



Contribution ID: 83

Type: Poster

Predictive analytics tools to adjust and monitor performance metrics for the ATLAS Production System

Thursday, 24 August 2017 16:25 (20 minutes)

Every scientific workflow involves an organizational part which purpose is to plan an analysis process thoroughly according to defined schedule, thus to keep work progress efficient. Having such information as an estimation of the processing time or possibility of system outage (abnormal behaviour) will improve the planning process, provide an assistance to monitor system performance and predict its next state.

The ATLAS Production System is an automated scheduling system that is responsible for central production of Monte-Carlo data, highly specialized production for physics groups, as well as data pre-processing and analysis using such facilities as grid infrastructures, clouds and supercomputers. With its next generation (ProdSys2) the processing rate is around 2M tasks per year that is more than 365M jobs per year. ProdSys2 evolves to accommodate a growing number of users and new requirements from the ATLAS Collaboration, physics groups and individual users. ATLAS Distributed Computing in its current state is the aggregation of large and heterogenous facilities, running on the WLCG, academic and commercial clouds, and supercomputers. This cyber-infrastructure presents computing conditions in which contention for resources among high-priority data analysis happens routinely, that might lead to significant workload and data handling interruptions. The lack of the possibility to monitor and predict the behaviour of the analysis process (its duration) and system's state itself caused to focus on design of the built-in situational awareness analytic tools.

Proposed suite of tools aims to estimate completion time (so called "Time To Complete", TTC) for every (production) task (i.e., prediction of the task duration), completion time for a chain of tasks, and to predict the failure state of the system (e.g., based on "abnormal" task processing times). Its implementation is based on Machine Learning methods and techniques, and besides the historical information about finished tasks it uses ProdSys2 job execution information and resources usage state (real-time parameters and metrics to adjust predicted values according to the state of the computing environment).

The WLCG ML R&D project started in 2016. Within the project the first implementation of the TTC Estimator (for production tasks) was developed, and its visualization was integrated into the ProdSys Monitor.

Primary authors: KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); GOLUBKOV, Dmitri (Institute for High Energy Physics (IHEP)-Unknown-Unknown); BARREIRO MEGINO, Fernando Harald (University of Texas at Arlington); GUBIN, Maksim (National Research Tomsk Polytechnic University (RU)); GRIGORYEVA, Maria (Institute for Theoretical and Experimental Physics (RU)); TITOV, Mikhail (National Research Centre Kurchatov Institute (RU)); BORODIN, Misha (University of Iowa (US)); PADOLSKI, Siarhei (BNL); MAENO, Tadashi (Brookhaven National Laboratory (US)); KORCHUGANOVA, Tatiana (National Research Tomsk Polytechnic University (RU))

Presenter: PADOLSKI, Siarhei (BNL)

Session Classification: Poster Session

Track Classification: Track 1: Computing Technology for Physics Research