

10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)



Contribution ID: 340

Type: **not specified**

Software for predicting a moment of task start in computer cluster by statistical analysis of jobs queue history

Monday, 3 July 2023 16:45 (15 minutes)

The state of queues in modern computing centers is such that a user's task can hold in the queue before start for weeks. However, even if the queuing system gives a forecast for the start of a task, the forecast often turns out to be incorrect. It is happening because during the week the task is in the queue, there are some events will occur that will change the moments tasks starts. We know the amount of resources required for the task but, it is difficult to get an answer about the moment the task started from the queuing system without placing the task in the queue.

The paper proposes to improve the previously created tool for predicting the characteristics of tasks. It is performed through the use of more modern software interfaces to the queuing system SchedMD (Slurm), as well as the use of various mathematical forecasting methods to predict the moment the task starts in the queue.

To compare and evaluate machine learning models, a Standard Workload Format (SWF) file was selected with task execution history data on a computing cluster of the University of Luxembourg with 59715 tasks. The following methods have been explored: linear regression with L2 regularization, support vector machines, random forest, LightGBM, CatBoost, LightGBM with parameter optimization, CatBoost with parameter optimization. To check the correctness of the prediction, a test bench was built based on the Slurm simulator (SUNY Center for Computational Research at the University of Buffalo, USA), which works on the basis of tasks execution and queuing saved logs.

Summary

Primary authors: SALNIKOV, Alexey (Lomonosov Moscow State University); YASHCHENKO, Igor (Lomonosov Moscow State University)

Presenter: YASHCHENKO, Igor (Lomonosov Moscow State University)

Session Classification: Distributed Computing Systems

Track Classification: Distributed Computing Systems