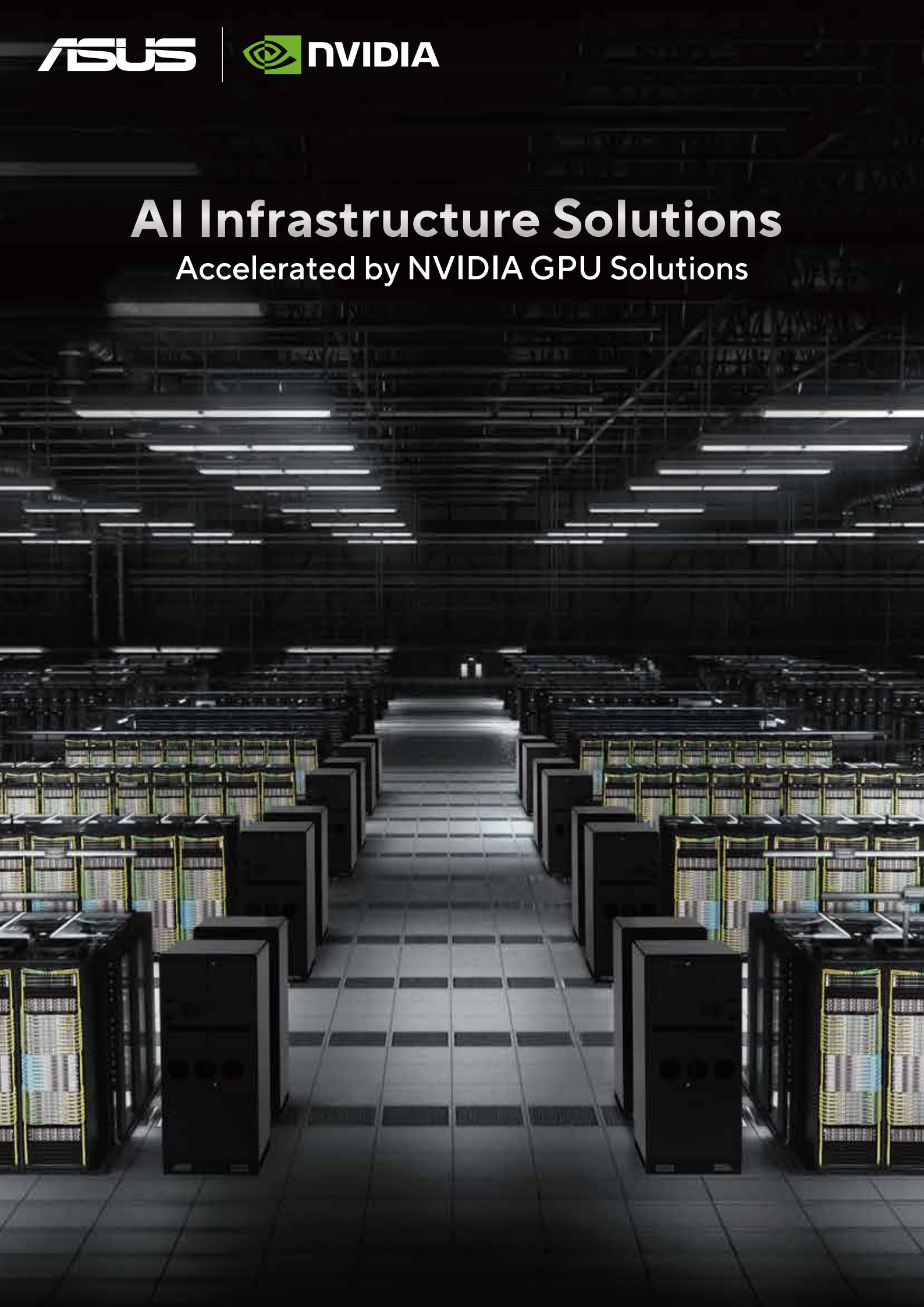
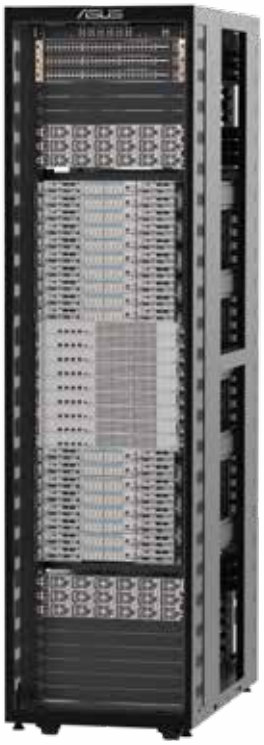
ASUS | NVIDIA

# AI Infrastructure Solutions

## Accelerated by NVIDIA GPU Solutions



ASUS AI POD with NVIDIA Blackwell GPUs

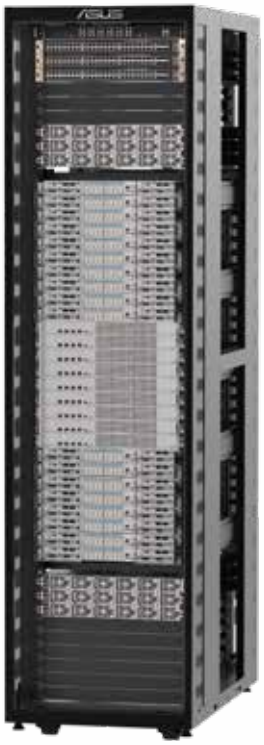


ESC NM2N721-E1

# ASUS AI POD with NVIDIA GB200 NVL72

**Unleash the power of AI with NVIDIA Grace Blackwell Superchips**

- 36 NVIDIA Grace™ CPUs
- 72 NVIDIA Blackwell GPUs
- 5th Gen NVIDIA NVLink™ technology
- Supports trillion-parameter LLM inference and training
- Scale-up ecosystem-ready
- ASUS infrastructure deployment center
- ASUS premium service suite



XA GB721-E2

# ASUS AI POD with NVIDIA GB300 NVL72

**Built for the age of AI reasoning**

- 36 NVIDIA Grace™ CPUs
- 72 NVIDIA Blackwell Ultra GPUs
- 5th Gen NVIDIA NVLink™ technology
- NVIDIA Blackwell Ultra GPU SXM7 and SOCAMM modules design for serviceability
- Support AI reasoning inference
- Scale-up ecosystem-ready
- ASUS infrastructure deployment center
- ASUS premium service suite

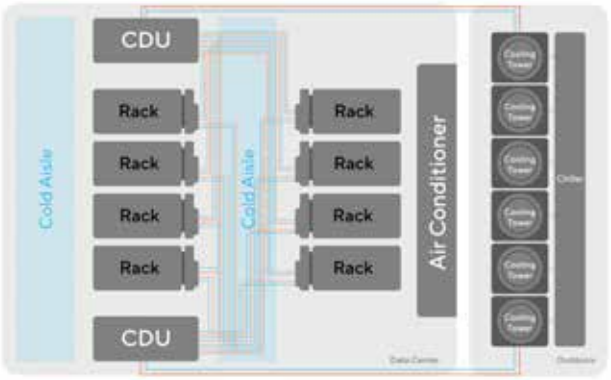
Apply for NVIDIA GB200 NVL72 Online Test

# ASUS AI POD infrastructure solutions and servers

## Cooling layout design and installation

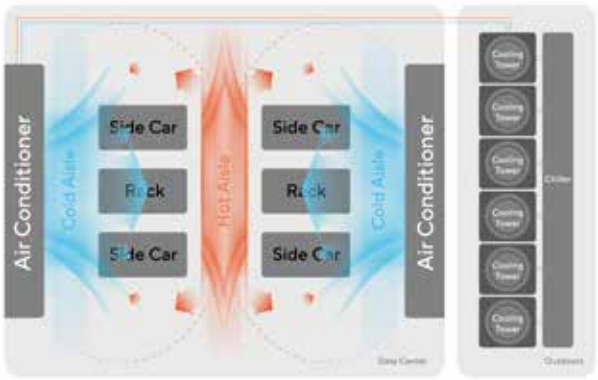
**Comprehensive cabinet-level liquid-cooling solutions**

### Liquid-to-liquid




- Ideal for large-scale, extensive infrastructure with high workloads
- Provides long-term, low PUE with sustained energy efficiency over time
- Reduces TCO for maximum value and cost-effective operations

### Liquid-to-air




- Ideal for small-scale data centers with compact facilities
- Designed to meet the needs of existing air-cooled data centers and easily integrate with current infrastructure
- Perfect for enterprises seeking immediate implementation and deployment

## Rack verification and deployment

**One-stop provider for building infrastructure, platforms, operations and applications**

- **ASUS Infrastructure Deployment Center (AIDC)**
- Automation, systemization and zero-touch onboarding: Service deployment, Software installation
- Centralized configuration control and management: System configuration, Network configuration
- Accelerated deployment for rack-scale infrastructure: Rack validation, OS image deployment

## Generative AI cloud services

**One-stop provider for building infrastructure, platforms, operations and applications**

TWS TAIWAN WEB SERVICE

- **TWSC AI Foundry Service**
- User-friendly platform: Professional service and intuitive usability
- AI workload-optimized design: Tailored for efficient AI processing
- Scalable architecture: Flexible and adaptable to growth
- AI governance: Secure, compliant, and trustworthy

Read our white paper and discover our comprehensive AI Infrastructure solution

# XA NB3I-E12

**NVIDIA HGX™ B300**
**Latest Blackwell platform for heavy AI workloads**

- Powered by dual Intel® Xeon® 6700/6700P processors, supporting a 350W TDP
- Seamless GPU-to-GPU interconnect via NVIDIA NVLink™, providing 1800GB/s bandwidth for optimized scalability
- Modular design with minimal cable usage, enhancing assembly speed and thermal efficiency
- Integration of advanced NVIDIA technologies, leveraging the full capabilities of NVIDIA GPUs, NVIDIA BlueField®-3 DPUs, NVIDIA NVLink™, NVLink Switch, and networking
- Exceptional power efficiency, supported by 5+5 80 PLUS® Titanium power supplies

# ESC NB8-E11

**NVIDIA HGX™ B200**
**Blackwell platform for unmatched AI performance**

- Equipped with NVIDIA Blackwell HGX™ B200 8-GPU
- Powered by dual 5th Gen Intel® Xeon® Scalable processors with support for 350W TDP
- Direct GPU-to-GPU interconnect via NVIDIA NVLink™ offering 1800GB/s bandwidth for optimized scaling
- Dedicated one-GPU-to-one-NIC topology, supporting up to eight NICs for maximum throughput in compute-intensive workloads
- Advanced NVIDIA technologies, including full integration of NVIDIA GPUs, NVIDIA BlueField®-3 DPUs, NVIDIA NVLink™ NVLink Switch, and networking
- Designed for generative AI with optimized server systems, data-center infrastructure, and AI software development capabilities integrated by ASUS

Learn more ASUS NVIDIA-Certified Systems™

# ESC N8-E11V

**NVIDIA HGX™ H200**
**End-to-end AI supercomputing platform**

- Powered by dual 5th Gen Intel® Xeon® Scalable processors with support for 350W TDP
- Direct GPU-to-GPU interconnect via NVIDIA NVLink™, offering 900GB/s bandwidth for optimized scaling
- Dedicated one-GPU-to-one-NIC topology, supporting up to eight NICs for maximum throughput in compute-intensive workloads
- Advanced NVIDIA technologies, including full integration of NVIDIA GPUs, NVIDIA BlueField®-3 DPUs, NVIDIA NVLink™, NVLink Switch, and networking
- Superior power efficiency with 4+2 80 PLUS® Titanium-rated power supplies

# ESC N8-E11

**NVIDIA HGX™ H100**
**End-to-end AI supercomputing platform**

- Powered by dual 5th Gen Intel® Xeon® Scalable processors with support for 350W TDP
- Direct GPU-to-GPU interconnect via NVIDIA NVLink™, offering 900GB/s bandwidth for optimized scaling
- Dedicated one-GPU-to-one-NIC topology, supporting up to eight NICs for maximum throughput in compute-intensive workloads
- Advanced NVIDIA technologies, including full integration of NVIDIA GPUs, NVIDIA BlueField®-3 DPUs, NVIDIA NVLink™, NVSwitch, and networking
- Superior power efficiency with 4+2 80 PLUS® Titanium-rated power supplies

Learn more ASUS NVIDIA-Certified Systems™

# ESC8000A-E13P

**NVIDIA H200 NVL PCIe, NVIDIA L40s PCIe, NVIDIA RTX PRO™ 6000 Blackwell Server Edition**
**Turbocharging generative AI and LLM workloads**

- Powered by AMD EPYC™ 9005 processors with 192 Zen 5c cores, 12-channel memory, DDR5 up to 6000 MHz, and a maximum TDP of 500W per socket
- Fully compatible with NVIDIA MGX™ architecture, enabling rapid and scalable deployment
- High-density 4U server supporting up to eight dual-slot NVIDIA H200 GPUs or NVIDIA RTX PRO™ 6000 Blackwell Server Edition, each with a power capacity up to 600W
- Optimized server configuration, five PCI 5.0 slots for high-bandwidth PCIe NICs and DPU to enable performance scaling
- ASUS-exclusive toolless design for simplified maintenance and operational efficiency
- ASUS Control Center IT-management software paired with a hardware-level Root-of-Trust solution for enhanced security

# ESC8000-E12P

**NVIDIA H200 NVL PCIe, NVIDIA L40s PCIe, NVIDIA RTX PRO™ 6000 Blackwell Server Edition**
**PCIe Ready for AI and HPC workloads**

- Powered by dual Intel® Xeon® 6 processors, supporting a 350W TDP
- Fully compatible with NVIDIA MGX™ architecture, enabling fast and large-scale deployment.
- High-density 4U server supporting up to eight dual-slot NVIDIA H200 GPUs or NVIDIA RTX PRO™ 6000 Blackwell Server Edition, each with a power capacity up to 600W
- Optimized server configuration, five PCI 5.0 slots for high-bandwidth PCIe NICs and DPU to enable performance scaling
- ASUS-exclusive toolless design empowers easy maintenance for maximum efficiency
- ASUS Control Center IT management software and hardware-level Root-of-Trust solution

Learn more ASUS NVIDIA-Certified Systems™

# ESC8000A-E12

**NVIDIA L40 GPU**
**Maximizing AI and graphics potential**

- AMD EPYC™ 9004 processors with 128 Zen 4c cores
- Supports up to eight dual-slot GPUs, NVIDIA NVLink, and BlueField® DPU for scalable performance
- Independent CPU and GPU airflow for thermal optimization
- Enhanced air cooling based on CPU TDP for diverse workloads
- Eight bays with tri-mode NVMe/SATA/SAS drives and 11 PCIe 5.0 slots for high bandwidth and upgrades

# ESC8000-E11

**NVIDIA L40 GPU**
**High-performance 4U server**

- 5th Gen Intel® Xeon® Scalable processors deliver up to 21% greater performance per watt
- Supports up to eight dual-slot GPUs, NVIDIA NVLink™, and BlueField® DPU for scalable performance
- Independent CPU and GPU airflow for thermal optimization
- Enhanced air cooling for versatile workloads based on CPU TDP
- Eight bays with Tri-Mode NVMe/SATA/SAS drives and 11 PCIe 5.0 slots for high bandwidth and upgrades

Learn more ASUS NVIDIA-Certified Systems™

# ASUS AI Total Infrastructure

Consult & Design

Install

Validation

Deployment (AIDC)

Maintain & Support

Find out more

Read the full technical insight articles