

**Building Data Lakes on AWS**

**Course Number:** AWS-118  
**Duration:** 1 day

**Overview**

This Building Data Lakes on AWS training is a one-day, advanced-level course that teaches attendees how to design, build, and operate a serverless data lake solution with AWS analytics services. Attendees learn the data lake architecture and how to ingest data from any data source at large scale, store the data securely and durably, enable data scientists to use the right tool to process data, and manage data. Students come away with a thorough understanding of the options available for analyzing data in near real-time.

Accelebrate is an AWS Training Partner (ATP) and this hands-on official AWS Classroom Training course is taught by an accredited Amazon Authorized Instructor (AAI).

**Prerequisites**

All students must have:

* Working knowledge of core AWS services and public cloud implementation
* One year of experience building data analytics pipelines or have completed the AWS [Data Analytics Fundamentals](https://www.aws.training/Details/eLearning?id=35364) digital course
* Attended one of [Architecting on AWS](file:////training/aws-architecture), [Developing on AWS](file:////training/aws-developing), or [Systems Operations on AWS](file:////training/systems-operations-aws)

**Materials**

All AWS Data Lake training students receive comprehensive courseware.

**Software Needed on Each Student PC**

A modern web browser and an Internet connection free of restrictive firewalls, so that the student can connect by SSH or Remote Desktop (RDP) into AWS virtual machines.

**Objectives**

* 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 permissions
* Ingest, store, and transform data in a data lake
* Query, analyze, and visualize data within a data lake

**Outline**

* 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
* 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
  + Set up a simple data lake
* 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 analyze data in a data lake
* 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
  + Build a data lake using AWS Lake Formation
* 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 FindMatches
  + Visualize data with Amazon QuickSight
  + Automate data lake creation using AWS Lake Formation blueprints
  + Data visualization using Amazon QuickSight
* Architecture and Course Review
  + Post-course knowledge check
  + Architecture review
  + Course review
* Conclusion