



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.passapply.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 has an application that runs on Amazon EC2 instances launched in a VPC. The company wants to build a solution that allows the security team to collect VPC Flow Logs and analyze network traffic. Which solution MOST cost-effectively meets these requirements?

- A. Publish VPC Flow Logs to Amazon CloudWatch Logs and use Amazon Athena for analytics.
- B. Publish VPC Flow Logs to Amazon CloudWatch Logs and stream log data to an Amazon OpenSearch Service cluster for analytics.
- C. Publish VPC Flow Logs to Amazon S3 in text format and use Amazon Athena for analytics.
- D. Publish VPC Flow Logs to Amazon S3 in Apache Parquet format and use Amazon Athena for analytics.

Correct Answer: B

QUESTION 2

A large fashion retailer wants to transform a source dataset to a consumable format. The retailer is building an ETL pipeline and needs to deduplicate the data because the retailer's various departments share similar customer and stock information. The retailer wants to build a data lake in Amazon S3 after the transformation and deduplication processes are completed.

Which solution MOST cost-effectively meets these requirements?

- A. Load the data into Amazon Redshift and build custom deduplication scripts by using SQL. Use the UNLOAD command in Amazon Redshift to store the data in Amazon S3.
- B. Use AWS Glue to transform the data and use FindMatches to deduplicate the data. Store the output in Amazon S3.
- C. Use Amazon EMR to transform the data. Deduplicate the data by using custom Spark SQL scripts and use EMRFS to store the output in Amazon S3.
- D. Use an Amazon Athena federated query to load the data from the sources. Build custom Athena SQL scripts to deduplicate and store the output to Amazon S3.

Correct Answer: C

QUESTION 3

A healthcare company ingests patient data from multiple data sources and stores it in an Amazon S3 staging bucket. An AWS Glue ETL job transforms the data, which is written to an S3-based data lake to be queried using Amazon Athena. The company wants to match patient records even when the records do not have a common unique identifier.

Which solution meets this requirement?

- A. Use Amazon Macie pattern matching as part of the ETL job



- B. Train and use the AWS Glue PySpark filter class in the ETLjob
- C. Partition tables and use the ETL job to partition the data on patient name
- D. Train and use the AWS Glue FindMatches ML transform in the ETLjob

Correct Answer: D

The FindMatches transform enables you to identify duplicate or matching records in your dataset, even when the records do not have a common unique identifier and no fields match exactly.

Reference: <https://docs.aws.amazon.com/glue/latest/dg/machine-learning.html>

QUESTION 4

A company analyzes historical data and needs to query data that is stored in Amazon S3. New data is generated daily as .csv files that are stored in Amazon S3. The company's analysts are using Amazon Athena to perform SQL queries against a recent subset of the overall data.

The amount of data that is ingested into Amazon S3 has increased substantially over time, and the query latency also has increased.

Which solutions could the company implement to improve query performance? (Choose two.)

- A. Use MySQL Workbench on an Amazon EC2 instance, and connect to Athena by using a JDBC or ODBC connector. Run the query from MySQL Workbench instead of Athena directly.
- B. Use Athena to extract the data and store it in Apache Parquet format on a daily basis. Query the extracted data.
- C. Run a daily AWS Glue ETL job to convert the data files to Apache Parquet and to partition the converted files. Create a periodic AWS Glue crawler to automatically crawl the partitioned data on a daily basis.
- D. Run a daily AWS Glue ETL job to compress the data files by using the .gzip format. Query the compressed data.
- E. Run a daily AWS Glue ETL job to compress the data files by using the .lzo format. Query the compressed data.

Correct Answer: BC

Reference: <https://www.upsolver.com/blog/apache-parquet-why-use> <https://aws.amazon.com/blogs/big-data/work-with-partitioned-data-in-aws-glue/>

QUESTION 5

A company uses the Amazon Kinesis SDK to write data to Kinesis Data Streams. Compliance requirements state that the data must be encrypted at rest using a key that can be rotated. The company wants to meet this encryption requirement with minimal coding effort.

How can these requirements be met?

- A. Create a customer master key (CMK) in AWS KMS. Assign the CMK an alias. Use the AWS Encryption SDK, providing it with the key alias to encrypt and decrypt the data.
- B. Create a customer master key (CMK) in AWS KMS. Assign the CMK an alias. Enable server-side encryption on the



Kinesis data stream using the CMK alias as the KMS master key.

C. Create a customer master key (CMK) in AWS KMS. Create an AWS Lambda function to encrypt and decrypt the data. Set the KMS key ID in the function's environment variables.

D. Enable server-side encryption on the Kinesis data stream using the default KMS key for Kinesis Data Streams.

Correct Answer: B

Reference: <https://aws.amazon.com/kinesis/data-streams/faqs/>

[Latest DAS-C01 Dumps](#)

[DAS-C01 PDF Dumps](#)

[DAS-C01 Practice Test](#)