

AWS DATA ENGINEER

With DATA ANALYTICS



Fundamentals of Cloud Computing & AWS

- ▶ Cybersecurity Introduction
- ▶ Concepts and Terminology
- ▶ Security Controls
- ▶ Security Principles
- ▶ Technical Terminology
- ▶ Identity and Access Management
- ▶ Risk Assessment

Data Engineering Concepts & Best Practices

- ▶ Core Principles of Data Engineering
- ▶ ETL vs. ELT - Key Differences
- ▶ Batch vs. Streaming Data Pipelines
- ▶ Data Warehousing vs. Data Lakes
- ▶ Cloud-Native Data Engineering Overview

Database Management & Data Modeling

- ▶ Introduction to SQL & NoSQL Databases
- ▶ Data Modeling for Structured & Unstructured Data
- ▶ Performance Optimization for SQL Queries
- ▶ Working with AWS Database Services (RDS, Aurora, DynamoDB)
- ▶ Data Indexing, Partitioning & Query Optimization

Working with Amazon S3 & Cloud Storage

- ▶ Object Storage & Lifecycle Policies
- ▶ Data Lake Architecture in AWS
- ▶ Data Security, Access Control & Encryption
- ▶ Integrating S3 with AWS Analytics & Processing Services

Building Data Pipelines with AWS Glue

- ▶ AWS Glue for ETL Automation
- ▶ Data Cataloging & Schema Evolution
- ▶ Data Transformation & Integration with AWS Services
- ▶ Orchestrating Workflows with AWS Glue Workflows & Step Functions

Real-Time Data Ingestion with Amazon Kinesis

- ▶ Streaming Data Overview & Use Cases
- ▶ Data Collection with Kinesis Data Streams & Firehose
- ▶ Managing Latency & Throughput in Streaming Pipelines
- ▶ Integration with AWS Lambda & Event-Driven Processing

Distributed Computing with AWS EMR

- ▶ Introduction to Hadoop, Spark & Presto on AWS
- ▶ Setting Up & Managing AWS EMR Clusters
- ▶ Running Big Data Workloads Using Spark
- ▶ Performance Tuning & Cost Optimization for EMR

Data Warehousing & Analytics with Redshift & Athena

- ▶ Introduction to Amazon Redshift
- ▶ Querying Data in Redshift & Athena
- ▶ Optimizing Performance with Distribution & Sorting Keys
- ▶ Cost-Efficient Query Execution Strategies

Data Processing & Transformation on AWS

- ▶ Apache Spark & Presto for Data Processing
- ▶ Data Transformation Techniques in AWS
- ▶ Serverless Query Processing with Athena
- ▶ Workflow Automation Using AWS Step Functions

Event-Driven Data Pipelines with Kafka & MSK

- ▶ Introduction to Managed Streaming for Apache Kafka (MSK)
- ▶ Event-Driven Architectures for Data Processing
- ▶ Integrating Kafka with AWS Analytics Services
- ▶ Monitoring & Managing High-Throughput Streaming

Operationalizing Machine Learning with SageMaker

- ▶ Introduction to AWS SageMaker for AI/ML
- ▶ Model Training, Deployment & MLOps Best Practices
- ▶ Batch & Real-Time Inference Using AWS AI Services
- ▶ Automating ML Pipelines Using AWS Step Functions

CI/CD & DevOps for Data Engineering

- ▶ Continuous Integration & Deployment for Data Pipelines
- ▶ Automating AWS Glue, Redshift & EMR Workloads
- ▶ Workflow Orchestration with Apache Airflow
- ▶ Infrastructure as Code (Terraform & AWS CloudFormation)

Data Analytics & Visualization with Amazon QuickSight

- ▶ Introduction to Business Intelligence & Data Visualization
- ▶ Connecting QuickSight to AWS Data Sources
- ▶ Building Interactive Dashboards
- ▶ Security & Performance Optimization for Data Reporting

Python for Cloud & Data Engineering

- ▶ Working with AWS SDK (Boto3) for Automation
- ▶ Scripting AWS Data Pipelines Using Python
- ▶ Managing Databases & Compute Resources Programmatically
- ▶ Logging, Monitoring & Debugging AWS Workloads

Security & Compliance for Cloud Data Pipelines

- ▶ Identity & Access Management (IAM) Best Practices
- ▶ Securing AWS Storage & Databases
- ▶ Data Encryption & Key Management (KMS)
- ▶ Threat Detection & Security Monitoring with AWS Tools

Monitoring & Observability in AWS

- ▶ Setting Up Monitoring with AWS CloudWatch
- ▶ Logging & Debugging AWS Glue & EMR Jobs
- ▶ Performance Tuning for Streaming Data Pipelines
- ▶ Automating Alerts & Incident Response

AWS Cloud Data Engineering Capstone Project

Project Scope:

- ▶ Design & Deploy a Scalable Data Pipeline using AWS services
- ▶ Develop Data Processing & Transformation Workflows
- ▶ Implement Real-Time & Batch Processing Pipelines
- ▶ Set Up Monitoring, Logging & Security Best Practices
- ▶ Optimize Performance & Cost Efficiency for Production Workloads



SOLUGENIX