

Computing at the cyclotron institute (CI) for 2023-2024

R. Burch, J. Gauthier, K. Hagel and Y.-W. Lui

Over the past year, the Cyclotron Institute's Scientific Compute Group (CI-SCG) has been dedicated to improving the institute's compute capacity. We've also been testing infrastructure to enable High Availability (HA) services for critical operations.

To do this, we added compute nodes and infrastructure to grow our compute capabilities and enhance our infrastructure for critical services.

Computers

The High Performance Research Computing (HPRC) group at Texas A&M University generously donated three full racks of computers to the Cyclotron Institute. These servers were previously part of the decommissioned HPRC ADA cluster. The CI-SCG expanded the institute's compute capacity by repurposing these racks. This expansion involves three racks, each containing six crates, with each crate housing twelve compute blades. Our existing compute capacity supports approximately 1200 job slots. The additional servers could provide an additional 5800 job slots if all servers are operational, representing a fivefold increase. The CI-SCG team is diligently integrating these servers into our infrastructure.

For the operating system installation, we deployed the Rocky Linux 8.x Operating System (OS) on all active compute blades. The installation process utilized Zero-Touch Provisioning, specifically leveraging the Preboot Execution Environment (PXE) Boot and our Infrastructure as Code (IaC) approach. Despite power and cooling constraints, we installed the OS crate by crate, enabling progress even before full power and cooling availability. We seamlessly integrated these servers into our metrics and observability system to monitor their performance and health, ensuring efficient operation. As power and cooling became available, we moved the three racks into the server room. Currently, twelve out of eighteen crates from these racks are in production, yielding 4024 additional job slots, resulting in a total of 5224 slots — a fourfold increase in compute capacity.

High Availability Services Environment

Our initiative to develop infrastructure for HA services is still in the testing phase. HA services are crucial for maintaining an acceptable level of service, even in the presence of faults and operational challenges. As our primary objective, we aim to minimize service interruptions and enhance reliability.

The implementation involves creating dedicated environments for development, testing, production, and backup. For our HA services, we've set up a three-manager-node, three-worker-node environment using FlatCar Linux for the OS and Docker Swarm for container management and operations. This setup ensures continuous availability of critical services. By leveraging Flatcar Linux and Docker Swarm, we establish a secure, scalable, and efficient infrastructure. This platform supports both development and production activities side-by-side and allows rigorous testing. Our efforts enhance

operational efficiency and anticipate a robust foundation for scaling and accommodating new services. Regular backup testing strengthens disaster recovery capabilities and contributes to overall business resilience.

To date, we have successfully deployed our Signage system in production and development modes on our test infrastructure. The Signage system plays a crucial role in communicating important information to our users. Additionally, we have deployed our LDAP service in this test infrastructure. LDAP provides centralized authentication and directory services, ensuring secure access to institute's resources and is critical to the usage of its compute infrastructure.

Cyclotron Institute Server computer Updates

We coded a graphical user interface application to more efficiently manage the Cyclotron computer user accounts. This GUI is written in Python, can create and modify LDAP entries, add user home directories and create logical volumes on the file server. The code can also pull information about existing users and logical volumes and allows to extend the volumes if users need more space. The application has been called CycAG (Cyclotron Account Generator). A second and simpler GUI has also been built, but unlike CycAG, it can't do any modifications and can only be used to search and display user information from the LDAP server. We named this second application CLS (Cyclotron LDAP Search).

Since CAS (Central Authentication Service) is now considered a legacy [1] protocol by TAMU Technology Services, we took the decision to set up the Cyclotron Institute web server with one of the more modern authentication protocols that are currently used by TAMU. We are now using OpenIDC [2] (also known as OpenID Connect) through Microsoft Azure to authenticate our users on most of our restricted webpages. We expect the CAS replacement to be fully completed on our web server by the end of 2024.

[1] Legacy: outdated computing system, hardware or software that is still in use.

[2] OpenIDC: it is an authentication standard built on top of OAuth 2.0.