Contribution ID: **14**   Type: **not specified**

The ATLAS Distributed Data Management System & Databases

Wednesday 29 May 2013 15:50 (20 minutes)

The ATLAS Distributed Data Management (DDM) System is responsible for the global management of petabytes of high energy physics data. The current system, DQ2, has a critical dependency on Relational Database Management Systems (RDBMS), like Oracle. RDBMS are well-suited to enforcing data integrity in online transaction processing applications, however, concerns have been raised about the scalability of its data warehouse-like workload. In particular, analysis of archived data or aggregation of transactional data for summary purposes is problematic. Therefore, we have evaluated new approaches to handle vast amounts of data. We have investigated a class of database technologies commonly referred to as NoSQL databases. This includes distributed filesystems, like HDFS, that support parallel execution of computational tasks on distributed data, as well as schema-less approaches via key-value stores, like HBase.

In this talk we will describe our use cases in ATLAS, share our experiences with various databases used in production and present the database technologies envisaged for the next-generation DDM system, Rucio. Rucio is an evolution of the ATLAS DDM system which addresses the scalability issues observed in DQ2.

Author: GARONNE, Vincent (CERN)

Co-authors: SERFON, Cedric (CERN); LASSNIG, Mario (University of Innsbruck (AT)); BARISITS, Martin (CERN); VIGNE, Ralph (University of Vienna (AT)); BEERMANN, Thomas (Bergische Universitaet Wuppertal (DE))

Presenter: GARONNE, Vincent (CERN)

Session Classification: Big Data Processing on Modern Architectures