

## Anomaly detection in the CERN cloud infrastructure

*Wednesday, 19 May 2021 11:16 (13 minutes)*

Anomaly detection in the CERN OpenStack cloud is a challenging task due to the large scale of the computing infrastructure and, consequently, the large volume of monitoring data to analyse. The current solution to spot anomalous servers in the cloud infrastructure relies on a threshold-based alarming system carefully set by the system managers on the performance metrics of each infrastructure's component. This contribution explores fully automated, unsupervised machine learning solutions in the anomaly detection field for time series metrics, by adapting both traditional and deep learning approaches. The paper describes a novel end-to-end data analytics pipeline implemented to digest the large amount of monitoring data and to expose anomalies to the system managers. The pipeline relies solely on open-source tools and frameworks, such as Spark, Apache Airflow, Kubernetes, Grafana, Elasticsearch. In addition, an approach to build annotated datasets from the CERN cloud monitoring data is reported. Finally, a preliminary performance of a number of anomaly detection algorithms is evaluated by using the aforementioned annotated datasets.

**Primary authors:** GIORDANO, Domenico (CERN); PALTENGGHI, Matteo (Politecnico di Milano (IT)); METAJ, Stiven (Politecnico di Milano (IT)); DVORAK, Antonin (Nuclear Physics Institute of the Czech Academy of Sciences, Řež, Czech Republic)

**Presenter:** METAJ, Stiven (Politecnico di Milano (IT))

**Session Classification:** Facilities and Networks

**Track Classification:** Distributed Computing, Data Management and Facilities