

21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015)



Contribution ID: 109

Type: oral presentation

Monitoring WLCG with lambda-architecture: a new scalable data store and analytics platform for monitoring at petabyte scale.

Tuesday 14 April 2015 15:15 (15 minutes)

Monitoring the WLCG infrastructure requires to gather and to analyze high volume of heterogeneous data (e.g. data transfers, job monitoring, site tests) coming from different services and experiment-specific frameworks to provide a uniform and flexible interface for scientists and sites. The current architecture, where relational database systems are used to store, to process and to serve monitoring data, has limitations in coping with the foreseen extension of the volume (e.g. higher LHC luminosity) and the variety (e.g. new data-transfer protocols and new resource-types, as cloud-computing) of WLCG monitoring events. This paper presents a new scalable data store and analytics platform designed by the Support for Distributed Computing (SDC) group, at the CERN IT department, which leverages on a stack of technology each one targeting specific aspects on big-scale distributed data-processing (commonly referred as lambda-architecture approach). Results on data processing on Hadoop for WLCG data transfers are presented, showing how the new architecture can easily analyze hundreds of millions of transfer logs in few minutes. Moreover, a comparison on data partitioning, compression and file format (e.g. CSV, AVRO) is presented, with particular attention on how the file structure impacts the overall MapReduce performance. In conclusion, the evolution of the current implementation, which focuses on data store and batch processing, towards a complete lambda-architecture is discussed, with consideration on candidate technology for the serving layer (e.g. ElasticSearch) and a description of a proof of concept implementation, based on Esper, for the real-time part which compensates for batch-processing latency and automates problem detection and failures.

Author: MAGNONI, Luca (CERN)

Co-authors: Prof. KHAN, Akram (Brunel University); Prof. SMITH, David (Brunel University); ANDREEVA, Julia (CERN); Mr SUTHAKAR, Uthayanath (Brunel University)

Presenter: MAGNONI, Luca (CERN)

Session Classification: Track 6 Session

Track Classification: Track6: Facilities, Infrastructure, Network