

DUNE Data Management

Steven Timm / Doug Benjamin / Igor Mandrichenko

DUNE Computing

30 Aug 2022

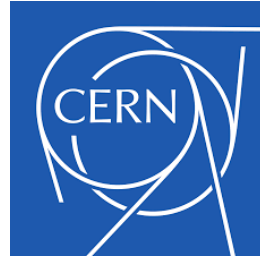
Data Management Structure

- DUNE Data Management group handles both development and operations for DUNE data management.
- Both US and UK contribute to development and operations
 - This US-DOE project covers some of Igor Mandrichenko's work on the various new data management software, which he will present a status on in the next talk.
 - FNAL and BNL staff cover:
 - Operations of services related to data management (B. White, D. Lee, R. Illingworth)
 - The co-leads (Steve Timm, Doug Benjamin)
 - Oregon State students wrote a graphical monitor (currently nonfunctional)
 - GridPP staff—Wenlong Yuan, SE operations and monitoring, James Perry, Rucio development.
- This session will focus on code development but integration, testing, and operations take a lot of time and effort too.

Data Management Interfaces

- ProtoDUNE DAQ group @ CERN
- DQM (data quality management) people
- Physics group coordinators
 - (What data to keep and for how long)
- Production group
 - (Know what output is getting generated, make sure it gets distributed)
- Individual site administrators
- Workflow Management

Data Challenge:

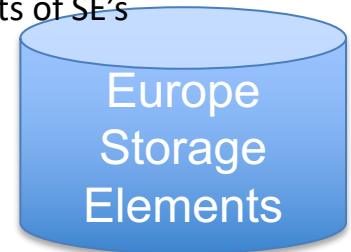


Data from EHN1 to EOSPUBLIC

Declare files

Copy to tape @FNAL and CERN

Distribute data to 3 other sets of SE's



WE ARE HERE



2nd copy of results to RAL and CCIN2P3

Return outputs to Fermilab (some via local writes first)

Reconstruct files streaming at sites without SE's, local for sites with SE's

Activities in past year

- Big picture for last year has been getting ready for 2nd ProtoDUNE beam runs
- Full replacement of legacy file transfer system with new system that uses Metacat for the metadata, Rucio for file transfers throughout.
- Details on this and more in Igor's talk.
- Data Challenge part 1 replicated some DUNE data around the world.
- Data Challenge Part 2 will now try to start processing it using the new workflow allocator and data dispatcher systems.
- After that, will move to make Rucio/Metacat be production system, hopefully before the NP04 beam run.

Progress needed before / during beam

- Another short challenge trying to catch real DAQ output running at full tilt.
- Training users on MetaCat/Rucio clients, especially those actively involved in NP04 beam runs
 - Expect this to be a significant barrier—different than using SAM
- Find faster way to make metadata—current way is a slow shell script that reads the whole file twice.
- Learn how to use Rucio data subscriptions to automate flow of data, especially the outputs from keep-up processing.
- Find out if DQM needs any help from data management.
- Making sure all the fields in SAM are required in MetaCat too plus the new retention fields.
- Finalize the schema to minimize confusion.

What comes after beam in 2023-2025?

- Metadata—Need better way to make it
 - Currently relies on slow shell script that reads the file twice
- HDF5 files, care and feeding:
 - Currently reading them with a POSIX streaming extension to xrootd libs.
 - Can we make it more structured.
 - Can we embed metadata in HDF5 files as we used to do in root?
- Working with huge time-window data like supernova events
 - How to merge, how to stitch together, etc.
 - This has implication on intra-DAQ analysis such as SN pointing too
- Understand the merging workflow that would be necessary if different APA wind up in different physical files
- Can we make a data manifest—which chunk of which trigger record is within which file? Forward and backward indices.
- ND-LAr test beam @ FNAL, how to ingest all that.
- RHEL9 / Centos/Alma/Rocky 9 support

Token Support

- All data challenge and ProtoDUNE beam running still using X.509 proxies to authenticate transfers
- Need to switch to JWT tokens (WLCG tokens a.k.a Scitokens)
- Already pilot jobs are being submitted via tokens to those sites that support it.
- Now need to get all our storage elements and data management middleware to support this too.
- Currently this is pending work by the core Rucio developers, the full token-exchange workflow of Rucio not expected until 2024.
- Preparing an informational talk on this for Sep. collaboration meeting.

Remote Access from HPC

- Everything we are talking about here—databases, Rucio, Metacat, Workflow, etc. all needs outbound network access from worker nodes so jobs can get to it.
- High performance computing worker nodes don't have that.
- There may be significant re-architecting of systems needed to make that work. Time to start thinking about it now rather than when we get our first grant on leadership-class facilities.

Conclusions:

- Integration and operations in DUNE data management has always needed more effort than the development side.
- This needs to be included in planning for this and future projects.
- A lot of data management packages that were vaporware 6 months ago are real and working now.
- There will be significant changes for regular users transitioning to the new systems (SAM->Metacat) and Workflow allocator. We will try to plan the transition so that there is one changeover and not several.
- After ProtoDUNE comes the real fun of dealing with the huge format files and/or merging and stitching smaller files together.
 - It is not clear anyone knows how to build a caching file system that is that capable at the moment. But we have to learn.