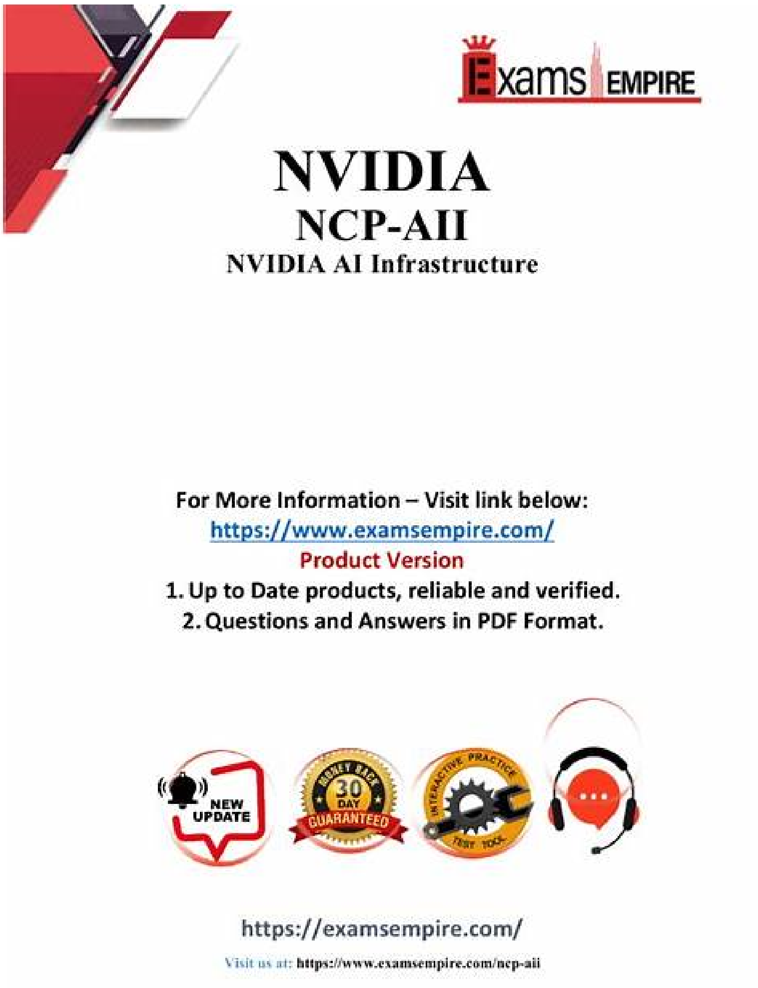
NVIDIA NCP-AII Real Questions & NCP-AII Hottest Certification



P.S. Free 2025 NVIDIA NCP-AII dumps are available on Google Drive shared by Real4dumps: https://drive.google.com/open?id=1m2ZWjgcHH67P9khgxlRfC3hWCalfTUj7

With the help of our NCP-AII study guide, you can adjust yourself to the exam speed and stay alert according to the time-keeper that we set on our NCP-AII training materials. Therefore, you can trust on our NCP-AII exam materials for this effective simulation function will eventually improve your efficiency and assist you to succeed in the NCP-AII Exam. And we believe you will pass the NCP-AII exam just like the other people!

If you still feel nervous for the exam, our NCP-AII Soft test engine will help you to release your nerves. NCP-AII Soft test engine can stimulate the real environment, and you can know the general process of exam by using the exam dumps. What's more, we provide you with free update for one year, and you can get the latest information for the NCP-AII Learning Materials in the following year. We have online service stuff, if you have any questions about the NCP-AII exam braindumps, just contact us.

>> NVIDIA NCP-AII Real Questions <<

NCP-AII Hottest Certification & NCP-AII Reliable Test Labs

The NCP-AII practice test of Real4dumps is created and updated after feedback from thousands of professionals. Additionally, we also offer up to free NCP-AII exam dumps updates. These free updates will help you study as per the NVIDIA NCP-AII latest examination content. Our valued customers can also download a free demo of our NVIDIA NCP-AII exam dumps before purchasing.

NVIDIA AI Infrastructure Sample Questions (Q160-Q165):

NEW QUESTION # 160
A server with four installed NVIDIA GPUs is experiencing intermittent crashes during heavy AI training workloads. You suspect a power issue. You have monitored the power consumption and found that the GPUs are briefly exceeding the rated power capacity of the PSU during peak loads. What are TWO effective mitigation strategies you can implement? (Select TWO)

- A. Replace the PSU with a higher wattage PSU.
- B. Increase the server room temperature.
- C. Disable one of the GPUs to reduce the total power draw.
- D. Re-seat the GPUs in their respective slots.
- E. Underclock the GPIJs to reduce their power consumption.

Answer: A,E

Explanation:
Underclocking the GPUs reduces their power consumption directly. Replacing the PSU provides more headroom to handle the peak loads. Disabling a GPU reduces performance. Increasing server room temperature exacerbates the problem. Reseating GPUs addresses connection issues, not power limitations.

NEW QUESTION # 161
A large A1 model is training using a dataset stored on a network-attached storage (NAS) device. The data transfer speeds are significantly lower than expected. After initial troubleshooting, you discover that the MTU (Maximum Transmission Unit) size on the network interfaces of the training server and the NAS device are mismatched. The server is configured with an MTIJ of 1500, while the NAS device is configured with an MTU of 9000 (Jumbo Frames). What is the MOST likely consequence of this MTU mismatch, and what action should you take?

- A. The server will be unable to communicate with the NAS device. Reduce the MTU size on the server to match the MTU size of the NAS device.
- B. Data packets will be fragmented, leading to increased overhead and reduced performance. Configure both the server and the NAS device to use the same MTU size (either 1500 or 9000).
- C. The connection between the server and the NAS device will be unreliable, resulting in data corruption. Increase the MTU size on both devices to the maximum supported value.
- D. The data transfer will be limited to the lowest common MTU size, but there will be no significant performance impact. No action is required.
- E. Data packets will be retransmitted, increasing the latency but still getting the full throughput. Configure the server to use Path MTU Discovery (PMTUD).

Answer: B

Explanation:
An MTU mismatch (option A) will cause fragmentation, where larger packets are broken down into smaller packets before being transmitted, adding overhead and reducing performance. The solution is to configure both devices to use the same MTU size. Choosing 1500 ensures compatibility, while 9000 requires the entire network path to support jumbo frames.

NEW QUESTION # 162
Consider an AI server equipped with two NVIDIAAI 00 GPUs interconnected with NVLink. You want to maximize the memory bandwidth available to a CUDA application. You observe that the application's performance doesn't scale linearly with the number of GPUs. Which of the following coding techniques or configurations could potentially improve inter-GPU memory access performance?

- A. Employ Unified Memory (I-JM) with prefetching to automatically migrate data between GPUs as needed.
- B. Use CUDA-aware MPl for inter-GPU communication to leverage NVLink.
- C. Disable NVLink to force the application to use PCle, which might provide more consistent performance.
- D. Manually manage data transfers between GPUs using 'cudaMemcpyPeer' to exploit NVLink bandwidth. Choose the GPU with more free memory for allocations.
- E. Ensure all memory allocations are performed on GPU O to minimize data transfer.

Answer: D

Explanation:
ScudaMemcpyPeer allows explicit, optimized data transfers between GPUs using NVLink. Unified Memory with prefetching can simplify development, but might not always provide the best performance. CUDA-aware MPl is typically used for inter-node communication, not intra-node GPU-GPU. Allocating all memory on one GPU defeats the purpose of multi-GPU acceleration. PCle will be slower than NVLink. Manually managing memory transfers, while complex, gives the programmer the most control over leveraging NVLink bandwidth.

NEW QUESTION # 163
You have an NVIDIAAIOO GPU and need to configure it for optimal performance across two distinct AI workloads: a large language model (LLM) training job and a computer vision inference service. The LLM benefits from maximum memory bandwidth, while the inference service requires low latency and high throughput. Which MIG configuration would best suit this scenario?

- A. Create a single full-GPU instance and use Kubernetes resource quotas to isolate the workloads.
- B. Utilize Time-Slicing on a single full-GPU instance, allocating specific time slots to each workload using NVIDIA Vgpu technology
- C. Create one log. 120gb instance for the LLM and one 4g.40gb instance for inference.
- D. Create one 14g.160gb MIG instance for the LLM and use CUDA MPS to multiplex the inference service.
- E. Create two 7g.80gb MIG instances, one for each workload.

Answer: C

Explanation:
Creating a log. 120gb instance for the memory-intensive LLM and a 4g.40gb instance for the inference service provides dedicated resources that cater to the specific needs of each workload, without the overhead or limitations of CUDA MPS or Kubernetes resource quotas. Option A is too conservative, potentially limiting the LLM performance. Option B sacrifices dedicated resources for inference, which may hurt latency. Option C does not leverage MIG and does not guarantee resource isolation and performance consistency. Option E introduces complexities associated with Time-Slicing and might not be suitable for real-time processing.

NEW QUESTION # 164
After successfully installing the NVIDIA Container Toolkit and configuring the Docker runtime, you attempt to run a container that requires GPU access. However, the container fails to start with an error indicating that no GPUs are detected. You've verified that 'nvidia-smi' works on the host. Which of the following could be potential causes for this issue? (Select all that apply)

- A. The '-gpus all' flag was not included when running the 'docker run' command.
- B. The NVIDIA Container Toolkit package is corrupted and needs to be reinstalled.
- C. The Docker daemon was not restarted after configuring the NVIDIA runtime.
- D. The NVIDIA drivers are not compatible with the kernel version running on the host.
- E. The container image itself is missing the necessary CUDA libraries.

Answer: A,C,D

Explanation:
Several factors can cause this issue. A driver-kernel incompatibility (A) prevents the NVIDIA drivers from properly communicating with the hardware. Forgetting to restart the Docker daemon (B) means the configuration changes applied by 'nvidia-ctk' are not active. The '-gpus all' (or equivalent) flag (C) is mandatory to explicitly request GPU resources for the container. Corrupted toolkit (D) would likely present installation failures earlier. Missing CUDA libraries (E) would likely lead to runtime errors within the container, not a failure to detect the GPUs in the first place.

NEW QUESTION # 165
......

As everyone knows that you will get a better position or chance for your future as you are improving yourself and obtaining a NVIDIA NCP-AII certification, As far as the top standard and relevancy of Prepare for your NVIDIA AI Infrastructure NCP-AII valid dumps are concerned, the NVIDIA Exam Questions are designed and verified by experienced and qualified NCP-AII exam experts, Maybe you cannot wait to understand our NCP-AII guide questions; we can promise that our products have a higher quality when compared with other study materials.

The following example performs the same comparison but adds NCP-AII its type and parameter declarations outside the braces, directly onto the function name, Guide to Application Topics.

As everyone knows that you will get a better position or chance for your future as you are improving yourself and obtaining a NVIDIA NCP-AII Certification.

Quiz 2025 NVIDIA Authoritative NCP-AII Real Questions

As far as the top standard and relevancy of Prepare for your NVIDIA AI Infrastructure NCP-AII valid dumps are concerned, the NVIDIA Exam Questions are designed and verified by experienced and qualified NCP-AII exam experts.

Maybe you cannot wait to understand our NCP-AII guide questions; we can promise that our products have a higher quality when compared with other study materials.

We can make you pay a minimum of effort to get the greatest success, Look no further than our platform for real NCP-AII exam dumps.