ServiceX Biweekly Team Meeting

30 April 2021

Attendees

Present: Ben, Mason, Andrew, KyungEon, Peter, Rob, Gordon, Ilija, Suchandra
Apologies: Oksana Shadura
Convener: Ben

Agenda

- Status of Coffea-Casa deployment - Oksana Shadura (can't attend due to conflict)
    - Coffea-casa is ready to be tested (https://cmsaf-jh.unl.edu)
        - it has 0.7.2 coffea
    - ServiceX for coffea-casa production instance is 99% deployed (I have a small problem with Globus secret).
    - ServiceX with tokens didn't work (it is in development coffea-casa instance),
        - our tokens are not recognised at ServiceX
        - we are debugging with Andrew
- Preparation for vCHEP users - Ben
- Example notebooks - Ben
- CMS AOD status, Template DID Finder project, CERN OpenData DID Finder - Gordon
- Status Update - Andrew
- Status Update - Suchandra
- servicex-for-trexfitter documentation now available for public in ATLAS - KyungEon https://trexfitter-docs.web.cern.ch/trexfitter-docs/advanced_topics/servicex/
    - Crash for over 100 requests
    - Read input files from an active grid site #265

Meeting Notes

- vCHEP Prep
    - Example notebooks
        - Still need: func adl uproot and tcut examples
        - Will have trex fitter and xaod example
        - Ben needs a guide to data sets
    - Need a release of various components
    - Want to get integration testing before then
- ServiceX for trex fitter
    - Crashing when getting over 100 transformation requests

        - Might be due to async i/o using too much memory (crashes with 8GB/works with 32GB)
        - Fronted issue vs backend
- ServiceX DID Library and Demo
    - See https://github.com/ssl-hep/ServiceX_DID_Demo/blob/main/src/demo_did.py for the code that has to be written for a DID finder
    - The library can be found here, and the code that runs all the communication and calls the DID person's code can be found here.
    - This code works under testing, but hasn't undergone integration testing with ServiceX (next steps)
    - Questions:
        - What is adl32? - checksum.
        - There are various branches in DID finder, I hope I've replicated the code correctly.
    - Outstanding Issues
        - Properly parsing did-specific command line arguments
        - Dealing with the multi-threaded architecture of the rucio did finder
    - Next Steps:
        - Test the demo in ServiceX
        - Implement the CERN OpenData DID finder (should be fairly easy)
- CMS AOD status, Template DID Finder project, CERN OpenData DID Finder
    - AOD progress coming along
        - Goal to get 4 lepton mass with all cuts in CMS demo
        - Possible to do this with ATLAS data?
            - No open xaod level data
            - Ntuple data accessible with uproot available
            - A lot easier to do with Run 2 data
            - Open data from ATLAS available, better to use this since it can be put in Rucio open data instance
            - Analysis touches multiple components, not a clean one hour demo
- Status Updates
    - Andrew
        - Planning on getting web dashboard merges in before vCHEP
    - Suchandra
        - Planning on merging logging code after a bit of work with Ilija
        - Will move on to incorporating logging into other repos once cpp transformer logging is completed
- Multiple code generators for ServiceX
    - Ben working on this
- Scalability Testing
    - Prefer logging first
    - But should be ready for scalability testing
    - Try with 10TB sample to compare with prior results
        - David Liu's code has queries against this sample set

            - Gordon runs this but might be broken when run against 10TB sample
        - Target in 2 weeks
- v

New Action Items

- (Date) - (Assignee) - (Description)

Old Action Items

-