The evolution of the Trigger and Data Acquisition System in the ATLAS experiment

A Krasznahorkay[1,2] on behalf of the ATLAS Collaboration

New York University, NY, USA

E-mail: `Attila.Krasznahorkay@cern.ch`

Abstract. The ATLAS experiment, aimed at recording the results of LHC proton-proton collisions, is upgrading its Trigger and Data Acquisition (TDAQ) system during the current LHC first long shutdown. The purpose of the upgrade is to add robustness and flexibility to the selection and the conveyance of the physics data, simplify the maintenance of the infrastructure, exploit new technologies and, overall, make ATLAS data-taking capable of dealing with increasing event rates. The TDAQ system used to date is organised in a three-level selection scheme, including a hardware-based first-level trigger and second- and third-level triggers implemented as separate software systems distributed on separate, commodity hardware nodes. While this architecture was successfully operated well beyond the original design goals, the accumulated experience stimulated interest to explore possible evolutions. We will also be upgrading the hardware of the TDAQ system by introducing new elements to it. For the high-level trigger, the current plan is to deploy a single homogeneous system, which merges the execution of the second and third trigger levels, still separated, on a unique hardware node. Prototyping efforts already demonstrated many benefits to the simplified design. In this paper we report on the design and the development status of this new system.

1. Introduction

The ATLAS experiment [1] is one of the major experiments of the Large Hadron Collider (LHC) [2]. It is a general purpose experiment, aiming at studying the Standard Model Higgs Boson, and looking for physics beyond the Standard Model of particle physics. It consists of various semi-conductor based charged particle tracker detectors (the Inner Detector), a liquid-argon based electromagnetic calorimeter, and a hadronic calorimeter made of steel and scintillator tiles. The outside of the detector is covered by a muon spectrometer for identifying muons leaving the detector.

The detector provides many millions of read-out channels, able to capture data every 25 nanosecods. This can not all be kept for further data analysis of course. The Trigger & Data Acquisition (TDAQ) system is needed to collect information from all parts of the detector, and select on average 400 events per second[3] to be keep for further processing.

The TDAQ system performed very well during LHC's Run 1 (2009-2012), in many aspects well beyond its design values. It allowed ATLAS to collect events relevant for physics analyses with a high efficiency and sustainable event recording rates. However it is clear that to do

[1] Currently at: European Laboratory for Particle Physics (CERN), Geneva, Switzerland

[2] On leave from: Institute of Nuclear Physics of the Hugnarian Academy of Sciences (ATOMKI), Hungary

[3] Average number for the 2012 data taking period.

the same for Run 2 (2015-2018), which will provide the LHC experiments with much more demanding conditions, the TDAQ system has to be improved to keep its high efficiency and low event-recording rate.

This document summarizes some of the hardware and software developments that are happening during LHC's Long Shutdown 1 (LS1, 2013-2014), and how they will impact ATLAS's data taking in LHC's Run 2.

1.1. The ATLAS TDAQ System in Run 1

ATLAS uses a 3-level trigger system, organised as follows:

- **Level-1** (LVL1) composed of specially designed, dedicated hardware components making decisions about the events based on simple reconstruction methods of information just from the calorimeters and the muon spectrometer.
- **Level-2** (LVL2) composed mainly of dedicated trigger algorithms running on commodity PCs, reconstructing physical objects primarily in small regions of the dector (Region of Interest, RoI), seeded by LVL1.
- **EventFilter** (EF) running mainly offline reconstruction code, having possible access to the full event information.

Level-2 and the EventFilter are referred to collectively as the High Level Trigger (HLT).

Figure 1 shows the logical structure of the TDAQ system, citing the original design parameters of the system, and the operating parameters that were in effect by the end of LHC's Run 1.

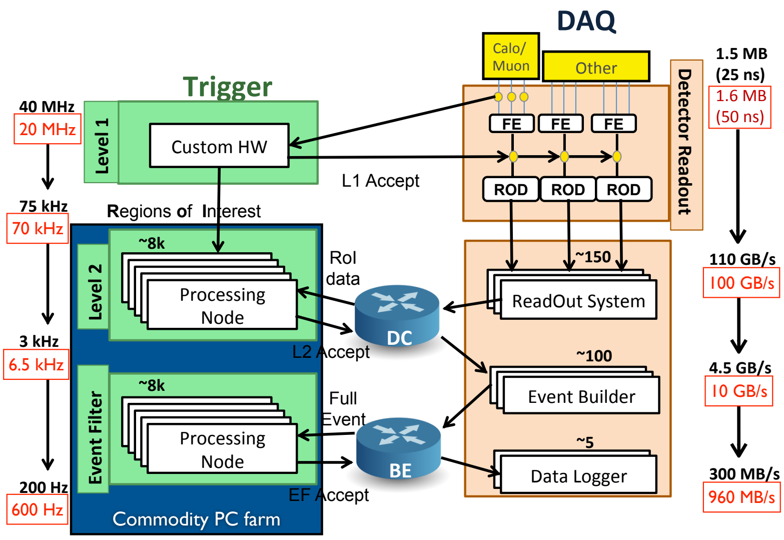

Figure 1. Schematic view of the ATLAS TDAQ system as used in Run 1. The (unframed) labels in black show the design parameters of the system, the (framed) labels in red show the parameters achieved at the end of Run 1.

The data acquisition network was partitioned into multiple segments, as shown in Figure 2. Two separate Data Collection (DC) networks were transporting Read Out System (ROS) data fragments to the LVL2 trigger processes running on the computing nodes (XPU). The two networks were coupled by the Event Building (SFI) nodes that collected all the fragments of the events accepted by LVL2, and sent the full events onward to the EF nodes. The events accepted by the EF, and were written to permanent storage via the Sub-Farm Output (SFO) nodes.

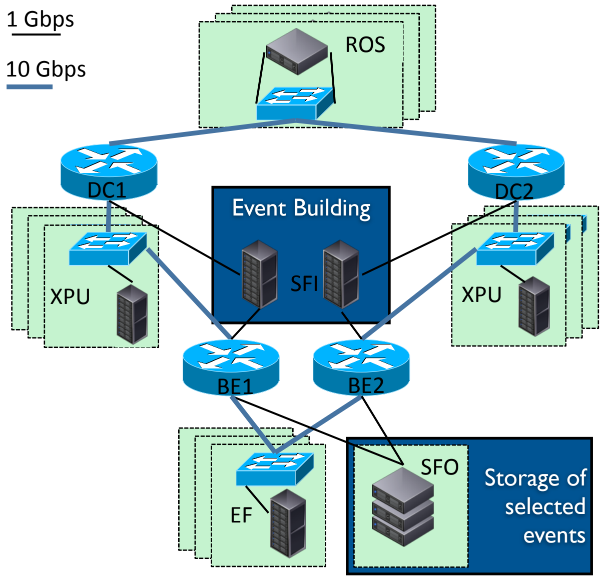

Figure 2. Schematic view of the ATLAS DAQ network as used in Run 1.

1.2. Changes for LHC's Run 2

LHC's Long Shutdown 1 is used to increase both the energy and the intensity of the proton-proton collisions in the machine. Table 1 summarizes the most important changes in the parameters of the machine.

Table 1. Characteristic properties of LHC's Run 1, and the expected properties of the machine for Run 2. The pile-up is expressed in average number of interactions per bunch crossing.

Period	E_{cm} [TeV]	# of bunches	Bunch separation [ns]	Peak lumi. [$cm^{-2}s^{-1}$]	Pile-up [#/BC]	Event size [MB]
Run 1	8	1380	50	$\sim 7 \times 10^{33}$	~ 35	~ 1.5
Run 2	~ 13	1380 - 2700	25 - 50	$1.5 - 2 \times 10^{34}$	up to 80	1.7 - 2.4

According to current expectations the machine will provide proton-proton collisions at 13-14 TeV, with several times the intensity of that in Run 1, with up to an expected 80 proton-proton interactions per bunch crossing. To achieve our physics goals, which include continuing the searches for new physics in electroweak final states, and performing precision measurements on the newly discovered Higgs boson, we will need to:

- Speed up the HLT code, to compensate for increased processing time due to high pileup and allow more use of slower offline reconstruction algorithms.
- Make the trigger and offline analysis selections more similar to each other, increasing the trigger selection efficiency, and simplifying analyses a little bit.
- Introduce more specialised, more selective triggers, to help reduce trigger rates.

The code optimisation is critical, as observations indicate a steeper than linear CPU time response as a function of the average number of pile-up vertices in the events. This means that with the current code we would not be able to fit into the resource limits set by the HLT farm.

Physics analyses will also need to use more specialised triggers than in Run 1, when it was still possible to afford selecting events with relatively simple requirements. As an example, in Run 2 it is expected that even the LVL1 trigger looking for 20 GeV muons will give an acceptance rate $\gg$10 kHz. Since the 100 kHz LVL1 readout rate is a hard limitation in the system, the trigger configuration will need a strategy combining higher thresholds, isolations and single muons in coincidence with some other condition(s) for selecting events for LVL2.

2. Hardware Upgrades

This section summarizes the hardware upgrades being made to ATLAS during LS1 that have an impact on the TDAQ system.

2.1. The Level-1 Topological Trigger

The Level-1 trigger system in Run 1 was only able to look for single objects (like jets and muons), or require simple combinations of such single objects to be present in the event. To extend the functionality of the LVL1 trigger, a new hardware component will be added to the system, the LVL1 Topological Trigger [3].

It will receive detailed information (such as energy and direction) about the candidate objects found by the LVL1 calorimeter and muon triggers. This information will then be processed by dedicated algorithms implemented in its FPGAs, looking for signatures such as:

- A muon close to a jet.
- An object with a certain transverse mass.
- An event with an effective mass above a given criterium.

The topological trigger will only be able to send the limited information of which signatures were found to the HLT due to the 100 kHz readout rate for this information. This is a complication for the HLT which would use the knowledge of which L1 object combination passed the topological criteria to seed the RoI-guided HLT reconstruction. The solution is to simulate the L1 topological trigger hardware: the FPGA code will also be implemented in C++, validated against the hardware response, and run online in the HLT.

2.2. Fast TracKing (FTK)

Another dedicated hardware component is now being developed to help with track finding and reconstruction in the HLT. Called the Fast Tracker ([4]), it will receive all the semi-conductor data from the inner detector charged particle tracker at each LVL1 accept signal, with up to 100 kHz rate. The hardware then finds and reconstructs charged track candidates using pattern matching on specially pre-processed data in an associative memory. This way the hardware provides tracking information for the whole detector very quickly, without the use of HLT processing time or readout bandwidth.

The information can then be used in the HLT in a number of ways:

- Calculating the isolation of identified leptons from other charged particles.
- Finding the primary vertices of the event using all the reconstructed tracks.
- Seeding the HLT track reconstruction to reconstruct charged tracks with high precision in a large area of the detector.
- The reconstructed tracks may be used directly in the b-tagging and tau reconstruction algorithms to improve their performance with the additional tracking information.

There is currently both a very active code development program underway to make use of the FTK tracks in the HLT reconstruction, and a large set of trigger performance studies going on to evaluate the performance of using FTK tracks as part of the object/event selection.

According to the current schedule, the FTK system will cover the barrel system by July 2015, and the entire detector by the end of 2015.

2.3. Simplified HLT & DAQ Achitecture

While the system architecture described in Section 1.1 served ATLAS well in Run 1, technology advancements allow it to be simplified for Run 2.

With the rolling replacement of core network routers during LS1, the simpler design shown in Figure 3 has been possible so that it still provides greater bandwidth to meet the needs of increased event data sizes and access patterns expected with the high pileup conditions of Run 2.

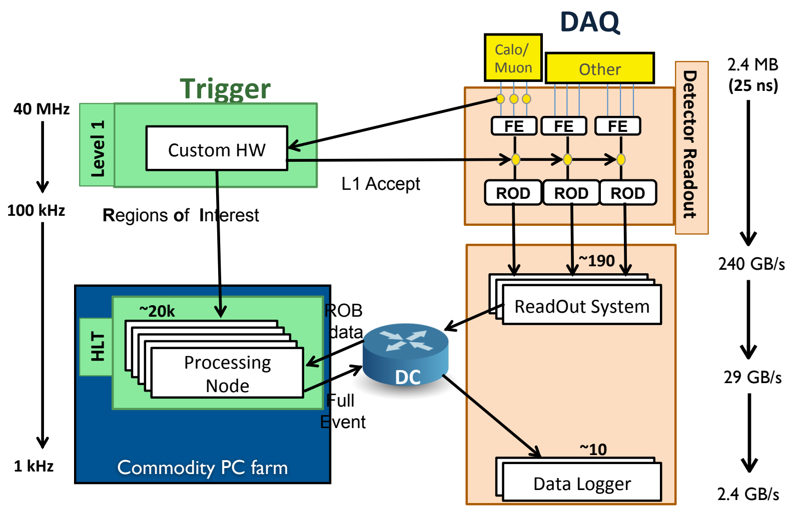

Figure 3. Schematic view of the ATLAS TDAQ system planned to be used in Run 2.

This has also made is possible to merge the L2, EF and event building functionality into a single HLT node. This is advantageous for balancing HLT resources and allows for a much simpler software design too, described in Section 3.1.

Figure 4 shows how the system will use a single network – with multiple switches – for transporting the event fragments to the HLT nodes, as the HLT nodes will now take care of both the LVL2 and EF decision making and the event building as well. The accepted, fully built events will be sent on the same network to the same type of SFO nodes that were used in the previous system as well.

This network design is easily scalable, in case the HLT farm needs to be expanded during Run 2.

3. Software Upgrades

ATLAS set out to simplify, and at the same time improve its trigger software significantly during LS1. There are more development projects than could be listed here (for instance updating the trigger Event Data Model, improving the trigger analysis tools, speeding up and improving the efficiencies of algorithms, etc.). For this reason this section just focuses on the two most important developments happening in the core software of the ATLAS trigger.

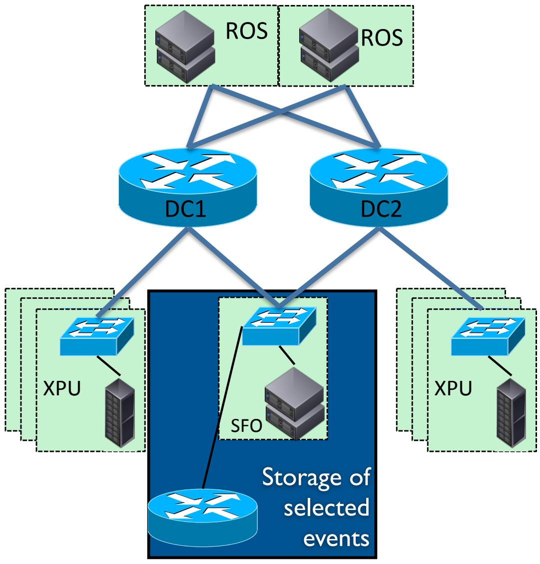

Figure 4. Schematic view of the ATLAS DAQ network as it will be used at the beginning of Run 2.

3.1. The Merged High Level Trigger

Having two High Level Trigger levels (LVL2 and EF) proved to give us very good flexibility during Run 1 to construct trigger configurations that could be used even at the highest LHC intensities. But it became evident by the end of the running period, that the system should be simplified somewhat in order to ease day-to-day operations of the data acquisition.

Some of the imperfections of the old system were:

- *The rigidity of the network configuration.* Computing resources needed to be assigned to either LVL2 or the EF, and re-configuring the network was a non-trivial task. This is addressed by the changes decribed in Section 2.3.
- *Calculation of the same quantities at LVL2 and the EF.* Because of the limitation of how much data could be sent from LVL2 to the EF, some reconstruction that were performed at LVL2, needed to be repeated in the EF.
- *Duplications in network transfers.* The information requested by LVL2 algorithms needed to be requested once more by the event building node in order to build the full event, increasing the request rate hitting the Read-Out System nodes by $\sim$10%.

The simplified software design, made possible by the merged HLT, is shown in Figure 5.

The new system will still separate algorithms running either at the LVL2 or the EF stage. But the algorithms themselves will be run on a single computing node, inside a single process. At the same time, the event building will also run on the same node. The separate LVL2 Processing Unit (L2PU) and EF Processing Task (PT) applications will be merged into a single HLT Processing Unit (HLTPU) application. The tasks of the Data Flow Manager (DFM), the Sub-Farm Input (SFI) and Event Filter Dataflow (EFD) components will be taken over by the Data Collection Manager (DCM).

The defining difference between the algorithms at that point will be that LVL2 algorithms will be executed before the full event would be built, while EF algorithms would have the ability to access the full event information, just like in the previous implementation.

The main improvements expected from the new system are the following:

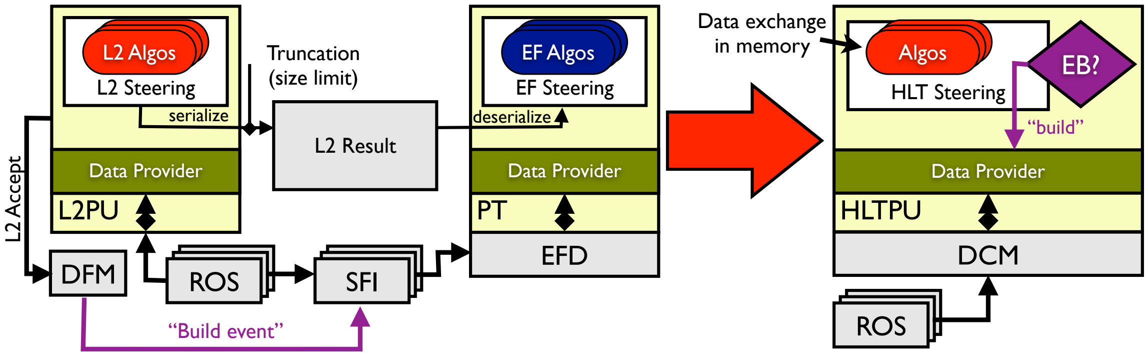

Figure 5. Schematic view of the structure of the HLT code as used in Run 1 shown on the left, and the schema under development for Run 2 is shown on the right. The components are explained in the referring text.

- As all the algorithms will be running inside the same process, they will be free to exchange as much information between each other as needed.
- The data collection system on the processing units will be capable of caching event fragments, so that once event building is requested for a given event, the HLT node will only need to collect the data not requested by one of the LVL2 algorithms already.
- The event building point will be much more flexible in the new system. It should be possible to find the optimal event building point inside any trigger configuration in order to optimise the network traffic, queue/buffer lengths and CPU time.

The ability to exchange much more information between algorithms than before allows us to explore new reconstruction strategies. One example is how the Inner Detector track reconstruction code is being re-designed. In the new system it will no longer be necessary to repeat some parts of the track finding at both levels. The system is being re-designed ([5]) to combine fast LVL2-like data unpacking and pattern recognition with subsequent refinements based on offline reconstruction tools.

3.2. Multi-Processing in the HLT

The clear trend in computing is to move toward computers with more and more CPU cores. Up to now we could take advantage of the multi-core CPUs by running multiple independent HLT processes on them. Unfortunately in the many-core era this strategy will not scale due to memory constraints.

HLT processes need about 1.5 GB of memory to run in our current system. This means that on an N-core system we need to start N HLT processes, and we need to provide N$\times$1.5 GB memory to them. With current servers being available with 48 CPU cores, and future servers on the horizon with O(100) cores, this setup will soon not be able to provide enough memory for the processes.

The problem is tackled by taking advantage of the fact that much of the data held in memory by the individual processes is the same, for example the magnetic field description and detector conditions data. The unique data needed by the processes is 300 MB/process, which is much more realistic to have available even on a many-core system.

Using the Linux kernel's copy-on-write feature, it is possible to start a single process that is forked into multiple processes after initialisation, allowing the running of many HLT processes in parallel, with much reduced memory requirements compared to starting them individually. This setup has been demonstrated successfully in recent tests, making it ATLAS's baseline model for running HLT processes in Run 2.

4. Summary

The ATLAS TDAQ system worked exceptionally well during LHC's Run 1. It allowed us to efficiently select the events needed for all the physics analyses and the discovery of the Higgs boson. With the almost 2 year first long shutdown of the LHC underway to upgrade the energy and intensity capabilities of the accelerator, we are in the process of making improvements to the system so that it will be able to serve us just as efficiently during LHC's Run 2.

There are many developments happening to simplify and improve both the hardware and the software design of the TDAQ system. These will allow us to have even more flexibility in developing and configuring how we select proton collisions that are interesting to physics analyses.

All the developments are well underway, with full scale tests of the system expected in 2013 and 2014.

References

[1] ATLAS Collaboration 2008 *Journal of Instrumentation* **3** S08003

[2] Evans L and Bryant P 2008 *Journal of Instrumentation* **3** S08001

[3] Simioni E *et al.* 2012 Topological and Central Trigger Processor for 2014 LHC luminosities Tech. Rep. ATL-DAQ-PROC-2012-041 https://cds.cern.ch/record/1459209 CERN Geneva

[4] Shochet M *et al.* 2013 Fast TracKer (FTK) Technical Design Report Tech. Rep. CERN-LHCC-2013-007 ATLAS-TDR-021 https://cds.cern.ch/record/1552953 CERN Geneva

[5] ATLAS Collaboration 2013 Studies for the development of the Inner Detector trigger algorithms at ATLAS Tech. Rep. ATL-DAQ-PUB-2013-002 https://cds.cern.ch/record/1602918 CERN Geneva