



ORGANISME DE FORMATION AUX TECHNOLOGIES ET METIERS DE L'INFORMATIQUE

Formation Data Warehousing with BigQuery: Storage Design, Query Optimization, and Administration

N° ACTIVITÉ : 11 92 18558 92

TÉLÉPHONE : 01 85 77 07 07

E-MAIL : inscription@hubformation.com

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.

Objectifs

- | Décrire les principes de base de l'architecture BigQuery
- | Implémenter des modèles de conception de stockage et de schéma pour améliorer les performances
- | Planifier des transferts de données pour ingérer des données
- | Améliorer l'efficacité de la lecture et optimiser les performances des requêtes
- | Gérer la capacité et automatiser les charges de travail
- | Identifier 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
- | Optimiser les modèles d'utilisation
- | Gérer les données et les ressources
- | Déployer plusieurs catégories de modèles de machine learning avec BigQuery ML.

Prérequis

- | Avoir suivi le cours " Google Cloud Platform - Les fondamentaux du Big Data et du Machine Learning" ou avoir des connaissances équivalentes.

Public

- | 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.

Programme de la formation

Principes de base de l'architecture BigQuery

- | Introduction
- | Infrastructure centrale BigQuery
- | Stockage BigQuery
- | Traitement des requêtes BigQuery
- | Shuffling des données BigQuery
- | Expliquer les avantages du stockage en colonne.
- | Comprendre comment BigQuery traite les données.

Référence	GCP200DWBQ
Durée	3 jours (21h)
Tarif	2 100 €HT

SESSIONS PROGRAMMÉES

A DISTANCE (FRA)

- du 15 au 17 mai 2024
- du 16 au 18 octobre 2024
- du 18 au 20 décembre 2024

[VOIR TOUTES LES DATES](#)

| Découvrir les principes de base du service de shuffling de BigQuery pour améliorer l'efficacité des requêtes.

Optimisations de stockage et de schéma

| Stockage BigQuery
| Partitionnement et clustering
| Champs imbriqués et répétés
| Syntaxe ARRAY et STRUCT
| Les meilleures pratiques
| Comparer les performances de différents schémas (snowflake, dénormalisé, et champs imbriqués et répétés).
| Partitionner et regrouper les données pour de meilleures performances.
| Améliorer la conception du schéma à l'aide de champs imbriqués et répétés.
| Décrire les meilleures pratiques supplémentaires telles que l'expiration des tables et des partitions

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

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
| Écrire des instructions DML.
| Résoudre les problèmes de performances et les goulots d'étranglement courants de DML.
| Identifier les Slowly Changing Dimensions (SCD) dans vos données et effectuez des mises à jour.

Améliorer les performances de lecture

| Cache de BigQuery
| Vues matérialisées
| BI Engine
| Lectures à haut débit
| API de lecture de stockage BigQuery
| 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 écueils liés à l'utilisation de sources de données externes.

Optimisation et dépannage des requêtes

| Exécution simple des requêtes
| SELECT et Agrégation
| JOIN et JOIN biaisés
| Filtrage et classement
| Meilleures pratiques pour les fonctions
| Interpréter les détails d'exécution de BigQuery et le plan de requête.
| Optimiser les performances des requêtes en utilisant les méthodes suggérées pour les instructions et les clauses SQL.
| Démontrer les meilleures pratiques pour les fonctions dans les cas d'utilisation métier.

Gestion de la charge de travail et tarification

| Emplacements BigQuery
| Modèles de tarification et estimations
| Réservations de créneaux
| Contrôle des coûts
| Définir un emplacement BigQuery.
| Expliquer les modèles de tarification et les estimations de tarification (interface utilisateur BigQuery, bq dry_run, API jobs).

- | Comprendre les réservations de créneaux, les engagements et les affectations.
- | Identifier les meilleures pratiques pour contrôler les coûts.

Journalisation et surveillance

- | Cloud Monitoring
- | BigQuery Admin Panel
- | Cloud Audit Logs
- | INFORMATION_SCHEMA
- | Chemin de requête et erreurs courantes
- | Utiliser Cloud Monitoring pour afficher les métriques BigQuery.
- | Explorez le BigQuery Admin Panel.
- | Utiliser les Cloud Audit Logs.
- | Utiliser les tables INFORMATION_SCHEMA pour obtenir des informations sur vos entités BigQuery.

Security in BigQuery

- | Ressources sécurisées avec IAM
- | Vues autorisées
- | Données sécurisées avec classification
- | Chiffrement
- | Découverte et gouvernance des données
- | 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 Data Catalog.
- | Créer des stratégies 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
- | Script
- | Procédures stockées
- | Intégration avec les produits Big Data
- | Planifier des requêtes.
- | Utiliser des scripts et des procédures stockées pour créer des transformations personnalisées.
- | Décrire comment intégrer les charges de travail BigQuery à d'autres produits de big data Google Cloud.

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
- | Décrire certaines des différentes applications de BigQuery ML.
- | Créer et déployer plusieurs catégories de modèles de machine learning avec BigQuery ML.
- | Utiliser AutoML Tables pour résoudre des problèmes commerciaux à forte valeur.

Méthode pédagogique

Chaque participant travaille sur un poste informatique qui lui est dédié. Un support de cours lui est remis soit en début soit en fin de cours. La théorie est complétée par des cas pratiques ou exercices corrigés et discutés avec le formateur. Le formateur projette une présentation pour animer la formation et reste disponible pour répondre à toutes les questions.

Méthode d'évaluation

Tout au long de la formation, les exercices et mises en situation permettent de valider et contrôler les acquis du stagiaire. En fin de formation, le stagiaire complète un QCM d'auto-évaluation.

Suivre cette formation à distance

Voici les prérequis techniques pour pouvoir suivre le cours à distance :

- | Un ordinateur avec webcam, micro, haut-parleur et un navigateur (de préférence Chrome ou Firefox). Un casque n'est pas nécessaire suivant l'environnement.

| Une connexion Internet de type ADSL ou supérieure. Attention, une connexion Internet ne permettant pas, par exemple, de recevoir la télévision par Internet, ne sera pas suffisante, cela engendrera des déconnexions intempestives du stagiaire et dérangera toute la classe.

| Privilégier une connexion filaire plutôt que le Wifi.

| Avoir accès au poste depuis lequel vous suivrez le cours à distance au moins 2 jours avant la formation pour effectuer les tests de connexion préalables.

| Votre numéro de téléphone portable (pour l'envoi du mot de passe d'accès aux supports de cours et pour une messagerie instantanée autre que celle intégrée à la classe virtuelle).

| Selon la formation, une configuration spécifique de votre machine peut être attendue, merci de nous contacter.

| Pour les formations incluant le passage d'une certification la dernière journée, un voucher vous est fourni pour passer l'examen en ligne.

| Pour les formations logiciel (Adobe, Microsoft Office...), il est nécessaire d'avoir le logiciel installé sur votre machine, nous ne fournissons pas de licence ou de version test.

| Horaires identiques au présentiel.

Accessibilité

Les sessions de formation se déroulent sur des sites différents selon les villes ou les dates, merci de nous contacter pour vérifier l'accessibilité aux personnes à mobilité réduite.

Pour tout besoin spécifique (vue, audition...), veuillez nous contacter au 01 85 77 07 07.