

# Institute for Digital Research and Education (IDRE) Facilities

## High Performance Computing and Storage, Operations and Support

IDRE's Computation and Storage program was established to promote multi-disciplinary computational research and serve as a catalyst for solving most challenging computational and data management and analysis problems. It addresses the high-end computational and storage needs of UCLA's students, researchers and faculty by seeding a pool of computing and storage hardware. The program provides a centralized platform where researchers from different disciplines can collaborate and form a community to maintain and share a common compute infrastructure. It is instrumental in supporting research grants and advancing a compute and data hosting facility in terms of administration and maintenance of scalable computing systems, scalable storage, and associated infrastructure and services.

Resources related to the IDRE Computation and Storage program include IDRE's Research Technology Group which comprises a set of very highly skilled people in system administration, system maintenance, user services, and high performance computing applications. The group leverages its expertise by maintaining and providing the Hoffman2 Shared Compute Cluster, associated high performance Shared Storage, the Cloud Archival Storage Service, and the three IDRE data centers to support novel computational science at UCLA under IDRE's Computation and Storage Program.

## Hoffman2 Shared Cluster

UCLA's Shared Hoffman2 Cluster currently consists of approximately 1,200 64-bit nodes and 13,000+ cores, with an aggregate of over 50TB of memory. Each node has 1GB Ethernet network and a DDR, QDR, or FDR Infiniband interconnect. The cluster includes a job scheduler, compilers for C, C++, Fortran 77, 90 and 95 on the current Shared Cluster architecture, applications and software libraries that offer languages, compilers and software specific to Chemistry, Chemical Engineering, Engineering, Mathematics, Visualization, Programming and an array of miscellaneous software. The current peak CPU performance of the cluster is approximately 160 Trillion Floating Point, double precision, operations per second (TFLOPS). Hoffman2 is currently the largest and most powerful cluster in the University of California system.

Additional Hoffman2 resources for researchers include complete system administration for contributed cores, cluster access through dual, redundant 10Gb network interconnects to the campus backbone, the capability to run large parallel jobs that can take advantage of the cluster's InfiniBand interconnect, and web access to the Hoffman2 Cluster through the UCLA Grid Portal, as well as access to a Panasas parallel filesystem and a NetApp storage system. Current HPC storage capacity is 2 petabytes.

The cluster is also an end point on the Globus Online service using the 10Gb network interconnect backbone, thus providing researchers a facility for fast and reliable data movement between Hoffman2 and most leadership class facilities across the USA.

## Dawson2 GPU Cluster

UCLA's Dawson2 GPU Cluster, ranked 148 in the 2010 Top500, comprises 96 HP ProLiant SL390 G7 systems, each having dual socket Intel Xeon X5650 processors, 3 Nvidia M2070 Graphics processors, and 48 GB of main memory giving peak performance of 150 double precision Trillion Floating Point operations per second (TFLOPS) and achieved 70Tflops on Linpak. The cluster uses QDR Infiniband networking and 160 Terabytes of high performance common disk space from Panasas for communication and storage respectively.

Resources on Dawson2 are available for GPU software development.

## Data Center System

IDRE has been instrumental in creating and implementing a strategic campus data center plan to effectively support research computing. IDRE's data center system includes space in the following facilities:

### *The Math Sciences Data Center*

The Math Sciences Data Center, which is shared with the campus administrative systems unit, houses research computing clusters, including part of the main Hoffman2 Cluster. The Math Sciences Data Center also provides backend services to Visualization Portal and Modeling Lab. Approximately 2,700 square feet of the Math Sciences data center's 5,000 square feet are dedicated to supporting IDRE research computing. The compute resources in this facility and the IDRE Performance Optimized Data Center (POD) are networked through 10gigabit Ethernet as well as wide-area Infiniband fabric. The Math Sciences facility is a Tier 3 data center space with greater than 600kW of power with full UPS and motor generator backup and 170 tons of redundant air conditioning capacity. The facility has a strong physical security, which includes 24 x 7 staff monitoring and physical presence in the center. In addition to cluster compute nodes, the Math Sciences Data Center houses critical administrative servers, storage subsystems (over 5PB of storage located on Nexsan, Panasas, and NetApp storage servers), an IBM Tivoli tape backup system, and the primary instance of the IDRE Cloud Storage system.

### *The IDRE Performance Optimized Data Center (POD)*

The capacity of the Math Sciences Data Center has been extended through the innovative use of an HP Performance Optimized Data Center (POD). The POD is a retrofitted 40' by 8' shipping container that has been extensively modified to provide space for an additional 1,535 nodes and 18,000+ cores, as well as associated network and interconnect equipment. The POD provides 380KW of power and 170 tons of air conditioning in a highly efficient manner at a fraction of the cost of a conventional data center. The POD is also the home of a replicated instance of the IDRE Cloud Storage system.

## Access to National High-End Computational Resources

IDRE is part of the NSF XSEDE Campus Champion Program, which provides information about national high performance computing opportunities and resources and assists researchers by:

- Providing information about high performance computing and XSEDE resources
- Assisting in getting researchers access to allocations of high performance computing resources
- Facilitating workshops about the use of high performance computing resources and services
- Providing contacts within the high performance computing community for quick problem resolution

IDRE collaborates with San Diego Supercomputer Center's HPC@UC program. It can assist with scaling issues and can help predict the timing on large computing resources. IDRE also has strong relationships with NSF and DOE centers, including NERSC, NASA, ALCC and INCITE.

## IDRE Research Network and UCLA Science DMZ

During 2017 IDRE, with assistance from the NSF, will place a new, high-performance upgrade to its internal network and external connections into production. Internally, bandwidth is being increased to 400 gigabits between the Hoffman2 cluster and our HPC and CASS storage systems. Externally, we are increasing bandwidth to redundant 100 gigabit connections to the UCLA Science DMZ. These links will enable direct, high-performance connections from UCLA schools, departments and labs to the IDRE data centers as well as high-throughput capabilities to national laboratories and remote collaborators.

## IDRE Storage Options

IDRE offers two main data storage services. Our HPC Shared Storage program provides high performance storage and scratch space for use with the Hoffman2 Shared Cluster and our Cloud Archival Storage Service (CASS) for backup, archival and sharing of research data. IDRE's Cloud and virtual computing research and development work is mainly focused on Infrastructure as a Service (IaaS) and Storage as a Service (SaaS). These services provide users with the ability to provision processing, storage, networks, and other fundamental computing resources to deploy and run arbitrary software, which can include operating systems and applications and the ability to create storage space on demand.

### *HPC Storage*

Our HPC Shared Storage provides NFS mounted file systems for home directories, collaborative projects, and scratch space for use with our Shared Cluster. For storage associated with jobs on the HPC clusters (both the Virtual Research Shared Cluster and the General Purpose Cluster), users have the option of paying a one-

time, per terabyte, charge for storage on the Panasas or NetApp storage system. This is particularly an important option for those that need more than the 20 GB directory space per user that is standard on the Research Virtual Shared or General Purpose Cluster or that want increased permanent space for large data sets to avoid recurring upload and transfer times.

#### *Cloud Storage Program*

IDRE's CASS is a 6+ petabyte storage resource for campus researchers. The goal of this program is to provide archival, backup and sharing of research data at a cost and service level that is better than commercial cloud providers, research groups, and departmental implementations.

### Facilities for Interdisciplinary Collaborations

#### *IDRE Technology Sandbox*

The Technology Sandbox is an interdisciplinary computing facility open to UCLA students, researchers, faculty and staff researching new technologies for digital scholarship. The Sandbox maintains a wide variety of computer modeling, GIS, web authoring, web programming, graphics, animation, image processing, compression and archiving applications, as well as a meeting and collaboration space.

#### *IDRE Portal*

The purpose of the IDRE Portal is to provide a state-of-the-art presentation and meeting facility for UCLA faculty researchers, to provide a meeting place for scheduled IT governance and policy groups, and to host IDRE-sponsored classes, seminars and presentations. Beyond these three main purposes, the Portal is made available for campus events on a scheduled basis that makes effective use of IDRE's Events and Infrastructure Support resources.