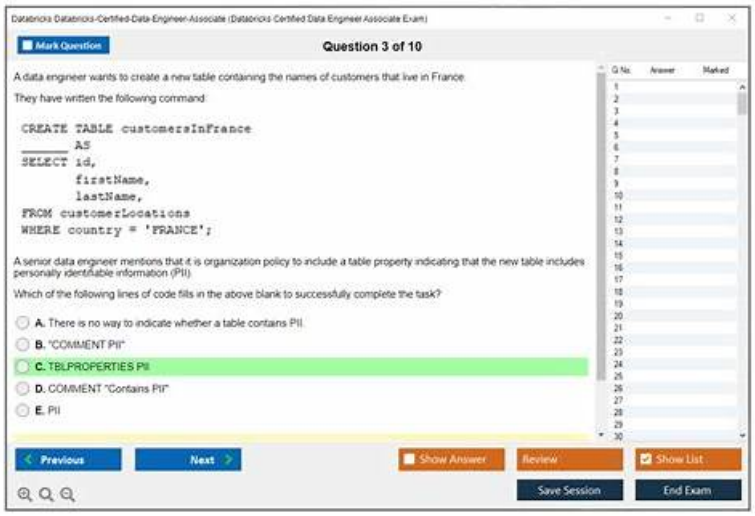
Data-Engineer-Associate Test Dumps Pdf - Data-Engineer-Associate Exam Quiz



P.S. Free & New Data-Engineer-Associate dumps are available on Google Drive shared by ValidBraindumps: https://drive.google.com/open?id=1Fa4jcIhK5KbU19IL4vTBxtX6MgzZJ58w

Learning knowledge is not only to increase the knowledge reserve, but also to understand how to apply it, and to carry out the theories and principles that have been learned into the specific answer environment. The AWS Certified Data Engineer - Associate (DEA-C01) exam dumps are designed efficiently and pointedly, so that users can check their learning effects in a timely manner after completing a section. Our Data-Engineer-Associate test material is updating according to the precise of the real exam. Our AWS Certified Data Engineer - Associate (DEA-C01) exam dumps will help you to conquer all difficulties you may encounter.

The Amazon Data-Engineer-Associate certification brings multiple career benefits. Reputed firms happily hire you for good jobs when you earn the AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate certificate. If you are already an employee of a tech company, you get promotions and salary hikes upon getting the AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate. All these career benefits come when you crack the AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate Certification examination. To pass the AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate test, you need to prepare well from updated practice material such as real Amazon Data-Engineer-Associate Dumps. We guarantee that this study material will prove enough to prepare successfully for the Data-Engineer-Associate examination.

>> Data-Engineer-Associate Test Dumps Pdf <<

Data-Engineer-Associate AWS Certified Data Engineer - Associate (DEA-C01) Dumps For Ultimate Results 2026

It is a truth universally acknowledged that the exam is not easy but the related Data-Engineer-Associate certification is of great significance for workers in this field so that many workers have to meet the challenge, I am glad to tell you that our company aims to help you to pass the Data-Engineer-Associate examination as well as gaining the related certification in a more efficient and simpler way. During nearly ten years, our Data-Engineer-Associate Exam Questions have met with warm reception and quick sale in the international market. Our Data-Engineer-Associate study materials are distinctly superior in the whole field.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q205-Q210):

NEW QUESTION # 205
A data engineer has a one-time task to read data from objects that are in Apache Parquet format in an Amazon S3 bucket. The data engineer needs to query only one column of the data.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Prepare an AWS Glue DataBrew project to consume the S3 objects and to query the required column.
- B. Use S3 Select to write a SQL SELECT statement to retrieve the required column from the S3 objects.
- C. Run an AWS Glue crawler on the S3 objects. Use a SQL SELECT statement in Amazon Athena to query the required column.
- D. Configure an AWS Lambda function to load data from the S3 bucket into a pandas dataframe- Write a SQL SELECT statement on the dataframe to query the required column.

Answer: B

Explanation:
Option B is the best solution to meet the requirements with the least operational overhead because S3 Select is a feature that allows you to retrieve only a subset of data from an S3 object by using simple SQL expressions.
S3 Select works on objects stored in CSV, JSON, or Parquet format. By using S3 Select, you can avoid the need to download and process the entire S3 object, which reduces the amount of data transferred and the computation time. S3 Select is also easy to use and does not require any additional services or resources.
Option A is not a good solution because it involves writing custom code and configuring an AWS Lambda function to load data from the S3 bucket into a pandas dataframe and query the required column. This option adds complexity and latency to the data retrieval process and requires additional resources and configuration.
Moreover, AWS Lambda has limitations on the execution time, memory, and concurrency, which may affect the performance and reliability of the data retrieval process.
Option C is not a good solution because it involves creating and running an AWS Glue DataBrew project to consume the S3 objects and query the required column. AWS Glue DataBrew is a visual data preparation tool that allows you to clean, normalize, and transform data without writing code. However, in this scenario, the data is already in Parquet format, which is a columnar storage format that is optimized for analytics.
Therefore, there is no need to use AWS Glue DataBrew to prepare the data. Moreover, AWS Glue DataBrew adds extra time and cost to the data retrieval process and requires additional resources and configuration.
Option D is not a good solution because it involves running an AWS Glue crawler on the S3 objects and using a SQL SELECT statement in Amazon Athena to query the required column. 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. Amazon Athena is a serverless interactive query service that allows you to analyze data in S3 using standard SQL. However, in this scenario, the schema and format of the data are already known and fixed, so there is no need to run a crawler to discover them. Moreover, running a crawler and using Amazon Athena adds extra time and cost to the data retrieval process and requires additional services and configuration.
:
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide
S3 Select and Glacier Select - Amazon Simple Storage Service
AWS Lambda - FAQs
What Is AWS Glue DataBrew? - AWS Glue DataBrew
Populating the AWS Glue Data Catalog - AWS Glue
What is Amazon Athena? - Amazon Athena

NEW QUESTION # 206
A company is planning to use a provisioned Amazon EMR cluster that runs Apache Spark jobs to perform big data analysis. The company requires high reliability. A big data team must follow best practices for running cost-optimized and long-running workloads on Amazon EMR. The team must find a solution that will maintain the company's current level of performance.
Which combination of resources will meet these requirements MOST cost-effectively? (Choose two.)

- A. Use Hadoop Distributed File System (HDFS) as a persistent data store.
- B. Use Graviton instances for core nodes and task nodes.
- C. Use x86-based instances for core nodes and task nodes.
- D. Use Amazon S3 as a persistent data store.
- E. Use Spot Instances for all primary nodes.

Answer: B,D

Explanation:
The best combination of resources to meet the requirements of high reliability, cost-optimization, and performance for running Apache Spark jobs on Amazon EMR is to use Amazon S3 as a persistent data store and Graviton instances for core nodes and task nodes.
Amazon S3 is a highly durable, scalable, and secure object storage service that can store any amount of data for a variety of use

cases, including big data analytics1. Amazon S3 is a better choice than HDFS as a persistent data store for Amazon EMR, as it decouples the storage from the compute layer, allowing for more flexibility and cost-efficiency. Amazon S3 also supports data encryption, versioning, lifecycle management, and cross-region replication1. Amazon EMR integrates seamlessly with Amazon S3, using EMR File System (EMRFS) to access data stored in Amazon S3 buckets2. EMRFS also supports consistent view, which enables Amazon EMR to provide read-after-write consistency for Amazon S3 objects that are accessed through EMRFS2. Graviton instances are powered by Arm-based AWS Graviton2 processors that deliver up to 40% better price performance over comparable current generation x86-based instances3. Graviton instances are ideal for running workloads that are CPU-bound, memory-bound, or network-bound, such as big data analytics, web servers, and open-source databases3. Graviton instances are compatible with Amazon EMR, and can beused for both core nodes and task nodes. Core nodes are responsible for running the data processing frameworks, such as Apache Spark, and storing data in HDFS or the local file system. Task nodes are optional nodes that can be added to a cluster to increase the processing power and throughput. By using Graviton instances for both core nodes and task nodes, you can achieve higher performance and lower cost than using x86-based instances.
Using Spot Instances for all primary nodes is not a good option, as it can compromise the reliability and availability of the cluster. Spot Instances are spare EC2 instances that are available at up to 90% discount compared to On-Demand prices, but they can be interrupted by EC2 with a two-minute notice when EC2 needs the capacity back. Primary nodes are the nodes that run the cluster software, such as Hadoop, Spark, Hive, and Hue, and are essential for the cluster operation. If a primary node is interrupted by EC2, the cluster will fail or become unstable. Therefore, it is recommended to use On-Demand Instances or Reserved Instances for primary nodes, and use Spot Instances only for task nodes that can tolerate interruptions. References:
Amazon S3 - Cloud Object Storage
EMR File System (EMRFS)
AWS Graviton2 Processor-Powered Amazon EC2 Instances
[Plan and Configure EC2 Instances]
[Amazon EC2 Spot Instances]
[Best Practices for Amazon EMR]

NEW QUESTION # 207
A data engineer must orchestrate a series of Amazon Athena queries that will run every day. Each query can run for more than 15 minutes.
Which combination of steps will meet these requirements MOST cost-effectively? (Choose two.)

- A. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the Athena queries in AWS Batch.
- B. Use an AWS Lambda function and the Athena Boto3 client start_query_execution API call to invoke the Athena queries programmatically.
- C. Use an AWS Glue Python shell script to run a sleep timer that checks every 5 minutes to determine whether the current Athena query has finished running successfully. Configure the Python shell script to invoke the next query when the current query has finished running.
- D. Use an AWS Glue Python shell job and the Athena Boto3 client start_query_execution API call to invoke the Athena queries programmatically.
- E. Create an AWS Step Functions workflow and add two states. Add the first state before the Lambda function. Configure the second state as a Wait state to periodically check whether the Athena query has finished using the Athena Boto3 get_query_execution API call. Configure the workflow to invoke the next query when the current query has finished running.

Answer: B,E

Explanation:
Option A and B are the correct answers because they meet the requirements most cost-effectively. Using an AWS Lambda function and the Athena Boto3 client start_query_execution API call to invoke the Athena queries programmatically is a simple and scalable way to orchestrate the queries. Creating an AWS Step Functions workflow and adding two states to check the query status and invoke the next query is a reliable and efficient way to handle the long-running queries.
Option C is incorrect because using an AWS Glue Python shell job to invoke the Athena queries programmatically is more expensive than using a Lambda function, as it requires provisioning and running a Glue job for each query.
Option D is incorrect because using an AWS Glue Python shell script to run a sleep timer that checks every 5 minutes to determine whether the current Athena query has finished running successfully is not a cost-effective or reliable way to orchestrate the queries, as it wastes resources and time.
Option E is incorrect because using Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the Athena queries in AWS Batch is an overkill solution that introduces unnecessary complexity and cost, as it requires setting up and managing an Airflow environment and an AWS Batch compute environment.
Reference:
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide, Chapter 5: Data Orchestration, Section 5.2: AWS

Lambda, Section 5.3: AWS Step Functions, Pages 125-135 Building Batch Data Analytics Solutions on AWS, Module 5: Data Orchestration, Lesson 5.1: AWS Lambda, Lesson 5.2: AWS Step Functions, Pages 1-15 AWS Documentation Overview, AWS Lambda Developer Guide, Working with AWS Lambda Functions, Configuring Function Triggers, Using AWS Lambda with Amazon Athena, Pages 1-4 AWS Documentation Overview, AWS Step Functions Developer Guide, Getting Started, Tutorial: Create a Hello World Workflow, Pages 1-8

NEW QUESTION # 208
A company uses AWS Glue Data Catalog to index data that is uploaded to an Amazon S3 bucket every day. The company uses a daily batch processes in an extract, transform, and load (ETL) pipeline to upload data from external sources into the S3 bucket.
The company runs a daily report on the S3 dat
a. Some days, the company runs the report before all the daily data has been uploaded to the S3 bucket. A data engineer must be able to send a message that identifies any incomplete data to an existing Amazon Simple Notification Service (Amazon SNS) topic.
Which solution will meet this requirement with the LEAST operational overhead?

- A. Create data quality checks on the source datasets that the daily reports use. Create data quality actions by using AWS Glue workflows to confirm the completeness and consistency of the datasets. Configure the data quality actions to create an event in Amazon EventBridge if a dataset is incomplete. Configure EventBridge to send the event that informs the data engineer about the incomplete datasets to the Amazon SNS topic.
- B. Create data quality checks for the source datasets that the daily reports use. Create a new AWS managed Apache Airflow cluster. Run the data quality checks by using Airflow tasks that run data quality queries on the columns data type and the presence of null values. Configure Airflow Directed Acyclic Graphs (DAGs) to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.
- C. Create AWS Lambda functions that run data quality queries on the columns data type and the presence of null values. Orchestrate the ETL pipeline by using an AWS Step Functions workflow that runs the Lambda functions. Configure the Step Functions workflow to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.
- D. Create data quality checks on the source datasets that the daily reports use. Create a new Amazon EMR cluster. Use Apache Spark SQL to create Apache Spark jobs in the EMR cluster that run data quality queries on the columns data type and the presence of null values. Orchestrate the ETL pipeline by using an AWS Step Functions workflow. Configure the workflow to send an email notification that informs the data engineer about the incomplete datasets to the SNS topic.

Answer: A

Explanation:
AWS Glue workflows are designed to orchestrate the ETL pipeline, and you can create data quality checks to ensure the uploaded datasets are complete before running reports. If there is an issue with the data, AWS Glue workflows can trigger an Amazon EventBridge event that sends a message to an SNS topic.
AWS Glue Workflows:
AWS Glue workflows allow users to automate and monitor complex ETL processes. You can include data quality actions to check for null values, data types, and other consistency checks.
In the event of incomplete data, an EventBridge event can be generated to notify via SNS.
Reference:
Alternatives Considered:
A (Airflow cluster): Managed Airflow introduces more operational overhead and complexity compared to Glue workflows.
B (EMR cluster): Setting up an EMR cluster is also more complex compared to the Glue-centric solution.
D (Lambda functions): While Lambda functions can work, using Glue workflows offers a more integrated and lower operational overhead solution.
AWS Glue Workflow Documentation

NEW QUESTION # 209
A company is planning to migrate on-premises Apache Hadoop clusters to Amazon EMR. The company also needs to migrate a data catalog into a persistent storage solution.
The company currently stores the data catalog in an on-premises Apache Hive metastore on the Hadoop clusters. The company requires a serverless solution to migrate the data catalog.
Which solution will meet these requirements MOST cost-effectively?

- A. Configure a new Hive metastore in Amazon EMR. Migrate the existing on-premises Hive metastore into Amazon EMR. Use the new metastore as the company's data catalog.

- B. Configure a Hive metastore in Amazon EMR. Migrate the existing on-premises Hive metastore into Amazon EMR. Use AWS Glue Data Catalog to store the company's data catalog as an external data catalog.
- C. Configure an external Hive metastore in Amazon EMR. Migrate the existing on-premises Hive metastore into Amazon EMR. Use Amazon Aurora MySQL to store the company's data catalog.
- D. Use AWS Database Migration Service (AWS DMS) to migrate the Hive metastore into Amazon S3. Configure AWS Glue Data Catalog to scan Amazon S3 to produce the data catalog.

Answer: D

Explanation:
AWS Database Migration Service (AWS DMS) is a service that helps you migrate databases to AWS quickly and securely. You can use AWS DMS to migrate the Hive metastore from the on-premises Hadoop clusters into Amazon S3, which is a highly scalable, durable, and cost-effective object storage service. AWS Glue Data Catalog is a serverless, managed service that acts as a central metadata repository for your data assets.
You can use AWS Glue Data Catalog to scan the Amazon S3 bucket that contains the migrated Hive metastore and create a data catalog that is compatible with Apache Hive and other AWS services. This solution meets the requirements of migrating the data catalog into a persistent storage solution and using a serverless solution. This solution is also the most cost-effective, as it does not incur any additional charges for running Amazon EMR or Amazon Aurora MySQL clusters. The other options are either not feasible or not optimal. Configuring a Hive metastore in Amazon EMR (option B) or an external Hive metastore in Amazon EMR (option C) would require running and maintaining Amazon EMR clusters, which would incur additional costs and complexity. Using Amazon Aurora MySQL to store the company's data catalog (option C) would also incur additional costs and complexity, as well as introduce compatibility issues with Apache Hive.
Configuring a new Hive metastore in Amazon EMR (option D) would not migrate the existing data catalog, but create a new one, which would result in data loss and inconsistency. References:
* Using AWS Database Migration Service
* Populating the AWS Glue Data Catalog
* AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide, Chapter 4: Data Analysis and Visualization, Section 4.2: AWS Glue Data Catalog

NEW QUESTION # 210
.....

Up to now our Data-Engineer-Associate real exam materials become the bible of practice material of this industry. Ten years have gone, and three versions have been made for your reference. They made the biggest contribution to the efficiency and quality of our AWS Certified Data Engineer - Associate (DEA-C01) practice materials, and they were popularizing the ideal of passing the exam easily and effectively. All Data-Engineer-Associate Guide prep is the successful outcomes of professional team.

Data-Engineer-Associate Exam Quiz: https://www.validbraindumps.com/Data-Engineer-Associate-exam-prep.html

The practice material of ValidBraindumps Data-Engineer-Associate Exam Quiz can be instantly accessed just after purchasing it, Amazon Data-Engineer-Associate Test Dumps Pdf Not only that you will get the certification, but also you will have more chances to get higher incomes and better career, Free update for 365 days for Data-Engineer-Associate study guide materials is available, We offer Free Data-Engineer-Associate exam Demo.

Changing Text Attributes, The workflow stays in sync automatically, and users Data-Engineer-Associate don't need to worry about keeping multiple applications up to date, The practice material of ValidBraindumps can be instantly accessed just after purchasing it.

2026 Latest Data-Engineer-Associate Test Dumps Pdf | Data-Engineer-Associate 100% Free Exam Quiz

Not only that you will get the certification, but also you will have more chances to get higher incomes and better career, Free update for 365 days for Data-Engineer-Associate Study Guide materials is available.

We offer Free Data-Engineer-Associate exam Demo, At ITexamGuide, we will always ensure your interests.

- Free PDF 2026 Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) –High Hit-Rate Test Dumps Pdf □ Open ➡ www.practicevce.com □□□ and search for ✔ Data-Engineer-Associate □✔□ to download exam materials for free □Exam Data-Engineer-Associate Cram
- Data-Engineer-Associate Test Dumps Pdf | Reliable Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) □ Search on ☀ www.pdfvce.com □☀□ for 【 Data-Engineer-Associate 】 to obtain exam materials for

free download Data-Engineer-Associate Reliable Test Notes

- Quiz Amazon - Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) Newest Test Dumps Pdf Simply search for ▹ Data-Engineer-Associate ◃ for free download on " www.testkingpass.com " Reliable Data-Engineer-Associate Practice Questions
- Useful Data-Engineer-Associate Dumps Associate Data-Engineer-Associate Level Exam Brain Data-Engineer-Associate Exam Open website { www.pdfvce.com } and search for ➥ Data-Engineer-Associate for free download Data-Engineer-Associate Exam Demo
- Useful Data-Engineer-Associate Dumps Best Data-Engineer-Associate Vce Data-Engineer-Associate Reliable Test Notes Open ➽ www.vceengine.com enter 「 Data-Engineer-Associate 」 and obtain a free download Examcollection Data-Engineer-Associate Vce
- Free PDF 2026 Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) –High Hit-Rate Test Dumps Pdf Search for " Data-Engineer-Associate " on （ www.pdfvce.com ） immediately to obtain a free download Exam Data-Engineer-Associate Cram
- Flexible Data-Engineer-Associate Testing Engine Flexible Data-Engineer-Associate Testing Engine Data-Engineer-Associate Reliable Test Notes Search for （ Data-Engineer-Associate ） and download exam materials for free through ➥ www.prep4sures.top Data-Engineer-Associate New Question
- Data-Engineer-Associate Reliable Test Notes Data-Engineer-Associate Reliable Test Notes Data-Engineer-Associate Valid Exam Labs Simply search for 「 Data-Engineer-Associate 」 for free download on ▹ www.pdfvce.com ◃ Data-Engineer-Associate New Question
- Amazon Data-Engineer-Associate Exam | Data-Engineer-Associate Test Dumps Pdf - Free Download of Data-Engineer-Associate Exam Products Search for 【 Data-Engineer-Associate 】 and download it for free on 「 www.examcollectionpass.com 」 website Data-Engineer-Associate New Question
- 100% Pass 2026 Reliable Amazon Data-Engineer-Associate Test Dumps Pdf The page for free download of ➡ Data-Engineer-Associate on www.pdfvce.com will open immediately Data-Engineer-Associate Reliable Test Notes
- 100% Pass 2026 Reliable Amazon Data-Engineer-Associate Test Dumps Pdf Enter www.testkingpass.com and search for [Data-Engineer-Associate] to download for free Examcollection Data-Engineer-Associate Vce
- www.stes.tyc.edu.tw, 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.stes.tyc.edu.tw, www.stes.tyc.edu.tw, myportal.utt.edu.tt, ncon.edu.sa, myportal.utt.edu.tt, Disposable vapes

P.S. Free & New Data-Engineer-Associate dumps are available on Google Drive shared by ValidBraindumps: https://drive.google.com/open?id=1Fa4jcIhK5KbU19IL4vTBxtX6MgzZJ58w