Compute Resources

- Overview
  - El Gato
  - Ocelote
  - Puma
- Free vs Buy-In
  - Compute System Details
- Example Resource Requests

Overview

El Gato

During the quarterly maintenance cycle on April 27, 2022 the ElGato K20s were removed because they are no longer supported by Nvidia.

Implemented at the start of 2014, it has been re provisioned with CentOS 7 and new compilers and libraries. From July 2021 it has been using Slurm for job submission. El Gato is a large GPU cluster, purchased by an NSF MRI grant by researchers in Astronomy and SISTA.

Ocelote

Implemented in the middle of 2016. It is designed to support all workloads on the standard nodes except:

1. Large memory workloads beyond 188GB are handled with the large memory node which has 2TB of memory.
2. GPU workload is supported on 46 nodes with Nvidia P100 GPU's.

Puma

Implemented in 2020, Puma is the biggest cat yet. Similar to Ocelote, it has standard CPU nodes (with 94 cores and 512 GB of memory per node), GPU nodes (with Nvidia V100) and two high-memory nodes (3 TB). Local scratch storage increased to ~1.4 TB. Puma runs on CentOS 7.

Free vs Buy-In

The HPC resources at UA are differentiated from many other universities in that there is central funding for a significant portion of the available resources. Each PI receives a standard monthly allocation of hours at no charge. There is no charge to the allocation for windfall usage and that has proven to be very valuable for researchers with substantial compute requirements.

Research groups can ‘Buy-In’ (adding additional compute nodes) to the base HPC systems as funding becomes available. Buy-In research groups will have highest priority on the resources they add to the system. If the expansion resources are not fully utilized by the Buy-In group they will be made available to all users as windfall.

Compute System Details

During the quarterly maintenance cycle on April 27, 2022 the ElGato K20s and Ocelote K80s were removed because they are no longer supported by Nvidia.

<table>
<thead>
<tr>
<th>Name</th>
<th>El Gato</th>
<th>Ocelote</th>
<th>Puma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>IBM System X iDataPlex dx360 M4</td>
<td>Lenovo NeXtScale nx360 M5</td>
<td>Penguin Altus XE2242</td>
</tr>
<tr>
<td>Year Purchased</td>
<td>2013</td>
<td>2016 (2018 P100 nodes)</td>
<td>2020</td>
</tr>
</tbody>
</table>
| **Node Count** | 131 | 400 | 236 CPU-only
| **Total System Memory (TB)** | 26TB | 82.6TB | 128TB
| **Processors** | 2x Xeon E5-2650v2 8-core (Ivy Bridge) | 2x Xeon E5-2695v3 14-core (Haswell) | 2x AMD EPYC 7642 48-core (Rome)
| **Cores / Node (schedulable)** | 16c | 28c (48c - High-memory node) | 94c
| **Total Cores** | 2160* | 11528* | 23616*
| **Processor Speed** | 2.66GHz | 2.3GHz (2.4GHz - Broadwell CPUs) | 2.4GHz
| **Memory / Node** | 256GB - GPU nodes | 192GB (2TB - High-memory node) | 512GB (3TB - High-memory nodes)
| **Accelerators** | 46 NVIDIA P100 (16GB) | 29 NVIDIA V100S
| **/tmp** | ~840 GB spinning /tmp is part of root filesystem | ~840 GB spinning /tmp is part of root filesystem | ~1440 TB NVMe /tmp
| **HPL Rmax (TFlop/s)** | 46 | 382
| **OS** | Centos 7 | CentOS 7 | CentOS 7
| **Interconnect** | FDR Infiniband | FDR Infiniband for node-node 10 Gb Ethernet node-storage | 1x 25Gb/s Ethernet RDMA (RoCEv2) 1x 25Gb/s Ethernet to storage

* Includes high-memory and GPU node CPUs

**Example Resource Requests**