VERTIV WHITE PAPER

Understanding Data Center Liquid Cooling Options and Infrastructure Requirements

Executive Summary

With more and more processing power being packed into servers supporting artificial intelligence and other processor-intensive applications, rack power requirements are exceeding 20 kilowatts (kW) in a growing number of facilities, and many organizations are now looking to deploy racks with requirements of 50 kW or more.

Air cooling systems have continually evolved to address higher densities with greater efficiency, but there is a point at which air simply does not have the thermal transfer properties required to provide sufficient cooling to high-density racks in an efficient manner. This can reduce the performance and reliability of specialized servers and becomes less energy efficient as rack power increases. As more high-power racks are deployed, air cooling becomes untenable from an economic and sustainability perspective.

As a result, more organizations are exploring the feasibility of bringing liquid to the rack to increase the capacity and efficiency of data center cooling. Liquid cooling leverages the higher thermal transfer properties of water or other fluids to support efficient and cost-effective cooling of high-density racks.

The approach has been proven in multiple applications, such as mainframes and gaming computers, but is not commonly used for rack-mounted servers. That is changing quickly. As high-density racks proliferate, liquid cooling is moving out of its niche in high-performance computing (HPC) centers and into the mainstream. Vertiv expects the technology to continue to build momentum in the coming years.

Extensive development work has already been conducted across data center equipment and infrastructure suppliers to support liquid cooling, and multiple solutions are available today to cool servers through rear door heat exchangers, conductive cold plates, and immersion. The approach that works best for a specific facility will depend on multiple factors. However, data center operators considering liquid cooling share a common challenge: integrating with data center and building infrastructure to support delivery of liquids to and from the rack.

This paper provides an overview of liquid cooling technology with a focus on how data center infrastructure has evolved to support liquid cooling.

The Inevitability of Liquid Cooling

There has long been the expectation that rack power densities would reach levels unsupportable by air cooling. Until recently, those expectations have been overly optimistic, as densities did not rise as quickly as expected. Nevertheless, those expectations triggered significant research and development (R&D) into solutions that could support extremely high rack densities, most notably conductive cold plate and immersive liquid cooling.

Now, a convergence of trends is driving rack power consumption to the levels previously predicted across a significant segment of the data center industry.

At the heart of rising rack densities is newer generation central processing units (CPUs) and graphics processing units (GPUs), which have thermal power densities much higher than previous generation architectures. After years of relatively stable power densities, Intel's Cannon Lake CPUs, for example, had thermal power densities double that of the previous generation CPUs introduced just two years earlier.

Simultaneously, server manufacturers are packing more CPUs and GPUs into each rack unit (U). With multiple high-performance servers in a rack, systems that deliver cooling air to racks are unable to provide adequate cooling capacity, even with containment. In addition, the strategy of spreading compute loads out is not feasible in processing-intensive applications because of the latency challenges created by physical distance that exist even within a single server. As a result, components are being compacted within devices, creating ultra-dense 1U servers that are driving rack densities to unprecedented levels.

The underlying trend driving these technology advances is the expanded use of artificial intelligence and HPC beyond their traditional use in science. These technologies are now being deployed in data centers supporting cloud-based HPC, finance, online gaming, healthcare, film editing, animation, and media streaming. As a result, high-density equipment racks are moving out of niche applications and becoming more mainstream, forcing thermal management systems to evolve to meet new requirements.

This is impacting data center design on multiple fronts. The first is new data centers being designed to rely exclusively on liquid cooling, creating smaller and more efficient data centers with massive compute capabilities. The second is data centers designed with air cooling, but which also include liquid cooling infrastructure to simplify the future transition. The third, and by far the most common, is data center operators integrating liquid cooling into existing air-cooled facilities, often transitioning some of the capacity of air systems to liquid. Finally, liquid cooling is emerging as a viable alternative for processing-intensive edge computing sites.

While each of these scenarios will utilize similar solutions, they also present distinct challenges that are best addressed by working closely with an infrastructure partner with deep domain expertise in data center thermal management and a broad portfolio of solutions for air and liquid cooling.

Legacy of Innovation

As a global leader in thermal management, Vertiv brings a holistic approach to liquid cooling supported by a legacy of innovation and a broad offering of solutions that simplify and accelerate the transition to liquid cooling. Our offering is the result of years of R&D and has been shaped by collaboration with leading universities, involvement in industry associations, and partnerships with emerging technology providers, including:

- Working with the Center for Energy-Smart Electronic Systems (ES2) partner universities to advance data center efficiency through liquid cooling and other methods. This type of partnership can assist in developing compatible materials; further our understanding of deployment, fluid hygiene filtration and containment methods; and optimizing controls for liquid cooling systems.
- Active participation in The Green Grid and Open Compute Project to help develop best practices for system design and deployment, and to ensure interoperability of liquid cooling system components.
- Working with Green Revolution Cooling on the introduction of a new line of immersion cooling tanks.

Through these efforts, and our liquid cooling R&D program, Vertiv is keeping pace with changing customer requirements to deliver a portfolio that supports hybrid air and liquid cooling as well as fully liquid-cooled data centers, including:

- Coolant distribution units (CDUs) and indoor chillers engineered to provide complete infrastructure solutions for data center liquid cooling
- Active and passive rear-door heat exchangers
- Innovative and efficient immersion cooling systems
- Heat rejection systems designed to work with liquid cooling CDUs and chillers
- Retrofit solutions that enable air cooling equipment to be modified to support liquid cooling
- Established practices and services for commissioning, startup, and operation of liquid cooling infrastructure

As market demand for liquid cooling grows, Vertiv is continuing to expand our capabilities, leveraging our scale and thermal management expertise to develop new solutions and services.

How Thermal Management Systems Are Evolving

Data center cooling technologies have always had to evolve to support the changing needs of the IT systems they support. In previous years, air-cooling systems have adapted to higher densities by moving cooling closer to the source of heat and employing containment. But these approaches deliver diminishing returns as rack densities rise above 20 kW (Figure 1). A variety of liquid-cooling technologies have emerged to meet the cooling requirements of high-density racks.

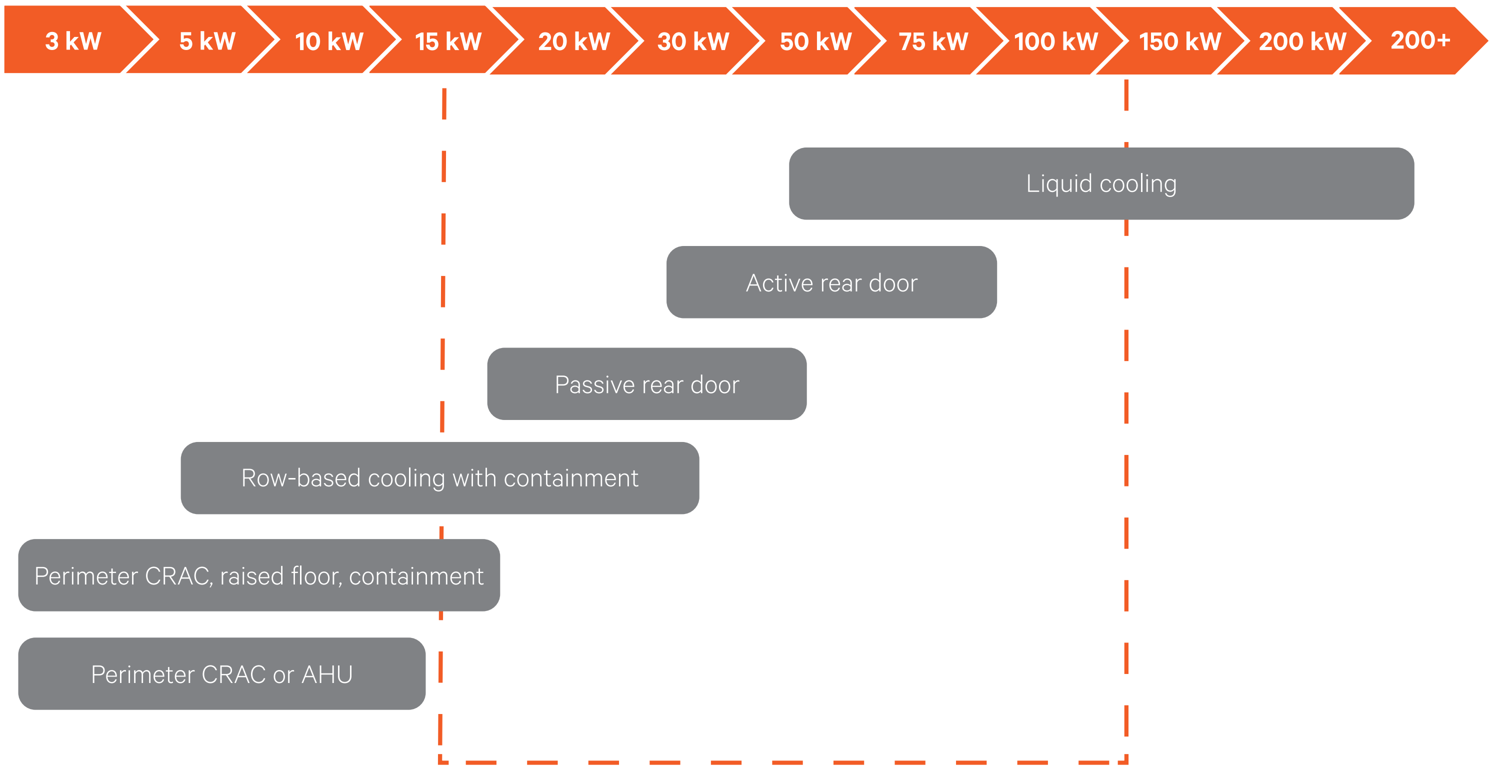

Figure 1. Air-based cooling systems lose their effectiveness when rack densities exceed 20 kW, at which point liquid cooling becomes the viable approach

Rear-door Heat Exchangers

Rear-door heat exchangers are a mature technology that provide a viable solution for managing densities above 20 kW. While these technologies don't bring liquid directly to the server, they do utilize the high thermal transfer properties of liquid and employ similar data center infrastructure as that required for direct liquid cooling.

Passive or active heat exchangers (Figure 2) replace the rear door of the IT equipment rack with a liquid heat exchanger. With a passive design, server fans expel heated air through a liquid-filled coil mounted in place of the rear door of the rack. The coil absorbs the heat before the air passes into the data center. Active heat exchangers include fans to pull air through the coils and remove heat from even higher density racks.

These systems can form the foundation for a hybrid approach to data center cooling in which liquid and air-cooling systems work together to cool environments with mixed rack densities. The next phase in this evolution is the move to direct liquid cooling. The two most common approaches to direct liquid cooling are direct-to-chip cold plates and immersion cooling.

Direct-to-Chip Liquid Cooling

Direct-to-chip cold plates sit atop the board's heat-generating components (CPUs, GPUs, memory modules) to draw off heat through single-phase cold plates or two-phase evaporation units.

Figure 2. Active and passive rear-door heat exchangers

Single-phase cold plates use a cooling fluid looped into the cold plate by a CDU to absorb heat from server components. The heated fluid is transferred outside the rack for heat rejection, which is typically performed by the CDU.

Fluid selection is determined by balancing the thermal capture properties and viscosity of the fluid. Water delivers the highest heat capture capacity, but is often mixed with glycol, which reduces heat capture but increases viscosity to enhance pumping efficiency. These systems can also use dielectric fluid to mitigate the damage from a leak; however, the dielectric fluid has a lower thermal transport capacity than the water/glycol mixture. Figure 3 compares the heat capture capacity of various mediums.

With two-phase cold plates, a low-pressure dielectric liquid flows into evaporators, and the heat generated by server components boils the fluid. The heat is released from the evaporator as vapor, which is transferred outside the rack for heat rejection.

Medium Type	Specific Heat (J/kgK)	Volume/kg	Joules/litre
Water (reference only)	4182	1L	4182
Hydrocarbon	2300	1.24L	1854.8
Fluorocarbon	1300	0.71L	1831.0
Air	1000	773.46L	1.3

Figure 3. Relative heat capture capacities by medium
Source: OCP's *Design Guidelines for Immersion-Cooled IT Equipment*

Direct-to-chip cooling technologies generally have higher heat removal capacities than rear-door heat exchangers, and a number of companies offer cold plate technology that can be integrated with existing and new servers. This approach, however, does not remove 100% of the heat from equipment in the rack. Typically, direct-to-chip cooling can remove 70-75% of the heat generated by the equipment in the rack, requiring a hybrid approach to cooling. This is because the cold plate works best when covering a flat surface such as the CPUs, GPUs, and memory modules and is not applied to other components such as power supplies and IC capacitors. Organizations deploying direct-to-chip liquid cooling should work with their infrastructure partner to determine whether some air-cooling systems can be retrofitted to support direct-to-chip cooling.

Immersion Cooling

With immersion cooling, servers and other components in the rack are submerged in a thermally conductive dielectric liquid or fluid. With this method, the need for air cooling is eliminated, including the fans within servers. This approach maximizes the thermal transfer properties of liquid and is the most energy efficient form of liquid cooling.

As with direct-to-chip cooling, single-phase and two-phase immersion cooling systems are available. In a single-phase system (Figure 4), servers are installed vertically in a dielectric fluid that is thermally conductive. Heat is transferred to the coolant through direct contact with server components and removed by heat exchangers in a CDU. In data center applications, the CDU will typically be a separate component located in proximity to the immersion tank or at the perimeter of the data center. However, “micro” immersion tanks are also available that integrate the CDU into the tank, providing a complete, self-contained cooling solution for high-density edge applications.

In two-phase immersion cooling, servers are immersed in a bath of dielectric fluid engineered to have a boiling point low enough to produce an adequate IT case temperature but high enough to reduce energy consumption while moving the heat outdoors. Heat from the servers boils the surrounding fluid, changing its phase. The rising vapor is condensed back to liquid by coils located at the top of the tank and falls by gravity back into the liquid tank.

Because the fluids used in these systems are very expensive and have environmental, health and safety concerns, it’s important to design two-phase immersion systems to minimize vapor loss. One approach is to seal the immersion tank to prevent vapor from escaping.

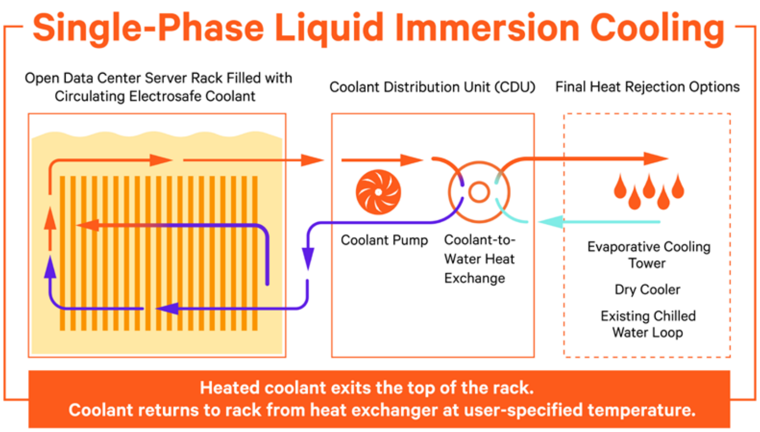

Figure 4. Coolant circulation and heat removal in a single-phase immersion cooling system

This method effectively produces de minimis vapor loss; however, access to the enclosed equipment requires liquid handling methods as outlined in Section 608 of the Federal Clean Air Act, complicating server maintenance.

The other approach is atmospheric pressure control utilizing breathing to vary the amount of evaporated vapor as the IT load increases or decreases. A bellows is used in conjunction with a low temperature condenser to condense the vapor and activated carbon to adsorb/desorb the vapor during the exhalation/inhalations. This approach provides easier access to servers in the bath, but adds complexity and can only minimize vapor loss.

Liquid Cooling Infrastructure

Whether rack-mounted equipment is being cooled indirectly through rear-door heat exchangers or directly through conductive or immersive technologies, dedicated infrastructure is required to create a fluid cooling loop that enables heat transfer between facility and secondary circuits and fluid other than facility water to be used for cooling.

Properly configuring the cooling infrastructure is key to making a successful transition to liquid cooling. When configuring a dedicated loop to support liquid cooling, consideration should be given to the ability of the infrastructure systems to ensure precise temperature control of the liquid being used for cooling, as well as the ability to respond to the sudden load increases common in HPC.

Dedicated liquid cooling infrastructure should also be designed to minimize fluid volume to reduce the consequences of a leak and to mitigate risks from facility pressure. The following are components of dedicated liquid cooling infrastructure.

Coolant Distribution Unit (CDU)

The CDU provides controlled cooling liquid for heat exchangers, direct-to-chip, and immersion cooling. The CDU creates an isolated secondary loop that is separate from the chilled water supply, enabling strict containment and precise control of the liquid cooling system. Normally, it maintains liquid cooling supply temperature above the data center dew point to prevent condensation and maximize sensible cooling. Heat from the rack is then removed by either a liquid-to-air or liquid-to-liquid heat exchanger within the CDU. CDUs are available for use in row or on the perimeter of the data center. A range of capacities are also available, with some models delivering capacities exceeding 1 MW to support larger deployments with a minimal number of CDUs.

When deploying indirect or direct liquid cooling in a hybrid environment, the CDU provides a complete solution for localized coolant delivery and heat removal. When supporting direct liquid cooling in a fully liquid-cooled environment, the CDU is matched with a heat rejection system. When used with rear door cooling, the use of a chiller is typical for delivering lower temperatures to the rear door units.

Indoor Chiller

For applications that do not have access to chilled water in the data center, or prefer not to tap into the existing chilled water system, a modular indoor chiller provides efficient and reliable support for liquid cooling. Variable speed pumps allow the flow of refrigerant to adapt to changing loads and internal controls maintain the leaving water temperature by controlling the speed of the pump. Indoor chillers designed to support liquid cooling use the same footprint as perimeter cooling units to simplify retrofits and future proof new data center designs.

Immersion Tank

The immersion tank houses vertically mounted servers in a dielectric bath and, working with a CDU, circulates cooling fluid to remove heat.

Heat Rejection

CDUs with liquid-to-liquid heat exchangers can be connected to the building's chilled water system for heat rejection. Where no building system is available, indoor chillers as described previously can be employed. It may be possible to use existing drycoolers or cooling towers for final heat rejection, but these systems often require modification to support liquid cooling. For example, depending on the location of the facility, the drycooler may require an adiabatic assist to maintain the lower supply temperatures required for liquid cooling throughout the year. Your infrastructure partner can help determine what changes are required to existing heat rejection systems

Benefits of Liquid Cooling

Facilities looking to deploy extremely high-density racks (>30 kW) will have little choice whether or not to use liquid cooling. No matter how the system is configured or optimized, air cooling will simply be unable to deliver the heat removal capacity required to maintain the reliability of IT systems. This will be true on the edge as well as in core data centers.

In cases where rack densities are gradually creeping up to the threshold at which liquid cooling becomes a necessity, facility operators will have to weigh the benefits that can be achieved by moving to liquid cooling against the costs and the risks associated with pushing air cooling systems beyond the point where they can effectively manage heat.

Some operators may also have concerns about bringing liquid to the IT rack. Today's liquid cooling infrastructure is designed to minimize both the likelihood and consequences of fluid leaks on data center operation. Discuss your concerns with your infrastructure provider to get a better understanding of the leak mitigation features specific systems employ.

Liquid cooling may not be right for every facility, but where appropriate it can deliver the following benefits.

Improved IT Reliability and Performance

A liquid cooling system will not only enable the desired reliability, but also deliver IT performance benefits. As CPU case temperatures approach the maximum safe operating temperature, as is likely to occur with air cooling, CPU performance is throttled back to avoid thermal runaway. Liquid cooling allows densely packed systems to operate continuously at their maximum voltage and clock frequency while avoiding overheating.

Improved Energy Efficiency

The higher thermal transfer properties of liquid compared to air, combined with the elimination of fans required to move air across the data center and through servers, can create significant energy savings in liquid-cooled data centers. The pumps required for liquid cooling consume less power than the fans needed to accomplish the same cooling.

Sustainability

Not only does liquid cooling create opportunities to reduce data center energy consumption and drive power usage effectiveness (PUE) down to near 1.0, it provides a more effective approach for re-purposing captured heat to reduce the demand on building heating systems. The return-water temperature from the systems can be 140 degrees Fahrenheit (60 C) or higher and the liquid-to-liquid heat transfer is more efficient than is possible with air-based systems.

Maximize Space Utilization

The density enabled by liquid cooling allows a facility to better use existing data center space, eliminating the need for expansions or new construction, or to build smaller-footprint facilities. It also enables processing-intensive edge applications to be supported where physical space is limited.

Lower Cost of Ownership

The American Society of Heating, Regrigerating and Air-Conditioning Engineers (ASHRAE) conducted a detailed cost of ownership analysis of air-cooled data centers versus a hybrid model air- and liquid-cooled data center and found that, while a number of variables can influence total cost of ownership (TCO), "liquid cooling creates the possibility for improved TCO through higher density, increased use of free cooling, improved performance and improved performance per watt."

Conclusion

Supporting the applications that a growing number of businesses are now using to create competitive advantage requires dense concentrations of processing power that exceed the physical limits of air cooling.

Liquid cooling offers a heat transfer capacity that is several orders of magnitude higher than air cooling and will increasingly be required to support processing-intensive applications at the core and the edge. Complete infrastructure solutions are available today for safe and effective support of liquid cooling technologies, enabling both the design of fully liquid-cooled facilities and the successful introduction of liquid cooling into existing air-cooled data centers.

Contact your local Vertiv representative for more information on liquid cooling solutions for your data center.

Vertiv.com | Vertiv Headquarters, 1050 Dearborn Drive, Columbus, OH, 43085, USA

SL-70807 (R06/21)