Demo Video: Click Here

Course Instructor:

A.K.M. Alfaz Uddin

Enterprise Data Engineering Lead Engineer,
Banglalink Digital Communications Ltd.
Former Lead Engineer, bKash Limited.
Former Senior Software Engineer, IMpulse (BD) Ltd
Former Specialist, BI/DW & CLM Systems, Robi Axiata Limited

Module 1: Introduction to Data Engineering: 1 hour

- What is Data? Importance of data.
- Introduction to Data Engineering
- Importance of Data-Driven Decision Making
- Component of Big Data
- Big Data Tools
- Data Engineering vs. Data Science vs. Data Analysis
- Skills required for Data Engineers
- Daily Role and Responsibility of a Data Engineer
- Challenges and Opportunities in Data Engineering
- Data Engineering Lifecycle
- Key Concepts: Big Data, Databases, Data Warehousing
- Question & Answer Session

Module 2: SQL(PostgreSQL) for Data Engineers: 12 hours

- Introduction to Databases:
  - Overview of databases.
  - Understanding relational database management systems (RDBMS).
  - OLAP vs OLTP.
- PostgreSQL setup & configuration
- SQL Basics
  - What is SQL?
  - Syntax and structure of SQL table creation.
  - Data types: numeric, string, date/time, etc.
  - Overview of DDL, DML and DCL
  - CRUD operations: SELECT, INSERT, UPDATE, DELETE.
- Querying Data
  - SELECT statement:
    - Retrieving data from table.
    - Filtering rows using WHERE clause.
    - Sorting results with ORDER BY.
    - Limiting results with LIMIT.
- Joins:
  - Inner joins, outer joins (left, right, full) joins.
- Aggregating Data
  - Aggregate functions: SUM, AVG, COUNT, MIN, MAX.
  - Grouping data with GROUP BY clause.
  - Filtering grouped data with HAVING clause.

- Subqueries
- Modifying Data
  - INSERT.
  - UPDATE.
  - DELETE.
  - MERGE.
- Working with Views
  - Creating and managing views.
  - Advantages of using views.
- Introduction to PL/pgSQL
  - Overview of PL/pgSQL as the procedural language for PostgreSQL.
  - Importance of stored procedures, functions, and triggers.
- PL/pgSQL Syntax Basics
  - Structure of PL/pgSQL blocks.
  - Declaration of variables and data types.
  - Comments in PL/pgSQL code.
- Flow Control Statements
  - Conditional statements
  - Looping
- Creating and Calling Functions
  - Syntax for creating user-defined functions in PL/pgSQL.
  - Defining function parameters and return types.
  - Calling functions from SQL queries or other PL/pgSQL code.
- Stored Procedures
  - Creating stored procedures in PL/pgSQL.
  - Difference between functions and stored procedures.
  - Advantages of using stored procedures for application logic.
- Normalization (1NF, 2NF, 3NF & BCNF)
- Indexes and Performance Optimization
  - Importance of indexes in database performance.
  - Creating and managing indexes.
  - Query optimization techniques.
- Question & Answer Session
- Assignment

Module 3: Python for Data Engineering: 08 Hours

- Python Basics
  - Introduction to Python and its relevance in Data Engineering.
  - Setting up Python development environment.
  - Basic syntax, variables, data types, and operators.
- Data Structures in Python
  - Lists, tuples, dictionaries, and sets
- Control Flow Structures and Functions
  - Conditional statements
  - Looping
  - Writing and calling functions.
  - Function parameters and return values.
- File Handling and Input/Output
  - Reading from and writing to files.
- Working with Libraries
  - Introduction to Python standard libraries.
  - Exploring Python libraries: NumPy, Pandas, Polars etc.
  - Installing and managing libraries using pip.
- Data Manipulation with Pandas
  - Introduction to Pandas library.
  - DataFrame basics.
  - Data loading and manipulating data using DataFrames.
  - Data cleaning, filtering, and transformation.
  - Handling missing data.
- NumPy for Numerical Computing
  - Basics of NumPy arrays.
  - Mathematical operations with NumPy.
- Working with SQL Databases in Python
  - Connecting to PostgreSQL using SQLAlchemy/psycopg2.
  - Executing SQL queries from Python.
- Question & Answer Session
- Assignment

Module 4: Data warehousing & ETL: 2 hours

- Introduction to Data Warehousing
  - Importance & understanding the concept of data warehousing.
  - Data warehousing architecture.
- Data Modeling for Data Warehousing
  - Dimensional modeling vs. relational modeling.
  - Star schema and snowflake schema.
  - Fact and dimension tables.
  - Slowly changing dimensions (SCDs).
- ETL Concepts and Processes
  - Understanding ETL and its role in data warehousing.
  - ETL vs. ELT
- Question & Answer Session

Module 5: Workflow Orchestration Tool: Apache Airflow: 06 hours

- Overview of popular ETL tools: Informatica, ODI, SSIS, Apache NiFi, Talend etc
- Batch Processing vs. Streaming Processing
- Setting up Apache Airflow environment.
- Components of workflow orchestration: tasks, dependencies, scheduling, etc.
- Directed Acyclic Graphs (DAGs) in Apache Airflow
- Introduction to operators in Apache Airflow.
- Types of operators.
- Defining tasks and dependencies between tasks.
- Automating ETL processes with scheduling and dependencies.
- Monitoring ETL pipelines.
- Question & Answer Session
- Assignment

Module 6: Big Data Technologies: 04 hours

- Introduction to Big Data
  - Understanding the concept of Big Data.
  - Characteristics of Big Data: volume, velocity, variety, veracity, and value.
  - Importance and applications of Big Data in various industries.
- Hadoop Ecosystem
  - Overview of Apache Hadoop and its components.

        - Understanding Hadoop Distributed File System (HDFS) for distributed storage.
        - Introduction to Hadoop MapReduce.
    - Apache Spark
        - Introduction to Apache Spark.
        - Hadoop vs. Apache Spark.
        - Basics of Spark programming using Python (PySpark).
    - Question & Answer Session
    - Assignment

Module 7: NoSQL Technologies: 04 hours

- Introduction to NoSQL Databases
    - Overview of NoSQL databases and their characteristics.
    - Comparison between NoSQL and relational databases.
- Introduction to MongoDB
    - Overview of MongoDB as a document-oriented NoSQL database.
    - Features and advantages of MongoDB.
    - Installation and setup of MongoDB
- MongoDB Data Model
    - Understanding the document-oriented data model.
    - Collections and documents.
- CRUD Operations in MongoDB
    - Basic CRUD operations (Create, Read, Update, Delete) using MongoDB.
- Querying and Aggregation
    - Query operators and expressions in MongoDB.
    - Aggregation pipeline.
- Question & Answer Session
- Assignment

Module 8: GCP & Google big query: 04 hours

- Introduction to GCP
    - Understanding of cloud computing
    - Types of cloud systems.
    - Overview of and GCP & understanding GCP services
- Introduction to Google BigQuery
    - What is BigQuery and its key features?
    - Exploring the BigQuery UI and running queries
- Data Loading & Manipulation

        - Loading data into BigQuery from various sources.
        - Writing basic SQL queries in BigQuery
        - Filtering, sorting, and aggregating data
    - Question & Answer Session
    - Assignment

Module 9: Capstone project

- Extract data from public API.
- Pre-processing, cleansing, and transformation of raw data.
- Loading in fact layer using Apache Airflow.
- Schedule workflow in Airflow.

Contact Details:

Mr. Sohan Khan

Course Coordinator at aiQuest Intelligence

Cell: +8801704265972 (Call/WhatsApp)

Watch Free Courses: https://www.aiquest.org/free-courses

Facebook Community: Join Our Community!

Visit Our Pages: Study Mart, aiQuest Intelligence