

Formation : Build Batch Data Pipelines on Google Cloud

Cours officiel, préparation aux certifications Google Cloud

Formation pratique - 1j - 7h00 - Réf. GBD

Prix : 990 € H.T.

NEW

Avec cette formation, vous apprendrez à concevoir et optimiser des pipelines de données batch sur Google Cloud, en utilisant Dataflow et Dataproc Serverless, tout en assurant la qualité, la supervision et la fiabilité des traitements.

Objectifs pédagogiques

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

- ✓ Déterminer si les pipelines de données batch sont adaptés à votre cas d'usage.
- ✓ Concevoir des pipelines batch scalables pour traiter de grands volumes de données.
- ✓ Mettre en place des contrôles de qualité des données pour garantir leur intégrité.
- ✓ Orchestrer, gérer et surveiller les pipelines batch, avec gestion des erreurs et monitoring.

Public concerné

Ingénieurs data, analystes data.

Prérequis

Connaissances de base en data warehousing et ETL/ELT, maîtrise de SQL, notions de programmation (Python recommandé) et familiarité avec GCloud CLI, la console et les services principaux de Google Cloud.

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

PARTICIPANTS

Ingénieurs data, analystes data.

PRÉREQUIS

Connaissances de base en data warehousing et ETL/ELT, maîtrise de SQL, notions de programmation (Python recommandé) et familiarité avec GCloud CLI, la console et les services principaux de Google Cloud.

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.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

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

1 Quand choisir les pipelines batch

- Comprendre le rôle du data engineer.
- Comprendre le rôle des composants.
- Comprendre le rôle le cycle de vie des pipelines batch.

Travaux pratiques

Description de la pratique Cas d'usage des pipelines batch, traitement et défis courants.

2 Concevoir et construire des pipelines batch

- Conception de pipelines batch.
- Transformations de données à grande échelle.
- Dataflow et Serverless pour Apache Spark.
- Connexions de données et orchestration.
- Exécution de pipelines Apache Spark.
- Optimisation des performances.

Travaux pratiques

Pipeline batch avec Serverless pour Apache Spark, pipeline batch avec Dataflow Job Builder.

3 Orchestrer et superviser les pipelines

- Orchestration des traitements batch.
- Cloud Composer.
- Observabilité et monitoring.
- Alertes et troubleshooting.
- Gestion visuelle des pipelines.

Travaux pratiques

Pipeline batch avec Serverless pour Apache Spark, pipeline batch avec Dataflow Job Builder

4 Contrôler la qualité des données

- Validation et nettoyage des données.
- Analyse des erreurs (logs).
- Évolution des schémas.
- Intégrité et déduplication.

Travaux pratiques

Validation de la qualité des données.

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.

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 : 30 sep., 9 déc.

PARIS LA DÉFENSE
2026 : 23 sep., 2 déc.