Computing Consortium Organization

H Schellman

August 2022

Now 2 - ½ day workshops

- https://indico.fnal.gov/event/55942/
- Talks by early career collaborators
- See where we are on multiple projects
- We need to report to DOE on what we are doing
  - Annual reports are due in early January
- Further in the future
  - actual meeting (in Oregon? At the actual dunes?)

US Consortium – (why we are here today)

- In March 2020 US collaborators submitted a proposal to DOE
  - "Essential computing and software development for the DUNE experiment"
- Originally 10 institutions, 3 were working on advanced pattern recognition and were not funded.
- Now have 7 – funding arrived at OSU 9/29/2021 – got through the paperwork 2 months later. But fiscal year starts on 5/1/2022….
- Argonne, Brookhaven, Colorado State, Fermilab, Minnesota, Oregon State and Wichita State
- Already done a lot – report here
- We need to complete what we said we would do and apply for longer term funding over the next year.

DOE guidance

- Priority should be given to Task 1 (Data bases), within Task II (Frameworks and data structures) to data modeling, data discovery, and signal processing, and within Task III (Common Software, Training, Standards and Infrastructure) to build systems and code standards.

Task I - Databases

- BNL – Paul Laycock, Lino Gerlach, conditions database
- CSU = Norm Buchanan, Ana Paula Vizcaya Hernandez
- Minnesota – Hajime Muramatsu, Marvin Marshak + Alex Wagner
- FNAL - Brandon White, Steve White, Igor Mandrichenko, Vladimir Podstavkov

Task II - Data model (large and small scale)

- Large Scale (FNAL and BNL)
  - MetaCat (Igor Mandrichenko)
  - Data Dispatcher (Igor M. , Steve Timm)
  - Hardware/production systems (Kirby, Ken Herner)
  - Rucio (Doug Benjamin, Steve Timm, Robert Illingworth, FNAL team)
- Small Scale (ANL, BNL, CSU, FNAL, OSU)
  - HDF5 and decoding -- Barnali Chowdhury, Amit Bashyal, Peter Van Gemmeren, Tom Junk, Kurt Biery, Saba, CSU postdocs, Jake Calcutt, David Adams, Brett Viren, Doug Benjamin

Frameworks

- This is intertwined with event model and with ND integration
  - FNAL (Kyle Knoepfel, Chris Jones, Andrew Norman and Tom Junk)
  - BNL (Paul Laycock, Brett Viren)
  - Argonne (Barnali Chowdhury and Amit Bashyal)
  - OSU (Jake Calcutt)
  - CSU
  - WSU (ND integration)
  - Wirecell people

Task III Code management/build systems/standards

- Code management and documentation
  - Tom Junk, Jake Calcutt, David Adams
- ND integration
  - Michael Muether, Palash Roy
- Ongoing training and documentation work
  - Ken Herner, Claire David, Mike Kirby, Tom Junk … and Dave DeMuth
  - HMS is now part of a second DOE graduate training grant!

Overall organization of DUNE offline

Consortium	Sites	• Collaboration
• Project management • Core software development • Core operations • Interfaces to DUNE consortia • Interfaces to other projects • Training • User support	• Storage • CPU • Networking	• Algorithms • Operations help • User support • Production group • Validation • Calibrations • New ideas

What problems are we trying to solve

- LAr TPC's have very large trigger records (200 MB for protoDUNE vs. < 10 MB for ATLAS/CMS). Need to be able to access small chunks of data efficiently.
- possible SuperNova -> 460 TB of data
- New detector technologies
- Many subsystems in ND
- We're supposed to use 75% non-FNAL computing
- Keep up with general chaos due to OS/Authentication/architecture/HPC evolution and people being people.

DUNE FD-Data for Supernova

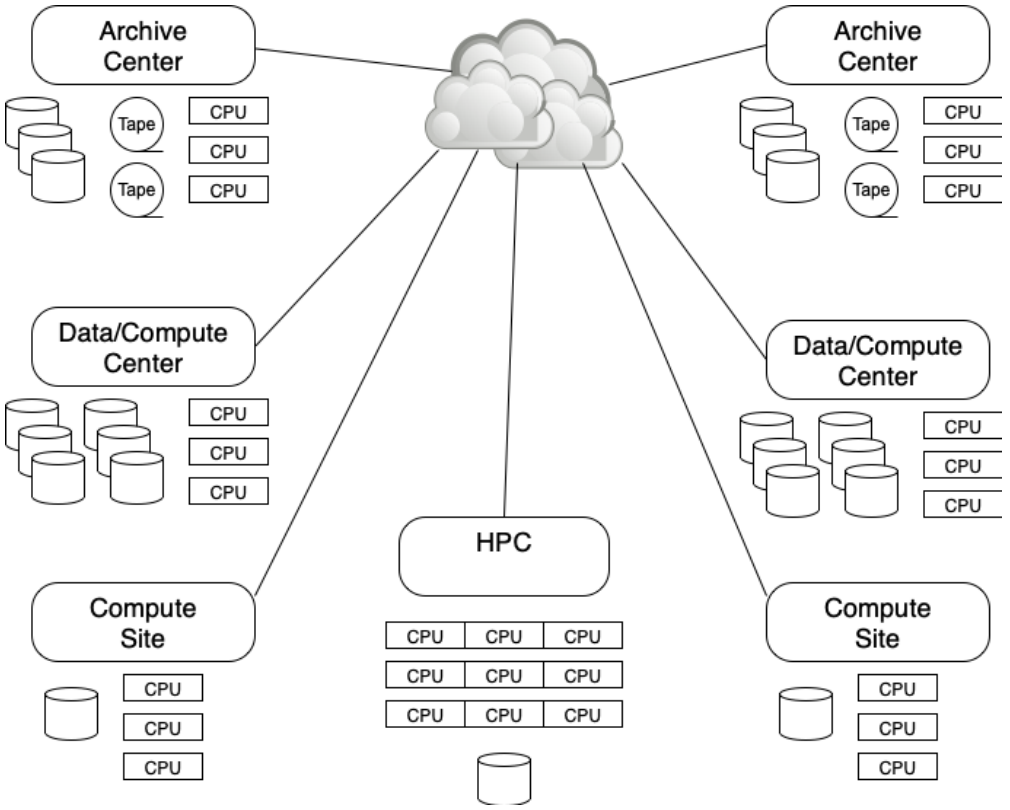
Pack 150 5 ms APA readouts
into a 6 GB file

Ship 20,000 time slices (x 4 modules)

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

Distributed computing model

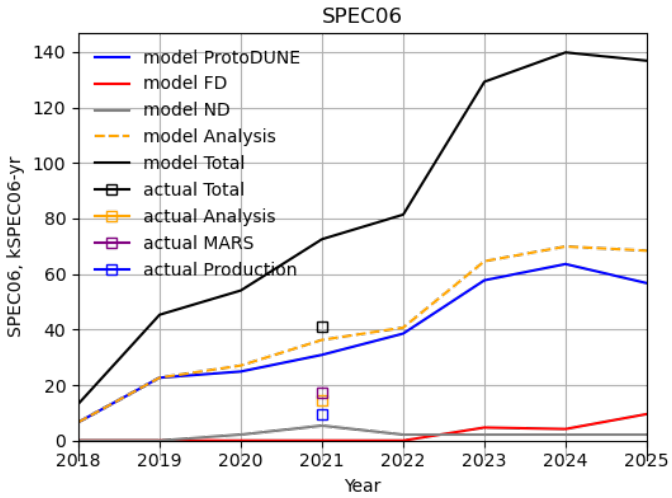
- Less “tiered” than current WLCG model → **DOMA**
- Collaborating institutions (or groups of institutions) provide significant disk resources (~1PB chunks)
- **Rucio** places multiple copies of datasets
- **We likely can use common tools:**
  - **But need our own contribution system**
  - **And may have different requirements for dataset definition and tracking**

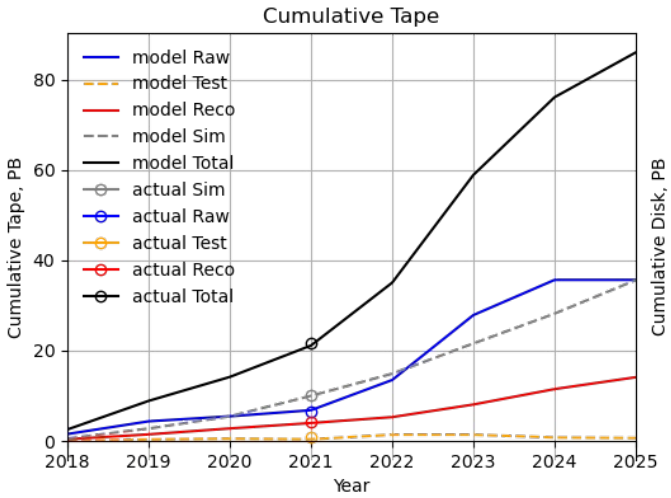


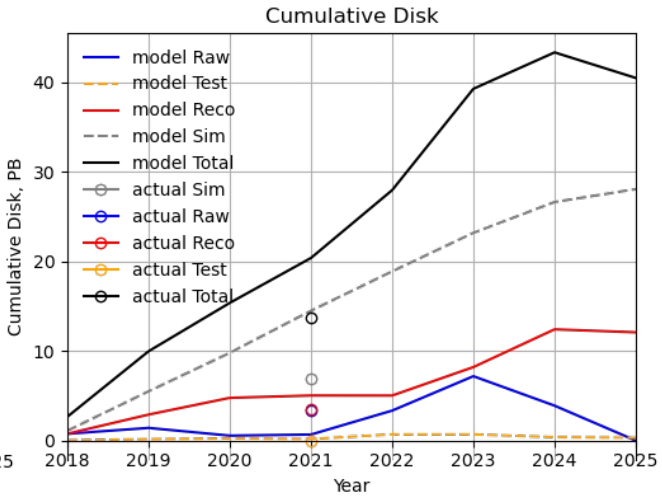
DEEP UNDERGROUND NEUTRINO EXPERIMENT

CDR - Resource estimates to 2025

- 2 copies of raw data on tape (6 months on disk)
- 1 copy of “test” data stored for 6 months
- 1 copy of reco/sim on tape
  - Currently assume 1 reco pass over all data and 1 sim pass/year
  - Assume reco/sim resident on disk for 2 years
- Assume 2 disk copies of reco and sim
  - impose shorter lifetimes on tests and intermediate sim steps.

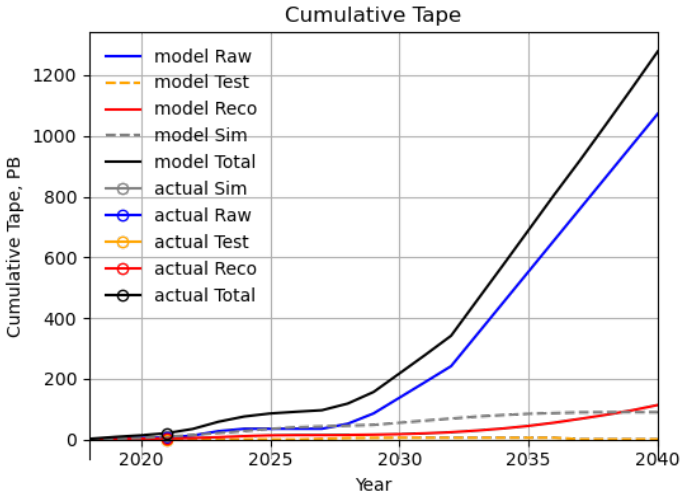


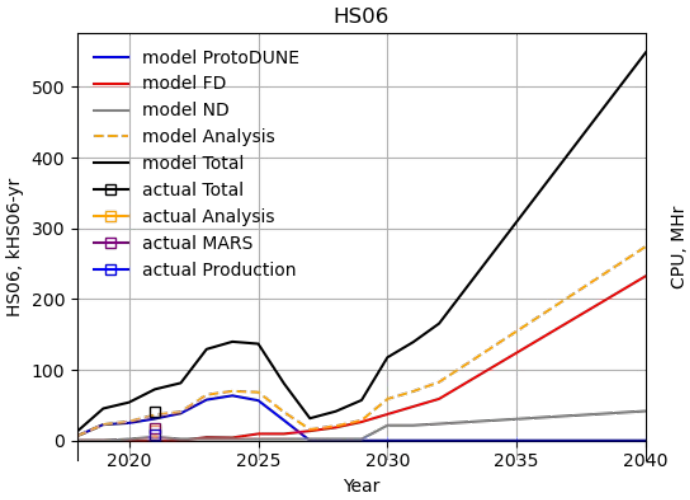


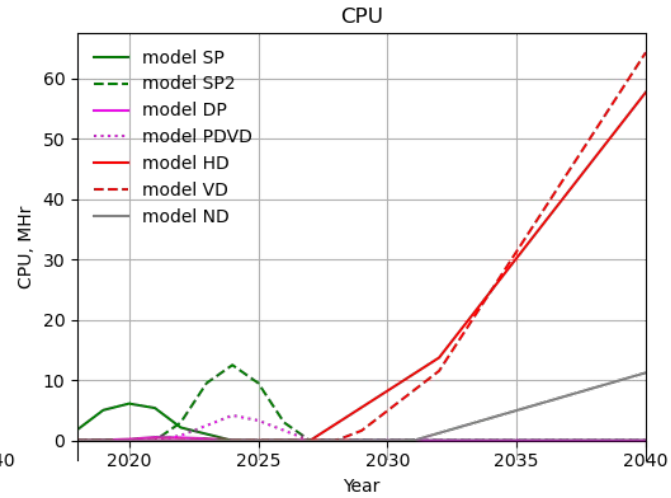


Longer term projections

- **VD assumed to be similar to HD**, raw data may be larger due to longer drift.
- 2 copies of raw data on tape (6 months on disk)
- 1 copy of "test" data stored for 6 months
- 1 copy of reco/sim on tape
  - Currently assume 1 reco **pass over all data** and 1 sim pass/year
  - Assume reco/sim resident on disk for 2 years
- Assume 2 disk copies of reco and sim





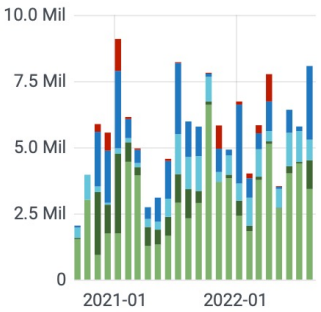


SCIENTIFIC COMPUTING SYSTEMS: DUNE

Total Wall Hours

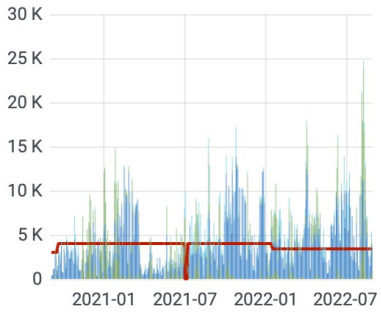
130 Mil

Monthly Wall Time



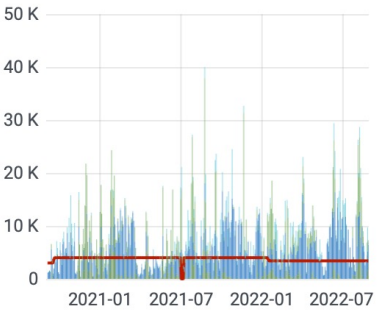
	avg	total
FermiGrid Analysis	2.86 Mil	71.43 Mil
FermiGrid Production	597 K	14.92 Mil
OSG Analysis	661 K	16.53 Mil
OSG Production	1.07 Mil	26.73 Mil
NERSC Analysis	0	0
NERSC Production	200 K	5.01 Mil

Running Batch Jobs



	max	avg
Production Onsite	9.20 K	448
User Onsite	13.83 K	3.54 K
Allocation	4.00 K	3.76 K
Production OSG	12.76 K	988
User OSG	4.54 K	532

Weighted Slots Claimed by Running Batch Jobs

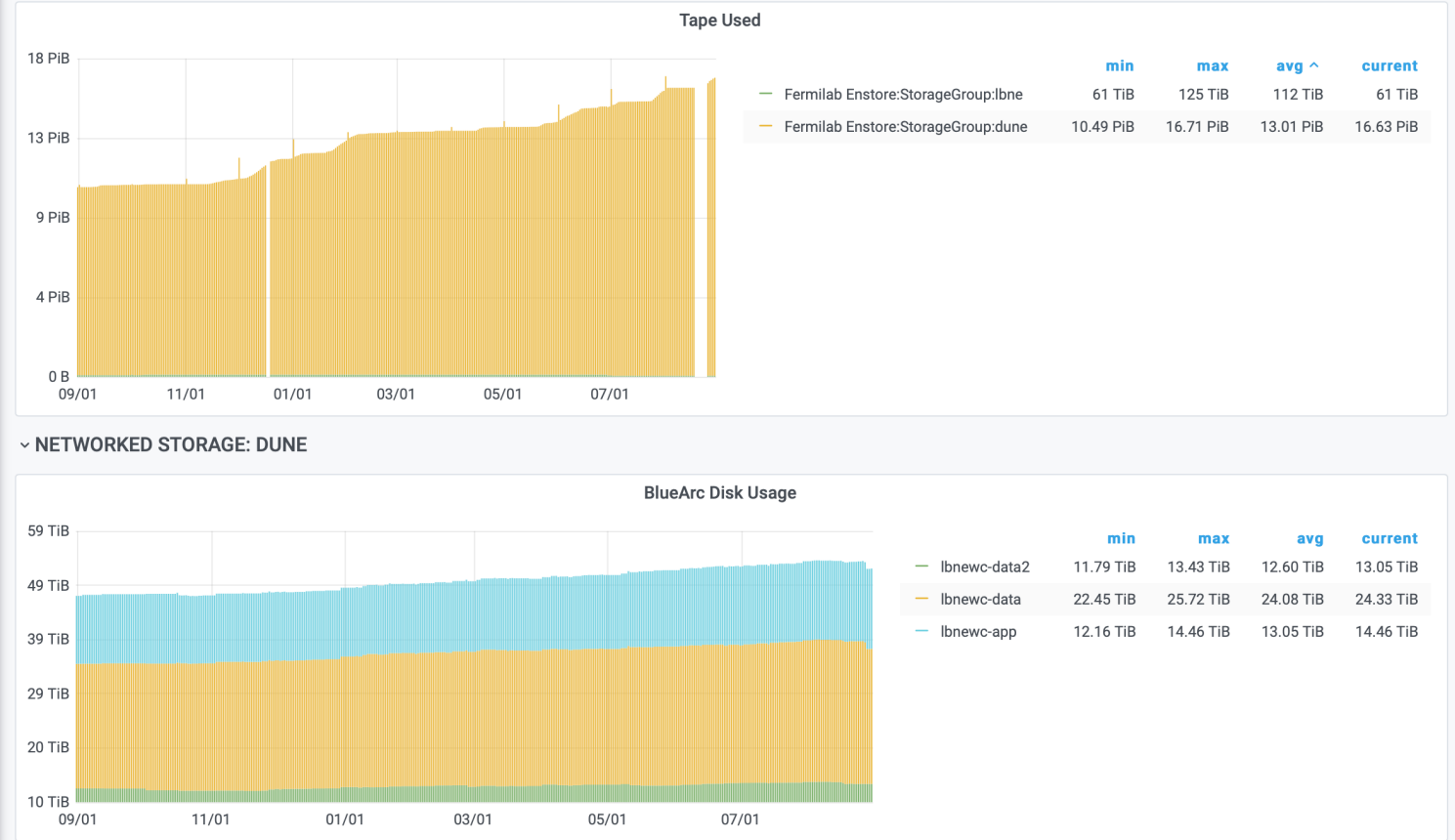


	max	avg
Production Onsite	15.09 K	959
User Onsite	15.90 K	4.69 K
Allocation	4.00 K	3.76 K
Production OSG	26.81 K	1.83 K
User OSG	11.08 K	1.04 K

Last 2 years Memory use

Last 2 years, CPU request use

Storage at FNAL



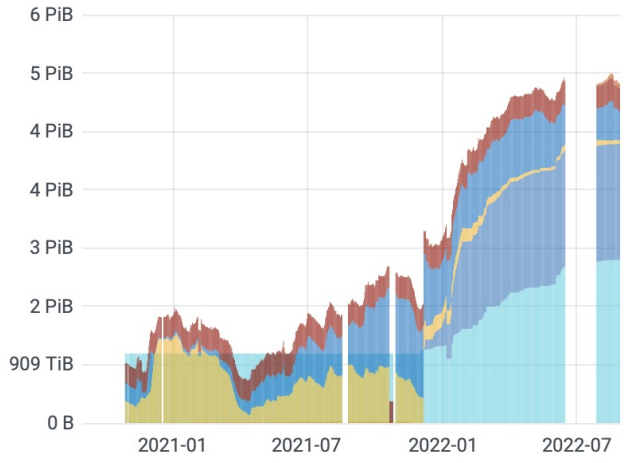
DEEP UNDERGROUND NEUTRINO EXPERIMENT

Storage at FNAL

SCIENTIFIC DATA STORAGE: DUNE

Max Disk Space Used (last 90 days)	New Tape Space Used	Total Data Transferred from dCache	Total Data Transferred to dCache
886 PiB	10.6 PiB	61.2 PiB	26.7 PiB

Disk Space Used by Pool Group

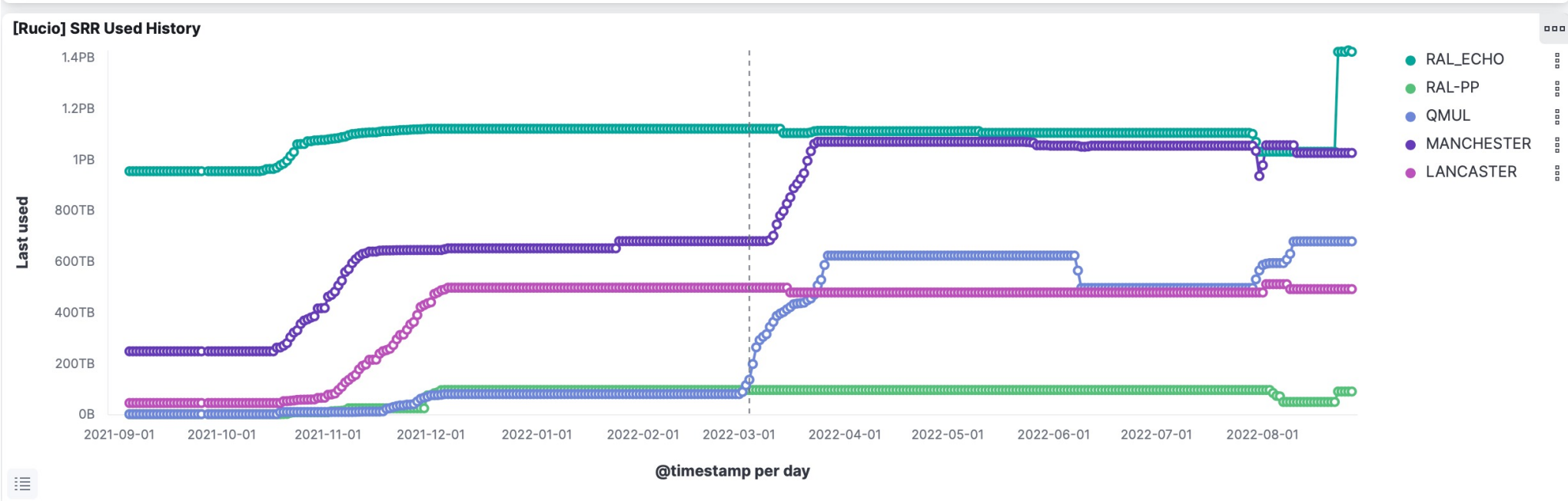


Fermilab Public dCache:StorageGroup:dune_readWritePools	11 MiB	1.369 PiB	487 TiB	55 TiB
Fermilab Public dCache:StorageGroup:dune_PublicScratchPools	172 MiB	1.233 PiB	568 TiB	447 TiB
Fermilab Public dCache:StorageGroup:dune_DuneAnalysisPools	335 MiB	374 TiB	344 TiB	355 TiB
Fermilab Public dCache:StorageGroup:dune_DunePhysicsPools	941 KiB	67 TiB	6 TiB	67 TiB
Fermilab Public dCache:StorageGroup:dune_CdfWritePools	307 GiB	13 TiB	8 TiB	
Fermilab Public dCache:StorageGroup:dune_SlowPublicScratchPools	63 GiB	10 TiB	851 GiB	8 TiB
Fermilab Public dCache:StorageGroup:dune_Geant4ReadWritePools	1 KiB	9 TiB	3 TiB	692 GiB
Fermilab Public dCache:StorageGroup:dune_SlowReadWritePools	1 KiB	9 TiB	3 TiB	692 GiB
Fermilab Public dCache:StorageGroup:lbne_DuneAnalysisPools	3 MiB	3 TiB	3 TiB	3 TiB
Fermilab Public dCache:StorageGroup:dune_NovaPrestagePools	246 GiB	484 GiB	316 GiB	307 GiB
Fermilab Public dCache:StorageGroup:dune_HsmWritePools	225 MiB	483 GiB	128 GiB	238 GiB

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Moving Data to Europe

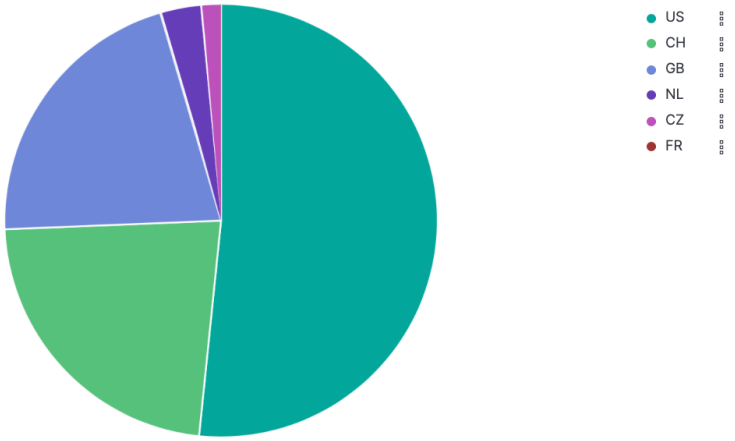
FNAL, BNL, Edinburg + OSU testing



6,216,645
Total replicas

21.17PB
Total bytes

[rucio] Replicas pie by country

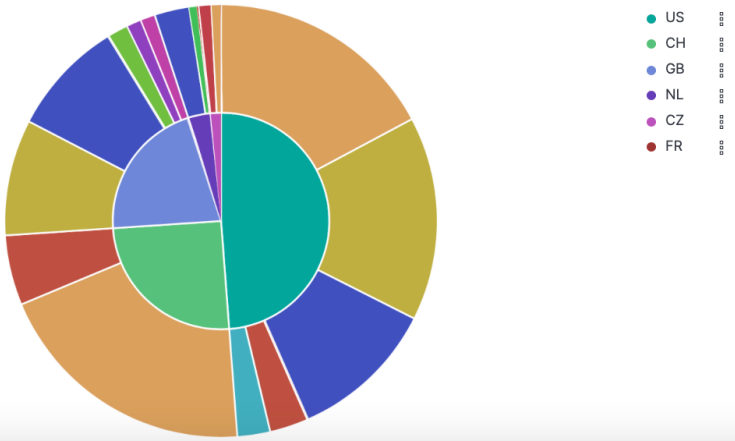


[rucio] Replicas by country

Export

RSE	Replicas	Total bytes
US	3,040,632	10.33PB
CH	1,339,415	6.0PB
GB	1,242,479	3.16PB
NL	177,176	399.04TB
CZ	89,079	511.38TB
FR	1,309	1.72TB

[rucio] Scopes by country



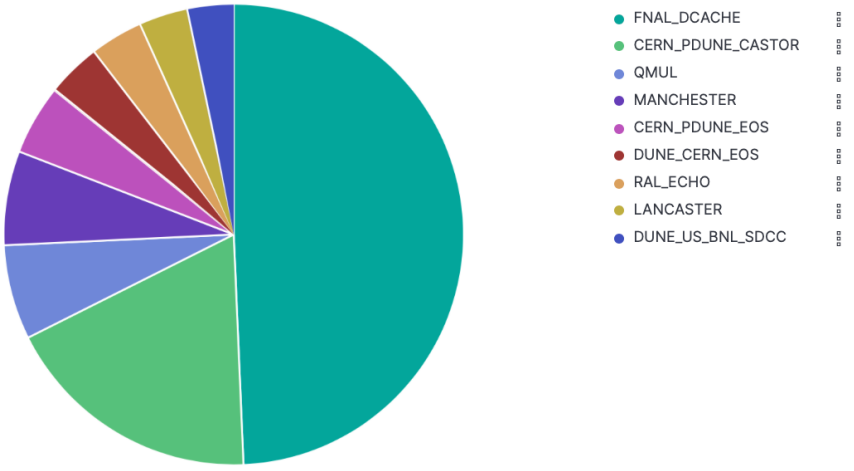
[rucio] Replicas per site

Export

RSE	Replicas	Total bytes
QMUL	383,418	806.91TB
MANCHESTER	381,433	968.05TB
CERN_PDUNE_EOS	282,793	975.33TB
DUNE_CERN_EOS	217,976	626.0TB
RAL_ECHO	214,746	811.42TB
LANCASTER	198,142	484.92TB
DUNE_US_BNL_SDCC	191,773	499.18TB
NIKHEF	177,176	399.04TB

1 2 3

[rucio] Replicas pie



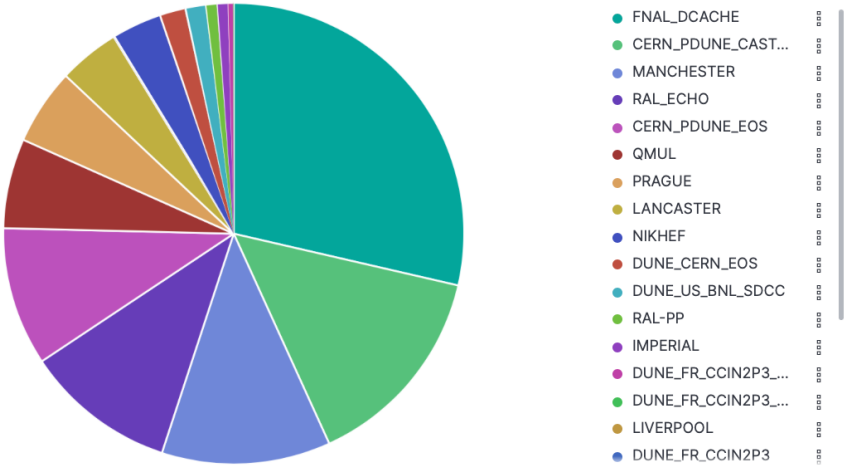
DUNE RSE USAGE

Export

RSE	Total Allocati...	Used	Free	Free(%)
PRAGUE	1.13PB	511.3TB	558.9TB	54.583%
MANCHESTER	1.08PB	1PB	46.3TB	4.728%
RAL_ECHO	1.0PB	1.4PB	-384.9TB	-42.321%
QMUL	1.0PB	679.9TB	291.1TB	32.006%
NIKHEF	981.02TB	399TB	529.3TB	59.324%
DUNE_ES_PIC	719.94TB	1.1GB	654.8TB	100%

1 2

Rucio used space per site pie



Short term efforts

- Test Rucio/Metacat/DataDispatcher for ProtoDUNE data taking
- Process data from coldboxes using new production interfacedto Rucio/Metacat/DataDispatcher
- Work on HDF5 read/write/sim
  - UK groups will test the Workload Manager
- Improve monitoring
- Configuration and slow controls → Conditions DB
- Integrate offline calibrations
- Support ProtoDUNE II running and analysis

Long term efforts

- HPC integration
- Framework improvements
- Continue to integrate hardware databases
- Stay on top of OS/authentication/Spack evolution
- Work with DOE/NSF for **multi-decade** operations program

FTE estimate. Does not include shared facility (storage etc.) costs

- Some effort (mainly operations – pastels at top) can be trained collaboration physicists.
- Rest requires experts
- Currently have around 14 FTE experts (FNAL + collab), all in-kind contributions except UK DUNE funded personnel and DOE.
- Expert need is greatest for ProtoDUNE 2 and pre-operations in 2024-2028.