Translating Analyses Into Prototype Analysis Systems

Brian O. Cruz Rodríguez
Jim Pivarski

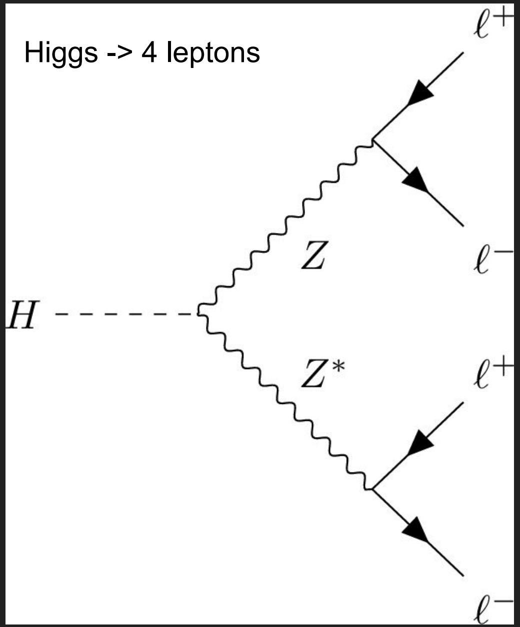
New Perspectives 2021
August 18, 2021

brian.cruz2@upr.edu
(University of Puerto Rico - Mayagüez Campus)

pivarski@princeton.edu
(Princeton University)

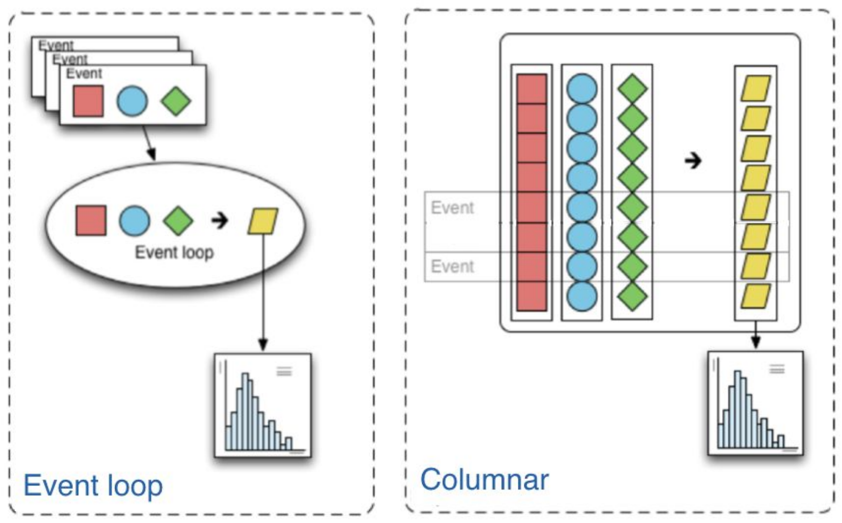
Introduction

- IRIS-HEP Fellowship program
  - Software R&D projects
- Analyze Higgs to 4 leptons decay
  - The four-leptons decay include: 4μ, 4e, and 2μ2e
  - Using CMS Open Data
  - Traditional analysis in ROOT (original)
  - Using COFFEA and Awkward-array (prototype)
- Compare the prototype and original analyses for:
  - Time-to-insight
  - Functionality
  - Reusability



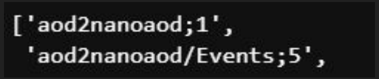
Original vs Prototype

- Original
  - Has the data as C++ objects
  - Goes event-by-event (row-wise)
- Prototype
  - Has the data as a numpy array
  - Ability to go column-wise



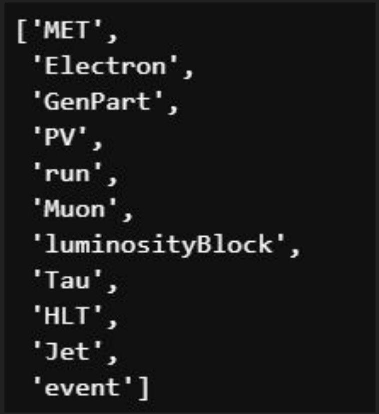
Original and Prototype

- ROOT files have a columnar format
  - Data is stored in a TTree class called Events
  - TTrees has TBranches
  - TBranches has “leaves”



```
['aod2nanoaod;1',
 'aod2nanoaod/Events;5',
```

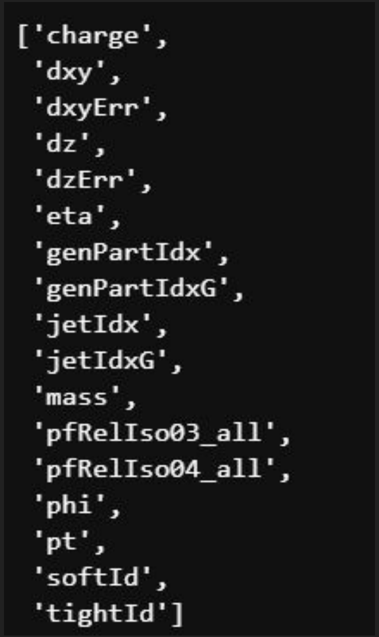
Tree



```
['MET',
 'Electron',
 'GenPart',
 'PV',
 'run',
 'Muon',
 'luminosityBlock',
 'Tau',
 'HLT',
 'Jet',
 'event']
```

Branches

Muon branch



```
['charge',
 'dxy',
 'dxyErr',
 'dz',
 'dzErr',
 'eta',
 'genPartIdx',
 'genPartIdxG',
 'jetIdx',
 'jetIdxG',
 'mass',
 'pfRelIso03_all',
 'pfRelIso04_all',
 'phi',
 'pt',
 'softId',
 'tightId']
```

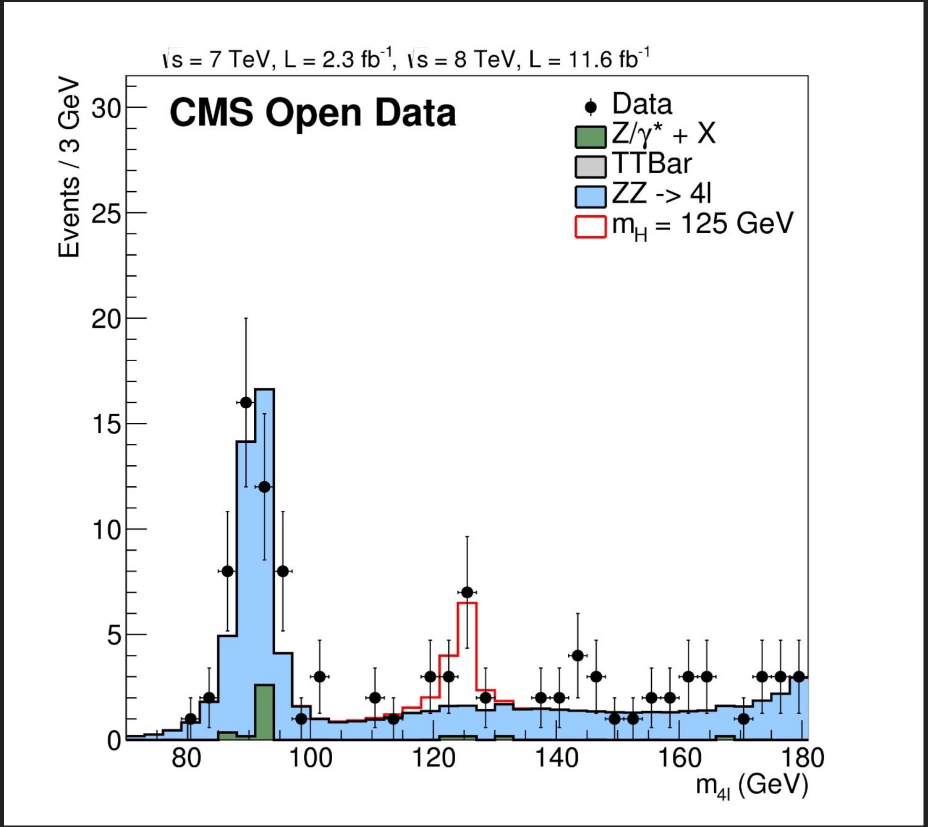
Leaves

Prototype Tools

- Uproot
  - Python implementation of the ROOT I/O
  - Independent of the ROOT toolkit
- Awkward-array
  - Library that can manipulate complex data structures with the efficiency of Numpy arrays
- COFFEA
  - Columnar Object Framework For Effective Analysis
  - Prototype package that uses Uproot, Awkward-array and the scientific python ecosystem
  - Provides an array-based syntax for HEP Event data manipulation

Objective

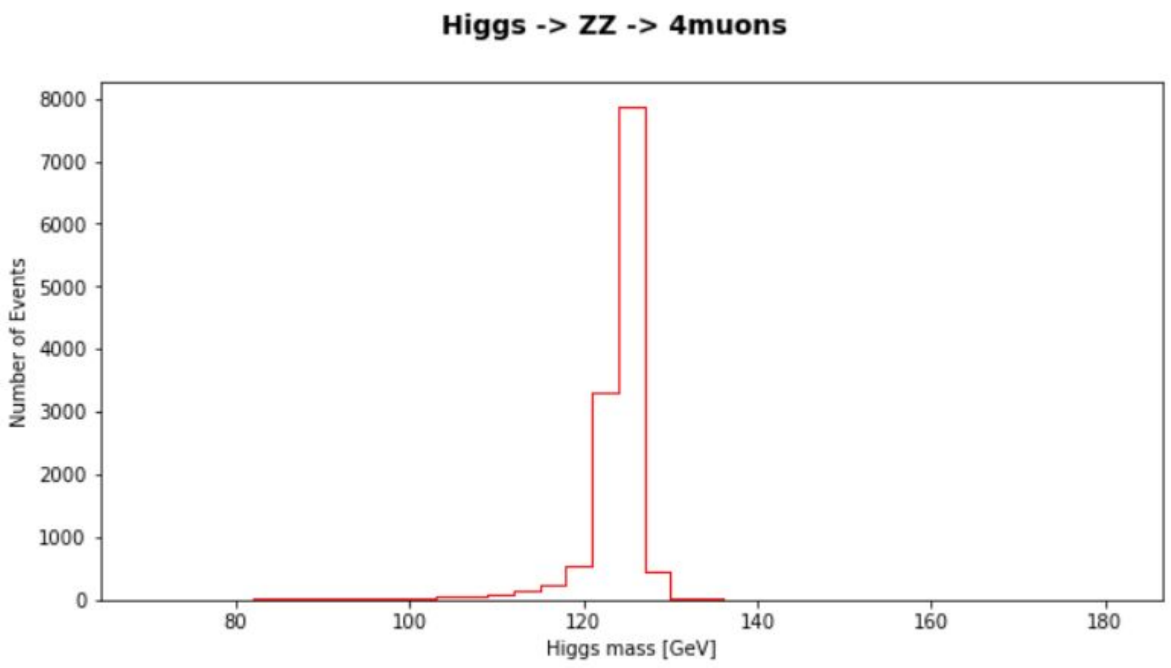
- To show that the prototype translation worked, we plotted the Higgs mass histogram (shown in red) for the 3 decays



Analysis Tools

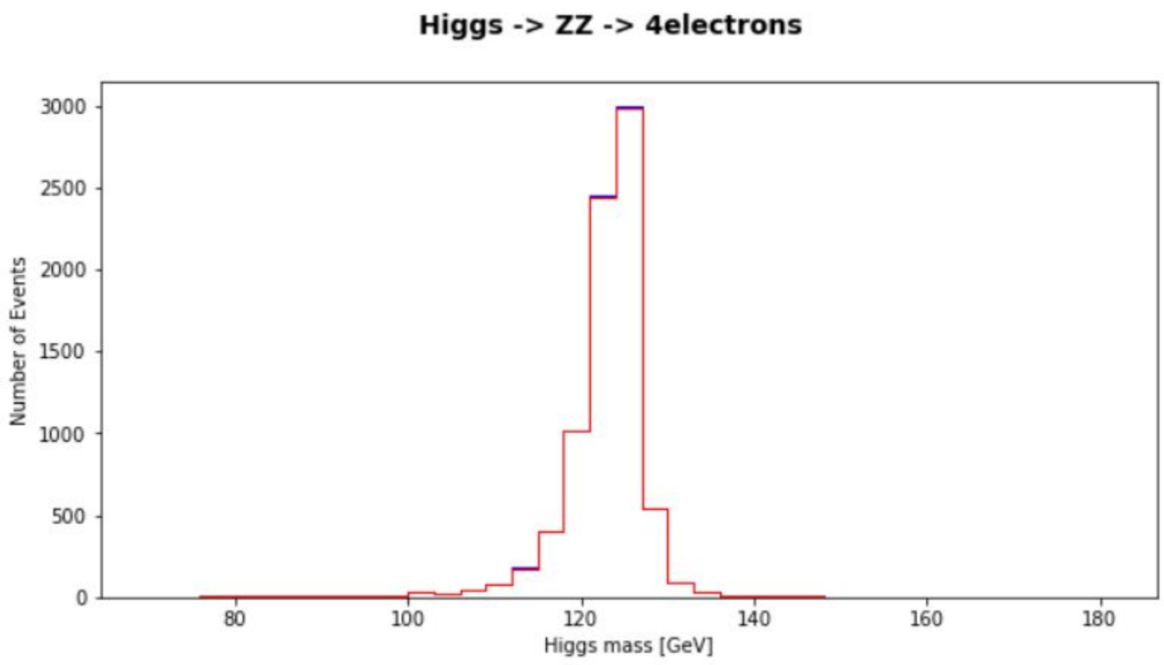
- Docker
  - Service that delivers self-sustaining software packages
  - Created a Docker container from the CMSSW_5_3_32 image
  - Download the cms-opendata-analyses/HiggsExample20112012 Github repository with the original code
  - Downloading the cms-opendata-analyses/AOD2NanoAODOutreachTool Github repository with a tool to convert AOD to NanoAOD
  - Download the Higgs to 4 leptons AODSIM dataset sample from the CERN Open Data Portal
  - Produce the NanoAOD of the sample
- JupyterLab
  - Interface for Jupyter notebooks
  - To interactively run the prototype code to compare it with the original

Higgs to 4μ mass histogram



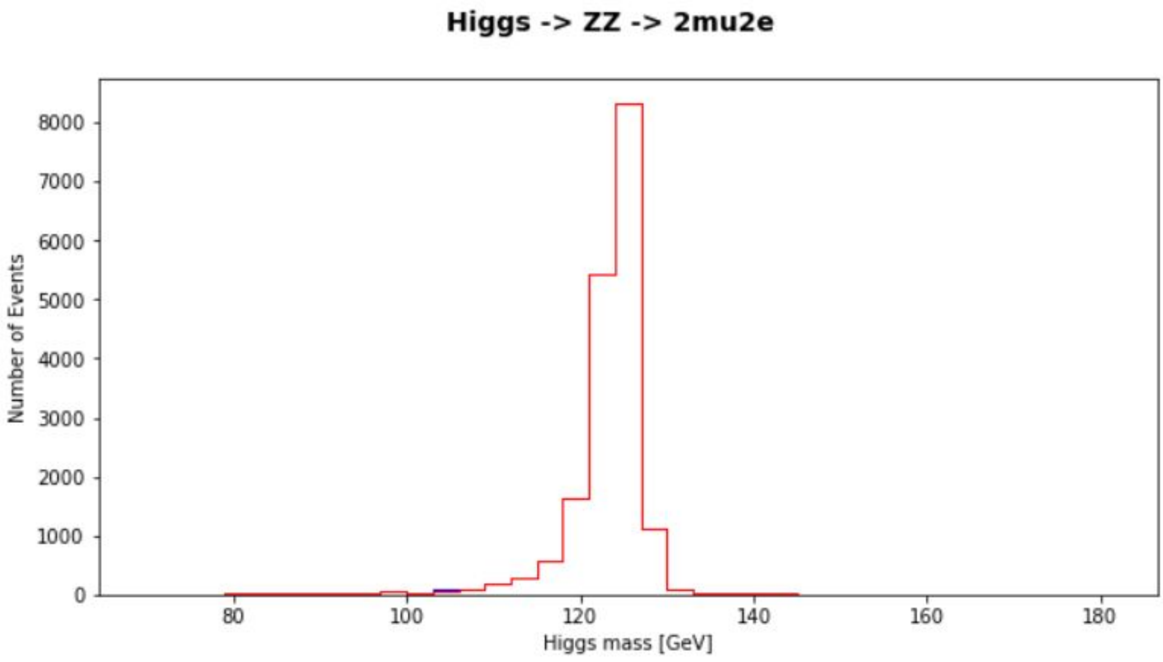
- Red: mass histogram from the original
- Blue: mass histogram from the prototype

Higgs to 4e mass histogram



- Red: mass histogram from the original
- Blue: mass histogram from the prototype

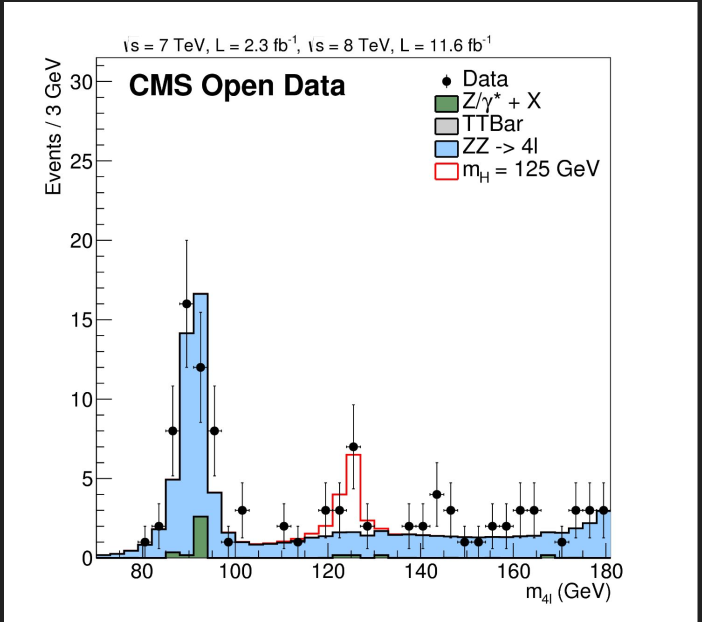
Higgs to 2μ2e mass histogram



- Red: mass histogram from the original
- Blue: mass histogram from the prototype

Next step

- Scale-up
- Analyze more datasets
  - Produce the NanoAODs of all the 21 Higgs analysis samples
  - Produce the Higgs plot shown earlier

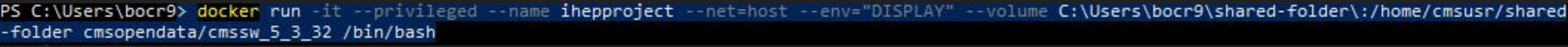


Thank you!
Any questions?

Backup Slides

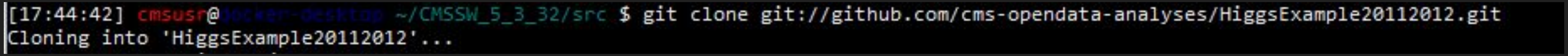
Set-up

- Creating a Docker container from the CMSSW_5_3_32 image



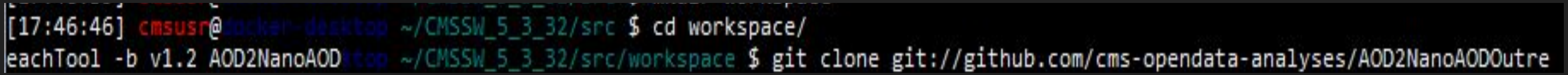
```
PS C:\Users\bocr9> docker run -it --privileged --name iheppproject --net=host --env="DISPLAY" --volume C:\Users\bocr9\shared-folder\:/home/cmsusr/shared
-folder cmsopendata/cmssw_5_3_32 /bin/bash
```

- Download the cms-opendata-analyses/HiggsExample20112012 Github repository in the Docker container and compile the codes with ***scram b***



```
[17:44:42] cmsusr@docker-desktop ~/CMSSW_5_3_32/src $ git clone git://github.com/cms-opendata-analyses/HiggsExample20112012.git
Cloning into 'HiggsExample20112012'...
```

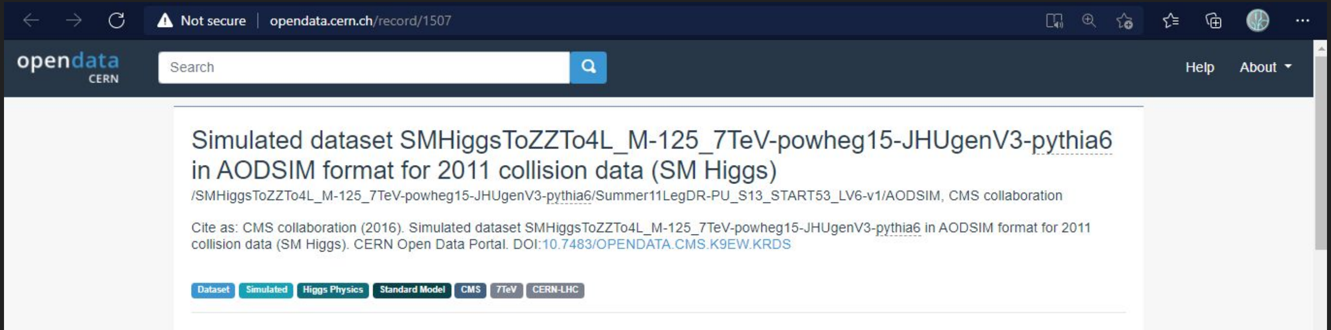
- Downloading the cms-opendata-analyses/AOD2NanoAODOutreachTool Github repository too, and compile the code as well



```
[17:46:46] cmsusr@docker-desktop ~/CMSSW_5_3_32/src $ cd workspace/
eachTool -b v1.2 AOD2NanoAODtop ~/CMSSW_5_3_32/src/workspace $ git clone git://github.com/cms-opendata-analyses/AOD2NanoAODOutre
```

Downloading the Dataset

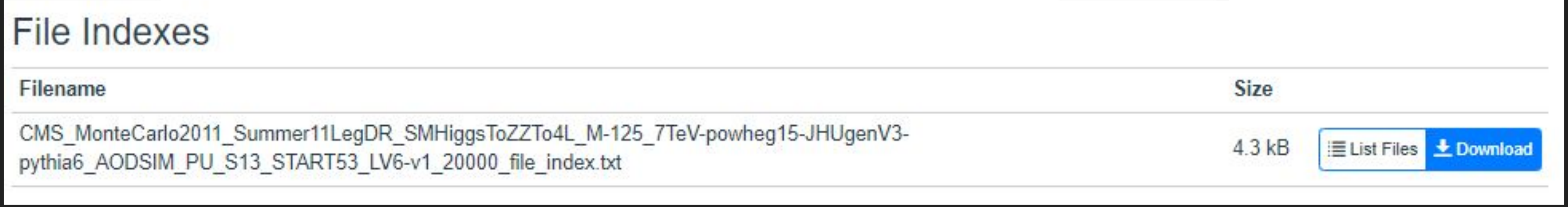
- Download the simulated Higgs to 4 leptons AODSIM sample from the CERN Open Data Portal into a Docker container



Simulated dataset SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6 in AODSIM format for 2011 collision data (SM Higgs)

/SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6/Summer11LegDR-PU_S13_START53_LV6-v1/AODSIM, CMS collaboration

Cite as: CMS collaboration (2016). Simulated dataset SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6 in AODSIM format for 2011 collision data (SM Higgs). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.K9EW.KRDS



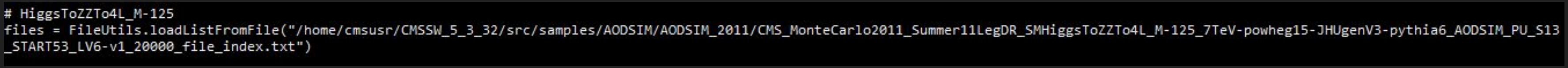
File Indexes

Filename	Size	
CMS_MonteCarlo2011_Summer11LegDR_SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6_AODSIM_PU_S13_START53_LV6-v1_20000_file_index.txt	4.3 kB	List Files Download

- Download the AODSIM index file in the Docker container

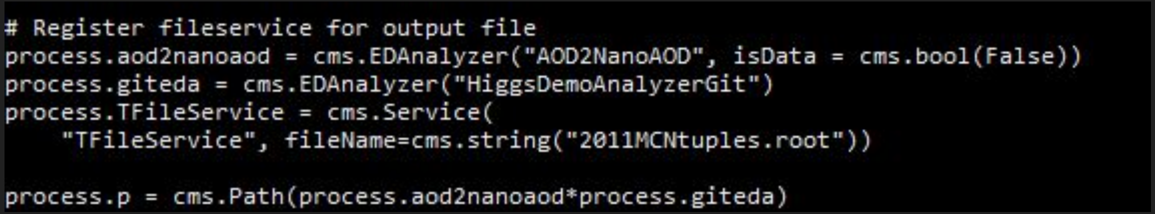
Producing the NanoAOD

- Add the index file as input to the proper Outreach Tool python configuration file, *simulation_cfg.py*



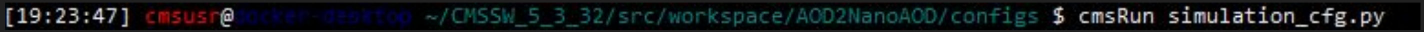
```
# HiggsToZZTo4L_M-125
files = FileUtils.loadListFromFile("/home/cmsusr/CMSSW_5_3_32/src/samples/AODSIM/AODSIM_2011/CMS_MonteCarlo2011_Summer11LegDR_SMHiggsToZZTo4L_M-125_7TeV-powheg15-JHUgenV3-pythia6_AODSIM_PU_S13
_START53_LV6-v1_20000_file_index.txt")
```

- Add the original EDAnalyzer, *HiggsDemoAnalyzerGit*, to the config file



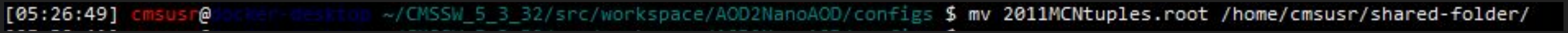
```
# Register fileservice for output file
process.aod2nanoaod = cms.EDAnalyzer("AOD2NanoAOD", isData = cms.bool(False))
process.giteda = cms.EDAnalyzer("HiggsDemoAnalyzerGit")
process.TFileService = cms.Service(
    "TFileService", fileName=cms.string("2011MCNtuples.root"))

process.p = cms.Path(process.aod2nanoaod*process.giteda)
```

- Run the config file to produce the NanoAOD



```
[19:23:47] cmsusr@docker-desktop ~/CMSSW_5_3_32/src/workspace/AOD2NanoAOD/configs $ cmsRun simulation_cfg.py
```

- Move the produced file to the shared folder



```
[05:26:49] cmsusr@docker-desktop ~/CMSSW_5_3_32/src/workspace/AOD2NanoAOD/configs $ mv 2011MCNtuples.root /home/cmsusr/shared-folder/
```