



NVIDIA Tesla P100 Supercharges HPC Applications by More Than 30X

Powered by Pascal Architecture, Tesla P100 Delivers Massive Leap in Data Center Throughput

ISC16 - To meet the unprecedented computational demands placed on modern data centers, NVIDIA today introduced the [NVIDIA® Tesla® P100 GPU accelerator](#) for PCIe servers, which delivers massive leaps in performance and value compared with CPU-based systems.

Demand for supercomputing cycles is higher than ever. The majority of scientists are unable to secure adequate time on supercomputing systems to conduct their research, based on National Science Foundation data.¹ In addition, [high performance computing \(HPC\) technologies](#) are increasingly required to power computationally intensive [deep learning applications](#), while researchers are applying AI techniques to drive advances in traditional scientific fields.

The Tesla P100 GPU accelerator for PCIe meets these computational demands through the unmatched performance and efficiency of the [NVIDIA Pascal™ GPU architecture](#). It enables the creation of "super nodes" that provide the throughput of more than 32 commodity CPU-based nodes and deliver up to 70 percent lower capital and operational costs.²

"Accelerated computing is the only path forward to keep up with researchers' insatiable demand for HPC and AI supercomputing," said Ian Buck, vice president of accelerated computing at NVIDIA. "Deploying CPU-only systems to meet this demand would require large numbers of commodity compute nodes, leading to substantially increased costs without proportional performance gains. Dramatically scaling performance with fewer, more powerful Tesla P100-powered nodes puts more dollars into computing instead of vast infrastructure overhead."

The Tesla P100 for PCIe is available in a standard PCIe form factor and is compatible with today's GPU-accelerated servers. It is optimized to power the most computationally intensive AI and HPC data center applications. A single Tesla P100-powered server delivers higher performance than 50 CPU-only server nodes when running the AMBER molecular dynamics code,³ and is faster than 32 CPU-only nodes when running the VASP material science application.⁴

Later this year, Tesla P100 accelerators for PCIe will power an upgraded version of Europe's fastest supercomputer, the [Piz Daint system](#) at the Swiss National Supercomputing Center in Lugano, Switzerland.

"Tesla P100 accelerators deliver new levels of performance and efficiency to address some of the most important computational challenges of our time," said Thomas Schulthess, professor of computational physics at ETH Zurich and director of the Swiss National Supercomputing Center. "The upgrade of 4,500 GPU-accelerated nodes on Piz Daint to Tesla P100 GPUs will more than double the system's performance, enabling researchers to achieve breakthroughs in a range of fields, including cosmology, materials science, seismology and climatology."

The Tesla P100 for PCIe is the latest addition to the [NVIDIA Tesla Accelerated Computing Platform](#). Key features include:

- **Unmatched application performance for mixed-HPC workloads** -- Delivering 4.7 teraflops and 9.3 teraflops of double-precision and single-precision peak performance, respectively, a single Pascal-based Tesla P100 node provides the equivalent performance of more than 32 commodity CPU-only servers.
- **CoWoS with HBM2 for unprecedented efficiency** -- The Tesla P100 unifies processor and data into a single package to deliver unprecedented compute efficiency. An innovative approach to memory design -- chip on wafer on substrate (CoWoS) with HBM2 -- provides a 3x boost in memory bandwidth performance, or 720GB/sec, compared to the [NVIDIA Maxwell™ architecture](#).
- **PageMigration Engine for simplified parallel programming** -- Frees developers to focus on tuning for higher performance and less on managing data movement, and allows applications to scale beyond the GPU physical memory size with support for virtual memory paging. Unified memory technology dramatically improves productivity by enabling developers to see a single memory space for the entire node.
- **Unmatched application support** -- With 410 GPU-accelerated applications, including nine of the top 10 HPC applications, the Tesla platform is the world's leading HPC computing platform.

Tesla P100 for PCIe Specifications

- 4.7 teraflops double-precision performance, 9.3 teraflops single-precision performance and 18.7 teraflops half-precision performance with NVIDIA GPU BOOST™ technology
- Support for PCIe Gen 3 interconnect (32GB/sec bi-directional bandwidth)
- Enhanced programmability with Page Migration Engine and unified memory
- ECC protection for increased reliability
- Server-optimized for highest data center throughput and reliability

- Available in two configurations:
 - 16GB of CoWoS HBM2 stacked memory, delivering 720GB/sec of memory bandwidth
 - 12GB of CoWoS HBM2 stacked memory, delivering 540GB/sec of memory bandwidth
- 16GB of CoWoS HBM2 stacked memory, delivering 720GB/sec of memory bandwidth
- 12GB of CoWoS HBM2 stacked memory, delivering 540GB/sec of memory bandwidth

Availability

NVIDIA Tesla P100 GPU accelerator for PCIe-based systems is expected to be available beginning in Q4 2016 from NVIDIA reseller partners and server manufacturers, including Cray, Dell, Hewlett Packard Enterprise, IBM, Lenovo and SGI.

Additional Resources

[Boost throughput for HPC](#) (video)

[Pascal deep dive](#) (blog)

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](#) (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/>.

1 Source: <https://portal.xsede.org/#/gallery>

2 CPU server: Dual socket Intel E5-2680v3 12 cores, 128GB DDR4 per node, FDR IB / GPU server: 8x Tesla P100 for PCIe with dual-socket Intel E5-2680v3

3 Simulations on SDSU Comet supercomputer.

4 VASP 5.4.1_05Feb16, Si-Huge Dataset. Sixteen, 32 nodes are estimated based on same scaling from four to eight nodes.

Certain statements in this press release including, but not limited to, statements as to: the impact, performance, benefits and availability of the NVIDIA Tesla P100 GPU accelerator; high performance computing technologies being increasingly required, and the role of accelerated computing for HPC and AI supercomputing 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 May 1, 2016. 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.

© 2016 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Tesla, Pascal, Maxwell, and NVIDIA GPU BOOST 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.

Hector Martinez
Corporate Communications
+1-408-486-3443
hmarinez@nvidia.com