

Entrepôt de données avec BigQuery : Conception du stockage, optimisation des requêtes et administration

Durée: 3 Jours **Réf de cours: GO8327** **Version: B** **Méthodes d'apprentissage: Classe à distance**

Résumé:

Dans ce cours, vous découvrirez les composants internes de BigQuery et les bonnes pratiques de conception, d'optimisation et d'administration de votre entrepôt de données.

Grâce à une combinaison de passages théoriques, de démonstrations et d'ateliers, vous découvrirez l'architecture BigQuery et découvrirez comment concevoir un stockage et des schémas optimaux pour l'ingestion et les modifications de données.

Ensuite, vous apprendrez des techniques pour améliorer les performances de lecture, optimiser les requêtes, gérer les charges de travail et utiliser les outils de journalisation et de surveillance. Vous découvrirez également les différents modèles de tarification.

Enfin, vous apprendrez différentes méthodes pour sécuriser les données, automatiser les charges de travail et créer des modèles de machine learning avec BigQuery ML.

Mis à jour 12/02/2025

Classe à Distance - site Client

Cette formation peut être suivie à distance en synchrone depuis n'importe quel site pourvu d'une connexion internet (2 Mb/s en symétrique recommandés). Le programme (théorie et pratique) suit le même déroulé pédagogique qu'en présentiel. La solution technologique adoptée permet aux apprenants à distance de suivre les présentations faites au tableau, de voir et d'entendre l'instructeur et les participants en temps réel, mais également d'échanger avec eux.

Public visé:

Cette formation est destinée aux analystes de données, data scientists, ingénieurs de données et développeurs qui effectuent des travaux à grande échelle nécessitant des connaissances internes avancées de BigQuery pour optimiser les performances.

Objectifs pédagogiques:

- Décrire les principes de base de l'architecture BigQuery.
- Implémenter des modèles de conception de stockage et de schéma pour améliorer les performances.
- Utiliser DML et planifier des transferts de données pour ingérer des données.
- Appliquer les meilleures pratiques pour améliorer l'efficacité de la lecture et optimiser les performances des requêtes.
- Gérer la capacité et automatiser les charges de travail.
- Comprendre les modèles par rapport aux anti-modèles pour optimiser les requêtes et améliorer les performances de lecture.
- Utiliser des outils de journalisation et de surveillance pour comprendre et optimiser les modèles d'utilisation.
- Appliquer les meilleures pratiques de sécurité pour gérer les données et les ressources.
- Créer et déployer plusieurs catégories de modèles de machine learning avec BigQuery ML.

Pré-requis:

- Avoir suivi la formation Google Cloud Platform sur le BigData et le Machine Learning (GO8325), ou posséder une expérience équivalente
- Avoir de l'expérience dans l'utilisation d'un langage de requête de type SQL pour analyser les données
- GO8324 - Les fondamentaux de Google Cloud Platform (CP100A)
- GO8325 - Google Cloud Platform : Fondamentaux du Big Data et du Machine Learning (CPB100)

Contenu:

Principes de base de l'architecture BigQuery

- Introduction
- Infrastructure centrale BigQuery
- Stockage BigQuery
- Traitement des requêtes BigQuery
- Shuffling des données BigQuery

Optimisations de stockage et de schéma

- Stockage BigQuery
- Partitionnement et clustering
- Champs imbriqués et répétés
- Syntaxe ARRAY et STRUCT
- Les meilleures pratiques

Ingestion de données

- Options d'intégration de données
- Ingestion par lots
- Ingestion de diffusion en continu
- Legacy Streaming API
- BigQuery Storage Write API
- Matérialisation des requêtes
- Interroger des sources de données externes
- Service de transfert de données

Modification des données

- Gestion du changement dans les entrepôts de données
- Gestion des Slowly Changing Dimensions (SCD)
- Déclarations DML
- Bonnes pratiques DML et problèmes courants

Améliorer les performances de lecture

- Cache de BigQuery
- Vues matérialisées
- BI Engine
- Lectures à haut débit
- API de lecture de stockage BigQuery

Optimisation et dépannage des requêtes

- Exécution simple des requêtes
- SELECT et Agrégation
- JOIN et JOIN biaisés
- Filtrage et classement
- Meilleures pratiques pour les fonctions

Gestion de la charge de travail et tarification

- Emplacements BigQuery
- Modèles de tarification et estimations
- Réservations de créneaux
- Contrôle des coûts

Journalisation et surveillance

- Cloud Monitoring
- BigQuery Admin Panel
- Cloud Audit Logs
- INFORMATION_SCHEMA
- Chemin de requête et erreurs courantes

Security in BigQuery

- Ressources sécurisées avec IAM
- Vues autorisées
- Données sécurisées avec classification
- Chiffrement
- Découverte et gouvernance des données

Automatisation des charges de travail

- Planifier des requêtes
- Script
- Procédures stockées
- Intégration avec les produits Big Data

Apprentissage automatique dans BigQuery

- Présentation de BigQuery ML
- Comment faire des prédictions avec BigQuery ML
- Comment créer et déployer un système de recommandation avec BigQuery ML
- Comment créer et déployer une solution de prévision de la demande avec BigQuery ML
- Modèles de séries temporelles avec BigQuery ML
- BigQuery ML Explainability

Autres moyens pédagogiques et de suivi:

- Compétence du formateur : Les experts qui animent la formation sont des spécialistes des matières abordées et ont au minimum cinq ans d'expérience d'animation. Nos équipes ont validé à la fois leurs connaissances techniques (certifications le cas échéant) ainsi que leur compétence pédagogique.
- Suivi d'exécution : Une feuille d'émargement par demi-journée de présence est signée par tous les participants et le formateur.
- En fin de formation, le participant est invité à s'auto-évaluer sur l'atteinte des objectifs énoncés, et à répondre à un questionnaire de satisfaction qui sera ensuite étudié par nos équipes pédagogiques en vue de maintenir et d'améliorer la qualité de nos prestations.

Délais d'inscription :

- Vous pouvez vous inscrire sur l'une de nos sessions planifiées en inter-entreprises jusqu'à 5 jours ouvrés avant le début de la formation sous réserve de disponibilité de places et de labs le cas échéant.
- Votre place sera confirmée à la réception d'un devis ou "booking form" signé. Vous recevrez ensuite la convocation et les modalités d'accès en présentiel ou distanciel.
- Attention, si cette formation est éligible au Compte Personnel de Formation, vous devrez respecter un délai minimum et non négociable fixé à 11 jours ouvrés avant le début de la session pour vous inscrire via moncompteformation.gouv.fr.

Accueil des bénéficiaires :

- En cas de handicap : plus d'info sur globalknowledge.fr/handicap
- Le Règlement intérieur est disponible sur globalknowledge.fr/reglement