Trigger Input to First-Year Analysis Model Working Group

And some soul searching…

Trigger Open Meeting – 29 July 2009

Introduction

- The First-Year Analysis Model (FYAM) working group is planning the use of data formats and computing resources for the coming year
  - https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AnalysisModelFirstYear
- Joerg Stelzer reporting today to the FYAM on the trigger needs
  - Input collected from each slice in this wiki: https://twiki.cern.ch/twiki/bin/view/Atlas/TriggerFirstYearAnalysisModel
  - I'll show Joerg's slides next
- Some questions need to be further developed, like how to distribute the data to ease efficient analysis
- This talk aims to keep you informed of the overall trigger needs and to generate discussion

Summary of Input from Trigger Slice Groups

- Input from triggers
  - Egamma, Muon, Tau, MET, Jet, Inner Detector
  - Thanks to Margherita Primavera, Andrew Hamilton, Stefania Xella, Diego Casadei, Patricia Conde Muino, John Baines,

- Discussion with John Baines, Ricardo Goncalo, Stefania Xella, Cristobal Padilla, and others. Thanks as well.

- Questionnaire about data access

https://twiki.cern.ch/twiki/bin/view/Atlas/TriggerFirstYearAnalysisModel

Info Collected from Trigger Slices

- Categorize data access by
  - Purpose
    - Trigger Tuning
      - LUT tables, cut tuning, etc., not trigger algorithm development
    - Performance Monitoring
      - Run by run monitoring: use of online and offline histograms by shifter
    - Performance Debugging
      - In case histograms (or other analysis) indicate performance issues
    - Efficiency calculation
  - Phase
    - Phase 1: HLT @ CAF only (1-2 weeks)
    - Phase 2: HLT @ P1 recording / no rejection (depends on lumi and trigger understanding)
    - Phase 3: HLT @ P1 triggering
    - Phase 4: stable trigger performance

- Needs for data production / access by trigger slice
  - Data type: RAW, ESD, AOD, Trigger Ntuple, Histograms
  - Access delay
  - Data volume, stream, selectiveness (on demand, certain percentage)

- How to summarize: by slice, phase, data type
  - By data type: easy to compare to Samira's summary of Perf. Group requirements

Phase 1: HLT @ CAF

- Special: trigger runs on all RAW data at CAF
  - All RAW data transferred to CAF
  - RAW to RAW-with-trigger
  - RAW-with-trigger to ESD
  - ESD to NT (contain trigger + offline data)
    - Storage of ESD and/or ntuples on Trigger Group space
  - Setup same as later at Tier0
- All streams needed
  - In this phase we need all streams at the CAF, need to check the trigger performance on all data
- Phase used to study the robustness of the trigger on real data
  - Also verify that trigger NT production, monitoring, performance checks, bug response, etc. work to satisfaction

Phase 2,3,4: HLT @ P1

- HLT result calculated at point 1 and contained in the RAW data
  - Processed at Tier0, where we expect production of Trigger NT
- Phase 2:
  - Trigger online, but not rejecting: vital to validate performance as quickly as possible so as to be able to switch on active trigger when needed - requires full statistics to be able to validate chains as quickly as possible.
- Phase 3:
  - Trigger online & active for the first time: need immediate (<24hrs) feedback for validation of triggers and spotting problems
- While online & tier0 monitoring will spot gross problems, detailed validation requires the NT/AOD

Access to RAW Data

Debugging, algorithm development, trigger studies (partially) require *running of the trigger on RAW data*. All slices agree that RAW data for the following is needed:

Fast debugging:

On demand access to problematic runs/events at CAF

- Runs that show trigger problems
  - Checking histograms, ntuples, large number of events in the debug stream, etc.
- Includes debug stream
  - Sent to the CAF at any rate
- Immediate access crucial
  - Especially in phases 3 and 4, when the trigger is actively rejecting
  - If problem spotted during data taking: start transfer of data to Tier0, then data to CAF
  - Should be as fast as possible
- Limited to experts

Development:

Trigger reconstruction can only run from RAW data.

- Requires access to a fraction of the RAW data streams
  - Event sample used as "signal"
  - Background sample for rates (e.g. enhanced bias sample used to validate new menus).
- Concern: how is the RAW data distributed and accessed

Access to ESD / PerfDPD Data

- Most Trigger information is in AOD and ESD.
- The exception is detailed LVL1 info and Event Filter tracks, where the TrackParticle but not the tracks are in AOD.
- LVL1 will use DPD. So the main requirement for ESD is for access to Track information. Bjet slice needs some ESD access for performance debugging.
- Jet slice also seem to rely on ESD, PerfDPD for performance analysis

Access to AOD / DPD Data

- AOD will be used for efficiency calculation
  - Trigger D3PD format being defined with the goal of common D3PD format for the trigger (goal to replace AOD for many studies)

AOD / DPD for performance studies:

- Performance validation & measurements will be done from AOD, but mostly not before phase 4.
  - Certainly Egamma, Bphysics, InDet, MET have said so.
  - Muons have no definite plans to move away from NT

- In the long run it is desirable to replace the trigger ntuples produced from ESD with AOD or better D3PD
  - Not possible for commissioning & early running where we need to rely on existing well tested tools.

Trigger NTuples

- Overview
  - Currently 5 slices use trigger ntuples
    - MinBias, Muon, Egamma, Tau
    - MET uses Tau ntuples
  - Inner Detector uses ID commissioning ntuples
  - Jet, BJet, B-Physics do not use ntuple
    - Performance studies on ESD, DPD, sometimes ESD need to be reproduced at CAF (with additional cell-containers)

- Ntuples are established and well tested and have become very sophisticated. They serve many purposes
  - Tuning of triggers
  - Performance debugging

  and are like the 'perfDPD for the Trigger'

- In the initial phases (<4) the ntuples are also important for monitoring
  - It is impossible to predict the conditions and problems that will be encountered in early running. Only experience will show us the full set of histograms needed, the flexibility of ntuples will be vital to producing the immediate feedback needed.
    - E.g. correlation for unforeseen variable pairs, different range/binning than anticipated.

Trigger Ntuples (II)

It is crucial for all triggers to have the ntuples available as quickly as possible. Especially in phases 3 and 4.

- Production of Ntuples at Tier0 (single file)
  - All streams in phase 3. For phase 4 not quite certain yet.
  - Setup for producing the trigger ntuples exists for Tier0 (single file), Tier1, CAF (muon will come today), but needs to be tested

- Analysis of Ntuples at Tier1/2
  - No access to conditions needed (suitable for Tier2)
  - Quick move of Ntuples from Tier0 to Tier1/2 required
  - Dedicated Tier2's might be the preferred model, rather then even distribution

Summary Data Types

- Trigger Validation primarily on **Ntuples**
  - Fast access needed on regular basis
- Trigger bug fixing, performance debugging, and development only with **RAW** data (not ESD)
  - On CAF for bug fixing
    - debug stream, specific events (e.g. BS problems)
  - On Tier1/2 (need COOL access)
- **ESD/perfDPD** not much under consideration yet, except Jet
- **AOD/DPD**: some studies, e.g. efficiencies calculation
  - Trigger specific AODs not yet defined, DPD will probably be preferred
- **D3PD**: Trigger joined the efforts of a common D3PD format
  - Goal to move toward D3PD based studies where applicable (e.g. efficiencies for custom analyses)

Summary (II)

- Main concern is the in-time access to the Ntuples for validation
  - More relaxed when trigger runs stable
  - More tight requirements when new trigger release is deployed

- Tools
  - Recommend use of the TrigDecisionTool for access of trigger data for trigger aware analysis
  - Metadata tools are under scrutiny
    - Trigger info (menu, algorithm setup, rates) for given runs
  - Trigger efficiencies become part of the insitu performance developments
  - Matching of trigger and offline reco objects

Summary (III)

- Similar issues for trigger and detector commissioning
- NTuple distribution to Tier2 and access
- Performance optimization and development of algorithms using RAW data at Tier2
  - Data distribution needs (evenly at Tier2 or dedicated sites) currently under discussion (counts also for ntuples)
  - COOL access, (TriggerDB access)
- Mechanisms need to be tested
  - Tier0/CAF production of ntuples
  - Distribution of ntuples to Tier2
- No validation of AOD content yet
  - Focus primarily on ntuples (or perfDPD) rather than AOD
- Most slices need unbiased samples (can't reduce ESD/AOD/NT significantly to fit size requirements of perfDPD)
- Reprocessing cycles are not an issue

Data access for commissioning

- How to assign data locations most efficiently?

- Grid roles and groups are expected to become active for LHC data taking period
  - Several possibilities exist to distribute data – not clear what will be the outcome in practice
  - E.g. CMS assigns particular Tier2's to physics and reconstruction groups
    - Seems to work well; generated a bloodshed when Tier2 assignments were discussed

- Some RAW data can be made available in Tier2's (How much? For what purpose?)
  - This data can be put in the most convenient Tier2's to provide faster and more convenient/reliable access to people working on particular issues

- There are ~70 Tier2's; not uniform – very variable size/capacity
  - Not efficient to have all data spread through every site
  - E.g. samples from tau stream could go to Niels Bohr Institute

- Commissioning ntuples and D3PDs can be copied to Tier3's, but for how long should a copy be kept (on disk? on tape?)

- Both commissioning ntuples and D3PDs can be re-made from ESD or AOD (probably more efficient to do this centrally by production team)