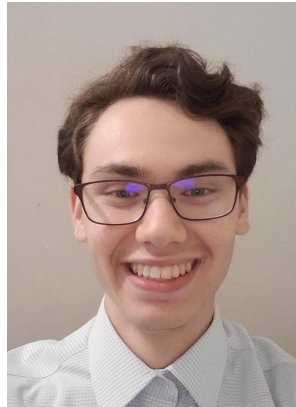


Project Proposal: Development of Experiment-Specific Data Schemas for Coffea



Brandeis
UNIVERSITY



By Sam Kelson



A little about me ...

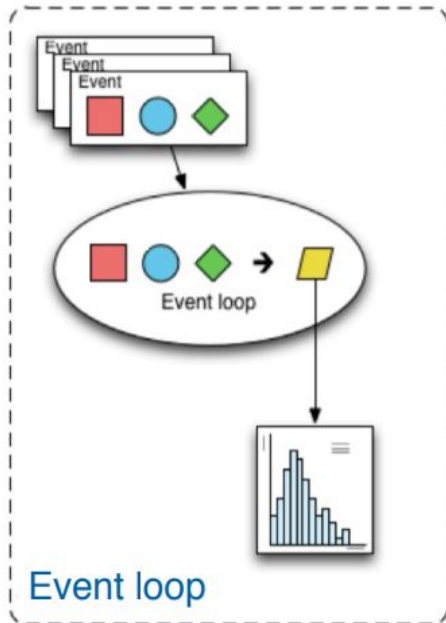


- IRIS-HEP Mentors: Lindsey Gray, Nick Smith, Matthew Feickert, Giordon Stark

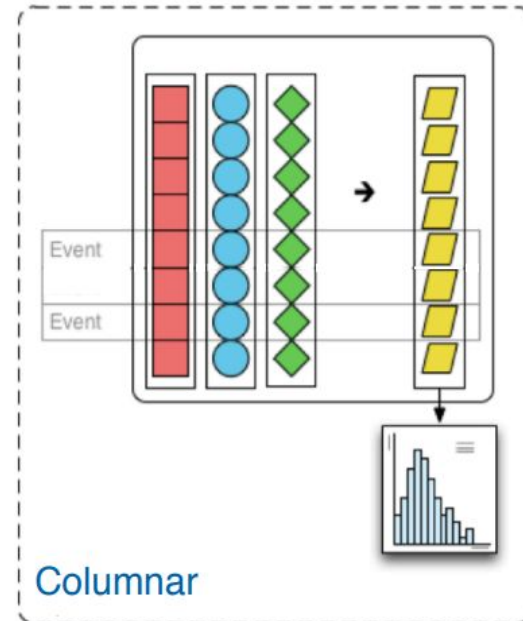
Intro: What is Coffea?



- Columnar Object Framework For Effective Analysis




ex: ROOT's RDataFrame



ex: coffea

Intro: Coffea in the PyHEP Ecosystem

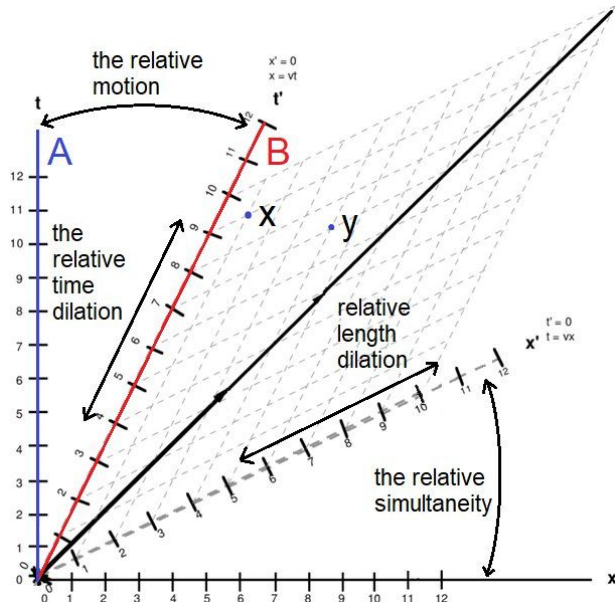


Visualization	 Coffea	 matplotlib	 mphep	
Algorithms	 SciPy	 Numba	 Coffea	
Array API	 APACHE ARROW	 NumPy	 Awkward Array	
Data ingestion	 Laurelin	 ServiceX	 uproot	
Task scheduler	 APACHE SPARK	 DASK	 Striped	 Parsl
Resource provisioning	 kubernetes	 HTCondor	 slurm workload manager	etc.



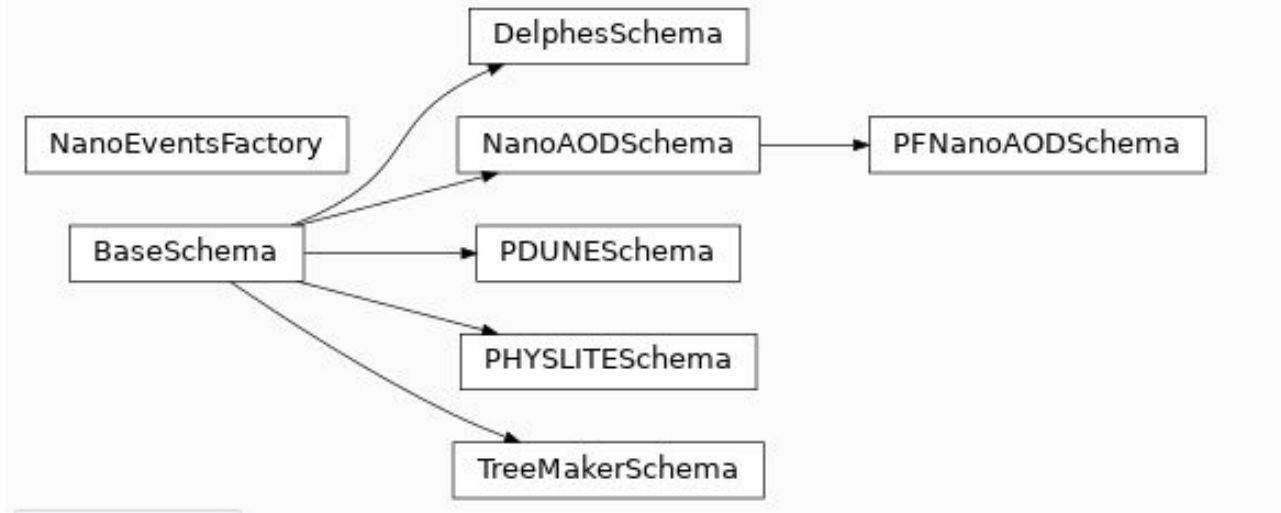
Intro: Coffea uses Awkward Array Behaviors

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





Class Inheritance Diagram



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)