Contribution ID: **769** Type: **Parallel session talk**

Enhancing CMS data analyses using a distributed high throughput platform

Saturday 20 July 2024 15:04 (17 minutes)

A flexible and dynamic environment capable of accessing distributed data and resources efficiently, is a key aspect for HEP data analysis, especially for the HL-LHC era. A quasi-interactive declarative solution, like ROOT RDataFrame, with scale-up capabilities via open-source standards like Dask, can profit from the "HPC, Big Data and Quantum Computing" Italian Center DataLake model under development. The starting point is a prototypal CMS high throughput analysis platform, offloaded on local Tier-2.
This contribution evaluates the scalability, identifies bottlenecks and explores the interactivity of such platform, on two use-cases: a CMS physics analysis with high-rate triggered events and a study of the CMS muon detector performance in phase-space regions driven by analysis needs, accessing detector datasets. The metrics used to evaluate the scaling and speed-up performance will be reported and results will be discussed, emphasising the differences with the legacy analysis workflows.

Alternate track

I read the instructions above

Yes

Authors: FANFANI, Alessandra (Universita e INFN, Bologna (IT)); BATTILANA, Carlo (Universita e INFN, Bologna (IT)); Prof. BONACORSI, Daniele (University of Bologna / INFN); DIOTALEVI, Tommaso (Universita e INFN, Bologna (IT))

Presenter: DIOTALEVI, Tommaso (Universita e INFN, Bologna (IT))

Session Classification: Computing and Data handling

Track Classification: 14. Computing, AI and Data Handling