

Laurelin: Java-native ROOT I/O for Apache Spark

Wednesday, 19 May 2021 17:53 (13 minutes)

Apache Spark is one of the predominant frameworks in the big data space, providing a fully-functional query processing engine, vendor support for hardware accelerators, and performant integrations with scientific computing libraries. One difficulty in adopting conventional big data frameworks to HEP workflows is the lack of support for the ROOT file format in these frameworks. Laurelin implements ROOT I/O with a pure Java library, with no bindings to the C++ ROOT implementation, and is readily installable via standard Java packaging tools. It provides a performant interface enabling Spark to read (and soon write) ROOT TTrees, enabling users to process these data without a pre-processing phase converting to an intermediate format.

Primary authors: MELO, Andrew Malone (Vanderbilt University (US)); SHADURA, Oksana (University of Nebraska Lincoln (US))

Presenter: MELO, Andrew Malone (Vanderbilt University (US))

Session Classification: Software

Track Classification: Distributed Computing, Data Management and Facilities