# NVIDIA HGX-2 GPU-Accelerated Platform Gains Broad Adoption

## Baidu, Tencent Adopt HGX-2 to Build More Powerful AI Services; Inspur, Lenovo, Huawei, Sugon Introduce and Support New HGX-2 Based Servers

GTC China -- NVIDIA today announced broad adoption of the NVIDIA HGX-2™ server platform, the world's most powerful accelerated server platform, for AI deep learning, machine learning and high performance computing.

Delivering two petaflops of compute performance in a single node, NVIDIA HGX-2 can run AI machine learning workloads nearly 550x faster, AI deep learning workloads nearly 300x faster and HPC workloads nearly 160x faster, all compared to a CPU-only server.

Leading technology companies around the world are taking advantage of HGX-2's record-setting performance. The newest adoptions announced today at GTC China include:

- Baidu and Tencent are using HGX-2 for a wide range of even more powerful AI services, both for their internal use and for their cloud customers.
- Inspur is China's first to build an HGX-2 server. Its Inspur AI Super-Server AGX-5 is designed to solve the performance expansion problem of AI, deep learning and HPC.
- Huawei, Lenovo and Sugon announced that they have become NVIDIA HGX-2 cloud server platform partners.

Previously announced HGX-2 support and adoption comes from leading global server makers, including Foxconn, Inventec, QCT, Quanta, Supermicro, Wistron and Wiwynn. Additionally, Oracle announced last month its plans to bring the NVIDIA HGX-2 platform to Oracle Cloud Infrastructure in both bare-metal and virtual machine instances, giving customers access to a unified HPC and AI computing architecture.

“Leading technology companies are quickly taking advantage of HGX-2, the most powerful cloud node in history," said Ian Buck, vice president and general manager of Accelerated Computing at NVIDIA. “With HGX-2's unmatched compute power and versatile design, companies around the world are able to build new products and services that can scale to solve immense compute challenges and address some of the world's most pressing problems."

The HGX-2 cloud server platform features multi-precision computing capabilities, providing unique flexibility to support the future of computing. It fuses high-precision FP64 and FP32 for accurate HPC, while also enabling faster, reduced-precision FP16 and INT8 for deep learning and machine learning.

The HGX-2 delivers unmatched compute power. It incorporates such breakthrough features as NVIDIA NVSwitch™ interconnect fabric, which seamlessly links 16 NVIDIA® Tesla® V100 Tensor Core GPUs to work as a single, giant GPU delivering two petaflops of AI performance. It also provides 0.5TB of memory and 16TB/s of aggregate memory bandwidth.

**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: NVIDIA HGX-2 being the world’s most powerful server platform and it gaining broad adoption; the use of HGX-2 to build more powerful AI services; and the introduction of HGX-2 based servers; the benefits, performance, features and abilities of NVIDIA’s HGX-2; leading technology companies taking advantage of HGX-2; the companies and reasons why they are adopting HGX-2, including for internal use, AI services, for cloud customers, and for deep learning, machine learning, and high performance computing in the cloud; Huawei, Lenovo and Sugon becoming NVIDIA HGX-2 cloud server platform partners; Oracle planning to bring NVIDIA HGX-2 platform to Oracle Cloud Infrastructure, giving customers access to a unified HPC and AI computing architecture; leading technology companies taking advantage of HGX-2, the most powerful cloud node in history; HGX-2 having unmatched compute power; companies around the world now being able to build new products and services that can scale to solve immense compute challenges and address some of the world’s most pressing problems; and HGX-2 cloud server platform’s features supporting the future of computing and providing accurate HPC while enabling faster reduced precision 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.

**Media Contacts**

Kristin Bryson
+1 203 241 9190
kbryson@nvidia.com