Training: Google Cloud
Build Batch Data Pipelines on Google Cloud

Google Cloud

TRAINING GOALS:

In this intermediate course, you will learn to design, build, and optimize robust batch data pipelines on Google Cloud. Moving beyond fundamental data handling, you will explore large-scale data transformations and efficient workflow orchestration, essential for timely business intelligence and critical reporting.

Get hands-on practice using Dataflow for Apache Beam and Serverless for Apache Spark (Dataproc Serverless) for implementation, and tackle crucial considerations for data quality, monitoring, and alerting to ensure pipeline reliability and operational excellence. A basic knowledge of data warehousing, ETL/ELT, SQL, Python, and Google Cloud concepts is recommended.

What you'll learn

- Determine whether batch data pipelines are the correct choice for your business use case.
- Design and build scalable batch data pipelines for high-volume ingestion and transformation.
- Implement data quality controls within batch pipelines to ensure data integrity.
- Orchestrate, manage, and monitor batch data pipeline workflows, implementing error handling and observability using logging and monitoring tools.

Audience

This course is for:

- Data Engineers
- Data Analysts

CONSPECT:

- When to choose batch data pipelines
    - Topics
        - You will learn the critical role of a data engineer in developing and maintaining

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

batch data pipelines, understand their core components and lifecycle, and analyze common challenges in batch data processing. You'll also identify key Google Cloud services that address these 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.

- Design and build batch data pipelines
  - Topics
    - You will design scalable batch data pipelines for high-volume data ingestion and transformation. You'll also optimize batch jobs for high throughput and cost-efficiency using various resource management and performance tuning techniques.
  - 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.

- Control data quality in batch data pipelines
  - Topics
    - You will develop data validation rules and cleansing logic to ensure data quality within batch pipelines. You'll also implement strategies for managing schema evolution and performing data deduplication in large datasets.
  - 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.

- Orchestrate and monitor batch data pipelines
  - Topics
    - You will orchestrate complex batch data pipeline workflows for efficient scheduling and lineage tracking. You'll also implement robust error handling, monitoring, and observability for batch data pipelines.
  - Objectives
    - Orchestrate complex batch data pipeline workflows for efficient scheduling and lineage tracking
    - Implement robust error handling, monitoring, and observability for batch data

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

pipelines

REQUIREMENTS:

- Basic proficiency with Data Warehousing and ETL/ELT concepts
- Basic proficiency in SQL
- Basic programming knowledge (Python recommended)
- Familiarity with gcloud CLI and the Google Cloud console
- Familiarity with core Google Cloud concepts and services

Difficulty level



CERTIFICATE:

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

This course is one of the four courses in the series “Data Engineering on Google Cloud” and is intended also to help you prepare for the Google Cloud Professional Data Engineer certification exam. Google Cloud certification exams are offered at Kryterion test centers worldwide. More information about Data Engineer exam https://cloud.google.com/learn/certification/data-engineer/

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