Buddy Scharfenberg Updated 11/23/2022

Initial Computing Model

Overview

The newest High Performance Computing (HPC) resource, Hellbender, has been provided through partnership with the Division of Research Innovation and Impact (DRII) and is intended to work in conjunction with DRII policies and priorities. This document will provide definitions about how fairshare, general access, priority access, and researcher contributions will be handled for Hellbender.

HPC has been identified as a continually growing need for researchers, as such DRII has invested in Hellbender to be an institutional resource. This investment is intended to increase ease of access to these resources, provide cutting edge technology, and grow the pool of resources available.

Fairshare

To understand how general access and priority access differs, fairshare must first be defined. Fairshare is an algorithm that is used by the scheduler to assign priority to jobs from users in a way that gives every user a fair chance at the resources available. This algorithm has several metrics to perform this calculation over for any given job waiting in the queue, such as job size, wait time, current and recent usage, and individual user priority levels. This allows administrators to tune the fairshare algorithm, to adjust how it determines which jobs are next to run once resources are available.

Resources available to everyone: General Access

General access will be open to any research or teaching faculty, staff, and students for any UM system campus. General access is defined as open access to all resources available to users of the cluster at an equal fairshare value. This means that all users will have the same level of access to the general resource.

Research users of the general access portion of the cluster will be given the RDE Standard Allocation to operate from. Larger storage allocations will be provided through RDE Advanced Allocations, and independent of HPC priority status.

Hellbender Advanced: Priority Access

When researcher needs are not being met at the general access level, researchers may request an advanced allocation on Hellbender to gain priority access.

Priority access will give research groups a limited set of resources that will be available to them without competition from ~~the~~ general access users.

Priority Access will be provided to a specific set of hardware through a priority partition which contains these resources. See appendices for hardware definitions and associated costs. This partition will be created, and limited to use by the user and their associated group. These resources will also be in an overlapping pool of resources available to general access users . This pool will be administered such that if a priority access user submits jobs to their priority access partition, any jobs running on those resources from the overlapping partition will be requeued and begin execution again on another resource in that partition if available, or return to wait in the queue for resources.

Priority access users will retain general access status, fairshare will still play a part in moderating their access to the general resource. Fairshare inside a priority partition determine which user's jobs are selected for execution next inside this partition. The jobs running inside this priority partition will also affect a user's fairshare calculations even for resources in the general access partition. Meaning that running a large amount of jobs inside a priority partition will lower a user's priority for the general resources as well.

Priority Designation

Hellbender Advanced Allocations are eligible for DRII Priority Designation. This means that DRII has determined the proposed use case (such as a core or grant-funded project) presents a strategic advantage or high priority service to the university. In this case, DRII fully subsidizes the resources used to create the Advanced Allocation.

Traditional Investment

Hellbender Advanced Allocation requests that are not approved for DRII Priority Designation may be treated as traditional investments with the researcher paying for the resources used to create the Advanced Allocation at the defined rate (see appendix A). These rates are subject to change based on the determination of DRII, and hardware costs.

Resource Management

Information Technology Research Support Solutions (ITRSS) will procure, set up, and maintain the resource. ITRSS will work in conjunction with MU Division of Information Technology and Facility Services to provide adequate infrastructure for the resource.

Resource Growth

Priority access resources will generally be made available from existing hardware in the general access pool and the funds will be retained for a future time to allow a larger pool of funds to accumulate for expansion of the resource. This will allow the greatest return on investment over time. If the general availability resources are less than 50% of the overall resource, an expansion cycle will be initiated to ensure all users will still have access to a significant amount of resources. If a researcher or research group is contributing a large amount of funding, it may trigger an expansion cycle if that is determined to be advantageous at the time of the contribution.

Appendix A: Hellbender Advanced Allocation

All hardware is subsidized by DRII at a rate of 25% regardless of DRII priority designation status. Any costs associated with a node definition have already been calculated to be the remaining unsubsidized amount.

- Compute Node Definitions
  - Dell C6525
    - CPU: AMD 7713 Epyc Milan Processor
      - Qty: 2
      - Cores: 64 physical per cpu (128 per node)
      - Base clock: 2.0GHz
      - Boost clock: 3.675GHz
      - L3 Cache: 256MB
    - RAM: 512 GB DDR4 - 3200
    - Local scratch: 1.6TB NVME dedicated
    - Network: HDR-200 Infiniband connection
      - Bandwidth: 200Gb/s
      - Latency: less than 600 nanoseconds
      - Purpose: MPI communications and access to all network storage.
    - Priority Cost: $2,701.99 per year per node
    - Current total Quantity: 112
- GPU Node Definitions
  - Dell R750XA
    - CPU: Intel Xeon Gold 6338 Ice Lake Processor
      - Qty: 2
      - Cores: 32 physical per cpu (64 per node)
      - Base clock: 2.0GHz
      - Turbo clock: 3.2GHz
      - L3 cache: 48MB

- GPU: Nvidia Ampere A100
  - Qty: 4
  - RAM: 80GB each (320GB total)
  - Cuda cores: 6912 each (27,648 total)
  - Base clock: 1065MHz
  - Boost clock: 1410MHz
- RAM: 256 GB DDR4 - 3200
- Local scratch: 1.6TB NVME dedicated
- Network: HDR-200 Infiniband connection
  - Bandwidth: 200Gb/s
  - Latency: less than 600 nanoseconds
  - Purpose: MPI communications and access to all network storage.
- Priority Cost: $7,691.38 per year per node
- Current total Quantity: 17