

Amazon Data-Engineer-Associate Exam Dumps Are Verified By Renowned Exam Trainers



P.S. Free & New Data-Engineer-Associate dumps are available on Google Drive shared by LatestCram:
https://drive.google.com/open?id=1vu0mXDgxSSkJg8_GNjTIPOHiP0xxMaA5

This desktop practice exam software completely depicts the Amazon Data-Engineer-Associate exam scenario with proper rules and regulations and any other plugins to access Amazon Data-Engineer-Associate Practice Test. One such trustworthy point about exam preparation material is that it first gains your trust, and then asks you to purchase it.

We are dedicated to helping you pass your exam just one time. Data-Engineer-Associate learning materials are high quality, and we have received plenty of good feedbacks from our customers, they thank us for helping the exam just one time. If you can't pass your exam in your first attempt by using Data-Engineer-Associate exam materials of us, we ensure you that we will give you full refund, and no other questions will be asked. In addition, we provide you with free demo for one year for Data-Engineer-Associate Exam Braindumps, and the update version for Data-Engineer-Associate exam materials will be sent to your email address automatically.

>> Data-Engineer-Associate Latest Exam Pass4sure <<

Free PDF 2026 Amazon High-quality Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Latest Exam Pass4sure

Similarly, the LatestCram Amazon Data-Engineer-Associate practice test creates an actual exam scenario on each and every step so that you may be well prepared before your actual AWS Certified Data Engineer - Associate (DEA-C01) examination time. Hence, it saves you time and money. LatestCram provides three months of free updates if you purchase the Amazon Data-Engineer-Associate Questions and the content of the examination changes after that.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q173-Q178):

NEW QUESTION # 173

A company uses Amazon DataZone as a data governance and business catalog solution. The company stores data in an Amazon S3

data lake. The company uses AWS Glue with an AWS Glue Data Catalog.

A data engineer needs to publish AWS Glue Data Quality scores to the Amazon DataZone portal.

Which solution will meet this requirement?

- A. Configure AWS Glue ETL jobs to use an Evaluate Data Quality transform. Define a data quality ruleset inside the jobs. Configure the Amazon DataZone project to have an Amazon Redshift data source. Enable the data quality configuration for the data source.
- B. Create a data quality ruleset with Data Quality Definition Language (DQDL) rules that apply to a specific AWS Glue table. Schedule the ruleset to run daily. Configure the Amazon DataZone project to have an Amazon Redshift data source. Enable the data quality configuration for the data source.
- C. Create a data quality ruleset with Data Quality Definition Language (DQDL) rules that apply to a specific AWS Glue table. Schedule the ruleset to run daily. Configure the Amazon DataZone project to have an AWS Glue data source. Enable the data quality configuration for the data source.
- D. Configure AWS Glue ETL jobs to use an Evaluate Data Quality transform. Define a data quality ruleset inside the jobs. Configure the Amazon DataZone project to have an AWS Glue data source. Enable the data quality configuration for the data source.

Answer: C

Explanation:

Publishing AWS Glue data quality scores to Amazon DataZone requires creating a DQDL ruleset, scheduling it to run regularly, and then linking the corresponding AWS Glue table as a data source in the DataZone project. The setup ensures that data quality scores from Glue are correctly published and accessible within Amazon DataZone:

"You can define DQDL rulesets for Glue tables and publish the data quality results to DataZone when the project is configured with an AWS Glue data source and the rulesets are scheduled."

-Ace the AWS Certified Data Engineer - Associate Certification - version 2 - apple.pdf Option C follows the expected flow without unnecessary complexity and aligns perfectly with the integration flow supported by AWS.

NEW QUESTION # 174

A security company stores IoT data that is in JSON format in an Amazon S3 bucket. The data structure can change when the company upgrades the IoT devices. The company wants to create a data catalog that includes the IoT data. The company's analytics department will use the data catalog to index the data.

Which solution will meet these requirements MOST cost-effectively?

- A. Create an Amazon Redshift provisioned cluster. Create an Amazon Redshift Spectrum database for the analytics department to explore the data that is in Amazon S3. Create Redshift stored procedures to load the data into Amazon Redshift.
- B. Create an Amazon Athena workgroup. Explore the data that is in Amazon S3 by using Apache Spark through Athena. Provide the Athena workgroup schema and tables to the analytics department.
- C. Create an AWS Glue Data Catalog. Configure an AWS Glue Schema Registry. Create AWS Lambda user defined functions (UDFs) by using the Amazon Redshift Data API. Create an AWS Step Functions job to orchestrate the ingestion of the data that the analytics department will use into Amazon Redshift Serverless.
- D. Create an AWS Glue Data Catalog. Configure an AWS Glue Schema Registry. Create a new AWS Glue workload to orchestrate the ingestion of the data that the analytics department will use into Amazon Redshift Serverless.

Answer: B

Explanation:

The best solution to meet the requirements of creating a data catalog that includes the IoT data, and allowing the analytics department to index the data, most cost-effectively, is to create an Amazon Athena workgroup, explore the data that is in Amazon S3 by using Apache Spark through Athena, and provide the Athena workgroup schema and tables to the analytics department.

Amazon Athena is a serverless, interactive query service that makes it easy to analyze data directly in Amazon S3 using standard SQL or Python¹. Amazon Athena also supports Apache Spark, an open-source distributed processing framework that can run large-scale data analytics applications across clusters of servers². You can use Athena to run Spark code on data in Amazon S3 without having to set up, manage, or scale any infrastructure. You can also use Athena to create and manage external tables that point to your data in Amazon S3, and store them in an external data catalog, such as AWS Glue Data Catalog, Amazon Athena Data Catalog, or your own Apache Hive metastore³. You can create Athena workgroups to separate query execution and resource allocation based on different criteria, such as users, teams, or applications⁴. You can share the schemas and tables in your Athena workgroup with other users or applications, such as Amazon QuickSight, for data visualization and analysis⁵.

Using Athena and Spark to create a data catalog and explore the IoT data in Amazon S3 is the most cost-effective solution, as you

pay only for the queries you run or the compute you use, and you pay nothing when the service is idle¹. You also save on the operational overhead and complexity of managing data warehouse infrastructure, as Athena and Spark are serverless and scalable. You can also benefit from the flexibility and performance of Athena and Spark, as they support various data formats, including JSON, and can handle schema changes and complex queries efficiently.

Option A is not the best solution, as creating an AWS Glue Data Catalog, configuring an AWS Glue Schema Registry, creating a new AWS Glue workload to orchestrate the ingestion of the data that the analytics department will use into Amazon Redshift Serverless, would incur more costs and complexity than using Athena and Spark. AWS Glue Data Catalog is a persistent metadata store that contains table definitions, job definitions, and other control information to help you manage your AWS Glue components⁶. AWS Glue Schema Registry is a service that allows you to centrally store and manage the schemas of your streaming data in AWS Glue Data Catalog⁷. AWS Glue is a serverless data integration service that makes it easy to prepare, clean, enrich, and move data between data stores⁸. Amazon Redshift Serverless is a feature of Amazon Redshift, a fully managed data warehouse service, that allows you to run and scale analytics without having to manage data warehouse infrastructure⁹. While these services are powerful and useful for many data engineering scenarios, they are not necessary or cost-effective for creating a data catalog and indexing the IoT data in Amazon S3. AWS Glue Data Catalog and Schema Registry charge you based on the number of objects stored and the number of requests made^{6,7}. AWS Glue charges you based on the compute time and the data processed by your ETL jobs⁸. Amazon Redshift Serverless charges you based on the amount of data scanned by your queries and the compute time used by your workloads⁹. These costs can add up quickly, especially if you have large volumes of IoT data and frequent schema changes. Moreover, using AWS Glue and Amazon Redshift Serverless would introduce additional latency and complexity, as you would have to ingest the data from Amazon S3 to Amazon Redshift Serverless, and then query it from there, instead of querying it directly from Amazon S3 using Athena and Spark.

Option B is not the best solution, as creating an Amazon Redshift provisioned cluster, creating an Amazon Redshift Spectrum database for the analytics department to explore the data that is in Amazon S3, and creating Redshift stored procedures to load the data into Amazon Redshift, would incur more costs and complexity than using Athena and Spark. Amazon Redshift provisioned clusters are clusters that you create and manage by specifying the number and type of nodes, and the amount of storage and compute capacity¹⁰. Amazon Redshift Spectrum is a feature of Amazon Redshift that allows you to query and join data across your data warehouse and your data lake using standard SQL¹¹. Redshift stored procedures are SQL statements that you can define and store in Amazon Redshift, and then call them by using the CALL command¹². While these features are powerful and useful for many data warehousing scenarios, they are not necessary or cost-effective for creating a data catalog and indexing the IoT data in Amazon S3. Amazon Redshift provisioned clusters charge you based on the node type, the number of nodes, and the duration of the cluster¹⁰. Amazon Redshift Spectrum charges you based on the amount of data scanned by your queries¹¹. These costs can add up quickly, especially if you have large volumes of IoT data and frequent schema changes. Moreover, using Amazon Redshift provisioned clusters and Spectrum would introduce additional latency and complexity, as you would have to provision and manage the cluster, create an external schema and database for the data in Amazon S3, and load the data into the cluster using stored procedures, instead of querying it directly from Amazon S3 using Athena and Spark.

Option D is not the best solution, as creating an AWS Glue Data Catalog, configuring an AWS Glue Schema Registry, creating AWS Lambda user defined functions (UDFs) by using the Amazon Redshift Data API, and creating an AWS Step Functions job to orchestrate the ingestion of the data that the analytics department will use into Amazon Redshift Serverless, would incur more costs and complexity than using Athena and Spark. AWS Lambda is a serverless compute service that lets you run code without provisioning or managing servers¹³. AWS Lambda UDFs are Lambda functions that you can invoke from within an Amazon Redshift query. Amazon Redshift Data API is a service that allows you to run SQL statements on Amazon Redshift clusters using HTTP requests, without needing a persistent connection. AWS Step Functions is a service that lets you coordinate multiple AWS services into serverless workflows. While these services are powerful and useful for many data engineering scenarios, they are not necessary or cost-effective for creating a data catalog and indexing the IoT data in Amazon S3. AWS Glue Data Catalog and Schema Registry charge you based on the number of objects stored and the number of requests made^{6,7}. AWS Lambda charges you based on the number of requests and the duration of your functions¹³. Amazon Redshift Serverless charges you based on the amount of data scanned by your queries and the compute time used by your workloads⁹. AWS Step Functions charges you based on the number of state transitions in your workflows. These costs can add up quickly, especially if you have large volumes of IoT data and frequent schema changes. Moreover, using AWS Glue, AWS Lambda, Amazon Redshift Data API, and AWS Step Functions would introduce additional latency and complexity, as you would have to create and invoke Lambda functions to ingest the data from Amazon S3 to Amazon Redshift Serverless using the Data API, and coordinate the ingestion process using Step Functions, instead of querying it directly from Amazon S3 using Athena and Spark. References:

What is Amazon Athena?

Apache Spark on Amazon Athena

Creating tables, updating the schema, and adding new partitions in the Data Catalog from AWS Glue ETL jobs Managing Athena workgroups Using Amazon QuickSight to visualize data in Amazon Athena AWS Glue Data Catalog 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?

[Creating and using AWS Lambda UDFs]

[Using the Amazon Redshift Data API]

[What is AWS Step Functions?]

AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 175

A company currently uses a provisioned Amazon EMR cluster that includes general purpose Amazon EC2 instances. The EMR cluster uses EMR managed scaling between one to five task nodes for the company's long-running Apache Spark extract, transform, and load (ETL) job. The company runs the ETL job every day.

When the company runs the ETL job, the EMR cluster quickly scales up to five nodes. The EMR cluster often reaches maximum CPU usage, but the memory usage remains under 30%.

The company wants to modify the EMR cluster configuration to reduce the EMR costs to run the daily ETL job.

Which solution will meet these requirements MOST cost-effectively?

- A. Change the task node type from general purpose EC2 instances to memory optimized EC2 instances.
- B. Reduce the scaling cooldown period for the provisioned EMR cluster.
- C. **Switch the task node type from general purpose EC2 instances to compute optimized EC2 instances.**
- D. Increase the maximum number of task nodes for EMR managed scaling to 10.

Answer: C

Explanation:

The company's Apache Spark ETL job on Amazon EMR uses high CPU but low memory, meaning that compute-optimized EC2 instances would be the most cost-effective choice. These instances are designed for high-performance compute applications, where CPU usage is high, but memory needs are minimal, which is exactly the case here.

* Compute Optimized Instances:

* Compute-optimized instances, such as the C5 series, provide a higher ratio of CPU to memory, which is more suitable for jobs with high CPU usage and relatively low memory consumption.

* Switching from general-purpose EC2 instances to compute-optimized instances can reduce costs while improving performance, as these instances are optimized for workloads like Spark jobs that perform a lot of computation.

NEW QUESTION # 176

A company is building a data lake for a new analytics team. The company is using Amazon S3 for storage and Amazon Athena for query analysis. All data that is in Amazon S3 is in Apache Parquet format.

The company is running a new Oracle database as a source system in the company's data center. The company has 70 tables in the Oracle database. All the tables have primary keys. Data can occasionally change in the source system. The company wants to ingest the tables every day into the data lake.

Which solution will meet this requirement with the LEAST effort?

- A. Create an Apache Sqoop job in Amazon EMR to read the data from the Oracle database. Configure the Sqoop job to write the data to Amazon S3 in Parquet format.
- B. Create an Oracle database in Amazon RDS. Use AWS Database Migration Service (AWS DMS) to migrate the on-premises Oracle database to Amazon RDS. Configure triggers on the tables to invoke AWS Lambda functions to write changed records to Amazon S3 in Parquet format.
- C. **Create an AWS Database Migration Service (AWS DMS) task for ongoing replication. Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.**
- D. Create an AWS Glue connection to the Oracle database. Create an AWS Glue bookmark job to ingest the data incrementally and to write the data to Amazon S3 in Parquet format.

Answer: C

Explanation:

The company needs to ingest tables from an on-premises Oracle database into a data lake on Amazon S3 in Apache Parquet format. The most efficient solution, requiring the least manual effort, would be to use AWS Database Migration Service (DMS) for continuous data replication.

Option C: Create an AWS Database Migration Service (AWS DMS) task for ongoing replication. Set the Oracle database as the source. Set Amazon S3 as the target. Configure the task to write the data in Parquet format.

AWS DMS can continuously replicate data from the Oracle database into Amazon S3, transforming it into Parquet format as it ingests the data. DMS simplifies the process by providing ongoing replication with minimal setup, and it automatically handles the

conversion to Parquet format without requiring manual transformations or separate jobs. This option is the least effort solution since it automates both the ingestion and transformation processes.

Other options:

Option A (Apache Sqoop on EMR) involves more manual configuration and management, including setting up EMR clusters and writing Sqoop jobs.

Option B (AWS Glue bookmark job) involves configuring Glue jobs, which adds complexity. While Glue supports data transformations, DMS offers a more seamless solution for database replication.

Option D (RDS and Lambda triggers) introduces unnecessary complexity by involving RDS and Lambda for a task that DMS can handle more efficiently.

Reference:

AWS Database Migration Service (DMS)

DMS S3 Target Documentation

NEW QUESTION # 177

A company uses Amazon RDS to store transactional data. The company runs an RDS DB instance in a private subnet. A developer wrote an AWS Lambda function with default settings to insert, update, or delete data in the DB instance.

The developer needs to give the Lambda function the ability to connect to the DB instance privately without using the public internet. Which combination of steps will meet this requirement with the LEAST operational overhead? (Choose two.)

- A. Configure the Lambda function to run in the same subnet that the DB instance uses.
- B. Update the security group of the DB instance to allow only Lambda function invocations on the database port.
- C. Turn on the public access setting for the DB instance.
- D. Attach the same security group to the Lambda function and the DB instance. Include a self-referencing rule that allows access through the database port.
- E. Update the network ACL of the private subnet to include a self-referencing rule that allows access through the database port.

Answer: A,D

Explanation:

To enable the Lambda function to connect to the RDS DB instance privately without using the public internet, the best combination of steps is to configure the Lambda function to run in the same subnet that the DB instance uses, and attach the same security group to the Lambda function and the DB instance. This way, the Lambda function and the DB instance can communicate within the same private network, and the security group can allow traffic between them on the database port. This solution has the least operational overhead, as it does not require any changes to the public access setting, the network ACL, or the security group of the DB instance.

The other options are not optimal for the following reasons:

* A. Turn on the public access setting for the DB instance. This option is not recommended, as it would expose the DB instance to the public internet, which can compromise the security and privacy of the data. Moreover, this option would not enable the Lambda function to connect to the DB instance privately, as it would still require the Lambda function to use the public internet to access the DB instance.

* B. Update the security group of the DB instance to allow only Lambda function invocations on the database port. This option is not sufficient, as it would only modify the inbound rules of the security group of the DB instance, but not the outbound rules of the security group of the Lambda function.

Moreover, this option would not enable the Lambda function to connect to the DB instance privately, as it would still require the Lambda function to use the public internet to access the DB instance.

* E. Update the network ACL of the private subnet to include a self-referencing rule that allows access through the database port. This option is not necessary, as the network ACL of the private subnet already allows all traffic within the subnet by default.

Moreover, this option would not enable the Lambda function to connect to the DB instance privately, as it would still require the Lambda function to use the public internet to access the DB instance.

1: Connecting to an Amazon RDS DB instance

2: Configuring a Lambda function to access resources in a VPC

3: Working with security groups

4: Network ACLs

NEW QUESTION # 178

.....

Do you worry about not having a long-term fixed study time? Do you worry about not having a reasonable plan for yourself? Data-Engineer-Associate exam dumps will solve this problem for you. Based on your situation, including the available time, your current level of knowledge, our study materials will develop appropriate plans and learning materials. Whatever you want to choose, you want to learn from which stage. In our study materials, you can find the right one for you. At the same time, the Data-Engineer-Associate Exam Prep is constantly updated. After you have finished learning a part, you can choose a new method according to your own situation. Our study materials are so easy to understand that no matter who you are, you can find what you want here.

Data-Engineer-Associate Reliable Braindumps Free: <https://www.latestcram.com/Data-Engineer-Associate-exam-cram-questions.html>

Perhaps after you have used Data-Engineer-Associate real exam once, you will agree with this point, The progress of previously given AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) practice tests are saved in the history so that the customers can assess it and avoid mistakes in future exams and pass AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) certification exam easily. Being an excellent people is a chronic process, but sometimes to get the certificates of great importance in limited time, we have to finish the ultimate task---pass the certificate quickly and effectively by using useful Data-Engineer-Associate training materials: AWS Certified Data Engineer - Associate (DEA-C01) in the market.

A lot of negotiating is attitude and presentation, First, Nesamoney explores Data-Engineer-Associate the current state of digital advertising, reviewing best practices for maximizing personalization through long-familiar data and targeting technologies.

Updated Data-Engineer-Associate Latest Exam Pass4sure offer you accurate Reliable Braindumps Free | Amazon AWS Certified Data Engineer - Associate (DEA-C01)

Perhaps after you have used Data-Engineer-Associate Real Exam once, you will agree with this point, The progress of previously given AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) practice tests are saved in the history so that the customers can assess it and avoid mistakes in future exams and pass AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) certification exam easily.

Being an excellent people is a chronic process, Data-Engineer-Associate Question Explanations but sometimes to get the certificates of great importance in limited time, we have to finish the ultimate task---pass the certificate quickly and effectively by using useful Data-Engineer-Associate training materials: AWS Certified Data Engineer - Associate (DEA-C01) in the market.

If you are still hesitating whether to select Amazon Data-Engineer-Associate VCE dumps, you can download our free dumps to determine our reliability, Most people spend much money and time to prepare the Data-Engineer-Associate dump torrent and Data-Engineer-Associate real pdf dumps, but the result is bad.

- Get latest AWS Certified Data Engineer - Associate (DEA-C01) Prepare Torrent Pass the AWS Certified Data Engineer - Associate (DEA-C01) Exam in the First Attempt - www.prepawaypdf.com □ Open ➔ www.prepawaypdf.com □ enter { Data-Engineer-Associate } and obtain a free download □ Data-Engineer-Associate Exam Lab Questions
- New Data-Engineer-Associate Exam Discount □ Data-Engineer-Associate Updated Demo □ Data-Engineer-Associate Valid Exam Registration □ Easily obtain free download of ✓ Data-Engineer-Associate □ ✓ □ by searching on (www.pdfvce.com) * Reliable Data-Engineer-Associate Test Preparation
- AWS Certified Data Engineer - Associate (DEA-C01) free prep material - Data-Engineer-Associate valid braindumps □ Search for { Data-Engineer-Associate } and download exam materials for free through ➔ www.examdiscuss.com □ □ □ Latest Data-Engineer-Associate Examprep
- The Best Accurate Data-Engineer-Associate Latest Exam Pass4sure – Find Shortcut to Pass Data-Engineer-Associate Exam □ Search on (www.pdfvce.com) for ➤ Data-Engineer-Associate □ to obtain exam materials for free download □ Data-Engineer-Associate Exam Lab Questions
- Data-Engineer-Associate Trusted Exam Resource □ Data-Engineer-Associate Exam Lab Questions □ Data-Engineer-Associate Exam Vce Free □ Search for ✓ Data-Engineer-Associate □ ✓ □ and download exam materials for free through ➔ www.practicevce.com □ □ Reliable Data-Engineer-Associate Exam Cost
- Data-Engineer-Associate Test Registration □ Reliable Data-Engineer-Associate Test Preparation □ Data-Engineer-Associate Updated Demo □ Search for ➤ Data-Engineer-Associate □ and download it for free immediately on [www.pdfvce.com] □ Reliable Data-Engineer-Associate Test Preparation
- Reliable Data-Engineer-Associate Test Preparation □ Data-Engineer-Associate Exam Vce Free □ Data-Engineer-Associate Updated Demo □ Search for ➤ Data-Engineer-Associate □ and download exam materials for free through □ www.troytecdumps.com □ □ Valid Data-Engineer-Associate Vce
- Data-Engineer-Associate Certification Exam Infor □ Valid Data-Engineer-Associate Vce □ Data-Engineer-Associate Valid Study Plan □ Open ➔ www.pdfvce.com □ □ □ and search for ➤ Data-Engineer-Associate □ □ □ to download exam materials for free □ Reliable Data-Engineer-Associate Test Preparation

P.S. Free & New Data-Engineer-Associate dumps are available on Google Drive shared by LatestCram:
https://drive.google.com/open?id=1vu0mXDgxSSkJg8_GNjTIPOHiP0xxMaA5