Contribution ID: **645** Contribution code: **contribution ID 645** Type: **Poster**

Distributed RDataFrame: leveraging Dask and latest optimisations

The declarative approach to data analysis provides high-level abstractions for users to operate on their datasets in a much more ergonomic fashion compared to imperative interfaces. ROOT offers such a tool with RDataFrame, which creates a computation graph with the operations issued by the user and executes it lazily only when the final results are queried. It has always been oriented towards parallelisation, with native support for multithreading execution on a single machine.

Recently, RDataFrame has been extended with a Python layer that is capable of steering and executing the RDataFrame computation graph over a set of distributed resources, requiring minimal code changes for an RDataFrame application to run distributedly. The new tool features a modular design, such that it can support multiple backends - a single interface can be then connected to multiple distributed computing frameworks.

Since v6.24, Distributed RDataFrame has already been included in ROOT as an experimental feature, and it is currently under heavy development. This presentation will show the current performance figures when running real analyses with two different computing frameworks: Apache Spark and Dask. Furthermore, the performance optimisations that are being applied to Distributed RDataFrame will be discussed, namely caching, exploitation of RDataFrame native multithreading and compilation of C++ kernels in the distributed worker processes.

Significance

References

Speaker time zone

Compatible with Europe

Authors: PADULANO, Vincenzo Eduardo (Valencia Polytechnic University (ES)); KABADZHOV, Ivan Donchev (Albert Ludwig University of Freiburg); TEJEDOR SAAVEDRA, Enric (CERN); GUIRAUD, Enrico (EP-SFT, CERN)

Presenter: PADULANO, Vincenzo Eduardo (Valencia Polytechnic University (ES))

Session Classification: Posters: Orange

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