

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

Cours officiel, préparation aux examens de certification Google Cloud

Cours Pratique de 3 jours - 21h
Réf : DWQ - Prix 2024 : 2 790€ HT

Avec cette formation, vous découvrirez les composants internes de BigQuery et les bonnes pratiques de conception, d'optimisation et d'administration de votre Data Warehouse. Vous découvrirez l'architecture BigQuery et comment concevoir un stockage et des schémas optimaux pour l'ingestion et les modifications de données. 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 et aussi différentes méthodes pour sécuriser les données, etc.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

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

LE PROGRAMME

dernière mise à jour : 10/2023

1) Principes de base de l'architecture BigQuery

- Expliquer les avantages du stockage en colonne.
- Comprendre comment BigQuery traite les données.
- Découvrir les principes de base du service de shuffling de BigQuery pour améliorer l'efficacité des requêtes.

PARTICIPANTS

Data analyst, data scientist, data engineer et développeurs qui effectuent des travaux à grande échelle nécessitant des connaissances avancées de BigQuery pour optimiser les performances.

PRÉREQUIS

Avoir suivi le cours « Google Cloud Platform Big Data and Machine Learning Fundamentals » Réf. GCV 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 stage, 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.

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.

2) Optimisations de stockage et de schéma

- Comparer les performances de différents schémas.
- 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.

3) Ingestion 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.

4) Modification des données

- 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 effectuer des mises à jour.

5) Améliorer les 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 écueils liés à l'utilisation de sources de données externes.

6) Optimisation et dépannage des requêtes

- 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.

7) Gestion de la charge de travail et tarification

- Définir un emplacement BigQuery.
- Expliquer les modèles de tarification et les estimations de tarification.
- Comprendre les réservations de créneaux, les engagements et les affectations.
- Identifier les meilleures pratiques pour contrôler les coûts.

8) Journalisation et surveillance

- Utiliser Cloud Monitoring pour afficher les métriques BigQuery.
- Explorer le BigQuery Admin Panel.
- Utiliser les Cloud Audit Logs.
- Utiliser les tables INFORMATION_SCHEMA pour obtenir des informations sur vos entités BigQuery.

9) Sécurité dans BigQuery

- Explorer la découverte de données à l'aide de Data Catalog.
- Découvrir 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.

10) Automatisation des charges de travail

- 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.

11) Apprentissage automatique dans BigQuery

- 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.

LES DATES

CLASSE À DISTANCE

2024 : 19 nov.