Contribution ID: **2**   Type: **not specified**

Efficient Data Management Tools for the Heterogeneous Big Data Warehouse

Friday, 2 October 2015 15:40 (15 minutes)

The traditional relational databases (aka RDBMS) having been consistent for the normalized data structures. RDBMS served well for decades, but the technology is not optimal for data processing and analysis in data intensive fields like social networks, oil-gas industry, experiments at the Large Hadron Collider, etc. Several challenges have been raised recently on the scalability of data warehouse-like workload against the trans-actional schema, in particular for the analysis of archived data or the aggregation of data for summary and accounting purposes. We have evaluated new approaches of handling vast amount of data. In particular, we have studied a new class of technologies commonly referred to as non-relational (NoSQL) databases. This includes schema-less approaches via key-value stores, like HBase, Cassandra, MongoDB. We studied performance, throughput and scalability of the above technologies for several scientific and industrial use-cases. The detailed studies and comparison make this project successful for different heterogeneous systems. This paper presents technologies and architectures we have studied, as well as the description of the back-end application that implements data uploading from RDBMS to NoSQL data warehouse, NoSQL database organization and how it could be used for data analytics in the further.

Primary author: Ms OSIPOVA, Victoriya (Tomsk Polytechnic University, Tomsk, Russia)

Co-authors: Mr ALEKSEEV, Alexey (Tomsk Polytechnic University, Tomsk, Russia); Mr KLIMENTOV, Alex-eyi (Brookhaven National Laboratory, Upton, USA National Research Center "Kurchatov Insitute", Moscow, Russia); Ms SEIDOVA, Asel (Tomsk Polytechnic University, Tomsk, Russia); Mr IVANOV, Maksim (Tomsk Polytechnic University, Tomsk, Russia); Ms GRIGOREVA, Nina (Tomsk Polytechnic University, Tomsk, Russia); Ms GERGET, Olga (Tomsk Polytechnic University, Tomsk, Russia)

Presenter: Ms OSIPOVA, Victoriya (Tomsk Polytechnic University, Tomsk, Russia)

Session Classification: Workload Management Systems in Applied Research and BigData