

Google Data Engineer

Duration: 8 Days

Pre-requisites:

- Knowledge on cloud computing, linux and Hadoop HDFS is a must

Training Objective

The training offers a deep understanding of google cloud platform, particularly focusses on architecting for data processing

1. Introduction

Theory, Practice and Tests

Lab: Setting Up A GCP Account

Lab: Using The Cloud Shell

2. Compute

About this section

Compute Options

Google Compute Engine (GCE)

Lab: Creating a VM Instance

More GCE

Lab: Editing a VM Instance

Lab: Creating a VM Instance Using The Command Line

Lab: Creating And Attaching A Persistent Disk

3. Google Container Engine - Kubernetes (GKE)

More GKE

Lab: Creating A Kubernetes Cluster And Deploying A Wordpress Container

App Engine

Contrasting App Engine, Compute Engine and Container Engine

Lab: Deploy And Run An App Engine App

Compute

4. Storage

Storage Options

Quick Take

Cloud Storage

Lab: Working With Cloud Storage Buckets

Lab: Bucket And Object Permissions

Lab: Life cycle Management On Buckets

Fix for AccessDeniedException: 403 Insufficient Permission

Lab: Running A Program On a VM Instance And Storing Results on Cloud Storage

5. Virtual Machines and Images

Live Migration

Machine Types and Billing
Sustained Use and Committed Use Discounts
Rightsizing Recommendations
RAM Disk
Images
Startup Scripts And Baked Images

6. VPCs and Interconnecting Networks

VPCs And Subnets
Global VPCs, Regional Subnets
IP Addresses
Lab: Working with Static IP Addresses
Routes
Firewall Rules
Lab: Working with Firewalls
Lab: Working with Auto Mode and Custom Mode Networks
Lab: Bastion Host

7. Cloud VPN

Lab: Working with Cloud VPN
Cloud Router
Lab: Using Cloud Routers for Dynamic Routing
Dedicated Interconnect Direct and Carrier Peering
Shared VPCs
Lab: Shared VPCs
VPC Network Peering
Lab: VPC Peering
Cloud DNS And Legacy Networks
Networking

8. Managed Instance Groups and Load Balancing

Managed and Unmanaged Instance Groups
Types of Load Balancing
Overview of HTTP(S) Load Balancing
Forwarding Rules Target Proxy and Url Maps
Preview
Backend Service and Backends
Load Distribution and Firewall Rules
Lab: HTTP(S) Load Balancing
Lab: Content Based Load Balancing
SSL Proxy and TCP Proxy Load Balancing
Lab: SSL Proxy Load Balancing
Network Load Balancing
Internal Load Balancing
Autoscalers
Lab: Autoscaling with Managed Instance Groups

9. Ops and Security

StackDriver

StackDriver Logging

Lab: Stackdriver Resource Monitoring

Lab: Stackdriver Error Reporting and Debugging

10. Cloud Deployment Manager

Lab: Using Deployment Manager

Lab: Deployment Manager and Stackdriver

11. Cloud Endpoints

Cloud IAM: User accounts, Service accounts, API Credentials

Cloud IAM: Roles, Identity-Aware Proxy, Best Practices

Lab: Cloud IAM

12. Data Protection

Operations and Security

13. Transfer Service

Lab: Migrating Data Using The Transfer Service

gcloud init

Lab: Cloud Storage Versioning, Directory Sync

14. Cloud SQL, Cloud Spanner ~ OLTP ~ RDBMS

Cloud SQL

Lab: Creating A Cloud SQL Instance

Lab: Running Commands On Cloud SQL Instance

Lab: Bulk Loading Data Into Cloud SQL Tables

15. Cloud Spanner

More Cloud Spanner

Lab: Working With Cloud Spanner

16. BigTable ~ HBase = Columnar Store

BigTable Intro

Columnar Store

Denormalised

Column Families

BigTable Performance

Getting the HBase Prompt

Lab: BigTable demo

17. Datastore ~ Document Database

Datastore

Lab: Datastore demo

18. BigQuery ~ Hive ~ OLAP

BigQuery Intro

BigQuery Advanced

Lab: Loading CSV Data Into Big Query

Lab: Running Queries On Big Query

Lab: Loading JSON Data With Nested Tables

Lab: Public Datasets In Big Query

Lab: Using Big Query Via The Command Line

Lab: Aggregations And Conditionals In Aggregations

Lab: Subqueries And Joins

Lab: Regular Expressions In Legacy SQL

Lab: Using The With Statement For SubQueries

19. Dataflow ~ Apache Beam

About this section

Data Flow Intro

Apache Beam

Lab: Running A Python Data flow Program

Lab: Running A Java Data flow Program

Lab: Implementing Word Count In Dataflow Java

Lab: Executing The Word Count Dataflow

Lab: Executing MapReduce In Dataflow In Python

Lab: Executing MapReduce In Dataflow In Java

20. Dataproc ~ Managed Hadoop

Data Proc

Lab: Creating And Managing A Dataproc Cluster

Lab: Creating A Firewall Rule To Access Dataproc

Lab: Running A PySpark Job On Dataproc

Lab: Running The PySpark REPL Shell And Pig Scripts On Dataproc

Lab: Submitting A Spark Jar To Dataproc

Lab: Working With Dataproc Using The GCloud CLI

21. Pub/Sub for Streaming

Pub Sub

Lab: Working With Pubsub On The Command Line

Lab: Working With PubSub Using The Web Console

Lab: Setting Up A Pubsub Publisher Using The Python Library

Lab: Setting Up A Pubsub Subscriber Using The Python Library

Lab: Publishing Streaming Data Into Pubsub

Lab: Reading Streaming Data From PubSub And Writing To BigQuery

Lab: Executing A Pipeline To Read Streaming Data And Write To BigQuery

Lab: Pubsub Source BigQuery Sink

22. Datalab ~ Jupyter

Data Lab

Lab: Creating And Working On A Datalab Instance

Lab: Importing And Exporting Data Using Datalab

Lab: Using The Charting API In Datalab

23. Composer ~ Airflow

Directed Acyclic Graph (DAG)?

Apache Airflow architecture

Google Cloud Platform: Cloud composer used as Apache Airflow

Understanding Apache Airflow program structure

Lab 1 : Create and submit Apache airflow DAG program

Lab 2: Using Template functionality in Apache Airflow program

Using Variables in Apache Airflow

Lab 3: Calling Bash script in different folder / different machine.

24. Cloud Functions

Virtual Machines - Cloud Functions

What is Cloud Functions?

Architecture of Cloud Function

Use cases of Cloud Functions

Cloud Functions Demo

25. Vision, Translate, NLP and Speech: Trained ML APIs

Lab: Taxicab Prediction - Setting up the dataset

Lab: Taxicab Prediction - Training and Running the model

Lab: The Vision, Translate, NLP and Speech API

Lab: The Vision API for Label and Landmark Detection

26. Additional topics in brief which are prerequisite for this course.

Appendix: Hadoop Ecosystem

Introducing the Hadoop Ecosystem

Hadoop

HDFS

MapReduce

Yarn

Hive

Hive vs. RDBMS

HQL vs. SQL

OLAP in Hive

Windowing Hive

Pig

Spark

Streams Intro

Microbatches

Window Types

Hadoop Ecosystem

Introduction

Theory, Practice and Tests

Lab: Setting Up A GCP Account

Lab: Using The Cloud Shell