



Free Questions for **DAS-C01**

Shared by **Hamilton** on **12-12-2023**

For More Free Questions and Preparation Resources

[Check the Links on Last Page](#)



Question 1

Question Type: MultipleChoice

A large energy company is using Amazon QuickSight to build dashboards and report the historical usage data of its customers. This data is hosted in Amazon Redshift. The reports need access to all the fact tables' billions of records to create aggregation in real time grouping by multiple dimensions.

A data analyst created the dataset in QuickSight by using a SQL query and not SPICE. Business users have noted that the response time is not fast enough to meet their needs.

Which action would speed up the response time for the reports with the LEAST implementation effort?

Options:

- A- Use QuickSight to modify the current dataset to use SPICE
- B- Use AWS Glue to create an Apache Spark job that joins the fact table with the dimensions. Load the data into a new table
- C- Use Amazon Redshift to create a materialized view that joins the fact table with the dimensions
- D- Use Amazon Redshift to create a stored procedure that joins the fact table with the dimensions. Load the data into a new table

Answer:

A

Question 2

Question Type: MultipleChoice

A hospital uses an electronic health records (EHR) system to collect two types of data:

- * Patient information, which includes a patient's name and address
- * Diagnostic tests conducted and the results of these tests

Patient information is expected to change periodically. Existing diagnostic test data never changes and only new records are added.

The hospital runs an Amazon Redshift cluster with four dc2.large nodes and wants to automate the ingestion of the patient information and diagnostic test data into respective Amazon Redshift

tables for analysis The EHR system exports data as CSV files to an Amazon S3 bucket on a daily basis Two sets of CSV files are generated One set of files is for patient information with updates, deletes, and inserts The other set of files is for new diagnostic test data only

What is the MOST cost-effective solution to meet these requirements?

Options:

- A- Use Amazon EMR with Apache Hudi. Run daily ETL jobs using Apache Spark and the Amazon Redshift JDBC driver
- B- Use an AWS Glue crawler to catalog the data in Amazon S3 Use Amazon Redshift Spectrum to perform scheduled queries of the data in Amazon S3 and ingest the data into the patient information table and the diagnostic tests table.
- C- Use an AWS Lambda function to run a COPY command that appends new diagnostic test data to the diagnostic tests table Run another COPY command to load the patient information data into the staging tables Use a stored procedure to handle create update, and delete operations for the patient information table
- D- Use AWS Database Migration Service (AWS DMS) to collect and process change data capture (CDC) records Use the COPY command to load patient information data into the staging tables. Use a stored procedure to handle create, update and delete operations for the patient information table

Answer:

B

Question 3

Question Type: MultipleChoice

A company collects data from parking garages. Analysts have requested the ability to run reports in near real time about the number of vehicles in each garage.

The company wants to build an ingestion pipeline that loads the data into an Amazon Redshift cluster. The solution must alert operations personnel when the number of vehicles in a particular garage exceeds a specific threshold. The alerting query will use garage threshold values as a static reference. The threshold values are stored in

Amazon S3.

What is the MOST operationally efficient solution that meets these requirements?

Options:

- A-** Use an Amazon Kinesis Data Firehose delivery stream to collect the data and to deliver the data to Amazon Redshift. Create an Amazon Kinesis Data Analytics application that uses the same delivery stream as an input source. Create a reference data source in Kinesis Data Analytics to temporarily store the threshold values from Amazon S3 and to compare the number of vehicles in a particular garage to the corresponding threshold value. Configure an AWS Lambda function to publish an Amazon Simple Notification Service (Amazon SNS) notification if the number of vehicles exceeds the threshold.
- B-** Use an Amazon Kinesis data stream to collect the data. Use an Amazon Kinesis Data Firehose delivery stream to deliver the data to Amazon Redshift. Create another Kinesis data stream to temporarily store the threshold values from Amazon S3. Send the delivery stream and the second data stream to Amazon Kinesis Data Analytics to compare the number of vehicles in a particular garage to the corresponding threshold value. Configure an AWS Lambda function to publish an Amazon Simple Notification Service (Amazon SNS) notification if the number of vehicles exceeds the threshold.
- C-** Use an Amazon Kinesis Data Firehose delivery stream to collect the data and to deliver the data to Amazon Redshift. Automatically initiate an AWS Lambda function that queries the data in Amazon Redshift. Configure the Lambda function to compare the number of vehicles in a particular garage to the corresponding threshold value from Amazon S3. Configure the Lambda function to also publish an Amazon Simple Notification Service (Amazon SNS) notification if the number of vehicles exceeds the threshold.
- D-** Use an Amazon Kinesis Data Firehose delivery stream to collect the data and to deliver the data to Amazon Redshift. Create an Amazon Kinesis Data Analytics application that uses the same delivery stream as an input source. Use Kinesis Data Analytics to compare the number of vehicles in a particular garage to the corresponding threshold value that is stored in a table as an in-application stream. Configure an AWS Lambda function as an output for the application to publish an Amazon Simple Queue Service (Amazon SQS) notification if the number of vehicles exceeds the threshold.

Answer:

A

Explanation:

This solution meets the requirements because:

It uses Amazon Kinesis Data Firehose to collect and deliver data to Amazon Redshift in near real time, without requiring any coding or server management¹.

It uses Amazon Kinesis Data Analytics to process and analyze streaming data using SQL queries or Apache Flink applications². It can also create a reference data source that allows joining streaming data with static data stored in Amazon S3³. This way, it can compare the number of vehicles in each garage with the corresponding threshold value from the reference data source.

It uses AWS Lambda to create a serverless function that can be triggered by Kinesis Data Analytics as an output destination⁴. The Lambda function can then publish an Amazon SNS notification to alert operations personnel when the number of vehicles exceeds the threshold⁵.

Question 4

Question Type: MultipleChoice

An ecommerce company is migrating its business intelligence environment from on premises to the AWS Cloud. The company will use Amazon Redshift in a public subnet and Amazon QuickSight. The tables already are loaded into Amazon Redshift and can be accessed by a SQL tool.

The company starts QuickSight for the first time. During the creation of the data source, a data analytics specialist enters all the information and tries to validate the connection. An error with the following message occurs: "Creating a connection to your data source timed out."

How should the data analytics specialist resolve this error?

Options:

- A- Grant the SELECT permission on Amazon Redshift tables.
- B- Add the QuickSight IP address range into the Amazon Redshift security group.
- C- Create an IAM role for QuickSight to access Amazon Redshift.
- D- Use a QuickSight admin user for creating the dataset.

Answer:

A

Explanation:

Connection to the database times out

Your client connection to the database appears to hang or time out when running long queries, such as a COPY command. In this case, you might observe that the Amazon Redshift console displays that the query has completed, but the client tool itself still appears to be running the query. The results of the query might be missing or incomplete depending on when the connection stopped.

Question 5

Question Type: MultipleChoice

A company stores Apache Parquet-formatted files in Amazon S3. The company uses an AWS Glue Data Catalog to store the table metadata and Amazon Athena to query and analyze the data. The tables have a large number of partitions. The queries are only run on small subsets of data in the table. A data analyst adds new time partitions into the table as new data arrives. The data analyst has been asked to reduce the query runtime.

Which solution will provide the MOST reduction in the query runtime?

Options:

- A- Convert the Parquet files to the csv file format..Then attempt to query the data again
- B- Convert the Parquet files to the Apache ORC file format. Then attempt to query the data again
- C- Use partition projection to speed up the processing of the partitioned table
- D- Add more partitions to be used over the table. Then filter over two partitions and put all columns in the WHERE clause

Answer:

C

Question 6

Question Type: MultipleChoice

A company is hosting an enterprise reporting solution with Amazon Redshift. The application provides reporting capabilities to three main groups: an executive group to access financial reports, a data analyst group to run long-running ad-hoc queries, and a data engineering group to run stored procedures and ETL processes. The executive team requires queries to run with optimal performance. The data engineering team expects queries to take minutes.

Which Amazon Redshift feature meets the requirements for this task?

Options:

- A- Concurrency scaling
- B- Short query acceleration (SQA)
- C- Workload management (WLM)
- D- Materialized views

Answer:

D

Explanation:

Materialized views:

Question 7

Question Type: MultipleChoice

A company receives datasets from partners at various frequencies. The datasets include baseline data and incremental data. The company needs to merge and store all the datasets without reprocessing the data.

Which solution will meet these requirements with the LEAST development effort?

Options:

- A- Use an AWS Glue job with a temporary table to process the datasets. Store the data in an Amazon RDS table.
- B- Use an Apache Spark job in an Amazon EMR cluster to process the datasets. Store the data in EMR File System (EMRFS).
- C- Use an AWS Glue job with job bookmarks enabled to process the datasets. Store the data in Amazon S3.
- D- Use an AWS Lambda function to process the datasets. Store the data in Amazon S3.

Answer:

C

Explanation:

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy to prepare and load data for analytics¹. It can process datasets from various sources and formats, such as JDBC, Amazon S3, Amazon RDS, etc.

AWS Glue job bookmarks are a feature that helps AWS Glue track data that has already been processed during a previous run of an ETL job. This can prevent the reprocessing of old data and

enable the processing of new data when rerunning on a scheduled interval². Job bookmarks can handle both baseline data and incremental data from different sources.

Amazon S3 is a highly scalable, durable, and secure object storage service that can store any amount and type of data³. It can be used as a data lake to store the merged and processed datasets from AWS Glue. It can also integrate with other AWS services, such as Amazon Athena, Amazon Redshift Spectrum, Amazon EMR, etc., for further analysis and processing.

Question 8

Question Type: MultipleChoice

A company is designing a data warehouse to support business intelligence reporting. Users will access the executive dashboard heavily each Monday and Friday morning

for 1 hour. These read-only queries will run on the active Amazon Redshift cluster, which runs on dc2.8xlarge compute nodes 24 hours a day, 7 days a week. There are

three queues set up in workload management: Dashboard, ETL, and System. The Amazon Redshift cluster needs to process the queries without wait time.

What is the MOST cost-effective way to ensure that the cluster processes these queries?

Options:

- A- Perform a classic resize to place the cluster in read-only mode while adding an additional node to the cluster.
- B- Enable automatic workload management.
- C- Perform an elastic resize to add an additional node to the cluster.
- D- Enable concurrency scaling for the Dashboard workload queue.

Answer:

D

Question 9

Question Type: MultipleChoice

A company has a process that writes two datasets in CSV format to an Amazon S3 bucket every 6 hours. The company needs to join the datasets, convert the data to Apache Parquet, and store the data within another bucket for users to query using Amazon Athen

a. The data also needs to be loaded to Amazon Redshift for advanced analytics. The company needs a solution that is resilient to the failure of any individual job component and can be restarted in case of an error.

Which solution meets these requirements with the LEAST amount of operational overhead?

Options:

A- Use AWS Step Functions to orchestrate an Amazon EMR cluster running Apache Spark. Use PySpark to generate data frames of the datasets in Amazon S3, transform the data, join the data, write the data back to Amazon S3, and load the data to Amazon Redshift.

B- Create an AWS Glue job using Python Shell that generates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift. Use an AWS Glue workflow to orchestrate the AWS Glue job at the desired frequency.

C- Use AWS Step Functions to orchestrate the AWS Glue job. Create an AWS Glue job using Python Shell that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift.

D- Create an AWS Glue job using PySpark that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift. Use an AWS Glue workflow to orchestrate the AWS Glue job.

Answer:

D

Explanation:

AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy to prepare and load data for analytics¹. It can process datasets from various sources and formats, such as CSV and Parquet, and write them to different destinations, such as Amazon S3 and Amazon Redshift².

AWS Glue provides two types of jobs: Spark and Python Shell. Spark jobs run on Apache Spark, a distributed processing framework that supports a wide range of data processing tasks³. Python Shell jobs run Python scripts on a managed serverless infrastructure⁴. Spark jobs are more suitable for complex data transformations and joins than Python Shell jobs.

AWS Glue provides dynamic frames, which are an extension of Apache Spark data frames. Dynamic frames handle schema variations and errors in the data more easily than data frames. They also provide a set of transformations that can be applied to the data, such as join, filter, map, etc.

AWS Glue provides workflows, which are directed acyclic graphs (DAGs) that orchestrate multiple

ETL jobs and crawlers. Workflows can handle dependencies, retries, error handling, and concurrency for ETL jobs and crawlers. They can also be triggered by schedules or events.

By creating an AWS Glue job using PySpark that creates dynamic frames of the datasets in Amazon S3, transforms the data, joins the data, writes the data back to Amazon S3, and loads the data to Amazon Redshift, the company can perform the required ETL tasks with a single job. By using an AWS Glue workflow to orchestrate the AWS Glue job, the company can schedule and monitor the job execution with minimal operational overhead.

Question 10

Question Type: MultipleChoice

A company has a mobile app that has millions of users. The company wants to enhance the mobile app by including interactive data visualizations that show user trends.

The data for visualization is stored in a large data lake with 50 million rows. Data that is used in the visualization should be no more than two hours old.

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

Options:

- A- Run an hourly batch process that renders user-specific data visualizations as static images that are stored in Amazon S3.
- B- Precompute aggregated data hourly. Store the data in Amazon DynamoDB. Render the data by using the D3.js JavaScript library.
- C- Embed an Amazon QuickSight Enterprise edition dashboard into the mobile app by using the QuickSight Embedding SDK. Refresh data in SPICE hourly.
- D- Run Amazon Athena queries behind an Amazon API Gateway API. Render the data by using the D3.js JavaScript library.

Answer:

A

Question 11

Question Type: MultipleChoice

A financial services firm is processing a stream of real-time data from an application by using

Apache Kafka and Kafka MirrorMaker. These tools run on premises and stream data to Amazon Managed Streaming for Apache Kafka (Amazon MSK) in the us-east-1 Region. An Apache Flink consumer running on Amazon EMR enriches the data in real time and transfers the output files to an Amazon S3 bucket. The company wants to ensure that the streaming application is highly available across AWS Regions with an RTO of less than 2 minutes.

Which solution meets these requirements?

Options:

- A- Launch another Amazon MSK and Apache Flink cluster in the us-west-1 Region that is the same size as the original cluster in the us-east-1 Region. Simultaneously publish and process the data in both Regions. In the event of a disaster that impacts one of the Regions, switch to the other Region.
- B- Set up Cross-Region Replication from the Amazon S3 bucket in the us-east-1 Region to the us-west-1 Region. In the event of a disaster, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region.
- C- Add an AWS Lambda function in the us-east-1 Region to read from Amazon MSK and write to a global Amazon
- D- DynamoDB table in on-demand capacity mode. Export the data from DynamoDB to Amazon S3 in the us-west-1 Region. In the event of a disaster that impacts the us-east-1 Region, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region.
- E- Set up Cross-Region Replication from the Amazon S3 bucket in the us-east-1 Region to the us-west-1 Region. In the event of a disaster, immediately create Amazon MSK and Apache Flink clusters in the us-west-1 Region and start publishing data to this Region. Store 7 days of data in on-premises Kafka clusters and recover the data missed during the recovery time from the on-premises cluster.

Answer:

A

Question 12

Question Type: MultipleChoice

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?

Options:

- 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

Answer:

B



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