

World's Largest Server Companies Announce NVIDIA Volta Systems Supercharged for AI

Dell EMC, Hewlett Packard Enterprise, IBM, Supermicro Announce Servers Based on NVIDIA Tesla V100 Accelerators -- World's Most Advanced Data Center GPU

SANTA CLARA, CA - NVIDIA (NASDAQ: NVDA) and its systems partners Dell EMC, Hewlett Packard Enterprise, IBM and Supermicro today unveiled more than 10 servers featuring NVIDIA® [Volta architecture](#)-based [Tesla® V100](#) GPU accelerators -- the world's most advanced GPUs for AI and other compute-intensive workloads.

NVIDIA V100 GPUs, with more than 120 teraflops of deep learning performance per GPU, are uniquely designed to deliver the computing performance required for AI deep learning training and inferencing, high performance computing, accelerated analytics and other demanding workloads. A single Volta GPU offers the equivalent performance of 100 CPUs, enabling data scientists, researchers and engineers to tackle challenges that were once impossible.

Seizing on the AI computing capabilities offered by NVIDIA's latest GPUs, Dell EMC, HPE, IBM and Supermicro are bringing to the global market a broad range of multi-V100 GPU systems in a variety of configurations.

"Volta systems built by our partners will ensure that enterprises around the world can access the technology they need to accelerate their AI research and deliver powerful new AI products and services," said Ian Buck, vice president and general manager of Accelerated Computing at NVIDIA.

V100-based systems announced include:

- [Dell EMC](#) -- The PowerEdge R740 supporting up to three V100 GPUs for PCIe, the PowerEdge R740XD supporting up to three V100 GPUs for PCIe, and the PowerEdge C4130 supporting up to four V100 GPUs for PCIe or four V100 GPUs for NVIDIA [NVLink™ interconnect technology](#) in an SXM2 form factor.
- [HPE](#) -- [HPE Apollo 6500](#) supporting up to eight V100 GPUs for PCIe and [HPE ProLiant DL380](#) systems supporting up to three V100 GPUs for PCIe.
- [IBM](#) -- The next generation of [IBM Power Systems](#) servers based on the POWER9 processor will incorporate multiple V100 GPUs and take advantage of the latest generation NVLink interconnect technology -- featuring fast GPU-to-GPU interconnects and an industry-unique [OpenPOWER](#) CPU-to-GPU design for maximum throughput.
- [Supermicro](#) -- Products supporting the new Volta GPUs include a 7048GR-TR workstation for all-around high-performance GPU computing, 4028GR-TXRT, 4028GR-TRT and 4028GR-TR2 servers designed to handle the most demanding deep learning applications, and 1028GQ-TRT servers built for applications such as advanced analytics.

These partner systems complement an [announcement yesterday](#) by China's leading original equipment manufacturers -- including Inspur, Lenovo and Huawei -- that they are using the Volta architecture for accelerated systems for hyperscale data centers.

Additional NVIDIA V100 Details

Each NVIDIA V100 GPU features over 21 billion transistors, as well as 640 Tensor Cores, the latest NVLink high-speed interconnect technology, and 900 GB/sec HBM2 DRAM to achieve 50 percent more memory bandwidth than previous generation GPUs.

V100 GPUs are supported by NVIDIA Volta-optimized software, including [CUDA® 9.0](#) and the newly updated deep learning SDK, including [TensorRT™ 3](#), [DeepStream SDK](#) and [cuDNN 7](#) as well as all major [AI frameworks](#). Additionally, hundreds of thousands of GPU-accelerated applications are available for accelerating a variety of data-intensive workloads, including AI training and inferencing, high performance computing, graphics and advanced data analytics.

Partner Quotes

"One of the core principles for Dell EMC is to deliver differentiated solutions to our customers so that they can leverage the most advanced technology for a competitive advantage. To that end, we are proud of the work we do with partners like NVIDIA to build PowerEdge servers ideal for compute-intensive workloads including data analytics, high-performance computing, machine learning and AI."

-- Armughan Ahmad, senior vice president and general manager of Hybrid Cloud and Ready Solutions at Dell EMC

"As deep learning continues to become more pervasive, technology advancements across systems and accelerators need to evolve in order to gain intelligence from large datasets faster than ever before. The HPE Apollo 6500 and HPE ProLiant DL380 systems combine the industry-leading GPU performance of NVIDIA Tesla V100 GPU accelerators and Volta architecture with HPE unique innovations in system design and manageability to deliver unprecedented levels of performance, scale and efficiency for high performance computing and artificial intelligence applications."

-- Bill Mannel, vice president and general manager of High Performance Computing and Artificial Intelligence at Hewlett Packard Enterprise

"IBM's upcoming POWER9 servers will support NVIDIA's Volta GPU, and will be the only one to support the latest generation of NVLink and PCIe 4.0, which will deliver maximum throughput. With accelerators like Volta, IBM will scale deep learning performance to new heights."

-- Brad McCredie, vice president and IBM Fellow, Cognitive Systems Development at IBM

"Supermicro designs the most application-optimized GPU systems and offers the widest selection of GPU-optimized servers and workstations in the industry. Our high performance computing solutions enable deep learning, engineering and scientific fields to scale out their compute clusters to accelerate their most demanding workloads and achieve fastest time-to-results with maximum performance per watt, per square foot and per dollar. With our latest innovations incorporating the new NVIDIA V100 PCI-E and V100 SXM2 GPUs in performance-optimized 1U and 4U architectures with next-generation NVLink, our customers can accelerate their applications and innovations to help solve the world's most complex and challenging problems."

-- Charles Liang, president and CEO of Supermicro

Keep Current on NVIDIA

Subscribe to the [NVIDIA blog](#), follow us on [Facebook](#), [Google+](#), [Twitter](#), [LinkedIn](#) and [Instagram](#), and view NVIDIA videos on [YouTube](#) and images on [Flickr](#).

About NVIDIA

[NVIDIA](#)'s (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI -- the next era of computing -- with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at <http://nvidianews.nvidia.com/>.

Certain statements in this press release including, but not limited to, statements as to: Dell EMC, HPE, IBM and Supermicro adopting servers featuring NVIDIA V100 GPU systems; and the performance, benefits, efficiency and abilities of NVIDIA V100 GPUs, Volta systems and GPU-accelerated applications 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 reports NVIDIA files with the Securities and Exchange Commission, or SEC, including its Form 10-Q for the fiscal period ended July 30, 2017. 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.

© 2017 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, CUDA, DGX, NVLink, TensorRT and Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. 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.

Kristin Bryson
Enterprise Data Center, AI/DL
+1-203-241-9190
kbryson@nvidia.com