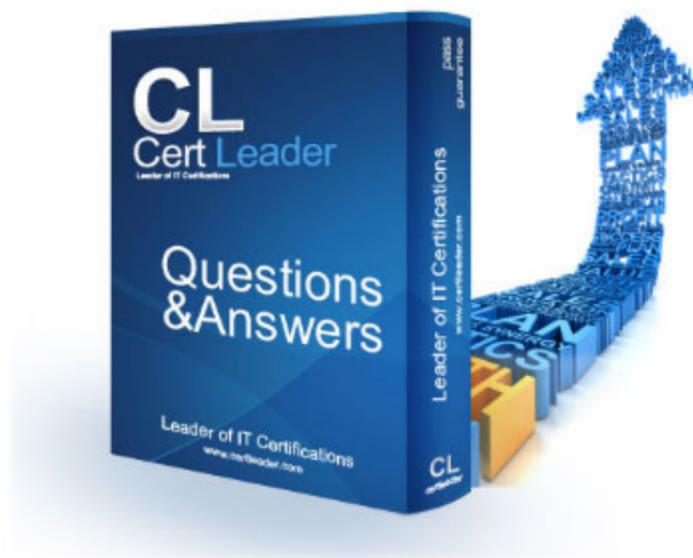


AIP-C01 Dumps

AWS Certified Generative AI Developer - Professional

<https://www.certleader.com/AIP-C01-dumps.html>



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

An ecommerce company is developing a generative AI application that uses Amazon Bedrock with Anthropic Claude to recommend products to customers.

Customers report that some recommended products are not available for sale on the website or are not relevant to the customer. Customers also report that the solution takes a long time to generate some recommendations.

The company investigates the issues and finds that most interactions between customers and the product recommendation solution are unique. The company confirms that the solution recommends products that are not in the company's product catalog. The company must resolve these issues.

Which solution will meet this requirement?

- A. Increase grounding within Amazon Bedrock Guardrail
- B. Enable Automated Reasoningcheck
- C. Set up provisioned throughput.
- D. Use prompt engineering to restrict the model responses to relevant product
- E. Use streaming techniques such as the InvokeModelWithResponseStream action to reduce perceived latency for the customers.
- F. Create an Amazon Bedrock knowledge base
- G. Implement Retrieval Augmented Generation RA
- H. Set the PerformanceConfigLatency parameter to optimized.
- I. Store product catalog data in Amazon OpenSearch Service
- J. Validate the model's product recommendations against the product catalog
- K. Use Amazon DynamoDB to implement response caching.

Answer: C

NEW QUESTION 3

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 application
- B. Create multiple pipelines in AWS CodePipeline
- C. Configure each pipeline to have its own settings for each FM
- 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 environment
- F. Configure the applications to invoke the Amazon Bedrock FMs by using the `aws_bedrock.FoundationModel.fromFoundationModelId()` method
- G. Create a separate pipeline in AWS CodePipeline for each environment.
- H. Create one AWS CDK application
- I. Configure the application to invoke the Amazon Bedrock FMs by using the `aws_bedrock.FoundationModel.fromFoundationModelId()` method
- 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 environment
- L. Configure the application to invoke the Amazon Bedrock FMs by using the `aws_bedrock.ProvisionedModel.fromProvisionedModelArn()` method
- M. Create a pipeline in AWS CodePipeline
- N. Configure the pipeline to deploy to the production environment by using an AWS CodeBuild deploy action
- O. For the development environment, manually recreate the resources by referring to the production application code.

Answer: C

NEW QUESTION 4

A company needs a system to automatically generate study materials from multiple content sources. The content sources include document files (PDF files, PowerPoint presentations, and Word documents) and multimedia files (recorded videos). The system must process more than 10,000 content sources daily with peak loads of 500 concurrent uploads. The system must also extract key concepts from document files and multimedia files and create contextually accurate summaries. The generated study materials must support real-time collaboration with version control.

Which solution will meet these requirements?

- A. Use Amazon Bedrock Data Automation (BDA) with AWS Lambda functions to orchestrate document file processing
- B. Use Amazon Bedrock Knowledge Bases to process all multimedia

- C. Store the content in Amazon DocumentDB with replicatio
- D. Collaborate by using Amazon SNS topic subscription
- E. Track changes by using Amazon Bedrock Agents.
- F. Use Amazon Bedrock Data Automation (BDA) with foundation models (FMs) to process document file
- G. Integrate BDA with Amazon Textract for PDF extraction and with Amazon Transcribe for multimedia file
- H. Store the processed content in Amazon S3 with versioning enable
- I. Store the metadata in Amazon DynamoD
- J. Collaborate in real time by using AWS AppSync GraphQL subscriptions and DynamoDB.
- K. Use Amazon Bedrock Data Automation (BDA) with Amazon SageMaker AI endpoints to host content extraction and summarization model
- L. Use Amazon Bedrock Guardrails to extract content from all file type
- M. Store document files in Amazon Neptune for time series analysi
- N. Collaborate by using Amazon Bedrock Chat for real-time messaging.
- O. Use Amazon Bedrock Data Automation (BDA) with AWS Lambda functions to process batches of content file
- P. Fine-tune foundation models (FMs) in Amazon Bedrock to classify documents across all content type
- Q. Store the processed data in Amazon ElastiCache (Redis OSS) by using Cluster Mode with shardin
- R. Use Prompt management in Amazon Bedrock for version control.

Answer: B

NEW QUESTION 5

A healthcare company is developing a document management system that stores medical research papers in an Amazon S3 bucket. The company needs a comprehensive metadata framework to improve search precision for a GenAI application. The metadata must include document timestamps, author information, and research domain classifications.

The solution must maintain a consistent metadata structure across all uploaded documents and allow foundation models (FMs) to understand document context without accessing full content.

Which solution will meet these requirements?

- A. Store document timestamps in Amazon S3 system metadat
- B. Use S3 object tags for domain classificatio
- C. Implement custom user-defined metadata to store author information.
- D. Set up S3 Object Lock with legal holds to track document timestamp
- E. Use S3 object tags for author informatio
- F. Implement S3 access points for domain classification.
- G. Use S3 Inventory reports to track timestamp
- H. Create S3 access points for domain classificatio
- I. Store author information in S3 Storage Lens dashboards.
- J. Use custom user-defined metadata to store author informatio
- K. Use S3 Object Lock retention periods for timestamp
- L. Use S3 Event Notifications for domain classification.

Answer: A

NEW QUESTION 6

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 7

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 8

A book publishing company wants to build a book recommendation system that uses an AI assistant. The AI assistant will use ML to generate a list of recommended books from the company's book catalog. The system must suggest books based on conversations with customers. The company stores the text of the books, customers' and editors' reviews of the books, and extracted book metadata in Amazon S3. The system must support low-latency responses and scale efficiently to handle more than 10,000 concurrent users. Which solution will meet these requirements?

- A. Use Amazon Bedrock Knowledge Bases to generate embedding
- B. Store the embeddings as a vector store in Amazon OpenSearch Service
- C. Create an AWS Lambda function that queries the knowledge base
- D. Configure Amazon API Gateway to invoke the Lambda function when handling user requests.
- E. Use Amazon Bedrock Knowledge Bases to generate embedding
- F. Store the embeddings as a vector store in Amazon DynamoDB
- G. Create an AWS Lambda function that queries the knowledge base
- H. Configure Amazon API Gateway to invoke the Lambda function when handling user requests.
- I. Use Amazon SageMaker AI to deploy a pre-trained model to build a personalized recommendation engine for book
- J. Deploy the model as a SageMaker AI endpoint
- K. Invoke the model endpoint by using Amazon API Gateway.
- L. Create an Amazon Kendra GenAI Enterprise Edition index that uses the S3 connector to index the book catalog data stored in Amazon S3. Configure built-in FAQ in the Kendra index
- M. Develop an AWS Lambda function that queries the Kendra index based on user conversation
- N. Deploy Amazon API Gateway to expose this functionality and invoke the Lambda function.

Answer: A

NEW QUESTION 9

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 10

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 10

An e-commerce 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 states
- 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 machines
- J. Use S3 references for the outputs as event data.

Answer: B

NEW QUESTION 12

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 simultaneously
- B. Use Amazon CloudWatch metrics to monitor latency, concurrency, and throughput
- 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 performance
- 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 assistant
- 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 meanings
- J. Run the test conversations in parallel by using Amazon Bedrock model evaluation jobs
- K. Apply similarity and hallucination thresholds
- L. Integrate the process into the CI/CD pipeline to block releases that fail.

Answer: D

NEW QUESTION 15

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 documents
- 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 documents
- 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 documents
- H. Implement verification logic to compare against retrieved sources and to cite sources.

Answer: B

NEW QUESTION 17

Example Corp provides a personalized video generation service that millions of enterprise customers use. Customers generate marketing videos by submitting prompts to the company's proprietary generative AI (GenAI) model. To improve output relevance and personalization, Example Corp wants to enhance the prompts by using customer-specific context such as product preferences, customer attributes, and business history.

The customers have strict data governance requirements. The customers must retain full ownership and control over their own data. The customers do not require

real-time access. However, semantic accuracy must be high and retrieval latency must remain low to support customer experience use cases. Example Corp wants to minimize architectural complexity in its integration pattern. Example Corp does not want to deploy and manage services in each customer's environment unless necessary. Which solution will meet these requirements?

- A. Ensure that each customer sets up an Amazon Q Business index that includes the customer's internal data
- B. Ensure that each customer designates Example Corp as a data accessor to allow Example Corp to retrieve relevant content by using a secure API to enrich prompts at runtime.
- C. Use federated search with Model Context Protocol (MCP) by deploying real-time MCP servers for each customer
- D. Retrieve data in real time during prompt generation.
- E. Ensure that each customer configures an Amazon Bedrock knowledge base
- F. Allow cross-account querying so Example Corp can retrieve structured data for prompt augmentation.
- G. Configure Amazon Kendra to crawl customer data source
- H. Share the resulting indexes across accounts so Example Corp can query each customer's Amazon Kendra index to retrieve augmentation data.

Answer: A

NEW QUESTION 22

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 24

A company uses Amazon Bedrock to implement a Retrieval Augmented Generation (RAG)-based system to serve medical information to users. The company needs to compare multiple chunking strategies, evaluate the generation quality of two foundation models (FMs), and enforce quality thresholds for deployment. Which Amazon Bedrock evaluation configuration will meet these requirements?

- A. Create a retrieve-only evaluation job that uses a supported version of Anthropic Claude Sonnet as the evaluator model
- B. Configure metrics for context relevance and context coverage
- C. Define deployment thresholds in a separate CI/CD pipeline.
- D. Create a retrieve-and-generate evaluation job that uses custom precision-at-k metrics and an LLM-as-a-judge metric with a scale of 1–5. Include each chunking strategy in the evaluation dataset
- E. Use a supported version of Anthropic Claude Sonnet to evaluate responses from both FMs.
- F. Create a separate evaluation job for each chunking strategy and FM combination
- G. Use Amazon Bedrock built-in metrics for correctness and completeness
- H. Manually review scores before deployment approval.
- I. Set up a pipeline that uses multiple retrieve-only evaluation jobs to assess retrieval quality
- J. Create separate evaluation jobs for both FMs that use Amazon Nova Pro as the LLM-as-a-judge model
- K. Evaluate based on faithfulness and citation precision metrics.

Answer: B

NEW QUESTION 28

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 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 31

A company uses Amazon Bedrock to build a Retrieval Augmented Generation (RAG) system. The RAG system uses an Amazon Bedrock Knowledge Bases that is based on an Amazon S3 bucket as the data source for emergency news video content. The system retrieves transcripts, archived reports, and related documents from the S3 bucket.

The RAG system uses state-of-the-art embedding models and a high-performing retrieval setup. However, users report slow responses and irrelevant results, which cause decreased user satisfaction. The company notices that vector searches are evaluating too many documents across too many content types and over long periods of time.

The company determines that the underlying models will not benefit from additional fine-tuning. The company must improve retrieval accuracy by applying smarter constraints and wants a solution that requires minimal changes to the existing architecture.

Which solution will meet these requirements?

- A. Enhance embeddings by using a domain-adapted model that is specifically trained on emergency news content for improved vector similarity.
- B. Migrate to Amazon OpenSearch Service
- C. Use vector fields and metadata filters to define the scope of results retrieval.
- D. Enable metadata-aware filtering within the Amazon Bedrock knowledge base by indexing S3 object metadata.
- E. Migrate to an Amazon Q Business index to perform structured metadata filtering and document categorization during retrieval.

Answer: C

NEW QUESTION 35

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 that have 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 chain of prompts
- E. Log model inputs and outputs to Amazon CloudWatch Log
- F. Use logs from Amazon 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 36

A financial services company is developing a generative AI (GenAI) application that serves both premium customers and standard customers. The application uses AWS Lambda functions behind an Amazon API Gateway REST API to process requests. The company needs to dynamically switch between AI models based on which customer tier each user belongs to. The company also wants to perform A/B testing for new features without redeploying code. The company needs to validate model parameters like temperature and maximum token limits before applying changes.

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

- A. Create AWS Systems Manager Parameter Store parameters for each configuration
- B. Use Lambda functions to poll for parameter update
- C. Use Amazon EventBridge events to trigger redeployments when configurations change.
- D. Store model configurations in Amazon DynamoDB table
- E. Optimize access patterns to retrieve configurations according to customer tier
- F. Configure Lambda functions to query DynamoDB at the beginning of each request to determine which model to use.
- G. Use AWS AppConfig to manage model configuration
- H. Use feature flags to perform A/B testing
- I. Define JSON schema validation rules for model parameter
- J. Configure Lambda functions to retrieve configurations by using the AWS AppConfig Agent.
- K. Create an Amazon ElastiCache (Redis OSS) cluster to store model configuration
- L. Set short TTL value
- M. Run custom validation logic in Lambda function
- N. Use Amazon CloudWatch metrics to monitor configuration usage.

Answer: C

NEW QUESTION 38

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 solutio
- D. Handle simple routing based on file extensions in application cod
- 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 typ
- 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 43

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 documen
- B. Create an Amazon Bedrock knowledge base with the documen
- C. Write a prompt that asks the FM to generate a response to each sample promp
- D. Use the RetrieveAndGenerate API to generate a report for each model.
- E. Combine the sample prompts into a single JSONL documen
- F. Store the document in an Amazon S3 bucke
- 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 outpu
- I. Run an evaluation job for each FM and select the FM as the generator.
- J. Combine the sample prompts into a single JSONL documen
- K. Store the document in an Amazon S3 bucke
- L. Create an Amazon Bedrock evaluation job that uses a judge mode
- M. Specify the S3 location as input and Amazon QuickSight as outpu
- N. Run an evaluation job for each FM and select the FM as the evaluator.
- O. Combine the sample prompts into a single JSON documen
- P. Create an Amazon Bedrock knowledge base from the documen
- Q. Create an Amazon Bedrock evaluation job that uses the retrieval and response generation evaluation typ
- R. Specify an Amazon S3 bucket as the outpu
- S. Run an evaluation job for each FM.

Answer: B

NEW QUESTION 45

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 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 configuration 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 47

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 expansio
- 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 querie
- 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 49

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 year
- 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 53

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

Answer: D

NEW QUESTION 56

A company uses AWS Lake Formation to set up a data lake that contains databases and tables for multiple business units across multiple AWS Regions. The company wants to use a foundation model (FM) through Amazon Bedrock to perform fraud detection. The FM must ingest sensitive financial data from the data lake. The data includes some customer personally identifiable information (PII).

The company must design an access control solution that prevents PII from appearing in a production environment. The FM must access only authorized data subsets that have PII redacted from specific data columns. The company must capture audit trails for all data access.

Which solution will meet these requirements?

- A. Create a separate dataset in a separate Amazon S3 bucket for each business unit and Region combination
- B. Configure S3 bucket policies to control access based on IAM roles that are assigned to FM training instance
- C. Use S3 access logs to track data access.
- D. Configure the FM to authenticate by using AWS Identity and Access Management roles and Lake Formation permissions based on LF-Tag expression
- E. Define business units and Regions as LF-Tags that are assigned to databases and tables
- F. Use AWS CloudTrail to collect comprehensive audit trails of data access.
- G. Use direct IAM principal grants on specific databases and tables in Lake Formation
- H. Create a custom application layer that logs access requests and further filters sensitive columns before sending data to the FM.
- I. Configure the FM to request temporary credentials from AWS Security Token Service
- J. Access the data by using presigned S3 URLs that are generated by an API that applies business unit and Regional filter
- K. Use AWS CloudTrail to collect comprehensive audit trails of data access.

Answer: B

NEW QUESTION 59

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 flow
- B. Configure Amazon CloudWatch metrics and alarms to monitor data quality

- C. Use a custom AWS Lambda function to pre-process the data
- 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 quality
- H. Load processed data into Amazon Bedrock.
- I. Use Amazon Comprehend to extract entities
- J. Create an AWS Lambda function to chunk text
- K. Run Amazon Athena to query and validate data quality
- 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 quality
- P. Load processed data into Amazon Bedrock.

Answer: B

NEW QUESTION 62

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 patterns
- B. Manually review outputs daily
- 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 datasets
- E. Configure AWS CodePipeline to run the evaluation jobs when prompt templates change
- 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 67

A financial services company is creating a Retrieval Augmented Generation (RAG) application that uses Amazon Bedrock to generate summaries of market activities. The application relies on a vector database that stores a small proprietary dataset with a low index count. The application must perform similarity searches. The Amazon Bedrock model's responses must maximize accuracy and maintain high performance. The company needs to configure the vector database and integrate it with the application. Which solution will meet these requirements?

- A. Launch an Amazon MemoryDB cluster and configure the index by using the Flat algorithm
- B. Configure a horizontal scaling policy based on performance metrics.
- C. Launch an Amazon MemoryDB cluster and configure the index by using the Hierarchical Navigable Small World (HNSW) algorithm
- D. Configure a vertical scaling policy based on performance metrics.
- E. Launch an Amazon Aurora PostgreSQL cluster and configure the index by using the Inverted File with Flat Compression (IVFFlat) algorithm
- F. Configure the instance class to scale to a larger size when the load increases.
- G. Launch an Amazon DocumentDB cluster that has an IVFFlat index and a high probe value
- H. Configure connections to the cluster as a replica set
- I. Distribute reads to replica instances.

Answer: B

NEW QUESTION 71

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 74

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 event
- 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 78

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

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
- 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 83

An insurance company uses existing Amazon SageMaker AI infrastructure to support a web-based application that allows customers to predict what their insurance premiums will be. The company stores customer data that is used to train the SageMaker AI model in an Amazon S3 bucket. The dataset is growing rapidly. The company wants a solution to continuously re-train the model. The solution must automatically re-train and re-deploy the model to the application when an employee uploads a new customer data file to the S3 bucket.

Which solution will meet these requirements?

- A. Use AWS Glue to run an ETL job on each uploaded file
- B. Configure the ETL job to use the AWS SDK to invoke the SageMaker AI model endpoint
- C. Use real-time inference with the endpoint to re-deploy the model after it is re-trained on the updated customer dataset.
- D. Create an AWS Lambda function and webhook handlers to generate an event when an employee uploads a new file
- E. Configure SageMaker Pipelines to re-deploy the model after it is re-trained on the updated customer dataset
- F. Use Amazon EventBridge to create an event bus
- G. Set the Lambda function event as the source and SageMaker Pipelines as the target.
- H. Create an AWS Step Functions Express workflow with AWS SDK integrations to retrieve the customer data from the S3 bucket when an employee uploads a new file to the S3 bucket

- I. Use a SageMaker Data Wrangler flow to export the data from the S3 bucket to SageMaker Autopilo
- J. Use the SageMaker Autopilot to re-deploy the model after it has been re-trained on the updated customer dataset.
- K. Create an AWS Step Functions Standard workflow
- L. Configure the first state to call an AWS Lambda function to respond when an employee uploads a new file to the S3 bucket
- M. Use a pipeline in SageMaker Pipelines to re-deploy the model after it has been re-trained on the updated customer dataset
- N. Use the next state in the workflow to run the pipeline when the first state receives a response.

Answer: D

NEW QUESTION 88

A company is using AWS Lambda and REST APIs to build a reasoning agent to automate support workflows. The system must preserve memory across interactions, share relevant agent state, and support event-driven invocation and synchronous invocation. The system must also enforce access control and session-based permissions.

Which combination of steps provides the MOST scalable solution? (Select TWO.)

- A. Use Amazon Bedrock AgentCore to manage memory and session-aware reasoning
- B. Deploy the agent with built-in identity support, event handling, and observability.
- C. Register the Lambda functions and REST APIs as actions by using Amazon API Gateway and Amazon EventBridge
- D. Enable Amazon Bedrock AgentCore to invoke the Lambda functions and REST APIs without custom orchestration code.
- E. Use Amazon Bedrock Agents for reasoning and conversation management
- F. Use AWS Step Functions and Amazon SQS for orchestration
- G. Store agent state in Amazon DynamoDB.
- H. Deploy the reasoning logic as a container on Amazon ECS behind API Gateway
- I. Use Amazon Aurora to store memory and identity data.
- J. Build a custom RAG pipeline by using Amazon Kendra and Amazon Bedrock
- K. Use AWS Lambda to orchestrate tool invocation
- L. Store agent state in Amazon S3.

Answer: AB

NEW QUESTION 91

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 95

An elevator service company has developed an AI assistant application by using Amazon Bedrock. The application generates elevator maintenance recommendations to support the company's elevator technicians. The company uses Amazon Kinesis Data Streams to collect the elevator sensor data. New regulatory rules require that a human technician must review all AI-generated recommendations. The company needs to establish human oversight workflows to review and approve AI recommendations. The company must store all human technician review decisions for audit purposes.

Which solution will meet these requirements?

- A. Create a custom approval workflow by using AWS Lambda functions and Amazon SQS queues for human review of AI recommendations
- B. Store all review decisions in Amazon DynamoDB for audit purposes.
- C. Create an AWS Step Functions workflow that has a human approval step that uses the waitForResource API to pause execution
- D. After a human technician completes a review, use an AWS Lambda function to call the SendTaskSuccess API with the approval decision
- E. Store all review decisions in Amazon DynamoDB.
- F. Create an AWS Glue workflow that has a human approval step
- G. After the human technician review, integrate the application with an AWS Lambda function that calls the SendTaskSuccess API
- H. Store all human technician review decisions in Amazon DynamoDB.
- I. Configure Amazon EventBridge rules with custom event patterns to route AI recommendations to human technicians for review
- J. Create AWS Glue jobs to process human technician approval queue
- K. Use Amazon ElastiCache to cache all human technician review decisions.

Answer: B

NEW QUESTION 98

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 available
- 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 103

A company is developing a customer communication platform that uses an AI assistant powered by an Amazon Bedrock foundation model (FM). The AI assistant summarizes customer messages and generates initial response drafts.

The company wants to use Amazon Comprehend to implement layered content filtering. The layered content filtering must prevent sharing of offensive content, protect customer privacy, and detect potential inappropriate advice solicitation. Inappropriate advice solicitation includes requests for unethical practices, harmful activities, or manipulative behaviors.

The solution must maintain acceptable overall response times, so all pre-processing filters must finish before the content reaches the FM.

Which solution will meet these requirements?

- A. Use parallel processing with asynchronous API call
- B. Use toxicity detection for offensive content
- C. Use prompt safety classification for inappropriate advice solicitation
- D. Use personally identifiable information (PII) detection without redaction.
- E. Use custom classification to build an FM that detects offensive content and inappropriate advice solicitation
- F. Apply personally identifiable information (PII) detection as a secondary filter only when messages pass the custom classifier.
- G. Deploy a multi-stage process
- H. Configure the process to use prompt safety classification first, then toxicity detection on safe prompts only, and finally personally identifiable information (PII) detection in streaming mode
- I. Route flagged messages through Amazon EventBridge for human review.
- J. Use toxicity detection with thresholds configured to 0.5 for all categories
- K. Use parallel processing for both prompt safety classification and personally identifiable information (PII) detection with entity redaction
- L. Apply Amazon CloudWatch alarms to filter metrics.

Answer: D

NEW QUESTION 105

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your AIP-C01 Exam with Our Prep Materials Via below:

<https://www.certleader.com/AIP-C01-dumps.html>