

NVIDIA

Exam Questions NCA-AIO

NVIDIA-Certified Associate AI Infrastructure and Operations



NEW QUESTION 1

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

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

Answer: C

NEW QUESTION 2

In training and inference architecture requirements, what is the main difference between training and inference?

- A. Training requires real-time processing, while inference requires large amounts of data.
- B. Training requires large amounts of data, while inference requires real-time processing.
- C. Training and inference both require large amounts of data.
- D. Training and inference both require real-time processing.

Answer: B

NEW QUESTION 3

An IT professional is considering whether to implement an on-prem or cloud infrastructure. Which of the following is a key advantage of on-prem infrastructure?

- A. Lower upfront costs and capital expenditure.
- B. Scalability and flexibility.
- C. Ensure data security and sovereignty.
- D. Easy remote management.

Answer: C

NEW QUESTION 4

How is the architecture different in a GPU versus a CPU?

- A. A GPU acts as a PCIe controller to maximize bandwidth.
- B. A GPU is architected to support massively parallel execution of simple instructions.
- C. A GPU is a single large and complex core to support massive compute operations.

Answer: B

NEW QUESTION 5

Which phase of deep learning benefits the greatest from a multi-node architecture?

- A. Data Augmentation
- B. Training
- C. Inference

Answer: B

NEW QUESTION 6

When deploying high-density workloads in a data center, what are the three main resource constraints that need to be considered?

- A. Processing speed, storage capacity, and network connectivity.
- B. Power, cooling, and physical space.
- C. Bandwidth, security, and redundancy.

Answer: B

NEW QUESTION 7

Which is the best PUE value for a data center?

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

Answer: A

NEW QUESTION 8

Which of the following statements is true about GPUs and CPUs?

- A. GPUs are optimized for parallel tasks, while CPUs are optimized for serial tasks.
- B. GPUs have very low bandwidth main memory while CPUs have very high bandwidth main memory.
- C. GPUs and CPUs have the same number of cores, but GPUs have higher clock speeds.
- D. GPUs and CPUs have identical architectures and can be used interchangeably.

Answer: A

NEW QUESTION 9

For which workloads is NVIDIA Merlin typically used?

- A. Recommender systems
- B. Natural language processing
- C. Data analytics

Answer: A

NEW QUESTION 10

In an AI cluster, what is the purpose of job scheduling?

- A. To gather and analyze cluster data on a regular schedule.
- B. To monitor and troubleshoot cluster performance.
- C. To assign workloads to available compute resources.
- D. To install, update, and configure cluster software.

Answer: C

NEW QUESTION 10

How many 1 Gb Ethernet in-band network connections are in a DGX H100 system?

- A. 1
- B. 2

Answer: C

NEW QUESTION 11

What is a significant benefit of using containers in an AI development environment?

- A. They increase the base accuracy of AI models by optimizing their algorithms.
- B. They ensure that AI applications run consistently across different computing environments.
- C. They can automatically generate AI datasets for machine learning model training.
- D. They directly increase the processing speed of GPUs used in AI computations.

Answer: B

NEW QUESTION 16

Which are three key features of InfiniBand networking technology?

- A. High reliability, high latency, and CPU offloads.
- B. High latency, high reliability, and high bandwidth.
- C. GPU offloads, low latency, high reliability.
- D. Low latency, high bandwidth, and CPU offloads.

Answer: D

NEW QUESTION 20

A customer is evaluating an AI cluster for training and is questioning why they should use a large number of nodes. Why would multi-node training be advantageous?

- A. The model is too large to fit into GPU memory.
- B. The model is being used by a large number of users.
- C. The model is being used for large-scale inference workloads.

Answer: A

NEW QUESTION 23

Which GPUs should be used when training a neural network for self-driving cars?

- A. NVIDIA H100 GPUs
- B. NVIDIA L4 GPUs
- C. NVIDIA DRIVE Orin

Answer: A

NEW QUESTION 28

What is the maximum number of MIG instances that an H100 GPU provides?

- A. 7
- B. 8
- C. 4

Answer: A

NEW QUESTION 33

What is a key value of using NVIDIA NIMs?

- A. They provide fast and simple deployment of AI models.
- B. They have community support.
- C. They allow the deployment of NVIDIA SDKs.

Answer: A

NEW QUESTION 36

Which aspect of computing uses large amounts of data to train complex neural networks?

- A. Machine learning
- B. Deep learning
- C. Inferencing

Answer: B

NEW QUESTION 40

What is the primary command for checking the GPU utilization on a single DGX H100 system?

- A. nvidia-smi
- B. ctop
- C. nvml

Answer: A

NEW QUESTION 44

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

NCA-AIIO Practice Exam Features:

- * NCA-AIIO Questions and Answers Updated Frequently
- * NCA-AIIO Practice Questions Verified by Expert Senior Certified Staff
- * NCA-AIIO Most Realistic Questions that Guarantee you a Pass on Your First Try
- * NCA-AIIO Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The NCA-AIIO Practice Test Here](#)