LE production

Elisabetta Pennacchio
Production meeting
September 26th

2) LE request : ticket, footprint

- 14 samples, required statistic is different for each sample (from 0.2M to 2.4M)
- FD1-HD, FD2-VD, with and without radiological background
- generation, g4stage1, g4stage2, detsim, reco1. Further processing once Pandora traning is complete
- 9M of events corresponding to 13 samples, 100M for one sample at generation stage, filtered to 1M after G4
- expected data volume: ~1.1 PB
- validation sample : 50k events for each sample (to allow training) , ~55TB

Plan:

- prepare jobscripts, run preliminary tests to validate workflows and metadata
  → ~1 week (update at the production meeting next week)

- next week during the production meeting we will discuss how to organize and run the campaign. Help is needed: call for shifters

Where we are:

- The samples to be generated correspond to FD1-HD, FD2-VD, with and without radiological background.

- For FD1-HD there are also two different configurations: *centralAPA* and *lateralAPA*, requiring not only a different generation fcl file, but also a different g4 fcl file

- FD1-HD workflows have 4 processing steps: *gen g4, detsim, reco*
- FD2-VD workflows have 5 processing steps: *gen, g4 stage1, g4 stage2, detsim, reco*

- Event size is very different among samples. For 10 events/output file (here)

  - ✓ No background: ~8MB (both geometry)
  - ✓ Background: FD1-HD ~500MB (central APA) , 350MB (lateralAPA)
  - ✓ FD2-VD ~2.5 GB

→ The number of events/file has to be different among samples. To get output file of a "reasonable" size

  - ✓ no background: 2000 events/file (4 samples) *2000x8MB/10 ~1.6 GB is it feasible?*
  - ✓ background FD1-HD 50 events/file (6 samples) *50x350(500)MB/10 ~ 1.7(2.5) GB*
  - ✓ FD2-VD 10 events/file (3 samples) *2.5 GB*

Implementation

- working dir*:* ***/exp/dune/app/home/dunepro/le_ritm2205749***

```
[dunepro@dunegpvm05 le_ritm2205749]$ pwd
/exp/dune/app/home/dunepro/le_ritm2205749
[dunepro@dunegpvm05 le_ritm2205749]$ ls -rtl
total 1
-rw-r--r-- 1 dunepro dune 263 Sep 24 13:40 SL7.sh
-rw-r--r-- 1 dunepro dune 387 Sep 24 13:40 setup-rucio-metacat.sh
drwxr-xr-x 2 dunepro dune   9 Sep 25 03:58 fdvd_marley_cc
drwxr-xr-x 2 dunepro dune   5 Sep 25 05:10 fdvd_radiological_marley_cc
drwxr-xr-x 2 dunepro dune   5 Sep 25 06:41 fdhd_radiological_central_marley_cc
drwxr-xr-x 2 dunepro dune   5 Sep 25 06:52 fdhd_radiological_lateral_marley_cc
drwxr-xr-x 2 dunepro dune   5 Sep 25 06:59 fdvd_marley_es
drwxr-xr-x 2 dunepro dune   5 Sep 25 07:15 fdhd_marley_cc
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:05 fdhd_marley_es
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:12 fdvd_radiological_marley_es
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:18 fdhd_radiological_central_marley_es
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:21 fdhd_radiological_lateral_marley_es
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:25 fdvd_radiological
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:30 fdhd_radiological_central
drwxr-xr-x 2 dunepro dune   5 Sep 25 08:35 fdhd_radiological_lateral
drwxr-xr-x 2 dunepro dune   5 Sep 25 09:52 fdhd_neutrons
[dunepro@dunegpvm05 le_ritm2205749]$
```

14 subdirs
one subdir/sample

How a directory looks like:

```
[dunepro@dunegpvm05 fdvd_marley_es]$ ls -rtl
total 24
-rwxr-xr-x 1 dunepro dune  4424 Sep 24 09:29 pdjson2metadata
-rw-r--r-- 1 dunepro dune   804 Sep 25 06:54 le_input.json
-rw-r--r-- 1 dunepro dune 10240 Sep 25 06:55 input.tar
-rw-r--r-- 1 dunepro dune  5731 Sep 25 12:24 le_fdvd_marley_es.jobscript
-rw-r--r-- 1 dunepro dune   936 Sep 25 12:25 submit.log
-rw-r--r-- 1 dunepro dune   577 Sep 26 03:47 fdvd_marley_es.json
[dunepro@dunegpvm05 fdvd_marley_es]$
```

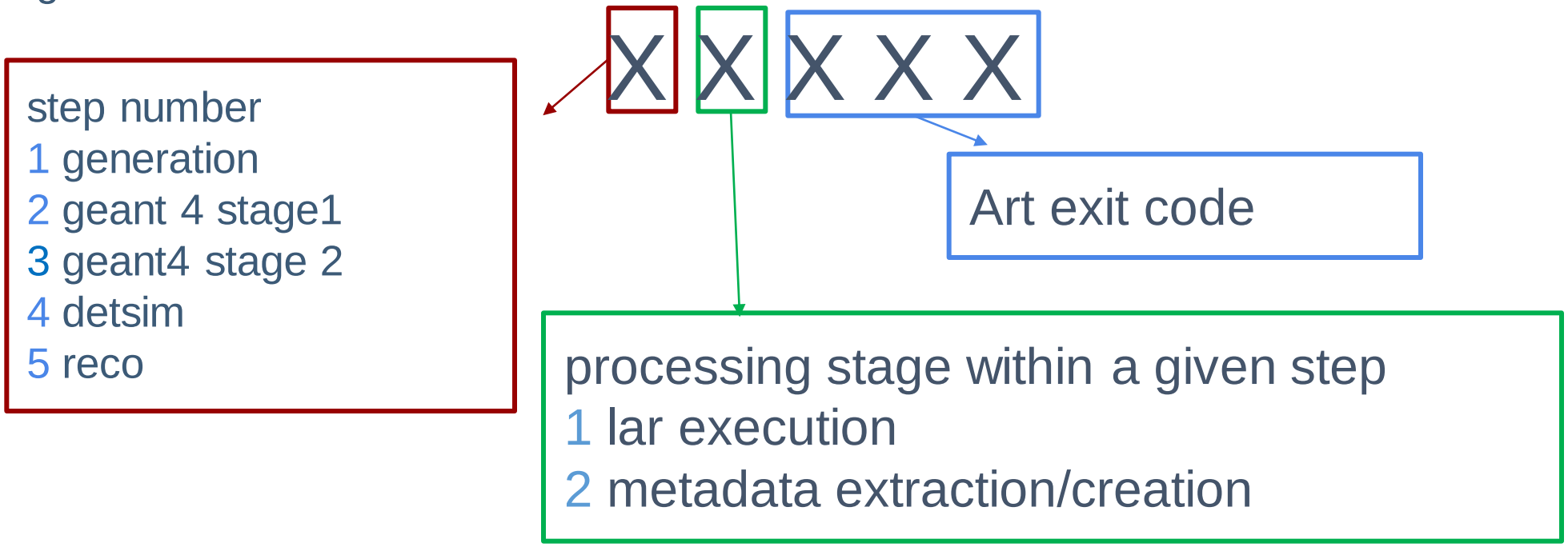
1. metadata generation

2. jobscript

3. json file used to create dataset used now to test metadata

```
{
  "namespace" : "usertests",
  "core.group" : "dune",
  "core.run_type": "fardet-vd",
  "dune.campaign": "le_mc_2024a",
  "core.application.version":"v09_91_04d00",
  "core.application.name": "reco",
  "core.data_tier": "hit-reconstructed",
  "core.file_type":"mc",
  "dune.config_file": "standard_reco1_dunevd10kt_1x8x14_3view_30deg.fcl",
  "dune.requestid": "ritm2205749",
  "dune_mc.detector_type": "fardet-vd",
  "dune_mc.gen_fcl_filename": "prodmarley_nue_es_flat_dunevd10kt_1x8x14_3view_30deg.fcl",
  "dune.output_status" : "confirmed",
  "core.data_stream" : "out1"
}
```

To make debug easier, a system of error codes is implemented in the jobscript.
A 5 digits exit code has been defined:



Error messages during execution:

ERROR: lar (generation) exit code: 11090
ERROR: lar (geant4 step1) exit code: 21090
ERROR: lar (geant4 step2) exit code: 31090
ERROR: lar (detsim) exit code: 41090
ERROR: lar (reco) exit code: 51090
ERROR: metadata generation 52002

Results

- All 14 jobscripts have been tested interactively, then workflows have been submitted to justIN to validate also metadata generation.
- Jobs generating 10 events have been submitted, to double check output file size and execution time.
- In the next page a summary table is presented for 13 samples over 14.
- Samples are presented in descending order of priority (ticket). For the last sample the jobscript has been prepared and tested as well, but results are not included in the table, because the validation sample of 50k events is nor required
- For each sample, a link to the relevant justIN workflow(s) is(are) presented

Summary table

1	FD2-VD MARLEY CC (0.8 M)	10	3404	5600<t<11000 sec	~8.5 MB
2	FD1-HD MARLEY CC (0.8M)	10	3435	~1600 sec	~6 MB
3	FD2-VD Radiological background + MARLEY CC (2.4M)	10	3421	~4 h	~3 GB
4	FD1-HD Radiological background central APA + MARLEY CC (0.8M)	10 50	3436 3463	~1.5 h ~7,30 h	~600 MB ~3 GB
5	FD1-HD Radiological background lateral APA + MARLEY CC (0.8M)	10 50	3437 3486	~1h	~370MB
6	FD2-VD MARLEY ES (0.2M)	10	3424	6000<t<10000 sec	~7.5 MB
7	FD1-HD MARLEY ES (0.2M)	10	3488	~1700 s	~5MB
8	FD2-VD Radiological background + MARLEY ES (0.6M)	10	3426	~4h	~3GB
9	FD1-HD Radiological background central APA + MARLEY ES (0.2M)	10 50	3439 3489	~2h	~300MB
10	FD1-HD Radiological background lateral APA + MARLEY ES (0.2M)	10 50	3440 3490	~1h	~370MB
11	FD2-VD Radiological background (1M)	10	3429	~4h	~3GB
12	FD1-HD Radiological central APA (0.5M)	10 50	3441 3491	~2h	~600MB
13	FD1-HD Radiological lateral APA (0.5M)	10 50	3456 3492	~1h	~360MB

What's next (before moving to validation samples production)

- Once running workflows finished, confirm that 50 event/job are ok (blue lines)
- How to proceed for the samples without background?
- Check carefully metadata: are all needed metadata included? Are there any errors?
- Shifters are needed: any volunteer? I can help in submitting and monitoring some workflows but not all
  contact me and Aaron

- A google doc has been prepared:
  - I'll add the table at page 9
  - please add your comments/suggestions

Comments here will be the starting point for the meeting next week