

# The CMS Data Quality Monitoring software: Experience and future improvements

**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 (DQMGUI). In this contribution the usage of the DQM Software in the different environments and its integration in the CMS Reconstruction Software Framework and in all production workflows are presented.

## 1. Overview

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

The system comprises:

- Tools for booking, filling, handling and archiving histograms and scalar monitor elements.
- Standardized interface for algorithms performing automated quality tests;
- Systems for: online monitoring of the detector, the trigger, the DAQ hardware status and data throughput; offline monitoring of the reconstruction; validation of the calibration results, the software releases, the simulated data;
- Visualization of the monitoring results (DQMGUI);
- 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 reach good detector and operation efficiency.

## 2. DQM GUI and Framework

### The DQM GUI

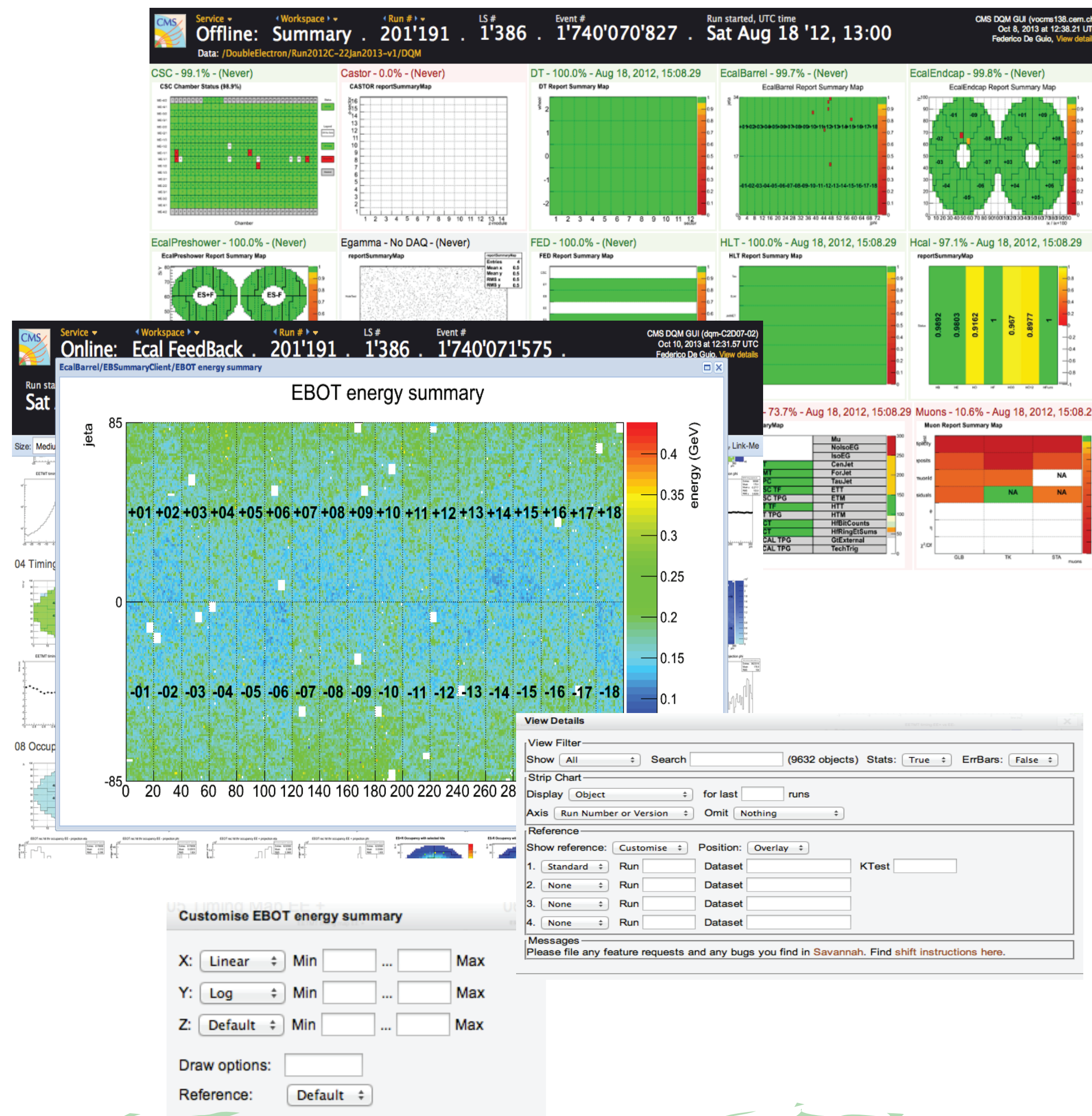
The central component of the data quality monitoring system of the CMS Experiment is a web site for browsing data quality histograms. It guarantees authenticated Worldwide access. It is a single customizable application capable of delivering visualization for all the DQM needs in all of CMS, for all subsystems, for live data taking as much as archives and offline workflows. The content is organized in workspaces from high-level summaries to shift views to expert areas, including even a basic histogram style editor. Within a workspace histograms can be organized into layouts to bundle related information together.

### 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 [4] 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. The content of the DQMStore is available for the whole duration of the job and can be persisted on disk in ROOT format.

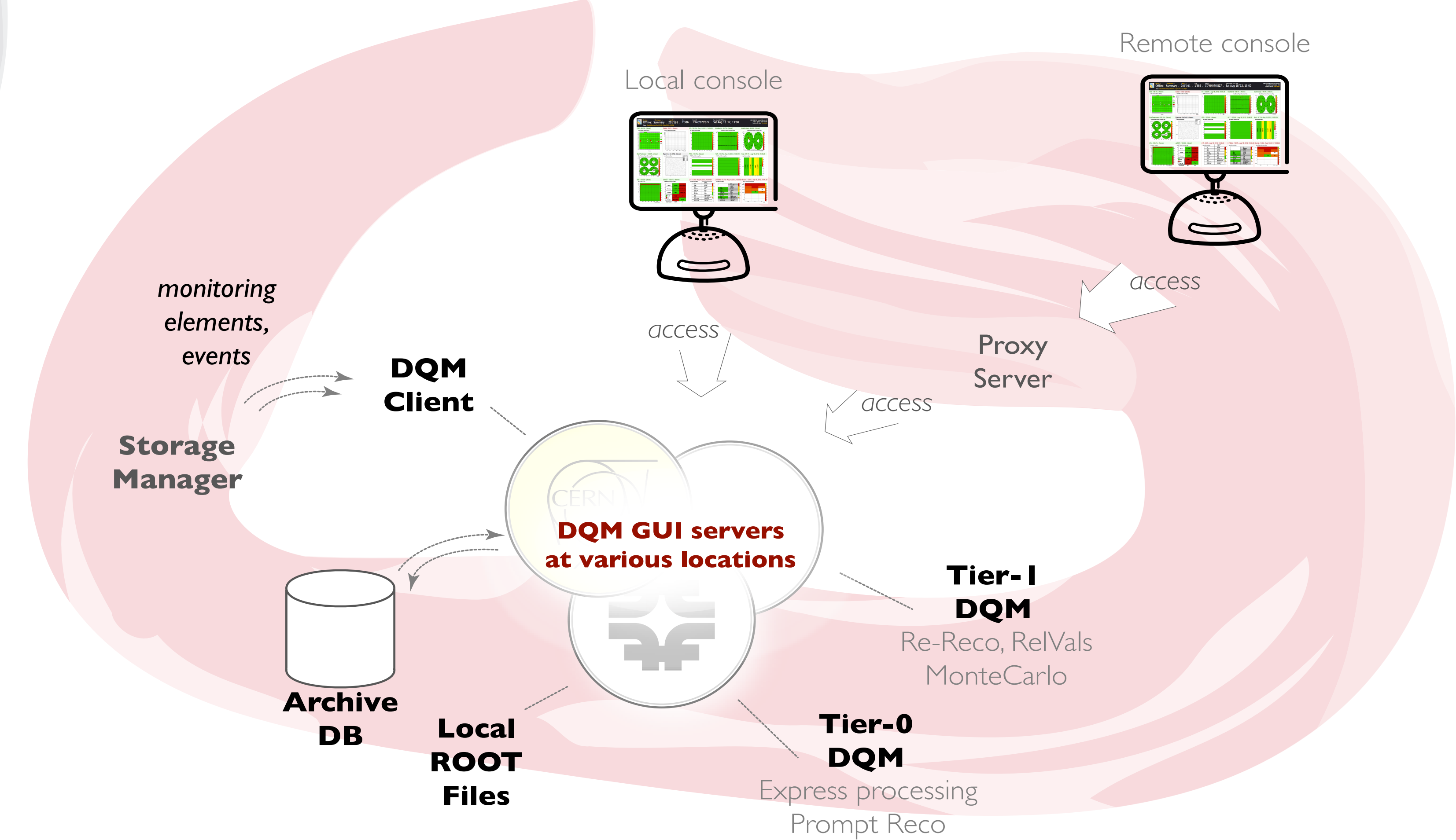
### The DQM framework

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



### CMS DQM main features

- DQM aggregates event level information that is sensitive to both detector and software problems in one central place (web server).
- Fast turn-around and small latency. Update of info frequently during data taking. Quality tests up to every lumi-section.
- Automatic alarms notify about problems.
- Synoptic overview of detector status (front page) and detailed histograms for inspection by the experts.
- The DQM information is key input to the creation of the good run list used for analysis.
- The web server is accessible from everywhere
- DQM is maintainable in a modular way by subsystems
- DQM runs in spy mode, in order to not interfere with the data taking.



The online DQM applications are an integral part of the rest of the event data processing at the cluster at CMS Point-5. 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 and generates its results in the form of monitoring elements and 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 efficiencies and checked for quality, in particular compared against reference distributions. 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.

## 3. Operation and experience

The DQM system is in production since 2008 and behaved very well during three years of data taking. More than three millions of histograms are accessible in different instances of the GUI in a quick and easy way. Moreover the GUI index is exposed via several APIs (PNGs, ROOT, JSONs [5] 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 have 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.

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.

## 4. Future development

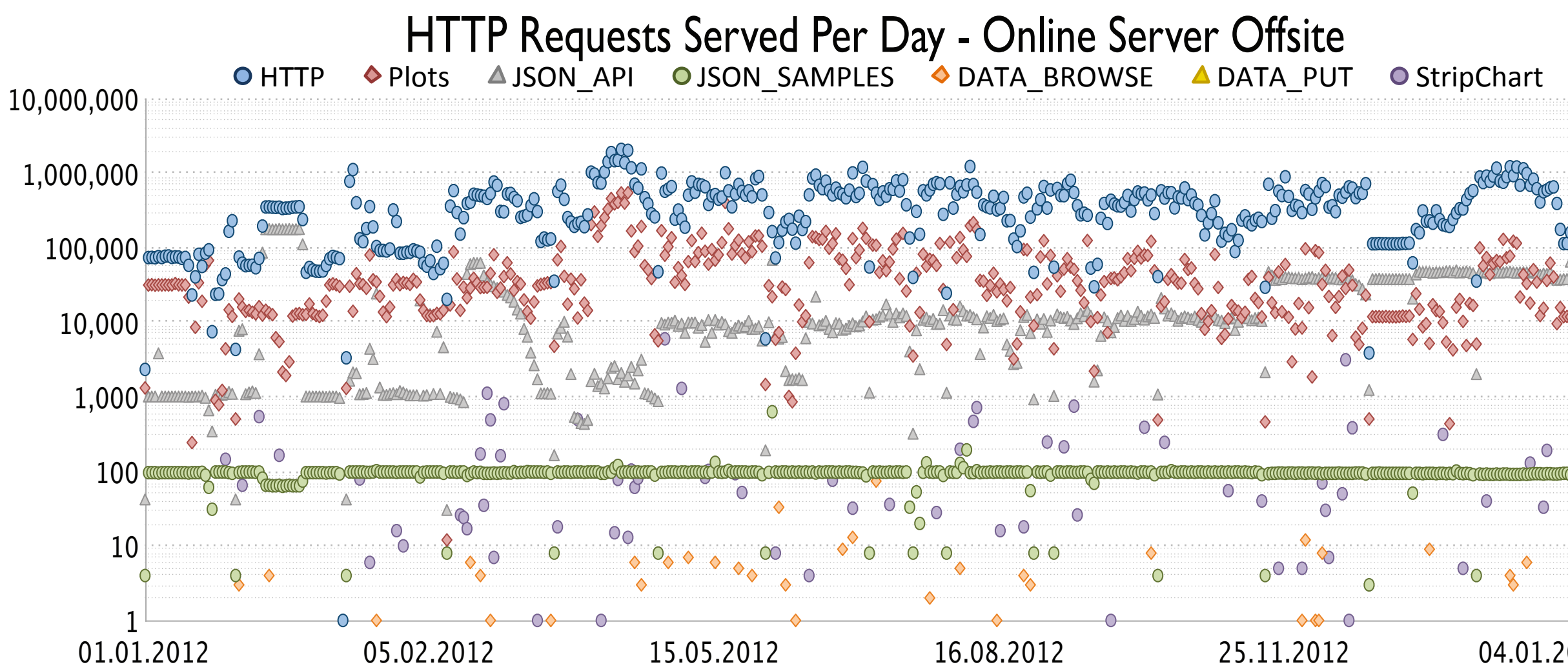
The DQM Framework performed well during LHC-Run 1 phase, yet some improvements are required to cope with foreseen changes in LHC conditions and to adopt the most recent technology trend.

On the CMS software side, the DQM framework will move to the new Multi-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.

On the online side, the possibility to save a set of key variables to a database will in principle give the possibility to plot 'everything vs everything' 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 full automatic certification of the quality of the acquired data. This is auspicious in order to minimize the risk of human errors and to speed up the certification procedure.



## References

- [1] CMS: the Compact Muon Solenoid technical proposal. LHC Tech. Proposal. CERN, Geneva, 1994.
- [2] CMS data quality monitoring: systems and experiences, L. Tuura et al, CHEP'09.
- [3] CMS Offline Web Tools, S. Metson et al, CHEP'07.
- [4] ROOT - A data analysis framework, <http://root.cern.ch>
- [5] <http://json.org>

