

# Data Engineering on Google Cloud Platform

This four-day, expert-led course provides you with a hands-on introduction to developing and implementing data processing systems on the Google Cloud Platform. Through a mix of presentations, live demos and interactive labs, you will learn how to design data processing systems, create end-to-end data pipelines, analyze data and apply machine learning. The focus is on structured, unstructured and streamed data to give you a comprehensive understanding of modern data processing.

## Course Contents

- 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 this course, you will learn the following skills:

- Develop data processing systems on Google Cloud Platform
- Process batch and streaming data by implementing autoscaling data pipelines on Cloud Dataflow
- Derive business insights from extremely large datasets using Google BigQuery
- Train, evaluate, and make predictions with machine learning models using TensorFlow and Cloud ML
- Leverage unstructured data with Spark and ML APIs on Cloud Dataproc
- Enable instant insights from streaming data

Official Google Cloud documentation.

## Target Group

This course is specifically designed for experienced developers who are responsible for big data transformations. This includes tasks such as:

- Extracting, loading, transforming, cleansing and validating data.
- Development of pipelines and architectures for data processing.
- Creation and maintenance of models for machine learning and statistics.
- Query data sets, visualize results and create reports.

## Prerequisites

- Basic knowledge of common query languages such as SQL
- Knowledge of data modeling, extraction, transformation and loading activities
- Knowledge of developing applications with a common programming language such as Python
- Familiarity with machine learning and/or statistics

Status 12/13/2025

## This Course in the Web



You can find the up-to-date information and options for ordering under the following link:

[www.experteach-training.com/go/GCDE](http://www.experteach-training.com/go/GCDE)

## Reservation

On our Website, you can reserve a course seat for 7 days free of charge and in a non-committal manner. This can also be done by phone under +49 6074/4868-0.

## Guaranteed Course Dates

To ensure reliable planning, we are continuously offering a wide range of guaranteed course dates.

## Your Tailor-Made Course!

We can precisely customize this course to your project and the corresponding requirements.

Training		Prices, excl. of V.A.T.
<b>Classes in Germany</b>	<b>4 Days</b>	<b>€ 2,595</b>
<b>Classes in Austria</b>	<b>4 Days</b>	<b>€ 2,595</b>
<b>Classes in Switzerland</b>	<b>4 Days</b>	
<b>Online Training</b>	<b>4 Days</b>	<b>€ 2,595</b>
Date/course venue	Course language German 	
20/01-23/01/26  München	14/07-17/07/26  Online	
20/01-23/01/26  Online	29/09-02/10/26  Berlin	
10/03-13/03/26  Online	29/09-02/10/26  Hamburg	
10/03-13/03/26  Wien	29/09-02/10/26  Online	
27/04-30/04/26  Düsseldorf	29/09-02/10/26  Online	
27/04-30/04/26  Online	10/11-13/11/26  Online	
14/07-17/07/26  Frankfurt	10/11-13/11/26  Zürich	



# Table of Contents

## 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.

