Anomaly Detection in dCache Storage System

The dCache project provides open-source software that is deployed worldwide to meet the increasingly demanding requirements of scientific storage. It offers a multifaceted approach to support various use cases using the same storage infrastructure, which includes high-throughput data ingestion, data sharing over wide area networks, efficient access from HPC clusters, and long-term data persistence on tertiary storage.

DESY-IT operates multiple dCache instances that generate gigabytes of log files each day. These log files contain information about data access and operations, which can be utilized to identify misbehavior by malicious users or system failures.

This project aims to explore possibilities to automate the processing of dCache log data and implement techniques for uncovering unusual patterns in system operations by applying state-of-the-art machine learning methods to detect anomalies effectively.

Prerequisites:

- Python programming skills
- Familiarity with Jupiter Notebooks, NumPy, Pandas and Kafka
- Fundamental knowledge of machine learning concepts is beneficial but not required
- · Experience with Linux OS is a plus

Group

IT

Project Category

B5. Computing

Special Qualifications

Prerequisites: - Python programming skills - Familiarity with Jupiter Notebooks, NumPy, Pandas and Kafka -Fundamental knowledge of machine learning concepts is beneficial but not required • Experience with Linux OS is a plus

DESY Site

Hamburg

Primary authors: Dr SCHWARZ, Kilian (IT (IT Scientific Computing)); MKRTCHYAN, Tigran (DESY-IT, Scientific Computing)

Presenter: MKRTCHYAN, Tigran (DESY-IT, Scientific Computing)