



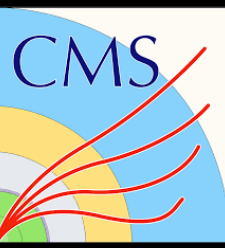
NTUPLE Processing With CMS Connect

Zach Flowers



Overview

- As some of you may know we are planning to process the next round of ntuples with CMS connect for many reasons: file limitations, job limitations, etc.
- This is intended to help people get up and running to spin these jobs
- The more people that can contribute the faster we can have new samples to look at!
- Note that a lot of this is taken from Erich's slides and experience:
 - https://indico.cern.ch/event/967079/contributions/4070053/attachments/2124897/3577346/Condor_with_CMS_Connect.pdf



Getting Started

- Setup an account with either an existing institutional ID like CERN or KU or an unaffiliated Globus account
 - <https://connect.uscms.org/signup>
- Make sure you are in the cms.org.ku group
- Note that connect uses ssh keys and there are two different login nodes
 - The node you will want to use is login-el7.uscms.org
- You will also need to have your grid certificate setup
- Other helpful links are in the backup



Framework Setup

- Once you have an account you can login to connect and navigate to your stash area: /stash/user/USERNAME/
- This is the main area that you should work out of
 - There no quotas on the number of files or space but try not to abuse that
- Here you will follow the same instructions for setting up the analysis code base as if you were working at some other site (UNL, LxPlus, LPC, etc.)
 - Note that the recommended version of CMSSW to use is CMSSW_10_6_5
 - <https://github.com/crogan/KUEWKinoAnalysis>



Framework Setup Pt. 2

- Instructions for setting up code can be found here:
<https://github.com/crogan/KUEWKinoAnalysis>
- Recommend using 10_6_5 as I've already made a sandbox that I think anyone can use
- After you have cloned the repos HiggsAnalysis, CombineHarvester, and KUEWKinoAnalysis come back here for RestFrames



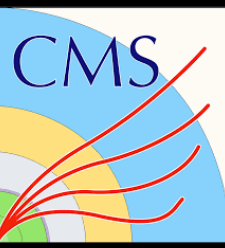
RestFrames

- Download the new version of RestFrames from this tarball with wget:
 - `wget --no-check-certificate http://stash.osgconnect.net/+zflowers/RestFrames_vNew.tar`
- Install RestFrames inside of your CMSSW/src area:
 - `tar -xvf RestFrames_vNew.tar`
 - `cd RestFrames/`
 - `./configure --prefix=$CMSSW_BASE/src/restframe_build && make && make install`



Compile Everything

- Go back to your CMSSW src folder
- If you haven't already run `cmsenv` and `scram b`
- `source RestFrames/setup_RestFrames.sh`
 - Note that there is also now a `setup_RestFrames.sh` in the KUEWKinoAnalysis framework. This for the jobs to use but can also be used for setting up RestFrames
- Go to KUEWKinoAnalysis and compile everything with `make cmssw`



Processing NTuples

- To submit jobs from the KUEWKinoAnalysis directory run the following:
 - `python scripts/CONDOR_unlt3_NANO_submit_connect.py`
`-split # -list list`
 - For split # I recommend between 5 and 25
 - Lists can be found in `samples/NANO/Lists/`
 - If the list is for data you need to add `--data` and for signals you need to add `--sms`



Monitoring Jobs

- Some of the jobs may get held due to not being able to find cvmfs
 - Release with `condor_release $USER`
- If you find the jobs repeatedly being held you should remove them as something probably went wrong in the submission
- Typically before submitting I will run `cmsenv`, source `RestFrames` and recompile (make `cmssw`) as this tends to fix most problems
- To check jobs run `condor_q $USER -batch`
 - There's also the monitoring website: https://grafana.mwt2.org/d/JiH0SAoZk/connect-servers?orgId=1&from=now-6M&to=now&var-Server=login-el7_uscms_org&refresh=1m
- Your files will be saved here: `/stash/user/USERNAME/NTUPLES/Processing/`



Other Items

- The files in here need to be added to KUEWKinoAnalysis
 - /home/t3-ku/z374f439/ForChris/
 - CONDOR_....py should go in scripts/
 - nano_....json should go in json/lwttnn/
 - Everything else should go in the main area (where make cmssw is ran)
- Problem with current HEM Veto: Always returns true (investigating)
- New lists
 - Added powheg ttbar, new SMS samples and staus
 - Waiting on filter efficiencies



Organization

- We need to process three types (data, bkg, sms) of datasets over three different years (2016, 2017, 2018)
 - Total of 9 lists
 - Biggest difference between each of these is number of jobs and files
- After processing we also need to hadd these and store them in a central place

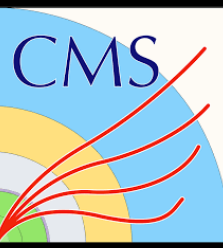


Backup



Other Links

- Twiki: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSConnect>
- SSH Keys: <https://ci-connect.atlassian.net/wiki/spaces/CMS/pages/57436024/Generate+SSH+key+pair+and+add+the+public+key+to+your+account>
- Handbook: <https://ci-connect.atlassian.net/wiki/spaces/CMS/overview>
- Support Email: cms-connect-support@cern.ch



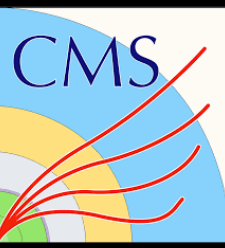
Sandboxes

- Type of sandbox used in our case for setting up the environment for the jobs
- One has already been made can be used as long as nothing that depends on CMSSW is changed
- Changes inside of KUEWKinoAnalysis should not require the creation of a new sandbox



Making A Sandbox

- Download the tool from git: <https://github.com/CMSConnect/cmssw-sandbox.git>
- Then run the following to create the sandbox: `cmssw_sandbox create -a CMSSW_X_Y_Z`
- Example:
 - Consider that you have a CMSSW release of version 10_2_11 with code inside of `CMSSW_10_2_11/src` that you want to put inside of a sandbox
 - Then run the following from outside of the `cmssw` area
 - `cmssw_sandbox create -a CMSSW_10_2_11`



Using The Sandbox

- Once you have the sandbox you need to copy it to the directory: `/stash/user/USERNAME/public/`
 - You may need to create the directory called public
- Give the directory 755 permissions and the sandbox file 644 permissions using `chmod`
- You should now be able to use the sandbox!
- To test it run: `wget --no-check-certificate http://stash.osgconnect.net/+USERNAME/SANDBOX_FILENAME`
- You also need to edit the `execute_script.sh` in `KUEWKinoAnalysis` to grab the correct sandbox