

Amazon-Web-Services

Exam Questions AIP-C01

AWS Certified Generative AI Developer - Professional



NEW QUESTION 1

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 2

A company is using Amazon Bedrock to design an application to help researchers apply for grants. The application is based on an Amazon Nova Pro foundation model (FM). The application contains four required inputs and must provide responses in a consistent text format. The company wants to receive a notification in Amazon Bedrock if a response contains bullying language. However, the company does not want to block all flagged responses.

The company creates an Amazon Bedrock flow that takes an input prompt and sends it to the Amazon Nova Pro FM. The Amazon Nova Pro FM provides a response.

Which additional steps must the company take to meet these requirements? (Select TWO.)

- A. Use Amazon Bedrock Prompt Management to specify the required inputs as variable
- B. Select an Amazon Nova Pro F
- C. Specify the output format for the respons
- D. Add the prompt to the prompts node of the flow.
- E. Create an Amazon Bedrock guardrail that applies the hate content filte
- F. Set the filter response to bloc
- G. Add the guardrail to the prompts node of the flow.
- H. Create an Amazon Bedrock prompt route
- I. Specify an Amazon Nova Pro F
- J. Add the required inputs as variables to the input node of the flo
- K. Add the prompt router to the prompts nod
- L. Add the output format to the output node.
- M. Create an Amazon Bedrock guardrail that applies the insults content filte
- N. Set the filter response to detec
- O. Add the guardrail to the prompts node of the flow.
- P. Create an Amazon Bedrock application inference profile that specifies an Amazon Nova Pro F
- Q. Specify the output format for the response in the descriptio
- R. Include a tag for each of the input variable
- S. Add the profile to the prompts node of the flow.

Answer: AD

NEW QUESTION 3

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 4

A company has set up Amazon Q Developer Pro licenses for all developers at the company. The company maintains a list of approved resources that developers must use when developing applications. The approved resources include internal libraries, proprietary algorithmic techniques, and sample code with approved styling.

A new team of developers is using Amazon Q Developer to develop a new Java-based application. The company must ensure that the new developer team uses the company's approved resources. The company does not want to make project-level modifications.

Which solution will meet these requirements?

- A. Create a Git repository that contains all of the approved internal libraries, algorithms, and code sample
- B. Include this Git repository in the application project locally as part of the workspace
- C. Ensure that the developers use the workspace context to retrieve suggestions from the Git repository.
- D. In the project root folder, create a folder named amazonq/rule
- E. Add the approved internal libraries, algorithms, and code samples to the folder.
- F. Create a folder in the application project named rule
- G. Store the guidelines and code in the folder for Amazon Q Developer to reference for code suggestions.
- H. Create an Amazon Q Developer customization that includes the approved data source
- I. Ensure that the developers use the customization to develop the application.

Answer: D

NEW QUESTION 5

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 6

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 request
- R. Configure exponential backoff retries, dynamic chunk sizing, and context-aware token limits.

Answer: B

NEW QUESTION 7

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 8

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 DynamoDB
- 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 state
- 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 processing
- 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 9

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 entities
- 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 10

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 Management
- E. Use Amazon Bedrock Flows to deploy the prompt variants with defined traffic allocation
- F. Configure Amazon Bedrock guardrails to monitor demographic fairness

- 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 CloudWatc
- 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 variant
- L. Enable model invocation logging in Amazon CloudWatc
- M. Set up CloudWatch alarms for InvocationsIntervened metrics with a dimension for each demographic group.

Answer: B

NEW QUESTION 10

An ecommerce 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 chang
- 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 configurations from the AWS AppConfig Agent for each user reques
- H. Run business logic in the Lambda function to select the appropriate FM for each reques
- 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 logi
- L. Route requests to model-specific Lambda functions for each Amazon Bedrock FM.

Answer: C

NEW QUESTION 15

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 flo
- B. Invoke AWS Lambda functions to run clarification logi
- 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 DynamoD
- 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 contex
- I. Store conversation history in Amazon S3 and configure server-side encryptio
- 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) cluste
- N. Configure encryption at rest.

Answer: B

NEW QUESTION 19

A company has a recommendation system running 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 experiences intermittent issues where some recommendations do not match customer preferences. The company needs an observability solution to monitor operational metrics and detect patterns of performance degradation compared to established baselines. The solution must 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 Insight
- B. Set up alarms for latency threshold
- C. Add custom token metrics using the CloudWatch embedded metric format.
- D. Implement AWS X-Ray
- E. Enable CloudWatch Logs Insight
- F. Set up AWS CloudTrail and create dashboards in Amazon QuickSight.
- G. Enable Amazon CloudWatch Application Insight
- H. Create custom metrics for recommendation quality, token usage, and response latency using the CloudWatch embedded metric format with dimensions for request types and user segment
- I. Configure CloudWatch anomaly detection on model metric
- J. Use CloudWatch Logs Insights for pattern analysis.

- K. Use Amazon OpenSearch Service with the Observability plugi
- L. Ingest metrics and logs through Amazon Kinesis and analyze behavior with custom queries.

Answer: C

NEW QUESTION 22

A pharmaceutical company is developing a Retrieval Augmented Generation 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 chunkin
- B. Set a 300-token maximum chunk size and a 10% overlap between chunk
- C. Use an appropriate Amazon Bedrock embedding model.
- D. Configure the knowledge base to use hierarchical chunkin
- E. Use parent chunks that contain 1,000 tokens and child chunks that contain 200 token
- F. Set a 50-token overlap between chunks.
- G. Configure the knowledge base to use semantic chunkin
- 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 chunkin
- J. Manually split each document into separate files before ingestio
- K. Apply post-processing reranking during retrieval.

Answer: B

NEW QUESTION 23

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 24

A company is creating a workflow to review customer-facing communications before the company sends the communications. The company uses a pre-defined message template to generate the communications and stores the communications in an Amazon S3 bucket. The workflow needs to capture a specific portion from the template and send it to an Amazon Bedrock model. The workflow must store model responses back to the original S3 bucket. Which solution will meet these requirements?

- A. Create a flow in Amazon Bedrock Flow
- B. Configure S3 action nodes at the beginning and end of the flow to retrieve and store the communications and the model response
- C. In the middle of the flow, configure an expression to parse each communicatio
- D. Configure an agent step to send the parsed input to the model for review.
- E. Create an AWS Step Functions Express workflow state machin
- F. Use an Amazon S3 integration GetObject step to retrieve the original communication
- G. Use an intrinsic function Pass step to parse the communications and to pass the results to an Amazon Bedrock InvokeModel ste
- H. Configure an Amazon S3 integration PutObject step to store the model responses back to the S3 bucket.
- I. Create an Amazon Bedrock agent that has an action grou
- J. Configure instructions to define how the agent should parse the communication
- K. Configure the action group to retrieve the communications from the S3 bucket, invoke the Amazon Bedrock model, and store the model responses back to the S3 bucket.
- L. Create an Amazon Bedrock agent that has a single action grou
- M. Configure three AWS Lambda functions in the action grou
- N. Configure the functions to retrieve the communications from the S3 bucket, parse the communications and invoke the Amazon Bedrock model, and store the model responses back to the S3 bucket.

Answer: A

NEW QUESTION 25

A healthcare company is developing an application to process medical queries. The application must answer complex queries with high accuracy by reducing semantic dilution. The application must refer to domain-specific terminology in medical documents to reduce ambiguity in medical terminology. The application must be able to respond to 1,000 queries each minute with response times less than 2 seconds. Which solution will meet these requirements with the LEAST operational overhead?

- A. Use Amazon API Gateway to route incoming queries to an Amazon Bedrock agen
- B. Configure the agent to use an Anthropic Claude model to decompose queries and an Amazon Titan model to expand querie
- C. Create an Amazon Bedrock knowledge base to store the reference medical documents.
- D. Configure an Amazon Bedrock knowledge base to store the reference medical document
- E. Enable query decomposition in the knowledge bas

- F. Configure an Amazon Bedrock flow that uses a foundation model and the knowledge base to support the application.
- G. Use Amazon SageMaker AI to host custom ML models for both query decomposition and query expansion.
- H. Configure Amazon Bedrock knowledge bases to store the reference medical document.
- I. Encrypt the documents in the knowledge base.
- J. Create an Amazon Bedrock agent to orchestrate multiple AWS Lambda functions to decompose queries.
- K. Create an Amazon Bedrock knowledge base to store the reference medical document.
- L. Use the agent's built-in knowledge base capabilities.
- M. Add deep research and reasoning capabilities to the agent to reduce ambiguity in the medical terminology.

Answer: B

NEW QUESTION 29

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 33

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 35

A company uses Amazon Bedrock to generate technical content for customers. The company has recently experienced a surge in hallucinated outputs when the company's model generates summaries of long technical documents. The model outputs include inaccurate or fabricated details. The company's current solution uses a large foundation model (FM) with a basic one-shot prompt that includes the full document in a single input.

The company needs a solution that will reduce hallucinations and meet factual accuracy goals. The solution must process more than 1,000 documents each hour and deliver summaries within 3 seconds for each document.

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

- A. Implement zero-shot chain-of-thought (CoT) instructions that require step-by-step reasoning with explicit fact verification before the model generates each summary.
- B. Use Retrieval Augmented Generation (RAG) with an Amazon Bedrock knowledge base.
- C. Apply semantic chunking and tuned embeddings to ground summaries in source content.
- D. Configure Amazon Bedrock guardrails to block any generated output that matches patterns that are associated with hallucinated content.
- E. Increase the temperature parameter in Amazon Bedrock.
- F. Prompt the Amazon Bedrock model to summarize each full document in one pass.

Answer: BC

NEW QUESTION 39

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 41

A company uses an organization in AWS Organizations with all features enabled to manage multiple AWS accounts. Employees use Amazon Bedrock across multiple accounts. The company must prevent specific topics and proprietary information from being included in prompts to Amazon Bedrock models. The company must ensure that employees can use only approved Amazon Bedrock models. The company wants to manage these controls centrally.

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

- A. Create an IAM permissions boundary for each employee's IAM role
- B. Configure the permissions boundary to require an approved Amazon Bedrock guardrail identifier to invoke Amazon Bedrock model
- C. Create an SCP that allows employees to use only approved models.
- D. Create an SCP that allows employees to use only approved model
- E. Configure the SCP to require employees to specify a guardrail identifier in calls to invoke an approved model.
- F. Create an SCP that prevents an employee from invoking a model if a centrally deployed guardrail identifier is not specified in a call to the model
- G. Create a permissions boundary on each employee's IAM role that allows each employee to invoke only approved models.
- H. Use AWS CloudFormation to create a custom Amazon Bedrock guardrail that has a block filtering policy
- I. Use stack sets to deploy the guardrail to each account in the organization.
- J. Use AWS CloudFormation to create a custom Amazon Bedrock guardrail that has a mask filtering policy
- K. Use stack sets to deploy the guardrail to each account in the organization.

Answer: CD

NEW QUESTION 42

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 45

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 AP
- J. Use Amazon Cognito to enforce OAuth 2.1.
- K. Use a Lambda layer to host the MCP serve
- 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 serve
- 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 47

A hotel company wants to enhance a legacy Java-based property management system (PMS) by adding AI capabilities. The company wants to use Amazon Bedrock Knowledge Bases to provide staff with room availability information and hotel-specific details. The solution must maintain separate access controls for each hotel that the company manages. The solution must provide room availability information in near real time and must maintain consistent performance during peak usage periods.

Which solution will meet these requirements?

- A. Deploy a single Amazon Bedrock knowledge base that contains combined data for all hotel
- B. Configure AWS Lambda functions to synchronize data from each hotel's PMS database through direct API connection
- C. Implement AWS CloudTrail logging with hotel-specific filters to audit access logs for each hotel's data.
- D. Create an Amazon EventBridge rule for each hotel that is invoked by changes to the PMS database
- E. Configure the rule to send updates to a centralized Amazon Bedrock knowledge base in a management AWS account
- F. Configure resource-based policies to enforce hotel-specific access controls.
- G. Implement one Amazon Bedrock knowledge base for each hotel in a multi-account structure
- H. Use direct data ingestion to provide near real-time room availability information
- I. Schedule regular synchronization for less critical information.
- J. Build a centralized Amazon Bedrock Agents solution that uses multiple knowledge bases
- K. Implement AWS IAM Identity Center with hotel-specific permission sets to control staff access.

Answer: C

NEW QUESTION 49

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 52

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 54

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 relationship
- 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 response
- 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 case
- K. Use a built-in Amazon Bedrock evaluation to track retrieval precision and hallucination rates.

Answer: D

NEW QUESTION 57

A company is building a generative AI (GenAI) application that processes financial reports and provides summaries for analysts. The application must run two compute environments. In one environment, AWS Lambda functions must use the Python SDK to analyze reports on demand. In the second environment, Amazon EKS containers must use the JavaScript SDK to batch process multiple reports on a schedule. The application must maintain conversational context throughout multi-turn interactions, use the same foundation model (FM) across environments, and ensure consistent authentication. Which solution will meet these requirements?

- A. Use the Amazon Bedrock InvokeModel API with a separate authentication method for each environmen
- B. Store conversation states in Amazon DynamoD
- C. Use custom I/O formatting logic for each programming language.
- D. Use the Amazon Bedrock Converse API directly in both environments with a common authentication mechanism that uses IAM role
- E. Store conversation states in Amazon ElastiCach
- F. Create programming language-specific wrappers for model parameters.
- G. Create a centralized Amazon API Gateway REST API endpoint that handles all model interactions by using the InvokeModel AP
- H. Store interaction history in application process memory in each Lambda function or EKS containe
- I. Use environment variables to configure model parameters.
- J. Use the Amazon Bedrock Converse API and IAM roles for authenticatio
- K. Pass previous messages in the request messages array to maintain conversational contex
- L. Use programming language-specific SDKs to establish consistent API interfaces.

Answer: D

NEW QUESTION 60

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 locatio
- C. Configure Amazon CloudWatch alarms to monitor Regional service usag
- 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 AP
- 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 throttlin
- H. Configure the Lambda functions to call the foundation model in the nearest secondary Region when the application reaches service quotas in the primary Regio
- I. Use intelligent routing to ensure compliance with data residency requirements.
- J. Configure provisioned throughput for Amazon Bedrock in multiple Region
- K. Implement failover logic in the application code to switch between Regions when throttling occur
- L. Use AWS Global Accelerator to route traffic to the appropriate endpoints based on user location.

Answer: B

NEW QUESTION 64

A company has a customer service application that uses Amazon Bedrock to generate personalized responses to customer inquiries. The company needs to establish a quality assurance process to evaluate prompt effectiveness and model configurations across updates. The process must automatically compare outputs from multiple prompt templates, detect response quality issues, provide quantitative metrics, and allow human reviewers to give feedback on responses. The process must prevent configurations that do not meet a predefined quality threshold from being deployed. Which solution will meet these requirements?

- A. Create an AWS Lambda function that sends sample customer inquiries to multiple Amazon Bedrock model configurations and stores responses in Amazon S3. Use Amazon QuickSight to visualize response pattern
- B. Manually review outputs dail
- C. Use AWS CodePipeline to deploy configurations that meet the quality threshold.
- D. Use Amazon Bedrock evaluation jobs to compare model outputs by using custom prompt dataset
- E. Configure AWS CodePipeline to run the evaluation jobs when prompt templates chang

- F. Configure CodePipeline to deploy only configurations that exceed the predefined quality threshold.
- G. Set up Amazon CloudWatch alarms to monitor response latency and error rates from Amazon Bedrock
- H. Use Amazon EventBridge rules to notify teams when thresholds are exceeded
- I. Configure a manual approval workflow in AWS Systems Manager.
- J. Use AWS Lambda functions to create an automated testing framework that samples production traffic and routes duplicate requests to the updated model version
- K. Use Amazon Comprehend sentiment analysis to compare results
- L. Block deployment if sentiment scores decrease.

Answer: B

NEW QUESTION 66

A financial services company is building a customer support application that retrieves relevant financial regulation documents from a database based on semantic similarity to user queries. The application must integrate with Amazon Bedrock to generate responses. The application must search documents in English, Spanish, and Portuguese. The application must filter documents by metadata such as publication date, regulatory agency, and document type. The database stores approximately 10 million document embeddings. To minimize operational overhead, the company wants a solution that minimizes management and maintenance effort while providing low-latency responses for real-time customer interactions. Which solution will meet these requirements?

- A. Use Amazon OpenSearch Serverless to provide vector search capabilities and metadata filtering
- B. Integrate with Amazon Bedrock Knowledge Bases to enable Retrieval Augmented Generation (RAG) using an Anthropic Claude foundation model.
- C. Deploy an Amazon Aurora PostgreSQL database with the pgvector extension
- D. Store embeddings and metadata in a table
- E. Use SQL queries for similarity search and send results to Amazon Bedrock for response generation.
- F. Use Amazon S3 Vectors to configure a vector index and non-filterable metadata field
- G. Integrate S3 Vectors with Amazon Bedrock for RAG.
- H. Set up an Amazon Neptune Analytics database with a vector index
- I. Use graph-based retrieval and Amazon Bedrock for response generation.

Answer: A

NEW QUESTION 69

A financial services company is developing a Retrieval Augmented Generation (RAG) application to help investment analysts query complex financial relationships across multiple investment vehicles, market sectors, and regulatory environments. The dataset contains highly interconnected entities that have multi-hop relationships. Analysts must examine relationships holistically to provide accurate investment guidance. The application must deliver comprehensive answers that capture indirect relationships between financial entities and must respond in less than 3 seconds. Which solution will meet these requirements with the LEAST operational overhead?

- A. Use Amazon Bedrock Knowledge Bases with GraphRAG and Amazon Neptune Analytics to store financial data
- B. Analyze multi-hop relationships between entities and automatically identify related information across documents.
- C. Use Amazon Bedrock Knowledge Bases and an Amazon OpenSearch Service vector store to implement custom relationship identification logic that uses AWS Lambda to query multiple vector embeddings in sequence.
- D. Use Amazon OpenSearch Serverless vector search with k-nearest neighbor (k-NN). Implement manual relationship mapping in an application layer that runs on Amazon EC2 Auto Scaling.
- E. Use Amazon DynamoDB to store financial data in a custom indexing system
- F. Use AWS Lambda to query relevant records
- G. Use Amazon SageMaker to generate responses.

Answer: A

NEW QUESTION 73

A media company is launching a platform that allows thousands of users every hour to upload images and text content. The platform uses Amazon Bedrock to process the uploaded content to generate creative compositions. The company needs a solution to ensure that the platform does not process or produce inappropriate content. The platform must not expose personally identifiable information (PII) in the compositions. The solution must integrate with the company's existing Amazon S3 storage workflow. Which solution will meet these requirements with the LEAST infrastructure management overhead?

- A. Enable the Enhanced Monitoring tool
- B. Use an Amazon CloudWatch alarm to filter traffic to the platform
- C. Use Amazon Comprehend PII detection to pre-process the data
- D. Create a CloudWatch alarm to monitor for Amazon Comprehend PII detection events
- E. Create an AWS Step Functions workflow that includes an Amazon Rekognition image moderation step.
- F. Use an Amazon API Gateway HTTP API with request validation templates to screen content before storing the uploaded content in Amazon S3. Use Amazon SageMaker AI to build custom content moderation models that process content before sending the processed content to Amazon Bedrock.
- G. Create an Amazon Cognito user pool that uses pre-authentication AWS Lambda functions to run content moderation checks
- H. Use Amazon Textract to filter text content and Amazon Rekognition to filter image content before allowing users to upload content to the platform.
- I. Create an AWS Step Functions workflow that uses built-in Amazon Bedrock guardrails to filter content
- J. Use Amazon Comprehend PII detection to pre-process the content
- K. Use Amazon Rekognition image moderation.

Answer: D

NEW QUESTION 76

A bank is building a generative AI (GenAI) application that uses Amazon Bedrock to assess loan applications by using scanned financial documents. The application must extract structured data from the documents. The application must redact personally identifiable information (PII) before inference. The application must use foundation models (FMs) to generate approvals. The application must route low-confidence document extraction results to human reviewers who are within the same AWS Region as the loan applicant. The company must ensure that the application complies with strict Regional data residency and auditability requirements. The application must be able to scale to handle 25,000 applications each day and provide 99.9% availability. Which combination of solutions will meet these requirements? (Select THREE.)

- A. Deploy Amazon Textract and Amazon Augmented AI within the same Region to extract relevant data from the scanned document
- B. Route low-confidence pages to human reviewers.
- C. Use AWS Lambda functions to detect and redact PII from submitted documents before inference
- D. Apply Amazon Bedrock guardrails to prevent inappropriate or unauthorized content in model output
- E. Configure Region-specific IAM roles to enforce data residency requirements and to control access to the extracted data.
- F. Use Amazon Kendra and Amazon OpenSearch Service to extract field-level values semantically from the uploaded documents before inference.
- G. Store uploaded documents in Amazon S3 and apply object metadata
- H. Configure IAM policies to store original documents within the same Region as each applicant
- I. Enable object tagging for future audits.
- J. Use AWS Glue Data Quality to validate the structured document data
- K. Use AWS Step Functions to orchestrate a review workflow that includes a prompt engineering step that transforms validated data into optimized prompts before invoking Amazon Bedrock to assess loan applications.
- L. Use Amazon SageMaker Clarify to generate fairness and bias reports based on model scoring decisions that Amazon Bedrock makes.

Answer: ABD

NEW QUESTION 79

A company is using Amazon Bedrock and Anthropic Claude 3 Haiku to develop an AI assistant. The AI assistant normally processes 10,000 requests each hour but experiences surges of up to 30,000 requests each hour during peak usage periods. The AI assistant must respond within 2 seconds while operating across multiple AWS Regions.

The company observes that during peak usage periods, the AI assistant experiences throughput bottlenecks that cause increased latency and occasional request timeouts. The company must resolve the performance issues.

Which solution will meet this requirement?

- A. Purchase provisioned throughput and sufficient model units (MUs) in a single Region
- B. Configure the application to retry failed requests with exponential backoff.
- C. Implement token batching to reduce API overhead
- D. Use cross-Region inference profiles to automatically distribute traffic across available Regions.
- E. Set up auto scaling AWS Lambda functions in each Region
- F. Implement client-side round-robin request distribution
- G. Purchase one model unit (MU) of provisioned throughput as a backup.
- H. Implement batch inference for all requests by using Amazon S3 buckets across multiple Regions
- I. Use Amazon SQS to set up an asynchronous retrieval process.

Answer: B

NEW QUESTION 81

A company is building an AI advisory application by using Amazon Bedrock. The application will provide recommendations to customers. The company needs the application to explain its reasoning process and cite specific sources for data. The application must retrieve information from company data sources and show step-by-step reasoning for recommendations. The application must also link data claims to source documents and maintain response latency under 3 seconds.

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

- A. Use Amazon Bedrock Knowledge Bases with source attribution enabled
- B. Use the Anthropic Claude Messages API with RAG to set high-relevance thresholds for sourced documents
- C. Store reasoning and citations in Amazon S3 for auditing purposes.
- D. Use Amazon Bedrock with Anthropic Claude models and extended thinking
- E. Configure a 4,000-token thinking budget
- F. Store reasoning traces and citations in Amazon DynamoDB for auditing purposes.
- G. Configure Amazon SageMaker AI with a custom Anthropic Claude model
- H. Use the model's reasoning parameter and AWS Lambda to process response
- I. Add source citations from a separate Amazon RDS database.
- J. Use Amazon Bedrock with Anthropic Claude models and chain-of-thought reasoning
- K. Configure custom retrieval tracking with the Amazon Bedrock Knowledge Bases API
- L. Use Amazon CloudWatch to monitor response latency metrics.

Answer: A

NEW QUESTION 85

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 89

A GenAI developer is evaluating Amazon Bedrock foundation models (FMs) to enhance a Europe-based company's internal business application. The company

has a multi-account landing zone in AWS Control Tower. The company uses Service Control Policies (SCPs) to allow its accounts to use only the eu-north-1 and eu-west-1 Regions. All customer data must remain in private networks within the approved AWS Regions.

The GenAI developer selects an FM based on analysis and testing and hosts the model in the eu-central-1 Region and the eu-west-3 Region. The GenAI developer must enable access to the FM for the company's employees. The GenAI developer must ensure that requests to the FM are private and remain within the same Regions as the FM.

Which solution will meet these requirements?

- A. Deploy an AWS Lambda function that is exposed by a private Amazon API Gateway REST API to a VPC in eu-north-1. Create a VPC endpoint for the selected FM in eu-central-1 and eu-west-3. Extend existing SCPs to allow employees to use the F
- B. Integrate the REST API with the business application.
- C. Deploy the FM on Amazon EC2 instances in eu-north-1. Deploy a private Amazon API Gateway REST API in front of the EC2 instance
- D. Configure an Amazon Bedrock VPC endpoint
- E. Integrate the REST API with the business application.
- F. Configure the FM to use cross-Region inference through a Europe-scoped endpoint
- G. Configure an Amazon Bedrock VPC endpoint
- H. Extend existing SCPs to allow employees to use the FM through inference profiles in Europe-based Regions where the FM is availabl
- I. Use an inference profile to integrate Amazon Bedrock with the business application.
- J. Deploy the FM in Amazon SageMaker in eu-north-1. Configure a SageMaker VPC endpoint
- K. Extend existing SCPs to allow employees to use the SageMaker endpoint
- L. Integrate the FM in SageMaker with the business application.

Answer: C

NEW QUESTION 90

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 paralle
- 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 onl
- E. Configure Amazon Bedrock guardrails to restrict prompt variatio
- 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 agen
- H. Store intermediate prompts in Amazon DynamoD
- I. Resolve agent ordering by using TTL-based backoff and retries.
- J. Use Amazon EventBridge Pipes to invoke resolvers based on Amazon CloudWatch log pattern
- K. Store response metadata in DynamoDB with TTL and versioned write
- L. Use Amazon Q Developer to dynamically generate fallback prompts.

Answer: A

NEW QUESTION 93

A company is using Amazon Bedrock to develop an AI-powered application that uses a foundation model (FM) 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 locatio
- C. Configure Amazon CloudWatch alarms to monitor Regional service usag
- D. Use Amazon SNS to send email alerts when usage approaches thresholds.
- E. Use Amazon Bedrock cross-Region inference profiles by specifying geographical codes in profile IDs when calling the InvokeModel AP
- 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 throttlin
- H. Configure the Lambda functions to call the FM in the nearest secondary Region when quotas are reached.
- I. Configure provisioned throughput for Amazon Bedrock in multiple Region
- J. Implement failover logic in application code to switch Regions when throttling occur
- K. Use AWS Global Accelerator to route traffic based on user location.

Answer: B

NEW QUESTION 95

.....

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](#)