Skillsoft
global knowledge™

DATA ENGINEERING ON GOOGLE CLOUD

Course Code: 5975

Get hands-on experience designing and building data processing systems on Google Cloud.

This 4days-course uses lectures, demos, and hands-on labs to show you how to design data processing systems, build end-to-end data pipelines, and analyze data. This course covers structured, unstructured, and streaming data.

This course consists of the following four modules:

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

What You'll Learn

After this course participants should be able to:

- Design and build data processing systems on Google Cloud.
- Process batch and streaming data by implementing autoscaling data pipelines on Dataflow.
- Derive business insights from extremely large datasets using BigQuery.
- Leverage unstructured data using Spark and ML APIs on Dataproc.
- Enable instant insights from streaming data.

Who Needs to Attend

- Data engineers
- Database administrators
- System administrators

Prerequisites

Participants should have:

- Prior Google Cloud experience 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.
- Experience developing applications using a common programming language such as Python.

DATA ENGINEERING ON GOOGLE CLOUD

Course Code: 5975

VIRTUAL CLASSROOM LIVE	$4,789 CAD	4 Day

Virtual Classroom Live Outline

Data engineering tasks and component

- 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 and/or the gcloud CLI.
  - Lab: Loading Data into BigQuery

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 the Storage Transfer Service.
- Explain the functionality and use cases for the Transfer Appliance.
- Understand the features and deployment of Datastream.
  - Lab: Datastream: PostgreSQL Replication to BigQuery

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

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

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 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
    - Lab: Creating a Streaming Data Pipeline for a Real-Time Dashboard with Dataflow

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

Introduction to data engineering

- Discuss the challenges of data engineering, and how building data pipelines in the cloud helps to address these.
- Review and understand the purpose of a data lake versus a data warehouse, and when to use which.
    - Lab: Using BigQuery to Do Analysis

Build a Data Lake

- Discuss why Cloud Storage is a great option for building a data lake on Google Cloud.
- Explain how to use Cloud SQL for a relational data lake.
    - Lab: Loading Taxi Data into Cloud SQL

Build a data warehouse

- Discuss requirements of a modern warehouse.
- Explain why BigQuery is the scalable data warehousing solution on Google Cloud.
- Discuss the core concepts of BigQuery and review options of loading data into BigQuery.
    - Lab: Working with JSON and Array Data in BigQuery
    - Lab: Partitioned Tables in BigQuery

Introduction to building batch data pipelines

- Review different methods of loading data into your data lakes and warehouses: EL, ELT, and ETL.
- Execute Spark on Dataproc

- Review the Hadoop ecosystem.
- Discuss how to lift and shift your existing Hadoop workloads to the cloud using Dataproc.
- Explain when you would use Cloud Storage instead of HDFS storage.
- Explain how to optimize Dataproc jobs.
    - Lab: Running Apache Spark Jobs on Dataproc

Serverless data processing with Dataflow

- Identify features customers value in Dataflow.
- Discuss core concepts in Dataflow.
- Review the use of Dataflow templates and SQL.
- Write a simple Dataflow pipeline and run it both locally and on the cloud.
- Identify Map and Reduce operations, execute the pipeline, and use command line parameters.
- Read data from BigQuery into Dataflow and use the output of a pipeline as a sideinput to another pipeline.
    - Lab: A Simple Dataflow Pipeline (Python/Java)
    - Lab: MapReduce in Beam (Python/Java)
    - Lab: Side Inputs (Python/Java

Manage data pipelines with Cloud Data Fusion and Cloud Composer

- Discuss how to manage your data pipelines with Cloud Data Fusion and Cloud Composer.
- Summarize how Cloud Data Fusion allows data analysts and ETL developers to wrangle data and build pipelines in a visual way.
- Describe how Cloud Composer can help to orchestrate the work across multiple Google Cloud services.
    - Lab: Building and Executing a Pipeline Graph in Data Fusion
    - Lab: An Introduction to Cloud Composer

Introduction to processing streaming data

- Explain streaming data processing.
- Identify the Google Cloud products and tools that can help address streaming data challenges.
- Serverless messaging with Pub/Sub
- Describe the Pub/Sub service.
- Explain how Pub/Sub works.
- Simulate real-time streaming sensor data using Pub/Sub.
    - Lab: Publish Streaming Data into Pub/Sub

Dataflow streaming features

- Describe the Dataflow service.
- Build a stream processing pipeline for live traffic data.
- Demonstrate how to handle late data using watermarks, triggers, and accumulation.
    - Lab: Streaming Data Pipelines

High-throughput BigQuery and Bigtable streaming features

- Describe how to perform ad-hoc analysis on streaming data using BigQuery and dashboards.
- Discuss Bigtable as a low-latency solution.
- Describe how to architect for Bigtable and how to ingest data into Bigtable.
- Highlight performance considerations for the relevant services.
    - Lab: Streaming Analytics and Dashboards
    - Lab: Generate Personalized Email Content with BigQuery Continuous Queries and Gemini
    - Lab: Streaming Data Pipelines into Bigtable

Advanced BigQuery functionality and performance

- Review some of BigQuery's advanced analysis capabilities.
- Discuss ways to improve query performance.
    - Lab: Optimizing Your BigQuery Queries for Performance

Virtual Classroom Live Labs

- Labs are provided for this course

Aug 25 - 28, 2026 | 9:00 AM - 5:00 PM EDT

DATA ENGINEERING ON GOOGLE CLOUD

Course Code: 5975

PRIVATE GROUP TRAINING	4 Day

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 7/1/2026 7:02:45 AM