



Contribution ID: 270

Type: **oral presentation**

Analysis of CERN Computing Infrastructure and Monitoring Data

Monday, April 13, 2015 2:00 PM (15 minutes)

Optimising a computing infrastructure on the scale of LHC requires a quantitative understanding of a complex network of many different resources and services. For this purpose the CERN IT department and the LHC experiments are collecting a large multitude of logs and performance probes, which are already successfully used for short-term analysis (e.g. operational dashboards) within each group.

The IT analytics working group has been created with the goal to bring data sources from different services and on different abstraction levels together and to implement a suitable infrastructure for mid- to long-term statistical analysis. It further provides a forum for joint optimization across single service boundaries and the exchange of analysis methods and tools.

To simplify access to the collected data, we implemented an automated repository for cleaned and aggregated data sources based on the Hadoop ecosystem. This contribution describes some of the challenges encountered, such as dealing with heterogeneous data formats, selecting an efficient storage format for map reduce and external access, and will describe the repository user interface.

Using this infrastructure we were able to quantitatively analyse the relationship between CPU/wall fraction, latency/throughput constraints of network and disk and the effective job throughput.

In this contribution we will first describe the design of the shared analysis infrastructure and then present a summary of first analysis results from the combined data sources.

Primary authors: NIEKE, Christian (Brunswick Technical University (DE)); DUELLMANN, Dirk (CERN)

Presenter: NIEKE, Christian (Brunswick Technical University (DE))

Session Classification: Track 6 Session

Track Classification: Track6: Facilities, Infrastructure, Network