WHITE PAPER

DIRECT LIQUID COOLING: A GAME CHANGER FOR DATA CENTER POWER

Executive Summary

The revolutionary benefits of Generative AI are driving an AI infrastructure buildout race for accelerated computing as never seen before. Cloud Service Providers, Enterprise customers, and sovereign countries are actively planning and deploying AI factories to produce intelligence by harnessing their valuable data. However, the high-performance GPUs underpinning these AI factories consume tremendous power, constraining where and how fast these AI factories can be built.

Geographic and Utility Considerations

Most of the new upside demand is only achievable in greenfield data center builds. However, locations with available power, land, fiber, and water at a reasonable price are challenging and competitive. Utilities indicate 2-3 year wait times to access large data center power requirements. Additionally, regional power constraints and government restrictions limit builds and capacities in areas such as Japan, Singapore, Taiwan, Netherlands, Ireland, and many cities in the US. As a result, new builds are targeting more rural areas that are adequate for AI training which is less sensitive to terrestrial latency. Favorable areas emerging are Hokkaido, Japan, south of Taoyuan in Taiwan, and the U.S. Midwest (North Dakota, Missouri, Indiana).

Direct Liquid Cooling Benefits …up to 40% reduction in power

For customers struggling to deploy AI infrastructure due to lack of power, Supermicro Direct Liquid Cooling (DLC) is the perfect solution, offering many benefits:

1) Helps customers save up to 40% of the electric utility demand (and bill), which is crucial in geographies that are power-limited.
2) Boosts data center computing performance by providing superior chip cooling, allowing chips to run at full performance without throttling.
3) Shortens customers' AI factory deployment time because less electrical power is required from the grid – significant power from the utility can take years.
4) Reduces carbon emissions due to lower power consumption, true to Supermicro's green computing mission, saving our one-and-only Mother Earth.

Data Center DLC Example: …15MW+ reduced to less than 10MW

Example DLC power savings impact: An 8192 NVIDIA HGX H100 GPU cluster in a traditional air-cooled environment requires a 15MW+ data center. Less than 10MW is required in an optimized DLC-cooled environment. This technology is a game changer in areas where the utility allocation is limited to 10MW.

Additionally, power and cooling costs drive the highest percentage of data center build CAPEX, so minimizing the IT power demand keeps more money in your pocket.

"DLC can be free with a big bonus!" Charles Liang, CEO of Supermicro.

As the pioneer of optimized DLC solutions, Supermicro is your one-stop shop, designing and manufacturing an end-to-end data center liquid cooling total solution spanning cold plates, cooling distribution manifolds, hoses, cooling distribution units, and the cooling tower, all managed and monitored by our data center level SuperCloud Composer software. Supermicro controls the supply chain and quality, enabling a 2-4 week lead time with a given forecast. Additionally, Supermicro's onsite deployment services optimize the customer's time to online (TTO).

Supermicro's DLC total solution is mature and deployed at major customer sites worldwide. One recent example is xAI's Memphis SuperCluster, which consists of thousands of Supermicro's DLC liquid-cooled racks and 100K NVIDIA HGX H100 GPUs, the world's most powerful AI training cluster. From start to finish, the onsite deployment took 19 days, a testament to the maturity and simplicity of Supermicro's DLC solution and onsite services.

"We were able to install and bring online a massive new training center in nineteen days. That's the fastest by far that anyone's been able to do that."
Elon Musk on Podcast with Dr. Jordan B Peterson 7/22/2024

Supermicro Direct Liquid Cooling Capabilities

Supermicro has been aggressively expanding its liquid cooling rack integration capacity in the US and Asia to drive industry-wide adoption and make large scale DLC deployments available to customers worldwide. Supermicro's capacity to ship 1,000 fully integrated DLC racks/month (since June 2024) is expanding to deliver 2,500 racks/month by the end of the year. Supermicro is driving liquid-cooled data centers from <1% market share a few months ago to an expected 30% market share in the next 12 months, with most of those deployments coming from Supermicro.

The more DLC infrastructure you deploy, the more energy is saved. Supermicro's optimized solution is the future of green AI factories. And the future is today. Join Supermicro in this green AI revolution!

Summary/For More Information

The reduced power demand decreases the overall cost of building a new data center. An increased number of high-performance AI servers within the same power budget as a traditional air-cooled data center is very real. Supermicro brings Direct Liquid Cooled rack-scale infrastructure to your data center's doorstep.

- https://money.udn.com/money/story/11162/8157744
- https://www.thestack.technology/xai-elon-musk-ai-supercomputer/
- Global Data Center Trends 2024 | CBRE
- Visit the Liquid Cooling web page
- Read the Liquid Cooling Whitepaper
- Read the Liquid Cooling Brochure and the Liquid Cooling Datasheet