Caltech HPC Cluster:
Initial planning thoughts to facilitate discussion

Kaushik Bhattacharya

24 February 2017
Background

• Need an institute-wide high performance computing
  – Essential tool across disciplines and divisions
  – Individual closet-sized clusters are inefficient and unsustainable
  – Enables critical research that is not otherwise possible

• Cloud is complementary, not yet a substitute

• National facilities are a complement, not a substitute

• Caltech has spent about $1.6M/year over the last 10 years on computing clusters through startups, gifts and grants
Proposed model

• Staggered model
  – Use $3-4M initial investment to seed and institute wide heterogeneous cluster
  – Add nodes in years 2 and 3 if sufficient new funds (startups, gifts, grants) become available; pool funds to subsequent year if funds are insufficient
  – Freeze the cluster after year three, run for 5-7 years
  – Start new cluster/evaluate transition to the cloud every 4-6 years

• Housed in Powell Booth/South Mudd

• Management
  – Managed by the Vice Provost
  – Run by IMSS
  – Faculty advisory committee: Dan Meiron (chair), Mitch Guttman, Jonathan Katz, Tom Miller, Mark Simons, Maria Spiropulu

• Faculty can buy nodes (fair share) through startups, grants, gifts
  – Fair share usage
  – Lower user fees
Faculty survey: December 2016

- 147 unique responses
- Snapshot (CPU use)

<table>
<thead>
<tr>
<th>Typical Core</th>
<th>BBE</th>
<th>CCE</th>
<th>EAS</th>
<th>GPS</th>
<th>HSS</th>
<th>PMA</th>
<th>U</th>
<th>Total</th>
<th>Heavy GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>500+</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td></td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>100-500</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>50-100</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td></td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>15-50</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>1-14</td>
<td>6</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>44</td>
<td></td>
</tr>
</tbody>
</table>

- Suggests we need 12,000 – 15,000 CPUs (consistent with installed capacity)
- GPU use is not that high, but growing
- Some users need small core count, but lots of memory
- Not all heavy users need high speed interconnect
$4M strawman based on Dell preliminary quote

- **Strawman**
  - 8000 cores, 1 petaflop
  - 250 CPU nodes,
  - 40 GPU nodes,
  - 2:1 infiniband connections,
  - 500 CPUs with extra memory

- **Caveat:** Skylake, NVLINK pricing is not finalized

- Likely to meet immediate campus demand
  - perhaps needs growth of GPU as FRAM (GPS cluster) shuts down
  - Assume that some of the new clusters continue to operate for 3-5 years
  - Some users are better served by the cloud

- HPC committee members have the spread-sheet and have been asked to configure the machine as they see fit
User fees

- Very early strawman

<table>
<thead>
<tr>
<th>Usage (K hours)</th>
<th>Owner $/core hour</th>
<th>Owner Annual ($)</th>
<th>Non owner $/core hour</th>
<th>Non owner Annual ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>50-500</td>
<td>0.75</td>
<td>3375</td>
<td>1.5</td>
<td>6750</td>
</tr>
<tr>
<td>500-3000</td>
<td>0.375</td>
<td>12750</td>
<td>1.125</td>
<td>34875</td>
</tr>
<tr>
<td>3000-10000</td>
<td>0.1875</td>
<td>25875</td>
<td>0.9375</td>
<td>100500</td>
</tr>
</tbody>
</table>

- Comparison

<table>
<thead>
<tr>
<th></th>
<th>Pod</th>
<th>AWS</th>
<th>Spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU-hour</td>
<td>7-9</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>GPU-hour</td>
<td>197</td>
<td>90</td>
<td>20</td>
</tr>
</tbody>
</table>

- Estimated fees based on strawman cluster: $1.13M
Timeline and communication

• March 2017
  – Confirmation of the availability of core funding $2M
  – Exploration of additional funds – faculty, divisions, startups …
  – Discussion with vendors

• April – May 2017
  – Negotiation with vendors
  – Order

• September 2017
  – Delivery

• Will post regular updates on imss.caltech.edu/hpc
• Please contact Bhattacharya, Meiron, Guttman, Katz, Miller, Simons, Spiropulu if you have questions/comments/suggestions