



Contribution ID: 328

Type: oral presentation

## Evaluation of NoSQL databases for DIRAC monitoring and beyond

*Monday, April 13, 2015 2:45 PM (15 minutes)*

Nowadays, many database systems are available but they may not be optimized for storing time series data. The DIRAC job monitoring is a typical use case of such time series. So far it was done using a MySQL database, which is not well suited for such an application. Therefore alternatives have been investigated.

Choosing an appropriate database for storing huge amounts of time series is not trivial as one must take into account different aspects such manageability, scalability, extensibility etc.

We compared the performance of Elasticsearch, OpenTSDB that is based on HBase and InfluxDB time series NoSQL databases using the same set of machines and the same data. We also evaluated the effort required for maintaining them.

Using the LHCb Workload Management System, based on DIRAC, as a use case we have setup a new monitoring system in parallel with the current MySQL system and we publish the same data into the databases under test. We have evaluated Grafana (for OpenTSDB) and Kibana (for Elasticsearch) metrics and graph editors for creating dashboards in order to have clear picture on the usability of each candidate. In this paper we present the result of this study and the performance of the selected technology. We also give an outlook of other potential applications of NoSQL databases with DIRAC project.

**Author:** MATHE, Zoltan (CERN)

**Co-authors:** CASAJUS RAMO, Adrian (University of Barcelona (ES)); STAGNI, Federico (CERN); TOMASSETTI, Luca (University of Ferrara and INFN)

**Presenter:** STAGNI, Federico (CERN)

**Session Classification:** Track 3 Session

**Track Classification:** Track3: Data store and access