CHESS DAQ* Introduction

Werner Sun (for the CLASSE IT group), Cornell University

23 July 2015, "The More You Know" Seminar Series

* DAQ = data acquisition
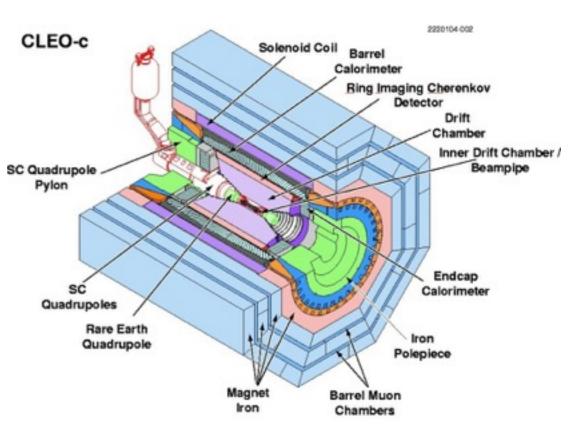**https://en.wikipedia.org/wiki/Data_acquisition**

- Historically, low data volumes:
  - Less than 1 GB per user group
  - Different storage solutions for each staff scientist
- But things are changing:
  - Higher flux with undulators
  - More advanced area detectors
  - Maia detector
  - Time-resolved experiments
  - Up to 400 MB/s (burst), 1 TB/week
  - NSF Data Sharing Policy
- Requires paradigm shift:
  - Central data storage
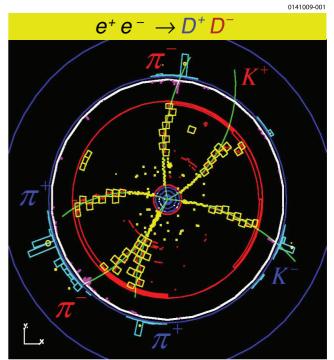 - On-site data reduction/analysis
  - Not unique to CHESS

A LIGHT FOR SCIENCE

ESRFnews

Number 65 December 2013

Big data in focus

New lattice shapes light-source scene
Superconductor predicted from first principles

ESRF

- CLEO experiment, 1979-2008
  - Collected billions of e^+e^- collision events
  - ~200 people, ~20 institutions, 500+ publications
  - Types of data:
    - Raw data: ~20 kB/event
    - Centrally reduced data: < 10 kB/event
    - Simulated data (typically 10-20x real data)
    - Personal analysis data
  - Total 2000-2008:
    - Raw data 80+ TB
    - Reduced + simulated data ~40 TB
  - Data analysis software developed by CLEO collaborators, analysis jobs run at Cornell
- CesrTA follows similar model to CLEO

$e^+e^- \rightarrow D^+D^-$

Maia Detector

- Key driver of growth in CHESS data volume

- Designed by particle physicists at CSIRO and BNL

  - Non-trivial data acquisition pipeline, tightly integrated with CLASSE computing

  - Binary logger daemon (blogd) runs on the CLASSE central infrastructure.

    - Receives data from detector (HYMOD)

    - Writes data to disk (CHESS DAQ filesystem)

  - Data processed with GeoPIXE on the CLASSE Compute Farm

    - Compute Farm has direct access to data (no file transfers necessary)

- Average ~500 GB / user group / run
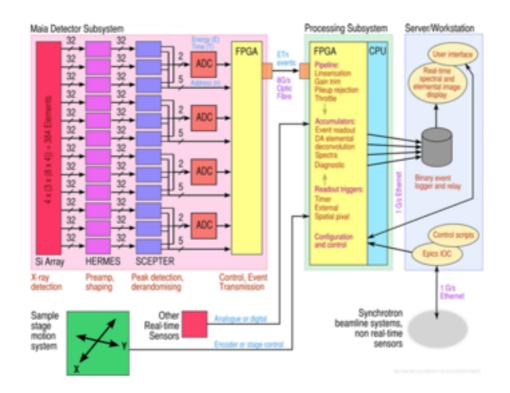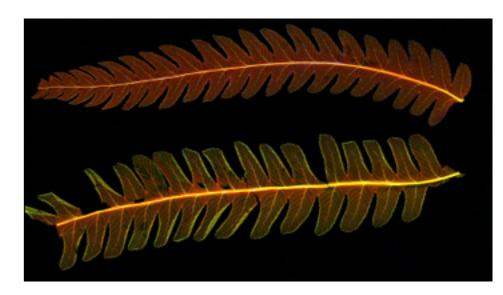
  - 8+ TB collected since early 2014

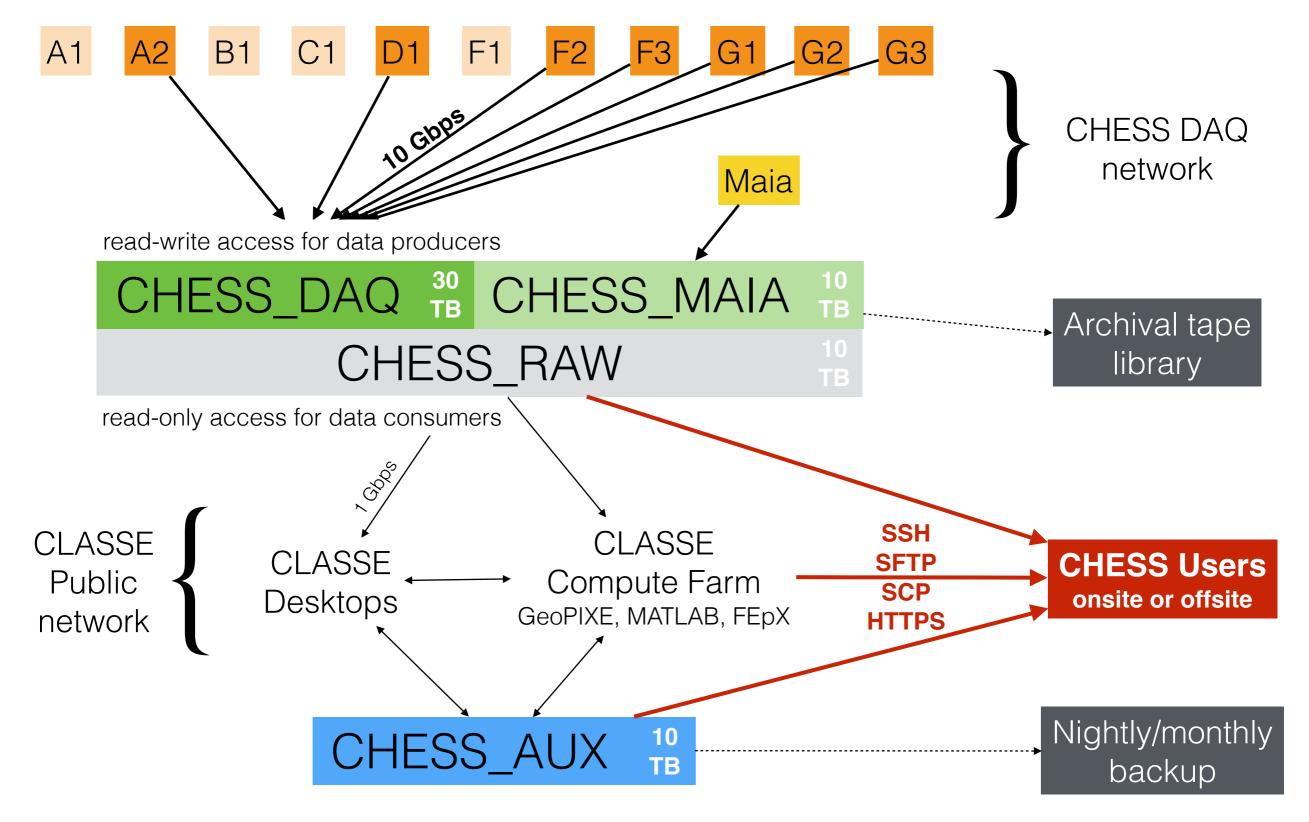Figure 1. Functional block diagram of the Maia system.

CHESS DAQ Basics

- Unlike CLEO, many data sources at CHESS, many types of user groups
- All data producers (eventually) write to two dedicated filesystems:
  - **CHESS_DAQ**: 30 TB
  - **CHESS_MAIA**: 10 TB
  - [**CHESS_RAW**: 10 TB, read-only interface, restore point for archived data]
- Additional filesystems for previous run(s)
  - **PREVIOUSDAQ** (30 TB) and **PREVIOUSMAIA** (10 TB)

- Auxiliary (meta)data, processed data, user data
  - **CHESS_AUX**: 10 TB

- Software packages for CHESS
  - **CHESS_OPT**: 500 GB
  - Anaconda, WIEN2k, GeoPIXE, tomopy, RAW (BioSAXS)

See https://wiki.classe.cornell.edu/CHESS/DataStorageManagement

Data Archival and Rotation

- At least two runs on disk at all times (current + previous)

- Archival and backup handled by Symantec NetBackup
  - Can request files to be restored from tape
    - Complies with NSF open access mandate and CHESS Data Management Plan
  - Data can be kept on disk upon request

- CHESS_DAQ / CHESS_MAIA:
  - After each run, two copies of data written to tape
    - One copy stored off-site (disaster recovery copy)
  - Data is archived indefinitely
  - Before next run:
    - Move data to PREVIOUSDAQ / PREVIOUSMAIA, make read-only
    - Prepare CHESS_DAQ / CHESS_MAIA directories for next run, update softlinks

- CHESS_AUX and CHESS_OPT:
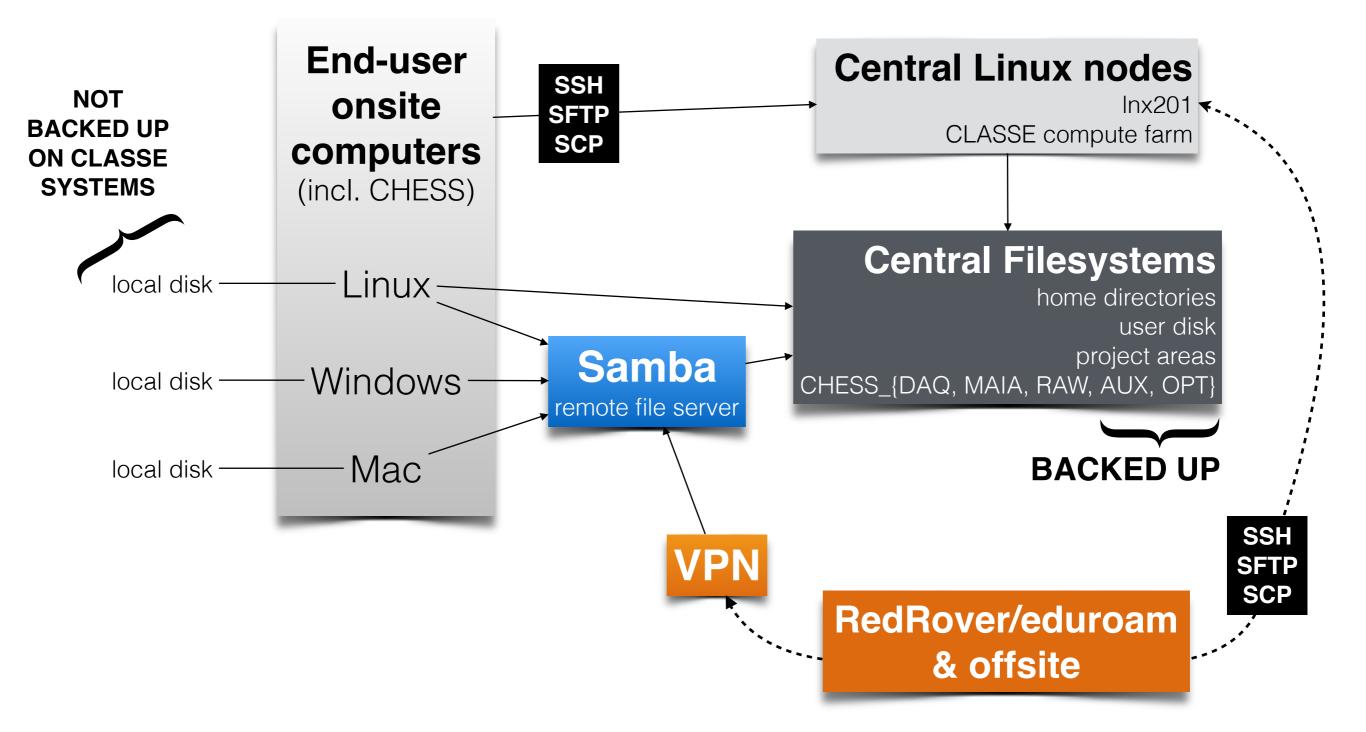 - Monthly full + nightly incremental backups

- Permissions
  - CHESS_DAQ / CHESS_MAIA:
    - Network-based permissions
    - Can be written to **only** from CHESS DAQ subnet
    - Analysis systems are on CLASSE Public network, read-only access
  - CHESS_AUX and CHESS_OPT:
    - Group-based permissions
    - Open to anyone in the "chess" security group
      - Currently includes staff, students, postdocs, users

- For CHESS users: data transfer kiosks at CHESS ops and reception
  - Can be unlocked by anyone with a CLASSE account

- Some station/analysis computers are now CLASSE-managed

- Centralized computing based on Linux (currently Scientific Linux 6)
  - Similar models at SLAC, FNAL, CERN, BNL, ESRF, APS, etc.
- CLASSE accounts: single sign-on for all central resources
- CLASSE Linux systems give direct access to:
  - CHESS DAQ filesystems (data, software, project space)
  - Compute Farm (includes MATLAB, Mathematica, python, etc.)
  - Personal web pages

- Access:
  - Log directly into Linux:
    - **ssh <username>@lnx201.classe.cornell.edu**
      - On Windows, use PuTTY or X2Go (see Computing homepage)
    - **cd /nfs/chess/raw**
  - Graphical file access from Windows: use Samba
    - Connect to the CLASSE VPN (see Computing homepage)
    - In Windows Explorer, browse to \\chesssamba.classe.cornell.edu\raw

CLASSE Connection Diagram

End-user onsite computers
(incl. CHESS)

SSH
SFTP
SCP

Central Linux nodes
lnx201
CLASSE compute farm

NOT BACKED UP ON CLASSE SYSTEMS

local disk — Linux

local disk — Windows

local disk — Mac

Samba
remote file server

Central Filesystems
home directories
user disk
project areas
CHESS_{DAQ, MAIA, RAW, AUX, OPT}

BACKED UP

VPN

RedRover/eduroam & offsite

SSH
SFTP
SCP

Usage Patterns

- Example #1
  - Log into lnx201
  - Develop C++ or python analysis on CHESS_AUX
  - Submit data analysis batch jobs to Compute Farm
  - Write output to CHESS_AUX
  - Further analysis with interactive jobs on Linux
  - **Files backed up by CLASSE IT**

- Example #2
  - Log into lnx201
  - Run interactive data processing/ visualization
  - Stage output to CHESS_AUX
  - Use Samba/SFTP/ SCP to copy files to personal computer
  - **User responsible for backing up files**

- Example #3
  - Use Samba/SFTP/ SCP to copy data to personal computer
  - Analyze data locally
  - **User responsible for backing up files**

CHESS DAQ Hardware

- 10 Gb storage area network (SAN)
  - Enterprise-class storage devices, servers, network switches
- 2 x Infortrend iSCSI storage devices
  - Dual controllers, redundant power
  - 24 x 4 TB drives/device configured in 2 x RAID 6
  - Total 128 TB usable
- Files served by CHESS DAQ cluster
  - 5 x IBM x3550 M4 servers
  - Each has 2 x 6-core Intel Xeon, 128 GB RAM
- Networking:
  - 10 Gb IBM Blade switches
  - New multi-mode optical fiber runs throughout CHESS
- Throughput:
  - Up to 900-1000 MB/s writes (sustained 600 MB/s)
  - Average 200-300 MB/s reads
- IBM tape library, total capacity 250 TB (uncompressed)

Current Status & Plans

- First deployment of CHESS DAQ in October 2014
  - 24 TB written so far
- CHESS DAQ network available at all beamlines
- 7 of 11 beamlines are now writing to central storage
  - Missing: B1, C1, and MacCHESS (except BioSAXS @ G1)
- Archived: 2014-3, 2015-1, plus some legacy data

- Plans for summer down:
  - Archive 2015-2 data
  - Investigate networking bottleneck at A2 with Pilatus 300K
  - Organize network switches on experimental floor
  - Convert B1 and C1 (possibly)
- Longer term: develop central software repository for CHESS

Additional Resources

- CHESS DAQ documentation:
  - https://wiki.classe.cornell.edu/CHESS/DataStorageManagement
  - Lyris mailing list: chess-daqadmin-l@cornell.edu
    - Meeting next Monday 27 July, 2 PM, Wilson 374

- CLASSE computing homepage:
  - https://wiki.classe.cornell.edu/Computing/WebHome

- Linux @ CLASSE:
  - https://wiki.classe.cornell.edu/Computing/LinuxSupport

- Central file systems:
  - https://wiki.classe.cornell.edu/Computing/DataStewardship

- CLASSE Compute Farm:
  - https://wiki.classe.cornell.edu/Computing/ComputeFarmIntro