A Web-based application for the collection, management and release of Alignment and Calibration configurations used in data processing at the Compact Muon Solenoid experiment

IOP Conf. Series: Journal of Physics: Conf. Series **898** (2017) 032034 doi:10.1088/1742-6596/898/3/032034

A Web-based application for the collection, management and release of Alignment and Calibration configurations used in data processing at the Compact Muon Solenoid experiment

Audrius Mecionis[2], Salvatore Di Guida[3], Giovanni Franzoni[1], Marco Musich[4], Gianluca Cerminara[1], Andreas Pfeiffer[1], Giacomo Govi[5] on behalf of the CMS Collaboration

[1]CERN, Geneva, Switzerland
[2]Vilnius University, Vilnius, Lithuania
[3]Università degli Studi Guglielmo Marconi, Roma, Italy
[4]Université catholique de Louvain, Louvain-la-Neuve, Belgium
[5]Fermi National Accelerator Laboratory, Batavia, USA

E-mail: audrius.mecionis@cern.ch

Abstract. The Compact Muon Solenoid (CMS) experiment makes a vast use of alignment and calibration measurements in several data processing workflows: in the High Level Trigger, in the processing of the recorded collisions and in the production of simulated events for data analysis and studies of detector upgrades. A complete alignment and calibration scenario is factored in approximately three-hundred records, which are updated independently and can have a time-dependent content, to reflect the evolution of the detector and data taking conditions. Given the complexity of the CMS condition scenarios and the large number (50) of experts who actively measure and release calibration data, in 2015 a novel web-based service has been developed to structure and streamline their management. The cmsDbBrowser provides an intuitive and easily accessible entry point for the navigation of existing conditions by any CMS member, for the bookkeeping of record updates and for the actual composition of complete calibration scenarios. This paper describes the design, choice of technologies and the first year of usage in production of the cmsDbBrowser.

1. Introduction

The Compact Muon Solenoid (CMS) is a multipurpose detector operated at the CERN Large Hadron Collider (LHC)[1]. Non event calibration and alignment data are essential to describe the evolving status of detector components and to maintain the performance of the experiment. There are several different workflows at CMS which consume non event data:

- the High Level Trigger (HLT)[2] to select the collisions to be recorded;
- the processing of the raw data of recorded collisions;
- the production of simulated events for data analysis and studies of detector upgrades.

A complete alignment and calibration scenario is factored in approximately three-hundred records, which are updated independently and can each have a time-dependent content, to reflect the evolution of the detector and data taking conditions. This paper gives an overview of a novel web-based service

called cmsDbBrowser which allows CMS condition experts who actively measure and release calibration data to streamline and structure the bookkeeping.

2. Conditions data in CMS

2.1. Conditions overview

Non event data, which describe the status of the CMS detector as it evolves in time, are crucial for the optimal performance of the reconstruction of collision events coming from simulated or real data, as well as for physics analysis. The non event data, also referred to as conditions, are the result of calibration and alignment procedures. Based on the CMS data model, the CMS conditions are stored into relational databases as a set of binary objects (BLOBs)[3], serialized using boost libraries within the CMS offline software framework. The CMS Condition Database system[4] relies on one ORACLE 11g Release 2 Real Application Cluster and its Active Standby copy, housing two main database services:

- **OMDS** allows to handle the configuration and condition data produced online by the sub-detectors in relational tables.

- **ORCON** serves a subset of condition data needed by HLT, data quality monitoring (DQM) and offline processing of recorded events. Non event data are stored and retrieved for consumption as C++ objects.

Two instances of the condition databases exist and are synchronized via the Oracle Active Data Guard (ADG) technology, in order to assure that the replicated content of their master copies, located in the private network of the CMS experiment, are also accessible from the CERN-wide general purpose network in a read-only manner.

2.2. Conditions data model

- **Payload** is the "atom" of conditions data, representing the set of parameters which holds a specific type of calibration or alignments constants consumed in any of the workflows of the data processing (e.g. reconstruction of raw data). It is associated to a user-defined C++ class in CMS software framework (*Payload Type*)[4]. The Payload data is exchanged and stored as an unstructured binary array, with no assumption on its internal layout. Each Payload is coupled to a unique **Payload hash**. This can be thought as a fingerprint of each payload. The hash is computed via SHA1 algorithm and depends on:
    - The actual conditions stored in the payload;
    - The version of the boost library used to serialize the payload when uploaded;

- **Record** is a dedicated C++ class in the CMS software framework which acts as entry point for a Payload by importing the database content on in the CMSSW executable program. Every data processing workflow involves a specific set of Records, each of them requiring valid conditions;

- **Interval Of Validity** (**IOV**) is the time interval during which a Payload has to be consumed. Time is represented by a Run number, luminosity section id or a universal timestamp;

- **Tag** is a label which identifies the fully defined history for a given calibration or alignment content; each tag comprises a set of IOVs and their associate Payloads covering the time span required by the workflow;

- **Global Tag** is a consistent set of Tags assigned to the Records involved in a given workflow. Thanks to global tags, the data processing managers can handle well defined collections of Tags and configure the CMSSW executable to consistently consume non event data for large-scale

data production. A Global Tags provides a calibration and alignment content, for instance, for a Monte Carlo production scenario or data reprocessing.

3. Overview of the CMS Conditions Browser

3.1. Introduction

The CMS conditions database browser service (cmsDbBrowser) was created in order to:

- provide an intuitive and easy way to inspect, navigate and make user-defined search queries to the existing conditions data and metadata; Any CMS member should be able to easily carry out such searches. As shown in figure 1, the result of the query is presented in a categorized and structured way and displays Tags, Global Tags and Payloads in data tables;
- ease the bookkeeping and management of the condition metadata by condition managers:
  - handle update requests submitted by detector experts;
  - compose complete calibration scenarios like Global Tags;
- provide a single entry point for monitoring all condition related services;
- allow to inspect the actual Payload data with the help of the Payload Inspector service.

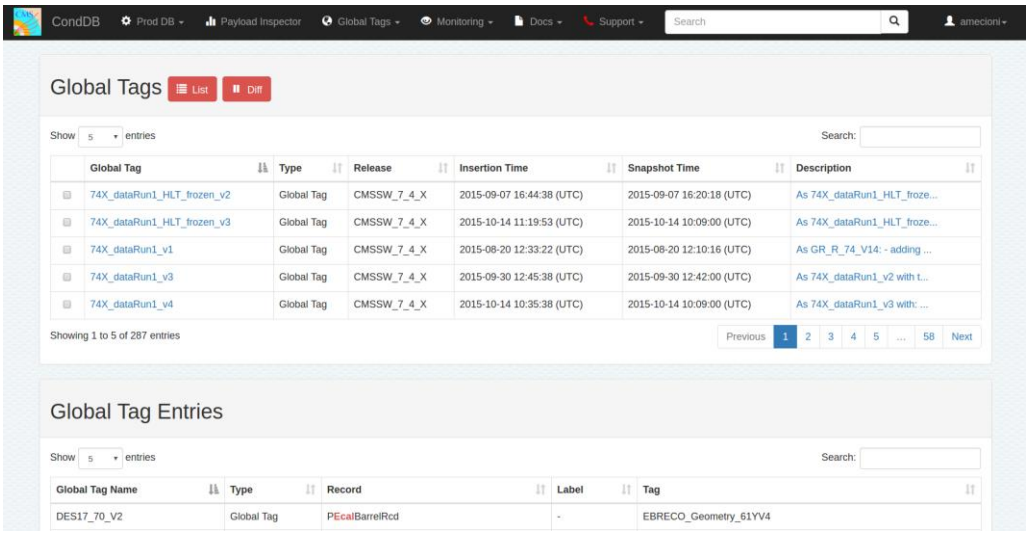

Figure 1: The CMS Condition Browser Service.

3.2. User types

There are three roles/types of users of the cmsDbBrowser service:

- CMS Collaborator - can easily navigate and browse all the conditions related data. Also he/she can draw various plots of deserialized Payload data, consult the monitoring of all the services the storage and consumption of condition rely upon;

- Calibration Expert - has a simple way to request a record updates for a specific production workflow, as well as create a special Global Tag type called a Candidate to test their requested Payload changes in 11g;

- Condition Manager - has the ability to manage all the condition metadata and to compose complete calibration scenarios (Global Tags) to be consumed in production workflows. Also, as shown in figure 2, to accept or reject record update requests done by the calibration experts.

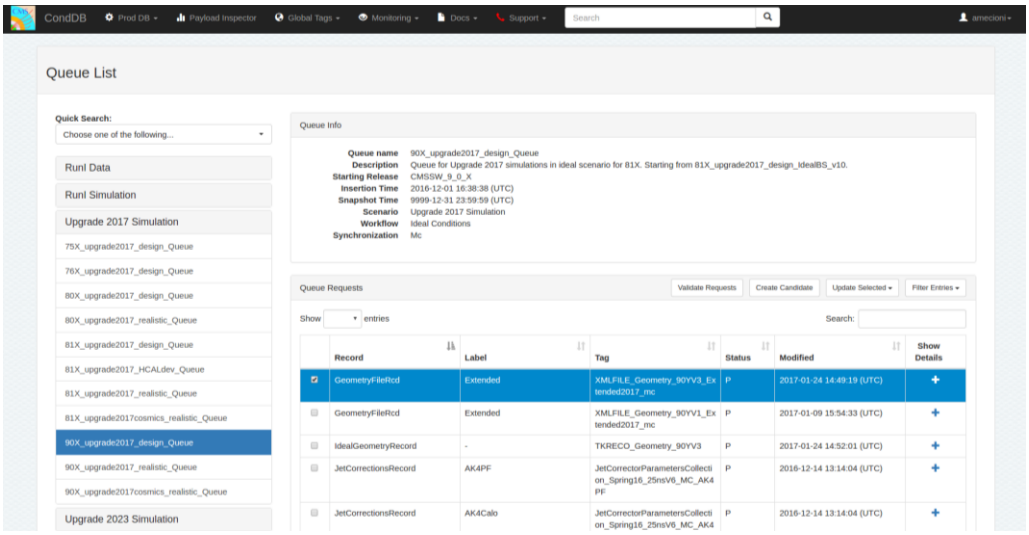

Figure 2: Condition metadata management in cmsDbBrowser.

3.3. Service architecture

To read condition data from the database cmsDbBrowser, together with similar services designed to work with condition data, uses the read-only ADG copy which stores replicated content of the master condition database. For security and reliability reasons, calibration data can be written in the condition database only from within the technical network of the experiment which is protected by a firewall separating it from the general purpose CERN network.

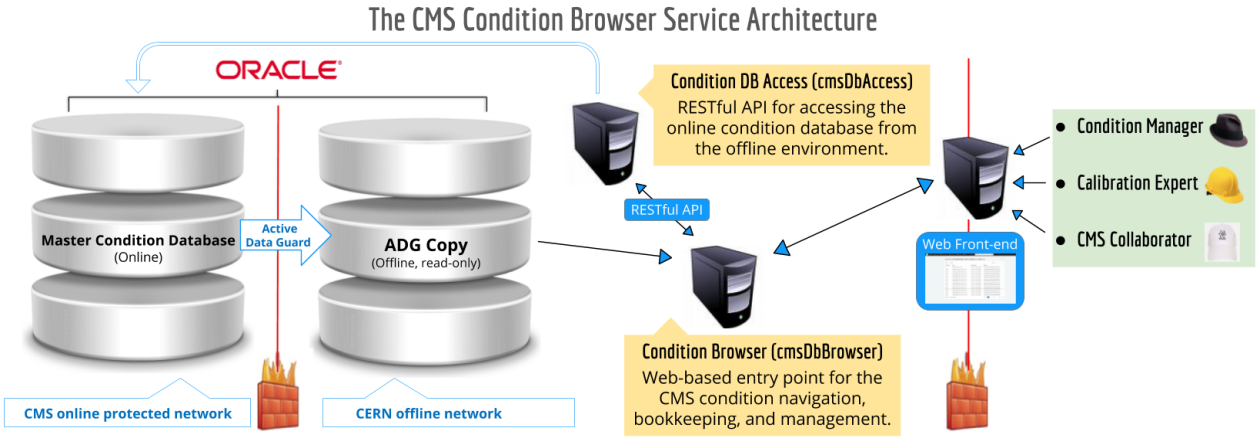

Figure 3: The CMS Condition Browser Service Architecture.

As shown in figure 3, this design does not allow a direct export to the master Oracle database where all the conditions data are stored. To solve this issue a RESTfull API to the main condition database was created as a separate service called cmsDbAccess. In this way the access to the master database is securely isolated and separated from the main cmsDbBrowser service. The cmsDbAccess API is deployed in the special machine which has an open gateway to the CMS technical environment and can access the main condition database directly.

3.4. Design and Implementation choices

The backend of the cmsDbBrowser service is implemented in Python programming language and Flask web framework. Python was chosen as the base programming language because of its simplicity and flexibility. As an Object Relational Mapper SQLAlchemy is used to handle all the database connections, manage transactions and abstract database querying. For creating a pleasant user experience the Bootstrap CSS framework is used together with the jQuery Javascript library, which allowed to create a modern user interface with advanced functionality. This set of technologies have proven to be reliable, efficient and secure. It meets our end-users requirements and provides our infrastructure with good performance and stability.

4. Conclusions

The CMS conditions database browser service (cmsDbBrowser) was created to streamline and structure the management of alignment and calibration sets which are fundamental to the CMS detector operated at the CERN Large Hadron Collider (LHC). In addition, the service was designed to provide a single entry point for any CMS member to query the conditions database and give an answer to all conditions-related needs. Given the complex CMS experiment environment, the latest state-of-the-art technologies was used to create a service which have proven to be stable and reliable. It meets the requirements of the CMS collaborators and saves a lot of valuable time for the experts and managers.

References

[1] CMS Collaboration, “The CMS experiment at the CERN LHC” (2008) JINST 3 S08004, doi:10.1088/1748- 0221/3/08/S08004
[2] CMS Collaboration, “CMS High Level Trigger” CERN/LHCC 2007-021 (2007)
[3] Pfeiffer A, Govi G, Ojeda M, “Multi-threaded Object Streaming”, *J. Phys.: Conf. Ser.* **664** 042044 (2015), doi:10.1088/1742-6596/664/4/042024
[4] Govi G, Di Guida S, Pfeiffer A, Ojeda M, “The CMS Condition Database system”, *J. Phys.: Conf. Ser.* **664** 042024 (2015), doi:10.1088/1742-6596/664/4/042024