Analysis Grand Challenge

Mentee: Christina Mondelli (Lewis)
Mentor: Andrew Wightman (University of Nebraska-Lincoln)

Overview

1. Background
   ➔ AGC
   ➔ IRIS-HEP
   ➔ Coffea Casa and Processor
   ➔ Columnar Analysis

2. Purpose

3. Timeline

5. Completed Tasks

6. Results

7. Next steps for the Future

8. Acknowledgements and References
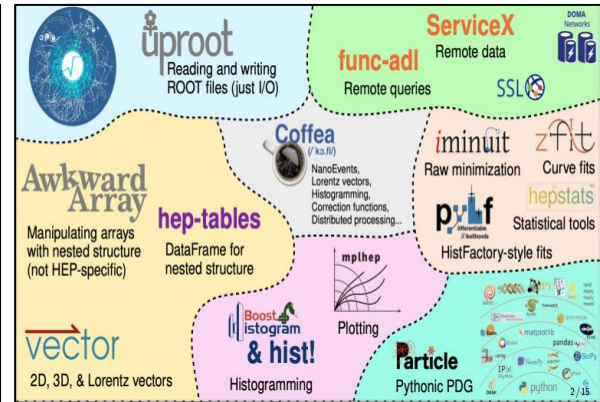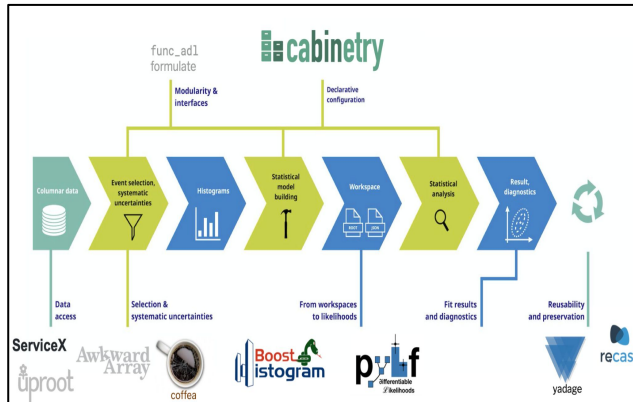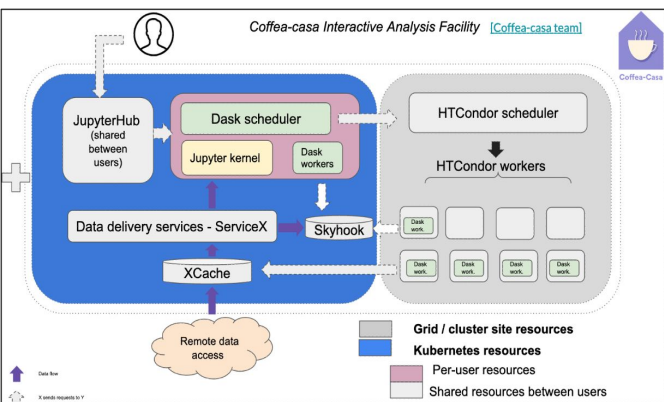
Analysis Grand Challenge

- Executes a High Energy Physics analysis that scales to the data volumes and complexity expected from the High Luminosity Large Hadron Collider
- The High-Luminosity LHC, projected to be operational in 2029, will start a search for discoveries that test the standard model and ask questions about what we know about dark matter
- Higher luminosity could present multiple problems
- IRIS-HEP collaboration is producing software and computing tools meant to help address these challenges

https://news.fnal.gov/2021/02/hl-lhc-accelerator-upgrade-project-receives-approval-to-move-full-speed-ahead-from-department-of-energy/

IRIS-HEP

- Has analysis tools and facilities

https://indico.cern.ch/event/1126109/contributions/4788591/
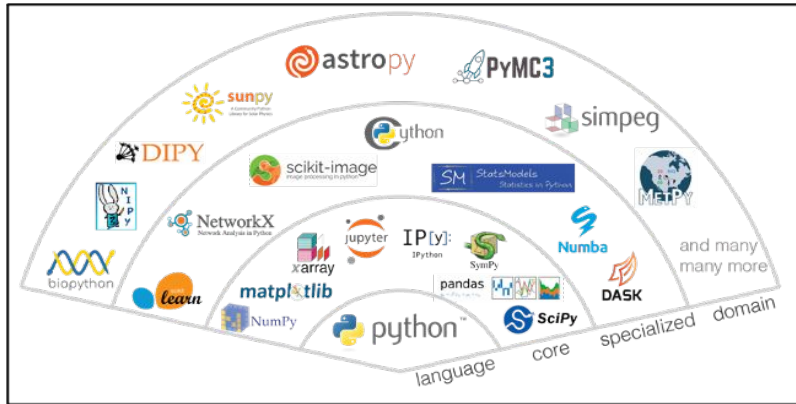
https://iris-hep.org/as.html

https://indico.cern.ch/event/1126109/contributions/4788591/attachments/2430323/4162323/Analysis%20Grand%20Challenge%20Tools%20Workshop%202022.pdf
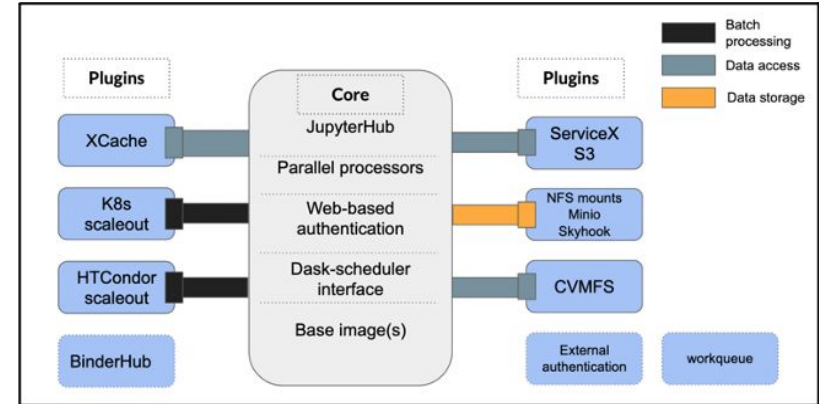
Coffea-Casa

- Can execute columnar data extraction, statistical inference, and turning data into histograms


https://jupytearth.org/jupyter-resources/introduction/ecosystem.html

https://iris-hep.org/projects/coffea-casa.html

Standard Processor

- Executes instructions from an operating system
  - → Events in this high energy physics research is the information that is read out by the detector that is associated with a single LHC beam crossing

- Processes rows and columns of data

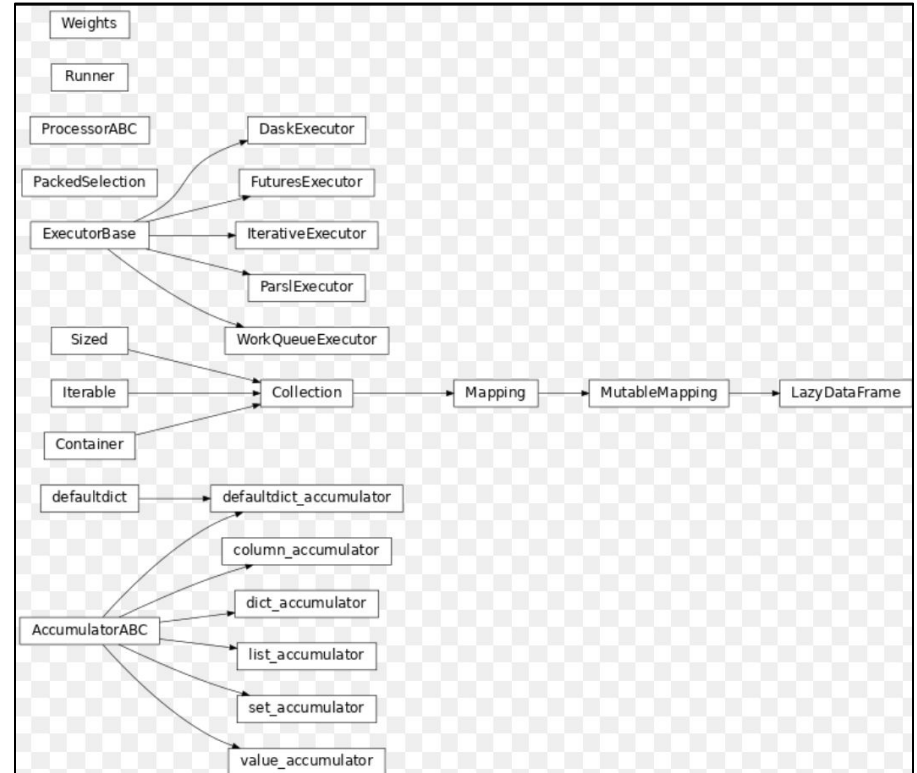https://www.gearnuke.com/what-is-a-cpu
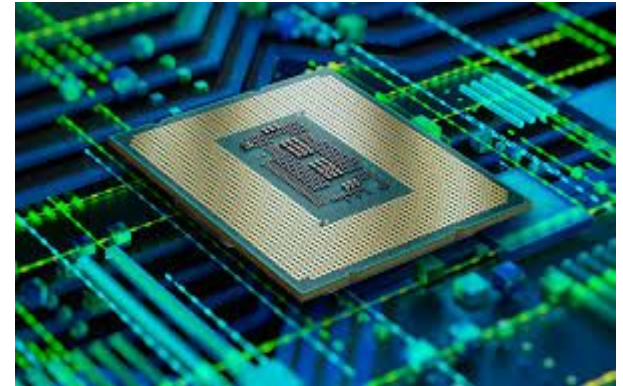
Coffea processor

- Has scale-out built into it which allows for an easier process

- Scale-out allows you to utilize larger CPU resources

- Each subset of events will be worked on by itself

https://coffeateam.github.io/coffea/modules/coffea.processor.html

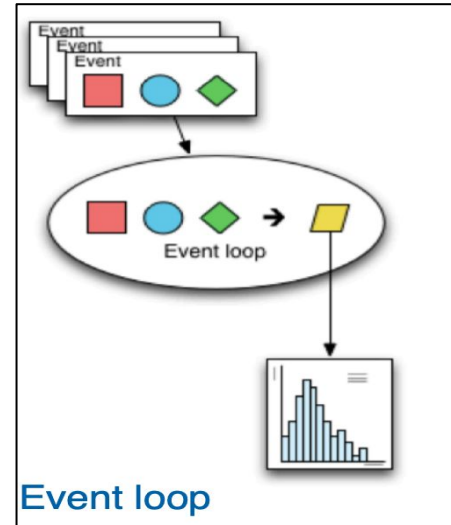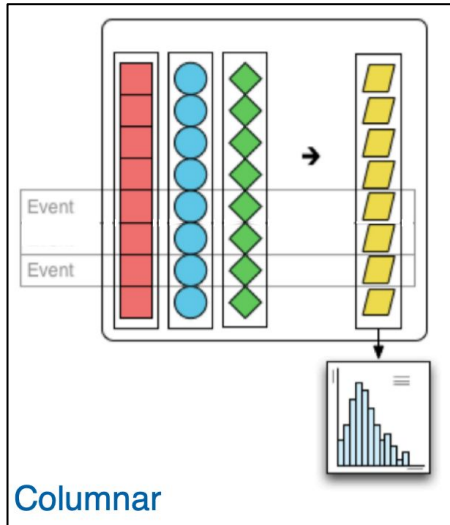9 August 2023   Christina Mondelli | Analysis Grand Challenge

Columnar Analysis

- Uses the basics of array programming
- Calculates quantities and statistics on entire rows of events simultaneously
- Ending result of an interactive environment and fast results while still being able to "scale to the entire dataset"
- There are powerful python libraries that implement array programming

9 August 2023   Christina Mondelli | Analysis Grand Challenge

https://indico.cern.ch/event/759388/contributions/3306852/attachments/1816027/2968106/ncsmith-how2019-co

Objective and Purpose

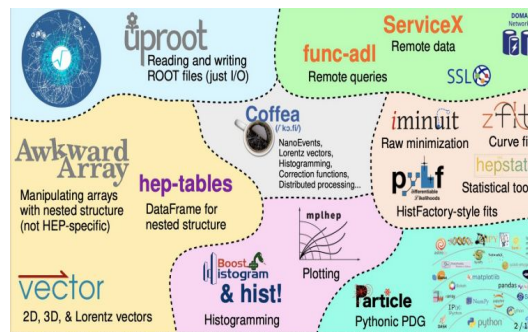- Objective: Convert an AGC open data analysis into one using CMS Run2 data and recipes
- Create a more "CMS-like" example analysis for people to be able to reference when they are trying to take over AGC/IRIS-HEP tools
- How can we create a more "CMS-like" example analysis?
➔ Improve CMS recipes for making corrections

https://iris-hep.org/as.html

https://indico.cern.ch/event/1126109/contributions/4788591/attachments/2430323/4162323/Analysis%20Grand%20Challenge%20Tools%20Workshop%202022.pdf
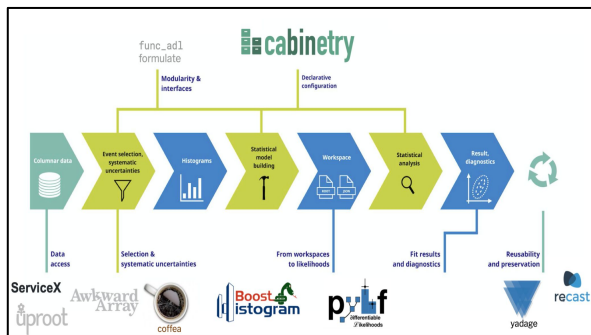
Timeline

Weeks 1 and 2	Computer programming training
Weeks 3 and 4	Account setup and preliminary research
Week 5	Obtain SingleMuon NanoAOD files for Run2018A and turn them into text files
Weeks 6	Start the creation of the python script and check to make sure ttbar-analysis-pipeline executes
Week 7	Build up the script, debug the AGC CMS ttbar notebook and get it running with CMS Run2 data on Coffea-Casa
Week 8	Modify script, fix the histogram output issue
Week 9	Create a new repo in Github, poster due
Week 10	Presentation

Christina Mondelli | Analysis Grand Challenge

Completed Tasks

1. Wrote a python script to convert the text files into a json file that looks like what Coffea-Casa would be expecting to see

2. Got acquainted with the AGC CMS ttbar notebook

3. Debugged and got it running on CMS Run2 data

4. Created a new repo in GitHub and put it in Coffea-casa

Results

- Once the AGC CMS ttbar notebook was running with CMS Run2 data, it produced two histograms
- These histograms are two MC samples
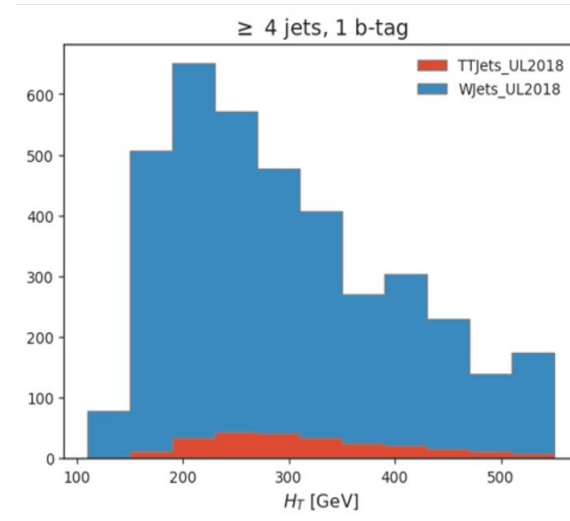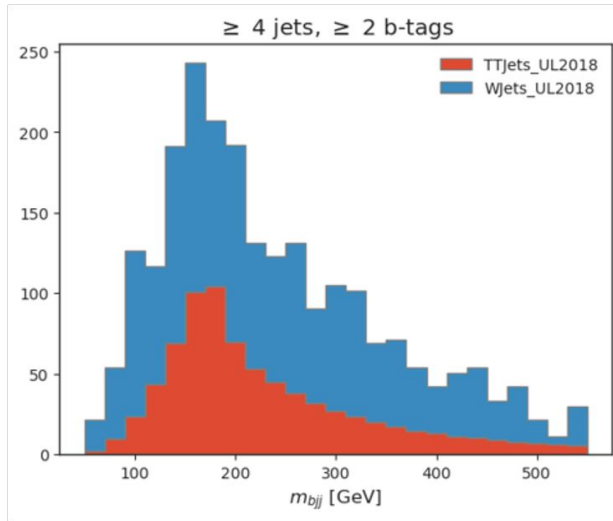
Next Steps for the Future

- Add performance tests
  - ➔ Only read a subset of branches
  - ➔ Timing and event throughout metrics
  - ➔ Benchmark performance of xcache
  - ➔ Efficient run workbook at different analysis facilities (LPC)
  - ➔ Enlarge the portability of AGC tools/setup
  - ➔ Test the portability of AGC tools/setup
- This allows for benchmarking
  - ➔ Comparison of performances
- The goal is to implement each of these tests and see how efficiently they would be able to work

https://www.javatpoint.com/performance-testing

9 August 2023   Christina Mondelli | Analysis Grand Challenge

Acknowledgements

References

Analysis systems. Institute for Research and Innovation in Software for High Energy Physics. (2023, August 2). https://iris-hep.org/as.html

Array Programming (2023) *Wikipedia*. Available at: https://en.wikipedia.org/wiki/Array_programming#Ada (Accessed: 02 August 2023)

Coffea conceptsÁ. Coffea concepts - coffea 0.1.dev1+ga33fc17 documentation. (n.d.). https://coffeateam.github.io/coffea/concepts.html#columnar-analysis

Coffea-Casa. Institute for Research and Innovation in Software for High Energy Physics. (2023b, August 2).
https://iris-hep.org/projects/coffea-casa.html

Ecosystem — Jupyter meets the Earth. (n.d.). https://jupytearth.org/jupyter-resources/introduction/ecosystem.html

High-Luminosity LHC. (2023, July 21). CERN. https://home.web.cern.ch/science/accelerators/high-luminosity-lhc

Home · Indico. (n.d.). https://indico.cern.ch/event/759388/contributions/3306852/attachments/1816027/2968106/ncsmith-how2019-columnar.pdf

Iris-Hep AGC Tools 2022 workshop. Indico. (n.d.). https://indico.cern.ch/event/1126109/contributions/4788591/

Jupyter meets the Earth. Ecosystem - Jupyter Meets the Earth. (n.d.). https://jupytearth.org/jupyter-resources/introduction/ecosystem.html

Leah. (2021, February 11). *HL-LHC accelerator upgrade project receives approval to move full-speed-ahead from Department of Energy*. News.
https://news.fnal.gov/2021/02/hl-lhc-accelerator-upgrade-project-receives-approval-to-move-full-speed-ahead-from-department-of-energy/