Journal of Physics: Conference Series

OPEN ACCESS

# The CMS data quality monitoring software: experience and future prospects

To cite this article: Federico De Guio and the Cms Collaboration 2014 *J. Phys.: Conf. Ser.* **513** 032024

View the article online for updates and enhancements.

## Related content

- CMS data quality monitoring: Systems and experiences
  L Tuura, A Meyer, I Segoni et al.
- Data Quality Monitoring of the CMS Tracker
  Suchandra Dutta and the CMS Tracker Collaboration
- CMS data processing workflows during an extended cosmic ray run
  CMS Collaboration

## Recent citations

- Continuous and fast calibration of the CMS experiment: design of the automated workflows and operational experience
  P Oramus *et al*

This content was downloaded from IP address 128.141.160.21 on 10/06/2018 at 19:20

Journal of Physics: Conference Series **513** (2014) 032024 doi:10.1088/1742-6596/513/3/032024

# The CMS data quality monitoring software: experience and future prospects

**Federico De Guio on behalf of the CMS Collaboration**

CERN, Geneva, Switzerland

E-mail: `federico.de.guio@cern.ch`

**Abstract.** The Data Quality Monitoring (DQM) Software proved to be a central tool in the CMS experiment. Its flexibility allowed its integration in several environments: Online, for real-time detector monitoring; Offline, for the final, fine-grained Data Certification; Release-Validation, to constantly validate the functionality and the performance of the reconstruction software; in Monte Carlo productions. The central tool to deliver Data Quality information is a web site for browsing data quality histograms (DQM GUI). In this contribution the structure of the DQM framework is described and the usage of the DQM software in the different environments and the performance of the system after the first years of data taking are presented.

## 1. Introduction

The Compact Muon Solenoid (CMS) [1] is a multi-purpose detector at the Large Hadron Collider [2] at CERN. Data Quality Monitoring (DQM) is critically important for the detector and operation efficiency and for a reliable certification of the recorded data. The CMS collaboration has adopted a single end-to-end DQM chain [3, 4].

The system comprises:

- tools for booking, filling, handling and archiving histograms and scalar monitor elements;
- standardised interface for algorithms performing automated quality and validity tests;
- systems for: online monitoring of the detector, the trigger, the DAQ hardware status and data throughput; offline monitoring of the reconstruction; validation of calibration results, software releases, simulated data;
- visualization of the monitoring results (DQM GUI);
- workflows and tools for the certification of datasets and subsets thereof for physics analyses;
- tools for retrieval of DQM quantities from the conditions database;
- standardization and integration of DQM components in CMS software releases.

The high-level goal of the system is to discover and pinpoint errors, problems occurring in detector hardware or reconstruction software, early, with sufficient accuracy and clarity to maintain good detector and operation efficiency. In the following sections the structure of the DQM framework is presented together with the performance achieved during the first years of data taking. In the last section the ongoing development are briefly described.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

## 2. DQM GUI and framework

### *2.1. The monitor element*

The histograms exist inside the DQM framework as monitor elements (ME) which are the central monitoring tools. The core of each ME is a ROOT [5] object which is enriched with additional information such as the output of quality tests, folder hierarchy, flags. The MEs are booked inside the DQMStore class which is the shared container that holds all the monitoring information (figure 1). The content of the DQMStore is available for the whole duration of the job and can be persisted on disk in ROOT format.

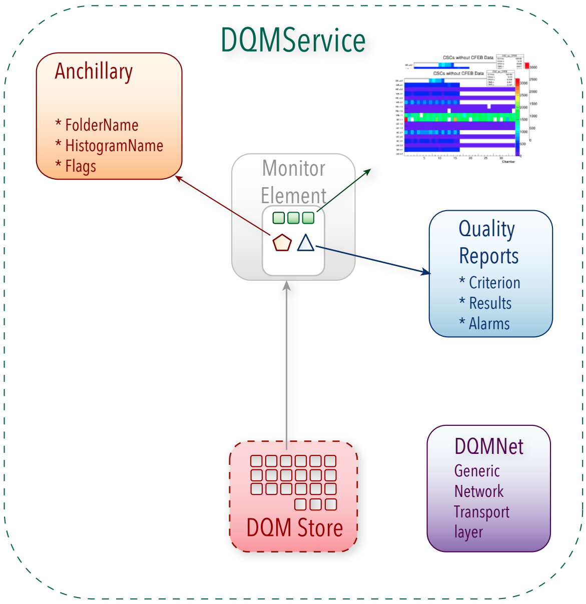



**Figure 1.** The Monitor Element is the central monitoring tool; the DQMStore is the shared containers that holds all the monitor elements.

### *2.2. The DQM framework*

The DQM framework is currently run based and can be ideally divided in two parts: the Online and the Offline DQM systems.

The online DQM applications are an integral part of the event data processing happening on the online cluster. They receive event data and trigger histograms at the rate of about 10-15 Hz. Each application receives events from the storage manager proxy over HTTP and runs its choice of algorithms and analysis modules generating results in the form of monitoring elements including reference histograms and quality test results.

Numerous offline workflows involve data quality monitoring. These systems vary considerably in location of the computing resources, data content and timing, but as far as DQM is concerned, CMS has standardised on a two-step process for all these activities. In the first step the MEs are created and filled with information from the CMS event data. This step is usually run in parallel for different events in the same run. During a second step the histograms are extracted from the files and summed across the entire run to yield full event statistics on the entire dataset. The final histograms are used to calculate quantities such as trigger efficiencies and are compared against reference distributions. Simple quality tests (i.e. histogram mean included in a given range) can be implemented to automatically check the content of the histograms and raise alarms (a system of histograms flagging) in case some of the conditions are not observed.

The histograms and the quality test results along with any alarms are output to a ROOT file, which is then uploaded to a central DQM GUI web server.

### *2.3. The DQM GUI*
The central component of the Data Quality Monitoring system is a web site for browsing data quality histograms. It guarantees authenticated Worldwide access. It is a customizable application capable of delivering visualization for all the DQM needs, for all subsystems, for live data taking as much as archives and offline workflows. Some snapshot of the DQM GUI is shown in figure 2. The content is organized in workspaces depending on the scope and ranging from high-level summaries to shift views to expert areas. Within a workspace histograms can be organized into layouts to bundle related information together.

## 3. Operation and experience
The DQM system is in production since 2008 and behaved very well during more than four years of data taking. A total of more than three million of unique histograms are accessible through different instances of the GUI in a quick and easy way. Moreover the GUI index is exposed via several APIs (PNGs, ROOT, JSONs available as URI or via CLI) that serve both private and central tools giving the possibility to customize analysis based on the DQM histograms.

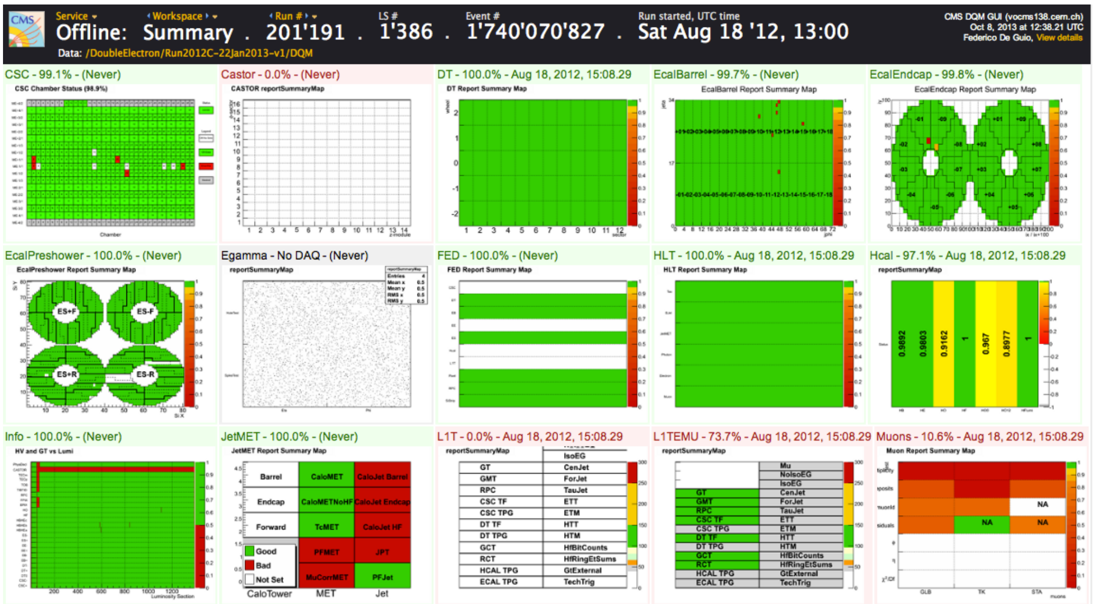
The performance of the GUI has been excellent: the response time was on average not above 100 ms despite the number of accesses per day to both the online and the offline servers was often exceeding one million and the number of monitored histograms is continuously increasing (figure 3). The DQM framework proved to be very flexible. Hundreds of modules compose the so called ‘DQM sequence’ and monitor both low level and high level variables at all the steps of the event reconstruction.

**Table 1.** Some figures about the Offline GUI server and the RelVal one. The latter is the instance of the GUI used to collect the histograms needed to validate the CMS software releases. The number of histograms stored per run in each instance of the DQM GUI is about 50000.

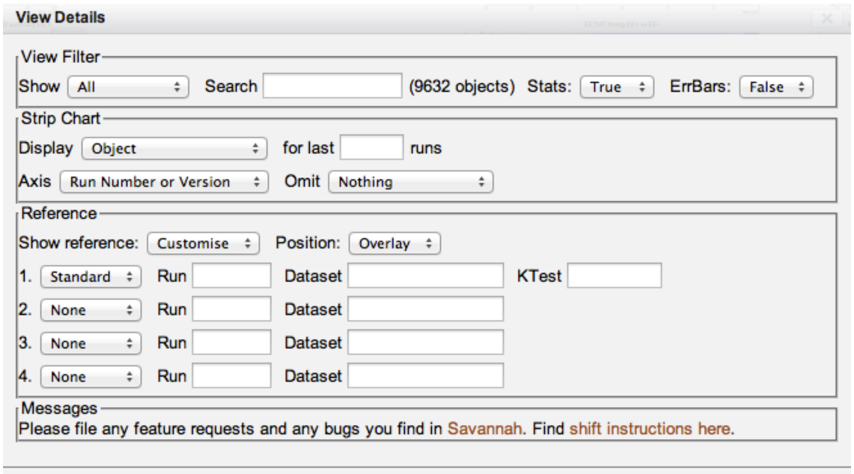
| Quantity | Offline server | RelVal server |
|---|---|---|
| Samples | 264,383 | 31,587 |
| Source files | 303,580 | 31,744 |
| Datasets | 2,068 | 18,379 |
| Unique objects (Mes) | 913,174 | 1,649,086 |
| Disk space (TB) | 2.5 | 0.38 |

## 4. Certification of the acquired data
CMS uses an application named Run Registry (a database with a front-end web application) as tracking and bookkeeping tool for the central workflow managing the creation of the final physics dataset certification result. The run registry is both a user interface managing the workflow and a persistent store of the information.

The certification process begins with the physicists on shift filling in the run registry with basic run information and adding any pertinent observations done using the Online DQM during the data taking. The analyzed statistic is partial in this case. A team of ‘offline shifters’ and sub-detectors experts takes care of cross checking the data quality once the whole statistics has been processed and made available in the offline instance of the DQM GUI. For each detector segment a single boolean flag describes the final quality result.




(a)




(b)

**Figure 2.** A snapshot of the Offline DQM GUI global summary (a). (b) shows the snapshot of the 'View Details' panel where is possible to add a filter on the name of the MEs, plot the mean or the RMS of the selected distribution as a function of the run number and overlay a reference distribution from a different dataset. A basic histogram style editor is also available in the GUI.

During the data taking the certification sign-off is done on a weekly basis and the final verdict on the quality of the collected data is delivered to the collaboration. The stored quality flags are meant to be used in any subsequent data re-processing and longer-term data quality evaluation.

## 5. Future development

The DQM Framework performed well during LHC-Run 1 phase, yet some developments are required to adopt foreseen changes in LHC conditions and to keep apace with the most recent technology trend. On the CMS software side, the DQM framework will move to a new Multi-

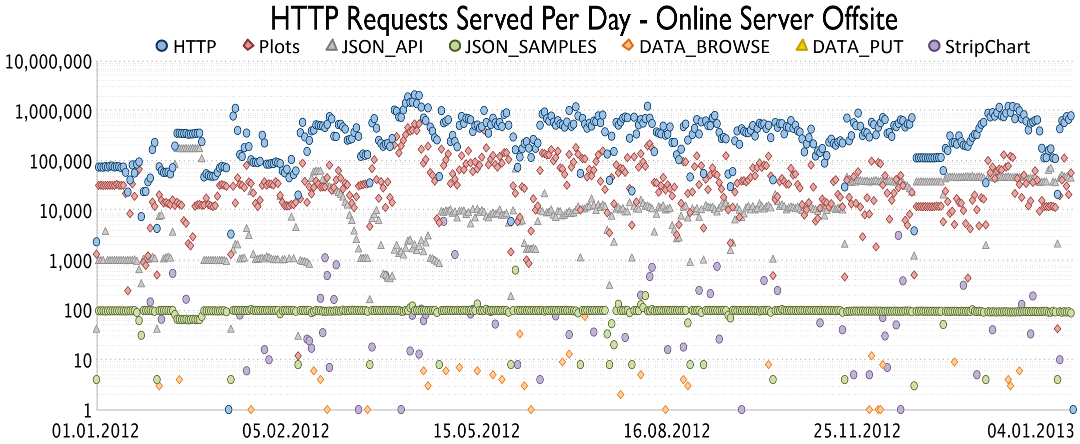



**Figure 3.** Number of different kind of server requests per day as a function of time for the 2012 run until the end of the data taking. The data refer to the server hosting the online instance of the DQM GUI.

Core Multi-Thread design. New functionalities will be also made available. The possibility to compare data and Montecarlo samples (an operation which is usually performed at analysis level) directly in the DQM GUI will be implemented together with the capability to stack and display more than one MC sample where each one can be previously normalized. The goal is to validate the simulation of the detector against the data in a central way and on a regular basis. On the online side, the possibility to save a set of key variables to a database will create the possibility to generate correlation plots between different monitored variables adding even more flexibility to a tool which is already integrated in different environments. Finally the review of the content of the histograms and the optimization and tuning of the quality criteria and the alarms go in the direction of having a fully automated certification of the quality of the acquired data. This is desirable in order to minimize the risk of human mistakes and to speed up the certification procedure.

## References

[1] The CMS Experiment at the CERN LHC, JINST 3 S08004, 2008.
[2] LHC Machine, JINST 3 S08001, 2008.
[3] CMS data quality monitoring: systems and experiences, L. Tuura et al, CHEP09.
[4] CMS Offline Web Tools, S. Metson et al, CHEP07.
[5] ROOT - A data analysis framework, http://root.cern.ch