Contribution ID: **51** Type: **Presentation**

SWAN and its analysis ecosystem

Tuesday 29 January 2019 11:35 (20 minutes)

CERN, and High Energy Physics (HEP) in general, face unprecedented challenges in data storage, processing and analysis. With the planned improvements to the Large Hadron Collider (LHC), including the High-Luminosity LHC, there is an expected increase of data in one order of magnitude. After processing and filtering these data, new tools and solutions, capable of dealing with such large datasets, are particularly important for the last phases of analysis.

SWAN (Service for Web-based ANalysis) is a service that provides an interactive interface to access data analysis tools from the web, allowing users to perform their work in a simpler way and with much larger datasets. Its integration with CERN's infrastructure, more precisely with users synchronized storage and syncing capabilities (via CERNBox), computing resources, experiments data (via EOS) and software, allows a seamless experience across all of our user scenarios and devices.

But even more data means even more resources and different approaches. SWAN has recently been integrated with CERN's Spark Clusters and we are already working to provide access to our Worldwide LHC Computing Grid (WLCG) batch service. We're also experimenting with the integration outside of CERN's infrastructure, using external cloud vendors to install and deploy our Science Box service (which bundles SWAN, CERNBox, EOS and CVMFS). These last experiments are also important in our pursuit of bringing SWAN to new education scenarios, namely via the UP2University European Project.

Authors: CASTRO, Diogo (CERN); MOSCICKI, Jakub (CERN); LAMANNA, Massimo (CERN)

Co-authors: BOCCHI, Enrico (CERN); TEJEDOR SAAVEDRA, Enric (CERN); PIPARO, Danilo (CERN); KOTHURI, Prasanth (CERN)

Presenter: CASTRO, Diogo (CERN)

Session Classification: Data science: applications and infrastructure

Track Classification: Cloud infrastructure and software stacks for data science