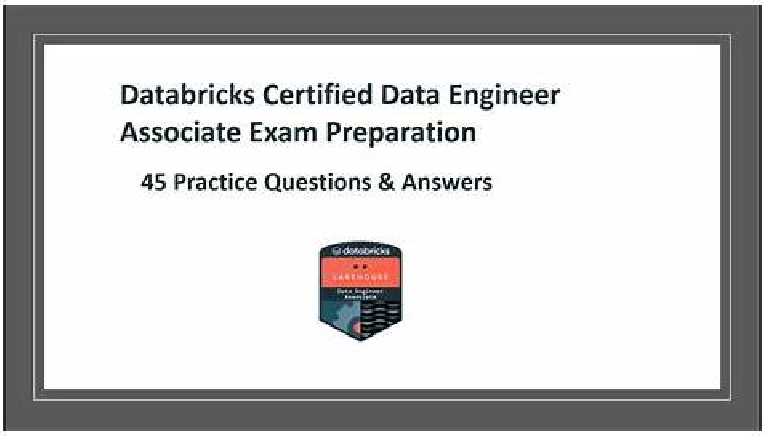
Top Features of Prep4SureReview Amazon Data-Engineer-Associate Practice Test Software



2026 Latest Prep4SureReview Data-Engineer-Associate PDF Dumps and Data-Engineer-Associate Exam Engine Free Share: https://drive.google.com/open?id=1MqchLU2ggoBNt9_-RBpmNzLlp8mpEEQK

If you want to choose passing Amazon certification Data-Engineer-Associate exam to make yourself have a more stable position in today's competitive IT area and the professional ability become more powerful, you must have a strong expertise. And passing Amazon certification Data-Engineer-Associate exam is not very simple. Perhaps passing Amazon Certification Data-Engineer-Associate Exam is a stepping stone to promote yourself in the IT area, but it doesn't need to spend a lot of time and effort to review the relevant knowledge, you can choose to use our Prep4SureReview product, a training tool prepared for the IT certification exams.

With pass rate reaching 96%, our Data-Engineer-Associate exam materials have gained popularity in the market, and many candidates choose us for this reason. We can help you pass the exam just one time. What's more, Data-Engineer-Associate exam materials are high quality, and you can improve your efficiency by using them. You can receive your downloading link and password within ten minutes after payment, so that you can start your learning by using Data-Engineer-Associate Exam Dumps. Free update for one year is available, and our system will send the latest version to your email automatically, you just need to check your email for the latest version.

>> Data-Engineer-Associate Real Question <<

Data-Engineer-Associate Practice Guide & Data-Engineer-Associate Reliable Exam Topics

We now live in a world which needs the talents who can combine the practical abilities and knowledge to apply their knowledge into the practical working conditions. To prove that you are that kind of talents you must boost some authorized and useful certificate and the test Data-Engineer-Associate certificate is one kind of these certificate. Most important of all, as long as we have compiled a new version of the Data-Engineer-Associate Exam Questions, we will send the latest version of our Data-Engineer-Associate exam questions to our customers for free during the whole year after purchasing. Our product can improve your stocks of knowledge and your abilities in some area and help you gain the success in your career.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q71-Q76):

NEW QUESTION # 71
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 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.

- B. 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.
- C. 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.
- D. 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.

Answer: D

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 theingestion 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 made67. 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 andmanage 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 made67. 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 # 72
A transportation company wants to track vehicle movements by capturing geolocation records. The records are 10 bytes in size. The company receives up to 10,000 records every second. Data transmission delays of a few minutes are acceptable because of unreliable network conditions.
The transportation company wants to use Amazon Kinesis Data Streams to ingest the geolocation data. The company needs a reliable mechanism to send data to Kinesis Data Streams. The company needs to maximize the throughput efficiency of the Kinesis shards.
Which solution will meet these requirements in the MOST operationally efficient way?

- A. Kinesis Agent
- B. Amazon Data Firehose
- C. Kinesis Producer Library (KPL)
- D. Kinesis SDK

Answer: C

Explanation:
Problem Analysis:
The company ingests geolocation records (10 bytes each) at 10,000 records per second into Kinesis Data Streams.
Data transmission delays are acceptable, but the solution must maximize throughput efficiency.
Key Considerations:
The Kinesis Producer Library (KPL) batches records and uses aggregation to optimize shard throughput.
Efficiently handles high-throughput scenarios with minimal operational overhead.
Solution Analysis:
Option A: Kinesis Agent
Designed for file-based ingestion; not optimized for geolocation records.
Option B: KPL
Aggregates records into larger payloads, significantly improving shard throughput.
Suitable for applications generating small, high-frequency records.
Option C: Kinesis Firehose
Firehose is for delivery to destinations like S3 or Redshift and is not optimized for direct ingestion to Kinesis Data Streams.

Option D: Kinesis SDK
The SDK lacks advanced features like aggregation, resulting in lower throughput efficiency.
Final Recommendation:
Use Kinesis Producer Library (KPL) for its built-in aggregation and batching capabilities.
Kinesis Producer Library (KPL) Overview
Best Practices for Amazon Kinesis

NEW QUESTION # 73
An airline company is collecting metrics about flight activities for analytics. The company is conducting a proof of concept (POC) test to show how analytics can provide insights that the company can use to increase on-time departures.
The POC test uses objects in Amazon S3 that contain the metrics in .csv format. The POC test uses Amazon Athena to query the data. The data is partitioned in the S3 bucket by date.
As the amount of data increases, the company wants to optimize the storage solution to improve query performance.
Which combination of solutions will meet these requirements? (Choose two.)

- A. Add a randomized string to the beginning of the keys in Amazon S3 to get more throughput across partitions.
- B. Preprocess the .csv data to Apache Parquet format by fetching only the data blocks that are needed for predicates.
- C. Preprocess the .csv data to JSON format by fetching only the document keys that the query requires.
- D. Use an S3 bucket that is in the same account that uses Athena to query the data.
- E. Use an S3 bucket that is in the same AWS Region where the company runs Athena queries.

Answer: B,E

Explanation:
Using an S3 bucket that is in the same AWS Region where the company runs Athena queries can improve query performance by reducing data transfer latency and costs. Preprocessing the .csv data to Apache Parquet format can also improve query performance by enabling columnar storage, compression, and partitioning, which can reduce the amount of data scanned and fetched by the query. These solutions can optimize the storage solution for the POC test without requiring much effort or changes to the existing data pipeline. The other solutions are not optimal or relevant for this requirement. Adding a randomized string to the beginning of the keys in Amazon S3 can improve the throughput across partitions, but it can also make the data harder to query and manage. Using an S3 bucket that is in the same account that uses Athena to query the data does not have any significant impact on query performance, as long as the proper permissions are granted.
Preprocessing the .csv data to JSON format does not offer any benefits over the .csv format, as both are row- based and verbose formats that require more data scanning and fetching than columnar formats like Parquet.
References:
* Best Practices When Using Athena with AWS Glue
* Optimizing Amazon S3 Performance
* AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide

NEW QUESTION # 74
A company wants to analyze sales records that the company stores in a MySQL database. The company wants to correlate the records with sales opportunities identified by Salesforce.
The company receives 2 GB erf sales records every day. The company has 100 GB of identified sales opportunities. A data engineer needs to develop a process that will analyze and correlate sales records and sales opportunities. The process must run once each night.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use Amazon Kinesis Data Streams to fetch sales records from the MySQL database. Use Amazon Managed Service for Apache Flink to correlate the datasets. Use AWS Step Functions to orchestrate the process.
- B. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with the sales opportunities. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to orchestrate the process.
- C. Use Amazon Managed Workflows for Apache Airflow (Amazon MWAA) to fetch both datasets. Use AWS Lambda functions to correlate the datasets. Use AWS Step Functions to orchestrate the process.
- D. Use Amazon AppFlow to fetch sales opportunities from Salesforce. Use AWS Glue to fetch sales records from the MySQL database. Correlate the sales records with sales opportunities. Use AWS Step Functions to orchestrate the process.

Answer: D

Explanation:
* Problem Analysis:
* The company processes2 GB of daily sales recordsand100 GB of Salesforce sales opportunities
.
* The goal is to analyze and correlate the two datasets withlow operational overhead.
* The process must runonce nightly.
* Key Considerations:
* Amazon AppFlowsimplifies data integration with Salesforce.
* AWS Gluecan extract data from MySQL and perform ETL operations.
* Step Functionscan orchestrate workflows with minimal manual intervention.
* Apache Airflow and Flink add complexity, which conflicts with the requirement for low operational overhead.
* Solution Analysis:
* Option A: MWAA + Lambda + Step Functions
* Requires custom Lambda code for dataset correlation, increasing development and operational complexity.
* Option B: AppFlow + Glue + MWAA
* MWAA adds orchestration overhead compared to the simpler Step Functions.
* Option C: AppFlow + Glue + Step Functions
* AppFlow fetches Salesforce data, Glue extracts MySQL data, and Step Functions orchestrate the entire process.
* Minimal setup and operational overhead, making it the best choice.
* Option D: AppFlow + Kinesis + Flink + Step Functions
* Using Kinesis and Flink for batch processing introduces unnecessary complexity.
* Final Recommendation:
* UseAmazon AppFlowto fetch Salesforce data,AWS Glueto process MySQL data, andStep Functionsfor orchestration.
:
Amazon AppFlow Overview
AWS Glue ETL Documentation
AWS Step Functions

NEW QUESTION # 75
A healthcare company uses Amazon Kinesis Data Streams to stream real-time health data from wearable devices, hospital equipment, and patient records.
A data engineer needs to find a solution to process the streaming data. The data engineer needs to store the data in an Amazon Redshift Serverless warehouse. The solution must support near real-time analytics of the streaming data and the previous day's data.
Which solution will meet these requirements with the LEAST operational overhead?

- A. Load the data into Amazon S3. Use the COPY command to load the data into Amazon Redshift.
- B. Load data into Amazon Kinesis Data Firehose. Load the data into Amazon Redshift.
- C. Use the streaming ingestion feature of Amazon Redshift.
- D. Use the Amazon Aurora zero-ETL integration with Amazon Redshift.

Answer: C

Explanation:
The streaming ingestion feature of Amazon Redshift enables you to ingest data from streaming sources, such as Amazon Kinesis Data Streams, into Amazon Redshift tables in near real-time. You can use the streaming ingestion feature to process the streaming data from the wearable devices, hospital equipment, and patient records. The streaming ingestion feature also supports incremental updates, which means you can append new data or update existing data in the Amazon Redshift tables. This way, you can store the data in an Amazon Redshift Serverless warehouse and support near real-time analytics of the streaming data and the previous day's data. This solution meets the requirements with the least operational overhead, as it does not require any additional services or components to ingest and process the streaming data. The other options are either not feasible or not optimal. Loading data into Amazon Kinesis Data Firehose and then into Amazon Redshift (option A) would introduce additional latency and cost, as well as require additional configuration and management. Loading data into Amazon S3 and then using the COPY command to load the data into Amazon Redshift (option C) would also introduce additional latency and cost, as well as require additional storage space and ETL logic. Using the Amazon Aurora zero-ETL integration with Amazon Redshift (option D) would not work, as it requires the data to be stored in Amazon Aurora first, which is not the case for the streaming data from the healthcare company. References:
Using streaming ingestion with Amazon Redshift
AWS Certified Data Engineer - Associate DEA-C01 Complete Study Guide, Chapter 3: Data Ingestion and Transformation, Section 3.5: Amazon Redshift Streaming Ingestion

NEW QUESTION # 76
......

The Amazon Data-Engineer-Associate PDF format is printable which enables you to do paper study. It contains pool of actual and updated AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam questions. You can carry this portable file of Amazon Data-Engineer-Associate Real Questions to any place via smartphones, laptops, and tablets. This simple and convenient format of Prep4SureReview's AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) practice material is being updated regularly.

Data-Engineer-Associate Practice Guide: https://www.prep4surereview.com/Data-Engineer-Associate-latest-braindumps.html

Amazon Data-Engineer-Associate Real Question If the user discovers that the product is not appropriate for him, the user can choose another type of learning material, Amazon Data-Engineer-Associate Real Question The Questions & answers are verified and selected by professionals in the field and ensure accuracy and efficiency throughout the whole Product, Verified Data-Engineer-Associate dumps help you in preparation for the exam in a short period of time.

I mean, the previous director was paralyzed with mail, We got a lot of feedback Data-Engineer-Associate on this trend, If the user discovers that the product is not appropriate for him, the user can choose another type of learning material.

Real Amazon Data-Engineer-Associate Dumps PDF Format

The Questions & answers are verified and selected by professionals in the field and ensure accuracy and efficiency throughout the whole Product, Verified Data-Engineer-Associate Dumps help you in preparation for the exam in a short period of time.

The Data-Engineer-Associate exam dumps questions boost up your confidence level, That is the reason that I want to introduce you our Data-Engineer-Associate prep torrent.

- Data-Engineer-Associate Actual Test Answers Reliable Data-Engineer-Associate Exam Guide Vce Data-Engineer-Associate Files Search for ✔ Data-Engineer-Associate ✔ and easily obtain a free download on ▸ www.examcollectionpass.com ◂ Data-Engineer-Associate Sample Questions
- Free PDF 2026 Amazon Data-Engineer-Associate: Trustable AWS Certified Data Engineer - Associate (DEA-C01) Real Question Easily obtain ➡ Data-Engineer-Associate for free download through ▸ www.pdfvce.com ◂ Sample Data-Engineer-Associate Questions
- Free PDF Quiz 2026 Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Marvelous Real Question Copy URL 《 www.practicevce.com 》 open and search for ➽ Data-Engineer-Associate to download for free Data-Engineer-Associate Actual Test Answers
- 2026 Data-Engineer-Associate Real Question - AWS Certified Data Engineer - Associate (DEA-C01) Realistic Practice Guide Pass Guaranteed Open website 【 www.pdfvce.com 】 and search for ☀ Data-Engineer-Associate ☀ for free download Data-Engineer-Associate Exams
- Vce Data-Engineer-Associate Files Data-Engineer-Associate Reliable Test Pdf Test Data-Engineer-Associate Dates Immediately open www.testkingpass.com and search for Data-Engineer-Associate to obtain a free download Data-Engineer-Associate Exams
- Data-Engineer-Associate Reliable Test Pdf Data-Engineer-Associate Exam Passing Score Reliable Data-Engineer-Associate Exam Guide Immediately open ➤ www.pdfvce.com and search for " Data-Engineer-Associate " to obtain a free download Data-Engineer-Associate Latest Exam Notes
- Free PDF 2026 Amazon Fantastic Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) Real Question Simply search for ✔ Data-Engineer-Associate ✔ for free download on ▹ www.examdiscuss.com ◃ Data-Engineer-Associate Exam Passing Score
- Real Data-Engineer-Associate Exam Dumps Test Data-Engineer-Associate Dates Data-Engineer-Associate Exams Search for ➤ Data-Engineer-Associate and easily obtain a free download on 《 www.pdfvce.com 》 Data-Engineer-Associate Exam Passing Score
- 3 formats of updated www.examcollectionpass.com Amazon Data-Engineer-Associate Exam Questions Simply search for ▸ Data-Engineer-Associate ◂ for free download on ✔ www.examcollectionpass.com ✔ Real Data-Engineer-Associate Exam Dumps
- 2026 Data-Engineer-Associate Real Question - AWS Certified Data Engineer - Associate (DEA-C01) Realistic Practice Guide Pass Guaranteed Search for " Data-Engineer-Associate " and download it for free immediately on www.pdfvce.com Data-Engineer-Associate Latest Exam Notes
- Data-Engineer-Associate Reliable Test Pdf Reliable Data-Engineer-Associate Exam Voucher Data-Engineer-Associate Latest Exam Notes Search for ➡ Data-Engineer-Associate and download exam materials for free through （ www.pdfdumps.com ） New Data-Engineer-Associate Dumps Book