



Contribution ID: 225

Type: Poster

Performances studies for a real time HEP data analysis

Thursday 27 October 2022 11:00 (30 minutes)

In recent years, new technologies and new approaches have been developed in academia and industry to face the necessity to both handle and easily visualize huge amounts of data, the so-called “big data”. The increasing volume and complexity of HEP data challenge the HEP community to develop simpler and yet powerful interfaces based on parallel computing on heterogeneous platforms. Good examples are 1) the pandas framework, which is an open source set of data analysis tools allowing the configuration and fast manipulation of data structures, and 2) the Jupyter Notebook, which is a web application that allows users to create and share documents that contain live executable code. Similarly to the python-based pandas, ROOT::RDataFrame offers another parallel data analysis tool also providing a C++ interface as well as Python bindings (thus compatible with the Jupyter Notebook).

In this contribution we aim to document our experience and performance studies in deploying an HEP analysis workflow, in a realtime analysis fashion, being developed within a Jupyter environment (from the selection criteria to extract the physical signal to the fitting tasks). For this purpose we exploit CMS Run1 Open Data to extract the signal associated with the decay of a beauty meson particle.

We will discuss how the combination of HEP specific tools and technologies coming from the much wider data analysis world may result in a powerful and easy-to-use tool for a HEP data analyst. Among these tools we will test the advantage of offloading some of the most compute intensive tasks on heterogeneous architectures through GooFit, a tool that exploits the computational capabilities of GPUs to perform maximum likelihood fits.

Significance

This contribution provides a performance study of cutting-edge HEP data analysis tools by comparing different approaches to the problem of speeding-up a standard analysis task on an heterogeneous computing platform, thus providing useful advice to the HEP analysts.

References

Experiment context, if any

Four out of five co-authors are CMS members; CMS open data are used.

Primary author: Mr HOANG, Dung (Rhodes College)

Co-authors: DI FLORIO, Adriano (Politecnico e INFN, Bari); POMPILI, Alexis (Universita e INFN, Bari (IT)); SOZ-BILIR, Umit (Universita e INFN, Bari (IT)); MASTRAPASQUA, Vincenzo (Universita e INFN, Bari (IT))

Presenter: SOZBLIR, Umit (Universita e INFN, Bari (IT))

Session Classification: Poster session with coffee break

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