



Contribution ID: 404

Type: **Oral presentation to parallel session**

Sequential Data access with Oracle and Hadoop: a performance comparison

Tuesday 15 October 2013 17:48 (20 minutes)

The Hadoop framework has proven to be an effective and popular approach for dealing with “Big Data” and, thanks to its scaling ability and optimised storage access, Hadoop Distributed File System-based projects such as MapReduce or HBase are seen as candidates to replace traditional relational database management systems whenever scalable speed of data processing is a priority. But do these projects deliver in practice? Does migrating to Hadoop’s “shared nothing” architecture really improve data access throughput? And, if so, at what cost?

We answer these questions—addressing cost/performance as well as raw performance—based on a performance comparison between an Oracle-based relational database and Hadoop’s distributed solutions like MapReduce or HBase for sequential data access. A key feature of our approach is the use of an unbiased data model as certain data models can significantly favour one of the technologies tested.

Author: BARANOWSKI, Zbigniew (CERN)

Co-authors: GRANCHER, Eric (CERN); CANALI, Luca (CERN)

Presenter: BARANOWSKI, Zbigniew (CERN)

Session Classification: Data Stores, Data Bases, and Storage Systems

Track Classification: Data Stores, Data Bases, and Storage Systems