University at Buffalo Center for Computational Research

The following is a short and long description of CCR Facilities for use in proposals, reports, and presentations. If desired, a letter of support for your proposal can also be written. In addition to the computing, storage, and services listed below, CCR staff members, who have extensive expertise in high-performance computing, high-capacity data storage, advanced database design, custom software development, bioinformatics support, and scientific visualization, are available to facilitate your research.

At a minimum, faculty using CCR resources are required to:

1. **Acknowledge CCR support in publications**,
2. Include **a percent credit to CCR** on the Sponsored Programs Approval Form. A typical percent credit is 5%, though this should be larger for grants requiring significant CCR resources. NOTE: Clicking the "use CCR" box on the Sponsored Programs Approval Form is not the same as including a percent credit to CCR.

Want your jobs to start sooner?
While there is no cost for UB faculty groups to use CCR compute resources, queue priority is based on contribution level. In order to receive a priority boost in your jobs, you must have a **funded grant** that meets the criteria above and **includes direct funds to CCR.** At minimum, you should budget $2,600 per year (roughly 40 hours/year of CCR support staff time at the SPA approved rate of $65/hour for typical projects - higher amounts as appropriate for projects that are expected to more heavily utilize CCR services). Boosting your group's priority will substantially increase your job throughput. Please enter your grant information in your ColdFront project and notify CCR Help with your grant information so we can provide you with the boost in queue priority.

CCR Acknowledgment: *This work was performed in part at the University at Buffalo's Center for Computational Research [1]. Use of the NIH sub-cluster should also acknowledge NIH award S10OD024973 ("High Performance Data and Computing Infrastructure"), and use of the NSF sub-cluster should acknowledge NSF award 1724891 ("MRI: Acquisition of High Performance Computing Infrastructure to Support Computational and Data-Enabled Science and Engineering").*
[1] Center for Computational Research, University at Buffalo, http://hdl.handle.net/10477/79221.

BiBTeX entry:

```
@misc{UBCCR,
   author= {{Center for Computational Research, University at Buffalo}},
   year  = {2020},
   title = {{UB CCR Support Portfolio,}},
   note  = {\url{http://hdl.handle.net/10477/79221}},
}
```

SHORT DESCRIPTION:

The Center for Computational Research (http://buffalo.edu/ccr), a leading academic supercomputing facility, maintains a high-performance computing environment, on-premise research cloud environment, remote visualization capabilities, and support staff with expertise in computing, visualization, and networking. The Center's extensive computing facilities, which are housed in a state-of-the-art 4000 sq ft machine room, include a generally accessible (to all UB researchers) Linux cluster with more than 30000 processor cores and high-performance Infiniband/Omni-Path networks, a subset of which contain NVidia Tesla V100 graphics processing units (GPUs). Industrial partners of the University have access to a cluster with more than 3400 processor cores and high-performance networking and storage. The Center maintains a 1.1PB Panasas parallel scratch storage solution plus a 2PB VAST all-flash shared network attached filesystem. Remote visualization capabilities are provided through five graphics-enabled GPU servers accessible through CCR's OnDemand web portal. A leading academic supercomputing facility, CCR has more than 1 PFlop/s of peak performance compute capacity. CCR additionally hosts a number of clusters and specialized storage devices for various departments, projects, and collaborations. Researchers interested in hosting services should contact CCR staff.

In addition to its computing and visualization resources, CCR has a support staff consisting of computational scientists, programmers, and system administrators with expertise in all areas of computing, including scientific and parallel computing, (big) data analytics, custom software development, database engineering, and scientific visualization.

MORE DETAILED DESCRIPTION:
CCR’s resources include the following:

x86_64 Linux Cluster: CCR's general purpose (and general access, see below for faculty/project specific sub-clusters) x86_64 Linux cluster consists of a mix of sub-clusters of varied capabilities:

Compute Nodes:

- 24 Dell C6420 quad-server chassis, each with 4 servers (96 total) containing two “Cascade Lake” Xeon Gold 6230 2.1GHz processors, 192GB memory, and Mellanox Infiniband (IB) interconnect,
- 86 Dell R440 servers with two “Skylake” Xeon Gold 6130 2.1GHz processors, 192GB of memory, Intel Omni-Path interconnect (OPA),
- 372 Dell C6100 servers, each with two Intel “Westmere” Xeon 2.40GHz (E5645) processors, 48GB of memory, QDR Infiniband,
- 128 IBM iDataPlex dx360 M2 servers, each of which has two Intel “Nehalem” Xeon 2.26GHz (L5520) processors, 24GB of memory, QDR Infiniband,
- 128 Dell C6100 servers containing two Intel “Westmere” Xeon 2.13GHz (L5630) processors and 24GB of memory, QDR Infiniband,
- 34 Dell R620 servers each with two Intel “Sandy Bridge” 2.26GHz Xeon (E5-2660) processors, QDR Infiniband,

A GPU enabled sub-cluster:

- 8 Dell R740 servers with two “Cascade Lake” Xeon Gold 6230 2.1GHz processors, 192GB memory, and Mellanox IB interconnect, and 16 Dell R740 servers with two Intel “Skylake” Xeon Gold 6130 2.1GHz sixteen core processors, 192GB of memory, and Intel OPA. All 24 servers are each outfitted with two NVidia “Volta” V100 graphics processing units (GPUs).

A “large memory” sub-cluster:

- 6 Dell C6420 server chasses (4 servers/chassis), each with two Intel “Cascade Lake” Xeon 6230 2.1GHz processors (40 cores/server total), 768GB memory, 16 Dell R640 servers with two Intel “Skylake” Xeon Gold 6130 2.1GHz processors (32-cores total), 768GB memory, and Omni-Path interconnect. Older nodes in this pool include a single server (Dell R910) with four Intel “Beckton” Xeon 2.0GHz (X7550) eight-core processors has 256GB of main memory. Additional nodes are available with 32-cores each and either 256GB (8 with Intel “Westmere” Xeon (E7-4830) processors, 8 with AMD “Magny-Cours” Opteron (6132HE) 2.2GHz processors) or 512GB of memory (2 with Intel “Westmere” Xeon (E7-4830) processors).

Core Network: In addition to the high-performance low-latency Infiniband/Omni-Path networks, all of CCR’s clusters are interconnected with a 10/40 gigabit Ethernet (10gigE/40gigE) core network from Arista (6508 switch), with each server connected to either 10gigE of 1gigE gigabit Ethernet leaf switches, uplinked by dual 10gigE or 40gigE links to the core. All nodes have local disk (typically 8:1 volume ratio compared to local memory), with the newest nodes utilizing higher performance solid state drives (SSDs). The peak performance of this general access cluster overall is approximately 1 PFlop/s.

Industrial Users/Economic Outreach Cluster: This cluster consists of 216 HP SL230 Gen8 servers of which 144 are "Parallel" compute nodes with an FDR InfiniBand interconnect. The remaining 72 nodes are "Serial" nodes with GigE connections. The nodes consist of 2 Intel "Ivy Bridge" Xeon 2.6GHz (E5-2650V2) 8-core processors (total of 3,456 cores), 64GB of memory and 500GB of local scratch. This cluster is primarily for industrial partners and the idle CPU cycles are made available to CCR users with the understanding that their jobs will be preempted if an industrial user job is submitted. The peak performance of this cluster overall is approximately 70 TFlop/s, and is due to be refreshed in 2020.

Faculty/Departmental Clusters: CCR hosts private clusters for researchers in various departments. Such faculty/departmental clusters are maintained similarly to the general production systems with a common operating environment. Access to these resources is generally reserved to the faculty/department that procured them, but also includes a pre-emptible scavenger partition/queue that allows the general research community to leverage them when underutilized by the faculty/department. Currently these clusters number more than 535 nodes of various configurations (some including GPUs) across more than 40 faculty in various departments. All such clusters are also integrated within the core CCR 10gigE Ethernet network and storage systems. Frequently such systems are used for specialized instruction in HPC topics, or for more general disciplinary use (e.g., a small cluster is routinely used for chemistry instruction on a shared enterprise WebMO service).

Panasas High Performance storage solution. The Panasas storage consists of two ASD-100 director modules containing a total of 6 directors and five ASU-100 disk modules with 1.4PB raw, 1.1PB usable space. This will be CCR's high performance parallel file system. With 8 x 40Gbit connections to the Arista core switch the storage provides 19GB/sec write and 13GB/sec read I/O performance. Designed for high performance and concurrent access, CCR's Panasas storage is primarily intended for generation and analyses of large quantities of short-lived data (scratch usage).

Vast Data Storage: The Vast Data storage system consists of over 2PB of usable flash storage, connected to the CCR core network via twelve 40GigE links to our core Arista switch. This storage system serves as the high reliability core storage for home and projects usage. The storage is designed to tolerate simultaneous failures, helping assure the 24x7x365 availability of the Center's primary storage. The storage provides up to 10GB/sec write and 36GB/sec read performance.

Remote Visualization Capabilities: CCR provides five graphics-enabled GPU servers for users that require use of an OpenGL application GUI. These servers are housed within the UB-HPC (academic) cluster but are dedicated for remote visualization use. Users access these servers through the OnDemand web portal which provides access to a visualization session for up to 24 hours. The servers supporting the remote visualization services vary in configuration, but all include two-four NVIDIA Quadro P4000 or K4200 GPUs per server and at least 256GB of memory.
Staff: CCR maintains a support staff of 10, including 2 Ph.D. level computational scientists, an HPC support specialist, 5 systems analysts/administrators, an economic outreach engagement specialist, and 1 administrative support staff. Currently an additional 10 research staff (including 4 Ph.D. level computational scientists) with additional expertise in software engineering, statistics, and computational biology are also available for engaging with sponsored research projects.

Research cloud: CCR's Lake Effect private cloud is a subscription-based Infrastructure as a

Service (IAAS) research cloud that provides users with root-level, on-demand access to virtual servers and storage. Backed by RedHat OpenStack software version 13 and RedHat Ceph version 3, Lake Effect is compatible with Amazon AWS and EC2 instances. It includes a variety of instance sizes, of which a 16 core 64 GB is currently the largest. Lake Effect is currently provisioned with 1088 available vCPU cores, 21 hypervisors, 720 TB of dedicated Ceph storage, and 2x Nvidia V100 32GPUs.

External Network: In addition to high-end computing, the on-premise cloud environment and remote visualization resources, high-speed network connectivity is provided to the university's networks and beyond. The University at Buffalo provides fiber-optic backbone networking between its three campuses. CCR currently has a 10Gb connection to UB's 100Gb core, with plans to upgrade to 10G/40G/100G in summer 2020. The University at Buffalo is an Internet2 member. CCR offers a data transfer node reserved for transferring files to/from CCR as well as to UBBox, and hosts a Globus (http://globus.org) endpoint for the University at Buffalo. Globus provides secure and fast data transfer to/from the CCR storage and other endpoints around the world.