Readiness of the ATLAS Trigger and Data Acquisition system for the first LHC beams

W. Vandelli[a]*, on behalf of ATLAS TDAQ Collaboration†

[a]CERN, Geneva, Switzerland

The ATLAS Trigger and Data Acquisition (TDAQ) system is based on O(2k) processing nodes, interconnected by a multi-layer Gigabit network, and consists of a combination of custom electronics and commercial products. In its final configuration, O(20k) applications will provide the needed capabilities in terms of event selection, data flow, local storage and data monitoring. In preparation for the first LHC beams, many TDAQ sub-systems already reached the final configuration and roughly one third of the final processing power has been deployed. Therefore, the current system allows for a sensible evaluation of the performance and scaling properties. In this paper we introduce the ATLAS TDAQ system requirements and architecture and we discuss the status of software and hardware component. We moreover present the results of performance measurements validating the system design and providing a figure for the ATLAS data acquisition capabilities in the initial data taking period.

1. ATLAS Trigger and Data Acquisition

ATLAS [2] is one of the experiments installed along the Large Hadron Collider at CERN. The ATLAS Trigger and Data Acquisition (TDAQ) [3] system is responsible for the selection and the conveyance of interesting physics data, reducing the initial bunch-crossing frequency of 40 MHz to a rate of stored events of ∼200 Hz. The TDAQ design evolved into a three-level selection scheme, with a “Region-of-Interest” (RoI) based second-level trigger. The final system will include up to 2000 nodes interconnected by a multi-stage Gigabit Ethernet network [4].

The ATLAS TDAQ system is based on in-house designed multi-threaded software, mostly written using the C++ and Java programming languages and running on a Linux operating system.

2. TDAQ Architecture

The system is mostly built around three independent Gigabit Ethernet networks which connect all the involved nodes and provide the means of conveyance of event data as well as control and monitoring information.

The initial event selection is performed in custom electronic modules (LVL1), which should provide an output rate of 100 kHz. Upon acceptance, the detector front-end boards send the data to the Read-Out Buffers (ROBs), via ∼1600 optical links. The ROBs are custom PCI boards, installed in dedicated PCs, which form the Read-Out System (ROS). Overall the ROS system accounts for roughly 550 ROBs and 150 PCs.

On the trigger path, the Region-of-Interest Builder (RoIB) assembles the RoI information provided by the LVL1. The RoI defines, event-by-event, the regions of the detector in which interesting physics features reside. The event is then assigned to a processing node of the second-level trigger (LVL2) farms. The LVL2 collects data fragments from the ROS system, via the data-collection network, and refines the LVL1 result with a local analysis inside the RoI regions. This mechanism ensures a fast event rejection with minimal data-flow demands, therefore reducing the networking and ROS computing power requirements.

After the LVL2 selection, at a rate of ∼3.5 kHz, the Event Builder (EB) is responsible for the collection of all the data fragments from the ROS

*This research project has been supported by a Marie Curie Early Stage Research Training Fellowship of the European Community’s Sixth Framework Programme under contract number (MRTN-CT-2006-035606)

†For the full author list see [1].

PCs and the assembling of complete events. The events are then made available to the Event Filter (EF) processors via the event-filter network, for the last selection stage. Given the foreseen event size of ∼1.5 MB, the Event Builder will have to sustain an aggregated I/O network bandwidth of $\mathcal{O}$(5GB/s). The building process is performed by a farm accounting roughly 100 applications.

The last filtering process is carried out by the Event Filter, the second component, together with the LVL2, of the High-Level Trigger (HLT) infrastructure. In the final TDAQ system the HLT farm is expected to include 2300 processing nodes. Event definitively accepted by the EF are transferred at 200 Hz (300 MB/s) to the Sub-Farm Output (SFO). This is a small farm of 5 nodes where the data are temporary buffered on local disks while waiting for the transmission to the mass storage.

In preparation for the 2008 data-taking period most of ATLAS TDAQ sub-systems have been fully deployed. In particular the networking, ROS, EB and SFO systems were in their final configuration. Concerning the HLT farm instead, 850 quad-core dual-CPU nodes have been installed and commissioned. Thanks to a double network connection, these nodes can be configured either as LVL2 or EF processors, increasing the flexibility of the system.

2.1. Controls and Configuration

Dedicated hardware and software are needed in order to manage, control and configure the large number of processing nodes and applications which form the ATLAS TDAQ system. In the final system, more than 100 computing units will be in charge of providing file server functionalities as well as monitoring and control capabilities. All these tasks are performed over the control network infrastructure and they are therefore completely decoupled from the data-flow.

The control, configuration and monitoring software is based on a implementation of the CORBA [5] inter-process communication standard. Notable features provided by control framework are: remote handling and configuration of distributed applications, software and hardware resource granting and an expert system for automatic recovery procedures.

During the first-beam data-taking period the typical data acquisition session included 7000 applications, distributed over 1500 nodes. The configuration of the system was stored in a database accounting roughly 100000 objects.

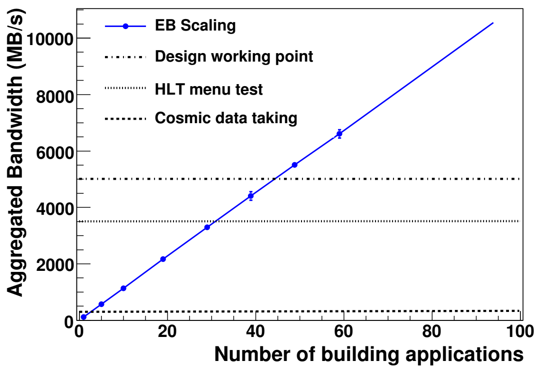

Figure 1. Event Builder aggregated bandwidth scaling as a function of the number of employed building applications. Measurements have been performed up to 59 applications. The continuous line provides the extrapolation up to the final building farm. The dashed lines show different working points of the TDAQ system.

3. First-beam data taking

The ATLAS TDAQ system was successfully used during the detector commissioning and in particular for the continuous cosmic and first-beam data-taking period, from August to November 2008. Overall roughly 1 PB of data have been collected over this period, confirming the functionality of the data acquisition system.

The usage during the data-taking period provided an invaluable feedback in terms of functionalities, stability and efficiency, however, as shown in Figure 1, the typical TDAQ working point was quite far from its design specification. Therefore dedicated performance measurements have been performed in order to back-up the data-taking results.

Important outcomes have been obtained while testing the $10^{31}\text{cm}^{-2}\text{s}^{-1}$ trigger menu. The full TDAQ chain was running with simulated data loaded into the ROS system, while the corresponding LVL1 results were seeding the LVL2 via a dedicated feeder. During this test we achieved a stable running condition at LVL1 rate of 60 kHz (Figure 2), with a trigger menu optimized for 10 kHz only. The Event Builder was driven by the LVL2 at an aggregated bandwidth of ~3 GB/s (Figure 1). Even more interesting are the results concerning the ROS system. In the final running conditions large differences are expected in the request rate across the ROS PCs, with the so-called "hot ROSs" experiencing up to 12 kHz of data requests. As shown in Figure 3, the present ROSs are able to sustain request rates up to roughly 30 kHz.

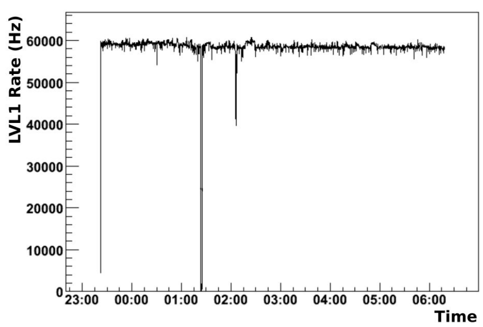

Figure 2. LVL1 trigger rate as a function of the time during the trigger-menu testing. The dips at 01:30 and 2 o'clock are due to system-wide automatic synchronization jobs.

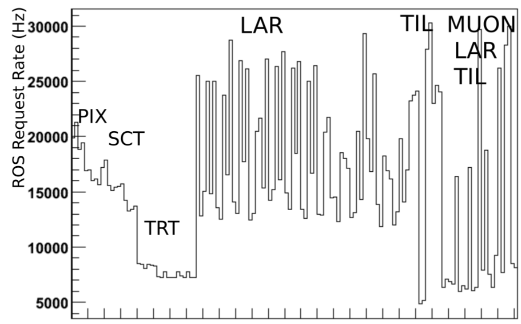

Figure 3. Distribution of the ROS request rate per sub-detector. Simulated data were loaded into the ROSs, which were serving event fragments to the LVL2 and the Event Builder.

4. Conclusions

Most of the ATLAS TDAQ sub-systems completed the deployment and commissioning phase and one third of the final HLT computing power is installed and functional. Future deployments in this area will follow the ATLAS needs. The TDAQ system is regularly and successfully used for the ATLAS commissioning and data-taking operations. Moreover, dedicated TDAQ performance tests are periodically carried-out in order to integrate the data-taking results. Based on the current achievements, the ATLAS TDAQ seems robust and scalable enough to fulfill the ATLAS requirements.

REFERENCES

1. ATLAS TDAQ Collaboration, "The ATLAS Trigger/DAQ Authorlist, version 1.0", ATL-DAQ-PUB-2008-004, CERN, Geneva, 2008, `http://cdsweb.cern.ch/record/1141516`.
2. ATLAS Collaboration, G. Aad et al, "The ATLAS experiment at the CERN Large Hadron Collider", JINST 3 S08003, 2008.
3. ATLAS Collaboration, "ATLAS, High-Level Trigger, Data Acquisition and Controls", CERN/LHCC/2003-022, Geneva, CERN, 2003
4. S. Stancu et al, "Networks for ATLAS Trigger and Data Acquisition", in *Proc.* 15^{th} CHEP, Mumbai, India, 2006.
5. M. Henning and S. Vinoski, "Advanced CORBA programming with C++", in ISBN 0-201-37927-9, Addison-Wesley, 1999.