Contribution ID: **495**   Type: **Talk**

Monitoring large-scale dCache installations with storage events using Kafka streams

Monday 21 October 2024 17:27 (18 minutes)

DESY operates multiple dCache storage instances for multiple communities. As each community has different workflows and workloads, their dCache installations range from very large instances with more than 100 PB of data, to instances with up to billions of files or instances with significant LAN and WAN I/O.

To successful operate all instances and quickly identify issues and performance bottlenecks, DESY IT relies for monitoring heavily on dCache own storage events. Each atomic operation in the distributed storage instances trigger a storage event with details to the corresponding transfer or service status change.

These events are collected and parsed through an Apache Kafka event streaming bus. From the Kafka event stream, the events are aggregated in an Elastic Search+Lucene based database and search engine for on the fly operational diagnostics and analytics. Beyond day to day operations, an on demand Apache Spark cluster on top the National Analysis Facility at DESY is used for in detail analyses of operational data to extract information over a wide time span and number of storage events. In a similar fashion, all dCache logging messages are also processed through Kafka stream allowing to employ a passive monitoring waiting for specific signature to raise an alarm. In the future ML and AI algorithms for predictive maintenance are in the development pipeline. Furthermore, additional matrices are collecting from the dCache pools themselves and also pushed to Kafka to generate an almost complete picture of the dCache instances.

In this talk, we present our aggregation and analyses pipelines and workflows and how they are enabling DESY IT to scale out dCache storages for heterogeneous user groups and use cases.

Primary authors: VOSS, Christian; SAHAKYAN, Marina; Mr MKRTCHYAN, Tigran (DESY)

Co-authors: TRAUTSCH, Alexander (DESY); SPERL, Christian (DESY); CHRISTIANS, Felix (DESY); HAN-NAPPEL, Juergen (DESY It); GRIZZO, Sandro (DESY); HARTMANN, Thomas (Deutsches Elektronen-Synchrotron (DE))

Presenter: VOSS, Christian

Session Classification: Parallel (Track 7)

Track Classification: Track 7 - Computing Infrastructure