Training: Google Cloud
Data Engineering on Google Cloud

Google Cloud

TRANING TERMS

2026-07-28 | 4 days | Virtual Classroom
2026-09-01 | 4 days | Virtual Classroom
2026-09-14 | 4 days | Virtual Classroom
2026-10-27 | 4 days | Virtual Classroom
2026-12-01 | 4 days | Virtual Classroom
2026-12-14 | 4 days | Virtual Classroom

TRAINING GOALS:

Get hands-on experience with designing and building data processing systems on Google Cloud. This course uses lectures, demos, and hands-on labs to show you how to design data processing systems, build end-to-end data pipelines, analyze data, and implement machine learning. This course covers structured, unstructured, and streaming data.

What you'll learn

- Design scalable data processing systems in Google Cloud.
- Differentiate data architectures and implement data lake house and pipeline concepts.
- Build and manage robust streaming and batch data pipelines.
- Utilize AI/ML tools to optimize performance and gain process and data insights.

Audience

This course is primarily intended for Data Engineers, Data Analysts, and Data Architects.

Products

- AlloyDB
- BigLake
- BigQuery
- Bigtable
- Cloud Composer

Compendium Education Center Ltd.
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

- Cloud Data Fusion
- Cloud Logging
- Cloud Monitoring
- Dataflow
- Dataform
- Dataplex Universal Catalog
- Dataproc
- Managed Service for Apache Kafka
- Pub/Sub
- Serverless for Apache Spark
- VertexAI

Course structure

This course is comprised of the following four courses:

- Introduction to Data Engineering on Google Cloud
- Build Data Lakes and Data Warehouses with Google Cloud
- Build Batch Data Pipelines on Google Cloud
- Build Streaming Data Pipelines on Google Cloud

CONSPECT:

- Data Engineering Tasks and Components
  - Topics
  - The role of a data engineer
  - Data sources versus data sinks
  - Data formats
  - Storage solution options on Google Cloud
  - Metadata management options on Google Cloud
  - Sharing datasets using Analytics Hub
  - Objectives
    - Explain the role of a data engineer.
    - Understand the differences between a data source and a data sink.
    - Explain the different types of data formats.

Compendium Education Center Ltd.
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

        - Explain the storage solution options on Google Cloud.
        - Learn about the metadata management options on Google Cloud.
        - Understand how to share datasets with ease using Analytics Hub.
        - Understand how to load data into BigQuery using the Google Cloud console or the gcloud CLI.
    - Activities
        - Lab: Loading Data into BigQuery
        - Quiz
- Data Replication and Migration
    - Topics
        - Replication and migration architecture
        - The gcloud command-line tool
        - Moving datasets
        - Datastream
    - Objectives
        - Explain the baseline Google Cloud data replication and migration architecture.
        - Understand the options and use cases for the gcloud command-line tool.
        - Explain the functionality and use cases for Storage Transfer Service.
        - Explain the functionality and use cases for Transfer Appliance.
        - Understand the features and deployment of Datastream.
    - Activities
        - Explain the baseline Google Cloud data replication and migration architecture.
        - Understand the options and use cases for the gcloud command-line tool.
        - Explain the functionality and use cases for Storage Transfer Service.
        - Explain the functionality and use cases for Transfer Appliance.
        - Understand the features and deployment of Datastream.
- The Extract and Load Data Pipeline Pattern
    - Topics
        - Extract and load architecture
        - The bq command-line tool
        - BigQuery Data Transfer Service
        - BigLake
    - Objectives
        - Explain the baseline extract and load architecture diagram.
        - Understand the options of the bq command-line tool.
        - Explain the functionality and use cases for BigQuery Data Transfer Service.

Compendium Education Center Ltd.
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

        - Explain the functionality and use cases for BigLake as a non-extract-load pattern
    - Activities
        - Lab: BigLake: Qwik Start
        - Quiz
- The Extract, Load, and Transform Data Pipeline Pattern
    - Topics
        - Extract, load, and transform (ELT) architecture
        - SQL scripting and scheduling with BigQuery
        - Dataform
    - Objectives
        - Explain the baseline extract, load, and transform architecture diagram.
        - Understand a common ELT pipeline on Google Cloud.
        - Learn about BigQuery's SQL scripting and scheduling capabilities.
        - Explain the functionality and use cases for Dataform.
    - Activities
        - Lab: Create and Execute a SQL Workflow in Dataform
        - Quiz
- The Extract, Transform, and Load Data Pipeline Pattern
    - Topics
        - Extract, transform, and load (ETL) architecture
        - Google Cloud GUI tools for ETL data pipelines
        - Batch data processing using Dataproc
        - Streaming data processing options
        - Bigtable and data pipelines
    - Objectives
        - Explain the baseline extract, transform, and load architecture diagram.
        - Learn about the GUI tools on Google Cloud used for ETL data pipelines.
        - Explain batch data processing using Dataproc.
        - Learn how to use Dataproc Serverless for Spark for ETL.
        - Explain streaming data processing options.
        - Explain the role Bigtable plays in data pipelines.
    - Activities
        - Lab: Use Dataproc Serverless for Spark to Load BigQuery (optional)
        - Lab: Creating a Streaming Data Pipeline for a Real-Time Dashboard with Dataflow
        - Quiz
- Automation Techniques

Compendium Education Center Ltd.
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

- Topics
  - Automation patterns and options for pipelines
  - Cloud Scheduler and Workflows
  - Cloud Composer
  - Cloud Run Functions
  - Eventarc
- Objectives
  - Explain the automation patterns and options available for pipelines.
  - Learn about Cloud Scheduler and Workflows.
  - Learn about Cloud Composer.
  - Learn about Cloud Run functions.
  - Explain the functionality and automation use cases for Eventarc.
- Activities
  - Lab: Use Cloud Run Functions to Load BigQuery (optional)
  - Quiz

- 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
    - 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.

Compendium Education Center Ltd.
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

        - 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
        - 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

Compendium Education Center Ltd.
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

        - Lab: Getting Started with BigQuery ML
        - Lab: Vector Search with BigQuery
- When to choose batch data pipelines
    - Topics
        - Batch data pipelines and their use cases
        - Processing and common challenges
    - Objectives
        - Explain the critical role of a data engineer in developing and maintaining batch data pipelines.
        - Describe the core components and typical lifecycle of batch data pipelines from ingestion to downstream consumption.
        - Analyze common challenges in batch data processing, such as data volume, quality, complexity, and reliability, and identify key Google Cloud services that can address them.
    - Activities
        - Quiz
- Design and Build Scalable Batch Data Pipelines
    - Topics
        - Design batch pipelines
        - Large scale data transformations
        - Dataflow and Serverless for Apache Spark
        - Data connections and orchestration
        - Execute an Apache Spark pipeline
        - Optimize batch pipeline performance
    - Objectives
        - Design scalable batch data pipelines for high-volume data ingestion and transformation.
        - Optimize batch jobs for high throughput and cost-efficiency using various resource management and performance tuning techniques.
    - Activities
        - Quiz
        - Lab: Build a Simple Batch Data Pipeline with Serverless for Apache Spark (optional)
        - Lab: Build a Simple Batch Data Pipeline with Dataflow Job Builder UI (optional)
- Control Data Quality in Batch Data Pipelines
    - Topics
        - Batch data validation and cleansing
        - Log and analyze errors
        - Schema evolution for batch pipelines

Compendium Education Center Ltd.
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

        - Data integrity and duplication
        - Deduplication with Serverless for Apache Spark
        - Deduplication with Dataflow
    - Objectives
        - Develop data validation rules and cleansing logic to ensure data quality within batch pipelines.
        - Implement strategies for managing schema evolution and performing data deduplication in large datasets.
    - Activities
        - Lab: Validate Data Quality in a Batch Pipeline with Serverless for Apache Spark (optional)
        - Quiz
- Orchestrate and Monitor Batch Data Pipelines
    - Topics
        - Orchestration for batch processing
        - Cloud Composer
        - Unified observability
        - Alerts and troubleshooting
        - Visual pipeline management
    - Objectives
        - Orchestrate complex batch data pipeline workflows for efficient scheduling and lineage tracking.
        - Implement robust error handling, monitoring, and observability for batch data pipelines.
    - Activities
        - Lab: Building Batch Pipelines in Cloud Data Fusion
        - Quiz
- Streaming use cases and reference architectures
    - Topics
        - Introduction to streaming data pipelines on Google Cloud
        - Streaming ETL
        - Streaming AI/ML
        - Streaming applications
        - Reverse ETL
    - Objectives
        - Understand various streaming use cases and their applications, including Streaming ETL, Streaming AI/ML, Streaming Application, and Reverse ETL
        - Identify and describe common sample architectures for streaming data, including

Compendium Education Center Ltd.
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

Streaming ETL, Streaming AI/ML, Streaming Application, and Reverse ETL.

- Activities
  - Quiz

- Product deep dives
  - Topics
    - Understanding the products
    - Architectural considerations for Pub/Sub and Managed Service for Apache Kafka
    - Dataflow: The processing powerhouse
    - BigQuery: The analytical engine
    - Bigtable: The solution for operational data
  - Objectives
    - Pub/Sub and Managed Service for Apache Kafka
    - Dataflow
    - BigQuery
    - Bigtable
  - Activities
    - Lab: Stream data with pipelines - Esports use case (optional)
    - Lab: Use Apache Beam and Bigtable to enrich esports downloadable content (DLC) data
    - Lab: Stream e-sports data with Pub/Sub and BigQuery
    - Lab: Monitor e-sports chat with Streamlit

REQUIREMENTS:

- Understanding of data engineering principles, including ETL/ELT processes, data modeling, and common data formats (Avro, Parquet, JSON).
- Familiarity with data architecture concepts, specifically Data Warehouses and Data Lakes.
- Proficiency in SQL for data querying.
- Proficiency in a common programming language (Python recommended).
- Familiarity with using Command Line Interfaces (CLI).
- Familiarity with core Google Cloud concepts and services (Compute, Storage, and Identity management).

Difficulty level

Compendium Education Center Ltd.
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

CERTIFICATE:

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

This course is intended to help you prepare for the Professional Data Engineer certification exam. Google Cloud certification exams are offered at Kryterion test centers worldwide. More information about Professional Data Engineer exam https://cloud.google.com/learn/certification/data-engineer

TRAINER:

Authorized Google Cloud Trainer.

Compendium Education Center Ltd.
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