

Distributed Columnar HEP analysis using coffea + dask

Tuesday 2 July 2024 15:00 (1 hour)

This talk will explore the recent advancements in the Coffea framework, a set of tools and wrappers designed to facilitate columnar Collider High-Energy Physics (HEP) analyses with user-friendly syntax. With the release of AwkwardArray 2.0, the Dask parallel processing library is now a core part of HEP analysis, and provides powerful computing abstractions through the task graphs it produces.

Coffea has been retooled to harness this new infrastructure. By integrating uproot, dask-awkward, and dask-histogram, analyses written using Coffea can be automatically optimized for data transport and distributed across clusters.

This talk will demonstrate the current capabilities of dask-awkward, dask-histogram, and Coffea through an extended Jupyter-based tutorial. Attendees will see practical examples showcasing the functionality of these tools and learn how to use them to develop their own analyses. The basics of task-graph building and lazy evaluation will be covered, along with tips and caveats for migrating analysis code from Coffea 0.7 to this enhanced toolset.

Authors: KROMMYDAS, Iason (Rice University (US)); GRAY, Lindsey (Fermi National Accelerator Lab. (US)); SMITH, Nick (Fermi National Accelerator Lab. (US))

Presenter: KROMMYDAS, Iason (Rice University (US))

Session Classification: Plenary Session Tuesday