IRIS-HEP Fellowship Proposal

Storm Lin

June 9, 2021

Proposal:

The IRIS-HEP Analysis Systems focus area has created a number of software tools that allow for new approaches to high energy physics analysis. In particular, these tools include a number of Python packages such as coffea, awkward-array, cabinetry, and uproot, which are used to efficiently interface and work with collider experiment data. These packages are fairly recent and still in active development, so there is testing and improvement that needs to be done. One way to gain useful insight into these new systems, their efficacy, and how they can be improved is to use them to recreate analyses that were previously done using a different analysis system and compare the two.

I am proposing a project of taking existing prototype analysis codes that use these new software tools, modifying them to further test interaction between the coffea and cabinetry packages, and running them on the newly-developed coffea-casa analysis facility. I would create an example analysis that makes use of coffea-cabinetry integration as well as set up and carry out benchmark test comparisons in order to study the performance of these tools. This project will involve working under the mentorship of members of the IRIS-HEP Analysis Systems team. Advisors for this project could include Kyle Cranmer, Alexander Held, Jim Pivarski, Gordon Watts, and/or Oksana Shadura. The data used for the demo analysis will most likely come from the ATLAS Open Data.

The objective of this work is to demonstrate the use of new software tools like coffea as well as new analysis facilities like coffea-casa in replicating the analyses performed by previous systems and to use benchmark tests to examine the relative performance of the prototype system which uses the new framework in comparison to the previous approach. This project builds towards the analysis grand challenge and would aid in the development of existing and new tools and frameworks for high energy physics by testing these projects in a system similar to how they would be used in future experimental analysis work. In particular, it would benefit the prototype-stage coffea, cabinetry, and coffea-casa by acting as a study of their

capabilities and limitations in a way that could lead to improvements that would further the development of these projects. This project would also contribute to the development of uproot, awkward-array, cabinetry, and other HEP analysis software tools.

Proposed Timeline:

Approximately 9 weeks from around June 14, 2021 to around August 13, 2021
Proposed total of 1.5 FTE-months with no other major activities during that time

Week 1-2: Study background and implementation of prototype analysis system. Become familiar with the use of coffea, awkward-array, cabinetry, and other relevant packages and tools needed for the project. Investigate available ATLAS Open Data analyses such as those found at http://opendata.atlas.cern/release/2020/documentation/physics/SL3.html and https://cds.cern.ch/record/2707171.

Weeks 3-5: Start to work with ATLAS Open Data analysis on the coffea-casa analysis facility. Develop or adapt analysis code to use the coffea framework. A possible analysis candidate for this would be https://github.com/atlas-outreach-data-tools/atlas-outreach-cpp-framework-13tev/tree/master/Analysis/TTbarAnalysis.

Week 6-8: Investigate and test existing coffea-cabinetry integration notebooks using the coffea-casa analysis facility. Develop a demo based on ATLAS Open Data analysis showing typical coffea-cabinetry workflow.

Week 9: Open pull requests in relevant github repositories (e.g. coffea-casa-tutorials.git). Document and present results.