Building Data Analytics Solutions Using Amazon Redshift

Amazon

- **Nível:**
- **Duração:** 7h

Sobre o curso

The course dives deep into the Amazon Redshift service and the current thinking in building and operating data analytics pipelines to turn data into insights.

In this course, you will build a data analytics solution using Amazon Redshift, a cloud data warehouse service. The course focuses on the data collection, ingestion, cataloging, storage, and processing components of the analytics pipeline. You will learn to integrate Amazon Redshift with a data lake to support both analytics and machine learning workloads. You will also learn to apply security, performance, and cost management best practices to the operation of Amazon Redshift.

Destinatários

This course is intended for data warehouse engineers, data platform engineers, and architects and operators who build and manage data analytics pipelines

---

Objetivos

In this course, you will learn to:

- 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

Pré-requisitos

- Students with a minimum one-year experience managing data warehouses will benefit from this course.
- We recommend that attendees of this course have:
    - Completed either AWS Technical Essentials or Architecting on AWS
    - Completed Building Data Lakes on AWS

Programa

- Overview of Data Analytics and the Data Pipeline
- Using Amazon Redshift in the Data Analytics Pipeline
- Introduction to Amazon Redshift
- Ingestion and Storage
- Processing and Optimizing Data
- Security and Monitoring of Amazon Redshift Clusters
- Designing Data Warehouse Analytics Solutions
- Developing Modern Data Architectures on AWS

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
- Interactive Demo 1: Touring the Amazon Redshift console
- Amazon Redshift features
- Practice Lab 1: Load and query data in an Amazon Redshift cluster

Ingestion and Storage

- Ingestion
- Interactive Demo 2: Connecting your Amazon Redshift cluster using a Jupyter notebook with Data API
- Data distribution and storage
- Interactive Demo 3: Analyzing semi-structured data using the SUPER data type
- Querying data in Amazon Redshift
- Practice Lab 2: Data analytics using Amazon Redshift Spectrum

Processing and Optimizing Data

- Data transformation
- Advanced querying
- Practice Lab 3: Data transformation and querying in Amazon Redshift
- Resource management
- Interactive Demo 4: Applying mixed workload management on Amazon Redshift
- Automation and optimization
- Interactive demo 5: 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
- Activity: Designing a data warehouse analytics workflow

Developing Modern Data Architectures on AWS

- Modern data architectures