

RITM2157117: CAF files production for latest LBL and atmospheric productions

- dunesw v09_91_01d00

1. cafmaker_dune10kt_1x2x6_runreco-nuenergy_geov5.fcl

*fardet-hd__fd_mc_2023a_reco2__full-reconstructed_** 6 datasets

2. cafmaker_dunevd10kt_1x8x6_3view_30deg_runreco-nuenergy_geov3.fcl

*fardet-vd__fd_mc_2023a_reco2__full-reconstructed_** 6 datasets

3. cafmaker_atmos_dune10kt_1x2x6_runreco-nuenergy-nuangular_geov5.fcl

*fardet-hd__full-reconstructed__v09_85_00d00__reco2_atmos_** 1 dataset

13 input samples (~500 TB , 350K files)

```

TrigReport ----- Event summary -----
TrigReport Events total = 100 passed = 100 failed = 0

```

```

TrigReport ----- Modules in End-path -----
TrigReport      Run      Success      Error Name
TrigReport      100      100          0 cafmaker

```

```

TimeReport ----- Time summary [sec] -----
TimeReport CPU = 38.887287 Real = 70.729220

```

```

MemReport ----- Memory summary [base-10 MB] -----
MemReport VmPeak = 2463.18 VmHWM = 910.897

```

FD1

Interactive tests on dunegpvm01

# evts/file	reco2	caf.root
100	~1GB	275 KB

Each job producing a file named *caf.root*

sample	#files	Caf files total size
FD1	105k	30GB
FD2	97k	27GB
FD1 atmo	146k	42GB

```

TrigReport ----- Event summary -----
TrigReport Events total = 100 passed = 100 failed = 0

```

```

TrigReport ----- Modules in End-path -----
TrigReport      Run      Success      Error Name
TrigReport      100      100          0 cafmaker

```

```

TimeReport ----- Time summary [sec] -----
TimeReport CPU = 12.772586 Real = 21.597696

```

```

MemReport ----- Memory summary [base-10 MB] -----
MemReport VmPeak = 2373.32 VmHWM = 800.948

```

FD2

```

TrigReport ----- Modules in End-path -----
TrigReport      Run      Success      Error Name
TrigReport      100      100          0 cafmaker

```

```

TimeReport ----- Time summary [sec] -----
TimeReport CPU = 8.489746 Real = 24.548577

```

```

MemReport ----- Memory summary [base-10 MB] -----
MemReport VmPeak = 2455.45 VmHWM = 852.41

```

atmo FD1

Questions to be addressed

➤ How to proceed?

1) run on each file input file, store individual *caf.root* file then merge them in several *caf* files of “appropriate” size

2) Proposal from Heidi in slack:

Heidi Schellman 12:52 AM

Hi folks [RITM2157117](#) is for CAF production each file takes 20s and produces 1/4 MB. I have a working merge program optimized for calibration files. In principle I can make one for this but it looks as if it might be better to start with jobs running over lists of 100-1000 files per process. Do we have that capability at the moment?

tested on 10 files (1000 events):

➤ `lar -c cafmaker_dunevd10kt_1x8x6_3view_30deg_runreco-nuenergy_geov3.fcl -S input_vd.list`

➤ `lar -c cafmaker_dune10kt_1x2x6_runreco-nuenergy_geov5.fcl -S input_hd.list`

output size ~1.9 MB

FD1-HD

```
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_530_20230825T19162
5Z_gen_g4_detsim_hitreco__20240222T021847Z_reco2.root_1709147143
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_539_20230825T19162
4Z_gen_g4_detsim_hitreco__20240222T001426Z_reco2.root_1709156047
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_557_20230825T19170
6Z_gen_g4_detsim_hitreco__20240221T163145Z_reco2.root_1709147907
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_584_20230825T19162
7Z_gen_g4_detsim_hitreco__20240221T140103Z_reco2.root_1709155474
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_631_20230825T19164
9Z_gen_g4_detsim_hitreco__20240221T163144Z_reco2.root_1709147400
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_676_20230825T19163
3Z_gen_g4_detsim_hitreco__20240222T001459Z_reco2.root_1709148321
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_741_20230825T19163
6Z_gen_g4_detsim_hitreco__20240222T001437Z_reco2.root_1709149056
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_787_20230825T19163
9Z_gen_g4_detsim_hitreco__20240222T001416Z_reco2.root_1709148376
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_832_20230825T19164
6Z_gen_g4_detsim_hitreco__20240221T163143Z_reco2.root_1709147818
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/tape_backed/dunepro//fardet-hd/full-reconstructed/2024/mc/out1/fd_mc_2023a_reco2/00/00/14/04/anie_dune10kt_1x2x6_1404_880_20230825T19164
8Z_gen_g4_detsim_hitreco__20240222T150417Z_reco2.root_1709148043
```

FD2-VD

```
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/88/0d/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_490_20230807T0003
35Z_gen_g4_detsim_hitreco__20240222T175513Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/42/f8/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_5_20230806T233235
Z_gen_g4_detsim_hitreco__20240222T225631Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/32/9e/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_550_20230807T0005
38Z_gen_g4_detsim_hitreco__20240222T225641Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/e1/63/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_55_20230806T23353
4Z_gen_g4_detsim_hitreco__20240222T225638Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/27/78/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_578_20230807T0006
35Z_gen_g4_detsim_hitreco__20240222T171329Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/8d/86/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_599_20230807T0008
35Z_gen_g4_detsim_hitreco__20240222T171253Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/66/cc/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_6_20230806T233234
Z_gen_g4_detsim_hitreco__20240222T171329Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/0d/82/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_68_20230806T23363
5Z_gen_g4_detsim_hitreco__20240222T175108Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/31/b7/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_7_20230806T233236
Z_gen_g4_detsim_hitreco__20240222T174952Z_reco2.root
root://fndca1.fnal.gov:1094/pnfs/fnal.gov/usr/dune/persistent/staging/fardet-vd/0c/dd/anu_numu2nue_nue2nutau_dunevd10kt_1x8x6_3view_30deg_979_731_20230807T0017
35Z_gen_g4_detsim_hitreco__20240222T171325Z_reco2.root
```

FD1-HD

```
TrigReport ----- Event summary -----  
TrigReport Events total = 1000 passed = 1000 failed = 0  
  
TrigReport ----- Modules in End-path -----  
TrigReport      Run      Success      Error Name  
TrigReport      1000      1000          0 cafmaker  
  
TimeReport ----- Time summary [sec] -----  
TimeReport CPU = 305.405268 Real = 397.244412  
  
MemReport ----- Memory summary [base-10 MB] -----  
MemReport VmPeak = 2735.48 VmHWM = 1102.97  
  
Art has completed and will exit with status 0.
```

FD2-VD

```
TrigReport ----- Event summary -----  
TrigReport Events total = 1000 passed = 1000 failed = 0  
  
TrigReport ----- Modules in End-path -----  
TrigReport      Run      Success      Error Name  
TrigReport      1000      1000          0 cafmaker  
  
TimeReport ----- Time summary [sec] -----  
TimeReport CPU = 138.390197 Real = 181.719808  
  
MemReport ----- Memory summary [base-10 MB] -----  
MemReport VmPeak = 2601.75 VmHWM = 968.442  
  
Art has completed and will exit with status 0.
```

Output caf.root ~ 1.9 MB

Both ways a priori work

- Method 1 → all output small files to be merged
- Method 2 → need to test a workflow submission on justIN, to identify an “ideal size” of the input list
small number of files to be merged

in both cases:

- Define a naming convention for *caf.root* files
- Metadata ?
 - define metadata for single files and for merged files
 - parents?

It is not possible right now to answer to these questions:

- Next week Heidi can present her merge program
- Need to understand data management needs/constraints

To discuss and share information/ideas:

Issue open on github: <https://github.com/DUNE/dune-prod-utils/issues/4>

Link to the [google doc](#)