

Lazy Data Loading with "Virtual Arrays" in Awkward

Monday 14 July 2025 10:35 (20 minutes)

High-energy physics (HEP) analyses frequently manage massive datasets that surpass available computing resources, requiring specialized techniques for efficient data handling. Awkward Array, a widely adopted Python library in the HEP community, effectively manages complex, irregularly structured ("ragged") data by mapping flat arrays into nested structures that intuitively represent physical objects like particles and their associated properties. Typically, analyses utilize only specific subsets of these objects and properties, presenting an important opportunity to reduce memory usage through lazy data loading strategies.

In this presentation, we will introduce and delve into Awkward Array's newly developed "Virtual Arrays" feature, explicitly designed for lazy loading of data buffers. Instead of immediately loading entire datasets into memory, Virtual Arrays defer data retrieval from disk until explicitly requested by computation. We will discuss in greater detail the underlying architecture, design considerations, and practical implementation of Virtual Arrays, highlighting their integration into analytical workflows.

We will illustrate how developers and analysts can seamlessly incorporate lazy data loading into their existing frameworks using Coffea—the Columnar Object Framework For Effective Analysis. Coffea facilitates efficient event data processing through columnar operations and transparently scales computations from personal laptops to extensive distributed computing environments without modifications to analysis code. Real-world examples from high-energy physics, including selective data processing and efficient histogramming, will underscore the technical implications and significant performance improvements provided by Virtual Arrays, accelerating data-intensive analysis and enhancing computational efficiency in collider experiments.

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

Co-authors: OSBORNE, Ianna (Princeton University); FACKELDEY, Manfred Peter (Princeton University (US))

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

Session Classification: Talks