Building examples for Future Circular Collider (FCC) analyses using the Columnar Framework For Effective Analysis (COFFEA) framework and developing the schema class implementation of FCC simulation samples in COFFEA

Prayag Yadav

June 2024

Mentor: David Lange (Princeton University), Co-mentor: Bhawna Gomber (University of Hyderabad)

The Future Circular Collider is a proposed successor to the Large Hadron Collider. It is expected to be significantly larger in size and more powerful than the LHC and is said to be the next discovery machine for standard model as well as beyond standard model physics. There are two proposed stages of development for the FCC: FCC-ee first stage, which is an electron-positron Higgs-EWK factory, and FCC-hh second stage, which will collide hadrons (protons) at $\sqrt{s} = 100\ TeV$.

A columnar approach to analysis is imperative to tackle the challenges of processing huge amounts of data from the FCC. A columnar approach takes advantage of parallel processing and is easily scalable for various computing capabilities. RDataFrame by ROOT interprets TTree data into dataframe tables and enables multithreading and parallel processing. Examples of analyses of FCC simulation samples utilizing RDataFrame are readily available.

The Columnar Framework For Effective Analysis (COFFEA) is another Python framework used for HEP analyses, which enables parallel computing and high scalability using industry-standard frameworks like Dask, Spark, Work Queue, etc. COFFEA has a strong interface with the Python ecosystem and leverages popular data science tools like NumPy, Jupyter, and Matplotlib to provide an easy interface for an effective analysis.

The purpose of this project is to build the necessary schema class for FCC simulation samples in COFFEA and create examples of simple analyses done in COFFEA utilizing FCC simulation samples. The schema classes for the NanoAOD data structure are already present in COFFEA. Developing a schema class for FCC samples would enable analyses to be performed in Python and take advantage of parallel processing with less time-to-insight.

References

1. M. Mangano, 'The future of Collider Physics', 2021. [Online]. Available: https://indico.cern.ch/event/1076616/contributions/4528111/attachments/2348705/4005694/Mangano_Iran.pdf.
2. "COFFEA" N. Smith et al 2020 EPJ Web Conf. 245 06012 and L. Gray et al 2023 J. Phys.: Conf. Ser. 2438 012033
3. FCCSW --- HEP-FCC https://hep-fcc.github.io/FCCSW/, [Accessed 12-06-2024]
4. FCCAnalyses --- HEP-FCC https://hep-fcc.github.io/FCCAnalyses/, [Accessed 12-06-2024]