AWS Machine Learning Associate: Hands On!

Study online at https://quizlet.com/_g5xoo5

Your company needs to store a massive volume of semi-structured data for future analysis. Which AWS storage solution would be the most cost-effective and flexible?	Amazon S3
Your application needs to provide low-latency access to frequently accessed data while minimizing costs. Which S3 storage class should you choose?	S3 Standard
Your company uses AWS Lambda to generate thumbnails from images uploaded to S3. The thumbnails are only needed temporarily. Which storage class should you choose for these thumbnails?	S3 One Zone-Infrequent Access
A large-scale real-time data processing application needs to analyze data from IoT devices. Which service should you use to ingest, process, and analyze this data in near real-time?	Amazon Kinesis Data Streams
Your company is implementing a data lake architecture on AWS. They need to enforce governance and control access to specific data products across multiple teams. Which approach should they use?	Data Mesh
A company needs to encrypt sensitive data at rest stored in Amazon S3. They also require auditability of key usage. Which encryption option should they choose?	SSE-KMS
Your organization needs to choose a data format for a large-scale ETL pipeline that processes batch data in Hadoop and Spark. The data should be optimized for analytics and efficient storage. Which format is most appropriate?	Parquet
Your application requires high-performance block storage for an EC2 instance that needs to persist data after termination. Which AWS storage solution should you choose?	Amazon EBS
A team is working on a shared development environment and needs a scalable, highly available file system that can be mounted on multiple EC2 instances across multiple Availability Zones. Which AWS service should they use?	Amazon EFS
During peak usage, your Kinesis Data Streams application is experiencing throttling errors. Which of the following actions should you take to resolve this issue?	Increase the number of shards
Your data engineering team is setting up a data pipeline to process large datasets using Hadoop and Spark on AWS. Which service would you choose to efficiently manage the cluster?	Amazon EMR
Your team is building a recommendation system for an e-commerce platform. Which SageMaker feature would help you efficiently handle and scale the training of your machine learning model?	SageMaker Training Jobs
A financial institution wants to detect anomalies in transaction data to identify potential fraud. Which AWS service could you use to deploy an outlier detection model, and monitor its performance over time?	SageMaker Model Monitor
Your company is developing a machine learning model to predict customer churn, but you're concerned about potential bias in the model's predictions. Which SageMaker feature helps detect and mitigate bias during model training?	SageMaker Clarify
You are tasked with deploying a machine learning model to edge devices in a factory. Which SageMaker feature would you use to optimize and deploy the model efficiently? SageMaker Neo	SageMaker Neo
You are managing a big data project that requires periodic processing of large datasets stored in Amazon S3. Which service	AWS Glue

would you use to schedule and manage the execution of these data processing jobs?	
A healthcare company needs to label large amounts of medical images for training a diagnostic AI model. Which SageMaker service would streamline this process by managing the labeling tasks?	SageMaker Ground Truth
A data scientist wants to experiment with different machine learning algorithms and compare their performance on a dataset. Which SageMaker feature provides an environment for such experimentation?	SageMaker Studio
You are working with a dataset that contains several missing values in a column representing customer ages. The data is relatively small, and the distribution of ages is skewed by a few extreme outliers. Which imputation technique should you choose to fill in the missing values?	Median Replacement
Your machine learning model is struggling to accurately predict a rare event, such as equipment failure, which occurs in less than 1% of your dataset. You want to generate synthetic samples to balance the dataset without simply duplicating existing data. Which technique should you use?	SMOTE (Synthetic Minority Oversampling Technique)
A media company wants to automatically generate subtitles for videos in real-time. Which AWS service is best suited for this task?	Amazon Transcribe
Your company needs to perform sentiment analysis on customer feedback provided in various languages. Which AWS service would best suit this requirement?	Amazon Comprehend
A financial institution needs to detect fraudulent transactions in real-time using historical transaction data. Which AWS service is most appropriate for this use case, while minimizing development effort?	Amazon Fraud Detector
A developer wants to build an application that can translate customer service chat logs from one language to another in real-time. Which AWS service would be most appropriate?	Amazon Translate
Your team needs to automatically label and classify images uploaded to an S3 bucket. Which AWS service should you use?	Amazon Rekognition
A news organization needs to automatically transcribe live news broadcasts into multiple languages simultaneously. Which combination of AWS services should they use?	Amazon Transcribe and Amazon Translate
A sports analytics company wants to identify player positions and movements in real-time during live broadcasts. Which AWS service should they utilize, while minimizing development effort?	Amazon Rekognition
You need to redact personally identifiable information (PII) from a large set of documents. Which AWS service should you use, while minimizing development and maintenance effort?	Amazon Comprehend
A retail company wants to monitor its key business metrics, such as daily sales and customer engagement, to quickly identify any unusual changes. Which AWS service should they use to minimize development effort?	Amazon Lookout for Metrics
An e-commerce platform wants to prevent "try before you buy" abuse, where users repeatedly purchase items and return them after use. The company has labeled data of previous fraudulent activities. Which AWS service should they use to address this issue?	Amazon Fraud Detector
Which SageMaker algorithm would be most appropriate for predicting the next word in a sequence of text?	Seq2Seq
	Amazon FSx for Lustre

AWS Machine Learning Associate: Hands On!

Study online at https://quizlet.com/_g5xoo5

Which SageMaker input mode is recommended for training when you need high-performance access to data stored in a single Availability Zone?	
Which SageMaker algorithm is best suited for clustering similar items without labeled data?	K-Means
You need to detect bounding boxes of multiple objects within an image and classify them. Which SageMaker algorithm should you use?	Object Detection
Which SageMaker algorithm is used to organize documents into topics in an unsupervised manner?	LDA
You're working with a sparse dataset and need to predict user-item interactions. Which SageMaker algorithm is most appropriate?	Factorization Machines
In SageMaker's XGBoost, which hyperparameter helps prevent overfitting by controlling the step size shrinkage?	The eta hyperparameter in XGBoost controls the step size shrinkage, which helps prevent overfitting by reducing the impact of individual trees.
You are tasked with training a SageMaker model to predict future values of a time series. Which algorithm should you choose?	DeepAR
Which input format is optimal for training a model using SageMaker's K-Means algorithm?	RecordIO-ProtoBUF
When using SageMaker's Object2Vec algorithm, what is the primary output?	Object2Vec produces low-dimensional dense embeddings of high-dimensional objects, useful for tasks like similarity searches and recommendation systems.
You are designing a model for sentiment analysis on customer reviews. Which neural network type is best suited for this task and why?	Recurrent Neural Network (RNN)
You have a dataset with 10,000 labeled images for training a CNN. To prevent overfitting, which technique should you consider implementing?	Apply dropout layers
You are deploying a deep learning model on an AWS EC2 instance. The model requires extensive GPU resources. Which instance type should you select for optimal performance?	p3.8xlarge is designed for deep learning tasks with powerful GPUs.
You are training a deep neural network and notice that the training accuracy is high, but the validation accuracy is much lower. What is the most likely issue?	Overfitting occurs when the model performs well on training data but fails to generalize to unseen data.
You have a model that predicts whether a customer will make a purchase. You notice a high rate of false negatives. Which metric should you focus on improving?	Recall
While tuning hyperparameters for a model in SageMaker, you want to optimize the model as efficiently as possible. Which strategy should you use of the given choices?	Bayesian optimization learns from each trial, making it more efficient.
You're using transfer learning to fine-tune a pre-trained BERT model for a custom NLP task. What is the most important hyperparameter to adjust during fine-tuning?	Learning rate
You need to process a massive dataset for training a deep learning model and want to speed up the training process. Which AWS service or feature should you consider using?	SageMaker Distributed Training
Your model training job is taking longer than expected on SageMaker. You suspect that the compute instances are not fully utilized. Which SageMaker feature can help you diagnose this issue?	SageMaker Debugger
When training a deep learning model, you notice that your model's loss is decreasing very slowly. What is a likely reason for this?	Learning rate is too low
When using a Transformer model for text classification, which component is responsible for capturing the relationships between different tokens in a sentence?	Self-Attention captures the relationships between tokens by considering the entire sentence context.

You are fine-tuning a pre-trained GPT model for a customer service chatbot. Which training technique is most appropriate if you want to adapt the model with minimal training data?	Freezing the initial layers and training the last few layers
Which AWS service would you use to quickly deploy a pre-trained model like GPT-J for text generation tasks?	Amazon SageMaker JumpStart
In a sequence-to-sequence transformer model used for translation, what is the role of the decoder?	The decoder generates the target sequence based on the encoded information from the source sequence.
When using masked self-attention in GPT models, what is the purpose of the mask?	The mask ensures that each token can only attend to past tokens, not future ones.
You are using AWS SageMaker JumpStart to deploy a model for text generation. Which model type would you choose for multilingual text generation?	Jurassic-2 is noted for its multilingual capabilities
During the training of a transformer model, which aspect allows the model to consider the position of each token within the sequence?	Positional encoding adds information about token positions.
You are tasked with summarizing large volumes of text data. Which model would you select from AWS's offerings?	Amazon Titan
In the context of self-attention, which matrices are used to compute the attention scores for each token?	Query and Key
What is a key advantage of using self-attention in transformer models over traditional RNNs?	Self-attention allows parallel processing, unlike RNNs.
You are developing a customer support chatbot using Amazon Bedrock. The chatbot needs to respond accurately to specific customer inquiries using recent company data. Which approach would be most effective for this scenario?	RAG is ideal for incorporating recent or proprietary information that the foundational model wasn't originally trained on, making it well-suited for a chatbot needing up-to-date responses.
You are using Amazon Bedrock to build an AI system for generating marketing content. The system must ensure the content generated adheres to brand guidelines and avoids inappropriate language. Which feature of Amazon Bedrock would best serve this purpose?	Content filtering guardrails
You're implementing a customer query system that uses Amazon Bedrock's LLM agents. The agent needs to use both a knowledge base and an external API for retrieving real-time data. How should you structure the agent to meet these requirements?	By using separate action groups, the LLM can be guided on when to use the knowledge base and when to call the API, making the system more flexible and maintainable.
While working with Amazon Bedrock, you need to frequently update your knowledge base with new company policies. What is the most efficient way to handle this?	Updating the vector database is faster and more efficient than fine-tuning when it comes to adding new information.
When setting up a new LLM agent in Amazon Bedrock, you want it to dynamically decide which tool to use based on the user's query. Which component is responsible for guiding the LLM in making this decision?	Action Groups
You're working on a Bedrock model that needs to incorporate sensitive customer data during fine-tuning. What steps should you take to ensure the security of this data?	Use a VPC and PrivateLink for data transfer during fine-tuning
Your team is planning to deploy a generative AI model in a serverless environment using Amazon Bedrock. What advantage does this provide?	Serverless environments in Amazon Bedrock provide automatic scaling, allowing your model to handle varying loads without the need for infrastructure management.
You are tasked with deploying a Bedrock model that requires high availability and low-latency responses. Which deployment option should you choose?	Provisioned Throughput
You are creating a custom model in Amazon Bedrock that will require additional training data. What type of training should you use if the data is not labeled?	Continued Pre-Training is used when the data is not labeled, allowing the model to familiarize itself with new information without the need for labeled training pairs.
	Action groups with a Code Interpreter

AWS Machine Learning Associate: Hands On!

Study online at https://quizlet.com/_g5xoo5

Your company wants to deploy a Bedrock-based AI system with the ability to write and execute Python code as part of its responses. Which feature should be enabled?	
Which step is most critical before deploying a SageMaker model using a Docker container?	Validate the model using SageMaker JumpStart
What is a primary use case for SageMaker Model Monitor?	Detecting data drift in deployed models
What is the correct way to structure a Dockerfile for a SageMaker training container?	Specify the training script using ENV SAGEMAKER_PROGRAM
You need to monitor a deployed model for feature attribution drift. Which metric should you monitor using SageMaker Model Monitor?	Normal Discounted Cumulative Gain (NDCG)
For which use case is SageMaker JumpStart most appropriate?	Deploying pre-trained models to pre-configured endpoints
Which deployment method in SageMaker is most suitable for handling infrequent or unpredictable traffic with the ability to scale down to zero?	Serverless Inference
Which AWS service is used to orchestrate steps in a machine learning pipeline, such as model training and deployment?	AWS Step Functions
What is a key advantage of using SageMaker's Managed Spot Training? Spot instances guarantee uninterrupted training.	Training cost can be reduced by up to 90%
Which SageMaker feature allows you to compare a new model's performance against an existing production model without impacting live traffic?	Shadow Testing
Which of the following is a key benefit of using Docker containers in Amazon SageMaker?	Containers allow the use of any script or algorithm regardless of runtime or language
You are deploying a SageMaker training job that handles sensitive data. Which of the following options would best ensure that the data is protected both at rest and in transit?	Enable KMS encryption for data at rest and TLS/SSL for data in transit
Which of the following is the most appropriate technique for ensuring that a specific SageMaker notebook instance has the least privilege required to perform its tasks?	Use IAM Access Analyzer to generate a custom least-privilege policy based on access activity.
What should you do to enable a SageMaker training job to access S3 resources in a secure manner?	Assign an appropriate IAM role to the SageMaker instance.
During a hyperparameter tuning job, you want to monitor the CPU and memory utilization of the SageMaker instance. Which service would you use for this?	AWS CloudWatch
You are configuring a SageMaker notebook that requires access to a private S3 bucket and other AWS services in the same VPC. What is the best practice to ensure secure and efficient access?	Using a VPC endpoint allows secure and efficient access to S3 and other AWS services without the need for internet access.
You are configuring network isolation for a SageMaker training job to prevent data leaks. What is a potential drawback of enabling network isolation?	The job will lose access to S3
Which encryption method is recommended for securing data in transit when using SageMaker?	TLS/SSL encryption
You need to grant cross-account access to a KMS-encrypted S3 bucket used in a SageMaker job. What should you configure?	Adding IAM roles from the other account to the KMS key policy allows cross-account access to the encrypted S3 bucket.
To enhance the security of your SageMaker environment, you decide to log and monitor all API calls. Which AWS service should you use?	AWS CloudTrail
You are tasked with setting up a highly secure SageMaker environment for training a sensitive dataset. What combination of practices would you implement?	Enabling network isolation, using KMS encryption, and assigning least-privilege IAM roles are all best practices for securing a sensitive SageMaker environment.
	CloudWatch Unified Agent

You are setting up CloudWatch Logs for an EC2 instance. Which agent should you use to collect both system-level metrics and log data?	
Your application is experiencing intermittent high latency. You need to analyze the logs to find patterns over time. Which CloudWatch feature would you use?	CloudWatch Logs Insights
You want to stream real-time log data from CloudWatch Logs to Amazon S3 for archival. What approach should you take?	Use CloudWatch Logs Subscriptions
You need to ensure that a specific S3 bucket is not publicly accessible. Which AWS service can help you audit this?	AWS Config
To optimize costs, you want to automatically shut down underutilized EC2 instances. Which CloudWatch feature should you use?	CloudWatch Alarms
Your organization wants to ensure that a specific instance type is used for all EC2 instances in a particular environment. Which AWS Config feature would you use?	AWS Config Rules
Which AWS service should you use to detect unusual activity, like unauthorized access attempts, in your account?	AWS CloudTrail Insights
Your company needs to monitor and manage costs associated with AWS usage. Which service should you use?	AWS Cost Explorer
To ensure your CloudWatch Logs are encrypted with a specific KMS key, what should you configure?	CloudWatch Logs Encryption
Your team uses Amazon QuickSight for business analytics. You want to enable anomaly detection on sales data to automatically identify unusual spikes in sales. Which feature of QuickSight should you use?	Machine Learning Insights
A team is deploying a machine learning model to production and is concerned about potential bias in the predictions. Which AWS service should they use to monitor for bias in real-time?	SageMaker Clarify
You are designing an ML system that must be able to scale during high traffic and ensure that model predictions are delivered within a set latency threshold. Which SageMaker deployment strategy should you consider?	Real Time Inference with Auto-scaling
Which of the following practices would help you minimize the environmental impact of your ML workloads on AWS?	Utilize Spot Instances and energy-efficient instance types
Your team needs to retrain an ML model whenever there is significant data drift. Which AWS service will help you automate the retraining process based on predefined conditions?	SageMaker Pipelines is correct because it allows for the automation of ML workflows, including retraining based on data drift detected by SageMaker Model Monitor.
A company wants to optimize its ML model to ensure minimal cost while maintaining adequate performance. What approach should they take?	Use Elastic Inference - Amazon Elastic Inference allows you to attach just the right amount of GPU-powered acceleration to any Amazon EC2 and Amazon SageMaker instance to reduce the cost of running deep learning inference by up to 75%. Amazon Elastic Inference supports TensorFlow, Apache MXNet, and ONNX models, with more frameworks coming soon.
You are tasked with ensuring the explainability of your ML model in production. Which tool should you use to generate explanations for predictions made by the model?	SageMaker Clarify
You need to establish a feedback loop that incorporates human review for certain predictions made by your model. Which AWS service would you use?	Amazon Augmented AI
You are setting up a CI/CD pipeline for machine learning in AWS and need to ensure model consistency between training and inference. Which service can help maintain feature consistency across these stages?	SageMaker Feature Store
	SageMaker Lineage Tracking and SageMaker Model Registry

Your team needs to track and audit the complete lineage of a machine learning model, from data ingestion to deployment. Which combination of AWS services should you use?	
A financial institution requires that all models be explainable to meet regulatory requirements. Which AWS service should be used to ensure that model predictions are understandable and compliant?	SageMaker Clarify
A healthcare company has developed a new version of its machine learning model that predicts patient readmission risk. The model is deployed on Amazon SageMaker, and the company wants to roll out this new version carefully to ensure it does not negatively impact the existing predictions made by the current model in production. The new model needs to be evaluated for its performance in a real-world setting with actual user traffic, but any potential errors should be minimized to avoid affecting patient care. The company wants to gradually introduce the new model to a small percentage of the traffic initially and then increase this percentage if the model performs well. It also needs a rollback strategy in case the new model performs worse than expected. Which deployment strategy should the company choose to deploy the new model version?	Canary Deployment is the most suitable strategy for this scenario. It allows the company to initially direct a small percentage of traffic to the new model version while monitoring its performance. If the new model performs well, the percentage of traffic can be gradually increased. This approach minimizes risk and allows for a controlled rollout, with the ability to quickly revert to the previous model version if issues arise.
A healthcare startup is developing a machine learning model to predict the risk of various diseases based on patient data. The dataset includes demographic information, such as age, gender, and ethnicity, as well as medical history. The data science team suspects that there might be a bias in the dataset related to the distribution of disease labels across different demographic groups. They decide to use the Difference in Proportion of Labels (DPL) metric to analyze this bias before training the model. Which type of analysis is best performed using the Difference in Proportion of Labels (DPL) metric, and why is it suitable for this scenario?	The best use of the Difference in Proportion of Labels (DPL) metric is to identify potential biases in the dataset by comparing the proportion of each disease label across different demographic groups. This analysis helps the data science team understand if there are disparities in the distribution of disease labels among different demographic groups, which is crucial for addressing bias in the machine learning model.
An e-commerce company uses an Amazon SageMaker endpoint to provide real-time product recommendations to users. Recently, the company's data science team noticed that the recommendation service is occasionally returning errors and taking longer to respond during peak traffic times. To ensure that the service maintains high availability and performance, the team needs to diagnose and troubleshoot these issues effectively. Which AWS service should the team primarily use to monitor the SageMaker endpoint, detect anomalies, and set up automatic alerts for any performance degradation or errors?	Amazon CloudWatch is the best tool for this scenario. It provides real-time monitoring of AWS resources, including SageMaker endpoints. The team can use CloudWatch to set up detailed logging, track performance metrics such as latency and error rates, and create alarms to automatically notify them when the service is not performing as expected. This allows for proactive management and quick response to any issues affecting the recommendation service.
A healthcare company is developing a machine learning model using Amazon SageMaker to predict whether patients are at risk of developing a rare but serious disease. The dataset is highly imbalanced, with only 1% of the records indicating that a patient has the disease. The model needs to ensure that high-risk patients are correctly identified as positive cases, as missing these cases could result in severe consequences for patient health. On the other hand, it is also important to minimize false positives to avoid unnecessary stress and medical expenses for patients. Which evaluation metric should the company use to best evaluate the performance of the model?	The F1 Score is the harmonic mean of precision and recall, making it a suitable metric when dealing with imbalanced datasets where both false positives and false negatives are important. In this scenario, identifying high-risk patients (true positives) while minimizing the number of false positives is crucial, making the F1 Score an appropriate choice as it balances these two aspects.
A mobile gaming company has developed a machine learning model using Amazon SageMaker to predict the likelihood of users making in-app purchases. The model is integrated into the game to offer personalized promotions to users based on their predicted purchasing behavior. The company expects a high volume of prediction requests that need to be processed immediately as users	Serverless Inference is the most appropriate choice for this scenario because it automatically scales to handle the fluctuating traffic without the need for the company to manage underlying infrastructure. It provides a cost-effective solution, especially during

interact with the game. However, user traffic fluctuates significantly throughout the day, with peak usage times in the evenings and low usage times in the early morning. Which SageMaker inference deployment endpoint should the company use to deploy this model?	low-traffic periods when fewer resources are needed. Additionally, it can handle real-time prediction requests during peak hours, ensuring low latency and immediate responses to user actions.
A retail company wants to predict the likelihood of customers making a purchase based on their online browsing behavior. The data available includes customer demographics, browsing history, product preferences, and time spent on different product pages. The dataset contains millions of records with both numerical and categorical features. The company needs a scalable solution that can handle large datasets efficiently and provide explainable results to understand feature importance. Which SageMaker built-in algorithm should the company use, and what key configuration settings should be applied to optimize the model performance for this scenario?	SageMaker XGBoost is a popular choice for binary classification tasks like predicting customer purchase likelihood. Setting the objective to 'binary' ensures the model is trained for binary classification. Adjusting max_depth to 6 helps prevent overfitting and improves model generalization. Scaling the pos_weight based on class imbalance helps address any bias in the dataset. Using 'ml.m5.4xlarge' instances for training and inference ensures efficient processing of large datasets.
A marketing company uses a machine learning model deployed on Amazon SageMaker to predict customer churn based on behavioral data. Recently, the company has noticed a significant drop in the model's performance metrics, such as accuracy and F1 score. After investigating, the data science team suspects that a change in the distribution of incoming data, possibly due to a new marketing campaign targeting different demographics, is affecting the model's performance. The team wants to analyze the impact of this change in data distribution on the model's predictions and ensure that the model remains unbiased across various demographic groups. Which tool or feature of Amazon SageMaker should the team use to address this issue?	SageMaker Clarify is specifically designed to detect bias in datasets and model predictions. It can be used to analyze how changes in data distribution affect the model's predictions and to ensure that the model remains fair across different demographic groups. This makes it the most suitable tool for the scenario described, where the impact of a data shift on model performance and fairness needs to be assessed.
A company is developing an object recognition model to identify various products in images taken from store shelves. The data science team is using Amazon SageMaker Data Wrangler to prepare the image data for training. The images contain a variety of lighting conditions, angles, and background clutter. Which data transformations should the team apply in SageMaker Data Wrangler to improve the performance of the object recognition model?	Resizing and cropping images to a fixed size helps standardize the input data for the model, while normalizing pixel values ensures consistency in the image data. Applying data augmentation techniques like flipping and rotation helps increase the diversity of the training data, which can improve the model's ability to generalize and perform well on unseen images.
A technology startup is developing a machine learning-based application using Amazon SageMaker. The application requires frequent model training on large datasets, but the workloads are not time-sensitive and can be interrupted if necessary. The startup is looking to optimize its infrastructure costs, as it has a limited budget and wants to maximize cost savings without compromising on the ability to scale when needed. Given these requirements, which purchasing option should the startup choose to optimize its costs for training machine learning models on Amazon SageMaker?	Spot Instances offer a cost-effective solution for running workloads that can tolerate interruptions. Since the startup's model training jobs are not time-sensitive and can be interrupted, Spot Instances are ideal. They provide significant cost savings (up to 90% off On-Demand prices) by utilizing unused EC2 capacity, making them the best choice for optimizing costs under the given conditions.
A customer service center records all phone calls for quality assurance purposes. The company wants to analyze these recordings to understand customer sentiment, identify frequently discussed topics, and extract key phrases related to customer complaints. The recordings are in various audio formats, and the company also wants to create text transcripts of these calls to store them in a searchable database for further analysis. Which combination of AWS services and configurations will best meet these requirements?	Using Amazon Transcribe to convert audio recordings to text ensures that the company can create searchable transcripts of the calls. Amazon Comprehend can then be used to analyze sentiment, detect entities, and extract key phrases from these transcripts, providing valuable insights into customer interactions. Storing the output in Amazon Elasticsearch allows for efficient search capabilities and further analysis of the data.
A retail company wants to store customer transaction data on AWS for analytical processing. The data is semi-structured and the analytics team requires efficient storage that supports schema evolution and provides high performance for batch and interactive analytics. Which data format should the company use?	Apache Parquet is a columnar storage format that is well-suited for storing semi-structured data with schema evolution requirements. It provides efficient storage, high performance for batch and interactive analytics, and supports complex nested data structures.

	This makes it a suitable choice for the retail company's analytical processing needs.
A media company is developing a deep learning model to automatically tag and classify videos based on their content. The training dataset consists of hundreds of terabytes of high-definition videos stored in a shared file system. The training jobs require high throughput and low-latency access to this large volume of data across multiple Amazon SageMaker instances. The data scientists want to ensure that the file system can handle this load efficiently. Which solutions could meet these requirements? (Select TWO)	Use Amazon FSx for Lustre Use Amazon EFS (Elastic File System) with Provisioned Throughput modes
A healthcare startup is developing a deep learning model using Amazon SageMaker to predict patient outcomes based on medical imaging data. They are using a deep neural network (DNN) with multiple hidden layers. After a few epochs of training, they notice that the accuracy on the training data starts to oscillate significantly, and the model is not improving. The team suspects that the model might be overfitting or learning too quickly due to high variance in the gradients during training. Which regularization technique should the team apply to stabilize the training process and improve model performance?	Apply Dropout regularization with a dropout rate of 0.5 to the hidden layers and use a lower learning rate to reduce the variance in gradients.
A financial services company wants to develop a sophisticated chatbot to assist customers with their investment inquiries. The chatbot should be able to provide detailed responses tailored to the company's specific product offerings, such as investment plans, retirement accounts, and market trends. The company has unique terminology and a specific style of communication that they want the chatbot to adopt to align with their brand voice. Additionally, the chatbot needs to be able to handle sensitive financial information securely and operate within the company's compliance guidelines. The company decides to use Amazon Bedrock to build this chatbot. Given these requirements, they want to use a language model that can be customized to their specific needs. Which approach should the company use to build the chatbot using Amazon Bedrock, and what key configurations should be applied to achieve the desired outcome?	Use a foundation model from Amazon Bedrock and fine-tune it with a dataset containing the company's product information, FAQs, and example customer interactions. Set the model parameters to include custom vocabularies and domain-specific terminology to ensure accurate and relevant responses.
A financial trading firm is developing a machine learning system to detect anomalies in stock market transactions in real-time. The system must continuously ingest streaming data from trading activities, process the data with minimal delay, and generate metrics for immediate analysis and decision-making. The firm wants to use AWS services to ensure the pipeline can handle high throughput and provide real-time insights without buffering delays. Which combination of AWS services should the firm implement to achieve this requirement?	Amazon Kinesis Data Streams with Managed Service for Apache Flink provides a scalable, real-time data processing solution that can handle high throughput and provide immediate insights without buffering delays. Managed Service for Apache Flink offers advanced stream processing capabilities, making it a suitable choice for real-time anomaly detection in stock market transactions.
A retail company has deployed several machine learning models using Amazon SageMaker to provide personalized product recommendations to customers on their website. The company has noticed that the accuracy of the recommendations seems to be degrading over time, potentially due to changes in customer behavior and preferences. The data science team wants to implement a solution to continuously monitor the models in production, detect any deviations in data quality or model performance, and automatically alert the team when significant issues are detected. Which AWS service should the team use to achieve this goal?	Amazon SageMaker Model Monitor
A healthcare organization is deploying a machine learning model on Amazon SageMaker to analyze patient data and predict potential health risks. Due to strict regulatory requirements, the organization must ensure that all data processing and model	

hosting occurs within a secure, isolated environment that prevents unauthorized access and protects sensitive patient information. The organization also wants to limit access to the model only to specific trusted IP addresses and other AWS services that are part of its internal network. Which AWS configuration should the organization implement to meet these security and isolation requirements?	Deploy the SageMaker model in a private subnet within a Virtual Private Cloud (VPC) and use security groups to restrict inbound traffic to specific IP addresses and allow access only from other services within the VPC.
A financial services company wants to automate its machine learning (ML) workflows to quickly respond to new training data arriving in its Amazon S3 bucket. The company needs to re-train its ML models whenever new data is added to the bucket. To automate this, they want to set up a continuous integration and continuous deployment (CI/CD) pipeline using AWS services. The solution should automatically trigger the ML model training pipeline and perform necessary preprocessing whenever new data is uploaded. Which combination of services and configurations should the company use to achieve this? (Select TWO.)	Use Amazon SageMaker Pipelines to define and manage the ML workflow, including data preprocessing, model training, and evaluation steps. Set up Amazon EventBridge to monitor the S3 bucket for object creation events and use it to trigger a SageMaker Pipeline execution.
A machine learning team is building a predictive model using historical sales data stored in Amazon S3. The data must be prepared by handling missing values and outliers before training. Which AWS tools can be used to perform these data preparation tasks efficiently? (Select TWO)	Amazon SageMaker Data Wrangler is a data preparation tool specifically designed for machine learning tasks. It allows users to clean, normalize, and preprocess data efficiently before training machine learning models. It is a suitable choice for handling missing values and outliers in historical sales data. AWS Glue DataBrew is a visual data preparation tool that can be used to clean and normalize data, handle missing values, and remove outliers efficiently. It provides a user-friendly interface for data preparation tasks without the need for complex coding.
A company focused on sustainable technology is training a deep learning model using Amazon SageMaker to develop a computer vision application. The training process involves a large dataset of high-resolution images and requires significant computational resources. The company's primary goal is to minimize energy consumption and carbon footprint while still achieving efficient training performance. Which instance type should the company select for their training job in Amazon SageMaker to maximize energy efficiency and meet their sustainability goals?	The ml.trn1.2xlarge instance type is designed for training deep learning models and offers a balance between computational power and energy efficiency. It provides sufficient resources for training a deep learning model with high-resolution images while minimizing energy consumption, making it the ideal choice for the company's sustainability goals.