

Developing Artificial Intelligence in the Cloud: the AI_INFN platform & InterLink

Friday 13 September 2024 08:45 (7 minutes)

The INFN CSN5-funded project AI_INFN ("Artificial Intelligence at INFN") aims to promote ML and AI adoption within INFN by providing comprehensive support, including state-of-the-art hardware and cloud-native solutions within INFN Cloud. This facilitates efficient sharing of hardware accelerators without hindering the institute's diverse research activities. AI_INFN advances from a Virtual-Machine-based model to a flexible Kubernetes-based platform, offering features such as JWT-based authentication, JupyterHub multitenant interface, distributed filesystem, customizable conda environments, and specialized monitoring and accounting systems. It also enables offloading mechanisms using Virtual Kubelet and InterLink API, synergizing with InterTwin. This setup can manage workflows across various providers and hardware types, which is crucial for scientific use cases that require dedicated infrastructures for different parts of the workload. The project aims to perform functional tests and benchmarks to validate its production applicability. Initial test results, emerging case studies and integration scenarios will be presented.

Summary

Presenter: PETRINI, Rosa (INFN Sezione di Pisa, Universita' e Scuola Normale Superiore, P)