**FACILITIES, EQUIPMENT AND OTHER RESOURCES**

**Information Technology Center (ITC)**



The CCIT Information Technology Center (ITC) is the primary data center for Clemson University. The ITC contains 20,000 sq. ft. of raised floor space, with roughly 10,000 sq. ft. allocated for Research Computing and Data Infrastructure (RCDI).

The data center’s electrical system features two 2.5MW transformers fed from redundant power feeds. In the event of a power outage, the facility has UPS and generator backup systems, including two 2.5MW generators with a guaranteed 24-hour supply of diesel fuel on-site. Equipment in the facility is cooled by three 250-ton chillers, which produce 500 tons of cooling capacity. Only two chillers are active simultaneously, leaving the third chiller available as a reserve in case of failure. 275 tons of liquid cooling capacity is provided to RCDI through a tertiary loop, with temperature maintained through building automation and three-way mixing valves.

The ITC is the home of RCDI's hardware resources, including the Palmetto HPC Cluster and the Indigo Data Lake. The ITC is also home to the Network Operations Center (NOC), which has staff on-site 24x7x365 to monitor Clemson’s critical IT infrastructure, including RCDI resources.

**High Performance Computing (HPC) – Palmetto Cluster**

|  |  |
| --- | --- |
| palmetto cluster nodes | Clemson’s high-performance computing resources include a “condominium” style cluster, known as Palmetto, developed to serve the university’s wide-ranging research needs. Designed and deployed by the RCD Infrastructure (RCDI) group in collaboration with faculty researchers across the university, Palmetto provides a shared platform that optimizes resources for the benefit of all users. Named for South Carolina’s state tree, the Palmetto Cluster was designed to suit many different research applications, with many powerful multi-core nodes, each with a significant |

amount of memory. The Palmetto Cluster currently benchmarks at 3.01 PetaFLOPS using 11,280 CPU cores and 2,611,200 accelerator cores, and was ranked #239 in the 61st edition of the Top500 list of HPC clusters. The benchmark utilized 220 compute nodes with either NVIDIA Tesla V100, or NVIDIA Tesla A100 GPU Accelerators, and HDR100 Interconnect. The Palmetto Cluster currently contains over 1800 compute nodes consisting of over 1000 CPU-only nodes, and more than 700 GPU accelerated nodes. GPU accelerated nodes are also interconnected with high speed Infiniband networking. The latest compute nodes contain dual 32-core Intel Ice Lake CPUs, 256GB of memory, dual NVIDIA Tesla A100 GPU Accelerators, and HDR100 100G Infiniband Networking. Palmetto also contains special resources such as large-memory nodes (1TB of memory or more), and NVIDIA DGX systems for large scale AI/ML workloads.

Storage on the Palmetto Cluster consists of approximately 3 PB of large-scale, backed-up project space, and approximately 2 PB of high-speed scratch space provided by a Beegfs parallel filesystem. There is an additional 120 TB scratch space consisting of all NVMe flash storage. As of July 2023, a new 500TB all-flash scratch space is available on Palmetto as part of the Indigo Data Lake, providing up to 5TB of space or 5 million files per user.

The Palmetto Cluster is housed at Clemson’s Information Technology Center (ITC). The ITC is a 24/7 monitored environment with proper power, cooling, and physical security. The Palmetto Cluster is UPS and generator backed to prevent unexpected interruptions to compute jobs.

**High Performance Storage (Indigo)**

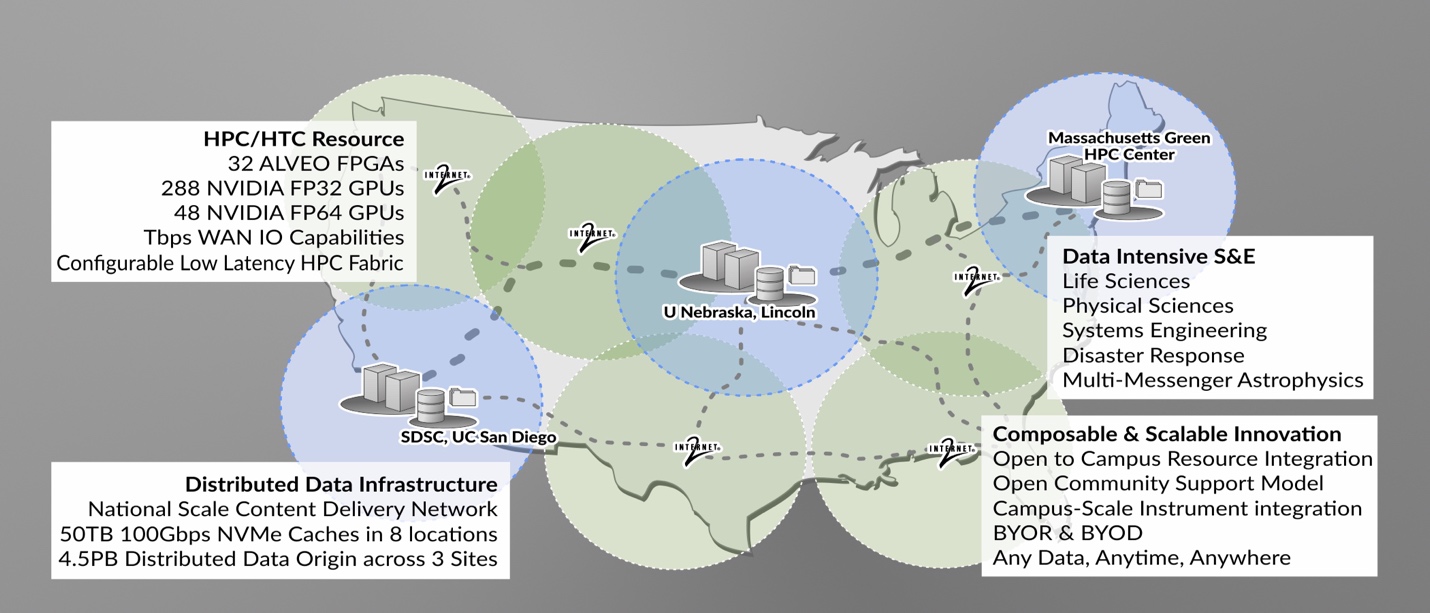
Indigo is a Clemson University research data repository and processing platform. The system is backed by an all-flash scalable storage system provided by VAST Data. Indigo has roughly 3PB of raw storage, but it is anticipated to exceed 5.5PB of data leveraging data reduction and de-duplication technologies. The data lake can support large numbers of simultaneous clients by leveraging an aggregate of 800 Gbps of network throughput over InfiniBand and 400 Gbps over ethernet. Indigo is accessible from all RCD-managed resources via NFS, SMB, or S3 protocols. Students, faculty, and staff can access their storage via the SMB protocol over the campus network. A 500TB partition of the filesystem has been made available as scratch space to all users of the Palmetto Cluster at no charge. Each user is limited to 5TB or 5 million files, whichever comes first. This partition is automatically purged of stale data and is not intended for long-term storage. Faculty can purchase persistent storage on Indigo that also features backups and snapshots. Indigo storage owners will no longer need to copy data to scratch space before processing and can run directly against their partition.

**Open Science Grid**Clemson University maintains a strong relationship with the [Open Science Grid](https://osg-htc.org) (OSG) community, contributing computing power to support OSG and sending out many computing jobs to take advantage of the community resource. Clemson’s CCIT team has dedicated personnel to support faculty and staff interacting with OSG resources.

**Cloud Computing Resources (CloudLab)**

Clemson University is one of the three major sites participating in the NSF-funded CloudLab project ([www.cloudlab.us](https://www.cloudlab.us)). The project enables researchers to provision experimental distributed infrastructure with administrative privilege for over 1,500 computing nodes and computing resources. Clemson University hosted resources for CloudLab include a 316-node compute cluster and a 6-node storage cluster. The CloudLab cluster nodes contain a mix of Intel Xeon, AMD EPYC, and IBM Power PC architectures, and also offer a variety of co-processors such as NVIDIA Tesla GPUs, NVIDIA BlueField-2 SmartNICs, and Xilinx FPGAs. The latest expansion of the CloudLab cluster contains 32 nodes with 36-core Intel Ice Lake CPUs featuring Speed Select and SGX capabilities, and 32 nodes with 32-core AMD EPYC Milan CPUs featuring SEV capabilities. All 64 new nodes contain 256GB of RAM, NVMe local storage, and 100GbE Networking. 4 of the storage nodes contain dual 10-core CPUs, 256GB memory, 8 1TB HDDs, and 12 4TB HDDs, and 2 of the storage nodes contain dual 6-core CPUs, 128GB memory, and 45 6TB HDDs.

**National Research Platform**



The National Research Platform (NRP) ([nationalresearchplatform.org](https://nationalresearchplatform.org)) is a partnership of more than 50 institutions-including Clemson- led by researchers and cyberinfrastructure professionals at UC San Diego, supported in part by awards from the National Science Foundation. NRP is the name of the set of programs, facilities, and policies that are designed for distributed growth and expansion. NRP’s primary computation, storage, and network resource is a ~300 node distributed cluster called Nautilus that hosts many network testing data transfer nodes including ones used by the Open Science Grid (OSG). Nautilus is a powerful nationally distributed computer system with CPUs, GPUs, and FPGAs, in two types of subsystems (“high-performance FP64/FPGA” and “FP32-optimized”), specialized for a wide range of data science, simulations, and machine learning or artificial intelligence, allowing data access through a federated national-scale content delivery network.

**Fabric**

|  |  |
| --- | --- |
| FABRIC (FABRIC is Adaptive ProgrammaBle Research Infrastructure for Computer Science and Science Applications) is an international infrastructure that enables cutting-edge experimentation and research at-scale in the areas of networking, cybersecurity, distributed computing, storage, virtual reality, 5G, machine learning, and science applications. | fabric sponsors |

The FABRIC infrastructure is a distributed set of equipment at commercial collocation spaces, national labs and campuses. Each of the 29 FABRIC sites has large amounts of compute and storage, interconnected by high speed, dedicated optical links. It also connects to specialized testbeds (5G/IoT PAWR, NSF Clouds), the Internet and high-performance computing facilities to create a rich environment for a wide variety of experimental activities.

**Networks**

At the core of Clemson’s local area network are two fully redundant, 100 Gbps-connected Juniper QFX10008's. These have multiple 40 Gbps-connected links to Cisco Nexus 7700's in diverse campus locations. The Nexus switches aggregate dual 10 Gbps connections from Cisco 9300 switch stacks that serve as building network distribution and access switches. The multi-gigabit Cisco 9300s allow end user connections of up to 5 Gbps. This network design has zero single points of failure in the core and distribution layers, is consistent across Clemson’s entire campus, is easy to troubleshoot, and behaves deterministically, should link or equipment failure occur.

The C-Light Network is Clemson University’s upstream connection to the national research community via direct fiber between Clemson, Atlanta, and Charlotte. C-Light connects to Internet2, a national high-speed research and education network in Atlanta and Charlotte, including a dedicated 100Gbps connection to Internet2's Advanced Layer 2 Service (AL2S) network, to provide Clemson University's research community high speed and redundant connections for their research needs. C-Light currently provides over 160Gbps upstream capacity to its membership with geographically redundant connections in Atlanta, Charlotte, Clemson, Anderson, and Columbia. C-Light's network brings Clemson the technological infrastructure that faculty and researchers need to collaborate nationally and internationally with colleagues and access resources, maintaining Clemson University's in the national research conversation.

**CCIT Research Computing and Data (RCD)**

The Research Computing and Data (RCD) group is a centrally funded support organization inside Clemson Computing and Information Technology (CCIT). RCD provides general and advanced research computing support, training, and outreach. RCD faculty and technical staff are a highly successful group of research scientists who lead use of high-performance computing applications, high throughput computing, high-performance networking, data access and interpretation, geospatial data, visualization, artificial intelligence (AI), machine learning, social and biological sciences, humanities computing, and software environments for cyber-communities.

**Clemson Center for Geospatial Technologies**

|  |  |
| --- | --- |
| three dimensional map of Clemson&#39;s campus | The Clemson Center for Geospatial Technologies (CCGT) offers specialized expertise and capabilities in geospatial research and support. It provides comprehensive solutions to meet various GIS-related needs, including technical workshops, project support, interdisciplinary collaboration, and innovative research at Clemson University and throughout the state of South Carolina. |

*Staffing and Engagement:* The center is staffed by a dedicated team of enterprise and desktop GIS experts who provide education and research support to the Clemson community. This team consists of both full-time and part-time staff members. The CCGT team actively supports and trains over 2,000 students, staff, and faculty annually through various academic and research initiatives. They engage in grant collaborations, manage cloud-computing and cyberGIS platforms, and establish partnerships with industry stakeholders to foster innovation and knowledge exchange.

*Software, Computing and Facilities:* CCGT manages ESRI's campus-wide license, providing the research community with access to a suite of geospatial analysis, mapping, and data management technologies. These technologies encompass desktop, enterprise, and cloud-based solutions, available at no cost. The center offers a hybrid cyberinfrastructure that enables researchers to utilize high-throughput computing through the local GIS cluster (GalaxyGIS), virtual machine (VM) technologies for desktop GIS, high-performance computing via the Palmetto cluster, and cloud GIS and data services through ArcGIS Online and locally managed servers. CCGT encompasses a dedicated training facility that serves as a space for GIS training, workshops, seminars, summer school, webinars, class support, and customized training sessions. The training room is equipped with 20 student workstations, each featuring dual monitors. Additionally, the center maintains and supports access to ground and air drone technologies for imagery and LiDAR data collection.

**OTHER RESOURCES: PERSONNEL**

The following are the non-funded project participants: