Data Engineering on Google Cloud Training

Duration: **4 Day**

BILGINC IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

Overview

Modern organisations depend on scalable data platforms capable of collecting, transforming, storing, and analysing vast amounts of information from multiple sources. Google Cloud provides a comprehensive ecosystem of managed services that enable data engineers to build reliable and high-performance data solutions.

Data Engineering on Google Cloud is a hands-on training programme designed to help participants design, implement, and manage end-to-end data processing solutions using Google Cloud technologies. Through instructor-led demonstrations and practical labs, learners gain experience with data lakes, data warehouses, ETL and ELT architectures, batch processing, streaming analytics, and pipeline automation.

The course covers Google's core data engineering services and prepares participants to develop production-ready data pipelines that support analytics, reporting, and business intelligence initiatives.

Google Cloud Services Covered

- BigQuery
- Bigtable
- Cloud Storage
- Cloud SQL
- Cloud Spanner
- Dataproc
- Dataflow
- Cloud Data Fusion
- Cloud Composer
- Pub/Sub

Prerequisites

Participants are recommended to have:

- Experience using Google Cloud Console and Cloud Shell
- Working knowledge of SQL
- Familiarity with ETL processes and data modelling concepts
- Programming experience in Python or another common language

Data Engineering on Google Cloud Training

Duration: **4 Day**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

Who Should Attend

This programme is intended for:

- Data Engineers
- Database Administrators
- System Administrators
- Cloud Engineers
- Analytics Engineers
- Data Platform Specialists
- Professionals building data solutions on Google Cloud

What You Will Learn

Upon successful completion of this course, participants will be able to:

- Design scalable data processing architectures on Google Cloud.
- Develop batch and streaming data pipelines using managed cloud services.
- Build autoscaling pipelines with Dataflow.
- Analyse massive datasets efficiently using BigQuery.
- Process structured and unstructured data with Apache Spark on Dataproc.
- Implement real-time analytics solutions using streaming technologies.
- Design modern data lakes and cloud data warehouses.
- Select appropriate Google Cloud services for storage, processing, orchestration, and governance.

Outline

Module 1 – Introduction to Data Engineering

- Responsibilities of a Data Engineer
- Data sources and synchronisation
- Data formats
- Google Cloud storage services
- Metadata management
- Sharing datasets through Analytics Hub

Module 2 – Data Replication and Migration

- Replication architectures
- gcloud CLI
- Data migration strategies
- Datastream

Module 3 – Extract and Load Pipelines

- EL architecture
- bq command-line tool
- BigQuery Data Transfer Service

BİLGİNÇ IT Academy

info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

- BigLake

Module 4 – ELT Pipelines

- Extract, Load and Transform architecture
- SQL scripting in BigQuery
- Scheduling jobs
- Dataform

Module 5 – ETL Pipelines

- ETL architecture
- Visual ETL tools
- Batch processing with Dataproc
- Streaming processing options
- Bigtable integration

Module 6 – Pipeline Automation

- Automation patterns
- Cloud Scheduler
- Workflows
- Cloud Composer
- Cloud Run Functions
- Eventarc

Module 7 – Modern Data Engineering Concepts

- Data engineering fundamentals
- Common implementation challenges
- BigQuery overview
- Data lakes and warehouses
- Transactional databases versus analytical platforms
- Data governance
- Access management
- Production-ready pipelines
- Google Cloud customer case studies

Module 8 – Building a Data Lake

- Data lake architecture
- Cloud Storage implementation
- Secure storage
- Managing multiple data formats
- Cloud SQL for transactional workloads

Module 9 – Building a Data Warehouse

- Modern warehouse architecture
- Working with BigQuery
- Loading datasets
- Schema design
- Nested and repeated fields

BİLGİNÇ IT Academy

info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

- Partitioning
- Clustering optimisation

Module 10 – Batch Data Processing

- EL, ELT and ETL approaches
- Data quality management
- BigQuery execution methods
- Performance considerations
- Data quality improvements

Module 11 – Apache Spark with Dataproc

- Hadoop ecosystem
- Running Spark workloads
- Cloud Storage integration
- Dataproc optimisation

Module 12 – Serverless Data Processing

- Dataflow fundamentals
- Pipeline development
- GroupByKey
- Combine
- Side inputs
- Windowing
- Dataflow templates

Module 13 – Pipeline Orchestration

- Cloud Data Fusion
- Pipeline components
- Wrangler
- Cloud Composer
- Apache Airflow
- DAG management
- Workflow scheduling
- Monitoring and logging

Module 14 – Messaging with Pub/Sub

- Pub/Sub architecture
- Push and Pull subscriptions
- Publishing messages

Module 15 – Streaming Data Processing

- Streaming architectures
- Windowing strategies
- Real-time analytics

Module 16 – BigQuery & Bigtable Streaming

BİLGİNÇ IT Academy

info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

- Streaming into BigQuery
- Bigtable optimisation
- High-throughput data ingestion

Module 17 – Advanced BigQuery

- Analytic window functions
- GIS capabilities
- Query and performance optimisation