

GIGAPOD Scalable AI Computing Cluster

GIGABYTE GIGAPOD is an AI computing cluster solution designed for exceptional scalability and high performance. It offers seamless adaptability for data centers facing growing AI demands, with optimized thermal or liquid cooling for peak computational power.



Unleash a Turnkey AI Data Center with High Throughput and an Incredible Level of Compute

GIGABYTE has been pivotal in providing its technology leaders with a supercomputing infrastructure built around powerful GIGABYTE GPU servers that house either NVIDIA HGX™ or AMD Instinct™ accelerators. GIGAPOD is a service that has professional help to create a cluster of racks all interconnected as a cohesive unit. An AI ecosystem platform thrives with a high degree of parallel processing as the GPUs are interconnected with blazing fast communication by NVIDIA NVLink™ or AMD Infinity Fabric™. With the introduction of the GIGAPOD, GIGABYTE now offers a one-stop source for data centers that are moving to an AI factory that runs deep learning models at scale. The hardware, expertise, and close relationship with cutting-edge GPU partners ensures the deployment of an AI supercomputer goes off without a hitch and minimal downtime.

GIGABYTE G Series Servers Built for 8-GPU Platforms

One of the most important considerations when planning a new Al data center is the selection of hardware, and in this Al era, many companies see the choice of the GPU/Accelerator as the foundation. Each of GIGABYTE's industry leading GPU partners (AMD, Intel, and NVIDIA) has innovated uniquely advanced products built by a team of visionary and passionate researchers and engineers, and as each team is unique, each new generational GPU technology has advances that make it ideal for particular customers and applications. This consideration of which GPU to build from is mostly based on factors: performance (Al training or inference), cost, availability, ecosystems, scalability, efficiency, and more. The decision isn't easy, but GIGABYTE aims to provide choices, customization options, and the know-how to create ideal data centers to tackle the demand and increasing parameters in Al/ML models.

The Ideal GIGAPOD for You

GIGABYTE enterprise products not only excel at reliability, availability, and serviceability. They also shine in flexibility, whether it be the choice of GPU, rack dimensions, or cooling method and more. GIGABYTE is familiar with every imaginable type of IT infrastructure, hardware, and scale of data center. Many GIGABYTE customers decide on the rack configuration based on how much power their facility can provide to the IT hardware, as well as considering how much floor space is available. So, this is why the service, GIGAPOD, came to be. Customers have choices. Starting with how the components are cooled and how the heat is removed, customers can select either traditional air-cooling or direct liquid cooling (DLC).

GIGAPOD with Air-cooled Configurations



SKU#	GPUs Supported	GPU Server (Form Factor)	GPU Servers per Rack	Power Consumption per Rack	No. of Racks per SU*	
1		5U	4	50kW	8+1 (42U)	
2	NVIDIA HGX™ H200 / H100 AMD Instinct™ MI300X	5U	4	50kW	8+1 (48U)	
3		5U	8	100kW	4+1 (48U)	
4	NVIDIA HGX™ B300 / B200 / H200 AMD Instinct™ MI350X / MI325X / MI300X	8U	4	66kW	8+1 (48U)	
5	NVIDIA HGX™ B200	80U	4	55kW	8+1 (44OU)	
6	Intel® Gaudi® 3	8U	4	62kW	8+1 (48U)	

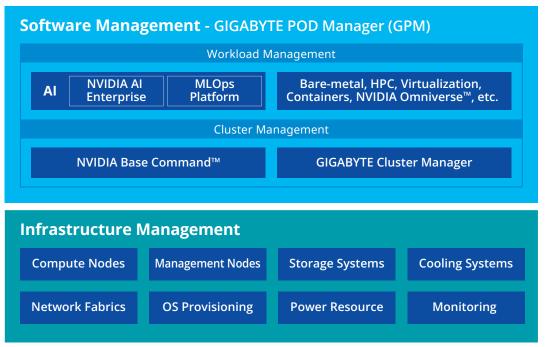
^{*}Compute Racks + 1 Management Rack

GIGAPOD with Liquid-cooled Configurations



SKU#	GPUs Supported	GPU Server (Form Factor)	GPU Servers per Rack	Power Consumption per Rack	No. of Racks per SU*	CDU
1	NVIDIA HGX [™] H200 / H100 AMD Instinct [™] MI300X	5U	8	100kW	4+1 (48U)	In-rack
2		5U	8	100kW	4+1 (48U)	In-row
3	NVIDIA HGX™ B300 / B200 / H200 AMD Instinct™ MI355X / MI325X / MI300X	4U	8	120kW	4+1 (42U)	In-rack
4		4U	8	120kW	4+1 (42U)	In-row
5		4U	8	120kW	4+1 (48U)	In-rack
6		4U	8	120kW	4+1 (48U)	In-row

Applications & Cloud Services





GIGABYTE POD Manager (GPM)

Streamlined POD Monitoring & Automation

Integrated with GIGABYTE POD Manager (GPM), our proprietary infrastructure and workflow management platform, GIGAPOD streamlines AI and HPC operations with real-time monitoring, orchestration, and automation—delivering complete data center solutions from hardware to cluster management.



Why is GIGAPOD the rack scale service to deploy?



Connections

GIGABYTE works closely with technology partners (AMD, Intel, and NVIDIA) to ensure a fast response to customers requirements and timelines.



Depth in **Portfolio**

GIGABYTE servers (GPU, Compute, Storage, & High-density) have numerous SKUs that are tailored for all imaginable enterprise applications.



Scale Up or Out

A turnkey high-performing data center has to be built with expansion in mind so new nodes or processors can effectively become integrated.



Performance

From a single GPU server to a cluster, GIGABYTE has tailored its server and rack design to guarantee peak performance with optional liquid cooling.



Experienced

GIGABYTE has successfully deployed large GPU clusters and is ready to discuss the process and provide a timeline that fulfills customers requirements.



Learn more at https://www.GigaComputing.com/en

- * All specifications are subject to change without notice. Please visit our website for the latest information.
- * The entire materials provided herein are for reference only. Giga Computing reserves the right to modify or revise the content at anytime without prior notice.



Designed by





