Exam Questions AIP-C01

AWS Certified Generative AI Developer - Professional

https://www.2passeasy.com/dumps/AIP-C01/

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
An ecommerce company is developing a generative AI application that uses Amazon Bedrock with Anthropic Claude to recommend products to customers. Customers report that some recommended products are not available for sale on the website or are not relevant to the customer. Customers also report that the solution takes a long time to generate some recommendations.
The company investigates the issues and finds that most interactions between customers and the product recommendation solution are unique. The company confirms that the solution recommends products that are not in the company??s product catalog. The company must resolve these issues.
Which solution will meet this requirement?

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

Answer: C

NEW QUESTION 3
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 4
A company deploys multiple Amazon Bedrock–based generative AI (GenAI) applications across multiple business units for customer service, content generation, and document analysis. Some applications show unpredictable token consumption patterns. The company requires a comprehensive observability solution that provides real-time visibility into token usage patterns across multiple models. The observability solution must support custom dashboards for multiple stakeholder groups and provide alerting capabilities for token consumption across all the foundation models that the company??s applications use.
Which combination of solutions will meet these requirements with the LEAST operational overhead? (Select TWO.)

A. Use Amazon CloudWatch metrics as data sources to create custom Amazon QuickSight dashboards that show token usage trends and usage patterns across FMs.
B. Use CloudWatch Logs Insights to analyze Amazon Bedrock invocation logs for token consumption patterns and usage attribution by applicatio
C. Create custom queries to identify high-usage scenario
D. Add log widgets to dashboards to enable continuous monitoring.
E. Create custom Amazon CloudWatch dashboards that combine native Amazon Bedrock token and invocation CloudWatch metric
F. Set up CloudWatch alarms to monitor token usage thresholds.

G. Create dashboards that show token usage trends and patterns across the company??s FMs by using an Amazon Bedrock zero-ETL integration with Amazon Managed Grafana.
H. Implement Amazon EventBridge rules to capture Amazon Bedrock model invocation event
I. Route token usage data to Amazon OpenSearch Serverless by using Amazon Data Firehos
J. Use OpenSearch dashboards to analyze usage patterns.

Answer: CD

NEW QUESTION 5
A company is building a legal research AI assistant that uses Amazon Bedrock with an Anthropic Claude foundation model (FM). The AI assistant must retrieve highly relevant case law documents to augment the FM??s responses. The AI assistant must identify semantic relationships between legal concepts, specific legal terminology, and citations. The AI assistant must perform quickly and return precise results.
Which solution will meet these requirements?

A. Configure an Amazon Bedrock knowledge base to use a default vector search configuratio
B. Use Amazon Bedrock to expand queries to improve retrieval for legal documents based on specific terminology and citations.
C. Use Amazon OpenSearch Service to deploy a hybrid search architecture that combines vector search with keyword searc
D. Apply an Amazon Bedrock reranker model to optimize result relevance.
E. Enable the Amazon Kendra query suggestion feature for end user
F. Use Amazon Bedrock to perform post-processing of search results to identify semantic similarity in the documents and to produce precise results.
G. Use Amazon OpenSearch Service with vector search and Amazon Bedrock Titan Embeddings to index and search legal document
H. Use custom AWS Lambda functions to merge results with keyword-based filters that are stored in an Amazon RDS database.

Answer: B

NEW QUESTION 6
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 7
A company needs a system to automatically generate study materials from multiple content sources. The content sources include document files (PDF files, PowerPoint presentations, and Word documents) and multimedia files (recorded videos). The system must process more than 10,000 content sources daily with peak loads of 500 concurrent uploads. The system must also extract key concepts from document files and multimedia files and create contextually accurate summaries. The generated study materials must support real-time collaboration with version control.
Which solution will meet these requirements?

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

Answer: B

NEW QUESTION 8
A healthcare company is developing a document management system that stores medical research papers in an Amazon S3 bucket. The company needs a comprehensive metadata framework to improve search precision for a GenAI application. The metadata must include document timestamps, author information, and research domain classifications.
The solution must maintain a consistent metadata structure across all uploaded documents and allow foundation models (FMs) to understand document context without accessing full content.
Which solution will meet these requirements?

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

Answer: A

NEW QUESTION 9
A medical company is creating a generative AI (GenAI) system by using Amazon Bedrock. The system processes data from various sources and must maintain end-to-end data lineage. The system must also use real-time personally identifiable information (PII) filtering and audit trails to automatically report compliance.
Which solution will meet these requirements?

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

Answer: A

NEW QUESTION 10
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 10
A healthcare company is using Amazon Bedrock to develop a real-time patient care AI assistant to respond to queries for separate departments that handle clinical inquiries, insurance verification, appointment scheduling, and insurance claims. The company wants to use a multi-agent architecture.
The company must ensure that the AI assistant is scalable and can onboard new features for patients. The AI assistant must be able to handle thousands of parallel patient interactions. The company must ensure that patients receive appropriate domain-specific responses to queries.
Which solution will meet these requirements?

A. Isolate data for each agent by using separate knowledge base
B. Use IAM filtering to control access to each knowledge bas
C. Deploy a supervisor agent to perform natural language intent classification on patient inquirie
D. Configure the supervisor agent to route queries to specialized collaborator agents to respond to department-specific querie
E. Configure each specialized collaborator agent to use Retrieval Augmented Generation(RAG) with the agent's department-specific knowledge base.
F. Create a separate supervisor agent for each departmen
G. Configure individual collaborator agents to perform natural language intent classification for each specialty domain within each departmen

H. Integrate each collaborator agent with department-specific knowledge bases onl
I. Implement manual handoff processes between the supervisor agents.
J. Isolate data for each department in separate knowledge base
K. Use IAM filtering to control access to each knowledge bas
L. Deploy a single general-purpose agen
M. Configure multiple action groups within the general-purpose agent to perform specific department function
N. Implement rule-based routing logic in the general-purpose agent instructions.
O. Implement multiple independent supervisor agents that run in parallel to respond to patient inquiries for each departmen
P. Configure multiple collaborator agents for each supervisor agen
Q. Integrate all agents with the same knowledge bas
R. Use external routing logic to merge responses from multiple supervisor agents.

Answer: A

NEW QUESTION 13
Company configures a landing zone in AWS Control Tower. The company handles sensitive data that must remain within the European Union. The company must use only the eu-central-1 Region. The company uses Service Control Policies (SCPs) to enforce data residency policies. GenAI developers at the company are assigned IAM roles that have full permissions for Amazon Bedrock.
The company must ensure that GenAI developers can use the Amazon Nova Pro model through Amazon Bedrock only by using cross-Region inference (CRI) and only in eu- central-1. The company enables model access for the GenAI developer IAM roles in Amazon Bedrock. However, when a GenAI developer attempts to invoke the model through the Amazon Bedrock Chat/Text playground, the GenAI developer receives the following error:
User arn:aws:sts:123456789012:assumed-role/AssumedDevRole/DevUserName Action: bedrock:InvokeModelWithResponseStream
On resource(s): arn:aws:bedrock:eu-west-3::foundation-model/amazon.nova-pro-v1:0 Context: a service control policy explicitly denies the action
The company needs a solution to resolve the error. The solution must retain the company's existing governance controls and must provide precise access control. The solution must comply with the company's existing data residency policies.
Which combination of solutions will meet these requirements? (Select TWO.)

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

Answer: BE

NEW QUESTION 17
A book publishing company wants to build a book recommendation system that uses an AI assistant. The AI assistant will use ML to generate a list of recommended books from the company's book catalog. The system must suggest books based on conversations with customers.
The company stores the text of the books, customers' and editors' reviews of the books, and extracted book metadata in Amazon S3. The system must support low-latency responses and scale efficiently to handle more than 10,000 concurrent users.
Which solution will meet these requirements?

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

Answer: A

NEW QUESTION 22
A university recently digitized a collection of archival documents, academic journals, and manuscripts. The university stores the digital files in an AWS Lake Formation data lake.
The university hires a GenAI developer to build a solution to allow users to search the digital files by using text queries. The solution must return journal abstracts that are semantically similar to a user's query. Users must be able to search the digitized collection based on text and metadata that is associated with the journal abstracts. The metadata of the digitized files does not contain keywords. The solution must match similar abstracts to one another based on the similarity of their text. The data lake contains fewer than 1 million files.
Which solution will meet these requirements with the LEAST operational overhead?

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

Answer: D

NEW QUESTION 23
An ecommerce company is using Amazon Bedrock to build a generative AI (GenAI) application. The application uses AWS Step Functions to orchestrate a multi-agent workflow to produce detailed product descriptions. The workflow consists of three sequential states: a description generator, a technical specifications validator, and a brand voice consistency checker. Each state produces intermediate reasoning traces and outputs that are passed to the next state. The application uses an Amazon S3 bucket for process storage and to store outputs.
During testing, the company discovers that outputs between Step Functions states frequently exceed the 256 KB quota and cause workflow failures. A GenAI Developer needs to revise the application architecture to efficiently handle the Step Functions 256 KB quota and maintain workflow observability. The revised architecture must preserve the existing multi-agent reasoning and acting (ReAct) pattern.
Which solution will meet these requirements with the LEAST operational overhead?

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

Answer: B

NEW QUESTION 24
A financial services company uses multiple foundation models (FMs) through Amazon Bedrock for its generative AI (GenAI) applications. To comply with a new regulation for GenAI use with sensitive financial data, the company needs a token management solution.
The token management solution must proactively alert when applications approach model- specific token limits. The solution must also process more than 5,000 requests each minute and maintain token usage metrics to allocate costs across business units.
Which solution will meet these requirements?

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

Answer: A

NEW QUESTION 26
A company has a recommendation system running on Amazon EC2 instances. The applications make API calls to Amazon Bedrock foundation models (FMs) to analyze
customer behavior and generate personalized product recommendations.
The system experiences intermittent issues where some recommendations do not match customer preferences. The company needs an observability solution to monitor operational metrics and detect patterns of performance degradation compared to established baselines. The solution must generate alerts with correlation data within 10 minutes when FM behavior deviates from expected patterns.
Which solution will meet these requirements?

A. Configure Amazon CloudWatch Container Insight
B. Set up alarms for latency threshold
C. Add custom token metrics using the CloudWatch embedded metric format.
D. Implement AWS X-Ra
E. Enable CloudWatch Logs Insight
F. Set up AWS CloudTrail and create dashboards in Amazon QuickSight.
G. Enable Amazon CloudWatch Application Insight
H. Create custom metrics for recommendation quality, token usage, and response latency using the CloudWatch embedded metric format with dimensions for request types and user segment
I. Configure CloudWatch anomaly detection on model metric
J. Use CloudWatch Logs Insights for pattern analysis.
K. Use Amazon OpenSearch Service with the Observability plugi
L. Ingest metrics and logs through Amazon Kinesis and analyze behavior with custom queries.

Answer: C

NEW QUESTION 30
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 32
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 37
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 38
A company wants to select a new FM for its AI assistant. A GenAI developer needs to generate evaluation reports to help a data scientist assess the quality and safety of various foundation models FMs. The data scientist provides the GenAI developer with sample prompts for evaluation. The GenAI developer wants to use Amazon Bedrock to automate report generation and evaluation.
Which solution will meet this requirement?

A. Combine the sample prompts into a single JSON documen
B. Create an Amazon Bedrock knowledge base with the documen
C. Write a prompt that asks the FM to generate a response to each sample promp
D. Use the RetrieveAndGenerate API to generate a report for each model.
E. Combine the sample prompts into a single JSONL documen

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

Answer: B

NEW QUESTION 41
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 42
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 44
A company uses Amazon Bedrock to generate technical content for customers. The company has recently experienced a surge in hallucinated outputs when the company??s model generates summaries of long technical documents. The model outputs include inaccurate or fabricated details. The company??s current solution uses a large foundation model (FM) with a basic one-shot prompt that includes the full document in a single input.
The company needs a solution that will reduce hallucinations and meet factual accuracy goals. The solution must process more than 1,000 documents each hour and deliver summaries within 3 seconds for each document.
Which combination of solutions will meet these requirements? (Select TWO.)

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

Answer: BC

NEW QUESTION 45
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 50
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:
python
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 55
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 57
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 62
A company is implementing a serverless inference API by using AWS Lambda. The API will dynamically invoke multiple AI models hosted on Amazon Bedrock. The company needs to design a solution that can switch between model providers without modifying or redeploying Lambda code in real time. The design must include safe rollout of configuration changes and validation and rollback capabilities.
Which solution will meet these requirements?

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

Answer: B

NEW QUESTION 64
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 69
A financial services company needs to pre-process unstructured data such as customer transcripts, financial reports, and documentation. The company stores the unstructured data in Amazon S3 to support an Amazon Bedrock application.
The company must validate data quality, create auditable metadata, monitor data metrics, and customize text chunking to optimize foundation model (FM) performance.
Which solution will meet these requirements with the LEAST development effort?

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

Answer: B

NEW QUESTION 73
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 76
A financial services company is developing a Retrieval Augmented Generation (RAG) application to help investment analysts query complex financial relationships across multiple investment vehicles, market sectors, and regulatory environments. The dataset contains highly interconnected entities that have multi-hop relationships. Analysts must examine relationships holistically to provide accurate investment guidance. The application must deliver comprehensive answers that capture indirect relationships between financial entities and must respond in less than 3 seconds.
Which solution will meet these requirements with the LEAST operational overhead?

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

Answer: A

NEW QUESTION 78
A media company is launching a platform that allows thousands of users every hour to upload images and text content. The platform uses Amazon Bedrock to process the uploaded content to generate creative compositions.
The company needs a solution to ensure that the platform does not process or produce inappropriate content. The platform must not expose personally identifiable information (PII) in the compositions. The solution must integrate with the company's existing Amazon S3 storage workflow.
Which solution will meet these requirements with the LEAST infrastructure management overhead?

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

Answer: D

NEW QUESTION 81
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 83

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 87
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 92
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 93
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 97
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 102
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 103
A company is using Amazon Bedrock to develop an AI-powered application that uses a foundation model (FM) that supports cross-Region inference and provisioned throughput. The application must serve users in Europe and North America with consistently low latency. The application must comply with data residency regulations that require European user data to remain within Europe-based AWS Regions.
During testing, the application experiences service degradation when Regional traffic spikes reach service quotas. The company needs a solution that maintains application resilience and minimizes operational complexity.
Which solution will meet these requirements?

A. Deploy separate Amazon Bedrock instances in North American and European Region
B. Use a custom routing layer that directs traffic based on user locatio
C. Configure Amazon CloudWatch alarms to monitor Regional service usag
D. Use Amazon SNS to send email alerts when usage approaches thresholds.
E. Use Amazon Bedrock cross-Region inference profiles by specifying geographical codes in profile IDs when calling the InvokeModel AP
F. Configure separate Amazon API Gateway HTTP APIs to direct European and North American users to the appropriate Regional endpoints.
G. Deploy a multi-Region Amazon API Gateway HTTP API and AWS Lambda functions that implement retry logic to handle throttlin
H. Configure the Lambda functions to call the FM in the nearest secondary Region when quotas are reached.
I. Configure provisioned throughput for Amazon Bedrock in multiple Region
J. Implement failover logic in application code to switch Regions when throttling occur

K. Use AWS Global Accelerator to route traffic based on user location.

Answer: B

NEW QUESTION 106
.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual AIP-C01 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the AIP-C01 Product From:

https://www.2passeasy.com/dumps/AIP-C01/

Money Back Guarantee

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