A Status Report on the Prototype
NYS Grid: September 21, 2006

Russ Miller
Comp Sci & Eng, SUNY-Buffalo
Hauptman-Woodward Medical Res Inst
Cyberinfrastructure (e-Science)

- CI, HPC, & CSE are Critical to 21st Century
  - Discovery
  - Economic Development
  - EOT
- Digital Data-Driven Society
- Knowledge-Based Economy
- Mission
  - Seamless, Ubiquitous, Secure, Interwoven, Dynamic
    - Compute Systems, Storage, Instruments, Sensors
    - Computational Methodologies (Algorithms)
    - Networking
    - HCI
- Immediate Goals Include
  - Develop Software, Algorithms, Portals, Interfaces
Grid Computing Overview

- 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
Grid Services and Applications

ACDC-Grid
Computational Resources

Applications
Shake-and-Bake  Apache  MySQL  Oracle

High-level Services and Tools
Globus Toolkit  MPI  MPI-IO  C, C++, Fortran, PHP  globusrun

Core Services
Metacomputing Directory Service  Globus Security Interface  GRAM  GASS

Local Services
Condor  Stork  LSF  MPI  PBS  Maui Scheduler  RedHat Linux  WINNT  TCP  UDP  Irix  Solaris

ACDC-Grid
Data Resources

Adapted from Ian Foster and Carl Kesselman
“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
NSF Middleware Initiative (NMI)

- Develop, improve, and deploy a suite of reusable software components for use in national-scale “cyberinfrastructure”.
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
Organization of Cyberinstitute at SUNY-Buffalo

HPC (Furlani: CCR)
- Computing
- Data
- Visualization
- Networking

CSE
- MultiScale
- Sciences
- Engineering
- Life Sciences
- Media

CI
- Scheduling
- Monitoring
- Virtual Reality

Enabling
- Programmers
- GUI Design
- Integration
Proposed Organization of CyberInstitute of New York State

CINYS

NEW YORK STATE

HPC
- Computing
- Data
- Visualization
- Networking

CSE
- MultiScale
- Sciences
- Engineering
- Life Sciences
- Media

CI
- Scheduling
- Monitoring
- Virtual Reality

Enabling
- Programmers
- GUI Design
- Integration

Don’t Panic – 1 of 4
Proposed Organization of CyberInstitute of New York State

CINYS

NEW YORK STATE

NYS Resources
- Computing
- Data Storage
- Networking
- Rendering
- Sensors
- Instruments

CSE
- MultiScale
- Sciences
- Engineering
- Life Sciences
- Media

CI
- Scheduling
- Monitoring
- Virtual Reality

Enabling
- Programmers
- GUI Design
- Integration

Don’t Panic – 2 of 4
Proposed Organization of CyberInstitute of New York State

NYS Resources
- Computing
- Data Storage
- Networking
- Rendering
- Sensors
- Instruments

NYS CSE
- MultiScale
- Sciences
- Engineering
- Life Sciences
- Media

NYS CI
- Scheduling
- Monitoring
- Networking
- Security

CINYS Enablers
- Programmers
- GUI Design
- Integration

Don’t Panic – 3 of 4
 Courtesy of NYSErNet
Proposed Organization of CyberInstitute of New York State

NYS Resources
- Computing
- Data Storage
- Networking
- Rendering
- Sensors
- Instruments

NYS CSE
- MultiScale
- Sciences
- Engineering
- Life Sciences
- Media

NYS CI
- Scheduling
- Monitoring
- Networking
- Security

CINYS Enablers
- Programmers
- GUI Design
- Integration

ExCom

CINYS Director

NYSERNet

NEW YORK STATE

Yell & Scream – 4 of 4
Proposed CINYS Budget

Participants
- Standard Resources
  - Compute Systems
  - Data Storage
  - Visualization Devices
  - Sensors
  - Internet-Ready Devices
- Percent FTE Sysadmin
- Faculty Research Groups
- Intellectual Capital

New York State
- Seed Funds
- Special Resources
  - Networking
  - Large Data Storage
- Operating Budget
  - Enablers/Programmers
  - Access Grid Nodes
  - General Operating
    - Travel
    - Training
Federal Funding Opportunity

**NSF “High-Performance Computing for Science and Engineering Research and Education: Operations (User Support, System Administration and Maintenance)”**

- NSF 06-599
- Due Nov 28
- $2-10M/year
- Integrate with TeraGrid
- Min of 5TF *Sustained*
- Min 50% of Machine
NYSGrid Status: Implementation Details

(Jon Bednasz & Steve Gallo)

- Getting Started (Build Cluster, etc.)
- Resource Manager
- Install OSG software stack
- Request/install Host certificate
- Configure OSG
- Local Site-verify.sh
- Grid-cat monitoring
- ACDC monitoring
- Remote Site-verify.sh
Getting Started

- 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

- 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
Obtain OSG Secure Certificate

- Request integration of new resource into OSG
- Request host certificate to identify resource
- Approved host certificate is imported into web browser
- Export host certificate from web browser
- Install host certificate on resource
Configure OSG

- Configure OSG
  - cd $VDT_LOCATION/monitoring
  - ./configure-osg.sh

- Determine and configure OSG to use range of firewall ports.
Local Site-Verify.sh

$ cd $VDT_LOCATION
$ source ./setup.sh
$ grid-proxy-init
Enter "Your Passphrase"
$ cd verify
$ perl site_verify.pl --
host=hostname.domain.tld
Remote Site-Verify.sh

- From another OSG site:

  $ cd $VDT_LOCATION
  $ source ./setup.sh
  $ grid-proxy-init
  Enter "Your Passphrase"
  $ cd verify
  $ perl site_verify.pl --
  host=hostname.domain.tld
Status I

- **On-Line**
  - RIT (later today)
  - SUNY-Albany
  - SUNY-Binghamton
  - SUNY-Buffalo
  - Syracuse University

- **Waiting on Host Cert**
  - Niagara University
  - NYU

- **Installing**
  - University of Rochester
  - Cornell University
Status II

- Building Gatekeeper Machine
  - Hauptman-Woodward Medical Research Institute (HWI)
  - SUNY-Geneseo
Acknowledgments

- Jon Bednasz, CCR
- Steve Gallo, CCR
- Mark Green, ITR/CCR
- Cathy Ruby, ITR
- Amin Ghadersohi, ITR
- Naimesh Shah, ITR
- Jason Rappleye, CCR
- Sam Guercio, CCR
- Martins Innus, CCR
- Cynthia Cornelius, CCR
- Tom Furlani, CCR
- NSF, NIH, NYS, NIMA, NTA, Oishei, Wendt, DOE
## Acknowledgments II

- **UAlbany**: Eric Warnke
- **RIT**: Rick Bohn
- **SU**: Jorge González Outeiriño
- **NYU**: Chris Grim
- **U of R**: Bill Webster
- **Cornell**: Resa Alvord
- **Binghamton**: Steaphan Greene

- **Niagara U**: Ann Rensel
- **HWI**: Steve Potter
- **Geneseo**: Kirk Anne
Acknowledgments III

- Cornell (CTC & Administration): Vision, Leadership, & Guts to bring us all together in the hope that by working together we, and many others in NYS, will all prosper.
- Cornell: Hosting Workshop I
- RPI: Hosting Workshop II