The Liquid Cooling Reckoning

Why enterprise data centers must act now.

Executive Summary

Around two years ago, Maintech began proactively building relationships with liquid cooling manufacturers and industry experts specifically to get ahead of the curve before client demand arrived. That investment has included training, certification, and working partnerships with specialist manufacturers, ensuring the technical teams are ready to deliver.

"Our focus at Maintech is to ensure that when [liquid cooling] does hit our client data centers... we can provide the support the same as if it was an air cool system. And from what we've seen and what we've learned, we're very comfortable that we're going to be able to do that."

— Bill D'Alessio, EVP, Maintech

As AI accelerates toward 4,000-watt chips by 2029, data centers are being asked to dissipate heat at a scale they were never built for. But AI is not the only pressure. Denser storage, high-throughput networking, and ordinary enterprise growth are all adding to the thermal load. For racks within manageable power envelopes, air cooling holds up well, particularly alongside direct-to-chip approaches. Where density tips past what air can realistically handle, liquid cooling stops being a specialist choice and starts being the obvious one - a technology with over 60 years of engineering behind it, tracing back to IBM systems in the 1960s. What has changed is the urgency.

Why it matters:

- Water is 3,000× more effective than air at absorbing heat.
- This is not just an AI problem; standard servers are hitting thermal limits.
- 22% of operators already use direct liquid cooling; 61% are considering it.
- Regulations are tightening. Efficiency is not optional; it is a requirement.
- Liquid cooling enables heat recovery and reuse, turning a waste product into a building heating asset.

Direct-to-chip and immersion cooling represent the two most practical paths forward for enterprises. Direct-to-chip integrates with existing infrastructure and is supported across major vendors; immersion - where servers are submerged in thermally conductive fluid - offers even greater heat removal for the highest-density environments and is gaining rapid traction across the industry.

This paper draws on insights from Maintech's recent industry roundtable, How Liquid Cooling Could Unlock Opportunity for Your Enterprise Data Centre (January 2026, London), featuring Dr. Jon Summers (RI.SE Institute Sweden), Stephen Zhao (Castrol Thermal Management), and Bill D'Alessio (EVP, Maintech). The perspectives shared at that event - combined with our own research and operational experience - inform the recommendations that follow.

Data Centers Have a Heat Problem

We have spent over fifty years inside data centers and seen every major technology shift from Y2K to the cloud. This is the one that worries us most, because most of the industry is pretending it is not happening.

But here is what most people do not realize: liquid cooling is not new technology. Researchers were writing about running liquid alongside electronics in 1968. IBM's earliest mainframes used nickel-cooled pipework under raised floors - which is the original reason many data centers have raised floors in the first place. The high-performance computing community has been running liquid-cooled systems for over twenty-five years. The top 100 supercomputers in the world are all liquid cooled. This is proven, mature infrastructure with a deep engineering pedigree.

What changed is the heat. The GPUs going into data centers today produce so much heat that air conditioning cannot keep up. The latest high-end GPUs generate 1,400 watts of heat each - roughly the same as a small space heater.[1] By 2027, a single rack of these will need more cooling than most data centers were built to provide for an entire row.[2] By 2029, individual GPUs are projected to exceed 4,000 watts.[3]

Meanwhile, most enterprise data centers were designed decades ago to handle a fraction of that heat using fans and cold air. That worked when servers were modest. It does not work when every business is racing to deploy AI and the hardware behind it runs hotter than anything the industry has ever seen.

The answer is not more fans or bigger air conditioning units. For many high-density environments, liquid cooling is becoming one of the most viable approaches - running water or specialised coolant directly to the hottest components, the same principle that cools a car engine. Water absorbs roughly 3,000 times more heat than air per unit of volume.

"...a thousand watts going in at one volt, that's a thousand amps and just goes straight into a piece of silicon and stops there. The only way out is heat, and so we have to manage that heat, and extracting that heat makes a lot of sense to do this with liquid."

Dr. Jon Summers, RI.SE Institute Sweden

An NVIDIA and Vertiv study, published by the American Society of Mechanical Engineers, found that switching from air to 75% liquid cooling improved overall energy efficiency (TUE) by 15.5%, cut total data center power by 10.2%, and slashed fan energy by 80%.[5] Those are not marginal improvements. That is a fundamental shift in operating economics.

Over the past two years, Maintech has worked directly with liquid cooling manufacturers, researchers, and operators to understand how these systems behave in production environments. Not in vendor demonstrations, but in the data centers where our engineers are already working.

Why This Matters Now

It is tempting to treat liquid cooling as a concern only for the tech giants - Google, Meta, Microsoft. That was a reasonable position three years ago, but it is not anymore.

The liquid cooling market doubled in 2025, **reaching close to $3 billion.**[3] Every major analyst forecasts it will grow to $15–20 billion within the next five years.[8] That kind of growth does not happen for niche technology. It happens when an entire industry recognizes that the old way of doing things has stopped working. As of early 2024, 22% of data center operators were already using direct liquid cooling, with 61% open to considering it. [14]

Every major server manufacturer now offers liquid-cooled options. Every large-scale AI deployment assumes liquid cooling as the baseline. But the transition is not just happening at hyperscale.

The Enterprise Case Nobody is Talking About

We are seeing a phenomenon that does not get enough attention:

Enterprises that are not running GPU workloads at all, but whose latest generation of standard servers simply runs hotter than the last. Telecom companies, healthcare organizations, and even cloud storage providers are hitting a ceiling where their existing cabinet air conditioning can not keep up.

The servers keep overheating, and in our experience the conversation quickly arrives at the same fork in the road: invest in a much more powerful, and expensive, HVAC system, or address the heat at the source with liquid cooling. The first option scales poorly. The second is increasingly the one operators are choosing.

We have talked to storage companies that offer mailbox and data backup services - no AI, no GPUs - who are exploring immersion cooling specifically to extend the lifespan of their storage disks. The economics of server procurement are shifting too: AI demand is not just inflating the cost of drives, SSDs, and NVMe storage, it is disrupting access to them entirely. Semiconductor constraints have pushed lead times out with no clear timeline for relief. If a server costs twice what it did three years ago and takes six months to replace, keeping it running longer stops being a nice-to-have and starts being a business decision.

The real question:

If your data center plan assumes air cooling will be enough for the next five years, what is that assumption based on?

And what is your backup plan if it is wrong?

What the Industry Gets Wrong

There is a lot of confident misinformation circulating about liquid cooling. Some of it comes from vendors with something to sell. Some of it comes from people who have not actually operated these systems. We have spent two years working with researchers, manufacturers, and coolant specialists to understand the real picture. Here is what we have found.

Myth	Reality
"Liquid cooling uses more water."	It does not. Most systems are closed-loop, and liquid cooling saves energy, which in turn reduces water use at the power plant level.
"Liquid cooling is automatically more efficient."	Not always. Newer chips require lower inlet temperatures, which can make cooling harder. Efficiency depends on full system design, not just cold plates.
"Air cooling is dead."	It is not. Many workloads (like storage or networking) still run just fine on air. Liquid cooling is for the hot zones, not everything.
"You need liquid cooling above 20 kW per rack."	Not necessarily. Rear-door heat exchangers can handle 50–60 kW. The real question is how much fan noise, vibration, and climate risk you are willing to accept.
"Just put your servers in a tank."	Immersion is not that simple. Standard cables fail, many parts are not compatible, and OEM warranties may not apply (e.g. NVIDIA air-cooled GPUs).
"Set your coolant and forget it."	Not anymore. Early fluids degraded fast. Now, 6-monthly testing and inline monitoring are standard. These systems still need maintenance.

Liquid Cooling Approaches Compared

“Liquid cooling” is a family of approaches, each with different trade-offs. There are now over 45 liquid cooling companies in the market, making differentiation genuinely difficult. Here is what you need to know about the major categories.

Direct-to-Chip (DTC)

For most enterprises, direct-to-chip is the most accessible starting point. Metal cold plates filled with coolant mount directly onto the hottest components - the processors. The coolant carries heat away to a cooling distribution unit (CDU) and recirculates. The rest of the server still uses air for secondary components. It is a hybrid approach, and it works because it targets the actual bottleneck without requiring you to rip everything out and start over. Every major server vendor supports it, and it is where most new liquid cooling is being deployed.

Cold plate design matters more than most buyers realize. There are micro-channels, skived fins, pin fins, spray jets, and other approaches - each with different thermal resistance characteristics. NVIDIA’s upcoming Vera Rubin architecture uses laser-welded micro-channels matched to individual GPU hotspots, achieving a thermal resistance of approximately 0.015 Kelvin per watt. That is how they claim 45°C inlet compatibility. Whether that works at production scale across hundreds of racks remains to be seen - but the engineering direction is clear.

Immersion Cooling

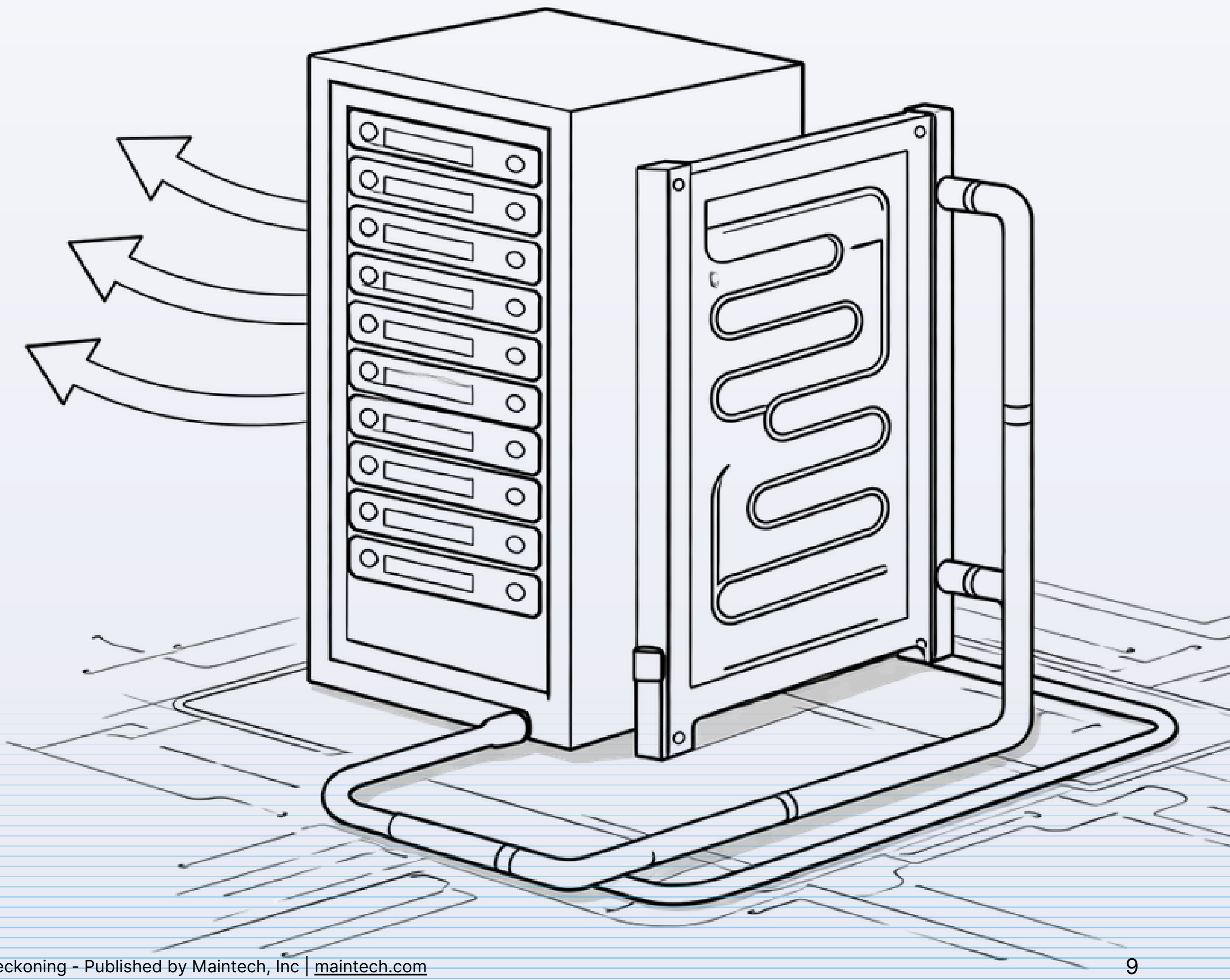
Immersion cooling submerges entire servers in a non-conductive (dielectric) fluid. Single-phase immersion, where the fluid remains liquid, is gaining real traction. Two-phase immersion, where the fluid boils on contact with hot components, is still early-stage and faces additional challenges: biodegradable fluids that change phase also tend to be flammable, and the fluorinated “forever chemicals” used in two-phase systems present environmental concerns.

Immersion has a genuine advantage for hardware longevity. It is not just that servers run cooler, it is that immersion tanks have massive thermal inertia, which absorbs the spiky power fluctuations that AI workloads produce. Dr. Jon Summers put it directly: it is not the high temperature that kills GPUs; it is the thermal fluctuation. There is evidence that air-cooled GPUs in AI data centers are showing failure after just eighteen months. In immersion, the thermal mass of the fluid dampens those fluctuations, which is why immersed IT consistently lasts longer.

A critical caveat: NVIDIA does not warranty air-cooled GPU hardware for immersion use. They have cited signal integrity concerns - dielectric fluids can interfere with high-frequency signals between components, potentially requiring additional retimers that consume extra power. Chassis-based immersion systems (like those from Isotope or AirSys) that allow server pull-out and swap like a standard rack are helping bridge the operational gap, but the warranty question must be addressed before any enterprise deployment.

Rear-Door Heat Exchangers

These mount a liquid-cooled radiator on the back of a standard rack, catching heat from the exhaust air. The servers themselves do not change. It is the easiest option to install, handles up to 50–60 kW per rack, and buys you time. But it does not address the core problem - the heat at the chip. The liquid loops you install for rear-door exchangers, however, are the same infrastructure you will need for DTC, so the investment is not wasted.



Enterprise Comparison

	Direct-to-Chip	Single-Phase Immersion	Two-Phase Immersion	Rear-Door Heat Exchanger
Best use case	Start here for most enterprises	Consider for very high density or new builds	Monitor; deploy only at extreme density	Use as a short-term bridge, not a long-term plan
Best suited for	AI, HPC, general enterprise	Very high density, noise-sensitive environments	Extreme density R&D, pilot program	Moderate density retrofit
Maturity	High — broad vendor support	Medium — growing rapidly	Early — limited production scale	High — simple to deploy
Power per rack	40–120+ kW	60–200+ kW	100–250+ kW	20–60 kW
NVIDIA warranty	Fully supported	Not warranted for air-cooled GPUs	Not warranted for air-cooled GPUs	No impact
Retrofit complexity	Moderate	High	High	Low

The Operational Reality Nobody Warns You About

The reality of operating liquid-cooled infrastructure day to day is more complex. Here is what we have learned from working with these systems and talking to the people who build them:

Coolant is Not Just Water

In direct-to-chip systems, you are typically running a propylene glycol/water mix (PG25). Pure water is a better heat conductor, but things grow in it very quickly. PG25 is naturally biostatic, which is why it is the reference fluid for NVIDIA's NVL architecture. But even PG25 needs monitoring: galvanic corrosion can occur where dissimilar metals meet inside cold plates and CDUs, bio-organic growth can accumulate in micro-channels despite the glycol, and the fluid's chemical properties degrade over time.

The industry standard is now six-monthly fluid sampling and testing. Newer formulations from companies like Castrol extend reliable life to around five years, but past that point, testing frequency should increase. Inline sensors are now available that provide continuous monitoring and can flag issues months before they become operational problems. If your liquid cooling vendor does not include an ongoing fluid management program, that is a red flag.

Material Compatibility Is Non-Negotiable

For immersion deployments, every component that contacts the dielectric fluid must be verified for compatibility. The most common early mistake is using standard patch cables or power leads. The plasticizer in standard cable insulation leaches into dielectric fluid within about six months, making cables brittle and prone to failure. Connectors, seals, and even certain components on the motherboard may not be compatible. Coolant suppliers provide compatibility checklists - use them and verify independently.

Concurrent Maintainability Is Harder Than You Think

In a traditional air-cooled environment, you can swap a failed component without affecting its neighbors. With direct-to-chip cooling, pipe work connects every cold plate in a rack to a shared coolant loop. Making that system concurrently maintainable at the rack level - meaning you can service one server without draining the entire rack - is very difficult. You can achieve concurrent maintainability at the CPU level with quick-disconnect fittings, but you will still need to power down some IT to drain and maintain the coolant system. In high-performance computing, the accepted practice is one planned maintenance day per year where the entire system goes down. For enterprise environments with strict uptime SLAs, this needs careful planning.

Your White Space Shrinks, Your Gray Space Grows

Densification is the whole point of liquid cooling: more compute in less physical space. But the cooling infrastructure itself - CDUs, pipework, manifolds, leak detection - takes up room. As your white space (where the IT sits) gets denser, your gray space (the mechanical plant) needs to grow. Factor this into capacity planning from the start, not after the CDUs arrive.

Raised Floor or Slab Floor?

The original reason data centers have raised floors is to accommodate pipework underneath. If you are routing large volumes of coolant across your data hall, a raised floor provides natural leak containment. Some new builds are running pipework on slab floors with overhead distribution, which is cheaper but carries a higher leak risk. For an industry that is fundamentally risk-averse, this is a design decision that deserves serious scrutiny.

What Operators Discover After Deployment

The challenges above are well-documented. What is less discussed is what maintenance teams encounter once systems are live. Based on our operational experience, these are the issues that consistently catch operators off guard:

Quick-disconnect fittings are a common failure point. They are essential for concurrent maintainability, but under repeated use they wear and drip. Teams that have not rehearsed fluid handling procedures treat a minor leak as an emergency. It is not, but only if you have trained for it.

CDU monitoring is underestimated. Operators focus on server health and overlook the CDU itself, such as flow rates, pressure differentials, and inlet temperatures all require regular attention. A CDU running outside specification will degrade cooling performance quietly, long before any alarm triggers.

Thermal sensor calibration drifts. Sensors that read accurately at installation can shift over time, particularly in systems where coolant chemistry has changed. Teams relying on threshold alerts without periodic physical verification are operating on data they should not fully trust.

Leak detection generates false positives. Early-generation sensors are prone to alerting on condensation or minor environmental changes. Without a calibrated response protocol, false positives create alert fatigue, which is exactly the wrong state to be in when a real leak occurs.

None of these are insurmountable. They are, however, the kind of operational detail that only emerges through hands-on experience, and that a service partner should be able to bring to the table before deployment, not after.

"...I think you should invest in modular infrastructure, definitely. If you can go modular as much as you can, it can grow."

— Dr. Jon Summers, RI.SE Institute Sweden

The Cost of Transition

The most common objection we hear is cost. Yes, liquid cooling requires upfront investment - piping, cooling units, leak detection.

But framing this as “liquid cooling costs more” misses the bigger picture.

The Energy Bill

Cooling typically accounts for up to **40% of a data center's total energy bill.**[9] McKinsey estimates that liquid cooling can reduce energy consumption by 31-37% compared to air-cooled systems.[8] For a mid-sized 10-megawatt facility, that difference works out to roughly **$3.5 million per year - $35 million over a standard ten-year cycle.** McKinsey puts the payback period at under one to three years, depending on local electricity costs.[8]

The Space Dividend

Liquid cooling supports rack densities well above 50 kilowatts - many times the industry average of 8 kilowatts typical of air-cooled facilities.[6] In expensive real estate markets - London, Singapore, Frankfurt, Northern Virginia - the cost of building or leasing additional data center space often dwarfs the cost of a cooling upgrade. Densification means you can avoid new construction entirely.

The Hardware Lifespan

Heat kills electronics - but it is not just about temperature. The data increasingly shows that it is thermal fluctuation that causes the most damage. Modern AI workloads are intensely spiky: GPUs ramp to full power for training bursts, then drop. Each cycle creates thermal stress on the silicon. There is growing evidence that air-cooled GPUs in AI data centers are showing failures after just eighteen months of operation.

Immersion cooling addresses this directly. The large volume of coolant in an immersion tank acts as a thermal buffer, absorbing power spikes and maintaining a much more consistent temperature. Supermicro's field data shows that air-cooled servers immersed in dielectric fluid significantly outlast their air-cooled counterparts. Direct-to-chip cooling also helps, running components cooler and eliminating the strain on server fans - one of the few moving parts in modern servers and a common source of failures.

Component lifespan roughly doubles for every 10°C drop in operating temperature - a principle rooted in the Arrhenius equation and widely observed in field data.[7] The result: longer-lasting equipment, fewer emergency replacements, and better returns on the hardware you have already bought.

> ***"I've been singing the praises of liquid cooling for 15 years, and when I first started in liquid cooling, people were saying, my year's researching that - it's got no life in it at all. And I'm somewhat vindicated today - as it's so popular now."***
>
> **Dr. Jon Summers, RI.SE Institute Sweden**

What the Smartest Operators Have in Common

Over the past two years, we have worked with enterprise clients across financial services, healthcare, technology, and government. The organizations handling this transition well share three things in common:

They Plan for What is Coming, Not What is Here

The best operators start by asking: "What will our data center need to support in three to five years?" Then they work backwards. They assess their current power capacity, floor loading, plumbing, and workloads before evaluating any technology. The most common mistake we see is the opposite - an enterprise hits a specific heat problem with a new AI cluster, scrambles for a quick fix, and ends up with a patchwork of approaches that do not add up to a strategy.

They Refuse to Be Locked In

The liquid cooling market is crowded and evolving fast. Many vendors offer proprietary systems that only work with their products. NVIDIA's Vera Rubin architecture, for example, is moving toward a closed system where the rack rolls in as a self-contained unit - just connect water, power, and data. That simplifies deployment but raises questions about what it means for third-party cold plate manufacturers and for the operator's ability to choose their own components.

The smartest operators insist on vendor-neutral assessments and open standards wherever possible. Your cooling infrastructure will need to support three or four generations of server hardware over its useful life. If it only works with one vendor's equipment, you have traded one form of inflexibility for another.

They Invest in the Skills Gap

Liquid cooling requires a different skill set from air-cooled maintenance. Technicians need to understand fluid handling, connector management, CDU maintenance, system flushing procedures, and leak response protocols. Most enterprise IT teams do not have this expertise today. The organizations getting ahead are either building this capability internally or partnering with a service provider that already has it - and verifying that the provider has actual operational experience, not just a slide deck.

How Maintech Supports the Transition

We have been inside data centers for over fifty years. We manage hundreds of thousands of servers across more than 110 data centers in 60+ countries, serving four of the top ten global banks. We say that not as a credential, but as context: when we talk about liquid cooling operations, we are not theorizing. We are describing work we do.

To strengthen our liquid cooling offering, Maintech has systematically built relationships with liquid cooling manufacturers, coolant specialists like Castrol, and innovative system integrators, including AirSys and Isotope. We have visited their labs, tested their products, and trained our field engineers on their systems. Our goal is straightforward: when liquid cooling arrives in our clients' data centers, they should not have to care whether a server is air-cooled or liquid-cooled. When they call us to report a server down, the experience should be the same.

Our Four-Stage Approach

Readiness Assessment

We analyze your current environment to determine readiness for liquid cooling - power capacity, floor loading, plumbing infrastructure, workload profiles. We identify gaps and set measurable objectives.

Vendor-Neutral Roadmap

We assess your specific requirements and recommend the best-fit solutions and partners for your workloads, space constraints, and sustainability targets. We foster introductions to the manufacturers and specialists we have vetted - and you can visit their labs to see the technology in person.

Managed Deployment

We coordinate deployment with leading vendors and manage the project end-to-end - from rack and stack through power-on testing to seamless integration with existing infrastructure.

Ongoing Operations and Maintenance

We provide 24/7 proactive monitoring, preventive maintenance, coolant testing and remediation, emergency response, and component replacement from our network of 100+ forward stocking locations. Real-time thermal data from DTC systems feeds directly into our predictive maintenance programs, enabling us to identify and address potential failures before they affect uptime.

Technology Partners in Action

Below are two of our current delivery partners whose technologies are already helping enterprises transition to efficient, scalable cooling infrastructure.

LiquidRack™ – Spray-Based Rack-Level Cooling

By Airsys

A compact, CDU-free rack system that uses precision spray cooling to dramatically improve thermal efficiency and reduce dielectric fluid use by up to 80% compared to immersion.

- Up to 175 kW cooling per rack
- No external CDU required - integrated pump and heat exchanger
- Ideal for high-density deployments with strict space or fluid constraints
- Supports free-cooling year-round at higher liquid temperatures
- Simplifies maintenance with modular server cassettes and quick disconnects

Why we partner with Airsys:

Most immersion systems require significant fluid volumes and facility modifications. LiquidRack's spray-based approach delivers comparable density with 80% less fluid - one example of how vendors are addressing the challenge of retrofitting existing facilities or operating in environments where fluid handling is a constraint.

Technology Partners in Action

Iceotope KUL BOX – Fully Integrated AI Edge Cluster

By Iceotope

ICEOTOPE®

A turnkey, fanless solution for AI inferencing at the edge. Combines compute, networking, memory, storage, and immersion cooling - all pre-assembled and ready to deploy.

- Compact and facility-independent (no external chillers or facility water)
- Quiet operation with no fans or airflow paths
- Extended hardware lifespan with stable thermal performance
- Ideal for telco, medical, or industrial edge environments
- Delivered with full installation, training, and lifecycle support

Why we partner with Iceotope:

Edge deployments fail when they are too complex for local teams to support. KUL BOX arrives as a sealed, pre-integrated unit - no chillers, no facility water, no fans to fail. For clients deploying AI inference in hospitals, factory floors, or telecom sites without dedicated cooling infrastructure, it removes the variables that typically derail edge projects.

The Regulation You Cannot Ignore

If the engineering and the economics do not convince you, regulation will.

The European Union now requires data centers above 500 kilowatts to report their energy efficiency, with binding targets for 2030 that are extremely difficult to meet with air cooling alone.[10] But the EU framework is just the baseline. Denmark requires new facilities to make waste heat available to district heating. The Netherlands imposed a moratorium on new construction in Amsterdam. Ireland has grid connection constraints limiting new capacity in Dublin.

Singapore only permits new data center construction if the proposed facility meets strict efficiency standards.[11] In the United States, roughly $64 billion worth of data center projects have been delayed or canceled due to community resistance and environmental concerns.[12]

Heat reuse is the opportunity most operators miss. Air cooling rejects heat at 30–35°C - too low to be useful. Liquid cooling systems running at 55°C or above can feed directly into district heating networks. Data centers in Stockholm, Helsinki, and Copenhagen are already supplying heat to thousands of homes, with some operators reporting heat sales offsetting 10–15% of electricity costs. In constrained markets, a heat reuse agreement can be the difference between planning approval and rejection.

The regulatory direction is clear: efficiency requirements will get stricter, not looser. The question is not whether these regulations will affect you – it is whether you will be ahead of them or catching up.

What is Coming Next

We are always talking with researchers and engineers who are shaping the next wave, and a few developments are worth watching.

The GPU Manufacturer as Infrastructure Dictator

NVIDIA's trajectory is clear: with each generation, they are integrating more of the cooling system into the reference architecture. The Vera Rubin design is essentially a closed system - you roll the rack in, connect water loops, power, and data, and you are running. The "bird's nest of pipework and cables" of previous generations is gone. That is operationally simpler, but it means NVIDIA is increasingly dictating the cooling architecture, not the operator. The main variable that remains in the operator's hands is the coolant and its ongoing management.

The Grid Problem

Here is a detail that reveals how much the industry has shifted: traditionally, a UPS protects the IT from the grid - smoothing out power fluctuations so servers run cleanly. In large AI data centers today, it is the reverse. The IT workloads are so spiky that UPS systems are now needed to protect the grid from the IT. Super-capacitors may be required to absorb the fluctuations. That is a sign of just how much power these systems consume and how fundamentally they are changing the relationship between data centers and the energy infrastructure they depend on.

The Reckoning is Here

The chip roadmap is set. NVIDIA's published trajectory through 2028 makes it clear that heat output per chip and per rack will keep climbing aggressively.[4] Every major business is deploying or planning to deploy AI infrastructure. Air cooling has hard physical limits that have already been reached for high-density workloads. And climate change is pushing outdoor temperatures beyond what many existing cooling systems were designed to handle.

The organizations taking action now - assessing their environments, building vendor-neutral roadmaps, starting with direct-to-chip cooling, investing in the operational skills to manage the transition, and working with partners who have actual hands-on experience - are positioning themselves for a decade of infrastructure advantage. More computing in less space. Lower energy costs. Longer-lasting hardware. Compliance with regulations that are only going to get tighter.

In five years, there will be two kinds of enterprise data centers: those that made the transition on their own terms and those that were forced into it. The window to choose is closing. What separates the two is operational readiness. The organisations that navigate this transition well will be those that planned for it, built the skills to support it, and chose partners who already understand how these environments behave in production. That is precisely where Maintech operates: as the team that keeps infrastructure running once the deployment is done.

"… if as an enterprise or as a data center you're thinking of adopting liquid cooling, it's a very good time to start going in because the industry is now more mature. I think a lot of these risks are quite easily managed. If you go to the right players, we can support you along your journey."

Stephen Zhao, Castrol Thermal Management

Sources & References

Hardware Roadmaps

1. Inside NVIDIA Blackwell Ultra: The Chip Powering the AI Factory Era | NVIDIA Technical Blog

2. Nvidia's Vera Rubin CPU, GPUs chart course for 600kW racks • The Register

3. Data Center Liquid Cooling Market to Approach $7 Billion by 2029 as AI Deployments Accelerate, According to Dell'Oro Group - Dell'Oro Group

4. Nvidia's roadmap shows just how deep Moores Law is buried • The Register

Technical & Engineering

5. Quantifying Data Center PUE When Introducing Liquid Cooling

6. AI and cooling: methods and capacities | Uptime Intelligence

7. Does a 10°C Increase in Temperature Really Reduce the Life of Electronics by Half? | Electronics Cooling

Market Data

8. Keeping cool in the data age | McKinsey & Company

9. Towards energy-efficient data centers: A comprehensive review of passive and active cooling strategies - ScienceDirect

Regulatory & Sustainability

10. Energy Efficiency Directive | eudca.org

11. Singapore's Data Centers are Tackling the Sustainability Dilemma | Schneider Electric

12. $64 billion of data center projects have been blocked or delayed amid local opposition — Data Center Watch

13. How the Bipartisan Dirty Data Center Boom is Leaving Taxpayers Without Power (Part 1 of 3) · Consumer Federation of America

14. Liquid cooling becoming essential as AI servers proliferate | Network World

Appendix A: A Glossary of Liquid Cooling Terms

A quick-reference glossary for key technologies and terminology used throughout this whitepaper.

DTC (Direct-to-Chip Cooling)

A method where cold plates are mounted directly onto hot components (like CPUs or GPUs), circulating liquid coolant over them to remove heat more efficiently than air cooling.

CDU (Coolant Distribution Unit)

A rack-level or room-level system that manages liquid flow, temperature, and pressure within a liquid cooling loop. Acts as the intermediary between facility cooling and IT equipment.

NVL (NVIDIA Liquid-Cooled Platform)

NVIDIA's liquid-ready platform designed for next-gen GPU workloads. NVL systems are pre-qualified for direct-to-chip cooling and support higher TDPs than standard air-cooled counterparts.

PG25 (Propylene Glycol 25%)

A common coolant mixture made of 25% propylene glycol and 75% deionized water. Used in liquid cooling systems for its low freezing point and corrosion resistance.

Rear-Door Heat Exchanger (RDHx)

A passive or active liquid-cooled panel mounted on the back of a server rack that captures and removes heat from exhaust air before it enters the data center environment.

TDP (Thermal Design Power)

The maximum amount of heat a component (like a CPU or GPU) is expected to generate under typical workloads is used to guide cooling system design.

TUE (Total Usage Effectiveness)

A metric that measures how efficiently energy is used in a data center. Unlike PUE, it includes both electrical and thermal efficiency, offering a broader view of energy performance.

About Maintech

Maintech is a global provider of IT infrastructure services, supporting data center, cloud, and end-user environments for mid-market and enterprise organizations. With more than 50 years of experience, Maintech manages hundreds of thousands of servers across 110+ data centers worldwide, operating through 15 regional hubs and 100+ forward stocking locations with an in-country direct labor workforce. Maintech is trusted by four of the top ten global banks and maintains a vendor-neutral approach across all major technology platforms.

The company is ISO 9001 certified and a DCA Corporate Partner.

Continue the conversation

Global Headquarters:

Maintech Incorporated, 100 Walnut Avenue, Suite 600, Clark, New Jersey, 07066

contact@maintech.com

+1 (800) 426-8324

EMEA Office:

Maintech Europe LTD, 16 St Johns Lane, Farringdon, London, UK EC1M 4BS

+1 (800) 426-8324

APAC Office:

Maintech SG PTE LTD, 12 Woodlands Square #02-75 Woods Square Tower 1, Singapore 737715

+1 (800) 426-8324

www.maintech.com