DCS
DATACENTRE SOLUTIONS

ISSUE IV 2026
AN ANGEL BUSINESS COMMUNICATIONS PUBLICATION
DATACENTRE.SOLUTIONS

POWERING THE AI ERA

An Interview with Matthew Baynes, Vice President for Secure Power and Data Centre Business

Schneider Electric

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

Accelerate AI Deployments with Prefabricated & Pre-integratedPodand Rack Solutions

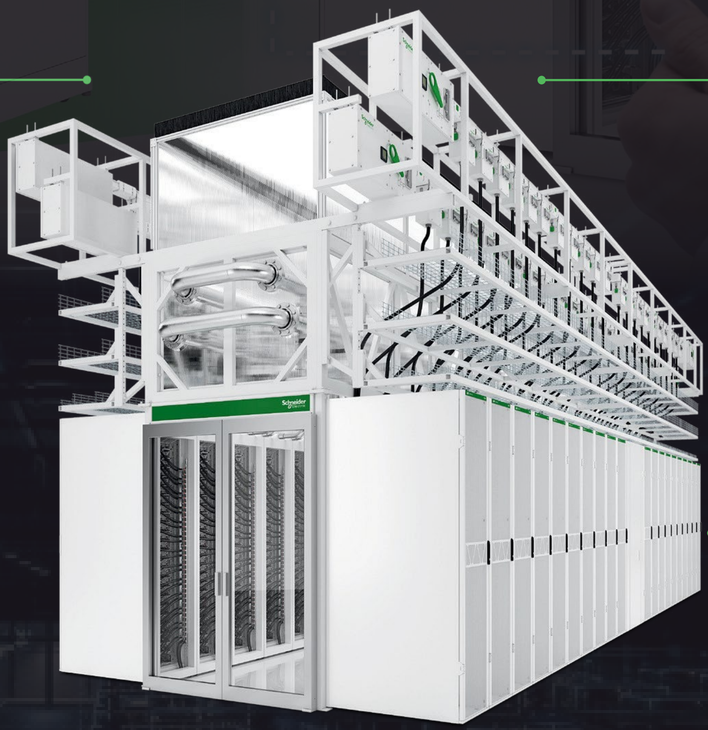
EcoStruxure™ Pod and Rack Solutions from Schneider Electric™ deliver a faster, lower-risk method to deploy AI and accelerated compute infrastructure at scale. How? By shifting complexity off-site.

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

VIEWPOINT

BY PHIL ALSOP
EDITOR

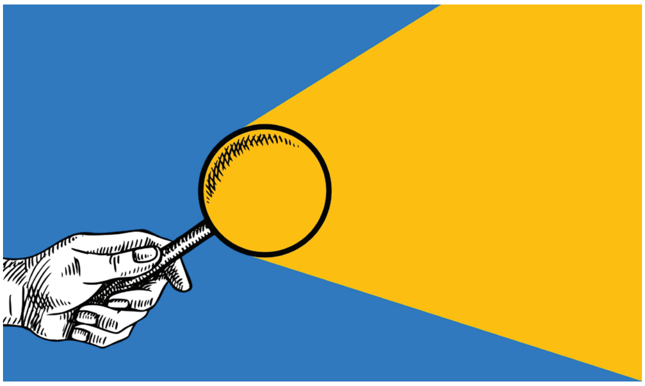
Outages decline, but the industry's risk profile is entering a new era

UPTIME Institute's newly released 8th Annual Outage Analysis Report 2026 offers a sobering reminder that while the data centre industry has made undeniable progress in reducing outages, the path to greater resiliency is becoming increasingly complex. For the fifth consecutive year, outage frequency on a persite basis has declined - a testament to better design practices, improved operational maturity, and the industry's relentless focus on uptime. Yet the pace of improvement is slowing, and the nature of outages is shifting in ways that should give operators pause.

What stands out most in this year's analysis is the rising prominence of failures originating outside the data centre. Publicly reported incidents increasingly point to external infrastructure - particularly grid instability, fibre failures, and connectivity disruptions - as the root cause of extended downtime. As Andy Lawrence of Uptime Intelligence notes, the industry is entering an era where outages are less likely to stem from a single point of failure and more likely to emerge from "complex interactions between systems, including software, networks, and external dependencies." In other words, the perimeter of responsibility is expanding, but control over that perimeter is not.

The financial stakes continue to escalate. According to Uptime's 2025 survey data, 57% of operators reported that their most recent major outage exceeded $100,000 in cost, while one in five saw losses surpass $1 million. These figures underscore a troubling reality: even as outages become less frequent, their consequences are becoming more severe.

Power remains the leading cause of impactful outages, driven by failures in UPS systems, transfer switches, and generators. But the report highlights a new layer of risk emerging from highdensity AI workloads and worsening grid



constraints - factors that will only intensify as demand for compute accelerates.

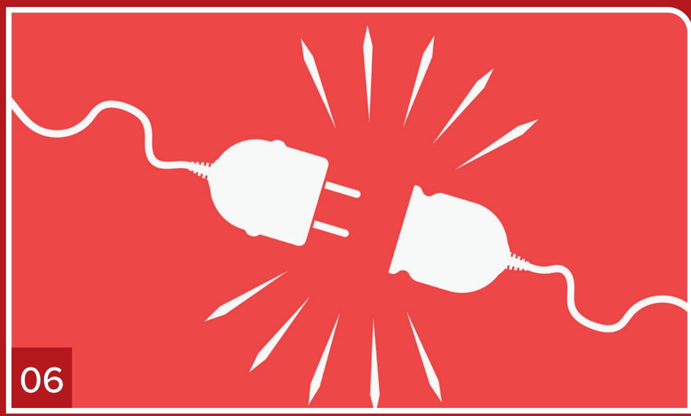
Operators are responding by investing more heavily in automation and advanced control systems. Yet automation introduces its own challenges, including new classes of failure modes that are not yet fully understood. Meanwhile, human error - particularly failures to follow established procedures - remains the top contributor to preventable outages in 2026.

Uptime's latest analysis makes one thing clear: the industry is not facing a crisis of reliability, but a crisis of complexity. As digital infrastructure becomes more distributed, interconnected, and dependent on external systems, the definition of resiliency must evolve. The next frontier is not simply building stronger data centres - it is building stronger ecosystems.

Contents

Cover Feature

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

ISSN 2756-1143 (Online)

Uptime publishes annual outage analysis report 2026

8th Annual Report analyses data on IT and data centre outages including causes, frequency, costs, and consequences.

UPTIME INSTITUTE has released its 8th Annual Outage trends report, an ongoing series from Uptime Institute Intelligence analyzing IT service resiliency. Outage prevention continues to be a central focus for data center operators as demand growth, AI-driven workloads and power constraints reshape risk profiles.

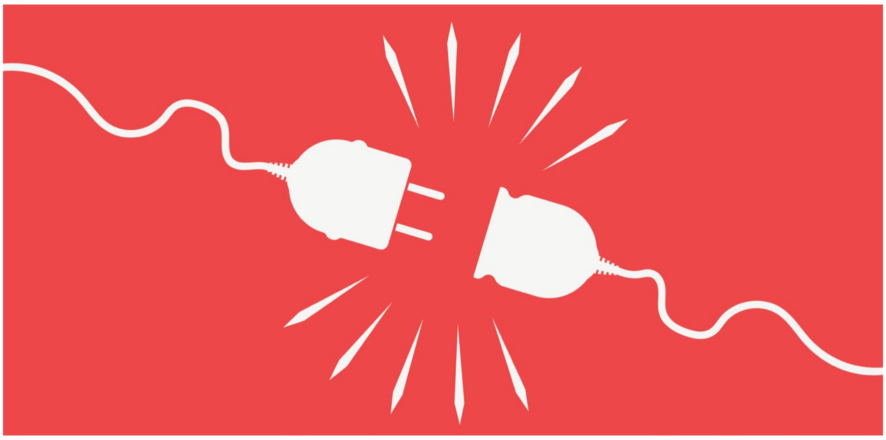
As design and operations improve, operators must still navigate greater system complexity, grid instability, deeper interdependencies and evolving external threats. The 8th Annual Outage Analysis 2026 report analyzes recent data on the causes, frequency and consequences of IT and data center outages.

For the fifth consecutive year, Uptime Intelligence Research suggests that outage frequency on a per-site basis is declining. However, the pace of improvement has slowed compared to previous years and approximately 1 in 10 note their last outage had serious or severe impacts.

In publicly reported outages, external infrastructure failures are becoming more prominent. Also, outages linked to fiber and connectivity issues are rising and more likely to result in extended disruptions.

"Outages overall have slowed down, and overall, digital infrastructure is remarkably resilient. But further resiliency gains are becoming harder to achieve," said Andy Lawrence, founding member and executive director, Uptime Intelligence. "We believe that over time, failures will increasingly not be the result of a single point of failure, but instead be linked to complex interactions between systems, including software, networks, and external dependencies. While site based electrical and mechanical infrastructure remain a critical building block that needs to be resilient, digital infrastructure is becoming more distributed with outages originating outside the data center, including those tied to power availability, network connectivity or the reliance on external cloud services playing a larger role."

Uptime's annual outage analysis is unique in the industry and based on data from a variety of sources, including publicly available reports (e.g., information reported in news and social media), Uptime surveys (such as the Annual Uptime Institute Global Data Center Survey and the Uptime Institute Data Center Resiliency Survey 2026). This information is further supplemented with information from Uptime Institute members and partners, and its database of publicly reported outages.



Key findings include:

- For the fifth consecutive year, outage rates on a per-site basis are declining. However, the pace of improvement has slowed.
- External infrastructure failures are becoming more prominent in publicly reported outages – possibly indicative of a long-term trend foreseen in Uptime predictions. Outages linked to fiber and connectivity issues are rising and more likely to result in extended disruptions.
- Outage costs continue to edge upward. In Uptime's 2025 Annual Survey, 57% of respondents said their most recent major outage cost more than $100,000. For the second consecutive year, 1 in 5 reported costs exceeding $1 million. Around one in ten say their last outage had serious or severe impacts.
- Power remains the leading cause of impactful outages, but the risks are evolving. Failures involving UPS systems, transfer switches and generators are dominant; however, worsening grid constraints and high-density workloads are introducing new pressure points.
- Operators are adapting investment strategies toward automation and control systems to manage complexity, while resiliency assessments remain more focused on internal systems than on external and systemic risks. However, more automation can cause different classes of problems.

For 2026, failures to follow established procedures remain the leading driver of human error-related outages. Issues such as inconsistent or unclear processes are also common, alongside installation and in-service errors.

Beware unintended consequences

CNDCP Paper outlines potential pitfalls in proposed rating scheme for data centres.

THE Climate Neutral Data Centre Pact (the Pact), a collaboration of over 120 data centre operators and associations working towards climate neutral data centres by 2030, has published a paper highlighting some areas of concern with the European Commission's proposal for a sustainability rating scheme for Europe's data centres. Whilst the Pact has supported the concept and the drafting of the proposal throughout, it believes that in key areas its advice and insight has not been clearly heard. Several areas in particular give cause for concern; the unintended consequences of which could damage Europe's competition and growth agendas in the digital era.

The Pact applauds the Commission for aiming to provide consistent, clear and comparable information to allow data centre customers to make informed choices. However, as it stands the proposed scheme offers some raw data and some 'classes' of efficiency without considering the interrelated nature of these systems. The Commission itself understand that to be effective the rating scheme needs a single consolidated label that accounts for all factors.

Europe's climate also varies significantly from north to south, yet the current proposal takes no account of this. Without climate normalisation, the rating scheme may inadvertently hinder or make economically unviable, data centre developments in the south even as demand there grows.

The CNDCP paper also outlines other areas where further consideration of the real-world benefits, and potential unintended consequences of the rating scheme should be more deeply examined. Measures that look good on paper may in fact not deliver the desired environmental benefits. Some may also favour the largest developers and operators and could potentially lock out smaller European competitors from this important market at a time when sovereignty and strategic autonomy are high on the pollical agenda.

Commenting on the paper, Lex Coors, chair of the CNDCP board, said "We have collaborated with the EU on this scheme from the outset, and, as the paper makes it clear, we are fully supportive of efficiency ratings for data centres. We are not trying to advocate for lower targets, but to ensure that the Commission understands the potential unintended consequences of some of its proposal. We offer these thoughts based on real-world data and we hope the Commission will use them to further adapt its proposal in the coming weeks."

Inadequate community engagement is slowing down data centre planning permission

A new Hoare Lea report finds that data centre planning applications are being delayed by an average of 490 days, driven largely by objections related to inadequate community engagement, unclear community benefits, design, infrastructure constraints, and energy use.

Hoare Lea's Societal Insights team analysed 33 disputed applications to understand the underlying reasons for rejection and delay. Rejections frequently cited policy non-compliance, unsuitable locations and insufficient energy strategies.

These findings identify the tension between the growing demand for data centres in the UK, that are supported by the government's AI Growth Zones, and the realities of navigating the planning system.

Carl Walker, Head of Societal Insights at Hoare Lea said: "Success will be judged not only by new infrastructure, but also by the skills, growth and opportunities delivered to local communities, who's needed must be recognised in the planning process."

Ambitions to build data centres in the UK already face significant challenges. Electricity grid capacity is already under strain, particularly in London and the M4 corridor, where data centres have delayed housing developments. Concerns also exist around environmental impacts, green belt land, and controversial government interventions overriding local planning decisions.

The report findings underscore the need for integrated approaches that combine renewable energy planning, transparent governance and meaningful community dividends. By engaging communities early, supporting local skills and infrastructure, and embedding social and environmental value, data centres can become catalysts for sustainable local growth rather than sources of conflict.

Europe's climate also varies significantly from north to south, yet the current proposal takes no account of this. Without climate normalisation, the rating scheme may inadvertently hinder or make economically unviable, data centre developments in the south even as demand there grows

Balancing cloud spend and AI deployment in Europe's digital landscape

European organisations confront a costly inefficiency in their cloud-first strategies, affecting AI deployment and infrastructure resilience.

EUROPEAN organisations are encountering challenges as they adopt AI, in the context of widely used cloud-first strategies. Research from Insight indicates that organisations are utilising, on average, 76% of their annual cloud capacity, with approximately 24% remaining unused. This unused capacity represents expenditure that could potentially be redirected towards infrastructure to support AI at scale.

Across the EMEA region, almost half of organisations spend up to €5 million annually on cloud services. For organisations with an average cloud spend of €3.75 million, this corresponds to approximately €901,000 in unused capacity each year, which may affect the level of investment available for AI platforms, data sovereignty measures, and infrastructure resilience.

These findings are outlined in Insight's Digital Sovereignty Trilemma report, which describes three areas organisations are balancing as AI adoption increases:

Economic efficiency:
Around a quarter of cloud capacity is not utilised, which can affect budgets available for AI development and data platforms.

Operational resilience:
47% of organisations report over-provisioning infrastructure to maintain availability.

Data sovereignty:
Regulatory requirements related to data residency and governance are influencing some organisations to place workloads on sovereign or dedicated platforms, requiring trade-offs between control, cost, and performance.

AI-related demand is contributing to a reported 12% year-on-year increase in hosting costs. At the same time, a growing proportion of organisations identify digital sovereignty as a strategic priority, with this figure expected to reach 82% within three years.

Over-provisioning (47%), limited visibility (47%), and inactive resources (46%) are identified as factors associated with increased cloud expenditure and reduced flexibility. Additionally, 56% of organisations do not conduct total cost of ownership (TCO) assessments before making major workload decisions, and 41% report constraints related to legacy applications, which can affect optimisation efforts.

In response, organisations are increasingly exploring sovereignty-aware hybrid architectures. Approximately 85% are evaluating or implementing dedicated infrastructure for AI. Approaches such as improving visibility, reducing inactive resources, and aligning cloud strategies with AI and data requirements are being considered to improve efficiency.

In the UK, 78% of organisations currently consider digital sovereignty important, with this figure expected to rise to 90% within one to two years and higher over the longer term. As cloud usage expands and AI-related costs increase, organisations are reassessing workload placement to manage costs while maintaining operational requirements.

Hydrogen backup power demonstration at 3 MW scale for data centres

A demonstration of hydrogen-fuelled engines has been completed as part of testing for data centre backup power applications.

INNIO GROUP, in partnership with the Net Zero Innovation Hub for Data Centers, recently completed a demonstration of backup power solutions using 100% hydrogen-fuelled engines at the 3 MW scale. The project is intended to test hydrogen-based approaches for data centre backup power.

The collaborative approach of the Innovation Hub involved participation from experts from Microsoft, Google, and Data4, who witnessed the live tests. Their presence supported assessment of whether the system met operational requirements for data centres.

INNIO's natural gas engines, known for rapid start-up and stable performance, were operated using hydrogen to assess performance under data centre conditions.

The Jenbacher engine was tested during 3 MW-class trials at INNIO's research facility. Testing included AI load profiles and rapid load fluctuations. The results were used to evaluate hydrogen engines as a potential backup option for data centre operations.

Findings from the structured innovation process reflect ongoing industry activity related to decarbonisation. Following an international Request for Information (RFI), clean fuels including hydrogen were identified as potential alternatives to diesel generators for data centre backup power.

The Innovation Hub, supported by organisations including Data4, Google, and Microsoft, enabled MW-scale validation of the technology. The collaboration is part of wider efforts to explore scalable approaches within the sector.

Veolia works with Amazon to develop reclaimed water for data centre cooling

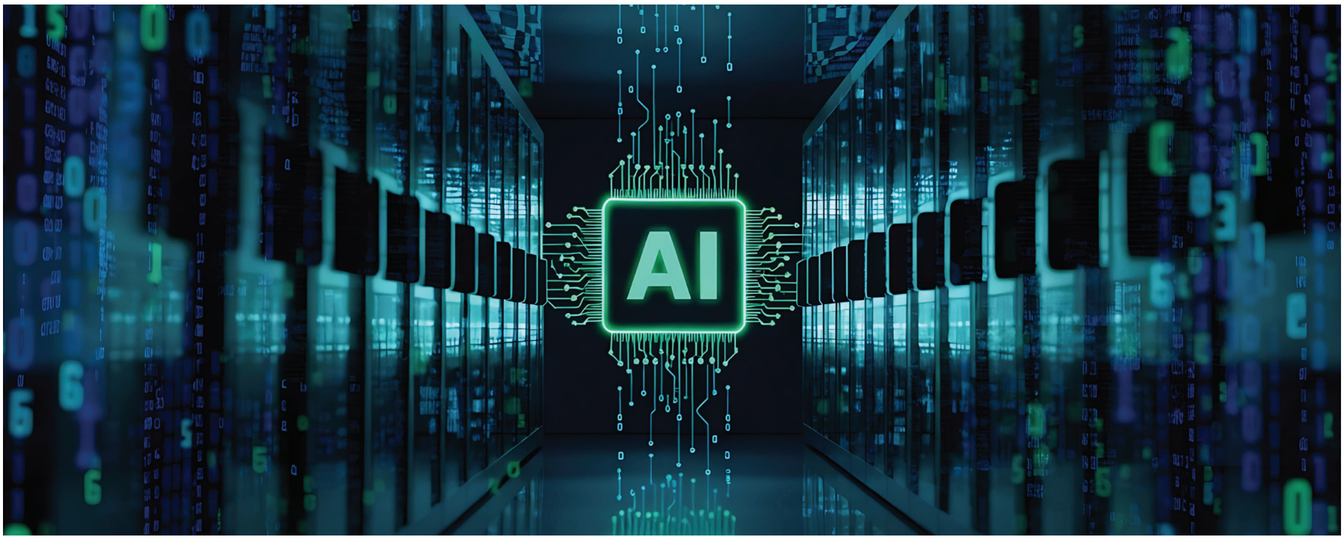
VEOLIA is working with Amazon to reduce data center water use and advance water reuse in Amazon's data center operations in Mississippi, contributing to local water resilience while supporting Amazon's goal to be water positive across its direct data center operations by 2030.

Turning wastewater into cooling power for data centers

The first facility is expected to be operational in 2027, making it the first Amazon data center in Mississippi to use reclaimed water for cooling. Veolia, world leader in water technologies, will deploy autonomous, innovative containerized treatment systems that will transform effluent from nearby wastewater treatment plants and other available sources into cooling water that meets the quality standards required for industrial cooling processes.

The project is expected to reuse more than 83 million gallons of potable water per year once fully operational, equivalent to the annual water use of approximately 760 U.S. homes — estimated to be equivalent to the volume of water the data center would otherwise draw from local groundwater and potable water supplies.

The modular, containerized design of Veolia's water treatment systems enables scalable deployment, allowing it to replicate the solution at Amazon facilities around the world where conditions are suitable, working alongside its industrial and municipal clients. This approach supports both partners' ambition to advance responsible water stewardship and more sustainable data centers operations.



Powering the AI Era

Schneider Electric's vision for data centre growth in the UK and Ireland – an interview with **Matthew Baynes, Vice President for Secure Power and Data Centre Business**

THE data centre industry is entering a transformative phase, driven by the rapid acceleration of artificial intelligence (AI), evolving cloud demands, and increasing pressure on energy infrastructure. At the forefront of this shift is Schneider Electric, which has recently appointed a new leader to guide its ambitions across the UK and Ireland.

As Vice President for Secure Power and Data Centre Business, Matthew Baynes steps into a role that reflects both the scale of opportunity and the complexity of the current market. His remit spans strategy, commercial leadership, and full accountability for Schneider Electric's data centre ambitions across the region - placing him squarely in the middle of one of the most dynamic sectors in digital infrastructure today.

The evolving data centre landscape

While AI dominates headlines, the reality of the data centre market is more nuanced. According to Matthew, cloud computing remains the foundational driver of growth in the UK. Hyperscale cloud infrastructure continues to underpin demand, providing the baseline upon which newer technologies, including AI, are built.

However, AI is undeniably reshaping the trajectory of the industry. Two distinct but interconnected trends are emerging: AI training and AI inference. Training, particularly for large language models, requires vast computational resources and often gravitates toward regions with lower power and land costs. In contrast, inference, where AI models are deployed in enterprise and real-world applications, tends to be more distributed and closer to end users.

This distinction is critical when considering the UK's role in the global AI ecosystem. While the country may not host the bulk of large-scale training workloads, it is well positioned to support inference-driven applications, enterprise AI adoption, and sovereign AI initiatives.

The UK's position in a global AI race

There has been ongoing debate about whether the UK risks falling behind larger markets such as the United States or parts of Asia in AI development. Matthew offers a balanced perspective: the outcome will largely depend on how the UK leverages its strengths and addresses its constraints.

One of the most significant challenges is the cost and availability of power. Compared to other regions, the UK faces relatively high energy costs, which can deter large-scale infrastructure investment - particularly for energy-intensive AI training workloads. However, this does not preclude growth. Instead, it shifts the focus toward strategic applications and targeted development.

Government-backed AI growth zones represent a key opportunity. These initiatives aim to enable the rapid development of large-scale campuses - potentially up to 500 megawatts - designed to support future AI workloads. What ultimately fills these campuses will depend on policy direction, particularly around sovereign AI and public-sector adoption.

For Schneider Electric, this creates both a challenge and an opportunity. The company must navigate existing constraints while positioning itself to support emerging demand in a market that is still defining its AI identity.

Innovation in the age of AI infrastructure

As AI workloads evolve, so too must the infrastructure that supports them. One of the most significant shifts is the increasing density of computing hardware, particularly GPUs. AI clusters, often built around technologies from companies like NVIDIA, are pushing the limits of traditional data centre design.

This has direct implications for power, cooling, and deployment models.

Schneider Electric is responding with a combination of innovation and strategic collaboration. Close alignment with NVIDIA allows the company to anticipate future hardware requirements and adapt its infrastructure solutions accordingly. This partnership-driven approach ensures that Schneider Electric remains aligned with the cutting edge of AI development.

Cooling is a particularly critical area. As GPU densities increase, traditional air cooling becomes insufficient. Liquid cooling technologies, such as coolant distribution units (CDUs), are rapidly becoming essential for next-generation data centres. Schneider Electric's investments and acquisitions in this space signal a clear commitment to supporting high-performance AI environments.

Speed and standardisation: Meeting demand at scale

Another defining characteristic of the current market is speed. The pace at which AI infrastructure needs to be deployed is unprecedented. Organisations are racing to build capacity, often under intense competitive pressure.

To meet this demand, Schneider Electric is focusing on standardisation and modularity. Prefabricated AI pods, reference designs, and pre-engineered solutions are becoming central to its strategy. These approaches enable faster deployment, reduced complexity, and greater scalability.

Rather than designing each data centre from scratch, customers can leverage proven architectures that are optimised for AI workloads. This not only accelerates time to market but also reduces risk - an important consideration in an environment where technology is evolving rapidly.

Growth in a expanding market

With demand surging, the question for established players like Schneider Electric is whether the focus should be on defending existing market share or pursuing new growth opportunities.

With demand surging, the question for established players like Schneider Electric is whether the focus should be on defending existing market share or pursuing new growth opportunities

Matthew's view is pragmatic: the scale of demand is so vast that growth is the primary objective. While increasing market share remains an ambition, the immediate priority is to scale alongside the market and ensure the company can meet customer needs.

This requires careful balance. Rapid growth can strain supply chains, engineering capacity, and service delivery. Schneider Electric's investments in R&D, manufacturing, and partnerships are therefore as much about resilience as they are about expansion.

Encouragingly, the company believes it is well positioned. Its heritage in both traditional data centre infrastructure

and emerging technologies provides a strong foundation for growth across multiple segments - from cloud to AI.

Becoming an energy technology partner

Perhaps the most significant shift in Schneider Electric's positioning is its ambition to be more than just an infrastructure provider. The company is aiming to become a true energy technology partner for the data centre industry.

This reflects a broader trend: energy is no longer just a supporting function, it is a central challenge. Data centres are among the most energy-intensive assets in the digital economy, and the rise of AI is only increasing that demand.

To address this, Schneider Electric is focusing on three key areas:

- **Efficiency:** Developing solutions that enable more efficient data centre design and operation, helping customers maximise performance while minimising energy consumption.
- **Future-proofing:** Creating flexible, scalable systems that can adapt to evolving workloads and technologies.
- **Services and software:** Providing tools that allow customers to monitor, manage, and optimise their infrastructure in real time.

This holistic approach recognises that the challenges facing data centre operators are interconnected. Power availability, sustainability, performance, and cost must all be addressed simultaneously.

Supporting customers through rapid change

The rapid growth of the data centre market brings with it a range of challenges for operators. From securing power connections to managing construction timelines and integrating new technologies, the pressure is significant.

Schneider Electric's strategy is to act as an enabler, helping customers navigate this complexity. By offering reference designs, modular solutions,

This holistic approach recognises that the challenges facing data centre operators are interconnected. Power availability, sustainability, performance, and cost must all be addressed simultaneously

and integrated services, the company aims to simplify deployment and reduce friction.

In an environment where delays can have significant financial and competitive implications, this support can be a critical differentiator.

Looking ahead

Although still early in his tenure, Matthew's priorities are clear. The focus is on leveraging Schneider Electric's existing strengths while adapting to the demands of a rapidly evolving market.

The UK and Ireland may face challenges in terms of energy costs and infrastructure constraints, but they also have significant opportunities, particularly in enterprise AI, sovereign capabilities, and innovation-led growth.

For Schneider Electric, the path forward lies in aligning technology, energy, and customer needs. By doing so, the company aims not only to grow with the market but to help shape its future.

As the AI revolution continues to unfold, one thing is certain: the data centre industry will remain at the heart of digital transformation. And companies that can combine innovation with practical delivery, particularly in energy and infrastructure, will play a defining role in what comes next.

Worldwide AI spending to grow 47% in 2026

Worldwide spending on AI is forecast to total $2.59 trillion in 2026, a 47% increase year-over-year, according to Gartner.

➤ Table 1: Worldwide AI Spending by Market, 2025-2027 (Millions of U.S. Dollars)

Source: Gartner (May 2026)

"THROUGH THE next several years, the need for capacity will make AI infrastructure, including AI-optimized IaaS, AI-optimized servers, AI network fabric, AI processing semiconductors and devices, the largest segment of the market, accounting for over 45% of spending, which will be driven by vendors," said John-David Lovelock, Distinguished VP Analyst at Gartner. "Within this segment, spending on AI-optimized servers will triple over the next five years to become the largest subsegment, as cloud services providers expand capacity in anticipation of the workloads created by GenAI models and agentic workflows."

Market	2025	2026	2027
AI Services	436,351	585,527	759,418
AI Cybersecurity	25,920	51,347	85,997
AI Software	282,897	453,209	638,431
AI Models	15,494	32,604	59,161
AI Platforms for Data Science and Machine Learning	21,292	29,928	42,639
AI Application Development Platforms	6,587	8,416	10,922
AI Data	826	3,126	6,480
AI Infrastructure	975,581	1,431,509	1,890,310
Total AI Spending	**1,764,947**	**2,595,667**	**3,493,358**

Enterprises will expand their use of both the GenAI models embedded in existing software applications and the new AI agents within multiple workflows. Model consumption will increase through multistep processes and integration into broad suites of tools as enterprises recognize the potential value of agentic automation. This dynamic means that the short-term outlook for AI models has been increased to 110% growth in 2026, adding $6 billion in spending for this year (see Table 1).

"Up to this point, AI spending has primarily been driven by technology companies and hyperscalers," said Lovelock. "Enterprises have yet to really flex their spending potential. That is coming and 2026 will be the inflection year. Currently, organizations show limited appetite for using AI to drive disruptive enterprise change. Instead, they favor tactical AI initiatives with incremental improvements in efficiency and productivity.

"For this reason, CIOs face challenges in proving the value from AI investments and demonstrate tangible business outcomes," said Lovelock. "Aligning AI initiatives with strategic business objectives is the essential step for success. This incremental approach persists despite AI hype and valuations that reflect aspirations to transform the broader economy."

AI observability to increase

Forty percent of organizations deploying AI will implement dedicated AI observability tools by 2028 to monitor model performance, bias and outputs, according to Gartner, Inc., a business and technology insights company.

"AI is everywhere, but most organizations are still figuring out how to monitor and trust these systems," said Padraig Byrne, VP Analyst at Gartner. "That visibility gap makes scaling risky and that's why observability matters. Unlike traditional software, AI's decision making is often hidden, making it hard to explain or trust, yet errors can cause substantial financial loss, reputational damage and regulatory scrutiny."

Gartner defines observability as the characteristic of software and systems that enables them to be understood based on their outputs and enables questions about their behavior to be answered. AI observability requires dedicated tools that manage and assess the behavior, decision-making and risks of an AI solution, such as model drift, bias and LLM logic.

"The shift to specialized AI observability tools is accelerating due to executive concern over risk management in complex AI models and agentic AI, not just for infrastructure or application health," said Byrne. "There's a growing need for predictive issue detection and real-time actionable insights in AI models. Failure to adopt these tools exposes organizations to significant governance risks."

According to Gartner research, AI observability also includes the ability to monitor the availability, performance and accuracy of the AI platforms beyond risk and trust, which becomes essential as enterprises increasingly rely on AI-driven outcomes for decision-making.

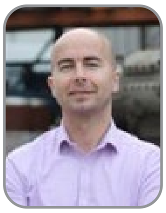
"Without clear, standardized model telemetry, infrastructure and operations (I&O) teams face prolonged incident resolution times for AI applications, which will require complex manual efforts to trace and debug the behaviors of opaque deep learning models," said Byrne. "Dedicated AI observability provides the necessary mechanisms to monitor and mitigate algorithmic risk, establishing the technical foundation for widespread enterprise AI trust and adoption."

"Unlike traditional software, AI's decision making is often hidden, making it hard to explain or trust, yet errors can cause substantial financial loss, reputational damage and regulatory scrutiny."
– *Padraig Byrne, VP Analyst at Gartner*

Gartner recommends I&O leaders factor the following steps into their AI platform strategies:

- Establish mandatory AI model monitoring policies for all production deployments, requiring continuous tracking of fairness, drift and data quality metrics.
- Standardize monitoring frameworks across data science, MLOps and engineering teams to ensure consistency and control. This mitigates organizational silos and streamlines issue resolution.
- Prioritize infrastructure capable of ingesting and analyzing high-volume model telemetry, focusing on specialized solutions that support distributed tracing of AI inference calls.
- Ensure IT strategies include provisions for future monitoring of AI platform performance, detection of shadow IT activity and cost management to address these challenges as the technology matures.

The benefits of free heating (and cooling) to data centres and their neighbours



Keeping data centres cool is a well-documented costly and energy intensive process. By creating a symbiotic relationship with adjacent buildings, and looking to chillers that utilise ambient temperatures, data centres can cut costs and carbon, while becoming beneficial to the communities in which they operate.

BY TIM MITCHELL, SALES DIRECTOR, KLIMA THERM

THE conversations around data centres are changing, technological developments and better use and management of the resources they use and generate, means that having a data centre as a neighbour can be a very positive thing, particularly in urban environments. Locating data centres within city boundaries brings logistical complexities – building density means limited space, strict zoning regulations, and heightened demand for reliable connectivity – there are undoubted benefits for neighbouring communities, that must be harnessed if data centres are to take their place in a productive, net zero future.

End users closest to the data centre will benefit from low latency and rapid access to digital services, making the zone around the data centre ideal for businesses and public services. The building and maintenance of the data centre creates jobs, and current planning models mean that data centre owners can't build with impunity; investment in infrastructure being part of the deal.

Take the heat off

From the data centre owner and builders' point of view, this approach is of course a 'good thing to do', ensuring ethical business operations, while bringing the locality on board, which can go someway to preventing any planning objections.

One of the more direct ways a data centre can give back is through heat recovery, which is as beneficial for the data centre as it is for adjacent buildings – a win win scenario. Data centres can recover nearly 97% of their consumed electrical energy in the form of heat. Integrating heat networks enables urban areas to use this by-product and reduce carbon emissions, a method that is considerably more effective than releasing excess heat into the atmosphere, providing measurable benefits to both the environment and the data centre.

Data centres typically generate heat output in water loops at temperatures between 30 and 35°C, which can be boosted to around 80°C using high temperature heat pumps. This warmed water can be redirected to provide domestic hot water and even heat entire structures.

Cool for free

To prevent heat wastage chillers that include free cooling, taking advantage of low ambient temperatures in winter months, are gaining in popularity. Free cooling can operate for 70 to 80% of the year with a chilled water setpoint of 20°C in a typical data centre in London. In colder regions like Scotland or northern England, that figure can be even higher.

For data centre operators, free cooling leads to reduced energy consumption and cost savings, including less intensive maintenance requirements. Free cooling chillers can reduce the wear and tear on mechanical cooling equipment, potentially extending their lifespan and reducing the frequency of replacement and repairs.

A data centre that reuses its waste heat, while utilising outside temperatures where possible, will operate efficiently and cost effectively, while taking a useful and cooperative place in the fabric of modern life.

www.klima-therm.co.uk

Data centre UPS batteries: 3 strategic considerations

As data centre power demands rise and space tightens, UPS battery strategy is moving beyond uptime and reliability toward efficiency, lifecycle cost, and sustainability considerations.

BY PAUL WILLOUGHBY, SALES DIRECTOR UK & IRELAND, ENERSYS

UPS BATTERIES remain essential to data centre resilience. But operating conditions are changing: power demand is climbing, scrutiny is rising, and physical constraints are tightening.

Batteries can't be specified as 'just another component' in a UPS. They should be treated as part of a holistic backup power strategy—supporting capacity growth, cost control, and compliance, not only uptime. Here are three key considerations to help inform a modern data centre UPS battery strategy.

Space is now a competitive constraint

In key European markets, demand is rising while usable capacity remains tight—reducing the margin for inefficient layouts. 'Hidden footprint' decisions—like how much space is allocated to power infrastructure—now carry real commercial impact. CBRE reports that in 2025, vacancy across FLAP-D markets fell year-over-year, while rental rates rose sharply. Governments are also encouraging data centre expansion beyond traditional hotspots to unlock new capacity. The UK's AI Growth Zones are one example of support for buildout in new regions beyond London. [1]

Why it matters for UPS batteries
Legacy, lower-energy-density batteries can occupy more space than necessary to meet reserve-power requirements—space that could otherwise be used for revenue-generating IT load. A modern strategy treats energy density, footprint, and layout as design variables, not afterthoughts.

Focus on lifecycle cost, not just purchase price

Cost pressures are not easing for operators. The Uptime Institute Global Data Centre Survey 2025 noted that cost issues remained a "top concern" in 2025, reflecting pressure across energy, cooling, maintenance, and other overhead. [2]

Why it matters for UPS batteries
The true cost of a battery system is rarely the purchase order alone. Total cost of ownership (TCO) includes energy overheads, monitoring, maintenance regimes, replacement cycles, and the operational friction created by unplanned interventions.

Even seemingly small efficiency losses can compound over years of continuous operation. Float charging keeps batteries at standby readiness via a constant trickle charge; inefficient float behaviour can quietly inflate electricity costs—and associated emissions—over time.

An effective UPS battery strategy should therefore consider long-term TCO over upfront cost.

Lifecycle and reporting expectations are tightening

Sustainability scrutiny is shifting from broad ambition statements to measurable reporting—and, in some regions, regulation. The EU's revised Energy Efficiency Directive raises expectations around monitoring and disclosure of data centre energy performance, and the Commission has signalled further policy attention. [3]

Battery systems sit inside this broader compliance story too. EU Regulation 2023/1542 introduces a lifecycle framework aimed at improving battery sustainability and circularity. While many requirements fall on manufacturers and importers, operators may need to apply due diligence in procurement, documentation, and end-of-life handling.

Why it matters for UPS batteries
A battery strategy that ignores end-of-life handling, documentation, and auditability risks becoming a compliance headache later. A holistic approach plans for replacement, recycling routes, and evidence packs from day one.

A holistic strategy supporting growth

In today's environment, uptime and reliability alone are not a complete UPS battery strategy. Operators that scale successfully will align UPS battery choices to space efficiency, lifecycle cost, and regulatory direction—so resilience supports growth, not just survival.

REFERENCES

- [1] UK Government, "AI Growth Zones." https://www.gov.uk/government/collections/ai-growth-zones.
- [2] Uptime Institute, Uptime Institute Global Data Center Survey Results 2025, 2025, https://uptimeinstitute.com/resources/research-and-reports/uptime-institute-global-data-center-survey-results-2025.
- [3] European Commission, "Focus: Data Centres – An Energy-Hungry Challenge." https://energy.ec.europa.eu/news/focus-data-centres-energy-hungry-challenge-2025-11-17_en.

PLATINUM PARTNER

Shrinking the 1: Why IT load is no longer untouchable

As AI workloads push data centre power demands to new extremes, attention is beginning to shift beyond cooling and infrastructure toward the IT load itself. New real-world evidence suggests server energy consumption can be reduced by around a third with no measurable performance impact, challenging long-held assumptions about efficiency optimisation.

BY JORDAN TOULSON, SENIOR PRODUCT MANAGER, QIO TECHNOLOGIES

REAL-WORLD evidence shows server energy consumption can be cut by around a third with no performance impact. It's the biggest untapped efficiency gain on the table. The metric we all rely on is the reason we've been missing it.

PUE has driven real progress. We've pushed average scores from around 2.5 to 1.4 by going after the numerator: cooling, UPS losses, lighting, airflow. That work matters. It also hit diminishing returns some time ago. The easy wins on the facility side have gone.

The denominator or the "1" in the PUE equation - the IT load itself, has been treated as untouchable throughout. Servers, the largest energy user inside that load, are revenue. They're the workload. You don't mess with them. That instinct made sense when there was nothing practical to be done about it. It no longer does.

Industry evidence now points clearly in the other direction. Lab work from World Wide Technology [1] and earlier research from Intel [2] established that server energy can be optimised in real time without compromising performance. More recent real-world validation has taken the argument further. Across a mixed fifty-host estate, independently metered at rack PDU level over a two-week window, a sustained 35% reduction in server energy consumption was recorded with no observable impact on workload performance [3].

That's around 62W saved per server on average, across mixed AMD and Intel silicon, with workloads, SLAs and operational policies left untouched. Apply the same per-server saving across a 2,000-server estate and the picture changes: roughly £260,000 off the annual electricity bill and around 185 tonnes of CO_2e avoided every year.

On a practical level this means that the long-standing instinct to treat server energy as sacred no longer has an evidence base behind it. Rack power came down. Workloads didn't notice. The trade-off the sector has assumed was there, simply wasn't.

So why aren't we chasing it? Partly because the dominant metric punishes us for trying. Cut the IT load and PUE often looks worse, even when total consumption has fallen. A smaller denominator pushes the ratio up. The optics go the wrong way while the electricity bill goes the right way. It's a known flaw, and one the DCA and other industry bodies are well placed to help the sector rethink.

For the C-suite the question is simple. Are we protecting a number on a slide, or taking cost out of the business? The power envelope of a rack is no longer fixed. AI-enabled control of server energy is commercially available, operationally proven and quick to deploy. The barrier is mindset, not technology.

We've spent twenty years shrinking everything else. The next phase belongs to operators who finally turn their attention to the one.

REFERENCES

- [1] (2024) ServerOptix lab evaluation, World Wide Technology (WWT)
- [2] (2022) Leveraging AI for smarter data centre power efficiency, Intel
- [3] (2026) Validating Scalable Server Energy Optimisation in Real-World Data Centre Conditions, Qnetix


«Simply better connected»
R&Mfreenet 4.0 – the structured cabling system for smart, modern Data Centres
Reichle & De-Massari AG
Binzstrasse 32, 8620 Wetzikon
+41 (0)44 933 81 11, hq@rdm.com
www.rdm.com
R&M

Designing intelligence into the electrical foundation of data centres

IoT-Enabled Components and the next phase of data center infrastructure.

BY GEORGE CONNELLY, STRATEGIC BUSINESS ANALYST AT WESGARDE

DATA CENTERS have always been critical infrastructure, but their role has expanded dramatically in recent years. As artificial intelligence, cloud computing, and real time digital services continue to scale, data centers are under pressure to operate with unprecedented levels of efficiency, reliability, and responsiveness. This shift is driving a quiet but fundamental transformation in how data center infrastructure is designed and managed—one increasingly shaped by IoT-enabled components embedded deep within power, cooling, and control systems.

Rather than relying on periodic inspections or static thresholds, modern data centers are moving toward continuous visibility. Sensors, displays, relays, and control electronics now generate streams of real time data that allow operators to monitor conditions remotely, anticipate failures before they occur, and fine tune performance across sprawling facilities. These capabilities are not limited to software layers alone. Much of the intelligence enabling this shift lives in the physical components themselves, where rugged electrical hardware is now paired with connectivity, cloud access, and advanced user interfaces.

The expanding role of data in physical infrastructure

At its core, IoT in data centers is about transforming physical conditions into actionable insight. Temperature, power draw, load distribution, and system health have always mattered, but they were historically difficult to measure continuously and at scale. Today, embedded sensors and smart displays make these variables visible in real time, often from hundreds of miles away.

This shift has particular importance in high density environments. As computing power increases, so does heat concentration, making thermal management one of the most critical challenges data centers face. IoT-enabled components allow operators to see how conditions vary across zones, racks, or even individual chips, revealing patterns that would otherwise remain hidden. Over time, this data becomes the foundation for smarter decisions—adjusting cooling strategies, balancing loads, and identifying inefficiencies before they escalate into outages.

From monitoring to predictive maintenance

One of the most significant outcomes of increased visibility is the rise of predictive maintenance. Rather than responding to failures after they occur, data centers are increasingly using historical and real time data to anticipate issues in advance. When patterns indicate abnormal behavior—such as sustained temperature increases in a specific area—operators can intervene

➤ IoT-enabled electrical infrastructure delivers continuous visibility into power and thermal conditions across modern high-density data centers.

➤ IoT-ready electrical enclosures pair rugged protection with certified ingress ratings, helping data centers safeguard critical electronics while maintaining reliable, long-life operation in demanding thermal and cooling environments.

early, minimizing downtime and preventing cascading failures.

This approach is especially valuable in facilities where uptime is non-negotiable. By segmenting floor plans and tracking performance trends across individual zones, operators gain a granular understanding of how systems behave over time. Predictive maintenance does not eliminate the need for on-site personnel, but it allows staffing resources to be used more strategically, reducing the need for constant manual oversight.

Importantly, predictive capabilities depend on the quality and accessibility of data. IoT-enabled components that integrate seamlessly with cloud platforms provide richer datasets and faster response times, enabling maintenance strategies that evolve alongside the infrastructure itself.

Certifications, reliability, and the physical realities of data centers

As IoT becomes more embedded in physical infrastructure, compliance and durability take on heightened importance. Data centers operate in demanding environments where water, heat, and electrical loads coexist at scale. Components must meet stringent certifications to ensure safe operation over long lifecycles.

Environmental and material standards—such as RoHS and REACH compliance—help ensure that components meet regulatory requirements across global markets. In parallel, ingress protection (IP) ratings are increasingly critical, particularly for enclosures and electronics exposed to cooling systems that rely heavily on liquid circulation. In some applications, enclosures must also withstand extreme conditions, including high heat or explosive debris, adding another layer of complexity to component selection.

These considerations underscore a key reality: IoT-enabled components must be as physically robust as they are digitally capable. Data alone is not enough if the hardware generating it cannot withstand the environment in which it operates.

> As IoT becomes more embedded in physical infrastructure, compliance and durability take on heightened importance

Engineering led integration in a growing market

As data centers evolve, so too does the role of component partners who support their design and operation. This is where companies like Wesgarde enter the conversation—not as broad line suppliers, but as engineering focused collaborators. Rather than offering one-size-fits-all product lists, Wesgarde applies its experience across power distribution, enclosures, and control electronics to help OEMs and engineers integrate components directly into their designs. This approach prioritizes customization and fit-for-purpose solutions, reducing the burden on engineering teams and allowing them to focus on system level challenges.

In the data center context, this often means adapting familiar components—such as power distribution units or enclosures—for new operational demands. Custom enclosure designs, for example, can incorporate transparent doors for visual inspection, tailored dimensions for specific installations, or precision drilling to accommodate complex wiring requirements. These adjustments may seem incremental, but at scale they can significantly improve usability, safety, and efficiency.

IoT, sustainability, and thermal control

Sustainability has become an unavoidable topic in discussions about data centers, particularly given their energy and water requirements. While headlines often oversimplify the issue, modern facilities are increasingly designed around closed-loop thermal systems that prioritize efficiency and conservation.

In these systems, large volumes of water are introduced once and carefully managed over time. Maintaining precise temperature control is essential—not only for protecting equipment, but also for avoiding costly and wasteful system flushes. IoT-enabled thermal controls play a crucial role here, allowing operators to monitor conditions continuously and respond before temperatures drift into unsafe ranges.

Predictive maintenance and real-time monitoring support sustainability goals by aligning operational efficiency with environmental responsibility. In many cases, reducing waste and conserving resources also lowers operating costs, creating incentives that reinforce one another rather than competing.

➤ IoT-enabled thermal controls help data centers fine-tune closed-loop cooling systems, maintaining precise water and air temperatures to protect equipment, conserve resources, and support long-term sustainability goals.

Looking ahead: Intelligence at the component level

The trajectory of data center infrastructure points toward deeper integration between physical hardware and digital intelligence. As manufacturers continue to embed connectivity and analytics into components, predictive capabilities are likely to become more sophisticated and more centralized within cloud platforms.

Distributors and engineering partners act as intermediaries in this process, relaying feedback from the field to manufacturers and helping translate emerging needs into practical solutions. Over time, this feedback loop accelerates innovation, ensuring that IoT-enabled components evolve in step with the environments they support.

For data centers navigating rapid growth and rising complexity, the path forward is not defined by any single technology. Instead, it lies in thoughtful integration—combining proven hardware with targeted intelligence to create systems that are resilient, adaptable, and prepared for what comes next.

TRUSTED POWER CONTROL FOR MISSION CRITICAL DATA CENTRES

DSE335 MKII
Auto Transfer Switch Control Module

Reliable power control is essential in critical operations where downtime is not an option. Whether in data centres, hospitals, or industrial environments, even the slightest power disruption can lead to serious safety, financial, or operational consequences. Robust, intelligent control systems ensure continuous performance, system protection, and complete confidence in essential infrastructure reliability.

The DSE**335 MKII** is an advanced auto transfer switch (ATS) controller designed for seamless power switching between two different power sources (S1/S2). Ideal for mission-critical applications, it offers configurable inputs/outputs, comprehensive monitoring, and a close transition feature for no-break power transfer. Its robust design and flexible programming make it a trusted choice for reliable power continuity.

Speed, scale, and AI

How modular construction is enabling data centre builders to meet the moment.

BY TES GROUP

THE SHEER scale of the AI data centre boom represents a once-in-a-generation opportunity for data centre builders. Worldwide, around £2.2 trillion will be spent on AI data centres between now and 2029. However, the unprecedented scale of demand and the speed at which AI infrastructure must come online to meet the moment presents a huge challenge. AI is not only changing the size of the facilities being built, but also how and where they're delivered. Increasingly, off-site manufacturing of vertically integrated modular electrical rooms is emerging as an essential tool in helping OEMs meet the scale of demand at speed.

The AI boom is here, and It's bigger Than anyone could have imagined

In 2025, the global market capacity of data centers was approximately 59 GW, with Goldman Sachs Research estimating that there will be around 122 GW of data center capacity online by the end of 2030.

While AI is the primary driver of this explosion in demand, the other technologies that drove data centre growth prior to the boom aren't slowing down either. Internet of Things (IoT) technology is pushing edge adoption, the ongoing 5G rollout, and an overall adoption of digital services continues to drive both hyperscale and regional cloud buildouts.

Nevertheless, AI's impact on data centre pipelines goes beyond the number of data centres being built. Facilities designed to support AI training and inference workloads are also pushing rack densities into the stratosphere.

Building the infrastructure to support the age of AI is as much a power challenge as it is one of securing GPUs or taming environmental impact. Power requirements per rack have already doubled, tripled, quadrupled, and could easily go 10× in the next few years. Pre-AI Boom, rack densities hovered around 10–15 kW; now they're hitting 40–80 kW in the same physical footprint and growing. The powertrain has to keep up.

Manufacturing, supply chain, and site selection: New challenges and old methods

The physical build-out of data centres has become another thorny problem. Developers are encountering mounting execution hurdles, with projects mired by lengthy equipment lead times ranging from eight to 24 months.

Speed is of the essence in the current data centre landscape. More conventional manufacturing processes simply can't keep up. These problems are further compounded by ongoing supply chain disruption and rising costs as Brexit and Trump's Tariffs make moving goods across borders an ever more expensive, cumbersome process.

Demand is also intensifying the pressure on an already undersized talent pool. While overall construction labour costs are rising, the most critical constraint is access to the highly specialised electrical engineers and skilled technicians required for complex, high-density power commissioning on-site. Securing this essential expertise is a major bottleneck, often commanding premium rates that cut directly into a builder's P&L and delay project completion.

Even accessing talent (at any price) is a problem. In Europe's biggest data centre markets (the FLAP-D), capacity saturation is already pushing data centre companies out into secondary markets. Frustrated by long lead times and regulatory hurdles, builders are flocking to Tier 2 markets like Lisbon, Madrid, and Milan. While approvals may be faster and land may be cheaper, new markets mean fewer local manufacturing capabilities and smaller talent pools.

The upshot is that demand for AI workloads outstrips the pace at which data centre builders can deliver their projects. Businesses can't afford to wait years for data centre construction to be completed, and every day lost impacts revenues. Traditional on-site construction methods are notoriously slow and cumbersome, which is where modular construction off-site offers a potential lifeline for the industry.

Modular construction allows for speed and scale of delivery

Once very much a niche application suited to small projects in unconventional locations, modular off-site construction is emerging as a vital source of speed at scale in what would be considered traditional data centre builds.

Off-site manufacturing supports faster and more efficient delivery by prefabricating all elements of the power infrastructure at dedicated campuses within skid-mounted, containerised units. Each pod contains a complete suite of power equipment, including high-voltage input, step-down transformers, low-voltage distribution, UPS systems, battery backup, and generator transfer capability. Controlled factory settings then mean they can be commissioned and tested before ever leaving the manufacturing facility. This represents a degree of vertical integration to which traditional data centre projects haven't had access.

Historically, power components arrived at the site from multiple different suppliers requiring lengthy, involved onsite integration and testing. The modular approach shifts this labour, time, and skill-intensive work, which is often vulnerable to local weather and logistical disruption, into a controlled factory environment. The result is improved quality, tighter tolerances, and the completion of rigorous testing, including Factory Acceptance Testing (FAT), to at least Level 1 before any units are shipped. This critical step removes significant specialist electrical engineering and commissioning requirements from the construction site, dramatically reducing project risk and accelerating final handover. This model also allows manufacturers to engineer for specific climates. Pods built for Nordic locations are designed to operate at minus thirty degrees Celsius, while units destined for southern Europe are prepared for temperatures of forty degrees or higher. As the climate crisis worsens, on-site manufacturing is going to face worse

and more frequent disruption due to extreme weather and natural disasters like flooding or heatwaves.

Completing the majority of manufacturing and testing off-site also shortens lead times by allowing power infrastructure to be assembled in parallel with other construction on-site. Modular power pods can be completed while the data hall is still under construction, allowing projects to progress at the same time, rather than one after another.

For data centre builders looking to build bigger and faster in new markets amid a labour shortage, it's easy to understand the appeal.

The approach also supports more efficient use of valuable floorspace. Land in proximity to power generation sources and population hubs where data can travel with minimal latency is rising sharply in price. Operators paying a premium for square footage look to maximise revenue by devoting as much area as possible to servers, rather than housing non-revenue-generating electrical infrastructure. Shifting the powertrain into external pods mounted on skids frees up space for more racks and therefore more income.

There is also an element of flexibility here; standardised pod designs can be deployed wherever demand arises, enabling customers to redirect modules to different sites as requirements evolve. Off-site manufacturing can run in parallel with civil works, accelerating delivery even further.

In an age of skyrocketing costs, modular powertrains also reduce up-front capital expenditure. Instead of building the full electrical backbone of a facility before clients commit, operators can deploy pods when needed and tailor them to each customer's specifications. Because pods can be lifted out and replaced, the modular approach also better supports phased expansion and helps data centres balance cap-ex and op-ex across the facility's lifecycle. Data centres can begin with an initial deployment and easily scale across subsequent phases using the same modular architecture.

For builders seeking efficient European access and a guarantee of high-specification manufacturing and precision engineering, this proven modular method provides a reliable and expedited solution for scaling AI capacity.

Modular in the mainstream

Modular construction is stepping out of its previous role as an enabler of somewhat niche applications like remote sites, disaster recovery facilities, and edge deployments. Modular is moving into the mainstream.

It's a strategy driven by the need for flexibility and speed at scale. The combination of high-capacity manufacturing, controlled conditions, and repeatable modular design enables producers to support dozens of concurrent builds and deliver outcomes that traditional construction approaches can't.

Fire safety at scale: Protecting modern data centre construction sites

As data center construction across North America accelerates, projects are becoming larger, faster and more complex than ever before – and the consequences of getting fire safety wrong have never been higher. Traditional approaches, developed for early, smaller-scale facilities, are being stretched beyond their limits on today's megasites. Drawing on his experience supporting tier-one contractors and data center developers, Aaron Velardi, head of North American development at Ramtech North America, explores how evolving risks, updated NFPA guidance and technology-enabled temporary protection are redefining fire safety during data center construction.

BETWEEN 2017 and 2021, United States (U.S.) fire departments responded to an estimated average of 4,440 fires in structures under construction per year. These fires caused an annual average of five civilian deaths, 59 civilian injuries and $370 million in direct property damage.

Since then, the risk landscape has evolved. Data centers have experienced significant AI-driven growth, accelerating rapidly from 2023 onward. As a result, construction sites have grown dramatically in size and complexity, as data centers have become the foundation of the digital infrastructure that powers the modern economy.

In 2024, there were 5,388 data centers in the U.S. - the most of any country worldwide and ten times more than China and most European countries - with a third of these located in just three states: Virginia (643), Texas (395) and California (319). At the end of 2025, nearly 3,000 were reported as planned or under construction.

In April last year, two people were hospitalized after a minor fire at the $11 billion Amazon data center construction site in New Carlisle, Indiana. At this level of investment, insurance exposure is significant and inadequate fire risk controls can lead to higher premiums, coverage restrictions or costly project delays.

While older facilities, built when the sector was still in its infancy, often spanned tens of thousands of square feet, modern megasites routinely exceed 500,000 square feet, with workforces of at least 850 employees and budgets frequently approaching $1 billion over construction periods of five years.

Once concentrated in urban hubs, these facilities are increasingly being built nearby to small towns and in rural regions across the U.S., yet they all face the same distinct risks - from protecting people up to a quarter mile away to preparing for active shooter incidents and maintaining clear communication across large construction zones.

The cost of neglecting safety

Industry attitudes remain one of the most significant barriers to effective safety on modern data center construction sites. Many owners and contractors default to minimum code compliance, prioritizing short-term cost savings over comprehensive safety measures. When the scale of risk is this high, treating safety as a cost rather than a core project value is a dangerous miscalculation.

Ironically, the very investments that could prevent catastrophic losses - from robust temporary fire protection to continuous monitoring systems - are often the first items cut when budgets are tight or schedules are accelerated. This mindset can leave megasites vulnerable to incidents that could have been prevented with proactive planning.

The combination of human density, complex logistics and high-risk equipment makes comprehensive planning essential rather than optional. The sheer size of modern data center construction sites introduces unique operational and safety challenges. Extensive high-voltage infrastructure, large-scale generators and temporary lithium-ion battery storage during construction further elevate fire risk early in the build, meaning a single incident can ripple outward with far-reaching consequences.

From financial loss and operational disruption to human injury and reputational damage, the cost of inaction is simply too high to ignore. Ensuring robust safety protocols, clear ownership and continuous vigilance is no longer optional - it is an essential component of responsible project management in today's high-stakes data center landscape.

Evolving regulatory frameworks

Regulatory frameworks are gradually evolving to address the unique challenges of modern construction megasites. While there are currently no fire safety regulations specific to data centers, the National Fire Protection Association (NFPA) Codes and Standards provide comprehensive guidance for fire prevention, detection and suppression across all types of construction environments.

NFPA 241, in particular, focuses on construction, alteration and demolition activities, providing detailed requirements for temporary fire protection during the most vulnerable phases of a project. The 2022 edition strengthened Fire Prevention Program requirements, clarified the roles and responsibilities of owners, contractors and designated Fire Prevention Program Managers (FPPMs), as well as introducing provisions for temporary protection systems that can be scaled to sites of any size.

> From financial loss and operational disruption to human injury and reputational damage, the cost of inaction is simply too high to ignore. Ensuring robust safety protocols, clear ownership and continuous vigilance is no longer optional

The forthcoming 2026 update is expected to further strengthen requirements for temporary fire protection, reinforce the role of Fire Prevention Program Managers and place greater emphasis on after-hours protection, site security and closer alignment with other NFPA codes – all directly relevant to large, complex data center construction sites.

These updates highlight a fundamental shift: fire prevention is no longer simply a regulatory checkbox or a responsibility left to local authorities or inspectors. On large, complex sites, ownership of fire safety is now an active, ongoing duty that falls squarely on project leaders, contractors and appointed FPPMs, requiring investment, training and proactive management from day one.

Implementing modern technology

To address the complex and evolving risks of modern construction megasites, the adoption of advanced wireless and IoT-connected fire safety systems has become essential. Traditional approaches - such as fire watches or air horns - rely heavily on human vigilance and are inherently limited. They are often unreliable in extreme weather, difficult to scale across sites and prone to delays or errors.

In contrast, modern technology provides continuous, auditable monitoring and early warning capabilities that can cover every corner of a high-risk construction environment, both during working hours and after-hours when the risk of fire or arson is greatest. Designed for rapid deployment, these systems can protect hundreds of thousands of square feet within days, automatically logging the exact location and cause of every alarm.

This real-time data allows site teams to respond immediately, preventing small incidents from escalating into catastrophic events. It also generates detailed records that can be analyzed to inform targeted prevention programs, track recurring risks and refine site-wide safety strategies. As a result, it extends far beyond simply alerting personnel to flames.

On megasites, these systems can coordinate alerts across vast distances, support efficient evacuations and integrate directly with first responders. For instance, alarms can be relayed to central monitoring centers via cellular or IP-based communications, enabling rapid escalation and coordinated response across large and geographically dispersed construction zones.

From reactive to proactive prevention

In data centers and large-scale infrastructure projects, where downtime can cost millions per day, early detection and rapid intervention are not just life-saving - they are essential risk management. As a result, technology-enabled fire safety delivers both operational and financial benefits for data center construction sites.

Gone are the days when data centers were a simple build; now, they are a high-stakes ecosystem where fire risks are as complex as the technology that will eventually be inside once construction is complete. The industry must move from reactive compliance to proactive prevention by leveraging technology, standardized codes and clear ownership.

Modern risks - lithium-ion storage, energized electrical systems, sheer site scale and personnel safety hundreds of yards away - demand modern solutions. Sites must ensure real-time monitoring, integrated communications and competent FPPMs capable of managing risk across these sprawling environments.

To truly safeguard people and assets, every strategy must be deliberate, coordinated and consistently enforced. In the face of growing complexity, no site should be left unprotected and no life or investment should be left at risk. Data center construction teams must standardize, train, implement - and never leave safety to chance.

Why location data is critical to smarter data centre planning

Data centres are some of the country's most important infrastructure. This is because they power the economy – from online streaming and mobile banking to the development and increasing use of AI. Which is why the government has designated them as critical national infrastructure. But data centres require significant power, water, and connectivity to operate, placing increasing pressure on energy providers, water companies, and telcos. This is why trusted and accurate location data is critical to building more sustainable data centres.

BY OS

The planning challenge

MANY OF these centres are being planned for industrial land near water sources, but the challenge lies in the suitability of those sites - from the quality and availability of water, to the treatment and recycling requirements, and the environmental impact of repurposing brownfield land. Understanding what the land was previously used for, and what habitats exist there, is also essential.

At the same time, there's a growing focus on building new housing, and the surrounding infrastructure needed to support them. They also require analysis of existing roads, substations and sewers.

The role of location data

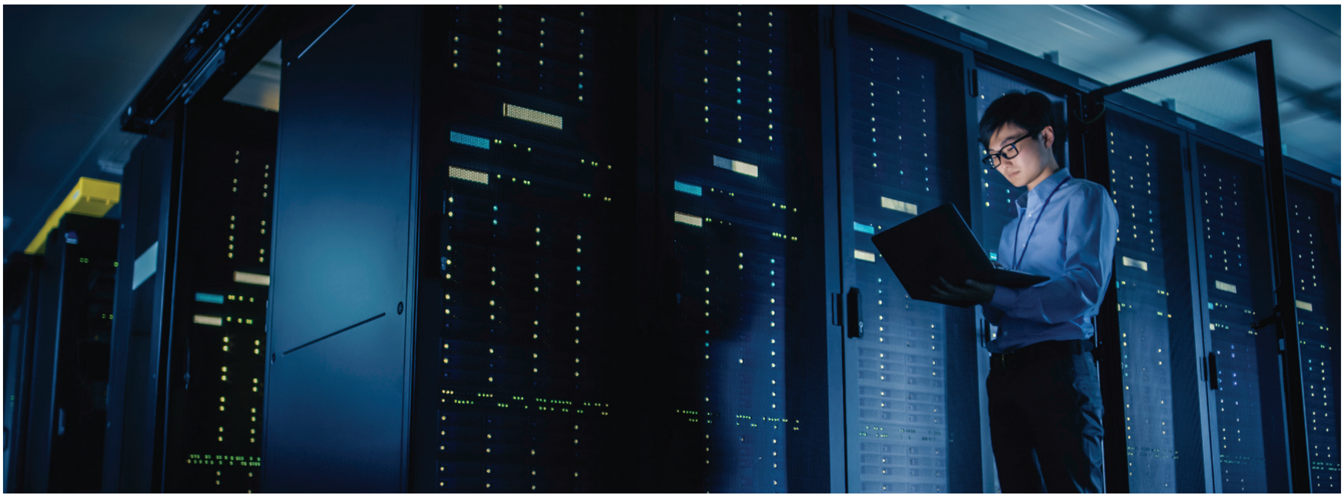
Location data can answer critical questions when integrated with utilities, transport, and environmental analysis. This means tech companies, investment firms, and planners can make informed decisions about where to build and where to invest.

For local authorities, it provides the assurance needed to support planning and ensure long-term viability. Stakeholders get reliable insights to decide where to build, in line with government policy while balancing socio-economic benefits and environmental impact for neighbouring communities.

There's also a real opportunity to use location data more proactively to inform legislation and policy, helping to answer not just where infrastructure should go, but how and why it should be developed.

Future energy networks and sustainability

With the rising demand for data centres, there will also be significant developments in the future energy network. Integrating renewable energy assets into existing systems will be crucial, as well as managing the growth of decentralised energy and ensuring network resilience. This will require smarter planning, and location data will be central to making them viable, especially when mapping terrain stability, and ecological sensitivity.



Data centres present a unique opportunity: location data can help identify areas where waste heat from these facilities could be repurposed to warm nearby homes, supporting government-backed initiatives and helping transition towards circularity in the energy sector for a more sustainable future.

Chris Wilton, OS Utilities Lead commented; "Infrastructure decisions must become truly data-led, informed not only by location, but by environmental impact, socio-economic benefit, and long-term capability. The challenge is around execution – Ordnance Survey's trusted location data provides the foundation for smarter planning and future resilience, helping meet the growing demand for sustainable data centres while supporting government policy."

Many of these data centres are being developed by global tech giants and private investment firms as a reliable long-term investment opportunity. However, building in the wrong location can expose facilities to flood risk, grid constraints, and environmental challenges. Effective planning helps prevent assets from becoming stranded and ensures facilities can continue to scale - for example, by considering the availability of renewable energy both now and in the future.

OS data can support this by providing detailed insight into the local environment, including land-use attributes, and help identify suitable locations for solar power. In addition, terrain and buildings data offer a long-term view that enables investors to make confident, data-led decisions that balance growth with sustainability.

Smaller, smarter data centres

On the other end of the data centre spectrum, innovators are developing solutions to reduce environmental impact and improve resilience offering alternatives that work alongside large-scale facilities. An example is SpaceD, a Geovation-backed start-up founded by Pegah Noori khah.

With AI driving exponential demand for data processing, Pegah saw a risk of reversing progress on decarbonisation. In London, for instance, clusters of data centres in the Isle of Dogs consume up to 75% of local electricity, delaying housing projects because grid capacity has been exhausted.

SpaceD's solution sits at the intersection of energy tech, PropTech, and climate innovation. Its platform aggregates data from electricity grids, fibre connectivity, and the built environment to identify empty commercial sites for smaller, smarter data centres.

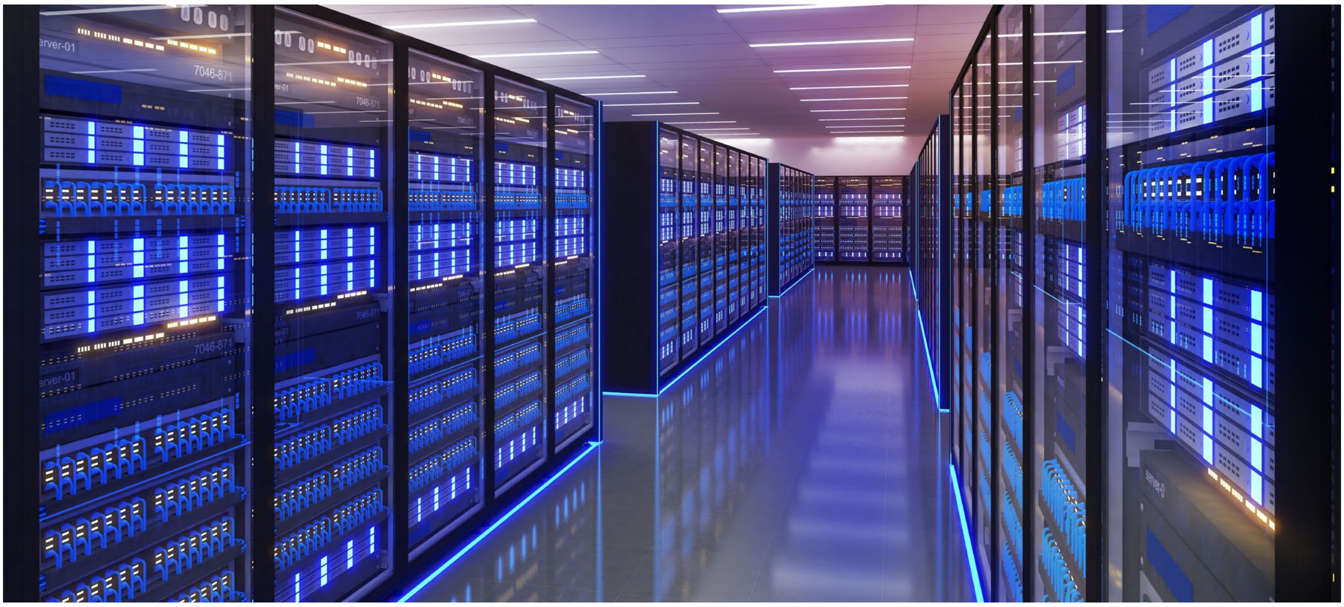
By leveraging OS building data alongside global connectivity datasets and grid information, SpaceD runs complex spatial algorithms. These analyses prioritise commercial buildings within local energy networks, aligning with the UK's shift toward localised energy markets.

"To be more resilient, we need to connect the dots - using integrated data to overcome fragmented infrastructure and unlock new opportunities. This depends on strong interoperability between systems, so AI solutions can work across silos rather than in isolation. It's not just about building large data centres; there's also value in creating micro data centres in the right locations. AI-powered platforms like SpaceD aggregate grid, connectivity, and building data through interoperable layers to identify optimal sites, and highlight how this complementary approach could increase resilience by keeping critical services running when needed." Pegah says.

"Take autonomous vehicles and EVs, for example. To reduce strain on local grids and support emergency technologies, we need decentralised data infrastructure. This eases pressure on the grid through smarter timing of data processing – which also generates usable heat- and places smaller data centres closer to where the data is generated. Doing so improves processing efficiency and supports sustainability by optimising energy use, while advancing technologies mean decentralised data centres can increasingly provide decentralised heat where it's needed."

Conclusion

The rise of data centres is reshaping the UK's infrastructure landscape, but it requires an ecosystem where trusted data, effective infrastructure planning, and innovation come together. Using location data alongside emerging technologies could ensure these critical facilities are built in the right places - resilient, efficient, and sustainable. In turn, this enables the growth of AI and digital services, reinforcing the cycle of innovation that drives economic growth. The key to underpinning the future of the UK's digital economy is getting the location right.



Future-proofing data centres

How silicone roofing cuts PUE and protects uptime.

BY ERROL BULL, P.E., CSI, APPLICATION DEVELOPMENT LEADER AT MOMENTIVE PERFORMANCE MATERIALS INC. MEMBER: IIBEC, ASTM C24, ISO TC59/SC8

DATA CENTRES live and die by their uptime, yet the roof over the servers is often the last thing on anyone's mind – until it leaks!

While the focus in data centre construction is largely on high-tech server stacks and sophisticated cooling systems, roofs are key to preventing water leaks, maintaining thermal performance, and delivering significant savings. As this feature explains, one of the ways to future-proof or restore your roof, save money and protect the mission-critical equipment inside is a silicone roof coating.

The cost of cooling

Cooling can be one of the highest operating costs for data centres, often accounting for 40% or more of total energy consumption, depending on the facility's age. Any material that effectively reduces the facility's thermal load is a direct investment in the bottom line. By reducing cooling energy use, silicone roof coatings may lower the Power Usage Effectiveness (PUE).

A roof's ability to reflect solar energy, known as its Solar Reflectance Index (SRI), is a crucial part of managing the building's heat gain. This is where the choice of roof coating is hugely important. White, highly reflective coatings provide solar reflectance and thermal emissivity.

By reflecting a significant portion of solar radiation rather than absorbing it as a dark roof does, a high-quality reflective, white silicone coating can lower the roof's temperature. This directly reduces heat transfer into the facility, cutting HVAC demand and saving an estimated 10% to 50% on cooling costs, depending on the environment. Overall, this reduction can yield an annual saving of 7% to 15% in cooling costs.

The reduced thermal load isn't just saving money on energy costs; it can also extend the life of the cooling equipment. By reducing stress and operational hours on chillers, fans and compressors, the coating effectively defers significant CapEx costs associated with replacing or overhauling these expensive HVAC components.

Building envelope integrity

Naturally, a data centre's roof is the first line of defence against environmental hazards, with water leaks being a large risk. A roof failure leading to leaks can cause catastrophic damage to mission-critical equipment, resulting in potentially unrecoverable downtime and data loss – unthinkable in the data centre world!

The right silicone coating can provide effective long-term protection against this. Liquid-applied silicone creates a seamless membrane over the entire roof surface. Many data centre roofs have sheet-applied membranes and metal substrates with vulnerable seam joints, whereas a seamless silicone coating provides a robust, durable barrier against water ingress.

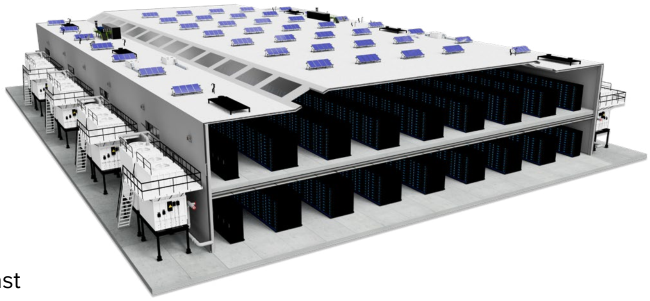
In many locations, data centre roofs are exposed to constant and intense UV radiation, which is a leading cause of premature failure in many organic roofing materials (like single-ply or modified bitumen). Silicone is an inorganic material with a chemical structure that is inherently resistant to UV degradation. This stability ensures that the coating maintains its structural integrity, providing a long-term, reliable asset protection strategy.

Construction efficiency and sustainability

The process of restoring or replacing a data centre roof can be highly disruptive and costly. Apart from the inevitable downtime, there is an inherent risk of dust and debris, and of exposing the highly sensitive equipment below.

Whether you are breaking ground on a new build or maintaining an existing facility, silicone roof coatings add an essential layer of protection. Applying a silicone coating during construction provides a seamless, waterproof barrier from day one, reinforcing the roof's integrity and ensuring the building envelope is watertight from the outset.

Silicone is compatible with almost any roof type (like metal, single-ply, or bitumen). This means you can restore your roof without the mess, the cost, or the significant risk of a complete tear-off. The coating can often be applied directly over the existing substrate with minimal preparation, saving time and money and, crucially, preserving uptime.



As data centres adopt on-site solar arrays, a critical design challenge emerges: the roof should last as long as the panels. With solar arrays typically lasting 25 years or more, they often outlast conventional roofing systems. A high-performance silicone roof coating provides a seamless layer with the elasticity needed to withstand the thermal stress of solar racking, ensuring the roof's lifespan better matches the solar panels'.

Lower energy consumption directly impacts the building's PUE. This reduction in cooling demand is key to achieving a smaller overall carbon footprint, helping the facility meet green building standards such as LEED, BREEAM, or EDGE.

Case in point: Restoring thermal performance and protection

The critical nature of this restorative approach was recently demonstrated at a major technology infrastructure site in Mexico. The facility's six-year-old single-ply PVC waterproofing system had failed prematurely due to constant UV exposure and hail damage.

To avoid costly downtime, the operators chose to apply a high-performance, liquid-applied, 100% alkoxy silicone coating over the 170,000 sq. ft. area. The silicone created a seamless, durable barrier that immediately addressed the risk of leaks while the white finish restored the roof's high solar reflectivity and thermal performance. By choosing this restoration strategy over a full replacement, the facility ensured continuous operations and restored thermal efficiency – all without the disruption, cost and downtime of a major construction project.

A strategic investment

For modern data centre designers and operators, a reflective silicone roof coating can be a low cost strategic investment, especially when restoring older roofs. The right coating supports long-term thermal performance, helps prevent catastrophic failure and reduces maintenance requirements throughout the building's working life. In a sector where avoiding downtime is the main priority, silicone roof coatings are an essential part of the solution.

Commissioning is about certification and verification, not box-ticking

CEO of Global Commissioning, Louis Charlton, discusses how we need to protect commissioning practices as the industry continues to evolve quickly. The certification discipline that covers performance, safety, and reliability across the data centre landscape is vital to ensure IT facilities perform as expected.

OVER THE past decade, the criticality of data centres has increased by an order of magnitude. From the cloud revolution to the AI boom, digital infrastructure growth has seen a global spike in compute capacity and facility size. In the US, the average data centre is expected to grow from around 40 MW today to 60 MW by 2028. Data centre designs are changing, too, with rising rack densities and the widespread adoption of liquid cooling to keep up with tech giants' appetite for AI.

As facilities get bigger, delivery timelines get shorter, and the complexity of the average data centre increases, risk tolerances are shrinking. Much of the success of this worldwide buildout rests on the commissioning industry.

In a world where delays and design mistakes can mean multi-million dollar disruptions to multi-billion dollar projects, commissioning agents are invaluable guarantors that design goals and execution are in alignment. Considering the impact the commissioning process has on every aspect of a project, from energy efficiency to operational reliability, the way our industry approaches and defines commissioning carries real consequences.

As the data centre sector has evolved over the past ten years, so too has the language of commissioning. What was once a clearly defined discipline — technical, structured, and evidence-based — has been increasingly diluted by boilerplate marketing. The result is a growing misunderstanding, both of what commissioning does, and its importance.

Today, almost every contractor or consultant seems to offer "commissioning management" as a service. The phrase has become a fixture on capability statements, often sitting neatly between "design management" and "handover support," as though commissioning were simply another administrative process within a project lifecycle. A box-ticking exercise.

But commissioning is not a box to be ticked at the tail end of a project. It is a critical source of assurance and verification. When the meaning of commissioning becomes blurred, so too does the standard of assurance that

clients expect and depend on. Today, that distinction matters more than ever.

An erosion of meaning

Commissioning has always been about assurance: proving that what was designed, built, and integrated actually performs as intended. As the data centre sector has expanded at extraordinary speed, however, common usage of the term has broadened to cover a vast range of interpretations. As a result, ask ten different data centre companies to define commissioning and you will likely get ten different (albeit overlapping) answers. Some see commissioning as document control or process management; others use it to describe systems testing and facility handover; and a few, still, understand it as a full lifecycle certification of performance.

This broadening of commissioning's definition has, unintentionally, diluted the credibility of the discipline. Clients hear the same word from multiple providers but rarely receive the same level of technical depth, independence, or rigour. The result is confusion and, in some cases, misplaced trust.

Too often, we see commissioning managers parachuted in close to the completion of a project. They are then handed a collection of spreadsheets, checklists, and reports, and tasked with "closing out" the commissioning process. Ticking boxes, in short. It is an approach designed for completion, not for verification. The client may successfully get sign-off on their facility, but such an approach does not equal proof. There is no true assurance.

True commissioning cannot simply be bolted onto the end of a programme or managed remotely from a desktop. It requires technical engagement, continuity, and structure from the earliest design stage through to performance testing. In short, real commissioning is built in at the foundations, not applied retrospectively like a coat of paint over a cracked facade.

Certification: The true goal of commissioning

At its heart, commissioning is a certification process. Commissioning agents interrogate a facility throughout its design and construction in order to deliver measurable, defensible evidence that said facility performs exactly as it was designed. Carbon reporting regulations are becoming more strict, and the price of everything from semiconductors to concrete, labour, and electricity are expected to continue rising through 2026. It should be obvious that any discrepancy between the standards at which a facility was designed to operate and the reality when it comes online can be a minefield of fines, delays, and lost revenue. Commissioning is the ultimate assurance layer between design, construction and operation. It is the point at which assumptions are tested and verified against real performance.

When commissioning is carried out properly, the result is not simply a completed handover file but a certified statement of reliability and readiness. Every sequence, every interface, and every system has been certifiably demonstrated to function safely, efficiently, and in harmony with the facility as a whole.

This is the foundation on which we built Global Commissioning. Our role is not to oversee commissioning, ticking boxes at the end of a project to meet a handover date, but rather to meaningfully certify performance. We validate that design intent has been achieved, that all performance standards are met, and that the asset operates in full accordance with its engineered purpose.

Our certification is not theoretical. It is recognised across EMEA by some of the world's largest owner-operators and end-user organisations for whom

> This is the foundation on which we built Global Commissioning. Our role is not to oversee commissioning, ticking boxes at the end of a project to meet a handover date, but rather to meaningfully certify performance

disruptions and downtime are not an abstract inconvenience but an existential risk. To them, commissioning is not paperwork; it is risk management.

Verification: The missing discipline

Verification is the cornerstone of genuine commissioning, and yet it is often the first element lost when the process is treated as an add-on service. Verification is the process that transforms commissioning from coordination and box-ticking into proof.

This is where technical expertise and independence matter most. Verification is not an administrative exercise; it is the practical observation and testing of systems under load. It means validating the sequence of operations, confirming redundancy, assessing control logic, and ensuring interoperability between disciplines. It demands engineers who understand not just how systems function individually, but how they behave collectively and how those relationships impact a facility's overall performance.

Behind every certification issued by Global Commissioning are professionals who have undergone these processes countless times. They have witnessed systems under real operating conditions, challenged assumptions, and resolved conflicts long before a client ever steps into the facility. Their work not only provides confidence but traceability: data-backed assurance that the design intent has been realised.

Verification may not be glamorous work, but it is essential. Without it, commissioning becomes opinion rather than evidence. Box-ticking without verification.

The role and responsibility of global commissioning

At Global Commissioning, commissioning is not some value-add service or afterthought; it is the reason we exist. Every process, every tool, and every training programme we develop is designed around one purpose: to certify performance.

Our name may include the word "commissioning," but what we actually deliver is assurance, verification, and validation. We validate that what was designed and constructed performs exactly as intended, and we certify that result through a disciplined, transparent, and repeatable methodology. That clarity of purpose is what gives our certification its strength and consistency.

We view commissioning as the last line of defence for an industry where the stakes are rising year-on-year. Commissioning is the point where documentation ends and evidence begins. It is the interface between project delivery and operational excellence. And in that moment, independence and integrity are everything.

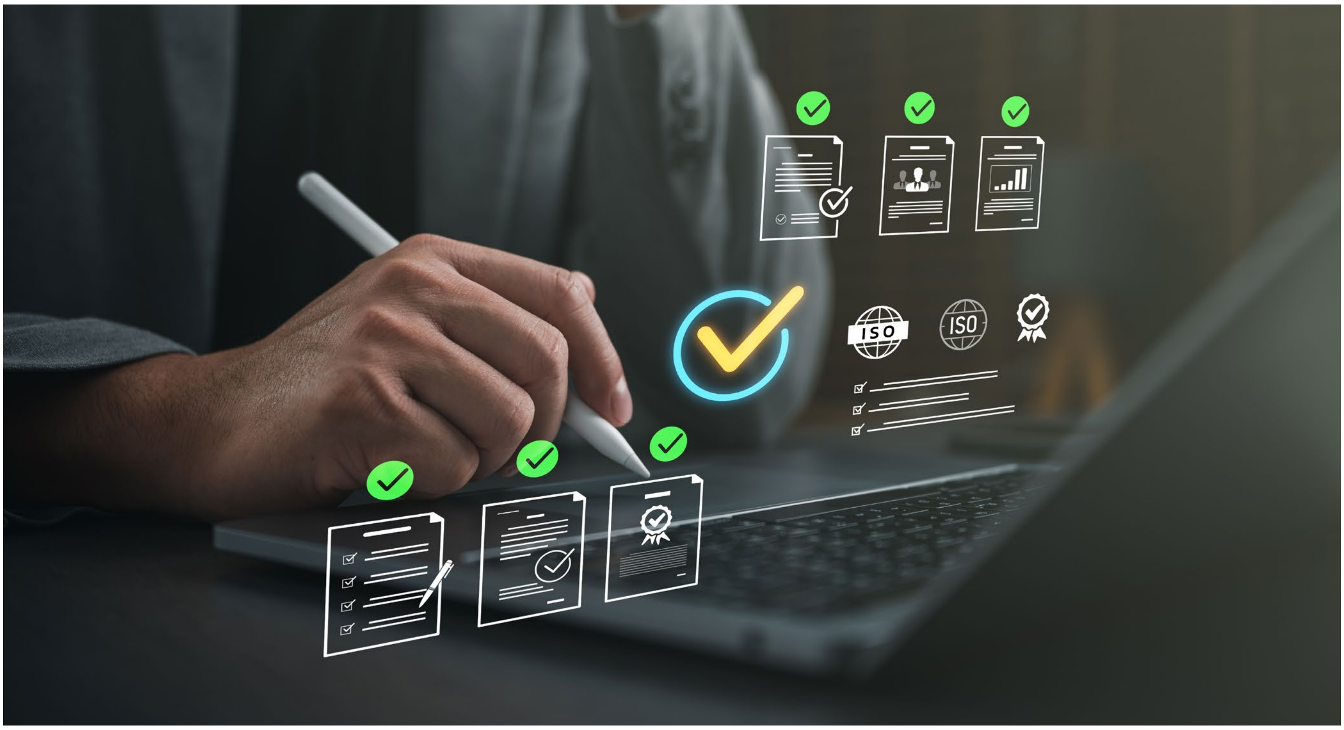
Proper commissioning, certification, and verification are not a luxury. Commissioning is the mechanism t hat ensures design intent survives contact with reality. It is the process that safeguards operational performance long after construction has finished.

Reclaiming the meaning of commissioning

Commissioning should never be reduced to another service line. Another box-ticking exercise. It is the industry's most important form of assurance, the certification that design intent has been achieved, and that the facility performs as expected.

At Global, we are proud to carry the name commissioning, because to us it still means what it always should have: validation, verification, and proof. It represents not a process to be managed, but a standard to be upheld.

As our industry continues to evolve, we all have a choice to make. We can treat commissioning as another box to tick, or we can protect it as the certification discipline that underpins performance, safety, and reliability across the data centre world.



Rethinking energy efficiency in data centres: From infrastructure to insight

As data centres pursue greater efficiency and resilience, attention is shifting beyond core IT and cooling systems toward the wider operational environment. Lighting is emerging as a strategic consideration, influencing not only energy consumption, but also maintenance effectiveness, visibility and long-term operational performance.

BY DAVID GOODCHILD, SENIOR PROJECT MANAGER, HOLOPHANE EUROPE

AS DATA CENTRE demand continues to grow, energy efficiency is becoming a defining factor in how facilities are designed, operated and future-proofed.

Traditionally, optimisation efforts have focused on high-load systems such as cooling and IT infrastructure. While these remain critical, there is increasing recognition that meaningful efficiency gains will come from a more integrated view of the entire environment.

This reflects a broader shift across the sector. Energy efficiency is no longer viewed purely through the lens of power consumption, but increasingly through operational performance, resilience and long-term value.

Supporting infrastructure systems are also receiving greater scrutiny, particularly where operational performance and energy optimisation intersect. Lighting is one example that can have a wider operational impact than energy consumption figures alone might suggest.

Visibility, visual comfort, and glare control all influence how effectively maintenance teams can operate within complex spaces. The requirement for "more light" is often misinterpreted as a need for more lumens, rather than better distribution of light where it is required.

In practice, poorly considered lighting can introduce delays, increase risk and reduce operational efficiency. This has led to greater focus on the quality, direction and control of light, rather than simply the quantity delivered.

The right light, in the right place, at the right time

Traditional warehouse luminaires are typically designed to deliver maximum illumination to the floor plane. Data centre whitespaces, however, often require strong vertical illumination across rack faces and equipment surfaces, not simply higher light levels at floor level.

When lighting distribution is not properly considered, spaces can become over-lit in some areas to compensate for poor visibility elsewhere. This not only reduces visual comfort but can also lead to unnecessary energy use.

Considering optical distribution alongside the efficiency of the light source can significantly influence both lighting system energy requirements and operational performance.

In operational terms, every watt consumed by a lighting system ultimately contributes to the thermal load within the environment. As facilities continue to optimise efficiency wherever possible, these considerations are becoming increasingly relevant.

Connected lighting systems integrated with wider building controls also allow environments to respond dynamically to operational activity. Areas can be illuminated only when required, adjusted based on occupancy patterns, and monitored over time to support ongoing optimisation.

Across the supply chain, there is growing focus on solutions that combine performance, control and longevity.

This includes lighting systems that:

- Deliver consistent illumination across both horizontal and vertical working planes
- Integrate with intelligent building and control platforms
- Support long service life with reduced maintenance intervention

These principles are already being applied in mission-critical environments, where the cost of inefficiency is measured not only in energy consumption, but also in downtime, disruption and operational risk.

Experience from mission-critical environments continues to demonstrate that relatively small infrastructure decisions can have wider operational consequences over the lifecycle of a facility.

Conclusion

Energy efficiency in data centres is increasingly defined by how well systems work together, rather than how they perform in isolation.

For decision-makers, the opportunity lies in adopting solutions that not only reduce energy consumption but also enhance operational performance, support maintenance and deliver long-term value.

As the sector continues to scale, organisations that can translate efficiency strategies into practical operational outcomes will play an important role in shaping the next generation of data centre infrastructure.

Cooling AI responsibly:
A new path to water resilience

New research shows water use by data centers worldwide will more than triple by 2050, impacting water-stressed regions the most. Water scarcity poses a material risk to data center operations and expansion. Addressing the water challenge – through smart location choices, technology and strategic partnerships – is critical in securing long-term growth.

BY MATTHEW PINE, PRESIDENT AND CHIEF EXECUTIVE OFFICER OF XYLEM, A GLOBAL WATER SOLUTIONS COMPANY

WHEN THE dry, hot months arrive, households turn on their taps, putting local water infrastructure under peak pressure. Increasingly, so do their new neighbors: data centers powering a worldwide surge in AI.

These data centers need the most water at exactly the moment when communities can spare the least. For operators, that means managing high-demand cooling periods, meeting uptime targets and navigating stricter local regulations. Around 40% of data centers are already in water-stressed regions. And by 2050, their demand for water will rise by an estimated 272%, according to the Watering the New Economy report published by Global Water Intelligence (GWI) and Xylem.

But there is a way data centers can avoid a trade-off with communities over a shared, essential resource. A way to advance AI innovation and protect people. This path requires a fundamental shift toward reusing more water, plugging leaks and forging partnerships that help municipalities modernize their infrastructure. The report refers to this as a broad-based, global water transition – and data centers are at the heart of it.

AI growth starts with water

The *Watering the New Economy* report gives us a comprehensive assessment of how AI is reshaping global water use, along with a fact-based framework to rethink how we manage and protect the water that AI innovation depends on. We now have a blueprint for transforming water from AI's hidden vulnerability into a source of resilience.

Water underpins the entire AI supercycle, from data centers and semiconductor manufacturing to power generation. Cooling drives almost all the onsite water demand at data centers.

Sector-wide cooling efficiency is improving and will continue to improve by as much as 46% by 2050, according to the report. Keeping servers from overheating is also a far less water-intensive process than fabricating computer chips (which requires filtered, ultrapure water) or generating offsite electricity.

But the reality remains. Aging water infrastructure is struggling to keep up with demand from both communities and a rapidly expanding network of AI data hubs – especially during summer months. Data centers are increasingly being built in dry or rural areas, where land and power are cheap, but water is far from abundant. And unlike fabrication plants, data centers consume (through evaporation) most of the water they draw, with even the most efficient cooling systems discharging just 15% of their withdrawals.

From strain to solutions

These converging risks help explain why giant tech companies – from Amazon to Microsoft, from Alphabet to Meta – are committing to ambitious 'water positive' goals by 2030, undertaking to replenish more water than they consume. There are growing industry calls for tech companies to be 'good neighbors' and to pursue AI expansion in a way that meets community needs.

Pushback from residents can lead companies to relocate, as was the case in Arizona where already-scarce water supplies were threatened by a semiconductor packaging plant. In Chile, protests and political debates have flared up over AI infrastructure expansion.

But states like Arizona also offer a model for the solution. In Chandler, Intel funded a water treatment facility that's achieving 96% water recovery and reuse. Treated water like this can be sent back to communities, agriculture, and neighboring industries.

The treatment plant is owned and operated by a local utility. This partnership allows the City of Chandler municipality to improve public infrastructure, while strengthening the region's water supply chain, which Intel relies on.

A moment to act

Strengthening shared systems also means stopping leaks. According to our report, global wastewater volumes total around 320 trillion liters a year, and up to 100 trillion liters lost to leaking pipes could be recovered – far exceeding the additional water demand the AI supercycle is expected to add in the next quarter century.

In Mexico for example, Amazon has partnered with local utilities to deploy smart water management solutions that use data, analytics, advanced pressure management and real-time leak detection.

Collaboration and targeted investment in intelligent infrastructure deliver resilience and protect community water systems. Operators are exploring innovative air-cooling and water reuse solutions at data centers. Some are even storing rainwater to ease pressure on local water systems.

These efficiency gains help decouple AI expansion from water availability. Another crucial element of the water transition is optimizing the energy mix to support the expected growth of renewable energy. Offsite power generation for data centers and fabrication plants accounts for roughly half (54%) of AI's current footprint. The GWI and Xylem report estimates renewables could save just over 100 trillion liters of water by 2050.

A clear imperative for water leadership

As AI drives the fastest industrial expansion in history, the spotlight has been on its hunger for energy – not its thirst for water. But with AI-driven water demand set to surge nearly 130% over the next 25 years, water security is increasingly becoming a strategic priority.

Leaders across the sector now have the chance to build systems that protect communities and secure long-term capacity. We can choose a path that supports AI innovation and protects water for communities – one built on reuse, efficiency and shared investment.

Ultimately, the new AI economy will require a successful water transition. Handled responsibly, this water transition can become a foundation for data center resilience – supporting uptime, permitting, community trust and long-term performance. The tools exist. What's needed now is joint commitment and cross-sector collaboration.

Rising pressure on data centres from AI is redefining cooling and water management requirements

AI is placing data centres under unprecedented strain. At the heart of the challenge, the demands of this scaling technology are pushing traditional cooling approaches beyond their practical limits. As energy requirements evolve, operators are being forced to rethink how facilities are designed and operated to maintain performance, efficiency and resilience.

BY JON HEALY, REGIONAL STRATEGIC OPERATIONS OFFICER, EMEA, AT SALUTE

LIQUID COOLING technologies are a critical asect of this next phase of growth, offering the efficiency required to support high-density AI environments. However, we're not talking about a simple upgrade here; the process requires careful planning around infrastructure design, water usage, operational processes and workforce capability. Without a comprehensive and coordinated approach from the start, the drive for capacity risks creating unnecessary delays and operational inefficiencies, all while undermining the steps towards a more sustainable industry already put in place.

The challenges with water management

Water, a primary and limited resource, is a major part of the cooling system and is quickly becoming a widely talked-about issue. While proper regulations around this are still developing, investors and customers won't wait and will ask the big questions: how are these big data centres going to be sustainable if they end up using so much water? From the operator's lens, this means respecting community resources, understanding usage levels and trying to recycle wherever possible. If these basic stakeholder concerns go unheard, data centres will be at a huge reputational and operational risk.

Thanks to increased power densities, grid constraints and planning delays, these challenges are becoming more complex. UK regulators, for example, are looking at the resilience and sustainability aspects of newer data centres, and similar pressures exist across the whole of Europe. The lesson here is that operators need to think holistically and beyond their immediate capacity needs. In practice, this means that operators should integrate their cooling strategies with water management from the beginning, and facilities should be designed keeping in mind future workloads and resilience, rather than chasing speed to market and risk harming the environment.

Fulfilling demand for now and the future

Closed-loop cooling systems are gaining a lot more traction, and for the right reasons. They offer a clear-cut path to sustainability as these systems enable water reuse, reduce overall energy demand, and limit environmental impact. Currently, mid-sized operators are constantly trying to get ahead of their competitors and are facing immense pressure to adopt modern cooling systems at speed while ensuring long-term reliability. The solution here is to partner with experienced delivery teams can help strike the balance between speed, cost, and sustainability.

Modular construction is also helping operators keep pace with demand. How? Pre-built modules include liquid cooling from the start, which reduces the need for on-site construction time and disruption to local communities. For example, in areas like the Thames Valley or the Midlands, where planning approvals can be slow, modular approaches can allow operators to deploy capacity at speed without compromising sustainability.

The important part is, modular designs embed sustainability into the build rather than having it as an add-on later. In this day and age, to meet both investor expectations and regulatory changes while still running on schedule, operators need to combine modular deployment with careful water and cooling strategies.

A new benchmark

Water and cooling have become vital foundations for the new data centres that will power the world's growing AI systems. Operators that treat water management and advanced cooling as strategic priorities, not as technical afterthoughts, will strengthen their ability adapt and respond to the increasingly demanding environment around them as it continues to change. This means designing facilities with long-term resilience in mind, embedding efficiency and reuse into cooling infrastructures, and ensuring teams are prepared to manage more complex systems safely and effectively.

The next generation of AI infrastructure will be judged not just on how quickly capacity can be delivered, but on how responsibly it is sustained. Those that combine collaboration and robust execution will be best positioned to support continued AI innovation while balancing performance and social responsibility.

Why liquid and hybrid cooling are powering the AI boom

As compute demands rise and sustainability pressures mount, adopting liquid and hybrid cooling isn't just a technical upgrade, it's a foundational step in enabling the next era of highperformance, sustainable datacentre design.

BY ALISTAIR BARNES, HEAD OF MECHANICAL ENGINEERING AT COLT DATA CENTRE SERVICES

ARTIFICIAL INTELLIGENCE (AI) is driving transformative change across both technology and the wider world. According to Gartner, worldwide spending on AI is forecast to reach $2.52 trillion in 2026, a 44% increase year-on-year.

However, as AI adoption expands, so does the thermal output of the hardware it relies on. This poses a growing challenge for data centre operators, who must ensure efficient thermal management to protect infrastructure and maintain performance, particularly as high-performance computing components, such GPUs generate significantly more heat than conventional IT equipment.

Cooling costs present a major challenge in their own right. The Global Cooling Pledge, introduced at COP28, includes a commitment to reduce coolingrelated emissions by 68% by 2050. This has major implications for data centres, where an estimated 40% of total energy consumption is dedicated to cooling systems.

Traditional airbased cooling is reaching its limits

While conventional aircooling systems were designed for far lower rack densities, today's compute requirements are pushing well beyond what air alone can effectively manage. AI workloads are already driving rack densities that exceed well beyond 100kW.

With overall data centre power demand expected to grow by 165% by 2030, this upward trend is set to continue. As a result, traditional aironly cooling systems simply cannot keep pace.

How liquid cooling drives efficient, sustainable data centres

Liquid cooling has emerged as an effective solution to tackle this challenge. Unlike traditional air-cooling systems, liquid cooling systems use fluids such as water or specialised coolants to absorb and transfer heat away from critical components. With far greater thermal capacity and conductivity than air, liquids can export heat at higher temperatures, enabling heat reuse and reduce the load on cooling systems.

There are several types of liquid cooling technologies in use today. Some systems circulate coolant directly through racks to extract heat from servers, while others use cold plates attached to heat-generating components. Some go further still by fully submerging entire servers in thermally conductive, non-electrically conductive liquid, enabling comprehensive heat management across all components.

By targeting heat at its source, liquid cooling allows data centres to support higher compute densities and operate more efficiently, a critical capability as AI-driven demand for capacity continues to surge.

Striking the right cooling balance

Despite its advantages, liquid cooling is not a complete replacement for air cooling. Even with liquid systems in place, residual heat still escapes into the surrounding environment. As a result, air cooling is still required to maintain optimal conditions within the data hall.

For many data centres, a hybrid cooling strategy which combines liquid and air systems delivers the best results. The ratio of liquid-to-air will differ from customer-to-customer, but this approach ensures high thermal performance while also optimising Power Usage Effectiveness (PUE) and reducing energy consumption across the facility.

Preparing for the next phase

As AI adoption accelerates, the pressure on datacentre infrastructure will continue to intensify. Rising rack densities, expanding compute requirements, and increasing sustainability commitments mean that traditional aircooling systems can no longer support the demands of modern workloads.

Liquid cooling offers the performance, efficiency, and environmental advantages needed to keep pace with this shift. But the future is not purely liquid. For most facilities, the most practical and scalable approach will be hybrid cooling architectures that combine the strengths of both technologies.

Ultimately, the organisations that embrace these nextgeneration cooling strategies will be best positioned to unlock the full potential of AI. As compute demands rise and sustainability pressures mount, adopting liquid and hybrid cooling isn't just a technical upgrade, it's a foundational step in enabling the next era of highperformance, sustainable datacentre design.


«Simply better connected»
R&Mfreenet 4.0 – the structured cabling system for smart, modern Data Centres
Reichle & De-Massari AG
Binzstrasse 32, 8620 Wetzikon
+41 (0)44 933 81 11, hq@rdm.com
www.rdm.com
R&M

Data centre cooling options

Modern data centres require advanced cooling solutions to operate at peak performance.

BY BALTIMORE AIRCOIL COMPANY (BAC)

AS ONE of the most rapidly growing industries in the world, data centers are changing how we develop new cooling technology at a record pace. The data center market is projected to continue growing by 8.5% per year over the next five years, to over $600 billion by 2029.*

Artificial intelligence (AI) and machine learning (ML) are driving increased data center demand, and the numbers are clear—this is only the beginning. Goldman Sachs Research estimates a nearly 200 terawatt-hour per year increase from 2024 to 2030, with AI representing about 19% of data center global power demand by 2028.**

➤ Data center cooling systems

It's safe to say that there is great emphasis placed on infrastructure development to build new and expand existing data centers. To optimize computing power and storage capacity in a smaller space, server rack densities are higher than ever, and the heat emitted requires constant and consistent heat-rejection to prevent overheating and component damage. The cooling system chosen plays a vital role in mitigating these risks while providing options to maximize facility efficiency and Total-power Usage Effectiveness (TUE).

Data center cooling systems are designed using equipment to cool the servers inside the building working together with equipment outside the building to reject the heat.

Cooling technologies inside the data center

There are two primary cooling processes used inside the data center, air cooled and liquid cooled.

Air cooled server racks

Air cooled server racks have been used extensively in data centers for years. They are ideal if you have a standard enterprise IT infrastructure or a storage based system and want a more traditional solution. However, the heat from increasing rack densities is growing beyond their ability to meet the needs of many data centers. There are also hotspots, higher TUE, and a higher carbon footprint than liquid cooling.

Liquid cooling

Liquid cooling can absorb and carry away significantly more heat and is now becoming necessary to handle the increased cooling demands.

Immersion cooling systems are born from remarkable ingenuity, allowing sensitive computer parts to be fully submerged in a non-conductive dielectric fluid that efficiently rejects heat. Immersion cooling tanks are often paired with an evaporative or dry cooler outside the data center to maximize cooling output with many additional benefits.

With the demand for servers in data centers at an all-time high, power consumption and server storage are huge concerns. With exceptional heat dissipation, immersion cooling tanks allow for servers to be stored closer together. Combined with outside cooling units without a chiller in between, TUE is also greatly improved. This system provides consistent cooling with substantial energy reduction, opening pathways to improved high-density computing.

Two liquid cooling alternatives are direct-to-chip cooling and rear door heat exchangers. **Direct-to-chip** cooling circulates cool liquid through a cold plate that contacts heat-generating components like the CPU or GPU. This method, however, is less efficient than immersion cooling and can be challenging to maintain. **Rear door heat exchangers** (RDHx) are another alternative that can add liquid cooling abilities to existing air cooled facilities. The RDHx cools warm air from equipment and expels it out the front or rear door, unlike traditional cooling systems that process air in a separate area of the data center. RDHx systems are less efficient and have limited heat removal capacities compared to immersion cooling.

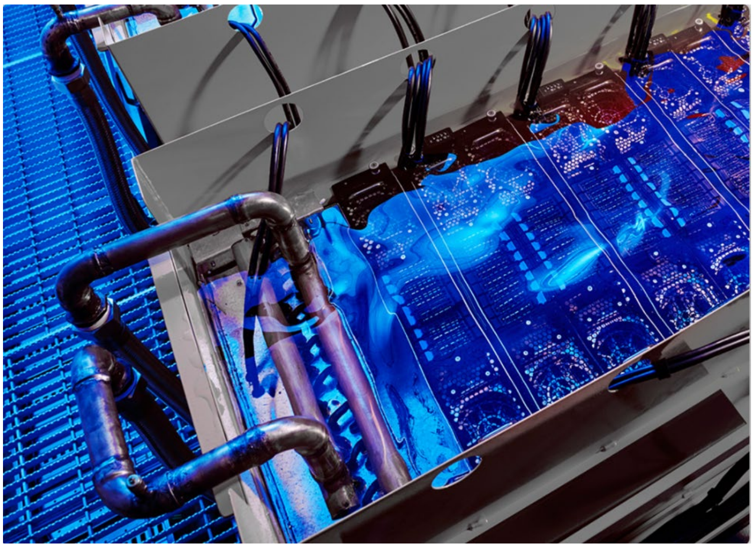
Heat rejection technologies outside the data center

There are also a wide arrange of heat rejection technology options to cool outside of the data center to meet site specific goals.

Dry cooling

Dry coolers are an increasingly common choice for data center cooling because they can effectively lower the inside temperature of facilities without using water. Sites that operate in water stressed areas may choose dry coolers for their industry-leading Water Usage Effectiveness (WUE).

Buildings located in regions with cool ambient air can utilize that outside air to lower internal temperatures, cooling the intended space in an energy-efficient transfer. Dry cooling inherently uses more energy than evaporative cooling, but many dry cooling units are constructed with access to renewable energy, creating a sustainable long-term solution.



(TOP) Liquid Cooling

(BOTTOM) Dry Cooling

Hybrid and adiabatic cooling

Hybrid and adiabatic coolers provide options to balance the trade-off between energy consumption and water usage. With the ability to operate in both dry and evaporative modes, using wet mode during hot design days and switching to dry mode when ambient temperatures drop, these systems optimize water and energy usage based on your facility's needs.

While there are functional differences between the two, both hybrid and adiabatic cooling provide users with the flexibility to reduce water or energy consumption throughout the year. Users facing water stress can still reap the energy-saving benefits of evaporative cooling while using much less total water.

Evaporative cooling

Evaporative cooling systems use the same principle as perspiration to provide cooling. A cooling tower evaporates water over fill and/or coils to efficiently reject heat and discharges warm air from the cooling

➤ **(TOP)** Hybrid and Adiabatic Cooling
➤ **(BOTTOM)** Evaporative Cooling

tower to the atmosphere, and are extremely energy efficient.

Evaporative fluid coolers, or closed circuit cooling towers, maintain a clean, contaminant-free system using two fluid circuits: an external one where spray water mixes with air, and an internal one where process fluid flows through a coil.

For data center sites with access to fresh water, evaporative cooling towers and fluid coolers save energy, allowing more to be directed to servers.

With widespread development of data centers around the world, there are many different environments that require solutions with either less water usage, less energy consumption, or some bridge between the two. Knowing how much water and energy is necessary for cooling is crucial to understand which outside cooling solution suits a data center best.

BAC specializes in cooling for the future, providing a system solution for data centers that combines immersion cooling with dry, adiabatic, hybrid and/or evaporative cooling options. Connect with a BAC representative to learn how you can cool your data center in ways that sustain our world.

REFERENCES

- *"Data Center - Worldwide: Statista Market Forecast." Statista, July 2024, www.statista.com/outlook/tmo/data-center/worldwide
- **"AI Is Poised to Drive 160% Increase in Data Center Power Demand." Goldman Sachs, 14 May 2024, www.goldmansachs.com/insights/articles/AI-poised-to-drive-160-increase-in-power-demand

Data centres striving to balance innovation with sustainability measures

Rather than viewing sustainability and innovation as opposing forces, leading operators now understand that energy efficiency and grid stewardship are integral to long-term success.

BY ADHUM CARTER WOLDE-LULE – DIRECTOR AT PRISM POWER GROUP

CRITICS ARGUE that data centres have a significant carbon footprint and could derail climate goals. However, others contend that relative to global energy use, data centre power draw is still modest (roughly a few percent) and must be weighed against benefits – for instance, AI services potentially delivering efficiency improvements elsewhere.

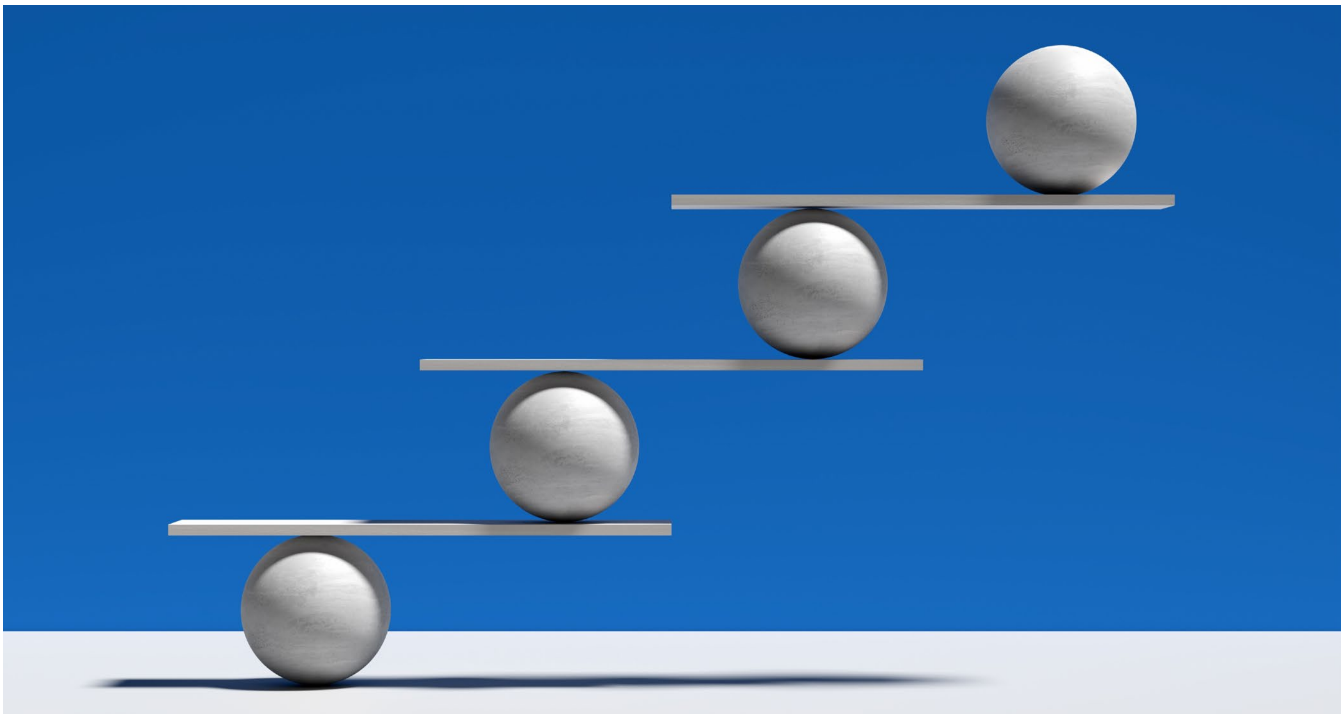
It's never that black and white, of course, but in response to sustainability critiques and grid challenges, the data centre industry is rapidly innovating to improve efficiency and green its power sources.

Major operators are adopting advanced designs to squeeze more work out of every watt. For example, cloud giants have invested in next-generation cooling: Microsoft is rolling out liquid immersion and evaporative cooling techniques to cut both energy and water usage, alongside AI-driven software to optimise server workloads.

Amazon Web Services has redesigned its facilities with high-efficiency proprietary hardware and real-time adaptive cooling, achieving industry-leading power usage effectiveness (PUE) and water usage effectiveness.

Data centre providers like Digital Realty report a broad shift from traditional air cooling to liquid cooling systems, which can support denser computing with lower power overhead – over half of Digital Realty's global sites and counting now utilise liquid cooling.

Beyond cooling, companies are also sourcing cleaner energy to run their servers. Tech firms have signed massive long-term renewable energy contracts and even begun exploring small-scale nuclear power agreements to ensure their data centres are powered by low-carbon sources.



Such deals can stimulate new wind and solar projects (even if the local grid mix remains dirty) and have enabled operators to claim carbon-neutral operations. Industry-wide, there is a trend toward integrating on-site generation and storage – co-locating solar panels, fuel cells, or battery banks with data centres to supply backup power and reduce peak grid draw.

Moreover, many experts see data centres as part of the solution to energy issues: their large, round-the-clock demand can provide a stable revenue base for new renewable projects, while their flexibility in timing certain computing tasks means they can adjust load to support grid stability.

Indeed, researchers note that AI and smart software are now helping to cut data centre energy waste (Google, for instance, used AI to slash cooling power by 40% in its facilities). The International Energy Agency argues that if deployed wisely, AI and data centres could save more emissions in other sectors (through efficiency gains) than they produce, highlighting examples from improved grid management to optimising industrial processes.

However, realising these benefits at scale requires continued investment and supportive policy – a point acknowledged even by optimistic reports. The balance between digital growth and sustainability is thus driving continuous innovation in data centre design, power sourcing and grid integration.

As a Director at Prism Power, I have witnessed first-hand how data centre energy strategy has evolved from a background concern to a boardroom priority. The criticisms raised in global media are not unfounded - communities and regulators are rightfully anxious about the carbon footprint and grid impact of large data centres. We acknowledge, for example, the reports of steeply rising energy usage and instances where data centre clusters taxed local grids.

These challenges underscore why our industry must be proactive. At Prism Power, our focus has always been on sustainable, efficient power solutions tailored to clients' needs. We've been partnering with data centre operators to implement many of the innovations discussed above: from intelligent power distribution systems that improve energy efficiency, to integrating on-site renewables and battery storage for peak shaving and backup.

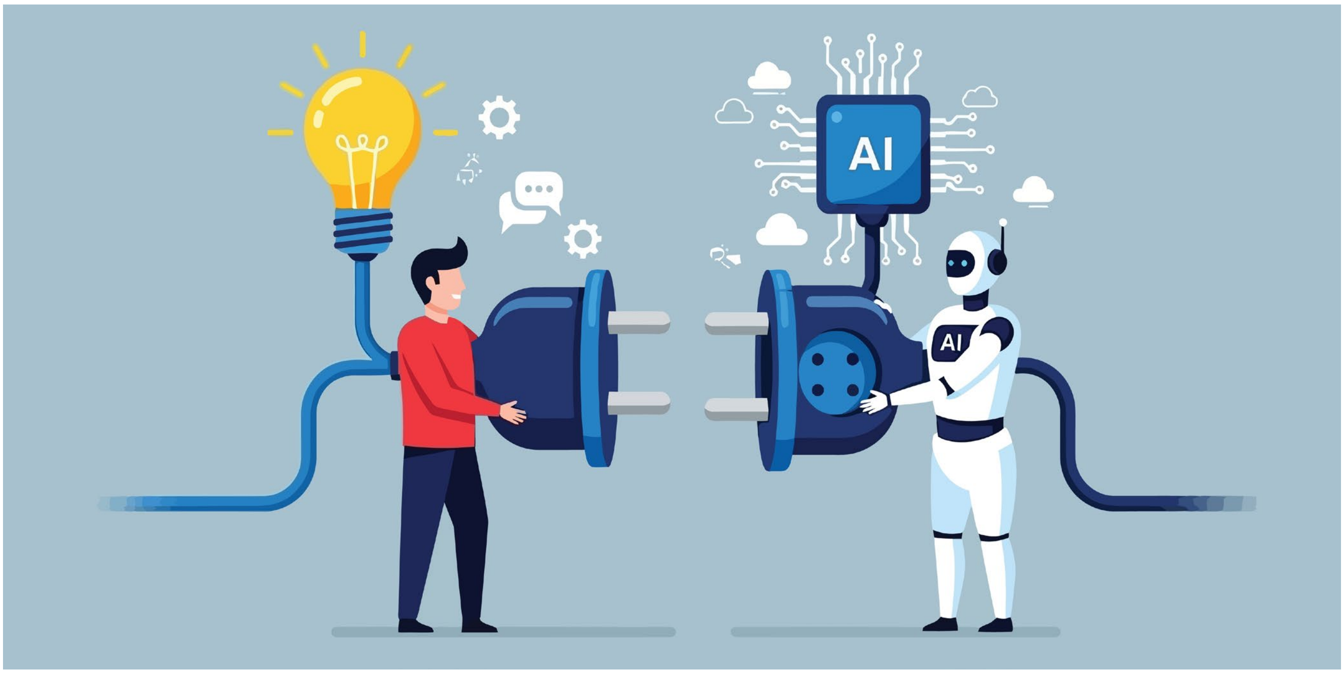
Such steps not only reduce environmental impact but also make facilities more resilient and grid-friendly. Plus, it's heartening to see major tech companies invest in greener practices - like the shift to liquid cooling, AI-driven optimisation, plus 24/7 renewable energy sourcing is accelerating. These innovations prove that the industry can adapt. Yet, we also recognise that progress must continue.

Addressing sustainability concerns is an ongoing journey: improving transparency in emissions reporting, utilising waste heat and recycled water, and working closely with utilities on demand-response are all areas of active development.

As an industry, we are moving toward a model where data centres support the energy transition – for instance, by scheduling non-urgent computing tasks to coincide with renewable output or by providing grid services through energy storage and smart controls.

At Prism Power, we see it as our responsibility to help bridge the gap between digital growth and sustainable infrastructure. By designing and delivering power systems that optimise efficiency, reduce waste and incorporate green energy, we aim to ensure that the next generation of data centres can meet booming demand without compromising on climate goals or grid reliability.

The conversation has shifted: rather than viewing sustainability and innovation as opposing forces, leading operators now understand that energy efficiency and grid stewardship are integral to long-term success.



Powering the future responsibly

Driving data centre growth with sustainability at the core.

BY DARREN ELLIS, MANAGING DIRECTOR, EMEA, PARK PLACE TECHNOLOGIES

WHEN THE UK government classified data centres as Critical National Infrastructure (CNI) in 2015, it acknowledged their dual role as both an economic asset and a national dependency. What was once a specialist IT domain has become the unseen backbone of modern society, powering innovation across industries and sustaining essential services.

As demand for digital services increases, data centre capacity is expanding at speed, driving a rise in energy consumption. UK data centres currently consume approximately 5 TWh of electricity each year, with demand expected to grow fivefold by 2030. *[NIA Report – Powering the Data Centre Book – December 2025]* Therefore, ensuring this growth is sustainable is crucial to safeguard the UK's economic competitiveness and the strength of its digital infrastructure.

Innovation in action: driving greener, more efficient data centres

As demand for data centres continues to grow, technology is playing a vital role in lowering their environmental impact. Far from being simple server storage facilities, modern data centres are sophisticated tech hubs that deliver immense computing power while minimising energy waste. Leading this transformation are advanced cooling solutions. Conventional air conditioning is increasingly being replaced by liquid cooling, immersion systems and free-cooling techniques that leverage outside temperatures to reduce energy use. These innovations enable data centres to operate at peak performance without depending solely on energy-intensive chillers. This reduces both electricity consumption and carbon emissions.

Beyond cooling, AI-powered workload management systems help operators to optimise computing output while enhancing Power Usage Effectiveness (PUE), the standard measure of a data centre's energy efficiency. By smartly distributing workloads across server clusters and dynamically adjusting performance based on demand, data centres can operate nearer to their ideal PUE, cutting unnecessary energy use.

Energy management platforms further boost efficiency by incorporating renewable energy sources and predictive grid load balancing. Instead of operating as passive consumers of electricity, modern data centres can actively control when, where and how they use power. By combining real-time analytics with long-term forecasting, these platforms allow operators to shift workloads to times of higher renewable availability, lowering dependence on carbon-heavy grid peaks and storing surplus clean energy for later use.

Smarter design further amplifies these gains. Data centres are increasingly sited and built with sustainability in mind. Modular layouts allow incremental scaling to minimise over-provisioning, while materials and layouts are optimised to reduce heat retention. Heat recovery systems can capture surplus energy and redirect it to nearby industrial or residential facilities. Across the UK, several data centres already heat local swimming pools, showing how waste energy can become a valuable community resource.

Shaping growth: policy and planning for sustainable data centres

As data centre expansion continues, policy and planning has never been more important. While rapid expansion is necessary to meet the escalating demand, unchecked growth risks environmental strain, community pushback and grid instability.

Planning frameworks that incentivise energy-efficient designs encourage developers to adopt best practices from the start. Clear, consistent guidelines reduce uncertainty for operators while protecting local communities and ecosystems. Policy reforms can simplify approvals for low-impact, sustainable projects, easing bottlenecks while maintaining environmental standards.

Incentives for facilities that integrate with renewable energy grids, recover waste heat or optimise PUE metrics encourage environmentally intelligent design. Similarly, national strategies that align data centre growth with grid capacity and urban planning priorities help avoid over-concentration.

As data centre expansion continues, policy and planning has never been more important. While rapid expansion is necessary to meet the escalating demand, unchecked growth risks environmental strain, community pushback and grid instability

Effective policy and planning reforms are not intended to restrict growth but to guide it responsibly. By incorporating sustainability, efficiency and community engagement into regulatory frameworks, the UK can foster the data centre expansion it requires while protecting environmental and social priorities.

Responsible growth: powering digital progress while protecting the environment

Balanced growth in the data centre sector means accepting that digital demand and environmental stewardship are interdependent priorities that must progress together. As societies rely more heavily on digital services, the need for additional capacity is unavoidable, so the way that capacity is delivered must continue to improve.

Building the next generation of capacity requires energy efficiency to be integrated from the start, through smarter design, low-carbon power sourcing and more efficient cooling and hardware strategies. Achieving balanced growth also calls for increased transparency and accountability, with clear reporting on energy use, water consumption and carbon emissions. The sector must continue to adopt circular practices, extending equipment lifecycles, minimising waste and leveraging opportunities such as heat reuse and resource recovery.

Balanced growth is not about limiting digital progress. It is about ensuring that the infrastructure underpinning modern life scales responsibly, supporting economic innovation while continuously strengthening environmental resilience for the long term.

Making data centres fit for BFSI: speed, security, and sovereignty

As banking and financial services infrastructure evolves, data centres are being pushed to deliver more than capacity and uptime. Speed, security and data sovereignty are reshaping how operators design, monitor and manage the physical layer.

SOPHIA BOROWKA, PHD, PRODUCT MANAGER DIGITAL SOLUTION, REICHLE & DE-MASSARI

FOR banking, financial services and insurance organisations, a data centre is more than a place to host applications. It is a foundation for trust, resilience, and competitive performance. In this sector, infrastructure must support money-moving decisions, sensitive customer records, regulatory scrutiny, and increasingly real-time digital services.

A design that is "good enough" for a general enterprise environment may not be sufficient for BFSI. Three key requirements define the challenge: speed and consistency, customer data protection, and data sovereignty. Together, they are forcing financial institutions and their infrastructure partners to look more closely at the physical layer: cabling, patching, monitoring, documentation, and site design. In BFSI, every route, connection and change can affect performance, auditability, and risk.

Requirement 1: latency

In high-frequency trading and exchange hosting, microseconds can influence outcomes. For retail and corporate banking, latency has a different but equally important role. Real-time fraud scoring, instant payments, risk checks and digital customer channels all depend on fast, predictable processing. Poor performance can have material consequences from delayed decisions, failed transactions, weaker customer experience and operational instability.

For this reason, BFSI data centres increasingly treat latency as a design constraint rather than a simple performance metric. The focus is not only on reducing average delay, but also on controlling jitter and tail latency. Critical systems need short, repeatable paths, fewer hops, and carefully managed interconnection points. In trading environments, even cross-connect choices can become a matter of fairness and control.

This places new importance on structured cabling. High-density fibre, pre-terminated connectivity, disciplined labelling, and controlled patching all help reduce variation and human error. Physical layouts must be latency-aware, with key systems placed close to switches, meet-me rooms, or distribution frames where appropriate. Separate fabrics or distinct routing paths may also be required for different traffic classes, helping to reduce unpredictable behaviour and support resilience.

Monitoring must be equally precise. BFSI operators need continuous visibility into latency, jitter, packet loss and time synchronisation health. In capital markets, accurate event sequencing and traceable timestamps are also part of the compliance picture.

Monitoring architectures must therefore provide evidence without disturbing the critical path. Passive visibility tools, such as optical taps, can support observability without introducing latency, provided optical loss is properly budgeted.

Requirement 2: customer data safety

Financial institutions process identity documents, transaction records, beneficial ownership information, sanctions screening outputs, insurance data and, in some cases, biometric or health-related information. This makes them high-value targets for cybercrime, fraud, and ransomware. The data centre must therefore support not only security, but proof of security.

In practice, this means strong physical segregation, strict privileged access control, pervasive encryption, immutable logging, and detailed audit trails. The challenge is complicated by competing lifecycle requirements. Data protection rules push firms towards minimisation and defensible deletion, while anti-money laundering and counter-terrorist financing obligations often require long-term retention, fast retrieval, and legal holds. Data centre platforms must support both obligations without creating gaps in control.

At the physical layer, this turns cabling and rack infrastructure into governed assets. Moves, additions and changes must be documented, authorised and traceable. A poorly recorded patching change is no longer just an operational inconvenience; it can become an audit weakness. This is why automated infrastructure management and digital twin technology are becoming more relevant in BFSI environments. They provide a single source of truth for connectivity, assets, and changes, helping teams prove what is connected, where it is located and who changed it.

Requirement 3: sovereignty

For European BFSI firms, keeping customer data on-premise or in-country is not always legally mandatory, but it often reduces risk. Localised infrastructure can simplify supervisory access, audit rights, operational control and exit planning. It can also reduce the complexity associated with international data transfers and remote administration from other regions.

This is driving demand for clearly defined in-country capacity, local resilience, and transparent operational boundaries. Data centres serving BFSI customers may need dedicated cages, controlled meet-me room cross-connects, rack-level segregation and country-local carrier paths. Management access, support operations and traffic egress must also be controlled. It is not enough to know where the servers are; operators must understand where management sessions terminate, where data can flow and how changes are evidenced.

The result is a more demanding design brief. BFSI data centres must be fast, secure, auditable, and adaptable. They need deterministic physical paths, high-density connectivity, resilient A/B architectures, real-time monitoring and documentation that stands up to inspection. They also need the ability to expand or upgrade without disrupting critical services.

This is where integrated physical-layer design becomes essential. These requirements can be addressed through high-density fibre and copper connectivity, modular distribution, pre-terminated systems, controlled patching platforms and creating digital twins. Solutions should support dense, organised patching with labelling and monitoring options. Passive TAP-based visibility can enable monitoring without adding latency. For distributed or edge environments handling sensitive workloads, secure rack and monitoring options can extend control beyond the core facility.

For new builds, this approach helps create a coherent architecture from the meet-me room and optical distribution frame through to rack-level connectivity. For existing facilities, it can support brownfield optimisation: consolidating patch fields, improving cable routing, introducing pre-terminated trunks with precisely modelled lengths, standardising labelling, and adding real-time port visibility. These improvements reduce operational risk while making infrastructure easier to scale and audit.

Supporting confidence

BFSI data centres are entering a period in which physical infrastructure will be judged not only by capacity and uptime, but by its ability to support regulatory confidence, customer trust and business speed. The organisations that succeed will be those that treat the physical layer as a strategic control point. In financial services, the right connection is no longer simply a cable between two ports. It is part of the evidence chain that proves performance, protection, and sovereignty.

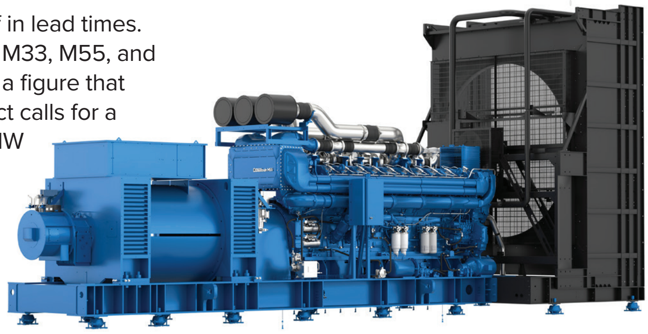
Speed, scale, reliability: How Baudouin is rewriting the rules for data centre power

As hyperscale data centre projects accelerate, every element of the backup power supply chain is under pressure to deliver at speed and scale. Baudouin's fully integrated French manufacturing operation combines rapid lead times, high-density performance, and mission-critical reliability for the demands of modern AI infrastructure.

DATA CENTRE projects move fast. Hyperscale campuses are breaking ground at an unprecedented pace, and the pressure on every link in the supply chain, including backup power, is relentless. Delivery windows are tight, and the entire project timeline depends on every component arriving on schedule.

It's in this environment that the question of where and how a genset is built stops being a footnote and becomes a deciding factor.

Baudouin's answer is unambiguous: design it in France, build it in France, and control every step in between. Three manufacturing sites, 35,000 m^2 of production space, and a fully integrated operation from engine manufacture through to genset assembly. A single industrial footprint, a single chain of accountability with full traceability and quality control across the entire range.



That structure pays off in lead times. Six months across the M33, M55, and M61 genset platforms, a figure that holds whether a project calls for a single unit or a multi-MW deployment across several locations. With over 190 MW of installed genset capacity across the EMEA region as of 2025, Baudouin has already delivered at the kind of scale hyperscale demands.

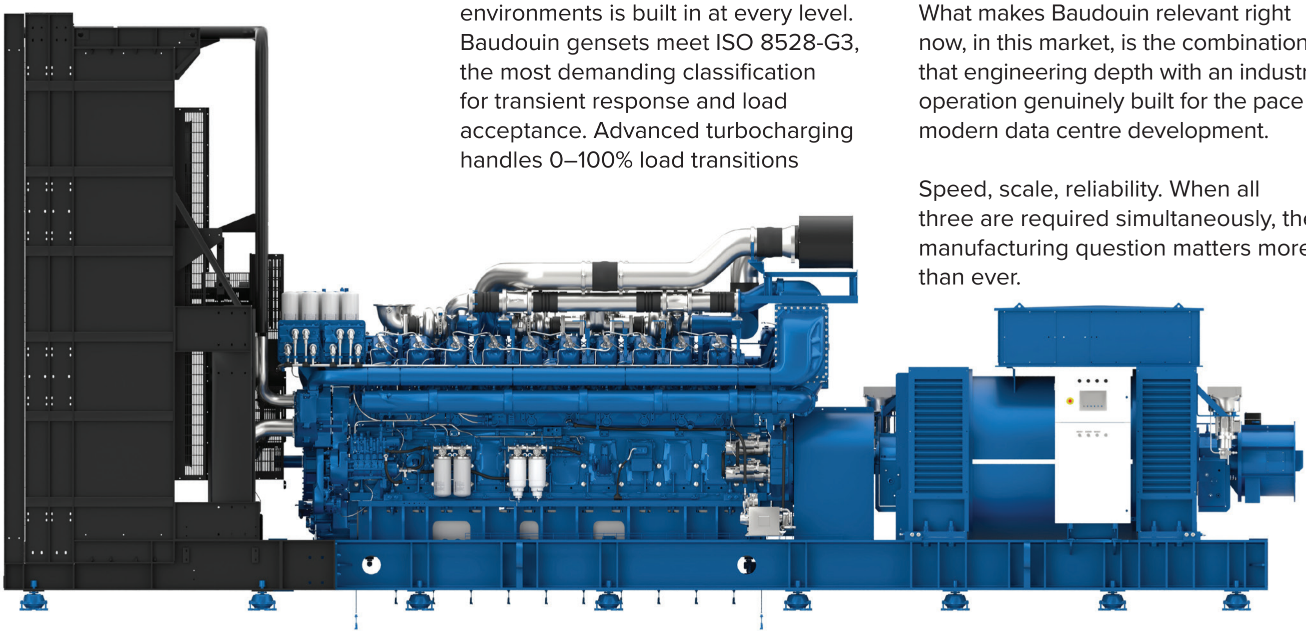
The platform range covers ground from 1 MW to the top of the market. The 20M55 and the latest released 20M61 (launched in 2025) push output to 6,250 kVA, currently the highest power rating available for data centre gensets. Power density sits at 45.8 kWe/L on the 20M55, and footprint density reaches 100 kWe/m^2, which matters when every square metre on a densely packed campus carries a real cost.

Performance in mission-critical environments is built in at every level. Baudouin gensets meet ISO 8528-G3, the most demanding classification for transient response and load acceptance. Advanced turbocharging handles 0–100% load transitions seamlessly, and dual starting motor redundancy guarantees start-up reliability when it matters most.

Compliance comes standard. CE marking, Uptime Institute pre-approval wwwfor Tier III and Tier IV applications, and full compatibility with 100% HVO fuel. For operators working against tightening sustainability targets, that last point is increasingly a procurement requirement and Baudouin meets it with zero performance compromise.

More than a century of standby power engineering sits behind these products. What makes Baudouin relevant right now, in this market, is the combination of that engineering depth with an industrial operation genuinely built for the pace of modern data centre development.

Speed, scale, reliability. When all three are required simultaneously, the manufacturing question matters more than ever.



Uptime Resilience

Network Security

IoT Monitoring

TRUSTED POWER CONTROL FOR MISSION CRITICAL DATA CENTRES

DSE335 MKII

Auto Transfer Switch Control Module

Reliable power control is essential in critical operations where downtime is not an option. Whether in data centres, hospitals, or industrial environments, even the slightest power disruption can lead to serious safety, financial, or operational consequences. Robust, intelligent control systems ensure continuous performance, system protection, and complete confidence in essential infrastructure reliability.

The DSE**335 MKII** is an advanced auto transfer switch (ATS) controller designed for seamless power switching between two different power sources (S1/S2). Ideal for mission-critical applications, it offers configurable inputs/outputs, comprehensive monitoring, and a close transition feature for no-break power transfer. Its robust design and flexible programming make it a trusted choice for reliable power continuity.

To find out more please visit **www.deepseaelectronics.com**

Connectors and connectivity in data centre and enterprise environments

As AI, hyperscale and enterprise networks scale in speed and density, connectivity infrastructure is emerging as a critical enabler of performance, resilience and future-ready digital environments.

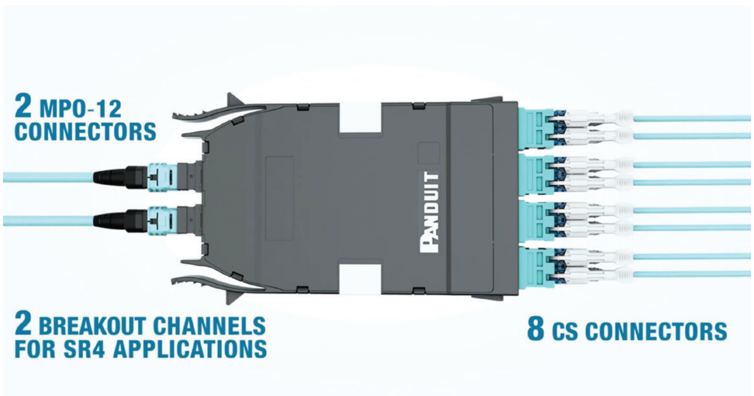
BY HANS OBERMILLACHER, PANDUIT

AS AI, high-performance computing, edge processing and IoT adoption continue to expand, the physical foundation of digital infrastructure is becoming increasingly critical. While much of the industry focus remains on processors, optical technologies and switching platforms, long-term performance depends just as heavily on the cabling, connectivity systems and enclosure architectures that support them.

Across enterprise campuses, hyperscale facilities and data centre environments, the move toward higher bandwidth, greater rack density and lower power consumption is forcing a rethink of network design. This means building structured, scalable and manageable infrastructure capable of supporting 400G, 800G and future 1.6T networks while maintaining reliability, operational simplicity and thermal resilience.

Fibre connectivity in AI-Driven data centres

AI training clusters and inference environments are driving unprecedented east-west traffic. Modern GPU-based servers no longer rely on traditional duplex fibre links. Instead, they commonly deploy multi-fibre parallel optics using high-density multi-fibre connectors terminated in modular cassette and panel systems.



➤ Panduit MPO Connectivity.

AI GPU servers typically utilise eight-fibre connections, while 800G channels operate over 16 fibres (eight duplex lanes). This represents at least a fourfold increase in fibre count compared to legacy duplex architectures, and in practical AI pod deployments, overall fibre density can increase by up to eight times.

This escalation in fibre count places enormous demands on connectivity hardware. High-density fibre trunks, precision-terminated multi-fibre connectors, and modular patching frames become essential to maintain order within increasingly constrained white space environments. Modern panel systems are engineered to maximise port density per rack unit while preserving bend radius control and front-access serviceability, allowing operators to scale fibre counts without expanding footprint.

Large AI clusters often consist of tightly coupled "super pods" which can span up to 50 metres. Within these architectures, factory-terminated trunk assemblies and modular cassette-based - distribution provide rapid deployment, consistent performance and simplified migration.

Addressing latency concerns

A common misconception is that structured cabling introduces additional latency due to extra connection points. In AI super pod architectures where link lengths are typically under 50 metres, light propagation delay is approximately 250 nanoseconds. In comparison, forward error correction (FEC), frame buffering and queuing delays contribute significantly more latency.

IEEE channel specifications allocate connector loss budgets of approximately 1.5 dB for multimode fibre and 2.5 dB for single-mode fibre. When connectors are manufactured to tight geometric tolerances, tested for insertion loss and return loss compliance, and maintained through disciplined cleaning procedures, structured connectivity does not compromise high-speed, or AI workload latency.

This shifts the design conversation toward performance assurance and lifecycle reliability. Precision-engineered ferrules, low-loss alignment sleeves and secure latching mechanisms ensure repeatable optical performance even in high-frequency MAC environments. Proper cable management accessories further protect connector interfaces from stress and contamination, safeguarding long-term link integrity.

Structured cabling versus point-to-point architectures

As AI clusters scale to thousands of GPUs, the debate between structured cabling and direct point-to-point (P2P) connectivity has intensified. While P2P may appear simpler at small scale, complexity escalates rapidly in high-density GPU environments.

Structured connectivity introduces modular patch panels, labelled adapter fields, high-count fibre trunks and horizontal and vertical cable managers designed to separate trunk and patch pathways. This modularity enables: simplified deployment and faster installation, clear traceability and documentation, reduced mean time to repair (MTTR), and minimal disruption during hardware refresh cycles

Hardware refresh in AI environments may occur every 12–18 months. A structured backbone allows servers to be replaced by reconfiguring short patch leads without disturbing core trunks, protecting capital investment and reducing downtime.

Rail-optimised topologies require deterministic mapping between GPU ports and leaf switches. Structured connectivity systems with colour-coded patching fields, port identification and physical separation of network domains make this deterministic mapping achievable and repeatable.

Managing density, pathways and airflow

High-density fibre infrastructures can overwhelm containment pathways if not carefully designed. Consolidating multiple fibre runs into high-count trunks can reduce pathway utilisation by up to 70%.

Purpose-built fibre routing systems, overhead pathways and rack-integrated management fingers help segregate power and data while maintaining minimum bend radius compliance. Properly designed connectivity frames support front and rear access, sliding trays and modular growth without disturbing adjacent live circuits.

Reducing congestion has measurable consequences:

- Improved airflow and thermal efficiency
- Lower cooling costs
- Reduced mechanical strain on cables
- Enhanced serviceability

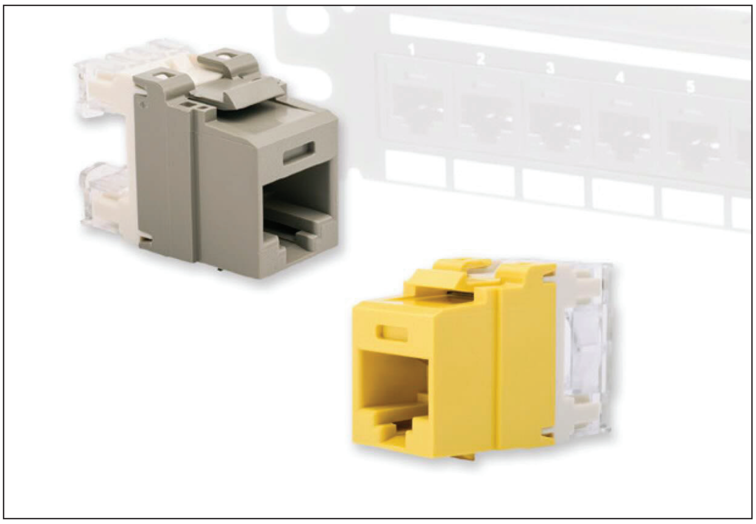
➤ High density structured cab.

As AI workloads demand greater rack power densities, airflow optimisation becomes inseparable from connectivity planning. Cable management is no longer cosmetic; it directly influences energy performance and reliability.

Copper cabling in enterprise and data centres

While fibre dominates high-speed AI interconnects, copper cabling remains indispensable across enterprise and data centre environments. Its cost efficiency, Power over Ethernet (PoE) capability and ease of termination ensure continued relevance.

Category 6A copper cabling supports 10Gbps over 100 metres and remains widely deployed in enterprise networks. High-performance modular



➤ Cat 6 Keystone.

jacks, patch panels and shielded connectivity options provide consistent electrical performance while supporting high-density installations.

In intelligent buildings, copper connectivity extends beyond data transport. Robust field-terminated plugs, high-cycle modular interfaces and zone enclosure solutions support flexible reconfiguration of workspaces, access points and building automation systems.

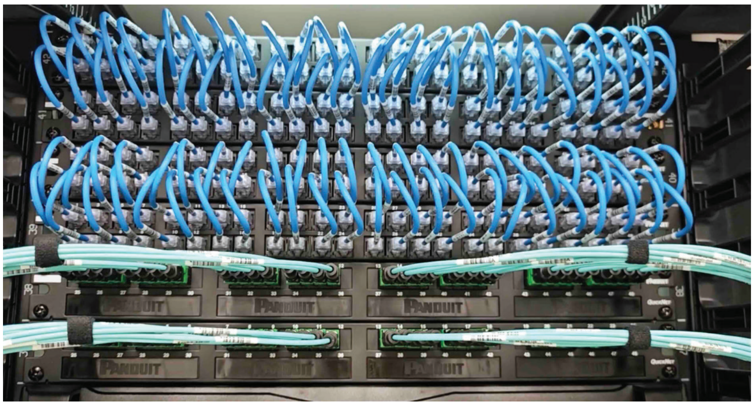
One of copper’s defining advantages is PoE compatibility. IEEE 802.3bt supports power delivery up to 100 watts, enabling: wireless access points, security cameras, LED lighting systems, IoT sensors and smart building controls.

Advanced copper connectors designed for high mating cycles and secure retention ensure reliable power and data delivery. Enhanced conductor designs and improved thermal dissipation characteristics allow cable bundles to sustain higher PoE loads.

Copper in data centre architectures

Copper plays a significant role within data centres, particularly for management networks and short-reach interconnects. Category 8 cabling supports 25G and 40G Ethernet over 30 metres, aligning with top-of-rack and end-of-row architectures.

➤ Quick Deployment Connectivity.



In GPU deployments, thousands of copper patch leads may be required solely for management connectivity. Structured copper patch panels with angled port orientation, integrated strain relief and cable segregation channels prevent disorganisation and simplify troubleshooting. Shielded systems further enhance electromagnetic compatibility in high-power rack environments, protecting sensitive management traffic from interference generated by high-current equipment.

Sustainability and energy considerations

Sustainability is increasingly central to infrastructure planning. Copper cabling offers recyclability and lifecycle transparency, supported by Environmental Product Declarations (EPDs). Within AI data centres, multimode fibre transceivers may consume up to 15% less power than single-mode equivalents for short-reach applications.

Emerging DC power distribution concepts using PoE-based architectures may reduce conversion losses that can approach 20% in traditional AC-to-DC systems. Connectivity design is increasingly intertwined with power architecture, making cable gauge, bundle design and thermal modelling strategic decisions rather than secondary considerations.

Next steps

The roadmap toward 1.6Tb/s and higher speeds will continue to reshape connector strategies. Parallel optics, higher fibre counts and advanced modulation schemes are enabling greater bandwidth over existing fibre infrastructure. Smaller form-factor connectors, ultra-high-density patching platforms and modular trunk systems are increasing fibre counts per rack unit while preserving manageability. Infrastructure deployed today must therefore accommodate higher fibre counts and modular upgrades without wholesale replacement.

In both enterprise and AI-driven environments, connectors are no longer passive components. They are performance enablers, density managers, airflow protectors and sustainability contributors. Connectivity architecture has become a strategic asset, forming the resilient, scalable foundation upon which next-generation digital infrastructure depends.

One of copper’s defining advantages is PoE compatibility. IEEE 802.3bt supports power delivery up to 100 watts, enabling: wireless access points, security cameras, LED lighting systems, IoT sensors and smart building controls.


AFL

Why Nickel-Zinc belongs in Europe's next retrofit discussion

As European data centres face tighter regulations, sustainability demands and ageing VRLA infrastructure, nickel-zinc technology is emerging as a practical retrofit path that reduces deployment risk, lowers lifecycle costs and avoids major facility modifications.

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

EUROPE'S DATA CENTRE operators know what they're moving away from: ageing VRLA systems with short replacement cycles and rising infrastructure overhead. ZincFive's nickel-zinc retrofit kit gives them a clear path forward, a drop-in replacement that aligns with EU Battery Regulation requirements, avoids the supply chain exposure of lithium-ion, and delivers a 15-year service life within existing cabinet and open rack infrastructure.

For European data centre operators approaching a VRLA replacement cycle, battery selection is increasingly defined by regulatory, sustainability, and retrofit constraints within existing physical architectures, including open rack designs. Lithium-ion was the assumed successor to lead-acid. That assumption is under pressure.

A regulatory shift with real procurement implications

The IEA's Global Critical Minerals Outlook 2025 identifies lithium among the supply chains most exposed to geopolitical concentration risk, a concern echoed in the European Parliament Research Service's 2025 brief on the EU's battery industry. The EU Battery Regulation (2023/1542), now in effect, goes further, requiring a digital battery passport that validates carbon footprint and sustainability metrics through accredited third parties and makes them publicly available. Compliance responsibility extends to operators and integrators, not manufacturers alone.

The challenge compounds in brownfield environments. More than 70% of global data centre capacity sits in existing facilities, and in European markets, permitting timelines, grid access constraints, and sustainability mandates are pushing operators to optimise what they have. Retrofits can cost 30% to 50% less than new construction and avoid much of the delay tied to greenfield development. Upgrading to lithium-ion in an existing facility, though, can require enclosure changes, fire suppression additions, and compliance reviews that expand project scope and erode the cost and timing advantages that made retrofit attractive in the first place.

A drop-In answer to a structural problem

ZincFive's nickel-zinc (NiZn) technology resolves the question on both fronts. The NiZn Retrofit Kit fits directly into existing cabinet and open rack footprints, enabling a chemistry transition without enclosure changes, room redesigns, or fire suppression

upgrades. The batteries did not exhibit thermal runaway in UL 9540A testing and align with the same regulatory framework as VRLA. By enabling right-sized backup power, the kit allows operators to deploy only the capacity they need, avoiding the hidden costs of oversized lead-acid systems.

On the sustainability side, the chemistry aligns with where EU regulation is heading. ZincFive's batteries are designed to meet EU Battery Regulation requirements, with 100% of the nickel and zinc recyclable for reuse, avoiding the critical mineral dependencies drawing scrutiny from European policymakers. For operators under ESG reporting pressure, this is not a peripheral consideration. Procurement teams in European markets are now asking about this specification.

The lifecycle economics reinforce the case. Rather than resetting on another five-to-seven-year VRLA replacement cycle, ZincFive's retrofit kit carries a 10-year warranty and a 15-year service life. As labour costs increase across Europe, fewer replacement events translate directly into lower operational expenditure and less disruption. Avoided cascading costs from facility modifications, thermal management upgrades, and repeated maintenance windows compound across multi-system environments into a materially lower total cost of ownership (TCO).

ZincFive engineered the retrofit kit around a flexible shelf architecture compatible with legacy VRLA enclosures and open rack configurations, giving service providers and integrators a turnkey, repeatable model that scales across sites without custom engineering at each one.

With three times the power density of VRLA, the retrofit kit reduces physical footprint while delivering the high-rate performance modern UPS environments demand. In many cases, operators can eliminate a cabinet or battery string altogether, reclaiming floor space without reengineering surrounding infrastructure. The kit uses the same nickel-zinc cell technology and monobloc batteries as ZincFive's BC Series systems, a foundation already proven in production environments. Operators are not evaluating chemistry alone. They are evaluating execution risk, deployment fit, and real-world performance.

Conclusion

The next VRLA replacement cycle is no longer a routine maintenance event. In Europe's evolving regulatory and infrastructure environment, operators are being forced to evaluate not only battery performance, but deployment risk, lifecycle economics, sustainability alignment, and retrofit feasibility within existing architectures. The next VRLA replacement cycle is a strategic inflection point for European data centres. In a market increasingly shaped by retrofit constraints, sustainability mandates, and infrastructure efficiency, nickel-zinc technology enables operators to upgrade chemistry without upgrading infrastructure.

"The next VRLA replacement cycle is no longer a routine maintenance event. In Europe's evolving regulatory and infrastructure environment, operators are being forced to evaluate not only battery performance, but deployment risk, lifecycle economics, sustainability alignment, and retrofit feasibility within existing architectures"

From training-centric to inference-driven infrastructure

AI infrastructure is moving beyond uniform GPU training clusters. While earlier deployments focused on dense, synchronous training fabrics, current fiber deployments for AI and cloud workloads are considerably more varied. Inference now accounts for the majority of AI compute demand, introducing architectures optimized for throughput serving, reasoning workloads, disaggregated compute tiers, and persistent contextual memory. This article explores a six-category planning matrix linking AI workloads to infrastructure requirements. For a glossary of terms and a deeper exploration of the concepts introduced here, please see Architecting AI at Scale: From Training Clusters to Inference-Driven Infrastructure

Synchronous training fabric

TRAINING workloads process massive datasets through repeated forward and backward passes while synchronizing gradient updates across thousands of GPUs. Because collective operations occur continuously across the cluster, network consistency and latency control become critical infrastructure requirements. To support this synchronized compute behavior, most training environments use dense Clos network topologies built on high-bandwidth Ethernet fabrics. Within each server, GPUs communicate over scale-up interconnects such as NVLink, while RoCE-based scale-out networks transport collective operations including allreduce and alltoall traffic between servers. As a result, training fabrics are engineered to deliver predictable throughput and tightly controlled latency.

Training clusters also create significant fiber density requirements because large deployments can require tens of thousands of optical connections across relatively short distances. Therefore, structured cabling systems must support high port counts, organized routing pathways, and operational accessibility for ongoing expansion and maintenance. Network behavior must also remain predictable during synchronized compute activity, with congestion or latency variation affecting operations across the cluster. Infrastructure planning consequently focuses on deterministic forwarding, congestion management, and tightly engineered fabric design.

Throughput inference pod

Throughput inference has become a primary production workload across many AI environments. Because these systems prioritize requests per second

and efficiency, infrastructure design focuses on scalable serving capacity rather than tightly synchronized compute behavior. Most inference pods operate with lower compute density than training clusters and commonly use mid-range GPUs, PCIe-attached accelerators, and distributed serving pools behind load balancers. Cross-node synchronization requirements are also reduced because many models can operate within a single server. This creates more balanced network behavior across the environment.

North-south traffic carries client requests and generated responses, while east-west communication supports model distribution, cache movement, and internal service coordination. As a result, many deployments rely on 100GbE or 200GbE RoCE fabrics with lower overall fiber density and greater emphasis on modular scalability. Infrastructure planning prioritizes flexibility, longer-reach connectivity, and incremental expansion without major redesign.

Disaggregated reasoning pod

Reasoning inference supports workloads such as long-context analysis, coding assistance, document synthesis, and multi-step generative tasks. Because these environments generate large token outputs and extended compute sequences, latency per token becomes a critical performance requirement. Most reasoning architectures separate prefill and decode into distinct compute pools. Prefill systems process prompts in parallel using GPU-dense infrastructure, while decode systems focus on sequential token generation and memory bandwidth efficiency. This separation introduces a dedicated traffic pattern across the AI fabric through KV-cache transfers between compute tiers.

KV-cache movement can reach gigabyte scale during long-context operations, creating large unidirectional traffic bursts between prefill and decode systems. This means network behavior depends on low tail latency, careful queue management, and tightly controlled congestion handling across RoCE fabrics. Fiber infrastructure also becomes more specialized because prefill and decode pools may span multiple racks, rows, or halls. Infrastructure planning, therefore, prioritizes shortest-path optical routing, controlled pathway allocation, and low-jitter interconnects.

Heterogeneous decode-accelerated pod

This architecture extends disaggregated inference by shifting decode execution from GPUs to specialized accelerators optimized for sequential token generation. These systems may include ASICs, FPGA-based devices, or custom memory-optimized hardware. Because of this separation, GPU-based systems continue to handle prefill workloads while decode execution is optimized for throughput efficiency and lower cost per token. The resulting infrastructure introduces heterogeneity across compute tiers, where different systems operate with varying NIC capabilities, MTU configurations, and buffering behavior. Therefore, the RoCE fabric must support consistent communication across non-uniform endpoints while maintaining stable congestion control and predictable latency.

KV-cache transfer remains the dominant inter-tier flow, with large burst traffic moving from prefill GPUs to decode accelerators. Asymmetry increases further due to differences in execution cadence and throughput between tiers, placing additional emphasis on queue management, adaptive routing, and traffic shaping across the network. Physical infrastructure also becomes denser at the decode layer, where higher rack-level concentration of accelerators increases local fiber demand at top-of-rack switching. These systems highlight how compute specialization extends into network and optical design, requiring tighter coordination between accelerator architecture, Ethernet behavior, and fiber deployment strategy.

Context-centric agentic pod

Emerging agentic AI systems introduce persistent contextual memory that spans ongoing interactions, workflows, and multi-step reasoning tasks. Because these systems depend on continuous access to stored state and intermediate data, a dedicated context-memory layer becomes a core part of the architecture. This context layer is typically implemented using RDMA-accessible DRAM, NVMe storage, or CXL-enabled memory systems. Prefill and decode tiers interact continuously with this shared memory pool to retrieve context, store intermediate results, and maintain session state.

The result is the introduction of a third primary traffic class within the AI fabric: compute-to-compute traffic, compute-to-context traffic, and north-south service traffic. Since delays in context retrieval directly impact response quality and system performance, the network must support concurrent flows with strict latency consistency. This means fiber infrastructure becomes more segmented, with dedicated optical paths often required to isolate context-memory traffic from bulk data movement elsewhere in the facility. In some instances, separate fiber domains or tightly controlled wavelength allocation are used to maintain predictable latency. These architectures reflect a shift toward infrastructure that integrates persistent memory systems and contextual retrieval directly into the operational AI fabric, rather than treating them as external supporting services.

Agent workflow and tool-execution estate

AI systems increasingly extend beyond inference into orchestration of external tools and services. Because these workflows execute multi-step sequences, the orchestration layer becomes a distinct operational domain alongside inference infrastructure. Traffic patterns resemble distributed service architectures rather than synchronized accelerator fabrics.

Standard Ethernet replaces RoCE in most deployments, since performance requirements operate at millisecond rather than microsecond scale. Flows are shorter-lived, packet sizes are smaller, and east-west communication behaves more like large-scale web infrastructure than tightly coupled compute clusters.

Although optical demands are less intensive than GPU-driven systems, co-location within shared facilities introduces additional complexity. Workflow systems, inference clusters, storage platforms, and training environments often share aggregation layers and physical pathways, requiring coordinated capacity planning. This creates a multi-fabric environment where independent workloads operate within the same physical infrastructure, each with distinct density, latency, and scaling characteristics that must be managed at the network and fiber layer.

Conclusion

AI infrastructure is moving toward specialized architectures aligned to distinct workloads. Each introduces unique network behaviors and distinct optical requirements. RoCE-based Ethernet is emerging as a common scale-out transport layer across these environments. At the same time, fiber infrastructure is shifting from a uniform backbone to coordinated connectivity domains that reflect different traffic patterns and performance needs.

These shifts mark a clear transition toward workload-aware infrastructure planning. Organizations that understand these architectural differences are better positioned to scale AI systems amidst rapidly growing and increasingly diverse compute demand.

Redefining hydronic infrastructure for D2C cooling

Rethinking hydronic design with polymer piping systems is key to efficient, scalable, and durable D2C liquid cooling in high-density data centers.

AS DATA CENTERS evolve to support artificial intelligence (AI) and high-performance computing, the demands placed on cooling infrastructure are changing fundamentally. Rack densities are rising beyond 100 kW, and with that, the need for more efficient, reliable, and scalable thermal management has never been greater. Direct-to-chip liquid cooling (DLC) is rapidly becoming the preferred solution, but unlocking its full potential requires more than just advanced cooling units and cold plates. It requires a rethinking of the entire hydronic infrastructure.

From component optimization to system thinking

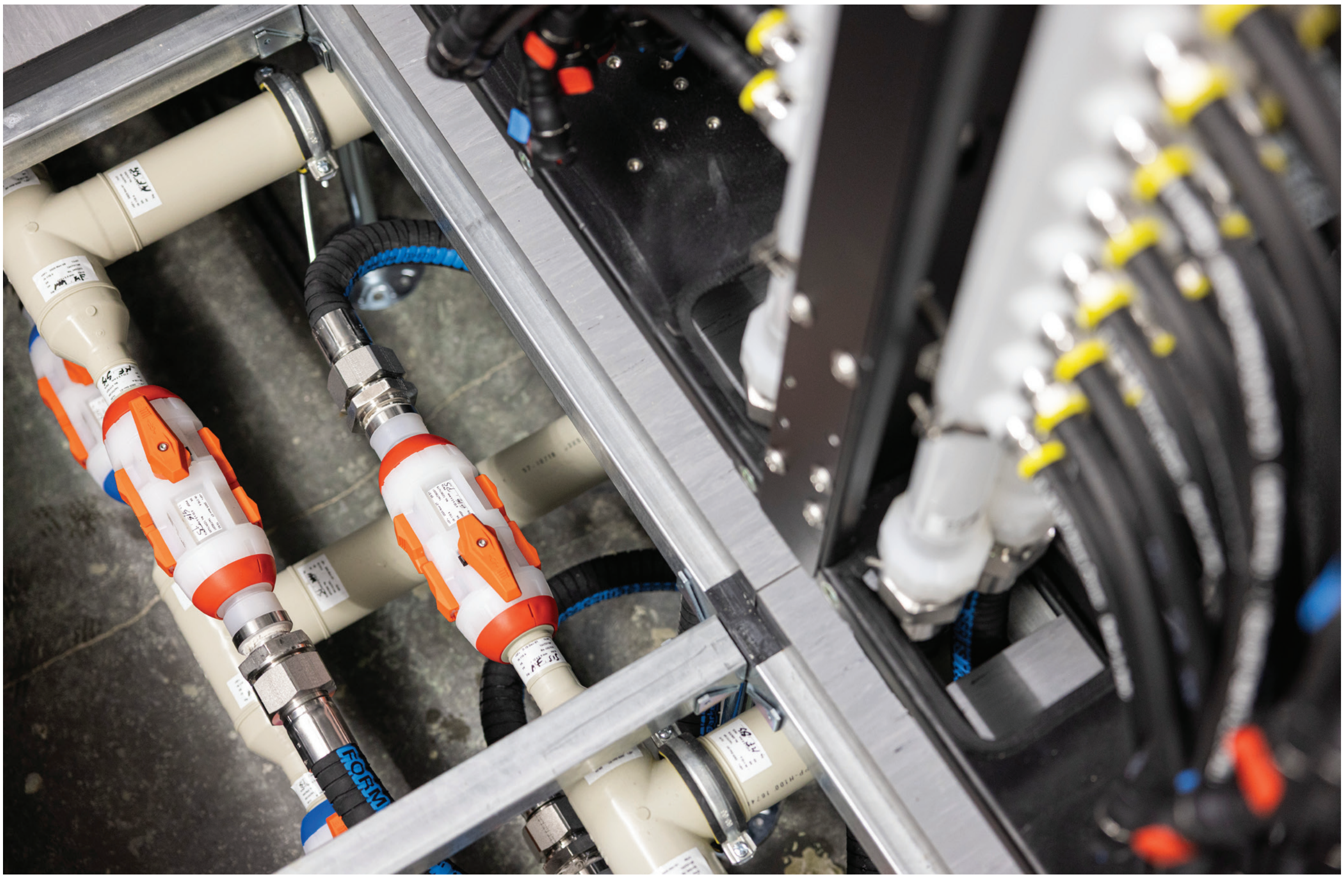
Much of the current discourse around liquid cooling focuses on visible and high-impact components: cold plates, cooling distribution units (CDUs), and heat exchangers. While these elements are critical, they represent only part of a much larger system. DLC operates as a closed-loop hydronic network, connecting facility water systems (FWS) with technology cooling systems (TCS) and delivering coolant directly to the chip. Within this loop, piping is not simply a passive transport medium. It is a core enabler of system performance, governing flow behavior, influencing energy efficiency, and playing a decisive role in maintaining coolant purity.

As system complexity increases, the limitations of traditional approaches to hydronic design are becoming more apparent.

The hidden constraints of conventional piping

Historically, metal piping has been the standard choice for data center cooling systems. Its widespread use is largely based on familiarity and established practices. However, in the context of modern DLC environments, this approach presents several challenges.

One of the most significant is corrosion. Metal piping systems are inherently susceptible to corrosion over time, particularly when exposed to water-based coolants and additives. This process leads to the release of particles into the fluid, which can accumulate and create deposits. In direct-to-chip cooling systems, where cold plates rely on microchannels to



transfer heat efficiently, even minimal contamination can lead to fouling, reduced performance, and increased maintenance requirements.

In addition to contamination risks, corrosion and scaling increase internal surface roughness, resulting in higher friction losses and reduced hydraulic efficiency. Over time, this can significantly impact system performance and energy consumption.

Installation challenges further compound these issues. Metal piping is heavy and typically requires welding and assembly on-site, extending project timelines and introducing variability.

As data center construction faces increasing pressure to deliver faster and more efficiently, these constraints are becoming increasingly critical.

A shift toward engineered hydronic solutions

High-performance polymer piping systems are gaining traction as a viable and advantageous alternative. Unlike metal, polymers are inherently corrosion-free, eliminating the risk of internal degradation and particle release. This ensures consistent coolant purity, which is essential for protecting sensitive components such as cold plates.

Polymer piping also offers exceptionally smooth internal surfaces, supporting stable and predictable hydraulic performance. Without the risk of corrosion-induced roughness or scaling, systems can maintain optimal flow conditions over time.

From a construction standpoint, polymer systems provide significant advantages. Their lightweight nature simplifies handling and transport, while their compatibility with prefabrication enables large sections of piping to be assembled off-site. These prefabricated modules can be installed quickly and efficiently, reducing on-site labor and shortening project timelines.

Supporting scalability and sustainability

Sustainability considerations are also driving material selection. Polymer piping systems typically offer a lower embodied carbon footprint compared to traditional metal solutions, supporting environmental targets as data centers scale.

At the same time, engineered hydronic systems enable greater flexibility. Modular and prefabricated designs allow infrastructure to expand alongside demand, ensuring scalability without compromising reliability.

Ultimately, piping is no longer a secondary component, it is a critical foundation for performance, efficiency, and long-term operation in modern data center cooling.

GF is a global leader in flow solutions, delivering mission-critical products and services for the safe and sustainable transport of water and other fluids. With more than 60 years of expertise in advanced polymer piping systems for high-tech industries, GF supports data center operators with reliable and efficient cooling infrastructure.

Founded in 1802 and headquartered in Switzerland, GF employs approximately 13,300 people across 46 countries. In 2025, GF generated sales of CHF 3 billion and is listed on the SIX Swiss Exchange.

DCA INSIGHTS

Steve Hone
CEO,The DCA
(Data Centre Alliance)

Data Centre Alliance

HERE we are in May already! The DCA team are preparing for a busy couple of months. The number of DC events is now absolutely astonishing – forget the usual work! We could all spend our time at industry events!

See below for the upcoming events in the sector.

The primary role of the UK Data Centre Trade Association is to ensure the industry is informed and up to date. We do this in collaboration with our Partners - publishing content, speaking / moderating and hosting knowledge sharing & networking event. We have also recently launched two incredibly useful tools for DCA Partners.

- The Regulatory Radar in collaboration with Clear Decisions provides AI-powered policy intelligence for EU & UK data centre operators.
- Power Trends Dashboard in collaboration with 1GibLabs analyses power distribution, availability, capacity, and regulatory trends for data centre location selection.

Both of these are fantastic tools for our Partners to use when logged into the Members Portal – if you need help accessing, let us know!

DCA events / hosted sessions

Get these DCA events and sessions in your diaries. Visit the events page on the DCA website for more information.

Up and coming industry events supported by The DCA

- **02 June 2026** - Datacloud Global Congress, Cannes
- **17 June 2026** – Major Energy Users Council MEUC
- **22 June 2026** - ISC High Performance 2026, Hamburg
- **23 June 2026** - Data Center Anti Conference Europe, London
- **25 June 2026** - Mixing IT Networking event, London
- **02 July 2026** - PowerEx Golf Day, Harrowgate
- **03 July 2026** - PowerEx Live North, Harrowgate

DCA INSIGHTS

This DCA feature comprises of articles from DCA Partners. My thanks to all the authors and their teams for providing these valuable contributions.

- **Flammable Gas? In MY Data Centre Cooling System? What ARE They Thinking?** - *Mike Hayes, Applications Specialist – FläktGroup*
  Mike shares his thoughts on data centres transitioning to "mildly flammable" A2L refrigerants to ensure environmental compliance while managing operational safety risks.

- **Beyond Power and Cooling: The Building Fabric's Role in Data Centre Resilience** - *Mo Aljan, Business Development Director - Saint-Gobain*
  Mo explains that the building fabric of a data centre no longer sits outside the infrastructure. It forms part of the facility's uptime architecture.

- **AI Is Redefining Data Center Infrastructure — And Exposing the Limits of Power and Cooling Architectures** - *Özlem Merhametsiz, Marketing Manager - Legrand Data Center Solutions*
  Özlem details how artificial intelligence (AI) is not simply increasing demand; it is fundamentally changing how data centers behave at a physical level.

- **Taking data centre power monitoring to the next level** - *James Kirkwood, Technical Account Director – EkkoSense*
  James informs us that data centre operators can unlock immediate cost savings and stranded capacity by optimising existing power usage before investing in new infrastructure.

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

Best regards,
Steve

www.dcauk.org

https://www.linkedin.com/company/data-centre-alliance

DCA EVENTS / HOSTED SESSIONS

- **30 JULY 2026 –** DCA 10X10 & Networking, London
- **20 OCTOBER 2026 –** Data Centre Transformation, Birmingham
- **18 NOVEMBER 2026 –** DCA 10X10 hosted at Data Centres, Ireland
- **10 DECEMBER 2026 –** DCA 10X10 hosted at PowerEx Live, London

Flammable gas? In MY data centre cooling system? What ARE they thinking?

As refrigerant regulations tighten, data centre cooling systems face a difficult balance between sustainability, efficiency and operational risk. Mike Hayes of FläktGroup explores how low-GWP refrigerants and evolving DX technologies could offer a compliant path forward without introducing flammability concerns into the data hall.

BY MIKE HAYES, APPLICATIONS SPECIALIST – FLÄKTGROUP

WITH acknowledgement to the BBC character 'The 13th Duke of Wybourne', such is the reaction from most DCA members upon hearing about the next generation of refrigerants. These must be referred to as "Mildly Flammable" and classified A2L, rather than the much more reassuring label of "Non-Flammable" or A1, that we are all familiar with. The purpose of this article is to look at the latest environmental issues surrounding refrigerants in our Direct Expansion (DX) cooling systems and determine methods of compliance that do not involve any increased risk of explosions!

Environmental concerns

At the end of the 1980s we had the Montreal Protocol, which successfully phased out chlorine-based refrigerants (CFCs) due to depletion of the ozone layer, which protects us all from higher levels of Ultraviolet (UV) radiation and hence from such nasties as skin cancer. Today though, the concern is increased global warming, caused by the release of Carbon Dioxide (CO_2) and other "Greenhouse" gases into the atmosphere. This layer of gases, like the glass in a greenhouse, traps the Infra-Red (IR) radiation from the sun and prevents much of it being reflected out into space. The greater the amount of greenhouse gases in the troposphere, the warmer the planet becomes.

EU refrigerant restrictions

The current popular refrigerant used in Computer Room Air Conditioners (CRAC) systems is R410A, which has a Global Warming Potential (GWP) of 2088, meaning that just 1 kg escaping is equivalent to 2088 kg of CO_2! This is why the recent Kigali Amendment has now decreed no new R410A products from 2028, all refrigerants to have a GWP <750 from 2028, moving down to <150 by 2032. An article such as this should also mention Polyfluoroalkyl substances (PFAS) otherwise known

as "Forever Chemicals" which even low GWP refrigerants contain or will break down into. There is, as yet, no legislation covering these, and it is muted that "Specialist" applications such as cooling for DCs will be exempt anyway.

Impact upon cooling systems

Unfortunately, there is no direct "Drop-in" low GWP replacement for R410A that does not carry a label reading "Mildly flammable" and hence A2L rating. Of course, flammability is of much less concern if you have a chilled water system, with an externally mounted air-cooled chiller. In this case any refrigerant leakage escapes to atmosphere and does not cause risk of ignition. This means that some chillers can even use highly flammable refrigerants such as propane (R290) which has a GWP of zero.

The big issue

For office or "Comfort" air conditioning systems, flammable refrigerants (A2L) are relatively easy to accommodate and so have been readily adopted by manufacturers. A simple calculation looks at the quantity of refrigerant that could potentially leak into the conditioned space and whether the resultant concentration of gas remains below the lower flammability limit (LFL) of the refrigerant. The problem is that Data Centres have a lot more cooling in a lot less space, hence a big issue!

The challenge

Proven over the decades, the humble DX split system CRAC units will always have their place in Data Centres. The refrigeration system is "Split" between indoor and outdoor units, with only small-bore gas and liquid refrigerant pipes running between the two. In this way, heat is taken directly from the air in the DC and rejected directly to outside air. This approach is popular mainly for reasons of reliability, autonomy and energy efficiency, together with speed and ease of installation. Therefore, a low GWP, non-flammable (A1) refrigerant for such DX split systems would doubtless be high on the priority list for many DCA members.

DX systems evolution

FlaktGroup, now an arm of Samsung, have been developing advanced refrigeration systems for 60 years, moving away from fixed speed motors and mechanical refrigeration controls over the last 15 years. For example, the latest Multi Denco range, incorporates only variable speed fans and compressors, together with electronic expansion valves and pressure transducers to monitor refrigeration performance. The good news here is that a simple software update is now all that is required to allow operation with a safe, and long-term tested, low GWP alternative refrigerant.

There is a solution

The proposed transition is from R410A to R513A, which remains non-flammable (A1) but has a GWP of only 633, getting us below the EU 2028 GWP <750 requirement. The fossils amongst us, me included, will remember R12 as a very good, relatively low-pressure CFC refrigerant which was replaced by the popular, single component "Drop In" R134a because of the Montreal Protocol. Now, the Kigali Amendment drives us towards low GWP R513A as an A1 "Drop In" for R134a. Although not straying too far from field proven technology, the switch to R513A requires a higher refrigerant flow rate than R410A. to achieve the same cooling capacity. Some models may therefore be fitted with physically larger compressors but running at lower speed to equal or improve upon R410A power inputs. Another big advantage of variable speed fans and compressors is that the power input reduces as the cube of the reduction in speed. This means that a system operating at 80% of the cooling capacity uses only 50% of the power input. At around 60% cooling capacity where many systems operate, the power input is reduced to only 20%.

Conclusions

A quick calculation of total equivalent warming impact (TEWI) will demonstrate that 85% of the CO2 output from a data centre cooling system is associated with the (UK) grid that powers it and hence has little to do with refrigerant GWP! For this reason, our eyes should remain firmly on system energy efficiency, as they have been for many years now. Indirect free cooling systems, such as Multi Denco DMF will provide an annualised pPUE of 1.1 and will operate on low GWP R513A. Nobody really knows what mildly flammable (A2L) refrigerants will bring, but at least you will not have to find out in your data centre!

Beyond power and cooling: The building fabric's role in data centre resilience

As data centres push for greater resilience, efficiency and sustainability, the building fabric is emerging as a critical part of infrastructure strategy. From thermal stability and climate defence to fire protection and lifecycle performance, the envelope is playing an increasingly important role in supporting uptime and long-term operational success.

BY MO ALJAN, BUSINESS DEVELOPMENT DIRECTOR - SAINT-GOBAIN

AS DATA CENTRES evolve to meet increasing performance, sustainability and resilience demands, the role of the building fabric is changing. No longer simply an enclosure, it is becoming part of the infrastructure, supporting thermal control, protecting against environmental risk and contributing to long-term operational performance.

Data centres are now widely recognised as critical national infrastructure, supporting cloud computing, artificial intelligence and the wider digital economy. As demand accelerates, operators and developers face increasing pressure to deliver facilities that are more resilient, more energy-efficient and capable of continuous operation.

Much of the industry's focus has understandably been on power and cooling systems. However, as operational demands increase, another factor is moving up the agenda: the performance of the building itself.

In modern data centres, the building fabric, understood here as the envelope and passive protection systems, is no longer simply the shell around digital infrastructure. It is becoming part of the infrastructure, influencing thermal stability, environmental resilience, fire performance and lifecycle value.

In effect, the building fabric no longer sits outside the infrastructure. It forms part of the facility's uptime architecture.

From structure to system

Historically, the building envelope has been treated as a secondary consideration, providing weather protection and basic environmental control while performance was driven by mechanical and electrical systems.

That distinction is becoming less clear.

Increasing rack densities, particularly driven by AI workloads, are placing greater demands on cooling systems and internal environmental control. At the same time, external pressures such as rising temperatures and more frequent extreme weather events are making stable operation more difficult to maintain.

Research from the Uptime Institute indicates that 45% of operators have experienced an extreme weather event that threatened continuous operations, with nearly 9% reporting an outage or significant disruption as a result. This

highlights that resilience is no longer solely a function of power and cooling systems.

In this context, the building fabric becomes more closely tied to operational performance.

Four ways the building fabric functions as infrastructure

A useful way to understand this shift is to consider four ways in which the building fabric now supports infrastructure performance.

Thermal control

As facilities move towards higher-density environments, maintaining consistent thermal conditions becomes more complex. Many operators are now managing hybrid cooling environments, where air-cooled and liquid-cooled systems coexist within the same facility. Commentary from Data Center Knowledge suggests that these mixed-density halls are often more operationally challenging than fully air- or fully liquid-cooled environments, because they require operators to manage different cooling architectures and risk profiles within the same facility.

Standards bodies such as ASHRAE define tight environmental conditions for IT equipment, reinforcing the importance of maintaining stable internal conditions through both cooling strategy and envelope performance.

In that context, the building envelope plays a critical supporting role. Heat gain through roofs and façades, air leakage and envelope inefficiencies can increase cooling load and reduce system efficiency. Designing for airtightness, insulation performance and reduced heat gain through the building envelope supports more stable internal conditions, reduces avoidable cooling load and improves operational reliability.

Climate defence

Climate resilience is now a core design requirement rather than a future consideration.

Data centres must operate reliably under increasingly variable environmental conditions, including extreme temperatures and flooding events. Water ingress remains one of the most persistent operational risks in data centres, with even relatively minor failures capable of affecting service continuity.

Industry risk guidance and operational commentary identify roof interfaces, plant zones and service entry points as areas of heightened vulnerability, where failures in detailing, waterproofing or long-term durability can lead directly to operational disruption, asset risk and costly remediation.

As a result, waterproofing, façade performance and interface design must be treated as critical components of infrastructure rather than secondary construction details.

Passive fire protection

Fire risk in data centres is less about large-scale events than about the consequences of localised incidents, particularly smoke spread and contamination.

Guidance on data-centre fire and smoke resilience from insurers such as Aviva emphasises that even relatively minor incidents can lead to disproportionate operational disruption, particularly when smoke spreads beyond the point of origin and beyond the intended compartment.

In practice, smoke can travel through relatively small openings if service penetrations are not adequately sealed or protected. This reinforces the need for robust compartmentation strategies, rated separations and consistent detailing across interfaces to maintain the integrity of fire protection systems.

Lifecycle performance and carbon

Sustainability considerations are also reshaping how data centres are designed.

While operational energy remains a major focus, embodied carbon and lifecycle performance are becoming increasingly significant. Data centres are complex assets, with a large proportion of their carbon footprint associated with mechanical and electrical systems. Industry initiatives such as the iMasons Climate Accord highlight the importance of assessing whole-life carbon across both building and infrastructure systems, while guidance from RICS reinforces the importance of durability and lifecycle value in reducing long-term environmental impact.

Within that context, the building fabric contributes through durability, material efficiency and reduced need for replacement or intervention over time. Systems that maintain performance over extended asset lifecycles, with minimal degradation, support both operational resilience and more efficient lifecycle carbon outcomes.

Why early integration matters

These factors point towards a broader shift in how data centres are designed and delivered.

Rather than treating building fabric, mechanical systems and operational requirements as separate elements, there is increasing recognition that they must be considered together from the earliest design stages. Early coordination helps ensure that envelope performance, fire strategy and environmental resilience align with cooling and operational requirements.

This approach reduces specification risk, improves performance outcomes and supports more efficient delivery, particularly in the context of accelerated construction programmes and phased expansion.

Organisations such as the Data Centre Alliance play an important role in supporting this integration, helping to share best practice and promote standards across the sector.

Conclusion

As data centres continue to evolve, the demands placed on their performance, resilience and sustainability will only increase.

In this environment, the building fabric is taking on a more defined role within the overall infrastructure of the facility. Its contribution to thermal control, environmental resilience, passive fire protection and lifecycle performance is becoming increasingly critical to long-term success.

For designers, developers and operators, this represents a necessary shift in perspective, recognising that reliable performance depends not only on the systems within the building, but on how the building itself is designed, detailed and protected to support them.

AI Is redefining data center infrastructure – and exposing the limits of power and cooling architectures

AI is reshaping data centre infrastructure beyond higher rack densities, exposing the limits of traditional power and cooling models. As workloads become more volatile, operators must rethink infrastructure as an integrated, responsive system built for dynamic stability.

BY ÖZLEM MERHAMETSIZ IS A MARKETING MANAGER AT LEGRAND DATA CENTER SOLUTIONS

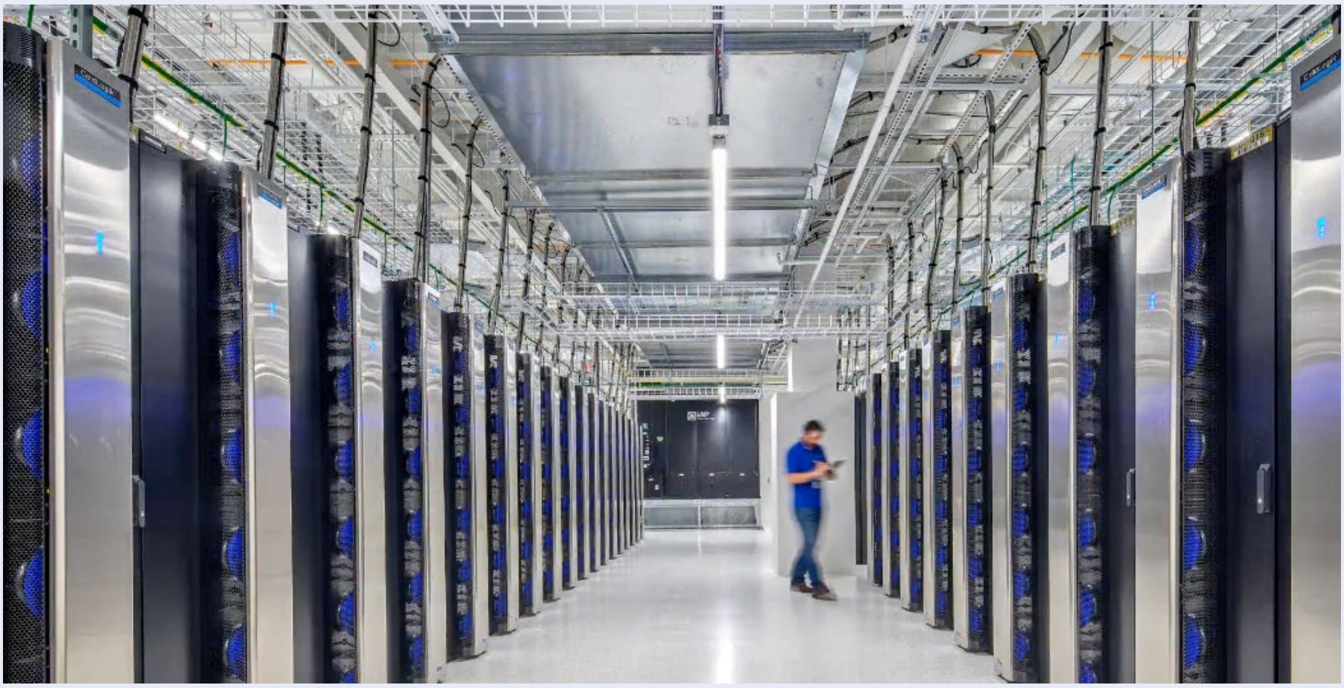
FOR MORE than two decades, data center infrastructure has been built around a stable equation: predictable workloads, gradual power variation, and cooling systems designed to maintain equilibrium. That equation no longer holds.

Artificial intelligence (AI) is not simply increasing demand; it is fundamentally changing how data centers behave at a physical level. Power consumption is no longer linear. Heat generation is no longer progressive. And the traditional separation between electrical and mechanical infrastructure is rapidly becoming obsolete.

What is emerging is a new challenge: dynamic instability. In AI and high-performance computing (HPC) environments, power and cooling systems must respond in real time to rapid, synchronized fluctuations driven by compute workloads. The data center is no longer static; it is a dynamic system that must continuously adapt.

The real disruption: volatility, not just density

Much of the industry's focus has been on density - more GPUs, higher rack



power, faster scaling. But deeper disruption lies in variability.

AI workloads generate rapid, high-amplitude power fluctuations driven by training cycles and synchronized compute activity. In modern GPU clusters, power demand can shift dramatically within milliseconds, leading to continuous micro-cycling across the infrastructure.

Traditional environments were designed for:

- Gradual load variation
- Predictable redundancy events
- Limited transient behavior

AI introduces:

- Sudden, synchronized power spikes across clusters
- High-frequency transitions between idle and peak states
- Persistent electrical volatility at scale

The implication is clear: infrastructure is no longer sized purely for capacity—it must be engineered for behavior. This shift impacts both power delivery and thermal management simultaneously.

Power isn't the only problem—or the only solution

The first response to AI-driven demand has been to reinforce electrical infrastructure—improving UPS efficiency, increasing overload capacity, and enhancing transient response.

These steps are necessary, but not sufficient.

Every electrical fluctuation has a direct thermal consequence. Power spikes generate immediate increases in heat output, and in AI environments, these spikes occur across entire clusters.

If power systems respond in milliseconds while cooling systems lag, instability does not disappear—it simply shifts:

From electrical instability → to thermal instability

This imbalance is becoming one of the defining constraints in modern data center design.

The cooling bottleneck: when airflow reaches Its limits

Traditional cooling architectures were not designed for this level of responsiveness.

Room-based systems—such as CRAH units, raised floors, and containment—operate on airflow distribution and thermal averaging. They assume stable and predictable heat loads.

AI workloads break these assumptions.

Heat is now:

- Highly localized, creating rack-level hotspots
- Highly dynamic, fluctuating in real time
- Highly concentrated, with rack densities anchored around 80 kW per rack and scaling to 100-200 kW in leading environments

Air, as a cooling medium, struggles to efficiently manage these conditions. Its thermal capacity and responsiveness are limited compared to the intensity and speed of AI-driven heat loads. The result is systemic inefficiency:

- Overcooling to compensate for uncertainty
- Persistent hotspots due to delayed response
- Increased energy consumption and reduced efficiency

Cooling is no longer a background function - it is becoming a critical control layer for maintaining infrastructure stability.

Rear door cooling: bringing thermal control to the source

To address this challenge, cooling strategies are shifting toward more localized approaches.

Rear door heat exchangers (RDHx), remove heat directly at the rack, capturing it before it recirculates into the room. By reducing thermal lag, they enable a more immediate response to changing workloads.

Modern rack-level cooling approaches offer several advantages:

- Faster thermal response, aligned with compute activity
- Higher density support, without relying on room airflow or requiring facility redesign
- Improved efficiency through targeted heat removal

These systems shift cooling from a passive, room-level function to an active, responsive component of infrastructure stability.

Engineering electrical stability in the AI era

At the same time, power infrastructure must evolve to handle increasingly aggressive electrical behavior.

AI workloads introduce rapid load cycling and extreme current transients, placing new stress on power conversion systems, conductors, and semiconductors. This has implications for both performance and long-term reliability.

Maintaining resilience in this environment requires infrastructure that can absorb and manage volatility - not just deliver peak capacity. Modern UPS architectures are adopting silicon carbide (SiC) semiconductors to deliver efficiencies exceeding 98% and reduce switching losses during rapid load transitions. Combined with internal busbar-based power distribution, this supports stable current flow, minimizes thermal hotspots, and provides consistent voltage behavior under highly dynamic conditions.

UPS systems are evolving beyond their traditional role as backup infrastructure. They are becoming active regulators of electrical behavior within the data center environment.

Beyond silos: toward integrated infrastructure

One of the clearest lessons from AI is that power and cooling cannot be treated as separate domains.

A power spike without a coordinated cooling response leads to thermal instability. A cooling system without electrical awareness introduces inefficiency and risk.

The future lies in integration—where power, cooling, and distribution operate as a coordinated system.

This requires:

- Alignment between electrical and mechanical design
- Greater visibility into real-time system behavior
- Control strategies that adapt dynamically

The data center must evolve from a collection of independent systems into a unified, responsive platform.

From performance to control

Historically, data center design has focused on performance metrics such as efficiency, capacity, and footprint.

These metrics remain important—but they are no longer enough. The real challenge in AI environments is maintaining stability under dynamic, unpredictable conditions.

This marks a fundamental shift: From maximizing performance ➔ to controlling behavior.

Control - over power fluctuations, thermal response, and system interaction is becoming the defining measure of infrastructure effectiveness.

Designing for dynamic stability

As AI adoption accelerates, infrastructure demands will continue to evolve. Power systems must handle greater variability. Cooling systems must operate closer to the source. And the entire environment must respond faster and more precisely.

The data center is no longer static—it is a dynamic system shaped by the workloads it supports.

Technologies such as modular power architectures and localized cooling are not incremental improvements. They are essential to enabling stability in this new environment.

Because in the era of AI, the key question is no longer how much power or cooling can you deliver - but how effectively can you control it.

> Maintaining resilience in this environment requires infrastructure that can absorb and manage volatility - not just deliver peak capacity. Modern UPS architectures are adopting silicon carbide (SiC) semiconductors to deliver efficiencies exceeding 98% and reduce switching losses during rapid load transitions

To explore more on how high-power modular UPS architectures and integrated rack-level cooling approaches are enabling greater stability in AI-driven data centers, visit Legrand.

Taking data centre power monitoring to the next level

Before investing significantly in new power generation, interconnection or site development, it’s important that data centre operators should work to optimise their existing use of power. Done smartly, this approach can help unlock immediate cooling energy cost savings and improve site power efficiency by opening up stranded capacity. So how can data centre operators put such a power optimisation approach into practice?

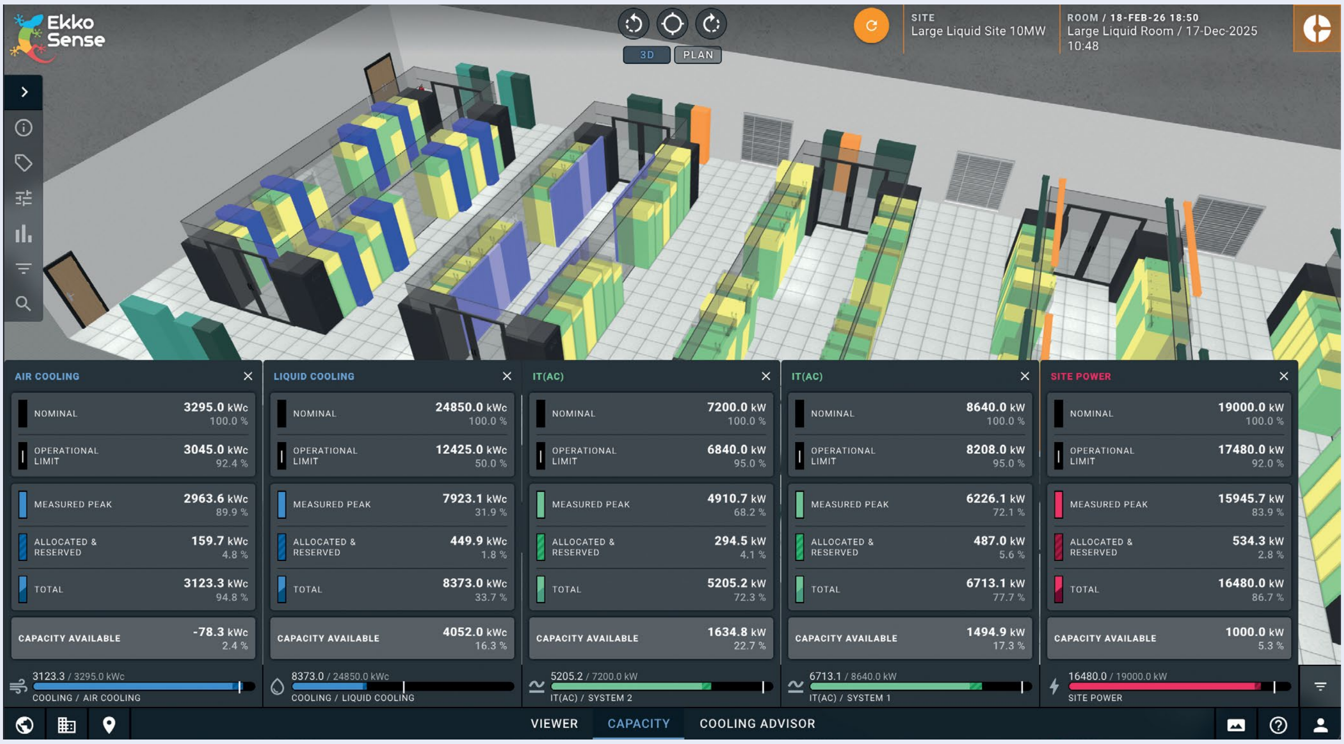
BY JAMES KIRKWOOD, TECHNICAL ACCOUNT DIRECTOR. EKKOSENSE

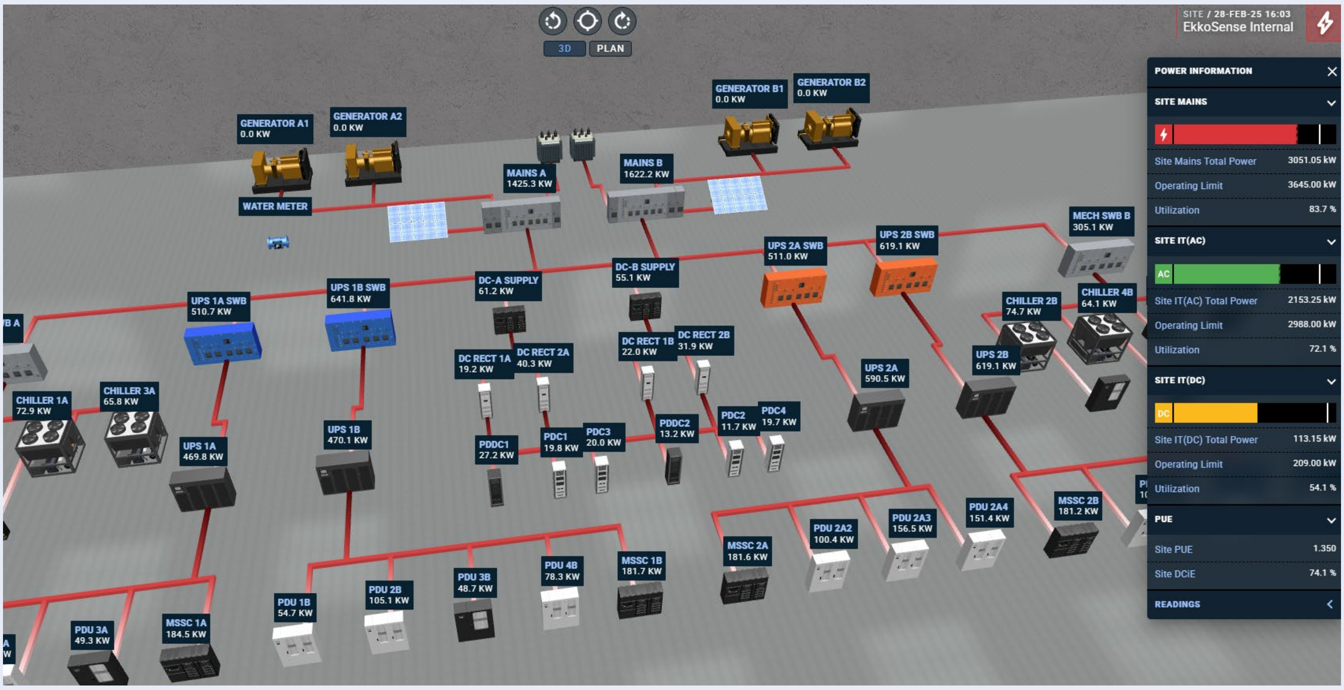
DATA CENTRE electricity demand shows no signs of slowing down. Real estate services company JLL predicts that overall global data centre capacity will double by 2030. IEA research echoes this, identifying that overall global data centre electricity consumption will almost double, scaling from 485 TWh in 202 to some 950 TWh in 2030. It also projects that consumption from ‘AI-focused data centres’ will grow around 3x – much faster than other areas.

Not surprisingly, securing new power and infrastructure to support these projected demands is proving challenging. According to analysis from Sightline Climate, between 30 to 50% of data centres scheduled to come online during 2026 may be delayed, with power constraints, equipment shortages and local opposition cited as key barriers. Long lead times for infrastructure upgrades to transformers and substations, along with inevitable regulatory hurdles are also proving a key delaying factor in bringing new capacity on board.

Transforming raw power data into actionable, operational insights

Given all this, there’s never been a greater requirement for operations teams to ensure that their existing data centre assets are performing optimally so that they are able to support as much workload as possible. However, to do this operations teams will need





to know exactly what's going on across their data centre estates. This requires real-time visibility into power consumption, capacity and risk across their estate – only then can they take the informed decisions needed to improve power efficiency, resilience and planning accuracy.

Unfortunately there are a number of issues here that need resolving if existing assets are to be maximised. Firstly, AI workloads are complex and typically require much tighter operational tolerances and much faster responses. Unlike traditional data centre workloads, high-density AI environments are potentially much more volatile. Power demand shifts quickly, thermal conditions change in real time and small issues can escalate fast if they're not seen and resolved early. For organisations running or planning AI at scale, having clear real-time operational insight is now essential.

However, while data centre investment is surging forward, many current legacy facilities and monitoring capabilities simply aren't keeping pace with evolving GPU and AI workloads. For example, most existing BMS and EPMS systems cannot forecast power and cooling bottlenecks effectively to identify any stranded power capacity, while lack of real-time insight means that traditional tools generally lack the level of holistic insight needed to maximise power inputs and optimise energy use.

The result for many organisations is that they're still running with very average Power Usage Effectiveness (PUE) levels. In its 2025 Global Data Centre Survey, Uptime Institute reported that average annual PUE has remained largely static for the last six years, ranging between 1.59 and 1.54. Whether that's under-utilised servers, inefficient cooling, or stranded capacity across an estate, it's clear that there's a significant opportunity for data centre teams to optimise their PUE scores and power, capacity and thermal performance through the continuous measurement and transformation of raw power data into actionable, operational insights.

Everything starts with power visibility

A successful power optimisation program begins with visibility. Many data centres still operate with fragmented monitoring systems that fail to reveal exactly where power and cooling capacity is being wasted. That's why the continuous measurement and analysis of electrical consumption across both IT and facility infrastructure is so important – particularly as facilities adapt to increasingly higher density workloads.

Effective data centre power monitoring works by continuously collecting and analysing power data, gathering the information and insight needed to unlock actionable insight and make informed decisions. Key requirements here include:

- **Real-time monitoring of data centre power consumption** – with live monitoring of power usage across racks, rows, rooms, and facilities to understand energy consumption and identify abnormal patterns before they become operational risks.

> Unfortunately there are a number of issues here that need resolving if existing assets are to be maximised. Firstly, AI workloads are complex and typically require much tighter operational tolerances and much faster responses. Unlike traditional data centre workloads, high-density AI environments are potentially much more volatile

- **Power usage analytics across both IT and facility infrastructure** - Correlating power data from UPS systems, PDUs, branch circuits, and IT load to provide a unified view of data centre power consumption and capacity utilisation.
- **The ability to govern power usage to maintain resilience and efficiency** – Applying power thresholds, alerts, and governance controls helps ensure changes are managed safely, protecting critical infrastructure while optimising energy efficiency

At EkkoSense we're committed to opening up enhanced visibility into data centre power consumption and capacity, helping operators to reduce waste, improve planning accuracy, and control energy costs. Given the DCA's goal of facilitating the sharing of knowledge and innovation across the sector, here are our Top Six recommendations for operators looking to take their power monitoring forward to the next level:

- **Simplify your upstream and downstream power decisions** – look out for solutions with a 3D Mechanical and Electrical One-line visualization capability to help you make more informed power distribution decisions.
- **Track your PUE scores in real-time** – there's now an increased regulatory focus on hitting PUE performance targets. Rather than collecting data manually, far better to opt for an approach that can automate this time-consuming process.
- **Remove Power Risk through real-time visibility** – Insist on granular monitoring of your electrical infrastructure, so that you can quickly identify emerging power risks such as overloads, stranded capacity, or misallocated loads.
- **Refine your Critical Data Centre Infrastructure Alerting** – you need estate-wide access to your critical power management data and key power metrics. Insist on immediate access to the kind of data that's simply not available from traditional BMS platforms.
- **Ensure Support for new AI workloads and larger & more complex rooms** – any power monitoring approach has to address the capacity planning needs of the largest operators. Make sure you can support multiple UPS and AC systems per rack, along with the ability to monitor up to eight PDUs for each rack
- **Keep your 3 Phase Power Balancing on track** – make sure your power monitoring approach can visualize 3 Phase loadings on your critical power distribution elements – it's important to keep them balanced.

> A successful power optimisation program begins with visibility. Many data centres still operate with fragmented monitoring systems that fail to reveal exactly where power and cooling capacity is being wasted. That's why the continuous measurement and analysis of electrical consumption across both IT and facility infrastructure is so important

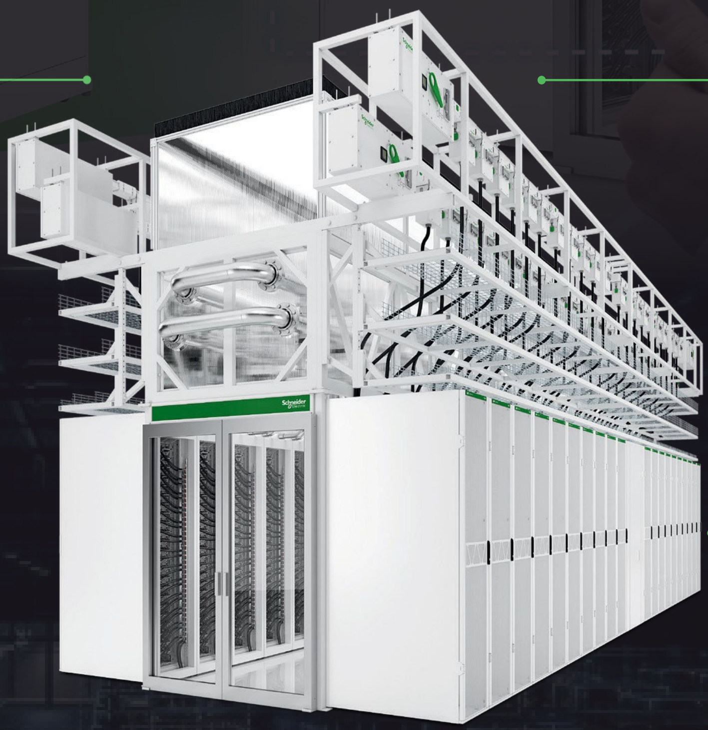
www.ekkosense.com

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