

Basic Physics Analyses Implemented Using Apache Spark.

Wednesday, September 14, 2022 3:00 PM (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.

Author: CANALI, Luca (CERN)

Presenter: CANALI, Luca (CERN)

Session Classification: Plenary Session Wednesday