ATLAS TDAQ Monitoring WG

Data Quality Monitoring Requirements

Document Version: 1.3
Document ID: ATLAS-TDAQ-2006-XXX
Document Date: 30.05.2006
Document Status: Draft

Abstract

The user requirements for Data Quality Monitoring of the ATLAS experiment are presented. They shall be the basis for the design and implementation of the Data Quality Monitoring Framework in the context of the ATLAS TDAQ/DCS system and offline, or Tier-0, system.

Institutes and Authors:

UCI, Irvine, USA: A. Corso-Radu, S. Kolos

SMU, Dallas, USA: H. Hadavand, B. Kehoe

CERN, Geneva, Switzerland: M. Hauschild

Table 1 Document Change Record

Title:	ATLAS TDAQ Monitoring WG – Data Quality Monitoring Requirements		
ID:	**ATLAS-TDAQ-2006-XXX**		
Version	**Issue**	**Date**	**Comment**
1	0	10/04/06	Initial version for comments.
1	1	18/05/06	Include comments from Monitoring WG
1	2	23/05/06	Update the Online Capabilities section, add 2 new requirements
1	3	30/05/06	Update the Tier-0 Capabilities section

1 Introduction

Data Quality (DQ) Monitoring is an important and integral part of the data taking process of HEP experiments. DQ Monitoring can occur both in the online and offline, or Tier-0 level. When used in the online environment, it provides the shift personnel with live information about the run which can be used to overcome problems early on and help avoid taking faulty data. In the event that the data taken is determined to be faulty, the run is marked accordingly as not to be used for physics analysis. At the Tier-0 level more sophisticated analysis of physics quantities are performed to ensure the quality of the run for use in physics analysis. Data Quality Monitoring involves analysis of monitoring data through user defined algorithms and relaying the summary of the analysis results while data is being processed. This information is used by shift operator or detector expert to make an assessment on the quality of the current run and is then archived for future reference.

1.1 Purpose of the document

This document presents the Requirements for Data Quality Monitoring of the ATLAS experiment. The collected requirements shall be the basis for the design and implementation of the Data Quality Monitoring framework (DQMF) in the context of the ATLAS Trigger DAQ, DCS, and Tier-0 system.

1.2 Glossary, acronyms and abbreviations

1.2.1 Glossary

See ATLAS TDAQ Glossary [1].

1.2.2 Acronyms and Abbreviations

ATLAS A Toroidal LHC Apparatus

DQM Data Quality Monitoring

DQP Data Quality Parameters

GUI Graphical User Interface

API	Application Program Interface
MDA	Monitoring Data Archive

1.3 References

1. ATLAS TDAQ Glossary: http://atlas.web.cern.ch/Atlas/GROUPS/DAQTRIG/glossary.html

2. Data Quality Monitoring and Automatic Checking – M. Hauschild – TDAQ week Oct. 2005 http://agenda.cern.ch/askArchive.php?base=agenda&categ=a051610&id=a051610s7t4/transparencies

3. Trigger and Monitoring Experiences at D0 – B. Kehoe – Cal. Trig. Software Workshop Jan. 2005, http://agenda.cern.ch/askArchive.php?base=agenda&categ=a045682&id=a045682s6t3/moreinfo

2 General Description

The purpose of the Data Quality Monitoring Framework (DQMF) is to apply specific analysis algorithms to various types of monitoring data (histograms, messages, counters, etc.) according to a particular configuration, which is defined by Detector experts. These algorithms operate on the histogram or vector output from monitoring tools at Athena, GNAM, ROD or DCS levels. The results of this analysis may generate alarms when deviations from the standard are encountered. A summary of these results will be displayed to the shifter and will also be archived for future retrieval. Using this summary, the shift operator will make a final data quality assessment for a given run. The archived results will permit a check or refinement of this assessment offline. DQMF will use already existing tools whenever it's possible, for example for displaying histograms, archiving information, etc.

2.1 Context

The Data Quality Monitoring Framework interacts with the Online Monitoring Services as well as with some other Online Services, provided as part of the ATLAS TDAQ software infrastructure (see Fig. 1), in order to be able to fulfill its objectives, in particular:

Information Service (IS)

The DQMF will use IS to retrieve the required monitoring data and also for making the DQ analysis results publicly available.

Online Histogramming Service (OH)

The DQMF will use OH to retrieve histograms produced in the current run and to transmit requests to histogram providers. Histograms generated within DQM algorithms will be transmitted to OH to enable their display.

Error Reporting/Message Reporting Service (ERS/MRS)

The DQMF will use ERS/MRS to send messages to the TDAQ system.

Controls Services (CS)

The DQMF will use CS to send messages to the TDAQ system in case a run needs to be paused, reconfigured, etc. in order to avoid taking faulty data.

Configuration Database (ConfDB)

The DQMF will use ConfDB to store and retrieve all the information, which is necessary to configure the DQM activity for the current run, i.e. algorithms to be used, monitoring data to be retrieved, etc.

Monitoring Data Archiving service (MDA)

The DQMF will use MDA to archive the DQ results.

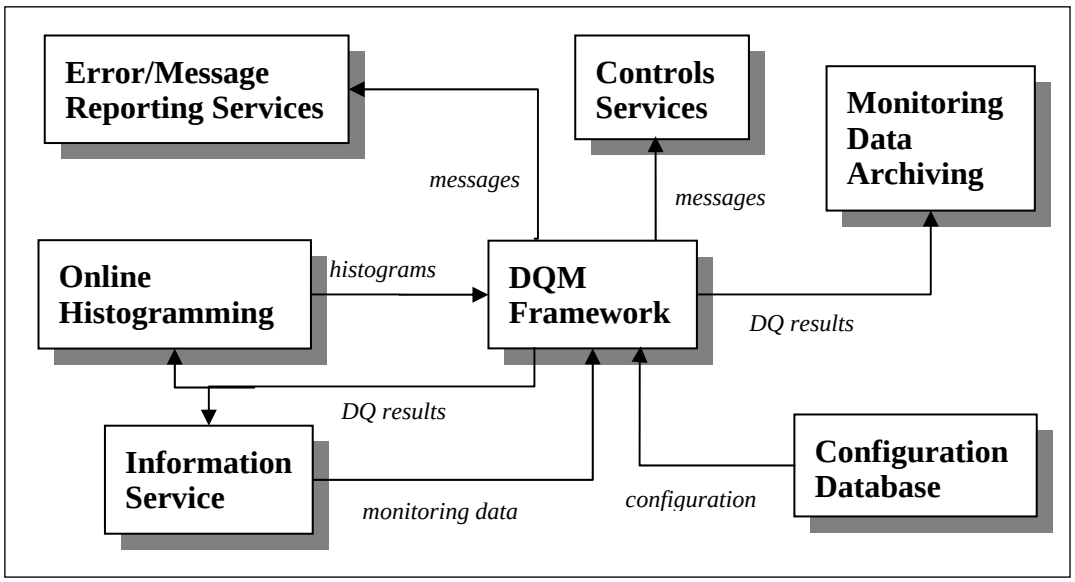

Fig. 1: DQMF Interaction with TDAQ components

2.2 General Capabilities

Data Quality Monitoring may be used in both the online and offline (Tier-0) environments. Both cases require event processing on multiple nodes. DQ monitoring will be done on the full statistics of the monitoring data and therefore it is assumed that monitoring information produced on multiple nodes has been gathered beforehand. This functionality is already provided by the basic monitoring services for the online data taking environment. It is assumed that the same infrastructure will be also provided for the offline environment.

2.2.1 Online Capabilities

In the online environment (i.e. in Point 1) the DQMF should be able to interact with the Online Services of all currently active TDAQ partitions, in the way which was described in chapter 2.1 of this document. DQMF will develop data quality decisions in the context of any single TDAQ partition and those decisions will be specific for this partition and independent of the information from other partitions.

In the context of a specific partition DQMF will perform data quality monitoring periodically (with configurable period) and store the history of the DQ results for all the periods of a specific Run. A possible minimal period for some of the DQ assessments is the Luminosity Block, which should take around 1 minute. A high granularity of the DQ decisions allows making fine separation between “good” and “bad” physics data within a given Run.

The DQMF should be able to use different reference values (e.g. reference histograms) depending on external conditions like luminosity, trigger pre-scaling factors, detector status information or run type (physics, calibration, etc.). Depending on such conditions the DQMF should be able to automatically switch from one set of references to another set, if such a different set is provided by the user. In addition, DQMF will log what set of references has been used at each check to allow reproducibility.

2.2.2 Tier-0 Capabilities

The monitoring scheme for the Tier0 environment depends on the computing model. The present model assumes latency for most data of up to 24 hours to update calibrations before the start of processing. This model does not allow fast feedback to Point-1 and thus, does not require continuous update of histograms and other monitoring data while processing. At the same time the current model also assumes existence of Express streams, which are foreseen to contain high priority events, such as Z dielectron and dimuon events, and require fast processing with high priority and quick feedback to Point-1, and thus, would need prompt reconstruction and monitoring while processing. Therefore, independent of the reconstruction model for the entire event data, online monitoring services for continuous updating and checking are required at least for the express streams. Thus for the express streams functionality for the Tier-0 monitoring is very similar to that for online monitoring in Point-1. In order to use the same tools as in the online case, Online Services are required to run on Tier-0. This would allow fully integrating Tier-0 monitoring into an ATLAS wide monitoring scheme. It is technically possible to run Online Services like the OH server on Tier-0 that would fit into the existing scheme.

For the non express stream data it may be sufficient to collect, merge and check histograms after processing has finished, which typically takes 3 hours. This purely offline approach to monitoring can be implemented using ROOT histogram files as input and does not require a presence of the Online Services. But this approach imposes some loss of functionality with respect to the online one; in particular the feedback (message or control) to the monitoring tasks from the DQMF will not be possible.

Finally, independent of the monitoring approach, a DQMF should be available in Tier-0 and provide the same general capabilities as described in section 2.2.1. A purely offline approach for monitoring may require development of a separate version of DQMF, which can be used in the absence of Online Services. As an alternative, a unified use of Online Services will make possible development of a single unique DQMF, which can be used for both express and non express streams. Such DQMF could also be used without any modifications if an alternative reconstruction model for non express streams would be used, in which case all data will be processed without latency such that event reconstruction takes place while the data is being taken.

2.2.3 Interfaces

The DQMF shall provide two types of user interfaces to its services:

Main graphical user interface, which has to allow:

- creating and modifying DQM configuration parameters;
- controlling the workflow of the DQM system;
- assigning data quality tags to the current run.

Read only Web interface, which will allow watching the DQ results for the ongoing run.

2.3 General constraints

Failure of DQM shall not disrupt data taking activities of the TDAQ system.

2.4 General assumptions and dependencies

The DQMF assumes the Online Services are up and running.

The DQMF assumes the Databases are operational and accessible.

The DQMF assumes that there is a technical possibility of providing some information (DQ results, messages, etc.) from Tier-0 to Point-1.

2.5 User characteristics

The DQMF shall be used by Detector experts for creating/modifying DQ configurations.

The DQMF shall be used by a normal shifter in day-by-day data taking activities to perform data quality monitoring.

3 Specific Constraints, Assumptions/Dependencies, Use Cases and Requirements

3.1 Constraints

CO001

The DQMF should be provided on all ALTAS T/DAQ supported platforms.

CO002

The DQMF should be compatible in terms of compiler versions and options with the algorithms packages it is using.

3.2 Use Cases

3.2.1 DQM Configuration

UC001 Identifying parameters which have to be monitored

The Detector expert identifies a set of *data quality parameters* (DQP) which have to be checked out at various phases of the DAQ session in order to verify the data quality during a run. The term *data quality parameters or* DQP is used to designate various types of data which are produced by different components of the DAQ system at the time of operation. A set of DQP may include histograms, counters, messages, alarms, physics data tags, etc.

UC002 Defining reference values and validity ranges

The Detector expert defines reference values or validity ranges for any DQP defined in the DQM configuration. In addition to that, the user may provide accepted deviation for any given DQP and an algorithm which has to be used for analysing the monitored data. An algorithm can be chosen from a set of predefined ones, provided by DQMF, or it can be a custom specific algorithm, supplied by the Detector expert as an extension of that set.

UC003 Defining the DQM reaction

The Detector expert may specify an action which has to be taken if the value of a specific DQP is out of the defined range. Possible actions may include: sending a message, generating an alarm, requesting additional histograms, asking the shifter for a decision, etc.

A message by definition doesn't cause any action and it can be called an information message whereas alarm message can cause an action, e.g. flashing light, sending email, etc.

UC004 Defining operational parameters

The Detector expert defines at which moment the DQP have to be checked. There might be a wide range of possible ways for checking a given DQP: at a specific Run Control transition (i.e. StartOfRun or EndOfRun); periodically with a configurable time interval; at the moment which is identified by occurrence of a specific event (i.e. when a certain command or message appears in the system).

UC005 Saving DQM configuration

The Detector expert saves a description of all the DQP to permanent storage and associates a unique tag with this set of DQP.

UC006 Modifying DQM configuration

The Detector expert may modify the existing set of DQP, save it to permanent storage as a new version of the existing DQP set and associate a unique tag with this version. The old version of DQP set should remain unchanged.

UC007 Linking DQM configuration with DAQ partition

The Detector expert associates a given DQP set with a given DAQ partition in which case the DQM system will use this DQP set for the next run of this partition.

3.2.2 DQM Operations

UC008 DQM system initialization

Being initialized in a context of a specific DAQ partition, the DQM system reads its own configuration using a DQP configuration tag associated with this partition. Then DQM subscribes for the monitoring data and loads analysis algorithms with respect to the DQM configuration.

UC009 DQM monitoring

The DQMF receives monitoring data from various parts of the T/DAQ system with respect to the subscriptions, which have been done during initialization. DQM analyses this data using an appropriate algorithm and may request monitoring data providers to produce some additional, more detailed histograms when it is necessary to diagnose a problem. The DQM produces summary plots based on the results of the DQ algorithm that is being run. These will be transmitted back to OH for display.

UC010 DQM reaction

The DQMF finds deviations between the monitoring data and reference values and performs the action, which was defined in the DQP configuration for that specific case. If no actions have been defined, DQM performs default action, which is also defined in the DQM configuration.

UC011 Changing reference histograms

The validity of the reference histograms defined in the DQM configuration could depend on some external conditions like: instantaneous luminosity, trigger pre-scaling factors, LHC background information or run type (physics, calibration, etc.). So, if an external condition has changed and the detector and physics experts have provided another set of reference histograms for this condition, one should switch automatically to that new set.

UC012 DQM in automatic mode

The DQMF makes data quality assessments, based on the results of the checks, which has been performed according to the DQP configuration. No human intervention is required by DQM in this mode.

UC013 DQM in semi-automatic mode

The DQMF asks shift operator intervention in case it is not able to make definite conclusion about some DQ results. The operator can browse all the information which is currently available in the DQMF system and assign a specific flag to the DQ result. The operator can also add comments which will be stored together with the DQM results to the permanent data storage.

3.2.3 DQM Output

UC014 DQM results

The DQMF enables histogram output for display of the results of the DQP monitoring to the Shift operator, writes these results to permanent data storage, and establishes a link between those results and the physics data from the DAQ run from which these results have been obtained.

UC015 DQM GUI

The Shift operator can visualize the DQ results using a GUI provided by the DQMF. This GUI can be used by the user to browse archived DQ results and to make decisions on data quality.

3.3 Functional Requirements

3.3.1 DQM Configuration

URF001 Monitoring data

The DQMF shall allow to define in its configuration what kind of data (DQP) will be used at run time for DQ analysis.

Priority High

URF002 Reference values

The DQMF shall allow defining in its configuration what reference values (e.g. histograms) will be used at run time for DQ analysis.

Priority High

URF003 Reference values sets

The DQMF shall allow defining different sets of reference values, which depend on the data taking conditions, e.g. run type, luminosity, trigger pre-scaling factors, etc.

Priority Medium

URF004 DQ algorithms

The DQMF shall allow defining DQ analysis algorithms in its configuration.

Priority High

URF005 Actions for DQ results

The DQMF shall allow defining reaction to the DQ results in its configuration

Priority High

URF006 Use database as configuration storage

The DQMF shall store its configuration in a database.

Priority High

URF007 Loading from database

The DQMF shall be able to load its configuration parameters (DQP) from a database.

Priority High

URF008 Update time

The DQMF shall be able to load the update time associated with each analysis algorithm from a database.

Priority High

URF009 Configuration versioning

The DQMF shall provide versioning for the DQM configuration database.

Priority Medium

3.3.2 DQM Operations

URF010 Obtaining monitoring data

The DQMF shall be able to retrieve monitoring data from the Online Monitoring Services.

Priority High

URF011 Algorithm execution

The DQMF shall provide a framework to execute algorithms to analyse monitored data.

Priority High

URF012 Predefined algorithms

The DQMF shall provide a list of predefined algorithms to analyse monitored data.

Priority Medium

URF013 Custom algorithms

The DQMF shall provide an API which can be used by the subsystem experts for custom algorithm development.

Priority High

URF014 Input data for DQ algorithms

The DQMF shall support custom DQ algorithms, which can be applied to different types of monitoring data.

Priority High

Note DQMF should define several interfaces for the user defined algorithms, which will allow developing different types of algorithms, i.e. to compare histograms, to analyze histograms, to compare values, etc…

URF015 Results of DQ algorithms

The DQMF shall support custom DQ algorithms, which can return different types of results.

Priority High

Note DQM shall be able to accept different values as results of the custom DQ algorithms, i.e. simple binary value (true/false), reliability (0-100%), histogram, etc…

URF016 DQM actions

The DQMF shall be able to respond to results of comparisons by sending messages to the Control services.

Priority High

URF017 Commands to monitoring data providers

The DQMF shall be able to send commands to monitoring data providers (i.e. PT application, RCD application, etc.)

Priority High

URF018 DQM alarms

The DQMF shall be able to respond to results of comparisons by issuing alarms when deviations are present.

Priority High

URF019 DQM reaction on external conditions

DQMF shall be able to receive and react to changes of external conditions like luminosity, pre-scaling factors, run type, etc.

Priority Medium

URF020 Change reference sets

Depending on actual external conditions, the DQMF should be able to automatically switch during a run from one set of reference values to another, if such a different set was provided by the expert.

Priority Medium

URF021 DQM checks logs

DQMF shall be able to log what set of references has been used at certain check to allow traceability.

Priority Medium

3.3.3 DQM output

URF022 Publish DQ results

The DQMF shall be able to publish DQ results (including histograms) to the Online Monitoring Services.

Priority High

URF023 History of DQ results

The DQMF shall keep the history of the DQ results over a run.

Priority High

URF024 Displaying DQ results

The DQMF shall present the DQ results to the shift operator in real time. This will facilitate examination of these results in run sub-units appropriate for final DQ judgments, such as luminosity blocks.

Priority High

URF025 Final DQ assessment

The DQMF shall develop and present to a human operator the final DQ assessment for all run sub-units at the end of the run.

Priority High

URF026 Confirmation for the final DQ assessment

The DQMF shall ask for confirmation from human operator before accepting the final DQ assessment for a given run.

Priority High

URF027 Editing DQ assessment

The DQMF shall provide the shift personnel with the possibility to modify the final DQ assessment before accepting it.

Priority High

URF028 Archiving DQ results

The DQMF shall archive the complete history of the DQ results to permanent data storage during a run.

Priority High

URF029 Browsing archived DQ results

The DQMF shall provide a possibility to browse the archived DQ results.

Priority Medium

3.3.4 DQM Graphical User Interfaces

URF030 Configuration GUI

The DQMF shall provide a GUI for creating and modifying configuration parameters.

Priority High

URF031 Control GUI

The DQMF shall provide a control GUI, which shall be able to:

- display the currently active DQM configuration
- display results of the DQ analysis algorithms
- allow to set the final DQ assessment for the current run

Priority High

URF032 Archive browser

The DQMF shall provide a GUI for browsing the DQ results from previous runs.

Priority Medium

URF033 Online Web access

A web interface shall be provided to have read-only access to the DQ results for the ongoing run.

Priority Medium

URF034 Offline Web access

A web interface shall be provided to browse the archive of the DQ results.

Priority Medium

3.4 Non-Functional Requirements

URN001 Scalability

The DQMF shall scale to the final size of the experiment.

Priority Medium

This document has been prepared using the Requirements Document Template provided and approved by the Atlas TDAQ and DCS Connect Forum. For more information, go to

http://atlas-connect-forum.web.cern.ch/Atlas-connect-forum/.

This template is based on the SDLT Single File Template that have been prepared by the IPT Group (Information, Process and Technology), IT Division, CERN (The European Laboratory for Particle Physics) and then converted to MS Word. For more information, go to http://framemaker.cern.ch/.