

REDCHALK
GROUP

THE DATA CENTER BOOM: EVOLUTION OF COOLING TECHNOLOGIES

+ A STRATEGIC GUIDE TO THE
THERMAL MANAGEMENT
REVOLUTION TRANSFORMING DATA
CENTERS

VOLUME 2

2025
NOVEMBER

AUTHORS

UMAIR HUSSAIN

PARTNER

UHUSSAIN@REDCHALK.COM

DUKE OLGES

ANALYST

DOLGES@REDCHALK.COM

The Evolution of Data Center Cooling: From Air to Immersion

A Strategic Guide to the Thermal Management Revolution Transforming Data Centers

Executive Summary

Artificial intelligence and machine learning workloads are proliferating at unprecedented rates, reshaping data centers in ways the industry has never seen before. The computational demands of AI generate heat loads that traditional air cooling simply can't handle - pushing the technology past its physical limits and driving a fundamental shift toward liquid cooling solutions. This transition creates both challenges and opportunities across the entire data center infrastructure ecosystem, from industrial automation and power infrastructure providers to HVAC manufacturers, building systems integrators, and equipment suppliers.

This white paper traces the complete thermal management evolution from traditional air cooling through direct-to-chip liquid cooling to immersion cooling systems. For companies with expertise in industrial controls, automation, and thermal management, understanding this progression is critical for positioning capabilities and identifying market opportunities in the rapidly expanding data center sector.

Part I: The Foundation Era - Traditional Air Cooling (1990s-2010s)

The Established Paradigm

For decades, data center thermal management was synonymous with air conditioning. The industry built its foundation on Computer Room Air Conditioning (CRAC) and Computer Room Air Handler (CRAH) systems, technologies that leveraged principles familiar to industrial thermal management experts. These systems represented the logical application of existing HVAC principles to the emerging challenge of cooling electronic equipment.

Technical Architecture

The traditional air cooling approach employed a straightforward methodology. CRAC and CRAH units chilled air before distributing it through raised floors into the "cold aisle" of server racks. The fundamental difference between these systems lay in their heat rejection mechanisms: CRAH units utilized chilled water with control valves, while CRAC units employed refrigerant-based compression systems - technologies well within the expertise of industrial HVAC and thermal management companies.

This approach was enhanced by hot and cold aisle containment strategies, aligning server racks in alternating rows to optimize airflow patterns. Chilled air entered through under-floor plenums and raised floor vents, circulated through electronic components, and exited into hot aisles before returning to air handlers. The simplicity and cost-effectiveness of this methodology made it the leading thermal management approach for data centers.

Performance Capabilities and Limitations

Traditional air cooling systems demonstrated clear thermal density limitations that are critical to understand. Legacy air-cooling techniques without hot/cold aisle containment could support rack thermal

densities up to approximately 15kW per rack. With the incorporation of in-row cooling and aisle containment, operators could extend this capability to roughly 25kW per rack.¹

While effective for conventional IT workloads, these systems exhibited several inherent inefficiencies. Air cooling required significant power consumption for fans and pumps merely to transport air throughout the facility. The low heat capacity of air compared to liquid coolants meant that substantial air volumes were necessary to achieve adequate heat removal. Additionally, the approach often necessitated overcooling due to uneven temperature distribution and insufficient heat dissipation, creating operational expense challenges.

Market Context and Applications

During this era, data centers primarily supported traditional enterprise workloads with relatively modest thermal requirements. Most racks operated at thermal power densities below 15kW, well within the capabilities of air cooling systems.² The technology's simplicity, established supply chains, and familiar maintenance requirements made it the natural choice for data center operators.

This period established the thermal management principles and supplier relationships that continue to influence the industry today. Companies with industrial cooling expertise developed deep knowledge of airflow management, heat exchanger design, and control systems that would prove valuable as the industry evolved toward more sophisticated cooling approaches.

Part II: The Transition Era - The Rise of Direct-to-Chip Liquid Cooling (2015-Present)

The Catalyst for Change

The emergence of artificial intelligence and machine learning workloads fundamentally altered data center thermal requirements. GPU-based servers, essential for AI model training, generate significantly higher heat loads than traditional CPU servers. Current projections indicate AI-based workloads will grow to approximately 13.5-20GW by 2028, accounting for roughly 15-20% of total data center workloads.³

This shift created thermal challenges that exceeded the capabilities of traditional air cooling. Modern AI servers regularly exceed 30-40kW per rack, with some configurations reaching beyond 50kW - far above the 25kW maximum sustainable with enhanced air cooling systems.⁴ The industry faced a critical inflection point: continue with inadequate cooling that limited performance or transition to fundamentally different thermal management approaches.

Direct-to-Chip Cooling: The Bridge Technology

Direct-to-chip (DTC) cooling emerged as the practical solution for high-density AI workloads, representing a hybrid approach that combines liquid and air cooling principles. This technology directly targets the silicon layer within IT equipment - the primary heat source - using liquid coolants with far superior heat capacity compared to air.

Technical Implementation

The DTC approach utilizes cold plates (specialized heat sinks) installed directly on chips, connected to a coolant distribution network. Coolant flows from Coolant Distribution Units (CDUs) through manifolds to individual cold plates, providing precision cooling at the point of heat generation. This targeted approach enables efficient heat removal from the most critical components while maintaining system reliability.

A key characteristic of DTC systems is their hybrid nature. Since they directly cool only the silicon components, DTC removes approximately 70-80% of rack-generated heat, requiring complementary air cooling infrastructure to manage the remaining 20-30% from power supplies, memory modules, and networking equipment.⁵

Performance Advantages

DTC cooling systems demonstrate remarkable thermal performance improvements. While traditional air cooling maxes out around 25kW per rack, DTC systems can effectively manage thermal loads up to 100kW per rack.⁶ This capability directly enables the high-density computing requirements of modern AI workloads without thermal throttling.

The technology offers additional operational benefits including reduced fan energy consumption, improved equipment reliability through consistent temperature control, and space efficiency gains by eliminating large air handling equipment. For retrofit applications, DTC systems can often be implemented in existing data centers without major infrastructure overhauls.

Market Adoption and Drivers

The DTC liquid cooling market is experiencing rapid growth, driven by several converging factors. Hyperscalers (e.g., Microsoft, Meta, Amazon, Google) and cloud service providers are aggressively scaling AI infrastructure, creating immediate demand for thermal solutions capable of supporting high-density workloads. The technology's compatibility with existing data center infrastructure makes it attractive for retrofit applications where complete facility redesigns are impractical.

Regulatory pressures around energy efficiency are accelerating adoption, as DTC systems significantly improve Power Usage Effectiveness (PUE) compared to traditional air cooling. Environmental regulations and corporate sustainability commitments are making energy-efficient cooling a compliance requirement rather than just an optimization opportunity.⁷

Strategic Market Opportunities

The transition to DTC cooling creates significant opportunities for companies with industrial thermal management expertise. This technology requires sophisticated heat exchanger design, precision fluid handling systems, and advanced control algorithms - capabilities that align closely with existing competencies in thermal management, process control, and industrial automation.

Key component opportunities include measurement instrumentation, cold plate manufacturing, coolant distribution systems, and intelligent thermal management controls. Companies with established thermal management, HVAC, and process control capabilities have the relevant technologies in heat exchangers, pumping systems, and process control that can be adapted for data center applications.

Part III: The Future Era - Immersion Cooling and Beyond (2025+)

The Next Thermal Frontier

As computational demands continue escalating, even direct-to-chip cooling will reach its limitations. The industry is now pioneering immersion cooling technologies that represent the ultimate thermal management solution for extreme-density computing applications. This approach completely reimagines data center cooling by submerging entire servers in specially formulated dielectric fluids.

Immersion Cooling Fundamentals

Immersion cooling systems eliminate air as a heat transfer medium entirely, instead utilizing engineered dielectric fluids that safely contact electronic components. These fluids possess excellent thermal conductivity properties while maintaining electrical insulation, enabling direct heat transfer from all components simultaneously.

The technology offers two primary implementations: single-phase and two-phase systems. Single-phase immersion maintains the coolant in liquid state throughout the thermal transfer process, while two-phase systems utilize phase change (boiling) to achieve superior heat transfer rates through latent heat absorption.

Technical Capabilities

Immersion cooling represents a quantum leap in thermal management capability. These systems can handle rack densities exceeding 100kW while achieving the lowest possible PUE ratings.⁸ Two-phase immersion systems demonstrate particular effectiveness for extreme applications, utilizing the high energy transfer rates associated with the phase change phenomenon.

A critical advantage of immersion cooling is its ability to remove 100% of generated heat without requiring ancillary air cooling infrastructure. This eliminates fans, reduces component failure rates, and enables unprecedented computational densities. The absence of air movement also creates silent operation and eliminates dust contamination concerns.

However, this technological shift presents a strategic inflection point for the industry. While immersion cooling solves critical operational challenges for data centers, it simultaneously disrupts the entire cooling equipment ecosystem and supply chain. Manufacturers who are heavily invested in traditional air cooling systems are facing a critical crossroad - as immersion cooling eliminates the need for the equipment these companies have built their businesses around. As a result, the competitive landscape is fundamentally reshaping, favoring players who can rapidly evolve their portfolios toward liquid cooling technologies, acquire emerging specialists, or establish strategic partnerships to remain relevant in the next generation of data center infrastructure.

Market Trajectory and Applications

Current market data indicates immersion cooling systems are experiencing rapid adoption, particularly in high-performance computing and AI applications. The total market opportunity for liquid cooling (e.g., direct-to-chip, immersion cooling) is projected to grow at 67% CAGR through 2028,⁹ with immersion systems capturing increasing market share as computational densities continue rising.¹⁰

Early adopters include hyperscale operators deploying AI training clusters, research institutions running high-performance computing workloads, and edge computing applications where space efficiency is paramount. The technology is particularly attractive for new-build facilities designed from inception for liquid cooling.

Economic and Environmental Drivers

Immersion cooling offers compelling total cost of ownership advantages despite higher initial capital requirements. Energy savings from eliminated fan power, reduced cooling infrastructure, and improved PUE metrics create significant operational expense reductions. Equipment longevity improvements from consistent thermal management and eliminated thermal cycling further enhance economic returns.

Environmental considerations are increasingly driving adoption decisions. Immersion systems enable heat recovery at higher temperatures (up to 80°C versus 30-45°C for air cooling), making waste heat suitable for district heating or other thermal applications.¹¹

Future Considerations

The immersion cooling sector continues evolving with innovations in dielectric fluid formulations, system designs, and deployment mechanisms. Vendors are developing biodegradable fluids with low global warming potential to address environmental concerns. Modular immersion systems are enabling easier deployment and scalability.

Emerging technologies like microconvective cooling are pushing thermal management capabilities even further, utilizing microchannel heat exchangers for precise chip-level thermal control. These innovations suggest that immersion cooling represents just the beginning of a broader transformation toward liquid-based thermal management.

This evolution is expected to accelerate significantly over the next five to ten years as data centers are rapidly scaled out and deployed to meet escalating computational demands. As liquid cooling gains broader market penetration, immersion cooling will capture an increasing share of that adoption - driven by the need for innovative thermal solutions capable of supporting extreme computational densities and AI-driven workloads.

The growth trajectory creates opportunities across two key stakeholder groups. On the supply side, immersion cooling benefits equipment manufacturers and solution providers including HVAC companies, specialized cooling technology firms, liquid cooling system integrators, and industrial equipment providers who can adapt their thermal management capabilities for data center applications. On the demand side, primary adopters include hyperscalers building AI infrastructure, colocation providers, high-performance computing facilities, cryptocurrency mining operations, and adjacent industries with extreme cooling requirements (e.g., telecommunications infrastructure, financial services trading operations, manufacturing with edge AI, scientific research institutions).

Strategic Implications for Industrial Leaders

Market Positioning Opportunities

The evolution toward immersion cooling creates unprecedented opportunities for companies with thermal management, fluid handling, and process control expertise. Immersion cooling systems require sophisticated fluid handling systems, advanced heat exchangers, and intelligent control systems - all areas where industrial leaders possess deep expertise.

Key opportunity areas include:

- **Fluid Management Systems:** Precision pumping, filtration, and distribution networks
- **Heat Exchanger Technology:** Advanced brazed plate and shell-and-tube heat exchangers optimized for dielectric fluids
- **Process Control Systems:** Intelligent thermal management with predictive analytics and automated optimization
- **Safety and Monitoring:** Leak detection, fluid quality monitoring, and emergency shutdown systems

Competitive Landscape

The immersion cooling market remains relatively fragmented, with opportunities for established industrial players to leverage their scale, manufacturing capabilities, and customer relationships. Current market leaders include specialized cooling companies (e.g., Vertiv), but established global conglomerates with thermal management and industrial infrastructure capabilities possess advantages in manufacturing scale, global service networks, and integrated system capabilities.

Companies with strong positions in thermal management, fluid handling, and process automation are well-positioned to capture market share as immersion cooling adoption accelerates. The key is understanding data center requirements and adapting existing technologies for this specialized application.

Conclusion: Preparing for the Thermal Management Revolution

The Inevitable Transformation

The data center industry is experiencing a fundamental shift in thermal management philosophy. Traditional air cooling, while adequate for conventional workloads, cannot meet the demands of modern AI and high-performance computing applications. Direct-to-chip cooling provides an essential bridge technology, but the industry's trajectory clearly points toward a need for more advanced cooling applications (e.g., immersion cooling) to handle extreme-density computing.

This evolution is not simply a technological shift - it represents a complete reimagining of data center design principles. Future facilities will be optimized for liquid cooling from inception, with thermal management systems integrated into fundamental infrastructure decisions. The companies that understand and prepare for this transformation will capture the greatest opportunities.

Strategic Imperatives

For industrial leaders, the cooling evolution represents both challenge and opportunity. The challenge lies in adapting existing capabilities to data center requirements while building new competencies in areas like dielectric fluid handling and high-density thermal management. The opportunity lies in applying decades of industrial thermal management expertise to a rapidly growing market.

Success in this market requires understanding not just the technology but the operational requirements of data center environments. Reliability, serviceability, and scalability take precedence over raw performance. Systems must operate continuously for decades with minimal maintenance, in environments where downtime costs millions of dollars per hour.

The Path Forward

Companies seeking to participate in the data center cooling revolution should focus on several key areas:

Market Understanding: Develop deep knowledge of data center operational requirements, working closely with key stakeholders (e.g., hyperscalers, colocation providers, architecture and engineering firms) to understand current pain points and future needs. This foundational understanding must come before any technology or capability investments to ensure alignment with evolving market demands.

Capability and Technology Development: Invest in both R&D capabilities and new product offerings specifically focused on liquid cooling technologies, with emphasis on immersion systems and emerging cooling approaches. This dual focus requires adapting existing thermal management technologies for data center applications - particularly around dielectric fluid compatibility and high-density heat transfer -

while simultaneously building entirely new capabilities to meet the heightening computational and thermal demands of AI-driven infrastructure.

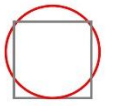
Partnership and M&A Strategy: Consider strategic partnerships or acquisitions of specialized cooling companies to combine complementary strengths and accelerate market entry. Partnerships enable access to liquid cooling expertise through collaboration, while targeted acquisitions provide immediate ownership of proven technologies, established customer relationships, and advanced operational capabilities.

The data center cooling market represents a significant growth opportunity for companies with relevant capabilities. This transformation from air to liquid cooling is inevitable, driven by fundamental physical limits and escalating computational demands. Those who act decisively to understand and serve this market will benefit from sustained growth in one of technology's most critical infrastructure sectors.

The question is not whether data centers will transition to liquid cooling, but how quickly the transformation will occur and which companies will capture the value created by this fundamental shift. For industrial leaders with thermal management expertise, the cooling revolution represents a historic opportunity to apply proven capabilities in a rapidly expanding market.

Footnotes:

1. Green Revolution Cooling, "The Definitive Guide to Immersion Cooling"
 2. Uptime Institute, "Uptime Institute Global Data Center Survey 2021"
 3. Schneider Electric, "The AI Disruption" Challenges and Guidance for Data Center Design"
 4. Uptime Institute, "Five Data Center Predictions for 2024"
 5. Datacenters.com "Why Liquid Cooling Is Becoming the Data Center Standard"
 6. Vertiv, "Vertiv™ CoolChip CDU, 121"
 7. Datacenters.com "Inside the Eu's Crackdown on Data Center Emissions: What It Means Globally", Vertiv, "How climate change rules are driving data center cooling innovation", IBM, "IT sustainability beyond the data center", Google, "Innovating across our operations and supply chain"
 8. SpringerOpen, "Enough hot air: the role of immersion cooling"
 9. Evercore ISI, "'DEEP DIVE' On Liquid Cooling Opportunity"
 10. Red Chalk Group Analysis
 11. Elsevier, "A review of data center cooling technology, operating conditions and the corresponding low-grade waste heat recovery opportunities"
-



REDCHALK
GROUP

