



Contribution ID: 237

Type: **Orale**

Integration of a smart Italian cache federation for CMS

Friday, 7 June 2019 09:10 (20 minutes)

A huge increase of both storage and computing requirements are foreseen for the HL-LHC era (factor ~20 for the storage, and ~30x for CPUs) and new kind of resource are covering a crescent amount of needs (e.g. private or public cloud and HPC facilities). Thus CMS experiment has been pushed towards evaluating an evolution for the optimization of the amount of space that is managed centrally and the CPU efficiency of the jobs that run on “storage-less” resources.

In particular “Tier2” sites layer, for the most part, can be instrumented to read data from a remote source eventually enabling the use of a geographically distributed cache storage based on unmanaged resources, with a consequent reduction of the operational efforts for maintaining managed custodial storages and increasing the flexibility of the system. The cache system will appear as distributed and shared file system populated with the most requested data; in case of missing information data access will fallback to the remote access.

Moreover in a possible future scenario where a data-lake model will be implemented, a protection layer against a central managed storages might be a key factor along with the control on data access latency. The cache storages used for such a layer will be by definition “non-custodial”, thus reducing the overall operational costs.

The objective of this contribution is to present the first integration experience of an INFN federation of cache servers spanning over the CNAF Tier-1, Bari and Legnaro Tier-2s that provided unmanaged storages organized under a common namespace with the use of XCache technology. The results in terms of performances on real CMS experiment workflows will be shown. In addition first studies on CMS metadata regarding the analysis jobs access pattern over Italian Tier2’s will be presented, leading to an estimation of the possible improvements provided by the introduction of the proposed solution at production scale for the Italian Tier2’s.

Furthermore the Proof-of-Concept deployment of a “smart decision service” platform to enable AI based cache operations will be shown. The objective is to enable ML-based management of the Cache content in order to reduce costs in term of hardware but also to optimize the operational costs thanks to the intelligence.

Future plans and evolution toward the direction of building Data Lakes prototypes in the context of the IDDLs and ESCAPE funded projects will be presented.

Primary authors: CIANGOTTINI, Diego (INFN Perugia); TRACOLLI, Mirco (PG)

Co-authors: SPIGA, Daniele (PG); CESINI, Daniele (CNAF); BOCCALI, Tommaso (PI)

Presenter: CIANGOTTINI, Diego (INFN Perugia)

Session Classification: Calcolo negli esperimenti

Track Classification: Modelli di calcolo per gli esperimenti e evoluzione delle infrastrutture