Contribution ID: **57**   Type: **Lecture**

Big Data technologies and distributed data processing with SQL (lecture)

Wednesday, 30 September 2020 15:10 (1 hour)

The interest of many users communities in solutions based on the Big Data ecosystem such as Hadoop is constantly increasing at CERN, including physics experiments, monitoring, and accelerator controls.

This lecture will introduce the participant to some of the Big Data technologies that CERN offers for distributed processing. It will cover some history, architecture, and specifics of selected technologies such as the Hadoop, Spark, and Presto for SQL-like data processing.

Moreover, this talk will cover the example of using the Big Data tools to speedup some computations on hundreds of terabytes of events coming from the Atlas experiment at CERN.

Exercise hours

Lecture hours

Presenter: KLESZCZ, Emil (CERN)