

## Introduction à l'ingénierie des données sur Google Cloud

**Durée: 1 Jour**    **Réf de cours: GO9093**    **Méthodes d'apprentissage: Intra-entreprise & sur-mesure**

### Résumé:

Dans ce cours, vous découvrirez l'ingénierie des données sur Google Cloud, les rôles et responsabilités des ingénieurs de données, et comment ceux-ci correspondent aux offres fournies par Google Cloud.

Vous apprendrez également des manières de relever les défis de l'ingénierie des données.

*Mis à jour 10/02/2026*

Formation intra-entreprise

Cette formation est délivrable en session intra-entreprise, dans vos locaux ou dans les nôtres. Son contenu peut être adapté sur-mesure pour répondre aux besoins de vos collaborateurs. Contactez votre conseiller formation Global Knowledge ou adressez votre demande à [info@globalknowledge.fr](mailto:info@globalknowledge.fr).

### Public visé:

Ingénieurs de données Administrateurs de bases de données Administrateurs système

### Objectifs pédagogiques:

- A l'issue de la formation, les participants seront capables de :
  - Comprendre comment créer et déployer des pipelines de données de divers modèles sur Google Cloud.
- Comprendre le rôle d'un ingénieur de données.
- Identifier les tâches d'ingénierie des données et les composants principaux utilisés sur Google Cloud.
- Identifier et utiliser diverses techniques d'automatisation sur Google Cloud.

### Pré-requis:

- Expérience préalable de Google Cloud au niveau fondamental en utilisant Cloud Shell et en accédant aux produits depuis la console Google Cloud.
- Maîtrise de base d'un langage de requête courant tel que SQL.
- Expérience de la modélisation de données et des activités ETL (extraire, transformer, charger).
- Expérience du développement d'applications à l'aide d'un langage de programmation courant tel que Python.

### Test et certification

- 

Après cette formation, nous vous conseillons le(s) module(s) suivant(s):

- Ingénierie des données sur Google Cloud (GO5975)

## Contenu:

### Tâches et composants de l'ingénierie des données

- Expliquer le rôle d'un ingénieur de données.
- Comprendre les différences entre une source de données et un récepteur de données.
- Expliquer les différents types de formats de données.
- Expliquer les options de solution de stockage sur Google Cloud.
- Apprendre les options de gestion des métadonnées sur Google Cloud.
- Comprendre comment partager facilement des ensembles de données avec Analytics Hub.
- Comprendre comment charger des données dans BigQuery en utilisant la console Google Cloud ou la CLI gcloud

### Réplication et Migration de données

- Expliquer l'architecture de base de réplication et de migration de données de Google Cloud.
- Comprendre les options et les cas d'utilisation de l'outil de ligne de commande gcloud.
- Expliquer la fonctionnalité et les cas d'utilisation du Service de transfert de stockage.
- Expliquer la fonctionnalité et les cas d'utilisation de Transfer Appliance.
- Comprendre les fonctionnalités et le déploiement de Datastream

### Modèle de Pipeline d'extraction et de chargement de données

- Expliquer le diagramme d'architecture de base d'extraction et de chargement.
- Comprendre les options de l'outil de ligne de commande bq.
- Expliquer la fonctionnalité et les cas d'utilisation du Service de transfert de données BigQuery.
- Expliquer la fonctionnalité et les cas d'utilisation de BigLake en tant que modèle sans extraction-chargement.

### Modèle de Pipeline d'extraction, de chargement et de transformation de données

- Expliquer le diagramme d'architecture de base d'extraction, de chargement et de transformation.
- Comprendre un pipeline ELT courant sur Google Cloud.
- Apprendre les capacités de scripting SQL et de planification de BigQuery.
- Expliquer la fonctionnalité et les cas d'utilisation de Dataform.

### Modèle de Pipeline d'extraction, de transformation et de chargement

#### de données

- Expliquer le diagramme d'architecture de base d'extraction, de transformation et de chargement.
- Apprendre les outils graphiques de Google Cloud utilisés pour les pipelines de données ETL.
- Expliquer le traitement de données par lots avec Dataproc.
- Apprendre à utiliser Dataproc Serverless pour Spark pour l'ETL.
- Expliquer les options de traitement de données en continu.
- Expliquer le rôle que joue Bigtable dans les pipelines de données.

### Techniques d'automatisation

- Expliquer les modèles et les options d'automatisation disponibles pour les pipelines.
- Apprendre sur Cloud Scheduler et Workflows.
- Apprendre sur Cloud Composer.
- Apprendre sur les fonctions Cloud Run.
- Expliquer la fonctionnalité et les cas d'utilisation d'automatisation pour Eventarc.

## Méthodes pédagogiques :

Un support de cours officiel Google en anglais sera fourni aux participants.

## Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.