

Google Cloud Fundamentals: Big Data and Machine Learning

Learn via: **Classroom/AFA**

Duration: **1 Day**

Overview

This one-day instructor-led course introduces participants to the big data capabilities of Google Cloud Platform. Through a combination of presentations, demos, and hands-on labs, participants get an overview of the Google Cloud platform and a detailed view of the data processing and machine learning capabilities. This course showcases the ease, flexibility, and power of big data solutions on Google Cloud Platform.

Target Audience

- Data analysts, Data scientists, Business analysts getting started with Google Cloud Platform.
- Individuals responsible for designing pipelines and architectures for data processing, creating and maintaining machine learning and statistical models, querying datasets, visualizing query results and creating reports.
- Executives and IT decision makers evaluating Google Cloud Platform for use by data scientists.

Prerequisites

To get the most out of this course, participants should have:

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

What You Will Learn

- Identify the purpose and value of the key Big Data and Machine Learning products in the Google Cloud Platform.
- Use Cloud SQL and Cloud Dataproc to migrate existing MySQL and Hadoop/Pig/Spark/Hive workloads to Google Cloud Platform.
- Employ BigQuery and Cloud Datalab to carry out interactive data analysis.
- Train and use a neural network using TensorFlow.
- Employ ML APIs.
- Choose between different data processing products on the Google Cloud Platform.

Training Outline

Module 1: Introducing Google Cloud Platform

- Google Platform Fundamentals Overview.
- Google Cloud Platform Data Products and Technology.
- Usage scenarios.
- Lab: Sign up for Google Cloud Platform.

Module 2: Compute and Storage Fundamentals

- CPUs on demand (Compute Engine).
- A global filesystem (Cloud Storage).
- CloudShell.
- Lab: Set up a Ingest-Transform-Publish data processing pipeline.

Module 3: Data Analytics on the Cloud

- Stepping-stones to the cloud.

- Cloud SQL: your SQL database on the cloud.
- Lab: Importing data into CloudSQL and running queries.
- Spark on Dataproc.
- Lab: Machine Learning Recommendations with SparkML.

Module 4: Scaling Data Analysis

- Fast random access.
- Datalab.
- BigQuery.
- Lab: Build machine learning dataset.
- Machine Learning with TensorFlow.
- Lab: Train and use neural network.
- Fully built models for common needs.
- Lab: Employ ML APIs

Module 5: Data Processing Architectures

- Message-oriented architectures with Pub/Sub.
- Creating pipelines with Dataflow.
- Reference architecture for real-time and batch data processing.

Module 6: Summary

- Why GCP?
- Where to go from here
- Additional Resources