

Building Data Lakes on AWS

AWS Classroom Training

Course 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.

- Course level: Intermediate
- Duration: 1 day

Activities

This course includes presentations, lecture, hands-on labs, and group exercises.

Course objectives

In this course, you will learn to:

- Apply data lake methodologies in planning and designing a data lake
- CO1. Plan and design a data lake using established data lake methodologies.
- CO2. Describe the components and services required for building a data lake on AWS.
- CO3. Explain how to secure a data lake on AWS using appropriate permissions.
- CO4. Compare the ways data can be ingested, stored, and transformed in a data lake on AWS.
- CO5. Analyze and visualize data stored in a data lake on AWS.
- CO6. Build and automate deployment of a data lake on AWS.
- CO7. Describe the role of a data lake within a modern data architecture.

Intended audience

This course is intended for:

- Data platform engineers
- Solutions architects
- IT professionals

Prerequisites

We recommend that attendees of this course have:

Building Data Lakes on AWS

AWS Classroom Training

- Completed the [AWS Technical Essentials](#) classroom course
- One year of experience building data analytics pipelines or have completed the [Data Analytics Fundamentals](#) digital course

Course outline

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, cataloging, 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 catalog
- Identify data formatting, partitioning, and compression for efficient storage and query

Module 3: Building a Data Lake with AWS Lake Formation

- 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 analyze data in a data lake
- Lab 01: Building a Data Lake with AWS Lake Formation

Module 4: Data Processing and Analysis

- 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

Building Data Lakes on AWS

AWS Classroom Training

Module 5: Additional Lake Formation configurations

- Explain the available built-in Blueprints to create and populate a new Lake Formation
- Describe methods for applying advanced permissions to secure data access and workflow.
- Describe fine-grained row/cell access control
- Explain the Lake Formation Tag-based access control mechanism and the different use cases for Named access control vs. Tag-based access control
- Describe access flow that enforces fine-grained access policies to both catalog metadata and underlying data resource for analytics services connecting to Lake Formation

Module 6: Modern Data Architecture

- Explain capabilities of a modern data architecture: Scalable data lakes, Purpose-built analytics services, Seamless data movement, unified governance, and performance and cost-effectiveness
- Articulate the typical data movement within a modern data architecture: Inside out, Outside in, Around the perimeter, and Sharing across.
- Describe focus of building and maintaining data products as a service.
- Describe a typical Data Mesh architecture using Lake Formation and the key enablers supporting this methodology
- Lab 3: Building and publishing a data product in Lake Formation

Module 7: Course Wrap Up

- Post course knowledge check
- Architecture review
- Course review