

Article Overview

Liquid cooling is now essential for high-density AI servers, enabling extreme computational performance while managing heat and power efficiently.

Why Liquid Cooling is Critical for AI Servers

Modern AI workloads, especially those using GPUs like NVIDIA Blackwell GB200/GB300 or high-power AI ASICs, generate extreme heat due to high thermal design power (TDP) exceeding 1,000W per accelerator . Traditional air cooling cannot handle these densities, making direct-to-chip (DLC) and immersion liquid cooling the baseline for AI data centers . Liquid cooling improves energy efficiency by up to 15%, reduces cooling costs, and allows higher server density while extending hardware lifespan .

Types of Liquid Cooling

- o Cold Plate Liquid Cooling (Direct-to-Chip): Coolant circulates through a copper micro-channel block mounted on CPUs or GPUs. Heat is transferred to a secondary loop connected to cooling distribution units (CDUs) and then rejected via heat exchangers or cooling towers . This method preserves standard server form factors and is widely deployed, accounting for roughly 65% of liquid cooling systems in 2026 .
- o Immersion Cooling: Entire servers are submerged in dielectric fluids, eliminating thermal resistance from heat spreaders and airflow. Single-phase immersion keeps the fluid liquid throughout operation, while two-phase immersion offers maximum thermal efficiency but is more complex . Immersion cooling achieves PUE values as low as 1.02-1.05 under optimized conditions.
- o Hybrid Systems: Combine liquid cooling for high-TDP accelerators with air cooling for lower-power components like NICs and SSDs, allowing incremental upgrades in existing data centers .

Infrastructure and Power Considerations

High-density AI servers demand resilient power systems. Rack densities now range from 40 kW to 140 kW, requiring robust UPS, PDUs, and electrical distribution to prevent hotspots or throttling . Integrated design from grid to chip and chip to chiller ensures synchronized power and thermal management, critical for maintaining performance and reliability at scale .

Sustainability and Resource Use

Liquid cooling enables heat reuse and supports energy-efficient operations, but it also increases water consumption. AI data centers can consume millions of gallons of water daily for cooling, highlighting the need for sustainable water management and renewable energy integration . Hybrid cooling systems can help balance efficiency with environmental impact.

AI computing power plus liquid-cooled servers

Operational Benefits

- o Higher server density and reduced space requirements
- o Improved energy efficiency and lower operational costs
- o Extended hardware lifespan due to stable thermal conditions
- o Scalable deployment with integrated monitoring and lifecycle support

Conclusion

Liquid-cooled servers are no longer optional for AI workloads--they are fundamental to achieving high computational performance, energy efficiency, and operational reliability. Direct-to-chip and immersion cooling, combined with robust power infrastructure and integrated design, allow AI data centers to meet the growing demands of generative AI and high-performance computing while addressing sustainability challenges .

High-density computing workloads like AI training and inference run too hot for traditional

Introduction AI training and inference servers use accelerators and processors with high thermal design power (TDP)¹. Air-cooling

Discover six key benefits of Direct Liquid Cooling for modern data centers supporting more

Liquid cooling, however, can maintain more compute power in traditional server form factors because it's so efficient at absorbing

By reducing dependence on chillers and enabling more efficient heat rejection, liquid cooling is driving the next

That's where liquid cooling comes in. Unlike conventional air-based systems, liquid cooling

Liquid cooling for AI data centers is a complex architecture now needed to address the increased AI data center heat

AI Computing Power Surging, "Side Car" as Transition Edward Kung, president of Taiwan Thermal Management

As power-hungry AI and HPC workloads rise, traditional air cooling falls short. Discover why hyperscale data

AI computing power plus liquid-cooled servers

centers

Although each generation offers significantly more computing power for each watt, full liquid-cooled AI compute infrastructure

The portfolio is based on the industry's first 100% fanless direct liquid cooling system architecture and spans every

Single-phase liquid cooling enables high-density server cooling, meeting the demands of AI and next-generation computing

Liquid Cooling Steps Up for High-Density Racks and AI Workloads Rack densities, AI workloads and energy costs keep climbing. Air

As rack densities surge and grid headroom tightens, liquid cooling is becoming the backbone of AI data centers. Here we unpack the

Pushing the bounds of computational power comes at a cost: increasing power consumption. The liquid-cooled servers needed to

Explore air vs. liquid cooling for AI data centers. Learn how hybrid cooling strategies support high-density racks and

Web: <https://kremgroup.co.za>

