

The archive solution for distributed workflow management agents of the CMS experiment at LHC

Monday 9 July 2018 12:00 (15 minutes)

The CMS experiment at the CERN LHC developed the Workflow Management Archive system to persistently store unstructured framework job report documents produced by distributed workflow management agents. In this talk we present its architecture, implementation, deployment, and integration with the CMS and CERN computing infrastructures, such as central HDFS and Hadoop Spark cluster. The system leverages modern technologies such as a document oriented database and the Hadoop eco-system to provide the necessary flexibility in order to reliably process, store, and aggregate ~1M documents on a daily basis. We will discuss the data transformation, the short and long term storage layers, the query language, along with the aggregation pipeline developed to visualize various performance metrics to assist CMS data operators in assessing the performance of the CMS computing system.

Primary author: KUZNETSOV, Valentin Y (Cornell University (US))

Presenter: KUZNETSOV, Valentin Y (Cornell University (US))

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling