COURSE OUTLINE

GC-DWB

DATA WAREHOUSING WITH BIGQUERY: STORAGE DESIGN, QUERY OPTIMIZATION, AND ADMINISTRATIONAL

Google Cloud Partner

DURATION	LEVEL	TECHNOLOGY	DELIVERY METHOD	TRAINING CREDITS
3 Days	Intermediate	Google Cloud	VILT & ILT	NA

INTRODUCTION

In this course, you learn about the internals of BigQuery and best practices for designing, optimizing, and administering your data warehouse. Through a combination of lectures, demos, and labs, you learn about BigQuery architecture and how to design optimal storage and schemas for data ingestion and changes. Next, you learn techniques to improve read performance, optimize queries, manage workloads, and use logging and monitoring tools. You also learn about the different pricing models. Finally, you learn various methods to secure data, automate workloads, and build machine learning models with BigQuery ML.

AUDIENCE PROFILE

Data analysts, data scientists, data engineers, and developers who perform work on a scale that requires advanced BigQuery internals knowledge to optimize performance.

PREREQUISITES

Big Data and Machine Learning Fundamentals

COURSE OBJECTIVES

This course teaches participants the following skills:

- Describe BigQuery architecture fundamentals.
- Implement storage and schema design patterns to improve performance.
- Use DML and schedule data transfers to ingest data.
- Apply best practices to improve read efficiency and optimize query performance.
- Manage capacity and automate workloads.
- Understand patterns versus anti-patterns to optimize queries and improve read performance.
- Use logging and monitoring tools to understand and optimize usage patterns.
- Apply security best practices to govern data and resources.
- Build and deploy several categories of machine learning models with BigQuery ML.

COURSE CONTENT

Lesson 1: BigQuery Architecture Fundamentals

Topics

- Introduction.
- BigQuery Core Infrastructure.
- BigQuery Storage.
- BigQuery Query Processing.
- BigQuery Data Shuffling.

Objectives

- Explain the benefits of columnar storage.
- Understand how BigQuery processes data.
- Explore the basics of BigQuery's shuffling service to improve query efficiency.

Activities

- Labs and demos.

Lesson 2: Storage and Schema Optimizations

Topics

- BigQuery Storage.
- Partitioning and Clustering.
- Nested and Repeated Fields.
- ARRAY and STRUCT syntax.

- Best Practices.

Objectives

- Compare the performance of different schemas (snowflake, denormalized, and nested and repeated fields).
- Partition and cluster data for better performance.
- Improve schema design using nested and repeated fields.
- Describe additional best practices such as table and partition expiration.

Activities

- Labs and demos.

Lesson 3: Ingesting Data

Topics

- Data Ingestion Options.
- Batch Ingestion.
- Streaming Ingestion.
- Legacy Streaming API.
- BigQuery Storage Write API.
- Query Materialization.
- Query External Data Sources.
- Data Transfer Service.

Objectives

- Ingest batch and streaming data.
- Query external data sources.
- Schedule data transfers.
- Understand how to use Storage Write API.

Activities

- Labs and demos.

Lesson 4: Changing Data

Topics

- Managing Change in Data Warehouses.
- Handling Slowly Changing Dimensions (SCD).
- DML statements.
- DML Best Practices and Common Issues.

Objectives

- Write DML statements.
- Address common DML performance problems and bottlenecks.
- Identify slowly changing dimensions (SCD) in your data and make updates.

Lesson 5: Improving Read Performance

Topics

- BigQuery's Cache.
- Materialized Views.
- BI Engine.
- High Throughput Reads.
- BigQuery Storage Read API.

Objectives

- Explore BigQuery's cache.
- Create materialized views.
- Work with BI Engine to accelerate your SQL queries.
- Use the Storage Read API for fast access to BigQuery-managed storage.
- Explain the caveats of using external data sources.

Activities

- Labs and demos

Lesson 6: Optimizing and Troubleshooting Queries

Topics

- Simple Query Execution.
- SELECTs and Aggregation.
- JOINs and Skewed JOINs.
- Filtering and Ordering.
- Best Practices for Functions.

Objectives

- Interpret BigQuery execution details and the query plan.
- Optimize query performance by using suggested methods for SQL statements and clauses.
- Demonstrate best practices for functions in business use cases.

Activities

- Labs and demos.

Lesson 7: Workload Management and Pricing

Topics

- BigQuery Slots.
- Pricing Models and Estimates.
- Slot Reservations.
- Controlling Costs.

Objectives

- Define a BigQuery slot.
- Explain pricing models and pricing estimations (BigQuery UI, bq dry_run, jobs API).
- Understand slot reservations, commitments, and assignments.
- Identify best practices to control costs.

Activities

- Demos.

Lesson 8: Logging and Monitoring

Topics

- Cloud Monitoring.
- BigQuery Admin Panel.
- Cloud Audit Logs.
- INFORMATION_SCHEMA.
- Query Path and Common Errors.

Objectives

- Use Cloud Monitoring to view BigQuery metrics.
- Explore the BigQuery admin panel.
- Use Cloud Audit logs.
- Work with INFORMATION_SCHEMA tables to get insights for your BigQuery entities.

Activities

- Labs and demos.

Lesson 9: Security in BigQuery

Topics

- Secure Resources with IAM.
- Authorized Views.
- Secure Data with Classification.
- Encryption.
- Data Discovery and Governance.

Objectives

- Explore data discovery using Data Catalog.
- Discuss data governance using DLP API and Data Catalog.

- Create IAM policies (e.g., authorized views) to secure resources.
- Secure data with classifications (e.g., row-level policies).
- Understand how BigQuery uses encryption.

Activities

- Labs and demos.

Lesson 10: Automating Workloads

Topics

- Scheduling Queries.
- Scripting.
- Stored Procedures.
- Integration with Big Data Products.

Objectives

- Schedule queries.
- Use scripting and stored procedures to build custom transformations.
- Describe how to integrate BigQuery workloads with other Google Cloud big data products.

Activities

- Demos.

Lesson 11: Machine Learning in BigQuery

Topics

- Introduction to BigQuery ML.
- How to Make Predictions with BigQuery ML.
- How to Build and Deploy a Recommendation System with BigQuery ML.
- How to Build and Deploy a Demand Forecasting Solution with BigQuery ML.
- Time-Series Models with BigQuery ML.
- BigQuery ML Explainability.

Objectives

- Describe some of the different applications of BigQuery ML.
- Build and deploy several categories of machine learning models with BigQuery ML.
- Use AutoML Tables to solve high-value business problems.

Activities

- Labs and demos.

ASSOCIATED CERTIFICATIONS & EXAM

There is no international certification linked to this course currently.