

Free PDF Quiz Amazon - Data-Engineer-Associate - AWS Certified Data Engineer - Associate (DEA-C01) Fantastic Valid Study Materials



P.S. Free & New Data-Engineer-Associate dumps are available on Google Drive shared by PassExamDumps:
<https://drive.google.com/open?id=1oqi-lnLY8WA9CpG2RzW7vDeexirWDpwj>

You can use Data-Engineer-Associate guide materials through a variety of electronic devices. At home, you can use the computer and outside you can also use the phone. Now that more people are using mobile phones to learn our Data-Engineer-Associate study materials, you can also choose the one you like. One advantage is that if you use our Data-Engineer-Associate Practice Questions for the first time in a network environment, then the next time you use our study materials, there will be no network requirements. You can open the Data-Engineer-Associate real exam anytime and anywhere.

The product we provide with you is compiled by professionals elaborately and boosts varied versions which aimed to help you learn the Data-Engineer-Associate study materials by the method which is convenient for you. They check the update every day, and we can guarantee that you can get a free update service from the date of purchase. Once you have any questions and doubts about the Data-Engineer-Associate Exam Questions we will provide you with our customer service before or after the sale, you can contact us if you have question or doubt about our exam materials and the professional personnel can help you solve your issue about using Data-Engineer-Associate study materials.

>> Valid Data-Engineer-Associate Study Materials <<

Data-Engineer-Associate Pass Test | Test Data-Engineer-Associate Discount Voucher

The second form is AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) web-based practice test which can be accessed through online browsing. The Data-Engineer-Associate web-based practice test is supported by browsers like Firefox, Microsoft Edge, Amazon Chrome, and Safari. You don't need to install any plugins or software to attempt the Data-Engineer-Associate web-based practice test. This online Amazon Data-Engineer-Associate exam is also compatible with all operating systems.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q123-Q128):

NEW QUESTION # 123

A company stores datasets in JSON format and .csv format in an Amazon S3 bucket. The company has Amazon RDS for Microsoft SQL Server databases, Amazon DynamoDB tables that are in provisioned capacity mode, and an Amazon Redshift cluster. A data engineering team must develop a solution that will give data scientists the ability to query all data sources by using syntax similar to SQL.

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

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

Answer: A

Explanation:

The best solution to meet the requirements of giving data scientists the ability to query all data sources by using syntax similar to SQL with the least operational overhead is to use AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use Amazon Athena to query the data, use SQL for structured data sources, and use PartiQL for data that is stored in JSON format.

AWS Glue is a serverless data integration service that makes it easy to prepare, clean, enrich, and move data between data stores¹. AWS Glue crawlers are processes that connect to a data store, progress through a prioritized list of classifiers to determine the schema for your data, and then create metadata tables in the Data Catalog². The Data Catalog is a persistent metadata store that contains table definitions, job definitions, and other control information to help you manage your AWS Glue components³. You can use AWS Glue to crawl the data sources, such as Amazon S3, Amazon RDS for Microsoft SQL Server, and Amazon DynamoDB, and store the metadata in the Data Catalog.

Amazon Athena is a serverless, interactive query service that makes it easy to analyze data directly in Amazon S3 using standard SQL or Python⁴. Amazon Athena also supports PartiQL, a SQL-compatible query language that lets you query, insert, update, and delete data from semi-structured and nested data, such as JSON. You can use Amazon Athena to query the data from the Data Catalog using SQL for structured data sources, such as .csv files and relational databases, and PartiQL for data that is stored in JSON format. You can also use Athena to query data from other data sources, such as Amazon Redshift, using federated queries. Using AWS Glue and Amazon Athena to query all data sources by using syntax similar to SQL is the least operational overhead solution, as you do not need to provision, manage, or scale any infrastructure, and you pay only for the resources you use. AWS Glue charges you based on the compute time and the data processed by your crawlers and ETL jobs¹. Amazon Athena charges you based on the amount of data scanned by your queries. You can also reduce the cost and improve the performance of your queries by using compression, partitioning, and columnar formats for your data in Amazon S3.

Option B is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, and use Redshift Spectrum to query the data, would incur more costs and complexity than using Amazon Athena. Redshift Spectrum is a feature of Amazon Redshift, a fully managed data warehouse service, that allows you to query and join data across your data warehouse and your data lake using standard SQL. While Redshift Spectrum is powerful and useful for many data warehousing scenarios, it is not necessary or cost-effective for querying all data sources by using syntax similar to SQL. Redshift Spectrum charges you based on the amount of data scanned by your queries, which is similar to Amazon Athena, but it also requires you to have an Amazon Redshift cluster, which charges you based on the node type, the number of nodes, and the duration of the cluster⁵. These costs can add up quickly, especially if you have large volumes of data and complex queries. Moreover, using Redshift Spectrum would introduce additional latency and complexity, as you would have to provision and manage the cluster, and create an external schema and database for the data in the Data Catalog, instead of querying it directly from Amazon Athena.

Option C is not the best solution, as using AWS Glue to crawl the data sources, store metadata in the AWS Glue Data Catalog, use AWS Glue jobs to transform data that is in JSON format to Apache Parquet or .csv format, store the transformed data in an S3 bucket, and use Amazon Athena to query the original and transformed data from the S3 bucket, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Glue jobs are ETL scripts that you can write in Python or Scala to transform your data and load it to your target data store. Apache Parquet is a columnar storage format that can improve the performance of analytical queries by reducing the amount of data that needs to be scanned and providing efficient compression and encoding schemes⁶. While using AWS Glue jobs and Parquet can improve the performance and reduce the cost of your queries,

they would also increase the complexity and the operational overhead of the data pipeline, as you would have to write, run, and monitor the ETL jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using AWS Glue jobs and Parquet would introduce additional latency, as you would have to wait for the ETL jobs to finish before querying the transformed data.

Option D is not the best solution, as using AWS Lake Formation to create a data lake, use Lake Formation jobs to transform the data from all data sources to Apache Parquet format, store the transformed data in an S3 bucket, and use Amazon Athena or Redshift Spectrum to query the data, would incur more costs and complexity than using Amazon Athena with PartiQL. AWS Lake Formation is a service that helps you centrally govern, secure, and globally share data for analytics and machine learning⁷. Lake Formation jobs are ETL jobs that you can create and run using the Lake Formation console or API. While using Lake Formation and Parquet can improve the performance and reduce the cost of your queries, they would also increase the complexity and the operational overhead of the data pipeline, as you would have to create, run, and monitor the Lake Formation jobs, and store the transformed data in a separate location in Amazon S3. Moreover, using Lake Formation and Parquet would introduce additional latency, as you would have to wait for the Lake Formation jobs to finish before querying the transformed data. Furthermore, using Redshift Spectrum to query the data would also incur the same costs and complexity as mentioned in option B. References:

What is Amazon Athena?

Data Catalog and crawlers in AWS Glue

AWS Glue Data Catalog

Columnar Storage Formats

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

AWS Glue Schema Registry

What is AWS Glue?

Amazon Redshift Serverless

Amazon Redshift provisioned clusters

[Querying external data using Amazon Redshift Spectrum]

[Using stored procedures in Amazon Redshift]

[What is AWS Lambda?]

[PartiQL for Amazon Athena]

[Federated queries in Amazon Athena]

[Amazon Athena pricing]

[Top 10 performance tuning tips for Amazon Athena]

[AWS Glue ETL jobs]

[AWS Lake Formation jobs]

NEW QUESTION # 124

A company analyzes data in a data lake every quarter to perform inventory assessments. A data engineer uses AWS Glue DataBrew to detect any personally identifiable information (PII) about customers within the data. The company's privacy policy considers some custom categories of information to be PII. However, the categories are not included in standard DataBrew data quality rules.

The data engineer needs to modify the current process to scan for the custom PII categories across multiple datasets within the data lake.

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

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

Answer: B

Explanation:

The data engineer needs to detect custom categories of PII within the data lake using AWS Glue DataBrew. While DataBrew provides standard data quality rules, the solution must support custom PII categories.

Option B: Implement custom data quality rules in DataBrew. Apply the custom rules across datasets.

This option is the most efficient because DataBrew allows the creation of custom data quality rules that can be applied to detect specific data patterns, including custom PII categories. This approach minimizes operational overhead while ensuring that the specific privacy requirements are met.

Options A, C, and D either involve manual intervention or developing custom scripts, both of which increase operational effort compared to using DataBrew's built-in capabilities.

Reference:

NEW QUESTION # 125

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. Run an AWS Glue crawler on the S3 objects. Use a SQL SELECT statement in Amazon Athena 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. 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.
- D. Prepare an AWS Glue DataBrew project to consume the S3 objects and 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.

References:

- * 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 # 126

Files from multiple data sources arrive in an Amazon S3 bucket on a regular basis. A data engineer wants to ingest new files into Amazon Redshift in near real time when the new files arrive in the S3 bucket.

Which solution will meet these requirements?

- A. Use the zero-ETL integration between Amazon Aurora and Amazon Redshift to load new files into Amazon Redshift.
- B. Use AWS Glue job bookmarks to extract, transform, and load (ETL) load new files into Amazon Redshift.
- **C. Use S3 Event Notifications to invoke an AWS Lambda function that loads new files into Amazon Redshift.**
- D. Use the query editor v2 to schedule a COPY command to load new files into Amazon Redshift.

Answer: C

Explanation:

For near real-time processing of new files in S3, event-driven ingestion is optimal. S3 Event Notifications can trigger AWS Lambda to immediately load data into Amazon Redshift, eliminating latency associated with batch scheduling.

"Event-based triggers using S3 notifications and Lambda functions are effective for near real-time ingestion pipelines into Amazon Redshift."

-Ace the AWS Certified Data Engineer - Associate Certification - version 2 - apple.pdf Option C (Glue bookmarks) is best for batch jobs, and zero-ETL applies to Aurora to Redshift, not S3.

NEW QUESTION # 127

A marketing company uses Amazon S3 to store marketing data. The company uses versioning in some buckets. The company runs several jobs to read and load data into the buckets.

To help cost-optimize its storage, the company wants to gather information about incomplete multipart uploads and outdated versions that are present in the S3 buckets.

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

- A. Use Amazon S3 Inventory configurations reports to gather the information.
- B. Use AWS CLI to gather the information.
- C. Use the Amazon S3 Storage Lens dashboard to gather the information.
- D. Use AWS usage reports for Amazon S3 to gather the information.

Answer: A

Explanation:

The company wants to gather information about incomplete multipart uploads and outdated versions in its Amazon S3 buckets to optimize storage costs.

* Option B: Use Amazon S3 Inventory configurations reports to gather the information. S3 Inventory provides reports that can list incomplete multipart uploads and versions of objects stored in S3. It offers an easy, automated way to track object metadata across buckets, including data necessary for cost optimization, without manual effort.

Options A (AWS CLI), C (S3 Storage Lens), and D (usage reports) either do not specifically gather the required information about incomplete uploads and outdated versions or require more manual intervention.

References:

* Amazon S3 Inventory Documentation

NEW QUESTION # 128

.....

For easy use, PassExamDumps provides you with different version Data-Engineer-Associate exam dumps. PDF version dumps are easy to read and reproduce the real exam. SOFT version dumps is a test engine which can measure what your preparations for the exam. If you want to know whether you prepare well for the Data-Engineer-Associate test, you can take advantage of the SOFT version dumps to measure your ability. So you can quickly know your weaknesses and shortcomings, which is helpful to your further study.

Data-Engineer-Associate Pass Test: <https://www.passexamdumps.com/Data-Engineer-Associate-valid-exam-dumps.html>

Data-Engineer-Associate online test engine allows candidates to learn in the case of mock examinations, They are waiting to give you the most professional suggestions on our Data-Engineer-Associate exam questions, One year free updating of our Data-Engineer-Associate exam dumps, Amazon Valid Data-Engineer-Associate Study Materials Why we can grow so fast, Amazon Valid Data-Engineer-Associate Study Materials So do not worry about anything, Whenever you have questions about our Data-Engineer-Associate learning quiz, you are welcome to contact us via E-mail.

This chapter reviews the process of subnetting IP networks, The `DataSource` object can then be used to obtain a `java.sql.Connection` object, Data-Engineer-Associate Online Test engine allows candidates to learn in the case of mock examinations.

100% Pass Quiz Amazon - Reliable Valid Data-Engineer-Associate Study Materials

They are waiting to give you the most professional suggestions on our Data-Engineer-Associate exam questions, One year free updating of our Data-Engineer-Associate exam dumps, Why we can grow so fast?

So do not worry about anything.

- Data-Engineer-Associate Positive Feedback □ Data-Engineer-Associate New Practice Materials □ Valid Exam Data-Engineer-Associate Practice □ □ www.exam4labs.com □ is best website to obtain ➔ Data-Engineer-Associate □ □ □ for free download □ Data-Engineer-Associate New Dumps Book
- Free PDF Quiz Amazon - Valid Data-Engineer-Associate - Valid AWS Certified Data Engineer - Associate (DEA-C01) Study Materials □ Easily obtain free download of ➔ Data-Engineer-Associate □ by searching on ☀ www.pdfvce.com □ ☀ □ □ Data-Engineer-Associate Test Dumps.zip
- 2026 First-grade Amazon Data-Engineer-Associate: Valid AWS Certified Data Engineer - Associate (DEA-C01) Study Materials □ Search for □ Data-Engineer-Associate □ and download it for free on (www.prepawaypdf.com) website □ Free Data-Engineer-Associate Practice Exams
- Data-Engineer-Associate Reliable Braindumps Sheet □ Valid Exam Data-Engineer-Associate Practice □ Data-Engineer-Associate New Practice Materials ✱ Search for ⇒ Data-Engineer-Associate ⇐ and download it for free on 《 www.pdfvce.com 》 website □ Popular Data-Engineer-Associate Exams
- 100% Pass Quiz 2026 Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Newest Valid Study Materials □ Enter ☀ www.examcollectionpass.com □ ☀ □ and search for ➔ Data-Engineer-Associate □ to download for free □ Data-Engineer-Associate Reliable Exam Review
- Data-Engineer-Associate Reliable Braindumps Sheet □ Data-Engineer-Associate Valid Dump □ Popular Data-Engineer-Associate Exams □ Immediately open ➤ www.pdfvce.com □ and search for □ Data-Engineer-Associate □ to obtain a free download □ Popular Data-Engineer-Associate Exams
- Free PDF Quiz Amazon - Valid Data-Engineer-Associate - Valid AWS Certified Data Engineer - Associate (DEA-C01) Study Materials □ Search for □ Data-Engineer-Associate □ and download it for free on 「 www.vce4dumps.com 」 website □ Best Data-Engineer-Associate Preparation Materials
- Free PDF Quiz Amazon - Valid Data-Engineer-Associate - Valid AWS Certified Data Engineer - Associate (DEA-C01) Study Materials □ Immediately open □ www.pdfvce.com □ and search for 「 Data-Engineer-Associate 」 to obtain a free download □ Data-Engineer-Associate Valid Test Experience
- 2026 First-grade Amazon Data-Engineer-Associate: Valid AWS Certified Data Engineer - Associate (DEA-C01) Study Materials □ Immediately open □ www.vce4dumps.com □ and search for [Data-Engineer-Associate] to obtain a free download □ Exam Data-Engineer-Associate Guide Materials
- Real Amazon Data-Engineer-Associate PDF Questions [2026]-The Greatest Shortcut Towards Success □ Search for ☀ Data-Engineer-Associate □ ☀ □ on ⇒ www.pdfvce.com ⇐ immediately to obtain a free download □ Data-Engineer-Associate Test Dumps.zip
- Data-Engineer-Associate Certification Sample Questions □ Data-Engineer-Associate Positive Feedback □ Latest Data-Engineer-Associate Test Sample □ Search for 《 Data-Engineer-Associate 》 and download exam materials for free through 《 www.practicevce.com 》 □ Data-Engineer-Associate Reliable Braindumps Sheet
- www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, wedacareer.com, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, Disposable vapes

DOWNLOAD the newest PassExamDumps Data-Engineer-Associate PDF dumps from Cloud Storage for free:
<https://drive.google.com/open?id=1oqi-lnLY8WA9CpG2RzW7vDeexirWDpwj>