



Contribution ID: 45

Type: **not specified**

Anomaly detection of large scale distributed storage system based on machine learning

Wednesday 30 January 2019 11:30 (20 minutes)

In a large scale storage system which consists of hundreds of servers, tens of thousands of clients, a variety of devices, anomaly detection is a nontrivial task. Traditional solutions which are still working in our cluster operations include setting static thresholds on KPIs, searching key words in system logs and so on. These methods highly depend on experience of system administrators, cannot adapt to new anomalies in the cluster. The machine learning communities has developed a wide range of algorithms which is able to do anomaly detection of high dimensional data by statistically learning over a large training data set. The monitoring system of IHEP cluster has accumulated billions of performance metrics entries in its database. It provides possibilities to train those machine learning models. This presentation will show how our preliminary works on anomaly detection of ganglia time serials monitoring data by machine learning algorithms including LSTM, HTM and Isolation Forest.

Primary author: WANG, Lu (Computing Center, Institute of High Energy Physics, CAS)

Presenter: WANG, Lu (Computing Center, Institute of High Energy Physics, CAS)

Session Classification: Scalable Storage Backends for Cloud, HPC and Global Science

Track Classification: Scalable Storage Backends for Cloud, HPC and Global Science