Contribution ID: **7**   Type: **not specified**

RForest: Evolution of ROOT Tree I/O

Monday, 10 September 2018 16:35 (20 minutes)

The ROOT TTree data format encodes hundreds of petabytes of HEP events. Its columnar layout drives rapid analyses, as only those parts (columns) that are really used in a given analysis need to be read from storage. Its unique feature is the seamless C++ integration, which allows users to directly store their event classes without explicitly defining data schemas.

In this contribution, we present an evolution for future ROOT event I/O. Along with the ROOT 7 interface modernization, we aim for robust, where possible compile-time safe C++ interfaces to read and write event data. On the implementation side, we show a prototype that combines the best of ROOT's current TTrees with recent advances in columnar data formats. A core ingredient is a strong separation of the high-level logical data layout (C++ classes) from the low-level physical data layout (arrays of simple types). We show how a a well-defined low-level data format speeds up serialization and deserialization and facilitates fast vectorized and bulk operations. This lets ROOT I/O run optimally on the upcoming ultra-fast NVRAM storage devices, as well as file-less storage systems such as object stores.

Primary author: BLOMER, Jakob (CERN)

Presenter: BLOMER, Jakob (CERN)

Session Classification: Input and Output of Scientific Datasets

Track Classification: Presentations