

Mis à jour le 05/09/2023

S'inscrire

Formation BigQuery avancé pour la gestion de datawarehouse

Présentation

Maîtrisez un service de data warehouse simple à configurer et à installer. BigQuery est un outil Google puissant et serverless pour vous permettre de combiner ainsi que d'analyser vos données business.

Durant notre formation BigQuery avancé, nous reviendrons sur la [structure de données Google Cloud Platform](#) pour maîtriser la gestion des organisations, des projets et des ensembles de données. Vous découvrirez la gestion des accès des utilisateurs et nous vous rappellerons les bases du SQL.

Nous vous apprendrons diverses méthodes d'ingestion de données (pour les fichiers CSV, les flux CDC ou le streaming). Vous saurez comment créer des fonctions avancées en SQL (REGEXP, UNNEST ou Window) ou encore l'utilisation des services data GCP (Dataproc ou encore Dataflow).

À la fin de ce cours, vous saurez analyser vos performances via diverses méthodes, comprendre l'architecture d'un datawarehouse efficient ou [optimiser vos coûts](#) via l'implémentation de quotas et de slots.

Objectifs

- Comprendre la structure des données dans GCP (organisation, projet, jeu de données)
- Maîtriser les concepts IAM et l'utilisation de gcloud/bq CLI
- Savoir évaluer les coûts de requêtes et utiliser SQL simple ainsi que les CTE
- Explorer les services de données GCP tels que Dataproc, Dataflow, Cloud Functions / Cloud Run
- Optimiser les performances en mesurant avec Collab, en gérant les clusters et partitions, et en analysant les plans de requêtes

Public visé

- Business analyst
- Financial analyst
- Data analyst
- Data scientist
- Data engineer
- Data miner

Pré-requis

- Connaissance du langage SQL
- Avoir un compte Google avec des crédits GCP disponibles
- Avoir de l'expérience sur BigQuery

Programme de notre formation BigQuery avancé

Rappels rapides

- Structure de données GCP (orga, project, dataset)
- IAM
- gcloud/bq cli
- Query costs
- SQL simple
- CTEs

Ingestion de données

- Fichiers (CSV, parquet)
- Streaming (streaming vs storage write)
- CDC datastreams
- External queries

SQL avancé

- SQL functions (Cast, REGEXP, DATE...)
- Nested fields (UNNEST et ARRAY_AGG)
- Window functions

Services data GCP

- Dataproc
- Dataflow
- Cloud functions / Cloud Run

Performance

- Mesures avec collab
- Cluster et partitions
- Analyse d'un query plan
- Materialized view

Organisation des transformations avec dataform

- Git basique
- Sources et models
- Tests
- Operations
- Includes

Architecture de datawarehouse

- Medaillon
- Star / Snowflake
- Data Vaultn

Gestion des coûts

- Information schema
- Mise en place de Quotas
- Utilisation de slots
- Atelier

Gestion d'accès

- Authorized view
- RLS
- DLP
- Dataplex

Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.