Contribution ID: **725** Contribution code: **contribution ID 725**   Type: **Poster**

Going fast on a small-size computing cluster

Fast turnaround times for LHC physics analyses are essential for scientific success. The ability to quickly perform optimizations and consolidation studies is critical. At the same time, computing demands and complexities are rising with the upcoming data taking periods and new technologies, such as deep learning.

We present a show-case of the HH->bbWW analysis at the CMS experiment, where we process O(1-10)TB of data on ˜100 threads in a few hours. This analysis is based on the columnar NanoAOD data format, makes use of the NumPy ecosystem and HEP specific tools, in particular Coffea and Dask.

Data locality, especially IO latency, is optimized by employing a multi-level caching structure using local file storage and on-worker SSD caches. We process thousands of events simultaneously within a single thread, thus enabling straightforward use of vectorised operations. Resource intensive computing tasks, such as GPU accelerated DNN inference and histogram aggregation in the O(10)GB regime, are offloaded to dedicated workers. The analysis consists of hundreds of distinctly different workloads and is steered through a workflow management tool ensuring reproducibility throughout the development process up to journal publication.

Significance

We show that fast turnaround times of a few hours can be achieved on only ˜100 CPU threads for a complex frontier physics analysis at the CMS experiment.

This high data throughput is achieved by an efficient combination of multiple modern tools, such as Dask, vectorised operations and SSD data caches. This show-case goes far beyond classical physics analyses and presents a novel way of performing an efficient LHC physics analysis.

References

Speaker time zone

Compatible with Europe

Primary authors: FISCHER, Benjamin (RWTH Aachen University (DE)); NOLL, Dennis (RWTH Aachen University (DE)); FACKELDEY, Manfred Peter (Rheinisch Westfaelische Tech. Hoch. (DE)); ERDMANN, Martin (Rheinisch Westfaelische Tech. Hoch. (DE)); EICH, Niclas Steve (Rheinisch Westfaelische Tech. Hoch. (DE)); DIEKMANN, Svenja (Rheinisch Westfaelische Tech. Hoch. (DE)); RATH, Yannik Alexander (RWTH Aachen University (DE))

Presenter: FACKELDEY, Manfred Peter (Rheinisch Westfaelische Tech. Hoch. (DE))

Session Classification: Posters: Walnut

Track Classification: Track 1: Computing Technology for Physics Research