



Contribution ID: 146

Type: Oral

Automatic log analysis with NLP for the CMS workflow handling

Tuesday 5 November 2019 14:45 (15 minutes)

The central Monte-Carlo production of the CMS experiment utilizes the WLCG infrastructure and manages daily thousands of tasks, each up to thousands of jobs. The distributed computing system is bound to sustain a certain rate of failures of various types, which are currently handled by computing operators a posteriori. Within the context of computing operations, and operation intelligence, we propose a machine learning technique to learn from the operators with a view to reduce the operational workload and delays. This work is in continuation of CMS work on operation intelligence to try and reach accurate predictions with machine learning. We present an approach to consider the log files of the workflows as regular text to leverage modern techniques from natural language processing (NLP). In general, log files contain a substantial amount of text that is not human language. Therefore, different log parsing approaches are studied in order to map the log files' words to high dimensional vectors. These vectors are then exploited as feature space to train a model that predicts the action that the operator has to take. This approach has the advantage that the information of the log files is extracted automatically and the format of the logs can be arbitrary. In this work the performance of the log file analysis with NLP is presented and compared to previous approaches.

Consider for promotion

Yes

Authors: LAYER, Lukas (Universita e sezione INFN di Napoli (IT)); ABERCROMBIE, Daniel Robert (Massachusetts Inst. of Technology (US)); BAKHSHIANSOHI, Hamed (Deutsches Elektronen-Synchrotron (DE)); SI, Weinan (University of California Riverside (US)); VLIMANT, Jean-Roch (California Institute of Technology (US)); Mr VARGAS HERNANDEZ, Andres (Catholic University of America (US)); AGARWAL, Sharad (University of Wisconsin Madison (US)); ADELMAN-MCCARTHY, Jennifer (Fermi National Accelerator Lab. (Fermilab)-Unknown-Unknown)

Presenter: LAYER, Lukas (Universita e sezione INFN di Napoli (IT))

Session Classification: Track 3 –Middleware and Distributed Computing

Track Classification: Track 3 –Middleware and Distributed Computing