


New

## Big Data II - Praxis-Vertiefung

Data Engineering Schulung für moderne Datenpipelines mit Spark, Kafka, dbt und DuckDB

 Live Online oder Präsenz



2 Tage (14 Stunden)

Preis : 1.590,00 € netto  
1.892,10 € inkl. 19 % MwSt.  
Inhouse-Paket : Auf Anfrage

Nr.: 54602

Dieses Big Data Training zeigt Ihnen, wie moderne Datenplattformen praktisch aufgebaut, erweitert und betrieben werden. Im Mittelpunkt steht die Umsetzung realer Datenpipelines: Sie laden Daten aus verschiedenen Quellen, transformieren und modellieren diese strukturiert und lernen, wie Daten zuverlässig, skalierbar und nachvollziehbar bereitgestellt werden.

In dieser Data Engineering Schulung arbeiten Sie mit aktuellen Werkzeugen wie dlt, dbt, DuckDB, Apache Spark und Apache Kafka. Sie lernen, wie moderne ELT-Workflows entstehen, wie Datenqualität sichergestellt wird und wie sich Batch- und Streaming-Prozesse in einer produktionsnahen Umgebung verbinden lassen.

Das Seminar legt den Fokus auf praktische Handlungsfähigkeit. Sie arbeiten mit realistischen Musterdaten, entwickeln End-to-End-Datenpipelines und erhalten Vorlagen, Patterns und Best Practices, die Sie auf eigene Projekte übertragen können.

### Ihre Vorteile / Nutzen

#### Für Teilnehmende:

Sie können moderne Datenpipelines eigenständig planen, entwickeln, testen und stabil betreiben. Sie gewinnen Sicherheit im Umgang mit wichtigen Data-Engineering-Tools und verstehen, wie Spark, Kafka, dbt, DuckDB und dlt in realen Datenplattformen zusammenspielen.

#### Für Unternehmen:

Unternehmen profitieren von skalierbaren, wartbaren und effizient betriebenen Datenprozessen. Dadurch entsteht eine verlässliche Grundlage für Analytics, Reporting, KI-Anwendungen und datenbasierte Produkte.

Dieses Seminar eignet sich für alle, die Big Data praktisch anwenden und moderne Datenpipelines selbst entwickeln möchten. Jetzt Platz sichern und teilnehmen.

### Seminarziel

Sie entwickeln die Fähigkeit, moderne Datenpipelines eigenständig zu planen, zu implementieren und zu betreiben. Dabei verbinden Sie Data Engineering Best Practices mit praktischer Umsetzung in den Bereichen Datenaufnahme, Transformation, Datenqualität, Batch-Verarbeitung, Streaming und Performanceoptimierung.

### Lernziele - Sie lernen:

- Daten aus APIs, Datenbanken und Dateien strukturiert zu laden, zu transformieren und aufzubereiten.
- Moderne ETL- und ELT-Workflows zu entwerfen und in Data-Engineering-Projekten umzusetzen.
- dlt für Data Ingestion, Schema Evolution und inkrementelles Laden einzusetzen.
- dbt für Datenmodellierung, Tests, Dokumentation und Data Quality zu nutzen.
- DuckDB für lokale Analytics, schnelle Datenexploration und OLAP-Szenarien einzusetzen.
- Apache Spark für verteilte Batch-Verarbeitung zu verwenden und Performanceprobleme zu erkennen.
- Apache Kafka für Event Streaming und Echtzeit-Datenpipelines zu konfigurieren.
- Datenpipelines zu überwachen, zu testen und produktionsnah zu betreiben.
- Performance- und Kostenfaktoren in Datenplattformen zu bewerten.
- Best Practices für stabile, wartbare und nachvollziehbare Datenpipelines anzuwenden.

### Kompetenzen

Das Seminar stärkt Ihre technisch-operative Kompetenz im Data Engineering. Sie bauen praktische Fähigkeiten im Umgang mit Datenpipelines, ELT-Workflows, Datenmodellierung, Spark, Kafka, dbt und DuckDB auf.

Gleichzeitig entwickeln Sie ein systematisches Architekturverständnis und stärken Ihre praktische Entscheidungsfähigkeit beim Aufbau datengetriebener Prozesse. Dazu gehören auch analytisches Denken, Fehlersuche, Qualitätsbewusstsein und der strukturierte Transfer in reale Projekte.

## Wer sollte teilnehmen:

### Zielgruppe

Das Seminar richtet sich an Data Engineers, Data Scientists, Softwareentwickler:innen, Analytics Engineers sowie technisch affine Anwender:innen, die bereits Grundlagen im Big-Data-Umfeld besitzen und nun praxisorientierte Umsetzungskompetenz aufbauen möchten.

Besonders geeignet ist das Training für Teams, die Datenpipeline-Prozesse professionalisieren, Data-Engineering-Kompetenz aufbauen oder moderne Tools wie Spark, Kafka, dbt und DuckDB in der Praxis kennenlernen möchten.

### Voraussetzungen

Grundlagenwissen zu Big-Data-Architekturen und Datenkonzepten wird vorausgesetzt, zum Beispiel durch die Teilnahme an [54601 - Big Data I - Grundlagen](#)

Basiskonzepte in SQL sind hilfreich. Erste Erfahrung mit Python oder einer anderen Programmiersprache ist von Vorteil, aber nicht zwingend erforderlich.

## Trainingsprogramm

### Modul 1: Architektur & Modern Data Stack

- ELT vs. ETL: Paradigmen, die heute zählen
- Überblick über dlt, dbt, DuckDB, Spark & Kafka
- Rolle von Storage, Compute und Metadaten

### Modul 2: Daten laden & vorbereiten (Ingestion & Staging)

- Quellen: APIs, Datenbanken, Flat Files
- Einsatz von dlt zur Datenaufnahme
- Versionierung & Change Tracking

### Modul 3: Datenmodellierung & Transformation mit dbt

- Lakehouse & Medallion Modellierung
- Datenqualität, Tests & Dokumentation
- Best Practices für Wiederverwendungsfähigkeit

### Modul 4: Datenanalyse & Exploration mit DuckDB

- Lokale Query Engines effizient nutzen
- Performancevorteile gegenüber klassischen Tools
- Analysepipeline entwickeln

### Modul 5: Skalierung & Betrieb

- Batch-Verarbeitung mit Apache Spark
- Streaming mit Apache Kafka
- Performance, Monitoring & Kostenoptimierung

### Modul 6: Stabiler Betrieb & Takeaways

- Pipeline-Orchestrierung und Deployment
- Logging, Alerting & Fehlermanagement
- Best Practices & Transfer in den eigenen Kontext



### Schulungsmethode

Das Seminar kombiniert Hands-on Labs, Live Coding, Mini-Projekte, technische Impulse, Diskussion realer Anwendungsfälle sowie Best Practices und Troubleshooting-Sessions.

Der Fokus liegt klar auf praktischer Umsetzung: Die Teilnehmenden arbeiten mit vorbereiteten Übungen, realitätsnahen Daten und nachvollziehbaren

Patterns. Dadurch entsteht eine direkte Verbindung zwischen Schulungsinhalten und typischen Anforderungen aus Data-Engineering-Projekten.

## Hinweis

### **Hinweis für Online-Schulungen:**

Wir empfehlen daher, sich im Vorfeld auf die Schulung mit einem zusätzlichen Bildschirm vorzubereiten, um das bestmögliche Lernerlebnis zu gewährleisten.

## Häufig gestellte Fragen (FAQ)

### **Was ist der Unterschied zwischen Big Data I und Big Data II?**

Big Data I vermittelt Grundlagen, Begriffe und Architekturverständnis. Big Data II ist die praxisorientierte Vertiefung und konzentriert sich auf die konkrete Umsetzung moderner Datenpipelines.

### **Für wen eignet sich dieses Big Data Training?**

Das Training eignet sich für Data Engineers, Data Scientists, Softwareentwickler:innen, Analytics Engineers und technisch affine Fachkräfte, die moderne Datenpipelines praktisch entwickeln möchten.

### **Ist Big Data II eine Data Engineering Schulung?**

Ja. Das Seminar ist als praxisnahe Data Engineering Schulung aufgebaut. Es behandelt Datenaufnahme, Transformation, Datenmodellierung, Spark, Kafka, Datenqualität und produktionsnahe Pipeline-Prozesse.

### **Welche Tools werden im Seminar eingesetzt?**

Zum Einsatz kommen unter anderem dlt, dbt, DuckDB, Apache Spark und Apache Kafka. Je nach Übung werden zusätzlich typische Konzepte moderner Datenplattformen behandelt.

### **Ist das Seminar auch als Spark Kafka Schulung geeignet?**

Ja. Spark und Kafka werden als wichtige Bausteine moderner Datenplattformen behandelt. Sie lernen die grundlegenden Konzepte, typischen Einsatzszenarien und das Zusammenspiel mit Datenpipelines kennen. Das Seminar vermittelt jedoch keine tiefgehende Spark- oder Kafka-Administration, sondern zeigt, wie beide Technologien im Kontext von Big Data und Data Engineering sinnvoll eingeordnet und praktisch genutzt werden.

### **Was lerne ich über Datenpipelines?**

Sie lernen, Daten aus verschiedenen Quellen zu laden, zu transformieren, zu modellieren, zu testen und bereitzustellen. Dabei entwickeln Sie ein Verständnis für robuste und wartbare End-to-End-Datenpipelines.

### **Benötige ich Programmiererfahrung?**

Basiskenntnisse in SQL sind hilfreich. Erste Erfahrung mit Python oder einer anderen Programmiersprache ist von Vorteil, aber nicht zwingend erforderlich.

### **Wird eine Schulungsumgebung bereitgestellt?**

Ja. Für das Seminar steht eine vorbereitete Schulungsumgebung zur Verfügung. Dadurch kann der Fokus auf den Übungen und den Data-Engineering-Konzepten liegen.

### **Kann das Seminar als Inhouse-Schulung durchgeführt werden?**

Ja. Das Seminar kann als Inhouse-Training durchgeführt und bei Bedarf auf Zielgruppe, Toollandschaft oder konkrete Unternehmensszenarien angepasst werden.

### **Welches Seminar folgt nach Big Data II?**

Als nächster Schritt eignet sich [54603 - Big Data III - Advanced Training](#), wenn Sie Themen wie Streaming-Architekturen, DataOps, Governance und produktionsreifen Betrieb vertiefen möchten.



## Termine und Orte



Nur noch wenige freie Plätze



Garantietermin

### Live Online Training

8. Okt 2026 bis 9. Okt 2026

10. Dez 2026 bis 11. Dez 2026