

Fluid Dynamics: Liquid Cooling's Role in the Future of Data Centers

By Sam Allen

The demand for data is higher than ever and driving the need for more efficient data centers. Liquid cooling technology is helping to transform data centers by revolutionizing these centers' infrastructure as the sector continues to grow. When making plans for the design and construction of data centers, having a road map on how to integrate liquid cooling will prove invaluable.



As yesterday's supercomputers get outpaced by today's superchips, the amount of compute load possible in the same footprint is growing exponentially and all that work needs cooling — and fast.

The aviation and space industries have used liquid cooling for their high-density applications for decades and NASA may be the most famous for its application of liquid cooling, as used in the Apollo spacesuits. These industries obviously know the value of physical space and recognize the optimal way to carry out computing applications is to densify load using compact liquid cooling solutions like those used in the evolution of avionic electronics.

Given the current landscape, private industry is applying the cooling concept on a massive scale in the data center market, with liquid-cooled data centers being constructed at 10 times the density of those cooled wholly by air.

Projections suggest that the total capacity of operational hyperscale data centers is expected to triple over the next

six years. Understanding what liquid cooling is, how it works and under what circumstances it should be used, is critical for project owners seeking to build data centers designed for the future.

Why Liquid Cooling?

Liquid cooling is beginning to manifest on a broad scale wherever systems with high compute loads are found, such as those that employ artificial intelligence (AI) or machine learning (ML).

The densification of compute and an increased need for power and telecom components obligates us to cool in a different, more efficient way.

The data center market has seen legacy operating temperatures rise and acceptable humidity ranges expand as information technology equipment (ITE) has changed, with some operators allowing for server inlet temperatures up to 90°F/32°C. However, air is still an inherent insulator and has a fairly low heat capacity, so when the power consumed and

the TDP (thermal design power) increases exponentially, the practicality of air as a cooling medium falls off quickly.

When changing the cooling medium over to a liquid, for example, the heat capacity increases by a factor of 3,200 per unit volume, and energy can be moved around the data center much more efficiently.

In the past, data centers exclusively relied on air cooling, but there are practical limitations to how much heat air can effectively dissipate. As AI/ML has evolved, the actual silicon chips and components needed to support these new loads have advanced to the point that manufacturers of these components are no longer permitting them to be cooled by air. Rather, these components must be liquid cooled.

As this chip technology continues to evolve with loads becoming more and more dense, liquid cooling is becoming increasingly prevalent in data centers, and this shift is expected to increase well into the future. Our next challenge in the architecture, engineering and construction (AEC) community is to evolve our thinking, strategy and approach as quickly as possible to deploy these shifts in cooling topology while still driving down relative energy consumption and conserving resources.

How Liquid-Cooled ITE Works

When considering deploying liquid-cooled ITE, an owner has a wide array of options of how to cool the increased loads. Instead of server racks being deployed at 8 or 12 kilowatts (kW) a rack, for example, there are solutions to deploy at

80 to 120+ kW a rack. This opens the door to a broad spectrum of possibilities to manage high-density applications. In this white paper, we will share what we have seen to be an emergence of three types of liquid cooling at the ITE level: rear door heat exchangers, direct-to-chip cooling and immersion cooling. (See Figure 1.)

Rear door heat exchangers (RDHX) are cooling devices that are installed at the back of server racks. They operate on the same principle as a radiator in a car, deploying a liquid-to-air heat exchange. Using this cooling approach, the ITE remains air cooled, but the heat generated is transferred to a liquid inside a finned tube coil (the radiator) on the rear of the server rack by a fan forcing air across the RDHX.

This liquid then carries the heat away to the facility's cooling infrastructure, which may include cooling distribution units (CDUs), liquid-to-liquid heat exchangers, and ultimately a heat rejection plant consisting of chillers, cooling towers, dry coolers or other equipment. The liquid inside the coil is referred to as the technology cooling system (TCS) and is refreshed at the CDU. As with many of the other cooling options, the location of the CDU is variable and may be rack mounted, located in an adjacent cabinet space or located in a nearby mechanical space. This liquid-cooled approach transfers approximately 100% of the server heat to the TCS and the discharge of the RDHX is typically the same temperature as the cold aisle, so containment and aisle relationships are not as critical as in a traditional air-cooled data center. This solution is an example of air-assisted liquid cooling (AALC).

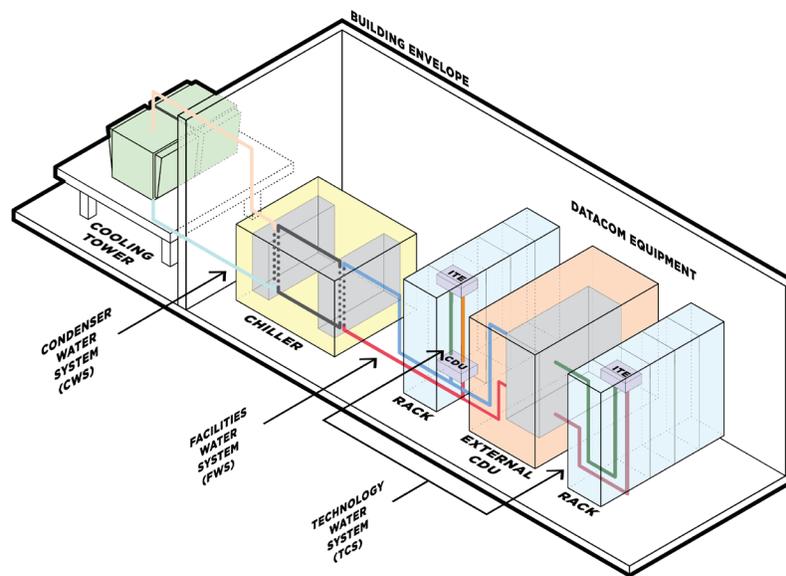


Figure 1: Anatomy of liquid-cooled data center infrastructure. Data center operations that optimize liquid cooling are the wave of the future.

Direct-to-chip cooling is an innovative approach in data centers where heat is dissipated directly from the chips to a TCS fluid, (sometimes a dielectric fluid), using application-specific heat sinks in the ITE. Because of the much higher heat-carrying capacity of liquids, this method boasts higher densities at the server level and therefore per rack. Similarly, this fluid carries the heat away from the servers to CDUs, mounted in a variety of locations, where the liquid is refreshed and made available again to the servers.

As a part of this approach, there is still some heat produced by the components that are not cooled by liquid and, therefore, must be cooled by other means, such as smaller air-cooling solutions. However, sometimes the opportunity to increase density per rack creates a situation where the air-cooled load, albeit a small percentage of the total load, requires similar infrastructure as seen in legacy data centers. It may be obvious, but this solution is an example of direct liquid cooling (DLC).

Immersion cooling is an advanced data center cooling technique that offers remarkable density, efficiency and compactness. When implementing immersion cooling, the entire server is immersed in a dielectric fluid that absorbs the heat generated by the servers entirely. Most immersion cooling solutions take the form of horizontal servers arranged in tanks filled with a dielectric fluid that is circulated throughout the tank. This dielectric fluid, otherwise known as the TCS fluid, is circulated through CDUs, creating the demarcation point between facility systems and technology systems. Immersion cooling holds the promise of the most densification, significantly enhancing data center efficiency and reducing physical footprints, making it a forward-looking solution. Finally, this solution is an example of total liquid cooling (TLC).

Deployments

There are many options and variations of these approaches, but the principles are the same. Different liquid cooling technologies are suited for varying levels of compute load.

Rear door exchangers generally are used for applications with loads of up to 50 kW per rack. Direct-to-chip cooling solutions often are employed in scenarios with larger loads, typically up to 100 kW per rack. For even larger loads — up to 1 MW per pod — immersion cooling often is the most suitable option.

It is worth noting that several cooling technologies can be combined within a single facility. For such versatility to take place, however, the data center itself must be flexible, agnostic and adaptable to diverse cooling approaches. What is important to remember is that the infrastructure required for liquid cooling systems has distinct characteristics

when compared to even the most efficient air-cooled data centers of today. These characteristics could involve modifications to water, power, telecom and other essential infrastructure components.

Even when deployed using 100% liquid cooled ITE, keep in mind some of the ITE power is converted to heat in the form of convection, which can affect a space's temperature and require the use of air cooling systems. This air cooling can range from 20% to 50% of the ITE, typically. Additionally, some of the networking racks are still air cooled, so a liquid-cooled data center is an inescapable combination of air- and liquid cooling solutions.

The choice of what type of liquid cooling to use is dependent on the answer to fundamental questions such as:

- What type of ITE is the owner planning to use and how will it be deployed?
- What is the amount of available power to the site now and planned for the future?
- What mechanical systems at a given facility will serve the ITE?
- Are there any physical site constraints?

While answering such questions, there are a number of factors to consider.

Water and Other Fluids

Liquid cooling involves the use of fluids on the white space floor, which has long been verboten or only permitted with containment and leak detection systems in place. However, there's no way around it with liquid cooling. The liquid-cooled environment is one of piping, cooling distribution units, a form of liquid heat exchange/sync with the ITE, liquid tight hoses, drip-proof connections, and the controls and valving to accompany the solution.

The densification of ITE has created a thermal footprint that requires a large volume of fluid being served directly to the ITE that is unique to the liquid-cooled market. This volume of fluid in the white space should prompt a conversation about the owner's perception of or tolerance of risk. There are options in fluid choices to complement an owner's perspective. Dielectric fluids, engineered fluids and glycol solutions are all common heat transfer fluids found in the technology cooling systems that serve liquid-cooled ITE.

One may be tempted to take facility water directly to the ITE, but this is commonly avoided for a few reasons. First, the chip sets require a much higher level of fluid purity than commonly found in traditional facility water systems. The other main

reason to avoid this is to limit the potential volume of water per circuit, such that if something were to happen a smaller volume would have to be cleaned up.

Regardless of the approach chosen for liquid cooling, TCS fluids must dissipate their heat to a data center's mechanical infrastructure, which is typically accomplished in the form of liquid-to-liquid exchange. As a result, it becomes very common to see large facility water system (FWS) piping mains in the mechanical spaces serving the white space and a significant volume of TCS piping located in the white space itself.

The choice of TCS fluid is crucial to maintain throughout the life of the infrastructure. Each TCS fluid has its own unique density, thermal capacity, viscosity and chemistry. If the fluid changes during the infrastructure's lifetime, changes to the rate of heat removal, CDU components and facility components should be anticipated.

There is an appropriate concern in the data center community about water consumption and water scarcity, with some jurisdictions prohibiting data centers from consuming any potable water whatsoever. The most accepted metric for measuring this is water usage effectiveness (WUE), which can be substantial in traditional air-cooled data centers that employ evaporative cooling, which typically consume millions of gallons of water per megawatt per year.

By transitioning to liquid cooling, the opportunity presents itself to drive the WUE to zero. The WUE of a liquid-cooled data center depends on the final source of heat rejection. For example, if dry coolers can be used for all hours in a specific climate, the WUE trends to zero.

Energy

Implementing liquid cooling introduces unique components into the mechanical and electrical infrastructure and changes the electrical distribution and power consumption profile of a data center. For example, CDUs effectively move the energy consumed by server fans over to the CDU pumps, migrating the energy consumption from denominator to numerator in traditional power usage effectiveness (PUE) calculations.

In other words, the actual ITE compute capacity of a liquid-cooled data center is higher than that of an air-cooled data center as the ITE capacity does not have to derate to account for server fans. There is a discussion happening in the industry that suggests the total usage effectiveness (TUE) may be a more appropriate comparative tool than the PUE. This is a key concept to understand when comparing and contrasting liquid cooling to air cooling.

Facility Infrastructure

Once a data center project owner decides to move forward with liquid cooling, much will change, starting with the initial site selection and choice of a design team partner. As a data center densifies and the white space for a given ITE load decreases, the need for power goes up while the need for land comes down. The challenge facing the market is finding appropriate sites for these densified campuses, as traditional data center hot spots are already facing increased pressures regarding water and power availability. The site selection due diligence process is changing and with it the need for a design team highly knowledgeable in this area.

The building makeup and infrastructure needs to change in order to meet the demands of a liquid-cooled environment. In a liquid-cooled data center, there is an abundance of piping, CDUs, heat exchangers, pumps, chillers and cooling towers with water-side economizers as opposed to the air-side economizers found in legacy data centers. Large facility water piping networks will replace the voluminous obligations of air-cooled systems and exterior equipment allocations will migrate toward an equivalent split of mechanical and electrical components.

The human resources side of the equation must also be examined. The operator's staff and mechanical contractors' skillsets need to change with the new infrastructure topography. As these mechanical systems start to look like large central utility or distributed redundant plants, the need for training on chillers, cooling towers, dry coolers, chemical treatments, large hydronic systems, pumps and hydronic accessories starts to present itself. Training and installation will clearly change but so too will startup and commissioning, operations and maintenance, troubleshooting, and emergency response protocols.

Heat Reuse

As with any data center, the compute power, when converted to waste heat, must be rejected somewhere. The transition to liquid cooling presents not only enhanced efficiency but also the potential to create higher grade waste heat with enough potential energy to still be useful. When looking at the central plant at data centers, the opportunity to evaluate that waste heat is significant. In particular, the potential for using the waste heat for something greater, such as sending it to a neighboring heat host, merits serious examination.

Forward-thinking private enterprises in the U.S. have begun harnessing waste heat and directing it toward establishments such as coffeehouses, agriculture facilities and other industrial facilities. For many of the same reasons that liquid presents itself as superior for handling large densities of



■ Burns & McDonnell Data Center and Power Experience

Figure 2: Data Center Growth. Data center capacity is expected to triple in the next six years. AEC firms that have broad data center, telecom and power experience can help responsibly manage that growth. (Map created January 2024.)

computation loads, the possibility of transmitting usable heat to neighboring entities starts to stand out. The inherent inefficiency of air-based heat transfer for other useful purposes can thereby be avoided.

It is worth noting that Europe is at the vanguard of heat reutilization endeavors largely due to regulatory obligations. In the U.S., opportunities are starting to present themselves not only as sustainable alternatives, but also as opportunities for additional revenue streams. Whether it is charging an internal customer or an external one for energy consumed, there is the opportunity to add to the profitability of data centers.

The Key to Successful Implementation

Implementing innovative liquid cooling technology in a data center raises many challenges, including determining strategies for retrofitting, addressing greenfield construction, managing financing, and handling workforce hiring and retraining. However, the risks associated with inaction and waiting for the existing technology used at legacy data centers to phase out can prove more costly long term.

No matter the type of liquid cooling a project team uses, it is crucial to maintain a responsible approach toward

the planning, construction and operation of data centers. Enhancing the efficiency of both mechanical and electrical systems and managing fluid systems thoughtfully plays a pivotal role in enabling companies to operate optimally and as responsible corporate citizens. (See Figure 2.)

Liquid cooling affords data center project owners the opportunity to get dense and efficient. But owners need help navigating the intricacies of these types of efforts. The place to find the needed knowledge for mission-critical data center projects is at the intersection of data center experience and utility know-how in collaboration with a well-seasoned architecture, engineering, and construction partner.

About Burns & McDonnell



Burns & McDonnell is a family of companies bringing together an unmatched team of engineers, construction and craft professionals, architects, and more to design and build our critical infrastructure. With an integrated construction and design mindset, we offer full-service capabilities. Founded in 1898 and working from dozens of offices globally, Burns & McDonnell is 100% employee-owned. For more information, visit burnsmcd.com.