



Contribution ID: 113

Type: Talk

## On-the-fly data set joins and concatenations with ROOT RNTuple

Wednesday 23 October 2024 14:24 (18 minutes)

With the large data volume increase expected for HL-LHC and the even more complex computing challenges set by future colliders, the need for more elaborate data access patterns will become more pressing. ROOT's next-generation data format and I/O subsystem, RNTuple, is designed to address those challenges, currently already showing a clear improvement in storage and I/O efficiency with respect to its predecessor, TTree. These improvements provide a solid baseline to introduce extensions that directly target common HENP workflow features not easily achievable before. Notably, many workflows benefit from the ability to join and concatenate data sets during application runtime, with the aim to reduce overall storage requirements and improve application ergonomics. The successful implementation of such compositions requires taking several factors into careful consideration, especially for large data sets that do not fit in memory. These factors include the transparent handling of (in)compatibility between different data sets, the rules that determine how data set compositions are processed, and their effects on runtime performance. In this contribution, we will present the ongoing work to support advanced composition of RNTuple data sets. We will discuss the main design considerations through a selection of concrete workflow use cases, the interfaces and internal machinery that enable the compositions, and an initial set of performance evaluation results.

**Primary authors:** DE GEUS, Florine (CERN/University of Twente (NL)); Dr PADULANO, Vincenzo Eduardo (CERN); BLOMER, Jakob (CERN); CANAL, Philippe (Fermi National Accelerator Lab. (US)); VARBANESCU, Ana-Lucia (University of Twente)

**Presenter:** DE GEUS, Florine (CERN/University of Twente (NL))

**Session Classification:** Parallel (Track 5)

**Track Classification:** Track 5 - Simulation and analysis tools