HEP-CCE: Fine-grained I/O and Storage (IOS)
Planning for Year 2

Peter van Gemmeren, Rob Ross

Year 2 Plans (from proposal)

4th quarter: Decide on optimization targets for memory infrastructures for phase 2.
- For example explicit synchronous/asynchronous CPU-GPU data transfers vs unified GPU/CPU memory architectures.

Phase 2: Prototyping

- From proposal: *We will prototype EDMs for the different storage solutions for reading and writing and the different memory architectures identified in phase 1 using our synthetic benchmarks for profiling*

Priority: Continuing Activities

- **Darshan for ROOT I/O in HEP workflows on HPC**
  - Study usage patterns from ATLAS, CMS and DUNE
  - Improve analysis capabilities
  - Start looking in depth at ROOT 7 / RNTuple?

- **Investigate HDF5 as intermediate event storage for HPC processing**
  - Prototyping framework to store ROOT serialized event data in HDF5

- **Developing testing/emulation framework**
  - To mimic I/O behavior of generic HEP workflows

New Activities
(Subject to effort resources)

- **Data Model and storage for CPU/GPU**
  - Cross educate first!
  - There are a number of HEP-oriented projects that are either investigating or that provide columnar access to nested data structures including:
    - AwkardArray
    - ROOT DataFrame
    - ATLAS xAOD, PHYS, PHYSlite DAOD
    - CMS NanoAOD
- **In close collaboration with PPS**

People Involved
(sorry if we missed you!)

- **High Energy Physics**
  - Doug Benjamin (ANL)
  - Paolo Calafiura (LBL)
  - Philippe Canal (FNAL)
  - Oliver Gutsche (FNAL)
  - Salman Habib (ANL)
  - Kenneth Herner (FNAL)
  - Patrick Gartung (FNAL)
  - Lisa Goodenough (FNAL)
  - Christopher Jones (FNAL)
  - Liz Sexton Kennedy (FNAL)
  - Kyle Knoepfel (FNAL)
  - Peter Van Gemmeren (ANL)

- **(More) High Energy Physics**
  - Tammy Walton (FNAL)
  - Torre Wenaus (BNL)

- **Computer Science**
  - Suren Byna (LBL)
  - Matthieu Dorier (ANL)
  - Rob Latham (ANL)
  - Rob Ross (ANL)
  - Saba Sehrish (FNAL)
  - Shane Snyder (ANL)
  - John Wu (LBL)