# SUPERMICRO RACK SCALE LIQUID COOLING SOLUTIONS

## LIQUID COOLING ADVANTAGES

UP TO

### 89%

### REDUCTION
in Server Cooling Component Electricity Costs

UP TO

### 40%

### REDUCTION
in Electricity Costs for Entire Data Center
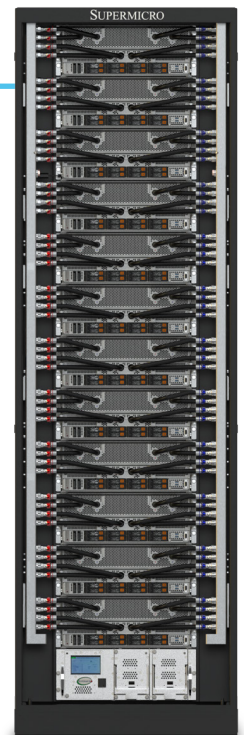
UP TO

### 55%

### REDUCTION
in Datacenter Server Noise

## SUPERMICRO DIRECT TO CHIP LIQUID COOLING SOLUTIONS

Liquid cooled rack solution that delivers superior performance and efficiency for large scale AI and cloud scale compute infrastructure

- Full turn-key single-source solution optimized from proven total solution blueprints of compute, GPU, storage, networking, and power & cooling reference designs, with integrated power management tools

- Support highest densities and highest TDP CPUs & GPUs with up to 100kW power and cooling per rack

- Fully validated and tested at system (L10), rack (L11) and cluster (L12) levels

- Accelerated lead times based on in-stock inventory with deployment in weeks not months

- Enterprise grade redundant cooling pump and power supplies, leak-proof connectors and leak detection

## DATA CENTER SCALE LIQUID COOLING REDUCES COSTS AND INCREASES PERFORMANCE

The latest servers with multiple CPUs and GPUs are creating a significant challenge for data center operators. High-end servers are now generating up to 10kW of heat, which must be removed from the server. Traditional air cooling through CRAC units, even with hot and cold aisle separation, is expensive and inefficient.

Servers that are application-optimized for AI, HPC, and Analytics require the latest in CPU and GPU technologies, which run hotter than previous generations. Multiple CPUs and GPUs per server are needed for performance-intensive computing, driving up the electricity requirements of the server as well as at the rack level.

## RACK SCALE



Rear Door Heat Exchanger



Immersion Cooling



Direct-to-Chip
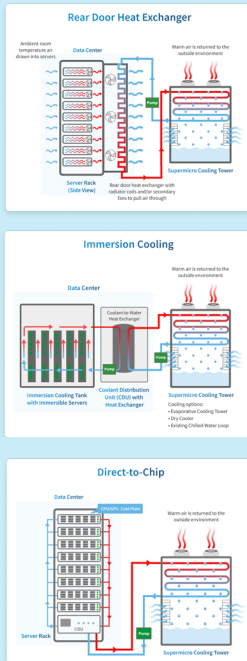
# COMPONENTS OF AN EFFICIENT LIQUID COOLING SOLUTION

Supermicro's liquid cooled rack solution consists of several components that are designed in-house to achieve the highest level of performance and reliability. All the components are integrated as a rack-level solution providing a true one-stop shop customer experience. The critical components of Supermicro's liquid cooled rack solution are:

**Coolant Distribution Unit (CDU)** – Contains the pumping system that circulates the coolant to the cold plates cooling down the CPUs, GPUs, and DIMMs. The Supermicro CDU integrates 2 hot-swappable and redundant pumping modules and power supply modules, guaranteeing nearly a 100% uptime to the operator. The CDU has cooling capacity of up to 100kW or 250kW (depending on CDU configuration) enabling extremely high rack densities. The CDU also offers an easy-to-use touch screen to monitor and control the rack operation, with WebUI access and integrated in Supermicro's Super Cloud



*Supermicro CDU*

Composer® data center management software. The CDU control system optimizes power consumption while ensuring efficient cooling is delivered to all CPUs, GPUs, and DIMMs. An effective anti-condensation strategy is adopted to prevent any hardware degradation.

**Coolant Distribution Manifold (CDM)** – The CDM contains the distribution pipes that supply coolant to each server and return the hot coolant back to the CDU.
There are two types of CDMs:

**Vertical** – Vertical manifolds are placed at the front or back of the rack and directly connected to the CDU with hoses. They deliver coolant to the cold plates on systems with inlet and outlet pipes.

**Horizontal** – Horizontal manifolds are placed at the front of the rack and occupy 1U of rack space each. They connect the vertical manifolds at the rear of the rack to cold plates on systems with inlet and outlet pipes at the chassis of the rack (SuperBlade and 8-GPU servers)



*Horizontal CDM*

**Quick Disconnect Couplings (QDC)** – Flexible QDC are used to bring the cold liquid to the CPU, GPU, and DIMM cold plates and return the hot liquid to the CDMs. Supermicro connectors are single handed, zero drip and quick-disconnect allowing operators to service liquid cooled systems safely and efficiently.
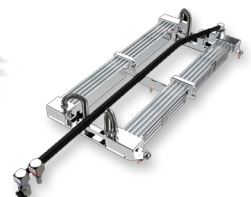
**Cold Plates** – The cold plates are placed on top of the CPUs, GPUs, and DIMMs and feature micro-sized channels through which coolant flows to cool down the chips with high efficiency. Supermicro cold plates are designed to reduce hot spots on the chip and achieve ultra-low thermal resistance.
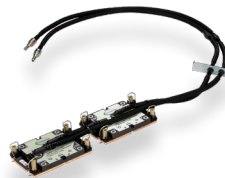


*CPU Cold Plate*



*GPU Cold Plate*



*DIMM Cold Plate*

# RACK SCALE ADVANCED ENGINEERING

As the Rack becomes the new unit for scalable computing, engineering a liquid cooled rack requires careful planning and expert assembly. A liquid cooling solution needs to be flexible enough to handle a wide range of servers, while being able to cool an entire rack of high performance systems.

## RACK SAMPLE CONFIGURATIONS
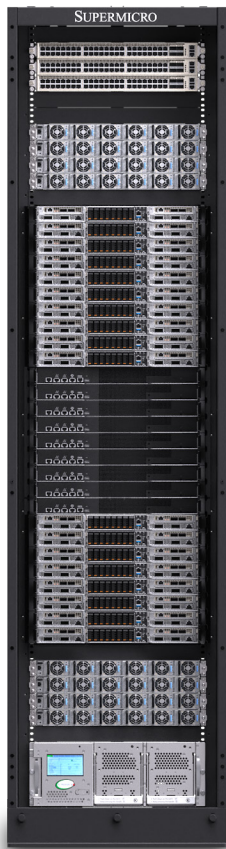
### NVIDIA GB200 NVL72

Up to 18 compute nodes / 72 NVIDIA Blackwell GPUs in a 52U rack

| Product | Qty. |
| --- | --- |
| 1U Liquid-cooled Server with 2 NVIDIA GB200 Grace™ Blackwell Superchips | 18 |
| 33kW Power Shelves | 8 |
| CDU | 1 |
| Vertical CDM | 2 |
| NVLink Switch | 9 |
| In-band & Out-of-band Management Switch | 1 |

### 1U Flagship Rackmount Architecture

Up to 44 servers in a 48U rack

| Product | Qty. |
| --- | --- |
| 1U Hyper Server | 44 |
| CDU | 1 |
| Vertical CDM | 1 |
| Switch | 1 |

### Purpose-Built HPC-at-Scale Solution

Up to 21 systems / 84 server nodes in a 48U rack

| Product | Qty. |
| --- | --- |
| FlexTwin™ Server (2U4N) | 21 |
| CDU | 1 |
| Vertical CDM | 1 |
| Switch | 1 |

## GPU-Optimized

Supermicro GPU-optimized systems provide maximum acceleration for large-scale AI training, large language models, and generative AI applications. The GPU-optimized systems have been completely re-engineered to take advantage of the latest interconnect, memory, storage, and cooling technologies to ensure significant performance increases over prior generations. The modular design includes a dedicated GPU tray which houses an SXM5 (H100/H200) or SXM6 (B200) 8-GPU baseboard and is easily accessible from the cold aisle, simplifying installation and servicing. For maximum GPU density, a liquid-cooled 4U architecture can be integrated using Supermicro's complete direct-to-chip liquid cooling solution, allowing up to 8 systems in a standard 48U rack for a total of 64 GPUs. Supermicro builds upon its proven AI-optimized system architecture with thermal-design improvement to handle the most powerful AI GPUs.

## BigTwin®

Supermicro BigTwin systems provide maximum performance and serviceability in a multi-node architecture, with dual Intel® Xeon® 6700 or 6500 series processors per node and a hot-swappable tool-less design. Optimized for density (2U4N) or storage (2U2N), BigTwin systems with shared components can be up to 8% more power efficient than standard 1U servers, with Supermicro's Resource Saving Architecture of shared power and cooling reducing TCO and TCE. All BigTwin systems can be air cooled, with liquid cooling options available to not only further reduce power consumption and noise levels, but also allow maximum compute density of up to eight 350W TDP CPUs in a 2U chassis.

## FlexTwin™

Supermicro FlexTwin is a new 2U 4-node platform designed for maximum performance density and serviceability in a multi-node architecture, featuring support for the latest CPU, memory, storage, and cooling technologies. Purpose-built to support demanding HPC workloads including financial services, manufacturing, scientific research, and complex modeling, FlexTwin can be customized to suit specific HPC applications and customer requirements thanks to Supermicro's modular Building Block architecture. Each hot-swappable FlexTwin node features direct-to-chip liquid cooling technology which can remove up to 90% of server-generated heat, which not only reduces data center cooling costs compared to traditional air cooling but also ensures maximum compute performance by reducing instances of thermal throttling under maximum load.

## SuperBlade®

Supermicro's GPU-enabled high-performance, density-optimized, and energy-efficient Super-Blade can significantly reduce initial capital and operational expenses for many organizations. SuperBlade utilizes shared, redundant components, including cooling, networking, power and chassis management, to deliver the compute performance of an entire server rack in a much smaller physical footprint. For even greater compute densities, dual-processor and single-width blade configurations can support 350W TDPs with optional direct-to-chip liquid cooling. The use of liquid cooling in data centers not only allows components to run at higher performance levels, but also reduces the need for Computer Room Air Conditioning (CRAC) units and improves overall efficiency, lowering OPEX, TCO and TCE.

## Hyper

The Hyper series brings flagship performance to Supermicro's range of rackmount servers, built to take on the most demanding workloads along with the storage & I/O flexibility that provide a custom fit for a wide range of application needs. With up to 8 PCIe 5.0 expansion slots in a 2U chassis that can accommodate up to 4 double-width GPUs, enterprises have the flexibility to add GPUs, DPUs, and other accelerators to increase workload performance. With Supermicro direct-to-chip liquid cooling solution, Hyper can support dual processors in 1U chassis with CPU TDP up to 500W and 12 hot-swappable NVMe drive bays.

| X14/H14 Direct-to-Chip Liquid Cooling Systems | | |
| --- | --- | --- |
| **Product Family** | **Server** | **Description** |
| **GPU** | SYS-422GA-NBRT-LCC | Dual Intel® Xeon® 6900 Series Processors with P-cores<br>4U, 24 DIMMs<br>NVIDIA HGX™ B200 8-GPU |
| | AS -4126GS-NBR-LCC | Dual AMD EPYC™ 9005/9004 Series Processors<br>4U, 24 DIMMs<br>NVIDIA HGX™ B200 8-GPU |
| | SAS -4126GS-NMR-LCC | Dual AMD EPYC™ 9005/9004 Series Processors<br>4U, 24 DIMMs<br>AMD Instinct™ MI325X Accelerators |
| | ARS-111GL-NHR-LCC | NVIDIA 72-core Grace CPU on GH200 Hopper™ Superchip<br>Onboard Memory and GPU |
| | ARS-111GL-DNHR-LCC | NVIDIA 72-core Grace CPU on GH200 Hopper™ Superchip<br>Onboard Memory and GPU |
| **FlexTwin™** | SYS-222FT-HEA-LCC | Dual Intel® Xeon® 6900 Series Processors with P-cores<br>2U 4-Node, 24 DIMMs |
| | AS -2126FT-HE-LCC | Dual AMD EPYC™ 9005/9004 Series Processors<br>2U 4-Node, 24 DIMMs |
| **Hyper** | SYS-122HA-TN-LCC | Dual Intel® Xeon® 6900 Series Processors with P-cores<br>1U, 24 DIMMs |

| X14 Optional Liquid Cooling Support Systems | | |
| --- | --- | --- |
| **Product Family** | **Server** | **Description** |
| **Hyper** | SYS-222HA-TN | Dual Intel® Xeon® 6900 Series Processors with P-cores<br>4U, 24 DIMMs<br>Support up to 4 double-width GPUs |
| | SYS-122H-TN | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>1U, 32 DIMMs<br>Support up to 3 single-width GPUs |
| | SYS-222H-TN | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U, 32 DIMMs<br>Support up to 4 double-width GPUs |

| Product Family | Server | Description |
|---|---|---|
| **BigTwin®** | SYS-222BT-DNR | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 2-Node, 16 DIMMs |
| | SYS-622BT-DNC8R | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 2-Node, 16 DIMMs |
| | SYS-622BT-HNC8R | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 4-Node, 16 DIMMs |
| | SYS-222BT-HNC8R | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 4-Node, 16 DIMMs |
| | SYS-222BT-HNC9R | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 4-Node, 16 DIMMs |
| | SYS-222BT-HNR | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 4-Node, 16 DIMMs |
| | SYS-222BT-HER | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores, or 6700 Series Processors with E-cores<br>2U 4-Node, 16 DIMMs |
| **SuperBlade®** | SBI-612BA-1NE34 | Single Intel® Xeon® 6900 Series Processor with P-cores<br>With 6U Enclosure<br>12 DIMMs |
| | SBI-422B-1NE14 | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores or E-cores<br>With 8U Enclosure<br>16 DIMMs |
| | SBI-612B-1NE34 | Single Intel® Xeon® 6700/6500 Series Processor with P-cores or E-cores<br>With 6U Enclosure<br>16 DIMMs |
| | SBI-622B-1NE34 | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores or E-cores<br>With 6U Enclosure<br>32 DIMMs |
| | SBI-622B-1NE38 | Dual Intel® Xeon® 6700/6500 Series Processors with P-cores or E-cores<br>With 6U Enclosure<br>32 DIMMs |

| X13/H13 Direct-to-Chip Liquid Cooling Systems | | |
| --- | --- | --- |
| **Product Family** | **Server** | **Description** |
| **GPU** | SYS-421GE-NBRT-LCC | Dual 5th/4th Intel® Xeon® Scalable Processors<br>4U, 32 DIMMs<br>NVIDIA HGX™ B200 8-GPU |
| | SYS -421GE-TNHR2-LCC | Dual 5th/4th Intel® Xeon® Scalable Processors<br>4U, 32 DIMMs<br>NVIDIA HGX™ H200/H100 8-GPU |
| | AS -4125GS-TNHR2-LCC | Dual AMD EPYC™ 9004/9005 Series Processors<br>4U, 24 DIMMs<br>NVIDIA HGX™ H200/H100 8-GPU |
| | AS -4125GS-TNMR2-LCC | Dual AMD EPYC™ 9004/9005 Series Processors<br>4U, 24 DIMMs<br>AMD Instinct™ MI300X Accelerators |
| | SYS-221GE-TNHT-LCC | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U, 32 DIMMs<br>NVIDIA HGX™ H100 8-GPU |
| **APU** | AS -2145GH-TNMR-LCC | Quad AMD Instinct™ MI300A Accelerated Processing Units (APUs)<br>2U, Onboard 512GB HBM3 stacks memory<br>4 Onboard GPUs |
| **GPU Workstation** | SYS-751GE-TNRT | Dual 5th/4th Intel® Xeon® Scalable Processors<br>Rackmount tower, 16 DIMMs<br>Support up to 4 double-width GPUs |

| X13 Optional Liquid Cooling Support Systems | | |
| --- | --- | --- |
| **Product Family** | **Server** | **Description** |
| **GPU** | SYS-421GU-TNXR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>4U, 32 DIMMs<br>NVIDIA HGX™ H200/H100 4-GPU |
| | SYS-521GE-TNRT | Dual 5th/4th Intel® Xeon® Scalable Processors<br>5U, 32 DIMMs<br>Support up to 10 double-width GPUs |
| | SYS-421GE-TNRT | Dual 5th/4th Intel® Xeon® Scalable Processors<br>4U, 32 DIMMs<br>Support up to 10 double-width GPUs (Dual-Root PCIe GPU) |
| | SYS-421GE-TNRT3 | Dual 5th/4th Intel® Xeon® Scalable Processors<br>4U, 32 DIMMs<br>Support up to 8 double-width GPUs (Direct-connect PCIe GPU) |

| Product Family | Server | Description |
|---|---|---|
| **GPU Workstation** | SYS-751GE-TNRT-NV1 | Dual 5th/4th Intel® Xeon® Scalable Processors<br>Rackmount tower, 16 DIMMs<br>Support up to 4 double-width GPUs |
| **Hyper** | SYS-121H-TNR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U, 32 DIMMs<br>Support up to 3 single-width GPUs |
| | SYS-221H-TNR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>1U, 32 DIMMs<br>Support up to 4 double-width GPUs |
| | SYS-221H-TN24R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U, 32 DIMMs<br>Support up to 2 double-width GPUs |
| | SYS-621H-TN12R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U, 32 DIMMs<br>Support up to 4 double-width GPUs |
| **BigTwin®** | SYS-221BT-DNTR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 2-Node, 16 DIMMs |
| | SYS-621BT-DNRT | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 2-Node, 16 DIMMs |
| | SYS-221BT-DNC8R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 2-Node, 16 DIMMs |
| | SYS-621BT-DNC8R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 2-Node, 16 DIMMs |
| | SYS-221BT-HNR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |
| | SYS-221BT-HNTR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |
| | SYS-221BT-HNC8R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |
| | SYS-221BT-HNC9R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |
| | SYS-621BT-HNC8R | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |
| | SYS-621BT-HNTR | Dual 5th/4th Intel® Xeon® Scalable Processors<br>2U 4-Node, 16 DIMMs |

| Product Family | Server | Description |
|---|---|---|
| **FatTwin®** | SYS-521E3-RTB | Single 5th/4th Intel® Xeon® Scalable Processors<br>4U 4-Node, 16 DIMMs |
| | SYS-F511E2-RT | Single 5th/4th Intel® Xeon® Scalable Processors<br>4U 8-Node, 16 DIMMs |
| **SuperBlade®** | SBI-411E-1G | Single 5th/4th Intel® Xeon® Scalable Processors<br>With 8U Enclosure<br>8 DIMMs |
| | SBI-621E-1NE34 | Dual 5th/4th Intel® Xeon® Scalable Processors<br>With 6U Enclosure<br>32 DIMMs |
| | SBI-621E-1NE38 | Dual 5th/4th Intel® Xeon® Scalable Processors<br>With 6U Enclosure<br>32 DIMMs |

# LLC Systems



**HGX™ GPU**
X14/H14/X13/H13



**Instinct™ GPU**
H14/H13



**FlexTwin™**
X14/H14



**Hyper**
X14



**Instinct™ APU**
H13



**GPU Workstation**
X13

# Optional LCC Systems



**HGX GPU**
X13



**PCIe GPU**
SYS-421GU-TNXR



**SuperBlade**
X14/X13



**Hyper**
X14/X13



**BigTwin**
X14/13



**FatTwin**
X13

## Liquid Cooling Tower

The Supermicro Liquid Cooling Tower solution offers a versatile and energy efficient method of removing the heat produced by today's latest servers. The modular design is available in different sizes with multiple cell configurations up to a maximum of 10MW per unit and scalable up to multiple cell configurations from 5MW to hundreds of MW. Each cell can be used separately, enabling cooling tower redundancy.

## Supermicro SuperCloud Composer® for Liquid Cooled Data Centers

SuperCloud Composer's LCCM (Liquid Cooling Consult Module) is a powerful tool to collect vital information on physical assets and sensor data from a CDU (Cooling Distribution Unit), including pressure, humidity, pump and valve status, and more. CDU data is presented in real-time, enabling users to monitor operating efficiency of their liquid cooled racks. Using these insights, SuperCloud Composer (SCC) helps the user to set up alerts, manage firmware updates, and more.
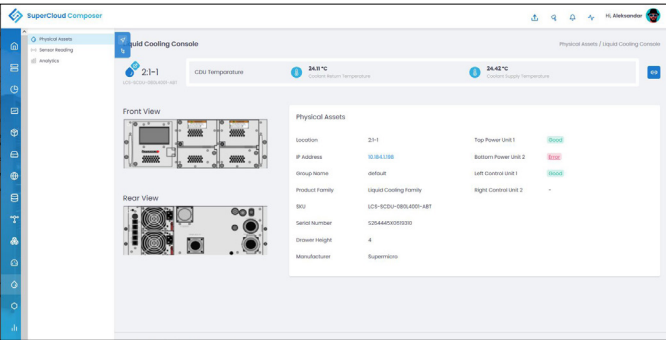
## Supermicro's Rack Integration Services, Turnkey Cluster Level Liquid Solutions

Supermicro's Rack Integration Services leverage application-optimized motherboards, chassis, cooling subsystems, networking components, cluster management tools, energy-efficient power supply technologies, and compact enclosures to design and develop customized and enterprise solutions. Supermicro understands the requirements of today's fast-paced business environments and customer requirements. We offer an end-to-end integration service that helps customers reduce overhead, maximize efficiency and quality, and gain a competitive advantage through reduced time-to-results.

Supermicro works with leading organizations in all geographies to design, install, and test various liquid cooling solutions. The Supermicro process involves a rigorous set of phases that ensure the most optimized and tested solutions for environments where liquid cooling is required for maximum performance.
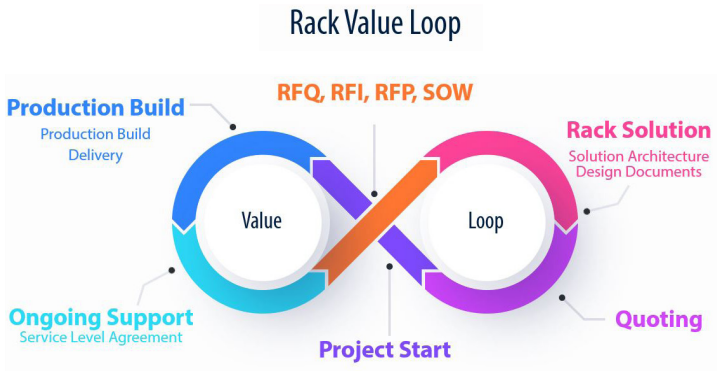
## ABOUT SUPER MICRO COMPUTER, INC.

Supermicro® (NASDAQ: SMCI), the leading innovator in high-performance, high-efficiency server technology is a premier provider of advanced server Building Block Solutions®  for Data Center, Cloud Computing, Enterprise IT, Hadoop/Big Data, HPC and Embedded Systems worldwide. Supermicro is committed to protecting the environment through its "We Keep IT Green®" initiative and provides customers with the most energy-efficient, environmentally-friendly solutions available on the market.

**www.supermicro.com/liquidcooling**