Contribution ID: **111**   Type: **Short Talk**

First experiences with a portable analysis infrastructure for LHC at INFN

Thursday, 20 May 2021 15:26 (13 minutes)

The challenges proposed by the HL-LHC era are not limited to the sheer amount of data to be processed: the capability of optimizing the analyser's experience will also bring important benefits for the LHC communities, in terms of total resource needs, user satisfaction and in the reduction of end time to publication. At the Italian National Institute for Nuclear Physics (INFN) a portable software stack for analysis has been proposed, based on cloud-native tools and capable of providing users with a fully integrated analysis environment for the CMS experiment. The main characterizing traits of the solution consist in the user-driven design and the portability to any cloud resource provider. All this is made possible via an evolution towards a "python-based" framework, that enables the usage of a set of open-source technologies largely adopted in both cloud-native and data-science environments. In addition, a "single sign on"-like experience is available thanks to the standards-based integration of INDIGO-IAM with all the tools. The integration of compute resources is done through the customization of a JupyterHUB solution, able to spawn identity-aware user instances ready to access data with no further setup actions. The integration with GPU resources is also available, designed to sustain more and more widespread ML based workflow. Seamless connections between the user UI and batch/big data processing framework (Spark, HTCondor) are possible. Eventually, the experiment data access latency is reduced thanks to the integrated deployment of a scalable set of caches, as developed in the context of ESCAPE project, and as such compatible with the future scenarios where a data-lake will be available for the research community.

The outcome of the evaluation of such a solution in action is presented, showing how a real CMS analysis workflow can make use of the infrastructure to achieve its results.

Primary authors: CIANGOTTINI, Diego (INFN, Perugia (IT)); BOCCALI, Tommaso (INFN Sezione di Pisa, Universita' e Scuola Normale Superiore, P); CECCANTI, Andrea (Universita e INFN, Bologna (IT)); SALOMONI, Davide (Universita e INFN, Bologna (IT)); SPIGA, Daniele (Universita e INFN, Perugia (IT)); TEDESCHI, Tommaso (Universita e INFN, Perugia (IT)); TRACOLLI, Mirco (INFN Perugia)

Presenter: CIANGOTTINI, Diego (INFN, Perugia (IT))

Session Classification: Virtualisation

Track Classification: Distributed Computing, Data Management and Facilities