

Golden Award AI Computing & Tech

NVIDIA Vera Rubin NVL72



NVIDIA

NVIDIA Corporation

香港商輝達香港控股有限公司台
灣分公司



Winning Reason

NVIDIA Vera Rubin NVL72 serves as the core of rack-scale AI supercomputing, integrating 72 Rubin GPUs and 36 Vera CPUs. Leveraging 6th-generation NVLink Switch technology, this system is purpose-built for the inference and training of trillion-parameter models. It delivers extreme compute density, a full liquid-cooling solution, and high-performance, ultra-fast interconnect bandwidth. Compared to its predecessor, the NVL72 offers a generational leap in performance, significantly reducing inference costs per token while maximizing computational efficiency.

Product Feature

Vera Rubin NVL72 is a new class of AI supercomputer designed for pre-training, post-training, and test-time scaling. This platform features an extreme co-design of six chips including the Vera CPU and Rubin GPU. Key innovations include 6th-gen NVLink connecting 72 GPUs to act as one massive system and a cable-free rack design that reduces compute tray assembly from 2 hours to 5 minutes. The 100% liquid-cooled architecture supports 45C water to save up to 60MW in large AI factories. Optimized for agentic AI, it delivers 10x higher inference performance per watt. With 220 trillion transistors and a 260TB/s NVLink copper spine, it provides the massive scale needed for trillion-parameter models.

<https://bcaward.computex.biz/>

Sustainable Tech Special Award AI Computing & Tech

NVIDIA Vera Rubin NVL72



NVIDIA

NVIDIA Corporation

香港商輝達香港控股有限公司台
灣分公司



Winning Reason

Achieving an extraordinary 10x improvement in performance-per-watt, our solution redefines modern computational power. The system features a revolutionary architecture that is entirely cable-free, hose-free, and fanless. By utilizing advanced liquid cooling technology, it seamlessly meets the intensive demands of modern workloads while delivering on ambitious energy-saving targets and operational sustainability.

Product Feature

Vera Rubin NVL72 is a new class of AI supercomputer designed for pre-training, post-training, and test-time scaling. This platform features an extreme co-design of six chips including the Vera CPU and Rubin GPU. Key innovations include 6th-gen NVLink connecting 72 GPUs to act as one massive system and a cable-free rack design that reduces compute tray assembly from 2 hours to 5 minutes. The 100% liquid-cooled architecture supports 45C water to save up to 60MW in large AI factories. Optimized for agentic AI, it delivers 10x higher inference performance per watt. With 220 trillion transistors and a 260TB/s NVLink copper spine, it provides the massive scale needed for trillion-parameter models.

<https://bcaward.computex.biz/>