[sfeir] Institute

Google Cloud | GCP200DATAPLEX

Managing a Data Mesh with Dataplex

La formation pour apprendre à créer et gérer vos data mesh avec Dataplex

2 jours / 14h

Objectifs pédagogiques

- Identifier l'importance d'une plateforme de données moderne
- Configurer et installer Dataplex
- Sécuriser les lacs de données, les zones et les ressources
- Mettre en œuvre le balisage des ressources et utiliser des étiquettes pour rechercher des ressources
- Traiter les données à l'aide des tâches Dataplex
- Concevoir, exécuter et générer des rapports sur les processus de qualité des données

Modalités d'évaluation : Les objectifs pédagogiques sont évalués à travers la réalisation des parties pratiques (labs dirigés) sous la supervision du formateur délivrant la session de formation.

Public Cible

Toutes personnes intéressées par gérer, de surveiller et de gouverner les données et les artefacts d'IA dans les lacs de données, les entrepôts et les bases de données avec Dataplex

Prérequis

Avoir suivi le cours Data Engineering on Google Cloud (Jour 1 et 2) du parcours d'apprentissage Data Engineer ou avoir une expérience équivalente avec Google Cloud.

Programme

Module 01 : Introduction to Dataplex

Sujets

- Plateformes de données modernes et conception orientée données
- Piliers de la gouvernance des données
- Qu'est-ce que Dataplex ?
- Fonctionnalités de Dataplex
- Comparaison de Dataplex avec d'autres produits sur Google Cloud

Objectifs

- Identifier l'importance d'une plateforme de données moderne
- Expliquer le rôle de Dataplex sur Google Cloud

Module 02 : Creating a Data Mesh on Dataplex

Sujets

- Qu'est-ce qu'un datamesh ?
- Concepts de Dataplex
- Création de lacs et de zones de données
- Ressources dans Dataplex

Objectifs

- Définir les concepts clés de Dataplex
- Configurer et installer Dataplex

Activités

- Lab : Provisionner un maillage de données à l'aide de Dataplex

Module 03 : Processing Data on Dataplex

Sujets

- Traitement des données sur Dataplex
- Tâches de préparation des données
- Tâches d'ingestion
- Tâches Dataflow et Spark

Objectifs

- Comprendre les différentes options de traitement des données dans Dataplex
- Configurer et exécuter des tâches de préparation des données sur Dataplex

Activités

- Lab : Standardiser les données à l'aide des tâches Dataplex

Module 04 : Managing Data Security through Dataplex

Sujets

- Autorisations et rôles IAM
- Sécurisation de votre lac de données
- Gestion des politiques
- Sécurité des métadonnées

Objectifs

- Sécurisez les lacs de données, les zones et les ressources dans Dataplex

Activités

- Lab : Gérer la sécurité des données à l'aide de Dataplex

Module 05 : Data Tagging and Data Catalog

Sujets

- Introduction au catalogue de données
- Métadonnées techniques et métadonnées métier
- Etiquettes et templates d'étiquettes
- Entrées et groupes d'entrées
- Lignée de données

Objectifs

- Implémenter l'étiquettage des ressources et utiliser des étiquettes pour rechercher des assets

Activités

- Lab : Catalogue de données et lignée de données

Module 06 : Data Quality and Profiling

Sujets

- Tâches de qualité des données et AutoDQ
- Rapports sur la qualité des données
- Profilage des données

Objectifs

- Concevoir, exécuter et rendre compte des processus de qualité des données

Activités

- Lab : Qualité des données et profilage de vos données dans BigQuery

Module 07 : Dataplex Best Practices

Sujets

- Bonnes pratiques
- Démonstration de bout en bout

Objectifs

- Mettre en œuvre les meilleures pratiques pour Dataplex

Activités

- Lab : Gérer un maillage de données avec Dataplex