



Welcome!

Data Analysis on Ranger
October 23-24



About the Center for Advanced Computing

Mission

*Enable the Success of Cornell Researchers, Collaborators,
and Supporters whose Work Demands Advanced-
Computing Solutions*



Organization Overview

- Cornell core facility providing advanced computing services
 - Consulting
 - Computing
 - Database & Data Storage Resources
 - Visualization
- Staff of expert consultants, systems administrators & programmers
- CAC Director reports to the Vice Provost for Research
- CAC has a Faculty Oversight Committee, chaired by VPR
 - Includes leaders in the fields of engineering, life sciences, social sciences, computer science, & business



Highly Skilled Staff

- Consulting
 - Offer high-quality technical support on demand
 - Partner with research groups, collaborators, and supporters on long term projects
- Systems Administration
 - Windows
 - Linux
 - Max OSX
- Programming
 - Parallelization
 - Optimization
 - Scalability
- Database
 - Schema Design
 - Data Pipelines & Workflows
- Rapid Adopters of New Technologies
 - Leverage new technologies to create new and innovative solutions
 - Develop prototypes using best-of-breed or emerging technologies



Data Analysis – The Growing HPC Challenge

- Modern Research is Producing Massive Amounts of Data
 - Microscopes
 - Telescopes
 - Gene Sequencers
 - Mass Spectrometers
 - Satellite Images
 - Distributed Weather Sensors
 - High Performance Computing (especially HPC Clusters)
- Research Communities Rely on Distributed Data Sources
 - Collaboration
 - Virtual Laboratory's
 - Laboratory Information Management Systems (LIMS)
- New Management and Usage Issues
 - Security
 - Reliability/Availability
 - Manageability
 - Data Locality – You can't ftp a petabyte to your laptop....



Data Intensive Application Characteristics

- Applications which generate or analyze large volumes of data
- Applications requiring a very large memory/core ratio
- Applications requiring interactive or on-demand access to resources
- Analysis of streaming data and distributed data collections



Agenda Overview – Day 1 – Thursday, 10/23/2008

- 8:30 Welcome - *David Lifka*
- 9:00 Visualization Systems and Software – *Andrew Dolgert*
- 10:00 Lab & Demo: Visualizing with ParaView and VisIT – *Andrew Dolgert*
- 11:00 Lab: Remote and Parallel Visualization
- 12:00 Lunch
- 1:15 Processing Radio Astronomy data from the
Arecibo Observatory – *John Zollweg*
- 2:15 Using Cayuga to Analyze Streaming Data – *David Lifka*
- 3:15 Break and room change
- 3:30 Relational Databases, Not your Father's Flat Files – *Linda Woodard*
- 4:30 Adjourn



Agenda Overview – Day 2 – Friday, 10/24/2008

- 8:30 Science Gateways Overview – *David Lifka*
- 9:30 Developing and Accessing Science Gateways for Large Data Applications – *David Lifka*
- 11:00 Guiding Your Computational Work:
HPC Application Steering – *Nate Woody*
- 12:00 Lunch
- 1:15 Data Formats for Visualization and Interoperability – *Steve Lantz*
- 2:15 Break
- 2:30 Data Transfer Options for Ranger – *John Zollweg*
- 3:30 Adjourn



Before We Get Started

- Getting an Allocation & Account on Ranger
 - TeraGrid Getting Started Guide
 - http://www.teragrid.org/userinfo/getting_started.php
 - Getting Started with Ranger Virtual Workshop
 - <https://www.cac.cornell.edu/ranger/registration/>
- TACC User Portal
 - <https://portal.tacc.utexas.edu/gridsphere/gridsphere>
- Transferring Data to and from Ranger
 - <http://www.teragrid.org/userinfo/data/gridftp.php>