OAK RIDGE National Laboratory

The Future of Physics Division Operated Computing

Robert Varner

The 2021 Physics Computing Committee

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Membership of the 2021 Physics Computing Committee

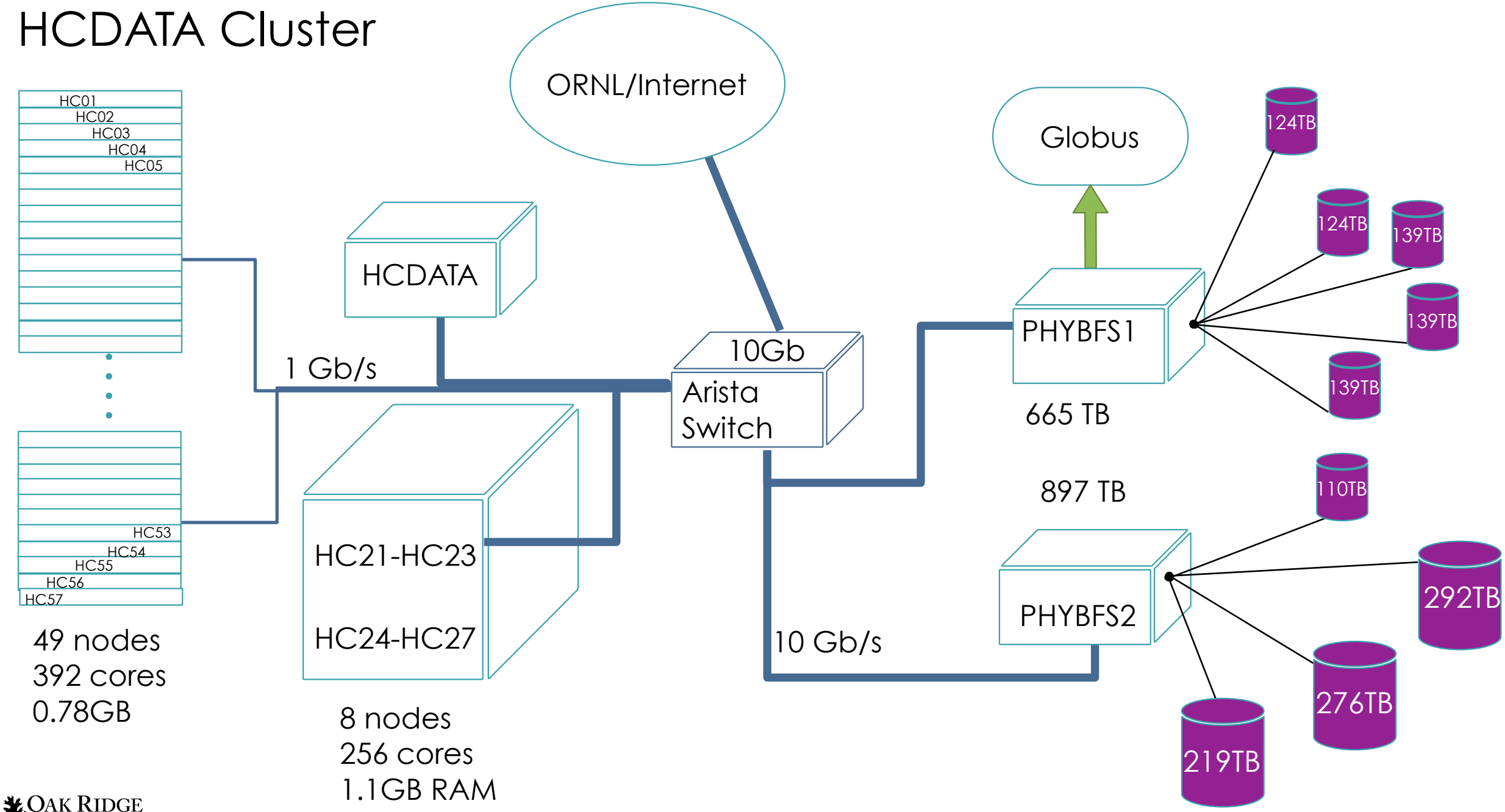
- Leah Broussard - Neutron Symmetries
- Raph Hix - Nuclear Theory
- Toby King - Nuclear Structure/Nuclear Astrophysics
- Jason Newby - COHERENT
- Ken Read - High Energy Nuclear Physics
- Michael Willis - ARDIDSA

- Robert Varner - Chair

Computing resources in the Physics Division

- What do we have
- What is the problem
- Where can we go from here

HCDATA Cluster

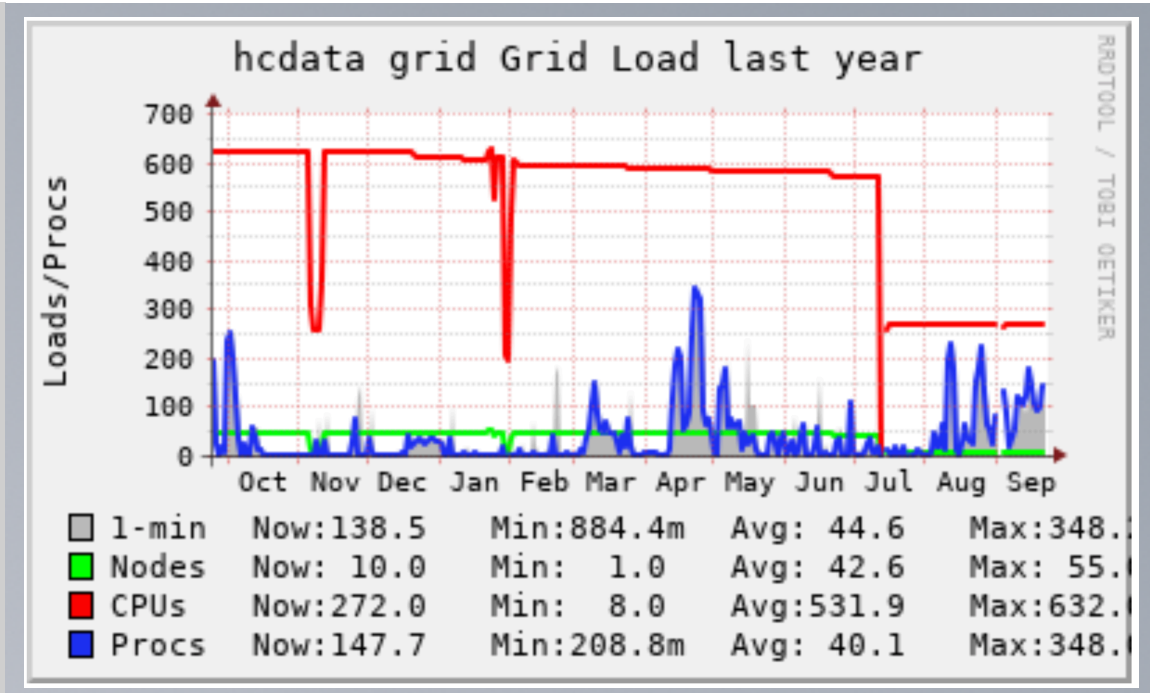


Physics Operated Computer Systems

- **HCDATA cluster - Open Research**
  - **HCDATA**(2009),
  - **PHYBFS1**(2015-2018) and
  - **PHYBFS2**(2019-2021) file servers
    - Total of **1.3 petabytes** of RAID6 disk
  - **HC01** to **HC57** -- 49 nodes (2009) + 8 nodes (2016) compute
  - 10Gbit Ethernet backbone, ARISTA switch (2015).
- *Only HC21-HC27 operating at this time*.

HCDATA usage

- 100 user accounts
- Less than 30 are ORNL staff
- Average number of processes: 40
  - Peak: 350
- Average I/O: 20MB/sec
  - Peak: 300 MB/s



HCDATA Cluster service to the division

COHERENT

- Data archiving
  - Raw data from several years of COHERENT measurements
- Login gateway
  - Enables all collaborators simple access to monitor and control the experiments
- Simulations
  - Simulations for a liquid Ar detector were performed on the cluster; simulations for future detectors are ongoing and planned.
- Data analysis
  - External collaborators, students and postdocs are processing data on the cluster
- Data transfer
  - Archiving at OLCF HPSS is ongoing, using the GLOBUS endpoint

HCDATA Cluster service to the division

MAJORANA/LEGEND

- Simulations
  - 60% of simulations for MAJORANA/LEGEND have been run on the HCDATA cluster (vs NERSC and institutional) and continue to run on the cluster
- Data analysis
  - ORNL analysis runs on the cluster, in parallel to MJD primary analysis
- Data storage
  - The full MJD dataset is available on the cluster
  - Enabled by the GLOBUS endpoint

HCDATA Cluster service to the division

PROSPECT

- Simulations
- Data Analysis
- Supporting UT students working with Galindo-Uribarri on Prospect

HCDATA Cluster service to the division

- nEDM
  - Planned data archiving and distribution to collaborators
- ARDIDSA
  - GEANT4 simulation cycles
- Nuclear Structure
  - Coulomb excitation and reorientation calculations for exotic beam measurement

Physics Division - those NOT using the cluster

Using outside resources
- University and other national laboratory systems
- e.g. BORAX at Livermore

Using desktop resources
- Powerful laptops and desktops, few TB disks
- Highly configured desktop workstations, e.g. Mac Pro

Using ORNL CADES cluster
- Neutron symmetries group
- Relativistic Nuclear Physics group
- ARDIDSA group
- Nuclear Theory (Gaute Hagen funded a condo with two nodes)

HCDATA Cluster - How much support does it take?

- ITSD services - 120 hours/year
  - OS updates, Security compliant configuration - often specified by processes internal to ITSD, not visible to us, software installation
  - RAID array configuration and diagnostics, new node configuration
  - Oversight of resources; disk replacement when Physics Staff unavailable
- Physics Division support - 200 hours/year
  - User support, software installation, system monitoring
  - Cluster specification, new hardware procurement
  - Failed disk replacement, maintaining spare disks

HCDATA Cluster - What is the problem?

- **Cost of operation**
  - ITSD was charging us around **$13k/year**
  - Sudden loss of our ITSD manager in April
    - ITSD revised that to **$80k/year**
  - ITSD cost scales with number of nodes - copies of the OS and user base

HCDATA Cluster - What else is a problem?

- **Aging infrastructure**
  - **The cluster has become a critical resource** for some programs
  - Most **compute nodes** are **12 years old**
  - **Youngest compute node is 5 years old**
  - **HCDATA** - login node and /home file server - **12 years old**
  - 10 nodes are shutdown for equipment issues
- No plan for routine update
  - occasional end-of-the-year funds
  - programmatic upgrades for disk space.
  - **Only the 10Gb switch has a maintenance contract**

ORNL new version of CADES - an opportunity

CADES = Compute and Data Environment for Science

- ORNL org burden funded cluster - operated by NCCS (ORNL supercomputer center)
- 18,000 cores - AMD EPYC
- Perhaps as many as 500 GPUs
- CPUs + modest /home file system + 2.3PB scratch file system

- Cost model: "Birthright", scheduling using a Fair-Share algorithm
- Access:
  - ORNL staff, and guests
  - Open Research
  - GLOBUS
- *Intention to support computation and data flow across ORNL projects*

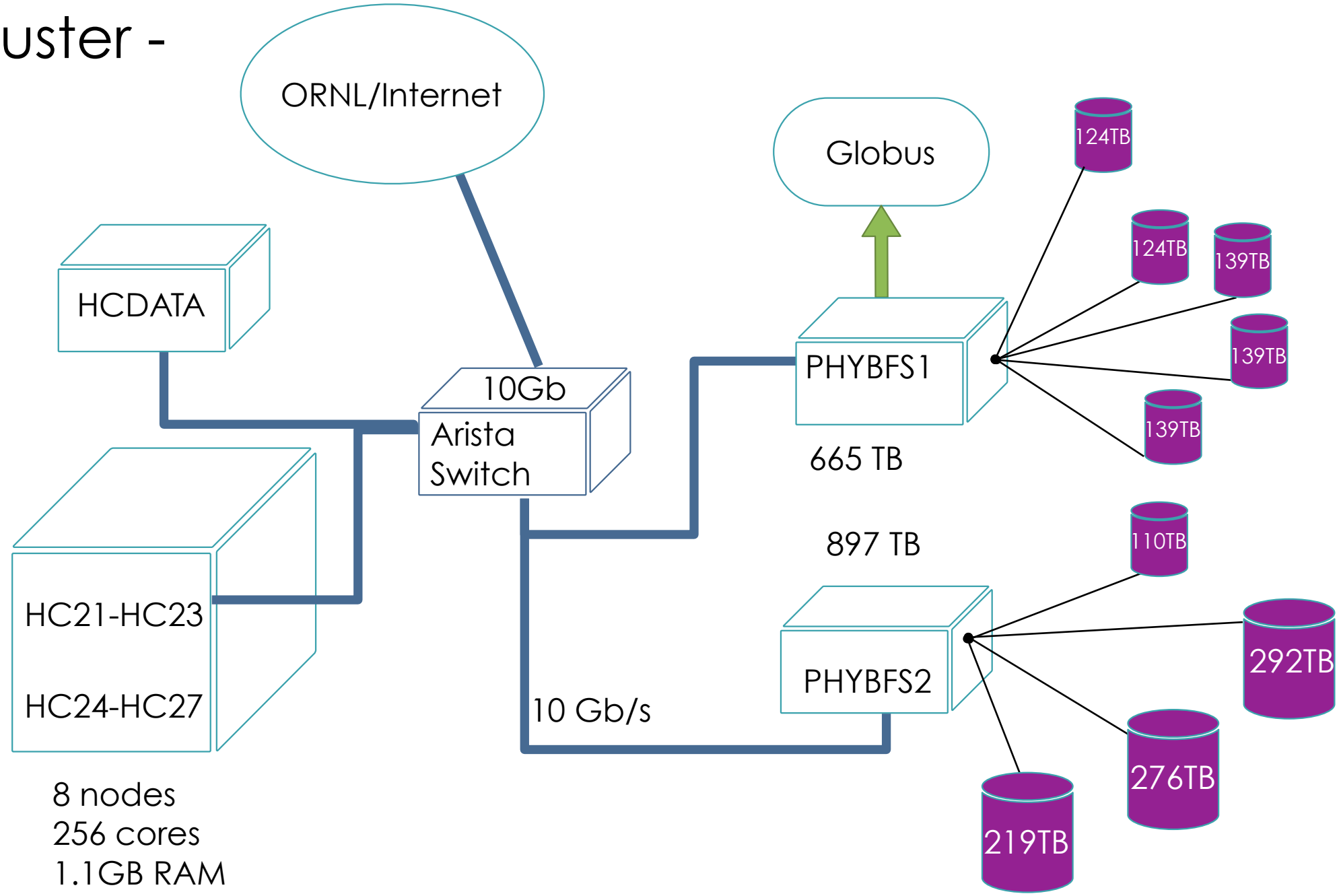
ORNL new version of CADES - an opportunity

- Beyond the birthright access
  - Well-defined projects will submit proposals for enhanced access within CADES - program advisory committee to judge
  - Well-funded projects will have the opportunity to purchase access to additional cycles beyond CADES - exact mechanism is not defined
- What about large datasets?
  - Developers want to address this.
  - 100 TB data sets not a problem
  - 1.5 PB needs more development

ORNL new version of CADES - not a copy of current CADES

- Transition from current CADES resources to new CADES is still being worked out.
  - I do not know what this means, really.
- The new CADES will not have any Moderate resources
  - Current Moderate condos will continue to operate as they have.
  - In the long term, this will be addressed

HCDATA Cluster - FY22

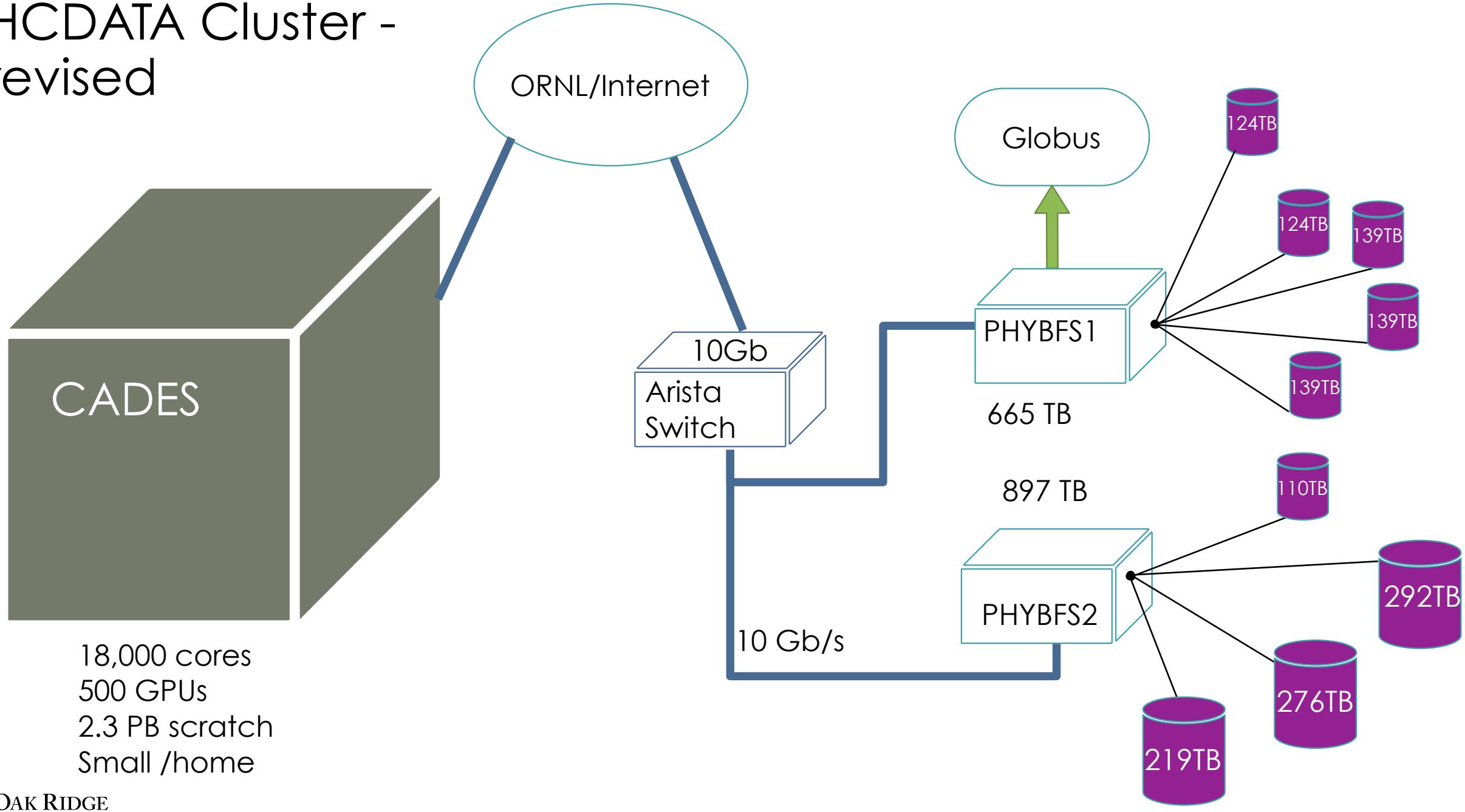


HCDATA - transition to CADES

Physics Division will support and encourage our staff and collaborators in the transition to CADES. The CADES developers are interested in supporting our workloads, even the data handling.

- A rough timeline
  - Downsize the cluster to fewer nodes - FY21 Q4
  - Begin utilizing the new CADES - FY22 Q3
  - Reevaluate the downsized cluster usage - FY23 Q1
  - Begin full transition to CADES - FY23 Q2,Q4
  - Complete the transition to CADES - FY23 Q4

HCDATA Cluster - revised



18,000 cores
500 GPUs
2.3 PB scratch
Small /home

HCDATA to CADES transition - details

- Using CADES should be very similar to using HCDATA - login hosts with batch queues
- Large-scale, long-term data storage at CADES is not even in the planning. We must operate the file servers with GLOBUS indefinitely.
- Access should be like that to NCCS, perhaps a little simpler. What that means is unclear.

- The New CADES is a new development, with many issues to be resolved, by new management. We can be part of helping define the solutions.

OAK RIDGE National Laboratory

Physics Computing Planning

- Wait!
- There's more to be done.

Other Physics Division Computer Systems

- **TROLL (2009)**- theory development system (Hagen)
- **ORPH02 (2013)** - Physics procurement web page, mailman email lists
- **PHYXWEB (2015)** - 2013 external web pages, workshops, "FTP", radware, fsnutown, Hawaii2018
- **MJDB (2016)** - MAJORANA database server, QA document server (Varner, etal)
- **PHYNFS3(2013)** - SMB file server (HRIBF, IRIS documents), former data acquisition fileserver, archival files
- **HCBACKUP (2010)** - backup for PHYNFS3
- Other data acquisition Linux systems in room T210 (2014-2016), mostly used for simulations now.

ORPH02

- *Physics Procurement System*
  - Database of procurement details, approvals
  - Web-based interface for
    - entering data,
    - approvals
    - Lookup prior purchase requests
    - Routing approvals to manager, requests to procurement staff
  - 20 years old - developed for Physics Division by Doro Wiarda
    - C++, Apache httpd, mysql

ORPH02 (continued)

- Procurement system options
  1. Forget about it, just redefine our processes around ORNL BUY
  2. Virtualize the existing system -
     - run the code in a container service at ORNL
     - Someone else has to learn it to maintain it. (C++, httpd, mysql)
  3. Work with ITSD to devise some kind of forms through SharePoint
     - Decide about approval stream
     - Decide about record keeping

ORPH02 (continued)

- Mailman
  - Easy mail lists, web interface, PHP.
  - phyall; employees; nufunsym-l; seminar, physicscluster, fnpbjournal, fnpbusers, phybrownbag
  - These an alternative server.
- Elogs
  - Computer system manager logs
  - HRIBF electronic logs

PHYXWEB - Linux Web server

- The Physics Division web page server - PD pages are mostly obsolete
  - Replaced by https://www.ornl.gov/division/pd ?
- Server of some workshops and conferences
  - QM09
  - Hawaii2018
- HRIBF - UPAK data analysis and data acquisition software
- radware.phy.ornl.gov
- fsnutown.phy.ornl.gov
- fds.phy.ornl.gov (now implemented at fds.ornl.gov)
- Decisions needed about its future

PHYNFS3 - Linux 16 TB fileserver

- File server
  - NFS files to workstations in Physics - remains of former DAQ network for HRIBF - majority of user files are for users no longer here
  - SMB files to PC's - mostly engineering and design data for Physics detector systems, HRIBF and IRIS ion sources (Dowling)
  - NIS credential server
- Plan:
  - virtualize the SMB files on a server run by ITSD
  - Plan to archive NFS user files, invite current users to save their files
  - Decommission use of NIS for authentication
- Shutdown the server - although it still works and I have spare disks.

HCBACKUP - LINUX fileserver 24TB

- Backup of ORPH02 and PHYNFS3
- Log server for orph02, phynfs3, and all the Workstation enclave systems in Room T210
- Spare parts for HCDATA, if necessary.
- Shutdown when ORPH02 and PHYNFS3 cease operations