

VERSION IT

Since 2001



Google Cloud Data Engineering+ AI Fundamentals



Best Training And Placements Institute



#No -46, SAP street, Behind Mytrivanam,
Ameerpet, Hyderabad-500038



+91 9848015399
+91 9391237284

www.versionit.co.in

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!

Training By Certified Instructors



Mock Interviews



Weekly Assignments



Project Training



Interview Cracking tips



Resume Preparation

GCP Cloud Basics

GCP Introduction

- **The need for cloud computing in modern businesses.**
- **Key features and offerings of Google Cloud Platform (GCP).**
- **Overview of core GCP services and products.**
- **Benefits and advantages of using cloud infrastructure.**
- **Step-by-step guide to creating a free-tier account on GCP.**

GCP Interfaces

Console

- **Navigating the GCP Console**
- **Configuring the GCP Console for Efficiency**
- **Using the GCP Console for Service Management**

Shell

- **Introduction to GCP Shell**
- **Command-line Interface (CLI) Basics**
- **GCP Shell Commands for Service Deployment and Management**

SDK

- **Overview of GCP Software Development Kits (SDKs)**
- **Installing and Configuring SDKs**
- **Writing and Executing GCP SDK Commands**

GCP Locations

Regions

- **Understanding GCP Regions**
- **Selecting Regions for Service Deployment**
- **Impact of Region on Service Performance**

Zones

- **Exploring GCP Zones**
- **Distributing Resources Across Zones**
- **High Availability and Disaster Recovery Considerations**

Importance

- **Significance of Choosing the Right Location**
- **Global vs. Regional Resources**
- **Factors Influencing Location Decisions**

GCP IAM & Admin

Identities

- **Introduction to Identity and Access Management (IAM)**
- **Users, Groups, and Service Accounts**
- **Best Practices for Identity Management**

Roles

- **GCP IAM Roles Overview**
- **Defining Custom Roles**
- **Role-Based Access Control (RBAC) Implementation**

Policy

- **Resource-based Policies**
- **Understanding and Implementing Organization Policies**
- **Auditing and Monitoring Policies**

Resource Hierarchy

- **GCP Resource Hierarchy Structure**
- **Managing Resources in a Hierarchy**
- **Organizational Structure Best Practices**

Linux Basics on Cloud Shell

- **Getting started with Linux**
- **Linux Installation**
- **Basic Linux Commands**
 - **Cloud shell tips**
 - **File and Directory Operations (ls, cd, pwd, mkdir, rmdir, cp, mv, touch, rm, nano)**
 - **File Content Manipulation (cat, less, head, tail, grep)**
 - **Text Processing (awk, sed, cut, sort, uniq)**
 - **User and Permission related (whoami, id, su, sudo, chmod, chown)**

Python for Data Engineer

Data Types

- **Strings**
- **Operators**
- **Numbers (Int, Float)**
- **Booleans**

Data Structures

- **Lists**
- **Tuples**
- **Dictionaries**
- **Sets**

Python Programming Constructs

- **if, elif, else statements**
- **for loops, while loops**
- **Exception Handling**
- **File I/O operations**

Modular Programming in Python

- **Functions & Lambda Functions**
- **Classes**

GCP Data Engineering Tools

Google Cloud Storage

- **Overview of Cloud Storage as a scalable and durable object storage service.**
- **Understanding buckets and objects in Cloud Storage.**
- **Use cases for Cloud Storage, such as data backup, multimedia storage, and website content**
- **Labs: using console & CLI to do below**
 - **Creating and managing Cloud Storage buckets.**
 - **Uploading and downloading objects to and from Cloud Storage.**
 - **Setting access controls and permissions for buckets and objects.**
 - **Data Transfer and Lifecycle Management**
 - **Versioning and Object Versioning**
- **Integration with Other GCP Services**
- **Monitoring and logging for Cloud Storage operations.**

Cloud SQL

- **Introduction to Cloud SQL**
- **Creating and Managing Cloud SQL Instances**
- **Configuring database settings, users, and access controls.**
- **Connecting to Cloud SQL instances using Cloud SQL studio, Shell, Workbenches**
- **Importing and exporting data in Cloud SQL.**
- **Backups and High Availability**

- **Integration with Other GCP Services**
- **Managing database user roles and permissions.**
- **Introduction to DMS**
- **End to End Database migration Project**
 - **Manual: Export and Import method**
 - **Automation: Cloud SQL DMS method**

BigQuery (SQL development)

- **Introduction to Big Query**
- **BigQuery Architecture**
- **Use cases for Big Query in business intelligence and analytics.**
- **Various method of creating table in Big Query**
- **BigQuery Data Sources and File Formats**
- **Native table and External Tables**
- **Working with complex data types**
 - **Working json data, nested, repeated and array data**
- **Data Integration and Export**
 - **Loading data into BigQuery from Cloud Storage, Cloud SQL, and other sources.**
 - **Exporting data from BigQuery to various formats.**
 - **Real-time data streaming into BigQuery.**
- **Configuring access controls and permissions in BigQuery.**

- **BigQuery Views:**
 - **Views**
 - **Materialized Views**
 - **Authorized Views**
- **Optimization techniques in BigQuery**
- **BigQuery Slots – on demand, flat-rate, flex-slots**
- **Case Study-1: implement a real-world analytics data platform for Spotify**
- **Case Study-2: Enterprise Social Media analytics platform**

DataProc (Pyspark Development)

- **Introduction to Hadoop and Apache Spark**
- **Understanding the difference between Spark and MapReduce**
- **What is Spark and Pyspark.**
- **Understanding Spark framework and its functionalities**
- **Overview of DataProc as a fully managed Apache Spark and Hadoop service.**
- **Cluster Creation and Configuration**
 - **Creating and managing DataProc clusters.**
 - **Configuring cluster properties for performance and scalability.**
 - **Preemptible instances and cost optimization.**
- **Learning Pyspark**
 - **How to read from multiple data sources – csv, text, json, parquet, database table, BigQuery tables**
 - **How to perform multiple transformations**
 - **How to write to multiple targets – csv, text, json, parquet, database table, BigQuery tables**

- **Running Jobs on DataProc**
 - **Submitting and monitoring Spark and Hadoop jobs on DataProc.**
 - **Use of initialization actions and custom scripts.**
 - **Job debugging and troubleshooting.**
- **Case study-1: Data Cleaning of Employee Travel Records**
- **Case study-2: Processing real-time patient health data**
- **Case study-3: Creating a pyspark job to support ML model creations**

DataFlow (Apache Beam development)

- **Introduction to DataFlow**
- **Use cases for DataFlow in real-time analytics and ETL.**
- **Understanding the difference between Apache Spark and Apache Beam**
- **How Dataflow is different from Dataproc**
- **Learning Apache Beam**
 - **How to read from multiple data sources – csv, text, json, parquet, database table, BigQuery tables**
 - **How to perform multiple transformations**
 - **How to write to multiple targets – csv, text, json, parquet, database table, BigQuery tables**
- **Case study-1: Template method of creating pipelines**
- **Case study-2: E-commerce Transaction Processing**
- **Case study-3: End to End Streaming Pipeline using Apache beam with Dataflow, Python app, PubSub, BigQuery, GCS**

Cloud Pub/Sub (Streaming)

- **Introduction to Pub/Sub**
- **Understanding the role of Pub/Sub in event-driven architectures.**
- **Key Pub/Sub concepts: topics, subscriptions, messages, and acknowledgments.**
- **Creating and Managing Topics and Subscriptions**
 - **Using the GCP Console to create Pub/Sub topics and subscriptions.**
 - **Configuring message retention policies and acknowledgment settings.**
- **Publishing and Consuming Messages**
 - **Writing and deploying code to publish messages to a topic.**
 - **Implementing subscribers to consume and process messages from subscriptions.**
- **Case study-1: Streaming use-case using Dataflow**

Cloud Composer (DAG Development)

- **Introduction to Composer/Airflow**
- **Overview of Airflow Architecture**
- **Use cases for Composer in managing and scheduling workflows.**
- **Creating and Managing Workflows**
 - **Creating and configuring Composer environments.**
 - **Defining and scheduling workflows using Apache Airflow.**
 - **Monitoring and managing workflow executions.**

- **Integration with Data Engineering Services**
 - **Orchestrating workflows involving BigQuery, DataFlow, and other services.**
 - **Coordinating ETL processes with Composer.**
 - **Integrating with external systems and APIs.**
- **Error Handling and Troubleshooting**
 - **Handling errors and retries in Composer workflows.**
 - **Debugging and troubleshooting failed workflow executions.**
 - **Logging and monitoring for Composer workflows.**
- **Level-1-DAG: Orchestrating the BigQuery pipelines**
- **Level-2-DAG: Orchestrating the DataProc pipelines**
- **Level-3-DAG: Orchestrating the Dataflow pipelines**
- **Deploy DAGs: Implementing CI/CD in Composer Using Cloud Build and GitHub**

Databricks on GCP

- **What is Lakehouse Architecture?**
- **Difference Between:**
 - **Data Lake**
 - **Data Warehouse**
 - **Data Lakehouse**
- **Introduction to Databricks**
 - **Overview of Databricks**
 - **Databricks Architecture**
 - **Setting up a Databricks Workspace**

- **Unity Catalog - Unified Data Governance**
 - **Introduction to Unity Catalog**
 - **Core Concepts: Metastore, Catalog, Schema, Tables, Volumes, Functions, Models**
 - **External Data Access**
 - **Storage Credentials**
 - **External Locations (S3, GCS, ADLS)**
 - **Lakehouse Federation - Foreign Catalog, Connection**
 - **Delta Sharing - Share, Recipient, Provider**
- **Databricks Clusters**
 - **Introduction to Clusters**
 - **Types of Clusters**
 - **Cluster Modes**
- **DBUtils Commands - File handling, Notebook Widgets, Secrets**
- **Introduction to Delta Lake**
 - **What is Delta Lake?**
 - **Creating Delta Lake Tables - Managed Tables, External Tables**
 - **Understanding Delta Lake Table Creation - Delta Log, Parquet Data Files**
 - **Advanced Delta Lake Features**
 - **Versioning**
 - **Time Travel**
 - **Compacting**
 - **Liquid clustering**
 - **Vacuum**

- **Databricks Jobs & Pipelines**
 - **Creating a Job**
 - **Scheduling Jobs**
 - **Setting Parameters**
 - **Managing Dependencies**
 - **Setting up Alerts**
- **Case study-1: Structured Streaming with Databricks**
- **Case study-2: Incremental Data Loading with Auto Loader**

Data Fusion (complementary)

- **Introduction to Data Fusion**
 - **Overview of Data Fusion as a fully managed data integration service.**
 - **Use cases for Data Fusion in ETL and data migration.**
- **Building Data Integration Pipelines**
 - **Creating ETL pipelines using the visual interface.**
 - **Configuring data sources, transformations, and sinks.**
 - **Using pre-built templates for common integration scenarios.**
- **Integration with GCP and External Services**
 - **Integrating Data Fusion with BigQuery, Cloud Storage, and other GCP services.**
- **Case study-1: End to End pipeline using Data fusion with Wrangler, GCS, BigQuery**

Cloud Functions(complementary)

- **Cloud Functions Introduction**
- **Setting up Cloud Functions in GCP**
- **Event-driven architecture and use cases**
- **Writing and deploying Cloud Functions**
- **Triggering Cloud Functions:**
 - **HTTP triggers**
 - **Pub/Sub triggers**
 - **Cloud Storage triggers**
- **Monitoring and logging Cloud Functions**
- **Usecase-1: Loading the files from GCS to BigQuery as soon as it is uploaded.**

Terraform(complementary)

- **Terraform Introduction**
- **Installing and configuring Terraform.**
- **Infrastructure Provisioning**
- **Terraform basic commands**
 - **Init, plan, apply, destroy**
- **Labs: Create Resources in Google Cloud Platform**
 - **GCS buckets**
 - **Dataproc cluster**
 - **BigQuery Datasets and tables**
 - **And more resources as needed**

AI Fundamentals for GCP Data Engineers (Current Market Demand)

- **Introduction to Artificial Intelligence & Generative AI**
 - **Understanding AI, ML, Deep Learning, Generative AI, and LLMs.**
 - **Real-world AI use cases in Data Engineering and Analytics.**
- **Machine Learning Fundamentals**
 - **Supervised, Unsupervised, and Reinforcement Learning.**
 - **Training, validation, testing, and model evaluation concepts.**
- **Prompt Engineering for Data Engineers**
 - **Designing effective prompts for Gemini, ChatGPT, and Vertex AI.**
 - **Using AI for SQL generation, code generation, documentation, and data analysis.**
- **Google Vertex AI Fundamentals**
 - **Overview of Vertex AI ecosystem.**
 - **Working with foundation models, Model Garden, and Gemini APIs.**
- **AI-Powered Data Pipelines**
 - **Integrating AI services with BigQuery, GCS, Dataflow, and Dataproc.**
 - **Building intelligent ETL/ELT pipelines using AI-driven transformations.**

Complementary End-to-End Projects:

- **Healthcare project on GCP of 8+ hours**
- **1Road Traffic project on Databricks of 5+ hours**

By the End of the course What Students can Expect

- **Proficient in SQL Development**
 - **Mastering SQL for querying and manipulating data within Google BigQuery and Cloud SQL.**
 - **Writing complex queries and optimizing performance for large-scale datasets.**
 - **Understanding schema design and best practices for efficient data storage.**
- **Pyspark Development Skills**
 - **Proficiency in using PySpark for large-scale data processing on Google Cloud.**
 - **Developing and optimizing Spark jobs for distributed data processing.**
 - **Understanding Spark's RDDs, Dataframes, and transformations for data manipulation.**
- **Apache Beam Development Mastery**
 - **Creating data processing pipelines using Apache Beam.**
 - **Understanding the concepts of parallel processing and data parallelism.**
 - **Implementing transformations and integrating with other GCP services.**
- **DAG Creations with Cloud Composer**
 - **Designing and implementing Directed Acyclic Graphs (DAGs) for orchestrating workflows.**
 - **Using Cloud Composer for workflow automation and managing dependencies.**
 - **Developing DAGs that integrate various GCP services for end-to-end data processing.**

- **Notebooks, Workflows with Databricks**
 - **Understand how to build and manage data pipelines using Databricks and Delta Lake.**
 - **Efficiently query and analyze large datasets with Databricks SQL and Apache Spark.**
 - **Implement scalable workflows and optimize performance within Databricks.**
- **Architecture Planning**
 - **Proficient in architecting end-to-end data solutions on GCP.**
 - **Understanding the principles of designing scalable, reliable, and cost-effective data architectures.**
- **Certification Readiness**
 - **Prepare for the Google Cloud Professional Data Engineer (PDE)**
 - **Associate Cloud Engineer (ACE) certifications through a combination of theoretical knowledge and hands-on experience.**
- **ML & AI ready**
 - **Students will understand how modern AI and Generative AI integrate with Google Cloud Data Engineering solutions and will be able to build AI-enabled data platforms using BigQuery, Vertex AI, and Gemini.**
- **Course Highlights:**
 - **Hands-on Labs & Real-world Use Cases**
 - **Live Q&A Sessions**
 - **Course Completion Certificate**
 - **Access to Study Materials & Code Repository**
 - **Industry-Standard Best Practices**

Thank You