2018 COMMUNITY CLUSTER PARTNER ANNUAL MEETING

PRESTON SMITH
DIRECTOR OF RESEARCH COMPUTING SERVICES
PSMITH@PURDUE.EDU

ALEXANDER YOUNTS
PRINCIPAL RESEARCH COMPUTING ENGINEER
AY@PURDUE.EDU
COMMUNITY CLUSTERS
Community Cluster Program

The Rules

• We build a cluster each year
  • Cluster is a service, buying in provides 5 years of access.
  • Since 2008, faculty invest approx. $2.5M of compute each year

• You get out at least what you put in
  – Buy 1 node or 100, you get a queue that guarantees access up to that many CPUs
Community Cluster Program

The Rules

• But wait, there’s more!!
  – What if your neighbor isn’t using his queue?
    • You can use it, but your job has to run in 4-hour chunks if he wants to run.

• You don’t have to do the work
  – Your grad student gets to do research rather than run your cluster.
    • Nor do you have to provide space in your lab for computers.
  – ITaP provides data center space, systems administration, application support.
  – Just submit jobs!
# 10 HPC SYSTEMS

<table>
<thead>
<tr>
<th>System</th>
<th>Cores</th>
<th>Installed</th>
<th>Departments</th>
<th>Faculty Investors</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEELE</td>
<td>7,216</td>
<td>May 2008</td>
<td></td>
<td></td>
<td>Retired Nov. 2013</td>
</tr>
<tr>
<td>COATES</td>
<td>8,032</td>
<td>May 2008</td>
<td>24</td>
<td>61</td>
<td>Retired Sep. 2014</td>
</tr>
<tr>
<td>ROSSMANN</td>
<td>11,088</td>
<td>Sept. 2010</td>
<td>17</td>
<td>37</td>
<td>Retired Sep. 2015</td>
</tr>
<tr>
<td>CARTER</td>
<td>10,368</td>
<td>April 2012 – 2017</td>
<td>26</td>
<td>60</td>
<td>#54 on June 2012 Top 500</td>
</tr>
<tr>
<td>CONTE</td>
<td>9,280 Xeon</td>
<td>August 2013 – 2018</td>
<td>26</td>
<td>62</td>
<td>#28 on June 2013 Top 500</td>
</tr>
<tr>
<td>DATA DEPOT</td>
<td>2.5 PB</td>
<td>Nov. 2014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RICE</td>
<td>13,200</td>
<td>May 2015</td>
<td>33</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>HALSTEAD</td>
<td>10,160</td>
<td>December 2016</td>
<td>39</td>
<td>62</td>
<td>#302 on Nov 2017 Top 500</td>
</tr>
<tr>
<td>BROWN</td>
<td>13,200</td>
<td>October 2017</td>
<td>36</td>
<td>70</td>
<td>$5039 for 5 years of service</td>
</tr>
</tbody>
</table>

**Notes:**
- **Rice:** 13,200 cores, Installed May 2015
- **Hansen:** 9,120 cores, Installed Sept. 2011
- **Carter:** 10,368 cores, Installed April 2012 – Retired 2017
- **Conte:** 9,280 Xeon cores (69,900 Xeon Phi), Installed August 2013 – Retired 2018
- **Data Depot:** 2.5 PB of disk storage, Installed Nov. 2014
- **Rice:** 13,200 cores, Installed May 2015
- **Halstead:** 10,160 cores, Installed December 2016
- **Brown:** 13,200 cores, Installed October 2017

**Important Dates:**
Purdue Community Clusters
Top500 Rankings
Cluster Program Partners

<table>
<thead>
<tr>
<th>Department</th>
<th>Cores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeronautics and Astronautics</td>
<td>5740</td>
</tr>
<tr>
<td>Mechanical Engineering</td>
<td>5556</td>
</tr>
<tr>
<td>CMS Tier2</td>
<td>5440</td>
</tr>
<tr>
<td>Electrical and Computer Engineering</td>
<td>4344</td>
</tr>
<tr>
<td>Earth, Atmospheric, and Planetary Sciences</td>
<td>2540</td>
</tr>
<tr>
<td>Materials Engineering</td>
<td>2064</td>
</tr>
<tr>
<td>Nuclear Engineering</td>
<td>1564</td>
</tr>
<tr>
<td>Other College of Engineering</td>
<td>980</td>
</tr>
<tr>
<td>Chemistry</td>
<td>824</td>
</tr>
<tr>
<td>Physics and Astronomy</td>
<td>820</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td>640</td>
</tr>
<tr>
<td>Other Executive Vice President for Research and Partnerships</td>
<td>600</td>
</tr>
<tr>
<td>Statistics</td>
<td>512</td>
</tr>
<tr>
<td>Chemical Engineering</td>
<td>424</td>
</tr>
<tr>
<td>Agricultural and Biological Engineering (Biological Engineering)</td>
<td>368</td>
</tr>
<tr>
<td>Biological Sciences</td>
<td>356</td>
</tr>
<tr>
<td>Industrial Engineering</td>
<td>296</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>276</td>
</tr>
<tr>
<td>Computer and Information Technology</td>
<td>248</td>
</tr>
<tr>
<td>Medicinal Chemistry and Molecular Pharmacology</td>
<td>248</td>
</tr>
<tr>
<td>Mathematics</td>
<td>232</td>
</tr>
<tr>
<td>Bioinformatics Core</td>
<td>200</td>
</tr>
<tr>
<td>Agronomy</td>
<td>180</td>
</tr>
<tr>
<td>ITaP</td>
<td>176</td>
</tr>
<tr>
<td>Computer Science</td>
<td>156</td>
</tr>
<tr>
<td>Horticulture and Landscape Architecture</td>
<td>156</td>
</tr>
<tr>
<td>Forestry and Natural Resources</td>
<td>96</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>40</td>
</tr>
<tr>
<td>Botany and Plant Pathology</td>
<td>40</td>
</tr>
<tr>
<td>Industrial and Physical Pharmacy</td>
<td>40</td>
</tr>
<tr>
<td>Brian Lamb School of Communication</td>
<td>32</td>
</tr>
<tr>
<td>Agricultural Economics</td>
<td>20</td>
</tr>
<tr>
<td>Animal Sciences</td>
<td>20</td>
</tr>
<tr>
<td>Food Science</td>
<td>20</td>
</tr>
<tr>
<td>Health Sciences</td>
<td>20</td>
</tr>
<tr>
<td>Other College of Pharmacy</td>
<td>20</td>
</tr>
<tr>
<td>Agricultural and Biological Engineering (Agricultural Systems Mgmt)</td>
<td>16</td>
</tr>
</tbody>
</table>

302M hours delivered in 2017
177 active (over 200 all-time) investors from **50 departments, from every College, and 3 Purdue campuses**

Today, the program is part of many departments’ faculty recruiting process.

_A selling point to attract people to Purdue!_
Research Awards for Research Computing Users

Expenditures Using Research Computing

Total Expenditures


Research Awards to HPC Partners
Brown – on the Floor Today

Specifications

*If we were to build a CPU cluster today, it would look exactly like Brown!*

**Base node: Dell R640**

- 24-core node, 2.6 GHz Intel Xeon Gold “Sky Lake” processors (Xeon Gold 6126)
  - 32 Flops per cycle!
- 96 GB DDR4 memory
  - 384 GB, 768 GB & 1536 GB options
- EDR Infiniband interconnect
  - 100 Gbps, 3:1 fat tree – very similar in speed to Halstead
  - Converged fabric – IP traffic uses Infiniband rail

There are still Brown nodes available at $5,039 per node (through 2022)

- now $4,480 per node

Next spring: Look for next CPU cluster based on “Cascade Lake” or AMD “Rome”
Interactive Data Science – Data Workbench

$300 per lab/year
37 labs using already!
ACCELERATED COMPUTING

FPGAS, MICs and GPUS
Past Forays into Accelerators

- FPGA (Brutus)
- Small number of K20-era GPUs (Carter)
- Xeon Phi (Conte)
It all comes full circle!

- Like with clusters, networks of GPU workstations are everywhere.
- For individual PIs, consumer GPUs are cost-effective for quite a bit of capability.
- However – Nvidia GeForce license doesn’t allow for data center use!

This is the 2018 version of “building a cluster in your office” and having your grad student run it!
Faculty Survey

- May 2018
- Over 50 invited to participate
- 20 responses
Survey Data

Top Challenges facing the research group

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not enough resources (GPUs) for the group</td>
<td>68</td>
</tr>
<tr>
<td>Power/Cooling issues</td>
<td>42</td>
</tr>
<tr>
<td>Software or system installation and maintenance</td>
<td>42</td>
</tr>
<tr>
<td>System Performance</td>
<td>35</td>
</tr>
<tr>
<td>Hardware maintenance</td>
<td>31</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>30</td>
</tr>
<tr>
<td>Access to Storage Capacity</td>
<td>20</td>
</tr>
<tr>
<td>Storage Performance</td>
<td>0</td>
</tr>
</tbody>
</table>

Top Priorities when Selecting GPU servers or workstations

<table>
<thead>
<tr>
<th>Priority</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per-GPU performance</td>
<td>32</td>
</tr>
<tr>
<td>Total system performance</td>
<td>31</td>
</tr>
<tr>
<td>Lowest total price per system</td>
<td>28</td>
</tr>
<tr>
<td>Memory per GPU</td>
<td>26</td>
</tr>
<tr>
<td>Single Precision FLOPs</td>
<td>17</td>
</tr>
<tr>
<td>Double Precision FLOPs</td>
<td>13</td>
</tr>
</tbody>
</table>

GPU numbers, by GPU type

<table>
<thead>
<tr>
<th>GPU Type</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tesla P100</td>
<td>40</td>
</tr>
<tr>
<td>Tesla V100</td>
<td>12</td>
</tr>
<tr>
<td>Tesla K80</td>
<td>4</td>
</tr>
<tr>
<td>Tesla K40</td>
<td>8</td>
</tr>
<tr>
<td>Tesla K40</td>
<td>0</td>
</tr>
<tr>
<td>GTX 10XX</td>
<td>89</td>
</tr>
<tr>
<td>Titan</td>
<td>124</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
</tr>
</tbody>
</table>

| Total GPUs per server                         | 293       |

GPUs per server
Mean of 4.4, stdev of 2.2
Survey Data

Server Counts
Lab-run: avg of 8.7 per PI, stdev of 7
IT-run: avg of 18 servers, serving 40 PIs

Memory Config per workstation/server

<table>
<thead>
<tr>
<th>Answer</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 GB or less</td>
<td>9.09%</td>
</tr>
<tr>
<td>33-96 GB</td>
<td>27.27%</td>
</tr>
<tr>
<td>97-192 GB</td>
<td>18.18%</td>
</tr>
<tr>
<td>193-256 GB</td>
<td>0.00%</td>
</tr>
<tr>
<td>257-512 GB</td>
<td>9.09%</td>
</tr>
<tr>
<td>&gt; 512 GB</td>
<td>36.36%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>
Key Takeaway – Campus Needs More GPUs

A Faculty Quote

“What is the point of having extensive studies of our strategy about Data Science when we don’t even have [enough] computers???

The Biggest Challenge, from Survey:

Having sufficient GPUs for the lab

Also, facility problems, and “taking care of all the computers” problems
2018 Community Cluster

Gilbreth

• GPU-based system ideal for machine learning, AI, big data science – as well as FEA, Chemistry, MD
  – .74 PF of computing!

• Annual subscription for the lab’s access to the entire cluster.

Hardware currently being received!

Prof. Lillian Moller Gilbreth
Open Bid Process – Requesting:

- Mellanox SwitchX-2 EDR for GPUDirect RDMA
- Intel Xeon Skylake Silver Processors
- 192GB of DDR4 Memory
- SSD boot drive, 250G or better
- Two Nvidia P100 16GB GPU per node

- Responses ranged from $14,100-14,700
Compute Node

- Base node: Dell R740
- 16-core node, 2.1-3.0 GHz Intel Xeon Sky Lake processors (Xeon Silver 4110)
  - 85w vs 125w
- 2x 16GB P100 GPUs per node
- 192 GB DDR4 memory
- Mellanox SwitchX-2 and ConnectX-4 EDR Infiniband interconnect
  - 100 Gbps
  - Converged fabric – IP traffic uses Infiniband rail
Existing GPU Hardware

- HalsteadGPU and BrownGPU nodes and owners will be merged into Gilbreth
  - (We’ll reach out and work with you directly to migrate)

- PLUS 3 new dedicated model-training nodes, funded in partnership with OVPR
  - 4x 32GB V100 GPUs per node, with NVLink
  - 8TB local NVMe flash

- Caffe, TensorFlow, and CNTK are up to 3x faster with Tesla V100 compared to P100
- Up to 125 TFLOPS of TensorFlow operations per V100 GPU
GPU Resource Breakdown

- 2 GPU-enabled frontend nodes (1 P100 GPU each)
- 4 interactive queue nodes (8 P100 GPUs)
- 40 batch queue nodes (80 P100 GPUs)
- 3 Training Nodes (12 32GB V100 GPUs)
  - NVLink, NVMe flash

88 P100 GPUs
12 V100 GPUs
.74 PFLOPS
Speeds and Feeds and Specs

Storage

Brown (550 nodes)
- 3 PB Lustre
- 40 GB/sec bandwidth
- 400k IOPS

Gilbreth (~50 nodes)
- 2 PB Lustre
- 30 GB/sec bandwidth
- 200k IOPS
- Expandable to 10PB and 72GB/s

Gilbreth Flash Burst Buffer
- ~100 TB NVMe
- 60 GB/sec bandwidth
- 1M IOPS
Application Performance

LAMMPS, GROMACS, Tensorflow, vs CPU

Remaining top code to benchmark:
- VASP
## Gilbreth - Pricing

### Costs

- Three ways to buy:

<table>
<thead>
<tr>
<th></th>
<th>Buy Node (5y)</th>
<th>Buy Node (annual)</th>
<th>Shared Queue Subscription</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>$14,144</td>
<td>$2,829/year</td>
<td>$1,599/year</td>
</tr>
</tbody>
</table>

Shared queue model provides cheapest annual cost to get access to high-end GPUs! A proven model with Halstead and Brown.

Subscription is not a “node” purchase – access to a large pool of nodes!
Community Cluster Guiding Principles

- You don’t have to do the work
- You get to benefit from the pool of unused resources (get back more than what you put in)
- ITaP provides high-end networking, storage sufficient to drive these powerful systems
- Your lab doesn’t have to power, cool, and listen to your GPU systems.
- ITaP manages the compute environment and software stacks so your students can get PhDs and write papers.
Gilbreth Capabilities

- The big benefit is scale, for throughput!
  - You and your students will have access to a bigger pool of resources than you can do by yourself.
  - If you’ve used Brown-GPU or Halstead-GPU, please give us your feedback on scheduling set-up!

- Gilbreth’s Pascal/Volta cards have more RAM than consumer cards.
- 200+ TB of Lustre scratch
- Additional shared flash
- Volta + nvlink training nodes
  - Local flash on training nodes
- GPU-enabled Jupyter notebooks
- CUDA 10
- Nodes dedicated to interactive use

Not a “node” purchase – it’s access to a large pool of nodes!
Managed Software Stacks

**ML Toolkit**

- ml-toolkit-gpu/all/all
- ml-toolkit-gpu/caffe/1.0.0
- ml-toolkit-gpu/cntk/2.3
- ml-toolkit-gpu/gym/0.10.5
- ml-toolkit-gpu/keras/2.1.5
- ml-toolkit-gpu/opencv/3.3.1
- ml-toolkit-gpu/pytorch/0.4.0
- ml-toolkit-gpu/tensorflow/1.4.0
- ml-toolkit-gpu/tflearn/0.3.2
- ml-toolkit-gpu/theano/1.0.2

**Non-CentOS Userland**

- Bring your own Linux userland with Singularity containers!

  (Ubuntu, Slackware, Fedora, other)
RESEARCH DATA
On-Campus Research Storage Solutions

Left to fend for themselves, researchers will find a way
Growth in Research Data Storage
The Research Data Depot

Impact

• Over 550 research labs are Depot partners!
  – 60% are not HPC users!
  – Thousands of individual users

• Over 2 PB sold

• A research group purchasing space has purchased, on average, nearly 10 TB.

Refresh/expansion underway to extend Depot for 5 more years!
Further enhancement (performance, scalability) to non-HPC users.
The Research Data Depot

$70
At $75 per TB/year

• Storage oriented around the PI’s research lab, with
  – Snapshots
  – Multi-site, active copies of data
  – Easy ways to do common research data management patterns
  – A scalable, expandable storage resource optimized for HPC

• Access to Globus data transfer service, and endpoint sharing

• **Our goal: enabling the frictionless use and movement of data**

Expansion: now with a tertiary copy of data for near real-time disaster protection, for less $$$ per TB!
Our key value – computational experts

- New faculty orientations
- One-on-one consultations
- UNIX, MPI, HPC, Python, R training offerings
- Cyberinfrastructure seminars

ADVANCED DOMAIN EXPERTISE

Chemistry
Physics
Astrophysics
Earth and Atmospheric Sciences
Computer Science
Chemical Engineering
Electrical and Computer Engineering
Cell and Molecular Biology
Entomology
Dedicated Cluster for HPC and Data Science

2035 students using Scholar to learn HPC and data science this semester!

Interactive GPUs now available for instructional use!
UPCOMING ENHANCEMENTS
Science DMZ to Data-Intensive Instruments

**NSF CC* Grant**

- Extend science DMZ to big data instrument “hot-spots”
- Create a cost-effective, sustainable architecture for research data networks.
- Develop research computing professionals.

The network is one part of all the tools we provide for supporting data-intensive instrumentation!
Cybersecurity Framework for Regulated Data

REED+

A managed research ecosystem with sufficient storage, high speed computing capability and security to efficiently and cost effectively handle Purdue’s controlled research data and processing needs in a manner compliant with the highest level of cybersecurity applicable to Unclassified data possessed by Purdue University and Purdue University researchers.

2 year, $600k NSF Award #1840043 to build framework
Changes for Off-Campus Access to Clusters

VPN Required for Off-Campus

- SSH
- Thinlinc remote desktops
- CIFS access to cluster scratch
THANK YOU

www.rcac.purdue.edu
psmouth@purdue.edu