# Data Center Building Block Solutions®

## For AI Factories

# Data Center Building Block Solutions® (DCBBS)

## End-to-End Solutions from System to Rack to Data Center Scale


Supermicro DCBBS delivers complete, modular AI infrastructure—from GPUs and networking to racks, site infrastructure, management software, and services.

**Rapid Deployment**
With First-to-Market

**Customized**
to Your Workloads

**Total Solutions**
with Onsite Services

**Tested and Validated**
Before Shipping

# DCBBS Components

## AI and Compute Systems

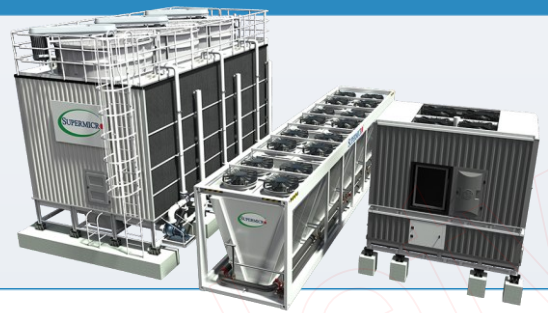
Industry's broadest AI and compute system portfolio with unmatched customization—from GPU systems to petascale storage, multi-node systems, and liquid cooling.

## In-Rack Solutions

Complete in-rack integration—cooling distribution units & manifolds, hose kits, networking switches, power shelf, and battery backup unit, RDHx for maximum density and efficiency.

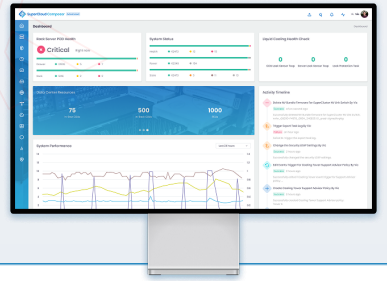
## In-Row Solutions

Scalable liquid cooling from 200kW Liquid-to-Air Sidecars to 1.8MW In-Row CDUs, plus pre-validated SuperCluster solutions with integrated networking for turnkey multi-rack deployment.



## Site Infrastructure Solutions

Site-level cooling infrastructure with scalable water and dry cooling towers for thermal and power management. Optimized cabling design and implementation for high-performance data pipeline.



## Management Software Suite

End-to-end SuperCloud software from bare metal to AI workloads—delivering unified infrastructure control, deployment automation, developer tools, and multi-tenant GPU cloud management.

## Services and Onsite Deployment

Comprehensive professional services from data center design and solution validation through on-site deployment and ongoing support, with 4-hour response options for mission-critical uptime.

# As Part of Data Center Building Block Solutions®

**Service Building Blocks**

- Onsite Deployment
- Full Cluster Testing
- Software Integration
- L11 & L12 Testing
- Networking & Cabling
- Racking & Stacking
- PoC Service

**Data Center Management Software**

**Networking Infrastructure**

Compute Network | In-band Network | OOB Network | Storage Network

**L11 Rack Level Building Blocks**


**L10 Systems**

GPU Server | Rackmount Server | Multi-node Server

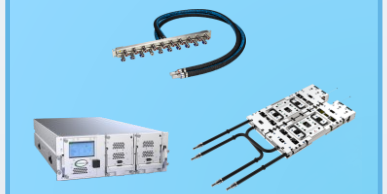
**L10 Storage**

Storage Server

**Liquid Cooling**

**Cooling Tower**

**Components**



Floor plan design | Power resource planning | Water resource planning | Datacenter Cooling Design | DC Project Management

# Supermicro DLC-2

## Up to 40% More AI per Watt with The Next Generation Direct-to-Chip Liquid Cooling Solutions

| Power Savings | Water Savings | System Heat Capture | Quiet Data Center | Space Savings |
|---|---|---|---|---|
| up to **40%** | up to **40%** | up to **98%** | as low as **50dB** | up to **60%** |
| Savings in entire data center (vs. Air-cooling) by using Supermicro DLC-2 | Savings with 45 °C warm water operation and eliminating chilled water and compressor | Heat capture in DLC-2 Liquid-cooling with CPU, GPU, PCIe Switch, DIMM, VRM, PSU, and more | Significantly reduces noise with less fans and fan speed. As quiet as a library | Savings with more than 2.5x compute density compared to air-cooled systems |


Supermicro DLC-2 reduces power costs by up to 40%, accelerate time-to-deployment and time-to-online, and allow data centers to run more efficiently with lower PUE, while **lowering TCO by up to 20%**

# AI and Compute Systems

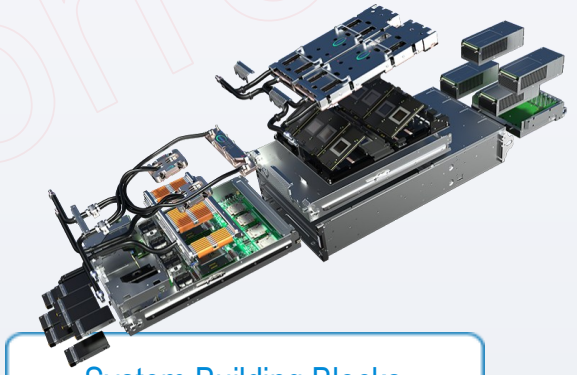
Industry's broadest AI and compute system portfolio with unmatched customization — from GPU systems to petascale storage, multi-node systems, and liquid cooling.

## GPU and CPU Systems

Broad range of density, efficiency, and scalability-optimized systems featuring the latest AI and compute technologies in various form factors designed for diverse workloads—from AI training and inference to high-performance computing.

## Storage Systems

Petascale and object storage servers with seamless integration across leading software-defined networking partners, delivering maximum throughput and scalability for data-intensive AI workloads.



## System Building Blocks

Supermicro-designed liquid cooling cold plates that efficiently remove up to 98% of heat from critical electronics in systems, plus modular building block components that enable precise system customization and optimization for specific workload requirements.

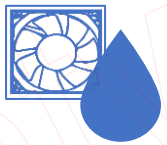
# GPU and CPU Systems

- Complete portfolio of NVIDIA GPU-accelerated systems featuring the latest NVIDIA Blackwell, and Grace Blackwell architectures, optimized for AI training, inference, and generative AI workloads at scale.
- Variety of DLC systems optimized for high performance GPU+CPU compute workloads
- Up to 98% heat capture coverage for system components through Supermicro DLC-2
- Lower fan speeds and fewer required fans save more power, and achieve a quiet data center at as low as 50dB

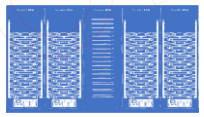
# The Most Powerful and Efficient NVIDIA Blackwell Architecture Solutions

The Supermicro Building Block portfolio provides the core infrastructure elements necessary to scale the NVIDIA Blackwell Platform, delivering the pinnacle of AI training and inference performance.

**Next-Gen**
Performance



**Air or Liquid**
Cooling Options

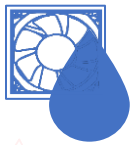


**Scalable**
Cluster-Ready

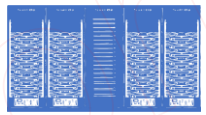
# Supermicro Blackwell Ultra Solutions

Supermicro integrates Blackwell Ultra generational advancements with system-level and rack-level engineering to create a solution that maximizes efficiency and performance at scale.

**Performance**
Up 288GB HBM3e per GPU



**Air or Liquid**
Cooling Options



**Scalable**
800G-Ready

# Supermicro's NVIDIA Blackwell Solutions Overview

## 4U Liquid-Cooled NVIDIA HGX B200 System

8-GPU SXM per system | 1.4TB HBM3e per SXM | Dual Socket CPUs up to 500W

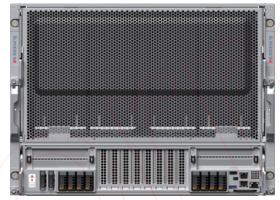
## 10U Air-Cooled NVIDIA HGX B200 System

8-GPU SXM per system | 1.4TB HBM3e per SXM | Dual Socket CPUs up to 500W

## 4U Liquid-Cooled Front I/O DLC-2 NVIDIA HGX B200 System

8-GPU SXM per system | 1.4TB HBM3e per SXM | Dual Socket CPUs up to 350W

## 8U Air-Cooled Front I/O NVIDIA HGX B200 System



8-GPU SXM per system | 1.4TB HBM3e per SXM | Dual Socket CPUs up to 350W

## 48U Liquid-Cooled NVIDIA GB200 NVL72

4 GPUs, 2 CPUs per compute tray | 13.4TB HBM3e per rack | NVIDIA Grace CPU

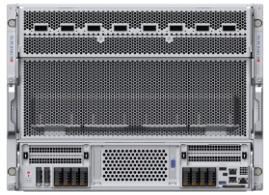
### Blackwell Ultra Features →

- Up to 288GB HBM3e per GPU
- Up to 800G network speeds via integrated NVIDIA ConnectX®-8 SuperNICs

## 4U Liquid-Cooled Front I/O DLC-2 NVIDIA HGX B300 System

8-GPU SXM per system | 2.1TB HBM3e per SXM | Dual Socket CPUs up to 350W

## 8U Air-Cooled Front I/O NVIDIA HGX B300 System



8-GPU SXM per system | 2.1TB HBM3e per SXM | Dual Socket CPUs up to 350W

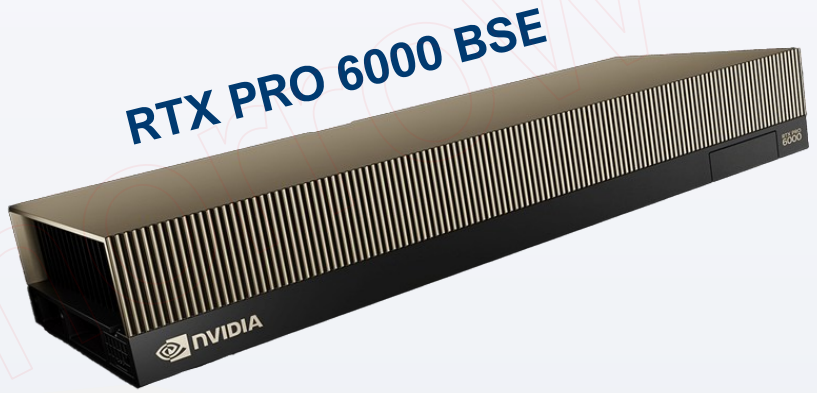
## 48U Liquid-Cooled NVIDIA GB300 NVL72

4 GPUs, 2 CPUs per compute tray | 21TB HBM3e per rack | NVIDIA Grace CPU

# Supermicro PCIe GPU System For Enterprise AI

## System Supporting NVIDIA RTX PRO 6000 BSE

- **Supermicro** offers broad range of systems optimized for latest **NVIDIA RTX PRO 6000 Blackwell Server Edition**, ideal for AI-enabled enterprise applications.
- These systems provide powerful and cost-effective **multi-workload acceleration** for large language model (LLM) inference and fine-tuning, visualization, graphics & rendering, and virtualization.




PCIe GPU

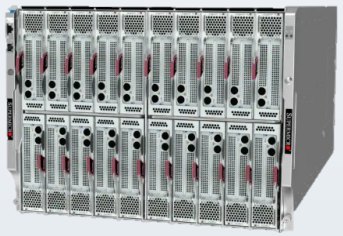
NVIDIA MGX

Edge-optimized

Rackmount

Multi Node

Blade Systems



**Note:** This is not an exhaustive list. Additional systems will be qualified based on customer demand and business ROI

# Supermicro's NVIDIA RTX PRO 6000 BSE

## PCIe-Optimized Systems Supporting NVIDIA RTX PRO 6000 Blackwell Server Edition

### 2U CloudDC / SYS-222C-TN — 2-GPU


Dual Xeon 6700 CPUs, 6 PCIe 5.0 x16/x8 slots, 2 kW CRPS PSU. Up to 24 NVMe front drives in a DC-MHS chassis.

2-2-3 architecture:

- **2 (Dual)** Intel Xeon 6700 series CPUs
- **2 GPU** PCIe per system (Up to 2-GPU)
- **3 NIC:** E/W: 2x BF3 B3140H, N/S: BF3 B3220
- **Alternate NIC**: CX7 2x200G

### 4U MGX / SYS-422GL-NR — 4-GPU


Eight GPU-ready slots; ship with four RTX PRO 6000s for a 2.4 kW draw. Plenty of NIC slots for 2× 200 GbE or IB.

2-4-3 architecture:

- **2 (Dual)** Intel Xeon 6900 series CPUs
- **4 GPU** PCIe per system (Up to 8-GPU)
- **3 NIC:** E/W: 2x BF3 B3140H, N/S: BF3 B3220
- **Alternate NIC**: CX7 2x200G

### 4U MGX / SYS-422GL-NR — 8-GPU


Dual-root PCIe 5 ×16 lanes feed every slot; four 3.2 kW PSUs keep all 8 GPUs locked at 600 W for steady peak performance.

2-8-5 architecture:

- **2 (Dual)** Intel Xeon 6900 series CPUs
- **8 GPU** PCIe per system (Up to 8-GPU)
- **5 NIC**: E/W: 4x BF3 B3140H, N/S: BF3 B3220
- **Alternate NIC:** CX7 2x200G

### 5U GPU SuperServer / SYS-522GA-NRT or AS -5126GS-TNRT2 — 8-GPU


Intel AMD

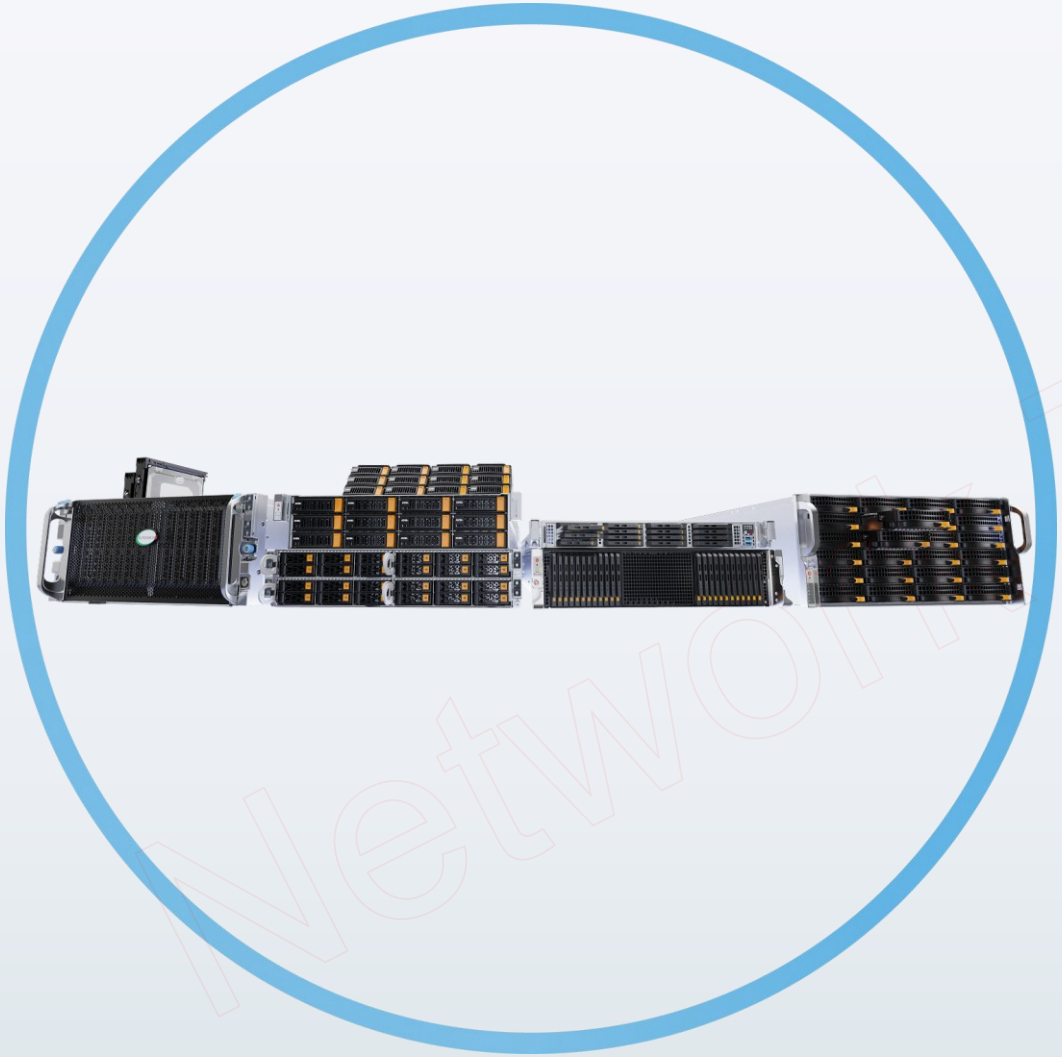
13 PCIe 5 x16 slots & 24 front U.2/U.3 NVMe bays. Taller 5U plenum + 8.6kW usable power keeps GPUs & DPUs cool together.

2-8-5 architecture:

- **2 (Dual)** Intel Xeon 6th / AMD EPYC 4th CPUs
- **8 GPU** PCIe per system (Up to 8-GPU)
- **5 NIC:** E/W: 4x BF3 B3140H, N/S: BF3 B3220
- **Alternate NIC**: CX7 2x200G

**Based on NVIDIA's Reference Architectures,** Digits refer to: # of sockets (CPUs), # of GPU, # of network adapters (For North/South and East/West)

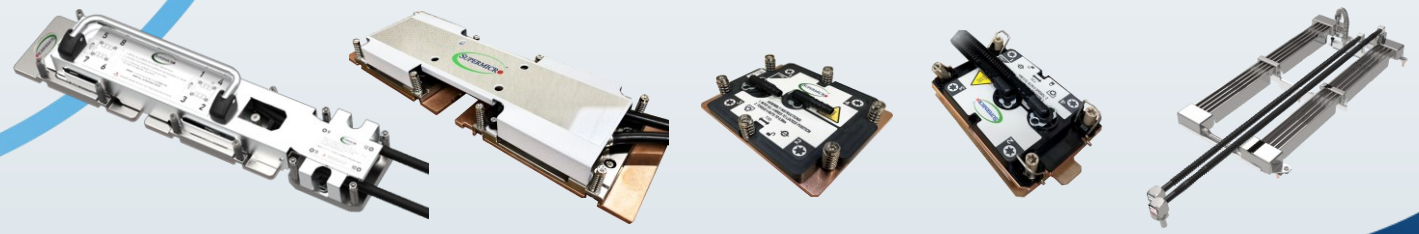
# Storage Systems



- **Petascale and Object Storage Systems** – High-capacity storage servers engineered for massive data sets, supporting petascale deployments with flexible object storage architectures optimized for AI data lakes and distributed workloads.
- **Software-Defined Storage and Memory Integration** – Seamless integration with leading software partners ensures optimized data flow, reduced latency, and maximum throughput across storage and compute infrastructure.
- **AI Workload Optimization** – Purpose-built for data-intensive AI applications including large-scale model training, high-volume inference, and real-time data pipeline operations requiring extreme scalability and performance.

# System Building Blocks and Cold Plates

- DLC-2 cold plates are mounted on top of CPUs, GPUs, DIMMs, VRM, PCIe switches, and more for up to 98% heat capture coverage
- GPU cold plate also helps cool other related components that require liquid cooling, such as GPU switches
- Advanced heat dissipation efficiency with micro-sized channels thinner than a human hair



# In-Rack Solutions

Complete in-rack integration—cooling distribution units & manifolds, hose kits, networking switches, power shelf, and battery backup unit, RDHx for maximum density and efficiency.

## CDUs and CDMs

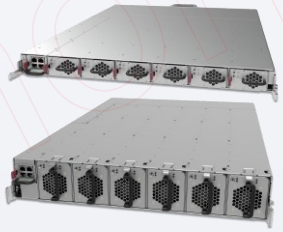
Coolant Distribution Units and Manifolds in various configurations, including vertical CDMs for higher server densities. In-Rack CDUs accommodate up to 250kW of cooling capacity for liquids up to 45°C, with redundant power supplies and pumps.

## Rear Door Heat Exchanger

Rear Door Heat Exchangers (RDHx) attach directly to server racks to remove heat at the source, significantly reducing data center air conditioning requirements while maintaining optimal operating temperatures.

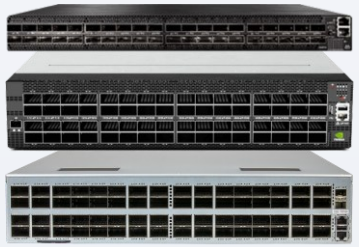
## Cooling Hose Kit

Flexible hose kits designed to connect liquid-cooled racks directly to data center primary water supply or cooling towers. Pre-installed hose kits enable plug-and-play, toolless deployment with seamless integration into new or existing facility piping.



## Power Shelves & BBUs

33kW Power Shelves engineered for the latest generation of high-density AI racks, combined with Battery Backup Units providing 90 seconds of 48V DC power for application checkpointing during power failures—eliminating the need for complete application restarts.



## Networking Switches

NVIDIA Ethernet and InfiniBand platforms including NVIDIA Spectrum-X and Quantum-2 for optimized AI cluster networking, as well as qualified partner solutions. Supermicro Ethernet switches with up to 800 GbE link speeds and 51.2 Tbps throughput are also available.

# In-Rack Coolant Distribution Unit (CDU)


- DLC-2's 250kW in-rack CDU capacity and the up to 98% heat capture allow for an increased inlet liquid temperature at up to 45 °C
- Warm water operation eliminates need for chilled water, chiller compressor, and reduces equipment cost, saving up to 40% of water consumption
- Intelligent monitoring and flexible management via CDU touch panel, remote access, and full integration with SuperCloud Composer
- 1.8 MW capacity in-row CDU also available

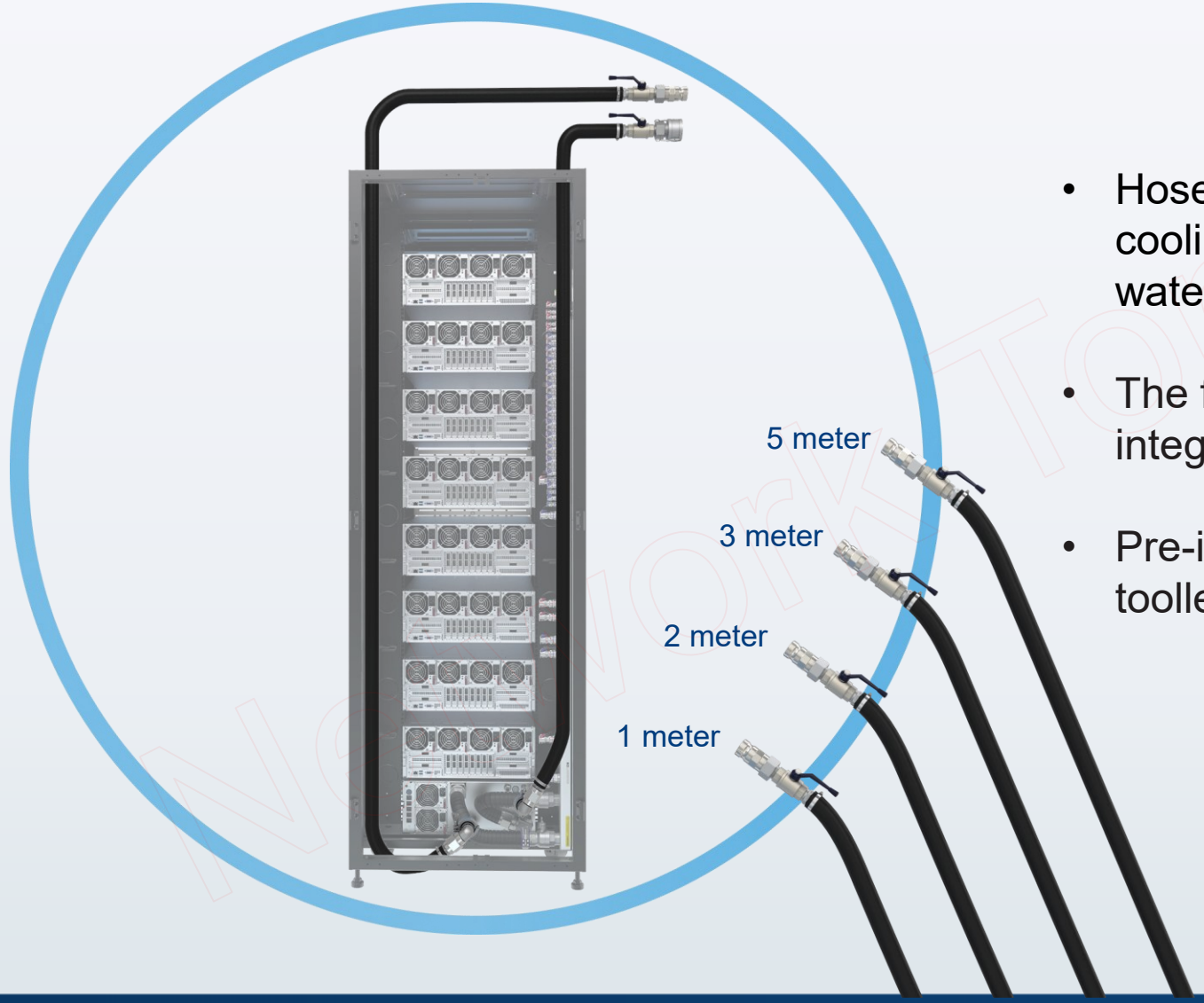
# Coolant Distribution Manifold (CDM)

- CDMs distribute coolant to each server in the rack and return the hot coolant back to the CDU
- DLC-2 vertical CDMs save rack space further and increase density
- Supports 42U, 48U, or 52U rack configuration
- Customizable spacing and size for Quick Disconnect Couplings (QDC) with liquid drip free, one-handed operation and maintenance

# Rear Door Heat Exchanger

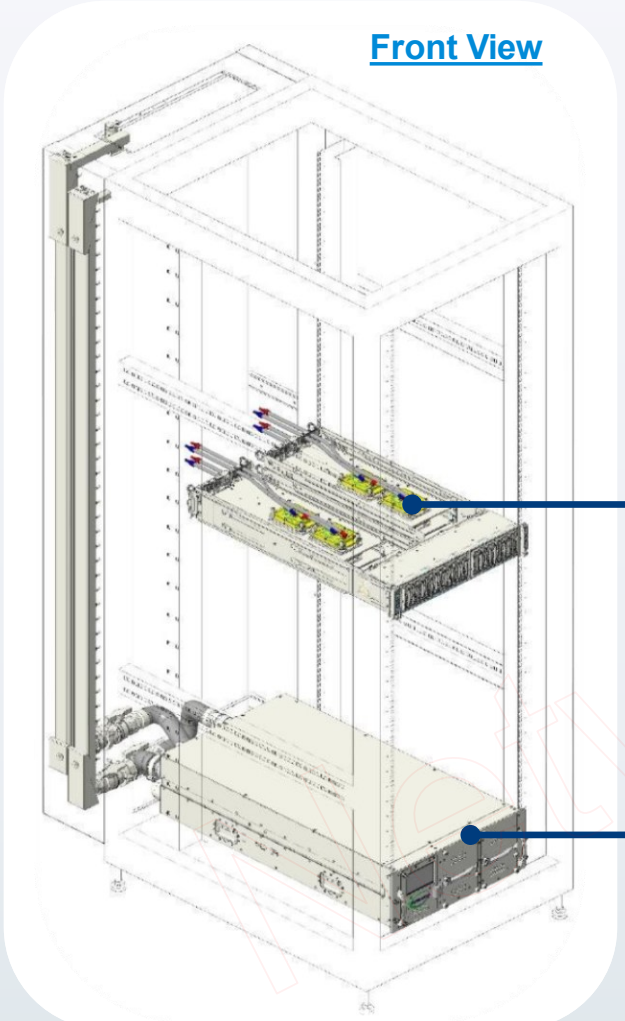
- Optional liquid-cooling solutions are also available to enable DLC infrastructure performance in challenging or restrictive environments
- Rear Door Heat Exchangers attach directly to server racks to remove heat at the source
- Significantly reduces data center air conditioning requirements while maintaining optimal operating temperatures for high-density AI infrastructure.

# Cooling Hose Kit



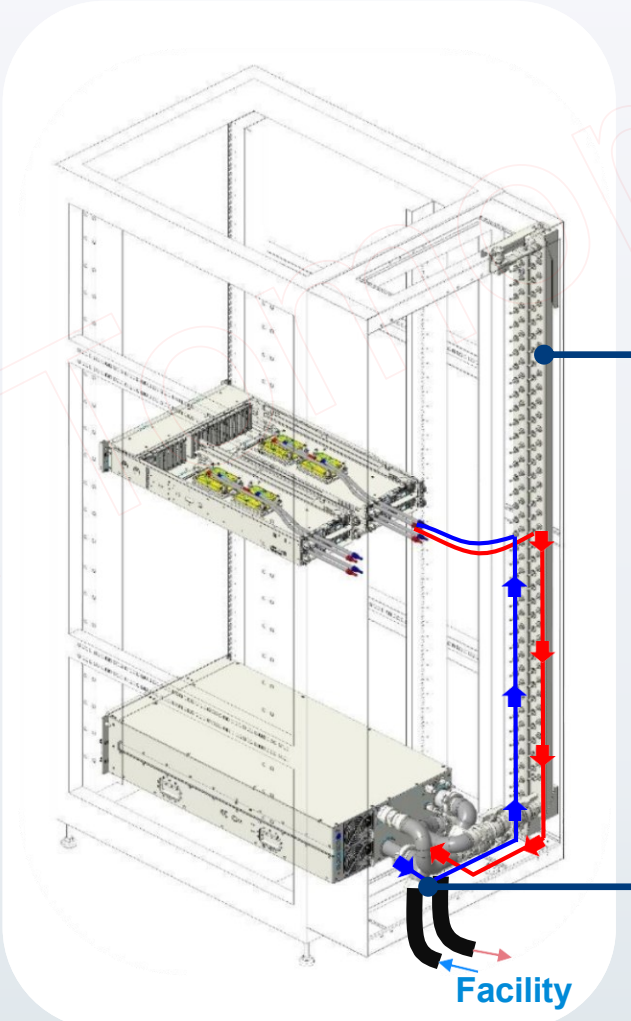

- Hose Kits are designed to easily connect liquid cooling racks directly to data center primary water supply or cooling towers
- The flexibility of hose kits ensure seamless integration with new or existing facility piping
- Pre-installed Hose-Kit in rack for plug-and-play toolless deployment

# Liquid-cooled Rack Overview




1 **Cold Plate**

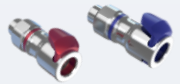
2 **CDU**
**(Coolant Distribution Unit)**




1 **CDM + QDC**

**CDM (Coolant Distribution Manifold)**
- Rear and front liquid I/O

**QDC (Quick Disconnect Couplings)**



2 **Hose Kit**
Connect between rack and datacenter facility

# Power Shelves and Battery Backup Units (BBUs)

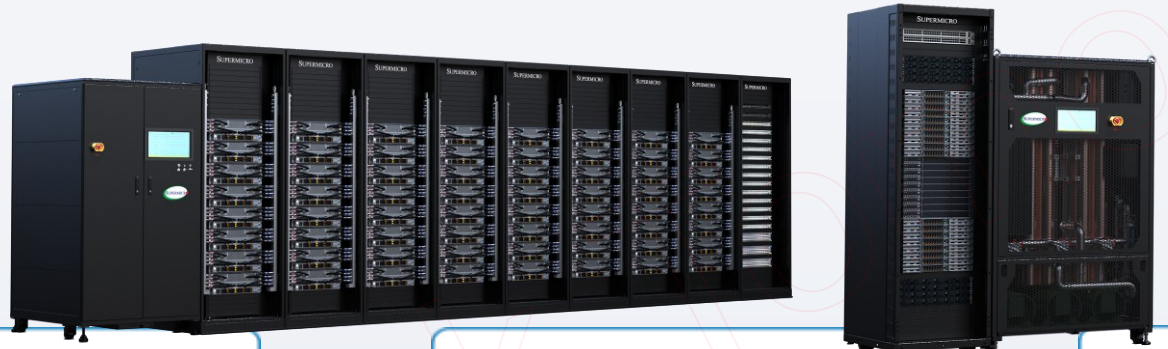
- Power Shelves for the latest generation of racks of integrated AI servers and switches, designed for 33kW per shelf in various rack sizes.
- Battery Backup Units (BBU) are critical for mission-critical workloads, providing 48V DC backup power of 33kW for 90 seconds in the event of a power failure

# Networking Switches

- NVIDIA Ethernet and InfiniBand platforms including NVIDIA Spectrum-X and Quantum-2 for optimized AI cluster networking, as well as qualified partner solutions.
- Supermicro Ethernet switches with up to 800 GbE link speeds and 51.2 Tbps throughput are also available for diverse workload requirements.

# In-Row Solutions

Scalable liquid cooling from 200kW Liquid-to-Air Sidecars to 1.8MW In-Row CDUs, plus pre-validated SuperCluster solutions with integrated networking for turnkey multi-rack deployment.





### In-Row CDUs

High-capacity In-Row Coolant Distribution Units designed to accommodate multiple high-powered liquid-cooled server racks, delivering up to 1.8MW of cooling capacity for scalable, multi-rack deployments in dense AI data center environments.

### Liquid-to-air (L2A) Sidecar

Independent cooling units that convert hot liquid from servers back to cool liquid within the data center, providing up to 200kW of cooling capacity with no external infrastructure modifications—ideal for retrofitting and rapidly deploying liquid-cooled systems.

### SuperCluster

Highly customizable, plug-and-play cluster solutions with multiple racks and integrated networking fabric. Each system undergoes L11 or L12 validation testing before shipping, ensuring turnkey deployments that enable seamless scaling from pilot projects to full production AI infrastructure.

# In-Row Coolant Distribution Unit (CDU)


- In-Row CDU supports cooling capacity up to 1.8 MW, empowering reliable high-density cluster deployments for AI/HPC workloads.
- Redundant N+1 pump design with Automatic Transfer Switch (ATS) ensures uptime and improves system reliability
- Redundancy control ensures power continuity during critical operations
- Touch screen and web interface simplify operation while enhancing user experience
- In-Row cabinet design delivers high-capacity cooling, centralized liquid distribution, and flexible row-level deployment

# Liquid-to-Air (L2A) Sidecar

- L2A Sidecar system supports cooling capacity up to 200 kW while enabling high-density deployments
- Redundant N+1 pump design with Automatic Transfer Switch (ATS) ensures uptime and improves system reliability
- Redundancy control ensures power continuity during critical operations
- Touch screen and web interface simplify operation while enhancing user experience
- Side-mounted CDU design enables efficient liquid circulation while optimizing heat rejection

# SuperCluster with L10 – L12 Testing and Validation

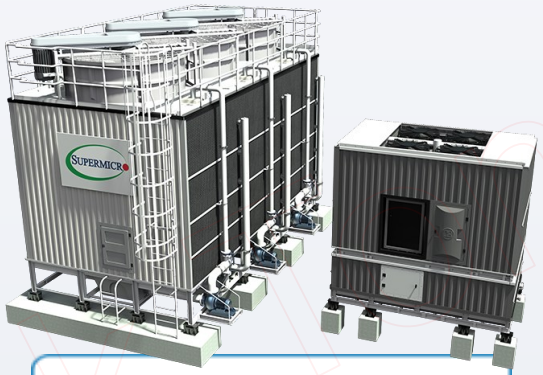

- DLC-2 enabled racks and clusters are tested and validated at L10 system level, L11 rack level and L12 cluster level before shipping
- L11 testing is conducted to ensure components meet performance, reliability, and integration standards before deployment, from compute nodes, networking, power distribution, cooling systems
- L12 testing evaluates cluster systems under different benchmarks, ensuring optimal configuration and functionality across all modern workloads to simplify deployment and accelerate time-to-online

# Site Infrastructure Solutions

Scalable liquid cooling from 200kW Liquid-to-Air Sidecars to 1.8MW In-Row CDUs, plus pre-validated SuperCluster solutions with integrated networking for turnkey multi-rack deployment.

## Cabling

Optimized traffic engineering with complete cabling design, documentation, and implementation services including routing, port mapping, and cable length optimization. Engineered solutions reduce time-to-online while ensuring high-performance data flow for cluster deployments.



## Water Cooling Towers

Primary liquid loop facility supporting 1 MW to 50+ MW capacity with modular design for faster deployment. Closed-loop with EC fans and N+1 pumps ensures reliable, energy-efficient operation with lower power and water usage.

## Dry Cooler (Coming Soon)

Delivers low PUE/WUE cooling for water-constrained regions, scaling from 1 MW to 50+ MW. Adiabatic-assisted air pre-cooling excels in high-temperature environments. Closed-loop design keeps coolant clean and non-corrosive for efficiency, minimal maintenance, and reliability.

# Cabling

- **Complete Cabling Design and Engineering** – Optimized traffic engineering with full design and documentation including routing, port mapping, and cable length optimization for maximum efficiency and performance.
- **Professional Implementation Services** – End-to-end cabling implementation that ensures tight integration across all infrastructure layers, reducing complexity and deployment risks.
- **Cost and Time Optimization** – Reduces time-to-online, material costs, and labor expenses while improving overall deployment quality and customer experience through engineered solutions.

# Water Cooling Tower

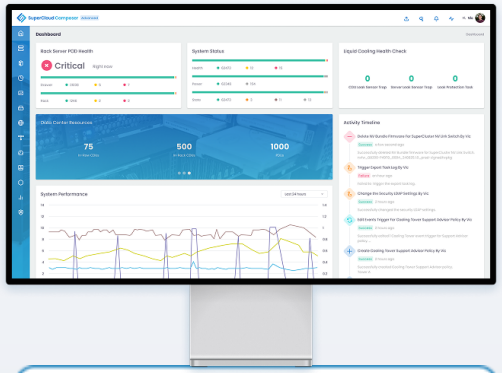

- Primary liquid loop facility supports scalable cooling capacity from 1 MW up to 50 MW or more
- Modular design enables easy deployment for faster time-to-operation
- Closed-loop design ensures reliability while minimizing risk and maintenance needs
- EC fan design maximizes energy savings
- N+1 submersible pump enhances reliability and delivers quieter operation
- Reduced power and water usage versus conventional air-cooled systems

# Dry Cooler (Coming Soon)


- Air cooled heat rejection system where water consumption is not required
- Closed loop design to optimize system reliability
- Simple mechanical set up due to use of air rather than water
- Integrated controls for real-time monitoring and system management

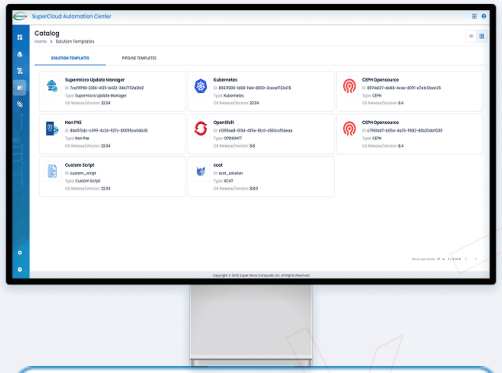
# Management Software Suite

End-to-end SuperCloud software from bare metal to AI workloads—delivering unified infrastructure control, deployment automation, developer tools, and multi-tenant GPU cloud management.



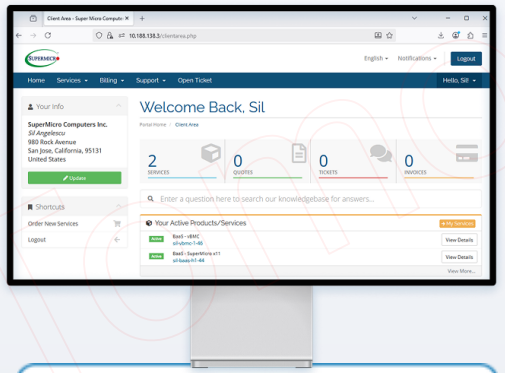
## SuperCloud Composer (SCC)

Unified rack-scale and liquid cooling management across servers, networks, PDUs, CDUs, and 3rd-party systems. Provides power management, advanced leak detection, protection, and alerts to safeguard GPU and cooling infrastructure. With observability and tuning at scale.



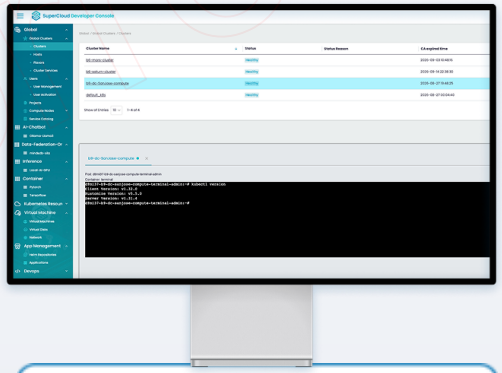
## SuperCloud Automation Center (SCAC)

Provides pre-built, enterprise grade automation for data center and edge infrastructure. It covers everything from firmware and OS provisioning to Kubernetes and AI workload enablement while ensuring security, scalability and governance.




## SuperCloud Developer Experience Console (SDX)

Developer-oriented AI experience console that brings GPU as a Service, Kubernetes, Machine Learning pipelines, and data services into one workspace. Delivers AI-native workflows, and empowering teams to go from code to model faster and safer.

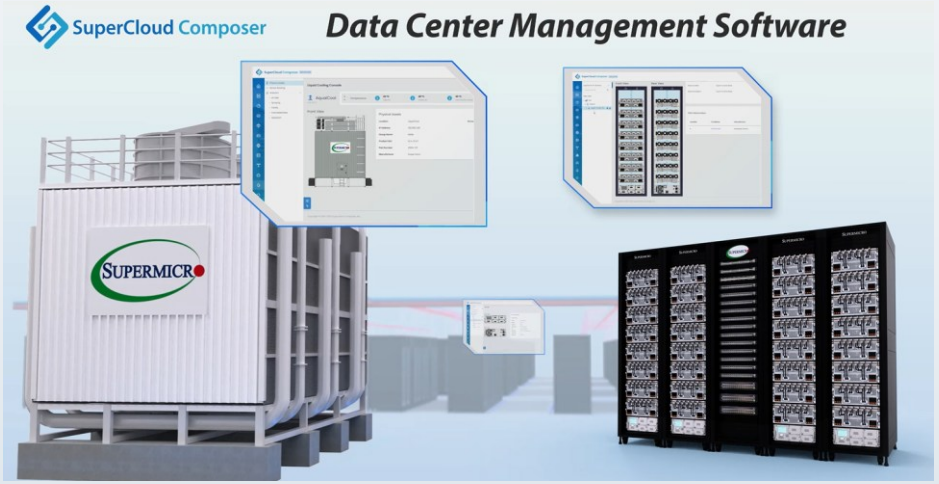


## SuperCloud Director (SCD)

Delivers multi-tenant AI cloud control with integrated bare metal, Ethernet & Infiniband network multi-tenancy, and storage management, optimizing performance, security, and developer agility. Purpose-built for GPU as a Service, AI factory deployments, and hyperscale operations.

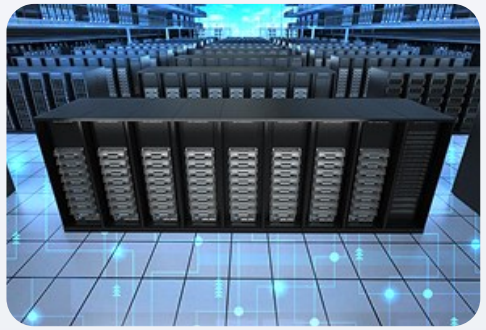
# SuperCloud Composer

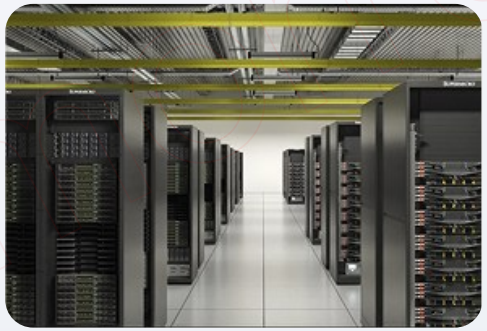

- SuperCloud Composer® (SCC) Liquid Cooling Consult Module collects and monitors vital real-time information from GPU, CPU, DIMM, CDU, PDU, and Cooling Tower to ensure maximum operating efficiency
- Provides lifecycle management and orchestration of liquid-cooled data centers



# Services and Onsite Deployment

Comprehensive professional services from data center design and solution validation through on-site deployment and ongoing support, with 4-hour response options for mission-critical uptime.




## Planning and Validation

Projects are thoroughly planned, including designs for the data center floor plan, rack layout, port mapping, and more. Before shipping, Supermicro goes beyond industry standards with testing that includes rack (L11) and cluster-level (L12) validation.

## Onsite Deployment

End-to-end deployment starts with Supermicro's white-glove delivery service, ensuring the utmost care. On-site service team members handle racking and stacking, with optional software installation available.

## Onsite Support

Supermicro ensures a smooth transition after deployment, with full documentation and ongoing support to ensure long-term success. Global Services provide options for on-site response time as short as 4 hours and parts replacement services.

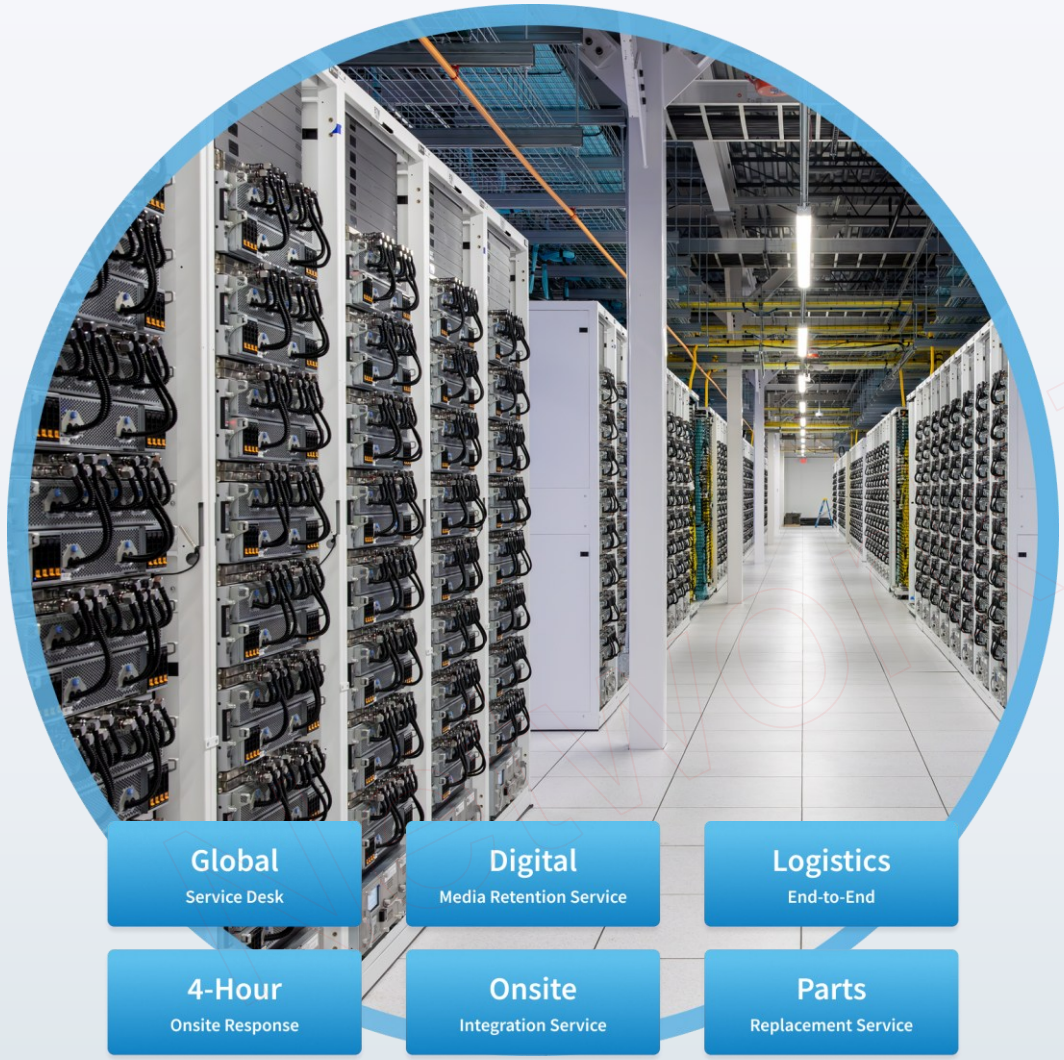
## Data Center Fit-Out Service

Turn-key solution transforming data center shells to operational environments. Includes design, engineering, permitting, electrical from PDUs to servers, mechanical systems with racks and CDUs, and liquid cooling infrastructure for complete readiness.

## Build Service

Buildout services significantly reduce lead time for new data center construction. Supermicro ensures success for bare-land builds, data center retrofitting for liquid-cooling with end-to-end project management.

# Supermicro Global Services




- **Professional Services Portfolio** – Comprehensive service-level building blocks including data center design, solution validation, and professional on-site deployment.
- **Continued On-Site Support** – Ongoing support ensures long-term success and optimal performance throughout the data center lifecycle.
- **Rapid Response Options** – 4-hour on-site response time available for mission-critical uptime and enterprise reliability.

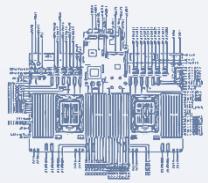
# Data Center Solutions and Services

- **Full-Spectrum Deployment** – Guide customers from bare-land builds to air-to-liquid cooling retrofits or co-location hosting services.
- **Accelerated, High-Quality Installations** – On-site buildout services reduce construction lead time and deliver higher-quality installations with faster time-to-online.

# The Supermicro Advantage

Supermicro specializes in building data center scale IT solutions, optimized for customer business goals at every step of our highly integrated supply chain with USA-based program/manufacturing.

With 30 years of technology leadership, Supermicro meets the needs of the entire data center with a portfolio of application-optimized systems, cluster-level design, and on-site services.



## Architecture

Extensive R&D capabilities to scale up and scale out data center and enterprise clusters.

**Designed on Industry Standards**

**Ease of transition and upgrade with cost saving benefits**

## Solution Design

Building block solution offers gen-to-gen compatible server designs

**High Flexibility and Customization**

**Product fit for wide range of customer use cases**

## Production

5,000 (2,000+ liquid cooled) racks per month worldwide accelerating time-to-market

**Short Lead Time**

**Sooner you get IT working the sooner IT works for you**

## Deployment

L10 to L12 integration from design to validation to deployment with global service

**Pain-free Data Center Plug and Play Deployment**

**Ready-to-go solution with minimal set up required**