



Contribution ID: 10

Type: **Oral Presentation**

COFFEA - Columnar Object Framework For Effective Analysis

Tuesday, July 30, 2019 2:00 PM (20 minutes)

The COFFEA Framework provides a new approach to HEP analysis, via columnar operations, that improves time-to-insight, scalability, portability, and reproducibility of analysis. It is implemented with the Python programming language and commodity big data technologies such as Apache Spark and NoSQL databases. To achieve this suite of improvements across many use cases, COFFEA takes a factorized approach, separating the analysis implementation and data delivery scheme. All analysis operations are implemented using the NumPy or awkward-array packages which are wrapped to yield user code whose purpose is quickly intuited. Various data delivery schemes are wrapped into a common front-end which accepts user inputs and code, and returns user defined outputs. We will present published results from analysis of CMS data using the COFFEA framework along with a discussion of metrics and the user experience of arriving at those results with columnar analysis.

Primary authors: GRAY, Lindsey (Fermi National Accelerator Lab. (US)); CREMONESI, Matteo (Fermi National Accelerator Lab. (US)); SMITH, Nick (Fermi National Accelerator Lab. (US)); REINSVOLD HALL, Allison (Fermilab); JAYATILAKA, Bo (Fermi National Accelerator Lab. (US)); PIVARSKI, Jim (Princeton University)

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

Session Classification: Computing, Analysis Tools, & Data Handling

Track Classification: Computing, Analysis Tools, & Data Handling