

NXCALS Data Extraction and Analysis with Apache Spark

Monday 29 May 2017 16:00 (20 minutes)

The CERN Accelerator Logging Service (CALs) was designed in 2001, and has been in production for 14 years. It is a mission-critical service for the operation of the LHC (Large Hadron Collider).

CALs uses an Oracle database for storage of technical accelerator data and persists approx 0.75 petabytes of data coming from more than 1.5 million pre-defined signals. These signals represent data related to CERN's core infrastructure such as electricity, cooling and ventilation, industrial data such as cryogenics, vacuum and control devices, and beam-related data such as beam positions, currents, losses, etc.

Over time, the scope of the service and the data mining requirements have evolved significantly resulting in the current infrastructure slowly reaching hard scalability limits. In order to address this, a next generation Hadoop based Logging System (NXCALS) is currently being developed. This new system provides a Data Analysis Platform based on Apache Spark.

This presentation will briefly introduce the background of the Logging Service, give a general overview of the new NXCALS system, and then go into more details regarding the use of Apache Spark.

Author: TSVETKOV, Nikolay (CERN)

Co-authors: WOZNIAK, Jakub (CERN); SOBIESZEK, Marcin (CERN)

Presenter: TSVETKOV, Nikolay (CERN)

Session Classification: Implementations & Technologies

Track Classification: Implementations & Technologies