



Big Data Training: Stream Analytics Basics

Was hat Stream Analytics mit Wasserpipelines auf dem Mars zu tun?

- Die Daten treffen in einem kontinuierlichen Strom rund um die Uhr ein.
- Die Verarbeitung und Analyse erfolgt in Echtzeit.
- Die Technologien sind neu.
- Stream Analytics Pipelines sind bereits auf der Erde erfolgreich im Einsatz.
- Die Technologien sind zukunftsorientiert und erfreuen sich wachsenden Interesses.

Data Streams sind unbegrenzt wachsende Files: Daten treffen rund um die Uhr ein und sollen Real-Time verarbeitet und analysiert werden. Stream Analytics und Stream Processing nicht mit herkömmlichen Technologien bewältigt werden. Neue Open Source Frameworks entstanden in den vergangenen Jahren und sind in vielen großen Unternehmen erfolgreich im Einsatz. Immer mehr kleinere und mittlere Unternehmen erkennen deren Potenzial zur Lösung drängender Herausforderungen.

Die Frameworks sind rasch heruntergeladen und auf einem Laptop installiert. Die APIs (Applications Programming Interfaces) sehen einfach aus, das erste Erfolgserlebnis stellt sich bald ein. Und damit kommen viele Fragen:

- Welche Tools und Frameworks zur Stream Analytics gibt es?
- Wodurch unterscheiden sich die einzelnen Frameworks?
- Welche Vor- und Nachteile haben sie?
- Wie können die Tools zu einer Pipeline kombiniert werden?
- Welche Fragen sollten in einer Evaluation der Stream Analytics Pipeline gestellt werden?
- Wie sieht der passende PoC (Proof of Concept) für meinen Anwendungsfall aus?

Dieses Training setzt bei diesen Fragen ein.

Die Teilnehmenden lernen die Grundlagen der Stream Analytics kennen und sind nach dem Kurs in der Lage, die Evaluation der Tools für ihren Anwendungsfall zu planen und einen Proof of Concept zielgerichtet durchzuführen.

Zielpublikum

Informatikerinnen und Informatiker, die

- ... sich für in Real-Time Stream Processing und Analytics interessieren und hinter die Kulissen der APIs blicken möchten;
- ... die ein Grundverständnis für die Komplexität der verteilten Big-Data-Systeme aufbauen möchten;
- ... die anhand eines soliden Verständnisses von Konzepten in der Lage sein wollen, selbständig Details zu vertiefen.

Lernziele

Die Teilnehmenden ...

- ... können erläutern, warum für das ausfallfreie Rechnen mit großen Datenbeständen verteilte Systeme eingesetzt werden;
- ... kennen die Trade-Offs, die verteilte Systeme mit sich bringen;
- ... kennen die wichtigsten Commit-Protokolle in verteilten Systemen und die Beschränkungen, die sich daraus ergeben;
- ... kennen die markantesten Architekturen verteilter Systeme;
- ... können die wichtigsten Open-Source-Frameworks für Real-Time Processing und Real-Time Analytics benennen und in Kategorien einteilen;
- ... kennen Herangehensweisen, wie einzelne Frameworks zu einer produktiven Pipeline kombiniert werden;
- ... haben Hands-On-Erfahrungen mit den APIs von Apache Kafka und Apache Spark und kennen die Herausforderungen bei der Implementierung der Exactly-Once-Semantik;
- ... kennen Windows und Watermarks als wichtige Konzepte bei der Real-Time Analytics;
- ... können das Vorgehen zur Evaluation einer Pipeline für Ihren spezifischen Use-Case planen und kennen die wichtigsten Fragen, die ihr PoC klären sollte.

Notwendige Vorkenntnisse

- Erfahrene Informatikerinnen und Informatiker, insbesondere Programmierinnen und Programmierer, Systemverantwortliche, Enterprise Architektinnen und Architekten.
- Grundkenntnisse Linux sind von Vorteil.
- Hands-On Übungen werden mit den Python-APIs durchgeführt.

Dozentin



Ursula Deriu führt an verschiedenen Fachhochschulen Vorlesungen zu den Themenkreisen Data Management, Big Data und Data Science durch. Sie ist Autorin des Buchs [Stream Analytics Pipelines](#).

Sprache

- Deutsch
- Englisch auf Anfrage
- Unterlage in deutscher Sprache

Form und Dauer

- Online Training auf einer geschlossenen Cloud-Plattform.
- In-House-Training auf Anfrage.
- Das Training dauert zwei Tage

Buchung und Daten

- Auskunft und Buchung über die Homepage von tirsus.com/stream-analytics-basics

