Microsoft

AI-300

Operationalizing Machine Learning and Generative AI Solutions

For More Information – Visit link below:

https://www.examsempire.com/

Product Version

1. **Up to Date products, reliable and verified.**
2. **Questions and Answers in PDF Format.**



Latest Version: 8.1

Topic 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.

Visit us at: https://www.examsempire.com/ai-300

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.

Question: 1

You need to standardize how Fabrikam Inc. manages machine learning assets.
Which action should you perform first?

A. Register assets in the Azure Machine Learning registry.
B. Create a shared Azure Machine Learning workspace.
C. Deploy a managed online endpoint.
D. Create a new Microsoft Foundry project.

Answer: B

Explanation:
In Azure Machine Learning, a workspace is the top-level resource that acts as a centralized hub for every ML activity. It stores experiments, pipelines, datasets, models, environments, and compute targets all in one place. Before you can register assets or deploy endpoints, the workspace itself must exist and be shared across the team. Fabrikam's core challenge is inconsistent experiment tracking, ad-hoc versioning, and no standardized deployment path — all symptoms of the absence of a single, governed workspace. Option A (registering assets) is a subsequent action only possible after the workspace exists. Option C (online endpoint) is a deployment concern, not an asset-management foundation. Option D (Foundry project) is appropriate for generative AI workloads, not for the traditional ML pipeline standardization described here. The workspace is the prerequisite that makes every other governance action possible.
Microsoft Learn Reference Topic: Azure Machine Learning workspaces – Microsoft Learn: Manage Azure Machine Learning workspaces

Question: 2

You need to isolate training workloads while remaining cost-aware to address Fabrikam Inc.'s issues, constraints, and technical requirements.
What should you implement?

A. Training jobs that run on a single shared compute cluster
B. Fixed-size compute cluster
C. Dedicated compute clusters per experiment
D. Managed compute targets with autoscaling

Answer: D

Explanation:
Imagine a hospital analytics firm with data scientists who kick off dozens of training jobs throughout the week. During peak hours, five jobs compete for the same GPU cluster and fail or queue for hours. On quiet nights, that cluster sits completely idle, burning money. Managed compute targets with autoscaling solve both problems: the cluster scales out automatically when multiple jobs arrive simultaneously and scales back to zero when idle. Option A (single shared cluster) is exactly the resource-contention problem Fabrikam already has. Option B (fixed-size cluster) wastes money during off-peak hours. Option C (dedicated per-experiment clusters) eliminates contention but is prohibitively expensive for a cost-conscious healthcare firm. Autoscaling managed compute is the cloud-native answer to variable workload demand.
Microsoft Learn Reference Topic: Azure Machine Learning compute targets – Configure autoscale for compute clusters

Question: 3

You need to recommend an experiment-tracking strategy that ensures consistent experiment results.
What should you recommend?

A. Azure Machine Learning job output logs
B. MLflow experiment tracking
C. Application Insights logs
D. Azure Monitor alerts

Answer: B

Explanation:
MLflow is the industry-standard open-source platform for experiment tracking, and Azure Machine Learning has first-class native integration with it. When you use MLflow within an Azure ML job, parameters, metrics, and artifacts are automatically logged to the run history of the AML workspace, making every run reproducible and comparable. Option A (AML job output logs) only captures console output and lacks structured parameter and metric logging. Option C (Application Insights logs) is designed for application-level telemetry, not ML experiment metadata. Option D (Azure Monitor alerts) is a reactive notification tool, not a tracking system. MLflow's autologging capability means that for common frameworks such as scikit-learn, XGBoost, and PyTorch, parameters and metrics are captured without a single line of custom code, directly answering Fabrikam's requirement for consistent experiment tracking.
Microsoft Learn Reference Topic: Track ML experiments with MLflow – Azure Machine Learning MLflow integration

Question: 4

HOTSPOT

RAG optimization

Requirement	Strategy
Domain specialization	Supervised fine-tuning Reinforcement fine-tuning Direct preference optimization
Poor response accuracy of a RAG-based solution	Apply prompt engineering. Adjust the temperature parameter. Increase the max_tokens parameter.

Answer:

RAG optimization

Requirement	Strategy
Domain specialization	Supervised fine-tuning Reinforcement fine-tuning Direct preference optimization
Poor response accuracy of a RAG-based solution	Apply prompt engineering. Adjust the temperature parameter. Increase the max_tokens parameter.

Question: 5

Fabrikam Inc. needs to improve the performance of a GPT-5 model based on the stated technical requirements.
Which action should you perform first?

A. Deploy the model to production to gather real-world feedback.
B. Evaluate the model output.
C. Fine-tune the model to improve accuracy.
D. Generate synthetic interaction data.

Answer: B

Explanation:

The first action should be to evaluate the model output. Fabrikam already has evaluation data containing manually defined expected responses, and the technical requirements explicitly state that the organization must use its existing evaluation datasets based on real data with input-output pairs. Microsoft Foundry supports evaluating a model against an existing curated dataset to establish measurable baseline performance before optimization. Evaluation results can identify weaknesses in accuracy, relevance, groundedness, safety, or other quality dimensions and provide objective evidence for subsequent changes.

This sequencing is critical because Fabrikam also requires that advanced fine-tuning be applied only when prompt engineering is insufficient. Therefore, immediately fine-tuning the model would bypass the required baseline evaluation and optimization process. Fine-tuning should be driven by demonstrated performance gaps, not applied automatically.

Deploying directly to production is also inappropriate because leadership requires low operational risk and gradual rollout. Production traffic should not be used as the first evaluation mechanism when curated evaluation data already exists. Synthetic interaction data is useful when evaluation coverage is insufficient, but Microsoft specifically positions synthetic generation as an option when an adequate test dataset is unavailable. Fabrikam already has suitable evaluation data.

Study Guide Optimize generative AI systems and model performance — evaluate model performance, benchmark against curated datasets, iterate on prompts, and apply advanced fine-tuning when justified.

Visit us at: https://www.examsempire.com/ai-300

Thank You for Trying Our Product

Special 16 USD Discount Coupon: NSZUBG3X

Email: support@examsempire.com

Check our Customer Testimonials and ratings available on every product page.

Visit our website.

https://examsempire.com/