Contribution ID: **15**                                                            Type: **Notebook talk**

# Basic Physics Analyses Implemented Using Apache Spark.

*Wednesday, 14 September 2022 15:00 (30 minutes)*

Apache Spark is a very successful open-source tool for data processing. This talk will focus on the use of Spark and its DataFrame API in the context of HEP. We will go through a few demos of some simple and outreach-style analyses implemented using Jupyter notebooks and the Spark Python API (PySpark). We will wrap up with a short discussion of the key features in Spark and its ecosystem that can be useful for Physics analysis and what still needs improvements.

**Primary author:**   CANALI, Luca (CERN)

**Presenter:**   CANALI, Luca (CERN)

**Session Classification:**   Plenary Session Wednesday