

Cornell University
Center for Advanced Computing



Cornell University
Center for Advanced Computing

Welcome!

Data Analysis on Ranger

October 12-13, 2009



Center for Advanced Computing (CAC)

- Mission:
 - *To Enable the Success of Cornell Researchers, Collaborators, and Supporters whose Work Demands Advanced-Computing Solutions*
- Organization:
 - Cornell core facility
 - 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



CAC Services

- Maintenance and support of general-purpose, high-performance computing systems
 - <http://www.cac.cornell.edu/services/HPCsystemsList.aspx>
 - Faculty and research group assistance in the preparation of proposals
- Research project participation and consulting
 - Parallel code development/optimization/extensive debugging or development
 - Database design, implementation and maintenance
 - Develop pipelines for data intensive research

ranger.tacc.utexas.edu

- Ranger is a open science research TeraGrid resource built and operated by TACC as the first of the new NSF Track2 HPC acquisitions.
- <http://www.tacc.utexas.edu/resources/hpcsystems/#constellation>
 - Operating System: Linux
 - Number of Nodes: 3,936
 - Number of Processing Cores: 62,976
 - Total Memory: 123TB
 - Peak Performance: 579.4TFlops
 - Total Disk: 1.73PB (shared)
31.4TB (local)



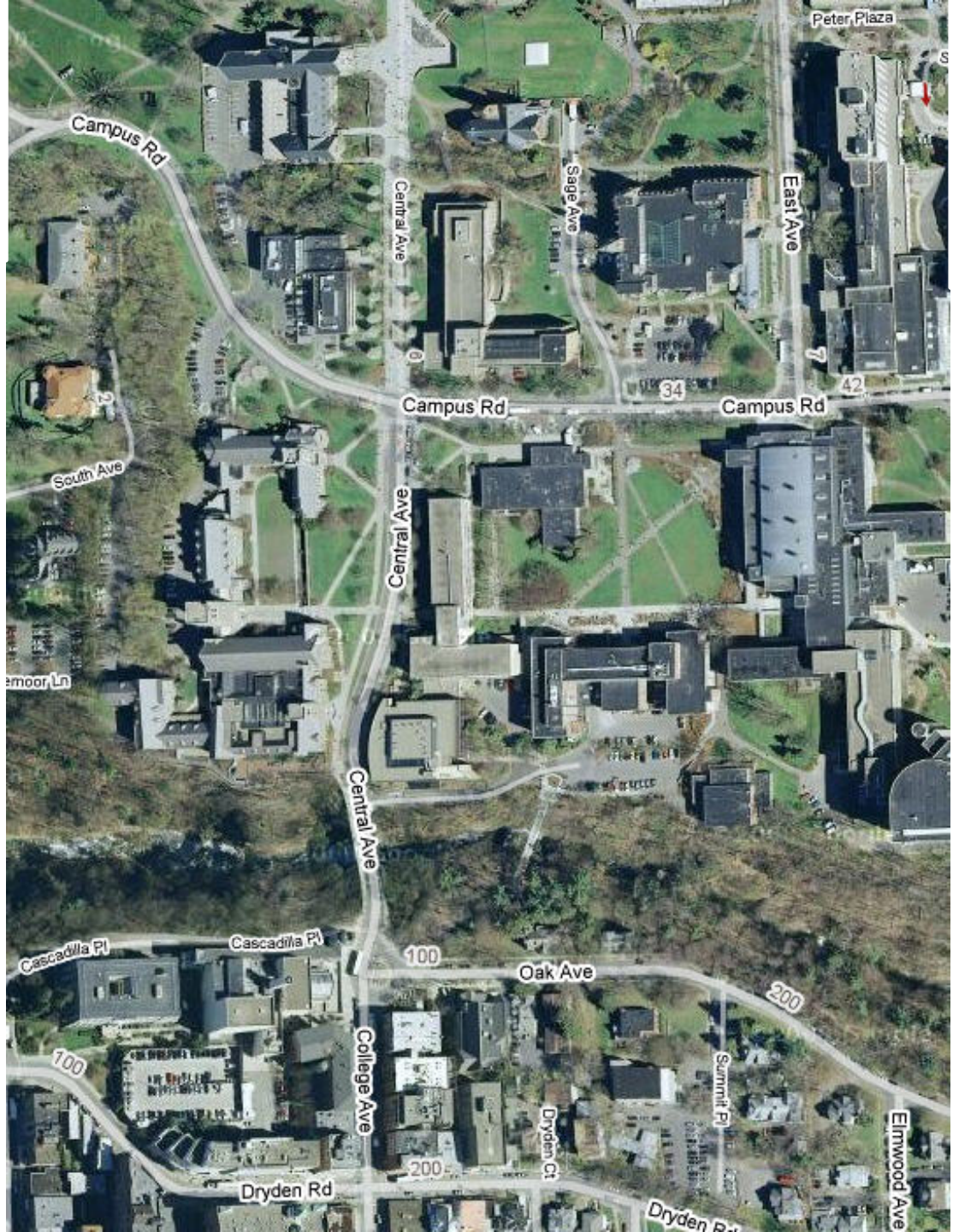
Computing Accounts

- Workshop/training accounts
 - Access to ranger.tacc.utexas.edu and spur.tacc.utexas.edu
 - ssh required
 - Account will be deleted after one week
- TeraGrid Allocations
 - PIs may request Startup, Education, or Research allocations.
 - <https://portal.teragrid.org/gridsphere/gridsphere>
 - Go to Allocations tab
- CAC Services
 - PI may set up projects for
 - compute time
 - consulting time
 - Maintenance
 - Storage
 - <http://www.cac.cornell.edu/services/>



Workshop Logistics

- Refreshments
 - Morning and afternoon breaks provided
 - No food or beverages permitted in training room
 - Lunch on your own
- Restrooms
- Room access
- Video recording





Workshop Web Page

<http://www.cac.cornell.edu/education/training/data09.aspx>

- Agenda
- Lecture slides
- Participants
- Links



Agenda	October 12
9:00	Welcome
9:30	HPC Environment
11:00	Data Transfer, Movement, and Storage
12:00	Lunch
1:00	Data and Database Formats
2:00	Data Analysis with MATLAB
3:00	Data Analysis with Python and R
4:00	MapReduce with Hadoop
5:00	Adjourn

3:45-4:15 + Q/A
G10 Biotech

Dr. Johannes Gehrke
Recent advances in scalability
for computer games,
simulations, and data mining.