

Formation Google Cloud Platform GCP200DE Ingénierie de données

Ce cours est disponible à distance
uniquement

Durée de la formation : 4 jour(s)

OBJECTIFS

A l'issue de la formation, l'apprenant sera capable de :

Savoir concevoir et déployer de pipelines et d'architectures pour le traitement des données

Comprendre comment obtenir des informations métier à partir de très grands ensembles de données à l'aide de Google BigQuery

Savoir tirer parti des données non structurées à l'aide de Spark et des API de Machine Learning sur Cloud Dataproc

Comprendre comment activer Instant Insights à partir des données par flux

PREREQUIS

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

PUBLIC

Développeurs expérimentés en charge des transformations du Big Data

Formation Google Cloud Platform GCP200DE Ingénierie de données

Durée de la formation : 4 jour(s)

Ce cours est disponible à distance
uniquement

PROGRAMME

INTRODUCTION À L'INGÉNIERIE DES DONNÉES

- Explorez le rôle d'un data engineer
- Analyser les défis d'ingénierie des données
- Introduction à BigQuery
- Data lakes et data warehouses
- Démo: requêtes fédérées avec BigQuery
- Bases de données transactionnelles vs data warehouses
- Démo: recherche de données personnelles dans votre jeu de données avec l'API DLP
- Travailler efficacement avec d'autres équipes de données
- Gérer l'accès aux données et gouvernance
- Construire des pipelines prêts pour la production
- Etude de cas d'un client GCP
- Lab: Analyse de données avec BigQuery

CONSTRUIRE UN DATA LAKE

- Introduction aux data lakes
- Stockage de données et options ETL sur GCP
- Construction d'un data lake à l'aide de Cloud Storage
- Démo: optimisation des coûts avec les classes et les fonctions cloud de Google Cloud Storage
- Sécurisation de Cloud Storage
- Stocker tous les types de données
- Démo: exécution de requêtes fédérées sur des fichiers Parquet et ORC dans BigQuery
- Cloud SQL en tant que data lake relationnel

CONSTRUIRE UN DATA WAREHOUSE

- Le data warehouse moderne
- Introduction à BigQuery
- Démo: Requête des TB + de données en quelques secondes
- Commencer à charger des données
- Démo: Interroger Cloud SQL à partir de BigQuery
- Lab: Chargement de données avec la console et la CLI
- Explorer les schémas
- Exploration des jeux de données publics BigQuery avec SQL à l'aide de INFORMATION_SCHEMA
- Conception de schéma
- Démo: Exploration des jeux de données publics BigQuery avec SQL à l'aide de INFORMATION_SCHEMA
- Champs imbriqués et répétés dans BigQuery
- Lab: tableaux et structures
- Optimiser avec le partitionnement et le clustering
- Démo: Tables partitionnées et groupées dans BigQuery
- Aperçu: Transformation de données par lots et en continu

INTRODUCTION À LA CONSTRUCTION DE PIPELINES DE DONNÉES PAR LOTS EL, ELT, ETL

Formation Google Cloud Platform GCP200DE Ingénierie de données

Durée de la formation : 4 jour(s)

Ce cours est disponible à distance
uniquement

Considérations de qualité
Comment effectuer des opérations dans BigQuery
Démo: ETL pour améliorer la qualité des données dans BigQuery
Des lacunes
ETL pour résoudre les problèmes de qualité des données

EXÉCUTION DE SPARK SUR CLOUD DATAPROC

L'écosystème Hadoop
Exécution de Hadoop sur Cloud Dataproc GCS au lieu de HDFS
Optimiser Dataproc
Atelier: Exécution de jobs Apache Spark sur Cloud Dataproc

TRAITEMENT DE DONNÉES SANS SERVEUR AVEC CLOUD DATAFLOW

Cloud Dataflow
Pourquoi les clients apprécient-ils Dataflow?
Pipelines de flux de données
Lab: Pipeline de flux de données simple (Python / Java)
Lab: MapReduce dans un flux de données (Python / Java)
Lab: Entrées latérales (Python / Java)
Templates Dataflow
Dataflow SQL

GESTION DES PIPELINES DE DONNÉES AVEC CLOUD DATA FUSION ET CLOUD COMPOSER

Création visuelle de pipelines de données par lots avec Cloud Data Fusion: composants, présentation de l'interface utilisateur, construire un pipeline, exploration de données en utilisant Wrangler
Lab: Construction et exécution d'un graphe de pipeline dans Cloud Data Fusion
Orchestrer le travail entre les services GCP avec Cloud Composer – Apache Airflow Environment: DAG et opérateurs, planification du flux de travail
Démo : Chargement de données déclenché par un événement avec Cloud Composer, Cloud Functions, Cloud Storage et BigQuery
Lab: Introduction à Cloud Composer

INTRODUCTION AU TRAITEMENT DE DONNÉES EN STREAMING

Traitement des données en streaming

SERVERLESS MESSAGING AVEC CLOUD PUB/SUB

Cloud Pub/Sub
Lab: Publier des données en continu dans Pub/Sub

FONCTIONNALITÉS STREAMING DE CLOUD DATAFLOW

Formation Google Cloud Platform GCP200DE Ingénierie de données

Ce cours est disponible à distance
uniquement

Durée de la formation : 4 jour(s)

Fonctionnalités streaming de Cloud Dataflow
Lab: Pipelines de données en continu

FONCTIONNALITÉS STREAMING À HAUT DÉBIT BIGQUERY ET BIGTABLE

Fonctionnalités de streaming BigQuery
Lab: Analyse en continu et tableaux de bord
Cloud Bigtable
Lab: Pipelines de données en continu vers Bigtable

FONCTIONNALITÉ AVANCÉES DE BIGQUERY ET PERFORMANCE

Analytic Window Functions
Utiliser des clauses With
Fonctions SIG
Démo: Cartographie des codes postaux à la croissance la plus rapide avec BigQuery GeoViz
Considérations de performance
Lab: Optimisation de vos requêtes BigQuery pour la performance
Lab: Création de tables partitionnées par date dans BigQuery

INTRODUCTION À L'ANALYTIQUE ET À L'IA

Qu'est-ce que l'IA?
De l'analyse de données ad hoc aux décisions basées sur les données
Options pour modèles ML sur GCP

API DE MODÈLE ML PRÉDÉFINIES POUR LES DONNÉES NON STRUCTURÉES

Les données non structurées sont difficiles à utiliser
API ML pour enrichir les données
Lab: Utilisation de l'API en langage naturel pour classer le texte non structuré

BIG DATA ANALYTICS AVEC LES NOTEBOOKS CLOUD AI PLATFORM

Qu'est-ce qu'un notebook
BigQuery Magic et liens avec Pandas
Lab: BigQuery dans Jupyter Labs sur IA Platform

PIPELINES DE PRODUCTION ML AVEC KUBEFLOW

Façons de faire du ML sur GCP
Kubeflow AI Hub
Lab: Utiliser des modèles d'IA sur Kubeflow

Formation Google Cloud Platform GCP200DE Ingénierie de données

Ce cours est disponible à distance
uniquement

Durée de la formation : 4 jour(s)

CRÉATION DE MODÈLES PERSONNALISÉS AVEC SQL DANS BIGQUERY ML

BigQuery ML pour la construction de modèles rapides

Démo: Entraîner un modèle avec BigQuery ML pour prédire les tarifs de taxi à New York

Modèles pris en charge

Lab: Prédire la durée d'une sortie en vélo avec un modèle de régression dans BigQuery ML

Lab: Recommandations de film dans BigQuery ML

CRÉATION DE MODÈLES PERSONNALISÉS AVEC CLOUD AUTOML

Pourquoi Auto ML?

Auto ML Vision

Auto ML NLP

Auto ML Tables

Formation Google Cloud Platform GCP200DE Ingénierie de données

Durée de la formation : 4 jour(s)

Ce cours est disponible à distance
uniquement

Prise en compte du handicap

Pour les personnes en situation de handicap, afin de nous permettre d'organiser le déroulement de la formation dans les meilleures conditions possibles, contactez-nous via notre formulaire de contact ou par mail (formation@access-it.fr) ou par téléphone (0320619506). Un entretien avec notre référente handicap pourra être programmé afin d'identifier les besoins et aménagements nécessaires.

Modalités et moyens Pédagogiques, techniques et d'encadrement mis en œuvre

Répartition théorie/pratique : 45%/55%. Cette formation se compose d'une alternance d'apports théoriques, de travaux pratiques s'articulant autour d'une application fil rouge, de démonstrations, de phases d'échanges entre participants et de synthèses de la part du formateur.

Formation accessible à distance de n'importe où et n'importe quand, via un ordinateur type PC disposant d'une connexion à Internet à haut débit (ADSL ou plus).

Pour assurer un démarrage dans les meilleures conditions au premier jour de la formation, notre service logistique se met systématiquement en relation, en amont, avec vous afin de réaliser un test de validation technique et de vous présenter l'environnement de formation.

Pendant toute la durée de la formation, le stagiaire dispose d'une assistance technique et pédagogique illimitée, par e-mail, avec un délai de prise en compte et de traitement qui n'excède pas 24h. En complément, le stagiaire peut planifier un rendez-vous pédagogique avec un formateur expert afin d'échanger sur des éléments de la formation.

La durée de la formation affichée sur cette page est une durée estimée qui peut varier en fonction du profil du stagiaire et de ses objectifs (notamment s'il souhaite valider sa formation par le passage d'un examen de certification).

Durant la formation, le formateur prévoit :

Des démonstrations organisées en modules et en séquences découpées le plus finement possible, en suivant le programme pédagogique détaillé sur cette page ;

Des énoncés et corrigés de travaux pratiques à réaliser tout au long de la formation ;

Des travaux pratiques sont proposés ; la plateforme prévoit l'environnement technique nécessaire à la réalisation de l'ensemble des travaux pratiques ;

Le formateur valide les connaissances acquises après chaque TP ;

Il est proposé un ou plusieurs livres numériques faisant office d'ouvrage(s) de référence sur le thème de la formation.

Validation et sanction de la formation

Une attestation mentionnant les objectifs, la nature et la durée de l'action et les résultats de l'évaluation des acquis de la formation sera remise au stagiaire à l'issue de sa formation par courrier électronique.

A la demande, il sera délivré un certificat de réalisation.

Type de formation

Professionalisante ayant pour objectif le perfectionnement, l'élargissement des compétences

Moyens permettant de suivre l'exécution de l'action

Le contrôle de la présence des stagiaires sera assuré par la vérification de l'assiduité des participants. Le stagiaire signera une feuille de présence par demi-journée de formation. Celle-ci sera également signée par le formateur.

Modalité d'évaluation des acquis

Durant la formation, le participant est amené à mettre en pratique les éléments du cours par la réalisation de travaux pratiques réalisés sur PC.

La validation des acquis du stagiaire est faite par le formateur à la fin de chaque atelier. Cette validation individuelle est possible du fait du faible nombre de participants par session de formation (6 personnes maximum).

A la fin de la formation, le stagiaire a donc atteint les objectifs fixés par la formation.

En complément, pour les stagiaires qui le souhaitent, certaines formations peuvent être validées officiellement par l'éditeur en passant un examen de certification.

Access it étant centre d'examen ENI, les examens peuvent être réalisés sur demande à distance ou dans nos locaux de Villeneuve d'Ascq.

Les candidats à la certification doivent produire un travail personnel important en vue de se présenter au passage de l'examen, le seul suivi de la formation ne constitue pas un élément suffisant pour garantir un bon résultat et/ou l'obtention de la certification

Assistance Post-Formation

Toute personne ayant suivi une formation avec Access it bénéficie d'une assistance post-formation d'une durée de 1 mois. Ce nouveau service d'accompagnement permet aux stagiaires rencontrant des difficultés dans la mise en œuvre des connaissances acquises de solliciter l'aide de nos formateurs sur des aspects relatifs aux programmes de formation suivis. Pour en bénéficier, il suffit de se rendre sur la page contact de notre site web et de remplir le formulaire. Une réponse est apportée par mail ou par téléphone dans un délai de 48 heures.

Centre d'Examen ENI

Notre centre de formation agréé Qualiopi bénéficie de l'agrément d'ENI pour les certifications informatiques. C'est pour nos clients la garantie de pouvoir suivre des formations préparant à des certifications officiellement reconnues par l'Etat (Répertoire spécifique) et de se certifier en un même lieu.

Aide à l'orientation

Pour chacune des grandes thématiques couvertes par notre offre de formation, nous proposons via nos spécialistes un rendez-vous physique ou téléphonique qui via un diagnostic permettra aux personnes souhaitant être accompagnées dans le choix d'un programme ou dans la définition d'un parcours de formation une orientation vers les programmes les plus adaptés à leurs besoins et à leur niveau.

Aspects Pratiques

Dès leur inscription, les participants sont contactés par nos services qui s'assurent que les débits internet constatés sur le lieu depuis lequel ils souhaitent se former sont suffisants pour suivre la formation dans des conditions optimales.

À l'occasion de cet appel, nos experts s'assurent également qu'ils disposent du matériel nécessaire pour suivre la formation (PC Portable, webcam, Micro-casque..).

Avant le début de la formation, les participants reçoivent un lien leur permettant d'accéder à la classe virtuelle et leurs identifiants personnels de connexion. Un aide à l'utilisation de la solution de visioconférence utilisée leur est également proposée.

Le jour de la formation, ils se connectent à la classe virtuelle depuis leur navigateur internet. Ils voient et entendent le formateur ainsi que les autres participants et peuvent à tout moment communiquer avec eux.

Ils participent aux échanges et réalisent les ateliers dans les mêmes conditions que s'ils étaient en salle. Grâce à nos outils de prise en main à distance, les formateurs peuvent à tout moment prendre la main sur leurs postes pour les aider ou vérifier leurs TP.

Tout au long de la formation, les participants peuvent bénéficier de l'assistance immédiate de nos experts en composant le numéro qui leur a été communiqué avant la formation.

Des bilans intermédiaires ont lieu en présence des participants du formateur et du référent pédagogique d'Access it afin de vérifier l'état d'avancement de la session, les difficultés rencontrées et permettre d'éventuels actions correctives.

Bénéfices pour les participants

Se former depuis leur lieu de travail ou leur domicile,

Accéder sans se déplacer à la qualité d'une formation délivrée par un formateur consultant ayant une expérience probante sur le sujet animé.

Bénéficier à distance de la richesse d'une formation interentreprises : échanges avec le formateur et les autres participants, partages d'expériences, ateliers pratiques...

Pouvoir se former en toutes circonstances et notamment en cas d'imprévus.

Bénéfices pour l'entreprise

Optimiser ses budgets en limitant les frais de déplacement et d'hébergement.

Proposer à tous ses collaborateurs, quelle que soit leur situation géographique, des formations de qualité (en Inter comme en Intra).

Limiter les temps de déplacement.

Proposer davantage de choix dans les formations à des collaborateurs peu mobiles.

Assurer la montée en compétences de ses collaborateurs quelles que soient les circonstances