CELEBRATING
14 YEARS

QualityThought®
Transforming Dreams! Redefining Future!

GCP

Cloud Data Engineer
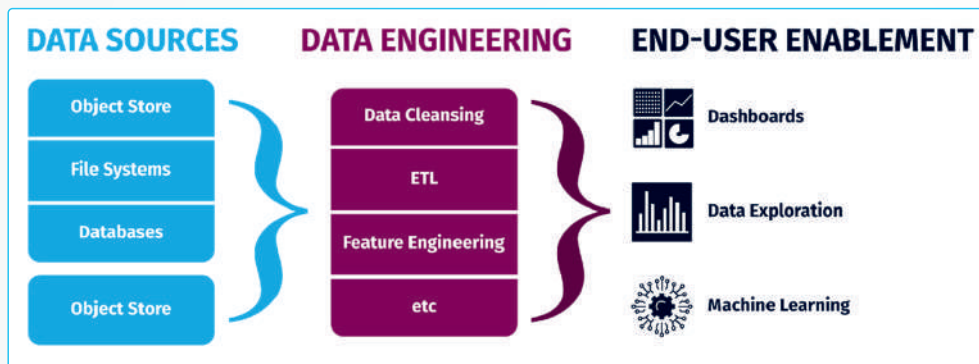
Google Cloud Certified
PROFESSIONAL
DATA ENGINEER

Google Cloud Certified
PROFESSIONAL
DATA ENGINEER

Cloud Data Engineer - GCP

Course Duration	total sessions hours
60 Days	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
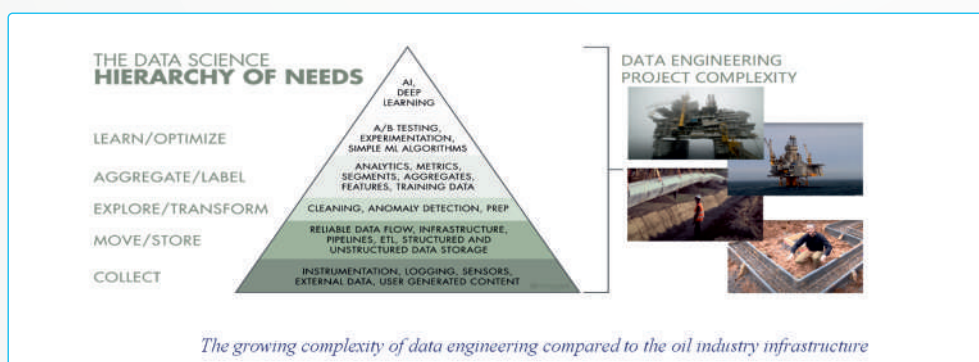

CLOUD DATA ENGINEER JOB DESCRIPTION

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.

The growing complexity of data engineering compared to the oil industry infrastructure

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

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

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

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

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)

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

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

NAME	PACKAGE	COMPANY	NAME	PACKAGE	COMPANY
Annapurna	11 LPA	Capgemini	Sindhu	9 LPA	Capgemini
Harika	8 LPA	TCS	teja	9 LPA	HCL
Bhargavi	7 LPA	LTI Mindtree	Krishna	11 LPA	IDC Placed
Kuvalya	8 LPA	IBM	Srikanth	14 LPA	L & T
Chandini	11 LPA	LTI Mindtree	Veeraj	9 LPA	L & T
Jhansi	12 LPA	Tech M	Nikhil	8 LPA	Softility
Dattatreya	11LPA	LTI Mindtree	Kishore b	12 LPA	Mindgraph
Mani rola	11.5 LPA	LTI Mindtree	Satish s	8 LPA	Mindtree
Prasanthi	11 LPA	Infogain	Ravi b	9 LPA	Softility
Sulthan	12.5 LPA	LTI Mindtree	Rakesh	10 LPA	Quantiphi
Venkateswarlu	9.5 LPA	Infinity	chandhra	13 LPA	Quantiphi
Anil	7 LPA	Softility	Siva	14 LPA	Quantiphi
Bharath	11 LPA	Virtusa	Satish	12 LPA	TCS
Hareesh	14 LPA	Emids	Amarnath	11.5 LPA	Vartusa
Sai	9 LPA	TCS	Prasanna	8 LPA	Gspann
Sai babu	10 LPA	LTI Mindtree	Praveen k	8.5 LPA	TCS
Thousif	12 LPA	Capgemini	Kishore	8 LPA	Capgemini
Mounika	11 LPA	Capgemini	Anish	9.45 LPA	Vertusa
Nandish	9.5 LPA	foray	satish	9 LPA	TCS
Uma mahesh	7.5 LPA	Infosys	Kishore	8.5 LPA	Capgemini
Prasanna	11 LPA	Gspann	Yamini	9 LPA	Capgemini
Haritha	12 LPA	Collabera	Pasha	7 LPA	Info graph
Naveen	9 LPA	Replicon	Vishw a	9 LPA	IBM
Hari prasad	54 LPA	PWC	Krishna	9 LPA	IDC
Priya	8 LPA	Foray	Manibau	14 LPA	Capgemini
Vinod	9 LPA	Capgemini	Raj kumar	11 LPA	LTI Mindtree
E. Praveen	7.5 LPA	Foray	Vardhan	22 LPA	CSS Global
Divya	12 LPA	Ford	Ranjit	9.5 LPA	Capgemini
Kiran	11 LPA	Cognizant	Ramesh	15 LPA	Cloud4c
Chiranjeevi	10.5 LPA	VHS consultancy	Guru	13 LPA	HSBC
Praneeth	9.5 LPA	Capgemini	Neha	10.5 LPA	Capgemini
Harika	9 LPA	Capgemini	Sindhu	9.5 LPA	Capgemini
Vijay	14 LPA	L&T	Satyanarayana	13.5 LPA	Ford
Subbarao	8 LPA	Capgemini	Chara n	12 LPA	LTI Mindtree
B. Venkatesh	11 LPA	Capgemini	Narayana	14 LPA	Gspann
Kishor	12 LPA	Capgemini	madhu	11 LPA	TCS
Venkat	11.5 LPA	Perimal	Sandha	9 LPA	Capgemini
Hareesh	10 LPA	Capgemini	Ravikumar	12 LPA	LTI Mindtree
Sunitha	9 LPA	Capgemini	santhosh	9 LPA	TCS
Deekshitha	11 LPA	TCS	Vernkateswar	8.5 LPA	Infinity
Sandhya	9 LPA	CSS Global	Veraju	11 LPA	LTI
Kalyan	12 LPA	Capgemini	Inti anil	7 LPA	Softility
Neha	9.45 LPA	Capgemini	Sudha	7.5 LPA	Tech M
Uma Lakshmi	9 LPA	HSBC	Sunanda	35 LPA	Motivity Labs
Sairam	8.5 LPA	Capgemini	Harish	14 LPA	Emids
K. Ramesh	9 LPA	Cloud4c	Bharath	9 LPA	LIT
Goutham	7 LPA	RNTB	Ashok	15 LPA	Cyron Cloud
Aneesh	12 LPA	ENY	Anitha	10 LPA	COGNIZENT
Vamsi	11 LPA	FORWAY	Ram	12 LPA	ANAL YBLI CS
Ramakrishna	14 LPA	SINGPORE COMP	Aravind	14 LPA	SONATA
Vishwanath	12 LPA	GSPANN	Sairam	11.5 LPA	AG LISIUM

STUDENT TRANSFORMATION STAGES

- Student's Enrollment
- 8Hours JOIP Training
- Mock Interview
- Interview Schedule
- Resume Preparation
- JOB
- 3-6 Months Internship
- JOB
- Interview Schedule

OUR STUDENTS ARE PLACED IN

wipro | LARSEN & TOUBRO | Apollo HOSPITALS | Cognizant | Microsoft | servicenow

amazon | S&P Global | Aadvi Tech Solutions | Medtronic | meslova | {actlogica} Actuarial | Finance | Technology

Sunseaz Technologies Pvt Ltd | CAPRUS IT UNLOCKING SMART SOLUTIONS | SOLUGENIX | Collabera Value. Accelerated. | i Prism Technologies | NOOR

VALUEMOMENTUM | PERVACIO | M MIRABEL TECHNOLOGIES | erwin | STIC Soft When you need the results | sumtotal A Skillsoft Company

Nendrasys Accelerating Business Transformation | magnasoft | Innominds Powering the Digital Next | High Noon Consulting. Possible Is Everything | JRD SYSTEMS | ValueLabs

&techforce.ai | ENVISION | splashBI | archents | {⁀} techolution | SOLIX Empowering Data Management

QualityThought

📞 90595 14148

Quality Thought Infosystems India (P) Ltd.

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