

Building Data Lakes on AWS

Course Duration: 8 Hours (1 Day)

Overview

The "Building Data lakes on AWS" course provides a comprehensive guide to creating, managing, and utilizing Data lakes on the AWS cloud platform. It is designed to help learners understand the value of Data lakes, differentiate them from data warehouses, and recognize the crucial components that make up a data lake. The course covers essential topics such as Data ingestion, Cataloguing, preparation, and processing using a variety of AWS services, including AWS Glue, Amazon Athena, and AWS Lake Formation. Learners will gain practical experience through hands-on labs, setting up a simple data lake, building a data lake with AWS Lake Formation, automating data lake creation, and Data Visualization using Amazon Quick Sight. By the end of the course, participants will have a solid understanding of building Data lakes on AWS and will be equipped with the skills to build a data lake on AWS effectively, ensuring they can leverage the full potential of their data assets in the cloud.

Audience Profile

This AWS data lake course offers in-depth training on setting up and managing data lakes, ideal for IT professionals focused on data management and analytics.

- Data Engineers
- Data Scientists
- Data Analysts
- Cloud Architects
- IT Managers
- Database Administrators
- Big Data Specialists
- Business Intelligence Professionals
- System Administrators
- Developers interested in data lake architectures

Course Syllabus

Module 1: Introduction to Data Lakes

- Describe the value of data lakes
- Compare data lakes and data warehouses
- Describe the components of a data lake

- Recognize common architectures built on data lakes

Module 2: Data Ingestion, Cataloguing, and Preparation

- Describe the relationship between data lake storage and data ingestion
- Describe AWS Glue crawlers and how they are used to create a data catalogue
- Identify data formatting, partitioning, and compression for efficient storage and query
- Lab 1: Set up a simple data lake

Module 3: Data Processing and Analytics

- Recognize how data processing applies to a data lake
- Use AWS Glue to process data within a data lake
- Describe how to use Amazon Athena to analyse data in a data lake

Module 4: Building a Data Lake with AWS Lake Formation

- Describe the features and benefits of AWS Lake Formation
- Use AWS Lake Formation to create a data lake
- Understand the AWS Lake Formation security model
- Lab 2: Build a data lake using AWS Lake Formation

Module 5: Additional Lake Formation Configurations

- Automate AWS Lake Formation using blueprints and workflows
- Apply security and access controls to AWS Lake Formation
- Match records with AWS Lake Formation Find Matches
- Visualize data with Amazon Quick Sight
- Lab 3: Automate data lake creation using AWS Lake Formation blueprints
- Lab 4: Data visualization using Amazon Quick Sight

Module 6: Architecture and Course Review

- Post course knowledge check
- Architecture review
- Course review