

# **AWS Certified Data Engineer - Associate (DEA-C01) exam collection, Data-Engineer-Associate actual test**



BONUS!!! Download part of Exam4PDF Data-Engineer-Associate dumps for free: [https://drive.google.com/open?id=1wQ0Edp2siuNczgEuLfMbBXzU7ml8a\\_sH](https://drive.google.com/open?id=1wQ0Edp2siuNczgEuLfMbBXzU7ml8a_sH)

For candidates who are going to buy Data-Engineer-Associate Exam Materials online, they may have the concern about the website safety. If you choose us, we will offer you a clean and safe online shopping environment. In addition, Data-Engineer-Associate exam dumps are high quality and accuracy, and you can pass your exam just one time. We apply the international recognition third party for the payment, therefore your money safety can also be guaranteed. In order to let you access to the latest information, we offer you free update for 365 days after purchasing, and the update version will be sent to your email automatically.

We are willing to provide all people with the demo of our Data-Engineer-Associate study tool for free. If you have any doubt about our products that will bring a lot of benefits for you. The trial demo of our Data-Engineer-Associate question torrent must be a good choice for you. By the trial demo provided by our company, you will have the opportunity to closely contact with our Data-Engineer-Associate Exam Torrent, and it will be possible for you to have a view of our products. More importantly, we provide all people with the trial demo for free before you buy our Data-Engineer-Associate exam torrent and it means that you have the chance to download from our web page for free; you do not need to spend any money.

**>> Data-Engineer-Associate Reliable Exam Materials <<**

## **First-rank Data-Engineer-Associate Exam Preparation: AWS Certified Data Engineer - Associate (DEA-C01) boosts the Most Efficient Training Dumps - Exam4PDF**

We offer you free demo to you to have a try before buying Data-Engineer-Associate study guide, therefore you can have a better understanding of what you are going to buy. Free demo can be find in our website, if you are quite satisfied with the free demo, just add the Data-Engineer-Associate study guide to shopping cart, after you buy it, our system will send the downloading link and password to you within ten minutes, and you can start your learning right now. Moreover, we offer you free update for one year after you buy the Data-Engineer-Associate Exam Dumps, therefore you can get the latest version timely.

## **Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q197-Q202):**

### **NEW QUESTION # 197**

A company uses an on-premises Microsoft SQL Server database to store financial transaction data. The company migrates the transaction data from the on-premises database to AWS at the end of each month. The company has noticed that the cost to migrate data from the on-premises database to an Amazon RDS for SQL Server database has increased recently.

The company requires a cost-effective solution to migrate the data to AWS. The solution must cause minimal downtime for the applications that access the database.

Which AWS service should the company use to meet these requirements?

- A. AWS Lambda
- B. AWS DataSync
- **C. AWS Database Migration Service (AWS DMS)**
- D. AWS Direct Connect

**Answer: C**

Explanation:

AWS Database Migration Service (AWS DMS) is a cloud service that makes it possible to migrate relational databases, data warehouses, NoSQL databases, and other types of data stores to AWS quickly, securely, and with minimal downtime and zero data loss<sup>1</sup>. AWS DMS supports migration between 20-plus database and analytics engines, such as Microsoft SQL Server to Amazon RDS for SQL Server<sup>2</sup>. AWS DMS takes over many of the difficult or tedious tasks involved in a migration project, such as capacity analysis, hardware and software procurement, installation and administration, testing and debugging, and ongoing replication and monitoring<sup>1</sup>. AWS DMS is a cost-effective solution, as you only pay for the compute resources and additional log storage used during the migration process<sup>2</sup>. AWS DMS is the best solution for the company to migrate the financial transaction data from the on-premises Microsoft SQL Server database to AWS, as it meets the requirements of minimal downtime, zero data loss, and low cost.

Option A is not the best solution, as AWS Lambda is a serverless compute service that lets you run code without provisioning or managing servers, but it does not provide any built-in features for database migration. You would have to write your own code to extract, transform, and load the data from the source to the target, which would increase the operational overhead and complexity.

Option C is not the best solution, as AWS Direct Connect is a service that establishes a dedicated network connection from your premises to AWS, but it does not provide any built-in features for database migration. You would still need to use another service or tool to perform the actual data transfer, which would increase the cost and complexity.

Option D is not the best solution, as AWS DataSync is a service that makes it easy to transfer data between on-premises storage systems and AWS storage services, such as Amazon S3, Amazon EFS, and Amazon FSx for Windows File Server, but it does not support Amazon RDS for SQL Server as a target. You would have to use another service or tool to migrate the data from Amazon S3 to Amazon RDS for SQL Server, which would increase the latency and complexity. Reference:

Database Migration - AWS Database Migration Service - AWS

What is AWS Database Migration Service?

AWS Database Migration Service Documentation

AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

## NEW QUESTION # 198

A company wants to analyze sales records that the company stores in a MySQL database. The company wants to correlate the records with sales opportunities identified by Salesforce.

The company receives 2 GB of sales records every day. The company has 100 GB of identified sales opportunities. A data engineer needs to develop a process that will analyze and correlate sales records and sales opportunities. The process must run once each night.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to fetch both datasets. Use AWS Lambda functions to correlate the datasets. Use AWS Step Functions to orchestrate the process.
- **B. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with sales opportunities. Use AWS Step Functions to orchestrate the process.**
- C. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use Amazon Kinesis Data Streams to fetch sales records from the MySQL database. Use Amazon Managed Service for Apache Flink to correlate the datasets. Use AWS Step Functions to orchestrate the process.
- D. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with the sales opportunities. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the process.

**Answer: B**

Explanation:

\* Problem Analysis:

\* The company processes 2 GB of daily sales records and 100 GB of Salesforce sales opportunities.

\* The goal is to analyze and correlate the two datasets with low operational overhead.

\* The process must run once nightly.

\* Key Considerations:

\* Amazon AppFlow simplifies data integration with Salesforce.

- \* AWS Glue can extract data from MySQL and perform ETL operations.
- \* Step Functions can orchestrate workflows with minimal manual intervention.
- \* Apache Airflow and Flink add complexity, which conflicts with the requirement for low operational overhead.
- \* Solution Analysis:
- \* Option A: MWAA + Lambda + Step Functions
- \* Requires custom Lambda code for dataset correlation, increasing development and operational complexity.
- \* Option B: AppFlow + Glue + MWAA
- \* MWAA adds orchestration overhead compared to the simpler Step Functions.
- \* Option C: AppFlow + Glue + Step Functions
- \* AppFlow fetches Salesforce data, Glue extracts MySQL data, and Step Functions orchestrate the entire process.
- \* Minimal setup and operational overhead, making it the best choice.
- \* Option D: AppFlow + Kinesis + Flink + Step Functions
- \* Using Kinesis and Flink for batch processing introduces unnecessary complexity.
- \* Final Recommendation:
- \* Use Amazon AppFlow to fetch Salesforce data, AWS Glue to process MySQL data, and Step Functions for orchestration.

:

Amazon AppFlow Overview

AWS Glue ETL Documentation

AWS Step Functions

### NEW QUESTION # 199

A retail company uses an Amazon Redshift data warehouse and an Amazon S3 bucket. The company ingests retail order data into the S3 bucket every day.

The company stores all order data at a single path within the S3 bucket. The data has more than 100 columns.

The company ingests the order data from a third-party application that generates more than 30 files in CSV format every day. Each CSV file is between 50 and 70 MB in size.

The company uses Amazon Redshift Spectrum to run queries that select sets of columns. Users aggregate metrics based on daily orders. Recently, users have reported that the performance of the queries has degraded.

A data engineer must resolve the performance issues for the queries.

Which combination of steps will meet this requirement with LEAST developmental effort? (Select TWO.)

- **A. Configure the third-party application to create the files in a columnar format.**
- B. Load the JSON data into the Amazon Redshift table in a SUPER type column.
- C. Develop an AWS Glue ETL job to convert the multiple daily CSV files to one file for each day.
- D. Configure the third-party application to create the files in JSON format.
- **E. Partition the order data in the S3 bucket based on order date.**

**Answer: A,E**

Explanation:

The performance issue in Amazon Redshift Spectrum queries arises due to the nature of CSV files, which are row-based storage formats. Spectrum is more optimized for columnar formats, which significantly improve performance by reducing the amount of data scanned. Also, partitioning data based on relevant columns like order date can further reduce the amount of data scanned, as queries can focus only on the necessary partitions.

\* A. Configure the third-party application to create the files in a columnar format:

\* Columnar formats (like Parquet or ORC) store data in a way that is optimized for analytical queries because they allow queries to scan only the columns required, rather than scanning all columns in a row-based format like CSV.

\* Amazon Redshift Spectrum works much more efficiently with columnar formats, reducing the amount of data that needs to be scanned, which improves query performance.

Reference: Amazon Redshift Spectrum and Columnar File Formats

C: Partition the order data in the S3 bucket based on order date:

Partitioning the data on columns like order date allows Redshift Spectrum to skip scanning unnecessary partitions, leading to improved query performance.

By organizing data into partitions, you minimize the number of files Spectrum has to read, further optimizing performance.

Reference: Best Practices for Amazon Redshift Spectrum Performance

Alternatives Considered:

B (Develop an AWS Glue ETL job): While consolidating files can improve performance by reducing the number of small files (which can be inefficient to process), it adds additional ETL complexity. Switching to a columnar format (Option A) and partitioning (Option C) provides more significant performance improvements with less development effort.

D and E (JSON-related options): Using JSON format or the SUPER type in Redshift introduces complexity and isn't as efficient as

the proposed solutions, especially since JSON is not a columnar format.

References:

Amazon Redshift Spectrum Documentation

Columnar Formats and Data Partitioning in S3

### NEW QUESTION # 200

A company stores sales data in an Amazon RDS for MySQL database. The company needs to start a reporting process between 6:00 A.M. and 6:10 A.M. every Monday. The reporting process must generate a CSV file and store the file in an Amazon S3 bucket.

Which combination of steps will meet these requirements with the LEAST operational overhead? (Select TWO.)

- A. Create and invoke an AWS Batch job that runs a script in an Amazon Elastic Container Service (Amazon ECS) container. Configure the script to generate the report and to save it to the S3 bucket.
- **B. Create an Amazon EventBridge Scheduler to run every Monday at 6:00 A.M.**
- C. Create and invoke an Amazon EMR Serverless job to generate the report and to save it to the S3 bucket.
- **D. Create and invoke an AWS Glue ETL job to generate the report and to save it to the S3 bucket.**
- E. Create an Amazon EventBridge rule to run every Monday at 6:00 A.M.

**Answer: B,D**

Explanation:

The Amazon EventBridge Scheduler offers a simple, serverless cron-based execution mechanism. It can trigger an AWS Glue ETL job that extracts data from Amazon RDS, formats it as CSV, and writes it to Amazon S3 - all without manual orchestration or servers.

"For scheduled data extraction and transformation, use AWS Glue jobs triggered by EventBridge Scheduler for fully managed, low-maintenance workflows."

- Ace the AWS Certified Data Engineer - Associate Certification - version 2 - apple.pdf Glue natively integrates with RDS and S3, avoiding the need to manage Batch or EMR infrastructure.

### NEW QUESTION # 201

A company has an Amazon Redshift data warehouse that users access by using a variety of IAM roles. More than 100 users access the data warehouse every day.

The company wants to control user access to the objects based on each user's job role, permissions, and how sensitive the data is.

Which solution will meet these requirements?

- A. Use the column-level security (CLS) feature of Amazon Redshift.
- B. Use dynamic data masking policies in Amazon Redshift.
- **C. Use the role-based access control (RBAC) feature of Amazon Redshift.**
- D. Use the row-level security (RLS) feature of Amazon Redshift.

**Answer: C**

Explanation:

Amazon Redshift supports Role-Based Access Control (RBAC) to manage access to database objects. RBAC allows administrators to create roles for job functions and assign privileges at the schema, table, or column level based on data sensitivity and user roles.

"RBAC in Amazon Redshift helps manage permissions more efficiently at scale by assigning users to roles that reflect their job function. It simplifies user management and secures access based on job role and data sensitivity."

- Ace the AWS Certified Data Engineer - Associate Certification - version 2 - apple.pdf RBAC is preferred over RLS or CLS alone because it offers a more comprehensive and scalable solution across multiple users and permissions.

### NEW QUESTION # 202

.....

We boost a professional expert team to undertake the research and the production of our Data-Engineer-Associate study materials. We employ the senior lecturers and authorized authors who have published the articles about the test to compile and organize the Data-Engineer-Associate study materials. Our expert team boosts profound industry experiences and they use their precise logic to verify the test. They provide comprehensive explanation and integral details of the answers and questions. Each question and answer

- [illegible]

pct.edu.pk, www.stes.tyc.edu.tw, Disposable vapes

DOWNLOAD the newest Exam4PDF Data-Engineer-Associate PDF dumps from Cloud Storage for free:  
[https://drive.google.com/open?id=1wQ0Edp2siuNczgEuLfMbBXzU7m18a\\_sH](https://drive.google.com/open?id=1wQ0Edp2siuNczgEuLfMbBXzU7m18a_sH)