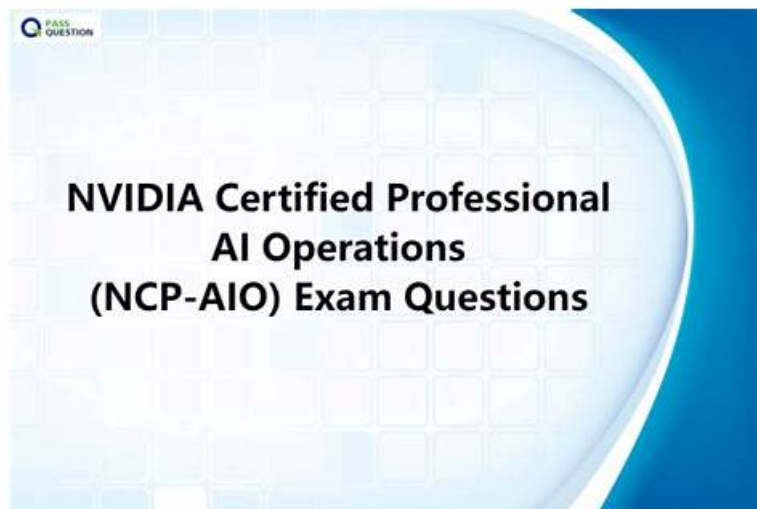


# NCP-AIO Exam Questions Fee, NCP-AIO Interactive Practice Exam



BTW, DOWNLOAD part of TestKingIT NCP-AIO dumps from Cloud Storage: [https://drive.google.com/open?id=1LXzPB5DRHAio3wqRWyYiZuowHsJGdPW\\_](https://drive.google.com/open?id=1LXzPB5DRHAio3wqRWyYiZuowHsJGdPW_)

Our team of experts updates actual NVIDIA AI Operations (NCP-AIO) questions regularly so you can prepare for the NVIDIA AI Operations (NCP-AIO) exam according to the latest syllabus. Additionally, we also offer up to 1 year of free NVIDIA AI Operations (NCP-AIO) exam questions updates. We have a 24/7 customer service team available for your assistance if you get stuck somewhere. Buy NVIDIA NCP-AIO Latest Questions of TestKingIT now and get ready to crack the NCP-AIO certification exam in a single attempt.

## NVIDIA NCP-AIO Exam Syllabus Topics:

Topic	Details
Topic 1	<ul style="list-style-type: none"><li>• Troubleshooting and Optimization: NVIThis section of the exam measures the skills of AI infrastructure engineers and focuses on diagnosing and resolving technical issues that arise in advanced AI systems. Topics include troubleshooting Docker, the Fabric Manager service for NVIDIA NVlink and NVSwitch systems, Base Command Manager, and Magnum IO components. Candidates must also demonstrate the ability to identify and solve storage performance issues, ensuring optimized performance across AI workloads.</li></ul>
Topic 2	<ul style="list-style-type: none"><li>• Installation and Deployment: This section of the exam measures the skills of system administrators and addresses core practices for installing and deploying infrastructure. Candidates are tested on installing and configuring Base Command Manager, initializing Kubernetes on NVIDIA hosts, and deploying containers from NVIDIA NGC as well as cloud VMI containers. The section also covers understanding storage requirements in AI data centers and deploying DOCA services on DPU Arm processors, ensuring robust setup of AI-driven environments.</li></ul>
Topic 3	<ul style="list-style-type: none"><li>• Administration: This section of the exam measures the skills of system administrators and covers essential tasks in managing AI workloads within data centers. Candidates are expected to understand fleet command, Slurm cluster management, and overall data center architecture specific to AI environments. It also includes knowledge of Base Command Manager (BCM), cluster provisioning, Run.ai administration, and configuration of Multi-Instance GPU (MIG) for both AI and high-performance computing applications.</li></ul>

Topic 4	<ul style="list-style-type: none"> <li>• <b>Workload Management:</b> This section of the exam measures the skills of AI infrastructure engineers and focuses on managing workloads effectively in AI environments. It evaluates the ability to administer Kubernetes clusters, maintain workload efficiency, and apply system management tools to troubleshoot operational issues. Emphasis is placed on ensuring that workloads run smoothly across different environments in alignment with NVIDIA technologies.</li> </ul>
---------	---

>> NCP-AIO Exam Questions Fee <<

## NCP-AIO Interactive Practice Exam - Reliable NCP-AIO Braindumps Book

To choose the IT industry is to choose a high salary and a brighter future. And few people can resist the temptation. So, more and more people are interested in the certification exams. NVIDIA NCP-AIO Certification is growing popular among IT fields. TestKingIT gives the candidates to provide the exam materials with best price and high quality practice tests. Our products are cost-effective and we will provide free updates for a year. Our certification training materials are available. We TestKingIT is a leading supplier of answer's dumps providing with the most accurate training materials --- questions and answers.

### NVIDIA AI Operations Sample Questions (Q34-Q39):

#### NEW QUESTION # 34

After installing Kubernetes on your NVIDIA hosts using BCM, you notice that the GPU metrics are not being collected by your monitoring system (e.g., Prometheus). You've confirmed that the NVIDIA Device Plugin is running correctly and GPUs are accessible to containers.

What is the next MOST likely component to investigate and how would you address it?

- A. The Kubernetes API server is throttling metrics requests. Increase the API server's throttling limits for metrics requests.
- B. The Prometheus service discovery is not configured to scrape metrics from the NVIDIA Device Plugin endpoint. Update the Prometheus configuration to include the device plugin's metrics endpoint.
- **C. The NVIDIA Data Center GPU Manager (DCGM) exporter is not deployed or configured correctly. Deploy and configure the DCGM exporter to expose GPU metrics in a Prometheus-compatible format.**
- D. The kubelet's resource usage metrics endpoint is not properly configured. Edit the kubelet configuration file to enable GPU metrics collection.
- E. The cluster's logging driver is interfering with metrics collection. Switch to a different logging driver (e.g., journald) that doesn't conflict with metrics collection.

**Answer: C**

Explanation:

The NVIDIA Data Center GPU Manager (DCGM) exporter is specifically designed to collect and expose GPU metrics in a format that Prometheus can consume. If GPU metrics are not being collected, the DCGM exporter is the most likely culprit. The other options are less directly related to GPU metric collection. Option A pertains more to core Kubernetes metrics, option C relates to generic prometheus service discovery which isn't specialized to GPU data. Logging drivers and API throttling are less likely to directly block metrics collection.

#### NEW QUESTION # 35

You have configured MIG instances for different users in a multi-tenant environment. One user complains that their application is running slower than expected, despite having a dedicated MIG instance. You suspect resource contention on the host system. Which of the following could be causing the slowdown, even with MIG in place?

- **A. Network bandwidth limitations. If the application relies on network communication, bandwidth limitations could be the bottleneck.**
- **B. Insufficient host memory. The overall host system might be running low on memory, causing swapping and slowing down all processes.**
- C. Insufficient power provided by the PSU.
- **D. CPU core oversubscription. Even with dedicated MIG instances, CPU cores might be oversubscribed, leading to performance degradation.**
- E. MIG guarantees complete isolation, so resource contention is impossible.

**Answer: A,B,D**

Explanation:

MIG provides GPU resource isolation, but it does not isolate other system resources. CPU oversubscription, insufficient host memory, and network bandwidth limitations can all contribute to performance slowdowns, even with dedicated MIG instances. It's important to monitor and manage these resources in addition to GPU resources.

#### NEW QUESTION # 36

You have a VMI container running on a cloud platform. You need to monitor the GPU utilization (e.g., GPU memory usage, GPU utilization percentage). Which of the following tools is MOST commonly used for this purpose?

- A. top
- B. nvidia-smi
- C. ps
- D. htop
- E. vmstat

**Answer: B**

Explanation:

'nvidia-smi' (NVIDIA System Management Interface) is the primary command-line utility for monitoring and managing NVIDIA GPUs. It provides detailed information about GPU utilization, memory usage, temperature, and other metrics.

#### NEW QUESTION # 37

Consider an HPC application heavily reliant on CODA. You plan to leverage MIG to optimize GPU resource allocation within your cluster.

Which configuration approach would BEST ensure the HPC application benefits from high GPU compute capability while coexisting with other workloads?

- A. Create MIG instances tailored to the HPC application's specific memory and compute needs, allocating the necessary resources without over-provisioning. Utilize the remaining resources for other workloads.
- B. Create a single, large MIG instance dedicated solely to the HPC application, maximizing its compute capacity.
- C. Configure all MIG instances with equal memory and compute allocation to provide a fair distribution of resources.
- D. Create multiple small MIG instances and distribute the HPC workload across them.
- E. Disable MIG and allow the HPC application to utilize the entire GPU for maximum performance.

**Answer: A**

Explanation:

Tailoring MIG instances to the HPC application's specific requirements ensures efficient resource allocation and allows other workloads to utilize the remaining GPU capacity. D is not ideal for concurrent workloads. A and E don't account for specific workload requirements.

#### NEW QUESTION # 38

You are deploying a stateful application to your Kubernetes cluster running on NVIDIA hardware provisioned through BCM. This application requires direct access to a persistent volume on a high-performance NVMe drive. Which of the following methods is MOST appropriate for providing this access while ensuring high performance and data consistency?

- A. Using a Network File System (NFS) share mounted on the host and exposed to the container via a PersistentVolume.
- B. Leveraging a local Persistent Volume with 'volumeBindingMode: WaitForFirstConsumer' and node affinity to ensure the pod is scheduled on the node with the NVMe drive.
- C. Configuring a standard Persistent Volume Claim backed by a software-defined storage solution like Ceph or Rook.
- D. Creating a PersistentVolumeClaim (PVC) backed by a cloud-based block storage service (e.g., AWS EBS, Azure Disk).
- E. Using a 'hostPath' volume, directly mapping the NVMe drive's path on the host node to the container.

**Answer: B**

Explanation:

BTW, DOWNLOAD part of TestKingIT NCP-AIO dumps from Cloud Storage: <https://drive.google.com/open?id=1LXzPB5DRHAio3wqRWyYiZuowHsJGdPW>