Data Warehousing on AWS

Course description

Data Warehousing on AWS introduces you to concepts, strategies, and best practices for designing a cloud-based data warehousing solution using Amazon Redshift, the petabyte-scale data warehouse in AWS. This course demonstrates how to collect, store, and prepare data for the data warehouse by using AWS services such as Amazon DynamoDB, Amazon EMR, Amazon Kinesis, and Amazon S3. Additionally, this course demonstrates how to use Amazon QuickSight to perform analysis on your data.

- Course level: Intermediate
- Duration: 3 days

Activities

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

Course objectives

In this course, you will:

- Discuss the core concepts of data warehousing, and the intersection between data warehousing and big data solutions
- Launch an Amazon Redshift cluster and use the components, features, and functionality to implement a data warehouse in the cloud
- Use other AWS data and analytic services, such as Amazon DynamoDB, Amazon EMR, Amazon Kinesis, and Amazon S3, to contribute to the data warehousing solution
- Architect the data warehouse
- Identify performance issues, optimize queries, and tune the database for better performance
- Use Amazon Redshift Spectrum to analyze data directly from an Amazon S3 bucket
- Use Amazon QuickSight to perform data analysis and visualization tasks against the data warehouse

Intended audience

This course is intended for:

- Database Architects
- Database Administrators
- Database Developers
- Data Analysts

- Data Scientists

Prerequisites

We recommend that attendees of this course have:

- Taken AWS Technical Essentials (or equivalent experience with AWS)
- Familiarity with relational databases and database design concepts

Enroll today

Visit aws.training to find a class today.

Course outline

Day 1

Module 1: Introduction to Data Warehousing

- Relational databases
- Data warehousing concepts
- The intersection of data warehousing and big data
- Overview of data management in AWS
- Hands-on lab 1: Introduction to Amazon Redshift

Module 2: Introduction to Amazon Redshift

- Conceptual overview
- Real-world use cases
- Hands-on lab 2: Launching an Amazon Redshift cluster

Module 3: Launching clusters

- Building the cluster
- Connecting to the cluster
- Controlling access
- Database security
- Load data
- Hands-on lab 3: Optimizing database schemas

Day 2

Module 4: Designing the database schema

- Schemas and data types
- Columnar compression
- Data distribution styles
- Data sorting methods

Module 5: Identifying data sources

- Data sources overview
- Amazon S3
- Amazon DynamoDB
- Amazon EMR
- Amazon Kinesis Data Firehose
- AWS Lambda Database Loader for Amazon Redshift
- Hands-on lab 4: Loading real-time data into an Amazon Redshift database

Module 6: Loading data

- Preparing Data
- Loading data using COPY
- Maintaining tables
- Concurrent write operations
- Troubleshooting load issues
- Hands-on lab 5: Loading data with the COPY command

Day 3

Module 7: Writing queries and tuning for performance

- Amazon Redshift SQL
- User-Defined Functions (UDFs)
- Factors that affect query performance
- The EXPLAIN command and query plans
- Workload Management (WLM)
- Hands-on lab 6: Configuring workload management

Module 8: Amazon Redshift Spectrum

- Amazon Redshift Spectrum
- Configuring data for Amazon Redshift Spectrum
- Amazon Redshift Spectrum Queries
- Hands-on lab 7: Using Amazon Redshift Spectrum

Module 9: Maintaining clusters

- Audit logging
- Performance monitoring
- Events and notifications
- Lab 8: Auditing and monitoring clusters
- Resizing clusters
- Backing up and restoring clusters
- Resource tagging and limits and constraints
- Hands-on lab 9: Backing up, restoring and resizing clusters

Module 10: Analyzing and visualizing data

- Power of visualizations
- Building dashboards
- Amazon QuickSight editions and features