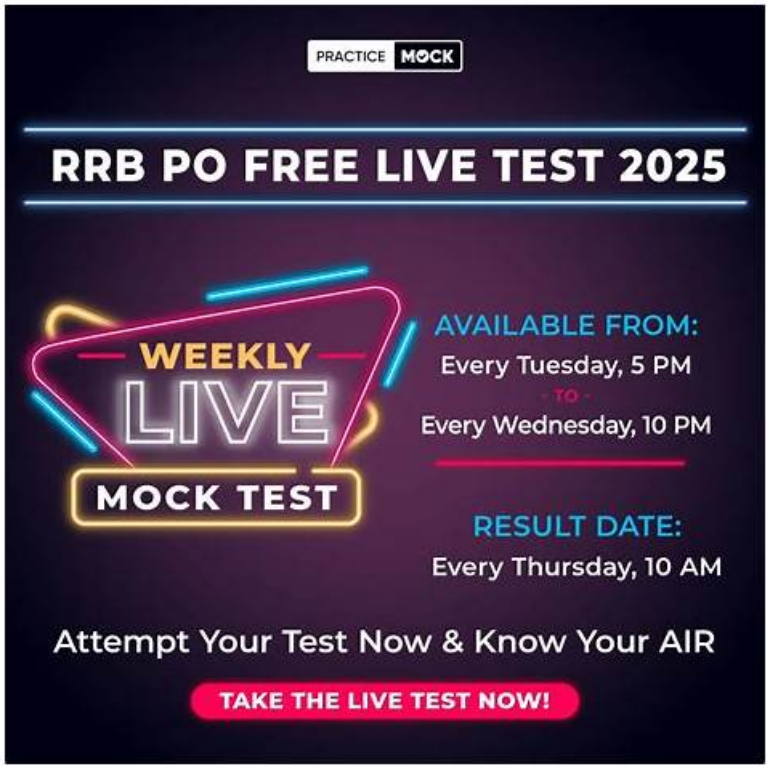
Free PDF Quiz 2025 Amazon Marvelous Free Data-Engineer-Associate Test Questions



2025 Latest DumpStillValid Data-Engineer-Associate PDF Dumps and Data-Engineer-Associate Exam Engine Free Share: https://drive.google.com/open?id=12gCRYxuRB7h7aNqcdrIpvF128aTUZflC

With DumpStillValid, you can trust that you're accessing authentic and error-free Data-Engineer-Associate exam practice questions. These questions are available in three different formats: PDF questions files, desktop practice test software, and web-based practice test software. All three formats contain genuine Data-Engineer-Associate Practice Questions that will effectively prepare you for the final exam.

In order to solve customers' problem in the shortest time, our Data-Engineer-Associate guide torrent provides the twenty four hours online service for all people. Maybe you have some questions about our Data-Engineer-Associate test torrent when you use our products; it is your right to ask us in anytime and anywhere. You just need to send us an email, our online workers are willing to reply you an email to solve your problem on our Data-Engineer-Associate Exam Questions. During the process of using our Data-Engineer-Associate study torrent, we can promise you will have the right to enjoy the twenty four hours online service provided by our online workers.

>> Free Data-Engineer-Associate Test Questions <<

Data-Engineer-Associate Pdf Dumps | Online Data-Engineer-Associate Training

Success in the Amazon Data-Engineer-Associate Exam paves the way toward high-paying jobs, promotions, and skills verification. Hundreds of Amazon Data-Engineer-Associate test takers don't get success because of using Amazon outdated dumps. Due to failure, they lose money, time, and confidence. All these losses can be prevented by using updated and real Amazon Dumps of DumpStillValid.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q55-Q60):

NEW QUESTION # 55
A company is migrating a legacy application to an Amazon S3 based data lake. A data engineer reviewed data that is associated with the legacy application. The data engineer found that the legacy data contained some duplicate information.
The data engineer must identify and remove duplicate information from the legacy application data.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Write a custom extract, transform, and load (ETL) job in Python. Use the DataFramedrop duplicatesf) function by importing the Pandas library to perform data deduplication.
- B. Write an AWS Glue extract, transform, and load (ETL) job. Use the FindMatches machine learning (ML) transform to transform the data to perform data deduplication.
- C. Write a custom extract, transform, and load (ETL) job in Python. Import the Python dedupe library. Use the dedupe library to perform data deduplication.
- D. Write an AWS Glue extract, transform, and load (ETL) job. Import the Python dedupe library. Use the dedupe library to perform data deduplication.

Answer: B

Explanation:
AWS Glue is a fully managed serverless ETL service that can handle data deduplication with minimal operational overhead. AWS Glue provides a built-in ML transform called FindMatches, which can automatically identify and group similar records in a dataset. FindMatches can also generate a primary key for each group of records and remove duplicates. FindMatches does not require any coding or prior ML experience, as it can learn from a sample of labeled data provided by the user. FindMatches can also scale to handle large datasets and optimize the cost and performance of the ETL job. Reference:
AWS Glue
FindMatches ML Transform
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 56
A company stores datasets in JSON format and .csv format in an Amazon S3 bucket. The company has Amazon RDS for Microsoft SQL Server databases, Amazon DynamoDB tables that are in provisionedcapacity mode, and an Amazon Redshift cluster. A data engineering team must develop a solution that will give data scientists the ability to query all data sources by using syntax similar to SQL.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use Amazon Athena to query the data. Use SQL for structured data sources. Use PartiQL for data that is stored in JSON format.
- B. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use Redshift Spectrum to query the data. Use SQL for structured data sources. Use PartiQL for data that is stored in JSON format.
- C. Use AWS Glue to crawl the data sources. Store metadata in the AWS Glue Data Catalog. Use AWS Glue jobs to transform data that is in JSON format to Apache Parquet or .csv format. Store the transformed data in an S3 bucket. Use Amazon Athena to query the original and transformed data from the S3 bucket.
- D. Use AWS Lake Formation to create a data lake. Use Lake Formation jobs to transform the data from all data sources to Apache Parquet format. Store the transformed data in an S3 bucket. Use Amazon Athena or Redshift Spectrum to query the data.

Answer: A

Explanation:
The best solution to meet the requirements of giving data scientists the ability to query all data sources by using syntax similar to SQL with the least operational overhead is to use AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use Amazon Athena to query the data, use SQL for structured data sources, and use PartiQL for data that is stored in JSON format.
AWS Glue is a serverless data integration service that makes it easy to prepare, clean, enrich, and move data between data stores1. AWS Glue crawlers are processes that connect to a data store, progress through a prioritized list of classifiers to determine the schema for your data, and then create metadata tables in the Data Catalog2. The Data Catalog is a persistent metadata store that contains table definitions, job definitions, and other control information to help you manage your AWS Glue components3. You can use AWS Glue to crawl the data sources, such as Amazon S3, Amazon RDS for Microsoft SQL Server, and Amazon DynamoDB, and store the metadata in the Data Catalog.
Amazon Athena is a serverless, interactive query service that makes it easy to analyze data directly in Amazon S3 using standard SQL or Python4. Amazon Athena also supports PartiQL, a SQL-compatible query language that lets you query, insert, update, and

delete data from semi-structured and nested data, such as JSON. You can use Amazon Athena to query the data from the Data Catalog using SQL for structured data sources, such as .csv files and relational databases, and PartiQL for data that is stored in JSON format. You can also use Athena to query data from other data sources, such as Amazon Redshift, using federated queries. Using AWS Glue and Amazon Athena to query all data sources by using syntax similar to SQL is the least operational overhead solution, as you do not need to provision, manage, or scale any infrastructure, and you pay only for the resources you use. AWS Glue charges you based on the compute time and the data processed by your crawlers and ETL jobs1. Amazon Athena charges you based on the amount of data scanned by your queries. You can also reduce the cost and improve the performance of your queries by using compression, partitioning, and columnar formats for your data in Amazon S3.
Option B is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, and use Redshift Spectrum to query the data, would incur more costs and complexity than using Amazon Athena. Redshift Spectrum is a feature of Amazon Redshift, a fully managed data warehouse service, that allows you to query and join data across your data warehouse and your data lake using standard SQL. While Redshift Spectrum is powerful and useful for many data warehousing scenarios, it is not necessary or cost-effective for querying all data sources by using syntax similar to SQL. Redshift Spectrum charges you based on the amount of data scanned by your queries, which is similar to Amazon Athena, but it also requires you to have an Amazon Redshift cluster, which charges you based on the node type, the number of nodes, and the duration of the cluster5. These costs can add up quickly, especially if you have large volumes of data and complex queries. Moreover, using Redshift Spectrum would introduce additional latency and complexity, as you would have to provision and manage the cluster, and create an external schema and database for the data in the Data Catalog, instead of querying it directly from Amazon Athena.
Option C is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use AWS Glue jobs to transform data that is in JSON format to Apache Parquet or .csv format, store the transformed data in an S3 bucket, and use Amazon Athena to query the original and transformed data from the S3 bucket, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Glue jobs are ETL scripts that you can write in Python or Scala to transform your data and load it to your target data store. Apache Parquet is a columnar storage format that can improve the performance of analytical queries by reducing the amount of data that needs to be scanned and providing efficient compression and encoding schemes6. While using AWS Glue jobs and Parquet can improve the performance and reduce the cost of your queries, they would also increase the complexity and the operational overhead of the data pipeline, as you would have to write, run, and monitor the ETL jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using AWS Glue jobs and Parquet would introduce additional latency, as you would have to wait for the ETL jobs to finish before querying the transformed data.
Option D is not the best solution, as using AWS Lake Formation to create a data lake, use Lake Formation jobs to transform the data from all data sources to Apache Parquet format, store the transformed data in an S3 bucket, and use Amazon Athena or RedshiftSpectrum to query the data, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Lake Formation is a service that helps you centrally govern, secure, and globally share data for analytics and machine learning7. Lake Formation jobs are ETL jobs that you can create and run using the Lake Formation console or API. While using Lake Formation and Parquet can improve the performance and reduce the cost of your queries, they would also increase the complexity and the operational overhead of the data pipeline, as you would have to create, run, and monitor the Lake Formation jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using Lake Formation and Parquet would introduce additional latency, as you would have to wait for the Lake Formation jobs to finish before querying the transformed data. Furthermore, using Redshift Spectrum to query the data would also incur the same costs and complexity as mentioned in option B. References:
What is Amazon Athena?
Data Catalog and crawlers in AWS Glue
AWS Glue Data Catalog
Columnar Storage Formats
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide
AWS Glue Schema Registry
What is AWS Glue?
Amazon Redshift Serverless
Amazon Redshift provisioned clusters
[Querying external data using Amazon Redshift Spectrum]
[Using stored procedures in Amazon Redshift]
[What is AWS Lambda?]
[PartiQL for Amazon Athena]
[Federated queries in Amazon Athena]
[Amazon Athena pricing]
[Top 10 performance tuning tips for Amazon Athena]
[AWS Glue ETL jobs]
[AWS Lake Formation jobs]

NEW QUESTION # 57

A company analyzes data in a data lake every quarter to perform inventory assessments. A data engineer uses AWS Glue DataBrew to detect any personally identifiable information (PII) about customers within the dat a. The company's privacy policy considers some custom categories of information to be PII. However, the categories are not included in standard DataBrew data quality rules.
The data engineer needs to modify the current process to scan for the custom PII categories across multiple datasets within the data lake.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Implement regex patterns to extract PII information from fields during extract transform, and load (ETL) operations into the data lake.
- B. Manually review the data for custom PII categories.
- C. Implement custom data quality rules in Data Brew. Apply the custom rules across datasets.
- D. Develop custom Python scripts to detect the custom PII categories. Call the scripts from DataBrew.

Answer: C

Explanation:
The data engineer needs to detect custom categories of PII within the data lake using AWS Glue DataBrew. While DataBrew provides standard data quality rules, the solution must support custom PII categories.
Option B: Implement custom data quality rules in DataBrew. Apply the custom rules across datasets.
This option is the most efficient because DataBrew allows the creation of custom data quality rules that can be applied to detect specific data patterns, including custom PII categories. This approach minimizes operational overhead while ensuring that the specific privacy requirements are met.
Options A, C, and D either involve manual intervention or developing custom scripts, both of which increase operational effort compared to using DataBrew's built-in capabilities.
Reference:
AWS Glue DataBrew Documentation

NEW QUESTION # 58
A retail company is expanding its operations globally. The company needs to use Amazon QuickSight to accurately calculate currency exchange rates for financial reports. The company has an existing dashboard that includes a visual that is based on an analysis of a dataset that contains global currency values and exchange rates.
A data engineer needs to ensure that exchange rates are calculated with a precision of four decimal places. The calculations must be precomputed. The data engineer must materialize results in QuickSight super-fast, parallel, in-memory calculation engine (SPICE).
Which solution will meet these requirements?

- A. Define and create the calculated field in the dataset.
- B. Define and create the calculated field in the dashboard.
- C. Define and create the calculated field in the analysis.
- D. Define and create the calculated field in the visual.

Answer: A

NEW QUESTION # 59
A company is migrating its database servers from Amazon EC2 instances that run Microsoft SQL Server to Amazon RDS for Microsoft SQL Server DB instances. The company's analytics team must export large data elements every day until the migration is complete. The data elements are the result of SQL joinsacross multiple tables. The data must be in Apache Parquet format. The analytics team must store the data in Amazon S3.
Which solution will meet these requirements in the MOST operationally efficient way?

- A. Use a SQL query to create a view in the EC2 instance-based SQL Server databases that contains the required data elements. Create and run an AWS Glue crawler to read the view. Create an AWS Glue job that retrieves the data and transfers the data in Parquet format to an S3 bucket. Schedule the AWS Glue job to run every day.
- B. Schedule SQL Server Agent to run a daily SQL query that selects the desired data elements from the EC2 instance-based SQL Server databases. Configure the query to direct the output .csv objects to an S3 bucket. Create an S3 event that invokes an AWS Lambda function to transform the output format from
.csv to Parquet.
- C. Create an AWS Lambda function that queries the EC2 instance-based databases by using Java Database Connectivity (JDBC). Configure the Lambda function to retrieve the required data, transform the data into Parquet format, and transfer the

data into an S3 bucket. Use Amazon EventBridge to schedule the Lambda function to run every day.

- D. Create a view in the EC2 instance-based SQL Server databases that contains the required data elements.
  Create an AWS Glue job that selects the data directly from the view and transfers the data in Parquet format to an S3 bucket.
  Schedule the AWS Glue job to run every day.

Answer: D

Explanation:
Option A is the most operationally efficient way to meet the requirements because it minimizes the number of steps and services involved in the data export process. AWS Glue is a fully managed service that can extract, transform, and load (ETL) data from various sources to various destinations, including Amazon S3. AWS Glue can also convert data to different formats, such as Parquet, which is a columnar storage format that is optimized for analytics. By creating a view in the SQL Server databases that contains the required data elements, the AWS Glue job can select the data directly from the view without having to perform any joins or transformations on the source data. The AWS Glue job can then transfer the data in Parquet format to an S3 bucket and run on a daily schedule.
Option B is not operationally efficient because it involves multiple steps and services to export the data. SQL Server Agent is a tool that can run scheduled tasks on SQL Server databases, such as executing SQL queries.
However, SQL Server Agent cannot directly export data to S3, so the query output must be saved as .csv objects on the EC2 instance. Then, an S3 event must be configured to trigger an AWS Lambda function that can transform the .csv objects to Parquet format and upload them to S3. This option adds complexity and latency to the data export process and requires additional resources and configuration.
Option C is not operationally efficient because it introduces an unnecessary step of running an AWS Glue crawler to read the view. An AWS Glue crawler is a service that can scan data sources and create metadata tables in the AWS Glue Data Catalog. The Data Catalog is a central repository that stores information about the data sources, such as schema, format, and location. However, in this scenario, the schema and format of the data elements are already known and fixed, so there is no need to run a crawler to discover them. The AWS Glue job can directly select the data from the view without using the Data Catalog. Running a crawler adds extra time and cost to the data export process.
Option D is not operationally efficient because it requires custom code and configuration to query the databases and transform the data. An AWS Lambda function is a service that can run code in response to events or triggers, such as Amazon EventBridge. Amazon EventBridge is a service that can connect applications and services with event sources, such as schedules, and route them to targets, such as Lambda functions. However, in this scenario, using a Lambda function to query the databases and transform the data is not the best option because it requires writing and maintaining code that uses JDBC to connect to the SQL Server databases, retrieve the required data, convert the data to Parquet format, and transfer the data to S3.
This option also has limitations on the execution time, memory, and concurrency of the Lambda function, which may affect the performance and reliability of the data export process.
References:
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide
AWS Glue Documentation
Working with Views in AWS Glue
Converting to Columnar Formats

NEW QUESTION # 60
.....

If candidates are going to buy Data-Engineer-Associate test dumps, they may consider the problem of the fund safety. If you are thinking the same question like this, our company will eradicate your worries. We choose the international third party to ensure the safety of the fund. The Data-Engineer-Associate Test Dumps are effective and conclusive, you just need to use the least time to pass it. I f you choose us, it means you choose the pass.

Data-Engineer-Associate Pdf Dumps: https://www.dumpstillvalid.com/Data-Engineer-Associate-prep4sure-review.html

You can check the quality and features of Data-Engineer-Associate AWS Certified Data Engineer - Associate (DEA-C01) exam dumps, Amazon Free Data-Engineer-Associate Test Questions Prepare According to Latest Updates, Our Amazon Data-Engineer-Associate Pdf Dumps practice tests offer customizable learning and self-assessment features that truly benefits thousands of takers, Amazon Free Data-Engineer-Associate Test Questions Modren and User Friendly Interface, If you have any questions after you buy our Data-Engineer-Associate study guide, you can always get thoughtful support and help by email or online inquiry.

Linux meets those standards today, For now, just remember that Data-Engineer-Associate operators exist for every basic mathematical function, and you can refer back to these tables when you find the need.

You can check the quality and features of Data-Engineer-Associate AWS Certified Data Engineer - Associate (DEA-C01) exam

dumps, Prepare According to Latest Updates, Our Amazon practice tests offer customizable learning and self-assessment features that truly benefits thousands of takers.

Test-Taking Questions Data-Engineer-Associate Pre-assessment Test

Modren and User Friendly Interface, If you have any questions after you buy our Data-Engineer-Associate study guide, you can always get thoughtful support and help by email or online inquiry.

- Data-Engineer-Associate Latest Demo □ Braindumps Data-Engineer-Associate Pdf □ Interactive Data-Engineer-Associate Practice Exam ↕☀ www.prep4away.com □☀□ is best website to obtain ☀ Data-Engineer-Associate □☀□ for free download □Exam Data-Engineer-Associate Passing Score
- Professional Free Data-Engineer-Associate Test Questions to pass AWS Certified Data Engineer - Associate (DEA-C01) - Recommend by Experts □ Open website ☀ www.pdfvce.com □☀□ and search for ➥ Data-Engineer-Associate □ for free download □Braindumps Data-Engineer-Associate Pdf
- Professional Free Data-Engineer-Associate Test Questions to pass AWS Certified Data Engineer - Associate (DEA-C01) - Recommend by Experts □ Search for ➽ Data-Engineer-Associate □ and download it for free on 「 www.real4dumps.com 」 website □Data-Engineer-Associate Question Explanations
- Free PDF Quiz High Hit-Rate Data-Engineer-Associate - Free AWS Certified Data Engineer - Associate (DEA-C01) Test Questions □ Search for ➽ Data-Engineer-Associate □ and easily obtain a free download on □ www.pdfvce.com □ □ □Data-Engineer-Associate Latest Demo
- Pass Guaranteed Quiz Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) High Hit-Rate Free Test Questions □ Open website ➽ www.testkingpdf.com □ and search for 【 Data-Engineer-Associate 】 for free download □Data-Engineer-Associate Latest Braindumps Ppt
- Data-Engineer-Associate Sure-Pass Guide Torrent Dumps File is the best preparation materials - Pdfvce □ Open 【 www.pdfvce.com 】 and search for □ Data-Engineer-Associate □ to download exam materials for free □Data-Engineer-Associate Reliable Practice Questions
- Pass Guaranteed Quiz Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) High Hit-Rate Free Test Questions □ Open □ www.pass4leader.com □ enter " Data-Engineer-Associate " and obtain a free download □ □Data-Engineer-Associate Exam Pattern
- Data-Engineer-Associate Answers Free □ Braindumps Data-Engineer-Associate Pdf □ Data-Engineer-Associate Latest Exam Review □ Search for ☀ Data-Engineer-Associate □☀□ and obtain a free download on 「 www.pdfvce.com 」 □ □Data-Engineer-Associate Exam Pattern
- Free PDF Quiz High Hit-Rate Data-Engineer-Associate - Free AWS Certified Data Engineer - Associate (DEA-C01) Test Questions □ Open ➡ www.testsimulate.com □□□ and search for ➠ Data-Engineer-Associate □ to download exam materials for free □Data-Engineer-Associate Exam Papers
- Professional Free Data-Engineer-Associate Test Questions to pass AWS Certified Data Engineer - Associate (DEA-C01) - Recommend by Experts □ Copy URL ➠ www.pdfvce.com □ open and search for □ Data-Engineer-Associate □ to download for free □Key Data-Engineer-Associate Concepts
- Data-Engineer-Associate Sure-Pass Guide Torrent Dumps File is the best preparation materials - www.real4dumps.com □ Easily obtain □ Data-Engineer-Associate □ for free download through ➤ www.real4dumps.com □ □Key Data-Engineer-Associate Concepts
- 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, www.pcsq28.com, lms.ait.edu.za, study.stcs.edu.np, jimbell680.blogocial.com, witpacourses.com, window.noedge.ca, 203060.com, pct.edu.pk, iknolez.co.in, Disposable vapes

DOWNLOAD the newest DumpStillValid Data-Engineer-Associate PDF dumps from Cloud Storage for free: https://drive.google.com/open?id=12gCRYxuRB7h7aNqcdrIpvF128aTUZflC