A data science team is tasked with building a recommendation system for a media streaming platform. They need to ingest user activity logs stored in Amazon S3, process it for real-time updates, and use the data for training machine learning models in SageMaker. The data is stored in CSV format but needs to be optimized for faster querying and lower storage storage costs. What steps should they take to improve the ingestion and storage of the data (select two answers)	1) Convert the data to Parquet format to reduce storage costs and improve query performance, 2) Use AWS Glue to catalog and prepare the data before ingestion into SageMaker.
You are developing a machine learning solution that requires real-time transformation of streaming data from IoT devices. The data must be transformed, deduplicated, and outliers removed before being used for feature engineering. Which AWS services would best meet this requirement? Select three answers.	Kinesis Data Streams, AWS Lambda and Apache Flink
An online retail company needs to ensure that its customer data complies with privacy regulations such as CCPA and GDPR. They are using AWS services to prepare and train machine learning models. The data must be encrypted, and customer PII needs to be masked. What are the best services and techniques to meet these compliance needs during data preparation (select two answers)?	A) Correct - AWS KMS ensures the encryption of customer data at rest, which is essential for compliance with CCPA and GDPR B) Correct - Amazon Macie can detect and mask PII in the dataset, ensuring compliance with privacy regulations.
A media company needs to merge its video metadata stored in DynamoDB with user interaction data stored in Amazon S3. The data needs to be processed and analyzed using Amazon Redshift. Which AWS services and techniques will help in efficient data and merging (select two answers)?	AWS Glue and Redshift Spectrum
A healthcare company is building a data lake on AWS to store patient records in different formats. The data is first ingested as JSON, and they want to convert it into a format that optimizes both query performance and storage costs. Arrange the steps to complete the data pipeline (select five responses and place them in the correct order)	A - ingest raw JSON data into S3 B - Convert the JSON data to Parquet format using AWS Glue C - Compress the Parquet data to reduce storage costs D - Store the Parquet data in Amazon S3 E - Use AWS Glue Data Catalog to manage and query the Parquet files
A retail company is merging customer transaction data from Amazon S3 and DynamoDB to create a unified dataset for a machine learning model in SageMaker. The team needs to handle data duplication and resolve conflicting records while maintaining optimal performance. What is the most appropriate solution for merging these datasets?	A) Use AWS Glue to handle the merge with deduplication B) Use Amazon EMR with Spark for merging and resolving conflicts
An e-commerce company is experiencing ingestion errors when collecting large volumes of real-time clickstream data from their website via Amazon Kinesis. They notice some data is being dropped due to capacity limits and need to improve the robustness of their pipeline. What strategies should they implement to handle this scenario? (select two answers)	B) Use Amazon Kinesis Data Firehose for automatic scaling and error retries, D) Enable Kinesis Shard Splitting to increase stream capacity
A healthcare company is implementing an ML model for analyzing patient data. The data needs to be compliant with security and encryption standards and it requires long-term archival. The team must also ensure that the infrastructure is highly available and able to scale based on processing demands. The task is to match the AWS storage options and security services to their corresponding compliance and stability needs.	Amazon S3 - compliance with encryption for healthcare data Amazon S3 Glacier - Long Term archival storage for medical data AWS Key Management Services (KMS) - encryption and key management for regulatory compliance Amazon FSx - High-performance access to large medical datasets Amazon EBS - High-performance block storage for mission-critical workloads
Your ML pipeline needs to process data from multiple sources and transform it into a structured format for model training. The data transformation needs to be optimized for performance and scalability using AWS services. Match the services to their correct roles.	Amazon EMR with Spark - distributed processing of large-scale data AWS Glue - validating data quality post-transformation AWS Glue Data Quality - Running SQL queries on transformed data Amazon S3 - storing and retrieving transformed datasets Amazon Athena - running SQL queries on transformed data
You are developing an ML pipeline that uses Amazon SageMaker to manage the features of your model. The pipeline should be de-	A) Use SageMaker Pipelines to automate feature scaling, normalization, and integration with training jobs.

signed to handle feature scaling and normalization as part of the transformation process. It also needs to integrate seamlessly with SageMaker training jobs. What is the best strategy to implement this solution? (select three answers)	B) Use SageMaker Data Wrangler to preprocess and scale the features E) Use Amazon EMR to preprocess and normalize features in parallel
You are tasked with building an end to end machine learning pipeline for a healthcare application that requires data from multiple data sources. The datasets are stored in Amazon S3, but there are frequent issues with missing values and format inconsistencies that affect model training. Your goal is to automate data quality checks before training the model. Which AWS services will help ensure data completeness and correctness in your pipeline (select two).	A) Use AWS Glue Data Quality to validate data against quality rules to check for completeness. D) Use AWS Glue DataBrew for visual transformation and detection of missing values.
Your organization is working on a credit scoring model using customer data from various regions. There's a concern about potential measurement bias due to regional differences in data collection practices. Which AWS services can help you identify and mitigate this bias in the model-building process (select 2)	A) Use AWS Glue Data Quality to validate data against quality rules to check for completeness D) Use SageMaker Data Wrangler to normalize data from different regions.
A retail company is using Amazon SageMaker to train a demand forecasting model. The dataset includes significant seasonal variations and imbalances across product categories. Which actions should be taken to ensure that your model avoids bias due to these imbalances? (select two answers)	A) Use Amazon SageMaker Data Wrangler to preprocess data and apply stratified sampling, ensuring all categories are equally represented E) Use stratified data sampling in SageMaker to balance the training dataset based on seasonal variations.
A multinational organization is using Amazon SageMaker to train models with sensitive personal data. What combination of AWS services and configurations ensures data is secured and compliant with data residency laws? Select up to three correct answers	A) Use region-specific AWS KMS keys for encryption at rest B) Configure SageMaker VPC Endpoints to restrict access to specific regions. D) Implement SageMaker training jobs with encryption in transit using SSL/TLS
You are training a distributed machine learning model using Amazon SageMaker with multiple instances. The dataset is stored in Amazon S3 and is too large to fit into the memory of a single instance. What techniques can you use to ensure efficient data loading across multiple instances? Select three correct answers.	A) Use Amazon FSx for Lustre to cache the dataset from S3 and enable high-speed access for distributed training B) enable distributed sharding to split the dataset evenly across all instances for parallel data loading C) Configure SageMaker to use Pipe mode, which streams data directly from Amazon S3 to the training instances, eliminating the need for local disk storage.
You are training a large-scale image classification model on Amazon SageMaker using the ImageClassification algorithm. The dataset is stored in Amazon S3 and consists of millions of high-resolution images. You need to distribute the training across multiple instances for speed and cost optimization. What steps can you take to ensure effective distributed training and cost efficiency on Amazon SageMaker (select two correct answers)	A) Use SageMaker's Pipe Mode to stream training data directly from Amazon S3 to instances without copying to EBS, B) Configure SageMaker to use a Spot instance fleet to reduce training costs
After deploying a machine learning model in production using Amazon SageMaker, you notice a gradual decrease in performance over time. The model's prediction no longer aligns with real-world outcomes. How can you detect the cause of this issue and resolve it? (select two possible answers)	A- Use SageMaker Model Monitor to track data drift D- Use SageMaker Debugger to investigate model performance issues.
A financial services company is managing multiple fraud detection models on Amazon SageMaker. The company frequently updates models and requires real-time performance monitoring and the ability to roll back to previous versions. They want to use SageMaker to ensure minimal disruption during deployment and monitor model drift over time. What strategies should be implemented (Select 3 answers)?	A. Use SageMaker Model Registry for version tracking and rollback C. Use SageMaker Model Monitor to track model drift over time D. Use Blue/Green Deployment for seamless transitions between versions
Your company runs machine learning models for real-time prediction tasks in a distributed architecture. The models are updated frequently, and performance must be monitored for drift and degradation over time. The team also wants to ensure all models are version controlled for reproducibility and auditing purposes.	A. Use SageMaker Model Registry for model version and audit tracking B. Enable SageMaker Model Monitor to track model drift over time E. Implement SageMaker Clarify to continuously detect model bias over time

What tools and configurations you implement to meet these requirements (select three answers)	
A retail company is preparing to deploy a machine learning model for real-time inventory predictions using Amazon SageMaker. The company needs to configure compute resources, optimize costs, and enable monitoring of the infrastructure. Arrange the following five steps in order.	1 - A: Provision EC2 instances in SageMaker for model training 2- B: Configure Spot Instances for training cost savings 3 - C: Enable auto-scaling for SageMaker inference endpoints 4 - D: Deploy the model for real-time inference 5 - E: Monitor training and inference performance using CW
A machine learning model deployed in SageMaker is showing signs of drift in prediction quality over time. You need to monitor the model's performance for data drift and prediction anomalies while ensuring notifications are sent when significant changes occur. Which services will you use? (Select two answers)	A. Use SageMaker Model Monitor to detect data drift over time C. Configure Amazon SNS to send notifications for model drift alerts
Your team needs to ensure compliance with strict data security regulations for your ML models. They require encryption of both training data and model artifacts while ensuring minimal access. What would be your security configuration (select two answers)	A. Use AWS KMS (Key Management Service) to encrypt both the training data and the model artifacts B. Configure IAM Policies with the least privilege principle for all roles interacting with the model
A financial services industry deploys multiple SageMaker models for real-time risk assessment. The model endpoints must auto-scale to handle fluctuating traffic and ensure each request is processed quickly. The team needs to configure thresholds for scaling based on traffic while monitoring resource usage. Match the tasks with the corresponding AWS services used for auto-scaling and monitoring.	A - AWS Auto Scaling - defining scaling policies based on traffic B - CloudWatch - monitoring system metrics like CPU, memory and request rate C- Amazon EC2 - scaling instances based on demand D - SageMaker Inference Recommender - Using the best instance types for inference E - Amazon SNS - notifying teams about scaling events
A financial services company has deployed a SageMaker model to predict loan defaults. To ensure regulatory compliance and consistent performance, the team must monitor model metrics and receive real-time alerts to metrics such as accuracy or inference time fall below acceptable levels. Match the services to their appropriate tasks for setting up monitoring dashboards and alerts.	A - SageMaker Model Monitor - sending notifications when performance metrics are breached B - Amazon CloudWatch - Monitoring real-time model accuracy and latency C - Amazon QuickSight - creating visual dashboards for model metrics D - Amazon SNS - sending notifications when performance metrics are breached E - AWS Lambda - Triggering workflows for performance remediation
A financial institution is developing a model to detect fradulent transactions using Amazon SageMaker. To ensure explainability in regulatory audits, they need a model that balances performance and transparency. Which SageMaker algorithmwould best address this challenge?	C - Implement the SageMaker Linear Learner for interpretable linear classification
A retail company wants to enhance its customer experience by integrating automatic language translation for its support system. They plan to handle live chat in multiple languages. Which AWS AI services should they use to implement real-time translations and ensure cost-effective scaling for this solution? (Select two answers)	A - Use Amazon Translate for real-time language translation and Amazon Lex for conversational AI in multiple languages C - Use AWS Lambda for scaling translation requests and integrate with Amazon Translate
A healthcare company is developing a predictive model for patient outcomes. The model must be interpretable for medical professionals and comply with healthcare regulations. They are using Amazon SageMaker and must decide between interpretable algorithms and appropriate explainability tools. How should they proceed to meet these requirements? (Select two answers)	A - Use SageMaker Linear Learner to provide transparent coefficients and explainability D - Leverage SageMaker Clarify for bias detection and model explanation with any algorithm
A startup is building a recommendation system using Amazon SageMaker and needs to scale the training process efficiently for their growing dataset. The team is deciding between various training options and configurations to balance cost and performance. What options should they choose to ensure optimal resource utilization (select two answers):	A - Use Amazon EC2 Spot Instances with SageMaker to reduce training costs while scaling D - Set up distributed training with PyTorch in SageMaker to handle large-scale datasets
A data analytics company is building a machine learning model using Amazon SageMaker to predict customer churn. They want to reduce training time while keeping model performance intact.	

What should they focus on to optimize the process (Select two answers)	A - Use early stopping to reduce the overall training time D - Implement SageMaker Automatic Model Tuning to optimize hyperparameters
A startup is working on an image recognition model to classify products. They are facing issues with overfitting and want to optimize their model to improve accuracy. Using SageMaker, they aim to implement regularization and feature selection. How should they proceed? (Select five responses and place them in the correct order)	A - perform feature selection using SageMaker data wrangler D - use SageMaker Automatic Model Tuning for hyperparameter optimization E - adjust learning rate and batch size in the training configuration C - Apply L2 regularization to reduce model complexity B - Usage SageMaker Debugger to monitor model performance
Your team has trained a classification model using Amazon SageMaker and you are tasked with evaluating the model performance. You decide to calculate the F1 score, precision, and recall to assess its effectiveness. What would be the next best step to gain deeper insights into the model's performance (select two answers)	A - Use a confusion matrix to visualize true positive and false positive rates C - Visualize the ROC curve to assess performance across thresholds
A marketing team is using SageMaker to train a model that predicts customer churn. They suspect that the model is underfitting based on poor training and validation performance. What techniques should be used in order to improve the model's capacity to learn complex patterns (select two answers)	A - increase the number of layers in the neural network C - Use SageMaker hyperparameter tuning to optimize parameters
A healthcare provider is developing an ML model to predict patient outcomes based on previous medical records. They need to establish performance baselines using historical patient data and create a system to monitor the model's fairness. The company uses SageMaker Clarify and SageMaker Model Registry to track bias and store baselines. Match the AI services to the tasks needed for performance baseline management.	A - 3 SageMaker Clarify - Provides tools for evaluating bias in models and generating fairness metrics, ensuring that healthcare models treat all patient demographics fairly B - 1 SageMaker Model Registry - Tracks different model versions and the associated baselines for future comparsions, ensuring repeatability and auditability C - 2 Amazon CloudWatch - Monitors performance drift and accuracy in real-time, helping to identify whether the model deviates from its established baseline D - 4 AWS Glue - Transforms historical patient data, preparing it for training and establishing performance baselines using clean, structured data E - 5 Amazon QuickSight - Visualizes bias and performance metrics, providing healthcare teams with insights into how the model compares to its baseline over time
A telco company wants to ensure their customer churn production model does not introduce bias based on age or income. They using SageMaker Clarify and SageMaker Model Monitor to detect and mitigate bias in the training data and during inference. Match the steps involved in bias detection and mitigation with the correct AWS services.	SageMaker Clarify - Detecting bias in training and inference A - 1 SageMaker Model Monitor - Monitoring real-time bias during inference B - 2 Amazon S3 - Storing historical training data C -3

	Amazon QuickSight - Visualizing bias detection results D - 4 AWS Glue - Preparing data for bias detection E - 5
An automotive company is training a vehicle failure prediction model on SageMaker. The model's loss fluctuates wildly during training, and the team suspects the issue is related to hyperparameters. What steps should they take using SageMaker Debugger to resolve this? (select two answers)	A - analyze the learning rate using SageMaker Debugger to detect learning rate decay and adjust it with SageMaker AMT B - Enable SageMaker Debugger to monitor training gradients for instability
A financial institution has developed a fraud detection model using TensorFlow and deployed it on SageMaker. They are facing integration losses while deploying the TensorFlow model using a custom container. What steps can they take to integrate the model into SageMaker successfully select 2 answers?	1) Use the SageMaker Inference Toolkit to simplify model integration and custom container handling 2) Build the custom container using Amazon ECR and integrate it into SageMaker
A healthcare company has multiple machine learning models trained to predict patient outcomes. They want to combine the output of these models using stacking to improve overall prediction accuracy. What steps can they take to integrate the model into SageMaker successfully select two answers?	A) Use SageMaker Pipelines to automate the training of base models and the meta-learner D) - Train the meta-learner separately after all base models are trained in SageMaker
A financial services firm is developing a fraud detection model with significant class imbalance. They want to focus on reducing false positives and false negatives to ensure accurate detection without wrongly flagging legitimate transactions. Which metrics should be prioritized, and how can they be implemented in SageMaker (Select two answers)	A - Usage precision to focus on reducing false positives D - Create a custom evaluation metric that balances precision and recall based on business risk
A large financial institution is using Amazon SageMaker to train various machine learning models for predictive analytics. They want to monitor and reduce the TCO while ensuring high availability and scalability for critical models. Which approaches should they consider to optimize TCO? (Select two answers)	A - Use SageMaker cost allocation tags to track costs by project B - Apply SageMaker endpoint auto-scaling for critical models
A retail company is using SageMaker Clarify to audit their demand forecasting model for bias. They want to ensure predictions are not biased against certain product categories. Arrange the following steps in the correct order (Select 4 Responses)	B - Upload the demand forecasting model and data to SageMaker A - Set up sensitive attributes like product category and region D - Run SageMaker Clarify bias detection to analyze the model C - Analyze the SHAP values to understand feature impacts
Your team has developed a large machine learning model that needs to be deployed on edge devices with limited resources. To optimize the model, you want to reduce its size while while maintaining accuracy. How would you approach model compression without significantly sacrificing performance? (select two answers)	A - use model pruning to remove unnecessary neurons and layers B - convert model weights to lower precision using SageMaker Neo
Your team is deploying a new machine learning model and wants to run it in a shadow variant in Amazon SageMaker to compare its performance with the existing production model. Which approaches will ensure a proper comparison? (Select two answers)	A - Set up SageMaker Model Monitor to track data quality between the production and the shadow models E - Route a small percentage of production traffic to the shadow model using SageMaker endpoints
A financial services firm is conducting multiple ML deployments on fraud detection models wants to ensure each model is auditable and reproducible. Which AWS services would they use for tracking model versions, storing artifacts, and automating model retraining?	A) SageMaker Model Registry - stores and manages model versions, making them auditable and reproducible B) AWS Lambda - Triggers retraining workflows, when specific events occur C) Amazon S3 - Stores model artifacts and logs to ensure that artifacts are preserved D) AWS Glue - Prepares and cleans data for retraining models to ensure consistency E) SageMaker Pipelines - automates retraining processes and manages workflows for versioned models
A media company needs to fine-tune a language model using a dataset of movie scripts to create a recommendation engine for	

their platform. They plan to use transfer learning and a custom dataset. Which steps should they follow to fine-tune their model using AWS services, ensuring both the data and the model are handled effectively?	Amazon S3 - storing the movie scripts dataset SageMaker JumpStart - selecting a pre-trained language model Amazon Bedrock - customizing the model with movie script data SageMaker Pipelines - automating the fine-tuning process SageMaker Training - fine-tuning the language model
A team is using Amazon SageMaker to train a large NLP model. The model takes too long to converge, and training resources are being over-utilized. They want to optimize resource allocation and identify the bottlenecks. What approach should they take? (select two answers)	A - Enable SageMaker Debugger to analyze CPU and memory usage during training D - Use SageMaker Debugger's profiling feature to optimize the resources for faster convergence
Your team is building a real-time inference on Amazon SageMaker. The model is expected to handle low-latency, high throughput requests. You need to decide on a cost-effective deployment strategy while ensuring that the system can scale automatically during peak demand. What is the best approach to deploy the model (select two answers)	A - Deploy the model on SageMaker real-time endpoints with automatic scaling configured. C - Deploy the model on SageMaker hosting with multi-model endpoints to reduce infrastructure cost
You are tasked with deploying a batch inference pipeline to process large amounts of image data stored in Amazon S3. The pipeline needs to be cost-efficient and run daily. How would you provision compute resources for the deployment?	B - deploy SageMaker batch transform jobs using Spot instances for cost efficiency
A media company needs to deploy a recommendation system that serves personalized content based on real-time user data. The traffic fluctuates, and they need to balance cost and performance. How should the endpoints be configured (select two answers)?	B - Deploy SageMaker real-time endpoints on CPU instances with auto-scaling D - Use SageMaker Model Monitor to track endpoint performance and adjust scaling
A logistics company is building an ML model to predict delivery times. They are considering using SageMaker BYOC (Bring Your Own Container) to ensure consistency across development and production. What challenges should they be aware of when deploying custom containers?	B - Troubleshooting issues such as container image corruption or network configuration errors
A logistics company is deploying models to predict package delivery times on edge devices located in trucks. They need to ensure minimal latency while considering resource constraints. How can they optimize the model and deployment (select two answers)?	A - Use SageMaker Neo to optimize the model for deployment on resource-constrained devices C - Manage and update models on the edge using IoT Greengrass
You are using AWS CloudFormation to automate the deployment of a SageMaker model endpoint in a secure VPC. You notice the deployment fails due to a misconfiguration of network settings. What could be the cause, and how can it be resolved (select two answers)?	A - Ensure that SageMaker is configured with the correct security group in CloudFormation B - Check that VPC settings allow inbound traffic from the required SageMaker IP ranges
A machine learning workflow requires consistent, repeatable infrastructure deployments for model training and inference. The team is deciding between AWS CloudFormation and AWS CDK. What is a key difference between them that should be considered when building the infrastructure for the ML workflow?	A - AWS CF uses YAML/JSON templates, while AWS CDK uses programming languages like Python or TypeScript
A machine learning team is deploying a CI/CD pipeline to automate model deployment using AWS CodePipeline, CodeBuild and CodeDeploy. Arrange the following steps for setting up this pipeline to ensure automated model deployment and monitoring - select five responses and place them in the correct order	A - set up a CodePipeline source stage to retrieve model code B - Configure CodeBuild to train and package the model E - Test the deployed model for performance metrics C - Deploy the trained model using CodeDeploy to a SageMaker endpoint D - Monitor pipeline execution using Amazon CloudWatch logs
A healthcare organization is implementing a CI/CD pipeline for deploying their ML models and managing model versions using Git and AWS services. They want to ensure automated versioning, deployment, and rollback mechanisms. Arrange the steps in the correct order to implement this pipeline (Select five responses and place them in the correct order)	A - set up a Git repository for version control of the ML models B - Create a source stage in CodePipeline to pull the code from Git C - Use CodeBuild to train and package the model D - Deploy the trained model to SageMaker using CodeDeploy E - Use Git branches to automate rollback in case of deployment issues
You are configuring a CI/CD pipeline for a machine learning model deployment using CodePipeline and Git version control. The team	A - Use CodePipeline to trigger deployments from the master branch to feature branches that are merged

uses Gitflow to manage feature branches and needs to ensure that failed deployments automatically roll back to the last stable version. How would automate this process? Select three answers	B - Configure AWS CodeDeploy to automate rollback if the latest version fails D - Set up AWS CodeBuild to run tests and validate model performance before deployment
A data science team is implementing a CI/CD pipeline for their machine learning models using SageMaker, CodeBuild, and CodeDeploy. During deployment, they encounter intermittent build failures and slow pipeline execution. They need to troubleshoot and resolve the issues while ensuring that the pipeline auto-scales to meet growing demand. Match the AWS services to their corresponding troubleshooting actions.	A - 3: CloudWatch logs are used to track and analyze logs from the pipeline and deployment stages B - 1: CodeBuild is responsible for building the models, and errors can be debugged here C - 5: CloudWatch Alarms trigger actions when certain thresholds are reached, like resource utilization or error rates D - 2: Auto Scaling adjusts infrastructure to meet pipeline demand E - 4: Trusted Advisor helps optimize costs and improve performance based on resource usage
An e-commerce company is integrating containers from their ML pipeline to run distributed training jobs. They plan to use Amazon SageMaker to train models, Amazon ECR for container image storage, and Amazon ECS for deployment. The team needs to configure auto-scaling and optimize container build performance. Match the AWS services with their specific functions in this pipeline.	A - 2: SageMaker is used for training distributed machine learning models B - 5: Amazon ECS deploys containers for running ML jobs in production C - 3: Amazon ECR stores container images needed for deployment D - 1: Auto Scaling manages scaling for containerized resources based on demand E - 4: AWS CodePipeline builds and deploys containers automatically in a CI/CD setup
A team wants to optimize the training costs of a large machine learning model by leveraging SageMaker Spot Instances. They need to track the actual savings realized during training and ensure that the job runs efficiently. What solutions can they implement? (Select two answers)	B - Configure SageMaker to use Managed Spot Training with cost tracking features C - Use AWS Cost Explorer to analyze cost savings and compare Spot vs. On-Demand training expenses
Your company is hosting multiple machine learning models on SageMaker endpoints within a VPC. Users report performance issues due to network latency. What steps can you take to optimize performance for these endpoints?	A - Use enhanced networking on EC2 instances to improve VPC performance C - Utilize AWS Global Accelerator to optimize traffic routing to the VPC
A machine learning model needs to be deployed across multiple environments, including a real-time inference service in production and batch processing in development. The deployment should minimize operational overhead while optimizing cost and scaling for each environment. How would you configure the deployments (Select 3 answers)?	A - deploy the model in Amazon SageMaker endpoints for real-time inference with automatic scaling in the production environment B - Use Amazon ECS with Fargate for batch inference in development to automatically scale container workloads D - Use Amazon EKS in both environments, leveraging Kubernetes orchestration for unified deployment
A company has deployed a machine learning model on SageMaker with a fixed number of instances. They now want to auto-scale based on latency and CPU utilization to handle dynamic workloads efficiently. Which two metrics should be prioritized for auto-scaling in this scenario (select two answers)?	A - Latency as tracked by CW B - CPU Utilization as tracked by CW
In a CI/CD Pipeline for an ML model, you aim to ensure end-to-end testing from data preprocessing to deployment. The pipeline uses AWS Glue for preprocessing, SageMaker for model training, and Lambda for inference. What is the best approach to automate end-to-end tests (select two answers)	A - Use AWS Step Functions to orchestrate the testing pipeline from data preprocessing to inference B - Implement Integration tests in CodePipeline to validate each component
A financial services organization is automating retraining for their ML models when model accuracy declines. They use SageMaker Pipelines and AWS Lambda to detect performance drops and initiate the retraining process. Arrange the steps to implement the automated training pipeline (select five responses and place them in the correct order)	B - Use SageMaker Model Monitor to track model accuracy E - Trigger a Lambda function using Amazon EventBridge to initiate retraining A - Retrain the model using SageMaker Pipelines C - Test the retrained model using CodeBuild for performance metrics D - Deploy the retrained model using CodeDeploy if tests pass

A financial institution is building a scalable ML pipeline using SageMaker endpoints for real-time fraud detection. They want to use spot instances to minimize to minimize costs and maintain continuous monitoring. Arrange the steps to implement this pipeline (Select 5 responses)	E - Use SageMaker Inference Recommender to select optimal instance configurations B - Use spot instances during model training to optimize costs A - Deploy the model using SageMaker endpoints C - Enable Auto Scaling on SageMaker endpoints to handle changing workloads D - Configure Amazon CloudWatch for performance monitoring
A tech company needs to ensure their ML models in production do not drift over time and that model quality remains high. How can they continuously monitor and automate drift detection in their AWS SageMaker environments?	B - Use Amazon SageMaker Model Monitor to detect data quality drift and integrate with CloudWatch for alerts
Your organization uses SageMaker to deploy ML models in production and wants to monitor model performance. They plan to implement A/B testing for model variants and use SageMaker Model Monitor to observe any performance degradation. CW dashboards will track resource usage and latency. Match the AWS services with their roles in this workflow.	A - 5: Amazon SageMaker deploys ML models into production environments B - 1: SageMaker Model Monitor continuously checks real-time model performance C - 3: Amazon CW tracks resource usage and latency for performance monitoring D - 2: A/B testing compares the performance of different model versions E - 4: SageMaker Endpoint hosts models for inference requests
A fintech startup is deploying models for fraud detection and needs to monitor infrastructure performance metrics such as CPU utilization, network latency and throughput. They plan to use AWS CloudWatch for setting alarms and CloudWatch Logs for storing logs, AWS X-Ray will analyze latency across different services. Match the AWS components with the infrastructure monitoring tasks they support.	A - 1: CloudWatch Logs enables performance metrics to be stored for further analysis, making it essential for tracking system health B - 5: AWS X-Ray provides distributed tracing capabilities, identifying latency issues across multiple services, which are critical for performance optimization C - 2: SageMaker Endpoint is responsible for hosting the fraud detection model and serving inference requests in production
You need to ensure that your machine learning system logs all SageMaker API Actions and tracks across to critical datasets stored in Amazon S3. What services would you use to ensure you meet auditing and compliance needs?	A - CloudTrail captures all API actions for compliance and auditing C - Amazon Macie helps monitor and classify sensitive data stored in S3 D - Athena allows querying of CloudTrail logs for further analysis
A healthcare ML application using SageMaker requires high inference speed and low latency for critical predictions. You need to choose the best instance type while ensuring that costs do not escalate unnecessarily. How would you balance performance and cost (select two answers)	A - Use SageMaker inference recommender to evaluate different instance types B - Leverage inference-optimized instances to reduce latency
Your company runs large-scale ML training jobs using SageMaker and stores model data in Amazon S3. Costs are steadily increasing, and you need a detailed analysis of resource usage and opportunities to save costs. Which AWS services and setups will help you reduce expenses and optimize resource usage (select three answers)?	A - set up AWS Cost Explorer to track and analyze cost spikes related to SageMaker training jobs B - Set up AWS Trusted Advisor to recommend S3 storage optimizations C - Create a Cost and Usage Report to get detailed insights into S3 and SageMaker usage
Your team is running an inference pipeline on Amazon SageMaker and experiencing high latency during peak traffic. You need to monitor latency, identify bottlenecks, and automate alerts when latency exceeds a threshold. What AWS tools and configurations would you implement (Select three answers)?	A) CloudWatch Lambda Insights provides granular metrics, making it effective for identifying latency during inference. B) AWS X-Ray helps visualize the execution flow, which is key to identifying bottlenecks in latency C) CloudWatch Alarms can trigger notifications when latency exceeds defined thresholds, helping automate detection
Your organization is experiencing latency issues in SageMaker model inference during peak traffic. You are tasked with reducing latency while maintaining cost efficiency. Which steps should you take using AWS services (select two answers)?	A - Use AWS X-Ray to identify latency bottlenecks in your SageMaker inference requests D - Enable Auto Scaling on SageMaker endpoints to adjust instance counts during high traffic
Your team is preparing to deploy a machine learning model to production, and you must ensure that only authorized users can access the deployed instances and datasets. How can you manage and audit access effectively (select three answers)?	A - Configure SageMaker endpoints with IAM roles that allow access to only authorized users B - Use bucket policies on Amazon S3 to restrict dataset access to specific SageMaker roles

	C - Enable AWS CloudTrail to track access to both SageMaker endpoints and data sources
During an audit of your CI/CD pipeline for deploying ML models on SageMaker, you discover that access to the pipeline has not been properly restricted. What security practices should you implement to resolve this issue?SELECT THREE ANSWERS	A- Use IAM policies with least privilege C- Use AWS CloudTrail to track access and modifications to pipeline resources E- Use AWS CodePipeline's integration with AWS KMS for securing artifacts in transit and at rest
An online education platform needs to monitor their ML models for both performance and budget. They want to set up a monitoring dashboard using CloudWatch and QuickSight with real-time alerts. Arrange the steps to implement this solution	A - Configure CW for SageMaker model performance metrics C - Create a CloudWatch dashboard for monitoring performance and cost B - Set up CloudWatch alarms for threshold breaches D - Use Amazon QuickSight for cost visualization E - Automate alerts in Amazon QuickSight for real-time updates
Your company is experiencing intermittent scaling issues with its SageMaker endpoints during peak traffic hours. The endpoint is unable to handle the incoming inference requests, leading to high latency and occasional failures. You have been tasked with setting up a monitoring and scaling mechanism to ensuring smooth scaling of the endpoint as traffic fluctuates. How would you troubleshoot a solution that dynamically handles traffic peaks without compromising cost efficiency? (Select three answers)	A - Use CW Alarms to monitor CPU and memory utilization on SageMaker instances. B - Configure Auto Scaling on the SageMaker endpoint to adjust the number of instances based on traffic E- Review CloudWatch logs to troubleshoot scaling failures and identify bottlenecks
Your organization is using multiple SageMaker endpoints for inference, but the growing costs have raised concerns. You have been asked to address the usage of these endpoints and implement cost-saving measures without compromising performance. How would you integrate AWS Trusted Advisor and related services to automate this cost optimization process? (Select two answers)	A - Use AWS Trusted Advisor to review endpoint configuration and receive recommendations on instance rightsizing D - Use SageMaker inference Recommender to analyze cost vs. performance for different instance types
A healthcare company plans to train multiple deep learning models using Amazon SageMaker Spot Instances. To ensure minimal interruption and maximize savings, they want to track Spot instance performance, monitor cost, and auto-scale inference resources during peak data analysis periods. Match the services with their respective roles.	A - SageMaker Spot Instances - Use Spot Instances for cost-effective training B - AWS Auto-scaling - scale inference resources automatically C - CloudWatch Alarms - Trigger alarms when costs exceed thresholds D - AWS Budgets - Monitor real-time cost savings E - SageMaker Endpoint - Deploy models for real-time predictions
A gaming company is deploying a SageMaker models for real-time player behavior predictions. They want to secure the endpoints within a VPC, encrypt the data transmitted between the clients and the endpoints, and monitor network activity. Match the AWS services with their respective security roles.	A - 2: AWS CloudTrail logs all API calls made to and from AWS resources, ensuring traceability B - 3: Security Groups restrict and allow network traffic, acting as a firewall for SageMaker endpoints C - 4: VPC Flow Logs monitor network traffic in and out of the VPC, allowing for network-level visibility D - 5: Amazon VPC provides network isolation, preventing, direct internet access to the SageMaker models E - 1: AWS Certificate Manager (ACM) handles SSL/TLS certificates, ensuring secure communication
An organization is using Amazon SageMaker needs to ensure continuous compliance monitoring for all machine learning workloads, especially data access and model version changes. They want automated alerts and audits for any unauthorized changes or security policy violations. What combination of services will best meet this need select 2 answers	A - Use AWS CloudTrail to track all changes to SageMaker model versions and data access B - Implement AWS Config to monitor IAM policy compliance and generate alerts on policy changes
Your organization is implementing a policy to ensure least privilege access for all SageMaker users. You need to configure IAM roles so that users only have the required access to perform their tasks. How should you proceed to implement least privilege access while auditing compliance (select two answers)?	A - Use AWS IAM to create fine-grained access policies for each user based on their job function B - Implement AWS Config to monitor IAM policy compliance and generate alerts on policy changes
Your organization requires all outbound traffic from ML systems hosted in Amazon SageMaker to be strictly controlled and monitored. The systems must also be isolated in a private network. What steps should you take to achieve this? (select two answers)	A) Set up a NAT gateway for controlled outbound traffic in a private subnet B) Enable VPC Flow Logs to monitor network traffic

You are tasked with securing a machine learning deployment that automates model training and inference using Amazon SageMaker. The pipeline runs within an AWS VPC, and you need to ensure all communications and access to resources are encrypted. What configuration should you use?	A - Use AWS KMS to encrypt all sensitive data stored in the pipeline C - Enable SSL/TLS encryption for all VPC endpoint communications
Your company recently experienced an unexpected cost surge from a SageMaker training job. You need to identify the root cause and ensure that failure cost overruns are prevented. What strategies should you implement? (Select two answers)	A - set up cost allocation tags and use AWS Cost Explorer to investigate the specific services and resources driving costs E - Set up AWS Budgets to alert your team before cost limits are exceeded
A machine learning team is deploying a real-time inference model using Amazon SageMaker. They want to ensure scalability, cost efficiency and monitoring of their infrastructure. Arrange the steps for proper deployment and monitoring in the correct order - five responses	C - Apply IAM policies for secure access to the SageMaker endpoint A - Deploy the model to a SageMaker endpoint and configure auto-scaling B - Set up CW metrics to monitor endpoint performance and cost D - Use AWS Budgets to monitor the cost of autoscaling E - Use AWS Cost Explorer to analyze the cost trends for endpoint usage
A retail company wants to build a demand forecasting model using SageMaker. They need to ingest sales and inventory data from Amazon DynamoDB and store it efficiently for machine learning. Which services and techniques should they use to ingest and store the data for this model? (Select two answers)	A - Use SageMaker Data Wrangler to perform feature scaling and one-hot encoding on the categorical variables C - Store the merged data in Parquet format on Amazon S3 to reduce storage costs
You are tasked with building a machine learning pipeline for a large ecommerce company. The project requires you to perform feature engineering on a dataset containing both numerical and categorical data. You need to handle feature scaling and encode the categorical variables before training the model. Which services and techniques are most appropriate for this task? Select two correct answers.	A - Use SageMaker Data Wrangler to perform feature scaling and one-hot encoding on the categorical variables D - Use SageMaker Feature Store to store the engineered features for later reuse during training
A healthcare company is preparing a dataset for training a machine learning model to predict patient readmission rates. The dataset contains sensitive medical information, and the company must ensure HIPAA compliance, mitigate potential bias in the data, and anonymize personal details. What is the most appropriate approach to address those requirements (select three answers)	A) Use AWS Glue to anonymize patient information B) Implement SsageMaker Clarify to detect bias in the dataset D) Use AWS Key Management Service (KMS) to encrypt the data
You are troubleshooting a machine learning pipeline where data extraction from Amazon S3 into Amazon SageMaker is running slower than expected. The dataset is very large, and network bandwidth is becoming a bottleneck. How can you improve the data extraction performance (select two answers).	A) Enable Amazon S3 Transfer Acceleration to reduce latency for cross-region data transfers C) Use Amazon SageMaker Pipe Mode to stream data directly from S3
A retail company is ingesting streaming data from multiple sources like Amazon Kinesis and S3 into SageMaker Data Wrangler for feature engineering. You need to implement the data preparation steps including normalization and feature selection. What is the correct order of steps to prepare the data for SageMaker training (select four answers)	A - ingest streaming data from Amazon Kinesis into SageMaker Data Wrangler D - Visualize and validate the data with built-in tools in Data Wrangler B - Apply normalization and scaling to the data using SageMaker Data Wrangler C - Use SageMaker Feature Store to store the features for reuse
An e-commerce platform needs to integrate product data from multiple sources, including Amazon RDS and Amazon S3 to create a comprehensive dataset for machine learning analysis. The team needs to address performance issues and avoid data duplication (select two answers)?	A) Use AWS Glue for merging datasets with deduplication C) Implement Amazon EMR with Apache Spark for high-performance data merging
A machine learning team is using Amazon S3 and AWS Glue for data ingestion. Recently, ingestion failures have been occuring	A) Use Amazon S3 access logs to identify the bottlenecks in request rates

due to hitting S3 reqeust rate limits, causing workflow disruptions. What steps should they take to troubleshoot and scale their ingestion pipeline (select two answers).	C) Use AWS CloudWatch to monitor and auto-scale the ingestion pipeline based on S3 request rates
A logistics company deploys an ML model to optimize delivery routes. They want to establish a baseline model using historical data and set up automated monitoring of the model's performance over time. The team uses SageMaker Model Monitor to detect any drift and CloudWatch for real-time metric tracking. Match the AWS services to their task in creating and maintaining performance baselines.	A - Amazon S3 Glacier - ideal for archive of historical purchasing data at low cost B - Amazon S3 Standard - Commonly used for frequently accessed data such as current purchasing data C - Amazon S3 Intelligent-Tiering - Automatically balances cost and performance by adjusting storage tiers based on access frequency D - Amazon FSx - provides high performance data access for analysis of large datasets E - Amazon EFS - Used for shared access to files and model data during real-time operations
Your team needs to build a machine learning model that requires labelling a large dataset for training. To streamline this, you decide to use SageMaker Ground Truth and automate the workflow for labeling a variety of data types such as text and images. The team also needs to ensure high-quality annotations and avoid data bias. Which tools should be integrated to achieve this goal (Select two)?	B) Use SageMaker Clarify to identify potential biases in the data labelling process C) Deploy an Amazon Mechanical Turk workforce to automate the process
A large e-commerce company is streamlining user clickstream data to Amazon S3 using Amazon Kinesis. There are frequent gaps and irregularities in the data due to network interruptions, and you need to ensure the streaming data is validated for completeness and format consistency before it's stored in S3. Which AWS services can you integrate into the data pipeline to automate these checks (select two)	A) Use AWS Glue Data Quality to run quality checks on the streaming data for completeness B) Implement Amazon Kinesis Data Analytics for real-time validation of streaming data
You are training a model using a large dataset from various user groups, and there is concern about imbalanced representation of these groups. To ensure fair predictions, you must preprocess the data to address this imbalance. Which methods and AWS services will you use (select TWO)	A) Use SageMaker Clarify to detect class imbalance in the dataset B) Use SageMaker Data Wrangler to perform data augmentation and upsample underrpresented groups
You are tasked with deploying a sentiment analysis model using Amazon SageMaker for a global customer base. You suspect that there may be regional biases in the dataset, leading to biased model outputs. Which strategies should you adopt to detect and mitigate bias during model training and deployment? Select three correct answers.	A) Use SageMaker Clarify to detect regional biases in the training dataset before model deployment. B) Implement SageMaker Model Monitor to continuously monitor the model for bias in regional sentiment prediction post deployment D) Use stratified sampling during training to ensure an equal representation of all regions
A media company is using Amazon SageMaker to train a recommendation system with user data stored in Amazon S3. The dataset is large, and loading times are causing significant delays in model training. What should be done to minimize data loading delays and optimize performance? (Select two correct answers)	A) Use Amazon FSx for Lustre to cache the dataset locally C) Implement SageMaker Pipe Mode to stream data directly from Amazon S3 to the training container, reducing memory usage
You are building a machine learning pipeline on AWS to classify image data using Amazon SageMaker. The dataset consists of large images, and you need to choose the appropriate ML algorithm. You must ensure that your model is scalable and handle overfitting. What should you do to optimize model training and avoid overfitting while ensuring stability? Select three correct answers	B) Use a convolutional neural network (CNN) and implement L2 regularization to reduce overfitting D) Implement early stopping in SageMaker to prevent overfitting E) Use an ensemble of models in SageMaker to improve model accuracy and scalability
Your team is training a deep learning model on SageMaker, but they have run into frequent training failures. Logs show OutOfMemory errors during the training process. What measures can be taken to troubleshoot and fix the issue?	A) Switch to a larger instance type with more memory as ml.p3.16xlarge B) Use SageMaker Model Debugger to analyze memory during the training job

You have trained a model to predict customer churn using Amazon SageMaker. During post-training analysis, you want to understand which features are most important to the model's predictions. What is the best way to achieve this?	A) Use SHAP (SHapley Additive exPlanations) to explain the feature importance
You are tasked with deploying a healthcare image analysis model on SageMaker. The model needs frequent updates and it's crucial to monitor its predictions in real-time and roll back if the new versions performs poorly. Furthermore, you need to track model versions for audit purposes. Which AWS services should you use to meet these requirements? (select two answers)	A) Use SageMaker Model Registry for version tracking and auditability D) Implement Blue/Green Deployment for seamless model version transitions
You are tasked with deploying a high-traffic recommendation model on Amazon SageMaker. The model must auto-scale based on demand, support real-time traffic and allow for real-time traffic, and allow for rapid rollback to previous versions in case of failure. It should also be monitored for prediction drift and performance issues over time. Which strategies and services should you use?	Select three answers - A) Deploy the model with SageMaker Endpoints and enable Blue/Green Deployment for rollback capability B) Implement SageMaker Model Monitor to track prediction drift and performance issues C) Use SageMaker Automatic Scaling to adjust resources based on real-time traffic demands
A financial services firm is setting up a CI/CD pipeline to automate the deployment of machine learning models on Amazon SageMaker. They need to automate dataset versioning, model deployment, and continuous monitoring. Arrange the following steps in the correct order to complete the pipeline setup (select five responses and place them in the correct order)	E - Version datasets using Amazon S3 and AWS Lambda triggers A - Use AWS CodePipeline to orchestrate the CI/CD process B - Set up SageMaker Model Registry to manage model versions C - Automate model deployment using CodeDeploy D - Integrate CloudWatch for continuous monitoring of the pipeline
Your ML team has deployed a model in production, but they are concerned about prediction anomalies and need a solution that tracks these anomalies and alerts the the team when they occur. Additionally, the team wants to ensure that model inputs do not experience unexpected changes over time. What will you configure (select two answers)?	A - Configure SageMaker Model Monitor to track prediction anomalies D - Set up Amazon SNS for alert notifications when Model Monitor detects anomalies
You need to secure access to a SageMaker endpoint deployed within a VPC to ensure that only specific roles and users can interact with the model. What would be the most appropriate security configurations to apply? (select two answers)	A - Use IAM policies to restrict SageMaker endpoint access to specific roles and users B - Configure Amazon VPC security groups to control inbound and outbound traffic
A retail company runs an ML model for real-time product recommendations, deployed on SageMaker. As the company grows, it needs to ensure consistent performance under various loads. The team must configure SageMaker endpoints to scale automatically based on traffic. Match the services to the appropriate tasks for configuring and monitoring auto-scaling for SageMaker endpoints.	Amazon CloudWatch - Monitoring endpoint latency and traffic AWS Auto Scaling - Automatically adjusting endpoint instance capacity AWS Lambda - Triggering scaling actions in response to alarms SageMaker Endpoint - Setting scaling policies based on performance metrics AWS Trusted Advisor - Providing cost and performance optimization recommendations
A media streaming company has deployed a recommendation system on SageMaker. The company wants to monitor model performance and user engagement metrics and set up a dashboard that can track these metrics. They also need to configure automatic alerts for any degradation in model performance. Match the AWS services to the task required to set up this monitoring system.	Amazon CloudWatch - Monitoring real-time engagement and model performance metrics SageMaker Model Monitor - detecting drift or degradation in model performance Amazon QuickSight - Visualizing engagement trends and performance metrics AWS Lambda - automating remediation actions for performance degradation Amazon SNS - automating alerts for performance degradation
A financial services company is developing a predictive model using Amazon SageMaker to detect fradulent transactions. The data is highly imbalanced, and the model must provide interpretable results to comply with regulatory requirements. The team also needs to automate model retraining when new data arrives via Amazon S3. Which approach should they use (please select two answers)	A - Use SageMaker XGBoost for handling enhanced tabular data and provide interpretability with SHAP values B - Implement SageMaker Linear Learner for interpretable coefficients and integrate with SageMaker Pipelines for automated retraining
A global media company wants to automate the process of converting spoken interviews into text for transcription and transla-	

tion into multiple languages. They need an efficient solution that integrates AI services and scales with demand, providing high accuracy for both transcription and translation. What should they use to achieve this? (Select two answers)	A - Use Amazon Transcribe for automatic speech-to-text conversion and Amazon Translate for multilingual translations C - Integrate AWS Lambda for automatically scaling transcription jobs with Amazon Transcribe
A retail company is developing a product recommendation system using Amazon SageMaker. The team needs to balance model interpretability with performance to provide insights to non-technical stakeholders. They are evaluating different machine learning models and considering tools for interpretability. What solution best fits their needs? (Select two answers)	B - Implement SageMaker Linear Learner for intepretable recommendations and use SageMaker Clarify for bias detection C - Train SageMaker XGBoost for high performance and explain using SHAP values
A healthcare startup is training a deep learning model using Amazon SageMaker. Due to the complexity of the model, they need to balance training time and performance without excessive costs. Which steps should they follow to optimize the training process? (Select two answers)	A - Use Spot instances to reduce costs while scaling training D - utilize early stopping in SageMaker to halt training when improvements are minimal
A team is developing an XGBoost model using Amazon SageMaker to predict customer churn. They want to automate the hyperparameter tuning process and ensure optimal performance. Arrange the following steps in the correct order to achieve this (five responses and place them in the correct order)	B, D, A, C, E A - SageMaker Automatic Model Tuning automates the process of optimizing hyperparameters, reducing manual efforts in the training process B - Specifying a hyperparameter search range for parameters like max_depth and eta is critical to exploring the space effectively C - Model performance evaluation based on metrics like AUC help to track the success of the model training process D - Choosing between random search or Bayesian optimization in SageMaker ensures that you use the right strategy for hyperparameter tuning E - Analyzing results from multiple trials allows you to identify and select the best-performing model configuration
A financial services company is using Amazon SageMaker to develop a credit risk prediction model. After training the the model, the team needs to assess whether the model is overfitting or underfitting. Which techniques would help them evaluate the model's performance (select two answers)	A) Compare training and validation loss curves over time E) Compare the F1 Score to assess balance between precision and recall
A retail company has developed a model to predict product demand using SageMaker. However, the model performs well on the training set but poorly on the validation set. How can the team reduce overfitting and improve the model's generalization to unseen data (select two answers)?	A) Add dropout layers during training to prevent overfitting C) Use early stopping to terminate training when validation loss increases
A logistics company deploys an ML model to optimize delivery routes. They want to establish a baseline model using historical data and set up automated monitoring of the model's performance over time. The team uses SageMaker Model Monitor to detect any drift and CW for real-time metric tracking. Match the AWS services to their tasks in creating and maintaining performance baselines.	A) Amazon S3 - Stores both the historical data and baseline metrics, ensuring that the data is readily available for future comparisons and model retraining B) SageMaker Model Monitor - Detects drift or any other changes in model performance, comparing the current metrics to the established baseline C) Amazon CW - Monitors real-time performance metrics, ensuring that deviations from the baseline can be detected quickly D) SageMaker Model Registry - Stores model versions and baseline metrics, enabling tracking and comparisons between different models and their baseline performances E) Amazon QuickSight - Visualizes both performance baselines and any drift detected by SageMaker Model Monitor, providing clear insights into model performance trends
A financial services company is using SageMaker to develop a fraud detection model. They need to ensure that the model is free from bias and does not unfairly target specific groups of customers based on demographics. They plan to use SageMaker Clarify for	A) 1 - B) 5, C) 4, D) 5, E) 3 A) SageMaker Clarify - detects bias in model predictions, ensuring that fraud detection models do not unfairly target specific demographic groups B) Amazon CloudWatch - Monitors real-time performance metrics, including bias detection reports, ensuring that models continue to perform fairly C) Amazon S3 - Stores data used for bias detection, making it

detecting and mitigating bias in both the training and inference phases. Match the tasks to the relevant AWS services that will support bias detection and mitigation efforts.	available for future analysis D) AWS Glue DataBrew - Prepares and transforms raw data, ensuring that it is clean and ready for bias detection processes E) SageMaker Model Monitor - Tracks bias in real-time data during inference, flagging any performance deviations that suggest unfair treatment of certain groups
A retail company uses Amazon SageMaker to deploy multiple machine learning models for sales forecasting. They need to manage model versions to ensure reproducibility, rollback, and performance comparison across different models. How can the company manage model versions effectively using AWS services (Select two answers)	A) Use SageMaker Model Registry to store and version models for audit and tracking E) Roll back to a previous mode version using SageMaker Model Registry's rollback feature
A retail company has built a recommendation system using PyTorch and wants to deploy it in SageMaker. They are considering different approaches to ensure smooth deployment and scaling of the model. What are the best practices for integrating their PyTorch model with SageMaker select two answers	A) Use SageMaker's PyTorch estimator to easily integrate the model into SageMaker C) Store the model artifacts in Amazon S3 for SageMaker to access during deployment
A retail company wants to boost the performance of their recommendation model by applying boosting algorithms. They are unsure whether to use SageMaker's built-in algorithms or to develop a custom model. What are the key considerations they should keep in mind when implementing boosting in SageMaker (2 answers)	A) Implement boosting using SageMaker's built-in XGBoost algorithm for scalability and ease C) Use SageMaker's Hyperparameter Tuning for optimizing boosting parameters
A healthcare company is building a classification model for disease diagnosis. Given the high-risk nature of incorrect classifications, they need to balance the business impact of false positives and false negatives carefully. Which metrics and approaches should they consider for model evaluation in SageMaker (Select two answers)?	A) Use F1 score to balance precision and recall in classification E) Create a custom evaluation metric to give different weights to false positives and false negatives
A gaming company is deploying multiple machine learning models in real time for player recommendations. They want to use a cost-effective solution that can scale with user traffic while maintaining low latency. What are the best deployment strategies they should implement using SageMaker (two answers)?	A) Use SageMaker multi-model endpoints to host multiple models on the same instance C) Enable auto-scaling on SageMaker endpoints to handle varying traffic effectively
You are setting up a hyperparameter tuning job in Amazon SageMaker to optimize a machine learning model. Arrange the following steps in the correct order to set up this hyperparameter tuning job (select five responses and place them in the correct order)	C,A,B,D,E C) Create a SageMaker training job to use in tuning A) Define the hyperparameter ranges for tuning B) Specify the objective metric to optimize (e.g. accuracy) D) Run the tuning job to explore different hyperparameters E) Review the best model based on the objective metric
You are tasked with deploying a complex neural network to an environment where compute resources are constrained. The goal is to minimize the model's size by applying efficient techniques without heavily sacrificing accuracy. What methods will you employ?	A) Convert the model to INT8 precision with SageMaker to reduce weight size
In a production environment, you are testing a shadow variant of a machine learning model alongside the current production model. What steps would you take to ensure that both models' performance is accurately compared without affecting the user experience (select three answers)	B) Monitor both the production and the shadow models using Amazon CloudWatch dashboards C) Configure SageMaker Model Monitor to track bias detection across both models D) Set up an A/B testing framework to split traffic between the live and shadow models
A retail company is developing an ML model for sales forecasting and needs to maintain repeatable training processes across multiple model versions. They alsowant to automate deployment using version control. Which AWS services would you match to each step of this version management process?	A) - 1, B) - 2, C) -3, D)-4, E)-5 A) SageMaker Model Registry - Manages different model versions, ensuring repeatability across versions B) Amazon S3 - Stores datasets used for model training and retraining C) AWS CodePipeline - Automates the deployment of model versions across different environments D) AWS Glue - Prepares and transforms data to ensure consis-

	tency between versions E) SageMaker Pipelines - Automates version management workflows, ensuring model retraining processes are consistent
A logistics company is using a pre-trained model to optimize delivery routes. They want to fine-tune the model with their proprietary delivery data to improve prediction accuracy. What AWS services should they use to handle data preparation, fine-tuning and model deployment?	A) - 1, B) -4, C) -5, D)-3, E)-2 A) Amazon S3 - Storing proprietary delivery data B) SageMaker Training - Fine-tuning the pre-trained model on delivery data C) SageMaker Endpoints - Deploying the fine-tuned model for prediction use D) AWS Glue - Preparing the proprietary data for model training E) SageMaker JumpStart - Selecting the pre-trained model for optimization
You are managing an ML project where reproducibility is a key requirement. Your team uses Amazon SageMaker to train models and wants to ensure that every experiment can be replicated with the same results. They want to track version metadata, manage version control for datasets and code, and repeat experiments efficiently. How would you configure SageMaker to achieve this? (Select two answers)	A) Use SageMaker Experiments to track and organize trials and metadata B) Store your training datasets in Amazon S3 with versioning enabled to ensure reproducibility
A financial institution wants to deploy a machine learning model that makes predictions on transactions in real-time to detect fraud. The solution must be low-latency, scalable and optimized for cost. They are considering using Amazon SageMaker and other AWS services. What would be the most appropriate deployment strategy?	C) Use SageMaker real-time endpoints with multi-model hosting and auto-scaling
Your team is deploying a model for image classification using Amazon SageMaker. The model should be cost-efficient but must handle both batch and real-time inference. How would you provision resources for both types of workloads? (select two answers)	A) Use SageMaker real-time endpoints with multi-model hosting for both batch and real-time inference D) Deploy SageMaker real-time endpoints and schedule batch jobs using Spot Instances
An e-commerce company is deploying a fraud detection model in SageMaker. They want to ensure low latency predictions with the ability to monitor and troubleshoot endpoint issues like high latency. How would you set up and maintain the endpoints?	D) Use SageMaker Model Monitor to track latency and resource utilization
An AI company wants to deploy a large-scale multi-container ML system using SageMaker, where each container will serve a different task in the inference pipeline. How can they achieve this efficiently while ensuring minimal downtime (Select two answers)?	A) Use SageMaker endpoints with multi-container support, assigning different tasks to each container D) Leverage ECR for container image management and deploy the images using SageMaker endpoints
A media streaming service wants to implement an auto-scaling solution for their SageMaker endpoints, which serve a real-time recommendation engine. They need to optimize for cost-efficiency while ensuring low-latency response during high-demand periods. How can they achieve this? (Select two answers)	A) Configure auto-scaling based on CPU utilization metrics E) Implement scaling based on invocations per instance
A company wants to automate the entire machine learning pipeline using AWS CDK. They need consistent infrastructure deployments, including SageMaker, S3 and Lambda. What should be prioritized to ensure a seamless IaC experience?	A) Use AWS CDK to define infrastructure in Python, ensuring version control
Your team is automating the provisioning of machine learning infrastructure using AWS CDK. What steps should be taken to integrate the deployment with continuous delivery pipelines (select two answers)	A) Define your infrastructure using AWS CDK and deploy it using AWS CodePipeline D) Integrate AWS CodeBuild to run tests and validate the infrastructure prior to deployment
A healthcare organization is using AWS CodePipeline to automate ML model deployments, but they need to integrate automated testing and model monitoring. Arrange the the steps to incorporate testing and monitoring into the CI/CD pipeline (Select five responses and place them in the correct order)	B - Use CodeBuild to train the model and run unit tests A - Set up a test stage in CodePipeline to validate the ML model C - Deploy the model using CodeDeploy to a SageMaker endpoint D - Monitor deployed model performance using SageMaker Model Monitor E - Track pipeline errors and model metrics using CW Logs
A media company is setting up a CI/CD pipeline for their ML workflows with version control, automated deployments, and rollback	A - Create a Git repository to store and version ML models B - Use CodeBuild to train the model and run unit tests

mechanisms using AWS services and Git. Arrange the steps to ensure the pipeline is fully automated and can handle rollbacks (Select five responses and place them in the correct order)	C - Deploy the model using CodeDeploy to a SageMaker endpoint D - Monitor deployed model performance using SageMaker model monitor E - Track pipeline errors and model metrics using CloudWatch logs
You are tasked with building an automated model deployment pipeline using SageMaker Pipelines and AWS CodePipeline. The pipeline should handle model retraining based on new data and deploy it automatically if it passes validation tests. How do you set this up to ensure retraining and automated deployment? (select two answers)	A) Use SageMaker Pipelines to orchestrate model retraining and deployment tasks C) Configure EventBridge to automatically trigger model retraining when model performance drops
A machine learning team is experiencing latency issues during the deployment of their models using a CI/CD pipeline. The models are hosted on SageMaker, and the deployment is managed via CodeDeploy. They want to implement monitoring and retry mechanisms to minimize deployment failures and reduce downtime. Match the AWS services with their correct roles in monitoring and retrying the deployment.	A) Amazon SageMaker - Host ML models B) AWS Lambda - Retry failed deployment step C) AWS CodeDeploy - deploy models across environments D) Amazon CW Logs - Monitor logs to detect latency issues E) AWS Step Functions - automate pipeline orchestration
A healthcare company needs to deploy custom ML models in containers using Amazon SageMaker for inference. They plan to store their custom containers in ECR and deploy them using ECS. The team also wants to monitor deployment performance and automatically trigger retries in case of deployment failures. Match the services with their appropriate role in the containerized ML workflow.	A) Amazon SageMaker - Handle containerized ML workloads B) Amazon ECR - Store containerized ML models C) Amazon ECS - Deploy containers for inference D) AWS CloudWatch - Monitor deployment metrics E) AWS Step Functions - Automate retries for failed tasks
You are running an Amazon SageMaker training job using Spot Instances and notice that your job experiences delays due to Spot instance interruptions. What steps can you take to balance performance and cost optimization in this scenario (Select three answers)?	A) Use Spot Instance pricing history to adjust job start times for higher availability B) Implement SageMaker Checkpointing to minimize rework after interruptions D) Use a combination of Spot and On-Demand Instances
You are responsible for deploying a SageMaker model endpoint within a VPC. Security policies require that the endpoint must not have direct internet access but should still be able to pull container images from ECR. How do you configure the VPC to meet this requirement select two answers.	B) Attach a NAT gateway to the private subnet to enable access to AWS services such as ECR D) Configure an interface VPC endpoint for ECR in the VPC
You are deploying a machine learning model using the SageMaker SDK, and you need to ensure that the model is deployed with the appropriate IAM permissions, inside a VPC, and auto-scales based on traffic. What are the correct configurations in the SageMaker SDK to meet this requirements (Select 2 answers)?	A) Use the VpcConfig parameter in the SageMaker SDK to specify private subnets and security groups B) Configure an IAM role with the necessary permissions and pass it through the ExecutionRoleArn parameter in the SageMaker SDK
During a high-demand period, your SageMaker endpoint starts to experience delays in processing requests. You want to auto-scale the endpoint using CW metrics. Which specific metric combination would best address this issue (Select two answers)	a) increase the number of instances when invocations per instance cross a threshold b) set scaling policies based on latency exceeding a specified threshold
You are tasked with ensuring continuous testing of an ML model deployment in a CI/CD pipeline using CodePipeline, CodeBuild and SageMaker. How can you automate performance monitoring of the model in production while ensuring test coverage (two answers)	A) Use SageMaker Model Monitor to track the performance of the deployed model C) Leverage Amazon CloudWatch Logs to capture model performance metrics for testing
A media company wants to automate retraining of their recommendation engine when model drift is detected. They plan to use SageMaker Model Monitor and Pipelines for automation and deploy models only after evaluating their performance. Arrange the following steps to implement this pipeline (FIVE RESPONSES)	B) Use SageMaker Model Monitor to detect drift in the recommendation engine E) Trigger SageMaker Pipelines for retraining upon drift detection A) Run tests on the retrained model using CodeBuild D) Use SageMaker Pipelines to retrain and evaluate the model C) Deploy the model using CodeDeploy after successful evaluation
A technology startup is designing a scalable ML solution for customer churn prediction. They plan to use SageMaker endpoints	D) use saagemaker inference recommender to choose optimal instance types B) enable spot instances for cost-effective model training E) set up auto scaling for sagemaker endpoints

with autoscaling and cost-efficient training using spot instances. Arrange the steps for this implementation in the correct order.	C) deploy the trained model to sagemaker endpoints A) set up CW for endpoint monitoring and logging
A fintech company uses Amazon SageMaker for training and deploying ML models. To ensure data quality for incoming datasets, they want to set up automated data validation checks before training starts. Which services can they integrate to automate data quality checks and detect anomalies (select two answers)	B) Set up AWS Glue Data Brew for cleaning and validation of datasets D) Integrate AWS Glue Data Quality to assess data completeness and accuracy
A fintech company deploys ML models in production using SageMaker and needs to analyze prediction performance over time. They also want to detect anomalies in model behavior, using real-time monitoring, and set up alarms for latency or memory usage spikes. Match the AWS services with the tasks they perform in the monitoring pipeline	A - 2: SageMaker Model Monitor detects performance degradation and tracks model drift B - 5: CloudWatch monitors real-time resource usage and provides metrics C - 4: SageMaker Endpoint serves production models for inference D - 1: SageMaker Clarify analyzes models for potential bias E - 3: CloudWatch Alarms trigger alerts for latency or memory issues
You are managing an ML model pipeline where model performance degradation is being detected. You want to log these degradations and trigger automated model retraining when performance falls below a threshold. What combination of services would achieve this? (Select two answers)	C) Use Amazon EventBridge to trigger automated retraining based on performance logs E) Set up SageMaker Model Monitor to track model performance and trigger retraining
Your machine learning model hosted on SageMaker is processing a large batch of requests, leading to latency issues. You want to optimize your infrastructure by selecting the appropriate instance types for inference while maintaining a cost-effective solution. What strategies would you use? (Select three answers)	B) Use memory-optimized instances to handle large batch sizes effectively C) Use SageMaker Multi-Model endpoints to deploy models on fewer instances D) Implement SageMaker Inference Recommender to evaluate the best instance type for batch processing
Your company is running multiple machine learning pipelines using Amazon SageMaker. To track and optimize costs, you want to ensure that all ML resources, including EC2 instances and S3 buckets, are tagged appropriately for cost tracking. How would you configure AWS tools to automate cost allocation select 2 answers	A) Use AWS Organizations to enforce tagging policies across all ML resources B) Configure AWS Billing and Cost Management to track costs by resource tag
Your machine learning inference endpoint is facing inconsistent latency, which impacts real-time predictions. You need to identify resource allocation bottlenecks and optimize the infrastructure for lower latency. How can AWS services help you? (Select two answers)	A) Use AWS Compute Optimizer to recommend instance optimizations B) Configure Amazon CloudWatch to monitor inference latency and throughput
You are tasked with rightsizing your SageMaker inference infrastructure to reduce costs while maintaining performance. Which AWS tools would you use to balance cost and performance (select 3 answers)	A) Use SageMaker Inference Recommender to optimize instance types based on cost and latency requirements B) Implement SageMaker Savings Plans for cost savings on long-running inference workloads C) Use AWS Compute Optimizer to get insights into the performance of current instance types
You are managing a multi-team SageMaker environment and want to enforce least privilege policies for each team. How would you ensure that each team's access is restricted only to the resources they need? (Select two answers)	A) Use IAM policies scoped to specific resources such as datasets, models, and SageMaker endpoints C) Enable SageMaker Role Manager to generate fine-tuned IAM policies for each team's resources
You need to monitor and audit security events in your CI/CD pipeline for deploying models on AWS. Which tools and techniques should you use to ensure the pipeline is secure and complies with best practices (select three answers)	A) Enable AWS CloudTrail to monitor API calls and detect unauthorized access B) Use IAM roles with strict least privilege policies for all pipeline components D) Configure AWS CodePipeline with VPC endpoints to restrict network traffic
Your team has set up an automated machine learning pipeline for a retail company to monitor the accuracy of demand forecasting models using Amazon SageMaker. You are tasked with setting up an event-driven system to trigger retraining when data drifts. The solution needs to integrate Amazon EventBridge for event handling and Amazon SageMaker Model Monitor to model drifts.	B) Set up Amazon SageMaker Model Monitor to monitor data quality and model performance D) Store training data in Amazon S3 and register the model in SageMaker Model Registry A) Create an EventBridge rule to detect data drift events from SageMaker Model Monitor

Arrange the steps in the correct order to complete the task of setting up this automated retraining process.	E) Configure an EventBridge rule to trigger an alert when drift is detected
A machine learning team at your organization deployed a SageMaker endpoint, but it experiences occasional spikes during high-traffic periods. The team suspects the resource allocation for the endpoint is insufficient to handle the workload during these peaks. To resolve this, they want to implement an Auto Scaling mechanism. How can Auto Scaling help optimize the resource allocation of SageMaker Endpoints based on traffic patterns and resource utilization (select 2 answers)	A) Configure Auto Scaling to add more EC2 instances behind the SageMaker endpoint during high traffic E) Configure Auto Scaling based on invocation metrics like invocations per instance and latency
After deploying several models on SageMaker endpoints, you notice significant cost overruns AWS Trusted Advisor has flagged several cost-related inefficiencies, and your team has been tasked with reducing costs while maintaining endpoint performance. What strategies and tools would you implement to optimize the cost efficiency of your SageMaker deployments? (Select three answers)	A) Use AWS Cost Explorer to analyze historical usage data and identify trends that lead to high costs B) Automate cost management using AWS Budgets to send alerts wen spending exceeds allocated amounts C) Implement AWS Trusted Advisor recommendations to rightsizing EC2 instances used by SageMaker endpoints
An AI startup is using SageMaker Spot Instances to train models for image recognition. They need to monitor infrastructure performance, track cost efficiency, and ensure scalability for inference workloads during peak demand Spot instance interruptions are a concern, so they want to ensure the right configurations are in place. Match the services with the roles they play.	A) SageMaker Spot Instances - Use Spot instances for cost savings B) AWS Auto Scaling - Scale resources dynamically for inference C) CloudWatch Alarms - set alarms for resource or cost anomalies D) SageMaker Model Monitor - Monitor model accuracy post-deployment E) AWS Budgets - Track budget and cost savings in real-time
A global e-commerce company is deploying a SageMaker model for product recommendations. To maintain data privacy and comply with global data protection laws, they need to restrict access to their model endpoints, encrypt all data in transit, and log network activity for audit purposes	A) Security Groups - control access to the endpoints B) Amazon VPC - secure the endpoint with a private network C) AWS Certificate Manager - Manage encryption for data in transit D) AWS CloudTrail - Capture API calls for audit E) VPC Flow Logs - log network traffic
You are managing a SageMaker-based ML pipeline for a financial organization that requires strict security controls. You notice that unauthorized API calls were made to your SageMaker endpoints. You want to identify the source of these calls and ensure all future actions are securely audited. How should you troubleshoot this issue please select two answers	A) Use AWS CloudTrail to review API call logs for all SageMaker actions B) Implement IAM role policies to restrict access to SageMaker endpoints
You are tasked with securing access to an Amazon S3 bucket that stores ML training model for your SageMaker models. To comply with least privilege access policies, only specific users should have readonly access to the data. What is the best approach to achieve this w/ 2 answers?	A) Use IAM roles with read-only access policies to allow specific users access to the S3 bucket B) Enable AWS Config to ensure the S3 bucket has restrictive access policies in place
You need to secure network access for an ML training environment in Amazon SageMaker hosted in a VPC. The environment must be isolated, but you still need access to Amazon S3 for model storage and training data. What network configuration should you use?	A) Use private subnets and a NAT gateway to enable secure access to Amazon S3
Your machine learning team is implementing a CI/CD pipeline using AWS CodePipeline to automate model deployments. They need to ensure that access to the pipeline is restricted, and all stages are monitored for unauthorized changes. What steps should they follow? Select three	A) Apply least privilege IAM roles for CodePipeline stages to limit access B) Use AWS CloudTrail to monitor and log all changes made to the pipeline E) Configure CW Alarms to monitor for changes in pipeline execution
A team is using AWS Compute Optimizer to rightsize their EC2 instances used in machine learning training jobs. They aso want to ensure that right metrics are in place for future cost tracking. What additional steps can they take to manage costs efficiently (Select three answers)	A) Set up AWS Budgets with cost alerts tied to EC2 instance usage patterns C) Enable Amazon CloudWatch to track EC2 instance-level usage and set thresholds for usage D) Apply cost allocation tags to all EC2 instances and use AWS Cost Explorer for analysis

Your team is building a machine learning (ML) CI/CD pipeline using AWS CodePipeline for automated deployments. You want to ensure security and cost-efficiency by applying best practices in access control and monitoring throughout the pipeline setup. List the steps to configure the secure CI/CD pipeline in the correct order (select four responses and place them in the correct order)	A) Create IAM roles with least privilege access for CodePipeline, SageMaker, and other services D) Set up VPC configurations to restrict pipeline access to private subnets B) Encrypt pipeline data at rest using AWS KMS C) Monitor pipeline security using Amazon CloudWatch logs
A logistics company is using Amazon Kinesis to ingest real-time sensor data from its fleet of delivery trucks. The data needs to be stored for later use in machine learning models that predict maintenance requirements. The company wants to ensure that the storage solution is both scalable and cost-effective, while also allowing the data to be queried efficiently. What are the best storage and ingestion options for this setup select 2 answers.	A) Use Amazon S3 to store the ingested data in Parquet format for scalability and cost savings D) Use Amazon Kinesis Firehose to deliver the data into Amazon S3 for long-term storage
You are working on a machine learning project that requires the use of various AWS services to process and transform large datasets before training. The data is initially stored in Amazon S3, and you need to apply feature scaling and detect any outliers. What is the best combination of services to prepare the data for machine learning (select 3 correct answers)	A) Use AWS Glue DataBrew to detect outliers and apply transformations such as scaling and normalization B) Use Amazon EMR with Apache Spark to process and transform large datasets, applying feature engineering techniques such as binning and log transformation C) Use SageMaker Data Wrangler to connect to S3 and apply feature scaling while detecting outliers
A multinational corporation is building a model for customer sentiment analysis. They need to anonymize customer feedback data, which contains PII, and ensure compliance with global data residency requirements. They also need to measure and mitigate bias in the dataset before training. What are the best tools and techniques to address these concerns (Select three answers)	A) Use SageMaker Clarify to measure and mitigate bias in the customer feedback data C) Use Amazon Macie to detect and mask sensitive information like email addresses E) Ensure that customer data is stored in AWS Regions that comply with local data residency laws
You are working on a machine learning pipeline using Amazon SageMaker to process large datasets stored in Amazon S3. The data needs to be transformed into the most efficient format for storage, query, performance, and cost-effectiveness for downstream ML tasks. Arrange the steps for transforming and selecting the appropriate data format in the correct order (four responses)	A) Use AWS Glue to transform data from CSV to Parquet for better query performance B) Compress data using AWS Glue to reduce storage costs on Amazon S3 D) Use AWS Glue Data Catalog to catalog and manage the Parquet-formatted data C) Load data into Amazon Redshift using Amazon S3 as the staging area
A healthcare firm is using SageMaker Data Wrangler to process patient data stored in Amazon S3. The data need to undergo cleaning, binning and feature selection before model training. Arrange the steps in the correct order.	A - ingest the raw patient data from Amazon S3 into SageMaker Data Wrangler B - Apply binding to categorical data for better model performance C - Normalize and scale the numerical data for consistency D - Select important features using builtin Data Wrangler tools E - Export the transformed data to Amazon S3 for SageMaker model training
A logistics company is combining real-time sensor data from Amazon Kinesis and historical shipment data from Amazon RDS for predictive modelling. They need to handle performance considerations while merging these large datasets and ensuring no data duplication. Which AWS services and techniques should they use three answers	A - Use AWS Glue for merging historical and real-time data B) Implement Amazon EMR with Spark for performance optimization during merging C - Use Amazon Kinesis Data Analytics to handle real-time data streams
A media company is ingesting video data from multiple sources into Amazon S3, but they are encountering frequent timeouts and ingestion errors due to inconsistent data upload speeds. They need a solution that can handle varying upload rates and automatically retry failed uploads. What AWS services and techniques should they uses? (Select two answers)	C) Use Amazon S3 Multipart Upload to handle large video files D) Enable S3 Transfer Acceleration to improve upload speeds