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 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 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 architecture is the core concept behind large language models?

A. BERT Large model
B. State space model
C. Transformer model
D. Attention model

Answer: C

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

A. Data Augmentation
B. Training
C. Inference

Answer: B

NEW QUESTION 8
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 9
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 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 10
What is a common tool for container orchestration in AI clusters?

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

Answer: A

NEW QUESTION 15
For which workloads is NVIDIA Merlin typically used?

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

Answer: A

NEW QUESTION 17
What is one key advantage that Cloud GPU Infrastructure has over On-Prem GPU infrastructure?

A. Lower cost barrier to entry.
B. Reduced cost of I/O traffic.
C. Greater flexibility for hardware orchestration.

Answer: A

NEW QUESTION 18
How is out-of-band management utilized by network operators in an AI environment?

A. It is used to remotely manage and troubleshoot network devices independently of the production network.
B. It is used to directly manage the AI model??s learning rate during training sessions.
C. It is used to increase the computational power of AI models by adapting additional processing resources.
D. It is used to manage the data throughput of AI applications by prioritizing network traffic.

Answer: A

NEW QUESTION 20
How many distinct network fabrics are in an AI cluster?

A. 3
B. 2
C. 4
D. 5

Answer: A

NEW QUESTION 25
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 28
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 31
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 36
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 37
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 40
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 41
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 46
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
.....

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