Buy-In

Overview

The University of Arizona’s High Performance Computing (HPC) clusters are servers (computing nodes) and associated high performance storage. There are additional nodes to meet specific needs like high amounts of memory or GPUs. All UA research faculty can sign up for a free monthly allocation. For researchers who need compute resource beyond the free standard allocation, we encourage ‘buy-in’ of additional compute nodes.

Benefits to Buy-In

Dedicated Research Compute. Research groups can ‘Buy-In’ (add resources such as processors, memory, etc.) to the base HPC systems as funding becomes available. Researchers receive 100% of the CPU*hour time their purchases create as a monthly high-priority allocation. This time receives the highest priority queue on the HPC systems.

Quality Environment. The Buy-In option allows research groups to take advantage of the central machine room space that is designed for maintaining high performance computing resources. The UITS Research Technologies group physically maintains the purchased nodes, applies updates and patches, monitors the systems for performance and security, and manages software. Additionally, Research Technologies staff is available for research support and questions through hpc-consult@list.arizona.edu. In short, essentially all costs associated with maintaining compute resources are covered by UITS rather than individual researchers.

Flexible Capacity. Buy-In research group members also benefit from their resources being integrated into a larger computing resource. This means the buy-in resources can be used in conjunction with the free allocation and resources provided to address computational projects that would be beyond the capacity of a group running an independent system alone.

Shared Resource. The University Research Computing Community as a whole benefits from buy-in expansions to the HPC systems. As mentioned above, researchers who buy-in receive 100% of the allocation of time for their purchase. However if the buy-in resources are not fully utilized, they are made available as windfall resources. This helps to ensure full use of all HPC resources and can be used to justify future purchases of computing resources.

Cost Competitiveness. Lower costs included in the grant proposals (i.e. hardware only, no operations costs) and evidence of campus costsharing give a positive advantage during funding agency review.

Pricing. For the year following the award the UA HPC request for proposal (RFP) pricing is locked in and is often considerably less than the "market price."

Buy-In Details

Puma 2020

There are several buy-in options for Puma:

1. CPU-only node: Penguin Computing Altus XE2242 CPU chassis
   a. There are 4 CPU nodes in an Altus XE2242 chassis
   b. Technical specs for 1 node of 4 in an Altus XE2242 chassis
      i. 96 cores: Dual socket AMD EPYC 7642 CPU (2x48 cores, 2.4 GHz, 225W)
      ii. 512GB RAM, DDR4-3200MHz REG, ECC, 2Rx4 (16 x 32GB)
      iii. 2TB SSD local hard drive, 2.5", NVMe, 4 Lane, 1 DWPD, 3D TLC

2. GPU node: Penguin Computing Altus XE2214GT GPU chassis
   a. GPU chassis have 4 GPUs in them
   b. Technical specs for the full XE2214GT chassis
      i. 96 cores: Dual socket AMD EPYC 7642 CPU (2x48 cores, 2.3 GHz, 225W)
ii. 4 NVIDIA Tesla V100S-PCIe, 32GB video memory, 5120 CUDA, 640 Tensor, 250W
iii. 512GB RAM, DDR4-3200MHz REG, ECC, 2Rx4 (16 x 32GB)
iv. 2TB SSD local hard drive, NVMe, 4 Lane, 1 DWPĐ, 3D TLC

3. High memory node: Penguin Computing Altus XE1212 high memory chassis
   a. Technical specs
      i. 96 cores: Dual socket AMD EPYC 7642 CPU (2x48 cores, 2.4 GHz, 225W)
      ii. 3072 GB RAM, DDR4-2933MHz LR, ECC, 4R (24 x 128GB)
      iii. 2TB SSD local hard drive, NVMe, 4 Lane, 1 DWPĐ, 3D TLC

---

**Cost and Allocations**

<table>
<thead>
<tr>
<th>Option Number</th>
<th>CPU cores</th>
<th>V100S GPU</th>
<th>RAM (GB)</th>
<th>Monthly High-priority Allocation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>512</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CPU-Only Options**

<table>
<thead>
<tr>
<th>Option Number</th>
<th>CPU cores</th>
<th>V100S GPU</th>
<th>RAM (GB)</th>
<th>Monthly High-priority Allocation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A - One CPU node</td>
<td>96</td>
<td>512</td>
<td>70,080</td>
<td></td>
<td>$8,037.50</td>
</tr>
<tr>
<td>1B - Two CPU nodes</td>
<td>192</td>
<td>512</td>
<td>140,160</td>
<td></td>
<td>$16,075.00</td>
</tr>
<tr>
<td>1C - Three CPU nodes</td>
<td>288</td>
<td>512</td>
<td>210,240</td>
<td></td>
<td>$24,112.00</td>
</tr>
<tr>
<td>1D - Full Altus XE2242</td>
<td>384</td>
<td>512</td>
<td>280,320</td>
<td></td>
<td>$32,150.00</td>
</tr>
</tbody>
</table>

**GPU Node Options**

<table>
<thead>
<tr>
<th>Option Number</th>
<th>CPU cores</th>
<th>V100S GPU</th>
<th>RAM (GB)</th>
<th>Monthly High-priority Allocation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A - 1/4 Altus XE2214GT</td>
<td>24</td>
<td>1</td>
<td>512</td>
<td>17,520</td>
<td>$8,523.75</td>
</tr>
<tr>
<td>2B - 2/4 Altus XE2214GT</td>
<td>48</td>
<td>2</td>
<td>512</td>
<td>35,040</td>
<td>$17,047.50</td>
</tr>
<tr>
<td>2C - 3/4 Altus XE2214GT</td>
<td>72</td>
<td>3</td>
<td>512</td>
<td>52,560</td>
<td>$25,571.25</td>
</tr>
<tr>
<td>2D - Full Altus XE2214GT</td>
<td>96</td>
<td>4</td>
<td>512</td>
<td>70,080</td>
<td>$34,095.00</td>
</tr>
</tbody>
</table>

**High Memory Node**

<table>
<thead>
<tr>
<th>Option Number</th>
<th>CPU cores</th>
<th>V100S GPU</th>
<th>RAM (GB)</th>
<th>Monthly High-priority Allocation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 - Full Altus XE1212</td>
<td>96</td>
<td>3072</td>
<td>70,080</td>
<td></td>
<td>$42,230.00</td>
</tr>
</tbody>
</table>

---

**Buy-in Policies**
University of Arizona can only purchase whole chassis units from Penguin Computing. That is 4 CPU nodes (option 1D), 1 GPU node with 4 GPUs (option 2D), or 1 high memory nodes (option 3). Research Computing will work to match partial node buy-in requests to make full nodes.

- Monthly high priority time is calculated as: (Number of CPUs * 24 hours * 365 year) / 12 months
- Purchasing GPUs expands the limit the PI has on number of GPUs that can be used at any time
- Buy-in high priority allocations will last the lifetime of the system. Puma was purchased in August 2020 and will be end-of-life August 2025.

The HPC Buy-in program is not designed to replace or compete with the very large-scale resources at national NSF and DOE facilities, e.g. XSEDE, the Open Science Grid. National resources are available at no financial cost to most US-based researchers through competitive proposal processes. Please contact hpc-consult@list.arizona.edu if you are interested in applying for these resources.

The HPC Buy-in program is designed to meet the needs of researchers with medium scale HPC requirements who want guaranteed, consistent access to compute resources.

High-Priority Allocation Policies

- Standard and high priority jobs will preempt windfall jobs when necessary.
- High priority jobs are run on both the buy-in nodes and the centrally-funded nodes. This is advantageous if there is a short-term project deadline.