Data Engineering

Course Outline:

Module 1 - Data Engineering on AWS/ Azure / GCP

Class 1

- Introduction to Data Engineering
- Learning about S3
- Learning about EC2 (Windows and Linux)
- Learning about RDS (MySQL and MSSQL)
- Setting up Databases on AWS
- Learning about AWS Glue (Data Catalog, Tables and Crawlers)

Class 2

- Using Spark on AWS Glue
- Creating Jobs on AWS Glue
- Working with AWS Lambda (Setting up Python Lambda Function)

Module 2 - Dashboard Reporting Automation

Class 3

- Using Tableau for your Business Intelligence
- Connecting to MySQL/MSSQL Database your setup on AWS data using Tableau Desktop / Tableau Server

Class 4

- Using Live and Extract connections in Tableau
- Working with MSSQL, MySQL, PostgreSQL
- Scalability is key: What to do when your data grows

Module 3 - Apache Airflow for workflow management

Class 5

- Using Airflow as the ETL orchestration tool
- Installation of Airflow
- Using Airflow to Schedule ETL workflows

Module 4 - **SQL to NoSQL Database**

Class 6

- Building Backend infrastructures - NodeJS for APIs design, and database communication

- Setup Apache Cassandra on Docker
- Working with Cassandra Database
- Connection data exchanges using APIs using Next.js, Database and Frontend

Module 5 Working with Apache Kafka and Apache Spark

Class 7:

- Introduction to Kafka
- Kafka Architecture
- Kafka Topics
- Kafka and Zookeeper
- Design Producer and Consumer Messaging system

Class 8

- Introduction to Spark
- Batch Processing
- Stream Processing

Mock Interviews: showcasing your skills

Certificate of Completion

Facilitator

Data Engineering

Isaac Omolayo is a Data Engineer with an American IT firm and has 4 years + experience in building ETL processes, data infrastructures, automated data systems, working with cloud engines, and data analytics. He has vast knowledge in working with Apache Kafka, Apache Spark, Apache Airflow, Databases (SQL/NoSQL), Python and many other Big Data technologies. Generally, Isaac is a lover of open-source technologies.