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

QUESTION 1

A company uses Amazon Redshift to store historical sales transactions. The company must encrypt data at rest in the Redshift cluster. The company also must store encryption keys by using an on-premises hardware security module (HSM). Which solution will meet these requirements with the LEAST operational overhead?

A. Create and store encryption keys by using AWS CloudHSM Classic. Launch a new Redshift cluster with the option to use CloudHSM Classic to store keys.

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

C. Establish an AWS Site-to-Site VPN connection between the existing VPC and the on-premises network. Create an HSM connection and a client certificate for the on-premises HSM. Configure the existing Redshift cluster in the VPC with the option to use the on-premises HSM to store keys. Reboot the Redshift cluster.

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

Correct Answer: B

QUESTION 2

A company has an application that uses the Amazon Kinesis Client Library (KCL) to read records from a Kinesis data stream.

After a successful marketing campaign, the application experienced a significant increase in usage. As a result, a data analyst had to split some shards in the data stream. When the shards were split, the application started throwing an

ExpiredIteratorExceptions error sporadically.

What should the data analyst do to resolve this?

A. Increase the number of threads that process the stream records.

B. Increase the provisioned read capacity units assigned to the stream\\'s Amazon DynamoDB table.

C. Increase the provisioned write capacity units assigned to the stream\\'s Amazon DynamoDB table.

D. Decrease the provisioned write capacity units assigned to the stream\\'s Amazon DynamoDB table.

Correct Answer: C

Reference: https://docs.aws.amazon.com/streams/latest/dev/troubleshooting-consumers.html#shard-iterator-expires-unexpectedly

QUESTION 3

A large company receives files from external parties in Amazon EC2 throughout the day. At the end of the day, the files are combined into a single file, compressed into a gzip file, and uploaded to Amazon S3. The total size of all the files is close to 100 GB daily. Once the files are uploaded to Amazon S3, an AWS Batch program executes a COPY command to load the files into an Amazon Redshift cluster.

Which program modification will accelerate the COPY process?

A. Upload the individual files to Amazon S3 and run the COPY command as soon as the files become available.

B. Split the number of files so they are equal to a multiple of the number of slices in the Amazon Redshift cluster. Gzip and upload the files to Amazon S3. Run the COPY command on the files.

C. Split the number of files so they are equal to a multiple of the number of compute nodes in the Amazon Redshift cluster. Gzip and upload the files to Amazon S3. Run the COPY command on the files.

D. Apply sharding by breaking up the files so the distkey columns with the same values go to the same file. Gzip and upload the sharded files to Amazon S3. Run the COPY command on the files.

Correct Answer: B

Reference: https://docs.aws.amazon.com/redshift/latest/dg/t_splitting-data-files.html

QUESTION 4

An online retail company stores Application Load Balancer (ALB) access logs in an Amazon S3 bucket. The analytics team wants to query the logs by using Amazon Athena to analyze traffic distribution and patterns. An unpartitioned table has been created in Athena. The most common query obtains log information by year, month, and day. However, as the size of the data increases, the response time for the same query increases.

Which solution should improve the query performance in Athena with the LEAST operational effort?

A. Create an AWS Glue job with a transform that infers the schema of all ALB access logs and populates the partition metadata by year, month, and day to the AWS Glue Data Catalog.

B. Create an AWS Glue crawler with a classifier that infers the schema of all ALB access logs and populates the partition metadata by year, month, and day to the AWS Glue Data Catalog.

C. Create an AWS Lambda function to transform all ALB access logs. Save the results to Amazon S3 in Apache Parquet format and partition the results by year, month, and day. Use Athena to query the transformed data.

D. Use AWS Data Pipeline to periodically start a transient Amazon EMR cluster with a Spark step. Use PySpark to create files in Amazon S3 and partition the files by year, month, and day.

Correct Answer: D

QUESTION 5

A marketing company has an application that stores event data in an Amazon RDS database. The company is replicating this data to Amazon Redshift for reporting and business intelligence (BI) purposes. New event data is continuously generated and ingested into the RDS database throughout the day and captured by a change data capture (CDC) replication task in AWS Database Migration Service (AWS DMS). The company requires that the new data be

replicated to Amazon Redshift in near-real time.

Which solution meets these requirements?

A. Use Amazon Kinesis Data Streams as the destination of the CDC replication task in AWS DMS. Use an AWS Glue streaming job to read changed records from Kinesis Data Streams and perform an upsert into the Redshift cluster.

B. Use Amazon S3 as the destination of the CDC replication task in AWS DMS. Use the COPY command to load data into the Redshift cluster.

C. Use Amazon DynamoDB as the destination of the CDC replication task in AWS DMS. Use the COPY command to load data into the Redshift cluster.

D. Use Amazon Kinesis Data Firehose as the destination of the CDC replication task in AWS DMS. Use an AWS Glue streaming job to read changed records from Kinesis Data Firehose and perform an upsert into the Redshift cluster.

Correct Answer: D