Exploring Supermicro AI Solutions Inside Out

Speaker: **Hendry Chang**, Senior Product Manager

Agenda

01 AMD Product Portfolio Lineup

02 Liquid Cooling Total Solutions

03 Software Management Solutions (SCC)

04 GPUaaS & AI Cloud

Industry Shifts: How do we design our Data Centers?

1960s — MAINFRAME COMPUTING

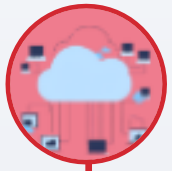
- Earliest attempts at deploying **Direct-to-Chip Liquid Cooling**

1990s — INTERNET WEB APPLICATIONS

- Supermicro introduces **building block solutions** to the world

2000s — CLOUD COMPUTING

- Platform became major players, driving widespread adoption of **virtualization** and **cloud**.

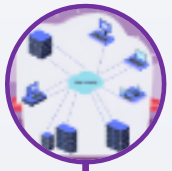


2010 — HPC & BIG DATA

- Supermicro introduces **Rack-Scale Solutions.**

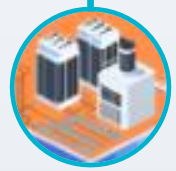
2018 — EDGE COMPUTING & 5G

- Supermicro introduces new **Edge Computing** and **IoT Solutions** at Embedded World



2023 & Beyond — GEN AI & LIQUID COOLING

- **Gnerative AI** challenges the world's power grid.
- Supermicro drives global **Direct-to-Chip Cooling** market adoption to 30%, integrating **Green Computing** as the core of AI supercomputing.



Supermicro H14 Systems with AMD EPYC™ 9005/9004 Series Processors

Hyper DP

Flagship Power and Flexibility Rackmount Server

CloudDC with **DC-SCM**

Ultimate Scalability and Flexibility for Cloud Data Center

GrandTwin®

High Density Combined with I/O Flexibility

FlexTwin™

Max Density, High-Performance Multi-Node System with Liquid Cooling

5U PCIe GPU System

GPU-Dense Servers Optimized for Compute Intensive Workloads

4U Liquid cooling Universal GPU System

For AI/ Deep Learning and HPC MI325X

8U Universal GPU System

Next Generation MI325 Machine Learning Platform

10U GPU System

Superior AI Applications and Large Language Models

* 500W support depends on thermal result

Advantage of Supermicro & Instinct™ MI300X

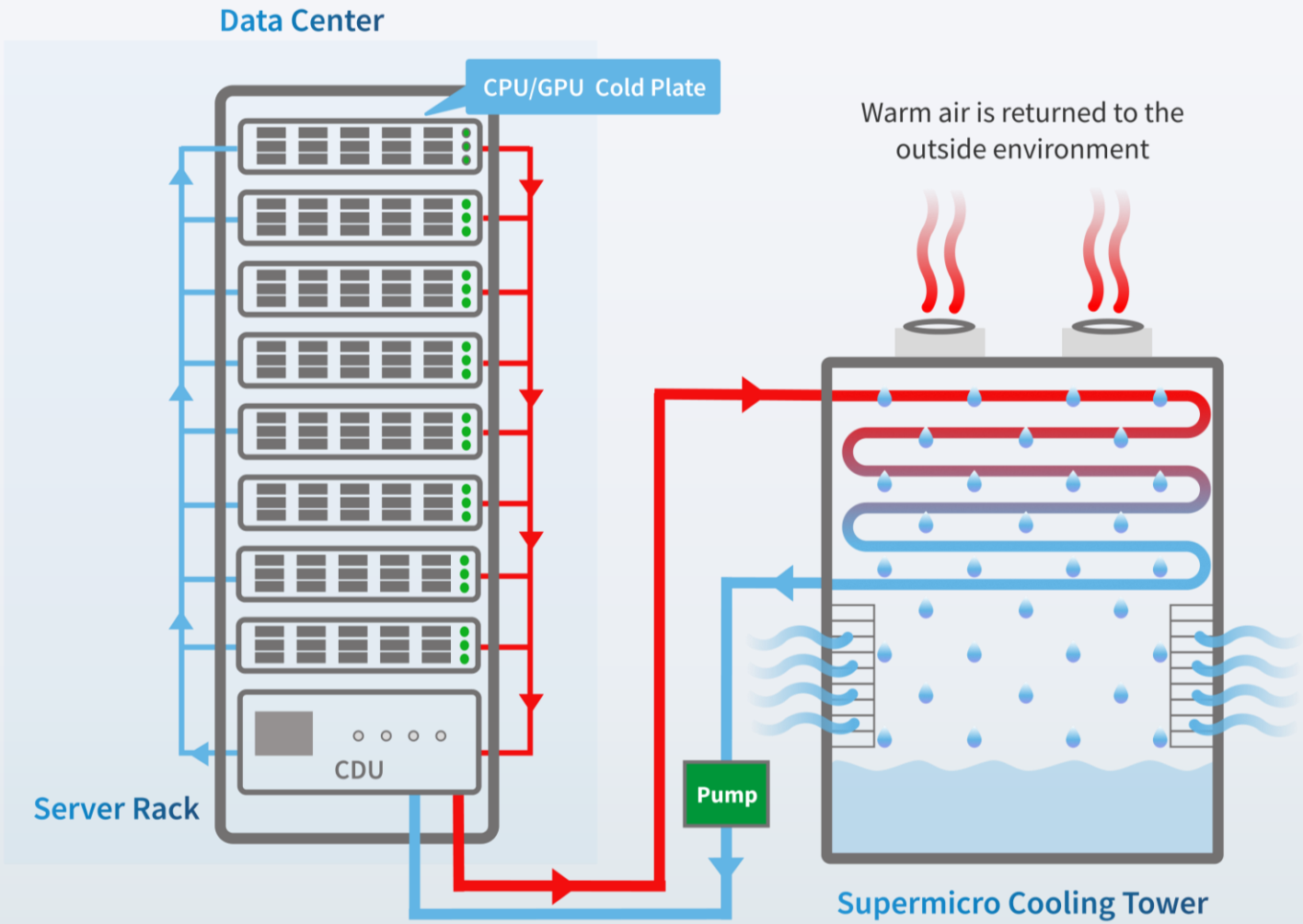
Supermicro Universal GPU Platform

- AI Training & Inferencing
- Front Access Design for Easy Maintenance
- Maximum Density for Space Saving
- Direct-to-Chip End-to-End Single Vendor Solution

Scan to learn more system detail!

AS-4125GS-TNMR2-LCC with 8-MI300X OAM Baseboard Shelf

Liquid Cooling Overview



How Direct-to-Chip Can Help

DLC Blooming AI Infra to Peak

- Allows GPUs to Reach Maximum Performance
- Lower Electricity Costs
- Lower Acoustic Noise
- Buy more, Save more!

Liquid Cooling Advantages

Up to **89%**

REDUCTION
Electricity Costs of Cooling Infrastructure in Server

Up to **40%**

REDUCTION
in Electricity Costs for Entire Data Center

Up to **55%**

REDUCTION
in Data Center Server Noise

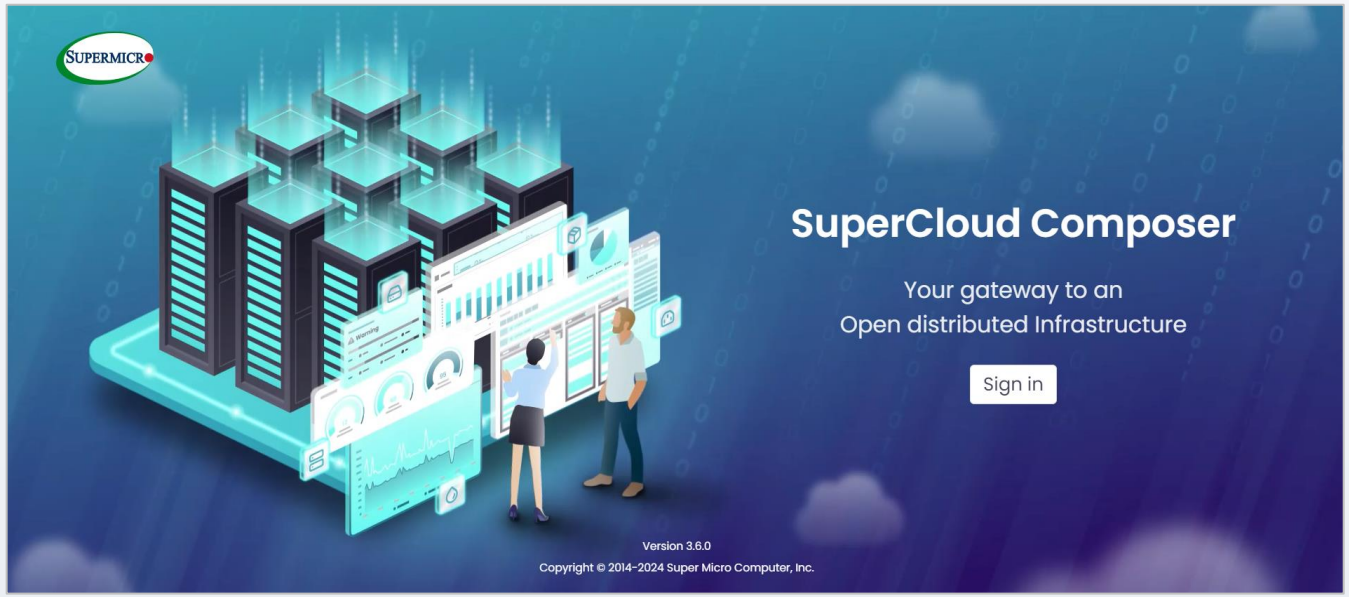
Liquid Cooling End-to-End Single Vendor

Design	Assembly & Configuration	Testing & Validation	Delivery & Deploy	Service Support
✓ Power Budget Consulting	✓ Rack level Installation of CDU, CDM, hoses, tubes etc.	✓ Complete rack level in house testing	✓ High quality packing and shipping	✓ Customer Training after Deployment
✓ Rack Layout Optimization	✓ Cabling and Labeling	✓ Multi-Vendor Compatibility test	✓ On site deployment services	✓ Seamless logistics for regional and global customers.
✓ BOM Creation	✓ Third Party Equipment Integration	✓ Full Rack Power Measurement	✓ Asset Labels, Unpacking Instructions, Customer User Manual	
✓ POC Available for large scale projects		✓ Performance Benchmarks		

Too Many Systems to Manage?

Supermicro's Got Your Back!

SuperCloud Composer (SCC)



Advantages:

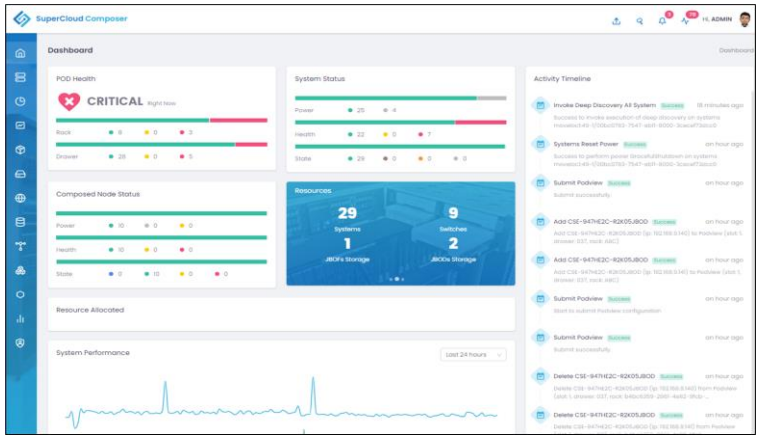
- GUI (Graphic User Interface) for easy server configuration
- Monitors Supermicro servers & 3rd party systems
- Server Monitoring with Historical Data
- Components to Cooling Tower Level Health Alert
- Reliable Scalability for servers of all sizes

SuperCloud Composer

SCC Liquid Cooling Management

SCC

IT Admin

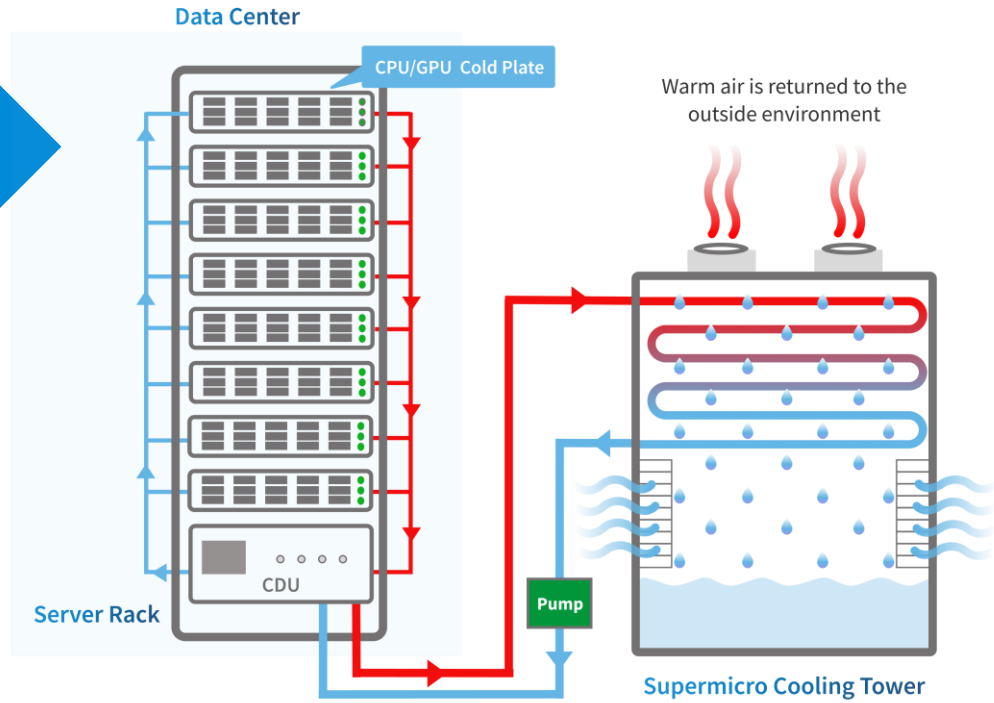


Management Network

SCC Features:

- Liquid Cooling Rack Space Management
- CDU and Cooling Tower Physical Assets Management
- CDU and Cooling Tower Live Sensor Reading with Historical Data

Supermicro Total Liquid Cooling Solution

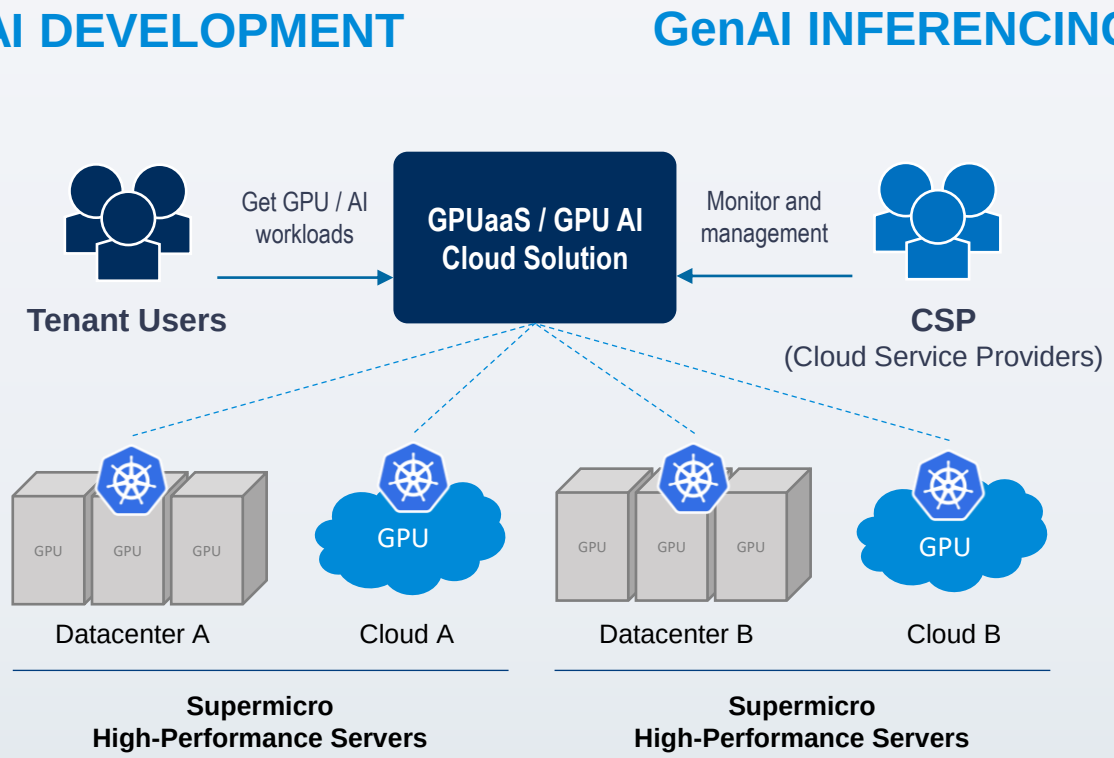


Supermicro
AI Cloud Platform Solution

Speed-up your GPU / AI Cloud Business

Enables Cloud Service Provider(CSPs) to Sell Compute Resources on Pay-as-You-Go Basis

- Efficient GPU allocation
- Unified containerized resource pooling
- Multi-tenant flexible shared/dedicated policy
- Streamlined GPU cloud operations



What is GPUaaS and AI Cloud?

GPUaaS — **Provides on-demand access to GPU resources over the cloud.**

Use Cases:

- AI / ML infrastructure provision
- High-performance data analysis

Customer Key Requirement:

- Scalability and Flexibility
- Cost Efficiency
- Multi-Tenant Architecture

AI Cloud — **Specifically designed to support AI workloads using GPUs.**

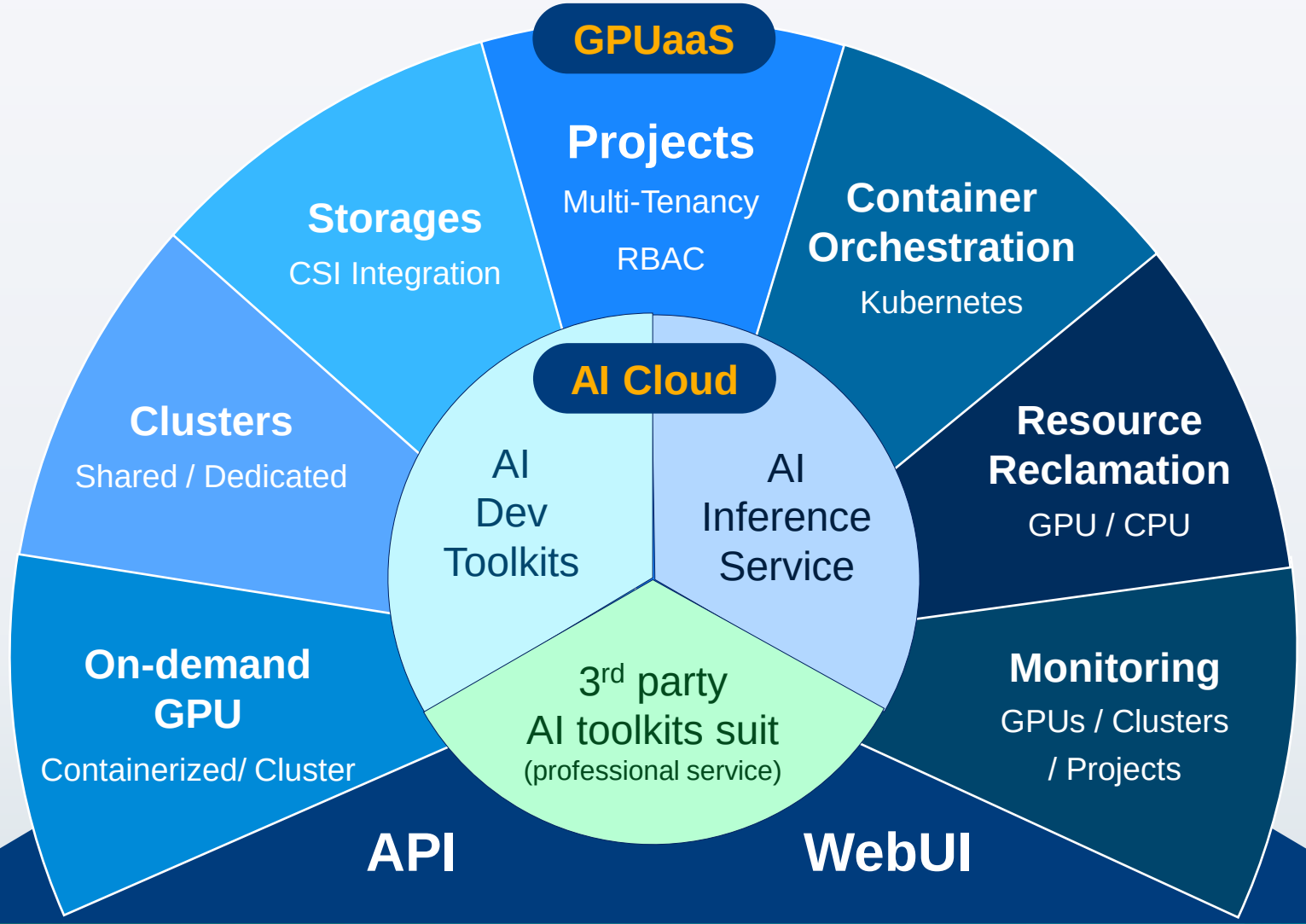
Use Cases:

- AI model training
- AI inferencing service

Key Benefits to End-user:

- End-to-End AI Environment
- Integrated AI Tools
- Generative AI & LLM Scalability
- All the GPUaaS requirements.

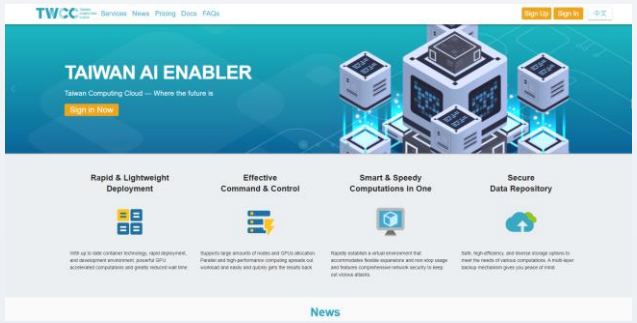
Supermicro GPUaaS / AI Cloud Platform Solution

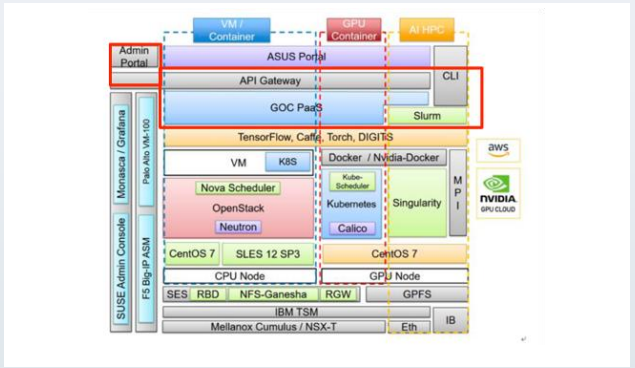


Success Story

Taiwan's Largest CSP for AI: NCHC

TWCC TAIWAN COMPUTING CLOUD



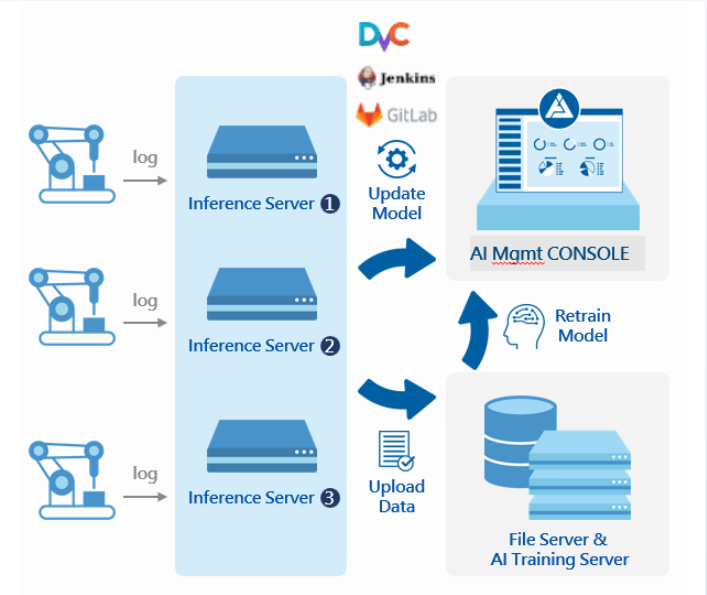


Well-Known Medical Institution Private GPU Cloud

北榮一號雲 (VGHTPE No.1 Cloud)

Enterprise Private GPU Cloud for Smart Manufactory

Vanguard International Semiconductor Corporation (VIS)




SUPERMICRO
AMD