NVIDIA

Exam Questions NCA-AIIO

NVIDIA-Certified Associate AI Infrastructure and Operations

NEW QUESTION 1
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?

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

NEW QUESTION 2
Which of the following aspects have led to an increase in the adoption of AI? (Choose two.)

A. Moore??s Law
B. Rule-based machine learning
C. High-powered GPUs
D. Large amounts of data

Answer: CD

NEW QUESTION 3
Which of the following statements is true about Kubernetes orchestration?

A. It is bare-metal based but it supports containers.
B. It has advanced scheduling capabilities to assign jobs to available resources.
C. It has no inferencing capabilities.
D. It does load balancing to distribute traffic across containers.

Answer: BD

NEW QUESTION 4
When training a neural network, what is the most common pattern of storage access?

A. Random write
B. Sequential read
C. Sequential write

Answer: B

NEW QUESTION 5
Which of the following statements correctly differentiates between AI, Machine Learning, and Deep Learning?

A. Machine Learning is a subset of AI, and AI is a subset of Deep Learning.
B. AI and Deep Learning are the same, while Machine Learning is a separate concept.
C. AI is a subset of Machine Learning, and Machine Learning is a subset of Deep Learning.
D. Deep Learning is a subset of Machine Learning, and Machine Learning is a subset of AI.

Answer: D

NEW QUESTION 6
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 7
What enables moving data between GPU memory and local or remote storage without using the CPU?

A. NVLink
B. GPUDirect P2P
C. InfiniBand
D. GPUDirect Storage

Answer: D

NEW QUESTION 8
When monitoring a GPU-based workload, what is GPU utilization?

A. The maximum amount of time a GPU will be used for a workload.

B. The GPU memory in use compared to available GPU memory.
C. The percentage of time the GPU is actively processing data.
D. The number of GPU cores available to the workload.

Answer: C

NEW QUESTION 9
Which NVIDIA parallel computing platform and programming model allows developers to program in popular languages and express parallelism through extensions?

A. CUDA
B. CUML
C. CUGRAPH

Answer: A

NEW QUESTION 10
Which two components are included in GPU Operator? (Choose two.)

A. Drivers
B. PyTorch
C. DCGM
D. TensorFlow

Answer: AC

NEW QUESTION 10
Which feature of RDMA reduces CPU utilization and lowers latency?

A. Increased memory buffer size.
B. Network adapters that include hardware offloading.
C. NVIDIA Magnum I/O software.

Answer: B

NEW QUESTION 13
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 15
What is a common tool for container orchestration in AI clusters?

A. Kubernetes
B. MLOps
C. Slurm
D. Apptainer

Answer: A

NEW QUESTION 18
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 22
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 26
When using an InfiniBand network for an AI infrastructure, which software component is necessary for the fabric to function?

A. Verbs
B. MPI
C. OpenSM

Answer: C

NEW QUESTION 28
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 29
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 33
How many Mellanox ConnectX-6 Single Port VPI cards are in a DGX A100 system?

A. 8
B. 16
C. 4

Answer: A

NEW QUESTION 38
What is the importance of a job scheduler in an AI resource-constrained cluster?

A. It allocates resources based on which job requests came first.
B. It ensures that all jobs in the cluster are executed simultaneously.
C. It increases the number of resources available in the cluster.
D. It allocates resources efficiently and optimizes job execution.

Answer: D

NEW QUESTION 40
In an AI cluster, what is the importance of using Slurm?

A. Slurm is used for data storage and retrieval in an AI cluster.
B. Slurm is responsible for AI model training and inference in an AI cluster.
C. Slurm is used for interconnecting nodes in an AI cluster.
D. Slurm helps with managing job scheduling and resource allocation in the cluster.

Answer: D

NEW QUESTION 42
What is the maximum number of MIG instances that an H100 GPU provides?

A. 7
B. 8
C. 4

Answer: A

NEW QUESTION 47
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 48
What is an advantage of InfiniBand over Ethernet?

A. InfiniBand always provides higher bandwidth than Ethernet.
B. InfiniBand supports RDMA while Ethernet does not.
C. InfiniBand offers lower latency than Ethernet.

Answer: C

NEW QUESTION 53
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons 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 FirstTry

* NCA-AIIO Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year