

Inhouse Trainings

Real-time Architekturen mit Spark und Kafka

Zielgruppe: *Software-Entwickler, Software-Architekten*

Dauer: *2 Tage*

Abstract

Bei immer mehr Anwendungsfällen werden Big Data Systeme mit der Herausforderung konfrontiert, Datenströme an Stelle von Dateien zu verarbeiten. Dies erfordert oft eine Umstellung (oder Neuentwicklung) der gesamten Architektur eines Big Data Systems, da von einem Batch-Processing auf eine Event-Streaming Paradigma gewechselt werden muss.

In diesem Training werden die Architektur-Prinzipien beleuchtet, welche benötigt werden, um Big Data Systeme zu betreiben, die große, in Echtzeit anfallende Datenmengen verarbeiten und hochverfügbar für Anfragen zur Verfügung stellen können. Hierzu richten die Teilnehmer ein exemplarisches Big Data System mit Spark und Kafka ein, welches in der Lage ist den Stream der Wikipedia-„Edits“ (individuelle Bearbeitungen einzelner Artikel) in Echtzeit zu verarbeiten.

Agenda:

- Event Streams
 - o Broker, Topics, Partitionen in Kafka
- Stream Processing
 - o Transformationen
 - o Verarbeitungsmuster
 - o Fehlerbehandlung (at-least-once vs. exactly once)
- Offloading / Archivierung von großen Datenmengen
 - o Lambda Architektur

- o Flume, Kafka Connect, Camus/Gobblin
- Result store
 - o Caches (HBase, Cassandra, Riak, Redis)
 - o Dashboards (ES, Kibana)
 - o Umgang mit historischen Daten