Entreposage de données avec Google BigQuery : Conception du stockage, optimisation des requêtes et administration (GCP200DWBQ)

Durée: 3 Jours **Réf de cours: GO9096** **Méthodes d'apprentissage: Intra-entreprise & sur-mesure**

Résumé:

Découvrez les composants internes de BigQuery et les meilleures pratiques pour la conception, l'optimisation et l'administration de votre entrepôt de données.

Grâce à une combinaison de conférences, de démonstrations et de travaux pratiques, vous découvrirez l'architecture de BigQuery et 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 diverses méthodes pour sécuriser les données, automatiser les charges de travail et créer des modèles d'apprentissage automatique avec BigQuery ML.
Mis à jour 09/03/2026

Formation intra-entreprise

Cette formation est délivrable en session intra-entreprise, dans vos locaux ou dans les nôtres. Son contenu peut être adapté sur-mesure pour répondre aux besoins de vos collaborateurs. Contactez votre conseiller formation Global Knowledge ou adressez votre demande à info@globalknowledge.fr.

Public visé:

Ce cours s'adresse aux analystes de données, datascientists, ingénieurs spécialisés et développeurs qui effectuent des travaux à une échelle nécessitant une connaissance avancée des composants internes de BigQuery pour optimiser les performances.

Objectifs pédagogiques:

- A l'issue de la formation, les participants seront capables de :
- Décrire les principes fondamentaux de l'architecture BigQuery.
- Mettre en œuvre 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 gouverner les données et les ressources.
- Créer et déployer plusieurs catégories de modèles d'apprentissage automatique avec BigQuery ML.

Pré-requis:

Avoir suivi le cours Google Introduction à l'ingénierie des données (GO9093) ou avoir des connaissances équivalentes

Test et certification

Après cette formation, nous vous conseillons le(s) module(s) suivant(s):

Pas de recommendation

Contenu:

Principes fondamentaux de l'architecture BigQuery

- Expliquer les avantages du stockage en colonnes.
- Comprendre comment BigQuery traite les données.
- Explorer les bases du service de répartition de BigQuery pour améliorer l'efficacité des requêtes.

Optimisation du stockage et du schéma

- Comparer les performances de différents schémas (flocon de neige, dénormalisé, et champs imbriqués et répétés).
- Partitionner et regrouper les données pour de meilleures performances.
- Améliorer la conception du schéma en utilisant des champs imbriqués et répétés.
- Décrire des pratiques exemplaires supplémentaires telles que l'expiration des tables et des partitions

Ingestion de données

- Ingérer des données par lots et en streaming.
- Interroger des sources de données externes.
- Planifier les transferts de données.
- Comprendre comment utiliser l'API Storage Write.

Modification des données

- Écrire des instructions DML.
- Résoudre les problèmes de performance et les goulots d'étranglement courants du DML.
- Identifier les dimensions à évolution lente (SCD) dans vos données et effectuer des mises à jour.

Amélioration des performances de lecture

- Explorer le cache de BigQuery.
- Créer des vues matérialisées.
- Travailler avec BI Engine pour accélérer vos requêtes SQL.
- Utiliser l'API Storage Read pour un accès rapide au stockage géré par BigQuery.
- Expliquer les inconvénients de l'utilisation de sources de données externes.

Optimisation et dépannage des requêtes

- Interpréter les détails d'exécution de BigQuery et le plan de requête.
- Optimiser les performances des requêtes en utilisant des méthodes suggérées pour les instructions et les clauses SQL.
- Démontrer les meilleures pratiques pour les fonctions dans des cas d'utilisation métier.

Gestion de la charge de travail et tarification

- Définir un slot BigQuery.
- Expliquer les modèles de tarification et les estimations de prix (UI BigQuery, bq dry_run, API des tâches).
- Comprendre les réservations, les engagements et les attributions de slots.
- Identifier les meilleures pratiques pour contrôler les coûts.

Journalisation et surveillance

- Utiliser Cloud Monitoring pour afficher les métriques de BigQuery.
- Explorer le panneau d'administration de BigQuery.
- Utiliser les journaux d'audit Cloud.
- Travailler avec les tables INFORMATION_SCHEMA pour obtenir des informations sur vos entités BigQuery.

Sécurité dans BigQuery

- Explorer la découverte de données à l'aide de Data Catalog.
- Discuter de la gouvernance des données à l'aide de l'API DLP et de Data Catalog.
- Créer des politiques IAM (par exemple, des vues autorisées) pour sécuriser les ressources.
- Sécuriser les données avec des classifications (par exemple, des politiques au niveau des lignes).
- Comprendre comment BigQuery utilise le chiffrement.

Automatisation des charges de travail

- Planifier des requêtes.
- Utiliser le scripting et les procédures stockées pour créer des transformations personnalisées.
- Décrire comment intégrer les charges de travail BigQuery avec d'autres produits big data de Google Cloud.

Apprentissage automatique dans BigQuery

- Décrire certaines des différentes applications de BigQuery ML.
- Créer et déployer plusieurs catégories de modèles d'apprentissage automatique avec BigQuery ML.
- Utiliser AutoML Tables pour résoudre des problèmes métier à forte valeur ajoutée.

Méthodes pédagogiques :

Un support de cours officiel sera fourni aux participants. Ce cours comporte des travaux pratiques

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