Netceed

White Paper

Liquid cooling

The next wave of data center evolution

netceed.com

AI computing is pushing data centers to a boiling point

While graphics processing units (GPUs) provide the scorching speeds companies and users want, the servers consume substantial energy and give off immense levels of heat. Air cooling simply can't keep up. Data centers that fail to evolve risk extreme overspending—or worse, overheating.

In the AI arms race, liquid cooling is a must. It meets the cost, ESG, and performance demands vital to resilient performance. But there's no one-size-fits-all solution. Only a partner that understands the intricacies can help you identify and execute the ideal approach in a timely manner.

Table of contents

Start with our deep dive into the three liquid cooling methodologies. Discover their unique strengths and implementation considerations. With expert insights guiding the way, you can seize every cutting-edge advantage.

Heat from GPU racks is surging past air cooling's limits

As AI's prominence grows, so does data center rack density. From 2016 to 2024, density doubled from 6.1 to 12 kW.[1] In hyper-scale data centers, the future generation of chips could see rack density reaching 250kW. The current NVIDIA GB200/Blackwell, which features 5nm chips, already achieves 130kW/rack density. Increased workloads will drive a 160% rise in data center power demands by 2030.[2]

AI consumes electricity at unprecedented rates[3]

Google search
0.3 watt-hours

ChatGPT query
2.9 watt-hours

Water has a **heat-carrying capacity 3500x higher** than air.[6]

Fans and AC can't handle the heat

Traditional air cooling
3-15 kW per rack[4]

Row-based air cooling with containment
5-30 kW per rack[4]

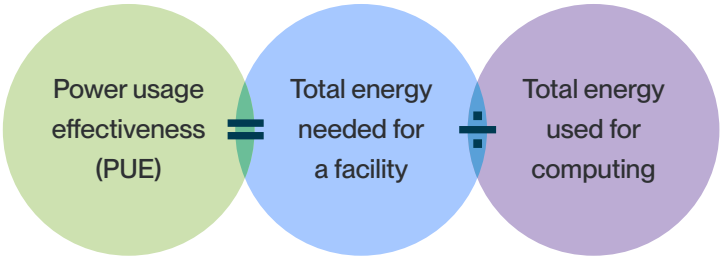
Liquid cooling
50-200 kW per rack[5]

Lower OpEx by going liquid

Air cooling isn't just a weaker approach. It's much more expensive. In older data centers that depend on fans and AC to protect servers, cooling eats up 40% of electricity demand.

40% of data center electricity goes to air cooling.[4]

Liquid cooling offers the opportunity to optimize power usage effectiveness (PUE), a ratio that denotes how much energy goes toward actual computing, to bring down overall operating expenses (OpEx).



The closer your score is to 1, the better. Liquid-cooled solutions usually reach around 1.10-1.15. Air-cooled data centers achieve PUEs as low as 1.5 in ideal conditions but can get as high as 2.9.[7,8]

Air cooling's poor PUE compounds OpEx

Let's say you pay $0.12/kWh for electricity. Even an efficient air-cooled approach could cost 3x more every month. If the system is deficient, the difference could skyrocket to 6x the cost of liquid cooling.

	Liquid	**Air** (lower range)	**Air** (upper range)
PUE	~1.125	1.5	2.9
Cooling energy per 100 kWh of total energy	11 kWh	33 kWh	65 kWh
Cost per KWh	$0.12	$0.12	$0.12
1-month cooling spend per 100 kWh of total energy	$963.60	$2,890.80	$5,694.00

Liquid cooling is essential to ESG

AI alone will consume 8% of U.S. power by 2030.[3] Liquid cooling's superior thermodynamics will be key to complying with sustainability regulations and achieving net-zero goals.

Data center electricity usage by 2026[4]

United States
>260 TWh

Worldwide
>1,000 TWh

Data centers can explore partnerships with utilities to apply advanced metering infrastructure and take part in demand response plans that protect electrical grids. There are even opportunities to design smart buildings and recycle water as well as the heat produced by servers. Taking extra steps to reduce waste will satisfy regulators and investors alike.

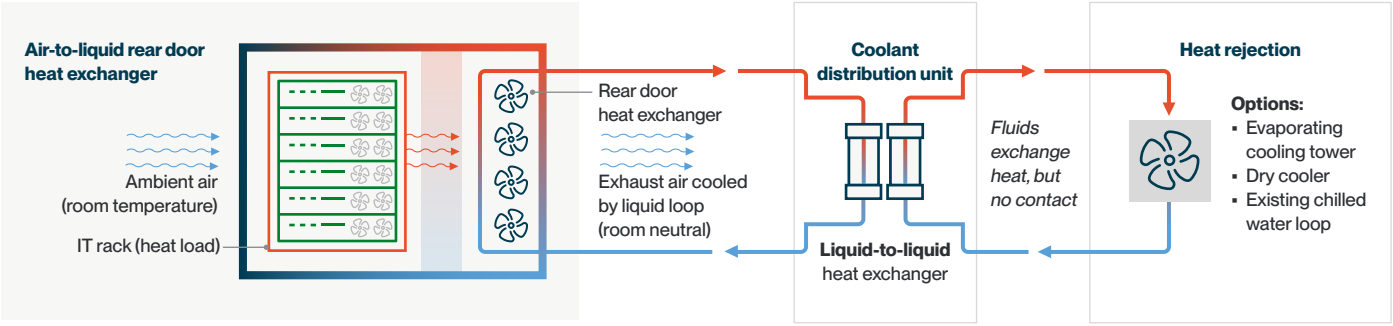
The three types of liquid cooling

1. Rear door heat exchangers (RDHx)

RDHx use both air and liquid to cool down servers. Units are mounted to the back of racks, and fans move air through the server and RDHx, picking up the heat along the way. Water running through coils within the RDHx unit absorbs the heat. Cooled air flows out the back of the exchanger and the water is piped back into the cooling distribution unit.

Active RDHx: Fans on the exchanger help carry air away from the rack.

Passive RDHx: No additional fans are included in the mounted unit.

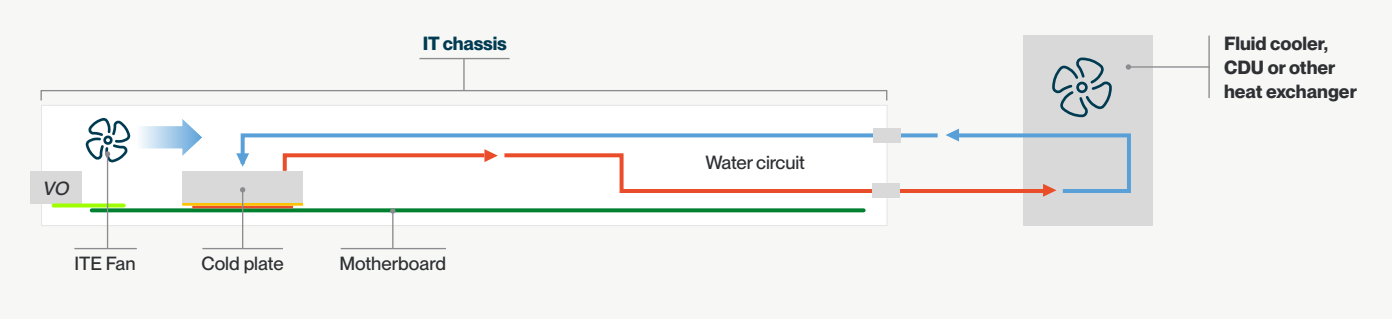


2. Direct-to-chip (DTC) cooling

DTC systems mount a cold plate right on top of the CPU or GPU. Non-conductive coolant flows through the cold plate, where a thermal interface conducts heat from the chip to the coolant. The coolant continues back into the plumbing system to dissipate the heat and be re-cooled for reuse.

Single-phase DTC: Coolant remains liquid during the entire process.

Two-phase DTC: Coolant boils in the cold plate at the point of contact. The resulting vapor carries the heat away and is eventually condensed back to liquid.

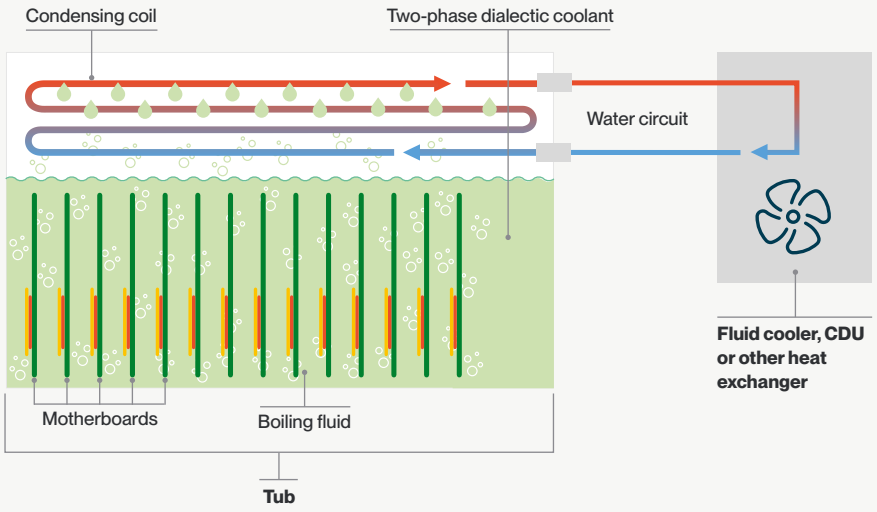


3. Immersion cooling

Entire racks are completely submerged in a conductive, dielectric liquid inside a sealed tank that connects to the chilled water loop. The liquid solution varies, but usually involves some type of glycol and other chemicals that may be subject to strict regulations. Immersion enables a more consistent temperature, ensuring IT components aren't subjected to sudden spikes. Like DTC cooling, immersion cooling can be done in a single- or two-phase approach.

Traditional: Immersion racks are placed inside data centers alongside or in place of air- and other liquid-cooled racks.

Modular: Self-contained units that are equipped to run a small number of servers and operate like "mini data centers." Their smaller footprint enables them to fit where conventional centers can't, opening avenues to optimal power usage and performance latency.



Unpacking the mysteries of liquid cooling strategies

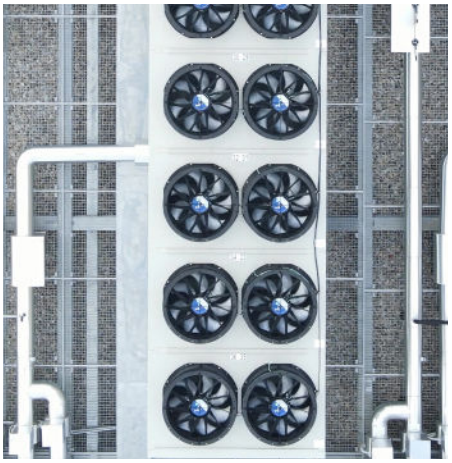
There's no universal choice or playbook

Each approach offers the ability to add cost-efficient processing power. Determining the best option comes down to the specifics:

- The GPUs or CPUs you plan to use
- Existing data center infrastructure
- CapEx and OpEx limitations
- Electricity and fiber costs
- Current lease agreement (if applicable)

No matter the path, there are several key factors to consider for each venture.

Retrofitting: Adding RDHx and/or DTC to existing servers



Start with potential power availability

Determine the number of GPUs you can run sustainably based on electricity limits. Power constraints can help you form parameters around cooling type(s), placement, plumbing, and more.

Reengineer your setup to blend approaches

It may be more economical to reserve liquid cooling for hotter GPUs while using air cooling for traditional storage. Plus, GPU racks are heavy. Some facilities will require flooring upgrades to accommodate the weight.

Prioritize plumbing so you only do it once

Plumbing is expensive, intricate work. Confirm how many cooling distribution units you'll need so you can execute one cost-effective installation.

Get your landlord involved

If you're nearing a lease renewal, you may be able to negotiate plumbing upgrades or increased power availability.

Greenfield builds: Starting fresh with future-proofed solutions

Keep one eye on the horizon

Build around the specific GPUs you intend to use to make the most of your initial investment. But don't paint yourself into a corner. Leave flexibility for upgrades and new technology to avoid becoming obsolete.

Balance power availability and fiber connectivity

Since power dissipates in transit, proximity to your supplier can keep energy costs down. At the same time, fiber installation represents a large investment (and in some places, fiber isn't available at all). Consider locations that strike a healthy equilibrium.

Be open to mixing it up

Consider ways that RDHx, DTC, and immersion cooling can work side by side. A combination of approaches could lead to the right balance of cost and cooling power.

Modular immersion cooling: Installing fully kitted 100kW units

Seek out a prime location

In AI, mere milliseconds separate the leaders from the afterthoughts. Modular units can fit where traditional data centers can't, so you may be able to get closer to the network and reduce latency.

Make the most of your power

Low-usage areas could ensure a steady, affordable supply. If your unit will be in a more crowded space, there may be opportunities to make right-of-way agreements that prioritize electricity access.

Research local regulations

There are multiple immersion coolants that successfully prevent overheating. But most contain toxic chemicals subject to varying regulations around the globe. Responsible usage and disposal must be a strategic consideration.

Prepare for repairs

Immersion tank maintenance is incredibly intricate compared to typical server racks. Prepare to invest in specialized training for staff, or find a partner with proven immersion cooling experience.

Four features of the ideal liquid cooling partner

1 **Value-added distributor:** You want more than a partner for dropping in new equipment: You need one that can provide end-to-end kitting, integration, and procurement. Seek out proven problem-solvers who will understand your unique needs. They'll be able to design and implement bundled solutions that align with your specific goals.

2 **Global reach:** Players in competitive markets such as AI development can't afford to wait: Hit the ground running with a partner that can tap into established manufacturer relationships and delivery networks.

3 **Technology and vendor agnostic:** There is no reason to lock into a single approach or manufacturer: Your partner should provide solutions built around your data center specifications, not a particular brand.

4 **Forward-looking:** Data centers evolve quickly: If your distributor can't keep up, you'll always be lagging. Ensure that the firm you partner with can navigate uncharted territory.

Bring your partner in early

Involve your partner in the design, planning, and engineering phases. They'll have insights into potential installation hurdles and cost optimization opportunities that can reshape your project.

Netceed: Exceed every expectation

Create a cost-optimized, sustainable liquid cooling network with a partner that puts you first. We'll handle the complexities of sourcing and implementing fit-for-purpose solutions through our:

- End-to-end design, logistics, and installation services
- State-of-the-art global distribution network
- Unmatched technical expertise

80 locations in 19 countries

1,500 sourcing and supply partners

2,000+ dedicated team members

30 years of industrial experience

Visit **netceed.com/liquidcooling** to connect with our liquid cooling specialists.

Netceed

Visit **netceed.com/liquidcooling** to connect with our liquid cooling specialists.

Sources

[1] AFCOM State of the Data Center Report

[2] Liquid Cooling: A Year in Review

[3] AI is poised to drive 160% increase in data center power demand

[4] Data centers warm up to liquid cooling

[5] Data Center Liquid Cooling for High-Density Compute

[6] New Liquid Cooling Designs Refine Options for Data Centers

[7] The Best Liquid Cooling Solution For Your Data Centre

[8] Enough hot air: the role of immersion cooling

Netceed is a leading one-stop shop for all materials, distribution, and logistics solutions for the broadband and telecommunications industry, supporting HFC, FTTH, FTTx, wireless, 5G/mobile, and data center technologies. We're redefining what a materials distributor and value-added services provider can be and do for our customers. But we're more than a distributor, we're a trusted partner for technical and logistics expertise.

Netceed is a global leader in distribution, logistics, technical services and product design supporting cable operators, telecommunications service providers, and contractors alike since 1993 with extensive industry expertise. Netceed has operations that span 19 countries and supports over 19,000 customers worldwide, which gives us a unique advantage of global purchasing power, driving costs down for our customers.

225 Raritan Center Parkway | Edison, NJ 08837 | 1-833-NETCEED | hello.us@netceed.com | netceed.com