Contribution ID: **10** Type: **Notebook talk**

Awkward Target for Kaitai Struct

Wednesday 11 October 2023 17:00 (30 minutes)

Data formats for scientific data often differ across experiments due the hardware design and availability constraints. To interact with these data formats, researchers have to develop, document and maintain specific analysis software which are often tightly coupled with a particular data format. This proliferation of custom data formats has been a prominent challenge for the Nuclear and High Energy Physics (NHEP) community. Within the Large Hadron Collider (LHC) experiments, this problem has largely been mitigated with the widespread adoption of ROOT.

However, not all experiments in the NHEP community use ROOT for their data formats. Experiments such as Cryogenic Dark Matter Search (CDMS) continue to use custom data formats to meet specific research needs. Therefore, simplifying the process of converting a unique data format to analysis code still holds immense value for the broader NHEP community. We propose adding Awkward Arrays, a Scikit-HEP library for storing nested, variable data into Numpy-like arrays, as a target language for Kaitai Struct for this purpose.

Kaitai Struct is a declarative language that uses a YAML-like description of a binary data structure to generate code, in any supported language, for reading a raw data file. Researchers can simply describe their data format in the Kaitai Struct YAML (.ksy) language just once. Then this KSY format can be converted into a compiled Python module (C++ files wrapped up in pybind11 and Scikit-Build) which takes the raw data and converts it into Awkward Arrays.

This talk will focus on introducing the recent developments in the Awkward Target for Kaitai Struct Language. It will demonstrate the use of given KSY to generate Awkward C++ code using header-only LayoutBuilder and Kaitai Struct Compiler.

Author: GOYAL, Manasvi (Princeton University (US))

Presenter: GOYAL, Manasvi (Princeton University (US))

Session Classification: Plenary Session Wednesday