

Monitoring performance of a highly distributed and complex computing infrastructure in LHCb

Monday, 10 October 2016 15:15 (15 minutes)

In order to ensure an optimal performance of the LHCb Distributed Computing, based on LHCbDIRAC, it is necessary to be able to inspect the behavior over time of many components: firstly the agents and services on which the infrastructure is built, but also all the computing tasks and data transfers that are managed by this infrastructure. This consists of recording and then analyzing time series of a large number of observables, for which the usage of SQL relational databases is far from optimal. Therefore within DIRAC we have been studying novel possibilities based on NoSQL databases (ElasticSearch, OpenTSDB and InfluxDB) as a result of this study we developed a new monitoring system based on ElasticSearch. It has been deployed on the LHCb Distributed Computing infrastructure for which it collects data from all the components (agents, services, jobs) and allows creating reports through Kibana and a web user interface, which is based on the DIRAC web framework.

In this paper we describe this new implementation of the DIRAC monitoring system. We give details on the ElasticSearch implementation within the DIRAC general framework, as well as an overview of the advantages of the pipeline aggregation used for creating a dynamic bucketing of the time series. We present the advantages of using the ElasticSearch DSL high-level library for creating and running queries. Finally we shall present the performances of that system.

Secondary Keyword (Optional)

Databases

Primary Keyword (Mandatory)

Monitoring

Tertiary Keyword (Optional)

Primary author: MATHE, Zoltan (CERN)

Co-authors: HAEN, Christophe (CERN); STAGNI, Federico (CERN)

Presenter: HAEN, Christophe (CERN)

Session Classification: Track 7: Middleware, Monitoring and Accounting

Track Classification: Track 7: Middleware, Monitoring and Accounting