Top New Data-Engineer-Associate Exam Objectives | Reliable Amazon Data-Engineer-Associate: AWS Certified Data Engineer - Associate (DEA-C01) 100% Pass

2026 Latest PDFDumps Data-Engineer-Associate PDF Dumps and Data-Engineer-Associate Exam Engine Free Share: https://drive.google.com/open?id=1AgDea5WC6KP1NlQcav3AHJF5agy8cNM_

The Amazon Data-Engineer-Associate PDF dumps file is the most convenient way to prepare for the examination. This document is a collection of most probable and realistic AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate dumps. With this PDF file, you have AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate questions that will appear in the real exam. You can immediately download our Data-Engineer-Associate PDF Questions from the PDFDumps website after payment. Without place and time limits, you can use the PDF format of AWS Certified Data Engineer - Associate (DEA-C01) Data-Engineer-Associate real exam questions via smartphones, tablets, and laptops.

With our software version of our Data-Engineer-Associate guide braindumps, you can practice and test yourself just like you are in a real exam for our Data-Engineer-Associate study materials have the advandage of simulating the real exam. The results of your Data-Engineer-Associate Exam will be analyzed and a statistics will be presented to you. So you can see how you have done and know which kinds of questions of the Data-Engineer-Associate exam are to be learned more.

>> New Data-Engineer-Associate Exam Objectives <<

Data-Engineer-Associate Latest Exam Guide | Data-Engineer-Associate Valid Exam Pattern

You may strand on some issues at sometimes, all confusions will be answered by the bountiful contents of our Data-Engineer-Associate exam materials. Wrong choices may engender wrong feed-backs, we are sure you will come a long way by our Data-Engineer-Associate practice questions. In fact, a lot of our loyal customers have became our friends and only relay on our Data-Engineer-Associate study braindumps. As they always said that our Data-Engineer-Associate learning quiz is guaranteed to help them pass the exam.

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

NEW QUESTION # 205
A company uses an Amazon Redshift provisioned cluster as its database. The Redshift cluster has five reserved ra3.4xlarge nodes and uses key distribution.
A data engineer notices that one of the nodes frequently has a CPU load over 90%. SQL Queries that run on the node are queued. The other four nodes usually have a CPU load under 15% during daily operations.

The data engineer wants to maintain the current number of compute nodes. The data engineer also wants to balance the load more evenly across all five compute nodes.
Which solution will meet these requirements?

- A. Change the sort key to be the data column that is most often used in a WHERE clause of the SQL SELECT statement.
- B. Change the primary key to be the data column that is most often used in a WHERE clause of the SQL SELECT statement.
- C. Change the distribution key to the table column that has the largest dimension.
- D. Upgrade the reserved node from ra3.4xlarqe to ra3.16xlarqe.

Answer: C

Explanation:
Changing the distribution key to the table column that has the largest dimension will help to balance the load more evenly across all five compute nodes. The distribution key determines how the rows of a table are distributed among the slices of the cluster. If the distribution key is not chosen wisely, it can cause data skew, meaning some slices will have more data than others, resulting in uneven CPU load and query performance.
By choosing the table column that has the largest dimension, meaning the column that has the most distinct values, as the distribution key, the data engineer can ensure that the rows are distributed more uniformly across the slices, reducing data skew and improving query performance.
The other options are not solutions that will meet the requirements. Option A, changing the sort key to be the data column that is most often used in a WHERE clause of the SQL SELECT statement, will not affect the data distribution or the CPU load. The sort key determines the order in which the rows of a table are stored on disk, which can improve the performance of range-restricted queries, but not the load balancing. Option C, upgrading the reserved node from ra3.4xlarge to ra3.16xlarge, will not maintain the current number of compute nodes, as it will increase the cost and the capacity of the cluster. Option D, changing the primary key to be the data column that is most often used in a WHERE clause of the SQL SELECT statement, will not affect the data distribution or the CPU load either. The primary key is a constraint that enforces the uniqueness of the rows in a table, but it does not influence the data layout or the query optimization. References:
Choosing a data distribution style
Choosing a data sort key
Working with primary keys

NEW QUESTION # 206
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. Use an S3 bucket that is in the same AWS Region where the company runs Athena queries.
- B. Preprocess the .csv data to JSON format by fetching only the document keys that the query requires.
- C. Add a randomized string to the beginning of the keys in Amazon S3 to get more throughput across partitions.
- D. Use an S3 bucket that is in the same account that uses Athena to query the data.
- E. Preprocess the .csv data to Apache Parquet format by fetching only the data blocks that are needed for predicates.

Answer: A,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 # 207
A data engineer needs to join data from multiple sources to perform a one-time analysis job. The data is stored in Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3.
Which solution will meet this requirement MOST cost-effectively?

- A. Copy the data from DynamoDB, Amazon RDS, and Amazon Redshift into Amazon S3. Run Amazon Athena queries directly on the S3 files.
- B. Use Redshift Spectrum to query data from DynamoDB, Amazon RDS, and Amazon S3 directly from Redshift.
- C. Use an Amazon EMR provisioned cluster to read from all sources. Use Apache Spark to join the data and perform the analysis.
- D. Use Amazon Athena Federated Query to join the data from all data sources.

Answer: D

Explanation:
Amazon Athena Federated Query is a feature that allows you to query data from multiple sources using standard SQL. You can use Athena Federated Query to join data from Amazon DynamoDB, Amazon RDS, Amazon Redshift, and Amazon S3, as well as other data sources such as MongoDB, Apache HBase, and Apache Kafka1. Athena Federated Query is a serverless and interactive service, meaning you do not need to provision or manage any infrastructure, and you only pay for the amount of data scanned by your queries.
Athena Federated Query is the most cost-effective solution for performing a one-time analysis job on data from multiple sources, as it eliminates the need to copy or move data, and allows you to query data directly from the source.
The other options are not as cost-effective as Athena Federated Query, as they involve additional steps or costs. Option A requires you to provision and pay for an Amazon EMR cluster, which can be expensive and time-consuming for a one-time job. Option B requires you to copy or move data from DynamoDB, RDS, and Redshift to S3, which can incur additional costs for data transfer and storage, and also introduce latency and complexity. Option D requires you to have an existing Redshift cluster, which can be costly and may not be necessary for a one-time job. Option D also does not support querying data from RDS directly, so you would need to use Redshift Federated Query to access RDS data, which adds another layer of complexity2. References:
Amazon Athena Federated Query
Redshift Spectrum vs Federated Query

NEW QUESTION # 208
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. Reduce the scaling cooldown period for the provisioned EMR cluster.
- B. Change the task node type from general purpose EC2 instances to memory optimized EC2 instances.
- 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.

Reference:
Managed Scaling: The EMR cluster's scaling is currently managed between 1 and 5 nodes, so changing the instance type will leverage the current scaling strategy but optimize it for the workload.
Alternatives Considered:
A (Increase task nodes to 10): Increasing the number of task nodes would increase costs without necessarily improving performance. Since memory usage is low, the bottleneck is more likely the CPU, which compute-optimized instances can handle better.
B (Memory optimized instances): Memory-optimized instances are not suitable since the current job is CPU-bound, and memory usage remains low (under 30%).
D (Reduce scaling cooldown): This could marginally improve scaling speed but does not address the need for cost optimization and improved CPU performance.
Amazon EMR Cluster Optimization
Compute Optimized EC2 Instances

NEW QUESTION # 209
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 a new AWS Glue workload 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 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.
- D. 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.

Answer: C

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 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 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 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 # 210

......

give you the most professional guidance.

Data-Engineer-Associate Latest Exam Guide: https://www.pdfdumps.com/Data-Engineer-Associate-valid-exam.html

Outside, you can use the APP version of our Data-Engineer-Associate study materials, Amazon New Data-Engineer-Associate Exam Objectives These test questions provide you with the experience of taking the actual test, Why Choose PDFDumps Data-Engineer-Associate Braindumps, It is well known that Data-Engineer-Associate Latest Exam Guide - AWS Certified Data Engineer - Associate (DEA-C01) exam is an international recognition certification, which is very important for people who are engaged in the related field, If you are still upset about how to pass exam with passing marks, come here and let us help you, choosing our Data-Engineer-Associate exam cram will be the first step to success of your career.

All Data-Engineer-Associate exam torrent does a lot of help for you to pass the Data-Engineer-Associate exam easily and successfully, This is another example of tools that are expected in non-web contexts, but not always available in the web world.

Download Free Updated PDFDumps Amazon Data-Engineer-Associate Exam Questions after Paying Affordable Charges

Outside, you can use the APP version of our Data-Engineer-Associate Study Materials, These test questions provide you with the experience of taking the actual test, Why Choose PDFDumps Data-Engineer-Associate Braindumps.

It is well known that AWS Certified Data Engineer - Associate (DEA-C01) exam is an international Data-Engineer-Associate recognition certification, which is very important for people who are engaged in the related field, If you arestill upset about how to pass exam with passing marks, come here and let us help you, choosing our Data-Engineer-Associate exam cram will be the first step to success of your career.

- Data-Engineer-Associate Exam Overview □ Data-Engineer-Associate Practice Exam Online □ Data-Engineer-Associate Exam Overview □ Immediately open ➠ www.validtorrent.com □ and search for " Data-Engineer-Associate " to obtain a free download □Minimum Data-Engineer-Associate Pass Score
- Valid Data-Engineer-Associate Exam Camp Pdf □ Data-Engineer-Associate Exam Flashcards □ Data-Engineer-Associate Latest Test Fee □ Search for [Data-Engineer-Associate] and download it for free immediately on 「 www.pdfvce.com 」 □Data-Engineer-Associate Hot Spot Questions
- Data-Engineer-Associate Latest Test Fee □ Minimum Data-Engineer-Associate Pass Score □ Data-Engineer-Associate Valid Test Voucher □ Easily obtain free download of ➻ Data-Engineer-Associate □ by searching on 《 www.dumpsquestion.com 》 □New Data-Engineer-Associate Dumps Ebook
- Valid Dumps Data-Engineer-Associate Book □ Exam Data-Engineer-Associate Simulator □ Data-Engineer-Associate Exam Overview □ Download ▸ Data-Engineer-Associate ◂ for free by simply searching on ➥ www.pdfvce.com □ □ □Vce Data-Engineer-Associate Download
- Perfect New Data-Engineer-Associate Exam Objectives, Ensure to pass the Data-Engineer-Associate Exam □ Search for [Data-Engineer-Associate] and download it for free on ☀ www.testkingpass.com □☀□ website □Data-Engineer-Associate Practice Exam Online
- Valid Amazon New Data-Engineer-Associate Exam Objectives Seriously Researched by Amazon Hard-working Trainers □ □ Search on 「 www.pdfvce.com 」 for ▸ Data-Engineer-Associate ◂ to obtain exam materials for free download □Data-Engineer-Associate Exam Questions Pdf
- Data-Engineer-Associate Exam Questions Pdf □ Data-Engineer-Associate Exam Questions Pdf □ Vce Data-Engineer-Associate Download □ Open （ www.prep4away.com ） enter [Data-Engineer-Associate] and obtain a free download □Test Data-Engineer-Associate Questions Answers
- Data-Engineer-Associate Hot Spot Questions □ Valid Data-Engineer-Associate Vce □ Data-Engineer-Associate Exam Overview □ Easily obtain ☀ Data-Engineer-Associate □☀□ for free download through ➻ www.pdfvce.com □ ✌New Data-Engineer-Associate Dumps Ebook
- Valid Data-Engineer-Associate Vce □ Vce Data-Engineer-Associate Download □ Vce Data-Engineer-Associate Download □ Immediately open " www.dumpsquestion.com " and search for [Data-Engineer-Associate] to obtain a free download □Test Data-Engineer-Associate Questions Answers
- Data-Engineer-Associate Exam Overview □ Data-Engineer-Associate Practice Exam Online □ Data-Engineer-Associate Exam Questions Pdf □ Download ➠ Data-Engineer-Associate □ for free by simply searching on □ www.pdfvce.com □ □Exam Data-Engineer-Associate Topic
- Pass Guaranteed Quiz 2026 Amazon Useful Data-Engineer-Associate: New AWS Certified Data Engineer - Associate (DEA-C01) Exam Objectives □ The page for free download of ➻ Data-Engineer-Associate □ on 《 www.troytecdumps.com 》 will open immediately □Exam Data-Engineer-Associate Simulator
- classifylist.com, jasontnya524434.westexwiki.com, bookmarkrange.com, tayacsxq633498.levitra-wiki.com, 8090.hhh1234.com, kiananile175841.myparisblog.com, delilahscgw150087.wikidirective.com, allenrrdu952090.blogsuperapp.com, justpaste.me, bookmarkgenius.com, Disposable vapes