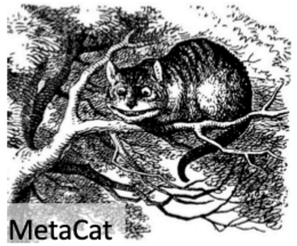
Brief introduction to justIN

New concepts

- **I highly encourage you to read about these new concepts**
- Location management **Rucio** https://rucio.github.io/
- Data catalog management **MetaCat** https://metacat.readthedocs.io/en/latest/
- Workflow management **justIN** https://justin.dune.hep.ac.uk/docs/



RUCIO

justIN

A few changes

- As you may have noticed, we are in a transition period at the production level we are transitioning from POMS to **justIN**
- At the user level we are transitioning from SAM data catalog to **MetaCat** (**Rucio**)
- To retrieve file location users used to do `samweb get-file-access-url`
- SAM locations are not supported anymore
- Instead, user needs to get the file location from Rucio
- `rucio -a $USER  list-file-replicas fardet-hd:nu_dune10kt_1x2x6_1062_166_20230823T120323Z_gen_g4_detsim_hitreco__20240222T231742Z_reco2.root --pnfs`

```
root://dcdndoor.sdcc.bnl.gov:1094/pnfs/sdcc.bnl.gov/data/dune/RSE/fardet-hd/91/66/nu_dune10kt_1x2x6_1062_166_20230823T120323Z_gen_g4_detsim_hitreco__20240222T231742Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/10/62/nu_dune10kt_1x2x6_1062_166_20230823T120323Z_gen_g4_detsim_hitreco__20240222T231742Z_reco2.root_1709158151
```

justIN

- justIN is a workflow system that processes data by satisfying the requirements of data location/data catalog, software distribution and job submission to the grid
- There is documentation on how to get started with justIN at https://justin.dune.hep.ac.uk/docs/tutorials.dune.md
- If you are a member of DUNE and have a Fermilab computing account, you should be able to run this brief tutorial
- **Before this, please READ the prerequisites carefully and check if you can login https://justin.dune.hep.ac.uk/docs/tutorials.dune.md and CONTINUE to read the DUNE tutorial to make sure you can run the examples**
- This tutorial will repeat every single step discussed in DUNE justIN tutorial with specific DUNE examples

Things you can do

- We will show how to do three basic tasks
  - Process data (submit a job to the grid) if you are using code from the base release and you don’t actually modify any of it
  - Process data (submit a job to the grid) if you are using code from the base release and you want to use a customized FCL file
  - Process data (submit a job to the grid) if you are NOT using code from the base release and you want to use customized code

Things you can do

- Process data (submit a job to the grid) if you are using code from the base release and you don't actually modify any of it
  - Once you have identified what data you want to process, you can see the most recent data (official data) sets available at https://wiki.dunescience.org/wiki/Data_Collections_Manager/data_sets
  - Let's say you want to run **`mergeana`** for electron neutrinos,
    - Where is the data?
    - In DUNE we provided datasets to easily identify a collection of files

```
fardet-hd:fardet-hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__prodgenie_n
ue_dune10kt_1x2x6__out1__validation
```

    - Dataset names tend to be self explanatory and includes the type of detector, which fcls files were used to produce, software version, data tier, and tag, in this case, the tag is validation

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don't actually modify any of it
  - Lets try to process `mergeana` in the first 100 files that in the data sets,
  - MetaCat relies on MQL queries to select a collection of files in this case to select the first 100 files of a given data set the query would be something like

```
"files from fardet-hd:fardet-hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__pro
dgenie_nue_dune10kt_1x2x6__out1__validation limit 100 ordered"
```

  - The flag "**ordered**" is crucial to ensure reproducibility
  - Now, how do you process this using justIN?
  - You need to provide a jobscirpt and run a justIN command line

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don’t actually modify any of it
  - The basic tasks of your jobscript are:
    - Set software env
    - Find where the data is
    - Process the data
    - Save the output in a given location
  - Examples of jobscripts can be found at https://github.com/DUNE/dune-prod-utils/tree/main/justIN-examples

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don't actually modify any of it
  - An example of a jobscript can be found at https://github.com/DUNE/dune-prod-utils/blob/main/justIN-examples/submit_ana.jobscript
  - Lets look at some of the code there

```
19    # fcl file and DUNE software version/qualifier to be used
20    FCL_FILE=${FCL_FILE:-standard_ana_dune10kt_1x2x6.fcl}
21    DUNE_VERSION=${DUNE_VERSION:-v09_81_00d02}
22    DUNE_QUALIFIER=${DUNE_QUALIFIER:-e26:prof}
23
44    # Setup DUNE environment
45    source /cvmfs/dune.opensciencegrid.org/products/dune/setup_dune.sh
46    setup dunesw "$DUNE_VERSION" -q "$DUNE_QUALIFIER"
47
61
62    lar -c $FCL_FILE $events_option -o $outFile "$pfn" > ${fname}_ana_${now}.log 2>&1
63    )
```

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don't actually modify any of it
  - As always, before sending 100s of jobs to the grid, it is highly recommended to test it interactively
  - AT THIS POINT, WE EXPECT THAT YOU HAVE SUCCESSFULLY REPRODUCED THE EXAMPLES IN the DUNE justIN tutorial
  - To test your jobscript interactively you can use the example from https://github.com/DUNE/dune-prod-utils/blob/main/justIN-examples/submit_ana.jobscript

```
$ source /cvmfs/dune.opensciencegrid.org/products/dune/setup_dune.sh
$ setup justin
$ justin-test-jobscript --mql "files from fardet-hd:fardet-
hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__prodgenie_nu_dune10
kt_1x2x6__out1__validation skip 10 limit 1 ordered " --jobscript
submit_ana.jobscript --env NUM_EVENTS=3
```

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don't actually modify any of it
  - If your interactively test was successful you can proceed to send your jobs
  - Crucial, make sure you have a valid proxy

```
$ export ROLE=Analysis
$ voms-proxy-init -rfc -noregen -voms=dune:/dune/Role=$ROLE -valid 120:00
```

  - Remember you have to specify where the output should go, for now the output can only go to /pnfs/dune/scratch/users/$USER

```
$ USERF=$USER

$ FNALURL='https://fndcadoor.fnal.gov:2880/dune/scratch/users'

$ justin simple-workflow --mql "files from fardet-hd:fardet-
hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__prodgenie_nu_dune10
kt_1x2x6__out1__validation skip 5 limit 5 ordered " --jobscript
submit_ana.jobscript --rss-mb 4000 --output-pattern "*_ana_*.root:$FNALURL/$USERF"
```

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you don't actually modify any of it

```
$ USERF=$USER
$ FNALURL='https://fndcadoor.fnal.gov:2880/dune/scratch/users'
$ justin simple-workflow --mql "files from fardet-hd:fardet-
hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__prodgenie_nu_dune
10kt_1x2x6__out1__validation skip 5 limit 5 ordered " --jobscript
submit_ana.jobscript --rss-mb 4000 --output-pattern "*_ana_*.root:$FNALURL/
$USERF"
```

- You can look at your job status by using justIN dashboard https://justin.dune.hep.ac.uk/dashboard/?method=list-workflows

Things you can do

- Process data (submit a job to the grid) if you are using code from the base release and you want to use a customized FCL file
- To do that, the best is to use the Rapid Code Distribution Service (RCDS) via cvmfs as explained in the tutorial
- Let's say you have a customized FCL file that you need to run over some datasets. As per instruction from the DUNE justIN tutorial you need to tar the files needed and put them in cvmfs

```
$ tar cvz my_fcls.tar my_fcls
$ source /cvmfs/dune.opensciencegrid.org/products/dune/setup_dune.sh
$ setup justin
$ rm -f /tmp/x509up_u`id -u`
$ kx509
$INPUT_TAR_DIR_LOCAL=`justin-cvmfs-upload my_fcls.tar`
```

- Wait a few minutes to check the files

```
$ ls -l $INPUT_TAR_DIR_LOCAL
```

Things you can do

- Process data (submit a job to the grid) if you are using code from the base release and you want to use a customized FCL file
- You can look at the example at https://github.com/DUNE/dune-prod-utils/blob/main/justIN-examples/submit_local_fcl.jobscript
- The key part of the code is the following

```
19    # fcl file and DUNE software version/qualifier to be used
20    FCL_FILE=${FCL_FILE:-$INPUT_TAR_DIR_LOCAL/my_code/my_ana.fcl}
21    DUNE_VERSION=${DUNE_VERSION:-v09_81_00d02}
22    DUNE_QUALIFIER=${DUNE_QUALIFIER:-e26:prof}
23
```

```
$ justin simple-workflow --mql "files from fardet-hd:fardet-
hd__fd_mc_2023a_reco2__full-
reconstructed__v09_81_00d02__standard_reco2_dune10kt_nu_1x2x6__prodgenie_nu_dune10
kt_1x2x6__out1__validation skip 5 limit 5 ordered “ --jobscript
submit_local_fcl.jobscript --rss-mb 4000 --env
INPUT_TAR_DIR_LOCAL="$INPUT_TAR_DIR_LOCAL"
```

Things you can do

- Process data (submit a job to the grid) if you are NOT using code from the base release and you want to use customized code
- Probably you are developing some reconstruction alg and you want to check the results in a large sample, before committing your software to GitHub
- You can use your customized software (e.g. local installation of dunereco) and use justIN to process the data with your new LArSoft module
- Similar to the previous part, you will need to provide all pieces in a tar file and put them in cvmfs

```
$ tar cvz my_code.tar my_code
```

- Here `my_code.tar` includes a directory with `my_fcls` files and one with my local products (e.g. `localProducts_larsoft_v09_85_00_e26_prof`) this is similar to what you used to do when using `jobsub` and using customized code

Things you can do

- Process data (submit a job to the grid) if you are NOT using code from the base release and you want to use customized code

- Next step is to send your tar file to cvmfs

```
$ tar cvz my_code.tar my_code
$ source /cvmfs/dune.opensciencegrid.org/products/dune/setup_dune.sh
$ setup justin
$ rm -f /tmp/x509up_u`id -u`
$ kx509
$INPUT_TAR_DIR_LOCAL=`justin-cvmfs-upload my_code.tar`
```

- Wait a few minutes to check the files

```
$ ls -l $INPUT_TAR_DIR_LOCAL
```

Things you can do

- Process data (submit a job to the grid) if you are just using code from the base release and you want to use customized software
- Next step is to send your tar file to cvmfs
- Once you files appear in cvmfs you can try use the following example
- https://github.com/DUNE/dune-prod-utils/blob/main/justIN-examples/submit_local_code.jobscript

```
19    # fcl file and DUNE software version/qualifier to be used
20    FCL_FILE=${FCL_FILE:-$INPUT_TAR_DIR_LOCAL/my_code/fcls/my_reco.fcl}
21    DUNE_VERSION=${DUNE_VERSION:-v09_85_00d00}
22    DUNE_QUALIFIER=${DUNE_QUALIFIER:-e26:prof}
```

```
44    # Setup DUNE environment
45    source /cvmfs/dune.opensciencegrid.org/products/dune/setup_dune.sh
46    source $INPUT_TAR_DIR_LOCAL/my_code/localProducts_larsoft_v09_85_00_e26_prof/setup
47    setup dunesw "$DUNE_VERSION" -q "$DUNE_QUALIFIER"
48    mrbslp
```

Things you can do

- Process data (submit a job to the grid) if you are NOT using code from the base release and you want to use customized software
- Next step is to send your tar file to cvmfs
- Once your files appear in cvmfs you can try the following example
- https://github.com/DUNE/dune-prod-utils/blob/main/justIN-examples/submit_local_code.jobscript

```
$ justin simple-workflow —-mql "files from fardet-hd:fardet-
hd__fd_mc_2023a__hit-
reconstructed__v09_78_01d01__standard_reco1_dune10kt_1x2x6__prodgenie_nu_dune10k
t_1x2x6__out1__v1_official limit 10 ordered" --jobscript /exp/dune/data/users/
higuera/data_collections/justIN/submit_local_code.jobscript --env
INPUT_TAR_DIR_LOCAL="$INPUT_TAR_DIR_LOCAL"
```

Questions

- Use the #computing-worfklow slack channel
- Justin commands
- $ export RUCIO_ACCOUNT= $USER
- $ finish-workflow --workflow-id <ID>