CAF production and LE request: planning for upcoming weeks

Production meeting, September 19th, 2024

Aaron Higuera, Rice University Elisabetta Pennacchio, IP2I 1. CAF production (13 samples, ~392K files to process) google doc

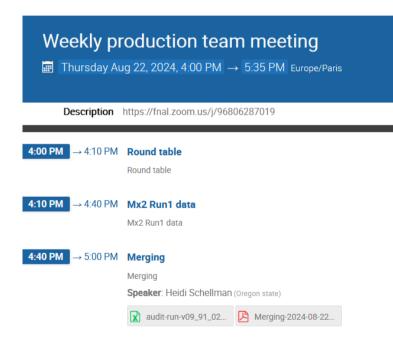
- shifters: Aaron Elisabetta
- where we are: next presentation (~392K files to process, 150k files already processed)
- what's next: 1) process FD1-HD and FD2-VD samples (12 samples, ~250k files)
 - minor modifications of the jobscript already used ~ 1 day (including one test workflow)
 - 2. need to identify sites where to run for the input samples $(6+6) \sim \text{timescale}$?
 - 3. once defined, 12 datasets can be processed in ~ 1 week, 10 days

2. and 3. can be done in parallel

	site
atmospheric	Prague
FD1-HD	
FD2-VD	

2) prepare next step: further merging of caf output files to obtain files of « good » size to be copied on tape (~1GB).

- Starting point: <u>this presentation</u>
- Heidi already prepared scripts used to merge calibration output from keep-up processing → identify what has to be changed in order to use these scripts also for CAF files timescale?



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
 \rightarrow ~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