Introduction to Data Engineering on Google Cloud

Duration: 1 Day Course Code: GO9093 Delivery Method: Company Event

Overview:

Learn how to design, build, and manage data processing solutions using Google Cloud services.

In this course, you learn about data engineering on Google Cloud, the roles and responsibilities of data engineers, and how those map to offerings provided by Google Cloud. You also learn about ways to address data engineering challenges.

Products:
- Analytics Hub
- BigQuery
- Storage Transfer Service
- Transfer Appliance
- Datastream
- BigQuery Data Transfer Service
- BigLake
- Dataform
- Dataproc
- Bigtable
- Dataflow
- Cloud Scheduler
- Workflows
- Cloud Composer
- Cloud Run functions
- Eventarc

Updated July 2026

Company Events

These events can be delivered exclusively for your company at our locations or yours, specifically for your delegates and your needs. The Company Events can be tailored or standard course deliveries.

Target Audience:

- Data engineers
- Database administrators
- System administrators

Objectives:

- **After this course participants should be able to:**
- Understand the role of a data engineer.
- Identify data engineering tasks and core components used on Google Cloud.
- Understand how to create and deploy data pipelines of varying patterns on Google Cloud.
- Identify and utilize various automation techniques on Google Cloud.

Prerequisites:

- Prior Google Cloud experience at the fundamental level using Cloud Shell and accessing products from the Google Cloud console.
- Basic proficiency with a common query language such as SQL.
- Experience with data modeling and ETL (extract, transform, load) activities.

Testing and Certification

None

- Experience developing applications using a common programming language such as Python

Follow-on-Courses:

No recommendation

Content:

Module 01: Data Engineering Tasks and Components

- 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.
- 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.
- Lab: Loading Data into BigQuery
- Quiz

Module 02: Data Replication and Migration

- 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.
- Lab: Datastream: PostgreSQL Replication to BigQuery (optional for ILT)
- Quiz

Module 03: The Extract and Load Data Pipeline Pattern

- 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.
- Explain the functionality and use cases for BigLake as a non-extract-load pattern
- Lab: BigLake: Qwik Start
- Quiz

Module 04: The Extract, Load, and Transform Data Pipeline Pattern

- 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.
- Lab: Create and Execute a SQL Workflow in Dataform
- Quiz

Module 05: The Extract, Transform, and Load Data Pipeline Pattern

- 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.
- Lab: Use Dataproc Serverless for Spark to Load BigQuery (optional for ILT)
- Lab: Creating a Streaming Data Pipeline for a Real-Time Dashboard with Dataflow
- Quiz

Module 06: Automation Techniques

- 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.
- Lab: Use Cloud Run Functions to Load BigQuery (optional for ILT)
- Quiz

Additional Information:

Official course book provided to participants

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK