



Contribution ID: 71

Type: **not specified**

## Evaluation of distributed open source solutions in CERN database use cases

*Tuesday 24 March 2015 17:20 (25 minutes)*

There are terabytes of data stored in a relational database (Oracle) at CERN which in fact does not need a relational model. Moreover, using a relational database management system very often brings a significant overhead in terms of resource utilization. The problem is notably observable for warehouse-type data sets. At the same time running analytical workloads on such data sets requires large amount of computing power combined with high storage throughput, a combination which can be achieved with a scalable database system. Introducing this kind of system will not only speed up processing but open new possibilities for data mining as well. This presentation will discuss advantages of using distributed architecture like Hadoop for scalable data processing of CERN data sets; such as stored in Oracle: DB LHC logging system, SCADA systems or even LHC experiments events data stored in Ntuples.

**Authors:** SURDY, Kacper (CERN); GRZYBEK, Maciej (Warsaw University of Technology (PL)); BARANOWSKI, Zbigniew (CERN)

**Co-authors:** LANZA GARCIA, Daniel (Univ. Extremadura, Cen. Uni. Merida (ES)); GRANCHER, Eric (CERN); CANALI, Luca (CERN)

**Presenter:** SURDY, Kacper (CERN)

**Session Classification:** Storage and File Systems

**Track Classification:** Storage & Filesystems