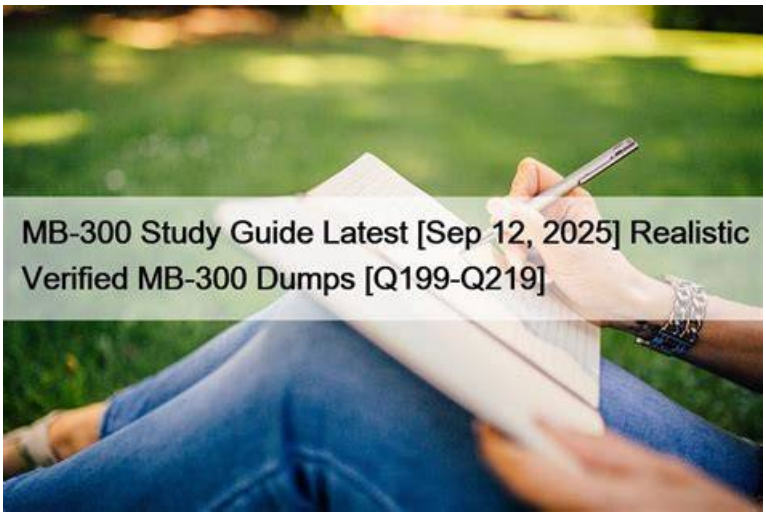
AI-300 Latest Dumps Free - Review AI-300 Guide



P.S. Free 2026 Microsoft AI-300 dumps are available on Google Drive shared by Actual4Cert: https://drive.google.com/open?id=1gLz4t6hKB5Z3AFb1SfkY-4tM3zDwX6Ig

Using Actual4Cert's AI-300 test certification training materials to pass AI-300 certification exam is easy. Our AI-300 test certification training materials is made up of senior IT specialist team through their own exploration and continuous practice and research. Our Actual4Cert's AI-300 test certification training materials can help you in your first attempt to pass AI-300 exam easily.

Microsoft AI-300 Exam Syllabus Topics:

Section	Objectives
Operationalizing machine learning solutions	- ML lifecycle management • 1. Model versioning and registry usage • 2. Model training and evaluation in Azure Machine Learning - Deployment and monitoring • 1. Monitor performance and drift • 2. Deploy models to endpoints - Scalability and performance optimization • 1. Cost optimization strategies • 2. Autoscaling AI workloads
Implement secure and scalable AI systems	- Security and governance • 1. Identity and access management for AI services • 2. Data privacy and compliance considerations
Design and implement generative AI solutions	- Large language model integration • 1. Prompt engineering and prompt flow design • 2. Use Azure OpenAI Service capabilities - RAG (Retrieval Augmented Generation) solutions • 1. Vector search integration • 2. Knowledge grounding and retrieval design

Plan and design AI solutions using Azure AI services	- Responsible AI design • 1. Responsible AI mitigation strategies • 2. Fairness, transparency, and accountability considerations - Requirements gathering and solution architecture • 1. Select appropriate Azure AI services • 2. Identify business requirements for AI solutions

>> AI-300 Latest Dumps Free <<

AI-300 Exam Guides - AI-300 Test Answers & AI-300 Exam Torrent

Now IT industry is more and more competitive. Passing Microsoft AI-300 exam certification can effectively help you entrench yourself and enhance your status in this competitive IT area. In our Actual4Cert you can get the related Microsoft AI-300 exam certification training tools. Our Actual4Cert IT experts team will timely provide you the accurate and detailed training materials about Microsoft Certification AI-300 Exam. Through the learning materials and exam practice questions and answers provided by Actual4Cert, we can ensure you have a successful challenge when you are the first time to participate in the Microsoft certification AI-300 exam. Above all, using Actual4Cert you do not spend a lot of time and effort to prepare for the exam.

Microsoft Operationalizing Machine Learning and Generative AI Solutions Sample Questions (Q110-Q115):

NEW QUESTION # 110
Hotspot Question
You have an Azure Machine Learning workspace named Workspace1.
You plan to train an image classification model by using Automated ML in Workspace1.
You need to complete the provided Azure Machine Learning Python SDK v2 code to bring labeled image data as input for model training.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

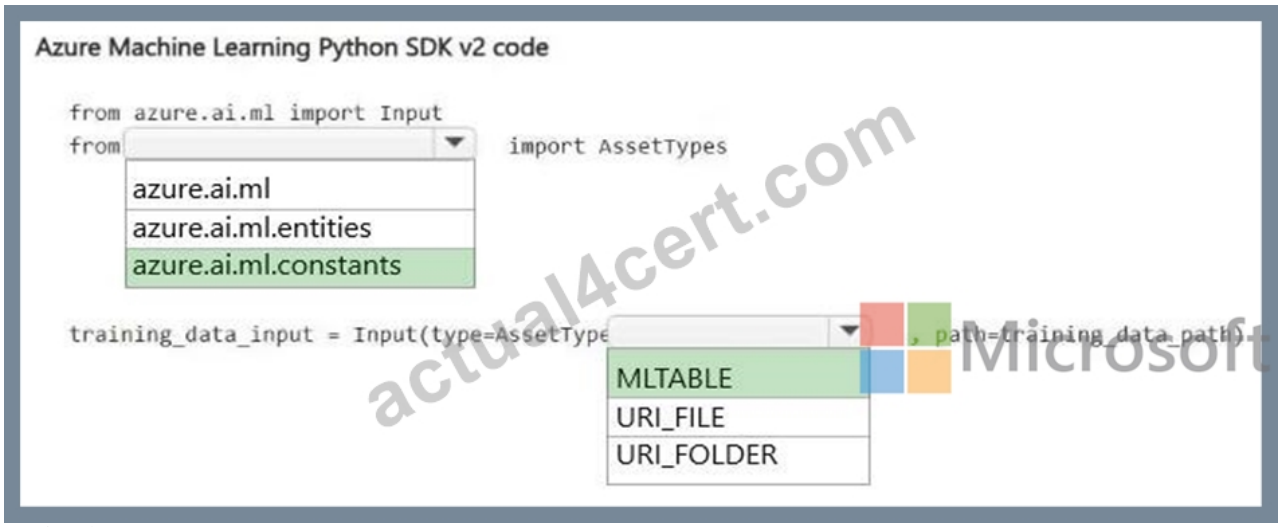
Azure Machine Learning Python SDK v2 code

```
from azure.ai.ml import Input
from [azure.ai.ml | azure.ai.ml.entities | azure.ai.ml.constants] import AssetTypes

training_data_input = Input(type=AssetType.[MLTABLE | URI_FILE | URI_FOLDER], path=training_data_path)
```


Answer:

Explanation:



Explanation:
Box 1: azure.ai.ml.constants
azure.ai.ml.constants is the official SDK v2 submodule where the AssetType enum resides.
Box 2: MLTABLE
AssetType.MLTABLE: Automated ML for Computer Vision tasks (such as image classification and object detection) specifically requires your input data and corresponding label annotations to be provided via an MLTable asset type. This structure points to a folder containing your dataset configurations and your .jsonl bounding box or classification files.
Reference:
https://learn.microsoft.com/en-us/azure/machine-learning/how-to-auto-train-image-models

NEW QUESTION # 111
Drag and Drop Question
An organization is deploying generative AI solutions by using Microsoft Foundry to support multiple production workloads.
The organization has the following workload requirements:
- One workload must be real-time, latency-sensitive, and have predictable global usage patterns that demand consistent performance.
- One workload must have variable performance and be optimized for cost-efficient operation.
You need to select a global deployment type for each workload.
Which type of deployment should you use for each workload requirement? To answer, move the appropriate deployment types to the correct requirements. You may use each deployment type once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Answer:

Explanation:

NEW QUESTION # 112

Case Study 1 - Fabrikam Inc.

Background

Fabrikam Inc. is a mid-sized healthcare analytics company that provides population health dashboards and predictive insights to regional hospital systems across the United States.

Fabrikam Inc. customers rely on near real time analytics to monitor patient flow, staffing needs, and readmission risks. They use multiple traditional forecasting machine learning models for predictions.

Fabrikam Inc. has an established Microsoft Azure footprint. The company uses Jupyter Notebooks that run on a local server as the primary development environment. The data science team is experiencing scalability, asset management and code management issues with the current development platform. Fabrikam Inc. plans to migrate to a cloud-based development environment to mitigate the issues.

Additionally, the company plans to implement a Retrieval-Augmented Generation (RAG)-based chat application for client support. Leadership requires the application to be developed and deployed with a low operational risk.

Current Environment

Fabrikam Inc. operates a single Azure subscription that has the following components:

* Azure Data Lake Storage Gen2 that contains de-identified clinical and operational datasets
* Azure AI Search indexing curated analytical documents and reference materials
* A small set of Python-based training scripts maintained by data scientists
* Azure OpenAI Service with deployed foundational models
* A Microsoft Foundry resource for building a RAG-based solution

Evaluation data has manually defined expected responses.

The current challenges faced by the data science team include the following:

* Model training jobs are run manually from notebooks.
* Experiment tracking is inconsistent
* Model versions are registered without standardized metadata.
* Deployment is performed manually by data scientists, with limited rollback capability.
* The team has no standardized evaluation process for generative AI outputs.

The environment currently allows public network access. Authentication relies on user accounts rather than managed identities. Compute targets are manually created and shared across experiments. This has led to resource contention during peak usage.

Business Requirements

Fabrikam Inc. has the following business requirements for the modernization initiative:

* Provide a conversational interface that answers analytics questions by using internal documents and datasets.
* Ensure that sensitive healthcare-related data is not exposed outside the Fabrikam Inc. Azure tenant.
* Enable repeatable and auditable model training and deployment processes.
* Support experimentation to compare prompt strategies and fine-tuned models.
* Align the model with the ranked preferences and optimize behavior for the long term.
* Minimize disruption to existing analytics workloads during rollout.

Technical Requirements

To support the business goals, Fabrikam Inc. identifies these technical requirements:

* Use Azure Machine Learning workspaces to centrally manage data assets, models, and environments.
* Implement experiment tracking and model versioning for all training jobs.
* Orchestrate training and evaluation by using pipelines rather than manually running notebooks.
* Deploy traditional machine learning models with support for staged rollout and rollback.
* Improve RAG-based solution output quality.
* Use the existing evaluation datasets that are based on real data with input-output pairs.
* Apply advanced fine-tuning techniques only when prompt engineering is insufficient Issues and Constraints Fabrikam Inc. must comply with internal security policies that require the company to restrict network access and avoid long-lived secrets. The data science team has limited Azure DevOps experience, so solutions must favor managed services and automation over custom infrastructure.

Cost predictability is important. Leadership prefers serverless or managed compute options where possible but is willing to approve dedicated compute for stable production workloads.

Problem Statement

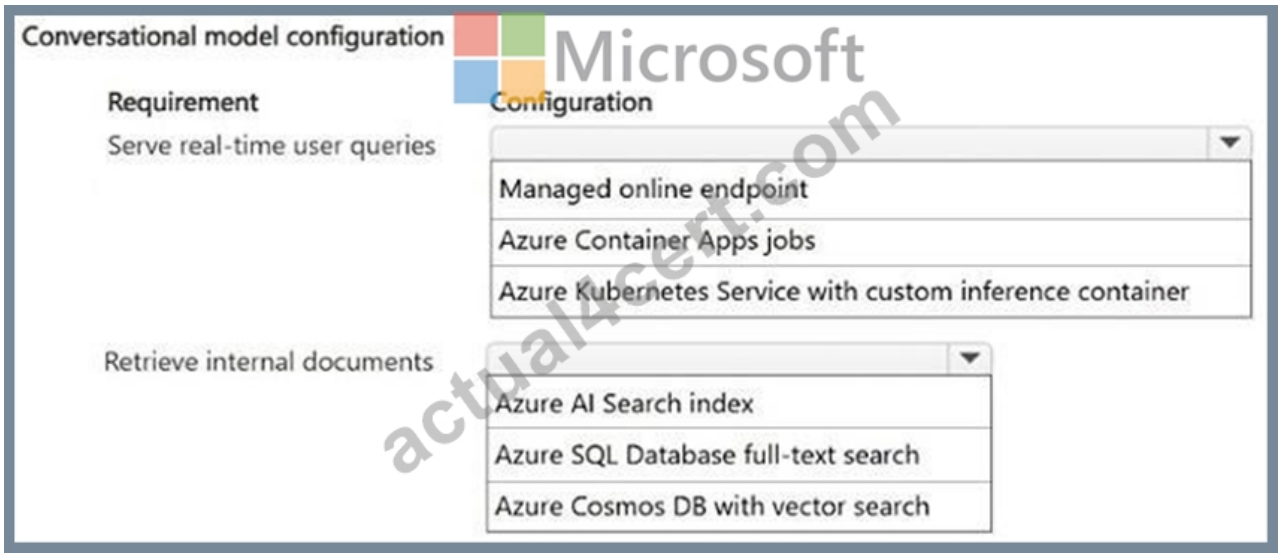
Fabrikam Inc. must design and implement an Azure-based AI operations solution that enables reliable training, evaluation, deployment, and iteration of generative AI models. The solution must support experimentation and gradual rollout while ensuring governance, security, and operational stability. The data science and platform teams must collaborate to deliver this solution by using Azure Machine Learning and Microsoft Foundry capabilities.

Hotspot Question

You need to deploy the RAG-based chat application that meets Fabrikam Inc.'s business and technical requirements.

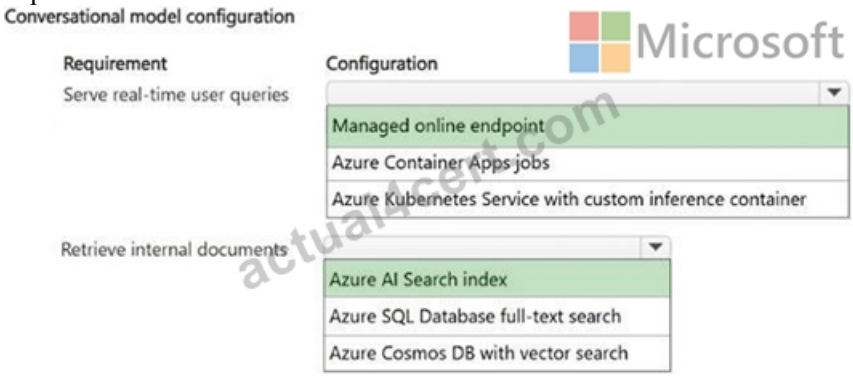
Which configuration should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION # 113

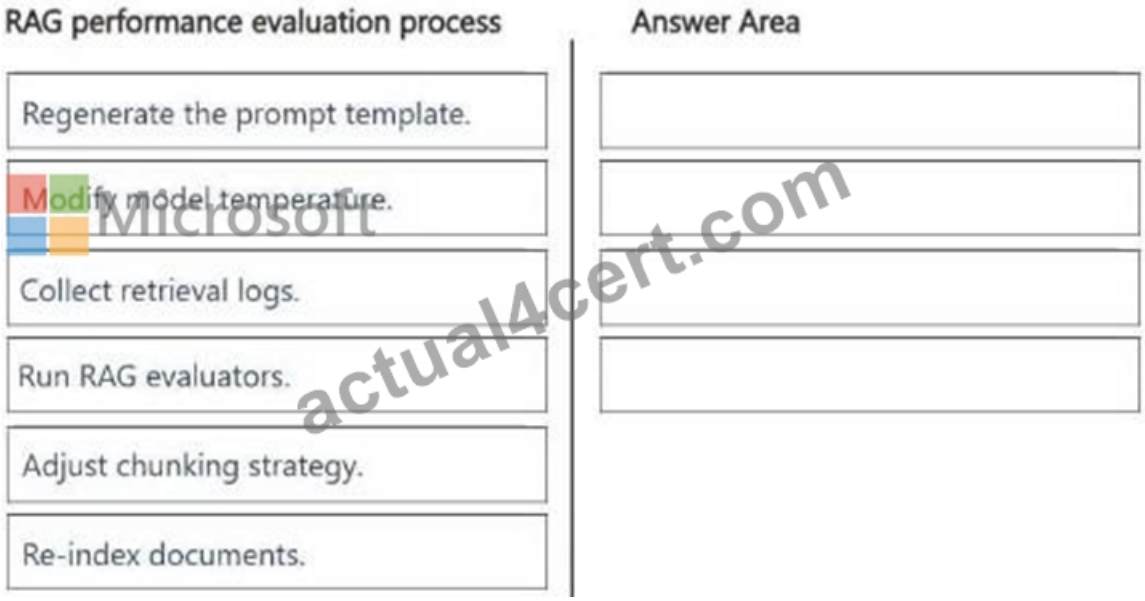
Drag and Drop Question

A team is developing a Retrieval-Augmented Generation (RAG) system.

The team requires improvements to the system's retrieval quality to ensure accurate, grounded responses.

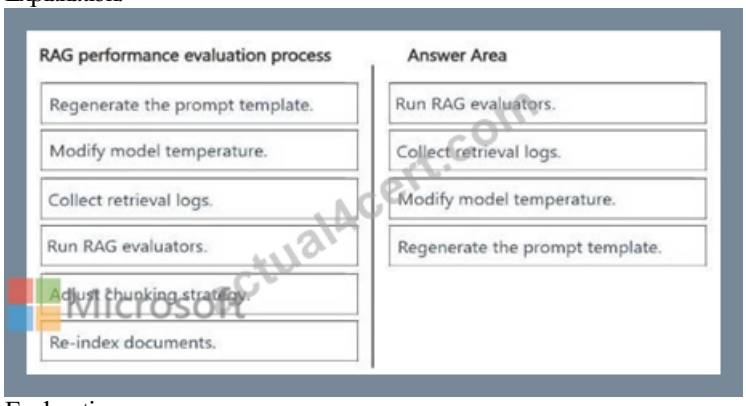
You need to assess RAG performance before you can suggest an improvement strategy.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:

Explanation:



Explanation:

To properly assess your Azure Retrieval-Augmented Generation (RAG) system's performance before implementing an improvement strategy, you should include the following four steps: Run RAG evaluators, Collect retrieval logs, Modify model temperature, and Regenerate the prompt template.

Step 1: Run RAG evaluators

Run RAG evaluators is the primary method to objectively measure system performance.

Evaluators calculate data-driven metrics like groundedness, relevance, and retrieval precision using tools like Azure AI Studio Evaluators.

Step 2: Collect retrieval logs

Collect retrieval logs provides the raw operational data needed for assessment. Analyzing these logs helps you identify exactly which documents were retrieved, their relevance scores, and where the retrieval pipeline failed to fetch the correct context.

Step 3: Modify model temperature

Modify model temperature: Adjusting the temperature during assessment helps isolate whether poor responses are caused by bad retrieval or by the LLM being too creative (high temperature) or too rigid (low temperature). Testing variations helps establish a performance baseline.

Step 4: Regenerate the prompt template

Regenerate the prompt template: Evaluating how different prompt variations alter the output allows you to assess if the current template is effectively forcing the model to rely only on the retrieved context, which is critical for identifying grounding issues.

Incorrect:

Adjust the chunking strategy

This is an improvement action, not an assessment step. You would perform this optimization strategy after your assessment reveals that information is being cut off or poorly contextualized.

Re-index documents: This is a heavy remediation step. Re-indexing is a time- and resource- consuming strategy used to fix issues once the assessment phase has already proven that the current index or embedding model is faulty.

Reference:

https://flytoleisure.medium.com/guideline-for-building-a-practical-and-effective-rag-retrieval-augmented-generation-application-f6cf50676e37

NEW QUESTION # 114

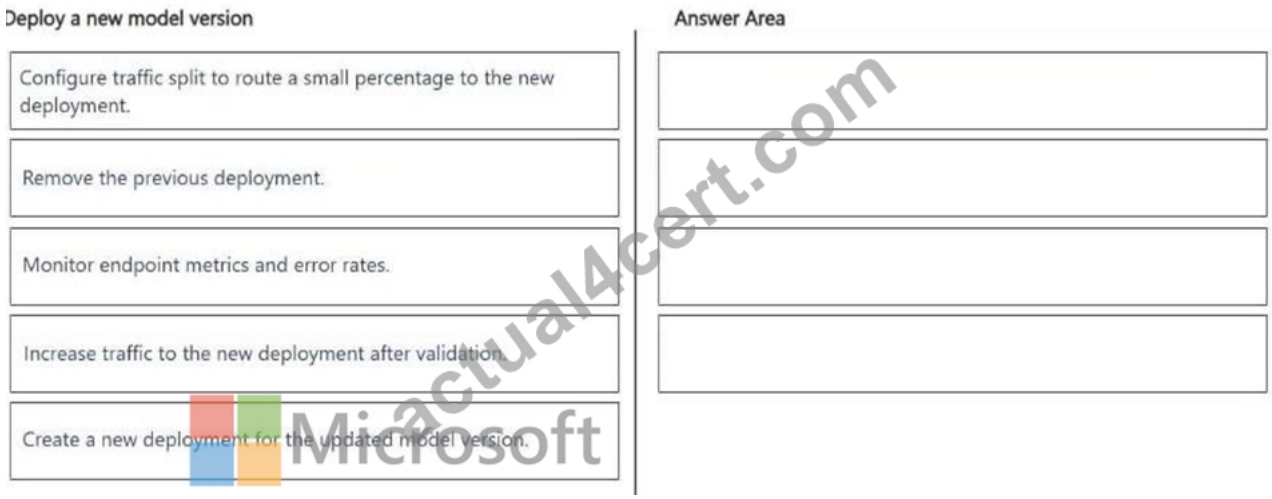
Drag and Drop Question

A team is deploying a new version of a customer scoring model to a production online endpoint in Azure Machine Learning.

The team must minimize risk by gradually introducing the new model version and ensuring that traffic can be reverted immediately if issues occur.

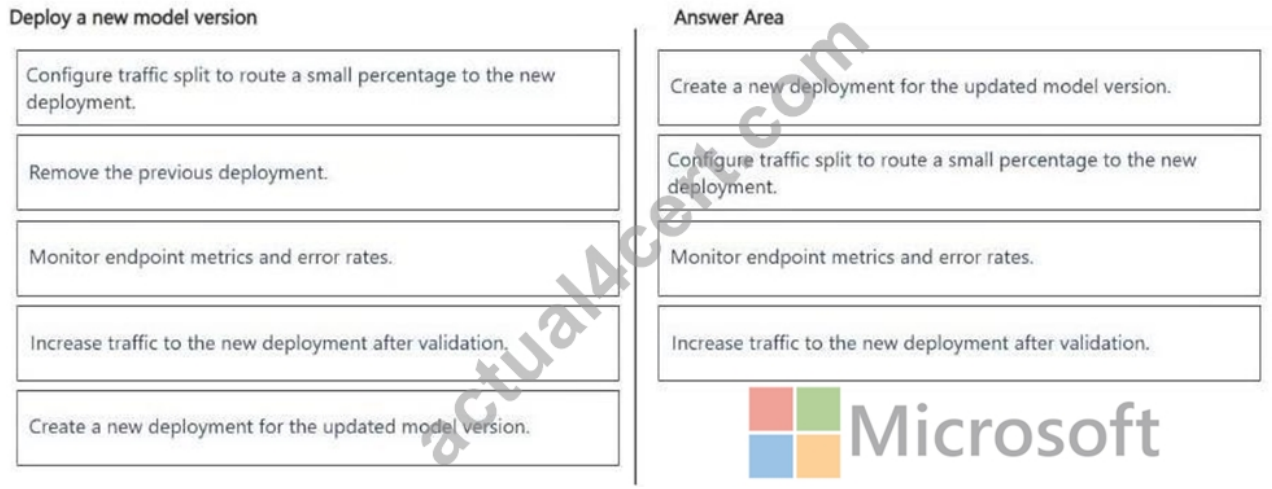
You need to roll out the new model according to the requirements.

Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:

Explanation:



Explanation:

To safely deploy your new customer scoring model with minimal risk, you should use Azure Machine Learning managed online endpoints with a blue-green deployment strategy. This allows you to test the new model (green) alongside the old model (blue) and route traffic gradually.

Here is the step-by-step process to execute this rollout.

Step 1: Create a New Deployment for the updated model version.

1. Create a New Deployment under the Existing Endpoint

Do not overwrite your current production deployment. Instead, deploy the new model version as a separate deployment within the same managed online endpoint.

Keep the existing deployment (e.g., blue-deployment) running at 100% traffic.

Create the new deployment (e.g., green-deployment) with the new model version.

Set the initial traffic allocation for the new deployment to 0%.

2. Test the New Deployment in Isolation [Step not in this question]

Before exposing real customers to the new model, test its stability and performance in the production environment.

Send test scoring requests directly to the green-deployment.

Use the X-MS-Routing-Override: green-deployment HTTP header in your SDK or REST requests to bypass the traffic split and target the new model directly.

Verify that the data schema, latency, and prediction outputs match your expectations.

Step 2: Configure traffic split to route a small percentage to the new deployment.

3. Gradually Shift Traffic (Canary Testing)

Once internal testing passes, begin introducing real production traffic to the new model in phases.

3a. Update the endpoint's traffic configuration to route a small percentage (e.g., 10%) to green- deployment. [Step 2] Step 3: Monitor endpoint metrics and error rates.

3b) Monitor system metrics closely, looking for spikes in HTTP 4xx/5xx errors, timeouts, or unusual scoring distributions. [Step 3]

Step 4: Increase traffic to the new deployment after validation

3c) If performance remains stable over a designated observation period, incrementally increase the traffic (e.g., to 30%, 50%, then

100%). [Step 4] Further future actions outside the context of the question:
4. Monitor App Insights and MetricsAzure ML integrates natively with Azure Monitor and Application Insights. Track these critical metrics during the shift:CPU/Memory Utilization: Ensure the new model does not exhaust container resources.
Model Latency: Compare the response times of the new deployment against the old one.
Tracking Token/Feature Drift: Ensure the live customer data matches the model's expected input feature distribution.
5. Instant Rollback Plan
If the new model causes errors, triggers high latency, or skews customer scores, you can revert traffic immediately.
Update the endpoint traffic settings to route 100% of traffic back to the original blue-deployment.
Set the green-deployment traffic to 0%.This configuration change takes effect almost instantly, isolating the faulty model without any downtime for your users.
Keep the green-deployment alive at 0% traffic while you pull the logs to investigate the root cause, then delete it once fixed.
Incorrect:
[Remove the previous deployment]
A step further in the future.
Reference:
https://chandras4bigdata.medium.com/mlops-interview-and-study-guide-489d13ff82fa

NEW QUESTION # 115
.....