

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.pass2lead.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 advertising company wants to perform sentiment analysis of social media data to measure the success of online advertisements. The company wants to implement an end-to-end streaming solution to continuously ingest data from various social networks, clean and transform the streaming data in near-real time, and make the data available for analytics and visualization with Amazon QuickSight. The company wants a solution that is easy to implement and manage so it can design better analytics solutions instead of provisioning and maintaining infrastructure.

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

- A. Use Amazon Kinesis Data Firehose to ingest the data. Author an AWS Glue streaming ETL job to transform the ingested data. Load the transformed data into an Amazon Redshift table.
- B. Use Apache Kafka running on Amazon EC2 instances to ingest the data. Create an Amazon EMR Spark job to transform the ingested data. Use the COPY command to load the transformed data into an Amazon Redshift table.
- C. Use Amazon Managed Streaming for Apache Kafka (Amazon MSK) to ingest the data. Create an Amazon EMR Spark job to transform the ingested data. Use the COPY command to load the transformed data into an Amazon Redshift table.
- D. Use Amazon Kinesis Data Streams to ingest the data. Author an AWS Glue streaming ETL job to transform the ingested data. Load the transformed data into an Amazon Redshift table.

Correct Answer: C

QUESTION 2

A company has an electronic healthcare system that contains data of patients. The data is consolidated from multiple systems and is stored in an Amazon S3 bucket in .csv format. The company has created an AWS Glue Data Catalog. The dataset contains duplicate data, and no unique keys exist to identify a patient. Fields do not match exactly across the systems.

A data analytics specialist must design a solution to identify and remove duplicates. The solution must minimize the amount of human intervention and code that are required.

The data analytics specialist starts by using labeled data to teach the FindMatches machine learning (ML) transform.

What must the data analytics specialist do next to meet these requirements?

- A. Identify matches in the dataset by using an AWS Glue ETL job with Spark distinct(). Review the output by using Amazon Redshift Spectrum.
- B. Identify matches in the dataset by using an AWS Glue ETL job with Spark distinct(). Create a Data Catalog of transformed results Review the output by using Amazon Athena.
- C. Identify matches in the dataset by using an AWS Glue ETL job that has a transform type of "find matching records." Create a Data Catalog of transformed results. Review the output by using Amazon Athena.
- D. Identify matches in the dataset by using an AWS Glue ETL job that has a transform type of "find matching records." Review the output by using Amazon Redshift Spectrum.

Correct Answer: C

QUESTION 3

A company wants to build a real-time data processing and delivery solution for streaming data. The data is being streamed through an Amazon Kinesis data stream. The company wants to use an Apache Flink application to process the data before writing the data to another Kinesis data stream. The data must be stored in an Amazon S3 data lake every 60 seconds for further analytics.

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

- A. Host the Flink application on an Amazon EMR cluster. Use Amazon Kinesis Data Firehose to write the data to Amazon S3.
- B. Host the Flink application on Amazon Kinesis Data Analytics. Use AWS Glue to write the data to Amazon S3.
- C. Host the Flink application on an Amazon EMR cluster. Use AWS Glue to write the data to Amazon S3.
- D. Host the Flink application on Amazon Kinesis Data Analytics. Use Amazon Kinesis Data Firehose to write the data to Amazon S3.

Correct Answer: A

QUESTION 4

A large marketing company needs to store all of its streaming logs and create near-real-time dashboards. The dashboards will be used to help the company make critical business decisions and must be highly available.

Which solution meets these requirements?

- A. Store the streaming logs in Amazon S3 with replication to an S3 bucket in a different Availability Zone. Create the dashboards by using Amazon QuickSight.
- B. Deploy an Amazon Redshift cluster with at least three nodes in a VPC that spans two Availability Zones. Store the streaming logs and use the Redshift cluster as a source to create the dashboards by using Amazon QuickSight.
- C. Store the streaming logs in Amazon S3 with replication to an S3 bucket in a different Availability Zone. Every time a new log is added in the bucket, invoke an AWS Lambda function to update the dashboards in Amazon QuickSight.
- D. Store the streaming logs in Amazon OpenSearch Service deployed across three Availability Zones and with three dedicated master nodes. Create the dashboards by using OpenSearch Dashboards.

Correct Answer: B

QUESTION 5

A marketing company wants to improve its reporting and business intelligence capabilities. During the planning phase, the company interviewed the relevant stakeholders and discovered that:

The operations team reports are run hourly for the current month's data.

The sales team wants to use multiple Amazon QuickSight dashboards to show a rolling view of the last 30 days based on several categories. The sales team also wants to view the data as soon as it reaches the reporting backend.

The finance team's reports are run daily for last month's data and once a month for the last 24 months of data.

Currently, there is 400 TB of data in the system with an expected additional 100 TB added every month. The company is looking for a solution that is as cost-effective as possible.

Which solution meets the company's requirements?

- A. Store the last 24 months of data in Amazon Redshift. Configure Amazon QuickSight with Amazon Redshift as the data source.
- B. Store the last 2 months of data in Amazon Redshift and the rest of the months in Amazon S3. Set up an external schema and table for Amazon Redshift Spectrum. Configure Amazon QuickSight with Amazon Redshift as the data source.
- C. Store the last 24 months of data in Amazon S3 and query it using Amazon Redshift Spectrum. Configure Amazon QuickSight with Amazon Redshift Spectrum as the data source.
- D. Store the last 2 months of data in Amazon Redshift and the rest of the months in Amazon S3. Use a long-running Amazon EMR with Apache Spark cluster to query the data as needed. Configure Amazon QuickSight with Amazon EMR as the data source.

Correct Answer: B

[Latest DAS-C01 Dumps](#)

[DAS-C01 Study Guide](#)

[DAS-C01 Braindumps](#)