Contribution ID: **21**   Type: **not specified**

Big Data at CERN

Thursday, 20 September 2018 10:00 (1 hour)

The LHC experiments continue to produce a wealth of valu-
able High Energy Physics data, which oer numerous possibilities for
new discoveries. The IT Department at CERN provides Hadoop and
Spark services and works closely with the scientic communities in their
quest to analyze and understand these vast amounts of physics and in-
frastructure data. The number of CERN teams using these services for
their systems has grown signicantly over the past years, since Big Data
technologies -such as Apache Spark- show great potential in speeding up
their existing workloads. The most signicant systems include the CMS
Data Reduction Facility which aims to reduce 1 PB of data produced
by the CMS Experiment to 1 TB of reusable data for physics analysis
through Spark, the Next CERN Accelerator Logging Service (NXCALS)
which will perform online and oine analysis over the data acquired from
each of the 20,000 devices that monitor the CERN accelerator complex,
as well as the monitoring system for the CERN Data Center and the
Worldwide LHC Computing Grid (WLCG) which consists of more than
170 dierent computing centers in 42 countries. This talk will provide
an overview of the current infrastructure based on Spark and other key
components of the Hadoop ecosystem, the active use cases on big data
analytics from various CERN communities, as well as the challenges in
the available data sources and their architecture.

Primary author: Dr MOTESNITSALIS, Vaggelis (CERN)

Presenter: Dr MOTESNITSALIS, Vaggelis (CERN)

Session Classification: Infrastructures