AMBIENT-IT

+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated on 10/20/2026

Sign Up

Cloud Data Fusion Training

2 days (14 hours)

Overview

Cloud Data Fusion is a Google Cloud service that enables you to create visual, scalable data pipelines tailored to your business needs. With its design interface, connectors, and Wrangler for data preparation, it makes it easy to build ETL flows between various sources and your analytics platforms.

Our Cloud Data Fusion training will help you structure your data processing workflows and master the essential steps from designing to operating data pipelines.

You'll learn how to configure connections, clean data with Wrangler, and combine joins, filters, and aggregations to feed BigQuery from your files and repositories.

By the end of the training, you'll be able to deploy configurable pipelines, schedule their executions, monitor the quality of results, and diagnose errors using available logs and metrics.

This course also covers service accounts and access permissions, computing resources, and best practices for transferring data between environments to ensure the reliability of your data flows while keeping costs under control.

Like all our training courses, this one will introduce you to **the latest stable version** of the technology and its new features.

Objectives

- Design a visual ETL pipeline tailored to a specific data integration need
- Prepare and transform data with Wrangler to feed an analytics destination

- Configure the connections and service accounts required for processing
- Configure and schedule pipelines using macros and execution arguments
- Diagnose errors and adjust resources based on available logs and metrics

Target Audience

- Data Engineer
- ETL Developer
- Business Intelligence Consultant
- Data Architect

Prerequisites

- Proficiency in SQL for querying and aggregating data
- Knowledge of ETL processes and data integration principles
- Basic experience with Google Cloud and its management console
- Basic understanding of Cloud Storage and BigQuery for data storage and analysis

Technical Prerequisites

- Computer with 8 GB of RAM recommended and an up-to-date Google Chrome or Mozilla Firefox browser
- Stable internet connection allowing access to the Google Cloud Console and the training interface
- A personal account providing access to a dedicated Google Cloud project, with active billing and available compute quotas
- Accessible Cloud Data Fusion instance and necessary permissions for Cloud Storage and BigQuery Set up before the session
- Service accounts and access to Cloud Logging and Cloud Monitoring configured for the exercises

Our Cloud Data Fusion Training Agenda

[Day 1 - Morning]

Architecture and initial data flows

- Positioning Cloud Data Fusion within an integration architecture and distinguishing between batch and real-time processing
- Organizing instances, namespaces, and pipelines in the console and Studio
- Pipeline execution cycle and the role of Apache Spark in distributed processing
- Configuring connections, sources, and destinations using available plugins
- Distinction between service agents and execution accounts, access rights, and the principle of least privilege
- Hands-on workshop: Explore the prepared instance, configure a Cloud Storage source, and run a first load to BigQuery.

[Day 1 - Afternoon]

Data preparation and transformation

- Exploring a sample in Wrangler and the limitations of visual preparation for batch pipelines
- Building recipes with directives for cleaning, type conversion, and missing value handling
- Defining input and output schemas and verifying their consistency across stages
- Enriching data streams through joins, filters, deduplication, and aggregations
- Validating and previewing the pipeline, checking results, and handling invalid records
- Hands-on workshop: Clean a sales file, enrich it using a BigQuery repository, and verify the data loaded into a target table.

[Day 2 - Morning]

Configuring and automating pipelines

- Building reusable pipelines with macros based on plugin-compatible properties
- Organizing preferences and execution arguments, value inheritance and priorities
- Configuring source paths, target tables, and processing dates
- Deploying and scheduling processes, dependencies, and triggers based on pipeline status
- Transferring between environments via JSON export and import, plugin verification, and recreation of triggers
- Hands-on workshop: Configure the sales pipeline for multiple files, schedule its execution, and trigger a consolidation pipeline upon successful completion.

[Day 2 - Afternoon]

Monitoring and ensuring the reliability of processes

- Reviewing execution history and metrics to monitor volumes and durations
- Analyzing logs in Cloud Logging and monitoring metrics in Cloud Monitoring
- Diagnosing errors related to permissions, schemas, connections, and resources
- Adjusting compute profiles, parallelism, and resources based on volume and cost
- Recovery strategies, duplicate prevention, and cleanup of training resources
- Hands-on workshop: Diagnose a triggered failure, correct the pipeline, verify that it restarts without duplicates, and compare metrics before and after the adjustment.

Target Audience

This training is intended for both individuals and companies—large or small—that wish to train their teams in new, advanced IT technology or to acquire specific business knowledge or modern methods.

Entry-Level Assessment

The pre-training assessment complies with Qualiopi quality standards. Upon final registration, the learner receives a self-assessment questionnaire that allows us to evaluate their estimated proficiency level across various types of technologies, as well as their expectations and personal goals for the upcoming training, within the limits imposed by the selected format. This questionnaire also allows us to anticipate certain connection or internal security issues within the company (intra-company or virtual classroom) that could pose challenges for monitoring and ensuring the smooth running of the training session.

Teaching Methods

Hands-On Training: 60% Practical, 40% Theoretical. Training materials distributed in digital format to all participants.

Organization

The course alternates between theoretical input from the instructor—supported by examples and reflection sessions—and group work.

Assessment

At the end of the session, a multiple-choice quiz is administered to verify that participants have successfully acquired the skills.

Certification

A certificate will be issued to each trainee who has completed the entire training program.

Training organization registered under number 11 75 54743 75. This registration does not constitute state accreditation.
© Ambient IT 2015-2026. All rights reserved. Paris, France - Switzerland - Belgium - Luxembourg