Course : Data Integration with Cloud Data Fusion

Official course, preparation for Google Cloud certification exams

Practical course - 2d - 14h00 - Ref. DID
Price : 2060 € E.T.

With this training course, you'll get to grips with Cloud Data Fusion, a data integration platform that lets you rapidly create and manage data pipelines. You'll learn about the challenges of data integration and the need for a platform. You'll discover Cloud Data Fusion's components, how to handle batch and streaming data in real time thanks to visual pipeline design, in-depth metadata and data lineage tracking, and how to deploy data pipelines on different runtime engines.

Teaching objectives

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

- Identify data integration needs and use cases for Cloud Data Fusion
- Understand the capabilities and main components of the Cloud Data Fusion platform
- Design and execute batch and real-time processing pipelines
- Use Wrangler and connectors to transform and integrate multi-source data
- Configure runtime environment, supervise, troubleshoot and understand metadata and lineage

Intended audience

Data engineers, data analysts.

Prerequisites

Completion of the "Big Data and Machine Learning Fundamentals" course Ref. GCD or equivalent knowledge.

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 ?

PARTICIPANTS

Data engineers, data analysts.

PREREQUISITES

Completion of the "Big Data and Machine Learning Fundamentals" course Ref. GCD or equivalent knowledge.

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.

Practical details

Teaching methods

Training in French. Official course material in English and digital format. Good understanding of written English.

Course schedule

1 Introduction to data integration and Cloud Data Fusion

- Understand the need for data integration.
- List the situations/cases where data integration can help companies.
- List available data integration platforms and tools.
- Identify data integration challenges.
- Understand how to use Cloud Data Fusion as a data integration platform.
- Create a Cloud Data Fusion instance.
- Discover the basic framework and main components of Cloud Data Fusion.

2 Building pipelines

- Understanding Cloud Data Fusion architecture.
- Define a data pipeline.
- Understand the DAG representation of a data pipeline.
- Learn how to use Pipeline Studio and its components.
- Design a simple pipeline using Pipeline Studio.
- Deploy and run a pipeline.

3 Building complex pipelines

- Perform branching, merging and joining operations.
- Execute the pipeline with execution arguments using macros.
- Working with error handlers.
- Execute pre- and post-pipeline runs using actions and notifications.
- Plan pipeline execution.
- Import and export existing pipelines.

4 Pipeline execution environment

- Understand the composition of a runtime environment.
- Configure your pipeline's execution environment, logging and metrics.
- Understand concepts such as calculation profile and provisioner.
- Create a calculation profile.
- Create pipeline alerts.
- Monitor the running pipeline.

5 Building transformations and preparing data with Wrangler

- Understand how to use Wrangler and its main components.
- Transform data using Wrangler's user interface.
- Transform data using CLI directives/methods.
- Create and use user-defined directives.

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.

6 Connectors and streaming pipelines

- Understand data integration architecture.
- List the different connectors.
- Use the Cloud Data Loss Prevention (DLP) API.
- Understand the reference architecture of streaming pipelines.
- Build and run a streaming pipeline.

7 Metadata and data lineage

- List metadata types.
- Differentiate between commercial, technical and operational metadata.
- Understand data lineage.
- Understand the importance of maintaining data lineage.
- Differentiate between metadata and data lineage.

Dates and locations

REMOTE CLASS
2026 : 19 Mar., 11 June, 24 Sep., 3 Dec.

PARIS LA DÉFENSE
2026 : 19 Mar., 11 June, 24 Sep., 3 Dec.