Contribution ID: **34**   Type: **not specified**

Integrating ROOT I/O with Apache Spark

Monday 10 September 2018 14:15 (15 minutes)

The DEEP-EST is the European Project building a new generation of the Modular Supercomputer Architecture (MSA). The MSA is a blueprint for heterogeneous HPC systems supporting high performance compute and data analytics workloads with highest efficiency and scalability.

Within the context of the project, we are working on the JVM based implementation of the ROOT File Format, spark-root/root4j, together with an Apache Spark Data Source. Current implementation allows to directly ingest HEP data, perform stream/batch processing and integrate Machine Learning pipelines with Apache Spark.

In this talk, we first discuss the intricacies and internals of the JVM-based implementation. Interesting examples of "bootstrapping ROOT" File Format will be provided as a proof of the robustness and simplicity of the structure of the format itself.

Furthermore, considering Apache Spark constitutes a query execution engine, comparisons of ROOT/c++ based workloads to Apache Spark based ones will be provided and discussed.

Primary author: KHRISTENKO, Viktor (CERN)

Presenter: KHRISTENKO, Viktor (CERN)

Session Classification: Parallelism, Heterogeneity and Distributed Data Processing

Track Classification: Presentations