Data Processing: Release 4 readiness

Stefano Lacaprara

Release 4

- Updated production script for new interface
    - https://agira.desy.de/browse/BIIDP-1776
        - **change GT handling in release 4**
            - **Support multiple GT in json (previously only one, the rest hardcoded)**
        - Remove SetDataFlagModule (not needed anymore)
            - isMC: 0
        - add SuppressErrorMask=0xFFFFFFFFFFFFFFFF to PXDUnpacker
        - EventsOfDoomBuster now in add_reconstruction()
        - Better handling of path for Cosmic
            - /group/belle2/dataprod/Data/release-04-00-00/DB00000711/Unofficial/e0010/**Cosmic**/
        - Other minor improvment.
- Question: path is now: /path/<release>/<**GTID**>/<Campaign>/<EXP>/<Type>
    - But we are not using a single GTID anymore
        - Eg: **data_reprocessing_prompt_rel4_patchb** and **online**
    - I'm using the first as GTID in the path, but that ID is not telling the full story. Should we revise it?
        - globalTag: data_reprocessing_prompt_rel4_patchb,online

Stefano Lacaprara INFN Padova

Release 4 test on collision data

- Tested on several exp 8 runs
  - https://agira.desy.de/browse/BII-5621
  - Found a number of [ERROR]
  - Full list in jira ticket
    - [ERROR] Correction factor=0 is very small/too large! Resetting to 1.0. { module: ECLShowerCorrector } ... message repeated 49 times [INFO] Processed: 1 runs, 2 events [ERROR] Correction factor=0 is very small/too large! Resetting to 1.0. { module: ECLShowerCorrector } https://agira.desy.de/browse/BII-5446
      - Patch forgotten in release-4
    - [ERROR] DHP data loss (CM=63) in 1220527 times, on 4 runs: 00224 00301 00799 02430 https://agira.desy.de/browse/BII-5622
      - ERROR should have been WARNING
    - Other about event too crowded for reconstruction or corrupted data

No major issues.

Release 4 test on Cosmic

- Tested on several exp 10 cosmic runs
  - **[FATAL] ERROR_EVENT : Invalid header size of FTSW data format(= 0x00000008 words). Exiting…**
    - https://agira.desy.de/browse/BIIDP-1932
    - inconsistency between RawFTSW unpacker merged on Aug. 20 and data which Nakao-san started sending with an updated data-format
  - https://agira.desy.de/browse/BII-5654
  - From run 1500 Nakao-san restored original RawFTSW format / version number, until the RawFTSW unpacker is further updated to be compatible
    - Tested and confirmed
    - RawFTSW is used to remove random trigger from processing
      - in principle we can use any RawXXX (all contains trigger type), but not always present.
  - **Error in <TStreamerInfo::Build>: Belle2::PXDRawROIs, discarding: int* m_rois, no [dimension]**
    - [Bjoern] Did someone **AGAIN** killed the root streamer by changing the comment to be "doxygen" conform. wow. i am impressed.
      - https://stash.desy.de/projects/B2/repos/software/pull-requests/5071/overview

Proc10 cdst processing

- Prepared json for cdst processing
  - Reminder: release4, only phase III data. Exp7+8
    - All runs (or only good ones? Maybe some can be magically recovered with new processing...)
  - https://agira.desy.de/browse/BIIDP-1928
  - https://stash.desy.de/projects/B2P/repos/data/pull-requests/73/overview
- Waiting for calibration GT from calib group
  - Should we wait also for patch release?
    - [ERROR] Correction factor=0 is very small/too large! Resetting to 1.0. { module: ECLShowerCorrector }
      - need to be fixed for proper ECL calibration maybe?

- Time estimate
  - Cdst processing for selected HLT skims (hadron / gamma-gamma / bhabha / mumu)
    - 11,000 hCPU / fb-1
    - With 400/1000/1500/2000 job slots: 1.2 / 0.5 / 0.3 / 0.23 day / fb-1 (we have 400 on **b2_prod**, more on **l**)
    - L=6.4 /fb (~6.0 /fb good runs)
      - Only b2_prod: 8 days
      - **With 1000 jobs, about 3 days (+1 contingency ?)**

To be discussed at B2GM

- **cdst processing via airflow?**
  - Oh, yeah!
    - Maybe at BNL/desy as well?

- **Good runs on the grid: how?**
  - We have subdir at KEKCC, so you can process only the good ones
    - /group/belle2/dataprod/Data/release-03-02-02/DB00000654/proc9/e0007/4S/**GoodRuns**
  - Not on the grid!
    - Dataset contains all runs, so it is up to the user to remove (a posteriori) the bad ones

- **HLT skim on the grid: how?**
  - Now we run all events on the grid, no HLT skim is available
    - Not a good way to encourage people to run on the grid!
  - At KEKCC: we run PromptSkim at quasi-online
    - We process the HLT raw skims
    - We process all events
  - On the GRID:
    - Full raw are copied to SE and then processed.
      - **a)** HLT skim full processing
      - **b)** copy to SE raw HLT skim (and process them)
      - **c)** run the HLT skim RAW->RAW on the grid (and process them)

Stefano Lacaprara INFN Padova