Data Engineering on Google Cloud

Course 4331– 32 Hours

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.

On Completion, Delegates 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.
- Understand ML APIs and BigQuery ML, and learn to use AutoML to create powerful models without coding.

Who Should Attend

- Extracting, loading, transforming, cleaning, and validating data.
- Designing pipelines and architectures for data processing.
- Integrating analytics and machine learning capabilities into data pipelines.
- Querying datasets, visualizing query results, and creating reports.

Prerequisites

To benefit from this course, participants should have completed “Google Cloud Big Data and Machine Learning Fundamentals” or have equivalent experience.
Participant should also have:

- Basic proficiency with a common query language such as SQL.
- Experience with data modeling and ETL (extract, transform, load) activities.
- Experience with developing applications using a common programming language such as Python. Familiarity with machine learning and/or statistics.

Course Contents

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