DATA CENTER Frontier

SPECIAL REPORT

Liquid cooling is in your future

Liquid cooling is the future of the data center. While the option exists to build future data centers from the ground up as liquid-cooled facilities, the reality is that the transition to liquid cooling is going to be a gradual process. Existing facilities will upgrade to liquid cooling options as equipment reaches EOL and they add support for new technologies on site, such as AI/ML (Machine Learning), and HPC.

Report Compiled By: David Chernicoff
Senior Editor, Data Center Frontier

brought to you by

nVent

Contents

Liquid cooling is in your future

Liquid cooling is the future of the data center. While the option exists to build future data centers from the ground up as liquid-cooled facilities, the reality is that the transition to liquid cooling is going to be a gradual process. Existing facilities will upgrade to liquid cooling options as equipment reaches EOL and they add support for new technologies on site, such as AI/ML (Machine Learning), and HPC.

WHY LIQUID COOLING IS IMPORTANT.

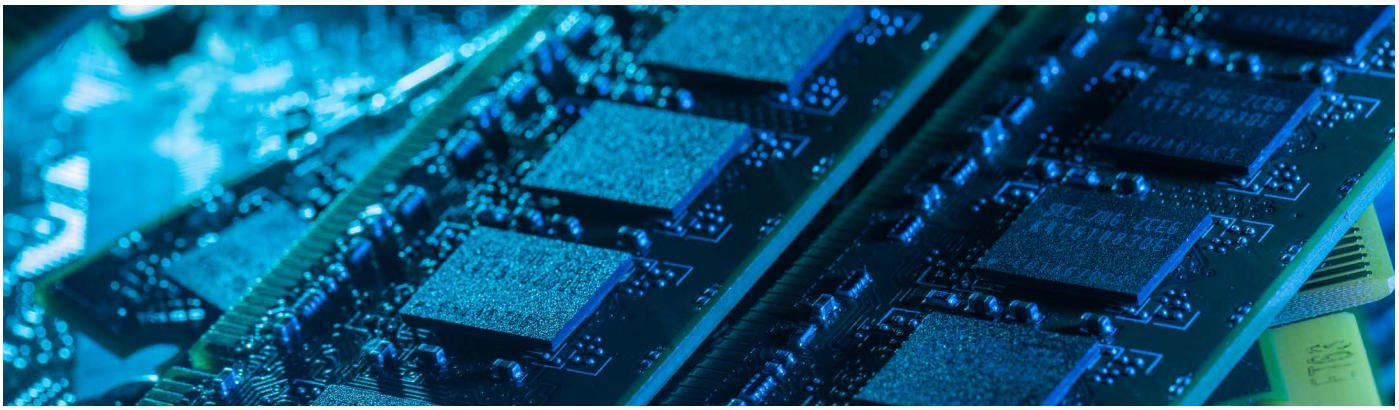
You've built a stable, effective data center for your operation. It meets your current needs and the future growth that was projected when the facility was designed. Perhaps most importantly, its operational costs are well established and easy to budget. So why do you need to consider major changes to the way the data center operates? Why is liquid cooling on the table?

First, air cooling on its own is no longer sufficient to simply meet your current operational needs. As your IT workload equipment moves to the end of its lifecycle, your replacement hardware is going to be more space efficient. The equipment deployed at the beginning of your last refresh cycle is going to be replaced with hardware that is faster, more effective, and more efficient at doing the necessary work.

This isn't new. You can walk into a data center built a decade ago and find that a data hall that housed a dozen racks may now only have a single rack that is handling a much higher workload than the original deployment. And the first thought to cross your mind is likely "Why do I need to add more space? Why aren't we using the free space here?" Quite a few data centers were constructed with the ability to increase the power available within the facility, but far fewer were built to handle the increased heat load brought on by higher rack densities. There's a good chance that adding additional capacity isn't possible because you will exceed the capacity of the air-to-air cooling systems built with the data center. And wholesale rip and replace is rarely a practical solution.

Second, environmental issues and a focus on sustainability means that you need to make your facilities more environmentally friendly. Cooling systems that are more efficient reduce energy demands, the number one complaint when evaluating data center changes. Additionally, reducing environmental footprint, be it noise, power, space requirements, or any other factor, must be on the list of data center improvements.

Third, technology will continue to change. Be it HPC, AI, or any new technology that places an increased demand on the IT workload supported by the data center, the demand for a more efficient data center will continue. While some changes can be planned well in advance, others, as highlighted by the rapid demand for supporting AI workloads, will come at a fast and furious pace. And you will need to be able to respond to those demands quickly to maintain business agility.



WHY DO I NEED TO ADD LIQUID COOLING?

While you may not be looking at making changes to your existing hardware deployments, the availability of liquid cooling options within your data centers will increase their flexibility and capabilities.

Upgrades or additions to your existing data centers should utilize hybrid liquid/air cooling technology. This means that your existing air-cooled operations won't be impacted and you'll have the option in new hardware deployments to determine the most cost effective and solution-focused choices as you place the technology necessary to solve your business problems.

If there is good justification for construction, building new data centers or additions to existing facilities that are completely liquid-cooled can be a smart choice. The wide selection of liquid cooling options available means that you will most likely end up with a hybrid cooling model, but should your IT workload demand it entirely liquid-cooled data halls that serve specific needs are a practical solution to a number of potential issues. In part, this is because liquid cooling isn't a single solution. There are a number of different liquid cooling solutions that can be deployed in concert with each other or to solve specific point problems.

WHAT ARE THE OPTIONS?

Let's look at the most common options for liquid cooling.

1. Liquid-to-air cooling is commonly the first choice when a data center workload is approaching the limits of its air-to-air cooling capabilities. In these systems, coolant liquid is circulated near heat sources, removing the heat from IT devices, then moving to a radiator which performs the liquid-to-air heat transfer. The liquid is then circulated back to the IT devices. This technology is most commonly represented by cold plates which sit directly on CPUs, GPUs and memory, but also applies to technologies like rear door heat exchangers which move heat from an entire rack to the external radiator system.

2. Liquid-to-liquid cooling operates in a similar fashion to liquid-to-air cooling but makes use of the higher heat load capabilities of liquids when compared to air. In this situation, there is often a closed loop system that performs the initial cooling of the IT equipment. That system is cooled by a secondary liquid system. The cooled liquid is then pumped back to the IT devices while the secondary system uses a radiator or cooling tower to remove the heat from that liquid so that it can then be cycled back to cool the primary system. This system adds a layer of complexity when compared to liquid-to-air cooling but is capable of cooling significantly higher power densities making it a more efficient solution for those environments.

3. Direct immersion cooling involves the hardening of the electronic components so the entire server can be submerged in a dielectric liquid coolant. The coolant, which can be as simple as mineral oil or as complex as specialized fluids designed for this application, is then circulated around the servers and then through external heat exchangers. This method of cooling is extremely effective at cooling high-performance systems and can be more effective than other forms of liquid cooling. Immersion cooling, while having its own specific complexities, does have the advantage of reducing the need for traditional cooling tools such as air conditioning, chillers, and the like. Because the fluid involved is non-conductive, any potential electronics issues due to a leak are minimized, as well.

None of these solutions are exclusive and can be mixed and matched to meet specific cooling needs.

IS THERE A LOGICAL PROGRESSION FOR DEPLOYING LIQUID COOLING SOLUTIONS?

Most enterprises are going to be looking at adding liquid cooling solutions to their data centers as the need for the additional cooling capabilities becomes clear. So where can we start and what are the goals?

Most data centers find that they need to support a specific task that requires high densities and increased cooling capabilities. Right now, this is likely to be a cluster deployment for high performance computing or to support artificial intelligence and machine learning. While you can utilize many of the standard data center techniques (hot aisle containment, etc.) dropping a significant source of heat into an existing facility can bring a new set of problems. That is what makes this the perfect opportunity to begin to introduce liquid cooling solutions to your data center. While choices such as in-row cooling, rear door heat exchangers, direct-to-chip cooling, and immersion cooling are all available options, starting with the simplest solution, such as a passive rear door heat exchanger, can minimize the impact on your data center while allowing optimal performance of your high-density computing solution.

This variety of liquid cooling options makes it possible for liquid cooling to be deployed gradually, rather than as a rip and replace solution. It also allows for interim choices. You may choose to use the rear door heat exchanger for a single rack while you build out a more complete liquid-to-liquid cooling solution that will be available as the high-density deployment grows. Or you can save the complex deployment for your next generation data center and have a progressively increased level of cooling solutions available in your existing space, choosing to add RDHx systems,

enclosure cooling solutions and liquid cooled cabinets that now give you a variety of solutions that can be matched to the demands of your IT workloads.

There is significant flexibility available with cooling choices that do not require reengineering your data halls or entire data center. Mixing and matching those solutions to the specific demands of the IT hardware can increase your efficiency in the data center while making more options available to meet those specific needs. An average server in your existing data center generates about 1.5kW of heat; according to Nvidia, a latest generation AI server using their GPUs can generate five or six times that much, so configuring your entire data center to support that level of cooling demand is unlikely to be efficient; finding the right solution to solve that point problem will be the short-term answer.

It is also important to note that not everything in the data center needs to be liquid cooled or even should be liquid cooled at this point in time. Devices like switches, routers, and network interface cards typically aren't liquid cooled, as heat generation is rarely an issue. Storage devices are just beginning to see the availability of specific tools to keep them operating at lower temperatures, as heat can reduce MTTF, though they don't see the same huge heat generation that a rack of AI GPUs can create, so adding specific cooling for storage is definitely a point solution that few will require. Other common data center equipment such power distribution units, backup batteries and the various other pieces of electronics found in the data center don't often require additional cooling, though if you chose to, using liquid cooled enclosures would enable you to cool any rack mounted equipment you would choose to install in such an enclosure. These devices are rarely the point of failure due to an increase in heat in the data center.

EVALUATING YOUR ENVIRONMENT FOR THE MOVE TO LIQUID COOLING.

The more things change, the more things stay the same in your data center environment. You still need to be able to deliver power and cooling to the IT equipment in the data center, regardless of the tasks being performed. What you want to avoid is that delivery becoming a roadblock to your business workflow. Like almost all business tasks, understanding the environment, both in terms of technical requirements and business process needs is the starting point for evaluating where and how the addition of liquid cooling capabilities can improve overall business workflow.

Changing the way you cool your data center workloads is a fairly fundamental change. Evaluating the environment isn't strictly about answering "how do I run this workload?" but also, "how do I run this workload optimally, taking into account not only the business workflow, but energy efficiency, sustainability, flexibility, being a good environmental citizen, and still gain the required businesses advantages of any change to my working environment?" Plus, the key takeaway here is that the capability of data centers is not going to be defined by how much power I have available, how high a power density my racks can support, or how much workload my IT equipment can handle, it's going to boil down to how much capacity can I effectively cool.

You are probably well aware of the power and cooling requirements for your existing infrastructure. You know how much power you use and how effectively you are currently cooling that workload. So, this gives you the starting point for evaluating the impact of adding these new, high-density, power and cooling intensive workloads. We know it's not going to be as simple as saying a certain rack is now going to draw this much more power and will be generating this much more heat. Your current infrastructure may, on paper, seem to be able to support the additional workload. But we are all familiar with problems that can crop up when adding heat to a data center. Reconfiguring the data center to support the additional workloads might be as simple as adding hot or cold aisle containment to an existing set of racks, or it could require adding rear door heat exchangers to all of the rack locations getting the new hardware. This is not something you want to find out after the fact.

Some reconfigurations may be necessary simply to stay ahead of the expected growth of your existing workloads. You know that higher power densities are coming. If your data center can deliver the required power your next step is to make sure that you can cool those densities so that the workload growth is sustainable. Make sure that you understand your current thermal workloads; it's the only way to be sure that you can plan for their growth. You know what tasks are currently being run in your data centers. Now you really need to be forward looking and determine whether future growth will be taking place on-premises or in the cloud. And if like most businesses you are building an even more complex hybrid cloud workflow, you need to assure the business side that the IT side is ready to handle the workload, especially when it is deployed on premises. This means being ready to deploy liquid cooling solutions, as needed, within your data center.

BUILDING FOR THE FUTURE

It is becoming clear that HPC and AI/ML are going to be a big part of the future of business, and being able to support those workloads is going to require that you make liquid cooling available in your data centers. Fortunately, the options for deploying liquid cooling are flexible enough that a business can start by deploying point solutions to solve a particular problem set, such as cooling a single rack or small set of racks where your AI/ML solution lives, without needing to deploy full scale data halls that are liquid cooled and support rack after rack of equipment running HPC/ML solutions.

This is also where your understanding of future business plans is necessary. While it can be very simple to solve a single problem with a dedicated liquid cooling solution, that same solution may not scale well if it needs to be implemented a dozen or more times. While it may work it is unlikely to be cost effective or energy efficient.

The advantage of these point solutions is that they enable you to dip a toe in the water without committing to deploying full scale solutions. But you need to have a plan in place that will enable you to deploy your solution at scale. This may mean planning your next generation data center to be fully liquid cooled or spending significant time in building your workload placement model for your hybrid cloud.

But the biggest advantage that planning for the future and understanding your workloads will bring is the realization that almost all of these solutions can be done in combinations, allowing you to completely match your power and cooling capabilities with what are everchanging demands from the IT workloads, as the business adapts to the next generation of must-have technologies. The real key to sustainable growth is having different options for supporting the next-generation workloads and liquid cooling gives you the greatest flexibility for that future.

Part of the solution will be finding the right partner to work with when building out your next-generation cooling solutions. This is especially true if you believe that you will need more than a single solution. This is not advocating for vendor lock-in by suggesting you find a single supplier for all of the various solutions but rather that you understand that there are not a lot of standards that apply to this equipment at this point in time. If you select, for example, a vendor for rear door heat exchangers, you are better served by having a single vendor who can provide any RDHx solution you need, rather than trying to mix and match solutions from multiple vendors. This also applies to plans to grow your cooling solutions within your environment. What is your point solution now is not something you want to have to retire if the next step is an incompatible solution that requires replacement of a successfully deployed point solution. Having that path to more capable solutions is a valuable capability.

IT'S TIME TO BE GETTING READY

Your environment is going to be changing. You're looking at high power density, higher temperature operations. But there's no new reason that you must jump directly into one solution right now. Your current data center is meeting your current needs; will it continue to do so in the future is the question you need to answer. You can determine how far your existing cooling model is going to take you and begin to plan for what the future of your data center needs to look like but remember you don't have to go directly to the latest and greatest in liquid cooling.

Getting a head start means that you can build test systems, experimental environments that let you evaluate the different flavors of liquid cooling, be it anywhere from the rear door to a full immersion system. You don't have to jump into any one thing right now. But you want to be positioned to be ready for the next wave of changes to your data center.