Training: Google Cloud
Build Data Lakes and Data Warehouses on Google Cloud

Google Cloud

TRAINING GOALS:

While the traditional approaches of using data lakes and data warehouses can be effective, they have shortcomings, particularly in large enterprise environments. This course introduces the concept of a data lakehouse and the Google Cloud products used to create one. A lakehouse architecture uses open-standard data sources and combines the best features of data lakes and data warehouses, which addresses many of their shortcomings.

What you'll learn

- Differentiate between data lakes and data warehouses.
- Explore use-cases for each type of storage and the available data lake and warehouse solutions on Google Cloud.
- Discuss the role of a data engineer and the benefits of a successful data pipeline to business operations.
- Examine why data engineering should be done in a cloud environment.

Audience

This course is designed for Data Engineers, Data Analysts and Data Architects.

CONSPECT:

- Introduction to Modern Data Engineering on Google Cloud
  - Topics
    - The classics: Data lakes and data warehouses
    - The modern approach: Data lake house
    - Choosing the right architecture
  - Objectives

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

        - Compare and contrast data lake, data warehouse, and data lake house architectures
        - Evaluate the benefits of the lake house approach
    - Activities
        - Quiz
- Building a data lake house with Cloud Storage, open formats, and BigQuery
    - Topics
        - Building a data lake foundation
        - Introduction to Apache Iceberg open table format
        - BigQuery as the central processing engine
        - Combining operational data in AlloyDB
        - Combining operational and analytical data with federated queries
        - Real world use case
    - Objectives
        - Discuss data storage options, including Cloud Storage for files, open table formats like Apache Iceberg, BigQuery for analytic data, and AlloyDB for operational data.
        - Understand the role of AlloyDB for operational data use cases.
    - Activities
        - Quiz
        - Lab: Federated Query with BigQuery
- Modernizing Data Warehouses with BigQuery and BigLake
    - Topics
        - BigQuery fundamentals
        - Partitioning and clustering in BigQuery
        - Introducing BigLake and external tables
    - Objectives
        - Explain why BigQuery is a scalable data warehousing solution on Google Cloud.
        - Discuss the core concepts of BigQuery.
        - Understand BigLake's role in creating a unified lakehouse architecture and its integration with BigQuery for external data.
        - Learn how BigQuery natively interacts with Apache Iceberg tables via BigLake.
    - Activities
        - Quiz
        - Lab: Querying External Data and Iceberg Tables
- Advanced lakehouse patterns and data governance
    - Topics
        - Data governance and security in a unified platform

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

        - Demo: Data Loss Prevention
        - Analytics and machine learning on the lakehouse
        - Real-world lakehouse architectures and migration strategies
    - Objectives
        - Implement robust data governance and security practices across the unified data platform, including sensitive data protection and metadata management.
        - Explore advanced analytics and machine learning directly on lakehouse data.
    - Activities
        - Quiz
- Labs and best practices
    - Topics
        - Review
        - Best practices
    - Objectives
        - Reinforce the core principles of Google Cloud's data platform
    - Activities
        - Lab: Getting Started with BigQuery ML
        - Lab: Vector Search with BigQuery

REQUIREMENTS:

- Basic understanding of the principles and activities associated with data engineering
- Familiarity with SQL or data management principles
- Familiarity with data warehouse or data lake architecture

Difficulty level



CERTIFICATE:

The participants will obtain certificates signed by Google Cloud (course completion).

TRAINER:

Authorized Google Cloud Trainer

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007