



Contribution ID: 72

Type: Oral

CMS data access and usage studies at PIC Tier-1 and CIEMAT Tier-2

Tuesday 5 November 2019 14:15 (15 minutes)

Computing needs projections for the HL-LHC era (2026+), following the current computing models, indicate that much larger resource increases would be required than those that technology evolution at a constant budget could bring. Since worldwide budget for computing is not expected to increase, many research activities have emerged to improve the performance of the LHC processing software applications, as well as to propose more efficient deployment scenarios and techniques which might alleviate the increase of expected resources for the HL-LHC. The massively increasing amounts of data to be processed leads to enormous challenges for HEP storage systems, networks and the data distribution to end-users. This is particularly important in scenarios in which the LHC data would be distributed from sufficiently small numbers of centers holding the experiment's data. Enabling data locality via local caches on sites seems a very promising approach to hide transfer latencies while reducing the deployed storage space and number of replicas elsewhere. However, this highly depends on the workflow I/O characteristics and available network across sites. A crucial assessment is to study how the experiments are accessing and using the storage services deployed in sites in WLCG, to properly evaluate and simulate the benefits for several of the new emerging proposals within WLCG/HSF. In order to evaluate access and usage of storage, this contribution shows data access and popularity studies for the CMS Workflows executed in the Spanish Tier-1 (PIC) and Tier-2 (CIEMAT) sites supporting CMS activities, based on local and experiment monitoring data spanning more than one year. Simulations of data caches for end-user analysis data, as well as potential areas for storage savings will be reviewed.

Consider for promotion

Yes

Author: FLIX MOLINA, Jose (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas)

Presenter: FLIX MOLINA, Jose (Centro de Investigaciones Energéticas Medioambientales y Tecnológicas)

Session Classification: Track 4 –Data Organisation, Management and Access

Track Classification: Track 4 –Data Organisation, Management and Access