Data Engineering on Google Cloud Platform

Google Cloud Partner

Data Engineering on Google Cloud Platform

Dieser vier Tage umfassende, von Experten geleitete Kurs bietet Ihnen eine praxisnahe Einführung in die Entwicklung und Umsetzung von Datenverarbeitungssystemen auf der Google Cloud Platform. Mit einer Mischung aus Präsentationen, Live-Demos und interaktiven Labs lernen Sie, wie Sie Datenverarbeitungssysteme entwerfen, End-to-End-Datenpipelines erstellen, Daten analysieren und maschinelles Lernen anwenden. Dabei stehen strukturierte, unstrukturierte und gestreamte Daten im Fokus, um Ihnen ein umfassendes Verständnis moderner Datenverarbeitung zu vermitteln.

Kursinhalt
- Introduction to Data Engineering
- Building a Data Lake
- Building a Data Warehouse
- Introduction to Building Batch Data Pipelines
- Executing Spark on Cloud Dataproc
- Serverless Data Processing with Cloud Dataflow
- Manage Data Pipelines with Cloud Data Fusion and Cloud Composer
- Introduction to Processing Streaming Data
- Serverless Messaging with Cloud Pub/Sub
- Cloud Dataflow Streaming Features
- High-Throughput BigQuery and Bigtable Streaming Features
- Advanced BigQuery Functionality and Performance
- Introduction to Analytics and AI
- Prebuilt ML model APIs for Unstructured Data
- Big Data Analytics with Cloud AI Platform Notebooks
- Production ML Pipelines with Kubeflow
- Custom Model building with SQL in BigQuery ML
- Custom Model building with Cloud AutoML

In diesem Kurs erwerben Sie die folgenden Fähigkeiten:

- Datenverarbeitungssysteme auf der Google Cloud Platform entwickeln
- Batch- und Streamingdaten durch die Implementierung von Autoscaling-Datenpipelines auf Cloud Dataflow verarbeiten
- Mit Google BigQuery Geschäftsinformationen aus extrem großen Datasets ableiten
- Modelle des maschinellen Lernens mit TensorFlow und Cloud ML trainieren, auswerten und mit ihnen Vorhersagen treffen
- Unstrukturierte Daten mit Spark und ML-APIs auf Cloud Dataproc nutzen
- Sofortige Informationsgewinnung aus Streamingdaten ermöglichen

Offizielle Google Cloud Unterlagen.

Zielgruppe
Dieser Kurs ist speziell für erfahrene Entwickler konzipiert, die Big-Data-Transformationen verantworten. Dazu gehören Aufgaben wie:

- Extraktion, Laden, Transformation, Bereinigung und Validierung von Daten.
- Entwicklung von Pipelines und Architekturen für die Datenverarbeitung.
- Erstellung und Wartung von Modellen für maschinelles Lernen und Statistik.
- Abfrage von Datensätzen, Visualisierung von Ergebnissen und Erstellung von Berichten.

Voraussetzungen
- Grundkenntnisse in gängigen Abfragesprachen wie SQL
- Kenntnisse in Datenmodellierung, Extraktion, Transformation und Ladeaktivitäten
- Kenntnisse im Entwickeln von Anwendungen mit einer gängigen Programmiersprache wie Python
- Vertrautheit mit maschinellem Lernen und/oder Statistik

Dieser Kurs im Web
Alle tagesaktuellen Informationen und Möglichkeiten zur Bestellung finden Sie unter dem folgenden Link: www.experteach.at/go/**GCDE**

Vormerkung
Sie können auf unserer Website einen Platz kostenlos und unverbindlich für 7 Tage reservieren. Dies geht auch telefonisch unter 06074 4868-0.

Garantierte Kurstermine
Für Ihre Planungssicherheit bieten wir stets eine große Auswahl garantierter Kurstermine an.

Ihr Kurs maßgeschneidert
Diesen Kurs können wir für Ihr Projekt exakt an Ihre Anforderungen anpassen.

Training		Preise zzgl. MwSt.
Termine in Deutschland	**4 Tage**	**€ 2.595,-**
Termine in Österreich	**4 Tage**	**€ 2.595,-**
Termine in der Schweiz	**4 Tage**	
Online Training	**4 Tage**	**€ 2.595,-**

Termin/Kursort — Kurssprache Deutsch

29.09.-02.10.26	HYBerlin	27.04.-30.04.27	HYOnline
29.09.-02.10.26	HYHamburg	20.07.-23.07.27	HYFrankfurt
29.09.-02.10.26	HYOnline	20.07.-23.07.27	HYOnline
29.09.-02.10.26	HYOnline	05.10.-08.10.27	HYBerlin
10.11.-13.11.26	ONOnline	05.10.-08.10.27	HYHamburg
26.01.-29.01.27	HYMünchen	05.10.-08.10.27	HYOnline
26.01.-29.01.27	HYOnline	05.10.-08.10.27	HYOnline
16.03.-19.03.27	HYOnline	23.11.-26.11.27	HYOnline
16.03.-19.03.27	HYWien	23.11.-26.11.27	HYZürich
27.04.-30.04.27	HYDüsseldorf		

Stand 23.06.2026

Google Cloud Platform

ExperTeach

Inhaltsverzeichnis

Data Engineering on Google Cloud Platform

Introduction to Data Engineering
Explore the role of a data engineer.
Analyze data engineering challenges.
Intro to BigQuery.
Data Lakes and Data Warehouses.
Demo: Federated Queries with BigQuery.
Transactional Databases vs Data Warehouses.
Website Demo: Finding PII in your dataset with DLP API.
Partner effectively with other data teams.
Manage data access and governance.
Build production-ready pipelines.
Review GCP customer case study.
Lab: Analyzing Data with BigQuery.

Building a Data Lake
Introduction to Data Lakes.
Data Storage and ETL options on GCP.
Building a Data Lake using Cloud Storage.
Optional Demo: Optimizing cost with Google Cloud Storage classes and Cloud Functions.
Securing Cloud Storage.
Storing All Sorts of Data Types.
Video Demo: Running federated queries on Parquet and ORC files in BigQuery.
Cloud SQL as a relational Data Lake.
Lab: Loading Taxi Data into Cloud SQL.

Building a Data Warehouse
The modern data warehouse.
Intro to BigQuery.
Demo: Query TB+ of data in seconds.
Getting Started.
Loading Data.
Video Demo: Querying Cloud SQL from BigQuery.
Lab: Loading Data into BigQuery.
Exploring Schemas.
Demo: Exploring BigQuery Public Datasets with SQL using INFORMATION_SCHEMA.
Schema Design.
Nested and Repeated Fields.
Demo: Nested and repeated fields in BigQuery.
Lab: Working with JSON and Array data in BigQuery.
Optimizing with Partitioning and Clustering.
Demo: Partitioned and Clustered Tables in BigQuery.
Preview: Transforming Batch and Streaming Data.

Introduction to Building Batch Data Pipelines,
EL, ELT, ETL.
Quality considerations.
How to carry out operations in BigQuery.
Demo: ELT to improve data quality in BigQuery.
Shortcomings.
ETL to solve data quality issues.

Executing Spark on Cloud Dataproc
The Hadoop ecosystem.
Running Hadoop on Cloud Dataproc.
GCS instead of HDFS.
Optimizing Dataproc.
Lab: Running Apache Spark jobs on Cloud Dataproc.

Serverless Data Processing with Cloud Dataflow
Cloud Dataflow.
Why customers value Dataflow.
Dataflow Pipelines.
Lab: A Simple Dataflow Pipeline (Python/Java).
Lab: MapReduce in Dataflow (Python/Java).
Lab: Side Inputs (Python/Java).
Dataflow Templates.
Dataflow SQL.

Manage Data Pipelines with Cloud Data Fusion and Cloud Composer
Building Batch Data Pipelines visually with Cloud Data Fusion.
Components.
UI Overview.
Building a Pipeline.
Exploring Data using Wrangler.
Lab: Building and executing a pipeline graph in Cloud Data Fusion.
Orchestrating work between GCP services with Cloud Composer.
Apache Airflow Environment.
DAGs and Operators.
Workflow Scheduling.
Optional Long Demo: Event-triggered Loading of data with Cloud Composer, Cloud Functions, Cloud Storage, and BigQuery.
Monitoring and Logging.
Lab: An Introduction to Cloud Composer.

Introduction to Processing Streaming Data
Processing Streaming Data.

Serverless Messaging with Cloud Pub/Sub
Cloud Pub/Sub.
Lab: Publish Streaming Data into Pub/Sub.

Cloud Dataflow Streaming Features
Cloud Dataflow Streaming Features.
Lab: Streaming Data Pipelines.

High-Throughput BigQuery and Bigtable Streaming Features
BigQuery Streaming Features.
Lab: Streaming Analytics and Dashboards.
Cloud Bigtable.
Lab: Streaming Data Pipelines into Bigtable.

Advanced BigQuery Functionality and Performance
Analytic Window Functions.
Using With Clauses.
GIS Functions.
Demo: Mapping Fastest Growing Zip Codes with BigQuery GeoViz.
Performance Considerations.
Lab: Optimizing your BigQuery Queries for Performance.
Optional Lab: Creating Date-Partitioned Tables in BigQuery.

Introduction to Analytics and AI
What is AI?.
From Ad-hoc Data Analysis to Data Driven Decisions.
Options for ML models on GCP.

Prebuilt ML model APIs for Unstructured Data
Unstructured Data is Hard.
ML APIs for Enriching Data.
Lab: Using the Natural Language API to Classify Unstructured Text.

Big Data Analytics with Cloud AI Platform Notebooks
Whats a Notebook.
BigQuery Magic and Ties to Pandas.
Lab: BigQuery in Jupyter Labs on AI Platform.

Production ML Pipelines with Kubeflow
Ways to do ML on GCP.
Kubeflow.
AI Hub.
Lab: Running AI models on Kubeflow.

Custom Model building with SQL in BigQuery ML
BigQuery ML for Quick Model Building.
Demo: Train a model with BigQuery ML to predict NYC taxi fares.
Supported Models.
Lab Option 1: Predict Bike Trip Duration with a Regression Model in BQML.
Lab Option 2: Movie Recommendations in BigQuery ML.

Custom Model building with Cloud AutoML
Why Auto ML?
Auto ML Vision.
Auto ML NLP.
Auto ML Tables.

ExperTeach Training & Consulting GmbH
Handelskai 94-96 · 1200 Wien · Telefon: +43 1 2350 383-0 · Fax: +43 1 2350 383-19 · info@experteach.at · www.experteach.at