


New

Big Data III - Advanced Training

Streaming Architekturen, Lakehouse, DataOps und MLOps produktiv einsetzen

 Live Online oder Präsenz



1 Tag (7 Stunden)

Preis : 890,00 € netto
1.059,10 € inkl. 19 % MwSt.
Inhouse-Paket : Auf Anfrage

Nr.: 54603

Dieses Big Data Advanced Training richtet sich an erfahrene Data Engineers, Data Architects, DevOps Engineers und Plattformteams, die moderne Datenplattformen produktionsreif betreiben, skalieren und weiterentwickeln möchten.

Im Mittelpunkt stehen Streaming Architekturen, Lakehouse-Konzepte, DataOps-Prozesse und MLOps-Strategien. Sie lernen, wie Echtzeitdaten zuverlässig verarbeitet, Datenpipelines automatisiert betrieben und moderne Datenplattformen sicher, skalierbar und wartbar gestaltet werden. Anhand praxisnaher Szenarien entwickeln Sie ein tiefes Verständnis für produktive Big-Data-Architekturen. Sie gewinnen Sicherheit bei Architekturentscheidungen, Governance-Fragen und dem stabilen Betrieb komplexer Datenplattformen.

Ihre Vorteile / Nutzen

Für Teilnehmende:

Sie können moderne Big-Data-Plattformen mit Fokus auf Streaming, Lakehouse, DataOps und MLOps sicher planen, bewerten und produktionsnah betreiben. Sie lernen, wie Datenpipelines automatisiert, überwacht und stabil weiterentwickelt werden.

Für Unternehmen:

Unternehmen profitieren von robusteren Datenplattformen, höherer Datenqualität, klareren Governance-Strukturen und besser beherrschbaren Betriebsprozessen. So entsteht eine tragfähige Grundlage für Analytics, KI, Machine Learning und datengetriebene Geschäftsmodelle. Dieses Seminar richtet sich an Teilnehmende, die Big-Data-Plattformen professionell skalieren und stabil betreiben möchten.

Jetzt Platz sichern und teilnehmen.

Trainingsziel:

Sie erwerben die Fähigkeit, moderne Big-Data-Plattformen stabil, sicher und skalierbar zu betreiben. Dabei vertiefen Sie Ihre Kenntnisse in Streaming Architekturen, Lakehouse-Architekturen, DataOps, MLOps, Governance und produktionsreifen Betriebsmodellen.

Lernziele - Sie lernen:

- Lakehouse-Architekturen im Unternehmenskontext zu bewerten und einzusetzen.
- Moderne Table-Formate wie Delta Lake, Apache Iceberg und Apache Hudi einzuordnen.
- Streaming Architekturen für Echtzeitdaten zu planen, zu bewerten und zu betreiben.
- Change Data Capture mit Tools wie Debezium in Datenpipelines zu integrieren.
- Apache Kafka für Event Streaming und Realtime-Szenarien professionell einzuordnen.
- Airflow oder Dagster für Orchestrierung, Monitoring und Betrieb einzusetzen.
- Data Contracts, Data Lineage und Governance-Modelle in Datenplattformen zu berücksichtigen.
- DataOps- und MLOps-Grundprinzipien für produktive Datenplattformen anzuwenden.
- Betriebsrisiken, Skalierungsgrenzen und typische Fehlerquellen frühzeitig zu erkennen.
- Eigene Architekturideen auf den Unternehmenskontext zu übertragen.

Kompetenzen

Sie vertiefen technische Kompetenzen im Bereich Streaming, Lakehouse-Architekturen, Orchestrierung, DataOps und MLOps. Gleichzeitig stärken Sie Ihre Fähigkeit, produktive Datenplattformen strategisch zu bewerten und technologisch fundierte Entscheidungen zu treffen.

Darüber hinaus entwickeln Sie Betriebs-, Governance- und Architekturkompetenzen, die für skalierbare, wartbare und sichere Datenplattformen entscheidend sind. Dazu gehören analytisches Denken, Qualitätsbewusstsein, Risikoeinschätzung und der strukturierte Transfer in reale Datenprojekte.

Wer sollte teilnehmen:

Zielgruppe

Das Seminar richtet sich an erfahrene Data Engineers, Data Architects, DevOps Engineers, Softwareentwickler, Tech Leads sowie technisch erfahrene Mitarbeitende, die Big-Data-Pipelines und Datenplattformen skalieren, automatisieren und professionell betreiben möchten.

Besonders geeignet ist das Training für Teams, die bereits Datenpipelines aufgebaut haben und nun Themen wie Streaming Architekturen, Lakehouse, DataOps, MLOps, Governance und Produktionsbetrieb vertiefen möchten.

Voraussetzungen

Sichere Grundlagen im Big-Data-Umfeld und erste praktische Erfahrung mit Datenpipelines werden vorausgesetzt.

Empfohlen wird die Teilnahme an [54602 - Big Data II - Praxis-Vertiefung](#) oder vergleichbare Kenntnisse in Data Engineering, SQL, Python, Spark, Kafka oder modernen Datenplattformen.

Trainingsprogramm

Modul 1: Lakehouse & Architekturentscheidungen

- Lakehouse & Medallion Layering
- Storage-Formate (Delta, Iceberg, Hudi) im Vergleich
- Governance- und Compliance-Anforderungen

Modul 2: Echtzeitdaten & Streaming-Architekturen

- Event Streaming mit Apache Kafka
- Topics, Partitionierung, Rebalancing
- Streaming vs. Microbatching

Modul 3: Change Data Capture & Realtime Synchronisation

- Einsatz von Debezium & Log-basierten CDC-Verfahren
- Umgang mit Update/Deletes in Streaming-Systemen
- Data Quality & Validierung in Streaming-Umgebungen

Modul 4: Orchestrierung & Betrieb

- Airflow / Dagster Workflows modellieren
- Monitoring, Alerting, Logging & SLA-Konzepte
- Fehlertoleranz & Wiederanfahen von Pipelines

Modul 5: DataOps & Governance

- Rollen, Verantwortlichkeiten & Betriebskonzepte
- Data Catalogues, Metadata & Data Lineage
- Data Contracts für unternehmensweite Standards

Modul 6: MLOps - KI-Modelle produktiv betreiben

- Versionierung, Deployment & Retraining-Zyklen
- Modellüberwachung und Drift-Erkennung
- Betriebsmodelle für KI in Data Plattformen

Häufig gestellte Fragen (FAQ)

Für wen eignet sich dieses Big Data Advanced Training?

Das Seminar eignet sich für erfahrene Data Engineers, Data Architects, DevOps Engineers, Tech Leads und Plattfornteam, die moderne Big-Data-Plattformen produktionsreif betreiben und skalieren möchten.

Wie unterscheidet sich Big Data III von Big Data II?

Big Data II fokussiert auf Aufbau und praktische Umsetzung moderner Datenpipelines. Big Data III geht einen Schritt weiter und behandelt Skalierung, Streaming Architekturen, Lakehouse, DataOps, MLOps, Governance und produktionsreifen Betrieb.

Ist das Seminar auch als Streaming Architekturen Seminar geeignet?

Ja. Streaming Architekturen sind ein zentraler Bestandteil des Seminars. Behandelt werden unter anderem Kafka, Event Streaming, Realtime-Datenverarbeitung, CDC und Architekturentscheidungen für Echtzeitdaten.

Was ist ein Lakehouse und warum ist es wichtig?

Ein Lakehouse kombiniert die Flexibilität eines Data Lakes mit Analyse-, Governance- und Qualitätsfunktionen klassischer Data-Warehouse-Ansätze. Moderne Datenplattformen nutzen dafür häufig Technologien wie Delta Lake, Apache Iceberg oder Apache Hudi.

Welche Rolle spielen DataOps und MLOps in diesem Seminar?

DataOps unterstützt den stabilen und automatisierten Betrieb von Datenpipelines. MLOps überträgt ähnliche Prinzipien auf Machine-Learning-Modelle, etwa durch Versionierung, Deployment, Monitoring und Retraining.

Welche Technologien werden im Seminar behandelt?

Je nach Übung und Schulungsumgebung werden unter anderem Apache Kafka, Debezium, Airflow oder Dagster, Delta Lake, Apache Iceberg, Apache Hudi sowie grundlegende MLOps-Werkzeuge und Konzepte behandelt.

Benötige ich Erfahrung mit Spark, Kafka oder dbt?

Erfahrung mit modernen Datenpipelines ist hilfreich. Empfohlen wird die vorherige Teilnahme an Big Data II oder vergleichbare Praxis in Data Engineering, SQL, Python, Spark, Kafka oder modernen Datenplattformen.

Wird eine Schulungsumgebung bereitgestellt?

Ja. Für das Seminar steht eine vorbereitete Schulungsumgebung zur Verfügung. Dadurch können sich die Teilnehmenden auf die Inhalte, Übungen und Architekturentscheidungen konzentrieren.

Ist das Seminar auch für Cloud- und Hybrid-Umgebungen relevant?

Ja. Die behandelten Konzepte sind auf lokale, hybride und cloudbasierte Datenplattformen übertragbar. Architektur- und Betriebsfragen werden technologieübergreifend betrachtet.

Kann das Seminar als Inhouse-Schulung durchgeführt werden?

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



Termine und Orte



Nur noch wenige freie Plätze



Garantietermin

Live Online Training

6. Nov 2026

18. Dez 2026