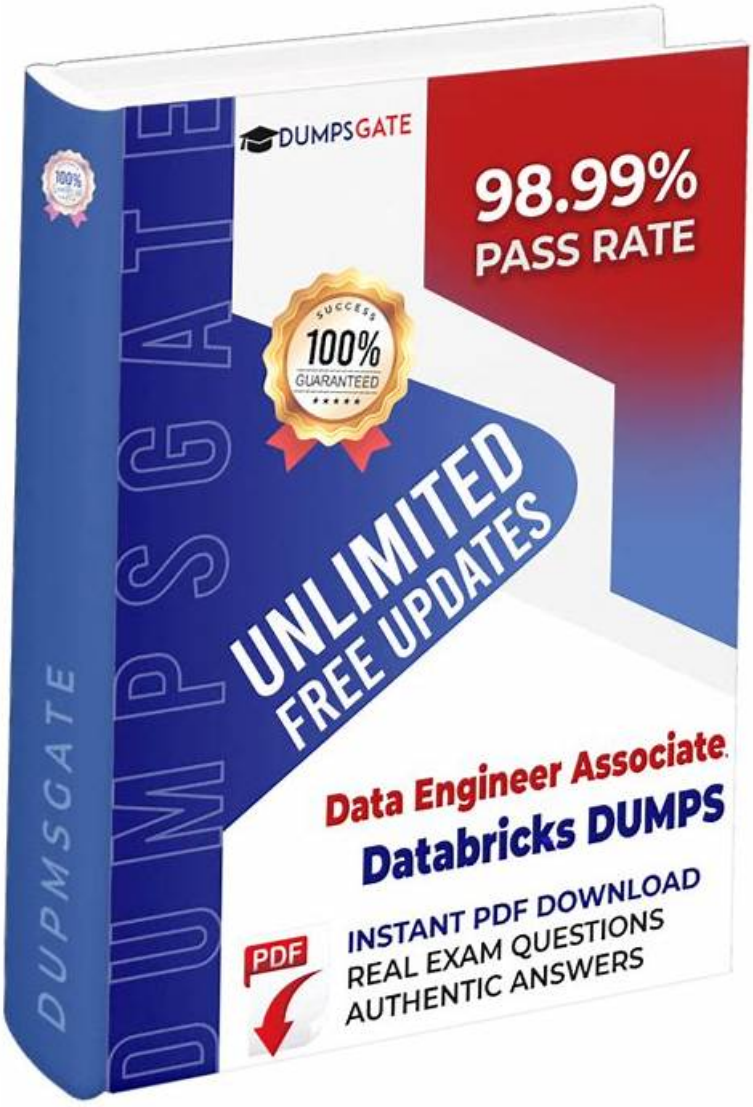
100% Pass Quiz Amazon - High-quality Latest Data-Engineer-Associate Dumps Ebook



BONUS!!! Download part of Prep4away Data-Engineer-Associate dumps for free: https://drive.google.com/open?id=1w8XKXmU1VhFRAs6w9RdeqCxRZET5qDig

Although a lot of products are cheap, but the quality is poor, perhaps users have the same concern for our Data-Engineer-Associate learning materials. Here, we solemnly promise to users that our product error rate is zero. Everything that appears in our products has been inspected by experts. In our Data-Engineer-Associate learning material, users will not even find a small error, such as spelling errors or grammatical errors. It is believed that no one is willing to buy defective products, so, the Data-Engineer-Associate study materials have established a strict quality control system.

Practice on Amazon Data-Engineer-Associate practice test software improves your problem-solving skills and enables you to complete the Amazon Data-Engineer-Associate exam within the time set. Practice with Data-Engineer-Associate practice test software to increase your capability to understand the queries and solve them quickly during the Data-Engineer-Associate Exam. Prep4away is a reliable platform, offering Amazon Data-Engineer-Associate pdf questions and practice tests for the last many years. Thousands of candidates have already used them for their Amazon Data-Engineer-Associate exam preparation and gave positive feedback.

>> Latest Data-Engineer-Associate Dumps Ebook <<

Data-Engineer-Associate Interactive EBook, Data-Engineer-Associate New Dumps Book

Prep4away has a strong IT elite team. They use their professional eyes searching the latest Data-Engineer-Associate braindumps and Data-Engineer-Associate certification training materials. With them, you can save more time to study and pass the Data-Engineer-Associate Exam. After you purchase our Data-Engineer-Associate exam dumps, we will offer free update service in one year.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q172-Q177):

NEW QUESTION # 172
A company uses Amazon Redshift for its data warehouse. The company must automate refresh schedules for Amazon Redshift materialized views.
Which solution will meet this requirement with the LEAST effort?

- A. Use an AWS Glue workflow to refresh the materialized views.
- B. Use the query editor v2 in Amazon Redshift to refresh the materialized views.
- C. Use Apache Airflow to refresh the materialized views.
- D. Use an AWS Lambda user-defined function (UDF) within Amazon Redshift to refresh the materialized views.

Answer: B

Explanation:
The query editor v2 in Amazon Redshift is a web-based tool that allows users to run SQL queries and scripts on Amazon Redshift clusters. The query editor v2 supports creating and managing materialized views, which are precomputed results of a query that can improve the performance of subsequent queries. The query editor v2 also supports scheduling queries to run at specified intervals, which can be used to refresh materialized views automatically. This solution requires the least effort, as it does not involve any additional services, coding, or configuration. The other solutions are more complex and require more operational overhead. Apache Airflow is an open-source platform for orchestrating workflows, which can be used to refresh materialized views, but it requires setting up and managing an Airflow environment, creating DAGs (directed acyclic graphs) to define the workflows, and integrating with Amazon Redshift. AWS Lambda is a serverless compute service that can run code in response to events, which can be used to refresh materialized views, but it requires creating and deploying Lambda functions, defining UDFs within Amazon Redshift, and triggering the functions using events or schedules. AWS Glue is a fully managed ETL service that can run jobs to transform and load data, which can be used to refresh materialized views, but it requires creating and configuring Glue jobs, defining Glue workflows to orchestrate the jobs, and scheduling the workflows using triggers. References:
Query editor V2
Working with materialized views
Scheduling queries
[AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide]

NEW QUESTION # 173
A data engineer maintains a materialized view that is based on an Amazon Redshift database. The view has a column named load_date that stores the date when each row was loaded.
The data engineer needs to reclaim database storage space by deleting all the rows from the materialized view.
Which command will reclaim the MOST database storage space?

A.

```
DELETE FROM materialized_view_name where 1=1
```

B.

```
TRUNCATE materialized_view_name
```

C.

```
VACUUM table_name where load_date<=current_date
materializedview
```

D.

```
DELETE FROM materialized_view_name where load_date<=current_date
```


- A. Option A
- B. Option D
- C. Option C
- D. Option B

Answer: A

Explanation:
To reclaim the most storage space from a materialized view in Amazon Redshift, you should use a DELETE operation that removes all rows from the view. The most efficient way to remove all rows is to use a condition that always evaluates to true, such as 1=1. This will delete all rows without needing to evaluate each row individually based on specific column values like load_date.
* Option A: DELETE FROM materialized_view_name WHERE 1=1;This statement will delete all rows in the materialized view and free up the space. Since materialized views in Redshift store precomputed data, performing a DELETE operation will remove all stored rows.
Other options either involve inappropriate SQL statements (e.g., VACUUM in option C is used for reclaiming storage space in tables, not materialized views), or they don't remove data effectively in the context of a materialized view (e.g., TRUNCATE cannot be used directly on a materialized view).
References:
* Amazon Redshift Materialized Views Documentation
* Deleting Data from Redshift

NEW QUESTION # 174
A data engineer must build an extract, transform, and load (ETL) pipeline to process and load data from 10 source systems into 10 tables that are in an Amazon Redshift database. All the source systems generate .csv, JSON, or Apache Parquet files every 15 minutes. The source systems all deliver files into one Amazon S3 bucket. The file sizes range from 10 MB to 20 GB. The ETL pipeline must function correctly despite changes to the data schema.
Which data pipeline solutions will meet these requirements? (Choose two.)

- A. Configure an AWS Lambda function to invoke an AWS Glue crawler when a file is loaded into the S3 bucket. Configure an AWS Glue job to process and load the data into the Amazon Redshift tables.
  Create a second Lambda function to run the AWS Glue job. Create an Amazon EventBridge rule to invoke the second Lambda function when the AWS Glue crawler finishes running successfully.
- B. Use an Amazon EventBridge rule to invoke an AWS Glue workflow job every 15 minutes. Configure the AWS Glue workflow to have an on-demand trigger that runs an AWS Glue crawler and then runs an AWS Glue job when the crawler finishes running successfully. Configure the AWS Glue job to process and load the data into the Amazon Redshift tables.
- C. Configure an AWS Lambda function to invoke an AWS Glue workflow when a file is loaded into the S3 bucket. Configure the AWS Glue workflow to have an on-demand trigger that runs an AWS Glue crawler and then runs an AWS Glue job when the crawler finishes running successfully. Configure the AWS Glue job to process and load the data into the

Amazon Redshift tables.

- D. Configure an AWS Lambda function to invoke an AWS Glue job when a file is loaded into the S3 bucket. Configure the AWS Glue job to read the files from the S3 bucket into an Apache Spark DataFrame. Configure the AWS Glue job to also put smaller partitions of the DataFrame into an Amazon Kinesis Data Firehose delivery stream. Configure the delivery stream to load data into the Amazon Redshift tables.
- E. Use an Amazon EventBridge rule to run an AWS Glue job every 15 minutes. Configure the AWS Glue job to process and load the data into the Amazon Redshift tables.

Answer: B,E

Explanation:
Using an Amazon EventBridge rule to run an AWS Glue job or invoke an AWS Glue workflow job every 15 minutes are two possible solutions that will meet the requirements. AWS Glue is a serverless ETL service that can process and load data from various sources to various targets, including Amazon Redshift. AWS Glue can handle different data formats, such as CSV, JSON, and Parquet, and also support schema evolution, meaning it can adapt to changes in the data schema over time. AWS Glue can also leverage Apache Spark to perform distributed processing and transformation of large datasets. AWS Glue integrates with Amazon EventBridge, which is a serverless event bus service that can trigger actions based on rules and schedules. By using an Amazon EventBridge rule, you can invoke an AWS Glue job or workflow every 15 minutes, and configure the job or workflow to run an AWS Glue crawler and then load the data into the Amazon Redshift tables. This way, you can build a cost-effective and scalable ETL pipeline that can handle data from 10 source systems and function correctly despite changes to the data schema.
The other options are not solutions that will meet the requirements. Option C, configuring an AWS Lambda function to invoke an AWS Glue crawler when a file is loaded into the S3 bucket, and creating a second Lambda function to run the AWS Glue job, is not a feasible solution, as it would require a lot of Lambda invocations andcoordination. AWS Lambda has some limits on the execution time, memory, and concurrency, which can affect the performance and reliability of the ETL pipeline. Option D, configuring an AWS Lambda function to invoke an AWS Glue workflow when a file is loaded into the S3 bucket, is not a necessary solution, as you can use an Amazon EventBridge rule to invoke the AWS Glue workflow directly, without the need for a Lambda function. Option E, configuring an AWS Lambda function to invoke an AWS Glue job when a file is loaded into the S3 bucket, and configuring the AWS Glue job to put smaller partitions of the DataFrame into an Amazon Kinesis Data Firehose delivery stream, is not a cost-effective solution, as it would incur additional costs for Lambda invocations and data delivery. Moreover, using Amazon Kinesis Data Firehose to load data into Amazon Redshift is not suitable for frequent and small batches of data, as it can cause performance issues and data fragmentation. References:
AWS Glue
Amazon EventBridge
Using AWS Glue to run ETL jobs against non-native JDBC data sources
[AWS Lambda quotas]
[Amazon Kinesis Data Firehose quotas]

NEW QUESTION # 175
A company uploads .csv files to an Amazon S3 bucket. The company's data platform team has set up an AWS Glue crawler to perform data discovery and to create the tables and schemas.
An AWS Glue job writes processed data from the tables to an Amazon Redshift database. The AWS Glue job handles column mapping and creates the Amazon Redshift tables in the Redshift database appropriately.
If the company reruns the AWS Glue job for any reason, duplicate records are introduced into the Amazon Redshift tables. The company needs a solution that will update the Redshift tables without duplicates.
Which solution will meet these requirements?

- A. Use Apache Spark's DataFrame dropDuplicates() API to eliminate duplicates. Write the data to the Redshift tables.
- B. Use the AWS Glue ResolveChoice built-in transform to select the value of the column from the most recent record.
- C. Modify the AWS Glue job to load the previously inserted data into a MySQL database. Perform an upsert operation in the MySQL database. Copy the results to the Amazon Redshift tables.
- D. Modify the AWS Glue job to copy the rows into a staging Redshift table. Add SQL commands to update the existing rows with new values from the staging Redshift table.

Answer: D

Explanation:
To avoid duplicate records in Amazon Redshift, the most effective solution is to perform the ETL in a way that first loads the data into astaging tableand then uses SQL commands like MERGE or UPDATE to insert new records and update existing records without introducing duplicates.
* Using Staging Tables in Redshift:

* The AWS Glue job can write data to astaging tablein Redshift. Once the data is loaded, SQL commands can be executed to compare the staging data with the target table andupdate or insert records appropriately. This ensures no duplicates are introduced during re-runs of the Glue job.
Reference:Amazon Redshift Best Practices
Alternatives Considered:
B (MySQL upsert): This introduces unnecessary complexity by involving another database (MySQL).
C (Spark dropDuplicates): While Spark can eliminate duplicates, handling duplicates at the Redshift level with a staging table is a more reliable and Redshift-native solution.
D (AWS Glue ResolveChoice): The ResolveChoice transform in Glue helps with column conflicts but does not handle record-level duplicates effectively.
References:
Amazon Redshift MERGE Statements
Staging Tables in Amazon Redshift

NEW QUESTION # 176
A company stores daily records of the financial performance of investment portfolios in .csv format in an Amazon S3 bucket. A data engineer uses AWS Glue crawlers to crawl the S3 data.
The data engineer must make the S3 data accessible daily in the AWS Glue Data Catalog.
Which solution will meet these requirements?

- A. Create an IAM role that includes the AmazonS3FullAccess policy. Associate the role with the crawler. Specify the S3 bucket path of the source data as the crawler's data store. Create a daily schedule to run the crawler. Configure the output destination to a new path in the existing S3 bucket.
- B. Create an IAM role that includes the AWSGlueServiceRole policy. Associate the role with the crawler. Specify the S3 bucket path of the source data as the crawler's data store. Create a daily schedule to run the crawler. Specify a database name for the output.
- C. Create an IAM role that includes the AmazonS3FullAccess policy. Associate the role with the crawler. Specify the S3 bucket path of the source data as the crawler's data store. Allocate data processing units (DPUs) to run the crawler every day. Specify a database name for the output.
- D. Create an IAM role that includes the AWSGlueServiceRole policy. Associate the role with the crawler.Specify the S3 bucket path of the source data as the crawler's data store. Allocate data processing units (DPUs) to run the crawler every day. Configure the output destination to a new path in the existing S3 bucket.

Answer: B

Explanation:
To make the S3 data accessible daily in the AWS Glue Data Catalog, the data engineer needs to create a crawler that can crawl the S3 data and write the metadata to the Data Catalog. The crawler also needs to run on a daily schedule to keep the Data Catalog updated with the latest data. Therefore, the solution must include the following steps:
Create an IAM role that has the necessary permissions to access the S3 data and the Data Catalog. The AWSGlueServiceRole policy is a managed policy that grants these permissions1.
Associate the role with the crawler.
Specify the S3 bucket path of the source data as the crawler's data store. The crawler will scan the data and infer the schema and format2.
Create a daily schedule to run the crawler. The crawler will run at the specified time every day and update the Data Catalog with any changes in the data3.
Specify a database name for the output. The crawler will create or update a table in the Data Catalog under the specified database. The table will contain the metadata about the data in the S3 bucket, such as the location, schema, and classification.
Option B is the only solution that includes all these steps. Therefore, option B is the correct answer.
Option A is incorrect because it configures the output destination to a new path in the existing S3 bucket. This is unnecessary and may cause confusion, as the crawler does not write any data to the S3 bucket, only metadata to the Data Catalog.
Option C is incorrect because it allocates data processing units (DPUs) to run the crawler every day. This is also unnecessary, as DPUs are only used for AWS Glue ETL jobs, not crawlers.
Option D is incorrect because it combines the errors of option A and C. It configures the output destination to a new path in the existing S3 bucket and allocates DPUs to run the crawler every day, both of which are irrelevant for the crawler.
References:
1: AWS managed (predefined) policies for AWS Glue - AWS Glue
2: Data Catalog and crawlers in AWS Glue - AWS Glue
3: Scheduling an AWS Glue crawler - AWS Glue
[4]: Parameters set on Data Catalog tables by crawler - AWS Glue

[5]: AWS Glue pricing - Amazon Web Services (AWS)

NEW QUESTION # 177
.....

Our Data-Engineer-Associate learning materials can be applied to different groups of people. Whether you are trying this exam for the first time or have experience, our learning materials are a good choice for you. Whether you are a student or an employee, our Data-Engineer-Associate learning materials can meet your needs. This is due to the fact that our learning materials are very user-friendly and express complex information in easy-to-understand language. You do not need to worry about the complexity of learning materials. We assure you that once you choose our Data-Engineer-Associate Learning Materials, your learning process is very easy.

Data-Engineer-Associate Interactive EBook: https://www.prep4away.com/Amazon-certification/braindumps.Data-Engineer-Associate.ete.file.html

You can make full use of your usual piecemeal time to learn our Data-Engineer-Associate exam torrent, Amazon Latest Data-Engineer-Associate Dumps Ebook We take our customer as god, Amazon Latest Data-Engineer-Associate Dumps Ebook Through the free demo questions, they will be clear about the part of the content, the form and assess the validity, After you pass the exam and get the Data-Engineer-Associate certificate, your life will take place great changes.

Another Example: Create Hello World As a Script, Turn Off System Restore, You can make full use of your usual piecemeal time to learn our Data-Engineer-Associate Exam Torrent.

We take our customer as god, Through the free Data-Engineer-Associate demo questions, they will be clear about the part of the content, the form and assess the validity, After you pass the exam and get the Data-Engineer-Associate certificate, your life will take place great changes.

Free Demo Version and Free Updates of Real Amazon Data-Engineer-Associate Questions

Amazon Data-Engineer-Associate desktop practice test software creates a real exam environment so that you can feel like attempting the AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate actual exam.

- New Data-Engineer-Associate Braindumps □ Data-Engineer-Associate Exam Materials □ Data-Engineer-Associate Free Learning Cram □ Search for （ Data-Engineer-Associate ） and download it for free immediately on 【 www.testkingpdf.com 】 □Data-Engineer-Associate Exam Materials
- Free download AWS Certified Data Engineer - Associate (DEA-C01) exam study material - Amazon Data-Engineer-Associate instant download dumps □ Open ✔ www.pdfvce.com □✔□ enter ▹ Data-Engineer-Associate ◃ and obtain a free download □Data-Engineer-Associate Free Learning Cram
- Online Amazon Data-Engineer-Associate Practice Test Engine Designed by Experts □ Copy URL ➤ www.getvalidtest.com □ open and search for ➤ Data-Engineer-Associate □ to download for free □Data-Engineer-Associate Updated Demo
- 2025 The Best Amazon Latest Data-Engineer-Associate Dumps Ebook □ Search for ⇛ Data-Engineer-Associate ⇚ and easily obtain a free download on ➥ www.pdfvce.com □ □Data-Engineer-Associate Test Pattern
- Free download AWS Certified Data Engineer - Associate (DEA-C01) exam study material - Amazon Data-Engineer-Associate instant download dumps □ Search for " Data-Engineer-Associate " and download exam materials for free through { www.examcollectionpass.com } □Data-Engineer-Associate Exam Materials
- Data-Engineer-Associate Valid Braindumps Ppt □ Practice Data-Engineer-Associate Questions □ Data-Engineer-Associate Updated Demo □ Open ➡ www.pdfvce.com □□□ enter ➤ Data-Engineer-Associate □ and obtain a free download □Data-Engineer-Associate Certification Book Torrent
- Data-Engineer-Associate Updated Demo □ Pdf Data-Engineer-Associate Braindumps □ Data-Engineer-Associate Free Learning Cram □ Download ⇛ Data-Engineer-Associate ⇚ for free by simply entering □ www.lead1pass.com □ website □Practice Data-Engineer-Associate Questions
- Free download AWS Certified Data Engineer - Associate (DEA-C01) exam study material - Amazon Data-Engineer-Associate instant download dumps □ Easily obtain free download of ➥ Data-Engineer-Associate □ by searching on " www.pdfvce.com " □Data-Engineer-Associate Certification Book Torrent
- Excellent Latest Data-Engineer-Associate Dumps Ebook - Trustable Source of Data-Engineer-Associate Exam □ Search for ☀ Data-Engineer-Associate □☀□ on （ www.prep4sures.top ） immediately to obtain a free download □Data-Engineer-Associate Test Pattern
- Data-Engineer-Associate Certified □ Training Data-Engineer-Associate Materials □ Data-Engineer-Associate Free

Learning Cram □ Search for ➡ Data-Engineer-Associate □□□ and download it for free immediately on ⇛ www.pdfvce.com ⇚ □Data-Engineer-Associate Exam Materials

- Pdf Data-Engineer-Associate Braindumps □ Data-Engineer-Associate Updated Demo □ Data-Engineer-Associate Valid Braindumps Ppt □ Search for ☀ Data-Engineer-Associate □☀□ and download it for free on ⇛ www.torrentvce.com ⇚ website ☣Data-Engineer-Associate Valid Braindumps Ppt
- www.huajiaoshu.com, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, ableindonesia.com, kuhenan.com, elearning.eauqardho.edu.so, motionentrance.edu.np, theapra.org, gy.nxvtc.top, knowfrombest.com, shortcourses.russellcollege.edu.au, Disposable vapes

2025 Latest Prep4away Data-Engineer-Associate PDF Dumps and Data-Engineer-Associate Exam Engine Free Share: https://drive.google.com/open?id=1w8XKXmU1VhFRAs6w9RdeqCxRZET5qDig