The Faculty of Arts and Sciences
High Performance Computing Core

Advanced Computational Support for Scientific Research at Yale

Andrew Sherman
HPC Specialist
April 9, 2010
Agenda

• What is HPC?
  – Application Drivers
  – Handling Massive Quantities of Data

• HPC at Yale
  – HPC Facilities
  – HPC Staff
  – HPC Resources & Assistance

• Yale HPC Video
  – Produced by ITS
  – Starring Steve Girvin, Richard Easther, and Dave Frioni
High Performance Computing

• What is it?
  – Use of most powerful/advanced computers to solve problems
  – Historically: Very large individual computers, often “1-off” machines
  – Today: Based on large “clusters” of commodity computers (PCs)

• What is it used for?
  – Research: Mainly physical and life sciences and engineering, but increasingly for data analysis and modeling in the social sciences
  – Applications: All sorts of important “real-world” applications (and not just the obvious scientific and engineering problems). Uses range from weather forecasting, to financial modeling, to data analysis.

• How is it accomplished?
  – Parallel Computing: Leverage a lot of independent processors to solve large problems quickly and (especially important) accurately
  – Difficulties:
    • Developing parallel approaches and well-tuned algorithms
    • Data: Sheer quantity, analysis, validation, …
Top 500 Systems by Application Area (11/2009)

Source: www.top500.org
Top 500 Performance by Application Area (11/2009)

Source: www.top500.org
Everyday Example: Weather Forecasting

Atmosphere modeled by dividing it into 3-dimensional cells.

Temperature, pressure, humidity, etc.

Calculations for each cell repeated many times to model passage of time.
Global Weather Forecasting Example

- Suppose whole global atmosphere divided into cells of size 0.5 mile × 0.5 mile × 0.5 mile to a height of 12 miles (24 cells high) ⇒ about $5 \times 10^9$ cells.

- Suppose each cell update uses ~200 arithmetic operations. ⇒ In one time step, ~$10^{12}$ arithmetic operations necessary.

- To forecast the weather for 7 days using 1-minute intervals, a computer operating at 10 Gflops ($10^{10}$ arithmetic operations/sec) would take $10^6$ seconds. ⇒ It would take over 10 days to simulate 7 days!

- To do this in 1 hour would require a computer 280 times faster ⇒ Computer speed of ~2.8 Tflops ($2.8 \times 10^{12}$ arithmetic ops/sec)
Top 500 Historical Performance Development

Source: www.top500.org

~2.8 TFlops
Data Drives HPC

• Looking ahead, data, not computing, will be the real driver
  – Currently Yale’s HPC clusters have nearly 1.5 PetaBytes of disk!
• Yale examples:
  – Astronomy
    • Chile telescope generates 100 GigaBytes per night now
    • Soon new CCD camera will generate 1 TeraByte (TB) per day
  – West Campus DNA Sequencing Center
    • 10-15 sequencing machines
    • 1 TB per day on average
    • Have just installed a 750 TB high-performance storage facility
  – Physics
    • Quantum particle systems: 100-200 TB from simulations
    • Laser modeling: 50-100 TB
  – Geophysical/Climate Modeling
    • Historical tropical ocean/atmospheric modeling: 150-200 TB
Understanding Data Requirements

HOW MUCH IS 750 TERABYTES?

750 TB is approximately...
163,404 DVD's or
192,000,000 songs or
129,024,000 photos

(That's like storing 24,5 photos every hour for a decade)

The BulldogN cluster, providing 750 TB of disk storage, is only one of several high-performance computing clusters at Yale. In all, the clusters incorporate about 1,456 TB of disk storage shared between computational, scratch, and backup storage. That's nearly twice the capacity depicted here!

HIGH-PERFORMANCE DISK STORAGE IN PERSPECTIVE

The expansion of Yale's High Performance Computing resources will introduce several new high-performance computing clusters and hundreds of Terabytes (TB) of additional disk storage for scientific data. The largest of these clusters, BulldogN, is dedicated to the West Campus Sequencing Center and will provide 750 TB of disk storage. But how can we wrap our minds around these numbers, and How much is 750 Terabytes? This graphic compares this storage capacity to a storage device that we commonly use: a DVD. A DVD holds 4.7 Gigabytes (GB) of data, which translates to approximately 1,175 songs or 790 digital photos.

Source: Adriana Corona/ITS Newsletter/Volume 3, Issue 4: November 2009
HPC at Yale

- At least 60 different projects (including FAS and YSM)
- More than 75 ladder faculty (plus research faculty, postdocs, …)
- Many students
- At least 23 different departments
- Research facilities: Peabody museum, ISPS, etc.
- As of January 2010, researchers were using over 75% of the installed HPC computational capacity

HPC at Yale is increasing rapidly. It appears that many new faculty in the sciences & engineering often see HPC facilities as of at least equal importance to physical research facilities.
Yale HPC Facilities

• Two “Logical” HPC Centers
  – Two physical locations: 300 George St. and West Campus A21
  – FAS Center: Serves main-campus science and engineering depts., including social sciences and non-biomedical life sciences
    • Currently operates 5 primary clusters, plus several smaller ones
    • Funded by a combination of university funds (mainly) and some grant funds (mainly from NSF)
    • Viewed as a “general purpose” facility for all science/engr faculty
  – Keck HPC Center: Serves the needs of biomedical research, both in the Yale Medical School and in several life science departments
    • Currently operates 2 primary clusters
    • Funded mainly by NIH grants
    • Restricted mainly to NIH-funded researchers
    • Close relationship with new biomedical research centers (e.g., West Campus gene sequencing center; Cell imaging center, etc.)
# FAS HPC Center Facilities

## Processing Power

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Date</th>
<th>Nodes (Intel Xeon)</th>
<th>Cores</th>
<th>Memory/Node</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>BulldogH</td>
<td>Fall 2008</td>
<td>70 (2x2x2.6GHz)</td>
<td>280</td>
<td>8 GB</td>
<td>Econ+</td>
</tr>
<tr>
<td>BulldogJ</td>
<td>Fall 2008</td>
<td>128 (2x4x2.83GHz)</td>
<td>1,024</td>
<td>16 or 32 GB</td>
<td>FAS</td>
</tr>
<tr>
<td>BulldogK</td>
<td>Spr 2009</td>
<td>192 (2x4x2.33GHz)</td>
<td>1,536</td>
<td>16 GB</td>
<td>FAS</td>
</tr>
<tr>
<td>BulldogL</td>
<td>Spr 2010</td>
<td>128 (2x4x2.26GHz)</td>
<td>1,024</td>
<td>48 GB</td>
<td>FAS</td>
</tr>
<tr>
<td>BulldogM</td>
<td>Spr 2010</td>
<td>128 (2x4x2.26GHz)</td>
<td>1,024</td>
<td>48 GB</td>
<td>Astro</td>
</tr>
</tbody>
</table>

## Storage Space

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Primary Networks</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>BulldogH</td>
<td>Gigabit Ethernet</td>
<td>40 GB scratch/node; 100 GB stor/group</td>
</tr>
<tr>
<td>BulldogJ</td>
<td>DDR InfiniBand</td>
<td>98 TeraByte parallel scratch storage</td>
</tr>
<tr>
<td>BulldogK</td>
<td>Gigabit Ethernet</td>
<td>Approx. 20 GigaBytes perm. storage/user</td>
</tr>
<tr>
<td>BulldogL</td>
<td>QDR InfiniBand</td>
<td>100 TeraByte parallel storage</td>
</tr>
<tr>
<td>BulldogM</td>
<td>Gigabit Ethernet</td>
<td>100 TeraByte parallel storage</td>
</tr>
</tbody>
</table>
West Campus Hot Aisle Containment Units ("Pods")
HPC Staff

• System Administration Staff
  – 4 people based at West Campus
  – Responsible for hardware and operating-system-level software, and for related user support issues
  – Shared with Keck HPC Center

• User Support Staff (FAS HPC Center)
  – Andy Sherman, based in Computer Science
    • Email: andrew.sherman@yale.edu
    • Phone: 436-9171 (office) or (203) 376-8144 (cell)
    • Office: AKW 015 (51 Prospect St.)
  – Responsible for application-level software and user support

• User Support Staff (Keck HPC Center)
  – Nick Carriero, Rob Bjornson, based in Computer Science
  – Responsible for biomedical HPC user support
  – Responsible for HPC application/data support for West Campus gene sequencing center
Online HPC Resources at Yale

• HPC portal
  – http://research.yale.edu/hpc/

• HPC wiki
  – http://hpc.research.yale.edu/wiki/index.php/Yale_HPC_Wiki

• Getting started page
  – http://hpc.research.yale.edu/wiki/index.php/Getting_started

• Cluster status
  – http://hpc-status.wss.yale.edu/ganglia/index.php

• For account information, or for help using the clusters
  – Email to hpc@yale.edu
Yale HPC Video

Video may be accessed at:

http://cmi2.yale.edu/hpc/video.html

Credits:

• Produced by Yale ITS
• Starring Steve Girvin, Richard Easther, Dave Frioni