Contribution ID: **16**   Type: **Talk**

Analysis Grand Challenge benchmarking tests on selected sites

Wednesday 15 May 2024 11:55 (20 minutes)

A fast turn-around time and ease of use are important factors for systems supporting the analysis of large HEP data samples. We study and compare multiple technical approaches.

This presentation will be about setting up and benchmarking the Analysis Grand Challenge (AGC) [1] using CMS Open Data. The AGC is an effort to provide a realistic physics analysis with the intent of showcasing the functionality, scalability and feature-completeness of the Scikit-HEP Python ecosystem.

I will present the results of setting up the necessary software environment for the AGC and benchmarking the analysis' runtime on various computing clusters: the institute SLURM cluster at my home institute, LMU Munich, a SLURM cluster at LRZ (WLCG Tier-2 site) and the analysis facility Vispa [2], operated by RWTH Aachen.

Each site provides slightly different software environments and modes of operation which poses interesting challenges on the flexibility of a setup like that intended for the AGC.

Comparing these benchmarks to each other also provides insights about different storage and caching systems. At LRZ and LMU we have regular Grid storage (HDD) as well as and SSD-based XCache server and on Vispa a sophisticated per-node caching system is used.

[1] https://github.com/iris-hep/analysis-grand-challenge
[2] https://vispa.physik.rwth-aachen.de/

Requested talk length

Author: KOCH, David Martin (Ludwig Maximilians Universitat (DE))

Presenter: KOCH, David Martin (Ludwig Maximilians Universitat (DE))

Session Classification: HSF