

The 8th International Conference "Distributed Computing and
Grid-technologies in Science and Education" (GRID 2018)



Contribution ID: 245

Type: **Sectional reports**

Cluster analysis of scientific payload to execute it efficiently in distributed computing environment

Monday, 10 September 2018 14:00 (15 minutes)

Every modern scientific experiment deals with the processing of large amounts of experimental data (up to exabytes) employing millions of computing processes and delivering corresponding scientific payloads. It raises the task of the efficient management of the batch processing of payloads, that should consider the well-defined grouping mechanism.

Understanding naturally occurring groupings of payloads would increase the processing rate and improve the scheduling. The automated discovery of the payloads groups should consider not only the descriptive parameters of the payload itself but the characteristics of its interaction processes with computing resources and the computing environment.

Our work is focused on applying machine learning methods to solve the stated problem, and particularly to evaluate and to use the approach of cluster analysis. Besides the ultimate goal, it will benefit the other related analytical services and processes aimed at analyzing particular payload parameters (e.g., prediction process) and/or set of parameters (e.g., correlations discovery).

Primary author: GUBIN, Maksim (Tomsk Polytechnic University)

Co-authors: GRIGORYEVA, Maria (National Research Centre «Kurchatov Institute»); TITOV, Mikhail (National Research Centre «Kurchatov Institute»)

Presenter: GUBIN, Maksim (Tomsk Polytechnic University)

Session Classification: 11. Big data Analytics, Machine learning

Track Classification: 11. Big data Analytics, Machine learning