Course : Data Engineering on Google Cloud Platform

Official course, preparation for Google Cloud certification exams

Practical course - 4d - 28h00 - Ref. DGC
Price : 3860 € E.T.

With this course, you'll learn how to design and build data processing systems on Google Cloud Platform. Through extensive hands-on work, you'll learn how to design data processing systems, build end-to-end data pipelines, analyze data and perform machine learning. The course covers structured, unstructured and streaming data.

Teaching objectives

At the end of the training, the participant will be able to:

- Design and develop data processing systems on Google Cloud
- Implement self-calibrating batch and streaming pipelines with Dataflow
- Analyze very large volumes of data with BigQuery
- Processing unstructured data with Spark, Dataproc and ML APIs
- Exploring and using ML tools: BigQuery ML and Cloud AutoML

Intended audience

Experienced developers responsible for managing megadata transformations, including data extraction, loading, transformation, cleansing and validation.

Prerequisites

Google Cloud Big Data/ML or equivalent. Proficiency in SQL, modeling and ETL, Python development and basic statistics.

Certification

We recommend you take this course if you want to prepare for certification as a "Google Cloud Professional Data Engineer".
Comment passer votre examen ?

Practical details

Teaching methods

Training in French. Official course material in English.

PARTICIPANTS

Experienced developers responsible for managing megadata transformations, including data extraction, loading, transformation, cleansing and validation.

PREREQUISITES

Google Cloud Big Data/ML or equivalent. Proficiency in SQL, modeling and ETL, Python development and basic statistics.

TRAINER QUALIFICATIONS

The experts who lead the training courses are specialists in the subjects covered. They are approved by the publisher and certified for the course. They have also been validated by our teaching teams in terms of both professional knowledge and teaching skills for each course they teach. They have at least three to ten years of experience in their field and hold or have held positions of responsibility in companies.

ASSESSMENT TERMS

Assessment of targeted skills prior to training.
Assessment by the participant, at the end of the training course, of the skills acquired during the training course.
Validation by the trainer of the participant's learning outcomes, specifying the tools used: multiple-choice questions, role-playing exercises, etc.
At the end of each training course, ITTCERT provides participants with a course evaluation questionnaire, which is then analysed by our teaching teams. Participants also complete an official evaluation of the publisher.
An attendance sheet for each half-day of attendance is provided at the end of the training course, along with a certificate of completion if the participant has attended the entire session.

Course schedule

1 Introduction to data engineering

- Explore the role of a data engineer.
- Analyze the challenges of data engineering.
- Introduction to BigQuery.
- Data lakes and data warehouses.
- Demonstration "Federated Queries with BigQuery".
- Transactional databases versus data warehouses.
- Demonstration "Search for personal data in your dataset with the DLP API".
- Work effectively with other data teams.
- Managing data access and governance.
- Build production-ready pipelines.
- Case study of a Google Cloud Platform (GCP) customer.

Hands-on work
Data analysis with BigQuery.

2 Building a data lake

- Introduction to data lakes.
- Data storage and ETL options on GCP.
- Building a data lake with Cloud Storage.
- Demonstration: optimizing costs with Google Cloud Storage classes and cloud functions.
- Securing Cloud Storage.
- Store all types of data.
- Demonstration: execution of federated queries on Parquet and ORC files in BigQuery.
- Cloud SQL as a relational data lake.

Hands-on work
Load the Taxis DB into the SQL Cloud.

3 Building a data warehouse

- The modern data warehouse.
- Introduction to BigQuery.
- Demonstration: Terabit data requests in a matter of seconds.
- Data loading.
- Demonstration: querying Cloud SQL from BigQuery.
- Explore patterns.
- Exploring public BigQuery datasets with SQL using INFORMATION_SCHEMA.
- Schematic design.
- Nested and repeated fields.
- Nested and repeated fields in BigQuery.
- Optimize partitioning and clustering.
- Demonstration: partitioned and grouped tables in BigQuery.
- Batch and continuous data transformation.

Hands-on work
Load data using the console and CLI. Working with arrays and structures.

TEACHING AIDS AND TECHNICAL RESOURCES
The teaching resources used are the publisher's official materials and practical exercises.

TERMS AND DEADLINES
Registration must be completed 24 hours before the start of the training course.

ACCESSIBILITY FOR PEOPLE WITH DISABILITIES
Do you have specific accessibility requirements? Contact Ms FOSSE, disability advisor, at the following address: psh-accueil@orsys.fr so that we can assess your request and its feasibility.

4 Introduction to building batch data pipelines

- EL, ELT and ETL (Extract, Load and Transform) integration approaches.
- Quality considerations.
- How to perform operations in BigQuery.
- Demonstration: ELT to improve data quality in BigQuery.
- Gaps.
- ETL to solve quality problems.

5 Running Spark on Dataproc Cloud

- The Hadoop ecosystem.
- Running Hadoop on Cloud Dataproc GCS instead of HDFS.
- Optimize Dataproc.

Hands-on work
Run Apache Spark jobs on Cloud Dataproc.

6 Serverless data processing with Cloud Dataflow

- Cloud Dataflow.
- Why do customers appreciate Dataflow?
- Data flow pipelines.
- Dataflow templates.
- Dataflow SQL.

Hands-on work
Simple dataflow pipeline (Python/Java). MapReduce in a data stream (Python/Java). Side inputs (Python/Java).

7 Data pipeline management with Cloud Data Fusion and Cloud Composer

- Visual creation of batch data pipelines with Cloud Data Fusion.
- Orchestrate work between GCP services with Cloud Composer - Apache Airflow Environment - DAG and operators.
- Demonstration: event-driven data loading with Cloud Composer, Cloud Functions, Cloud Storage...
- Monitoring and logging.

Hands-on work
Building and running a pipeline graph in Cloud Data Fusion (components, user interface presentation, building a pipeline, data exploration using Wrangler). Use of Cloud Composer.

8 Introduction to streaming data processing

- Streaming data processing.

9 Serverless messaging with Cloud Pub/Sub

- Presentation of Cloud Pub/Sub.

Hands-on work
Publish streaming data in Pub/Sub.

10 Dataflow Cloud streaming features

- Cloud Dataflow streaming features.

Hands-on work
Continuous data pipelines.

11 BigQuery and Bigtable high-speed streaming features

- BigQuery streaming features.
- Cloud Bigtable.

Hands-on work
Continuous analysis and dashboards. Continuous data pipelines to Bigtable.

12 Advanced BigQuery features and performance

- Features "Analytic Window".
- Use of With clauses.
- GIS functions.
- Demonstration: mapping the fastest-growing zip codes with BigQuery GeoViz.
- Performance considerations.

Hands-on work
Optimize your BigQuery queries for performance. Create date-partitioned tables in BigQuery (optional).

13 Introduction to analytics and artificial intelligence

- What is artificial intelligence (AI)?
- From ad hoc data analysis to data-driven decisions.
- Options for machine learning (ML) models on Google Cloud Platform.

14 Predefined ML model APIs for unstructured data

- Unstructured data is difficult to use.
- ML API for data enrichment.

Hands-on work
Use the natural language application programming interface (API) to classify unstructured text.

15 Big Data Analytics with Cloud AI Platform notebooks

- What is a notebook?
- BigQuery Magic and links with Pandas.

Hands-on work
BigQuery in Jupyter Labs on IA Platform.

16 Machine learning production pipelines with Kubeflow

- Ways to do machine learning (ML) on Google Cloud Platform.
- Kubeflow AI Hub.
- Artificial Intelligence (AI) Hub.

Hands-on work
Using AI models on Kubeflow.

17 Creating custom templates with SQL in BigQuery ML

- BigQuery ML for rapid model building.
- Demonstration: training a model with BigQuery ML to predict cab fares in New York.
- Supported models.

18 Creating custom templates with Cloud AutoML

- Why AutoML?
- Auto ML Vision.
- Auto ML Natural Language Processing (NLP).
- Auto ML Tables.

Dates and locations

REMOTE CLASS
2026 : 26 May, 28 July, 29 Sep., 24 Nov.

PARIS LA DÉFENSE
2026 : 26 May, 28 July, 29 Sep., 24 Nov.