



MLS-C01^{Q&As}

AWS Certified Machine Learning - Specialty (MLS-C01)

Pass Amazon MLS-C01 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.pass4itsure.com/aws-certified-machine-learning-specialty.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 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

Correct Answer: C

Reference: <https://worldwidescience.org/topicpages/i/imputing+missing+values.html>

QUESTION 2

A growing company has a business-critical key performance indicator (KPI) for the uptime of a machine learning (ML) recommendation system. The company is using Amazon SageMaker hosting services to develop a recommendation model

in a single Availability Zone within an AWS Region.

A machine learning (ML) specialist must develop a solution to achieve high availability. The solution must have a recovery time objective (RTO) of 5 minutes.

Which solution will meet these requirements with the LEAST effort?

- A. Deploy multiple instances for each endpoint in a VPC that spans at least two Regions.
- B. Use the SageMaker auto scaling feature for the hosted recommendation models.
- C. Deploy multiple instances for each production endpoint in a VPC that spans at least two subnets that are in a second Availability Zone.
- D. Frequently generate backups of the production recommendation model. Deploy the backups in a second Region.

Correct Answer: C

QUESTION 3

A retail company stores 100 GB of daily transactional data in Amazon S3 at periodic intervals. The company wants to identify the schema of the transactional data. The company also wants to perform transformations on the transactional data that is in Amazon S3.

The company wants to use a machine learning (ML) approach to detect fraud in the transformed data.



Which combination of solutions will meet these requirements with the LEAST operational overhead? (Select THREE.)

- A. Use Amazon Athena to scan the data and identify the schema.
- B. Use AWS Glue crawlers to scan the data and identify the schema.
- C. Use Amazon Redshift to store procedures to perform data transformations
- D. Use AWS Glue workflows and AWS Glue jobs to perform data transformations.
- E. Use Amazon Redshift ML to train a model to detect fraud.
- F. Use Amazon Fraud Detector to train a model to detect fraud.

Correct Answer: BDF

To meet the requirements with the least operational overhead, the company should use AWS Glue crawlers, AWS Glue workflows and jobs, and Amazon Fraud Detector. AWS Glue crawlers can scan the data in Amazon S3 and identify the schema, which is then stored in the AWS Glue Data Catalog. AWS Glue workflows and jobs can perform data transformations on the data in Amazon S3 using serverless Spark or Python scripts. Amazon Fraud Detector can train a model to detect fraud using the transformed data and the company's historical fraud labels, and then generate fraud predictions using a simple API call. Option A is incorrect because Amazon Athena is a serverless query service that can analyze data in Amazon S3 using standard SQL, but it does not perform data transformations or fraud detection. Option C is incorrect because Amazon Redshift is a cloud data warehouse that can store and query data using SQL, but it requires provisioning and managing clusters, which adds operational overhead. Moreover, Amazon Redshift does not provide a built-in fraud detection capability. Option E is incorrect because Amazon Redshift ML is a feature that allows users to create, train, and deploy machine learning models using SQL commands in Amazon Redshift. However, using Amazon Redshift ML would require loading the data from Amazon S3 to Amazon Redshift, which adds complexity and cost. Also, Amazon Redshift ML does not support fraud detection as a use case. References: AWS Glue Crawlers AWS Glue Workflows and Jobs Amazon Fraud Detector

QUESTION 4

A company maintains a 2 TB dataset that contains information about customer behaviors. The company stores the dataset in Amazon S3. The company stores a trained model container in Amazon Elastic Container Registry (Amazon ECR).

A machine learning (ML) specialist needs to score a batch model for the dataset to predict customer behavior. The ML specialist must select a scalable approach to score the model.

Which solution will meet these requirements MOST cost-effectively?

- A. Score the model by using AWS Batch managed Amazon EC2 Reserved Instances. Create an Amazon EC2 instance store volume and mount it to the Reserved Instances.
- B. Score the model by using AWS Batch managed Amazon EC2 Spot Instances. Create an Amazon FSx for Lustre volume and mount it to the Spot Instances.
- C. Score the model by using an Amazon SageMaker notebook on Amazon EC2 Reserved Instances. Create an Amazon EBS volume and mount it to the Reserved Instances.
- D. Score the model by using Amazon SageMaker notebook on Amazon EC2 Spot Instances. Create an Amazon Elastic File System (Amazon EFS) file system and mount it to the Spot Instances.

Correct Answer: B



QUESTION 5

An ecommerce company is collecting structured data and unstructured data from its website, mobile apps, and IoT devices. The data is stored in several databases and Amazon S3 buckets. The company is implementing a scalable repository to store structured data and unstructured data. The company must implement a solution that provides a central data catalog, self-service access to the data, and granular data access policies and encryption to protect the data.

Which combination of actions will meet these requirements with the LEAST amount of setup? (Choose three.)

- A. Identify the existing data in the databases and S3 buckets. Link the data to AWS Lake Formation.
- B. Identify the existing data in the databases and S3 buckets. Link the data to AWS Glue.
- C. Run AWS Glue crawlers on the linked data sources to create a central data catalog.
- D. Apply granular access policies by using AWS Identity and Access Management (IAM). Configure server-side encryption on each data source.
- E. Apply granular access policies and encryption by using AWS Lake Formation.
- F. Apply granular access policies and encryption by using AWS Glue.

Correct Answer: ACE

<https://docs.aws.amazon.com/lake-formation/latest/dg/what-is-lake-formation.html>

Lake Formation provides a single place to manage access controls for data in your data lake. You can define security policies that restrict access to data at the database, table, column, row, and cell levels. These policies apply to IAM users and roles, and to users and groups when federating through an external identity provider. You can use fine-grained controls to access data secured by Lake Formation within Amazon Redshift Spectrum, Athena, AWS Glue ETL, and Amazon EMR for Apache Spark. Whenever you create IAM identities, make sure to follow IAM best practices.

[Latest MLS-C01 Dumps](#)

[MLS-C01 VCE Dumps](#)

[MLS-C01 Brindumps](#)