Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

HSF Report on DUNE framework requirements

Paul Laycock, (Heidi, Andrew and Kirby are way less guilty)

August 30th, 2021

@BrookhavenLab

DUNE Framework requirements

Recap of the plan:

- Write a requirements document based on physics use cases
  - https://docs.dunescience.org/cgi-bin/private/ShowDocument?docid=21934
- Get feedback from the HSF
  - https://indico.cern.ch/event/1038551/timetable/?view=standard
- Make an implementation plan
  - You are here

Requirements document (reminder)

DUNE members - David Adams (BNL), Adam Aurisano (U. Cinc), Chris Backhouse (UCL), Mary Bishai (BNL), Claire David (York), Tom Junk (FNAL), Tom LeCompte (ANL), Chris Marshall (LBL), Brett Viren (BNL)

Advisors - Brian Bockelman (Madison), Chris Jones (FNAL), Kyle Knoepfel (FNAL), Liz Sexton-Kennedy (FNAL), Vakho Tsulaia (LBL), Peter Van Gemmeren (ANL)

Converged on outstanding items and added an executive summary highlighting the 43 requirements in broad categories:

- Configuration requirements
- Concurrency and Multithreading
- Reproducibility and provenance
- Random numbers, machine learning and conditions
- Data and I/O layer
- Memory management
- Physics analysis

HSF feedback process

- First request from HSF was to annotate the requirements document to provide some context for non-DUNE people
  - The document was updated to provide an introduction

- HSF coordination very much liked the idea of giving feedback
  - Scope was defined early on - no recommendation for a specific solution, rather focus on ensuring requirements are complete and well understood

- HSF frameworks group convened a task force of framework experts from various experiments and fields in a two-day workshop in early June

HSF report

General comments

Missing requirements

Requirements that need clarification

Overly constraining requirements

Novel requirements

Further items for DUNE’s consideration

Appendix on prior experience from other experiments

Missing requirements

- Timescale - why now?
  - We should make a timescale and focus on the next 5 years
- Backwards compatibility
  - Will we be constrained by ProtoDUNE?
- Are there constraints from online?
  - Considered outside the scope of the requirements doc, but…
- Programming languages
  - Important if a lot of interoperability is required
- HPC usage
  - We should be specific about use cases to pin down framework requirements

Requirements that need work

- Specify the use cases more quantitatively (be complete, especially wrt memory), and specify which of those require a framework
- One framework for both performing core reconstruction and analysis is ambitious, even if there are examples. Focus more on a component model
- Reproducibility is a common problem, bit-wise reproducibility is generally considered impossible in practical situations. With that in mind, specify what the real reproducibility requirements are

Novel requirements

- Multi-node processing from laptop to HPC scales is not something current frameworks can do well

- Overlapping processing atoms is not common in current frameworks (it is usually assumed that the atoms are independent and there is no need to care about correlations, c.f. APAs and supernovae)

- Fluid data-processing hierarchy (trigger records and slices). If this needs to be supported by the framework, work will be needed

Framework meets production system

- Framework job scheduling has evolved from single-threaded to multi-threaded to multi-node with coordinated communication (MPI)

- Using MPI highlights the need to understand the interface between what the framework does vs what the production system does
  - What does the interface look like?

- The ATLAS Event Service is an interesting example of such an interface
  - I read this as a subtle “hint hint”

What do we do next?

Proposal:

- Focus now on a strategy (timescale)
  - Evolution not revolution, we have something for ProtoDUNE II
    - Maintain the look and feel, evolve the foundations as needed
  - Investigate a component architecture (reco vs analysis)
- Confirm novel requirements are necessary and investigate solutions
  - Be more quantitative about use cases
  - Specify our requirements from an MPI framework standpoint
  - Have a meeting (obligatory requirement)
- Write this up in the CDR