2025 Excellent Data-Engineer-Associate New Braindumps Files Help You Pass Data-Engineer-Associate Easily

DOWNLOAD the newest DumpTorrent Data-Engineer-Associate PDF dumps from Cloud Storage for free: https://drive.google.com/open?id=1y3R78A_lRNoXr5uVMmqyspDxJYFkURLN

DumpTorrent also offers a demo of the Amazon Data-Engineer-Associate exam product which is absolutely free. Up to 1 year of free AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) questions updates are also available if in any case the sections of the Amazon Data-Engineer-Associate actual test changes after your purchase. Lastly, we also offer a full refund guarantee according to terms and conditions if you do not get success in the Amazon AWS Certified Data Engineer - Associate (DEA-C01) Certification Exam after using our Data-Engineer-Associate product. These offers by DumpTorrent save your time and money. Buy AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) practice material today.

If you are determined to purchase our AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate valid exam collection materials for your companies, if you pursue long-term cooperation with site, we will have some relate policy. Firstly we provide one-year service warranty for every buyer who purchased Amazon Data-Engineer-Associate valid exam collection materials.

>> Data-Engineer-Associate New Braindumps Files <<

Pass Guaranteed Quiz Amazon - Data-Engineer-Associate - Fantastic AWS Certified Data Engineer - Associate (DEA-C01) New Braindumps Files

Though our Data-Engineer-Associate study guide has three formats which can meet your different needs, PDF version, software version and online version, i love the PDF version to the best. If you choose the PDF version, you can download our Data-Engineer-Associate exam material and print it for studying everywhere. And you can take notes on them as long as any new thoughts come to you. If a new version of the Data-Engineer-Associate learning guide comes out, we will send you a new link to your E-mail box and you can download it again.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q171-Q176):

NEW QUESTION # 171
A company has a data lake in Amazon S3. The company collects AWS CloudTrail logs for multiple applications. The company stores the logs in the data lake, catalogs the logs in AWS Glue, and partitions the logs based on the year. The company uses Amazon Athena to analyze the logs.
Recently, customers reported that a query on one of the Athena tables did not return any data. A data engineer must resolve the issue.
Which combination of troubleshooting steps should the data engineer take? (Select TWO.)

- A. Confirm that Athena is pointing to the correct Amazon S3 location.
- B. Delete and recreate the problematic Athena table.
- C. Use the MSCK REPAIR TABLE command.
- D. Restart Athena.
- E. Increase the query timeout duration.

Answer: A,C

Explanation:
The problem likely arises from Athena not being able to read from the correct S3 location or missing partitions. The two most relevant troubleshooting steps involve checking the S3 location and repairing the table metadata.
* A. Confirm that Athena is pointing to the correct Amazon S3 location:
* One of the most common issues with missing data in Athena queries is that the query is pointed to an incorrect or outdated S3 location. Checking the S3 path ensures Athena is querying the correct data.
Reference:Amazon Athena Troubleshooting
C: Use the MSCK REPAIR TABLE command:
When new partitions are added to the S3 bucket without being reflected in the Glue Data Catalog, Athena queries will not return data from those partitions. The MSCK REPAIR TABLE command updates the Glue Data Catalog with the latest partitions.
Reference:MSCK REPAIR TABLE Command
Alternatives Considered:
B (Increase query timeout): Timeout issues are unrelated to missing data.
D (Restart Athena): Athena does not require restarting.
E (Delete and recreate table): This introduces unnecessary overhead when the issue can be resolved by repairing the table and confirming the S3 location.
References:
Athena Query Fails to Return Data

NEW QUESTION # 172
A gaming company uses Amazon Kinesis Data Streams to collect clickstream data. The company uses Amazon Kinesis Data Firehose delivery streams to store the data in JSON format in Amazon S3. Data scientists at the company use Amazon Athena to query the most recent data to obtain business insights.
The company wants to reduce Athena costs but does not want to recreate the data pipeline.
Which solution will meet these requirements with the LEAST management effort?

- A. Change the Firehose output format to Apache Parquet. Provide a custom S3 object YYYYMMDD prefix expression and specify a large buffer size. For the existing data, create an AWS Glue extract, transform, and load (ETL) job. Configure the ETL job to combine small JSON files, convert the JSON files to large Parquet files, and add the YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.
- B. Create an Apache Spark job that combines JSON files and converts the JSON files to Apache Parquet files. Launch an Amazon EMR ephemeral cluster every day to run the Spark job to create new Parquet files in a different S3 location. Use the ALTER TABLE SET LOCATION statement to reflect the new S3 location on the existing Athena table.
- C. Integrate an AWS Lambda function with Firehose to convert source records to Apache Parquet and write them to Amazon S3. In parallel, run an AWS Glue extract, transform, and load (ETL) job to combine the JSON files and convert the JSON files to large Parquet files. Create a custom S3 object YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.
- D. Create a Kinesis data stream as a delivery destination for Firehose. Use Amazon Managed Service for Apache Flink (previously known as Amazon Kinesis Data Analytics) to run Apache Flink on the Kinesis data stream. Use Flink to aggregate the data and save the data to Amazon S3 in Apache Parquet format with a custom S3 object YYYYMMDD prefix. Use the ALTER TABLE ADD PARTITION statement to reflect the partition on the existing Athena table.

Answer: A

Explanation:
Step 1: Understanding the Problem
The company collectsclickstream datavia Amazon Kinesis Data Streams and stores it inJSON formatin Amazon S3 using Kinesis Data Firehose. They useAmazon Athenato query the data, but they want toreduce Athena costswhile maintaining the same data pipeline.
Since Athena charges based on the amount of data scanned during queries, reducing the data size (by converting JSON to a more efficient format likeApache Parquet) is a key solution to lowering costs.
Step 2: Why Option A is Correct
* Option Aprovides a straightforward way to reduce costs withminimal management overhead:
* Changing the Firehose output format to Parquet: Parquet is a columnar data format, which is more compact and efficient than JSON for Athena queries. It significantly reduces the amount of data scanned, which in turn reduces Athena query costs.
* Custom S3 Object Prefix (YYYYMMDD): Adding a date-based prefix helps in partitioning the data, which further improves query efficiency in Athena by limiting the data scanned to only relevant partitions.
* AWS Glue ETL Job for Existing Data: To handle existing data stored in JSON format, a one- time AWS Glue ETL job can combine small JSON files, convert them to Parquet, and apply the YYYYMMDD prefix. This ensures consistency in the S3 bucket structure and allows Athena to efficiently query historical data.
* ALTER TABLE ADD PARTITION: This command updates Athena's table metadata to reflect the new partitions, ensuring that future queries target only the required data.
Step 3: Why Other Options Are Not Ideal
* Option B (Apache Spark on EMR)introduces higher management effort by requiring the setup of Apache Spark jobsand anAmazon EMR cluster. While it achieves the goal of converting JSON to Parquet, it involves running and maintaining an EMR cluster, which adds operational complexity.
* Option C (Kinesis and Apache Flink)is a more complex solution involvingApache Flink, which adds a real-time streaming layer to aggregate data. Although Flink is a powerful tool for stream processing, it adds unnecessary overhead in this scenario since the company already uses Kinesis Data Firehose for batch delivery to S3.
* Option D (AWS Lambda with Firehose)suggests usingAWS Lambdato convert records in real time.
While Lambda can work in some cases, it's generally not the best tool for handling large-scale data transformations like JSON-to-Parquet conversion due to potential scaling and invocation limitations.
Additionally, running parallel Glue jobs further complicates the setup.
Step 4: How Option A Minimizes Costs
* By usingApache Parquet, Athena queries become more efficient, as Athena will scan significantly less data, directly reducing query costs.
* Firehosenatively supports Parquet as an output format, so enabling this conversion in Firehose requires minimal effort. Once set, new data will automatically be stored in Parquet format in S3, without requiring any custom coding or ongoing management.
* TheAWS Glue ETL jobfor historical data ensures that existing JSON files are also converted to Parquet format, ensuring consistency across the data stored in S3.
Conclusion:
Option A meets the requirement toreduce Athena costswithout recreating the data pipeline, using Firehose's native support forApache Parquetand a simple one-timeAWS Glue ETL jobfor existing data. This approach involvesminimal management effortcompared to the other solutions.

NEW QUESTION # 173
A company uses Amazon Athena to run SQL queries for extract, transform, and load (ETL) tasks by using Create Table As Select (CTAS). The company must use Apache Spark instead of SQL to generate analytics.
Which solution will give the company the ability to use Spark to access Athena?

- A. Athena workgroup
- B. Athena query settings
- C. Athena data source
- D. Athena query editor

Answer: C

Explanation:
Athena data source is a solution that allows you to use Spark to access Athena by using the Athena JDBC driver and the Spark SQL interface. You can use the Athena data source to create Spark DataFrames from Athena tables, run SQL queries on the DataFrames, and write the results back to Athena. The Athena data source supports various data formats, such as CSV, JSON, ORC, and Parquet, and also supports partitioned and bucketed tables. The Athena data source is a cost-effective and scalable way to use Spark to access Athena, as it does not require any additional infrastructure or services, and you only pay for the data scanned by Athena.

The other options are not solutions that give the company the ability to use Spark to access Athena. Option A, Athena query settings, is a feature that allows you to configure various parameters for your Athena queries, such as the output location, the encryption settings, the query timeout, and the workgroup. Option B, Athena workgroup, is a feature that allows you to isolate and manage your Athena queries and resources, such as the query history, the query notifications, the query concurrency, and the query cost. Option D, Athena query editor, is a feature that allows you to write and run SQL queries on Athena using the web console or the API. None of these options enable you to use Spark instead of SQL to generate analytics on Athena. Reference:
Using Apache Spark in Amazon Athena
Athena JDBC Driver
Spark SQL
Athena query settings
[Athena workgroups]
[Athena query editor]

NEW QUESTION # 174
A company maintains a data warehouse in an on-premises Oracle database. The company wants to build a data lake on AWS. The company wants to load data warehouse tables into Amazon S3 and synchronize the tables with incremental data that arrives from the data warehouse every day.
Each table has a column that contains monotonically increasing values. The size of each table is less than 50 GB. The data warehouse tables are refreshed every night between 1 AM and 2 AM. A business intelligence team queries the tables between 10 AM and 8 PM every day.
Which solution will meet these requirements in the MOST operationally efficient way?

- A. Use an AWS Glue Java Database Connectivity (JDBC) connection. Configure a job bookmark for a column that contains monotonically increasing values. Write custom logic to append the daily incremental data to a full-load copy that is in Amazon S3.
- B. Use an AWS Database Migration Service (AWS DMS) full load migration to load the data warehouse tables into Amazon S3 every day Overwrite the previous day's full-load copy every day.
- C. Use AWS Glue to load a full copy of the data warehouse tables into Amazon S3 every day. Overwrite the previous day's full-load copy every day.
- D. Use an AWS Database Migration Service (AWS DMS) full load plus CDC job to load tables that contain monotonically increasing data columns from the on-premises data warehouse to Amazon S3.
  Use custom logic in AWS Glue to append the daily incremental data to a full-load copy that is in Amazon S3.

Answer: D

Explanation:
The company needs to load data warehouse tables into Amazon S3 and perform incremental synchronization with daily updates. The most efficient solution is to use AWS Database Migration Service (AWS DMS) with a combination of full load and change data capture (CDC) to handle the initial load and daily incremental updates.
* Option A: Use an AWS Database Migration Service (AWS DMS) full load plus CDC job to load tables that contain monotonically increasing data columns from the on-premises data warehouse to Amazon S3. Use custom logic in AWS Glue to append the daily incremental data to a full-load copy that is in Amazon S3.DMS is designed to migrate databases to AWS, and the combination of full load plus CDC is ideal for handling incremental data changes efficiently. AWS Glue can then be used to append the incremental data to the full data set in S3. This solution is highly operationally efficient because it automates both the full load and incremental updates.
Options B, C, and D are less operationally efficient because they either require writing custom logic to handle bookmarks manually or involve unnecessary daily full loads.
References:
* AWS Database Migration Service Documentation
* AWS Glue Documentation

NEW QUESTION # 175
A data engineer must orchestrate a data pipeline that consists of one AWS Lambda function and one AWS Glue job. The solution must integrate with AWS services.
Which solution will meet these requirements with the LEAST management overhead?

- A. Use an Apache Airflow workflow that is deployed on Amazon Elastic Kubernetes Service (Amazon EKS). Define a directed acyclic graph (DAG) in which the first task is to call the Lambda function and the second task is to call the AWS

Glue job.

- B. Use an AWS Glue workflow to run the Lambda function and then the AWS Glue job.
- C. Use an Apache Airflow workflow that is deployed on an Amazon EC2 instance. Define a directed acyclic graph (DAG) in which the first task is to call the Lambda function and the second task is to call the AWS Glue job.
- D. Use an AWS Step Functions workflow that includes a state machine. Configure the state machine to run the Lambda function and then the AWS Glue job.

Answer: D

Explanation:
AWS Step Functions is a service that allows you to coordinate multiple AWS services into serverless workflows. You can use Step Functions to create state machines that define the sequence and logic of the tasks in your workflow. Step Functions supports various types of tasks, such as Lambda functions, AWS Glue jobs, Amazon EMR clusters, Amazon ECS tasks, etc. You can use Step Functions to monitor and troubleshoot your workflows, as well as to handle errors and retries.
Using an AWS Step Functions workflow that includes a state machine to run the Lambda function and then the AWS Glue job will meet the requirements with the least management overhead, as it leverages the serverless and managed capabilities of Step Functions. You do not need to write any code to orchestrate the tasks in your workflow, as you can use the Step Functions console or the AWS Serverless Application Model (AWS SAM) to define and deploy your state machine. You also do not need to provision or manage any servers or clusters, as Step Functions scales automatically based on the demand.
The other options are not as efficient as using an AWS Step Functions workflow. Using an Apache Airflow workflow that is deployed on an Amazon EC2 instance or on Amazon Elastic Kubernetes Service (Amazon EKS) will require more management overhead, as you will need to provision, configure, and maintain the EC2 instance or the EKS cluster, as well as the Airflow components. You will also need to write and maintain the Airflow DAGs to orchestrate the tasks in your workflow. Using an AWS Glue workflow to run the Lambda function and then the AWS Glue job will not work, as AWS Glue workflows only support AWS Glue jobs and crawlers as tasks, not Lambda functions. References:
AWS Step Functions
AWS Glue
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide, Chapter 6: Data Integration and Transformation, Section 6.3: AWS Step Functions

NEW QUESTION # 176
.....

Since it was founded, our DumpTorrent has more and more perfect system, more rich questiondumps, more payment security, and better customer service. Now the Data-Engineer-Associate exam dumps provided by DumpTorrent have been recognized by masses of customers, but we will not stop the service after you buy. We will inform you at the first time once the Data-Engineer-Associate Exam software updates, and if you can't fail the Data-Engineer-Associate exam we will full refund to you and we are responsible for your loss.

Data-Engineer-Associate New Dumps Sheet: https://www.dumptorrent.com/Data-Engineer-Associate-braindumps-torrent.html

Amazon Data-Engineer-Associate New Braindumps Files Q3: Can I pass the exam with Q&As only, Amazon Data-Engineer-Associate New Braindumps Files Golden service: 7/24 online service support, Amazon Data-Engineer-Associate New Braindumps Files It is unbelievable, right, We provide you with free update for one year and the update version for Data-Engineer-Associate exam torrent will be sent to your email automatically, Amazon Data-Engineer-Associate New Braindumps Files The exam covers all the areas described in the Exam Summary.

This sample chapter helps you learn the tools and methods necessary to Data-Engineer-Associate keep systems patched, how to troubleshoot the problems the patches may cause, and understand the reasons tools may not work correctly.

High Hit Rate Amazon Data-Engineer-Associate New Braindumps Files | Try Free Demo before Purchase

Weve written many times in the past on this topic, Q3: Latest Data-Engineer-Associate Exam Online Can I pass the exam with Q&As only, Golden service: 7/24 online service support, It is unbelievable, right?

We provide you with free update for one year and the update version for Data-Engineer-Associate Exam Torrent will be sent to your email automatically, The exam covers all the areas described in the Exam Summary.

- Lab Data-Engineer-Associate Questions Data-Engineer-Associate Valid Test Pattern Data-Engineer-Associate

Dump Torrent □ Search on ⇒ www.exams4collection.com ⇐ for ➽ Data-Engineer-Associate □ to obtain exam materials for free download □Related Data-Engineer-Associate Certifications

- Data-Engineer-Associate Reliable Braindumps Sheet □ Data-Engineer-Associate Reliable Test Preparation □ Data-Engineer-Associate Valid Test Pattern □ The page for free download of ⇒ Data-Engineer-Associate ⇐ on ✔ www.pdfvce.com □✔□ will open immediately □Data-Engineer-Associate Reliable Braindumps Sheet
- Latest Data-Engineer-Associate New Braindumps Files - 100% Pass Data-Engineer-Associate Exam □ Search for □ Data-Engineer-Associate □ and easily obtain a free download on 《 www.prep4pass.com 》 □Data-Engineer-Associate Reliable Exam Pattern
- Latest Data-Engineer-Associate New Braindumps Files - 100% Pass Data-Engineer-Associate Exam □ Copy URL 【 www.pdfvce.com 】 open and search for [Data-Engineer-Associate] to download for free □Related Data-Engineer-Associate Certifications
- Data-Engineer-Associate Real Test Practice Materials - Data-Engineer-Associate Test Prep - www.torrentvce.com □ Copy URL 「 www.torrentvce.com 」 open and search for ▶ Data-Engineer-Associate ◀ to download for free □Data-Engineer-Associate Reliable Braindumps Sheet
- Quiz 2025 Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Pass-Sure New Braindumps Files □ Search for 「 Data-Engineer-Associate 」 and download it for free immediately on □ www.pdfvce.com □ □Lab Data-Engineer-Associate Questions
- Data-Engineer-Associate Reliable Exam Pattern □ Lab Data-Engineer-Associate Questions □ Data-Engineer-Associate Reliable Braindumps Sheet □ Search for ➡ Data-Engineer-Associate □□□ and obtain a free download on ✔ www.lead1pass.com □✔□ □Data-Engineer-Associate Test Free
- Data-Engineer-Associate New Braindumps Files - Valid Amazon Data-Engineer-Associate New Dumps Sheet: AWS Certified Data Engineer - Associate (DEA-C01) □ Copy URL ☀ www.pdfvce.com □☀□ open and search for □ Data-Engineer-Associate □ to download for free □Dump Data-Engineer-Associate File
- Data-Engineer-Associate Tests Dumps, Data-Engineer-Associate Test Exam, Data-Engineer-Associate Valid Dumps □ Immediately open ⇒ www.dumpsquestion.com ⇐ and search for ➤ Data-Engineer-Associate □ to obtain a free download □ □Updated Data-Engineer-Associate Dumps
- Quiz 2025 Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Pass-Sure New Braindumps Files □ Search for □ Data-Engineer-Associate □ on { www.pdfvce.com } immediately to obtain a free download □Related Data-Engineer-Associate Certifications
- Data-Engineer-Associate Tests Dumps, Data-Engineer-Associate Test Exam, Data-Engineer-Associate Valid Dumps □ Search for 「 Data-Engineer-Associate 」 on ➠ www.prep4pass.com □ immediately to obtain a free download □Data-Engineer-Associate Reliable Exam Pattern
- www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, universityofapprointernational.com, blog.primeitservice.com, vxlxemito123.obsidianportal.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, myportal.utt.edu.tt, pct.edu.pk, www.stes.tyc.edu.tw, yy.hackp.com.cn, Disposable vapes

BTW, DOWNLOAD part of DumpTorrent Data-Engineer-Associate dumps from Cloud Storage: https://drive.google.com/open?id=1y3R78A_lRNoXr5uVMmqyspDxJYFkURLN