SW Review (CMS top)

Status Report

4-MARCH-2022

Main version of the classes are ready

- Gitlab repo updated: https://gitlab.cern.ch/cdozen/nanoaodrdframe-ul/-/tree/UL_TTLFV

  ➡Before making any changes/modification try to compile and run it!

  ❖Detailed instruction: NanoAODframework instruction

- BaseAnalyser.cpp & BaseAnalyser.h —> Default version
  - User need to rename the source and header files according to the their analysis
    - BaseAnalyser.cpp —> TprimeAnalyser.cpp or SingletopAnalyser.cpp
    - BaseAnalyser.h —> TprimeAnalyser.h or SingletopAnalyser.h

Object ID definition

- In NanoAODAnalyzerrdframe.cpp
- ElectronID added as an example. User need to define other object IDs (MuonID, JetID, etc.)

```
//cut-based ID Fall17 V2 (0:fail, 1:veto, 2:loose, 3:medium, 4:tight)
std::string NanoAODAnalyzerrdframe::ElectronID(int cutbasedID){

if(cutbasedID<0 || cutbasedID>4){
    std::cout<< "Error Wrong Electron ID requested == " << cutbasedID << "!! Can't be applied" <<std::endl;
    std::cout<< "Please select number from 0 to 4 " <<std::endl;
}

//Rdataframe look for the variables in the intput Ttree..
std::string output = Form("Electron_cutBased == %d ",cutbasedID);

return output;
}
```

Object ID definition

- In BaseAnalyser.cpp
- Call ElectronID

```
void BaseAnalyser::selectElectrons()
{
    cout << "select good electrons" << endl;
    //string veto_electron_str = Form ("Electron_pt>%f && abs(Electron_eta)<%f && %s",cutdb->Md_Electron_pT,cutdb->Md_Electron_eta, ElectronID(2).c_str());
    //_rlm = _rlm.Define("goodElectrons", veto_electron_str.c_str());

    _rlm = _rlm.Define("goodElectrons", ElectronID(2)); //without pt-eta cuts      You, last week • cleaned main class and added electron ID informat…
    _rlm = _rlm.Define("goodElectrons_pt", "Electron_pt[goodElectrons]")
               .Define("goodElectrons_eta", "Electron_eta[goodElectrons]")
               .Define("goodElectrons_phi", "Electron_phi[goodElectrons]")
               .Define("goodElectrons_mass", "Electron_mass[goodElectrons]")
               .Define("goodElectrons_idx", ::good_idx, {"goodElectrons"})
               .Define("NgoodElectrons", "int(goodElectrons_pt.size())");

    //-----------------------------------------------------
    //generate electron 4vector
    //-----------------------------------------------------
    _rlm = _rlm.Define("goodElectron_4Vecs", ::generate_4vec, {"goodElectrons_pt", "goodElectrons_eta", "goodElectrons_phi", "goodElectrons_mass"});
```

Running on a range of entries

- BaseAnalyser.cpp:
- In Rdataframe
  - Range() : filter based on entry number

```
//============================Event Selection=============================================//
// Cuts to be applied in order
//========================================================================================//
void BaseAnalyser::defineCuts()
{
   // count some entries using ranges
    auto Nentry = _rlm.Count();
    // This is how you can express a range of the first 100 entries
    _rlm = _rlm.Range(0, 100);
    auto Nentry_100 = _rlm.Count();
    // This is how you pick all entries from 10 onwards
    _rlm = _rlm.Range(10, 0);
    auto Nentry_10_end = _rlm.Count();
    // We can use a stride too, in this case we pick an event every 3
    _rlm = _rlm.Range(10, 0, 3);
    auto Nentry_10_end_3 = _rlm.Count();

    cout << "Usage of ranges:\n"
        << " - All entries: " << *Nentry << endl
        << " - Entries from 0 to 100: " << *Nentry_100 << endl
        << " - Entries from 10 onwards: " << *Nentry_10_end << endl
        << " - Entries from 100 onwards in steps of 3: " << *Nentry_100_end_3 << endl;
```

Running on Max number of files: - -nrootfiles

- Select number of files in the command line:
  - - -nrootfiles (or -N) 4 //4 files will be selected in the inputfile directory

```
time ./processnanoaod.py --allinone  -Y 2018 --nrootfiles 4 --isDATA 0 roottestfilesUL18/ testBaseANA_UL18.root >testBaseUL.out
```

Running on Data / MC and UL / ReReco Selection

- DATA and MC selection can be done by user : on the command line
  - - -isDATA 0 : means MC
  - - -isDATA 1 : means DATA
  - - - isDATA -1 or nothing defined : it will be selected according to the input file “genweight” branch

```
time ./processnanoaod.py --allinone  -Y 2018 --nrootfiles 4 --isDATA 0 roottestfilesUL18/ testBaseANA_UL18.root >testBaseUL.out
```

- UL and ReReco selection can be done by user : on the command line
  - - -runtype (or -R) UL
  - - - runtype (or -R) ReReco

```
time ./processnanoaod.py --allinone  -Y 2018 --runtype UL --isDATA 0 roottestfilesUL18/ testBaseANA_UL18.root >testBaseUL.out
```

Running options

- Structure:
  - ./processnanoaod.py [option] [inputdir] [outputdir]
  - ./processnanoaod.py [options] [inputDir] {$outputroot} -> for “- - allinone” option
    - [option]:
      - **- - split** : How many jobs to split into. it will allow multithreading you to use multiple CPUs to process the files. (e.g. —split=5)
      - **[inputdir]:** Inputfiles directory : **roottestfilesUL18/**
      - **[outputdir]**: output directory : **testBaseANA_UL : no need to create an output directory, it will be created if there isn’t**

```
./processnanoaod.py  --split 4  -Y 2018 --runtype UL  - -nrootfile 4 --isDATA 0  roottestfilesUL18/ testBaseANA_UL
```

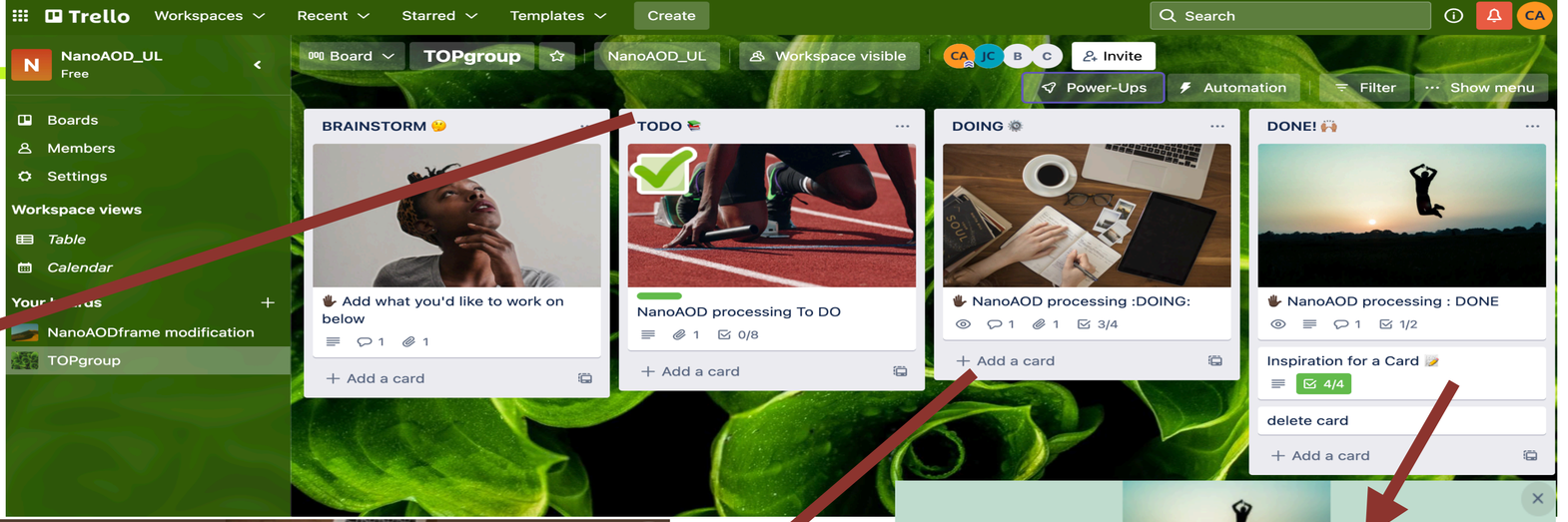
      - **--allinone** (-A) : will let you get a **single output root file** for all the files in the inputDir.
      - {**outputroot}**: should be a root file name : **testBaseANA_UL18.root**

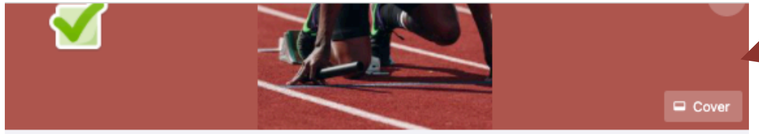
```
./processnanoaod.py --allinone  -Y 2018 --runtype UL --nrootfiles 4 --isDATA 0 roottestfilesUL18/ testBaseANA_UL18.root
```

Bonus slide

Web-based working space : Trello

[Trello_NanoUL_cmstop_working_space]





NanoAOD processing To DO

in list TODO

Labels

Description Edit

NanoAOD Processing
Tasklist :
https://indico.in2p3.fr/event/26311/contributions/105849/attachments/68731/96719/SB_RooDataFrame_9Fev2022.pdf

weekly meeting Lyon CMS NanoAOD top group indico:
Software Review (CMS Top)

To DO: Hide checked items Delete

38%

- [x] ~~Data OR MC should be a Boolean which is given as INPUT (or guessed by the input file name)~~
- [ ] JSON file should be read only on demand: if we have already processed a skim, the JSON should have been already applied.--->It is the responsibility of the analyzer to know what to ask on demand à BUT crucial to keep track of the information à Usage of input text file or code it somewhere else...
- [ ] Why do we load the PU when we do the skim, but if we do not dothe skim, the load of JEC/SF is done? PU/JEC/SF should be loaded on demand depending on what we will use (our current NanoAOD will need only PU as all the rest is applied)
- [ ] For JEC/SF, all files should be ready for any years we want to do
- [ ] Not sure to have understood the need of: WeightCalculatorFromHistogram
- [ ] Overlap between jets/lepton, need to understand if there is nothing in nanoAOD itself...
- [x] ~~input: cannot be a full directory à Need to be more adaptable: File list, single file is the baseline~~
- [x] ~~We want to run 100k of each background to make a quick check, currently it is too difficult to do...~~

Add an item

Suggested: Join

Add to card: Members, Labels, Checklist, Dates, Attachment, Custom Fields

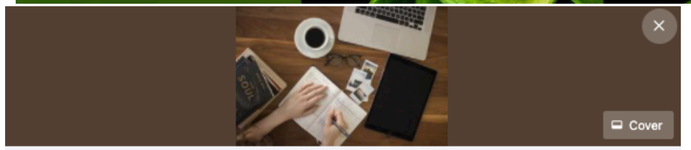
Add dropdowns, text fields, dates, and more to your cards.

Start free trial

Power-Ups: Add Power-Ups

Automation: In progress, Add button

Actions: Move, Copy, Make template, Watch, Archive, Share



NanoAOD processing :DOING:

in list DOING

Description

Add a more detailed description...

Attachments

Running late for a flight by Svitlana Dudarenko
Added 5 Jul 2019 at 00:06 - Comment - Delete - Edit
Make cover

Add an attachment

Checklist Hide checked items Delete

33%

- [x] ~~The ObjectID definition should be coded in NanoAODAnayzerrdframe.cpp~~
- [ ] cut000 definition should be change in the class with meaningfull name
- [ ] debug error mesagges should be added in the code

Add an item

Data or MC input files selection should be done by user Delete

0%

Add an item

Activity Show details

Suggested: Join

Add to card: Members, Labels, Checklist, Dates, Attachment, Custom Fields

Add dropdowns, text fields, dates, and more to your cards.

Start free trial

Power-Ups: Add Power-Ups

Automation: In progress, Add button

Actions: Move, Copy, Make template, Watch

NanoAOD processing : DONE

in list DONE!

Description Edit

NanoAOD Processing
Tasklist :
https://indico.in2p3.fr/event/26311/contributions/105849/attachments/68731/96719/SB_RooDataFrame_9Fev2022.pdf

weekly meeting Lyon CMS NanoAOD top group indico:
Software Review (CMS Top)

What needs to be done! Hide checked items Delete

100%

- [x] ~~TASK 1: Add a Boolean for UL/Re-Reco case (2018/2018UL))~~
- [x] ~~TASK2: Need to put a 'max of file' and to print clear how many events have been processed~~
- [x] ~~TASK3: Data or MC input files should be selected by user~~
- [x] ~~TASK4: Object ID defintion should be int~~
- [x] ~~TASK5: year should be integer~~

Add an item

Activity Show details

Write a comment...

Candan Dozen Altuntas 16 Feb at 18:13 (edited)

Task 1: Add a Boolean for UL/Re-Reco case (2018/2018UL))

- **processonefile.py**:
  Add an option as a run-type for UL or RecReco
  Add run-type option in "aproc" for BaseAnalyser

```
parser.add_option("-T", "--runtype",  dest="json", type='
aproc = ROOT.BaseAnalyser(t, outfile, options.year,optio
```

Suggested: Join

Add to card: Members, Labels, Checklist, Dates, Attachment, Custom Fields

Add dropdowns, text fields, dates, and more to your cards.

Start free trial

Power-Ups: Add Power-Ups

Automation: In progress, Add button

Actions: Move, Copy, Make template, Watch, Archive