


Google Cloud Platform - Introduction au Data Engineering sur Google Cloud

Découvrez les bases du Data Engineering avec Google Cloud Platform

 Présentiel ou en classe à distance



1 jour (7 h)

Réf.: CC382

Cette formation **Google Cloud** dédiée à l'ingénierie des données **permet de développer des compétences clés pour concevoir, déployer et automatiser des pipelines de données performants**. Grâce à une approche pratique et orientée projet, les participants explorent les services essentiels de Google Cloud tels que **BigQuery, Dataflow, Dataproc, Cloud Composer** ou **Datastream**.

Idéale pour réussir ses projets de **data engineering dans le cloud**, cette formation donne les clés pour migrer, transformer et valoriser les données à grande échelle, tout en intégrant des solutions d'automatisation robustes. Elle s'adresse aux professionnels souhaitant maîtriser l'écosystème data de Google Cloud et améliorer la qualité, la fiabilité et la scalabilité de leurs flux de données.

A qui s'adresse cette formation ?



Pour qui

- Ingénieurs de données
- Administrateurs de bases de données
- Administrateurs système



Prérequis

- Bonne compréhension des bases de Google Cloud au niveau fondamental en utilisant Cloud Shell
- Maîtrise des fondamentaux du langage SQL,
- Expérience de la modélisation des données et des activités ETL (extraction, transformation, chargement)
- Expérience du développement d'applications à l'aide d'un langage de programmation courant tel que Python

Programme

1 - Tâches et composants de l'ingénierie des données

- Rôle d'un ingénieur de données
- Différences entre une source de données et un puits de données
- Identifier les principaux types de formats de données
- Solutions de stockage disponibles sur Google Cloud
- Se familiariser avec la gestion des métadonnées dans Google Cloud
- Partager facilement des ensembles de données à l'aide d'Analytics Hub
- Charger des données dans BigQuery via l'interface ou la CLI gcloud Atelier

Chargement de données dans BigQuery

2 - Réplication et migration des données

- Décrire l'architecture fondamentale de la réplication et de la migration de données sur Google Cloud
- Options et les cas d'utilisation de l'outil de ligne de commande gcloud
- Identifier les fonctionnalités et les usages types de Storage Transfer Service

- Fonctionnalités et cas d'utilisation de Transfer Appliance
- Fonctionnalités et le déploiement de Datastream
Atelier

Datastream : réplication PostgreSQL vers BigQuery (facultatif pour ILT)

Quiz

3 - Le modèle de pipeline Extract and Load

- Diagramme de base d'architecture d'extraction et de chargement
- Comprendre les fonctionnalités de l'outil en ligne de commande bq
- Fonctionnalités et les cas d'utilisation du service de transfert de données BigQuery
- Fonctionnalités et les cas d'utilisation de BigLake en tant que modèle de chargement sans extraction
Atelier

BigLake: Qwik Start

Quiz

4 - Le modèle de pipeline Extract, Load and Transform

- Architecture type d'un processus ELT (Extraction, Chargement, Transformation)
- Explorer un pipeline ELT standard dans l'environnement Google Cloud
- Fonctionnalités de programmation et de script SQL de BigQuery
- Comprendre les fonctionnalités clés et les usages de Dataform
Atelier

Créer et exécuter un workflow SQL dans Dataform

Quiz

5 - Modèle de pipeline Extract, Transform and Load

- Diagramme d'architecture : extraction, transformation et chargement de base
- Découvrir les outils visuels de conception de pipelines sur Google Cloud
- Expliquer le traitement par lots avec Dataproc
- Utiliser Dataproc Serverless pour les tâches Spark/ETL
- Identifier les options de traitement de données en streaming
- Comprendre la place de Bigtable dans les pipelines de données
Atelier

Utiliser Dataproc Serverless pour Spark pour charger BigQuery (facultatif pour ILT)

Créer un pipeline de données en continu pour un tableau de bord en temps réel avec Dataflow

Quiz

6 - Techniques d'automatisation

- Modèles d'automatisation pour orchestrer les pipelines de données
- Fonctionnalités de Cloud Scheduler et Workflows
- Découvrir Cloud Composer pour l'orchestration de pipelines de données
- Fonctions Cloud Run
- Identifier les cas d'usage de l'automatisation via Eventarc
Atelier

Utiliser les fonctions Cloud Run pour charger BigQuery (facultatif pour ILT)

Quiz



Les objectifs de la formation

- Identifier les rôles et responsabilités d'un data engineer sur Google Cloud
- Différencier les sources et récepteurs de données dans une architecture cloud

- Choisir les formats et solutions de stockage adaptés sur Google Cloud
- Charger des données dans BigQuery via la console ou l'interface en ligne de commande
- Configurer des pipelines de réplication avec Datastream pour la migration de données
- Mettre en oeuvre des techniques d'automatisation pour le traitement des données



Evaluation

- Pendant la formation, le formateur évalue la progression pédagogique des participants via des QCM, des mises en situation et des travaux pratiques. Les participants passent un test de positionnement avant et après la formation pour valider leurs compétences acquises.



Les points forts de la formation

- Une vision complète des rôles et responsabilités des data engineers
- Un parcours structuré autour des modèles ETL, ELT et streaming
- Des ateliers pratiques sur les données et les principaux composants utilisés sur Google Cloud