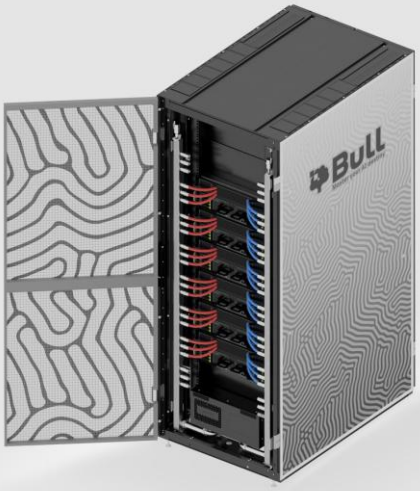
BullSequana AI 642-H03

A high-performance AI accelerator designed for training AI models and performing inference with large models



BullSequana AI 642-H03 overview

The BullSequana AI 642-H03 server is an AI-optimised compute node designed to empower industries, national HPC centres, government bodies, academia, and start-ups to execute accelerated, low-latency AI inference and precision fine-tuning for mid-sized language models (~100M–70B parameters) and multimodal AI workloads.

The BullSequana AI 642-H03 offers superior speed and performance, with up to 48 AMD Instinct™ MI355X GPUs per rack across six servers. The system delivers 64 TB of global GPU bandwidth, enabling high-speed communication pathways and direct memory exchange without routing through the CPU. The AMD Instinct™ MI355X GPU architecture balances GPU-CPU performance for AI workloads, supports a wide range of precision formats for training and inference including low-precision inference for mid-sized language models and offers 2.3TB HBM3e with 288GB of on-board memory per GPU.

Powered by AMD's latest AI technology, the BullSequana AI 642-H03 combines AMD EPYC processors with high-performance GPUs, making it an ideal solution for building training clusters that require eight GPUs per server. The BullSequana AI 642-H03 server is hybrid-cooled for better energy efficiency and sustained performance across demanding workloads, featuring a manifold and a rack-mounted CDU, with up to 70% direct liquid cooling (DLC). The remaining 30% is cooled by air via fans.

Designed for performance, built for functionality

The rack is purposely designed, customised, sized, and optimised to ensure the most effective power distribution unit (PDU), coolant distribution unit (CDU), effectively dissipating the hot air generated within compact 4U servers housed in 42U racks. For even higher energy efficiency, an optional rear-door heat exchanger can be added, targeting compute-only workloads with enhanced thermal performance.

The standardised 800 mm-wide rack allows for easier cable and accessory integration, and the in-rack CDU and manifold provide direct liquid cooling. Additionally, it features 80+ titanium-certified power supplies and PWM cooling fans.

With its fully pre-integrated, pre-tested and pre-configured design, the BullSequana AI 642-H03 server offers fast deployment, streamlined installation and reduces space and costs. This turnkey solution provides flexibility and efficiency for both rapid brownfield rollouts and greenfield deployments.

Technical Specifications

Values are given per server; rack-level figures should be for six servers.

GPU Architecture	• Hybrid-cooled 8 GPUs AMD MI355X UBB8 • Memory: 2.3TB HBM3e / (288 GB per GPU) • Interconnect: 8TB/s of global GPU bandwidth (1TB/s GPU-to-GPU) • GPU TDP: 11,200W (1,400W per GPU)
CPU Architecture	• Dual AMD EPYC™ 4th and 5th Gen processors • 12-channel DDR5 RDIMM, 3TB total size
Storage	• Front: 8 x 2.5" NVMe SSDs • Internal: 2 x M.2 NVMe SSDs
I/O Options	• Interconnect BXI V3 / Ethernet, InfiniBand XDR and NDR • 12 x PCIe Gen5 x16 • 2x 1 GbE RJ45, 2x USB 3.1, 1x VGA port and 2x 10Gb LAN ports
Power	• Full-server power consumption: 16 kilowatts
Availability and RAS features	• Smart Crisis Management and Protection, Dual ROM Architecture • Hot-swap devices (PSUs, PCIe blades, NVMe drives)

System Management	• BMC based on Aspeed AST2600
	Supported Operating Systems and Software
	• Operating Systems: Red Hat Enterprise Linux
	• Hardware Tooling: Hardware Abstraction Layer (HAL) and monitoring exporters & dashboards
	• Software Management Stack:
	  • Smart Management Center xScale
	  • Smart Management Center
	  • Smart Maintenance Management Suite
	• Smart Performance Management Suite:
	  • Bull OpenMPI and GPU programming environments
	  • Bull Slurm
	  • IBMS (Interconnect faBric Monitoring System)
	  • Singularity PRO and Enterprise
	• Cluster & Job Monitoring
	  • Performance and efficiency: BullSequana ARGOS

As part of the BullSequana AI 600H server family, the BullSequana AI 642-H03 is meticulously assembled and rigorously tested in Bull's factory, backed by robust quality-assurance processes that ensure reliability and give customers complete confidence in the product's performance and durability.

Software infrastructure for empowering AI teams

AI developers can utilise software infrastructure to build, train, and optimize their models effectively. By leveraging the full capabilities of the AMD ROCm 7 software stack, AI engineers can significantly enhance both usability and performance for MI355X GPUs. This enhancement leads to faster results compared to competing solutions and accelerates the model-development workflow.

AI system administrators benefit from Bull's HPC Software Suite, which provides a unified platform for running, tuning, and managing nodes and clusters at scale. It integrates orchestration, profiling, intelligent resource control, and advanced power monitoring and management to optimise performance, enhance energy efficiency, and ensure cost-effective, reliable operations. This approach aligns fully with your team's commitment to digital sobriety and sustainable infrastructure.

Security Features	• Secure Boot, Secure firmware, and TPM 2.0
Power Supply Unit (PSU)	• 80 PLUS Titanium, Full redundancy • AC Input: - 200-240V~/ 25A, 50-60Hz
Physical Characteristics	• 4U form factor • Total Weight: 86,8 Kg
Regulation and Safety	• Global Compliance: RoHS, REACH • Certification: CE, ErP Lot 9, FCC

Proven track record and expertise

Bull holds a leading position as one of Europe's key supercomputer manufacturers, setting it apart in the global market. We provide end-to-end solutions — designing, delivering, installing and configuring cluster systems for leading organizations worldwide. Our factory in France plays a strategic role in this mission, enabling high-quality production that supports innovation and digital sovereignty in AI and HPC infrastructure. With decades of experience, Bull has earned the trust of industrial, academic, and HPC sectors across the globe.

Connect with us
bull.com