DAS-C01 Q&As

AWS Certified Data Analytics - Specialty (DAS-C01)

Pass Amazon DAS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.geekcert.com/das-c01.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Amazon Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

An online retail company uses Amazon Redshift to store historical sales transactions. The company is required to encrypt data at rest in the clusters to comply with the Payment Card Industry Data Security Standard (PCI DSS). A corporate governance policy mandates management of encryption keys using an on-premises hardware security module (HSM).

Which solution meets these requirements?

A. Create and manage encryption keys using AWS CloudHSM Classic. Launch an Amazon Redshift cluster in a VPC with the option to use CloudHSM Classic for key management.

B. Create a VPC and establish a VPN connection between the VPC and the on-premises network. Create an HSM connection and client certificate for the on-premises HSM. Launch a cluster in the VPC with the option to use the on-premises HSM to store keys.

C. Create an HSM connection and client certificate for the on-premises HSM. Enable HSM encryption on the existing unencrypted cluster by modifying the cluster. Connect to the VPC where the Amazon Redshift cluster resides from the on-premises network using a VPN.

D. Create a replica of the on-premises HSM in AWS CloudHSM. Launch a cluster in a VPC with the option to use CloudHSM to store keys.

Correct Answer: B

Reference: https://docs.aws.amazon.com/redshift/latest/mgmt/security-key-management.html

QUESTION 2

A gaming company is collecting cllckstream data into multiple Amazon Kinesis data streams. The company uses Amazon Kinesis Data Firehose delivery streams to store the data in JSON format in Amazon S3 Data scientists use Amazon Athena to query the most recent data and derive business insights. The company wants to reduce its Athena costs without having to recreate the data pipeline. The company prefers a solution that will require less management effort.

Which set of actions can the data scientists take immediately to reduce costs?

A. Change the Kinesis Data Firehose output format to Apache Parquet Provide a custom S3 object YYYYMMDD prefix expression and specify a large buffer size For the existing data, run an AWS Glue ETL job to combine and convert small JSON files to large Parquet files and add the YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table.

B. Create an Apache Spark Job that combines and converts JSON files to Apache Parquet files Launch an Amazon EMR ephemeral cluster daily to run the Spark job to create new Parquet files in a different S3 location Use ALTER TABLE SET LOCATION to reflect the new S3 location on the existing Athena table.

C. Create a Kinesis data stream as a delivery target for Kinesis Data Firehose Run Apache Flink on Amazon Kinesis Data Analytics on the stream to read the streaming data, aggregate ikand save it to Amazon S3 in Apache Parquet format with a custom S3 object YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table

D. Integrate an AWS Lambda function with Kinesis Data Firehose to convert source records to Apache Parquet and write them to Amazon S3 In parallel, run an AWS Glue ETL job to combine and convert existing JSON files to large

Parquet

files Create a custom S3 object YYYYMMDD prefix Use ALTER TABLE ADD PARTITION to reflect the partition on the existing Athena table.

Correct Answer: D

Reference: https://docs.aws.amazon.com/firehose/latest/dev/record-format-conversion.html
https://docs.aws.amazon.com/firehose/latest/dev/data-analysis.html

QUESTION 3

An IoT company is collecting data from multiple sensors and is streaming the data to Amazon Managed Streaming for Apache Kafka (Amazon MSK). Each sensor type has its own topic, and each topic has the same number of partitions.

The company is planning to turn on more sensors. However, the company wants to evaluate which sensor types are producing the most data so that the company can scale accordingly. The company needs to know which sensor types have

the largest values for the following metrics: BytesInPerSec and MessagesInPerSec.

Which level of monitoring for Amazon MSK will meet these requirements?

A. DEFAULT level

B. PER_TOPIC_PER_BROKER level

C. PER_BROKER level

D. PER_TOPIC level

Correct Answer: D

QUESTION 4

A reseller that has thousands of AWS accounts receives AWS Cost and Usage Reports in an Amazon S3 bucket. The reports are delivered to the S3 bucket in the following format: //yyyymmdd-yyyymmdd/.parquet An AWS Glue crawler crawls the S3 bucket and populates an AWS Glue Data Catalog with a table. Business analysts use Amazon Athena to query the table and create monthly summary reports for the AWS accounts. The business analysts

are experiencing slow queries because of the accumulation of reports from the last 5 years. The business analysts want the operations team to make changes to improve query performance. Which action should the operations team take to meet these requirements?

A. Change the file format to .csv.zip

B. Partition the data by date and account ID

C. Partition the data by month and account ID

D. Partition the data by account ID, year, and month

Correct Answer: A

Reference: https://docs.aws.amazon.com/cur/latest/userguide/access-cur-s3.html

QUESTION 5

An education provider\\'s learning management system (LMS) is hosted in a 100 TB data lake that is built on Amazon S3. The provider\\'s LMS supports hundreds of schools. The provider wants to build an advanced analytics reporting platform using Amazon Redshift to handle complex queries with optimal performance. System users will query the most recent 4 months of data 95% of the time while 5% of the queries will leverage data from the previous 12 months.

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

A. Store the most recent 4 months of data in the Amazon Redshift cluster. Use Amazon Redshift Spectrum to query data in the data lake. Use S3 lifecycle management rules to store data from the previous 12 months in Amazon S3 Glacier storage.

B. Leverage DS2 nodes for the Amazon Redshift cluster. Migrate all data from Amazon S3 to Amazon Redshift. Decommission the data lake.

C. Store the most recent 4 months of data in the Amazon Redshift cluster. Use Amazon Redshift Spectrum to query data in the data lake. Ensure the S3 Standard storage class is in use with objects in the data lake.

D. Store the most recent 4 months of data in the Amazon Redshift cluster. Use Amazon Redshift federated queries to join cluster data with the data lake to reduce costs. Ensure the S3 Standard storage class is in use with objects in the data lake.

Correct Answer: C

Reference: https://aws.amazon.com/redshift/pricing/