Advanced Cooling and Power Solutions for High-Density AI Data Centers

By David Pike for Mouser Electronics

MOUSER ELECTRONICS | Amphenol

Artificial intelligence (AI) has changed the physical reality of the data center. Training clusters and high-performance inference systems pack a significant number of graphics processing unit (GPU) cards and accelerators into compact rack spaces. Along with this, increasing chipset performance has raised electrical demand, and nearly all of that energy ultimately becomes heat that must be managed. This management is critical, as stable operating conditions determine whether data centers deliver sustained performance or encounter throttling, downtime, or premature failure.

Managing this thermal load, however, is becoming more expensive, as cooling infrastructure and power distribution networks account for a substantial share of both capital expense and operating costs. A large portion of facility energy is spent on cooling rather than IT systems. In fact, cooling infrastructure can consume 30 to 40 percent of the total energy budget, with less efficient facilities exceeding that range.[1]

Rack power density is moving from tens of kilowatts toward several hundred kilowatts, with some deployments approaching 1MW. Rack performance is now constrained as much by environmental control as by computing power. With heat and power management becoming part of data center system design, rather than just a facilities concern, this white paper examines engineered thermal, sensing, and connector solutions for high-density compute systems.

Designing for Efficiency Within Recognized Standards

A data center's efficiency directly affects operations. Equipment must remain within temperature and humidity envelopes specified by manufacturers, while total facility energy consumption remains controlled. Industry standards define the measurable efficiency targets.

The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Technical Committee 9.9 (TC 9.9) defines the recommended environmental ranges for data processing equipment and includes guidance on sensor deployment.[2] There is also the ASHRAE Standard 90.4, which addresses the role of energy efficiency in data centers and supports the use of continuous monitoring tied to control systems.[3] In addition to international standards, regional regulations add further requirements. For example, California's Title 24 demands the use of monitoring systems for heating, ventilation, and air conditioning (HVAC) and lighting in certain jurisdictions. Broader frameworks, such as EN 50600, ANSI/TIA-942, and ISO 14001, offer guidance on infrastructure, operations, and environmental management to build reliable and sustainable data centers.[4]

Accurate measurements are a crucial part of each of these standards. High-density AI systems create heat that moves with the workload, so fixed assumptions about room temperature and airflow are unreliable. Meeting efficiency targets requires more than just knowing the conditions—it requires systems that can respond to them.

HVAC Feedback Loops and Precision Monitoring

Cooling performance depends on feedback. Sensors report temperature and humidity, while control systems adjust the flow of air, chilled water, or cooling liquid in response to this feedback. Without continuous data, cooling systems either overcompensate and waste energy or underperform, exposing equipment to thermal stress.

Room-level measurements alone cannot capture the full picture. Within a single rack, vertical gradients can develop. The air entering at the front may be within the expected temperature range, but exhaust air at the rear could exceed recommended limits. Additionally, restricted airflow areas can trap heat around power supplies or accelerator modules, while raised floor cavities may conceal localized hot zones.

For monitoring to be effective, sensors must be distributed throughout the installation. Engineers should position temperature sensors at different heights within each rack, near high-power components, and in regions where airflow paths narrow. Dispersing feedback capture in this way ensures monitoring extends beyond room-level measurements to include the equipment itself.

That kind of distributed monitoring depends on sensors that can be placed close to the equipment without getting in the way. **Amphenol Advanced Sensors JI and JIC** waterproof IP68 negative temperature coefficient (NTC) thermistors support this model. They have a compact format that allows for installation in confined spaces, such as server enclosures, switchgear compartments, and cooling assemblies. Meanwhile, the IP68 sealing protects against condensation or moisture. When connected to building management or rack-level controllers, these thermistors provide stable measurements that strengthen closed-loop HVAC control.

The Transition to Liquid Cooling

Conventional computing relies on airflow, heatsinks, and fans for cooling. But higher rack densities have forced engineers to revisit these approaches. Air cooling is effective at modest power levels, but as power densities increase, its limitations

become evident. Air has a low specific heat capacity, so removing higher thermal loads requires higher airflow velocity; however, higher fan speeds increase energy consumption and acoustic noise. On top of this, the tightly packed arrangements of AI servers further reduce efficiency by restricting flow paths.

The term liquid cooling may evoke images of equipment completely submerged in fluid, but this is only one type of liquid cooling, known as immersion cooling, and it supports extremely high densities. With the right preparation, servers can operate fully submerged in a dielectric liquid that transfers heat directly to an external heat exchanger.

But immersion cooling requires infrastructure planning. Facilities must be built to support fluid containment, handling, and long-term maintenance. Retrofitting existing installations can be complex. For greenfield sites designed around immersion from the outset, this liquid cooling method can support long-term thermal objectives. However, for many operators, incremental alternatives offer a more practical starting point.

Liquid cooling does not require the submersion of hardware. Rear door heat exchangers represent a more easily adopted transition. Installed on the back of a rack, they capture hot exhaust air before it disperses into the server room. A liquid loop within the door extracts heat via an integrated exchanger, reducing the need for room-level cooling without redesigning internal server layouts. An advantage of this approach is that operators can deploy it selectively on high-density racks.

Direct-to-chip cooling is an even more focused liquid cooling approach. Cold plates mount directly onto processors or accelerators, and coolant flows through internal channels within the plate to remove heat at its source (**Figure 1**). This technique reduces dependence on large heatsinks and high-speed fans. Moreover, the space created within the server becomes available for additional compute modules or improved airflow management for the remaining components.
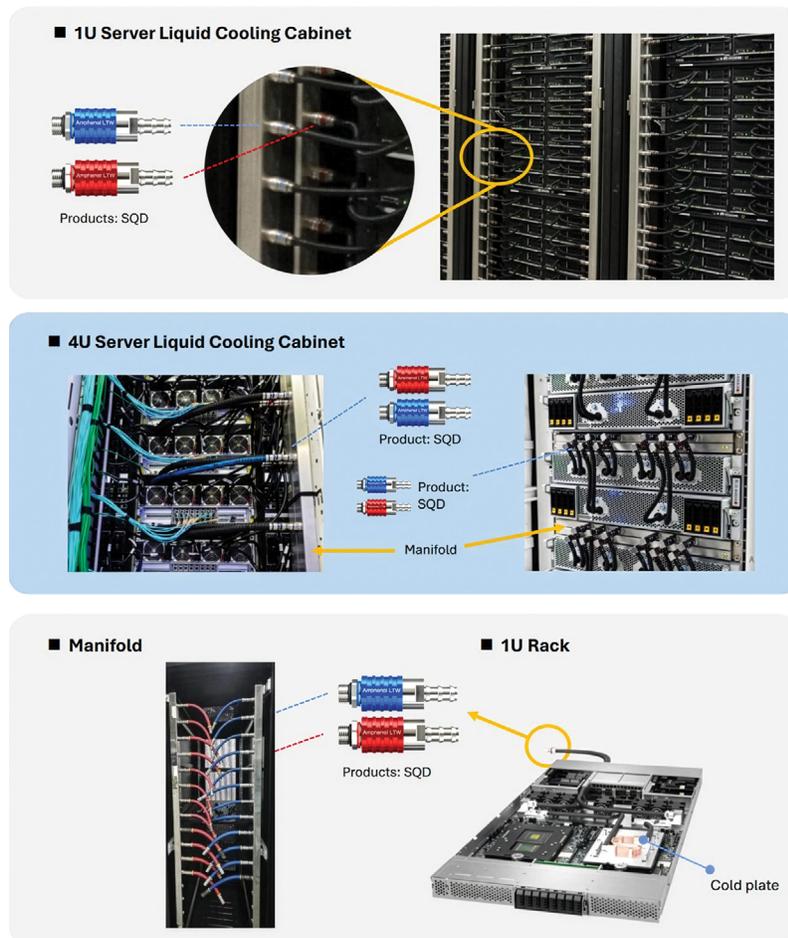

Figure 1: *Direct-to-chip liquid cooling architecture showing rack-level manifolds supplying cold plates within dense server configurations. (Source: Amphenol LTW)*

Liquid cooling benefits from liquids' higher thermal conductivity and heat capacity compared with air. Liquids transport energy more efficiently within a confined space. Further, capturing heat at the device interface prevents it from accumulating within the chassis.

Taking advantage of liquid cooling's benefits calls for new solutions to critical design considerations. Liquid cooling methods require different connectors than air cooling. Tubing and couplings must maintain sealing integrity under pressure, and connectors must function reliably within compact spaces already populated by high-speed signal paths and power distribution hardware.

Amphenol LTW SnapQD liquid cooling connectors address the needs of cold plate systems. Their compact design supports dense assemblies, while quick-disconnect functionality simplifies maintenance and component replacement. Also, their secure sealing mechanisms help maintain leak-free operation under sustained load.

As liquid distribution extends beyond individual cold plates, fluid routing becomes another important design consideration. Engineers must route fluid lines through tight spaces without blocking nearby components. To help achieve this, connector orientation can simplify the overall layout.

Amphenol Industrial UQD and UQDB liquid cooling connectors offer a range of sizes and geometries, including right-angle formats suited to confined rack environments. With durable materials and secure coupling mechanisms, these connectors support continuous operation in high-density installations.

Most facilities adopt a mixed cooling strategy. While air cooling works well for low-power components and for maintaining stable room conditions, liquid cooling is used where the heat is most concentrated, such as in GPUs. This approach makes it easier to gradually scale as compute density increases.

Power Density and Electrical Distribution

Thermal management is directly linked to power architecture. High-performance accelerators routinely consume several hundred watts per device, and a rack populated with accelerators, storage devices, and networking hardware can approach several hundred kilowatts. Industry roadmaps and hyperscale deployments now show rack power densities approaching 1MW for AI-focused installations.[5]

Electricity enters the facility at high voltage. It then passes through uninterruptible power supplies and distribution units before reaching individual servers. Each interface along that path contributes resistance. According to $P = I^2R$, energy loss increases with the square of current. At elevated current levels, small increases in resistance translate into greater energy loss and measurable heat, making connector choice all the more important.

Connector design affects both power efficiency and thermal load. In today's data centers, low contact resistance must be maintained over many mating cycles. Compact footprints are necessary in 1U and 2U power shelves where space remains limited.

Amphenol Positronic OCP ORV3 universal AC input connectors conform to Open Compute Project (OCP) Open Rack V3 (ORV3) specifications for components used in rack system deployments. These connectors' slim profile makes them suitable for rack-mounted power systems. In addition, high-conductivity alloys reduce terminal resistance, while recessed contacts enhance safety in high-current environments.

Converging Power and High-Speed Interfaces

In all data center installations, the pressure to manage board space grows as system density increases. Combining power delivery and high-speed signaling into shared connector interfaces reduces routing complexity and preserves printed circuit board (PCB) area. But combining power and signals in the same connector housing poses risks for signal integrity.

PCIe Gen 6 and related interconnect technologies employ advanced modulation schemes, such as pulse amplitude modulation with four levels (PAM4), to increase throughput without doubling channel frequency. These techniques demand careful control of impedance, insertion loss, and crosstalk.

Amphenol FCI Hyper Cool Edge connectors support hot-pluggable connectivity for accelerator, Compute Express Link (CXL), and network interface modules while accommodating modern high-speed signaling. They can deliver up to 200W of power, and the dedicated two-pin power receptacle can supply an additional 400W. Integrating power and data within a shared connector supports higher board-level density and simplifies system integration.

Conclusion

AI-driven workloads pack more computing power into smaller physical footprints. As rack density increases, heat generation rises, and efficiency becomes harder to maintain, further intertwining thermal and power limits. Decisions in one area directly affect the others.

While standards provide structure, performance depends on implementation. Distributed temperature sensing strengthens HVAC feedback, but what remains essential is the system response to

that feedback. Liquid cooling restores thermal headroom where air cooling approaches its limits, and low-resistance power connectors and integrated signal interfaces help manage electrical density.

Computing capability can scale only as much as the supporting architecture will allow. As rack targets move toward higher power levels, disciplined environmental control and efficient power delivery remain necessary for reliable operation and sustainable growth.

Sources

1.   https://sustainabilitymag.com/news/infosys-how-much-energy-are-data-centres-really-consuming

2.   https://www.ashrae.org/file%20library/technical%20resources/bookstore/ashrae_tc0909_power_white_paper_22_june_2016_revised.pdf

3.   https://www.ashrae.org/file%20library/about/government%20affairs/advocacy%20toolkit/virtual%20packet/standard-90.4-2022-fact-sheet.pdf

4.   https://blog.amphenol-sensors.com/industrial-blog/data-center-design-standards

5.   https://cloud.google.com/blog/topics/systems/enabling-1-mw-it-racks-and-liquid-cooling-at-ocp-emea-summit