Download NVIDIA NCA-AIIO Exam Dumps Free

Shared by Sandoval on 17-06-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

Which is the best PUE value for a data center?

Options:

A- PUE of 1.2
B- PUE of 3.5
C- PUE of 5.0
D- PUE of 2.0

Answer:

A

Explanation:

Power Usage Effectiveness (PUE) measures data center efficiency, with an ideal value of 1.0 (all power used by IT equipment). A PUE of 1.2, indicating only 20% overhead, is highly efficient and closer to the ideal than 2.0 (100% overhead), 3.5, or 5.0, making it the best among the options for energy-conscious AI deployments.

(Reference: NVIDIA AI Infrastructure and Operations Study Guide, Section on Data Center Efficiency)

Question 2

Question Type: MultipleChoice

A company is implementing a new network architecture and needs to consider the requirements and considerations for training and inference. Which of the following statements is true about training and inference architecture?

Options:

A- Training architecture and inference architecture have the same requirements and considerations.

B- Training architecture is only concerned with hardware requirements, while inference architecture is only concerned with software requirements.
C- Training architecture is focused on optimizing performance while inference architecture is focused on reducing latency.
D- Training architecture and inference architecture cannot be the same.

Answer:

C

Explanation:

Training architectures are designed to maximize computational throughput and accelerate model convergence, often by leveraging distributed systems with multiple GPUs or specialized accelerators to process large datasets efficiently. This focus on performance ensures that models can be trained quickly and effectively. In contrast, inference architectures prioritize minimizing response latency to deliver real-time or near-real-time predictions, frequently employing techniques such as model optimization (e.g., pruning, quantization), batching strategies, and deployment on edge devices or optimized servers. These differing priorities mean that while there may be some overlap, the architectures are tailored to their specific goals---performance for training and low latency for inference.

(Reference: NVIDIA AI Infrastructure and Operations Study Guide, Section on Infrastructure Considerations for AI Workloads; NVIDIA Documentation on Training and Inference Optimization)

Question 3

Question Type: MultipleChoice

Which type of GPU core was specifically designed to realistically simulate the lighting of a scene?

Options:

A- Tensor Cores
B- CUDA Cores
C- Ray Tracing Cores

Answer:

C

Explanation:

Ray Tracing Cores, introduced in NVIDIA's RTX architecture, are specialized hardware units built to accelerate ray-tracing computations---simulating light interactions (e.g., reflections, shadows) for photorealistic rendering in real time. CUDA Cores handle general-purpose parallel tasks, and Tensor Cores optimize matrix operations for AI, but only Ray Tracing Cores target lighting simulation.

(Reference: NVIDIA GPU Architecture Whitepaper, Section on Ray Tracing Cores)

Question 4

Question Type: MultipleChoice

What factors have led to significant breakthroughs in Deep Learning?

Options:

A- Advances in hardware, availability of fast internet connections, and improvements in training algorithms.
B- Advances in sensors, availability of large datasets, and improvements to the ''Bag of Words'' algorithm.
C- Advances in hardware, availability of large datasets, and improvements in training algorithms.
D- Advances in smartphones, social media sites, and improvements in statistical techniques.

Answer:

C

Explanation:

Deep learning breakthroughs stem from three pillars: advances in hardware (e.g., GPUs and TPUs) providing the compute power for large-scale neural networks; the availability of large datasets offering the data volume needed for training; and improvements in training algorithms (e.g., optimizers like Adam, novel architectures like Transformers) enhancing model efficiency and accuracy. While internet speed, sensors, or smartphones play roles in broader tech, they're less directly tied to deep learning's core advancements.

(Reference: NVIDIA AI Infrastructure and Operations Study Guide, Section on Deep Learning Advancements)