Contribution ID: **120**   Type: **Sectional reports**

Grid datasets popularity estimation using gradient boosting

Thursday, 7 July 2016 13:30 (15 minutes)

In this paper we present machine learning based technique to
estimate access patterns for grid datasets. We have analyzed 3-year
historical data of Kurchatov Institute Tier1 site and applied gradient
boosting algorithm to predict nearest future access patterns for the
dataset based on its previous access statistics. We show our method to
be effective for ATLAS data popularity estimation. Our method can be
used to optimize grid storage in two ways: to move unpopular
datasets to tape storage to save disk space and to increase number of
replicas of popular datasets to reduce access latency.

Primary authors: TESLYUK, Anton (Borisovich); CHERNYUK, Daria (Andreevna)

Presenter: CHERNYUK, Daria (Andreevna)

Session Classification: 10. Databases, Distributed Storage systems, Big data Analytics

Track Classification: 10. Databases, Distributed Storage systems, Big data Analytics