

NVIDIA and VMware to Accelerate Machine Learning, Data Science and AI Workloads on VMware Cloud on AWS Accelerated by NVIDIA GPUs

Highly scalable, more secure hybrid cloud platform to help enterprises migrate, modernize and deploy next-generation applications

VMworld—NVIDIA and VMware today announced their intent to deliver accelerated GPU services for VMware Cloud on AWS to power modern enterprise applications, including AI, machine learning and data analytics workflows. These services will enable customers to seamlessly migrate VMware vSphere-based applications and containers to the cloud, unchanged, where they can be modernized to take advantage of high-performance computing, machine learning, data analytics and video processing applications.

Increasingly businesses are applying artificial intelligence (AI) technologies to differentiate and advance their processes and offerings. Enterprises are rapidly adopting AI⁽¹⁾ and implementing new AI strategies that require powerful computers to create predictive models from petabytes of corporate data. Across industries, enterprises are implementing machine learning applications such as image and voice recognition, advanced financial modeling and natural language processing using neural networks that rely on NVIDIA GPUs for faster training and real-time inference. Additionally, VMware recently acquired Bitfusion, which enables VMware to make GPU capabilities efficiently available for AI and machine learning workloads in the enterprise.

Through this partnership, VMware Cloud on AWS customers will gain access to a new, highly scalable and secure cloud service consisting of Amazon EC2 bare metal instances to be accelerated by NVIDIA T4 GPUs and new [NVIDIA Virtual Compute Server \(vCS\) software](#).

“From operational intelligence to artificial intelligence, businesses rely on GPU-accelerated computing to make fast, accurate predictions that directly impact their bottom line,” said Jensen Huang, founder and CEO, NVIDIA. “Together with VMware, we’re designing the most advanced GPU infrastructure to foster innovation across the enterprise, from virtualization, to hybrid cloud, to VMware’s new Bitfusion data center disaggregation.”

“Our customers are embracing the unique value of VMware Cloud on AWS to accelerate the migration and modernization of business-critical applications,” said Pat Gelsinger, CEO, VMware. “Through new innovations driven by partnerships we have with industry leaders such as NVIDIA and AWS, we will bring best-in-class GPU acceleration services for the most intense data-driven workloads and modern applications across the hybrid cloud.”

Benefits of VMware Cloud on AWS with NVIDIA GPU for AI, ML and Data Analytics

Once available businesses will be able to leverage an enterprise-grade hybrid cloud platform to accelerate application modernization. They will be able to unify deployment, migration and operations across a consistent VMware infrastructure from data center to the AWS cloud in support of most compute-intensive workloads, including AI, machine learning and data analytics. Benefits will include:

- **Seamless portability:** Customers will be able to move workloads powered by NVIDIA Virtual Compute Server software and GPUs with a single click of a button and no downtime using VMware HCX. This will give customers more choice and flexibility to execute training and inference in the cloud or on-premises.
- **Elastic AWS infrastructure:** With the ability to automatically scale VMware Cloud on AWS clusters accelerated by NVIDIA T4, administrators will be able to grow or shrink available training environments depending on the needs of their data scientists.
- **Accelerated computing for modern applications:** NVIDIA T4 GPUs feature Tensor Cores for acceleration of deep learning inference workflows. When these are combined with [Virtual Compute Server](#) software for GPU virtualization businesses have the flexibility to run GPU-accelerated workloads like AI, machine learning and data analytics in virtualization environments for improved security, utilization and manageability.
- **Consistent Hybrid Cloud Infrastructure and Operations:** With VMware Cloud on AWS, organizations can establish consistent infrastructure and consistent operations across the hybrid cloud, leveraging VMware industry-standard vSphere, vSAN and NSX as a foundation for modernizing business-critical applications. IT operators will be able to manage GPU-accelerated workloads within vCenter, right alongside GPU-accelerated workloads running on vSphere on-premises.
- **Seamless, end-to-end data science and analytics pipeline:** The NVIDIA T4 data center GPU supercharges mainstream servers and accelerates data science techniques using [NVIDIA RAPIDS™](#), a collection of NVIDIA GPU acceleration libraries for data science including deep learning, machine learning and data analytics.

1. Gartner, "AI and ML Development Strategies," July 15, 2019.

About NVIDIA

NVIDIA (NASDAQ: NVDA) is a computer technology company that has pioneered GPU-accelerated computing. It targets the world's most demanding users — gamers, designers and scientists — with products, services and software that power amazing experiences in virtual reality, artificial intelligence, professional visualization and autonomous cars. More information at <http://nvidianews.nvidia.com/>.

About VMware

VMware software powers the world's complex digital infrastructure. The company's cloud, networking and security, and digital workspace offerings provide a dynamic and efficient digital foundation to customers globally, aided by an extensive ecosystem of partners. Headquartered in Palo Alto, California, VMware is committed to being a force for good, from its breakthrough innovations to its global impact. For more information, please visit <https://www.vmware.com/company.html>.

Certain statements in this press release including, but not limited to, statements as to: the benefits and impact of the partnership between NVIDIA and VMware to deliver GPU services for VMware Cloud, including providing a platform to help enterprises migrate, modernize and deploy next-generation applications, and customers gaining access to new cloud services; the GPU services offered by NVIDIA and VMware enabling customers to migrate VMware applications and containers to the cloud where they can be modernized; enterprises adopting and implementing AI, AI strategies, and machine learning applications; businesses relying on GPU-accelerated computing to make predictions that impact their bottom line; NVIDIA and VMware designing the most advanced and high performing GPU-accelerated hybrid cloud infrastructure to foster innovation across the enterprise; VMware customers embracing the VMware Cloud on AWS to accelerate the migration and modernization of business-critical applications; VMware, through its partnerships, bringing best-in-class GPU acceleration services for data-driven workloads and modern applications across the hybrid cloud; the benefits of VMware Cloud on AWS with NVIDIA GPUs, including businesses ability to leverage an enterprise-grade hybrid cloud platform to accelerate application modernization, unifying deployment, migration and operations across a VMware infrastructure, customers being able to move workloads quickly and it giving them more choice and flexibility, administrators being able to grow or shrink training environments depending on the needs of their data scientists, businesses flexibility running GPU-accelerated workloads, organizations establishing consistent infrastructure and operations across the hybrid cloud and ability to manage GPU-accelerated workloads within vCenter alongside vSphere, and NVIDIA GPUs supercharging servers and accelerating data science techniques are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

© 2019 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. VMware and VMware Cloud are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and other jurisdictions. Other company and product names may be trademarks of the respective companies with which they are associated. Features, pricing, availability and specifications are subject to change without notice. This article may contain hyperlinks to non-VMware websites that are created and maintained by third parties who are solely responsible for the content on such websites.

Gail Laguna
Professional Visualization
NVIDIA Corp.
+1-408-386-2435
glaguna@nvidia.com
Roger T. Fortier
VMware
+1-408-348-1569
rfortier@vmware.com