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

Cursusduur: 3 Dagen **Cursuscode: GO9096**

Beschrijving:

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.

Updated 26/02/2026

Doelgroep:

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

Doelstelling:

- **After this course participants should be able to:**
- 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.

Vereiste kennis en vaardigheden:

Basic knowledge of data analytics concepts and familiarity with SQL and Google Cloud fundamentals are recommended.

Examens en certificering

None

Vervolgcursussen:

No recommendation

- GO9093 - Introduction to Data Engineering on Google Cloud

Cursusinhoud:

BigQuery Architecture Fundamentals

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

Storage and Schema Optimizations

- 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

Ingesting Data

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

Changing Data

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

Improving Read Performance

- 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.

Optimizing and Troubleshooting Queries

- 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.

Workload Management and Pricing

- 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.

Logging and Monitoring

- 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.

Security in BigQuery

- 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.

Automating Workloads

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

Machine Learning in BigQuery

- 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

Extra informatie:

Official course book provided to participants.

Nadere informatie:

Neem voor nadere informatie of boekingen contact op met onze Customer Service Desk 030 - 60 89 444

info@globalknowledge.nl

www.globalknowledge.com/nl-nl/

Iepenhoeve 5, 3438 MR Nieuwegein