REPORT REPRINT

Is air cooling running out of breath?

APRIL 12 2019

By Daniel Bizo

The unexpected downsizing at Munters' datacenter division is concentrating minds in the datacenter sector on the future of cooling. The big question for makers of air systems is which way the technology and the market will go – is the only way down?

Introduction

When Swedish cooling specialist Munters announced the closing of its European datacenter business in February – and potentially quitting entirely in response to poor profitability – much of the datacenter sector found the news perplexing. Munters, after all, was one of the more successful makers of datacenter cooling systems. While some of the reasons behind the move are specific to the company, it serves as a warning for the rest of the industry that the datacenter equipment market is a difficult one. The big question for makers of air systems is which way the technology and the market will go – is the only way down?

451 TAKE

Air cooling is predominant in datacenters today, and for good reason. The engineering is well-understood, the products are well-honed and the operational regime around them is mature. Arguably, it has also become outstandingly efficient, leaving it up to the datacenter operator how aggressively they want to pursue energy efficiency, weighing it against other business objectives. However, the market appears to be saturated, and while there is growth for the time being, structural and technological shifts in the market point to a looming change of direction in the not-so-distant future. The best days for air cooling are probably over, and its solid place in datacenters will likely give way to a new flow of liquid cooling systems as we enter the next decade.

Context

Over the course of the last three decades, building and operating datacenters has become a big business. Cooling (including humidity controls) remains a major expenditure in datacenter budgets, which translates to a market worth billions of dollars, predominantly in various forms of climatic (air) systems. What has driven this spending is the long-held consensus that temperatures and humidity need to be tightly controlled for IT systems in order to minimize failure rates.

As IT systems and components became relatively inexpensive commodities, datacenter operators and their IT tenants warmed up to the idea of more relaxed temperature controls, which in turn opened up the datacenter market to a new type of cooling: evaporative air economizers. These air systems use ambient air as their primary mode of cooling, as opposed to compressors. This is also known as free cooling.

In the last 10 years, direct (external and internal air mixes) and indirect (no mixing, only heat exchange) evaporative air systems have become popular in commercial and webscale datacenters. The numerous suppliers that have entered the fray include major players such as Schneider Electric, Nortek and Vertiv, as well as specialists Excool, Energy Labs (acquired by Vertiv in 2018), Munters, Stulz, Silent-Aire and many others.

Munters, which sells climatic systems across a range of verticals, is one of the suppliers that rode high on the back of this trend. Some hyperscale operators favored its cooling systems, especially in Dublin, where the climate allows for highly efficient and completely mechanical-free operations. The Swedish company has sold nearly two thousand units of free-cooling systems with a combined cooling capacity of hundreds of megawatts.

Its retreat from the datacenter market (shutting European operations while maintaining some restructured US presence) stands in stark contrast to these project wins, which raises some questions about the broader market. Munters' issues cannot simply be dismissed as isolated due to, for example, poor performance.

Some key factors in the company's decision are specific to Munters, such as its stand-alone organizational structure for datacenters. A manufacturing cost base in Belgium probably didn't help it turn stronger profits in good times, and may have contributed to losses in bad times. Worse still, its heavy reliance on a handful of massive projects also likely undermined its pricing power, while the pressure on its factory to meet deadlines at scale exposed it to the financial risks stemming from delays or any systemic quality issues.

At minimum, the tale of Munters' datacenter business is a reminder of the difficult environment all equipment makers face. A structural shift in demand from a high number of one-off enterprise deals to a relatively small number of high-value projects challenges the consistency of sales and the health of margins due to a combination of pricing pressure and straining manufacturing and supply chains.

Munters' failure to achieve wider success and profitability via a more diverse customer base also points to another hard truth about the datacenter market: Business reality trumps academic technology discussions. On paper, direct and (particularly) indirect air economizers should completely dominate the market by now. But operators often stick to existing practices such as chilled-water systems because using air economizers would necessitate a major redesign of the facility layout, in addition to requiring extra space and retraining of staff. There is also some loss of flexibility in that the lack of chilled water means no support for close-coupled cooling units for high-density cabinets. At the same time, chillers have become much more efficient, operators have relaxed water temperatures and most suppliers offer free-cooling modes.

Have we reached peak air cooling?

This raises the question: What's next for cooling? When energy efficiency of best-in-class free-cooled facilities is already outstanding by historical standards, further optimization potential is marginal. Indeed, while expectations are much higher than 10 years ago, we don't see aggressive efficiency gains high on the agendas of operators. Air-cooling technology appears to have run its course. Products still get better, lower cost, smaller and easier to operate, but the days of major gains are over and the rule of ever-diminishing returns has kicked in. This points to less differentiation over time and growing cost competition.

Equally, market dynamics offer little reason for much optimism. Yes, datacenter capacity is still growing rapidly, somewhere around 7.5% per year on average in the next five years, as witnessed by 451 Research's Worldwide and Regional Datacenter Monitor. However, the reality of a small number of powerful buyers commanding a growing share of the market is inescapable. Other vendors may achieve healthier margins in their dealings with hyperscalers than Munters did, but it is hard to see any upside potential.

At the same time, processors and accelerators have become much more power-hungry. Intel's latest-generation server platform, the de facto standard in datacenters, now supports processors with up to 400W of thermal power, roughly double of the previous maximum, to deliver more performance. Cloud providers and tech companies have a tendency to prefer high-power processors because they improve overall infrastructure performance, which allows them to support more revenue-generating business. NVIDIA's chips for deep learning acceleration and high-performance computing jobs are similar beasts, and tend to end up densely packed for high-speed data sharing.

These chips are also opportunistic in their performance profile. Simply put, more cooling capacity on a chip level translates to more performance via higher sustained clock speeds. Cooling such systems at scale, if it becomes mainstream, calls for liquid cooling, not air. But what could really seal the deal for large players is the ability to free up power capacity and repurpose it for more IT. With fewer or no fans in servers and in the facility, and no compressors, 15-30% higher IT load can fit into a given site power envelope. This could be increasingly valuable as cloud infrastructure expands into increasingly more metro areas, some of them in short supply of readily available power for new datacenters.

Given all these factors, hyperscalers and some major colocation providers are actively looking at various new liquid cooling options for a potential change for their next-generation designs – engineering around liquid cooling has come a long way in recent years to offer more flexibility and reliability at scale, and is more mindful of maintenance and operations. Sooner or later, some of them will make the change and others will follow, 451 Research believes. Some, like French cloud and hosting provider OVH, already have.

Precisely when this shift will start remains the big question, but the business of air cooling, while still dominant, is fast approaching its peak, 451 Research believes. From there, the reversal of trends might be more sudden than what the history of datacenters guides for.