

CASE STUDY:

Large University System Enables Secure Generative AI Infrastructure Using VMware Private AI Foundation with NVIDIA



Matthew Marden
 Research Vice President,
 Business Value Strategy Practice, IDC

BUSINESS VALUE HIGHLIGHTS

Company

Higher education system with more than 100,000 students

Location

United States

Challenge

Enhance AI and generative AI (GenAI) capabilities while managing infrastructure costs and ensuring data privacy

Solution

VMware Private AI Foundation with NVIDIA

Current Financial and Quantitative Benefits

- 30%–50% lower infrastructure costs
- 50% higher IT infrastructure team efficiency
- 1–2 days required to train GenAI models
- 25% improved staff efficiencies for GenAI use case development
- 10%–30% enhancement in governance team efficiencies
- 70% reduced risk associated with GenAI use cases

Overview

For this study, IDC interviewed a large higher education system in the United States comprising a number of universities at different locations and several health institutions supporting over 100,000 students.

An IT director working with the higher education system explained that his organization recognized a need to enhance its AI capabilities while optimizing infrastructure costs and ensuring data privacy. He explained that his organization chose VMware Private AI Foundation with NVIDIA as its private cloud infrastructure for AI-based activities because it would help the organization to efficiently manage costs, leverage its existing datacenter investments and in-house VMware skills, and maintain control over its data, all while providing a scalable and secure GenAI environment.

VMware Private AI Foundation with NVIDIA is a GenAI platform that enables enterprises to run RAG workflows, fine-tune and customize large language models (LLMs), and perform inference workloads in their datacenters. The VMware platform seeks to address key concerns that GenAI adopters often have, such as privacy, cost, performance, and compliance, and incorporates solutions such as NVIDIA AI Enterprise and VMware Cloud Foundation.

According to the IT director, his organization's AI and GenAI strategies focus on leveraging advanced AI capabilities to enhance research, student services, and administrative functions. The IT director explained that his organization is using VMware Private AI Foundation with NVIDIA to create a scalable, secure AI environment that leverages its existing VMware-based datacenter infrastructure and staff experience. While its use of VMware Private AI Foundation with NVIDIA is still relatively recent, he focused on results achieved so far and the expected impact and benefits, including infrastructure cost savings, enhanced AI capabilities for research and student services, and improved productivity through efficient resource management and strong data governance.

Implementation

The IT director explained that his organization considered several criteria when selecting a platform for developing and deploying GenAI models and use cases.

The primary criteria included:

- **Considering existing infrastructure and cost management:** The higher education system already had a substantial VMware-based datacenter environment that it wanted to leverage to the highest extent possible. This factor led it to favor infrastructure approaches that incorporate on-premises elements based on VMware solutions.
- **In-house skill set and resource optimization:** The higher education system already had significant expertise in VMware technologies. Leveraging this existing skill set was crucial for a smooth transition and efficient ongoing operations. The IT director noted, *"We already had a lot of VMware skills, so our team knew how to do all that."*
- **Security and data control:** Ensuring data security and maintaining control over sensitive information were paramount. VMware's private AI solution allows an organization to keep data within its own environment, mitigating risks associated with data breaches and unauthorized access. The IT director highlighted, *"If individuals were taking things into the private cloud from our environment, sometimes we tend to lose control of the data that leaves our environment."*
- **Scalability and multi-tenancy:** The ability to support multiple institutions with a single solution was essential. VMware's multi-tenancy capabilities allow each institution to have a customized experience while sharing the same underlying infrastructure. The IT director

stated, *“We provide a shared service ... Many of these other institutions can use our service and not have to go out and individually get that.”*

- **Cost efficiency and predictability:** The need for predictable costs and efficient resource utilization was a significant consideration. The ability to fractionally use GPUs and manage costs effectively was a key advantage of the VMware solution. The IT director explained, *“We have the management stack on one side and the AI stack on the other ... This way, we can efficiently allocate all the GPUs across the board.”*

The higher education system began deploying VMware Private AI Foundation with NVIDIA in 2024, starting with a three-month engagement at a high-performance computing center. Following a successful proof of concept, it fully implemented the solution by early 2025, involving a team of three full-time equivalents (FTEs) and leveraging its existing VMware skills and infrastructure.

Benefits

Deploying VMware Private AI Foundation with NVIDIA enables the higher education system to optimize infrastructure costs, enhance AI capabilities for research and student services, and improve productivity through efficient resource management and strong performance. The expected benefits include impactful AI and GenAI solutions, robust data privacy, and significant operational efficiencies across its institutions.

Infrastructure Cost Optimization and Staff Efficiencies

The IT director explained that his organization has already achieved significant benefits in terms of infrastructure costs and staff time efficiencies by deploying VMware Private AI Foundation with NVIDIA. He expects these benefits to increase alongside GenAI use case growth. He noted, *“One of the main reasons we chose this solution is because we can have a predictable price.”* By leveraging its existing datacenters and using VMware’s technology to fractionally allocate GPU resources, the higher education system has been able to optimize infrastructure-related costs and establish predictable cost models for the GenAI models and use cases it builds and deploys.



One of the main reasons we chose [VMware Private AI Foundation with NVIDIA] is because we can have a predictable price.”

IT Director
Higher Education System

The efficient use of compute, storage, networking, and virtualization resources with VMware Private AI Foundation with NVIDIA has been an advantage. These efficiencies have allowed the higher education system to run and manage its AI workloads more effectively, reducing the need for additional hardware and minimizing operational costs.

In terms of staff time requirements, the higher education system has also realized benefits from efficiencies with VMware Private AI Foundation with NVIDIA. The IT director commented, *“We only have 3 FTEs involved, and they’re not even full-time dedicated to this project.”* He compared this baseline level of efficiency to other organizations that require more full-time staff to manage similar AI deployments. He also noted that the use of VMware tools and governance features has streamlined management processes, allowing the existing team to handle the AI infrastructure without the need for additional resources.

The use of VMware Private AI Foundation with NVIDIA has already enabled the higher education system to achieve a more cost-effective and efficient IT footing for its GenAI use cases, with further benefits expected. The predictable pricing, efficient resource allocation, and reduced staff time requirements have contributed to the early success of the organization’s GenAI initiatives, positioning it to potentially be a leader in leveraging AI technology in higher education.

Development of GenAI Models and GenAI Use Cases

Using VMware Private AI Foundation with NVIDIA has enhanced the higher education system’s ability to develop GenAI models efficiently and rapidly. The IT director explained that streamlined processes allow his organization to focus more on the models themselves rather than the underlying infrastructure, leading to faster cycles.

The efficiency gains are evident in the productivity of the team responsible for creating AI models. The IT director noted, *“We couldn’t have done it with a small team like we’re doing now without VMware Private AI Foundation with NVIDIA.”* The existing team, which manages over 1,000 VMs in addition to the new AI infrastructure, has become more productive without the need for additional resources and can deliver GenAI models in a matter of days.

The ability to fractionally use GPUs has further optimized the resource allocation required for the development of GenAI models, enabling the higher education system to develop models more rapidly. This flexibility in resource management has allowed the higher education system to efficiently allocate GPU capacity based on need, ensuring that hardware limitations do not hinder development processes as it seeks to create and use GenAI models for use cases that will have a broad-based positive impact on operations.

Governance and Security for GenAI

Using VMware Private AI Foundation with NVIDIA has significantly enhanced the governance, privacy, and security of GenAI models and use cases for the higher education system. According to the IT director, the built-in governance features of VMware's solution have been particularly beneficial, allowing the higher education system to automate the rapid deployment of models while ensuring compliance. This has enabled the introduction of AI into existing governance structures without creating new ones, streamlining the process and maintaining consistency.

Privacy and security have also been strengthened through VMware's solution. The higher education system can manage data more effectively, ensuring that sensitive information remains within its control.

"We provided a place to contain and manage that risk," the IT director explained, highlighting the importance of keeping data secure within its environment. Features such as VPN connections and two-factor authentication for user access have further enhanced security, reducing the risk of data breaches and unauthorized access. According to the IT director, *"We don't have to go and build a tool to manage containers, to manage the GenAI models, something to deploy AI, or something to fractionalize the GPUs with VMware Private AI Foundation with NVIDIA."*

Operational Impact of GenAI Use Cases

The IT director explained that the higher education system has already developed several GenAI use cases using VMware Private AI Foundation with NVIDIA, with considerable early impact on operations, employees, and students. One of the most notable use cases is the creation of virtual AI-enabled tutors. *"Each subject, what we're trying to do is to create a virtual AI-enabled tutor,"* according to the IT director. These AI tutors help students by providing explanations and answering questions related to their course material, enhancing their learning experience and accessibility to academic support.

Another main early use case is the automation of administrative tasks, such as budgeting and benefits analysis. The IT director explained, *"HR does benefits analysis. For healthcare insurance and things like that, we have to analyze what is the best offering, what's the best combination."* By leveraging AI to streamline these processes, the higher education system can improve efficiency and reduce the time required for complex analyses, ultimately benefiting both staff and students.



We don't have to go and build a tool to manage containers, to manage the GenAI models, something to deploy AI, or something to fractionalize the GPUs with VMware Private AI Foundation with NVIDIA."

IT Director
Higher Education System

The impact of these GenAI use cases extends beyond operational efficiencies. The IT director noted, *“The benefit is how the new workforce is able to do their work and understanding how to improve even if peaking in performance.”* The deployment of GenAI solutions has led to productivity gains across various business units, allowing employees to perform their tasks more effectively and students to receive better academic support. Overall, the use of VMware Private AI Foundation with NVIDIA has enabled the higher education system to enhance its operations, improve employee productivity, and provide valuable resources to its students.

Quantitative Benefits

The implementation of VMware Private AI Foundation with NVIDIA has already resulted in quantitative and financial benefits for the higher education system.

At the time of the interview, these benefits include:

- **Infrastructure costs:** The IT director estimated that infrastructure costs will be between 30%–50% lower with VMware Private AI Foundation with NVIDIA, as his organization is able to divide compute and GPU capacity more readily to various GenAI development efforts and use cases and take advantage of its existing VMware-based infrastructure.
- **IT infrastructure staff efficiencies:** The IT director put expected efficiencies for the higher education system’s IT infrastructure team at about 50%, as he expects to use about half of the staff resources to run and manage the VMware Private AI Foundation with NVIDIA environment compared with another environment for GenAI model development and GenAI use cases.
- **GenAI model training:** The IT director reported that his organization has created four GenAI models so far on VMware Private AI Foundation with NVIDIA, and each model only took one to two days.
- **GenAI use case development:** The IT director explained that VMware Private AI Foundation with NVIDIA has been an efficient platform for use case development so far, with estimated staff efficiencies at around 25%.
- **GenAI governance:** The IT director estimated 10%–30% efficiencies so far in terms of AI governance with VMware Private AI Foundation with NVIDIA, as these teams benefit from the governance tools included with the platform.

Overall, the use of VMware Private AI Foundation with NVIDIA has enabled the higher education system to enhance its operations, improve employee productivity, and provide valuable resources to its students.

- **GenAI risk:** The IT director linked functionalities of VMware Private AI Foundation with NVIDIA with a lower risk profile for GenAI use cases of around 70%, which will help the higher education system to avoid potential security events with high financial and reputational costs as the use of GenAI expands.
- **GenAI operational impact:** The IT director gave examples of early GenAI use cases that have been developed and are running on VMware Private AI Foundation with NVIDIA that are having a positive impact on how employees work, such as the legal team requiring minutes rather than hours to write policies.

Conclusion

The higher education system's adoption of VMware Private AI Foundation with NVIDIA has proven to be a strategic and beneficial decision. By leveraging existing infrastructure, optimizing resource utilization, and ensuring data security, the higher education system has successfully developed and deployed AI solutions that enhance student support, administrative efficiency, and research capabilities. The quantitative and financial benefits observed further validate the effectiveness of this implementation, positioning the higher education system as a leader in the use of AI in higher education.

Business Value Analysis and Methodology

IDC interviewed an IT director for the higher education system to understand how using VMware Private AI Foundation with NVIDIA has impacted his organization's infrastructure costs, staff time requirements, and development of GenAI models and GenAI use cases as well as the risks and outcomes associated with GenAI use cases.

About the IDC Analyst



Matthew Marden

Research Vice President, Business Value Strategy Practice, IDC

Matthew Marden is responsible for carrying out custom business value research engagements and consulting projects for clients in several technology areas, focusing on determining the return on investment of their use of enterprise technologies. Marden's research often analyzes how organizations are leveraging investment in digital technology solutions and initiatives to create value through efficiencies and business enablement.

[More about Matthew Marden](#)

Message from the Sponsor



VMware Private AI Foundation with NVIDIA is a joint GenAI platform by Broadcom and NVIDIA that helps enterprises unlock GenAI and unleash productivity with lower TCO. With this platform, enterprises can deploy AI workloads faster, fine-tune and customize LLM models, deploy retrieval augmented generation (RAG) workflows, and run inference workloads in their data centers, addressing privacy, choice, cost, performance, and compliance concerns.

[Learn more](#)

IDC Custom Solutions

IDC Custom Solutions produced this publication. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis that IDC independently conducted and published, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. This IDC material is licensed for external use and in no way does the use or publication of IDC research indicate IDC's endorsement of the sponsor's or licensee's products or strategies.



IDC Research, Inc.
140 Kendrick Street, Building B, Needham, MA 02494, USA
T +1 508 872 8200

[idc.com](https://www.idc.com)

[in @idc](https://www.linkedin.com/company/idc)

[X @idc](https://twitter.com/idc)

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight helps IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.

©2025 IDC. Reproduction is forbidden unless authorized. All rights reserved. [CCPA](#)