

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

OBJECTIFS

A l'issue de la formation le participant sera capable de :

Démontrer comment Apache Beam et Dataflow fonctionnent ensemble pour répondre aux besoins de traitement des données de votre organisation.

Résumer les avantages de Beam Portability Framework et activer-le pour vos pipelines Dataflow.

Activer Shuffle et Streaming Engine, respectivement pour les pipelines batch et streaming, pour des performances maximales.

Activer la planification flexible des ressources pour des performances plus rentables.

Sélectionner la bonne combinaison d'autorisations IAM pour votre tâche Dataflow.

Mettre en œuvre les meilleures pratiques pour un environnement de traitement de données sécurisé.

Sélectionner et ajuster les E/S de votre choix pour votre pipeline Dataflow.

Utiliser des schémas pour simplifier votre code Beam et améliorer les performances de votre pipeline.

Développer un pipeline Beam en utilisant SQL et DataFrames.

Effectuer la surveillance, le dépannage, les tests et la CI/CD sur les pipelines Dataflow.

PREREQUIS

Pour tirer le meilleur parti de ce cours, les participants doivent :

Avoir suivi le modules « Créer des pipelines de données par lots » dans le cours Data Engineering on Google Cloud ou avoir des connaissances équivalentes

Avoir suivi le module « Créer des systèmes d'analyse de flux résilients » dans le cours Data Engineering on Google Cloud ou avoir des connaissances équivalentes

Cette formation ne peut être financée que dans le cadre d'un projet d'entreprise (prise en charge entreprise ou OPCO). Les dossiers à financement personnel et CPF ne sont pas pris en compte.

PUBLIC

Data Engineer

Data Analysts et Data Scientists aspirant à développer des compétences en ingénierie des données

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

PROGRAMME

Introduction

Sujets

Présentation du cours

Actualisation des faisceaux et des flux de données

Objectifs

Présentation des objectifs du cours.

Démontrer comment Apache Beam et Dataflow fonctionnent ensemble pour répondre aux besoins de traitement des données de votre organisation.

Portabilité de Beam

Sujets

Portabilité de Beam

Runner v2

Environnements de conteneurs

Transformations Cross-Language

Objectifs

Résumer les avantages du Beam Portability Framework.

Personnaliser l'environnement de traitement des données de votre pipeline à l'aide de conteneurs personnalisés.

Examiner les cas d'utilisation pour les transformations Cross-Language.

Activez le Beam Portability Framework pour vos pipelines Dataflow.

Activités

Quiz

Séparer le calcul et le stockage avec Dataflow

Sujets

Dataflow

Dataflow Shuffle Service

Dataflow Streaming Engine

Flexible Resource Scheduling

Objectifs

Activez Shuffle et Streaming Engine, respectivement pour les pipelines batch et streaming, pour des performances maximales.

Activez la planification flexible des ressources pour des performances plus rentables.

Activités

Quiz

IAM, Quotas et Permissions

Sujets

IAM

Quota

Objectifs

Sélectionner la bonne combinaison d'autorisations IAM pour votre tâche Dataflow.

Déterminer vos besoins en capacité en inspectant les quotas pertinents pour vos tâches Dataflow.

Activités

Quiz

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

Sécurité

Sujets

Localité des données
Shared VPC
IPs privées
CMEK

Objectifs

Sélectionner votre stratégie de traitement des données zonales à l'aide de Dataflow, en fonction de vos besoins en matière de localisation des données.

Mettre en œuvre les meilleures pratiques pour un environnement de traitement de données sécurisé.

Activités

Lab pratique et quiz

Revue des concepts de BEAM

Sujets

Les bases Beam
Transformations utilitaires
Cycle de vie DoFn

Objectifs

Passer en revue les principaux concepts d'Apache Beam (Pipeline, PCollections, PTransforms, Runner, lecture/écriture, Utility PTransforms, side inputs), les bundles et le cycle de vie DoFn.

Activités

Lab pratique et quiz

Windows, Watermarks, Triggers

Sujets

Windows
Watermarks
Triggers

Objectifs

Implémenter une logique pour gérer vos données tardives.

Passer en revue les différents types de déclencheurs.

Passer en revue les principaux concepts de diffusion en continu (unbounded PCollections, windows).

Activités

Lab pratique et quiz

Sources and Sinks

Sujets

Sources et Sinks
Text IO et File IO
BigQuery IO
PubSub IO
Kafka IO
Bigable IO
Avro IO
Splittable DoFn

Objectifs

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

Écrire sur les IO de votre choix pour votre pipeline Dataflow.

Ajuster votre transformation Source/Sink pour des performances maximales.

Créer des Sources et des sinks personnalisés à l'aide de SDF.

Activités

Quiz

Schémas

Sujets

Beam Schemas

Exemples de code

Objectifs

Introduire des schémas, qui donnent aux développeurs un moyen d'exprimer des données structurées dans leurs pipelines Beam.

Utiliser des schémas pour simplifier votre code Beam et améliorer les performances de votre pipeline.

Activités

Lab pratique et quiz

État et Timers

Sujets

State API

Timer API

Summary

Objectifs

Identifier les cas d'utilisation pour les implémentations d'API d'état et de timer

Sélectionner le bon type d'état et de timers pour votre pipeline

Activités

Quiz

Bonnes pratiques

Sujets

Schémas

Gestion des données non traitables

La gestion des erreurs

Générateur de code AutoValue

Traitement des données JSON

Utiliser le cycle de vie DoFn

Optimisations de pipeline

Objectifs

Implement best practices for Dataflow pipelines.

Activités

Lab pratique et quiz

Dataflow SQL et DataFrames

Sujets

Dataflow et Beam SQL

Windowing in SQL

Beam DataFrames

Objectifs

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

Développer un pipeline Beam en utilisant SQL et DataFrames.

Activités

Lab pratique et quiz

Beam Notebooks

Sujets

Beam Notebooks

Objectifs

Prototyper votre pipeline en Python à l'aide des notebooks Beam.

Lancer une tâche dans Dataflow à partir d'un notebooks.

Activités

Quiz

Monitoring

Sujets

Job List

Job Info

Job Graph

Job Metrics

Metrics Explorer

Objectifs

Accéder à l'interface utilisateur des détails de la tâche Dataflow.

Interpréter les graphiques de métriques de travail pour diagnostiquer les régressions du pipeline.

Définir des alertes sur les tâches Dataflow à l'aide de Cloud Monitoring.

Activités

Quiz

Monitoring

Sujets

Logging

Rapport d'erreur

Objectifs

Utiliser les journaux Dataflow et les widgets de diagnostic pour résoudre les problèmes de pipeline.

Activités

Quiz

Dépannage et débogage

Sujets

Flux de travail de dépannage

Types de problèmes

Objectifs

Utiliser une approche structurée pour déboguer vos pipelines Dataflow.

Examiner les causes courantes des défaillances de pipeline.

Activités

Lab pratique et quiz

Performance

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

Sujets

Conception de pipelines
Forme des données
Source, Sinks et systèmes externes
Shuffle and Streaming Engine

Objectifs

Comprendre les considérations de performances pour les pipelines.
Tenir compte de la façon dont la forme de vos données peut affecter les performances du pipeline.

Activités

Quiz

Testing et CI/CD

Sujets

Présentation des tests et CI/CD
Tests unitaires
Tests d'intégration
Construction d'artefacts
Déploiement

Objectifs

Approches de test pour votre pipeline Dataflow.
Passez en revue les frameworks et les fonctionnalités disponibles pour rationaliser votre flux de travail CI/CD pour les pipelines Dataflow.

Activités

Lab pratique et quiz

Fiabilité

Sujets

Introduction à la fiabilité
Surveillance
Géolocalisation
Reprise après sinistre
Haute disponibilité

Objectifs

Mettre en œuvre les bonnes pratiques en matière de fiabilité pour vos pipelines Dataflow.

Activités

Quiz

Flex Templates

Sujets

Modèles classiques
Modèles flexibles
Utiliser les Flex Templates
Modèles fournis par Google

Objectifs

Utiliser des Flex Templates pour standardiser et réutiliser le code du pipeline Dataflow.

Activités

Lab pratique et quiz

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

Conclusion

Sujets

Synthèse

Objectifs

Récapitulatif rapide des sujets de formation

Formation Google Cloud Platform GCP300DATAFLOW Serverless Data Processing with Dataflow

Durée de la formation : 3 jour(s) / 21 heure(s)

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).

À 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

Notre centre de formation agréé Qualiopi bénéficie de l'agrément d'ENI et ICDL pour les certifications informatiques. C'est pour nos clients la garantie de pouvoir suivre des formations préparant à des certifications professionnelles.

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'imprévu.

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).

Limites 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