



Contribution ID: 422

Type: **Parallel**

ATLAS DDM/DQ2 & NoSQL databases: Use cases and experiences

Monday, May 21, 2012 5:50 PM (25 minutes)

The Distributed Data Management System DQ2 is responsible for the global management of petabytes of ATLAS physics data. DQ2 has a critical dependency on Relational Database Management Systems (RDBMS), like Oracle, as RDBMS are well suited to enforce data integrity in online transaction processing application. Despite these advantages, concerns have been raised recently on the scalability of data warehouse-like workload against the relational schema, in particular for the analysis of archived data or the aggregation of data for summary purposes. Therefore, we have considered new approaches of handling very large amount of data. More specifically, we investigated a new class of database technologies commonly referred to as

NoSQL databases. This includes distributed file system like HDFS that support parallel execution of computational tasks on distributed data, as well as schema-less approaches via key-value/document stores, like HBase, Cassandra or MongoDB. These databases provide solutions to particular types of problems: for example, NoSQL databases have demonstrated horizontal scalability, high throughput, automatic fail-over mechanisms, and provide easy replication support over LAN and WAN.

In this talk, we will describe our use cases in ATLAS, and share our experiences with NoSQL databases in a comparative study with Oracle.

Primary author: ATLAS, Collaboration (Atlas)

Co-authors: MOLFETAS, Angelos (CERN); DIMITROV, Gancho (Brookhaven National Laboratory (US)); STEWART, Graeme Andrew (CERN); CANALI, Luca (CERN); LASSNIG, Mario (CERN); BARISITS, Martin (Vienna University of Technology (AT)); GARONNE, Vincent (CERN)

Presenter: LASSNIG, Mario (CERN)

Session Classification: Software Engineering, Data Stores and Databases

Track Classification: Software Engineering, Data Stores and Databases (track 5)