QUESTION 158
A company needs to quickly make sense of a large amount of data and gain insight from it. The data is in different formats, the schemas change frequently, and new data sources are added regularly. The company wants to use AWS services to explore multiple data sources, suggest schemas, and enrich and transform the data. The solution should require the least possible coding effort for the data flows and the least possible infrastructure management.
Which combination of AWS services will meet these requirements?

A. Amazon EMR for data discovery, enrichment, and transformation
    Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
    Amazon QuickSight for reporting and getting insights
B. Amazon Kinesis Data Analytics for data ingestion
    Amazon EMR for data discovery, enrichment, and transformation
    Amazon Redshift for querying and analyzing the results in Amazon S3
C. AWS Glue for data discovery, enrichment, and transformation
    Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
    Amazon QuickSight for reporting and getting insights
D. AWS Data Pipeline for data transfer
    AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation
    Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
    Amazon QuickSight for reporting and getting insights

Answer: A

QUESTION 159
A company is converting a large number of unstructured paper receipts into images. The company wants to create a model based on natural language processing (NLP) to find relevant entities such as date, location, and notes, as well as some custom entities such as receipt numbers.
The company is using optical character recognition (OCR) to extract text for data labeling. However, documents are in different structures and formats, and the company is facing challenges with setting up the manual workflows for each document type. Additionally, the company trained a named entity recognition (NER) model for custom entity detection using a small sample size. This model has a very low confidence score and will require retraining with a large dataset.
Which solution for text extraction and entity detection will require the LEAST amount of effort?

A. Extract text from receipt images by using Amazon Textract.
    Use the Amazon SageMaker BlazingText algorithm to train on the text for entities and custom entities.
B. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace.
    Use the NER deep learning model to extract entities.
C. Extract text from receipt images by using Amazon Textract.

Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.
D. Extract text from receipt images by using a deep learning OCR model from the AWS Marketplace. Use Amazon Comprehend for entity detection, and use Amazon Comprehend custom entity recognition for custom entity detection.

Answer: C
Explanation:
https://aws.amazon.com/blogs/machine-learning/building-an-nlp-powered-search-index-with-amazon-textract-and-amazon-comprehend/

QUESTION 160
A company is building a predictive maintenance model based on machine learning (ML). The data is stored in a fully private Amazon S3 bucket that is encrypted at rest with AWS Key Management Service (AWS KMS) CMKs. An ML specialist must run data preprocessing by using an Amazon SageMaker Processing job that is triggered from code in an Amazon SageMaker notebook. The job should read data from Amazon S3, process it, and upload it back to the same S3 bucket. The preprocessing code is stored in a container image in Amazon Elastic Container Registry (Amazon ECR). The ML specialist needs to grant permissions to ensure a smooth data preprocessing workflow.
Which set of actions should the ML specialist take to meet these requirements?

A. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs, S3 read and write access to the relevant S3 bucket, and appropriate KMS and ECR permissions. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job from the notebook.
B. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Create an Amazon SageMaker Processing job with an IAM role that has read and write permissions to the relevant S3 bucket, and appropriate KMS and ECR permissions.
C. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs and to access Amazon ECR. Attach the role to the SageMaker notebook instance. Set up both an S3 endpoint and a KMS endpoint in the default VPC. Create Amazon SageMaker Processing jobs from the notebook.
D. Create an IAM role that has permissions to create Amazon SageMaker Processing jobs. Attach the role to the SageMaker notebook instance. Set up an S3 endpoint in the default VPC. Create Amazon SageMaker Processing jobs with the access key and secret key of the IAM user with appropriate KMS and ECR permissions.

Answer: D

QUESTION 161
A data scientist has been running an Amazon SageMaker notebook instance for a few weeks. During this time, a new version of Jupyter Notebook was released along with additional software updates. The security team mandates that all running SageMaker notebook instances use the latest security and software updates provided by SageMaker.
How can the data scientist meet this requirements?

A. Call the CreateNotebookInstanceLifecycleConfig API operation
B. Create a new SageMaker notebook instance and mount the Amazon Elastic Block Store (Amazon EBS) volume from the original instance
C. Stop and then restart the SageMaker notebook instance
D. Call the UpdateNotebookInstanceLifecycleConfig API operation

Answer: C
Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/nbi-software-updates.html

MLS-C01 Exam Dumps MLS-C01 Exam Questions MLS-C01 PDF Dumps MLS-C01 VCE Dumps

https://www.braindump2go.com/mls-c01.html

QUESTION 162
A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.
The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition. they need to be encrypted in transit. The library also must ensure that the images are not used to improve Amazon Rekognition as a service.
How should a machine learning specialist architect the solution to satisfy these requirements?

A. Enable server-side encryption on the S3 bucket.
Submit an AWS Support ticket to opt out of allowing images to be used for improving the service, and follow the process provided by AWS Support.
B. Switch to using an Amazon Rekognition collection to store the images.
Use the IndexFaces and SearchFacesByImage API operations instead of the CompareFaces API operation.
C. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare faces.
Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.
D. Enable client-side encryption on the S3 bucket.
Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.

Answer: B

QUESTION 163
A company is building a line-counting application for use in a quick-service restaurant. The company wants to use video cameras pointed at the line of customers at a given register to measure how many people are in line and deliver notifications to managers if the line grows too long. The restaurant locations have limited bandwidth for connections to external services and cannot accommodate multiple video streams without impacting other operations.
Which solution should a machine learning specialist implement to meet these requirements?

A. Install cameras compatible with Amazon Kinesis Video Streams to stream the data to AWS over the restaurant's existing internet connection.
Write an AWS Lambda function to take an image and send it to Amazon Rekognition to count the number of faces in the image.
Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
B. Deploy AWS DeepLens cameras in the restaurant to capture video.
Enable Amazon Rekognition on the AWS DeepLens device, and use it to trigger a local AWS Lambda function when a person is recognized. Use the Lambda function to send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
C. Build a custom model in Amazon SageMaker to recognize the number of people in an image.
Install cameras compatible with Amazon Kinesis Video Streams in the restaurant.
Write an AWS Lambda function to take an image. Use the SageMaker endpoint to call the model to count people.
Send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.
D. Build a custom model in Amazon SageMaker to recognize the number of people in an image.
Deploy AWS DeepLens cameras in the restaurant.
Deploy the model to the cameras.
Deploy an AWS Lambda function to the cameras to use the model to count people and send an Amazon Simple Notification Service (Amazon SNS) notification if the line is too long.

Answer: A

QUESTION 164
A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.

How can the ML team solve this issue?

A. Decrease the cooldown period for the scale-in activity.
    Increase the configured maximum capacity of instances.
B. Replace the current endpoint with a multi-model endpoint using SageMaker.
C. Set up Amazon API Gateway and AWS Lambda to trigger the SageMaker inference endpoint.
D. Increase the cooldown period for the scale-out activity.

Answer: A
Explanation:
https://aws.amazon.com/blogs/machine-learning/configuring-autoscaling-inference-endpoints-in-amazon-sagemaker/

QUESTION 165
A telecommunications company is developing a mobile app for its customers. The company is using an Amazon
SageMaker hosted endpoint for machine learning model inferences.
Developers want to introduce a new version of the model for a limited number of users who subscribed to a preview
feature of the app. After the new version of the model is tested as a preview, developers will evaluate its accuracy. If a
new version of the model has better accuracy, developers need to be able to gradually release the new version for all
users over a fixed period of time.
How can the company implement the testing model with the LEAST amount of operational overhead?

A. Update the ProductionVariant data type with the new version of the model by using the
    CreateEndpointConfig operation with the InitialVariantWeight parameter set to 0.
    Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the
    preview feature.
    When the new version of the model is ready for release, gradually increase InitialVariantWeight
    until all users have the updated version.
B. Configure two SageMaker hosted endpoints that serve the different versions of the model.
    Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the
    TargetVariant query string parameter.
    Reconfigure the app to send the TargetVariant query string parameter for users who subscribed
    to the preview feature.
    When the new version of the model is ready for release, change the ALB's routing algorithm to
    weighted until all users have the updated version.
C. Update the DesiredWeightsAndCapacity data type with the new version of the model by using the
    UpdateEndpointWeightsAndCapacities operation with the DesiredWeight parameter set to 0.
    Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the
    preview feature.
    When the new version of the model is ready for release, gradually increase DesiredWeight until
    all users have the updated version.
D. Configure two SageMaker hosted endpoints that serve the different versions of the model.
    Create an Amazon Route 53 record that is configured with a simple routing policy and that points
    to the current version of the model.
    Configure the mobile app to use the endpoint URL for users who subscribed to the preview
    feature and to use the Route 53 record for other users.
    When the new version of the model is ready for release, add a new model version endpoint to
    Route 53, and switch the policy to weighted until all users have the updated version.

Answer: D

QUESTION 166
A company offers an online shopping service to its customers. The company wants to enhance the site's security by
requesting additional information when customers access the site from locations that are different from their normal
location. The company wants to update the process to call a machine learning (ML) model to determine when additional
information should be requested.
The company has several terabytes of data from its existing ecommerce web servers containing the source IP
addresses for each request made to the web server. For authenticated requests, the records also contain the login

name of the requesting user.
Which approach should an ML specialist take to implement the new security feature in the web application?

A. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed
   access attempt.
   Use Amazon SageMaker to train a binary classification model using the factorization machines
   (FM) algorithm.
B. Use Amazon SageMaker to train a model using the IP Insights algorithm.
   Schedule updates and retraining of the model using new log data nightly.
C. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed
   access attempt.
   Use Amazon SageMaker to train a binary classification model using the IP Insights algorithm.
D. Use Amazon SageMaker to train a model using the Object2Vec algorithm.
   Schedule updates and retraining of the model using new log data nightly.

Answer: C

QUESTION 167
A retail company wants to combine its customer orders with the product description data from its product catalog. The structure and format of the records in each dataset is different. A data analyst tried to use a spreadsheet to combine the datasets, but the effort resulted in duplicate records and records that were not properly combined. The company needs a solution that it can use to combine similar records from the two datasets and remove any duplicates.
Which solution will meet these requirements?

A. Use an AWS Lambda function to process the data.
   Use two arrays to compare equal strings in the fields from the two datasets and remove any
   duplicates.
B. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog.
   Call the AWS Glue SearchTables API operation to perform a fuzzy-matching search on the two
   datasets, and cleanse the data accordingly.
C. Create AWS Glue crawlers for reading and populating the AWS Glue Data Catalog.
   Use the FindMatches transform to cleanse the data.
D. Create an AWS Lake Formation custom transform.
   Run a transformation for matching products from the Lake Formation console to cleanse the data
   automatically.

Answer: D
Explanation:
https://aws.amazon.com/lake-formation/features/

QUESTION 168
A company provisions Amazon SageMaker notebook instances for its data science team and creates Amazon VPC interface endpoints to ensure communication between the VPC and the notebook instances. All connections to the Amazon SageMaker API are contained entirely and securely using the AWS network. However, the data science team realizes that individuals outside the VPC can still connect to the notebook instances across the internet.
Which set of actions should the data science team take to fix the issue?

A. Modify the notebook instances' security group to allow traffic only from the CIDR ranges of the
   VPC.
   Apply this security group to all of the notebook instances' VPC interfaces.
B. Create an IAM policy that allows the sagemaker:CreatePresignedNotebookInstanceUrl and
   sagemaker:DescribeNotebookInstance actions from only the VPC endpoints.
   Apply this policy to all IAM users, groups, and roles used to access the notebook instances.
C. Add a NAT gateway to the VPC. Convert all of the subnets where the Amazon SageMaker
   notebook instances are hosted to private subnets.
   Stop and start all of the notebook instances to reassign only private IP addresses.

D. Change the network ACL of the subnet the notebook is hosted in to restrict access to anyone outside the VPC.

Answer: B
Explanation:
https://gmoein.github.io/files/Amazon%20SageMaker.pdf

QUESTION 169
A company will use Amazon SageMaker to train and host a machine learning (ML) model for a marketing campaign. The majority of data is sensitive customer data. The data must be encrypted at rest. The company wants AWS to maintain the root of trust for the master keys and wants encryption key usage to be logged.
Which implementation will meet these requirements?

A. Use encryption keys that are stored in AWS Cloud HSM to encrypt the ML data volumes, and to encrypt the model artifacts and data in Amazon S3.
B. Use SageMaker built-in transient keys to encrypt the ML data volumes. Enable default encryption for new Amazon Elastic Block Store (Amazon EBS) volumes.
C. Use customer managed keys in AWS Key Management Service (AWS KMS) to encrypt the ML data volumes, and to encrypt the model artifacts and data in Amazon S3.
D. Use AWS Security Token Service (AWS STS) to create temporary tokens to encrypt the ML storage volumes, and to encrypt the model artifacts and data in Amazon S3.

Answer: C

QUESTION 170
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.

Answer: C

QUESTION 171
A company sells thousands of products on a public website and wants to automatically identify products with potential durability problems. The company has 1.000 reviews with date, star rating, review text, review summary, and customer email fields, but many reviews are incomplete and have empty fields. Each review has already been labeled with the correct durability result.
A machine learning specialist must train a model to identify reviews expressing concerns over product durability. The first model needs to be trained and ready to review in 2 days.
What is the MOST direct approach to solve this problem within 2 days?

A. Train a custom classifier by using Amazon Comprehend.

B. Build a recurrent neural network (RNN) in Amazon SageMaker by using Gluon and Apache MXNet.
C. Train a built-in BlazingText model using Word2Vec mode in Amazon SageMaker.
D. Use a built-in seq2seq model in Amazon SageMaker.

Answer: B