

The logo for AHEAD, featuring the word "AHEAD" in a white, sans-serif font. The letter "E" is replaced by a blue horizontal bar with a white outline, and the letter "A" is also in blue with a white outline.

AHEAD

The Need for *Liquid Cooling* in Modern Computing

How AHEAD is helping organizations adapt their infrastructure for the density demands of the future with direct-to-chip liquid cooling

Computing demands continue to rise as organizations deploy AI and other advanced workloads. This means data centers are managing increasingly dense configurations with high-performance processors, GPUs, and other components. However, traditional air-cooling methods struggle to keep pace with the thermal requirements of these modern, high-density data centers.

Although [liquid cooling has been around for a long time](#), it has recently become a standard technology for high-density data centers. Enterprises and data center operators supporting modern workloads are now investing in liquid cooling as a way to manage high-density racks while boosting performance.

In this whitepaper, we'll discuss the trends driving the adoption of liquid cooling, the benefits of direct-to-chip solutions in particular, and how AHEAD helps organizations implement liquid cooling using a turnkey approach.





Trends Driving Liquid Cooling Innovation

Liquid cooling technology innovations are being driven by many factors, including the increased deployment of workloads like AI and high-performance computing (HPC), the higher thermal requirements of cutting-edge chipsets, and the need to maximize existing resources while modernizing data center infrastructure.

Modern Workloads

Many enterprises are now deploying applications and workloads with much higher processing and power requirements. For example, HPC workloads that simulate or model complex processes such as financial modeling, genomics, and drug discovery require enormous computational power in ranges of petaflops to exaflops.

In addition, newer AI models require massive computational power for training, often running on GPUs with thousands of cores. Serving these AI models and implementing real-time inferencing also requires efficiency and high-capacity CPUs and GPUs to handle tasks quickly.

The size and complexity of AI workloads is also increasing dramatically, as breakthroughs related to generative AI and large language models (LLMs) require even more computing resources than traditional AI. In fact, newer generative AI models are pre-trained on enormous amounts of data – estimated at over a petabyte in the case of OpenAI’s GPT-4 model – which requires incredibly powerful hardware and supporting infrastructure.

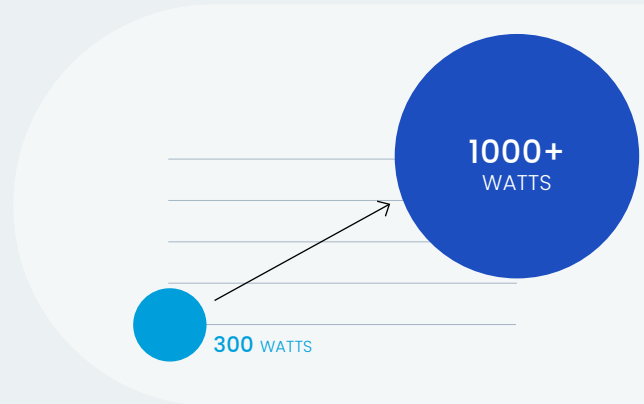
Thermal Design

The increasing compute density of the latest components is leading to greater thermal design power (TDP), which is a measure of the maximum potential heat generated.

GPUs

In only five years, the TDP of a typical GPU has increased from 300 watts to over 1000 watts for newer superchips like NVIDIA Blackwell.

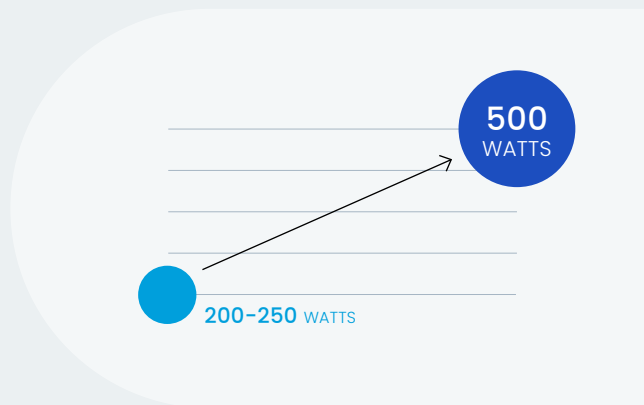
- **Performance Enhancements:** The push for higher computational power, particularly for AI and machine learning applications, has led to GPUs with increased power consumption.
- **Architectural Advancements:** While newer architectures offer improved efficiency, the emphasis on maximizing performance also results in higher overall TDPs.
- **Increased Core Counts and Clock Speeds:** Modern GPUs feature more cores and higher clock speeds, contributing to greater heat generation and higher TDPs.



CPUs

The maximum TDP of CPUs has also doubled, from 200–250 watts in 2019 to as much as 500 watts today.

- **Higher Core Counts:** Modern server CPUs incorporate more cores to handle parallel processing tasks, leading to increased power consumption and higher TDPs.
- **Enhanced Performance:** To meet the demands of data-intensive applications, CPUs are designed with higher clock speeds and improved architectures, which contribute to greater heat generation.
- **Advanced Manufacturing Processes:** While newer fabrication technologies aim to improve efficiency, the push for maximum performance often results in higher power densities and increased TDPs.



The trend towards more powerful components with higher TDPs will only continue into the future. While traditional air-cooling methods have sufficed for the density demands of the past, if an organization is looking to expand its capabilities, efficient thermal design of power-hungry systems is paramount.

Resource Maximization

As AI and HPC workloads lead to increased power demands, companies are faced with the challenge of leveraging their current infrastructure and data center footprints while also adapting as efficiently as possible in today's constantly advancing technology landscape.

Many data centers report average rack densities of around 12kW per rack, with power density topping out at 15-20kW per rack in well-designed facilities utilizing traditional air cooling. However, modern GPU based systems operating at 10-14+kW per system would only be able to utilize 25% or less of available rack space due to these air-cooling constraints.

This means data centers will need to rethink their designs and adapt to accommodate greater hardware density. At the same time, organizations need to align investments in liquid cooling and newer chipsets with existing hardware refresh cycles to maximize ROI.



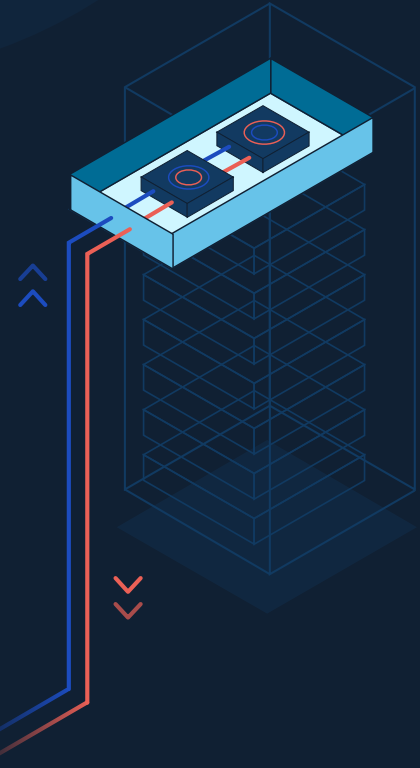
The Benefits of Direct-to-Chip Liquid Cooling

Here are some of the key benefits of liquid cooling technologies, with a focus on direct-to-chip solutions.

Increased Cooling Efficiency and Performance

Liquid cooling is the process of using liquids like water, specialized coolants, or dielectric fluids to absorb and dissipate heat generated by electronic components. The higher thermal conductivity and heat capacity of liquid compared to air make it a superior choice for transferring heat away from hot components.

Direct-to-chip and other liquid cooling solutions minimize the risk of overheating and help maintain lower operating temperatures, which is crucial for the reliability and longevity of hardware components. These solutions enable up to a 15% performance increase and up to 35% greater energy efficiency than traditional air cooling.

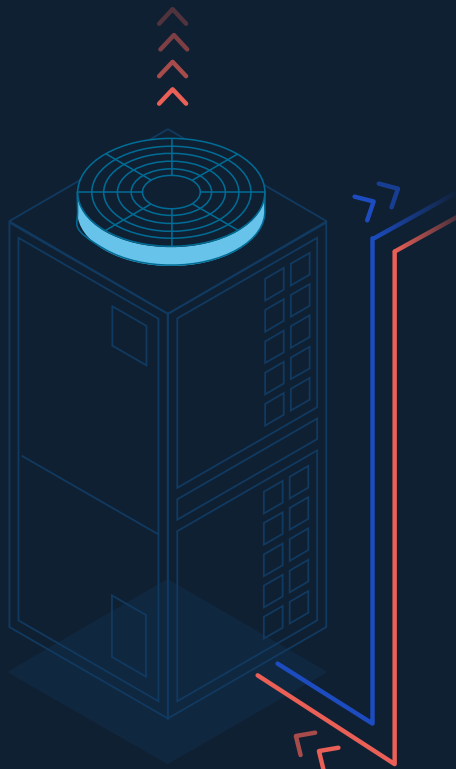


Support for High-Density, High-Performance Workloads

While retrofitting an entire data center with liquid cooling can be cost prohibitive, a hybrid approach with liquid and air cooling together is a more affordable and straightforward alternative. Smaller-scale liquid cooling options like server level direct-to-chip designs and rack level immersion cooling solutions do not require expensive data center infrastructure upgrades.

More specifically, direct-to-chip liquid cooling solutions — where coolant flows through pipes to small cold plates attached to CPUs and GPUs — can support extremely dense server configurations. This makes it ideal for advanced AI processing, machine learning, and large-scale analytics workloads, which might have thermal limitations with traditional air cooling alone.

As technology advances and chip architectures continue to evolve toward greater processing capabilities, liquid cooling systems are critical for handling the increased heat output. Direct-to-chip coolant distribution units (CDUs) are offered in many sizes and can scale with the compute needs of the data center.



Energy Savings and Cost Reduction

Since liquid cooling has a higher thermal conductivity and heat capacity compared to traditional air-cooling, less power is needed to keep systems at safe operating temperatures. This can lead to both energy savings and reduced greenhouse gas emissions for data center operations.

Some liquid cooling solutions allow the captured heat to be reused, either for heating buildings or other industrial processes. This approach transforms waste into a valuable resource, further improving the efficiency and sustainability profile of the data center. For these reasons, many large companies with sustainability commitments now see liquid cooling as a path to achieving their environmental goals.

AHEAD HATCH® STREAMLINES INFRASTRUCTURE MANAGEMENT

[AHEAD Hatch IT Lifecycle Management Platform](#) can help you track your hardware assets from procurement and manufacturing through to deployment and support. Hatch also provides consolidated data at the site level for your infrastructure and can easily integrate with ServiceNow® and other leading enterprise platforms to share data in real time.

Hatch can manage thousands of devices and components, including their firmware levels, BIOS, software image revisions, harvested hardware data, warranties and service contracts, and more. Real time data with Hatch streamlines order processing, inventory management, logistics, and asset management. This accelerates every aspect of planning, executing, and supporting your large-scale infrastructure initiatives.



Adopting Liquid Cooling with AHEAD

As you can see, liquid cooling offers a more efficient and sustainable alternative to traditional air cooling strategies. Organizations can integrate direct-to-chip liquid cooling into their servers and racks to optimize their infrastructure for the demands of modern, high-performance computing environments.

Although direct-to-chip liquid cooling solutions are much easier to deploy than large scale alternatives, it's still helpful to work with an experienced partner. AHEAD is an end-to-end infrastructure solution provider that can help your organization implement [liquid cooling](#). In fact, we operate a state-of-the-art facility alongside [AHEAD Foundry™](#) manufacturing, warehousing, and global logistics facilities, designed specifically for the complexities of direct-to-chip liquid cooled racks.

AHEAD Foundry provides end-to-end services for building and integrating hardware and networking infrastructure at scale in our facilities rather than on-site. This means we can help you design and deploy new systems or modernize your existing data centers using a plug-and-play approach. AHEAD Professional Services and Managed Services teams can get your large-scale infrastructure projects across the finish line and maintain them throughout their lifecycle.

[Contact AHEAD](#) to learn more about our liquid cooling rack integration and data center modernization services.



AHEAD

Combining cloud-native capabilities in software and data engineering with an unparalleled track record of modernizing infrastructure, we're uniquely positioned to help accelerate the promise of digital transformation.

Visit us at ahead.com.

National Hubs

CHICAGO

444 W. Lake Street
Suite 3000
Chicago, IL 60606

NEW YORK

500 5th Avenue
Floor 17
New York NY 10010

ATLANTA

1117 Perimeter Center
W406
Atlanta, GA 30338

SAN FRANCISCO

2000 Crow Canyon Place
Suite 250
San Ramon, CA 94583