

Formation : Introduction to Data Engineering on Google Cloud

Cours officiel, préparation aux examens de certification Google Cloud

Cours pratique - 1j - 7h00 - Réf. GDE

Prix : 850 € H.T.

NEW

Avec cette formation, vous pourrez acquérir les connaissances et compétences fondamentales en ingénierie des données sur Google Cloud Platform (GCP) afin de comprendre, concevoir et mettre en œuvre des pipelines de données efficaces. Vous apprendrez à utiliser les principaux services de GCP pour l'ingestion, le stockage, la transformation et l'analyse des données, tout en adoptant les bonnes pratiques de performance, de sécurité et de gouvernance des données.

Objectifs pédagogiques

À l'issue de la formation, le participant sera en mesure de :

- ✓ Comprendre le rôle d'un ingénieur de données
- ✓ Identifier les tâches d'ingénierie des données et les principaux composants utilisés sur Google Cloud
- ✓ Comprendre comment créer et déployer des pipelines de données de différents modèles sur Google Cloud
- ✓ Identifier et utiliser diverses techniques d'automatisation sur Google Cloud

Public concerné

Ingénieurs de données, administrateurs de bases de données, administrateurs système.

Prérequis

Connaissance de base de l'environnement Google Cloud. Maîtrise de base d'un langage de requête courant. Expérience en modélisation de données et activités ETL et en développement d'applications.

PARTICIPANTS

Ingénieurs de données,
administrateurs de bases de données,
administrateurs système.

PRÉREQUIS

Connaissance de base de l'environnement Google Cloud. Maîtrise de base d'un langage de requête courant. Expérience en modélisation de données et activités ETL et en développement d'applications.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils sont agréés par l'éditeur et sont certifiés sur le cours. Ils ont aussi été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum trois à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Évaluation des compétences visées en amont de la formation. Évaluation par le participant, à l'issue de la formation, des compétences acquises durant la formation. Validation par le formateur des acquis du participant en précisant les outils utilisés : QCM, mises en situation...

À l'issue de chaque formation, ITTCERT fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Les participants réalisent aussi une évaluation officielle de l'éditeur. Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le participant a bien assisté à la totalité de la session.

Certification

Nous vous recommandons de suivre cette formation si vous souhaitez préparer la certification "Google Cloud Professional Data Engineer".

[Comment passer votre examen ?](#)

Méthodes et moyens pédagogiques

Méthodes pédagogiques

Animation de la formation en français. Support de cours officiel au format numérique et en anglais. Bonne compréhension de l'anglais à l'écrit.

Modalités d'évaluation

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

Programme de la formation

1 Tâches et composants de l'ingénierie des données

- Comprendre le rôle d'un ingénieur de données.
- Comprendre les différences entre une source de données et un puits de données.
- Découvrir les différents types de formats de données.
- Expliquez les options de solution de stockage sur Google Cloud.
- Découvrir les options de gestion des métadonnées sur Google Cloud.
- Découvrir comment partager facilement des ensembles de données à l'aide d'Analytics Hub.
- Découvrir comment charger des données dans BigQuery à l'aide de la console Google Cloud ou de la gcloud CLI.

Travaux pratiques

Lab : chargement de données dans BigQuery.

2 Réplication et migration des données

- Découvrir l'architecture de base de réplication et de migration des données de Google Cloud.
- Comprendre les options et les cas d'utilisation de l'outil de ligne de commande gcloud.
- Découvrir les fonctionnalités et les cas d'utilisation de Storage Transfer Service.
- Découvrir les fonctionnalités et les cas d'utilisation de Transfer Appliance.
- Découvrir les fonctionnalités et le déploiement de Datastream.

Travaux pratiques

Lab : Réplication PostgreSQL vers BigQuery.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les ressources pédagogiques utilisées sont les supports et les travaux pratiques officiels de l'éditeur.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

3 Le modèle de pipeline Extract and Load

- Extraire et charger l'architecture.
- Comprendre les options de l'outil de ligne de commande bq.
- Découvrir les fonctionnalités et les cas d'utilisation du service de transfert de données BigQuery.
- Découvrir les fonctionnalités et les cas d'utilisation de BigLake en tant que modèle de chargement sans extraction.

Travaux pratiques

Lab : Introduction à BigLake.

4 Le modèle de pipeline Extract, Load and Transform

- Expliquer le diagramme de base d'architecture d'extraction, de chargement et de transformation.
- Comprendre un pipeline ELT courant sur Google Cloud.
- Découvrir les fonctionnalités de programmation et de script SQL de BigQuery.
- Expliquer les fonctionnalités et les cas d'utilisation de Dataform.

Travaux pratiques

Lab : Créer et exécuter un workflow SQL dans Dataform.

5 Modèle de pipeline Extract, Transform and Load

- Découvrir le diagramme d'architecture d'extraction, de transformation et de chargement de base.
- Découvrir les outils d'interface utilisateur graphique sur Google Cloud utilisés pour les pipelines de données ETL.
- Expliquer le traitement de données par lots à l'aide de Dataproc.
- Utiliser Dataproc Serverless pour Spark pour ETL.
- Explorer les options de traitement de données en streaming.
- Comprendre le rôle que joue Bigtable dans les pipelines de données.

Travaux pratiques

Lab : Utiliser Dataproc Serverless pour Spark pour charger BigQuery. Lab : Créer un pipeline de données en continu pour un tableau de bord en temps réel avec Dataflow.

6 Techniques d'automatisation

- Découvrir les modèles et options d'automatisation disponibles pour les pipelines.
- Découvrir Cloud Scheduler et Workflows.
- Découvrir Cloud Composer.
- Découvrir les fonctions de Cloud Run.
- Découvrir les cas d'utilisation des fonctionnalités et de l'automatisation pour Eventarc.

Travaux pratiques

Lab : Utiliser les fonctions Cloud Run pour charger BigQuery.

Dates et lieux

2026 : 18 juin, 1 oct., 10 déc.

2026 : 18 juin, 1 oct., 10 déc.