30-Day Course Syllabus

Day 1:

Introduction to Cloud Computing

Basics of AWS Serverless architecture

Introduction to Data engineering using AWS Services

Day 2:

OLAP VS OLTP

Popular Datawarehouse

Day 3:

Basic EC2 & Linux VM Types Overview

Create IAM User and Role for Data Engineering

VPC & Security Group Configuration

Day 4:

Data Lake & other storage

Setting up Sample Data on Amazon S3

Day 5:

Understanding Spark Architecture Basics

Day 6:

RDD and DataFrame Fundamentals

Day 7:

Exploring DataFrame Spark API and Data Source Spark API

Day 8:

Reading and Processing CSV, JSON, Parquet Files with Spark

Day 9:

AWS Glue - Architecture Overview

Writing AWS Glue Job Scripts and Properties

Day 10:

Common Transformation Techniques in PySpark

Deep Dive into Transformations and Actions in PySpark

Directed Acyclic Graphs (DAG)

Day 11:

Partition Techniques in Spark

Day 12:

Handling Complex JSON and Struct Data Types

Day 13:

AWS Glue and AWS Secret Manager

Day 14:

Understanding and Creating UDFs (User-Defined Functions)

Day 15:

Using AWS Glue - Data Catalog and Databases

Day 16:

Overview of Amazon RDS MySQL Management & Connectivity

AWS Glue Connect to Amazon RDS MYSQL

Day 17:

Nosql Database

AWS Glue & DynamoDB

Day 18:

Developing Data Catalog with AWS Athena

Athena create table by crawler

Day 19:

AWS Athena - Data Types and DDL Statements

Day 20:

Hive catalog and Delta Lake

Day 21:

Table Manipulation with Delta Lake and Delta Table

Day 22:

Spark UI for Performance Monitoring

Day 23:

Developing AWS Glue Jobs with Advanced Network Configuration

Day 24:

Spark Optimization Techniques

Adaptive Query Execution

Day 25:

Addressing Data Skew and Salting in Spark

Day 26:

Building a Lambda function

Day 27:

AWS Lambda Function with Amazon S3

Day 28:

Lambda Function with Amazon DynamoDB

Lambda Function with Amazon SNS

Day 29:

SCD Implementation

Day 30:

Medalion Architecture