

Course : Build Batch Data Pipelines on Google Cloud

Official Course: Preparation for Google Cloud Certifications

Practical course - 1d - 7h00 - Ref. GBD

Price : 990 € E.T.

NEW

In this course, you'll learn how to design and optimize batch data pipelines on Google Cloud using Dataflow and Dataproc Serverless, while ensuring the quality, monitoring, and reliability of your processing workflows.

Teaching objectives

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

- ✓ Determine whether batch data pipelines are suitable for your use case.
- ✓ Design scalable batch pipelines to process large volumes of data.
- ✓ Implement data quality controls to ensure data integrity.
- ✓ Orchestrate, manage, and monitor batch pipelines, including error handling and monitoring.

Intended audience

Data engineers, data analysts.

Prerequisites

Basic knowledge of data warehousing and ETL/ELT, proficiency in SQL, programming skills (Python recommended), and familiarity with the GCloud CLI, the console, and the main Google Cloud services.

Course schedule

1 When to Choose Batch Pipelines

- Understanding the role of a data engineer.
- Understand the role of the components.
- Understanding the role and lifecycle of batch pipelines.

Hands-on work

Practice Description: Use Cases for Batch Pipelines, Processing, and Common Challenges.

PARTICIPANTS

Data engineers, data analysts.

PREREQUISITES

Basic knowledge of data warehousing and ETL/ELT, proficiency in SQL, programming skills (Python recommended), and familiarity with the GCloud CLI, the console, and the main Google Cloud services.

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.

2 Designing and Building Batch Pipelines

- Design of batch pipelines.
- Large-scale data processing.
- Dataflow and Serverless for Apache Spark.
- Data Connections and Orchestration.
- Running Apache Spark pipelines.
- Performance Optimization.

Hands-on work

Batch pipeline with Serverless for Apache Spark, batch pipeline with Dataflow Job Builder.

3 Orchestrate and oversee the pipelines

- Orchestration of batch processing.
- Cloud Composer.
- Observability and monitoring.
- Alerts and Troubleshooting.
- Visual pipeline management.

Hands-on work

Batch pipeline with Serverless for Apache Spark, batch pipeline with Dataflow Job Builder

4 Check Data Quality

- Data validation and cleaning.
- Error analysis (logs).
- Changes in Schemas.
- Integrity and deduplication.

Hands-on work

Data Quality Validation.

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.

This programme is an original creation, developed by the teaching teams at ORSYS Formation. Any reproduction, representation, adaptation or use, in whole or in part, without the prior written authorisation of ORSYS, is strictly prohibited. ORSYS reserves the right to take any action necessary to protect its intellectual property rights.

Dates and locations

REMOTE CLASS

2026 : 9 Dec.

2027 : 18 Mar., 23 June, 30 Sep., 2 Dec.

PARIS LA DÉFENSE

2027 : 18 Mar., 23 June, 30 Sep., 2 Dec.