TechyEdz Solutions

Training | Consulting | Developement | Outsourcing

GCP Professional Data Engineer

9032803832

9032803832

contact@techyedz.com

www.techyedz.com

GCP Professional Data Engineer

Course Overview:

A Professional Data Engineer enables data-driven decision making by collecting, transforming, and publishing data. A Data Engineer should be able to design, build, operationalize, secure, and monitor data processing systems with a particular emphasis on security and compliance; scalability and efficiency; reliability and fidelity; and flexibility and portability. A Data Engineer should also be able to leverage, deploy, and continuously train pre-existing machine learning models.

Course Outline:

1. Designing data processing systems

1.1 Selecting the appropriate storage technologies

- Mapping storage systems to business requirements
- Data modeling
- Tradeoffs involving latency, throughput, transactions
- Distributed systems
- Schema design

1.2 Designing data pipelines

- Data publishing and visualization (e.g., BigQuery)
- Batch and streaming data (e.g., Cloud Dataflow, Cloud Dataproc, Apache Beam, Apache Spark and Hadoop ecosystem, Cloud Pub/Sub, Apache Kafka)
- Online (interactive) vs. batch predictions
- Job automation and orchestration (e.g., Cloud Composer)

1.3 Designing a data processing solution

- Choice of infrastructure
- System availability and fault tolerance
- Use of distributed systems
- Capacity planning
- Hybrid cloud and edge computing
- Architecture options (e.g., message brokers, message queues, middleware, service-oriented architecture, serverless functions)
- At least once, in-order, and exactly once, etc., event processing

1.4 Migrating data warehousing and data processing

- Awareness of current state and how to migrate a design to a future state
- Migrating from on-premises to cloud (Data Transfer Service, Transfer Appliance, Cloud Networking)
- Validating a migration

2. Building and operationalizing data processing systems

2.1 Building and operationalizing storage systems

- Effective use of managed services (Cloud Bigtable, Cloud Spanner, Cloud SQL, BigQuery, Cloud Storage, Cloud Datastore, Cloud Memorystore)
- Storage costs and performance
- Lifecycle management of data

2.2 Building and operationalizing pipelines

- Data cleansing
- Batch and streaming
- Transformation
- Data acquisition and import
- Integrating with new data sources

2.3 Building and operationalizing processing infrastructure

- Provisioning resources
- Monitoring pipelines
- Adjusting pipelines
- Testing and quality control

3. Operationalizing machine learning models

3.1 Leveraging pre-built ML models as a service

- ML APIs (e.g., Vision API, Speech API)
- Customizing ML APIs (e.g., AutoML Vision, Auto ML text)
- Conversational experiences (e.g., Dialogflow)

3.2 Deploying an ML pipeline

- Ingesting appropriate data
- Retraining of machine learning models (Cloud Machine Learning Engine, BigQuery ML, Kubeflow, Spark ML)
- Continuous evaluation

3.3 Choosing the appropriate training and serving infrastructure

- Distributed vs. single machine
- Use of edge compute
- Hardware accelerators (e.g., GPU, TPU)

3.4 Measuring, monitoring, and troubleshooting machine learning models

- Machine learning terminology (e.g., features, labels, models, regression, classification, recommendation, supervised and unsupervised learning, evaluation metrics)
- Impact of dependencies of machine learning models
- Common sources of error (e.g., assumptions about data)

4. Ensuring solution quality

4.1 Designing for security and compliance

- Identity and access management (e.g., Cloud IAM)
- Data security (encryption, key management)
- Ensuring privacy (e.g., Data Loss Prevention API)
- Legal compliance (e.g., Health Insurance Portability and Accountability Act (HIPAA), Children's Online Privacy Protection Act (COPPA), FedRAMP, General Data Protection Regulation (GDPR))

4.2 Ensuring scalability and efficiency

- Building and running test suites
- Pipeline monitoring (e.g., Stackdriver)
- Assessing, troubleshooting, and improving data representations and data processing infrastructure
- Resizing and autoscaling resources

4.3 Ensuring reliability and fidelity

- Performing data preparation and quality control (e.g., Cloud Dataprep)
- Verification and monitoring
- Planning, executing, and stress testing data recovery (fault tolerance, rerunning failed jobs, performing retrospective re-analysis)
- Choosing between ACID, idempotent, eventually consistent requirements

4.4 Ensuring flexibility and portability

- Mapping to current and future business requirements
- Designing for data and application portability (e.g., multi-cloud, data residency requirements)
- Data staging, cataloging, and discovery

Prerequisites:

- If you're already a data scientist, a data engineer, data analyst, machine learning engineer or looking for a career change into the world of data, the Google Cloud Professional Data Engineer Certification is for you.

- Being able to use cloud technologies is becoming a requirement for any kind of data focused role.
- Learners should be familiar with the fundamentals of cloud computing and relevant practical experience

Who Should Attend:

- Google recommends 3+ years of industry experience and 1+ years designing and managing solutions using GCP for professional level certifications.
- Data professionals

Number of Hours: 40hrs

Certification: GCP Professional Cloud Collaboration Engineer (GCP PDE)

Key Features:

- One to One Training
- Online Training
- Fastrack & Normal Track
- Resume Modification
- Mock Interviews
- Video Tutorials
- Training Materials
- Real Time Projects
- Virtual Live Experience
- Preparing for Certification
- Life time Access