WHITEPAPER · DATA CENTERS & INFRASTRUCTURE

Liquid-Cooled Hyperscale Data Centers

A Thermal, Energy & Sustainability Analysis

An engineering and economic framework for cooling AI-class racks — heat transport from die to facility loop, the efficiency metrics that actually matter (PUE, WUE), and the sustainability trade-offs of going wet.

AUTHORS
Root Digit Infrastructure Lab

PUBLISHED
2026

VERSION
1.0

CLASSIFICATION
Public

rootdigit.com

WHITEPAPER

Abstract & Contents

ABSTRACT

The arrival of dense AI accelerators has broken the assumptions on which a generation of air-cooled data centers was built. Where a conventional enterprise rack drew 5–10 kW, an AI training rack now routinely demands 40–120 kW, with roadmaps pointing beyond. At those densities air is no longer a sufficient heat-transport medium: the volume of cold air required, and the fan power consumed to move it, become physically and economically untenable. Liquid cooling — rear-door heat exchangers, direct-to-chip cold plates, and immersion — is the response. This whitepaper presents a thermal and economic analysis of liquid-cooled hyperscale facilities. We trace heat from the silicon die through the coolant loop, the cooling distribution unit (CDU), and the facility water loop to final rejection; define the efficiency metrics that govern operating cost (PUE, partial PUE, and WUE); and compare air, direct liquid cooling (DLC), and immersion across a transparent, modeled scenario. The findings indicate that well-engineered liquid cooling can lower facility PUE from a typical air-cooled 1.4–1.6 toward ~1.1, support rack densities an order of magnitude higher, and unlock high-grade waste-heat reuse — but that water consumption, coolant global-warming potential, and embodied carbon must be designed for, not assumed away.

KEY FINDINGS AT A GLANCE

- Air cooling becomes impractical above roughly 30–50 kW per rack; cooling-fan and chiller energy grow faster than the IT load they serve.
- Direct-to-chip and immersion cooling can support 100+ kW racks and push modeled facility PUE toward ~1.1, versus 1.4–1.6 for typical air designs.
- The dominant efficiency lever moves from the chiller plant to the coolant approach temperature — warmer coolant enables free cooling for more of the year.
- Sustainability is a balance: liquid cooling can cut water use with dry coolers and enable waste-heat reuse, but two-phase refrigerants carry high GWP that must be managed.

Contents

SECTION ONE

1. Introduction

For three decades the data center was an air-handling problem. The defining design question was how to move enough cold air past racks of a few kilowatts each. That era is ending, and the cause is the accelerator.

The generative-AI build-out has changed the thermal profile of the data center more abruptly than any prior workload transition. A single modern training accelerator can dissipate 700–1,000 watts at the package, and a rack densely populated with eight-GPU servers and their attendant switching now draws power that would once have served an entire row. Rack densities that sat comfortably at 5–10 kW through the 2010s have climbed to 40–60 kW for inference clusters and 80–120 kW for the densest training pods, with vendor roadmaps openly discussing 250 kW and beyond. The IT load has not merely grown; its *concentration* has grown, and concentration is precisely what air struggles with.

The physics are unforgiving. The rate at which a coolant can remove heat scales with its mass flow, its specific heat capacity, and the temperature rise it can tolerate. Air has a low volumetric heat capacity, so removing tens of kilowatts from a single rack demands enormous air volumes and the fan power to move them. Beyond roughly 30–50 kW per rack, the air-side solution becomes self-defeating: the energy spent moving and chilling air begins to rival the savings, hot-aisle temperatures become difficult to contain, and the silicon throttles to protect itself. The accelerator that the facility exists to serve is the first casualty of inadequate cooling.

Liquid offers a way out. Water carries roughly three to four thousand times the heat per unit volume of air, and engineered dielectric fluids are not far behind. By bringing liquid to — or into — the heat source, operators can remove an order of magnitude more heat from the same footprint while spending far less energy doing so. Liquid cooling is not new to high-performance computing, but the economics that once confined it to supercomputers now apply to mainstream AI infrastructure, and that is why it is moving rapidly into hyperscale practice.

This whitepaper analyses that transition through three lenses. The first is **thermal engineering**: how heat actually travels from the die to the outside air, and where the avoidable losses hide. The second is **energy efficiency**: the metrics — power usage effectiveness (PUE), its partial variant, and water usage effectiveness (WUE) — that determine the operating cost and environmental footprint of a facility. The third is **sustainability**: water, waste heat, refrigerant chemistry, and the embodied carbon of the cooling plant itself, which a power-only view systematically ignores.

Our purpose is decision-useful clarity for the infrastructure leaders who must choose a cooling architecture for a facility that will operate for fifteen years. We compare air, direct liquid cooling, and immersion on a single transparent model, state every assumption, and present results as the shape of the trade-offs rather than as a verdict for any one site. The figures throughout are illustrative and modeled from representative

parameters, not measurements of a specific facility; Section 6 makes every input explicit so the reader can substitute their own.

SECTION TWO

2. Cooling Technologies

There is no single liquid-cooling technology, but a ladder of them, each bringing the coolant closer to the heat and each removing more of it. The choice is an engineering and economic trade, not a binary upgrade.

At the bottom of the ladder sits **air cooling** in its familiar forms — raised-floor or in-row computer-room air handlers with hot-aisle/cold-aisle containment. It is mature, cheap to deploy, and demands nothing unusual of the server. Its ceiling is thermal: contained air designs serve perhaps 15–25 kW per rack comfortably and can be pushed toward 30–50 kW with aggressive containment and high airflow, but only by spending fan and chiller energy that erodes efficiency.

A first step toward liquid keeps the server air-cooled but moves the heat exchanger to the rack itself. **Rear-door heat exchangers (RDHx)** replace the cabinet's back door with a liquid-fed coil; server fans push hot exhaust through it and the air leaves the rack at near room temperature. RDHx is attractive because servers need no modification, yet it captures heat at the rack and can support 40–80 kW. Its limit is that heat still travels from chip to air to coil — an extra thermal hop with its own losses.

Direct-to-chip cooling, also called direct liquid cooling (DLC), closes that gap. A cold plate sits directly on the CPU and GPU packages, and a coolant — typically treated water or a water-glycol mix — flows through it, carrying heat away at the source. DLC commonly captures 70–80% of the rack's heat into liquid, with the remainder (memory, voltage regulators, drives) still handled by a modest air loop. It supports 100 kW racks and beyond, retains a conventional serviceable server form factor, and is the dominant choice for current AI deployments.

At the top of the ladder is **immersion cooling**, in which entire servers are submerged in a dielectric fluid. *Single-phase* immersion circulates a liquid that stays liquid, drawing heat to an external exchanger; it captures essentially all the heat, eliminates server fans, and tolerates very high density. *Two-phase* immersion uses a fluid that boils at the chip surface, exploiting the latent heat of vaporisation for an exceptionally high heat-transfer coefficient — but the engineered fluids involved have historically carried high global-warming potential and raised regulatory and supply concerns, which has slowed adoption.

Cooling technologies compared

Technology	Heat to liquid	Density supported	Trade-off
Air + containment	0%	15–50 kW/rack	Mature; fan/chiller energy caps efficiency
Rear-door heat exchanger	~70–95%*	40–80 kW/rack	No server change; extra air-to-coil hop
Direct-to-chip (DLC)	~70–80%	100+ kW/rack	Best balance; needs CDU + residual air loop
Single-phase immersion	~100%	Very high	No fans; tank handling, fluid cost
Two-phase immersion	~100%	Highest	Superior transfer; fluid GWP & supply risk

*RDHx captures rack-exhaust heat to liquid rather than chip heat directly; the captured fraction depends on coil sizing and approach temperature.

The practical conclusion is that the ladder is not a strict ranking — it is a menu matched to density, serviceability tolerance, and sustainability constraints. Most hyperscale AI deployments in 2026 land on direct-to-chip with a residual air loop, reserving immersion for the densest or most space-constrained sites and retaining air for the lower-density, latency-tolerant remainder of the estate.

SECTION THREE

3. Thermal Engineering: Die to Facility Loop

Cooling a chip is a relay of temperature differences. Heat will only flow from one stage to the next if each is colder than the last, and every junction in that relay costs a few degrees that the facility must ultimately pay for in energy.

Begin at the source. A modern accelerator package dissipates its power across a die of roughly a few square centimetres, producing a heat flux on the order of 50–150 W/cm² locally and far higher at hot spots. That heat must cross the silicon, the thermal interface material, and the integrated heat spreader before it ever reaches the coolant. Each interface adds thermal resistance, and the product of total resistance and heat load sets the junction temperature — which must stay below the manufacturer's limit (commonly 90–105 °C) or the part throttles. The engineering goal of every cooling stage is to minimise the temperature rise per watt removed.

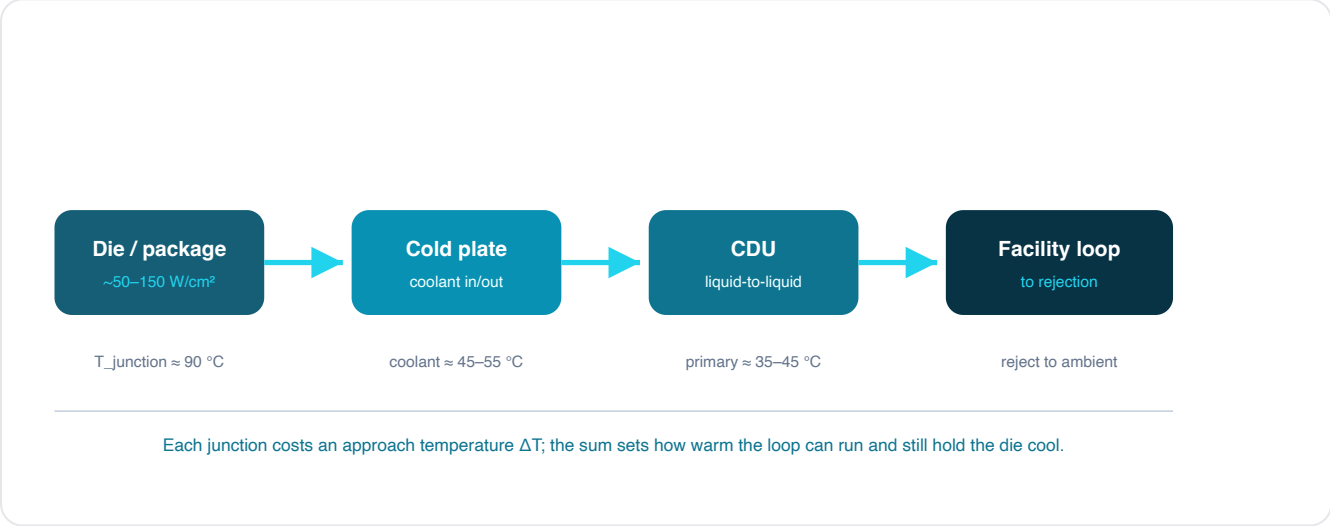

Figure 3.1 — The thermal relay from die to facility loop. Heat flows only down the temperature ladder; minimising each ΔT lets the loop run warmer and the chiller work less.

From the cold plate the heat enters the **technology cooling system (TCS)** loop — the closed circuit of treated coolant that serves the racks. This loop is deliberately isolated from the building's water by a **cooling distribution unit (CDU)**, a liquid-to-liquid heat exchanger with pumps, filtration, and controls. The CDU's job is twofold: it transfers heat from the clean, controlled TCS loop into the facility water loop, and it isolates the sensitive in-rack circuit from the larger building system, protecting it from contamination, pressure transients, and corrosion.

The key design parameter at every exchange is the **approach temperature** — the difference between the hot fluid entering and the cold fluid that cools it. A tight approach (a few degrees) means a large, expensive exchanger but lets the system run with warmer coolant; a loose approach is cheaper to build but forces colder, more energy-intensive supply. Because the cooling energy is dominated by how cold the facility ul-

timately has to make its water, approach temperatures are where capital cost and operating cost trade against each other most directly.

The facility water loop carries the collected heat to final rejection. Here the architecture branches. In a **chilled-water** design a mechanical chiller actively cools the loop — effective in any climate but energy-hungry. In a **free-cooling** design the loop rejects heat through cooling towers (evaporative) or dry coolers (sensible only) whenever the ambient is cold enough, with the chiller as backup. The decisive advantage of liquid cooling is that it permits much warmer supply temperatures than air: where air systems may need 18–24 °C supply, a DLC loop can hold the die safe with 30–45 °C coolant. That higher set-point dramatically widens the window in which free cooling alone suffices, and free cooling is where the largest efficiency gains live.

> **DESIGN PRINCIPLE**
>
> Run the loop as warm as the silicon safely allows. Every degree of usable warmth in the coolant extends the hours per year the facility can free-cool and shrinks the chiller's duty — the single largest lever on operating energy.

SECTION FOUR

4. Energy & Efficiency Metrics

A facility cannot be improved on what it does not measure. The metrics that govern data-center efficiency are deceptively simple to define and surprisingly easy to game — which is why understanding what they include matters as much as the number itself.

The headline metric is **power usage effectiveness (PUE)**: the ratio of total facility energy to the energy delivered to the IT equipment. A PUE of 1.0 would mean every watt entering the building reached a server; in practice the overhead — cooling, power conversion losses, lighting — pushes it higher. Air-cooled enterprise facilities commonly sit at 1.4–1.6; well-run liquid-cooled facilities can approach 1.1. The arithmetic is direct.

```
# Power Usage Effectiveness and its partial variant
PUE     = total_facility_energy / IT_equipment_energy
overhead = total_facility_energy - IT_equipment_energy
# overhead = cooling + power-conversion losses + lighting + ...

# Partial PUE isolates one subsystem (e.g. the cooling plant)
pPUE_cooling = (IT_energy + cooling_energy) / IT_energy

# Water Usage Effectiveness (litres per kWh of IT energy)
WUE = annual_water_litres / annual_IT_energy_kWh
```

Listing 4.1 — The core efficiency identities. PUE measures energy overhead; partial PUE isolates a subsystem; WUE captures the water cost the energy number hides.

PUE is powerful but blunt. Because it is a ratio, it improves whenever the IT load rises even if cooling efficiency does not — a busy facility looks better than an idle one of identical design. It also conceals *where* the overhead goes. **Partial PUE** (pPUE) addresses this by isolating a single subsystem, most usefully the cooling plant, so that a chiller upgrade or a switch to liquid can be evaluated on its own contribution rather than lost in the building total.

The cooling overhead itself is not monolithic. In an air-cooled facility it divides roughly among the chiller compressors (the largest share), the cooling-tower or condenser fans, the chilled-water pumps, and the air-handler fans that move air through the white space. Liquid cooling attacks this distribution at its root: by removing heat at the chip it largely eliminates the white-space air-mover energy, and by tolerating warm coolant it slashes or eliminates compressor energy for much of the year. What remains is comparatively modest pump and dry-cooler fan power.

The metric that air-centric analysis most often omits is **water usage effectiveness (WUE)** — the litres of water consumed per kilowatt-hour of IT energy. It matters because the cheapest path to a low PUE has historically been evaporative cooling, which trades electricity for water. A facility can post an enviable PUE

while consuming millions of litres a year through its cooling towers. WUE makes that trade visible, and in water-stressed regions it can be the binding constraint regardless of how good the energy number looks.

> **MEASUREMENT CAUTION**
>
> PUE and WUE pull in opposite directions. Evaporative cooling lowers PUE by spending water (raising WUE); dry coolers lower WUE by spending electricity (raising PUE). A credible efficiency claim states both, at a defined IT load, over a full annual cycle — never a single best-day PUE in isolation.

The metrics therefore form a system, not a leaderboard. A liquid-cooled facility designed for warm coolant and dry-cooler rejection can deliver a low PUE *and* a low WUE simultaneously — the outcome that genuinely matters — but only if the design is evaluated against both numbers across the full range of seasonal conditions it will face.

SECTION FIVE

5. Sustainability & Carbon

Efficiency and sustainability are related but not identical. A facility can be energy-efficient and still consume scarce water, leak high-impact refrigerants, or carry an oversized embodied-carbon burden. A complete sustainability view accounts for all four.

Water. The most immediate sustainability lever is water consumption. Evaporative cooling is thermodynamically attractive — it exploits the latent heat of vaporisation to reject large heat loads with little electricity — but it consumes water continuously, and at hyperscale that consumption is measured in millions of litres per facility per year. In water-stressed regions this is increasingly the gating constraint on whether a facility can be built at all. Liquid cooling helps in two ways: by enabling warm-water loops it widens the window for closed-loop dry coolers that consume essentially no water, and by raising heat-rejection temperatures it makes those dry coolers practical where they once were not. The trade is electricity for water, and in many climates and grids it is now the right trade.

Waste-heat reuse. Air-cooled facilities reject heat at temperatures too low to be useful — lukewarm exhaust is hard to do anything with. Liquid cooling changes this. Because DLC and immersion can run warm loops, they reject heat at 45–60 °C, a grade high enough to feed district heating networks, warm greenhouses, or preheat industrial processes. Several northern-European facilities already export waste heat to municipal heating systems. Reuse converts a disposal cost into a community benefit and improves the total energy balance, though it requires a nearby heat off-taker and the distribution infrastructure to reach them — a planning consideration best addressed at site selection.

Sustainability levers and their trade-offs

Lever	Benefit	Cost / caveat
Closed-loop dry coolers	Near-zero water use	More electricity in hot climates
Warm-water DLC / immersion	Enables free cooling & heat reuse	Higher CDU and plumbing capex
Waste-heat export	Offsets external fossil heating	Needs nearby off-taker & piping
Single-phase dielectric fluid	Full heat capture, low GWP	Fluid cost, tank serviceability
Two-phase fluid	Highest heat transfer	High-GWP risk; regulatory pressure

Refrigerant and coolant chemistry. The fluids themselves carry an environmental signature. Mechanical chillers and two-phase immersion systems can use refrigerants and engineered fluids with

high **global-warming potential (GWP)** — some legacy fluorinated compounds have GWP values in the thousands, meaning a small leak has an outsized climate impact. Regulatory pressure (notably the phasing-down of certain fluorinated gases) and supply-chain restrictions are pushing the industry toward low-GWP refrigerants and toward single-phase dielectric fluids or plain treated water wherever the thermal duty allows. Coolant selection is therefore not only a thermal decision but a climate and compliance one.

Embodied versus operational carbon. Finally, a credible carbon account separates the two halves of a facility's footprint. *Operational* carbon comes from the electricity (and any water treatment) consumed over the facility's life and is reduced by efficiency and clean-grid procurement. *Embodied* carbon is locked in at construction — the steel, concrete, copper, CDUs, and the cooling plant itself — and is paid once regardless of how the facility runs. Liquid cooling shifts this balance: it can add embodied carbon (more plumbing, exchangers, and fluids) while reducing operational carbon (less cooling energy over fifteen years). On a clean grid the embodied share grows relatively larger and deserves explicit accounting; on a carbon-intensive grid the operational savings dominate and liquid cooling pays back its embodied premium quickly.

> **WHOLE-LIFE VIEW**
>
> Optimising any single number — PUE, water, or capex — can quietly worsen another. Sustainable design optimises the whole-life total: energy, water, refrigerant impact, and embodied carbon together, across the realistic operating life of the facility.

SECTION SIX

6. Methodology: The Comparative Model

A comparison is only as honest as its assumptions. This section states the model, its scenarios, and its inputs in full so that the reader can substitute their own figures and reproduce the results.

We model a single reference white-space of fixed IT load and compare three cooling architectures serving it: a contained **air** baseline, a **direct-to-chip (DLC)** design with a residual air loop, and a **single-phase immersion** design. The IT load is held constant across scenarios so that differences arise from cooling architecture alone, not from workload. For each architecture we compute the cooling overhead, the resulting PUE, the supportable rack density, the annual water consumption, and a five-year total cost of ownership combining capital and energy.

```
# Comparative cooling model (per white-space, constant IT load)
PUE        = (IT_power + cooling_power) / IT_power
cooling_power = chiller_kW + pump_kW + fan_kW          # varies by architecture
free_cool_hours = hours_ambient_below(coolant_setpoint)
annual_energy = IT_power × 8760 × PUE_seasonal
annual_water  = WUE × IT_power × 8760

# Five-year total cost of ownership
TCO = capex_cooling + (5 × annual_energy × tariff)
                    + (5 × annual_water × water_rate)
                    + (5 × maintenance)
```

Listing 6.1 — The comparative model. Every term is an input the reader can set; defaults are tabulated below. PUE is computed seasonally to credit free-cooling hours.

Modeled input assumptions (illustrative defaults)

Parameter	Value	Parameter	Value
Reference IT load	2.0 MW	Electricity tariff	$0.10 / kWh
Air coolant set-point	18–24 °C	DLC coolant set-point	30–45 °C
Air-cooled PUE	1.45 (modeled)	DLC PUE	1.12 (modeled)
Immersion PUE	1.08 (modeled)	Water rate	$2.50 / m³
Density: air / DLC / immersion	25 / 110 / 150 kW	Horizon	5 years

These figures are representative of a temperate-climate hyperscale white-space and are *illustrative, not field-measured*. The PUE values assume competent design and a climate with meaningful free-cooling hours; a hot, humid site would compress the gap between air and liquid, while a cold site would widen it. The purpose is a transparent, adjustable model — not a claim about any specific facility. The charts in Section 7 are generated from these assumptions.

SCOPE NOTE

The model isolates cooling architecture at constant IT load and temperate climate. It deliberately excludes grid carbon intensity, waste-heat revenue, and embodied-carbon accounting, all of which are site-specific and discussed qualitatively in Sections 5 and 8 rather than reduced to a single modeled number.

SECTION SEVEN

7. Findings: PUE, Density & TCO

Three results recur across the modeled scenarios: liquid cooling lowers PUE substantially, it raises supportable density by roughly an order of magnitude, and its higher capital cost is recovered through energy savings over the operating life.

~1.1	~110 kW	~25%	~3 yr
MODELED LIQUID-COOLED PUE	DLC RACK DENSITY SUPPORTED	FACILITY ENERGY SAVED VS AIR	COOLING-CAPEX PAYBACK

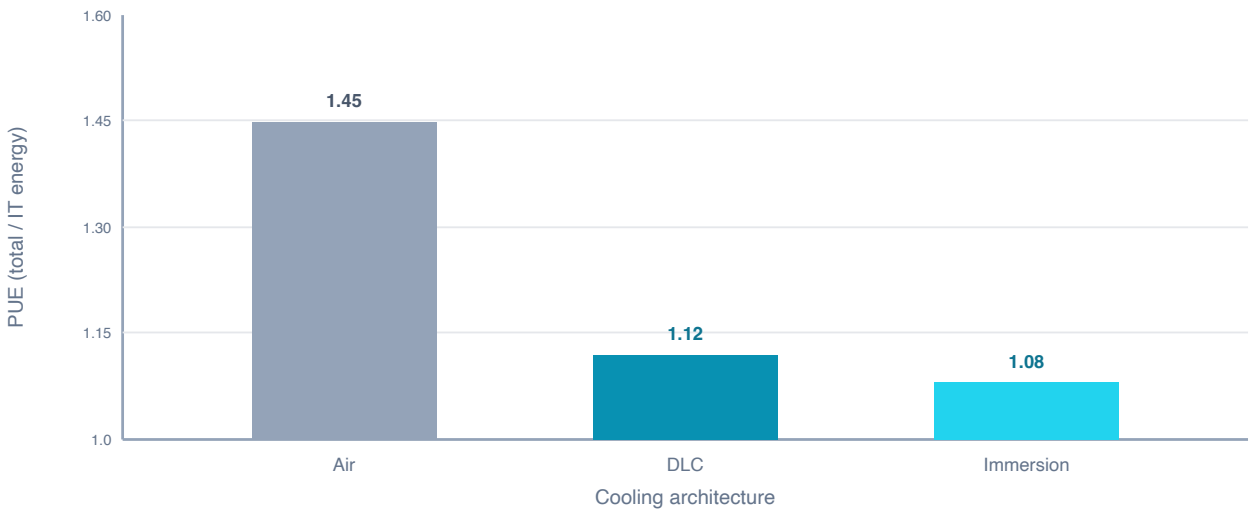

Figure 7.1 — Modeled facility PUE by cooling architecture. Lower is better; the air-cooled baseline carries roughly four times the overhead of single-phase immersion.

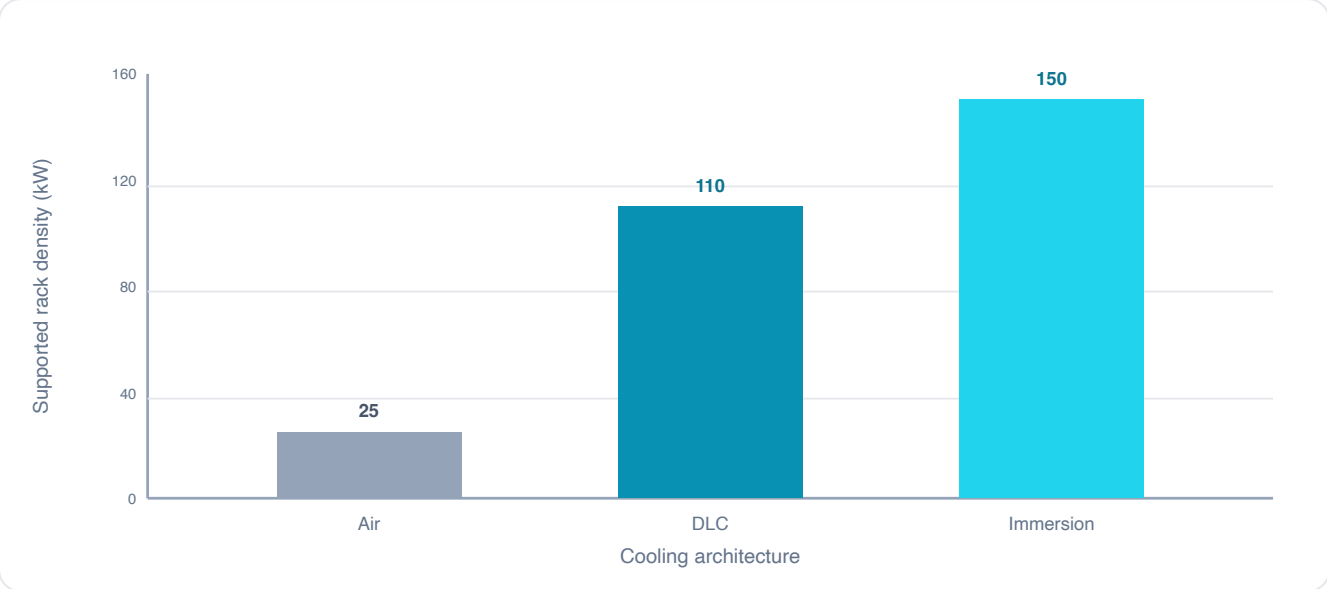

Figure 7.2 — Modeled supportable rack density by architecture. Liquid cooling lifts density by roughly an order of magnitude, compressing the same compute into far less floor.

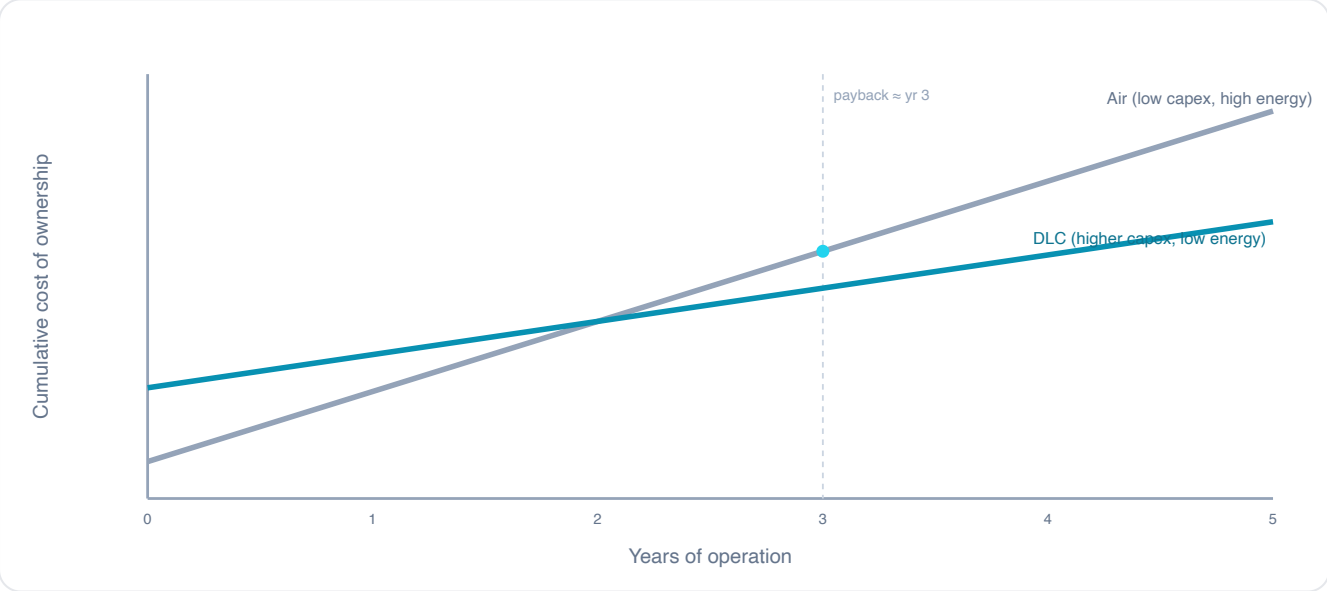

Figure 7.3 — Modeled cumulative cost of ownership. The DLC line starts higher (capital) but rises more slowly (energy); the curves cross near year three, after which liquid cooling is cheaper to own.

The PUE result (Figure 7.1) is the headline. In the reference model the air-cooled baseline carries an overhead of 0.45 (a PUE of 1.45), of which the chiller plant and white-space fans are the dominant components. Direct-to-chip cooling cuts that overhead to roughly 0.12 by eliminating most fan energy and enabling warm-water free cooling; single-phase immersion trims a little further by removing server fans entirely. The marginal gain from DLC to immersion is small relative to the leap from air to DLC — which is why most operators capture the large step first.

The density result (Figure 7.2) explains why the transition is happening regardless of energy economics. Air tops out near 25 kW per rack in this model; DLC supports roughly 110 kW and immersion 150 kW. For an AI cluster the implication is spatial: the same compute that would sprawl across a large air-cooled hall

fits into a fraction of the floor when liquid-cooled, with corresponding savings in building shell, land, and power-distribution runs.

The TCO result (Figure 7.3) addresses the obvious objection — that liquid cooling costs more to build. It does: the CDUs, manifolds, and plumbing raise capital cost. But the steeper energy slope of the air baseline overtakes that premium near year three in the reference model, after which the liquid-cooled facility is cheaper to own for the remainder of its life. Over a fifteen-year facility lifetime the cumulative advantage is substantial.

RESULT SUMMARY

- Modeled PUE falls from 1.45 (air) to ~1.12 (DLC) and ~1.08 (immersion); the largest single step is from air to direct-to-chip.
- Supportable rack density rises roughly an order of magnitude — from ~25 kW (air) to 110–150 kW (liquid) — compressing the same compute into far less floor.
- The cooling-energy share shifts from compressor-dominated to pump-and-dry-cooler-dominated, with warm coolant unlocking free-cooling hours.
- Higher cooling capital is recovered through energy savings near year three in the reference model; the advantage compounds over a fifteen-year life.

SECTION EIGHT

8. Discussion, Limitations & Deployment

A model is a lens, not a measurement. The results above should be read as structure — the shape of the trade-offs and the rank-order of the levers — rather than as precise predictions for any one facility.

Limitations. The comparative model holds IT load and climate constant to isolate cooling architecture; real facilities face seasonal swings, partial-load operation, and grid carbon that vary the picture materially. The modeled PUE values assume competent design and a temperate climate with meaningful free-cooling hours — a hot, humid site compresses the air-versus-liquid gap and may force evaporative assistance, while a cold site widens it. The TCO comparison excludes grid carbon pricing, waste-heat revenue, and embodied-carbon accounting, each of which is site-specific and can shift the crossover point in either direction. The density figures are representative ceilings, not guarantees; achievable density also depends on power distribution, structural floor loading, and serviceability.

Deployment considerations. Retrofitting liquid cooling into an existing air-cooled hall is materially harder than designing for it from the start: floor loading, leak detection, manifold routing, and CDU placement all want to be planned early. Serviceability changes too — DLC adds quick-disconnect fittings and leak-management discipline, while immersion changes how technicians physically handle servers. Operators should also plan for fluid management as an ongoing function: water treatment and filtration for DLC loops, and fluid inventory, monitoring, and end-of-life handling for immersion. None of these is prohibitive, but each is a competency that an air-only operations team will need to build.

Key risks. The dominant risks are organisational and design-timing rather than thermal. Choosing a cooling architecture after the building shell and power topology are frozen forecloses the warm-water, free-cooling design that delivers most of the benefit. Selecting a high-GWP two-phase fluid without a regulatory and supply-risk assessment can strand an asset. And optimising PUE in isolation can quietly raise water consumption to unsustainable levels in a stressed region. Each of these is avoidable with up-front, whole-life design.

> IN PRACTICE
>
> The most reliable predictor of a successful liquid-cooled facility is not the cooling technology chosen but whether the operator set the coolant temperature strategy and the rejection method — chiller, tower, or dry cooler — before the site and shell were locked. Those two decisions cascade into nearly every efficiency and sustainability outcome that follows.

Recommendations

1. **Design for the densest rack you will ever deploy.** Cooling architecture is hard to change later; size the loop, CDUs, and floor loading for the roadmap, not the first install.
2. **Run the loop as warm as the silicon safely allows.** Warm coolant is the master lever — it maximises free-cooling hours, shrinks the chiller, and enables waste-heat reuse.
3. **Report PUE and WUE together, annually.** A low PUE bought with millions of litres of water is not an efficient facility; measure both across a full seasonal cycle.
4. **Prefer single-phase dielectric or treated water** over high-GWP two-phase fluids unless the thermal duty genuinely requires it, and assess regulatory and supply risk before committing.
5. **Account for whole-life carbon, not just energy.** On a clean grid the embodied carbon of the cooling plant matters; on a dirty grid the operational savings dominate. Know which regime applies.

Conclusion

Air cooling carried the data center for thirty years, but the AI accelerator has moved heat density past the point where moving air is either physical or economic. Liquid cooling answers that challenge: it lowers facility PUE toward 1.1, lifts rack density by an order of magnitude, and — designed for warm loops and water-frugal rejection — can do so while cutting both energy and water and unlocking high-grade waste-heat reuse. The premium it adds in capital is recovered in a few years of energy savings and compounds over a facility's life. The operators who treat cooling as a whole-life design problem — set the coolant strategy and the rejection method up front, and report energy and water honestly — will cool the rack and reclaim the watts. Those who bolt liquid onto an air-shaped facility will capture a fraction of the benefit and pay for the rest.

References

1. The Green Grid — PUE: A Comprehensive Examination of the Metric (and the Water Usage Effectiveness companion metric).
2. ASHRAE Technical Committee 9.9 — Thermal Guidelines for Liquid-Cooled Data Processing Environments.
3. ASHRAE TC 9.9 — Thermal Guidelines for Data Processing Environments (air-cooling classes and allowable ranges).
4. Open Compute Project — Advanced Cooling Solutions: cold-plate and immersion design specifications.
5. IEC 30134-2 / 30134-9 — Information technology data centres: Key performance indicators (PUE and related KPIs).
6. Root Digit Infrastructure Lab — Internal comparative cooling-model methodology (2026).

-rootdigit®

Cool the rack.
Reclaim the watts.

Root Digit designs, integrates, and operates liquid-cooled data-center infrastructure — thermal engineering from die to facility loop, warm-water free-cooling strategy, CDU and immersion design, and the energy-and-water instrumentation that keeps PUE and WUE honest. We bring the model; you keep the efficiency.

Talk to our infrastructure engineers →

rootdigit.com

General: info@rootdigit.com · **Consultation:** rootdigit.com/contact/consultation
Dubai International Financial Centre (DIFC) · Wyoming, USA