Course Duration: 40 hours (5 Days)

Data to Decisions: ML and Analytics on AWS

This intensive 5-day course provides participants with a deep dive into the tools, services, and workflows available in AWS for Machine Learning and Data Analytics. Participants will explore data ingestion, transformation, storage, and analysis using services like Amazon S3, AWS Glue, Amazon Athena, and Amazon Redshift. On the ML side, learners will be introduced to Amazon SageMaker for building, training, and deploying machine learning models. The course combines theoretical instruction with real-world use cases and hands-on labs to ensure a well-rounded learning experience for professionals aiming to develop or scale analytics and ML workloads on AWS.

Course objectives

By the end of this course, learners will be able to:

- Understand the core AWS services for data engineering and machine learning.
- Ingest, store, and query data using AWS analytics services.
- Use AWS Glue and Amazon Athena for data transformation and ad hoc querying.
- Design and optimize data lakes and data pipelines.
- Train, evaluate, and deploy machine learning models using Amazon SageMaker.
- Apply ML models to solve classification, regression, and forecasting problems.
- Integrate analytics and ML solutions into real-world scenarios.

Prerequisites

- Architecting on AWS, Machine Learning fundamentals, Data Analytics fundamentals
- Familiarity with Python or SQL

Target Audience

- Data Analysts and Data Engineers
- Machine Learning Engineers and Data Scientists
- Cloud Engineers and Solution Architects
- IT professionals transitioning into cloud-based ML/analytics roles
- Anyone interested in using AWS for data-driven applications

KOENIG
step forward

Course outline

Day-1 & 2
AWS Sagemaker MLOPS using Python

1.1 Introduction to MLOps
- 1.1.1 Processes
- 1.1.2 People
- 1.1.3 Technology
- 1.1.4 Security and governance
- 1.1.5 MLOps maturity mode

1.2 Repeatable MLOps: Repositories
- 1.2.1 Managing data for MLOps
- 1.2.2 Version control of ML models
- 1.2.3 Code repositories in ML

1.3 Introduction to Machine Learning and the ML Pipeline
- 1.3.1 End-to-end orchestration with AWS Step Functions
- 1.3.2 End-to-end orchestration with SageMaker Projects
- 1.3.3 Governance and security

1.4 Introduction to Amazon SageMaker
1.5 Problem Formulation
1.6 Preprocessing
1.7 Model Training and Evaluation
1.8 Deployment Pipeline
1.9 MLOps: Monitoring
- 1.9.1 The importance of monitoring in ML
- 1.9.2 Operations considerations for model monitoring
- 1.9.3 Remediating problems identified by monitoring ML solutions

Day-3
AWS Datalake formation

2.1 Introduction to data lakes
2.2 Data ingestion, cataloging, and preparation
2.3 Data processing and analytics
2.4 Building a data lake with AWS Lake Formation
2.5 Terminologies
2.6 Lake Formation Integrations
2.7 Configure Data Lake
2.8 Using crawlers in AWS Glue
2.9 Query Data

Day-4

AWS Glue and Databrew

3.1 Introduction to AWS Glue
3.2 Understanding Data Transformation with AWS Glue
3.3 Working with Data Sources and Targets
3.4 Managing and Monitoring AWS Glue Jobs
3.5 Working with AWS Glue DataBrew
3.6 Data Profiling
3.7 Data Quality
3.8 Transformation
3.9 Feature Engineering

Day-5

Amazon Redshift

4.1 Warehouse creation and connecting with Query Editor
4.2 Table creation, one time data load
4.3 Continuous load with stored procedures
4.4 Data sharing from data producer (serverless) to consumer (provisioned cluster)
4.5 Create Machine Learning model with Redshift ML features
4.6 User Redshift Spectrum features to query data in S3 data lake
4.7 Redshift monitoring and audit logging