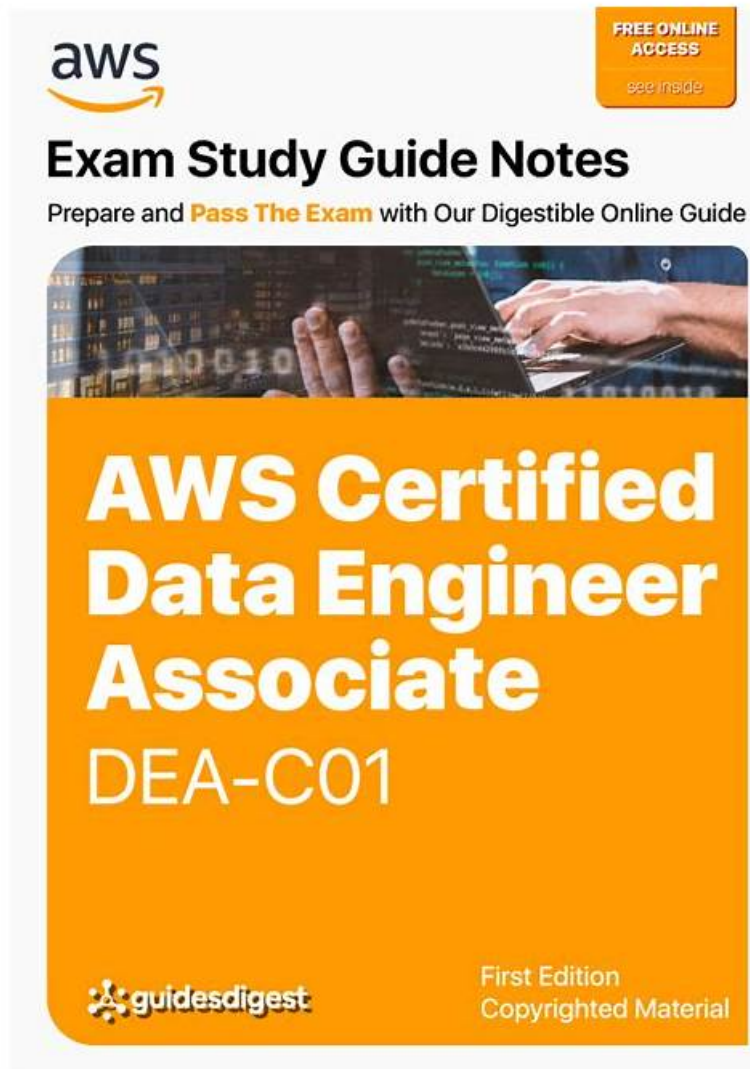


Pass Guaranteed Amazon - Data-Engineer-Associate - Authoritative AWS Certified Data Engineer - Associate (DEA-C01) New Study Plan



P.S. Free 2026 Amazon Data-Engineer-Associate dumps are available on Google Drive shared by PassReview:
https://drive.google.com/open?id=1NLSzdQcsy7Zod3195YCT0E_WEd2U_j1b

Are you an aspiring Amazon professional looking to pass the AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam? Look no further than our platform for real Data-Engineer-Associate exam dumps. Many candidates struggle to find reliable study materials, leading them to prepare with outdated material and ultimately waste their resources. But with our platform, you can access updated Amazon Data-Engineer-Associate Practice Questions and pass the certification test on your first try. Don't let a lack of credible study materials hold you back - trust our platform to help you achieve your career goals.

To do this you just need to pass the Amazon Data-Engineer-Associate certification exam. Are you ready to accept this challenge? Looking for the proven and easiest way to crack the Amazon Data-Engineer-Associate certification exam? If your answer is yes then you do not need to go anywhere. Just download Data-Engineer-Associate exam practice questions and start AWS Certified Data Engineer - Associate (DEA-C01) (Data-Engineer-Associate) exam preparation without wasting further time. The PassReview Amazon Data-Engineer-Associate Dumps will provide you with everything that you need to learn, prepare and pass the challenging Data-Engineer-Associate exam with flying colors. You must try PassReview Amazon Data-Engineer-Associate exam questions today.

Data-Engineer-Associate Reliable Exam Vce, Test Data-Engineer-Associate King

The chance to examine the content of the Data-Engineer-Associate practice material before purchasing it will give you peace of mind. So, try a free demo to evaluate the authenticity of the Amazon Data-Engineer-Associate Exam product. PassReview forewarns you that the topics of the Amazon Data-Engineer-Associate test change from time to time.

Amazon AWS Certified Data Engineer - Associate (DEA-C01) Sample Questions (Q64-Q69):

NEW QUESTION # 64

A company has a data warehouse in Amazon Redshift. To comply with security regulations, the company needs to log and store all user activities and connection activities for the data warehouse.

Which solution will meet these requirements?

- A. Create an Amazon Elastic File System (Amazon EFS) file system. Enable logging for the Amazon Redshift cluster. Write logs to the EFS file system.
- **B. Create an Amazon S3 bucket. Enable logging for the Amazon Redshift cluster. Specify the S3 bucket in the logging configuration to store the logs.**
- C. Create an Amazon Elastic Block Store (Amazon EBS) volume. Enable logging for the Amazon Redshift cluster. Write the logs to the EBS volume.
- D. Create an Amazon Aurora MySQL database. Enable logging for the Amazon Redshift cluster. Write the logs to a table in the Aurora MySQL database.

Answer: B

Explanation:

Problem Analysis:

The company must log all user activities and connection activities in Amazon Redshift for security compliance.

Key Considerations:

Redshift supports audit logging, which can be configured to write logs to an S3 bucket.

S3 provides durable, scalable, and cost-effective storage for logs.

Solution Analysis:

Option A: S3 for Logging

Standard approach for storing Redshift logs.

Easy to set up and manage with minimal cost.

Option B: Amazon EFS

EFS is unnecessary for this use case and less cost-efficient than S3.

Option C: Aurora MySQL

Using a database to store logs increases complexity and cost.

Option D: EBS Volume

EBS is not a scalable option for log storage compared to S3.

Final Recommendation:

Enable Redshift audit logging and specify an S3 bucket as the destination.

Reference:

Amazon Redshift Audit Logging

Storing Logs in Amazon S3

NEW QUESTION # 65

A company is migrating its database servers from Amazon EC2 instances that run Microsoft SQL Server to Amazon RDS for Microsoft SQL Server DB instances. The company's analytics team must export large data elements every day until the migration is complete. The data elements are the result of SQL joins across multiple tables. The data must be in Apache Parquet format. The analytics team must store the data in Amazon S3.

Which solution will meet these requirements in the MOST operationally efficient way?

- A. Create an AWS Lambda function that queries the EC2 instance-based databases by using Java Database Connectivity

(JDBC). Configure the Lambda function to retrieve the required data, transform the data into Parquet format, and transfer the data into an S3 bucket. Use Amazon EventBridge to schedule the Lambda function to run every day.

- **B. Create a view in the EC2 instance-based SQL Server databases that contains the required data elements. Create an AWS Glue job that selects the data directly from the view and transfers the data in Parquet format to an S3 bucket. Schedule the AWS Glue job to run every day.**
- C. Use a SQL query to create a view in the EC2 instance-based SQL Server databases that contains the required data elements. Create and run an AWS Glue crawler to read the view. Create an AWS Glue job that retrieves the data and transfers the data in Parquet format to an S3 bucket. Schedule the AWS Glue job to run every day.
- D. Schedule SQL Server Agent to run a daily SQL query that selects the desired data elements from the EC2 instance-based SQL Server databases. Configure the query to direct the output .csv objects to an S3 bucket. Create an S3 event that invokes an AWS Lambda function to transform the output format from .csv to Parquet.

Answer: B

Explanation:

Option A is the most operationally efficient way to meet the requirements because it minimizes the number of steps and services involved in the data export process. AWS Glue is a fully managed service that can extract, transform, and load (ETL) data from various sources to various destinations, including Amazon S3. AWS Glue can also convert data to different formats, such as Parquet, which is a columnar storage format that is optimized for analytics. By creating a view in the SQL Server databases that contains the required data elements, the AWS Glue job can select the data directly from the view without having to perform any joins or transformations on the source data. The AWS Glue job can then transfer the data in Parquet format to an S3 bucket and run on a daily schedule.

Option B is not operationally efficient because it involves multiple steps and services to export the data. SQL Server Agent is a tool that can run scheduled tasks on SQL Server databases, such as executing SQL queries. However, SQL Server Agent cannot directly export data to S3, so the query output must be saved as .csv objects on the EC2 instance. Then, an S3 event must be configured to trigger an AWS Lambda function that can transform the .csv objects to Parquet format and upload them to S3. This option adds complexity and latency to the data export process and requires additional resources and configuration.

Option C is not operationally efficient because it introduces an unnecessary step of running an AWS Glue crawler to read the view. 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. However, in this scenario, the schema and format of the data elements are already known and fixed, so there is no need to run a crawler to discover them. The AWS Glue job can directly select the data from the view without using the Data Catalog. Running a crawler adds extra time and cost to the data export process.

Option D is not operationally efficient because it requires custom code and configuration to query the databases and transform the data. An AWS Lambda function is a service that can run code in response to events or triggers, such as Amazon EventBridge. Amazon EventBridge is a service that can connect applications and services with event sources, such as schedules, and route them to targets, such as Lambda functions. However, in this scenario, using a Lambda function to query the databases and transform the data is not the best option because it requires writing and maintaining code that uses JDBC to connect to the SQL Server databases, retrieve the required data, convert the data to Parquet format, and transfer the data to S3. This option also has limitations on the execution time, memory, and concurrency of the Lambda function, which may affect the performance and reliability of the data export process.

Reference:

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

AWS Glue Documentation

Working with Views in AWS Glue

Converting to Columnar Formats

NEW QUESTION # 66

A company is planning to use a provisioned Amazon EMR cluster that runs Apache Spark jobs to perform big data analysis. The company requires high reliability. A big data team must follow best practices for running cost-optimized and long-running workloads on Amazon EMR. The team must find a solution that will maintain the company's current level of performance.

Which combination of resources will meet these requirements MOST cost-effectively? (Choose two.)

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

Answer: B,C

Explanation:

The best combination of resources to meet the requirements of high reliability, cost-optimization, and performance for running Apache Spark jobs on Amazon EMR is to use Amazon S3 as a persistent data store and Graviton instances for core nodes and task nodes.

Amazon S3 is a highly durable, scalable, and secure object storage service that can store any amount of data for a variety of use cases, including big data analytics¹. Amazon S3 is a better choice than HDFS as a persistent data store for Amazon EMR, as it decouples the storage from the compute layer, allowing for more flexibility and cost-efficiency. Amazon S3 also supports data encryption, versioning, lifecycle management, and cross-region replication¹. Amazon EMR integrates seamlessly with Amazon S3, using EMR File System (EMRFS) to access data stored in Amazon S3 buckets². EMRFS also supports consistent view, which enables Amazon EMR to provide read-after-write consistency for Amazon S3 objects that are accessed through EMRFS².

Graviton instances are powered by Arm-based AWS Graviton² processors that deliver up to 40% better price performance over comparable current generation x86-based instances³. Graviton instances are ideal for running workloads that are CPU-bound, memory-bound, or network-bound, such as big data analytics, web servers, and open-source databases³. Graviton instances are compatible with Amazon EMR, and can be used for both core nodes and task nodes. Core nodes are responsible for running the data processing frameworks, such as Apache Spark, and storing data in HDFS or the local file system. Task nodes are optional nodes that can be added to a cluster to increase the processing power and throughput. By using Graviton instances for both core nodes and task nodes, you can achieve higher performance and lower cost than using x86-based instances.

Using Spot Instances for all primary nodes is not a good option, as it can compromise the reliability and availability of the cluster. Spot Instances are spare EC2 instances that are available at up to 90% discount compared to On-Demand prices, but they can be interrupted by EC2 with a two-minute notice when EC2 needs the capacity back. Primary nodes are the nodes that run the cluster software, such as Hadoop, Spark, Hive, and Hue, and are essential for the cluster operation. If a primary node is interrupted by EC2, the cluster will fail or become unstable. Therefore, it is recommended to use On-Demand Instances or Reserved Instances for primary nodes, and use Spot Instances only for task nodes that can tolerate interruptions. Reference:

Amazon S3 - Cloud Object Storage

EMR File System (EMRFS)

AWS Graviton² Processor-Powered Amazon EC2 Instances

[Plan and Configure EC2 Instances]

[Amazon EC2 Spot Instances]

[Best Practices for Amazon EMR]

NEW QUESTION # 67

A company currently stores all of its data in Amazon S3 by using the S3 Standard storage class.

A data engineer examined data access patterns to identify trends. During the first 6 months, most data files are accessed several times each day. Between 6 months and 2 years, most data files are accessed once or twice each month. After 2 years, data files are accessed only once or twice each year.

The data engineer needs to use an S3 Lifecycle policy to develop new data storage rules. The new storage solution must continue to provide high availability.

Which solution will meet these requirements in the MOST cost-effective way?

- A. Transition objects to S3 One Zone-Infrequent Access (S3 One Zone-IA) after 6 months. Transfer objects to S3 Glacier Deep Archive after 2 years.
- **B. Transition objects to S3 Standard-Infrequent Access (S3 Standard-IA) after 6 months. Transfer objects to S3 Glacier Deep Archive after 2 years.**
- C. Transition objects to S3 One Zone-Infrequent Access (S3 One Zone-IA) after 6 months. Transfer objects to S3 Glacier Flexible Retrieval after 2 years.
- D. Transition objects to S3 Standard-Infrequent Access (S3 Standard-IA) after 6 months. Transfer objects to S3 Glacier Flexible Retrieval after 2 years.

Answer: B

Explanation:

To achieve the most cost-effective storage solution, the data engineer needs to use an S3 Lifecycle policy that transitions objects to lower-cost storage classes based on their access patterns, and deletes them when they are no longer needed. The storage classes should also provide high availability, which means they should be resilient to the loss of data in a single Availability Zone¹. Therefore, the solution must include the following steps:

* Transition objects to S3 Standard-Infrequent Access (S3 Standard-IA) after 6 months. S3 Standard-IA is designed for data that is accessed less frequently, but requires rapid access when needed. It offers the same high durability, throughput, and low latency as S3 Standard, but with a lower storage cost and a retrieval fee². Therefore, it is suitable for data files that are accessed once or twice each month. S3 Standard-IA also provides high availability, as it stores data redundantly across multiple Availability Zones¹.

* Transfer objects to S3 Glacier Deep Archive after 2 years. S3 Glacier Deep Archive is the lowest-cost storage class that offers

secure and durable storage for data that is rarely accessed and can tolerate a 12- hour retrieval time. It is ideal for long-term archiving and digital preservation³. Therefore, it is suitable for data files that are accessed only once or twice each year. S3 Glacier Deep Archive also provides high availability, as it stores data across at least three geographically dispersed Availability Zones¹.

* Delete objects when they are no longer needed. The data engineer can specify an expiration action in the S3 Lifecycle policy to delete objects after a certain period of time. This will reduce the storage cost and comply with any data retention policies.

Option C is the only solution that includes all these steps. Therefore, option C is the correct answer.

Option A is incorrect because it transitions objects to S3 One Zone-Infrequent Access (S3 One Zone-IA) after 6 months. S3 One Zone-IA is similar to S3 Standard-IA, but it stores data in a single Availability Zone. This means it has a lower availability and durability than S3 Standard-IA, and it is not resilient to the loss of data in a single Availability Zone¹. Therefore, it does not provide high availability as required.

Option B is incorrect because it transfers objects to S3 Glacier Flexible Retrieval after 2 years. S3 Glacier Flexible Retrieval is a storage class that offers secure and durable storage for data that is accessed infrequently and can tolerate a retrieval time of minutes to hours. It is more expensive than S3 Glacier Deep Archive, and it is not suitable for data that is accessed only once or twice each year³. Therefore, it is not the most cost-effective option.

Option D is incorrect because it combines the errors of option A and B. It transitions objects to S3 One Zone- IA after 6 months, which does not provide high availability, and it transfers objects to S3 Glacier Flexible Retrieval after 2 years, which is not the most cost-effective option.

References:

* 1: Amazon S3 storage classes - Amazon Simple Storage Service

* 2: Amazon S3 Standard-Infrequent Access (S3 Standard-IA) - Amazon Simple Storage Service

* 3: Amazon S3 Glacier and S3 Glacier Deep Archive - Amazon Simple Storage Service

* [4]: Expiring objects - Amazon Simple Storage Service

* [5]: Managing your storage lifecycle - Amazon Simple Storage Service

* [6]: Examples of S3 Lifecycle configuration - Amazon Simple Storage Service

* [7]: Amazon S3 Lifecycle further optimizes storage cost savings with new features - What's New with AWS

NEW QUESTION # 68

A data engineering team is using an Amazon Redshift data warehouse for operational reporting. The team wants to prevent performance issues that might result from long- running queries. A data engineer must choose a system table in Amazon Redshift to record anomalies when a query optimizer identifies conditions that might indicate performance issues.

Which table views should the data engineer use to meet this requirement?

- A. STL USAGE CONTROL
- B. STL PLAN INFO
- C. STL QUERY METRICS
- **D. STL ALERT EVENT LOG**

Answer: D

Explanation:

The STL ALERT EVENT LOG table view records anomalies when the query optimizer identifies conditions that might indicate performance issues. These conditions include skewed data distribution, missing statistics, nested loop joins, and broadcasted data. The STL ALERT EVENT LOG table view can help the data engineer to identify and troubleshoot the root causes of performance issues and optimize the query execution plan. The other table views are not relevant for this requirement. STL USAGE CONTROL records the usage limits and quotas for Amazon Redshift resources. STL QUERY METRICS records the execution time and resource consumption of queries. STL PLAN INFO records the query execution plan and the steps involved in each query.

References:

STL ALERT EVENT LOG

System Tables and Views

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

NEW QUESTION # 69

.....

It is known to us that the Data-Engineer-Associate exam has been increasingly significant for modern people in this highly competitive world, because the test certification can certify whether you have the competitive advantage in the global labor market or have the ability to handle the job in a certain area, especial when we enter into a newly computer era. Therefore our Data-Engineer-Associate practice torrent is tailor-designed for these learning groups, thus helping them pass the exam in a more productive and

www.xunshuzhilian.com, www.stes.tyc.edu.tw, www.stes.tyc.edu.tw, Disposable vapes

BONUS!!! Download part of PassReview Data-Engineer-Associate dumps for free: https://drive.google.com/open?id=1NLSzdQcsy7Zod3195YCT0E_WEd2U_j1b