Amazon-Web-Services

Exam Questions AIP-C01

AWS Certified Generative AI Developer - Professional

NEW QUESTION 1
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 2
A company upgraded its Amazon Bedrock–powered foundation model (FM) that supports a multilingual customer service assistant. After the upgrade, the assistant exhibited inconsistent behavior across languages. The assistant began generating different responses in some languages when presented with identical questions.
The company needs a solution to detect and address similar problems for future updates. The evaluation must be completed within 45 minutes for all supported languages. The evaluation must process at least 15,000 test conversations in parallel. The evaluation process must be fully automated and integrated into the CI/CD pipeline. The solution must block deployment if quality thresholds are not met.
Which solution will meet these requirements?

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

Answer: D

NEW QUESTION 3
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 4
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 5
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 6
A company is building a video analysis platform on AWS. The platform will analyze a large video archive by using Amazon Rekognition and Amazon Bedrock. The platform must comply with predefined privacy standards. The platform must also use secure model I/O, control foundation model (FM) access patterns, and provide an audit of who accessed what and when.
Which solution will meet these requirements?

A. Configure VPC endpoints for Amazon Bedrock model API call
B. Implement Amazon Bedrock guardrails to filter harmful or unauthorized content in prompts and response
C. Use Amazon Bedrock trace events to track all agent and model invocations for auditing purpose
D. Export the traces to Amazon CloudWatch Logs as an audit record of model usag
E. Store all prompts and outputs in Amazon S3 with server-side encryption with AWS KMS keys (SSE-KMS).
F. Define access control by using IAM with attribute-based access control (ABAC) to map departments to specific permission
G. Configure VPC endpoints for Amazon Bedrock model API call
H. Use IAM condition keys to enforce specific GuardrailIdentifier and ModelId value
I. Configure AWS CloudTrail to capture management and data events for S3 objects and KMS key usage activitie
J. Enable S3 server access logging to record detailed file-level interactions with the video archive
K. Send all CloudTrail logs to AWS CloudTrail Lak
L. Set up Amazon CloudWatch alarms to detect and alert on unexpected activity from Amazon Bedrock, Amazon Rekognition, and AWS KMS.
M. Restrict access to services by using VPC endpoint policie
N. Use AWS Config to track resource changes and compliance with security rule
O. Use server-side encryption with AWS KMS keys (SSE-KMS) to encrypt data at res
P. Store the model??s I/O in separate Amazon S3 bucket
Q. Enable S3 server access logging to track file-level interactions.
R. Configure AWS CloudTrail Insights to analyze API call patterns across accounts and detect anomalous activity in Amazon Bedrock, Amazon Rekognition, Amazon S3, and AWS KM
S. Deploy Amazon Macie to scan and classify the video archiv
T. Use server-side encryption with AWS KMS keys (SSE-KMS) to encrypt all stored dat
. Configure CloudTrail to capture KMS API usage events for audit purpose
. Configure Amazon EventBridge rules to process CloudTrail Insights anomalies and Macie finding
. Use CloudWatch alarms to trigger automated notifications and security responses when potential security issues are detected.

Answer: B

NEW QUESTION 7
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 8
An ecommerce company is developing a generative AI (GenAI) solution that uses Amazon Bedrock with Anthropic Claude to recommend products to customers. Customers report that some recommended products are not available for sale or are not relevant. Customers also report long response times for some recommendations.
The company confirms that most customer interactions are unique and that the solution recommends products not present in the product catalog.
Which solution will meet this requirement?

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

Answer: C

NEW QUESTION 9
A company provides a service that helps users from around the world discover new restaurants. The service has 50 million monthly active users. The company wants to implement a semantic search solution across a database that contains 20 million restaurants and 200 million reviews. The company currently stores the data in PostgreSQL.
The solution must support complex natural language queries and return results for at least 95% of queries within 500 ms. The solution must maintain data freshness for restaurant details that update hourly. The solution must also scale cost-effectively during peak usage periods.
Which solution will meet these requirements with the LEAST development effort?

A. Migrate the restaurant data to Amazon OpenSearch Servic
B. Implement keyword-based search rules that use custom analyzers and relevance tuning to find restaurants based on attributes such as cuisine type, features, and 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 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: B

NEW QUESTION 10
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 10
A company is using Amazon Bedrock to develop an AI-powered application that uses a foundation model that supports cross-Region inference and provisioned throughput. The application must serve users in Europe and North America with consistently low latency. The application must comply with data residency regulations that require European user data to remain within Europe-based AWS Regions.
During testing, the application experiences service degradation when Regional traffic spikes reach service quotas. The company needs a solution that maintains application resilience and minimizes operational complexity.
Which solution will meet these requirements?

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

Answer: B

NEW QUESTION 14
An ecommerce company is building an internal platform to develop generative AI applications by using Amazon Bedrock foundation models (FMs). Developers need to select models based on evaluations that are aligned to ecommerce use cases. The platform must display accuracy metrics for text generation and summarization in dashboards. The company has custom ecommerce datasets to use as standardized evaluation inputs.
Which combination of steps will meet these requirements with the LEAST operational overhead? (Select TWO.)

A. Import the datasets to an Amazon S3 bucke
B. Provide appropriate IAM permissions and cross-origin resource sharing (CORS) permissions to give the evaluation jobs access to the datasets.
C. Import the datasets to an Amazon S3 bucke
D. Provide appropriate IAM permissions and a VPC endpoint configuration to give the evaluation jobs access to the datasets.
E. Configure an AWS Lambda function to create model evaluation jobs on a schedule in the Amazon Bedrock consol
F. Provide the URI of the S3 bucket that contains the datasets as an inpu
G. Configure the evaluation jobs to measure the real world knowledge (RWK) score for text generation and BERTScore for summarizatio
H. Configure a second Lambda function to check the status of the jobs and publish custom logs to Amazon CloudWatc
I. Create a custom Amazon CloudWatch Logs Insights dashboard.
J. Use Amazon SageMaker Clarify on a schedule to create model evaluation job
K. Useopen source frameworks to create and run standardized evaluation
L. Publish results to Amazon CloudWatch namespace
M. Use an AWS Lambda function to check the status of the jobs and publish custom logs to Amazon CloudWatc
N. Create a custom Amazon CloudWatch Logs Insights dashboard.
O. Run an Amazon SageMaker AI notebook job on a schedule by using the fmvelos or ragas framework to run evaluations that use the datasets in the S3 bucke
P. Write Python code in the notebook that makes direct InvokeModel API calls to the FMs and processes their responses for evaluatio
Q. Publish job status and results to Amazon CloudWatch Logs to measure the real world knowledge (RWK) score for text generation and toxicity for summarization as metrics for accurac
R. Create a custom CloudWatch Logs Insights dashboard.

Answer: BC

NEW QUESTION 18
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 20
A company is creating a generative AI (GenAI) application that uses Amazon Bedrock foundation models (FMs). The application must use Microsoft Entra ID to authenticate. All FM API calls must stay on private network paths. Access to the application must be limited by department to specific model families. The company also needs a comprehensive audit trail of model interactions.
Which solution will meet these requirements?

A. Configure SAML federation between Microsoft Entra ID and AWS Identity and Access Managemen
B. Create department-specific IAM roles that allow only the required ModelId value
C. Create AWS PrivateLink interface VPC endpoints for Amazon Bedrock runtime service
D. Enable AWS CloudTrail to capture Amazon Bedrock API call
E. Configure Amazon Bedrock model invocation logging to record detailed model interactions.
F. Create an identity provider (IdP) connection in IAM to authenticate by using Microsoft Entra I
G. Assign department permission sets to control access to specific model familie
H. Deploy AWS Lambda functions in private subnets with a NAT gateway for egress to Amazon Bedrock public endpoint

I. Enable CloudWatch Logs to capture model interactions for auditing purposes.
J. Create a SAML identity provider (IdP) in IAM to authenticate by using Microsoft Entra I
K. Use IAM permissions boundaries to limit department roles' access to specific model familie
L. Configure public Amazon Bedrock API endpoints with VPC routing to maintain private network connectivit
M. Set up CloudTrail with Amazon S3 Lifecycle rules to manage audit logs of model interactions.
N. Configure OpenID Connect (OIDC) federation between Microsoft Entra ID and IA
O. Use attribute-based access control to map department attributes to specific model access permission
P. Apply SCP policies to restrict access to Amazon Bedrock FM families based on departmen
Q. Use Microsoft Entra ID's built-in logging capabilities to maintain an audit trail of model interactions.

Answer: A

NEW QUESTION 23
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 25
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 26
A legal research company has a Retrieval Augmented Generation (RAG) application that uses Amazon Bedrock and Amazon OpenSearch Service. The application stores 768- dimensional vector embeddings for 15 million legal documents, including statutes, court rulings, and case summaries.
The company's current chunking strategy segments text into fixed-length blocks of 500 tokens. The current chunking strategy often splits contextually linked information such as legal arguments, court opinions, or statute references across separate chunks. Researchers report that generated outputs frequently omit key context or cite outdated legal information.
Recent application logs show a 40% increase in response times. The p95 latency metric exceeds 2 seconds. The company expects storage needs for the application to grow from 90 GB to 360 GB within a year.
The company needs a solution to improve retrieval relevance and system performance at scale.
Which solution will meet these requirements?

A. Increase the embedding vector dimensionality from 768 to 4,096 without changing the existing chunking or pre-processing strategy.
B. Replace dynamic retrieval with static, pre-written summaries that are stored in Amazon S3. Use Amazon CloudFront to serve the summaries to reduce compute demand and improve predictability.
C. Update the chunking strategy to use semantic boundaries such as complete legal arguments, clauses, or sections rather than fixed token limit
D. Regenerate vector embeddings to align with the new chunk structure.
E. Migrate from OpenSearch Service to Amazon DynamoD
F. Implement keyword-based indexes to enable faster lookups for legal concepts.

Answer: C

NEW QUESTION 27
A company is developing a customer support application that uses Amazon Bedrock foundation models (FMs) to provide real-time AI assistance to the company??s employees. The application must display AI-generated responses character by character as the responses are generated. The application needs to support thousands of concurrent users with minimal latency. The responses typically take 15 to 45 seconds to finish.
Which solution will meet these requirements?

A. Configure an Amazon API Gateway WebSocket API with an AWS Lambda integratio
B. Configure the WebSocket API to invoke the Amazon Bedrock InvokeModelWithResponseStream API and stream partial responses through WebSocket

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

Answer: A

NEW QUESTION 30
A financial services company is developing a customer service AI assistant application that uses a foundation model (FM) in Amazon Bedrock. The application must provide transparent responses by documenting reasoning and by citing sources that are used for Retrieval Augmented Generation (RAG). The application must capture comprehensive audit trails for all responses to users. The application must be able to serve up to 10,000 concurrent users and must respond to each customer inquiry within 2 seconds.
Which solution will meet these requirements with the LEAST operational overhead?

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

Answer: A

NEW QUESTION 34
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 endpoin
E. Integrate the REST API with the business application.
F. Configure the FM to use cross-Region inference through a Europe-scoped endpoin
G. Configure an Amazon Bedrock VPC endpoin
H. Extend existing SCPs to allow employees to use the FM through inference profiles in Europe-based Regions where the FM is availabl
I. Use an inference profile to integrate Amazon Bedrock with the business application.
J. Deploy the FM in Amazon SageMaker in eu-north-1. Configure a SageMaker VPC endpoin
K. Extend existing SCPs to allow employees to use the SageMaker endpoin
L. Integrate the FM in SageMaker with the business application.

Answer: C

NEW QUESTION 37
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 39
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

AIP-C01 Practice Exam Features:

* AIP-C01 Questions and Answers Updated Frequently

* AIP-C01 Practice Questions Verified by Expert Senior Certified Staff

* AIP-C01 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* AIP-C01 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year