

Data & Analysis Preservation: status update

Maxim Potekhin

Nuclear and Particle Physics Software Group



Brookhaven™
National Laboratory

PHENIX DAP Meeting

11/04/2021



Overview

- Maxim has started work on sPHENIX
- HEPData
- Direct photons analysis - REANA

HEPData

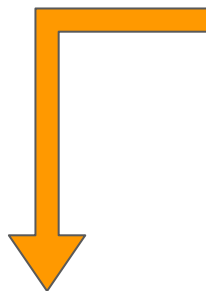
- ppg202 close to final (thanks Gabor)
- ppg212 will be ready in a week
- ppg241, ppg238 in the pipeline (Julia, Sanghwa need to approve)

REANA - sequential part

- Since last report, ran sequential parts of the workflow from two steps of the analysis
 - *Block 1, MB and ERT datasets. Block 2, creating histograms.*

```
version: 0.0.1
inputs:
  directories:
    - ./raw_taxiData
    - ./sim_Pi0Histogram
  files:
    - ./pi0extraction.cc
    - ./WGRatio.cc
    - ./generationRM_Pi0.cc
    - ./Convolution_Pi0.cc
    - ./universal.h
    - ./setup_env.csh
    - ./driver.csh
workflow:
  type: serial
  specification:
    steps:
      - environment: 'registry.sdcc.bnl.gov/sdcc-fabric/rhic_sl7_ext:1.3'
      - commands:
          - mkdir -p output_plots/pdf output_plots/txt output_plots/root
          - chmod +x ./driver.csh
          - ls -l > output.txt
          - ./driver.csh >> output.txt
          - ls -l >> output.txt
    outputs:
      files:
        - output.txt
```

Submission
YAML



```
#!/bin/tcsh
source ./setup_env.csh
```

Driver

```
root -l -b -q 'pi0extraction.cc("MB", "PbSc", 4,5)'

echo root -l -b -q 'pi0extraction.cc("ERT", "PbSc", 4,5)'
root -l -b -q 'pi0extraction.cc("ERT", "PbSc", 4,5)'

echo root -l -b -q 'WGRatio.cc'
root -l -b -q 'WGRatio.cc'

echo root -l -b -q 'generationRM_Pi0.cc'
root -l -b -q 'generationRM_Pi0.cc'
```

```
setenv OFFLINE_MAIN /cvmfs/phenix.sdcc.bnl.gov/x8664_sl7/release/release_new/new
setenv ONLINE_MAIN /cvmfs/phenix.sdcc.bnl.gov/x8664_sl7/release/release_new/new
setenv ROOTSYS /cvmfs/phenix.sdcc.bnl.gov/x8664_sl7/opt/phenix/core/root-5.34.36
setenv G4_MAIN /cvmfs/phenix.sdcc.bnl.gov/x8664_sl7/opt/phenix/core/geant4.10.00.p02

source /opt/phenix/core/stow/opt_phenix_scripts/bin/phenix_setup.csh

setenv LD_LIBRARY_PATH :$LD_LIBRARY_PATH

setenv ODBCINI $(PWD)/afs/rhic.bnl.gov/phenix/etc/odbc.ini
setenv PG_PHENIX_DBNAME Phenix_phnxdbrcf2_C
```

Setup

REANA - Parallel Part

- Original workflow calls for Condor submission, 60 jobs in parallel
- Optimal way to re-implement this in REANA is to use non-linear - parallel - workflow mechanism, which is described using a very different YAML syntax
- One of the two options is the “CWL” language, the other is “yadage”
- Status - studying CWL - there is a level of complexity
- Example of documentation: http://www.commonwl.org/user_guide/
- Once done, will need to “stitch together” the serial and the parallel parts