Contribution ID: **152**   Type: **Presentation**

Utilizing RDataFrame for Data Preservation and Open Publishing Data and Analyzes Software for HEP

Tuesday 12 March 2024 15:00 (15 minutes)

CERN produces, analyzes and archives vast amounts of data. To conduct an analysis a lot of software in the form of scripts and code is produced. As the time goes by and new approaches supersede the old ones, the aforementioned artifacts may become hard to understand and setting up and running them can be challenging. This may be a crucial concern when trying to publish the data in an open repository like CERN OpenData. Furthermore, an old code cannot leverage new technological advancements which could potentially enhance its performance.

To address this issue an effort to restore data and analysis from LHC Run 1 has been conducted. This work describes the process of transforming data regarding the analysis and code from the LHC Run 1 at the TOTEM experiment. It utilizes RDataFrame –a modern data processing tool –to transcribe C++ scripts into a form of a comprehensible Jupyter notebook. As a result, the number of lines of code has been greatly reduced, thus enhancing the readability. In addition, the notebook can be run on a novel serverless engine architecture.

The process described in this work shows potential applicability for further data preservation and publication efforts.

Authors: KRUCZKIEWICZ, Pawel (AGH University of Krakow (PL)); GRZANKA, Leszek (AGH University of Krakow (PL)); AVATI, Valentina (AGH University of Krakow (PL)); BURKIEWICZ, Kamil Krzysztof (AGH University of Science and Technology (PL)); MALAWSKI, Maciej (AGH University of Krakow (PL))

Presenter: KRUCZKIEWICZ, Pawel (AGH University of Krakow (PL))

Session Classification: FAIR Data Management

Track Classification: Main sessions: User Voice: Innovative Applications, Data Science Environments & Open Data