



Contribution ID: 160

Type: Poster

Creating a model for dynamic resource optimization for the CMS Workflow Management System

The CMS experiment hosted at CERN relies every day on the computing resources provided by the Worldwide LHC Computing Grid consortium (WLCG) to process data and produce Monte-Carlo simulated event samples. In such a context, utilizing the heavily distributed system of computing resources granted by the WLCG to process non-local data represents a challenging task. In addition, the CMS Workflow management system does not impose a strict boundary between physics driven and underlying infrastructure driven requirements, as it needs to deal with a plethora of configurations consisting of sets of parameters tied to both sides. This results in a distribution of workflow resource requirements and duration with a large variance, which undermines the predictability of the time and resources needed for delivering events.

In this framework, multiple negative effects that might contribute to the overall system limitations could be expected, such as decreased event throughput and resource pool fragmentation, suboptimal exploitation of the available resources, failed data transmission, and data loss. It is crucial, then, to define a minimal set of parameters that heavily affect the behavior of the system. A non-exhaustive list of such parameters could be data scattering and locality, workflow length and payload efficiency, resource requirements estimates and more.

This contribution shows an analysis of a significant amount of historical data to identify the minimal set of variables governing the duration of a CMS workflow and necessary to predict suboptimal behaviors in terms of resource exploitation. With such input features, it is possible to model the dependency of the workflow lifetime and the event throughput from the available resource pool, paving the way for outlining a dynamic optimization of the resource allocation to maximize these metrics. Such an approach would be beneficial in terms of minimizing the potential negative effects on the utilization of a heavily diverse resource pool with non-local data. This could eventually lead to a maximization of the event throughput while boosting the resource utilization efficiency.

Significance

References

Experiment context, if any

CMS experiment

Author: CMS COLLABORATION

Presenter: CMS COLLABORATION

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research