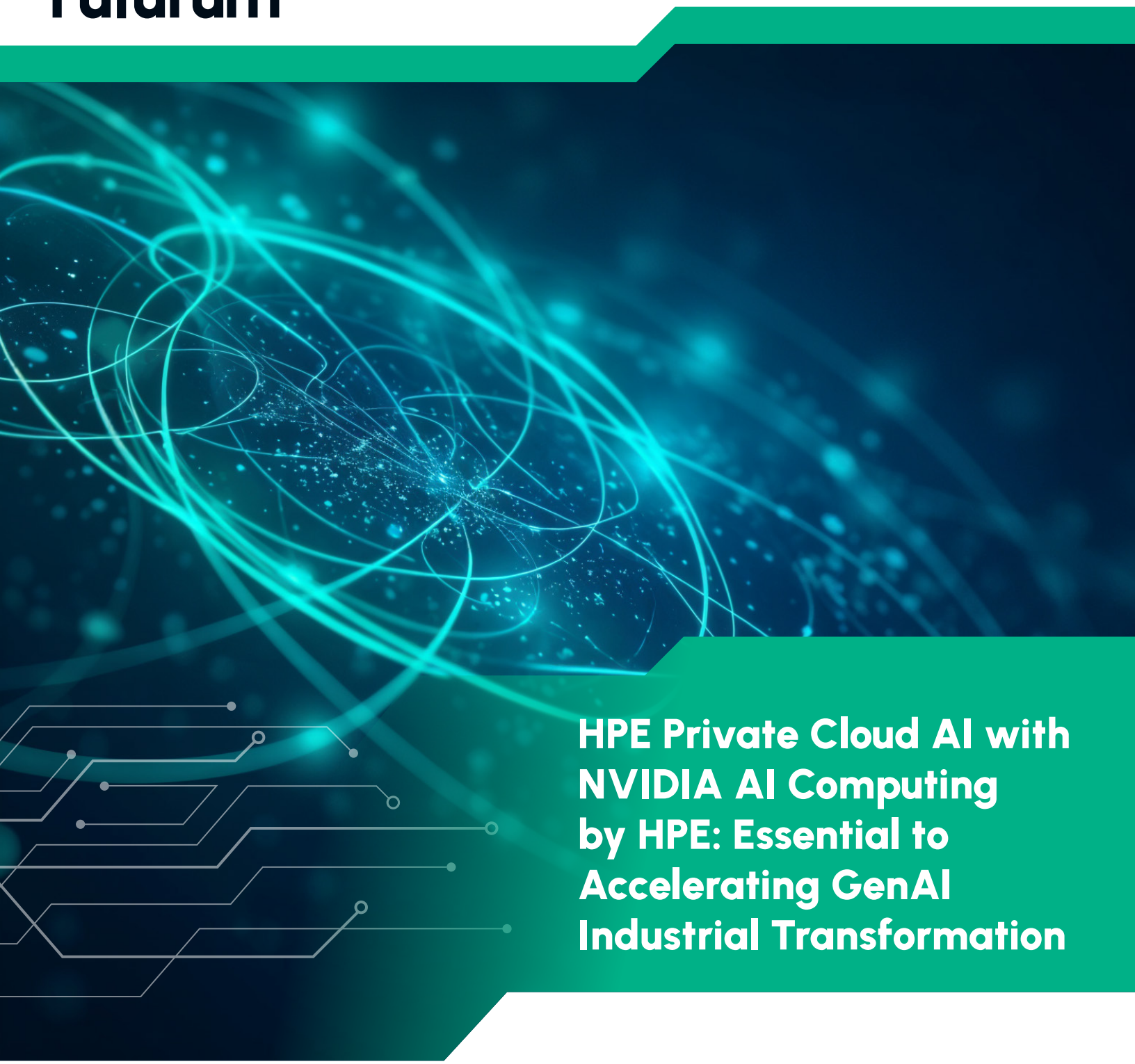
Futurum



HPE Private Cloud AI with NVIDIA AI Computing by HPE: Essential to Accelerating GenAI Industrial Transformation

AUTHORS

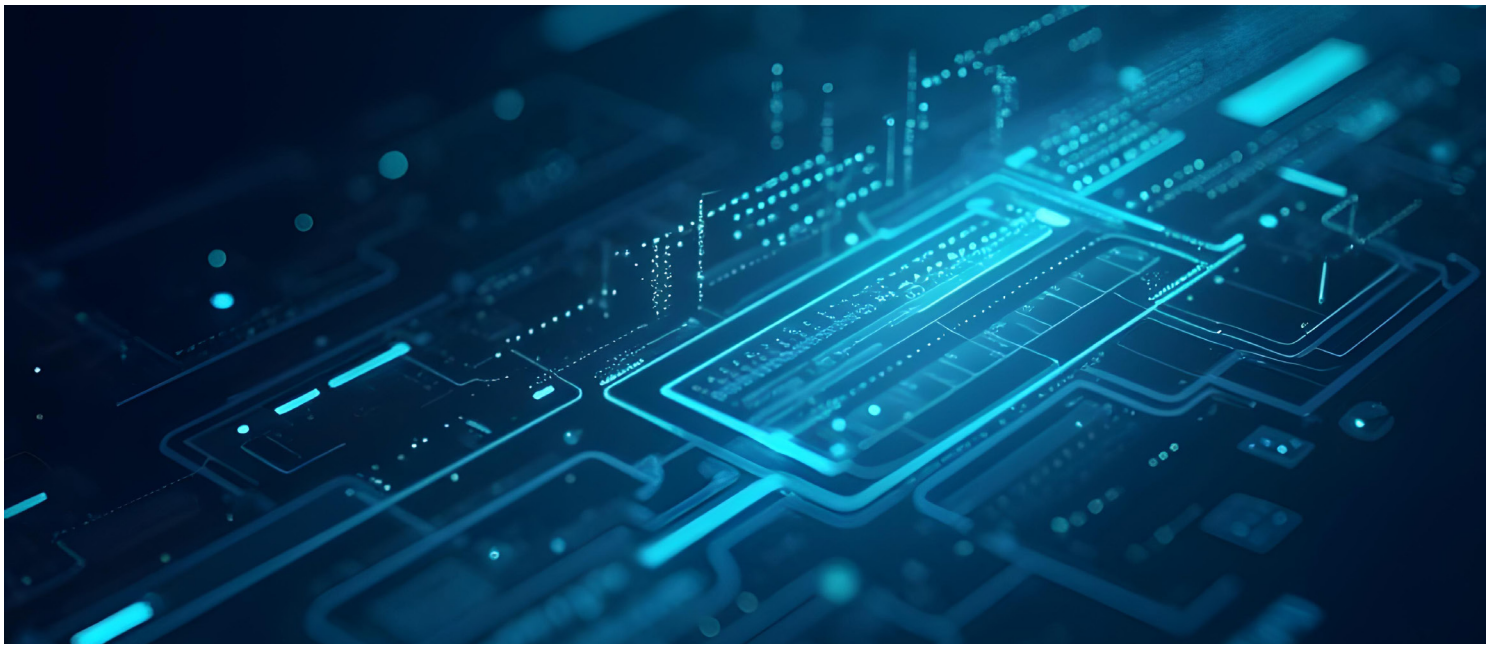
Nick Patience
VP & Practice Lead, AI | The Futurum Group

Ron Westfall
Research Director | The Futurum Group

JANUARY 2025

IN PARTNERSHIP WITH

Hewlett Packard Enterprise



Introduction

HPE Private Cloud AI Delivers Major Benefits for Organizations Looking to Use AI Alongside Controlling Their Data

The adoption of Generative AI (GenAI) across enterprises is proving to be a complex and challenging process. Many organizations are finding that pilot programs and do-it-yourself (DIY) approaches can take several months to reach productivity with numerous initiatives stalling or failing to progress beyond the experimental stage. This slow pace of implementation is often due to lack of clear objectives, data quality issues, and the need for specialized skills to effectively deploy and scale AI solutions.

Moreover, enterprises attempting to make GenAI production-ready are finding that their public cloud experiments are proving counterproductive. Not only are these initiatives failing to yield the desired results, but they are also incurring substantial expenses along the way. This dual challenge of ineffectiveness and high costs is becoming a significant concern for enterprises exploring GenAI implementations in public cloud environments.

In response to such challenges, there is a burgeoning demand for purpose-built AI solutions that directly address the unique requirements of enterprises. These solutions aim to provide easy button capabilities, streamlining the integration process and offering pre-configured tools that fully align with top priority business objectives. By adopting purpose-built solutions, enterprises can accelerate their AI adoption, reduce the risk of failed pilots, and more quickly realize the benefits of GenAI across their operations.

Data safety and governance factors are driving enterprise decision-making in selecting AI technologies and solutions. Many organizations are prioritizing on-premises private cloud solutions for their AI workloads, driven by factors such as data gravity and data sovereignty needs. For instance, most organizations (71%) plan to reevaluate their cloud workloads in 2025, focusing on optimizing placement between private and public clouds. Top drivers for workload changes include improved control and security at 55% and compliance or regulatory requirements at 20%[1].

1 Futurum Intelligence CIO Insights Survey for Q1 2025

This expanding preference for private cloud deployments is further reinforced by the limitations of public cloud options, which may not sufficiently address data protection and governance obligations. Moreover, enterprises are wary of the hidden costs associated with public cloud AI services, leading them to seek more controlled and cost-effective alternatives that ensure the security and regulatory compliance of their AI initiatives.

Enterprises need turnkey, private cloud AI solutions that enable them to ease adoption of the top AI-driven use cases integral to improving their business outcomes. Futurum Intelligence identifies these top use cases as of mid-2024 including conversational AI as the biggest use case today at 15.6%, closely followed by development Tools at 15%. Additionally, the use cases for development tools and text analysis & generation are forecast to post high CAGRs of 14.3% and 13.7% respectively, driven by the high demand for tools that enable the development of multimodal AI solutions (integrating text, images, and audio), as well as those with MLOps capabilities that streamline model deployment, monitoring, and lifecycle management for enterprises[2].

Section 1: HPE Private Cloud AI Advantages

HPE Private Cloud AI directly addresses the top challenges enterprises face in adopting AI across their entire organization. HPE Private Cloud AI provides a scalable, pre-tested, and AI-optimized private cloud solution. It empowers IT and AI teams to experiment, iterate, and scale AI projects efficiently. By leveraging a comprehensive ecosystem of AI models and development tools, organizations can maintain control over costs and mitigate financial risks.Co-developed with NVIDIA, this turnkey private cloud can enable enterprises to focus on developing new AI use cases that boost productivity and unlock new revenue streams.

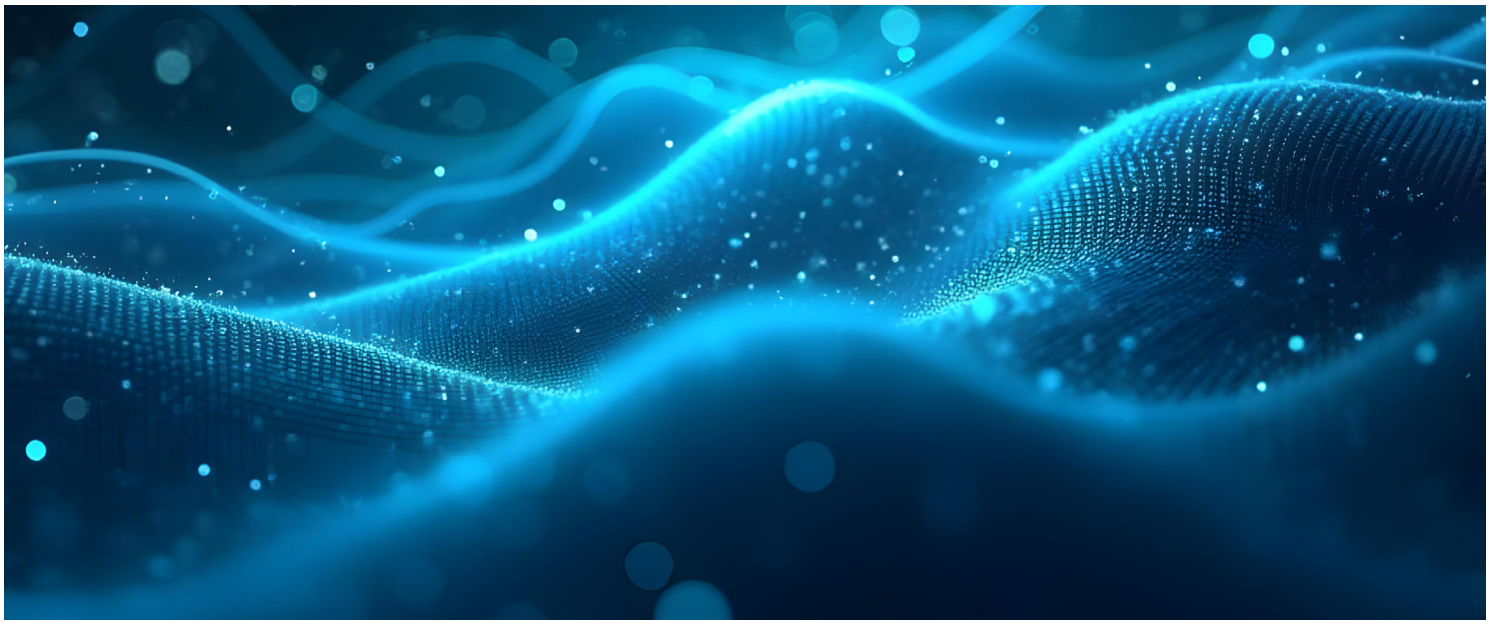
HPE Private Cloud AI provides instant AI productivity that can prove transformative for organizations looking to accelerate their time to value in the realm of AI. By providing self-service experience to essential AI tools, companies can dramatically speed up developer activity, with productivity gains of up to 90%. This approach empowers data scientists and AI practitioners to quickly experiment, iterate, and deploy AI models without the traditional bottlenecks associated with resource allocation and tool procurement. The result is a significant reduction in the time it takes to move from concept to production, allowing businesses to realize the benefits of AI investments more rapidly.

2 Futurum Intelligence Q2 2024 AI Software/Tools Market Report

Enterprise-grade confidence and control are paramount when implementing AI solutions at scale. A private cloud experience tailored for AI workloads offers the ability to manage, secure, and govern data, models, and infrastructure with the rigor required by large organizations. This level of control ensures that AI initiatives comply with regulatory requirements, maintain data integrity, and align with corporate governance policies. Moreover, NVIDIA AI Enterprise provides a comprehensive software suite for AI development and deployment, while NIM inference microservices offer pre-trained models that can be easily customized for specific enterprise needs. By providing a robust framework for AI development and deployment, companies can mitigate risks associated with data breaches, model drift, and unauthorized access, fostering trust in AI-driven decision-making processes across the enterprise.

Unifying access to all data is a critical component of successful AI implementations. By enabling access to data across the enterprise through a single global namespace, organizations can break down data silos and create a cohesive data ecosystem. This approach, often facilitated by an embedded data lakehouse architecture, enables AI teams to work with diverse data types and sources without the need for complex data integration projects. The result is a more agile and responsive AI development environment, where insights can be derived from a comprehensive view of the organization's data assets, leading to more accurate AI models that improve business outcomes.

As organizations increasingly prioritize data privacy alongside AI innovation, a cloud strategy that guarantees data protection is crucial. By deploying AI infrastructure on-premises while designing for hybrid scenarios, companies can maintain strict control over sensitive data while still leveraging cloud technologies and economics. Such a modular and flexible approach allows organizations to scale their AI capabilities as needed, taking advantage of built-in cloud efficiencies without compromising on data sovereignty or security. The result produces a best-of-both-worlds scenario where businesses can accelerate their AI initiatives while maintaining the highest standards of data privacy and compliance.



Section 2: HPE Private Cloud AI - Business Strengths

HPE GreenLake is integral to HPE Private Cloud AI, offering a comprehensive approach to data privacy and security and ensuring that sensitive information remains protected on-premises. By implementing multi-layered controls, organizations can safeguard their data and models continuously, maintaining reliability and performance without compromising security. This approach allows businesses to leverage the power of AI and advanced analytics while keeping their valuable assets under tight control. Furthermore, the HPE GreenLake dashboard manages the entire AI lifecycle by providing a unified cloud experience across all stages, from data preparation to model deployment and inferencing.

HPE Private Cloud AI offers a flexible ecosystem of NVIDIA and HPE AI models and tools, supported by scalable and pretested infrastructure. This flexibility allows organizations to experiment with various AI projects across a diverse range of models and development tools. HPE Private Cloud AI enables the seamless integration of custom and ISV AI tools and frameworks, allowing organizations to leverage their existing investments and expertise.By providing a solid foundation for AI initiatives, HPE Private Cloud AI enables businesses to scale their AI capabilities efficiently, adapting to evolving technological landscapes and business needs.

Deloitte's C-Suite AI for CFOs with NVIDIA AI and HPE Private Cloud AI uses GenAI to reimagine stale, static executive reporting to become a dynamic, on-demand and interactive experience. CFOs and executive decision makers can retrieve timely and relevant financial information, insightful analysis and alerts to what's trending, scenario analysis and explainability of what might happen and why.

Section 3: Unique NVIDIA Partnership in Co-Developing Private Cloud AI Portfolio

HPE and NVIDIA have strategically expanded their partnership to offer NVIDIA AI Computing by HPE. This comprehensive portfolio of AI solutions and joint go-to-market strategies aims to accelerate the adoption of GenAI within enterprises. The collaboration delivers a deep integration of NVIDIA's advanced AI computing capabilities with HPE's robust infrastructure and cloud solutions.

HPE Private Cloud AI integrates NVIDIA AI computing, networking, and software with HPE's AI storage, compute, and the HPE GreenLake cloud platform. Through flexible consumption models, organizations can transform the way they approach AI adoption and growth, easily adjusting resources to minimize upfront costs while maintaining the agility to respond to changing market demands. These models provide a fast path to consuming technology flexibly, allowing businesses to elastically scale their AI capabilities as opportunities arise and requirements evolve. This approach is particularly valuable in the rapidly evolving field of AI, where the ability to quickly access and deploy cutting-edge technologies can be a significant competitive advantage.

Built-in ITOps and AIOps capabilities are enhancing the cloud experience and boosting productivity for IT teams. These integrated features automate routine tasks, provide predictive maintenance, and offer intelligent insights into system performance. By leveraging machine learning and advanced analytics, AIOps can identify patterns and anomalies in real-time, enabling proactive problem-solving and optimizing resource allocation. This not only reduces the workload on IT staff but also improves overall system reliability and efficiency, creating a more streamlined and responsive IT environment.

NVIDIA AI Enterprise Software is accelerating data science pipelines and simplifying the development and deployment of production-grade copilots and generative AI applications. This comprehensive suite includes NVIDIA NIM, which enable rapid scaling of AI models in production environments. By providing optimized frameworks, pre-trained models, and development tools, the solution empowers data scientists and developers to bring AI solutions to market faster. The inclusion of NIM inference microservices further enhances the ability to deploy and manage AI at scale, making it easier for organizations to integrate advanced AI capabilities into their existing workflows and applications.

HPE AI Essentials complements NVIDIA AI Enterprise Software by offering a ready-to-run set of curated AI and data foundation tools. This software package provides a unified control plane that delivers adaptable solutions, ongoing enterprise support, and trusted AI services. By augmenting NVIDIA's offerings, HPE AI Essentials creates a comprehensive ecosystem for AI development and deployment. The combination of these tools enables organizations to build, train, and deploy AI models more efficiently, with the added benefits of enterprise-grade support and security. This integrated approach helps businesses navigate the complexities of AI implementation while ensuring they have access to the latest advancements in AI technology.

This comprehensive suite of offerings underscores HPE's commitment to simplifying AI adoption for enterprises. With a focus on ease of use, HPE claims that its Private Cloud AI can be set up with just a few clicks, providing a seamless, self-service cloud experience. The solution supports both standalone on-premises deployment and hybrid models, offering flexibility to meet diverse business needs.

Section 4: HPE GreenLake Observability and Management Benefits

HPE GreenLake introduced a self-service cloud feature that transforms hybrid environment management by offering a single, unified control plane that provides comprehensive manageability and observability across endpoints, workloads, and data. This approach enables automation, orchestration, and management of resources throughout hybrid environments, empowering organizations to streamline their operations and enhance overall efficiency.

In response to growing environmental concerns, HPE GreenLake has incorporated sustainability metrics into its platform. These metrics allow organizations to monitor and optimize the environmental impact of their workloads and endpoints. By providing visibility into sustainability-related data, HPE GreenLake enables businesses to make informed decisions that align with their environmental goals and corporate social responsibility initiatives.

In our view, the integration of OpsRamp's IT operations with HPE GreenLake cloud marks a significant advancement in infrastructure management. This combination brings advanced observability and AIOps capabilities to all HPE products and services. The OpsRamp AI Infrastructure Observability and Copilot Assistant enhance the platform's ability to monitor, analyze, and optimize IT operations, providing users with powerful tools to maintain peak performance and quickly resolve issues.

OpsRamp's observability capabilities have been extended to cover the entire NVIDIA-accelerated computing stack. This comprehensive coverage includes NVIDIA NIM/AI software, NVIDIA Tensor Core GPUs, and AI clusters. By providing end-to-end visibility into this advanced technology stack, OpsRamp enables organizations to maximize the performance and efficiency of their NVIDIA-powered infrastructure, ensuring optimal utilization of these powerful computing resources.

To further boost productivity in operations management, HPE GreenLake uses NVIDIA's accelerated computing platform for analyzing large datasets. This integration includes a conversational assistant that expedites intuitive interaction with the data analysis tools. By cultivating the power of NVIDIA's technology, organizations can quickly derive valuable insights from their data, leading to more informed decision-making and improved operational efficiency. This productivity-enhancing feature demonstrates HPE GreenLake's commitment to providing practical tools that drive business success in the modern data-driven landscape.

Section 5: Conclusions & Recommendations

We believe that HPE Private Cloud AI is at the forefront of digital transformation, enhancing capabilities in AI, hybrid cloud, and mission-critical workloads. This innovative platform enables organizations to rapidly deploy GenAI applications, streamlining their operations and accelerating their journey toward AI integration.

Recommendations and Call to Action

- **Data Privacy and Security:** HPE Private Cloud AI merits top consideration by enterprises due to its ability to ensure that data remains private and secure by keeping it on-premises, offering multi-layered controls to protect data and models and ensuring reliability and performance.

- **HPE Private Cloud AI's Competitive Advantages:** IT decision-makers should consider implementing HPE Private Cloud AI across hybrid cloud and critical workload environments to achieve a secure, AI-native network. This solution is purpose-built with AI and for AI, offering a comprehensive approach to using AI in network infrastructure and operations.

- **Productivity Gains:** NVIDIA AI Computing by HPE leverages NVIDIA's accelerated computing platform that ensures the analysis of large datasets through a conversational assistant, enhancing productivity in operations management. Plus, HPE AI Essentials complement NVIDIA AI Enterprise and NIM by offering a curated set of AI and data tools with a unified control plane, providing adaptable solutions, enterprise support, and trusted AI services that are key to assuring productivity gains.

Important Information About this Report

CONTRIBUTORS

Nick Patience
VP & Practice Lead, AI | The Futurum Group

Ron Westfall
Research Director | The Futurum Group

PUBLISHER

Daniel Newman
CEO | The Futurum Group

INQUIRIES

Contact us if you would like to discuss this report and The Futurum Group will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "The Futurum Group." Non-press and non-analysts must receive prior written permission by The Futurum Group for any citations

LICENSING

DISCLOSURES

The Futurum Group provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

Hewlett Packard Enterprise

ABOUT HPE

HPE combines technology insights, financial expertise, and a deep rooted focus on sustainability to create smarter IT lifecycles for customers and partners of all sizes. Working across the entire tech estate, from edge to cloud to end-user, our collaborative approach delivers asset management solutions that not only free up capital and maximize capacity, but also advance sustainable practices globally and consistently. For more information, visit: hpe.com

Futurum

ABOUT THE FUTURUM GROUP

The Futurum Group is an independent research, analysis, and advisory firm, focused on digital innovation and market-disrupting technologies and trends. Every day our analysts, researchers, and advisors help business leaders from around the world anticipate tectonic shifts in their industries and leverage disruptive innovation to either gain or maintain a competitive advantage in their markets.

Futurum

CONTACT INFORMATION

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |