Formation GCP – Professional Data Engineer

Informations

Durée : 4 jours (28h.)

Tarif* : Nous consulter

Réf : GCDE

Niveau : Difficile

intra

Mise à jour le 18/12/25

*tarif valable jusqu'au 31/12/2026

Prochaines sessions

Contactez-nous pour connaitre nos futures sessions.

Pré-requis

- Bonnes notions de bases de données relationnelles et NoSQL
- Connaissances en SQL (indispensable pour BigQuery)
- Notions de programmation (Python recommandé)
- Connaissance des fondamentaux du Cloud (niveau Digital Leader ou Associate)

Objectifs

Objectifs pédagogiques :

- Comprendre l'écosystème Data de Google Cloud
- Maîtriser la conception et la gestion d'architectures de traitement de données dans GCP
- Savoir déployer des pipelines batch et temps réel
- Apprendre à sécuriser, superviser et optimiser les plateformes de données
- Préparer la certification Google Cloud Professional Data Engineer

Objectifs opérationnels :

- Concevoir, déployer et exploiter une architecture Data complète sur GCP : ingestion temps réel ou batch, traitement de données, stockage (BigQuery, NoSQL), intégration ML, sécurité et optimisation de performance.

Programme

Jour 1 – Fondations de la Data dans GCP

Introduction au rôle de Data Engineer dans GCP
Présentation des services Data : BigQuery, Pub/Sub, Dataflow, Dataproc, Firestore, Bigtable
Modélisation de données dans le Cloud
Lab pratique : chargement et interrogation de données avec BigQuery

Jour 2 – Pipelines de données et traitement batch/streaming

Pub/Sub : ingestion d'événements et messages en temps réel
Dataflow : pipelines batch et streaming (Apache Beam)
Dataproc : clusters Hadoop/Spark pour le Big Data
Orchestration de pipelines avec Composer (Airflow)
Lab pratique : construire un pipeline Pub/Sub → Dataflow → BigQuery

Jour 3 – Bases de données et intégration IA/ML

Bases de données relationnelles avec Cloud SQL
Bases NoSQL : Firestore et Bigtable
Utilisation de Vertex AI pour entraîner et déployer un modèle
Intégration IA dans une architecture Data
Cas pratique : pipeline Data enrichi par un modèle ML (prédictions temps réel)

Jour 4 – Sécurité, Optimisation et Certification

Sécurité des données dans GCP : IAM, DLP, KMS
Gouvernance des données : Data Catalog, contrôle d'accès
Supervision et optimisation des performances BigQuery et Dataflow
Projet fil rouge : conception d'une plateforme Data complète (ingestion, traitement, stockage, visualisation)
Préparation à la certification Professional Data Engineer : quiz, études de cas