

Dynamo - The dynamic data management system for the distributed CMS computing system

Wednesday, October 12, 2016 11:30 AM (15 minutes)

The upgraded Dynamic Data Management framework, Dynamo, is designed to manage the majority of the CMS data in an automated fashion. At the moment all CMS Tier-1 and Tier-2 data centers host about 50 PB of official CMS production data which are all managed by this system. There are presently two main pools that Dynamo manages: the Analysis pool for user analysis data, and the Production pool which is used by the production systems to run (re)-reconstruction and produce Monte Carlo simulation and organize dedicated data transfer tests and tape retrieval. The first goal of the Dynamic Data Management system, to facilitate the management of the data distribution, had already been accomplished shortly after its first deployment in 2014. The second goal of optimizing the accessibility of data for the physics analyses has made major progress in the last year. Apart from the historic data popularity we are now also using the information from analysis jobs queued in the global queue to optimize the data replication for faster analysis job processing. This paper describes the architecture of all relevant components and details the experience of the upgraded system and running it over the last half year.

Tertiary Keyword (Optional)

Secondary Keyword (Optional)

Primary Keyword (Mandatory)

Distributed data handling

Primary author: IYAMA, Yutaro (Massachusetts Inst. of Technology (US))

Co-authors: PAUS, Christoph (Massachusetts Inst. of Technology (US)); GONCHAROV, Maxim (Massachusetts Inst. of Technology (US))

Session Classification: Track 4: Data Handling

Track Classification: Track 4: Data Handling