Contribution ID: **760** Contribution code: **contribution ID 760** Type: **Poster**

# New storage and data access solution for CMS experiment in Spain towards HL-LHC era

The Large Hadron Collider (LHC) will enter a new era for data acquisition by 2026 within the High-Luminosity Large Hadron Collider (HL-LHC) program, where the LHC will increase the proton-proton collisions up to unprecedented levels. This increase will imply a factor 10 in terms of luminosity as compared to the current values, having an impact in the way the experimental data is stored and analyzed. Currently, more than 1 Exabyte of simulated and real data have been produced by the LHC experiments, being stored on disk and tape and processed in a worldwide distributed computing infrastructure comprising 170 centers in 35 countries, known as WLCG (World-wide LHC Computing Grid), with sites categorized by their size and functions, namely Tier-0 (at CERN), 13 Tier-1s and 150 Tier-2s. The high estimates of data produced by LHC experiments in the HL-LHC period and the computation required to process this data will not fit the extrapolated available compute budget. This situation led scientists to search for new mechanisms in order to alleviate these increases so they can be affordable by the community. Facing these challenges, the LHC experiments have launched an extensive Research and Development program to reduce the overall cost of storage, in terms of hardware and operations. Also, enable more effective use of storage resources, and efficiently deliver data at scale to large, remote and heterogeneous computing resources that are expected to be integrated into a network-centric and global infrastructure (Data Lake). This program is developed with other data-intensive sciences that have similar computational challenges, since they all use the same computational clusters available worldwide.

At the LHC there are several experiments associated with their respective detectors that take real experimental data that are stored together with the simulated data on the WLCG sites.
There are four big detectors at the LHC, namely: ATLAS, CMS, LHCb and ALICE. All of these experiments are working together in these R&D activities.
One of the experiments is the Compact Muon Solenoid (CMS), responsible for the discovery of the Higgs boson together with the LHC ATLAS experiment.
The evolution of the Data Storage System in CMS, among the other LHC experiments, is crucial to be addressed by the WLCG community due the huge loads on data acquisition to be stored in the High Luminosity phase of the LHC.

One of the proposed solutions is the so-called WLCG-Data Lake model. This model presents a way to optimize the cost operations and reduce the hardware deployed by consolidating the storage resources in a few sites. The sites would be also chosen for concentrating their efforts and investments specializing on running large computing farms, storage systems or both. Among the benefits of the model stand out is that storage systems can be deployed as a distributed service accessed by remote facilities as a single-entry endpoint. Deploying resources in this manner a certain number of sites could share and offer resources as data federations, a collection of disparate storage resources that are transparently accessible across a wide area via a common namespace, taking into account geographic distance and latency.

Due to the nature of execution tasks to process and analyze data of CMS experiment, that involve the access and reading of data in the order of terabytes, new data management strategies are necessary to fluently explode WLCG resources. PIC and CIEMAT CMS sites have spent the last years investing efforts in different areas of remote data access and processing with the aim to contribute to these R&D activities for the HL-LHC program, and also to explore the possibility of resource consolidation in the region.

Following the trend in WLCG to federate computing resources, in PIC and CIEMAT we demonstrated for both centres, the Spanish Tier-1 and Tier-2 sites for CMS, that resource sharing can be achieved with a single approach in the context of a WLCG single VO (Virtual Organisation). That work showed how compute nodes can be shared between two sites, dynamically increasing the capacity of one site with idle execution of slots

from the remote side. This was performed by making use of the existing CMS xrootd federation infrastructure and the common CE/batch technology in both sites.
These studies are critical to explore the future scenarios, where storage resources would be concentrated in a reduced number of sites.

Another aspect to consider besides resources sharing is the capability of the existing Grid centers to expand and integrate compute resources, either from Cloud resources or nearby HPC centers. In this sense, PIC has been making efforts to transparently integrate commercial (Amazon Web Service 'AWS'and Google Cloud Platform 'GCP') and private Cloud resources (through HNSciCloud and ESCAPE projects), as well as resources from the Barcelona Supercomputing Center (BSC). A collaboration agreement with the BSC has been signed that allows PIC and CIEMAT to exploit HPC resources to execute jobs for all LHC experiments in Spain. Usually, these facilities do not have direct access to the network from their computing nodes, nor sufficient storage systems. The case of BSC is particularly challenging due its restrictive network setup. One of the solutions developed to integrate the resource in production consisted in using pre-placed Singularity containers with application software in the HPC facility, sharing the file-system with conditions data files. Also, Workload Management System HTCondor was extended to relay the communications between running pilot jobs and HTCondor daemons through the HPC shared file-system.

In both PIC and CIEMAT sites we have deployed cache solutions in order to study the benefits that these systems will bring to the execution tasks. To ensure the success of these tasks a content delivery caching network that minimizes the latency is needed in order to bring data to the sites attached to the Data lake. For this reason, a step forward was taken in the research deploying the first instance of XCache in computing nodes at PIC and CIEMAT. XCache is the preferred cache service for the CMS community because it is based on XrootD, the main protocol used in WLCG for scientific data access. This new technology could also make it easier to integrate opportunistic resources, such as Cloud services or HPC. In the case of the BSC, in which the LHC job submission is in the development phase, this new technology would make it possible to bring the necessary data for its execution, exercising the function of a storage system. Following this goal, the use and efficiency of caches constitutes a cornerstone in order to consolidate the CMS data region in the Spanish region.

The main objective is to research and develop new solutions to the current CMS data storage model in Spain that serves data to the computing nodes in Tier-1 of the PIC and Tier-2 of CIEMAT based on caches. Due to the increase in data collection for the HL-LHC, we explore new solutions and architectures for access to data and storage. The new architecture should allow to maintain the current CPU efficiency at the sites or improve it, as well as reduce the latency of access to data and the deployment of storage hardware. It is also intended to find a new architecture that allows us to run jobs on sites without having a local storage system.

In this contribution we propose a model of access and storage of CMS data in Spain based on an architecture of computing nodes with XCache caching systems in PIC and CIEMAT. This system can be deployed only in PIC Tier-1 or in CIEMAT Tier-2 as well.

We present dedicated studies to optimize the use, size and configuration of the XCaches, based on CMS Data popularity and access at both PIC and CIEMAT sites. The effects on the job efficiencies, and the implications on the use of the network, have been as well addressed. In this work we also explore solutions to optimize the use, size and configuration of caches. To do this, we explore issues such as the popularity of data in CMS in the region and optimal cache settings. Finally, we analyze how the addition of this element to the architecture affects the efficiency of jobs, latency and its possible implications in the use of the network.
This work is focused on the activities that are carried out within the CMS experiment, and in particular in test beds that are being carried out in the Spanish region, in PIC and CIEMAT compute facilities.

**The authors acknowledge support provided by Spanish funding agency AEI, grant PID2019-110942RB-C21, the Ministerio de Economía, Industria y Competitividad MINECO-SPAIN under contract TIN2017-84553-C2-1-R and by the Generalitat de Catalunya GenCat-DIUiE (GRR) with the project 2017-SGR-313.**

## Significance

We present a solution for the storage and access to data from the CMS experiment in the Spanish region through a model that would save on the deployment of storage systems in smaller sites. Also, it would allow HPC centers to run CMS jobs without a persistent storage system deployed within.

## References

[1] WLCG project, Consulted on 6th Jun of 2016: http://wlcg.web.cern.ch/.

[2] J. Albrecht, et al, “A Roadmap for HEP Software and Computing RD for the 2020s”, Computing and Software for Big Science volume 3, Article number: 7 (2019) https://doi.org/10.1007/s41781-018-0018-8

[3] X. Espinal, et al, ”The Quest to solve the HL-LHC data access puzzle. The first year of the DOMA ACCESS Working Group”, International Conference on Computing in High Energy and Nuclear Physics (CHEP), Adelaide, Australia, 4-8 november 2019, viewed 5 november 2019.

[4] C. Acosta-Silva, A. Delgado Peris, J. Flix, J. M. Guerrero, J. M. Hernández, A. Pérez-Calero Yzquierdo, F. J. Rodriguez Calonge, J. Gómez del Pulgar Ruano A 2019 “Lightweight site federation for CMS support”, International Conference on Computing in High Energy and Nuclear Physics (CHEP), Adelaide, Australia, 4-8 november 2019, viewed 3rd June of 2021.

[5] Delgado Peris, A., Flix Molina, J., Hernández J., Pérez-Calero Yzquierdo, A., Pérez Dengra, C., Planas, E., Rodríguez Calonge, J., Sikora, A 2019 “CMS data access and usage studies at PIC Tier-1 and CIEMAT Tier-2”, EPJ Web Conf., 245 (2020) 04028.

## Speaker time zone

Compatible with Europe

**Author:** PEREZ DENGRA, Carlos (PIC-CIEMAT)

**Co-authors:** Mr FLIX MOLINA, José (PIC-CIEMAT); Mrs SIKORA, Anna (UAB (Autonomous University of Barcelona))

**Presenter:** PEREZ DENGRA, Carlos (PIC-CIEMAT)

**Session Classification:** Posters: Windmill

**Track Classification:** Track 1: Computing Technology for Physics Research