

Orchestrate BigQuery Workloads with Dataform

The best training on Dataform to orchestrate BigQuery workloads

1 jour / 7h

Learning outcomes

- Understand the components of Dataform core.
- Create tables and views in BigQuery using Dataform.
- Document BigQuery tables and views.
- Understand BigQuery security settings using Dataform.
- Use assertions to validate data in Dataform workflows.
- Execute Dataform SQL workflows in an automated fashion.

Who this course is for

Data Analysts, Data Engineers and everyone interested to orchestrate BigQuery Workloads with Dataform

Prerequisites

Knowledge of SQL data analysis and BigQuery as discussed in BigQuery for Data Analysis.

Course Outline

Module 01 : Dataform Core Components

Topics

- SQL workflow
- Repositories and workspaces
- Default files and folders
- Compiled graphs

Objectives

- Understand the components of Dataflow core.

Module 02 : Table Definitions and Dependencies

Topics

- Declare a data source.
- Create a table.
- Create an incremental table.
- Set partitioning and clustering options.
- Create an empty table.
- Create an external BigLake table.
- Create views and materialized views.
- Define dependencies

Objectives

- Create tables and views in BigQuery using Dataform.

Module 03 : Document BigQuery Tables and Views

Topics

- Use column descriptions.
- Use globally defined JavaScript constants.
- Add labels.

Objectives

- Document BigQuery tables and views.

Activities

- Lab: Build SQL Workflows with Dependencies in Dataform

Module 04: BigQuery Security Settings

Topics

- IAM dataset and table/view access
- Column-level security
- Row-level security

Objectives

- Understand BigQuery security settings using Dataform.

Module 05: Assertions

Topics

- Use built-in assertions.
- Create manual assertions.

Objectives

- Use assertions to validate data in Dataform workflows

Activities

- Lab: Work with Assertions and BigQuery Security Settings in Dataform

Module 06: SQL Workflow Executions

Topics

- Dataform code lifecycle.
- What happens during compilation.
- Customize and schedule compilation results.
- Execute workflows (UI, Cloud Scheduler, Cloud Composer).
- Logging and monitoring

Objectives

- Execute Dataform SQL workflows in an automated fashion.

Activities

- Lab: Automate and Monitor SQL Workflow Executions in Dataform

Module 07: Advanced Use Cases

Topics

- Create a BigLake table after file upload using Cloud Run functions.
- Build a Machine Learning pipeline with BigQuery ML.

- Work with Slowly Changing Dimensions Type 2.

Objectives

- Explore additional use cases for Dataform.

Activities

- Lab: Create a BigLake Table with Dataform Using Cloud Run Functions