Data Quality Monitoring Framework for the ATLAS experiment: Performance achieved with colliding beams at the LHC

The ALTAS Collaboration

A Corso-Radu[1], H Hadavand[2], Y Illchenko[2], S Kolos[1], H Okawa[1], K Slagle[1], A Taffard[1]

[1] University of California Irvine, Irvine, CA 92697, USA
[2] South Methodist University, Dallas TX 75275, USA

E-mail: ataffard@cern.ch

Abstract. This first year of data taking has been of great interest, not only for the physics outcome, but also for operating the system under the environment it was designed for. The online data quality monitoring framework (DQMF) is a highly scalable distributed framework which is used to assess the operational conditions of the detector and the quality of the data. DQMF provides quick feedback to the user about the functioning and performance of the sub-detectors by performing over 75,000 advanced data quality checks, with rates varying depending on histogram update frequency. The DQM display (DQMD) is the visualisation tool with which histograms and their data quality assessments can be accessed. It allows for great flexibility for displaying histograms, their reference when applicable, configurations used for the automatics checks, data quality flags and much more. The DQM configuration is stored in a database that can be easily created and edited with the DQM Configurator tool (DQMC). This paper is describing the design and implementation of the DQMF and its display as well as the data quality performance achieved during this first year of data taking.

1. Introduction

The ATLAS experiment is one of the general-purpose particle physics detectors designed and built to record the proton-proton collisions provided by the Large Hadron Collider (LHC) at CERN, Geneva. The LHC will produce proton head-on collisions with center-of-mass energy of 14 TeV at a rate of 40MHz, rate at which ATLAS has to be able to analyze and filter the information provided by its approximately one hundred and forty million channels.

Given that the rate for interesting physics events is much lower than the collision rate and that the large number of channels makes the event size of about 1.5MB a powerful trigger system is needed to select the events that will be recorded. The data acquisition system (DAQ) is based on a three-level trigger architecture [1] to achieve a final event rate of 200Hz, from the 40MHz collision rate. This year with a collision rate of ~1MHz, the typical output rate used was ~350Hz. At each consecutive level of the trigger chain, more information is available. The data flow system (DF) is responsible for collecting data fragments, serving them to trigger processors and sending them to mass storage.

Data quality monitoring is part of the Monitoring Infrastructure of the ATLAS experiment. Monitoring data quality is an important and integral part of the data taking process in High Energy Physics experiments. It is performed throughout the data acquisition and then the offline processing of fully reconstructed data. Assessment of the quality of incoming data is made during continuous monitoring and is archived for retrieval at the physics analysis stage. Due to the complexity of the ATLAS experiment, a framework for automatic data quality assessments of incoming data and a visualization tool for easy identification of problems are essential. A highly scalable distributed data quality monitoring framework (DQMF) has been developed and is being used to monitor the quality of the data as well as operational conditions of hardware and software elements of the detector, trigger, and data acquisition systems. Online, the framework permits to avoid recording faulty data by automatically checking via predefined algorithms thousands of histograms from all sub-systems at all stages of data flow and alerting the shift crew as problems occur. The shift crew interacts with DQMF via the data quality monitoring display (DQMD), which serves as a platform to alert of problems and debug them. Offline, the DQMF is used to perform the data quality monitoring on a smaller "express" subset of the data reconstructed within 24hrs of the data being recorded to verify calibrations and alignments and then after the bulk reprocessing at Tier-0 from which the final data quality assessment is done. Figure 1 shows a schematic of the organization of ATLAS monitoring infrastructure used to obtain the final data quality assessments. This paper describes the implementation of the DQMF and DQMD, focusing on the online environment and the overall data quality performance achieved during this first year of data taking.
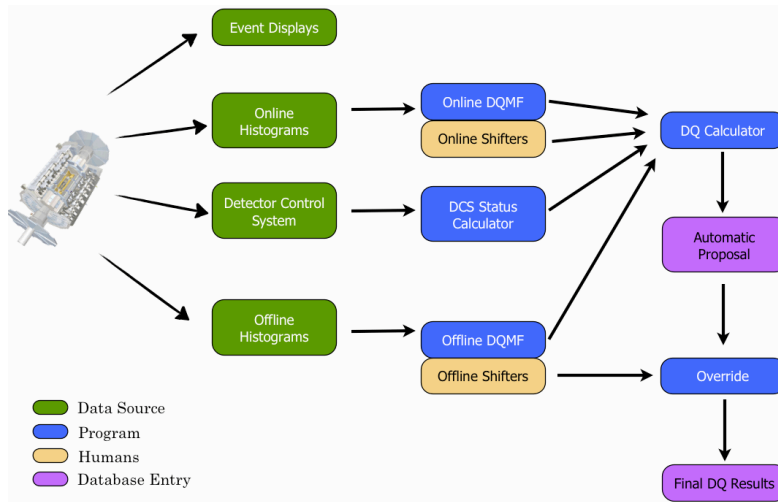

Figure 1: Schematic of the organization of ATLAS monitoring infrastructure used to obtain to final data quality assessments for retrieval at the analysis stage.

2. The data quality monitoring framework

A subset of the data acquisition system, constituting of 32 machines, is devoted to monitor the performance of each detector sub-systems, trigger chains and performances of the combined systems into physics objects. The information being monitored vary from operational condition, such as errors from readout electronics, distributions of ADC count or deposition of energy in the calorimeters, to distributions of triggered objects, and is published into a central Online Histogramming Service.

The Data Quality Monitoring Framework is a data driven distributed and scalable framework to monitor data quality both online and offline. Single data quality tests are defined by DQParameters. Each DQParameter specifies what input histogram(s) to use, what algorithm (e.g. empty histogram, χ^2)

and parameters (DQAlgorithm) to apply and the thresholds to classify the result (DQResult) as good or bad. All the DQParameters are grouped in different DQRegions, which in turn can be combined in more general DQRegions, thus forming a hierarchy, the DQTree. The logic to combine the DQResults of the sub-parameters and sub-regions as well as the information specific to each DQParameter, are defined in a configuration database. Within this structure each sub-detector of the ATLAS experiment is described as a top DQRegion that host several tiers of DQRegions and DQParameters to check the performance of its hardware and data being recorded. For each top DQRegion, an application runs the data quality checks and outputs the color-coded DQResults relaying the quality of the data to a central Information Service. In order to accommodate the diversity of information sources and destinations, the framework implements the input and output as plug-ins. This allows reading data from the configuration database, from some information service or from a ROOT file. Similarly, output can be sent to a conditions database, information services or a ROOT file.

Figure 2 depicts the interaction of the DQMF with the rest of the monitoring services. The incoming histograms, encapsulating data from the detector, are processed by the DQMF. The results of the checks are published to the online Information Service and archive for future reference. The information from both the Online Histogramming Service and the Information Service can then be accessed by the Data Quality Monitoring Display (DQMD) for visualization.
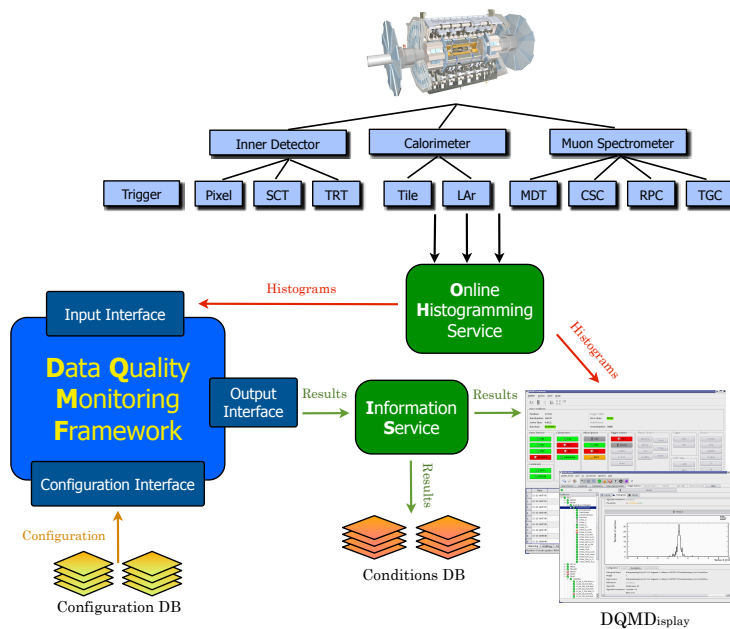

Figure 2: The Data Quality Monitoring Framework in the online infrastructure.

3. The data quality monitoring display

The Data Quality Monitoring Display is a graphical interface implemented using the QTROOT library [2][3]. The strengths of this implementation are that the performance is scalable, and the complete functionality of ROOT histograms is readily available for an interactive display. To achieve the necessary interoperability with the Data Quality Monitoring Framework and other online services, the display has been built using CORBA (Common Object Request Broker Architecture) technology provided in the software wrapper, called Inter Process Communication (IPC), in the scope of the ATLAS DAQ system [4]. CORBA provides the abstractions and services for development of portable distributed applications enabling the exchange of information, independent of hardware platforms, programming languages, and operating systems. It also provides the complete mechanism required for

distributed objects to be able to communicate with one another, whether locally or on remote hosts, without having to worry about low-level details of inter-process communication. The use of IPC greatly simplifies the implementation, hiding the details of the CORBA communication layer and its complicated API.

The display allows easy navigation between Regions and Parameters providing great flexibility for visualization of Results produced by DQMF. For any particular data acquisition configuration, all the available DQRegions and DQParameters are organized in a tree where each element is colored according to the result of the algorithm applied. For each DQRegion, the status color is inferred from the status of underlying DQParameters taken with some weight coefficients defined in configuration that determine importance of those parameters for that DQRegion. Each region being defined to represent a particular set of parameters and/or regions stands in the tree as a branch node with a summary result status for that set.
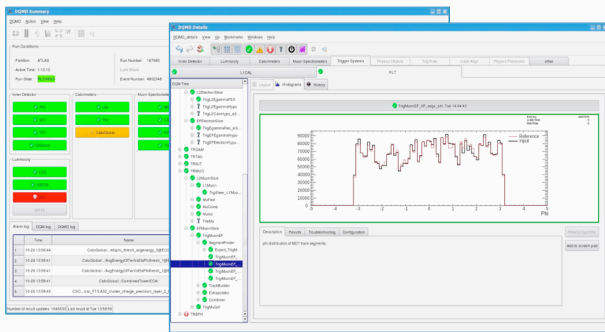

Figure 3: Overlay of the DQMD Summary Panel and Detailed Panel showing the histogram and reference for a given DQParameter.
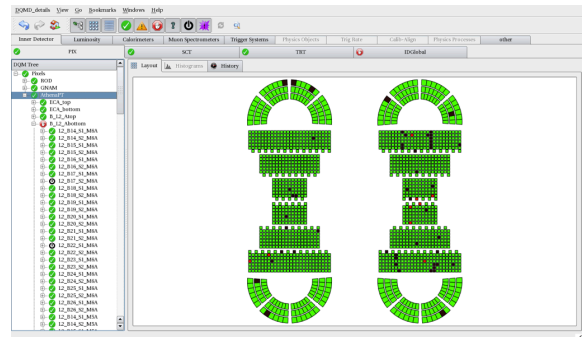
Figure 4: A layout of a sub-detector in the DQMD.

The DQMD graphic user interface consists of two windows, the summary and detailed panels, which implement its functionality (see Figure 3 and Figure 4). The summary panel is used to show the overall status of current run and detector sub-systems while the detailed panel allows the shifter to see all necessary details of data quality assessment. Results and information visible in those windows are continuously refreshed upon new results being made available.

3.1. Summary Panel
The summary panel is divided into three parts. The upper parts shows the run control conditions such as run number, running conditions etc. The central part provides the overall status per sub-system in the form of buttons painted with the colour of the corresponding result. Those buttons are grouped into sub-detector categories, which when toggled bring up the detailed panel window of the corresponding sub-detector tree and layout. Finally the bottom part contains three tabs, two for logging errors originating from either reading the configuration database or retrieval of information by the DQMD and one for alarms of results changing to worse statuses. Clicking on any of the alarm items brings up the detailed panel window highlighting the origin of the error.

3.2. Detailed Panel
The detailed panel organizes the underlying sub-system structure in a tree of regions and parameters to simplify navigation. It provides essential information related to data quality by means of a set of two tiered tabs, one tab per button of the summary panel. In each tab, a window with two sections can be found: the DQTree on the left hand side and the corresponding information of the selected element on the right hand side.

The right hand side window has three tabs: layout, histograms and history. In the layout tab a graphical representation of the system can be found. Each shape in the layout represents a sub-region

or sub-parameter and is color-coded according to its data quality status. This provided a more natural way to browse the information and facilitates "at a glance" status summary. In the histogram tab, all information regarding the result is made available: the monitored and reference histograms, current color coded status and time stamp, configuration information and algorithm parameters. Two additional tabs provide detailed description on the quantity monitored and how to interpret the information as well as what actions to take in case of problems. Finally the history tab provides graphs of the time evolution of the results values and their corresponding color-coded data quality statuses.

4. The data quality monitoring configurator

The data quality layout allows for easier understanding of the data quality status of a given sub-system and faster navigation. However, designing these layouts and translating them into a configuration language is non trivial. The data quality monitoring configurator, DQMC, was designed for this reason (see Figure 5). The DQMC allows experts to create or modify an existing configuration and design the layouts of each sub-detector systems. The configuration database is written in in platform-independent extensible markup language (XML). XML allows storing hierarchically structured data and provides ability to validate it with respect to the schema. Using the set of basic shape and arrangement scheme provided, the experts can easily create advanced graphical representation of any detector layout, which are then displayed in DQMD.
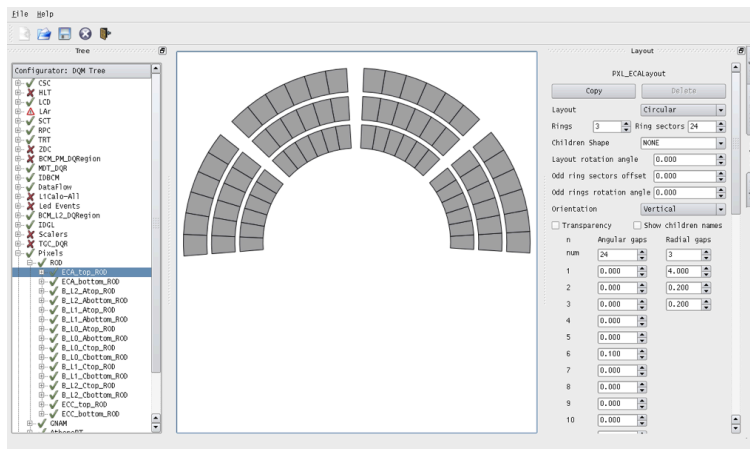

Figure 5: Data Quality Monitoring Configurator

5. Performance achieved with colliding beam

The DQMF permits to monitor more than 80 millions of channels available from all sub-detectors, where more than 97% of channels are operational for each. Overall, twenty applications runs in parallel to continuously assess the data quality of over 75,000 DQParameters regrouped in over 15,000 DQRegions, producing more than 150,000 DQResults per minute. All shifters at all desks in the ATLAS control room use the DQMD.

The data acquisition system has been implemented to handle significant changes in running condition and hence in data quality thus allowing to minimize the amount of data being lost or marked as bad for analysis. At the beginning of an LHC fill, the ATLAS detector is kept in standby, with silicon and muon detectors ramped to a low voltage setting thus protecting them from potential beam induced damage. Once "Stable beam" is declared, silicon and muon detectors voltages can be rapidly ramped to their nominal voltages thus minimizing downtime. The DAQ luminosity-weighted efficiency for the 7 TeV data up to end of September is 94.4%. If a problem occurs during data taking the DAQ permits to remove temporarily from the data taking the faulty fraction of the detector, which can then be fixed on the side and re-included into to run. The monitoring of the data quality is evaluated over luminosity block corresponding to about two minutes interval. The granularity of the data quality assessment permits to reduce the amount of data thrown away. Typical problems

encountered during data taking are related to power supply trips, error in configuration condition of the detector or coherent noise bursts in the calorimeter.

Inner Tracking Detectors			Calorimeters				Muon Detectors			
Pixel	SCT	TRT	LAr EM	LAr HAD	LAr FWD	Tile	MDT	RPC	CSC	TGC
99.2	99.5	100	90.2	96.5	97.5	99.9	99.9	99.9	99.9	99.5

Luminosity weighted relative detector uptime and good quality data delivery during 2010 stable beams at √s=7 TeV between March 30th and September 28th (in %)

Figure 6: Luminosity weighted relative fraction of good quality data delivery by the various ATLAS subsystems during LHC fills with stable beams in pp collisions at √s=7 TeV, and after switching the tracking detectors on, for runs between March 30th and September 28th.

The final data quality assessment is obtained by combining the data quality flags from the online and offline DQMF automatic assessments, the slow control information (e.g. high voltage) and the online and offline data quality shifters (see Figure 1). Preliminary data quality decisions are performed within 36 hours of the data being recorded and the final decision from the initial Tier-0 processing is available within 92 hours **Error! Reference source not found.**. Figure 6 shows the luminosity weighted relative fraction of good quality data delivered by the various ATLAS subsystems once all systems have been powered. Typically about 85% of the delivered luminosity is used for physics.

6. Summary
The Data Quality Monitoring Framework is a data driven distributed and scalable framework to monitor data quality both online and offline. Its visualization application is widely used in the ATLAS control room to ensure good quality data taking and chase down any issues that may arise. The framework is pivotal in assuring that good quality data is recorded and is an essential ingredient in providing the final data quality results to the analyzers. During this first year of data taking the framework has been proven to meet the stringent ATLAS requirements for data quality assessment.

References
[1]   The ATLAS Collaboration, ATLAS DAQ, EF, LVL2, and DCS, 1998 CERN/LHCC/98-16.
[2]   Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/.
[3]   QT-based ROOT implementation, http://root.bnl.gov/QtRoot/QtRoot.html
[4]   Applications of CORBA in the ATLAS prototype DAQ, R. Jones, S. Kolos, P. Livio Mapelli, Y. Ryabov, IEEE Trans. Nucl. Sci. 47 (2000) 331-6
[5]   "Commissioning of ATLAS data quality infrastructure with first collision data", J. Frost for the ATLAS Collaboration, in these proceedings.