CHESS Facility and Cyberinfrastructure Overview

Devin Bougie and Werner M. Sun
Cornell University, Ithaca, New York 14853, USA

June 19, 2017

Abstract

We present an overview of the Cornell High Energy Synchrotron Source (CHESS), focusing on the facility and key cyberinfrastructure components.

1 CHESS Facility

The Cornell High Energy Synchrotron Source (CHESS) is a NSF-funded National User Facility located on the Cornell University campus in Ithaca, New York. The mission of CHESS is to provide a national hard x-ray synchrotron radiation facility for individual investigators, on a competitive, peer reviewed, proposal basis. With 11 experimental stations, the facility is used by approximately 1,100 investigators per year from over 150 academic, industrial, government, non-profit, and international institutions. CHESS impacts a wide range of disciplines, serving researchers from the physical, biological, engineering, and life sciences, as well as cultural specialists such as anthropologists and art historians. CHESS users conduct studies encompassing, but not limited to, the atomic and nanoscale structure, properties, operando, and time-resolved behavior of electronic, structural, polymeric and biological materials, protein and virus crystallography, environmental science, radiography of solids and fluids, and micro-elemental analysis, and other technologies for x-ray science.

The CHESS facility is hosted by the Cornell Laboratory for Accelerator-based Sciences and Education (CLASSE), which also operates the Cornell Electron Storage Ring (CESR) as the x-ray source for CHESS. Computing services for CHESS are provided centrally by the CLASSE-IT department. The primary computing services used by CHESS are:

- high-speed data acquisition for x-ray detectors at the CHESS experimental stations
- access to and long-term storage of x-ray data collected by CHESS users
- software libraries and parallel computation resources for CHESS staff and users.

More information about the CHESS facility may be found at http://www.chess.cornell.edu.

2 CHESS Cyberinfrastructure

The CLASSE cyberinfrastructure (CI) consists of an interconnected series of high-availability server clusters (HACs), data acquisition systems, control systems, compute farms, and workstations. Most of these systems run either Scientific Linux or Windows on commodity 64-bit Intel-based hardware and are centrally managed using Puppet. The median age of key CI components is approximately 5 years, with an average refresh rate of once every 10 years. The CLASSE CI components most relevant to CHESS are described below and are shown in Figure 1.

2.1 Central Infrastructure

The central Linux infrastructure cluster runs the core CLASSE infrastructure services, including name services, file systems, databases, and web services. Recently, a dedicated oVIrt cluster has been commissioned to run centrally-provisioned virtual machines. These clusters utilize shared 10Gb iSCSI storage domains, and they provide file systems and other basic services to the rest of the lab.

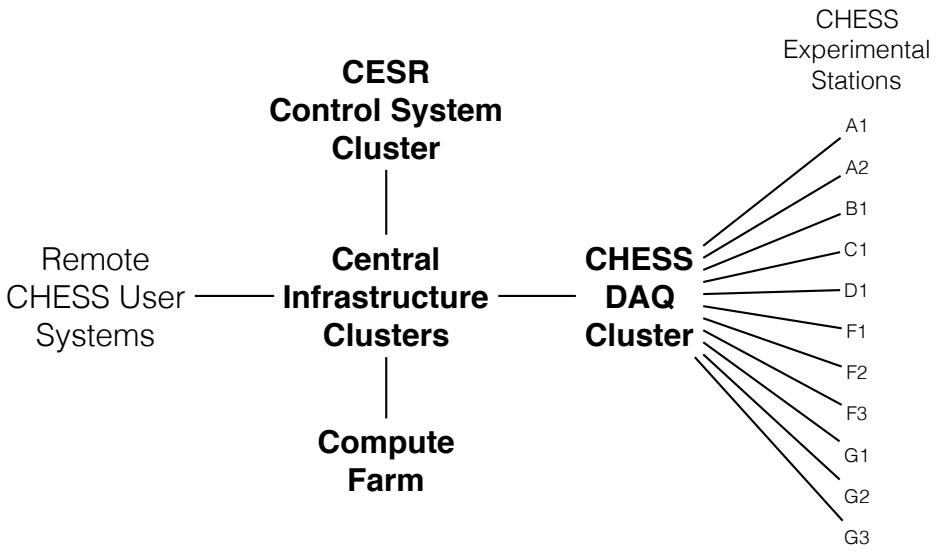

Figure 1: Main components of the CHESS cyberinfrastructure.

2.2 CHESS Data Acquisition (DAQ)

The CHESS data acquisition system runs on a dedicated HAC and provides 10Gb network connections to each experimental station. Data collected at the stations are written directly to the data acquisition system over either NFS or Samba, where it can then be processed on the CLASSE Compute Farm or end-user workstations. CHESS users can also download their data remotely using a Globus server endpoint or via SFTP.

2.3 Compute Farm

The CLASSE Compute Farm is a central resource consisting of approximately 60 enterprise-class Linux nodes (with around 400 cores) with a front-end queueing system that distributes jobs across the Compute Farm nodes. This queueing system supports interactive, batch, parallel, and GPU jobs, and it ensures equal access to the Compute Farm for all users.

2.4 CESR Control System

The CESR control system, responsible for running the particle accelerator that produces x-rays for CHESS, consists of a dedicated Linux HAC. Although the CESR, CLASSE, and CHESS DAQ clusters are essentially identical, the CESR cluster runs many more control system services and is able to operate independently from the CLASSE central infrastructure. This isolation ensures continuity of CESR operations in the event of a power failure or general network outage.

2.5 User Connectivity

Based on their requirements, CHESS users are either granted restricted "external" CLASSE accounts (providing access to station computers and remote access to data) or full CLASSE accounts (providing access to the CLASSE Compute Farm and full interactive desktops, both local and remote).

While collecting data at the experimental stations, CHESS users generally connect their instruments and experimental equipment to a private subnet that is selectively firewalled from the rest of the CLASSE infrastructure. If users require direct write access to the CHESS DAQ filesystems, they may use dedicated station and kiosk computers located at the experimental stations and in other restricted-access locations. Outside the experimental stations, CHESS user data is made available for read-only access through the CLASSE public network.