



AWS Data Engineering

AWS Data Engineer					
Components				Why Choose Us?	
Module 1		Module 2			
1	SQL, Spark SQL	7	AWS Lambda	✓	100% Practical, Real-time
2	Linux Concepts	8	AWS Glue	✓	Step by Step Examples
3	Python Concepts	9	AWS Athena	✓	Realtime Scenarios
4	Spark, PySpark	10	AWS Kinesis	✓	Resume Guidance
5	AWS Services	11	AWS Redshift	✓	Concept FAQs
6	AWS S3	12	End to End Project Work	✓	Project Work
Total Duration: 3 Months				Trainer : Praveen, 16+ Yrs Exp	

Chapter 1: Data Engineering Introduction

- What is Data Engineering
- Different Data Engineering Technologies
- Importance of Python in Data Engineering
- Importance of SQL in Data Engineering
- Importance of Linux in Data Engineering
- Cloud Computing Introduction

Chapter 2: SQL for Data Engineering

- SQL Introduction
- Provisioning MySQL using RDS
- Create, Alter, Drop Database
- Create, Alter, Insert tables
- Key Constraints
- Select Queries
- Joins

Chapter 3: Python for Data Engineering

- Python Introduction
- Python Programming Introduction
 - Interactive Mode Development &
 - Script Mode Development
- Python Installation on Windows Operating System & First Python Program using Python Shell
- Anaconda Python Installation and First Python Application using Jupyter Notebook & Spider IDE
- Python Editors & IDE Software's Installation & First Python Application using Editors & IDE's – Notepad++, PyCharm IDE, VSCode, etc..
- Python Program Indentation rules & Examples
- Language Fundamentals with Different Examples
 - Identifiers, Keywords, Datatypes & Many More...
- Flow Control Statements with Different Examples
 - (if, if-else, for, while, break, continue, etc..)
- Collection Datatypes with Different Examples
- Modules, Packages & Libraries with Different Example
- Procedure Oriented Concepts like Functions & Lambda Functions with Different Examples
- Object Oriented Concepts like Class & Object with Different Examples

Chapter 4: Linux for Data Engineering

- Linux Introduction
- Linux Filesystem Architecture
- Linux Installation on EC2 Instance
- Connecting Linux Machine
- Basic Linux Commands
- Linux File & Directory Permissions
- Linux Filter Commands

Chapter 5: AWS Basic Services

- AWS Introduction
- AWS Global Infrastructure (Regions & Availability Zones)
- AWS Account Creation
- AWS Free-tier
- AWS EC2
- AWS IAM

Chapter 6: PySpark

- PySpark Foundation
- PySpark Core Programming – RDD Programming, Transformations & Actions
- PySpark SQL – DataFrames, Tables, DSL & Native SQL
- PySpark Streaming Programming
- PySpark RDS Integration
- PySpark S3 Integration

Chapter 7: AWS S3

- AWS S3 Introduction
- AWS S3 Bucket using AWS Web Console
- Create AWS S3 Bucket
- Setup Data Set locally to upload into AWS s3
- Adding AWS S3 Buckets and Objects using AWS Web Console
- Version Control of AWS S3 Objects or Files
- AWS S3 Cross-Region Replication for fault tolerance
- Overview of AWS S3 Storage Classes or Storage Tiers
- Overview of Glacier in AWS s3
- Managing AWS S3 buckets and objects using AWS CLI
- AWS S3 Integration with PySpark

Chapter 8: AWS Lambda

- Hello World using AWS Lambda
- Setup Project for local development
- Deploy Project to AWS Lambda console
- Develop download functionality using requests
- Using 3rd party libraries in AWS Lambda
- Validating s3 access for local development
- Develop upload functionality to s3
- Validating using AWS Lambda Console
- Run using AWS Lambda Console
- Validating files incrementally
- Reading and Writing Bookmark using s3
- Maintaining Bookmark using s3
- Review the incremental upload logic
- Deploying lambda function
- Schedule Lambda Function using AWS Event Bridge

Chapter 9: AWS Glue

- AWS Glue Components
- Create Crawler and Catalog Table
- Create and Run the Glue Job

- Create and Run Glue Trigger
- Create Glue Workflow
- Run Glue Workflow and Validate
- Setup Spark History Server on AWS
- Build Glue Spark UI Container
- Update IAM Policy Permissions
- Start Glue Spark UI Container
- Steps for Creating Catalog Tables
- Create Glue Catalog Database
- Crawling Multiple Folders
- Crawling Multiple Folders
- Managing Glue Catalog using AWS CLI
- Managing Glue Catalog using AWS CLI

Chapter 10: AWS Athena

- Amazon Athena Introduction
- Glue Catalog Databases and Tables
- Access Glue Catalog Databases and Tables using Athena Query Editor
- Create Database and Table using Athena
- Populate Data into Table using Athena
- Using CTAS to create tables using Athena
- Amazon Athena Architecture
- Amazon Athena Resources and relationship with Hive
- Create Partitioned Table using Athena
- Develop Query for Partitioned Column
- Insert into Partitioned Tables using Athena
- Validate Data Partitioning using Athena
- Drop Athena Tables and Delete Data Files
- Drop Partitioned Table using Athena
- Data Partitioning in Athena using CTAS

Chapter 11: AWS Kinesis

- Building Streaming Pipeline using Kinesis
- Rotating Logs
- Setup Kinesis Firehose Agent
- Create Kinesis Firehose Delivery Stream
- Planning the Pipeline
- Create IAM Group and User
- Granting Permissions to IAM User using Policy
- Configure Kinesis Firehose Agent
- Start and Validate Agent
- Building Simple Streaming Pipeline by Integrating PySpark

Chapter 12: AWS Redshift - 1

- Amazon Redshift - Introduction
- Create Redshift Cluster using Free Trial
- Connecting to Database using Redshift Query Editor
- Get list of tables querying information schema
- Run Queries against Redshift Tables using Query Editor
- Create Redshift Table
- CRUD Operations
- Insert Data into Redshift Tables
- Update Data in Redshift Tables
- Delete data from Redshift tables
- Redshift Saved Queries using Query Editor

Chapter 13: AWS Redshift - 2

- Deleting Redshift Cluster
- Copy Data from s3 to Redshift - Introduction
- Setup Data in s3 for Redshift Copy
- Copy Database and Table for Redshift Copy Command
- Run Copy Command to copy data from s3 to Redshift Table
- Copy JSON Data from s3 to Redshift table using IAM Role
- Redshift Architecture
- Create multi-node Redshift Cluster
- Connect to Redshift Cluster using Query Editor
- Create Redshift Database
- Create Redshift Database User
- Create Redshift Database Schema
- Integrating PySpark with Redshift

Chapter 14: End to End Realtime Project

- End to End Project Work
- Resume Guidance
- Interview FAQs
- Mock Interview (On-Request)

Reach Us Now, for Free demo !

contact@sqlschool.com

Reach Us : +91 995144 0801, +91 966644 0801

Next Schedules: www.sqlschool.com/Register

Thank you for contacting our SQL School.

We shall look forward for your demo, course enrolment. Trust us, we shall ensure the best learning experience for your profile !