

Building a Serverless Data Lake

Course#: AWS-BSDL

Duration: 1 Day

Price: 0.00

Course Description

In this one-day, advanced course, you will learn to design, build, and operate a serverless data lake solution with AWS services. This course will include topics such as ingesting data from any data source at large scale, storing the data securely and durably, enabling the capability to use the right tool to process large volumes of data, and understanding the options available for analyzing the data in near-real time.

Objectives

Collect large amounts of data using services such as Kinesis Streams and Firehose and store the data durably and securely in Amazon Simple Storage Service.

Create a metadata index of your data lake.

Choose the best tools for ingesting, storing, processing, and analyzing your data in the lake.

Apply the knowledge to hands-on labs that provide practical experience with building an end-to-end solution.

Configure Amazon Simple Notification Service (Amazon SNS) to audit, monitor, and receive event notifications about activities in the data warehouse

Prepare for operational tasks, such as resizing Amazon Redshift clusters and using snapshots to back up and restore clusters

Use a business intelligence (BI) application to perform data analysis and visualization tasks against your data

Audience

Solutions architects

Big Data developers

Data architects and analysts

Data analysis practitioners

Prerequisites

Good working knowledge of AWS core services, including Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3)

Some experience working with a programming or scripting language

Familiarity with the Linux operating system and command line interface

Requires a laptop to complete lab exercises tablets are not appropriate

Content

This course covers the following concepts:

Key services that help enable a serverless data lake architecture

A data analytics solution that follows the ingest, store, process, and analyze workflow

Repeatable template deployment for implementing a data lake solution

Building a metadata index and enabling search capability

Setup of a large-scale data-ingestion pipeline from multiple data sources

Transformation of data with simple functions that are event triggered

Data processing by choosing the best tools and services for the use case

Options available to better analyze the processed data

Best practices for deployment and operations