Contribution ID: **587** Contribution code: **contribution ID 587** Type: **Oral**

Evaluating awkward arrays, uproot, and coffea as a query platform for High Energy Physics data

Thursday, 2 December 2021 11:00 (20 minutes)

Query languages for High Energy Physics (HEP) are an ever present topic within the field. A query language that can efficiently represent the nested data structures that encode the statistical and physical meaning of HEP data will help analysts by ensuring their code is more clear and pertinent. As the result of a multi-year effort to develop an in-memory columnar representation of high energy physics data, the numpy, awkward arrays, and uproot python packages present a mature and efficient interface to HEP data. Atop that base, the coffea package adds functionality to launch queries at scale, manage and apply experiment-specific transformations to data, and present a rich object-oriented columnar data representation to the analyst. Recently, a set of Analysis Description Language (ADL) benchmarks has been established to compare HEP queries in multiple languages and frameworks. In this paper we present these benchmark queries implemented within the coffea framework and discuss their readability and performance characteristics. We find that the columnar queries perform as well or better than the implementations given in previous studies.

Significance

This paper contains new results not given in any incremental or project status update that also present a fair and accurate performance evaluation of this framework on open data. The results serve to further justify the viability of scientific python ecosystem tools for HEP analysis.

References

https://arxiv.org/abs/2008.12712

Speaker time zone

Compatible with America

Primary authors: SMITH, Nick (Fermi National Accelerator Lab. (US)); THE CMS COLLABORATION

Presenter: SMITH, Nick (Fermi National Accelerator Lab. (US))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research