



Colorado State University



Brookhaven™
National Laboratory



DUNE Database: Offline Tasks

Ana Paula Vizcaya Hernandez, Lino Gerlach, Norm Buchanan,
Paul Laycock (and FNAL effort is implied !)

US Consortium Startup Workshop - December 14th, 2021



@BrookhavenLab

DUNE Databases effort

- Ana Paula (CSU) fully funded by the FOA (3 years)
- Lino Gerlach (BNL) fully funded by the FOA (3 years)

- Norm Buchanan (CSU) funded by research grants, on sabbatical in 2022
- Paul Laycock (BNL) funded at 25% by FOA, BNL contact for DUNE software and computing
- New hire at U Minnesota (pending) funded by FOA

- FNAL DB team (Vladimir, Steve, Igor, Brandon, Robert)

Part of the discussion is to understand how effort will be distributed in the future

The DUNE Databases Workshop 2021

- A dense but very useful workshop:
 - <https://indico.fnal.gov/event/49543/>
- These tasks are largely a follow up from the actions we wrote up following the workshop
- We expect these to evolve as work really starts to ramp up but this is a good place to start

DUNE hardware database

- The hardware database for construction, and the “Parts identifier” in particular, was undoubtedly the highest priority item coming out of the workshop
- We made significant progress on that this year, converging on requirements, and an implementation plan, this absorbed a lot of effort from BNL and FNAL
- We identified the “physicist in charge” (Jim Stewart, BNL) who will direct the work across the consortia, working with the Database group and consortia contacts to make sure work is coordinated
- Hajime’s talk will cover his related work (using the HWDB) in his talk
- From the offline side we still care about the transition from construction to operations when we will need to transfer some data, and retain its provenance to allow analysis of QA/QC data to understand operations problems

Highest priority for offline - APIs for retrieving all metadata

- Task - Understand what database sources we need to retrieve data from, what the data looks like and how we get it - define the requirements. Define studies of Run 1 data to guide granularity needs for offline processing.
- This is the broad starting point, we need to understand the big picture of the many sources of data that needs to be retrieved from databases to enable data processing. This will likely be the focus of Database effort for at least the first year.
- Our proposal is that Ana Paula and Lino work together with Jake to ramp up on the basics of looking at ProtoDUNE data. This initial task will spawn several child tasks and we can coordinate who looks at what to avoid duplication of effort but exploit the synergies.

Some ideas for institute topics:

DAQ - CSU

- Maintain DAQ (Run configuration) access tools.
- A lot of work was done here already by CSU for ProtoDUNE, continue to follow this up through ProtoDUNE II, and undoubtedly add other sources to the CSU list.

Some ideas for institute topics: Slow Controls (BNL)

- Voltages, currents and other slow controls and SC-like data (e.g. LAr temperature), operations generally needs much finer granularity than offline data processing and experience shows this can cause issues if the use cases aren't separated.
- Study TimescaleDB for Slow Controls (already used at BNL for Belle II)
- Address questions like:
 - Can we use the SC system to write both online archive and offline granularity data?
 - How do we separate and support the different use cases: online needs 24/7 support for writing and should only be exposed to expert online reading needs. Offline needs are different.

The big list of database sources (David Adams talk in the workshop)

Categories

First pass at list of categories (~DB tables)

- [Run configuration](#)
- Event metadata
- Geometry
- Event data file metadata (now in SAM)
- Datasets
- Good run (event?) lists
- [Electron lifetime](#)
- LAr temperature
- Cathode and anode voltages and currents
 - [Unstable voltages](#)
- Detector status (APAs, FEMBs, ...)
- [Channel status](#)
- [TPC CE channel calibration](#)
- [TPC dQ/Dx calibration](#)
- Beam status

- Slowly but surely we need to work through this list.

Other Tasks (not exhaustive)

- Interface to the software framework, piggyback off HSF work on defining best practice to ensure scalability (BNL)
- Database tools and documentation for users (CSU)
- Database design and deployment, separating use cases, including keeping connections to HWDB (Consortia)

Other Contacts

- Identify simulation expert and invite to Database group meetings to tell us about simulation needs (Paul)
- Reach out to the DQM team, understand the link to offline DQ (Norm)
- Reach out to the DRA group, ensure databases are a consideration when assessing physics performance - not-per-trigger-record ! (Ana Paula)
- Work with CERN ProtoDUNE groups on the ground and support the critical “Data Prep” work of David Adams (Lino)

Summary

- In summary, we have work to do !
- Different timezones are a challenge, but we're a small group
 - Regular meetings of the people involved in addition to Database group meetings and meetings of this consortia
- Expect things to evolve quickly at first as effort ramps up, hopefully this is enough material to start a discussion