Vendor: Amazon

Exam Code: MLA-C01

Exam Name: AWS Certified Machine Learning Engineer - Associate MLA-C01 Exam

Version: DEMO

QUESTION 1
Case Study
A company is building a web-based AI application by using Amazon SageMaker. The application will provide the following capabilities and features: ML experimentation, training, a central model registry, model deployment, and model monitoring.
The application must ensure secure and isolated use of training data during the ML lifecycle. The training data is stored in Amazon S3.
The company needs to use the central model registry to manage different versions of models in the application.
Which action will meet this requirement with the LEAST operational overhead?

A. Create a separate Amazon Elastic Container Registry (Amazon ECR) repository for each model.
B. Use Amazon Elastic Container Registry (Amazon ECR) and unique tags for each model version.
C. Use the SageMaker Model Registry and model groups to catalog the models.
D. Use the SageMaker Model Registry and unique tags for each model version.

Answer: C

QUESTION 2
Case Study
A company is building a web-based AI application by using Amazon SageMaker. The application will provide the following capabilities and features: ML experimentation, training, a central model registry, model deployment, and model monitoring.
The application must ensure secure and isolated use of training data during the ML lifecycle. The training data is stored in Amazon S3.
The company is experimenting with consecutive training jobs.
How can the company MINIMIZE infrastructure startup times for these jobs?

A. Use Managed Spot Training.
B. Use SageMaker managed warm pools.
C. Use SageMaker Training Compiler.
D. Use the SageMaker distributed data parallelism (SMDDP) library.

Answer: B

QUESTION 3
Case Study
A company is building a web-based AI application by using Amazon SageMaker. The application will provide the following capabilities and features: ML experimentation, training, a central model registry, model deployment, and model monitoring.
The application must ensure secure and isolated use of training data during the ML lifecycle. The training data is stored in Amazon S3.
The company must implement a manual approval-based workflow to ensure that only approved models can be deployed to production endpoints.
Which solution will meet this requirement?

A. Use SageMaker Experiments to facilitate the approval process during model registration.
B. Use SageMaker ML Lineage Tracking on the central model registry. Create tracking entities for the approval process.
C. Use SageMaker Model Monitor to evaluate the performance of the model and to manage the approval.
D. Use SageMaker Pipelines. When a model version is registered, use the AWS SDK to change the

approval status to "Approved."

Answer: D

QUESTION 4
Case Study
A company is building a web-based AI application by using Amazon SageMaker. The application will provide the following capabilities and features: ML experimentation, training, a central model registry, model deployment, and model monitoring.
The application must ensure secure and isolated use of training data during the ML lifecycle. The training data is stored in Amazon S3.
The company needs to run an on-demand workflow to monitor bias drift for models that are deployed to real-time endpoints from the application.
Which action will meet this requirement?

A. Configure the application to invoke an AWS Lambda function that runs a SageMaker Clarify job.
B. Invoke an AWS Lambda function to pull the sagemaker-model-monitor-analyzer built-in SageMaker image.
C. Use AWS Glue Data Quality to monitor bias.
D. Use SageMaker notebooks to compare the bias.

Answer: A

QUESTION 5
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

Explanation:
Amazon EMR with Spark is an excellent choice for aggregating, processing, and transforming large datasets from multiple sources (e.g., Amazon S3 and on-premises MySQL database). Spark jobs can handle both structured and unstructured. While Lake Formation is great for managing data lakes, it doesn't provide the ETL and data processing capabilities required to aggregate and transform datasets from multiple sources.

QUESTION 6
Hotspot Question

A company stores historical data in .csv files in Amazon S3. Only some of the rows and columns

in the .csv files are populated. The columns are not labeled. An ML engineer needs to prepare and store the data so that the company can use the data to train ML models.
Select and order the correct steps from the following list to perform this task. Each step should be selected one time or not at all. (Select and order three.)

```
- Create an Amazon SageMaker batch transform job for data cleaning and
feature engineering.
- Store the resulting data back in Amazon S3.
- Use Amazon Athena to infer the schemas and available columns.
- Use AWS Glue crawlers to infer the schemas and available columns.
- Use AWS Glue DataBrew for data cleaning and feature engineering.
```

Step 1: Select...
- Select...
- Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
- Store the resulting data back in Amazon S3.
- Use Amazon Athena to infer the schemas and available columns.
- Use AWS Glue crawlers to infer the schemas and available columns.
- Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 2: Select...
- Select...
- Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
- Store the resulting data back in Amazon S3.
- Use Amazon Athena to infer the schemas and available columns.
- Use AWS Glue crawlers to infer the schemas and available columns.
- Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 3: Select...
- Select...
- Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
- Store the resulting data back in Amazon S3.
- Use Amazon Athena to infer the schemas and available columns.
- Use AWS Glue crawlers to infer the schemas and available columns.
- Use AWS Glue DataBrew for data cleaning and feature engineering.

Answer:

Step 1: Select...
> Select...
> Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
> Store the resulting data back in Amazon S3.
> Use Amazon Athena to infer the schemas and available columns.
> **Use AWS Glue crawlers to infer the schemas and available columns.**
> Use AWS Glue DataBrew for data cleaning and feature engineering.

Step 2: Select...
> Select...
> Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
> Store the resulting data back in Amazon S3.
> Use Amazon Athena to infer the schemas and available columns.
> Use AWS Glue crawlers to infer the schemas and available columns.
> **Use AWS Glue DataBrew for data cleaning and feature engineering.**

Step 3: Select...
> Select...
> Create an Amazon SageMaker batch transform job for data cleaning and feature engineering.
> **Store the resulting data back in Amazon S3.**
> Use Amazon Athena to infer the schemas and available columns.
> Use AWS Glue crawlers to infer the schemas and available columns.
> Use AWS Glue DataBrew for data cleaning and feature engineering.

QUESTION 7
An ML engineer has trained a neural network by using stochastic gradient descent (SGD). The neural network performs poorly on the test set. The values for training loss and validation loss remain high and show an oscillating pattern. The values decrease for a few epochs and then increase for a few epochs before repeating the same cycle.
What should the ML engineer do to improve the training process?

A. Introduce early stopping.
B. Increase the size of the test set.
C. Increase the learning rate.
D. Decrease the learning rate.

Answer: D

QUESTION 8
An ML engineer needs to process thousands of existing CSV objects and new CSV objects that are uploaded. The CSV objects are stored in a central Amazon S3 bucket and have the same number of columns. One of the columns is a transaction date. The ML engineer must query the data based on the transaction date.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use an Amazon Athena CREATE TABLE AS SELECT (CTAS) statement to create a table based on the transaction date from data in the central S3 bucket. Query the objects from the table.
B. Create a new S3 bucket for processed data. Set up S3 replication from the central S3 bucket to the new S3 bucket. Use S3 Object Lambda to query the objects based on transaction date.
C. Create a new S3 bucket for processed data. Use AWS Glue for Apache Spark to create a job to query the CSV objects based on transaction date. Configure the job to store the results in the new S3 bucket. Query the objects from the new S3 bucket.

D. Create a new S3 bucket for processed data. Use Amazon Data Firehose to transfer the data from the central S3 bucket to the new S3 bucket. Configure Firehose to run an AWS Lambda function to query the data based on transaction date.

Answer: A

QUESTION 9
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

QUESTION 10
A company needs to run a batch data-processing job on Amazon EC2 instances. The job will run during the weekend and will take 90 minutes to finish running. The processing can handle interruptions. The company will run the job every weekend for the next 6 months.
Which EC2 instance purchasing option will meet these requirements MOST cost-effectively?

A. Spot Instances
B. Reserved Instances
C. On-Demand Instances
D. Dedicated Instances

Answer: A

QUESTION 11
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 model. Transfer the copy to Account B.
B. Create a resource-based IAM policy. Use the Amazon Comprehend ImportModel API operation to copy the model to Account B.
C. Use AWS DataSync to replicate the model from Account A to Account B.
D. Create an AWS Site-to-Site VPN connection between Account A and Account ?to transfer the model.

Answer: B

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

★ More than **99,900** Satisfied Customers Worldwide.

★ Average **99.9%** Success Rate.

★ **Free Update** to match latest and real exam scenarios.

★ **Instant Download** Access! No Setup required.

★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.

★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.

★ **100%** Guaranteed Success or **100%** Money Back Guarantee.

★ **Fast**, helpful support **24x7**.

View list of all certification exams: http://www.lead2pass.com/all-products.html

10% Discount Coupon Code: ASTR14