Data Engineering on Google Cloud Platform

Category: AI, Data & Analytics | **Vendor:** Google Cloud

Duration: 32.00 hours (4 days) **26.0 CPD Hours** **Rating:** ⭐ 4.6 (5,878 reviews)

Course Information

Delivery Format: Instructor Led - Online

Course Overview

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.

About This Course

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.

Who Should Attend

- Data engineers
- Database administrators
- System administrators

Prerequisites & Entry Requirements

General Prerequisites:

- 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

Learning Outcomes

Upon successful completion of this course, participants will 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.

Detailed Course Outline

Module 01 - Data engineering tasks and components

Topics:

- The role of a data engineer
- Data sources versus data syncs
- Data formats
- Storage solution options on Google Cloud
- Metadata management options on Google Cloud
- Share 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.

- 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.

Activities:

- Lab: Loading Data into BigQuery

Module 02 - 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 the Storage Transfer Service.

- Explain the functionality and use cases for the Transfer Appliance.

- Understand the features and deployment of Datastream.

Activities:

- Lab: Datastream: PostgreSQL Replication to BigQuery

Module 03 - 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 the BigQuery Data Transfer Service.
- Explain the functionality and use cases for BigLake as a non-extract-load pattern.

Activities:

- Lab: BigLake: Qwik Start

Module 04 - 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

Module 05 - 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 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
- Lab: Creating a Streaming Data Pipeline for a Real-Time Dashboard with Dataflow

Module 06 - Automation techniques

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

Module 07 - Introduction to data engineering

Topics:

- Data engineer’s role
- Data engineering challenges
- Introduction to BigQuery
- Data lakes and data warehouses
- Transactional databases versus data warehouses
- Effective partnership with other data teams
- Management of data access and governance
- Building of production-ready pipelines
- Google Cloud customer case study

Objectives:

- 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.

Activities:

- Lab: Using BigQuery to Do Analysis

Module 08 - Build a Data Lake

Topics:

- Introduction to data lakes
- Data storage and ETL options on Google Cloud
- Building of a data lake using Cloud Storage
- Secure Cloud Storage
- Store all sorts of data types
- Cloud SQL as your OLTP system

Objectives:

- 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.

Activities:

- Lab: Loading Taxi Data into Cloud SQL

Module 09 - Build a data warehouse

Topics:

- The modern data warehouse
- Introduction to BigQuery
- Get started with BigQuery
- Loading of data into BigQuery
- Exploration of schemas
- Schema design
- Nested and repeated fields
- Optimization with partitioning and clustering

Objectives:

- 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.

Activities:

- Lab: Working with JSON and Array Data in BigQuery
- Lab: Partitioned Tables in BigQuery

Module 10 - Introduction to building batch data pipelines

Topics:

- EL, ELT, ETL
- Quality considerations
- Ways of executing operations in BigQuery
- Shortcomings

- ETL to solve data quality issues

Objectives:

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

Module 11 - Execute Spark on Dataproc

Topics:

- The Hadoop ecosystem
- Run Hadoop on Dataproc
- Cloud Storage instead of HDFS
- Optimize Dataproc

Objectives:

- 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.

Activities:

- Lab: Running Apache Spark Jobs on Dataproc

Module 12 - Serverless data processing with Dataflow

Topics:

- Introduction to Dataflow
- Reasons why customers value Dataflow
- Dataflow pipelines
- Aggregating with GroupByKey and Combine
- Side inputs and windows
- Dataflow templates

Objectives:

- 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 side-input to another pipeline.

Activities:

- Lab: A Simple Dataflow Pipeline (Python/Java)
- Lab: MapReduce in Beam (Python/Java)
- Lab: Side Inputs (Python/Java)

Module 13 - Manage data pipelines with Cloud Data Fusion and Cloud Composer

Topics:

- Build batch data pipelines visually with Cloud Data Fusion

- Components

- UI overview

- Building a pipeline

- Exploring data using Wrangler

- Orchestrate work between Google Cloud services with Cloud Composer

- Apache Airflow environment

- DAGs and operators

- Workflow scheduling

- Monitoring and logging

Objectives:

- 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.

Activities:

- Lab: Building and Executing a Pipeline Graph in Data Fusion

- Lab: An Introduction to Cloud Composer

Module 14 - Introduction to processing streaming data

Topics:

- Process streaming data

Objectives:

- Explain streaming data processing.
- Identify the Google Cloud products and tools that can help address streaming data challenges.

Module 15 - Serverless messaging with Pub/Sub

Topics:

- Introduction to Pub/Sub
- Pub/Sub push versus pull
- Publishing with Pub/Sub code

Objectives:

- Describe the Pub/Sub service.
- Explain how Pub/Sub works.
- Simulate real-time streaming sensor data using Pub/Sub.

Activities:

- Lab: Publish Streaming Data into Pub/Sub

Module 16 - Dataflow streaming features

Topics:

- Steaming data challenges
- Dataflow windowing

Objectives:

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

Activities:

- Lab: Streaming Data Pipelines

Module 17 - High-throughput BigQuery and Bigtable streaming features

Topics:

- Streaming into BigQuery and visualizing results
- High-throughput streaming with Bigtable
- Optimizing Bigtable performance

Objectives:

- 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.

Activities:

- Lab: Streaming Analytics and Dashboards
- Lab: Generate Personalized Email Content with BigQuery Continuous Queries and Gemini
- Lab: Streaming Data Pipelines into Bigtable

Module 18 - Advanced BigQuery functionality and performance

Topics:

- Analytic window functions
- GIS functions
- Performance considerations

Objectives:

- Review some of BigQuery’s advanced analysis capabilities.
- Discuss ways to improve query performance.

Activities:

- Lab: Optimizing Your BigQuery Queries for Performance

Skills You Will Learn:

- 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.

Frequently Asked Questions

Q: What delivery options are available for Data Engineering on Google Cloud Platform?

We offer multiple delivery formats:

- Live Instructor-Led Classroom Online (Virtual/Live Online)
- Traditional Instructor-Led Classroom Training (ILT)
- On-site delivery at your offices anywhere in United Kingdom
- Private dedicated courses customized for your team

Q: How many CPD hours does this course provide?

The 4-day Data Engineering on Google Cloud Platform course provides up to 26.0 CPD hours of structured learning. CPD certificates can be provided upon request.

Q: What is the duration of the Data Engineering on Google Cloud Platform training?

The training takes place over 4 day(s), with each day lasting approximately 32.00 hours including breaks for lunch and refreshments.

Q: Do you provide corporate training for Data Engineering on Google Cloud Platform?

Yes, we provide corporate training, dedicated training, and closed classes for Data Engineering on Google Cloud Platform. Training can take place anywhere in United Kingdom including London, Manchester, Birmingham, Edinburgh, or live online allowing teams from across United Kingdom or internationally to attend.

Q: What related terms do people search for?

Popular related searches include: Data Engineering on Google Cloud Platform, Data Engineering, DEGCP

Q: Why choose Nexus Human for Data Engineering on Google Cloud Platform?

Nexus Human is recognized as one of the leading training providers. Our trainers have won multiple awards including:

- Small Firms Best Trainer Award
- National Training Partner of the Year (Ireland) - Multiple Years
- Global Top 30 Instructor Awards (2012, 2019, 2021)
- Tech Excellence Award Nominations
- Learning Performance Institute (LPI) External Training Provider Sponsor 2024

Q: Are there any discount codes available?

Yes! Use discount code **PENPAL5** when booking your Data Engineering on Google Cloud Platform training. Please note that only one discount code can be used per booking and cannot be combined with other special offers.

Nexus Human

Professional Training & Development

Email: info@nexushuman.com

Website: www.nexushuman.com

Phone: +353 1 889 8700 (Ireland) | +44 20 8142 8961 (UK)

© 2026 Nexus Human. All rights reserved. This brochure was generated on 13/11/2026.