



AWS – BIG DATA

CONTENIDOS

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

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

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

ARCHITECTURE AND COURSE REVIEW

- ▶ Post course knowledge check
- ▶ Architecture review
- ▶ Course review



OVERVIEW OF DATA ANALYTICS AND THE DATA PIPELINE

- ▶ Data analytics use cases
- ▶ Using the data pipeline for analytics

INTRODUCTION TO AMAZON EMR

- ▶ Using Amazon EMR in analytics solutions
- ▶ Amazon EMR cluster architecture
- ▶ Cost management strategies

DATA ANALYTICS PIPELINE USING AMAZON EMR: INGESTION AND STORAGE

- ▶ Storage optimization with Amazon EMR
- ▶ Data ingestion techniques

HIGH-PERFORMANCE BATCH DATA ANALYTICS USING APACHE SPARK ON AMAZON EMR

- ▶ Apache Spark on Amazon EMR use cases
- ▶ Why Apache Spark on Amazon EMR
- ▶ Spark concepts
- ▶ Transformation, processing, and analytics
- ▶ Using notebooks with Amazon EMR
- ▶ Practice Lab 1: Low-latency data analytics using Apache Spark on Amazon EMR

PROCESSING AND ANALYZING BATCH DATA WITH AMAZON EMR AND APACHE HIVE

- ▶ Using Amazon EMR with Hive to process batch data
- ▶ Transformation, processing, and analytics
- ▶ Introduction to Apache HBase on Amazon EMR

SERVERLESS DATA PROCESSING

- ▶ Serverless data processing, transformation, and analytics
- ▶ Using AWS Glue with Amazon EMR workloads

SECURITY AND MONITORING OF AMAZON EMR CLUSTERS

- ▶ Securing EMR clusters
- ▶ Interactive Demo 3: Client-side encryption with EMRFS
- ▶ Monitoring and troubleshooting Amazon EMR clusters

DESIGNING BATCH DATA ANALYTICS SOLUTIONS

- ▶ Batch data analytics use cases



AWS - DATA WAREHOUSING ON AWS

DEVELOPING MODERN DATA ARCHITECTURES ON AWS

- ▶ Modern data architectures

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



AWS - DATA WAREHOUSING ON AWS

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