



We offer free update service for one year!
<https://www.certqueen.com>

Exam : MLS-C01

**Title : AWS Certified Machine
Learning - Specialty**

Version : DEMO

1.A machine learning specialist stores IoT soil sensor data in Amazon DynamoDB table and stores weather event data as JSON files in Amazon S3. The dataset in DynamoDB is 10 GB in size and the dataset in Amazon S3 is 5 GB in size. The specialist wants to train a model on this data to help predict soil moisture levels as a function of weather events using Amazon SageMaker.

Which solution will accomplish the necessary transformation to train the Amazon SageMaker model with the LEAST amount of administrative overhead?

- A. Launch an Amazon EMR cluster. Create an Apache Hive external table for the DynamoDB table and S3 data. Join the Hive tables and write the results out to Amazon S3.
- B. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output to an Amazon Redshift cluster.
- C. Enable Amazon DynamoDB Streams on the sensor table. Write an AWS Lambda function that consumes the stream and appends the results to the existing weather files in Amazon S3.
- D. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output in CSV format to Amazon S3.

Answer: D

2.A Machine Learning Specialist is working for an online retailer that wants to run analytics on every customer visit, processed through a machine learning pipeline. The data needs to be ingested by Amazon Kinesis Data Streams at up to 100 transactions per second, and the JSON data blob is 100 KB in size.

What is the MINIMUM number of shards in Kinesis Data Streams the Specialist should use to successfully ingest this data?

- A. 1 shards
- B. 10 shards
- C. 100 shards
- D. 1,000 shards

Answer: A

3.A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences and trends to enhance the website for better service and smart recommendations.

Which solution should the Specialist recommend?

- A. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
- B. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database
- C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database
- D. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database

Answer: C

4.A company has raw user and transaction data stored in AmazonS3 a MySQL database, and Amazon RedShift A Data Scientist needs to perform an analysis by joining the three datasets from Amazon S3,

MySQL, and Amazon RedShift, and then calculating the average-of a few selected columns from the joined data

Which AWS service should the Data Scientist use?

- A. Amazon Athena
- B. Amazon Redshift Spectrum
- C. AWS Glue
- D. Amazon QuickSight

Answer: A

5.An online reseller has a large, multi-column dataset with one column missing 30% of its data A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data.

Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

- A. Listwise deletion
- B. Last observation carried forward
- C. Multiple imputation
- D. Mean substitution

Answer: C