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.leads4pass.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

QUESTION 1

An ecommerce company wants to train a large image classification model with 10,000 classes. The company runs multiple model training iterations and needs to minimize operational overhead and cost. The company also needs to avoid loss of work and model retraining.

Which solution will meet these requirements?

A. Create the training jobs as AWS Batch jobs that use Amazon EC2 Spot Instances in a managed compute environment.

B. Use Amazon EC2 Spot Instances to run the training jobs. Use a Spot Instance interruption notice to save a snapshot of the model to Amazon S3 before an instance is terminated.

C. Use AWS Lambda to run the training jobs. Save model weights to Amazon S3.

D. Use managed spot training in Amazon SageMaker. Launch the training jobs with checkpointing enabled.

Correct Answer: D

https://docs.aws.amazon.com/sagemaker/latest/dg/model-managed-spot-training.html

Managed spot training can optimize the cost of training models up to 90% over on-demand instances. SageMaker manages the Spot interruptions on your behalf.

"Spot instances can be interrupted, causing jobs to take longer to start or finish. You can configure your managed spot training job to use checkpoints. SageMaker copies checkpoint data from a local path to Amazon S3. When the job is restarted, SageMaker copies the data from Amazon S3 back into the local path. The training job can then resume from the last checkpoint instead of restarting."

QUESTION 2

A data engineer needs to provide a team of data scientists with the appropriate dataset to run machine learning training jobs. The data will be stored in Amazon S3. The data engineer is obtaining the data from an Amazon Redshift database and is using join queries to extract a single tabular dataset. A portion of the schema is as follows:

1.

TransactionTimestamp (Timestamp)

2.

CardName (Varchar)

3.

CardNo (Varchar)

The data engineer must provide the data so that any row with a CardNo value of NULL is removed. Also, the TransactionTimestamp column must be separated into a TransactionDate column and a TransactionTime column. Finally, the CardName column must be renamed to NameOnCard.

The data will be extracted on a monthly basis and will be loaded into an S3 bucket. The solution must minimize the

effort that is needed to set up infrastructure for the ingestion and transformation. The solution also must be automated and must minimize the load on the Amazon Redshift cluster.

Which solution meets these requirements?

A. Set up an Amazon EMR cluster. Create an Apache Spark job to read the data from the Amazon Redshift cluster and transform the data. Load the data into the S3 bucket. Schedule the job to run monthly.

B. Set up an Amazon EC2 instance with a SQL client tool, such as SQL Workbench/J, to query the data from the Amazon Redshift cluster directly Export the resulting dataset into a file. Upload the file into the S3 bucket. Perform these tasks monthly.

C. Set up an AWS Glue job that has the Amazon Redshift cluster as the source and the S3 bucket as the destination. Use the built-in transforms Filter, Map, and RenameField to perform the required transformations. Schedule the job to run monthly.

D. Use Amazon Redshift Spectrum to run a query that writes the data directly to the S3 bucket. Create an AWS Lambda function to run the query monthly.

Correct Answer: C

https://docs.aws.amazon.com/glue/latest/dg/aws-glue-programming-python-transforms.html

QUESTION 3

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.

Correct Answer: D

AWS Glue can connect with DynamoDB and join both data sets together via Glue Studio. Requiring minimal overheads

QUESTION 4

An energy company has wind turbines, weather stations, and solar panels that generate telemetry data. The company wants to perform predictive maintenance on these devices. The devices are in various locations and have unstable internet connectivity.

A team of data scientists is using the telemetry data to perform machine learning (ML) to conduct anomaly detection and predict maintenance before the devices start to deteriorate. The team needs a scalable, secure, high-velocity data ingestion mechanism. The team has decided to use Amazon S3 as the data storage location.

Which approach meets these requirements?

A. Ingest the data by using an HTTP API call to a web server that is hosted on Amazon EC2. Set up EC2 instances in an Auto Scaling configuration behind an Elastic Load Balancer to load the data into Amazon S3.

B. Ingest the data over Message Queuing Telemetry Transport (MQTT) to AWS IoT Core. Set up a rule in AWS IoT Core to use Amazon Kinesis Data Firehose to send data to an Amazon Kinesis data stream that is configured to write to an S3 bucket.

C. Ingest the data over Message Queuing Telemetry Transport (MQTT) to AWS IoT Core. Set up a rule in AWS IoT Core to direct all MQTT data to an Amazon Kinesis Data Firehose delivery stream that is configured to write to an S3 bucket.

D. Ingest the data over Message Queuing Telemetry Transport (MQTT) to Amazon Kinesis data stream that is configured to write to an S3 bucket.

Correct Answer: C

QUESTION 5

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

Correct Answer: C

In natural language processing, the latent Dirichlet allocation (LDA) is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

Amazon SageMaker Random Cut Forest (RCF) is an unsupervised algorithm for detecting anomalous data points within a data set

Neural network is used for image detection.

QUESTION 6

A car company is developing a machine learning solution to detect whether a car is present in an image. The image dataset consists of one million images. Each image in the dataset is 200 pixels in height by 200 pixels in width. Each image is labeled as either having a car or not having a car.

Which architecture is MOST likely to produce a model that detects whether a car is present in an image with the highest accuracy?

A. Use a deep convolutional neural network (CNN) classifier with the images as input. Include a linear output layer that outputs the probability that an image contains a car.

B. Use a deep convolutional neural network (CNN) classifier with the images as input. Include a softmax output layer that outputs the probability that an image contains a car.

C. Use a deep multilayer perceptron (MLP) classifier with the images as input. Include a linear output layer that outputs the probability that an image contains a car.

D. Use a deep multilayer perceptron (MLP) classifier with the images as input. Include a softmax output layer that outputs the probability that an image contains a car.

Correct Answer: D

QUESTION 7

A data scientist wants to build a financial trading bot to automate investment decisions. The financial bot should recommend the quantity and price of an asset to buy or sell to maximize long-term profit. The data scientist will continuously stream financial transactions to the bot for training purposes. The data scientist must select the appropriate machine learning (ML) algorithm to develop the financial trading bot.

Which type of ML algorithm will meet these requirements?

A. Supervised learning

B. Unsupervised learning

C. Semi-supervised learning

D. Reinforcement learning

Correct Answer: C

QUESTION 8

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 form 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 lhat 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

Correct Answer: C

QUESTION 9

A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena The dataset contains more than 800.000 records stored as plaintext CSV files Each record contains 200 columns and is approximately 1 5 MB in size Most queries will span 5 to 10 columns only

How should the Machine Learning Specialist transform the dataset to minimize query runtime?

A. Convert the records to Apache Parquet format

B. Convert the records to JSON format

C. Convert the records to GZIP CSV format

D. Convert the records to XML format

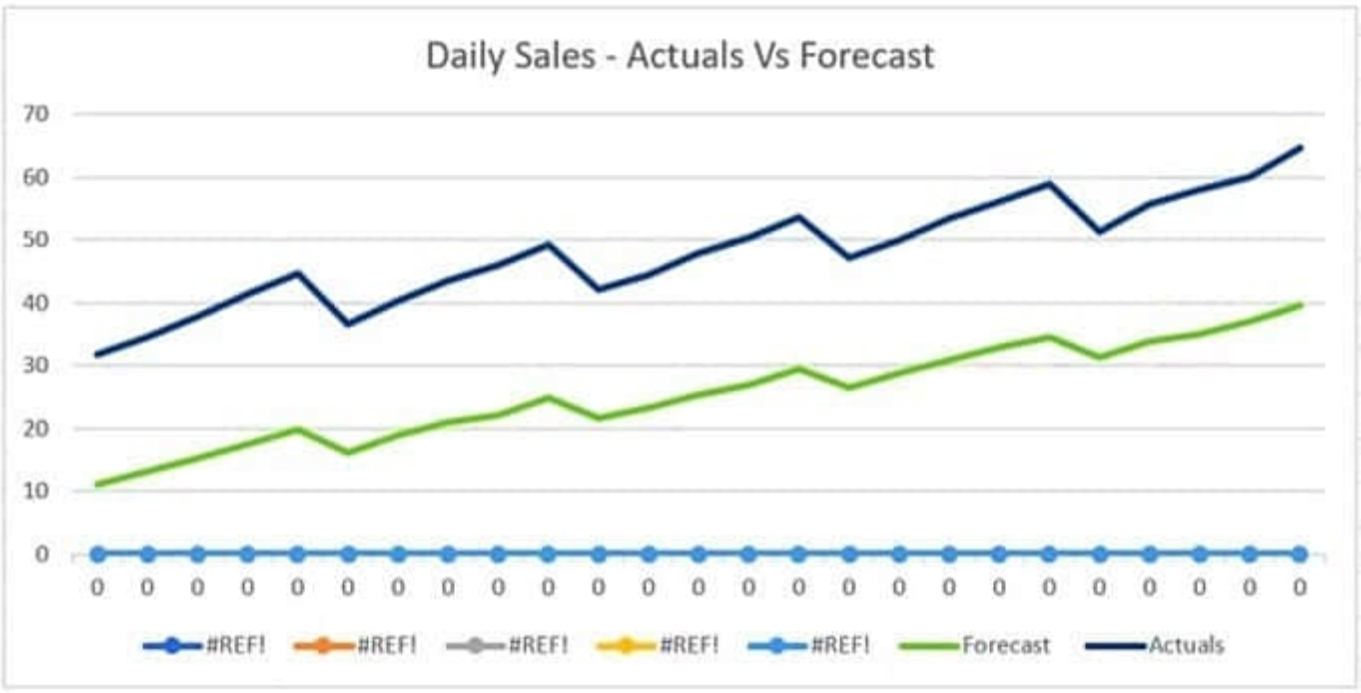
Correct Answer: A

Using compressions will reduce the amount of data scanned by Amazon Athena, and also reduce your S3 bucket storage. It\\'s a Win-Win for your AWS bill. Supported formats: GZIP, LZO, SNAPPY (Parquet) and ZLIB.

Reference: https://www.cloudforecast.io/blog/using-parquet-on-athena-to-save-money-on-aws/

QUESTION 10

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.

Correct Answer: A

QUESTION 11

A company is using Amazon Textract to extract textual data from thousands of scanned text-heavy legal documents daily. The company uses this information to process loan applications automatically. Some of the documents fail business validation and are returned to human reviewers, who investigate the errors. This activity increases the time to process the loan applications.

What should the company do to reduce the processing time of loan applications?

A. Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth. Perform a manual review on those words before performing a business validation.

B. Use an Amazon Textract synchronous operation instead of an asynchronous operation.

C. Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (Amazon A2I). Perform a manual review on those words before performing a business validation.

D. Use Amazon Rekognition\\'s feature to detect text in an image to extract the data from scanned images. Use this information to process the loan applications.

Correct Answer: C

QUESTION 12

An automotive company uses computer vision in its autonomous cars. The company trained its object detection models successfully by using transfer learning from a convolutional neural network (CNN). The company trained the models by

using PyTorch through the Amazon SageMaker SDK.

The vehicles have limited hardware and compute power. The company wants to optimize the model to reduce memory, battery, and hardware consumption without a significant sacrifice in accuracy.

Which solution will improve the computational efficiency of the models?

A. Use Amazon CloudWatch metrics to gain visibility into the SageMaker training weights, gradients, biases, and activation outputs. Compute the filter ranks based on the training information. Apply pruning to remove the low-ranking filters. Set new weights based on the pruned set of filters. Run a new training job with the pruned model.

B. Use Amazon SageMaker Ground Truth to build and run data labeling workflows. Collect a larger labeled dataset with the labelling workflows. Run a new training job that uses the new labeled data with previous training data.

C. Use Amazon SageMaker Debugger to gain visibility into the training weights, gradients, biases, and activation outputs. Compute the filter ranks based on the training information. Apply pruning to remove the low-ranking filters. Set the new weights based on the pruned set of filters. Run a new training job with the pruned model.

D. Use Amazon SageMaker Model Monitor to gain visibility into the ModelLatency metric and OverheadLatency metric of the model after the company deploys the model. Increase the model learning rate. Run a new training job.

Correct Answer: C

QUESTION 13

A manufacturing company wants to use machine learning (ML) to automate quality control in its facilities. The facilities are in remote locations and have limited internet connectivity. The company has 20 ?? of training data that consists of labeled images of defective product parts. The training data is in the corporate on-premises data center.

The company will use this data to train a model for real-time defect detection in new parts as the parts move on a conveyor belt in the facilities. The company needs a solution that minimizes costs for compute infrastructure and that maximizes the scalability of resources for training. The solution also must facilitate the company\\'s use of an ML model in the low-connectivity environments.

Which solution will meet these requirements?

A. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Deploy the model on a SageMaker hosting services endpoint.

B. Train and evaluate the model on premises. Upload the model to an Amazon S3 bucket. Deploy the model on an

Amazon SageMaker hosting services endpoint.

C. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.

D. Train the model on premises. Upload the model to an Amazon S3 bucket. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.

Correct Answer: C

Using S3 for scalable training and SageMaker Neo for compiling model for edge devices

QUESTION 14

A machine learning specialist is developing a regression model to predict rental rates from rental listings. A variable named Wall_Color represents the most prominent exterior wall color of the property. The following is the sample data, excluding all other variables:

Property_ID	Wall_Color
1000	Red
1001	White
1002	Green

The specialist chose a model that needs numerical input data.

Which feature engineering approaches should the specialist use to allow the regression model to learn from the Wall_Color data? (Choose two.)

A. Apply integer transformation and set Red = 1, White = 5, and Green = 10.

B. Add new columns that store one-hot representation of colors.

C. Replace the color name string by its length.

D. Create three columns to encode the color in RGB format.

E. Replace each color name by its training set frequency.

Correct Answer: BE

QUESTION 15

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

Correct Answer: D

Reference: https://towardsdatascience.com/deep-learning-personal-notes-part-1-lesson-2-8946fe970b95