Date de publication : 27/11/2025

Formation : Data Integration with Cloud Data Fusion

Cours officiel, préparation aux examens de certification Google Cloud

Formation pratique - 2j - 14h00 - Réf. DID
Prix : 2060 € H.T.

Avec cette formation, vous prendrez en main Cloud Data Fusion, une plateforme d'intégration de données, permettant de créer et de gérer rapidement des pipelines de données. Vous aborderez les défis liés à l'intégration de données et la nécessité d'une plateforme. Vous découvrirez les composants de Cloud Data Fusion, comment traiter les données par lots et les données de diffusion en temps réel grâce à une conception de pipelines visuelle, un suivi approfondi des métadonnées et de la lignée des données, et comment déployer des pipelines de données sur différents moteurs d'exécution.

Objectifs pédagogiques

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

- Identifier les besoins d'intégration de données et les cas d'usage adaptés à Cloud Data Fusion
- Comprendre les capacités et les principaux composants de la plateforme Cloud Data Fusion
- Concevoir et exécuter des pipelines de traitement batch et temps réel
- Utiliser Wrangler et les connecteurs pour transformer et intégrer des données multi-sources
- Configurer l'environnement d'exécution, superviser, dépanner et comprendre métadonnées et lignage

Public concerné

Data engineers, data analysts.

Prérequis

Avoir suivi le cours « Big Data and Machine Learning Fundamentals » Réf. GCD ou avoir des connaissances équivalentes.

Certification

Nous vous recommandons de suivre cette formation si vous souhaitez préparer la certification "Google Cloud Professional Data Engineer".
Comment passer votre examen ?

PARTICIPANTS

Data engineers, data analysts.

PRÉREQUIS

Avoir suivi le cours « Big Data and Machine Learning Fundamentals » Réf. GCD ou avoir des connaissances équivalentes.

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.

Méthodes et moyens pédagogiques

Méthodes pédagogiques

Animation de la formation en français. Support de cours officiel en anglais et au format numérique. 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 Introduction à l'intégration de données et Cloud Data Fusion

- Comprendre le besoin d'intégration de données.
- Lister les situations/cas où l'intégration de données peut aider les entreprises.
- Lister les plateformes et outils d'intégration de données disponibles.
- Identifier les défis liés à l'intégration des données.
- Comprendre l'utilisation de Cloud Data Fusion en tant que plateforme d'intégration de données.
- Créer une instance Cloud Data Fusion.
- Découvrir le framework de base et les principaux composants de Cloud Data Fusion.

2 Construire des pipelines

- Comprendre l'architecture de Cloud Data Fusion.
- Définir ce qu'est un pipeline de données.
- Comprendre la représentation DAG d'un pipeline de données.
- Apprendre à utiliser Pipeline Studio et ses composants.
- Concevoir un pipeline simple à l'aide de Pipeline Studio.
- Déployer et exécuter un pipeline.

3 Construire des pipelines complexes

- Effectuer des opérations de branchement, de fusion et de jointure.
- Exécuter le pipeline avec des arguments d'exécution à l'aide de macros.
- Travailler avec des gestionnaires d'erreurs.
- Exécuter des exécutions pré- et post-pipeline à l'aide d'actions et de notifications.
- Planifier l'exécution des pipelines.
- Importer et exporter des pipelines existants.

4 Environnement d'exécution du pipeline

- Comprendre la composition d'un environnement d'exécution.
- Configurer l'environnement d'exécution, la journalisation et les métriques de votre pipeline.
- Comprendre des concepts tels que le profil de calcul et l'approvisionneur.
- Créer un profil de calcul.
- Créer des alertes de pipeline.
- Surveiller le pipeline en cours d'exécution.

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.

5 Construire des transformations et préparer des données avec Wrangler

- Comprendre l'utilisation de Wrangler et de ses principaux composants.
- Transformer les données à l'aide de l'interface utilisateur de Wrangler.
- Transformer les données à l'aide de directives/méthodes CLI.
- Créer et utiliser des directives définies par l'utilisateur.

6 Connecteurs et pipelines de streaming

- Comprendre l'architecture d'intégration de données.
- Lister les différents connecteurs.
- Utiliser l'API Cloud Data Loss Prevention (DLP).
- Comprendre l'architecture de référence des pipelines de streaming.
- Construire et exécuter un pipeline de streaming.

7 Métadonnées et lignage des données

- Répertorier les types de métadonnées.
- Différencier les métadonnées commerciales, techniques et opérationnelles.
- Comprendre ce qu'est le lignage des données.
- Comprendre l'importance de maintenir la lignée de données.
- Différencier les métadonnées et le lignage des données.

Ce programme est une création originale, conçu par les équipes pédagogiques d'ORSYS Formation. Toute reproduction, représentation, adaptation ou exploitation, totale ou partielle, sans autorisation préalable écrite d'ORSYS, est strictement interdite. ORSYS se réserve le droit d'engager toute action nécessaire à la protection de ses droits de propriété intellectuelle.

Dates et lieux

CLASSE À DISTANCE
2026 : 3 déc.

PARIS LA DÉFENSE
2026 : 3 déc.