Contribution ID: **354** Type: **Sectional reports**

Search for Anomalies in the Computational Jobs of the ATLAS Experiment with the Application of Visual Analytics

Tuesday 11 September 2018 16:00 (15 minutes)

ATLAS is the largest experiment at the LHC. It generates vast volumes of scientific data accompanied with auxiliary metadata. These metadata represent all stages of data processing and Monte-Carlo simulation, as well as characteristics of computing environment, such as software versions and infrastructure parameters, detector geometry and calibration values. The systems responsible for data and workflow management and metadata archiving in ATLAS are called Rucio, ProdSys2, PanDA and AMI. Terabytes of metadata were accumulated over the many years of systems functioning. These metadata can help physicists carrying out studies to evaluate in advance the duration of their analysis jobs. As all these jobs are executed in a heterogeneous distributed and dynamically changing infrastructure, their duration may vary across computing centers and depends on many factors, like memory per core, system software version and flavour, volumes of input datasets and so on.

Ensuring the uniformity in jobs execution requires searching for anomalies (for example, jobs with too long execution time) and analyzing the reasons of such behavior to predict and avoid the recurrence in future. The analysis should be implemented on the basis of all historical jobs metadata that are too large to be processed and analyzed by standard means. Detailed analysis of the archive can benefit from application of visual analytics methods providing more easy way of navigation within the multiple internal data correlations. Presented research is the starting point in this direction. The slice of ATLAS jobs archive was analyzed visually, demonstrating the most and the less efficient computing sites. Then, the efficient sites will be compared to inefficient to find out parameters affecting jobs execution time or indicating possible time delays. Further work will concentrate on the increasing of the amount of analyzed jobs and the development of the interactive 3-dimensional visual models, facilitating the interpretation of analysis results.

Primary authors: Dr KLIMENTOV, Alexei (Brookhaven National Lab); Ms MARIA, Grigorieva (NRC KI); MILMAN, I.E. (National Research Nuclear University "MEPHI"); TITOV, Mikhail (National Research Centre «Kurchatov Institute»); KORCHUGANOVA, Tatiana (National Research Tomsk Polytechnic University); PILYUGIN, Victor (National Research Nuclear University "MEPhI", Moscow, Russian Federation)

Presenter: Ms MARIA, Grigorieva (NRC KI)

Session Classification: 2. Operation, monitoring, optimization in distributed computing systems

Track Classification: 2. Operation, monitoring, optimization in distributed computing systems