



Data Management for extreme scale computing

XFEL Status All Hands Meeting



Jürgen Starek
juergen.starek@desy.de

Use Case Goals

- ✕ Data creator: European XFEL
 - ▢→ High-performance X-Ray Light Source
 - ▢→ extremely high data rates (bursty; between 60 MB/s and 30 GB/s)
 - ▢→ Requirements:
 - ▢→ low-latency local calibration,
 - ▢→ followed by shipping calibrated data to remote sites
 - ▢→ Second analysis site Kurchatov Institute, Moscow
- ✕ Architecture provider and XDC partner: DESY IT
- ✕ Overarching goals:
 - ▢→ Storage events
 - ▢→ Quality of Service on the dCache storage system

Use Case Architecture

- ✕ Create flexible storage event handling for dCache
 - ▢▢▢▢➔ Perspective: Storage activity triggers both data management and processing tasks
 - ▢▢▢▢➔ Actual data movement through Indigo PaaS Orchestrator, Rucio
 - ▢▢▢▢➔ Technology: Kafka streams, Server-Sent Events (SSE)
- ✕ Function-as-a-Service architecture demonstrating event-driven actions
 - ▢▢▢▢➔ Events consumed by OpenWhisk, triggering (for now) processing tasks
 - ▢▢▢▢➔ FaaS as both demo and technology preview

FaaS demonstrator

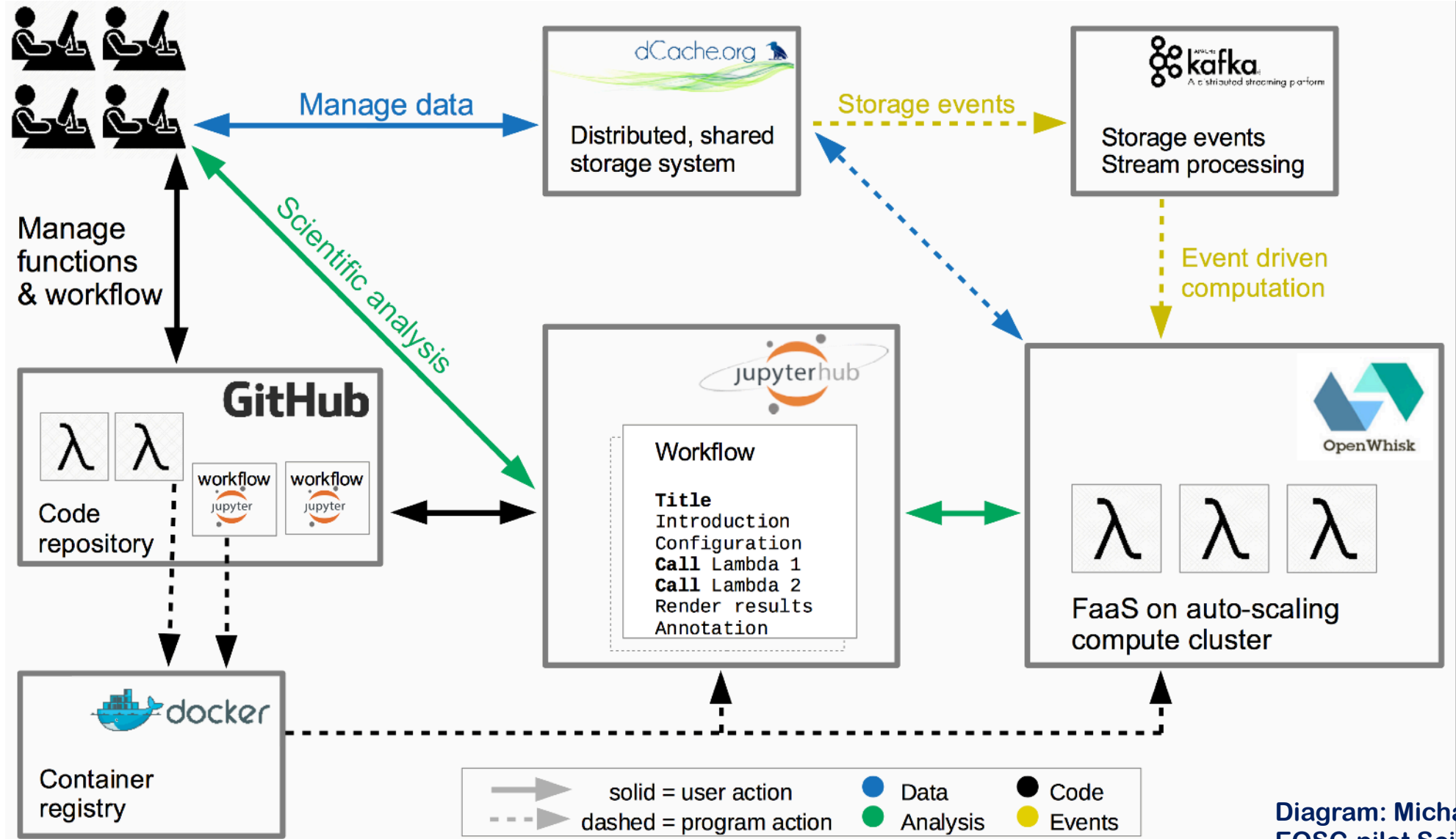


Diagram: Michael Schuh (DESY)
EOSC-pilot Science Demonstrator

Use Case Update

- ✕ Implementation status: On track with development of events framework and FaaS platform
- ✕ WRT Best Practices (D2.3):
 - ▬→ XFEL uses HDF5 data containers, (own) metadata catalogue, DOIs
 - ▬→ Data management plan (DMP) coming
 - ▬→ DESY remains active in RDA, EOSC-pilot
- ✕ FaaS architecture under development, nearing completion.
- ✕ Testbed: Interoperability tests between CERN, INFN, DESY products (dCache, EOS, Xcache, FTS)

Potential Stoppers




- ✗ We depend on the Orchestrator being ready in time
 - ▢▢▢▢➔ but can use our current alternative architecture if needed
- ✗ DESY will present architecture to XFEL and its remote partners as a proposal
 - ▢▢▢▢➔ which are not XDC members, acceptance is at their discretion

Workplan

- ✗ Provide FaaS demo for DI4R
- ✗ Integration with PaaS Orchestrator and XDC testbed
- ✗ Storage events (e.g. Kafka) for dCache across all relevant components
- ✗ Report and workshop „XDC Solutions for EOSC Research Communities“ (D2.7 and 2.8)

Potential Demo

FaaS demo:

-  Incoming files trigger storage events
-  event routing
-  dynamic processing on scalable cloud system
(Openstack / Openwhisk)