Contribution ID: **12**   Type: **not specified**

Overtaking PROOF: The Future of Distributed Analysis in ROOT

Wednesday, 12 September 2018 11:45 (15 minutes)

Widespread distributed processing of big datasets has been around for more than a decade now thanks to Hadoop, but only recently higher-level abstractions have been proposed for programmers to easily operate on those datasets, e.g. Spark. ROOT has joined that trend with its RDataFrame tool for declarative analysis, which currently supports local multi-threaded parallelisation. However, RDataFrame's programming model is general enough to accommodate multiple implementations or backends: users could write their code once and execute it as is locally or distributedly, just by selecting the corresponding backend. Moreover, the programmer should be provided with ergonomic interfaces, possibly web-based, which allow to dynamically plug in new resources, as well as to write, execute, monitor and debug distributed applications in the most intuitive way possible.

Primary authors: TEJEDOR SAAVEDRA, Enric (CERN); CERVANTES VILLANUEVA, Javier (CERN)

Presenter: CERVANTES VILLANUEVA, Javier (CERN)

Session Classification: End User Perspective

Track Classification: Presentations