Monitoring and Optimizing AI Datacenter Facilities

Sean Graham
Research Director,
Cloud to Edge Datacenter Trends

Executive Summary

Abstract

This IDC report addresses the critical challenges facing datacenter operators as they navigate the competing demands of growth and environmental sustainability. With the rapid expansion of generative AI (GenAI) and digital transformation, datacenter energy consumption is projected to rise sharply, increasing both costs and carbon emissions. Despite organizations' efforts to improve energy efficiency and invest in renewable technologies, sustainability initiatives are not keeping pace with growing demands. The report also underscores the importance of AI and intelligent systems in enhancing datacenter efficiency, with many organizations viewing these technologies as essential to their sustainability strategies.
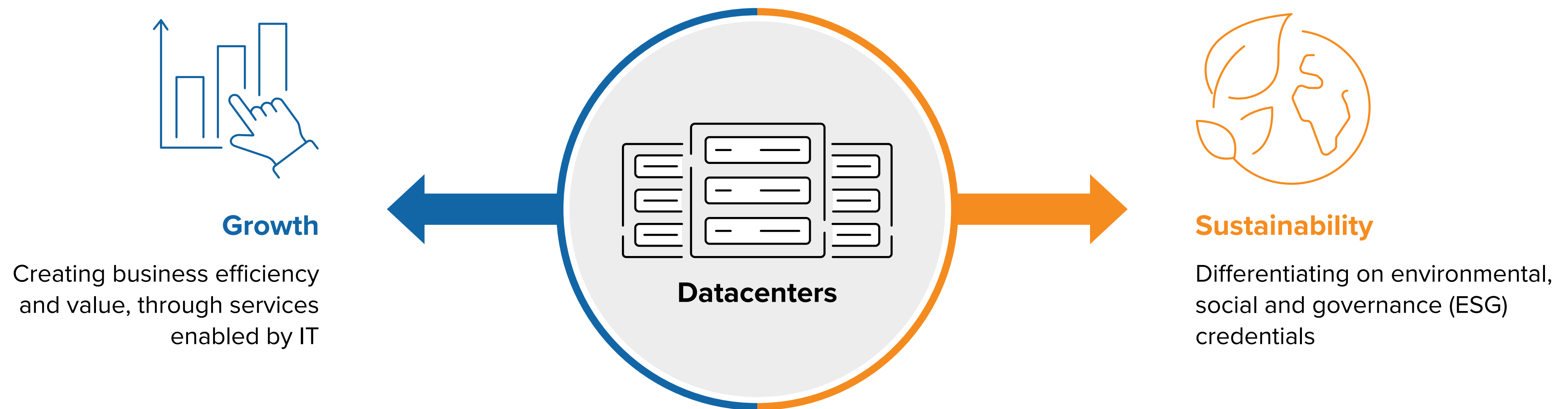
Key Findings

- Datacenter Energy consumption is expected to post a 19.5% compound annual growth rate (CAGR) from 2023 to 2028, rising from 352 terawatt hours (TWh) to 857 TWh.

- The cost of energy is becoming the largest operational expense for datacenter facilities due to increased consumption.

- The datacenter industry's carbon emissions are predicted to record a CAGR of 14.2% from 2023 to 2028, outpacing many sustainability efforts.

- According to an IDC datacenter operator survey, the average power utilization of a datacenter is 43%.

- 77% of organizations view AI as critical for their sustainability journey.

Recommendations

- **Understand the Current Environment:** Start with visibility into your existing datacenter energy consumption, energy costs, and carbon emissions.

- **Implement a Comprehensive Datacenter Management System:** Track key performance indicators (KPIs) related to energy and environmental factors and use design tools to unlock stranded capacity.

- **Implement AI for Datacenter Efficiency:** Empower datacenter operators to leverage AI algorithms to monitor and manage energy consumption.

- **Choose Projects with a Short-Term ROI:** Projects with a quick payback period focused on automation and AI are more likely to be funded, and the resulting savings will fuel future investments.

The preeminent challenge of the datacenter industry is the struggle between growth and environmental sustainability.

Growth

Creating business efficiency and value, through services enabled by IT

Datacenters

Sustainability

Differentiating on environmental, social and governance (ESG) credentials

The proliferation of generative AI and continued investment in digital transformation is natural-resource intensive, and both enterprises and colocation providers need additional datacenter capacity. However, organizations are also differentiating via sustainability efforts that align to customer, investor, and stakeholder values.

With the potential to drive business outcomes from GenAI and digital transformation, datacenter operators have notably shifted toward business-centric initiatives.

Q. What are your organization's top 3 initiatives related to datacenter priorities currently and over the next two years?
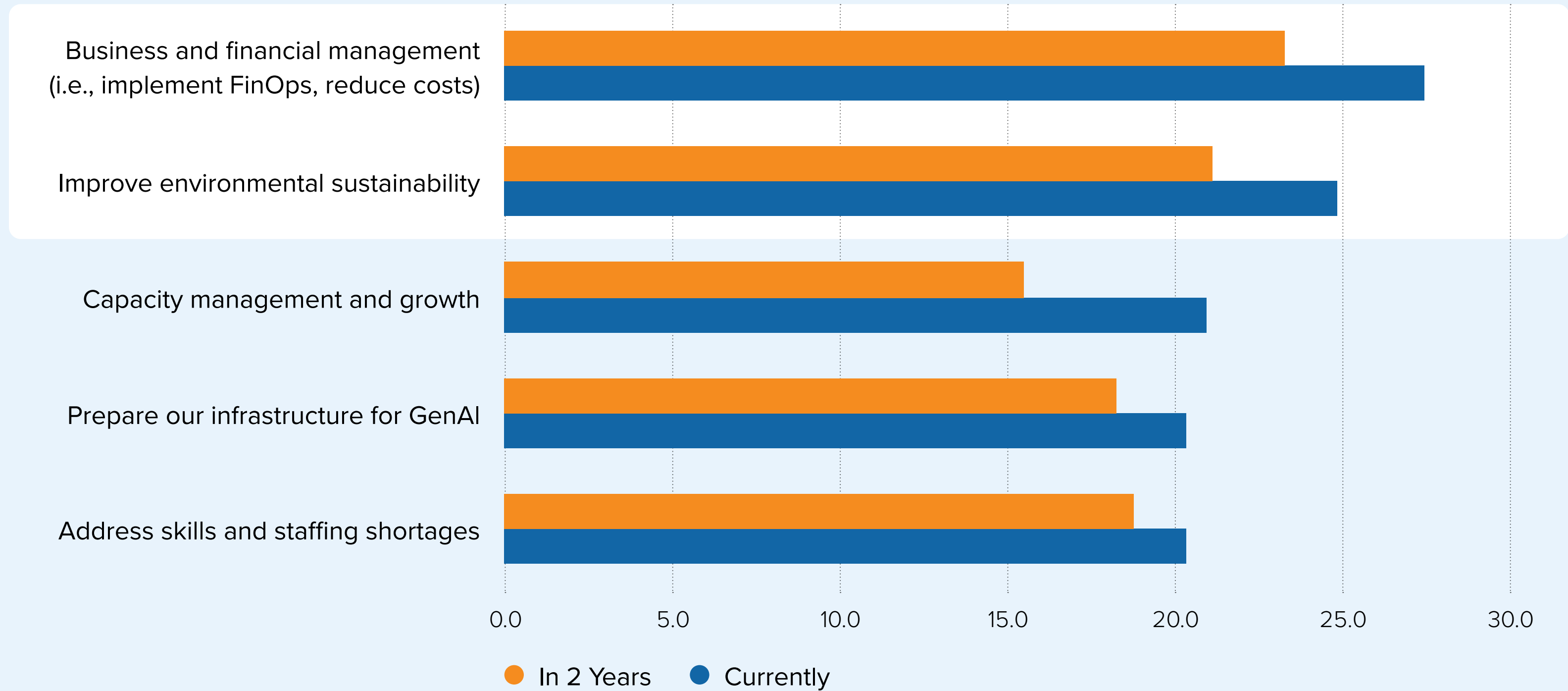

The top priorities for enterprise datacenter operators are *business and financial management of datacenters* and *improving environmental sustainability*. Datacenter operators expect these to remain top priorities, given their importance and complexity.

Datacenter operators will favor solutions that promote energy efficiency, as they address both of their top initiatives.
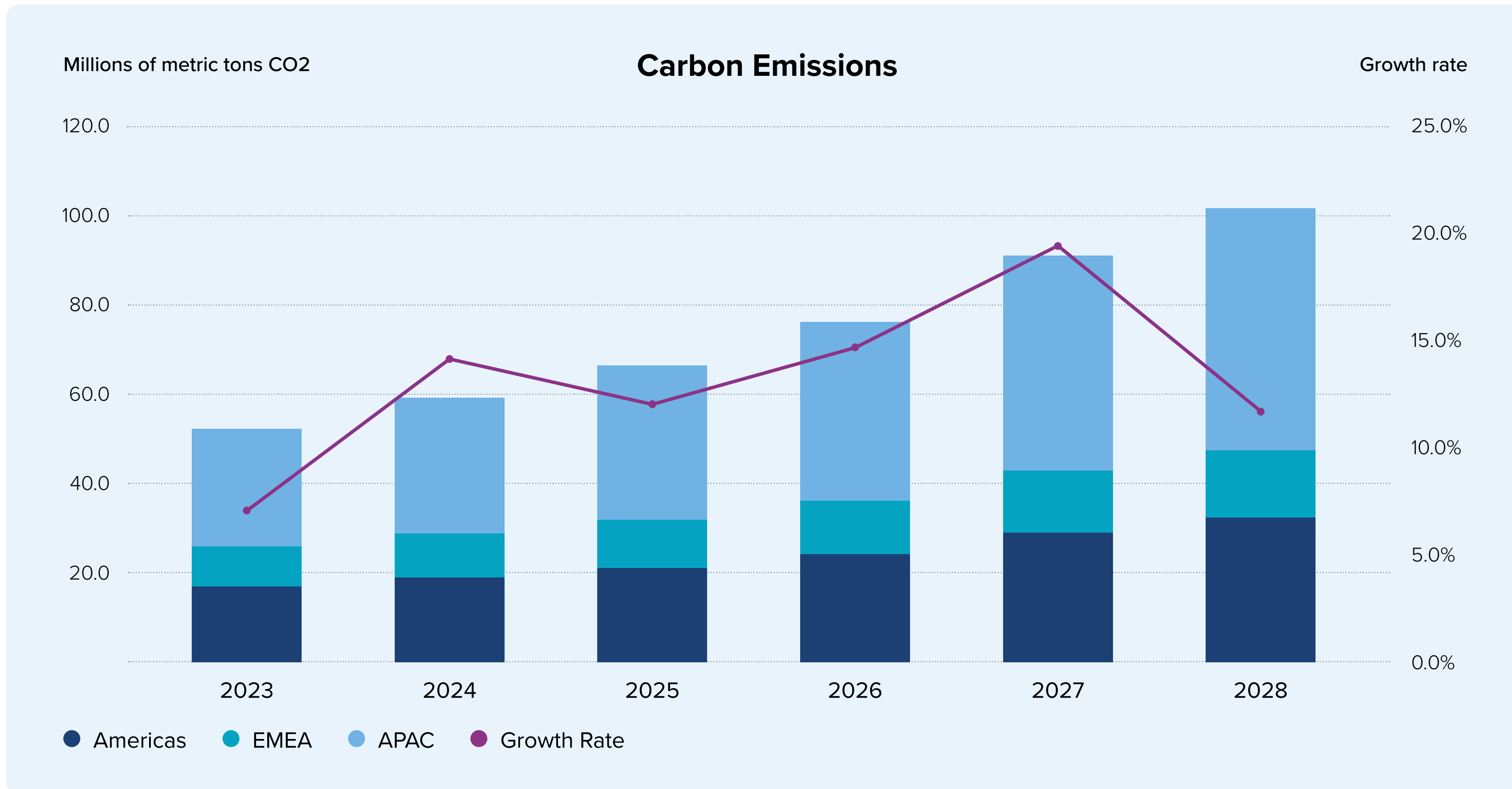
Datacenter energy consumption is expected to grow 2.4 times between 2023 and 2028.

With the proliferation of generative AI and continued investment in digital transformation, datacenter energy consumption is predicted to record a CAGR of 19.5%, rising from 352 Terawatt Hours (TWh) to 857 TWh, by 2028.
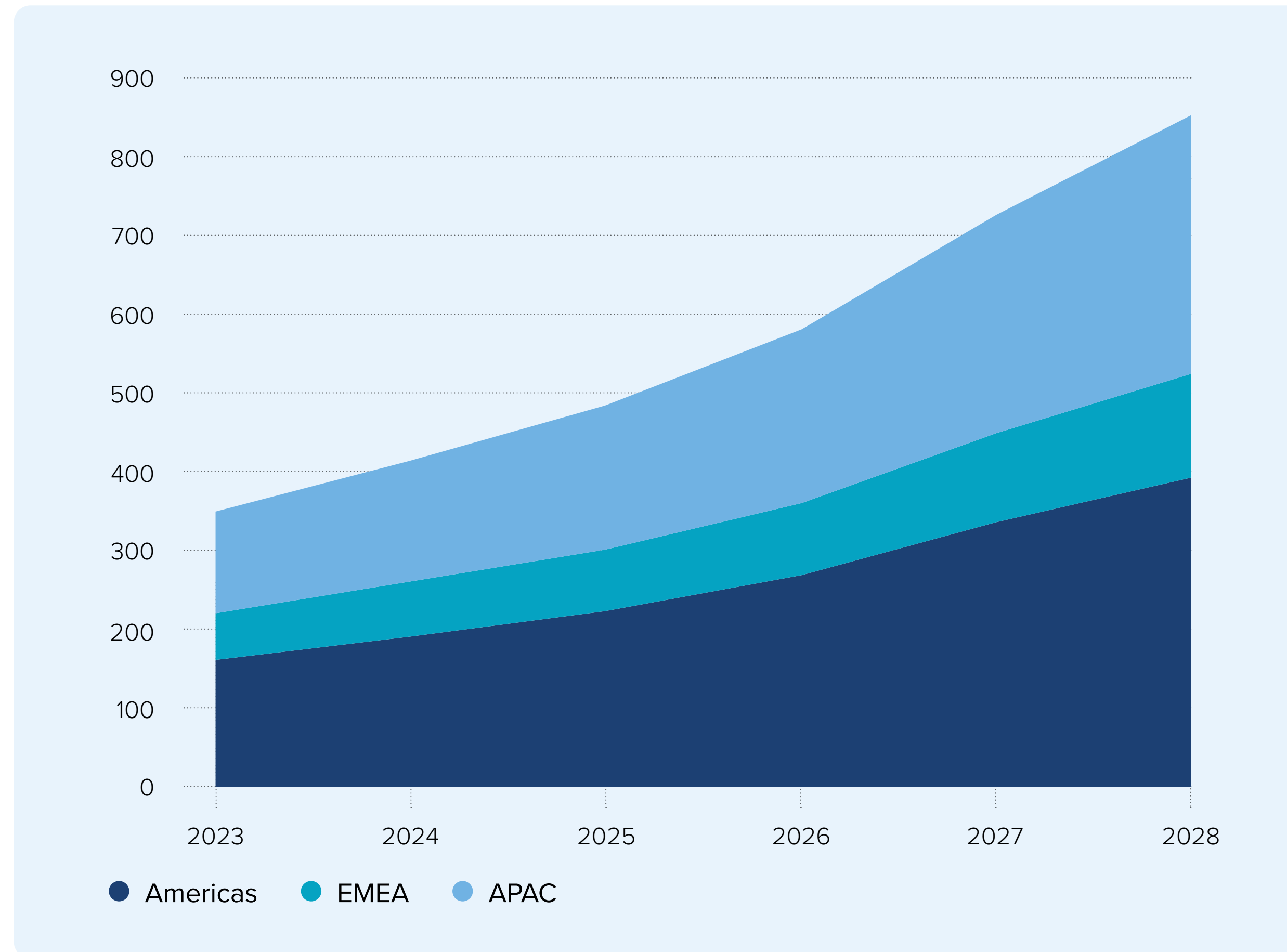
Energy is the largest operating cost for datacenter facilities. The industry is experiencing a doubling in energy costs due to increased consumption. In addition, the growing scarcity of power is driving up the costs of other limited resources, exacerbating the financial burden on the sector.

Datacenter operators are working to become more sustainable, but workload demand is outpacing sustainability innovations.

Carbon Emissions

Millions of metric tons CO2 / Growth rate

Legend: Americas, EMEA, APAC, Growth Rate

The investments in sustainability are evident: carbon emissions have a CAGR of 14.2% over the 2023-2028 period, compared to an energy growth CAGR of 19.5% in that same period.

Despite energy efficiency improvements and investments in renewable technology, carbon emissions are still expected to record a CAGR of 11.7% .by 2028
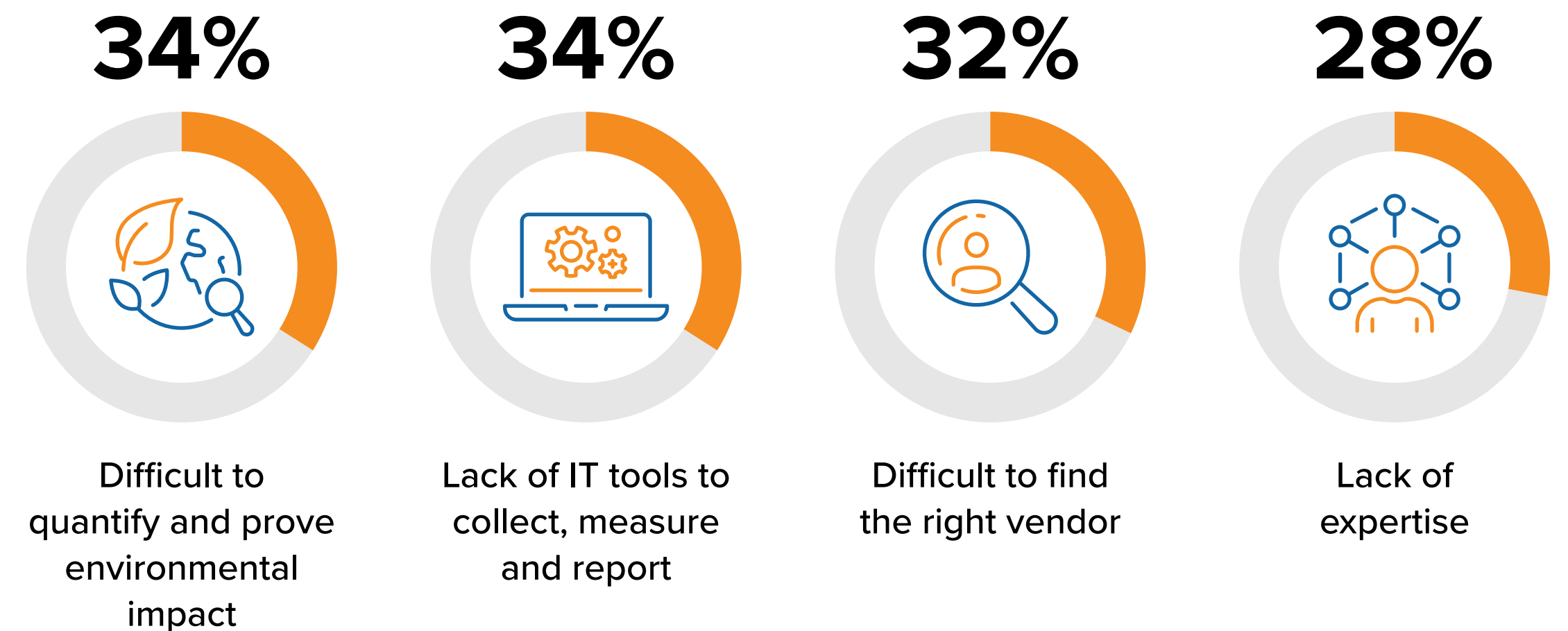
While the desire to reduce costs and become more sustainable is pervasive, datacenter operators face many challenges in execution.

Q. What is your company's average power usage effectiveness (PUE) currently, and what do you expect it will be over the next 2 years?
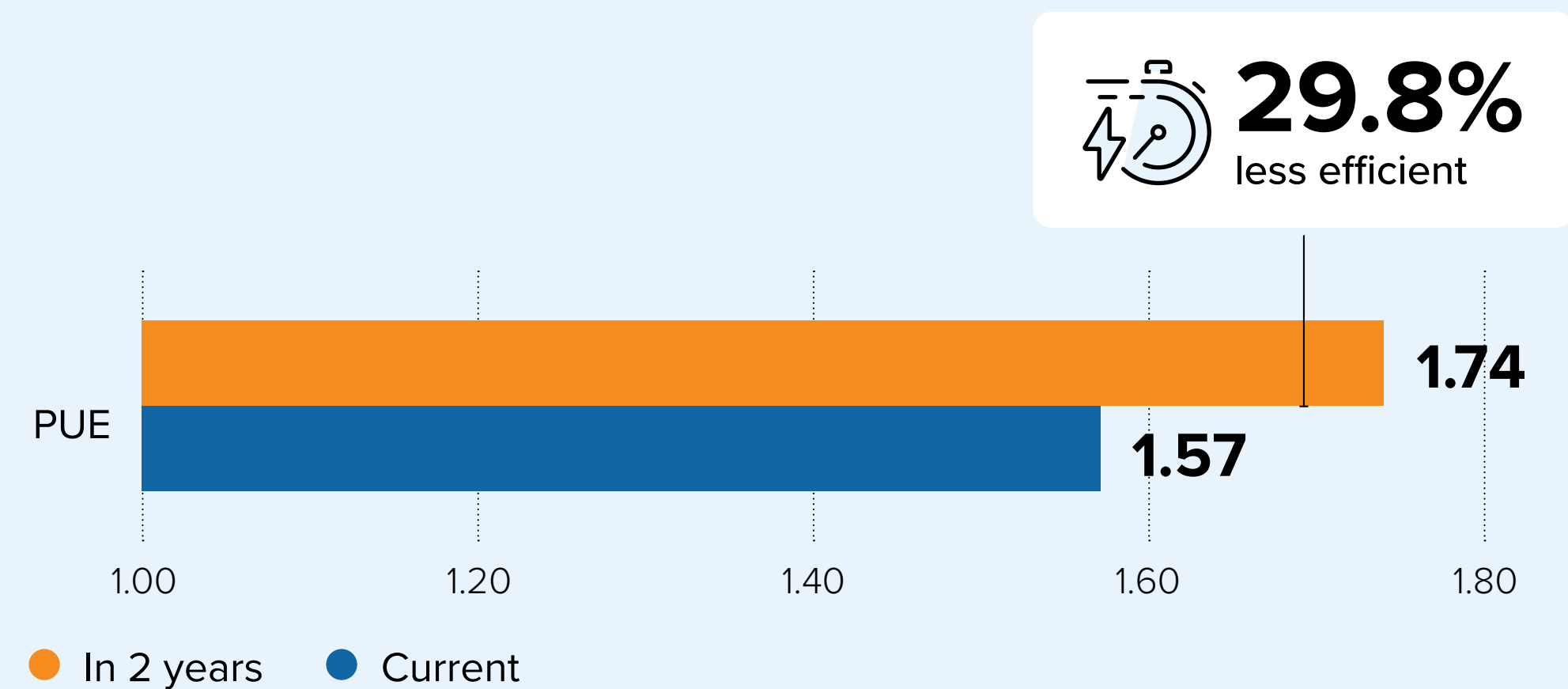
29.8%
less efficient

1.74

PUE

1.57

| 1.00 | 1.20 | 1.40 | 1.60 | 1.80 |

● In 2 years ● Current

Base: Enterprise datacenter respondents (n = 510)
Source: Datacenter Operations and Sustainable Survey, IDC, March 2024

**Q. What are the biggest challenges you face in implementing sustainability measures?**

34%

Difficult to quantify and prove environmental impact

34%

Lack of IT tools to collect, measure and report

32%

Difficult to find the right vendor

28%

Lack of expertise

While organizations value sustainability, they find implementation challenging and remain skeptical about achieving future energy efficiency gains due to high costs and uncertain technological advancements.

Stranded capacity also affects costs and sustainability.

Q. What is the current average power utilization/consumption as a percentage of capacity in your datacenters?
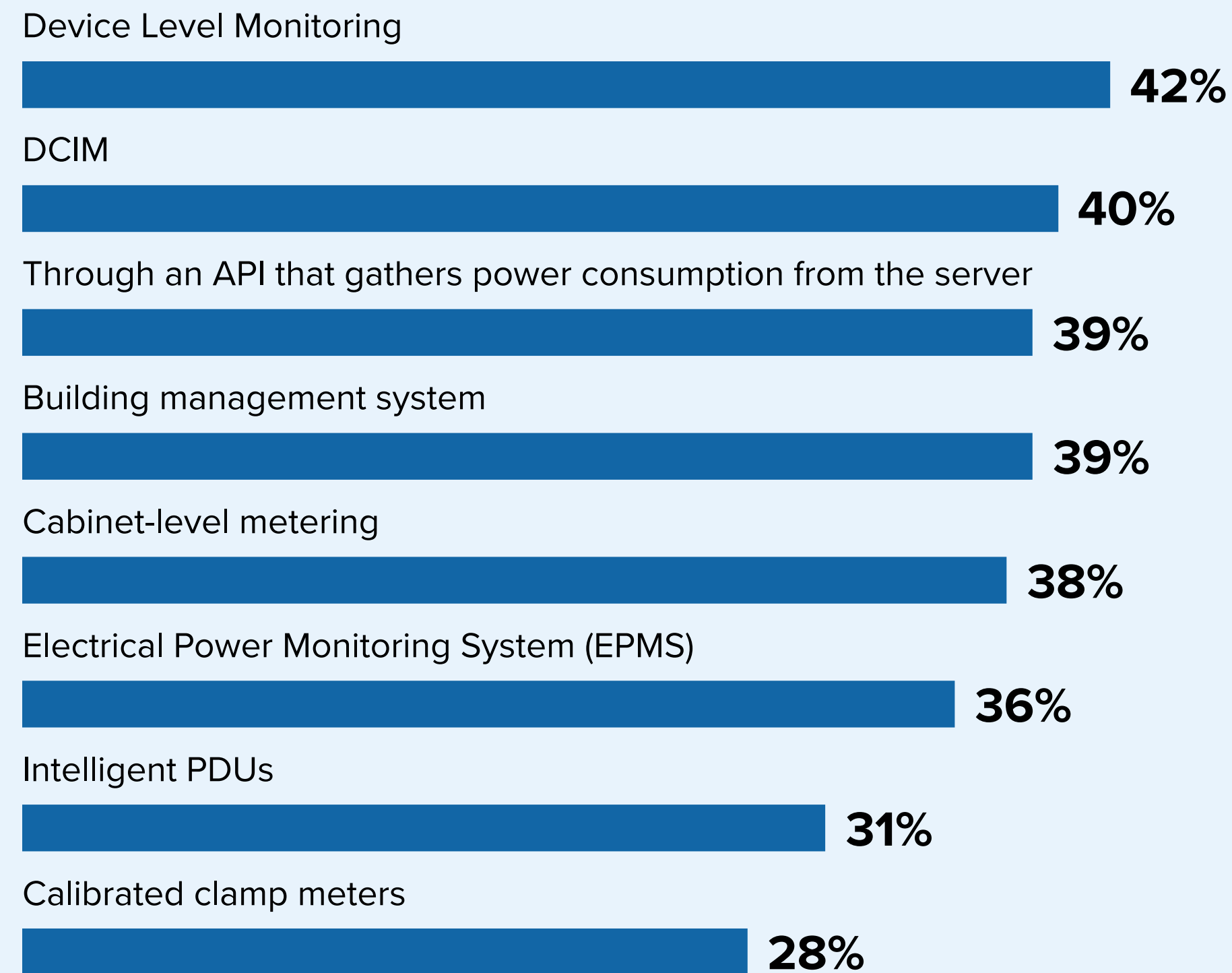
43%
Utilization
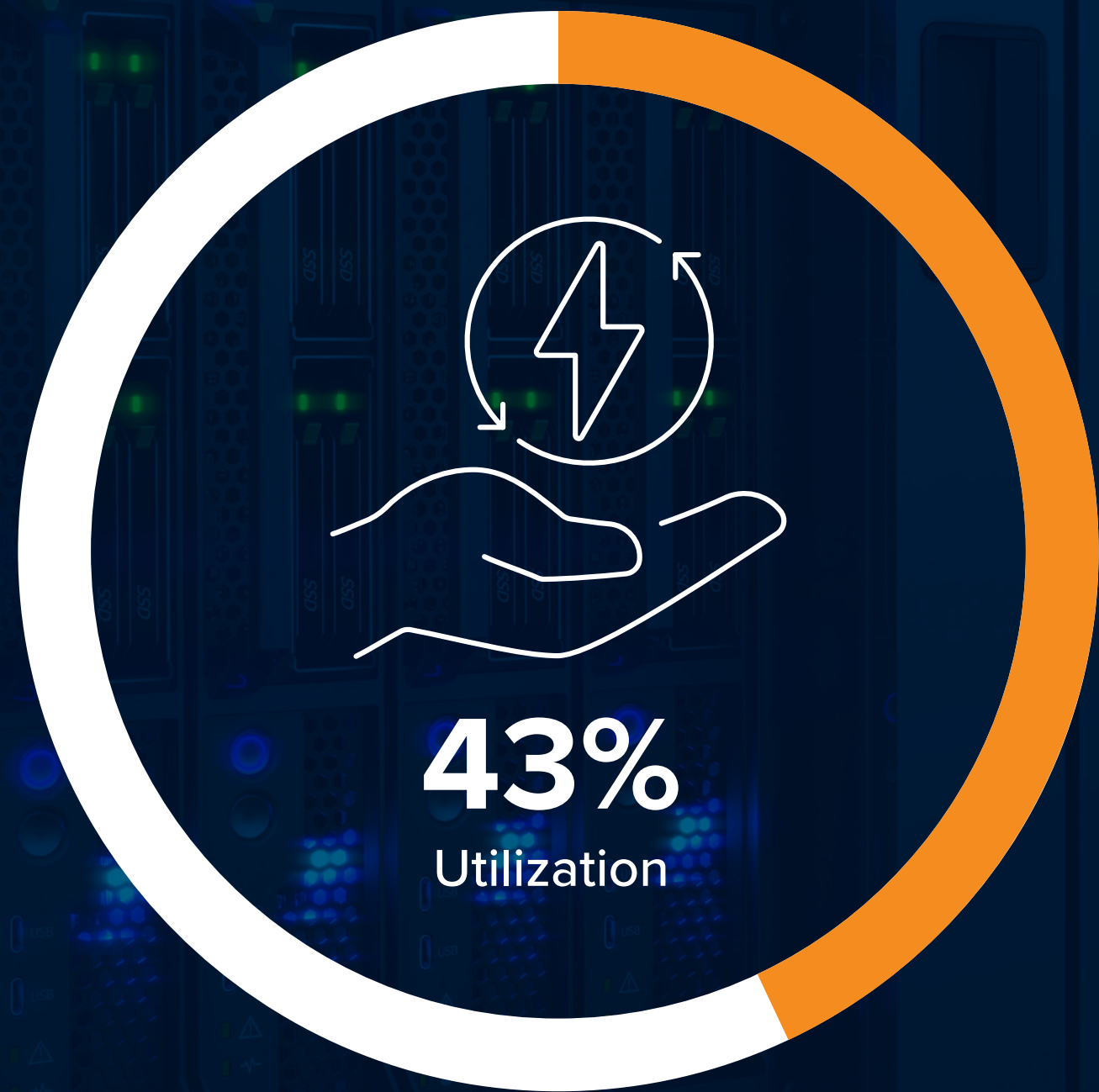
Stranded capacity, caused by overprovisioning, inefficient power distribution, unbalanced cooling, and space limitations, is the excess power available that is not being put to use, which increases costs, decreases efficiency, and limits scalability.

To address the problem, organizations should improve their capacity planning, optimize power distribution, match cooling capacity with IT load, optimum size IT equipment, and implement datacenter management tools.

Base: Enterprise datacenter respondents (n = 510)
Source: Datacenter Operations and Sustainable Survey, IDC, March 2024

The tool paradox: Too many monitoring tools result in a lack of information and efficiency.

Q. How do you monitor and measure power consumption in your datacenters?

Device Level Monitoring
42%

DCIM
40%

Through an API that gathers power consumption from the server
39%

Building management system
39%

Cabinet-level metering
38%

Electrical Power Monitoring System (EPMS)
36%

Intelligent PDUs
31%

Calibrated clamp meters
28%

## Benefits of a single source of truth

Improved decision-making: Everyone in the organization is basing their choices on the same verified information, thereby reducing confusion and ensuring strategic alignment.

Enhanced efficiency: No time is wasted searching for or verifying data from scattered sources.

Reduced errors: The possibility of basing decisions on inaccurate or outdated information is eliminated.
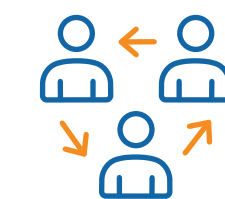
Stronger collaboration: Teams can work together more effectively when they have a shared understanding of the situation.

Increased accountability: With a clear reference point, everyone is held to the same standard.

# Hope remains — organizations see the promise of AI and intelligent systems to overcome their challenges.

Q. How important do you consider AI and its derivatives to be for your organization's sustainable transformation journey?

Not important
0%

Not very important
1%

Critical
32%

Very important
45%

Somewhat important
22%
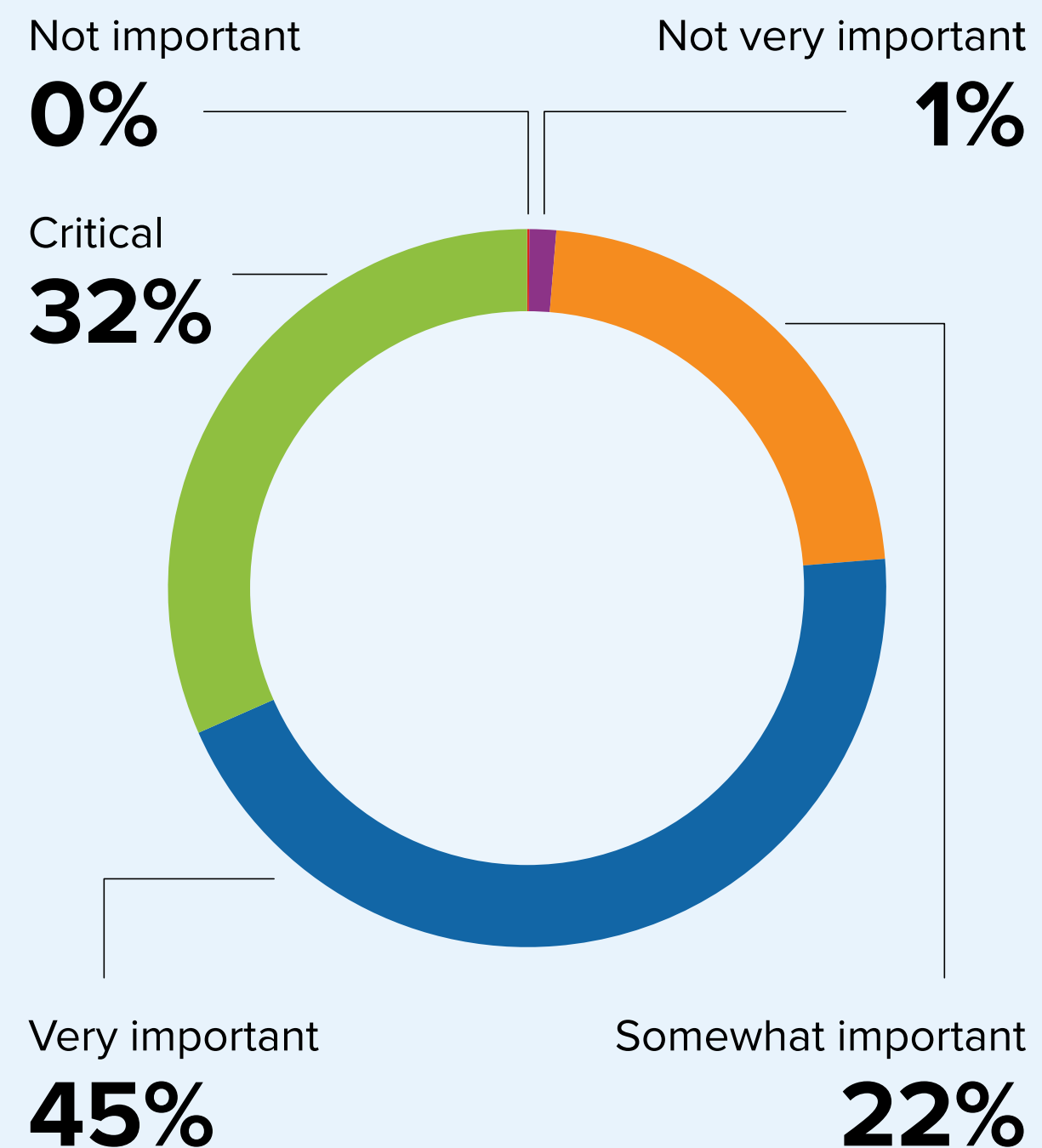
More than
75% of organizations see AI as critical or very important for their sustainability journey.

With datacenters, operators envision intelligent datacenters as the future of the industry and are anticipating the vendor community to innovate so it is commonplace.

Q. What is your company's datacenter workload and operations control and monitoring approach?

58%
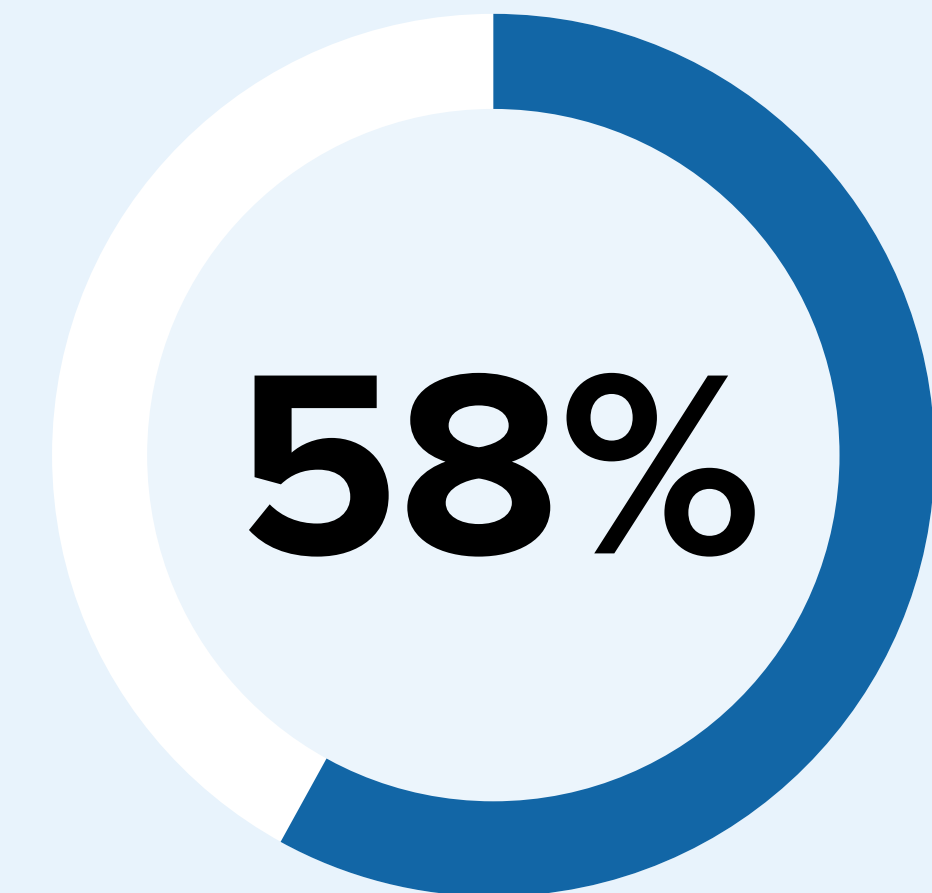
#1 response Intelligent datacenter as the industry's future, and this will be "table stakes" in the next two years.

Organizations should quantify the financial and environmental impact of applying AI for datacenter efficiency.

Example: Evaluation of the expense and carbon emissions of a 1-megawatt datacenter with 50% IT utilization with an industry average PUE of 1.57 two years from now

- Status quo (do nothing); PUE grows to 1.74[1]

- 15% PUE decrease to 1.33 by applying AI to datacenter facilities[2]

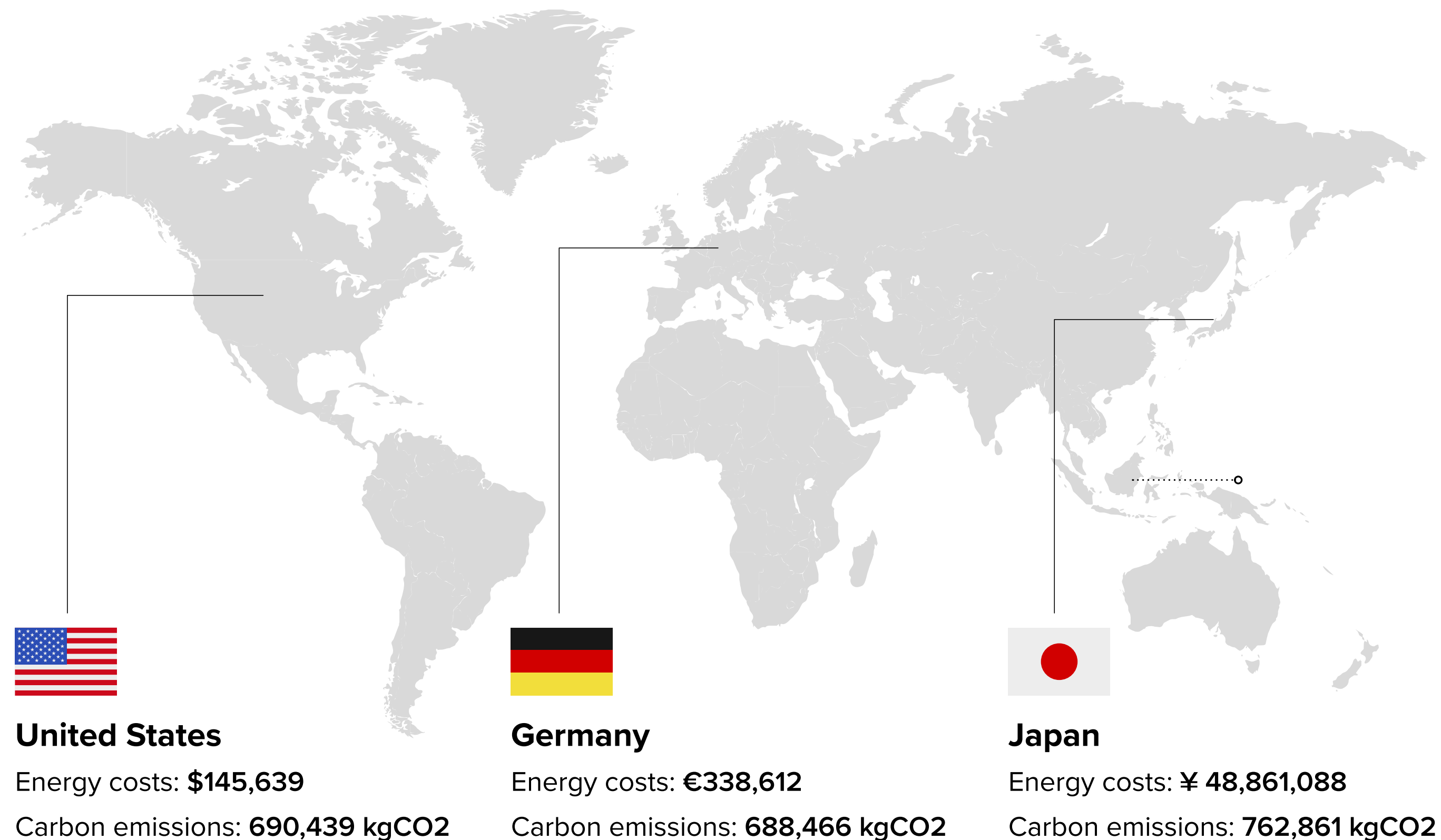| Scenario	Energy (kwh)
Status quo datacenter energy consumption (PUE of 1.74)	7,621,200
Datacenter Energy consumption with AI in the datacenter (PUE of 1.33)	5,845,110
Annual Energy Savings	1,776,090

1. Source: Datacenter Operations and Sustainable Survey, IDC, March, 2024

2. Source: DeepMind AI Reduces Google Data Centre Cooling Bill by 40% - Google DeepMind

3 Source: Industry end-user prices for electricity in selected countries, 2019-March 2023 – Charts – Data & Statistics – IEA

4 Source: 2023_07_international_factors_release_11.xlsx (live.com)

Annual savings for select countries[3,4]

United States
Energy costs: **$145,639**
Carbon emissions: **690,439 kgCO2**

Germany
Energy costs: **€338,612**
Carbon emissions: **688,466 kgCO2**

Japan
Energy costs: **¥ 48,861,088**
Carbon emissions: **762,861 kgCO2**

Recommendations for Datacenter Operators

Understand the current environment. Start by understanding your existing datacenter energy consumption, energy costs, and carbon emissions. This will establish a baseline for forecasting the benefits and measuring the energy and carbon efficiency of moving to public cloud, helping you gauge potential reductions.

Implement a comprehensive datacenter monitoring and management system. Track KPIs related to energy consumption, temperature, water usage, and cooling efficiency and use design tools to help unlock stranded capacity to address power scarcity.
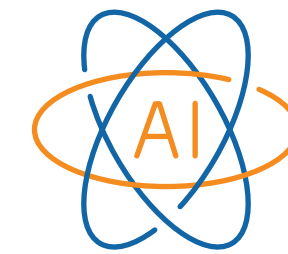
Implement AI for datacenter efficiency. Datacenter operators should leverage AI algorithms to monitor and manage energy consumption. AI can predict demand patterns, optimize cooling, and dynamically adjust power usage to match the IT load while reducing operational risk. Organizations should consider proven solutions to speed time to value and maximize efficiency.

Choose projects with a short-term ROI. Projects with a quick payback period focused on automation and AI are more likely to be funded, and the resulting savings will fuel future investments.

Message from the Sponsor

EkkoSense: Taking your data center optimization to the next level

Managing the transition to GPU-intensive AI workloads and associated liquid cooling requirements is challenging enough for data center teams — and that's before you factor in the pressure to cut energy and carbon usage as part of corporate ESG initiatives.

Resolving this requires absolute real-time white space visibility. EkkoSense, with its award-winning AI-powered EkkoSoft Critical 3D visualization and analytics SaaS solution, resolves this in a light-touch, vendor-agnostic way. Unlike traditional DCIM approaches, the advantages for data center operations teams are clear. EkkoSense benefits include identifying cooling, power and capacity inefficiencies, unlocking cooling energy and carbon savings, and freeing capacity to support additional workloads.

With EkkoSense you can take your data center optimization to the next level, bringing together your IT and M&E spaces, and making the invisible visible. All with an exceptional ROI – often of less than 12 months. Find out more and see a demo of EkkoSoft Critical in action at **www.ekkosense.com**

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets.

With more than 1,300 analysts worldwide, IDC offers global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries. IDC's analysis and insight help IT professionals, business executives, and the investment community to make fact-based technology decisions and to achieve their key business objectives.

Founded in 1964, IDC is a wholly-owned subsidiary of International Data Group (IDG, Inc.), the world's leading tech media, data, and marketing services company.

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. As a premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets, IDC's Custom Solutions group helps clients plan, market, sell, and succeed in the global marketplace. We create actionable market intelligence and influential content marketing programs that yield measurable results.

IDC UK
5th Floor, Ealing Cross, 85 Uxbridge Road, London, W5 5TH, United Kingdom
T 44.208.987.7100

X @idc   in @idc   idc.com

Privacy Policy | CCPA