Google Data Integration with Cloud Data Fusion

Duration: 2 Days Course Code: GO8334

Overview:

This 2-day course introduces learners to Google Cloud's data integration capability using Cloud Data Fusion.

In this course, we discuss challenges with data integration and the need for a data integration platform (middleware). We then discuss how Cloud Data Fusion can help to effectively integrate data from a variety of sources and formats and generate insights. We take a look at Cloud Data Fusion's main components and how they work, how to process batch data and real time streaming data with visual pipeline design, rich tracking of metadata and data lineage, and how to deploy data pipelines on various execution engines.

Updated 20/05/2026

Target Audience:

- Data Engineer
- Data Analysts

Objectives:

- **In this course participants will learn:**
- Identify the need of data integration
- Identify use cases for possible implementation with Cloud Data Fusion
- Design and execute batch and real-time data processing pipelines
- Use connectors to integrate data from various sources and formats
- Understand the relationship between metadata and data lineage
- Understand the capabilities Cloud Data Fusion provides as a data integration platform
- List the core components of Cloud Data Fusion
- Work with Wrangler to build data transformations
- Configure execution environment; Monitor and troubleshoot pipeline execution

Prerequisites:

- Completed "Introduction to Data Engineering"

Testing and Certification

None

Follow-on-Courses:

None recommended

Content:

01 Introduction

- Introduce the course objectives

02 Introduction to data integration and Cloud Data Fusion

- Understand the need for data integration
- List the situations/cases where data integration can help businesses
- List the available data integration platforms and tools
- Identify the challenges with data integration
- Understand the use of Cloud Data Fusion as a data integration platform
- Create a Cloud Data Fusion instance
- Familiarize with core framework and major components in Cloud Data Fusion

03 Building pipelines

- Understand Cloud Data Fusion architecture
- Define what a data pipeline is
- Understand the DAG representation of a data pipeline
- Learn to use Pipeline Studio and its components
- Design a simple pipeline using Pipeline Studio
- Deploy and execute a pipeline

04 Designing complex pipelines

- Perform branching, merging, and join operations
- Execute pipeline with runtime arguments using macros
- Work with error handlers
- Execute pre- and post-pipeline executions with help of actions and notifications
- Schedule pipelines for execution
- Import and export existing pipelines

05 Pipeline execution environment

- Understand the composition of an execution environment
- Configure your pipeline's execution environment, logging, and metrics. Understand concepts like compute profile and provisioner
- Create a compute profile
- Create pipeline alerts
- Monitor the pipeline under execution

06 Building Transformations and Preparing Data with Wrangler

- Understand the use of Wrangler and its main components
- Transform data using Wrangler UI
- Transform data using directives/CLI methods
- Create and use user-defined directives

07 Connectors and streaming pipelines

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

08 Metadata and data lineage

- List types of metadata
- Differentiate between business, technical, and operational metadata
- Understand what data lineage is
- Understand the importance of maintaining data lineage
- Differentiate between metadata and data lineage

09 Summary

- Review the course objectives ; concepts

Additional Information:

“Official course book provided to participants.”

Further Information:

For More information, or to book your course, please call us on 00 20 (0) 2 2269 1982 or 16142

training@globalknowledge.com.eg

www.globalknowledge.com/en-eg/

Global Knowledge, 16 Moustafa Refaat St. Block 1137, Sheraton Buildings, Heliopolis, Cairo