Publication date:
21 Apr 2021
Author:
Vladimir Galabov, Head of Cloud and Data Center Research Practice
Greg Cline, Principal Analyst, Data Center Compute

Immersion cooling is heating up

Table of Contents :

Table of Figures :

Omdia view

Summary

Liquid cooling was a key discussion topic in Omdia's annual Data Center (DC) Thermal Management report, published in December 2020. We forecast the market for liquid cooling products to roughly double over the next five years, growing 16% per year, on average. We include direct to chip (D2C) and immersion cooling methods in the liquid category. Of the two, immersion cooling is the less developed and fastest growing segment.

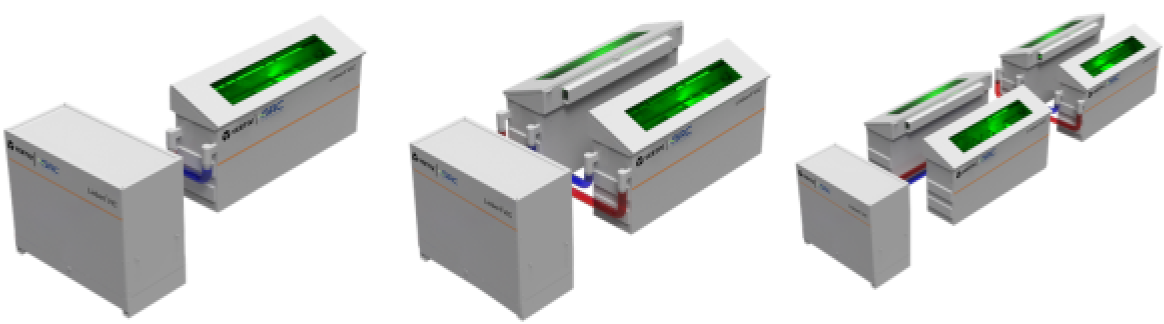
It is becoming increasingly likely that we see immersion cooling move into larger scale commercial deployments. We will also see more vendors entering the market and large end users such as hyperscale cloud service providers (CSPs) looking to deploy immersion cooling to decrease DC power consumption.

Immersion cooling: The basics

Immersion cooling can be implemented in a single-phase or two-phase solution. Single-phase (1P) immersion cooling encapsulates the server in a sealed chassis, and the system can be configured in a rackmount or standalone format. Two-phase (2P) immersion cooling places the server in the liquid, but the liquid changes state during the cooling process. As the fluid heats up and turns to condensation, the water circuit and heat exchanger remove the heat.

In 2P immersion cooling, dielectric cooling fluid inside a steel holding tank packed with servers is boiled and condensed in a closed cooling system, exponentially increasing heat transfer efficiency. One liquid cooling vendor whose solution we have been able to familiarize ourselves with is Wiwynn. The fluid inside its immersion tank is harmless to electronic equipment and engineered to boil at 122 degrees Fahrenheit, 90 degrees lower than the boiling point of water. The boiling effect, which is generated by the work the servers are doing, carries heat away from computer processors. The low-temperature boiling point enables the servers to operate continuously at full power without risk of failure due to overheating.

1. Figure 1: Example immersion cooling configurations



Source: Vertiv

Past hesitancy gives way to an increased sense of urgency

Historically, DC operators have been reluctant to adopt liquid cooling unless it was deemed absolutely necessary. So, what has changed?

Global demand for compute is showing no signs of slowing down. The proliferation of artificial intelligence (AI) techniques, complex analysis and simulation models, and cryptocurrency mining has created a market for servers configured with specialized co-processors such as GPUs. This has accelerated power consumption and computing density in the DC, increasing the need for advanced cooling technologies. Then there's global warming. The realization that real actions are needed from governments, corporations, and individuals to tackle the climate emergency is fueling not just sustainability commitments but also real actions.

Earlier this month, Microsoft revealed that it is testing Wiwynn's 2P immersion cooling technology, described above, in a production environment at its DC in Quincy, Washington. In prior testing, the CSP found that using liquid cooling reduced the power consumption of any given server between 5% and 15%. Microsoft highlighted two main reasons for expanding its investment in liquid cooling:

- **It is it the only way to offset the slowdown of Moore's Law**. Transistor width has already shrunk to the atomic scale and will soon reach a physical limit. Thus, to satisfy performance demand, DC processors such as CPUs and GPUs will get bigger and bigger, increasing the power requirements. Microsoft noted that the power requirements of the CPUs they use have already increased from 150 watts to 300 watts.
- **It is an effective way to cope with upticks in demand**. Microsoft runs a software-defined DC and foresees a future where management software allocates sudden spikes in compute demand to the servers in liquid cooled tanks, taking advantage of the ability to overclock CPUs. In essence, this would allow the CSP to run its servers "hotter" when demand spikes, and "cooler" when demand is low.

Something not highlighted by Microsoft but worth noting is that by reducing the energy consumption of the IT equipment, liquid cooling reduced the DC operator's electricity bill. This is a big expense for hyperscale CSPs.

2. Figure 2: The immersion cooling deployed by Microsoft

Source: Wiwynn

Microsoft and Wiwynn are not the only cloud and DC ecosystem players announcing liquid cooling innovations this month. Vertiv just introduced a 1P immersion cooling solution, called Liebert VIC. It can support high-density loads of up to 100kW per rack, utilizing warm water. It also uses a dielectric liquid coolant, which is odorless, non-toxic, and electrically and chemically inert and has 1,200 times the heat capacity of air. The vendor estimates Liebert VIC can reduce cooling energy costs by up to 95%.

Omdia expects a steady flow of partnerships and announcements as more large DC operators turn to liquid cooling. For example, Wiwynn recently invested in LiquidStack to develop new 2P immersion cooling solutions for large and edge DCs alike.

Bottom line: Skate where the puck will be

Hall of fame hockey player Wayne Gretzky described his method for success in the quote "I skate to where the puck is going to be, not to where the puck has been." Although not yet at an industry inflection point, the long-term trend is clear: the need for liquid cooling is increasing. Average server rack density has nearly quadrupled in the past nine years from 2.4kW per rack in 2011 and 5.6kW per rack in 2017 to 8.4kW per rack in 2020, according to the 10th annual *Uptime Institute Global Data Center Industry Survey*, which was released in July 2020.

Most liquid cooling deployments in 2019 were still small in scale, with many still evaluating early proof-of-concept deployments. Larger and repeat deployments are increasing but remain a small portion of deployments. Still, the trend toward liquid cooling is real, and it is clear where the puck is sliding.

Some questions still remain. If liquid cooling enables higher compute density, with a power envelope of 100kW per rack, how do we power a DC with thousands of 100kW racks? Is a hyperscale DC that needs as much electricity as an entire town even feasible?

Appendix

Further reading

Data Center Thermal Management Report – 2020 Analysis (December 2020)

"Uptime Institute 10th Annual Global Data Center Survey Shows Increasing Complexities, Outages," Uptime Institute, https://uptimeinstitute.com/about-ui/press-releases/uptime-institute-10th-annual-global-data-center-survey, retrieved April 2021.

Data Center Server Equipment Market Tracker – 4Q20 Analysis (January 2021)

The Canary in the Coal Mine for Liquid Cooling is Server Shipments with Coprocessors (November 2018)

Authors

Greg Cline, Principal Analyst, Cloud and Data Center Research Practice

Vlad Galabov, Research Director, Cloud and Data Center Research Practice

askananalyst@omdia.com

Citation policy

Request external citation and usage of Omdia research and data via citations@omdia.com.

Omdia consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help you. For more information about Omdia's consulting capabilities, please contact us directly at consulting@omdia.com.

Copyright notice and disclaimer

CONTACT US

omdia.com

askananalyst@omdia.com