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](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 an 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
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00</td>
<td>Welcome</td>
</tr>
<tr>
<td>9:30</td>
<td>HPC Environment</td>
</tr>
<tr>
<td>11:00</td>
<td>Data Transfer, Movement, and Storage</td>
</tr>
<tr>
<td>12:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:00</td>
<td>Data and Database Formats</td>
</tr>
<tr>
<td>2:00</td>
<td>Data Analysis with MATLAB</td>
</tr>
<tr>
<td>3:00</td>
<td>Data Analysis with Python and R</td>
</tr>
<tr>
<td>4:00</td>
<td>MapReduce with Hadoop</td>
</tr>
<tr>
<td>5:00</td>
<td>Adjourn</td>
</tr>
</tbody>
</table>

3:45-4:15 + Q/A
G10 Biotech
Dr. Johannes Gehrke
Recent advances in scalability for computer games, simulations, and data mining.