



DUNE/HEP-CCE Roadmap

A.Norman

Roadmap Meeting Nov 2023

Introduction

- Priorities for HEP-CCE work that impacts DUNE are aligned with white paper we generated in June 2023.

Major thrusts are:

- **HPC Facility APIs and Integration**
- **Expand PPS acceleration work beyond wirecell algorithms**
- **Expand data representation & novel storage compatibility of DUNE**
- **Hybrid/Advanced workflow scheduling, provisioning and management**

Goal for DUNE

- DUNE's need is to be able to perform the first stages of signal processing and data reduction @ HPC centers
 - HEP-CCE PPS {
 - Initial signal processing is HIGHLY accelerable w/ GPUs
 - Initial hit find is accelerable w/ GPUs
 - HEP-CCE IOS {
 - Memory footprints are large
(compared to typical grid hardware)
 - Initial and Intermediate data representations are naturally parallelizable along detector regions (APAs)
 - HEP-CCE Workflows {
 - Complex workflows w/ ML are used for reconstruction
 - Event data is reduced by factors of ~100x coming out of initial reconstruction (i.e. going from raw signals high level tracking objects)
 - Going from GB/evt --> MB/evt

Facility APIs

- **HPC Facility APIs and Integration**

- Integration of newer site specific access patterns and protocols into a common API layer
 - Goal is to provide a common layer that CMS/DUNE can use with their global pooling model to add in HPC centers
- Required for provisioning of the resources DUNE needs in a configuration that DUNE needs.
 - need more advanced provisioning than single node slot model of GlideinWMS
 - need methods to “prime” the site for data
 - need methods to manage edge services and local services (think instance servers)
- These are all essentially interfaces that let us setup the complex workflows components

PPS Aligned Work

- Expand PPS acceleration work beyond wirecell algorithms
 - “Cookbook” deliverable (from DUNE perspective) are intended to port more core algorithms according to standard patterns.
 - Provide templates for collaboration effort to expand on.
- Target is the LArSoft algorithms layer
 - Integrates with framework decisions
 - essentially “porting” of core concepts/algorithms to accelerated variants
 - Goal really is to target different “classes” of algorithm patterns so there are representative examples to build off
- May [must] require thought about framework processing model and data representation

Storage Aligned Work

- Expand data representation & novel storage compatibility of DUNE
 - Includes further HDF5 work
 - Exploration of RNTuple representations
 - Continue exploring novel data stores and transient data exchange technologies
- Goal is to solidify the HDF5 representation and understand how this interacts with the PPS algorithms
 - Then benchmark an RNTuple cast of the data to understand how it would interact w/ algorithms
- Novel datastores are work on how to handle intermediate representations and data sharing across complex workflows
 - Feeds into the ML portions of reco and data reduction from the infrastructure side

Complex Workflows

- Hybrid/Advanced workflow scheduling, provisioning and management
 - DUNE workflows are evolving to be less linear in flow and external tool use than previous generations of neutrino analysis.
 - Natural parallelism
 - AI/ML classification call outs
 - Highly asymmetric data passing between stages
 - ...
- The overall workflows naturally want to map to more complex arrangement of resources (i.e. inference servers, pipelined stages, splitting of evt data, intermediate data stores)
- Need to have a workflow/resource provisioning engine that can manage this
 - Natural overlaps w/ what HEPCloud was intended to do
 - But doesn't exist yet today except as one-off demonstrations