Real Time and Trigger Analysis (RTTA) Issues
Analysis and Ecosystem Workshop

"Stone Soup" Discussion

Contributors so far: Mike Sokoloff and Giulio Eulisse (session conveners), Graeme Stewart, Caterina Doglioni, Marco Montella, Conor Fitzpatrick, Daniel Craik, Eduardo Rodrigues [add your name here]

General challenges
What are the AE challenges that come from doing Real Time & Trigger Analysis?

For heterogeneous hardware in triggers, how do we (the WLCG) guarantee that trigger software and offline software produce the same (indistinguishable) results?? What level of difference is tolerable, and what tests do we need to demonstrate that we can tolerate the differences?

Analysis moving towards smaller and smaller data formats. How to be sufficiently flexible and accommodate the following needs:

- RTTA in ATLAS and CMS generally requires some re-calibration prior/during analysis workflow, so it still needs intermediate larger data formats with appropriate variables
    - More information on ATLAS calibration for RTA (slight changes expected for Run-3 but the overall idea is there) in this talk.
- Understanding anomalous events requires larger data formats (coming from dedicated triggers)

Expansions of quasi-real-time alignment and calibration efforts
How does this interact with MC in terms of Analysis Ecosystems?

ATLAS (potentially similar for CMS but CMS experts should confirm):

- Alignment: what can't be done at the level of the trigger can be neglected for analysis use cases
- Calibration: [case of jets] not all calibration steps can be done in (quasi-)real-time since they use lower-statistics processes and need (almost) the whole dataset to be derived prior to application. Could think of iterative calibration in the trigger for at least central-forward calibration.

LHCb:

- What is different is how the experiment uses this rather than how it is "done"

- MC is not produced with run block specific calibrations/alignments, it is produced with a baseline alignment typically for the entire year/data taking period. Data is aligned/calibrated to within a predetermined acceptable margin of this. Ties into 'what is sufficient' discussion on TLA.

ALICE:

- Similar to LHCb (two pass reco);
- particularly TPC needs good conditions to reconstruct - Graeme's guess! Giulio should comment

Expansions of "Quality Control"

studies performed in real time to better understand both detector performance and particle physics questions;

ATLAS:

- In Run 3, aim to test entire workflow from raw data to plots in online and offline monitoring using the separate RTA stream [→ how to handle different streams in analysis facilities?]
    - Run-2: only tested events that would have ended up in the stream, coming from the standard stream

LHCb: [to be edited by Dan Craik, Hlt1]

- [Mike will find someone to fill this in more generally for Hlt2, maybe Nicole]
- RTA shifter for alignment and calibration [Sylvia B]; (RTA shifter for trigger code)

Expansions of real-time reconstruction of objects to be used in offline analysis;

ATLAS/CMS:

- Run-3 plan for a wider set of objects to be reconstructed (Run-2: jets for ATLAS and CMS and muons for CMS). Source: talks at https://indico.cern.ch/event/1007936/

Analysis workflows instantiated in the trigger software

E.g. make histograms, Dalitz plots, whatever that are persisted for later analysis in place of
persisting the underlying events;

ATLAS:

- There is in principle the functionality to do this in the L1 trigger (https://arxiv.org/abs/2105.01416)
- Challenges of doing this at HLT
    - Performance limitation (see calibration point above)
    - CPU cost limitations often surpass storage limitations, so no incentive to do histograms if the event size is negligible → can accelerators help?
- If this is done, think how to reconcile making final plot with usual blind analysis
    - E.g. make plot but don't look at it until the end of data taking or similar

LHCb: Dan Craik, parallel session talk for Hlt1;
Anyone for Hlt2? [Patrick Robbe et. al developing automated analysis in Monet (online monitoring)]

Could ML instantiated in accelerators help? If so, would WLCG need accelerators?

CMS/inter-experiment: see https://indico.fnal.gov/event/46746/contributions/210997/attachments/141316/177896/hls4ml-cpad21-ngadiuba.pdf

Anomaly detection-like trigger / RTA streams

Topic under development…assuming an algorithm exists, then may need to:

- Write small (very small) records for all (or many random, unbiased) events to eliminate anomalies other than the interesting ones
    - How unbiased is a zero-bias trigger?
- If "anomaly" detected, turn on a dedicated trigger stream
- Record events that have been rejected by all triggers.

Outlier detection for monitoring in CMS: https://arxiv.org/abs/1808.00911
Ongoing challenge (DarkMachines/ML4Jets): https://indico.cern.ch/event/980214/contributions/4413658/attachments/2278124/3870358/ml4jets_data_challenge.pdf

Usage of accelerators in online environments.

Inform HL-LHC developments with Run-3 experiences in CMS (Patatrack) and LHCb (HLT1).
For potential whitepaper, report on benchmarking against CPU.

—---------- this space reserved for discussion May 24 —----------

For LHCb, how “negligible” are differences between Hlt1 and Hlt2 reconstructions? What are the associated systematics for dark photon searches in Hlt1?

Is there (what is) the method for doing a blind analysis using data collected during the trigger? What are the control samples?

DC: in principle, Gaudi probably supports n-dimensional “histograms”

Simulation of workflows on GPU (/FPGA or other hybrid architectures) on WLCG:

- Not a must (not cost effective right now) but we can profit from them when they arrive
    - More workflows (generation/simulation/?) will be on GPUs so why not