
CTR-703 AWS Big Data

Overview

Course Duration: 5 Days

About This Course

The AWS Big Data course focuses on the tools and services AWS offers for managing and analyzing large datasets. It covers data ingestion, storage, processing, and visualization, using services like Amazon S3, Redshift, EMR, and QuickSight.

Audience Profile

This course is intended for data engineers, data analysts, and IT professionals who need to process and analyze large datasets using AWS.

At Course Completion

- Implement data ingestion and processing pipelines using AWS services.
- Store and manage large datasets in scalable AWS storage solutions.
- Analyze and visualize big data with AWS analytics and BI tools.

Course Outline

Module 1: Introduction to Big Data and AWS

- Overview of Big Data Concepts
- The Importance of Big Data in Modern Enterprises
- Introduction to AWS Big Data Services
- Setting up the AWS Environment for Big Data
- Introduction to AWS Data Lakes and Analytics

Module 2: Data Ingestion and Collection

- Data Sources and Types (Structured, Unstructured, Semi-Structured)
- AWS Data Ingestion Tools: Kinesis, Data Pipeline, Glue

- Real-Time Data Streaming with AWS Kinesis
- Batch Data Ingestion Techniques
- Best Practices for Data Ingestion in AWS

Module 3: Data Storage and Management

- Storing Data with Amazon S3 and Glacier
- Data Warehousing with Amazon Redshift
- NoSQL Databases: DynamoDB
- Managing Data in Relational Databases: RDS and Aurora
- Data Archiving and Lifecycle Policies

Module 4: Data Processing with Apache Spark on AWS

- Introduction to Apache Spark
- Running Apache Spark on AWS EMR (Elastic MapReduce)
- Spark Core Concepts: RDDs, DataFrames, and Datasets
- Processing Large Datasets with Spark
- Optimizing Spark Jobs on AWS EMR
- Hands-On Lab: Building a Spark Application on AWS

Module 5: Data Processing and Analytics

- Data Transformation and ETL with AWS Glue
- Querying Data with Amazon Athena
- Real-Time Analytics with Amazon Kinesis Analytics
- Combining Spark with Other AWS Big Data Services
- Big Data Processing Architectures on AWS

Module 6: Data Security and Compliance

- Securing Data in AWS: Encryption, IAM, and Policies
- Compliance Standards and AWS Compliance Programs
- Data Privacy and Governance
- Managing Access to Data with AWS IAM
- Monitoring and Auditing Data Access

Module 7: Big Data Visualization and Reporting

- Data Visualization Tools in AWS: QuickSight
- Integrating Big Data with BI Tools
- Building Dashboards and Reports
- Real-Time Reporting with AWS Big Data Services
- Best Practices in Data Visualization

Module 8: Advanced Big Data Techniques

- Machine Learning with Big Data on AWS
- Predictive Analytics with AWS Services
- Data Lake Architecture and Implementation
- Serverless Big Data Processing with AWS Lambda
- Handling Streaming Data and Complex Workloads

Module 9: Cost Management and Optimization

- Cost Management Tools and Services in AWS
- Optimizing Big Data Workloads for Cost Efficiency
- Best Practices for Budgeting and Forecasting
- Implementing Cost-Effective Data Processing Pipelines
- Using AWS Cost Explorer and Trusted Advisor

Module 10: Real-World Applications and Case Studies

- Industry-Specific Use Cases for AWS Big Data
- Success Stories and Lessons Learned
- Group Project: Implementing a Big Data Solution on AWS
- Challenges and Solutions in Big Data Projects
- Future Trends in Big Data and AWS

work environment: AWS S3, AWS EMR, AWS Glue, AWS Kinesis, AWS Redshift, AWS QuickSight, Apache Spark

Prerequisites

- Basic understanding of data processing concepts.
- Familiarity with databases and SQL.
- Experience with AWS services is recommended but not required.