

## Introduction and Current Efforts

With the increasing level of complexity and demand for IT services, Clemson University has developed a series of robust and innovative techniques to address the growing needs of cyberinfrastructure practitioners both at a national and local level. Clemson has deployed a collaborative campus strategy addressing the rigorous demands of a national research institution. By offering a comprehensive portfolio of cyberinfrastructure services, Clemson has been able to strategically overlay the campus CI plan, synchronizing the multifold mission of Clemson Computing and Information Technology: to provide robust day-to-day services to the campus community while maintaining the high end infrastructure that fosters nationally competitive research and innovation.

Clemson has developed its cyberinfrastructure plan with the following principles:

1. Clemson University is committed to developing and maintaining a world-class campus cyberinfrastructure, seamlessly integrating a variety of disciplines to foster advances in research and education.
2. Clemson University will align its cyberinfrastructure objectives to harmonize the multifold mission of Clemson Computing and Information Technology while also supporting research that synchronizes with statewide and national initiatives.
3. Clemson has and will continue to reverse the de-evolution of the IT organization into an administrative agency, and will capitalize on the far-reaching impacts of campus cyberinfrastructure to re-energize the academic and educational impacts of IT.

Clemson has already taken numerous steps to address the growing demands for campus cyberinfrastructure creating inclusionary and friction free science and has developed partnerships capitalizing on the human capital and intellectual ability available at the University. The University has developed the Clemson University *Center for Next Generation Computing*; The *Center for Next Generation Computing* leverages a number of primary focus areas in which Clemson has developed expertise. The *Center for Next Generation Computing* initiates a “big tent environment”, formally integrating non-IT faculty and researchers with IT operations and development.

Clemson University has provisioned a number of campus cyberinfrastructure resources upon which researchers can conduct science. Primarily, Clemson has deployed a modern enterprise campus network comprised of two fully redundant core switches that aggregate 1 Gb/s or 10 Gb/s connections from each campus building’s dual-switch stacks. This state of the art, fully redundant, and centralized network provides the backbone to the execution of high-level computationally based science. Clemson has enabled IPv6 transport on all wired campus networks, and intends to support it to the extent allowed by the deployed technology on all production and research networks. Robust upstream IPv6 peering is enabled with major ISPs, R&E networks, and other external peers. Clemson is

an early adopter of both Shibboleth and InCommon that allows for statewide, cross-institutional access to IRB materials. Clemson has also deployed the highly sustainable Palmetto Cluster based on the condominium model of resource sharing. The Palmetto Cluster is made available to anyone (student, faculty or staff) at Clemson.

In parallel to the development of the infrastructure initiatives, the development of the human infrastructure capitalizes on the physical investments made. The Clemson Cyberinfrastructure and Technology Integration (CITI) group is a centrally funded support organization providing training to faculty, staff and students in the basic use of computing and data resources, programing compiling, and code parallelization/optimization. This outreach and workforce development is a core component of Clemson University's long-term CI objectives.

## **Moving forward: Clemson University's CI Roadmap**

While Clemson develops scalable and resilient campus cyberinfrastructure, continued efforts in innovation and research remain. To execute on this mission, Clemson has identified a number of focus areas for its cyberinfrastructure roadmap.

### **High Performance Computing**

The Advanced Computing Infrastructure team maintains Clemson's cluster known as the Palmetto Cluster. Palmetto is currently utilized by 31 out of 52 academic departments and is based on the "condominium" model, which is a community-owned model leveraging shared resources (with faculty ownership totaling about 50% of nodes) with university investments to maximize cluster resources.

Palmetto currently contains around 2,000 compute nodes with 496 TFlops of compute power. Palmetto will continue to leverage the presence of dual NVidia GPU technology, and will continue to expand at a rate of 75-100 nodes per fiscal year, representing a roughly 150-200 TFlop increase in computer power. Palmetto has been principally involved in \$83.1MM grants and awards since FY10, with \$34.1MM of that occurring in FY13.

### **Identity and Access Management**

Clemson's identity management systems are tightly integrated. Clemson integrates systems from the OS level to the enterprise application level (which include student learning systems, HR systems etc.), leveraging Shibboleth and InCommon as our primary authentication solutions. CAS (Central Authentication Service) is also utilized in situations that require it. Clemson has developed a collaborative provisioning system known as "Central.Clemson", enabling self-service provisioning of group resources to file systems, mailing lists, and other applications (which include Palmetto Cluster Services). Clemson has participated in the development (and subsequent deployment) of Junetsu, which Clemson utilizes as the primary network registration system. There are plans for the University to integrate the identity management systems into a Cloud Desktop; a fully self-service VDI solution based on OpenStack.

### **Advanced Networking**

Clemson's campus network is a state of the art modern enterprise network with a centralized and unified architecture, redundancy at all levels that eliminates single points of failure and with a professional 24/7/365 network operations center (NOC) to monitor activity. Clemson has deployed the C-Light regional network that serves as Clemson's connection to the research community VIA direct fiber between Clemson, Atlanta and Charlotte, providing access to high speed national and international research networks such as Internet2. Clemson connects to the Internet2 Innovation platform, providing 100 Gb/s access to the campus community through a science DMZ. The utilization of this DMZ allows Clemson to carve out spaces in the existing Palmetto cluster and connect chunks of compute nodes directly to users and instead of through proxy servers. Clemson is a charter GENI OpenFlow site, and a number of staff members have served as mentors to other national universities as part of NSF's GENI-enabling initiative.

Clemson (as part of NSF award #1244936 "CC-NIE Integration: Clemson-NextNet) is enhancing current cyberinfrastructure by extending our existing SDN connections to the I2 Innovation Platform, to 20-campus building at 10 Gb/s per drop and 40 Gb/s per building. Clemson's use of SDN will provide frictionless end-to-end network connections and will create SDN circuits to connect subsets of the Palmetto Cluster to campus locates and nationally via the I2 innovation platform. Additionally, Clemson plans to integrate the InCommon Identity Manager directly to the OpenFlow controller to manage network access and maintain a secure, fast and efficient network.

### **Visualization**

Visualization plays a significant role in the exploration and understanding of data across all disciplines with a universal goal: gaining insight into the complex relationships that exists within the data.

#### **Goals**

- Facilitate student involvement and utilization of Cyberinfrastructure resources at Clemson through visualization initiatives
- Foster collaborations with Clemson Faculty and other universities
- Enhance K-16 literacy and engagement with science and technology through visualization

#### **Objectives**

- Mentor undergraduate students in visualization as part of the University Professional Internship and Co-Op program at Clemson University as part of Clemson's 2020 Mission
- Recruit and mentor women and members of underrepresented groups to participate in visualization initiatives like the REU Site in visualization
- Provide training in tools and technologies that are common in visualization

- Acquire funding to build a community of educators who work collectively to incorporate visualization principles and techniques into STEM education (K-16)

### **Data Science**

In November 2013, Clemson deployed a dedicated Hadoop Cluster to assist researchers with their “big data” needs. To coordinate, Clemson initiated bi-weekly training meetings where students, faculty and staff engaged an open forum discussion. This is in addition to the new user introduction sessions conducted by our data science team. Clemson is upgrading the cluster to Hadoop 2.0, providing easier access to multiple file systems across campus, better scheduling of resources and increased flexibility in designing pipelines. The cluster is built using large memory nodes and will be coupled with the deployment of the Berkeley Data Analytics Stack, allowing us to utilize these nodes for in-memory data analysis. This provides an easy-to-access environment for new users while providing optimization tools for advanced users.

### **Outreach and Workforce Development**

Clemson’s Cyberinfrastructure and Technology Integration group (CITI) fosters high-quality educational and training support for use and integration of advanced computing infrastructure. This group works closely with faculty, researchers, students, partners and the Clemson community, supporting high-quality curriculum and professional development for all disciplines, while sharing the resources of physical resources of Clemson Computing and Information Technology and the human capital of CITI.

#### **Goals:**

- Identify and improve HPC education initiatives across the country, participating with partners to build a network of shared ideas, experiences, and best practices for the betterment of the entire educational system (K-20)
- Provide and facilitate an innovative knowledge network of educational assets to continuously improve HPC curriculum, instruction, assessment, professional development, leadership, and community engagement.
- Identify and evaluate funding opportunities and actively participate in development campaigns
- Organize and/or sponsor activities that support HPC at Clemson University
- Inspire the next generation of innovators and creators through K-16 STEM initiatives

#### **Objectives**

- Guide Clemson University and partner HPC users to content experts for mentoring support
- Create a network/repository of tutorials and training opportunities

- Communicate regularly with HPC community about events, training, case studies, etc
- Broaden participation in HPC across Clemson University through interdisciplinary projects
- Create/sponsor/support annual Clemson University HPC events
  - STEM K-16
  - Speaker series
  - Technology conference