

KTest

高品質 より良いサービス



練習問題

<https://www.ktest.jp>

1年間無料アップデート

Exam : HPE0-S62

Title : HPE Compute Architect

Version : DEMO

1.A manufacturer has validated an LLM for explaining service instructions. The validated LLM must be retained, and model training or parameter updates are outside the approved scope of this phase. Its internal manuals change weekly, and the application team can integrate a service endpoint but wants to minimize ongoing ownership of a custom retrieval backend. The customer has confirmed that the Knowledge Bases capability described in HPE's April 15, 2025 article is available in its proposed deployment.

Which proposal best fits this next project phase?

- A. An HPE ProLiant GPU-server solution with an application team maintaining a custom RAG stack
- B. An HPE Private Cloud AI proposal using Knowledge Bases in HPE AI Essentials Software
- C. An HPE Private Cloud AI proposal with recurring fine-tuning jobs to incorporate each manual revision
- D. An HPE ProLiant GPU-server solution running a standalone foundation-model inference service

Answer: B

Explanation:

- B: The proposed managed retrieval capability fits both the changing knowledge and the team's preferred integration boundary.

2.A retailer is replacing an HPE compute platform to improve checkout responsiveness. The agreed business target is a 95th-percentile response time below two seconds while processing 400 transactions per second during the evening peak. The previous platform passes its CPU benchmark but exceeds the response-time target at that load.

Which acceptance test provides the most relevant evidence that the replacement achieves the requested outcome?

- A. Compare the maximum CPU benchmark scores of the old and new hosts.
- B. Run representative checkout transactions at the target peak load and measure the response-time distribution.
- C. Measure the time needed to restart an idle application after a host failure.
- D. Measure average checkout response time during the lowest-traffic hour.

Answer: B

Explanation:

- B: This test reproduces the agreed load and measures the percentile response-time outcome.

3.A customer will deploy a RAG assistant for internal engineering documents. Its approved policy requires the documents, embeddings, prompts, and generated responses to be stored and processed at the customer's facility. Outbound platform management traffic is permitted if it contains none of that content. The customer can maintain the document collection, but has no team available to operate the production RAG runtime. It requires a contracted provider to perform that runtime operation. The following proposals have all passed the same functional, capacity, and supportability validation. All table commitments and costs are stipulated facts for this scenario.

Which proposal meets both requirements at the lowest three-year cost?

Proposal	AI deployment	Location of documents, embeddings, prompts, and responses	Production RAG runtime operator	Three-year cost units
A	Provider-hosted RAG on HPE GPU servers	Provider's facility	Contracted provider	80
B	Customer-installed RAG on HPE GPU servers	Customer's facility	Customer's team	90
C	HPE Private Cloud AI plus a separately scoped HPE AI Services operating contract	Customer's facility	Contracted provider	120
D	Custom RAG stack on HPE GPU servers plus a separately scoped provider operating contract	Customer's facility	Contracted provider	150

For C, the local-content commitment and application operating duties are explicitly included in the proposal. They are not inferred from the term "cloud-managed" or from standard PCAI support. Cost units are fictional and comparable across these complete proposals.

- A. Proposal B
- B. Proposal C
- C. Proposal A
- D. Proposal D

Answer: B

Explanation:

- B: Proposal C explicitly commits to local AI content and contracted runtime operation. Its three-year cost of 120 units is the lowest among proposals that meet both requirements. The scoped commitments establish that the proposed service meets the customer's content-location and operational requirements.

4.A customer is selecting an HPE ProLiant host for an application licensed per populated physical CPU socket. The same annual charge applies to every populated socket, regardless of core count. Four vendor-validated configurations below all satisfy the required application performance, memory, support and availability conditions.

Which configuration minimises the annual application licence charge?

- A. One populated socket with a 64-core processor.
- B. Two populated sockets with 24-core processors.
- C. Two populated sockets with 32-core processors.
- D. Four populated sockets with 16-core processors.

Answer: A

Explanation:

- A: The proposed one-socket configuration is already validated for the required workload. Under the stipulated per-populated-socket licence metric, one populated socket incurs one annual charge, giving the lowest qualifying licence cost.

5. A team is following HPE's May 6, 2026 natural-language-to-SQL tutorial with HPE AI Essentials 1.12+ and OpenWebUI 0.6.31+. The database is populated, its registered connection is healthy, and the team's authorized user can access the data. The model endpoint and chat integration also pass their tests. Before approving the assistant's manufacturing-data mapping, the team needs evidence about which tables exist in the connected schema and what their records contain.

Which next platform action most directly supplies that evidence?

- A. Register an additional database connection using the credentials from the successful connection test.
- B. Inspect the packaged model's tokenizer and runtime metadata in the model-management workflow.
- C. Generate another model-endpoint API token and repeat the completed chat connection test.
- D. Explore the connected source in Data Catalog and use Data Preview on the relevant tables.

Answer: D

Explanation:

- D: This uses the existing data connection to discover the tables and inspect their contents before approving the assistant's mapping.