

Updated 05/17/2024

Sign up

Training for GCP Professional Data Engineer certification

3 days (21 hours)

Presentation

Our GCP Professional Data Engineer training course is designed to equip you with the essential skills needed to create, manage and optimize large-scale data solutions on Google Cloud Platform.

During this training course, you will develop in-depth expertise in the field of data engineering, exploring in detail the key features and concepts of the GCP platform.

You'll learn how to apply this knowledge practically to design efficient, highly scalable data solutions.

This course will give you an in-depth understanding of fundamental concepts such as data storage, real-time processing and distributed data architectures.

You will be trained to set up an optimal development environment and master the deployment of data solutions on Google Cloud. Throughout the course, we will focus on the effective use of GCP tools and services specific to data engineering.

As with all our training courses, this one will introduce you to the [latest GCP resources](#).

Objectives

- Gain an in-depth understanding of data lifecycles, technical aspects and structure types
- Understanding GCP services and hosted technologies

- Learn how to design and operate efficient storage systems
- Develop expertise in the design of GCP data pipelines
- In-depth design and choice of data processing structure

Target audience

- Cloud professionals
- Data engineers

Prerequisites

Basic knowledge of data engineering and familiarity with cloud services would be an asset.

Our GCP Professional Data Engineer training program

Selection of GCP storage technologies

- Data Lifecycle (Ingest, Store, Process and Analyze, Explore and Visualize)
- Technical Aspects of Data: Volume, Velocity, Variation, Access, and Security
- Types of Structure
 - Structured
 - Semi-Structured
 - Unstructured
- Schema Design Considerations (Relational Design & NoSQL Design)

Building and operating storage systems

- Cloud SQL
- Cloud Spanner
- Cloud Bigtable
- Cloud Firestore
- BigQuery
- Cloud Memorystore
- Cloud Storage
- Unmanaged Databases

Designing data pipelines

- An overview of data pipelines (Steps / Data types)
- GCP pipeline components / services
- Migrating Hadoop and Spark to GCP

Processing / Data processing

- Design and choice of structure
- Distributed processing design
- Migrating a DataWarehouse

Processing infrastructure configuration

- Provisioning and tuning of processing resources (Compute Engine, Kubernetes, Cloud Bigtable, Dataproc)
- Supervision of processing resources (Stackdriver Monitoring, Stackdriver Logging, Stackdriver Trace)

Security / Administration on GCP

- Identity and access management with Cloud IAM
- Using IAM with storage and processing services
- Data Security
- Confidentiality and the Data Loss Prevention API

Database design for reliability, scalability and availability

- Bigtable database design (Scalability and Reliability)
- Designing Cloud Spanner databases
- BigQuery database design

Understanding data operations for greater flexibility and portability

- Catalog and discovery with Data Catalog
- Data processing with Data Prep
- Data visualization with Data Studio (Looker studio)
- Data mining with Cloud Datalab
- Orchestrating workflows with Cloud Composer

Deploying machine learning pipelines

- ML pipeline structure
 - Data Ingestion
 - Data Preparation
 - Data Segregation
 - Model Training
 - Model Evaluation
 - Model Deployment
 - Model Monitoring
- GCP options for ML pipeline deployment
 - Cloud AutoML
 - BigQuery ML
 - Kubeflow
 - Spark Machine Learning

Choosing the right training infrastructure

- Hardware gas pedals (GPUs, TPUs, CPUs)
- Distributed infrastructure and single machine
- Edge Computing with GCP (Cloud IOT, Edge TPU)

Supervision and alerts

- The three types of machine learning algorithms
- Deep Learning
- Machine learning model engineering
- Common sources of error in machine learning models

Use prebuilt models as a service

- Visio (Vision AI, Video AI)
- Conversation (Dialogflow, Cloud Text-to-Speech API, Cloud Speech-to-Text API)
- Language (Translation, Natural language)
- Structured data (Recommendations AI API, Cloud Inference API)

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or learning difficulties.

in-company security (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.