Data Engineering auf Cloud-Plattformen

Sonstiges · Anbieter: NEXIO Online Akademie GmbH

FÖRDERUNG & KOSTEN

Kurspreis: 0 €

Förderfähig über: Bildungsgutschein

AZAV-zertifizierter Bildungsträger — Förderung durch die Agentur für Arbeit möglich

Kursdaten auf einen Blick

Durchführung	Online
Lernformat	Zeitgebunden
Dauer	12 Wochen in Vollzeit
Unterrichtssprache	Deutsch
Kursstart	Startet jeden Monat
Vorkenntnisse	Grundlegende Kenntnisse
Abschluss	Internationales Zertifikat AWS (nach bestandener Prüfung) und trägerinternes Zertifikat bzw. Teilnahmebescheinigung

Kurzbeschreibung

In diesem Kurs erarbeitest Du die Grundlagen und Anwendungsbereiche des Data Engineering auf Cloud-Plattformen. Du steigst mit Cloud-Datenplattformen und Architekturmustern wie Data Lake, Data Warehouse und Lakehouse ein und vertiefst anschließend SQL für analytische Auswertungen. Du …

Bildungsträger & Kontakt

NEXIO Online Akademie GmbH

Am Herrschaftsweiher 11, 67071 Ludwigshafen am Rhein

Telefon: 0800 100 87 34

E-Mail: beratung@nexio-akademie.de

Web: https://nexio-akademie.de

Kurs online ansehen

01 Kursbeschreibung

In diesem Kurs erarbeitest Du die Grundlagen und Anwendungsbereiche des Data Engineering auf Cloud-Plattformen. Du steigst mit Cloud-Datenplattformen und Architekturmustern wie Data Lake, Data Warehouse und Lakehouse ein und vertiefst anschließend SQL für analytische Auswertungen. Du lernst, Daten aus Batch-, Streaming- und Change-Data-Capture-Quellen aufzunehmen, mit verteilten Verarbeitungsverfahren aufzubereiten und über ETL- sowie ELT-Strecken in modellierte Zielstrukturen zu überführen. Weitere Schwerpunkte sind Streaming-Verarbeitung mit verteilten Clustern, Datenmodellierung, Speicherstrategien, Datenqualität und Validierung sowie Datenlebenszyklus und Katalogisierung. Im weiteren Verlauf richtest Du den Pipeline-Betrieb samt Orchestrierung und Monitoring ein und arbeitest Sicherheit und Datengovernance auf Cloud-Plattformen aus. Den Abschluss bildet ein Praxisprojekt, in dem Du eine Ende-zu-Ende-Datenpipeline aus Aufnahme, Transformation, Modellierung, Qualitätssicherung, Orchestrierung und überwachter Auslieferung umsetzt.

02 Lehrplan & Module

1. Cloud-Datenplattformen

Du lernst den Aufbau von Cloud-Datenplattformen kennen, etwa Speicher-, Rechen- und Analysedienste, und vergleichst Bereitstellungsmodelle wie IaaS, PaaS und SaaS. So kannst du Plattformkomponenten für Datenprojekte gezielt auswählen.

2. Data Lake, Data Warehouse und Lakehouse

Du vergleichst die Architekturmuster Data Lake, Data Warehouse und Lakehouse hinsichtlich Datenstruktur, Abfrageleistung und Einsatzzweck. Dabei lernst du auch offene Tabellenformate kennen.

3. Datenaufnahme aus verschiedenen Quellen

Du bindest Datenquellen über Batch-Ladeprozesse, Streaming und Change Data Capture an. Zudem lernst du, mit APIs, Datenbanken und Dateiformaten wie CSV, JSON und Parquet umzugehen.

4. Datenverarbeitung und ETL/ELT

Du entwickelst ETL- und ELT-Prozesse zur Bereinigung, Transformation und Anreicherung von Daten. Dabei setzt du SQL und Python für wiederholbare Transformationen ein.

5. Verteilte Cluster- und Stream-Verarbeitung

Du lernst die Prinzipien verteilter Datenverarbeitung auf Clustern kennen, zum Beispiel mit Apache Spark. Außerdem verarbeitest du Datenströme nahezu in Echtzeit.

6. Datenmodellierung: Normalisierung und Dimensional Modeling

Du modellierst Daten nach den Regeln der Normalisierung und nach dimensionalen Ansätzen mit Fakten- und Dimensionstabellen. So entstehen Strukturen, die für Analysen und Reporting geeignet sind.

7. Datenspeicher und Speicherstrategien

Du vergleichst Objektspeicher, relationale Datenbanken und spaltenorientierte Speicher und wählst passende Speicherstrategien. Partitionierung und Komprimierung helfen dir, Leistung und Kosten zu optimieren.

8. Datenqualität und Validierung

Du definierst Regeln für Datenqualität und prüfst Vollständigkeit, Konsistenz und Plausibilität automatisiert. So erkennst du fehlerhafte Daten, bevor sie in Analysen einfließen.

9. Datenlebenszyklus, Katalogisierung und Kostensteuerun

Du steuerst Daten über ihren Lebenszyklus von der Erfassung bis zur Archivierung und Löschung. Mit Datenkatalogen dokumentierst du Datenbestände, und du lernst, Cloud-Kosten zu überwachen und zu senken.

10. Pipeline-Betrieb, Orchestrierung und Monitorin

Du orchestrierst Datenpipelines mit Workflow-Werkzeugen, planst Abhängigkeiten und richtest Monitoring und Alarmierung ein. Dadurch laufen Pipelines zuverlässig und Fehler werden schnell erkannt.

11. Sicherheit und Datengovernance auf Cloud-Plattformen

Du sicherst Datenplattformen mit Zugriffsrechten, Verschlüsselung und Netzwerkabschottung ab. Ergänzend lernst du Konzepte der Data Governance wie Verantwortlichkeiten, Klassifizierung und Datenschutzvorgaben kennen.

03 Enthaltene Zertifikate

Trägerzertifikat: Internationales Zertifikat AWS (nach bestandener Prüfung) und trägerinternes Zertifikat bzw. Teilnahmebescheinigung

04 Voraussetzungen

Vorkenntnisse

Niveau: Grundlegende Kenntnisse

Vorausgesetzt werden Deutschkenntnisse in Wort und Schrift auf Niveau B1 sowie Englischkenntnisse in Wort und Schrift auf Niveau B2 (Fachdokumentation). Erforderlich sind solide IT-Grundkenntnisse, sicherer Umgang mit dem Betriebssystem, Grundkenntnisse in Datenverarbeitung sowie ein Grundverständnis von Cloud Computing. Programmiererfahrung in einer höheren Sprache, idealerweise Python, ist von Vorteil.

05 Zielgruppen & Berufsperspektiven

- Arbeitssuchende

Mögliche Berufsbilder: Data Engineer, Dateningenieur, Cloud-Ingenieur

NÄCHSTER SCHRITT

Kursdetails online ansehen und Anbieter kontaktieren

NEXIO Online Akademie GmbH · Am Herrschaftsweiher 11, 67071 Ludwigshafen am Rhein
0800 100 87 34 · beratung@nexio-akademie.de · https://nexio-akademie.de

Online ansehen