Amazon-Web-Services

Exam Questions AIP-C01

AWS Certified Generative AI Developer - Professional

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

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 bucke
I. Use AWS Lambda to process logs in real tim
J. Manually update CloudWatch alarm thresholds based on trends identified by the Lambda function.

Answer: C

NEW QUESTION 2
A company is building a serverless application that uses AWS Lambda functions to help students around the world summarize notes. The application uses Anthropic Claude through Amazon Bedrock. The company observes that most of the traffic occurs during evenings in each time zone. Users report experiencing throttling errors during peak usage times in their time zones.
The company needs to resolve the throttling issues by ensuring continuous operation of the application. The solution must maintain application performance quality and must not require a fixed hourly cost during low traffic periods.
Which solution will meet these requirements?

A. Create custom Amazon CloudWatch metrics to monitor model error
B. Set provisioned throughput to a value that is safely higher than the peak traffic observed.
C. Create custom Amazon CloudWatch metrics to monitor model error
D. Set up a failover mechanism to redirect invocations to a backup AWS Region when the errors exceed a specified threshold.
E. Enable invocation logging in Amazon Bedroc
F. Monitor key metrics such as Invocations, InputTokenCount, OutputTokenCount, and InvocationThrottle
G. Distribute traffic across cross-Region inference endpoints.
H. Enable invocation logging in Amazon Bedroc
I. Monitor InvocationLatency, InvocationClientErrors, and InvocationServerErrors metric
J. Distribute traffic across multiple versions of the same model.

Answer: C

NEW QUESTION 3
A company has a generative AI (GenAI) application that uses Amazon Bedrock to provide real-time responses to customer queries. The company has noticed intermittent failures with API calls to foundation models (FMs) during peak traffic periods.
The company needs a solution to handle transient errors and provide detailed observability into FM performance. The solution must prevent cascading failures during throttling events and provide distributed tracing across service boundaries to identify latency contributors. The solution must also enable correlation of performance issues with specific FM characteristics.
Which solution will meet these requirements?

A. Implement a custom retry mechanism with a fixed delay of 1 second between retrie
B. Configure Amazon CloudWatch alarms to monitor the application??s error rates and latency metrics.
C. Configure the AWS SDK with standard retry mode and exponential backoff with jitte
D. Use AWS X-Ray tracing with annotations to identify and filter service components.
E. Implement client-side caching of all FM response
F. Add custom logging statements in the application code to record API call durations.
G. Configure the AWS SDK with adaptive retry mod
H. Use AWS CloudTrail distributed tracing to monitor throttling events.

Answer: B

NEW QUESTION 4
A retail company is using Amazon Bedrock to develop a customer service AI assistant. Analysis shows that 70% of customer inquiries are simple product questions that a smaller model can effectively handle. However, 30% of inquiries are complex return policy questions that require advanced reasoning.
The company wants to implement a cost-effective model selection framework to automatically route customer inquiries to appropriate models based on inquiry complexity. The framework must maintain high customer satisfaction and minimize response latency.
Which solution will meet these requirements with the LEAST implementation effort?

A. Create a multi-stage architecture that uses a small foundation model (FM) to classify the complexity of each inquir
B. Route simple inquiries to a smaller, more cost-effective mode
C. Route complex inquiries to a larger, more capable mode
D. Use AWS Lambda functions to handle routing logic.
E. Use Amazon Bedrock intelligent prompt routing to automatically analyze inquirie
F. Route simple product inquiries to smaller models and route complex return policy inquiries to more capable larger models.
G. Implement a single-model solution that uses an Amazon Bedrock mid-sized foundation model (FM) with on-demand pricin
H. Include special instructions in model prompts to handle both simple and complex inquiries by using the same model.
I. Create separate Amazon Bedrock endpoints for simple and complex inquirie
J. Implement a rule-based routing system based on keyword detectio

K. Use on-demand pricing for the smaller model and provisioned throughput for the larger model.

Answer: B

NEW QUESTION 5
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 6
A GenAI developer is building a Retrieval Augmented Generation (RAG)-based customer support application that uses Amazon Bedrock foundation models (FMs). The application needs to process 50 GB of historical customer conversations that are stored in an Amazon S3 bucket as JSON files. The application must use the processed data as its retrieval corpus. The application??s data processing workflow must extract relevant data from customer support documents, remove customer personally identifiable information (PII), and generate embeddings for vector storage. The processing workflow must be cost- effective and must finish within 4 hours.
Which solution will meet these requirements with the LEAST operational overhead?

A. Use AWS Lambda and Amazon Comprehend to process files in parallel, remove PII, and call Amazon Bedrock APIs to generate vector
B. Configure Lambda concurrency limits and memory settings to optimize throughput.
C. Create an AWS Glue ETL job to run PII detection scripts on the dat
D. Use Amazon SageMaker Processing to run the HuggingFaceProcessor to generate embeddings by using a pre-trained mode
E. Store the embeddings in Amazon OpenSearch Service.
F. Deploy an Amazon EMR cluster that runs Apache Spark with user-defined functions (UDFs) that call Amazon Comprehend to detect PI
G. Use Amazon Bedrock APIs to generate vector
H. Store outputs in Amazon Aurora PostgreSQL with the pgvector extension.
I. Implement a data processing pipeline that uses AWS Step Functions to orchestrate a workload that uses Amazon Comprehend to detect PII and Amazon Bedrock to generate embedding
J. Directly integrate the workflow with Amazon OpenSearch Serverless to store vectors and provide similarity search capabilities.

Answer: D

NEW QUESTION 7
A company has deployed an AI assistant as a React application that uses AWS Amplify, an AWS AppSync GraphQL API, and Amazon Bedrock Knowledge Bases. The application uses the GraphQL API to call the Amazon Bedrock RetrieveAndGenerate API for knowledge base interactions. The company configures an AWS Lambda resolver to use the RequestResponse invocation type.
Application users report frequent timeouts and slow response times. Users report these problems more frequently for complex questions that require longer processing.
The company needs a solution to fix these performance issues and enhance the user experience.
Which solution will meet these requirements?

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

Answer: A

NEW QUESTION 8
A company is building a generative AI (GenAI) application that produces content based on a variety of internal and external data sources. The company wants to ensure that the generated output is fully traceable. The application must support data source registration and enable metadata tagging to attribute content to its original source. The application must also maintain audit logs of data access and usage throughout the pipeline.
Which solution will meet these requirements?

A. Use AWS Lake Formation to catalog data sources and control acces
B. Apply metadata tags directly in Amazon S3. Use AWS CloudTrail to monitor API activity.

C. Use AWS Glue Data Catalog to register and tag data source
D. Use Amazon CloudWatch Logs to monitor access patterns and application behavior.
E. Store data in Amazon S3 and use object tagging for attributio
F. Use AWS Glue Data Catalog to manage schema informatio
G. Use AWS CloudTrail to log access to S3 buckets.
H. Use AWS Glue Data Catalog to register all data source
I. Apply metadata tags to attribute data source
J. Use AWS CloudTrail to log access and activity across services.

Answer: D

NEW QUESTION 9
A company has a recommendation system. The system's applications run on Amazon EC2 instances. The applications make API calls to Amazon Bedrock foundation models (FMs) to analyze customer behavior and generate personalized product recommendations.
The system is experiencing intermittent issues. Some recommendations do not match customer preferences. The company needs an observability solution to monitor operational metrics and detect patterns of operational performance degradation compared to established baselines. The solution must also generate alerts with correlation data within 10 minutes when FM behavior deviates from expected patterns.
Which solution will meet these requirements?

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

Answer: C

NEW QUESTION 10
A healthcare company is using Amazon Bedrock to build a system to help practitioners make clinical decisions. The system must provide treatment recommendations to physicians based only on approved medical documentation and must cite specific sources. The system must not hallucinate or produce factually incorrect information.
Which solution will meet these requirements with the LEAST operational overhead?

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

Answer: B

NEW QUESTION 10
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 dat
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 custome
D. Retrieve data in real time during prompt generation.
E. Ensure that each customer configures an Amazon Bedrock knowledge bas
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 11

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

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

Answer: B

NEW QUESTION 16
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 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 mode

B. Configure metrics for context relevance and context coverag
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 datase
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 combinatio
G. Use Amazon Bedrock built-in metrics for correctness and completenes
H. Manually review scores before deployment approval.
I. Set up a pipeline that uses multiple retrieve-only evaluation jobs to assess retrieval qualit
J. Create separate evaluation jobs for both FMs that use Amazon Nova Pro as the LLM-as-a-judge mode
K. Evaluate based on faithfulness and citation precision metrics.

Answer: B

NEW QUESTION 23
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 26
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 Servic
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 29
A financial technology company is using Amazon Bedrock to build an assessment system for the company??s customer service AI assistant. The AI assistant must provide financial recommendations that are factually accurate, compliant with financial regulations, and conversationally appropriate. The company needs to combine automated quality evaluations at scale with targeted human reviews of critical interactions.
What solution will meet these requirements?

A. Configure a pipeline in which financial experts manually score all responses for accuracy, compliance, and conversational qualit
B. Use Amazon SageMaker notebooks to analyze results to identify improvement areas.
C. Configure Amazon Bedrock evaluations that use Anthropic Claude Sonnet as a judge model to assess response accuracy and appropriatenes
D. Configure custom Amazon Bedrock guardrails to check responses for compliance with financial policie
E. Add Amazon Augmented AI (Amazon A2I) human reviews for flagged critical interactions.
F. Create an Amazon Lex bot to manage customer service interaction
G. Configure AWSLambda functions to check responses against a static compliance databas
H. Configure intents that call the Lambda function
I. Add an additional intent to collect end-user reviews.
J. Configure Amazon CloudWatch to monitor response patterns from the AI assistan
K. Configure CloudWatch alerts for potential compliance violation
L. Establish a team of human evaluators to review flagged interactions.

Answer: B

NEW QUESTION 31
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 configuratio
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 tie
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 testin
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 33
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 34
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 39
A company is planning to deploy multiple generative AI (GenAI) applications to five independent business units that operate in multiple countries in Europe and the Americas.
Each application uses Amazon Bedrock Retrieval Augmented Generation (RAG) patterns with business unit-specific knowledge bases that store terabytes of unstructured data.
The company must establish well-architected, standardized components for security controls, observability practices, and deployment patterns across all the GenAI applications. The components must be reusable, versioned, and governed consistently.
Which solution will meet these requirements?

A. Configure Amazon API Gateway REST API endpoints for the GenAI application
B. Deploy common security, observability, and RAG patterns based on the AWS Well- Architected Generative AI Lens in standardized AWS CloudFormation template
C. Use CloudFormation Guard after deployment to validate policy compliance in each business unit.
D. Create standardized AWS CloudFormation templates to implement security, observability, and RAG patterns based on the AWS Well-Architected Generative AI Len
E. Establish a centralized repository for version contro
F. Integrate a CI/CD pipeline with CloudFormation Guard to enforce consistent and repeatable deployments across business units.
G. Use AWS Service Catalog to define standardized portfolios and versioned products for each business uni

H. Use the portfolios to enforce security, observability, and RAG patterns based on the AWS Well-Architected Generative AI Len
I. Require business units to use the Service Catalog console to deploy resources.
J. Document security controls, observability requirements, and RAG patterns based on the AWS Well-Architected Generative AI Lens in a shared design documen
K. Use Amazon Macie to enforce deploymen
L. Delegate implementation responsibility to each business unit.

Answer: B

NEW QUESTION 42
A finance company is developing an AI assistant to help clients plan investments and manage their portfolios. The company identifies several high-risk conversation patterns such as requests for specific stock recommendations or guaranteed returns. High-risk conversation patterns could lead to regulatory violations if the company cannot implement appropriate controls.
The company must ensure that the AI assistant does not provide inappropriate financial advice, generate content about competitors, or make claims that are not factually grounded in the company's approved financial guidance. The company wants to use Amazon Bedrock Guardrails to implement a solution.
Which combination of steps will meet these requirements? (Select THREE)

A. Add the high-risk conversation patterns to a denied topics guardrail.
B. Configure a content filter guardrail to filter prompts that contain the high-risk conversation patterns.
C. Configure a content filter guardrail to filter prompts that contain competitor names.
D. Add the names of competitors as custom word filter
E. Set the input and output actions to block.
F. Set a low grounding score threshold.
G. Set a high grounding score threshold.

Answer: ADF

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 capabilitie
M. Add deep research and reasoning capabilities to the agent to reduce ambiguity in the medical terminology.

Answer: B

NEW QUESTION 51
A financial services company is developing a real-time generative AI (GenAI) assistant to support human call center agents. The GenAI assistant must transcribe live customer speech, analyze context, and provide incremental suggestions to call center agents while a customer is still speaking. To preserve responsiveness, the GenAI assistant must maintain end-to-end latency under 1 second from speech to initial response display. The architecture must use only managed AWS services and must support bidirectional streaming to ensure that call center agents receive updates in real time.
Which solution will meet these requirements?

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

Answer: B

NEW QUESTION 52
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 inpu
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 layou
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 dat
I. Create a second application that is optimized for unstructured dat
J. Configure each application to use a tailored chunking strategy that is suited to the application's content typ
K. Implement logic to link queries to the appropriate sources.

Answer: B

NEW QUESTION 56
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 categorie
B. Set up denied topics for investment advice and include sample phrases to bloc
C. Set up sensitive information filters that apply the block action for all PII entitie
D. Apply the guardrail to all model inference calls.
E. Configure multiple guardrails by using tiered policie
F. Create one guardrail and set content filters to hig
G. Configure the guardrail to block PII for public interaction
H. Configure a second guardrail and set content filters to mediu
I. Configure the second guardrail to mask PII for internal us
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 conten
L. Set updenied topics for investment advice and include clear definitions and sample phrases to bloc
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 cas
P. Create one guardrail that applies a harmful content filte
Q. Create a guardrail to apply topic filters for investment advic
R. Create a guardrail to apply sensitive information filters to block PI
S. Use AWS Step Functions to chain the guardrails sequentially.

Answer: C

NEW QUESTION 61
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 latenc
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 chainin
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 ElastiCach
H. Use AWS Lambda functions to pre- process metadata and to trace end-to-end latenc
I. Use AWS X-Ray to identify and remediate performance bottlenecks.
J. Use Amazon Kendra to improve roast log retrieval accurac
K. Store normalized prompt metadata within Amazon DynamoD
L. Use AWS Step Functions to orchestrate multi-step prompts.

Answer: A

NEW QUESTION 66
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 rol
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 mode
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 polic
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 polic
K. Use stack sets to deploy the guardrail to each account in the organization.

Answer: CD

NEW QUESTION 69
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 serve
B. Grant the AI agent Lambda functions permission to invoke the Lambda function that hosts the MCP serve
C. Configure the AI agent??s MCP client to invoke the MCP server asynchronously.
D. Use a Lambda function to host the MCP serve
E. Grant the AI agent Lambda functions permission to invoke the Lambda function that hosts the MCP serve
F. Configure the AI agent to use the STDIO transport with the MCP server.
G. Use a Lambda function to host the MCP serve
H. Create an Amazon API Gateway HTTP API that proxies requests to the Lambda functio
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 comman
O. Specify the user credentials as environment variables.

Answer: C

NEW QUESTION 72
A pharmaceutical company is developing a Retrieval Augmented Generation (RAG) application that uses an Amazon Bedrock knowledge base. The knowledge base uses Amazon OpenSearch Service as a data source for more than 25 million scientific papers. Users report that the application produces inconsistent answers that cite irrelevant sections of papers when queries span methodology, results, and discussion sections of the papers.
The company needs to improve the knowledge base to preserve semantic context across related paragraphs on the scale of the entire corpus of data.
Which solution will meet these requirements?

A. Configure the knowledge base to use fixed-size 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 75
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 folde
B. In each file, create access control entries that specify the IAM Identity Center group that should have access to that department's dat
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 bucke
E. Add access control entries that map each department's S3 prefix to its corresponding IAM Identity Center grou
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 bucke
H. Add a StringNotEquals condition key to the permissions set for each group that specifies the department each group is associated wit
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 bucke
K. Add anAccessControlList object to the file that specifies the S3 path of each department's prefi
L. Specify the IAM Identity Center group that should have access to each department's prefi
M. Reference the file location in the data source metadata settings.

Answer: B

NEW QUESTION 80
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 83
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 84
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 Bedroc
H. Use Amazon EventBridge rules to notify teams when thresholds are exceede
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 versio
K. Use Amazon Comprehend sentiment analysis to compare result
L. Block deployment if sentiment scores decrease.

Answer: B

NEW QUESTION 86
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 algorith
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) algorith
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) algorith
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 valu
H. Configure connections to the cluster as a replica se
I. Distribute reads to replica instances.

Answer: B

NEW QUESTION 87
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 fil
B. Configure the ETL job to use the AWS SDK to invoke the SageMaker AI model endpoin
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 fil
E. Configure SageMaker Pipelines to re-deploy the model after it is re-trained on the updated customer datase
F. Use Amazon EventBridge to create an event bu
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 bucke
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 workflo
L. Configure the first state to call an AWS Lambda function to respond when an employee uploads a new file to the S3 bucke
M. Use a pipeline in SageMaker Pipelines to re-deploy the model after it has been re-trained on the updated customer datase
N. Use the next state in the workflow to run the pipeline when the first state receives a response.

Answer: D

NEW QUESTION 90
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 reasonin
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 EventBridg
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 managemen
F. Use AWS Step Functions and Amazon SQS for orchestratio
G. Store agent state in Amazon DynamoDB.
H. Deploy the reasoning logic as a container on Amazon ECS behind API Gatewa
I. Use Amazon Aurora to store memory and identity data.
J. Build a custom RAG pipeline by using Amazon Kendra and Amazon Bedroc
K. Use AWS Lambda to orchestrate tool invocation
L. Store agent state in Amazon S3.

Answer: AB

NEW QUESTION 91
A medical company uses Amazon Bedrock to power a clinical documentation summarization system. The system produces inconsistent summaries when handling complex clinical documents. The system performed well on simple clinical documents.
The company needs a solution that diagnoses inconsistencies, compares prompt performance against established metrics, and maintains historical records of prompt versions.
Which solution will meet these requirements?

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

Answer: B

NEW QUESTION 92
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 Regio
B. Configure the application to retry failed requests with exponential backoff.
C. Implement token batching to reduce API overhea

D. Use cross-Region inference profiles to automatically distribute traffic across available Regions.
E. Set up auto scaling AWS Lambda functions in each Regio
F. Implement client-side round-robin request distributio
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 Region
I. Use Amazon SQS to set up an asynchronous retrieval process.

Answer: B

NEW QUESTION 94
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 recommendation
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 waitForTaskToken API to pause executio
D. After a human technician completes a review, use an AWS Lambda function to call the SendTaskSuccess API with the approval decisio
E. Store all review decisions in Amazon DynamoDB.
F. Create an AWS Glue workflow that has a human approval ste
G. After the human technician review, integrate the application with an AWS Lambda function that calls the SendTaskSuccess AP
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 revie
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 97
A company provides a service that helps users from around the world discover new restaurants. The service has 50 million monthly active users. The company wants to implement a semantic search solution across a database that contains 20 million restaurants and 200 million reviews. The company currently stores the data in a PostgreSQL database.
The solution must support complex natural language queries and return results for at least 95% of queries within 500 ms. The solution must maintain data freshness for restaurant details that update hourly. The solution must also scale cost-effectively during peak usage periods.
Which solution will meet these requirements with the LEAST development effort?

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

Answer: D

NEW QUESTION 101
A company is developing a generative AI (GenAI) application by using Amazon Bedrock. The application will analyze patterns and relationships in the company??s data. The application will process millions of new data points daily across AWS Regions in Europe, North America, and Asia before storing the data in Amazon S3. The application must comply with local data protection and storage regulations. Data residency and processing must occur within the same continent. The application must also maintain audit trails of the application??s decision-making processes and provide data classification capabilities.
Which solution will meet these requirements?

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

Answer: C

NEW QUESTION 102
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 conten
C. Use prompt safety classification for inappropriate advice solicitatio
D. Use personally identifiable information (PII) detection without redaction.
E. Use custom classification to build an FM that detects offensive content and inappropriate advice solicitatio
F. Apply personally identifiable information (PII) detection as a secondary filter only when messages pass the custom classifier.
G. Deploy a multi-stage proces
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 mod
I. Route flagged messages through Amazon EventBridge for human review.
J. Use toxicity detection with thresholds configured to 0.5 for all categorie
K. Use parallel processing for both prompt safety classification and personally identifiable information (PII) detection with entity redactio
L. Apply Amazon CloudWatch alarms to filter metrics.

Answer: D

NEW QUESTION 104
A company is designing a solution that uses foundation models (FMs) to support multiple AI workloads. Some FMs must be invoked on demand and in real time. Other FMs require consistent high-throughput access for batch processing.
The solution must support hybrid deployment patterns and run workloads across cloud infrastructure and on-premises infrastructure to comply with data residency and compliance requirements.
Which combination of steps will meet these requirements? (Select TWO.)

A. Use AWS Lambda to orchestrate low-latency FM inference by invoking FMs hosted on Amazon SageMaker AI asynchronous endpoints.
B. Configure provisioned throughput in Amazon Bedrock to ensure consistent performance for high-volume workloads.
C. Deploy FMs to Amazon SageMaker AI endpoints with support for edge deployment by using Amazon SageMaker Ne
D. Orchestrate the FMs by using AWS Lambda to support hybrid deployment.
E. Use Amazon Bedrock with auto-scaling to handle unpredictable traffic surges.
F. Use Amazon SageMaker JumpStart to host and invoke the FMs.

Answer: BC

NEW QUESTION 105
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 110
.....

Relate Links

100% Pass Your AIP-C01 Exam with Exambible Prep Materials

https://www.exambible.com/AIP-C01-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/