

Automatised Event-Workflows with Offloading to Batch Systems

Thursday, 27 August 2020 16:00 (30 minutes)

DESY provides significant storage and computing resources with more than 30PB of data about 50 000 cores to its users. It is one of the largest sites in the High-Energy computing Grid and provides the computing and storage infrastructure to the European XFEL as well as local experiments. As with such a the large user base, DESY's goal is to provide an easy and efficient access to the resources to enable user workflows.\newline Established workflow management systems are mostly polling-like, so that they regularly sample the states of the connected systems and initiate further steps based on the current sampling.

We present our push-based approach to enable scalable workflow chains based on events and functions. Here events are flowing on an Apache Kafka/Confluent message backbone and can trigger predefined functions in an OpenWhisk framework. With dCache storage events as prime example, this allows for automatic processing chains, where file updated or newly written to the storage system initiates its own processing by a predefined function.

However, as lambdas in Function-As-A-Service (FaaS) frameworks are intended to be low-latency and fast result operations, computational or I/O intensive processing jobs need to be run on better suited systems. Thus, we demonstrate how to further offload such computational heavy workloads to our HTCondor batch system by function chains.\newline

This enables us to interlink our storage and computing resources even closer.

Due to the generic approach, such automatised workflow chains are not limit to storage events but can be extended to other event types as well.

Primary author: Dr HARTMANN, Thomas (DESY)

Co-authors: Dr VOSS, Christian (DESY Hamburg); Mr HANNAPPEL, Juergen (DESY); SAHAKYAN, Marina (DESY); Mr SCHUH, Michael (DESY); Dr FUHRMANN, Patrick (DESY/dCache.org)

Presenter: Mr SCHUH, Michael (DESY)

Session Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators Session

Track Classification: Converging High Performance infrastructures: Supercomputers, clouds, accelerators