RAISETECH

GCP DATA ENGINEER With AI

Design, Build, Automate.
Drive Insights at Scale

MASTER THE FUTURE

End-to-end expertise in building scalable, secure and cost-optimized data solution on Google Cloud platform

CLOUD NATIVE

Leverage the power of Google Cloud services

DATA PIPELINES

Ingest, process and transform data seamlessly

DATA WAREHOUSING

Build scalable warehouses for smarter analytics

ANALYTICS AT SCALE

Deliver insights with high performance and reliability

SECURE & RELIABLE

Enterprise-grade security and governance

#304, 3rd floor, Nilagiri Block
Beside Metro station, Ameerpet, Hyderabad.

96422 10326

RAISETECH

CLOUD DATA ENGINEER – GCP

Course Duration
60 DAYS

Total Sessions Hours
100 HRS

WHAT IS A CLOUD DATA ENGINEER?

- A cloud data engineer is like a swiss army knife in the data space; there are many roles and responsibilities that data engineers are capable of, depending on the particular needs of the organization.
- In short, data engineers set up and operate the organization's data infrastructure preparing it for further analysis by data analysts and scientists.

DATA SOURCES

- Object Store
- File Systems
- Databases
- Cloud Store

DATA ENGINEERING

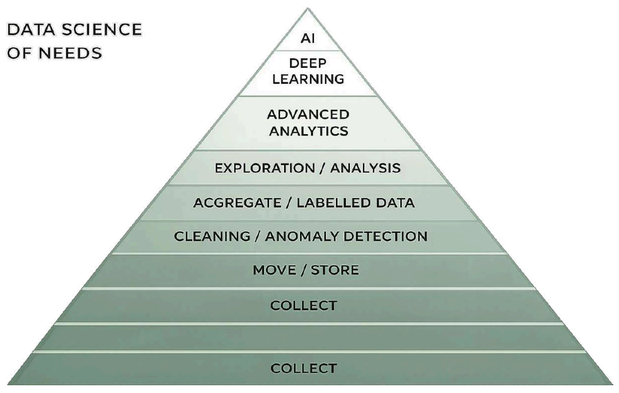
- Data Cleaning
- ETL
- Feature Engineering
- ... etc.

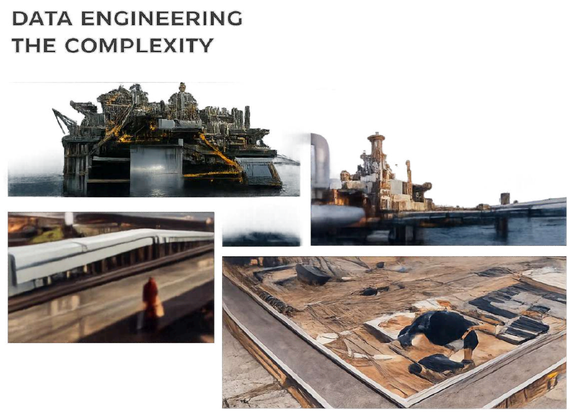
END-USER ENABLEMENT

- Dashboards
- Data Exploration
- Machine Learning

CLOUD DATA ENGINEER JOB DESCRIPTION

- Migrate on-premises corporate applications and related data to the cloud
- Designing and deploying new applications directly in the cloud
- Identifying best practices for cloud services monitoring and management
- Researching and implementing cloud services to support cloud apps and maintain cloud services
- Monitoring cloud app performance for potential bottlenecks and resolving performance issues
- Identifying and implementing cost-saving strategies to reduce ongoing cloud expenses
- Automating key services and tasks across cloud systems to increase efficiency and further reduce cloud costs
- Formulating a recovery plan and executing the plan in the event of cloud downtime or failure
- If you look at the Data Science Hierarchy of Needs, you can grasp a simple idea: The more advanced technologies like machine learning or artificial intelligence are involved, the more complex and resource-heavy data problems become.





"The growing complexity of data engineering compounds the oil industry's infrastructure."

96422 10326

#304, 3rd floor, Nilgiri Block
Beside Metro station, Ameerpet, Hyderabad.

RAISE TECH

CLOUD DATA ENGINEER - GCP

PREREQUISITES

- Basic SQL knowledge
- Any Basic programming Knowledge (Java/Python/C)

WHO CAN ATTEMPT THIS COURSE?

- Database Engineers
- BigData/Hadoop Engineers
- ETL/Data Warehouse Engineers
- Any Application Programmers
- Test Engineers
- Data Analysts

GCP DATA ENGINEERING WITH GCP DATA ANALYTICS

- Introduction to Cloud Computing
- Roles and Responsibilities of Cloud Data Engineer
- Overview of Cloud Platforms
- Overview of Google Cloud Platform
- Overview of Analytics Services on GCP
- Setup GCP for individual Account
- Overview of GCP Project & GCP Credits & Billing
- How to access GCP services with Google Cloud Shell
- How to access GCP services with Google Cloud SDK

GOOGLE CLOUD STORAGE (GCS) [DATALAKE SETUP]

- Introduction to Google Cloud Storage
- Create/Delete/Upload Buckets,Folders,Files using GCS Web UI
- Create/Delete/Upload Buckets,Folders,Files using gsutil commands
- Create/Delete/Upload Buckets,Folders,Files using Python
- Setup Google Cloud Libraries in Python Virtual Environment
- Handling multiple files in GCS using Python
- Data Processing in GCS using Pandas
- Data conversions and Write to GCS using Pandas
- Validate Files in GCS using Python & gsutil & Pandas

GOOGLE BIG QUERY [DWH SETUP]

- Introduction to Google BigQuery
- Overview of CRUD Operations in Google BigQuery
- Merge/Upsert operations into Google BigQuery Tables
- DB operations in Google BigQuery using UI
- Create Table in Google BigQuery using Command
- Overview of Loading Data from Files into BigQuery Tables
- Execution Plan of BigQuery
- Partitioned tables in BigQuery
- Clustered tables in BigQuery
- Google BigQuery External Tables
- External Queries/External Connections on Google BigQuery
- Integration between Google BigQuery and Python
- SQL operations in Google BigQuery [Basics to Advanced]
- Pandas Integration with Google BigQuery
- Postgres DB integrations with BigQuery
- Views & Materialized Views

GCP DATAPROC [BIGDATA PROCESSING]

- Introduction to GCP Dataproc
- Setup Dataproc Cluster for Development
- Overview of HDFS Commands & gsutil on Dataproc
- Handling Local Files in HDFS on Dataproc
- Handling GCS Files in HDFS on Dataproc
- CLI connectivity in Dataproc Cluster using Pyspark
- Scala/Spark SQL
- ETL Datapipeline creation using GCP Dataproc
- GCP Dataproc Jobs using Spark SQL & Scripts
- GCP Dataproc Workflow
- Dataproc Jobs/Workflows handling with
- gcloud Commands Run and Validate ELT Data Pipeline using Dataproc

96422 10326

#304, 3rd floor, Nilgiri Block
Beside Metro station, Ameerpet, Hyderabad.

RAISETECH

CLOUD DATA ENGINEER - GCP

DATAPROC ON GCP [BIGDATA PROCESSING]

- Introduction to Databricks on GCP
- Setup Databricks on GCP
- Databricks Architecture
- Setup Databricks CLI and run Commands
- Data Operations in DBFS using Databricks Spark SQL
- Build ETL Pipeline using Databricks Job in Workflows
- Databricks Workflows
- Create and Run Orchestrated Pipeline using Databricks Job
- Review Execution details of ETL Data Pipeline using Job

GOOGLE PUB/SUB

- Introduction to Google Pub/Sub
- Google Pub/Sub Architecture
- Pub/Sub messages to Pub/Sub
- Stream data from Google Pub/Sub to BigQuery
- Integration between Google Pub/Sub and BigQuery
- Overview of GCP pipelines on GCP

SPARK ON GOOGLE DATAPROC AND BIGQUERY

- Review Spark Google BigQuery Connector
- Spark on Dataproc and BigQuery using PySpark, CLI
- Spark on Dataproc and BigQuery using Notebook
- Spark Application Code to Write to BigQuery Table
- Spark Application PySpark with BigQuery Integration using Client Mode
- Spark Application PySpark with BigQuery Integration using Cluster Mode
- Spark Application deployment with BigQuery Integration in GCS
- Run Spark Application as Dataproc Job using Web UI
- Run Spark Application as Dataproc Job by using Commands
- Review Dataproc Jobs and Spark Application using Dataproc UI

GOOGLE CLOUS COMPOSER [DATAPIPELINE ORCHESTRATOR

- Introduction to Google Cloud Composer
- Setup Airflow on Cloud Composer Environment
- Overview of Airflow Architecture
- Airflow DAG for Cloud Composer
- Deploy and Run First Airflow DAG in Google Cloud Composer
- Run Airflow Commands in Cloud Composer using gcloud
- Integration of GCP Dataproc Workflows using Airflow
- Deploy and Run GCP Dataproc Workflow using Airflow
- Deploy and Run Airflow DAG with Variables
- Deploy Data Pipeline or Airflow DAG using Dataproc Jobs
- Deploy and Run Airflow DAG with Dataproc Jobs
- GOOGLE BIGTABLE
- Introduction to Google BigTable
- Integration between PySpark and BigTable

GOOGLE BIGTABLE

- Introduction to Google BigTable
- Integration between PySpark and BigTable

96422 10326

#304, 3rd floor, Nilgiri Block
Beside Metro station, Ameerpet, Hyderabad.

RAISETECH

CLOUD DATA ENGINEER – GCP

DATAWAREHOUSE CONCEPTS

- Introduction to DWH
- Architecture of DWH
- Difference between OLTP and OLAP
- Dimension and Fact tables
- Types of Dimensions and Facts
- Slowly Changing Dimensions (Type - 1,2,3)

CLOUD RUN & FUNCTIONS

- Introduction to Cloud Run
- Cloud Run Architecture
- Deploying Containerized Applications
- Service Configuration & Revisions
- Traffic Management & Scaling

LOOKER STUDIO

- Overview of Business Intelligence (BI)
- Connecting Data Sources
- Google Sheets Integration
- BigQuery Integration
- Data Blending & Joins

DATABASE CONCEPTS

- SQL Fundamentals
- DDL Statements
- DML Statements
- Logical operations
- Arithmetic operations
- Group & Aggregation functions
- String functions
- Format functions
- Cast functions
- Conditional expressions
- Set Operators (Union, Intersect, Minus)
- Case, Coalesce, Nullify
- Inner join
- Outer Join
- Self Join
- Cross Join
- OLAP Functions
- Rank, Cume, Norm, Ntile (Row Number)

OTHER CONCEPTS

- Agile Process (JIRA, Scrum, Sprint)
- BI IP process – Cock/Sprintz
- Confluence - Documents
- Requirements Understanding
- Go Live/Prod deployment process
- End to End Use cases
- RESUME & Interview PREPARATION

96422 10326

#304, 3rd floor, Nilagiri Block
Beside Metro station, Ameerpet, Hyderabad.

RAISETECH

OUR COURSES

100%
PLACEMENT ASSISTANCE

- Java with AI
- Data Science with AI
- Cyber Security with AI
- Devops / AWS
- Digital Marketing
- WITH Python with AI
- Gcp- Cloud Data Engineer with AI

CLOUD NATIVE
Leverage the power of Google Cloud services

DATA PIPELINES
Ingest, process and transform data seamlessly

DATA WAREHOUSING
Build scalable warehouses for smarter analytics

ANALYTICS AT SCALE
Deliver insights with high performance and reliability

SECURE & RELIABLE
Enterprise-grade security and governance

#304, 3rd floor, Nilagiri Block
Beside Metro station, Ameerpet, Hyderabad.

96422 10326