NVIDIA.

NVIDIA Announces Ready-Made NVIDIA DGX SuperPODs, Offered by Global Network of Certified Partners

World's Most Advanced AI System Now Available in 20-Node Building Block Increments; First Installments Shipping by Yearend to Korea, UK, Sweden and India

GTC — NVIDIA today announced the NVIDIA DGX SuperPOD™ Solution for Enterprise, the world's first turnkey AI infrastructure, making it possible for organizations to install incredibly powerful AI supercomputers with extraordinary speed — in many cases in just a few weeks' time.

Available in cluster sizes ranging from 20 to 140 individual NVIDIA DGX A100™ systems, DGX SuperPODs are now shipping and expected to be installed in Korea, the U.K., Sweden and India before the end of the year.

Sold in 20-unit modules interconnected with NVIDIA Mellanox® HDR InfiniBand networking, DGX SuperPOD systems start at 100 petaflops of AI performance and can scale up to 700 petaflops to run the most complex AI workloads.

"Traditional supercomputers can take years to plan and deploy, but the turnkey NVIDIA DGX SuperPOD Solution for Enterprise helps customers begin their AI transformation today," said Charlie Boyle, vice president and general manager of DGX systems at NVIDIA. "State-of-the-art conversational AI, recommender systems and computer vision workloads rapidly exceed the capabilities of traditional infrastructure, and our new solution gives customers a fast track to the world's most advanced, scalable AI infrastructure and NVIDIA expertise."

Global Innovators Adopt DGX SuperPOD Solution for AI Centers of Excellence

Visionary organizations are creating AI centers of excellence with the DGX SuperPOD Solution for Enterprise. Those unveiling new DGX SuperPOD AI supercomputers today include:

- **NAVER**, the leading search engine in Korea, has created with LINE, Japan's No. 1 messaging service, the AI technology brand NAVER CLOVA. NAVER CLOVA is using its DGX SuperPOD built with 140 DGX A100 systems to scale out research and development of natural language processing models and conversational AI services on its AI platform with the NVIDIA TensorRT™ SDK for high-performance deep learning inference.
- **Linköping University**, in Sweden, is building BerzeLiUs, a DGX SuperPOD of 60 DGX A100 systems. BerzeLiUs will be a powerful resource to advance AI research and boost collaboration between academia and Swedish industry across research programs financed by the Knut and Alice Wallenberg Foundation, such as the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program and initiatives in the life sciences and quantum technology.
- **C-DAC**, the Centre for Development of Advanced Computing operating under the Ministry of Electronics and Information Technology in India, is commissioning India's fastest and largest HPC-AI supercomputer, called PARAM Siddhi – AI. Built with 42 DGX A100 systems, the supercomputer will help address nationwide and global challenges in healthcare, education, energy, cybersecurity, space, automotive and agriculture through research partnerships and collaboration across academia, industry and startups.

Additionally, NVIDIA separately announced today plans to build Cambridge-1, an 80-node DGX SuperPOD with 400 petaflops of AI performance. Once deployed by the end of the year, it will be the fastest supercomputer in the U.K. The system will be used for collaborative research within the U.K. AI and healthcare community across academia, industry and startups.

Cambridge-1 will help accelerate diverse healthcare workloads, including drug development with the NVIDIA Clara™ healthcare application framework. It will also enable researchers to rapidly analyze volumes of medical information using natural language processing with the specialized NVIDIA BioMegatron model available on the NVIDIA NGC™ software hub.

World-Leading Infrastructure for AI Innovation

The DGX SuperPOD Solution for Enterprise was developed through years of research and development in creating the world's most advanced AI system to power NVIDIA's own engineering in automotive, healthcare, conversational AI, recommender systems, data science and computer graphics.

NVIDIA Selene, a 280-node DGX SuperPOD, set the bar high for AI with top marks on both TOP500 and MLPerf results published earlier this year. Its DGX SuperPOD architecture also delivers breakthrough efficiency with record-setting Green500 performance of 20 gigaflops/watt.

AI infrastructure requires extremely high-speed storage to handle a variety of data types in parallel, such as text, tabular data, audio and video. The NVIDIA DGX SuperPOD Solution for Enterprise features all-flash storage that is optimized to meet

customers' specific requirements as well as the unique demands of AI workloads. DDN is the first NVIDIA-qualified storage partner for the DGX SuperPOD Solution for Enterprise.

Fully Integrated AI Deployments Across Systems to Software
From customized capacity planning and data center design services to application performance testing and developer operations training, the DGX SuperPOD Solution for Enterprise provides the fastest path to AI innovation at scale. Each DGX SuperPOD is fully racked, stacked and configured by NVIDIA-Certified partners. These NVIDIA AI experts ensure installs are easy, even when building out AI infrastructure with dozens or hundreds of nodes connected by extensive cabling.

Following installation, NVIDIA and certified experts work with customers to ensure their AI workloads are optimized with the latest NVIDIA software available on the NGC hub of cloud-native, GPU-optimized containers, models and industry-specific SDKs.

Availability
The DGX SuperPOD Solution for Enterprise is available from select NVIDIA partners worldwide. Learn more at www.nvidia.com/dgxsuperpod.

In addition to the new DGX SuperPOD Solution for Enterprise, the DGX SuperPOD blueprint is available to serve as an industry guide for NVIDIA-Certified partners to plan and deploy their own DGX SuperPOD offerings, complete with services and certified support for NGC software.

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: the benefits, performance and availability of the NVIDIA DGX SuperPOD Solution for Enterprise; NVIDIA's plans to build Cambridge-1 as the fastest supercomputer in the U.K. once deployed; and the benefits and availability of the Cambridge-1 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.

Shannon McPhee
+1-310-920-9642
smcphee@nvidia.com