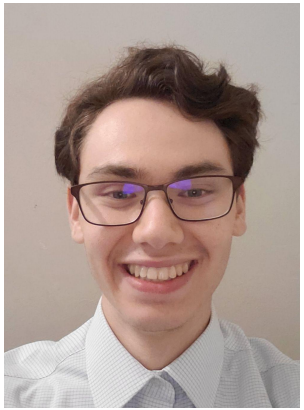
Project Proposal: Development of Experiment-Specific Data Schemas for Coffea



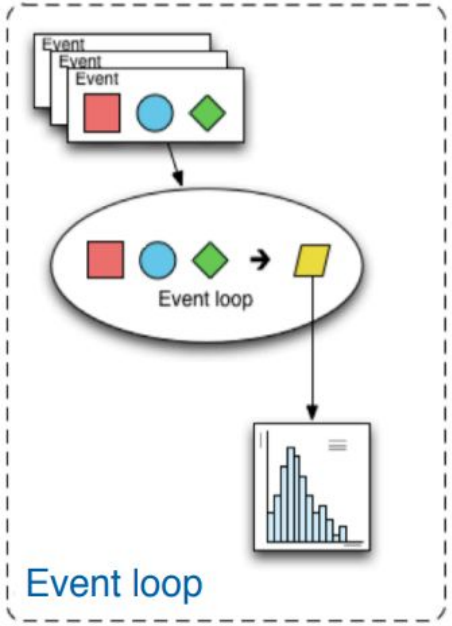
By Sam Kelson

A little about me …

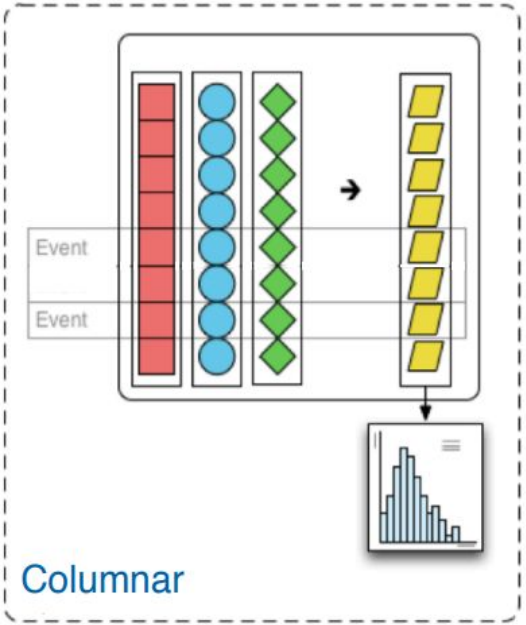
- IRIS-HEP Mentors: Lindsey Gray, Nick Smith, Matthew Feickert, Giordon Stark

Intro: What is Coffea?

- **C**olumnar **O**bject **F**ramework **F**or **E**ffective **A**nalysis



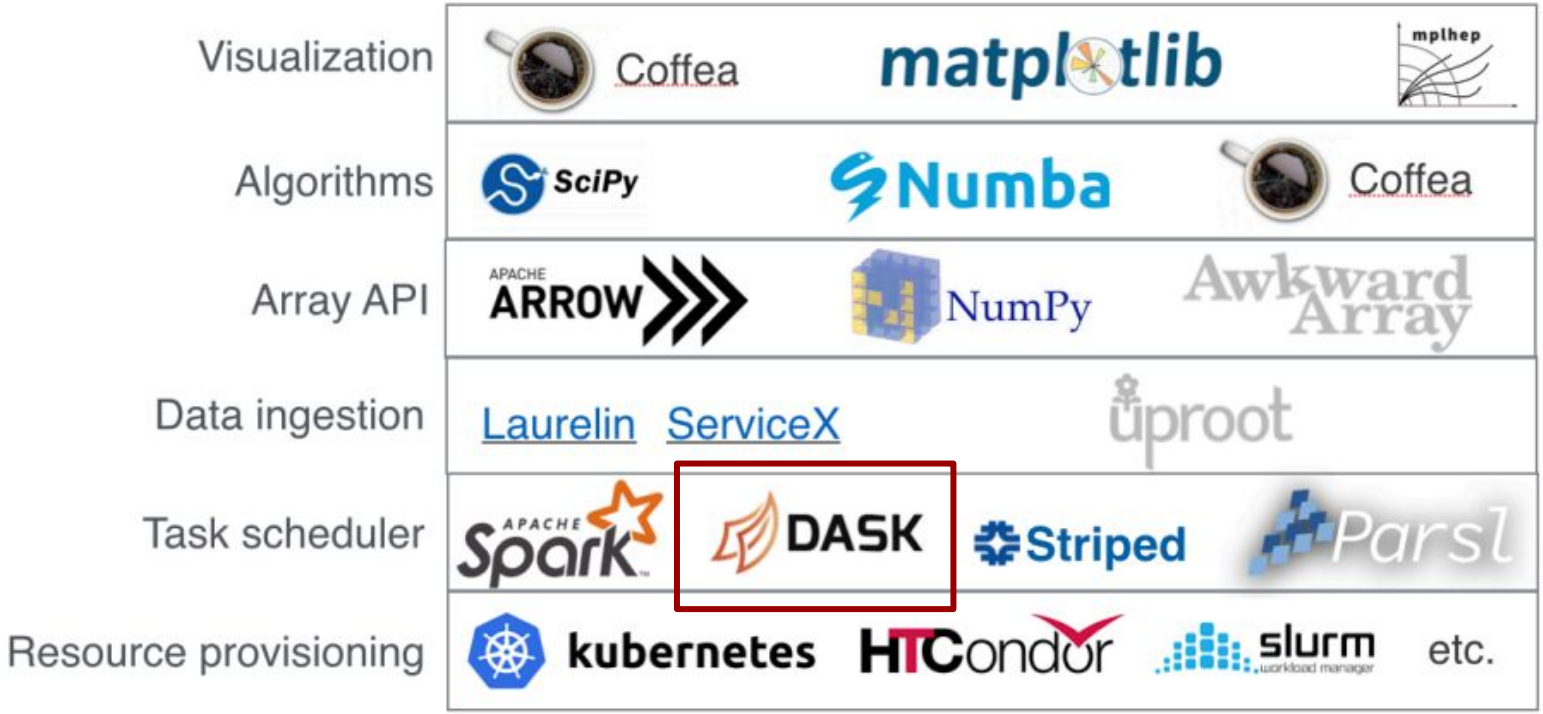
ex: ROOT's RDataFrame



ex: coffea

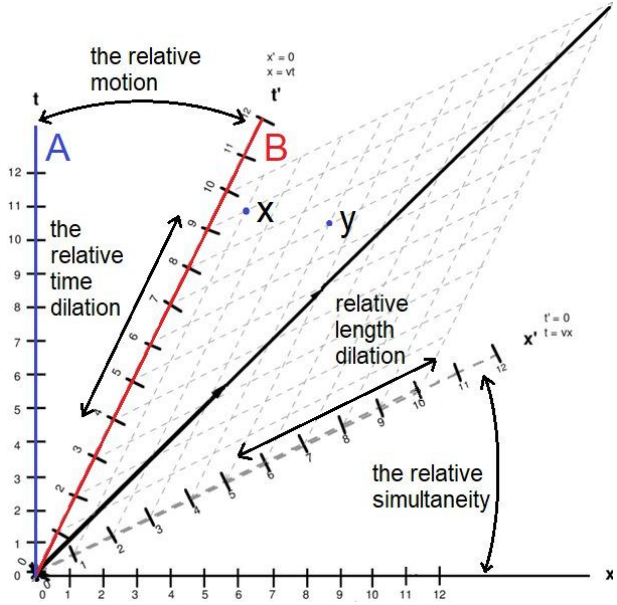
img credit: Nick Smith

Intro: Coffea in the PyHEP Ecosystem

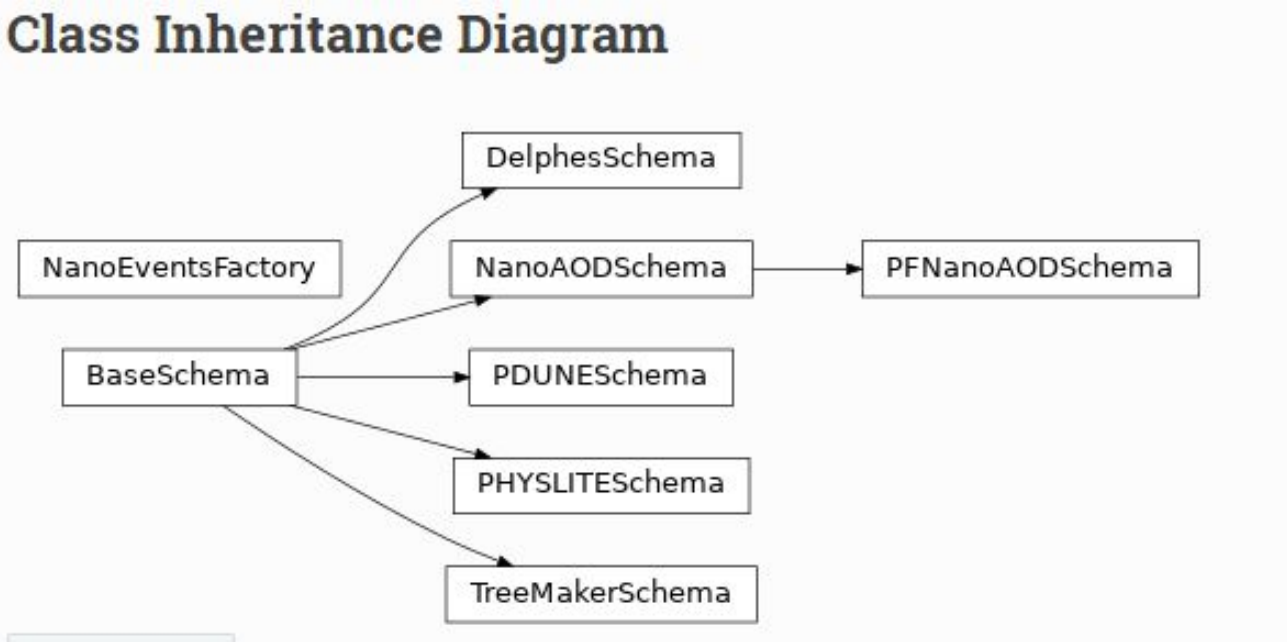


Intro: Coffea uses Awkward Array Behaviors

- A data structure is defined by the information it encodes and the ways in which it can be used.
- Ex: Lorentz vectors



Goal: Experiment-Independent Functionality for Coffea



Conclusion: Project Deliverables

1. Updates to ATLAS PHYSLITE schema
2. Test suite for future data schemas
3. Community reference benchmarks for Coffea scaling
4. Improved public documentation for new developers
5. General schema improvements (time permitting)