

Certifiante

Formation Data Engineering sur Google Cloud Platform (DEGCP)

Réf. SF34392

Maîtriser la conception et l'implémentation de pipelines de données performants dans le cloud Google !

4 j Durée	3 400 € HT Inter / participant	8 000 € HT Intra (≤12 pers.)	0 Sessions à venir
--------------	-----------------------------------	---------------------------------	-----------------------

Objectifs pédagogiques

- ✓ Concevoir et développer des systèmes de traitement de données sur Google Cloud Platform.
- ✓ Mettre en œuvre des pipelines évolutifs pour des traitements batch et temps réel avec Cloud Dataflow.
- ✓ Exploiter BigQuery pour analyser de grands volumes de données et en extraire des insights métier.
- ✓ Créer, entraîner, évaluer et déployer des modèles de machine learning avec TensorFlow et Cloud ML.
- ✓ Traiter des données non structurées à l'aide de Spark et des API ML sur Cloud Dataproc.
- ✓ Mettre en place des architectures de streaming pour obtenir des analyses instantanées.

Public visé & prérequis

Public visé Data analysts, Data scientists, Développeurs, Ingénieurs Big Data

Prérequis Avoir suivi le cours Google Cloud Fundamentals : Big Data and Machine Learning ou disposer d'une expérience équivalente. Connaissance pratique d'un langage de requête tel que SQL. Expérience des activités ETL et de la modélisation de données. Maîtrise d'un langage de programmation courant, tel que Python. Notions de base en statistiques et machine learning.

Programme détaillé

Module 1 : Introduction à Google Cloud Dataproc

Création et gestion de clusters
Utilisation de types de machines personnalisés et de nœuds préemptibles
Mise à l'échelle et suppression des clusters

Module 2 : Exécution de jobs Dataproc

Lancement de jobs Pig et Hive
Séparation du stockage et du calcul

Module 3 : Intégration de Dataproc avec Google Cloud Platform

Personnalisation des clusters avec des actions d'initialisation
Intégration avec BigQuery

Module 4 : Exploitation des données non structurées avec les API ML de Google

Présentation des API de machine learning
Cas d'usage courants
Appels aux API ML

Module 5 : Analyse de données serverless avec BigQuery

Introduction à BigQuery

Écriture de requêtes et utilisation de fonctions
Chargement et exportation des données
Gestion des champs imbriqués et répétés
Interrogation de tables multiples
Optimisation des performances et coûts

Module 6 : Pipelines de données autoscalables avec Dataflow

Le modèle de programmation Beam
Pipelines en Python et Java
Traitements distribués et MapReduce
Gestion du streaming et architecture de référence

Module 7 : Introduction au Machine Learning

Principes fondamentaux et typologies du ML
Généralisation et jeux de données adaptés

Module 8 : Création de modèles ML avec TensorFlow

Débuter avec TensorFlow
Construction et entraînement de modèles
Graphes, boucles et mécanismes de monitoring

Module 9 : Passage à l'échelle avec Cloud ML

Intérêt et cas d'usage de Cloud ML
Emballage et déploiement d'un modèle TensorFlow
Entraînement local et sur le cloud

Module 10 : Feature Engineering

Conception de variables pertinentes
Transformation et création de variables synthétiques
Prétraitement avec Cloud ML

Module 11 : Architectures de pipelines de streaming

Enjeux du traitement de flux
Gestion de volumes variables et données désordonnées ou retardées

Module 12 : Ingestion de volumes variables avec Cloud Pub/Sub

Concepts de topics et abonnements
Principes de fonctionnement

Module 13 : Implémentation de pipelines de streaming

Défis spécifiques au traitement en flux
Gestion des données tardives (watermarks, triggers, accumulation)

Module 14 : Analytique en flux et tableaux de bord

De la donnée brute à la prise de décision
Interrogation de données en streaming avec BigQuery
Découverte de Google Data Studio

Module 15 : Performance et faible latence avec Bigtable

Présentation de Cloud Spanner
Conception d'un schéma Bigtable

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

Note moyenne des participants de la filière Data Engineering & Big Data : 4,9/5 — 13 avis

Arnaud M. — That was really useful for me, even though I am a PDA. The trainer was very patient with me (I needed more explanation as I was the only...

Thierry R. — Formation idéale pour appréhender la technologie nosql Bon support de formation. Bonne pédagogie

Ils nous font confiance

Cette année notre filière Data Engineering & Big Data a accompagné plus de 73 entreprises dont :

SOPRA STERIA CAPGEMINI TECHNOLOGY SERVICES AMADEUS SAS GROUPAMA SUPPORTS & SERVICES AUBAY FRAIKIN FRANCE LBC FRANCE
ORANGE BUSINESS SERVICES SA ACATUS INFORMATIQUE ACENSI SANA +63 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com