Formation Google Cloud Platform - Ingénierie de données

Concevoir et développer des systèmes de traitement de données sur GCP

N° ACTIVITÉ : 11 92 18558 92 | TÉLÉPHONE : 01 85 77 07 07 | E-MAIL : inscription@hubformation.com

Si les bénéfices liés à l'adoption du Cloud sont aujourd'hui nombreux (disponibilité, agilité, adaptabilité, gains financiers,...), l'une des principales vertus de cette révolution est de pouvoir accéder à des outils et à des puissances de traitement qu'il serait bien souvent difficile de s'offrir (et de rentabiliser) dans un modèle traditionnel. C'est notamment dans le cadre du traitement et de l'analyse de très importants volumes de données (Big Data) que se mesure sans doute le mieux cet avantage. Comme tous les grands acteurs du Cloud, Google propose ainsi de nombreux services dédiés au traitement et à l'analyse dite de type Big Data ainsi que des solutions permettant de tirer par exemple parti du Machine Learning. Les participants à cette formation apprendront à concevoir des systèmes de traitement de données, à créer des pipelines de données, à réaliser des traitements sur des données structurées et non structurées et enfin à exploiter ces données grâce à de puissants outils d'analyse.

Référence	GCP200DE
Durée	4 jours (28h)
Tarif	3 250 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

du 22 au 25 septembre 2025
du 1er au 4 décembre 2025

VOIR TOUTES LES DATES

Objectifs

| concevoir et développer des systèmes de traitement de données sur Google Cloud Platform
| traiter des données par lot ou par flux en implémentant des pipelines de données d'autoscaling sur Cloud Dataflow
| obtenir des informations métier à partir de très grands ensembles de données à l'aide de Google BigQuery
| établir des prédictions avec des modèles de Machine Learning en utilisant Tensorflow et Cloud ML
| tirer parti des données non structurées à l'aide de Spark et des API de Machine Learning sur Cloud Dataproc
| activer Instant Insights à partir des données par flux

Public

| Développeurs expérimentés en charge des transformations du Big Data

Prérequis

| Avoir suivi la formation "Google Cloud Platform - Les fondamentaux du Big Data et du Machine Learning" ou bénéficier d'une expérience équivalente
| Maîtriser les principes de base des langages de requête courants tels que SQL
| Avoir de l'expérience en modélisation, extraction, transformation et chargement des données
| Savoir développer des applications à l'aide d'un langage de programmation courant tel que Python
| Savoir utiliser le Machine Learning et/ou les statistiques
| Pour suivre cette formation dans des conditions optimales, nous vous recommandons de venir en formation avec un ordinateur portable

Programme de la formation

Présentation de Google Cloud Dataproc
| Créer et gérer des clusters
| Exploiter des types de machines personnalisés et des noeuds de calcul

préemptifs
| Faire évoluer et supprimer des clusters
| Atelier : Créer des clusters Hadoop avec Google Cloud Dataproc

Exécution de tâches Dataproc
| Exécuter des tâches Pig et Hive
| Séparer le stockage et le calcul
| Atelier : Exécuter des tâches Hadoop et Spark avec Dataproc
| Atelier : Envoyer et surveiller des tâches

Intégrer Dataproc à Google Cloud Platform
| Personnaliser un cluster à l'aide d'actions d'initialisation
| Assistance BigQuery
| Atelier : Exploiter les services Google Cloud Platform

Analyse des données non structurées avec les API de Machine Learning de Google
| API de Machine Learning Google
| Cas d'utilisation courants du ML
| Appeler des API de ML
| Atelier : Ajouter les capacités de Machine Learning à l'analyse Big Data

Analyse des données sans serveur avec BigQuery
| Qu'est-ce que BigQuery ?
| Requêtes et fonctions
| Atelier : Écrire des requêtes dans BigQuery
| Charger des données dans BigQuery
| Exporter des données à partir de BigQuery
| Atelier : Charger et exporter des données
| Champs imbriqués et répétés
| Interroger plusieurs tables
| Atelier : Requêtes complexes
| Performances et tarification

Pipelines de données d'autoscaling sans serveur avec Dataflow
| Le modèle de programmation Beam
| Pipelines de données dans Beam Python
| Pipelines de données dans Beam Java
| Atelier : Écrire un pipeline Dataflow
| Traitement adaptable du Big Data avec Beam
| Atelier : MapReduce dans Dataflow
| Intégrer des données supplémentaires
| Atelier : Entrées secondaires
| Gérer des données par flux
| Architecture de référence GCP

Premiers pas avec le Machine Learning
| Qu'est-ce que le Machine Learning (ML) ?
| ML efficace : concepts et types
| Ensembles de données de ML : généralisation
| Atelier : Explorer et créer des ensembles de données de ML

Créer des modèles de ML avec TensorFlow
| Premiers pas avec TensorFlow
| Atelier : Utiliser tf.learn
| Graphiques et boucles TensorFlow + atelier
| Atelier : Utiliser TensorFlow de bas niveau + arrêt prématuré
| Surveiller l'apprentissage ML
| Atelier : Graphiques d'apprentissage TensorFlow

Faire évoluer les modèles de ML avec Cloud ML
| Pourquoi utiliser Cloud ML ?
| Créer le package d'un modèle TensorFlow
| Apprentissage de bout en bout
| Atelier : Exécuter un modèle de ML localement et sur le Cloud

Extraction de caractéristiques
| Créer des fonctionnalités pertinentes
| Transformer les entrées
| Fonctionnalités synthétiques
| Prétraitement avec Cloud ML
| Atelier : Extraction de caractéristiques

Architecture des pipelines d'analyse par flux
| Défis du traitement des données par flux
| Gérer les volumes de données variables
| Gérer les données non triées/en retard
| Atelier : Concevoir un pipeline de flux de données

Ingestion de volumes de variables
| Qu'est-ce que Cloud Pub/Sub ?
| Fonctionnement : Thèmes et abonnements
| Atelier : Simulateur

Mise en oeuvre de pipelines par flux
| Défis du traitement par flux
| Gestion des données en retard : filigranes, déclenchements et accumulation
| Atelier : Pipeline de traitement des données par flux pour le trafic en direct

Analyse de flux de données et tableaux de bords
| Analyses de flux de données : prendre des décisions à partir des données
| Interroger les données par flux avec BigQuery
| Qu'est-ce que Google Data Studio ?
| Atelier : Créer un tableau de bord en temps réel pour visualiser les données traitées

Haut débit et faible latence avec Bigtable
| Qu'est-ce que Cloud Spanner ?
| Concevoir un schéma Bigtable
| Ingestion dans Bigtable
| Atelier : Flux dans Bigtable

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

| Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.
| Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.
| Privilégier une connexion filaire plutôt que le Wifi.
| Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.
| Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).
| Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.

| Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.
| Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.
| Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.
Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.