DATA ENGINEERING COURSE SYLLABUS

GETIN TECHNOLOGIES - TIRUNELVELI

9, Prasanna Vinayagar Kovil Street, 1st Floor, Palayamkottai, Tirunelveli, Tamil Nadu 627002.

Landmark: Above Bharani Poly Clinic / Near Medai Police Station

Mobile NO: 8925831821

GETIN TECHNOLOGIES - KOVILPATTI

Door No: 971G/6, 1st Floor, Kalki Street, Krishna Nagar, Kovilpatti - 628502.

Landmark: After Six Veedu Line

Mobile NO: 8925831826

GETIN TECHNOLOGIES - VIRUDHUNAGAR

1st Floor, Venkatachalapathy Complex, Kooraikundu,

Landmark: Opposite Of Virudhunagar Collectorate / Above Burma Kadai Virudhunagar, Tamil Nadu 626002.

Mobile NO: 8925831828

WEBSITE: www.getintech.in

DATA ENGINEERING SYLLABUS

Introduction to Data Engineering:

- **Overview of data engineering and its role in data-driven organizations**

- **Introduction to key concepts and terminology**

- **Overview of common tools and technologies used in data engineering**

Data Ingestion:

- **Techniques for collecting data from various sources**

- **Introduction to streaming and batch processing**

- **Hands-on exercises with data ingestion tools (e.g., Apache Kafka, AWS Kinesis)**

Design Basics:

- **Overview of different types of data storage solutions (relational databases, NoSQL databases, data lakes, etc.)**

- **Hands-on exercises with relational and NoSQL databases (e.g., MySQL, MongoDB)**

Python Programming:

- **Getting started with Python programming for Data Processing**

- **Data Types**

- **Python Packages**

- **Loops and Conditional Statements**

- **Functions**

- **Collections**

- **String Handling**

- **File handling**

- **Exceptional Handling**

- **MySQL Integration**

- **INSERT, READ, DELETE, UPDATE, COMMIT, ROLLBACK operations**

- **MongoDB Integration Pre-processing, Cleaning, and Transforming Data**

Big Data Tools:

- **Linux OS**
- **Apache Hadoop**
- **HDFS**
- **Hadoop Cluster on GCP - Dataproc**

Apache Kafka:

- **Big Data and Apache Kafka**
- **Producers and Consumers**
- **Clusters Architectures**
- **Kafka Streams**
- **Kafka pipeline transformations**

Data Transformation:

- **Introduction to ETL (Extract, Transform, Load) processes.**
- **Techniques for data transformation and cleaning**
- **Hands-on exercises with ETL tools (e.g., Apache Airflow)**

Data Engineering with AWS:

- **Introduction to AWS platform, Creation of free account**
- **Walk through the platform and services offered by AWS**
- **IAM - Identity and Access Management**
- **Intro to AWS Data Warehouses, Data Marts, Data Lakes, and ETL/ELT pipelines**
- **Configuring the AWS Command Line Interface tool**
- **Creating an S3 bucket**
- **Working with Databases and various File formats (Data Lakes)**
- **Amazon Database Migration Service (DMS) for ingesting data**
- **Amazon Kinesis and Amazon MSK for streaming data**
- **AWS Lambda for transforming data**
- **AWS Glue for orchestrating big data pipelines**

- Consuming data - Amazon Redshift & Amazon Athena for SQL queries

Data Engineering with Azure:

- Introduction to Microsoft Azure platform, Creation of free account
- Walk through the platform and services offered by Azure
- IAM - Identity and Access Management
- Azure Data Lake - Managing Data
- Securing and Monitoring Data
- Introduction to Azure Data Factory (ADF)
- Building Data Ingestion Pipelines Using Azure Data Factory
- Azure Data Factory Integration Runtime
- Configuring Azure SQL Database
- Processing Data with Azure Databricks
- Introduction to Azure Synapse Analytics
- Data Transformations with Azure Synapse Dataflows
- Monitoring And Maintaining Azure Data Engineering Pipelines

Data Engineering with GCP:

- introduction to GCP platform, Creation of free account
- Walk through the platform and services offered by GCP
- IAM - Identity and Access Management
- Bigdata Solutions with GCP Components
- Data Warehouse – BigQuery
- Processing ETL/ELT pipelines with Data Fusion
- Connecting BI tool for visualizing Data with Looker Studio
- Architecting Data Pipelines
- CI/CD On Google Cloud Platform for Data Engineers

Data Quality and Testing

- Importance of data quality in data engineering
- Techniques for assessing and improving data quality

- **Hands-on exercises with data quality tools and frameworks**

Scalability and Performance

- **Strategies for scaling data pipelines**
- **Techniques for optimizing performance**
- **Hands-on exercises with performance tuning**

Reliability and Monitoring

- **Introduction to monitoring and logging in data engineering**
- **Strategies for ensuring reliability and fault tolerance**
- **Hands-on exercises with monitoring tools (e.g., Prometheus, Grafana)**

Data Security and Privacy

- **Overview of data security and privacy considerations**
- **Techniques for securing data pipelines and storage**
- **Hands-on exercises with data encryption and access control**

Deployment and Operations

- **Best practices for deploying and managing data engineering systems**
- **Introduction to containerization and orchestration tools (e.g., Docker, Kubernetes)**
- **Hands-on exercises with containerized data pipelines**