

Miller's Cyberinfrastructure Laboratory (MCIL)

Russ Miller

Director, MCIL

Dept of Comp Sci & Eng, SUNY-Buffalo
Hauptman-Woodward Inst.



NSF, NIH, DOE, NIMA, NYS, HP
www.cse.buffalo.edu/faculty/miller/CI/



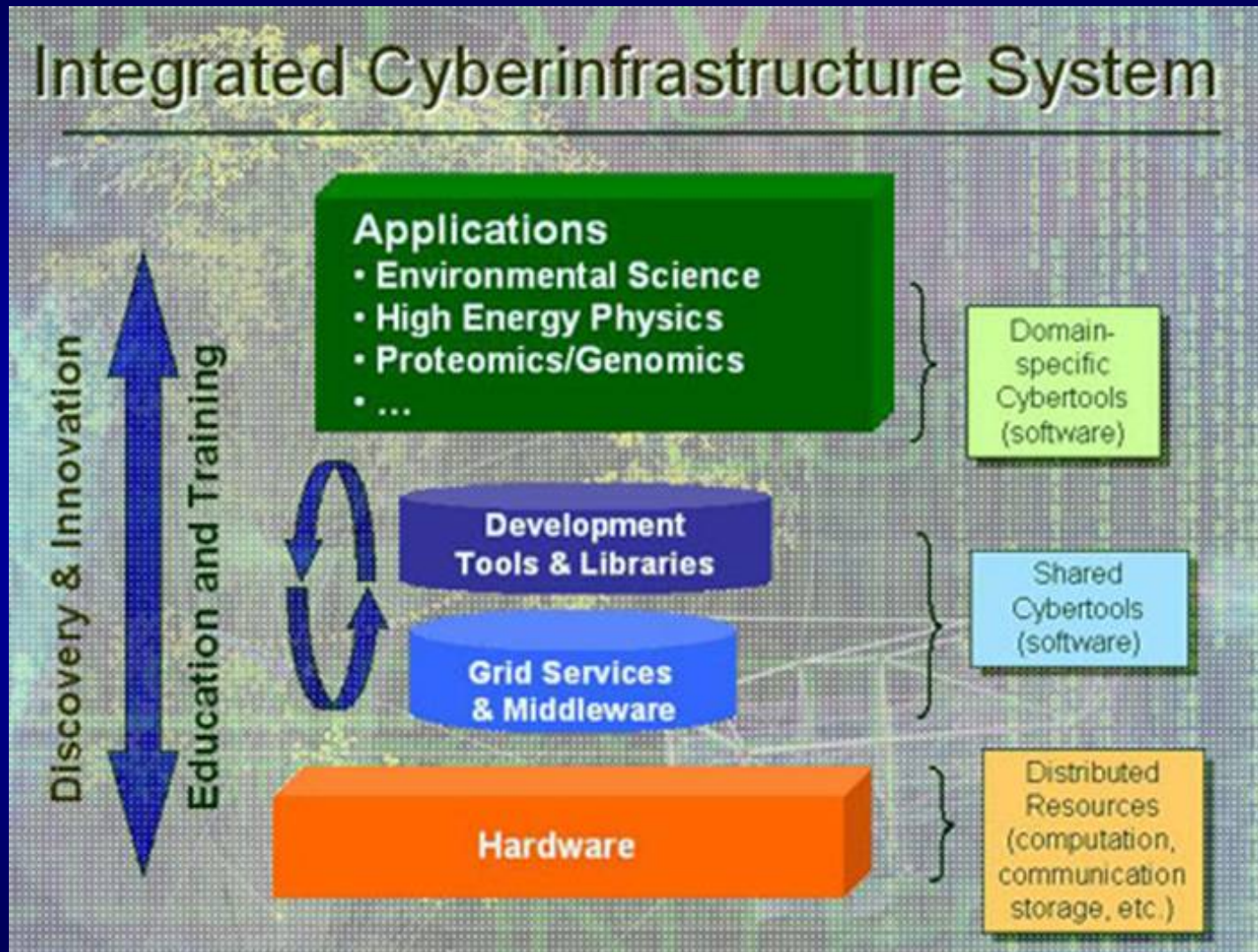
Advanced
Center for Computational Research
Data
Center

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**

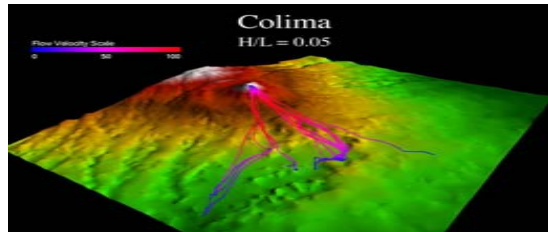


NSF Integrated Cyberinfrastructure

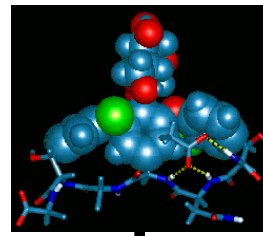


NSF Director Arden L. Bement: "leadership in cyberinfrastructure may determine America's continued ability to innovate – and thus our ability to compete successfully in the global arena."

Grid Computing Overview



Data Acquisition



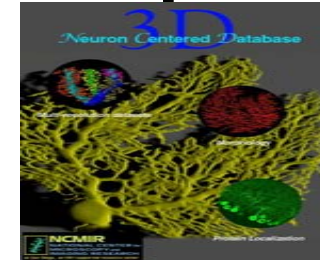
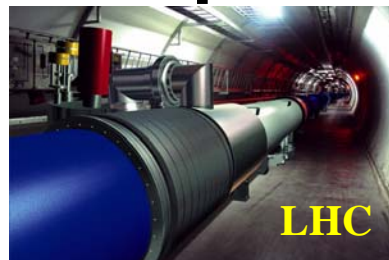
Advanced Visualization



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




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

CI Lab

- Biography
- Photos/Videos
- Media Coverage
- Research
 - Overview
 - Grids
- CI Lab
 - Overview
 - Equipment
 - Documentation
 - Projects
 - News
 - Publications
 - Personnel
 - WNY/NYS Grid
 - Collaborations
- Publications
- Presentations
- SnB
- CCR
- Teaching
- Personal Info
- Utilities
- Contact Info
- Home

Introduction

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, the Cyberinfrastructure Laboratory (CI Lab) is dedicated to the integration of research in disciplinary domains, including science, engineering, and biomedicine, with research in enabling technologies and interfaces. The goal is to 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 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.

Projects in the CI Lab are currently supported by an NSF ITR grant, an NSF CRI grant, and the Center for Computational

Miller's Cyberinfrastructure Lab

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



MCIL Equipment

■ Experimental Equipment (1.25 TF; 140 Cores; 22TB)

❑ Clusters

○ Head Nodes: Dell 1950 (Intel)

○ Workers: Intel 8×2 ×4 & AMD 8×2×2

❑ Virtual Memory Machines (2 × Intel 4×4)

❑ Dell GigE Managed Switches

❑ InfiniBand

❑ 22 TB Dell Storage (2)

❑ Condor Flock (35 Intel/AMD)

❑ *In process: 40-50TF system*

■ Production Equipment

❑ Dell Workstations; Dell 15 TB Storage

❑ Access to CCR equipment (13TF Dell/Intel clusters)



Evolution of CI Lab Projects

■ Buffalo-Based 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

■ Western New York Grid

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

■ New York State 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}



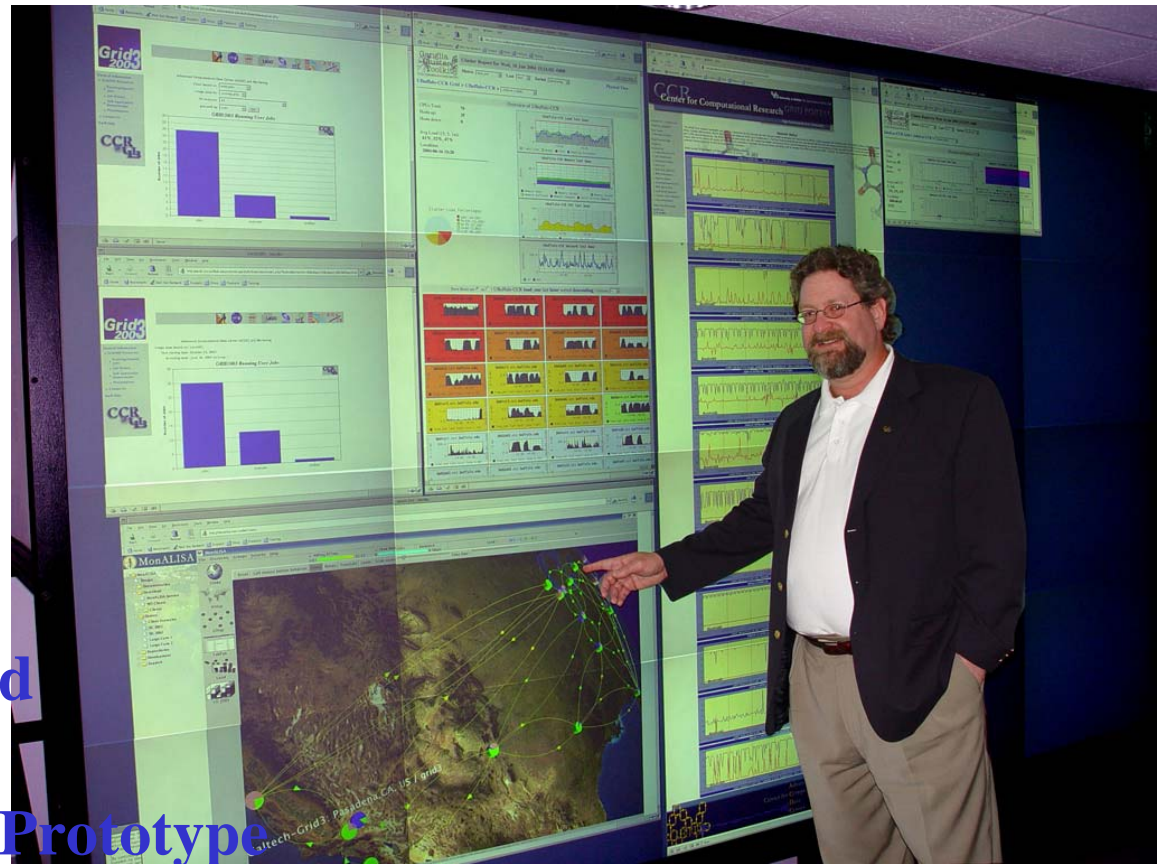
NYS Grid Resources

- Albany: 8 Dual-Processor Xeon Nodes
- Binghamton: 15 Dual-Processor Xeon Nodes
- Buffalo: 1050 Dual-Processor Xeon Nodes
- Cornell: 30 Dual-Processor Xeon Nodes
- Geneseo State: Sun/AMD with 128 Compute Cores
- Hauptman-Woodward Institute: 50 Dual-Core G5 Nodes
- Marist: 9 P4 Nodes
- Niagara University: 64 Dual-Processor Xeon Nodes
- NYU: 58 Dual-Processor PowerPC Nodes
- RIT: 4 Dual-Processor Xeon Nodes
- Syracuse: 8 Dual-Processor Xeon Nodes



CI Lab Collaborations

- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
 - Only External Member
- Open Science Grid
 - GRASE VO
- 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**

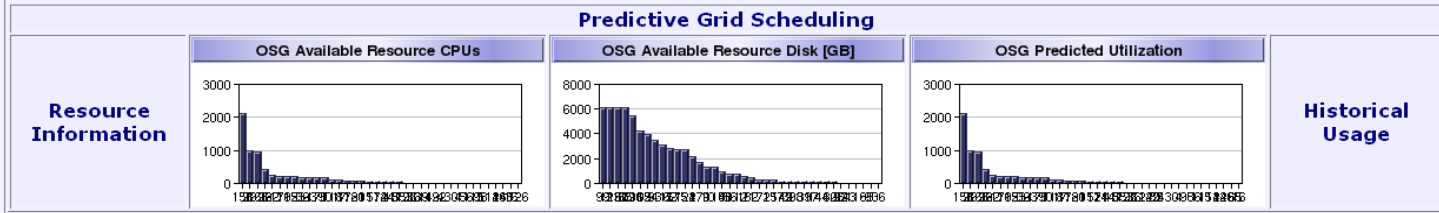
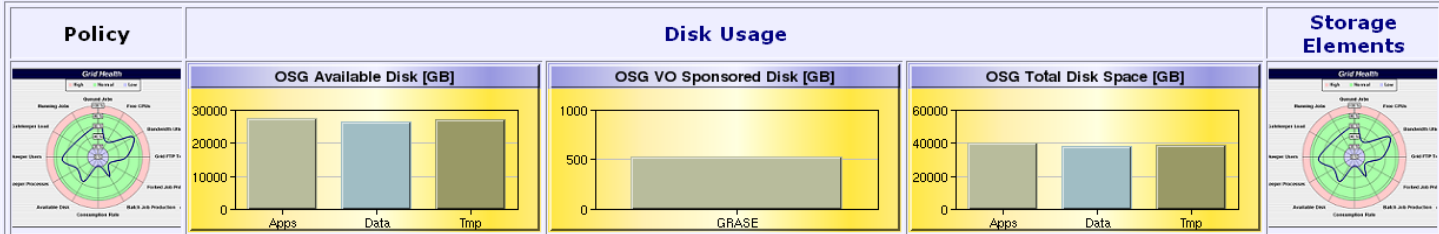
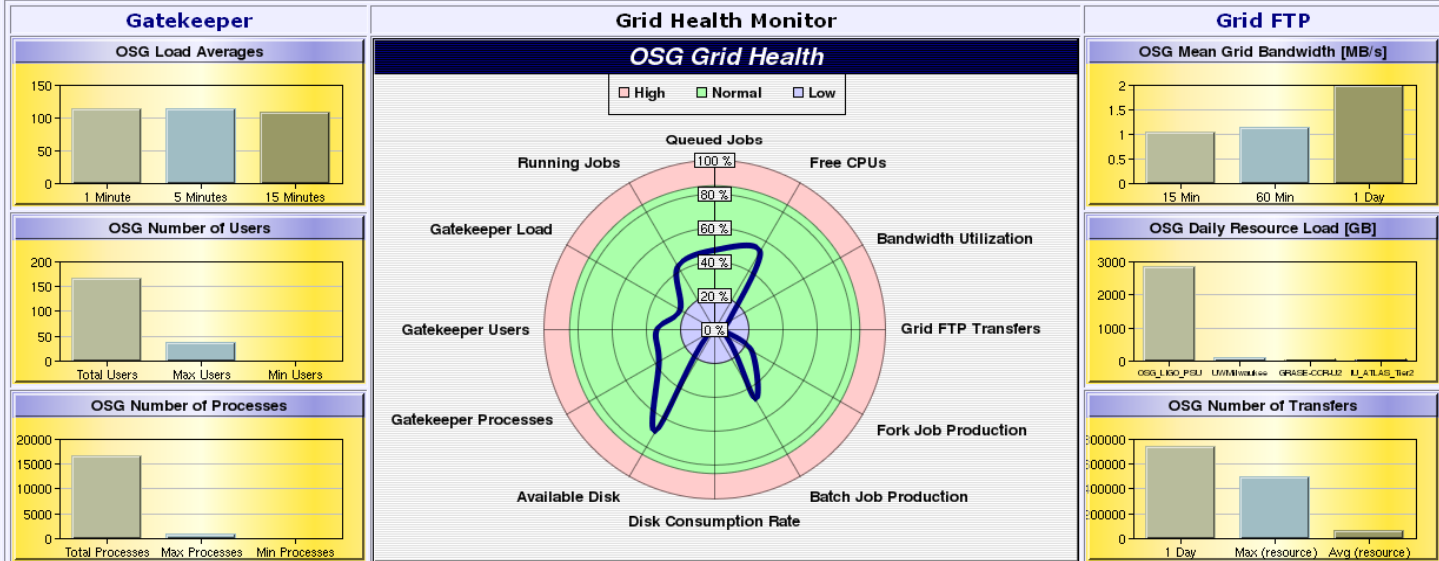
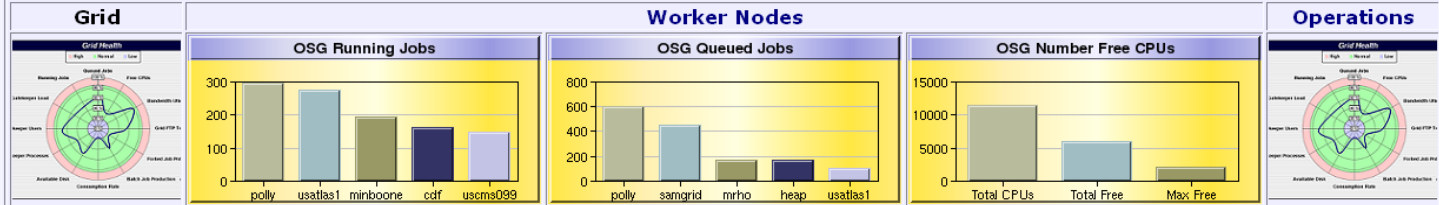


CI Lab

- 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

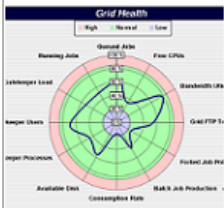
OSG-ITB OSG ACDC TeraGrid Ad-Hoc



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	

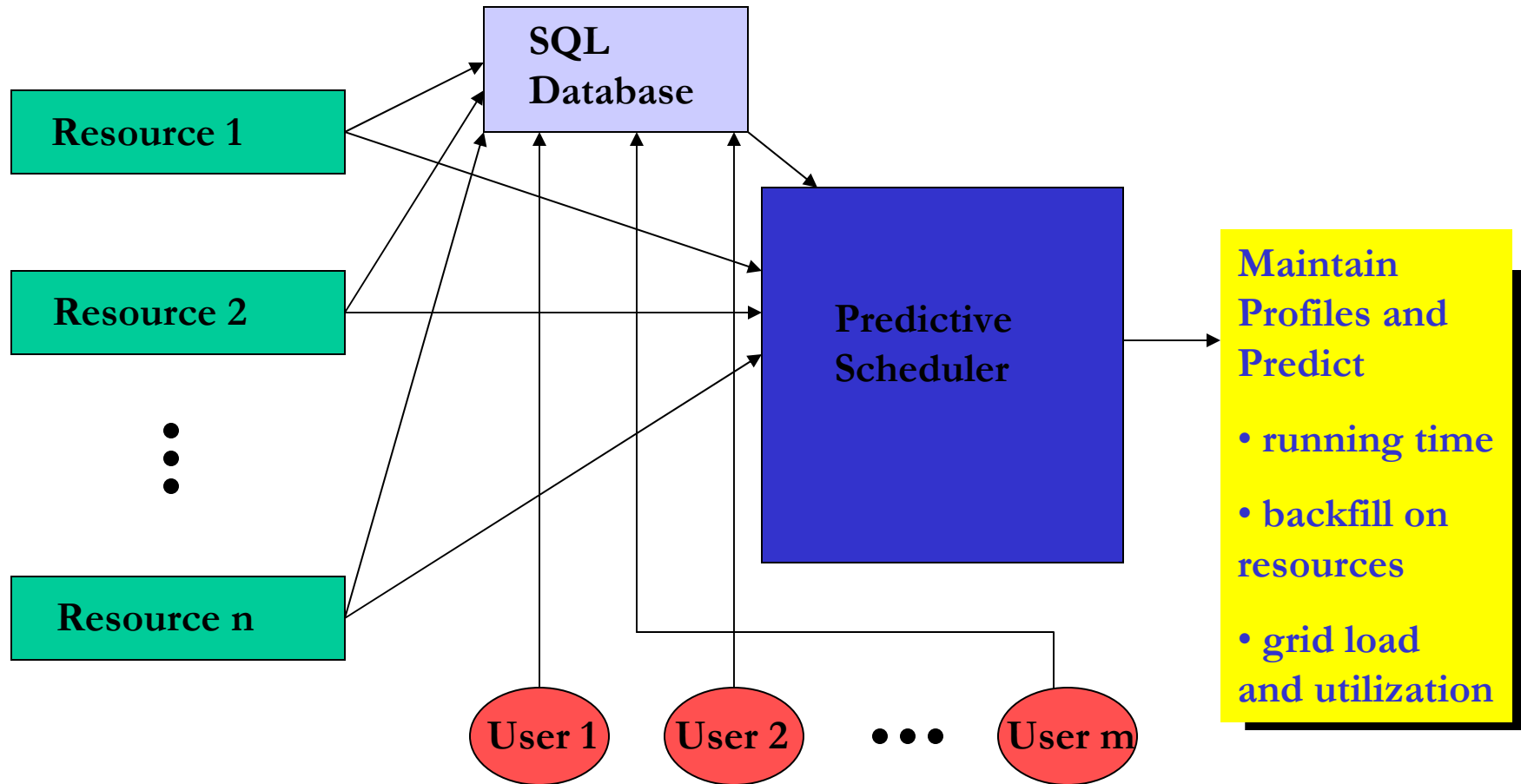
CI Lab Operations Dashboard

Predictive Scheduler

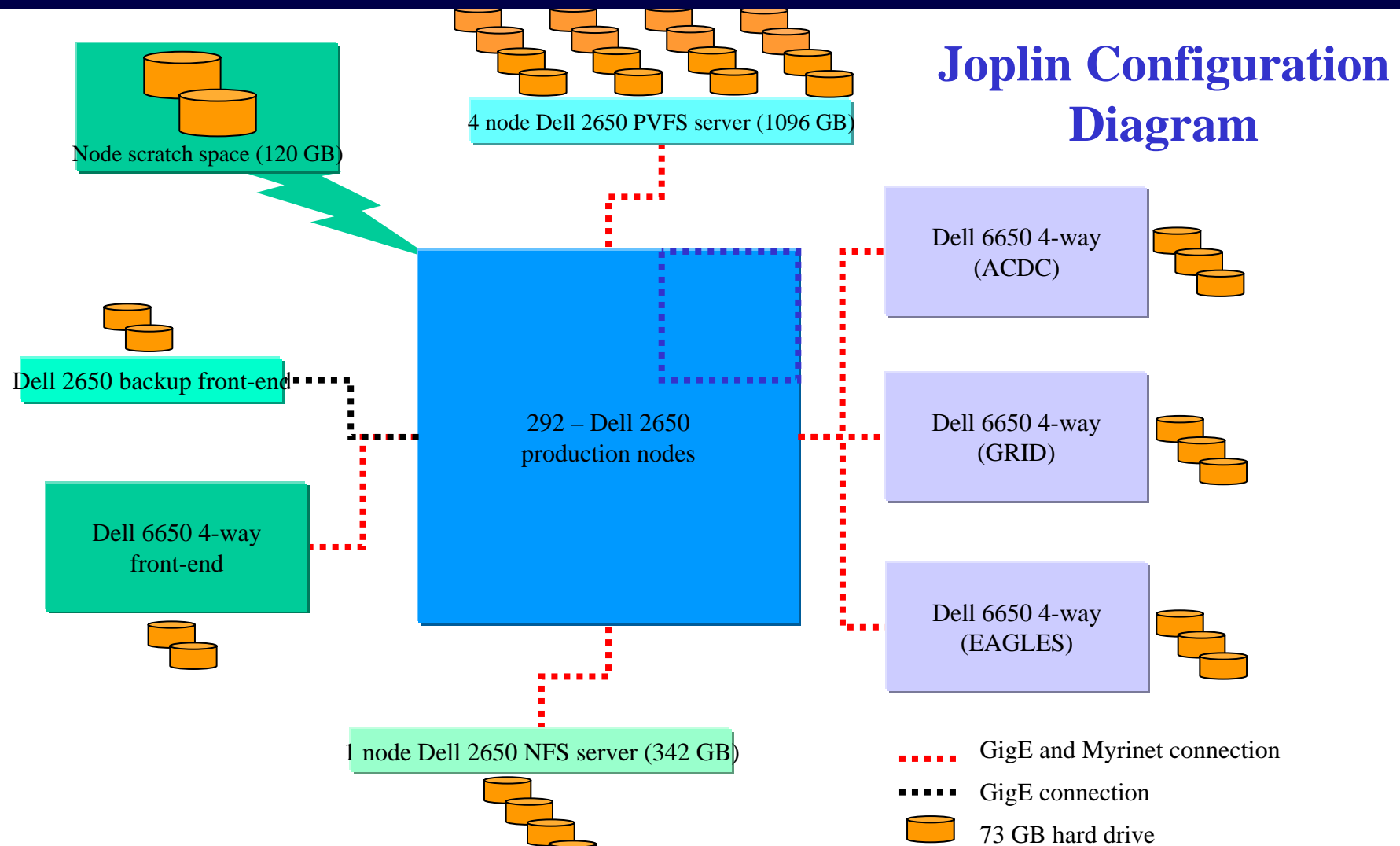
- **Build profiles based on statistical analysis of logs of past jobs**
 - Per User/Group
 - Per Resource
- **Use these profiles to predict runtimes of new jobs**
- **Make use of these predictions to determine**
 - Resources to be utilized
 - Availability of Backfill



System Diagram



ACDC-Grid Dynamic Resource Allocation



Grid Administration

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

Center for Computational Research **GRID PORTAL**
High Performance Grid Computing

Grid Site Administration

PORTAL LOGOUT
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

Users
Groups
Portal Event Log
Database Job List

Organizations (add, edit, delete)
Resources (view, refresh, ping, delete, create host certificate)

Globus Administration
Reports (machine usage, user access to machines, etc.)

Generate Globus grid-mapfile

Specifying an optional include file will cause the contents of this file to be included at the top of the generated grid-mapfile. If a grid-mapfile path is specified a copy of the generated file will be saved into this location. The generated file will be staged to the grid nodes unless the box is checked.

Optional include file:

Optional grid-mapfile path:

Do not stage the file to the grid nodes

Generate Reset

CCR Grid Computing Services: Grid Admin - Resources - Microsoft Internet Explorer

Center for Computational Research **GRID PORTAL**
High Performance Grid Computing

MDS Resource Update Status

Current Time: 16-September-2003 10:58:12

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

Return to the Grid Resource Admin menu.
Return to the Grid Admin menu.

Advanced
Center for Computational Research
Data
Center

CCR Grid Computing Services: Database Job Admin - Microsoft Internet Explorer

Center for Computational Research **GRID PORTAL**
High Performance Grid Computing

Create New Database Job

Create a new database job that can be run by the portal. Job scripts must reside in `/home/griddev/www/jobscripts` prior to creating the database job entry.

Job Name:

Full Path To Script:

Accepts Arguments:

Run Script:

Run As User:

Create Job Reset

Return to the Database Job Admin menu.
Return to the Grid Admin menu.



Data Grid

■ Motivation:

- ❑ Large data collections are emerging as important community resources.
- ❑ Data Grids complement Computational Grids.

■ Definition: *A data grid is a network of distributed storage resources, including archival systems, caches, and databases, which are linked logically to create a sense of global persistence.*

■ Goal: Design and implement transparent management of data distributed across heterogeneous resources.



ACDC-Grid Data Grid

CCR Grid Computing Services: Data Management - Microsoft Internet Explorer

File Edit View Favorites Tools Help

UB University at Buffalo The State University of New York

CCR Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGOUT

- User Tools
 - » Manage Account
- Grid General Info
- Projects
- Resources
 - » Computational Grid
 - » Job Submission
 - » Job/Queue Status
 - » Data Grid
 - » Network Status
 - » Running/Queued Jobs
 - » PBS Job History
 - » Grid Portal Statistics
 - » Conder Flock Statistics
 - » User Information
- Education/Outreach
- Staff Only
- CCR HOME

VIEW Group GROUP miller UserList rappleye

- rappleye
 - KeyMaster
 - Morpheus
 - Tank
 - Agent
 - Rabbit
 - Tank
 - Morpheus
 - Oracle.m
 - Neo
 - Neo
 - Cypher
 - Neo
 - Morpheus
 - Oracle

Advanced
Center for Computational Research
Data

Browser view of "miller" group files published by user "rappleye"



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

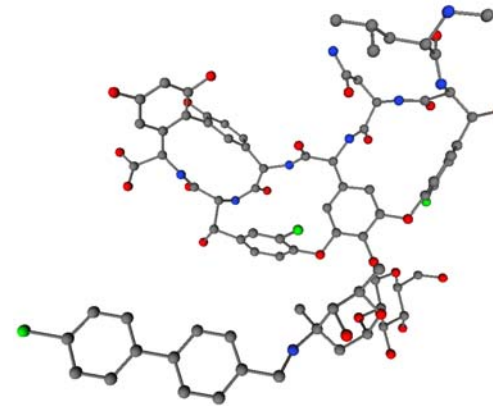
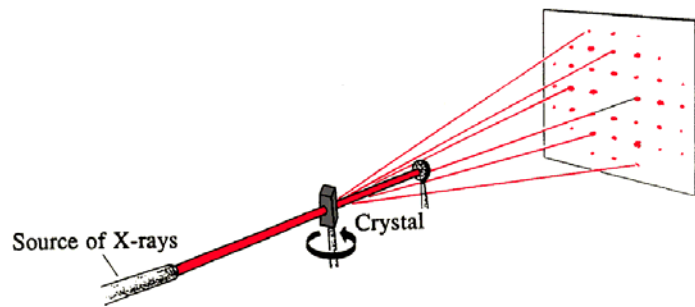
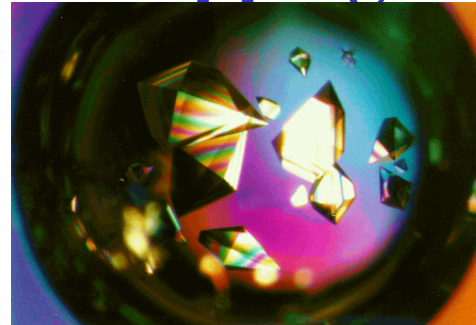


X-Ray Crystallography

- **Objective:** Provide a 3-D mapping of the atoms in a crystal.

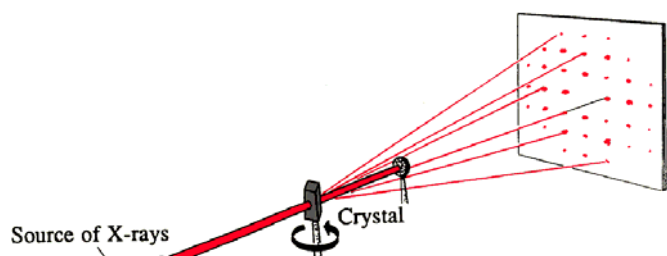
- **Procedure:**

1. Isolate a single crystal.
2. Perform the X-Ray diffraction experiment.



3. Determine molecular structure that agrees with diffraction data.

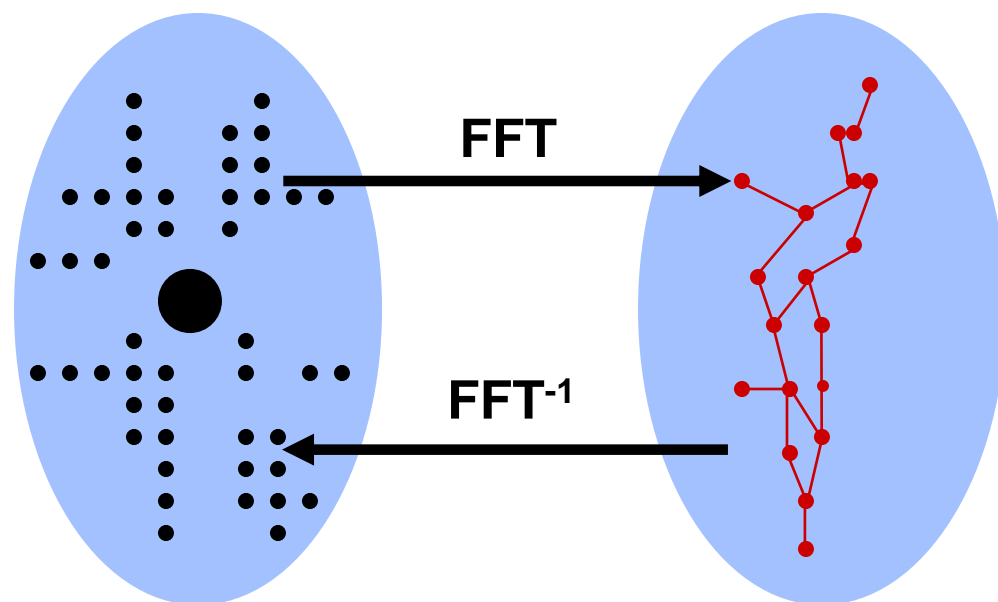
X-Ray Data & Corresponding Molecular Structure



- Experiment yields reflections and associated intensities.
- Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.
- *Phase angles are lost in experiment.*
- *Phase Problem: Determine the set of phases corresponding to the reflections.*

Reciprocal or
“Phase” Space

Real Space

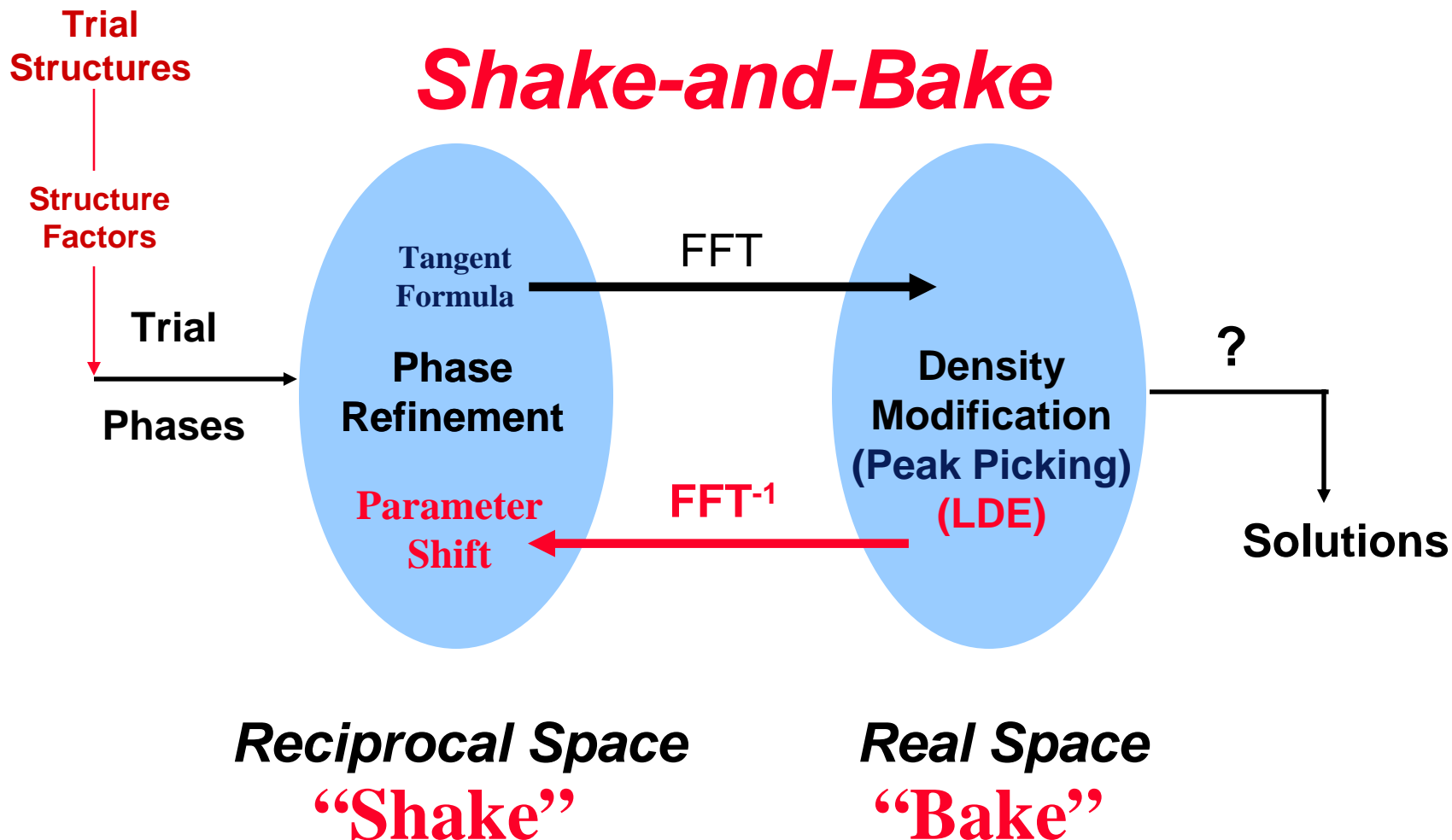


X-Ray Data

Molecular
Structure



Shake-and-Bake Method: Dual-Space Refinement



Grid Enabled *SnB*

- **Problem Statement**
 - ❑ Use all available resources for determining a single structure
- **Grid Enabling Criteria**
 - ❑ Run on heterogeneous set of resources
 - ❑ Store results in *SnB* database
 - ❑ Mine database (and automagically deploy new jobs) to improve parameter settings
- **Runtime Parameters Transparent to User**
 - ❑ Assembling Necessary Files
 - ❑ Number of Processors
 - ❑ Trials per Processor
 - ❑ Appropriate Queue and Running Times



Grid Enabled *SnB* Execution

❑ User

- defines Grid-enabled *SnB* job using Grid Portal or *SnB*
- supplies location of data files from Data Grid
- supplies *SnB* mode of operation

❑ Grid Portal

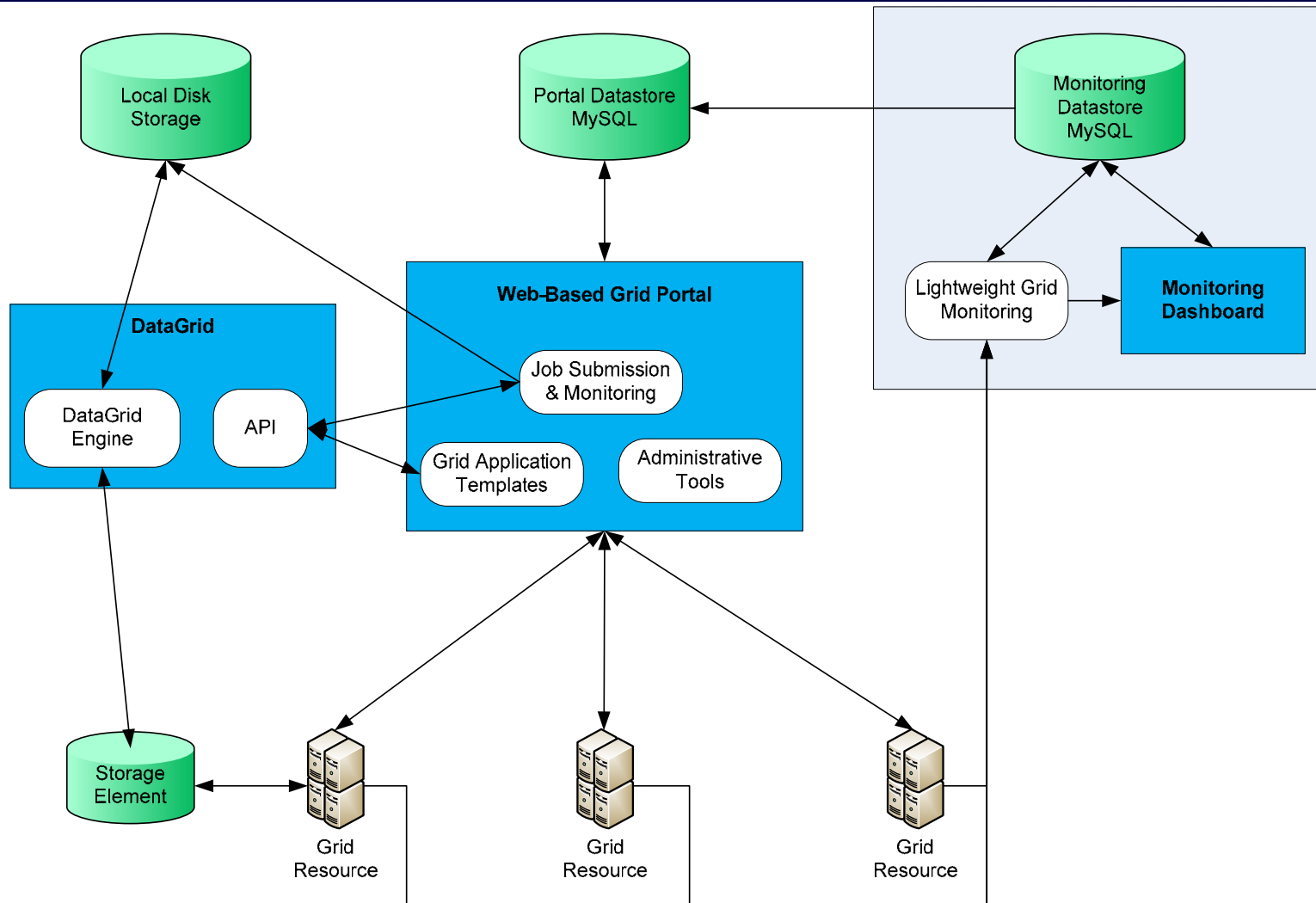
- assembles required *SnB* data and supporting files, execution scripts, database tables.
- determines available ACDC-Grid resources.

❑ ACDC-Grid job management includes:

- automatic determination of appropriate execution times, number of trials, and number/location of processors,
- logging/status of concurrently executing resource jobs, &
- automatic incorporation of *SnB* trial results into the molecular structure database.




NYS Grid Portal



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

CCR Grid Computing Services: Portal Job Submission - Microsoft Internet Explorer

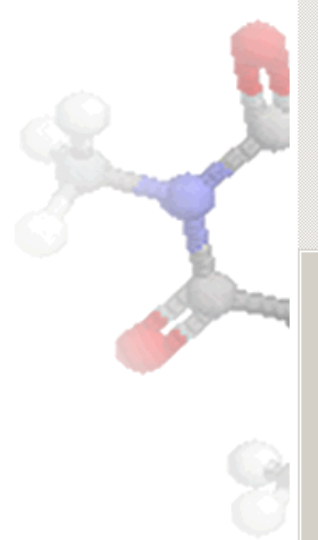
File Edit View Favorites Tools Help

Address <https://griddev.ccr.buffalo.edu/jobs/submit/index.php>

Initial Data Sets

Select dataset to delete	
Datasets	Dataset 1
Name (8 chars max):	<input type="text" value="iledhkl"/>
Dataset Type:	<input type="text" value="Native"/>
File Name (*.hkl) :	<input type="text"/> <input type="button" value="Browse"/>
File Type:	<input type="text" value="F, Sig(F)"/>
Wavelength:	<input type="text" value="1.5418"/>
Max. Resolution:	<input type="text" value="0.94"/>
Anomalous Dispersion:	<input type="text" value="Not Measured"/>
Heavy Element Type:	<input type="text"/>
Nat. Element Replaced:	<input type="text"/>
No. Expected Sites:	<input type="text"/>
F Prime (f'):	<input type="text"/>
F Double Prime (f''):	<input type="text"/>

[Return to the Grid Job Menu](#)



Default Parameters (cont'd)

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

Trial Summary

Grid Job 447 Trial Summary

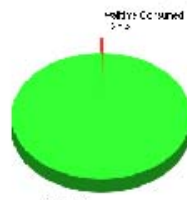
Number of Trials Complete: 285 (28.5%)



Walltime Summary

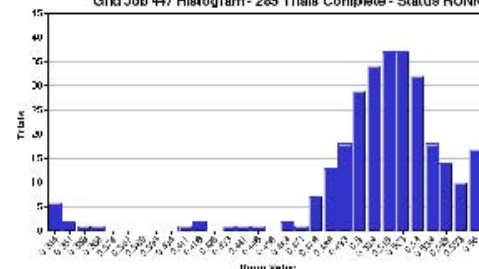
Grid Job 447 Walltime Summary

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

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

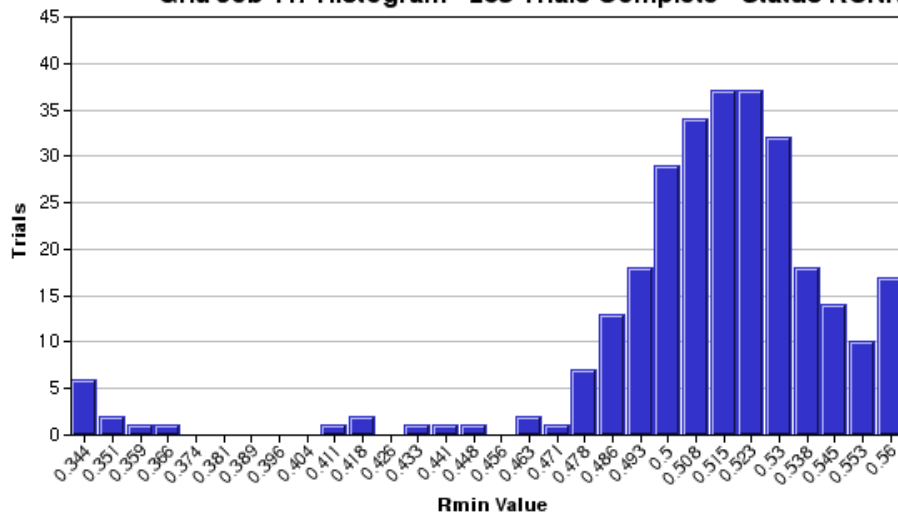
Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Grid Job 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

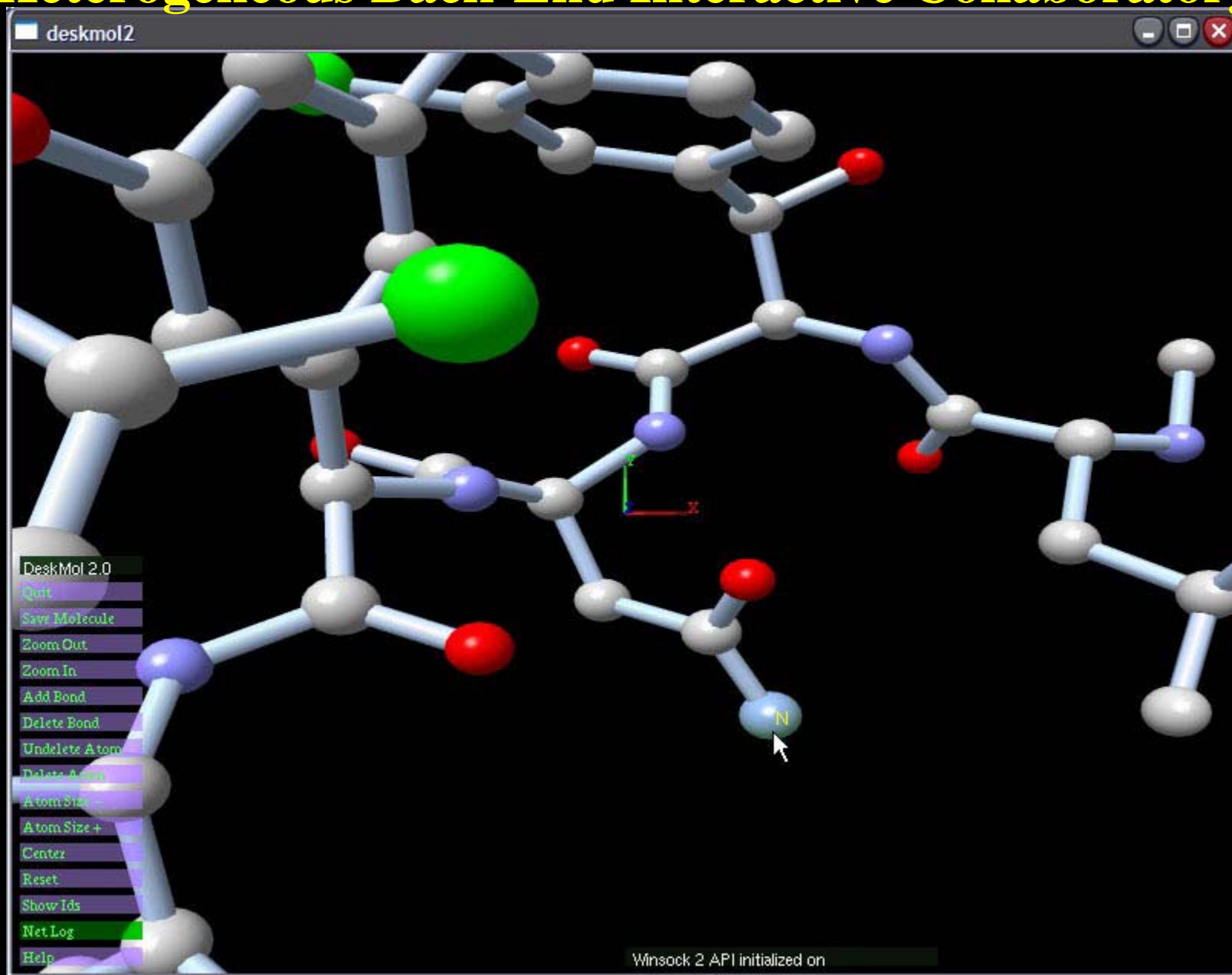
Filter Job List

SnB

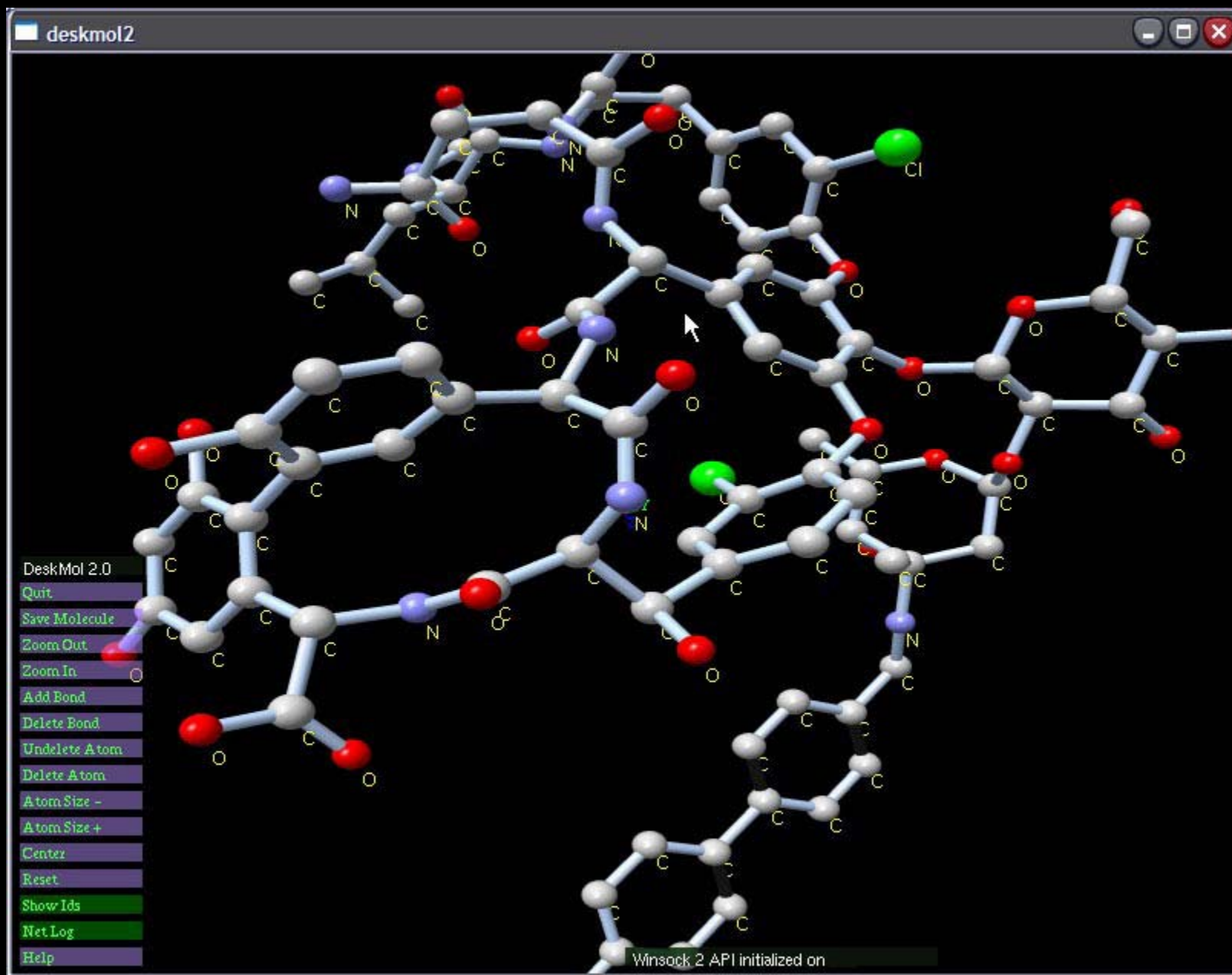
Job Id	Job Name	Resource	Num Procs	Status	Percent Complete	Last Update	Cancel Job	Drilldown
447	iledhkl	clearwater.ccr.buffalo.edu	5	RUNNING	28.5	15-Mar-2005 10:22:00	<input type="checkbox"/>	
446	trilys	clearwater.ccr.buffalo.edu	10	RUNNING	1	15-Mar-2005 10:22:00	<input type="checkbox"/>	
444	64chkl	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.

New York State Grass Roots Cyberinfrastructure Initiative

- Miller's NYS Grid used as fundamental infrastructure.
- Currently an initiative of NYSErNet.
- Open to academic and research institutions.
- **Mission Statement: To create and advance collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in NYS.**
- **Enable Research, Scholarship, and Economic Development in NYS.**
- **Currently, no significant utilization.**



TRUN: Transborder Research University Network

- **Ontario: York, Toronto, Western Ontario, McMaster, Queen's, Waterloo, Guelph**
- **NYS: Buffalo, Rochester, Syracuse, Cornell, Albany, RIT**
- **Mission Statement: Expand and support cooperation among research universities in the border region of Province of Ontario and NYS:**
 - ❑ Collaborative/consortial research
 - ❑ Joint applications for external funding
 - ❑ Cooperative academic programs
 - ❑ Faculty and student exchanges
 - ❑ Shared facilities
 - ❑ Joint conferences, symposia, workshops



www.trun.ca



TRUN: Transborder Research University Network

■ Current Focus

- Great Lakes Sustainable Energy
- IT-Supported Disciplinary Research
- High Performance Computing
- Canada-U.S. Policy and Standardization of Binational Data



www.trun.ca

■ General Issues

- Public Policy Issues, Regional Governance
- Border Security and Mobility
- Economic and Workforce Development
- University Partnerships with Government and Industry
- Health Care and Policy
- Basic Research and Technology Transfer



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