DCS
DATACENTRE SOLUTIONS

ISSUE II 2026
AN ANGEL BUSINESS COMMUNICATIONS PUBLICATION
DATACENTRE.SOLUTIONS

SCALING LIQUID COOLING FOR THE AI ERA

An Interview with Andrew Whitmore,
Chief Commercial Officer at Motivair by Schneider Electric

Schneider Electric


ZincFive
BC 2 AI
Designed for AI.
No Tradeoffs in
Safety, Space,
or Sustainability.
LEARN MORE
zincfive.com

VIEWPOINT

BY PHIL ALSOP
EDITOR

AI in the data centre – no longer just a ‘nice to have’?

AI IS NO longer a futuristic add-on to the data-centre stack; it’s becoming the defining layer that separates merely functional facilities from truly optimised ones. For an industry under relentless pressure - energy constraints, rising densities, sustainability targets, and unyielding uptime expectations - intelligence isn’t a luxury. It’s the next competitive battleground.

Data centres already generate oceans of telemetry, but the sector has reached the limits of what human operators can meaningfully interpret. AI changes that equation. By ingesting and correlating thousands of signals - thermal behaviour, power fluctuations, component performance, workload patterns - it can surface insights that would otherwise remain buried. Predictive maintenance is the most immediate win. Instead of swapping components on fixed cycles or waiting for failures, operators can act on early-stage degradation signals. The result is fewer service interruptions, longer asset life, and a more efficient use of both hardware and human resources.

Operational optimisation is where AI’s impact becomes even more compelling. Intelligent workload placement, dynamic cooling adjustments, and real-time power tuning allow facilities to run closer to their true performance envelope without compromising resilience. For hyperscale operators, shaving even a fraction off PUE translates into enormous cost savings and measurable sustainability gains. For colocation providers, AI-driven efficiency becomes a differentiator in a market where margins are tight and customers increasingly scrutinise environmental performance.

Energy management may prove to be AI’s most transformative contribution. As grids strain and renewable availability fluctuates, data centres must become far more adaptive. AI can forecast power demand, optimise cooling strategies hour-by-hour, and even shift workloads to align

with greener energy windows. This isn’t just operational fine-tuning; it’s a pathway to meeting regulatory expectations and corporate sustainability commitments without sacrificing performance.

But perhaps the most significant shift is cultural. AI pushes data-centre teams to evolve from reactive operators to strategic orchestrators. Instead of firefighting, they can focus on long-term resilience, capacity planning, and innovation. The facility becomes a living, learning system - one that continuously refines itself.

The challenge now is adoption. AI requires clean data, trust in automation, and a willingness to rethink long-standing operational habits. Yet the direction of travel is unmistakable. The data centres that embrace intelligence will be the ones that scale sustainably, deliver consistent performance, and stay ahead of customer expectations in an increasingly demanding digital economy.

Contents

NEWS

DCS DATACENTRE SOLUTIONS

Editor
Philip Alsop
+44 (0)7786 084559
philip.alsop@angelbc.com

Senior B2B Event & Media Executive
Mark Hinds
+44 (0)2476 718971
mark.hinds@angelbc.com

Director of Logistics
Sharon Cowley
+44 (0)1923 690200
sharon.cowley@angelbc.com

Design & Production Manager
Mitch Gaynor
+44 (0)1923 690214
mitch.gaynor@angelbc.com

Graphic Design & Multimedia Assistant
Harvey Watkins
harvey.watkins@angelbc.com

Publisher
Jackie Cannon
+44 (0)1923 690215
jackie.cannon@angelbc.com

Circulation & Subscriptions
+44 (0)1923 690214
circ@angelbc.com

Directors
Scott Adams: CTO
Sukhi Bhadal: CEO

Published by: Angel Business Communications Ltd, 6 Bow Court, Burnsall Road, Coventry CV5 6SP. UK
T: +44 (0)2476 718970 E: info@angelbc.com

Angel BUSINESS COMMUNICATIONS

Datacentre Solutions is published 10 times a year on a controlled circulation basis in Europe, Middle East and Africa only. Subscription rates on request. All information herein is believed to be correct at time of going to press. The publisher does not accept responsibility for any errors and omissions. The views expressed in this publication are not necessarily those of the publisher. Every effort has been made to obtain copyright permission for the material contained in this publication. Angel Business Communications Ltd will be happy to acknowledge any copyright oversights in a subsequent issue of the publication.

The paper used within this magazine is produced by chain of custody certified manufacturers, guaranteeing sustainable sourcing.
ISSN 2756-1143 (Online)

North America data centre market: capacity growth and frontier expansion

The North American data centre industry faces record-low vacancy and surging expansion into new markets. Texas is poised to surpass Virginia by 2030.

THE NORTH AMERICAN data centre sector is undergoing a period of significant change, reflected in a record-low vacancy rate of 1% for the second consecutive year. This position highlights continued expansion and ongoing structural shifts across the market.

A recent report from JLL outlines these developments. It states that 64% of the 35 gigawatt (GW) construction pipeline now extends beyond traditional mature markets. As a result, Texas is positioned to potentially overtake Virginia as the world's largest data centre market by 2030.

JLL's updated report reflects an evolution in its market analysis, including expanded coverage of hyperscale activity and the addition of more than 40 frontier markets.

Available capacity remains limited to small and fragmented blocks, leaving limited flexibility for large-scale deployments. As a result, tenants are increasingly securing space for delivery in 2027 or 2028, demonstrating sustained demand within the sector.

More than half of the current construction activity is taking place in frontier markets. Texas accounts for 6.5 GW of ongoing construction, supporting projections that it could become the largest global market by 2030. Other states, including Tennessee, Wisconsin and Ohio, are also benefiting from strong energy availability and business-friendly operating conditions.

Supply constraints continue to support rent growth. Data centre rents increased by 9% in 2025, in line with a five-year compound annual growth rate (CAGR) of 10%, reinforcing consistent pricing momentum across the sector.

These conditions have enabled landlords to achieve higher rental levels on renewals, while tenants are experiencing increased costs when negotiating new leases.

Hyperscalers are driving significant investment, with $710 billion of planned capital expenditure for 2026. However, infrastructure constraints, including grid connection timelines that average around four years, are influencing how major tenants plan and deliver new data centre capacity.

This situation is contributing to expansion into frontier markets with greater access to power resources. Collaboration with utility providers is becoming more common as developers explore approaches such as flexible load planning and phased power deployment to accelerate connections.

The combination of strong debt availability, growing AI-related demand and increased investment activity has positioned data centres as a prominent segment within North American commercial real estate. Asset-backed security volumes exceeded $17 billion in 2025, nearly doubling from the previous year, while single-asset single-borrower lending volumes rose to over $11 billion, representing a threefold increase. These figures reflect increasing complexity and maturity in transaction structures across the sector.

atNorth acquired by CPP Investments and Equinix

atNorth, a Nordic data centre provider, is being acquired by CPP Investments and Equinix in a US$4 billion deal, with the company set to continue operating under its own brand while expanding its data centre operations across the region.

ATNORTH, a Nordic high-density colocation and bespoke data centre provider, is the subject of a joint acquisition by Canada Pension Plan Investment Board (CPP Investments) and Equinix. The transaction is based on a US$4 billion enterprise value and is subject to customary closing conditions, including regulatory approvals.

Under the agreement, CPP Investments will hold approximately 60% of atNorth, while Equinix will hold around 40%. A provisional US$4.2 billion financing package, provided by European and Canadian financial institutions, underpins the deal.

The acquisition highlights atNorth's role in the Nordic region, which is increasingly recognised for its digital infrastructure and emerging AI capabilities. The investment aims to support the expansion of data centre capacity in the region, which is noted for its focus on sustainable practices, renewable energy, and technical expertise.

atNorth operates eight data centres across Denmark, Finland, Iceland, Norway, and Sweden, with additional sites under development. The company plans to secure 1 GW of power capacity to meet growing demand for AI and high-performance computing. Several facilities are equipped with liquid cooling technology to accommodate high-density computing workloads.

Sustainability is integrated into atNorth's operations, with renewable energy sourcing, efficient modular design, and heat reuse initiatives forming part of the company's approach. The company will continue to operate under the atNorth brand while benefiting from partnerships with global enterprises.

For CPP Investments and Equinix, the acquisition represents an opportunity to expand in the data centre sector. The Nordics' stable economy and access to renewable energy make it a suitable environment for data infrastructure growth.

The region's focus on sustainable projects and naturally cool climate further supports its development as a hub for digital infrastructure and innovation in the coming years.

Nscale secures $1.4bn loan for AI infrastructure expansion in Europe

NSCALE has secured a $1.4 billion Delayed Draw Term Loan (DDTL) financed by a syndicate of asset managers and banks, including PIMCO, Blue Owl, and LuminArx Capital Management. The financing is intended to support Nscale's AI infrastructure deployments across Norway, Portugal, Iceland, and the UK.

The GPU-backed DDTL provides capital flexibility and allows Nscale to use debt financing for AI infrastructure projects. The loan structure supports the company's approach to meeting demand for scalable AI infrastructure across Europe.

The loan is designed to fund capital expenditure primarily related to the acquisition and installation of multiple GPU clusters. These clusters are intended for customers under contract and for pipeline cluster developments, contributing to the company's AI infrastructure expansion.

The participation of financial institutions reflects investor interest in AI infrastructure as a growing sector. With backing from asset managers and banking partners, the funding supports deployment of AI-related infrastructure and associated compute capacity.

Current market trends show increasing demand for AI infrastructure as organisations rely on artificial intelligence for operational and data processing requirements across sectors such as healthcare and financial services.

Although the funding is allocated to specific geographic regions, it may have broader economic implications, including infrastructure development and related activity within those markets.

Through the $1.4 billion Delayed Draw Term Loan, Nscale has secured financing to support its AI infrastructure expansion plans in Europe. The funding arrangement enables continued deployment of GPU clusters and associated infrastructure across its target markets.

> The GPU-backed DDTL provides capital flexibility and allows Nscale to use debt financing for AI infrastructure projects

Data centre microgrid deployed in Europe

Pure Data Centres Group collaborates with AVK to deploy 110MW data centre microgrid at its Dublin campus.

PURE DATA Centres Group, in partnership with AVK, has launched a large-scale data centre microgrid at its Dublin campus. The microgrid features a 110MW on-site deployment, designed to address grid limitations in Ireland and support potential future regional expansions.

The microgrid provides dispatchable power to support data centre operations without immediate reliance on the national grid. It comprises three interconnected energy centres, each capable of generating up to 30MW, alongside an integrated Battery Energy Storage System (BESS) to enhance operational efficiency and energy reliability.

With data centre development in Europe often constrained by grid capacity, this deployment offers a model that could be applied in other markets, including Germany, the Netherlands, and the UK. By enabling phased delivery, the microgrid can help reduce pressure on national grids while supporting technological deployments such as cloud computing and AI.

The microgrid is designed to support long-term grid connections and the integration of renewable energy. It provides an alternative approach for data centres facing energy limitations, allowing scalable operations and improved energy management.

Through this project, Pure DC and AVK aim to provide a framework for data centre infrastructure that balances on-site energy provision with broader operational and sustainability considerations.

Vertiv and Zagreb University launch data centre design programme

VERTIV, a provider of critical digital infrastructure, has partnered with the Zagreb University of Applied Sciences (TVZ) to launch a Data Center Design programme within its Electrical Engineering Department. The course is designed to provide participants with skills to support Europe's digital infrastructure and respond to the growing demand for trained professionals in the sector.

The programme provides an in-depth overview of data centre design and optimisation, with a focus on power systems. Participants engage in practical exercises, work with industry-standard software tools, and solve real-world engineering problems to gain the knowledge needed to design, implement, and maintain systems that meet recognised standards for energy efficiency and safety.

Led by Associate Professor Dr. Stjepan Tvorić, the programme is aligned with current industry standards and technologies. Through the collaboration with Vertiv, students have access to professionals in electrical engineering and data centre infrastructure who provide insights from real-world projects. The programme is aimed at electrical engineers seeking to advance their skills, as well as professionals from related technical fields.

Dr. Tvorić said, "Collaborating with Vertiv means we are engaged in discussions about current engineering challenges, using real project examples and case studies, and implementing tools and methods required by today's industry."

Dr. Tvorić said, "Collaborating with Vertiv means we are engaged in discussions about current engineering challenges, using real project examples and case studies, and implementing tools and methods required by today's industry."

Participants gain hands-on experience with CAD tools, designing power systems and practising installation and maintenance techniques. The programme is fully funded through the Croatian Employment Service, making it accessible to eligible candidates pursuing careers in high-demand roles.

Vedran Brzić, Vice President for Infrastructure Solutions at Vertiv, noted that the programme provides practical, career-focused experience. "Vertiv and the Zagreb University of Applied Sciences have built a strong long-term partnership through various joint projects," he said. The initiative addresses the need for trained data centre professionals across industries, helping students acquire skills aligned with market demand.

Regulatory Radar Platform launched for data centre regulatory insight

Regulatory Radar is a new platform designed to provide structured regulatory intelligence and real-time insight for organisations operating in the data centre sector.

THE DATA CENTRE ALLIANCE and Clear Decisions have collaborated to launch Regulatory Radar, a regulatory intelligence platform designed specifically for the data centre sector.

The platform has been developed in response to rapid changes in areas such as energy reform, planning policy and sustainability regulation.

Regulatory Radar provides a unified source of real-time regulatory updates across the UK and EU. It uses an AI-curated approach with a human-in-the-loop model to combine automated analysis with expert oversight.

This is intended to ensure that regulatory information is current, contextually interpreted and relevant for operational and strategic use, rather than simply aggregated data.

By bringing together fragmented policy developments into one continuously updated platform, the system enables organisations to link regulatory changes directly to commercial and operational impacts. The aim is to support a shift from monitoring policy developments to adopting a more proactive approach to regulatory management, providing leadership teams with greater visibility when making decisions.

Regulatory Radar provides a unified source of real-time regulatory updates across the UK and EU. It uses an AI-curated approach with a human-in-the-loop model to combine automated analysis with expert oversight

Key features include:

- Forward-looking analysis of planning, energy and sustainability reforms
- Integrated tracking of AI infrastructure and digital policy strategy
- Identification of emerging compliance and reporting requirements
- Interpretation of regulatory direction of travel
- Dynamically refreshed insights validated through expert review

Steve Hone and Max Weidl have highlighted the importance of adapting to evolving regulatory conditions, while Louisa Cilenti has noted the transition from reactive to proactive regulatory awareness. She stated that by turning complexity into clearer insight, data centre operators can better navigate fast-moving regulatory areas across the UK and Europe.

Regulatory Radar represents a structured approach to regulatory intelligence for the industry, focusing on interpretation and strategic application of policy change.

Global Public Cloud spending to surpass $1 trillion in 2026

Global spending on public cloud services is forecast to surpass $1 trillion in 2026—growing over 21% - and is expected to double by 2029, according to the latest update of International Data Corporation's (IDC) Worldwide Software and Public Cloud Services Spending Guide.

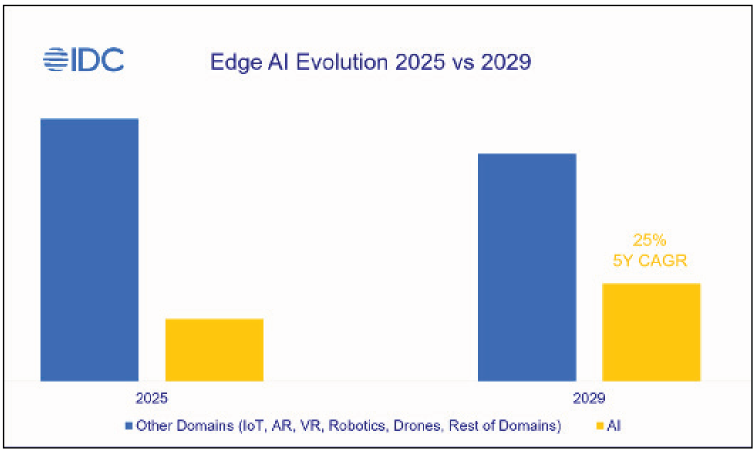
GROWTH IS driven by enterprise-wide application modernization, accelerating adoption of AI-enabled platforms, and rising demand for secure, scalable digital infrastructure across industries.

What is happening in the global public cloud market in 2026?

Growth of public cloud services is led by continued expansion of Software-as-a-Service (SaaS) and a sharp acceleration in Platform-as-a-Service (PaaS) adoption as organizations scale AI development, data platforms, and cloud-native application environments.

Global public cloud market at a glance

- Total public cloud services spending: >$1 trillion (+21% YoY)
- Largest deployment category: SaaS, more than half of total spend
- Fastest-growing deployment category: PaaS, +37% YoY
- Top spending industries: Banking, Software and Information Services, Retail (25% combined)
- Fastest-growing industries: Capital Markets, Banking, Software and Information Services
- Leading regions: United States ($647B), Western Europe ($255B), APeJC ($84B)
- Five-year CAGR above 20%: Middle East and Africa, Latin America, APeJC, Central and Eastern Europe, Western Europe



Why is public cloud spending accelerating?

SaaS Anchors Market Scale

SaaS will remain the largest public cloud deployment category in 2026, accounting for more than half of global spending. Looking at the largest secondary markets, investments are concentrated in IaaS, Enterprise Resource Management (ERM), and security software, reflecting enterprise priorities around cloud migration and cybersecurity, and core modernization.

PaaS Fuels Growth Momentum

PaaS will be the fastest-growing deployment model, expanding over 37% year-over-year in 2026. Growth is driven by rising demand for artificial intelligence (AI) platforms and application development software, as organizations adopt cloud-native environments to support generative and agentic AI, real-time analytics, and data-intensive workloads.

Industry Trends: Where is growth strongest?

Banking, software and information services, and retail will be the three largest public cloud–spending industries in 2026. The next five largest industries — Professional and Personal Services, Capital Markets, Media and Entertainment, Telecommunications, and Healthcare Providers — will account for more than a quarter of global spending.

Three fastest-growing industry for PaaS

- **Banking:** Institutions are modernizing core systems and deploying AI-driven risk, fraud, and real-time banking services.
- **Retail:** PaaS supports rapid development of applications for dynamic pricing, inventory optimization, and digital commerce.
- **Aerospace and Defense:** Increased defense budgets and geopolitical tensions are driving

investment in secure cloud platforms supporting analytics, AI-enabled intelligence, and mission-critical systems.

Regional Outlook

- **United States:** Spending will reach $647 billion in 2026, supported by large-scale enterprise migrations, hyperscaler AI infrastructure investment, and strong demand from regulated industries.
- **Western Europe:** The market will total $255 billion, driven by cloud modernization programs, sovereign cloud adoption, and regulatory-led investment in data protection and AI governance.
- **APeJC:** Spending will reach $84 billion, fuelled by expanding digital economies, midmarket cloud adoption, and government-led digital transformation initiatives.

Analyst perspective

"Public cloud spending will continue to grow as cloud migration remains central for agility, resilience, and efficiency," said Andrea Minonne, research manager at IDC. "In aerospace and defense, higher budgets and rising geopolitical tensions are driving demand for secure, cloud-based platforms supporting advanced analytics and mission-critical systems, with increased defense spending across NATO members and escalating tensions in the Middle East accelerating investment in AI-enabled and security-focused cloud environments."

What is IDC's outlook?

Public cloud services spending will maintain strong double-digit growth through 2029, as AI-enabled platforms, cloud-native application development, and industry-specific cloud solutions become embedded in enterprise digital strategies. Growth could accelerate with broader AI monetization, while risks include regulatory fragmentation, skills shortages, and cloud cost optimization pressures.

FAQs

Why is public cloud spending growing faster than overall IT spending?

Public cloud enables rapid modernization, scalable AI deployment, and operational resilience. Enterprises increasingly prioritize cloud as core infrastructure rather than discretionary IT, supporting automation, analytics, and digital service delivery.

Which deployment models are shaping the market?

SaaS remains the largest category due to widespread adoption of ERM, security, and CRM applications, while PaaS is driving growth as organizations build and scale AI-driven and cloud-native applications.

Which industries are leading adoption?

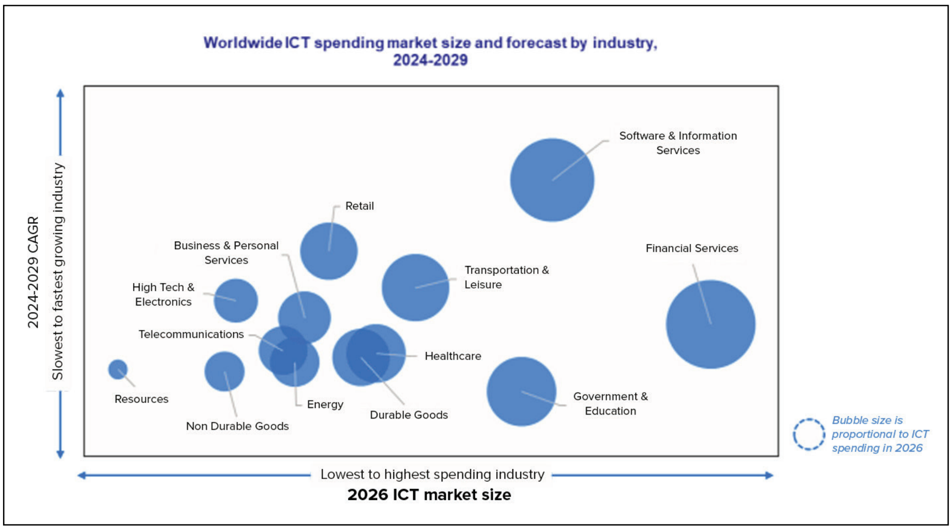
Banking, software and information services, and retail lead in total spending, while Capital Markets, Banking, and Software and Information Services, Insurance, and Aerospace and Defence are growing the fastest.

Edge computing global spending to grow at 15%, reaching $450 billion by 2029

Worldwide spending on edge computing reached $265 billion in 2025, expected to nearly double by 2029, reflecting a pivotal expansion fueled by rapid Edge AI advancements that accelerate enterprise transformation and open new opportunities for service providers, according to the IDC Worldwide Edge Spending Guide.

"The combination of maturing edge architectures and rapid AI development is fundamentally redefining how organizations process and act on data," said Alexandra Rotaru, Data & Analytics Manager and WW Edge Spending Guide Product Lead at IDC. "We see enterprises and service providers shifting toward intelligent, distributed systems capable of realtime decisioning and automation at scale. Edge AI is no longer experimental, its impact is already visible across industrial automation, smart retail, connected vehicles, and nextgeneration healthcare."

Worldwide spending on edge computing reached $265 billion in 2025 and is expected to nearly double by 2029, according to IDC.



Source: IDC's Worldwide ICT Spending Guide: Enterprise and SMB by Industry, February (2026 V1)

Edge as the foundational layer enabling multiple technology domains

IDC segments edge spending across more than 1,000 named enterprise use cases spanning six domains—AI, IoT, AR, VR, drones, and robotics—highlighting the expanding role of edge infrastructure in enabling advanced intelligent workloads. Artificial Intelligence is one of the fastestgrowing domains in the forecast, reflecting the increasing need to process data and run complex models directly at the edge. As organizations deploy more Aldriven applications that require lowlatency inference, realtime context, and resilient distributed architectures, edge computing becomes a critical enabler of innovation and value creation across industries.

Which edge-related technology dominates the market?

Hardware remains the dominant investment category early in the forecast period, propelled by the rapid adoption of Alaccelerated infrastructure and increasingly sophisticated edge systems. This momentum reflects the growing need for realtime computing at the point of data creation, as organizations deploy heavier edge architectures capable of supporting advanced AI workloads.

However, the aggregate Services segments (including Provisioned and Professional Services) are expected to surpass hardware's share by 2029. Within Provisioned Services, Infrastructure as a Service (IaaS) remains the fastestgrowing category as organizations increasingly rely on scalable, flexible, and costefficient consumption models to meet rising Aldriven compute requirements at the edge. As more workloads shift toward distributed and Alenabled environments, asaservice offerings provide a critical foundation for managing capacity, optimizing costs, and accelerating deployment across diverse industry use cases.

Impact across industries

The Retail & Services sector remains the largest contributor to global edge spending, driven by rapid adoption of AI and IoT-related use cases such as video analytics and realtime operational optimization. Manufacturing & Resources follows, with Alenabled quality control, predictive maintenance, and autonomous material handling increasingly dependent on realtime edge processing. Financial services is the fastestgrowing sector, supported by Aldriven fraud analysis requiring lowlatency, distributed architectures.

Service Providers also continue to scale investments in Already edge platforms—including MEC, CDN and virtualized network functions—which are on track to represent nearly onethird of the total market by 2029.

Regional development driven by the USA and Western Europe

North America will remain the top region for edge spending, driven by rapid adoption of EdgeAI workloads and advanced infrastructure deployments. Western Europe and China follow, supported by strong industrial and publicsector investment in Alenabled applications. The United States and Latin America are expected to record the fastest growth, as organizations across both regions scale lowlatency, Aldriven use cases that benefit from distributed edge architectures.

Global ICT spend to reach $4 trillion in 2026

Global spending on information and communications technology (ICT) is forecast to reach $4 trillion in 2026 and surpass $6 trillion by 2029, according to the latest update of the International Data Corporation's (IDC) Worldwide ICT Spending Guide Enterprise and SMB by Industry. Excluding the consumer segment, the ICT market is projected to grow 10% in 2026, driven by rapid adoption of Artificial Intelligence (AI) platforms, which will grow over 70% by the end of the year.

What are the main technology drivers for ICT spending?

Software emerges as the largest technology group in 2026, absorbing more than 33% of global ICT spending. This surge is primarily fueled by robust investments in enterprise resource management (ERM) applications, security software, and production and operations applications, which together will account for over one-third of total global software expenditure. These categories reflect enterprises' ongoing priorities around operational efficiency and cybersecurity resilience. Hardware is set to be the fastest growing technology group, with a projected year-over-year growth rate of 15% in 2026. This expansion is driven by significant momentum in non-x86 servers, the proliferation of wearables, and increased adoption of Infrastructure-as-a-Service (IaaS). This underscores the rising demand for specialized compute infrastructure, edge devices, and scalable cloud resources to support next-generation workloads and hybrid environments.

Global ICT Market at a Glance

- Total ICT spending: $4 trillion (+10% YoY)
- Software: Largest tech group, >33% of spend
- Hardware: Fastest growth, +15% YoY
- Top industries: Software and Information Services, Banking, Retail (>$1 trillion combined)
- Fastest-growing sectors: Software and Information Services, Media and Entertainment, Retail
- Leading regions: USA ($2 trillion), Western Europe ($908 billion), China ($355 billion)
- Five-year CAGR >10%: USA, China, Latin America

“We are entering a new phase of the AI-everywhere journey: the era of expectations and reckoning,” said Andrea Siviero, senior director at IDC. “The excitement of experimentation is giving way to a sharper focus on accountability, value creation, and productivity impact. In 2026, enterprises will demand that AI and digital investments demonstrably improve processes, accelerate decision making, and ultimately drive business growth across the organization.”

Key global dynamics

Major global events in 2025, including escalating trade tariffs, security threats, and the US government shutdowns, prompted organizations to accelerate investments in AI-driven optimization and security. Geopolitical tensions and supply chain disruptions accelerated digital transformation and security spending. The United States will lead global ICT spending in 2026, reaching $2 trillion, driven by its large enterprise base and rapid adoption of cloud and AI technologies. Western Europe will be the second largest market at $908 billion, supported by regulatory-driven modernization and accelerated AI adoption in sectors like banking and manufacturing. China will follow with $355 billion, fueled by government-led digital infrastructure and expansion of smart manufacturing and AI platforms.

What are the key industry trends in ICT spending?

Software and information services, banking, and retail will be the three largest spending industries and together will represent over $1 billion in ICT spending in 2026. The next five largest industries—federal/central government, telecommunications, media and entertainment, healthcare provider, and high tech and electronics—will together account for more than 20% of worldwide spending.

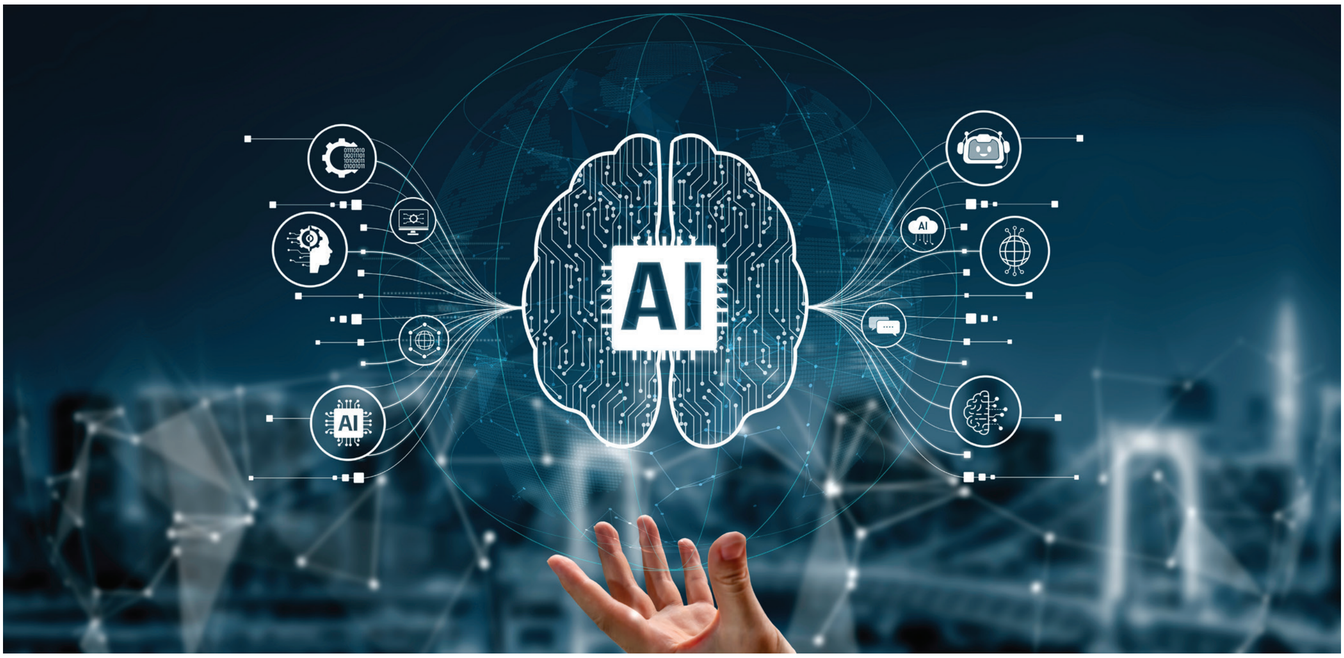
“As companies continue to invest in automation, industries such as aerospace and defense, insurance, and software and information services are poised to accelerate spending in AI platforms the fastest,” said Andrea Minonne, research manager at IDC. “In aerospace and defense, escalating geopolitical cross-regional tensions and heightened security concerns are prompting governments to expand defense budgets”

“As companies continue to invest in automation, industries such as aerospace and defense, insurance, and software and information services are poised to accelerate spending in AI platforms the fastest,” said Andrea Minonne, research manager at IDC. “In aerospace and defense, escalating geopolitical cross-regional tensions and heightened security concerns are prompting governments to expand defense budgets”.

Looking at other fast-growing vertical markets, insurance firms are leveraging AI platforms to manage rising claims volumes, improve risk modelling, and deliver real-time health insights, responding to both regulatory demands and evolving customer expectations. Software and information services are at the forefront of deploying generative and agentic AI to automate business processes, enable intelligent workflow orchestration, and support scalable digital transformation initiatives.

Meanwhile, high tech and electronics, automotive, and consumer goods are navigating ongoing frictions including trade tariffs and talent shortages and are prioritizing nearshoring strategies, factory automation, and advanced analytics. These investments are enabling organizations to mitigate supply chain risks, optimize production, and maintain competitiveness in a volatile global market.

Software and information services, banking, and retail will be the three largest spending industries and together will represent over $1 billion in ICT spending in 2026. The next five largest industries will account for more than 20% of worldwide spending.



Scaling liquid cooling for the AI Era

One year on from Schneider Electric's strategic investment in Motivair, the combined business is already demonstrating the impact of aligning specialist cooling expertise with global infrastructure scale. As the data centre industry accelerates into an AI-driven future, the partnership is positioning itself at the centre of one of the sector's most critical transitions: the shift toward high-density compute and advanced cooling architectures.

AN INTERVIEW WITH ANDREW WHITMORE, CHIEF COMMERCIAL OFFICER AT MOTIVAIR BY SCHNEIDER ELECTRIC

MOTIVAIR'S INTEGRATION into the Schneider Electric ecosystem has been characterised by rapid expansion, operational scaling, and a clear focus on enabling next-generation workloads. The past year has not only validated the strategic rationale behind the deal, but also highlighted how quickly the data centre landscape is evolving under the pressure of AI.

A year of globalisation and growth

Since joining Schneider Electric, Motivair has significantly extended its manufacturing and delivery capabilities. Production has been industrialised across multiple regions, with facilities now operating in Italy and Bangalore alongside its established footprint. This geographic diversification reflects the increasingly global nature of data centre demand, particularly as AI deployments expand beyond traditional hyperscale hubs.

The company has also secured project wins across key regions including the Far East, the Middle East, Europe, and North America. This broad market traction underscores both the universality of the cooling challenge and the growing recognition of liquid cooling as a viable and necessary solution in high-performance environments.

Crucially, the integration has been described as collaborative and non-disruptive an outcome often cited as elusive in large-scale acquisitions. Cultural alignment between the two organisations has played a central role here, with both businesses sharing a focus on engineering excellence and customer-centric innovation. The result is a partnership that combines Motivair's deep domain expertise with Schneider Electric's global reach, supply chain strength, and end-to-end infrastructure capabilities.

AI drives the cooling conversation

If the past year has had a defining theme, it is undoubtedly the rise of artificial intelligence. AI workloads particularly those driven by GPUs are fundamentally reshaping data centre design, pushing rack densities to unprecedented levels and forcing operators to rethink traditional cooling strategies.

While liquid cooling has been discussed in industry circles for over a decade, the AI boom has propelled it into the mainstream. Today, it is no longer a niche consideration but a central topic at industry events and among operators planning new builds or retrofits.

For Motivair, however, this is not new territory. The company has been supplying liquid cooling solutions for nearly 15 years, placing it in a strong position to guide customers through what can often be a complex transition. Rather than simply promoting technology, the focus has been on education, standardisation, and responsible deployment.

Industry collaboration is a key part of this effort. Motivair and Schneider Electric are actively engaged with organisations such as ASHRAE and its Technical Committee 9.9, which focuses on mission critical facilities. They are also working closely with leading chipmakers including NVIDIA and AMD to help define best practices, testing methodologies, and performance benchmarks.

This ecosystem approach is essential. As liquid cooling adoption grows, consistency and interoperability will be critical to ensuring reliability and scalability across the industry.

Not the end of air cooling

Despite the growing prominence of liquid-based systems, it would be a mistake to view this as a wholesale replacement of air cooling. In reality, the future is far more nuanced.

Air cooling remains highly effective for many workloads and will continue to play a significant role, particularly in facilities operating at lower rack densities. Even as some environments move from 5 kW racks to 20–30 kW, there is often no immediate requirement to transition fully to liquid cooling.

Instead, hybrid approaches are emerging as a practical pathway forward. Technologies such as liquid-to-air systems allow operators to introduce liquid cooling incrementally within existing air-cooled environments. These solutions combine elements of coolant distribution units (CDUs) with rear-door heat exchangers, enabling higher performance without wholesale infrastructure changes.

This flexibility is critical in an industry where retrofitting legacy facilities can be both costly and operationally complex. By offering intermediary steps, vendors like Motivair are helping to de-risk the transition and broaden adoption.

Direct-to-chip leads the way

Within the liquid cooling landscape itself, multiple approaches are competing for attention. However, current trends suggest a clear frontrunner.

Direct-to-chip, single-phase liquid cooling is widely regarded as the most viable solution for near-term AI deployments. This approach involves delivering coolant directly to high-heat components such as CPUs and GPUs, providing efficient heat removal at the source.

The direction of travel is largely dictated by the silicon roadmap. As processors from companies like NVIDIA and AMD continue to increase in power density - often exceeding 1 kW per chip - cooling technologies must evolve accordingly. Not all approaches are capable of meeting these demands.

Immersion cooling, for example, has generated significant interest but faces practical limitations in high-density AI environments. Single-phase immersion systems can struggle to handle the thermal loads of modern GPUs, while two-phase systems introduce additional complexities, including environmental considerations around certain fluids and constraints on scalability.

This is not to say immersion has no role to play. It may prove effective in specific use cases, particularly at lower densities or in purpose-built facilities. However, for mainstream AI deployments, direct-to-chip remains the dominant and most immediately applicable solution.

Unlocking the value of waste heat

One of the often-cited advantages of liquid cooling is its ability to capture and reuse waste heat more effectively

than traditional air-based systems. This capability is attracting growing interest, particularly as sustainability becomes a central concern for data centre operators and local communities alike.

In parts of mainland Europe, district heating networks are already leveraging data centre waste heat to warm homes and businesses. Countries such as Denmark have been at the forefront of this approach, integrating data centres into broader energy ecosystems.

The potential applications are diverse. Beyond residential heating, recovered heat can be used for greenhouse agriculture, enabling year-round food production, or for snow-melting systems in colder climates. There is also ongoing exploration into more advanced use cases, such as water desalination and even electricity generation.

Importantly, these initiatives highlight a shift in how data centres are perceived. Rather than being viewed solely as energy-intensive facilities, they are increasingly seen as potential contributors to local infrastructure and sustainability goals.

However, challenges remain. Proximity to heat users, infrastructure costs, and regulatory frameworks all influence the viability of such projects. While the UK may currently lag behind some European counterparts, the underlying technology is capable of transporting heat over longer distances, suggesting that broader adoption is possible with the right investment and planning.

Innovation at scale

Looking ahead, the combined capabilities of Motivair and Schneider Electric are set to drive further innovation across both power and cooling domains.

Schneider Electric's expertise spans the entire energy chain - from grid to chip - while Motivair brings deep specialisation in thermal management. Together, they are developing solutions that address the full lifecycle of data centre operations.

On the power side, emerging technologies such as 800 V DC architectures are being explored to support the next generation of high-performance computing. Meanwhile, on the cooling front, Motivair continues to expand its portfolio to meet the demands of increasingly large-scale deployments.

A recent example is the launch of the MCDU70, a 2.5 MW coolant distribution unit designed as a modular building block for hyperscale and campus environments. These units can be scaled to support facilities ranging from tens to hundreds of megawatts, and even gigawatt-scale developments.

This modular approach aligns with broader industry trends, where flexibility and scalability are essential to accommodate rapid growth and evolving technology requirements.

Edge AI: same fundamentals, different scale

As AI continues to mature, attention is also turning to the edge - where inferencing workloads are deployed closer to end users. While this represents a shift in deployment models, the underlying infrastructure requirements remain largely consistent.

Edge environments still require reliable power delivery and effective cooling, even if the scale is smaller. The transition from large-scale generative AI training to distributed inferencing may result in changes to density profiles, but it does not fundamentally alter the physics of heat generation.

Every unit of electrical energy consumed by IT equipment ultimately becomes heat and managing that heat efficiently remains a core challenge regardless of location. In this sense, edge AI is less a departure from existing paradigms and more an extension of them.

A pivotal moment for the industry

The convergence of AI, high-density compute, and sustainability is creating a pivotal moment for the data centre sector. Cooling, once considered a supporting function, is now a strategic priority - one that directly influences performance, efficiency, and environmental impact.

For Motivair and Schneider Electric, the past year has laid a strong foundation. By combining innovation with global scale, and by actively engaging in industry-wide collaboration, they are helping to shape the standards and solutions that will define the next generation of data centres.

As AI continues to drive demand, the importance of getting cooling right cannot be overstated. Whether through direct-to-chip liquid systems, hybrid approaches, or advanced heat reuse strategies, the industry is entering a new era - one where thermal management is central to unlocking the full potential of digital infrastructure.

Speed Meets Certainty: The Fastest Path to AI-Ready Infrastructure

Accelerate AI Deployments with Prefabricated & Pre-integratedPodand Rack Solutions

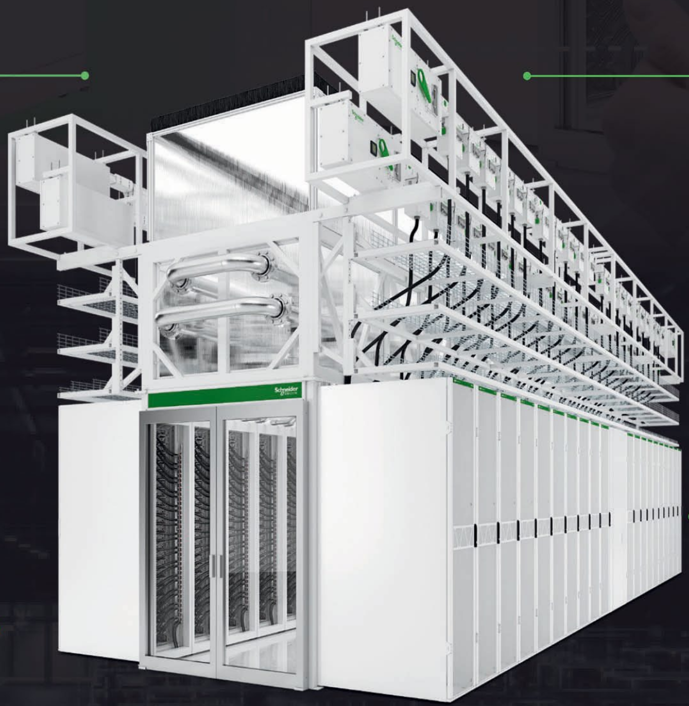
EcoStruxure™ Pod and Rack Solutions from Schneider Electric™ deliver a faster, lower-risk method to deploy AI and accelerated compute infrastructure at scale. How? By shifting complexity off-site.

Our solutions arrive fully engineered, factory-assembled, and compute-ready, compressing deployment timelines from months to days, enabling rapid, predictable, and globally scalable AI capacity without comprimising reliability.

Prefabricated for Speed
Factory-built and tested pods and racks reduce onsite work and accelerate time-to-compute.

Deploy in days...
Not weeks.
Not months.

Global Expertise, Delivered Anywhere
Engineering and service experts to support fast, high-density rollouts across 100+ countries.



AI-Ready Power & Cooling
Engineered for the next generation of liquid-cooled AI architectures, with integrated technical water and high-density power pathways built to evolve as workloads grow.

Scalable Pod Architecture
Flexible by design, allowing pod infrastructures to scale from initial deployments to multi-megawatt buildouts with consistent quality across global sites.

Pre-integrated Racks
EIA, ORV3, and NVIDIA MGX racks ship integration-ready for rapid deployment and reliable performance.

Discover the IT Pod

Schneider Electric™

Powered Land: Rewiring data centre strategy for the AI era

In the AI era, “powered land” is redefining data centre strategy by making secure, scalable energy access the primary driver of site selection, investment and long-term infrastructure planning.

BY CHRISTIAN GOLDSMITH, DATA CENTRE GLOBAL SOLUTIONS LEAD, ARCADIS

AS AI WORKLOADS accelerate, driven by large language models (LLMs), high-performance computing and AI gigafactories, energy has become one of the defining features of digital infrastructure strategy.

In several established markets, including West London, grid connection timelines now exceed build schedules. Transmission networks built for a different era are ageing, capacity is tight, and public scrutiny over power allocation is intensifying. It’s not enough to ‘find land and then secure power’. Instead, land and power need to be integrated from day one.

This is the logic behind Powered Land; sites with credible, bankable routes to reliable power, whether through grid connection, private wire, on-site generation or hybrid solutions. For developers, hyperscalers and investors, the starting point for any investment case must be securing power.

From grid connection to energy strategy

This shift is reshaping data centre geography. Because AI training workloads are less latency-sensitive, compute is moving toward where power is available, namely secondary markets, remote campuses, and energy-adjacent locations.

Yet Powered Land is more than a site selection tactic. It’s a strategic integration challenge across sovereignty, equity, permitting, grid infrastructure, sustainability and speed-to-market. Power generation timelines, transmission upgrades and regulatory approvals often operate on 5 to 10-year horizons. Private equity, by contrast, seeks returns in half that time. Bridging this gap calls for phased power solutions, including temporary generation, hybrid supply, and ramping-up staged capacity, all aligned with build and revenue schedules.

This requires insight across the entire value chain: investors, utilities, transmission operators, hyperscalers, supply chains and local authorities. It is not about owning every component, but about integrating them, and aligning capital, engineering, regulatory pathways and community engagement into a coherent investment proposition.

The Next Frontier: inference at the edge

Today’s remote AI training campuses are laying the foundation for tomorrow’s digital economy. But inference (the real-time application of AI models) increasingly needs to sit closer to population centres. Autonomous mobility, healthcare diagnostics, industrial automation and smart infrastructure cannot tolerate excessive latency. That means a new generation of distributed, lower-latency inference data centres embedded within or adjacent to urban environments. Ironically, these are the very locations where power scarcity and public sensitivity are most acute.

The future of Powered Land therefore equates to portfolios of strategically located sites, supported by behind the meter modular generation, resilient grid connections, energy storage and SMRs, integrated with district heating, hydrogen production and broader decarbonization strategies. It is not a single mega-project; it is a coordinated energy-and-digital ecosystem.

Orchestrating the powered land ecosystem

Delivering Powered Land at scale requires orchestration.

Power providers do not want to become data centre operators, or vice versa. Investors want certainty and accelerated returns, while utilities and regulators operate on longer infrastructure timelines. All these factors need to be aligned.

Arcadis operates at that intersection, acting as an integrator across power, place and digital infrastructure.
By aligning equity, engineering, sustainability strategy and delivery, we help transform fragmented assets into coherent, bankable propositions. In a market defined by complexity, it is this ability to connect the whole system that truly unlocks value.

AI at Scale: The next phase of UK data centre growth

AI-driven data centre growth in the UK is entering a new phase, where scalable retrofits, modular design and regional expansion are essential to deliver resilient, high-density infrastructure beyond London.

BY CLAIRE KEELAN, MANAGING DIRECTOR UK, ONNEC GROUP UK LTD

SCALING AND upgrading legacy data centres to support AI and advanced compute is not a simple refresh cycle. It takes careful planning, modular upgrades, and flexible designs that minimise disruption to live environments.

London can do this at scale because it already has the ecosystem, the installed base and the policy tailwinds. In practice, it remains the starting point for many AI-driven capacity plans, especially where latency, resilience and regulatory expectations are non-negotiable.

Home to 40% of the nation's data centres, London remains the UK's key data centre hub. That dominance reflects investor appetite, political support, and a proven ecosystem serving both enterprise and financial services. It is also being reinforced by government initiatives like the AI Growth Zones and forthcoming planning reforms.

But this reliance is why retrofitting and expanding to new regional hubs has moved up the agenda.

Building AI-ready capacity without downtime

For data centre operators, rapid growth, regulatory compliance, and AI adoption make retrofitting legacy sites a practical necessity. Retrofitting also goes far beyond swapping ageing hardware. It means reworking power paths, cooling, and cabling to cope with the extreme demands of AI workloads, which can exceed 40kW per rack.

This is where modular, flexible architectures matter. They allow facilities to evolve over time, supporting AI, hybrid workloads and future expansion without lengthy downtime. Where retrofits cannot stretch quickly enough, colocation remains an option. It offers a way to add scalable, secure and compliant capacity – without the cost and complexity of building and running every environment in-house.

Even so, London cannot absorb the demand for AI on its own, particularly as power and planning constraints tighten and national compute needs keep rising.

A distributed future takes shape

While London continues to anchor the market, operators are increasingly looking beyond the M25 to meet growing national demand. Regions such as Manchester, South Wales and Scotland are attracting investment, but building new data centres into these areas brings a different set of challenges. Less experienced local planning authorities, grid limitations and logistical constraints can slow delivery unless operators bring the right expertise to the table.

At the same time, regional "AI growth zones" are emerging as engines of national capacity. Lower land costs, access to renewable energy, and stronger links to universities and research centres make these locations increasingly attractive. The result is a more distributed and resilient UK footprint that balances power use, eases local bottlenecks and reduces pressure on London and the M4 corridor.

The shift is clear. The next phase is not just about where capacity sits, but how reliably it can be delivered across a mixed portfolio of retrofits and new regional builds.

Why partners matter

Given the scale and complexity of AI-driven infrastructure, operators cannot go it alone. The right partners bring power, cooling, cabling and IT expertise into a single coordinated approach, avoiding the silos that undermine flexibility and slow change.

Trusted partners provide continuity through rolling upgrades, rapid mobilisation and regional deployments. In many projects, the biggest advantage they bring is repeatability. This means taking what works in London's toughest environments, including tight planning conditions, phased upgrades and high-density deployments, and applying it consistently across new regional footprints.

That continuity reduces design cycles, accelerates approvals, and helps operators maintain consistent operational standards across multiple sites, so infrastructure can keep pace with AI demands today, and whatever comes next.

Nuclear Energy: The backbone for AI infrastructure

The projected exponential increase in using artificial intelligence by 2030 is not widely disputed.

BY DOREEN ABEYSUNDRA, FOUNDER, FRESCO CLEANTECH SOLUTIONS

2026 is that paradigm shift in AI's utility.

THERE IS NO AI without energy, specifically electricity for data centers. From today's 1.5% of global electricity consumption to IEA's April 2025 projection of between 2% to 4.4% by 2030, data centers need energy infrastructure to increase by no less than 0.5% to at least double in the next four years!

Cooling is also a concern for today's data center. Increasingly efficient solutions are being designed to improve a campus' power usage effectiveness (PUE). From air, liquid or immersion cooling systems offering least to highest PUE, all need electricity to reduce the expansive heat generated by increasingly powerful GPUs needed to run AI models.

The 2015 Paris Agreement aimed to cap global warming by 1.5 degrees Celsius compared to before the Industrial Revolution and the Carbon Border Adjustment Mechanism upheld in the EU since 2023 - an acknowledgement that clean water and air are fundamental for Planet Earth.

What's the solution to powering AI with limited environmental pollution?

The proposal of nuclear energy for non-intermittent electricity supply, complemented by renewable energy – that's not just solar and wind but hydropower, geothermal, tidal – matched with non-lithium battery storage and AI-enhanced grid balancing optimization.

Following the 1954 inception of grid-connected nuclear power, the industry is shifting from traditional gigawatt-scale plants toward Small Modular Reactors (SMRs). China anticipates commissioning a 125MW SMR mid-year with new facilities early 2026. The USA aims to demonstrate advanced modular reactor (AMR) technologies by July, targeting a quadrupling of nuclear capacity by 2050.

Why do history and projections matter?
They demonstrate nuclear as a credible energy source for uninterrupted electricity supply; until technology allows for data centers in space powered by the sun and cooled by the atmosphere!

Perceived challenges?
Nuclear plants cannot load-follow and are too expensive. Not true - they can turn up or down to adjust to cyclical demand spikes and typically offer the lowest cost of baseload grid capacity.

Glaring challenge?
Data centers need electricity but most new nuclear is not operational. That's why grid optimization alongside national infrastructure planning for low-carbon energy generation and usage is fundamental, until units become commercially available. Now is the time to eliminate dark data: unused and forgotten material stored virtually, estimated to comprise up to 80% of many enterprises' storage!

Operational challenges?
Lack of grid infrastructure and connection, particularly in Europe. One workaround is behind-the-meter usage, where SMR or AMR plants connect directly whilst nations improve on connectivity.

Financial challenges?
Insufficient comprehension of the Risk-Return ratio is a stumbling block in matching today's ample uninvested liquidity with nuclear power investment. Fresco CleanTech Solutions expounds the pairing of sound due diligence – legal, insurance, technical and cyber security – with effective financial structuring to achieve bankable solutions, alongside apolitical and non-siloed national infrastructure planning for intensive energy procurement and deployment.

Kudos to The DCA for incorporating the nuclear discussion into its forums and knowledge hub.

> There is no AI without energy, specifically electricity for data centers.

AI Workloads break power systems – Here's the solution

AI's volatile, high-density workloads are destabilising traditional power systems, positioning mechanical inertial stabilisation as a credible solution to maintain grid and infrastructure resilience.

BY KEITH MACLEAN-MARTIN TECHNICAL DIRECTOR, PILLER UK LIMITED

AI IS THE BUZZ across the whole data centre industry, driving speculation, development and huge investment – the numbers sometimes are quite unbelievable. This includes the power profiles from the GPU servers which are described as dynamic, volatile and cyclic – nothing like conventional data centre loads. This is breaking site power systems new and old, and stressing the DNO Utility level. Many have suggested batteries are the solution... others are backing super or ultra capacitors... but no viable solution has been identified – until now.

Langley Power Solutions Division showcased at the DCA 10x10 event, held at PowerEx Live London, Piller Power System's ShieldX ™ technology that protects the power infrastructure from these demanding loads. The Dynamic Power Stabiliser has been specifically developed to combat AI's volatile loads, evolved from UPS technology that has been protecting systems around the world for decades.

When GPU servers are called into operation, load steps of 100% are typical. If this was one server the infrastructure could handle it, but when 1000's of servers simultaneously step from 50% load to 150% and back again, all in just 100ms, this breaks power systems.

The technology uses mechanical inertia to instantly respond to these dynamic changes, fixing the system supply quality and shielding power source from load instability – no electronic switching nor chemical reaction. This in turn shields other loads from the propagation of supply instability. The response derived from natural inertia in a synchronous machine uses no complex controls, avoiding the risk of conflicting tuned electronic control circuits.

The inertial system response is supported by not just the synchronous machine, but includes flywheels connected using robust bi-directional thyristor converters to enable full power flow in both directions. This extends the protection and maintains support through the cyclic profile, not just for the first few spikes. This cycling, of every 1-2 seconds, has had a dramatic impact on previously deployed solutions, cutting life to only 1 or 2 years before complete replacement is required. The robust rotating machinery is not impacted by this cycling, retaining more than 20 years of operational life.

Large utility supplies with a lot of inertia can support these demands without any instability but we are losing inertia all the time and the grid isn't always available. Low inertia grids such as a microgrids, renewable sources, or back up generation, the frequency changes rapidly – leading to loss of supply. The rotating technology is strong enough to be grid forming maintaining the operation all assets even if islanded.

The world has been disrupted by AI. Power demand is off the chart, grids have lost their inertia and onsite generation is frequently being used. The demands of AI are breaking the power systems, so far with no answer - not even AI can answer this... I've tried. Relying on the grid, using batteries or capacitors is not adequate. It seems the only viable solution that is not adversely affected by the load profile is a mechanical inertial system - Piller's ShieldX ™ delivers stability and protection to power systems using proven technology to deliver dynamic power stabilization.

Is colocation the catalyst for digital transformations?

Colocation has always been the go-to for businesses that need a safe and dependable home for their data. But now, it's not just about servers. It's about creating a vibrant space where companies can thrive, push boundaries, and conquer the constantly evolving digital world.

BY FRANCESCO FONTANA, ENTERPRISE MARKETING & ALLIANCES DIRECTOR AT ARUBA S.P.A

IT'S NOT just about the physical space anymore. It's about building a thriving colocation ecosystem. That means super-fast connections, a variety of cloud providers, and a network of businesses working together to spark new ideas and breakthroughs. As tech keeps evolving, colocation providers who get behind AI, sustainability, and community building will be the ones paving the way forward.

An increasing risk of data deluge

Data centres are the engines that drive our digital world, handling the ever-growing need for computing power. But with society producing more and more data – from social media, streaming services, and online shopping to 5G networks, cloud computing, and even the Internet of Things (IoT) – the pressure is on for data centres to keep up and keep delivering. That being said, companies are ready to embrace data and become even more data-driven, with plans to grow and revamp their IT systems. But here's the catch: many of these companies have grand plans for digital transformation, but they're forgetting one crucial ingredient – their data centre.

Something needs to change to ensure their capacity and connections can keep up in the long run and to keep everything secure and running smoothly.

Could colocation be the solution?

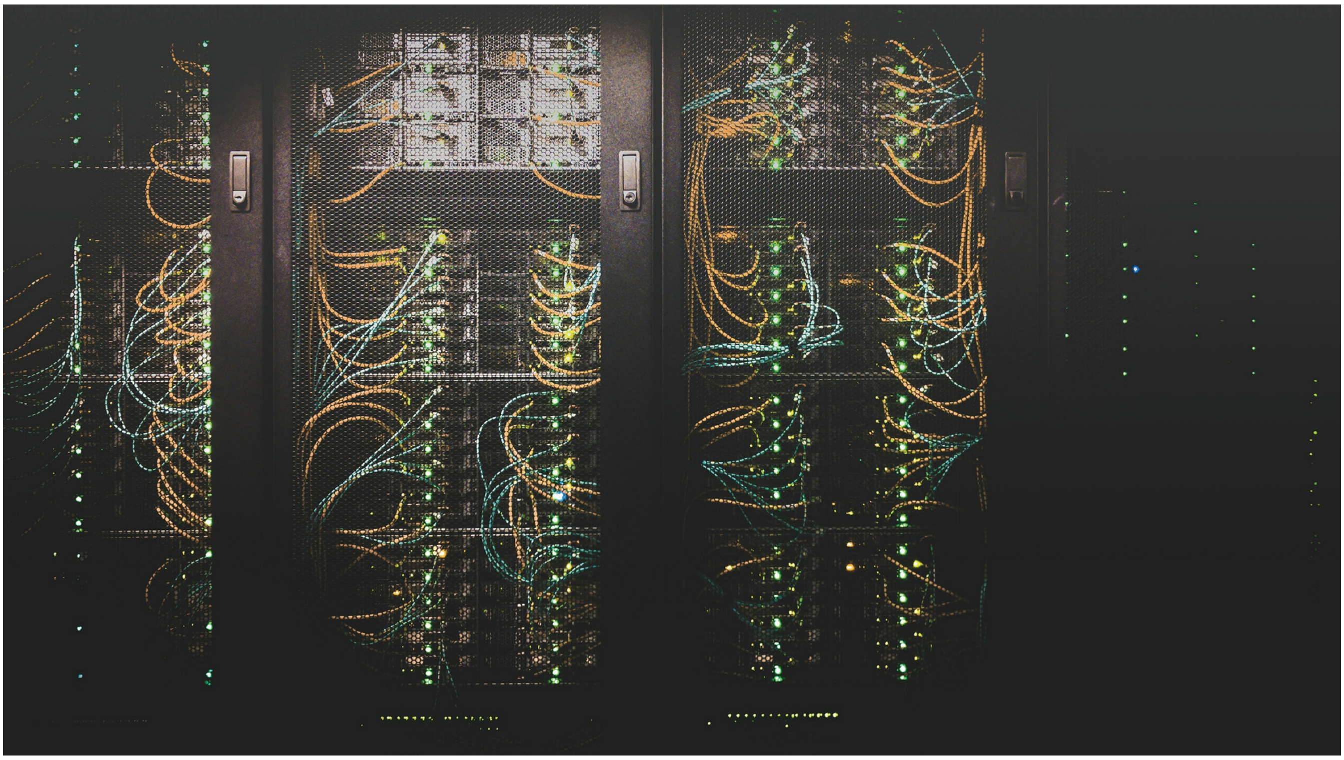
If companies want to grow, they need rock-solid, dependable IT systems that can handle their everyday workloads and then some. If their systems can't keep up, they risk major headaches like downtime, which can mean lost productivity, lost money, and even unhappy customers.

To tackle this issue, some companies have looked to the cloud as their storage saviour, hoping to power new applications. While the cloud's impressive computing and operational space can certainly fit the bill for big enterprises, there are a few factors to consider for the rest, and affordability is a primary consideration.

Cloud storage costs can quickly accumulate and become difficult to manage due to unpredictable pricing models and data egress fees. As a result, many companies have brought some of their workloads back home to on-premise data centres, using a hybrid approach for their IT setup. This way, they get the best of both worlds: on-site data storage and processing power combined with the benefits of the cloud.

But keep in mind that this setup is complex and needs significant networking muscle to keep the data flowing smoothly. And that's where data centre colocation steps in as a real game-changer for big companies.

Think of colocation as renting an apartment for your data. Companies pay for space in a data centre, which gives them the power, cooling, and a safe place to keep their IT gear instead of having to host it all on-site. But colocation isn't just about 'rack space' anymore. It's become a major player in making digital transformation happen, giving businesses the flexibility, scalability, connections, and security they need to push their digital projects forward. By using these professional



data facilities, companies get a secure and reliable infrastructure without the hassle of building and maintaining their own space.

How colocation enables digital transformations

Digital transformation requires companies to be quick on their feet and ready to grow. Colocation offers the power to scale IT resources up or down whenever needed. That doesn't just mean adjusting their infrastructure as things change; it's also crucial for supporting new applications, diving deeper into the cloud, and handling all that growing data.

In addition, colocation data centres create ecosystems of connectivity. Companies get direct links to multiple internet providers, cloud services, and other businesses. This makes digital transformation much smoother because it's easier to build powerful IT systems, tap into cloud services without a hitch, and run applications lightning-fast, no matter where you are. And by taking away the hassle of running data centres, colocation simplifies IT management, freeing up valuable time and resources. That means business leaders can focus on other digital projects that actually help the company grow.

While location isn't usually a problem for connectivity, it can be a real game-changer. The closer your data is to you, the faster you can get to it. It is for this reason we see colocation data centres operating near major cloud regions. Countries like Italy and Spain, for example, benefit from strong connectivity to North African and Middle Eastern markets, as well as North and West Europe. We've seen real results from this: lower latency (less lag time) and better connections between companies. That translates to a smoother user experience and even lower broadband costs for those power-hungry applications.

And there's more. Colocation also supports growing worldwide ESG requirements. Going green is a significant priority already, but with all the pressure on businesses to balance digital transformation with being carbon neutral, colocation becomes crucial. Companies need a future-proof plan. Colocation providers understand that, and they've had sustainability as their top priority for years.

The future is already here

In our world, where data is at the heart of everything, IT systems need to be lightning-fast, adaptable, and dependable if businesses want to go digital without a hitch. Colocation is the perfect answer, giving companies the power of all those cutting-edge technologies, such as the cloud, without the exorbitant price tag.

As more companies jump on the digital transformation train, the colocation market will keep moving forward, growing steadily. As it grows, expect to see even more services emerge, such as the highly sought-after AI-powered resource management and eco-friendly energy solutions.

In our world, where data is at the heart of everything, IT systems need to be lightning-fast, adaptable, and dependable if businesses want to go digital without a hitch. Colocation is the perfect answer, giving companies the power of all those cutting-edge technologies, such as the cloud, without the exorbitant price tag

2026's Perfect Storm: AI growth vs. power availability

TES Power's Managing Director, Michael Beagan, explores the **top five power challenges** the data centre industry may face in 2026, in light of the surge in AI. With a huge increase in demand, power is becoming the most precious commodity across the data centre industry, and this piece delves into the details of why.

WORLDWIDE DEMAND for digital infrastructure continues to surge, predominantly driven by the AI boom. However, data centre builders looking to capitalise on this generational hunger for greater compute capacity will enter 2026 facing myriad challenges, new and old. AI data centres are consuming extraordinary amounts of power, and AI workloads are starting to reshape the digital infrastructure that hosts them. Increasingly, these trends are coming into conflict with constrained power grids, supply chain woes, and stricter ESG requirements. The upshot is that data centres face unprecedented pressure on their power infrastructure.

In 2026, the industry will have to navigate this convergence of grid limitations, supply chain constraints, and evolving regulatory environments. Highlighting this rapidly evolving and uncertain landscape, we're going to take a look at the five biggest power-related challenges shaping the data centre landscape, and the emerging strategies organisations are using to stay ahead.

T1 Market power constraints

Goldman Sachs Research forecasts global power demand from data centers will increase 50% by 2027 and by as much as 165% by the end of the decade, reflecting the massive worldwide surge in compute capacity required to support the booming AI sector. The problem is that the AI, data centre, and energy industry all move at different speeds.

Data centres can typically be built in one to two years. That's still slower than AI companies would like, but the real challenge is that the electricity infrastructure needed to support new data centres takes much longer to get connected. The timelines required to expand transmission networks, substations, and generation capacity stretch far further into the future. In the EU, wait times for securing a grid connection can range from two to ten years. More and more, this mismatch is becoming a critical bottleneck.

Meanwhile, hyperscale and AI-focused developers continue to cluster near major cities where demand is highest, placing additional pressure on already stressed local grids.

Racks supporting AI training workloads can exceed 40 kW today, trending toward 85 kW, with 200–250 kW per rack projected by 2030, and long-term visions discussed by Google exceeding 1 MW racks. These power densities are running up against hard physical and regulatory limits. Several major hubs, including Dublin and Amsterdam, have already paused new grid connections due to capacity constraints before the AI boom. Increasingly saturated markets in the FLAP-D region are likely to follow suit.

Put simply, Tier 1 markets can no longer support growth at the scale or speed required by AI infrastructure. Developers can't continue building in the same locations in the same ways as before.

In 2026 and beyond, data centre companies need to look seriously at new geographic hubs with direct access to available power. Regions such as the Northern Ireland coast could offer proximity to generation assets and more favourable grid availability. This macro-level strategy helps decongest Tier 1 markets and unlock faster development cycles.

AI training is redefining data-centre power consumption

It's no secret that AI increases the overall energy demand of data centres. However, less well known is the fact it also reshapes how and where power is consumed. AI training workloads require enormous, sustained electricity use but are (a bit) less latency-reliant than cloud or AI inference workloads, enabling facilities to be placed far from population centres. As a result, there's a new generation of data centres dedicated especially to training large AI models that, while physically larger and more power-intensive, can be located in more remote areas. Outside congested, largely urban, hubs, data centres can capture advantages like lower land costs, access to water, and abundant renewable power.

However, building any data centre (let alone one big enough to support mass AI training and similar applications) anywhere off the beaten track presents logistical issues. Doing so creates complexities around construction, manufacturing logistics, and commissioning, and traditional approaches to building in Tier 1 markets often fall short.

Data centres must find ways to deliver massive compute capacity in remote locations while overcoming logistical challenges in construction and deployment. AI-focused sites require a different design and build methodology. Companies that bring modular construction expertise and experience delivering remote, large-scale facilities can help accelerate deployment, reduce risk, and simplify the thorny prospect of building far outside established industrial hubs.

Land scarcity demands greater flexibility

Tied into the first two challenges is the problem of land. Data centres are big business, but more prosaically, they're very big buildings, often with attendant outbuildings for hosting UPS and backup generators. Securing suitable plots for data centre development is becoming increasingly difficult and expensive, especially in or near major metro areas. Sites with the right combination of access to power, fibre, zoning, and environmental compliance are increasingly scarce, driving up costs across multiple markets.

Developers are finding themselves faced with interminable permitting cycles, not to mention growing community scrutiny around noise, water usage, and sustainability.

These land constraints often force operators to make trade-offs: building more capacity in multiple locations with smaller footprints, compromising on ideal site layouts, or locating closer to power generation at the expense of also being close to an increasingly hostile general public.

Flexibility is becoming a key differentiator. Leveraging modular, highly customisable power systems makes it easier to adapt infrastructure to unconventional or constrained sites. This helps reduce the burden of site selection and enables operators to maximise usable capacity within limited or irregular parcels of land.

Tightening cross-border regulations and supply chain pressures

The global data-centre supply chain is already under significant strain due to surge in demand. Now, data centre companies (often international firms with sprawling supply chains) are having to contend with heightening cross-border complexity. Continued post-Brexit fallout is snarling up the border between the UK and Europe, and US tariffs — bad enough on their own — have ushered in a new era of frosty trade relations and more isolationist policies. Setting aside the increased cost of moving construction materials and data centre components around the world, customs complexities are also driving up lead times. Critical components such as switchgear, transformers, batteries, and high-density compute hardware all face growing bottlenecks. Deloitte reported recently that large original equipment manufacturers are increasingly locking themselves into multi-year agreements to produce key components, like transformers, switchgear, power management equipment, and power generation systems with single clients. As a result, these critical components are sold out, with restocks not expected to arrive for several years.

All of these challenges mean steep increases in lead times across core electrical infrastructure. These delays directly threaten deployment schedules for new builds and expansions.

To navigate this environment, data centres are leaning on partners with localised manufacturing, diversified supply networks, and the ability to operate across regulatory borders. Companies with operations in Northern Ireland, for example, enable clients to take advantage of UK-based trade agreements (including with the US) as well as Northern Ireland's unique relationship with the EU offering operators a way to mitigate cross border complexity Partnering with companies that have in-house engineering and assembly capabilities also reduces reliance on constrained international supply lines.

Evolving ESG regulations increase operational and design pressures

Governments are accelerating ESG mandates as net-zero deadlines approach, with the EU typically leading the way. For data centres, this means tighter scrutiny on power efficiency, emissions, water usage, generator runtime, and backup-power resilience.

At the same time, operators must secure access to a growing but still limited pool of clean energy sources. Variability in renewable generation introduces added complexity for UPS systems and backup infrastructure, creating concerns about reliability under fluctuating loads, especially when regulatory frameworks restrict generator size and design specifications.

Adopting modular, easily serviceable power systems allow for rapid replacement and minimal downtime. Predictive monitoring capabilities enhance visibility across power-distribution components, improving both "time to detect" and "time to resolve."

Flexibility and resilience in the face of a changing power landscape

In 2026, the power landscape for data centres will be defined by scarcity, complexity, and accelerating demand from AI. Operators that explore new locations, embrace modularity, rethink supply chains, and keep a watchful eye on evolving ESG requirements will be best positioned to deliver reliable capacity at scale.

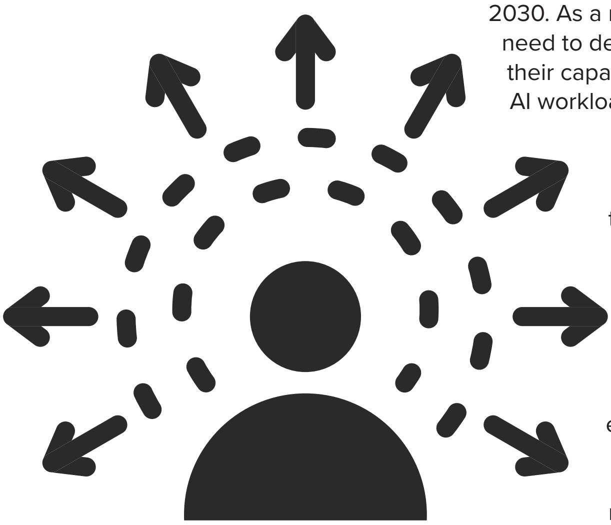
Partnering with organisations that have expertise related to remote builds, modular power systems, and cross-border operations is becoming an important competitive edge for data centre companies looking to navigate the challenges to come.

Bridging the knowledge gap: The role of data centres in shaping AI's future

In this new era of AI, data centres will not only power the digital services we rely on but will also be central to shaping the future of technology and society at large.

BY MARK PESTRIDGE, EXECUTIVE VICE PRESIDENT AND GENERAL MANAGER, TELEHOUSE EUROPE

AS PART OF A renewed focus on technological leadership, earlier this year, UK Prime Minister Sir Keir Starmer unveiled an ambitious AI strategy aiming to secure the UK's position among the world's AI frontrunners. Central to this vision is the expansion of data centre infrastructure, which will be essential to delivering the high-performance computing power needed to support advanced AI systems. Demand for AI-ready infrastructure is projected to rise by 33% each year until 2030. As a result, data centres will need to dedicate around 70% of their capacity to support advanced AI workloads.



In recognition of their growing importance, the UK government has classified data centres as part of the nation's Critical National Infrastructure (CNI), highlighting their role in sustaining the economy and enabling digital services. This shift signifies the increasing reliance on data centres to ensure the smooth functioning of national systems. Consequently, these facilities must be adequately protected from potential threats, including cyber-attacks, IT disruptions, and extreme weather events.

Widespread lack of awareness

Despite these advancements, public understanding of data centres remains alarmingly low. New research by Telehouse revealed that 51% of UK consumers have never heard of the term 'data centre', and 67% don't know what role they play in everyday life.

Among those who are at least somewhat familiar, 48% recognise data centres as essential for services like video streaming and online shopping. However, the scale of their operations remains largely unknown. 43% of the UK public are unaware of the vast network of people, applications, and data supported by data centres.

This lack of awareness is particularly concerning given the growing dependence on data centres to support AI, cloud computing, and digital infrastructure. As the government continues to prioritise data centres as critical national infrastructure, it's essential that public understanding keeps pace.

The importance of public perception and economic benefits

Public support is crucial for the expansion of data centre infrastructure. Without it, growth could face resistance, especially from communities concerned about the impact on local utilities and services. While the UK government's AI roadmap includes growth zones to streamline planning for these facilities, much of the discussion around data centres continues to focus on challenges such as energy consumption and resource demands. To address this, the data centre industry must educate the public on its essential role, along with its sustainability efforts, including renewable energy use and waste heat recovery, in order to build trust and strengthen community relationships.

At the same time, there is growing recognition of the significant economic and societal benefits that data centres support, particularly in relation to AI. According to IMF projections, AI could boost UK productivity by up to 1.5 percentage points annually, contributing £47 billion to the economy each year for the next decade. Beyond economic growth, AI promises to advance healthcare, enabling faster, more accurate diagnoses and enhancing remote care services. As demand for AI infrastructure rises, so does the need for skilled workers, with the data centre sector expected to create over 40,000 high-paying jobs by 2035. With these developments accelerating, the UK's AI market is projected to become the third-largest in the world by 2035, behind only the US and China.

In addition to these opportunities, the UK faces intense competition from global players. European and East Asian nations are already attracting significant AI investments, with Nordic countries benefiting from cheaper renewable energy and abundant land, making them ideal locations for large-scale AI operations. South Korea, for example, has announced plans for the world's largest AI-focused data centre, a $35bn project, while the US is progressing with its $500bn Stargate initiative. France, too, has committed €109bn in private-sector investments for AI development. To maintain its position as a global leader in AI, the UK must ensure it attracts and retains investment in AI and data centre infrastructure. Without this focus, the UK risks falling behind in the global AI race.

Emphasising sustainability

The data centre industry has already made significant strides to address its environmental impact. Many facilities are transitioning to renewable energy sources, while investing in innovations such as liquid cooling and hot and cold aisle containment systems to help reduce energy consumption. These solutions increase efficiency, lower operational costs, and decrease carbon emissions.

Free cooling methods, such as using outside air to regulate temperature, are also being integrated into existing facilities to reduce water and energy use. These sustainability efforts benefit the environment while helping data centres manage operational costs, ensuring their long-term viability as demand for AI infrastructure grows.

Data centres are also increasingly investing in apprenticeship programmes and technical training to equip the next generation with the skills needed to support the industry's expansion. Community engagement initiatives, including volunteering and STEM workshops, are further strengthening ties between data centres and local populations.

Enablers of AI ambitions

The UK's data centre industry is set to play a key role in realising the country's AI ambitions. However, this progress hinges on overcoming the knowledge gap among the general public. By educating citizens about the importance of data centres and their role in building a sustainable future, the industry can garner the support it needs to thrive.

The need for data centres as critical infrastructure will only increase as AI continues to evolve. By embracing innovation and sustainability, the industry can ensure it remains at the forefront of the digital revolution, driving economic growth and delivering lasting societal benefits. In this new era of AI, data centres will not only power the digital services we rely on but will also be central to shaping the future of technology and society at large.

Architecting for the AI era

How to build smarter, scalable data centres to meet evolving technology demands

BY TATE CANTRELL, CTO, VERNE

AI COMPUTE is accelerating power demand, which is projected to more than double by 2030 around 945 terawatt-hours (TWh) globally, applying pressure on data centres to meet rising capacity requirements.

At the same time, new EU regulations are pushing for greater efficiency and reduced environmental footprint as data centres with IT power demand of 500 kW or above must now publish their environmental KPIs, including energy usage, annually.

These new developments are reshaping expectations and prompting a reassessment of what data centre flexibility, scalability, and reliability should look like now, and in the future. The next era of data infrastructure must adapt to this demand for capacity and efficiency, requiring a reimagination of how data centres are designed, constructed, and operated.

Smarter cooling solutions

With the rising demand of AI and high-performance computing, traditional cooling methods are no longer sufficient for evolving data centres. These workloads require much more powerful hardware, particularly densely packed GPUs, which generate significantly more heat per task than conventional systems. Typical air-based systems rely on large volumes of chilled air and energy-intensive equipment. As a result, liquid cooling is becoming a pivotal component of data centre infrastructure, redefining how data centres tackle thermal challenges.

Liquid cooling uses fluids to absorb heat directly from the source. Beyond improving thermal efficiency, this approach permits higher operating temperatures, significantly reducing data centre's reliance on chillers and compressors. This innovation elevates cooling from a technical operation to a strategic differentiator in data centre design.

NVIDIA has already proposed equipment running at peak performance using 45°C liquid cooling, unlocking a future of greater efficiency across the industry. Such advancements would make it possible to eliminate compressors and their associated F-gases in all but the most extreme climates.

Consequently, operating at higher temperatures isn't a concession – it's a technological leap. Imagine getting rid of the air conditioning but still keeping the room cool. Liquid cooling is the smarter, simpler, and far more sustainable way forward.

Higher density, smaller footprint

The rise of AI is also forcing a rethink of how infrastructure is designed and deployed. Data centres will continue to see more concentrated computing power, with specialised hardware delivering greater processing capability per rack. And while data centre facilities continue to sprawl, higher-density architectures can reduce the overall impact by making more efficient use of space.

Notably, higher density compute means more efficient use of capital by maximising return on physical infrastructure investment. It also provides opportunities for better operational predictability by demanding tighter integration with customer workloads.

Density-led design is setting a new benchmark for performance, efficiency, and cost-effectiveness. AI is accelerating this shift, not by shrinking data centre footprints, but by redefining how space is engineered and optimised.

Bringing IT and facility systems together

The shift in design and thermal strategy is just one part of a broader transformation. As data centre demands evolve, IT and facility systems are being integrated to operate as a single, coordinated engine.

Traditionally, data centre infrastructure was designed and operated independently from the IT equipment it supports but the relationship between IT and facility systems is shifting to enable the next generation of compute.

This separation is no longer sufficient and combining servers, power, cooling, and controls into a unified, intelligent infrastructure is now essential for delivering the performance, efficiency and reliability high-intensity compute demands.

Joining these systems together represents not only a change in infrastructure but also in attitude. Operators are treating IT hardware less like general-purpose tools and more like special-purpose factory equipment. These systems are highly specialised, precisely engineered, and tightly integrated with the systems that support it.

This approach necessitates a fresh perspective on how operators coordinate space, power, cooling, and orchestration. With better integration of systems, data centres unlock smarter energy use, faster problem resolution, and scalable, high-density deployments.

Sustainable design strategies

With the rapid rise of power-hungry AI workloads, there is more pressure than ever for data centres to address their energy intensive activities, especially as AI workloads grow in scale and intensity. Innovations in liquid cooling, energy optimisation, and efficient hardware design are helping to reduce environmental impact. But their impact is limited if the underlying energy source remains carbon intensive.

That's why operators must consider where compute happens when managing these demands. Locating AI infrastructure in data centres powered by renewable energy offers a smarter, lower-carbon impact future. Locations such as the Nordics are rich in clean energy, enabling organisations to significantly reduce the carbon cost of innovation and high-density compute.

Designing for the future

Against a backdrop of rapid technological, economic, and social change, adaptability in data centre design isn't just beneficial – it's essential.

Those who embrace innovation in build processes, design infrastructure for flexibility at scale, and integrate sustainability not as an afterthought, but as a foundational principle, hold the key to success. These strategies aren't passing trends. They're fast becoming the new standard for high-performing, future-ready data centres.

Data centres are the beating heart of the UK's AI ambitions, but first they need decarbonising

A future built on AI must be powered by clean, robust infrastructure, which means treating data centres as the foundation to sustainable economic strategy.

BY JON HEALY, MANAGING DIRECTOR, EMEA AT SALUTE

IT'S NO SECRET that the UK has big ambitions when it comes to artificial intelligence. National strategies are being backed by major R&D investments in both the private and public sectors, so there's clear momentum. But powering that vision requires infrastructure. Specifically, data centres. These facilities are essential for training AI models, storing vast datasets and keeping digital services running around the clock. Data centres are fundamental to the UK's future as an AI leader, but they come with a cost that extends beyond the business' purse strings.

Data centres consume massive amounts of electricity. As AI applications expand, so too will the carbon footprint of the systems supporting them. If we carry on with this trajectory without considering the environmental impact, this growth risks clashing with the UK's climate goals and undermining its credibility as a leader in sustainable innovation. The ambition to be a leader in one area should not displace another.

The International Energy Agency's special report 'Energy and AI' projects that electricity demand from data centres worldwide is set to more than double by 2030 to around 945 terawatt-hours (TWh). To put that in perspective, that's slightly more than the current total electricity consumption in Japan.

Our data centres are critical, but they need green infrastructure. Energy efficiency cannot be pushed aside in favour of fulfilling AI ambitions. There is a real opportunity for the UK to not only take its place amongst global leaders in AI, but stand above the rest as the champion of the green infrastructure powering this rising technology. To do this, we must weave sustainable practices and solutions into every aspect of data centre design and operation, from how facilities are cooled to how power is sourced. We're seeing common themes across nations as the world contemplates this growing demand, the main ones being scaling energy grids, the challenge of building next-generation facilities and complying

with increasingly strict environmental regulations – all within constrained energy markets.

If the UK wants to lead, it needs to grab with both hands the opportunity to align the two global agendas that are already shaping what the world will look like in 50 years' time. A study conducted by the Grantham Institute and the London School of Economics suggests that by 2035, artificial intelligence might help cut CO2 emissions by up to 5.4 billion tonnes. It all comes down to data centres.

A problem shared is a problem halved

Pressure on policy is growing throughout the region. Regulations like the Corporate Sustainability Reporting Directive (CSRD) in the EU are increasing the bar for emissions reporting's standardisation and transparency. At the same time, operators find it more difficult to meet investment standards without deeper decarbonisation initiatives owing to the EU Taxonomy's stringent energy-efficiency guidelines on what constitutes sustainable infrastructure.

Data centre decarbonisation is not limited to sustainability specialists or facilities teams. Operations, engineering, compliance and policy should all work in conjunction on this. From choosing the site and building it to sourcing power and maintaining equipment, each of these areas aids in reducing emissions over the course of a site's lifecycle. Until a collaborative approach is achieved, we'll keep seeing fragmented progress.

We also need to face the fact that just being more efficient isn't enough anymore. As AI training and demand grow, so will the need for energy. Operators need to look at a whole-systems approach focusing on renewable energy, designs that reuse materials, and clear tracking of all types of emissions.

The barriers in play

There are already bottlenecks in the infrastructure required to support AI at scale. Grid capacity is constrained in many high-demand regions, including the South East and London. According to some providers, new connections won't be accessible for another ten years. As a result, the UK's potential to boost local computing capacity is restricted.

The development of new data centres is hampered by delays in planning and uncertainty about long-term energy availability. The UK runs the risk of turning into a net importer of computing power if these problems continue, thereby losing control and innovation to nations with more developed infrastructure plans. The UK cannot afford this situation if it hopes to maintain its global position.

The UK is not alone, in the Netherlands and Ireland, local resistance, environmental concerns, and limited grid headroom have resulted in moratoriums or delayed permissions for new data centre builds. Without cross-border coordination and long-term energy investment, EMEA's AI ambitions could run into the same constraints already visible in the UK.

Aligning two global agendas

While funding for AI research is essential, data centre capacity and sustainability investments must also be treated with equal importance. The incentives of embracing clean energy, renewing old infrastructure, and creating quicker grid connections must also be considered by policymakers. Clear planning frameworks can help construct new and effective facilities in underserved areas.

Aligning the growth of AI infrastructure with the UK's climate commitments is also necessary. There is a chance that advancements in one area could jeopardise commitments in another if there is no clear guidance on how compute growth fits into national carbon budgets.

To stay competitive and climate-conscious, EMEA governments must align AI growth targets with realistic infrastructure planning. This requires strategic coordination between regulators, operators and power providers - not just within borders, but across them.

Data centre operators are currently facing a two-fold problem of dealing with complex infrastructure while also trying to cut down on emissions. Many operators are looking into options like modular construction, advanced cooling technologies, and the use of renewable energy sources. But constraints from outdated decisions about existing infrastructure and inconsistent reporting standards across the industry often slow down progress. Data centres need to be more open and have stringent ways of measuring emissions during both the construction and operational phases. To address these problems, setting benchmarks, writing detailed reports and sharing information across the industry are the need of the hour.

> The development of new data centres is hampered by delays in planning and uncertainty about long-term energy availability

Embrace the opportunity or undermine national ambition

The UK's data centre sector, as the key to achieving the UK's AI goals, is in a strong position to shape the path we take next. As demand for AI capabilities accelerates and the urgency of decarbonisation intensifies, the opportunity and the responsibility to lead through innovation has never been clearer. It needs a coordinated, whole-systems approach, one that brings together government, industry leaders, energy providers and local authorities – all working as one. Thorough planning, grid readiness and incentives for sustainable design must be central to policy. Without this alignment, the UK risks falling behind in both climate commitments and market competitiveness.

Beyond our national borders, closer coordination across the EMEA region could help drive wider consideration and adoption to greener infrastructure as data centres continue to be built across all countries to fulfil the AI demand. Shared standards, joint investment in low-carbon grids and cross-border innovation frameworks will be essential to scaling sustainably. A future built on AI must be powered by clean, robust infrastructure, which means treating data centres as the foundation to sustainable economic strategy.

Reinventing data centre design with AI and modular solutions

The rapid evolution of AI, generative workloads, and high-density compute environments is pushing the industry toward highly modular, AI-optimized, and sustainable data centre designs.

BY TOD HIGINBOTHAM, CEO, ZINCFIVE

WITH THE EXPLOSION of AI as driver, hyperscalers and enterprises alike are pivoting from traditional, boxy data centers and sprawling server farms to modular, AI-optimized, and vertically scalable data centers. According to the 2025 Data Center Energy Storage Industry Insights Report, 86% of respondents anticipate a shift toward more customized power solutions to accommodate AI demands.

Construction of data centers has increased more than sevenfold in two years, based on findings from commercial property giant JLL. Rising demand for high-performance compute (HPC) environments, rapid scalability, and energy efficiency is shaping the next generation of infrastructure. Data centers must now balance power, heat, and space constraints while ensuring rapid deployment and cost efficiency. In response, the industry is embracing:

- Modular Configurations for Scalability and Efficiency
- Vertical Expansion in Urban and Suburban Markets
- AI-Driven Cooling and Power Innovations

Modular configurations: A scalable solution for AI

Prefabricated modular data centers (MDCs) are becoming the preferred solution for hyperscalers and enterprises looking to rapidly scale AI workloads. According to the same Energy Storage Industry Insights Report, two-thirds (68%) of respondents use modular power solutions and plan to continue, while 22% deploy them at all locations. Unlike traditional fixed infrastructure, modular components can be preassembled, transported, and deployed quickly, reducing build times while optimizing efficiency.

The Open Compute Project Foundation (OCP) pioneered modular innovation with its Open Rack standard, which expands airflow, optimizes cooling, and increases rack density for AI and cloud workloads. Similarly, Tesla's Dojo supercomputer is a case study in modular efficiency, with its Training Tile architecture enabling incremental scaling of AI infrastructure.

Meanwhile, the NCAR-Wyoming Supercomputing Center (NWSC) highlights how modularity supports sustainability and energy efficiency. Located in Wyoming, the facility leverages natural cooling and local wind energy, proving that modular designs can drive both performance and environmental responsibility.

Vertical expansion: The rise of high-density data centers

The industry is seeing a paradigm shift in data center architecture as rising land costs and urban proximity demands push operators to build vertically rather than horizontally.

Traditionally, data centers have been single-story, sprawling structures, but today, multi-story designs are becoming the norm, especially in high-density metro areas. According to a recent Wall Street Journal report, major operators like Equinix are developing 12-story data centers in Amsterdam and eight- to 10-story facilities across the U.S. to keep infrastructure close to users and reduce latency. Suburban expansion follows a similar trend. Markets like Chicago, Dallas, and Atlanta are seeing a surge in three- to five-story data centers as land scarcity and zoning restrictions drive new designs.

Two- and three-story data centers have become the norm in cities, according to JLL.

AI-Optimized cooling and power infrastructure

AI workloads dramatically increase power density, with racks now exceeding 100kW and generating massive heat loads – far beyond what traditional air cooling can handle. The solution? AI-driven cooling and next-gen power systems.

With air cooling maxing out at 20-30kW per rack, hyperscalers are rapidly adopting liquid cooling to meet AI demands. Companies like Flex and Inspur are leading with direct-to-chip and immersion cooling technologies, ensuring efficient heat dissipation while maximizing compute performance.

Power innovations: Modular battery systems for cost-efficient deployment

Cost-efficient power solutions are a critical component of modular data center growth. In this environment, modular battery cabinets powered by nickel-zinc (NiZn) technology emerge as a cost-effective and scalable alternative to traditional power backup systems.

The total cost of data center buildouts is a key concern, particularly for enterprises looking for scalable power solutions. Modular power units using nickel-zinc-battery-based immediate power technology significantly reduce costs in multiple ways:

- **Smaller footprint:** Nickel-zinc battery technology is 3x more power dense than VRLA and 2x more power dense than Lithium-ion; therefore, requiring less linear space per module—a savings of up to $10,000 per linear foot.
- **Inherent chemistry characteristics:** NiZn chemistry is inherently safe at the cell level, as well as non-toxic, and performs well at temperatures up 35 degrees Celsius. With these characteristics NiZn chemistry decreases the need for cooling and fire suppression equipment unlike lithium-ion alternatives.
- **Factory-sealed, build-ready design:** Unlike lithium-ion batteries, which require on-site installation, NiZn battery cabinets arrive pre-packaged and ready for immediate deployment.
- **Lower shipping and installation costs:** The modular, plug-and-play design minimizes installation complexity, streamlining data center expansion and upgrades.

By leveraging modular power solutions like NiZn battery cabinets, data center operators can reduce infrastructure costs, accelerate deployment, and enhance overall energy efficiency.

The future of AI data centers: modular, vertical, and sustainable

The rapid evolution of AI, generative workloads, and high-density compute environments is pushing the industry toward highly modular, AI-optimized, and sustainable data center designs. Companies like Equinix, Inspur, and ZincFive are demonstrating how modular configurations, vertical expansion, and AI-driven power solutions can meet the next decade's infrastructure demands.

Powering intelligence: GIS and the global race to build AI data centres

The global race to power AI has triggered an unprecedented boom in data centre construction, reshaping energy grids, telecommunications networks, and the digital economy itself. This surge is remodelling infrastructure demands across the telecommunications and utilities sectors – particularly in the UK, Germany, and the USA.

BY BUDDIE CERONIE, GM FOR TELECOMMUNICATIONS, VERTIGIS

THE USA is experiencing rapid growth in hyperscale and colocation data centres, driven by AI workloads and cloud migrations. AI data centres are projected to push US energy demand up by 26% by 2028, requiring over $2 trillion in new energy generation resources. However, utilities are struggling to predict how much of this demand will materialise, creating uncertainty around infrastructure investment and grid planning.

Germany is planning 43 new data centres, including both hyperscale and colocation facilities. This expansion complements its existing portfolio of 187 operational data centres, reinforcing its position as one of Europe's leading digital infrastructure hubs.

In September 2025, the USA announced significant investment in the UK's AI sector through the UK–US Tech Prosperity Deal. This landmark agreement aims to deepen collaboration on AI, quantum computing, and nuclear technologies, backed by £31 billion in private investment from major US tech firms, including Microsoft, Google, NVIDIA, OpenAI, and CoreWeave. This agreement marks a pivotal moment, positioning the UK as a central hub in the global AI ecosystem.

As a result, the UK government is planning nearly 100 new data centres by 2030. This represents a 20% increase that will solidify its position as the third-largest global data centre market after the USA and Germany.

Reinventing infrastructure amid decades of underinvestment

While the surge in AI data centres is driven by demand for AI processing power, it raises concerns over energy consumption, water usage, and sustainability. It also puts pressure on utilities, which must reinvent themselves to meet surging demand while balancing reliability, affordability, and environmental impact. AI workloads require high-density computing, translating into massive power draw and cooling needs. This shift is forcing utilities to rethink grid architecture, invest in renewables, and accelerate permits for new generation capacity.

Simultaneously, the data centre boom is having a major impact on telecommunication networks. They are facing pressure to support ultra-low latency, high-bandwidth, and edge computing capabilities for AI inferencing. The surge is driving demand for fibre expansion, 5G densification, and network resilience to support real-time AI applications. AI data centres also require robust interconnectivity, pushing telcos to upgrade backbone infrastructure and peering arrangements.

The race to power AI is more than a technical one – it has significant geopolitical, economic, and environmental ramifications. All sectors must collaborate to avoid bottlenecks and ensure scalability, while governments and regulators are being urged to streamline approvals and incentivise sustainable builds.

However, both the telecommunications and utilities sectors have suffered from decades of underinvestment, with ageing infrastructure stifling innovation. To prepare for the AI-driven surge, they must accelerate digital transformation and modernise IT systems to keep pace.

GIS key to AI infrastructure planning

This is where GIS (Geographic Information Systems) technology becomes a game-changer. It helps telcos and utilities modernise and optimise network design, deployment, and lifecycle management. In the race to build and deploy AI-powered data centres, GIS plays a critical role by providing spatial intelligence, operational oversight, and infrastructure optimisation, from site selection to infrastructure mapping to operational monitoring. Here is how GIS supports these developments:

- **Site selection and planning**
  GIS helps identify optimal locations for new data centres by analysing proximity to power grids, fibre networks, and water sources. It also evaluates environmental risks such as flood zones, seismic activity, and other natural hazards. Additionally, GIS helps organisations understand regulatory constraints and land use patterns, ensuring strategic placement that balances performance, cost, and sustainability.
- **Infrastructure Mapping and Visibility**
  Using GIS technology, real-time mapping of assets, such as power lines, substations, cooling systems, and fibre routes, allows operators to visualise dependencies and vulnerabilities across the entire ecosystem, including utility interconnects and telecommunications backbones. This spatial awareness is critical for identifying single points of failure, planning redundancies, and ensuring seamless service delivery. GIS platforms support dynamic layering of infrastructure data, enabling teams to overlay environmental factors, zoning constraints, and traffic patterns for informed decision-making.

  As AI data centres become more distributed and resource-intensive, this visibility empowers stakeholders to optimise routing, anticipate bottlenecks, and coordinate upgrades across jurisdictions and providers.

- **Operational Monitoring**
  GIS integration with IoT sensors and SCADA (Supervisory Control and Data Acquisition) systems provides continuous, real-time monitoring of critical infrastructure. This includes tracking energy consumption, cooling efficiency, and emissions, enabling operators to maintain performance and compliance.

  By visualising operational data spatially, GIS helps pinpoint inefficiencies, detect anomalies, and anticipate failures before they occur. It supports predictive analytics to forecast maintenance needs and infrastructure stress, reducing downtime and extending asset life. Automated alerts triggered by threshold breaches allow for rapid response and targeted interventions.

  GIS can also layer historical data with live telemetry to uncover trends, optimise resources, and inform upgrades, all within a unified, map-based interface.

- **Risk Management and Resilience**
  Models powered by GIS simulate climate impacts, cyber threats, and supply chain disruptions, supporting contingency planning and disaster recovery through the mapping of alternate routes and backup systems. These simulations enable utilities and telcos to identify vulnerabilities proactively and develop robust contingency plans.

  By visualising alternate routes, backup facilities, and interdependencies across networks, GIS supports faster disaster recovery, ensures service continuity, and helps prioritise investments in areas most susceptible to disruption. This spatial intelligence empowers decision-makers to balance risk, cost, and sustainability in real time.

- **Stakeholder Collaboration**
  With a diverse array of stakeholders, from utilities and telcos to regulators, developers, and local authorities, the complexity of AI data centre expansion demands more than technical oversight. Transparent reporting and strict compliance with environmental and zoning regulations are essential for responsible growth and public trust.

  GIS dashboards enable seamless cross-functional coordination, offering a unified view of infrastructure assets, regulatory boundaries, and project timelines. This spatial intelligence empowers decision-makers to align priorities, mitigate risks, and accelerate approvals while maintaining accountability across the ecosystem.

As AI data centres become more complex and resource-intensive, GIS transforms from being viewed as a simple mapping tool into a strategic infrastructure command centre, essential for scaling responsibly and sustainably.

The AI data centre boom is not just a technological revolution; it's a test of how well nations can align digital ambition with physical infrastructure. GIS stands at the heart of this transformation, ensuring growth that is not only rapid but also resilient, intelligent, and sustainable.

By visualising operational data spatially, GIS helps pinpoint inefficiencies, detect anomalies, and anticipate failures before they occur. It supports predictive analytics to forecast maintenance needs and infrastructure stress, reducing downtime and extending asset life

How data centres are adapting to the demands of AI

Artificial intelligence (AI) workloads are already transforming data centres, and at speed. From power and cooling to deployment timelines and security, the infrastructure beneath modern compute is being rethought. But for many operators, that pace is creating pressure points across the board.

We spoke with Jon Abbott, Technologies Director, Global Strategic Clients at Vertiv, about how the infrastructure conversation is shifting, what risks are emerging, and where operators are finding smarter ways to keep up.

Q: What's the biggest pressure AI is putting on data centre infrastructure right now?

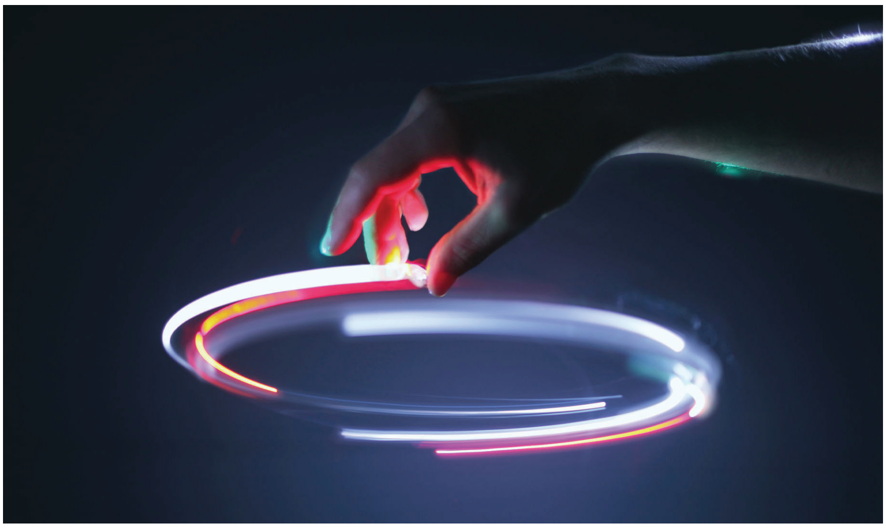
IT'S THE GAP between how fast AI is scaling and how slow IT infrastructure is to respond. Compute teams are pushing for denser, more powerful systems that can be stood up quickly. But physical infrastructure still has long lead times, not just for procurement, but for power upgrades, cooling integration, and site-level planning.

The result is a growing number of operators realising that what worked two years ago no longer matches the demands coming their way. AI isn't something that can be layered onto existing systems without consequence. It changes the shape, the speed and the risk profile of the entire stack.

Q: How is cooling infrastructure adapting to AI-scale deployments?

Precision and flexibility are both essential. AI workloads are highly dense and thermally intense, which pushes air-based systems to their limits. Liquid cooling, whether rear-door or direct-to-chip, is no longer a future consideration. It's something which is being specified now.

What we're seeing now is a shift to hybrid environments. Air continues to handle lower-density and peripheral equipment, while liquid cooling systems target the highest density loads. But integrating those two modes introduces complexity. There's a greater need for synchronised monitoring, fluid network management, and clear maintenance protocols that align with new risk profiles.



Cooling infrastructure is increasingly tied to workload reliability and performance, so it's a critical system that must be tuned, monitored and managed with the same level of oversight as compute and power.

Q: Is there a typical infrastructure model that's emerging, or is every site different?

The pressure to move faster is making prefabrication and modularity more attractive. Operators want systems that arrive ready to deploy, that reduce variables on site, and integrate power and cooling into a known, tested unit.

There's also more joint planning between disciplines as well as third parties such as technology partners, integrators and utilities. Power, thermal and software teams are sitting down earlier to map out system behaviour - not just specifications. And the more AI becomes business-critical, the more that kind of coordination becomes non-negotiable.

You still see variation between colocation, hyperscale and enterprise sites, but the trend is clear: integration over isolation, speed over customisation, visibility over assumptions.

Q: Are operators rethinking power in response to AI too?

Definitely. Higher rack power density changes everything - from uninterruptible power supply (UPS) provisioning to distribution pathing. But it's also about behaviour. AI inference workloads, for instance, can cause sudden fluctuations in how much power is required. So power systems need to respond in real time, not just hold steady under flat loads.

Beyond resilience, there's growing interest in how power infrastructure can provide operational and financial flexibility. That includes grid participation, demand response, or integrating battery assets to manage volatility. In Europe, especially, these capabilities are becoming a serious part of long-term planning.

Q: What operational risks are going unnoticed in some AI deployments?

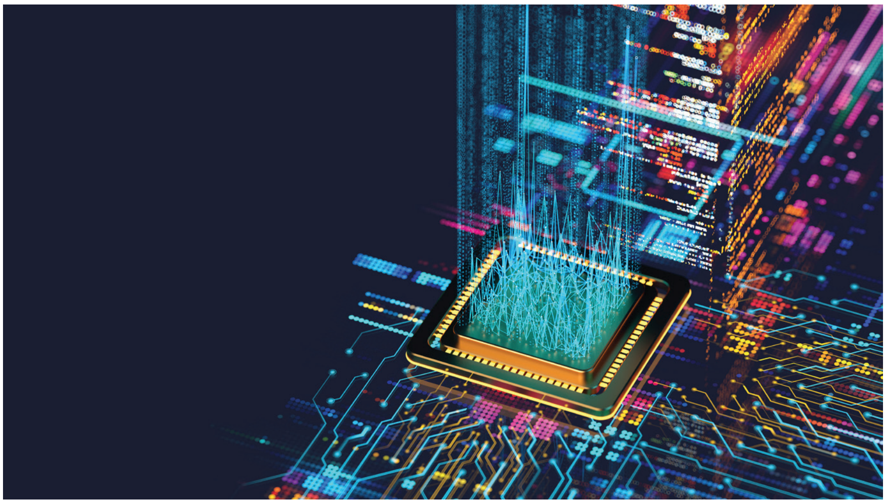
Hybrid cooling - integrating different cooling systems (air and liquid) to achieve seamless operation - can be challenging if not managed properly. Therefore, careful planning is required, which means collaboration between IT, facility, power teams and specialist expertise. If the airflow and liquid systems aren't coordinated, efficiency and performance could be negatively impacted.

Another issue is inconsistent commissioning. When upgrades are layered onto legacy infrastructure under time pressure, systems often drift out of sync. That could be mismatched firmware, overlapping alarms, or gaps between facility-level and IT-level telemetry.

These risks can be managed, but only if integration is planned from the start, and operations teams have access to full-stack visibility.

Q: Is physical security evolving alongside this shift?

Yes, particularly in sites supporting AI model training or housing high-value infrastructure. Operators are revisiting access control, real-time surveillance, and rack-level authorisation. There's more interest in role-based policies that align with operational states - so



access privileges can be time-bound or workload-specific.

In automated or lightly staffed environments, the integration between physical security and operational telemetry becomes even more important. If a door opens outside an expected maintenance window, that should generate the same level of response as a power fault. Security can't sit outside the main operational workflow anymore.

Q: How are sustainability and regulatory expectations influencing infrastructure decisions now?

Sustainability strategy is no longer an isolated workstream. It's increasingly tied to risk management, cost control, and even market access. Operators are now expected to demonstrate how infrastructure choices align with energy efficiency goals, emissions targets and, in some regions, mandatory reporting requirements.

> The data centre that succeeds won't necessarily be the one with the most power or space - it'll be the one that can change direction quickly, bring new capacity online fast, and manage it intelligently

Cooling systems are being evaluated not just on effectiveness, but on refrigerant type, lifecycle emissions and water use. Power procurement strategies are shifting to include cleaner sources, on-site generation and more intelligent load balancing. Even material selection and equipment disposal are under review.

These changes are being driven by more than ESG pressure. In many cases, they're becoming prerequisites for securing energy connections, local planning approval or customer contracts. For forward-looking operators, this is prompting a more integrated approach where performance, resilience and sustainability are designed together from the outset.

Q: What defines a future-ready data centre in this environment?

Adaptability. The data centre that succeeds won't necessarily be the one with the most power or space - it'll be the one that can change direction quickly, bring new capacity online fast, and manage it intelligently.

That means tighter integration between systems, more use of modular and prefabricated components, and operational visibility that extends from the power room to the workload layer. It also means rethinking risk - not only in terms of uptime, but in terms of latency, energy cost, security, and speed to deploy.

AI is a demanding tenant. Infrastructure has to evolve to match that expectation - not in the abstract, but right now.

From digital dreams to physical reality

The differentiator in the AI race is no longer just data science or software talent. It is the readiness of the data centre behind it.

BY DARREN WATKINS, CHIEF REVENUE OFFICER AT VIRTUS DATA CENTRES

ARTIFICIAL intelligence (AI) continues to dominate boardroom conversations across the UK. In January this year Gov. UK announced that the UK AI sector attracted £200 million a day in private investment since July 2024. The scale of the ambition is clear, yet a critical disconnect is emerging. Organisations are investing heavily in AI talent, data strategies, and cutting-edge algorithms, yet many are discovering a fundamental truth: AI can't run on aspiration alone.

This mismatch is born out by the numbers. According S&P Global Market Intelligence, the share of businesses scrapping most of their AI initiatives rose to 42%, up from 17% last year. The problem is rarely the model or the data scientists - it is the lack of physical environments capable of sustaining AI workloads. Running AI at scale requires a new breed of data centre, those that are purpose-built facilities engineered for high density, advanced cooling and resilience.

This is a major part of why AI initiatives might stall. Enterprises may have the right data scientists and the right models, but without access to an AI-ready data centre, they hit a physical barrier. Inadequate infrastructure – lack of appropriate hardware, software or cloud infrastructure to manage data and deploy AI models – is hindering AI progress. EY's AI Pulse Survey found that 67% of senior business leaders admitted that their current infrastructure is actively slowing down their AI adoption.

So what infrastructure is required to help AI initiatives be successful?

AI workloads require specialist infrastructure that resides in AI-optimised data centres and run by data centre experts. These AI-optimised data centres house specific hardware such as graphics processing units (GPUs), tensor processing units (TPUs), and application-specific integrated circuits (ASICs) for their high parallel processing capabilities, which is essential for training and running complex AI models. Therefore, it is much more practical and cost effective to deploy AI applications in purpose-built facilities operated by data centre specialists.

AI readiness starts with construction

Much of the conversation about AI and infrastructure focuses on algorithms, processors, or cloud models. But AI readiness is grounded in the physical construction of the data centre itself. The design decisions made at build stage directly determine whether a facility can host and scale for AI.

In short, an AI-ready data centre is not just another building with racks. It is a purpose-engineered environment where design, materials and systems converge to support workloads that would overwhelm legacy facilities.

The data centre is no longer a neutral warehouse for IT, it has become an optimisation layer in the AI value chain.

Three data centre priorities for AI-ready infrastructure

High-density power readiness

AI racks are designed to handle processing power far beyond traditional levels. They have to have a high load capacity, ample space, configurable options, optimised airflow and cable management features. Because of this, they are much heavier than previous iterations and can weigh several tonnes.

Many existing facilities were never built to accommodate these heavier racks, or the added weight of coolant and water required for liquid cooling, so the base build of new AI facilities need to be enhanced. This means reinforced floors to cope with increased structural weight, modular halls, and layouts designed for density.

High-density workloads also require not just more power, but smarter power. Redundant distribution paths, advanced uninterruptible power supply systems (UPS), and the ability to segment power delivery by rack are critical.

Precision cooling
AI applications demand extensive computational power to process vast amounts of data and perform complex tasks. This computational intensity translates into significant heat generation.

Traditional air cooling can sometimes be insufficient when racks are running above 50 kilowatts. Liquid cooling technologies play a pivotal role in enhancing performance, increasing energy efficiency and improving the reliability of AI-centric operations. The adoption of advanced liquid cooling technologies not only optimises heat management and reuse, but it also contributes to reducing environmental impact by enhancing energy efficiency and enabling the integration of renewable energy sources into data centre operations.

However, liquid cooling system are complex. Unlike air-based cooling systems, liquid cooling solutions require specialised components, such as cooling distribution units (CDU's) which must be carefully integrated into data centre infrastructure. Designing a facility with liquid and immersion cooling systems capable of continuous performance means embedding pipework, pumps and containment systems into the architecture.

Today, many data centres are transitioning from 100% air cooling to a hybrid model encompassing air and liquid-cooled solutions to improve efficiency, performance, and sustainability.

Data and compute proximity
A centrally located facility may deliver economies of scale, but it cannot overcome physics. If the compute is too far from the user, latency follows. And the tolerance for latency has collapsed in the AI era. AI models are only as good as the data they are trained and run on. Yet data often sits in multiple locations: across cloud platforms, enterprise systems and edge devices. When compute is physically distant from the data, latency rises and model accuracy suffers.

Internal teams may tolerate a half-second delay when running a batch job, but customers interacting with an AI-powered chatbot or real-time fraud system will not. Increasingly, enterprises are deploying AI applications in distributed data centres that bring compute closer to the user base.

For use cases such as autonomous vehicles, fraud detection or personalised services, milliseconds matter. If inference workloads are running in a facility hundreds of miles away from the data source, the user experience collapses. This is why proximity to data has become a performance factor in its own right.

The principle is straightforward - reduce physical distance and latency is reduced, and accuracy is improved. In practice, this reshapes the geography of data centre demand. Enterprises are beginning to choose data centre locations not just on cost or capacity, but on closeness to key datasets and user bases. Facilities near financial hubs, media clusters and population centres are becoming strategic assets.

> AI is often framed around algorithms, skills and regulation. But without the physical foundations, none of these deliver value. Enterprises cannot run AI reliably or sustainably in legacy facilities

Data centres as strategic AI enablers

AI is often framed around algorithms, skills and regulation. But without the physical foundations, none of these deliver value. Enterprises cannot run AI reliably or sustainably in legacy facilities. The future lies in AI-ready data centres - environments built for density, proximity and flexibility.

The organisations moving fastest on AI are those treating data centre strategy as core business strategy. They are aligning their AI objectives with facilities capable of delivering power, cooling and resilience at scale.

The differentiator in the AI race is no longer just data science or software talent. It is the readiness of the data centre behind it.

Bridging the talent gap to power the UK's AI future

Addressing the data center talent shortage is crucial in supporting the UK's broader digital and AI ambitions.

BY VIVEK SWAMINATHAN, DIRECTOR OF PRODUCTS AND SOLUTIONS, DIGITAL WORKPLACE SOLUTIONS, UNISYS

THE UK IS on pace to accelerate its artificial intelligence (AI). It recently announced plans to unlock investments for AI Growth Zones nationwide by improving access to power and creating highly skilled jobs. To accomplish these goals, data centers are emerging as critical infrastructure, and the U.K. has proposed sites to host 500 megawatts of AI infrastructure by 2030.

As these AI ambitions grow, success does not hinge solely on innovation. Putting in place the appropriate infrastructure and ensuring there are enough trained workers to operate these data centers is imperative. Without it, the industry could fall short of its potential.

While the U.K. government has pledged £1.2 billion in investments and the creation of thousands of new tech jobs by 2028, a growing talent shortage in the data center sector threatens to derail progress. Furthermore, with AI workloads surging and data center demand projected to grow 20% by 2030, the industry faces a pressing question: Who will build and run the digital backbone of its future?

The data center talent gap could cause undue bottlenecks and losses in revenue

As recently as 2024, 68% of U.K. IT leaders cited insufficient skills and expertise as the top barrier to AI adoption. This challenge is especially acute in the data center sector, where such roles as infrastructure architects, systems engineers and facilities managers require specialised, often under-promoted skillsets.

As a result, traditional recruitment strategies are failing and operational bottlenecks are forming. If left untreated, hiring in this landscape will only become more competitive, making it more difficult for organisations to find the right people. This will lead to delays in deployments, increased downtime at data centers due to system breakdowns and wasted time and resources.

Businesses that invest in upskilling will lead the UK's AI landscape

Forward-thinking companies are not waiting for the right talent to find them. Rather, they are creating and implementing proactive recruitment plans and developing their existing workforce to address tomorrow's technological needs. Global tech companies like NVIDIA and Salesforce are clear leaders in this space, investing in scalable training platforms to build internal capabilities and reduce reliance on hiring external candidates.

Bypassing traditional recruitment limitations, firms like Hearst UK are mirroring these efforts by launching in-house academies and partnering with training providers to create tailored learning pathways. These proactive strategies are helping businesses attract and retain talent in a market where job-hopping and skills mismatches are increasingly common. By offering clear development opportunities and career progression paths, companies can foster loyalty and reduce turnover — two important factors in maintaining operational continuity in high-demand environments like data centers.

Training programs should not be viewed as costs in today's digital economy — they are strategic assets. Businesses that embed continuous learning into their operations are better positioned to adapt to evolving technologies and customer demands. Additionally, companies committed to making high-quality AI training resources widely available — including Salesforce, IBM, Microsoft, and BT — will significantly benefit the U.K.'s workforce, by upskilling 7.5 million workers by 2030. These efforts support internal workforce development and ensure businesses can keep pace with rapid technological change. For data centers, this means faster deployment, better uptime and stronger innovation pipelines.

Collaboration Is key to solving talent challenges

While many companies are taking necessary, proactive steps to ensure the workforce is enabled to use AI, no single organisation can solve the talent gap alone. To meet industry demands, Public-private partnerships between government agencies and industry leaders are essential to meet the demand at scale.

While many companies are taking necessary, proactive steps to ensure the workforce is enabled to use AI, no single organisation can solve the talent gap alone

Launching the Isambard-AI supercomputer in the U.K. city of Bristol is a prime example of how these partnerships can succeed. Enabled by collaboration between NVIDIA, the University of Bristol and the U.K. government, this supercomputer is expected to provide unprecedented capabilities for researchers and industry to harness AI's immense potential in robotics, big data, climate research and drug discovery. Furthermore, the project showcases how shared infrastructure and talent development can accelerate innovation and workforce readiness.

Salesforce's AI Center in London is another model for success. Designed not only to support enterprise AI adoption, the center also serves as a hub for workforce development—offering training, community engagement and digital skills programs that reach beyond the tech sector.

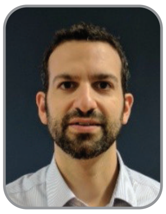
Meanwhile, organisations like techUK are working to ensure that AI and digital infrastructure growth is matched by talent development. By advocating for scalable, business-led training models that can be replicated across regions and sectors, these efforts are helping to build a more inclusive and resilient digital workforce — all of which will be critical to advance data centers in the U.K.

Building the workforce behind the infrastructure

Addressing the data center talent shortage is crucial in supporting the U.K.'s broader digital and AI ambitions. While critical steps are being taken every day, sustained investment in skills development and industry collaboration will be essential to ensure long-term resilience.

By focusing on practical training and inclusive pathways into the sector, the U.K. can strengthen the foundations of its digital economy and remain competitive in a rapidly evolving global landscape. The future of AI doesn't just depend on algorithms — it depends on the people who power the infrastructure behind them.

Innovation without compromise: The case for digital twins



Ultimately, in an industry undergoing scrutiny and accelerating demand, digital twins offer a smarter, more adaptive way forward.

BY MARK FENTON, PRODUCT ENGINEERING DIRECTOR AT CADENCE

POWER GRIDS keep countries ticking, generating and transporting power. However, a staggering 70% of data center facility leaders around the globe say their national power grid is being stretched to its limits. One key driver of this growing pressure is artificial intelligence (AI). As AI adoption accelerates, powering everything from drug discovery to fraud detection and autonomous assistants, the strain on the data centers that support it is becoming harder to ignore. In fact, three-quarters (74%) of data center operators report that their data center is experiencing increased demand from AI-driven workloads.

As AI reshapes industries and intensifies infrastructure demands, data centers must find ways to scale responsibly. That means reducing the load on shared energy and water resources to help safeguard national grids and the communities that depend on them.

The good news is that AI can equally empower data center leaders, too. With tools like digital twins—virtual models of physical operations—operators can better test, monitor, and fine-tune data center performance. Even better, AI-enhanced digital twins offer more predictive insights and smarter automation, helping to reduce AI's environmental impact and boost overall efficiency. It's ultimately a more intelligent approach that brings greater control to an increasingly complex industry.

Uncovering pain points

Before making any changes, data center operators must first determine exactly where efficiency challenges reside. Commonly, this starts with identifying stranded capacity—installed capacity in the data center that cannot be used. Imagine it's like a game of Tetris, where data centers are playing five levels at once, trying to fit all the blocks (systems) into the data center. Operators are constantly trying to slot systems into place, often without full visibility of available capacity. As a result, the facility risks falling short of its intended performance, while inefficiencies quietly rack up costly consequences.

Currently, stranded capacity is reported as a constraint by 29% of data center operators. In addition, 60% of data center leaders say overprovisioning—allocating more resources than necessary—still occurs within their data centers. This is understandable, as they're seeking reliability, especially during peak periods. However, it wastes energy, driving up the data center's footprint and operational costs.

Moreover, high-density servers, which are designed to deliver exceptional performance, can become part of the overprovisioning issue if not managed correctly. Therefore, it's crucial to ensure these systems are running efficiently, minimizing stranded capacity and avoiding over-provisioning wherever possible.

Hailed as a game-changer for data center operations, digital twins play a critical role in tackling stranded capacity and overprovisioning, ensuring resources are allocated correctly. By providing a real-time, virtual model of the data center's infrastructure and operations, they allow operators to visualize power and space utilization with pinpoint accuracy, identifying underused systems and imbalances in workload distribution. When enhanced with AI, digital twins can go even further, using predictive insights to test how to scale operations without disruption to surrounding resources.

Keeping cool

Whilst stranded capacity and overprovisioning contribute to widespread inefficiencies, cooling is another area of potential concern. Here, data center operators should rethink legacy practices. This is where digital twins continue to come into their own. By creating simulations of the physical infrastructure with digital twins and AI, operators gain a powerful tool when planning for new and improved cooling systems. With the immense amount of energy and heat generated by AI-intensive operations, traditional air cooling is being pushed to its limits. Operators are now turning to liquid cooling, which can manage higher thermal loads more efficiently.

Whilst stranded capacity and overprovisioning contribute to widespread inefficiencies, cooling is another area of potential concern. Here, data center operators should rethink legacy practices

Currently, 45% of data center decision-makers have adopted liquid cooling, with another 19% planning to introduce it within the next year. This highlights a shift driven by increasingly dense server racks and AI workloads that exceed the thresholds of conventional systems. While air cooling handles loads up to around 20kW per rack, hybrid approaches that incorporate liquid cooling are far better suited for racks exceeding 25kW.

Similarly to visualizing power, digital twins enable operators to simulate these systems in advance. As a result, they offer visibility into variables like cooling performance, airflow, and heat recovery potential that would otherwise be a challenge. This helps identify the most effective approach for each facility's specific environment, reducing guesswork and risk. What's more, operators can test multiple cooling scenarios, compare performance, and make informed investment decisions all before physical implementation.

Beyond this, digital twins support innovations like waste heat recovery, further contributing to sustainable operations. In this way, they're not just tools for optimization, they're enablers of more responsible innovation.

The catalyst for change

Data centers are making clear strides toward operational improvement, but turning ambition into action is often difficult, as it involves highlighting existing areas of inefficiency. AI-powered digital twins are emerging as the industry's breakthrough solution here, and 73% of operators already agree. A good starting point is to use digital twins to identify and address stranded capacity and overprovisioning, optimize existing air cooling systems, and test the impact of implementing liquid cooling. By unlocking greater visibility and control, operators can modernize and maximize the efficiency of existing infrastructure before investing in additional capacity.

Ultimately, in an industry undergoing scrutiny and accelerating demand, digital twins offer a smarter, more adaptive way forward. As data centers navigate the realities of the AI era, those that embrace this kind of intelligent, responsible innovation will be best positioned to lead the way.

The visibility gap that could undermine data centre growth

Meeting the demands of AI-driven computing will require more than additional power. It will depend on better precision and deeper insight across every layer of infrastructure.

BY RAMI JEBARA, CTO AND CO-FOUNDER AT HYPERVIEW

THE DATA CENTRE industry is bracing for an energy reckoning. By 2030, electricity demand from global data centres is expected to grow by more than 165%, according to Goldman Sachs. In the US alone, the Electric Power Research Institute projects that data centres could consume nearly 9% of national electricity generation within five years, compared to just 4% in 2023.

Artificial intelligence is now the primary force reshaping infrastructure demands. Training and inference workloads are driving increases in compute density, straining cooling systems, and pushing power consumption to levels few data centres were designed to handle. Hyperscalers are re-engineering entire facilities to keep up, while everyone else is racing to avoid being left behind.

As the industry focuses on expansion, a deeper and less visible problem is emerging. Operators lack real-time insight into what is happening inside these environments, and that blind spot is growing faster than capacity itself.

Legacy systems weren't built for AI-scale complexity

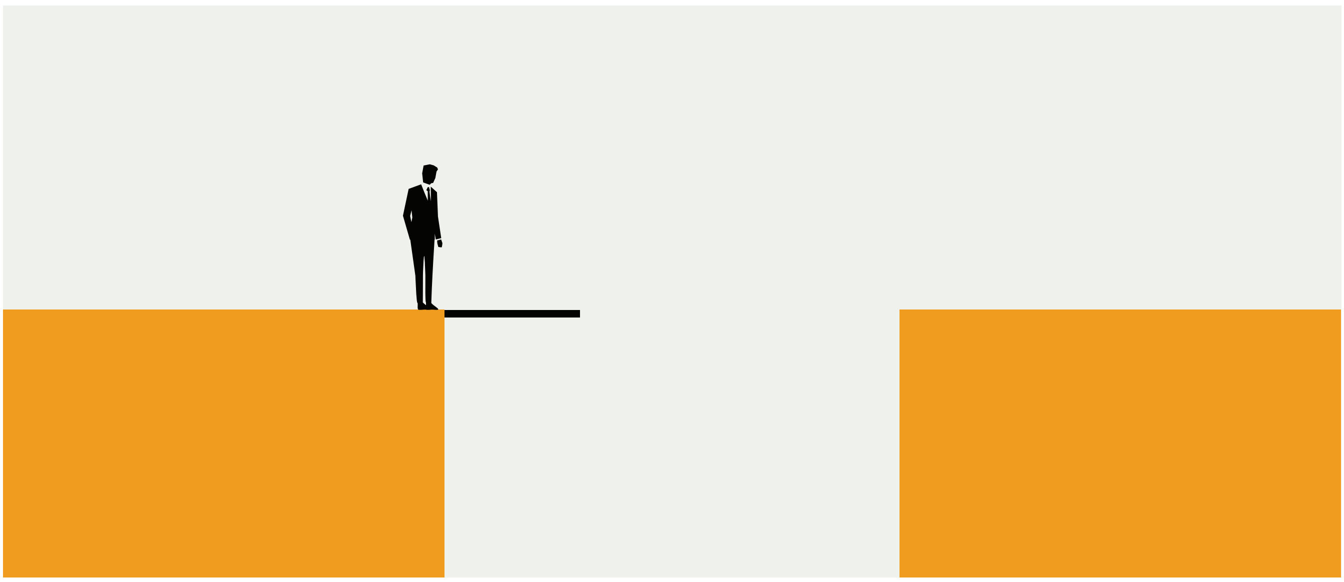
The behaviour of modern data centres has diverged significantly from the environments that legacy Data Centre Infrastructure Management (DCIM) platforms were originally designed to support. AI-driven workloads and modern applications are increasingly decentralised, operate at much higher intensity, and place unpredictable demands on power and cooling infrastructure. As edge and cloud deployments expand, the scope of operational oversight continues to shift and fragment.

Despite this evolution, many operators continue to use on-premise DCIM systems developed for a more static, centrally managed infrastructure model. These platforms often struggle to deliver the speed, interoperability, and forecasting capabilities required to manage today's hybrid estates effectively.

Their limitations in providing comprehensive visibility, adapting to dynamic workload patterns, and delivering timely operational insight are increasingly at odds with the performance, sustainability, and compliance standards data centres must meet. What may have once been considered a technical limitation now represents a far-reaching operational risk.

The cost of operational guesswork is rising

With each additional kilowatt of AI compute, the pressure on infrastructure



intensifies. Although rack densities are increasing, the Uptime Institute's 2024 survey reports that the average remains below 8 kilowatts per rack. Cabinets exceeding 30 kilowatts are still uncommon across most facilities. Liquid and immersion cooling are becoming essential for higher intensity deployments, but they introduce new environmental and operational complexities. In these conditions, managing infrastructure through static spreadsheets or delayed dashboards is no longer sustainable.

Operators need the ability to monitor, model, and predict performance in real time. This includes clear visibility into where power is being wasted, where cooling systems are overcompensating, and where capacity is underutilised. Without this insight, even well-resourced teams are forced to make costly decisions based on incomplete data.

Overbuilding or overcooling as a safeguard remains a common practice. However, this approach drives up capital expenditure, increases energy consumption, raises carbon emissions, and undermines opportunities to optimise.

Sustainability and compliance demand greater visibility

A lack of precision in infrastructure monitoring is making it more difficult for organisations to meet increasingly strict regulatory and sustainability requirements. Across Europe and North America, data centres must now comply with detailed reporting mandates covering energy use, emissions, and resource efficiency. ESG transparency is now recognised as a core performance measure, not a secondary concern.

Compliance depends on accurate measurement. Without systems that benchmark performance and generate auditable records, data centre teams struggle to keep pace with rising expectations.

In this context, continuous visibility into operational environments has become a valuable asset. With access to current, actionable insights, teams can report more accurately, strengthen internal controls, and improve accountability across the organisation.

AI at the edge demands smarter operations at the core

AI-driven decentralisation is leading operators to manage an increasing number of smaller, distributed facilities, each facing its own cooling, power, and network challenges. Attempting to oversee these environments individually is no longer practical, which is why centralised, cloud-based DCIM platforms with built-in intelligence are quickly becoming standard practice.

Today's infrastructure management tools must go beyond basic asset tracking. They need to anticipate problems, adapt to changing conditions, and support long-term operational resilience under growing pressure.

Capabilities such as anomaly detection, risk modelling, live heat mapping, and capacity forecasting enable operators to transform infrastructure data into meaningful, real-time insights. With broader visibility and stronger control, teams are better positioned to make faster, more informed decisions across complex environments.

A new baseline for modernisation

The industry's default response to rising pressure has often been physical expansion. However, in an environment shaped by energy constraints, supply chain challenges, and increasing scrutiny around sustainability, this approach is becoming less viable.

Operators positioned to thrive in this changing landscape are those who can maintain visibility, act with speed, and make efficient scaling decisions grounded in real operational insight. Achieving this begins with modernising the less visible layers of infrastructure management.

AI continues to drive significant innovation across the sector and is also revealing critical gaps in how infrastructure is monitored, managed, and planned. The challenge is not a lack of investment in new capacity, but a lack of control over how existing resources are operated.

Meeting the demands of AI-driven computing will require more than additional power. It will depend on better precision and deeper insight across every layer of infrastructure.


ZincFive
BC 2 AI
Designed for AI. No Tradeoffs in Safety, Space, or Sustainability.
LEARN MORE
zincfive.com

AI Data Centers need immediate power. Nickel-Zinc delivers.

AI workloads generate AI Dynamic Power profiles that legacy batteries were never designed to manage. ZincFive's BC 2 AI Battery Cabinet brings nickel-zinc chemistry to the UPS layer to absorb rapid load changes and protect the grid.

BY BRANDON SMITH, VP OF GLOBAL SALES AND PRODUCT, ZINCFIVE

SINCE THE INCEPTION of data centers, backup power infrastructure operated on a quiet, unquestioned premise: keep the lights on when the grid fails. Batteries sat in the background; Uninterruptible Power Supply (UPS) systems stood watch. The workload was steady. The system was predictable. In that model, batteries functioned as passive reserves.

Artificial Intelligence (AI) changed that assumption entirely.

AI requires something different: Immediate Power Solutions (IPS) — systems that actively absorb and release power in real time, not just during outages.

A GPU cluster running an inference workload doesn't draw power like a row of web servers; it spikes. Power demand fluctuates from an idle baseline to peak in milliseconds, then drops and surges again. In some cases, clusters reach up to 15 times their idle power level in a fraction of a second, cycling repeatedly during a single processing task — a behavior now commonly described as AI Dynamic Power.
The instability doesn't stay local. When those surges hit UPS infrastructure that wasn't designed to absorb them, the impact travels outward: degraded power quality, stressed grid connections, reduced compute efficiency. The UPS, long treated as an afterthought, has become a critical point of failure in the AI data center.

A chemistry problem, not a capacity problem

The reflex answer is more batteries. More capacity. More backup. But AI Dynamic Power is not a capacity problem. It is a chemistry problem — an immediate power problem.

The challenge is the ability to absorb and inject high power in milliseconds, repeatedly, without degradation, excess heat, or added infrastructure burden. What the AI data center needs is a battery that can intercept extreme bursts of energy, release them just as quickly, sustain millions of high-intensity cycles, and do it all without introducing new safety or operational risk.

Lead-acid batteries, still common in older deployments, were built for slow, steady discharges. They degrade fast under rapid cycling and fall well short of what AI loads demand.

Lithium-ion offers faster response and greater energy density, but its chemistry has a known vulnerability: repeated high-frequency surges accelerate thermal stress, drive faster wear, and introduce thermal runaway risk. Managing that risk requires overbuilding battery capacity to dilute the power and thermal load, while still demanding fire suppression systems, specialized room design, and heightened insurance exposure.

Supercapacitors react in an instant but store so little energy they can't support anything beyond a brief transient.

Each of these chemistries solves only one piece of the puzzle.

Why Nickel-Zinc solves what other chemistries don't

ZincFive's nickel-zinc (NiZn) battery technology was designed for exactly the demands AI imposes. NiZn delivers up to three times the power density of legacy chemistries at half the footprint and one-third the weight. That combination matters: high-intensity surges get intercepted at the UPS level before they propagate through the system, and operators get that capability without sacrificing rack space or floor area.

The safety profile is inherent to the chemistry itself. NiZn operates across a wide temperature range with no thermal runaway risk, even if cooling systems fail. That eliminates the suppression

infrastructure that lithium-ion requires. On the sustainability side, NiZn produces 25 to 50% lower lifecycle greenhouse gas emissions than lead-acid or lithium-ion, uses materials that are highly recyclable, and lasts three times longer than lead-acid under real-world cycling conditions. Fewer replacements mean less waste and less downtime.

Those numbers have started to matter in procurement decisions. The 2025 Data Center Energy Storage Industry Insights Report, done by Data Center Frontier in partnership with ZincFive, found that 87 percent of operators now rank sustainability as a top factor when selecting power systems, and 72 percent report meaningful cost reductions tied to their sustainability investments. Performance and responsibility no longer compete with each other.

BC 2 AI: Built for the AI power profile

ZincFive built its new BC 2 AI UPS Battery Cabinet on the foundation of its proven BC Series platform, with a specific design goal: address the AI Dynamic Power problem from the inside out.

As an Immediate Power Solution, BC 2 AI is engineered not just to back up the load, but to actively shape it, smoothing volatility in real time while maintaining traditional UPS protection.

At the center of the system is an advanced Battery Management System (BMS) capable of millisecond response to transient surges, with patented sync charge and discharge capability and real-time pulse response built into an intelligent monitoring interface. The 90Ah NiZn battery at its core supports millions of high-intensity cycles and carries a 10-year warranty. Eliminating thermal runaway risk also removes the need to overbuild battery capacity to manage thermal load, a hidden cost that consistently inflates lithium-ion deployments.

BC 2 AI is dual-use by design, supporting AI dynamic power demands while delivering reliable backup for conventional IT workloads in the same compact cabinet. Operators deploy one system — not separate solutions for AI and traditional infrastructure.

Backward and forward compatible with megawatt-class UPS systems, BC 2 AI integrates seamlessly into existing data center infrastructure originally designed for VRLA batteries. That brownfield-friendly architecture reflects where the market stands: most operators need AI-ready power that fits within their existing infrastructure, without a mandate to rebuild from scratch.

Power that keeps pace

AI is not just changing data center workloads. It is rewriting what every supporting system needs to do. Power infrastructure that once sat passively in the background now has to actively stabilize operations, absorb unpredictable surges, and meet tightening regulatory requirements around carbon footprint and battery lifecycle management.

BC 2 AI was built for that shift. It's fast enough to catch millisecond spikes, sustainable enough to satisfy ESG benchmarks, compact enough to fit where operators need it, and compatible enough to work within the infrastructure they already own.

The data centers that keep pace with AI won't be the ones with the most power on tap. They will be the ones that manage it with the right chemistry. And increasingly, that means Immediate Power Solutions — high-density, fast-acting systems designed to respond at the speed AI demands. The ability to right-size those systems, free from the overbuilding that thermal risk demands of lithium, is what makes that scale practical.

Neoclouds and the physical constraints of AI infrastructure

AI infrastructure is evolving along multiple trajectories. Emerging, specialized AI cloud operators (Neoclouds) continue to influence the deployment and distribution of compute capacity, while hardware innovation pushes the limits of fiber density, power efficiency, and optical networking. This article examines the rise of Neocloud providers alongside the infrastructure advances enabling scalable AI systems. Further reading: For more information on hyperscale market shifts, read AI, Neoclouds, and the New Limits of Data Centers and Inside the AI Rack: Power, Fabric, and Fiber.

Neoclouds: Specialist operators building AI factories

NEOCLOUDS DELIVER accelerated compute capacity for training and inference as a service. These specialists operate more like AI utilities than general-purpose cloud warehouses, with economic viability sustained by high utilization. Neoclouds include:

- **Accelerator-First Clouds:** Focused platforms built around large GPU fleets and standardized pod architecture. Differentiation centers on factors such as operational maturity, stable network fabrics, and rapid deployment cycles.
- **Power-Advantaged Operators:** Prioritizing electrical capacity, with power availability and distance from network aggregation hubs influencing site topology, network architecture, and reliance on coherent DCI.
- **Converted-Infrastructure Entrants:** Originating from land acquisition, permitting, substation, or cooling expertise, and now transitioning into AI operations with varied technical maturity (constrained by technical complexity challenges).

Financial risk and engineering implications

Neocloud operators face GPU pricing volatility, elevated exposure, and utilization variability, with market slowdowns introducing risks such as paused construction, contract renegotiation, and consolidation.

Engineering strategies must therefore align with these conditions, favoring modular architectures, standardized interfaces, and block-scalable designs that limit exposure. By contrast, bespoke and tightly optimized layouts built on assumptions of continuous capital availability introduce systemic fragility. Within this context, fiber infrastructure becomes part of the uptime surface area.

Structural differences between neoclouds and hyperscalers

While the hyperscalers can negotiate supply at massive scale, amortize engineering investment across diverse services, and absorb a degree of overbuild, Neocloud operators function within narrower economic and operational margins.

- **Resources and Engineering Depth:** Hyperscalers engineer redundancy across large teams and mature processes, while Neoclouds often concentrate operational knowledge in small groups or individuals.

- **Supply Chain Posture:** Hyperscale platforms impact component roadmaps and supply allocation, while Neocloud operators procure hardware earlier in lifecycle stages and in smaller volumes.
- **Operating Economics:** Sustained GPU utilization remains a critical economic driver. In models where margins depend on utilization, downtime generates direct pricing pressure, emphasizing operational reliability.

Emerging Technologies: Density, power Limits, and the fiber roadmap

A range of emerging technologies continues to guide the next phase of AI infrastructure. Accelerators and switching platforms attract significant attention. However, critical advancements increasingly occur deeper within the system fabric, particularly in interconnect density and power delivery efficiency.

Progress across these domains significantly influences practical scaling limits, while shifts in physical layer design enhance efficiency and inspire next-generation topologies.

Smaller interfaces and higher-density connectivity

Next-generation hyperscale data center fabrics (400G and 800G toward 1.6T), increase emphasis on physical density alongside raw bandwidth. Ongoing reductions in pluggable optics size, combined with more compact electrical and optical interfaces, enable higher port density per switch face and greater termination density per rack unit.

Connector ecosystems continue to advance, supporting dense patching through high-fiber-count MPO interfaces and compact multifiber connectors such as MMC, without expanding patch panel footprint.

Operational constraints drive this transition. Rack-scale AI deployments operate within tightly constrained mechanical envelopes shaped by liquid cooling systems, power distribution infrastructure, and service access requirements.

Limited physical space for connectivity increases demand for architectures capable of delivering higher bandwidth within reduced footprints. Under these conditions, structured cabling, pre-terminated assemblies, and disciplined port allocation and breakout strategies transition from optional optimizations to core architectural requirements, ensuring serviceability, scalability, and reliability within dense, high-performance environments.

A range of emerging technologies continues to guide the next phase of AI infrastructure. Accelerators and switching platforms attract significant attention. However, critical advancements increasingly occur deeper within the system fabric, particularly in interconnect density and power delivery efficiency

A broader AI and optical supplier landscape

NVIDIA continues to play a significant role in AI platform design. However, the broader hyperscale ecosystem now includes alternative accelerator platforms, additional vendors, and emerging optical component suppliers.

Advancements in scale-out fabrics and interconnect technologies introduce both new opportunities and increased infrastructure complexity. Growing

competition drives innovation across optics, switch design, and power efficiency, but also expands variability in form factors, lane mapping schemes, and interface standards. In response, connectivity teams abstract inter-pod and DCI layers while optimizing last-mile equipment connections to align with evolving hardware profiles.

Resulting infrastructure models segment environments into fixed, vendor-agnostic trunk layers and flexible, generation-specific connection layers. Such segmentation enables fiber plants, cable pathways, and termination points to support multiple generations of optical technology without repeated physical redesign or rework.

Power and cooling as the primary constraints

Power delivery and thermal management efficiency constrain hyperscale data center growth. Incremental improvements in link efficiency, port density, and thermal performance deliver disproportionate value as rack power levels approach 100 kW and pod-scale deployments extend into multi-megawatt ranges.

Optical technologies that reduce electrical consumption or limit electrical reach directly influence achievable fiber density and long-term operational stability. These dynamics move optical interfaces closer to compute and accelerate adoption of Linear Pluggable Optics (LPO) and Co-Packaged Optics (CPO), enabling higher capacity without increases in space, power consumption, or operational complexity.

Multicore fiber as a potential disruptor

Within the fiber technology landscape, Multicore Fiber (MCF) represents a disruptive option positioned for medium-term adoption. In contrast to traditional single-core fiber, MCF incorporates multiple independent cores within a single cladding, with each core transmitting a separate optical signal, allowing capacity expansion without increasing cable diameter.

AI campuses and pod architectures with extreme density requirements benefit from reduced fiber counts between distribution points, lowering congestion within trays, conduits, and duct systems. This means higher-capacity cabling becomes more practical compared to continuous pathway expansion.

Initial adoption remains selective as testing standards, connectorization approaches, and repair methodologies continue to mature. Even within these constraints, MCF supports architectural expansion without increasing cable volume, while alignment with pluggable optics and CPO designs featuring native multicore support may accelerate broader adoption.

Hollow Core Fiber: Long-term potential

Hollow Core Fiber (HCF) offers reduced nonlinear effects for high-quality, low-latency performance advantages. While these features suit hyperscale data centers supporting advanced AI and cloud workloads, factors such as high manufacturing complexity, elevated costs, challenging splicing requirements, and constrained supply limit HCF's near-term adoption.

Anchoring the Stack

Across advances in accelerators, switches, and optics, a consistent architectural principle persists: active technologies evolve rapidly, while the physical fiber layer must remain compatible across multiple technology cycles. Infrastructure designs emphasizing scalable pathways, disciplined physical structure, and forward-compatible fiber selections achieve greater longevity than designs optimized solely for current-generation components.

Emerging technology progress is cumulative rather than singular. Incremental gains in density, efficiency, and structural simplicity define realistic future scalability, with backbone fiber infrastructure serving as a primary determinant of achievable growth.

Conclusion

A consistent pattern emerges across market structure, architectural design, and physical infrastructure. AI growth increasingly reflects practical constraints (density, power availability, and fiber capacity) rather than theoretical performance limits. The next decade of hyperscale expansion will hinge on these realities. As operators diversify and architectures evolve, the physical layer will remain the anchor determining realistic scalability.

AFL

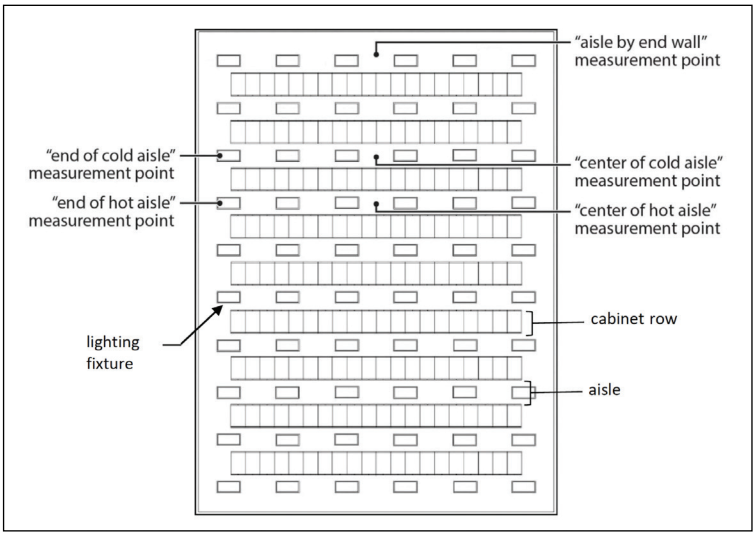
Move into the light

White server cabinets can materially improve data centre lighting efficiency, cutting fixture counts and reducing lighting energy costs by up to 25 percent while improving aisle visibility

BY MICHAEL AKINLA, PANDUIT

FOR YEARS, black server cabinets have been treated as the default choice in data centres, comms rooms and technical environments. They look familiar, align with the visual language of enterprise IT, and rarely attract much debate. Yet the question of cabinet colour is more consequential than it first appears. In high-density data centres and intelligent buildings alike, lighter cabinet finishes can influence lighting performance, visibility, working conditions and energy use in ways that deserve more serious attention.

➤ Figure 1 – Illuminance Measurement Points



The strongest case for white cabinets is not aesthetic. It is practical. Dark cabinet surfaces absorb a large proportion of the light falling on them, while white or lighter surfaces reflect more of that light back into the room. In a space defined by long rows of cabinets, narrow aisles and repeated vertical surfaces, that difference changes the lighting environment materially. A lighter cabinet finish can raise overall illuminance, reduce the number of fittings needed to achieve a target lux level and lower the continuous lighting power load.

That is the core point that should frame the discussion. White cabinets are not a shortcut to better cooling simply because they are white. Their most defensible benefit is that they improve the efficiency of the lighting environment around the IT estate. In a modelled 5,000 sq ft (465m²) data centre, Panduit researchers found, switching from black cabinets to white increased illuminance across

Scenario	Lighting load with dark cabinets	Lighting load with white cabinets	Reported lighting saving	Fixture reduction
Baseline hot aisle/cold aisle	5,120W	3,840W	25%	20
Cold aisle containment	7,168W	5,376W	25%	20
Vertical exhaust duct configuration	5,120W	3,328W	35%	28

all measured areas. In the centres of the hot and cold aisles, illuminance rose by 56 percent, while the remaining monitored areas saw a 32 percent increase. Those are not cosmetic differences; they are significant enough to affect both lighting design and day-to-day operational visibility.

The practical effect of that increased reflectance is straightforward. If the same target lighting level can be achieved with fewer luminaires, then both installation cost and ongoing electrical load can be reduced. In the same modelled scenarios, white cabinets allowed lighting layouts to be reduced while maintaining a minimum target of 300 lux. In a standard hot-aisle/cold-aisle configuration, lighting electrical load fell from 5,120 W to 3,840 W, a reduction that translated into a 25 percent lighting cost saving. In a cold-aisle containment configuration, the load fell from 7,168 W to 5,376 W, again equivalent to a 25 percent saving. In a vertical exhaust duct configuration, the load dropped from 5,120 W to 3,328 W, producing a 35 percent saving.

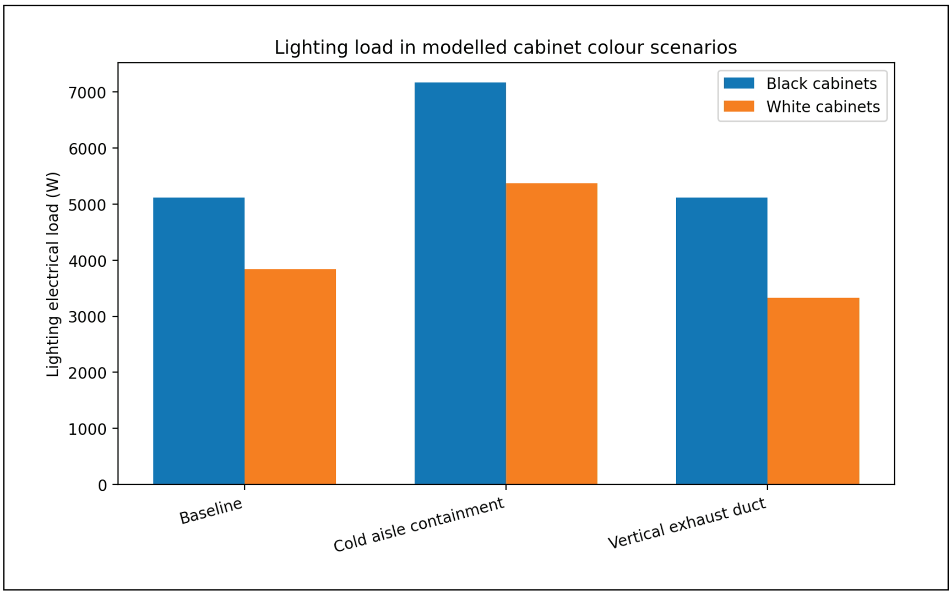
There was also an installation benefit. The white-cabinet scenarios required fewer fixtures to meet the same illuminance threshold: 20 fewer in the baseline and cold-aisle-containment cases, and 28 fewer in the vertical exhaust duct case. Using the cost assumptions in the study, that equated to fixture savings of many thousands of euros or pounds respectively. This matters because it shows that white cabinets can contribute to both lower CapEx and lower OpEx when the lighting design is planned around their higher reflectance instead of being left unchanged.

Containment adds another layer to the story. White cabinets still outperform black cabinets when containment is introduced, but containment itself changes how light behaves in the room. In cold-aisle containment, the centre of the cold aisle became 44 to 48 percent darker than in the baseline scenario, regardless of cabinet colour. Even so, white cabinets still produced 48 percent higher illuminance in the contained cold aisle than black cabinets. In vertical exhaust duct arrangements, the effect was especially pronounced in the hot aisle, where white cabinets and ducts delivered around 182 percent higher illuminance than black cabinets with ducts.

That is a valuable insight for designers. White cabinets can improve lighting performance in contained and chimneyed environments, but they do not eliminate the need for careful lighting layout. In fact, the study makes clear that optimisation is more complex than simply removing a number of fittings. Fixture spacing, aisle length, room reflectance and the practical constraint of fitting whole numbers of luminaires all affect what is achievable in the real world. The right lesson is not "white solves everything", but "white improves the design envelope".

There is also an operational case for lighter cabinets that goes beyond energy. Better reflected light improves the visual conditions in working aisles. It makes cabinets easier to inspect, labels easier to read and patching areas easier to navigate. In dense rooms where technicians are tracing ports or checking terminations, that extra brightness can reduce the gloomy, low-contrast effect that often comes with rows of black infrastructure. This benefit

➤ Figure 2 – Energy Load Reduction



is especially relevant in intelligent buildings, where technical spaces are often smaller, less well-lit and used intermittently rather than continuously. White cabinets in telecoms rooms, BMS hubs, security spaces and edge environments can make modest lighting schemes more usable without increasing general lighting demand.

It is important not to overstate the effect inside the cabinet itself. One of the more useful findings in the study is that the internal visibility gain is highly dependent on whether the room lighting is reduced to capture the energy saving. If the room lighting is left unchanged, the rear of a white cabinet can be around 80 percent brighter than the rear of a black one. However, if the room lighting is reduced to exploit the higher reflectivity and save energy, the improvement inside the cabinet falls to around 16 percent. The practical conclusion is clear: white cabinets improve ambient conditions, but they do not remove the need for local task lighting when detailed work is being carried out inside the enclosure.

That makes the best strategy a combined one. The reflected light from white cabinets should be used to lower the baseline ambient lighting requirement, while dedicated task lighting is reserved for the point of intervention. In practical terms, that means designing lighter cabinet environments alongside door-triggered or occupancy-triggered lighting, rather than relying on permanently high-output room lighting. This is especially compelling in intelligent buildings and edge environments, where technical rooms are often occupied only briefly. In such cases, the combination of reflective cabinet surfaces and smart controls can help avoid over-lighting the whole space simply to make occasional maintenance easier.

There is, however, one further caution. A room that relies too heavily on white or near-white surfaces can create glare and poor contrast, which may become uncomfortable in frequently occupied environments. The answer is balance. White cabinets work best as part of a broader lighting strategy with controlled reflectance from walls, ceilings and floors, and with photometric design that considers how the space is actually used. The aim is not maximum whiteness; it is efficient and usable brightness.

The continuing case for white server cabinets is a strong one, when discussed in context of the individual site. They do not transform cooling performance by colour alone, and they should not be sold on that basis. What they do offer is more credible and more useful: higher reflectance, lower lighting load, reduced fixture count, better aisle visibility, improved support for smart lighting strategies and more usable technical environments. In a sector where energy performance and operational efficiency are under constant scrutiny, that makes white cabinets more than a stylistic alternative. It makes them a practical design choice that still deserves serious consideration.

DATACENTRE SOLUTIONS ROADSHOW

Based on insights from the Data Centre Solutions Roadshow in Newcastle, 25 February, where senior leaders explored regional data centre, energy, infrastructure trends.

From Grid to GPU: Can the North East deliver the UK's AI infrastructure future?

The Data Centre Solutions Roadshow arrived in Newcastle at a pivotal moment for the industry. AI is no longer an emerging trend. It is a force that is reshaping infrastructure demand, investment strategy and regional development.

The conversation throughout the day made one thing clear. The question is no longer whether the UK can support AI infrastructure at scale. It is whether it can deliver it fast enough.

Across the Fireside and four panel discussions, a consistent narrative emerged. The North East has the ingredients to become a leading AI infrastructure region. Energy, land, connectivity and skills are all present. Investment is already underway.

Yet delivery remains constrained.

What followed was a structured examination of the problem. From regional positioning and power constraints to design complexity and execution challenges, each session built on the last.

The result is a clear picture of both opportunity and risk.

A region ready to lead

The Fireside session set the strategic context.

The North East is being positioned as one of the UK's most credible locations for large-scale AI infrastructure. This is not speculative. It is already in motion.

The £10 billion QTS development in Northumberland stands as a defining example. A hyperscale campus of significant scale, backed by Blackstone, it represents one of the largest inward investments in UK data centre infrastructure.

Its location is no accident. Access to energy is the primary driver. Offshore wind, interconnectors and legacy infrastructure combine to create a level of power availability that is increasingly rare.

Alongside this sits a coordinated regional ecosystem. Universities are producing talent in AI and engineering. Industrial sectors such as defence, healthcare and advanced manufacturing are generating demand. Local authorities are aligned around growth.

The concept of the AI Growth Zone reflects this convergence.

The ambition is not simply to host data centres. It is to build a regional economy around them.

Demand is real, but not yet deployed

The Fireside discussion also highlighted a critical tension.

Demand exists. Defence organisations, healthcare systems and government departments are preparing to deploy AI workloads. New infrastructure providers are entering the market, offering GPU-based compute services at speed.

Yet this demand is not translating into deployment.
The reason is consistent throughout the day. Infrastructure is not yet ready at the scale required.

This creates a structural bottleneck. Supply depends on demand certainty. Demand depends on operational infrastructure. Both are waiting.

Resilience redefined

Panel 1 shifted the focus from regional strategy to infrastructure reality.

The concept of resilience was fundamentally reframed.

Historically, resilience has been defined by redundancy. Backup systems, failover design and uptime metrics have been the foundation of data centre engineering.

That model is no longer sufficient.

Resilience is now defined by access. Specifically, access to power.

For large-scale AI deployments, the primary challenge is not how systems recover from failure. It is whether they can be powered and scaled in the first place.

This marks a significant shift in thinking.

Power as the first constraint

The panel made it clear that power has moved from an operational consideration to a strategic determinant.

AI workloads require levels of energy that exceed traditional planning assumptions. High-density environments, rapid load fluctuations and continuous scaling requirements place unprecedented demand on infrastructure.

Microgrids are emerging as one response. By securing power upfront and distributing it across a campus, operators can reduce dependence on incremental grid upgrades.

However, this approach introduces new complexity. It requires significant investment, careful integration and a redefinition of the relationship between operators and utilities.

At the same time, grid dependency remains a constraint. Connection timelines and infrastructure upgrades cannot keep pace with demand.

Connectivity and workload behaviour

Panel 1 also reinforced the importance of connectivity.

The North East's growing network of cable landing stations enables direct international links, reducing dependence on traditional hubs. This strengthens the region's case as a location for hyperscale and AI deployments.

Alongside this, workload behaviour is changing.

AI introduces volatility. Power demand can move from low to high in milliseconds. Cooling systems must respond instantly. Thermal management becomes a dynamic challenge rather than a steady-state problem.

This drives the need for new approaches to both power buffering and cooling design.

Energy becomes the defining variable

Panel 2 brought the discussion back to energy, but with a sharper focus on scale, cost and delivery.

The level of demand reaching utilities is unprecedented. Requests for capacity linked to data centres alone are exceeding historic peak demand.

This is not incremental growth. It is a step change.

At the same time, infrastructure delivery remains constrained by physical realities. Grid upgrades, transformer availability and network expansion require long lead times.

The result is a growing gap between demand and supply.

Location is now determined by energy

For operators, this shift is already reshaping strategy.

Energy availability is now the primary driver of site selection. Regions with access to large-scale renewable power are gaining advantage, while traditional hubs face increasing constraints.

The North East benefits from this shift. Offshore wind and energy inflows position it as a viable alternative to more established locations.

However, this advantage must be balanced against cost.

The cost challenge

Energy cost emerged as a critical issue.

While sustainability remains important, economic reality is becoming more influential. Lower-cost markets such as Norway are attracting workloads due to cheaper and more abundant renewable energy.

This creates a competitive challenge for the UK.

Operators must balance sustainability requirements with cost pressures. In a constrained environment, access to power may take precedence over ideal sourcing.

Design struggles to keep pace

Panel 2 also highlighted the challenges of infrastructure design.

The industry is still adapting to AI requirements. There is no single established model. Power distribution, cooling strategies and system integration are all evolving.

The gap between theoretical design and practical delivery remains significant.

This creates uncertainty. Projects must be delivered in an environment where standards are still forming and best practice is not yet defined.

Designing for volatility

Panel 3 moved deeper into this challenge.

The industry is no longer designing for predictable workloads. It is designing for volatility.

AI environments require constant visibility, rapid response and tight integration between systems.

Traditional management approaches are no longer sufficient.

Operational tolerance has reduced significantly. Where systems once had margin for error, AI environments operate within seconds.

While sustainability remains important, economic reality is becoming more influential. Lower-cost markets such as Norway are attracting workloads due to cheaper and more abundant renewable energy

This demands a shift toward predictive management.

Fragmentation as a constraint

One of the most important insights from Panel 3 was the continued fragmentation of infrastructure management.

Power, cooling and IT systems are still often managed separately. Data is siloed. Integration remains limited.

This fragmentation becomes a critical weakness in high-density environments.

To operate effectively, data centres must function as a single system.

The limits of physics

The discussion also highlighted physical constraints.

Cooling systems are operating closer to their limits. Liquid cooling introduces new dynamics, including rapid thermal changes and reduced inertia.

Even with advanced control systems, there are limits to how quickly infrastructure can respond.

This challenges assumptions around performance and reliability.

Speed versus certainty

A clear divide emerged between different approaches to delivery.

In the United States, the focus is on speed. Infrastructure is built quickly and refined through operation. Failure is accepted as part of the process.

In Europe, the approach is more cautious. Greater emphasis is placed on design certainty and risk reduction.

While this reduces immediate risk, it can slow deployment and limit adaptability.

This difference has implications for competitiveness.

Execution as the core challenge

Panel 4 brought the discussion to its conclusion.

The central issue is not technology. It is execution.

The delivery model for data centres remains fragmented. Multiple stakeholders are involved, each with their own priorities and processes.

This creates friction.

Projects are slowed by misalignment, delayed decision making and the need to revisit earlier stages.

The need for early collaboration

The panel emphasised the importance of early engagement.

Stakeholders must be involved from the outset. Utilities, developers, operators and contractors need to align before design decisions are fixed.

This enables more realistic planning and reduces the need for rework.

Without this, projects are forced into reactive adjustments.

Cultural change and shared responsibility

Delivering at speed also requires cultural change.

Traditional models emphasise control and separation between organisations. AI infrastructure requires collaboration and shared ownership.

Breaking down silos is essential.

This includes greater openness, trust and willingness to adapt as projects evolve.

Balancing speed and risk

The panel also addressed the tension between speed and risk.

Faster delivery introduces uncertainty. Design may be less refined. Operational risks may increase.

The challenge is to manage these trade-offs effectively.

Clear decision making and shared understanding are critical.

From strategy to action

The final message of the day was clear.

The industry understands the challenge. It has the technology, the investment and the demand.

What it lacks is coordinated execution.

The North East has the potential to lead. It has energy, land, connectivity and a collaborative ecosystem.

The question is whether these elements can be aligned into a functioning delivery model.

Conclusion: The infrastructure test case

The Newcastle Roadshow revealed more than a regional opportunity. It highlighted a national challenge.

The UK is at a point where ambition must translate into delivery.

AI infrastructure is not limited by innovation. It is limited by the ability to connect systems, align stakeholders and execute at scale.

The North East represents a test case.

If it can bring together power, infrastructure and demand, it can demonstrate how the UK can compete in the global AI economy.

If not, the risk is clear. Opportunity will remain just out of reach, and other regions will move ahead.

Final thought

The industry is no longer asking what is possible.

It is asking what can be built.

The answer will define the next decade of digital infrastructure.

DCA INSIGHTS

Steve Hone
CEO,The DCA
(Data Centre Alliance)

Data Centre World 2026 – What an event!

The DCA team rated this year as CloserStill Media's best ever show!
THE DCA had a huge amount of involvement again this year – placing speakers on panels, hosting the Keynote Theatre, judging Awards and running our own panel and debate sessions! Thank to everyone that supported these endeavours and ensured they were a success. I'd like also thank Data Centre Solutions Magazine for their support filming interviews with DCA Partners on both days of the show. The DCA stand was overwhelmed with visitors for much of the event with the informal drinks party at the end of the first day resulting in us taking over the aisles! It was great to meet up with so many DCA Members and Partners – we hope you enjoyed it and found it productive.

Advisory Council - I am delighted to announce that the DCA Data Centre Advisory Council is now officially up and running. The launch brunch at DCW was very well supported, with over 60 Council Members attending. The next steps - the Advisory Council Board chaired by Liam Round have convened and are now in the process of both mapping out and prioritising initiatives that the group will be focusing on in the coming months. I will keep you posted on progress!

Regulatory Radar - The DCA and Clear Decisions have jointly launched Regulatory Radar, the world's first dedicated regulatory intelligence platform built specifically for the data centre industry. At a time when energy reform, planning policy, sustainability regulation and AI infrastructure strategy are reshaping the operating environment, Regulatory Radar provides data centre owners, operators and infrastructure investors with a single, structured source of real-time regulatory insight across the UK and EU – all DCA Members have access to the Radar through The DCA website. Find out more here

For information and more details, please visit www.dcauk.org or follow The DCA on LinkedIn.

DCA Insights in Data Centre Solutions – ISSUE 02 2026
This DCA Insight feature is comprised of articles from DCA Partners and industry experts. Thank you to all the authors for their contributions.

- **Practical Considerations for Designing a Future-Ready AI Data Centre** - *Alastair Waite, Commscope* details how designing, implementing and managing data cabling in an AI data centre involves unique challenges due to the introduction of high-performance processor chips to the servers.

> The DCA and Clear Decisions have jointly launched Regulatory Radar, the world's first dedicated regulatory intelligence platform built specifically for the data centre industry

- **Why AI is Forcing a Shift Beyond Air Cooling- A Structural Shift, Not a Tactical Upgrade!** - *Kate Steele, Director EMEA HPC/AI, Lenovo.* Kate informs us of how rising AI data centre thermal demands are driving a structural shift from traditional air cooling to liquid cooling.
- **From Predictive to Agentic: The Next Era of Data Centre Automation** - *Jeff Safovich, CTO RiT Tech* explains how the next necessary step in data centre operations is the move from predictive AI to agentic AI.

If you'd like to find out more about The DCA and how we support the sector and those working in it, drop me an email, steveh@dcauk.org

Best regards,
Steve

KEY DC INDUSTRY EVENTS Q2 2026

- **16 APRIL –** NDCA Data Centre 'Power @ The Edge', Royal Leamington Spa
- **21 APRIL –** MEUC Buying & Using Utilities, London
- **14 MAY –** Data Centre Solutions Awards, London
- **2-4 JUNE –** Datacloud Global Congress 2026, Cannes
- **25 JUNE –** Mixing IT, London

Practical considerations for designing a future-ready AI data centre

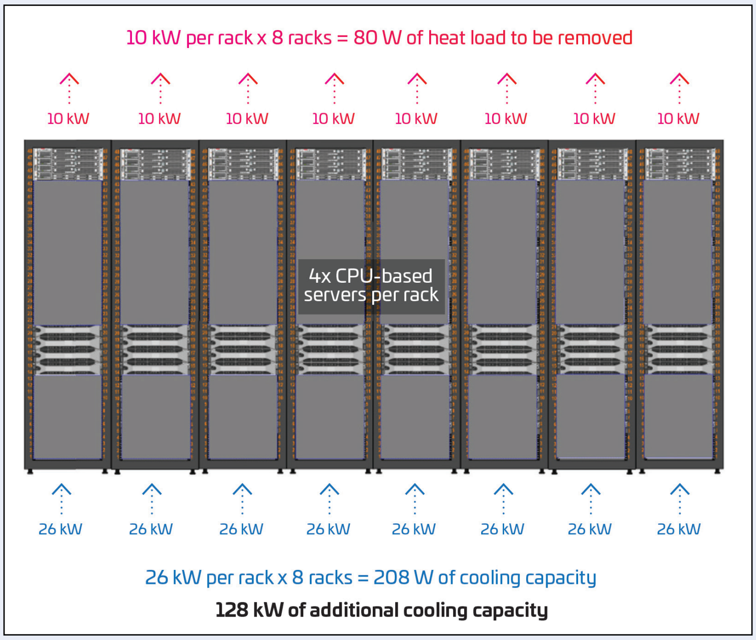
AI-driven rack densities are forcing data centre design to balance high-speed connectivity, cooling and cabling at unprecedented scale.

BY ALASTAIR WAITE, SENIOR MANAGER, EUROPEAN DATA CENTRE MARKET DEVELOPMENT

DESIGNING, implementing and managing data cabling in an AI data centre involves unique challenges due to the introduction of high-performance processor chips to the servers (i.e., Accelerator Application Specific Integrated Circuits like GPUs and TPUs). These integrated circuits process huge amounts of data in parallel, and combined with High Bandwidth Memory componentry in a Rack Scale System (RSS) can achieve 3.6 exaflops of compute performance.

Today's highest-performing AI superchips can process at speeds of up to 40 Petaflops (floating point operations per second). For comparison, traditional central processing unit (CPU) of the type commonly found in data centre servers provide theoretical peak performance around 1,200 times slower than the latest GPU.

In an AI cluster, an Accelerator Application Specific Integrated Circuit (AASIC) does not operate on its own but works in parallel with multiple other AASICs and supporting hardware like High-Bandwidth Memory (HBM), scale-up and scale-out switching hardware, power supplies and cooling infrastructure. Often, an RSS will be configured with up to 72 GPUs working together in a chassis or rack, each GPU requiring huge amounts of bandwidth via 400G, 800G and even 1.6T channels to maintain peak performance.



➤ Figure 1: 4x CPU-based servers per rack: Air-cooled capacity in a data centre using CPU-based servers. This scenario demonstrates surplus air handling capacity.

The cooling challenge

Increased performance exacerbates cooling challenges impacting the compute cabinet and its space available for fiber-optic cabling. Since the increased performance of these systems comes with higher power consumption, cooling becomes even more important; it must be evacuated from the servers and racks to keep the data centre hardware at optimum performance.

Previously, compute rack power demand of 8 to 10 kW produced heat that could be easily transferred away using chilled air to maintain an ambient temperature of around 25° C. This has been sufficient for most data centres for the last 20 years.

As higher-performing compute hardware with higher thermal densities have arrived, rack power requirements have moved beyond 10 kW and keep rising as high as 80 or 100 kW with some cloud providers forecasting rack level power loads of 0.5 MW, levels that air cooling alone cannot support. This has pushed liquid cooling to the fore. Water and other fluids can be 50 to 100 times more efficient than air cooling, promising a solution to the thermal challenges of modern AI clusters. That fluid must be pumped around the data halls, into the rows and racks, and finally into the server chassis hosting the powerful accelerators.

Here is where the physical layer cabling design matters most, to coherently connect all those accelerators in a way that doesn't impede the performance of the overall architecture. This is complicated by the introduction of additional cooling hardware (fluid piping), alongside a dramatic increase in the number of fibers required to fit into a small footprint, whether a server chassis or an entire rack.

Figures 1 and 2 illustrate the cooling challenge facing designers.

Additional DTC pipework infrastructure consumes interior rack space that was once available for structured cabling, stressing the interior space's capacity even as the number of fiber cables rises.

Key points in designing a physical layer

- **Consider additional networks**
  In traditional data centre design, we typically have the East-West network fabric cabling (e.g., server to T0 switch, and T1 to T2 switching layers), plus the out-of-band management and dedicated SAN links.However, an AI cluster deploying the new NVIDIA NVL72 RSS sees each rack typically having these scale-out fiber network links:
  - 72x compute back-end/GPU fabric
  - 36x storage + in-band front-end/CPU fabric
  - 4x out-of-band management (OOB) links,
  - Plus 8x Cat6 or Cat6a copper links for the Admin Network
  - And 20x additional fiber cabling links to support redundancy

  It's quickly apparent that the traditional cabling approach for compute cabinets used in non-AI environments is not a sustainable strategy where liquid cooling infrastructure is deployed alongside fiber cabling.

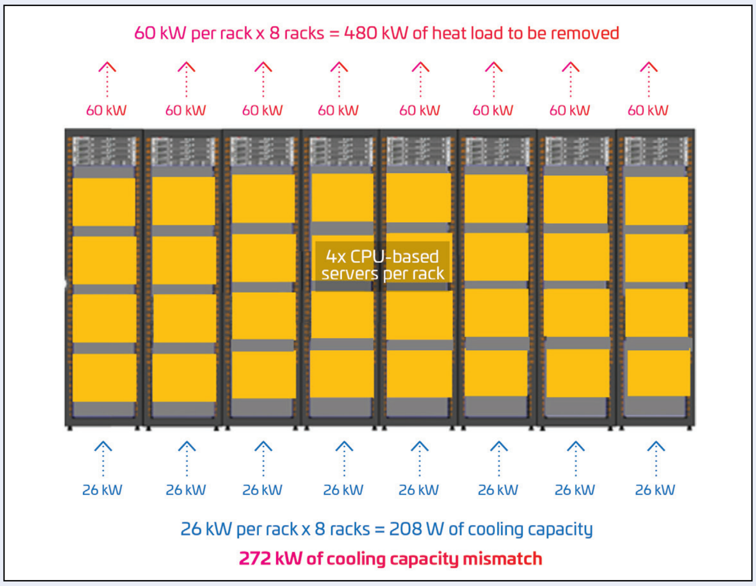
- **Overhead pathway zoning**
  Having individual zones above cabinet rows can prevent conflicts between data cabling, power feeds and the cooling liquid piping. Protecting fiber cabling across the data centre hall and observing maximum run lengths for AI-enabled servers will help maximize performance, so inclusion of a fiber raceway like the FiberGuide® system is an effective solution.

- **Cabinet selection and cable management**
  Server racks come in 600 mm and 800 mm widths. While the smaller footprint to conserves space, in the world of AI, the vertical riser space must be sufficiently large to enclose all the additional infrastructure required. Selecting a wider/deeper cabinet may be more prudent. However, this brings challenges that make I difficult for traditional rear-of-cabinet patching, because there's less space to access the rear. Designing in-cabinet patching solutions with all-front access platforms, like the Rapid Fiber Connect™ platform, can significantly improve installation time and easier day 2 moves, adds and changes (MACs).

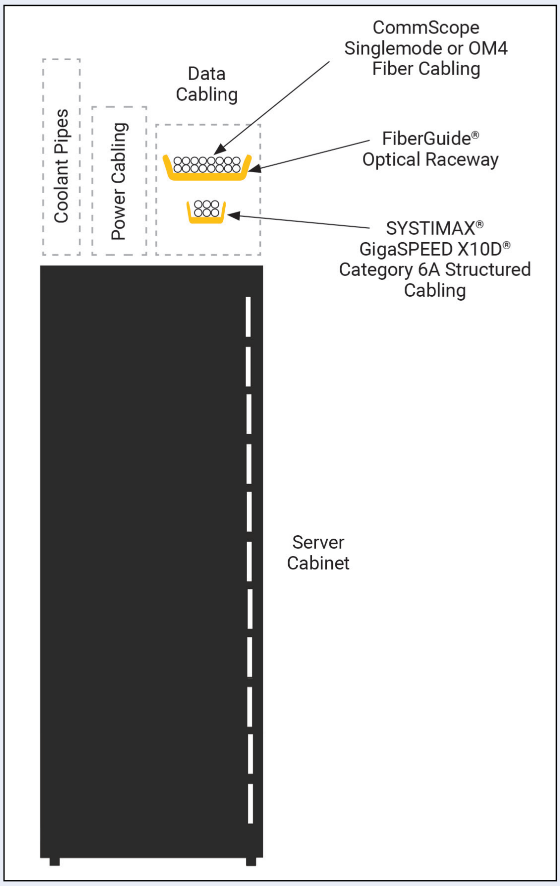
- **Maximizing GPU performance**
  Essential to AI performance is the amount of low-latency, high-bandwidth data provided to the GPUs. High-quality data cabling and optical end-face performance are considered "table stakes" today. As fabric bandwidths continue to increase, singlemode and OM4 structured cabling is essential. Many optical channels require so much bandwidth that parallel optics are needed, necessitating parallel connectivity like MPO16 and MPO12/8 supplied with APC polishing, to prevent optical reflections that impact link performance.

- **Cable organization.** Using structured cabling simplifies MACs and streamlines future scalability, especially important with the rapid pace of AI hardware evolution.

Prioritizing these areas of focus early in AI data centre builds can promote efficient, reliable and scalable data cabling solutions to support their computational demands. As GPU power needs increase to support their growing performance levels, data centres must keep a future-focused view on their buildouts.



(above) Figure 2: 4x GPU-enabled servers per rack: Air-cooling system is no longer sufficient and requires more cooling capacity, such as from liquid cooling in the form or direct-to-chip (DTC) and its additional pipework for each cabinet, rack and row.



Why AI Is forcing a shift beyond air cooling

Lenovo As AI drives extreme rack densities, air cooling is reaching its limits, forcing a strategic shift toward liquid cooling as a core enabler of performance, efficiency and future data centre scalability.

BY KATE STEELE, DIRECTOR EMEA HPC/AI, LENOVO

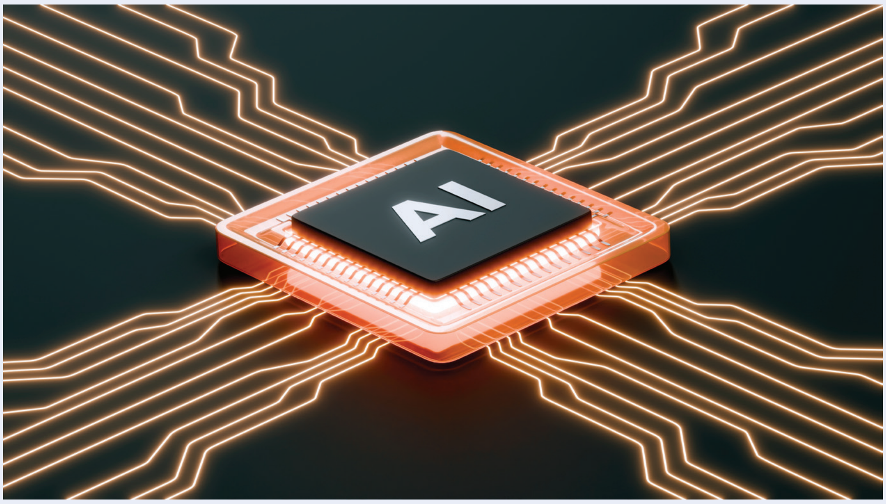
COOLING HAS moved from a facilities issue to a board-level decision with implications for capital planning, site selection and long-term competitiveness. As AI systems scale and rack densities rise, the method used to extract heat is becoming a strategic constraint. Air cooling, long the default for enterprise data centres, is approaching its physical and economic limits. In practical terms, it influences how much compute can be deployed per square metre, how quickly new capacity can be brought online and how efficiently existing sites can be upgraded.

GPU-based AI systems and high-performance computing (HPC) clusters are consuming more power per node than traditional enterprise workloads ever demanded. Operators are being asked to deliver greater compute density, faster time to insight and tighter sustainability performance, often within existing footprints, planning restrictions and grid constraints. Cooling architecture now determines whether legacy approaches can keep up.

Air cooling has reached its practical ceiling

Air-cooled systems can support up to approximately 70 kilowatts per rack in the data center, but hit a physical limit when it comes to removing heat known as ‘specific heat capacity’. Beyond this point, increasing airflow delivers diminishing returns. Operators are forced to deploy higher-powered fans, redesign containment and push more cold air through ever-denser racks, all of which increases energy overhead and operational complexity. Therefore, operators need to look at alternative cooling methods. Adding to the complexity is the rise in demand for AI workloads across every industry, requiring five times the amount of power.

The industry is now seeing a shift from chips operating at 120 watts to 600 watts or more. At that level, a small number of AI servers are generating a significantly larger amount of heat. The thermal load is no longer distributed evenly across the floor; it is concentrated in high-density clusters that strain conventional airflow design and increase the risk of thermal throttling or unplanned downtime. In high-performance environments, even marginal thermal instability can reduce system efficiency, shorten component lifespan and undermine service-level commitments.



This change is pushing air cooling beyond the threshold at which it is viable. It mirrors a previous shift in the automotive industry, where manufacturers had to adopt liquid cooling for engines as consumers demanded more power. Citroen developed the last mass-produced air-cooled engine in the 1970s, with a few holdouts such as Porsche still using air cooling into the last years of the 20th century. Air-cooled cars are now largely limited to vintage or specialist equipment, whereas modern cars use liquid cooling.

Similar lessons can be applied to the data center industry. Water is the perfect choice because it has the highest capacity to retain and move heat out of any common liquid or gas. It is more than 3,000 times more efficient at transferring heat by volume, meaning far greater amounts of thermal energy can be removed using far less space and energy input. To put that in context, if walking was 3,000 times more efficient, you could stroll from London to Rome in the time it takes to brew a cup of coffee.

Water makes an enormous difference to the efficiency of data centers, which use as much as half of their total energy on cooling systems. Liquid cooling techniques, such as direct-to-node cooling, can remove up to 95% of heat from servers by capturing it directly at the processor rather than attempting to cool the entire room. Because heat

is removed at source, the residual load on room-level air conditioning drops significantly. This heat can then be reused to heat facilities. Modern cooling systems are designed to be serviceable and safe, making them practical in a wide variety of environments.

Why liquid changes the equation

Liquid cooling is not an exclusive technology reserved for the biggest hyperscalers like Amazon, Google and Microsoft. As demand for AI grows, along with concerns over sustainable computing, we are seeing universities and Tier 2 and Tier 3 data centers (rated for reliability) adopting liquid cooling to keep pace. For those switching to liquid cooling, there is no need for a complete overhaul: the technology integrates neatly with existing infrastructure. The latest water-cooled systems work across multiple configurations, including hybrid systems using both air and liquid, which can make adoption easier.

With the energy used by data centers predicted to rise 150% in Europe in the coming decade, according to think tank Ember, measuring the efficiency of data centers in power use (or Power Usage Effectiveness) will come under scrutiny from regulators, investors and increasingly sustainability-conscious customers.

Liquid cooling is the easiest way to stay compliant while boosting performance, and warm-water cooling, which does not waste water to keep it cool, also avoids putting pressure on local water supplies. In this model, a greater proportion of energy is directed toward computation rather than cooling overhead, improving both performance and sustainability outcomes.

A structural shift, not a tactical upgrade

It is tempting to frame the move from air to liquid as a technical upgrade.

> Liquid cooling is not an exclusive technology reserved for the biggest hyperscalers like Amazon, Google and Microsoft. As demand for AI grows, along with concerns over sustainable computing, we are seeing universities and Tier 2 and Tier 3 data centers (rated for reliability) adopting liquid cooling to keep pace.

In reality, it is closer to a structural transition in how data centres are designed. Air cooling enabled decades of steady IT expansion. It was flexible, familiar and adequate for moderate densities. AI has altered the equation. When processors draw several hundred watts each and racks push beyond 70kW, cooling becomes the gating factor for performance, efficiency and sustainability.

The industry is approaching a thermal tipping point. Continuing to scale AI on air-cooled foundations will mean higher energy overheads, tighter density constraints and increasing operational complexity. Liquid cooling, by contrast, aligns with the trajectory of chip design and power demand.

For organisations investing heavily in AI, the choice is clear. Cooling strategy must be considered alongside compute strategy from the outset. Those that adapt early will gain efficiency, flexibility and headroom for future innovation. Those that delay may find that their infrastructure, not their ambition, sets the limit.

In the era of AI, competitive advantage will not only be measured in model performance or data access. It will also be determined by who can manage heat most effectively, turning it from a constraint into an opportunity.

From predictive to agentic: The next era of data center automation

As AI-driven density accelerates, data centres must move beyond human-led operations toward agentic AI, enabled by unified, physically grounded infrastructure data.

BY JEFF SAFOVICH, CTO, RIT TECH

WHEN NASA lands a rover on Mars, engineers face a fundamental limitation of physics: communication delay. Depending on planetary alignment, it can take up to 20 minutes for a signal to travel from Earth to Mars, and another 20 minutes for the response to return.

Because of this 40-minute round trip, NASA cannot manually drive a rover with a joystick. If the rover had to wait for human instructions every time it encountered an obstacle or a steep drop, it would crash. Instead, the rover must observe its environment and make decisions autonomously.

Today, the data center industry is facing its own Mars rover moment.

As we deploy massive, high-density AI clusters to support generative models and compute-intensive workloads, the physical dynamics of data center facilities are changing. The time it takes for a localized thermal issue to escalate, from a minor anomaly to a critical hardware throttling event, has shrunk from hours to seconds.

In this new high-density reality, the communication delay is no longer the network. The latency is the human operator.

When an AI training cluster suddenly spikes, the thermal reaction happens too quickly for a human-in-the-loop approval process. To protect both the infrastructure and the workload, systems can no longer wait for human permission. They must be able to detect conditions and execute safe, autonomous actions.

The evolutionary path - beyond the predictive dashboard

Over the last decade, the data center software industry has focused heavily on predictive analytics and AIOps platforms. We have built increasingly sophisticated systems that monitor server telemetry, analyze trends, and send probabilistic alerts to centralized dashboards.

These predictive tools function like advanced alarm clocks. They can detect signs such as a degrading cooling pump or a spike in CPU utilization and alert a facility manager that a thermal

While closed-loop agentic control is the long-term destination, the journey there unlocks massive, immediate value for operators managing today's enterprise and colocation environments. By establishing a UIIM foundation, organizations can deploy "assistive automation" right now to solve their most pressing capacity and operational challenges

issue is likely to occur in a specific zone. But that is where the automation ends. The system raises the alert and then waits. A human operator must read it, manually cross-reference data across siloed IT and Building Management Systems (BMS), decide what to do, and then carry out the change.

The next necessary step in data center operations is the move from predictive AI to agentic AI.

Agentic AI refers to autonomous software agents that can plan, reason, and execute complex, multi-step workflows without human intervention. Instead of simply flagging a thermal anomaly, an agentic AI system could autonomously migrate critical inference workloads to a cooler zone to prevent hardware throttling, while at the same time sending a command to the facilities side to adjust coolant flow.

However, true agentic control over critical life-support systems is not something we will hand over overnight. Giving an autonomous agent that level of authority requires a very high degree of operational trust, and building that trust will require a deliberate, step-by-step evolution.

The missing foundation - Grounding AI in physical reality

You cannot allow an autonomous system to drive if its sensors are blind, and you cannot allow Agentic AI to manage a data center if its underlying data is fragmented.

The reality is that most enterprise AIOps tools today have a major physical blind spot. They understand the logical layer - applications, server workloads, and network traffic - but they have little deterministic awareness of physical power constraints, localized cooling dynamics, or precise spatial relationships within the facility. If an AI agent only sees the software layer, it will eventually make a decision that trips a breaker or overheats a room.

Preparing for an agentic future requires grounding AI in physical reality. This demands a structural shift in how infrastructure data is managed, moving toward Universal Intelligent Infrastructure Management (UIIM).

UIIM is an engineering paradigm that breaks down the traditional silos between IT operations and facilities management. Its goal is to create a single, deterministic digital twin of the entire environment. Instead of relying on probabilistic assumptions, UIIM ensures that every server, power connection, cooling asset, and network link is mapped, tracked, and physically correlated within a unified data model.

By establishing this single source of truth, we create the essential physical guardrails that any future AI system will need in order to operate safely.

The steps toward autonomy

While closed-loop agentic control is the long-term destination, the journey there unlocks massive, immediate value for operators managing today's enterprise and colocation environments. By establishing a UIIM foundation, organizations can deploy "assistive automation" right now to solve their most pressing capacity and operational challenges.

Resolving stranded capacity

- Most enterprises believe they are out of space or power, but in reality, up to 30 percent of capacity is simply "stranded" – paid for, but unusable due to fragmented tracking. By correlating power, cooling, and space data in a single reliable data lake, assistive AI can automatically identify these hidden pockets of capacity and recommend safe placements for new hardware, maximizing the lifespan of legacy facilities.

Automating capacity planning

- Instead of engineers spending weeks manually cross-referencing floor plans and power spreadsheets to deploy a new cluster, intelligent planning engines can analyze the unified data store in seconds. The system recommends the exact optimal rack location that balances power redundancies, cooling availability, and network proximity.

Cross-system root cause analysis

- When a complex outage occurs, IT and facility teams often spend hours in war rooms looking at different dashboards. With a UIIM foundation, an assistive AI can instantly correlate a sudden voltage drop on a facility PDU with a logical application failure happening at the exact same millisecond. It bypasses the noise and identifies the physical root cause, significantly reducing mean time to resolution.

Trusting the machine

As data centers scale to meet the demands of artificial intelligence, our management tools must evolve to keep pace with the speed and density of these new workloads.

The future of data center operations will undoubtedly be agentic. Over time, human engineers will move away from manually reacting to flashing dashboard alerts and instead focus on higher-value responsibilities such as strategic design, capacity planning, and defining the operational rules that guide autonomous systems.

However, that future must be built on a foundation of absolute data accuracy. The organizations that will successfully operate the agentic systems of tomorrow are those doing the difficult work today: breaking down silos, establishing unified data models, and bridging the divide between the physical and digital layers of infrastructure.