DCS

DATACENTRE SOLUTIONS

DEVELOPING DIGITAL INFRASTRUCTURE IN A HYBRID WORLD

FRANKFURT SPECIAL EDITION

HTTPS://DATACENTRE.SOLUTIONS

SILICON CARBIDE

THE NEW STANDARD FOR HIGH-EFFICIENCY UPS SYSTEMS

legrand

AN ANGEL BUSINESS COMMUNICATIONS PUBLICATION

VIEWPOINT

BY PHIL ALSOP, EDITOR

AI or sustainability – time to choose?

I AM fairly sure that much of the IT industry, and, indeed, the wider world, has already made its decision on this one. AI and all things digital are to be embraced and sustainability must not be allowed to disturb this brave new world – that seems to be the increasingly strong direction of travel. Just as well as, despite the significant progress made by the data centre industry when it comes to sustainability over the past few years, the predicted AI explosion, alongside ever-expanding digital consumption, means that global power consumption is going to increase massively over the next few years. A slightly embarrassing future if sustainability really is the number one objective.

There is to be no environmental revolution – at least not just yet – rather a continued, steady focus on energy efficiency, no matter how much extra energy is going to be consumed over the next few decades.

Of course, renewable energy resources, electrification and other 'green' developments (carbon capture anyone?!) will play an increasingly important role in the AI-driven digital world, but there is to be no slowing down of consumption – and without such a shift in focus, there will continue to be a not so gradual depletion of the planet's finite resources and likely increased climate volatility.

There are entertaining conversations to be had as to how many ways and by how much AI can help reduce corporate and private carbon footprints. But building more, ever larger data centres to house the ever-higher density computers demanded by AI and the quantum computing that follows seems to be at odds with any credible sustainability agenda. In the same way that there is a major resetting of global geopolitics right now, with 80 years of received wisdom heading for the waste pile (maybe to be recycled at some stage in the future?!), perhaps it is time to accept a new truth when it comes to AI and sustainability?

No longer the need to pretend that AI and sustainability are compatible. Instead, an acceptance that AI and digitalisation are the number one priority, and sustainability must know its place in such a new world order.

Of course, energy efficiency will continue to be a major focus for the data centre sector – after all, it makes complete financial success. However, those individuals who somehow imagined that, faced with a choice of a more digital or a more sustainable world, the planet's future would take priority for governments and businesses alike, are destined for disappointment.

This new normal is great news for the data centre industry. It can continue to innovate, continue to underpin AI and digital activities and all without any Jiminy Cricket-like character asking any potentially awkward questions.

I hope you enjoy the articles in the Best of DCS magazine. They are proof that, no matter the actual priorities when it comes to the twin drivers of AI and sustainability, both are fantastic catalysts for data centre technology innovation.

DCS DATACENTRE SOLUTIONS

06 Silicon Carbide (SiC): The New Standard for High-Efficiency UPS Systems

Unlock next-gen UPS efficiency with Silicon Carbide – smarter, smaller, and built for high-demand environments

08 Ending the industry's reliance on generators

The UK government recently stated that data centres are to be classified as critical national infrastructure alongside those like the NHS and power grid. This means they will get additional government support in the case of a major incident.

10 Enterprise data centres in the AI age

As enterprises deepen their investment in AI-driven (artificial intelligence) workloads and high-performance computing (HPC), data centre strategies must evolve

12 Challenges of data centre fire safety

The pivotal role of data centres and server rooms and their fundamental contribution to modern life is undeniable. And, with the amount of data stored doubling every 18 months, it is a rate that is only increasing with the ongoing development of AI

14 Sustainable power - solving the energy challenges of AI

The AI revolution is well underway, but it brings with it challenges for data centre energy consumption, power availability, and carbon emissionsx

18 Thermal imaging for data centres

Data center maintenance teams have a big share in safeguarding the critical resource that customers and businesses depend upon

22 Preparing data centres for a much cooler, more fluid future

Data centre operators know they must constantly adapt to the changing needs of their clients. With the relentless advance of AI applications.

24 How AI is transforming the facilities management industry

Artificial intelligence (AI) is reshaping the facilities management (FM) industry at breakneck pace

26 New technologies, new frontiers

Revolutionising data centre maintenance for the digital age

28 Liquid assets

The significance of liquid cooling for data centre sustainability

30 The data centre of 2030: smarter, faster and more sustainable

Data centres are poised for an era of rapid change, and by the end of this decade they will be very different, with the data centre of 2030 built for sustainability as well as computing power

32 Smart modular data centres

Future-proof IT infrastructure solutions for the digital transformation

36 Data security meets power security

Where information flows in real time and digital operations are mission-critical, data centers have become the backbone of the global economy

38 Light in the data centre: efficiency meets future-proofing

Why intelligent lighting solutions matter today - and set tomorrow's standards evolving demands- lighting has long been part of the solution

42 Specifying liquid-cooled load banks for next-generation data centers

As the demand for high-performance computing and artificial intelligence continues to grow, data centers are adopting liquid cooling technologies to manage thermal loads more efficiently

44 Modular integrated data centres: a solution to the challenges

Modular Integrated Data Centres (MIDCs) offer a transformative approach to DC infrastructure

48 Revolutionising data centre sustainability

The power of liquid immersion cooling technology

52 Base 16 Technology in the data centre – what are the benefits?

As the data centre market transitions to higher data speeds, it is important to consider what solutions provide the speed and bandwidth necessary.

54 It's time to rethink data centre power

The AI revolution is fully under way. But without a top-to-bottom reassessment of power solutions, its progress could slow to a crawl

56 AI and GPU data centres: Navigating the networking challenge

AI for IT operations (AI Ops), a practice that leverages AI to optimise and automate network management tasks, is widely expected to revolutionise network operations

58 Navigating power demand in the age of AI

As the world increasingly relies on digital services and AI, data centres face a growing problem in terms of power demand

62 How must data centres evolve to meet AI workloads?

Artificial Intelligence (AI) and Machine Learning (ML) are becoming mainstream and they can't be avoided or ignored

64 Embracing prefabricated modular data centres for scalable growth

Prefabricated modular data centres have emerged as a powerful solution in the digital age.

68 Schneider Electric continues to drive the sustainability agenda

Schneider Electric, explain the introduction of new model based, automated sustainability reporting features within the company's EcoStruxure IT DCIM software

72 Transforming data centres to meet AI's evolving demands

While 2023 marked a pivotal moment in recognising the vast potential of artificial intelligence, 2024 kicked of what is becoming a truly transformative period

74 Data centres and augmented reality: a blueprint for industry transformation

The days of physically configuring server rooms filled with cryptic codes and blinking lights are becoming a thing of the past as Augmented Reality (AR) begins to make its mark on the data centre sector

78 How eCommerce and fintech firms are leveraging the transformative power of AI-enabled data centres

The impact AI, blockchain and process automation is having on the eCommerce and fintech sectors

DCS DATACENTRE SOLUTIONS

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

Senior B2B Event & Media Executive
Mark Hinds
+44 (0)2476 7189701
mark.hinds@angelbc.com

Sales Manager
Peter Davies
+44 (0)2476 718970
peter.davies@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

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

Circulation & Subscriptions
+44 (0)2476 718970 circ@angelbc.com

Directors
Sukhi Bhadal: Chief Executive Officer
Scott Adams: Chief Technical Officer
E: info@angelbc.com

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

Silicon Carbide (SiC):

The new standard for high-efficiency UPS systems

Unlock next-gen UPS efficiency with Silicon Carbide—smarter, smaller, and built for high-demand environments.

WITH GLOBAL data center electricity consumption expected to more than double by 2030 – and artificial intelligence (AI) workloads pushing power demands to new extremes – the efficiency of power backup systems is under the spotlight like never before. A report by the International Energy Agency (IEA) indicates that data centers consumed 415 terawatt-hours (TWh) of electricity in 2024, accounting for roughly 1.5% of the world's total electricity use. This figure is projected to soar to 945 TWh by 2030 – almost equivalent to Japan's current annual electricity consumption. The primary driver behind this surge is the rapid growth of AI, which is dramatically increasing the compute intensity of workloads.

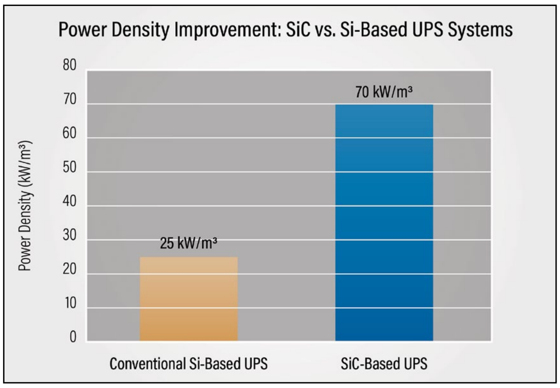
As organizations strive for sustainability, scalability, and resiliency, traditional silicon-based Uninterruptible Power Supply (UPS) systems are reaching their performance limits. To address these escalating demands, Silicon Carbide (SiC), a next-generation semiconductor technology, is transforming the UPS landscape by delivering unparalleled efficiency, power density, and thermal performance. According to Omdia, the market for SiC semiconductors is expected to grow from just 3% of the power semiconductor market in 2022 to 13.1% by 2027.

Why SiC outperforms traditional silicon in UPS applications

While silicon has been the foundation of UPS systems for decades, it has inherent material limitations. Its lower thermal conductivity, restricted high-temperature capability, and higher switching losses diminish its effectiveness for today's high-demand applications. In contrast, Silicon Carbide, a wide-bandgap semiconductor, addresses these limitations, offering significantly improved electrical performance, thermal management, and long-term reliability.

Efficiency and performance

With a wide bandgap of 3.26 eV (vs. silicon's 1.12 eV), SiC can operate at higher voltages, switching frequencies, and temperatures, all while generating drastically lower energy losses. The result is UPS systems that achieve over



98% efficiency in power conversion, significantly reducing waste and long-term energy consumption. This allows data centers to enhance energy efficiency and meet carbon reduction goals.

Power density and size

SiC devices enable the creation of high-density UPS systems that are 2-3 times more compact than traditional silicon-based models due to their thinner drift region and lower on-state resistance. This compactness minimizes the need for bulky components and enhances scalability, making SiC ideal for space-constrained environments such as AI-driven data centers. The more efficient use of physical space allows data centers to allocate more room to revenue-generating compute infrastructure – such as CPUs and GPUs – while also simplifying airflow design and rack layout.

Ultimately, this space optimization contributes to increased processing power per square foot and lower total cost of ownership.

Thermal management

SiC's thermal conductivity is over three times that of silicon, reducing the need for complex and bulky cooling systems. This not only simplifies UPS design but also lowers cooling costs and extends system lifespan.

SiC in action: From niche innovation to industry standard

Initially adopted in the aerospace and electric vehicle sectors due to their extreme operating conditions, SiC is rapidly becoming the standard in UPS systems. Today's SiC-based UPS systems feature:

- Switching frequencies greater than 100 kHz
- Power densities exceeding 70 kW/m^3
- Operational temperatures up to 300°C
- Online double-conversion efficiency ranging from 97–99%

With wider adoption, improved manufacturing processes, and cost reductions, SiC-based UPS solutions have become both technically superior and economically viable.

The real-world impact: Efficiency, reliability, and sustainability

SiC-based UPS systems provide unmatched reliability and thermal resilience in critical data center environments where uptime is essential. With reduced failure rates, simplified cooling requirements, and longer service life, these systems offer a lower total cost of ownership (TCO). They also align perfectly with sustainability goals by minimizing switching and conduction losses while significantly lowering cooling demands, which can account for up to 40% of a data center's energy use. In addition to improved energy efficiency and reduced thermal management needs, SiC-based UPS architectures often feature modular designs that support rapid scaling without requiring complete system redesigns, making them ideal for growing, high-density data centers focused on performance and sustainability.

Understanding the TCO advantage

With fewer components, smaller footprints, and reduced cooling demands, SiC-based UPS systems dramatically lower operating expenses. In many deployments, organizations have reported:

- Energy and cooling cost savings of up to 20% over the UPS lifecycle
- Reduced service downtime due to predictive diagnostics and enhanced reliability
- More efficient use of floor space, lowering real estate and facility costs

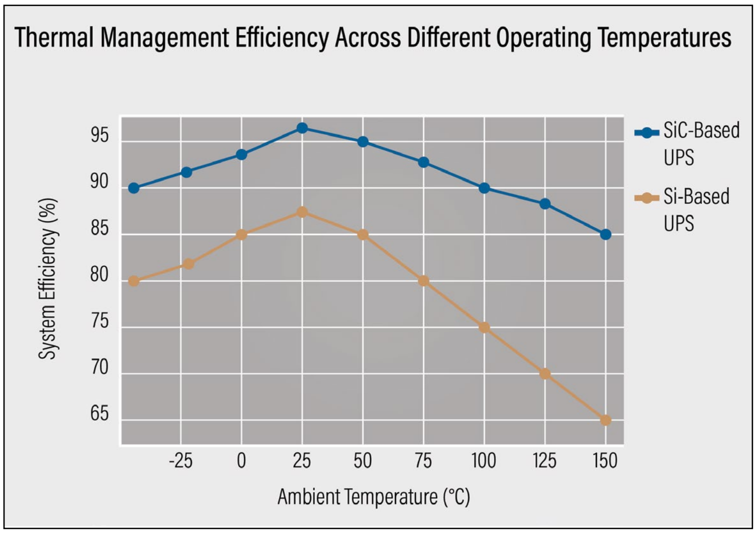
Ready for the future

Silicon Carbide technology is no longer a niche innovation– it's the new standard. Providing best-in-class efficiency, power density, and reliability, SiC-based UPS systems represent the future of resilient and sustainable power infrastructure.

In our increasingly digital and electrified world, SiC enables smarter, smaller, and more efficient UPS designs. As this technology evolves, businesses seeking future-proof solutions should look to SiC to meet tomorrow's power challenges – today.

The road ahead: Embracing smarter power infrastructure

As the data center landscape evolves to support AI, edge computing, and 24/7 digital demands, power infrastructure must evolve too. Silicon Carbide is enabling a new generation of UPS design– systems that are more compact, efficient, and more adaptive, modular, and sustainable.



One example is Legrand's Keor FLEX UPS, a state-of-the-art modular solution demonstrating how SiC-based innovation can translate into real-world resilience and efficiency. With up to 98.4% online double-conversion efficiency, true hot-swap modularity, and intelligent diagnostics based on actual component life cycles, Keor FLEX highlights the potential of smarter infrastructure to reduce both downtime and operating costs. Its front-access design, redundant ventilation, and predictive maintenance features minimize disruption while maximizing uptime – critical in dynamic, always-on data center environments. The system is scalable from 100 kW up to 1.2 MW in a single system, and parallel operation of up to four units reaching 4.8 MW, offering future-ready flexibility for growing demands.

For organizations rethinking their power strategies to improve uptime, reduce emissions, and optimize cost, Silicon Carbide is proving to be more than just an incremental upgrade – it's a foundational shift.

As power demands continue to surge, those that embrace this transformation today will be best positioned to deliver smarter, more scalable infrastructure tomorrow.

To learn more about how Keor FLEX redefines power protection for modern data centers, visit here.

Ending the industry's reliance on generators

The UK government recently stated that data centres are to be classified as critical national infrastructure alongside those like the NHS and power grid. This means they will get additional government support in the case of a major incident. To improve resiliency, data centres have often relied upon generators. However, this does not mean they are necessary for every data centre build. In fact, many data centres can reach industry reliability and availability standards without a single generator on campus.

BY JAMIE CAMERON, ASSOCIATE DIRECTOR, CUNDALL

DAs STATED by Uptime Institute, an unbiased advisory whose mission is to report on the technical requirements for resilience in data centres: On-site power production should be considered the primary power source. Local utility power should be seen as an economical alternative only.

Because of this and the desire from major data centre developers for a standardised reference design, we've seen many operators believe that they need full generator backup for every project. If, instead, they were to consider sites on a case-by-case basis, they would realise they could hit their resiliency benchmarks while improving their bottom line and sustainability targets for some developments.

The industry has largely set itself the target of net zero carbon. While many changes are happening across the full scope of data centre design, an

important issue that is being overlooked is the need for generators. It's time to reevaluate the status quo of data centre design, and that means asking: Do we need generators for our data centre?

Utility supplies and data centres

Data centre operators target a standard reliability metric, '5 9s', which refers to an uptime of 99.999%; this is calculated by looking at the average time it takes equipment to fail (Mean Time to Failure) and the average time it takes to repair (Mean Time to Repair).

Data centre campuses have grown from 10MW IT to 100MW IT to 1000MW IT and above scales. Particularly with machine learning and artificial intelligence development, these sites are becoming connected to the grid at higher voltage levels. This significantly changes the reliability of the utility networks supplying these campuses.

Typical distribution network connections operate between 11-33kV and are usually fed from the same substation; as such, when you get a utility fault, it would usually take out both your A and B utility supplies. This should not be mistaken with transmission networks, which instead operate at a much higher voltage of 132kV or above. Additionally, they have a much higher level of resilience with multiple supply routes, which allows the transmission grid operator to carry out remote switching and resupply sites via alternative routes.

To understand reliability, developers must look at their data. When this information is available from the utility, data centre developers can evaluate the predicted system's reliability by reviewing the historical performance of upstream networks and substations. This includes the mean time between failures (MTBF) and mean time to repair (MTTR) and inputting this real-life data into their data centre reliability calculations. In reviewing the electrical system holistically, you are better able to determine if generator backup capacity is even needed.

The Energy Networks Association (ENA) engineering recommendation P2 Security of Supply recognises the increased resilience of connecting to the network at higher voltages. We are currently seeing proposals for the large data centre campuses that would fall into Class E, referring to connections over 300MW and up to 1500MW capacity. Under the ENA P2 guidelines, the first circuit outage should be restored immediately, and a second circuit outage should be restored immediately to two-thirds of the capacity. It should be noted that immediately is defined as within 60 seconds to allow for switching of the grid.

Progression through reduction

For progression to occur and for our industry to meet the commitments of the COP21 Paris Agreement, we must adopt a reductionist mentality. First comes dismantling the belief that generators are necessary for every data centre site. The end-tenants must lead this change as it is their leasing agreements which enforce full generator backup.

Changing the perception of these tenants by showing them the data will be the first step to achieving change in the industry however, for this the utility companies must be more transparent with their network resilience as well.

A way of solving this is for end-customers to increase the scope of their site selection process, which includes reviewing the grid capacity and adjacency of major transmission substations. This is where the higher reliability of the connections can be provided, as well as reviewing the historical reliability data for these substations. Considering the grid first gives operators the opportunity to make more strategic decisions about location. They can choose data centre locations close to major substations, which enables resiliency to be provided through the grid rather than backup generators.

After this, the industry can take a broader look at what resiliency means. At this point, operators reviewing a site for data centre construction rarely look beyond resilience at a site level. Taking a step back to look at their data centre network as a whole means they can see where several discrete sites connect at different parts of the grid. Through this lens, they can provide redundancy through their IT infrastructure and enable generator backup only for vital network traffic. This would mean that if one site goes down, the network traffic can then be diverted to other data centres connected to either discrete parts of the grid or separate grids entirely. They still achieve resiliency and remove their reliance on generators alone.

Introducing generators to a site costs the operator financially, the designer in complexity, and the environment the embodied carbon stemming from the production and emissions during routine testing. In my own experience, I have worked on data centre sites in which over 50 generators have never been switched on beyond routine testing and maintenance. This is incredibly inefficient. At the same time, some data centre operators have smartly selected sites with highly resilient utility connections that mitigate the need for generators by achieving '5 9s' availability without them.

We have seen real-life data dispel the myth that generators are a necessity. Yet, the industry still chooses to mandate full generator backup throughout the leasing agreements. Despite being able to find other sites better suited to their projects and reaping the benefits of cost savings and better sustainability metrics, belief in generators is being maintained. It is a fact that generators will still be a must for a lot of projects, however, it is only when operators look at their projects individually that they determine the best solution for them. It's time to break the industry's perceived reliance on generators and adopt a better mindset.

Enterprise data centres in the AI age

As enterprises deepen their investment in AI-driven (artificial intelligence) workloads and high-performance computing (HPC), data centre strategies must evolve.

BY PETER MILES, VP OF SALES VIRTUS DATA CENTRES

THE DISCUSSION is not just about choosing between public cloud and private infrastructure but about refining the right mix of solutions to meet increasing performance, security and cost pressures. IT leaders are rethinking infrastructure strategies to ensure they can support the scale and speed required by AI and data-intensive applications while maintaining operational control and regulatory compliance.

AI's growing demands and infrastructure implications

AI workloads demand far greater computational power than traditional enterprise applications, requiring high-density processing, high-speed storage and low-latency networking. Many organisations initially turned to hyperscale cloud providers for AI capabilities, leveraging their scalable compute instances. However, as AI projects move from experimentation to production at scale, enterprises are encountering new challenges - soaring cloud costs, complex security considerations and an increasing need for predictable performance.

For AI training and inference workloads that require sustained, high-performance computing, colocation and private infrastructure often present a more cost-efficient alternative. AI models need uninterrupted access to vast datasets, and enterprises are realising that keeping critical workloads closer to their data sources, rather than constantly moving them in and out of the cloud, reduces cost and latency.

Additionally, as AI applications expand into industries such as healthcare, finance and manufacturing, the need for real-time decision-making is accelerating. This has increased demand for edge computing capabilities that bring AI processing closer to the point of data generation, ensuring lower latency and higher efficiency.

The cost of scale: Managing AI workloads beyond the cloud

Cloud computing transformed IT economics by shifting enterprises from CapEx-heavy investments to OpEx-based consumption models. But as enterprises scale AI applications, the limitations of hyperscale pricing structures are becoming apparent.

High egress costs, unpredictable price fluctuations, and the overhead of continuously running GPU-intensive (graphics processing unit) workloads in the cloud can make long-term AI deployments financially unsustainable.

In response, organisations are segmenting workloads based on cost, performance and compliance needs. Many are adopting hybrid models, leveraging colocation or private cloud for sustained, high-compute AI workloads while using public cloud resources for burst capacity and distributed applications. Strategic workload placement is becoming essential - not as a reaction to cost pressures but as a way to align infrastructure with business priorities.

The role of colocation in enterprise AI and HPC

Modern colocation facilities are no longer just space-and-power providers. They have evolved into critical enablers of hybrid cloud strategies, offering low-latency interconnection to hyperscalers while providing enterprises with dedicated infrastructure for high-performance workloads.

Key factors driving colocation adoption include:

- **AI and GPU Processing –** The demand for AI-ready infrastructure has led colocation providers to build facilities optimised for high-density GPU deployments, liquid cooling and enhanced power availability.
- **Direct Cloud Interconnectivity –** Enterprises are leveraging colocation hubs to establish high-speed, direct links between private infrastructure and hyperscale cloud environments, reducing latency and cloud transfer costs.
- **Data Sovereignty and Compliance –** Many industries face strict regulatory requirements for data locality. Colocation allows enterprises to maintain control over sensitive data while ensuring compliance with jurisdictional regulations.
- **Security Considerations –** By leveraging colocation as part of a hybrid strategy, businesses can build customised security postures that combine the agility of cloud services with the control of private environments.

Performance, scalability and security considerations

As enterprises expand AI-driven initiatives, they are prioritising not only where workloads run but also how infrastructure adapts to dynamic requirements. AI models depend on more than just compute power - they require fast access to storage, proximity to datasets and scalable networking.

High-bandwidth, low-latency environments are critical, which is why many enterprises are moving towards colocation solutions that offer high-speed interconnects to cloud platforms. Security is another key concern, as AI workflows often involve proprietary models and confidential datasets. Many enterprises are using colocation to establish controlled environments that integrate seamlessly into their broader security frameworks, ensuring strict access control, encryption and monitoring.

Scalability remains a challenge for enterprises running AI workloads. Unlike traditional business applications, AI models require an adaptable infrastructure capable of scaling up or down as demand fluctuates. Colocation data centres with modular expansion capabilities allow enterprises to deploy and scale AI clusters more efficiently, avoiding unnecessary expenditure on underutilised cloud resources.

The future of enterprise data centre strategy

As enterprises continue to refine their infrastructure strategies, several key trends will shape the next phase of data centre evolution:

- **AI-Specific Data Centres –** Purpose-built facilities optimised for AI workloads, featuring high-power densities, liquid cooling solutions and dedicated networking capabilities, will become increasingly common.
- **Sustainable AI Infrastructure –** The growing energy demands of AI are pushing organisations to explore renewable energy sources, power-efficient architectures and smarter workload scheduling to manage electricity consumption.
- **Workload Portability and Flexibility –** Enterprises will prioritise infrastructure solutions that allow workloads to move seamlessly between private, colocation and cloud environments as needs evolve.
- **Edge and Distributed AI Models –** As AI becomes more embedded in operational workflows, enterprises will look to edge data centres to decentralise processing and improve response times in latency-sensitive applications.

Preparing for an AI-Driven future

The rigid infrastructure strategies of the past are being replaced by adaptive, workload-driven models. Enterprises that integrate hybrid infrastructure - balancing public cloud, colocation and private environments - will be best positioned to scale AI initiatives efficiently while maintaining cost control, performance, and security.

The challenge for IT leaders is no longer about choosing a single infrastructure model but rather about optimising a mix of solutions that align with evolving business and technological demands. Enterprises must embrace flexibility, automation and interconnectivity to build sustainable, AI-ready data centre environments that will drive innovation into the next decade.

As businesses deepen their reliance on AI, success will depend on infrastructure strategies that are adaptable, scalable, and resilient - enabling enterprises to leverage AI not just for immediate gains but as a long-term competitive advantage.

The challenges of data centre fire safety

The pivotal role of data centres and server rooms and their fundamental contribution to modern life is undeniable. And, with the amount of data stored doubling every 18 months, it is a rate that is only increasing with the ongoing development of AI. The oft-quoted figure of an estimated 90% of data being created in the last two years helps to put this in perspective. It is therefore relatively easy to appreciate just how devasting a data loss can be. The financial consequences for a business can be considerable, as can the potential reputational damage, with business closure in the event of a serious data loss being an all too regular occurrence.

BY ROBERT YATES, HEAD OF BUILDING PRODUCTS FIRE SAFETY UK, SIEMENS

A BASIC CONSIDERATION in maintaining the operation of a data centre is protection from fire, with a solid fire safety concept an essential component in any data centre's business continuity strategy.

Business continuity is a fundamental tenet in the world of the data centre. With companies so reliant on their information systems to run their operations, any downtime can have significant and wide-ranging repercussions. A reliable infrastructure for data centre operations is therefore crucial, not only for the data centre itself but also for the customers that it serves. Integrity and functionality of the hosted computer environment are prime considerations in this process for a service in which such a high level of availability (business continuity) needs to be achieved.

Single source option

The importance of business continuity is perhaps best illustrated by the requirement for a Tier 4 data centre to be fully fault tolerant and have an uptime of 99.995%. In terms of fire safety, this means that a detection, alarm and extinguishing system needs to be designed to keep the business functioning, even if a fire does occur. As such, early detection plays an integral part: the earlier a fire can be detected, the earlier the operators can be notified of the event and the earlier the required technical and organisational measures can be initiated.

Fundamental to this early response in the event of a fire is the interface between detection, alarming, control and extinguishing. Communication between the different elements is essential for the effectiveness of the complete system. This is a strong argument for adopting an integrated fire safety system from a single source.

Aspirating smoke detection

The heavy power loads that are a characteristic of data centres need to be considered in establishing an effective fire safety system. Such loads or a defective component in equipment can quickly lead to overheating or a short circuit. Electrical fires will often start slowly with a long period of overheating and smouldering before flames even occur. To detect overheating and avoid the onset of flames, very early smoke detection is required. If smoke is greatly diluted by high ventilation, aspirating smoke detection (ASD) will provide the earliest possible warning even when the smoke is undetectable by the human eye. Air samples are continuously taken at the danger spots, usually in the circular airflow as well as among the server racks, and carried to the sampling device. As soon as smoke particles are detected by the air sampler, a pre-alarm or an alarm is triggered, depending on the smoke concentration level. The response characteristic is determined according to the application. Sensitivity ranges from normal to high, allowing even a minimal smoke concentration to be identified unequivocally at an early stage.

Aspirating systems can be up to a thousand times more sensitive than a standard point detection system. By combining this level of sensitivity with an environmental learning capability, it is possible for such a system to provide and maintain the optimum operating level and keep unwanted or 'false' alarms to an absolute minimum. Aspirating systems also usually have the capability to monitor their own integrity and, in the event of the system's ability to detect smoke being compromised for any reason, an alert is raised.

Shutting down equipment at the earliest indication of fire will stop even corrosive combustion gases from developing further. In a "gentle" shutdown, intelligent server management is activated to divert valuable data to neighbouring server racks. This can only be achieved by combining an appropriate software/hardware environment with the earliest possible fire detection. The final shutdown of power only takes place once the transfer of data is complete.

If such a "gentle" shutdown is considered too risky, an alternative method – aspirating smoke detection with verification by point-type detectors – may be used. In this system, the cooling system is shut down after pre-alarm while the point-type detectors verify the presence of combustion and trigger the extinguishing system.

ASD systems are available which operate through a dual wavelength technology to verify that particles aspirated in very low concentrations are actually smoke from a fire. Full integration of the device into the fire safety or management system ensures the safest operation possible because all the ASD warnings and possible maintenance messages are available at the management level to allow corrective measures to be taken. Recent developments have seen the introduction of newer generations of ASD technologies which mean that a single device can now cover an impressive area of up to 6,700 m2.. Even for Class A installations with the highest sensitivity requirements, up to 2,000 m2 of detection coverage is achievable with a single detector. This makes the technology ideally suited to applications typified by their large open areas, data centres being a prime example.

Preventing damage to HDDs

One of the fundamental considerations in designing an extinguishing system for a data centre application is to ensure that the chosen agent extinguishes the fire without harming sensitive electronic equipment. Water should therefore be avoided at all times. Furthermore, the agent must be environmentally friendly, safe for people working in the protected area, and cause no harm to the HDDs in operation.

Even though dry extinguishing systems are the best choice to protect data centres, the latest technological findings show that in very rare cases computers and HDDs can face problems after the extinguishing process has been triggered. These problems may range from automatic shutdown of an HDD with no damage after restart to more severe disturbances. Research has shown that the main cause of these problems was the high noise level caused by the discharge of the agent during the extinguishing process.

To address this issue, silent extinguishing technologies have been developed specifically for data centres and server rooms. This enables reliable protection of IT operations and minimises the risk of business interruptions following a fire extinguishing system discharge.

Experience and expertise

The value of the global data centre market was estimated at $187.35 billion in 2020 with growth projections suggesting it will reach $517.17 billion by 2030. Given the fundamental role that data centres play in modern business and life in general, minimising disruption has to be a prime concern. To ensure the highest possible fire safety, it is important to control the interfaces and use the latest scientific findings for the best solution. Having a complete fire safety system from a single supplier with extensive experience in data centre applications can provide the optimum solution in terms of detection, alarming, control and extinguishing.

Sustainable power - solving the energy challenges of AI

The AI revolution is well underway, but it brings with it challenges for data centre energy consumption, power availability, and carbon emissions. By utilising innovative technologies, data centres can provide the foundations for a greener future, powered by AI.

BY MARK YEELES, VP, SECURE POWER DIVISION, SCHNEIDER ELECTRIC UK&I

TODAY, many of the macro trends accelerating data centre adoption are being driven by increasingly compute intensive applications. Generative artificial intelligence (GenAI) and AI workloads, large scale data analytics, and accelerated computing are all impacting demand for physical space, connectivity, power and cooling.

In certain geographies such as Dublin and London this has already led to constraints, and now, both government and industry are working to solve the challenges.

Energy demand

There is no simple answer to the complex issues surrounding power. However, a careful strategy built around the most advanced and energy efficient equipment, AI-enhanced remote monitoring and predictive analytics, and underpinned by innovative energy systems will not only ensure continued, uninterruptible power, but provide a sustainable foundation for the data centres and grids of the future.

According to the IEA, electricity consumption from data centres, AI and cryptocurrency processing could double by 2026, and data centres are currently said to account for around 1% of global electricity consumption. In larger economies such as the US, China, and the EU, data centres account for around 2-4% of total electricity consumption, and in Ireland it could be over 20%. This level of growth has had a measurable impact on emissions too, with some hyperscalers experiencing a carbon footprint increase of 30% due to indirect emissions from data centres.

Geographical impact

Once seen as a centre for data centre excellence, growth and development, Dublin is in a challenging position. A combination of pressure on the national energy grid and ambitious renewable energy targets has effectively halted new data centre connections since 2021. Whilst the consultation process led by the Commission for the Regulation of Utilities (CRU) is ongoing, what's clear is that greater collaboration between government and

industry is needed to solve both the data centre, energy and sustainability challenges facing the country. Similarly in London, a recent report has found that there are up to 400GW of grid requests that may be effectively holding up data centre developments, the strong majority of which (70%) may never see approval. The UK National Grid CEO John Pettigrew has said that «bold action» is needed to create a network able to cope with growing demand from AI.

A new approach to energy generation

Despite an apparent similarity of situations, the two geographies are taking very different approaches. In Dublin the constraints on energy availability has led the CRU to propose strict new rules for energy generation, where new facilities connecting to the grid will be required to provide generation and/or storage capacity to match their requested demand. Additionally, these new distributed energy systems would be required to participate in the electricity market, improving the resiliency and reliability of the grid, while reporting on their annual consumption of renewable energy and their associated CO_2 emissions.

In the UK, encouraged by the measures set out in the UK Government's AI Opportunities Action Plan, operators are beginning to buildout in a host of new locations, including the North of England. In Greater Manchester, for example, Kao Data has been working with local government to bring its £350 million facility to life, while Microsoft's £100 million facility in Leeds brings with it new significant advantages for the Northwest. In Wales, Vantage's multi-billion pound Bridgeport development may also become one of Europe's largest data centre campuses and the company plans to invest over £12 billion in data centres across the UK – creating over 11,500 jobs in the process.

Many of these facilities are likely to deploy renewable energy sources (RES), while using innovative energy infrastructure such as battery energy storage systems (BESS), high-density Uninterruptible Power Supplies (UPS) and distributed energy resources to accelerate both their infrastructure deployments, while overcoming energy constraints.

Energy independence

Due to widespread issues of energy constraints, data centre operators are increasingly planning for greater energy independence. Instead of having the ability to operate for just a few hours the event of a power outage, they're now looking to technology to enable extended periods of self-sufficiency.

This is supported, in part, by the development of new UPS equipment to provide battery backup and advanced power protection. The latest technologies, for example, feature ultra-compact, pioneering high-density designs, and fault-tolerant architecture that can maximise availability, while delivering up to 99% energy efficiency. The ultimate expression of this growing need for energy independence can be found in microgrids - small-scale power systems that operate independently of the grid and generate electricity for a host of use-cases including university campuses, hospitals, industrial manufacturing systems and data centres.

Additionally, microgrids can work in conjunction with traditional power grids, or function autonomously, delivering power from distributed energy resources (DERs) such as solar panels, wind turbines, fuel cells and energy storage systems to exactly where it's needed. Furthermore, they can provide additional grid support by providing energy balancing capabilities due to the variability and intermittency of large scale RES.

Small Modular Reactors (SMRs)

With energy challenges continuing apace, some operators are exploring nuclear power generation. Amazon, Google, and Microsoft, for example, have all engaged in direct partnerships to either situate infrastructure near to these systems, or have made commitments to consume energy from existing nuclear power sources. Another direct avenue for the nuclear path, however, is the development of small modular reactors (SMRs). Based on decades of experience with applications as submarines and icebreakers, there are now around 80 development projects around the world looking, generating capacities of 300MW per system up-to around 500MW.

This new generation of SMRs are based around technologies that are safer and produce less waste, while enjoying a high degree of recyclability in the fuels used. Designs employing technologies such as molten salt, Thorium cycle, and other non-traditional approaches mean safer, more reliable reactors, and are seen as ideal for high demand data centres, and as a basis for microgrids – reducing energy losses, and virtually eliminating the emissions associated with power generation. Further, all of these innovative energy solutions are being combined with AI-enhanced management systems and predictive analytics built on decades of experience in managing complex networks. As such, they allow data centre operators to meet the challenging demand of the AI boom, with high-density and sustainable power solutions, while supporting national grids to develop and bring on more renewable energy in time.

With innovative technologies now providing a vehicle for greater energy independence, the data centres of the future can provide a platform for the responsible and sustainable growth of AI, while playing a critical role in advancing the UK and Ireland's digital economies. Further, by harnessing the power of the ecosystem, data centres can be a catalyst for the green transition, and act as a fourth utility that accelerates the development of renewable energy.

Thermal imaging for data centres

Data center maintenance teams have a big share in safeguarding the critical resource that customers and businesses depend upon. Fortunately, they have one secret weapon that enables them to spot issues in an early stage before they turn into big problems: FLIR thermal imaging.

BY FLIR

THE DATA CENTER MARKET has seen a massive growth in recent years. Spurred by a growing adoption of cloud technologies, AI, IoT, 5G and big data, new data centers are being built across all continents at fast pace.

Whether they are in-house data centers for some of the largest and most influential companies, or built by specialized vendors offering infrastructure services, data centers fulfill a critical role in maintaining the continuity of a business. Downtime of a data center can have an enormous economic impact and needs to be avoided at all costs. In addition to financial loss, reputational damage can be equally impactful, especially when the data center is supporting customer-facing services.

Guaranteeing uptime has become increasingly complex for data centers. With so much mechanical, electrical and electronic infrastructure under one roof, overheating is a major concern, not only because the infrastructure is not using the energy efficiently, but also because overheating can cause a complete shutdown of servers, impacting users around the world, or even data or equipment loss. One of the most reported incidents is the 2013 overheating of a Microsoft data center operating some of its cloud services, including Outlook, which led to services being lost for 16 hours.

➤ Right: FLIR Exx-Series

Maintenance inspections with thermal imaging cameras

The maintenance of a data center today involves much more than IT operations. Power distribution systems and cooling infrastructure are also essential for keeping the data center up and running, and for preventing mechanical or electrical failures and resulting outages.

Many systems that are critical for the data center's operation heat up before they fail.

Temperature is an important indicator of energy consumption and equipment operation, which is why infrared thermography (thermal imaging) is an ideal tool to inspect power consumption, electrical installations, cooling equipment and computing hardware.

Periodic inspections with a thermal imaging camera have become indispensable in predictive and preventive maintenance programs. Thermal cameras help maintenance staff to detect problems in electrical switchgear, motors, HVAC infrastructure, uninterruptible power supplies (UPS), power distribution units (PDU), batteries and generator equipment and all electrical devices that feed the server systems, before these problems turn into serious failures or downtime.

With cloud computing becoming the new normal, and as data centers are growing to great scales, the need for higher computing density and power efficiency is growing as well. Data center owners are seeking ways to increase their capacity, but they also want to reduce costs and energy. Thermal imaging can give them important information on how to optimize energy and space requirements, without causing overheating.

In short, regular inspections with thermal imaging cameras can help maintenance staff to:

- Find and fix hidden problems before they turn into unplanned downtime.
- Reduce the chance of component degradation going unnoticed due to overloaded circuits or loose connections.
- Prevent equipment breakdowns
- Optimize energy management and space allocation

What is thermal imaging?

A thermal camera is a non-contact device that detects infrared energy (heat) and converts it into a visual image. Infrared radiation lies between the visible and microwave portions of the electromagnetic spectrum.

Any object that has a temperature above absolute zero (-273.15 degrees Celsius or 0 Kelvin) emits radiation in the infrared region. Even objects that we think of as being very cold, such as ice cubes, emit infrared radiation. Thermal cameras turn this invisible energy into something that can be seen on a screen and measured.

The benefits of thermal imaging

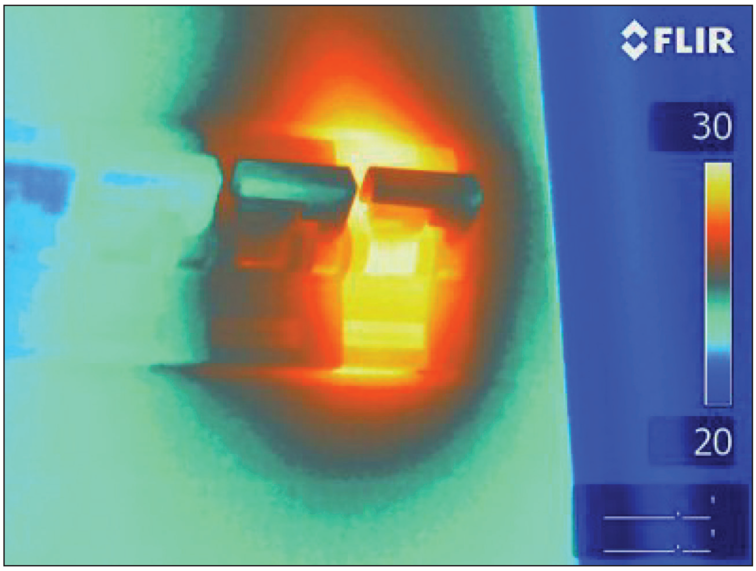
Why would you choose a FLIR thermal imaging camera? Sure, there are other technologies available to help you measure temperatures – infrared thermometers or thermocouples to name only two. But no other tool is as powerful and efficient as a FLIR thermal imaging camera.

See the whole picture

Unlike IR thermometers or thermocouples, thermal imaging cameras enable you to scan large areas for hot spots or temperature differences. Without a thermal camera, it's easy to miss critical parts like air leakages, areas with insufficient insulation or water intrusion. A thermal imaging camera can scan entire electrical installations, buildings, heating or HVAC installations. It never misses a potential problem area no matter how small this might be. They also allow you to compare temperatures of components in the same environment more easily.

Save time and costs

Maintenance of data center installations can be labor-intensive. Because they can easily see larger surfaces, thermal imaging cameras can be the solution to reduce maintenance time speed up inspection rounds, and still see all impending failures, before they turn into costly defects.



Overheated circuit breaker

Inspect without shutting down

Thermal imaging is a non-contact technology. This is a safe method, because maintenance personnel can keep a distance without having to touch hot items. But it also means that inspections can easily be carried out while the equipment is still running or under load. There's no need to foresee costly downtime. Some inspections, for example rotary UPS systems, can only be done during operation, which makes the thermal imaging camera an ideal tool for online inspections.

Report like a pro

Thermal imaging cameras allow users to make professional, more insightful reports of their inspections that also look great for management and customers. Users can compare current inspections with historical data and discover trends. Features like templates, batch processing, image editing and route planning further enhance the use-friendliness of today's reporting solutions.

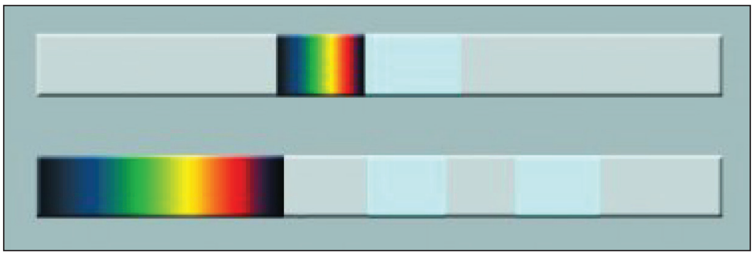
Thermal imaging applications

Thermal imaging is the perfect technology to tackle the wide range of maintenance and inspection jobs for data centers.

Electrical and mechanical systems

Thermal imaging cameras can be used to inspect a variety of electrical or power generation-related systems. Heat is an important indicator for defects in electrical installations. When current passes through a resistive element, it generates heat.

Over time, the resistance of electrical connections can increase, due to loosening and corrosion for instance. The corresponding rise in temperature can cause components to fail, resulting in unplanned outages.



The themal spectrum

Electrical systems can also suffer from load imbalances and increases in impedance to current. Thermal inspections can quickly locate hot spots, determine the severity of the problem, and help establish the time frame in which the equipment should be repaired.

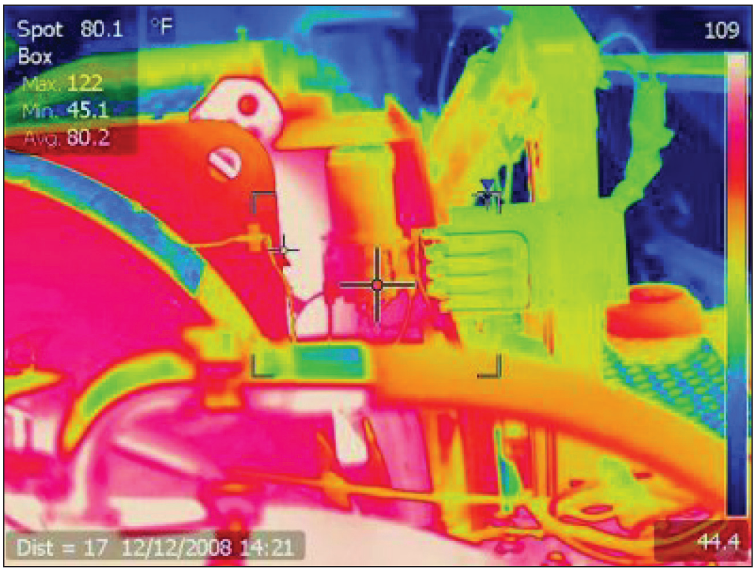
Thermal imaging cameras will help you to spot issues with:

- Overheated connections
- Overloaded or imbalanced circuits
- Damaged switches
- Faulty fuses
- Power supplies
- Battery systems
- Generator systems
- Uninterruptible Power Supplies (UPS)
- Transformers
- Electrical panels
- Resistive load banks

HVAC and cooling systems

To run smoothly and efficiently, data centers need perfectly air-cooled conditions. Data centers are typically making use of a hot aisle/cold aisle layout principle. Server racks are lined up in aisles with the front sides facing each other.

The cold aisles get cold air directly from the Computer Room Air-Conditioning (CRAC) unit from the bottom of the raised floor. The cool air cools down the servers on the racks. Meanwhile, the backs of the servers vent out hot air in the hot aisle, which then returns to the CRAC unit.



Thermal imaging wind turbine transmission survey performed at a 50-meter height

Thermal imaging has become increasingly important to verify the proper hot aisle/cold aisle operation, especially because data centers today are condensing more servers into their racks. Thermal cameras will enable users to see problems such as misaligned ductwork and electrical faults, and then make decisions about corrective actions.

HVAC inspection with a thermal imaging camera can help to:

- Monitor server rack temperature distribution patterns
- Locate misrouted and leaking ducts
- See electrical or mechanical CRAC unit defects
- Confirm the source of energy losses
- Find missing insulation
- Discover AC condensate leaks
- Find internal server fans which are inoperable or damaged

Renewable energy

increasingly improving their use of renewable energy sources, including solar and wind power. These renewable power sources enable data centers to reduce their environmental impact, whilst meeting long-term sustainability goals.

Solar power

The solar panel, the most important part of a solar system, must be reliable and capable of producing electricity for years on end. Unfortunately, solar panels are susceptible to wear. Maintenance professionals therefore use thermal imagers to inspect solar panels installed on rooftops or in solar parks to quickly pinpoint solar panel problems down to cell level.

Anomalies can clearly be seen on a crisp thermal image and - unlike most other methods – thermal imaging cameras can be used to scan installed solar panels during normal operation. Thermal imaging cameras also allow users to scan large areas within a short time frame.

Wind power

Wind turbine components are susceptible to wear and can break down. That's why preventive maintenance and periodic inspections are so important. Thermal imaging is the only technology that allows users to inspect all electrical and mechanical components of the wind turbine and the surrounding electrical system, so they can detect a problem before a breakdown occurs.

Fire protection

Although data center fires are relatively rare, they can have a devastating impact. Data centers may be equipped with fire alarms and firefighting systems, but once a fire starts, asset damage is almost certain. Fixed thermal imaging cameras can identify

Thermal imaging solutions from teledyne FLIR

Handheld thermography cameras

Armed with a FLIR thermal imaging camera, maintenance crews can easily diagnose a wide range of problems across their data center facility.

Available in a wide variety of sizes and image resolutions, FLIR thermography cameras always offer the highest accuracy and user-friendliness to meet the needs of maintenance professionals.

Thermal studio software with route creator

FLIR Thermal Studio Suite is state-of-the-art analysis and reporting software designed to help data center maintenance teams manage thousands of thermal images and videos. Whether you use handheld thermal cameras or unmanned aircraft systems (UAS), the FLIR Thermal Studio suite of software provides the automation and processing capabilities you need to streamline workflow and increase productivity.

The optional Route Creator plugin allows users to plan inspection routes in advance, complete inspections more efficiently and reduce reporting time by 50%.

FLIR CARE

Keep your thermal camera in optimal performance condition and avoid unforeseen downtime with a FLIR CARE service package. Whether you are looking for a performance verification service or a traceable calibration adjustment service, FLIR CARE's traceable calibration services have you covered.

FLIR PROTECT

FLIR PROTECT Extended Warranty and Service Packages give you the peace of mind that your camera is protected from material defects and issues for three years after your factory warranty expires. Choose FLIR Protect+, FLIR Protect Pro, or FLIR Total Protect to bundle a discounted FLIR CARE Service with your FLIR PROTECT extended warranty.

hot spots before they ignite and provide an early warning response to avoid a full-on destructive fire before assets are harmed or safety is compromised.

Physical security

Thermal cameras not only detect hot spots or temperature differences. They also help protect a physical perimeter against unwanted intruders or trespassers. As data centers operate 24/7, they need effective solutions to help them monitor the premises and detect threats around the clock.

Thermal security cameras that offer high contrast, high resolution and long detection ranges are ideal for data center deployments. Unlike standard video cameras, thermal cameras can see in most adverse weather conditions– such as light rain, fog, smoke, or total darkness.

When integrated with video analytics, thermal cameras can distinguish between a human or vehicle. When paired with radar, customers gain redundancy and reduce the likelihood of a false positive. By combining thermal cameras with HD visible cameras, remote operators can review thermal and visible video streams of the scene for improved alarm verification and intruder identification.

Preparing data centres for a much cooler, more fluid future as AI advances

Data centre operators know they must constantly adapt to the changing needs of their clients. With the relentless advance of AI applications, the demand for high-capacity computing is set to grow significantly. The high processing speeds and energy demands of this technology present increased pressure on cooling systems, necessitating new, more robust approaches.

BY OZGUR DUZGUNOGLU, DESIGN AND ENGINEERING DIRECTOR, TELEHOUSE EUROPE

MOST SENIOR PEOPLE within the IT world, regardless of industry, know compute requirements are set to increase substantially as AI and edge computing gain greater adoption. Recent Telehouse research revealed that 89% of IT decision-makers anticipate they will need high-performance, high-density computing systems by 2030. AI workloads are not currently large but could expand quickly, requiring low latency, high-bandwidth connectivity from data centres.

From the outset, data centres have always incorporated cooling mechanisms. Traditionally, the favoured technique has been air cooling, a method that relies on the simple physics of circulating cold air around operational hardware to mitigate the heat produced. But the modern workloads we witness, particularly those spearheaded by AI, are stretching the capacities of air cooling to their very limits.

With natural boundaries to heat transfer capabilities, air cooling technology could hinder the rollout of newer, more energy-intensive, heat-generating services to meet the new demand. This could cap the total volume of AI workloads a data centre can handle unless there is a change to a more efficient cooling technology. This is where liquid cooling technology comes into play.

The inherent limitations of air cooling systems in terms of heat transfer efficiency pose significant challenges. As we look to the near-future, ever-increasing computing power is likely to side-line air cooling. Operators will need to think very seriously about liquid cooling technology.

Although the current demand for liquid cooling remains low & steadyit is poised to assume a central role in the future of digital infrastructure.

With intensifying pressure for cooling systems to be sustainable and eco-friendly, liquid cooling is on the brink of becoming the industry standard. Data centre providers need to ready themselves now for what is an impending shift.

Understanding liquid cooling

Two primary forms of liquid cooling are set to dominate the data centre sector. First, conductive liquid cooling harnesses the potential of liquids to directly extract heat from processor components. This system uses heat sinks attached straight to heat-generating units like central processors. These are then connected via tubes, facilitating fluid circulation and ensuring efficient heat removal.

Conversely, immersive liquid cooling necessitates the complete submersion of servers in a purpose-designed, non-conductive fluid. This enables the heat to be effectively dissipated into the liquid medium. However, this approach requires specific modifications to the servers to guarantee safe immersion. Adopting liquid cooling brings a plethora of benefits to the table. It allows for the augmentation of rack densities, with some reaching an impressive 100KW per rack. This capability enables innovation-driven clients to implement power-hungry workloads vital for their growth. Furthermore, these cooling methods typically have a lower energy consumption, mitigating operators' worries about rising energy costs.

The reduced energy consumption translates into reduction in an improved PUE (power usage effectiveness) rating and reduction in the overall carbon footprint. Additional advantages include freeing up data centre space due to the removal of CRAH/CRAC units and a reduction in noise levels with the elimination of fans.

Yet, the transition to liquid cooling is not without hurdles. The introduction brings its own set of complexities, especially during the design and installation phases. Potential leaks pose significant threats, leading to catastrophic hardware destruction or data loss. The quality of water used in the building's cooling system requires rigorous monitoring, and the financial implications of potential damage and maintenance should not be underestimated. A comprehensive approach, involving specialised equipment, must be considered during the planning phase by the design team.

Planning for implementation

Given this level of investment, how should operators position themselves in anticipation of the rising demand for liquid cooling? A robust and diverse supply chain is crucial. Diversifying suppliers can act as a safeguard against potential component shortages. Additionally, fostering close ties with customers is essential. Through transparent dialogue, operators can gather insights about expected workload trajectories, ensuring all parties are on the same wavelength. AI is also set to become one of the cutting-edge technologies within this process. Operators will lean heavily on advanced systems to manage data centre functions and power consumption. AI will monitor building temperatures and recommend optimisation tactics. As the demand curve rises, the layouts of data centres will undergo a series of transformations, seamlessly incorporating liquid cooling solutions.

Anticipating the liquid future

While some operators may currently be satisfied with their well-established set-ups, especially if their customers are not crying out for demand-intensive workloads, such equilibrium is likely to be fleeting. An increasing number of organisations of all types will soon rely on high-density computing services, aiming to offer more advanced and competitive solutions to their customers.

Given the surging demand generated by AI applications for significant computational power, operators need to be one step ahead. Liquid cooling stands as the most effective means to meet these heavy heat-reduction demands. Companies need to start open dialogues with digital infrastructure providers about their anticipated needs so they ensure that the supply is in sync with demand. Everyone should be looking ahead to the next ten years.

Liquid cooling's appeal is multifaceted, not restricted to its efficiency alone. By closely collaborating with suppliers and establishing a harmonious relationship with the broader supply chain, operators can guarantee the on-time acquisition of essential components. Continued open communication with customers will be utterly essential, so operators have insight into their changing objectives and are able to plan with confidence, meeting sustainability goals in the process. Looking ahead, tools powered by AI will be invaluable, enabling operators to consistently monitor ambient temperatures and proactively implement liquid cooling solutions without compromising on emissions and energy efficiency targets. The clarion call is clear – the time to prepare for liquid cooling has arrived.

How AI is transforming the facilities management industry

Artificial intelligence (AI) is reshaping the facilities management (FM) industry at breakneck pace. Once stereotyped as a highly traditional industry, FM companies are turning their attention to advanced technologies to create smarter, efficient, and more sustainable built environments.

BY DAVID DE SANTIAGO, GROUP AI & DIGITAL SERVICES DIRECTOR AT OCS

THIS SHIFT is happening on a global level; according to a Gartner report, 70% of facilities managers have already elevated digital transformation to a core strategic position. From integrating generative AI language models to implementing multi-agent systems, FM is moving away from manual, reactive practices and becoming a leader in transformative technologies.

While many large FM companies have already adopted tech-first strategies, the future belongs to those that can harness these technologies to benefit their workforce, customers, and the environments they manage. Innovations such as smart buildings, satellite technology, and digital twinning are not only transforming service delivery but also positioning FM as an exciting space for IT, technology, and AI professionals.

AI application in FM

FM companies traditionally deliver a wide range of services categorised into "soft services" and "hard services." Soft services, such as cleaning, catering, and security, focus on making buildings safer and more comfortable for people, while hard services, such as HVAC, plumbing, and maintenance, focus on a building's physical infrastructure. Historically, these areas have operated in silos, each with distinct methodologies and goals. However, technology is now blurring the lines between soft and hard services. With the integration of IoT sensors, computer vision, and AI-powered platforms, the synergies between these two domains are becoming increasingly apparent. For example, environmental monitoring technologies that track air quality (a hard service) can now directly impact safety and comfort (a soft service). This convergence is influencing FM providers to rethink the value they deliver to customers by offering holistic, integrated solutions that enhance the lifecycle of the building while simultaneously improving user experience.

Multi-agent systems

Generative AI has quickly become mainstream, with many people treating it like the next version of the internet. Yet a significant challenge for FM companies is leveraging this technology to enhance job roles, workplaces, and the broader industry. The concept of multi-agent AI systems—where large language models collaborate with smaller, domain-specific AIs—addresses this challenge by providing nuanced, contextual solutions. In FM, this shift is enabling the adoption of innovative service models that are commonplace in the tech industry. For instance, the emergence of the "Facility Success Manager" role, powered by AI, exemplifies how FM is borrowing from technology-driven industries. Armed with insights from IoT sensors, historical data, and real-time analytics, a Facility Success Manager can contextualise information and provide tailored recommendations for a building's unique needs. This role goes beyond predictive maintenance to encompass personalised services, such as optimising energy consumption while improving occupant comfort, thereby reimagining the relationship between technology and human oversight.

Satellite technology

Advancements in satellite technology are also bringing new perspectives to FM—both figuratively and literally. High-resolution satellite imagery combined with AI-driven analysis has enabled FM companies to detect heat loss, monitor infrastructure projects, and map building usage with exceptional accuracy. However, the potential of this approach extends beyond satellites alone. The convergence of various technologies, such as near-infrared imaging, IoT sensors, blockchain, and hyperspectral cameras, is revolutionising how FM services are delivered. For example, motion-detecting IoT sensors integrated with computer vision can optimise workspace utilisation and predict traffic patterns within buildings. Blockchain technology ensures tamper-proof data logs, enhancing transparency and trust in critical areas like compliance. These developments enable FM companies to evolve into strategic consultants for building functionality, offering data-driven insights that extend far beyond the traditional scope of hard and soft services.

A striking example of this convergence is the transformation of security systems. Cameras, traditionally used for surveillance, are now capable of doubling as retail shelf analytics tools, providing valuable insights into stock levels and customer behaviour. Similarly, hyperspectral imaging, once reserved for industrial use, can detect microplastics in water systems, advancing sustainability efforts. These innovations highlight how FM companies are redefining their role, not just as service providers but as pioneers of integrated, technology-driven solutions that enhance the value and functionality of built environments.

Blockchain and augmented reality

The potential of AI extends even further as it intersects with emerging technologies like blockchain and augmented reality (AR). Blockchain is increasingly being used to streamline FM contracts and ensure transparency in procurement processes. For instance, blockchain technology can authenticate energy credits traded under sustainability initiatives, while AI analyses consumption patterns to recommend improvements. Augmented reality, on the other hand, is transforming how FM teams approach maintenance and training. By overlaying building schematics onto physical spaces, AR allows technicians to troubleshoot issues remotely, reducing downtime and improving efficiency. Extended reality (XR) is also gaining traction as a tool for immersive facility tours and workforce training, helping FM professionals upskill in real time.

Looking ahead

As AI adoption accelerates, the FM industry is poised to evolve dramatically over the next five to ten years. Proactive asset management, powered by predictive analytics, will become the norm, enabling FM teams to address potential issues before they arise. Digital twins will play a pivotal role in this shift, allowing FM companies to simulate the impact of environmental changes on building performance and recommend preventive measures. Fully automated workspaces, equipped with IoT sensors and AI integration, will adapt in real time to occupant needs, optimising everything from lighting to catering operations. Beyond this, FM companies are expanding their service scope to encompass areas like employee well-being and workplace productivity, measuring and enhancing factors such as stress levels, collaboration patterns, and satisfaction through AI-driven insights.

An example of this future is already taking shape through pilot projects. Leading FM companies are

exploring AR-driven live maintenance guidance combined with AI diagnostics to enhance repair accuracy. Another use case is the application of hyperspectral imaging being used to detect environmental contaminants in real time, contributing to safer and more sustainable buildings. These initiatives illustrate how FM companies are transitioning from reactive service providers to proactive innovators, fundamentally changing how built environments are managed and interacted with.

While these advancements are transformative, they also raise important questions about responsible AI deployment and workforce readiness. AI is not about replacing jobs; it is about augmentation. The role of technology is to empower the workforce, equipping them with tools to perform their roles more effectively. For example, a Facility Success Manager enhanced by AI can manage more complex systems while maintaining a human-centric approach to building management. Economic data underscores this potential: McKinsey Global estimates that AI could create up to 50 million new jobs by 2030, emphasising the importance of continuous learning and adaptation. Ethical considerations, particularly around data privacy and emotional AI, also demand attention as FM companies embrace these technologies.

Frameworks such as the European Union's AI Act provide essential guidelines for ensuring responsible development and deployment. Looking ahead, it is clear that AI will continue to redefine FM. The convergence of digital twinning, multi-agent paradigms, and emerging technologies will reshape how built environments are managed, moving the industry toward a future marked by adaptability, innovation, and sustainability. FM companies that approach this transformation with both ambition and responsibility will not only thrive but also set new benchmarks for excellence in service delivery. As the AI revolution unfolds, the potential to create smarter, more sustainable built environments is boundless. For FM, this is not just an evolution—it is a reinvention.

New technologies, new frontiers

Revolutionising data centre maintenance for the Digital Age

BY ANNA MAZZOLENI, GLOBAL PRODUCT MANAGER, ELECTRIFICATION SERVICE, ABB

IN 2023, it was estimated that there were now 5.4 billion people online globally. With more of our lives increasingly being lived virtually, businesses, governments and consumers are now undeniably reliant on digital services. Forming the foundation of these are droves and droves of data centers which have become an integral part of our data-driven society, making the stakes for their operators higher than ever.

Cybersecurity risks notwithstanding, having a firm grasp on inherent operational and structural risks is critical. In the context of data centers, this means being able to provide a constant and reliable stream of data services. Yet, the threat of power failures that cause data center outages remains a significant challenge — and the resulting downtime that comes with it is not only highly disruptive, but costly.

With our dependence on data centers showing no signs of abating, we need to get ahead of these challenges. For one, enhanced approaches to asset servicing and maintenance through the use of innovative technologies, such as augmented reality (AR), are a way to mitigate these risks. Together, these offer a promising path forward for safer, smarter and more reliable data center operations.

Enabling the next frontier in servicing innovation
It's well-chronicled across a myriad of industries that the pandemic catalyzed novel ways of doing business — asset servicing and maintenance are no exception to this, and we've seen this specifically through the use of AR.

With an AR-enhanced maintenance app on their smartphones, technicians can see more than what meets the eye: technical information and servicing guidance through images, instructional videos and documentation are overlaid on the physical equipment in front of them.

This is done in such a way that enhances their perception and interaction with their surroundings, all while preserving the necessary level of environmental awareness to ensure the safest user experience while minimizing human error.

Such maintenance apps can also be designed to be device-agnostic, accessible across mobile, tablet and desktop, but also hands-free devices such as industrial smart glasses. This would enable engineers and technicians to operate hands-free, improving both safety and productivity as they conduct their assessments and repairs in the harshest environments.

Scaling servicing capabilities

Through AR, technicians also benefit from real-time, remote support from experts, assisting them with complex repairs and maintenance tasks. For data centers, especially, where every second of downtime translates to lost revenue, remote support is especially vital as it provides access to near instantaneous expert support and solutions to the technical issue at hand. This enhances first-time fix rates while also extending service reach to remote and under-served areas.

With data centers often located in remote locations, the benefits as it pertains to cost- and time-efficiencies are clear. But these equally extend to sustainability considerations: Remote servicing saves up to 171g of CO_2 emissions per passenger-kilometer, totaling to 332 tons of CO_2 emissions per year by reducing up to a third of customer site visits by field service engineers.

At the same time, the use of technology to power remote support capabilities equally translates to more opportunities for immersive remote training which is proven to deliver better knowledge acquisition and retention. This, compounded by self-learning on virtual systems fosters a culture of first-rate support and servicing. AR can also power hands-on training experiences, enabling technical servicing teams to scale their junior talent. Junior technicians benefit from practical training opportunities through a virtual system before applying their skills in the field.

Taking the smart approach to maintenance

Businesses tend to maximize their capital expenditure by running their equipment, unknowingly, to the point of failure, which leads to longer term losses. In fact, this can cost up to ten times more than investing in regular maintenance and often contribute to more severe outages — and the older the equipment is, the greater the severity. According to the Uptime Institute's 2023 Global Data Center Survey, such outages can cost from US$250,000 to more than US$1 million.

Regular monitoring of facilities, especially mission-critical equipment, is essential for guaranteeing reliable service provision — and smart maintenance can better support that.

A preventative approach involves replacing older, non-digital circuit breakers with intelligent, sensor-enabled breakers linked to cloud-computing

platforms. Real-time data and analysis on asset condition and performance can prevent potential issues before they arise.

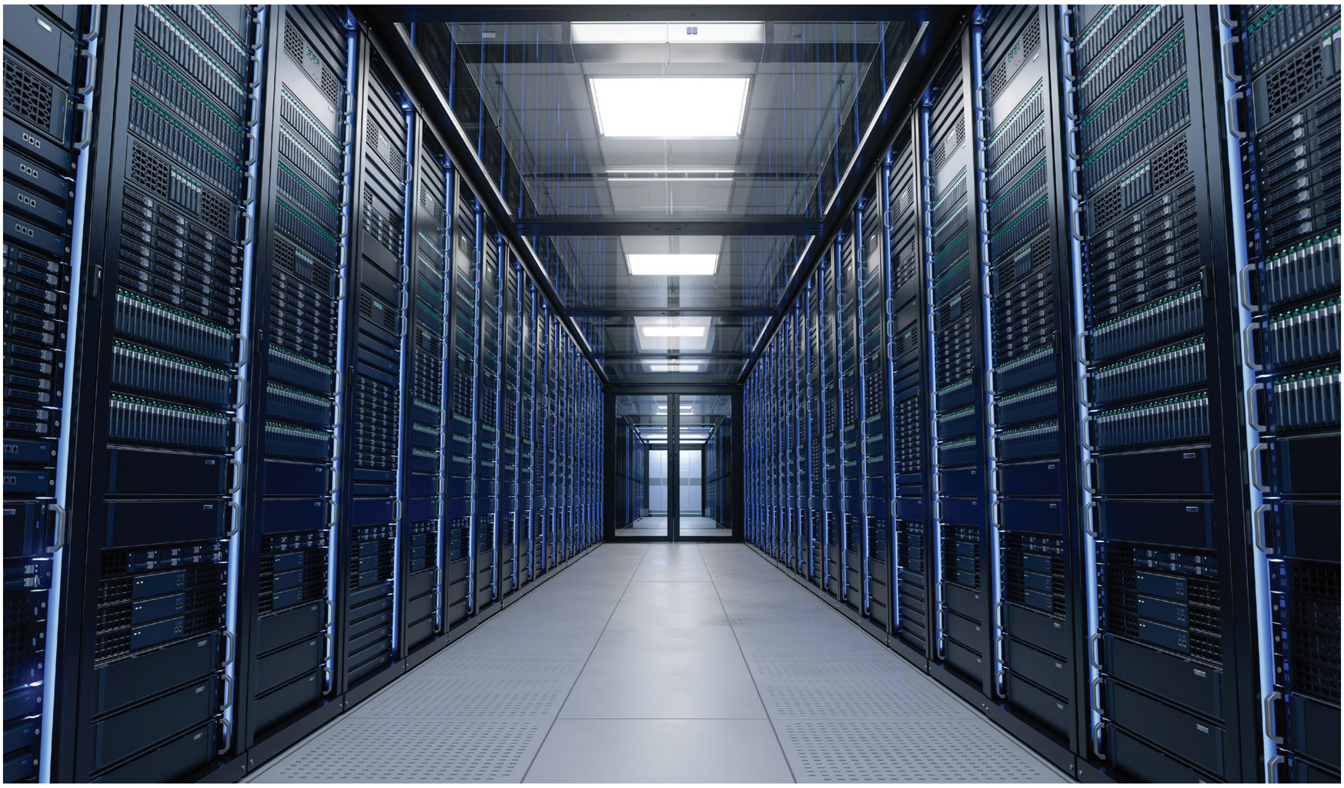
According to the Deloitte Analytics Institute, predictive maintenance increases productivity by 25%, reduces breakdowns by 70%, and lowers maintenance costs by 25%. Despite these benefits, less than 50% of global manufacturers use predictive maintenance technologies. The gap itself presents an opportunity for businesses to increase productivity and build a competitive advantage.

Digitally driven, digitally resilient

For businesses today, the incorporation of more digitally-integrated approaches that can ensure greater reliability is no longer a luxury but a necessity — and it's simply no different for data centers. With the cost of outages increasing and data center downtime costing nearly US$8,000 per minute, every second of downtime translates to revenue lost and reputational damage that's hard to come back from.

As the backbone of our increasingly digital world, data centers are pivotal in ensuring the smooth operation of critical services, be it in healthcare or financial services. To keep them running, we need to rethink our approach to servicing and maintenance, bringing outdated methods into the future. Powerful innovations such as augmented reality unlock new opportunities for predictive, data-driven maintenance all while also significantly reducing environmental impact and operational costs.

Ultimately, taking a more digitally-enabled approach to asset maintenance is essential to navigating the challenges associated with increased digitalization. Thankfully, innovative technologies are set to play a pivotal part in ensuring data centers to remain resilient and operationally-ready to meet the growing demands of data services.



Liquid assets

The significance of liquid cooling for data centre sustainability

BY ALISTAIR BARNES, HEAD OF MECHANICAL ENGINEERING AT COLT DCS

AMONGST the escalating demand for more energy from data centre (DC) customers, maintaining a low Power Usage Efficiency (PUE) rating by DC providers is paramount. With low PUE comes the ability to identify cost saving measures and recognise areas for improvement, whilst also reducing carbon footprints and impacts on the environment.

Yet, as more organisations deploy high performance computing (HPC) and greater IT workloads to support their business requirements, the by-product of heat from power usage is also on the rise. In order for DC providers to ensure that their equipment and systems remain in top condition to serve demand, suitable cooling and heat management is a must.

One effective strategy is building DCs in cold climates and leverage free cooling methods. When the external temperature of a DC is considerably lower than the internal temperature of a facility (as little at 2°C), natural air can be used to directly cool the site and therefore, reduce – if not completely eliminate – the need of mechanical cooling processes. As a result, energy consumption is decreased significantly, and equipment life can be extended reducing embedded carbon. However, geographical constraints such as proximity to key markets and infrastructure mean that not every organisation can benefit from this advantage. For those unable to build in cooler regions, advanced cooling techniques offer a promising alternative.

Among these, liquid cooling has emerged as a particularly effective solution, offering significant advantages over traditional air-based methods. Let's delve into the options and benefits of liquid cooling in modern DCs.

What is liquid cooling?

Rather than utilising traditional air-cooling technology, liquid cooling is a method that uses liquids such as water to conduct heat away from IT equipment.

Compared to air, liquid has a much higher heat capacity (1.004 vs 4.18). Therefore, by using liquid cooling methods, the absorption and transfer rate of heat released from IT equipment will be much higher and more effective when dealing with increasing levels of power.

The different types of liquid cooling

Several liquid cooling methods are available for DC providers. Rack-based liquid cooling circulates coolant through racks to absorb and remove heat directly from servers. Direct-to-chip cooling, including liquid-to-chip methods, targets heat sources by circulating cool liquid through a 'cold plate' in direct contact with components like GPUs. In contrast, immersion cooling submerges entire servers in a thermally conductive liquid, efficiently dissipating heat across all components. Let's delve into this in more detail.

Rack-based liquid cooling:

Rack-based liquid cooling is a method where coolant is circulated through the racks that house servers, absorbing and removing heat directly from the equipment. This approach is highly efficient, allowing for better heat management and enabling higher server densities within DCs. By reducing reliance on traditional air-based cooling, it can significantly lower energy consumption and operating costs.

However, implementing rack-based liquid cooling requires significant upfront investment and infrastructure modifications. Additionally, managing potential leaks and maintaining the cooling system can be complex, posing challenges for DCs not initially designed with this technology in mind.

Immersion Cooling:

Immersion cooling submerges servers in a thermally conductive, dielectric fluid that absorbs excess heat and stabilises IT equipment temperatures.

The process can be single-phase, where the liquid remains in a constant state, or two-phase, where the liquid evaporates and re-condenses to remove heat. Immersion cooling offers significant benefits, such as drastically lowering PUE levels – sometimes as low as 1.10 – by eliminating the need for traditional air cooling systems like CRAC units, which also reduces noise and saves space. However, implementing immersion cooling requires specialized infrastructure and careful management of the dielectric fluid.

Direct-to-Chip Cooling:

Direct-to-chip cooling involves circulating cool liquid through a system that directly contacts the chips and other heat-generating components. This method is more efficient than immersion cooling in dissipating heat at the chip level because the circulating fluid is typically cooler.

Direct-to-chip cooling allows for higher compute densities without needing additional space, making it an excellent option for upgrading existing air-cooled DCs. However, it still requires supplementary cooling for other IT equipment, typically using chilled air, which adds some complexity to the overall cooling strategy.

Hybrid approaches for a balanced PUE

While liquid cooling is an innovative solution, this technology is not yet in a position to completely replace air cooling in DCs. Even if equipment is cooled by liquid, heat will be transferred to it and some of this will be dissipated into a room or surrounding space where air will be required to remove this.

Therefore, a hybrid approach is the best option where liquid and air techniques are used together to offer the best balance of PUE performance.

Furthermore, it is important for organisations to know what options are available to them by working with a DC partner that fully understands their needs. DC providers must be able to review and manage their operations in real-time to ensure efficiency and to meet corporate sustainability goals.

By using modern data analytics tools, businesses can monitor power usage, internal and external temperatures, and electricity usage for cooling to optimise processing loads for cost-effectiveness and proactively monitor equipment maintenance.

Hybrid solutions are increasingly being adopted to combine the strengths of different cooling methods, improve PUE, and pave the way for a more energy-efficient future.

The data centre of 2030: smarter, faster and more sustainable

Data centres are poised for an era of rapid change, and by the end of this decade they will be very different, with the data centre of 2030 built for sustainability as well as computing power.

BY IAN JEFFS, UK&I COUNTRY GENERAL MANAGER AT LENOVO INFRASTRUCTURE SOLUTIONS GROUP

THE REASON for this is simple, the power demands of data centres are rising rapidly. Data centre technology already accounts for 2% of energy use worldwide, according to the International Energy Authority (IEA), and by 2026, that figure could well double, with the demands of data centres equalling the electricity consumption of Japan. In Britain, the National Grid predicts that data centre power consumption could increase six-fold over the coming decade. These predictions are set to drive a surge in demand for sustainable computing.

One of the reasons behind the exponential rise in electricity demand is artificial intelligence (AI) and in particular generative AI, alongside other innovative technologies such as quantum computing. Generative AI's appetite for energy is enormous. Research estimates that generating just one image with AI uses as much energy as fully charging a smartphone. But this must be weighed against the important role AI has to play in the battle against climate change, with AI expected to deliver breakthroughs in clean energy (for example designing nuclear fusion reactors) and in other areas such as limiting methane emissions from waste. The challenge is clear: the world must find a balance to enjoy the benefits of AI, while also containing its impact on the environment.

So how can we build data centres fit for a cleaner future? Business leaders need to take an overview of the real energy impact of data centres, taking

in everything from how servers are cooled and how that energy might be reused, to how they are shipped and the mix of energy used. Understanding the true impact of data centres is the first step towards a smarter, more sustainable future.
Powering progress

With generative AI appearing in software ranging from email apps to internet search, the energy demands of the AI industry is going to continue rocketing, with one study in the journal Cell suggesting the power demand of AI alone could match the demand of the Netherlands by 2027. The compute power required to train AI doubles every six months, and Gartner® predicts[1] that 'by 2030, AI could consume up to 3.5% of the world's electricity'. The IEA's report suggests that adding generative AI to search (as companies including Google are rushing to do at present) could multiply the energy demands of internet searches by 10.

All of this requires the technology industry to design carefully for sustainability, not just at the chip level, but at the server level and data centre level. It's also worth bearing in mind that there is a flipside to this in terms of the environmental benefits of innovations sparked by new technologies. Emerging technology such as quantum computing will be more energy-efficient and could also mean that problems are solved exponentially faster than classical computers. Both quantum and AI are expected to drive rapid innovation in everything from demand response in the electricity grid to photovoltaics to electricity generation technology. The 'smart grids' of the future will be powered by AI. In building decarbonisation, to take one example, Mckinsey estimates that AI can accelerate the process 100-fold compared to existing technologies.

Smarter systems

The data centre of the future will be designed from the bottom up with sustainability in mind. Technologies such as warm water cooling enable high performance with far less energy use, provided a host of power consumption benefits. For instance, in data centres that use warm water-cooling, there is much less need for high-speed fans to dissipate heat. Air-cooling systems can often consume vast amounts of energy in themselves.

Furthermore, any wasted heat can be reused effectively because warm water-cooling systems produce heat waste at a temperature that is more easily reused for other purposes. Not only will this improve energy efficiency across an entire data centre facility, but this energy can be successfully recycled in sustainable ways in the wider community. By 2030, recycling the heat from data centres will become the norm, from heating nearby buildings and swimming pools to piping warm liquid under roads and walkways to melt ice.

The data centres of the future will also be built around renewable energy, from renewable sources to solar panels on roofs, driving towards a future of carbon-neutral or even carbon-negative operations. Organisations will adopt 'as a service' approaches to AI to improve efficiency, and increasingly businesses will harness the power of AI to optimise electricity consumption in data centres themselves.

Just as AI will assist in demand response in the electricity grid, within the data centre, algorithms can help with optimisation, predictive maintenance and energy management. This can reduce energy consumption by improving cooling efficiency, minimising waste and optimising resource allocation.

Circular thinking

When designing and building a data centre fit for the future, it's key for business leaders to take a view across the whole lifecycle of their data centre and the servers that will work in it. Everything from how components are designed and manufactured to how they are shipped, deployed and disposed of at the end of their lives matters, and taking a holistic view is key to making real sustainability gains.

Through this decade, asset recovery services (ARS) and recycling of computer equipment will grow in importance. Other markets in Europe have already implemented legislation to reduce the environmental footprint of digital technology.

For example, France introduced a mandate that 20% of IT devices bought by organisations need to be refurbished, with a target of 40% by 2040.

Manufacturing products regionally to cut shipping miles will also be a key differentiator. The current shift towards 'as-a-service' approaches to everything from hardware to software will continue, with business leaders focusing on avoiding overprovisioning to cut carbon emissions.

Towards net zero

The challenge facing the technology sector is clear. AI's hunger for energy is set to catalyse a race to adopt smarter and more sustainable approaches in the data centre as this decade unfolds. By 2030, we will see a new kind of data centre powered by renewables, and integrated thoroughly into the community around it with excess heat efficiently re-used to heat buildings and swimming pools. They have the potential to power a new era of quantum and AI, which will help unearth breakthroughs to tackle global issues, such as climate change.

FURTHER READING / REFERENCE

- 1. Gartner Says CIOs Must Balance the Environmental Promises and Risks of AI. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission.

Smart modular data centres

Future-proof IT infrastructure solutions for the digital transformation

BY DATWYLER IT INFRA

DIGITISATION is progressing at a rapid pace – and posing enormous challenges for businesses. With automation, robotics, 5G, the increase of intelligent devices (IoT) and especially the sudden expansion of artificial intelligence (AI), more and more data are having to be analysed, processed and handled.

Having enough computing power – anywhere and at any time – has become a central prerequisite for business development. Powerful access nodes, right at the edge of the network, in the immediate vicinity of the data sources, are important and must nowadays meet the maximum requirements for speed, short latency times, reliability and security.

With its Smart Modular Data Centres (SMDC) Datwyler offers innovative, scalable and future-proof all-in-one IT infrastructure solutions, which have been developed especially for edge computing applications. These pre-configured plug-and-play solutions enable companies to process their data efficiently and securely directly at the point of origin.

From collecting and analysing production data to hosting local AI applications: The SMDC solutions from Datwyler provide a versatile way of covering all modern requirements.

In addition, these solutions comply with international industry standards for data centres and ensure that each individual component and element can be taken out of service for maintenance or repair without affecting the critical environment or IT processes.

Real-time data processing at the edge

The key advantage of an SMDC is its short latency times: Data are processed in near to real time thus guaranteeing fast reactions and optimum performance. At the same time, the operation of the data centre remains independent of the network, enabling local aggregation and storage of critical data. This increases reliability and reduces the risk of data loss.

Every Datwyler SMDC is a complete data centre solution that has been tailored to meet the user's exact requirements – fully integrated and from a single provider. It includes a flexible rack configuration – from a single rack, through multi-rack, to multi-module solutions – with integrated power distribution, an uninterruptible power supply, an efficient cooling system, fire extinguishing system, blanking panels and sealing strip as well as a data centre infrastructure management (DCIM) system that enables complete remote monitoring and control.

The entire infrastructure can be deployed or relocated within a few hours. Thanks to its modular design, every system has the flexibility to adapt to growing demands ("pay as you grow").

Energy efficiency

An important aspect of the SMDCs is their energy efficiency. The Datwyler solutions are setting new standards with a power usage effectiveness (PUE) value as low as 1.3. This low PUE value is achieved thanks to a highly efficient cooling system, optimised airflow and intelligent energy distribution. This reduces not only the energy consumption but also the CO_2 footprint, thus contributing to a sustainable IT strategy.

The focus is always on security: Integrated access control with both card and keypad input, intelligent PDUs for energy management and real-time status monitoring ensure reliable, secure operation.

In short: Datwyler's Smart Modular Data Centre is an ideal solution for companies who, in an increasingly digital world, need powerful, scalable and reliable IT infrastructures. With the SMDC you are opting for a future-proof solution that will provide you with maximum flexibility and reliability.

About Dätwyler IT Infra AG

Datwyler IT Infra is an international company with headquarters in Switzerland and affiliates in Europe, the Middle East and Asia. Datwyler enables organisations around the globe to run their IT and OT infrastructures seamlessly and scale their business with ease. The well-established company operates as a provider of innovative system solutions, products and services for data centres, fibre networks and intelligent buildings, as well as acting as a subcontractor or general contractor covering the entire value-added chain with tailor-made solutions and outstanding expertise. Datwyler, established in 1915, has a global team of approximately 1,000 employees.

ITinfra.datwyler.com

Data security meets power security

In a hyperconnected world, where information flows in real time and digital operations are mission-critical, data centers have become the backbone of the global economy. In these environments, energy is not just a resource – it's a guarantee of continuity, security, and trust.

BY HIMOINSA

NOTHING CAN FAIL. Every second of downtime translates into financial losses, reputational damage, and operational risks. That's why having a safe and reliable energy solution is not an option; it's an absolute necessity.

HIMOINSA, an strategic partner for Data Center power supply

In this context, the genset plays a key role as a critical power supply backup, ensuring that even in the event of a utility power failure, the data center maintains uninterrupted operations. In this regards, HIMOINSA is an strategic partner for power supply in Data Center sector. Why?

HIMOINSA, part of the Yanmar Group, is a manufacturer of Power Technology Solutions that designs and produces generator sets, battery storage systems, lighting towers, automatic transfer switches, monitoring controls and accessories, for backup and continuous power supply.

From mission-critical to backup and continuous power, HIMOINSA provides complete, reliable and efficient diesel and gas power generation solutions wherever reliability is needed. With thousands of installations worldwide and different applications and sectors such as healthcare, hotels, data centres, manufacturing, mining, construction and independent power plants, the company has delivered power solutions for the international market during the last 4 decades ensuring cutting-edge technology, high performance, lower NOx, the most efficient systems and simplified maintenance.

Data centres consume a significant amount of electricity. Due to the fast-growing artificial intelligence and digitalization, the electricity needs from data centres exceeded 2% of the world's demand and projected to rise to 13% by 2030.

HIMOINSA explores energy optimization strategies for Data Centers back-up & continuous power. Emphasizing the recovery of waste heat to be used in our generator sets. It highlights how this recovered heat can significantly reduce operational fix costs, enhance energy efficiency, and improve sustainability. Its engineering team advise about how the integration of heat exchangers, and the role of solar energy in auxiliary power solutions impact on Power Usage Effectiveness (PUE).

The company has recently launched the HGY series, which has born as a key player in the power generation industry, combining cutting-edge technology in power generation, ensuring highest reliability and superior performance, leading the market in critical mission projects. Packed by powerful GY engines, the HGY series covers power nodes from 1250 kVA to 4000 kVA, and is ready for alternative fuels like hydrotreated vegetable oil (HVO), and gas and hydrogen in the future.

Engine: heart of outstanding performance

At the heart of every genset there lies an internal combustion engine, converting fuel into energy. The GY series is a new family of 12, 16 and 20 cylinder engines managed by state-of-the-art electronic control, and with a common rail injection system. It combines the unique combustion technology for high-speed engines with the reliability of medium-speed engines, ensuring high performance and efficiency, supporting operations in extreme conditions, compact footprint, multiple fuel capabilities and low emissions.

The GY engine series is a perfect example of Quality, Efficiency, Reliability and Durability. Common-rail Fuel Injection Systems and High-Pressure Pump: generates up to 2,200-bar injection pressure for maximum efficiency, optimizing the combustion pressure curve through multiple injections.

High Power Density: The HGY Series delivers exceptional performance in terms of emissions and power density (up to 37.9kWm/L). The high specific power density reduces footprint and has a great impact on the total life cycle carbon footprint.

Single Cylinder Prototype: The design of the piston head and the fluid analysis is key for engine efficiency and performance. Yanmar has spent thousands of engineering hours in this stage of the development.

Fast Response: Fuel injection makes a huge difference in the power generation application where the speed of response is vital. The ECU can react to sudden changes in speed in milliseconds. Engine starting in less than 8 seconds. This engine can perform within ISO8528-5 G3 class conditions.

Sustainability & pue reduction (power usage effectiveness)

Considering that sustainability has become a top priority, our HGY Series has been designed to ensure a sustainable power supply for mission-critical project:

Emission control

To cope with the increasing scrutiny for environmental impacts, our HGY generators are equipped with exhaust after-treatment systems to meet European, German, and British regulations for mid-combustion plants operating more than 300 or 500 hours. The HGY series includes an EPA Tier 2 certified version for emergency applications in the United States, as well as compliance with NEA regulations in Singapore.

Fuel flexibility

HGY generators are compatible with renewable fuels such as HVO, its vegetable origin reduces CO2 emissions and lowers the impact on the environment. The HGY series will accommodate the future development of gas and hydrogen generators, significantly reducing emissions and carbon footprint. This is a remarkable step towards environmental responsibility and will help support end-users in their strategy to attain Net Zero.

Environmental Product Declaration (EPD)

To validate the reduction in emissions, EPD is provided, which includes comprehensive information on the environmental impact at each stage of a product's life cycle, helps customers to make sustainable purchasing decisions and implement low-carbon strategies.

FUEL CONSUMPTION | BEST-IN-CLASS

Outstanding performance with optimized fuel consumption is achieved through GY engines' new common rail high-pressure fuel injection system, piston design and high-power density. The HGY series boasts outstanding, state-of-the-art technology that significantly reduces operation costs and emissions. Thanks to their remarkably low fuel consumption, GY engines are ranked best in class in terms of fuel oil consumption (FOC) optimisation, making them one of the most efficient and competitive options on the market. Fuel consumption remains unchanged with HVO.

Reliability

In data centres, any power failure may affect the continuous operation of IT equipment, cooling systems and related infrastructure, directly leads to potential data loss and innumerable revenue impacts. Reliable power supply is absolutely critical for data centre security where every second counts. The HGY series is designed to provide comprehensive protection against power failures, offering a secure, continuous and reliable power supply. With a response time of less than eight seconds, the HGY series is fully compliant with ISO8528-5 G3 class, ensuring no interruption in the flow of data, minimises the risk of failure and maximises uptime.

Serviceability & extended services intervals

Easy maintenance and accessibility are seamlessly integrated into the design of the HGY series. The HGY Series offers extended service intervals, featuring a top overhaul interval of 10.000 hours and a major overhaul interval of up to 30,000 hours for continuous operation.

Light in the data centre: Efficiency meets future-proofing

Why intelligent lighting solutions matter today - and set tomorrow's standards. Evolving demands - lighting has long been part of the solution

BY TRILUX

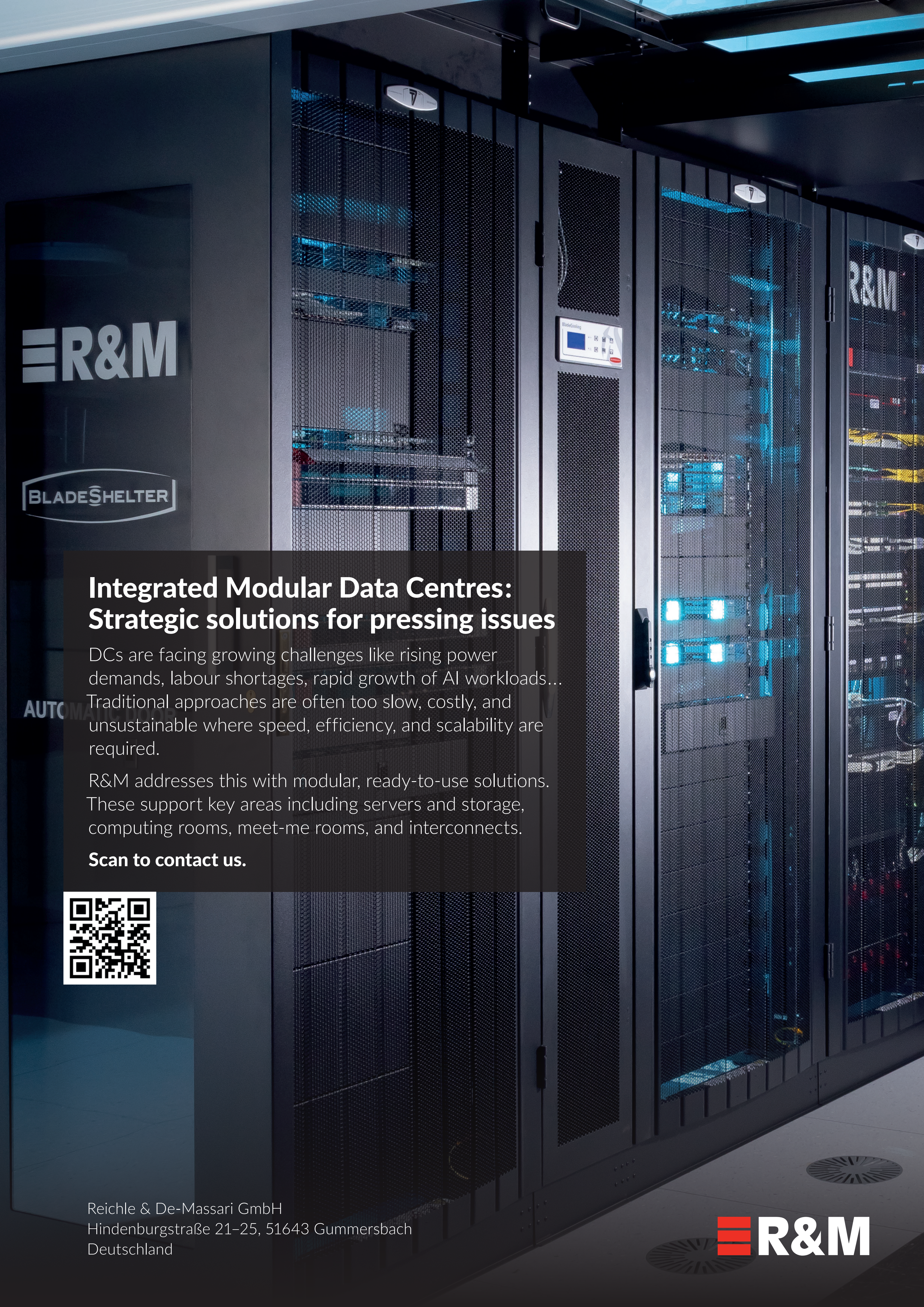
THE ENERGY REQUIREMENTS of digital infrastructure are constantly increasing. Data centres are among the most energy-intensive consumers in modern societies. While server technology, cooling and building systems have made significant progress in recent years, the focus is now shifting to lighting. It directly impacts not only energy use but also thermal load – and, in turn, the overall energy efficiency of the entire system.

"A key factor is replacing conventional fluorescent tubes with modern LED technology, which is up to 85% more efficient, generates less heat, and lasts significantly longer- a clear contribution to energy savings and reduced maintenance costs," explains Claudia Lüdenbach, Head of Application Management Industry Marketing at TRILUX, speaking at the Data Centre World Frankfurt trade fair, adding: "But efficiency alone is not enough. Smart control, sustainable materials and design, and a good working environment for staff are equally important."

Light in the server room: precise, efficient and reliable

Good lighting is crucial for accurate work in the aisles between the server racks - whether for maintaining components or reading displays. High vertical illuminance, glare-free light and excellent colour rendering (CRI > 90) are essential. At the same time, the lighting system must produce minimal heat to reduce strain on the air conditioning system.

As a pioneering technology and service provider in the European lighting market, TRILUX offers LED luminaires with efficient thermal management and high dust resistance to meet these requirements. Modular continuous line systems like E-Line Pro optimally light server aisles and surrounding areas.


R&M
BLADESHELTER
Integrated Modular Data Centres: Strategic solutions for pressing issues
DCs are facing growing challenges like rising power demands, labour shortages, rapid growth of AI workloads… Traditional approaches are often too slow, costly, and unsustainable where speed, efficiency, and scalability are required.
R&M addresses this with modular, ready-to-use solutions. These support key areas including servers and storage, computing rooms, meet-me rooms, and interconnects.
Scan to contact us.
Reichle & De-Massari GmbH
Hindenburgstraße 21–25, 51643 Gummersbach
Deutschland
R&M

LiveLink lighting management - At a glance

- Helps to save up to 85 % energy through demand-driven control
- Seamless integration into building management systems
- Can be combined with sensors, scheduling and zone control
- Time-saving commissioning
- Intelligent functions

Smart Lighting Solutions for Every Area

TRILUX ensures excellent lighting not just in server rooms, but throughout the entire data centre – including offices, technical areas, outdoor spaces, and circulation zones. Each has unique lighting needs, and TRILUX provides tailored solutions that combine comfort, safety, and energy efficiency - all from a single source. Intelligent control systems allow the lighting to be regulated as required - for optimum conditions and low operating costs throughout the building complex.

Office Areas

Create optimal work environments with smart lighting controls that adapt to user needs while reducing energy use.

Plant Rooms

The building technology in data centres is designed for maximum reliability. TRILUX supplies robust, higher protection rated lighting solutions to meet the demands.

Circulation Areas

TRILUX lighting solutions help to create easy orientation in corridors and staircases and a feel-good atmosphere in break rooms.

The Building Surround

Lighting for entrances, façades, walkways, and car parks enhances safety and integrates the facility into its urban surroundings.

An integrated lighting management system such as LiveLink can network all these areas - indoor and outdoor -controlling them centrally and improving overall building efficiency and quality.

Smart positioning and light control create a uniform, eye-friendly light distribution that increases safety and minimises the risk of errors during maintenance work.

Focus on sustainability - a cycle not a short life cycle

Sustainability in data centres means much more than just reducing CO_2. It is also about the service life, reparability and recyclability of technical components. That's why TRILUX develop durable, long-life lighting systems that can be adapted to future requirements - for example, with retrofittable modules or intelligent control systems.

Reusable housings, modular construction and service-friendly designs already make many modern lighting solutions fit for the circular economy. They also fulfil the requirements for sustainability certifications such as LEED or BREEAM - increasingly important for operators. The economic benefits are also clear: low-maintenance products reduce operational costs and minimise downtime, improving efficiency and offering greater planning security.

Continuous operation requires absolute reliability

Data centres operate 24/7 - and so must their lighting. To minimise maintenance costs and failure risks, operators rely on durable luminaires with an integrated emergency lighting function and high electromagnetic compatibility. All systems must work reliably even under extreme conditions - this applies to lighting as well as control and integration into the building automation system.

"Modular architectures and robust materials allow systems to be flexibly adapted and expanded as required. This not only creates technological security, but also investment protection. A lighting solution that grows with you offers long-term added value," summarises Claudia Lüdenbach, Head of Application Management Industry Marketing at TRILUX.

Expertise, planning skills and international experience are particularly crucial in data centre construction. Lighting technology and service providers such as TRILUX, who have set standards in large-scale industrial projects, contribute their expertise to lighting design for data centres - at every stage: from needs analysis to lighting design and implementation.

Thanks to its global presence and close networking with planners and operators, TRILUX develops customised concepts that are economical, scalable and future-proof. Efficiency, safety, and service always take centre stage, which is complemented by the flexibility to react quickly to new requirements. Visitors to the "Data Centre World Frankfurt" trade fair will be able to discuss lighting systems in data centres in more detail with the lighting experts at the TRILUX stand (H063 | Hall 8) on 4-5 June 2025. www.trilux.com/data-center

Lighting solutions with vision

Future-proof lighting for data centres combines efficiency with intelligent control and sustainability with adaptability - delivering reliability in complex environments. Whether it's LED technology, modular continuous line lighting systems or networked lighting management, planning for today means preparing for tomorrow.

Simplify Your Light - also in the data centre.

We connect and protect

LTA Coolant Distribution Unit

The trusted AI enablers

LIQUID-TO-AIR SIDECAR

>1GW OF Liquid Cooling Deployed Since 2020

Specifying liquid-cooled load banks for next-generation data centers

As the demand for high-performance computing and artificial intelligence continues to grow, data centers are adopting liquid cooling technologies to manage thermal loads more efficiently. Traditional air-cooling systems often struggle to dissipate the immense heat generated by modern high-density servers and hardware.

BY AVTRON POWER SOLUTIONS

LIQUID COOLING has superior thermal transfer properties, offering a more sustainable and effective solution. However, this shift also necessitates advanced testing and validation tools, such as liquid-cooled load banks, to ensure reliability and optimal performance.

How liquid cooling works in data centers

Liquid cooling utilizes a water and glycol mixture to absorb heat directly from server components. Compared to traditional air cooling, this method is significantly more effective due to the superior thermal conductivity of liquids. As computing power grows, so does thermal output, making liquid cooling essential for maintaining optimal temperatures and protecting high-value IT equipment from overheating.

Key methods:

- **Direct-to-Chip Cooling:** Coolant flows through a network of microchannels embedded in cold plates attached to heat-generating components. Heat is transferred to the coolant and then circulates to a heat exchanger or Coolant Distribution Unit (CDU).

- **Immersion Cooling:** Servers are fully or partially submerged in a dielectric fluid that absorbs heat. The fluid is circulated through a cooling loop to dissipate heat externally.
- **Rear-Door Heat Exchangers:** Mounted on the back of server racks, these units use liquid cooling to capture and remove heat from exhaust air, often combined with air cooling for hybrid solutions.
- **Coolant Distribution Systems:** CDUs regulate coolant temperature, flow rate, and pressure while serving as the interface between the facility's primary cooling loop and the server racks.

Using liquid cooled Load Banks

Liquid-cooled load banks are critical tools for testing and commissioning by simulating real-world electrical and thermal loads, allowing operators to validate the infrastructure's performance under various scenarios before going live and ensuring correct operation. Using load banks is a cost-effective solution for testing the system without risking the data server racks and data.

Key applications:

- **Thermal Validation and Heat Dissipation Testing:** They assess the cooling system's ability to handle specific thermal loads. Replicating heat output and coolant flow rates of servers, enabling engineers to verify that the liquid cooling loops are functioning correctly and efficiently dissipating heat.
- **Electrical Load Testing:** They simulate the electrical demands of a fully operational data center, ensuring that power distribution units (PDUs), backup generators, and uninterruptible power supplies (UPS) can handle peak loads without failure.
- **Commissioning and System Integration:** Before data centers go live, they use liquid-cooled load banks to commission the facility. This process ensures that all components – from cooling systems to electrical infrastructure –

work seamlessly together under realistic conditions.

- **Performance Optimization:**
  They help identify inefficiencies in cooling and power systems, enabling operators to fine-tune their infrastructure ensuring the data center operates at peak efficiency, minimizing energy consumption and operational costs.
- **Redundancy Testing:**
  They facilitate the testing of redundant systems, such as backup cooling and power systems. Simulating failure scenarios verifies the facility's resilience and ability to maintain uptime during unexpected events.

Building Load Banks to meet the requirements

Liquid-cooled load bank suppliers must adhere to several technical and operational requirements to adapt to changing demands.

Load Bank Features:

- **High Thermal Dissipation Capacity:**
  They must handle the high heat density characteristics of modern servers.
- **Scalability and Modularity:**
  They must be designed to allow for flexible testing of partial or full loads, accommodating future expansions. Connecting them to a network of liquid and/or air-cooled load banks will streamline commissioning and ensure the correct validation of multiple ancillaries. Also, having the ability to connect to and be controlled by the building management system is beneficial.
- **Compatibility with Liquid Cooling Systems:**
  They may need to integrate seamlessly with various liquid cooling architectures, including direct-to-chip cooling, immersion cooling, and cold plate systems. The connection points and ancillaries of the load bank should accurately match the data center to ensure real-life testing.
- **Precision Control and Monitoring:**
  Precise load bank control over electrical and thermal loads enables detailed testing and analysis. Real-time monitoring and data logging capabilities data can be used for comparisons with other liquid-cooled testing sites to improve the commissioning process.
- **Fine Load Control:**
  The ability to apply precise electrical load steps as fine as 5kW enables a broader range of test applications, allowing engineers to fine-tune load conditions.
- **Robust and Reliable Design:**
  Built to withstand rigorous testing environments with a heavy-duty frame that protects internal components and robust valves that can be cycled many times without leaks. Additionally, stainless steel corrosion-resistant piping and tanks are used to prevent rust or other corrosive materials from contaminating the system.
- **Sensors and Monitoring:**
  Integrated mechanisms, such as overtemperature protection and over-pressure systems, are essential to prevent damage to the load bank and the data center infrastructure.
- **Fast Cable Connections:**
  During commissioning, load banks are relocated within the data center to test various components. By using quick-connect power receptacles to make fast and efficient power connections minimizes downtime and improves testing efficiency.
- **Multiple Cable Connections:**
  Having the ability to connect multiple cables per phase allows the use of finer cable gauges and provides flexibility to connect to multiple busways.

Benefits of liquid cooled Load Banks in data centers

- **Improved Reliability:**
  Ensures all systems are tested under realistic conditions, reducing the likelihood of failures during operation.
- **Enhanced Energy Efficiency:**
  Identifies inefficiencies and enables operators to optimize power and cooling systems.
- **Reduced Commissioning Time:**
  Speeds up the process of validating and bringing new data centers online.
- **Environmental Sustainability:**
  Supports the transition to energy-efficient liquid cooling, aligning with sustainability goals.

Conclusion

As data centers evolve to meet the demands of modern computing, liquid cooling is becoming a cornerstone of their design. Liquid-cooled load banks play a vital role in ensuring these facilities operate efficiently, reliably, and sustainably. By providing realistic load simulations and enabling detailed testing, they empower data center operators to deliver high-performance environments that meet the needs of today and tomorrow.

Modular integrated data centres: a solution to the challenges of today and tomorrow

Modular Integrated Data Centres (MIDCs) offer a transformative approach to DC infrastructure. Prefabricated systems integrate IT, power, cooling, and networking components into compact, efficient, ready-to-deploy modules. This helps address current and upcoming challenges in an efficient, scalable, rapidly deployable way.

BY CARSTEN LUDWIG, MARKET MANAGER, REICHLE & DE-MASSARI

RAPID GLOBAL digitalisation and fast-developing technologies have significantly boosted demand for flexible, scalable DC solutions. Traditional construction methods struggle to keep up, due to lengthy build times, escalating costs, energy inefficiencies, and skilled labour shortages. MIDCs, however, offer a robust solution, with significantly shorter deployment timelines, reduced costs, and enhanced sustainability. These solutions combine the configurability and scalability of high standardization with the benefits of a fully pre-tested, right-first-time solution.

Market trends and requirements

Growing interest in MIDCs is driven by the increasing need for AI workload processing power, edge deployments, 5G network expansion, and digital transformation. Adoption of IoT, AI, and real-time analytics require positioning of DCs closer to users, to support lower latency and ultra-fast data processing. MIDCs can fulfil exactly this need.

Across various sectors, from cloud services and telecommunications to industrial automation and media streaming, MIDCs can address common critical requirements: rapid deployment (even in harsh environments), reduced PUE, renewable energy integration, scalability, advanced networking, and integrated AI-driven remote management systems. Traditional DCs can take years to construct, whereas MIDCs can be operational within weeks or months.

Leveraging prefabricated components standardises manufacturing, simplifies logistics, and minimises delays associated with traditional construction methods, significantly lowering overall capital and operational expenses. Advanced cooling technologies such as liquid cooling, precision airflow management, and modular uninterruptible power supplies substantially reduce energy consumption. That makes it easier to meet sustainability goals without compromising performance.

Cool and resilient

MIDCs provide several specific advantages when it comes to cooling. They are designed with efficient, built-in cooling systems such as in-row or direct liquid cooling, which help reduce energy consumption and boost efficiency. As the DC expands, cooling capacity can be scaled by adding modules, ensuring cooling needs are met as heat loads increase. MIDCs also offer precision cooling, targeting high-density server configurations to maintain optimal temperatures for critical equipment.

Additionally, the use of advanced, energy-efficient cooling techniques results in lower energy consumption, reducing costs and carbon footprint. The modular design allows cooling systems to be customized to the specific needs of each unit,


Galaxy VS

optimizing performance according to real-time demands. Another notable advantage is resilience against supply chain disruptions. Factory-controlled construction using standardised, pre-approved components reduces dependency on volatile material markets and complex onsite schedules.

Finding the right balance between customisation and standardisation is key. Custom modules provide tailored functionality to unique customer needs, while standardised components offer cost-effectiveness and easier integration.

Ready-to-deploy modules avoid multiple supplier dependencies, mitigating risk from material shortages and price fluctuations. Furthermore, prefabrication moves much of the construction workload to controlled factory environments where skilled labour is more readily available and specialised. Onsite work focuses largely on assembling pre-tested units, requiring fewer personnel and shortening project timelines significantly.

Implementation and operational excellence

It's important to realise that successfully integrating MI-DCs involves meticulous planning, particularly in the design and implementation phases. Efficient use of physical space is paramount, especially in edge environments characterised by constrained spaces. Solutions include high-density server racks and integrated floor cabling systems that accommodate extensive fibre or copper connectivity, maximising usable space.

Precise planning of power distribution, cooling, and redundancy ensures reliable performance and thermal management efficiency.

Advanced Data Centre Infrastructure Management (DCIM) software plays a pivotal role in MIDC effectiveness, providing remote monitoring, management, and automation of power, cooling, and IT assets. Real-time data-driven insights optimise capacity planning, avoiding over-provisioning or under-utilisation, enhancing reliability, and reducing operational costs through predictive maintenance.

Advanced DCIM ensures operational stability and energy efficiency across all modular units, tracking energy consumption to achieve low PUE, even under fluctuating workloads. What's more, predictive analytics can detect potential failures or inefficiencies in advance, helping maintain high availability through proactive maintenance and redundancy planning.

Future-proofing infrastructure

Scalability is a defining characteristic of MIDCs. This approach enables incremental expansion in alignment with evolving business needs, ensuring adaptability and future-proofing investments. Organisations can incrementally scale their infrastructure, significantly reducing upfront capital expenditure and minimising operational costs, aligning closely with key performance indicators (KPIs).

Integrated digital twin technology further enhances MI-DC deployment, enabling advanced simulation and planning, predicting performance accurately before physical implementation. This capability supports seamless integration and facilitates efficient management of infrastructure growth, reducing risk and optimising operational performance.

Comprehensive and integrated approach

MI-DCs present a powerful response to the pressing challenges faced by modern data centres, offering rapid deployment, cost-effectiveness, sustainability, and adaptability. They enable businesses across multiple industries to maintain agility, manage energy consumption responsibly, and scale infrastructure efficiently.

A successful MIDC implementation involves comprehensive coordination of multiple disciplines, including engineering, customisation, manufacturing, logistics, installation, security, connectivity, power management, and DCIM software configuration. Having a single point of contact managing interactions between various suppliers and specialists ensures consistency and interoperability throughout the project.

Seamless integration between infrastructure components, or with existing installations, demands expert engineering, particularly regarding precise cable routing and inter-module connectivity to maintain low latency and high-speed connections. Employing a multidisciplinary 'rainbow team' of experts facilitates highly structured communication and coordinated execution.

This holistic approach addresses unique operational requirements, such as optimising edge connectivity and data processing capabilities, integrating advanced containment systems for enhanced cooling efficiency, and incorporating renewable energy solutions for sustainability compliance.

In short...

MIDCs offer greater flexibility, scalability, and faster deployment than traditional data centres. They are ideal for applications requiring rapid expansion, such as cloud services, edge computing, AI workloads, and disaster recovery. MIDCs provide pre-fabricated, standardized units that streamline construction, reduce setup time, and make expanding – or downsizing - easy. When specifying a solution, it's important to consider power, cooling, and connectivity requirements, as well as future growth potential and customization needs. Additionally, cost and location factors should be evaluated to ensure optimal deployment.

Revolutionising data centre sustainability

The power of liquid immersion cooling technology.

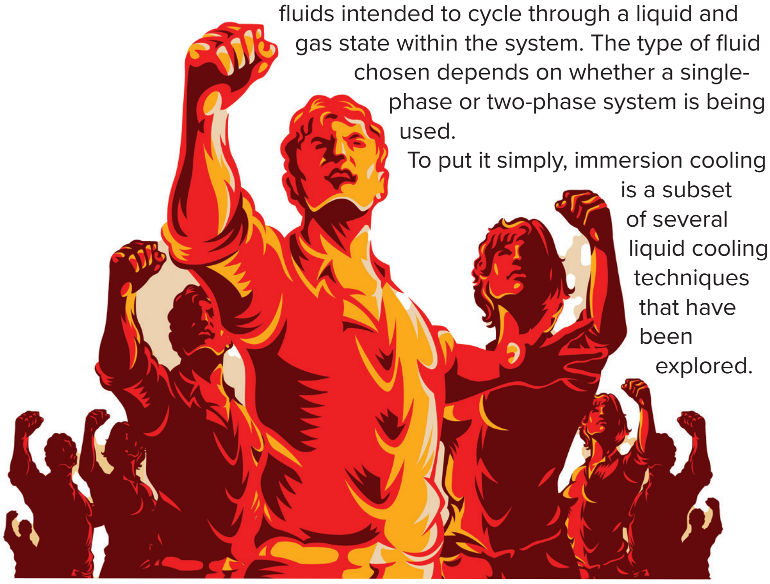
BY CHRIS CARREIRO, CTO, PARK PLACE TECHNOLOGIES

IMMERSION COOLING is a type of liquid cooling used to moderate data center equipment temperature by submerging it in a cooling fluid. Server immersion cooling helps to dissipate heat and keep components like CPUs performing optimally. Immersion cooling systems prove to be more efficient than traditional data center cooling methods (like computer room air conditioning, or CRAC) due to the increased thermal conductivity of most liquids compared to air.

Because 1-1.5% of electricity use across the globe is attributable to data centers, companies have been innovating to find a liquid cooling solution that can reduce that energy demand. Dielectric fluid immersion cooling is one solution that could increase CPU density in data centers while consuming less energy. Dielectric liquid cooling depends on the use of a thermally conductive but not electrically conductive fluid that will not disrupt the function of electrical components like servers. Examples of dielectric fluids include mineral oil hydrocarbons, synthetic fluorocarbons, and silicone fluids.

Immersion cooling technology can rely on dielectric fluids purpose-chosen to remain in a liquid state, or fluids intended to cycle through a liquid and gas state within the system. The type of fluid chosen depends on whether a single-phase or two-phase system is being used.

To put it simply, immersion cooling is a subset of several liquid cooling techniques that have been explored.



Other types of liquid cooling include direct-to-chip (DTC), rear-door server rack cooling, waterborne data center cooling, and evaporative cooling.

With immersion cooling, whole data center components are directly submerged into a specially designed tank. In contrast, water-cooled server racks look very similar to traditional rack-mount servers, but they are networked with waterblocks and tubing that circulates fluid to help dissipate heat.

The “phases” in single-phase and two-phase immersion are a reference to states of matter, and not physical stages in the system. The physical footprints of these two immersion cooling tanks are not drastically different, but their cooling cycles and contained fluids set them apart.

In single-phase immersion cooling, heat from the immersed server components is transferred directly to the surrounding fluid. However, the dielectric fluid does not undergo a “phase change” from a liquid to a gas. Instead, the fluid is cycled out of the immersion tank by a coolant pump that runs through a heat exchanger and is returned to the immersion tank at a lower temperature where it continues this heat transfer cycle.

In two-phase immersion cooling, heat from immersed server components causes the special immersion fluid to boil. The resulting steam heats a condenser coil in the top of the sealed chamber. The coolant in the condenser coil is cycled out of the chamber to a heat rejection mechanism (cooling tower, etc.). Then, the coolant is sent back to the sealed chamber at a lower temperature, ready to continue the heat transfer cycle.

Because of the steam from the phase change in two-phase liquid immersion cooling, the chamber must be sealed during operation. This means that performing maintenance requires a cooling and unsealing process that costs valuable operation time (which can cost as much as $5,600 per minute). The average power usage effectiveness ratio (PUE) within a data center can be measured by

dividing total energy consumed by energy used by computing equipment. This means that as PUE gets closer to 1, efficiency is improving. According to The Register, PUE for a traditional data center in 2022 was approximately 1.58, while single-phase immersion was able to bring this number down to the 1.05 to 1.10 range.

Not only does immersion cooling improve the energy efficiency of data centers, but it can save valuable space as well. According to a 2023 research article, immersion cooling only requires about one-third of the space to that of an air-cooled configuration.

One of the main contributors to this efficiency is the improved rack power density from not having to allow for air flow within servers. CRAC, one of the main traditional cooling methods for data centers, is reliant on the use of fans. This means that traditional data centers are very loud. Immersion cooling server configurations don't rely on fans and air flow for cooling. Because of their liquid cooling function, immersion cooling has proven to reduce data center noise.

Launching a liquid cooling solution comes with a list of challenges, like preparing existing hardware for immersion, training maintenance staff (or finding qualified third-party maintainers) on the repair process for immersed gear, and managing vendors for the tanks, dielectric fluid, and more. But the ROI and sustainability benefits make liquid cooling an exciting part of future data center planning.


HIMOINSA
A YANMAR COMPANY
HGY
SERIES
HIMOINSA
YANMAR
YANMAR
CRUCIAL BACKED UP DATA DEMANDS
THE BEST BACK-UP POWER
POWER & RESPONSIBILITY

Base 16 Technology in the data centre: what are the benefits?

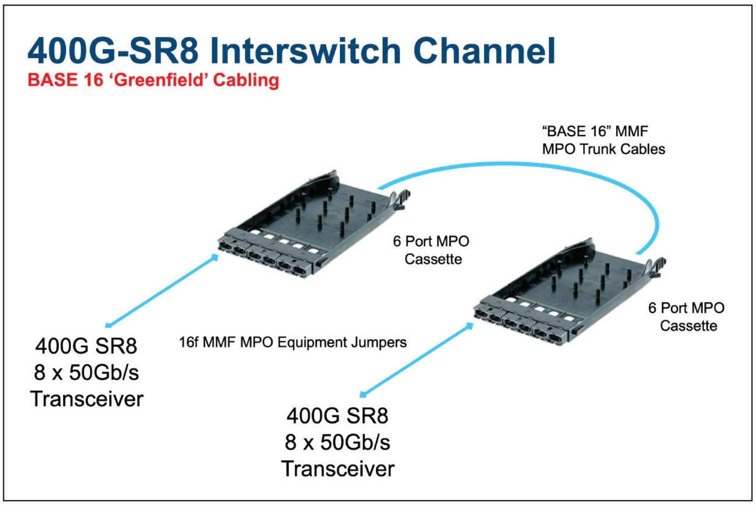
As the data centre market transitions to higher data speeds, it is important to consider what solutions provide the speed and bandwidth necessary to enable those applications that require faster and larger datasets.

BY MICHAEL AKINLA, PANDUIT

HYPERSCALE AND EDGE data centers are installing or upgrading infrastructure fabrics to 400G and even 800G (2x400G) for switch to server and leaf-spine connectivity, for which there are several MMF and SMF transceiver options to choose from. Among these options, many customers are evaluating multimode parallel optics for short-reach network connectivity. 400GBASE-SR8 transceivers over 16 parallel MMF strands (50Gbps on each strand of MMF), require Base 16 MMF structured cabling systems, providing significant value for short-reach connectivity systems interconnecting MOR/EOR switches to servers.

Newer technologies and applications that require faster interconnect, increasing bandwidth, and network capacity require innovative solutions to enable operators to offer leading-edge platforms. There is keen interest in 400GBASE-SR8 systems utilizing multimode transceivers, that for switch-to-server connections provide lower cost and power consumption than SMF alternatives such as 400GBASE-DR4.
Base-16 structured cabling, with small form factors cassettes, MPO-16, and Base 16 cabling simplify network deployment in greenfield networks as shown in Fig. 1. To take full advantage of multimode short reach variant, Base 16 interconnection with SR8 transceiver, can be utilised over existing Base-12 structured cabling infrastructure conversion cassettes.

➤ Figure 1. 16 fibre MPO Solution



We are currently seeing single mode traffic in terms of some customers moving from 400G to 800G (2x 400G), however, our customer interactions indicate that the multimode 16 fibre MPO modular cassette solution certainly has a place in the market, particularly for MOR/EOR switch to server interconnections due to the previously mentioned advantages. This is especially true in respect of preterm assembly offering speed of deployment, with highly flexible cassette configurations.

Expanding capacity

The key benefit that enabling high density breakouts of 50G server ports using SR8 is the level of expansion if offers. When you have Base-16 fibre from an SR8 perspective it offers 50G breakout, for example, from a 400G SR8 optic, being delivered as a 32 port 1RU modular 400G switch, will then enable 256, 50G breakout ports in terms of a high Radix.

A benefit of this approach is a reduction in the number of optical links and supporting cable plant (patch panels and connectors) as the deployment is 400G to 400G and then using breakout technology to get the 50G at the top of the rack (TOR) server end without using an actual TOR switch. This allows operators to deploy less fibre infrastructure into the data center environment. A further benefit therefore is the fourfold increase in respect of aggregator switch bandwidth, and in some situations a lower initial installed cost, depending on the SKU.

Moreover, the replacement of several TOR switches with a few MOR/EOR with more capable and efficient switches that produce significant reduction in power consumption.

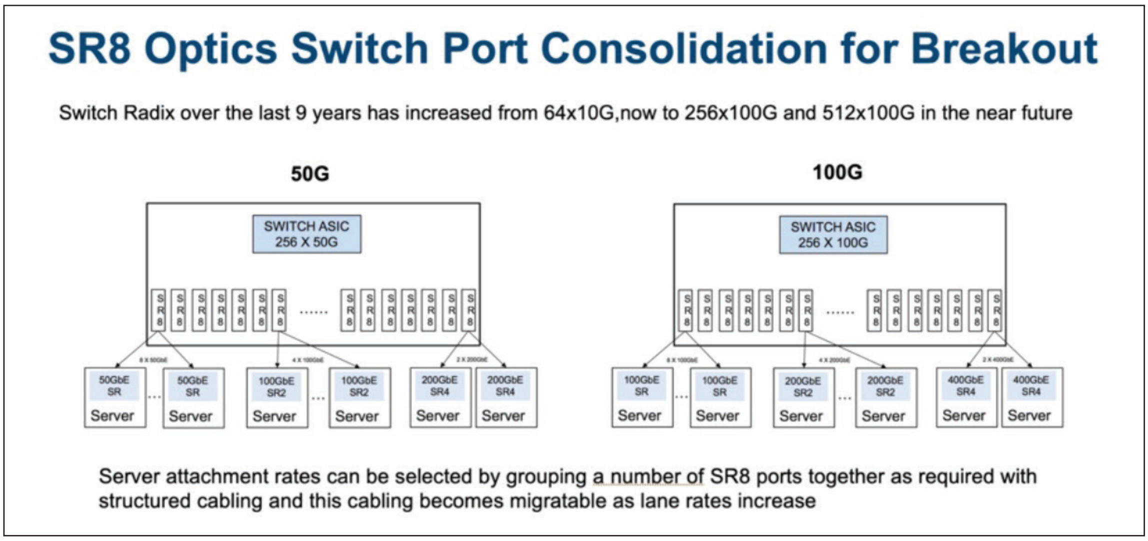


Figure 2. SR8 Optics Switch Port Consolidation for Breakout

The 400G breakout to 50G servers appears to be a sweet spot from a multimode perspective, using a variety of options such as the conversion cassettes, for customers who want to use existing infrastructure across legacy Base-8 and Base-12 fibre cable plant in conjunction with SR8 Base-16 optics.

400G DR4 transceivers can also breakout into 4 streams of 100G. This can be useful to implement the spine-and-leaf connections between switches. Breakout cassettes as the one shown in Fig. 3 can be used. However, contingent on the deployment length and application, SR8 multimode can provide advantages compared to single mode fibre DR4.

SR8 technology helps to simplify the server pod build providing different options. Data centers can deploy patch panels above the servers which breakout of the 400G transceiver with Base 16 MPO coming into the fibre panel into duplex LC ports, and then use patch cable to connect to the servers at 50G. Another option is to breaks out, with a hydra, from the Base 16 MPO delivering 50G into the server environment, together with an EOR/MOR Switch with 400G ports that break out onto the patch panel and feeds the servers. These deployments enable pre-terminated overhead cable providing simplified cable distribution in pre-populated cabinets. Overhead infrastructure is advantageous in a rolling rollout environment allowing easier access to end of life cabinets and hardware (Fig. 43.

Conclusion

As the data center market transitions to higher data speeds, we need to consider what solutions provide the speed and bandwidth necessary to enable those applications that require faster and larger datasets. Essential to this is the consideration of how the devices that enable this connectivity to interoperate with existing networks, and how their introduction will move the capabilities forward. Base 16 has a place in that discussion as business drivers for data center operator include new network upgrades and increasing bandwidth while reducing latency.

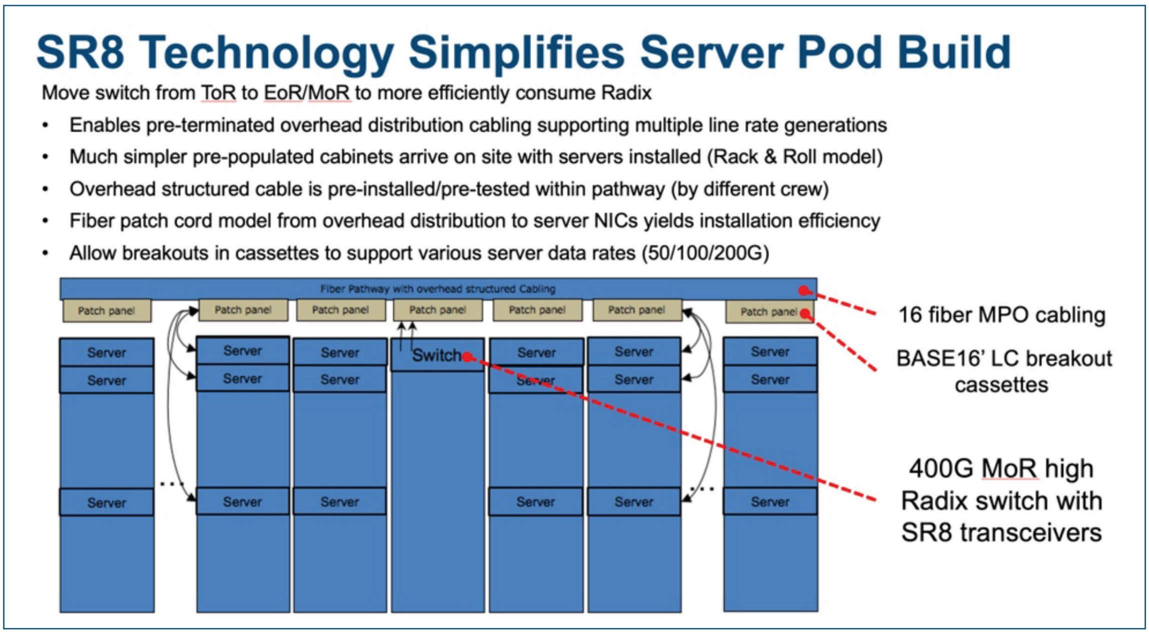


Figure3. Simplify Server Pod Build , SR8 Technology Simplifies Server Pod Build

It's time to rethink data centre power

The AI revolution is fully under way. But without a top-to-bottom reassessment of power solutions, its progress could slow to a crawl.

BY TOD HIGINBOTHAM, COO, ZINCFIVE

WITH BUSINESSES clamoring to harness the potential of AI, data center operators have a mandate to outfit their facilities with the latest CPUs, GPUs and other components that power the most demanding workloads. That level of power, however, has sparked a surge in electricity needs. And meeting those needs is harder than it sounds.

Already, data center projects are being hamstrung by power problems. Data center operators need more power -- and they need it from clean power sources. They also need a transmission system that can handle a heavier load. The burden, however, doesn't belong solely to utility providers. Data center operators also need to innovate. By rethinking data center design, operators can minimize their power needs. Meanwhile, they need to strategically monitor and manage their power usage to ensure optimal operations.

The AI revolution is fully under way. But without a top-to-bottom reassessment of power solutions, its progress could slow to a crawl.

A surge in demand

Insufficient power resources are already slowing projects down by years, according to a recent report from Cushman & Wakefield, a global commercial real estate services firm.

"Over the past year, power has become the number one consideration for data center operators as they conduct site selection to rapidly grow their portfolios," the firm's 2024 Global Data Center Market Comparison says. "Many utility providers are suggesting wait times of 2-3 years or more for sizable power to be delivered to their developments."

Data center operators are expanding their footprint in anticipation of the growth of AI. Currently, hyperscalers need around 10 kW to 14 kW per rack, the commercial property consultancy Newmark noted in a recent report. However, AI workloads will push that requirement up to 40 kW to 60 kW per rack. All told, Newmark expects AI to drive US data center demand to 35 GW by 2030, up from 17 GW in 2022.

As they anticipate greater power demands driven by AI, data center architects should consider what types of AI loads they are building for -- in other words, whether a data center operator will manage primarily inference or training. With AI training, there may be extreme load steps that require larger utility feeds or the use of innovative battery storage. There is interest in using batteries to offload these peaks and spare the uninterruptible power supply (UPS) from these load steps. These batteries could be located relatively close to AI servers and allow for a significant reduction in utility MWs to the facility.

Securing access to different power sources, transmission lines

When data center architects plot out new builds, they aren't just thinking about power availability but also the source of that power. Across the globe, more stringent regulatory environments as well as pressure from corporate stakeholders are driving data center operators to step up their reliance on renewable energy.

The demand for renewable power is reflected in forecasts for U.S. energy production: Solar installations will account for "almost all growth" in US power generation in 2024-2025, according to the U.S. Energy Information Administration (EIA). Utilities are opting for solar installations, Reuters notes, thanks to tax credits available from the 2022 Inflation Reduction Act.

While the shift to clean energy is happening, data center developers still need to think strategically about siting new facilities. Even with the increase in solar installations, the EIA said that by 2025, solar will still only account for 7% of power production in the US.

Along with power generation, data center developers need to consider transmission. On both of these matters, developers can collaborate directly with utility providers to ensure they can access the power they need.

The US electric grid faces capacity shortfalls, the North American Electric Reliability Corporation (NERC) warned in December, due to increases in demand and fossil fuel generators coming offline.

Delivering sustainable backup power

Transmission lines should also be assessed for reliability. Aging transmission lines, as well as transmission lines impacted by extreme weather events, can potentially cause costly outages. Power issues are consistently the most common cause of serious and severe data center outages, the Uptime Institute found in its annual outage analysis. More than half of the respondents surveyed by the Institute said their most recent significant, serious or severe outage cost more than $100,000.

The costly nature of outages underscores the importance of backup power systems. As with other power sources in the data center, developers face an imperative to make backup systems sustainable. Nickel-zinc (NiZn) batteries, an innovation in battery technology led by ZincFive, can power UPS systems more sustainably than traditional lead-acid batteries or lithium-ion batteries. In comparison to other chemistries, NiZn produces lower GHG emissions and offers a smaller water footprint and energy footprint. Specifically, NiZn batteries' lifetime greenhouse gas emissions are 4x lower than lead-acid and 6x lower than lithium-ion emissions. Nickel-zinc batteries use common, widely available, conflict-free materials. They're also highly recyclable.

Meanwhile, major corporations are exploring a range of alternative energy sources for backup generators. Microsoft, for instance, has been testing the viability of using large-format hydrogen fuel cells to supply data center backup power. Microsoft is also installing a "resiliency microgrid," which relies on renewable natural gas, for backup power at its San Jose, Calif. data center.

Dealing with heat

Data center developers and operators should also be looking for ways to more efficiently manage the heat generated by their infrastructure. Cooling systems are all the more critical as businesses adopt AI and HPC systems that emit greater levels of heat. Typically, cooling systems consume around 40% of a data center's power.

In addition to offering sustainable backup power, NiZn batteries allow data center operators to reduce the footprint of their cooling systems and other safety infrastructure. Nickel-zinc batteries exhibit no thermal runaway at the cell level and are thus non-flammable, unlike other UPS battery alternatives.

Conclusion

While power plays second fiddle to innovations in computational hardware and increasingly sophisticated AI workloads, it's a fundamental part of the data center. And as with other elements in the data center, it's rapidly evolving. To build a data center that can meet the demands of tomorrow's workloads, data center developers and operators should reconsider their power infrastructure, top to bottom.

AI and GPU data centres: Navigating the networking challenge

The rise of artificial intelligence (AI) and its integration into industries has increasingly become a focal point worldwide. AI for IT operations (AI Ops), a practice that leverages AI to optimise and automate network management tasks, is widely expected to revolutionise network operations. However, to be effective, it requires a flexible, software-defined network control plane paired with secure remote access for provisioning, orchestration, management, and remediation.

BY ALAN STEWART-BROWN, VP EMEA AT OPENGEAR

TO DELIVER to its full potential, AI relies on immense computational power, much of which is delivered through modern data centres. These data centres, equipped with advanced graphics processing units (GPUs), have become the backbone of AI innovation. Powered by Moore's Law, GPUs have been critical in supporting the growing demands of AI workloads. According to MarketsandMarkets, the global data centre GPU market size was valued at US $14.3 billion in 2023, and it is estimated to reach US $63 billion by 2028, growing at a compound annual growth rate of 34.6 during the forecast period from 2023 to 2028.

The elephant in the data centre: Networking bottlenecks

GPUs have revolutionised AI development due to their ability to process vast amounts of data simultaneously. This parallel processing is ideal for the complex computations required by deep learning and large language models like GPT.

Yet as these models grow in complexity and size, they generate "elephant flows" – substantial data chunks that strain traditional ethernet networks. This leads to congestion and increased latency, creating bottlenecks that hamper performance. Ethernet, while ubiquitous and cost-effective, wasn't originally designed to handle such voluminous and high-speed data transfers.

This networking bottleneck has ignited a debate within the data centre community: Should the industry continue to rely on traditional ethernet networks, or explore alternative solutions better suited for AI workloads? Some argue that enhanced ethernet technologies, such as remote direct memory access (RDMA) over converged ethernet (RoCE), offer low-latency data transfer capabilities that can mitigate these issues. Others believe that entirely new networking paradigms may be necessary to meet the demands of AI-driven data centres.

Amid this technological tug-of-war, network management within GPU data centres faces its own challenges. Traditional network switches typically include console management ports for straightforward configuration, but many newer, high-speed switches lack these ports, relying instead on ethernet management interfaces. This discrepancy necessitates a re-evaluation of management strategies to ensure seamless operation regardless of the underlying networking technology.

Adapting network management for AI's future

Independent overlay management networks emerge as a viable solution, providing a unified management layer that interfaces with both ethernet and serial connections. This approach ensures data centre operators maintain robust control over their networks, enabling secure remote access for provisioning, orchestration, management, and remediation tasks. By decoupling the management plane from the data plane, these overlay networks offer the flexibility and resilience required in the evolving landscape of GPU data centres.

However, as networks grow in complexity, relying solely on in-band management can be risky. This is where out of band management becomes crucial, providing a dedicated pathway that operates independently of the primary network infrastructure.

In the event of network failures or disruptions caused by heavy AI data loads, out of band access allows administrators to remotely manage and troubleshoot devices without relying on the main network. This ensures minimal downtime and maintains operational continuity, critical when dealing with AI workloads where any interruption can lead to significant productivity losses.

Integrating out of band management solutions enhances resilience, ensuring continuous operations even under strain. This dedicated channel allows swift issue resolution, safeguarding AI application performance and reliability.

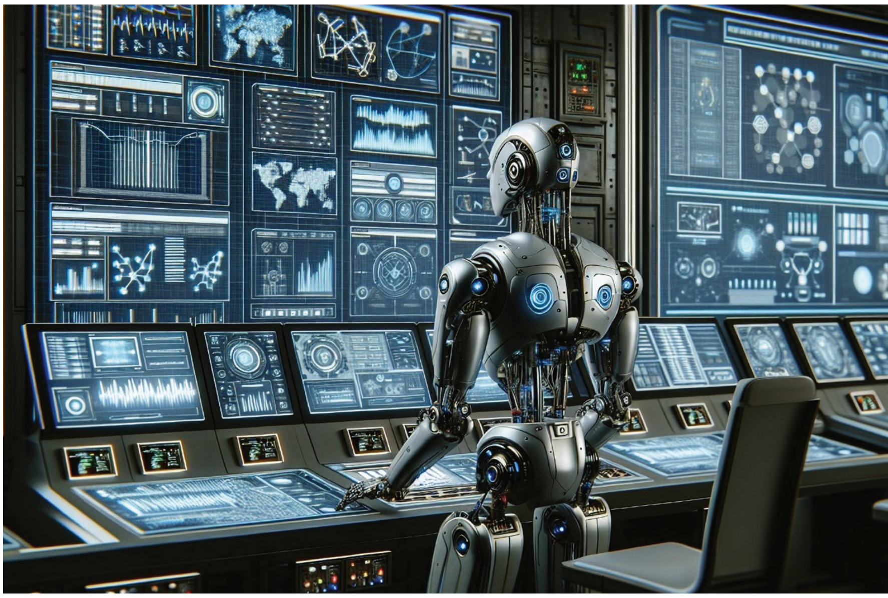
The broader challenge lies not just in selecting specific networking technologies but in designing infrastructure capable of meeting the ever-increasing demands of AI workloads. Data centres must prioritise flexibility, scalability, and security in their network designs. Embracing software-defined networking (SDN) creates a flexible control plane that dynamically adjusts to shifting workloads and network conditions. This adaptability is crucial for handling the unpredictable nature of AI data flows.

As edge computing and IoT devices generate more data at the network's periphery, data centres must extend capabilities beyond centralised locations. This expansion highlights the need for resilient infrastructure and cost-effective edge deployments. Implementing robust network management solutions across distributed environments ensures data integrity and availability, regardless of where data is generated or consumed. Without these measures, including both in-band and out-of-band management, the vast volumes of data risk being underutilised, limiting their potential to produce actionable insights and meaningful change.

Navigating the networking challenge in AI-powered data centres requires a commitment to innovation and agility. Organisations must remain open to adopting new technologies that enhance network performance and management. Integrating AI into operations can improve efficiency and reduce the likelihood of human error. This proactive approach enables businesses to pre-emptively address issues before they escalate into significant problems.

Yet there is no one-size-fits-all solution. The choice of networking infrastructure may vary based on specific use cases, budget constraints, and scalability requirements. What remains constant is the need for a robust, flexible network management strategy that accommodates current demands and future growth.

In this rapidly evolving landscape, robust network management remains essential for ensuring performance, scalability, and security in AI-powered data centres. Strategic use of out of band management, combined with innovative technologies, enables data centres to handle the growing demands of AI workloads while maintaining operational continuity. By adopting flexible and resilient infrastructure, organisations can unlock AI's full potential, driving innovation and thriving in an increasingly data-driven world.



Navigating power demand in the age of AI

As the world increasingly relies on digital services and AI, data centres face a growing problem in terms of power demand. ABB explores what data centre professionals can do to meet surging demand, while staying competitive and meeting their sustainability goals.

DANEL TURK, DATA CENTRES PORTFOLIO MANAGER AT ABB

THE CONTINUOUS GROWTH of AI services since the launch of ChatGPT in 2022 means that data centre power demands will continue to rise. In fact, the International Energy Agency's (IEA) Electricity 2024 report predicted that global data centre electricity use could double by 2026, reaching 1000 terawatt-hours (TWh) – that's equivalent to the annual electricity consumption of Japan.

So, data centre professionals face several challenges when looking to meet that demand in a cost-effective, sustainable way. Here are a few considerations.

Think modular

If you're building or enlarging a data centre, take advantage of a modular approach. Modularity offers a way of scaling sustainably in a way that meets the demand for power and availability while simplifying the specification and installation process.

Rather than constructing or expanding a data centre all at once, you can build it up in blocks. As an example, you can build a projected 200-megawatt (MW) data centre in 20 MW sections – onboarding customers as demand increases. That makes it more cost-effective because it avoids wasted capacity along the way.

It's also an approach that minimises downtime and risk, especially with prefabricated solutions – like skids and eHouses – that are ready for quick and easy installation as they are constructed off-site and then factory-tested before delivery. Besides, buying one prefabricated product instead of a range of individual components that require assembly saves money and time. This approach is favoured by local utilities and governments too, with the smaller incremental growth of a data centre more likely to be approved than the full-scale build of a new one.

nVent

We connect and protect

LTA Coolant Distribution Unit

The trusted AI enablers
LIQUID-TO-AIR SIDECAR

>1GW OF Liquid Cooling Deployed Since 2020

Consider shifting to medium voltage

Medium-voltage (MV) equipment has become increasingly viable and cost-effective especially in the case of rising power demand.

MV UPSs, for example, can be installed modularly. As an example, ten 2.5 MW UPS blocks can be wired in parallel to create a 25 MW system. This enables faster deployment and increases overall system capacity without any additional complexity. It also avoids the challenge of having multiple LV UPSs that need to be regularly maintained and serviced. In addition, the lower currents at MV level mean cables can have a smaller cross-section, leading to additional savings.

Furthermore, MV UPSs are often more energy efficient than their low-voltage (LV) counterparts, and can provide power protection for the entire data centre, not just the server racks. So, switching to MV can offer long-term cost savings.

Be more sustainable with BESS

Meeting sustainability goals is a challenge for data centre operators when it comes to the increase in power demand.
One solution is using battery energy storage systems (BESS) to reduce reliance on diesel gensets as they can integrate renewables, like wind and solar energy, into the local energy mix. By using a BESS, excess energy from solar panels on the roof can be stored and used at another time, optimising renewables usage. Using BESS can also provide load shifting and frequency response services to the grid, further aiding negotiations with the local utility, and creating potential new revenue streams via give-back schemes during peak demand.

First, though, it's important to plan carefully and consider what your energy demands are to avoid over-specifying – after all, specifying for five hours of uptime would require a lot of batteries.

Make use of AI

The onward development of AI might be driving the increase in power demand, but it can also help to make data centres more reliable and efficient.

This could happen through optimised cooling, predictive condition-based maintenance, data access and transfer, and demand balancing.

The key to this is connectivity. An automation system running an AI suite can keep track of the hundreds of thousands of monitoring points deployed in a typical mid-to-large data centre. This provides operators a 360 overview of the data centre's performance, energy use and asset health at any given point.

This data can then be used to make efficiency improvements. Take the example of a cooling system. While the upstream chiller and the distribution system are often viewed separately – leading to operational inefficiencies when attempts are made to make it more efficient – an automation system lets operators see holistically how one part affects the other. As a result, operators can make more informed decisions to improve overall efficiency.

Invest in SF6-free equipment

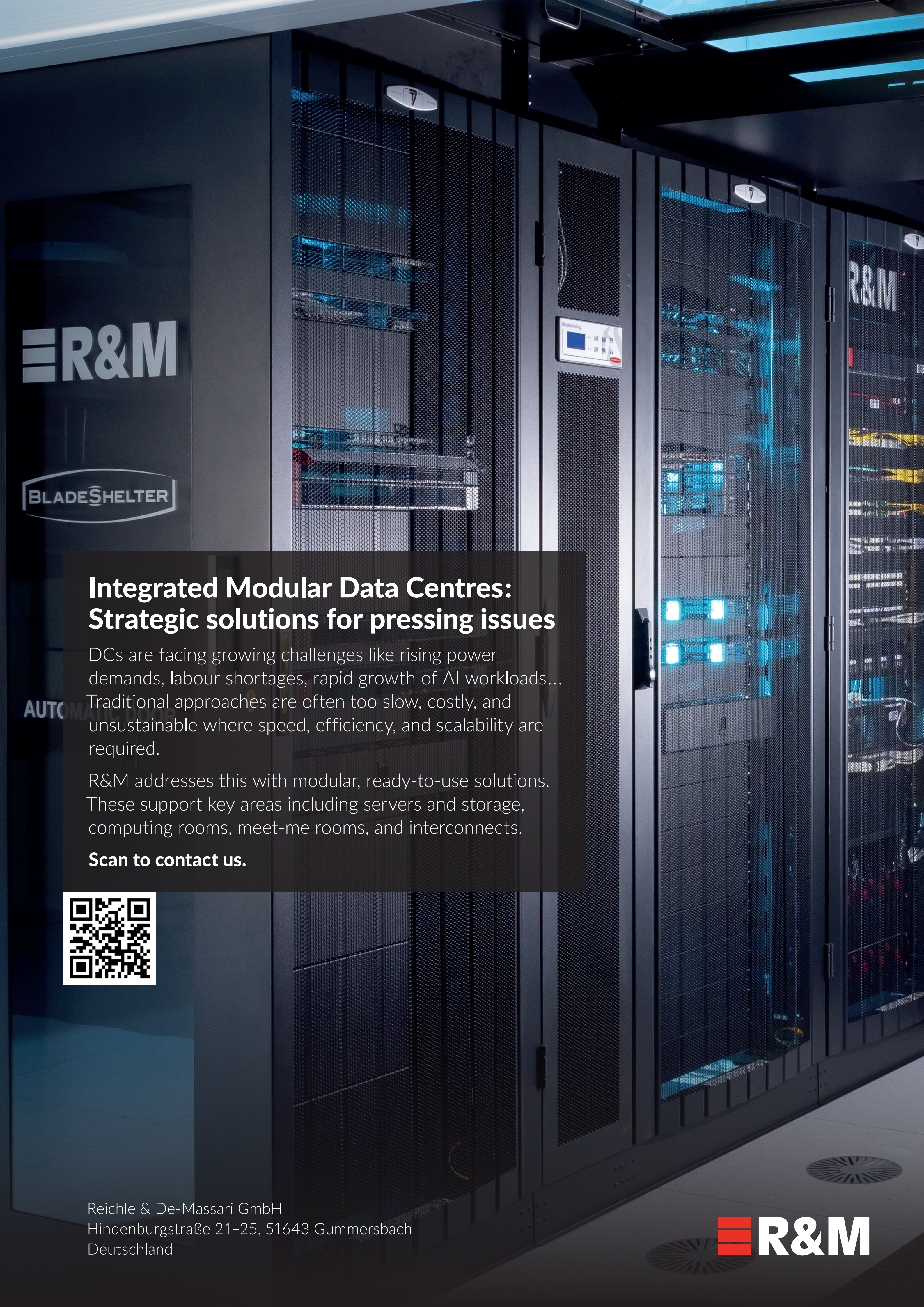
Moving to SF6-free equipment now will ensure that you're prepared for incoming regulations on the use of electrical equipment which contains the insulating gas. This will make your sustainability reporting easier. Regulations have been proposed because SF6 can leak to the atmosphere, where it has a global warming potential around 25,000 times greater than CO_2.

Adopt a TCO mindset

When considering how to grow efficiently and sustainably, data centre managers need to adopt a total cost of ownership (TCO) mindset.
This involves calculating and assessing all the direct and indirect costs of an asset over its entire lifecycle to determine its cost, as opposed to a more traditional view of separating capital and operational expenditure. By viewing expenditure with a TCO mindset, future operating cost savings are seen as net present value.

A simple way of looking at this is where you have a machine that will be running continuously for around 20 years; by investing in a more energy efficient model now, you will see significant cost and emission savings over the long term.

Power demands might be growing as AI and digital services expand exponentially, but data centre managers and operators can meet these challenges with the right strategies in place. While these challenges will become more urgent as demand increases, acting now, with a TCO mindset, will help to ensure data centres continue to make a positive impact.


R&M
BLADESHELTER
Integrated Modular Data Centres:
Strategic solutions for pressing issues
DCs are facing growing challenges like rising power demands, labour shortages, rapid growth of AI workloads… Traditional approaches are often too slow, costly, and unsustainable where speed, efficiency, and scalability are required.
R&M addresses this with modular, ready-to-use solutions. These support key areas including servers and storage, computing rooms, meet-me rooms, and interconnects.
Scan to contact us.
Reichle & De-Massari GmbH
Hindenburgstraße 21–25, 51643 Gummersbach
Deutschland
R&M

How must data centres evolve to meet AI workloads?

Artificial Intelligence (AI) and Machine Learning (ML) are becoming mainstream and they can't be avoided or ignored. This is having a knock-on effect on the infrastructure that powers our lives as these technologies will only work if there are powerful computers that can process millions of data points every single second. As a result, the growth of AI and ML, along with the continuing increase in cloud and enterprise workloads, means that the need for computing power is also growing - and the most efficient way to achieve computing power at scale is in data centres.

BY DARREN WATKINS, CHIEF REVENUE OFFICER AT VIRTUS DATA CENTRES

DEMAND for data centre space has been growing at an exponential rate for decades, driven by the digital economy. However, the overlay of AI and ML deployments are increasing demand further and faster across the world. Statista predicts that the data centre market will reach US$349.20bn in 2024, growing to US4438.70bn by 2028, causing data centre providers to re-evaluate their strategies.

Traditionally the backbone of many technological advancements, as computing power is the fuel of our technologically advanced society, data centres are now faced with the imperative to be more than infrastructure providers. They have a new critical challenge where they need to provide more than the essential network and infrastructure supporting data storage, management and cloud services in an always-on manner.

The rapid growth of AI and ML means that data centres need to be to be even more agile, innovative and collaborative to power this new era. This includes managing sustainable power at scale and implementing designs that support rapid and scaleable AI deployments, whilst consciously aligning with values that benefit the data centre provider, the customer and wider society responsibilities.

Location: where to build?

Many existing European hyperscale facilities simply aren't capable of meeting the short-term future demands of AI and there is a shortage of the right type of supply i.e. large scale facilities with renewable energy close to, but not in, a major European metropolitan city. So where can data centres be built to accommodate this growing demand for computing power?

Over the past few years, location has been a very specific, deliberate choice as the technological landscape was meticulously mapped to minimise latency. Today, with power constrained central metropolitan areas and the integration of AI and ML workloads which are less latency sensitive, are orchestrating a shift in priorities. These advanced AI workloads challenge the traditional principles that often dictated optimal data centre locations. The result is a profound reconsideration of what defines an ideal site, and new locations are opening up as potential locations for data centres.

This shift isn't about lessening the importance of low latency - it's about recognising the evolving needs of integrating AI and ML. The move towards larger campuses is a calculated strategy that acknowledges the non-linear cost relationship inherent in these operations; larger megascale campuses capable of 200-500MWs can often afford providers - and therefore customers - greater efficiencies. This bold step challenges the long-standing industry norm, presenting a compelling argument that prioritising sheer scale over proximity and access to renewable energy can yield more efficient and sustainable outcomes.

Sustainability is even more important

While some may consider access to power, water and connectivity traditional requirements, from a customer's perspective that will remain unchanged. However, for data centre providers, with the increase in computing power required to enable AI and ML workloads, comes an increase in the power needed to operate facilities.

Worldwide, data centres consume about 200 terawatt hours of power per year - more than some countries. And the forecast is for significant growth over the next decade, with some predicting that by 2030, computing and communications technology will consume between eight percent and 20 percent of the world's electricity, with data centres accounting for a third of that. This requires the technology industry and data centre providers to be even more innovative to lower Power Usage Effectiveness (PUE) and Water Usage Effectiveness (WUE) and in turn reduce their reliance on diesel generators.

Power must come from a renewable source and be used efficiently, and this means the facility must be designed to be as efficient and sustainable as possible. Sourcing only 100% renewable energy and contracting with Power Purchase Agreements (PPAs) to use dedicated solar and wind farms to power data centres are all critical initiatives which the most sustainable data centre providers are embracing. In some countries like Germany there are laws regarding the power usage effectiveness (PUE) of data centres to drive responsible behaviour.

With regards to cooling, unfortunately it is impossible to beat the laws of thermodynamics; heat generated by the computer systems still needs to be removed from a facility using power. However, there are methods that are more efficient than others; for example, removing the heat direct from the chip involves liquid and not air – and it requires design changes to the infrastructure to enable this. Another alternative is immersion cooling which is generally more suited to a bare-metal solution.

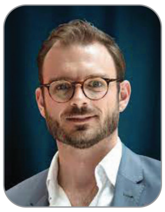
Reuse of waste heat has been in the headlines recently and is another way to achieve PUE. It's an interesting discussion as traditional data centres do not produce heat of a high enough grade to be very useful. However, higher density solutions to support the new AI / ML workloads will provide useful heat and, in some countries legislation is being introduced into municipalities to ensure they invest in the capability to reuse waste heat.

In this new era the industry is also placing an unprecedented emphasis on the benefits a data centre can bring to the local community beyond waste heat reuse. This includes striving to build facilities that are harmonious with the local environment, reducing the negative aesthetics of data centre buildings, providing local employment and potential upgrades to the local infrastructure.

The spotlight on sustainability is not just a buzzword but a strategic acknowledgment that data centres, powered by renewable energy, are integral to a future where efficiency and environmental consciousness go hand in hand. The technology industry and data centres must demonstrate a real commitment to sustainability and recognise the crucial role energy efficiency plays in the ongoing transformation of data centre operations. And the move towards larger campuses needs to align seamlessly with the imperative to reduce environmental impact.

It is clear that the data centre landscape is undergoing a profound evolution. The integration of AI and ML workloads, the redefinition of scalability, and the strategic development of AI ready megascale campuses collectively mark a new chapter in the story of data centres. This is not merely about keeping up with demand; it's about steering a course towards a data-driven future that is as dynamic as it is sustainable.

Embracing prefabricated modular data centres for scalable growth



Prefabricated modular data centres have emerged as a powerful solution in the digital age, offering unparalleled benefits for rapid and scalable growth in the data centre industry.

BY ALEX BREW, REGIONAL DIRECTOR, NORTHERN EUROPE AT VERTIV

BOOSTED BY the rise of artificial intelligence (AI), the digital era is ushering in a new age of technological advancements and unprecedented demand for data processing and storage. In this rapidly evolving landscape, data centres play a pivotal role as the nerve centres of our interconnected world. To meet the evolving needs of hyperscalers, enterprises and digital services, data centres are undergoing a transformative shift towards optimising efficiency and adaptability in both construction and operation.

One key driving force behind this transformation is the widespread adoption of prefabricated modular data centres. These innovative solutions offer a host of benefits, revolutionising the way data centres are designed, built and operated.

The power of this approach

Prefabricated modular data centres, commonly referred to as PFM data centres or integrated solutions, offer a multitude of advantages that revolutionise the way we approach data centre infrastructure. These benefits extend across various aspects of design, construction, performance, and geographic deployment.

One of the most prominent advantages of PFM data centres lies in their ability to expedite the design and deployment process. This is achieved through a unique prefabrication process, involving the off-site manufacturing and assembly of capacity units. Unlike conventional on-site construction, this approach allows for parallel activities, making it significantly more efficient. As a result, these modular data centres can be designed, tested and made operational in a remarkably shorter time frame compared to traditional construction methods. This rapid design capability is crucial in addressing the pressing need for data centre capacity, especially in the face of unpredictable and growing demand.

Furthermore, the modular nature of these data centres allows for a more agile response to demand. Instead of planning for unforeseen growth, organisations can build capacity to precisely match their current requirements. The modular approach enables them to scale their data centre infrastructure in a modular fashion, effectively aligning capacity with business demand. This minimises the risks associated with either overbuilding or underbuilding, ensuring a more efficient allocation of resources.

legrand®

We go further

UPS without compromise

Keor FLEX three-phase modular UPS

98.4% VFI Efficiency

Efficient
98.4% Double conversion efficiency

Scalable
Flexible & adaptable power systems 1.2MW parallelable up to 4.8MW

Available
Modular hot-swap & sync

Compact
1.2MW only in 2.2m

Resilient
Innovative & robust busbar design

Smart
Grid Support Predictive diagnostic

Scan here
to discover more

legrand.com/datacenter/gb-en/grey-space/ups-sts

#Legrand**ImprovingLives**

Another notable benefit of is the optimisation of component performance within a holistic system. These data centres take an integrated approach, enhancing the performance and efficiency of individual components by tightly integrating various systems, including power, thermal management, and IT components. By designing, configuring, and fabricating these components off-site, they work seamlessly together, contributing to the reliability and overall performance of the data centre.

Quality control is a crucial aspect of data centre construction, and prefabricated data centres excel in this area. The manufacturing and assembly of units in controlled environments result in higher quality controls. This not only increases the reliability of the components but also extends the geographic reach of data centre deployment. Prefabricated units can be transported and assembled in various locations, even in regions where traditional construction methods might face challenges. This flexibility in deployment enhances the adaptability and scalability of prefabricated modular data centres.

Delving into the distinction between standardisation and localisation

While the approach of standardisation is widely recognised for its role in streamlining and enhancing data centre deployment, it is essential to appreciate the subtle yet significant disparities between standardisation and localisation. These two methodologies each bring their own set of considerations and implications to the table.

Standardisation is rooted in the principles of uniformity and consistency. It involves the adoption of pre-established designs, components, and practices that are designed for replication across diverse deployments. Embracing standardisation empowers data centre operators to swiftly deliver a consistent and seamless experience in terms of infrastructure, functionality, and operational procedures. This approach not only facilitates the harmonious integration with existing systems but also paves the way for efficient scalability.

Conversely, localisation places a strong emphasis on customisation and adaptability to meet regional requirements and preferences. This approach is particularly pertinent in regions where regulations exhibit significant variations. It acknowledges that different geographical locations may be subject to distinct building codes, regulations, standards, and environmental considerations that wield a substantial influence over data centre operations.

Unlocking long-term success

So, while there are certainly subtleties to consider, as evidenced by the difference between standardisation and localisation, we can clearly see that as the data centre industry continues to evolve, the benefits of prefabricated modular data centres become increasingly evident. Accelerated deployment timelines, reduced costs, improved operational efficiency, compliance with regional requirements all contribute to the long-term success and establishment of this trend.

Prefabricated modular data centres have emerged as a powerful solution in the digital age, offering unparalleled benefits for rapid and scalable growth in the data centre industry. By embracing prefabricated modular data centres, or standardised elements of the approach, companies can harness their potential to unlock new opportunities, enable efficient and reliable data processing and storage, and drive economic growth.


HIMOINSA
A YANMAR COMPANY
HGY
SERIES
HIMOINSA
YANMAR
YANMAR
CRUCIAL BACKED UP DATA DEMANDS
THE BEST BACK-UP POWER
POWER & RESPONSIBILITY

Sustainability is good for business and good for the planet, and DCIM is critical to reduce energy consumption and waste

The introduction of new model based, automated sustainability reporting features within the company's EcoStruxure IT DCIM software - discussing the importance of accurate performance metrics as data centres respond to the business imperatives of energy efficiency and performance optimisation as well as the march of environmental legislation - with the EU Energy Directive already making an impact.

BY KEVIN BROWN, SENIOR VICE PRESIDENT, ECOSTRUXURE SOLUTIONS, SECURE POWER DIVISION, AND ALISON MATTE, SUSTAINABILITY LEAD ECOSTRUXURE IT, BOTH AT SCHNEIDER ELECTRIC

DCS: *Schneider Electric has introduced some new features to the* EcoStruxure *IT DCIM. It would be good to understand the industry background that has led to the need for these new features?*

KB: It's an interesting time for the industry because it really sits at the intersection where IT and OT meet. And that is fundamentally where DCIM tools exist. One of the things that we're seeing in the industry is a bigger focus on sustainability. Now, it varies in different parts of the world. Obviously, in Europe, you have the EED and some of the reporting requirements that are coming in. But in general, when you look at the energy consumption of IT and as well as data centres, and certainly with the recent AI boom, you're seeing a greater focus coming from governments and industry on what are we doing to ensure that the energy consumption of IT is being done in a responsible and professional way. And we really feel that DCIM is in a unique position to give insight for customers on that energy consumption, because we already have a lot of the data. If somebody's deployed a DCIM solution, we have the data.

But what we realised, and it came from a project we did with our own CIO team, is the tools weren't oriented in a 'user-friendly' way - it wasn't as easy to get that data and to organise it as required, number one. Number two, sometimes the data was incomplete. As a result, we've done some work recently so that customers can meet a very simple challenge of: how much energy is my IT consuming and can I report on it so that I can manage it more effectively? We're really quite excited about what we've brought to the market, and we think we're getting some pretty good response to it as well.

DCS: *Schneider Electric rightly has a well-earned reputation for being at the forefront when it comes to sustainability. It would be good to understand what you are seeing sustainability-wise within the data centre industry, and on a global basis, as to the actions required. Is progress fast enough or are we where we need to be, for example?*

AM: Everyone realises that data centres are consuming quite a significant amount of energy and electricity, and there is a massive strain on the grid to respond to the needs of every customer and every company. Where data centre growth is exponential in certain countries, we're seeing regulations coming into those countries as well to limit the amount of data centres that are being built, but also to start controlling the energy consumption of these data centres. Europe is at the forefront right now with the Energy Efficiency Directive and with other regulations. The EU Commission has set the EED, that's coming into force in September. Additionally, Germany, for example, is adopting its own set of rules. The Netherlands is doing the same. We're seeing regulations spread throughout not only the European continent, but we're also international regulations as well.

Sustainability is a way to limit the energy consumption. It is also a way to reduce costs related to energy and to reduce waste in data centres where all these IT assets that are being managed. We're seeing DCIM play a role, and we are seeing an opportunity to help customers manage those IT assets and also to report on the regulations that are coming in.

DCS: *In terms of the legislation that's now coming in, is it because the industry hasn't moved fast enough and so it's been given a little nudge, or do you think it was going to happen anyway? However well the industry performed, legislation was inevitable?*

AM: One of the main metrics that the industry has been tracking is PUE. That's been around for decades, and it's stagnating. I do think the regulations coming in is to push customers to start actually reporting, but also to align on key metrics and ways to understand what these data centres are actually consuming, because a lot of the times there are assumptions that are being made. Regulations are creating baselines for companies to be able to actually make progress. And regulations do help push things along.

KB: Just to augment what Alison is saying, all the metrics that are the EED, they all make sense from our point of view, and they really came from the industry, and they came from things that we're doing. As Alison is saying, the EED is formalising it and putting a structure around it. Where there's going to be a lot of discussion is around how good does your data need to be in order to report against that? Because part of the reason PUE is stagnating, using Alison's term, is because sometimes you don't have all the data. You get into some of the smaller data centres, server rooms, wiring closets (which, aren't covered by the EED, but we think they will be over time), and it's not easy to get some of these metrics, and I think that's really what's going to be interesting for those in Europe, right? And the other thing for customers in Europe concerns what level of detail do I need in order to report against the EED? And this is where we've been developing tools to model it out and make it a little easier.

You know, in the other parts of the world, like in the US, I would make the argument to many CIOs, the metrics that the EED has adopted are the ones that really we would consider best practise. So, why not use those to report everywhere? Our philosophy on it is, look, even if it's model-based, it's good to start getting a benchmark. It's good to start tracking this information, even down to the smallest sites, and then you can get better over time. So, your question is quite interesting. Is it because the industry wasn't doing the right things that the regulations came in? No, I think it's because the industry was doing the right things. Now, because it's becoming so prominent in the discussions, the EED is formalising it. For those in Europe, it's going to be a real question about what's the level of detail that I need in order to comply and be auditable against the regulation? But for other people in the world where you don't have quite the same strict regulation in place, the argument is that you should be doing this anyway. We have examples of our own CIO team implementing our tools and getting some great results from a business standpoint.

DCS: *In terms of the DCIM features that have been introduced by Schneider Electric, they are providing more of the required metrics and visibility. I believe the company talks about enhanced visibility, historical data analysis, and things like fast intuitive reporting. Fundamentally, the objective is to provide more of this information and, whether you developed it specifically for the directive or not, it's a lot of the information that people will be required to report on. Is that right?*

KB: Yes. Much of the learning that we did and what we built in was as a result of watching our own CIO team do their own green IT initiative internally. They started using our own tools. They were doing

Sustainability is a way to limit the energy consumption. It is also a way to reduce costs related to energy and to reduce waste in data centres where all these IT assets that are being managed

a refresh of their IT infrastructure. And the EED was in the background because we saw the EED coming. So obviously, those two things really played together. Much of the work we did was around trying to make it easier for somebody to get those metrics. And again, as I mentioned, one of the big challenges is that people don't always have the data, and we did not limit it. I think the EED, if I remember, it's like 500 kilowatts and above, right? Certain countries might have a lower threshold. We actually think you should be implementing this at even the smallest server rooms and wiring closets, and that's what our own CIO team was doing. When you have incomplete data, what do you do? The reality was that we've spent the last 10 to 20 years working on data centres and modelling out PUE and what that looks like.

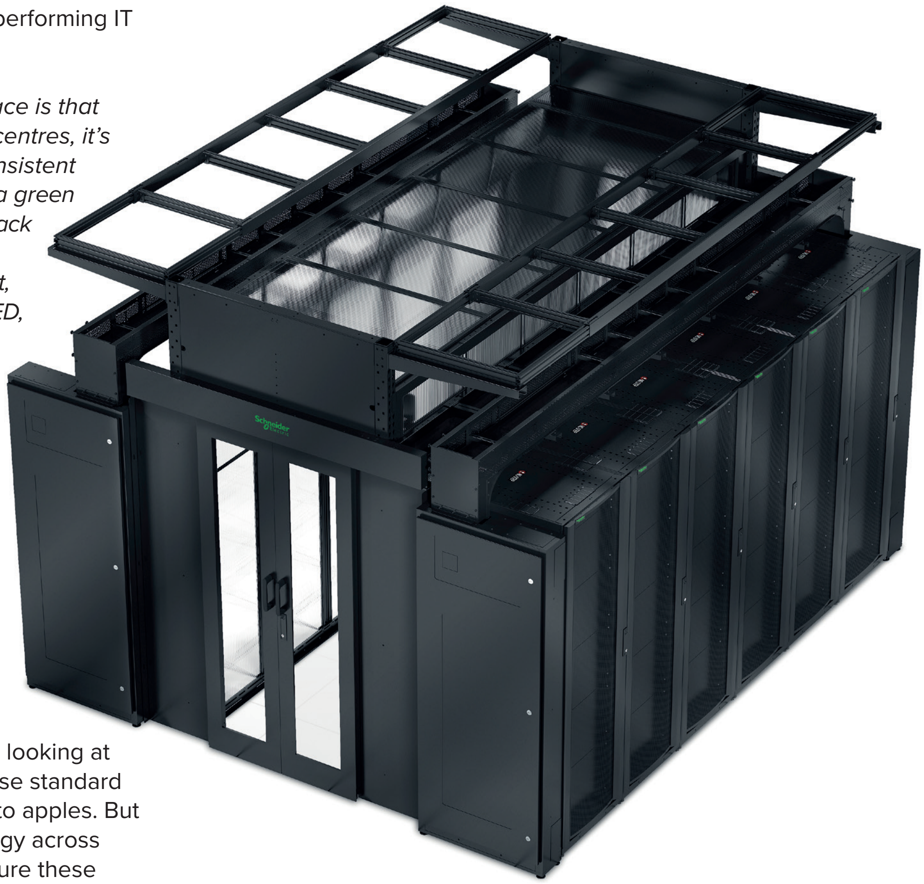
We've also spent the last 10 or 15 years, becoming a leader in sustainability as a business. You take that combined expertise - we say it's model-based, data-based modelling. Let's call it AI, because everybody has to use AI these days as a term! We looked at our knowledge of what we had internally, and we did use algorithms to help people start modelling out what these sites look like. And now you can obtain this information very easily - with the push of a button, you can get the PUE of a site in effect. Based on the model, we give people a level of confidence about what is the PUE at that site. And if we get more data, the confidence level gets better. And we're going to keep working on this because we have data coming into our cloud which we can continue to work on and analyse. These models are going to get better and better. And so in effect, it's a very easy way to get a baseline of what is your energy consumption, including the losses of the physical infrastructure. What we're most excited about is that we are enabling people to have a very fast and easy way to get started on this journey of really measuring what is your performance.

DCS: *I guess the next logical step at some stage, and it may already be happening to a certain extent, is once you've got all this information, clearly you can take actions based on it. I'm just wondering what the balance is at the moment and maybe plans for the future in terms of humans versus AI. AI has the potential to take a load of data and maybe recommend some actions based on it. What are you doing with it or what plans do you have with the decision making beyond the raw input of data that you're collecting?*

AM: I can give you an example from our own CIO team project. On our Lexington site, which had about 10 network closets, 10 IT rooms (so not a full data centre), they were able to start monitoring and tracking their energy consumption. They actually found zombie servers, and they found certain inefficiencies. With the recommendations that we were providing with our DCIM, alongside some upgrades to their assets, they were able to reduce their energy consumption by 20% in the first year. In carrying out that exercise, reducing 2 % their energy consumption, they reduced about 30% of costs. When you look at the sustainability side, they were able to reduce their carbon emissions by 17 tonnes that first year, just on that site, which is massive.

KB: The way this is going to play out is, when people go out and start measuring, it's going to point them in a direction about where's your worst performing site. And in the example Alison shared, they knew there was a lot of opportunities. Now, our software wasn't smart enough to tell them, hey, go virtualize this and go do that...you still need a human in there. But because they started running these tools and looking at the data, it pointed them in a direction to go. And what I think will happen is over time, as we start learning more from customers about what are their opportunities, you can see a day not too far in the future where we can sit there and go, based on what we see and based on benchmarks of other things that are similar, we think you may have a 30% opportunity at this site, and you should go look there. So that's really where we are in terms of the sophistication of the tool - it is giving people a very easy way to get a very broad view of their complex hybrid infrastructure and enabling them to start finding out where they should start looking and then really tracking the subsequent continuous improvement exercise.

Some of the savings that they got at our Lexington factory were because they went and did some higher-level IT stuff. They bought some new IT equipment, they virtualized things, and they were able to achieve, again, a 30% energy reduction, but that was a tremendous ROI from a business standpoint. Sustainability is good for business. I think that's an important message, and one we're trying to promote - when someone takes a sustainability initiative, a green IT initiative, they end up with a better solution that's more resilient and has a lower energy consumption. We maybe don't have don't have all the data as yet to support that assertion, but I think there's a lot of evidence that this will be the case. We're excited about this, not just because of the EED, but we think it's also going to lead to less waste and better performing IT infrastructure.



DCS: *One of the problems end users face is that when they're looking at different data centres, it's very difficult to obtain, shall we say, consistent statistics. Everyone will say, we've got a green data centre, and they might be able to back it up with some data, maybedepending on how they interpret it. Do you think that, with the combination of DCIM and the EED, at some stage we might arrive in a great place where you can compare five different data centres with the exact same set of metrics? You will know the actual performance of each data centre, as opposed to some of the marketing/ greenwashing that might still be going on?*

AM: They're using Cenelec methodology, which has been around for years, to calculate these metrics. There is a lot of standardisation. With these regulations that are coming into force, and the EED in particular, if we're looking at the same methodology to calculate these standard metrics, then you can compare apples to apples. But if you're not using the same methodology across your sites to be able to track and measure these metrics, you're not going to be able to compare. You're not going to be able to see, to have a good baseline. So that's the power of regulations and standards. It allows you to do so.

KB: Historically, we've all seen there's two things that have happened in the industry where not everyone's PUE was calculated the same way, and people would play some games with it. I think that's fair to say. And the second thing is, there's been a self-selection bias, when people go and publish 'we have a PUE of 1.15 or whatever'. Is that just for one data centre on the best day of the year? Is that an average number? What is that number? We don't know. And that's where I think, rightfully so, there's been some suspicion about some of these marketing initiatives, talking about how green data centres are. And that is a role where regulations and standards come in. It's all defined. So now the question is going to be, if you have to comply with the regulation and it's auditable, that's going to take away, let's call it, some of the game plan. And that's one of the things we did with the tools. We built in that methodology so that when you run a report, it is consistent with the standards that are in place, and it should give you a pretty good estimate.

And there's going to be a debate as to whether or not a model-based solution is good enough? Good enough to be able to comply with the regulations? But our broader point here is that, if I use these standards and I use a model-based solution for those areas where maybe I don't have all the metering in place that you need to comply with the regulations, it's still going to ensure that you're running better infrastructure, and that it's as energy-efficient as possible, it's as robust as possible. Let's not lose sight of this goal for the industry. We are looking at that broader goal as well as the narrower goal of can I report against the EED? We help you with that. But there's a broader goal about using these model-based, data-based AI engines, to help people get a handle on their whole infrastructure. Keep in mind, our best estimate is that at least half of energy consumption is outside of data centres.

If you look at energy consumption, about half of it is in server rooms, wiring closets, networking closets. And today, the EED doesn't cover that, but that's half the problem. It doesn't get the headlines that the large data centres get, but it's still there. And we would argue that over time, regulators like to stay busy, too. I mean, they're going to keep lowering this threshold. That's what I would guess. They're going to go from 500 kilowatts to 250 to 100. It's going to keep coming down. Why not get ahead of that on best practises and make sure that you are looking at your infrastructure, particularly when it's very cost-effective and very easy to do so?

This Q and A article is based on a video interview between DCS and Schneider Electric. The video can be viewed via the DCS website at: https://datacentre.solutions/videos/4660/schneider-electric-continues-to-drive-the-data-centre-sustainability-agenda

Transforming data centres to meet AI's evolving demands

While 2023 marked a pivotal moment in recognising the vast potential of artificial intelligence, 2024 kicked of what is becoming a truly transformative period. AI's broad applications, ranging from machine learning and deep learning to natural language processing have seamlessly integrated into our everyday lives, revolutionising how we live, work and connect.

BY KAMLESH PATEL, VP DATA CENTER MARKET DEVELOPMENT AT COMMSCOPE

As AI's popularity soars, data centre managers and their teams are grappling with the challenge of managing, not only the surge of petabytes of data flooding their networks, but also the need for ultra-low latency. Additionally, they are attempting to tackle the increased power demands and higher fibre counts needed to support the advancements that come from supporting AI.

Similarly, the rise of artificial intelligence has caused a fundamental shift in data centre design, significantly impacting network infrastructure in areas such as cabling, connectivity, architecture, resilience and adaptability.

Here are the key challenges and opportunities that I believe come with cabling AI data centres, some best practices and tips for success.

The unstoppable surge in power demand

Regions that house data centres are experiencing a surge in power demand. In the Republic of Ireland for example, data centres now consume over 20% of the country's electricity, a significant increase from just 5% in 2015. Consequently, for the first time ever, there is no longer a guarantee that the power needed to support data centre operations can be reliably supplied.

Recently, the 'net zero' goals of major tech companies have been challenged by this increasing power demand, a direct consequence of AI and energy-hungry data centres. Google reported a 48% increase in its greenhouse gas emissions over the past five years, largely due to the growth of its data centres, while Microsoft's Scope 3 emissions have risen by over 30% since 2020. To strike a balance between enhancing sustainability and expanding capacity and performance, data centres will require support from their infrastructure technology partners.

Ultra low latency meets ultra high connectivity solutions

Because the models used to train and run AI consume significant processing capacity and are typically too much for a single machine to handle, processing these large AI models requires numerous interconnected GPUs distributed across multiple servers and racks. This presents a unique challenge for the cabling infrastructure that links everything together to keep data flowing.

For instance, GPU servers demand significantly higher connectivity between servers, but due to power and heat limitations, fewer servers can be housed per rack. As a result, AI data centres require more inter-rack cabling compared to traditional data centres. Each GPU server is linked to a switch within the same row or room, with these connections needing 400G and 800G speeds over distances that traditional copper cables like DACs, AECs or ACCs can't handle. Moreover, every server must also be connected to the switch fabric, storage, and out-of-band management.

In an ideal setup, GPU servers in an AI cluster would be close together, because AI and machine learning algorithms - like high-performance computing (HPC) - are highly sensitive to latency. It's estimated that 30% of the time spent running a large training model is due to network latency, while 70% is spent on compute time. To reduce latency, AI clusters strive to keep GPU servers in close proximity, with most links limited to 100 metres. However, not all data centres can place GPU server racks in the same row. These racks require over 40 kW to power a GPU server, far more than typical server racks, forcing traditional data centres to space them out accordingly.

Although extra space isn't feasible in the densely packed server rack layouts of modern data centres, managing the narrow, congested pathways and the added cabling complexities brought by AI is made possible through innovations like rollable ribbon fibre.

The innovative design allows for the installation of up to six 3,456 fibre cables within a single four-inch duct, providing more than double the density compared to traditionally packed fibres.
In the rollable ribbon fibre cable, the fibres are attached intermittently to form a loose web. This design makes the ribbon more flexible, allowing the fibres to flex with a degree of independence from one another. The fibres can now be "rolled" into a cylinder, making much better use of space when compared with flat ribbons.

While the cables are lighter and simplify handling and installation, their intermittent bonding enables installers to position the fibres naturally into a smaller cross-section making it perfect for splicing.

In an ideal setup, GPU servers in an AI cluster would be close together, because AI and machine learning algorithms - like high-performance computing (HPC) - are highly sensitive to latency

Data centre architecture of the future

Looking to the future, the value proposition for data centres will hinge on their extensive processing and storage capabilities and operators need to thoughtfully select the optical transceivers and fibre cables for their AI clusters.

In an AI cluster, the optics cost is primarily driven by the transceiver due to its short links. Transceivers that utilise parallel fibres are particularly beneficial because they eliminate the need for optical multiplexers and demultiplexers, which are typically required for wavelength division multiplexing (WDM). This results in reduced costs and lower power consumption for transceivers with parallel fibre.

Links up to 100 metres are supported by both singlemode and multimode fibre applications and advances such as silicon photonics have lowered the cost of singlemode transceivers.

In many AI clusters, active optical cables (AOCs) are used to interconnect GPUs spread over many servers and racks. These cables are usually designed for short distances and are commonly used with multimode fibre and VCSELs. The transmitters and receivers in an AOC may be the same as in analogous transceivers but are the castoffs. These components don't need to meet stringent interoperability requirements since they are only required to work with the specific unit attached to the other end of the cable. Additionally, since the optical connectors are not accessible to the installer, there is no need for specialised skills to clean and inspect fibre connectors.

Strategic planning for AI cluster cabling

In summary, data centres must evolve and adapt to meet the growing demands of artificial intelligence in business applications and customer service delivery. Infrastructure designers and planners must focus on improving efficiency, scalability, and sustainability. Key to these advancements is the upgrade of cabling systems, which will help reduce costs, energy usage, and installation times. By embracing these innovations, data centre facilities will be well-equipped to manage both current and future AI-driven workloads.

Data centres and augmented reality:
a blueprint for industry transformation

The days of physically configuring server rooms filled with cryptic codes and blinking lights are becoming a thing of the past as Augmented Reality (AR) begins to make its mark on the data centre sector. This transformative technology is taking root at the beating heart of the digital world, completely revamping data centre operations.

BY JAB JEBRA, PRESIDENT AND CEO OF HYPERVIEW

THIS IS NOT MERELY a glimpse into the future, but a tangible revolution reshaping the present, and we are only just witnessing the beginning of how AR will transform business operations across all sectors. According to a McKinsey report, 2.7 billion deskless workers, representing roughly 80% of the global workforce, could become the primary users of immersive reality technology. This potential isn't just confined to distant possibilities, it's unfolding right now, with data centres creating the blueprint for other industries.

AR is not just streamlining data centre operations, but it also emerging as a powerful force in driving sustainability initiatives for organisations. With the infusion of AR in business, we may be looking at a world where most business travel comes to an end. AR is proving to be a secret weapon that propels us towards a reality where business operations are more efficient, more sustainable, and more responsible.

The infusion of AR and Data Centres

But first, the marriage between AR with Data Center Infrastructure Management (DCIM) starts with identifying the need for AR in data centre operations. The next step involves selecting compatible AR and DCIM platforms and connecting

them, usually via APIs. AR models of the data centre infrastructure are then developed and overlaid onto the physical world when viewed through an AR device. The integration is tested for accuracy, and data centre staff are trained on how to use the new system. The final step involves continuous monitoring and improvement of the AR system to ensure its effectiveness and efficiency in managing operations.

AR lets users explore a digital twin of the entire data centre. Conveniently scan QR code, bar codes and tags with a glance, conduct audits and inspections remotely, collaborate with experts worldwide – all while troubleshooting issues swiftly. Downtime, a data centre's worst nightmare, and typically caused by human error, is drastically reduced. Faster problem-solving and preventative maintenance minimise the risk and duration of outages, ensuring smooth operation and satisfied customers.

Without physically being there, AR enables data centre operators to monitor server health, temperature, and energy usage, among other critical metrics. This enhanced visibility is a game-changer for decision-making processes, with operators able to make informed decisions based on real-time data, rather than relying on periodic reports or manual checks.

Imagine a technician facing a massive server rack. No longer just metal and wires, the equipment comes alive with real-time information floating within view due to an AR headset. Data flows and energy usage are visualised, helping identify areas for optimisation.

In essence, AR empowers data centres with unprecedented visibility and control. It bridges the physical and digital worlds, providing technicians with the information they need, exactly where they need it. Ultimately, this leads to improved decision-making, faster problem-solving, increased operational efficiency, and reduced downtime, all of which are crucial for the success and competitiveness of data centres in today's digital age.

Sustainability: A shared responsibility

Beyond purely operational benefits, AR also plays an important role in driving sustainability efforts within data centres. As environmental consciousness is on the minds of business and society as a whole, data centres are embracing this technology to help reduce their environmental impact.

For example, business travel is important for employees so that they can have visibility of their operational hardware, but on the flip side, business travel is proving a major detriment to the future of our planet. According to the IEA, air travel from business trips contributes about 2% of the world's harmful emissions. This is where AR can help.

Remote monitoring and troubleshooting made possible by AR dramatically lessen the need for on-site visits. This translates directly to a smaller carbon footprint as emissions tied to travel are reduced.

This use of AR in data centres provides a blueprint for other industries to follow suit to strive for a greener future, especially when it comes to reducing travel. It's a solution that is universally applicable, given that business travel is a staple for most companies—whether for hosting crucial meetings in distant locales or necessitating in-person inspections and audits—its applicability spans across all sectors.

Furthermore, the insights gained through AR data play an important role in identifying energy inefficiencies within the data centre. This information allows operators to optimise operations and reduce energy consumption, making the data centre a more sustainable and responsible part of our digital infrastructure.

In short, AR encourages remote collaboration and virtual experiences, providing a green solution for many industries. It leads the way to a future with less unnecessary travel and more energy efficiency, where business and sustainability work hard in hand.

Time to turn the tide with AR

Looking ahead, integrating Augmented Reality (AR) is set to become a cornerstone of successful business operations across numerous industries. Data centres that are adopting AR, serve as a compelling blueprint for other sectors. By embracing this transformative technology, businesses can reimagine their operational efficiency and contribute to a more sustainable and responsible future.

The time for cautious observation has passed. AR adoption as a companion to DCIM is no longer a question of "if" but "when." Businesses that fail to embrace AR risk being left behind as their competitors reap the benefits of increased efficiency, improved decision-making, and a more sustainable future.

How eCommerce and fintech firms are leveraging the transformative power of AI-enabled data centres

Andy Connor, EMEA Channel Director at Subzero Engineering, examines the impact AI, blockchain and process automation is having on the eCommerce and fintech sectors.

THE FINTECH and eCommerce sectors are booming. By 2026, the global eCommerce market is expected to be worth more than $8.1 trillion1 annually. Meanwhile a recent study by McKinsey revealed there are now more than 272 fintech unicorns, with a combined market cap of $936 billion – a sevenfold increase in just five years.

Technologies such as artificial intelligence (AI), blockchain and process automation remain critical to sustaining this growth, with companies such as NVIDIA turning to GPU-powered servers to overcome the limitations of conventional data centre infrastructure.

Designed originally to accelerate computer graphics and image processing, GPUs perform complex three-dimensional vector calculations, enabling them to perform multiple operations simultaneously.

Making sense of vast amounts of data

This ability to multitask on a huge scale makes GPUs an ideal tool for managing and analysing vast volumes of data. For example, GPUs, can process neural network training data up to 250 times faster than conventional CPUs. They can also do the job more accurately, making them particularly well suited to high-quality data-driven decision making. For example, credit specialist Capital One uses a suite of GPU-optimised data libraries to accelerate its data science and analytics pipelines. The firm has not only achieved a 100-fold increase in data model training times, it has also reduced its costs by nearly 98%2.

This uptick in processing power means forward-thinking firms now have an opportunity to explore data models faster and with greater confidence. They can also do so in a more cost-effective and energy-efficient manner, and with a faster time to ROI.

Automating and improving processes in financial services

Financial services firms are among those with the most to gain from the huge potential of GPU-powered AI, and many companies are already leveraging this technology to automate and improve mission-critical processes. These use cases include.

- **Algorithmic trading:** analysing historical market and stock data to generate investment strategies, build portfolios and automatically buy and sell investments. Established banks which have developed algorithmic trading strategies include BMP Paribas, Deutsche Bank and Credit Suisse.
- **Detecting fraud:** combatting the most sophisticated types of transaction and identity fraud, increasing fraud detection accuracy, and boosting anti-money laundering and know-your-customer regulation. American Express, BNY Mellon and PayPal are already using a form of AI called natural language processing (NLP) to detect and prevent financial fraud.
- **Accelerating payments:** Fintech payment firms such as PayPal are using machine learning to improve payment authorization rates on their platforms. One way it does this is by predicting

and efficiently managing instances where a bank could decline a payment.

Achieving competitive advantage in eCommerce
eCommerce companies are also waking up to the power of AI, using the technology in innovative ways to achieve competitive advantage. Among some of the innovative use cases in the eCommerce sector include:

- **Predicting and managing customer churn:** AI can analyse customer behaviour and identify which customers are most likely to make repeat purchases This enables sales and marketing teams to target and better allocate resources to more profitable customers.
- **Dynamic pricing:** eCommerce giants such as Amazon and eBay are long-term advocates of AI-powered dynamic pricing. They use the technology to analyse market demand, competitor pricing and other factors so they can adjust their prices in response.
- **Enabling visual and voice search techniques:** Retailers such as ASOS, Forever 21 and Home Depot are using AI to free their customers from their keyboards, find the products they really want and accelerate their path to purchase.

It's interesting to note that financial services and ecommerce, as well as many other sectors, are exploring how AI can drastically improve online customer interactions. Thanks to generative AI (GenAI), clunky chatbots will soon be a thing of the past. Instead, GenAI-powered solutions are particularly good at finding the best answers to customer questions and sharing that information in a human-like way. Deutsche Bank, American Express and Wells Fargo are among the banks that are starting to go live with such GenAI-powered solutions.

Using AI to optimise the data centre

There is clearly huge scope for using GPU-powered AI to improve products and services within the eCommerce and financial services sectors, but the benefits don't end there. The technology is also transforming how data centres critical to these sectors are managed and optimized, how uptime is ensured, and higher levels of sustainability are achieved.

AI is not only helping data centre managers identify, troubleshoot and mitigate outages in a reactive way, it is also automating this process, predicting faults and triggering self-healing mechanisms. For instance, AI can be trained to identify unusually slow traffic within a particular node, and then re-boot a process or the entire node fix the issue. Other tasks such as energy management (including cooling and power management), inventory management and systems update management can also be automated in a similar way using AI.

AI is also being used to boost data centre visibility and decision making, helping managers to identify opportunities to optimize resource allocation and improve both workload management and capacity planning. Such processes can reveal golden opportunities to right-size data centre infrastructure, cut power consumption and reduce environmental impact.

Research by Schroders3 suggests that AI-related data centre power consumption is likely to increase sevenfold to 7GW by 2026. However, this spike in energy consumption and associated carbon emissions can be reduced, at least in part, through the careful use of GPUs. That's because GPUs are more powerful, fewer servers are needed, and data centre physical footprint and cooling requirements are reduced.

Simplifying and accelerating financial transactions with blockchain

Blockchain is another GPU-reliant technology helping to disrupt both the financial services and ecommerce sectors. Fintech firms such as OpenZeppelin are harnessing the power of smart contracts to simplify complex financial transactions. Smart contracts, which are powered by blockchain, automatically execute when certain conditions are met. They remove layers of intermediaries, reduce cost, and speed up contract execution. For example, a smart contract can be programmed to make a payment when a product or service has been successfully delivered.

Meanwhile Ripple, one of the best-known blockchain-based payment systems, enables banks, corporations and crypto exchanges to transfer money without the need for a third-party processor. The firm's solution has made cross-border payments significantly easier, faster, cheaper and more secure.
The data centre as the backbone of innovation
As transformative technologies such as AI and blockchain continue to evolve and become more integral to the success of the fintech and eCommerce sectors, the role of data centres becomes increasingly critical.

Data centres equipped with GPUs deliver the necessary computational power that AI applications require. As we continue to push the boundaries of what AI can achieve, the role of data centres will only grow in importance. They are not just a supporting infrastructure, but a vital component in the journey towards a more AI-driven future.

FURTHER READING / REFERENCE

- 1. https://www.forbes.com/advisor/business/ecommerce-statistics/
- 2. https://blogs.nvidia.com/blog/ai-accelerating-financial-services/
- 3 How AI is set to accelerate demand for data centres (schroders.com)