



Contribution ID: 202

Type: Poster

High performance analysis with RDataFrame and the python ecosystem: Scaling and Interoperability

Thursday 27 October 2022 11:00 (30 minutes)

The unprecedented volume of data and Monte Carlo simulations at the HL-LHC will pose increasing challenges for data analysis both in terms of computing resource requirements as well as “time to insight”. Precision measurements with present LHC data already face many of these challenges today. We will discuss performance scaling and optimization of RDataFrame for complex physics analyses, including interoperability with Eigen, Boost Histograms, and the python ecosystem to enable this.

Significance

Performance optimizations in this work are critical to enable higher complexity analyses while maintaining fast turnaround time. Identification of issues and bottlenecks are driving important and ongoing improvements in Root, Numba, and other libraries. Progress and impact on performance are being tracked and incorporated into benchmarking and implementation.

References

<https://indico.fnal.gov/event/23628/contributions/237985/attachments/154987/201732/highPerfAnalysis-May11-2022.pdf>

Experiment context, if any

CMS

Authors: BENDAVID, Josh (CERN); LONG, Kenneth (Massachusetts Inst. of Technology (US))

Presenters: BENDAVID, Josh (CERN); LONG, Kenneth (Massachusetts Inst. of Technology (US))

Session Classification: Poster session with coffee break

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