Nvidia Announces Its Most Powerful GPU - Blackwell Ultra, Designed for Training Agent AI

Within the framework of the Computex 2025 event taking place in Taipei, Taiwan, Nvidia officially introduced Blackwell Ultra - the company's latest generation GPU.

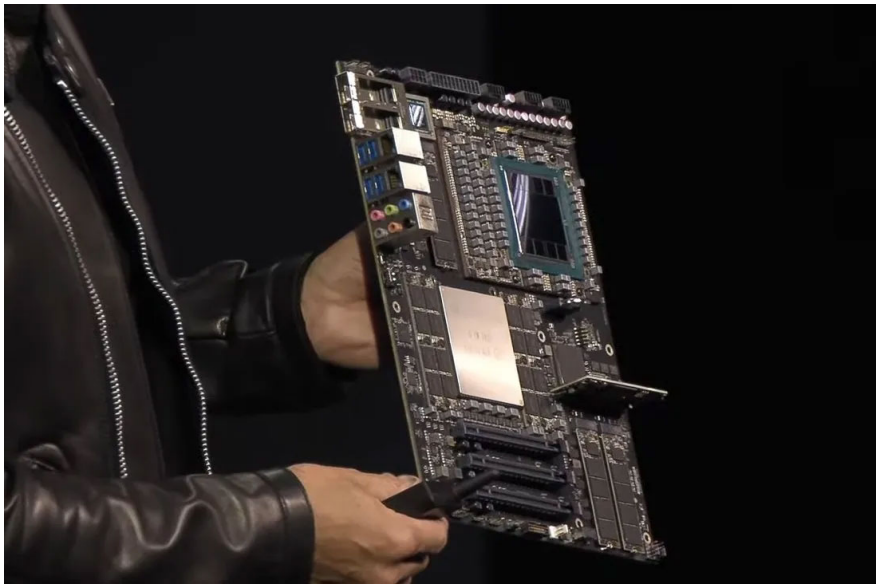
At Computex 2025 in Taipei, Taiwan, Nvidia officially introduced Blackwell Ultra - its latest generation GPU, built for large-scale AI tasks, autonomous AI, and advanced inference.

Superior performance with Blackwell Ultra architecture

Blackwell Ultra is the "heart" of the **GB300 NVL72** system - a liquid-cooled server cluster integrating **72 Blackwell Ultra GPUs** and **36 Nvidia Grace CPUs** . According to Nvidia, this system delivers **1.5 times** higher performance than the previous generation (GB200 NVL72) and increases revenue opportunities **by 50 times** compared to systems using Hopper GPUs.

Together with **the HGX B300 NVL16 (using NVLink Fusion technology), Nvidia claims that this new system has 11 times faster** inference speed on large language models (LLM) , **7 times more powerful** computing power and **4 times more** memory than the Hopper generation.

Speaking about the newly launched product, Jensen Huang - Founder and CEO of Nvidia - shared:
"AI has taken a giant leap forward - autonomy and inference capabilities require orders of magnitude more computing power. Blackwell Ultra is designed for this moment - a flexible platform that can easily handle AI pretraining, post-training, and inference."



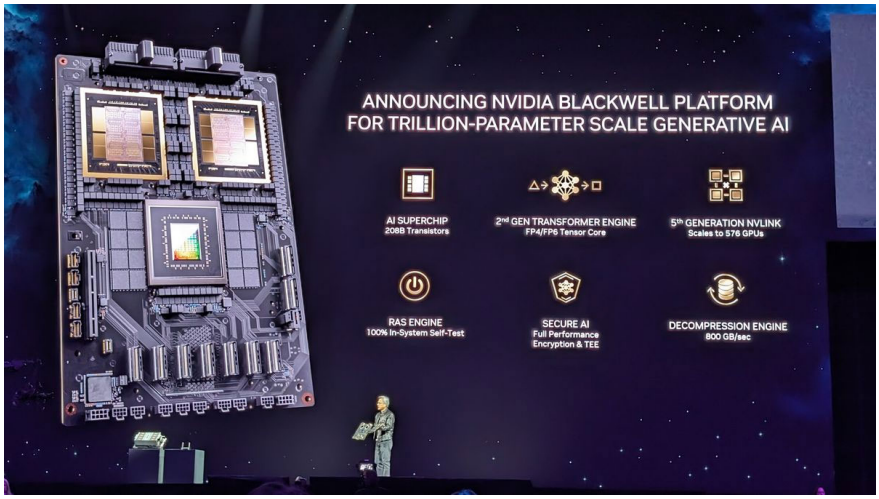
Autonomous AI and Physical AI

Nvidia says Blackwell Ultra is specifically designed to build **agentic AI** and **physical AI** . Among them:

1. **Autonomous AI** uses critical thinking and iterative planning to solve complex problems, taking a multi-step approach to decision making.
2. **Physical AI** allows companies to train robots and autonomous vehicles virtually at scale before deploying them in the real world.

The Blackwell Ultra system can integrate with the **Spectrum-X Ethernet** platform and **Quantim-X800 InfiniBand** , achieving 800Gb/s data transfer rates **per GPU** thanks to **the ConnectX-8 SuperNIC** - a network accelerator also launched at the event.

To optimize inter-GPU bandwidth, ConnectX-8 integrates **48 PCIe Gen6 lanes** with a PCIe Gen6 switch, consolidating **GPU-to-GPU** and **GPU-to-NIC** connectivity into one high-performance device, replacing previous dedicated PCIe switches.



Blackwell Ultra products will be available from Nvidia's traditional partners such as **Cisco, Dell, HPE, Lenovo and Supermicro** from **the second half of 2025** .

Additionally, Blackwell-based cloud computing services will be available on **AWS, Microsoft Azure, Google Cloud, Oracle Cloud,** and cloud GPU providers such as **CoreWeave, Crusoe, Lambda, Nebius, Nscale, Yotta, and YTL . The GB300 NVL72** system will also be available on Nvidia's DGX Cloud platform.