

# Orchestrate BigQuery Workloads using Dataform

La meilleure formation pour prendre en main Dataform dans Google Cloud

1 jour / 7h

## Objectifs pédagogiques

- Comprendre les composants de Dataform core.
- Créer des tables et des vues dans BigQuery à l'aide de Dataform.
- Documenter les tables et les vues BigQuery.
- Comprendre les paramètres de sécurité BigQuery à l'aide de Dataform.
- Utiliser des assertions pour valider les données dans les workflows Dataform.
- Exécuter les workflows Dataform SQL de manière automatisée.

## Public Cible

Data Analysts, Data Engineers

Tous clients souhaitant orchestrer des flux de travaux SQL complexes pour transformer des données dans BigQuery.

## Prérequis

Connaissance de l'analyse de données SQL et de BigQuery

# Programme

## Module 01 : Composants Dataform Core

### Sujets

- Workflow SQL
- Référentiels et espaces de travail
- Fichiers et dossiers par défaut
- Graphes compilés

### Objectifs

- Comprendre les composants de Dataform Core

## Module 02 : Définitions de tables et dépendances

### Sujets

- Déclarer une source de données.
- Créer une table.
- Créer une table incrémentielle.
- Définir les options de partitionnement et de clustering.
- Créer une table vide.
- Créer une table BigLake externe.
- Créer des vues et des vues matérialisées.
- Définir les dépendances.

### Objectifs

- Créez des tables et des vues dans BigQuery à l'aide de Dataform.

## Module 03 : Documenter les tables et les vues BigQuery

### Sujets

- Utilisez des descriptions de colonnes.
- Utilisez des constantes JavaScript définies globalement.
- Ajoutez des étiquettes.

### Objectifs

- Documentez les tables et les vues BigQuery.

### Activités

- Lab : Créer des workflows SQL avec des dépendances dans Dataform

## **Module 04 : Paramètres de sécurité BigQuery**

### **Sujets**

- Accès aux jeux de données IAM et aux tables/vues
- Sécurité au niveau des colonnes
- Sécurité au niveau des lignes

### **Objectifs**

- Comprendre les paramètres de sécurité de BigQuery à l'aide de Dataform.

## **Module 05 : Assertions**

### **Sujets**

- Utiliser des assertions intégrées.
- Créer des assertions manuelles.

### **Objectifs**

- Utiliser des assertions pour valider les données dans les flux de travail Dataform.

### **Activités**

- Lab : Utilisation des assertions et des paramètres de sécurité BigQuery dans Dataform

## **Module 06 : Exécutions de workflow SQL**

### **Sujets**

- Cycle de vie du code Dataform.
- Ce qui se passe pendant la compilation.
- Personnaliser et planifier les résultats de la compilation.
- Exécuter des workflows (interface utilisateur, Cloud Scheduler, Cloud Composer).
- Journalisation et surveillance.

### **Objectifs**

- Exécutez les flux de travail Dataform SQL de manière automatisée.

### **Activités**

- Lab : automatiser et surveiller les exécutions de workflow SQL dans Dataform

## **Module 07 : Cas d'utilisation avancés**

## **Sujets**

- Créez une table BigLake après le téléchargement de fichier à l'aide de fonctions Cloud Run.
- Créez un pipeline de Machine Learning avec BigQuery ML.
- Travaillez avec des dimensions à évolution lente de type 2.

## **Objectifs**

- Découvrez des cas d'utilisation supplémentaires pour Dataform.

## **Activités**

- Lab : Créer une table BigLake avec Dataform à l'aide de fonctions Cloud Run