Data-Engineer-Associate Complete Exam Dumps - Valid Study Data-Engineer-Associate Questions

What's more, part of that Itcertmaster Data-Engineer-Associate dumps now are free: https://drive.google.com/open?id=1x9RX_5KeZeKeyHcxNpth26u-bWdTKdNG

our Data-Engineer-Associate actual exam has won thousands of people's support. All of them have passed the exam and got the certificate. They live a better life now. Our Data-Engineer-Associate study guide can release your stress of preparation for the test. Our Data-Engineer-Associate Exam Engine is professional, which can help you pass the exam for the first time. If you can't wait getting the certificate, you are supposed to choose our Data-Engineer-Associate study guide.

How can you quickly change your present situation and be competent for the new life, for jobs, in particular? The answer is using our Data-Engineer-Associate practice materials. From my perspective, our free demo of Data-Engineer-Associate exam questions is possessed with high quality which is second to none. This is no exaggeration at all. Just as what have been reflected in the statistics, the pass rate for those who have chosen our Data-Engineer-Associate Exam Guide is as high as 99%, which in turn serves as the proof for the high quality of our Data-Engineer-Associate practice torrent.

>> Data-Engineer-Associate Complete Exam Dumps <<

Data-Engineer-Associate Complete Exam Dumps - 100% Useful Questions Pool

Prepared by experts and approved by experienced professionals, our Data-Engineer-Associate exam torrent is well-designed high quality products and they are revised and updated based on changes in syllabus and the latest developments in theory and practice. With the guidance of our Data-Engineer-Associate Guide Torrent, you can make progress by a variety of self-learning and self-assessing features to test learning outcomes. And as the high pass rate of our Data-Engineer-Associate exam questions is 99% to 100%, you will be bound to pass the Data-Engineer-Associate exam with ease.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q21-Q26):

NEW QUESTION # 21

A company needs to set up a data catalog and metadata management for data sources that run in the AWS Cloud. The company will use the data catalog to maintain the metadata of all the objects that are in a set of data stores. The data stores include structured sources such as Amazon RDS and Amazon Redshift. The data stores also include semistructured sources such as JSON files and .xml files that are stored in Amazon S3.

The company needs a solution that will update the data catalog on a regular basis. The solution also must detect changes to the source metadata.

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

- A. Use the AWS Glue Data Catalog as the central metadata repository. Use AWS Glue crawlers to connect to multiple data stores and to update the Data Catalog with metadata changes. Schedule the crawlers to run periodically to update the metadata catalog.
- B. Use the AWS Glue Data Catalog as the central metadata repository. Extract the schema for Amazon RDS and Amazon Redshift sources, and build the Data Catalog. Use AWS Glue crawlers for data that is in Amazon S3 to infer the schema and to automatically update the Data Catalog.
- C. Use Amazon Aurora as the data catalog. Create AWS Lambda functions that will connect to the data catalog. Configure the Lambda functions to gather the metadata information from multiple sources and to update the Aurora data catalog. Schedule the Lambda functions to run periodically.
- D. Use Amazon DynamoDB as the data catalog. Create AWS Lambda functions that will connect to the data catalog. Configure the Lambda functions to gather the metadata information from multiple sources and to update the DynamoDB data catalog. Schedule the Lambda functions to run periodically.

Answer: A

Explanation:

This solution will meet the requirements with the least operational overhead because it uses the AWS Glue Data Catalog as the central metadata repository for data sources that run in the AWS Cloud. The AWS Glue Data Catalog is a fully managed service that provides a unified view of your data assets across AWS and on- premises data sources. It stores the metadata of your data in tables, partitions, and columns, and enables you to access and query your data using various AWS services, such as Amazon Athena, Amazon EMR, and Amazon Redshift Spectrum. You can use AWS Glue crawlers to connect to multiple data stores, such as Amazon RDS, Amazon Redshift, and Amazon S3, and to update the Data Catalog with metadata changes.

AWS Glue crawlers can automatically discover the schema and partition structure of your data, and create or update the corresponding tables in the Data Catalog. You can schedule the crawlers to run periodically to update the metadata catalog, and configure them to detect changes to the source metadata, such as new columns, tables, or partitions12.

The other options are not optimal for the following reasons:

* A. Use Amazon Aurora as the data catalog. Create AWS Lambda functions that will connect to the data catalog. Configure the Lambda functions to gather the metadata information from multiple sources and to update the Aurora data catalog. Schedule the Lambda functions to run periodically. This option is not recommended, as it would require more operational overhead to create and manage an Amazon Aurora database as the data catalog, and to write and maintain AWS Lambda functions to gather and update the metadata information from multiple sources. Moreover, this option would not leverage the benefits of the AWS Glue Data Catalog, such as data cataloging, data transformation, and data governance.

* C. Use Amazon DynamoDB as the data catalog. Create AWS Lambda functions that will connect to the data catalog. Configure the Lambda functions to gather the metadata information from multiple sources and to update the DynamoDB data catalog. Schedule the Lambda functions to run periodically. This option is also not recommended, as it would require more operational overhead to create and manage an Amazon DynamoDB table as the data catalog, and to write and maintain AWS Lambda functions to gather and update the metadata information from multiple sources. Moreover, this option would not leverage the benefits of the AWS Glue Data Catalog, such as data cataloging, data transformation, and data governance.

* D. Use the AWS Glue Data Catalog as the central metadata repository. Extract the schema for Amazon RDS and Amazon Redshift sources, and build the Data Catalog. Use AWS Glue crawlers for data that is in Amazon S3 to infer the schema and to automatically update the Data Catalog. This option is not optimal, as it would require more manual effort to extract the schema for Amazon RDS and Amazon Redshift sources, and to build the Data Catalog. This option would not take advantage of the AWS Glue crawlers' ability to automatically discover the schema and partition structure of your data from various data sources, and to create or update the corresponding tables in the Data Catalog.

References:

* 1: AWS Glue Data Catalog

* 2: AWS Glue Crawlers

* : Amazon Aurora

* : AWS Lambda

* : Amazon DynamoDB

NEW QUESTION # 22

A data engineer is using Amazon Athena to analyze sales data that is in Amazon S3. The data engineer writes a query to retrieve sales amounts for 2023 for several products from a table named sales_data. However, the query does not return results for all of the products that are in the sales_data table. The data engineer needs to troubleshoot the query to resolve the issue.

The data engineer's original query is as follows:

```
SELECT product_name, sum(sales_amount)
FROM sales_data
WHERE year = 2023
GROUP BY product_name
```

How should the data engineer modify the Athena query to meet these requirements?

- A. Remove the GROUP BY clause
- B. Change WHERE year = 2023 to WHERE extractlyear FROM sales data) = 2023.
- C. Add HAVING sumfsales amount) > 0 after the GROUP BY clause.
- D. Replace sum(sales amount) with count(*J for the aggregation.

Answer: B

Explanation:

The original query does not return results for all of the products because the year column in the sales_data table is not an integer, but a timestamp. Therefore, the WHERE clause does not filter the data correctly, and only returns the products that have a null value for the year column. To fix this, the data engineer should use the extract function to extract the year from the timestamp and compare it with 2023. This way, the query will return the correct results for all of the products in the sales_data table. The other options are either incorrect or irrelevant, as they do not address the root cause of the issue. Replacing sum with count does not change the filtering condition, adding HAVING clause does not affect the grouping logic, and removing the GROUP BY clause does not solve the problem of missing products. References:

* Troubleshooting JSON queries - Amazon Athena (Section: JSON related errors)

* When I query a table in Amazon Athena, the TIMESTAMP result is empty (Section: Resolution)

* AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide (Chapter 7, page 197)

NEW QUESTION # 23

A company has a data processing pipeline that includes several dozen steps. The data processing pipeline needs to send alerts in real time when a step fails or succeeds. The data processing pipeline uses a combination of Amazon S3 buckets, AWS Lambda functions, and AWS Step Functions state machines.

A data engineer needs to create a solution to monitor the entire pipeline.

Which solution will meet these requirements?

- A. Configure an Amazon EventBridge rule to react when the execution status of a state machine changes.Configure the rule to send a message to an Amazon Simple Notification Service (Amazon SNS) topic that sends notifications.
- B. Configure the Step Functions state machines to store notifications in an Amazon S3 bucket when the state machines finish running. Enable S3 event notifications on the S3 bucket.
- C. Configure the AWS Lambda functions to store notifications in an Amazon S3 bucket when the state machines finish running. Enable S3 event notifications on the S3 bucket.
- D. Use AWS CloudTrail to send a message to an Amazon Simple Notification Service (Amazon SNS) topic that sends notifications when a state machine fails to run or succeeds to run.

Answer: A

Explanation:

AWS Step Functions natively emitsstate change eventstoAmazon EventBridge, which can trigger an Amazon SNS notification. This is the most direct and real-time way to alert on success/failure without relying on custom logging or polling.

"Step Functions automatically emits status changes that EventBridge can capture to trigger alerts or workflows. Use EventBridge to invoke an SNS topic for real-time alerts on job status."

-Ace the AWS Certified Data Engineer - Associate Certification - version 2 - apple.pdf This provides real-time alerting and the least operational overhead.

NEW QUESTION # 24

A data engineer runs Amazon Athena queries on data that is in an Amazon S3 bucket. The Athena queries use AWS Glue Data Catalog as a metadata table.

The data engineer notices that the Athena query plans are experiencing a performance bottleneck. The data engineer determines that the cause of the performance bottleneck is the large number of partitions that are in the S3 bucket. The data engineer must resolve the performance bottleneck and reduce Athena query planning time.
Which solutions will meet these requirements? (Choose two.)

- A. Transform the data that is in the S3 bucket to Apache Parquet format.
- B. Create an AWS Glue partition index. Enable partition filtering.
- C. Bucketthe data based on a column thatthe data have in common in a WHERE clause of the user query
- D. Use the Amazon EMR S3DistCP utility to combine smaller objects in the S3 bucket into larger objects.
- E. Use Athena partition projection based on the S3 bucket prefix.

Answer: B,E

Explanation:
The best solutions to resolve the performance bottleneck and reduce Athena query planning time are to create an AWS Glue partition index and enable partition filtering, and to use Athena partition projection based on the S3 bucket prefix.
AWS Glue partition indexes are a feature that allows you to speed up query processing of highly partitioned tables cataloged in AWS Glue Data Catalog. Partition indexes are available for queries in Amazon EMR, Amazon Redshift Spectrum, and AWS Glue ETL jobs. Partition indexes are sublists of partition keys defined in the table. When you create a partition index, you specify a list of partition keys that already exist on a given table. AWS Glue then creates an index for the specified keys and stores it in the Data Catalog. When you run a query that filters on the partition keys, AWS Glue uses the partition index to quickly identify the relevant partitions without scanning the entiretable metadata. This reduces the query planning time and improves the query performance1.
Athena partition projection is a feature that allows you to speed up query processing of highly partitioned tables and automate partition management. In partition projection, Athena calculates partition values and locations using the table properties that you configure directly on your table in AWS Glue. The table properties allow Athena to 'project', or determine, the necessary partition information instead of having to do a more time-consuming metadata lookup in the AWS Glue Data Catalog. Because in-memory operations are often faster than remote operations, partition projection can reduce the runtime of queries against highly partitioned tables. Partition projection also automates partition management because it removes the need to manually create partitions in Athena, AWS Glue, or your external Hive metastore2.
Option B is not the best solution, as bucketing the data based on a column that the data have in common in a WHERE clause of the user query would not reduce the query planning time. Bucketing is a technique that divides data into buckets based on a hash function applied to a column. Bucketing can improve the performance of join queries by reducing the amount of data that needs to be shuffled between nodes. However, bucketing does not affect the partition metadata retrieval, which is the main cause of the performance bottleneck in this scenario3.
Option D is not the best solution, as transforming the data that is in the S3 bucket to Apache Parquet format would not reduce the query planning time. 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 schemes. However, Parquet does not affect the partition metadata retrieval, which is the main cause of the performance bottleneck in this scenario4.
Option E is not the best solution, as using the Amazon EMR S3DistCP utility to combine smaller objects in the S3 bucket into larger objects would not reduce the query planning time. S3DistCP is a tool that can copy large amounts of data between Amazon S3 buckets or from HDFS to Amazon S3. S3DistCP can also aggregate smaller files into larger files to improve the performance of sequential access. However, S3DistCP does not affect the partition metadata retrieval, which is the main cause of the performance bottleneck in this scenario5. References:
Improve query performance using AWS Glue partition indexes
Partition projection with Amazon Athena
Bucketing vs Partitioning
Columnar Storage Formats
S3DistCp
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 25
A company needs to partition the Amazon S3 storage that the company uses for a data lake. The partitioning will use a path of the S3 object keys in the following format: s3://bucket/prefix/year=2023/month=01/day=01.
A data engineer must ensure that the AWS Glue Data Catalog synchronizes with the S3 storage when the company adds new partitions to the bucket.
Which solution will meet these requirements with the LEAST latency?

- A. Use code that writes data to Amazon S3 to invoke the Boto3 AWS Glue create partition API call.
- B. Run the MSCK REPAIR TABLE command from the AWS Glue console.
- C. Manually run the AWS Glue CreatePartition API twice each day.

- D. Schedule an AWS Glue crawler to run every morning.

Answer: D

Explanation:
The best solution to ensure that the AWS Glue Data Catalog synchronizes with the S3 storage when the company adds new partitions to the bucket with the least latency is to use code that writes data to Amazon S3 to invoke the Boto3 AWS Glue create partition API call. This way, the Data Catalog is updated as soon as new data is written to S3, and the partition information is immediately available for querying by other services. The Boto3 AWS Glue create partition API call allows you to create a new partition in the Data Catalog by specifying the table name, the database name, and the partition values1. You can use this API call in your code that writes data to S3, such as a Python script or an AWS Glue ETL job, to create a partition for each new S3 object key that matches the partitioning scheme.
Option A is not the best solution, as scheduling an AWS Glue crawler to run every morning would introduce a significant latency between the time new data is written to S3 and the time the Data Catalog is updated. 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. Crawlers can be scheduled to run periodically, such as daily or hourly, but they cannot run continuously or in real-time. Therefore, using a crawler to synchronize the Data Catalog with the S3 storage would not meet the requirement of the least latency.
Option B is not the best solution, as manually running the AWS Glue CreatePartition API twice each day would also introduce a significant latency between the time new data is written to S3 and the time the Data Catalog is updated. Moreover, manually running the API would require more operational overhead and human intervention than using code that writes data to S3 to invoke the API automatically.
Option D is not the best solution, as running the MSCK REPAIR TABLE command from the AWS Glue console would also introduce a significant latency between the time new data is written to S3 and the time the Data Catalog is updated. The MSCK REPAIR TABLE command is a SQL command that you can run in the AWS Glue console to add partitions to the Data Catalog based on the S3 object keys that match the partitioning scheme3. However, this command is not meant to be run frequently or in real-time, as it can take a long time to scan the entire S3 bucket and add the partitions. Therefore, using this command to synchronize the Data Catalog with the S3 storage would not meet the requirement of the least latency. Reference:
AWS Glue CreatePartition API
Populating the AWS Glue Data Catalog
MSCK REPAIR TABLE Command
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 26
......

The Amazon Data-Engineer-Associate certification exam helps you in getting jobs easily. Itcertmaster offers real Data-Engineer-Associate exam questions so that the students can prepare in a short time and crack the Data-Engineer-Associate exam with ease. These Data-Engineer-Associate Exam Questions are collected by professionals by working hard for days and nights so that the customers can pass Data-Engineer-Associate certification exam with good scores.

Valid Study Data-Engineer-Associate Questions: https://www.itcertmaster.com/Data-Engineer-Associate.html

Amazon Data-Engineer-Associate Complete Exam Dumps We pay deep attention to relevancy because out of context course content puts a lot of pressure on the learners, In order to let you obtain the latest information for the exam, we offer you free update for 365 days after buying Data-Engineer-Associate exam materials, and the update version will be sent to your email automatically, The information we have could give you the opportunity to practice issues, and ultimately achieve your goal that through Amazon Data-Engineer-Associate Exam Content exam certification.

Using the default Pastel Medium Tip preset, Data-Engineer-Associate Latest Test Materials block in large areas of value, starting with the midtones, A full-term male hashypospadias, We pay deep attention to relevancy Data-Engineer-Associate because out of context course content puts a lot of pressure on the learners.

Marvelous Data-Engineer-Associate Complete Exam Dumps | Amazing Pass Rate For Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) | Fantastic Valid Study Data-Engineer-Associate Questions

In order to let you obtain the latest information for the exam, we offer you free update for 365 days after buying Data-Engineer-Associate exam materials, and the update version will be sent to your email automatically.

The information we have could give you the opportunity to practice issues, and ultimately achieve your goal that through Amazon Data-Engineer-Associate Exam Content exam certification.

Questions & Answers are compiled by a group of Senior IT Professionals, Data-Engineer-Associate exam questions have a very high hit rate, of course, will have a very high pass rate.

- Exam Data-Engineer-Associate Book □ Valid Data-Engineer-Associate Test Objectives □ Data-Engineer-Associate Latest Test Cost □ Search for 「Data-Engineer-Associate」 and download exam materials for free through （www.pass4test.com） □Exam Data-Engineer-Associate Book
- Perfect Data-Engineer-Associate Prep Guide will be Changed According to The New Policy Every Year - Pdfvce □ Download ☀ Data-Engineer-Associate □☀□ for free by simply entering □ www.pdfvce.com □ website □Valid Data-Engineer-Associate Test Objectives
- Valid Braindumps Data-Engineer-Associate Ebook □ Data-Engineer-Associate Latest Test Cost □ Valid Data-Engineer-Associate Exam Topics □ Immediately open ▶ www.testkingpass.com ◀ and search for ▶ Data-Engineer-Associate ◀ to obtain a free download □New Exam Data-Engineer-Associate Materials
- Data-Engineer-Associate Latest Test Cost □ Valid Data-Engineer-Associate Test Objectives □ Dumps Data-Engineer-Associate Download □ Download ▶ Data-Engineer-Associate ◀ for free by simply entering □ www.pdfvce.com □ website □Data-Engineer-Associate Simulation Questions
- Valid Data-Engineer-Associate Test Objectives □ Data-Engineer-Associate Latest Test Cost □ Data-Engineer-Associate Verified Answers □ Copy URL □ www.exam4labs.com □ open and search for ☀ Data-Engineer-Associate □☀□ to download for free □Data-Engineer-Associate Reliable Exam Testking
- New Data-Engineer-Associate Braindumps Files □ Data-Engineer-Associate Verified Answers ✴ Data-Engineer-Associate Latest Test Cost □ Search for ✔ Data-Engineer-Associate □✔□ and download exam materials for free through （www.pdfvce.com） □Dumps Data-Engineer-Associate Download
- Data-Engineer-Associate Reliable Exam Pdf □ Valid Data-Engineer-Associate Exam Topics □ Data-Engineer-Associate Reliable Exam Testking □ Download ✔ Data-Engineer-Associate □✔□ for free by simply searching on 「www.pdfdumps.com」 □Valid Data-Engineer-Associate Test Objectives
- Make {Useful Study Notes} With Amazon Data-Engineer-Associate PDF Questions □ Simply search for ➥ Data-Engineer-Associate □ for free download on ▶ www.pdfvce.com ◀ □New Exam Data-Engineer-Associate Materials
- Valid Braindumps Data-Engineer-Associate Ebook □ Data-Engineer-Associate Detail Explanation □ Valid Data-Engineer-Associate Guide Files □ 「www.practicevce.com」 is best website to obtain □ Data-Engineer-Associate □ for free download □Data-Engineer-Associate Verified Answers
- Valid Data-Engineer-Associate Guide Files □ Data-Engineer-Associate Latest Test Cost □ Data-Engineer-Associate Sample Questions □ Simply search for ▶ Data-Engineer-Associate ◀ for free download on ☀ www.pdfvce.com □☀□ □ □Data-Engineer-Associate Sample Questions
- Data-Engineer-Associate Detail Explanation □ Data-Engineer-Associate Sample Questions □ Dumps Data-Engineer-Associate Download □ Open website 「www.examcollectionpass.com」 and search for （Data-Engineer-Associate） for free download □Data-Engineer-Associate Verified Answers
- app.parler.com, myportal.utt.edu.tt, stackblitz.com, www.stes.tyc.edu.tw, ready4interview.shop, www.stes.tyc.edu.tw, www.athworthacademy.in, bbs.t-firefly.com, bbs.t-firefly.com, Disposable vapes

P.S. Free 2026 Amazon Data-Engineer-Associate dumps are available on Google Drive shared by Itcertmaster: https://drive.google.com/open?id=1x9RX_5KeZeKeyHcxNpth26u-bWdTKdNG