Shaking-and-Baking on a Grid

Russ Miller

Director, Center for Computational Research

UB Distinguished Professor, Computer Science & Engineering

Senior Research Scientist, Hauptman-Woodward Medical Inst

University at Buffalo

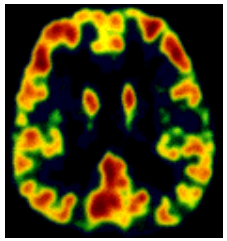
The State University of New York

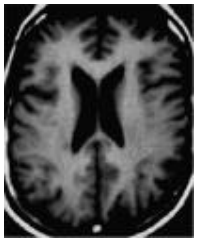
Outline

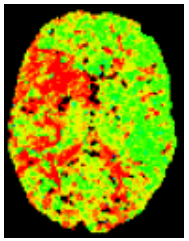
- **Buffalo's Center of Excellence in Bioinformatics**
- **Supercomputing & Visualization in CCR**
- **Grid Computing Overview**
- **WNY Computational & Data Grids**
  - ***Shake-and-Bake:* Computational Crystallography**
  - **ECCE: Computational Chemistry**

Biomedical Advances

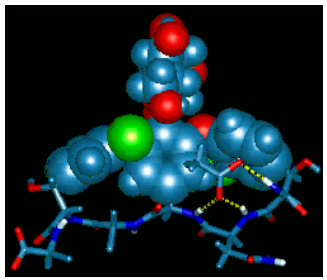
- PSA Test (screen for Prostate Cancer)
- Avonex: Interferon Treatment for Multiple Sclerosis
- Artificial Blood
- Nicorette Gum
- Fetal Viability Test
- Implantable Pacemaker
- Edible Vaccine for Hepatitis C
- Timed-Release Insulin Therapy
- Anti-Arrythmia Therapy
  - Tarantula venom







- Direct Methods Structure Determination
  - Listed on "Top Ten Algorithms of the 20th Century"
  - Vancomycin
  - Gramacidin A



- High Throughput Crystallization Method: Patented
- NIH National Genomics Center: Northeast Consortium
- Howard Hughes Medical Institute: Center for Genomics & Proteomics

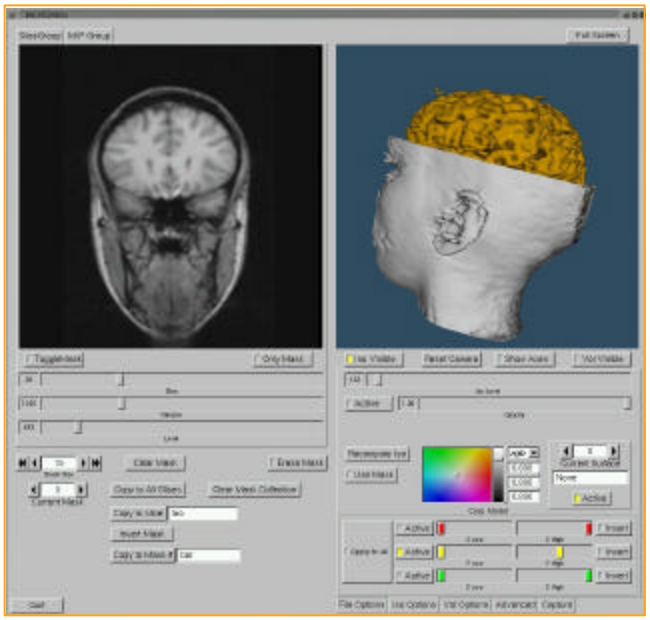
Major CCR Resources

- Dell Linux Cluster: #22 ® #25 ® #38
  - 600 P4 Processors (2.4 GHz)
  - 600 GB RAM; 40 TB Disk; Myrinet
- Dell Linux Cluster: #187 ® #368 ® off
  - 4036 Processors (PIII 1.2 GHz)
  - 2TB RAM; 160TB Disk; 16TB SN
  - Restricted Use (Skolnick)

- SGI Origin3800
  - 64 Processors (400 MHz)
  - 32 GB RAM; 400 GB Disk
- IBM RS/6000 SP
  - 78 Processors
  - 26 GB RAM; 640 GB Disk
- Sun Microsystems Cluster
  - 48 Sun Ultra 5s (333MHz)
  - 16 Dual Sunblades (750MHz)
  - 30 GB RAM, Myrinet
- SGI Intel Linux Cluster
  - 150 PIII Processors (1 GHz)
  - 75 GB RAM, 2.5 TB Disk Storage
- Apex Bioinformatics System
  - Sun V880 (3), 6800, 280R (2), PIIIs
  - Sun 3960: 7 TB Disk Storage
- HP/Compaq SAN (4Q03)
  - 75 TB Disk; 200 TB Tape

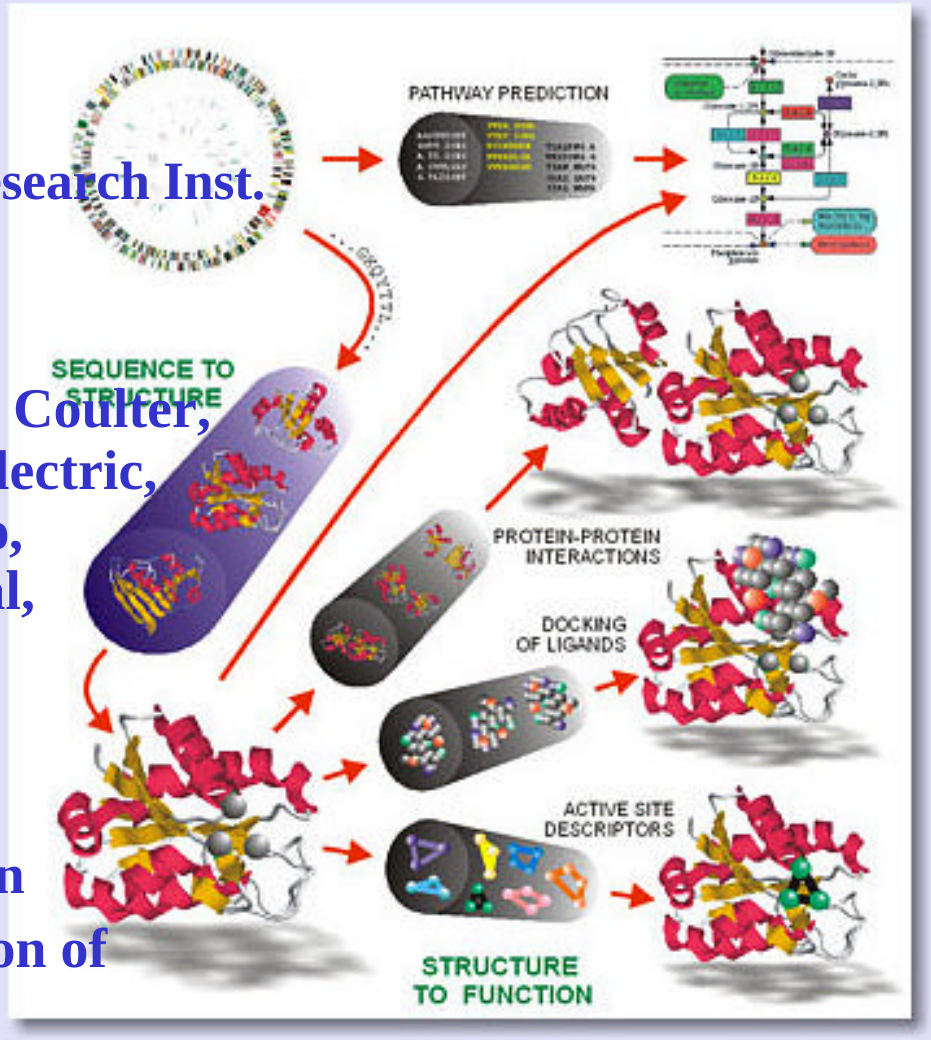
Bioinformatics in Buffalo
A $360M Initiative

- New York State: $121M
- Federal Appropriations: $13M
- Corporate: $146
- Foundation: $15M
- Grants & Contracts: $64M



Bioinformatics Partners

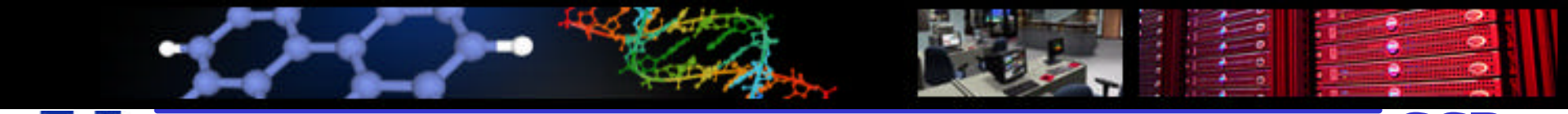
- **Lead Institutions**
  - University at Buffalo (UB)
  - Hauptman-Woodward Medical Research Inst.
  - Roswell Park Cancer Institute
- **Corporate Partners**
  - Amersham Pharmacia, Beckman Coulter, Bristol Myers Squibb, General Electric, Human Genome Sciences, Immco, Invitrogen, Pfizer Pharmaceutical, Wyeth Lederle, Zeptometrix
  - Dell, HP, SGI, Stryker, Sun
  - AT&T, Sloan Foundation
  - InforMax, Q-Chem, 3M, Veridian
  - BioPharma Ireland, Confederation of Indian Industries



UB Bioinformatics Snapshot (2002-03)

- 7/02: Jeff Skolnick, Director
  - Brought 13 addit'l staff with him
  - Authorized to hire 10 additional research groups
- 4/03: Norma Nowak, co-Dir
  - Authorized to hire 10 additional research groups
- 9/03: Daniel Fischer, Dir of Ed
- Additional Members TBD
- External Funding ($0)
  - Applications submitted
- Deliverables
  - 12 scientific papers

- Resources (Capaldi, Holm, Penksa, Miller, et al.)
  - Building
  - 6TF ® 10TF Compute Cluster

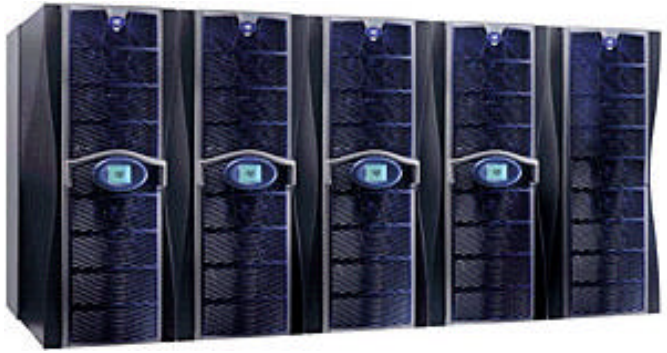


Related Academic Programs (Pitman, et al.)

- Bachelor's & Master's Program in Bioinformatics
- Related Disciplines
  - Chemical Biology (Sloan Support)
  - Computational Chemistry (Sloan Support)
  - Environmental Analysis (Sloan Support)
  - Medical Informatics (Graduate Certificate)
  - Pharmacometrics
  - UB-HWI Department of Structural Biology (Ph.D.)
- Advanced Degrees under Development
  - Biophotonics
- Complementary Degrees in WNY
  - Canisius College & RIT
  - Niagara University NYS $5M Center of Excellence in Bioinformatics (degrees in development)

Center for Computational Research

- **High-Performance Computing and High-End Visualization**
  - 110 Research Groups in 27 Depts
  - 25 Companies and Institutions
- **Sample Areas**
  - Urban Visualization and Simulation
  - Computational Chemistry
  - Ground Water Modeling
  - Geophysical Mass Flows
  - Networked Multimedia
  - Medical Imaging
- **Training**
  - Workshops; Courses
  - Degree Programs



CCR 1999-2003 Snapshot

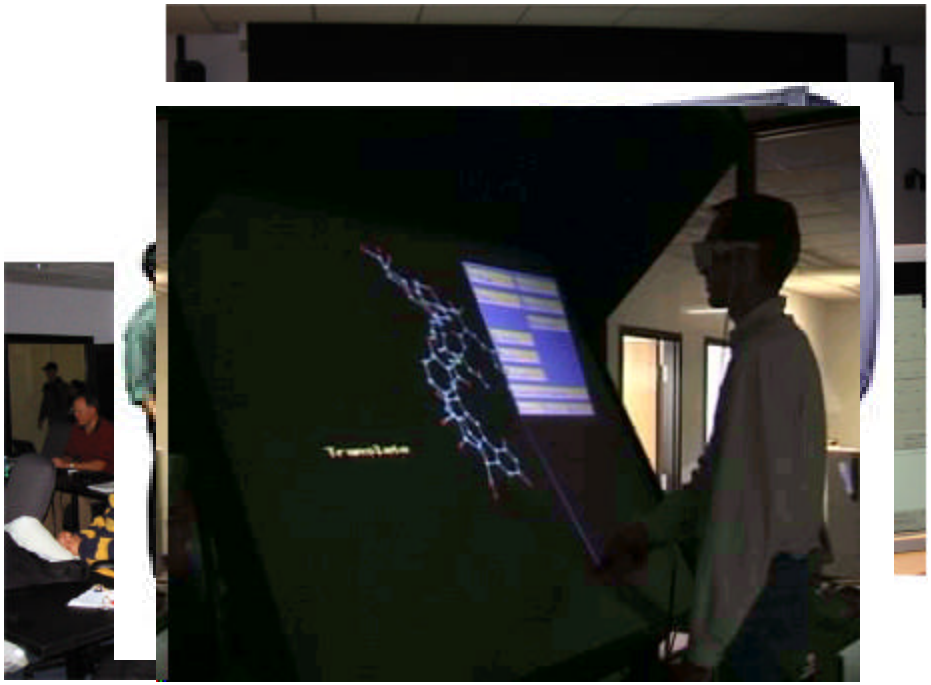
- **Personnel**
  - 18 State-Supported Staff (15 Technical / 3 Clerical)
  - 2 Grant-Supported Staff
  - Undergraduates (REUs, Workstudy)
- **External Funding**
  - $111M External Funding
    - $13.5M as lead
    - $97.5M in support
  - $41.8M Vendor Donations
- **Deliverables**
  - 350+ Publications
  - Software, Media, Algorithms, Consulting, Training, CPU Cycles…

Sample Computational Research

- **Computational Chemistry** (King, Kofke, Coppens, Furlani, Tilson, Lund, Swihart, Ruckenstein, Garvey)
  - **Algorithm development & simulations**
- **Groundwater Flow Modeling** (Rabideau, Jankovic, Becker, Flewelling)
  - **Predict contaminant flow in groundwater & possible migration into streams and lakes**
- **Geophysical Mass Flows** (Patra, Sheridan, Pitman, Bursik, Jones, Winer)
  - **Study of geophysical mass flows for risk assessment of lava flows and mudslides**
- **Bioinformatics** (Zhou, Miller, Hu, Szyperski – NIH Consortium, HWI)
  - **Protein Folding: computer simulations to understand the 3D structure of proteins**
  - **Structural Biology; Pharmacology**
- **Computational Fluid Dynamics** (Madnia, DesJardin, Lordi, Taulbee)
  - **Modeling turbulent flows and combustion to improve design of chemical reactors, turbine engines, and airplanes**
- **Physics** (Jones, Sen)
  - **Many-body phenomena in condensed matter physics**
- **Chemical Reactions** (Mountziaris)
- **Molecular Simulation** (Errington)

Visualization Resources

- **Fakespace ImmersaDesk R2**
  - Portable 3D Device
- **Tiled-Display Wall**
  - 20 NEC projectors: 15.7M pixels
  - Screen is 11’ ´ 7’
  - Dell PCs with Myrinet2000
- **Access Grid Node**
  - Group-to-Group Communication
  - Commodity components
- **SGI Reality Center 3300W**
  - Dual Barco’s on 8’ ´ 4’ screen
- **VREX VR-4200 Stereo Imaging Projector**
  - Portable projector works with PC

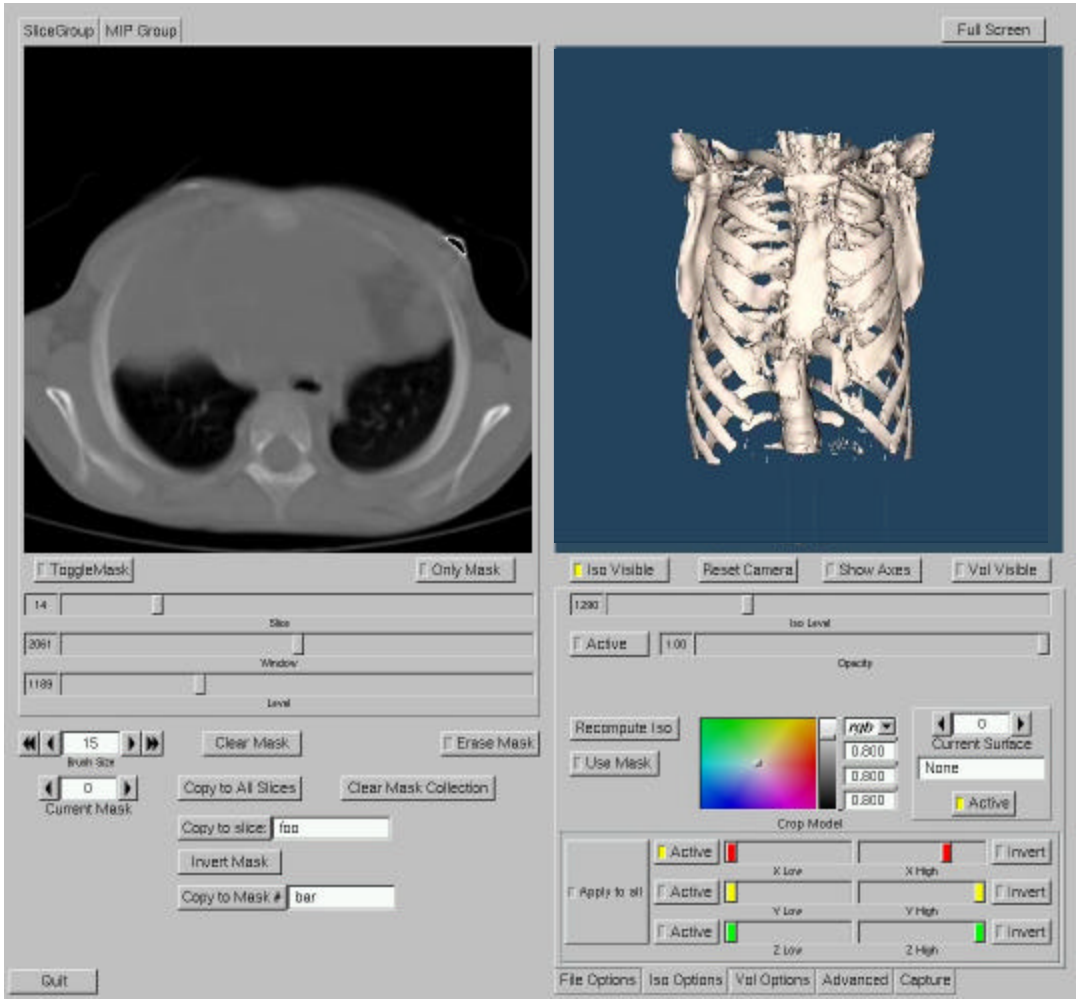


Sample Visualization Areas

- **Computational Science** (Patra, Sheridan, Becker, Flewelling, Baker, Miller, Pitman)
  - **Simulation and modeling**
- **Urban Visualization and Simulation** (CCR)
  - **Public projects involving urban planning**
- **Medical Imaging** (Hoffmann, Bakshi, Glick, Miletich, Baker)
  - **Tools for pre-operative planning; predictive disease analysis**
- **Geographic Information Systems** (CCR, Bisantz, Llinas, Kesavadas, Green)
  - **Parallel data sourcing software**
- **Historical Reenactments** (Paley, Kesavadas, More)
  - **Faithful representations of previously existing scenarios**
- **Multimedia Presentations** (Anstey, Pape)
  - **Networked, interactive, 3D activities**

3D Medical Visualization App

- Collaboration with Children's Hospital
  - Leading miniature access surgery center
- Application reads data output from a CT Scan
- Visualize multiple surfaces and volumes
- Export images, movies or CAD representation of model

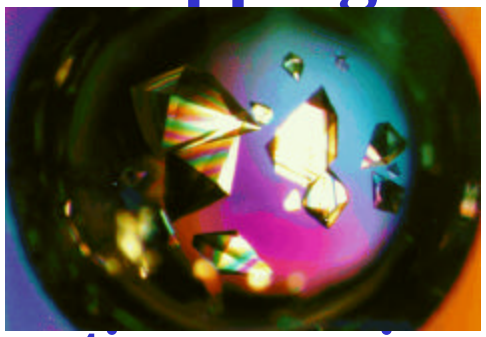


StreetScenes® Demo

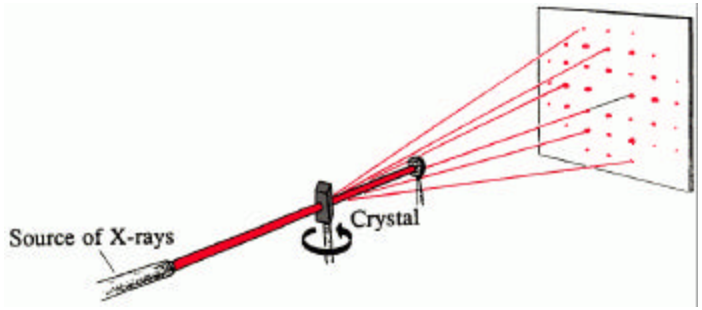
- ***StreetScenes***® **is a Virtual Reality (VR) software solution for 3D visualization of surface traffic**
- **3D model of proposed soccer stadium in Rochester**
- **Used *StreetScenes*® to import output file from Synchro traffic simulation**

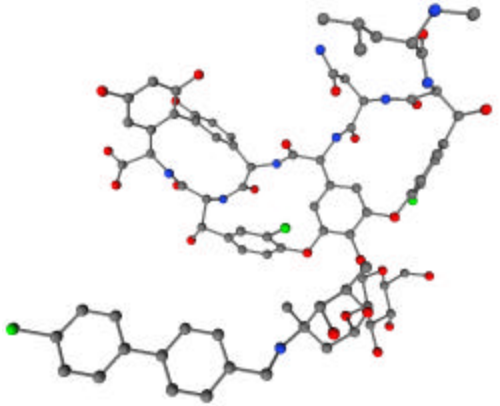
X-Ray Crystallography

- **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.**



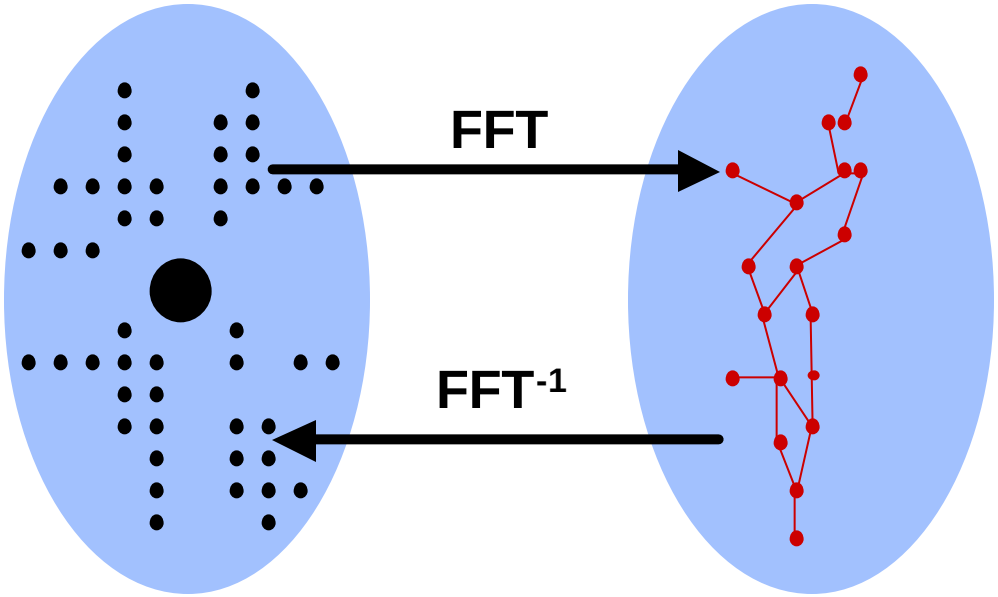


  3. **Determine molecular structure that agrees with diffration data.**

X-Ray Data & Corresponding Molecular Structure

Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.

Reciprocal or "Phase" Space

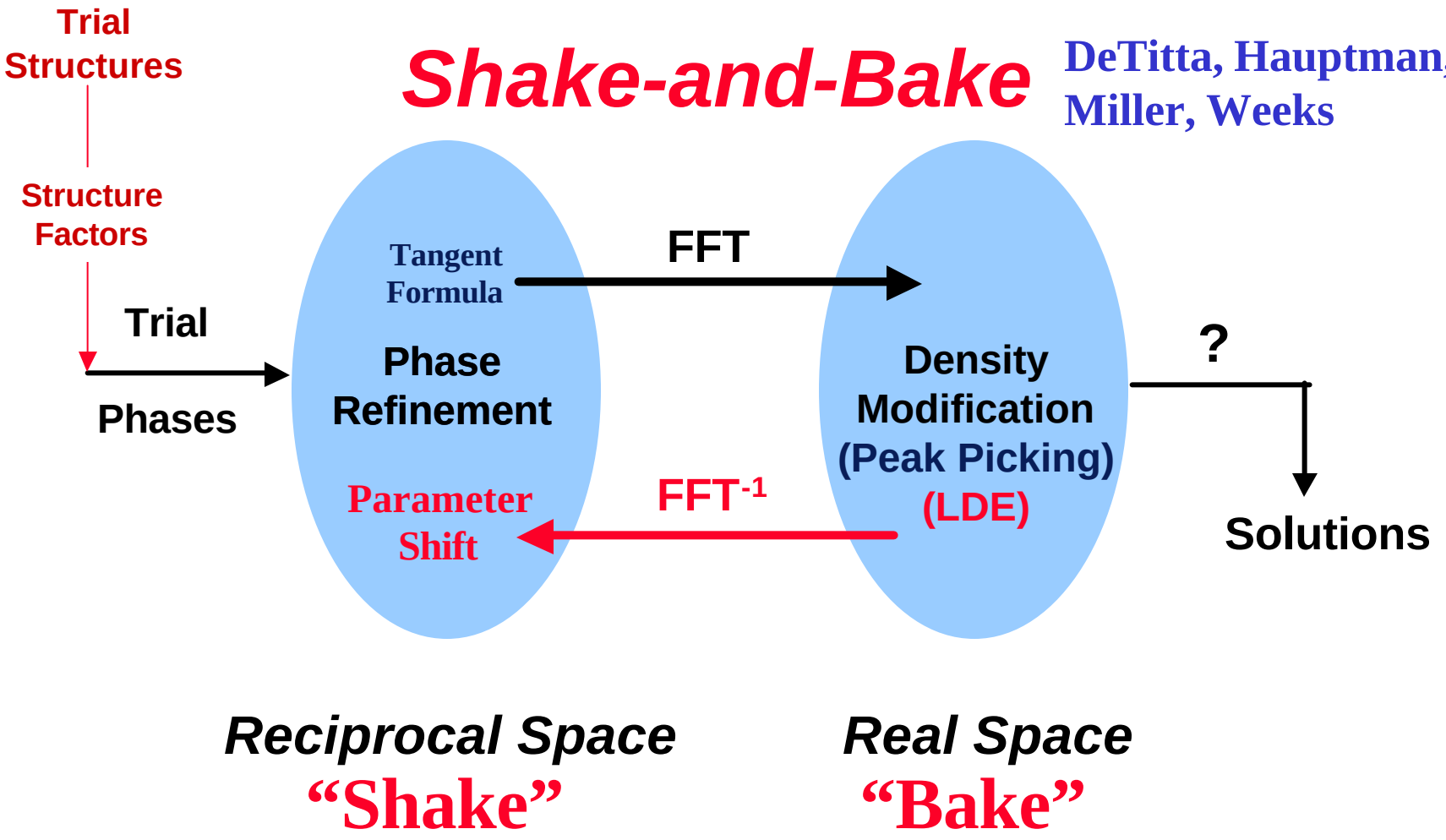


Real Space

X-Ray Data *Molecular Structure*

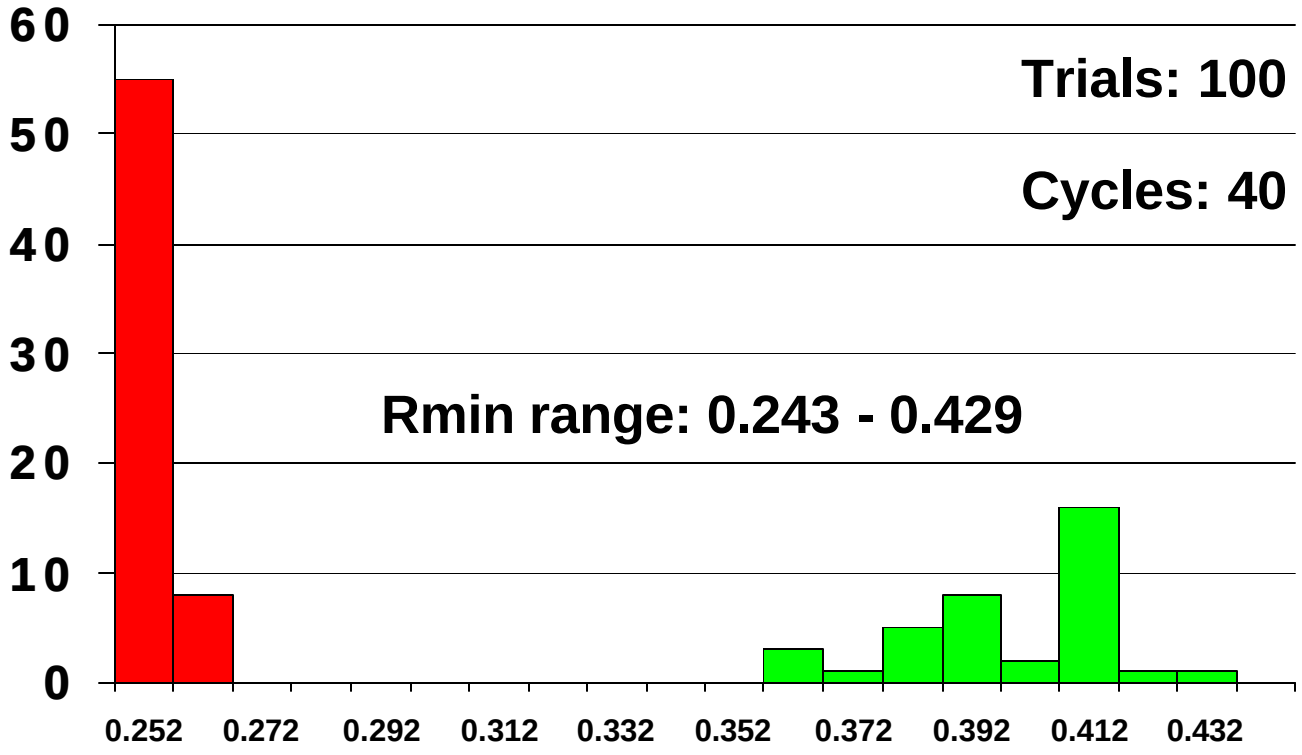
- **Phases lost during the crystallographic experiment.**
- ***Phase Problem*: Determine phases of the reflections.**

Shake-and-Bake Method: Dual-Space Refinement

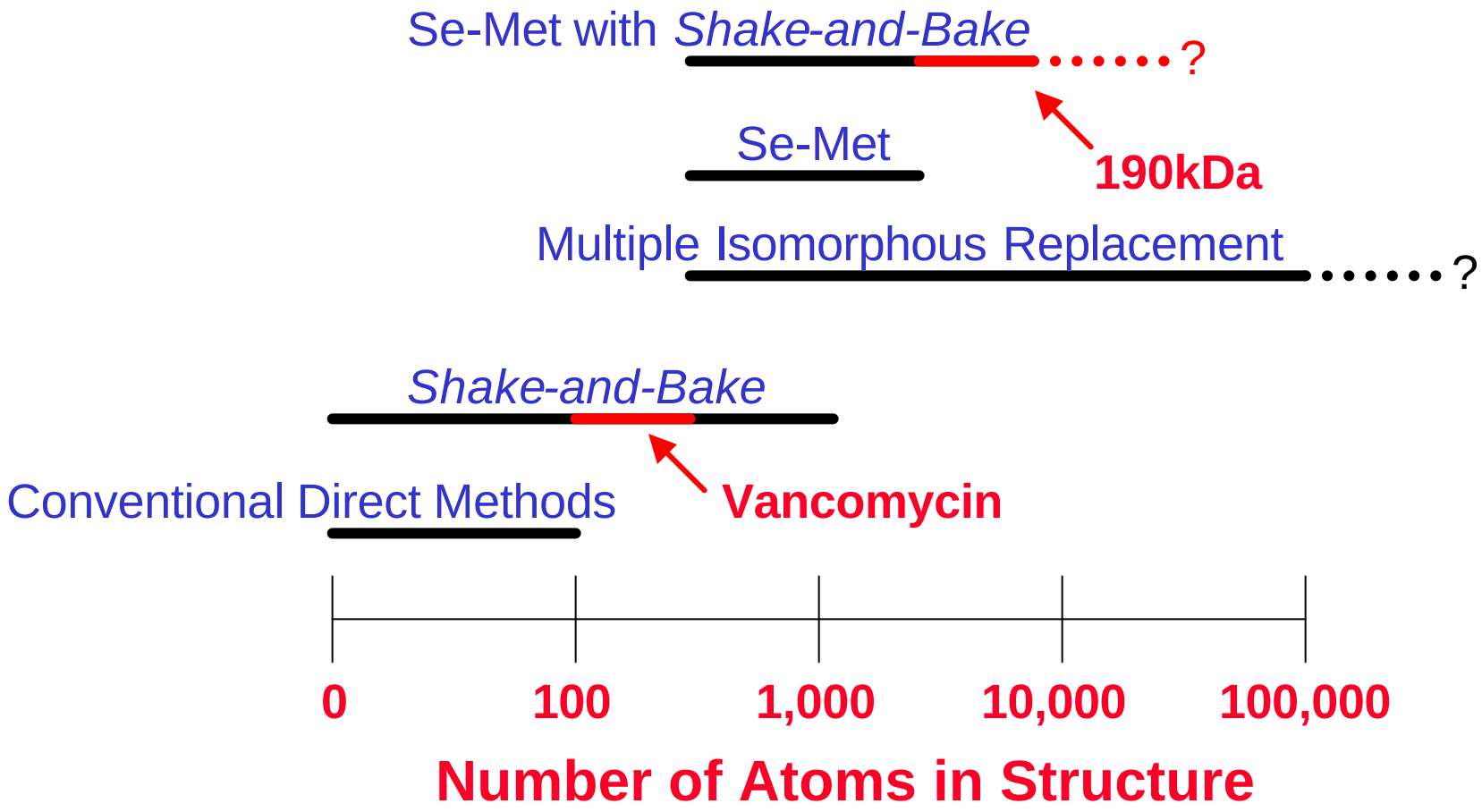


Ph8755: *SnB* Histogram

Atoms: 74 **Phases: 740**
Space Group: P1 **Triples: 7,400**



Phasing and Structure Size



Grid Computing 2003

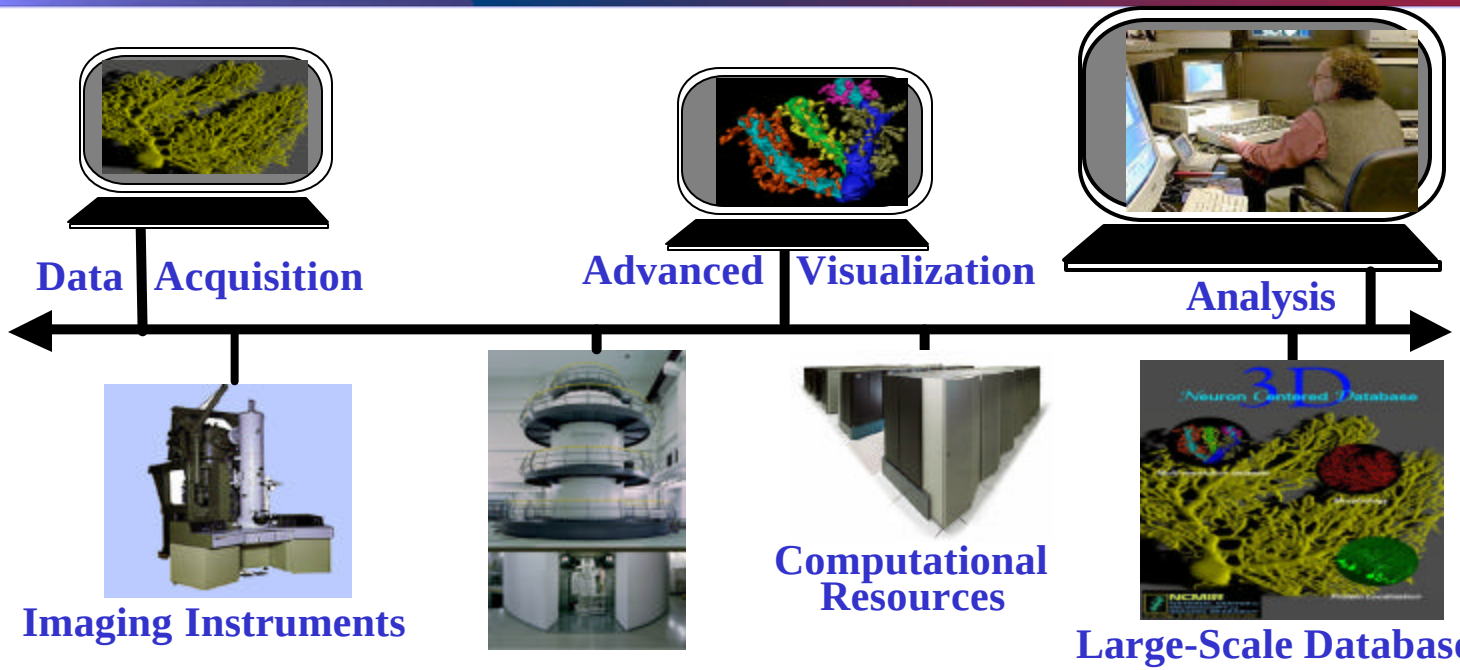
DISCOM

SinRG

APGrid

IPG …

Grid Computing Overview



Thanks to Mark Ellisman

- 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

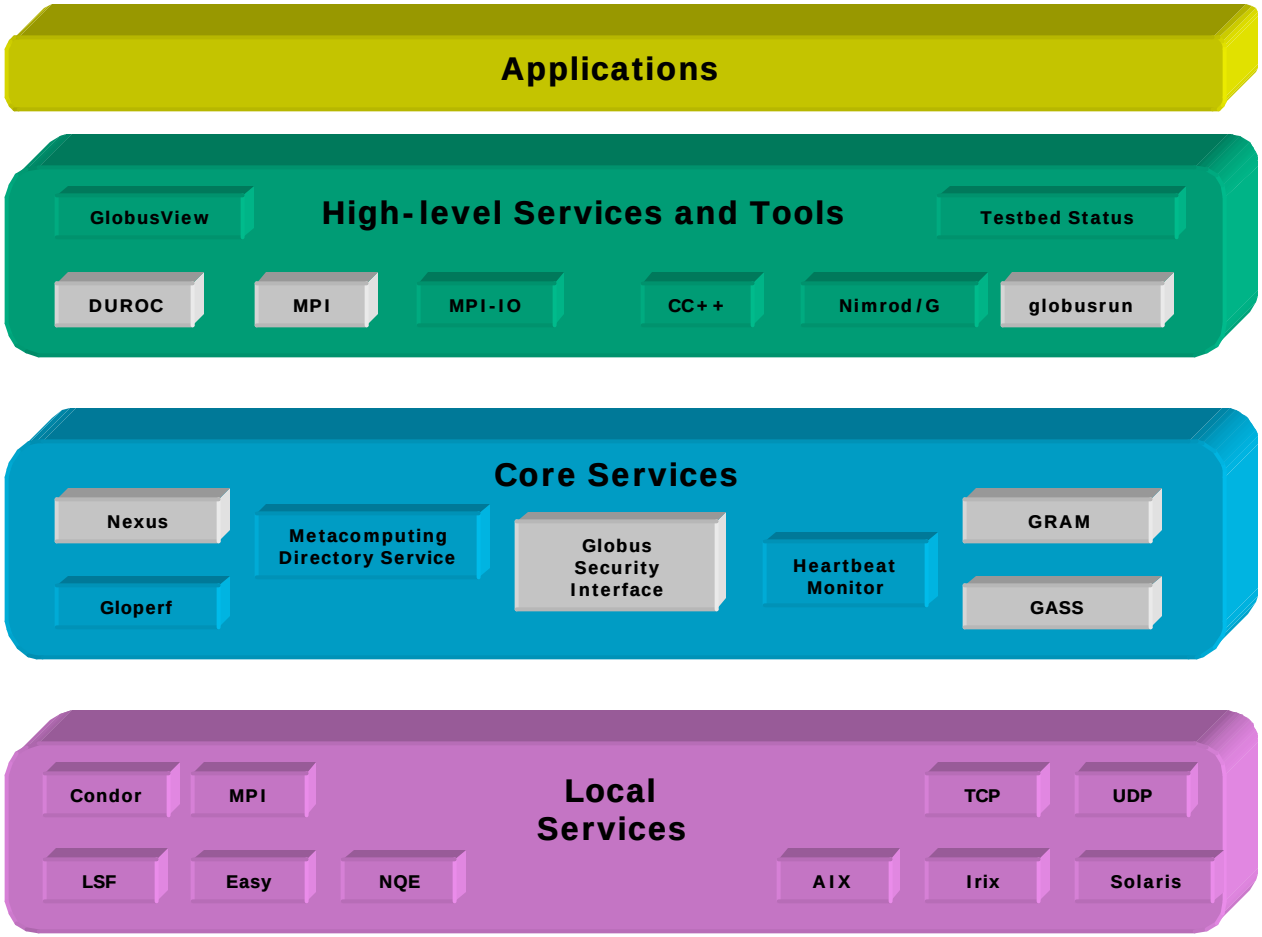
Factors Enabling the Grid

- **Internet is Infrastructure**
  - **Increased network bandwidth and advanced services**
- **Advances in Storage Capacity**
  - **Terabyte costs less than $5,000**
- **Internet-Aware Instruments**
- **Increased Availability of Compute Resources**
  - **Clusters, supercomputers, storage, visualization devices**
- **Advances in Application Concepts**
  - **Computational science: simulation and modeling**
  - **Collaborative environments ® large and varied teams**
- **Grids Today**
  - **Moving towards production; Focus on middleware**

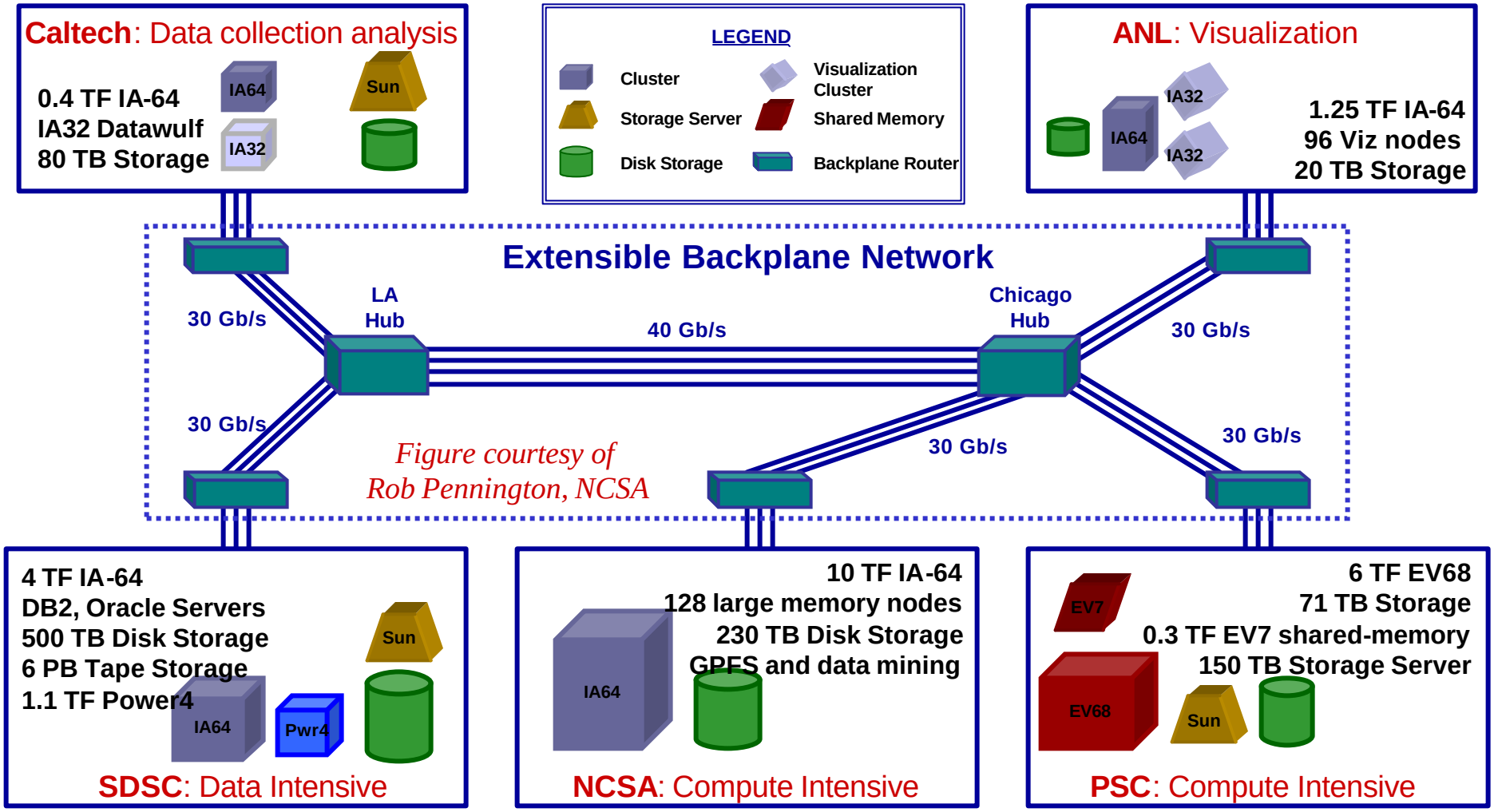
The Globus Project (Ian Foster and Carl Kesselman)

- **Globus model focuses on providing key Grid services**
  - Resource access and management
  - Grid FTP
  - Information Service
  - Security services
    - Authentication
    - Authorization
    - Policy
    - Delegation
  - Network reservation, monitoring, control

The Grid as a Layered Set of Services



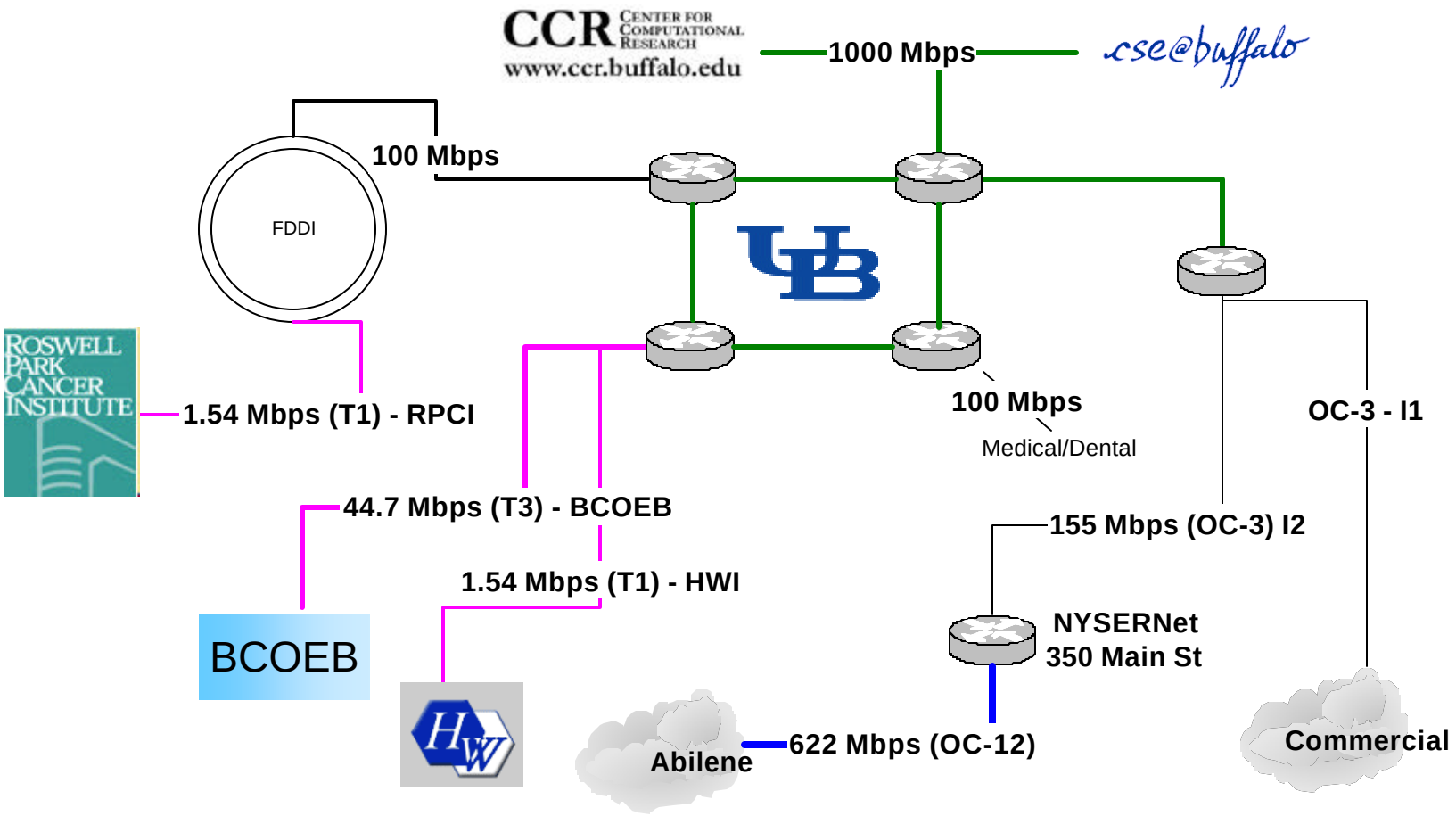
NSF Extensible TeraGrid Facility



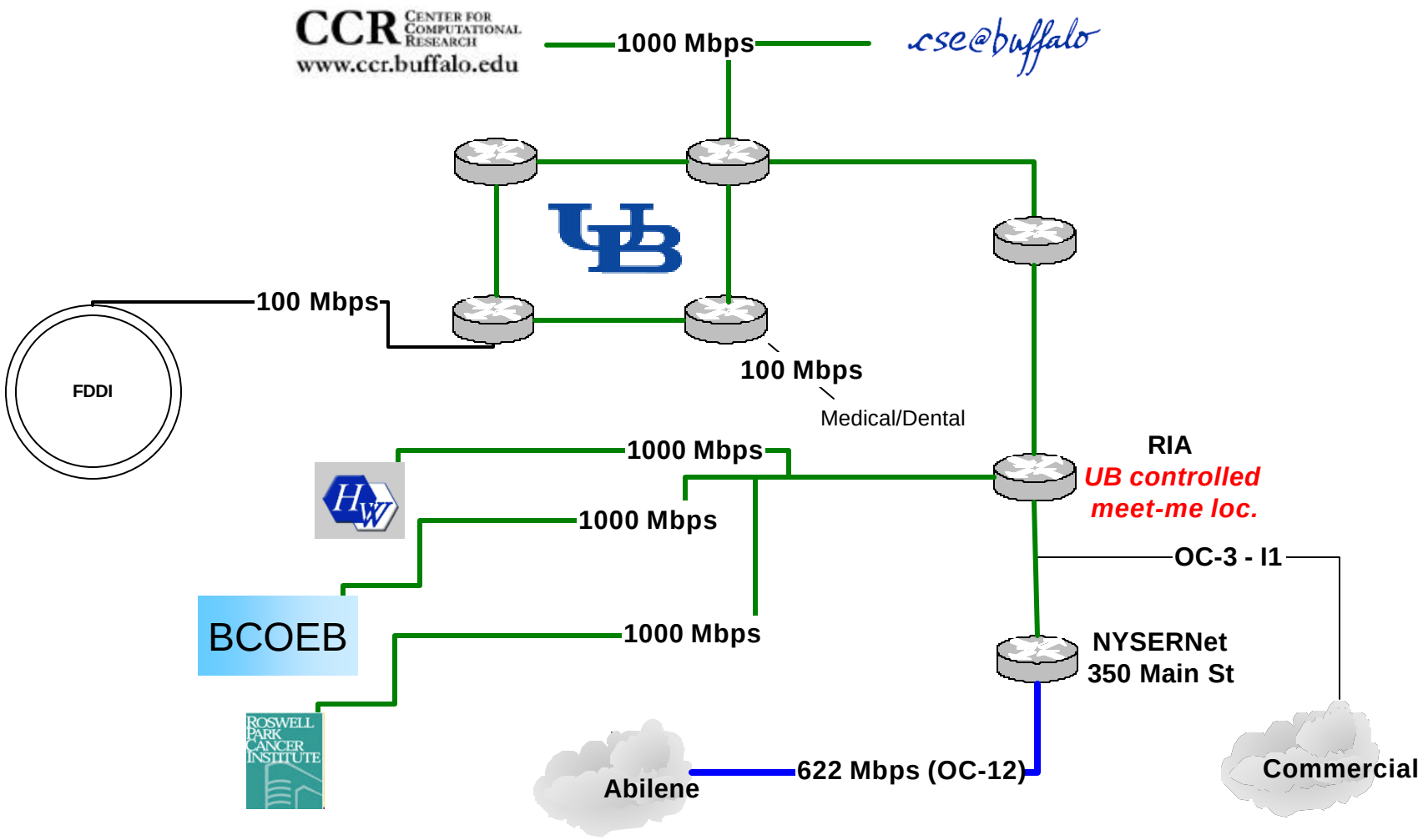
Critical Resources:
WNY Computational & Data Grids

- **Computational & Data Resources (CCR)**
  - **10TF Computing & 78TB Storage**
- **Instruments (HWI, RPCI)**
  - **Microarray; Diffractometer; NMR**
  - **High-Throughput Crystallization Laboratory**
- **Data Generation (HWI)**
  - **7TB per year**
- **Databases (UB-N, UB-S, BGH, CoE)**
  - ***SnB*; Multiple Sclerosis; Protein/Genomic**

Network Connections



Network Connections (New)



Advanced CCR Data Center (ACDC) Computational Grid Overview

Nash: Compute Cluster

75 Dual Processor
1 GHz Pentium III
RedHat Linux 7.3
1.8 TB Scratch Space

Joplin: Compute Cluster

300 Dual Processor
2.4 GHz Intel Xeon
RedHat Linux 7.3
38.7 TB Scratch Space

Mama: Compute Cluster

9 Dual Processor
1 GHz Pentium III
RedHat Linux 7.3
315 GB Scratch Space

ACDC: Grid Portal

4 Processor Dell 6650
1.6 GHz Intel Xeon
RedHat Linux 9.0
66 GB Scratch Space

Young: Compute Cluster

16 Dual Sun Blades
47 Sun Ultra5
Solaris 8
770 GB Scratch Space

Crosby: Compute Cluster

SGI Origin 3800
64 - 400 MHz IP35
IRIX 6.5.14m
360 GB Scratch Space

Fogerty: Condor Flock Master

1 Dual Processor
250 MHz IP30
IRIX 6.5

Expanding

RedHat, IRIX, Solaris, WINNT, etc

CCR

19 IRIX, RedHat, & WINNT Processors

T1 Connection

Computer Science & Engineering

25 Single Processor Sun Ultra5s

School of Dental Medicine

9 Single Processor Dell P4 Desktops

Hauptman-Woodward Institute

13 Various SGI IRIX Processors

Note: Network connections are 100 Mbps unless otherwise noted.

ACDC Data Grid Overview

182 GB Storage

Joplin: Compute Cluster

300 Dual Processor
2.4 GHz Intel Xeon
RedHat Linux 7.3
38.7 TB Scratch Space

Nash: Compute Cluster

75 Dual Processor
1 GHz Pentium III
RedHat Linux 7.3
1.8 TB Scratch Space

100 GB Storage

70 GB Storage

Mama: Compute Cluster

9 Dual Processor
1 GHz Pentium III
RedHat Linux 7.3
315 GB Scratch Space

56 GB Storage

ACDC: Grid Portal

4 Processor Dell 6650
1.6 GHz Intel Xeon
RedHat Linux 9.0
66 GB Scratch Space

100 GB Storage

Young: Compute Cluster

16 Dual Sun Blades
47 Sun Ultra5
Solaris 8
770 GB Scratch Space

136 GB Storage

Crosby: Compute Cluster

SGI Origin 3800
64 - 400 MHz IP35
IRIX 6.5.14m
360 GB Scratch Space

CSE Multi-Store
2 TB

Storage Area Network
75 TB

Network Attached Storage
480 GB

Fogerty: Condor Flock Master

1 Dual Processor
250 MHz IP30
IRIX 6.5

CCR
19 IRIX, RedHat, & WINNT Processors

T1 Connection

Computer Science & Engineering
25 Single Processor Sun Ultra5s

School of Dental Medicine
9 Single Processor Dell P4 Desktops

Hauptman-Woodward Institute
13 Various SGI IRIX Processors

Note: Network connections are 100 Mbps unless otherwise noted.

WNY Grid Highlights

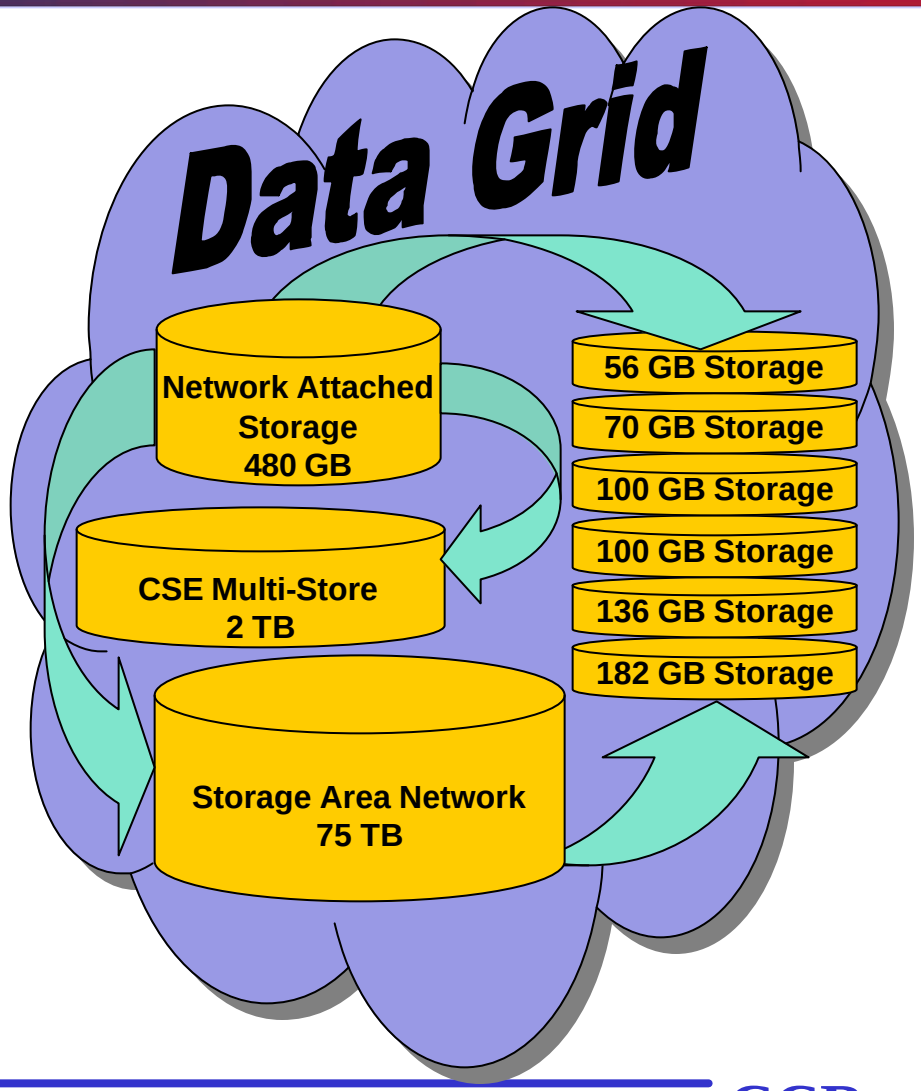
- **Heterogeneous Computational & Data Grid**
- **Currently in Beta with *Shake-and-Bake***
- **WNY Release in March**
- **Bottom-Up General Purpose Implemenation**
  - **Ease-of-Use User Tools**
  - **Administrative Tools**
- **Back-End Intelligence**
  - **Backfill Operations**
  - **Prediction and Analysis of Resources to Run Jobs (Compute Nodes + Requisite Data)**

Data Grid Motivation & Goal

- **Motivation:**
  - **Large data collections are emerging as important community resources.**
  - **Data Grids inherently complements Computational Grids, which manipulate data.**
  - ***A data grid denotes a large network of distributed storage resources such as archival systems, caches, and databases, which are linked logically to create a sense of global persistence.***
- **Goal:**
  - **To design and implement transparent management of data distributed across heterogeneous resources, such that the data is accessible via a uniform web interface.**

Data Grid Summary

- 544 GB Storage
  - Located on 6 heterogeneous ACDC-Grid resources
- 480 GB Storage
  - Located on 1 dual processor Dell PowerVault server
- 75,000 GB Storage (10/03)
  - Served by 4 – 16 processor HP GS1280 servers
- 2,000 GB Storage
  - Served by Sun Ultra-60 servers
- 78,024 GB Total Data Grid Storage available and accessible from the ACDC-Grid Portal

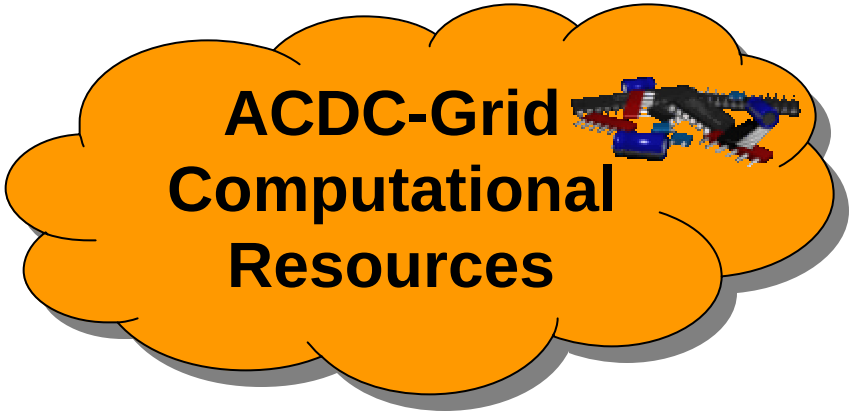


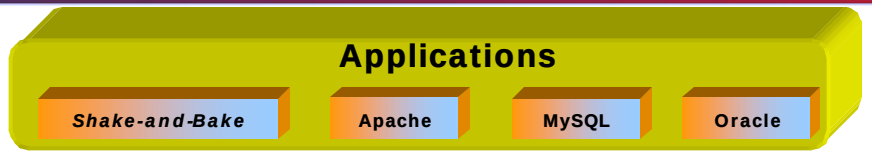
Grid-Based *SnB* Objectives

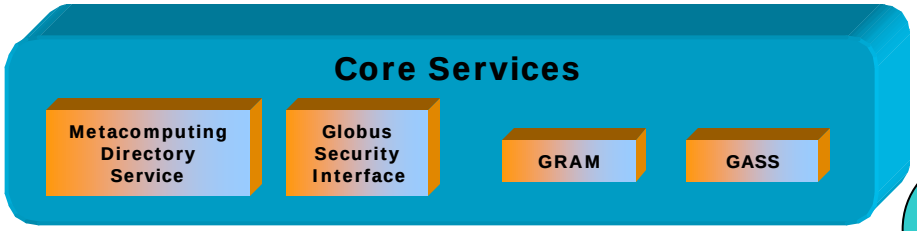
- **Install Grid-Enabled Version of *SnB***
- **Job Submission and Monitoring over Internet**
- ***SnB* Output Stored in Database**
- ***SnB* Output Mined through Internet-Based Integrated Querying Tool**

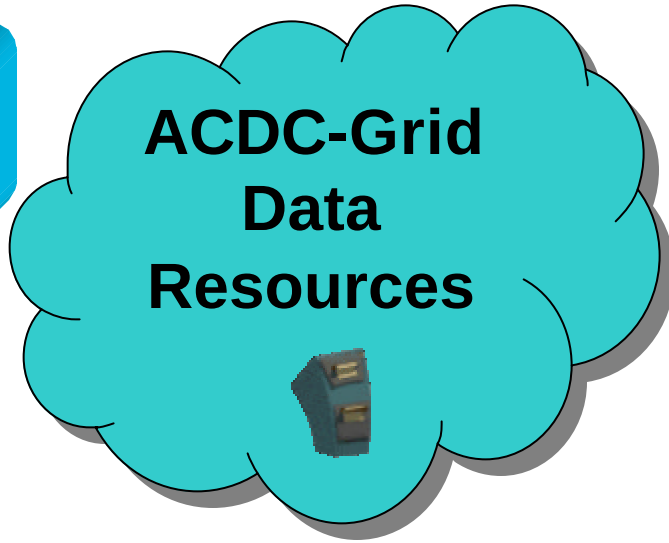
- **Serve as Template for Chem-Grid & Bio-Grid**
- **Experience with Globus and Related Tools**

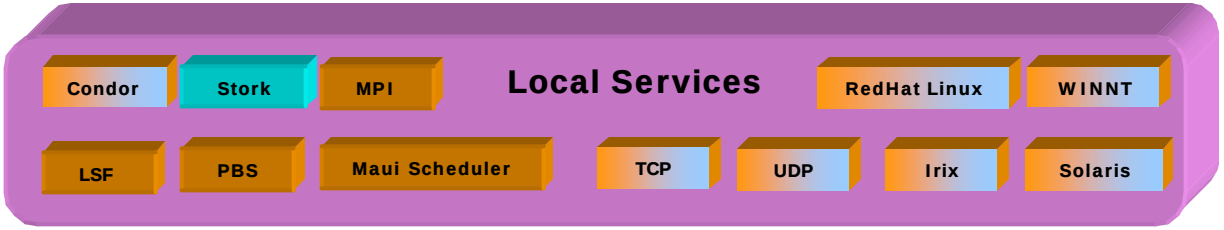
Grid Services and Applications









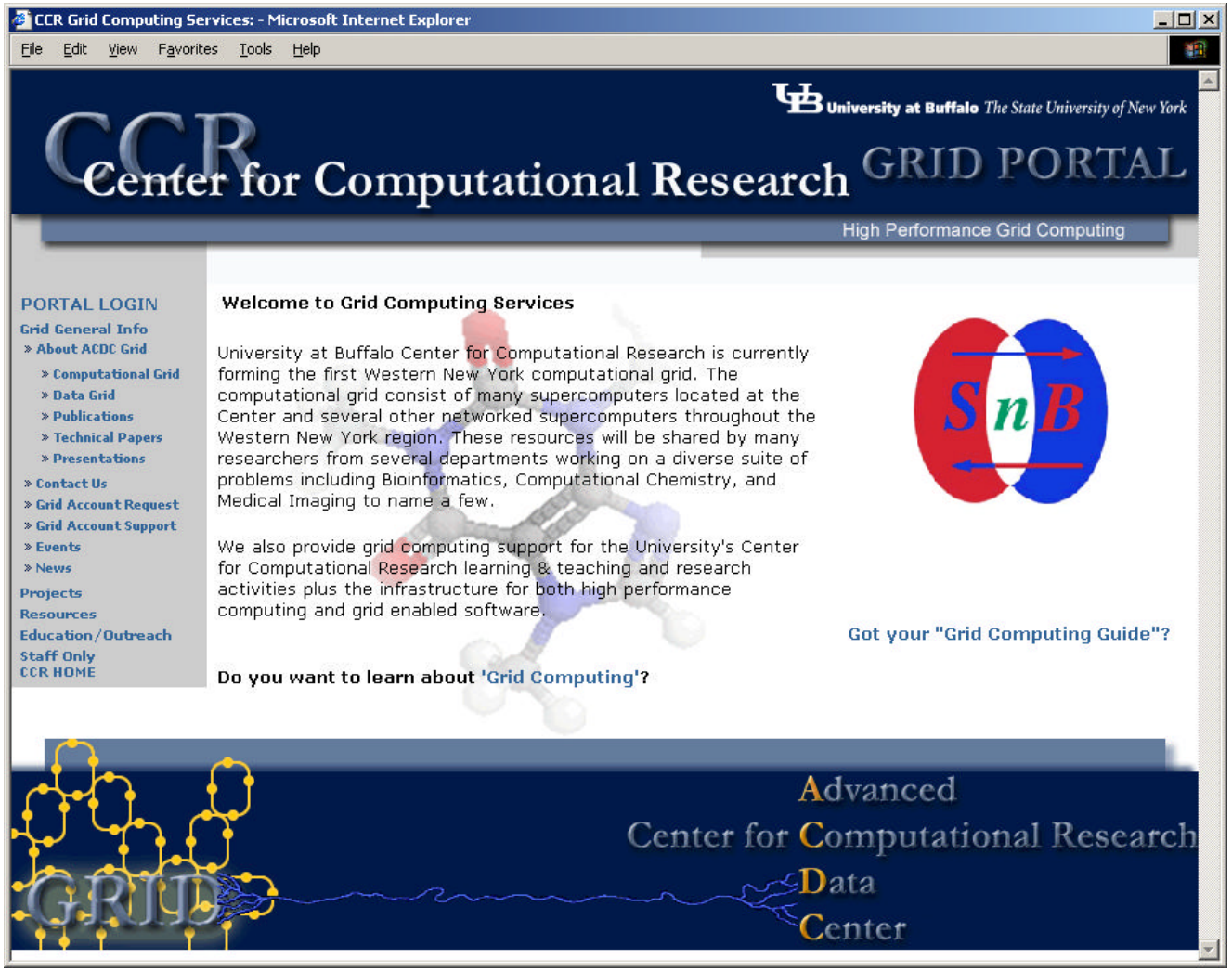


Adapted from Ian Foster and Carl Kesselman

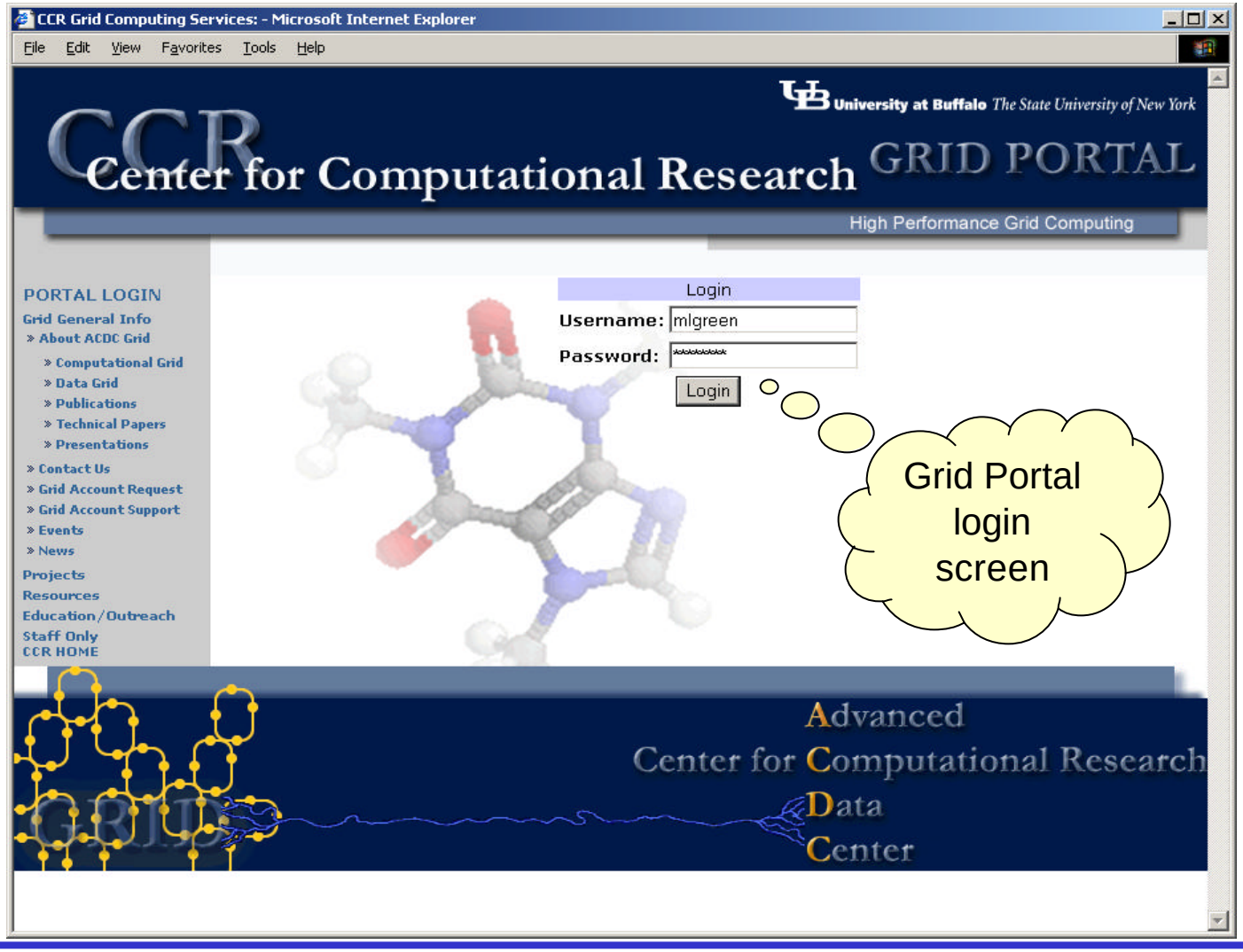
Grid Enabled *SnB* Execution

- **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.**

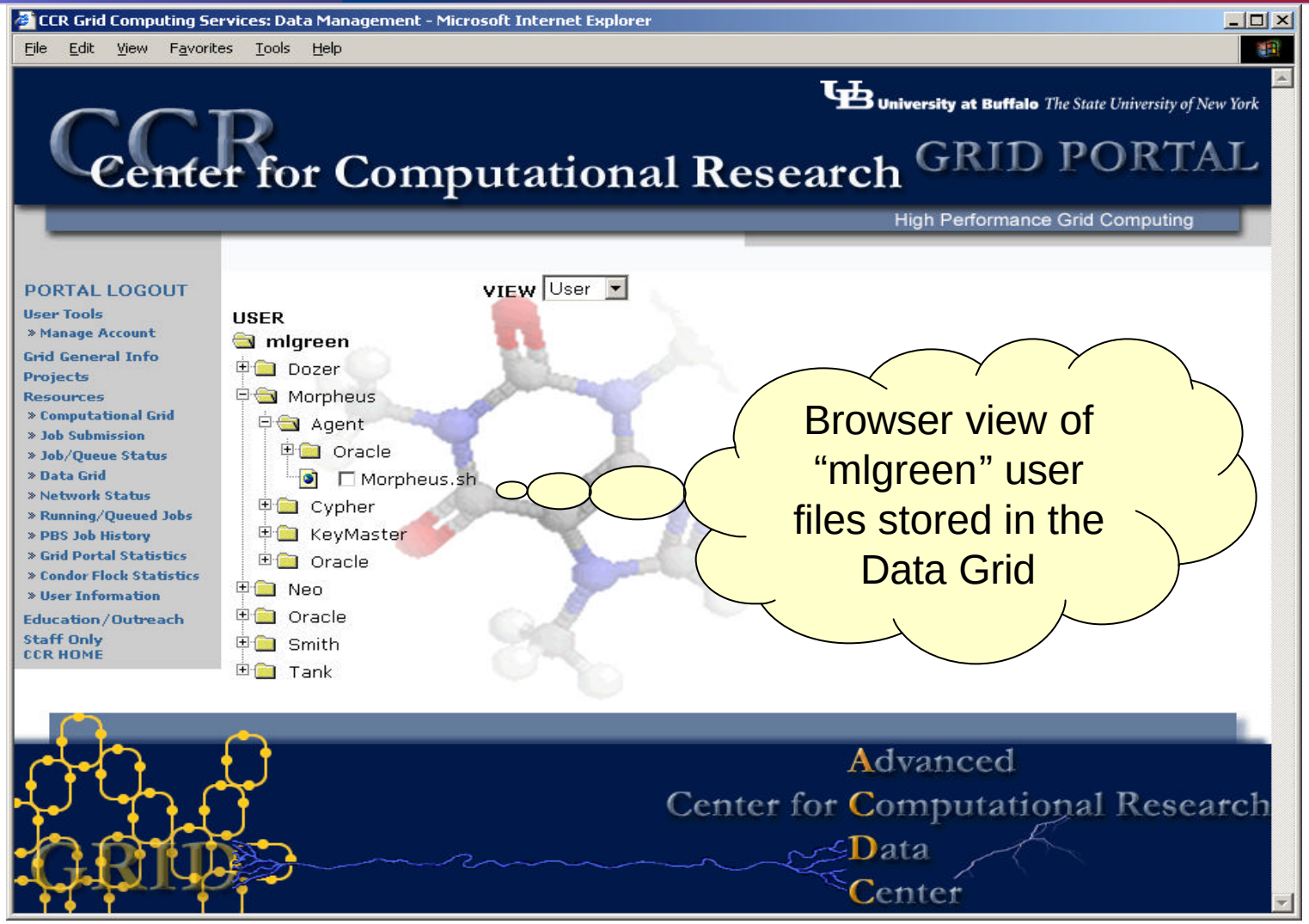
ACDC-Grid Portal



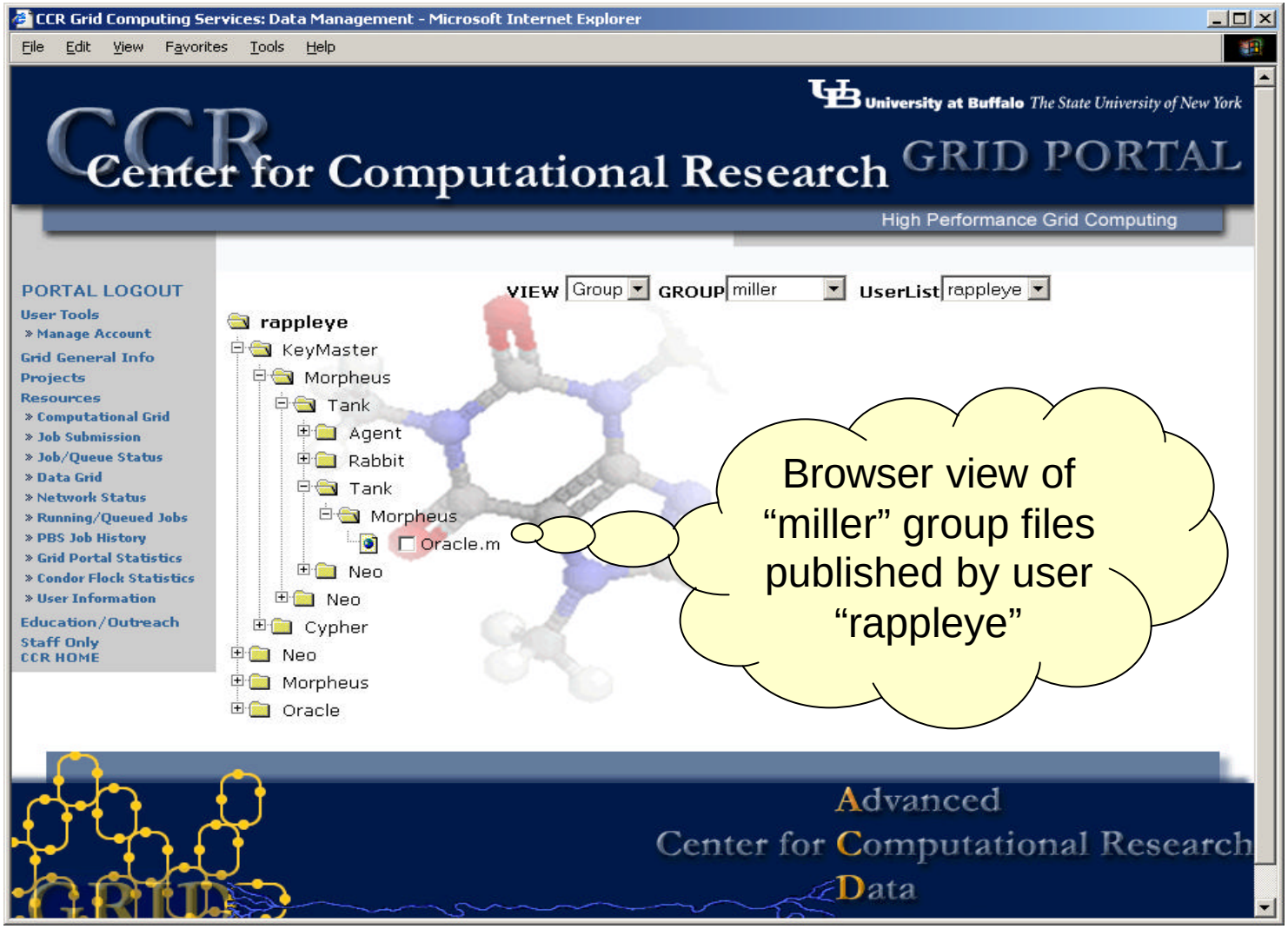
ACDC-Grid Portal Login



Data Grid Capabilities

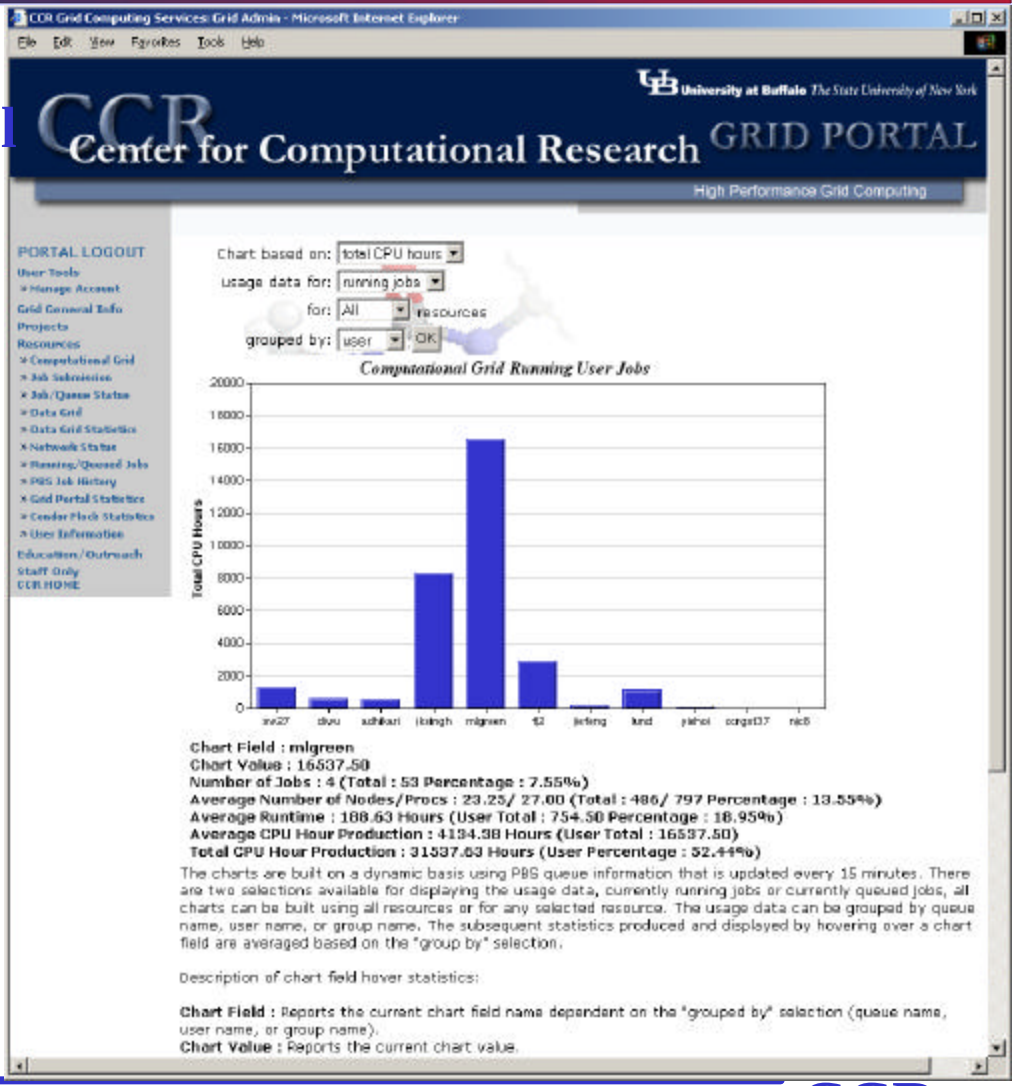


Data Grid Capabilities



Grid Portal Job Status

- **Grid-enabled jobs can be monitored using the Grid Portal web interface dynamically.**
  - **Charts are based on:**
    - **total CPU hours, or**
    - **total jobs, or**
    - **total runtime.**
  - **Usage data for:**
    - **running jobs, or**
    - **queued jobs.**
  - **Individual or all resources.**
  - **Grouped by:**
    - **group, or**
    - **user, or**
    - **queue.**



Grid Portal Job Status

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

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
- 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

Chart based on: total CPU hours

usage data for: queued jobs

for: All resources

grouped by: user OK

Computational Grid Queued User Jobs

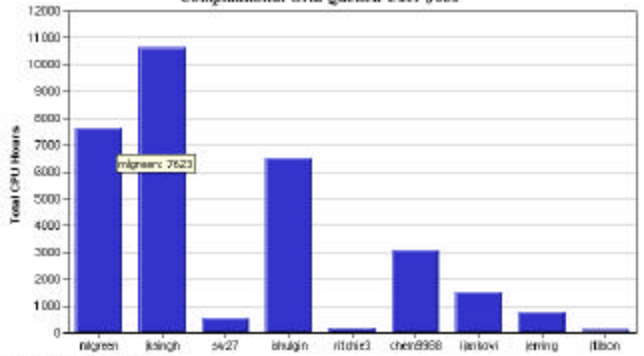


Chart Field : mlgreen
Chart Value : 7623.17
Number of Jobs : 27 (Total : 60 Percentage : 45.00%)
Average Number of Nodes/Procs : 1.96/ 3.93 (Total : 1641/ 3206 Percentage : 3.31%)
Average Runtime : 71.92 Hours (User Total : 1941.75 Percentage : 71.90%)
Average CPU Hour Production : 282.34 Hours (User Total : 7623.17)
Total CPU Hour Production : 31015.17 Hours (User Percentage : 24.58%)

The charts are built on a dynamic basis using PBS queue information that is updated every 15 minut
are two selections available for displaying the usage data, currently running jobs or currently queue
charts can be built using all resources or for any selected resource. The usage data can be groupe
name, user name, or group name. The subsequent statistics produced and displayed by hovering ov
field are averaged based on the "group by" selection.

Description of chart field hover statistics:

Chart Field : Reports the current chart field name dependant on the "grouped by" selection (queue
user name, or group name).
Chart Value : Reports the current chart value.

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

File Edit View Favorites Tools Help

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
- 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

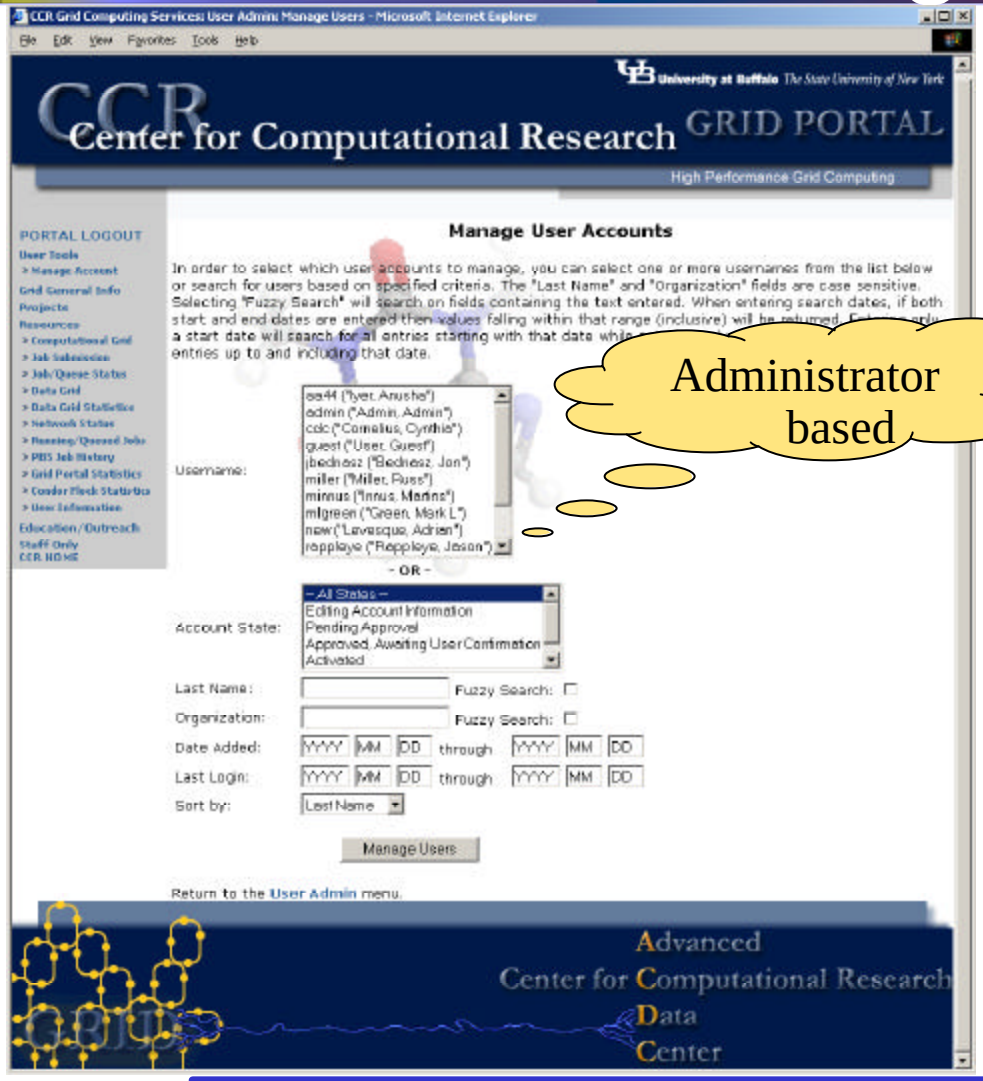
Chart based on: total CPU hours
usage data for: queued jobs
for: all resources
grouped by: user mlgreen

Job_Num	Job_Id	Username	Groupname	Queue	Node_Cnt	Proc_Cnt	Rank	Request_Time(hr)	Start_Time(hr)
1	23364	mlgreen	miler	grid	1	2	1	71.916666666667	queued
2	23365	mlgreen	miler	grid	1	2	2	71.916666666667	queued
3	23366	mlgreen	miler	grid	2	4	3	71.916666666667	queued
4	23367	mlgreen	miler	grid	2	4	4	71.916666666667	queued
5	23368	mlgreen	miler	grid	2	4	5	71.916666666667	queued
6	23369	mlgreen	miler	grid	3	6	6	71.916666666667	queued
7	23370	mlgreen	miler	grid	2	4	7	71.916666666667	queued
8	23371	mlgreen	miler	grid	2	4	8	71.916666666667	queued
9	23372	mlgreen	miler	grid	2	4	9	71.916666666667	queued
10	23373	mlgreen	miler	grid	2	4	10	71.916666666667	queued
11	23374	mlgreen	miler	grid	2	4	11	71.916666666667	queued
12	23375	mlgreen	miler	grid	2	4	12	71.916666666667	queued
13	23376	mlgreen	miler	grid	2	4	13	71.916666666667	queued
14	23377	mlgreen	miler	grid	2	4	14	71.916666666667	queued
15	23378	mlgreen	miler	grid	2	4	15	71.916666666667	queued
16	23379	mlgreen	miler	grid	2	4	16	71.916666666667	queued
17	23380	mlgreen	miler	grid	2	4	17	71.916666666667	queued
18	23381	mlgreen	miler	grid	2	4	18	71.916666666667	queued
19	23382	mlgreen	miler	grid	2	4	19	71.916666666667	queued
20	23383	mlgreen	miler	grid	2	4	20	71.916666666667	queued
21	23384	mlgreen	miler	grid	2	4	21	71.916666666667	queued
22	23385	mlgreen	miler	grid	2	4	22	71.916666666667	queued
23	23386	mlgreen	miler	grid	2	4	23	71.916666666667	queued
24	23387	mlgreen	miler	grid	2	4	24	71.916666666667	queued
25	23388	mlgreen	miler	grid	2	4	25	71.916666666667	queued
26	23389	mlgreen	miler	grid	2	4	26	71.916666666667	queued
27	23393	mlgreen	miler	grid	2	4	27	71.916666666667	queued

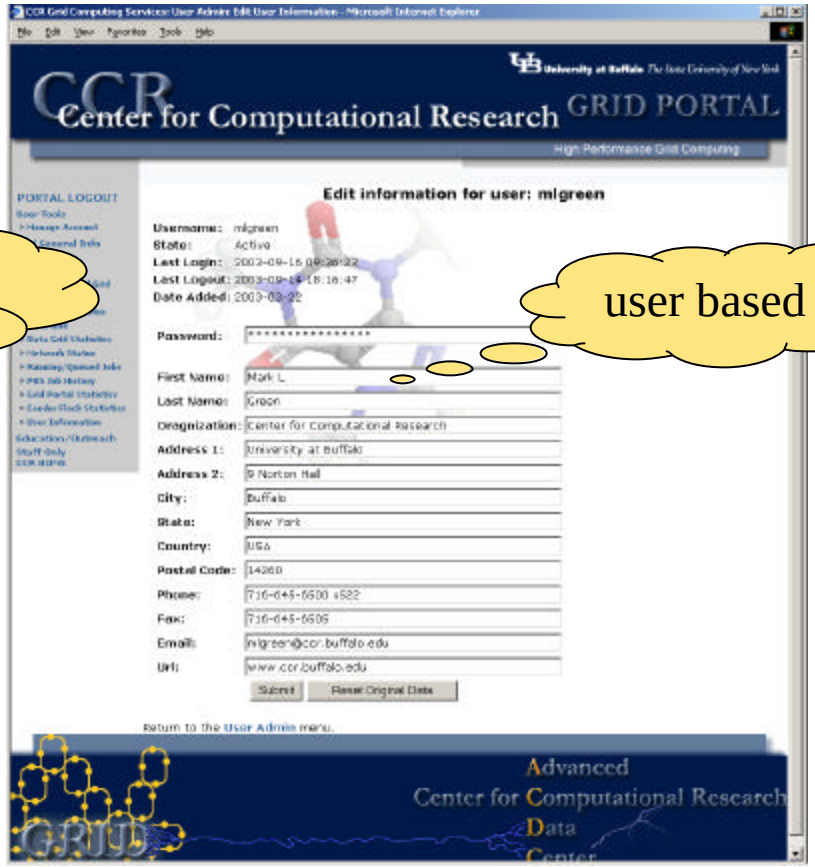
Description of table fields:

Job_Num : Sequential counter for row number.
Job_Id : PBS job identification number corresponding to local PBS queue job number.
Username : PBS job owner local username.
Groupname : PBS job owner local primary group name.

ACDC-Grid Portal User Management

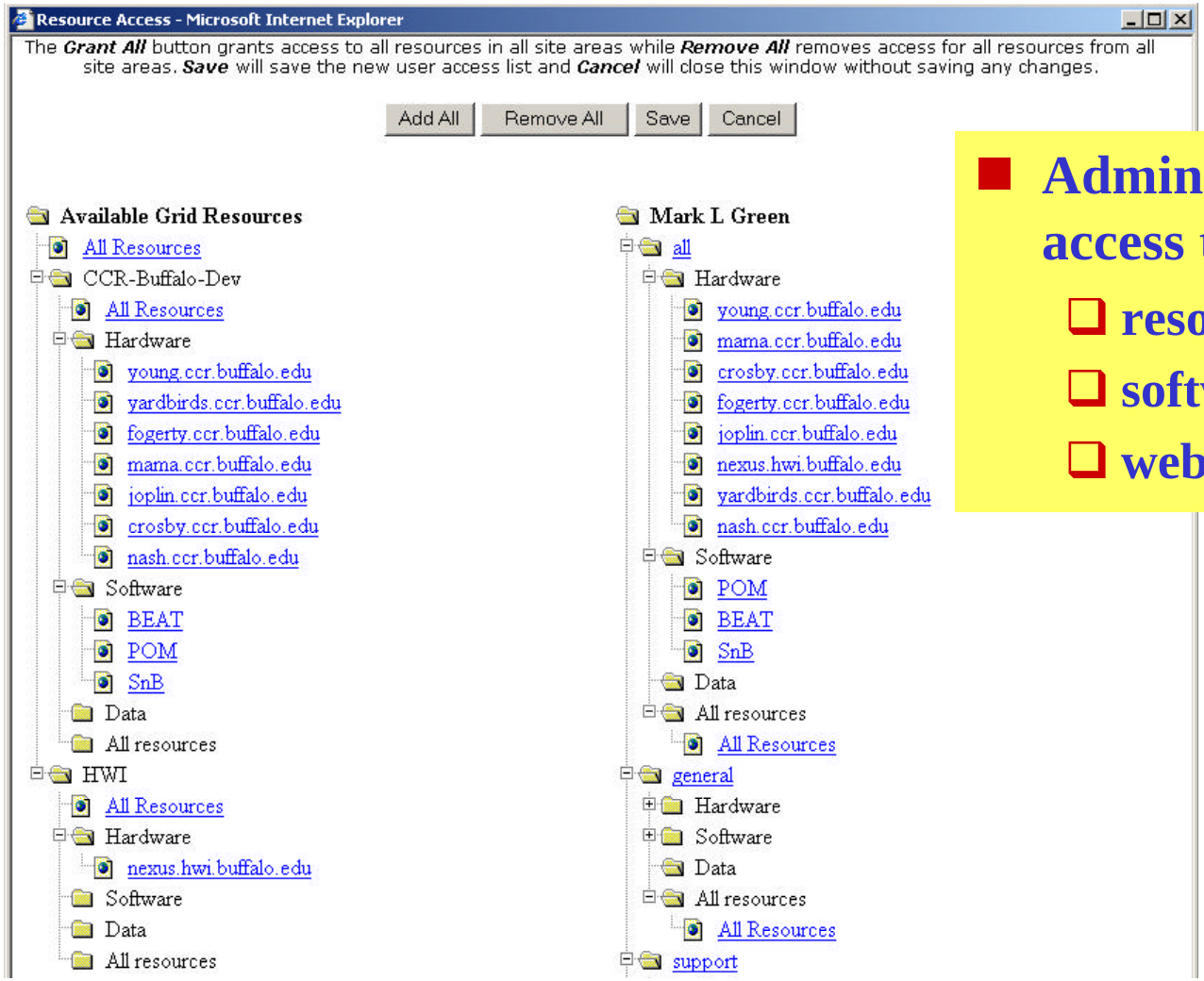


Administrator based



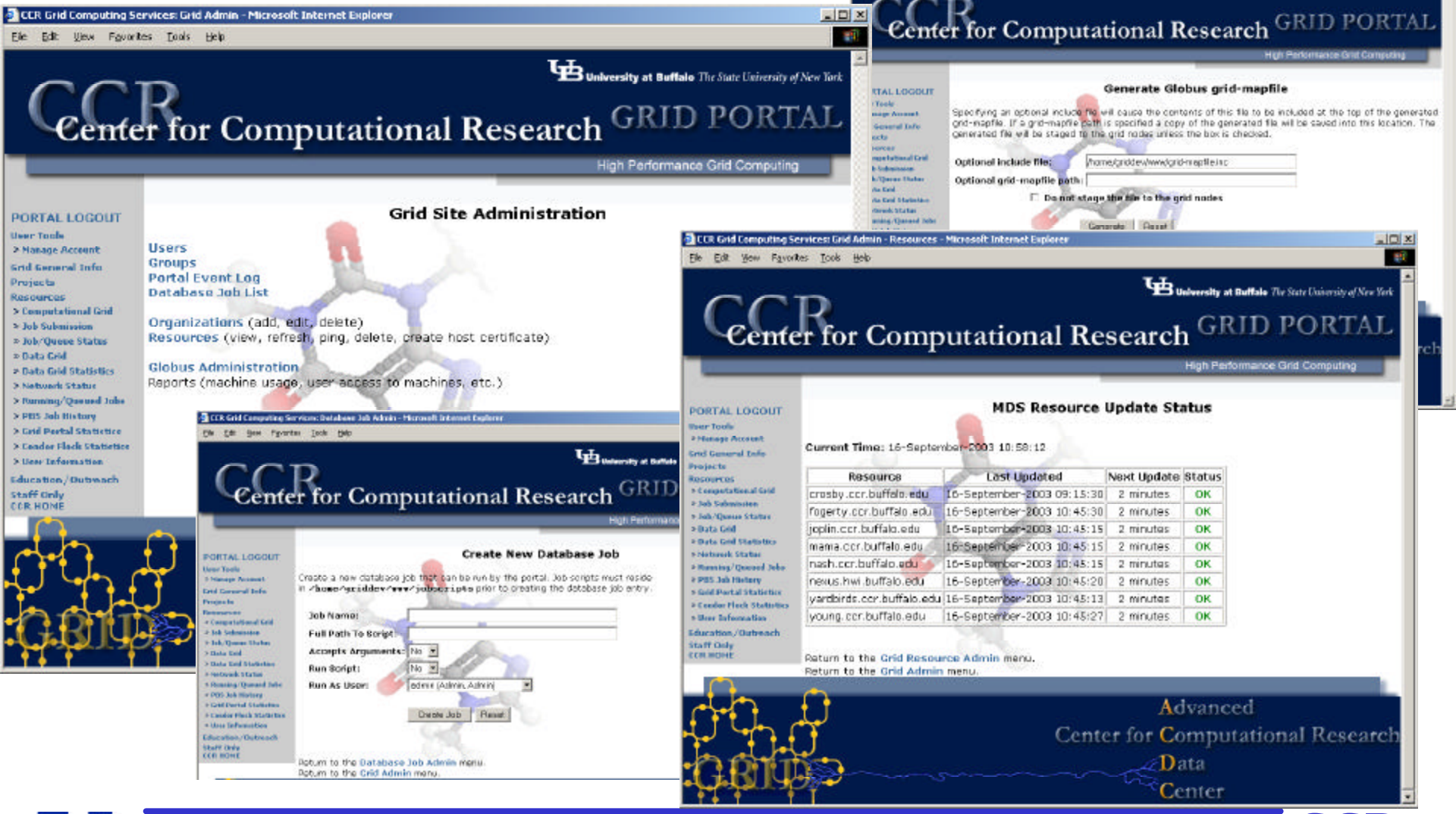
user based

ACDC-Grid Portal
Resource Management



- **Administrator grants a user access to ACDC-Grid**
  - **resources,**
  - **software, and**
  - **web pages.**

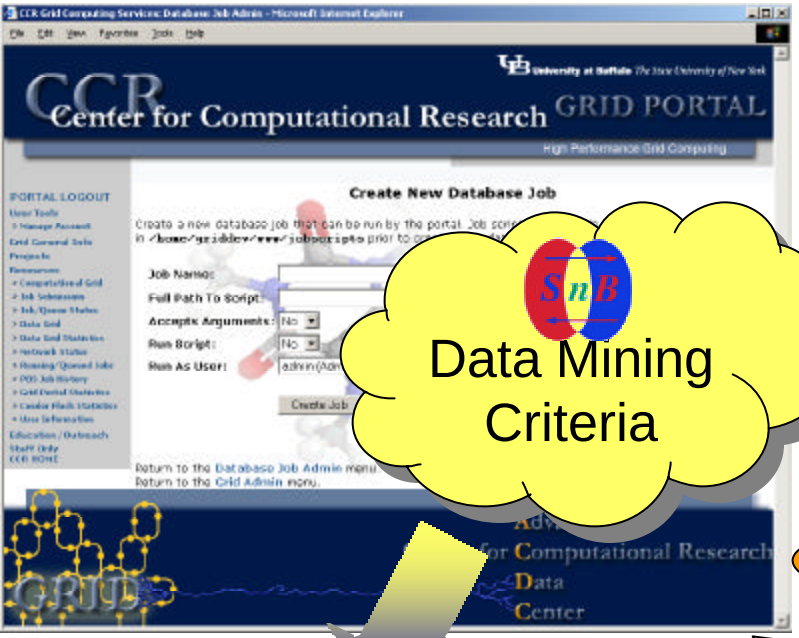
ACDC-Grid Administration

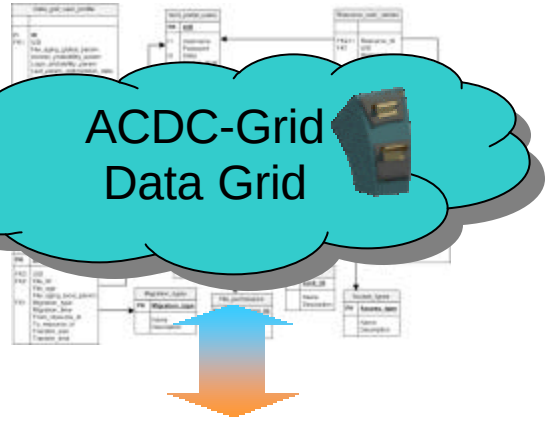


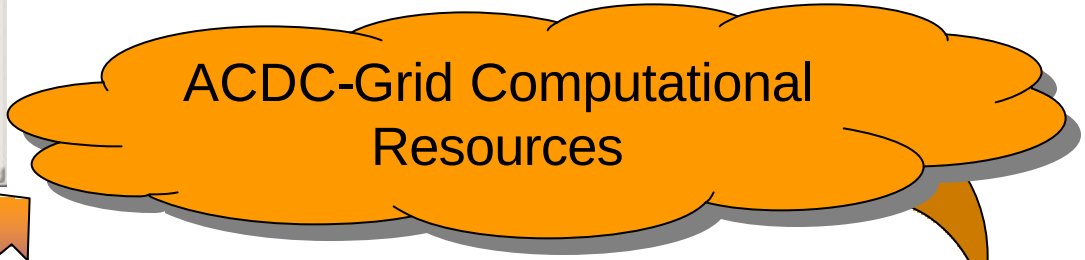
Grid Enabled Data Mining

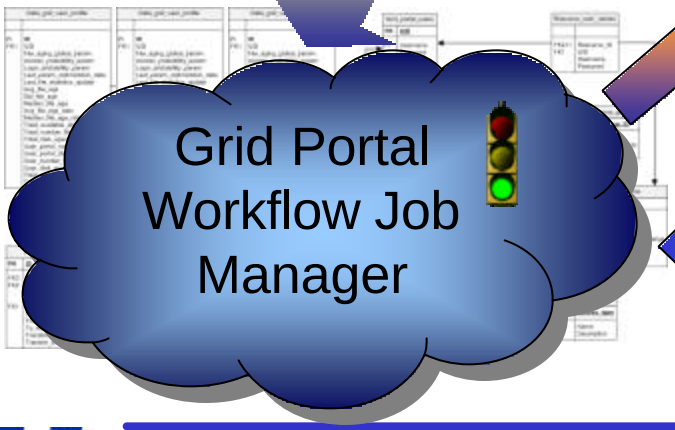
- **Problem Statement**
  - **Use all available resources in the ACDC-Grid for executing a data mining genetic algorithm optimization of *SnB* parameters for molecular structures having the same space group.**
- **Grid Enabling Criteria**
  - **All heterogeneous resources in the ACDC-Grid are capable of executing the *SnB* application.**
  - **All job results obtained from the ACDC-Grid resources are stored in a corresponding molecular structure databases.**

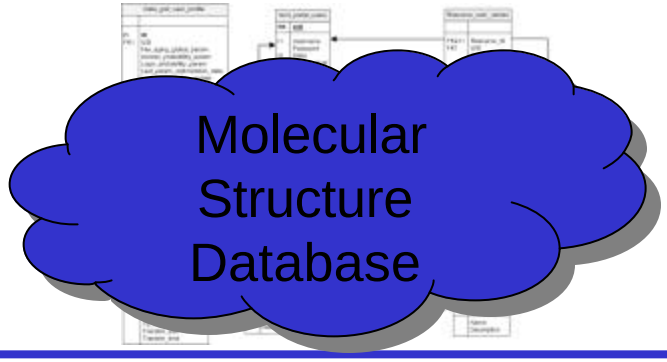
Grid Enabled Data Mining



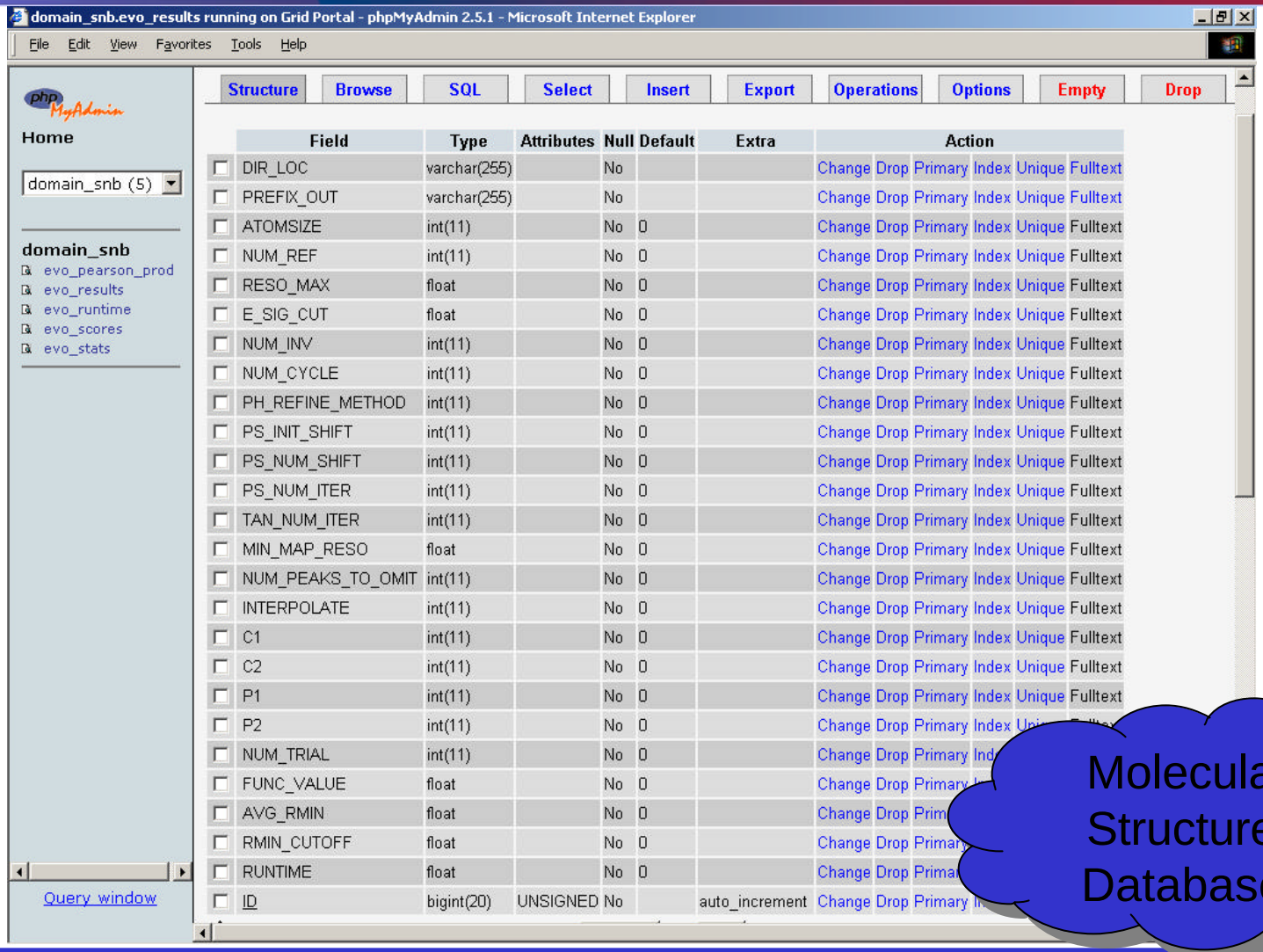






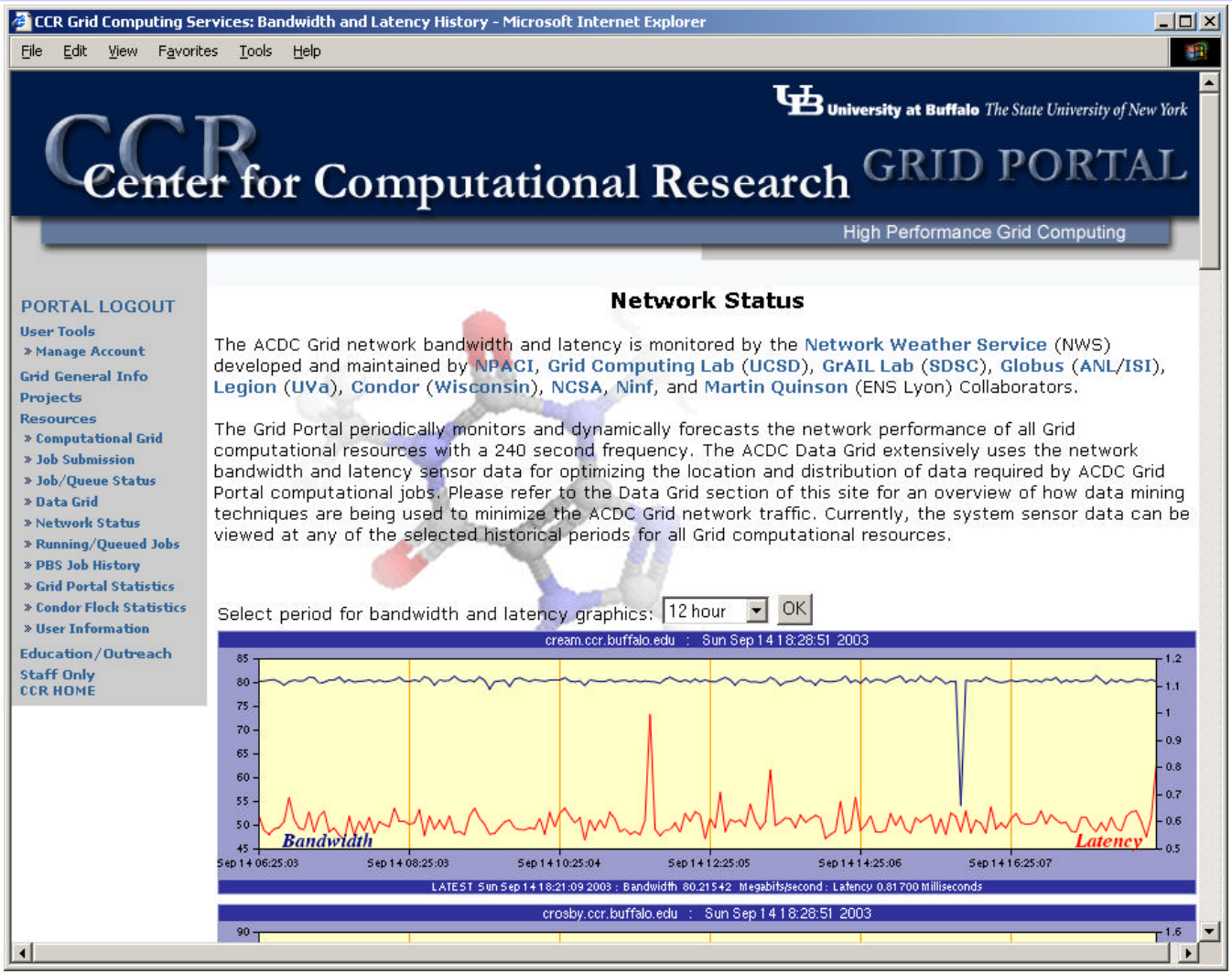


SnB Molecular Structure Database



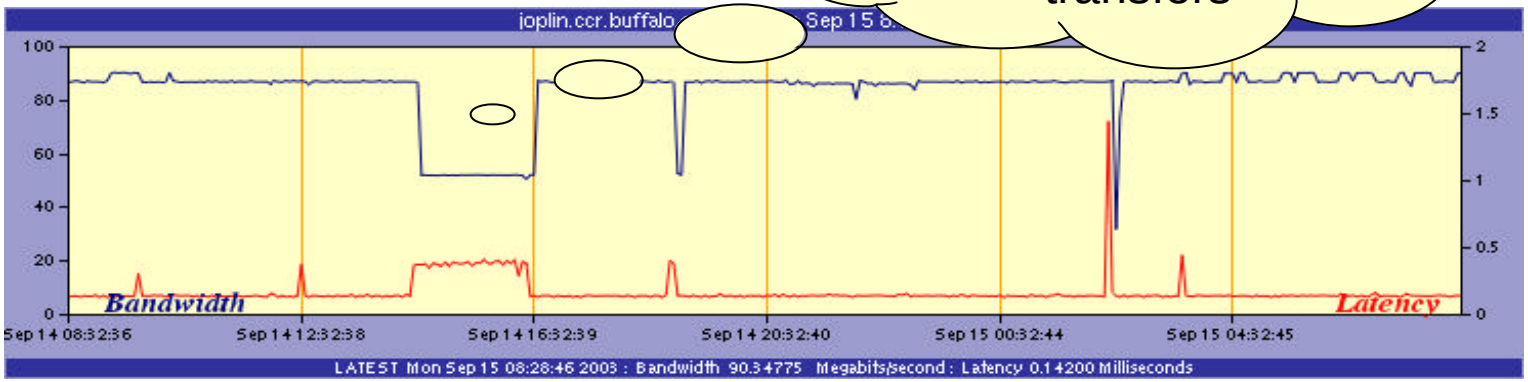
Molecular Structure Database

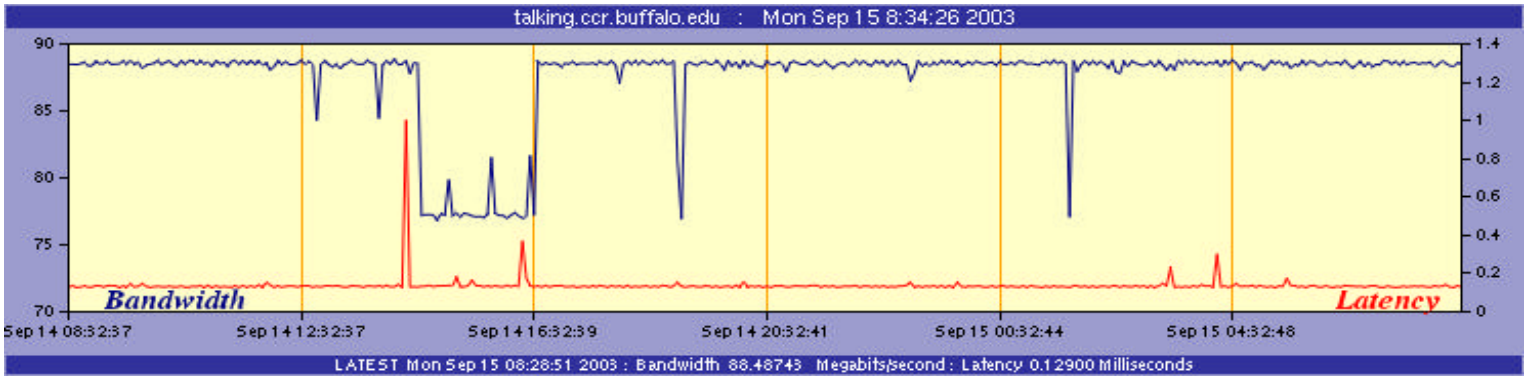
Data Grid Resource Info



Data Grid Resource Info

Both platforms have reduced bandwidth available for additional transfers





Data Grid File Age

Database *data_grid* - Table *file_management* running on *Grid Portal*

Structure | Browse | SQL | Select | Insert | Export | Operations | Options | Empty | Drop

Showing rows 0 - 29 (13000 total, Query took 0.0022 sec)

SQL-query : [Edit] [Explain SQL] [Create PHP Code]
SELECT `File_ID` , resource_id, filename, dir_id, access_time, file_age
FROM `file_management`
WHERE 1 LIMIT 0 , 30

Show : 30 row(s) starting from record # 30
in horizontal mode and repeat headers after 100 cells > >> Page number: 1

		File_ID Unique File ID Value	resource_id	filename	dir_id	access_time	file_age
Edit	Delete	62844	10	Dozer.csh	54030	2003-08-16 11:24:52	925832
Edit	Delete	57120	10	Tank.mpg	53304	2003-04-01 01:45:23	12837001
Edit	Delete	57121	10	Neo.ksh	53505	2003-06-08 06:41:05	6904459
Edit	Delete	57122	10	Trinity.m	53499	2003-07-30 11:34:54	2394030
Edit	Delete	57123	10	Rabbit.ksh	53541	2003-07-13 01:43:34	3898310
Edit	Delete	57124	10	Neo.ksh	53407	2003-06-22 06:19:03	5696181
Edit	Delete	57049	10	Agent.ppt	53928	2003-02-24 12:15:39	15909585
Edit	Delete	61710	10	Neo.txt	52724	2003-07-26 09:44:48	2746236
Edit	Delete	61711	10	Morpheus.sh	52710	2003-07-31 07:03:43	2367101
Edit	Delete	61712	10	Morpheus.ppt	52761	2003-08-26 08:04:38	117046
Edit	Delete	61713	10	Tank.jpg	52929	2003-06-26 09:59:37	5337347
Edit	Delete	61714	10	Rabbit.dat	52624	2003-08-26 05:57:43	124661
Edit	Delete	61715	10	KeyMaster.mpg	52770	2003-06-17 04:16:44	6178720

- **File age, access time, and resource id denote:**
  - **the amount of time since a file was accessed,**
  - **when the file was accessed, and**
  - **where the file currently resides respectively.**

ACDC-Grid Development/Maintenance

- **Development Requirements**
  - 7 – Person months for Grid Services Coordinator
    - Including Grid and Database conceptual design and implementation
  - 5 – Person months for Grid Services Programmer
    - Web portal programming
  - 5 – Person months for System Administrator
    - Globus, NWS, MDS, etc. installations
  - 3 – Person months for Database Administrator
    - Grid Portal Database implementation
- **Minimum Maintenance Requirements**
  - 1 – Grid Services Coordinator
    - 100% level of effort
  - 1 – Grid Services Programmer
    - 100% level of effort
  - 1 – System Administrator
    - 50% level of effort
  - 1 – Database Administrator
    - 10% level of effort

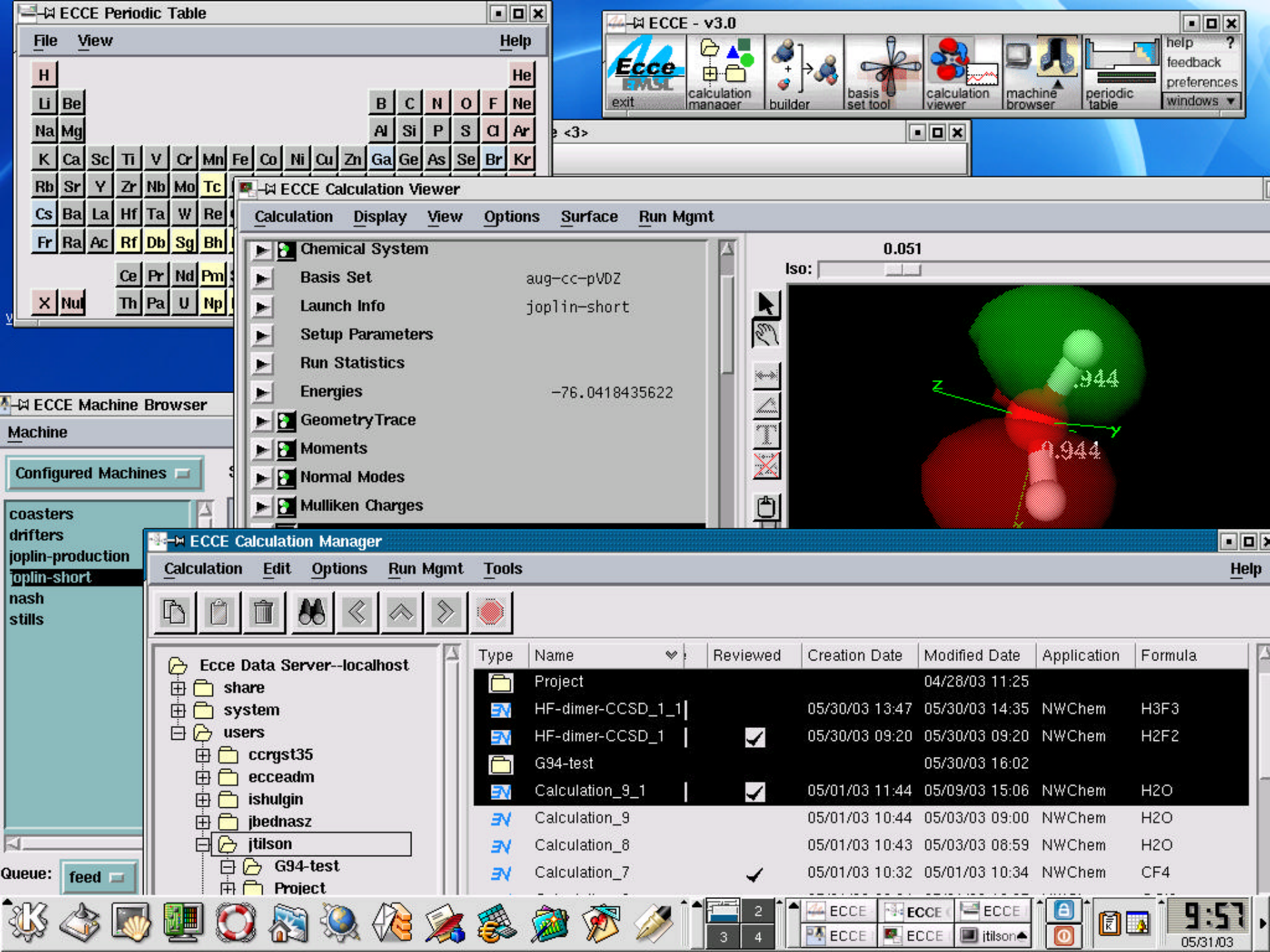
Future ACDC Applications

- Princeton Ocean Model (POM)
- Genetic Algorithms for Earthquake Structural Design
- Bioinformatics
- Computational Chemistry (Q-Chem)
- Environmental Engineering Applications

ECCE "Grid" at CCR

- Computational Chemistry
  - Relativistic effects/Heavy elements
  - Algorithm development
  - Theoretical physical chemistry
- Structural/Systems Biology
  - Protein structure
  - Enzyme catalysis
- Chemical Engineering
  - Condensed phases/Mixed phase predictions
  - Catalysis
- Geology, Pharmacology, Medical School

- Import Scientific Information
  - Application independent input
  - ECCE automatically formats for target application (Gaussian98, NWChem)
- Computing at CCR
  - 881 available CPUs (>2.5TFlops)
    - (Xeon, P3, Power3, R12K)
  - Uniform access to all platforms via ECCE "job launcher"
- Chemical Analysis
  - Full complement of visual tools for understanding data/publication quality graphics


ECCE Periodic Table
File View Help
ECCE - v3.0
exit
calculation manager
builder
basis set tool
calculation viewer
machine browser
periodic table
help
feedback
preferences
windows
ECCE Calculation Viewer
Calculation Display View Options Surface Run Mgmt
Chemical System
Basis Set aug-cc-pVDZ
Launch Info joplin-short
Setup Parameters
Run Statistics
Energies -76.0418435622
GeometryTrace
Moments
Normal Modes
Mulliken Charges
0.051
Iso:
ECCE Machine Browser
Machine
Configured Machines
coasters
drifters
joplin-production
joplin-short
nash
stills
Queue: feed
ECCE Calculation Manager
Calculation Edit Options Run Mgmt Tools Help
Ecce Data Server--localhost
share
system
users
ccrgst35
ecceadm
ishulgin
jbednasz
jtilson
G94-test
Project
Type Name Reviewed Creation Date Modified Date Application Formula
Project 04/28/03 11:25
HF-dimer-CCSD_1_1 05/30/03 13:47 05/30/03 14:35 NWChem H3F3
HF-dimer-CCSD_1 05/30/03 09:20 05/30/03 09:20 NWChem H2F2
G94-test 05/30/03 16:02
Calculation_9_1 05/01/03 11:44 05/09/03 15:06 NWChem H2O
Calculation_9 05/01/03 10:44 05/03/03 09:00 NWChem H2O
Calculation_8 05/01/03 10:43 05/03/03 08:59 NWChem H2O
Calculation_7 05/01/03 10:32 05/01/03 10:34 NWChem CF4
9:57
05/31/03

Acknowledgments

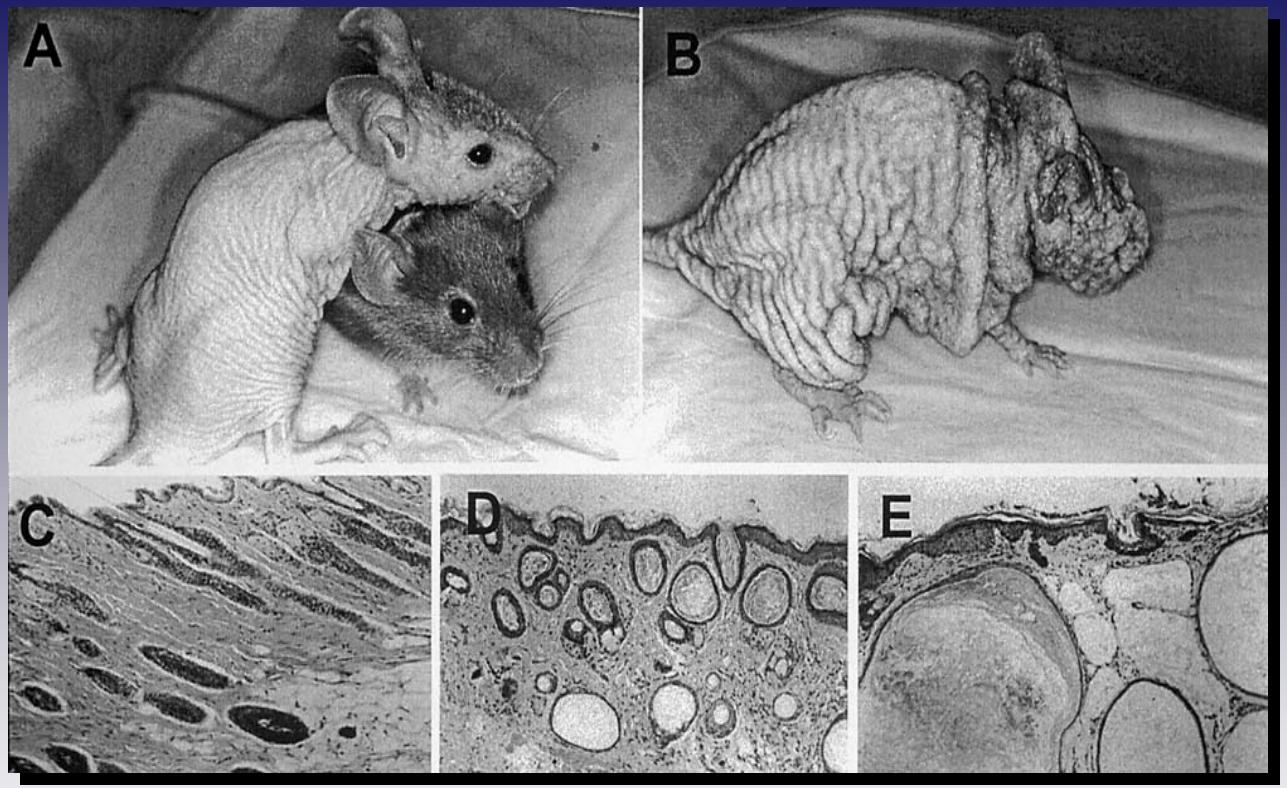
- Mark Green
- Steve Gallo
- Jason Rappleye
- Jeff Tilson
- Martins Innus
- Cynthia Cornelius

- Betty Capaldi
- Bruce Holm
- Janet Penksa

- George DeTitta
- Herb Hauptman
- Charles Weeks
- Steve Potter

- Philip Glick

Contact Information



miller@buffalo.edu
www.ccr.buffalo.edu