



Contribution ID: 176

Type: **Parallel**

Implementing data placement strategies for the CMS experiment based on a popularity mode

Thursday, 24 May 2012 17:25 (25 minutes)

During the first two years of data taking, the CMS experiment has collected over 20 PetaBytes of data and processed and analyzed it on the distributed, multi-tiered computing infrastructure on the WorldWide LHC Computing Grid. Given the increasing data volume that has to be stored and efficiently analyzed, it is a challenge for several LHC experiments to optimize and automate the data placement strategies in order to fully profit of the available network and storage resources and to facilitate daily computing operations.

Building on previous experience acquired by ATLAS, we have developed the CMS Popularity Service that tracks file accesses and user activity on the grid and will serve as the foundation for the evolution of their data placement. A fully automated, popularity-based site-cleaning agent has been deployed in order to scan Tier2 sites that are reaching their space quota and suggest obsolete, unused data that can be safely deleted without disrupting analysis activity. Future work will be to demonstrate dynamic data placement functionality based on this popularity service and integrate it in the data and workload management systems: as a consequence the pre-placement of data will be minimized and additional replication of hot datasets will be requested automatically.

This paper will give an insight into the development, validation and production process and will analyze how the framework has influenced resource optimization and daily operations in CMS.

Primary authors: SPIGA, Daniele (CERN); Dr GIORDANO, Domenico (CERN); Dr KARAVAKIS, Edward (CERN); BARREIRO MEGINO, Fernando Harald (Universidad Autonoma de Madrid (ES)); Dr GIRONE, Maria (CERN); CINQUILLI, Mattia (Univ. of California San Diego (US)); MAGINI, Nicolo (CERN); MANCINELLI, Valentina (Sezione di Perugia (INFN)-Universita e INFN)

Presenters: Dr GIORDANO, Domenico (CERN); BARREIRO MEGINO, Fernando Harald (Universidad Autonoma de Madrid (ES))

Session Classification: Distributed Processing and Analysis on Grids and Clouds

Track Classification: Distributed Processing and Analysis on Grids and Clouds (track 3)