

CELEBRATING
13 YEARS

Quality Thought®



GCP Cloud Data Engineer



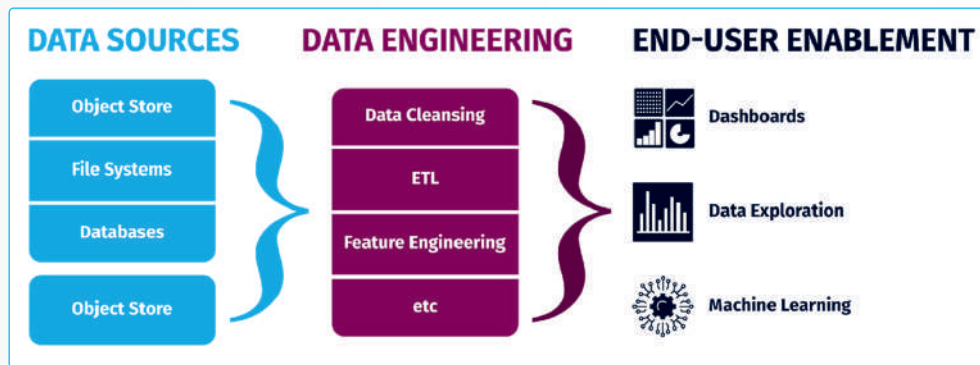
Cloud Data Engineer - GCP

Course Duration
45 Days

total sessions hours
300 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

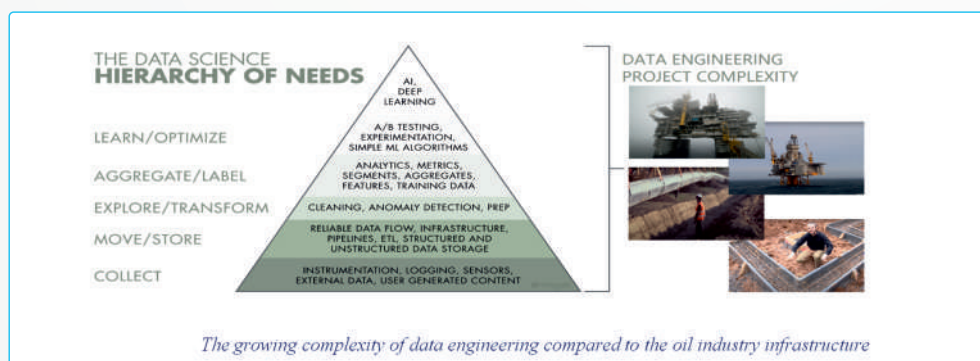


CLOUD DATA ENGINEER JOB DESCRIPTION

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.





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

GCP Cloud SQL [Postgres DB Setup]

- ⇒ Introduction to GCP Cloud SQL
- ⇒ Setup Postgres DB using GCP Cloud SQL
- ⇒ DB operations in GCP Cloud SQL with Postgres
- ⇒ Integration of GCP Cloud SQL Postgres with Python
- ⇒ Integration of GCP Cloud SQL Postgres with 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/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

Databricks 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 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 submit with BigQuery Integration using Client Mode
- ⇒ Spark Application submit 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



Datwarehouse 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)

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
- ⇒ Introduction to Spark
- ⇒ Spark Architecture
- ⇒ MR vs Spark performance comparison
- ⇒ 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

Python Concepts

- ⇒ Introduction to Python
- ⇒ Features of Python
- ⇒ Python Datatypes
- ⇒ Python Collections [List,Set,Tuple,Dictionary]
- ⇒ Python Operators
- ⇒ Python While Loops
- ⇒ Python For Loops
- ⇒ Python Functions
- ⇒ Python Exception Handling
- ⇒ Create/Delete/Write/Read operations on files
- ⇒ Introduction to Pandas
- ⇒ Data structures in Pandas
- ⇒ Series in Pandas
- ⇒ Dataframes in Pandas
- ⇒ Operations on Pandas Dataframes
- ⇒ Overview of other modules in Python

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

Recently Placed Students

NAME	PACKAGE	COMPANY
Annapurna	11 LPA	Capgemini
Harika	8 LPA	TCS
Bhargavi	7 LPA	LTI Mindtree
Kuvalya	8 LPA	IBM
Chandini	11 LPA	LTI Mindtree
Jhansi	12 LPA	Tech M
Dattatreya	11 LPA	LTI Mindtree
Mani rola	11.5 LPA	LTI Mindtree
Prasanthi	11 LPA	Infogain
Sulthan	12.5 LPA	LTI Mindtree
Venkateswarlu	9.5 LPA	Infinity
Anil	7 LPA	Soffility
Bharath	11 LPA	Virtusa
Hareesh	14 LPA	Emids
Sai	9 LPA	TCS
Sai babu	10 LPA	LTI Mindtree
Thousif	12 LPA	Capgemini
Mounika	11 LPA	Capgemini
Nandish	9.5 LPA	foray
Uma mahesh	7.5 LPA	Infosys
Prasanna	11 LPA	Gspann
Haritha	12 LPA	Collabera
Naveen	9 LPA	Replicon
Hari prasad	54 LPA	PWC
Priya	8 LPA	Foray
Vinod	9 LPA	Capgemini
E. Praveen	7.5 LPA	Foray
Divya	12 LPA	Ford
Kiran	11 LPA	Cognizant
Chiranjeevi	10.5 LPA	VHS consultancy
Praneeth	9.5 LPA	Capgemini
Harika	9 LPA	Capgemini
Vijay	14 LPA	L&T
Subbarao	8 LPA	Capgemini
B. Venkatesh	11 LPA	Capgemini
Kishor	12 LPA	Capgemini
Venkat	11.5 LPA	Perimal
Hareesh	10 LPA	Capgemini
Sunitha	9 LPA	Capgemini
Deekshitha	11 LPA	TCS
Sandhya	9 LPA	CSS Global
Kalyan	12 LPA	Capgemini
Neha	9.45 LPA	Capgemini
Uma Lakshmi	9 LPA	HSBC
Sairam	8.5 LPA	Capgemini
K. Ramesh	9 LPA	Cloud4c
Goutham	7 LPA	RNTB
Aneesh	12 LPA	ENY
Vamsi	11 LPA	FORWAY
Ramakrishna	14 LPA	SINGAPORE COMP
Vishwanath	12 LPA	GSPANN

NAME	PACKAGE	COMPANY
Sindhu	9 LPA	Capgemini
teja	9 LPA	HCL
Krishna	11 LPA	IDC Placed
Srikanth	14 LPA	L & T
Veeraj	9 LPA	L & T
Nikhil	8 LPA	Soffility
Kishore b	12 LPA	Mindgraph
Satish s	8 LPA	Mindtree
Ravi b	9 LPA	Soffility
Rakesh	10 LPA	Quantiphi
chandhra	13 LPA	Quantiphi
Siva	14 LPA	Quantiphi
Satish	12 LPA	TCS
Amarnath	11.5 LPA	Vartusa
Prasanna	8 LPA	Gspann
Praveen k	8.5 LPA	TCS
Kishore	8 LPA	Capgemini
Anish	9.45 LPA	Vertusa
satish	9 LPA	TCS
Kishore	8.5 LPA	Capgemini
Yamini	9 LPA	Capgemini
Pasha	7 LPA	Info graph
Vishw a	9 LPA	IBM
Krishna	9 LPA	IDC
Manibau	14 LPA	Capgemini
Raj kumar	11 LPA	LTI Mindtree
Vardhan	22 LPA	CSS Global
Ranjit	9.5 LPA	Capgemini
Ramesh	15 LPA	Cloud4c
Guru	13 LPA	HSBC
Neha	10.5 LPA	Capgemini
Sindhu	9.5 LPA	Capgemini
Satyanarayana	13.5 LPA	Ford
Chara n	12 LPA	LTI Mindtree
Narayana	14 LPA	Gspann
madhu	11 LPA	TCS
Sandha	9 LPA	Capgemini
Ravikumar	12 LPA	LTI Mindtree
santhosh	9 LPA	TCS
Vernkateswar	8.5 LPA	Infinity
Veraju	11 LPA	LTI
Inti anil	7 LPA	Soffility
Sudha	7.5 LPA	Tech M
Sunanda	35 LPA	Motivity Labs
Harish	14 LPA	Emids
Bharath	9 LPA	LIT
Ashok	15 LPA	Cyron Cloud
Anitha	10 LPA	COGNIZENT
Ram	12 LPA	ANAL YBLI CS
Aravind	14 LPA	SONATA
Sairam	11.5 LPA	AG LISIUM

STUDENT TRANSFORMATION STAGES



OUR STUDENTS ARE PLACED IN



QualityThought

90595 14148
1800 309 2939

Quality Thought Infosystems India (P) Ltd.

#302, Nilgiri Block, Ameerpet, Hyderabad-500016 | www.qualitythought.in | info@qualitythought.in