# Learning Guide for Professional Data Engineer Certification


Get Trained

Acquire Hands-On Experience

Gain Solution Design Experience

Review Documentation, Blogs and Whitepapers

Prepare for the Exam

Optional: Attend a Certification Preparation Workshop


## Get Trained

- ❏ Review the exam guide and take the practice exam to understand the scope of the certification exam and technical areas to focus.
- ❏ Complete the Data Engineer track trainings either through Coursera or Classroom Instructor-Led offerings. The training curriculum and content is the same across on-demand and instructor-led offerings.

  **On-Demand (Coursera)**

  Complete the Data Engineering on Google Cloud Platform Specialization that consists of the following 5 courses:

  1. Google Cloud Platform Big Data and Machine Learning Fundamentals
  2. Leveraging Unstructured Data with Cloud Dataproc on Google Cloud Platform
  3. Serverless Data Analysis with Google BigQuery and Cloud Dataflow
  4. Serverless Machine Learning with TensorFlow on Google Cloud Platform
  5. Building Resilient Streaming Systems on Google Cloud Platform

  or

  **Classroom Instructor-Led**

  Attend the following 2 classroom offerings:

  1. Google Cloud Platform Fundamentals: Big Data & Machine Learning
  2. Data Engineering on Google Cloud Platform

## Acquire Hands-On Experience

- ❏ Complete a set of self-paced labs around Data Engineering to gain hands-on experience.

  **Qwiklabs Quests**

  Completion of the following Qwiklabs quests are highly recommended:

  1. Advanced: Machine Learning APIs (8 labs)
  2. Advanced: Data Science on the Google Cloud Platform (9 labs)
  3. Advanced: Scientific Data Processing (7 labs)

4. Expert: Google Cloud Solutions II: Data and Machine Learning (10 labs)

## Gain Solution Design Experience

❏ Review the data engineering solutions at Google Cloud Solutions under the categories of data processing, data warehousing, analytics and visualization, IoT, etc.

A. Data Processing
- Data Lifecycle on Google Cloud Platform
- Build a Data Lake on Google Cloud Platform
- Migrating Hadoop Jobs from On-Premises to Google Cloud Platform
- Migrating HDFS Data from On-Premises to Google Cloud Platform
- Architecture: Apache Spark & Hadoop on Google Cloud Platform
- Running RStudio® Server on a Cloud Dataproc Cluster
- Architecture: Complex Event Processing

B. Data Warehouse
- BigQuery for Data Warehouse Practitioners
- Performing ETL from a Relational Database into BigQuery
- Build a Marketing Data Warehouse on Google Cloud Platform

C. Business Intelligence (Analytics and Visualization)
- Creating an Authorized View in BigQuery
- Architecture: Optimizing Large-Scale Ingestion of Analytics Events and Logs
- Real-Time Data Analysis with Kubernetes, Cloud Pub/Sub, and BigQuery
- Building a Mobile Gaming Analytics Platform - a Reference Architecture
- Creating Custom Interactive Dashboards with Bokeh and BigQuery
- Visualizing BigQuery Data Using Google Cloud Datalab
- Visualizing BigQuery Data Using Google Data Studio

D. Machine Learning
- Building a Serverless Machine Learning Model
- Architecture of a Serverless Machine Learning Model
- Using Cloud Dataflow for Batch Predictions with TensorFlow
- Running R at Scale on Compute Engine
- Using Distributed TensorFlow with Cloud ML Engine and Cloud Datalab
- Creating an Object Detection Application Using TensorFlow
- Using Machine Learning on Compute Engine to Make Product Recommendations
- Optical Character Recognition (OCR) Tutorial
- An Image Search Application that Uses the Cloud Vision API
- Considerations for Sensitive Data within Machine Learning Datasets

E. IoT

- Overview of Internet of Things
- Architecture: Real-Time Stream Processing for IoT
- Automating IoT Machine Learning: Bridging Cloud and Device Benefits with Cloud ML Engine
- Oil and Gas Equipment Monitoring and Analytics
- Designing a Connected Vehicle Platform on Cloud IoT Core

## Review Documentation, Blogs and Whitepapers

- ❏ Review the Pricing Calculator, Product Pricing, Cost Comparison Calculator and the Always Free Usage Limits.
- ❏ Read the Google Cloud Platform security whitepapers. For example: Infrastructure Security and Encryption at Rest.
- ❏ Read the Site Reliability Engineering Book, especially the Chapter 25 (Data Processing Pipelines), Chapter 26 (Data Integrity: What You Read Is What You Wrote).
- ❏ Explore the current Google Cloud Platform Marketplace solution offerings.
- ❏ In general, review the Google Cloud Platform Documentation and the Google Cloud Platform Blogs.

## Prepare for the Exam

- ❏ Review the 2 case studies in detail
    1. Flowlogistic
    2. MJTelco
- ❏ Re-take the practice exam

## Optional: Attend a Certification Preparation Workshop

- ❏ Register and attend the 2-day ROI Training: Google Cloud Certification Workshop for Data Engineer (Course 796). This is delivered by ROI Training, a GCP Authorized Training Partner.