AGC-RDF on REANA: A RECAST-like Reinterpretation Workflow with NanoAOD

Fellow: Dmytro Skorubskyi

Project Mentors: Oksana Shadura, Alex Held, Pallabi Das

Project Description

Modern particle physics analyses rely on reproducible, scalable workflows to maximize scientific impact and facilitate reinterpretation. The ROOT RDF Analysis Grand Challenge (AGC) provides benchmarks and reference implementations for high-throughput, declarative data analysis using ROOT's RDataFrame (RDF) interface. Meanwhile, the REANA platform offers a powerful, container-based environment for executing complex workflows at scale. This project aims to integrate the AGC implementations into REANA, enabling community users to run standardized RDF analyses in a reproducible manner. Moreover, we will extend the framework by developing a RECAST-like reinterpretation workflow: users can re-run AGC analyses with modified selections or models to explore new physics scenarios. Finally, the workflow will incorporate an automated NanoAOD production step, converting public OpenData MiniAOD samples into NanoAOD within the same REANA pipeline.

Key objectives:

- Containerize and deploy ROOT RDF AGC code on REANA.
- Design and implement a RECAST-inspired branching workflow to support reinterpretation of AGC analyses.
- Integrate NanoAOD generation from MiniAOD datasets as a preparatory step in the pipeline.

Deliverables

1. **REANA Workflow for ROOT RDF AGC**
    - A fully containerized REANA workflow encapsulating AGC benchmark (e.g., https://github.com/root-project/analysis-grand-challenge/). Users will launch analyses with a single configuration file.
2. **NanoAOD Production Step**
    - A pre-processing job in the REANA pipeline that ingests public MiniAOD samples (e.g., CMS OpenData) and produces NanoAOD files, ready for fast analysis.

3. **RECAST-like Reinterpretation Module**
    - A modular extension to the workflow allowing users to specify alternative selection cuts or physics models. Outputs will include comparative plots and summary tables.
4. **Documentation and User Guide**
    - Comprehensive README, quick-start tutorial, and example use cases. Includes instructions for setup, execution, customization, and interpretation of results.
5. **Summary Presentations**
    - Status and final results slides to be presented at IRIS-HEP collaboration meetings and relevant conferences.

Timeline

Weeks	Goals
1–3	Review AGC implementations and REANA platform. Acquire data and computing credentials.
4–6	Containerize AGC code and basic MiniAOD to NanoAOD conversion tools.
7–9	Implement core REANA workflow for RDF AGC; validate end-to-end execution. Begin documentation.
10–12	Develop RECAST-like branching logic; test reinterpretation scenarios.
13-14	Refine performance and scalability; finalize documentation; prepare presentations.

By deploying ROOT RDF benchmarks on REANA and adding reinterpretation and NanoAOD production capabilities, this project will deliver reusable, scalable workflows that empower both HEP and broader scientific communities to perform reproducible analyses and model tests efficiently.