Contribution ID: **264** Type: **Poster**

CMS Workflow Failures Recovery Panel, Towards AI-assisted Operation

Tuesday 10 July 2018 16:40 (20 minutes)

The central production system of CMS is utilizing the LHC grid and effectively about 200 thousand cores, over about a hundred computing centers worldwide. Such a wide and unique distributed computing system is bound to sustain a certain rate of failures of various types. These are appropriately addressed with site administrators a posteriori. With up to 50 different campaigns ongoing concurrently, the range of diversity of workload is wide and complex, leading to a certain amount of mis-configurations, despite all efforts in request preparation. Most of the 2000 to 4000 datasets produced each week are done so in full automation, and datasets are delivered within an agreed level of completion. Despite effort of reducing the rate of failure, there remains a good fraction of workflows that requires non trivial intervention. This work remains for computing operators to do. We present here a tool, which was developed to facilitate and improve this operation, in a view to reduce delays in delivery. A dense and comprehensive representation of what errors occurred during the processing of a request helps expediting the investigation. Workflows that suffered from similar failures are bundled and presented as such to the operator. A realistically simplified operating panel front-end is connected to a backend automatizing the technical operation required for ease of operation. The framework was built such that it is collecting both the decision and the information available to the operator for taking that decision. It is therefore possible to employ machine learning technique to learn from the operator by training on labelled data. The operator's procedure is automatized further by applying the decisions that are predicted with acceptable confidence. We present this tool that improves operational efficiency and will lead to further development in handling failures in distributed computing resources using machine learning.

Authors: ABERCROMBIE, Daniel Robert (Massachusetts Inst. of Technology (US)); REINSVOLD HALL, Allison (University of Notre Dame (US)); ROZO BERNAL, Paola Katherine (Universidad de los Andes (CO)); VLIMANT, Jean-Roch (California Institute of Technology (US)); NGUYEN, Thong (California Institute of Technology (US)); CONTRERAS CAMPANA, Christian (DESY, Hamburg (Germany)); CREMONESI, Matteo (Fermi National Accelerator Lab. (US))

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

Session Classification: Posters

Track Classification: Track 3 –Distributed computing