VERSION IT

VERSION IT

Since 2001

23rd ANNIVERSARY

GCP DATA ENGINEER

Best Training And Placements Institute

211, 2nd Floor, Anapurna Block,
Adithya Enclave, Ameerpet.

+91 9848015399
+91 9391237284

www.versionit.co.in

VERSION IT

About Version IT

Version IT is not a mere software training institute, a team of IT professionals developed it as the best knowledge centre for hundreds of career-building conscious young people. Our training academy is the best training institute in Hyderabad offering various software courses with aptly placement orientation. We proudly announce that we achieved 100% placements in every batch we have taken up in the past two decades. Version IT Academy's strength is our academic excellence with which we have been placed in the top position among the software training institute in Hyderabad.

Corporate Training

Class Room Training

Online Training

Why Choose Us!

VERSION IT

Training By Certified Instuctors

Mock Interviews

Weekly Assignments

Project Training

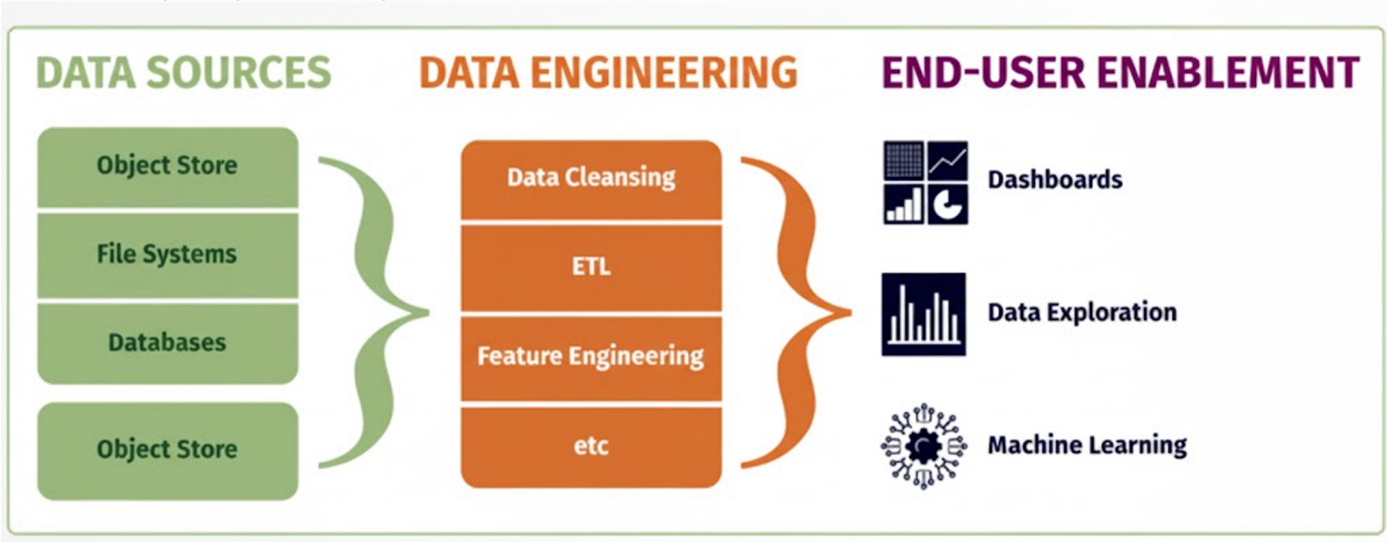
Interview Cracking tips

Resume Preparation

What Is a Cloud DATA ENGINEER?

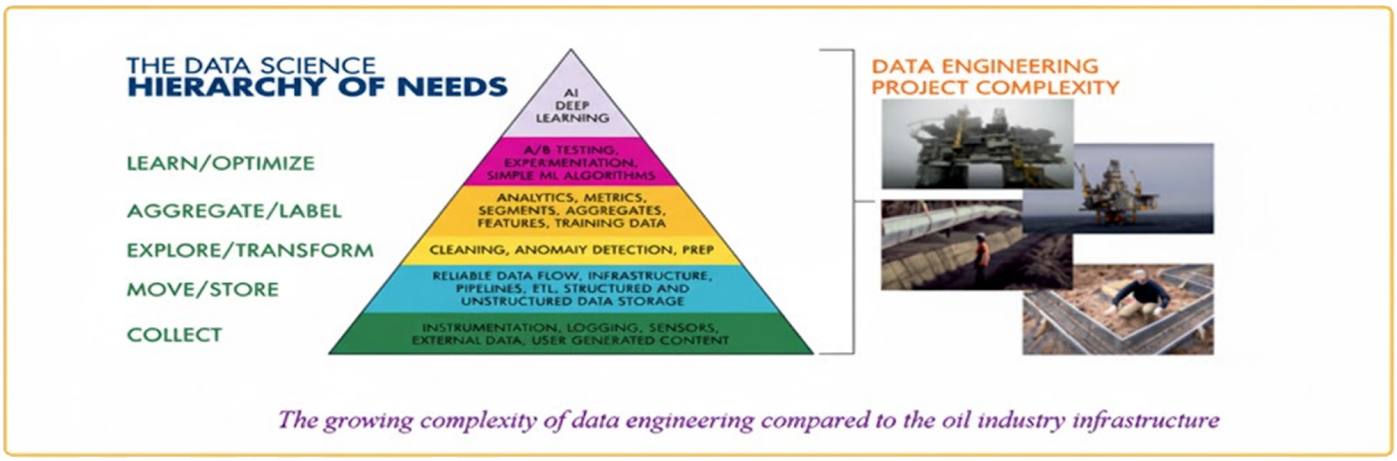
→ A cloud data engineer is like a swiss army knife in the data space; there are many roles and responsibilities that data engineers are capable of, depending on the particular needs of the organization.

→ In short, data engineers set up and operate the organization's data infrastructure preparing it for further analysis by data analysts and scientists



GCP DATA ENGINEER JOB DESCRIPTION

Specific responsibilities expected of a cloud data engineer can include any or all of the following:
→ Migrating 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 and promoting these best practices across the corporation
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 platforms become.



VERSION IT

Prerequisites

- Basic SQL knowledge
- Any Basic programing 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
- Clusterd 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/Spark 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

VERSION IT

Databricks on GCP [Bigdata Processing]

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

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 sybmit with BigQuery Integration using Client Mode
- Spark Application sybmit 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 Cloud Composer [Data Pipeline Orchestration]

- Introduction to Google Cloud Composer
- Setup Airflow or Cloud Composer Environment
- Overview of Airflow Architecture
- Airflow DAGs 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 Workflow using Airflow
- Deploy and Run GCP Dataproc Workflow using Airflow
- Deploy and Run Airflow DAGs 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 Pub/Sub

- Introduction to Google Pub/Sub
- Google Pub/Sub Architecture
- Publish messages to Pub/Sub
- Stream data from Google Pub/Sub to BigQuery
- Integration between Google Pub/Sub and Bigquery and Spark

Overview of CI/CD pipelines on GCP

VERSION IT

Datawarehouse Concepts

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

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, Csum, Msum, Mdiff, Row Number)

BigData Ecosystem Concepts

- Overview of Bigdata Concepts
- Overview of Hadoop Concepts
- HDFS commands
- Introudction to Spark
- Spark Architecture
- MR vs Spark performance comparision
- PySpark Dataframe operations
- PySpark Read and Write operations
- PySpark Transformations
- Handling different types of files using Spark
- Submitting Spark application in client/cluster mode
- Spark SQL
- Spark - Performance Tuning

Other Concepts

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