# UAB IT Research Computing (UABRC) Resources and Cybersecurity Facilities Document

## Introduction

UAB IT Research Computing (UABRC) maintains the Research Computing System (RCS), an integrated computing platform that provides access to enhanced compute, storage, and network capacity for UAB investigators and their collaborators. The RCS compute resources include a high-performance compute (HPC) cluster for large scale modeling and analysis workloads, an on-site cloud platform for highly customizable virtual machine (VM) based workloads, and a container orchestration platform for cloud-native workloads. The RCS storage resources include a high-speed GPFS file system attached to the HPC cluster to support data analysis workloads and large-capacity block and object storage to support the cloud and container workloads. The RCS networking infrastructure connects the computing and storage systems via 800 Gb/s Ethernet interconnect that provides ample capacity for data access and movement between the compute and storage resources. RCS networking also includes connectivity to the UAB Campus Network, a dedicated EDR/HDR Infiniband fabric for low-latency data exchange on the HPC cluster, and a 40 Gb/s ScienceDMZ for high-speed data transfers with national research and education networks.

These RCS resources combine to provide a low-friction application hosting platform that enables research teams to build and deploy preferred tools without enforced refactoring to adopt applications to campus resources. The RCS resources are deployed spanning two data centers. The on-campus data center is in the recently constructed Technology Innovation Center (TIC). The off-campus data center is located at a nearby (less than 10 km) commercial facility operated by DC BLOX, a regional data center provider offering a Tier III colocation facility in Birmingham with adequate power to support the high-density power requirements for expanding the compute capacity of the RCS and a resilient physical infrastructure designed to withstand natural disasters like tornados, which are common in the region. The commercial facility is connected to the campus data center via a dedicated, diverse fiber path lit with dual 100 Gb/s optics that leverages the University of Alabama System Regional Optical Network (UASRON). UABRC designs and maintains the RCS resources in open collaboration with the campus research community to ensure that the system addresses research needs and has the capacity to meet research demand.

## High-performance Computing Resources (Cheaha)

Cheaha is a campus high-performance computing (HPC) resource dedicated to enhancing research computing productivity at UAB. Cheaha is managed by UAB IT Research Computing (UABRC) and is available to members of the UAB community in need of increased computational capacity. Cheaha supports both high-performance computing (HPC) and high-throughput computing (HTC) paradigms. Cheaha provides 10752 CPU cores, 112 GPUs, and 88 TB of memory across 159 compute nodes interconnected via an EDR/HDR InfiniBand network, providing over 1.1 PFLOP/s of aggregate theoretical peak performance. A high-performance parallel filesystem running is connected to these HPC compute nodes via the InfiniBand fabric. Node details follow.

* Gen 7 (2017): 18 nodes, 2x14 core (504 cores total) 2.4 GHz Intel Xeon E5-2680 v4 compute nodes with 256 GB RAM, four NVIDIA Tesla P100 16 GB GPUs per node, and an EDR InfiniBand interconnect.
* Gen 8 (2019): 35 nodes, 2x12 core (840 cores total) 2.60 GHz Intel Xeon Gold 6126 compute nodes with 21 compute nodes at 192 GB RAM, 10 nodes at 768 GB RAM and 4 nodes at 1.5 TB RAM, and an EDR InfiniBand interconnect.
* Gen 9 (Q2 2021): 52 nodes, 2x24 core (2496 cores total) 3.0 GHz Intel Xeon Gold 6248R compute nodes each with 578 GB RAM and EDR InfiniBand interconnect.
* Gen 10 (Q4 2021): 34 nodes, 2x64 core (4352 cores total) 2.0 GHz AMD Epyc 7713 Milan compute nodes each with 512 GB RAM and EDR InfiniBand interconnect.
* Gen 11 (Q4 2023): 20 nodes, 2x64 core (2560 cores total) 2.0 GHz AMD Epyc 7713 Milan compute nodes each with 512 GB RAM and 2 A100 80 GB GPUs.

Cheaha provides researchers with both a web-based interface, via open OnDemand, and a traditional command-line interactive environment, via SSH. These interfaces provide access to many scientific tools that can leverage a dedicated pool of on-premises compute resources via the SLURM batch scheduler. The on-premises compute pool provides access to four recent generations of hardware based on the x86 64-bit architecture.

## High-performance Computing Software Tools (Cheaha)

General research computing and scientific programming software are available on Cheaha, including Anaconda, R and RStudio, and MATLAB through the Lmod environment module system. RStudio, MATLAB, Jupyter Notebook server, and Jupyter Lab are all available on our Open OnDemand web portal as interactive applications, along with a general-use desktop environment via no-VNC, directly in the browser. Researchers are enabled to develop and share their own custom interactive applications through a sandbox application feature within Open OnDemand.

The UAB Center for Clinical and Translational Science (CCTS) Informatics group has installed and supports a variety of bioinformatics tools that are available to be run from Cheaha. Standalone packages are available for quality control (fastQC, Picard Tools), alignment (Abyss, Velvet, BWA, Bowtie) visualization (IGV), variant calling (GATK, SnpEff, annoVar), RNAseq (Cufflinks, Cuffdiff, TopHat) and microbiome and metagenomic analysis (QIIME, HUMAnN, MEGAN).

Additional scientific domain-specific software is also available, including Relion for cryo-electron microscopy analysis, AFNI for fMRI analysis, and ANSYS for simulations for research efforts of the UAB School of Engineering. Many other software packages are installed and maintained, and we encourage and facilitate researchers installing their own additional software using Anaconda, R and MATLAB package management where possible.

## On-premises Cloud Resources (Cloud.rc)

UABRC operates a production private cloud called cloud.rc based on OpenStack, which echoes the research workload support goals of the NSF’s Jetstream2 resource part of the ACCESS network. The Cloud.RC platform has been used to support application development and DevOps processes to research labs across campus and is increasingly being leveraged to support many aspects of research IT operations. This fabric is composed of five Dell R640 48 core 192 GB RAM compute nodes for 240 cores and 960 GB of standard cloud compute resources. In addition, the fabric features four NVIDIA DGX A100 nodes that include 8 A100 GPUs and 1 TB of RAM each. These resources are available to the research community for provisioning on demand via the OpenStack services (Ussuri release). The production implementation further supports researchers by making their hosted services available beyond campus, while adhering to standard UAB Campus Network security practices. Scientific software developers have access to the full stack for limitless development opportunities, with a frictionless migration path to public cloud providers as needed for specific research projects. A Kubernetes environment was deployed in Q3 2022 to allow for development workflows using containers. The compute resources of the Kubernetes environment are a duplicate of the cloud resources. The OpenStack and Kubernetes resources are deployed via Canonical’s Charmed operations stack enabling node migration between platforms in response to capacity demand.

## Storage Resources

The compute nodes on Cheaha are backed by high-performance, 7 PB GPFS storage on DDN SFA14KX hardware connected via an EDR /FDR InfiniBand fabric. Two additional storage systems are available to support research operations and application design. They are based on the Ceph storage technology and provide different hardware configurations to address different usage scenarios. The fabrics include a 6.9 PB archive storage fabric for long term storage (LTS) built using 12 Dell DSS7500 nodes, and an 11 PB nearline storage fabric built with 14 Dell R740xd2 nodes and 248 TB SSD cache storage fabric (Q3 2023) built with 8 Dell 840 nodes. These storage systems provide block and object storage services to the OpenStack and Kubernetes platforms. Additionally, the object storage services are empowering research applications with cloud-native data management and availability capabilities.

## Network Resources

The RCS networking infrastructure connects the on- and off-campus computing and storage systems via 800 Gb/s Ethernet interconnect that provides capacity for data access and movement between the compute and storage resources. RCS networking also includes a dedicated EDR/HDR Infiniband fabric for the HPC platform and a 40 Gb/s ScienceDMZ for high-speed data transfers with national research and education networks. The ScienceDMZ supports direct connection to campus and high-bandwidth regional networks via 40 Gb/s Globus Data Transfer Nodes (DTNs) providing the capability to connect data intensive research facilities directly with the high-performance computing and storage services of the RCS. This network can support very high-speed secure connectivity between nodes connected to it for high-speed file transfer of very large data sets without the concerns of interfering with other traffic on the campus backbone, ensuring predictable latencies. The Science DMZ interface with (DTNs) includes Perfsonar measurement nodes and a Bro security node connected directly to the border router that provide a “friction-free” pathway to access external data repositories as well as computational resources.

The UAB Campus Network backbone is based on a 40 Gb/s redundant Ethernet network with 480 Gb/s back-planes on the core L2/L3 Switch/Routers. For efficient management, a collapsed backbone design is used. Each campus building is connected using 10 Gb/s ethernet links over single mode optical fiber. Desktops are connected at 1 Gb/s speed. The campus wireless network blankets classrooms, common areas and most academic office buildings. UAB connects to the Internet2 high-speed research network via the University of Alabama System Regional Optical Network (UASRON), a University of Alabama System owned and operated DWDM Network offering 100 Gb/s ethernet to the Southern Light Rail (SLR)/Southern Crossroads (SoX) in Atlanta, Ga. The UASRON also connects UAB to UA, and UAH, the other two University of Alabama System institutions, and the Alabama Supercomputer Center. UAB is also connected to other universities and schools through Alabama Research and Education Network (AREN).

## Data Management and Transfer Resources

A traditional POSIX filesystem is available on all Cheaha HPC nodes through GPFS for data requiring computational analysis, with separate data, scratch, and shared storage. Object storage is available via our Long-Term Storage (LTS) S3 interface. A REST endpoint is provided for LTS and exposed to the Internet to facilitate hosting research data products for external use. Block storage is available to support storage needs for our cloud and Kubernetes fabrics.

All faculty, staff and students who create a Research Computing account have immediate access to 5 TB of personal GPFS storage and may request an additional 5 TB of LTS storage. Research PI groups, Core facilities, and other research groups at UAB may request up to 25 TB of GPFS storage and 75 TB of LTS storage for shared collaboration spaces.

Globus High-Assurance (HA) endpoints are maintained on the RCS platform to facilitate internal and external data transfers. Connectors to our enterprise Box.com instance and our LTS S3 interface are made available as part of our Globus subscription. A controlled-access Science DMZ partition of our hardware is available to facilitate high-throughput, parallel batch data transfers over the available 40 Gb/s connection to the external internet. Standard data transfer software such as Rclone, AWSCLI and s3cmd, and UAB-specific documentation and support, are provided to researchers to facilitate data transfers whenever Globus is infeasible.

## Facilitation, Outreach, Documentation, and Support

UABRC provides research computation facilitation services for researchers using RCS. These services exist to reduce friction for researchers seeking to scale workflows from desktop and workstation scale up to HPC scale. Part of facilitation serves include computational outreach efforts within UAB, including facilitating lesson design for courses making use of our platform, teaching a Data Science Journal Club course, providing how-to-use-HPC lessons at University events, and proactively identifying opportunities for education and efficiency improvements using our internal observability stack.

Extensive documentation of computational capabilities, good practices for system use, references and tutorials are all available on our documentation website, publicly available on the Internet. Links to common, standard external educational resources are provided and encouraged where applicable, including to, the Software Carpentries and Data Carpentries lessons, and GitHub and GitLab version control documentation.

We provide coverage for support requests during standard business hours, and greater coverage for outages and security incidents. Tickets are tracked using ServiceNow software, and most requests are addressed within 1-2 business days, with faster responses for critical incidents. Support requests covered include software installation and update requests, new researcher training, hardware and software errors, data transfer and shared storage requests, and facilitation of collaborative, grant-funded research computation projects ranging from individuals, through labs, to core facilities.

## Research Computing Personnel

UAB IT Research Computing currently maintains a support staff of 14 led by the Assistant Vice President for Research Computing and includes one HPC Architect-Manager, one Research Facilitation and Data Manager, four Software developers, four Research Facilitation Scientists, and three system administrators.

## UAB Cybersecurity Policies and Practices

UAB IT maintains a comprehensive privacy and information security program that preserves and protects the confidentiality, availability and integrity of all information assets including patients, research, customer and business data. The security program upholds values and provides high standards of service, trust, confidentiality and responsiveness. The security program include the following:

* IT security policies designed to help ensure a secure state of operations and information management.
* Technical security standards that document baseline security requirements for key technologies and platforms such as major operating systems, databases, network device operating systems, firewalls, web-server security, email, encryption, secure file transfer protocols, virus defense, media reuse and media disposal.
* A comprehensive risk management program based on the NIST Cybersecurity Framework (CSF).
* A data classification rule to assist the UAB community in the classification of data and systems to determine the appropriate level of security. All UAB data stored, processed, or transmitted must be classified in accordance with this rule. Based on classification; users are required to implement appropriate security controls.
* A data protection rule to assist the UAB research community in the protections requirements of data and systems. All UAB research data stored, processed, or transmitted must be protected in accordance with these standards. Based on the regulatory or contractual requirements of the award; researchers are required to implement appropriate security controls.
* A computer security incident response plan that is supported by cross-functional response and recovery teams.
* User system access is tightly controlled and meets standards required by various regulations (e.g. HIPAA, FERPA, etc.). Two-factor authentication is utilized for many of the shared systems. Users must agree to the requirements set forth in the password rule. We also must routinely demonstrate compliance with Federal granting agencies and the corresponding security requirements such as the NIH, FISMA and the VA.
* An enterprise firewall platform for perimeter, datacenter and customer layered protection and segmentation. The firewall also consists of an Intrusion Prevention System (IPS) as well as content filtering and malware sandboxing.
* An enterprise centralized logging solution (aka SIEM tool) to store logs and generate alerts to the Security Operations team for items that might need attention.
* In addition to the Security Operations team (SOC) who handles incident response, forensics, EDR management, firewall management, etc., we also have a Threat and Vulnerability Team (TVT) that handles penetration testing, vulnerability scanning, phishing simulations and threat hunting. We also have a Risk Management team that handles application/business unit risk assessments, 3rd party risk assessments, MFA administration, Security Awareness and Training. We also have a Security Architect who handles architectural discussions for all sorts of projects across the campus.
* UAB maintains an IoT policy to ensure all IoT devices that reside on the UAB network (wired or wireless) must be proactively managed and adhere to the university’s IoT security strategy.
* Email is monitored with the highest level of defense offered from Microsoft with their XDR platform. This includes safe attachments, impersonation and other anti-phishing methods. UAB also utilizes DMARC, DKIM and SPF for additional email best practices. We also maintain email guidelines for the UAB community to adhere to.
* Encrypted VPN tunnels for business associates, staff remote access, or partner VPN connectivity.
* Capability to support encrypted secure file transfers with Globus for HPC and other cloud-based solutions for traditional cloud-based storage needs.
* Endpoint Detection / Response (EDR) agents and comprehensive patch management programs installed on all UAB managed computer workstations and servers to protect against malware infections. The EDR logs are monitored 24x7x365 by a third party to give full oversight and attention to needed items on the systems.
* A vulnerability management rule describes the process used by University of Alabama at Birmingham Information Technology (UAB IT) in mitigating the risks from computer security vulnerabilities.
* In-depth security awareness training that is provided for all Faculty, Staff.
* UAB also maintains an IT Governance program that will, in consultation with strategic campus partners, provide a review that evaluates the risk associated with the above items. This review is also intended to prevent duplication of technology while complying with security and regulatory requirements.