

The background of the top half of the image is a server room with multiple racks of server units. The units are illuminated with blue light, and there are glowing blue lines and dots overlaid on the scene, suggesting data flow or network activity. The Gigabyte logo is positioned in the top left corner of this section.

GIGABYTE™

GIGAPOD

Scalable, Turnkey AI Supercomputing Solution

One GIGAPOD is composed of multiple racks populated with GIGABYTE GPU servers acting as one powerful cluster that accelerates everything AI.

GIGAPOD: Optimized Rack-Level Solutions

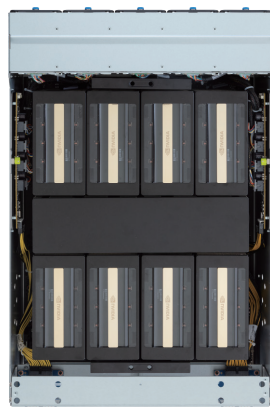
Delivers full integrated and tested racks for Data Center requirement.

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™ B200/H200 or AMD Instinct™ MI300 Series 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 Multi-GPU Platforms

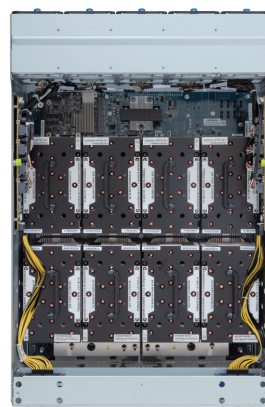
One of the most important considerations when planning a new AI data center is the selection of hardware, and in this AI era, many companies see the choice of the GPU/Accelerator as the foundation. Each of GIGABYTE's industry leading GPU partners (AMD, Intel, and NVIDIA) has innovated uniquely advanced 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 (AI 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 AI/ML models.



NVIDIA HGX™
B300 NVL16 / B200



NVIDIA HGX™
H200 / H100



AMD Instinct™
MI325X / MI300X



Intel® Gaudi® 3

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).



Direct Liquid Cooling

SKU#	GPUs Supported	GPU Server (Form Factor)	GPU Servers per Rack	Rack	Power Consumption per Rack	CDU
1	NVIDIA HGX™ H100/H200 AMD Instinct™ MI300X	5U	8	5 x 48U	100kW	In-rack
2		5U	8	5 x 48U	100kW	In-row
3	NVIDIA HGX™ H200/B200/B300 NVL16 AMD Instinct™ MI300X/MI325X	4U	8	5 x 42U	120kW	In-rack
4		4U	8	5 x 42U	120kW	In-row
5		4U	8	5 x 48U	120kW	In-rack
6		4U	8	5 x 48U	120kW	In-row

Air Cooling

SKU#	GPUs Supported	GPU Server (Form Factor)	GPU Servers per Rack	Rack	Power Consumption per Rack
1	NVIDIA HGX™ H100/H200 AMD Instinct™ MI300X	5U	4	9 x 42U	50kW
2		5U	4	9 x 48U	50kW
3		5U	8	5 x 48U	100kW
4	NVIDIA HGX™ H200/B200/B300 NVL16 AMD Instinct™ MI300X/MI325X	8U	4	9 x 48U	80kW
5	NVIDIA HGX™ H200/B200/B300 NVL16	80U	4	9 x 440U	66kW
6	Intel® Gaudi® 3	8U	4	9 x 48U	80kW

Why is GIGAPOD the rack scale service to deploy?



Industry 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.



High 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.

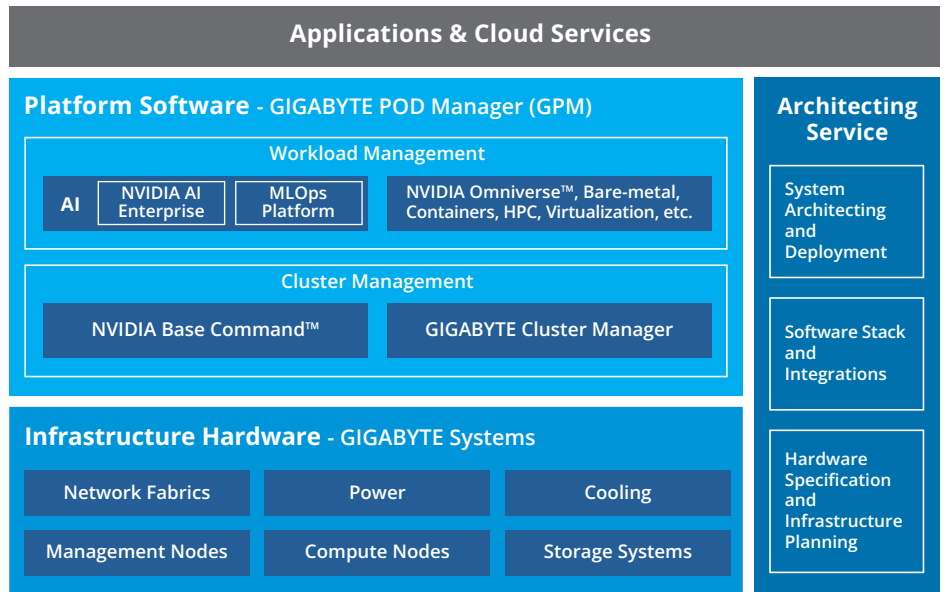
GIGABYTE POD Manager (GPM)

In the field of high-performance computing and AI applications, the concept of a Performance Optimized Datacenter (POD) has emerged as one of the most competitive solutions. It combines powerful compute systems with high scalability and cost efficiency. GIGAPOD, is more than just a scalable hardware design - it integrates infrastructure hardware, platform software, and end-to-end architecting services, covering everything from concept consulting to system validation.

At the core of managing and controlling system status and workflow is GIGABYTE POD Manager (GPM), a powerful software suite designed to enhance operational efficiency, streamline management, and optimize resource utilization. This software-centric approach ensures organizations can effectively manage their data center environments while adapting to modern workload demands.

Key Capabilities:

1. Datacenter Infrastructure Management
2. Operating System Provisioning
3. Orchestration & Workload Deployment
4. Real-Time Monitoring & Alerting
5. Customizable Ecosystem



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.
* All trademarks and logos are the properties of their respective holders.

Designed by

