# Exam Questions MLA-C01

AWS Certified Machine Learning Engineer - Associate

**https://www.2passeasy.com/dumps/MLA-C01/**

**NEW QUESTION 1**
An ML engineer needs to implement a solution to host a trained ML model. The rate of requests to the model will be inconsistent throughout the day.
The ML engineer needs a scalable solution that minimizes costs when the model is not in use. The solution also must maintain the model's capacity to respond to requests during times of peak usage.
Which solution will meet these requirements?

A. Create AWS Lambda functions that have fixed concurrency to host the mode
B. Configure the Lambda functions to automatically scale based on the number of requests to the model.
C. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster that uses AWS Fargat
D. Set a static number of tasks to handle requests during times of peak usage.
E. Deploy the model to an Amazon SageMaker endpoin
F. Deploy multiple copies of the model to the endpoin
G. Create an Application Load Balancer to route traffic between the different copies of the model at the endpoint.
H. Deploy the model to an Amazon SageMaker endpoin
I. Create SageMaker endpoint auto scaling policies that are based on Amazon CloudWatch metrics to adjust the number of instances dynamically.

**Answer:** D

**NEW QUESTION 2**
A company is building a deep learning model on Amazon SageMaker. The company uses a large amount of data as the training dataset. The company needs to optimize the model's hyperparameters to minimize the loss function on the validation dataset.
Which hyperparameter tuning strategy will accomplish this goal with the LEAST computation time?

A. Hyperbaric!
B. Grid search
C. Bayesian optimization
D. Random search

**Answer:** A

**NEW QUESTION 3**
A company has historical data that shows whether customers needed long-term support from company staff. The company needs to develop an ML model to predict whether new customers will require long-term support.
Which modeling approach should the company use to meet this requirement?

A. Anomaly detection
B. Linear regression
C. Logistic regression
D. Semantic segmentation

**Answer:** C

**NEW QUESTION 4**
A company uses Amazon Athena to query a dataset in Amazon S3. The dataset has a target variable that the company wants to predict.
The company needs to use the dataset in a solution to determine if a model can predict the target variable.
Which solution will provide this information with the LEAST development effort?

A. Create a new model by using Amazon SageMaker Autopilo
B. Report the model's achieved performance.
C. Implement custom scripts to perform data pre-processing, multiple linear regression, and performance evaluatio
D. Run the scripts on Amazon EC2 instances.
E. Configure Amazon Macie to analyze the dataset and to create a mode
F. Report the model's achieved performance.
G. Select a model from Amazon Bedroc
H. Tune the model with the dat
I. Report the model's achieved performance.

**Answer:** A

**NEW QUESTION 5**
A company is planning to create several ML prediction models. The training data is stored in Amazon S3. The entire dataset is more than 5 in size and consists of CSV, JSON, Apache Parquet, and simple text files.
The data must be processed in several consecutive steps. The steps include complex manipulations that can take hours to finish running. Some of the processing involves natural language processing (NLP) transformations. The entire process must be automated.
Which solution will meet these requirements?

A. Process data at each step by using Amazon SageMaker Data Wrangle
B. Automate the process by using Data Wrangler jobs.
C. Use Amazon SageMaker notebooks for each data processing ste
D. Automate the process by using Amazon EventBridge.
E. Process data at each step by using AWS Lambda function
F. Automate the process by using AWS Step Functions and Amazon EventBridge.
G. Use Amazon SageMaker Pipelines to create a pipeline of data processing step
H. Automate the pipeline by using Amazon EventBridge.

**Answer:** D

**NEW QUESTION 6**
A company has a large, unstructured dataset. The dataset includes many duplicate records across several key attributes.
Which solution on AWS will detect duplicates in the dataset with the LEAST code development?

A. Use Amazon Mechanical Turk jobs to detect duplicates.
B. Use Amazon QuickSight ML Insights to build a custom deduplication model.
C. Use Amazon SageMaker Data Wrangler to pre-process and detect duplicates.
D. Use the AWS Glue FindMatches transform to detect duplicates.

**Answer:** D

**NEW QUESTION 7**
A company uses a hybrid cloud environment. A model that is deployed on premises uses data in Amazon 53 to provide customers with a live conversational engine.
The model is using sensitive data. An ML engineer needs to implement a solution to identify and remove the sensitive data.
Which solution will meet these requirements with the LEAST operational overhead?

A. Deploy the model on Amazon SageMake
B. Create a set of AWS Lambda functions to identify and remove the sensitive data.
C. Deploy the model on an Amazon Elastic Container Service (Amazon ECS) cluster that uses AWS Fargat
D. Create an AWS Batch job to identify and remove the sensitive data.
E. Use Amazon Macie to identify the sensitive dat
F. Create a set of AWS Lambda functions to remove the sensitive data.
G. Use Amazon Comprehend to identify the sensitive dat
H. Launch Amazon EC2 instances to remove the sensitive data.

**Answer:** C

**NEW QUESTION 8**
A company has deployed an XGBoost prediction model in production to predict if a customer is likely to cancel a subscription. The company uses Amazon SageMaker Model Monitor to detect deviations in the F1 score.
During a baseline analysis of model quality, the company recorded a threshold for the F1 score. After several months of no change, the model's F1 score decreases significantly.
What could be the reason for the reduced F1 score?

A. Concept drift occurred in the underlying customer data that was used for predictions.
B. The model was not sufficiently complex to capture all the patterns in the original baseline data.
C. The original baseline data had a data quality issue of missing values.
D. Incorrect ground truth labels were provided to Model Monitor during the calculation of the baseline.

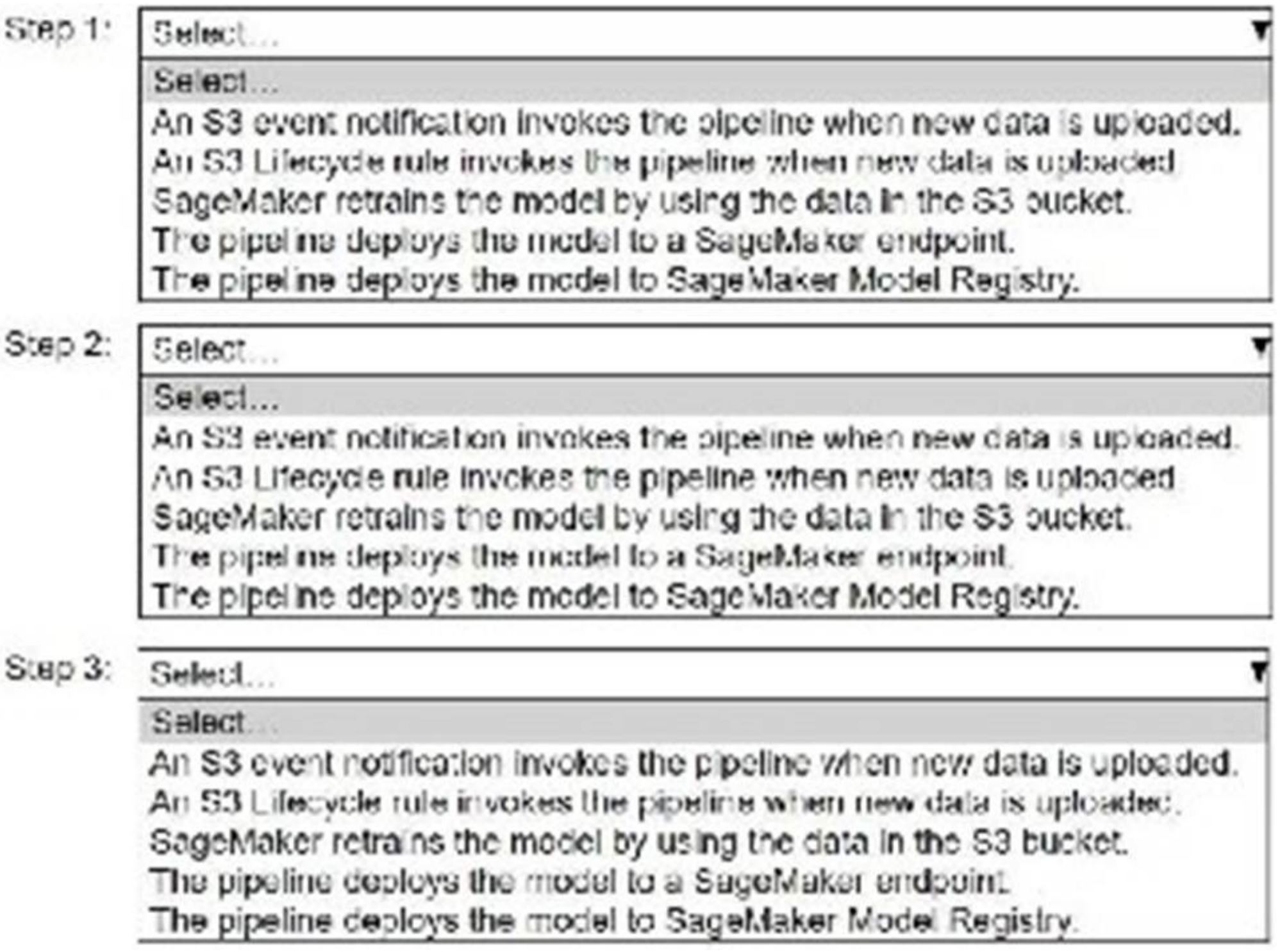
**Answer:** A

**NEW QUESTION 9**
A company runs an Amazon SageMaker domain in a public subnet of a newly created VPC. The network is configured properly, and ML engineers can access the SageMaker domain.
Recently, the company discovered suspicious traffic to the domain from a specific IP address. The company needs to block traffic from the specific IP address.
Which update to the network configuration will meet this requirement?

A. Create a security group inbound rule to deny traffic from the specific IP addres
B. Assign the security group to the domain.
C. Create a network ACL inbound rule to deny traffic from the specific IP addres
D. Assign the rule to the default network Ad for the subnet where the domain is located.
E. Create a shadow variant for the domai
F. Configure SageMaker Inference Recommender to send traffic from the specific IP address to the shadow endpoint.
G. Create a VPC route table to deny inbound traffic from the specific IP addres
H. Assign the route table to the domain.

**Answer:** B

**NEW QUESTION 10**
HOTSPOT
A company wants to host an ML model on Amazon SageMaker. An ML engineer is configuring a continuous integration and continuous delivery (Cl/CD) pipeline in AWS CodePipeline to deploy the model. The pipeline must run automatically when new training data for the model is uploaded to an Amazon S3 bucket.
Select and order the pipeline's correct steps from the following list. Each step should be selected one time or not at all. (Select and order three.)
• An S3 event notification invokes the pipeline when new data is uploaded.
• S3 Lifecycle rule invokes the pipeline when new data is uploaded.
• SageMaker retrains the model by using the data in the S3 bucket.
• The pipeline deploys the model to a SageMaker endpoint.
• The pipeline deploys the model to SageMaker Model Registry.




A. Mastered
B. Not Mastered

**Answer:** A

**Explanation:**


**NEW QUESTION 10**
A company has an ML model that generates text descriptions based on images that customers upload to the company's website. The images can be up to 50 MB in total size.
An ML engineer decides to store the images in an Amazon S3 bucket. The ML engineer must implement a processing solution that can scale to accommodate changes in demand.
Which solution will meet these requirements with the LEAST operational overhead?

A. Create an Amazon SageMaker batch transform job to process all the images in the S3 bucket.
B. Create an Amazon SageMaker Asynchronous Inference endpoint and a scaling polic
C. Run a script to make an inference request for each image.
D. Create an Amazon Elastic Kubernetes Service (Amazon EKS) cluster that uses Karpenter for auto scalin
E. Host the model on the EKS cluste
F. Run a script to make an inference request for each image.
G. Create an AWS Batch job that uses an Amazon Elastic Container Service (Amazon ECS) cluste
H. Specify a list of images to process for each AWS Batch job.

**Answer:** B

**NEW QUESTION 13**
An ML engineer has an Amazon Comprehend custom model in Account A in the us-east-1 Region. The ML engineer needs to copy the model to Account in the same Region.
Which solution will meet this requirement with the LEAST development effort?

A. Use Amazon S3 to make a copy of the mode
B. Transfer the copy to Account B.
C. Create a resource-based IAM polic
D. Use the Amazon Comprehend ImportModel API operation to copy the model to Account B.
E. Use AWS DataSync to replicate the model from Account A to Account B.
F. Create an AWS Site-to-Site VPN connection between Account A and Account to transfer the model.

**Answer:** B

**NEW QUESTION 17**
A company has a conversational AI assistant that sends requests through Amazon Bedrock to an Anthropic Claude large language model (LLM). Users report that when they ask similar questions multiple times, they sometimes receive different answers. An ML engineer needs to improve the responses to be more consistent

and less random.
Which solution will meet these requirements?

A. Increase the temperature parameter and the top_k parameter.
B. Increase the temperature paramete
C. Decrease the top_k parameter.
D. Decrease the temperature paramete
E. Increase the top_k parameter.
F. Decrease the temperature parameter and the top_k parameter.

**Answer:** D

**NEW QUESTION 18**
An ML engineer is using a training job to fine-tune a deep learning model in Amazon SageMaker Studio. The ML engineer previously used the same pre-trained model with a similar dataset. The ML engineer expects vanishing gradient, underutilized GPU, and overfitting problems.
The ML engineer needs to implement a solution to detect these issues and to react in predefined ways when the issues occur. The solution also must provide comprehensive real-time metrics during the training.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use TensorBoard to monitor the training jo
B. Publish the findings to an Amazon Simple Notification Service (Amazon SNS) topi
C. Create an AWS Lambda function to consume the findings and to initiate the predefined actions.
D. Use Amazon CloudWatch default metrics to gain insights about the training jo
E. Use themetrics to invoke an AWS Lambda function to initiate the predefined actions.
F. Expand the metrics in Amazon CloudWatch to include the gradients in each training ste
G. Use the metrics to invoke an AWS Lambda function to initiate the predefined actions.
H. Use SageMaker Debugger built-in rules to monitor the training jo
I. Configure the rules to initiate the predefined actions.

**Answer:** D

**NEW QUESTION 20**
A company has used Amazon SageMaker to deploy a predictive ML model in production. The company is using SageMaker Model Monitor on the model. After a model update, an ML engineer notices data quality issues in the Model Monitor checks.
What should the ML engineer do to mitigate the data quality issues that Model Monitor has identified?

A. Adjust the model's parameters and hyperparameters.
B. Initiate a manual Model Monitor job that uses the most recent production data.
C. Create a new baseline from the latest datase
D. Update Model Monitor to use the new baseline for evaluations.
E. Include additional data in the existing training set for the mode
F. Retrain and redeploy the model.

**Answer:** C

**NEW QUESTION 24**
An ML engineer is developing a fraud detection model by using the Amazon SageMaker XGBoost algorithm. The model classifies transactions as either fraudulent or legitimate.
During testing, the model excels at identifying fraud in the training dataset. However, the
model is inefficient at identifying fraud in new and unseen transactions.
What should the ML engineer do to improve the fraud detection for new transactions?

A. Increase the learning rate.
B. Remove some irrelevant features from the training dataset.
C. Increase the value of the max_depth hyperparameter.
D. Decrease the value of the max_depth hyperparameter.

**Answer:** D

**NEW QUESTION 29**
Case study
An ML engineer is developing a fraud detection model on AWS. The training dataset includes transaction logs, customer profiles, and tables from an on-premises MySQL database. The transaction logs and customer profiles are stored in Amazon S3.
The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data.
Before the ML engineer trains the model, the ML engineer must resolve the issue of the imbalanced data.
Which solution will meet this requirement with the LEAST operational effort?

A. Use Amazon Athena to identify patterns that contribute to the imbalanc
B. Adjust the dataset accordingly.
C. Use Amazon SageMaker Studio Classic built-in algorithms to process the imbalanced dataset.
D. Use AWS Glue DataBrew built-in features to oversample the minority class.
E. Use the Amazon SageMaker Data Wrangler balance data operation to oversample the minority class.

**Answer:** D

**NEW QUESTION 34**
FILL IN THE BLANK

A company stores time-series data about user clicks in an Amazon S3 bucket. The raw data consists of millions of rows of user activity every day. ML engineers access the data to develop their ML models.
The ML engineers need to generate daily reports and analyze click trends over the past 3 days by using Amazon Athena. The company must retain the data for 30 days before archiving the data.
Which solution will provide the HIGHEST performance for data retrieval?

A. Keep all the time-series data without partitioning in the S3 bucke
B. Manually move data that is older than 30 days to separate S3 buckets.
C. Create AWS Lambda functions to copy the time-series data into separate S3 bucket
D. Apply S3 Lifecycle policies to archive data that is older than 30 days to S3 Glacier Flexible Retrieval.
E. Organize the time-series data into partitions by date prefix in the S3 bucke
F. Apply S3 Lifecycle policies to archive partitions that are older than 30 days to S3 Glacier Flexible Retrieval.
G. Put each day's time-series data into its own S3 bucke
H. Use S3 Lifecycle policies to archive S3 buckets that hold data that is older than 30 days to S3 Glacier Flexible Retrieval.

**Answer:** C

**NEW QUESTION 35**
A company uses Amazon SageMaker for its ML workloads. The company's ML engineer receives a 50 MB Apache Parquet data file to build a fraud detection model. The file includes several correlated columns that are not required.
What should the ML engineer do to drop the unnecessary columns in the file with the LEAST effort?

A. Download the file to a local workstatio
B. Perform one-hot encoding by using a custom Python script.
C. Create an Apache Spark job that uses a custom processing script on Amazon EMR.
D. Create a SageMaker processing job by calling the SageMaker Python SDK.
E. Create a data flow in SageMaker Data Wrangle
F. Configure a transform step.

**Answer:** D

**NEW QUESTION 38**
An ML engineer needs to use AWS services to identify and extract meaningful unique keywords from documents.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use the Natural Language Toolkit (NLTK) library on Amazon EC2 instances for text pre- processin
B. Use the Latent Dirichlet Allocation (LDA) algorithm to identify and extract relevant keywords.
C. Use Amazon SageMaker and the BlazingText algorith
D. Apply custom pre-processing steps for stemming and removal of stop word
E. Calculate term frequency-inverse document frequency (TF-IDF) scores to identify and extract relevant keywords.
F. Store the documents in an Amazon S3 bucke
G. Create AWS Lambda functions to process the documents and to run Python scripts for stemming and removal of stop word
H. Use bigram and trigram techniques to identify and extract relevant keywords.
I. Use Amazon Comprehend custom entity recognition and key phrase extraction to identify and extract relevant keywords.

**Answer:** D

**NEW QUESTION 40**
An ML engineer trained an ML model on Amazon SageMaker to detect automobile accidents from dosed-circuit TV footage. The ML engineer used SageMaker Data Wrangler to create a training dataset of images of accidents and non-accidents.
The model performed well during training and validation. However, the model is underperforming in production because of variations in the quality of the images from various cameras.
Which solution will improve the model's accuracy in the LEAST amount of time?

A. Collect more images from all the camera
B. Use Data Wrangler to prepare a new trainingdataset.
C. Recreate the training dataset by using the Data Wrangler corrupt image transfor
D. Specify the impulse noise option.
E. Recreate the training dataset by using the Data Wrangler enhance image contrast transfor
F. Specify the Gamma contrast option.
G. Recreate the training dataset by using the Data Wrangler resize image transfor
H. Crop all images to the same size.

**Answer:** B

**NEW QUESTION 41**
An ML engineer normalized training data by using min-max normalization in AWS Glue DataBrew. The ML engineer must normalize the production inference data in the same way as the training data before passing the production inference data to the model for predictions.
Which solution will meet this requirement?

A. Apply statistics from a well-known dataset to normalize the production samples.
B. Keep the min-max normalization statistics from the training se
C. Use these values to normalize the production samples.
D. Calculate a new set of min-max normalization statistics from a batch of production sample
E. Use these values to normalize all the production samples.
F. Calculate a new set of min-max normalization statistics from each production sampl
G. Use these values to normalize all the production samples.

**Answer:** B

**NEW QUESTION 43**
An ML engineer needs to use AWS CloudFormation to create an ML model that an Amazon SageMaker endpoint will host.
Which resource should the ML engineer declare in the CloudFormation template to meet this requirement?

A. AWS::SageMaker::Model
B. AWS::SageMaker::Endpoint
C. AWS::SageMaker::NotebookInstance
D. AWS::SageMaker::Pipeline

**Answer:** A

**NEW QUESTION 44**
An ML engineer needs to use an ML model to predict the price of apartments in a specific location.
Which metric should the ML engineer use to evaluate the model's performance?

A. Accuracy
B. Area Under the ROC Curve (AUC)
C. F1 score
D. Mean absolute error (MAE)

**Answer:** D

**NEW QUESTION 49**
Case study
An ML engineer is developing a fraud detection model on AWS. The training dataset includes transaction logs, customer profiles, and tables from an on-premises MySQL database. The transaction logs and customer profiles are stored in Amazon S3.
The dataset has a class imbalance that affects the learning of the model's algorithm. Additionally, many of the features have interdependencies. The algorithm is not capturing all the desired underlying patterns in the data.
Which AWS service or feature can aggregate the data from the various data sources?

A. Amazon EMR Spark jobs
B. Amazon Kinesis Data Streams
C. Amazon DynamoDB
D. AWS Lake Formation

**Answer:** A

**NEW QUESTION 52**
A company has AWS Glue data processing jobs that are orchestrated by an AWS Glue workflow. The AWS Glue jobs can run on a schedule or can be launched manually.
The company is developing pipelines in Amazon SageMaker Pipelines for ML model development. The pipelines will use the output of the AWS Glue jobs during the data processing phase of model development. An ML engineer needs to implement a solution that integrates the AWS Glue jobs with the pipelines.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use AWS Step Functions for orchestration of the pipelines and the AWS Glue jobs.
B. Use processing steps in SageMaker Pipeline
C. Configure inputs that point to the Amazon Resource Names (ARNs) of the AWS Glue jobs.
D. Use Callback steps in SageMaker Pipelines to start the AWS Glue workflow and to stop the pipelines until the AWS Glue jobs finish running.
E. Use Amazon EventBridge to invoke the pipelines and the AWS Glue jobs in the desired order.

**Answer:** C

**NEW QUESTION 53**
A company is planning to use Amazon Redshift ML in its primary AWS account. The source data is in an Amazon S3 bucket in a secondary account.
An ML engineer needs to set up an ML pipeline in the primary account to access the S3 bucket in the secondary account. The solution must not require public IPv4 addresses.
Which solution will meet these requirements?

A. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC with no public access enabled in the primary accoun
B. Create a VPC peering connection between the account
C. Update the VPC route tables to remove the route to 0.0.0.0/0.
D. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC with no public access enabled in the primary accoun
E. Create an AWS Direct Connect connection and a transit gatewa
F. Associate the VPCs from both accounts with the transit gatewa
G. Update the VPC route tables to remove the route to 0.0.0.0/0.
H. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC in the primary accoun
I. Create an AWS Site-to-Site VPN connection with two encrypted IPsec tunnels between the account
J. Set up interface VPC endpoints for Amazon S3.
K. Provision a Redshift cluster and Amazon SageMaker Studio in a VPC in the primary accoun
L. Create an S3 gateway endpoin
M. Update the S3 bucket policy to allow IAM principals from the primary accoun
N. Set up interface VPC endpoints for SageMaker and Amazon Redshift.

**Answer:** D

**NEW QUESTION 54**
An ML engineer needs to deploy ML models to get inferences from large datasets in an asynchronous manner. The ML engineer also needs to implement

scheduled monitoring of the data quality of the models. The ML engineer must receive alerts when changes in data quality occur.
Which solution will meet these requirements?

A. Deploy the models by using scheduled AWS Glue job
B. Use Amazon CloudWatch alarms to monitor the data quality and to send alerts.
C. Deploy the models by using scheduled AWS Batch job
D. Use AWS CloudTrail to monitor the data quality and to send alerts.
E. Deploy the models by using Amazon Elastic Container Service (Amazon ECS) on AWS Fargat
F. Use Amazon EventBridge to monitor the data quality and to send alerts.
G. Deploy the models by using Amazon SageMaker batch transfor
H. Use SageMaker Model Monitor to monitor the data quality and to send alerts.

**Answer:** D

**NEW QUESTION 57**
An ML engineer needs to use an Amazon EMR cluster to process large volumes of data in batches. Any data loss is unacceptable.
Which instance purchasing option will meet these requirements MOST cost-effectively?

A. Run the primary node, core nodes, and task nodes on On-Demand Instances.
B. Run the primary node, core nodes, and task nodes on Spot Instances.
C. Run the primary node on an On-Demand Instanc
D. Run the core nodes and task nodes on Spot Instances.
E. Run the primary node and core nodes on On-Demand Instance
F. Run the task nodes on Spot Instances.

**Answer:** D

**NEW QUESTION 58**
A credit card company has a fraud detection model in production on an Amazon SageMaker endpoint. The company develops a new version of the model. The company needs to assess the new model's performance by using live data and without affecting production end users.
Which solution will meet these requirements?

A. Set up SageMaker Debugger and create a custom rule.
B. Set up blue/green deployments with all-at-once traffic shifting.
C. Set up blue/green deployments with canary traffic shifting.
D. Set up shadow testing with a shadow variant of the new model.

**Answer:** D

**NEW QUESTION 59**
.....

# THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual MLA-C01 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the MLA-C01 Product From:

**https://www.2passeasy.com/dumps/MLA-C01/**

## Money Back Guarantee

## MLA-C01 Practice Exam Features:

* MLA-C01 Questions and Answers Updated Frequently

* MLA-C01 Practice Questions Verified by Expert Senior Certified Staff

* MLA-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* MLA-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year