HPE7-S01 Dumps

Shared by Lee on 11-11-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A prospective HPE Private Cloud AI customer has data stored across several locations, including a Cassandra database, a Hive data warehouse, and a data lakehouse on S3 object storage.

Which statement correctly describes HPE Private Cloud AI's capability to query data across these sources?

Options:

A- It can run queries on each of these data sources, but it can only query one data source at a time.
B- It can run queries on the data lakehouse but not on the Cassandra database and Hive data warehouse.
C- It can run queries across all three data sources.
D- It can run queries on the Cassandra database and the Hive data warehouse but not on the data lakehouse.

Answer:

C

Explanation:

HPE Private Cloud AI incorporates HPE AI Essentials and its EzPresto SQL query engine to provide federated data access across heterogeneous enterprise data platforms.

EzPresto is based on PrestoDB and supports distributed SQL execution against multiple connected data sources without requiring organizations to consolidate all information into a single database.

HPE AI Essentials includes connector capabilities for Cassandra, Hive, and supported S3-based data sources.

Cassandra provides distributed NoSQL data storage, while Hive supports SQL-based analytics over large datasets. S3-compatible object storage can contain data lakehouse tables and supported file formats accessed through appropriate connectors.

Once administrators configure the necessary connections, users can discover accessible datasets and execute SQL queries across supported data sources.

Federated querying enables analytical operations, including joins and data transformations involving tables from different underlying systems.

HPE documentation also describes the Data Lakehouse Gateway, which provides access to supported Apache Iceberg tables stored in compatible S3 and NFS environments.

Option A is incorrect because EzPresto supports federated queries across multiple data sources. Option B incorrectly excludes Cassandra and Hive connectivity. Option D incorrectly excludes supported S3-based data lakehouse access.

Option C correctly identifies the platform's ability to query information across all three source types.

Official HPE Reference: HPE AI Essentials Software --- Data Engineering and EzPresto Federated Queries; Connecting Cassandra; Hive Connection Parameters; Connecting to External S3 Data Sources.

P2P exams

Question 2

Question Type: MultipleChoice

An HPE Private Cloud AI customer is interested in fine-tuning using Ray and asks how to submit jobs to the Ray cluster running on HPE AI Essentials.

What should you explain?

Options:

A- They can use the HPE AI Essentials API Ray endpoint.
B- They can submit jobs directly from the Ray dashboard, which opens from the HPE AI Essentials Tools & Frameworks page.
C- They can install the Ray Tune framework from the administration settings.
D- They can run a notebook server on HPE AI Essentials and use its Ray kernel.

P2P exams

Answer:

D

Explanation:

HPE AI Essentials provides an integrated notebook development environment that supports distributed AI and machine learning workloads using Ray.

The platform includes supported Jupyter notebook images with Ray libraries and an associated Ray kernel. This enables data scientists to develop Python applications, connect to the Ray

cluster, and submit distributed workloads without separately installing an entire Ray environment.

For model fine-tuning, a user can create a notebook server, activate the Ray-specific kernel, configure the required runtime dependencies, and connect to the Ray cluster running within HPE AI Essentials.

HPE's documentation recommends using Ray's JobSubmissionClient for submitting jobs. The client communicates with the Ray head service through its job-submission interface, allowing users to specify entry points, runtime environments, and resource requirements.

The Ray dashboard is primarily documented as an operational interface for inspecting resource consumption, tasks, logs, execution status, and Ray Tune experiments.

Option A does not identify the documented HPE-supported submission procedure clearly enough. Option B incorrectly presents the dashboard as the primary job-submission mechanism. Option C is unnecessary because the platform's supported Ray notebook environment includes the relevant Ray software libraries.

Option D accurately describes the integrated development environment through which users can prepare and submit Ray workloads.

Official HPE Reference: HPE AI Essentials Software 1.11.x --- Connecting to Ray Cluster; Using JobSubmissionClient to Submit Ray Jobs; Support Matrix --- Notebook Images and Ray Kernel Packages.

HPE AI Essentials Software 1.11.x Documentation

Question 3

Question Type: MultipleChoice

An HPE Private Cloud AI customer needs to archive Prometheus metrics collected on the solution. What should you recommend?

Options:

A- Set up an external OpenTelemetry (OTel) endpoint and register it in HPE AI Essentials.
B- Configure the Prometheus framework and enlarge the kubeprom-operator volume to at least 100G.
C- Install the Grafana framework on HPE Private Cloud AI and set up Prometheus as a data source.
D- Change the PV in the Prometheus namespace to use Retain mode.

Answer:

A

Explanation:

HPE AI Essentials integrates OpenTelemetry (OTel) to collect and export observability data from platform infrastructure and application workloads.

Prometheus provides metrics collection for Kubernetes components, nodes, containers, networking, storage, and supported AI applications. However, operational requirements may demand that historical metrics be retained beyond the platform's local retention period.

HPE AI Essentials addresses this requirement by supporting an external OpenTelemetry endpoint.

Administrators can configure an OTel exporter through the platform's administration settings. The exporter forwards collected telemetry to an external OTel collector, which can process and persist the information in an appropriate long-term monitoring or archival system.

HPE documents that Prometheus metrics are available through OTel, including resource metrics and application-specific measurements associated with KServe, MLflow, Ray, and other services.

Option B only increases local storage capacity and does not establish a comprehensive archival process.

Option C provides a visualization mechanism but does not automatically establish long-term metric retention.

Option D controls Kubernetes persistent-volume reclamation behavior and does not provide a reliable metrics archival strategy.

An external OpenTelemetry endpoint enables integration with enterprise observability infrastructure and supports retention policies independent of the local platform.

Official HPE Reference: HPE AI Essentials Software 1.11.x --- OpenTelemetry (OTel); HPE AI Essentials Software 2026.07.1 --- Configuring Endpoints: OTel Endpoint.

HPE AI Essentials Software 1.11.x Documentation

Question 4

Question Type: MultipleChoice

What is one key security consideration for AI workloads?

Options:

A- Redacting personally identifiable information from training datasets
B- Using a secure enclave processor to handle AI inferencing workloads
C- Deploying Istio to clients to regulate access to LLMs
D- Using custom container images for AI inferencing and training

Answer:

A

Explanation:

Protecting personally identifiable information is a fundamental security and governance requirement when developing, training, and deploying enterprise AI models.

Training datasets may contain sensitive information such as customer names, contact details, account identifiers, medical information, employee records, and confidential business data.

If this information is incorporated into training datasets without appropriate protection, it can create privacy, regulatory, and information-disclosure risks.

Redaction removes sensitive information, while masking, tokenization, and anonymization can provide alternative protection mechanisms depending on the intended use case.

HPE's guidance on responsible enterprise AI emphasizes data classification, minimization, and removing unnecessary sensitive information before it reaches AI models.

These controls should be implemented during data preparation and reinforced through access restrictions, encryption, audit logging, and appropriate retention policies.

Option B is incorrect because a secure enclave processor is not a universal requirement for AI inference workloads.

Option C is incorrect because Istio is generally deployed as part of the Kubernetes service-mesh infrastructure rather than installed directly on client applications.

Option D is incorrect because custom container images do not inherently provide stronger security. Their effectiveness depends on image provenance, vulnerability management, configuration, and runtime controls.

Redacting personally identifiable information directly addresses a major AI data-security risk.

Official HPE Reference: HPE --- Responsible AI in Enterprise Applications: A Practical Security and Governance Guide; HPE Inclusive, Responsible, Ethical Data Commitment.

Hewlett Packard Enterprise Community

Question 5

Question Type: MultipleChoice

A customer has policies that require sensitive information to be removed from AI fine-tuning data.

Which solution will help address this requirement?

Options:

A- Kubeflow Katib
B- NVIDIA NeMo Evaluator
C- Ray Tune
D- NVIDIA NeMo Curator

Answer:

D

Explanation:

NVIDIA NeMo Curator is a data preparation and curation framework designed to improve the quality, consistency, and suitability of datasets used for training and fine-tuning generative AI models.

A significant data-governance requirement is identifying and removing sensitive information before training begins. Such information may include personally identifiable information (PII), financial identifiers, confidential records, and other protected content.

NeMo Curator provides data-processing capabilities for filtering, cleansing, deduplication, and dataset preparation. Appropriate releases and workflows also support PII identification and removal, allowing organizations to sanitize sensitive training material before submitting it to model training pipelines.

Within the HPE Private Cloud AI architecture, NeMo Curator is identified as a component of the NVIDIA AI Enterprise software ecosystem.

Option A is incorrect because Kubeflow Katib automates hyperparameter optimization rather than sensitive-data removal.

Option B is incorrect because NVIDIA NeMo Evaluator assesses model quality and performance rather than serving as the primary dataset sanitization framework.

Option C is incorrect because Ray Tune focuses on distributed hyperparameter optimization.

NeMo Curator is the appropriate choice because it addresses data preparation before fine-tuning. Administrators must verify that the selected release supports the required PII processing workflow, since capabilities differ between versions.

Official HPE Reference: HPE Private Cloud AI --- Accelerate Your AI Path with a Turnkey AI Private Cloud, Figure 2, identifying NeMo Curator within the NVIDIA AI Enterprise software architecture.

Hewlett Packard Enterprise

Verification qualification: HPE confirms NeMo Curator's inclusion in the solution architecture. The HPE publication inspected does not independently specify its PII-removal implementation; the detailed capability should be checked against the exact NeMo Curator release before publishing this as fully verified.

Question 6

Question Type: MultipleChoice

A customer intends to implement machine learning models but has limited visibility into where business data is produced and how it is managed across departments.

During the readiness assessment, which step would best support an accurate evaluation of their AI deployment potential?

Options:

A- Begin by installing a standard data lake connector and mapping legacy tables to object storage.
B- Focus the session on reviewing backup retention schedules and recovery point objectives.
C- Recommend bypassing initial workshops to begin with a small AI pilot on anonymized data.
D- Identify data sources, formats, ownership, and movement patterns across the organization.

Answer:

D

Explanation:

An effective AI readiness assessment requires a comprehensive understanding of the organization's data landscape before selecting machine learning tools, infrastructure, or

deployment architectures.

The architect should begin by identifying enterprise data sources, their formats, ownership responsibilities, storage locations, access mechanisms, and movement patterns between departments.

This process establishes whether the organization possesses accessible, relevant, sufficiently reliable data to support its intended machine learning applications.

Understanding data ownership is particularly important because it identifies the stakeholders responsible for authorizing access, maintaining quality, enforcing governance policies, and determining permitted uses.

Data-flow analysis also reveals dependencies between transactional systems, warehouses, analytical platforms, and downstream applications. These dependencies affect data ingestion architecture, network capacity, workload placement, security controls, and operational requirements.

HPE's Artificial Intelligence Transformation Workshop explicitly includes identifying data sources, dependencies, and organizational readiness as fundamental planning activities.

Option A prematurely introduces technical integration before identifying the actual data landscape. Option B addresses business continuity but does not adequately assess AI data readiness. Option C bypasses essential discovery and may produce an unrepresentative pilot.

Option D establishes the information needed to assess feasibility, identify risks, and prepare accurate customer requirements for the proposed AI solution.

Official HPE Reference: HPE Artificial Intelligence Transformation Workshop --- Data Sources, Dependencies, and AI Roadmap; HPE AI Services --- Data Assessment and AI Strategy.

HPE EUROPE

Question 7

Question Type: MultipleChoice

Which factors should be considered when determining the GPU memory needed to perform inference with an AI model?

Options:

A- The number of parameters in the model and the floating point precision

B- The memory available on the host and the number of convolutional layers in the model

C- The memory available on the host and the number of parameters in the model
D- The floating point precision and the number of layers in the model

Answer:

A

Explanation:

The two principal factors determining the baseline GPU memory requirement for model inference are the number of model parameters and the numerical precision used to represent those parameters.

Each parameter occupies a specified number of bytes according to its data type. For example, FP32 uses four bytes per parameter, whereas FP16 and BF16 generally use two bytes per parameter. Lower-bit quantized representations can reduce the memory required for model weights.

The approximate memory requirement for storing model parameters is calculated as:

\[\text{Model memory} \approx \text{Parameter count}\times\text{Bytes per parameter} \]

A model containing 70 billion parameters requires approximately 140 GB for its weights when represented in FP16, excluding additional runtime memory.

Inference also consumes memory for key-value caches, intermediate activations, execution buffers, and framework overhead. These requirements depend on factors such as sequence length, batch size, model architecture, and concurrent requests.

Option B incorrectly prioritizes host memory and convolutional-layer count. Option C omits the critical influence of numerical precision. Option D considers precision but excludes the total parameter count.

HPE's published AI networking technical guidance explicitly relates accelerator memory requirements to model size and numerical representation.

Official HPE Reference: HPE Networking Technical Articles --- LLM Inference: Hardware and Software Optimizations, particularly Accelerator Memory Requirements; GPU Fabrics for GenAI Workloads.

Technical articles by HPE Networking

Question 8

Question Type: MultipleChoice

You are leading a workshop to assist a customer in deploying a pilot AI application.

Which data-related task should you help the customer with?

Options:

A- Pulling data out of Retrieval Augmented Generation (RAG) databases into standard SQL databases
B- Identifying where data is stored and planning to run AI workloads nearby
C- Converting all unstructured data such as PDFs into structured data for training datasets
D- Storing data unencrypted in a single location so that it is ready for training

Answer:

B

Explanation:

Data locality is an important architectural consideration when designing AI pilots because the location of source datasets influences application latency, data-transfer costs, infrastructure utilization, and security requirements.

During the workshop, the architect should identify the repositories containing the customer's relevant datasets, determine their physical or cloud locations, and establish how AI applications will access them.

Where practical, deploying AI compute resources close to frequently accessed data reduces unnecessary transfers across wide-area networks and can improve processing efficiency.

This is particularly beneficial for data-intensive workflows involving model training, retrieval-augmented generation, large-scale analytics, and repeated access to enterprise information.

HPE emphasizes a data-first approach to AI planning, including evaluating existing data infrastructure, understanding dependencies, and aligning AI workloads with available storage and networking capabilities.

Option A incorrectly assumes that data from RAG vector databases must be transferred into conventional SQL databases. Option C is unnecessarily restrictive because AI workflows can process unstructured information through document parsing, embedding generation, and other techniques without converting every document into relational data.

Option D directly contradicts enterprise data-protection requirements by recommending unencrypted storage.

Option B correctly prioritizes data discovery and appropriate workload placement before

infrastructure selection.

Official HPE Reference: HPE AI Services --- AI Readiness and Data Assessment; HPE Artificial Intelligence Transformation Workshop --- Data Sources and Dependencies; HPE Private Cloud AI Architecture.

HPE

Question 9

Question Type: MultipleChoice

You are configuring a solution in One Config Advanced (OCA).

What is one use case for the Customer Intent Document (CID)?

Options:

A- Specifying the units of measurement for HPE GreenLake billing tiers
B- Providing special testing instructions associated with a Factory Express Service
C- Receiving guidance for cabling layouts for integrated solutions
D- Indicating the support term for HPE Complete Care Services

Answer:

B

Explanation:

The Customer Intent Document (CID) captures customer-specific implementation requirements that cannot be adequately communicated through standard product selections or configuration part numbers. Within HPE One Config Advanced (OCA), it supports the translation of the proposed solution into precise instructions for factory integration and customization.

HPE Factory Express services use customer-intent information to establish requirements for system configuration, software image installation, firmware settings, verification, and acceptance testing. When a customer requires specialized testing, those instructions must be documented so the factory team can perform the agreed activities before shipment.

For example, a customer may require customized operating system images, prescribed firmware revisions, particular RAID configurations, or additional acceptance tests. The CID helps identify these requirements and supports coordination between the customer, HPE sales, configuration

specialists, and factory personnel.

Option A concerns commercial consumption measurements, which are not the purpose of the CID. Option C describes cabling guidance rather than the specific factory-testing use case. Cabling requirements may appear in integration documentation, but that does not make C the best answer. Option D relates to support contract selection.

HPE Factory Express documentation identifies the CID as the mechanism for conveying customer-specific configuration and service instructions.

Official HPE Reference: HPE Factory Express Image Load and Software Management Services --- System Verification and Acceptance Testing; HPE Factory Express Product Customization Services --- Customer Intent and Ordering Requirements.

Question 10

Question Type: MultipleChoice

You are developing a proposal for HPE ProLiant DL servers to serve as VMware ESXi hosts.

How does a customer requirement to support Fault Tolerance (FT) influence your plan?

Options:

A- You must ensure that the servers have sufficient memory to reserve 20% for FT.
B- You must ensure that the servers' storage controllers are compatible with SMP-FT.
C- You must ensure that the servers' adapters are RoCEv2 capable.
D- You must ensure that the servers' processors are compatible with SMP-FT.

Answer:

D

Explanation:

VMware vSphere Fault Tolerance provides continuous availability for protected virtual machines by maintaining synchronized execution between a primary virtual machine and a secondary virtual machine running on separate ESXi hosts.

If the primary host fails, the secondary virtual machine can continue operating with minimal interruption to the protected workload.

Modern VMware Fault Tolerance uses Symmetric Multiprocessing Fault Tolerance (SMP-FT), which

supports eligible virtual machines with multiple virtual CPUs.

Consequently, processor compatibility is an essential consideration when selecting HPE ProLiant DL servers for an ESXi environment that requires Fault Tolerance.

The architect must confirm that the selected processors, server models, firmware configurations, and ESXi releases satisfy the applicable compatibility requirements.

HPE publishes VMware ESXi support matrices identifying supported ProLiant platforms, processor families, and minimum software releases. These matrices establish the server-level compatibility baseline, while SMP-FT must also be validated for the specific CPU architecture.

Option A is incorrect because VMware FT does not impose a universal requirement to reserve exactly 20% of server memory.

Option B incorrectly associates SMP-FT compatibility with storage controllers. Option C incorrectly makes RoCEv2 networking a prerequisite.

Option D identifies the critical processor compatibility requirement that must be considered during the server selection and validation process.

Official HPE Reference: HPE VMware ESXi Support Matrix --- ProLiant DL Platforms, Processor Support, and Minimum ESXi Versions.

Verification qualification: The HPE compatibility matrix supports platform and processor-version selection. It does not independently establish SMP-FT support for every specific CPU SKU; that final capability check remains necessary.

HPE7-S01 --- Advanced HPE Compute Architect

Batch 6: Questions 51--59

The uploaded Word document contains 59 questions, so this is the final batch. The questions and answer options below retain the original wording, with the A--D formatting standardized and no bullet markers.

Question 11

Question Type: MultipleChoice

A customer is evaluating HPE Private Cloud AI's Data Lakehouse Gateway and needs help deciding whether the gateway will work with their data lakehouse.

What is one requirement that should be explained?

Options:

A- The data lakehouse must use ARVO files.
B- The data lakehouse must provide native integration with EzPresto.
C- The data lakehouse must be hosted in AWS.
D- The data lakehouse must be deployed on S3 or NFSv4 storage.

Answer:

D

Explanation:

HPE Private Cloud AI Data Lakehouse Gateway provides unified access to enterprise data stored across supported data lakehouse environments.

The gateway uses a global namespace to make distributed datasets accessible through compatible analytical and AI development tools.

According to HPE AI Essentials documentation, Data Lakehouse Gateway supports connections to AWS S3, S3-compatible object storage, and NFSv4 storage.

Therefore, the customer's data lakehouse must use one of these supported storage interfaces to establish a connection through the gateway.

Once configured, administrators can register supported tables, manage access permissions, and expose data to analytical applications through the platform's integrated services.

The gateway incorporates an EzPresto Iceberg catalog, enabling SQL-based access to supported Apache Iceberg tables. However, the external data lakehouse does not need to provide its own native EzPresto implementation.

Option A is incorrect because HPE documentation explicitly identifies AVRO files as unsupported by the gateway.

Option B incorrectly assumes that EzPresto must be integrated into the external data lakehouse.

Option C is incorrect because the gateway supports compatible on-premises and third-party S3 implementations in addition to AWS.

Option D accurately identifies the supported storage connectivity requirements. Administrators must additionally validate network accessibility, credentials, permissions, and table-format compatibility.

Official HPE Reference: HPE AI Essentials Software --- Data Lakehouse Gateway; Connecting Data Lakehouses --- Supported Vendors and Connection Requirements.

HPE AI Essentials Software 1.9.x Documentation

Question 12

Question Type: MultipleChoice

A customer is assessing whether to run their AI proof of concept (PoC) in the cloud or on-premises. They store sensitive data and need low-latency performance.

Which factor supports running the PoC in the data center?

Options:

A- Desire to minimize staff training on container platforms
B- Latency considerations for sensitive data stored locally
C- Need to procure the resources as quickly as possible
D- Preference for short-term licensing agreements

Answer:

B

Explanation:

Deploying AI workloads close to sensitive enterprise data can improve application performance while reducing unnecessary data movement and strengthening organizational control over information.

When an AI proof of concept processes data stored in an on-premises data center, running the computing infrastructure locally reduces reliance on external network connections and can minimize data-transfer latency.

This is especially important for latency-sensitive inference applications, real-time analytics, and AI systems that repeatedly access large datasets during processing.

Local deployment can also reduce the requirement to transfer confidential information into public cloud environments. This helps organizations maintain control over data residency, access policies, and security configurations.

HPE Private Cloud AI provides an integrated AI infrastructure that can operate within a customer's data center, combining compute, storage, networking, and AI software capabilities.

HPE explicitly positions the solution for organizations requiring data sovereignty, predictable performance, and low-latency processing near enterprise datasets.

Option A does not establish a clear on-premises advantage because container orchestration

knowledge may still be necessary. Option C generally favors cloud-based resource provisioning when rapid infrastructure availability is the overriding requirement. Option D concerns commercial flexibility rather than the stated security and performance requirements.

Option B directly addresses the customer's need for local data processing and low-latency AI performance.

Official HPE Reference: HPE Private Cloud AI --- Optimized AI Performance and Economics; Overview of the HPE Private Cloud AI Engineered System; Make Your Data AI-Ready.

HPE

To Get Premium Files for HPE7-S01 Visit
https://www.p2pexams.com/products/hpe7-s01

For More Free Questions Visit
https://www.p2pexams.com/hp/pdf/hpe7-s01

