CLUSTERS AND YOU
AN INTRODUCTION TO CLUSTER COMPUTING

Engineering IT BrownBag Series
29 October, 2015

Gianni Pezzarossi
Linux Systems Administrator

Mark Smylie Hart
Research Technology Facilitator
WHAT WE’LL COVER TODAY

• Intro to clusters
• Cluster types
• Illinois Campus Cluster Program (ICCP)
  – How it works
  – How you can get started
• Support Options
• Our vision of the future
WHAT IS A CLUSTER?

- HPC
- HTC
- VMs
- The Cloud
- GPUs / Phis
- Data Intensive
- Big Servers
- Alternate Platforms

Open Science Grid

Virtual Machines

Cloud computing
CLUSTER CARE AND FEEDING

• Infrastructure
  – Space
  – Power
  – Cooling

• Network Connectivity
  – ‘Infiniband’

• Resource Management Engine(s)

DON’T FORGET THE PEOPLE
WHO USES CLUSTERS?

- Nuclear Physics: 6.60%
- Combustion: 7.98%
- Industry: 1.58%
- Astrophysics: 17.55%
- Biology: 6.6%
- Chemistry: 4.35%
- Climate: 11.77%
- Computer Science: 1.32%
- Materials: 12.93%
- Fusion Energy: 13.29%
- Geosciences: 1.32%
- High Energy Physics: 13.72%
- Atomic Physics: 0.99%
(SOME OF) THE MANY FLAVORS OF CLUSTER

• High Availability / Load Balancing
  – Robustness through redundancy

• High Performance Computing (HPC)
  – Tightly coupled

• High Throughput Computing (HTC)
  – Loosely coupled
CLUSTER TYPES - HA

DATABASE

EMAIL SERVERS

[Diagram showing the relationship between database and email servers with a red cross indicating a failure or unavailability.]
CLUSTER TYPES - HTC

- DATA RESOURCE
- COMPUTE RESOURCE
CLUSTER TYPES - HPC

DATA NODES  COMPUTE NODES
QUESTIONS SO FAR?
CLUSTER TYPES - HPC

DATA NODES

COMPUTE NODES
BLUE WATERS

• Not JUST a Linux Cluster
  – Proprietary hardware and software
  – Tuned OS

• Scale!
  – Not all research problems will see benefits

• How is it different than ICCP
  – No MATLAB
  – Must submit an application to use (XSEDE.org)
ILLINOIS CAMPUS CLUSTER PROGRAM

• ICCP is using High Performance Computing
  – Infiniband interconnect (optional)
  – 2x 10-core Processors
    • New purchases are 2x12 Haswell
  – 64-256 GB RAM
  – 2x NVIDIA TESLA K40 GPU
• Also has a small Hadoop instance
NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

- NCSA runs Blue Waters (BW) and ICCP
- BW is the most capable supercomputer on ANY campus
- ICCP is a pooled resource
- Dedicated support teams for each
  - More than 25 years of experience with HPC support
  - 24/7 helpdesk to field questions
- NCSA has code support available
  - Help with optimizing/parallelizing code
VISUALISING CLUSTER USES

DATA INTENSIVE

Home Grown/VM

Campus Cluster

COMPUTE INTENSIVE

Blue Waters
HOW IT WORKS

Submitted
Job

Job Scheduler

Resource
Manager

Happy data

Job being
crunched
HOW IS THIS DIFFERENT?

• One node = 4 to 5 desktops
• Purpose built for speed and robustness
• It won’t run WeatherBug
HOW DO YOU GET STARTED?

• Purchase Nodes
  • On your own (as individual or unit) or…
  • Partnered with others to save costs
  • https://campuscluster.illinois.edu/invest/pricing.html

• Pay By The Hour (*coming soon*)
  • Will have limited options initially
  • Cost will be $15.84 per core-month

• For Free Through CSE
  • Competition for time
  • http://cse.illinois.edu/research/computing-resources
Node Investment

- Node lifetime officially defined as 4-5 years…
  - But work is underway to extend that

- Investors gain access to secondary queue (idle nodes)

- Investors get a seat on the Investor Forum which represents the interests of all investors in the Campus Cluster
STORAGE OPTIONS

• Investment nodes have scratch space only
  – 30TB (raw) storage blocks available

Active Data Service

• $100 / TB / year
• Regular snapshots (goof-proof, not bomb-proof)
• ADS will see Campus Cluster and vice-versa
SUPPORT OPTIONS

We’ve asked around, and we’re listening

• Improving training
• Improving access
• Improving options
SUPPORT OPTIONS

• Videos to help lower the barrier to entry
• Training offerings via CSE
• Lynda.com
• HPC University
FUTURE VISION

Improving ourselves

• Pet cluster projects in Engineering IT
  – Exploring multiple, flexible configuration(s)

• Cross-training opportunities with NCSA

• Will help us become better partners for you
<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overview: Engineering IT &amp; Campus</td>
<td></td>
<td>Research IT Support Contracts and You!</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Using Skype for Business</td>
<td></td>
<td>Purchasing: Hardware and Software</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Using Outlook to Manage Email and Calendaring</td>
<td></td>
<td>File Service: Options, Backups and Best Practices*</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>IT Security: Protecting Your Systems on Campus</td>
<td></td>
<td>Introduction to Cluster Computing</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Wireless Networking Clinic</td>
<td></td>
<td>Research Group Access Control: Using the Portal Groups Tool*</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Engineering Managed Linux Environment</td>
<td></td>
<td>No Brown Bag</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Data Center Shared Services</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*will be in Deere Pavilion, MEL