Coffea-Casa Analysis Facility IRIS-HEP Fellow Proposal

Goal: Contribute to the further development of the Coffea-Casa Analysis Facility (AF) at University of Nebraska-Lincoln (UNL), to expand a gallery of Coffea-Casa analysis samples with existing analysis from CMS adapted to be executed in AF@UNL. I will facilitate the use of Coffea-Casa AF for UNL and Boston University physicists currently working with CMS NanoAOD datasets. My goal will be as well to explore the use of the SkyHook and ServiceX services together deployed at Coffea-casa to be integrated in developed analysis examples.

Motivations: Improving Analysis Facility's stability and efficiency is important for adapting to the new era of HL-LHC with a significant increase of data volume. This past spring we showed proof-of-principle using the measurement of lepton fake rates in simulation as a test case. We propose to now scale this up towards a full analysis. Currently, I am part of a CMS analysis group led by Frank Golf (UNL).

I have some prior analysis experience working on a search for flavor changing neutral current decays of top quarks. I also have some experience with the Coffea-Casa tools and I am becoming familiar with Zora Che's work. I will continue to build on what Zora Che did and help the UNL and BU groups to include these tools in their analyses and provide feedback on the experience and use cases.

Research Plan: I expect to work on this project full time for the summer (equivalent to 3 full-time months). I am excited to join the IRIS-HEP community and work with Frank Golf, Oksana Shadura, Brian Bockelman, and Ken Bloom.

Technically, I'm interested in continuing the work we started customizing dask for parallel computing, the integration of services (ServiceX, Skyhook) and the assessment of performances therein. This project would enhance my own understanding of data organization and delivery system for analysis, as well as supporting coffea users (especially those within my group).

I would pivot my deliverables to lie within the quarterly goals of the Coffea-Casa project covering 12-13 weeks during summer 2021. During this period I will also be a liaison between UNL+BU coffea users and gather feedback for improvement and maintenance.

- Add tools for the application of lepton fake rates in simulation and the studies of their physics performance (May)
- Add tools for the measurement and application of lepton fake rates in data. This adds some additional complexity. For example, to remove contamination in the measurement of the lepton fake rate from prompt leptons (from leptonic decays of W and Z bosons) we perform a templated fit of the transverse mass distribution. (June)
- Expand the above tools to include using the Dask backend from `coffea` on Coffea-casa Analysis Facility@UNL and make performance measurements. Add it in the gallery of examples for the Coffea-casa project. We will also continue to explore other tools such as ServiceX, Arrow Dataset API with SkyHook DM and provide feedback on experience and performance. (July)
- Expand analysis examples using ServiceX and SkyHook services together deployed at Coffea-casa, compare different setups with different combinations of services (e.g. ServiceX versus ServiceX with Skyhook) and deliver performance comparison.. Write documentation, finalize to-do items, prepare final presentation (August)