DATAFLOW IN RUN3

Renaud Amalric - LPNHE (Paris)

01/12/2022
DATAFLOW IN RUN3

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s
0.5-1.5 MHz
70-200 GB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

REAL-TIME ALIGNMENT & CALIBRATION

BUFFER

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS

1.6 GB/s

5.9 GB/s

26% FULL EVENTS

OFFLINE PROCESSING

68% TURBO EVENTS

2.5 GB/s

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
NO MORE HARDWARE TRIGGER

Run I:
- Hardware trigger high \( p_T, E_T \)
- 40 MHz
- 1 MHz
- Time from collision: \( \mu s \)
- 1st software trigger partial reco
- 100 kHz
- Online: near detector
- 2nd software trigger full reco
- 5 kHz
- “Offline”: grid computing
- Reconstruction
  - Align + Calib
- Analysis
  - 5 kHz

Run II:
- Hardware trigger high \( p_T, E_T \)
- 40 MHz
- 1 MHz
- Time from collision: \( \mu s \)
- 1st software trigger partial reco
- 100 kHz
- Online: near detector
- 2nd software trigger full reco
- 100 kHz
- “Offline”: grid computing
- Analysis
  - (Turbo)

LHC BUNCH CROSSING (40 MHz)
- 5 TB/s
- 30 MHz non-empty pp

REAL-TIME ALIGNMENT & CALIBRATION

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)
- 5 TB/s
- 0.5-1.5 MHz
- 70-200 GB/s

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)
- 10 GB/s

OFFLINE PROCESSING
- 1.6 GB/s
- 5.9 GB/s
- 2.5 GB/s

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb
Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
QUICK WORD ON TRIGGERS

→ Gather the information from all the sub-detectors
→ Issue: we produce much more information than what we can actually store to disk to be analyze later on
→ Need triggers to select in real time what you actually want to store and do this as efficiently as possible
→ 2 possibilities
  ➤ Hardware triggers
  ➤ Software triggers

\[ \mathcal{L} = 2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \text{ (ATLAS/CMS)} \]
\[ \mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1} \text{ (LHCb)} \]
\[ \sqrt{s} = 14 \text{ TeV} \]
Why no more hardware triggers?

- Trigger for many hadronic channels saturated already at Run 1–2 luminosity
- Cannot effectively trigger on heavy flavour using hardware signatures

⇒ Fully software triggers
DATAFLOW IN RUN3

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz
70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

BUFFERS

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS

1.6 GB/s

5.9 GB/s

26% FULL EVENTS

2.5 GB/s

68% TURBO EVENTS

OFFLINE PROCESSING

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
FROM THE SUB-DETECTORS TO THE COMPUTING FARMS

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz
70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS

1.6 GB/s

5.9 GB/s

26% FULL EVENTS

68% TURBO EVENTS

OFFLINE PROCESSING

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR

Upgrade Computing Model TDR
New architecture to transmit data collected from every bunch crossing all the way to the event-building computing farms.

The sub-detectors are connected to the data center through long-distance (~250m) optical fibres installed in the PM85 shaft.
To perform the selection, the full-software LHCb trigger requires the complete event information from all the sub-detectors.

Detector’s data received by ~500 FPGAs

Regrouped to the same destination - to one server for event-building.

Event Building: combining the raw data to form single cohesive events

Event builder farm: ~170 servers (with 3 FPGAs each)

Adding GPUs on those servers to apply the first step of selections → HLT1 and then send the data downstream
FROM THE SUB-DETECTORS TO THE COMPUTING FARMS

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz
70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS
1.6 GB/s

26% FULL EVENTS
5.9 GB/s

68% TURBO EVENTS
2.5 GB/s

OFFLINE PROCESSING

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb Upgrade Trigger and Online TDR Upgrade Computing Model TDR
HIGH LEVEL TRIGGER 1

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT
5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)
0.5-1.5 MHz
70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

BUFFER

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)
10 GB/s

OFFLINE PROCESSING

6% CALIB EVENTS
1.6 GB/s

26% FULL EVENTS
5.9 GB/s

68% TURBO EVENTS
2.5 GB/s

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
Fast and partial reconstruction

- Charged particle track and vertex reconstruction
- Electron and muon identification

Up to ~500 GPUs (3 slots available per EB servers)

- Manageable amount of algorithms
- Parallel tasks
- No detailed knowledge of magnetic field & detector required

Software: Allen project
What do the GPUs have to do?
Tracks in VELO expected to be straight lines.

«Search by Triplet» algorithm:
- Find 3 hits in neighbouring VELO modules
- Format the triplet to other layers
- Do this for all the triplets found

Extrapolate the tracks to the beamline: PV’s found that the intersection between the track and the beamline.
The UT allows to reconstruct charged particles which decay after the VELO
- low momentum tracks bending out of the magnetic field region
- Extrapolate the VELO track to the UT silicon strips
- Account for small magnetic field
- Provides a first momentum estimate
- Requires (at least) 3 hits in the UT
→ Forward-tracking
→ Extrapolate the VEL0xUT tracks to the 12 layers of the SciFi
→ Extrapolation using B field’s parametrised trajectory
→ Search of hits from an extrapolated tracks in windows fixed by the momentum estimation in the previous stage
→ Reconstruct tracks with \( p > 3 \) GeV
→ Momentum resolution around 1%
Without the UT, no initial momentum estimate and no information on the charge of the particle

Higher $p$ and $p_T$ requirements

Double search windows around the VELO track extrapolation to identify the charge

Similar reconstruction efficiency for high momentum ($p > 5$ GeV and $p_T > 1$ GeV) but with an increase in the ghost rate

*Another solution also commissioned: Seeding + Matching*
- Used to improve the estimates of the momentum and the track’s impact parameter.

- Method for track fitting, iterates over all hits on a track. For every hit, estimate the state of the track at that location (predictions + measurements).

- Include the previous momentum estimate with the detector description to precisely estimate noise due to multiple scattering and energy loss.

- Gives the best linear estimator for track state.

- At HLT1 level only applied using VELO-parametrisation.
HLT1 SEQUENCE - MUON IDENTIFICATION

→ Forward tracks matched with hits on Muon stations

→ Important for the selection of decays with muons in final state

HLT1 SEQUENCE - CALORIMETER RECONSTRUCTION

→ Look for energetic clusters.

→ Enable’s photon and electron reconstruction at HLT1 level - for the first time.
→ Selecting events with HLT1 trigger lines for different physics purposes

→ Successfully reduced the input rate to reach 1 MHz of output
HIGH LEVEL TRIGGER 1

LHC BUNCH CROSSING (40 MHz)

- 5 TB/s
- 30 MHz non-empty pp

FULL DETECTOR READOUT

- 5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz

- 70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

BUFFER

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

6% CALIB EVENTS

1.6 GB/s

26% FULL EVENTS

5.9 GB/s

68% TURBO EVENTS

2.5 GB/s

OFFLINE PROCESSING

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb Upgrade Trigger and Online TDR Upgrade Computing Model TDR

From 30 MHz to 1MHz
All numbers related to the dataflow are taken from the LHCb:

- Upgrade Trigger and Online TDR
- Upgrade Computing Model TDR
Serves two purposes:

➤ Hold events selected by HLT1 while the Real-Time alignment and calibrations are performed
➤ Allows differ processing of the HLT1 selected events in between LHC fills

Optimal buffer size of around 30PB, which can buffered 80 hours of LHC collisions at an HLT1 1MHz output rate
All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
REAL-TIME ALIGNMENT AND CALIBRATION

LHC BUNCH CROSSING (40 MHz) → 5 TB/s 30 MHz non-empty pp → FULL DETECTOR READOUT 5 TB/s → PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1) 0.5-1.5 MHz 70-200 GB/s → BUFFER → FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2) 10 GB/s → 6% CALIB EVENTS 1.6 GB/s → 26% FULL EVENTS 5.9 GB/s → 68% TURBO EVENTS 2.5 GB/s → OFFLINE PROCESSING → ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
REAL-TIME ALIGNEMENT & CALIBRATION

- Use of calibration samples selected by HLT1 stored in buffer
- Real-time calculation of alignment and calibration constants
  - Constants that are used for the reconstruction and selections
  - Ensures measurements of physics parameters to the best resolution possible
- Used to reach offline-quality reconstruction at the HLT2 level

- Alignment for Tracking system and for the RICH mirrors
- Calibration for RICH and CALO
REAL-TIME ALIGNMENT AND CALIBRATION

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz
70-200 GB/s

BUFFER

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

OFFLINE PROCESSING

- 6% CALIB EVENTS
- 26% FULL EVENTS
- 68% TURBO EVENTS

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb
Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
HIGH LEVEL TRIGGER 2

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT
5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

REAL-TIME ALIGNMENT & CALIBRATION

0.5-1.5 MHz
70-200 GB/s

BUFFER

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

OFFLINE PROCESSING

6% CALIB EVENTS
1.6 GB/s

26% FULL EVENTS
5.9 GB/s

68% TURBO EVENTS
2.5 GB/s

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
→ Full reconstruction → offline-quality reconstruction
  ➤ Aligned and calibrated detector (see previous step)
  ➤ Full particle identification with RICH reconstruction
  ➤ Full track fit, with detailed magnetic field and detector description

→ Around ~1000 selections algorithms
  ➤ Run2 Stripping (offline) moved to the HLT2 level for Run3 (online)
→ The selections are tuned to different signal topology and physics analysis
→ Runs on CPUs
→ Implemented on Moore

*More on Moore for Run3 with Jonathan*
All numbers related to the dataflow are taken from the LHCb Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
MOVING TO OFFLINE

LHC BUNCH CROSSING (40 MHz)

5 TB/s
30 MHz non-empty pp

FULL DETECTOR READOUT

5 TB/s

PARTIAL DETECTOR RECONSTRUCTION & SELECTIONS (GPU HLT1)

0.5-1.5 MHz
70-200 GB/s

REAL-TIME ALIGNMENT & CALIBRATION

BUFFERS

FULL DETECTOR RECONSTRUCTION & SELECTIONS (CPU HLT2)

10 GB/s

OFFLINE PROCESSING

6% CALIB EVENTS
1.6 GB/s

26% FULL EVENTS
5.9 GB/s

68% TURBO EVENTS
2.5 GB/s

ANALYSIS PRODUCTIONS & USER ANALYSIS

All numbers related to the dataflow are taken from the LHCb

Upgrade Trigger and Online TDR
Upgrade Computing Model TDR
SPRUCING

→ Used for physics selections that cannot go straight to disk
→ Intermediate step between the tape storage and the disk storage to reduce the data
→ Also based on the Moore project
→ No offline reconstruction - the reconstruction comes from HLT2

→ Typical case that need sprucing are inclusive HLT2 lines or to data selection/processing algorithms that are too intensive to be run online.
→ In Run 2:
  ➤ 70% of the events passed to offline processing
  ➤ 30% to Turbo
→ In Run 3:
  ➤ 68% passed to Turbo (Baseline in Run3)
  ➤ 32% to offline processing
→ Turbo → bypass the offline processing steps and stripping (save in storage and computing power)
→ Saves only the signal candidate
NTUPLE MAKING

→ Software: DaVinci

→ Used in the previous runs but majors changes
  ➤ From LoKi to ThOr functors
  ➤ From TupleTools to FunTuples

→ FunTuples:
  ➤ More flexibility on the choice of variables
  ➤ Reduces the number of unused variables
  ➤ Reduces storage and computing use

More on DaVinci for Run3 with Jonathan

ANALYSIS PRODUCTION

→ Move to central analysis production from the user specific jobs

→ To central production managed by DIRAC

→ Analysis Productions:
  ➤ Automatic testing
  ➤ Automatic preservation of config details
  ➤ Automatic error interpretation
  ➤ Web interface
THANK YOU FOR LISTENING!