

Data Engineer

Course Outline

Module 1: Preparing for the Google Cloud Professional Data Engineer

Topics:

- Designing Data Processing Systems
- Building and Operationalizing Data Processing Systems
- Operationalizing Machine Learning Models
- Security, Policy, and Reliability

Module 2: Google Cloud Big Data and Machine Learning Fundamentals

Topics:

- Introduction
- Big Data and Machine Learning on Google Cloud
- Data Engineering for Streaming Data
- Big Data with BigQuery
- Machine Learning Options on Google Cloud
- The Machine Learning Workflow with Vertex AI

Hands-On:

- Vertex AI: Qwik Start
- Exploring a BigQuery Public Dataset
- Vertex AI: Predicting Loan Risk with AutoML

Module 3: Modernizing Data Lakes and Data Warehouses with Google Cloud

Topics:

- Introduction
- Introduction to Data Engineering
- Building a Data Lake
- Building a Data Warehouse

Hands-On:

- BigQuery: Qwik Start - Command Line
- Creating a Data Warehouse Through Joins and Unions
- Build and Execute MySQL, PostgreSQL, and SQLServer to Data Catalog Connectors

Module 4: Building Batch Data Pipelines on Google Cloud

Topics:

- Introduction to Building Batch Data Pipelines
- Executing Spark on Dataproc
- Serverless Data Processing with Dataflow
- Manage Data Pipelines with Cloud Data Fusion and Cloud Composer

Hands-On:

- Dataflow: Qwik Start - Templates
- Dataflow: Qwik Start - Python
- Dataproc: Qwik Start - Console
- Cloud Composer: Copying BigQuery Tables Across Different Locations

Module 5: Building Resilient Streaming Analytics Systems on Google Cloud

Topics:

- Introduction to Processing Streaming Data
- Serverless Messaging with Pub/Sub
- Dataflow Streaming Features
- High-Throughput BigQuery and Bigtable Streaming Features
- Advanced BigQuery Functionality and Performance

Hands-On:

- Building an IoT Analytics Pipeline on Google Cloud
- ETL Processing on Google Cloud Using Dataflow and BigQuery
- Creating Date-Partitioned Tables in BigQuery
- Troubleshooting and Solving Data Join Pitfalls
- Working with JSON, Arrays, and Structs in BigQuery

Module 6: Smart Analytics, Machine Learning, and AI on Google Cloud

Topics:

- Introduction to Analytics and AI
- Prebuilt ML Model APIs for Unstructured Data
- Big Data Analytics with Notebooks
- Production ML Pipelines with Kubeflow
- Custom Model Building with SQL in BigQuery ML
- Custom Model Building with AutoML

Hands-On:

- Dataprep: Qwik Start
- Creating a Data Transformation Pipeline with Cloud Dataprep
- Predict Visitor Purchases with a Classification Model in BQML
- Cloud Natural Language API: Qwik Start
- Google Cloud Speech API: Qwik Start
- Video Intelligence: Qwik Start