

Software and Experience with Managing Workflows for the Computing Operation of the CMS Experiment

Tuesday, October 11, 2016 11:00 AM (15 minutes)

We present a system deployed in the summer of 2015 for the automatic assignment of production and reprocessing workflows for simulation and detector data in the frame of the Computing Operation of the CMS experiment at the CERN LHC. Processing requests involves a number of steps in the daily operation, including transferring input datasets where relevant and monitoring them, assigning work to computing resources available on the CMS grid, and delivering the output to the Physics groups. Automatization is critical above a certain number of requests to be handled, especially in the view of using more efficiently computing resources and reducing latencies. An effort to automatize the necessary steps for production and reprocessing recently started and a new system to handle workflows has been developed. The state-machine system described consists in a set of modules whose key feature is the automatic placement of input datasets, balancing the load across multiple sites. By reducing the operation overhead, these agents enable the utilization of more than double the amount of resources with robust storage system. Additional functionalities were added after months of successful operation to further balance the load on the computing system using remote read and additional resources. This system contributed to reducing the delivery time of datasets, a crucial aspect to the analysis of CMS data. We report on lessons learned from operation towards increased efficiency in using a largely heterogeneous distributed system of computing, storage and network elements.

Secondary Keyword (Optional)

Distributed workload management

Primary Keyword (Mandatory)

Data processing workflows and frameworks/pipelines

Tertiary Keyword (Optional)

Primary author: Dr VLIMANT, Jean-Roch (California Institute of Technology (US))

Presenter: Dr VLIMANT, Jean-Roch (California Institute of Technology (US))

Session Classification: Track 3: Distributed Computing

Track Classification: Track 3: Distributed Computing