Fall 2017 ITS-CI Faculty Advisory Committee Meeting

Gwen Jacobs, Michelle Choe, Sean Cleveland, Jennifer Geis, Ouida Meier, Ron Merrill, David Schanzenbach

September 25, 2017
Outline

• Cray Cluster Statistics for Fiscal Year 2017
• Looking forward for the Cray Cluster
• Cray Cluster re-configuration and management system overhaul
  – Repeal and Replace
• Data Science Initiative
• Return on Investment
• Lustre Scratch file system and purge policy
• Workshops
• Q/A
Fiscal Year 2017 Statistics

Top 3 Organizational Units by CPU Usage
- Manoa -- School of Ocean and Earth Science and Technology (35.6 Mill Hrs)
- Manoa -- College of Natural Science (2.3 Mill Hrs)
- Manoa -- College of Tropical Agriculture and Human Resources (0.5 Mill Hrs)

HPC Users
- 452 Accounts
- 237 Active
- 173 New accounts

OUs on HPC
- 76 OUs
- 3 Campuses

HPC Usage
- 39.43 Million CPU hours
- 726,783 Compute Jobs
- 75.95% Utilization

Engagements
(software: install, consulting, dev)
- 998 Help tickets
- > 208 Meetings

Training
- 32 Trainings
- 181 Attendees (HPC 101, ‘Ike Wai, Agave)
Looking Forward

- Cray warranty expires **March 15, 2019**
  - ~$40k per year if available
  - Only Co-term for whole cluster
- Software licensing
  - Renew Intel compilers November 9, 2017
  - TAH Matlab license for UHM
    - See https://www.hawaii.edu/sitelic/matlab/matlab.html
- 24-ports remain on the core IB switch
  - Next potential purchase deadline is February 7, 2018
Cluster Reorganization

• Move towards Virtualized frontend
  – Dedicated login and data transfer VMs
• Tighter security
  – Multi-factor authentication
• Upgrade of operating system from CentOS 6.5 to CentOS 7.x
• Overhaul the cluster management system
Overhaul of Cluster Management system

• Cray Management System Problems
  – No upgrades without paying Cray
  – Deprecated - not available for new hardware

• Replace with open source alternative -- xCAT
  – OS upgrade possible
  – Non vendor locked
  – Process could take over a year to complete
    • Why? It’s like rebuilding a flying plane...
Data Science Initiative

• Possibly bootstrap a cluster #2
• Focus on high throughput computing and data science applications
• Will continue to utilize a condo model
• Will utilize the existing Lustre scratch file system and job scheduler
• Will leverage planned VM infrastructure and management system
• Rent existing standard community nodes
  – Possibly up to 100 Nodes
• Node rental price would be equivalent to $1,320 per year (⅕ of full node price).
• Possibly provide an option to rent part of the scratch file system with node rental
  – Possibly $200 TB/yr
Purge Policy

• Current process and policy is ineffective
  – Some users don’t heed our emails and use a vast majority of the file system (50+ TB)
• Current process is semi-automated
  – Email large users
  – Run purge notification scripts
    • collect information and email users
  – Scripted move and deletion of files, but manual verification of validity
## Top 20 users of Lustre as of 09/25

<table>
<thead>
<tr>
<th>Rank</th>
<th>Usage</th>
<th>File Count</th>
<th>Average File Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>73.38 TB</td>
<td>7,201,075</td>
<td>10.68 MB</td>
</tr>
<tr>
<td>2</td>
<td>45.55 TB</td>
<td>3,982,362</td>
<td>11.99 MB</td>
</tr>
<tr>
<td>3</td>
<td>28.21 TB</td>
<td>585,386</td>
<td>50.53 MB</td>
</tr>
<tr>
<td>4</td>
<td>22.25 TB</td>
<td>1,777,091</td>
<td>13.13 MB</td>
</tr>
<tr>
<td>5</td>
<td>21.36 TB</td>
<td>9,227,397</td>
<td>2.43 MB</td>
</tr>
<tr>
<td>6</td>
<td>17.92 TB</td>
<td>19,586</td>
<td>959.44 MB</td>
</tr>
<tr>
<td>7</td>
<td>15.06 TB</td>
<td>285,301</td>
<td>55.35 MB</td>
</tr>
<tr>
<td>8</td>
<td>10.05 TB</td>
<td>162,095</td>
<td>64.99 MB</td>
</tr>
<tr>
<td>9</td>
<td>9.70 TB</td>
<td>376,583</td>
<td>27.01 MB</td>
</tr>
<tr>
<td>10</td>
<td>8.76 TB</td>
<td>405,885</td>
<td>22.64 MB</td>
</tr>
<tr>
<td>11</td>
<td>8.45 TB</td>
<td>604,301</td>
<td>14.66 MB</td>
</tr>
<tr>
<td>12</td>
<td>7.89 TB</td>
<td>81,783</td>
<td>101.17 MB</td>
</tr>
<tr>
<td>13</td>
<td>7.73 TB</td>
<td>153,147</td>
<td>52.93 MB</td>
</tr>
<tr>
<td>14</td>
<td>7.23 TB</td>
<td>71,492</td>
<td>106.07 MB</td>
</tr>
<tr>
<td>15</td>
<td>6.39 TB</td>
<td>255,344</td>
<td>26.23 MB</td>
</tr>
<tr>
<td>16</td>
<td>6.10 TB</td>
<td>506,210</td>
<td>12.64 MB</td>
</tr>
<tr>
<td>17</td>
<td>4.64 TB</td>
<td>119,345</td>
<td>40.80 MB</td>
</tr>
<tr>
<td>18</td>
<td>4.23 TB</td>
<td>1,039,958</td>
<td>4.27 MB</td>
</tr>
<tr>
<td>19</td>
<td>3.42 TB</td>
<td>135,191</td>
<td>26.55 MB</td>
</tr>
<tr>
<td>20</td>
<td>3.35 TB</td>
<td>138,881</td>
<td>25.30 MB</td>
</tr>
</tbody>
</table>
Purposed Purge Policy Change

• Complete automation of the process
  – Runs automatically based on global and/or per user thresholds (usage, file count, age, etc)
• Do not provide grace periods or email notifications to users
• Files deleted oldest to newest
• Requires re-training users
Fiscal Year 2018 Training

- Software carpentry (date(s): TBD)
- Agave (date(s): TBD)
- Google Cloud Platform (Oct 6, 2017)
- HPC 101 (1 or 2 times a month)

Question: What workshops would people be interested in attending?
Questions?

Issues and concerns from the ITS-CI Faculty Advisory Committee?