Contribution ID: **533** Type: **Poster**

Modeling Allocation Utilization Strategies on Supercomputers

Tuesday 10 July 2018 16:40 (20 minutes)

Most supercomputers provide computing resources that are shared between users and projects, with utilization determined by predefined policies, load and quotas. The efficiency of the utilization of resources in terms of user/project depends on factors such as particular supercomputer policy and dynamic workload of supercomputer based on users' activities. The load on a resource is characterized by the number and parameters of jobs: the number of required nodes, required execution time (walltime), and jobs generation rate.
In this work we identify execution strategies geared towards the goal the maximizing the probability of utilization of allocated resources on a supercomputer. The execution strategies consist of find the optimal set of essential job parameters: number, size, length, rate. A simplified model for utilization of allocation time and a simulator based on queueing theory (with corresponding supercomputer Titan's requirements) were designed, the model was tested on both synthetic and real log data over many months of Titan's real work, identified strategies were compared with other possible strategies.
Experiments conducted using the simulator, showed that in most cases identified strategies increase the probability of utilizing allocation faster than a random choice of job processing parameters. We also find that the accuracy of the model will be higher if the amount of resources for utilization is larger, analyzed time intervals are longer and supercomputer's state is steadier over these intervals.

Authors: POYDA, Alexey (National Research Centre Kurchatov Institute (RU)); TITOV, Mikhail (National Research Centre Kurchatov Institute (RU)); KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); Dr WELLS, Jack C. (Oak Ridge National Laboratory (US)); ORAL, Sarp (Oak Ridge National Laboratory (US)); DE, Kaushik (University of Texas at Arlington (US)); OLEYNIK, Danila (University of Texas at Arlington (US)); JHA, Shantenu (Rutgers University (US))

Presenters: POYDA, Alexey (National Research Centre Kurchatov Institute (RU)); TITOV, Mikhail (National Research Centre Kurchatov Institute (RU))

Session Classification: Posters

Track Classification: Track 3 –Distributed computing