

TECHNICAL VALIDATION

# Accelerate AI Business Value With NVIDIA Computing by HPE

Shorten Time to Value for AI-driven Applications

By Justin Boyer, Technical Validation Analyst  
Enterprise Strategy Group

March 2025

# Contents

Introduction .....	3
Background .....	3
HPE Private Cloud AI .....	4
Enterprise Strategy Group Technical Validation .....	5
Use the Power of AI Immediately .....	5
Enterprise Strategy Group Testing .....	5
Flexibility to Grow With the Business .....	8
Enterprise Strategy Group Testing .....	8
Partnership With Equinix .....	8
Complete Control Over Data Pipelines .....	10
Enterprise Strategy Group Testing .....	10
Conclusion .....	12

## Introduction

This Technical Validation by Informa TechTarget's Enterprise Strategy Group details our evaluation of the HPE Private Cloud AI platform. This validation explores how HPE Private Cloud AI addresses the challenges that impede the production deployment of enterprise AI initiatives.

## Background

Organizations have embraced AI and continue to use it to generate value for consumers and increase efficiency and revenue for the business. While many businesses are exploring various AI applications, generative AI has emerged as a key area of focus, demonstrating significant advancements in capabilities. According to Enterprise Strategy Group research, 64% of survey respondents indicated that they will invest significantly in generative AI over the next 12 months. However, those who have already embarked on this journey have experienced many challenges, including ensuring data privacy and protection, security, high implementation and operational costs, and a lack of internal expertise. These challenges have led to increased costs and time to market, along with lengthy project rework or modifications.<sup>1</sup>

Implementing a successful AI project requires various skills and expertise that can be challenging to find or develop with technology that is so new. Organizations require AI-specific development skills and data scientists to develop accurate models. Additionally, they require infrastructure expertise to build and maintain an environment capable of performing the massive computation necessary to make generative AI or other AI use cases viable, particularly if they have not gone through a digital transformation journey.

Figure 1 shows the generative AI skills gap present in many organizations today. Among surveyed organizations, we see that AI system deployment and cloud computing, two infrastructure-focused skills, top the list of skills needed to grow internal generative AI talent.<sup>2</sup> This implies that, while organizations may excel in software development, they frequently need specialized support to build the production-ready infrastructure that AI systems demand.

---

<sup>1</sup> Source: Enterprise Strategy Group Research Report, [Navigating the Generative AI Partner and Alliance Landscape](#), November 2024.

<sup>2</sup> Ibid.

**Figure 1. Infrastructure Skills Deficient Among Enterprises Seeking Generative AI**



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Organizations seeking to take advantage of the power of AI-based applications need help to develop and deploy the required infrastructure without increasing rework, time to market, and overall costs.

## HPE Private Cloud AI

NVIDIA AI Computing by HPE focuses on laying the foundation elements required for long-term AI application success with access to the processing power and scalability needed to build robust data pipelines. The first offering of this portfolio, HPE Private Cloud AI, provides a turnkey, scalable, AI-optimized private cloud environment organizations can use as the foundation for their AI projects.

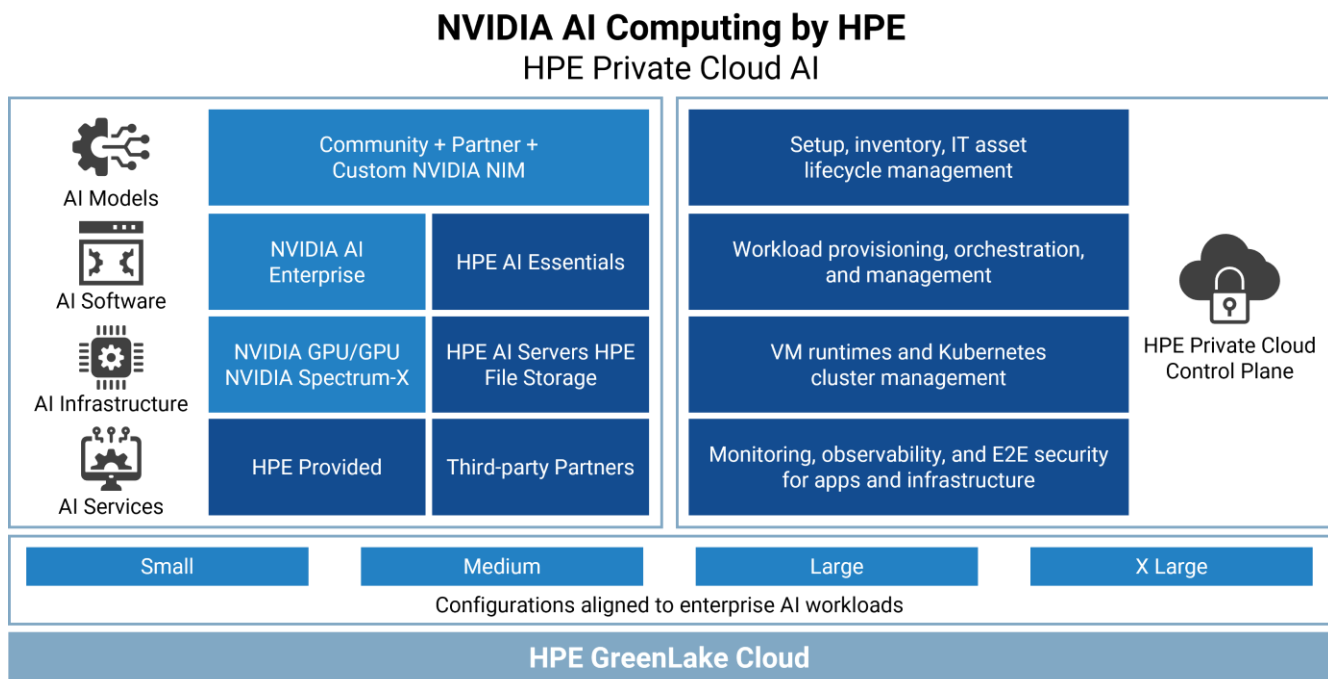
Rather than providing complex reference architectures, HPE Private Cloud AI, co-developed with NVIDIA and powered by HPE GreenLake, delivers a complete, pre-configured solution for enterprise AI. This turnkey system integrates compute, storage, networking, and AI software, enabling you to bypass the hurdles of piecing together disparate components and accelerate your AI initiatives.

HPE Private Cloud AI has been engineered to provide:

- Increased AI time to market and performance.
- Choice of scalable starting points that grow with needs.
- Ultimate control of private data and models.
- Built-in AI compliance and explainability.
- A colocation option in partnership with Equinix.

Figure 2 outlines the HPE Private Cloud AI architecture. HPE Private Cloud AI creates and runs the virtualized and containerized infrastructure required for AI-driven applications, featuring NVIDIA AI-optimized GPU processors, networking, and AI Enterprise software powered by HPE GreenLake and an HPE Private Cloud AI control plane for management. Organizations can choose from four starting points and scale up as needed. HPE's intelligent configuration tool ensures customers get the right-sized system for their AI workloads, optimizing performance and cost.

Figure 2. HPE Private Cloud AI



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

## Enterprise Strategy Group Technical Validation

Enterprise Strategy Group validated how HPE Private Cloud AI, powered by HPE GreenLake, helps organizations overcome many of the challenges they face when trying to build AI-based applications. We evaluated how HPE Private Cloud AI increases time to market, provides flexibility to grow with the business, and allows organizations complete control over their data pipelines, helping them realize the benefits of AI-driven applications.

### Use the Power of AI Immediately

Enterprise Strategy Group validated how HPE Private Cloud AI accelerates time to market and enables organizations to quickly create AI-driven applications.

### Enterprise Strategy Group Testing

In the last few years, AI-driven applications have taken hold across multiple industries. Generative AI, in particular, has grown in power, availability, and in the number of applicable use cases. As generative AI's capabilities increase, organizations look to these technologies to gain a business advantage.

However, many organizations require generative AI applications modeled on their proprietary data stores, leading to more accurate and effective insights based on their needs. Generalized AI applications, such as ChatGPT, don't fulfill these requirements. Even though building a homegrown AI-driven application might serve the business well, it's a challenging undertaking for many.

HPE Private Cloud AI delivers the infrastructure required to develop and run advanced AI-driven applications and the software needed to successfully deploy applications and manage resources. It's an end-to-end turnkey pre-

provisioned solution that provides organizations with the freedom to focus on gaining business value from AI-driven applications. With HPE Private Cloud AI, organizations can be ready to deploy AI-driven applications in a single day.

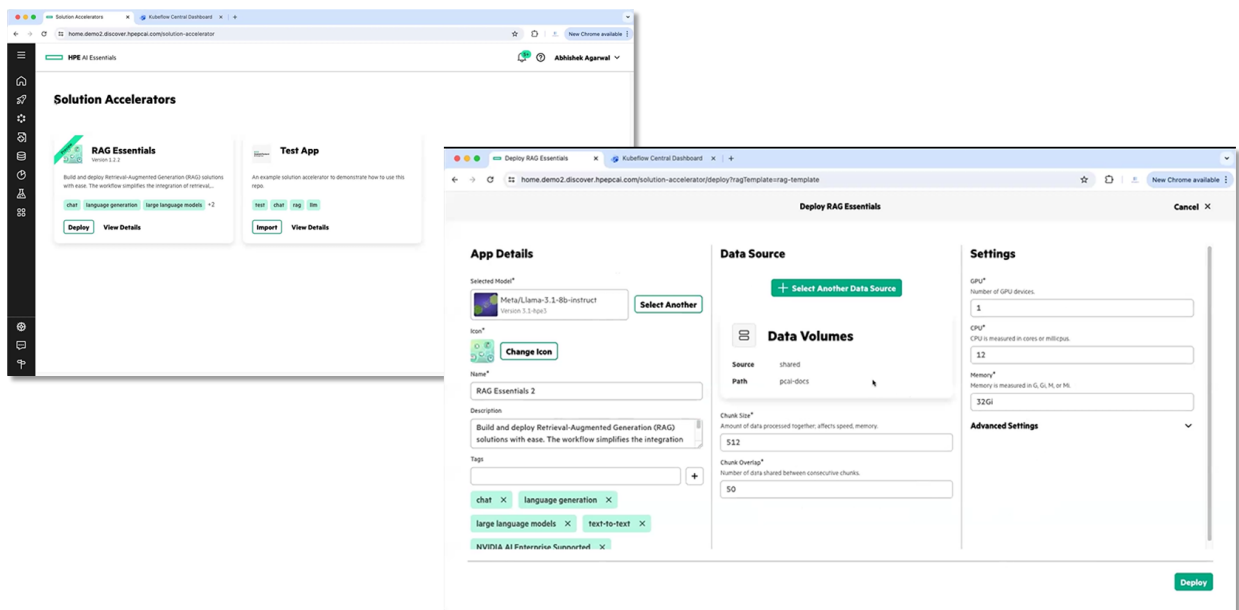
HPE Private Cloud AI runs popular open-source solutions for model development. The solution simplifies Kubeflow deployment by providing a fully managed, security-hardened environment for scalable and reproducible machine learning workflows. It enables hassle-free Apache Airflow deployment with automated setup, centralized orchestration, and security policies to streamline workflow automation. HPE Private Cloud AI integrates MLflow, offering centralized tracking, model management, and secure experiment logging at scale. It also offers a fully managed Jupyter notebook environment with enterprise-grade security, seamless collaboration, and scalable compute resources.

Kubernetes is widely recognized as a robust solution for orchestrating containers, but building and maintaining a Kubernetes cluster independently can be complex, time-consuming, and costly. It typically involves provisioning infrastructure, configuring virtual machines, installing tools like kubectl, kubeadm, kubelet, and container runtimes such as Docker. Additionally, setting up secure connections, certificate management, and continuous monitoring can be overwhelming. HPE Private Cloud AI manages and executes all of these steps, enabling organizations to focus on leveraging AI solutions without worrying about the underlying Kubernetes infrastructure. HPE Private Cloud AI leverages Kubernetes clusters powered by HPE AI Essentials, enabling organizations to benefit from the popular container management system without having to manage it themselves.

Enterprise Strategy Group walked through the process of creating an AI-driven application using HPE Private Cloud AI Solution Accelerators. Solution Accelerators are preconfigured solutions customers can deploy for an out-of-the-box architecture built for the selected use case. In our solution, we created a retrieval-augmented generation (RAG) application using the RAG Essentials Solution Accelerator (see Figure 3).

After navigating to the Solution Accelerators screen, we clicked the Deploy button in the RAG Essentials tile. On the next screen, we selected the desired model, data source, and compute settings. When the user clicks the Deploy button here, HPE AI Essentials Software creates the entire data pipeline along with API endpoints available for use by applications.

**Figure 3.** Three-step Deployment With Solution Accelerators



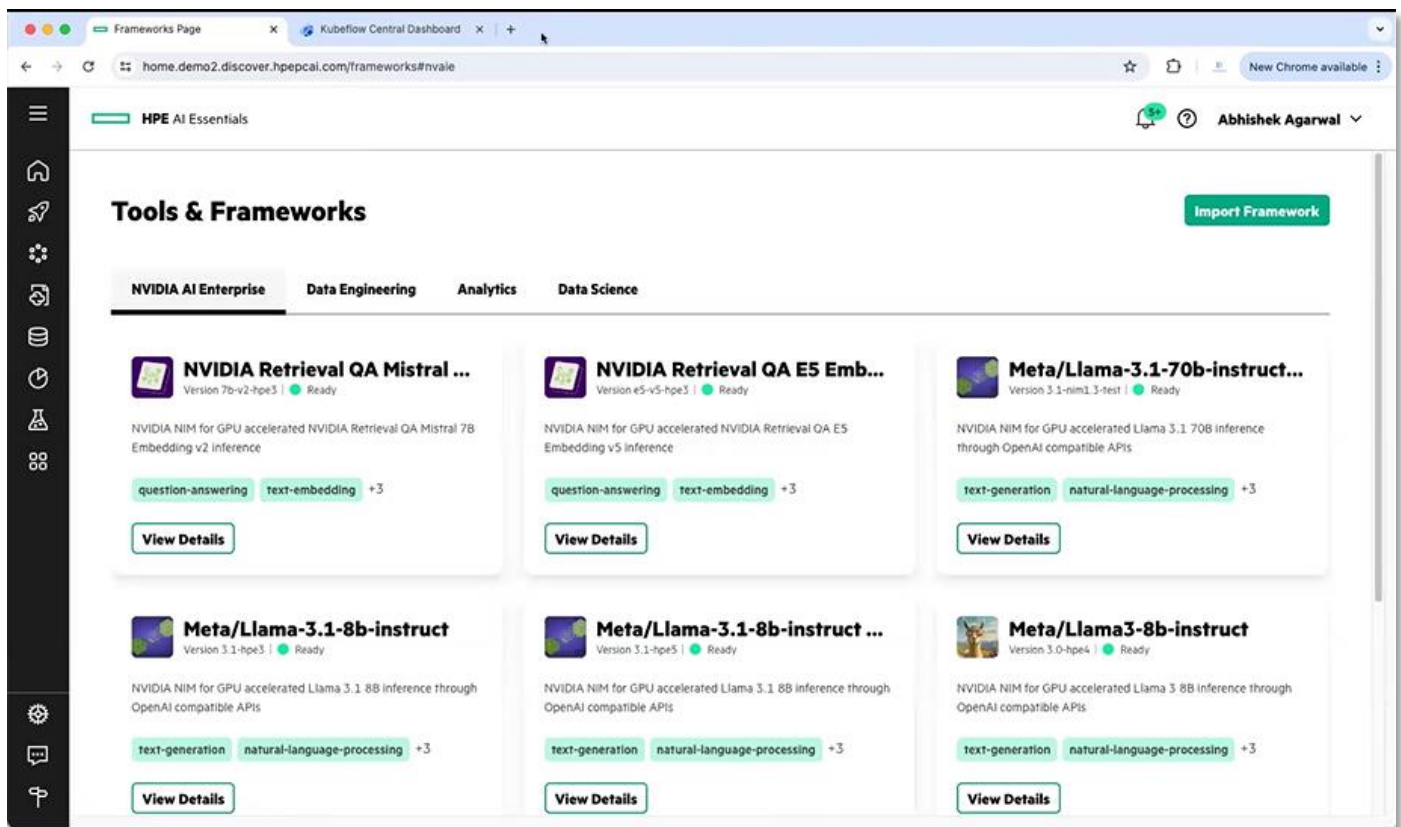
Source: Enterprise Strategy Group, a division of TechTarget, Inc.

HPE AI Essentials Software provides a comprehensive, security-enabled environment for deploying and managing industry-leading open source solutions for AI and data science. These built-in integrations help businesses get the most value out of their environments.

HPE AI Essentials Software features several out-of-the-box models, including popular large language models (LLMs) such as Llama (see Figure 4). These are available to deploy and use immediately. HPE Private Cloud AI also offers several data engineering tools, such as Apache Airflow for workflow and pipeline automation and EzPresto, a SQL query engine used to create federated queries across several data sources and part of the HPE Private Cloud AI software stack.

HPE AI Essentials Software also supports Apache Spark for analytics processing, enabling businesses to run an existing Spark application within the environment or create a new one seamlessly. Additionally, HPE Private Cloud AI provides fully managed, security hardened environments for various data science tools, including Kubeflow, a machine learning (ML) toolkit for Kubernetes; Ray, an AI Compute Engine; and MLflow, a unified ML operations platform. HPE also offers their Machine Learning Inference Software (MLIS) tool, a no-code solution used to deploy LLMs and interact with them. Organizations can import other frameworks to meet their specific needs. Finally, HPE AI Essentials Software features a fully managed Jupyter Notebook environment with enterprise-grade security, advanced collaboration features, and scalable compute resources.

Figure 4. Out-of-the-box AI Tools and Frameworks



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

HPE Private Cloud AI makes applying updates seamless. HPE maintains an available updates list, and admins can apply updates to hardware or software through the interface. As for identity and access management (IAM), broad access to HPE Private Cloud AI clusters is given via roles from HPE GreenLake through a centralized IAM solution

such as Azure AD. Fine-grained permissions, grouped into policies, are maintained within HPE Private Cloud AI for specific data sources or applications.

## Why This Matters

Despite the potential benefits of AI-driven applications and the appetite to use them, challenges abound in bringing them to fruition. According to Enterprise Strategy Group research, 62% of survey respondents indicated that they have faced moderate to extensive challenges moving AI models from development into production. Some cited reasons that stand out are scalability (37%), data integration (31%), monitoring and maintenance (26%), and infrastructure limitations (20%).<sup>3</sup> Organizations struggling to get their AI-driven applications to production need help to build a secure, reliable infrastructure to get them started.

Enterprise Strategy Group validated that HPE AI Essentials Software provides the tools and frameworks, along with the management software, necessary to help organizations build, deploy, and see immediate value from their AI-driven applications. We validated how Solution Accelerators help to deploy the necessary architecture for various AI use cases. We saw how the various built-in integrations with widely adopted tools and frameworks ease the transition into HPE Private Cloud AI and enable organizations to use the tools with which they're already familiar. Powered by HPE GreenLake, HPE Private Cloud AI is a turnkey solution that creates all of the necessary infrastructure to successfully run and scale an AI-based application.

With HPE Private Cloud AI, organizations can significantly reduce the time to market and time to value for deploying their AI-driven applications since they won't have to manage the infrastructure but instead can focus on developing the application and realizing business value. Further, HPE Private Cloud AI helps organizations overcome skills gaps in AI-driven application development by providing a fully managed, security-enabled environment on which to build, with built-in access to industry-leading AI development tools.

## Flexibility to Grow With the Business

Enterprise Strategy Group evaluated HPE Private Cloud AI for its flexibility and capabilities designed to help organizations scale their AI infrastructure with their business.

## Enterprise Strategy Group Testing

According to Enterprise Strategy Group research, 44% of survey respondents indicated that AI, data science, and machine learning have become significantly more important to their organization's future over the past two years, up 16% year over year.<sup>4</sup> Even though the appetite exists for AI, some organizations might not be able to build or scale the physical infrastructure necessary for proper AI system deployment.

HPE AI Essentials Software features pre-defined starting points to provide a solid baseline of GPU capabilities based on an organization's current needs and budget. These configurations range from a small configuration suitable for a broad set of data centers and basic use cases, including AI pilot programs, to large, 2-rack systems built for more advanced use cases. The starting points, provided by HPE GreenLake, are scalable when and if an organization needs it. Therefore, organizations can start with a small configuration for a pilot program, then scale up to a medium or large configuration based on their needs for a fully deployed production application.

## Partnership With Equinix

HPE GreenLake helps organizations experience the scalability and flexibility of the cloud using their own data centers. However, some organizations have challenges in creating the physical infrastructure necessary to

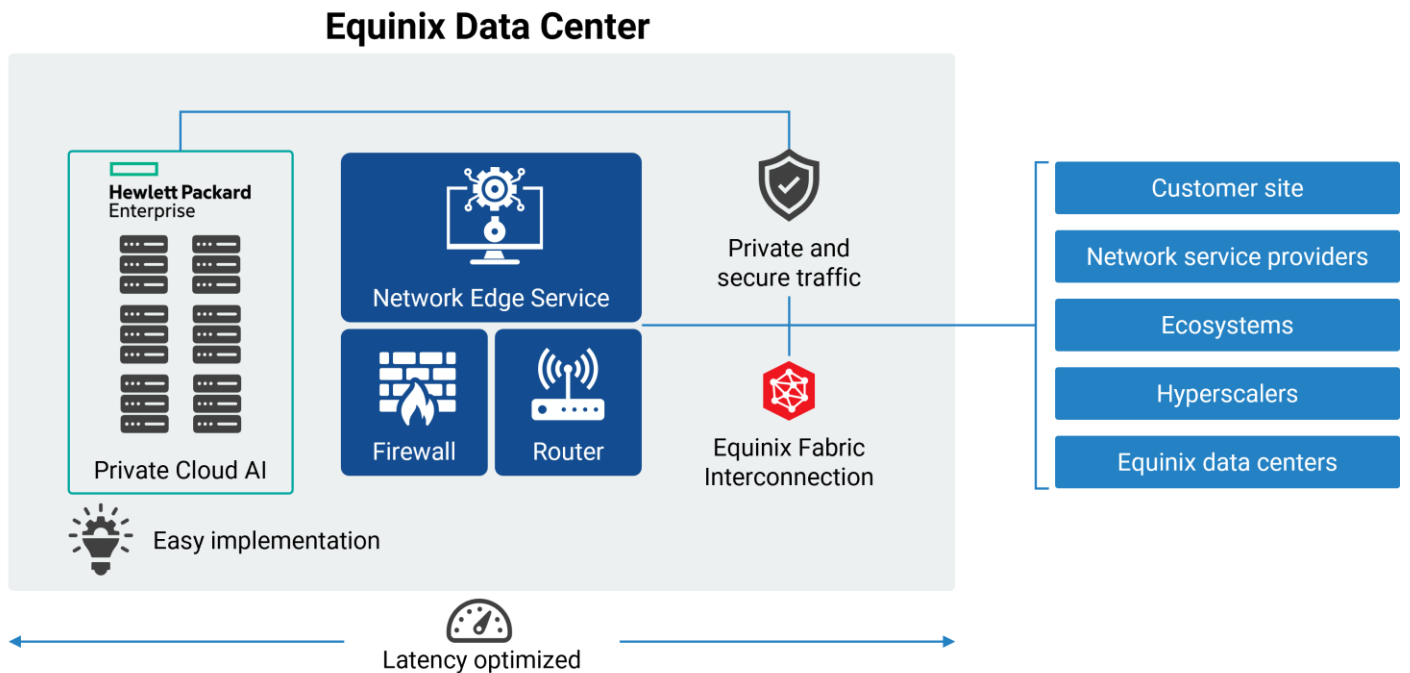
<sup>3</sup> Source: Enterprise Strategy Group Research Report, [Navigating Build-versus-buy Dynamics for Enterprise-ready AI](#), January 2025.

<sup>4</sup> Source: Enterprise Strategy Group Research Report, [2025 Technology Spending Intentions Survey](#), December 2024.

successfully bring an AI-driven application to production. To address this need, HPE partners with Equinix to provide a collocated data center that organizations can use to deploy HPE Private Cloud AI with a pay-as-you-go program.

Figure 5 shows a high-level architecture of an Equinix data center running HPE Private Cloud AI. Equinix data centers are secure, sustainable, and compliant with low-latency connections to cloud and ecosystem partners. Equinix offers this service in seven global locations across North America, Europe, and APAC. Equinix provides the physical infrastructure, running HPE GreenLake for cloud capabilities, on which HPE Private Cloud AI can be deployed for development and production workloads.

**Figure 5.** HPE and Equinix Partner to Provide Colocation Services



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

HPE offers hosted trials for HPE Private Cloud AI through Equinix’s Silicon Valley data center. These trials enable potential customers to quickly deploy infrastructure and evaluate Private Cloud AI for their needs. Featuring rapid deployment of AI workloads, HPE’s partnership with Equinix enables organizations to go from pilot to production easily.

## Why This Matters

Achieving business value through AI-driven applications requires powerful computing capabilities and strong infrastructure support. According to Enterprise Strategy Group research, compute components, real-time processing capabilities, and servers were among the top important infrastructure requirements cited for supporting AI at scale.<sup>5</sup> Some organizations desiring to take advantage of AI may not have the necessary data center infrastructure to support it.

Enterprise Strategy Group validated that HPE's partnership with Equinix provides an option for resource-challenged organizations. Equinix data centers can host Private Cloud AI in its seven colocation facilities across the world. Equinix data centers provide secure, scalable, and sustainable colocation solutions so organizations can benefit from AI without managing the physical infrastructure required to host Private Cloud AI. Additionally, hosted trials allow organizations to pilot HPE Private Cloud AI quickly and to start on the path to production workloads, with Equinix maintaining the physical infrastructure needed to provide high-performance computing power for AI-driven applications.

Using Equinix colocation facilities for their HPE Private Cloud AI deployments, organizations see faster time to value without requiring expensive data center buildouts.

## Complete Control Over Data Pipelines

Enterprise Strategy Group validated how HPE Private Cloud AI enables organizations to maintain complete control over their data pipelines and performance, helping them realize the benefits of their deployed applications.

## Enterprise Strategy Group Testing

When using a managed solution for infrastructure services, such as HPE GreenLake and Private Cloud AI, maintaining control over data pipelines, compute power, and even the underlying Kubernetes cluster becomes a paramount concern. By using HPE GreenLake to provide a private cloud environment, organizations maintain control over their data. HPE Private Cloud AI is designed to help organizations realize the benefits of AI while maintaining privacy, security, and sovereignty over their data.

HPE AI Essentials Software offers complete access control over data sources and deployed applications. It features a built-in data lake house that integrates, organizes, and governs enterprise data with one global namespace while maintaining data integrity and compliance. In turn, organizations allow private LLMs to access their data through secure connections, helping models provide more accurate and specific guidance while maintaining privacy. Additionally, organizations maintain full autonomy and transparency over where data goes and how LLMs process it with comprehensive logging. Integrated LLM guardrails ensure compliance with organizational generative AI policies. With HPE AI Essentials Software, organizations can fully automate AI pipelines with complete lineage, traceability, and auditability.

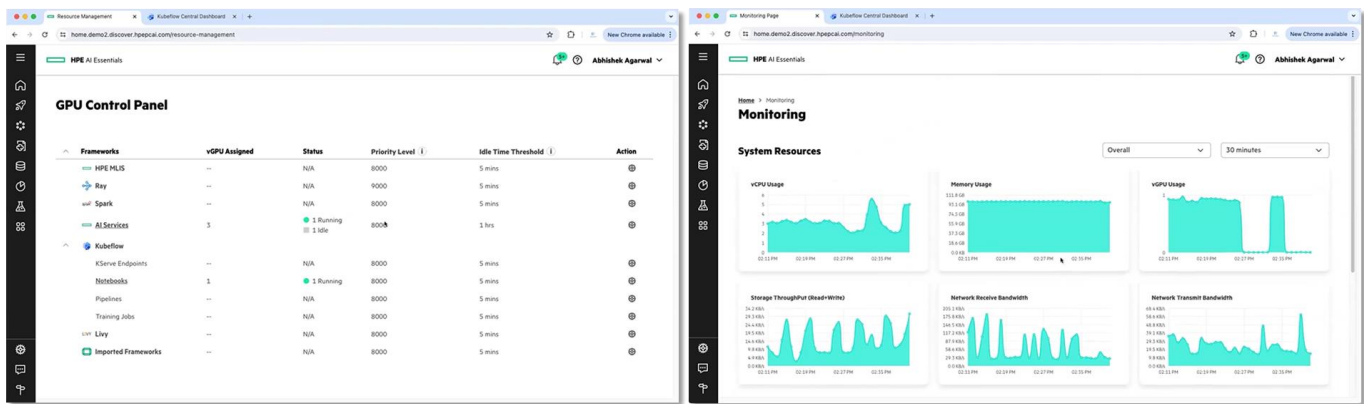
Enterprise Strategy Group research showed that organizations utilize AI-driven applications across several business areas to optimize decision-making and business value. Forty-six percent of respondents indicated that they are currently scaling and optimizing AI solutions across the business.<sup>6</sup> Doing so makes the best use of the models organizations have trained over time, increasing efficiency and reducing costs. To facilitate this transformation, HPE AI Essentials Software enables multi-tenant collaboration and end-to-end experiment tracking. Through this capability, various teams across the enterprise can see other's work and iterate accordingly, thus enabling business areas across the enterprise to collaborate on AI-driven solutions.

<sup>5</sup> Source: Enterprise Strategy Group Research Report, [Navigating Build-versus-buy Dynamics for Enterprise-ready AI](#), January 2025.

<sup>6</sup> Ibid.

If challenges arise within the infrastructure, developers and administrators have several tools at their disposal to manage them. Developers have visibility into the underlying Kubernetes cluster through Jupyter Notebooks. They can create notebooks with specific commands, with access to tools used to manage clusters. Administrators have access to performance monitoring through the HPE AI Essentials Software interface (see Figure 6). Through the GPU Control Panel, administrators can prioritize and adjust GPU allocation, enabling them to effectively manage resource contention and ensure that critical applications receive the necessary compute power. The Monitoring dashboard enables administrators to view utilization in real time and adjust accordingly.

**Figure 6.** Admin Monitoring and Management



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

However, organizations are not left alone in maintaining HPE Private Cloud AI or HPE GreenLake. HPE GreenLake features advanced network-based security, and Global HPE and partner services experts provide full-stack support and SLA response times for the entire software and hardware stack. If challenges arise, organizations can lean on HPE’s support team for help.

### Why This Matters

Organizations looking to keep their data private while utilizing trained models for AI-driven applications face several challenges. They might choose to forgo public cloud services and instead host a private cloud. However, keeping up with the infrastructure necessary to facilitate AI-driven applications is a challenge, as is the time and effort necessary to maintain it and fix any problems that arise.

HPE Private Cloud AI strikes a balance between complete autonomy and data privacy and support for challenges that arise. With HPE AI Essentials Software, customers have complete visibility into what data their LLMs are using for training and how that data is being used. All AI pipelines have complete traceability and auditability. HPE Private Cloud AI respects generative AI policies set by the organization and provides logging and visibility down to the Kubernetes cluster. Additionally, Global HPE support and partner support services provide the technical help needed when challenges within the environment arise.

With HPE Private Cloud AI, organizations maintain privacy, compliance, and security throughout the software and hardware stack, with fast help provided by HPE support resources when needed.

## Conclusion

According to Enterprise Strategy Group research, 64% of survey respondents indicated that they will invest significantly in generative AI over the next 12 months.<sup>7</sup> AI-driven applications will continue to thrive in the coming years, and as they grow more sophisticated, more businesses will attempt to reap the benefits. However, several challenges arise when organizations attempt to move AI workloads into production, stalling projects before any real business value can be realized. One such challenge is building and hosting the vast infrastructure required to make a production AI workload a reality while maintaining privacy and compliance.

HPE Private Cloud AI, a part of NVIDIA Computing by HPE, offers an alternative to organizations looking to benefit from AI without the desire or ability to manage the infrastructure required to scale and see business benefits quickly. HPE Private Cloud AI is a turnkey, end-to-end solution designed to create the infrastructure necessary for many common AI-driven use cases, enabling organizations to focus on training their models and deploying applications ready for use. The infrastructure required, including a Kubernetes cluster for container management and access to powerful NVIDIA GPUs built for AI, is hosted and managed by HPE GreenLake. The HPE Private Cloud AI software stack, HPE AI Essentials Software, provides a self-service interface through which developers can decide what models to use; integrate with industry-leading data science, analytics, and data engineering tools; and deploy applications within hours.

Enterprise Strategy Group validated that HPE Private Cloud AI enables organizations to begin seeing business value through AI quickly and efficiently, while maintaining privacy and complete visibility and control over their data. Through a three-step deployment process, organizations can deploy Solution Accelerators to create the required infrastructure for well-known use cases, such as RAG applications. Organizations maintain control of their data through secure data source connections, granular IAM controls over data sources and applications, and logging for visibility and auditability regarding the data used to train models. Through their partnership with Equinix, HPE provides a fast pilot-to-production solution for businesses that struggle to create the physical infrastructure required within their own data centers by utilizing colocation centers deployed throughout the world.

If your organization is investing in generative AI or other AI-driven applications, Enterprise Strategy Group recommends you consider HPE Private Cloud AI, an offering from NVIDIA Computing by HPE, to help accelerate your time to value for AI applications.

---

<sup>7</sup> Source: Enterprise Strategy Group Research Report: [Navigating the Generative AI Partner and Alliance Landscape](#), November 2024.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.


Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at [cr@esg-global.com](mailto:cr@esg-global.com).

---

#### About Enterprise Strategy Group

TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

 [contact@esg-global.com](mailto:contact@esg-global.com)

 [www.esg-global.com](http://www.esg-global.com)