Maslak Mahallesi, Maslak Meydan Sk. No:5. Spring Giz Plaza. Maslak/İstanbul
+90 212 282 7700 - info@bilginc.com

Data Engineering on Google Cloud Platform

Learn via: **Classroom/Virtual**

Duration: **4 Days**

Overview

This four-day instructor-led class provides participants a hands-on introduction to designing and building data processing systems on Google Cloud Platform.

Prerequisites

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

- Completed Google Cloud Fundamentals: Big Data & Machine Learning OR have equivalent experience
- 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

Who Should Attend

This class is intended for experienced developers who are responsible for managing big data transformations including:

- Extracting, Loading, Transforming, cleaning, and validating data
- Designing pipelines and architectures for data processing
- Creating and maintaining machine learning and statistical models
- Querying datasets, visualizing query results and creating reports

What You Will Learn

- Design and build data processing systems on Google Cloud Platform
- Process batch and streaming data by implementing autoscaling data pipelines on Cloud Dataflow
- Derive business insights from extremely large datasets using Google BigQuery
- Train, evaluate and predict using machine learning models using Tensorflow and Cloud ML
- Leverage unstructured data using Spark and ML APIs on Cloud Dataproc
- Enable instant insights from streaming data

Outline

Day 1: Serverless Data Analysis

- Module 1: Serverless data analysis with BigQuery
- Module 2: Serverless, autoscaling data pipelines with Dataflow

Day 2: Leveraging unstructured data

- Module 3: Google Cloud Dataproc Overview
- Module 4: Running Dataproc Jobs
- Module 5: Integrating Dataproc with Google Cloud Platform
- Module 6: Making Sense of Unstructured Data with Google's Machine Learning APIs

Day 3: Serverless Machine Learning

- Module 7: Getting started with Machine Learning
- Module 8: Building ML models with Tensorflow
- Module 9: Scaling ML models with CloudML
- Module 10: Feature Engineering
- Module 11: ML architectures

Day 4: Resilient streaming systems

- Module 12: Need for real-time streaming analytics
- Module 13: Architecture of streaming pipelines
- Module 14: Stream data and events into PubSub
- Module 15: Build a stream processing pipeline
- Module 16: High throughput and low-latency with Bigtable
- Module 17: Building Dashboards