

**Building Data Analytics Solutions Using Amazon Redshift**

**Course Number:** AWS-158  
**Duration:** 1 day

**Overview**

In this Building Data Analytics Solutions Using Amazon Redshift training course, attendees build a data analytics solution using Amazon Redshift, a cloud data warehouse service. The course focuses on the analytics pipeline's data collection, ingestion, cataloging, storage, and processing components. Students learn to integrate Amazon Redshift with a data lake to support both analytics and machine learning workloads. Participants also learn to apply security, performance, and cost management best practices to the operation of Amazon Redshift.

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**

* Completed either [AWS Technical Essentials](file:////training/aws-technical-essentials) or [Architecting on AWS](file:////training/aws-architecture)
* Completed [Building Data Lakes on AWS](file:////training/building-data-lakes-on-aws)

**Materials**

All Advanced AWS training students receive comprehensive courseware.

**Software Needed on Each Student PC**

A modern web browser and an Internet connection that allows connections by SSH or Remote Desktop (RDP) into AWS virtual machines.

**Objectives**

* Compare the features and benefits of data warehouses, data lakes, and modern data architectures
* Design and implement a data warehouse analytics solution
* Identify and apply appropriate techniques, including compression, to optimize data storage
* Select and deploy appropriate options to ingest, transform, and store data
* Choose the appropriate instance and node types, clusters, auto-scaling, and network topology for a particular business use case
* Understand how data storage and processing affect the analysis and visualization mechanisms needed to gain actionable business insights
* Secure data at rest and in transit
* Monitor analytics workloads to identify and remediate problems
* Apply cost management best practices

**Outline**

* Overview of Data Analytics and the Data Pipeline
  + Data analytics use cases
  + Using the data pipeline for analytics
* Using Amazon Redshift in the Data Analytics Pipeline
  + Why Amazon Redshift for data warehousing?
  + Overview of Amazon Redshift
* Introduction to Amazon Redshift
  + Amazon Redshift architecture
  + Touring the Amazon Redshift console
  + Amazon Redshift features
  + Load and query data in an Amazon Redshift cluster
* Ingestion and Storage
  + Ingestion
  + Connecting your Amazon Redshift cluster using a Jupyter notebook with
  + Data API
  + Data distribution and storage
  + Analyzing semi-structured data using the SUPER data type
  + Querying data in Amazon Redshift
  + Data analytics using Amazon Redshift Spectrum
* Processing and Optimizing Data
  + Data transformation
  + Advanced querying
  + Data transformation and querying in Amazon Redshift
  + Resource management
  + Applying mixed workload management on Amazon Redshift
  + Automation and optimization
  + Amazon Redshift cluster resizing from the dc2.large to ra3.xlplus cluster
* Security and Monitoring of Amazon Redshift Clusters
  + Securing the Amazon Redshift cluster
  + Monitoring and troubleshooting Amazon Redshift clusters
* Designing Data Warehouse Analytics Solutions
  + Data warehouse use case review
  + Designing a data warehouse analytics workflow
* Developing Modern Data Architectures on AWS