

Amazon-Web-Services

Exam Questions AIP-C01

AWS Certified Generative AI Developer - Professional



NEW QUESTION 1

A company is developing a generative AI (GenAI) application that uses Amazon Bedrock foundation models. The application has several custom tool integrations. The application has experienced unexpected token consumption surges despite consistent user traffic.

The company needs a solution that uses Amazon Bedrock model invocation logging to monitor InputTokenCount and OutputTokenCount metrics. The solution must detect unusual patterns in tool usage and identify which specific tool integrations cause abnormal token consumption. The solution must also automatically adjust thresholds as traffic patterns change.

Which solution will meet these requirements?

- A. Use Amazon CloudWatch Logs to capture model invocation log
- B. Create CloudWatch dashboards for token metric
- C. Configure static CloudWatch alarms with fixed thresholds for each tool integration.
- D. Store model invocation logs in Amazon S3. Use AWS Glue and Amazon Athena to analyze token usage trends.
- E. Use Amazon CloudWatch Logs to capture model invocation log
- F. Create CloudWatch metric filters to extract tool-specific invocation pattern
- G. Apply CloudWatch anomaly detection alarms that automatically adjust baselines for each tool's token metrics.
- H. Store model invocation logs in an Amazon S3 bucket
- I. Use AWS Lambda to process logs in real time
- J. Manually update CloudWatch alarm thresholds based on trends identified by the Lambda function.

Answer: C

NEW QUESTION 2

A company deploys multiple Amazon Bedrock-based generative AI (GenAI) applications across multiple business units for customer service, content generation, and document analysis. Some applications show unpredictable token consumption patterns. The company requires a comprehensive observability solution that provides real-time visibility into token usage patterns across multiple models. The observability solution must support custom dashboards for multiple stakeholder groups and provide alerting capabilities for token consumption across all the foundation models that the company's applications use.

Which combination of solutions will meet these requirements with the LEAST operational overhead? (Select TWO.)

- A. Use Amazon CloudWatch metrics as data sources to create custom Amazon QuickSight dashboards that show token usage trends and usage patterns across FMs.
- B. Use CloudWatch Logs Insights to analyze Amazon Bedrock invocation logs for token consumption patterns and usage attribution by application
- C. Create custom queries to identify high-usage scenarios
- D. Add log widgets to dashboards to enable continuous monitoring.
- E. Create custom Amazon CloudWatch dashboards that combine native Amazon Bedrock token and invocation CloudWatch metrics
- F. Set up CloudWatch alarms to monitor token usage thresholds.
- G. Create dashboards that show token usage trends and patterns across the company's FMs by using an Amazon Bedrock zero-ETL integration with Amazon Managed Grafana.
- H. Implement Amazon EventBridge rules to capture Amazon Bedrock model invocation events
- I. Route token usage data to Amazon OpenSearch Serverless by using Amazon Data Firehose
- J. Use OpenSearch dashboards to analyze usage patterns.

Answer: CD

NEW QUESTION 3

A financial services company is deploying a generative AI (GenAI) application that uses Amazon Bedrock to assist customer service representatives to provide personalized investment advice to customers. The company must implement a comprehensive governance solution that follows responsible AI practices and meets regulatory requirements.

The solution must detect and prevent hallucinations in recommendations. The solution must have safety controls for customer interactions. The solution must also monitor model behavior drift in real time and maintain audit trails of all prompt-response pairs for regulatory review. The company must deploy the solution within 60 days. The solution must integrate with the company's existing compliance dashboard and respond to customers within 200 ms.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Configure Amazon Bedrock guardrails to apply custom content filters and toxicity detection
- B. Use Amazon Bedrock Model Evaluation to detect hallucination
- C. Store prompt-response pairs in Amazon DynamoDB to capture audit trails and set a TTL
- D. Integrate Amazon CloudWatch custom metrics with the existing compliance dashboard.
- E. Deploy Amazon Bedrock and use AWS PrivateLink to access the application securely
- F. Use AWS Lambda functions to implement custom prompt validation
- G. Store prompt-response pairs in an Amazon S3 bucket and configure S3 Lifecycle policies
- H. Create custom Amazon CloudWatch dashboards to monitor model performance metrics.
- I. Use Amazon Bedrock Agents and Amazon Bedrock Knowledge Bases to ground responses
- J. Use Amazon Bedrock Guardrails to enforce content safety
- K. Use Amazon OpenSearch Service to store and index prompt-response pairs
- L. Integrate OpenSearch Service with Amazon QuickSight to create compliance reports and to detect model behavior drift.
- M. Use Amazon SageMaker Model Monitor to detect model behavior drift
- N. Use AWS WAF to filter content
- O. Store customer interactions in an encrypted Amazon RDS database
- P. Use Amazon API Gateway to create custom HTTP APIs to integrate with the compliance dashboard.

Answer: A

NEW QUESTION 4

A company uses an AI assistant application to summarize the company's website content and provide information to customers. The company plans to use Amazon Bedrock to give the application access to a foundation model (FM).

The company needs to deploy the AI assistant application to a development environment and a production environment. The solution must integrate the environments with the FM. The company wants to test the effectiveness of various FMs in each environment. The solution must provide product owners with the ability to easily switch between FMs for testing purposes in each environment.

Which solution will meet these requirements?

- A. Create one AWS CDK applicatio
- B. Create multiple pipelines in AWS CodePipelin
- C. Configure each pipeline to have its own settings for each F
- D. Configure the application to invoke the Amazon Bedrock FMs by using the `aws_bedrock.ProvisionedModel.fromProvisionedModelArn()` method.
- E. Create a separate AWS CDK application for each environmen
- F. Configure the applications to invoke the Amazon Bedrock FMs by using the `aws_bedrock.FoundationModel.fromFoundationModelId()` metho
- G. Create a separate pipeline in AWS CodePipeline for each environment.
- H. Create one AWS CDK applicatio
- I. Configure the application to invoke the Amazon Bedrock FMs by using the `aws_bedrock.FoundationModel.fromFoundationModelId()` metho
- J. Create a pipeline in AWS CodePipeline that has a deployment stage for each environment that uses AWS CodeBuild deploy actions.
- K. Create one AWS CDK application for the production environmen
- L. Configure the application to invoke the Amazon Bedrock FMs by using the `aws_bedrock.ProvisionedModel.fromProvisionedModelArn()` metho
- M. Create a pipeline in AWS CodePipelin
- N. Configure the pipeline to deploy to the production environment by using an AWS CodeBuild deploy actio
- O. For the development environment, manually recreate the resources by referring to the production application code.

Answer: C

NEW QUESTION 5

A company has deployed an AI assistant as a React application that uses AWS Amplify, an AWS AppSync GraphQL API, and Amazon Bedrock Knowledge Bases. The application uses the GraphQL API to call the Amazon Bedrock RetrieveAndGenerate API for knowledge base interactions. The company configures an AWS Lambda resolver to use the RequestResponse invocation type.

Application users report frequent timeouts and slow response times. Users report these problems more frequently for complex questions that require longer processing.

The company needs a solution to fix these performance issues and enhance the user experience.

Which solution will meet these requirements?

- A. Use AWS Amplify AI Kit to implement streaming responses from the GraphQL API and to optimize client-side rendering.
- B. Increase the timeout value of the Lambda resolve
- C. Implement retry logic with exponential backoff.
- D. Update the application to send an API request to an Amazon SQS queu
- E. Update the AWS AppSync resolver to poll and process the queue.
- F. Change the RetrieveAndGenerate API to the InvokeModelWithResponseStream AP
- G. Update the application to use an Amazon API Gateway WebSocket API to support the streaming response.

Answer: A

NEW QUESTION 6

A publishing company is developing a chat assistant that uses a containerized large language model (LLM) that runs on Amazon SageMaker AI. The architecture consists of an Amazon API Gateway REST API that routes user requests to an AWS Lambda function. The Lambda function invokes a SageMaker AI real-time endpoint that hosts the LLM.

Users report uneven response times. Analytics show that a high number of chats are abandoned after 2 seconds of waiting for the first token. The company wants a solution to ensure that p95 latency is under 800 ms for interactive requests to the chat assistant.

Which combination of solutions will meet this requirement? (Select TWO.)

- A. Enable model preload upon container startu
- B. Implement dynamic batching to process multiple user requests together in a single inference pass.
- C. Select a larger GPU instance type for the SageMaker AI endpoint
- D. Set the minimum number of instances to 0. Continue to perform per-request processin
- E. Lazily load model weights on the first request.
- F. Switch to a multi-model endpoint
- G. Use lazy loading without request batching.
- H. Set the minimum number of instances to greater than 0. Enable response streaming.
- I. Switch to Amazon SageMaker Asynchronous Inference for all request
- J. Store requests in an Amazon S3 bucke
- K. Set the minimum number of instances to 0.

Answer: AD

NEW QUESTION 7

A medical company is creating a generative AI (GenAI) system by using Amazon Bedrock. The system processes data from various sources and must maintain end-to-end data lineage. The system must also use real-time personally identifiable information (PII) filtering and audit trails to automatically report compliance.

Which solution will meet these requirements?

- A. Use AWS Glue Data Catalog to register all data sources and track lineag
- B. Use Amazon Bedrock Guardrails PII filter
- C. Enable AWS CloudTrail logging for all Amazon Bedrock API calls with Amazon S3 integratio
- D. Use Amazon Macie to scan stored data for sensitive information and publish findings to Amazon CloudWatch Log
- E. Create CloudWatch dashboards to visualize the findings and generate automated compliance reports.
- F. Use AWS Config to track data source configurations and change
- G. Use AWS WAF with custom rules to filter PII at the application layer before Amazon Bedrock processes the dat
- H. Configure Amazon EventBridge to capture and route audit events to Amazon S3. Use Amazon Comprehend Medical with scheduled AWS Lambda functions to analyze stored outputs for compliance violations.
- I. Use AWS DataSync to replicate data sources to track lineag
- J. Configure Amazon Macie to scan Amazon Bedrock outputs for sensitive informatio
- K. Use AWS Systems Manager Session Manager to log user interaction
- L. Deploy Amazon Textract with AWS Step Functions workflows to identify and redact PII from generated reports.
- M. Configure Amazon Athena to query data sources to analyze and report on data lineag
- N. Use Amazon CloudWatch custom metrics to monitor PII exposure in Amazon Bedrock responses and establish AWS X-Ray tracing to generate an audit trai
- O. Use an Amazon Rekognition Custom Labels model to detect sensitive information in the data that Amazon Bedrock processes.

Answer: A

NEW QUESTION 8

A healthcare company uses Amazon Bedrock to deploy an application that generates summaries of clinical documents. The application experiences inconsistent response quality with occasional factual hallucinations. Monthly costs exceed the company's projections by 40%. A GenAI developer must implement a near real-time monitoring solution to detect hallucinations, identify abnormal token consumption, and provide early warnings of cost anomalies. The solution must require minimal custom development work and maintenance overhead.

Which solution will meet these requirements?

- A. Configure Amazon CloudWatch alarms to monitor InputTokenCount and OutputTokenCount metrics to detect anomalies
- B. Store model invocation logs in an Amazon S3 bucket
- C. Use AWS Glue and Amazon Athena to identify potential hallucinations.
- D. Run Amazon Bedrock evaluation jobs that use LLM-based judgments to detect hallucination
- E. Configure Amazon CloudWatch to track token usage
- F. Create an AWS Lambda function to process CloudWatch metrics
- G. Configure the Lambda function to send usage pattern notifications.
- H. Configure Amazon Bedrock to store model invocation logs in an Amazon S3 bucket
- I. Enable text output logging
- J. Configure Amazon Bedrock guardrails to run contextual grounding checks to detect hallucination
- K. Create Amazon CloudWatch anomaly detection alarms for token usage metrics.
- L. Use AWS CloudTrail to log all Amazon Bedrock API calls
- M. Create a custom dashboard in Amazon QuickSight to visualize token usage patterns
- N. Use Amazon SageMaker Model Monitor to detect quality drift in generated summaries.

Answer: C

NEW QUESTION 9

A company configures a landing zone in AWS Control Tower. The company handles sensitive data that must remain within the European Union. The company must use only the eu-central-1 Region. The company uses Service Control Policies (SCPs) to enforce data residency policies. GenAI developers at the company are assigned IAM roles that have full permissions for Amazon Bedrock.

The company must ensure that GenAI developers can use the Amazon Nova Pro model through Amazon Bedrock only by using cross-Region inference (CRI) and only in eu-central-1. The company enables model access for the GenAI developer IAM roles in Amazon Bedrock. However, when a GenAI developer attempts to invoke the model through the Amazon Bedrock Chat/Text playground, the GenAI developer receives the following error:

```
User arn:aws:sts:123456789012:assumed-role/AssumedDevRole/DevUserName Action: bedrock:InvokeModelWithResponseStream
```

```
On resource(s): arn:aws:bedrock:eu-west-3::foundation-model/amazon.nova-pro-v1:0 Context: a service control policy explicitly denies the action
```

The company needs a solution to resolve the error. The solution must retain the company's existing governance controls and must provide precise access control.

The solution must comply with the company's existing data residency policies.

Which combination of solutions will meet these requirements? (Select TWO.)

- A. Add an AdministratorAccess policy to the GenAI developer IAM role
- B. Extend the existing SCPs to enable CRI for the eu.amazon.nova-pro-v1:0 inference profile
- C. Enable Amazon Bedrock model access for Amazon Nova Pro in the eu-west-3 Region
- D. Validate that the GenAI developer IAM roles have permissions to invoke Amazon Nova Pro through the eu.amazon.nova-pro-v1:0 inference profile on all European Union AWS Regions that can serve the model
- E. Extend the existing SCP to enable CRI for the eu-* inference profile

Answer: BE

NEW QUESTION 10

A company has a recommendation system. The system's applications run on Amazon EC2 instances. The applications make API calls to Amazon Bedrock foundation models (FMs) to analyze customer behavior and generate personalized product recommendations.

The system is experiencing intermittent issues. Some recommendations do not match customer preferences. The company needs an observability solution to monitor operational metrics and detect patterns of operational performance degradation compared to established baselines. The solution must also generate alerts with correlation data within 10 minutes when FM behavior deviates from expected patterns.

Which solution will meet these requirements?

- A. Configure Amazon CloudWatch Container Insights for the application infrastructure
- B. Set up CloudWatch alarms for latency threshold
- C. Add custom metrics for token counts by using the CloudWatch embedded metric format
- D. Create CloudWatch dashboards to visualize the data.
- E. Implement AWS X-Ray to trace requests through the application component
- F. Enable CloudWatch Logs Insights for error pattern detection
- G. Set up AWS CloudTrail to monitor all API calls to Amazon Bedrock
- H. Create custom dashboards in Amazon QuickSight.
- I. Enable Amazon CloudWatch Application Insights for the application resource
- J. Create custom metrics for recommendation quality, token usage, and response latency by using the CloudWatch embedded metric format with dimensions for request types and user segment
- K. Configure CloudWatch anomaly detection on the model metric
- L. Establish log pattern analysis by using CloudWatch Logs Insights.
- M. Use Amazon OpenSearch Service with the Observability plugin
- N. Ingest model metrics and logs by using Amazon Kinesis
- O. Create custom Piped Processing Language (PPL) queries to analyze model behavior patterns
- P. Establish operational dashboards to visualize anomalies in real time.

Answer: C

NEW QUESTION 10

A university recently digitized a collection of archival documents, academic journals, and manuscripts. The university stores the digital files in an AWS Lake Formation data lake.

The university hires a GenAI developer to build a solution to allow users to search the digital files by using text queries. The solution must return journal abstracts that are semantically similar to a user's query. Users must be able to search the digitized collection based on text and metadata that is associated with the journal abstracts. The metadata of the digitized files does not contain keywords. The solution must match similar abstracts to one another based on the similarity of their text. The data lake contains fewer than 1 million files.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Use Amazon Titan Embeddings in Amazon Bedrock to create vector representations of the digitized file
- B. Store embeddings in the OpenSearch Neural plugin for Amazon OpenSearch Service.
- C. Use Amazon Comprehend to extract topics from the digitized file
- D. Store the topics and file metadata in an Amazon Aurora PostgreSQL database
- E. Query the abstract metadata against the data in the Aurora database.
- F. Use Amazon SageMaker AI to deploy a sentence-transformer model
- G. Use the model to create vector representations of the digitized file
- H. Store embeddings in an Amazon Aurora PostgreSQL database that has the pgvector extension.
- I. Use Amazon Titan Embeddings in Amazon Bedrock to create vector representations of the digitized file
- J. Store embeddings in an Amazon Aurora PostgreSQL Serverless database that has the pgvector extension.

Answer: D

NEW QUESTION 14

A company is building a generative AI (GenAI) application that uses Amazon Bedrock APIs to process complex customer inquiries. During peak usage periods, the application experiences intermittent API timeouts that cause issues such as broken response chunks and delayed data delivery. The application struggles to ensure that prompts remain within token limits when handling complex customer inquiries of varying lengths. Users have reported truncated inputs and incomplete responses. The company has also observed foundation model (FM) invocation failures.

The company needs a retry strategy that automatically handles transient service errors and prevents overwhelming Amazon Bedrock during peak usage periods. The strategy must also adapt to changing service availability and support response streaming and token-aware request handling.

Which solution will meet these requirements?

- A. Implement a standard retry strategy that uses a 1-second fixed delay between attempts and a 3-retry maximum for all error
- B. Handle streaming response timeouts by restarting stream
- C. Cap token usage for each session.
- D. Implement an adaptive retry strategy that uses exponential backoff with jitter and a circuit breaker pattern that temporarily disables retries when error rates exceed a predefined threshold
- E. Implement a streaming response handler that monitors for chunk delivery timeout
- F. Configure the handler to buffer successfully received chunks and intelligently resume streaming from the last received chunk when connections are re-established.
- G. Use the AWS SDK to configure a retry strategy in standard mode
- H. Wrap Amazon Bedrock API calls in try-catch blocks that handle timeout exception
- I. Return cached completions for failed streaming request
- J. Enforce a global token limit for all user
- K. Add jitter-based retry logic and lightweight token trimming for each request
- L. Resume broken streams by requesting only missing chunks from the point of failure
- M. Maintain a small in-memory buffer of the most recent chunks.
- N. Set Amazon Bedrock client request timeouts to 30 seconds
- O. Implement client-side load shedding
- P. Buffer partial results and stop new requests when application performance degrades
- Q. Set static token usage caps for all requests
- R. Configure exponential backoff retries, dynamic chunk sizing, and context-aware token limits.

Answer: B

NEW QUESTION 16

A company runs a generative AI (GenAI)-powered summarization application in an application AWS account that uses Amazon Bedrock. The application architecture includes an Amazon API Gateway REST API that forwards requests to AWS Lambda functions that are attached to private VPC subnets. The application summarizes sensitive customer records that the company stores in a governed data lake in a centralized data storage account. The company has enabled Amazon S3, Amazon Athena, and AWS Glue in the data storage account.

The company must ensure that calls that the application makes to Amazon Bedrock use only private connectivity between the company's application VPC and Amazon Bedrock.

The company's data lake must provide fine-grained column-level access across the company's AWS accounts.

Which solution will meet these requirements?

- A. In the application account, create interface VPC endpoints for Amazon Bedrock runtime
- B. Run Lambda functions in private subnet
- C. Use IAM conditions on inference and data-plane policies to allow calls only to approved endpoints and role
- D. In the data storage account, use AWS Lake Formation LF-tag-based access control to create table-level and column-level cross-account grants.
- E. Run Lambda functions in private subnet
- F. Configure a NAT gateway to provide access to Amazon Bedrock and the data lake
- G. Use S3 bucket policies and ACLs to manage permission
- H. Export AWS CloudTrail logs to Amazon S3 to perform weekly reviews.
- I. Create a gateway endpoint only for Amazon S3 in the application account
- J. Invoke Amazon Bedrock through public endpoint
- K. Use database-level grants in AWS Lake Formation to manage data access
- L. Stream AWS CloudTrail logs to Amazon CloudWatch Log
- M. Do not set up metric filters or alarms.
- N. Use VPC endpoints to provide access to Amazon Bedrock and Amazon S3 in the application account
- O. Use only IAM path-based policies to manage data lake access
- P. Send AWS CloudTrail logs to Amazon CloudWatch Log
- Q. Periodically create dashboards and allow public fallback for cross-Region reads to reduce setup time.

Answer: B

NEW QUESTION 21

An ecommerce company is using Amazon Bedrock to build a generative AI (GenAI) application. The application uses AWS Step Functions to orchestrate a multi-agent workflow to produce detailed product descriptions. The workflow consists of three sequential states: a description generator, a technical specifications validator, and a brand voice consistency checker. Each state produces intermediate reasoning traces and outputs that are passed to the next state. The application uses an Amazon S3 bucket for process storage and to store outputs.

During testing, the company discovers that outputs between Step Functions states frequently exceed the 256 KB quota and cause workflow failures. A GenAI Developer needs to revise the application architecture to efficiently handle the Step Functions 256 KB quota and maintain workflow observability. The revised architecture must preserve the existing multi-agent reasoning and acting (ReAct) pattern.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Store intermediate outputs in Amazon DynamoD
- B. Pass only references between state
- C. Create a Map state that retrieves the complete data from DynamoDB when required for each agent's processing step.
- D. Configure an Amazon Bedrock integration to use the S3 bucket URI in the input parameters for large output
- E. Use the ResultPath and ResultSelector fields to route S3 references between the agent steps while maintaining the sequential validation workflow.
- F. Use AWS Lambda functions to compress outputs to less than 256 KB before each agent stat
- G. Configure each agent task to decompress outputs before processing and to compress results before passing them to the next state.
- H. Configure a separate Step Functions state machine to handle each agent's processin
- I. Use Amazon EventBridge to coordinate the execution flow between state machine
- J. Use S3 references for the outputs as event data.

Answer: B

NEW QUESTION 22

A company upgraded its Amazon Bedrock-powered foundation model (FM) that supports a multilingual customer service assistant. After the upgrade, the assistant exhibited inconsistent behavior across languages. The assistant began generating different responses in some languages when presented with identical questions. The company needs a solution to detect and address similar problems for future updates. The evaluation must be completed within 45 minutes for all supported languages. The evaluation must process at least 15,000 test conversations in parallel. The evaluation process must be fully automated and integrated into the CI/CD pipeline. The solution must block deployment if quality thresholds are not met.

Which solution will meet these requirements?

- A. Create a distributed traffic simulation framework that sends translation-heavy workloads to the assistant in multiple languages simultaneousl
- B. Use Amazon CloudWatch metrics to monitor latency, concurrency, and throughpu
- C. Run simulations before production releases to identify infrastructure bottlenecks.
- D. Deploy the assistant in multiple AWS Regions with Amazon Route 53 latency-based routing and AWS Global Accelerator to improve global performanc
- E. Store multilingual conversation logs in Amazon S3. Perform weekly post-deployment audits to review consistency.
- F. Create a pre-processing pipeline that normalizes all incoming messages into a consistent format before sending the messages to the assistan
- G. Apply rule-based checks to flag potential hallucinations in the output
- H. Focus evaluation on normalized text to simplify testing across languages.
- I. Set up standardized multilingual test conversations with identical meanin
- J. Run the test conversations in parallel by using Amazon Bedrock model evaluation job
- K. Apply similarity and hallucination threshold
- L. Integrate the process into the CI/CD pipeline to block releases that fail.

Answer: D

NEW QUESTION 25

A healthcare company is using Amazon Bedrock to build a system to help practitioners make clinical decisions. The system must provide treatment recommendations to physicians based only on approved medical documentation and must cite specific sources. The system must not hallucinate or produce factually incorrect information.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Integrate Amazon Bedrock with Amazon Kendra to retrieve approved document
- B. Implement custom post-processing to compare generated responses against source documents and to include citations.
- C. Deploy an Amazon Bedrock Knowledge Base and connect it to approved clinical source document
- D. Use the Amazon Bedrock RetrieveAndGenerate API to return citations from the knowledge base.
- E. Use Amazon Bedrock and Amazon Comprehend Medical to extract medical entitie
- F. Implement verification logic against a medical terminology database.
- G. Use an Amazon Bedrock knowledge base with Retrieve API calls and InvokeModel API calls to retrieve approved clinical source document
- H. Implement verification logic to compare against retrieved sources and to cite sources.

Answer: B

NEW QUESTION 27

A retail company has a generative AI (GenAI) product recommendation application that uses Amazon Bedrock. The application suggests products to customers based on browsing history and demographics. The company needs to implement fairness evaluation across multiple demographic groups to detect and measure bias in recommendations between two prompt approaches. The company wants to collect and monitor fairness metrics in real time. The company must receive an alert if the fairness metrics show a discrepancy of more than 15% between demographic groups. The company must receive weekly reports that compare the performance of the two prompt approaches.

Which solution will meet these requirements with the LEAST custom development effort?

- A. Configure an Amazon CloudWatch dashboard to display default metrics from Amazon Bedrock API call
- B. Create custom metrics based on model output
- C. Set up Amazon EventBridge rules to invoke AWS Lambda functions that perform post-processing analysis on model responses and publish custom fairness metrics.
- D. Create the two prompt variants in Amazon Bedrock Prompt Managemen
- E. Use Amazon Bedrock Flows to deploy the prompt variants with defined traffic allocatio
- F. Configure Amazon Bedrock guardrails to monitor demographic fairnes
- G. Set up Amazon CloudWatch alarms on the GuardrailContentSource dimension by using InvocationsIntervened metrics to detect recommendation discrepancy threshold violations.

- H. Set up Amazon SageMaker Clarify to analyze model output
- I. Publish fairness metrics to Amazon CloudWatch
- J. Create CloudWatch composite alarms that combine SageMaker Clarify bias metrics with Amazon Bedrock latency metrics.
- K. Create an Amazon Bedrock model evaluation job to compare fairness between the two prompt variants
- L. Enable model invocation logging in Amazon CloudWatch
- M. Set up CloudWatch alarms for InvocationsIntervened metrics with a dimension for each demographic group.

Answer: B

NEW QUESTION 32

A financial services company uses multiple foundation models (FMs) through Amazon Bedrock for its generative AI (GenAI) applications. To comply with a new regulation for GenAI use with sensitive financial data, the company needs a token management solution. The token management solution must proactively alert when applications approach model-specific token limits. The solution must also process more than 5,000 requests each minute and maintain token usage metrics to allocate costs across business units. Which solution will meet these requirements?

- A. Develop model-specific tokenizers in an AWS Lambda function
- B. Configure the Lambda function to estimate token usage before sending requests to Amazon Bedrock
- C. Configure the Lambda function to publish metrics to Amazon CloudWatch and trigger alarms when requests approach threshold
- D. Store detailed token usage in Amazon DynamoDB to report costs.
- E. Implement Amazon Bedrock Guardrails with token quota policies
- F. Capture metrics on rejected requests
- G. Configure Amazon EventBridge rules to trigger notifications based on Amazon Bedrock Guardrails metrics
- H. Use Amazon CloudWatch dashboards to visualize token usage trends across models.
- I. Deploy an Amazon SQS dead-letter queue for failed requests
- J. Configure an AWS Lambda function to analyze token-related failures
- K. Use Amazon CloudWatch Logs Insights to generate reports on token usage patterns based on error logs from Amazon Bedrock API responses.
- L. Use Amazon API Gateway to create a proxy for all Amazon Bedrock API calls
- M. Configure request throttling based on custom usage plans with predefined token quotas
- N. Configure API Gateway to reject requests that will exceed token limits.

Answer: A

NEW QUESTION 36

An e-commerce company operates a global product recommendation system that needs to switch between multiple foundation models (FM) in Amazon Bedrock based on regulations, cost optimization, and performance requirements. The company must apply custom controls based on proprietary business logic, including dynamic cost thresholds, AWS Region-specific compliance rules, and real-time A/B testing across multiple FMs. The system must be able to switch between FMs without deploying new code. The system must route user requests based on complex rules including user tier, transaction value, regulatory zone, and real-time cost metrics that change hourly and require immediate propagation across thousands of concurrent requests. Which solution will meet these requirements?

- A. Deploy an AWS Lambda function that uses environment variables to store routing rules and Amazon Bedrock FM ID
- B. Use the Lambda console to update the environment variables when business requirements change
- C. Configure an Amazon API Gateway REST API to read request parameters to make routing decisions.
- D. Deploy Amazon API Gateway REST API request transformation templates to implement routing logic based on request attributes
- E. Store Amazon Bedrock FM endpoints as REST API stage variables
- F. Update the variables when the system switches between models.
- G. Configure an AWS Lambda function to fetch routing configurations from the AWS AppConfig Agent for each user request
- H. Run business logic in the Lambda function to select the appropriate FM for each request
- I. Expose the FM through a single Amazon API Gateway REST API endpoint.
- J. Use AWS Lambda authorizers for an Amazon API Gateway REST API to evaluate routing rules that are stored in AWS AppConfig
- K. Return authorization contexts based on business logic
- L. Route requests to model-specific Lambda functions for each Amazon Bedrock FM.

Answer: C

NEW QUESTION 41

A financial services company needs to build a document analysis system that uses Amazon Bedrock to process quarterly reports. The system must analyze financial data, perform sentiment analysis, and validate compliance across batches of reports. Each batch contains 5 reports. Each report requires multiple foundation model (FM) calls. The solution must finish the analysis within 10 seconds for each batch. Current sequential processing takes 45 seconds for each batch. Which solution will meet these requirements?

- A. Use AWS Lambda functions with provisioned concurrency to process each analysis type sequentially
- B. Configure the Lambda function timeouts to 10 seconds
- C. Configure automatic retries with exponential backoff.
- D. Use AWS Step Functions with a Parallel state to invoke separate AWS Lambda functions for each analysis type simultaneously
- E. Configure Amazon Bedrock client timeout
- F. Use Amazon CloudWatch metrics to track execution time and model inference latency.
- G. Create an Amazon SQS queue to buffer analysis requests
- H. Deploy multiple AWS Lambda functions with reserved concurrency
- I. Configure each Lambda function to process different aspects of each report sequentially and then combine the results.
- J. Deploy an Amazon ECS cluster that runs containers that process each report sequentially
- K. Use a load balancer to distribute batch workload
- L. Configure an auto-scaling policy based on CPU utilization.

Answer: B

NEW QUESTION 45

A wildlife conservation agency operates zoos globally. The agency uses various sensors, trackers, and audiovisual recorders to monitor animal behavior. The agency wants to launch a generative AI (GenAI) assistant that can ingest multimodal data to study animal behavior. The GenAI assistant must support natural language queries, avoid speculative behavioral interpretations, and maintain audit logs for ethical research audits. Which solution will meet these requirements?

- A. Ingest raw videos into Amazon Rekognition to detect animal postures and expression
- B. Use Amazon Data Firehose to stream sensor and GPS data into Amazon S3. Prompt an Amazon Bedrock FM using basic templates stored in AWS Systems Manager Parameter Store
- C. Use IAM for access control
- D. Use AWS CloudTrail for audit logging.
- E. Use Amazon SageMaker Processing and Amazon Transcribe to pre-process multimodal data
- F. Ingest curated summaries into an Amazon Bedrock Knowledge Base
- G. Apply Amazon Bedrock guardrails to restrict speculative output
- H. Use AWS AppConfig to manage prompt template
- I. Use AWS CloudTrail to log research activity for audits.
- J. Use Amazon OpenSearch Serverless to index behavioral logs and telemetry
- K. Use Amazon Comprehend to extract entities
- L. Use Amazon Bedrock to answer questions over indexed data
- M. Use IAM for access control and CloudTrail for audit logging.
- N. Configure Amazon OpenSearch to federate data across Amazon S3, Amazon Kinesis, and Amazon SageMaker Feature Store
- O. Use EventBridge for ingestion orchestration
- P. Use custom AWS Lambda functions to filter LLM outputs for ethical compliance.

Answer: B

NEW QUESTION 47

A company is developing a generative AI (GenAI)-powered customer support application that uses Amazon Bedrock foundation models (FMs). The application must maintain conversational context across multiple interactions with the same user. The application must run clarification workflows to handle ambiguous user queries. The company must store encrypted records of each user conversation to use for personalization. The application must be able to handle thousands of concurrent users while responding to each user quickly. Which solution will meet these requirements?

- A. Use an AWS Step Functions Express workflow to orchestrate conversation flow
- B. Invoke AWS Lambda functions to run clarification logic
- C. Store conversation history in Amazon RDS and use session IDs as the primary key.
- D. Use an AWS Step Functions Standard workflow to orchestrate clarification workflow
- E. Include Wait for a Callback patterns to manage the workflow
- F. Store conversation history in Amazon DynamoDB
- G. Purchase on-demand capacity and configure server-side encryption.
- H. Deploy the application by using an Amazon API Gateway REST API to route user requests to an AWS Lambda function to update and retrieve conversation context
- I. Store conversation history in Amazon S3 and configure server-side encryption
- J. Save each interaction as a separate JSON file.
- K. Use AWS Lambda functions to call Amazon Bedrock inference API
- L. Use Amazon SQS queues to orchestrate clarification step
- M. Store conversation history in an Amazon ElastiCache (Redis OSS) cluster
- N. Configure encryption at rest.

Answer: B

NEW QUESTION 50

A company is using Amazon Bedrock to develop a customer support AI assistant. The AI assistant must respond to customer questions about their accounts. The AI assistant must not expose personal information in responses. The company must comply with data residency policies by ensuring that all processing occurs within the same AWS Region where each customer is located.

The company wants to evaluate how effective the AI assistant is at preventing the exposure of personal information before the company makes the AI assistant available to customers.

Which solution will meet these requirements?

- A. Configure a cross-Region Amazon Bedrock guardrail to apply sensitive information filter
- B. Set the guardrail to detect mode during development and testing
- C. Switch to block mode for production deployment.
- D. Configure an Amazon Bedrock guardrail to apply sensitive information filter
- E. Set the guardrail to mask mode during development and testing
- F. Switch to block mode for production deployment
- G. Deploy a copy of the guardrail to each Region where the company operates.
- H. Configure an Amazon Bedrock guardrail to apply content and topic filter
- I. Set the guardrail to detect mode during development, testing, and production
- J. Disable invocation logging for the Amazon Bedrock model.
- K. Configure a cross-Region Amazon Bedrock guardrail to apply a set of content and word filter
- L. Set the guardrail to detect mode during development and testing
- M. Switch to mask mode for production deployment.

Answer: B

NEW QUESTION 54

A company is building a video analysis platform on AWS. The platform will analyze a large video archive by using Amazon Rekognition and Amazon Bedrock. The platform must comply with predefined privacy standards. The platform must also use secure model I/O, control foundation model (FM) access patterns, and provide an audit of who accessed what and when.

Which solution will meet these requirements?

- A. Configure VPC endpoints for Amazon Bedrock model API call
- B. Implement Amazon Bedrock guardrails to filter harmful or unauthorized content in prompts and response
- C. Use Amazon Bedrock trace events to track all agent and model invocations for auditing purpose
- D. Export the traces to Amazon CloudWatch Logs as an audit record of model usage
- E. Store all prompts and outputs in Amazon S3 with server-side encryption with AWS KMS keys (SSE-KMS).
- F. Define access control by using IAM with attribute-based access control (ABAC) to map departments to specific permission
- G. Configure VPC endpoints for Amazon Bedrock model API call
- H. Use IAM condition keys to enforce specific GuardrailIdentifier and ModelId value
- I. Configure AWS CloudTrail to capture management and data events for S3 objects and KMS key usage activities
- J. Enable S3 server access logging to record detailed file-level interactions with the video archive
- K. Send all CloudTrail logs to AWS CloudTrail Lake
- L. Set up Amazon CloudWatch alarms to detect and alert on unexpected activity from Amazon Bedrock, Amazon Rekognition, and AWS KMS.
- M. Restrict access to services by using VPC endpoint policies
- N. Use AWS Config to track resource changes and compliance with security rules
- O. Use server-side encryption with AWS KMS keys (SSE-KMS) to encrypt data at rest
- P. Store the model's I/O in separate Amazon S3 buckets
- Q. Enable S3 server access logging to track file-level interactions.
- R. Configure AWS CloudTrail Insights to analyze API call patterns across accounts and detect anomalous activity in Amazon Bedrock, Amazon Rekognition, Amazon S3, and AWS KMS
- S. Deploy Amazon Macie to scan and classify the video archive
- T. Use server-side encryption with AWS KMS keys (SSE-KMS) to encrypt all stored data
- . Configure CloudTrail to capture KMS API usage events for audit purpose
- . Configure Amazon EventBridge rules to process CloudTrail Insights anomalies and Macie findings
- . Use CloudWatch alarms to trigger automated notifications and security responses when potential security issues are detected.

Answer: B

NEW QUESTION 56

A financial services company uses an AI application to process financial documents by using Amazon Bedrock. During business hours, the application handles approximately 10,000 requests each hour, which requires consistent throughput.

The company uses the CreateProvisionedModelThroughput API to purchase provisioned throughput. Amazon CloudWatch metrics show that the provisioned capacity is unused while on-demand requests are being throttled. The company finds the following code in the application:

```
response = bedrock_runtime.invoke_model(modelId="anthropic.claude-v2", body=json.dumps(payload))
```

The company needs the application to use the provisioned throughput and to resolve the throttling issues.

Which solution will meet these requirements?

- A. Increase the number of model units (MUs) in the provisioned throughput configuration.
- B. Replace the model ID parameter with the ARN of the provisioned model that the CreateProvisionedModelThroughput API returns.
- C. Add exponential backoff retry logic to handle throttling exceptions during peak hours.
- D. Modify the application to use the invokeModelWithResponseStream API instead of the invokeModel API.

Answer: B

NEW QUESTION 57

A company developed a multimodal content analysis application by using Amazon Bedrock. The application routes different content types (text, images, and code) to specialized foundation models (FMs).

The application needs to handle multiple types of routing decisions. Simple routing based on file extension must have minimal latency. Complex routing based on content semantics requires analysis before FM selection. The application must provide detailed history and support fallback options when primary FMs fail.

Which solution will meet these requirements?

- A. Configure AWS Lambda functions that call Amazon Bedrock FMs for all routing logic
- B. Use conditional statements to determine the appropriate FM based on content type and semantics.
- C. Create a hybrid solution
- D. Handle simple routing based on file extensions in application code
- E. Handle complex content-based routing by using an AWS Step Functions state machine with JSONata for content analysis and the InvokeModel API for specialized FMs.
- F. Deploy separate AWS Step Functions workflows for each content type with routing logic in AWS Lambda function
- G. Use Amazon EventBridge to coordinate between workflows when fallback to alternate FMs is required.
- H. Use Amazon SQS with different SQS queues for each content type
- I. Configure AWS Lambda consumers that analyze content and invoke appropriate FMs based on message attributes by using Amazon Bedrock with an AWS SDK.

Answer: B

NEW QUESTION 60

A company wants to select a new FM for its AI assistant. A GenAI developer needs to generate evaluation reports to help a data scientist assess the quality and safety of various foundation models (FMs). The data scientist provides the GenAI developer with sample prompts for evaluation. The GenAI developer wants to use Amazon Bedrock to automate report generation and evaluation.

Which solution will meet this requirement?

- A. Combine the sample prompts into a single JSON document
- B. Create an Amazon Bedrock knowledge base with the documents
- C. Write a prompt that asks the FM to generate a response to each sample prompt
- D. Use the RetrieveAndGenerate API to generate a report for each model.
- E. Combine the sample prompts into a single JSONL document
- F. Store the document in an Amazon S3 bucket
- G. Create an Amazon Bedrock evaluation job that uses a judge mode
- H. Specify the S3 location as input and a different S3 location as output
- I. Run an evaluation job for each FM and select the FM as the generator.

- J. Combine the sample prompts into a single JSONL document
- K. Store the document in an Amazon S3 bucket
- L. Create an Amazon Bedrock evaluation job that uses a judge mode
- M. Specify the S3 location as input and Amazon QuickSight as output
- N. Run an evaluation job for each FM and select the FM as the evaluator.
- O. Combine the sample prompts into a single JSON document
- P. Create an Amazon Bedrock knowledge base from the document
- Q. Create an Amazon Bedrock evaluation job that uses the retrieval and response generation evaluation type
- R. Specify an Amazon S3 bucket as the output
- S. Run an evaluation job for each FM.

Answer: B

NEW QUESTION 63

An ecommerce company operates a global product recommendation system that needs to switch between multiple foundation models (FMs) in Amazon Bedrock based on regulations, cost optimization, and performance requirements. The company must apply custom controls based on proprietary business logic, including dynamic cost thresholds, AWS Region-specific compliance rules, and real-time A/B testing across multiple FMs. The system must be able to switch between FMs without deploying new code. The system must route user requests based on complex rules including user tier, transaction value, regulatory zone, and real-time cost metrics that change hourly and require immediate propagation across thousands of concurrent requests. Which solution will meet these requirements?

- A. Deploy an AWS Lambda function that uses environment variables to store routing rules and Amazon Bedrock FM ID
- B. Use the Lambda console to update the environment variables when business requirements change
- C. Configure an Amazon API Gateway REST API to read request parameters to make routing decisions.
- D. Deploy Amazon API Gateway REST API request transformation templates to implement routing logic based on request attribute
- E. Store Amazon Bedrock FM endpoints as REST API stage variable
- F. Update the variables when the system switches between models.
- G. Configure an AWS Lambda function to fetch routing configuration from the AWS AppConfig Agent for each user request
- H. Run business logic in the Lambda function to select the appropriate FM for each request
- I. Expose the FM through a single Amazon API Gateway REST API endpoint.
- J. Use AWS Lambda authorizers for an Amazon API Gateway REST API to evaluate routing rules that are stored in AWS AppConfig
- K. Return authorization contexts based on business logic
- L. Route requests to model-specific Lambda functions for each Amazon Bedrock FM.

Answer: C

NEW QUESTION 64

A medical device company wants to feed reports of medical procedures that used the company's devices into an AI assistant. To protect patient privacy, the AI assistant must expose patient personally identifiable information (PII) only to surgeons. The AI assistant must redact PII for engineers. The AI assistant must reference only medical reports that are less than 3 years old. The company stores reports in an Amazon S3 bucket as soon as each report is published. The company has already set up an Amazon Bedrock Knowledge Bases. The AI assistant uses Amazon Cognito to authenticate users. Which solution will meet these requirements?

- A. Enable Amazon Macie PII detection on the S3 bucket
- B. Use an S3 trigger to invoke an AWS Lambda function that redacts PII from the report
- C. Configure the Lambda function to delete outdated documents and invoke knowledge base syncing.
- D. Invoke an AWS Lambda function to sync the S3 bucket and the knowledge base when a new report is uploaded
- E. Use a second Lambda function with Amazon Comprehend to redact PII for engineer
- F. Use S3 Lifecycle rules to remove reports older than 3 years.
- G. Set up an S3 Lifecycle configuration to remove reports that are older than 3 years
- H. Schedule an AWS Lambda function to run daily syncs between the bucket and the knowledge base
- I. When users interact with the AI assistant, apply a guardrail configuration selected based on the user's Cognito user group to redact PII from responses when required.
- J. Create a second knowledge base
- K. Use Lambda and Amazon Comprehend to redact PII before syncing to the second knowledge base
- L. Route users to the appropriate knowledge base based on Cognito group membership.

Answer: C

NEW QUESTION 66

A financial services company is developing a real-time generative AI (GenAI) assistant to support human call center agents. The GenAI assistant must transcribe live customer speech, analyze context, and provide incremental suggestions to call center agents while a customer is still speaking. To preserve responsiveness, the GenAI assistant must maintain end-to-end latency under 1 second from speech to initial response display. The architecture must use only managed AWS services and must support bidirectional streaming to ensure that call center agents receive updates in real time. Which solution will meet these requirements?

- A. Use Amazon Transcribe streaming to transcribe call
- B. Pass the text to Amazon Comprehend for sentiment analysis
- C. Feed the results to Anthropic Claude on Amazon Bedrock by using the InvokeModel API
- D. Store results in Amazon DynamoDB
- E. Use a WebSocket API to display the results.
- F. Use Amazon Transcribe streaming with partial results enabled to deliver fragments of transcribed text before customers finish speaking
- G. Forward text fragments to Amazon Bedrock by using the InvokeModelWithResponseStream API
- H. Stream responses to call center agents through an Amazon API Gateway WebSocket API.
- I. Use Amazon Transcribe batch processing to convert calls to text
- J. Pass complete transcripts to Anthropic Claude on Amazon Bedrock by using the ConverseStream API
- K. Return responses through an Amazon Lex chatbot interface.
- L. Use the Amazon Transcribe streaming API with an AWS Lambda function to transcribe each audio segment
- M. Call the Amazon Titan Embeddings model on Amazon Bedrock by using the InvokeModel API

N. Publish results to Amazon SNS.

Answer: B

NEW QUESTION 68

A financial services company wants to develop an Amazon Bedrock application that gives analysts the ability to query quarterly earnings reports and financial statements. The financial documents are typically 5–100 pages long and contain both tabular data and text. The application must provide contextually accurate responses that preserve the relationship between financial metrics and their explanatory text. To support accurate and scalable retrieval, the application must incorporate document segmentation and context management strategies.

Which solution will meet these requirements?

- A. Use a direct model invocation approach that uses Anthropic Claude to process each financial document as a single input
- B. Use fine-tuned prompts that instruct the model to parse tables and text separately.
- C. Use Amazon Bedrock Knowledge Bases to create a Retrieval Augmented Generation (RAG) application that retrieves relevant information from contextually chunked sections of financial document
- D. Segment documents based on their structural layout
- E. Include citations that reference the original source materials.
- F. Deploy an Amazon Bedrock agent that has an action group that calls custom AWS Lambda functions to analyze financial document
- G. Configure the Lambda functions to perform fixed-size chunking when a user submits a query about financial metrics.
- H. Create one specialized Amazon Bedrock application that is optimized for structured data
- I. Create a second application that is optimized for unstructured data
- J. Configure each application to use a tailored chunking strategy that is suited to the application's content type
- K. Implement logic to link queries to the appropriate sources.

Answer: B

NEW QUESTION 72

A financial services company is developing a customer service AI assistant by using Amazon Bedrock. The AI assistant must not discuss investment advice with users. The AI assistant must block harmful content, mask personally identifiable information (PII), and maintain audit trails for compliance reporting. The AI assistant must apply content filtering to both user inputs and model responses based on content sensitivity.

The company requires an Amazon Bedrock guardrail configuration that will effectively enforce policies with minimal false positives. The solution must provide multiple handling strategies for multiple types of sensitive content.

Which solution will meet these requirements?

- A. Configure a single guardrail and set content filters to high for all categories
- B. Set up denied topics for investment advice and include sample phrases to block
- C. Set up sensitive information filters that apply the block action for all PII entities
- D. Apply the guardrail to all model inference calls.
- E. Configure multiple guardrails by using tiered policies
- F. Create one guardrail and set content filters to high
- G. Configure the guardrail to block PII for public interaction
- H. Configure a second guardrail and set content filters to medium
- I. Configure the second guardrail to mask PII for internal use
- J. Configure multiple topic-specific guardrails to block investment advice and set up contextual grounding checks.
- K. Configure a guardrail and set content filters to medium for harmful content
- L. Set up denied topics for investment advice and include clear definitions and sample phrases to block
- M. Configure sensitive information filters to mask PII in responses and to block financial information in input
- N. Enable both input and output evaluations that use custom blocked messages for audits.
- O. Create a separate guardrail for each use case
- P. Create one guardrail that applies a harmful content filter
- Q. Create a guardrail to apply topic filters for investment advice
- R. Create a guardrail to apply sensitive information filters to block PII
- S. Use AWS Step Functions to chain the guardrails sequentially.

Answer: C

NEW QUESTION 77

A specialty coffee company has a mobile app that generates personalized coffee roast profiles by using Amazon Bedrock with a three-stage prompt chain. The prompt chain converts user inputs into structured metadata, retrieves relevant logs for coffee roasts, and generates a personalized roast recommendation for each customer.

Users in multiple AWS Regions report inconsistent roast recommendations for identical inputs, slow inference during the retrieval step, and unsafe recommendations such as brewing at excessively high temperatures. The company must improve the stability of outputs for repeated inputs. The company must also improve app performance and the safety of the app's outputs. The updated solution must ensure 99.5% output consistency for identical inputs and achieve inference latency of less than 1 second. The solution must also block unsafe or hallucinated recommendations by using validated safety controls.

Which solution will meet these requirements?

- A. Deploy Amazon Bedrock with provisioned throughput to stabilize inference latency
- B. Apply Amazon Bedrock guardrails with semantic denial rules to block unsafe output
- C. Use Amazon Bedrock Prompt Management to manage prompts by using approval workflows.
- D. Use Amazon Bedrock Agents to manage chains
- E. Log model inputs and outputs to Amazon CloudWatch Log
- F. Use logs from CloudWatch to perform A/B testing for prompt versions.
- G. Cache prompt results in Amazon ElastiCache
- H. Use AWS Lambda functions to pre-process metadata and to trace end-to-end latency
- I. Use AWS X-Ray to identify and remediate performance bottlenecks.
- J. Use Amazon Kendra to improve roast log retrieval accuracy
- K. Store normalized prompt metadata within Amazon DynamoDB
- L. Use AWS Step Functions to orchestrate multi-step prompts.

Answer: A

NEW QUESTION 80

An ecommerce company is developing a generative AI (GenAI) solution that uses Amazon Bedrock with Anthropic Claude to recommend products to customers. Customers report that some recommended products are not available for sale or are not relevant. Customers also report long response times for some recommendations.

The company confirms that most customer interactions are unique and that the solution recommends products not present in the product catalog. Which solution will meet this requirement?

- A. Increase grounding within Amazon Bedrock Guardrail
- B. Enable automated reasoning check
- C. Set up provisioned throughput.
- D. Use prompt engineering to restrict model responses to relevant product
- E. Use streaming inference to reduce perceived latency.
- F. Create an Amazon Bedrock Knowledge Bases and implement Retrieval Augmented Generation (RAG). Set the PerformanceConfigLatency parameter to optimized.
- G. Store product catalog data in Amazon OpenSearch Service
- H. Validate model recommendations against the catalog
- I. Use Amazon DynamoDB for response caching.

Answer: C

NEW QUESTION 82

A company uses AWS Lambda functions to build an AI agent solution. A GenAI developer must set up a Model Context Protocol (MCP) server that accesses user information. The GenAI developer must also configure the AI agent to use the new MCP server. The GenAI developer must ensure that only authorized users can access the MCP server.

Which solution will meet these requirements?

- A. Use a Lambda function to host the MCP server
- B. Grant the AI agent Lambda functions permission to invoke the Lambda function that hosts the MCP server
- C. Configure the AI agent's MCP client to invoke the MCP server asynchronously.
- D. Use a Lambda function to host the MCP server
- E. Grant the AI agent Lambda functions permission to invoke the Lambda function that hosts the MCP server
- F. Configure the AI agent to use the STDIO transport with the MCP server.
- G. Use a Lambda function to host the MCP server
- H. Create an Amazon API Gateway HTTP API that proxies requests to the Lambda function
- I. Configure the AI agent solution to use the Streamable HTTP transport to make requests through the HTTP API
- J. Use Amazon Cognito to enforce OAuth 2.1.
- K. Use a Lambda layer to host the MCP server
- L. Add the Lambda layer to the AI agent Lambda function
- M. Configure the agentic AI solution to use the STDIO transport to send requests to the MCP server
- N. In the AI agent's MCP configuration, specify the Lambda layer ARN as the command
- O. Specify the user credentials as environment variables.

Answer: C

NEW QUESTION 85

A pharmaceutical company is developing a Retrieval Augmented Generation (RAG) application that uses an Amazon Bedrock knowledge base. The knowledge base uses Amazon OpenSearch Service as a data source for more than 25 million scientific papers. Users report that the application produces inconsistent answers that cite irrelevant sections of papers when queries span methodology, results, and discussion sections of the papers.

The company needs to improve the knowledge base to preserve semantic context across related paragraphs on the scale of the entire corpus of data.

Which solution will meet these requirements?

- A. Configure the knowledge base to use fixed-size chunking
- B. Set a 300-token maximum chunk size and a 10% overlap between chunks
- C. Use an appropriate Amazon Bedrock embedding model.
- D. Configure the knowledge base to use hierarchical chunking
- E. Use parent chunks that contain 1,000 tokens and child chunks that contain 200 tokens
- F. Set a 50-token overlap between chunks.
- G. Configure the knowledge base to use semantic chunking
- H. Use a buffer size of 1 and a breakpoint percentile threshold of 85% to determine chunk boundaries based on content meaning.
- I. Configure the knowledge base not to use chunking
- J. Manually split each document into separate files before ingestion
- K. Apply post-processing reranking during retrieval.

Answer: B

NEW QUESTION 88

A software company is using Amazon Q Business to build an AI assistant that allows employees to access company information and personal information by using natural language prompts. The company stores this information in an Amazon S3 bucket.

Each department in the company has a dedicated prefix in the S3 bucket. Each object name includes the S3 prefix of the department that it belongs to. Each department can belong to only a single group in AWS IAM Identity Center. Each employee belongs to a single department.

The company configures Amazon Q Business to access data stored in an S3 bucket as a data source. The company needs to ensure that the AI assistant respects access controls based on the user's IAM Identity Center group membership.

Which solution will meet this requirement with the LEAST operational overhead?

- A. Create a JSON file named acl.json in each department folder
- B. In each file, create access control entries that specify the IAM Identity Center group that should have access to that department's data
- C. Indicate the location of the JSON file in the Access Control section of the data source settings.
- D. Create a single JSON file named acl.json at the top level of the S3 bucket
- E. Add access control entries that map each department's S3 prefix to its corresponding IAM Identity Center group

- F. Indicate the location of the JSON file in the Access Control section of the data source settings.
- G. For each IAM Identity Center group, create a separate permissions set that denies access to all prefixes in the S3 bucket
- H. Add a StringNotEquals condition key to the permissions set for each group that specifies the department each group is associated with
- I. Attach the permissions sets to the Identity Center groups.
- J. Create a metadata file named metadata.json at the top level of the S3 bucket
- K. Add anAccessControlList object to the file that specifies the S3 path of each department's prefix
- L. Specify the IAM Identity Center group that should have access to each department's prefix
- M. Reference the file location in the data source metadata settings.

Answer: B

NEW QUESTION 92

A company provides a service that helps users from around the world discover new restaurants. The service has 50 million monthly active users. The company wants to implement a semantic search solution across a database that contains 20 million restaurants and 200 million reviews. The company currently stores the data in PostgreSQL.

The solution must support complex natural language queries and return results for at least 95% of queries within 500 ms. The solution must maintain data freshness for restaurant details that update hourly. The solution must also scale cost-effectively during peak usage periods.

Which solution will meet these requirements with the LEAST development effort?

- A. Migrate the restaurant data to Amazon OpenSearch Service
- B. Implement keyword-based search rules that use custom analyzers and relevance tuning to find restaurants based on attributes such as cuisine type, features, and location
- C. Create Amazon API Gateway HTTP API endpoints to transform user queries into structured search parameters.
- D. Migrate the restaurant data to Amazon OpenSearch Service
- E. Use a foundation model (FM) in Amazon Bedrock to generate vector embeddings from restaurant descriptions, reviews, and menu items
- F. When users submit natural language queries, convert the queries to embeddings by using the same FM
- G. Perform k-nearest neighbors (k-NN) searches to find semantically similar results.
- H. Keep the restaurant data in PostgreSQL and implement a pgvector extension
- I. Use a foundation model (FM) in Amazon Bedrock to generate vector embeddings from restaurant data
- J. Store the vector embeddings directly in PostgreSQL
- K. Create an AWS Lambda function to convert natural language queries to vector representations by using the same FM
- L. Configure the Lambda function to perform similarity searches within the database.
- M. Migrate restaurant data to an Amazon Bedrock knowledge base by using a custom ingestion pipeline
- N. Configure the knowledge base to automatically generate embeddings from restaurant information
- O. Use the Amazon Bedrock Retrieve API with built-in vector search capabilities to query the knowledge base directly by using natural language input.

Answer: B

NEW QUESTION 94

A healthcare company is using Amazon Bedrock to build a Retrieval Augmented Generation (RAG) application that helps practitioners make clinical decisions. The application must achieve high accuracy for patient information retrievals, identify hallucinations in generated content, and reduce human review costs.

Which solution will meet these requirements?

- A. Use Amazon Comprehend to analyze and classify RAG responses and to extract medical entities and relationships
- B. Use AWS Step Functions to orchestrate automated evaluation
- C. Configure Amazon CloudWatch metrics to track entity recognition confidence score
- D. Configure CloudWatch to send an alert when accuracy falls below specified thresholds.
- E. Implement automated large language model (LLM)-based evaluations that use a specialized model that is fine-tuned for medical content to assess all responses
- F. Deploy AWS Lambda functions to parallelize evaluation
- G. Publish results to Amazon CloudWatch metrics that track relevance and factual accuracy.
- H. Configure Amazon CloudWatch Synthetics to generate test queries that have known answers on a regular schedule, and track model success rate
- I. Set up dashboards that compare synthetic test results against expected outcomes.
- J. Deploy a hybrid evaluation system that uses an automated LLM-as-a-judge evaluation to initially screen responses and targeted human reviews for edge cases
- K. Use a built-in Amazon Bedrock evaluation to track retrieval precision and hallucination rates.

Answer: D

NEW QUESTION 98

A company is implementing a serverless inference API by using AWS Lambda. The API will dynamically invoke multiple AI models hosted on Amazon Bedrock. The company needs to design a solution that can switch between model providers without modifying or redeploying Lambda code in real time. The design must include safe rollout of configuration changes and validation and rollback capabilities.

Which solution will meet these requirements?

- A. Store the active model provider in AWS Systems Manager Parameter Store
- B. Configure a Lambda function to read the parameter at runtime to determine which model to invoke.
- C. Store the active model provider in AWS AppConfig
- D. Configure a Lambda function to read the configuration at runtime to determine which model to invoke.
- E. Configure an Amazon API Gateway REST API to route requests to separate Lambda functions
- F. Hardcode each Lambda function to a specific model provider
- G. Switch the integration target manually.
- H. Store the active model provider in a JSON file hosted on Amazon S3. Use AWS AppConfig to reference the S3 file as a hosted configuration source
- I. Configure a Lambda function to read the file through AppConfig at runtime to determine which model to invoke.

Answer: B

NEW QUESTION 101

A company is designing a canary deployment strategy for a payment processing API. The system must support automated gradual traffic shifting between multiple Amazon Bedrock models based on real-time inference metrics, historical traffic patterns, and service health. The solution must be able to gradually increase traffic to new model versions. The system must increase traffic if metrics remain healthy and decrease traffic if the performance degrades below acceptable thresholds.

The company needs to comprehensively monitor inference latency and error rates during the deployment phase. The company must also be able to halt deployments and revert to a previous model version without any manual intervention. Which solution will meet these requirements?

- A. Use Amazon Bedrock with provisioned throughput to host model version
- B. Configure an Amazon EventBridge rule to invoke an AWS Step Functions workflow when a new model version is release
- C. Configure the workflow to shift traffic in stages, wait for a specified time period, and invoke an AWS Lambda function to check Amazon CloudWatch performance metric
- D. Configure the workflow to increase traffic if metrics meet thresholds and to trigger a traffic rollback if performance metrics fall below thresholds.
- E. Use AWS Lambda functions to invoke various Amazon Bedrock model version
- F. Use an Amazon API Gateway HTTP API with stage variables and weighted routing to shift traffic graduall
- G. Use Amazon CloudWatch to monitor performanc
- H. Use external logic to adjust traffic and roll back if performance falls below thresholds.
- I. Use Amazon SageMaker AI endpoint variants to represent multiple Amazon Bedrock model version
- J. Use variant weights to shift traffi
- K. Use Amazon CloudWatch and SageMaker Model Monitor to trigger rollback
- L. Use EventBridge to roll back deployments if an anomaly is detected.
- M. Use Amazon OpenSearch Service to track inference log
- N. Configure OpenSearch Service to invoke an AWS Systems Manager Automation runbook to update Amazon Bedrock model endpoints to shift traffic based on inference logs.

Answer: A

NEW QUESTION 105

A bank is developing a generative AI (GenAI)-powered AI assistant that uses Amazon Bedrock to assist the bank??s website users with account inquiries and financial guidance. The bank must ensure that the AI assistant does not reveal any personally identifiable information (PII) in customer interactions. The AI assistant must not send PII in prompts to the GenAI model. The AI assistant must not respond to customer requests to provide investment advice. The bank must collect audit logs of all customer interactions, including any images or documents that are transmitted during customer interactions. Which solution will meet these requirements with the LEAST operational effort?

- A. Use Amazon Macie to detect and redact PII in user inputs and in the model response
- B. Apply prompt engineering techniques to force the model to avoid investment advice topic
- C. Use AWS CloudTrail to capture conversation logs.
- D. Use an AWS Lambda function and Amazon Comprehend to detect and redact PI
- E. Use Amazon Comprehend topic modeling to prevent the AI assistant from discussing investment advice topic
- F. Set up custom metrics in Amazon CloudWatch to capture customer conversations.
- G. Configure Amazon Bedrock guardrails to apply a sensitive information policy to detect and filter PI
- H. Set up a topic policy to ensure that the AI assistant avoids investment advice topic
- I. Use the Converse API to log model invocation
- J. Enable delivery and image logging to Amazon S3.
- K. Use regex controls to match patterns for PI
- L. Apply prompt engineering techniques to avoid returning PII or investment advice topics to customer
- M. Enable model invocation logging, delivery logging, and image logging to Amazon S3.

Answer: C

NEW QUESTION 110

A financial services company needs to pre-process unstructured data such as customer transcripts, financial reports, and documentation. The company stores the unstructured data in Amazon S3 to support an Amazon Bedrock application. The company must validate data quality, create auditable metadata, monitor data metrics, and customize text chunking to optimize foundation model (FM) performance. Which solution will meet these requirements with the LEAST development effort?

- A. Use Amazon SageMaker Data Wrangler to create a data flo
- B. Configure Amazon CloudWatch metrics and alarms to monitor data qualit
- C. Use a custom AWS Lambda function to pre-process the dat
- D. Load processed data into Amazon Bedrock.
- E. Set up an AWS Glue crawler to catalog data source
- F. Create AWS Glue ETL jobs to run custom transformation script
- G. Use AWS Glue Data Quality to validate and monitor data qualit
- H. Load processed data into Amazon Bedrock.
- I. Use Amazon Comprehend to extract entitie
- J. Create an AWS Lambda function to chunk tex
- K. Run Amazon Athena to query and validate data qualit
- L. Load processed data into Amazon Bedrock.
- M. Create an AWS Step Functions workflow to orchestrate data pre-processing task
- N. Run custom code on Amazon EC2 instance
- O. Use Amazon SageMaker Model Monitor to monitor data qualit
- P. Load processed data into Amazon Bedrock.

Answer: B

NEW QUESTION 113

A company is using Amazon Bedrock to develop an AI-powered application that uses a foundation model that supports cross-Region inference and provisioned throughput. The application must serve users in Europe and North America with consistently low latency. The application must comply with data residency regulations that require European user data to remain within Europe-based AWS Regions. During testing, the application experiences service degradation when Regional traffic spikes reach service quotas. The company needs a solution that maintains application resilience and minimizes operational complexity. Which solution will meet these requirements?

- A. Deploy separate Amazon Bedrock instances in North American and European Region
- B. Use a custom routing layer that directs traffic based on user location
- C. Configure Amazon CloudWatch alarms to monitor Regional service usage
- D. Use Amazon SNS to send email alerts to the company when usage approaches specified thresholds.
- E. Use Amazon Bedrock cross-Region inference profiles by specifying geographical codes in profile IDs when the application calls the InvokeModel API
- F. Configure separate Amazon API Gateway HTTP APIs to direct European and North American users to the appropriate Regional endpoints.
- G. Deploy a multi-Region Amazon API Gateway HTTP API and AWS Lambda functions that implement retry logic to handle throttling
- H. Configure the Lambda functions to call the foundation model in the nearest secondary Region when the application reaches service quotas in the primary Region
- I. Use intelligent routing to ensure compliance with data residency requirements.
- J. Configure provisioned throughput for Amazon Bedrock in multiple Regions
- K. Implement failover logic in the application code to switch between Regions when throttling occurs
- L. Use AWS Global Accelerator to route traffic to the appropriate endpoints based on user location.

Answer: B

NEW QUESTION 115

A company is creating a generative AI (GenAI) application that uses Amazon Bedrock foundation models (FMs). The application must use Microsoft Entra ID to authenticate. All FM API calls must stay on private network paths. Access to the application must be limited by department to specific model families. The company also needs a comprehensive audit trail of model interactions. Which solution will meet these requirements?

- A. Configure SAML federation between Microsoft Entra ID and AWS Identity and Access Management
- B. Create department-specific IAM roles that allow only the required ModelId value
- C. Create AWS PrivateLink interface VPC endpoints for Amazon Bedrock runtime service
- D. Enable AWS CloudTrail to capture Amazon Bedrock API calls
- E. Configure Amazon Bedrock model invocation logging to record detailed model interactions.
- F. Create an identity provider (IdP) connection in IAM to authenticate by using Microsoft Entra ID
- G. Assign department permission sets to control access to specific model families
- H. Deploy AWS Lambda functions in private subnets with a NAT gateway for egress to Amazon Bedrock public endpoints
- I. Enable CloudWatch Logs to capture model interactions for auditing purposes.
- J. Create a SAML identity provider (IdP) in IAM to authenticate by using Microsoft Entra ID
- K. Use IAM permissions boundaries to limit department roles' access to specific model families
- L. Configure public Amazon Bedrock API endpoints with VPC routing to maintain private network connectivity
- M. Set up CloudTrail with Amazon S3 Lifecycle rules to manage audit logs of model interactions.
- N. Configure OpenID Connect (OIDC) federation between Microsoft Entra ID and IAM
- O. Use attribute-based access control to map department attributes to specific model access permissions
- P. Apply SCP policies to restrict access to Amazon Bedrock FM families based on departments
- Q. Use Microsoft Entra ID's built-in logging capabilities to maintain an audit trail of model interactions.

Answer: A

NEW QUESTION 118

An enterprise application uses an Amazon Bedrock foundation model (FM) to process and analyze 50 to 200 pages of technical documents. Users are experiencing inconsistent responses and receiving truncated outputs when processing documents that exceed the FM's context window limits. Which solution will resolve this problem?

- A. Configure fixed-size chunking at 4,000 tokens for each chunk with 20% overlap
- B. Use application-level logic to link multiple chunks sequentially until the FM's maximum context window of 200,000 tokens is reached before making inference calls.
- C. Use hierarchical chunking with parent chunks of 8,000 tokens and child chunks of 2,000 tokens
- D. Use Amazon Bedrock Knowledge Bases built-in retrieval to automatically select relevant parent chunks based on query context
- E. Configure overlap tokens to maintain semantic continuity.
- F. Use semantic chunking with a breakpoint percentile threshold of 95% and a buffer size of 3 sentences
- G. Use the RetrieveAndGenerate API to dynamically select the most relevant chunks based on embedding similarity scores.
- H. Create a pre-processing AWS Lambda function that analyzes document token count by using the FM's tokenize API
- I. Configure the Lambda function to split documents into equal segments that fit within 80% of the context window
- J. Configure the Lambda function to process each segment independently before aggregating the results.

Answer: C

NEW QUESTION 122

A medical company uses Amazon Bedrock to power a clinical documentation summarization system. The system produces inconsistent summaries when handling complex clinical documents. The system performed well on simple clinical documents.

The company needs a solution that diagnoses inconsistencies, compares prompt performance against established metrics, and maintains historical records of prompt versions.

Which solution will meet these requirements?

- A. Create multiple prompt variants by using Prompt management in Amazon Bedrock
- B. Manually test the prompts with simple clinical documents
- C. Deploy the highest performing version by using the Amazon Bedrock console.
- D. Implement version control for prompts in a code repository with a test suite that contains complex clinical documents and quantifiable evaluation metrics
- E. Use an automated testing framework to compare prompt versions and document performance patterns.
- F. Deploy each new prompt version to separate Amazon Bedrock API endpoints
- G. Split production traffic between the endpoints
- H. Configure Amazon CloudWatch to capture response metrics and user feedback for automatic version selection.
- I. Create a custom prompt evaluation flow in Amazon Bedrock Flows that applies the same clinical document inputs to different prompt variants
- J. Use Amazon Comprehend Medical to analyze and score the factual accuracy of each version.

Answer: B

NEW QUESTION 125

A company provides a service that helps users from around the world discover new restaurants. The service has 50 million monthly active users. The company wants to implement a semantic search solution across a database that contains 20 million restaurants and 200 million reviews. The company currently stores the data in a PostgreSQL database.

The solution must support complex natural language queries and return results for at least 95% of queries within 500 ms. The solution must maintain data freshness for restaurant details that update hourly. The solution must also scale cost-effectively during peak usage periods.

Which solution will meet these requirements with the LEAST development effort?

- A. Migrate the restaurant data to Amazon OpenSearch Service
- B. Implement keyword-based search rules that use custom analyzers and relevance tuning to find restaurants based on attributes such as cuisine type, feature, and location
- C. Create Amazon API Gateway HTTP API endpoints to transform user queries into structured search parameters.
- D. Migrate the restaurant data to Amazon OpenSearch Service
- E. Use a foundation model (FM) in Amazon Bedrock to generate vector embeddings from restaurant descriptions, reviews, and menu items
- F. When users submit natural language queries, convert the queries to embeddings by using the same FM
- G. Perform k-nearest neighbors (k-NN) searches to find semantically similar results.
- H. Keep the restaurant data in PostgreSQL and implement a pgvector extension
- I. Use a foundation model (FM) in Amazon Bedrock to generate vector embeddings from restaurant data
- J. Store the vector embeddings directly in PostgreSQL
- K. Create an AWS Lambda function to convert natural language queries to vector representations by using the same FM
- L. Configure the Lambda function to perform similarity searches within the database.
- M. Migrate the restaurant data to an Amazon Bedrock knowledge base by using a custom ingestion pipeline
- N. Configure the knowledge base to automatically generate embeddings from restaurant information
- O. Use the Amazon Bedrock Retrieve API with built-in vector search capabilities to query the knowledge base directly by using natural language input.

Answer: D

NEW QUESTION 128

A company is developing a customer support application that uses Amazon Bedrock foundation models (FMs) to provide real-time AI assistance to the company's employees. The application must display AI-generated responses character by character as the responses are generated. The application needs to support thousands of concurrent users with minimal latency. The responses typically take 15 to 45 seconds to finish.

Which solution will meet these requirements?

- A. Configure an Amazon API Gateway WebSocket API with an AWS Lambda integration
- B. Configure the WebSocket API to invoke the Amazon Bedrock InvokeModelWithResponseStream API and stream partial responses through WebSocket connections.
- C. Configure an Amazon API Gateway REST API with an AWS Lambda integration
- D. Configure the REST API to invoke the Amazon Bedrock standard InvokeModel API and implement frontend client-side polling every 100 ms for complete response chunks.
- E. Implement direct frontend client connections to Amazon Bedrock by using IAM user credentials and the InvokeModelWithResponseStream API without any intermediate gateway or proxy layer.
- F. Configure an Amazon API Gateway HTTP API with an AWS Lambda integration
- G. Configure the HTTP API to cache complete responses in an Amazon DynamoDB table and serve the responses through multiple paginated GET requests to frontend clients.

Answer: A

NEW QUESTION 131

A media company must use Amazon Bedrock to implement a robust governance process for AI-generated content. The company needs to manage hundreds of prompt templates. Multiple teams use the templates across multiple AWS Regions to generate content. The solution must provide version control with approval workflows that include notifications for pending reviews. The solution must also provide detailed audit trails that document prompt activities and consistent prompt parameterization to enforce quality standards.

Which solution will meet these requirements?

- A. Configure Amazon Bedrock Studio prompt template
- B. Use Amazon CloudWatch dashboards to display prompt usage metrics
- C. Store approval status in Amazon DynamoDB
- D. Use AWS Lambda functions to enforce approvals.
- E. Use Amazon Bedrock Prompt Management to implement version control
- F. Configure AWS CloudTrail for audit logging
- G. Use AWS Identity and Access Management policies to control approval permissions
- H. Create parameterized prompt templates by specifying variables.
- I. Use AWS Step Functions to create an approval workflow
- J. Store prompts in Amazon S3. Use tags to implement version control
- K. Use Amazon EventBridge to send notifications.
- L. Deploy Amazon SageMaker Canvas with prompt templates stored in Amazon S3. Use AWS CloudFormation for version control
- M. Use AWS Config to enforce approval policies.

Answer: B

NEW QUESTION 136

A financial services company is developing a customer service AI assistant application that uses a foundation model (FM) in Amazon Bedrock. The application must provide transparent responses by documenting reasoning and by citing sources that are used for Retrieval Augmented Generation (RAG). The application must capture comprehensive audit trails for all responses to users. The application must be able to serve up to 10,000 concurrent users and must respond to each customer inquiry within 2 seconds.

Which solution will meet these requirements with the LEAST operational overhead?

- A. Enable tracing for Amazon Bedrock Agent
- B. Configure structured prompts that direct the FM to provide evidence presentation
- C. Integrate Amazon Bedrock Knowledge Bases with data sources to enable RAG

- D. Configure the application to reference and cite authoritative content
- E. Deploy the application in a Multi-AZ architecture
- F. Use Amazon API Gateway and AWS Lambda functions to scale the application
- G. Use Amazon CloudFront to provide low-latency delivery.
- H. Enable tracing for Amazon Bedrock agent
- I. Integrate a custom RAG pipeline with Amazon OpenSearch Service to retrieve and cite source
- J. Configure structured prompts to present retrieved evidence
- K. Deploy the application behind an Amazon API Gateway REST API
- L. Use AWS Lambda functions and Amazon CloudFront to scale the application and to provide low latency
- M. Store logs in Amazon S3 and use AWS CloudTrail to capture audit trails.
- N. Use Amazon CloudWatch to monitor latency and error rate
- O. Embed model prompts directly in the application backend to cite source
- P. Store application interactions with users in Amazon RDS for audits.
- Q. Store generated responses and supporting evidence in an Amazon S3 bucket
- R. Enable versioning on the bucket for audit
- S. Use AWS Glue to catalog retrieved documents
- T. Process the retrieved documents in Amazon Athena to generate periodic compliance reports.

Answer: A

NEW QUESTION 139

A company is developing a generative AI (GenAI) application by using Amazon Bedrock. The application will analyze patterns and relationships in the company's data. The application will process millions of new data points daily across AWS Regions in Europe, North America, and Asia before storing the data in Amazon S3. The application must comply with local data protection and storage regulations. Data residency and processing must occur within the same continent. The application must also maintain audit trails of the application's decision-making processes and provide data classification capabilities. Which solution will meet these requirements?

- A. Deploy the application in each Region with local IAM policies
- B. Use Amazon Bedrock cross-Region inference to distribute the workload
- C. Use Amazon CloudWatch to log AI decision-making processes
- D. Manually track compliance certifications across Regions.
- E. Use SCPs with AWS Organizations to manage location-specific permissions
- F. Use AWS CloudTrail immutable logs to audit decision-making processes
- G. Import a custom model into Amazon Bedrock and deploy the model to each Region.
- H. Use Amazon S3 Object Lock with Region-specific S3 bucket policies
- I. Pre-process the data points within the Region based on geographic origin before sending the data points to Amazon Bedrock
- J. Use Amazon Macie to classify the data
- K. Use AWS CloudTrail immutable logs to audit the decision-making processes.
- L. Create separate AWS accounts for each Region with individual compliance frameworks
- M. Use Amazon SageMaker AI with custom monitoring
- N. Create manual compliance reports for each regulatory jurisdiction.

Answer: C

NEW QUESTION 143

A company is building a multicloud generative AI (GenAI)-powered secret resolution application that uses Amazon Bedrock and Agent Squad. The application resolves secrets from multiple sources, including key stores and hardware security modules (HSMs). The application uses AWS Lambda functions to retrieve secrets from the sources. The application uses AWS AppConfig to implement dynamic feature gating. The application supports secret chaining and detects secret drift. The application handles short-lived and expiring secrets. The application also supports prompt flows for templated instructions. The application uses AWS Step Functions to orchestrate agents to resolve the secrets and to manage secret validation and drift detection. The company finds multiple issues during application testing. The application does not refresh expired secrets in time for agents to use. The application sends alerts for secret drift, but agents still use stale data. Prompt flows within the application reuse outdated templates, which cause cascading failures. The company must resolve the performance issues. Which solution will meet this requirement?

- A. Use Step Functions Map states to run agent workflows in parallel
- B. Pass updated secret metadata through Lambda function output
- C. Use AWS AppConfig to version all prompt flows to gate and roll back faulty templates.
- D. Use Amazon Bedrock Agents only
- E. Configure Amazon Bedrock guardrails to restrict prompt variations
- F. Use an inline JSON schema for a single agent's workflow definition to chain tool calls.
- G. Use a centralized Amazon EventBridge pipeline to invoke each agent
- H. Store intermediate prompts in Amazon DynamoDB
- I. Resolve agent ordering by using TTL-based backoff and retries.
- J. Use Amazon EventBridge Pipes to invoke resolvers based on Amazon CloudWatch log patterns
- K. Store response metadata in DynamoDB with TTL and versioned writes
- L. Use Amazon Q Developer to dynamically generate fallback prompts.

Answer: A

NEW QUESTION 145

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

AIP-C01 Practice Exam Features:

- * AIP-C01 Questions and Answers Updated Frequently
- * AIP-C01 Practice Questions Verified by Expert Senior Certified Staff
- * AIP-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * AIP-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The AIP-C01 Practice Test Here](#)