

Exploring the self-service model to visualize the results of the ATLAS Machine Learning analysis jobs in BigPanDA with Openshift OKD3

Thursday 20 May 2021 15:13 (13 minutes)

A large scientific computing infrastructure must offer versatility to host any kind of experiment that can lead to innovative ideas. The ATLAS experiment offers wide access possibilities to perform intelligent algorithms and analyze the massive amount of data produced in the Large Hadron Collider at CERN. The BigPanDA monitoring is a component of the PanDA (Production ANd Distributed Analysis) system and its main role is to monitor the entire lifecycle of a job/task running in the ATLAS Distributed Computing infrastructure. Because many scientific experiments now rely upon Machine Learning algorithms, the BigPanDA community desires to expand the platform's capabilities and fill the gap between Machine Learning processing and data visualization. In this regard, BigPanDA partially adopts the cloud-native paradigm and entrusts the data presentation to MLFlow services running on Openshift OKD. Thus, BigPanDA interacts with the OKD API and instructs the containers orchestrator how to locate and expose the results of the Machine Learning analysis. The proposed architecture also introduces various DevOps-specific patterns, including continuous integration for MLFlow middleware configuration and continuous deployment pipelines that implement rolling upgrades. The Machine Learning data visualization services operate on demand and run for a limited time, thus optimizing the resource consumption.

Primary authors: STAN, Ioan-Mihail (University Politehnica of Bucharest (RO)); PADOLSKI, Siarhei (BNL); Mr LEE, Chris (Stony Brook University (US))

Presenter: STAN, Ioan-Mihail (University Politehnica of Bucharest (RO))

Session Classification: Monitoring

Track Classification: Distributed Computing, Data Management and Facilities