Course Outline

# Google Cloud Platform Big Data and Machine Learning Fundamentals

## Course Summary

### Description

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.

### Objectives

At the end of this course, students will be able to:

- Identify the purpose and value of the key Big Data and Machine Learning products in the Google Cloud Platform
- Use CloudSQL 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
- Choose between Cloud SQL, BigTable and Datastore
- Train and use a neural network using TensorFlow
- Choose between different data processing products on the Google Cloud Platform

### Topics

- Introducing Google Cloud Platform
- Compute and Storage Fundamentals
- Data Analytics on the Cloud
- Scaling Data Analysis
- Machine Learning
- Data Processing Architectures
- Summary

### Audience

This course is intended for the following participants:

- 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, visualising query results and creating reports
- Executives and IT decision makers evaluating Google Cloud Platform for use by data scientists

### Prerequisites

To get the most of out of this course, you 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

### Duration

One day

Course Outline

# Google Cloud Platform Big Data and Machine Learning Fundamentals

## Course Outline

*I. Introducing Google Cloud Platform*
- A. Google Platform Fundamentals Overview
- B. Google Cloud Platform Big Data Products

*II. Compute and Storage Fundamentals*
- A. CPUs on demand (Compute Engine)
- B. A global filesystem (Cloud Storage)
- C. CloudShell
- D. Lab: Set up a Ingest-Transform-Publish data processing pipeline

*III. Data Analytics on the Cloud*
- A. Stepping-stones to the cloud
- B. Cloud SQL: your SQL database on the cloud
- C. Lab: Importing data into CloudSQL and running queries
- D. Spark on Dataproc
- E. Lab: Machine Learning Recommendations with Spark on Dataproc

*IV. Scaling Data Analysis*
- A. Fast random access
- B. Datalab
- C. BigQuery
- D. Lab: Build machine learning dataset

*V. Machine Learning*
- A. Machine Learning with TensorFlow
- B. Lab: Carry out ML with TensorFlow
- C. Pre-built models for common needs
- D. Lab: Employ ML APIs

*VI. Data Processing Architectures*
- A. Message-oriented architectures with Pub/Sub
- B. Creating pipelines with Dataflow
- C. Reference architecture for real-time and batch data processing

*VII. Summary*
- A. Why GCP?
- B. Where to go from here
- C. Additional Resources