

Trainings

## Real-time Architectures with Spark and Kafka



Training course on the architectural principles involved in operating big data systems

Dauer: 2 days

Zielgruppe: Software Architects, Software Developer



***The training sessions are usually held in German. Please contact us if you are interested in training sessions in English.***

Training course on the architectural principles involved in operating big data systems

With the number of use cases increasing, big data systems are now facing the challenge of processing data streams instead of files. As this circumstance necessitates transitioning from a batch processing to an event streaming model, it often means reorganizing (or even redeveloping) the entire architecture of a big data system.

This training course examines the architectural principles needed to operate big data systems capable of processing large amounts of real-time data and making it highly available for queries. In order to explore these, course participants will use Spark and Kafka to set up a sample big data system capable of processing the Wikipedia edit stream, a real-time data stream containing every single edit made to every single Wikipedia article.

### **Agenda:**

- Event streams (brokers, topics, partitions in Kafka)
- Stream processing (transformations, processing patterns, error handling (at-least-once vs

exactly once)

- Offloading / archiving large volumes of data (Lambda Architecture, Flume, Kafka Connect, Camus/Gobblin)
- Storing analysis results (caches (HBase, Cassandra, Riak, Redis), dashboards (ES, Kibana), handling historical data)