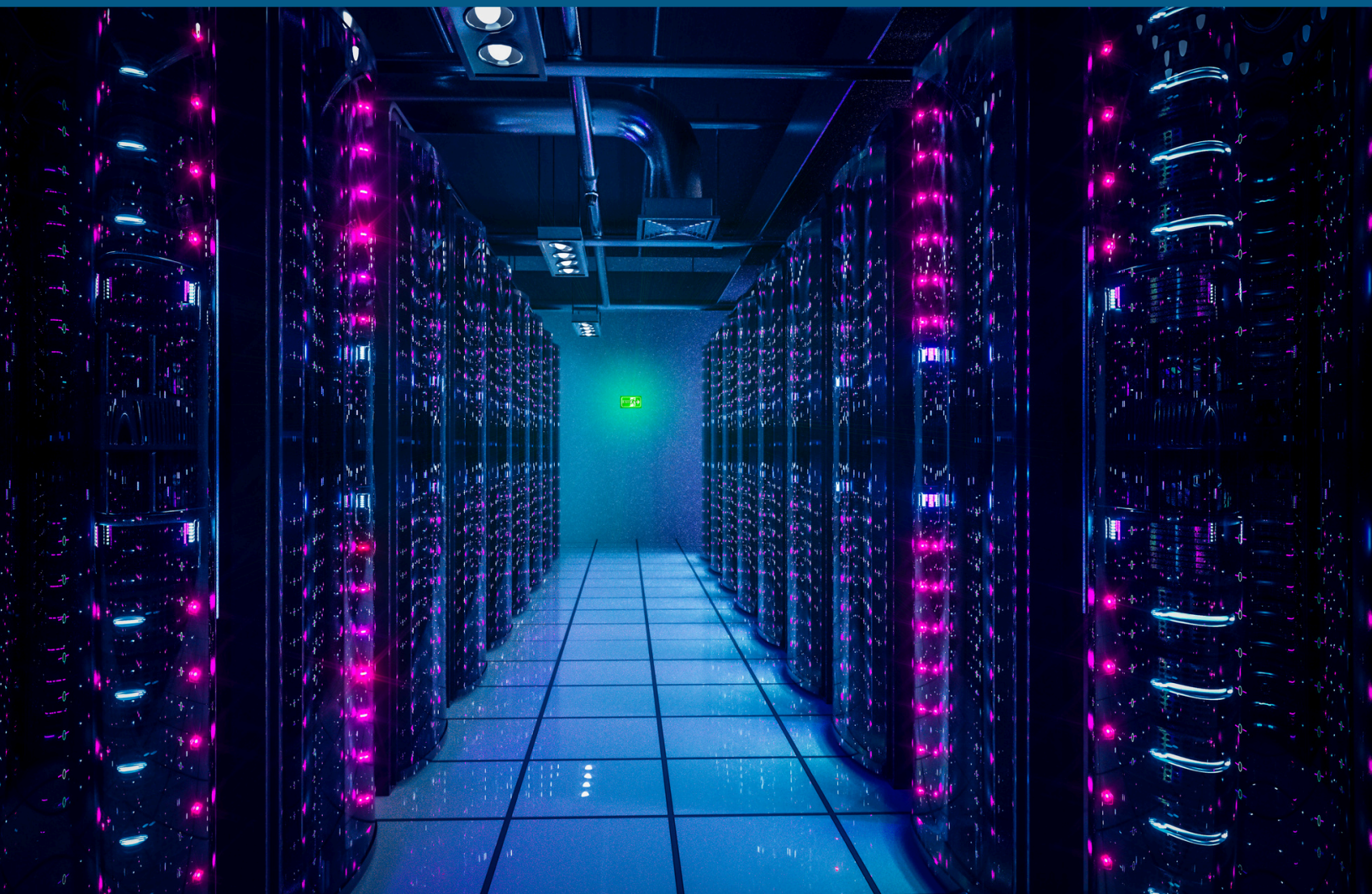




HOW AI WORKLOADS ARE DRIVING COOLING INNOVATION TO NEW HEIGHTS

BY: DREW ROBB

How AI Workloads Are Driving Cooling Innovation to New Heights

“Cooling vendors once had years before the next version of their equipment came out. Over the last couple of years, the timeline has shrunk due to the demands of AI and GPUs.

“AI is giving us a much faster timeline so we have entered a period of accelerated innovation,” said Srdan Mutabdzija, Liquid Cooling Product Manager at [Vertiv](#). “Engineers can figure anything out, but we need to solve things faster.”

What will that mean for chillers? Some may prefer to have several stages of heat exchange inside the data center while others will deploy large chillers outside the facility.

“We will see major advanced in heat rejection over the next few years,” said Mutabdzija. “We need to make it easy for data centers to install liquid cooling in their facilities.”

But the advantages are clear. He noted that only 1% of power is used in liquid cooling.

The Cooling Repercussions of the AI Wave

[Omdia's](#) Principal Analyst Shen Wang laid out the cooling repercussions of the AI wave. Air cooling hit its limit around 2022, he said. The consensus is that it can deliver up to 80 Watts per cm².

Beyond that range, single-phase direct-to-chip (DtC) cooling — in which water or a fluid is taken to cold plates that sit directly on top of computer chips to remove heat — is needed. Single-phase DtC can go as high as 140 W/cm².



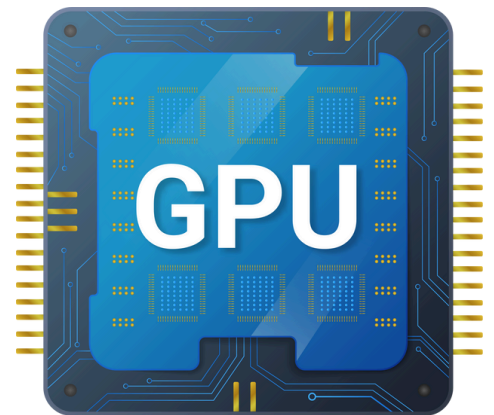
Shen Wang, Principal Analyst, Omdia

“Single-phase DtC is the best way to cool chips right now,” Wang said. “By 2026, the threshold for single-phase DtC cooling will be exceeded by the latest racks.”

That’s when two-phase liquid cooling and immersion cooling should begin to see a ramp-up in adoption rates. Two-phase cooling runs fluids at higher temperatures to the chip, causing them to turn to vapor as part of the cooling process, thereby increasing cooling efficiency.

GPUs Are Hot

AI is driving liquid cooling innovation and deployment. In turn, AI has ramped up the rate of adoption of GPUs. Yet CPUs dominate overall chip sales by a large margin and will continue to do so for many years to come. Wang cited Omdia figures from the end of 2024. More than 20 million chips in the 0- to 199-watt range were sold that year. About 15 million more were sold in the 200- to 399W range. In comparison, there were only a few million each for chips from 400- to-599w and from 600w+.



“The bulk of chips shipped are still CPUs by far and that will continue to be the case through 2028 at least, but number of GPUs will rise dramatically,” said Wang.

It is the higher-powered chips that drive the cooling market. Those below 200w have no need for liquid cooling. Chips between 200 and 399w have less than 5% adoption of liquid cooling and this number is not expected to change as a percentage through 2028. The higher you go in wattage, the stronger is deployment of various forms of liquid cooling.

“Advanced chips in the 600 watt and above range are seeing the heaviest adoption of liquid cooling,” Wang said. “By 2028, 80% of chips in that category will utilize liquid cooling, up from 50% today.”

Omdia believes DtC will remain the liquid cooling technology of choice at least over the short term. Out of around \$2 billion a year in liquid cooling revenue in 2024, more than \$1.5 billion was for DtC. By 2028, the total market size for liquid will rise to \$6 billion with almost \$5 billion going to DtC. By then, immersion will have grown to \$1 billion per year.

Immersion cooling is still evolving as a technology and will take some time to become established,” said Wang.

How To Deal with Cooling System Upgrades

Facilities planning to introduce GPUs and denser racks to operate AI applications should carefully plan how to introduce liquid. Luca Cutrone, Director of Business Development at CoolIT, suggested beginning with rear door heat exchangers (RDHX) as the simplest way to bring in liquid. RDHXs can be put into one or more racks without requiring much reconfiguration of the data center.

Next up, he said, try out liquid-to-air Coolant Distribution Units (CDUs) as they can be installed more easily as they do not require water pipes to be connected to the building system for heat rejection. They are a good way to bring localized liquid cooling to high-output IT equipment.

Cutrone cautioned data center managers to let the hyperscalers and big colocation providers operate on the cutting edge of liquid cooling. They are the ones that can afford the high price tag of innovation and are willing to make mistakes as they search for an optimum design. Their efforts will bleed down to the broader market as mature designs.

“A few of things we are going to see are chillers getting bigger and operating at higher temperatures, as well as heat reuse being integrated into total system design,” said Cutrone.

Immersion Cooling Complexity

How about immersion cooling? Phil Lawson-Shanks, Chief Innovation Officer at [Aligned Data Centers](#), believes immersion cooling is far from ready for broad deployment. He noted that it takes a long time to put in the plumbing lines for immersion. If a technician then has to change a blade, it becomes a maintenance headache.

“Vendors need to develop modular systems that make it simple to change out components due to failures and as the market evolves into the next generation of GPUs and servers,” said Lawson-Shanks.

Each segment of the data center market has different needs. Hyperscalers are on the cutting edge and will be the early adopters of advanced liquid cooling systems such as immersion cooling. Colos, on the other hand, want to utilize whatever generates market value while lowering risk. Enterprise data centers focus more on availability and see new technology through that lens. But liquid cooling is coming – and fast. Its heat rejection potential is so much higher than air that there is no question that it will grow in the data center.



Air Forever

Meanwhile, air cooling is not going away anytime soon. Mutabdzija advised data center managers not to forget about air cooling. Future data centers will need plenty of liquid and air cooling.

Lawson-Shanks concurs. “Single-phase DtC will dominate the liquid market for now, but it will be a long time before we remove air entirely from the data center,” he said.

Case in point: [Equinix](#) recently completed an air-cooling project on 70 kW racks without the need for liquid. Thus, air cooling innovation continues in parallel with liquid cooling, pushing the bounds of air cooling even higher.

“Air cooling is still very much the standard,” she said Ana Marie Ordonez, Principal Solutions Architect, Equinix. “We build large, highly dense data centers with liquid and air cooling.”

Power constraints are being given plenty of attention in the news. But she believes cooling is a bigger concern.

“It is never about how much power we can deliver to the rack but how can we cool it,” said Ordonez.



Looking for more information, check out the [Data Center World Knowledge Hub](#)