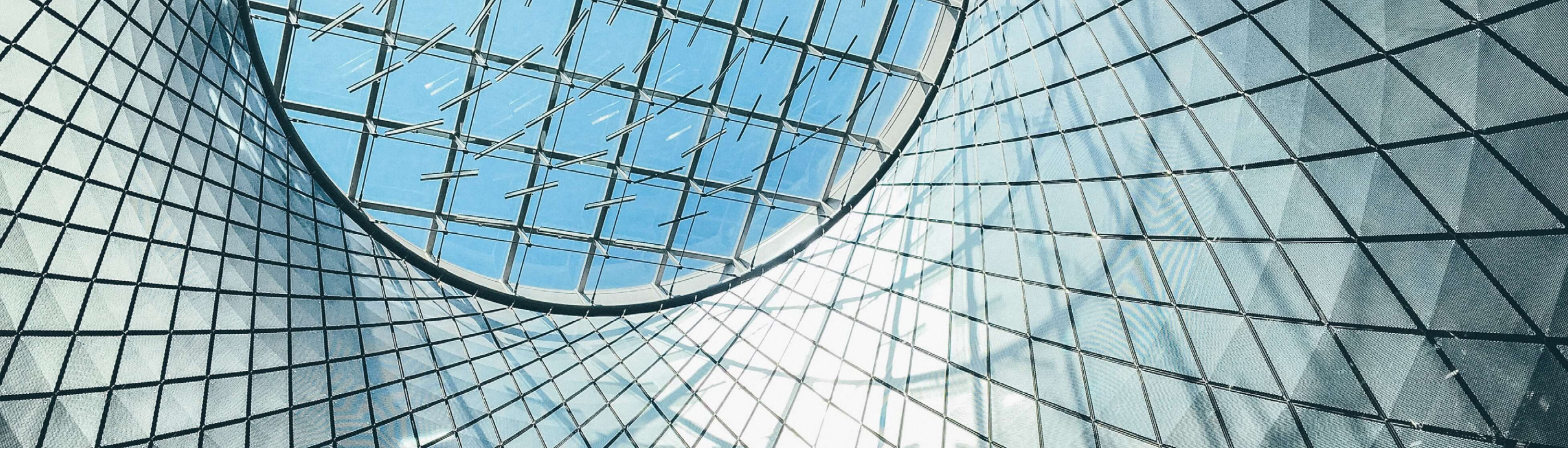

Market Insight Report Reprint

Liquid cooling is not what you think – it is becoming a flood of options

February 13 2020

Daniel Bizo

Customers identify liquid cooling with specific implementations they've come across (and likely rejected) rather than the broad category it is, with a diverse set of possible approaches to fitting out and running a datacenter. This view needs to change if liquid cooling is to become standard in datacenters.

451 Research

S&P Global
Market Intelligence

Introduction

Direct liquid cooling is mounting a comeback as the next big thing in datacenters. Vendors, established and emerging alike, sense an opportunity for (or threat from) real change this time around as the technology and business environment puts datacenters under increasingly more pressure to deliver online services for less money and less energy. A key obstacle, beyond even the huge market inertia around air-cooled facilities, is perception: Customers identify liquid cooling with specific implementations they've come across (and likely rejected) rather than the broad category it is, with a diverse set of possible approaches to fitting out and running a datacenter. This view needs to change if liquid cooling is to become standard in datacenters.

THE 451 TAKE

Datacenter operators prefer to design and operate their facilities as they see fit, and will not budge under pressure from mere laws of physics that dictate that liquids are orders of magnitude better at heat transfer than air. Past attempts at popularizing liquid cooling in datacenters (outside supercomputing) have largely failed because prescriptions, no matter the benefits, don't go down well with customers, who, of all things, distrust change the most. And change is exactly what many have tried to sell, and in wholesale quantities. Current attempts are not only cleverer and more sophisticated in their engineering compared with their predecessors, but also offer much-reduced business friction for operators in using liquids to cool their datacenter IT. This is moving the sector closer to mass adoption more than any cutting-edge but inflexible liquid-cooling system.

Context

Some emerging technologies benefit from novelty value, such as electric cars – even when there is not much new about them, their latest incarnation manages to get customers excited. There are other emerging technologies that don't enjoy that benefit because they've been around for ages, and therefore leave potential customers uninterested even when they achieve clear economic, energy-efficiency and carbon-reduction advantages. Unfortunately, cooling datacenter electronics directly with liquid (as opposed to using air as a medium to transfer heat from electronics) belongs to the latter category despite a clear scientific and an ever-stronger business case in favor of the shift.

There are numerous reasons, minor and major, why most datacenters use air as a cooling medium. Chief among them is organizational inertia in an industry that is extremely averse to change because it's rewarded not for innovation and efficiency, but for not making mistakes – especially big ones that could lead to a service outage. Then there is the missing alignment between datacenter facilities and IT teams, which leads to inefficiencies in the datacenter because IT is by and large disinterested in the cost inflicted on the facilities side, be it capital or operational. For a strategic move to liquid cooling, close cooperation between IT and facilities is necessary, including respective supply chains.

However, more recent developments signal that datacenters might finally approach financial, social and technical breaking points with current design practices. Even the most efficiently built air-cooled facilities carry as much as a million dollars' worth of extra cost per megawatt IT load for the additional electrical capacity as a reserve to cover the peak power requirements of air cooling systems (facility and server fans, water pumps, compressors). The added burden is not purely budgetary either, but of access to power. As multi-megawatt datacenters mushroom in cities, immediately available free power capacity can quickly become a scarcity. Limited available power capacity delays (substation upgrades can take years) or downsizes (to fit into existing substation envelopes) datacenter capacity development.

Additionally, there are growing social and environmental concerns over the size, noise and resource footprint of datacenter buildings, which in some cases lead to policy action against datacenters – Paris and Amsterdam are prime examples. Public policy increasingly scrutinizes carbon emissions from the electricity consumption of datacenters. Any energy not spent on compute and storage is wasted, and most modern datacenters still waste 20-40%. On the technical side, processor technology (general purpose and accelerators alike) is pushing air cooling to its practical limits, and silicon power density will continue to escalate into the 2020s as scale of integration outpaces transistor energy-efficiency gains. Liquid cooling not only supports this trajectory, but also makes chips perform better for the same money.

The evolving hydrography of liquid cooling

Computing has known direct liquid cooling for many decades, starting with high-performance mainframes in the 1960s. Arguably, liquid cooling dominated server-class computing multiple times through the following decades, most recently in the early 1990s before mainframe processors made the transition to complementary metal-oxide semiconductor technology (the currently dominant branch of semiconductor fabrication) from bipolar, which began to generate too much thermal power even for liquid coolers.

This long history in mainframes spurred the later development of the various water-cooled cold-plate techniques that permeate the high-performance computing market today, and led French hosting and cloud company OVH to adopt water cooling as its standard in its facilities and to become possibly the world's largest operator of liquid-cooled datacenter infrastructure. Water cooling has seen success primarily because water (and water-based fluid mixes) is low-cost and offers high thermal performance, which makes it ideal for high-density computing clusters that tend to pack server hardware at extreme densities.

While many in the datacenter industry associate liquid cooling with water, the market has seen numerous streams of technology development in the area that all share a common feature: the use of dielectrics instead of water. The single biggest problem with water is the piping work required and the statistically unlikely, but ultimately inevitable, leaks that might lead to system failures. While OVH's use has proved it is certainly possible to operate water cooling at scale, its engineering choices are specific to its own facilities (it almost exclusively operates its IT infrastructure out of its own datacenters) and mode of operation, where it controls every aspect of the stack, including manufacturing of hardware. Its design is simply not applicable for others.

Dielectrics have let ideas flow in many directions, inspiring both academia and industry to develop novel approaches. As a result, liquid-cooling options today cover a wide range. Cold plates that use dielectric can be much simpler and lower cost (close to being immaterial compared with the value of the IT hardware in some cases) – the chance of leakage and the damage caused by leakage are negligible compared with water systems. Israeli startup ZutaCore is such an example. The company uses a two-phase dielectric with a low boiling point to create a self-regulating system that requires a low amount of dielectric coolant fluid and operates below atmospheric pressure. The more thermal power the processor dissipates, the faster the coolant evaporates in the small chamber of the cold plate and makes its way to the condenser.

British liquid-cooling specialist Iceotope, embraced by Schneider Electric, developed a flexible framework that supports both cold-plate cooling and total cooling in sealed server chassis via precision delivery of single-phase dielectric coolant (combining cold-plate techniques and total liquid cooling by allowing the coolant to overflow onto the server board), allowing customers to mix various techniques in the same cabinet. Additionally, there are immersion tanks coming in many flavors from established companies such as GRC, as well as more recent challengers Submer, Asperitas and TMGcore, which are taking different approaches to engineering (largely differing in what coolant is used and how it's circulated for heat exchange), positioning (cost-performance trade-offs) and business model (fully productized or available as IP and engineering services). Some major server OEMs have also invested in their own nonwater liquid-cooling systems, but typically to limited commercial success, if any.

The goal here is not to list all the options and vendors available (there are plenty more), but to demonstrate that there is a growing range of possibilities. As engineering and businesses evolve, datacenter operators will have more flexibility to make the change to liquid cooling at the speed and in the form of their choosing. This is perhaps the most important development of recent years in direct liquid cooling – vendors have woken up to the reality that datacenter customers won't rip out and replace their entire way of designing, building and operating mission-critical datacenter infrastructure, including their trusted suppliers.

A growing number of liquid-cooling vendors speak not only of customization options, but also of co-engineering and licensing intellectual property to customers and equipment makers, to be able to integrate into datacenter supply chains. Air cooling does not limit equipment-buying choices or tie the hands of facility engineers, and neither should liquid cooling. Those vendors that understand this and adapt their engineering and business models accordingly should see their chances improve.

CONTACTS

The Americas
+1 877 863 1306
market.intelligence@spglobal.com

Europe, Middle East & Africa
+44 20 7176 1234
market.intelligence@spglobal.com

Asia-Pacific
+852 2533 3565
market.intelligence@spglobal.com

www.spglobal.com/marketintelligence