



Contribution ID: 659 Contribution code: **contribution ID 659**

Type: **Poster**

RNTuple Performance: Status and Outlook

Upcoming HEP experiments, e.g. at the HL-LHC, are expected to increase the volume of generated data by at least one order of magnitude. In order to retain the ability to analyze the influx of data, full exploitation of modern storage hardware and systems, such as low-latency high-bandwidth NVMe devices and distributed object stores, becomes critical.

To this end, the ROOT RNTuple I/O subsystem has been designed to address performance bottlenecks and shortcomings of ROOT's current state of the art TTree I/O subsystem. RNTuple provides a backwards-incompatible redesign of the TTree binary format and access API that evolves the ROOT event data I/O for the challenges of the upcoming decades. It focuses on a compact data format, on performance engineering for modern storage hardware, for instance through making parallel and asynchronous I/O calls by default, and on robust interfaces that are easy to use correctly.

In this contribution, we evaluate the RNTuple performance for typical HEP analysis tasks. We compare the throughput delivered by RNTuple to popular I/O libraries outside HEP, such as HDF5 and Apache Parquet. We demonstrate the advantages of RNTuple for HEP analysis workflows and provide an outlook on the road to its use in production.

Significance

We compare the throughput delivered by RNTuple in typical HEP analyses to popular I/O libraries outside HEP, such as HDF5 and Apache Parquet, demonstrating the advantages of RNTuple for HEP analysis workflows.

References

- <https://arxiv.org/abs/2003.07669>
- <https://arxiv.org/abs/2107.07304>

Speaker time zone

Compatible with Europe

Primary authors: LOPEZ GOMEZ, Javier (CERN); BLOMER, Jakob (CERN)

Presenter: LOPEZ GOMEZ, Javier (CERN)

Session Classification: Posters: Orange

Track Classification: Track 2: Data Analysis - Algorithms and Tools