

# CAF files production for LBL and atmospheric productions

([google doc](#))

- “standard campaign” input file → one (or more) output file(s)
- CAF campaign: one input file ~1GB producing 1 output file ~ 270KB in 20 sec
- Need to reduce number of small output files to be handled → 2 steps of “merging”
  1. run lar on Ni input files producing 1 output caf file (Ni to be quantified)
  2. merge No output caf files, to get a large of “acceptable” size to be put on tape

(CRAB meeting August 23)

- To maximize efficiency we are going to run files at the closest site to the data, e.g. atmospheric sample is located at Prague therefore they would be process locally at Prague

Following slides present preliminary results on step 1

→ the goal is to define  $N_i$  , check and validate metadata

Workflow id	Input dataset to justIN	input list to lar (« max » value)	# output files	Output dataset size
2950	1000 files	10 files	110	160 MB
2954	2000 files	10 files	215	338 MB
2959	1000 files	20 files	74	140 MB
2960	5000 files	20 files		

*mainly related to step 1  
Finalization of jobscript  
Impact on justIN, metacat,...*

*related to step 2  
Need to test merging files (and MD) script*

## Example:

```
justin simple-workflow --mql "$MQL_QUERY2000" --jobscript caf_10_atmo.jobscript --max-distance 0 --rss-mib 4000 --env INPUT_DIR=/cvmfs/fifeuser1.opensciencegrid.org/sw/dune/464abc63152e6d913844675d99c8988ab313ea1c --scope usertests --lifetime-days 10 --output-pattern 'caf_*.root:fardet-hd-caf_$JUSTIN_WORKFLOW_ID'
```

## Jobscript:

lar command →

```
lar -c "${CAF_FCL}" -S file.list :
```

## metadata generation:

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

```
ls -rtl /cvmfs/fifeuser1.opensciencegrid.org/sw/dune/464abc63152e6d913844675d99c8988ab313ea1c
```

```
total 33
```

```
-rwxr-xr-x 1 cvmfs cvmfs 393 Jul 24 19:55 CheckSum.py
```

```
-rwxr-xr-x 1 cvmfs cvmfs 2285 Jul 24 19:55 CheckConfiguration.py
```

```
-rwxr-xr-x 1 cvmfs cvmfs 21087 Jul 24 19:55 mergeMetaCat.py
```

```
-rwxr-xr-x 1 cvmfs cvmfs 6107 Jul 24 19:55 TypeChecker.py
```

```
-rwxr-xr-x 1 cvmfs cvmfs 2169 Jul 24 19:55 TimeUtil.py
```

```
-bash-4.2$
```

(scripts by Heidi)

## Good efficiency, but output files not located in a unique RSE

## 2950 RSEs used

Name	Inputs	Outputs
PRAGUE	1007	94
RAL_ECHO	107	11
DUNE_ES_PIC	14	5

## File states per stage

Stage ID	Files	Finding	Unallocated	Allocated	Outputting	Processed	Not found	Failed
1	1000	0	0	0	0	997	0	3

## 2954 RSEs used

Name	Inputs	Outputs
PRAGUE	1927	177
RAL_ECHO	267	26
DUNE_FR_CCIN2P3_DISK	82	8
DUNE_ES_PIC	30	4

## File states per stage

Stage ID	Files	Finding	Unallocated	Allocated	Outputting	Processed	Not found	Failed
1	2000	0	0	0	0	2000	0	0

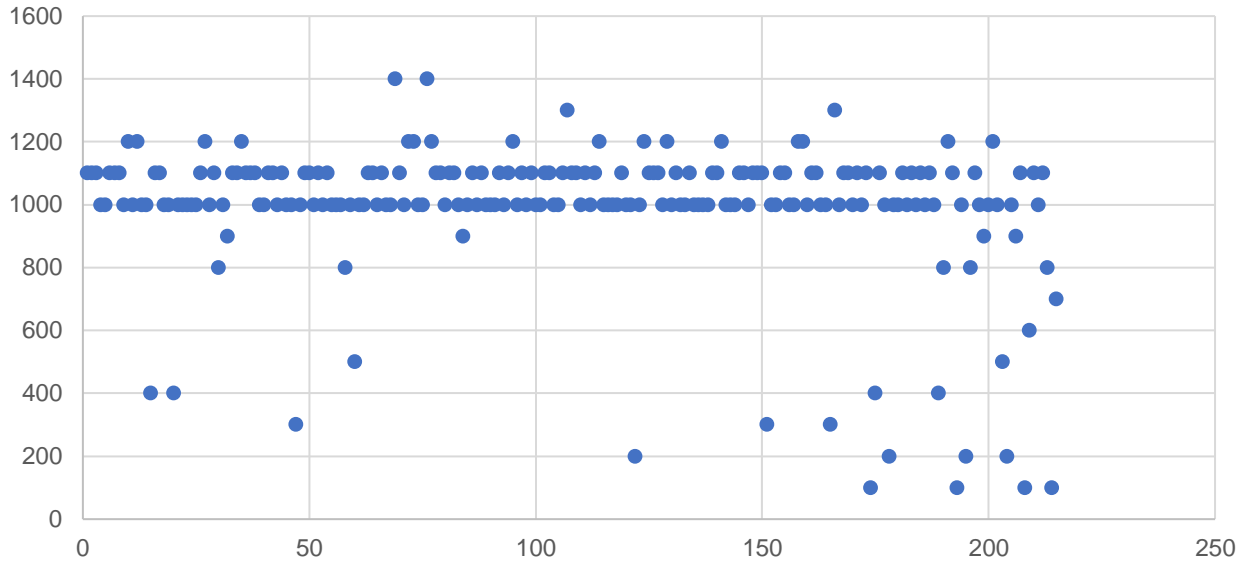
## 2959 RSEs used

Name	Inputs	Outputs
PRAGUE	1000	54
RAL_ECHO	97	15
DUNE_FR_CCIN2P3_DISK	24	2
DUNE_ES_PIC	12	3

## File states

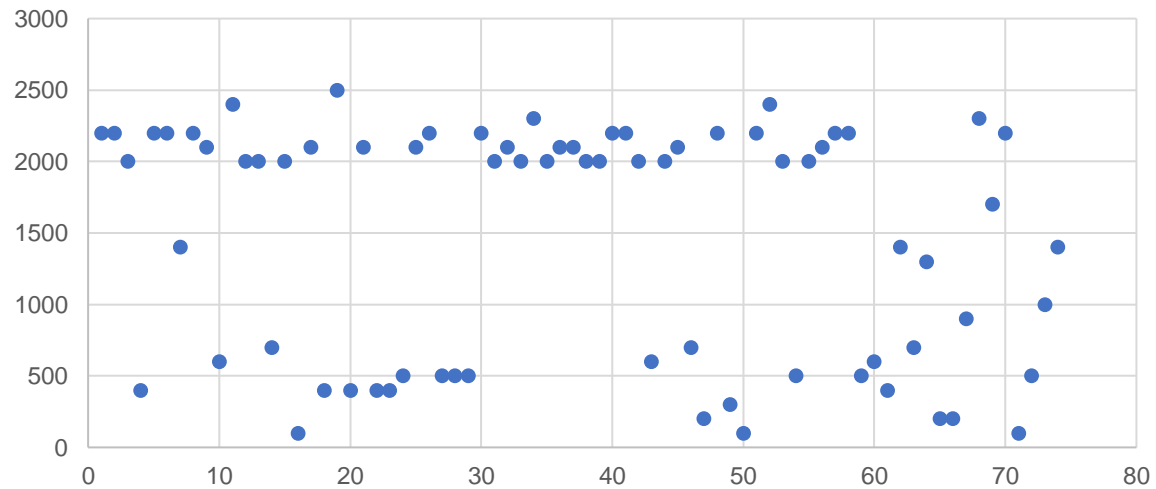
Total files	Finding	Unallocated	Allocated	Outputting	Processed	Not found	Failed
1000	0	0	0	0	1000	0	0

Workflow ID 2954



Output files can have different number of events

Workflow ID 2959





# [metadata](#) example (usertests:caf\_20240828T200714Z.root)

## File Information

**File ID** [vE4NcMNwQH22kGwb](#) 

**Namespace** [usertests](#) 

**Name** [caf\\_20240828T200714Z.root](#) 











**DID** [usertests:caf\\_20240828T200714Z.root](#) 

**Size** 1835562 (1.751 MB)

**Created** 2024-08-28 15:31:37.605377-05:00 by [dunepro](#)

**Updated** 2024-08-28 15:32:07.926706-05:00 by [dunepro](#)

**Checksums** adler32: 4bd109da

**Parents** [fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_50572032\\_48\\_20231203T040021Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240508T075221Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_50649927\\_202\\_20231208T045017Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240510T065156Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_6171714\\_125\\_20231126T014219Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240507T182658Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_6374820\\_330\\_20231201T083159Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240507T203533Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_6406500\\_16\\_20231202T122929Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240507T220722Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_6491894\\_470\\_20231208T004703Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240510T031250Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_66049673\\_730\\_20231207T160046Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240510T045525Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_74494003\\_838\\_20231201T170358Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240507T201530Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_74496296\\_863\\_20231201T213352Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240508T051330Z\\_reco2.root](#)   
[fardet-hd:atmnu\\_max\\_weighted\\_randompolicy\\_dune10kt\\_1x2x6\\_74511302\\_107\\_20231203T125127Z\\_gen\\_g4\\_detsim\\_hitreco\\_20240508T073206Z\\_reco2.root](#) 

## Children

## Metadata

Parameter	Value
art.file_format_era	ART_2011a
art.file_format_version	15
art.process_name	Reco2
core.application	art.reco2
core.application.family	art
core.application.name	reco2
core.application.version	v09_85_00d00
core.data_stream	physics
core.data_tier	root-tuple
core.end_time	1724877059
core.event_count	1100
core.file_format	root
core.file_type	mc
core.first_event_number	1601
core.group	dune
core.last_event_number	86400
core.run_type	fardet-hd
core.runs	[50572032, 50649927, 6171714, 6374820, 6406500, 6491894, 66049673, 74494003, 74496296, 74511302]
core.runs_subruns	[5057203200001, 5064992700001, 617171400001, 637482000001, 640650000001, 649189400001, 6604967300001, 7449400300001, 7449629600001, 7451130200001]
core.start_time	1724877059
dune.campaign	fd_mc_he_2023a_reco2
dune.config_file	reco2_atmos_dune10kt_1x2x6_geov5.fcl
dune.output_status	confirmed
dune.requestid	ritm2032831
dune.workflow	{'user': 'epennacc@fnal.gov', 'job_id': '63050.31@justin-prod-sched02.dune.hep.ac.uk', 'cpuinfo': 'AMD EPYC 7763 64-Core Processor', 'hostname': 'dune001-1967402.0-lcg2674.gridpp.rl.ac.uk', 'stage_id': 1, 'site_name': 'UK_RAL-Tier1', 'os_release': 'Scientific Linux release 7.9 (Nitrogen)', 'workflow_id': 2954, 'jobscript_start': 1724875620, 'jobscript_finish': 1724877065, 'jobscript_cpu_seconds': 309, 'jobscript_real_seconds': 1444}
dune_mc.detector_type	fardet-hd
dune_mc.gen_fcl_filename	prodgenie_atmnu_max_weighted_randompolicy_dune10kt_1x2x6.fcl

