

Formation : Serverless Data Processing with Dataflow

Cours officiel, préparation aux examens de certification Google Cloud

Cours pratique - 3j - 21h00 - Réf. SDD

Prix : 2890 € H.T.

Avec cette formation, vous approfondissez votre maîtrise de Dataflow pour faire évoluer vos applications de traitement de données. Vous découvrez comment Apache Beam et Dataflow fonctionnent ensemble sans risque de verrouillage fournisseur. Vous apprenez à transformer votre logique métier en pipelines Dataflow, puis vous maîtrisez les opérations essentielles : supervision, dépannage, tests et fiabilité.

Objectifs pédagogiques

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

- ✓ Expliquer l'usage conjoint de Beam et Dataflow pour traiter les données efficacement
- ✓ Activer Portability, Shuffle/Streaming Engine et Flexible Scheduling pour optimiser coûts et performance
- ✓ Choisir les bons accès IAM et appliquer les bonnes pratiques de sécurité des pipelines
- ✓ Configurer et optimiser l'I/O, les schémas, SQL/DataFrames pour simplifier et accélérer le pipeline
- ✓ Assurer le suivi, les tests, le dépannage et le CI/CD des pipelines Dataflow

Public concerné

Data engineers, data analysts et data scientists aspirant à développer des compétences en ingénierie des données.

Prérequis

Avoir suivi le cours "Data Engineering on Google Cloud Platform" Réf DGC 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 et data scientists aspirant à développer des compétences en ingénierie des données.

PRÉREQUIS

Avoir suivi le cours "Data Engineering on Google Cloud Platform" Réf DGC 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 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 Portabilité de Beam

- Portabilité Beam.
- Runner v2.
- Environnements de conteneurs.
- Transformations inter-langages.

2 Séparer le calcul et le stockage avec Dataflow

- Dataflow Streaming Engine.
- Planification flexible des ressources.
- Dataflow.
- Service Dataflow Shuffle.

3 IAM, quotas et permissions

- IAM.
- Quotas.

4 Sécurité

- Localisation des données.
- Shared VPC.
- IP privées.
- CMEK.

5 Revue des concepts de Beam

- Bases de Beam.
- Transformations utilitaires.
- Cycle de vie d'un DoFn.

6 Fenêtres, watermarks, triggers

- Fenêtres (Windows).
- Watermarks.
- Déclencheurs (Triggers).

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.

7 Sources et Sinks

- Sources et Sinks.
- Text IO et File IO.
- BigQuery IO.
- Pub/Sub IO.
- Kafka IO.
- Bigtable IO.
- Avro IO.
- Splittable DoFn.

8 Schémas

- Schémas Beam.
- Exemples de code.

9 État et minuteurs

- API State.
- API Timer.

10 Bonnes pratiques

- Schémas.
- Gestion des données non traitables.
- Gestion des erreurs.
- Générateur de code AutoValue.
- Gestion des données JSON.
- Utiliser le cycle de vie des DoFn.
- Optimisations de pipeline.

11 Dataflow SQL et DataFrames

- Dataflow et Beam SQL.
- Fenêtrage en SQL.
- Beam DataFrames.

12 Notebooks Beam

- Notebooks Beam.

13 Monitoring

- Liste des jobs.
- Informations sur les jobs.
- Graphe du job.
- Métriques du job.
- Metrics Explorer.

14 Journalisation et rapport d'erreurs

- Logging.
- Error Reporting.

15 Dépannage et debug

- Processus de dépannage.
- Types de problèmes.

16 Performance

- Conception du pipeline.
- Structure des données.
- Sources, Sinks et systèmes externes.
- Shuffle et Streaming Engine.

17 Tests et CI/CD

- Vue d'ensemble des tests et de la CI/CD.
- Tests unitaires.
- Tests d'intégration.
- Construction d'artifacts.
- Déploiement.

18 Fiabilité

- Introduction à la fiabilité.
- Supervision.
- Géolocalisation.
- Reprise après sinistre.
- Haute disponibilité.

19 Flex Templates

- Templates classiques.
- Flex Templates.
- Utilisation des Flex Templates.
- Templates fournis par Google.

Dates et lieux

CLASSE À DISTANCE

2026 : 16 juin, 29 sep., 1 déc.

PARIS LA DÉFENSE

2026 : 16 juin, 29 sep., 1 déc.