**Facilities, Equipment, and Other Resources**

***San Diego State University (SDSU)***

SDSU’s main campus data center is a 5,000 square foot facility located in room EBA-111 that offers colocation server rack space with appropriate power, cooling, and security to host computing hardware for business, research, and academic purposes. The data center includes server racks with redundant power distribution to each rack; redundant battery backup systems backed by a diesel generator; redundant computer room air conditioning units; multizone data center fire suppression system (Halon 1301 - SDSU was grandfathered in, prior to the EPA ban in 1994); data center infrastructure management (DCIM) system which provides monitoring and alerting for environmental sensors for temperature, humidity and water leaks; redundant high speed 10G and 100G networking; redundant next generation firewall systems with highly segmented tiered zones; multi factor data center access control with key card readers and biometric palm readers, an electronic key box for server rack access control; and video surveillance inside and outside of the data center.

At time of writing, the main campus data center is in the process of a major upgrade which includes increasing overall power capacity from 200kVA to 450kVA, installation of two redundant 8.5kVA rack power distribution units in each rack, and 36 new server rack cabinets with cold air containment. The completed project will also include four redundant 20-ton computer room air conditioning units. The first phase of construction is complete which included installation of 18 new server racks for campus business systems as well as an air conditioning upgrade. The second phase of construction is underway and includes 18 additional server racks for research and academic purposes as well as the installation of two new redundant data center power distribution units to increase overall data center power capacity.

SDSU’s Computational Science and Research Center (CSRC) data center is a 700 square foot facility located in the basement of North Life Science (NLS) building room 29E, shown within the red rectangle of Figure 1.

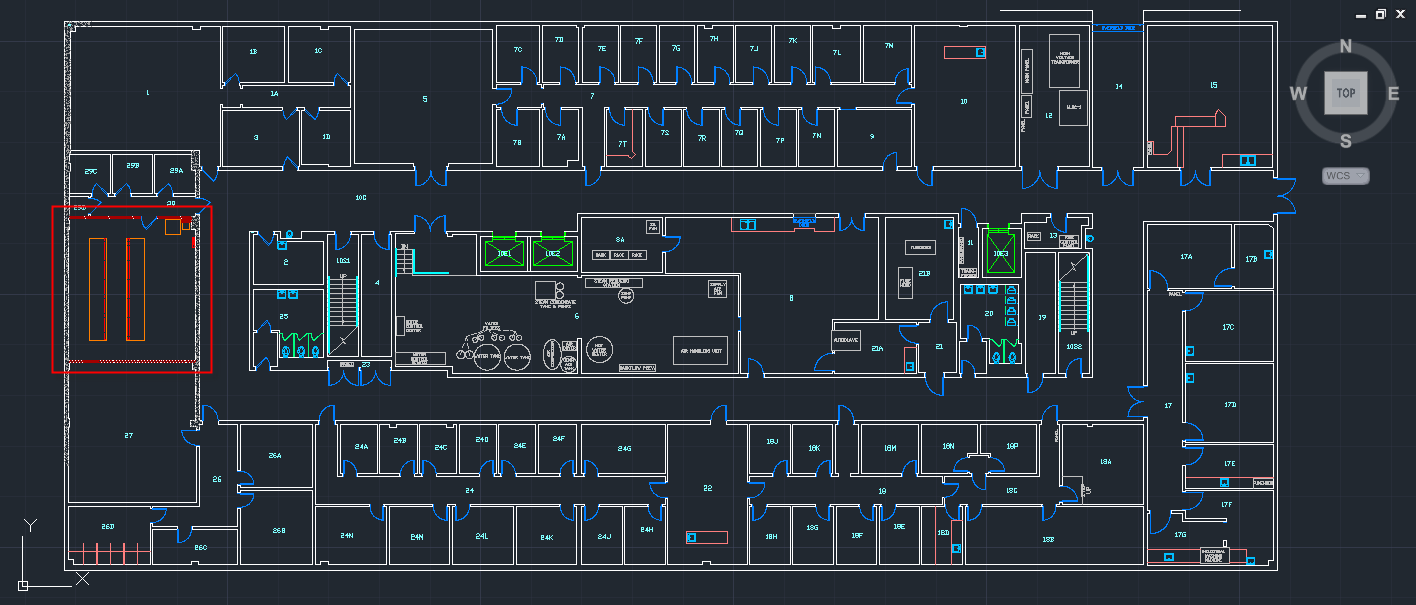


Figure . Floor plan of the SDSU North Life Science (NLS) building basement. The room shown within the red rectangle is room NLS-29E, the CSERC Data Center. The two smaller orange rectangles represent the location of two equipment rows, each accommodating 11 full-height server racks.

The two smaller rectangles in the floor plan represent the location of two equipment rows, each accommodating 11 full-height server racks. The CSRC data center houses the campus core Science DMZ switching infrastructure, a resource that enables high-speed scientific data transfer between collaborators at other research universities and national laboratories. The Science DMZ network is physically separate from the San Diego State University enterprise network and is configured to optimize the performance of scientific applications and scientific data sharing. Globus GridFTP is installed on Science DMZ Data Transfer Notes (“DTNs”) to enable efficient transfer of large datasets. The Science DMZ was established and funded through NSF Office of CyberInfrastructure CC-NIE Grant 1245312 "*CC-NIE Network Infrastructure: Implementation of a Science DMZ at San Diego State University to Facilitate High-Performance Data Transfer for Scientific Applications*". A Network connectivity diagram of the Science DMZ is shown in Figure 2. The CSRC currently maintains research computing and network infrastructure in eight different buildings on the SDSU campus. These areas are shown in Figure 3 and identified as

1. NLS-29E in the North Life Science building where the CSERC Data Center resides

Graphical user interface

Description automatically generated

Figure . Network connectivity diagram for the San Diego State University Science DMZ.

1. P-244 located in the Physics building where computational science doctoral students have personal office space
2. PA-213 located in the Physics-Astronomy building where Dr. Karen Campbell in Mathematical Statistics has her research laboratory
3. PS-106A located in the Physical Sciences building where Prof. Robert Edwards in Biology has his research laboratory
4. ENG-207B located in the Engineering building where the cluster mixcoatl.sdsu.edu resides,
5. CSL-121A located in the Chemical Sciences Laboratory where the San Diego State University Visualization Center resides
6. GMCS-214 located in the Geology, Mathematics, and Computer Science building where new office space for CSRC doctoral students reside and GMCS-218 where Geological Science faculty and doctoral students connect to the Science DMZ
7. EBA-109 located in the Education and Business Administration building where the CSRC visualization lab and the cluster dulcinea.sdsu.edu resides and EBA-111 where the SDSU main data center and the HPC cluster *fermi.sdsu.edu* resides.

The CSRC data center houses several campus High Performance Computing Resources, such as *beehive.sdsu.edu* and *mixcoatl.sdsu.edu* that are currently used by the SDSU researchers listed in Project Description section (b) for running numerical codes that support individual science drivers.

Campus Map.tif

Figure . Excerpt of the SDSU campus map showing room locations of Computational Science and Research Center computer server rooms and Science DMZ network access points.

"Mixcoatl" is a College of Engineering owned Linux cluster DTN running CentOS release 6.5 that resides in Engineering room 207B. Shown in Figure 4, Mixcoatl is a Dell PowerEdge M1000e blade enclosure with 16 Dell PowerEdge M620 compute nodes, where each node has two Intel Xeon E5-2680 (2.70GHz, 346 GFLOPs) octa-core CPUs, yielding a total core count of 256. Each node has 128 GB of physical memory, yielding a total physical memory capacity of approximately 2TB. Local scratch storage is provided by two 100GB SSD HDDs/node. Network connectivity to the Science DMZ is 10 Gbps Ethernet, while inter-blade connectivity is via 4x Fourteen Data Rate (FDR) InfiniBand (54.54 Gbps) through Mellanox Connect X3 PCI mezzanine cards. Mixcoatl is used for running carbon capture, utilization, and geologic sequestration codes, ANSYS Fluent, ANSYS HFSS, SIMULIA ABAQUS, OpenFOAM, ConvergeCFD, MATLAB, and the HIT3DP pseudospectral DNS code. Theoretical TFLOPs = 88.58.



Figure . mixcoatl.sdsu.edu. Dell PowerEdge M1000e blade enclosure with 16 compute nodes.

"Fermi" is a Department of Aerospace and Department of Mechanical Engineering owned Linux cluster running CentOS release 6.8 that resides in Education and Business Administration (EBA) building 111, which is the SDSU main data center. Fermi consists of a head server with four cores and four Microway Xeon 2U Twin² Servers that each house four compute-nodes, giving a total of 16 compute nodes plus a head node. Two Twin² Servers have 12 cores per compute-node and the other two Twin² Servers have 20 cores per compute-node. The total core count is 4 + 8(12) + 8(20) = 260. The 12-core nodes have Intel Xeon E5-2630 v2 @ 2.60GHz CPUs and the 20-core nodes have Intel Xeon E5-2650 v3 @ 2.30GHz CPUs. SDSU researcher John Abraham uses Fermi for running high-fidelity simulations of turbulent combustion in two-phase mixtures with applications to internal combustion engines.

“Dulcinea” is a CSRC owned Linux cluster running CentOS release 5.5 that resides in Education and Business Administration building 109, which is the CSRC Visualization Lab. Dulcinea consists of a head node and 14 compute nodes. Compute nodes 1 through 10 each have 8 physical cores via dual Intel E5520 @ 2.27GHz CPUs and 24GB RAM. Compute nodes 11 and 12 each have 8 physical cores via dual Intel E5620 @ 2.40GHz CPUs with 48GB RAM. Compute nodes 13 and 14 each have 16 physical cores via dual Intel E5-2650 @ 2.60GHz CPUs with 64GB RAM. The total core count is 8 + 12(8) + 2(16) = 136. Compute nodes 1 through 12 each have two Nvidia Tesla M1060 GPU cards. SDSU researcher Gustaaf Jacobs uses Dulcinea for running high-order PSIC method codes for simulation of pulse detonation engines.

The CSRC implemented and now manages the SDSU Science DMZ network using infrastructure acquired through NSF CC-NIE award 1245312 “*Network Infrastructure: Implementation of a Science DMZ at San Diego State University to Facilitate High-Performance Data Transfer for Scientific Applications*” (12/01/2012 – 05/31/2016). PI Paolini successfully implemented a Science DMZ at SDSU using Alcatel-Lucent and Brocade MLXe-4 switching infrastructure, two independent 10 Gbps uplinks to the CENIC CalREN-DC ("Digital California") and CalREN-HPR ("High-Performance Research") networks, and a 100 Gbps uplink to the HPR network acquired through SDSU’s participation in the NSF funded project *Pacific Research Platform*. Science DMZ performance monitoring is accomplished using the perfSONAR Performance Toolkit. Our perfSONAR host is publically accessible via URL <http://perfsonar.sdsu.edu/> (IPv6 and IPv4). Scientific datasets are transferred daily between our Science DMZ and XSEDE systems using Globus GridFTP. Access to cyberinfrastructure residing within our Science DMZ is accomplished using the CILogin Service and our university Shibboleth identity management system. CILogon is a research and scholarship service provider in the InCommon federation. Our Science DMZ identity management system supports the InCommon Research and Scholarship Category to provide collaborative services for researchers and scholars via their federated identities. Figure 2 shows a network connectivity diagram of the San Diego State University Science DMZ. The California Research and Education Network (“CalREN”) is a multitiered, advanced network-services fabric that is managed by the Corporation for Education Network Initiatives in California (“CENIC”) and serves research and education institutions in the state of California. The CalREN-DC ("Digital California") network provides connectivity to the commercial Internet. Most educational institutions (K-12, community colleges, and universities) obtain Internet connectivity through the CalREN-DC network tier. The CalREN-HPR ("High-Performance Research") network provides 10 Gbps and 100 Gbps connectivity among research universities and institutions in California, and connectivity to Internet2, the National LambdaRail, and ESnet, over a 100 Gbps backbone.

The CSRC has a Mellanox SX6036 36-port Non-blocking Managed 56Gb/s InfiniBand/VPI Switch, along with nine InfiniBand cables, to connect the *beehive.sdsu.edu* BeeGFS cluster components and provide 56Gb/s Fourteen Data Rate ("FDR") InfiniBand connectivity to Linux stations located in the CSRC data center.

“Beehive” is a 2.4 petabyte (PB) parallel storage cluster that runs the open-source BeeGFS file system. In 2017, SDSU coinvestigators Christopher Paolini, Jose Castillo, Gustaaf Jacobs, John Abraham, and Aram Alimohammad Kalhor received a $199,998 grant from the NSF Directorate for Computer and Information Science and Engineering to support their project *CC\* Storage Implementation of a Distributed, Shareable, and Parallel Storage Resource at CSRC to Facilitate High Performance Computing for Climate Science*. Climate Science in the XXI century is a complex enterprise. Scientists analyzing atmospheric, geologic and hydrological data have to perform calculations with sophisticated equipment that produces massive amounts of information at a very fine resolution and from multiple variables. Global and regional climate models look at the atmosphere, the ocean, and land and sea ice to make projections of the future climate. To study microfractures in rocks, scientists use a subsurface flow application capable of generating petascale datasets. Petascale refers to computer performance in excess of one petaflops, i.e. one quadrillion floating point operations. The massive amount of data and operations required in these fields of science would be staggering for a personal computer connected to the network. In fact, the individual processes need to be run simultaneously on network-connected computer processors. These applications are frequently I/O bound operations, meaning the time it takes to complete a computation is determined principally by the period spent waiting for input/output operations to be completed. To resolve this issue SDSU faculty worked with a consortium of researchers in Engineering, Computer, and Earth Sciences, and created a two 2.4PB parallel storage cluster using the open-source BeeGFS file system. The high-performance storage system allows relegating computational runtime to reasonable durations facilitating practical computing on shared clusters used by multiple research groups. Runtime performance of I/O bound numerical simulations is substantially improved by using a parallel file system that distributes file data across multiple servers and provides for concurrent access by multiple tasks of a parallel application, this permits multiple, distributed processes to sequentially perform read and write operations.  The parallel storage cluster supports multiple climate science research groups on campus studying subsurface CO2 sequestration, multi-phase turbulent combustion, coastal ocean modeling, and environmental effects of greenhouse gas emissions (Figure 6). Furthermore, the storage cluster has large capacity to store data. It can store real-time carbon flux measurements of atmospheric carbon dioxide and methane concentrations transmitted from an array of fast response (10 Hz) three-dimensional sonic anemometer-thermometer instruments and open path gas analyzers. These devices are installed in towers across the north slope of Alaska along a 300 km transect from Barrow to Ivotuk. The cluster is capable of 6 GiB/s (1GiB ≈ 1.074GB) throughput during read and write operations. This transfer speed improves continuous data acquisition processes and reduces numerical simulation runtime. This high-performance storage cluster has significantly advanced scientific discovery in several areas of climate science research at SDSU.

A picture containing text, electronics

Description automatically generated

Figure . Architecture of *beehive.sdsu.edu*, a 2.4 petabyte (PB) parallel storage cluster that runs the open-source BeeGFS file system for performing parallel I/O during HPC computation.

CC Storage Slide.tif

Figure . The SDSU parallel storage cluster *beehive.sdsu.edu* is used to perform parallel I/O using MPI-IO, parallel NetCDF, and Parallel HDF5, for storing and generating climate science data during distributed-parallel numerical simulations.