Google
Cloud Platform Data Engineer

SimpleGuru

Google Cloud Platform

Cloud Computing Concepts

- ➤ Introduction to Cloud Computing
- ➤ Overview of Cloud Service Models (IaaS, PaaS, SaaS)
- ➤ Overview of Cloud Platforms (AWS, Azure, GCP)
- ➤ Roles and Responsibilities of a Cloud Data Engineer

GCP Overview

- ➤ Introduction to Cloud Computing
- ➤ Overview of Cloud Service Models (IaaS, PaaS, SaaS)
- ➤ Overview of Cloud Platforms (AWS, Azure, GCP)
- ➤ Roles and Responsibilities of a Cloud Data Engineer

Connecting to Data Source

- ➤ Types of Connections
- ➤ Connect to data from File
- ➤ Connect to data from different Data sources
- ➤ Change Data Source Settings
- ➤ Direct Query and Import and Live connection

Google Cloud Storage (GCS)

- ➤ Introduction to Google Cloud Storage
- ➤ Buckets: Create/Delete/Upload using UI, gsutil, and Python
- ➤ Organizing Data: Folders and Object Structure
- ➤ Setting up Python Virtual Environment with GCP Libraries
- ➤ File Handling with Python (Upload/Download/Delete)
- ➤ Data Processing with Pandas on GCS Files
- ➤ Data Conversions and Validations using Pandas

Cloud SQL - PostgreSQL Setup

- ➤ Introduction to Cloud SQL
- ➤ Creating and Configuring PostgreSQL Instances
- ➤ CRUD Operations in Cloud SQL
- ➤ Integration with Python and Pandas
- ➤ Secure Connections and SSL Setup

Databricks on GCP - Big Data Processing

- ➤ Introduction to Databricks on GCP
- ➤ Databricks Architecture and Setup
- ➤ Databricks CLI Setup and Usage
- ➤ DBFS Operations using Spark SQL
- ➤ ELT Pipeline Creation with Databricks Workflows
- ➤ Orchestration and Job Execution
- ➤ Monitoring and Logs in Databricks

Spark with Dataproc and BigQuery

- ➤ Introduction to Google Dataproc
- ➤ Spark Integration with BigQuery
- ➤ Submitting Spark Applications with CLI and Notebooks
- ➤ Running Applications in Client and Cluster Modes
- ➤ Storing Application Code in GCS
- ➤ Running Spark Jobs using Dataproc Web UI and CLI
- ➤ Job Monitoring and Validation

Google Cloud Composer - Orchestration

- ➤ Introduction to Airflow and Cloud Composer
- ➤ Setting up Cloud Composer Environment
- ➤ Authoring and Deploying Airflow DAGs
- ➤ Running Composer DAGs via gcloud
- ➤ Dataproc Integration with Composer
- ➤ Variable Handling and DAG Execution Management

Google BigQuery - Data Warehousing

- ➤ Introduction to Google BigQuery
- ➤ CRUD Operations and SQL Queries
- ➤ Loading Data from GCS and Local Files
- ➤ Partitioned and Clustered Tables
- ➤ External Tables and External Queries
- ➤ Views and Materialized Views
- ➤ Integration with Python and Pandas
- ➤ Connecting BigQuery with PostgreSQL

Google BigQuery - Data Warehousing

- ➤ HDFS and gsutil Overview on Dataproc
- ➤ Local and GCS File Handling in Dataproc
- ➤ Connecting with Pyspark, Scala, and Spark SQL
- ➤ ETL Pipeline Creation
- ➤ Dataproc Job Submission via CLI and UI
- ➤ Workflow Templates and Monitoring

Google Pub/Sub

- Introduction to Google Pub/Sub
- Architecture and Use Cases
- Publishing and Subscribing to Topics
- Streaming to BigQuery
- Integration with Spark and BigQueryDatabricks

Google Bigtable

- Introduction to Google Bigtable
- Architecture and Data Modelling
- Integration with PySpark
- Real-Time Data Storage and Retrieval

Projects, Certification

- End-to-End Data Engineering Project on GCP
- Use Case: ELT Pipeline from GCS to BigQuery using Dataproc
- Real-Time Streaming Project with Pub/Sub and BigQuery
- Hands-on with Composer Orchestration and Monitoring
- Associate Cloud Engineer & Professional Data Engineer Exam Preparation
- Practice Labs

DevOps and CI/CD on GCP

- Introduction to CI/CD Concepts
- Cloud Source Repositories
- Cloud Build and Artifact Registry
- Cloud Deploy
- Automating Pipelines with GitHub Actions
- Deployment Strategies: Blue/Green, Canary
- Pipelines, datasets, triggers, Linked service, notebooks etc.

Monitoring, Logging, and Security

- Cloud Monitoring and Logging
- Cloud Trace, Debugger, and Profiler
- IAM and Security Policies
- Data Encryption and VPC Service Controls
- Security Command Center Overview