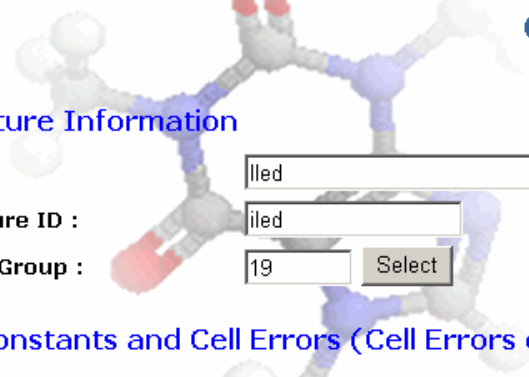
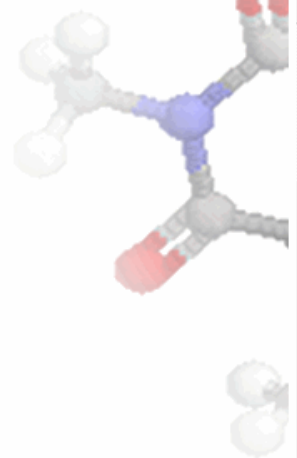
Return to the Grid Job Menu

Full Structure / Substructure Template Selection

USER TOOLS

- » Manage Account
- Grid General Info
- Projects
- Computational Grid
- » Job Submission
- » Job/Queue Status
- » MDS Information
- » Network Status
- » Running/Queued Jobs
- » PBS Job History
- » NYS Grid
- » Condor Flock Statistics
- Data Grid
- Education/Outreach
- Staff Only
- CCR HOME
- Printer Friendly

General Information

Structure Information

Title:

Structure ID :

Space Group :

Cell Constants and Cell Errors (Cell Errors optional)

A: +/-

B: +/-

C: +/-

Alpha: +/-

Beta: +/-

Gamma: +/-

Native Asymmetric Unit Contents

No Residues (Optional):

ASU Contents : (examples: C6H12O6 OR C6 H12 O6)

Initial Data Sets

| | |
|---------------------------------|--------------------------------------|
| Select dataset to delete | ○ |
| Datasets | Dataset 1 |
| Name (8 chars max): | <input type="text" value="iledhkl"/> |

Default Parameters Based on Template

User Tools
 » Manage Account
 Grid General Info
 Projects
 Computational Grid
 » Job Submission
 » Job/Queue Status
 » MDS Information
 » Network Status
 » Running/Queued Jobs
 » PBS Job History
 » NYS Grid
 » Condor Flock Statistics
 Data Grid
 Education/Outreach
 Staff Only
 CCR HOME
 Printer Friendly

SnB Job Review

| | |
|--|----------------------------|
| Grid Job ID: | 447 |
| Selected resource: | clearwater.ccr.buffalo.edu |
| Number of processors: | 5 |
| Wallclock time requested: | 720 |
| Number of triplet invariant to use: | 8400 |
| Start Phases From: | Random Atoms |
| Random seed (prime): | 11909 |
| Number of trials: | 1000 |
| Starting Trial: | 1 |
| Input Phase File: | Unused |
| Input Atom File: | Unused |
| Keep complete (every trial) peak file? : | Yes |
| Number of Shake-and-bake cycles: | 20 |
| Keep complete (every cycle) trace file? : | No |
| Terminate trials failing the R-Ratio test? : | No |
| R-Ratio cutoff: | Unused |
| Phase Refinement Method: | Parameter Shift(Fast) |
| Number of passes through phase set: | 3 |
| Phase shift: | 90.0 |
| Number of shifts: | 2 |
| Number of peaks to select: | 84 |
| Minimum interpeak distance: | 3 |
| Minimum distance between symmetry-related peaks: | 3.0 |
| Number of special position peaks to keep: | 0 |
| Fourier grid size: | 0.31 |
| Perform extra cycles with more peaks? : | No |
| Number of extra cycles: | Unused |
| Number of peaks: | Unused |
| Trials for E-Fourier filtering (fourier refinement)? : | None |
| Number of cycles: | Unused |
| Number of peaks: | Unused |
| Minimum E : | Unused |

SnB Review (Grid job ID: 447)

Details for Grid Job 447 - iledhkl

Job Detail Information

Status: **RUNNING**

Rmin Min: 0.344 Rmin Max: 0.56

Last Updated: 15-Mar-2005 10:22:00

Total Trials: 1000

Complete Trials: 285

Resource: clearwater.ccr.buffalo.edu Processors: 5

Best Trial Number: 34

Best Trial Rmin: 0.344

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Trial Summary

Grid Job 447 Trial Summary

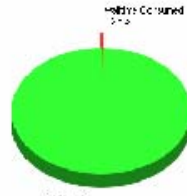
Number of Trials Complete: 285 (28.5%)



Walltime Summary

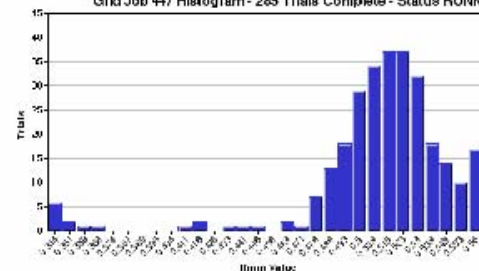
Grid Job 447 Walltime Summary

Walltime Consumed: 2 (0.3%)



Grid Job Trial Histogram

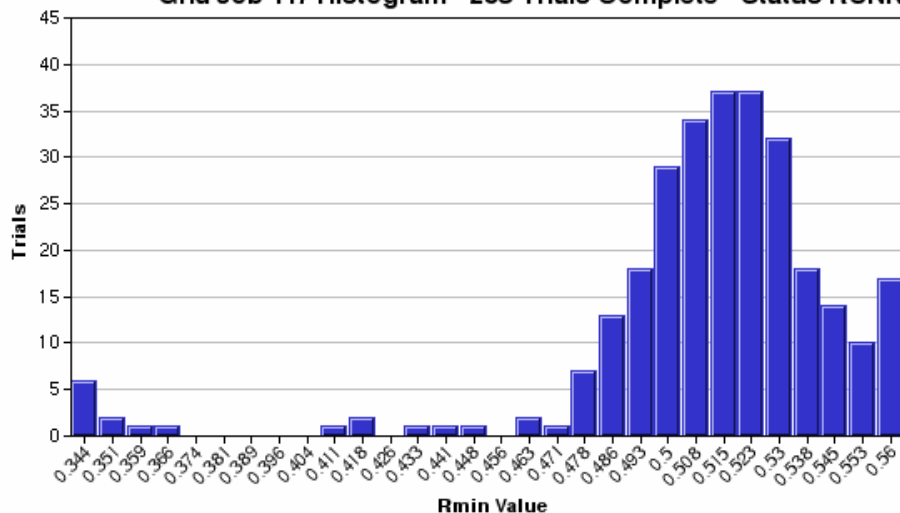
Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING



Click on image for enlarged view.

Graphical Representation of Intermediate Job Status

Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING



Histogram of Completed Trial Structures

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Grid Job Status

15-Mar-2005 10:23:49

Job Filter Criteria

Show GATs

BnP Auto Run
EADR
Ostrich
POM
Q-Chem
SnB
SnB DREAR

Job State

DEFINITION
STAGING
QUEUED
RUNNING
UPLOADING
COMPLETE
INCOMPLETE

Sort By

Job Id
Job Name
Resource
Num Procs
Status
Percent Complete
Last Update

Descending

Ascending

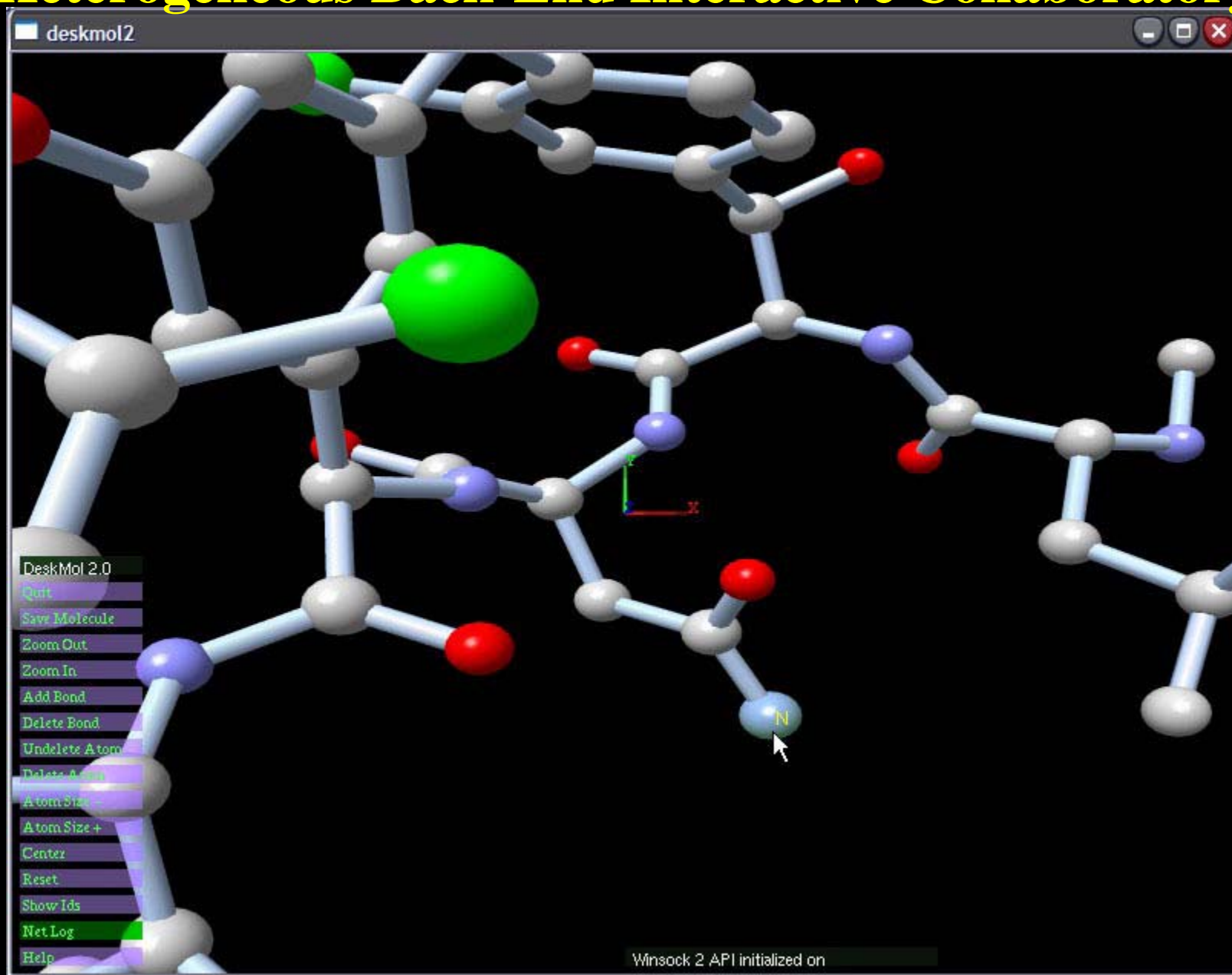
Filter Job List

SnB

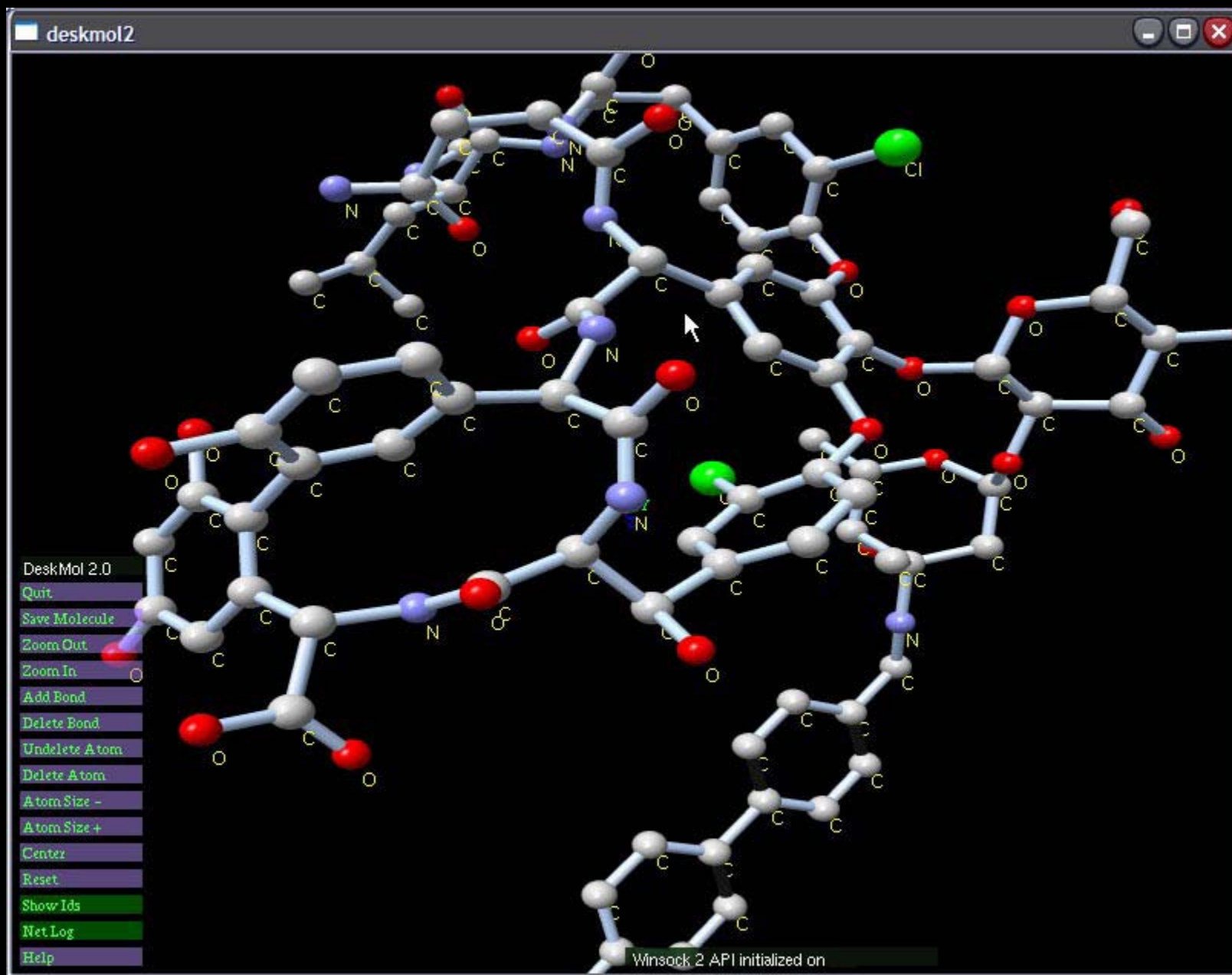
| Job Id | Job Name | Resource | Num Procs | Status | Percent Complete | Last Update | Cancel Job | Drilldown |
|--------|----------|----------------------------|-----------|-----------------|------------------|----------------------|--------------------------|-----------|
| 447 | iledhkl | clearwater.ccr.buffalo.edu | 5 | RUNNING | 28.5 | 15-Mar-2005 10:22:00 | <input type="checkbox"/> | |
| 446 | trilys | clearwater.ccr.buffalo.edu | 10 | RUNNING | 1 | 15-Mar-2005 10:22:00 | <input type="checkbox"/> | |
| 444 | 64chkl | nash.ccr.buffalo.edu | 3 | COMPLETE | 100 | 14-Mar-2005 22:00:01 | | |
| 443 | trilys | clearwater.ccr.buffalo.edu | 10 | COMPLETE | 100 | 10-Mar-2005 22:48:00 | | |
| 442 | pr435hkl | nash.ccr.buffalo.edu | 5 | COMPLETE | 100 | 10-Mar-2005 17:26:01 | | |
| 441 | vancohkl | clearwater.ccr.buffalo.edu | 10 | COMPLETE | 100 | 10-Mar-2005 18:08:01 | | |
| 434 | 16chkl | clearwater.ccr.buffalo.edu | 5 | COMPLETE | 100 | 10-Mar-2005 14:42:01 | | |
| 433 | 16chkl | clearwater.ccr.buffalo.edu | 5 | COMPLETE | 100 | 10-Mar-2005 14:38:01 | | |

Status of Jobs

Heterogeneous Back-End Interactive Collaboratory



User starts up – default image of structure.



Molecule scaled, rotated, and labeled.

NYSGrid.org

- **Grass-Roots Cyberinfrastructure Initiative in NYS**
- **Open to academic, research, government, and industrial organizations.**
- **Goal is to allow transparent collection, management, organization, analysis, and visualization of data, while ignoring location.**
- **Enable Research, Scholarship, and Economic Development in NYS.**
- **Mission Stmt: To create and advance collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in NYS.**



NYSGrid: Home - Netscape

File Edit View Go Bookmarks Tools Window Help

http://www.nysgrid.org/main/index.maml

Mail Home My Netscape

New Tab NYSGrid: Home



New York State Grid

Cyberinfrastructure for the 21st Century

[Home](#)
[Contact Us](#)

[About Us](#) [Organization](#) [Infrastructure](#) [User Information](#) [Events](#) [Media & News](#) [Grid Access](#)

In the 21st century, leading academic institutions will embrace our digital data-driven society and empower students to compete in this knowledge-based economy. In order to support research, scholarship, education, and community outreach, a grass-roots cyberinfrastructure initiative has been formed in New York State that will integrate research in disciplinary domains, including science, engineering, and biomedicine, with research in enabling technologies and interfaces. This initiative will allow students and scientists to transparently collect, manage, organize, analyze, and visualize data without having to worry about details such as where the data is stored, where the data is processed, where the data is rendered, and so forth. This ease of use and high availability of data and information processing tools will allow for revolutionary advances in all areas of science, engineering, and beyond.

Mission:

To create an advanced collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in New York State.

Cyberinfrastructure sits at the core of modern simulation and modeling, which allows for entirely new methods of investigation that allow scholars to address previously unsolvable problems. Specifically, the development of necessary software, algorithms, portals, and interfaces that will enable research and scholarship by freeing end-users from dealing with the complexity of various computing environments is critical to extending the reach of high-end computing, storage, networking, and visualization to the general user community.

The Cyberinfrastructure Initiative consists of resources at institutions throughout the state. The initiative is open to all interested parties and more information can be found on some of the accompanying pages.

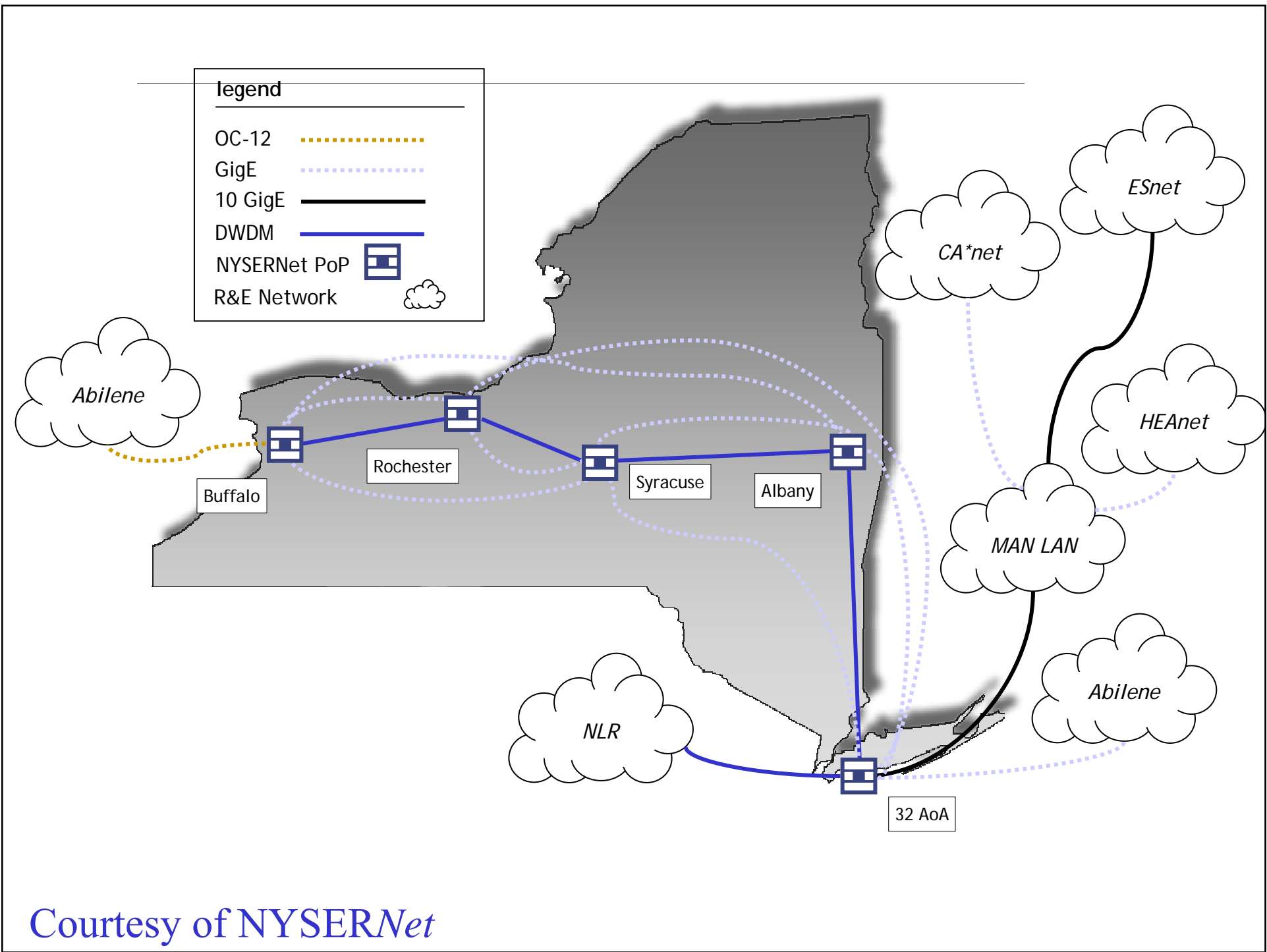
Current NYS Grid Participation

- Albany • ✓
- Alfred
- Binghamton • ✓
- Brookhaven
- Buffalo • ✓
- Columbia •
- Cornell • ✓
- Geneseo • ✓
- Hauptman-Woodward • ✓
- Iona •
- Marist • ✓
- Memorial Sloan-Kettering
- NYU • ✓
- Niagara • ✓
- RIT • ✓
- Rochester • ✓
- RPI •
- Stony Brook • ✓
- Syracuse • ✓
- NYSERNet

• - expressed interest in NYS Grid

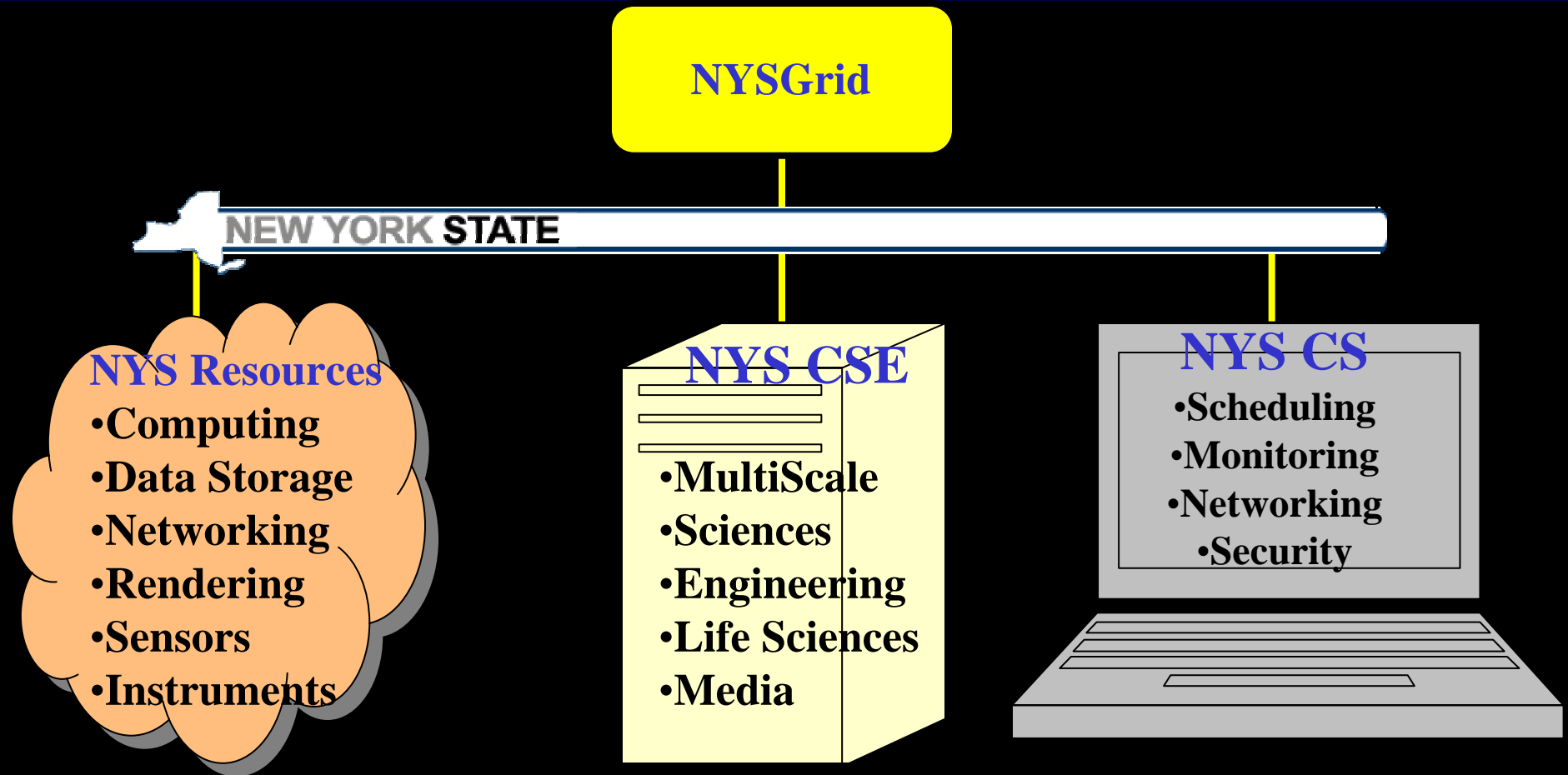
✓ - on NYS Grid





Courtesy of NYSErNet

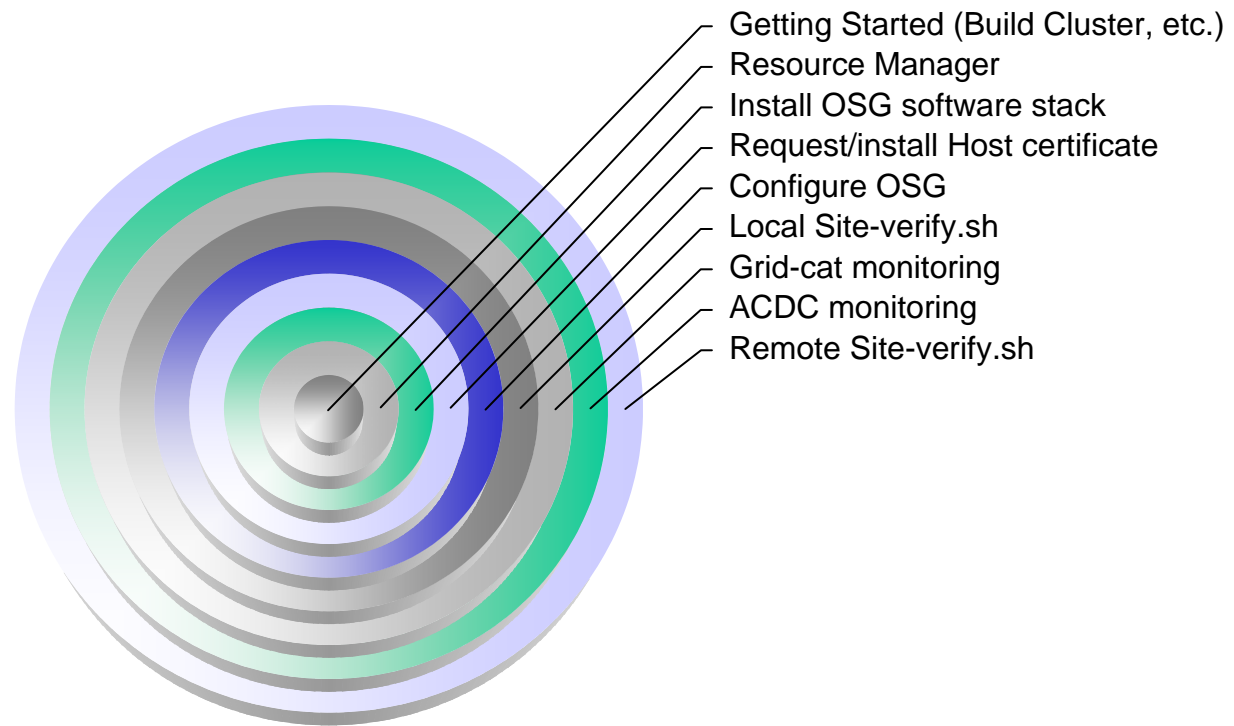
NYSGrid.org Organization



NSF Director Arden L. Bement: "leadership in cyberinfrastructure may determine America's continued ability to innovate – and thus our ability to compete successfully in the global arena."

NYS Grid Implementation Details

Figure Courtesy of Jon Bednasz, CCR/UB



Getting Started

(Courtesy of Jon Bednasz & Steve Gallo, CCR/UB)

- **Physically build a cluster**
 - ❑ 1 head node
 - ❑ 4+ compute nodes
- **Install Cluster Software**
 - ❑ Operating System (Red Hat)
 - ❑ Drivers for Interconnect (Myrinet, Infiniband, etc.)
 - ❑ Resource Manager (PBS, LSF, Condor, SGE)
- **Identify Gatekeeper Node for OSG Software**
 - ❑ Either stand alone machine or co-resident on Head Node
 - ❑ 5GB of space in /opt/grid
 - ❑ 5GB of space in /grid-tmp
- **Need to have ability to adjust firewalls**
- **Need to have ability to add users**



Installing OSG Stack on Gatekeeper

■ Installs are done via PACMAN

- ❑ `wget http://physics.bu.edu/pacman/sample_cache/tarballs/pacman-3.16.1.tar.gz`

■ Install OSG software

- ❑ `pacman -get OSG:ce`

■ Install (1) Package for your Resource Manager

- ❑ `pacman -get OSG:Globus-Condor-Setup`

- ❑ `pacman -get OSG:Globus-PBS-Setup`

- ❑ `pacman -get OSG:Globus-LSF-Setup`

- ❑ `pacman -get OSG:Globus-SGE-Setup`



NYSGrid.org Technical Group

- Jon Bednasz, Buffalo, Chair
- Steve Gallo, Buffalo
- Eric Warnke, Albany
- Steaphan Greene, Binghamton
- Ken Smith, Columbia
- Resa Alvord, Cornell
- Kirk Anne, Geneseo
- Steve Potter, Hauptman-Woodward
- Robert Schiaffino, Iona
- Earle Nietzel, Marist
- Ann Rensel, Niagara
- Chris Grim, NYU
- Rick Bohn, RIT
- Bill Webster, Rochester
- Lindsay Todd, RPI
- Ajay Gupta, Stony Brook
- Jorge González Outeiriño, Syracuse



NYSGrid.org Activities & Board

Activities

- Technical Working Group
- Middleware
- User Support and Services / EOT
- Communications
- Infrastructure
- Resource Providers
- Funding

Board

- Russ Miller
- Gurcharan Khanna
- Linda Callahan
- Mark Shephard
- Tim Lance
- (Heather Stewart)
- Jim Davenport
- Chris Haile



Technical WG Current Efforts

(Led by Steve Gallo and Jon Bednasz)

- **NYS Grid is Available**
- **CCR/UB & CTC/Cornell**
 - **Streamline bringing users onto NYS Grid**
 - **Documentation**
 - **Recommendations**
- **Need Early Adopters**
 - 1. Current Grid Users**
 - 2. New Users to Grid with HPC Needs**



Middleware WG Current Efforts

- **Discussions on current state of Middleware at Buffalo, Binghamton, & RPI**
 - **Scheduling**
 - **Portals**
 - **Monitoring**
 - **Fault Tolerance**
 - **Checkpoint/Restart**



Acknowledgments

- Mark Green
 - Cathy Ruby
 - Amin Ghadersohi
 - Naimesh Shah
 - Steve Gallo
 - Jason Rappleye
 - Jon Bednasz
 - Sam Guercio
 - Martins Innus
 - Cynthia Cornelius

 - George DeTitta
 - Herb Hauptman
 - Charles Weeks
 - Steve Potter
- Alan Rabideau
 - Igor Janckovic
 - Michael Sheridan
 - Abani Patra
 - Matt Jones

 - NSF ITR
 - NSF CRI
 - NSF MRI
 - NYS
 - CCR





www.cse.buffalo.edu/faculty/miller