

Exam Questions AI-200

Developing AI Cloud Solutions on Azure

<https://www.2passeasy.com/dumps/AI-200/>



NEW QUESTION 1

You are developing a microservices-based application that uses Azure Container Apps. The application consists of several containerized services that handle tasks, such as processing orders, managing inventory, and generating reports. You must secure the container apps. All apps must reside in the same virtual network, share the same Dapr configuration, and share the same logging location. Apps must support the configuration of the amount of memory and compute resources available to containers. You need to configure the Azure Container App. How should you complete the CLI command? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

`-n MyContainerApp -g MyResourceGroup --location eastus2`

▼

env

create

add-on

▼

env

create

add-on

`-n MyContainerApp -g MyResourceGroup --location eastus2`

▼

--enable-mtls
--internal-only
--enable-workload-profiles

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Verified Answer: ``az containerapp env create ... --enable-workload-profiles``.

Detailed Explanation: The shared boundary for Container Apps is the managed environment. Apps in one environment can share the environment's virtual network integration, observability destination, and Dapr-related environment configuration. Enabling workload profiles provides selectable compute profiles and resource sizing options for workloads in that environment. The CLI must therefore create a Container Apps environment and enable workload profiles. Creating an individual container app alone would not establish the common environment-level network, logging, and resource-profile boundary required by the scenario. Study Guide Alignment: Containerized Azure workloads: registry builds, App Service containers, Container Apps revision/scaling behavior, and AKS deployment choices.

Official Microsoft Learn References: AI-200 Study Guide | Azure Container Apps environments

NEW QUESTION 2

You develop an ASP.NET Core app that uses Azure App Configuration. You also create an App Configuration containing 100 settings. The app must meet the following requirements:

- Ensure the consistency of all configuration data when changes to individual settings occur.
- Handle configuration data changes dynamically without causing the application to restart.
- Reduce the overall number of requests made to App Configuration APIs

You must implement dynamic configuration updates in the app.

What are two ways to achieve this goal? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Create and configure Azure Key Vault
B. Implement the Azure Key Vault configuration provider
C. Register all keys in the App Configuration store. Set the `refreshAll` parameter of the `Register` method to `false`.
D. Create and register a sentinel key in the App Configuration store.
E. Set the `refreshAll` parameter of the `Register` method to `true`.
F. Decrease the App Configuration cache expiration from the default value.
G. Increase the App Configuration cache expiration from the default value.
H. Create and implement environment variables for each App Configuration store setting.

Answer: CE

Explanation:

Detailed Explanation: A sentinel key allows the application to check one value and refresh all related configuration only when that sentinel changes, preserving a consistent set of settings instead of refreshing 100 keys independently. Increasing the refresh/cache interval reduces how frequently the application contacts App Configuration APIs. Together these measures support dynamic refresh without application restart while controlling request volume. Registering every key separately or decreasing the cache interval would increase API traffic, and Key Vault/environment variables do not solve coordinated App Configuration refresh. Study Guide Alignment: Security and operations: Key Vault, App Configuration, managed identity, OpenTelemetry, Azure Monitor, and KQL-based troubleshooting. Official Microsoft Learn References: AI-200 Study Guide | Dynamic configuration in ASP.NET Core

NEW QUESTION 3

You need to address the known issue resulting from vector similarity queries.

Which two actions should you perform? Each correct answer presents part of the solution. Choose two. NOTE: Each correct selection is worth one point.

- A. Add a composite index on the vector fields and metadata properties of the container.
B. Reduce the numeric precision of the vector embeddings where supported.
C. Set the account consistency level to Strong.
D. Change the vector index type from flat to quantizedFlat or diskANN.

Answer: BD

Explanation:

The objective is to reduce RU consumption from vector similarity queries. Microsoft guidance identifies vector numeric precision and vector-index selection as major cost and performance levers. Using a lower supported vector precision can reduce storage and processing cost, while quantizedFlat and DiskANN are designed to reduce latency and RU consumption compared with flat search for appropriate data sizes. Strong consistency would increase resource cost rather than solve vector-search efficiency. A regular composite index is not a substitute for the vector index. Option B should be read as reducing vector numeric precision, not changing a generic ??indexing precision?? setting.

Study Guide Alignment: AI data-management workloads: Cosmos DB, PostgreSQL, caching, vector storage, vector retrieval, consistency, and connection optimization.

Official Microsoft Learn References: AI-200 Study Guide | Vector search in Azure Cosmos DB | Optimize Cosmos DB vector search performance

NEW QUESTION 4

You need to implement trace correlation according to the business requirements.

Which three 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.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Actions

⋮ Redeploy the instrumented services.

⋮ Configure a trace exporter in the OpenTelemetry SDK.

⋮ Instrument the application code by using the OpenTelemetry SDK.

⋮ Call TelemetryClient.TrackEvent() within service methods.

⋮ Implement a view in the OpenTelemetry SDK.

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Trace correlation requires the application to create OpenTelemetry spans and an exporter to send those spans to Azure Monitor. Instrumentation must therefore be present before the updated service is deployed. The previous Application Insights SDK instrumentation does not satisfy the case-study requirement that all tracing use OpenTelemetry. Configuring an OpenTelemetry view is unrelated to basic distributed-trace export, and calling TrackEvent is an Application Insights SDK pattern rather than the required OpenTelemetry approach.

Study Guide Alignment: Security and operations: Key Vault, App Configuration, managed identity, OpenTelemetry, Azure Monitor, and KQL-based troubleshooting.

Official Microsoft Learn References: AI-200 Study Guide | Enable Azure Monitor OpenTelemetry

NEW QUESTION 5

You are building a semantic search feature for a chatbot. You store document embeddings in Redis. You review the following Python code that connects to Redis and stores an embedding value:

```
01 import numpy as np
02 from redis import Redis
03 embedding = np.array([0.12, 0.98, 0.44], dtype=np.float32).tobytes()
04 r = Redis(
05     host="myredis.redis.cache.windows.net",
06     port=6380,
07     password="accesskey",
08     ssl=True
09 )
10 r.hset("doc:1", mapping={"embedding": embedding})
11 r.expire("doc:1", 600)
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Vector search fundamentals			
Statement	Yes	No	
The embedding is stored as a hash field.	☐	☐	
The code enables similarity search on the embedding field.	☐	☐	
The key will expire after 10 minutes.	☐	☐	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

``HSET`` stores the binary embedding as a field in the Redis hash keyed by ``doc:1``, so the first statement is true. The code does not create a RediSearch/Redis Query Engine vector index or define a vector field schema; merely storing bytes does not enable similarity search, so the second statement is false. ``EXPIRE doc:1 600`` sets a 600-second lifetime, which is ten minutes, making the third statement true.
Study Guide Alignment: AI data-management workloads: Cosmos DB, PostgreSQL, caching, vector storage, vector retrieval, consistency, and connection optimization.
Official Microsoft Learn References: AI-200 Study Guide | Vector search in Azure Managed Redis

NEW QUESTION 6

You optimize an AI inference API that uses Redis caching. You must reduce the risk of serving outdated data while minimizing cache management overhead. You need to implement the caching strategy that satisfies the requirements. What should you do?

- A. Reset expiration on each read.
- B. Disable expiration for cache keys.
- C. Set eviction policy to allkeys-lru.
- D. Trigger invalidation when source data changes

Answer: D

Explanation:

If the priority is to minimize stale results, source-driven invalidation is the direct control: when the underlying record changes, delete the corresponding cache entry so the next request repopulates it from current data. Sliding expiration can keep a frequently accessed stale item alive, disabling expiration is worse, and an allkeys-LRU policy evicts according to memory pressure rather than data freshness. Invalidation therefore best addresses correctness without requiring constant cache-wide maintenance.
Study Guide Alignment: AI data-management workloads: Cosmos DB, PostgreSQL, caching, vector storage, vector retrieval, consistency, and connection optimization.
Official Microsoft Learn References: AI-200 Study Guide | Azure Managed Redis documentation

NEW QUESTION 7

A Python API running in ACA must send distributed traces to Azure Monitor. The API creates spans. However, no traces appear in Azure Monitor. You need to configure the OpenTelemetry SDK pipeline to export traces to Azure Monitor. What should you do? To answer, move the appropriate actions to the correct requirements. You may use each action 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.

Actions

Enable log sampling.

Call `tracer.start_as_current_span()`

Initialize the application's TracerProvider for tracing.

Configure a span processor to send spans to the exporter.

Create the Azure Monitor component that sends trace data.

OpenTelemetry configuration mapping

Requirement

Register a global tracer provider.

Export traces to Azure Monitor.

Connect the exporter to the provider.

Generate spans in the application code.

Action

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

OpenTelemetry tracing is a pipeline: the application obtains a tracer from a configured provider, creates spans, passes ended spans through a span processor, and exports them using the Azure Monitor exporter. Creating spans alone is insufficient; without the provider/processor/exporter chain, no telemetry reaches Azure Monitor. Log sampling and legacy Application Insights TrackEvent calls are unrelated to the required OpenTelemetry trace-export path.
Study Guide Alignment: Security and operations: Key Vault, App Configuration, managed identity, OpenTelemetry, Azure Monitor, and KQL-based troubleshooting.
Official Microsoft Learn References: AI-200 Study Guide | Enable Azure Monitor OpenTelemetry

NEW QUESTION 8

You need to redesign the configuration approach so that:
configuration supports dynamic refresh;
secrets are stored outside App Configuration;
the application can securely access both App Configuration and Key Vault.
You need to enable secure dynamic configuration management for the platform. Which three actions should you perform? Each correct answer presents part of the solution. Choose three.
NOTE: Each correct selection is worth one point.

- A. Store configuration values in environment variables.
- B. Use a service principal secret for both App Configuration and Key Vault access.
- C. Store API keys in Key Vault.
- D. Allow configuration updates with a polling interval.
- E. Use managed identity for both App Configuration and Key Vault access.

Answer: CDE

Passing Certification Exams Made Easy

visit - https://www.2PassEasy.com

Explanation:

Explanation
The original answer omitted the secret-storage requirement. API keys and other secrets should be held in Key Vault, while App Configuration should retain non-secret configuration and Key Vault references. Dynamic refresh is implemented through the App Configuration refresh/polling mechanism rather than requiring an application restart. Managed identity provides credential-free access to both services in Azure. Environment variables do not provide centralized dynamic refresh, and a stored service-principal secret creates exactly the credential-management problem the design is intended to eliminate.
Study Guide Alignment: Security and operations: Key Vault, App Configuration, managed identity, OpenTelemetry, Azure Monitor, and KQL-based troubleshooting.
Official Microsoft Learn References: AI-200 Study Guide | Dynamic configuration in ASP.NET Core | Use Key Vault references in App Configuration | Managed identities for Azure resources

NEW QUESTION 9

You are developing an AI search API that caches semantic search results in Redis.
Search results must remain cached for 10 minutes. If the underlying data changes, cached entries must NOT be returned.
You need to implement a cache aside strategy to ensure data consistency.
Which two actions should you perform? Each correct answer presents part of the solution. Choose two
NOTE: Each correct selection is worth one point.

- A. Configure a 10-minute Time to Live on each key.
- B. Implement sliding expiration based on key access.
- C. Configure a cache notification for key space events.
- D. Delete related cache keys when the source data changes

Answer: AD

Explanation:

The cache-aside design needs both bounded lifetime and explicit invalidation. A ten-minute TTL ensures cached search results do not persist indefinitely, while deleting related keys when source data changes prevents known-stale entries from being returned during that ten-minute window. Sliding expiration would extend lifetime based on access and can preserve stale data. Keyspace notifications report Redis key events; they do not replace application logic that invalidates entries when the authoritative source changes.
Study Guide Alignment: AI data-management workloads: Cosmos DB, PostgreSQL, caching, vector storage, vector retrieval, consistency, and connection optimization.
Official Microsoft Learn References: AI-200 Study Guide | Azure Managed Redis documentation

NEW QUESTION 10

You are developing a solution that uses several Azure Service Bus queues. You create an Azure Event Grid subscription for the Azure Service Bus namespace.
You use Azure Functions as subscribers to process the messages.
You need to ensure that events can be sent to Azure Event Grid from the queues. The solution must use the principle of least privilege and minimize costs.
Which Azure Service Bus role and tier level should you use? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

Configuration	Value
Tier level	Basic Standard Premium
Role	Azure Service Bus Data Receiver Azure Service Bus Data Sender Azure Service Bus Data Owner

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Microsoft documents Event Grid integration for Service Bus queues and topics as requiring a Service Bus Premium namespace. The Azure Function subscriber must then receive messages from the queue; the least-privilege built-in data-plane role for that purpose is Azure Service Bus Data Receiver. Data Sender does not permit receiving, and Data Owner is broader than necessary. Premium therefore satisfies the feature prerequisite, while Data Receiver satisfies the stated least-privilege requirement.
Study Guide Alignment: Azure service integration: Service Bus, Event Grid, Azure Functions triggers /bindings, and event-driven processing.
Official Microsoft Learn References: AI-200 Study Guide | Service Bus to Event Grid integration | Authenticate Service Bus with Microsoft Entra ID

NEW QUESTION 10

You are developing several microservices to run on Azure Container Apps. External HTTP ingress traffic has been enabled for the Microservices.
A deployed microservice must be updated to allow users to test new features. You have the following requirements:
* Enable and maintain a single URL for the updated microservice to provide to test users.
* Update the microservice that corresponds to the current microservice version. You need to configure Azure Container Apps.
Which features should you configure? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Requirement:
Single URL for test users

Feature

Revision Label

Revision mode

Container image

Container registry

Revision Label

Revision mode

Container image

Container registry

Current microservice activation

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Azure Container Apps revision labels provide a stable label-specific URL that can be reassigned to a different revision, which is appropriate for a consistent test endpoint. Revision mode controls whether one or multiple revisions can be active at the same time and therefore governs how the current and updated application versions are activated. The container image and registry identify what is deployed, but they do not provide the stable test URL or the application-level revision activation behavior requested. Study Guide Alignment: Containerized Azure workloads: registry builds, App Service containers, Container Apps revision/scaling behavior, and AKS deployment choices. Official Microsoft Learn References: AI-200 Study Guide | Azure Container Apps revisions | Azure Container Apps revision labels

NEW QUESTION 11

You deploy a new revision of an app in Azure Container Apps. You need to gradually shift production traffic to the revision while monitoring performance. In addition, you need to be able to quickly roll back. Which two actions should you perform? Each correct answer presents part of the solution. Choose two. NOTE: Each correct selection is worth one point.

- A. Use traffic splitting
- B. Increase replica count.
- C. Restart the revision.
- D. Use single revision mode
- E. Enable multiple revision mode.

Answer: AE

Explanation:

Gradual production rollout requires more than one revision to remain active and a mechanism to divide traffic between them. Multiple revision mode permits concurrent active revisions, and traffic splitting lets a specified percentage of requests flow to the new revision while the remainder stays on the proven revision. This also enables rapid rollback by returning traffic to the prior revision. Increasing replicas or restarting a revision changes capacity or process state, not release traffic distribution. Study Guide Alignment: Containerized Azure workloads: registry builds, App Service containers, Container Apps revision/scaling behavior, and AKS deployment choices. Official Microsoft Learn References: AI-200 Study Guide | Azure Container Apps revisions

NEW QUESTION 13

A semantic search application queries Azure Database for PostgreSQL and stores document embeddings and metadata in a table with the following columns:

- embedding (pgvector)
- department
- created_at

The application must return the top five most similar documents for a given query embedding only from the finance department. You need to implement semantic retrieval with metadata filtering. Which query components should you select? To answer, move the appropriate query components to the correct requirements. You may use each query component once, more than once, or not at all. You may need to move the split bar between panes or scroll to view content.

Query components

WHERE department = 'Finance'

WHERE department LIKE 'dep%'

ORDER BY created_at DESC LIMIT 5

ORDER BY embedding <=> query_embedding LIMIT 5

Vector similarity search with metadata filtering

Requirement

Filter rows to finance.

Rank by similarity and return the top five.

Query component

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The metadata predicate must restrict the candidate rows to the Finance department, so the `WHERE department = 'finance'` component supplies the required filter. The `pgvector cosine-distance` operator `<=>` orders rows by vector distance to the supplied query embedding, and `LIMIT 5` keeps only the five nearest matches. Ordering by creation date would rank recency rather than semantic similarity, and a wildcard department filter would not satisfy the Finance-only requirement.

Study Guide Alignment: AI data-management workloads: Cosmos DB, PostgreSQL, caching, vector storage, vector retrieval, consistency, and connection optimization.

Official Microsoft Learn References: AI-200 Study Guide | Vector similarity search with Azure PostgreSQL

NEW QUESTION 14

A container in an AKS cluster repeatedly restarts.

Pod events show probe failures, although node-level CPU and memory metrics are normal.

You need to diagnose the cause of the repeating restarts.

What should you do first?

- A. Scale the deployment to more replicas.
- B. Decrease the `initialDelaySeconds` for the container liveness probe.
- C. Drain and reboot the node hosting the pod.
- D. Inspect the pod events and container logs.

Answer: D

NEW QUESTION 16

Drag and Drop

You are developing a new page for a website that uses Azure Cosmos DB for data storage. The feature uses documents that have the following format:

```
{
  "name": "John",
  "city": "Seattle"
}
```

You must display data for the new page in a specific order. You create the following query for the page:

```
SELECT *
FROM People p
ORDER BY p.name, p.city DESC
```

You need to configure an Azure Cosmos DB policy to support the query.

How should you configure the policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

JSON segments	Answer Area
orderBy	<pre>{ "automatic": true, "ngMode": "Consistent", "includedPaths": [{ "path": "/" }], "excludedPaths": [], "JSON segment": [{ "path": "/name", "order": "descending" }, { "path": "/city", "order": "JSON segment" }] }</pre>
sortOrder	
ascending	
descending	
compositeIndexes	

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: compositeIndexes

Azure Cosmos DB requires a defined composite index for any query utilizing an ORDER BY clause with multiple properties.

Box 2: descending

In an ORDER BY clause, properties without an explicit direction default to ascending order. Your query ORDER BY p.name, p.city DESC translates to an ascending sort on /name and a descending sort on /city.

Reference:

<https://docs.azure.cn/en-us/cosmos-db/index-policy>

NEW QUESTION 20

You have an Event Grid subscription that triggers an Azure Function.

You need to prevent loss of events in case the endpoint returns an HTTP 400 response.

Which action should you perform?

- A. Implement optimistic batchin
B. Implement asynchronous handshake validatio
C. Configure a dead-letter destinatio
D. Configure a retry policy.

Answer: C

Explanation:

Configuring a dead-letter destination is exactly the right step to take.

By default, when an endpoint returns an HTTP 400 (Bad Request) or HTTP 413 (Payload Too Large) response, Azure Event Grid immediately stops delivery attempts and drops the message. It treats these specific errors as non-transient, client-side issues, meaning it skips its standard 24-hour retry policy.

Configuring a dead-letter destination ensures that these dropped events are safely preserved for future troubleshooting and reprocessing.

Reference:

<https://turbo360.com/blog/azure-event-grid-dead-letter-monitoring>

NEW QUESTION 22

You deploy an AI application across multiple Azure regions.

The application must be able to view writes across all regions within a predictable time window.

You need to determine the appropriate consistency level.

What are two consistency levels you can use to achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Session
B. Strong

C. Bounded staleness
D. Eventual

Answer: BC

Explanation:

The two appropriate consistency levels for a globally distributed AI application needing to view writes within a predictable time window are Bounded Staleness and Strong.

Bounded Staleness: This provides a guarantee that reads will lag behind writes by a user-specified "staleness window" (either in time or number of operations). It delivers high availability and predictable read consistency across all regions, balancing performance with strict read-order guarantees.

Strong: This provides linearizable and absolute consistency, ensuring that writes are seen synchronously across all regions. It features an exact zero-second predictable time window, though it comes at the expense of higher latency for multi-region operations due to the time required to replicate globally.

Reference:

<https://learn.microsoft.com/en-us/azure/cosmos-db/consistency-levels>

NEW QUESTION 27

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual AI-200 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the AI-200 Product From:

<https://www.2passeasy.com/dumps/AI-200/>

Money Back Guarantee

AI-200 Practice Exam Features:

- * AI-200 Questions and Answers Updated Frequently
- * AI-200 Practice Questions Verified by Expert Senior Certified Staff
- * AI-200 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * AI-200 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year