

Building Data Lakes on AWS

Course Summary

Description

In this course, you will learn how to build an operational data lake that supports analysis of both structured and unstructured data. You will learn the components and functionality of the services involved in creating a data lake. You will use AWS Lake Formation to build a data lake, AWS Glue to build a data catalog, and Amazon Athena to analyze data. The course lectures and labs further your learning with the exploration of several common data lake architectures.

Objectives

At the end of this course, students will be able to:

- Apply data lake methodologies in planning and designing a data lake.
- Articulate the components and services required for building an AWS data lake.
- Secure a data lake with appropriate permission.
- Ingest, store, and transform data in a data lake.
- Query, analyze, and visualize data within a data lake.

Topics

- Introduction to data lakes
- Data ingestion, cataloging, and preparation
- Data processing and analytics
- Building a data lake with AWS Lake Formation
- Additional Lake Formation configurations
- Architecture and course review

Audience

Those who can benefit from this Building Data Lakes on AWS course include:

- Data platform engineers
- Solutions Architects
- IT professionals

Prerequisites

It is recommended that attendees of this course have:

- Working knowledge of core AWS services and public cloud implementation.
- One year of experience building data analytics pipelines or have completed Data Analytics Fundamentals.
- Attended one of Architecting on AWS, Developing on AWS, or Systems Operations on AWS.

Duration

One day

Building Data Lakes on AWS

Course Outline

- I. Introduction to data lakes*
 - A. Describe the value of data lakes
 - B. Compare data lakes and data warehouses
 - C. Describe the components of a data lake
 - D. Recognize common architectures built on data lakes

- II. Data ingestion, cataloging, and preparation*
 - A. Describe the relationship between data lake storage and data ingestion
 - B. Describe AWS Glue crawlers and how they are used to create a data catalog
 - C. Identify data formatting, partitioning, and compression for efficient storage and query
 - D. Lab 1: Set up a simple data lake

- III. Data processing and analytics*
 - A. Recognize how data processing applies to a data lake
 - B. Use AWS Glue to process data within a data lake
 - C. Describe how to use Amazon Athena to analyze data in a data lake

- IV. Building a data lake with AWS Lake Formation*
 - A. Describe the features and benefits of AWS Lake Formation
 - B. Use AWS Lake Formation to create a data lake
 - C. Understand the AWS Lake Formation security model
 - D. Lab 2: Build a data lake using AWS Lake Formation

- V. Additional Lake Formation configurations*
 - A. Automate AWS Lake Formation using blueprints and workflows
 - B. Apply security and access controls to AWS Lake Formation
 - C. Match records with AWS Lake Formation FindMatches
 - D. Visualize data with Amazon QuickSight
 - E. Lab 3: Automate data lake creation using AWS Lake Formation blueprints
 - F. Lab 4: Data visualization using Amazon QuickSight

- VI. Architecture and course review*
 - A. Post course knowledge check
 - B. Architecture review
 - C. Course Review