

Google Cloud Data Engineering (3 Months)

Duration: 3 Months (~120 Hours)

Mode: Live Online / Classroom

Tools & Technologies: BigQuery, Dataflow, Dataproc, Pub/Sub, Cloud Composer, Cloud Storage, Looker Studio

Certifications Prepared: Google Professional Data Engineer

Syllabus

Week 1: Data Storage & Lakehouse

- Cloud Storage for Data Lake
- Partitioning, lifecycle policies
- Best practices for Data Lakes in GCP
- Hands-on: Setup Data Lake in Cloud Storage
- Assignment

Week 2: BigQuery – Data Warehousing

- BigQuery datasets, tables, partitions
- Loading data from Cloud Storage
- SQL queries in BigQuery
- Partitioned & clustered tables
- Hands-on: Query structured & semi-structured data
- Assignment

Week 3: Data Integration with Dataflow

- Apache Beam concepts (batch & streaming pipelines)
- Dataflow pipelines for ETL
- Hands-on: Transform & load data from Pub/Sub → BigQuery
- Assignment

Week 4: Dataproc for Big Data

- Cluster setup (Spark, Hadoop)
- Submitting Spark jobs
- Integration with Cloud Storage & BigQuery
- Hands-on: ETL pipeline using Dataproc + Spark
- Assignment
- Mock Interview 1

Week 5: Real-Time Data Processing with Pub/Sub

- Pub/Sub topics & subscriptions
- Push vs pull delivery models
- Integrating Pub/Sub with Dataflow
- Hands-on: Real-time streaming pipeline (Pub/Sub → Dataflow → BigQuery)
- Assignment

Week 6: Orchestration with Cloud Composer

- Managed Apache Airflow overview
- DAGs in Airflow for data workflows
- Hands-on: Schedule BigQuery ETL job with Cloud Composer
- Assignment

Week 7: Visualization & BI

- Looker Studio basics (dashboards, data sources)
- Connecting Looker Studio to BigQuery
- Building reports for business insights
- Hands-on: Create interactive BI dashboard
- Assignment

Week 8: Capstone Project & Interview Prep

- Project: Real-time Data Pipeline (Pub/Sub → Dataflow → BigQuery → Looker Studio)
- Assignment
- Mock Interview 2

Learning Outcomes

- Design and manage scalable data lakes with Cloud Storage
- Build data warehouses and run analytics with BigQuery
- Develop ETL and streaming pipelines with Dataflow and Pub/Sub
- Implement Spark-based big data pipelines with Dataproc
- Orchestrate workflows with Cloud Composer and visualize data with Looker Studio
- Prepare for Google Professional Data Engineer certification