KOENIG
step forward

aws partner network
Training Partner

Data Processing and Orchestration on AWS

Course Duration: 16 Hours (2 Days)

Overview

Unlock the full potential of AWS with our Data Processing and Orchestration on AWS course. Over just 2 days (16 hours), gain a deep understanding of Data pipelines, orchestration, and key AWS services for efficient data processing. From Data ingestion and Storage to processing and Visualization, you'll follow a step-by-step approach that emphasizes best practices for security, Cost optimization, and disaster recovery. Practical labs, like Incremental data load from S3 to Redshift and Creating a data lake with Lake Formation, ensure hands-on learning. This course suits those looking to master data workflows using AWS Glue, Step Functions, and CloudWatch, among others.

Audience Profile

Koenig Solutions' Data Processing and Orchestration on AWS course equips IT professionals with the skills to efficiently manage data pipelines, orchestration, and AWS services for end-to-end data processing.

- Data Engineers
- Data Scientists
- Cloud Architects
- Solution Architects
- IT Managers
- Database Administrators
- DevOps Engineers
- Big Data Analysts
- System Integrators
- IT Security Specialists
- Data Analysts
- Software Developers specializing in cloud computing
- Business Intelligence (BI) Professionals
- AWS Cloud Practitioners
- IT Consultants focusing on cloud solutions

Course Syllabus

Module 1: Introduction

- 1.1 Data Pipelines and Orchestration in AWS
- 1.2 Choosing the Right Service: An Overview of AWS Data Processing Services

- 1.2.1 Data Warehousing: Redshift vs. Athena
- 1.2.2 NoSQL Databases: DynamoDB
- 1.2.3 Streaming Data Ingestion: Kinesis Firehose
- 1.2.4 Data Lakes: Building and Managing with Lake Formation
- 1.3 Introduction to Multitenancy in Data Management
- Key AWS Services for Multitenant Data Management
- Security Considerations in a Multitenant Environment
- 1.4 Benefits of Using Best Practices

Module 2: Data Ingestion

- 2.1 Batch Data Ingestion with Amazon S3
- 2.2 Streaming Data Ingestion with Amazon Kinesis Data Firehose
- 2.2.1 Delivery Streams and Transformations
- 2.2.2 Buffering and Error Handling
- 2.3 Real-Time Data Ingestion with AWS Greengrass and IoT Core

Module 3: Data Storage and Management

- 3.1 Data Warehousing for Analytics: Redshift vs. Athena
- 3.1.1 Redshift: Columnar Storage for Scalable Analytics
- 3.1.2 Athena: Serverless Analytics on S3
- 3.2 NoSQL Data Storage: DynamoDB for Scalable Applications
- 3.3 Data Lake Management with AWS Lake Formation
- 3.3.1 Data Lifecycle Management
- 3.3.2 Access Control and Security

Module 4: Data Processing and Transformation

- 4.1 Serverless Data Processing with AWS Glue
- 4.1.1 Data Catalog and Schema Management
- 4.1.2 ETL Jobs with AWS Glue (Extract, Transform, Load)
- 4.2 Data Filtering and Transformation with AWS Lambda
- 4.2.1 Triggering Lambda Functions with Amazon CloudWatch Events
- 4.2.2 Custom Logic for Data Processing

Module 5: Data Visualization and Analytics

- 5.1 Cloud-Native Visualization with Grafana
- 5.2 Introduction to Databricks
- Ingesting Data and Building a Lakehouse for Analyzing Customer Product Usage
- Querying a Lakehouse Using SQL Queries

Module 6: Orchestration and Monitoring

- 6.1 AWS Step Functions for Orchestrating Workflows
- 6.1.1 Chaining Data Processing Steps with Step Functions
- 6.1.2 Error Handling and Retries
- 6.2 Monitoring Pipeline Health with Amazon CloudWatch
- 6.2.1 Logging and Metrics for Data Processing Jobs
- 6.2.2 Setting Up Alerts and Notifications

Module 7: Security and Best Practices

- 7.1 IAM Roles and Permissions for Secure Data Access
- 7.2 Encryption at Rest and in Transit
- 7.3 Cost Optimization Strategies for Data Pipelines
- 7.4 Disaster Recovery and Backup Strategies

Labs

1. Incremental Data Load from Amazon S3 to Amazon Redshift Using AWS Glue
2. Data Ingestion from Amazon S3 to Amazon Redshift
3. Creating a Data Lake with AWS Lake Formation, Data Crawling Using AWS Glue, and Querying with Amazon Athena
4. Accessing Data from Amazon S3 to Amazon Redshift Using Redshift Spectrum
5. Using AWS Lambda UDF with Amazon Redshift
6. Sending VPC Flow Log Data to Splunk Using Amazon Kinesis Data Firehose