



# **Amazon-Web-Services**

## **Exam Questions MLS-C01**

AWS Certified Machine Learning - Specialty

## About ExamBible

*[Your Partner of IT Exam](#)*

## Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

## Our Advances

### \* 99.9% Uptime

All examinations will be up to date.

### \* 24/7 Quality Support

We will provide service round the clock.

### \* 100% Pass Rate

Our guarantee that you will pass the exam.

### \* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

**NEW QUESTION 1**

A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training. The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII). The dataset:

- \* Must be accessible from a VPC only.
- \* Must not traverse the public internet. How can these requirements be satisfied?

- A. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.
- B. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.
- C. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.
- D. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance.

**Answer: B**

**NEW QUESTION 2**

A Machine Learning Specialist built an image classification deep learning model. However the Specialist ran into an overfitting problem in which the training and testing accuracies were 99% and 75% respectively.

How should the Specialist address this issue and what is the reason behind it?

- A. The learning rate should be increased because the optimization process was trapped at a local minimum.
- B. The dropout rate at the flatten layer should be increased because the model is not generalized enough.
- C. The dimensionality of dense layer next to the flatten layer should be increased because the model is not complex enough.
- D. The epoch number should be increased because the optimization process was terminated before it reached the global minimum.

**Answer: D**

**NEW QUESTION 3**

A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes.

Which prior probability distribution should the ML Specialist use for this variable?

- A. Poisson distribution ,
- B. Uniform distribution
- C. Normal distribution
- D. Binomial distribution

**Answer: D**

**NEW QUESTION 4**

An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.

What should the Specialist do to meet these requirements?

- A. Create one-hot word encoding vectors.
- B. Produce a set of synonyms for every word using Amazon Mechanical Turk.
- C. Create word embedding factors that store edit distance with every other word.
- D. Download word embedding's pre-trained on a large corpus.

**Answer: A**

**NEW QUESTION 5**

A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:

Total number of images available = 1,000 Test set images = 100 (constant test set)

The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.

Which techniques can be used by the ML Specialist to improve this specific test error?

- A. Increase the training data by adding variation in rotation for training images.
- B. Increase the number of epochs for model training.
- C. Increase the number of layers for the neural network.
- D. Increase the dropout rate for the second-to-last layer.

**Answer: B**

**NEW QUESTION 6**

A large JSON dataset for a project has been uploaded to a private Amazon S3 bucket The Machine Learning Specialist wants to securely access and explore the data from an Amazon SageMaker notebook instance A new VPC was created and assigned to the Specialist

How can the privacy and integrity of the data stored in Amazon S3 be maintained while granting access to the Specialist for analysis?

- A. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enabled Use an S3 ACL to open read privileges to the everyone group
- B. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Copy the JSON dataset from Amazon S3 into the ML storage volume on the SageMaker notebook instance and work against the local dataset
- C. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Define a custom S3 bucket policy to only allow requests from your VPC to access the S3 bucket
- D. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enable
- E. Generate an S3 pre-signed URL for access to data in the bucket

**Answer:** B

#### NEW QUESTION 7

A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.

Which storage scheme is MOST adapted to this scenario?

- A. Store datasets as files in Amazon S3.
- B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
- C. Store datasets as tables in a multi-node Amazon Redshift cluster.
- D. Store datasets as global tables in Amazon DynamoDB.

**Answer:** A

#### NEW QUESTION 8

A city wants to monitor its air quality to address the consequences of air pollution A Machine Learning Specialist needs to forecast the air quality in parts per million of contaminates for the next 2 days in the city As this is a prototype, only daily data from the last year is available

Which model is MOST likely to provide the best results in Amazon SageMaker?

- A. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the single time series consisting of the full year of data with a predictor\_type of regressor.
- B. Use Amazon SageMaker Random Cut Forest (RCF) on the single time series consisting of the full year of data.
- C. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full yearof data with a predictor\_type of regressor.
- D. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full yearof data with a predictor\_type of classifier.

**Answer:** C

#### NEW QUESTION 9

A Machine Learning Specialist wants to determine the appropriate SageMakerVariant Invocations Per Instance setting for an endpoint automatic scaling configuration. The Specialist has performed a load test on a single instance and determined that peak requests per second (RPS) without service degradation is about 20 RPS As this is the first deployment, the Specialist intends to set the invocation safety factor to 0.5

Based on the stated parameters and given that the invocations per instance setting is measured on a per-minute basis, what should the Specialist set as the sageMakervariantinvocationsPerinstance setting?

- A. 10
- B. 30
- C. 600
- D. 2,400

**Answer:** C

#### NEW QUESTION 10

An Machine Learning Specialist discover the following statistics while experimenting on a model.

Experiment 1
Baseline model
Train error = 5%
Test error = 16%
Experiment 2
The Specialist added more layers and neurons to the model and received the following results:
Train error = 5.2%
Test error = 15.7%
Experiment 3
The Specialist reverted back to the original number of neurons from Experiment 1 and implemented regularization in the neural network, which yielded the following results:
Train error = 4.7%
Test error = 9.5%

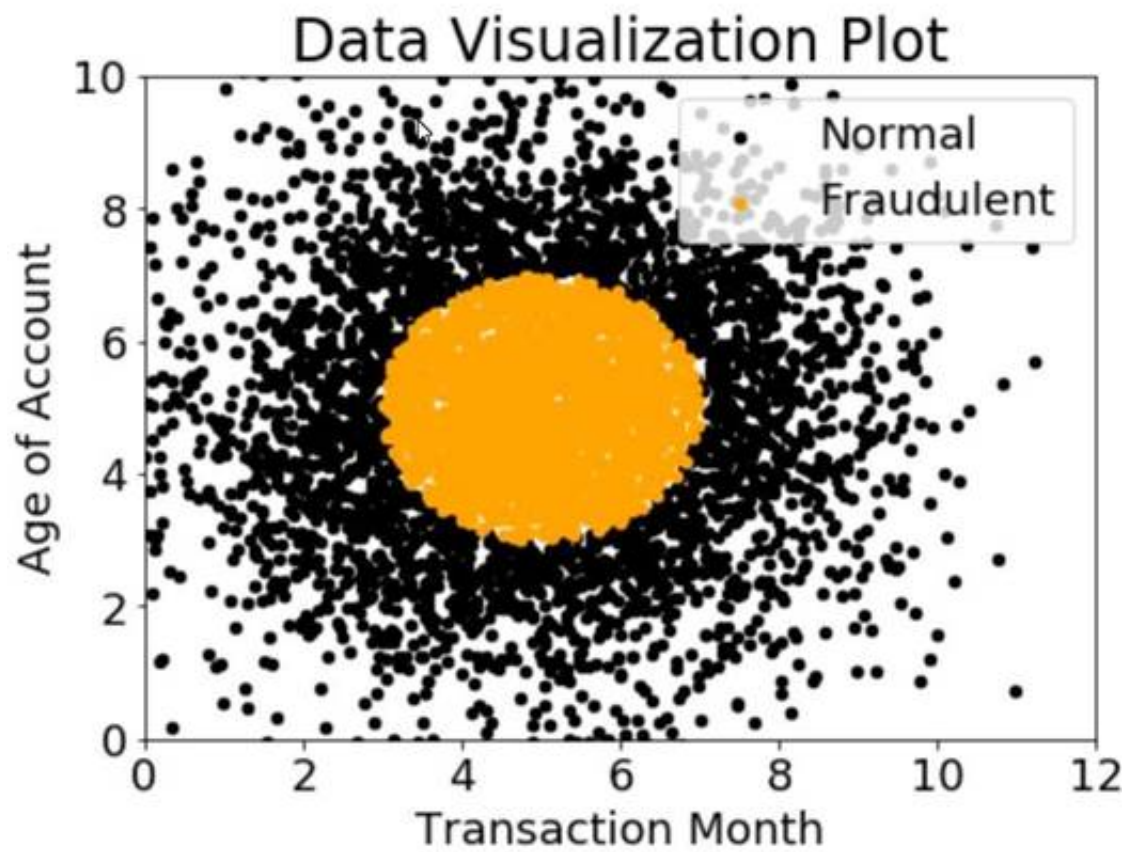
What can the Specialist from the experiments?

- A. The model In Experiment 1 had a high variance error lthat was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal bias error in Experiment 1
- B. The model in Experiment 1 had a high bias error that was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal variance error in Experiment 1
- C. The model in Experiment 1 had a high bias error and a high variance error that were reduced in Experiment 3 by regularization Experiment 2 shows thai high bias cannot be reduced by increasing layers and neurons in the model
- D. The model in Experiment 1 had a high random noise error that was reduced in Expenment 3 by regularization Expenment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model

**Answer:** C

#### NEW QUESTION 10

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information which model would have the HIGHEST accuracy?

- A. Long short-term memory (LSTM) model with scaled exponential linear unit (SELL)
- B. Logistic regression
- C. Support vector machine (SVM) with non-linear kernel
- D. Single perceptron with tanh activation function

**Answer: B**

#### NEW QUESTION 14

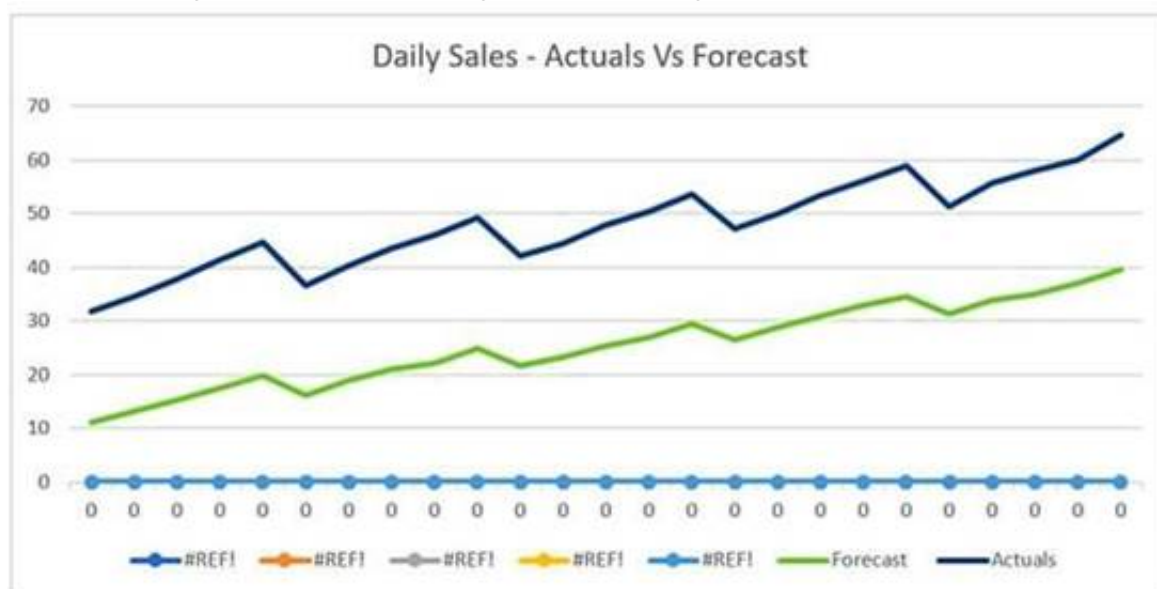
A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training. The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs. What does the Specialist need to do?

- A. Bundle the NVIDIA drivers with the Docker image.
- B. Build the Docker container to be NVIDIA-Docker compatible.
- C. Organize the Docker container's file structure to execute on GPU instances.
- D. Set the GPU flag in the Amazon SageMaker CreateTrainingJob request body

**Answer: A**

#### NEW QUESTION 16

The displayed graph is from a forecasting model for testing a time series.



Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?

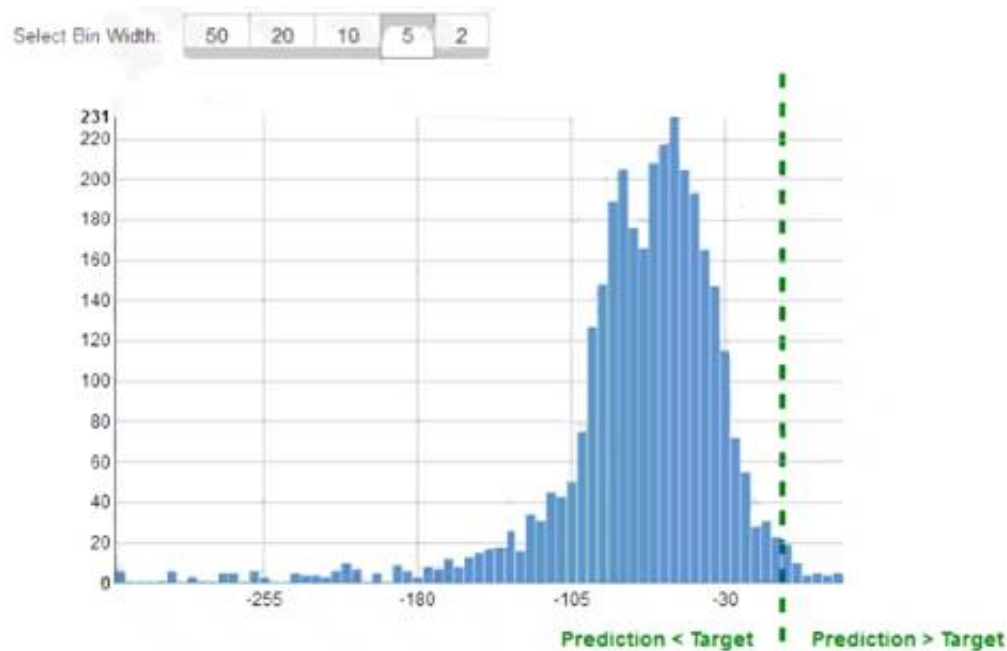
- A. The model predicts both the trend and the seasonality well.
- B. The model predicts the trend well, but not the seasonality.
- C. The model predicts the seasonality well, but not the trend.
- D. The model does not predict the trend or the seasonality well.

**Answer: D**

#### NEW QUESTION 17

While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown. What does this mean?





- A. The model might have prediction errors over a range of target values.
- B. The dataset cannot be accurately represented using the regression model
- C. There are too many variables in the model
- D. The model is predicting its target values perfectly.

**Answer:** D

#### NEW QUESTION 20

An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen

Which combination of algorithms would provide the appropriate insights? (Select TWO )

- A. The factorization machines (FM) algorithm
- B. The Latent Dirichlet Allocation (LDA) algorithm
- C. The principal component analysis (PCA) algorithm
- D. The k-means algorithm
- E. The Random Cut Forest (RCF) algorithm

**Answer:** CD

#### Explanation:

The PCA and K-means algorithms are useful in collection of data using census form.

#### NEW QUESTION 24

A Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences The aim is to then run Word2Vec to generate embeddings of the sentences and enable different types of predictions

Here is an example from the dataset

"The quck BROWN FOX jumps over the lazy dog "

Which of the following are the operations the Specialist needs to perform to correctly sanitize and prepare the data in a repeatable manner? (Select THREE)

- A. Perform part-of-speech tagging and keep the action verb and the nouns only
- B. Normalize all words by making the sentence lowercase
- C. Remove stop words using an English stopword dictionary.
- D. Correct the typography on "quck" to "quick."
- E. One-hot encode all words in the sentence
- F. Tokenize the sentence into words.

**Answer:** ABD

#### NEW QUESTION 27

While working on a neural network project, a Machine Learning Specialist discovers that some features in the data have very high magnitude resulting in this data being weighted more in the cost function What should the Specialist do to ensure better convergence during backpropagation?

- A. Dimensionality reduction
- B. Data normalization
- C. Model regularization
- D. Data augmentation for the minority class

**Answer:** D

#### NEW QUESTION 30

A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist.

Which machine learning model type should the Specialist use to accomplish this task?

- A. Linear regression
- B. Classification
- C. Clustering

D. Reinforcement learning

**Answer:** B

**Explanation:**

The goal of classification is to determine to which class or category a data point (customer in our case) belongs to. For classification problems, data scientists would use historical data with predefined target variables AKA labels (churner/non-churner) – answers that need to be predicted – to train an algorithm. With classification,

businesses can answer the following questions:

- > Will this customer churn or not?
- > Will a customer renew their subscription?
- > Will a user downgrade a pricing plan?
- > Are there any signs of unusual customer behavior?

**NEW QUESTION 31**

An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models

During the model evaluation the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images

Which of the following should be used to resolve this issue? (Select TWO)

- A. Add vanishing gradient to the model
- B. Perform data augmentation on the training data
- C. Make the neural network architecture complex.
- D. Use gradient checking in the model
- E. Add L2 regularization to the model

**Answer:** BD

**NEW QUESTION 36**

A Machine Learning Specialist is using Apache Spark for pre-processing training data As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE )

- A. Download the AWS SDK for the Spark environment
- B. Install the SageMaker Spark library in the Spark environment.
- C. Use the appropriate estimator from the SageMaker Spark Library to train a model.
- D. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
- E. Use the sageMakerMode
- F. transform method to get inferences from the model hosted in SageMaker
- G. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

**Answer:** DEF

**NEW QUESTION 39**

A Machine Learning Specialist is configuring automatic model tuning in Amazon SageMaker

When using the hyperparameter optimization feature, which of the following guidelines should be followed to improve optimization?

Choose the maximum number of hyperparameters supported by

- A. Amazon SageMaker to search the largest number of combinations possible
- B. Specify a very large hyperparameter range to allow Amazon SageMaker to cover every possible value.
- C. Use log-scaled hyperparameters to allow the hyperparameter space to be searched as quickly as possible
- D. Execute only one hyperparameter tuning job at a time and improve tuning through successive rounds of experiments

**Answer:** C

**NEW QUESTION 43**

A monitoring service generates 1 TB of scale metrics record data every minute A Research team performs queries on this data using Amazon Athena The queries run slowly due to the large volume of data, and the team requires better performance

How should the records be stored in Amazon S3 to improve query performance?

- A. CSV files
- B. Parquet files
- C. Compressed JSON
- D. RecordIO

**Answer:** B

**NEW QUESTION 48**

A Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors While exploring the data, the Specialist notices that the magnitude of the input features vary greatly The Specialist does not want variables with a larger magnitude to dominate the model What should the Specialist do to prepare the data for model training'?

- A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
- B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude
- C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
- D. Apply the orthogonal sparse Diagram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

**Answer:** C

#### NEW QUESTION 52

A Machine Learning Specialist is building a prediction model for a large number of features using linear models, such as linear regression and logistic regression. During exploratory data analysis, the Specialist observes that many features are highly correlated with each other. This may make the model unstable. What should be done to reduce the impact of having such a large number of features?

- A. Perform one-hot encoding on highly correlated features.
- B. Use matrix multiplication on highly correlated features.
- C. Create a new feature space using principal component analysis (PCA).
- D. Apply the Pearson correlation coefficient.

**Answer:** C

#### NEW QUESTION 57

A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor, and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset. Which tool should be used to improve the validation accuracy?

- A. Amazon Comprehend syntax analysts and entity detection.
- B. Amazon SageMaker BlazingText allow mode.
- C. Natural Language Toolkit (NLTK) stemming and stop word removal.
- D. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizers.

**Answer:** D

#### NEW QUESTION 59

A Data Scientist wants to gain real-time insights into a data stream of GZIP files. Which solution would allow the use of SQL to query the stream with the LEAST latency?

- A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.
- B. AWS Glue with a custom ETL script to transform the data.
- C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.
- D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

**Answer:** A

#### NEW QUESTION 60

A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:

- \* Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
- \* Support event-driven ETL pipelines.
- \* Provide a quick and easy way to understand metadata. Which approach meets these requirements?

- A. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
- B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.
- C. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
- D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.

**Answer:** B

#### NEW QUESTION 63

A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100,000 images per defect type for training. During the initial training of the image classification model, the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90%. It is known that human-level performance for this type of image classification is around 90%. What should the Specialist consider to fix this issue?

- A. A longer training time.
- B. Making the network larger.
- C. Using a different optimizer.
- D. Using some form of regularization.

**Answer:** D

#### NEW QUESTION 65

A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked. Which services are integrated with Amazon SageMaker to track this information? (Select TWO.)

- A. AWS CloudTrail
- B. AWS Health
- C. AWS Trusted Advisor
- D. Amazon CloudWatch



E. AWS Config

**Answer:** AD

#### NEW QUESTION 69

During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?

- A. The class distribution in the dataset is imbalanced
- B. Dataset shuffling is disabled
- C. The batch size is too big
- D. The learning rate is very high

**Answer:** D

#### NEW QUESTION 70

A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC. Why is the ML Specialist not seeing the instance visible in the VPC?

- A. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account, but they run outside of VPCs.
- B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
- C. Amazon SageMaker notebook instances are based on EC2 instances running within AWS service accounts.
- D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

**Answer:** C

#### NEW QUESTION 74

A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data

Which type of data repository is the MOST cost-effective solution?

- A. An Amazon EBS-backed Amazon EC2 instance with hourly directories
- B. An Amazon RDS database with hourly table partitions
- C. An Amazon S3 data lake with hourly object prefixes
- D. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes

**Answer:** C

#### NEW QUESTION 78

A Machine Learning Specialist is training a model to identify the make and model of vehicles in images The Specialist wants to use transfer learning and an existing model trained on images of general objects The Specialist collated a large custom dataset of pictures containing different vehicle makes and models

- A. Initialize the model with random weights in all layers including the last fully connected layer
- B. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
- C. Initialize the model with random weights in all layers and replace the last fully connected layer
- D. Initialize the model with pre-trained weights in all layers including the last fully connected layer

**Answer:** B

#### NEW QUESTION 80

A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket A Machine Learning Specialist wants to use SQL to run queries on this data. Which solution requires the LEAST effort to be able to query this data?

- A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
- B. Use AWS Glue to catalogue the data and Amazon Athena to run queries
- C. Use AWS Batch to run ETL on the data and Amazon Aurora to run the quenes
- D. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

**Answer:** D

#### NEW QUESTION 85

A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results The Specialist wants to use hyperparameter optimization to increase the model's accuracy

Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

- A. Launch multiple training jobs in parallel with different hyperparameters
- B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters
- C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
- D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter

**Answer:** B

NEW QUESTION 87

.....

## Relate Links

**100% Pass Your MLS-C01 Exam with ExamBible Prep Materials**

<https://www.exambible.com/MLS-C01-exam/>

## Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>