



Doctoral Thesis

CMS grid computing and monitoring of CMS data quality using selected data samples

Author(s):

Chen, Zhiling

Publication Date:

2011

Permanent Link:

<https://doi.org/10.3929/ethz-a-006444689> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Dissertation ETH NO. 19429

CMS Grid Computing
and
Monitoring of CMS Data Quality
Using Selected Data Samples

A dissertation submitted to
ETH ZURICH
for the degree of
DOCTOR OF SCIENCES

presented by

Zhiling Chen

M.Sc. Graduate School of the Chinese Academy of Sciences

born on July 31th, 1978

citizen of China

accepted on the recommendation of
Prof. Dr. Felicitas Pauss (examiner),
Prof. Dr. Günther Dissertori (co-examiner)

Abstract

The Large Hadron Collider (LHC) at CERN, the European Center for Particle Physics in Geneva, is a proton-proton collider, designed to operate at a center-of-mass energy of 14 TeV at a nominal luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The main physics goals of the LHC experiments are the search for the Standard Model Higgs boson and new physics phenomena beyond the Standard Model. The design and construction of the Compact Muon Solenoid (CMS) experiment at the LHC had to meet unprecedented challenges both for the detector operation as well as for the data handling. Due to the high event rate and large event size, the LHC experiments generate large amount of data of about 15 petabyte (10^{15} bytes) per year at the design luminosity, which thousands of scientists at hundreds of research institutes and universities around the world access and analyse. In addition, detailed Monte Carlo simulations of various physics processes also require large-scale computing power and huge amount of mass storage. To meet these requirements, a novel globally distributed model for data storage and CPU power was chosen: the Worldwide LHC Computing Grid (WLCG). The WLCG collaboration in Switzerland provides computing infrastructure and resources to physicists from Swiss institutions involved in the LHC experiments as well as to the experimental collaborations, by operating a high-performance Tier-2 center at CSCS in Manno and the Swiss CMS Tier-3 center at the Paul Scherrer Institute (PSI) in Villigen.

This thesis reports on my work for the Swiss Tier-2 and the CMS Tier-3 centers. The two facilities passed several benchmarks, were upgraded continuously over the past years and show excellent operation performance since the start-up of the LHC on 30 March 2010, providing proton-proton collisions at a center-of-mass energy of 7 TeV.

In the second part of this thesis studies of the detector performance and data quality monitoring are described, which are key issues for the physics output, especially at the start of data taking. A data sample of $Z \rightarrow e^+e^-$ recorded up to September 2010 is selected to study the performance of the CMS detector and to monitor the data quality. Electrons and positrons are reconstructed and identified in the electromagnetic calorimeter requiring a matching track in the tracking system. The measured invariant mass distribution obtained from the selected electron-positron pairs show a clear Z mass peak with very little background. This result is in good agreement with the Monte Carlo predictions and illustrates the good data quality at the start of the CMS operation.

Zusammenfassung

Der Large Hadron Collider (LHC) am Europäischen Zentrum für Teilchenphysik CERN in Genf ist ein Beschleuniger der für Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 14 TeV und einer Design Luminosität von $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ gebaut wurde. Die wichtigsten wissenschaftlichen Ziele der Experimente am LHC sind die Suche nach dem Standardmodell Higgs Boson und neuen physikalischen Phänomenen jenseits des Standardmodells. Design und Konstruktion des Compact Muon Solenoid (CMS) Experiments am LHC war eine grosse Herausforderung sowohl für den Betrieb des Detektors als auch für die Verarbeitung der aufgezeichneten Daten. Aufgrund der hohen Kollisionsrate und der grossen Menge der pro Kollision anfallenden Daten werden die LHC Experimente noch nie dagewesene Datenvolumen von etwa 15 Petabyte (10^{15} Bytes) pro Jahr erzeugen. Tausende von Wissenschaftlern an mehreren hundert Universitäten weltweit werden auf diese Daten zugreifen um sie zu analysieren. Zusätzlich benötigt auch die detaillierte Simulation der verschiedenen physikalischen Prozesse immense Rechen- und Speicherkapazitäten. Um diesen Anforderungen gerecht zu werden wurden die für die Analyse der Daten benötigten Rechen- und Speicherkapazitäten auf der ganzen Welt verteilt und als 'Worldwide LHC Computing Grid (WLCG)' vernetzt. Die WLCG-Gruppe der Schweiz stellt den Physikern an Schweizer Instituten sowie den LHC Kollaborationen Infrastruktur und Rechenkapazitäten für die Verarbeitung der LHC Daten zur Verfügung. Ein wichtiger Teil dieser Infrastruktur ist das der Hochleistungs-Tier-2-Cluster am Schweizer Supercomputing-Zentrum (CSCS) in Manno und das Schweizer CMS Tier-3 Zentrum am Paul Scherrer Institut (PSI) in Villigen.

Die vorliegende Doktorarbeit beschreibt meine Arbeit für den Schweizer Tier-2 und den Schweizer CMS Tier-3 Computer-Cluster. Diese zwei Tier Zentren bestanden mehrere Leistungstests und wurden kontinuierlich in den vergangenen Jahren erweitert. Sie bewährten sich bestens seit dem Start von Proton-Proton Kollisionen bei einer Schwerpunktsenergie von 7 TeV am 30. März 2010.

Ein weiterer Teil dieser Doktorarbeit beschreibt die Untersuchungen zur Leistungsfähigkeit des CMS Detektors und Qualität der aufgezeichneten Daten. Solche Untersuchungen sind besonders in den ersten Monaten der Datennahme von zentraler Wichtigkeit. Zu diesem Zweck wurden Ereignisse aus den Daten welche bis zum September 2010 aufgenommen wurden, analysiert, bei denen ein Z-Boson in ein Elektron und ein Positron zerfällt ($Z \rightarrow e^+e^-$). Elektronen und Positronen werden durch ein charakteristisches Signal im elektromagnetischen Kalorimeter identifiziert, wobei eine passende Spur in der zentralen Spurkammer gefordert wird. Die gemessene Verteilung der invarianten Masse dieser Ereignisse zeigt ein klares Z Signal mit sehr kleinem Untergrund. Dieses Ergebnis ist in guter Übereinstimmung mit der Vorhersage von Monte Carlo Simulationen und illustriert die sehr gute Qualität der aufgezeichneten Daten.