

Google Cloud Platform Big Data and Machine Learning Fundamentals

Overview

This course introduces participants to the Big Data and Machine Learning capabilities of Google Cloud Platform (GCP). It provides a quick overview of the Google Cloud Platform and a deeper dive of the data processing capabilities.

Prerequisite Comments

Before enrolling in this course, participants should have roughly one (1) year of experience with one or more of the following: A common query language such as SQL Extract, transform, load activities Data modeling Machine learning and/or statistics Programming in Python

Target Audience

This class is intended for the following:

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.

Course Objectives

This course teaches students the following skills:

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.

Course Outline

1 - Introducing Google Cloud Platform

Google Platform Fundamentals Overview.

Google Cloud Platform Big Data Products.

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.

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 Spark on Dataproc.

4 - Scaling Data Analysis

Fast random access.
Datalab.
BigQuery.
Lab: Build machine learning dataset.

5 - Machine Learning

Machine Learning with TensorFlow.
Lab: Carry out ML with TensorFlow
Pre-built models for common needs.
Lab: Employ ML APIs.

6 - Data Processing Architectures

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

7 - Summary

Why GCP?
Where to go from here
Additional Resources