

Content Delivery Network solutions for the CMS experiment: the evolution towards HL-LHC

Friday, March 29, 2024 11:30 AM (20 minutes)

The Large Hadron Collider at CERN in Geneva is poised for a transformative upgrade, preparing to enhance both its accelerator and particle detectors. This strategic initiative is driven by the tenfold increase in proton-proton collisions anticipated for the forthcoming high-luminosity phase scheduled to start by 2029. The vital role played by the underlying computational infrastructure, the World-Wide LHC Computing Grid, in processing the data generated during these collisions underlines the need for its expansion and adaptation to meet the demands of the new accelerator phase. The provision of these computational resources by the worldwide community remains essential, all within a constant budgetary framework. While technological advancements offer some relief for the expected increase, numerous research and development projects are underway. Their aim is to bring future resources to manageable levels and provide cost-effective solutions to effectively handle the expanding volume of generated data. In the quest for optimised data access and resource utilisation, the LHC community is actively investigating Content Delivery Network (CDN) techniques. These techniques serve as a mechanism for the cost-effective deployment of lightweight storage systems that support both, traditional and opportunistic compute resources. Furthermore, they aim to enhance the performance of executing tasks by facilitating the efficient reading of input data via caching content near the end user. A comprehensive study is presented to assess the benefits of implementing data cache solutions for the Compact Muon Solenoid (CMS) experiment. This in-depth examination serves as a use-case study specifically conducted for the Spanish compute facilities, playing a crucial role in supporting CMS activities. Data access patterns and popularity studies suggest that user analysis tasks benefit the most from CDN techniques. Consequently, a data cache has been introduced in the region to acquire a deeper understanding of these effects. In this contribution, the details of the implementation of a data cache system in the PIC Tier-1 compute facility is presented. It includes insights into the developed monitoring tools and discusses the positive impact on CPU usage for analysis tasks executed in the region. The study is augmented by simulations of data caches, with the objective of discerning the most optimal requirements in both size and network connectivity for a data cache serving the Spanish region. Additionally, the study delves into the cost benefits associated with deploying such a solution in a production environment. Furthermore, it investigates the potential impact of incorporating this solution into other regions of the CMS computing infrastructure.

Primary author: FLIX, Josep (PIC / CIEMAT)

Co-authors: PÉREZ DENGRA, Carles (PIC/CIEMAT); Dr SIKORA, Anna (UAB); SERRANO SIERRA, Paula (UAB)

Presenter: FLIX, Josep (PIC / CIEMAT)

Session Classification: Physics & Engineering Application

Track Classification: Track 1: Physics and Engineering Applications