

## 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 recognition credit to CCR** on the Sponsored Programs Approval Form. A typical percent recognition credit is 5%, though this should be larger for grants requiring significant CCR resources.

### Want your jobs to start sooner?

While there is no cost for UB faculty groups to use the compute resources in the CCR academic cluster, [job 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](#), 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 4,000 sq ft machine room, include a generally accessible (to all UB researchers) Linux cluster with more than 30,000 processor cores and high-performance Infiniband/Omni-Path networks, a subset of which contain NVidia Tesla H100, A100, and V100 graphics processing units (GPUs). Industrial partners of the University have access to a cluster with more than 5,500 processor cores (including GPUs) and high-performance networking and storage. The Center maintains a 4.5PB VAST all-flash shared network attached filesystem. Remote visualization capabilities are provided through three graphics-enabled GPU servers accessible through CCR's OnDemand web portal. A leading academic supercomputing facility, CCR has more than 3 PFlop/s of peak performance compute capacity. CCR also hosts several 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, system administrators, and research computing and data science facilitators with expertise in all areas of computing, including scientific and parallel computing, 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 the faculty/project specific cluster) 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)

**GPU enabled nodes:**

- 8 Dell R740 servers with two "Cascade Lake" Xeon Gold 6230 2.1GHz processors, 192GB memory, Mellanox IB interconnect, and two NVidia "Volta" V100 graphics processing units (GPUs) with 32GB of GPU memory
- 17 Dell R740 servers with two Intel "Skylake" Xeon Gold 6130 2.1GHz sixteen core processors, 192GB of memory, Intel OPA, and two NVidia "Volta" V100 graphics processing units (GPUs) with 16GB of GPU memory

**"Large memory" nodes:**

- 6 Dell C6420 server chassis (4 servers/chassis), each with two Intel "Cascade Lake" Xeon 6230 2.1GHz processors (40 cores/server total), 768GB memory, and Mellanox IB interconnect
- 16 Dell R640 servers with two Intel "Skylake" Xeon Gold 6130 2.1GHz processors (32-cores total), 768GB memory, and Omni-Path interconnect

*Core Network:* In addition to the high-performance low-latency Infiniband/Omni-Path networks, all of CCR's cluster nodes are connected at 10GigE or 40GigE to rack switches, then 100GigE interconnected to the 100GigE Arista core switch. All nodes have local storage utilizing high performance solid state drives (SSDs) (typically 8:1 volume ratio compared to local memory).

*Industrial Users/Economic Outreach Cluster:* This cluster consists of two generations of Dell servers. The first generation of 99 nodes are connected with an HDR InfiniBand interconnect. The majority of these nodes consist of 2 Intel "Ice Lake" Xeon Gold 6330 2.0GHz 28-core processors, 512GB of memory and 960GB of local scratch. Sixteen of the nodes are large memory nodes with 1TB of RAM and 16 of the nodes contain dual NVIDIA A100 GPUs. The second generation consists of 88 servers with 2 Intel "Sapphire Rapids" Xeon Gold 6448Y 2.1GHz 32-core processors and a minimum of 512GB main memory connected with NDR200 Infiniband. Sixteen of these servers have 1TB of memory and four are equipped with NVidia H100 GPUs. In total, there are 11,176 available cores. 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.

*Faculty/Departmental Cluster:* CCR hosts a private cluster for researchers in various departments. The faculty/departmental cluster is 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 this cluster includes nearly 500 nodes (over 21,000 cores) of various configurations (some including GPUs) across more than 50 faculty in various departments. This cluster is also integrated within the core CCR 10GigE/40GigE

Ethernet network and storage system. Frequently such systems are used for specialized instruction in HPC topics, or for more general disciplinary use.

*Vast Data Storage:* The Vast Data storage system consists of over 4.5PB of usable flash storage, connected to the CCR core network via twelve 100GigE 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 more than 10GB/sec write and 36GB/sec read performance and supports serial and parallel I/O, tested with OpenMPI and Intel MPI. This storage is used for both permanent project storage and global scratch directories.

*Remote Visualization Capabilities:* CCR provides two 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 Open OnDemand web portal which provides access to a visualization session for up to 24 hours. The servers include three NVIDIA A16 or two A40 GPUs per server and at least 512GB of memory.

*Staff:* CCR maintains a support staff of 12, including 2 Ph.D. level computational scientists, 6 systems analysts/administrators, 2 research computing and data science facilitators, an economic outreach engagement specialist, and 1 administrative support staff. Currently an additional 10 research staff (including 4 Ph.D. level computational scientists) with expertise in software engineering, statistics, and various disciplines are also available for engaging with sponsored research projects.

*Research cloud:* CCR's Lake Effect cloud is a subscription-based Infrastructure-as-a-Service (IAAS) on-premise private cloud that provides research users with root-level, on-demand access to virtual servers and storage. Backed by OpenStack software version 23 (Wallaby) and Ceph version 16 (Pacific), Lake Effect is comparable with Amazon AWS EC2. It includes a variety of instance sizes (flavors), of which a 32 core with 128 GB of RAM is currently the largest. Lake Effect is currently provisioned with more than 1200 available vCPU cores, 25 hypervisors, 7TB of local RAM, and 900 TB of dedicated Ceph storage. Two servers include Nvidia V100 32GB GPUs. Funds collected for the use of this service are used to purchase additional hardware, so the CCR cloud is always expanding.

*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's data center has a 100Gb connection to UB's 100Gb core, with internal racks connected at 40G/100G. The University at Buffalo is an Internet2 member. CCR offers a Globus data transfer service for transferring files to/from CCR, as well as to UBBox and Microsoft OneDrive. Globus provides secure and fast data transfer to/from the CCR storage and other endpoints around the world.