GIGA COMPUTING

All-AMD AI 伺服器方案解析：從硬體整合到系統管理的價值主張

GIGA COMPUTING

GIGABYTE GIGAPOD

Scalable and Turnkey Data Center Solution

GIGABYTE™

AMD

GIGAPOD

- Centralized Management of Cluster Resources
- Well-integrated Software and Hardware
- Optimizing IT for Stability and Efficiency

Enhanced Reliability

Strengthens IT infrastructure to minimize downtime and ensure consistent service availability.

Improved Observability

Strengthens IT infrastructure to minimize downtime and ensure consistent service availability.

Scalability and Flexibility

Adapts to growing business needs, ensuring long-term value.

Operational Efficiency

Automates routine tasks, reduces manual intervention, and optimizes resource utilization.

Actionable Insights

Offers in-depth analysis to facilitate data-driven decisions and proactive issue resolution.

Streamlined Workflows

Simplifies complex IT processes, enabling smoother operations and faster task completion.

GIGAPOD

Applications & Cloud Services

Platform-software – GIGABYTE POD Manager (GPM)

Workload Management

AI | AMD Infinity Hub | MLOps Platform

Bare-metal, Containers, virtualization, HPC, etc.

Cluster Management

AMD Enterprise AI Ecosystem

GIGABYTE Cluster Manager

Infrastructure Hardware – GIGABYTE Systems

Network Fabrics | Power | Cooling

Management Nodes | Compute Nodes | Storage Systems

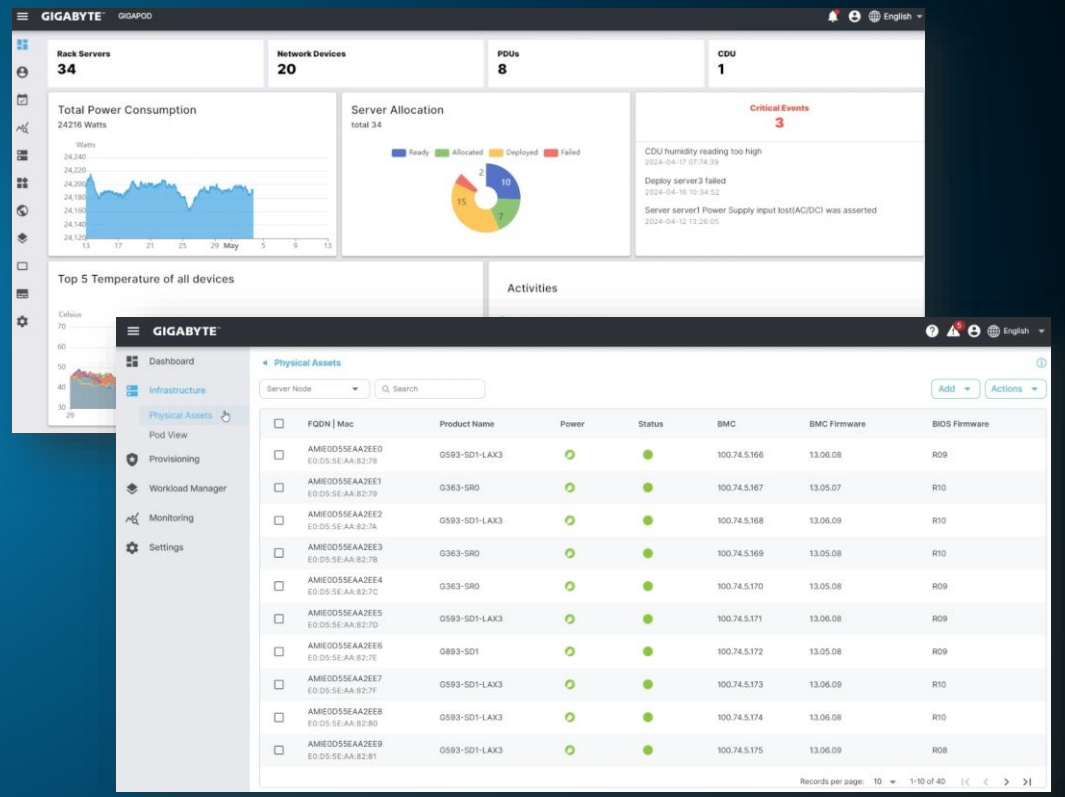
Architecting Service

- System Architecting and Deployment
- Software Stack and Integrations
- Hardware Specification and Infrastructure Planning

GIGABYTE POD Manager (GPM)

Infrastructure Management

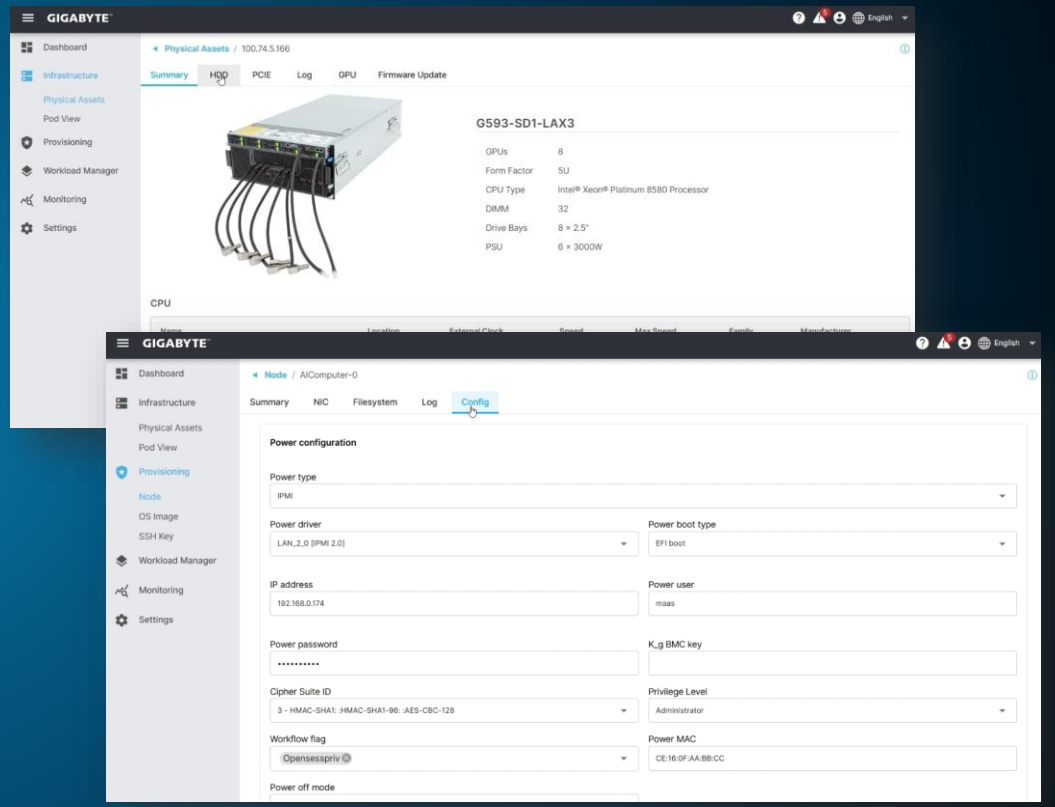
- Centralized inventory of servers, network switches, and storage devices.
- Real-time visualization of resource health, utilization, and physical location.
- POD view for virtualized physical location of each device in data centers.



GIGABYTE POD Manager (GPM)

Node Provisioning

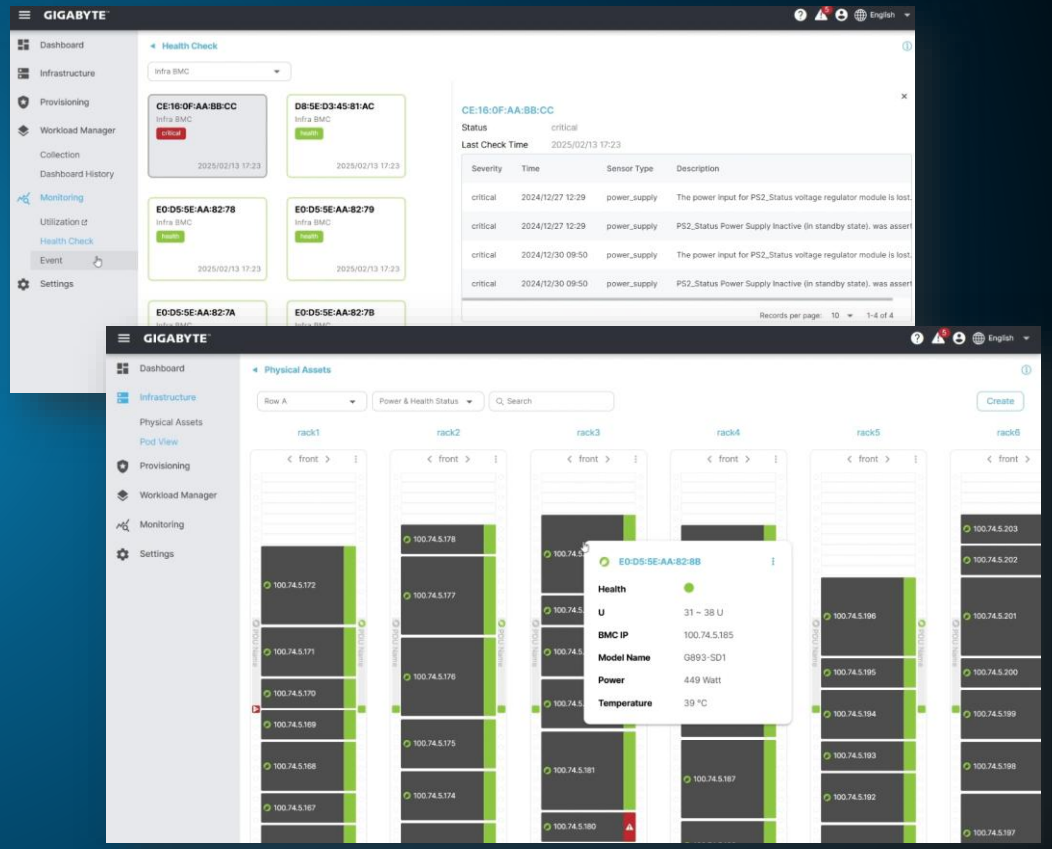
- Automated discovery of new devices within the network for quick onboarding.
- Predefined and customizable templates for OS installation and configuration.
- Batch deployment capabilities to install OS across multiple devices simultaneously.



GIGABYTE POD Manager (GPM)

POD Monitoring & Audit

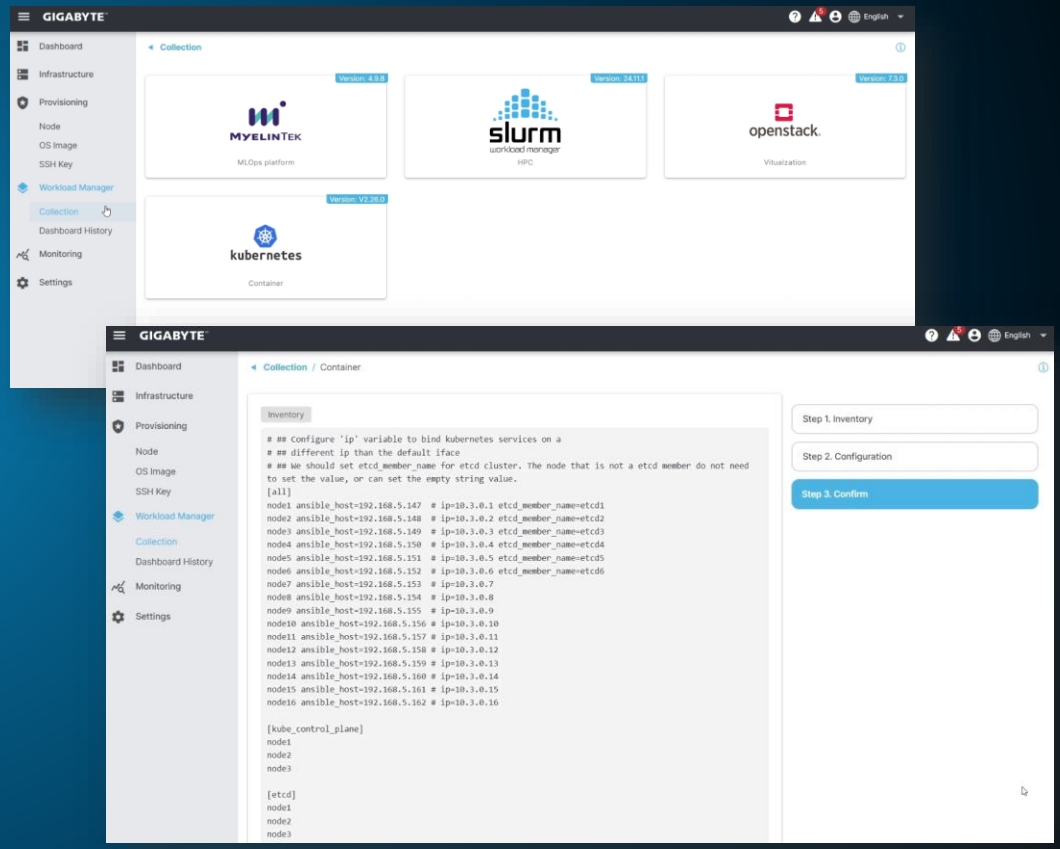
- Real-time monitoring dashboards with customizable metrics from physical devices to operating systems and applications.
- Configurable alert thresholds and notifications via email, webhook, or integrated chat systems.
- Event management tools for logging, categorizing, and resolving issues efficiently.



GIGABYTE POD Manager (GPM)

Orchestration Workload Deployment

- For bare-metal as a service, HPC as a service, containers as a service, virtualization as a service, and AI as a service.
- Support for automatically deploying and managing clustered applications such as Kubernetes, Hadoop, or OpenStack.



GIGABYTE POD Manager (GPM)

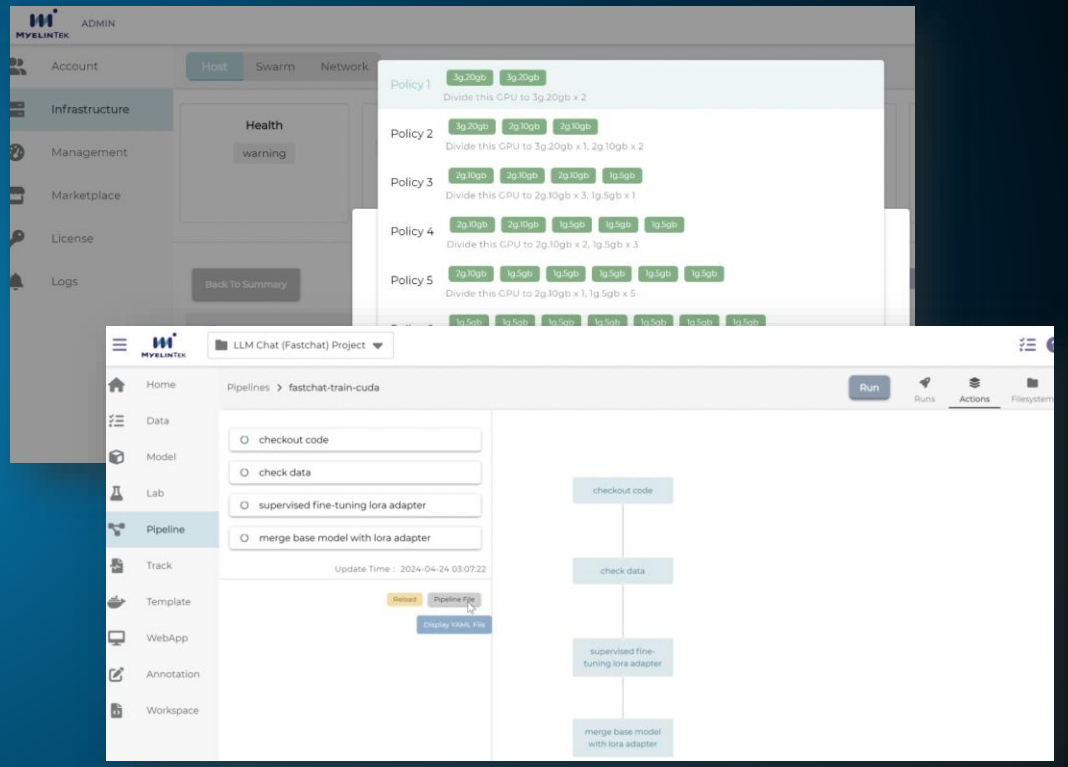
GPU Resource Management

- GPU Pooling and Sharing

AI Workflow Management

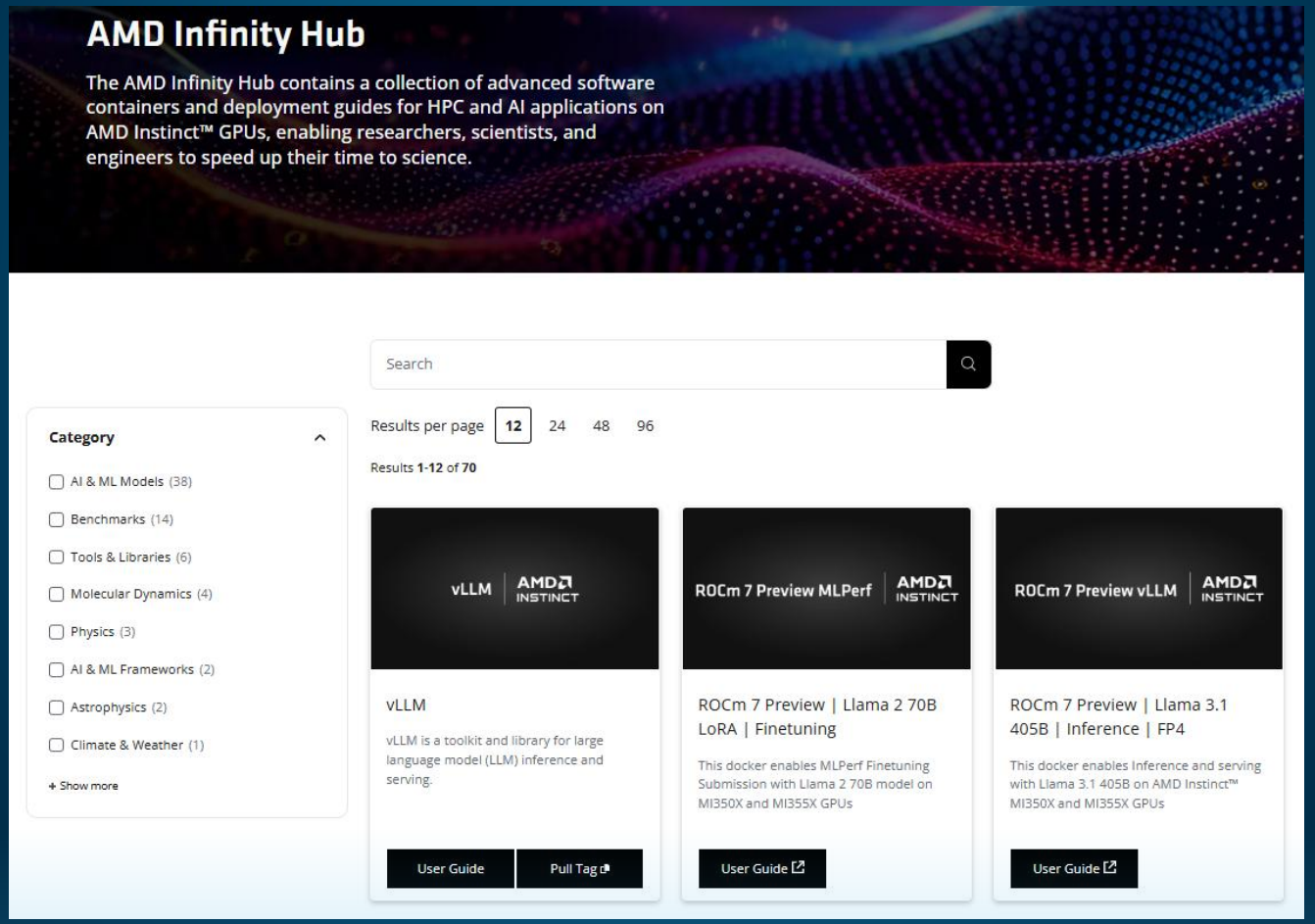
- MLOps Workflow (Pipeline) Automation for Model Pre-training, Fine-tuning & Deployment

- Templates for Data Processing, Model Development & Deployment



GIGABYTE POD Manager (GPM)

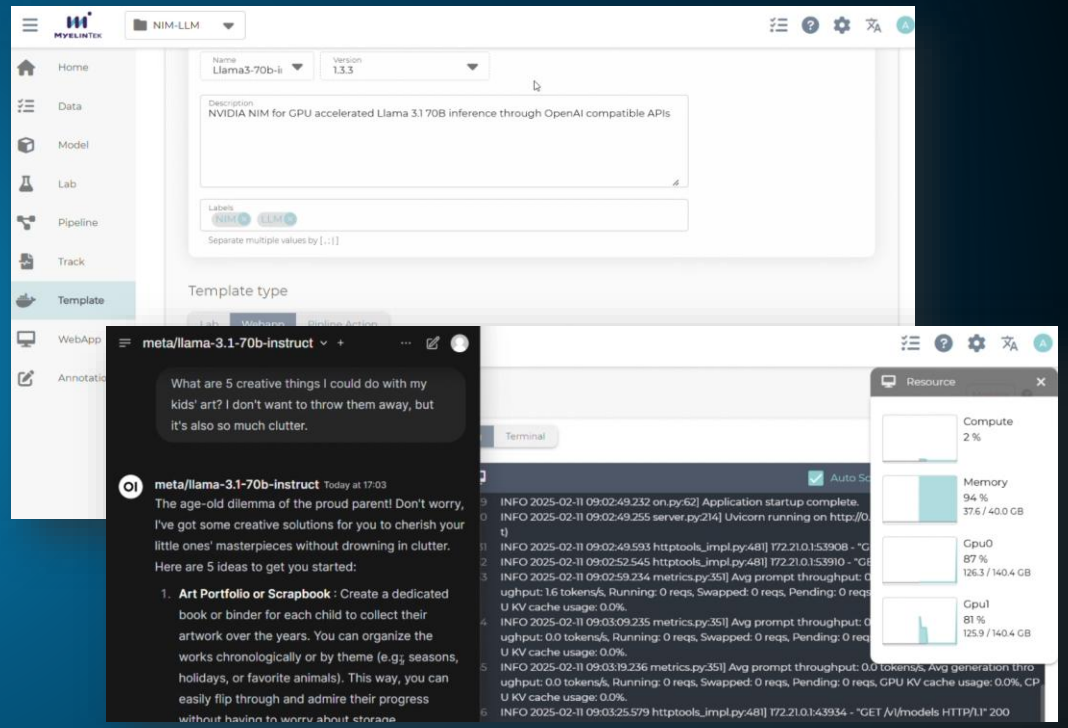
AMD Infinity Hub on GIGAPOD



GIGABYTE POD Manager (GPM)

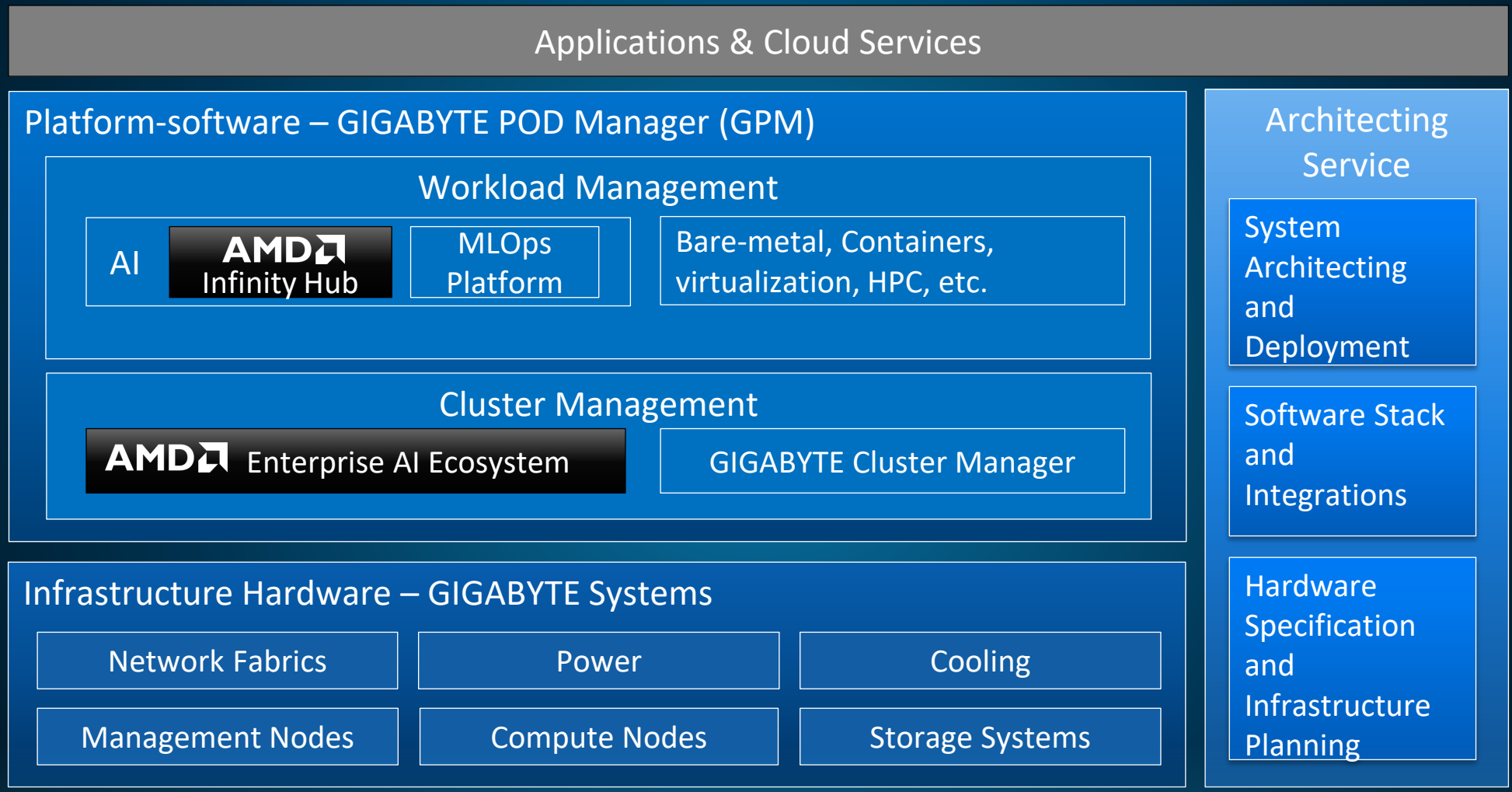
LLM Management

- LLM Readiness & Safety
- LLM Fine-tuning & Deployment



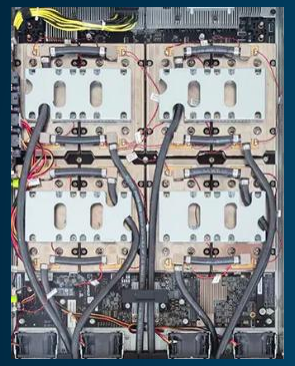
vLLM + AMD ROCm

GIGAPOD

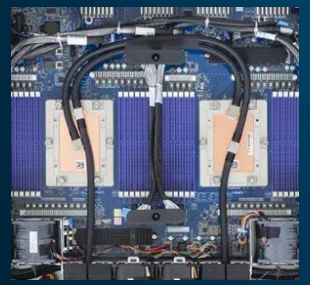

Applications & Cloud Services
Platform-software – GIGABYTE POD Manager (GPM)
Workload Management
AI
AMD Infinity Hub
MLOps Platform
Bare-metal, Containers, virtualization, HPC, etc.
Cluster Management
AMD Enterprise AI Ecosystem
GIGABYTE Cluster Manager
Infrastructure Hardware – GIGABYTE Systems
Network Fabrics
Power
Cooling
Management Nodes
Compute Nodes
Storage Systems
Architecting Service
System Architecting and Deployment
Software Stack and Integrations
Hardware Specification and Infrastructure Planning

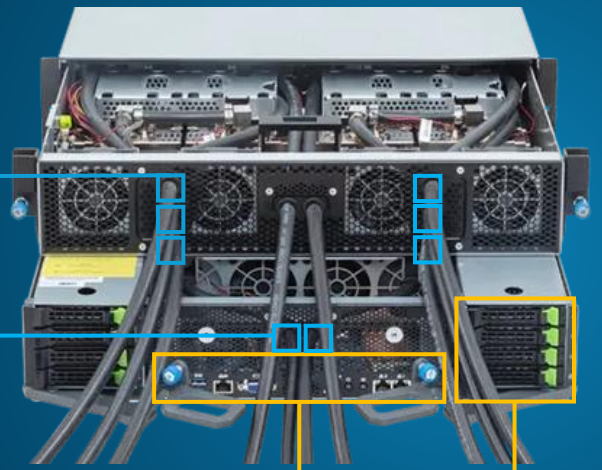
DLC G4L3-ZX1 MI325X/MI355X Server

6 x Hose for GPU Coldplate



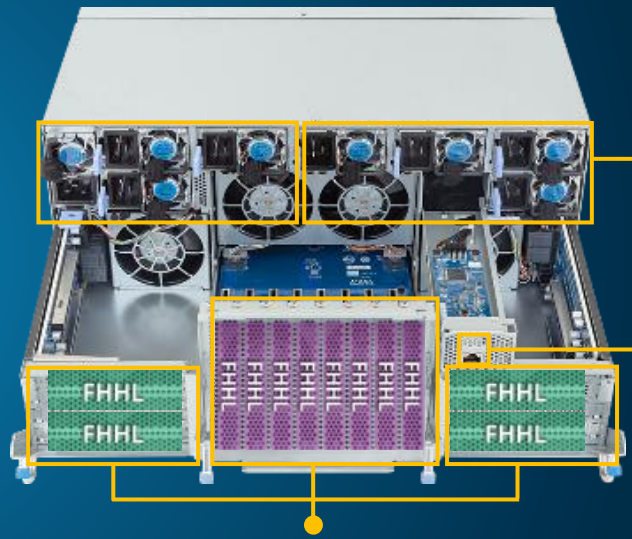
2 x Hose for CPU Coldplate





2 x USB 3.2 Gen1
1 x Front MLAN
1 x VGA
2 x 10Gbps LAN

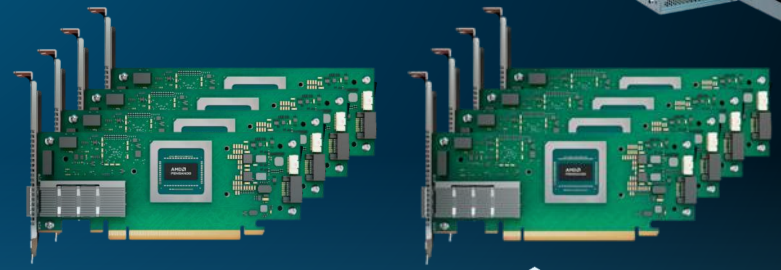
8 x GPU DERICT NVMe



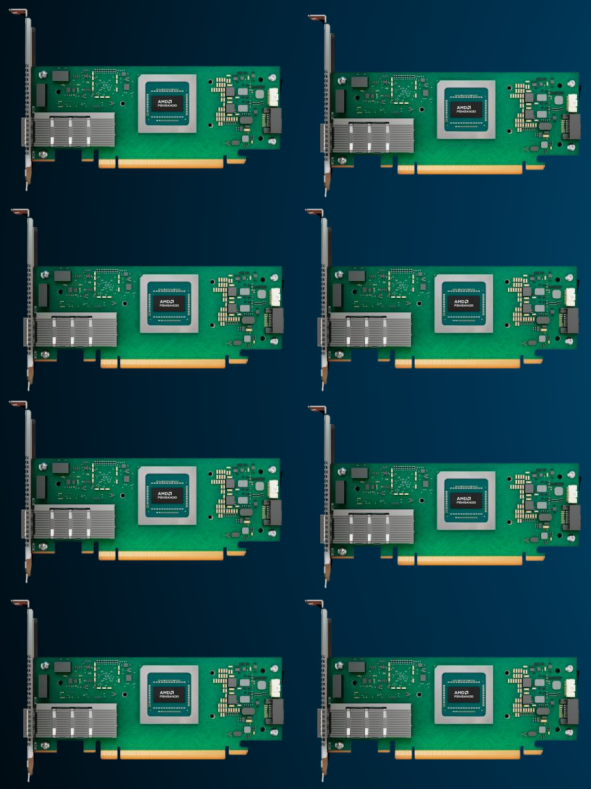
Balanced 4+4
3000W
fulfill peak
power

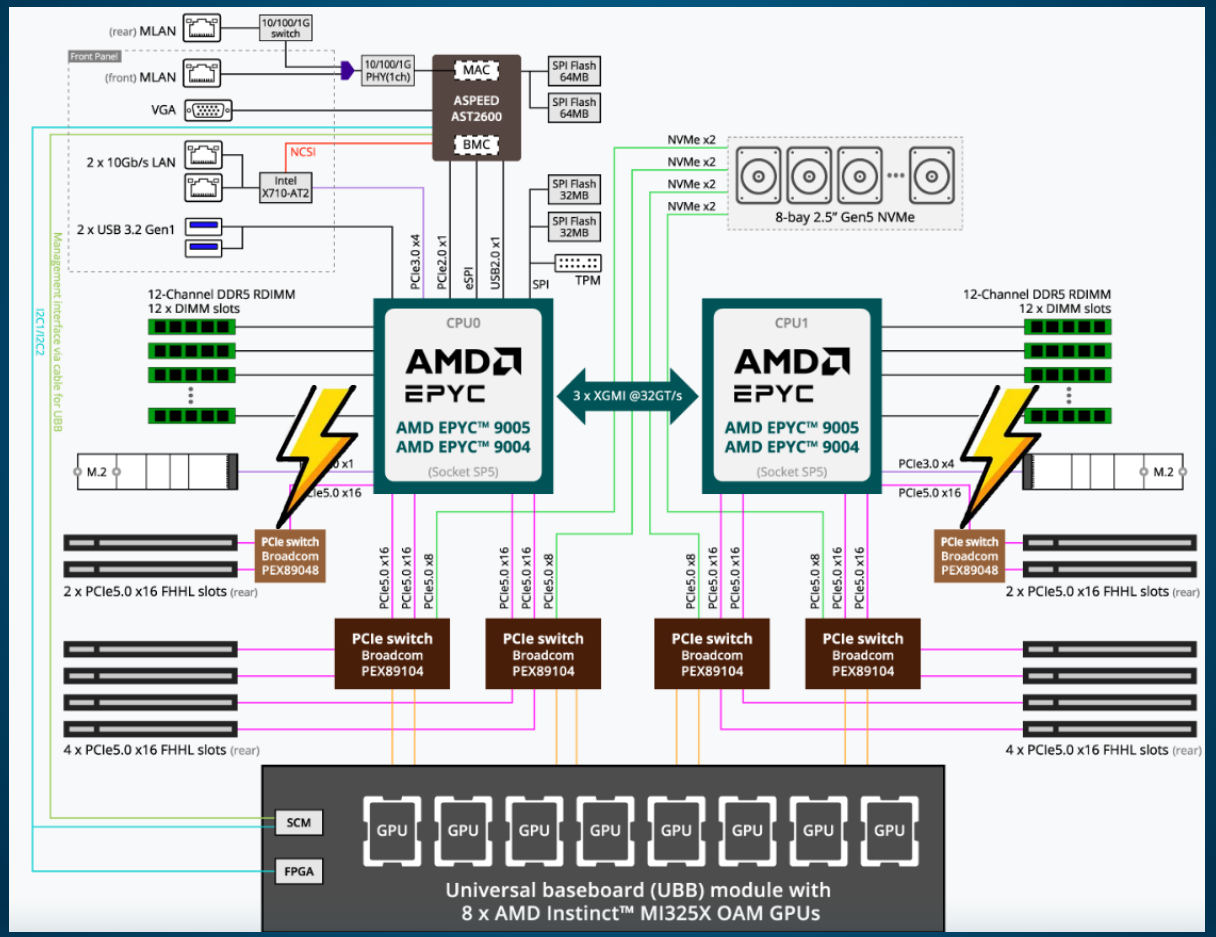
1 x Rear MLAN

3 x PCIe module cage

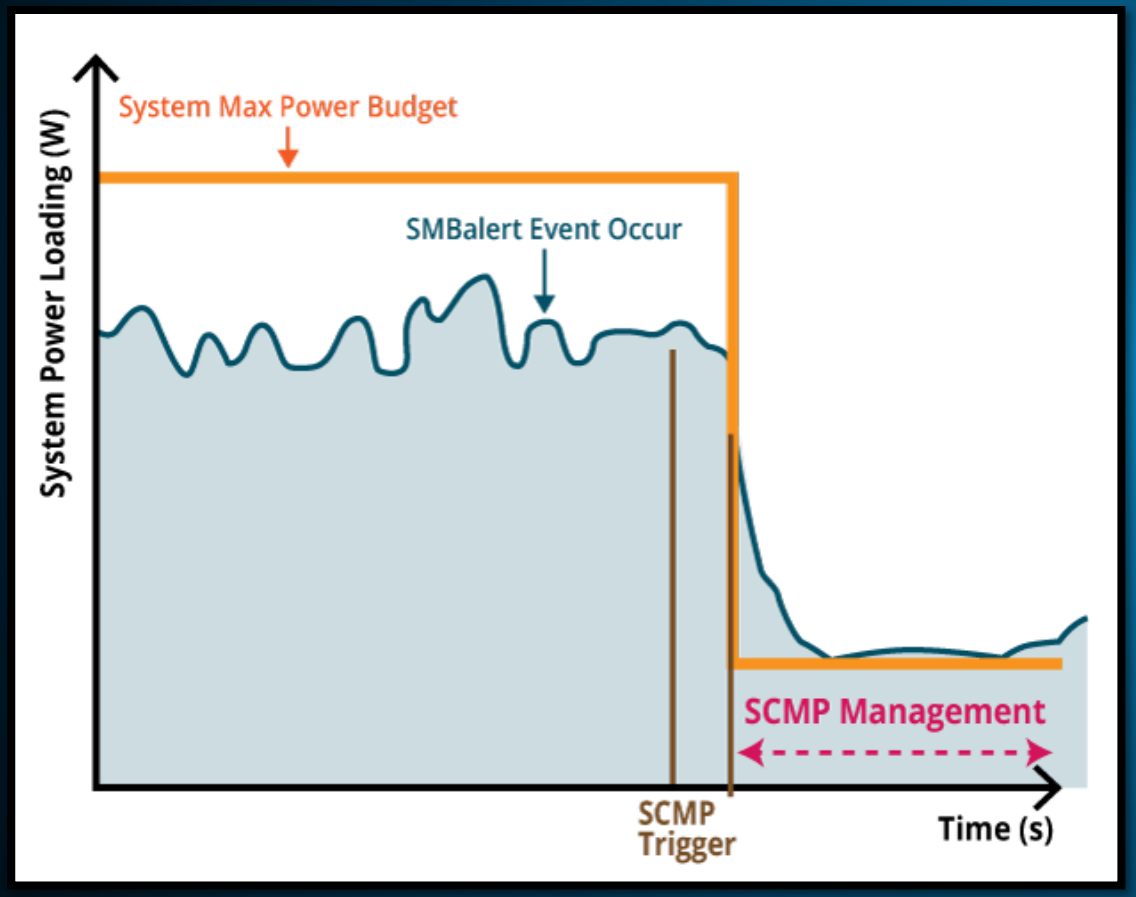


With full bandwidth PCIe Slots





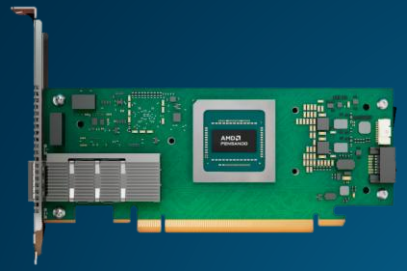
Crisis management while losing the power



Smart Crises Management and Protection (SCMP)

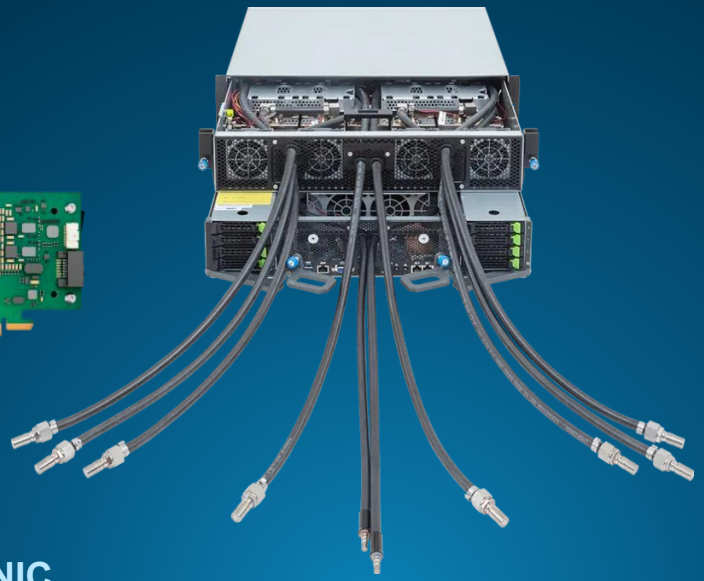
SCMP is a GIGABYTE patented feature which is deployed in servers with non-fully redundant PSU design. With SCMP, in the event of faulty PSU or overheated system, the system will force the CPU into an ultra-low power mode that reduces the power load, which prevents the system from unexpected shutdown and avoids component damage or data loss.

ETH Switches



Pensando™

Pollara 400 AI NIC



AMD CPU + GPU SERVER

GIGAPOD

WEKA / VAST storage

Rack / CDU

Object storage

Coldplate / Senser

GIGAPOD Partners

Infrastructure

Schneider Electric, VERTIV.

Comprehensive design across cooling, power and construction for better delivery the proper environment for server cluster.

Hardware Platform

AMD

Cooling Partner

motivair, USystems

Cooling unit and hardware-level part design for the best cooling efficiency.

Networking

BROADCOM, AMD, CISCO

Network switch partner for networking switch and topology design.

Software and storage

MYELINTEI — Certified Platform ready for sell

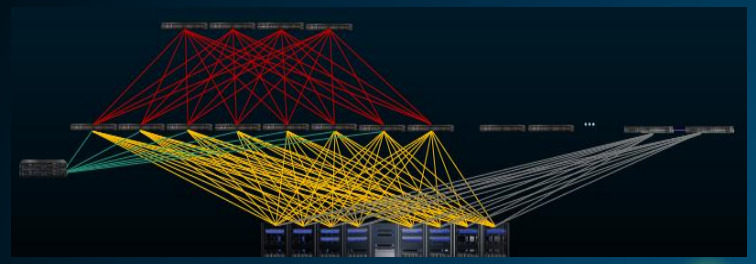
ddn

WEKA — Certified server ready for sell

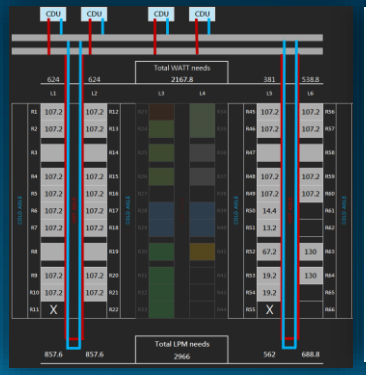
VAST — Certified server ready for sell

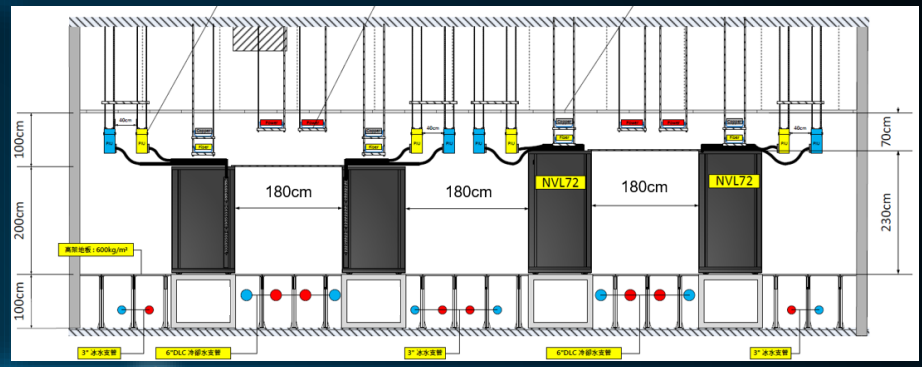
Software and storage with the right bandwidth design, features and performance for GIGAPOD.

network design

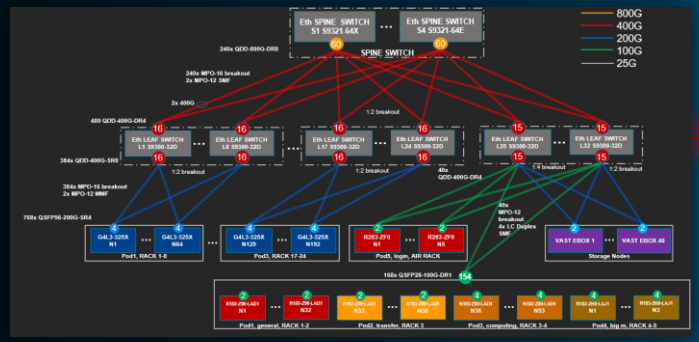


Infrastructure design includes placement, cooling and power

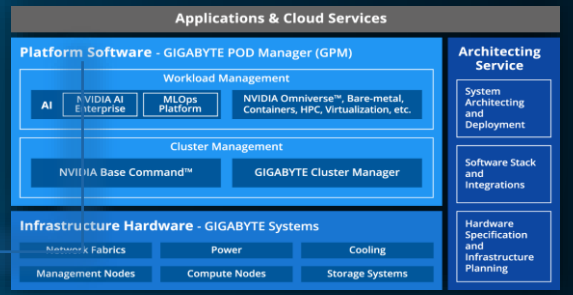




network design



POD Manager Platform



GIGA COMPUTING

感 謝 聆 聽