



Free Questions for **DAS-C01**

Shared by **Mills** on **22-07-2024**

For More Free Questions and Preparation Resources

[Check the Links on Last Page](#)



Question 1

Question Type: MultipleChoice

A gaming company is collecting clickstream 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?

Options:

- 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 it, and 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.

Answer:

D

Question 2

Question Type: MultipleChoice

A company needs to collect streaming data from several sources and store the data in the AWS Cloud. The dataset is heavily structured, but analysts need to perform several complex SQL queries and need consistent performance. Some of the data is queried more frequently than the

rest. The company wants a solution that meets its performance requirements in a cost-effective manner.

Which solution meets these requirements?

Options:

- A- Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon S3. Use Amazon Athena to perform SQL queries over the ingested data.
- B- Use Amazon Managed Streaming for Apache Kafka to ingest the data to save it to Amazon Redshift. Enable Amazon Redshift workload management (WLM) to prioritize workloads.
- C- Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon Redshift. Enable Amazon Redshift workload management (WLM) to prioritize workloads.
- D- Use Amazon Kinesis Data Firehose to ingest the data to save it to Amazon S3. Load frequently queried data to Amazon Redshift using the COPY command. Use Amazon Redshift Spectrum for less frequently queried data.

Answer:

B

Question 3

Question Type: MultipleChoice

A company operates toll services for highways across the country and collects data that is used to understand usage patterns. Analysts have requested the ability to run traffic reports in near-real time. The company is interested in building an ingestion pipeline that loads all the data into an Amazon Redshift cluster and alerts operations personnel when toll traffic for a particular toll station does not meet a specified threshold. Station data and the corresponding threshold values are stored in Amazon S3.

Which approach is the MOST efficient way to meet these requirements?

Options:

- A- Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneously. Create a reference data source in Kinesis Data Analytics to temporarily store the threshold values from Amazon S3 and compare the count of vehicles for a particular toll station against its corresponding threshold value. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.

B- Use Amazon Kinesis Data Streams to collect all the data from toll stations. Create a stream in Kinesis Data Streams to temporarily store the threshold values from Amazon S3. Send both streams to Amazon Kinesis Data Analytics to compare the count of vehicles for a particular toll station against its corresponding threshold value. Use AWS Lambda to publish an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met. Connect Amazon Kinesis Data Firehose to Kinesis Data Streams to deliver the data to Amazon Redshift.

C- Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift. Then, automatically trigger an AWS Lambda function that queries the data in Amazon Redshift, compares the count of vehicles for a particular toll station against its corresponding threshold values read from Amazon S3, and publishes an Amazon Simple Notification Service (Amazon SNS) notification if the threshold is not met.

D- Use Amazon Kinesis Data Firehose to collect data and deliver it to Amazon Redshift and Amazon Kinesis Data Analytics simultaneously. Use Kinesis Data Analytics to compare the count of vehicles against the threshold value for the station stored in a table as an in-application stream based on information stored in Amazon S3. Configure an AWS Lambda function as an output for the application that will publish an Amazon Simple Queue Service (Amazon SQS) notification to alert operations personnel if the threshold is not met.

Answer:

D

Question 4

Question Type: MultipleChoice

A company is creating a data lake by using AWS Lake Formation. The data that will be stored in the data lake contains sensitive customer information and must be encrypted at rest using an AWS Key Management Service (AWS KMS) customer managed key to meet regulatory requirements.

How can the company store the data in the data lake to meet these requirements?

Options:

A- Store the data in an encrypted Amazon Elastic Block Store (Amazon EBS) volume. Register the Amazon EBS volume with Lake Formation.

B- Store the data in an Amazon S3 bucket by using server-side encryption with AWS KMS (SSE-KMS). Register the S3 location with Lake Formation.

C- Encrypt the data on the client side and store the encrypted data in an Amazon S3 bucket. Register the S3 location with Lake Formation.

D- Store the data in an Amazon S3 Glacier Flexible Retrieval vault bucket. Register the S3 Glacier Flexible Retrieval vault with Lake Formation.

Answer:

B

Question 5

Question Type: MultipleChoice

A gaming company is building a serverless data lake. The company is ingesting streaming data into Amazon Kinesis Data Streams and is writing the data to Amazon S3 through Amazon Kinesis Data Firehose. The company is using 10 MB as the S3 buffer size and is using 90 seconds as the buffer interval. The company runs an AWS Glue ETL job to merge and transform the data to a different format before writing the data back to Amazon S3.

Recently, the company has experienced substantial growth in its data volume. The AWS Glue ETL jobs are frequently showing an OutOfMemoryError error.

Which solutions will resolve this issue without incurring additional costs? (Select TWO.)

Options:

- A- Place the small files into one S3 folder. Define one single table for the small S3 files in AWS Glue Data Catalog. Rerun the AWS Glue ETL jobs against this AWS Glue table.
- B- Create an AWS Lambda function to merge small S3 files and invoke them periodically. Run the AWS Glue ETL jobs after successful completion of the Lambda function.
- C- Run the S3DistCp utility in Amazon EMR to merge a large number of small S3 files before running the AWS Glue ETL jobs.
- D- Use the groupFiles setting in the AWS Glue ETL job to merge small S3 files and rerun AWS Glue ETL jobs.
- E- Update the Kinesis Data Firehose S3 buffer size to 128 MB. Update the buffer interval to 900 seconds.

Answer:

A, D

Explanation:

The groupFiles setting is a feature of AWS Glue that enables an ETL job to group files when they are read from an Amazon S3 data store. This can reduce the number of ETL tasks and in-memory partitions, and improve the performance and memory efficiency of the job¹. By using the groupFiles setting in the AWS Glue ETL job, the gaming company can merge small S3 files and

avoid the OutOfMemoryError error.

The Kinesis Data Firehose S3 buffer size and buffer interval are parameters that determine how much data is buffered before delivering it to Amazon S3. Increasing the buffer size and buffer interval can result in larger files being delivered to Amazon S3, which can reduce the number of small files and improve the performance of downstream processing². By updating the Kinesis Data Firehose S3 buffer size to 128 MB and buffer interval to 900 seconds, the gaming company can create fewer, larger S3 files and avoid the OutOfMemoryError error.

Question 6

Question Type: MultipleChoice

A company wants to use automatic machine learning (ML) to create and visualize forecasts of complex scenarios and trends.

Which solution will meet these requirements with the LEAST management overhead?

Options:

- A- Use an AWS Glue ML job to transform the data and create forecasts. Use Amazon QuickSight to visualize the data.
- B- Use Amazon QuickSight to visualize the data. Use ML-powered forecasting in QuickSight to create forecasts.
- C- Use a prebuilt ML AMI from the AWS Marketplace to create forecasts. Use Amazon QuickSight to visualize the data.
- D- Use Amazon SageMaker inference pipelines to create and update forecasts. Use Amazon QuickSight to visualize the combined data.

Answer:

B

Question 7

Question Type: MultipleChoice

A company is building an analytical solution that includes Amazon S3 as data lake storage and Amazon Redshift for data warehousing. The company wants to use Amazon Redshift Spectrum to query the data that is stored in Amazon S3.

Which steps should the company take to improve performance when the company uses Amazon Redshift Spectrum to query the S3 data files? (Select THREE)

Use gzip compression with individual file sizes of 1-5 GB

Options:

- A- Use a columnar storage file format
- B- Partition the data based on the most common query predicates
- C- Split the data into KB-sized files.
- D- Keep all files about the same size.
- E- Use file formats that are not splittable



Answer:

B, C, D



To Get Premium Files for DAS-C01 Visit

<https://www.p2pexams.com/products/das-c01>

For More Free Questions Visit

<https://www.p2pexams.com/amazon/pdf/das-c01>

20%
DISCOUNT

P2P
exams