

Application Note

High-Density Data Center Cooling - The Changes & Challenges



Table of Contents

Introduction	3
The Cooling Challenges	3
Liquid Cooling to The Rescue	4
Pumps - The Heart of the Cooling System	4
So Who Decides?	5
What Really Matters?	5
Reliability	5
Not Off-the-Shelf; Made in the USA & Configured to Perform	5
Conclusion	6
About Tark Thermal Solutions	6
Contact Tark Thermal Solutions	6

Introduction

High-performance computing (HPC) has emerged as a significant catalyst for the growth of high-density data centers, (HDDC) creating the requirement for advanced facilities to handle the surge in data computation and speed required for Artificial Intelligence, Machine Learning, and Large Language Model processing. High-density data centers offer unparalleled performance and efficiency in compact spaces; however, they face a critical challenge of managing the intense heat generated by densely packed servers.

As computing demands continue to rise traditional cooling methods often fall short. This application note explores the unique cooling challenges facing high-density data centers and introduces advanced thermal management solutions designed to maintain optimal operating conditions, prevent hardware failures, and maximize energy efficiency in these cutting-edge facilities.



The Cooling Challenges

Data centers face several significant cooling challenges due to increasing computing power, higher densities, and sustainability concerns. Modern data centers house high-performance servers and GPUs, generating significant heat. Efficiently cooling these dense environments is a growing challenge.

Traditional air cooling systems are reaching their physical limits at server rack densities of about 70 kilowatts. The rapidly increasing server rack densities are now the benchmark for state-of-the-art AI data center facilities. Many data center professionals are dissatisfied with the performance of their current cooling solutions, with 35% regularly having to make adjustments due to insufficient cooling capacity. Uneven airflow and poor airflow management can create localized hot spots, leading to hardware failures and inefficiencies.

High energy consumption is another problem with traditional air-cooled systems. Data centers are projected to consume approximately 2.5-4.0% of global electricity by 2026, creating an urgent need to enhance energy efficiency. Heat management for AI/ML workloads is putting increased strain on data center cooling systems, requiring more effective solutions to handle higher thermal loads.

Cooling failures lead to catastrophic downtime. With reported cost estimates of up to \$10,000/minute of downtime per rack, it is imperative that backup cooling systems ensure maximum uptime while maintaining cost efficiency.

Water-hungry evaporative cooling systems, common in warmer climates, have raised environmental and water usage concerns about the impact of data centers on local water resources.

Edge and modular data centers are smaller and more remote, often lacking the space for traditional cooling solutions. These DCs require innovative compact cooling techniques.

While liquid cooling (immersion or direct-to-chip) is more efficient, retrofitting existing data centers with this technology can be complex and costly.

Liquid Cooling to The Rescue

To address these challenges, data centers are exploring several strategies and advanced cooling methods, including liquid cooling, AI-driven cooling optimization, free air cooling (using outside ambient air), phase-change cooling, and renewable energy-powered cooling solutions.

Currently, liquid cooling, particularly two-phase and direct-to-chip systems, is emerging as a key solution for managing high-density racks and AI workloads. These systems can reduce energy consumption by up to 30% or more, aligning with sustainability goals.

Direct-to-Chip (D2C) cooling via centrifugal water pumps is a liquid cooling method used for 50 – 200 kW densities in HPC infrastructure of high-density data centers. It involves circulating a liquid coolant directly to a cold plate that is in contact with a computer chip (CPU, GPU, FPGA) then channeling the heat through a liquid coolant network to a rack-mounted Cooling Distribution Unit (CDU) for more efficient dissipation than traditional air cooling. The CDU removes the incoming heat and uses a pump to circulate the chilled coolant back into the closed-loop system. **Water-based pumps are the most popular use cases (80%) in the data center market with regenerative turbine refrigerant pumps being the secondary choice.**

Two-phase cooling is an advanced thermal management technique where the cooling liquid absorbs heat and undergoes a phase change boiling from a liquid to vapor. The vapor rises and reaches a heat exchanger, cools down and condenses back into a liquid and repeats the cycle. Immersion cooling and two-phase cold plate cooling are common types of solutions. Although considered to be the most efficient cooling method, these systems come with high initial costs, require expensive specialized coolants and a custom tank infrastructure which makes retrofitting existing data centers costly.

As the industry continues to contend with these cooling challenges, the focus remains on developing more efficient, sustainable, and scalable cooling solutions to support the growing demands of modern data centers.

For easier integration and high-performance workloads, D2C cooling is a more practical solution than two-phase cooling or immersion cooling, offering the lowest total cost of ownership. It delivers targeted cooling directly to the heat-generating components, avoiding excess energy waste. It is easier to integrate with existing infrastructure and is adaptable to traditional air-cooled rack-mounted server designs. Since it uses a closed-loop system, D2C cooling reduces the need for fluid replacement and maintenance, while also reducing the reliance on air conditioning. Some airflow is still needed for cooling other components; however, overall, D2C cooling is better for High-Performance Computing (HPC) and is well-suited for AI, ML, and GPU-intensive workloads.

Pumps – The Heart of the Cooling System

The number of pumps used to cool a typical data center varies based on factors like the data center's size, layout, and cooling requirements. Multiple pump types are typically employed in a data center cooling system, including coolant pumps for circulating refrigerant or cooling fluid, water pumps for systems using water-based cooling, and specialized liquid cooling pumps for direct-to-chip or immersion cooling systems.

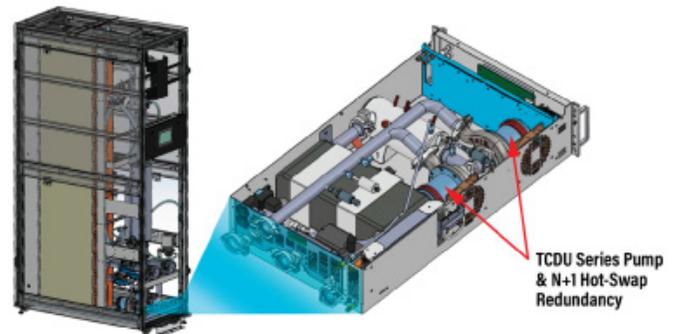


Regenerative Turbine
Refrigerant Pump



Centrifugal Water Pump

Liquid cooling systems may have dedicated pumps for each rack or row of servers. Redundancy is a critical factor in data center design where 24/7/365 operation is a must. There are often backup hot-swappable pumps installed to ensure continuous operation in case of failure or scheduled preventative maintenance.



There are three key components in a pump that affect the overall performance, and influences the pump manufacturer's ability to meet the data center specifications and deliver product in a timely manner.

1. The motor makes up 60-80% (new internal data from quoted motor suppliers) of the total pump cost. It is the heart of the pump and defines the achievable flow in conjunction with the impeller diameter. As the flow needs to increase in each application, the motor provides the power needed to reach the desired performance.
2. The impeller diameter determines what is needed to meet the required flow.
3. The 3rd key component is the motor controller which controls several critical functions to ensure optimal cooling efficiency, reduce energy consumption and prevent equipment failures.

Motor controllers in HDDC cooling pumps are used to monitor and adjust the motor's speed control, flow rates, energy consumption, voltage and current, temperature, pressure, diagnostics and fault detection

A common communication protocol for motor control applications in high density DCs is Modbus RS485. Modbus RS485 provides excellent signal noise immunity making it ideal for harsh industrial environments like data centers and it supports bus lengths up to 1200 meters (4000 feet) without additional hardware, allowing for extensive coverage in large data center facilities.

So Who Decides?

When it comes to pumps, there are no universal data center or market standards. Pumps are specified by the OEM based on market considerations from DC operators, integrators, and hyperscalers. Key titles of interest influencing the cooling system and type of pump specified include the principal, design engineering, R&D, manufacturing, test engineers, quality engineers, product managers, sourcing and supply chain managers and any Director, VP, or C-Suite stakeholders.

What Really Matters?

Reliability, efficiency, sustainability and energy management are key focal points for the DC market. There is a DC goal to reach a Power Usage Effectiveness (PUE) of 1.02. The DC market is currently operating between 1.1 - 1.2. PUE is a metric that measures the energy efficiency of DCs. Introduced in 2007, it has become the global standard for assessing DC energy consumption.

PUE formula = Total Facility Energy / IT Equipment Energy

On average, 40% of all DC operating expenses can be attributed to cooling and ventilation systems. Pumps contribute significantly to the 40% of DC energy consumption as they typically generate 100-500 watts of heat energy.

Reliability

Data center cooling pumps are required to perform 24/7/365 and are mission critical for continuous operation. DCs allow 5-7 minutes annually to hot-swap a pump. Depending on the DC design, there may be **one pump per cabinet, but a cabinet may have multiple servers per rack. N+1 redundancy is required** as uptime is mission critical. A comprehensive range of pumping solutions, such as Tark Thermal Solutions' TCDU Series of water pumps can provide superior performance for Coolant Distribution Units (CDUs), Economizers, and Immersion Cooling systems.

Designed, assembled, and tested to meet the rigorous demands of modern data centers, these pumps ensure precision performance, reliability, and energy efficiency.

Tark Thermal Solutions' four base TCDU models of water pumps leverage a hermetically sealed pump combined with a motor controller for precision cooling in water-based direct-to-chip applications. TCDU Series cooling pumps offer high efficiency, extended life, and quiet operation with the added advantage of precise speed controls to deliver variable flow and energy savings to data center operations.

Centrifugal pumps are ideal for moving large volumes of coolant with low-to-medium pressure requirements. **TCDU Series Cooling Pumps**

These pumps excel in CDU and economizer applications where steady flow is critical. The pump's cavitation prevention feature is a critical point for maintaining performance and continual operation.

- Features: High flow rates, smooth operation, low maintenance
- Applications: Economizers, CDUs, and large-scale cooling distribution
- Benefits: Cost-effective, energy-efficient, and suitable for various cooling loop configurations



Not Off-the-Shelf; Made in the USA & Configured to Perform

Many cooling pumps used in DCs are off-the-shelf pumps with some not being UL Listed. This has created unintended consequences for OEM CDU manufacturers and data center operators. **Cavitation and corrosion issues on the rotor and impeller have created unanticipated quality issues, resulting in pumps not lasting as long as specified.**

The TCDU Series of pumps has a series of core components and a streamlined design as the basis, with the option for configured-to-order inlet, outlets, and control modules. The four configured-to-order base models can account for more flow or pressure based on the selected motor or impeller. Tark Thermal Solutions can respond quickly with unique inlet or outlet connections, communication protocols, motor controllers, and flag or terminal connections to meet specific requirements.

Tark Thermal Solutions's pumps for DCs are UL-Listed, CE Certified for global sales and EMC-certified for motor controls and communication protocols. **These pumps are made in Dayton, OH and provide a reliable source of supply.**

Conclusion

Many modern high-density data centers, especially those handling AI/ML, HPC, and cloud services, need advanced cooling solutions that support 50 kW to 200 kW per rack with 24/7/365 performance. As data centers continue to scale, cooling efficiency remains crucial to reducing energy consumption, minimizing operational costs, maximizing uptime and ensuring long-term reliability. The trend towards more efficient and diverse cooling solutions suggests that pump configurations in data centers will continue to evolve.

Learn more about Tark Thermal Solutions' TCDU water pumps for high-density DCs and other pump solutions for data center applications.

<https://tark-solutions.com/products/pumps>

About Tark Thermal Solutions

Tark Thermal Solutions develops thermal management solutions for demanding applications across global medical, industrial, transportation, and telecommunications markets. We manufacture one of the most diverse product portfolios in the industry, ranging from active thermoelectric coolers and assemblies to temperature controllers and liquid cooling systems. Our engineers use advanced thermal modeling and management techniques to solve complex heat and temperature control problems. By offering a broad range of design, prototyping, and in-house testing capabilities, we partner closely with our customers across the entire product development lifecycle to reduce risk and accelerate their time-to-market. Our global manufacturing and support resources help customers maximize productivity, uptime, performance and product quality. Tark Thermal Solutions is the optimum choice for standard or custom thermal solutions.

Have a question or need more information about us?
Please contact us via tark-solutions.com