BNL Scientific Data and Computing Center (SDCC) Site Report

Caramarcu Costin <caramarc@bnl.gov>

HEPiX Spring 2021 – Online Workshop

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center

U.S. DEPARTMENT OF ENERGY

Scientific Data and Computing Center Overview

Located at Brookhaven National Laboratory (BNL) on Long Island, New York

SDCC was initially formed at BNL in the mid-1990s as the RHIC Computing Facility (RCF)

- Tier-0 computing center for the RHIC experiment
- RHIC celebrating 21 years of success. RHIC run 21 started in Jan 2021
- US Tier-1 Computing facility for the ATLAS experiment at the LHC
- Also one of two ATLAS shared analysis (Tier-3) facilities in the US
- Data center for US Belle II experiment
- BNL selected as the site for the upcoming major new facility Electron-ion Collider (EIC/eRHIC)
- sPHENIX - scheduled to start taking data in 2023

SDCC Overview (Cont.)

- Also providing computing resources for various smaller / R&D experiments at BNL:
  - DUNE, EIC, LSST, etc.
- Serving more than 2,000 users from > 20 projects

Besides providing computing / storage resources for our user community, we've recently expanded our emphasis on developing and administrating new collaborative tools:

- Invenio, Jupyter, BNL Box, Discourse, Gitea, Mattermost, etc.

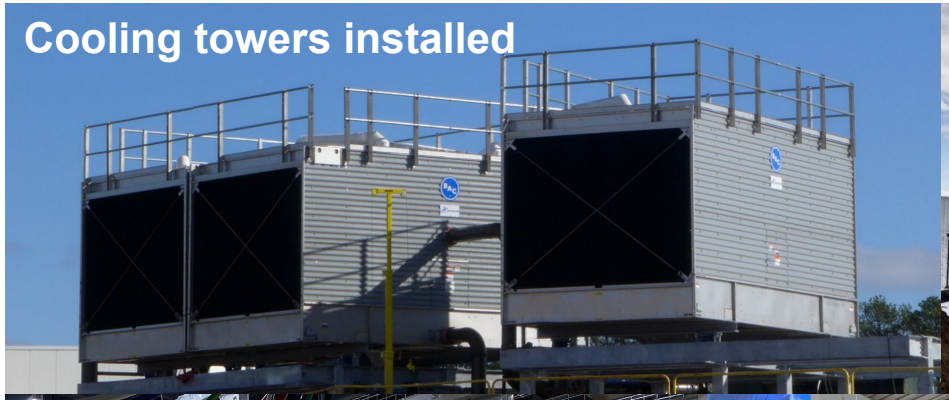
Currently we are 38 full time employees.

Hiring if you're potentially interested in joining SDCC:

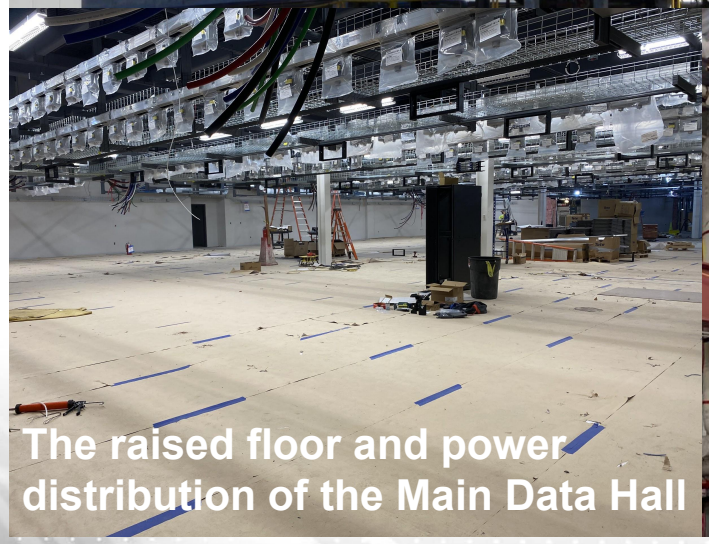
- Check https://jobs.bnl.gov for updates

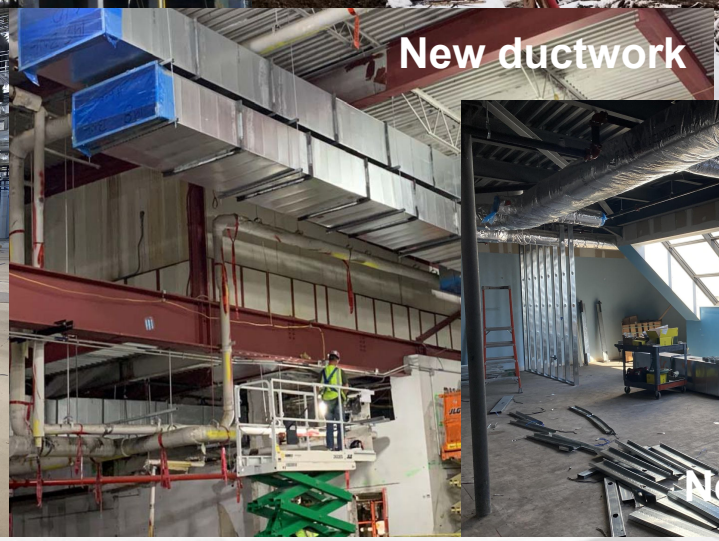
B725 Datacenter Construction

July 2020 - Mar 2021: construction is going ahead after 3 months of delay in 2020Q2 due to COVID-19. The early occupancy of B725 datacenter is expected to begin in June 2021 (network equipment deployment is expected to start in April-May 2021)


Cooling towers installed

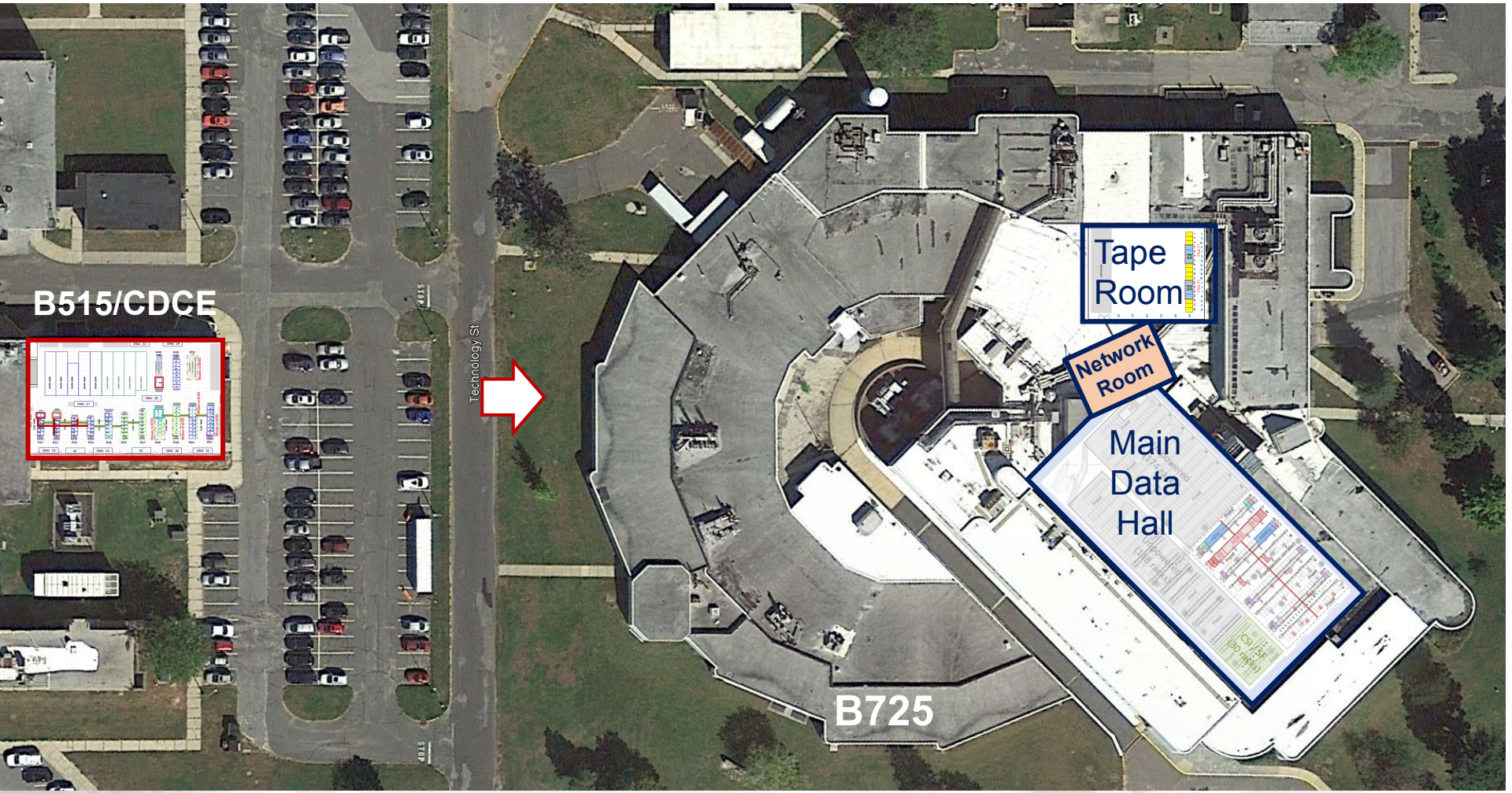

New generator yard


The raised floor and power distribution of the Main Data Hall


New ductwork


New office areas

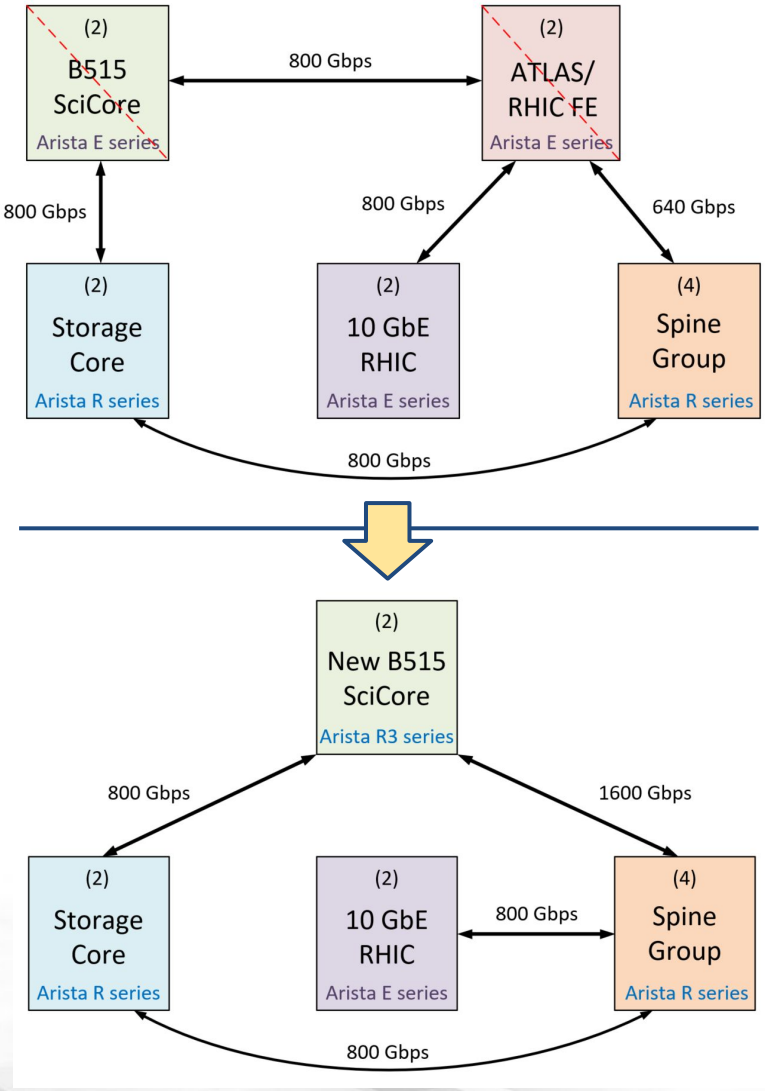
New datacenter is being designed & constructed for the SDCC Facility in B725 in FY19-21 period, with migration of most of the spinning disk storage and all of the compute capability (mostly via gradual HW refresh process) to the new datacenter from the existing B515 based datacenter to happen in FY21-23



Preparation for the Beginning of B515 / B725 Datacenter Operations in 2021Q3

SDCC Facility Wide Network Intervention of Dec 1, 2020:

- Most elaborate network intervention in SDCC Facility since 2017
- Aiming at performing the hardware refresh of central networking system of B515 datacenter and making it ready for the *transparent* transition to the combined B515 / B725 datacenter operations expected in 2021Q3.
- Performed successfully during the 6:00-18:00 BNL LT time window.
- Routing and interswitch links are optimized across the SciZone; service blocks consolidated as a result.
- Likely being the largest scale network intervention we are going to need to perform for the SDCC Facility until the next hardware refresh of the central network equipment expected in FY25-26.

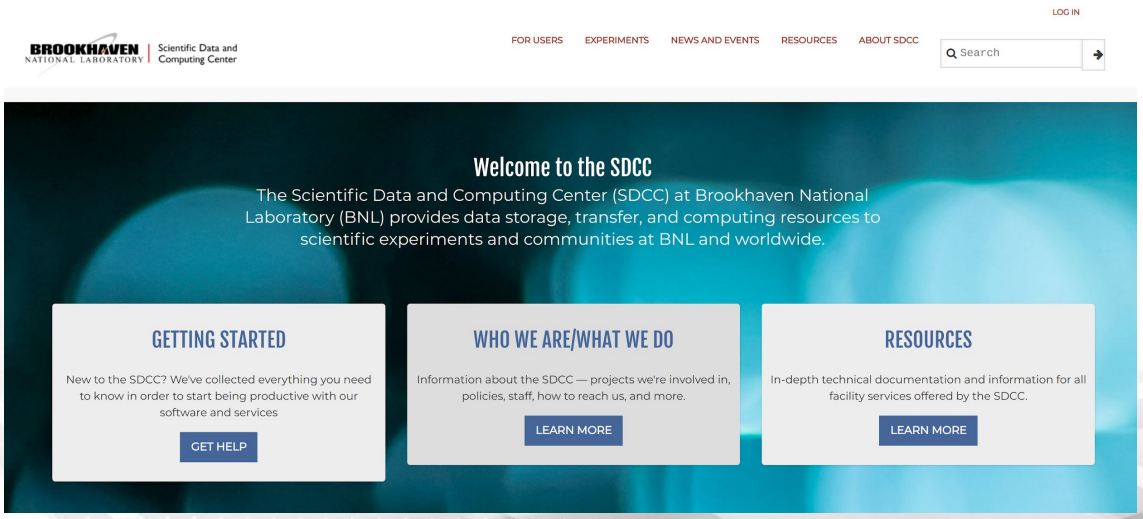


NSLS-II Activities

- NSLS-II is a user facility with 29 active beamlines and is expected to host more than 4000 users/year
- Apache Guacamole used as Remote desktop gateway
- Lustre is the primary file system available via Globus/SFTP/SSHFS/SMB.
- Globus Endpoint “NSLS2” running latest Globus version 5.4.
- New RHEV virtualization cluster with 7 dedicated hypervisors able to host 100s of virtual machines.
- User SSH gateways require Active directory credentials and DUO for Multi-factor authentication.
- Robinhood policy engine used for auditing and complete lifecycle management.
- More details on Will Wednesday presentation
    - Supporting a new Light Source at BNL

SDCC Drupal

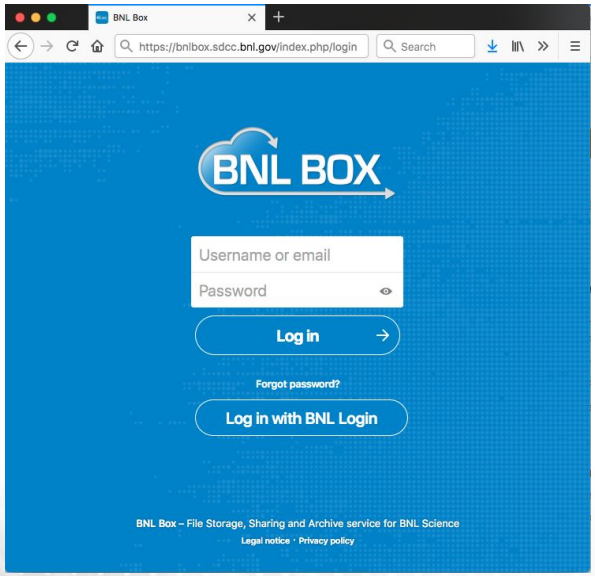
- Several Drupal instances have been deployed for both experiments and the SDCC
- The most recent site to be deployed using our Drupal model is the SDCC public facing page “sdcc.bnl.gov” which includes some key features:
  - Authentication easily managed through usage of Keycloak
  - Enabled Drupal roles and page access, allowing for secure private pages based on user role and open public pages for all visitors
  - Leveraged advanced web technology to maintain existing services
  - Updated and modernized theme

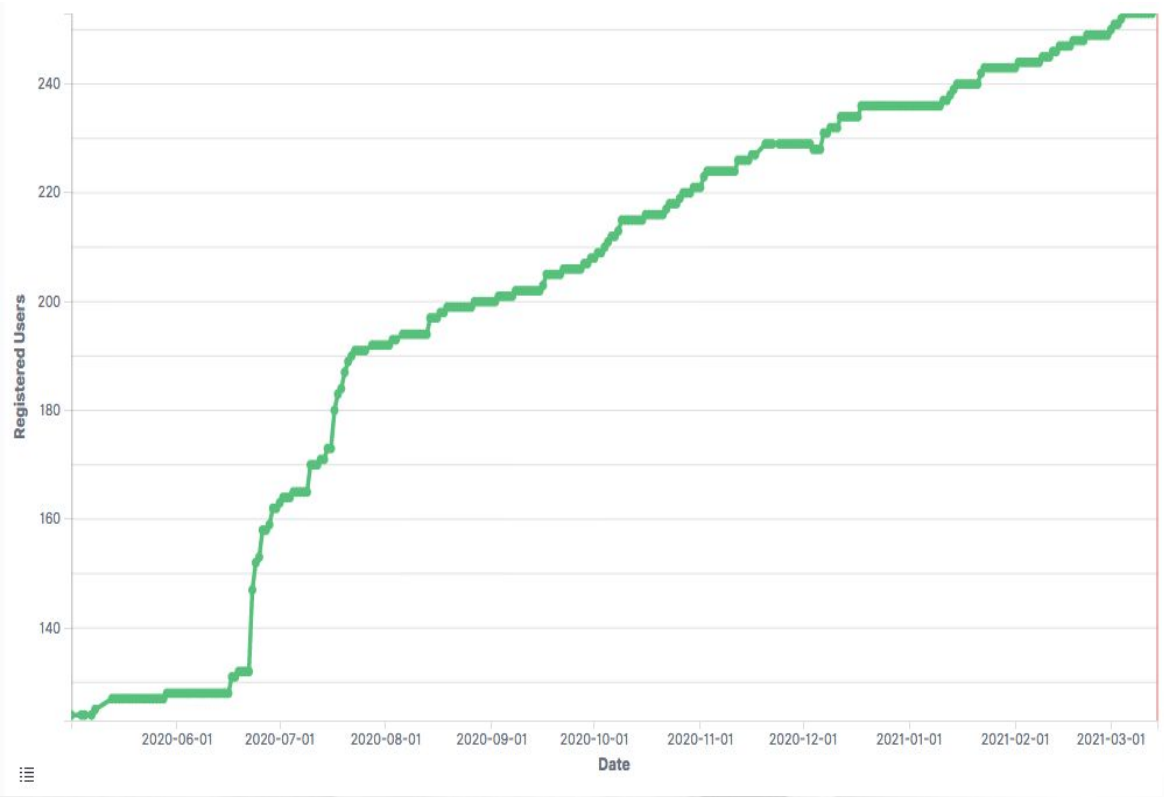


BNL Box

- Continued growth in use of Nextcloud-based cloud storage service from across the BNL scientific landscape
- Currently 253 users storing ~3 TB and ~4M files
- **New:** hourly ClamAV antivirus scanning leveraging Lustre changelog/robinhood
- Testing integration with Jupyterlab (via fuse/rclone)

https://bnlbox.sdcc.bnl.gov



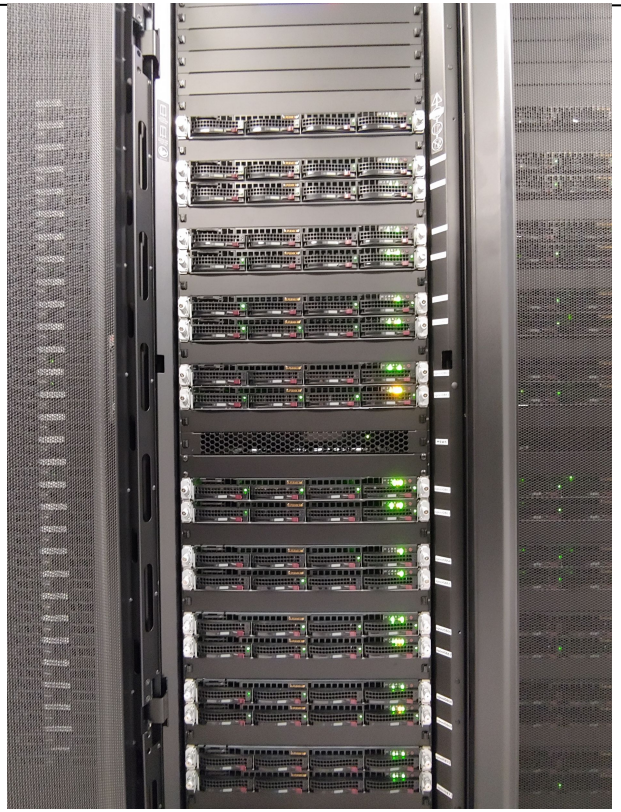


High Throughput Computing

Providing our users with ~2,000 HTC nodes:
- ~80,000 logical cores
- ~890 kHS06

73 new Supermicro SYS-6019U-TR4 1U servers brought online in July 2020
- Dual Intel Xeon Cascade Lake 6252 CPUs @ 2.4 GHz (96 log. cores total)
- 12 x 16 GB (192 GB total) DDR4-2933 MHz RAM
- 4 x 1.8 TB SSDs
- 1U form factor
- 1140 HS06/node = ~83 kHS06 total

All nodes running Scientific Linux 7 for some time
- SL6 Singularity containers provided to experiments which still require this OS



New Cascade Lake-based Supermicro 6019U-TR4 Servers

High Performance Computing

Currently supporting 5 HPC clusters

Institutional Cluster (IC)
216 HP XL190r Gen9 nodes with EDR IB
- 2x Intel Broadwell Xeon E5-2695v4 CPUs (36 cores total)
- 256 GB RAM (DDR-2400)
- 2x K80 or P100 GPUs

Skylake Cluster
64 Dell PowerEdge R640 nodes with EDR IB
- 2x Intel Skylake Xeon Gold 6150 CPUs (36 cores total)
- 192 GB RAM (DDR4-2666)

KNL Cluster
142 KOI S7200AP nodes with dual rail Omnipath Gen.1 interconnect
- 1x Intel Xeon Phi 7230 CPU (256 log. Cores total)
- 192 GB RAM (DDR4-1200)

ML Cluster
5 HP Proliant XL270d Gen10 nodes with EDR IB
- 2x Intel Xeon Gold 6248 CPUs (40 cores total)
- 768 GB RAM (DDR4-2933)
- 8x V100 GPUs

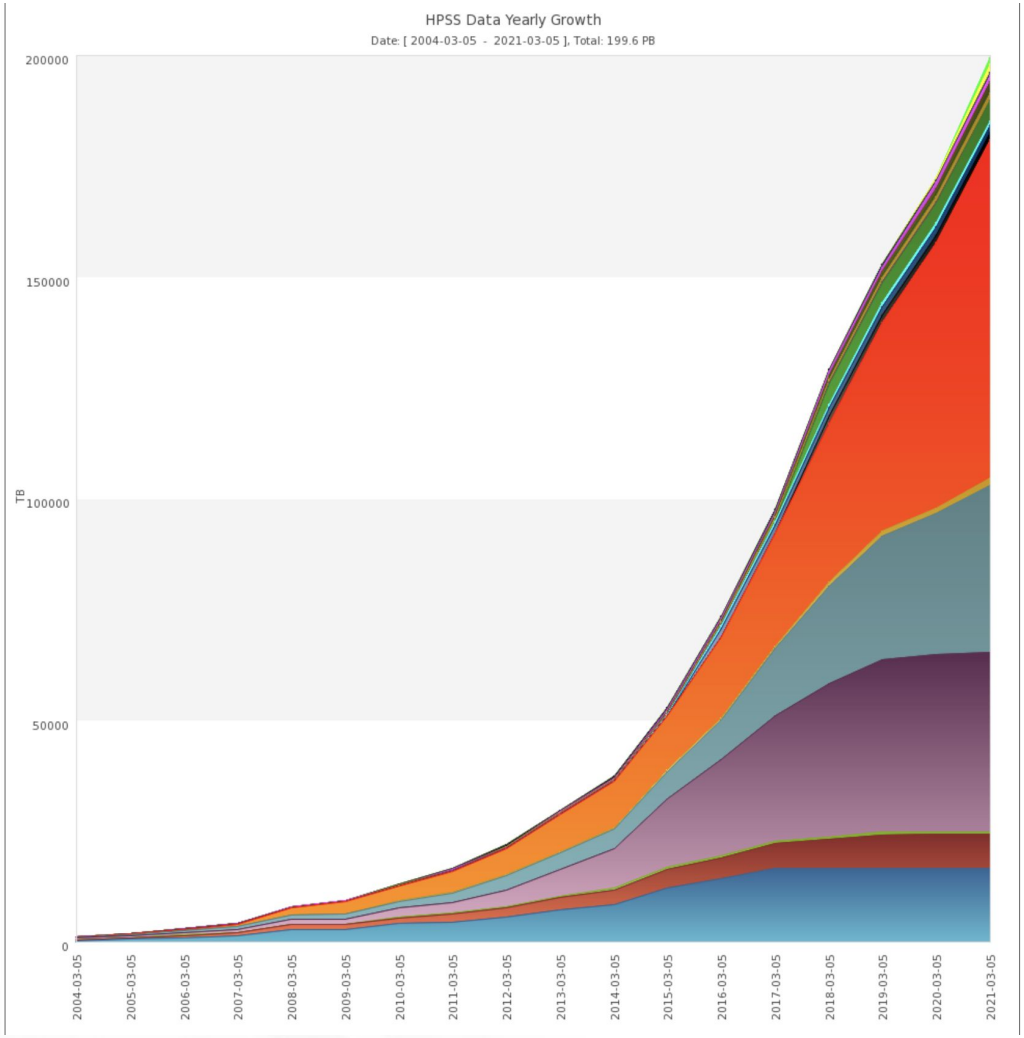
New HPC cluster for NSLS2
30 1U nodes with EDR IB
- 2x Intel Cascade Lake Xeon Gold 6252 (48 cores total
- 768 GB RAM (DDR4-2933)
- 12 of the hosts with 2x V100 GPUs

NSLS2 Cluster

New institutional resources at BNL see Tony Wong presentation later

Tape Storage (HPSS)

- Running HPSS v7.4.3.2 since Dec 2018
- Approx. 202.6 PB accumulated data
- Newly acquired TSM TS4500 tape library and data mover are in production since March 2021
- Upgrade to HPSS v9.x in July 2021
- Evaluation for new 20K-slot tape library
- Relocation to new data center

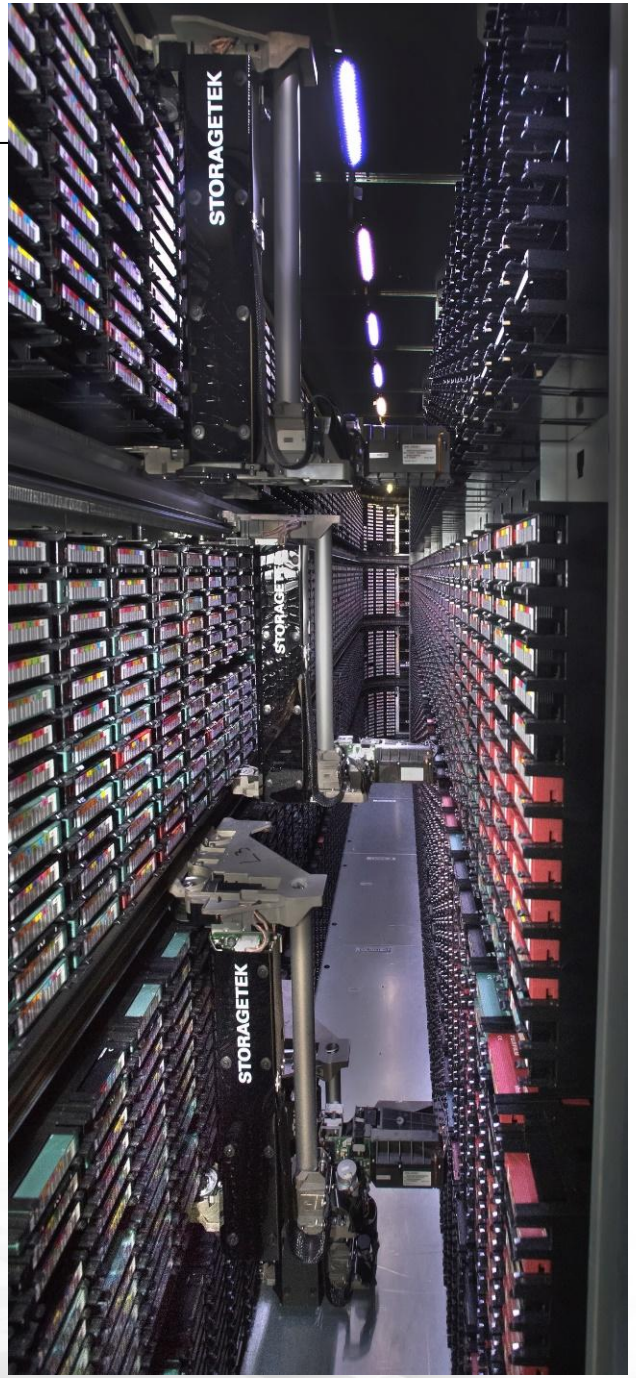


Tape Storage (HPSS)

- Upgrade LTO7 to LTO8 and LTO-8 deployment after HPSS upgrade in July
- LQCD and Belle2 requirements have dramatically increased
- PHENIX DST remains on LTO-6, due to lower data volume

Data Volume in 2020

- Data import : 22.4 PiB; 30,550,438 files
- Data export : 30.66 PiB; 16,310,528 files



Central Disk Storage

- Currently have 7 GPFS filesystems
  - Total of 14PB of raw storage and > 1.5 billion files
- GPFS version 4.2.3.x is end of service as of Sept 2020
  - Updated to GPFS version 5.0.5-3 on all clients/servers
- Increased Lustre 2.12.6 footprint to 7 PB
- Using Lustre CopyTool for HPSS tape archiving
  - Good performance and stability, easy to manage with Robinhood policy engine
- Still in the early phases of evaluating S3 storage in favor of POSIX storage

dCache/XROOTD

dCache
Managing over 55 PB of data total

- ATLAS (v6.2)
  - Joined WLCG DOMA SRM-HTTP test recently
- Belle II (v6.2)
  - 1 PB new storage added recently
- PHENIX (v5.2)
  - Mix of central and farm node storage
- sPHENIX (v6.2) recently added in production
- QoS v6.2 testbed
  - BNL and FNAL distributed filesystems
- Simons Foundation (v5.2)
- DUNE (v5.2)

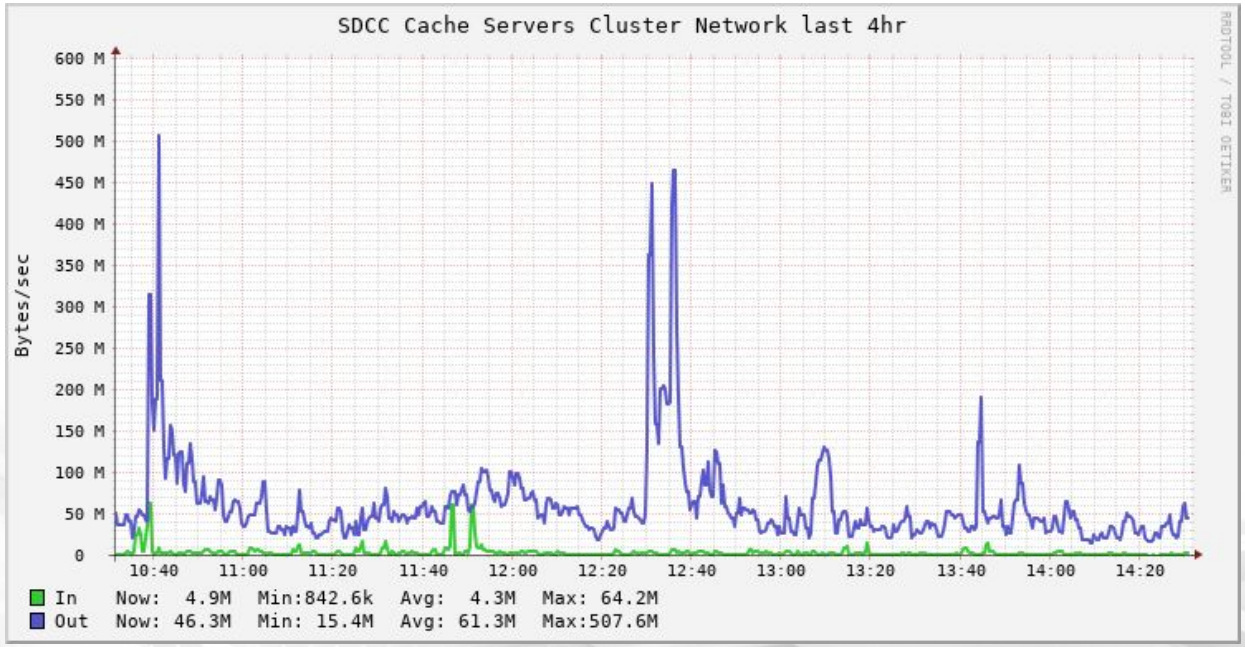
XROOTD
~11 PB total storage for STAR

- Mix of central and farm node storage
- Recently enabled write capability on central storage portion, previously it was read only

Running version 4.7.1

CVMFS

- Stratum Zero in production
  - 13 local repositories for BNL-based experiments and groups
- We use the stratum zero for publishing ATLAS python environments with SLAC
- All servers running version 2.8.0 (latest)
- Stratum One continues to grow in size & utilization
  - 30 TB of data in 104 replicated repositories



REANA

- REANA is platform for reproducible scientific analysis
  - Users run workflows in containers for reproducibility
  - Deployment/orchestration via k8s
- We've deployed a usable test v0.7 REANA cluster at SDCC/BNL
  - Running on our staff k8s cluster
    - Utilized provided helm chart to deploy
  - Ideally would run on our Openshift/OKD cluster instead
    - Issue with helm chart and unprivileged deployment
- Discussion with developers in REANA gitter
- Analysis and Data preservation is part of SDCC mission

Overleaf

- Growing interest in Overleaf (online collaborative LaTeX editor) at BNL across multiple groups (HEP/NP, Photon Sciences, Computational Sciences, etc.)
- SDCC is centralizing support for Overleaf at BNL
- Local installation due to confidential/sensitive nature of some documents
- Part of a growing portfolio of collaborative tools (MatterMost, BNLBox, Gitea, Invenio, etc.)

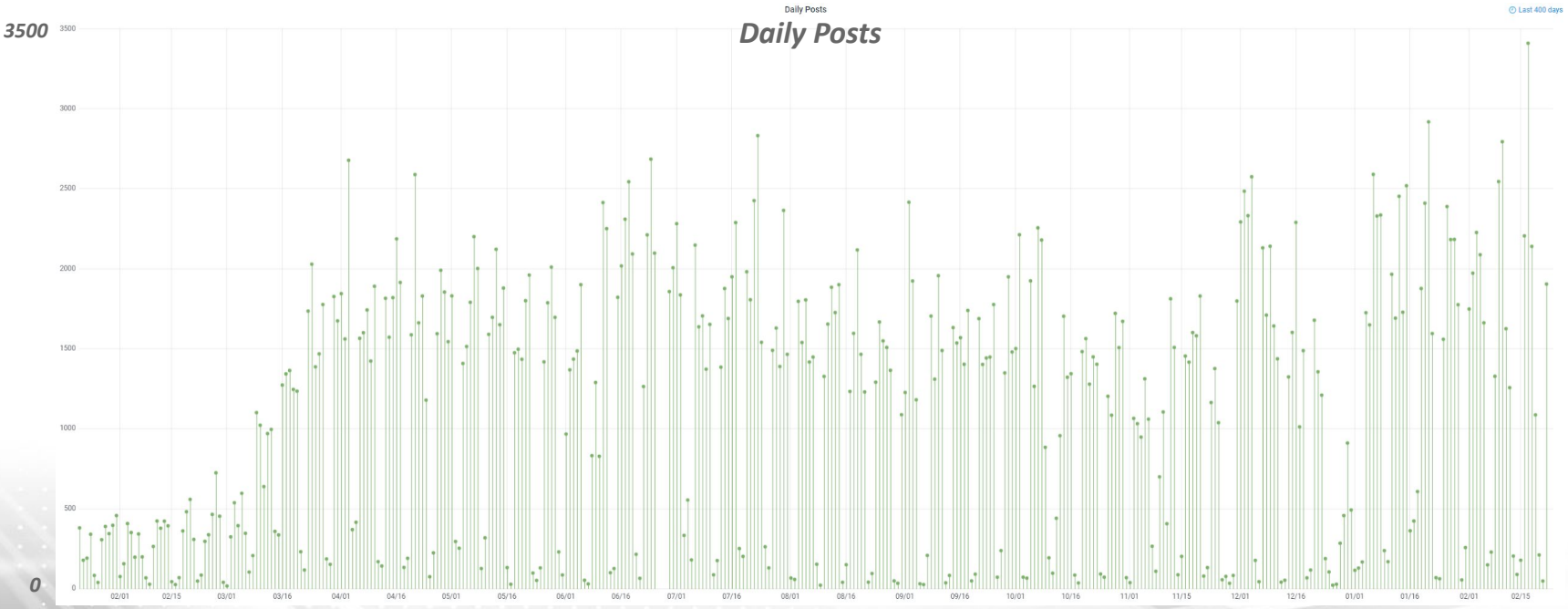
SDCC Talks @ HEPiX Spring 2021

- NSLS-II support activities (William Strecker-Kellogg)
- Magnetic Tape for Mass Storage in HEP (Shigeki Misawa)
- A Unified approach towards Multi-factor Authentication (MFA) (Masood Zaran)
- New institutional resources at BNL (Tony Wong)
- SDCC operations during transition to new data center (Alexandr Zaytsev)

Questions?

Thanks to the following people at BNL for contributing to this presentation:

John De Stefano, Alexandr Zaytsev, Costin Caramarcu, Tim Chou, Tejas Rao, Ofer Rind, Jason Smith, Will Strecker-Kellogg, Tony Wong, Iris Wu, Louis Pelosi, Chris Hollowell, Christian Lepore, Jane Liu and Robert Hancock



Mattermost Chat usage after lockdown (February 2020 - March 2021)