Contribution ID: **74** Type: **Oral**

HPC Friendly HEP data model and RNTuple in HEP-CCE

Monday, 11 March 2024 17:50 (20 minutes)

As the role of High Performance Computers (HPC) increases in the High Energy Physics (HEP) experiments, the experiments will have to adopt HPC friendly storage format and data models to efficiently utilize these resources. In its first phase, the HEP-Center for Computational Excellence (HEP-CCE) has demonstrated that the complex HEP data products can be stored in the HPC native storage backends, such as HDF5, after converting them into byte stream serialization buffers. To efficiently leverage the HPC resources including compute accelerators such as GPUs, the storage format has to allow efficient I/O on parallel file systems used on HPC and the data models have to be capable of being offloaded to the GPUs for processing without conversions. In its second phase, HEP-CCE is studying the design and development of the HEP data models that will be HPC friendly and relevant for the future HEP experiments. At the same time, ROOT, an open data analysis framework, widely used by the HEP community, has been developing a new I/O subsystem called ROOT::RNTuple. RNTuple optimizes performance and minimizes storage, which requires a more streamlined design than the current I/O subsystem (ROOT::TTree) and hence has limited support on data model complexity. When designing data models suitable for offloading to compute accelerators, we also consider their storage in both HPC native backends (such as HDF5) and the more typical HEP persistence in ROOT::RNTuple. Both offloading and storage technologies have different restrictions to construct HEP data models. Only those data models that can take these restrictions into account can be truly HPC friendly and fulfill the requirements of future HEP experiments (including processing using grid resources). In this paper, we will show our results and ongoing works related to data model design and persistence of future HEP experimental data.

Significance

Implementation and scaling test of I/O of HEP data in HPC friendly storage like HDF5.
Design of HEP data models that are HPC friendly, investigation of persistence in both HPC friendly format and RNTuple

References

https://indico.jlab.org/event/459/contributions/11807/attachments/9286/13474/CHEP2023%20Parallel%20IO.pdf
Amit Bashyal et. al., “Data Storage for HEP Experiments in the Era of High-Performance Computing”, 2022 Snowmass Summer Study, arXiv:2203.07885.

Experiment context, if any

Study targeted for HL-LHC and DUNE era experiments where the role of HPCs will further grow.

Primary author: BASHYAL, Amit

Co-authors: KNOEPFEL, Kyle (Fermi National Accelerator Laboratory); BHATTACHARYA, Meghna (Fermilab); VAN GEMMEREN, Peter (Argonne National Laboratory (US)); SEHRISH, Saba (Fermilab)

Presenter: BASHYAL, Amit

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research