PROCEEDINGS OF SCIENCE

Data Preparation for the CMS Detector at 8TeV at the LHC

Jean-Roch Vlimant*
On behalf of the CMS Collaboration
European Organization for Nuclear Research, CERN
E-mail: vlimant@cern.ch

The CMS detector, currently taking data at the LHC in Geneva, is a very complex apparatus composed of more than 70 million acquisition channels. Fast and efficient methods for the calibration and the alignment of the detector are a key asset to exploit its full physics potential. Moreover, a reliable infrastructure for the monitoring of the data quality and for their validation are instrumental to ensure timely preparation of results for conferences and publications.

The CMS experiment has set up a powerful framework in order to cope with all these requirements and in 2012 it had to consolidate and optimize all the workflows to withstand the higher luminosity and energy delivered by the LHC machine. The reconstruction algorithms have been optimized for the higher occupancies without compromising the physics performance. A Monte-Carlo production with a statistic comparable to the collision data has been prepared and fully validated.

36th International Conference on High Energy Physics,
July 4-11, 2012
Melbourne, Australia

*Speaker.

PoS(ICHEP2012)535

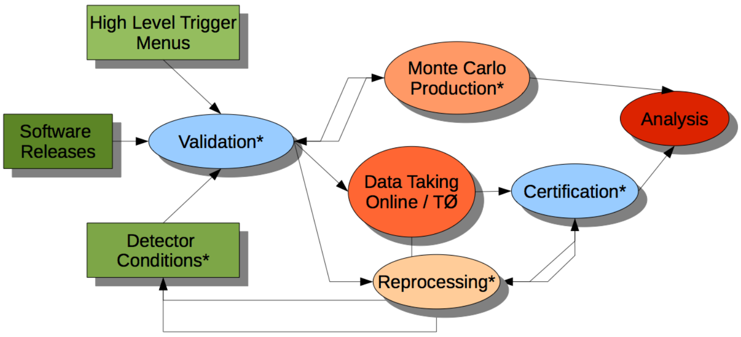

Figure 1: Overview of the CMS validation and certification operation. The items labelled with a * are part of a group created from past experience in operation, dedicated to Physics Performance and Dataset. Certification is the process of creation of a mask of bad data with a granularity of about one minute of data taking.

1. Introduction

The Compact Muon Solenoid (CMS) [1] is a general purpose particle physics detector installed at the Large Hadron Collider (LHC) [2] facility. CMS is a complex apparatus build around a large super-conducting solenoid, creating a 3.8T magnetic field, the outside flux of which is funneled back through a massive return yoke instrumented with muon detector chambers. The electromagnetic and hadronic calorimeters are located inside the solenoid, together with the charged particle tracking device, so as to limit the superfluous loss of energy in the solenoid material. This leads to a very good energy resolution and energy flow identification. Operating such a device requires very fast event selection, prompt and precise calibration, just-in-time event reconstruction and timely data availability, together with production of a large amount of simulated events. We address how such operation has been prepared for the year 2012.

2. Overview of CMS Offline Operation

There are several key ingredients (figure 1) to the efficient delivery of good physics results with CMS data that need to undergo validation scrutiny before being put in production. The High Level Trigger (HLT), the reconstruction software and the detector calibrations are the ones we are reporting on in this document. In addition to technical verification on software and database content, the validation consists in verifying that the physics deliverable of data taking and event reconstruction are kept at their best performance from an analysis point of view.

To mimic the flow from ideas for more accurate data to delivery to analysis, we are going to report on the validation strategy, then successively on the preparation of software, HLT and calibration for the 8TeV data taking at the LHC. We will continue with description of changes in computing infrastructure and data certification, last steps before delivery of data to the analyzers.

3. Data & Monte-Carlo Validation Organisation

The diversity and complexity of the event triggering, event reconstruction and detector calibration requires a variety of inspections of the deliverable of event processing to analysis. Given the large amount of verifications that needs to be performed, on a bi-weekly frequency to a large extent, a maximum amount of automatization is required, and good communication needs to be enforced. The checks are performed on a variety of simulated events and reprocessed data samples [3], covering a wide range of event type and acquisition eras. Such automation, centralized communication and book-keeping have been put in place [4] to simplify and expedite the collection of reports (figure 2). From the organized production of test samples, both simulated and data events, the Data Quality Monitoring (DQM) infrastructure [5] allows for production of automated comparison of detector, object and physics oriented distributions for a strong core of dedicated collaborators to scrutinize.

This established set of procedures is in constant consolidation to further scrutinize under-covered areas and will be the baseline validation for preparation of data taking after the LHC long shutdown in 2013.

4. 2012 LHC Collisions Environment

In order for the LHC to provide higher instantaneous luminosity, the number of parasitic soft proton-proton interaction per bunch crossing has increased over the years of operations (figure 3). Going from an average 2 additional interaction in 2010, this number was expected to be about 20 in 2012.

Software Development for 2012 LHC Collisions

Two stages of improvement of the software (figure 3) were performed in the fall of 2011 and in the spring of 2012 to cope with increasing number of pile-up and allow for the reconstruction of the events acquired in the CMS detector . A selected, small amounts of the events are reconstructed with no delay in order to derive calibrations. The reconstruction (prompt reconstruction) [6] is started within a 48 hours delay from acquisition so as to incorporate these prompt calibrations (prompt calibration loop). Event processing has to happen just-in-time, within allowed computing resources so as to not create back-log of unprocessed data. The improvements mostly came from

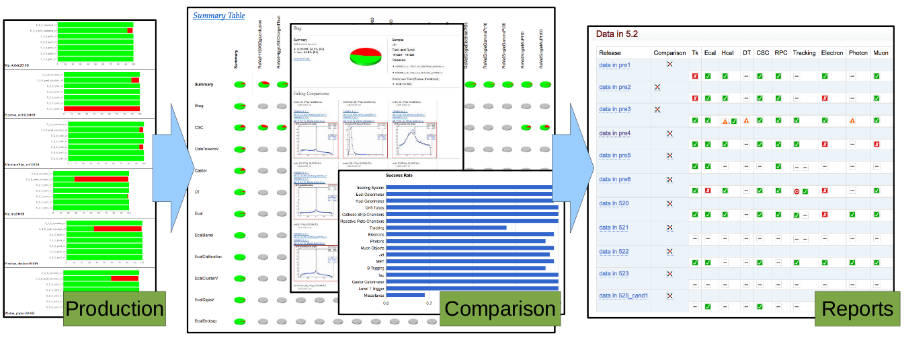

Figure 2: Typical workflow for validation. Software release in this particular case, but it applies to HLT and calibrations. Centralized productions are used for comprehensive and automatic comparisons for validators to scrutinize and report via a book keeping database.

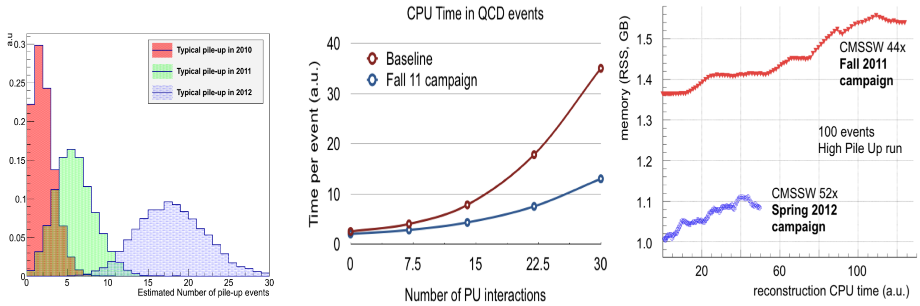

Figure 3: From left to right. Typical number of proton-proton interactions (PU) per bunch crossing in years of LHC operation. Timing of the event reconstruction as a function of number of PU, in the fall of 2011 compared to previously, measure on simulated events. Memory usage versus timing of the event reconstruction in the spring 2012 compared to the status in the fall 2011, measured from data of very high PU.

optimized algorithms for charged particle reconstruction (track reconstruction) [7], while many other smaller relative improvements added up to substantial contribution (figure 3). The physics validation has been performed throughout the integration of these improvements, in order to guarantee the delivery of good analysis with 2012 CMS data.

High Level Trigger Preparation

In CMS, the event acquisition triggering has a Level 1 (L1) hardware trigger [8], and the subsequence HLT [9] is fully based on software. Dedicated software improvements in addition to those reported in the previous section, have been achieved [10] to reduce the timing of HLT event reconstruction and selection. Furthermore, the improved resolution of the algorithms and specific pile-up subtractions allowed for reduced background rates. However, there is an expected inevitable increase of HLT processing time due to intrinsic filtered event reconstruction in HLT. The computing farm for HLT has consequently been upgraded with 50% more CPU available. The effective limitation on the HLT rate (about 300Hz [10]) is set by the amount of event which can be fully reconstructed after data acquisition. Consequently, and in order to increase the amount of data taken by CMS, twice to eight times the same amount of data has been acquired [10] and set aside for delayed reconstruction during the LHC shutdown of 2013.

Alignment and Calibration for Data Taking and Simulation

There are three types of calibration [11] for the data taking and event reconstruction in CMS. Calibrations required prior to collision data taking such as alignment; in addition to being recomputed for reprocessing of the data, using collision data, the initial aligment is computed using muons from atmospheric cosmic showers passing through the detector.

Conditions that are computed using prompt calibration loop, within 48 hours and used in the prompt reconstruction such as bad channel statuses, beam spot positioning and electromagnetic crystal calorimeter calibration for transparency loss. Improvement in condition database, monitoring and procedures allowed for simplified detector condition operation, and a large fraction of good data delivered to analysis.

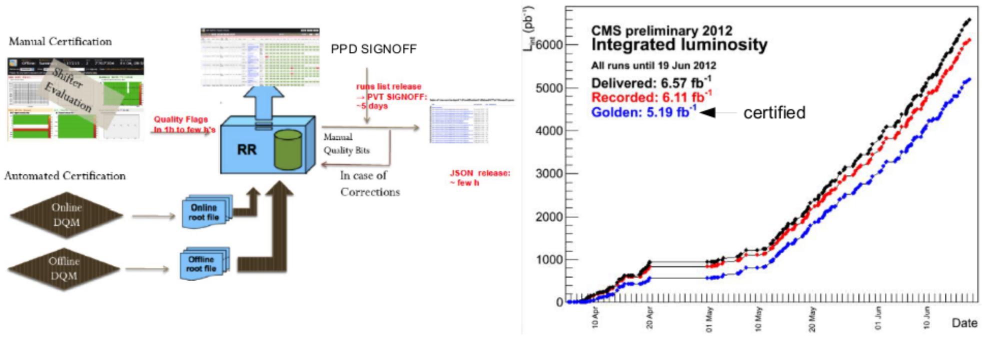

Figure 4: Left: workflow of the data certification in CMS. Intervention from shifters are collected in the Run Registery (RR) database and aggregated into a lumisection masking. Right: amount of integrated luminosity delivered, recorded and certified as a function of time.

Eventually, for campaigns of data reprocessing, ever improved calibration and alignments are computed using the data available. It should be noted that the historical CMS announcement at the ICHEP 2012 conference were made from dataset calibrated within the 48 hour delay of prompt reconstruction.

5. Computing Infrastructure Upgrade for Increased Pile-Up

The effective rate of data acquisition being limited to even reconstruction timing, which stayed roughly constant over the years, the number of events to be handled by the computing infrastructure has been roughly constant. However, with increased pile-up the event size grows non-linearly and therefore the size of the data to be handled, both for data and simulation had been predicted to be much larger than in the previous years. Computing infrastructure and organization were adapted accordingly [12]. A major new framework of data processing system has been deployed in the fall of 2011, commissioned and put in production for the 2012 data operation. With improved operation, traceability and reliability of event simulation and reprocessing during the year 2012, computing went over expectations.

With the change of energy in the center of mass from 7 TeV to 8 TeV in 2012, the totality of the simulated events (4 Billion events) had to be redone, from generation to reconstruction. In addition, as shown in figure 3 with increase pile-up, and to ensure statistical significance of the MC samples through event reweighing, the pile-up simulation had to be adjusted to match the expectation. Event simulation and data reprocessing were scheduled adequately to match major physics conference deadlines through improved coordination between analysis and operation.

Lessons learned from the improved operation during the year 2012 already spawned new projects for ever improved productivity and physics delivery of the CMS Collaboration.

6. Data Certification

In preparation for the 2012 data taking, the certification procedure (figure 4) [13] has been improved in how much of the computing memory it was taking in event processing. The consecutive runs of CMS data acquisitions, have been certified on a weekly basis for prompt analysis of the data. Chunks of data recovered with event reprocessing were certified within 2 weeks, in a timely manner. For ICHEP 2012 in early July, the whole 2012 data set (up to June 19, when LHC went into a scheduled technical stop) went through prompt reconstruction and certification and were included in most analysis. 85 % of the recorded luminosity was certified usable.

7. Conclusions

The framework of validation of new conditions, new HLT menus and algorithm and new reconstruction software has been consolidated from past experience for faster delivery of even more refined data to analysis. Performance improvements from many different areas of the data processing in CMS were put together to sustain the 2012 LHC beam condition, without compromising the physics analysis. Detector alignments and calibrations were scheduled, validated and delivered in a timely manner to ensure the best quality of CMS data at ICHEP 2012. Improved computing resource management system has been commissioned for event simulation and data reprocessing in 2012 and lead to increased productivity in the delivery of datasets for this conference. The preparation, procedure and experience gained with 2012 operation will live on through the LHC long shutdown towards operation of the upgraded CMS detector in 2014.

References

[1] R. Adolphi et al. [CMS Collaboration], The CMS experiment at the CERN LHC, JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004

[2] Lyndon Evans and Philip Bryant, LHC Machine, JINST 3 (2008) S08001, doi:10.1088/1748-0221/3/08/S08001

[3] Oliver Gutsche, Validation of software releases for CMS, 2010 J. Phys.: Conf. Ser. 219 042040

[4] Danilo Piparo, Automated quality monitoring and validation of the CMS reconstruction software, 2012 J. Phys.: Conf. Ser. 368 012008

[5] L Tuura et al, CMS data quality monitoring web service, 2010 J. Phys.: Conf. Ser. 219 072055

[6] David Lange, The CMS Reconstruction Software, CMS CR-2011/002, CHEP 2010

[7] Giacomo Sguazzoni, CMS reconstruction improvements for the tracking in large pile-up events, to be published on Journal of Physics: Conference Series. Prepared for CHEP 2012

[8] S. Dasu et al. [CMS Collaboration], CERN-LHCC-2000-038

[9] P. Sphicas, (ed) [CMS Collaboration], CERN-LHCC-2002-026

[10] Stephanie Beauceron, The CMS High Level Trigger, CMS CR-2012/355, ICHEP 2012

[11] Gianluca Cerminara, Alignment and Calibration of the CMS Detector, PoS EPS-HEP2011 (2011) 186

[12] Stuart Wakefield, The CMS workload management system, CMS CR-2012/105, CHEP 2012

[13] Valdas Rapsevicius, CMS Run Registry: Data Certification Bookkeeping and Publication System, 2011 J. Phys.: Conf. Ser. 331 042038