+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated 05/28/2024

Sign up

Advanced BigQuery training for datawarehouse management

3 days (21 hours)

Presentation

Master a data warehouse service that's easy to set up and install. BigQuery is a powerful, serverless Google tool for combining and analyzing your business data.

During our BigQuery advanced training course, we'll go back over the Google Cloud Platform data structure to master the management of organizations, projects and datasets. You'll discover how to manage user access, and we'll remind you of the basics of SQL.

We'll teach you various data ingestion methods (for CSV files, CDC feeds or streaming). You'll also learn how to create advanced SQL functions (REGEXP, UNNEST or Window) and how to use GCP data services (Dataproc or Dataflow).

At the end of this course, you'll be able to analyze your performance using various methods, understand the architecture of an efficient datawarehouse or optimize your costs by implementing quotas and slots.

Objectives

- Understand data structure in GCP (organization, project, dataset)
- Master IAM concepts and the use of gcloud/bq CLI
- Know how to evaluate query costs and use simple SQL and CTE
- Explore GCP data services such as Dataproc, Dataflow, Cloud Functions / Cloud Run
- Optimize performance by measuring with Collab, managing clusters and partitions, and analyzing query plans

Target audience

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

Prerequisites

- Knowledge of SQL language
- Have a Google account with GCP credits available
- Experience with BigQuery

Our BigQuery advanced training program

Quick reminders

- GCP data structure (orga, project, dataset)
- IAM
- gcloud/bq cli
- Query costs
- Simple SQL
- CTEs

Data ingestion

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

Advanced SQL

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

GCP data services

- Dataproc
- Dataflow
- Cloud functions / Cloud Run

Performance

- Collaborative measurements
- Clusters and partitions
- Query plan analysis
- Materialized view

Organizing transformations with dataform

- Basic Git
- Sources and models
- Tests
- Operations
- Includes

Data warehouse architecture

- Medallion
- Star / Snowflake
- Data Vaultn

Cost management

- Information schema
- Setting up Quotas
- Using slots
- Workshop

Access management

- Authorized view
- RLS
- DLP
- Dataplex

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.