Contribution ID: **30** Type: **Short Talk**

AwkwardForth: accelerating Uproot with an internal DSL

Wednesday, May 19, 2021 6:32 PM (13 minutes)

File formats for generic data structures, such as ROOT, Avro, and Parquet, pose a problem for deserialization: it must be fast, but its code depends on the type of the data structure, not known at compile-time. Just-in-time compilation can satisfy both constraints, but we propose a more portable solution: specialized virtual machines. AwkwardForth is a Forth-driven virtual machine for deserializing data into Awkward Arrays. As a language, it is not intended for humans to write, but it loosens the coupling between Uproot and Awkward Array. AwkwardForth programs for deserializing record-oriented formats (ROOT and Avro) are about as fast as C++ ROOT and 10–80× faster than fastavro. Columnar formats (simple TTrees, RNTuple, and Parquet) only require specialization to interpret metadata and are therefore faster with precompiled code.

Primary authors: PIVARSKI, Jim (Princeton University); OSBORNE, Ianna (Princeton University); DAS, Pratyush (Institute of Engineering and Management, Kolkata); LANGE, David (Princeton University (US)); ELMER, Peter (Princeton University (US))

Presenter: PIVARSKI, Jim (Princeton University)

Session Classification: Software

Track Classification: Offline Computing