Google Cloud Data Engineer Course

Course Duration: 40 Hours **Course code: GCDE**

1. Course Overview

This five-day course focuses on designing, building, and managing data processing systems on Google Cloud Platform (GCP). It provides hands-on experience with data ingestion, transformation, storage, and analysis using services such as BigQuery, Dataflow, Pub/Sub, Dataproc, and Cloud Storage. Learners will gain the skills required to build scalable, reliable, and cost-efficient data pipelines and support data-driven decision-making in modern enterprises.

2. What you'll learn?

By the end of the course, you will be able to:

- Design and implement data pipelines on Google Cloud
- Process batch and streaming data efficiently
- Use BigQuery for data warehousing and analytics
- Work with data ingestion tools like Pub/Sub and Dataflow
- Manage structured and unstructured data storage
- Ensure data quality, security, and governance
- Optimize data processing performance and cost
- Monitor and troubleshoot data pipelines

3. Target Audience

- Data engineers and data analysts
- Cloud engineers and architects
- DevOps professionals working with data pipelines
- Software developers transitioning to data engineering

Phone: +91-999-911-1686 **Mail Us:** info@ssdntech.com

4. Pre-Requisites

Before taking this course, you should have:

- Basic knowledge of SQL and databases
- Understanding of data processing concepts (ETL/ELT)
- Familiarity with programming (Python/Java preferred)
- Basic knowledge of cloud computing concepts

5. Course content

Module 1: Course Introduction

- Introduction and course logistics
- Overview of data engineering on GCP
- Course objectives and lab setup

Module 2: Data Engineering Fundamentals

- Role of a data engineer
- Data pipeline architectures (batch vs streaming)
- Data lifecycle and processing stages
- Introduction to ETL and ELT concepts

Module 3: Google Cloud Platform Overview

- Overview of GCP services for data engineering
- Setting up GCP environment and projects
- Identity and Access Management (IAM)
- Using Cloud Shell and SDK

Module 4: Data Storage Solutions on GCP

- Cloud Storage (object storage)
- BigQuery (data warehouse)
- Cloud SQL and Spanner (relational databases)
- Choosing the right storage solution

Module 5: Data Ingestion Techniques

- Batch data ingestion
- Streaming data ingestion with Pub/Sub
- Data transfer services
- Handling structured and unstructured data

Module 6: Data Processing with Dataflow

- Introduction to Apache Beam
- Building batch and streaming pipelines
- Transformations and windowing
- Error handling in pipelines

Module 7: Data Processing with Dataproc

- Introduction to Dataproc (Hadoop/Spark)
- Running Spark and Hadoop jobs
- Cluster management
- Use cases and comparisons with Dataflow

Module 8: Data Warehousing with BigQuery

- BigQuery architecture and features
- Loading and querying data
- Optimizing queries and performance
- Partitioning and clustering

Module 9: Workflow Orchestration

- Introduction to Cloud Composer (Airflow)
- Designing workflows and DAGs
- Scheduling and monitoring pipelines
- Managing dependencies

Module 10: Data Quality and Governance

- Data validation techniques
- Ensuring data consistency
- Data governance and compliance
- Metadata management

Module 11: Security and Access Control

- IAM roles and permissions
- Securing data at rest and in transit
- Data encryption and masking
- Managing sensitive data

Module 12: Monitoring and Troubleshooting

- Monitoring pipelines using Cloud Monitoring
- Logging with Cloud Logging
- Debugging pipeline failures
- Performance tuning

Module 13: Performance Optimization and Cost Management

- Optimizing data pipelines
- Cost control strategies
- Efficient resource utilization
- Query and storage optimization

Module 14: Real-Time Analytics and Visualization

- Streaming analytics with Pub/Sub and Dataflow
- Integrating with BI tools (Looker, Data Studio)
- Building dashboards and reports
- Use cases for real-time insights

Module 15: Capstone Project and Real-World Implementation

- Designing a complete data pipeline
- Implementing batch and streaming workflows
- Data storage, processing, and visualization
- Final project and assessment

Phone: +91-999-911-1686 **Mail Us:** info@ssdntech.com