CMS Alignment and Calibration Framework

Matthias Edelhoff (on behalf of the CMS Collaboration)

I. Physikalisches Institut B, RWTH Aachen, Sommerfeldstr. 14, D-52074 Aachen

Abstract. Precise and prompt alignment and calibration will be crucial to the performance of the CMS detector. A software and computing infrastructure has been commissioned, which accomplishes both tasks in a swift and reliable manner. First, incoming data from the experiment is processed and skimmed automatically at the Tier-0 site. Reduction of data size is of prime interest in this stage, to ensure efficient use of computing facilities and fast turnaround-times. Second, a versatile computing farm (CERN Analysis Farm) enables a fast execution of the alignment and calibration workflows. Both steps are coordinated and carried out for a large number of interdependent processes. The resulting database payloads are used for all subsequent reconstruction steps, providing optimised measurement precision to physics analysis. The framework is described alongside to the validation of its implementation. To this end some results using cosmic data taken in 2008 are shown.

Keywords: LHC CMS Alignment Calibration Framework
PACS: 29.85.-c, 29.85.Ca

INTRODUCTION

High precision in alignment and calibration is needed in order to reach the design performance of the Compact Muon Solenoid (CMS) experiment [1] at the Large Hadron Collider (LHC). To enable a fast turnaround of data for analysis, this precision has to be reached on a short timescale. Some of the calibration measurements can be carried out directly at the experimental site (Point 5) on a run-by-run basis (e.g. measurements of pedestals or gains). As opposed to these "online conditions", others are measured with larger latencies and are referred to as "offline conditions". In order to determine conditions of both categories with the required precision and speed, the alignment and calibration framework of CMS provides:

- Fast flow of specific data needed for alignment and calibration. This is achieved using dedicated data streams (express and calibration/monitoring streams).
- Fast turnaround for alignment and calibration tasks, including the ability to handle many tasks in parallel. The CERN Analysis Facility (CAF) provides this ability to all tasks determining offline conditions.
- Database infrastructure for handling of the alignment and calibration conditions, which is ensured by a combination of three databases situated at Point 5 and the CERN site.

REDUCED DATA FORMATS FOR ALIGNMENT AND CALIBRATION

There are two reduced data formats in use in the alignment and calibration workflow. One is referred to as AlCaReco and is a skim of the recorded data which is used by one or more alignment or calibration tasks. The other one is called AlCaRaw. Its purpose is to transport high rate data from Point 5 to the Tier-0 when bandwidth limitations do not allow to use the standard data format.

Alignment and calibration skims (AlCaReco)

Several alignment and calibration tasks need to run or even iterate on a large number of events. The reduced AlCaReco formats were conceived, in order to retain fast turnaround of the tasks, minimize input/output overhead and maximize their storage efficiency.
As CMS uses very different detector technologies, the alignment and calibration tasks are diverse. This diversity leads to widely differing input to the separate tasks. For each of the tasks the minimal event content needed for its execution was defined. Based on these minimal requirements, the AlCaReco formats were created so that each of the AlCaReco

skims can be used by one or more tasks. The content of the AlCaReco skims ranges from reconstructed objects, like tracks and clusters used for the track based alignment of the inner tracker and the muon system, to non event data like measurements of the laser alignment system in the silicon strip tracker. Furthermore, information normally not available after event reconstruction can be stored in the AlCaReco skims. For example the digitized measurements of the DT muon chambers is stored and used for drift velocity calibration.
Data Quality Monitoring (DQM) information is stored in each AlCaReco skim in order to verify the contained information. For all AlCaReco formats the reduction takes place by filtering events and event content. Events are filtered exploiting first High Level Trigger (HLT) and then offline reconstruction information. The HLT filtering is configured using the same database infrastructure that alignment and calibration conditions are stored in. Thus, flexible adjustments can be made in the face of changing HLT menu definitions during the early phase of collision data taking. Event content is filtered to meet the minimal requirements of the alignment and calibration tasks.

Reduced raw data format (AlCaRaw) for alignment and calibration

Some calibrations use event signatures occurring at a very high rate that would saturate the available bandwidth between the detector and the Tier-0 site if the full raw event content were transferred. For this reason, dedicated data streams, referred to as AlCaRaw streams are set up, which contain only the event content required for the corresponding calibration procedure, allowing the data in these streams to be transferred at a high rate.

Storage and distribution of alignment and calibration conditions

There are two main goals of the database infrastructure distributing alignment and calibration conditions. First, CMS data taking has to be independent of any off-site networks. This means that all databases needed for online operations have to be physically present at Point 5 in the online network. Second, conditions need to be accessible from the offline network and the whole Grid infrastructure [2].
To meet these goals three databases have been deployed (Figure 1 on the left):

OMDS (Online Master Database System) Central registry for online conditions.

ORCON (Offline Reconstruction Condition DB ONline subset) For offline conditions read by the online network.

ORCOFF (Offline Reconstruction Condition DB OFFline subset) For offline conditions read by the offline network.

OMDS is filled directly from the respective online calibrations. Part of its content is transferred to ORCON using the POPCON service [3]. POPCON is also used to update ORCON with offline conditions derived at the CAF after sign-off. In the process an interval of validity is defined and stored with each of the sets of conditions. After each update the high level trigger can make use of the improved conditions. This way the first goal is met by placing OMDS and ORCON at Point 5.
The ORCOFF database is kept in sync with the ORCON database via Oracle streaming. As a result, a replica of all conditions is available to the offline network at any time. This information is distributed through a hierarchy of proxies (Frontier) throughout the Grid. This ensures the second of the stated goals.

WORKFLOW FOR ALIGNMENT AND CALIBRATION

The alignment and calibration workflow is embedded into the general CMS computing model [4] for data taking and first processing (Figure 1 on the right).
After data has been taken at Point 5 it is transferred in three different types of stream:

Physics stream consists of events on the physics HLT menu.

Express stream includes events selected for alignment and calibration based on the HLT decision.

Calibration/Monitoring streams contain measurements by specialized hardware and AlCaRaw produced at Point 5.

The content of all three streams can be used to create AlCaReco skims at the Tier-0 or any other Tier in the Grid hierarchy. Those AlCaReco skims are then transferred to the CERN Analysis Facility (CAF) where offline alignment

and calibration tasks are executed.
The CAF is a batch farm currently providing 736 cores and hosting 1.3 PB of disk storage which is in part used to store AlCaReco skims and intermediate data. Dedicated job queues and the large disk storage enable fast task execution, validation, and debugging.

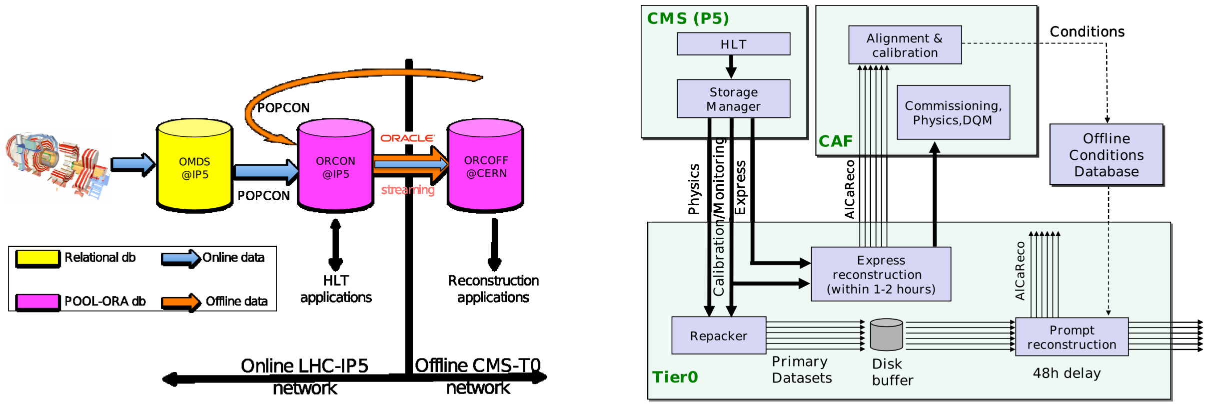

FIGURE 1. Left: Conditions databases infrastructure. Right: The CMS alignment and calibration workflow.

Prompt calibration loop

The prompt calibration loop is a special case of the regular calibration workflow. The goal of prompt calibration is to derive a set of alignment and calibration conditions before the bulk of the data is reconstructed. Subsequently, these conditions ensure precise measurements during the first reconstruction of the full physics stream (prompt reconstruction). In order to derive those prompt conditions the contents of the express and calibration stream are reconstructed and AlCaReco skims are created (express reconstruction). Those are used to carry out a subset of the alignment and calibration tasks which can be executed in about 24 h. During this process, the bulk of the data in the physics stream is cached on disk at the Tier-0.

TESTS OF THE ALIGNMENT AND CALIBRATION FRAMEWORK

The alignment and calibration framework has been tested using both simulated collision data and real cosmic ray data. One large test using simulated collision data has taken place during the computing, software and analysis challenge in 2008. Many alignment and calibration tasks have been successfully executed. Also the interdependence as well as time taken to perform the tasks has been studied [5].

Cosmic ray data taken in 2008

Additional tests were performed using cosmic ray data taken in 2008. Over the course of 23 days, 270 million cosmic ray triggered events were collected with the central magnetic field at its nominal value of 3.8 T [6]. An example of a cosmic ray muon passing through the inner tracker is shown on the left of Fig. 2.
A great variety of alignment and calibration tasks have been successfully carried out using these data. One example is the track based alignment of the inner tracking system [7]. The conditions obtained in this exercise are the positions and orientations of the 16588 detector modules. After the alignment, the most sensitive coordinate shows a precision of 3 to 4 (3 to 14) microns RMS in the barrel (end cap). This leads to significant improvements of the track fit quality (Figure 2 on the right) and thus improved measurement of the track parameters.

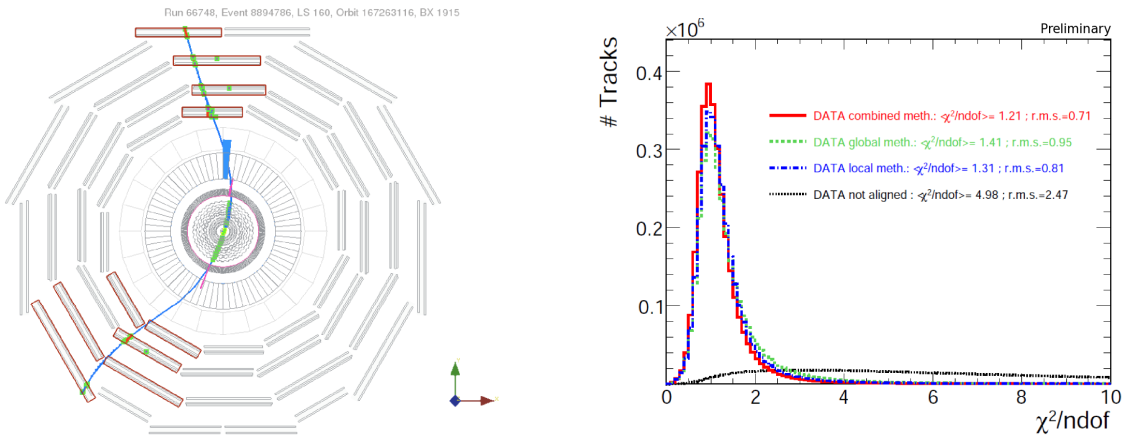

FIGURE 2. Left: Example of a reconstructed cosmic ray muon track. Right: Distributions of the χ^2/ndf of the tracks passing through the inner tracker before alignment (dotted line) and after alignment (solid line). (Details are discussed in [7])

Further Exercises toward LHC start-up in 2009

In 2009 again cosmic ray data were taken and the alignment and calibration workflow was executed. For the first time the prompt calibration loop was exercised successfully for three tasks all of which obtained and validated conditions in time for prompt reconstruction. The results of the alignment and calibration tasks obtained using cosmic ray data will be used as a starting point for the reconstruction of LHC collision data.

CONCLUSION

The alignment and calibration of the detectors is a challenging task at all LHC experiments. CMS has developed and commissioned a powerful framework to meet this challenge. This includes the use of reduced and dedicated data formats for both data transport (AlCaRaw) and usage in alignment and calibration algorithms (AlCaReco). The production of the AlCaReco skims is automated and takes place at the Tier-0. Offline alignment and calibration tasks are carried out at the CERN Analysis Facility which is a versatile batch farm enabling fast task execution and validation. Alignment and calibration conditions are stored in a distributed database scheme which ensures independence of CMS data acquisition from off-site networks while providing reliable access to the conditions in offline network.

A subset of the alignment and calibration tasks is carried out before all of the physics data are reconstructed. Thus, already analysis of the first (prompt) reconstruction of the full data sample will benefit from updated alignment and calibration constants.

The framework has been tested extensively on simulated and real cosmic ray data. The experience and results gained in those exercises will be applied during the first period of collision data taking.

REFERENCES

1. CMS Collaboration, *JINST* **3 S08 004** (2008).
2. I. Bird et. al., *LHC Computing Grid Technical Design Report*, CERN, 2005.
3. M. de Gruttola et. al. (For the CMS Collaboration), “First Experience in Operating the Population of the Condition Database for the CMS Experiment,” in *Proceedings of CHEP*, JPCS, 2009.
4. D. Hufnagel (For the CMS Collaboration), “First Few Days in the Life of CMS Events,” in *Proceedings of ICCMSE*, AIP, 2009.
5. D. Futyan, R. Mankel and C. Paus (For the CMS Collaboration), “The CMS Computing, Software and Analysis Challenge,” in *Proceedings of CHEP*, JPCS, 2009.
6. CMS Collaboration, *Commissioning of the CMS Experiment and the Cosmic Run at Four Tesla*, CMS-CFT-09-008. To be submitted to JINST, 2009.
7. CMS Collaboration, *submitted to JINST* (2009), `physics.ins-det/0910.2505`.