

CAF production (shifter notes)

Last week (production meeting and discussion with Heidi)

- it has been validated the possibility to run *lar* on an input files list to produce *caf.root* file (100 and 1K input files)
- Metadata for a *caf.root* from 1K input events have been created by using [*mergeMetaCat.py*](#)

This week

- *justIN* jobscrip to run *lar* on N files and run [*mergeMetaCat.py*](#) has been prepared and tested

working dir: `/exp/dune/app/home/dunepro/caf/justin`

two steps:

1. run from my account, and output files in scratch area `/pnfs/dune/scratch/users/epennacc/`
2. run as `dunepro`

caf_2.jobscript

```
for nf in {1..2}
do
  DID_PFN_RSE=`$JUSTIN_PATH/justin-get-file`
  ##Check that any file was returned
  if [ "${DID_PFN_RSE}" == "" ] ; then
    echo "Could not get file"
    exit 0
  fi

  FILE=`echo ${DID_PFN_RSE} | cut -f2 -d' '`
  DID=`echo ${DID_PFN_RSE} | cut -f1 -d' '`
  echo ${DID} >> did.list
  echo ${FILE} >> file.list
done
```

build file.list, and did.list, each one with 2 entries (the same)

fi

```
now=$(date -u +"%Y%m%dT%H%M%SZ")
```

```
####Run cafmaker
```

```
echo "now run lar "
```

```
lar -c "${CAF_FCL}" -S file.list > caf_${now}.log
```

```
python $INPUT_DIR/mergeMetaCat.py --fileName caf.root --nameSpace usertest --fileList did.list
```

```
# echo "$FILE" > justin-processed-pfns.txt
cat file.list > justin-processed-pfns.txt
```

```
cp justin-processed-pfns.txt caf_${now}.pfns
cat did.list >> caf_${now}.did
cat file.list >> caf_${now}.file
cp caf.root caf_${now}.root
cp caf.root.json caf_${now}.root.json
```

All files (list and output) are copied back

Example of lists:

```
<dunegpvm01.fnal.gov> more caf_20240725T075648Z.file  
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/23/ee/nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1003_149_20230809T224150Z_gen_g4_detsim_hitreco__20240223T010852Z_reco2.root  
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/91/c6/nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1253_722_20230803T002135Z_gen_g4_detsim_hitreco__20240222T150055Z_reco2.root
```

```
<dunegpvm01.fnal.gov> more caf_20240725T075648Z.did  
fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1003_149_20230809T224150Z_gen_g4_detsim_hitreco__20240223T010852Z_reco2.root  
fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1253_722_20230803T002135Z_gen_g4_detsim_hitreco__20240222T150055Z_reco2.root  
<dunegpvm01.fnal.gov> █
```

Workflows summary:

| workflow | # input file /jobs | Input dataset size | # submitted jobs | # jobs ended (*) |
|----------|--------------------------------|-----------------------------|------------------|------------------|
| 2622 | 2 (<i>caf_2.jobscript</i>) | 2 (<i>query.sh</i>) | 1 | 1 |
| 2623 | 2 (<i>caf_2.jobscript</i>) | 10 (<i>query_10.sh</i>) | 5 | 5 |
| 2624 | 10 (<i>caf_10.jobscript</i>) | 20 (<i>query_20.sh</i>) | 2 | 1 |
| 2627 | 10 (<i>caf_10.jobscript</i>) | 40 (<i>query_40.sh</i>) | 4 | 3 |
| 2628 | 20 (<i>caf_20.jobscript</i>) | 100 (<i>query_100.sh</i>) | 5 | 2 |

➤ Workflow example:

```
justin simple-workflow --mql "$MQL_QUERY20" --jobscript caf_10.jobscript --max-distance 0 --rss-mib 4000 --env INPUT_DIR=/cvmfs/fifeuser4.opensciencegrid.org/sw/dune/fc52fc798038c86afea45ebe468bfdec150e6a8c --output-pattern "caf_*.*: $FNALURL/$USERF"
```

2624

→ variable *\$MQL_QUERYn* is defined by sourcing *query_n.sh*

(*) distance set to zero, no checks on where files are, so some files may not be available

```
<dunegpvm01.fnal.gov> ls -rtl /cvmfs/fifeuser4.opensciencegrid.org/sw/dune/fc52fc798038c86afea45ebe468bfdec150e6a8c
total 33
-rw-r--r-- 1 cvmfs cvmfs  393 Jul 24 13:36 CheckSum.py
-rw-r--r-- 1 cvmfs cvmfs 2285 Jul 24 13:36 CheckConfiguration.py
-rw-r--r-- 1 cvmfs cvmfs 6107 Jul 24 13:36 TypeChecker.py
-rw-r--r-- 1 cvmfs cvmfs 2169 Jul 24 13:36 TimeUtil.py
-rw-r--r-- 1 cvmfs cvmfs 21087 Jul 24 13:36 mergeMetaCat.py
<dunegpvm01.fnal.gov>
```

<https://github.com/DUNE/data-mgmt-ops/blob/master/utilities/mergeMetaCat.py>

data-mgmt-ops / utilities / mergeMetaCat.py 

 **hschellman** dataset input for mergeRoot

Code Blame 529 lines (436 loc) · 20.6 KB

```
1 ##### metadata helper #####
```

Example of outputs (/pnfs/dune/scratch/users/epennacc/02628/1)

```
Jul 25 04:25 caf_20240725T090914Z.log
Jul 25 04:25 caf_20240725T090914Z.root
Jul 25 04:25 caf_20240725T090914Z.pfn
Jul 25 04:25 caf_20240725T090914Z.root.json
Jul 25 04:25 caf_20240725T090914Z.file
Jul 25 04:25 caf_20240725T090914Z.did
Jul 25 04:28 caf_20240725T091427Z.log
Jul 25 04:28 caf_20240725T091427Z.root.json
Jul 25 04:28 caf_20240725T091427Z.root
Jul 25 04:28 caf_20240725T091427Z.pfn
Jul 25 04:28 caf_20240725T091427Z.file
Jul 25 04:28 caf_20240725T091427Z.did
```

```
<dunegpvm01.fnal.gov> tail caf_20240725T090914Z.log
TrigReport      Run      Success      Error      Name
TrigReport      2000      2000          0      cafmaker

TimeReport ----- Time summary [sec] -----
TimeReport CPU = 790.816955 Real = 914.659315

MemReport ----- Memory summary [base-10 MB] -----
MemReport VmPeak = 2125.48 VmHWM = 917.176

Art has completed and will exit with status 0.
<dunegpvm01.fnal.gov> du -hs caf_20240725T090914Z.root
3.4M   caf_20240725T090914Z.root
```

caf.root file need to be merged

Running as dunepro → ok

Workflow 2632 (and 2633)

| | | |
|--------------------|--|----------------|
| Input files | fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1003_149_20230809T224150Z_gen_g4_detsim_hitreco__20240223T010852Z_reco2.root fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1253_722_20230803T002135Z_gen_g4_detsim_hitreco__20240222T150055Z_reco2.root | |
| Jobscript | Exit code | 0 |
| | Real time | 0m (35s) |
| | CPU time | 0m (31s = 88%) |
| Outputting started | 2024-07-25 13:02:41 | |
| Output files | usertests:caf_20240725T130208Z.root | |

Output in [metacat](#) (2633)



DUNE/ProtoDUNE MetaCat Production

[users](#) [roles](#) [namespaces](#) [datasets](#) [find file](#) [categories](#) [query](#) [named queries](#) [filters](#)

[docs](#) logged in as El

File Information

File ID [xxd5Jt5yR0W9gqKg](#)

Namespace [usertests](#)

Name [caf_20240725T130208Z.root](#)

DID [usertests:caf_20240725T130208Z.root](#)

Size 448765 (0.428 MB)

Created 2024-07-25 08:02:51.249938-05:00 by [dunepro](#)

Updated 2024-07-25 08:03:00.960813-05:00 by [dunepro](#)

Checksums **adler32:** f3108143

Parents [fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1003_149_20230809T224150Z_gen_g4_detsim_hitreco__20240223T010852Z_reco2.root](#)
[fardet-vd:nu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_1253_722_20230803T002135Z_gen_g4_detsim_hitreco__20240222T150055Z_reco2.root](#)

Children

Next steps...

1. 13 input datasets to be processed, 6+1 for HD, 6 for VD ~400k input files:
 - datasets have to be on disk
 - decide where to run jobs, where to store caf.root files waiting to be merged
2. Check metadata files
https://metacat.fnal.gov:9443/dune_meta_prod/app/gui/show_file?did=usertests%3Acaf_20240725T130208Z.root
https://metacat.fnal.gov:9443/dune_meta_prod/app/gui/show_file?did=usertests%3Acaf_20240725T130550Z.root
3. How to merge caf.root files ? Can be a dedicated workflow?
4. Input list of 20 files tested now. Can be increased?



- Points 1, 2 can be started right now
- Point 3 require some dedicated work to understand needed resources (people, time, disk....)
- Point 4 is very critical. More tests to « stress » the system have to be performed : can be done only when all key people available. Given the fact that many of us will not be around in the next weeks, propose to postpone this point to the second half of August

For production campaigns run this year, a pdf file from shifter notes has been created and uploaded on DocDB (SP 2GeV still to be done)

1. [DUNE-doc-31455-v1](#) (Jake) **Production report: PDHD 1GeV MC**
2. [DUNE-doc-31458-v4](#) (Jake + Doug) **Production report: LBL reco2**
3. [DUNE-doc-31461-v1](#) (Aaron) **Production report: reco2 atmospheric**
4. [DUNE-doc-31464-v1](#) (Elisabetta) **Production report: PDVD validation samples**