



Contribution ID: 481

Type: **Oral Presentation**

IRIS-HEP Tutorial: Fast columnar data analysis with data science tools (Part 1)

Monday, July 29, 2019 2:00 PM (1h 30m)

In this tutorial session, we introduce the scientific python ecosystem and extensions thereof that have been developed as part of the IRIS-HEP initiative to better fit the needs of particle physicists. This hands-on tutorial will introduce:

- Scientific programming with Numpy and various tools in its ecosystem: SciPy, Pandas, Scikit-Learn, etc.
- Tools to accelerate python when Numpy is not expressive or fast enough: Numexpr, Numba, GPU acceleration, etc.
- How to efficiently get data from ROOT files into this ecosystem via the uproot library
- Tools to deal with non-trivial data structures in columnar array format, such as jagged arrays, arrays-of-struct, etc. via the awkward-array library
- Existing and forthcoming tools to deal with histograms as data structures

Presenters: PIVARSKI, Jim (Princeton University); SMITH, Nick (Fermi National Accelerator Lab. (US))

Session Classification: Computing, Analysis Tools, & Data Handling