



Contribution ID: 500

Type: Oral

## Scalable processing for storage events and automation of scientific workflows

*Monday, November 4, 2019 3:00 PM (15 minutes)*

Low latency, high throughput data processing in distributed environments is a key requirement of today's experiments. Storage events facilitate synchronisation with external services where the widely adopted request-response pattern does not scale because of polling as a long-running activity. We discuss the use of an event broker and stream processing platform (Apache Kafka) for storage events, with respect to automatised scientific workflows starting from file system events (dCache, GPFS) as triggers for data processing and placement.

In a brokered delivery, the broker provides the infrastructure for routing generated events to consumer services. A client connects to the broker system and subscribes to streams of storage events which consist of data transfer records for files being uploaded, downloaded and deleted. This model is complemented by direct delivery using W3C's Server-Sent Events (SSE) protocol. We also address the shaping of a security model, where authenticated clients are authorised to read dedicated subsets of events.

On the compute side, the messages feed into event-driven work-flows, either user supplied software stacks or solutions based on open-source platforms like Apache Spark as analytical framework and Apache OpenWhisk for Function-as-a-Service (FaaS) and more general computational microservices. Building on cloud application templates for scalable analysis platforms, desired services can be dynamically provisioned on DESY's on-premise OpenStack cloud as well as in commercial hybrid cloud environments. Moreover, this model supports also the integration of data management tools like Rucio to address data locality e.g. to move files subsequent to processing by event-driven work-flows.

### Consider for promotion

No

**Primary authors:** HANNAPPEL, Jurgen Manfred (Deutsches Elektronen-Synchrotron DESY); STAREK, Jürgen (Deutsches Elektronen-Synchrotron DESY); SCHUH, Michael (Deutsches Elektronen-Synchrotron DESY); FUHRMANN, Patrick (Deutsches Elektronen-Synchrotron DESY); MILLAR, Paul (Deutsches Elektronen-Synchrotron DESY); HARTMANN, Thomas (Deutsches Elektronen-Synchrotron DESY)

**Presenter:** SCHUH, Michael (Deutsches Elektronen-Synchrotron DESY)

**Session Classification:** Track 3 – Middleware and Distributed Computing

**Track Classification:** Track 3 – Middleware and Distributed Computing