Contribution ID : **9**   Type : **Talk**

Developing Artificial Intelligence in the Cloud: the AI_INFN platform

Wednesday, 5 June 2024 11:40 (25)

The INFN CSN5-funded project AI_INFN ("Artificial Intelligence at INFN") aims at fostering the adoption of ML and AI within INFN by providing support on multiple aspects, including the provision of state-of-the-art hardware for AI and ML, leveraging on cloud native solutions in the context of INFN Cloud, to share hardware accelerators as effectively as possible without compromising on the diversity of the research activities of the Institute. AI_INFN evolves the Virtual-Machine-based model towards a more flexible platform built on top of Kubernetes. This is meant to be a composable toolkit and currently features: JWT-based authentication, JupyterHub multitenant interface, distributed filesystem, customizable conda environments, and a specialized monitoring and accounting system. Last but not least, the platform is an enabler to implement the offloading mechanism based on Virtual Kubelet and interLink API, a synergy with InterTwin. Preliminary results and applications will be presented.

Primary author(s) : SPIGA, Daniele (INFN); CIANGOTTINI, Diego (INFN); BIANCHINI, Giulio (INFN); AN-DERLINI, Lucio (INFN); PETRINI, Rosa (INFN); DAL PRA, Stefano (INFN)

Presenter(s) : PETRINI, Rosa (INFN)

Session Classification : High Performance Computing

Track Classification : High Performance Computing