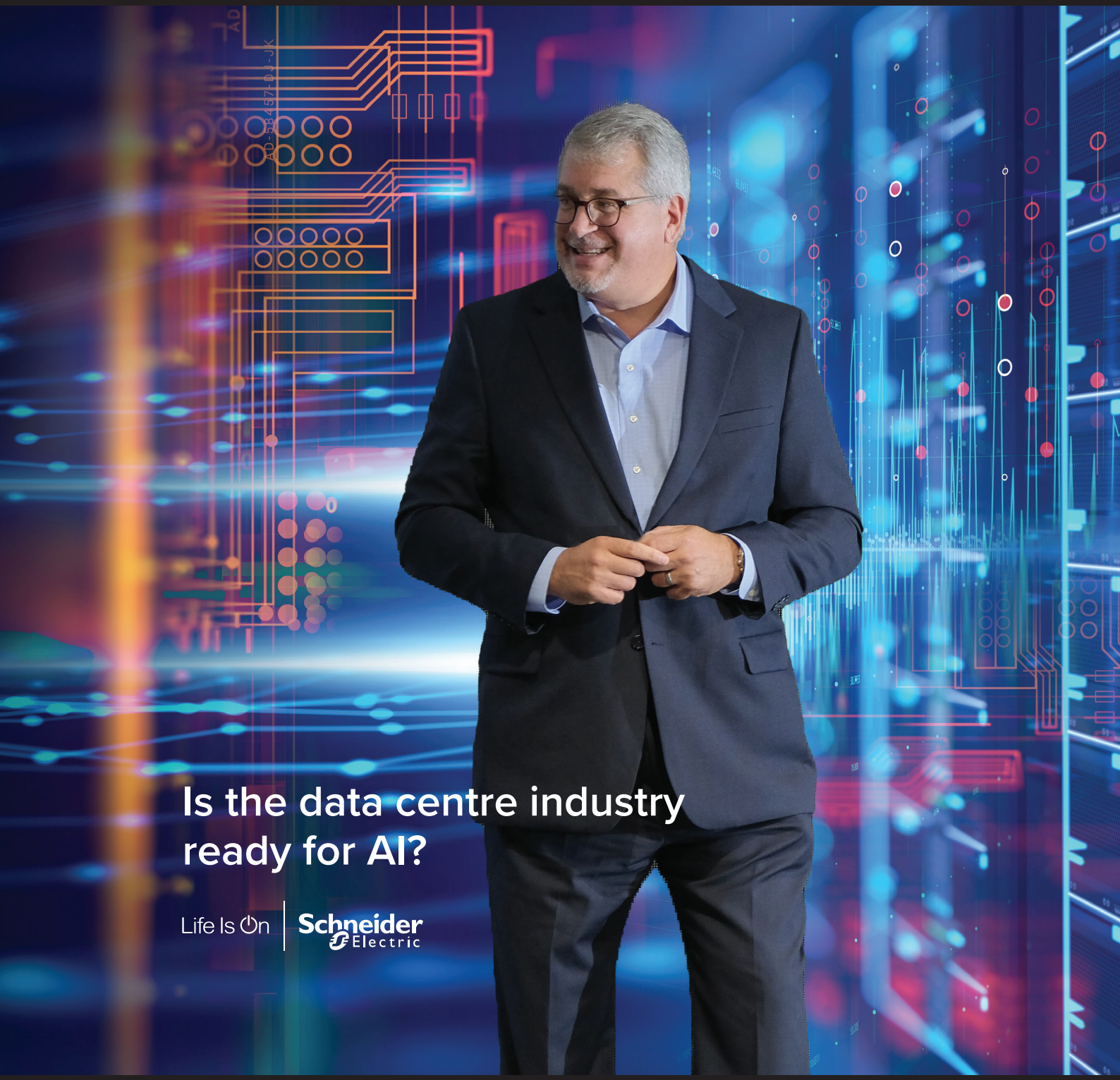
DCS DATACENTRE SOLUTIONS

DEVELOPING DIGITAL INFRASTRUCTURE IN A HYBRID WORLD

ISSUE IX 2023

HTTPS://DATACENTRE.SOLUTIONS



Cloud | Design + Optimisation | Energy Management | Infrastructure + Asset Management
Hosting + Colocation | Infrastructure + Hardware | Power & Cooling

AN ANGEL BUSINESS COMMUNICATIONS PUBLICATION

EcoStruxure™
Innovation At Every Level

Realize your company's IT potential

Software to design, monitor, and manage your IT space

62% of IT outages can be attributed to IT infrastructure failure[1]. Our Data Center Infrastructure Management (DCIM) 3.0 offer provides device monitoring, health assessments, and more so you can:

- Run simulated impacts to expose vulnerabilities in IT infrastructure and address them immediately
- Reduce physical threats by monitoring IT environmental conditions
- Improve sustainability efforts by tracking PUE, energy, and carbon emissions

#CertaintyInAConnectedWorld

apc.com/edge

EcoStruxure™ IT modernized DCIM

APC Smart-UPS™ Modular Ultra

[1]Uptime Institute Global Data Center Survey, 2018

Life Is On | Schneider Electric

VIEWPOINT

By Phil Alsop, Editor

AI – about to rewrite the data centre rules?

AS OUR COVER STORY demonstrates, the advent of AI promises a level of disruption within the data centre and wider IT sectors perhaps never witnessed before. In recent memory, first virtualisation, then cloud, brought with them a significant degree of change. But, for the data centre at least, the job remained largely the same – providing a suitable environment in which to house actual IT hardware (on top of which may well be sitting various levels of software virtualisation). And, of course, when it comes to AI, the data centre is, once more, expected to house the necessary IT hardware. Only, this time, the hardware in question is dramatically different from the servers, storage and switches required to underpin virtualisation and the cloud. We're talking rack and cabinet densities never seen before (well maybe in a few high-performance compute environments), and at significant scale. All of which means, existing data centres can be adapted to host some AI workloads, but almost certainly with a great deal of empty space around the requisite IT hardware, else the generated heat, and required cooling, would be off the charts.

And so, data centres built specifically for AI workloads are required. And, right now, there aren't nearly enough of them to cope with the anticipated AI demand. Of course, no one quite knows just how far AI will reach into the world of 'every day' IT, hence the actual demand, but plenty of expert judges think the AI requirement and, hence the data centre infrastructure required, will dwarf the cloud 'explosion'.

Add in the sustainability focus, and it's not difficult to see a data centre industry struggling to both meet customers' AI demands whilst also reducing carbon footprint on the way to Net Zero. The optimists will rightly point out how the data centre industry has to date managed to respond successfully to the challenge it has faced to date. The pessimists will underline the fact that never has their been such a significant IT revolution placing a new level of demands on the data centre. The truth, as ever, almost certainly lies somewhere in the middle.

I have no magical insights or answers, but would highly recommend close reading of the cover story to understand how AI threatens to re-shape the data centre industry. Forewarned is forearmed!

A Happy Festive Season to one and all!

CONTENTS

ISSUE IX 2023

DCS DATACENTRE SOLUTIONS

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

Marketing & Logistics Executive
Eve O'Sullivan
+44 (0)2476 823 123
eve.osullivan@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

Datacentre Solutions is published 6 times a year on a controlled circulation basis in Europe, Middle East and Africa only. Subscription rates on request. All information herein is believed to be correct at time of going to press. The publisher does not accept responsibility for any errors and omissions. The views expressed in this publication are not necessarily those of the publisher. Every effort has been made to obtain copyright permission for the material contained in this publication. Angel Business Communications Ltd will be happy to acknowledge any copyright oversights in a subsequent issue of the publication. Angel Business Communications Ltd. ©

ISSN 2756-1143 (Online)

Europe's top data centre hubs produce 1.2 million homes-worth of emissions

Relocating data centre infrastructure from existing hubs to emerging hotspots with greener energy networks could cut their associated CO2 emissions by up to 91%, according to new research from DataVita.

THE DATA CENTRE and multi-cloud services provider, working with digital sustainability consultancy Posetiv, analysed the 'carbon intensity' of Europe's top five data centre markets: London, Frankfurt, Amsterdam, Paris, and Dublin1.

Carbon intensity is a measure of the amount of CO2 generated per kilowatt-hour of electricity used. Countries with more renewable sources in their energy mix tend to score low, while those relying on fossil fuels have a higher associated carbon intensity.

DataVita found that the electricity usage of the five cities' data infrastructure amounted to a combined annual 6.3 million tonnes of CO2 – equivalent to the carbon footprint of powering 1.2 million average UK households2. Transitioning all five cities' collective data centre footprint to Scotland could cut total emissions by more than 80%. Frankfurt had the highest annual carbon footprint associated with data centre infrastructure, at 2.4 million tonnes of CO2 – relocating its capacity to Scotland could reduce emissions by 91%.

Southern Scotland – including the country's central belt between Edinburgh and Glasgow – had an average carbon intensity of 54.97 grams during 2022, according to National Grid data. This is around one-quarter of London's 200.46 grams and nearly nine times lower than Frankfurt's 473.62 grams.

Scotland is emerging as a data centre hub, with new facilities such as the country's first 'metro' data centre in Glasgow opening this year at 177 Bothwell Street, a newly developed office building in the city centre. Danny Quinn, MD of DataVita, said:

"If organisations are serious about reducing the environmental impact of their IT services, they need to take a close look at where the critical infrastructure they use is located. Looking to a market like Scotland – where renewable sources account for a much larger portion of the energy mix – could cut related emissions by 84-91%. "Data centres have made great strides optimising power usage effectiveness, water usage, waste heat recycling, and other efficiency measures, but sheer processing demand growth continues to outpace these improvements. The challenge will only increase as organisations increasingly rely on IT services and net zero targets draw nearer.

"There is no advantage in terms of latency and performance by using a data centre in these markets for the majority of IT workloads – particularly when it comes to London and Dublin. Any Scotland-based companies looking at large-scale cloud adoption should reflect on these metrics and consider whether potentially quadrupling their emissions, by choosing a provider in a one of the main markets, is a sensible strategy when delivering your services." Mark Butcher, MD of Posetiv, added: "Scotland's energy system offers prime sustainability for these data-intensive cities. In fact, transitioning all five cities' collective data centre footprint to Scotland could cut total emissions by over 80% – placing sustainability firmly alongside surging digital demand.

"Renewable-rich locations offer a way to fundamentally pair commercial growth with absolute emissions reductions – an essential combination in the climate era. While no single solution will address the industry's emissions alone, strategically building out lower-carbon hubs in locations like Scotland is an important piece of the net-zero puzzle."

Data centre sector eyes key emerging markets to deliver on soaring demand

Average global cost increase of six percent from 2022 to 2023, with some emerging markets seeing as much as a 22 percent jump.

AN UPSURGE in data centre construction in emerging markets is driving up costs, according to research from global professional services company Turner & Townsend.
Analysing 46 global markets, the Data Centre Cost Index 2023 looks at cost trends across the sector. The research analyses data centre benchmarking costs from over 200 projects in over 20 countries, alongside insight from 246 industry experts.
Longstanding hotspots top the rankings of the most expensive places for constructing data centres. The top five locations by cost - Tokyo, Zurich, Silicon Valley, New Jersey and Singapore – remain unchanged overall from 2022, though different in order. Tokyo, at US$13.7 per watt, has displaced Zurich, now US$13 per watt, to take the top spot.

The more striking trend highlighted by the report is of rapid cost escalation in emerging markets. The average cost of data centre construction globally has risen by six percent over the last year, softening slightly from eight percent in 2022. By comparison, cost increases of between 11 and 22 percent have been seen in seven markets across Asia, Africa and Latin America.

The trend has been notable in south-east Asian markets. Jakarta (US$10.5 per watt) has risen to the seventh most expensive market globally for data centre construction, while Kuala Lumpur (US$10 per watt) has climbed to the 13th position in the rankings. The rapid digitalisation and economic growth in parts of Asia is creating new sources of opportunity in under-served markets.
Reflecting the rising demand and costs in emerging markets, Riyadh in the Kingdom of Saudi Arabia (KSA) is a new entrant to the index with an average cost of US$10 per watt. The growth in this market is being driven by strong investment in digital connectivity and a growing range of giga-projects to support the government's nation-building agenda.

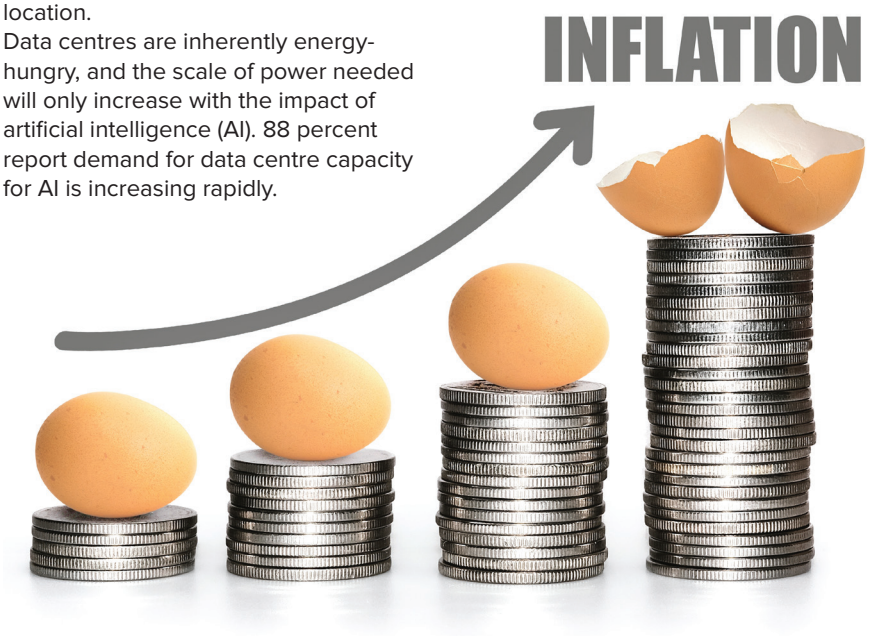
The largest year-on-year increase in price was seen in Cape Town, up from US$6.5 to US$7.9 per watt. Tokyo saw the second largest increase from US$11.4 to US$13.7 per watt. The overall results point to a narrowing of cost variation between traditional hubs and some emerging ones.

Soaring demand for data centres is being constrained by labour shortages – fuelling increased costs. 94 percent of survey respondents report skills shortages and 85 percent report 'hot' or 'overheating' conditions. The report identifies power availability as an increasingly dominant factor in driving investment decisions. Nine in 10 (92 percent) of respondents to Turner & Townsend's survey said that access to power is now more important for data centres than their geographical location.
Data centres are inherently energy-hungry, and the scale of power needed will only increase with the impact of artificial intelligence (AI). 88 percent report demand for data centre capacity for AI is increasing rapidly.

Despite the challenges facing data centre construction, respondents are optimistic about the resilience of the industry. 79 percent of those surveyed see the sector as recession-proof.

Commenting on the research, Rebecca Best, Director, UK Data Centre Cost Management lead at Turner & Townsend, said: "The data centre sector continues to provide immense opportunity. As individuals and businesses, our reliance on data is set to be further accelerated through the growth of artificial intelligence and machine learning. Demand for data storage is booming.

"However, on the other side of the coin there are growing challenges to delivery. Power availability, supply chain issues, skills shortages and increases in construction cost, alongside an urgent need to be more sustainable are converging to place hurdles in front of construction delivery.



Fewer than 50% of organisations expect to meet decarbonisation targets by 2030

Siemens launches study of 1,400 executives globally revealing regional, city and industry insights regarding the infrastructure transition across energy systems, mobility and buildings.

SIEMENS SMART INFRASTRUCTURE has released key insights into the divisive nature of the infrastructure transition in a new report, titled "Siemens Infrastructure Transition Monitor 2023: The Great Divide on The Path to Net Zero." Data from the report reveals that there is limited alignment on priorities and how best to progress towards a decarbonized and resource-efficient world. Whilst more than half of people surveyed believe the infrastructure transition is accelerating in their region, a quarter of participants - senior executives from seven major industry groups - said that progress is "too slow", while 29 percent believe progress is "coordinated", and 31 percent describe it as "on target".

The study set out to measure the current state of the infrastructure transition, including developments within the systems, services, buildings, and structures that are needed for industries, cities, and countries to function effectively. Data was collected through a global survey of 1,400 senior executives from 22 countries, as well as a series of in-depth interviews with leaders and experts.

The guiding principles behind the research outlined in the report include the necessity of the infrastructure transition to have a positive impact beyond decarbonization. Secondly, smarter infrastructure integration is mandatory to affect change. Finally, action must be undertaken urgently and at top speed to avert disastrous global consequences.

Matthias Rebellius, managing board member of Siemens AG and CEO of Smart Infrastructure, said: "The infrastructure transition is accelerating, putting pressure on systems worldwide – from energy, to mobility, to buildings.

Evolving the world's infrastructure is of the utmost importance to enable progress towards decarbonization, resource efficiency, and social wellbeing. Technology and digitalization are instrumental to achieving this transition in a smart and sustainable way. At Siemens Smart Infrastructure we have already taken the first steps, creating innovative products, systems, solutions, and services to support the present and future challenges of urbanization and climate change."

Change is not happening fast enough at the regional (country) level

Despite the acceleration of the infrastructure transition, faster progress is needed at the regional (country) level to support a low-carbon world. Energy is a key priority as almost three quarters of global greenhouse gas emissions come from production, use and transportation of energy. According to the report, less than 10 percent believe their region (or country) to be "advanced, fully integrated, full-scale" on major energy goals of the transition. According to McKinsey, to decarbonize the world's energy system would require an estimated USD 275 trillion to make deep changes to electrical power generation, distribution, and consumption.

Regulatory authorities are seen as having the greatest responsibility here (according to 31 percent of respondents), closely followed by the ultimate owners of assets, investors/ shareholders (25 percent). Businesses (17 percent), politicians (13 percent), and citizens (13 percent) are all described as having some responsibility, but significantly less.

Decarbonization is a competitive advantage for cities

In the fight against climate change, cities have a major role to play. In the survey, half of respondents (51 percent) believe that being ahead in decarbonization is a competitive advantage for a city. Decarbonizing mobility, including public transport networks and commercial and private vehicles is a priority to reduce emissions. 45 percent of respondents feel their cities have made progress to encourage the use of public transport.

However, according to the report, 44 percent also believe that the privatization of public transport would speed up decarbonization. In terms of feasible mobility policies, 46 percent of executives believe that subsidies or taxes should be used to make electric cars cheaper than combustion engine vehicles. Currently, the lack of charging infrastructure was found to be the biggest barrier to widespread adoption of electric vehicles.

Only 40 percent of organizations expect to reach decarbonization targets this year

Businesses are under pressure to decarbonize their business models, assets, and infrastructure. According to the report, nearly half have targets for Scope 1 and 2 emissions (47 percent). Only 40 percent think it is likely that they will meet their targets for the year ahead and just 44 percent expect to meet their 2030 targets. The report indicates that there could be a correlation between confidence in organizational growth prospects and confidence in decarbonization targets.

Lack of cloud management wastes IT resources

Aptum has recently released part two of its annual Cloud Impact Study 2023 - Maximizing Value: Controlling costs and optimizing cloud spend. According to the study, 71% of IT professionals surveyed stated that cloud-related costs make up 30% or more of their total IT spend.

THE STUDY INVOLVED 400 senior IT professionals from organizations with 250+ employees across the U.S., Canada, and UK. It explores the trends in cloud cost management and how these trends have evolved over time.

In today's challenging economic climate, organizations are becoming increasingly aware of the financial implications of their business operations. While the cloud has brought benefits such as flexibility, scalability, agility, and cost efficiency to organizations in recent years, there are still unforeseen costs. More than half (52%) of IT professionals admitted their organizations have wasted significant IT spend due to inefficiencies with cloud platforms and services.

In fact, 73% of IT respondents reported that their cloud investment has resulted in higher-than-expected IT costs within the last 12 months, marking a 28% increase compared to 2021 figures. Additionally, the majority (92%) of IT professionals stated that they are likely to conduct a full ROI (return-on-investment) analysis of cloud spending, a rise from 89% in 2022. But what is causing these increased costs?

The survey identified several factors, including poor planning, a lack of internal expertise, the acceleration of cloud adoption, and a lack of familiarity with complex cloud solutions. These challenges are considerably magnified when organizations implement a hybrid, multi-cloud strategy that incorporates both on-premises and public cloud services for different data services and workloads.

Currently, 62% of respondents agree that a lack of internal expertise has prevented them from expediting cloud implementations. However, to mitigate

these escalating costs, developing skills and knowledge in cloud strategy is crucial.

"Many organizations believe they are excelling in the cloud, but in reality, like flying in a cloud, they have to rely on instrumentation, which is usually inadequate to begin with, and piloting experience which takes time to develop," said Ian Rae, CEO & President at Aptum. "It isn't the cloud that is costly but the lack of observability and governance which hinder management and lead to cost overruns. This shortfall significantly hampers the ability to unlock the cloud's potential for enhancing operational efficiency while controlling costs."

The results demonstrate that organizations struggle to fully realize the value of their cloud spending as they expand their adoption of cloud services. The lack of a comprehensive, multi-cloud strategy poses a challenge for many IT leaders, preventing them from fully harnessing the transformative capabilities of the cloud. But continued cost savings rely on a clear cloud investment strategy that focuses on visibility and control of costs – especially for those accelerating cloud investment and moving into more complex, multi- and hybrid cloud environments. FinOps principles are key to ensuring organizations are operating efficiently in the cloud:

1.Cost Visibility and Accountability: FinOps helps gain visibility into cloud expenditures, allowing teams to be more conscious of their spending.
2.Optimized Resource Usage: FinOps practices help identify underutilized or unused resources, leading to termination or downsizing.
3.Enhanced Decision-Making and Forecasting: With clear insights, informed decisions can be made based on usage analysis, reducing the likelihood of unexpected overruns.

4.Cultural Change: A FinOps culture encourages cost to be an organizational mindset and a standard part of operational discussions.
5.Continuous Improvement: The iterative nature of FinOps ensures continuous improvement to cloud financial management practices, leading to ongoing savings.

IT teams confident about hitting net-zero targets, yet challenges still persist

Telehouse research finds 81% of IT decision-makers are confident about achieving net-zero targets.

IN LIGHT OF mounting global pressure to curb emissions and the increasing importance of energy-efficient operations, a promising 81% of IT decision-makers are optimistic about meeting their set net-zero target dates. However, the path towards sustainability is less than smooth. Only 19% of organisations have already achieved net-zero, while less than half (44%) are in the position where they are not net-zero but plan to reach this milestone in the future.

These insights are the findings of a new report titled 'Vision 2030: Your guide to addressing sustainability in digital infrastructure', from Telehouse International. 250 UK IT decision-makers were surveyed to understand their perspectives on digital infrastructure challenges, opportunities, and their interplay with sustainability goals for the next decade.

Economic and Technological Barriers to Sustainability

A closer inspection of the hurdles reveals cost as a predominant concern for 60% of IT decision-makers who aren't confident their organisation will meet net-zero goals by their target year. In tandem, half of these professionals (50%) highlight the lack of requisite technological infrastructure as a primary impediment, while 30% face resistance from stakeholders in adopting sustainable practices.

The Quest for Renewable Energy Sources

A distinct challenge surfaces around Scope 2 emissions targets, related to indirect greenhouse emissions from consumed electricity, heat, or steam. An overwhelming 92% of organisations surveyed focussing on Scope 2 emissions report they are grappling to identify energy sources that align with renewable standards and regulatory compliances.

Infrastructure Choices Shaped by Sustainability

The research further delves into the IT infrastructure procurement process, uncovering that a significant 84% of organisations perceive sustainability as an important consideration.

43% of respondents actively seek out providers and solutions with strong environmental credentials, while 23% only consider providers/solutions with proven sustainability credentials and transparent reporting on their environmental impact.

Mark Pestridge, Executive Vice President & General Manager of Telehouse Europe, said: "Organisations are committed to creating more sustainable digital infrastructure despite facing challenges in reducing and demonstrating their environmental impact. Leveraging the right digital infrastructure, nurturing sustainability initiatives, and fostering diversity are pivotal to advancing these objectives more swiftly."

Keysource launches 'State of the Industry' report

Keysource has launched its sixth annual State of the Industry Report which gathers the views and insights from over 250 IT Directors and Senior Datacentre professionals.

THE REPORT highlights the trilemma the industry is facing in dealing with the competing and compelling challenges of developing and delivering on sustainability targets; the pressure to speed up project delivery to remain competitive; and the continuing supply chain and skills issues. This is against a background of rising costs and new EU and imminent UK Regulation changes. Jon Healy, COO at Keysource, comments: "The pressure to speed up project delivery is perhaps the most concerning finding of the report with 75% of those surveyed identifying quality issues which could reasonably have been identified or better managed earlier as a result. Certainly, we are seeing some organisations prioritising speed above all else which is at best risky, especially considering our respondents' strong concerns about getting the correct advice as the skills shortage continues to bite."

The report shows that the skills shortage continues with competing demand both 'client side' and within the supply chain for the same people. This is reflected by the fact that only a third of respondents are confident in the quality of the information that is being provided which negatively impacts the ability to make informed decisions.

As a result, nearly half of respondents chose to sub-contract more projects or services than they had planned, as the industry turns even more to supply chain partners to keep to programme timescales. According to the majority of respondents this approach had a positive impact including better quality and quicker delivery, with the inevitable trade off of a higher cost.

There are some encouraging findings around sustainability with 69% of respondents having a seat at the table when discussing sustainability targets and over half having a separate 'green budget' that can be used for

sustainable solutions and initiatives. However, this positive progress is at odds with the just 17% who consider sustainability to be a high priority and the fact that less than a third said they were making significant progress with their sustainability strategy – with over half still not having one at all! In addition, 64% of respondents haven't evaluated the carbon impact of existing data centre services and solutions and 57% aren't intending to evaluate future investments, meaning missed opportunities to make both carbon and financial savings.

Jon Healy concludes: "The inrush of available capital that we have seen enter the data centre market is reflective of its promising returns and the comparative performance of other markets.

This is coupled with a relatively low risk given its resilience through recent years, which is overall very positive for the short and medium term of the industry.

This said, regional conditions can quickly change given the influences this sector has from a range of areas such as technology, regulation, energy resources, corporate governance, and a lack of skilled people.

"So, whilst this year's state of the industry report shows that the data centre and related sectors continue to grow despite these challenges and the current global unrest, political scepticism, and economic uncertainty, it also flags that a number of common challenges still remain and these are forcing decision makers to operate differently. Our industry has a clear trilemma and the need to solve all three are equally important!"

As a result, nearly half of respondents chose to sub-contract more projects or services than they had planned, as the industry turns even more to supply chain partners to keep to programme timescales

Challenging decade for digital infrastructure

Telehouse research finds 95% of IT decision-makers believe digital infrastructure is a business risk.

IN THE FACE of rapidly increasing data volumes, relentless cyber threats, a critical shortage of skills, and the intricate demands of regulatory compliance, a startling 95% of businesses perceive their digital infrastructure as a risk to their operations.

The findings form part of a new report titled 'Vision 2030: Overcoming your digital infrastructure connectivity challenges and requirements', from Telehouse International Corporation of Europe, a leading global data centre service provider. 250 UK IT decision-makers were surveyed to gauge their opinions on the digital infrastructure challenges and opportunities they are likely to face over the coming decade.

Data Deluge and Infrastructure Needs
Looking ahead to 2030, a resounding nine-in-ten (89%) of respondents anticipate the need for high-density, high-performance computer systems to harness the massive volumes of data generated by the IoT, widespread AI adoption, machine learning, advanced data analytics, and the expansion of cloud-based remote work. Moreover, 75% of organisations expect their data management responsibilities to increase significantly.

Data Vulnerability Looms Large
Highlighting the growing apprehension among IT professionals, a significant 42% of respondents pinpointed software as the most vulnerable aspect of their digital infrastructure over the next decade. This finding underscores the urgency for organisations to bolster their software defences.

Preparedness Gap: The IoT and Edge Computing Challenge
The survey exposed a readiness gap, with more than half (55%) of respondents acknowledging their partial readiness to grapple with the challenges posed by emerging technologies like the Internet of Things (IoT) and edge computing. 11% of organisations admitted to having limited capabilities, a shortcoming that could significantly hamper their growth and competitiveness if not promptly addressed.

Skills Shortage Persists
The scarcity of IT skills remains a pressing issue, particularly regarding emerging technologies. Nearly a third (29%) of decision-makers identified artificial intelligence (AI) as the area where their organisation faces the most significant skills deficit. Meanwhile, 20% identified a shortage of cloud-related skills, and 14% expressed concerns about security expertise. To bridge this skills gap, 35% of organisations have initiated internal training programs focusing on AI, edge computing, and cyber security.

AI Integration and Security Concerns
The research also highlighted a shifting landscape of challenges for IT decision-makers. Since Telehouse's 2020 research, the percentage of senior IT professionals foreseeing the integration of AI and analytics as their most significant infrastructure challenge has grown from 23% to 33%.

Notably, cyber security emerged as a prominent source of anxiety, with 33% expressing that cyber-attacks are their top concern regarding downtime. This anxiety is exacerbated by the evolving landscape of regulations and compliance standards.

Colocation on the Rise
As the complexities of security and compliance intensify, 54% of organisations are opting for colocation services over on-premises IT infrastructure. This represents a notable increase from the 33% reported in Telehouse's 2020 research.

Investment and Partnerships: The Path Forward
Amid these challenges, 61% of respondents plan to increase their investment in data centre infrastructure over the next decade, while 33% expect investment levels to remain steady. Additionally, one-fifth of organisations are exploring commercial relationships with colocation providers to bolster their connectivity capabilities in the face of the impending data deluge.

Mark Pestridge, Executive Vice President & General Manager of Telehouse Europe commented, "Our research underscores the formidable digital infrastructure challenges that organisations are grappling with as they navigate emerging technologies amidst a persistent shortage of specialised IT skills and mounting concerns about cyber risks, downtime, and regulatory compliance. The exponential growth of data and the growing demand for digital connectivity make it imperative for businesses to adopt robust models and establish the right partnerships to harness connectivity opportunities."

organisations are exploring commercial relationships with colocation providers to bolster their connectivity capabilities in the face of the impending data deluge

KAO DATA

Inspired by hyperscale. Built for industrial scale.

Award winning multi-site colocation.

Kao Data develop and operate advanced data centres for high performance computing. With a new 10MW facility operational in Harlow, and an industrial scale, 40MW data centre planned for Greater Manchester, we provide enterprise, cloud, HPC and AI customers with a world-class home for their compute.

info@kaodata.com ▪ kaodata.com

NVIDIA PREFERRED PARTNER

Gartner Identifies the Top 10 strategic technology trends for 2024

Gartner has published its list of 10 top strategic technology trends that organizations need to explore in 2024

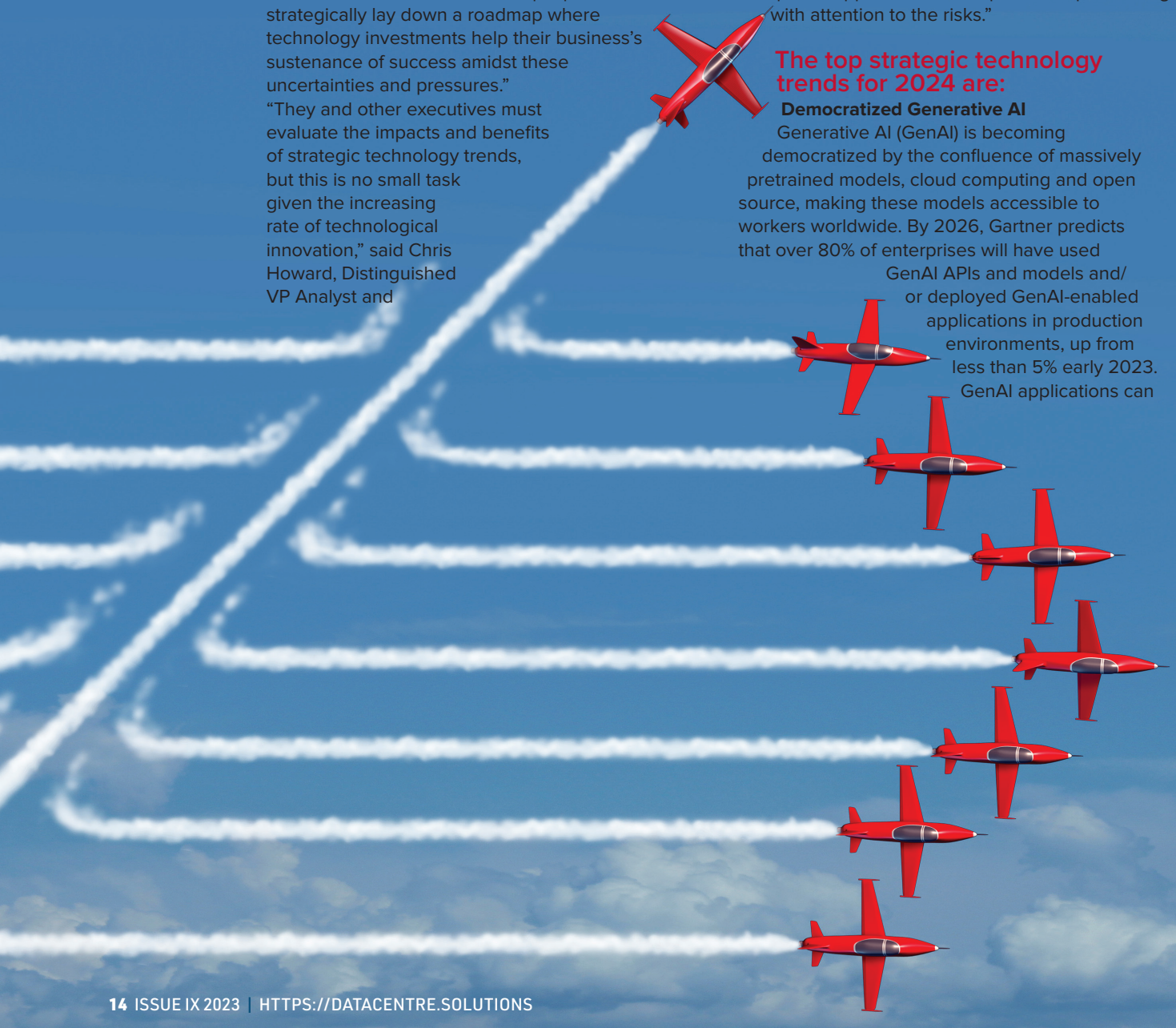
"TECHNOLOGY DISRUPTIONS and socioeconomic uncertainties require willingness to act boldly and strategically enhance resilience over ad hoc responses," said Bart Willemsen, VP Analyst at Gartner. "IT leaders are in a unique position to strategically lay down a roadmap where technology investments help their business's sustenance of success amidst these uncertainties and pressures."

"They and other executives must evaluate the impacts and benefits of strategic technology trends, but this is no small task given the increasing rate of technological innovation," said Chris Howard, Distinguished VP Analyst and Chief of Research at Gartner. "For example, generative and other types of AI offer new opportunities and drive several trends. But deriving business value from the durable use of AI requires a disciplined approach to widespread adoption along with attention to the risks."

The top strategic technology trends for 2024 are:

Democratized Generative AI

Generative AI (GenAI) is becoming democratized by the confluence of massively pretrained models, cloud computing and open source, making these models accessible to workers worldwide. By 2026, Gartner predicts that over 80% of enterprises will have used GenAI APIs and models and/or deployed GenAI-enabled applications in production environments, up from less than 5% early 2023. GenAI applications can



make vast sources of information — internal and external — accessible and available to business users. This means the rapid adoption of GenAI will significantly democratize knowledge and skills in the enterprise. Large language models enable enterprises to connect their workers with knowledge in a conversational style with rich semantic understanding.

AI Trust, Risk and Security Management
The democratization of access to AI has made the need for AI Trust, Risk and Security Management (TRiSM) even more urgent and clear. Without guardrails, AI models can rapidly generate compounding negative effects that spin out of control, overshadowing any positive performance and societal gains that AI enables. AI TRiSM provides tooling for ModelOps, proactive data protection, AI-specific security, model monitoring (including monitoring for data drift, model drift, and/or unintended outcomes) and risk controls for inputs and outputs to third-party models and applications. Gartner predicts that by 2026, enterprises that apply AI TRiSM controls will increase the accuracy of their decision making by eliminating up to 80% of faulty and illegitimate information.

AI-Augmented Development
AI-augmented development is the use of AI technologies, such as GenAI and machine learning, to aid software engineers in designing, coding and testing applications. AI-assisted software engineering improves developer productivity and enables development teams to address the increasing demand for software to run the business. These AI-infused development tools allow software engineers to spend less time writing code, so they can spend more time on more strategic activities such as the design and composition of compelling business applications.

Intelligent Applications
Intelligent applications include intelligence — which Gartner defines as learned adaptation to respond appropriately and autonomously — as a capability. This intelligence can be utilized in many use cases to better augment or automate work. As a foundational capability, intelligence in applications comprises various AI-based services, such as machine learning, vector stores and connected data. Consequently, intelligent applications deliver experiences that dynamically adapt to the user.
A clear need and demand for intelligent applications exists. Twenty-six percent of CEOs in the 2023 Gartner CEO and Senior Business Executive Survey cited the talent shortage as the most damaging risk for their organization. Attracting and retaining talent is CEOs' top workforce priority, while AI was named the technology that will most significantly impact their industries over the next three years.

Augmented-Connected Workforce
The augmented-connected workforce (ACWF) is a strategy for optimizing the value derived from human workers. The need to accelerate and scale talent is driving the ACWF trend. The ACWF uses intelligent applications and workforce analytics to provide everyday context and guidance to support the workforce's experience, well-being, and ability to develop its own skills. At the same time, the ACWF drives business results and positive impact for key stakeholders.

Through 2027, 25% of CIOs will use augmented-connected workforce initiatives to reduce time to competency by 50% for key roles.

Continuous Threat Exposure Management
Continuous threat exposure management (CTEM) is a pragmatic and systemic approach that allows organizations to evaluate the accessibility, exposure and exploitability of an enterprise's digital and physical assets continually and consistently. Aligning CTEM assessment and remediation scopes with threat vectors or business projects, rather than an infrastructure component, surfaces not only the vulnerabilities, but also unpatchable threats.
By 2026, Gartner predicts that organizations prioritizing their security investments based on a CTEM program will realize a two-thirds reduction in breaches.

Machine Customers
Machine customers (also called ‹custobots›) are nonhuman economic actors that can autonomously negotiate and purchase goods and services in exchange for payment. By 2028, 15 billion connected products will exist with the potential to behave as customers, with billions more to follow in the coming years. This growth trend will be the source of trillions of dollars in revenues by 2030 and eventually become more significant than the arrival of digital commerce. Strategic considerations should include opportunities to either facilitate these algorithms and devices, or even create new custobots.

Sustainable Technology
Sustainable technology is a framework of digital solutions used to enable environmental, social and governance (ESG) outcomes that support long-term ecological balance and human rights. The use of technologies such as AI, cryptocurrency, the Internet of Things and cloud computing is driving concern about the related energy consumption and environmental impacts. This makes it more critical to ensure that the use of IT becomes more efficient, circular and sustainable. In fact, Gartner predicts that by 2027, 25% of CIOs will see their personal compensation linked to their sustainable technology impact.

Platform Engineering
Platform engineering is the discipline of building and operating self-service internal development platforms. Each platform is a layer, created and maintained by a dedicated product team, designed to support the needs of its users by interfacing

with tools and processes. The goal of platform engineering is to optimize productivity, the user experience and accelerate delivery of business value.

Industry Cloud Platforms
By 2027, Gartner predicts more than 70% of enterprises will use industry cloud platforms (ICPs) to accelerate their business initiatives, up from less than 15% in 2023. ICPs address industry-relevant business outcomes by combining underlying SaaS, PaaS and IaaS services into a whole product offering with composable capabilities. These typically include an industry data fabric, a library of packaged business capabilities, composition tools and other platform innovations. ICPs are tailored cloud proposals specific to an industry and can further be tailored to an organization's needs.

Gartner unveils top predictions for IT organizations and users

Gartner has revealed its top strategic predictions for 2024 and beyond. Gartner's top predictions explore how generative AI (GenAI) has changed executive leaders' way of thinking on every subject and how to create a more flexible and adaptable organization that is better prepared for the future.

"GenAI presents an opportunity to accomplish things never before possible in the scope of human existence," said Daryl Plummer, Distinguished VP Analyst at Gartner. "CIOs and executive leaders will embrace the risks of using GenAI so they can reap the unprecedented benefits.

"This is the first full year with GenAI at the heart of every strategic decision, and every other technology-driven innovation has been pushed out of the spotlight," added Leigh McMullen, Distinguished VP Analyst at Gartner. "GenAI has broken the mold and has kept building more excitement."

Gartner analysts presented the top 10 strategic predictions during Gartner IT Symposium/Xpo, taking place here through Thursday.

By 2027, the productivity value of AI will be recognized as a primary economic indicator of national power
National governments have a strong commitment to AI and are prioritizing strategies and plans that recognize AI as a key technology in both private and public sectors. Incorporating AI into long-term national planning is being reinforced through the implementation of corresponding acts and regulations to bolster AI initiatives.

"Implementation at a national level will solidify AI as a catalyst for enhancing productivity to boost the digital economy," said Plummer. "Successful implementation of large-scale AI initiatives necessitates the support and collaboration of diverse stakeholders, showcasing the mobilization and convening ability of national resources."

By 2027, GenAI tools will be used to explain legacy business applications and create appropriate replacements, reducing modernization costs by 70%
"The maturity of large language models (LLMs) offers an opportunity for CIOs to find credible and long-awaited mechanism for modernizing legacy business applications in a cost-effective manner," said Plummer. "CIOs can create dedicated testing units to test the output generated by GenAI LLMs, while establishing change management and upskilling processes to enable the workforce to maximize productivity throughout the modernization cycle."

By 2028, enterprise spend on battling malinformation will surpass $30 billion, cannibalizing 10% of marketing and cybersecurity budgets to combat a multifront threat
The most effective malinformation influences humans' and machines' decision-making mechanisms and can be extremely hard to detect and shut down. Malinformation presents threats across three disparate functional areas: cybersecurity, marketing and AI.

"The rapid rise of GenAI has put fire under the feet of regulators about including malinformation as one of the risks associated with the increasing power and availability of GenAI to bad actors," said Plummer. "Enterprises who maintain a close watch on bad actors, regulators and providers of tools and technology that help combat malinformation are likely to gain significant advantage over competitors."

By 2027, 45% of chief information security officers (CISOs) will expand their remit beyond cybersecurity, due to increasing regulatory pressure and attack surface expansion
Responsibilities for security management and digital assets are fragmented across multiple divisions and teams, with the CISO overseeing the overall digital asset portfolio. This creates inconsistencies in support for regulatory disclosures, assurance of digital security and effective management of security incidents, reducing the overall performance of the organization.
Expanding the portfolio of the CISO will enable a unification of security management, providing oversight of the consolidated security incident management process throughout the organization.

By 2028, the rate of unionization among knowledge workers will increase by 1,000%, motivated by the adoption of GenAI
Executives are quick to call out AI as a cause of positions being eliminated. Therefore, it is important for executive leaders to communicate clearly with their employees their intent for internal AI deployments. This will avoid the unintended consequences of AI anxiety building among staff. Organizations that adopt GenAI and fail to clearly

address AI anxiety amongst their knowledge workers will experience 20% higher rates of turnover.

"Organizations should focus their AI efforts on worker augmentation to improve productivity and quality of work, rather than role automation," said Plummer. "Stay grounded in what the technology can and cannot deliver, because there remains a substantial amount of hype influencing board expectations."

In 2026, 30% of workers will leverage digital charisma filters to achieve previously unattainable advances in their career

A digital charisma filter prompts and sifts communications to make them more socially effective in various situations. They nudge in the moment of and before and after interactions to make leaders and co-workers more effective in the social circumstances where they wish to excel. Digital charisma filters will improve organizations' abilities to expand hiring to include more diverse workers.

"Organizations can expand their talent pool by incorporating the use of digital charisma filter assistants to improve the congruency of interactions at all phases of recruiting and employment," said Plummer. "Accelerate access to digital charisma assistants by pressing enterprise productivity and application vendors on how they are incorporating these capabilities into their roadmaps."

By 2027, 25% of Fortune 500 companies will actively recruit neurodivergent talent across conditions like autism, ADHD and dyslexia to improve business performance

"Organizations that hire and retain neurodivergent talent will experience increased employee engagement, productivity and innovation across the workforce," said Plummer.

Fortune 500 companies are already investing in neurodiversity hiring programs and are seeing impacts on engagement and business outcomes. Organizations need to establish an outreach program to boost the discoverability of neurodiverse talent. Fast-track efforts by leveraging best practices from experts and lessons from leading organizations already working on neurodiversity.

"Include neurodivergent people in company leadership positions," said Plummer. "Having openly neurodivergent leadership fosters a culture of inclusion and can be the most valuable action to take from the perspective of neurodivergent employees."

Through 2026, 30% of large companies will have a dedicated business unit or sales channels to access fast-growing machine customer markets

Machine customers will force a reshaping of key functions such as supply chain, sales, marketing, customer service, digital commerce and customer experience. In fact, by 2025, more than 25% of sales and service centers in large organizations will be fielding calls from machine customers.

"Organizations can expand their talent pool by incorporating the use of digital charisma filter assistants to improve the congruency of interactions at all phases of recruiting and employment,"

"Machine customers will need their own sales and service channels because they make transactions at high speeds and the volume of decision variables they use far exceed human capabilities," said Plummer. "Machine customers will require different talent, skills and processes that may not exist in a human-customer focused division."

By 2028, there will be more smart robots than frontline workers in manufacturing, retail and logistics due to labor shortages

Most manufacturing, retail and logistics companies cannot find or retain enough people to support their day-to-day operations. This will cause supply chain organizations to struggle to find enough front-line workers over the next decade. Robots will help fill this gap. A December 2022 Gartner survey found that 96% of supply chain technology workers have either deployed or plan to deploy cyber-physical automation and 35% have already deployed robots, with 61% piloting or in the middle of their first implementation.

"Robotic technology is advancing rapidly, making robots viable for a growing number of front-line jobs from the factory floor to the warehouse to the retail store and beyond," said Plummer.

By 2026, 50% of G20 members will experience monthly electricity rationing, turning energy-aware operations into either a competitive advantage or a major failure risk

Aging grid infrastructures are limiting the ability to add electricity generating capacity, yet demand for electricity continues to increase. Enterprises are assessing energy price and accessibility as a competitiveness, which means stable access to electricity for customers will become a competitive advantage. Because of this, executive leaders are creating energy-aware operations through optimization and direct investment in energy generation.

"Leverage energy efficiency to establish long-term competitive advantage by structurally reducing energy consumptions," said Plummer. "Assess enterprise investment by including current and future anticipated costs of energy."

Is the data centre industry ready for AI?

STEVEN CARLINI, VP FOR INNOVATION AND DATA CENTRE AT SCHNEIDER ELECTRIC, explains how the massive, continuing AI explosion is threatening to re-write the rules when it comes to data centre design and operation. The company recently published a Blueprint for AI data centres, which explores pioneering approaches to data centre design, and Steve shares the four key AI attributes and trends that will underpin physical infrastructure requirements in the era of machine learning.

THE MASSIVE, ongoing AI explosion promises to have a major impact on the data centre. But does this impact require some slight adjustments, so business pretty much as normal, only much more of it, or will it potentially mean the rulebook needs to be thrown out and data centre design will need to be 'reinvented'?

Steve responds: "There is an ongoing, insatiable capacity for data centres. But these new data centres, especially the AI training data centres, are much, much different than the data centres that we have today - the X86 type of server farms that are massive in scale, but which tend to be lower density and spread out more, have a larger footprint. The new servers with the GPU accelerators, or ASICs, are much different. They require a lot more power. All of the processors run in parallel, so you package these in servers, but they're not acting as an individual server, they're acting as one giant server. So, you could have 10,000 GPU accelerators operating in parallel in these data centres. So it is, from a training perspective, a need for an incredible amount of capacity because of all the parameters that we're putting into these to train the models and also the different formats of the parameters."

He adds: "At the beginning it was just text, but now you're feeding them with images and even videos, and they're outputting different formats as well. So multimodal AI is here and it's much different from a data centre perspective, from a capacity standpoint - very high capacity, the need for a lot of data centres, so very high scale and also very high density."

The obvious place to start to meet this AI demand would be to upgrade existing legacy data centres, but is this economically viable and/or practical? Steve says that it is possible to retrofit existing sites, but if you don't have a lot of reserve power or excess power feeding that data centre, they're going to, number one, have to source more power, and number two, have to densify that power. In other words, running all of the power to a smaller, smaller area. So, when you're thinking of the power train and you're thinking of bringing it down from medium voltage, which can be 10,000 to 39,000 volts, down to the server, it's optimal to reduce the number of transformations that you have. Therefore, it is more economical to start from scratch and then have your powertrain optimised for that layout.

We're not talking about cables anymore, we're talking about a very high power busway, and then distributing that to the servers is also very complicated, so there's a tremendous amount of work involved, and that's just the power. We haven't even started to talk about the cooling. So, a legacy data centre could be adapted for AI workloads, but you could potentially end up with a lot of excess floor space in these data centres. Each hall would have one little corner of AI, high density computing, or cluster AI.

Bearing in mind the problems converting legacy data centres to be AI-capable, and the number of data centre owners and operators who see the opportunity and want to be able to tell customers that they can accept AI workloads, what can end users expect from their data centre provider? In other words, if a provider is saying that their facility

is perfectly fitted for a customer's AI application, what might they actually mean?

Steve comments: "From a physical infrastructure perspective, you see the companies that have been doing it for a little while, a lot of them actually paused construction until they could figure out exactly how to build them at the correct density. But I think you're talking about maybe some of the service providers that are retrofitting some of their sites and building cages or clusters to accommodate these AI servers. And in most cases, and we've been dealing with a lot of these companies, they are designing their cages to be able to bring the appropriate amount of power and the appropriate amount of cooling to those cages. Or if they're building the cluster themselves with the racks, they could build out the entire system with the servers, or build out the power distribution to the racks and the cooling and the piping and the manifolds all the way to the racks as well. And you could put in your own servers. But we're seeing a big move.

There's, like I said, insatiable demand for this capacity. So, they are retrofitting many of the sites and we've been working with them for years on this and it never really was very slow going! Now, all of a sudden it's been like a light switch going on in terms of the demand.

"The other thing I should mention which is interesting is the traditional data centre workloads were variable. So, based on your business and based on the amount of processing and storage that needed to be done, the power levels would go up and down. When you start AI training - and I'm still talking about training models - when you're training a model, a large language model, your servers are on in this parallel design and they're running at capacity or very close to capacity. You have to be very careful to design these. If you design it for 100 racks and you put in 100 racks of IT, it's going to run at 100 the entire time. There's no built-in buffer like the old days!"

Talking of which, with data centres already under pressure to source ever more power just to meet the demands of existing digital transformation, let alone the 'new' AI boom. And, of course, the power needs to be renewable.

There has been a lot of work within the industry to make the data centres as efficient as possible and run the data centres with the most renewable sources as possible. And for every data centre that's being permitted or being constructed after the permit, there has to be an identified source of renewable power for these sites. And an increasing number of countries are getting very stingy with their permitting. It's a huge challenge and it started a couple of years ago and last year was not as bad as a lot of people predicted, with the plan brownouts and planned blackouts. Nevertheless, there is a very real concern.

There are a couple of issues. The renewable capacity that was planned to be built, there were different reasons why it didn't get built. There were some legal issues, there were some capacity issues with offshore wind, with the number of rigs that they had to be able to put them in, and then the whole supply chain. Wind and solar has not favourable for the last couple of years, but there is now a big acceleration in the amount of renewables this year. And there are a lot of government programmes and the data centre industry has been funding a lot of these projects directly or indirectly through PPAs and energy credits and the like.

From a technology perspective, there's going to have to be a bigger focus on energy storage, either on the grid or at these data centre sites. If there's a situation, for example, where you're running primarily on renewables and there's excess renewables, like during the day when the sun is shining, the wind is blowing, you would want to take advantage, instead of curtailing that power, to put energy storage on your site and put all that power into batteries where you could better utilise it or give it back to the utility when they need it. As a result, the whole grid is becoming much more distributed and much more complicated, and there's going to be new and innovative ways to manage the grid.

Keeping cool

With the increased power demand for the increased compute loads comes whole new levels of heat being generated. Which begs the question as to whether or not liquid cooling will have an increased role to play. Does the whole industry have to move in this direction or does air cooling still have a significant role to play? And what's the best AI cooling?

Steve explains: "If you look at the biggest models that have been trained, that are in production today, they were all trained primarily on air cooled servers. So, it is possible to use air cooled servers, but from a sustainability and efficiency perspective, it's not optimal. And from a cost perspective, it's not optimal

because of the more you spread out the load, the more expensive it gets, the more real estate that you take up. The current designs, the designs that are coming up, are going to be more densified, so the advantages are going to be they are going to be more sustainable and lower cost, so you can do it with air. But it's a funny looking data centre where you have a big IT rack and one or two servers in each one and it's all spread out. And in terms of the Infiniband and the fibre, you can't run 20 gig or 40 gig connectivity, you're going to be running 400 or 800, which is much more expensive because like I said, these are all running in parallel, so the connectivity between all of them has got to be lightning fast."

When you deploy in the field, the AI models don't have to be connected back to the core data centre, because once the model is trained, you're going to deploy a subset or a compressed version of that model in the field as an inference model

One of the potential pluses of moving to the liquid cooling option is the opportunity for waste heat reuse. Which begs the questions, if someone is building a new data centre with AI loads in mind, will they try and build it in a location where the heat transfer is easy to achieve? And/or, will government at whatever level demand such waste heat reuse? Waste heat reuse is seen a lot more in Europe than in the US and in parts of Asia, whether through ither voluntary or mandatory standards, like Germany just initiated a standard that 20% of the weight of the heat had to be reused. Everyone talks about district heating, which is heating of office buildings or homes, and because of the AI training models running, 100% of the time, the heat is more constant. In one example, Schneider Electric built a data centre where it was next to a plant that did pellets, the pellets for pellet stoves. These were made from either wood chips or excess newspapers, back when we had newspapers, but they would dry out these things and compress them into pellets, using data centre waste heat. Those types of applications are ideal because you don't really need a consistent heat all the time. At the same level, it could vary a little bit and not have a dramatic effect. Overall, there is more and more planning of how to position the data centres next to areas where the heat could be reused.

Data centre location is also important in terms of taking advantage of favourable climatic conditions (ie naturally cool/cold temperatures) and a plentiful supply of renewable energy sources. Might we end up with AI factories in these favourable geographical locations – the Nordics for example – because it makes more sense financially and sustainably, or is latency still an issue?

Steve outlines: "Everyone's familiar with kind of the cloud computing model, where you have kind of the core centralised data centre, regional data centres, and then edge data centres, and they're all sharing data, they're all sharing control planes. There are latency issues when these models are just churning and they're not really connected and the data doesn't need to be duplicated everywhere. I like to talk about integrating the supply chain and dynamically building based on the information from other IT systems, but when you're using that, it has to be more real time, so you can't have this big lag like when you're writing a book report. You could have those data centres located somewhere else and you could wait five minutes. But if you're running grid management too, or grid automation to supply power, those need to be real time as well. "When you deploy in the field, the AI models don't have to be connected back to the core data centre, because once the model is trained, you're going to deploy a subset or a compressed version of that model in the field as an inference model. And based on how fast you need the response or the decision, and based on how accurate you need the decision and how comprehensive it needs to be, that would dictate the size of the model that you need to deploy. So, you have these giant LLM models and you're going to deploy, in some cases, a model that's only 5% of the size of that in the field, and it doesn't need to be connected to the core data centre at that point either."

Higher densities, high data centre impact

We've covered power and we've covered cooling related to the new AI workloads But there is also the not so small matter of the cabinets and racks. If we're looking at higher densities, is there a need to re-engineer data centre infrastructure solutions to make them more robust?

The HPS servers are much larger, much heavier and deeper. As a result normal racks have already been stressed, and in a lot of cases, there's going to be custom racks that are higher, that can hold a lot more weight. A recent Schneider Electric White Paper includes some of the recommendations for how much weight it's going to need to hold, how deep it's going to need to be, and it's going to put tremendous pressure on existing infrastructure. For example, you either like raised floors or you don't. Thew dense workloads make it really challenging to deploy in raised floor environments. On the software side, there's an incredible amount of software that's looking at the GPUs and the health of the GPUs and the speed, so there's a tremendous amount of focus on the computational side of these AI models. So there needs to be an equal amount of focus on the physical infrastructure supporting these models as they're training and as they're running, especially as organisations start deploying more and more in the field and at the edge. Edge computing has been talked about for years and it may well be that AI is actually going to be one of the catalysts

where we're going to see much faster and larger scale deployment at the edge.

How is the industry responding to the AI challenge?

Steve is confident that the data centre industry is ready to meet the AI demand. He says: " Right now the industry is in a very steep uptick and is responding by adding a lot of capacity and adding the capacity to support these types of loads and some of the service providers are adapting to change and to support these type of loads as well. The stress at the moment is really for the enterprise that wants to run some of their legacy applications in cloud environments. A lot of the colo environments and a lot of the colos are transitioning to support the AI loads and it's really causing an economic supply and demand issue where you may have had your application in your colo for the past five years, and now all of a sudden there's such demand for that space that your next contract negotiation may be a lot different than it was in the past."

AI also present a challenge in terms of sustainability. Many, if not all, of the companies that are deploying AI workloads do have carbon neutral commitments and a lot of them are on the road to net zero. There is a tremendous scale out. Data centres are becoming are more power hungry, but they're doing it in a way where they're offsetting the increase in capacity with different vehicles, whether they're credits or PPAs or building the renewable supply themselves. Data centre operators are investing in a lot of wind farms and solar farms, and also they're offsetting water usage as well. They have to meet their carbon neutral and water neutral supply targets, to be waste neutral. As a result, there is a tremendous amount of focus on building these data centres in a sustainable way. Schneider Electric is projecting that, even though the capacity over the next ten years is going to double, the carbon emissions are going to be relatively flat. Data centres are going to scale and scale very quickly, but the amount of carbon emissions is not going to scale with it.

It is also worth bearing in mind that data centres are being used by virtually all industry sectors to combat their own carbon emissions as well. For example, coordinating transportation, logistics and charging with the most renewable sources and also even the carbon output. A lot of these companies are using data centres and AI to actually reduce their carbon footprint. So, even though there's more data centre capacity being deployed, some of that's being used to fight carbon emissions in other industries as well.

Schneider's role

Schneider Electric itself is focused on data centres of all capacity. It is helping with the largest providers of AI data centre capacity, the internet giants. Helping them with their power systems, helping them with the software management of those power systems and of their cooling control. The company is also involved with the service providers, the

colo providers and the new machine learning as a service providers that are emerging. There's a new category of data centre operators that are just building AI training data centres with the intention of having other companies come to them to train their models. Schneider Electric is putting the solutions together to be able to enable the new types of AI servers to be deployed from a power perspective, from a cooling perspective and from a management perspective. Liquid cooling is definitely desirable from a sustainability perspective, from an efficiency perspective - even though it's been around for a long time, it's quite immature with the number of deployments to date.

In summary, Schneider Electric is helping a lot of companies with the whole idea of reimagining what the data centre is going to look like for these environments is something that we're working on. It is working across the board as a partner and as a solution provider as well, to be able to deliver the power equipment, the cooling commitment, the equipment for these types of applications.

Will AI break the industry?!

A question that has to be asked. If everyone does AI as much as it's perhaps predicted, at some point the sign will go up – data centres are full, no more capacity. Or will the industry always come up with the right answers?

Steve replies: "I think it has the potential to break the industry, but there are certain controls, like the amount of available power, the amount of available renewable power are very limiting. The fact that a lot of countries are requiring a lot more diligence for permitting is another. And then the whole supply chain issue. The equipment providers like Schneider, we can build so much equipment, so it is almost natural kind of regulation to how fast these things are going to be deployed. I was at conference a few weeks ago and there were a couple of speakers that said if we could deliver 20,000 of these data centres tomorrow, we would have customers for them. There is this insatiable demand and the industry is reacting to deliver it as fast as it possibly can. But I don't think it will break the industry. I think it'll be more of a controlled rollout based on the factors that I have talked about."

Cooling the heat

A journey through data centre cooling systems.

BY AASHI MISHRA, SR. CONTENT WRITER, RESEARCH NESTER.

DATA CENTERS play a vital role in the digital era by facilitating the storage, processing and distribution of vast amounts of information. As our world becomes more interconnected the demand for data centers keeps rising, which calls for efficient and sustainable cooling methods. It has been noticed that 42 % of a data centers energy consumption is devoted to cooling. With the growth in data storage and processing needs it is essential to find efficient and effective cooling solutions. Inadequate cooling can lead to downtime, reduced equipment lifespan and increased energy expenses. That's why staying updated on the advancements, in data center cooling is vital to remain ahead of the curve.

Why Data Center Cooling is Important

Data centers are the backbone of modern computing landscape driving various aspects such, as social media platforms and financial institutions. However, these facilities produce an amount of heat, which can result in serious consequences if not effectively controlled. In fact, inadequate cooling can lead to a myriad of problems, including reduced efficiency, frequent instances of system downtime and even potential equipment malfunctions. Experts in the industry have found that even a single hour of downtime can cause a sized data center to lose up to $100,000. Moreover, with an estimated cost of $9,000 per minute for data center downtime each year it becomes evident that any disruption, to operations can result in financial consequences. This is why the cooling systems and technologies employed in data centers play such a crucial role. They ensure smooth and efficient operation while minimizing the chances of downtime and maximizing uptime.

Types of Data Center Cooling Systems

When it comes to data center cooling systems, there are generally three types; air based systems, liquid based systems and hybrid systems. Each type has its set of pros and cons that need to be carefully considered during the design phase of a data center cooling system.

Air based cooling systems are the widely used type of cooling system in data centers. These systems utilize air as the cooling medium and involve the use of computer room air conditioning (CRAC) units. Air based systems are known for their simplicity and cost effectiveness in terms of installation and maintenance. However, they may not be ideal, for high density data centers that demand efficient cooling solutions. On the side, liquid based systems

utilize water or other liquids as their main cooling medium. These systems prove to be more effective than air based ones and're particularly suitable for data centers, with high density requirements. However, it's worth noting that they do come with added complexity and higher installation and maintenance costs.

Hybrid systems bring together air based and liquid based cooling technologies creating a balance. They offer cooling solutions for high density data centers while also being cost effective and straightforward to maintain. It's no wonder that hybrid systems are gaining popularity in modern data centers.

Market Overview

According to a report from Research Nester, the Data Center Cooling Market is projected to reach a value of USD 220 billion by 2035 with an estimated compound growth rate (CAGR) of around 15% between 2023 and 2035. In 2022 the data center cooling industry was valued at USD 16 billion due to the growing demand for data centers across sectors. The global expenditure on IT data centers is anticipated to reach USD 222 billion in 2023. Currently there are 8,000 data centers worldwide with around one third of them located in the United States.

Key players in the data center cooling market include Schneider Electric, Vertiv Co., STULZ GmbH, Rittal GmbH & Co. KG, and Airedale International Air Conditioning Ltd. These companies offer a range of cooling solutions, from traditional air conditioning units to more innovative liquid cooling systems. Some key players and their recent developments in market are:

- In April, 2022 , Schneider Electric, a company specializing in the digital transformation of energy management and automation has recently unveiled its latest lineup of Uniflair Chillers. These advanced chillers, equipped with inverter screw compressors are specifically designed to cater to the cooling needs of large scale data centers. With their efficiency precise temperature control and adaptable configurations these chillers are well equipped to tackle both present and future challenges, in data center cooling.
- In May 2022, Alibaba Cloud, a technology and innovation firm has recently unveiled its third data center in Germany. This state of the art facility offers a suite of cloud computing solutions, including storage, networking and databases. Notably the data center incorporates an approach to cooling by utilizing dry coolers and leveraging naturally cool ambient air instead of relying solely on mechanical refrigeration. This friendly method allows for over 7,000 hours of free cooling, per year.

Market Growth Drivers

- **Explosive Growth in Data Usage:** The continuous growth of data usage fueled by cloud computing, Internet of Things (IoT) devices, streaming videos and big data analytics plays a role, in driving the Data Center Cooling market. Between 2000 and 2023 internet usage has seen an increase of 1300%. As we reach 2023, Asia remains the region with the highest number of internet users accounting for 52% of the global population. In order to keep up with the escalating demand, data centers need to expand their capacities and invest in cooling solutions.
- **Energy Efficiency Regulations:** Governments and environmental organizations around the world

are implementing measures to control energy usage and decrease carbon emissions. As an example the European Union has made a commitment to enhance energy efficiency with a goal of achieving a 20 % improvement in 2020 and at 32.5%, by 2030. Data centers, known for their energy consumption are facing increasing demands to adopt cooling technologies that are more energy efficient in order to comply with these regulations.

- **Rising Awareness of Sustainability:** Nowadays a lot of companies are placing an emphasis on sustainability and environmentally friendly practices. It's interesting to note that around 62% of businesses have implemented a sustainability strategy. Additionally, 68% of these companies have made the shift towards using more sustainable materials like recycled products and lower emission alternatives. In line with this trend data center operator are also joining the movement by adopting eco cooling solutions such, as free cooling, liquid cooling and renewable energy sources.

Cooling Technologies

Direct expansion cooling is a widely used option for data centers because of its efficiency and simplicity. This system works by pumping refrigerant through a compressor and evaporator coils, which absorb heat from the air and cool it before sending it into the data center. Although this method is effective it can be costly to maintain, may not be ideal for larger data centers.

Another common technique's chilled water cooling, where water circulates through pipes and gets cooled by a chiller unit before being distributed to cooling units across the data center. Compared to expansion cooling this method is more energy efficient and can be scaled up for larger facilities. However, it does require an infrastructure of pipes and pumps making maintenance more challenging. A newer technology called evaporative cooling that utilizes water evaporation to cool the air. With this system water is sprayed onto a surface that absorbs heat from the air and then evaporates the water resulting in air that circulates throughout the data center. This approach offers efficiency at an affordable cost but may not be suitable for areas, with high humidity levels.

Cooling System Design Considerations

When creating a cooling system for a data center there are crucial factors to keep in mind. One of the important considerations is capacity planning. The cooling system needs to be capable of handling the heat generated by the IT equipment in the data center, which can vary depending on the type and density of the equipment. It's also essential to plan for growth and expansion so that the cooling system can accommodate additional equipment without becoming overwhelmed.

Another significant aspect to consider is redundancy. A data centers cooling system should have components to ensure uninterrupted operation even if one component fails. This includes chillers, pumps and cooling towers. Additionally, it's essential for the design of the cooling system to include paths for air and water flow. This way if one path becomes blocked or compromised the system can still function effectively.

Lastly, energy efficiency plays a role in any data centers cooling system. Cooling often consumes an amount of energy in a data center setup; thus minimizing energy consumption has a substantial impact on operating costs. Best practices for designing an energy cooling system involve utilizing free cooling when viable optimizing airflow throughout the data center space and employing high efficiency equipment, like variable speed drives and electronically commutated (EC) fans.

Future of Data Center Cooling

The prospects for data center cooling in the future are truly thrilling and brimming with possibilities. Ground breaking advancements like immersion cooling and magnetic refrigeration present fresh avenues to cool data centers effectively all the while lowering energy usage. These remarkable innovations not enhance the overall performance of data centers but also play a crucial role, in minimizing their environmental footprint.

New technologies, like edge computing, artificial intelligence and the Internet of Things (IoT) are creating a demand, for cooling solutions that are both localized and efficient. For instance, In June 2022 , Microsoft and Meta have followed Googles lead by incorporating AI technology to manage their data centers. This decision was prompted by the heat produced by their new tensor processing units (TPUs) which surpassed the capabilities of their previous cooling systems. With data centers becoming increasingly interconnected and decentralized there is a rising need for inventive cooling solutions to meet the growing demands of this industry.

OFC

The future of optical networking and communications

Register Today

A Hybrid Conference – In-Person and Virtual Presentations

Technical Conference: 24 – 28 March 2024
Exhibition: 26 – 28 March 2024
San Diego, California, USA

Plenary Speakers

Anita Döhler
Chief Executive Officer, NGMN Alliance

Inder Monga
Executive Director, ESnet, and Scientific Networking Division Director, Berkeley Lab

David Richardson
Partner Researcher, Microsoft

OFCConference.org

IEEE ComSoc IEEE Communications Society
IEEE Photonics Society
OPTICA Formerly OSA

Data center performance measurement that adds up to improved efficiency

Modern data center operations extend beyond merely ensuring IT functions run smoothly. Today, data center operators face the growing challenge of optimizing their facilities for maximum efficiency.

BY MARK FENTON, SENIOR PRODUCT MARKETING MANAGER AT CADENCE

WHILE THE TERM “efficiency” is a recurring theme in the data center industry, its interpretation can vary. Is efficiency about conserving energy use, cost reduction, water conservation, carbon emissions or something else entirely?

Each person you ask may give you a different answer. Still, when speaking to senior management, all these factors are likely to contribute to their definition of efficiency.

By employing physics-based analysis, data center managers can truly understand how to tackle all the different factors that are in play in order to ensure their facility runs as efficiently as possible.

Managing the demands of a data center

Delivering higher processing throughput, whilst meeting the various demands across a facility – such as constrained space and sustainability goals – is a common challenge data center operators face for a number of reasons.

Firstly, in a modern data center, there is often a growing frequency of high-density IT equipment deployment requests that puts huge pressure on the facility systems. This high-density equipment often necessitates additional power and cooling resources that older facilities, particularly legacy ones, aren’t designed to accommodate.

In addition to this challenge, some data centers lack the technology to predict the impact that operational changes can have elsewhere. For example, a manager could deploy high-density IT hardware and be unaware of how this will affect the current cooling and power system provisioning.

This lack of visibility can increase risk that can only be comfortably resolved by over-provisioning once the new equipment is in place. This would have a negative impact on the data center’s sustainability posture as it would mean a rise in the carbon

emissions produced. With the introduction of regulations, such as the Corporate Sustainability Reporting Directive in Europe, which necessitate stringent tracking and reporting of power and cooling use, facility managers can't afford to have their carbon footprints surge unnecessarily.

Faced with balancing new sustainability requirements with demands already in play, how can managers ensure their data centers run at the best performance level possible? One route is through measurement, which can help identify how different demands interact with one another and have the most significant effects.

Scientific data center performance measurement
Measuring data center performance with The Green Grid's Performance Indicator (PI) plus computational fluid dynamics (CFD) simulation is a proven approach to enhance efficiency in a facility and troubleshoot any issues.

The Green Grid's PI is a powerful tool that empowers managers to establish the data centers' ability to meet IT cooling needs and be efficient so they can improve performance. It does this by delivering visibility into how different factors – including energy use, cooling effectiveness, and IT equipment performance – impact each other.

More specifically, the Green Grid's PI can help weigh-up effective cooling during normal operation (IT Thermal Conformance) and adequate cooling during failure or maintenance within design parameters (IT Thermal Resilience), without compromising efficiency (PUE ratio).

Meanwhile, CFD technology simulates cooling systems so data center managers can virtually test operational changes and measure the impacts before they're made in the physical facility. Combined, this approach offers managers a holistic view of how any given change will impact the data centers' performance. This means they can better plan, test, and validate changes, taking into account both risk and efficiency, before any change is made in the real data center.

Improving data center efficiency

Uncovering small improvements that can be made in a data center, can massively enhance a site's overall efficiency and, therefore, the data center's performance. The first step in discovering what these adjustments are is measurement, which quantifiable approaches, like

The Green Grid's PI and CFD simulations, make simpler and more effective. Executed well, performance assessments empower managers to balance the conflicting demands within their facility, ultimately making their job of handling operations simpler and the data center more efficient.

Hydrogen fuel cells and the role they play in data centre design

As data centre operators continually look for ways to decarbonise their business operations, hydrogen has become a highly anticipated solution for storing and releasing low or zero-carbon energy.

BY EUAN CUTTING, ELECTRICAL ENGINEER, BLACK & WHITE ENGINEERING

WHEN CONSIDERING OPTIONS for onsite electrical generators, there are very few practical options which can fit in a data centre campus and produce no carbon emissions. When excluding renewables from our options, it leaves us with gas generators and fuel cells. Of the two remaining options, fuel cells have higher efficiency. The cost-effectiveness depends on how they are used, whether it is for backup or continuous power. In terms of cost, hydrogen at present is prohibitively expensive, not to mention the infrastructure required to store it, and the significant storage difficulties due to its low density. While the cost of hydrogen is predicted to reduce over the next decade, making it more accessible, the issues around large-scale storage may prevent usage on many data centre sites. In the future, there will be locations with gas pipelines which could provide hydrogen, but this may not be an economical option when compared to consuming electricity from the utility grid.

As hydrogen can react with oxygen to release energy with no carbon emissions, it is no surprise that current predictions demonstrate that by 2030, the hydrogen economy could be worth $500 billion. Hydrogen fuel cells combine hydrogen and oxygen to generate low or zero-carbon electricity which can be used to power a data centre. The majority of fuel cells currently in use in data centres are solid oxide fuel cells (SOFCs), providing constant power. SOFCs can generate power through the conversion of fuels such as natural gas and biogas, into hydrogen, which is then reacted in the fuel cell to generate power. While using natural gas still results in carbon emissions, SOFCs are able to generate power with higher efficiency than combustion engines. SOFCs can also be fuelled directly with hydrogen, although this is not the norm due to hydrogen's cost and availability.

While it is capable of providing constant power, hydrogen fuel cells are also being considered for providing backup power to data centres. This is greatly appealing to data centre operators as a more environment-friendly replacement for traditional diesel generators. This change would see the use of fast-start fuel cells, such as proton exchange membrane (PEM) fuel cells which could take the place of diesel generators.

SOFCs and PEM fuel cells differ from one another in their construction, materials, and operation. In a high-level view, the primary differences are the electrolyte materials (where the hydrogen and oxygen react) and operating temperatures. SOFCs operate at high temperatures, requiring longer start-up times and as a result, only being suitable for continuous power supply. PEMs, by contrast, operate at lower temperatures and are capable of fast-start or continuous operation, but are a more expensive option.

How exactly fuel cells will be used in future data centres is still up for debate, with operators considering both backup and constant power options. There are significant trade-offs with each option; using fuel cells for backup power requires large quantities of hydrogen to be stored onsite, which is costly, space consuming, and high-risk, particularly when compared to current onsite diesel storage. Constant power usage, on the other hand, would see the fuel cells connected to a future hydrogen gas pipeline. The major disadvantage

with this option is operating costs – as current data centre operators state that they intend to use zero-carbon "green hydrogen," using renewable electricity to split water via electrolysis. This is almost guaranteed to be more expensive than just using utility electricity directly due to electrolyser efficiencies and losses in hydrogen distribution and storage.

Sustainability – a key driver in the use of hydrogen fuel cells

The key benefit and primary motivation for installing hydrogen fuel cells within a data centre is to reduce carbon emissions. As stated, some fuel cells such as SOFCs can use natural gas - while it is less damaging to the environment than diesel, it still results in significant carbon emissions. This may be attractive for data centre operators due to the cost of natural gas for industrial consumers versus the cost of utility electricity. Given the high efficiency of fuel cells, this allows data centre operators to produce lower cost electricity onsite, with the downside of carbon emissions when using natural gas and added energy system complexity.

Fuel cells powered directly with hydrogen rather than natural gas have the opportunity to be sustainable, provided that the hydrogen does not originate from fossil fuels. Depending on its source, hydrogen has commonly referred to 'colour' classifications to help differentiate how it has been produced and therefore how sustainable it is; hydrogen produced directly from natural gas is referred to as 'grey,' while hydrogen produced from natural gas with a carbon capture, utilization, and storage system (CCUS) is referred to as 'blue.' 'Green' hydrogen, produced via electrolysis with renewable energy, is often hailed as the gold standard. Additionally, there is 'pink' hydrogen - it is also produced via electrolysis like 'green' hydrogen however, it uses nuclear energy rather than renewable energy.

The goal to meet net zero-carbon targets, combined with the ongoing energy crisis, has also created a renewed interest in nuclear power which can provide vast quantities of low carbon electricity. In addition to the low carbon electricity source which can be used for electrolysis, there are proposals to utilise waste heat from high temperature nuclear reactors, significantly reducing the energy required for electrolysis. This potentially results in more efficient and economical hydrogen production.

However, companies are increasingly facing pressure from investors to comply with Environmental, Social, and Governance (ESG) standards which is another reason why hydrogen fuel cells are appealing to data centre operators. This pressure is already pushing companies, including data centre operators, to explore alternative technologies such as fuel cells, which may attract more investment and improve public image.

The practicalities of hydrogen fuel cell usage in data centres

Implementing hydrogen fuel cells in place of diesel generators results in a significant change to the overall energy system of a data centre. As data centre designers, it is our responsibility to design the electrical and mechanical infrastructure to provide constant power and cooling to the servers. From an electrical perspective, fuel cells and diesel generators have fundamentally different properties. To begin with, diesel generators produce alternating current (AC), whereas fuel cells produce direct current (DC). Additionally, they have varying properties when powering up and assuming high loads. This may provide novel options for DC systems within a data centre, although depending on the size of the data centre and the distances that electricity may have to travel, this could become inefficient.

From a mechanical perspective, designing a hydrogen storage system is significantly more complex than a diesel storage system. Hydrogen has more storage options available however, it presents higher risks than diesel such as greater flammability and explosivity, higher pressures, potential for low temperature, or chemical storage methods which are all hazardous. This therefore requires the mechanical design for such a system to comply with rigorous safety standards. Consequently, local planning restrictions may prevent a data centre from storing hydrogen onsite - a particular issue for data centres in Europe which are often located in urban and industrial environments.

There are various ways that hydrogen fuel cells could be implemented within a data centre. As mentioned, there are significant costs associated with adopting this technology, and in many cases, they are not a feasible replacement for existing diesel generators as current data centre sites have not been designed with hydrogen in mind. This is less of an issue in markets such as in United States where data centres have a greater access to available land. In typical European data centre locations, land comes at a premium and many data centres are located in relatively urban environments.

One suggestion is for greater integration of energy systems, which would see data centres located adjacent to energy industries or having data centres integrated with hydrogen generating plants and fuel cells. This solution sidesteps planning problems by locating data centres alongside low carbon energy industries. A major issue with this (aside from the available land) is blurring the lines between the data centre operators, utility providers, and energy companies. While this gives data centre operators direct access to low carbon energy, there will have to be a clear demarcation between data centre operators and utility operators, as the data centre operators' primary business is data rather than energy.

In addition to the above design considerations, cost is another hurdle for hydrogen uptake. While it is hard to predict exactly how costs will vary over the coming decade, especially with recent cost fluctuations due to the energy crisis, there are a range of investments and subsidies being launched. These aim to produce cost-competitive hydrogen at scale. This could take the form of subsidies such as contracts for difference (CfDs), which can help increase investment in low carbon technology. In the last decade, CfDs have played a major role in bringing down the cost of renewables for developers. This has been a success story in the UK, with the grid gaining large quantities of renewable energy.

Design-first greener, cleaner data centres

There are countries in which the local utility grid generates power from highly polluting fossil fuels such as coal. In this scenario, the carbon emissions per kWh of electricity from these girds are higher than the emissions from a fuel cell utilising natural gas. This allows a data centre to utilise SOFCs with a natural gas supply, which reduces or eliminates the data centres' demand on the local utility grid, while also reducing the carbon footprint of the data centre (when compared to a scenario in which the data centre used electricity from a carbon intensive grid). There is also the potential to utilise waste heat in a vapour absorption machine (VAM) chiller, to provide cooling and power for further carbon mitigation. Black & White Engineering has designed data centres that use SOFCs in this capacity, which can provide the dual benefit of reducing carbon emissions and costs, while additionally mitigating demands on local utility grids.

However, a caveat to the sustainability benefits of this kind of onsite power generation is, as a country begins to decarbonise its utility grid, the emissions reduction which the data centre operator previously had will be diminished, until eventually they may generate more carbon emissions per kWh than the electricity available on the utility grid.

While this is highly dependent on the country in question, a good example of this would be the combined heat and power (CHP) plants in the UK. A decade ago, these plants provided "low carbon" electricity in comparison to the grid at the time, but now in many cases, emit more carbon than local grids. Countries which already have decarbonised grids, France, Sweden, and Scotland, for example, will not benefit from a continuous system which uses natural gas to begin with.

An ideal scenario for a data centre operator utilising this system would be for a zero-carbon hydrogen pipeline to become available as the local utility grid becomes decarbonised. However, this also requires the hydrogen supply to be cost-competitive with utility electricity. As more fast-start PEM fuel cells enter the market, there will be greater flexibility on how fuel cells can be used. PEM fuel cells would allow for dynamic changes in power generation, potentially allowing backup power options or the ability to provide services for the utility grid. Again, this would necessitate significant onsite hydrogen storage or a 'green' hydrogen pipeline.

In the coming years, if 'green' hydrogen that is made from renewable energy) can be made as cost-competitive and as available as current fossil fuels, it will offer a sustainable alternative to diesel for data centre operators. While hydrogen technologies offer a promising solution, they are complex technologies which must be carefully implemented to ensure real carbon mitigation, practicality, and cost-effectiveness.

EcoStruxure™
Innovation At Every Level

Realize your company's IT potential

Software to design, monitor, and manage your IT space

62% of IT outages can be attributed to IT infrastructure failure[1]. Our Data Center Infrastructure Management (DCIM) 3.0 offer provides device monitoring, health assessments, and more so you can:

- Run simulated impacts to expose vulnerabilities in IT infrastructure and address them immediately
- Reduce physical threats by monitoring IT environmental conditions
- Improve sustainability efforts by tracking PUE, energy, and carbon emissions

#CertaintyInAConnectedWorld

apc.com/edge

EcoStruxure™ IT modernized DCIM

APC Smart-UPS™ Modular Ultra

[1]Uptime Institute Global Data Center Survey, 2018

Life Is On | Schneider Electric

Magnetic North: Scotland's renewable future as a data centre hub

With green credentials that include the UK's lowest carbon energy intensity, increasing access to RERs, low-cost land and political backing, are the pieces in place for Scotland to fulfil its data centre sector potential?

BY KERR JOHNSTONE, DIRECTOR, I3 SOLUTIONS GROUP

SCOTLAND has been talked about as a location for big and hyperscale data centres for almost two decades. The geography, climate, access to renewables, and improving subsea connectivity, tick many of the development criteria boxes for data centres.

All reinforce the huge potential for the country to attract multiple large scale data centre developments.

Scotland is green. A major positive for Scotland is its increasing access to renewable power generation resources. Scotland's renewable energy capacity reached 13.6GW in September 20221, a rise of 11.7% on the previous year due in most part to more onshore and offshore wind coming on stream.

The electricity used in the south of Scotland – which includes the central belt – has been the greenest (measured as grams of carbon dioxide equivalent produced per kilowatt-hour of electricity generated (gCO_2e/kWh) of anywhere in Great Britain since the turn of the decade. This is according to independent research commissioned by Scotland's leading data centre and multi-cloud services provider DataVita.

The data showed that the south of Scotland had the lowest average gCO_2e/kWh rate of any region within The United Kingdom at just over 47 grams per hour since January 2020.

DataVita says in particular periods an IT workload hosted in Scotland could be 11 times less carbon intense that the worst performing location in the UK and three times less intense than a workload running in London.

On land, under the sea

Scotland's land cost and site quality are attracting interest.

In May 2023 Scottish Futures Trust/Host in Scotland, Crown Estate Scotland and Scottish Enterprise issued an updated Site Selection Report.
In March 2021 Host in Scotland, partnering with Scottish Enterprise (SE) and Crown Estate Scotland (CES), commissioned a report which drew up a shortlist of sites it says are ready for green data centre development projects. From a long list of 80 suitable sites, it named 36 prime locations from the borders to Inverness based on criteria including available power, renewables access, land scale and

connectivity.

The updated 2023 Site Selection report allowed previous sites selected to be checked for their availability and updates on their development status to be included. [Since 2021] new sites for potential data centre development have been sought by contacting local authorities, various government agencies and property agents. This has resulted in an increase of 5 new shortlisted sites representing the best sites for data centre development across the country. Some of these sites are suitable for urban colocation use whilst other sites are considered more suitable for rural hyperscale development.

Scotland's mountainous geography is also a plus. In March 2023 energy giant SSE said it would invest £100m in a pumped hydro energy storage facility in the Scottish Highlands. And going all the way back to 2009 when the Crown Estate licensed its part of the seabed in Pentland firth to tidal power developer MeyGen plc, stories of the potential for 800MW data centres running on clean power have surfaced regularly. Swiss power giant ABB provides the grid connection for MeyGen in a deal announced in 2014.

Diverse fiber connectivity to Scotland is improving. Tampnet Carrier is a Nordic based high-speed network operator responsible for over 30% of the traffic between Norway, the UK and Europe. The company has two routes from Scotland to London. One from its PoP in Edinburgh connects directly to London offering diversity for data transfers.

Another route from Edinburgh utilizes Tampnet's subsea network in the North Sea connecting with the burgeoning data centre market in the Nordics. FARICE-1 is a submarine communications cable connecting Iceland, the Faroe Islands and Scotland. The cable has been in use since January 2004. Within the country the Scottish Government says £1 billion has been invested in such programmes as Digital Scotland Superfast Broadband (DSSB), Reaching 100% (R100) and Scottish 4G Infill (S4GI). These, alongside extensive commercial investment, have greatly extended the reach and capacity of fibre networks across the country.

Time to scale

Scotland's ability to attract hyperscale data centre developments would appear to reside in its renewable power generation capacity and energy storage potential.

As i3 Solutions Group GHG abatement initiative shows, evaluation of clean on-site power generation; Power storage such as battery (chemical), Kinetic, Pumped Hydro, and Gravity; Microgrids (islanded and integrated operation) and new revenue potential of being a grid power supplier are fundamental considerations for large

scale developments. Scotland's initial attractiveness will be viewed through access to thousands of megawatts of renewable energy power generation plus its low carbon intensity grid.

However, while power remains the dominant factor, other advantageous considerations around planning, location, design and operation cannot be ignored.

The planning environment for data centre development expects integration with local economies, Scotland's broader economy and adherence to sustainability goals. This could work in favour of developers.

For example, in 2021 the Scottish Government passed The Heat Networks (Scotland) Act 2021 to accelerate the deployment of district heating in population centres. This could make metro data centre developments attractive. Use cases for data centre heat reuse in rural locations include Scotland's historically large and rapidly growing agriculture and aqua culture sectors. Scotland's skills base is built on a historical engineering prowess in areas such as transport and, oil and gas which are transitioning to leadership in hydrogen and sustainable fuel and energy alternatives.

On the mechanical engineering design front the country's low ambient temperatures and soft water resources provide options for adiabatic cooling and enhanced free cooling opportunities to improve efficiency through low PUE (Power Usage Effectiveness).

Scotland is also sparsely populated. Covering an area of 78,782 square kilometres (30,418 sq mi), its population density is listed at 67.2 people per km2 (174/sq mi). Compare this with England where the population density is over 430 per km2 . That means for site selection there is high quality land available which in turn creates opportunities for modern buildings with highly efficient electrical and mechanical infrastructure design. Digital infrastructure requires long term sustainability. All large-scale data centre developments are seeking a clean energy future but many will also require the combination of benefits that Scotland has in abundance.

Skills shortages – the threat remains

Any industry faced with a shortage of qualified professionals to service the labour requirements of its component enterprises and those that service it, will face difficulties. The shortage of qualified professionals in the data centre industry can reduce its ability to meet the growing demands of enterprise, and could lead to increased operational costs, decreased efficiency and potential security vulnerabilities.

BY DAN SHEEHAN, CONSULTANT AT BCS

THERE ARE several reasons contributing to the shortage of skilled staff in the data centre industry. The exponential growth of data and the increasing reliance on digital infrastructure have outpaced the availability of trained professionals, particularly those that can evolve to adopt new technologies and operations. At the same time, a lack of specialised training programs means there may be a gap in the multidisciplinary skills education required for data centre operations.

In addition, skilled professionals in the data centre industry are in high demand, and often have multiple opportunities available to them, making it a challenge for data centre companies to attract and retain top talent.

It should be noted that these threats are relatively longstanding, rising from the dislocation of the pace of growth of the industry allied to time lags associated with attracting and training skilled labour resource. In recent years, the impact of the Covid pandemic has exacerbated these problems via the imposition of lockdown restrictions on freedom of movement and environments for learning.

With this is mind we asked a number of questions around this topic as part of our annual survey of 3500 senior European datacentre professionals. Their responses showed real and ongoing concerns with 98% believing that the coming year will see a decline in supply of staff, a slight rise on the 96% reporting this in winter 2022, and above the 93%

who reported the same in winter 2020. Further adding to the problem, some 92% believe that this will be accompanied by a rise in demand for staff with these skillsets.

Lying at the heart of the debate around shortages of skilled professionals is the potential impact on the delivery of new stock and subsequent consequences for the end user. The survey findings unequivocally support the notion that these shortages have already resulted in tangible consequences and have directly affected our respondents. When asked about the impacts they had experienced in the past year due to these shortages, the majority of respondents cited multiple factors.

For the last four surveys the most reported impact has been that these skills shortages have led to a greater workload on existing staff. In this survey we have seen a slight decline in the proportion of respondents, with some 85% citing it, down from almost 90% six months ago and pushing it back to second place. Just taking its place at the top, though by a very slim margin, is increased operating and labour costs, which is now reported by 86% of participants, roughly the same proportion as reported last winter.

These shortages have also contributed to the growing popularity of outsourcing options, with approximately 45% of respondents acknowledging it as a factor. However, it's worth noting that this percentage has decreased from the 52% recorded in our previous survey, and the more extreme consequence of skills shortages - lost orders – remains at just 9%, in line with the proportion identified six months ago. The number of respondents who found it problematic to resource existing work this year has remained un-changed, with 43% stating that they had experienced difficulties in meeting deadlines or client objectives. It should be noted that this remains well below the 70% who cited it as factor in Summer 2020, when the effects of the Covid pandemic and subsequent lockdowns were being felt across the world. In addition, just less than a third stated that shortages had led to delays to developing new products/ innovations, down on the 48% recording this in our last survey, whilst the proportion that noted they had ceased offering certain products or services has risen slightly to 22% from 17%.

Conclusion

I was lucky enough to have been taken on by BCS as part of the first year of the apprenticeship scheme in 2019 where I have benefited from a hybrid of academic study and a structured programme including hands on experience, supported by fantastic company mentors. The company has created a learning culture, giving its apprentices the opportunity to express themselves and have their voices heard, whilst ensuring they are involved in every aspect of the BCS business. This includes working across the entire client portfolio supporting and learning from the experienced team and seeing first-hand hyper-scale data centre construction projects from the blueprint stages right through to the delivery. So, whilst I understand that there are many actions that need to be taken by the global industry to address the skills shortage, my experience of an apprenticeship approach has been extremely positive and without it, I doubt that I would be working in this fastmoving and exciting sector.

Lying at the heart of the debate around shortages of skilled professionals is the potential impact on the delivery of new stock and subsequent consequences for the end user.

Best practices for hiring

How should a business set about recruiting for senior positions? And what are the pitfalls to avoid to ensure securing the right person for the job?

BY PETER HANNAFORD, SENIOR PARTNER OF PORTMAN PARTNERS

THE PAST DECADE has seen data center capacity grow to unprecedented levels, particularly in the last few years, due to the surge in internet usage and low-interest-rate financing for major capital projects. The increasing use of artificial intelligence means that an almost unimaginable increase in internet growth is now predicted over the next five years and beyond. The largest internet companies are now engaged in an accelerating race to secure data center capacity in strategic geographies around the world – and are falling over themselves to be the winner in the game. Billions of dollars are being invested each week in growing the global digital infrastructure. The stakes have never been higher. We all know that the cost of poor leadership is staggering for businesses, so the pressure is on for those firms, their boards and their investors to find the best leaders to fill critical roles. It's certainly not an easy task, especially when there could be a large pool of potential candidates to consider.

Assessing talent, however, is much more complicated than simply looking for a candidate that ticks all the typical boxes, and often the best people are overlooked because they don't meet a narrow criteria.

So how should a business set about recruiting for senior positions? And what are the pitfalls to avoid to ensure securing the right person for the job?

Partner with your search firm

If you are working with an external search firm to find the best candidates that you can, (and obviously we'd recommend that approach!) then you should view the firm as a project partner, temporarily extending your own team and working together in full transparency and collaboration to work towards a common goal. There should be no territorial behaviour. All candidates, whether identified by the in-house recruiter or referred by a member of the existing staff should be considered alike on a level playing field. A senior executive at one of the hyperscalers urges both in-house and external search partners to think of the relationship as a market trade. External search partners offer market, competitor and candidate intelligence in trade for the in-house recruiter who offers the currency of hiring manager relationships, project management know how, and key messages from the employer brand. If both trade well, we all walk away with a successful search outcome.

The CV only tells you what people have done, not what they are capable of doing

Ask most people what they would cite as the worst hiring decision ever and they'd probably quote poor John Sculley's appointment at Apple. He'd enjoyed great success with PepsiCo but failed at Apple. Compare him to Jony Ive. He worked for a small design agency in London designing, among other things, toothbrushes, microwave ovens and toilets. He was hired by Apple and went on to design the iPod, iPhone and iPad – and was subsequently knighted. What did Steve Jobs see in Jony that wasn't in his CV? His HR team must have thought he'd gone mad. When it comes to picking a leader, one of the biggest mistakes is assuming that the right person needs to have had a particular job title or have specific academic qualifications. This may be true in certain cases (for example, technical roles where the requisite skills are only acquired with experience). However, a candidate may have all the skills needed to succeed in your leadership role but may simply not have had a chance to put them into practice - yet. Intellect, motivation, values, and behaviours are difficult to change. Experience is the one thing that constantly changes. Look beyond the CV.

Embrace Diversity

Diversity and inclusion are often lumped together as a single matter for consideration. Diversity is who organisations are, but inclusion is what they do to make diversity work. While many organisations have been busy ticking the diversity box by extending

opportunities to traditionally under-represented groups (in the digital infrastructure sector, diversity is often couched in gender terms) the underpinning rationale for this is often missed. While a data center business may cultivate some kudos by hiring more women, for example, getting more women into traditionally male dominated roles is not the end goal. Real value is generated from cognitive diversity – the heterogeneity of knowledge and intellectual perspectives that come from different life experiences and socio-cognitive variables like gender, cultural background, age, nationality, or even occupational expertise.

So, forget the box ticking often associated with surface level DE&I initiatives and get thinking about how systems, processes and culture can support diversity of thinking. And remember:

- Diversity reflects the real world
- Diversity means looking at the same idea in a different light
- Diversity brings disruptive ideas and innovation
- Your clients and customers are diverse.

Seek culture contribution, not culture fit

Research undertaken at Stanford University showed that, in a poll, 100% of participants put culture-fit above skills and intellect when reviewing candidates. And a study of 200 start-ups found those who put culture-fit first were least likely to fail and more likely to IPO. This is understandable as all had the same mindset, the same values and were uniquely motivated. However, the study also found that after going public the businesses grew at a slower rate than those who didn't rate culture-fit so highly. They were thinking as a group and had no diversity of thought. So culture-fit can be advantageous in some scenarios but not in all. Embracing diversity helps but ask yourself what's missing from your culture and look for candidates who can help improve it.

Have a clear, timely interview process

As stated earlier, if the candidate you're presented with is good then he or she will likely be targeted by other firms competing for their services.
Time is of the essence. Six or seven (or more) separate unstructured interviews are not only a burden on the candidate, but they are also ineffective with the same questions often being asked and the candidate having to repeat what they said in an earlier round. If the candidate is excited and motivated at first interview, they can, understandably, be somewhat demotivated by the seventh.

We've had candidates discounted in later stage citing "a tiredness or disinterest" as the reason for rejection. If you need multiple people involved in the decision making, then it is surely better to arrange interviews in a panel format. This is easy to organise using Teams or Zoom, or in the boardroom at later stages.

Better questions

When interviewing the candidate, avoid questions such as "why do you want to join us". Remember that you/we have approached the candidate. They didn't apply for the job. During the interview process you are still trying to get them to want to join.
A better question would be "what do you think you could do for us".

By all means test that they've done their homework; that they know exactly what your firm does and what the current challenges might be.

First impressions are important. What's your gut feeling when you spoke with or met this person for the first time? The way they looked, acted, spoke, and, importantly, listened. Were you impressed? As the saying goes, you only get one chance to make a first impression. What was your first impression? Having said that, don't decide that "this is the one" too soon. Keep the first impression but continue to ask probing questions.

Likewise, don't be unduly influenced by academic qualifications. Remember that Steve Jobs, Mark Zuckerberg, Michael Dell, and Bill Gates had no college degrees. Someone with industry experience is, of course, good but it's not predictive of success. If the person is locked into a fixed mindset and approach, industry experience can be a drawback. Old habits die hard, and you may have a big job to re-train the person to your company's way of doing things.

Make a shortlist

Ideally no more than three. Then make sure you check references.
It's likely that the candidate will still be employed, and you need to respect the confidentiality of the situation. Clearly you should never contact anyone without the candidate's express permission. If you are working with a specialist industry search consultant, they would be able to get discrete feedback from their own network. It's likely that the candidate will be known to them anyhow.

Some form of psychrometric test may be useful at this stage. Psychometric tests are used by employers to assess intelligence, abilities, potential and personality. Portman uses the GC Index. This is not a psychometric test but an organimetric, which measures and describes five different ways in which people are inclined to make an impact and contribution.

Finally

When you think you have found the right person, arrange for them to meet the board or their peers in a social environment. Perhaps a dinner. People act differently when relaxed and undoubtedly this person will be required to attend formal and informal events on behalf of your firm at some time. It also gives you the opportunity for this person to shine in the company of their peers.

Unlocking data centres: sustainable refurbishments

Sustainable refurbishments of existing data centres can provide much-needed space for data centres, minimise the embodied carbon impact, and lower the carbon footprint.

BY STEVE DAVIDSON, TECHNICAL DIRECTOR, PARMARBROOK

THE GROWTH of the digital age has provided a market where both personal and professional sectors rely on data centres to function. As such, the demand for data centre facilities has skyrocketed. The uptime of these data centres is critical, as a power outage can lead to significant losses for businesses and put sensitive data at risk. However, as technology advances and capacity requirements increase, many data centres built in the 80s and early 90s are no longer fit for purpose, with outdated and inefficient buildings, unsuitable floor plans, and inadequate structural integrity – especially for the evolution of high density, power intensive HPC and AI we are witnessing.

Despite these challenges, the potential of existing legacy data centres across the UK remains untapped. Developers are increasingly opting for new-build data centres, but there is significant potential in unlocking the potential of existing, disused facilities through meticulously designed retrofits. This is especially relevant in traditional data centre hubs – like Slough and West London – where land for new builds is restrictive, and existing data centres have often been over-engineered and built in a way that fails to optimise the space or available power, the potential of which can be better realised with a retrofit.

One company that recognises this opportunity is Kao Data, who develops and operates high performance data centres for advanced computing, providing enterprise, cloud, HPC, and AI customers with a world-class home for their compute. Today the company is working with civil and structural engineers Parmarbrook to deliver on unlocking data centres, sustainably, across the UK.

Data centres and sustainable construction

Data centres are known for their high energy consumption, and the construction industry has a responsibility to minimise carbon emissions wherever possible. Demolishing existing data centres that can be sustainably refurbished comes with environmental costs – making a significant contribution to carbon emissions. Reusing and repurposing these structures allows for the preservation of embodied carbon, offsetting the environmental impact of the data centres operations.

“Today the data centre and construction industries have a vital role to play in the climate change challenge, and going forward, we must work collaboratively to deliver the highest possible sustainability standards for our sectors,” says Paul Finch, Chief Technology Officer at Kao Data. “Part of the approach must be to recognise the intrinsic value that legacy infrastructures can offer both to

the businesses and consumers, and to ensure that efficient and sustainable operations are built-in from the ground up."

The challenges

· Lack of available space to construct new data centres with sufficient electrical supply capacity within popular data centre hubs like West London, Manchester and Slough
· Identifying existing stock of under-utilised legacy data centres with available renewable power supply, yet unknown load capacity and structural merit
· Understanding the load capacity of the existing structure to maximise the layout of the data halls within them

- Architectural impact of the fit-out to ensure hot or cold aisles are economised to maximise the use of natural free-air cooling
- Capacity of the existing foundations for increased load allowance
- Existing roof space to accommodate revised design to deliver desired loads and cooling
- Existing plant replacement strategy: viability t to accommodate logistical requirements for plant upgrade
- Structural constraints; floor to ceiling heights, column spacing, accommodation of enhanced openings, riser relocation
- Change in levels across floor plate
- Retention of live environments and floor plates
- Capacity of the existing structure for increased load allowance, however, it is generally the foundations that govern and drive the cost
- Party wall restrictions
- Fire and building regulation changes
- Impact on the surrounding environment and local community

While refurbishing disused data centres is a viable solution, there are several challenges to overcome. These challenges include a lack of disused industrial and available space to construct new data centres, the unknown load capacity and structural integrity of existing vacant legacy facilities, and the need to maintain tenancy in live data centres during refurbishments in sub-let spaces. The capacity of the existing foundations and structure for load increase allowance must be considered, along with party wall restrictions and fire regulation changes.
There's also a key requirement to consider the data centre impact on the environment and the local community, and to ensure that any modernisation programmes consider the needs of local stakeholders nearby.

The solution

The key to unlocking the potential of disused, or partially used, data centres is to forensically establish a thorough understanding of the existing structure. Through a series of early investigations, structural modelling, and interrogation of archive information, we can gain a comprehensive understanding of the structure from the feasibility stage. The process includes archive searches

of existing planning and building control portals, seeking copies of archive drawings from the original design team, and conducting intrusive and non-intrusive investigations to understand the existing structure.

To limit the increase in load on the existing foundations, we aim to restrict the increase to 14%, removing the need for expensive and time-consuming foundation strengthening or underpinning works. Any increase of the design load applied to the existing structural frame would also need a detailed investigation, as the existing structural frame would need to be strengthened to support the increase in load.

Parmarbrook has decades of experience in designing and executing extensive refurbishments in commercial space where tenants are in

> Today the data centre and construction industries have a vital role to play in the climate change challenge, and going forward, we must work collaboratively to deliver the highest possible sustainability standards for our sectors

situ, and boast a portfolio of historic building optimisations. Retrofitting a live environment brings a set of challenges that our team is well-versed in overcoming. Much like with data centres, where uninterrupted power supply is critical, we ensure that refurbishments are carried out with no disruption to tenants, their businesses, and the people whose data the centres are holding and protecting.

"Parmarbrook has a wealth of experience in the cut and carve refurbishments of existing buildings in and around London, quite often working on live projects that remain fully or partially in occupation during the full term of the construction works. It is this experience that lends Parmarbrook to working on the refurbishment and retrofit of complex and constrained legacy data centres. As quite often, the project will have to be delivered and phased, ensuring no impact is experienced by the neighbouring data halls and ancillary support systems." – Steven Davidson, Technical Director at Parmarbrook.

Sustainable refurbishments not only unlock the potential of existing data centres, but they also contribute to the reduction of carbon emissions and help businesses achieve their sustainability goals.

As pressures increase on the availability of suitable land for data centres, it is essential to consider sustainable refurbishments as a viable option, and we must work together as an industry to ensure that we build and refurbish buildings with the lowest possible carbon emissions. By unlocking the potential of under-utilised data centres and repurposing these structures, we can maximise their potential and provide data centres with a reduced carbon footprint, due to the saving of the existing structures' embodied carbon.

Data centres play a crucial role in the digital age, and their uptime is critical. However, outdated and inefficient buildings, unsuitable floor plans, and inadequate structural integrity mean that many legacy data centres are no longer fit for purpose, especially as the computing envelope of density and power intensity increases.

Sustainable refurbishments of existing data centres can provide much-needed space for data centres, minimise the embodied carbon impact, and lower the carbon footprint. With careful planning and consideration with like-minded Architectural and MEPH consultant partners, it is possible to unlock the potential of disused, or partially used, data centres and provide modern data centres that meet the requirements of today's advanced computing customers across a plethora of industries.

A small early expenditure for investigation works can provide an overall cost saving by de-risking the structure and can help identify the potential of a large volume of legacy properties.

OFC

The future of optical networking and communications

Register Today

A Hybrid Conference – In-Person and Virtual Presentations

Technical Conference: 24 – 28 March 2024
Exhibition: 26 – 28 March 2024
San Diego, California, USA

Plenary Speakers

Anita Döhler
Chief Executive Officer, NGMN Alliance

Inder Monga
Executive Director, ESnet, and Scientific Networking Division Director, Berkeley Lab

David Richardson
Partner Researcher, Microsoft

OFCConference.org

IEEE ComSoc IEEE Communications Society
IEEE Photonics Society
OPTICA Formerly OSA

Redefining liquid cooling from the server to the switch

Liquid cooling has long been a focal point in discussions surrounding data centers, and rightfully so, as these facilities are at the epicenter of an unprecedented data explosion.

BY NATHAN BLOM, CCO, ICEOTOPE

THE EXPLOSIVE GROWTH of the internet, cloud services, IoT devices, social media, and AI has fueled an unparalleled surge in data generation, intensifying the strain on rack densities and placing substantial demands on data center cooling systems. In fact, cooling power alone accounts for a staggering 40% of a data center's total energy consumption.

However, the need for efficient IT infrastructure cooling extends beyond data centers. Enterprise organizations are also looking for ways to reduce costs, maximize revenue and accelerate sustainability objectives. Not to mention the fact that reducing energy consumption is rapidly becoming one of the top priorities for telcos with thousands of sites in remote locations, making the reduction of maintenance costs key as well.

Liquid cooling technologies have emerged as a highly efficient solution for dissipating heat from IT equipment, regardless of the setting. Whether it's within a data center, on-premises data hall, cloud environment, or at the edge, liquid cooling is proving its versatility. While most applications have centered on cooling server components, new applications are rapidly materializing across the entire IT infrastructure spectrum.

Earlier this month, BT Group, in a groundbreaking move, initiated trials of liquid cooling technologies across its networks to enhance energy efficiency and reduce consumption as part of its commitment to achieving net-zero status. BT kicked off the trials with a network switch cooled using Iceotope's Precision Liquid Cooling technology and Juniper Networks® QFX Series Switches. With 90% of their overall energy consumption coming from networks, it's easy to see why reducing energy consumption is such a high priority.

In a similar vein, Meta released a study last year confirming the practicality, efficiency and effectiveness of Precision Liquid Cooling technology to meet the cooling requirements of high-density storage disks. Global data storage is growing at such a rate there is an increased need for improved thermal cooling solutions. Liquid cooling for high-density storage is proving to be a viable alternative as it can mitigate for variances and improve consistency. Ultimately, it lowers overall power consumption and improves ESG compliance.
Liquid cooling technologies are changing the game when it comes to removing heat from the IT stack. While each of the technologies on the market today have their time and place, there is a reason we are seeing Precision Liquid Cooling in trials that are broadening the use case for liquid cooling. Precision Liquid Cooling ensures maximum efficiency and reliability as it uses a small amount of dielectric coolant to precisely target and remove heat from the hottest components of the server. This approach not only eliminates the need for traditional air-cooling systems, but it allows for greater flexibility in designing IT solutions than any other solution on the market today. There are no hotspots that can slow down performance, no wasted physical space on unnecessary cooling infrastructure, and minimal need for water consumption.

As the demand for data increases, the importance of efficient and sustainable IT infrastructure cooling cannot be overstated. Liquid cooling, and Precision Liquid Cooling in particular, is at the forefront of this journey. Whether it's reducing the environmental footprint of data centers, enhancing the energy efficiency of telecommunication networks, or meeting the ever-increasing demands of high-density storage, liquid cooling offers versatile and effective solutions. These trials and applications are not just milestones; they represent a pivotal shift toward a future where cooling is smarter, greener, and more adaptable, empowering businesses to meet their evolving IT demands while contributing to a more sustainable world.

KAO DATA

Inspired by hyperscale. Built for industrial scale.

Award winning multi-site colocation.

Kao Data develop and operate advanced data centres for high performance computing. With a new 10MW facility operational in Harlow, and an industrial scale, 40MW data centre planned for Greater Manchester, we provide enterprise, cloud, HPC and AI customers with a world-class home for their compute.

info@kaodata.com ▪ kaodata.com

NVIDIA PREFERRED PARTNER

Why hybrid cooling is the future for data centre

Rising rack and power densities are driving significant interest in liquid cooling for many reasons. Yet, the suggestion that one size fits all ignores one of the most fundamental aspects of potentially hindering adoption - that many data centre applications will continue to utilize air as the most efficient and cost-effective solution for their cooling requirements. The future is undoubtedly hybrid, and by using air cooling, containment, and liquid cooling together, owners and operators can optimise and future-proof their data centre environments.

BY GORDON JOHNSON, SENIOR CFD MANAGER, SUBZERO ENGINEERING

TODAY, many data centres are experiencing increasing power density per IT rack, rising to levels that just a few years ago seemed extreme and out of reach, but today are considered both common and typical while simultaneously deploying air cooling. In 2020 for example, the Uptime Institute found that due to compute-intensive workloads, racks with densities of 20 kW and higher are becoming a reality for many data centres.

This increase has left data centre stakeholders wondering if air-cooled IT equipment (ITE), along with containment used to separate the cold supply air from the hot exhaust air, has finally reached its limits and if liquid cooling is the long-term solution. The answer is not as simple as yes or no, however.

Moving forward it's expected that data centres will transition from 100% air cooling to a hybrid model encompassing air and liquid-cooled solutions with all new and existing air-cooled data centres requiring containment to improve efficiency, performance, and sustainability. Additionally, those moving to liquid cooling may still require containment to support their mission-critical applications, depending on the type of server technology deployed.

One might ask why the debate of air versus liquid cooling is such a hot topic in the industry right now? To answer this question, we need to understand what's driving the need for liquid cooling, the other options, and how can we evaluate these options while continuing to utilize air as the primary cooling mechanism.

Can Air and Liquid Cooling Coexist?

For those who are newer to the industry, this is a position we've been in before, with air and liquid cooling successfully coexisting while removing substantial amounts of heat via intra-board air-to-water heat exchangers. This process continued until the industry shifted primarily to CMOS technology in the 1990s, and we've been using air cooling in our data centres ever since.

With air being the primary source used to cool data centres, ASHRAE (American Society of Heating, Refrigeration, and Air Conditioning Engineers) has worked towards making this technology as efficient and sustainable as possible.

Since 2004, its published a common set of criteria for cooling IT servers with the participation of ITE and cooling system manufacturers entitled "TC9.9 Thermal Guidelines for Data Processing Environments".

ASHRAE has focused on the efficiency and reliability of cooling the ITE in the data centre. Several revisions have been published with the latest being released in 2021 (revision 5). This latest generation TC9.9 highlights a new class of high-density air-cooled ITE (H1 class) which focuses more on cooling high-density servers and racks with a trade-off in terms of energy efficiency due to lower cooling supply air temperatures recommended to cool the ITE.

As to the question of whether or not air and liquid cooling can coexist in the data centre white space, it's done so for decades already, and moving forward, many experts expect to see these two cooling technologies coexisting for years to come.

What Do Server Power Trends Reveal?

It's easy to assume that when it comes to cooling, a one-size will fit all in terms of power and cooling consumption, both now and in the future, but that's not accurate. It's more important to focus on the actual workload for the data centre that we're designing or operating.

In the past, a common assumption with air cooling was that once you went above 25 kW per rack it was time to transition to liquid cooling. But the industry has made some changes in regards to this, enabling data centres to cool up to and even exceed 35 kW per rack with traditional air cooling.

Scientific data centres, which include largely GPU-driven applications like machine learning AI and high analytics like crypto mining, are the areas of the industry that typically are transitioning or moving towards liquid cooling. But if you look at some other workloads like the cloud and most businesses, the growth rate is rising but it still makes sense for air cooling in terms of cost. The key is to look at this issue from a business perspective, what are we trying to accomplish with each data centre?

What's Driving Server Power Growth?

Up to around 2010 businesses utilized single-core processors, but once available, they transitioned to multi-core processors, however, there still was a relatively flat power consumption with these dual and quad-core processors. This enabled server manufacturers to concentrate on lower airflow rates for cooling ITE, which resulted in better overall efficiency.

Around 2018, with the size of these processors continually shrinking, higher multi-core processors became the norm and with these reaching their performance limits, the only way to continue to achieve the new levels of performance by compute-intensive applications is by increasing power consumption. Server manufacturers have been packing in as much as they can to servers, but because of CPU power consumption, in some cases, data centres were having difficulty removing the heat with air cooling., creating a need for alternative cooling solutions, such as liquid.

Server manufacturers have also been increasing the temperature delta across servers for several years now, which again has been great for efficiency since the higher the temperature delta the less airflow that's needed to remove the heat. However, server manufacturers are, in turn, reaching their limits, resulting in data centre operators having to increase the airflow to cool high-density servers and to keep up with increasing power consumption.

Additional Options For Air Cooling

Thankfully, there are several approaches the industry is embracing to cool power densities up to and even greater than 35 kW per rack successfully, often with traditional air cooling. These options start with deploying either cold or hot aisle containment. If no containment is used typically, rack densities should be no higher than 5 kW per rack, with additional supply airflow needed to compensate for recirculation air and hot spots.

What about lowering temperatures? In 2021, ASHRAE released their 5th generation TC9.9 which highlighted a new class of High-Density Air-Cooled IT equipment, which will need to use more restrictive supply temperatures than the previous class of servers.

At some point, high-density servers and racks will also need to transition from air to liquid cooling, especially with CPUs and GPUs expected to exceed 500 watts per processor or higher in the next few years. But this transition is not automatic and isn't going to be for everyone.

Liquid cooling is not going to be the ideal solution or remedy for all future cooling requirements. Instead, the selection of liquid cooling instead of air cooling has to do with a variety of factors, including specific location, climate (temperature/humidity), power densities, workloads, efficiency, performance, heat reuse, and physical space available.

This highlights the need for data centre stakeholders to take a holistic approach to cooling their critical systems. It will not and should not be an approach where we're considering only air or only liquid cooling moving forward. Instead, the key is to understand the trade-offs of each cooling technology and deploy only what makes the most sense for the application.

Increasing data centre power consumption in an energy crisis

The role of generator maintenance in ensuring power resilience.

BY CRESTCHIC

THE GLOBAL ENERGY CRISIS is affecting the profitability and operational efficiency of businesses from all sectors, but for data centres and their inherently higher power consumption levels, the impacts are polarised. For the data centre market to achieve its projected growth trajectory, it must continue to adopt power-intensive next-generation IT technologies. And, it must do that in the face of an unstable grid, potential planned blackouts and suggested power rationing. That means rigorous investment in backup power and generator maintenance, at a time when energy bills are consuming investable profit. Paul Brickman, Commercial Director, for Crestchic Loadbanks, explores the often-overlooked necessity of loadbank testing in securing data centre power in an energy crisis.

Hiked power prices as a result of Russia's fuel sanctions, and the threat of planned blackouts or

power rationing to better manage an unstable grid, are forcing data centres to redistribute investment away from key business functions like training and recruitment, and towards improving energy efficiency and securing power resilience.

The current geopolitical challenges and their far-reaching implications are recognised in the 2023 Uptime Institute Report – Five Data Center Predictions for 2023 – in which the authors acknowledge that the Russian fuel sanctions, along with technological challenges "will make the planning of data centre development and operation more difficult."

Powering next-generation data centres

According to the Uptime Institute, IT hardware has been fairly standard in terms of its draw on mainstream server capacity for a few decades now, creating technical stability and relatively constant power and cooling requirement. This has enabled data centre designers to accommodate several IT refreshes without major upgrades to server technology and the associated hike in power consumption that would come with it. Essentially offering the latest technologies, without an increase in energy requirement.

This grace period is now over. Power requirements for next-generation IT hardware are far higher, rack power density is increasing, and "hotter" processors are putting pressure on the performance parameters of existing infrastructure. This rapid rise in IT power density means server power consumption is on a steep climb, creating a need for more power, at a time when the grid is unstable, and energy prices are at their highest in decades.

Extreme-density racks are now commonplace in technical computing too, as well as high-performance analytics and artificial intelligence training. If data centre operators want to successfully penetrate these niche markets, they will need to foot the heightened energy bill and maintain a robust power resilience regime until the situation stabilises.

Data centres take action

Demand for next-generation IT technologies will not slow. As a result, data centres are forging ahead with the essential upgrades required to UPS, batteries, switchgear and generators to accommodate increased power density.

Extreme-density racks are now commonplace in technical computing too, as well as high-performance analytics and artificial intelligence training

This is a risk. With an unstable grid and planned blackouts still on the agenda for many governments, upgrades will need to be bolstered with a watertight energy resilience plan to protect against power fluctuations and total outages.

The importance of a power resilience strategy in an energy crisis

According to research from the Ponemon Institute, in its third "Cost of Data Center Outages" report, the total cost of downtime has continued to rise over the last six years – rising a staggering 38% to $740,357 per incident. That equates to nearly $9,000 per minute – a figure that many data centres will have to swallow if the grid becomes too unstable to sustain their increasing power consumption.

Many data centres will already have a robust generator testing and maintenance regime in place, but the use of load banks is often overlooked, especially when budgets are stretched.

At a time when power outages are more likely, load bank testing should play an integral role in a data centre's energy resilience strategy and it would be prudent to evaluate your strategy in line with the current landscape to ensure it is robust enough.

With this in mind, what would be considered best practice for testing a backup power system?

Best practice load bank testing in times of crisis

Ideally, as a minimum all generators should be tested annually for real-world emergency conditions using a resistive-reactive 0.8pf load bank. Best practice dictates that all gensets (where there are multiple) should be run in a synchronised state, ideally for 8 hours but for a minimum of 3.

Where a resistive-only load bank is used, testing should be increased to 2-4 times per year at 3 hours per test. In carrying out this testing and maintenance, fuel, exhaust and cooling systems are effectively tested and system issues can be uncovered in a safe, controlled manner without the cost of major failure or unplanned downtime.

The alternator is not thoroughly tested though, with a resistive-only test and therefore a resistive-reactive test would always be recommended. It may be advisable to test more frequently during times of crisis for added peace of mind.

Why is resistive-reactive the best approach?

Capable of testing both resistive and reactive loads, this type of load bank provides a much clearer picture of how well an entire system will withstand changes in load patterns while experiencing the level of power that would typically be encountered under real operational conditions.

Furthermore, the inductive loads used in resistive/reactive testing will show how a system will cope with a voltage drop in its regulator. This is particularly important in any application which requires generators to be operated in parallel (prevalent in larger business infrastructures such as hyperscale data centres) where a problem with one generator could prevent other system generators from working properly or even failing to operate entirely.

This is something which is simply not achievable with resistive-only testing.

Navigating growth when power is scarce

No matter the geopolitical challenges and the effect it is having on power availability, data centres have no choice but to grow.

Demand will not cease, and power-intensive next-generation technologies are unavoidable. Ensuring power resilience via a watertight backup power supply and a robust testing and maintenance regime will enable data centre designers and operators to grow, safe in the knowledge that, should planned blackouts, power rationing or grid fluctuations happen, the power will always remain on.

For support with improving your load bank testing regime, search Crestchic Load Banks online.

EcoStruxure™
Innovation At Every Level

Realize your company's IT potential

Software to design, monitor, and manage your IT space

62% of IT outages can be attributed to IT infrastructure failure[1]. Our Data Center Infrastructure Management (DCIM) 3.0 offer provides device monitoring, health assessments, and more so you can:

- Run simulated impacts to expose vulnerabilities in IT infrastructure and address them immediately
- Reduce physical threats by monitoring IT environmental conditions
- Improve sustainability efforts by tracking PUE, energy, and carbon emissions

#CertaintyInAConnectedWorld

apc.com/edge

EcoStruxure™ IT modernized DCIM

APC Smart-UPS™ Modular Ultra

[1]Uptime Institute Global Data Center Survey, 2018

Life Is On | Schneider Electric

UPS evolution and why sustainability is the next frontier

As an industry, we now need to look seriously at sustainability and how it can be genuinely achieved within datacentres and other critical organisations.

BY DAVID BOND, CHAIRMAN, CENTIEL UK

In the beginning...

TO UNDERSTAND the future, often we must look to the past. Modern humans evolved from primates, in various stages, over several hundred thousand years with each evolutionary stage having significant advantages over the previous stage. The uninterruptible power supply (UPS) industry has undergone a similar evolutionary process, but in a rather shorter time frame.

In the beginning (let's say 40 years ago), there were stand-alone UPS systems. These were huge monolithic units the size of a double wardrobe with a typical 60kVA (48kW) UPS weighing around 600kg.

When increased power availability was required (let's say 35 years ago) because of increasingly critical loads, parallel redundant UPS systems evolved. Back then, the only way to achieve parallel redundancy was to place two or more stand-alone UPS next to each other and install communication cables between each UPS.

When increased operating efficiency was required (let's say 30 years ago) because of increasing "environmentally friendly" awareness and the rising cost of energy, transformer-less UPS systems evolved. This evolution increased UPS operating efficiency by around 5%, decreased their size and weight, reduced their cost and enabled the next stage of evolution – modular UPS.

Around 20 years ago the first true modular UPS were developed, heralding the arrival of increased system availability, scalability, flexibility and maintainability. It is the development of these true modular UPS that we shall focus on now.

Where are we now?

Power protection systems must be available to provide clean, guaranteed power to their critical load every second of every day, so maximizing system availability must be the overriding objective of any installation. To achieve maximum availability, all single points of failure need to be removed.

Centiel's 4th Generation true modular UPS system, CumulusPowerTM, provides a significant improvement over previous designs. Each UPS module in the system contains all the power elements of a UPS – rectifier, inverter, static switch, display - and critically - all control and monitoring circuitry so there are no single points of failure.

Availability for Centiel's CumulusPowerTM UPS is further improved by the incorporation of Centiel's Distributed Active Redundant Architecture (DARA) and their Triple Mode communications bus. Duplication and redundancy of UPS components must also apply to the means of communication between UPS modules. The earliest, and simplest, communications bus is a single cable. However, if this breaks or becomes disconnected, the entire UPS system is compromised. For this reason, the ring circuit was introduced so if the communication cable was cut in one direction the signals could communicate the other way around the ring. The introduction of Centiel's Triple Mode communications bus increases system availability even further.

As the name suggests, Centiel's Triple Mode communications bus provides three separate paths of communication between UPS modules and frames made up of three separate ring circuits and three separate brains communicating with three other separate brains. It's the belt, braces and buttons approach where single points of failure are completely removed.

It is the combination of no single points of failure redundancy, DARA, Triple Mode bus and Centiel's safe/hot swap technologies that gives Centiel's 4th generation true modular UPS systems their class leading 'nine-nines' or 99.9999999% availability.

Scalability and flexibility are also important when installing a system to ensure the continual 'right sizing' of the UPS. A system which is too small will be overloaded, compromising reliability and availability, whilst one which is too large will be inefficient, will waste energy and will be unnecessarily costly to run. CumulusPowerTM has a

flat efficiency curve and provides greater than 97% efficiency even at low loads.

The next evolutionary stage

As we can see, the prevailing need has driven important developments in UPS technology over the years. The need facing us all now, and hence in UPS development, is Sustainability. Datacentres, for example, are an essential part of modern life but burn massive amounts of energy and use huge volumes of water. The earth is getting warmer and we are beginning to realise that some resources are finite. We must all work towards developing more sustainable solutions.

Centiel's development team has long been at the forefront of solving prevailing technology challenges. In the late 1980s I wrote a paper imagining the perfect UPS. I said it would be 100% efficient (we are now close to 98%), it would offer 100% availability (we are now at 99.9999999%), it would present a perfect load to the mains supply (it is now unity power factor with input current total harmonic distortion - THDi - of <1%) and would last forever. The development team behind Centiel, who created the first transformerless UPS and the first three phase true modular UPS, have pretty much solved the first three challenges to create the (almost) perfect UPS, and for the past four years they have been innovating to develop a UPS with a design life to match the design life of a datacentre (typically 30 years) and can be recycled to make UPS more sustainable.

Earlier this year, Centiel launched its new UPS, StratusPowerTM, which shares all the benefits of its award-winning three phase, true modular UPS CumulusPowerTM - including "9 nines" (99.9999999%) availability to effectively eliminate system downtime but now also offers class leading 97.6% on-line efficiency to minimise running costs and high quality, long-life components to improve sustainability.

Higher quality components cost more but this is just tens of pounds per module, compared to thousands to replace an entire system with a shorter design life. Combined with Centiel's approach as trusted advisors, it means they can help organisations take steps to move away from a "throw away" culture with a genuinely sustainable offering whilst helping to reduce Total Cost of Ownership (TCO) at the same time.

And finally...

Like all Centiel's UPS, StratusPowerTM is manufactured at its factory in Switzerland. However, uniquely, it includes even higher quality components so instead of replacing filter capacitors and cooling fans every four years, they now need replacing every 15 years, or just once during StratusPower's entire 30 year design life. Furthermore, at end of life, StratusPowerTM can also be almost 100% recycled.

Like Centiel's whole range of UPS solutions, UPS cabinets are designed with scalability and flexibility in mind, and future load changes are easily accommodated by adding or removing UPS modules as required. An organisation will never outgrow a well specified StratusPowerTM UPS and it can be constantly rightsized to ensure it always operates at the optimal point in its efficiency curve.

Centiel's UPS solutions also work with Lithium Iron Phosphate (LiFePO4) batteries which tolerate higher ambient temperatures, further reducing the need for cooling, and will also only need to be replaced once in a 30 year design life.

As an industry, we now need to look seriously at sustainability and how it can be genuinely achieved within datacentres and other critical organisations. It is the next step in UPS evolution and StratusPowerTM is a significant move towards this goal.

Centiel's experienced team is always available to advise on the most appropriate options for organisations looking to improve their approach to sustainability and reduce their carbon footprint while maintaining the highest level of availability for critical power protection. Centiel can help organisations take steps to move towards genuine sustainability.

The team at Centiel has been at leading-edge of UPS development and evolution over many decades. Centiel's goal is clear: to achieve the ultimate availability of power for its client base. Its quality technology, backed-up with comprehensive maintenance contracts ensure Centiel's clients' have the very best power protection at all times, now and in the future.

For further information please see:
www.centiel.co.uk

The Digital Age, built by Ironopolis

Durata Modular Data Centre Solutions operates in a market where there is a growing demand for its services that will not be going away.

THE TEAM at Middlesbrough-based Durata believes Teesside - an area commonly referred to locally as Ironopolis over the years - can quickly become a major player in the Digital Age.

The dream is for that part of North-East England, rich in history for iron and steel production, to become the home of the modular data centre. There is every chance that will come true.

The Application – faster, efficient and scalable

The modular data centre market is estimated to achieve a market size of USD 113.6bn by 2032, growing from an accounted USD 22.3bn last year. This is a result of anticipated modernisation of networks from telecom companies expected to drive growth for years to come through to MW power requirements for the ever-expanding hyperscale operators.

“We cater for the data centre market as a collective, whether that is constructing for compute or power,” said Lewis Cobb, Durata’s head of modular data centre solutions projects.

“Here on Teesside, we pull together Durata’s expertise in mechanical and electrical engineering and combine it with the 170-year infrastructure in steelworks and fabrication to deliver our Modular Data Centre Solutions that will enhance data centre infrastructure for our clients.”

The Execution – mechanical, electrical and fabrication

Teesside, deeply proud of its industrial heritage, has already established itself as a place for modular data centre solutions to be built before being delivered around the world.

Lewis said: “The data centre industry needs modular solutions. By operators investing in modular solutions for compute and power, they can build faster, deliver efficiently and deploy where required. “In order for this success, a key requirement for a modular solution provider is infrastructure.”

That outlook formed part of the vision when John McGee, the managing director, set up Durata ten years ago. Now that dream is fast becoming a reality.

Having grown from strength to strength in recent years, Durata, specialists in critical power infrastructure and modular data centre solutions, is now expanding its services across Europe from

The Durata Innovation Centre marks a leap in experience for our clients. Our long-standing goal is to turn Teesside into a home for modular data centre solutions, and by having a design studio for collaborative experiences

a base in Frankfurt, Germany, and securing new contracts across different countries having attracted the attention of global players across Europe, the Middle East and Africa (EMEA).

Durata has the infrastructure in place to support the modular solutions required by the data centre sector.

Concept to completion

Lewis said: "The concept to completion phrase is important because within that there are variables that make us a very attractive proposition.
"Our in-house design and Computer Aided Design (CAD) team allow us to develop a solution, we purchase steel from source and construct our own modules using Teesside infrastructure, followed by fitout from our Durata trained mechanical and electrical engineers.

"This is all overseen by our Durata project managers. We own the complete process with full control to ensure our clientele receive engineering excellence and continuity for scalable projects."
Durata have already welcomed various multi-national clients and hyperscale data centre builders to Teesside for an insight into how these modules can be constructed and transported by road, sea, air and rail.

By welcoming and showing their clients a host of facilities, they are demonstrating a manufacturing process paired with engineering intellect, and 2024 holds expansion plans to enhance that experience.
Lewis said: "The Durata Innovation Centre marks a leap in experience for our clients. Our long-standing goal is to turn Teesside into a home for modular data centre solutions, and by having a design studio for collaborative experiences, by having an unrivalled manufacturing process and to show how we can deploy globally via the Tees port, we believe we are the best positioned partner for modular solutions.
"We are growing. As well as extending our office in Middlesbrough to create an Innovation Centre, our office in Frankfurt now means we have a presence that opens the door to the rest of Europe, the Middle East and Asia.

"Frankfurt shows we are committed to the European market. Germany is a prominent market in the data centre industry and aids us in tapping into the continent.

"We have an expanding internal network of international partners that are opening doors and centralising enquiries to our Teesside headquarters. We have a turnkey execution and that is exciting for our clients, both existing and potential.

"Even though our headquarters are in Middlesbrough. We are taking the Durata brand and our modular solutions everywhere."

Durata has live projects in Europe and the Middle East, and soon Africa. It is operating in critical power, providing services for low yield uninterruptible power supplies (UPS) through to MW power for the hyperscale market.

Durata is also servicing the compute sector of the industry too, whether that is edge computing through to high density AI requirements. There is also a conscious effort to adapt with the changing landscape of density requirements, including Immersion technologies within the portfolio of solutions.

Lewis said: "Because of our flexibility we can cater for fast growing markets that need that flexibility.
"It is an exciting proposition that we are manufacturing data centre solutions here in Teesside that will aid in shaping the digital age, and deploying globally matching our clients business models.

"By having our Teesside hub of excellence we can continually distribute highly engineered solutions across a global landscape.

"Durata's modular data centre solutions are engineering excellence. And with preferred equipment collaboration, the result is a flexible, scalable, innovative modular solution for the data centres of today and tomorrow."

*For further information check out **duratauk.com**

The DCA News

By DCA CEO Steve Hone

THIS MONTH'S DCA News is a collaboration of articles from Venessa Moffat and the Data Centre Team at QIO Technologies.

This specialist guest feature comprises of a number of important pieces. Beginning with an overview of QIO Technologies and their work related to energy efficiency / sustainability in Data Centres.

This is followed by two informative pieces, the first *'Cracking the Sustainability Challenge',* looks into how DC Owners & Operators need to gear up to meet the increasing demand for digital services and how the ability to do so sustainably will be key to commercial viability.

The second article *'A Road Map for Action'* details the complexity of new reporting requirements and explains the importance of what operators need to measure and why.

Finally, we have a focus piece *'The role for AI'* and how this will be the key to unpicking the future sustainability challenge for data centres.

I'd like to thank the team at QIO, including Venessa Moffat, Katrina Koffler for taking the time to compile this informative feature, I hope it will be of interest.

Cracking the Sustainability Challenge for Data Centers

David McGuirk, Director of data center and telecoms, QiO technologies

QiO Profile
QiO TECHNOLOGIES is an award-winning climate tech company delivering IoT AI software that supports energy-intensive Data Center, Telecom and Industrial companies to reduce energy costs, improve resource efficiency, and meet the evolving demands of sustainability and net zero reporting.

We harness the power of data to identify operational sustainability, leveraging AI to analyse large disparate sources of operational data and identify the best operational practices without the need for an army of data scientists. Our AI continuously tracks and monitors real-time data to optimise resource use and cut GHG emissions while maintaining profitability and output.

For our data center customers, we have engineered a solution in partnership with Intel to transform server power management by targeting CPUs to reduce energy consumption and auto-adjust time-of-day workloads without SLA impact. This innovative approach takes the focus away from PUE and IT overhead efficiency, moving it to the whole IT stack to reduce energy consumption, reduce emissions and deliver a cascading sustainability effect: less power used equals less cooling required.
As both data center operators and enterprises gear up to meet the increasing demand for digital services, we are working with them to help control costs, deliver net zero targets and meet the rising demands and complexity of reporting requirements.

Built for data centers

At a time of energy market volatility and global inflationary pressures, our data center customers benefit from a sustainability journey which delivers a return on investment in the short and long term.

a. Improving observability
We collect telemetry data from chipsets to rapidly baseline and identify energy savings and emissions across server fleets at individual device level.

b. Boosting server sustainability to reduce energy costs by up to 52%
We analyse server workload and energy consumption patterns to improve power management efficiency, delivering closed-loop automation which adjusts performance/sleep states to match time of day workload.

Our AI also continuously tracks and monitors real-time data to maintain performance, providing predictive analytics that extends capex by extending the remaining useful life of assets within a fleet.

c. Improving facility efficiency to reduce cost by 5-15%
Our AI works across the complex 'systems of systems' within data centers tied to power and cooling, outside temperature, water usage and waste heat reuse to identify HVAC Setpoint Optimisation and Hybrid Energy Mix Optimisation.

d. Automated reporting
Continuous real-time data monitoring enables our customers to reduce, track and report Scope 1 and Scope 2 emissions in preferred reporting mechanisms. In the face of rising complexity, this makes the task of reporting more straightforward. Our reporting is aligned to all relevant standards including EnergyStar, EPC, ISO 50001, TCFD and EU Directive 2014/95,

Cracking the Sustainability Challenge for Data Center Operators

OUR FAST-PACED WORLD is digital. Behind our banking, shopping, entertainment, universities, governments and jobs lies an enormous IT infrastructure housed in data centers designed to handle huge volumes of traffic.

As data center operators and enterprise owners of data centers gear up to meet increasing demand for these digital services, the ability to do so sustainably will be key to commercial viability with pressure coming on five different fronts.

Sustainability standards and reporting
The first major challenge for data center operators comes from the need for public reporting on energy use and sustainability.

The EU is leading the way with the need for operators of data centers above 500 kW to make their environmental performance public at least once a year under the Energy Efficiency Directive (EED). Separately reporting for the Corporate Sustainability Reporting Directive requires exacting data measurement and disclosure across nine areas of resource use and IT equipment - in other words, the whole IT stack.

With other geographies set to follow suit and voluntary sustainability reporting becoming more commonplace, every data center operator or enterprise owner of data centers will need to accurately measure sustainability performance in the future.

This isn't just a consideration for the 55,000 or so businesses who fall into the scope of the EU's new regulations but for firms who are part of their supply chains and will have to report their own sustainability performance in the future.

Hitting the buffers of current sustainability efforts
Until now, a main benchmark and focal point for data center efficiency has been Power Usage Effectiveness (PUE) – the ratio of the total amount of power delivered to a facility compared to that provided to the computing equipment.

While the quest for lower PUE figures contributed to greater data center efficiency ten years ago, data center operators have to find new ways of improving

efficiency as demand for infrastructure grows. This will require more nuanced interrogation of performance metrics such as IT Equipment Energy Efficiency (ITEEsv) and IT Equipment Utilisation (ITEUsv) and the modelling and measurement of changes which can deliver efficiency gains.

As data centers are inherently complex 'systems of systems', with different control systems and different transport protocols tied to environmental systems such as power and cooling, outside temperature, water usage and waste heat reuse, this will demand new skills, knowledge and operational practice.

New constraints on resource use
While demand for data center infrastructure in the US and the EU is expected to double over the next seven years, data center operators are already hitting limits of sustainable resource use or going beyond it.

In Ireland, power supply cannot meet energy demand for planned data centers until 2028 at the earliest. In the US, communities are pushing back against water-use by data center operators, while analysis from Virginia Tech has found that one in five US data centers takes water from moderately or highly stressed watersheds.

Data center operators must therefore anticipate new limitations on their use of resources in the

future, with increasing scrutiny on the efficiency and sustainability of current operations.

The commercial case for sustainability
While the energy price spike of the past 18 months has reminded businesses that conservative use of resource isn't just good for the planet but it is key to commercial viability and competitiveness, the financial case for sustainability goes beyond the ability to manage costs and profitability.

Increasingly, access to capital is driven by environmental and sustainability performance with institutional and private equity investors seeing these factors as key indicators of long-term value. Going forwards, enterprise and owner-operators of data centers who do not or cannot report on sustainability performance will have to fight for a shrinking number of investors.

Meeting net-zero commitments
The final area of pressure for data center operators comes from the international commitments to net zero emissions targets.

As countries strive to freeze net greenhouse gas emissions by 2025, and then reduce them by 55% by 2030 before reaching net zero by 2050, businesses must not only make science-based targets to support these goals but also show they have a way of meeting them in the context of current and future operations. Increasingly, firms are facing scrutiny and criticism where plans rely on unproven technology or unreliable offsetting schemes.

A Roadmap for Action

TAKEN TOGETHER, the pressing nature of these challenges means there is a clear imperative for data center operators to reliably measure, deliver and report concrete improvements in operational sustainability in the next twelve months. To achieve this, three critical shifts need to take place.

Understand the reporting landscape and its demands on your business
The complexity of new reporting requirements in Europe, as well as those emerging in the UK and US, means that operators must move quickly to understand what they need to measure and why, now and in the future.

From now on, data centers above 500 kW need to report on environmental performance connected to energy consumption, power utilisation, temperature, cooling efficiency ratios, water usage, heat utilisation, energy reuse and use of renewable energy in line with CEN/CENELEC EN 50600-4 standards.

Data centers with a total rated energy input exceeding 1 MW also need to recover their waste heat, or at least prove that it's not possible for them to do so.

All data centers need to recognise the level of transparency and outside scrutiny that new reporting will bring. Under the EED and CSRD initiatives, the European Commission will introduce a European database on data centers that will include all aggregated data provided by each data center. As a matter of urgency, operators need to respond by setting the right benchmarks that will enable them to measure the right things. And just as importantly, create an operating framework that will drive the improvements required.

They will also need to keep a watching brief on changes to directives. For example, The European Commission has said that it will use the data it collects in the first rounds of reporting to develop a revision of the EED by May 2025. This is likely to include a roadmap towards net-zero emission data centers.

Prepare for business change
To drive this level of action, operators will need to change the way they are structured and work. This starts with leadership and accountability.

As a first step, operators that haven't already done so need to appoint a Head of Sustainability or other senior person with specific responsibility for environmental performance improvement. Such leaders will need to take full responsibility for understanding new reporting requirements, auditing the sustainability landscape and identifying gaps that could see them fail to improve and comply with the wide range of new standardised metrics.

Secondly, operators must work to foster a new culture of collaboration between departments that are currently siloed. Today, responsibility for a data center typically falls between two teams with conflicting agendas: a facilities or estates team

responsible for managing the energy budget; and an IT team tasked with maintaining a high-performing, high-availability IT estate. Both teams need to work together towards shared sustainability goals.

Thirdly, while integrating departments and systems to provide real-time data for transparency across the estate is a new must, real-time data integrations create real-time security risks. Operators must take immediate action by appointing a Chief Information Security Officer with specific responsibility for mitigating these risks; without blocking progress; and ensuring that the new restructured arrangements for improving sustainability don't expose new vulnerabilities.

All of these teams need to work together with a clear mandate and targets. To achieve this, operators need to review their internal organisation, restructure and incentivise all parties to work together towards the same goals.

Heads of Sustainability are well placed to take ownership of this new arrangement, and should be backed with control of energy budgets that have traditionally been managed by facilities teams. Ideally, these should be zero-based budgets, with a requirement that initiatives pay back within a year in both financial and energy savings.For everyone, the mantra needs to be that sustainability is not part of the business plan: it is the business plan.

Use AI tools to accelerate progress

Operators must also recognise that new efficiencies will be key to delivering reporting for CSRD, EED and other sustainability standards. For this to happen, they need to progress quickly on a number of technical fronts in order to measure the right things and then identify, test and iterate the improvements needed.

This task requires interrogating large amounts of data, modelling use cases and then seeing how they perform in real life. But even if organisations could hire and deploy big teams of data scientists, the scale of the task ahead of them is simply too big. Across large estates of servers, interrogating vast amounts of data to predict, test and measure performance in different use case scenarios (and manually evaluating each server type and build) is beyond the current capability and time-resources of data center operators. The process is also vulnerable to significant levels of human error.

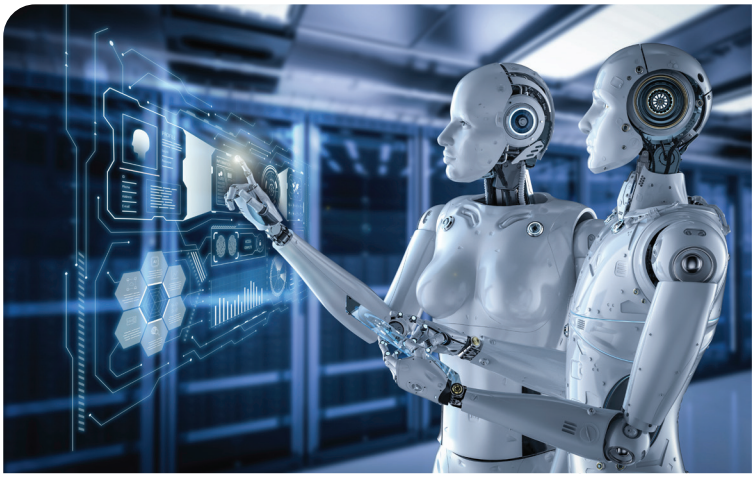
This means operators need to evaluate new technical solutions that will help them achieve their new goals. In particular, new AI systems designed specifically for data centers are uniquely placed to help operators by producing real-time data on resource use (energy and water consumption) and all other relevant factors required by the new standard at multiple levels – with particular focus directed towards the IT stack.

As a matter of urgency, operators need to understand the emerging ecosystem of climate tech and sustainability tech suppliers that play in this space, what they can deliver, how they fit together and how they will support future growth and viability.

The Role for AI

GIVEN the pressing nature of the challenge, not to mention new reporting deadlines, AI will be the key to unpicking the future sustainability challenge for data centers.

Suites of ready-to-deploy AI products with in-built algorithms are already available that can integrate with existing data center infrastructure management systems. These can recommend and implement optimal control settings on HVAC and cooling systems, as as well as CPU processors, helping to reduce energy consumption and costs, and carbon emissions.



Key benefits

The breakthrough benefits include autonomous control of IT systems that transform efficiency – and with no impact on business continuity or service levels. For example. AI systems can automatically adjust performance and sleep states to match time-of-day workloads while maintaining Quality of Service. Operators can run systems on open or closed-loop, so they can either manually implement suggested state changes or allow automatic implementation. Energy mix optimisation is also key. AI systems can use location specific weather forecast data to accurately predict energy generation from onsite solar or wind distributed energy sources. This helps to reduce fuel consumption and energy demand during peak tariff times, resulting in reduced emissions and lower running costs.

Energy efficiency indexing is another important benefit of AI. Operators can reduce energy and water consumption and thermal management costs in real-time without spending hours manually gathering and analysing time-series usage and workload data.

AI-driven energy efficiency indexing algorithms can recommend fan speed, temperature and even processor power state settings based on real-time server workload, resulting in quantified cost savings and reduced carbon emissions.

Predictive analytics within AI systems can also improve efficiency by helping operators to predict the failure of equipment – helping them to sweat assets for longer and reduce costs further. Modelling and optimisation of all equipment is an advantage too. Like a digital twin for buildings and HVAC – AI systems can plan the power and cooling distribution systems and model them within a piece of software. "What if" scenarios show how a change in one piece of equipment will affect the whole system of systems. Other scenarios can look at performance of systems and help companies model the most efficient energy setup.

Enhanced security is another plus. While integrating departments and systems to provide real-time data for transparency across the estate is a new must, real-time data integrations create real-time security risks. Crucially, AI systems can solve this problem by automating threat detection by detecting anomalies and securing all endpoints from cyber and physical threats.

This helps to support CSIOs that are responsible for managing emerging threat vectors that will inevitably increase following the integrations required to meet new reporting standards.

Energy cost reduction

FOR ALL THOSE who are involved in improving data center efficiency to comply with new regulations – which will include IT, finance, facilities and sustainability teams - perhaps the most eye catching additional theme throughout these benefits is the potential for reducing costs amid rising energy prices.

Based on our own technology, here's an example of how this works technically through QiO Foresight Optima DC+™, a new product we have developed in collaboration with Intel for data center optimisation at the server level that has a major impact on energy consumption:

- For each CPU core in the data center, Foresight OptimaDC+™ plots an Energy Efficiency Index (EEI) based on telemetry data from past processes to determine which is the most energy efficient process and why.
- The EEI then becomes a standard metric for comparison across different machines in different locations and can be used to establish best practices.
- Data used to develop the EEI metric also trains Foresight Optima DC+™, which is then deployed to identify strategies for reducing energy consumption.
- Foresight Optima DC+™ includes a reinforcement learning AI model that can perceive and interpret its environment to learn through trial and error, training itself to come up with a solution to optimise power consumption.
- Foresight Optima DC+™ will search for the most efficient way to operate the servers.
- With a full understanding of the performance needs, it can also recommend adjustments to sleep / power states of that processor to reduce the energy consumption during times the CPU is idle or could deliver the required performance using less energy.

Conclusion

Data centers have worked hard to minimise costs and energy consumption in the past decade, even as their workload volumes have grown dramatically. As the demand for data continues to grow, they will find it difficult to continue that trend without a change in approach and the tools they use.

It is now critical they work even harder to make critical savings in their energy budget, comply with new sustainability regulations and reduce carbon emissions. AI is the game changing technology that will enable the shift from 'business as usual' to data driven decision making and automation for a sustainable business within a net zero economy.

Visit www.qio.io to learn how you can leverage AI to optimise operational sustainability, reduce costs and cut carbon emissions.