uptime INTELLIGENCE

Uptime Institute Global Data Center Survey 2025

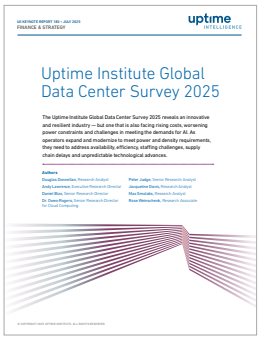
The Uptime Institute Global Data Center Survey 2025 reveals an innovative and resilient industry — but one that is also facing rising costs, worsening power constraints and challenges in meeting the demands for AI. As operators expand and modernize to meet power and density requirements, they need to address availability, efficiency, staffing challenges, supply chain delays and unpredictable technological advances.

Authors

Douglas Donnellan, Research Analyst
Andy Lawrence, Executive Research Director
Daniel Bizo, Senior Research Director
Dr. Owen Rogers, Senior Research Director for Cloud Computing
Peter Judge, Senior Research Analyst
Jacqueline Davis, Research Analyst
Max Smolaks, Research Analyst
Rose Weinschenk, Research Associate

Key points

- Cost issues remain the top concern for digital infrastructure management teams in 2025 — but concerns about forecasting future capacity requirements have grown significantly.
- Average PUE levels show little change for the sixth consecutive year, with improvements constrained by legacy infrastructure and region-specific barriers to efficient cooling.
- Average server rack power densities continue to rise slowly, driven by greater adoption of racks in the 10 kW to 30 kW range. Few facilities exceed 30 kW, and extreme densities remain rare.
- The collection and reporting of key sustainability metrics have not improved in 2025 — due, in part, to rising power demand and easing regulatory pressure in some regions.
- Trust in AI for data center operations depends on the use case: most operators would allow its use for analyzing sensor data and predictive maintenance tasks, but not configuration changes, controlling equipment or managing staff.
- Impactful data center outages are gradually becoming less frequent — but one in 10 outages still cause serious or severe disruption, underlining the need for continued investment.
- Enterprises continue to adopt hybrid IT strategies, but on-premises data centers remain foundational, with 45% of IT workloads still residing in corporate facilities.
- Staffing challenges persist in 2025. Nearly two-thirds of operators report difficulty retaining staff, finding qualified candidates — or both.



Contents

Uptime Institute Intelligence is an independent unit of Uptime Institute dedicated to identifying, analyzing and explaining the trends, technologies, operational practices and changing business models of the mission-critical infrastructure industry. For more about Uptime Institute Intelligence, visit uptimeinstitute.com/ui-intelligence or contact research@uptimeinstitute.com.

Figures

Introduction

The 15th annual Uptime Institute Global Data Center Survey is the most comprehensive and longest-running study of its kind. The survey reveals the state of the industry in terms of resiliency, sustainability, efficiency, staffing, cloud and AI.

The survey was conducted online from April 2025 to May 2025 and collected responses from more than 800 data center owners and operators, and over 1,000 vendors and consultants. This report focuses on owners and operators of digital infrastructure — an analysis of the experiences and views of vendors and consultants will be published separately. For more details, including demographics, see the **Appendix**.

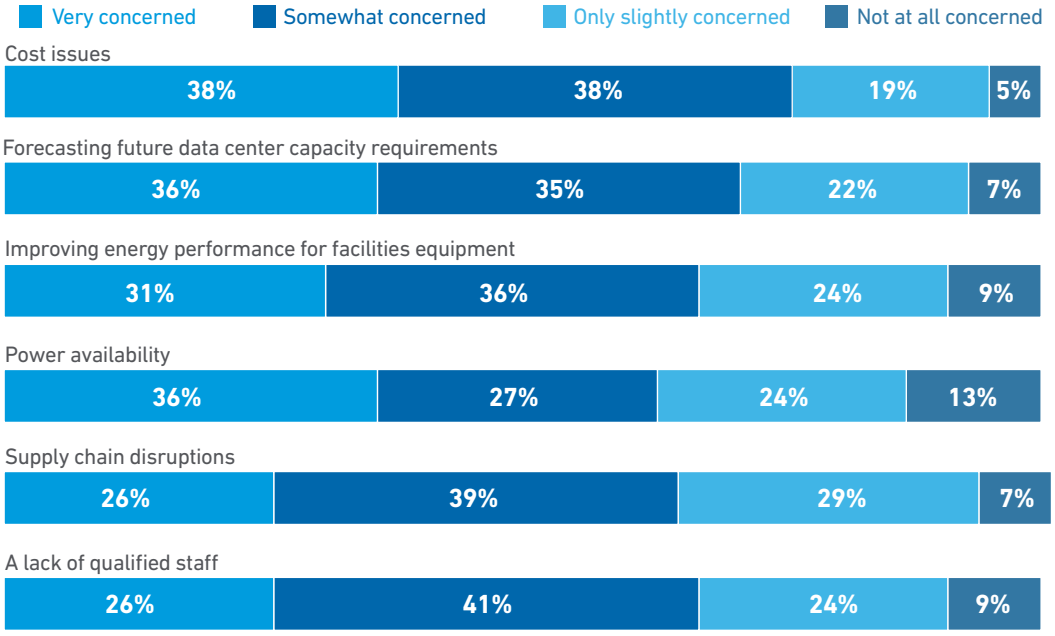
For the third consecutive year, Uptime's survey asked data center operators to identify their management team's top concerns related to digital infrastructure. In 2025, new response options were added to reflect the evolving challenges surrounding power availability, supply chain disruptions and demand for AI.

Costs remain the leading concern, although by a narrower margin than in 2024 (see **Figure 1**). At the same time, the share of operators whose management is "very concerned" about forecasting future data center capacity requirements increased by a significant nine percentage points since 2023. These shifts highlight the uncertainty facing operators in 2025 as they balance rising costs — driven by inflation, labor shortages and energy prices — with increasingly complex capacity planning shaped by market volatility, geopolitical instability and ongoing supply chain disruptions.

Figure 1

Costs and capacity planning are top management concerns in 2025

Looking at the next 12 months, how concerned is your digital infrastructure management regarding each of the following issues? (n=638)



(Only the top six response categories are shown.)

A growing interest in AI workloads adds further complexity. Although approximately one-third of data center owners and operators currently perform some AI training or inference, a significantly greater proportion plans to do so in the future (see *Data center AI strategies are mixed in early 2025*). Uncertainty about the appropriate or likely venues for AI workloads, and apprehension about the power demands of projected Nvidia GPU systems, is likely to be contributing to capacity concerns.

This uncertainty leaves many operators grappling with how the impact of AI will align with business goals, how much infrastructure demand will increase, and whether existing infrastructure can support higher power and cooling requirements. It is likely that, because AI is still in the exploratory or early planning phase for many operators, two of the new response options relating to integrating AI software/hardware and meeting AI infrastructure needs did not rank among the top six management concerns.

Response options relating to power availability and supply chain issues, which were also introduced in 2025, are nearly as concerning to management as improving energy performance for facilities infrastructure. Whether for AI adoption or general capacity expansion, securing adequate electrical power and timely access to infrastructure hardware are becoming more significant barriers to growth.

The remaining response options, staffing shortages and improving energy performance for IT, have seen little change from previous years. Roughly two-thirds of operators say management remains at least somewhat concerned about both issues, indicating their continued relevance to operational planning.

Industry benchmarks

The 2025 Uptime Institute Global Data Center Survey marks 15 years of tracking key high-level operational and design metrics from across the data center industry. Respondents shared information on their facility energy performance, equipment redundancy and rack power density. Although these metrics each have limitations and are best understood in context, data from our large sample highlights broader trends across the data center landscape.

In the next few years, deployments of generative AI infrastructure will place new demands on facility design. Engineering choices to meet these demands will affect energy efficiency, density and equipment redundancy — for a growing, but currently relatively small, number of data centers. The 2025 data also indicates that while densified IT hardware is beginning to reshape the data center landscape, the influence of densified IT infrastructure remains moderated by the large and still growing installed base of more conventional IT.

PUE improves beneath the surface

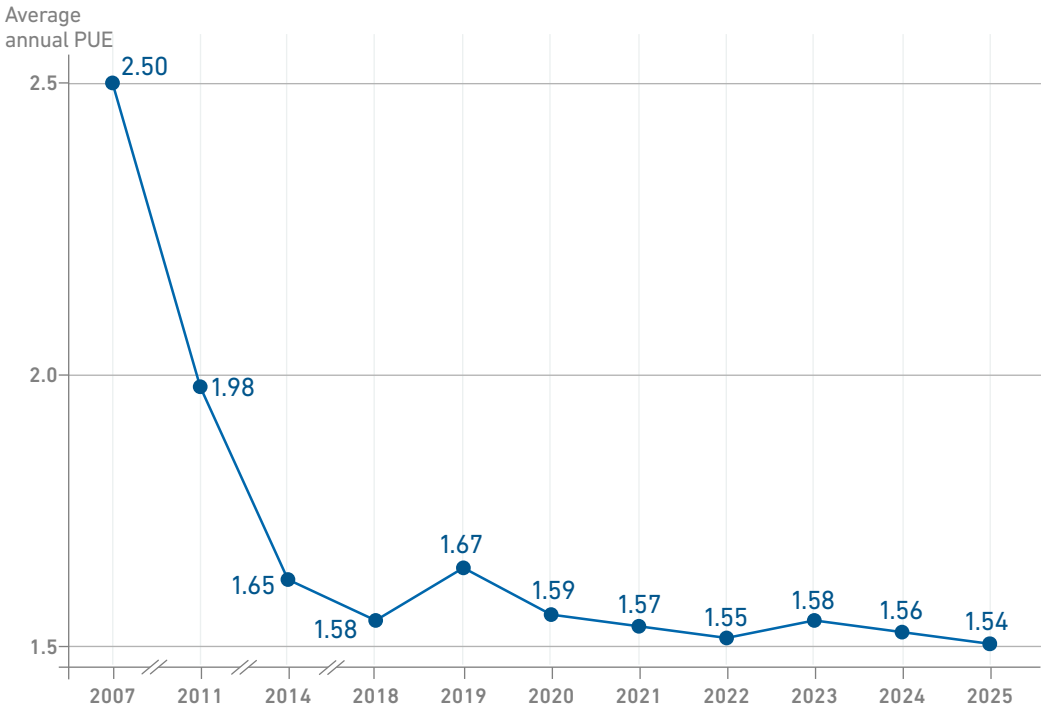
Introduced by The Green Grid in 2007, PUE is a metric operators use to track a facility's energy efficiency over time. PUE is calculated by dividing total facility power by the power consumption of IT equipment. PUE calculation is straightforward, which helped make it common practice — although it excludes important elements of efficiency from its scope, such as facility water use or IT efficiency.

Uptime's annual data center survey has collected PUE data from a large, diverse sample of operators since the metric's introduction, charting a facility energy efficiency trend for the industry. In 2025, respondents' annual PUE had a weighted average of 1.54 (see **Figure 2**), marking the sixth consecutive year that this headline figure has virtually stood still. The single average cannot capture diversity in the sample, but a closer examination of facility groupings by size, age, and region reveals variation between these categories.

Figure 2

Weighted average PUE hovers above 1.5

What is the average annual PUE for your data center? (n=681)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

The industry's average dropped quickly — from 2.5 in 2007 to 1.65 in 2014 — as operators first deployed the most practical and economical changes. Operators favored upgrades to older electrical systems, better airflow management, and optimized cooling system controls because these changes were often straightforward and less disruptive to live workloads — and paid dividends in cost savings, as well as facility efficiency.

In existing facilities — especially in older legacy data centers — further upgrades often present a much weaker business case. High demand still incentivizes operators to run their data centers for longer, which is why legacy facilities remain common: over a quarter (27.5%) of respondents report working in data center facilities that are 16 years or older. New data centers, however, present opportunities to optimize for efficiency and include efficient heat rejection systems early in the design process, allowing them to achieve better-than-average PUEs.

Although the weighted average PUE has remained relatively stable for several years, the outlook is improving. Facilities commissioned within five years of the survey (from around early 2020) achieved an average PUE of 1.48. Notably, larger data centers (20 MW and above) averaged a PUE of 1.44 globally — in part, because the largest facilities tend to be newer. In addition, larger data centers can often justify significant investment in efficient power and cooling infrastructure, which benefit from economies of scale.

Climatic conditions constrain operator selection of cooling systems, making a data center's region an important consideration when evaluating its PUE. Many new facilities in high latitudes in North America and Europe have PUEs of 1.3 or better. This year, 15% of respondents report their PUE in this range. Uptime expects the share of more efficient facilities to grow as higher temperature set points take effect, particularly with the increased adoption of liquid-cooled IT hardware.

There appears to be no hard boundary that would prevent global average PUE from finally dropping below the 1.5 mark, as many facilities have already achieved this. However, the average is likely to move slowly (if at all) over the next few years. Many legacy facilities will continue to operate, and climatic conditions will mean the most energy-efficient heat rejection systems will remain untenable in some regions. PUE could make incremental gains as construction of large new data centers continues, incorporating efficiency optimizations and economies of scale — provided these gains are not offset by development in more challenging climates. Additionally, in several major regions, new regulations are mandating lower design and operational PUEs — forcing some older data centers to either close or undertake major retrofits.

Rack densities on the move

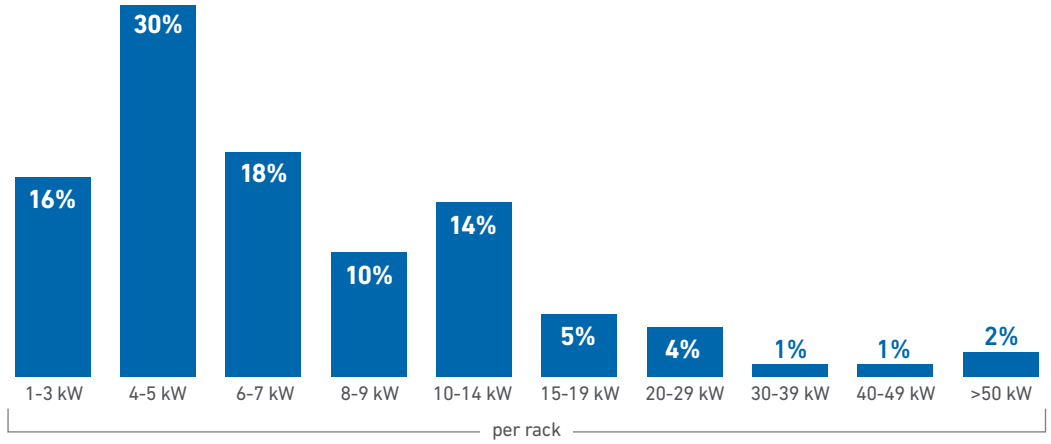
In last year's survey (2024), Uptime observed that typical (modal) rack power densities had shifted visibly, albeit gradually, towards higher bands. This trend continues in 2025 (see **Figure 3**), largely in the form of strong adoption of racks between 10 kW and 30 kW. The average of the modal densities in our total sample reaches almost 9 kW, up slightly from 8.3 kW last year.

However, a small number of high-density facilities (which we define as having typical density of 30 kW or above) distort the average figures. Average modal density excluding these outliers comes to 7.5 kW in our 2025 survey, up from 6.8 kW in 2024. Directionally, the trend is the same.

Figure 3

Most operators still see single-digit modal average density

What is the most common (modal average) rack density deployed in your data center? (n=710)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Given the sharp rise in per-server and per-CPU socket power consumption, this trend is not surprising. When all systems are highly exercised, it does not take many servers (along with storage shelves and top of rack switches) to cross the 10 kW mark. Many mainstream servers now approach (or even surpass) 1 kW of power when richly configured to drive application performance or support heavy workload consolidation. Uptime expects this trend to continue as fleets of servers are refreshed with newer, more performant and efficient, but also more power-hungry, hardware.

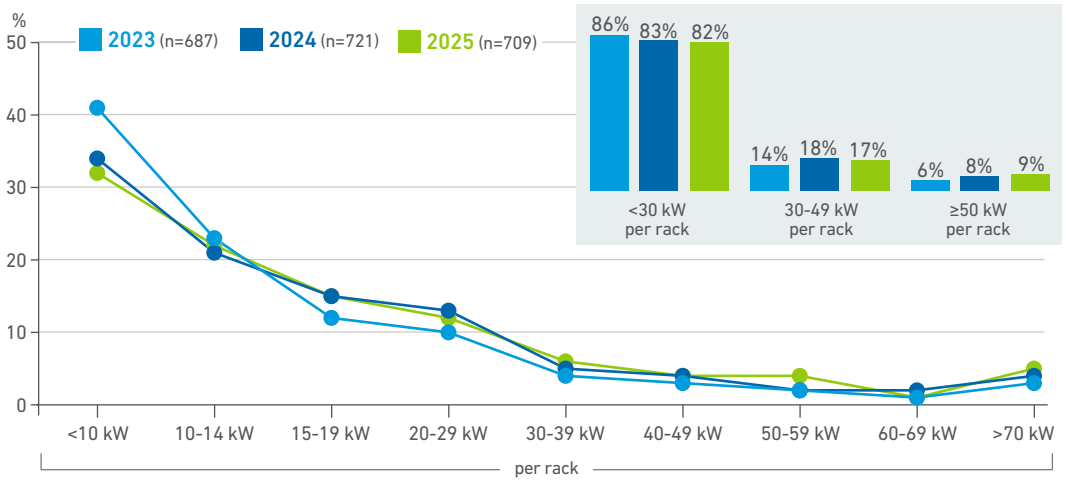
The 2025 survey results also indicate that densification is not occurring everywhere at the same pace — and in some sectors, is not measurably occurring at all. Many data centers have pressed ahead with densification faster than their peers, flattening the distribution curve — rather than shifting the curve towards higher rack power. This pattern is most pronounced in the North America and Asia-Pacific regions. At the same time, low-density facilities (under 4 kW per rack average) remain common even in those geographies, with China being the sole exception.

Perhaps counterintuitively, the highest density racks have seen little change. More than 80% of the operators responding to our survey say their facility has no racks above 30 kW — about the same share as last year (see **Figure 4**). This suggests that high-performance computing (HPC) systems, including AI training compute, are concentrated in relatively few sites. Around one in eight facilities reports having some dense racks in the 30 kW to 59 kW range. Our survey identified even higher densities, including some cabinets exceeding 100 kW, but these are still rare.

Figure 4

Most operators' highest density racks still peak below 30 kW

What is the highest server rack density deployed in your data center?



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

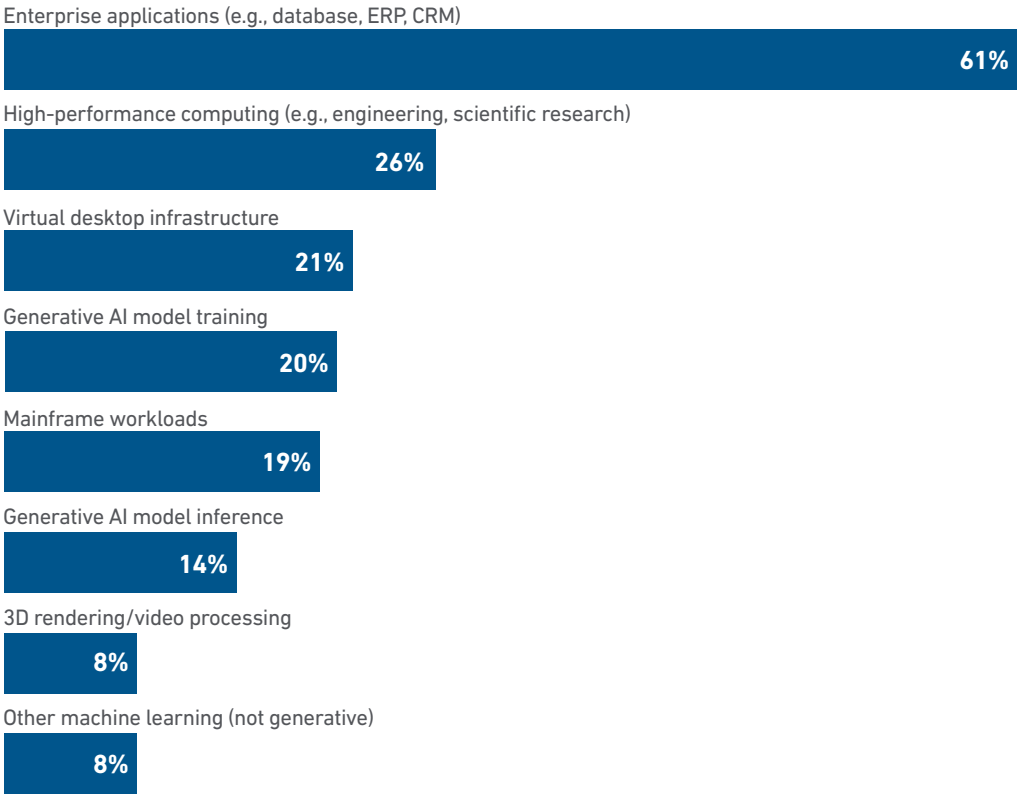
The workloads that drive the installation of the highest density racks are diverse (see **Figure 5**). As expected, the most common applications in the 10 kW to 30 kW band are enterprise software. Examples include databases, enterprise resource planning and specialist enterprise workloads, such as mainframe applications and virtual desktop infrastructure. Furthermore, one in four respondents already run HPC or AI compute at these density points.

The distribution of workloads starts to shift markedly above 30 kW towards more HPC or AI, but it is only above 60 kW per rack that these computationally intense workloads take primacy over enterprise software as the main driver of densified infrastructure.

Figure 5

High-density workloads are diverse

Which of the following types of applications are driving the highest density deployments in your data center? Choose all that apply (n=685)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Sustainability and metrics

After years of gradual increases, Uptime's annual survey results show that sustainability data collection for nearly all metrics has stagnated or even declined since 2024. While PUE and energy consumption remain widely tracked, the broad pattern is static (see **Figure 6**). This development is somewhat unexpected.

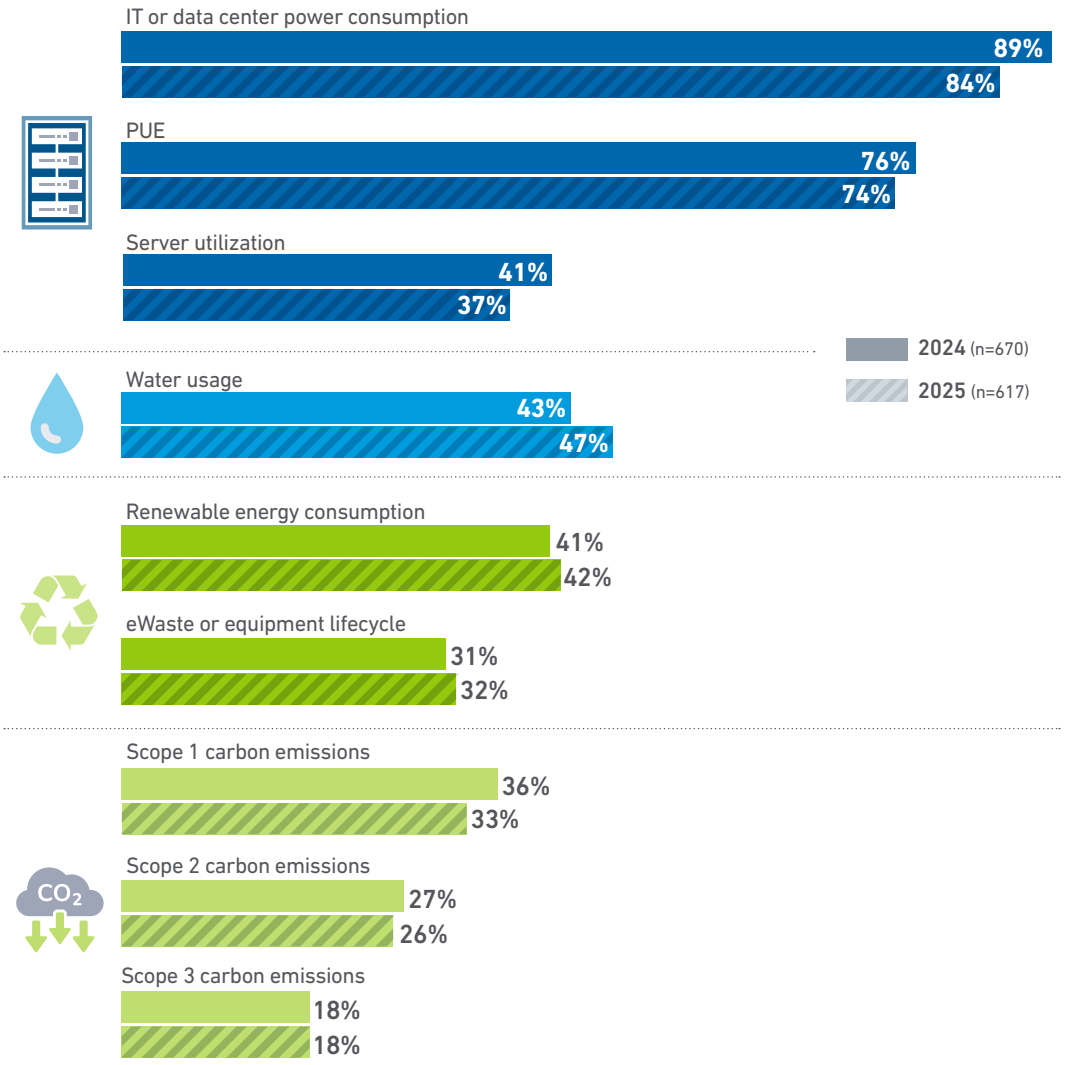
The only metric to show a measurable (but slight) increase in collection is water usage, led by a 5 percentage point increase in Europe, presumably motivated by the EU's Energy Efficiency Directive (EED), which requires reporting of water usage.

Energy consumption and PUE, the most commonly tracked efficiency/sustainability metrics, remain widely reported at 84% and 74% of respondents, respectively. This is because PUE and energy use are easy to collect, often required by regulators or customers, and are directly tied to operational costs. However, even these popular metrics have declined slightly from 2024 levels (89% and 76%, respectively). The drop in energy data collection is especially notable in North America, where it fell by nearly 9%.

Figure 6

Sustainability reporting falters

Which of the following IT or data center metrics does your organization collect for corporate sustainability purposes? Choose all that apply.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

More laissez-faire governance regimes in some regions are undoubtedly contributing to this decrease in data collection. For example, the US Securities and Exchange Commission abandoned efforts to require climate risk reporting — including greenhouse gas (GHG) emissions — early in 2024 (see *Should operators continue to prepare for climate risk reporting?*).

Meanwhile, with its Omnibus simplification measure, the European Commission has delayed GHG reporting under the Corporate Sustainability Reporting Directive and intends to exempt many organizations (those with fewer than 1,000 employees) from its scope (see *EU climate reporting: simplification is not simple*). However, while these may play a role, there are counter trends and regulations in many geographies that would have the opposite effects — especially in the EU.

In 2025, there has been high-profile criticism of sustainability measures from some governments. Sustainability data collection has declined slightly more in the US than in other regions, but the shifts are small, and it is too early to attribute the trend to any specific political change.

Practical considerations may be a stronger obstacle to data collection. Large expansion projects, such as AI data centers in development or recently commissioned, have high resource consumption. This makes sustainability goals more challenging to achieve and reduces operators' appetite for collecting and reporting the data.

Apart from PUE and energy usage, less than half of survey respondents track any other key sustainability metrics. Renewable energy usage and water consumption data are required in Europe under the EED; reporting levels in Europe are currently slightly higher than in other regions. These levels are expected to rise as the EED becomes more fully implemented. Worldwide, data collection for renewable energy usage and water consumption remains low, at 42% and 47%, respectively.

Water usage is the only metric where the percentage reporting data grew, increasing by 4%. Although most data centers in the Northern Hemisphere do not consume a significant amount of water compared with other industrial sites, water scarcity remains a growing public concern. Uptime Institute's Cooling Systems Survey 2024 revealed that cooling systems consuming water are prevalent worldwide (see *Water is a local issue: site selection and facility design*). As more data centers larger than 100 MW are built, water usage is likely to become a contentious issue in areas that are currently water-stressed (or may become so due to climate change).

With increasing scrutiny of water-intensive cooling systems, especially in arid regions, collecting water usage data is a sensible measure to anticipate rising stakeholder expectations — even in jurisdictions without formal mandates.

GHG and renewables reporting remain static

Carbon emissions are the primary cause of climate change and reducing them begins with tracking key metrics: renewable energy usage (since energy consumption is generally the most significant source of data center carbon emissions) and GHG protocols.

Renewable energy reporting is effectively flat this year, after a 7% increase in the collection of renewable energy consumption data between 2023 and 2024. This lack of enthusiasm may be due to the difficulty of procuring renewable energy and reporting its usage accurately.

To reduce the carbon footprint of their data centers, many operators use energy attribute certificates (EACs) or renewable energy certificates (RECs). The effectiveness of these methods to offset the carbon intensity of electricity useage is debatable, and the offsets are also becoming increasingly expensive.

Uptime's data from previous years show that many organizations are either rolling back targets or delaying net-zero deadlines. Given that many data center companies rely heavily on offsets to achieve these goals, these higher costs may have led some operators to quietly retire net-zero targets or to stop tracking and reporting emissions entirely.

Fewer organizations collect carbon emissions under the GHG protocols than any other metric. In 2025, reporting is static for all three categories: Scope 1 (direct emissions), Scope 2 (emissions embedded in electricity) and Scope 3 (emissions attributable to the supply chain and embedded in products). From 2023 to 2024, Scopes 1 and 2 rose by 8% and Scope 3 by 4%.

Collecting data on GHG emissions and renewable energy is challenging, but the task remains worthwhile. The need for reducing emissions is clear and, as a step in that direction, pressure to report data will continue from many stakeholders — including governments.

Reporting legislation is increasing

Despite the publicity surrounding measures, such as the US government's policy changes and the EU's simplification of regulations, the number of sustainability standards requiring reporting has increased during 2025 (see *Data center sustainability standards grow globally*), with many modeled (at least partly) on the EED.

China and Japan have introduced minimum performance standards for PUE. In the US, despite federal disavowal of climate change measures, state-level regulations continue, with California's Senate Bills 253 and 261 set to require GHG emissions reporting. Other countries, including Brazil and the UK, are aligned with reporting standards based on the International Sustainability Standards Board (ISSB).

With increased awareness of the role of digital infrastructure, several governments are classifying data centers as critical national infrastructure (CNI). While more focused on resiliency, CNI status will inevitably require closer alignment with government strategies including climate change mitigation and reporting.

Even in the absence of reporting regulations, some investors, stakeholders and customers are asking operators to report environmental impacts using voluntary platforms, including the Carbon Disclosure Project (CDP) and Global Reporting Initiative (GRI). Data center companies that voluntarily provide sustainability information are better positioned to meet possible future regulatory requirements and may also attract new business from organizations with environmental goals.

At the same time, local resistance to new data center projects is increasing, especially in areas where energy, land or water use is contentious (see *Data center resource use will raise deep questions — and opposition*). Collecting clear environmental data will be crucial to addressing these criticisms and preparing for future sustainability requirements.

Innovation and impact

Three years on since the public release of ChatGPT (in November 2022), AI continues to be the hottest topic among data center investors and many operators, dominating media coverage and conversations at industry events. Adoption of generative and other forms of AI in business applications is accelerating, which in turn has driven a wave of investment in data centers and created an expectation of more AI-based functionality in all facility operations.

For some operators, initial skepticism in the capabilities of AI-based software is being replaced with cautious interest.

AI is primarily an efficiency tool in operations

As the number of AI-based software deployments grows, information about the capabilities and limitations of AI in the workplace is becoming more widely available. More operators are also learning about the challenges and implications of hosting AI as a workload.

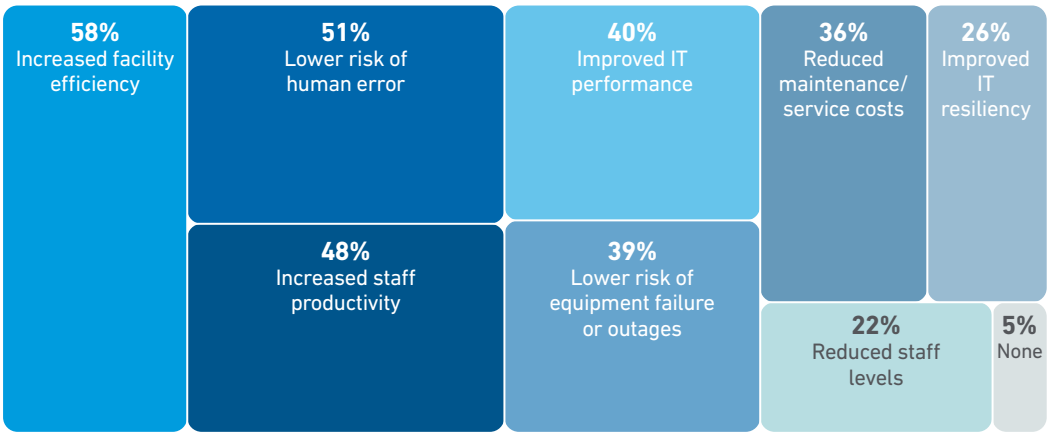
Greater exposure is also fostering better awareness of the many types of AI and their suitability to specific tasks. Some AI approaches are more reliable and transparent than large language model (LLM) systems that dominate the headlines. AI-based products and services for data center management and operational support are available, and most of the industry is entering a period of careful adoption, testing and validation. Data centers are slow and careful in adopting new technologies, and AI will not be an exception.

In 2025, the three primary drivers that motivate AI-based software deployments for use within data center operations remain the same as the previous year: to increase facility efficiency (58%), to reduce human error (51%) and to improve staff productivity (48%); see **Figure 7**.

Figure 7

The perceived benefits of using AI in operations

Which of the following (if any) do you consider benefits of using AI in your data center operations? Choose all that apply. (n=633)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Since 2024, Uptime has added two potential benefits to the option list, which have sparked interest among respondents: to improve IT resiliency (26%) and to reduce average staffing levels (22%). Today's AI technologies, most operators believe, are unlikely to replace the staff required to run data center facilities, setting the industry apart from many others, where this substitution is expected to take place — and respondents' opinions reflect this.

Trust depends on the application

Since 2022, trust in AI as a tool for data center operations has been tracked by Uptime as a binary choice. This year, we have introduced a range of potential data center use cases. The results show trust in AI as a tool varies considerably depending on the specific application of this technology.

In 2025, 73% of respondents say they would trust an adequately trained AI system to run automated analytics on sensor data or to prioritize alarms. However, just 14% would allow these systems to make configuration changes mechanical and electrical equipment (see **Figure 8**).

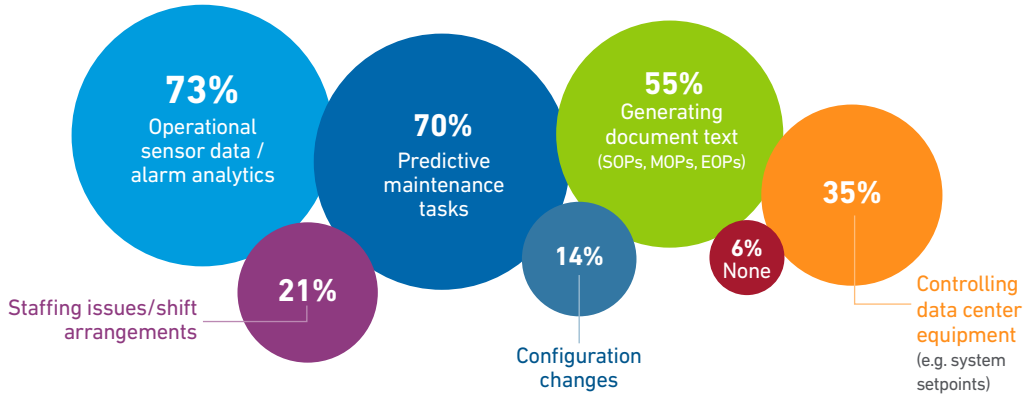
AI-based data center optimization software — an emerging data center management software category with significant potential — can use models to set parameters such as cooling equipment set points within pre-defined ranges. This particular application would be permitted by 35% of respondents.

Only 21% of survey respondents would allow AI systems to create (staffing) shift arrangements — even though it is a task that LLMs appear to be well-suited for, and for which there are over-the-counter software products readily available.

Figure 8

Confidence in AI for operations depends on use case

For which of the following (if any) would you allow AI to make operational decisions in a data center, assuming it has been adequately trained with historical data? Choose all that apply. (n=630)



("Other" responses are not included.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Some operators have started to consider AI-based controls as part of their sustainability initiatives, since the technology can be used to reduce the power consumption of cooling equipment and improve PUE. A few colocation companies are beginning to use AI-based controls as a differentiator to attract customers. However, the majority of respondents continue to believe that AI software capable of controlling equipment poses an unacceptable level of risk and will not use this software in their facilities.

Resiliency and outages

Data center operators continue to prioritize resiliency through ongoing investment in redundancy, maintenance and operational best practices. However, outages remain a persistent challenge — especially as infrastructure becomes more complex and capacity demands increase. External factors, such as extreme weather and grid disturbances or failures, may also have a greater impact than in recent years.

Tracking and analyzing outages are inherently difficult. What constitutes an outage, its severity, and the challenges in identifying its root cause can vary significantly from one site to another. At the same time, outages are often under-reported or disclosed with limited details. This lack of transparency makes it more difficult for industry professionals to learn how best to prevent and mitigate similar failures in the future.

Despite these challenges, Uptime survey results consistently show that impactful data center outages (those significant enough to record, even if the resulting costs were negligible) are becoming less frequent relative to the overall growth in IT. This trend holds true despite the rise in headline coverage of data center failures — a reflection not of worsening performance but of the growing visibility and criticality of digital infrastructure in daily life.

Maintaining this progress will require ongoing vigilance from operators. Aging grid infrastructure, extreme weather events and rapidly increasing workload demands — including those driven by AI — are placing new stresses on both power systems and operational processes. Although distributed and software-defined approaches to resiliency are advancing and show promise, they can also introduce complexity and new risks. As a result, the industry's resiliency gains remain under threat from an expanding range of challenges.

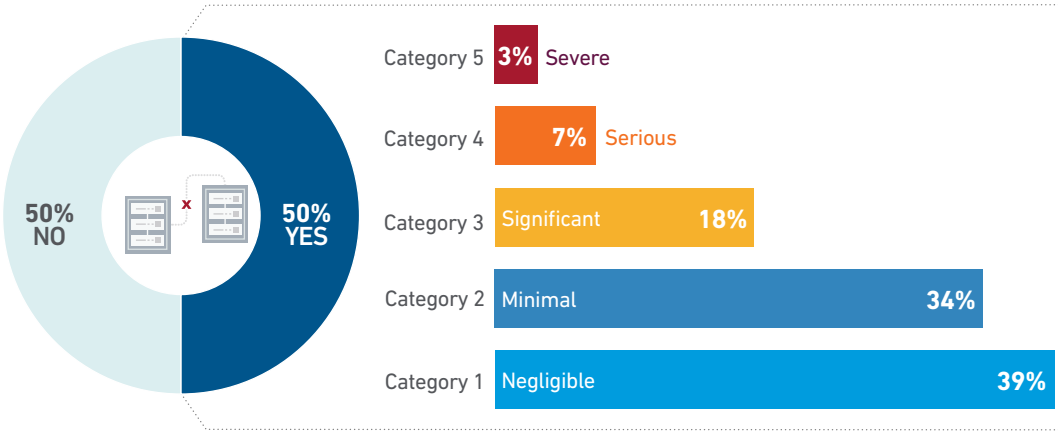
Outage frequency continues to decline

For the fifth consecutive year, Uptime's annual survey results indicate that, on a per-site basis, the frequency of impactful outages is decreasing. However, the pace of improvement has slowed. In 2025, 50% of data center operators reported experiencing at least one impactful outage at their facility over the past three years (see **Figure 9**), which is down from 53% in 2024 (see **Figure 10**).

Figure 9

Half of operators report no outages in the past three years

Has your organization had an impactful outage in the past three years? If so, how would you classify the most impactful outage on a scale of 1 (negligible) to 5 (severe)? (n=754)



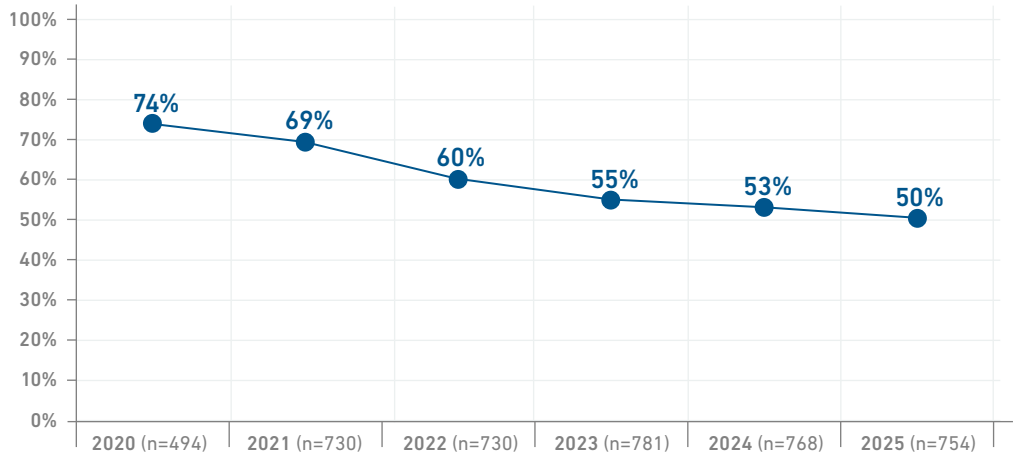
(Responses for "Don't know" are not included.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Figure 10

Gradual improvements in outage frequency continue

Has your organization had an impactful outage in the past three years ("Yes" respondents)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

As noted in the Uptime Intelligence report, *Annual outage analysis 2025*, this slowdown in improvement does not suggest operator complacency. Instead, it reflects a maturing industry facing a more complex risk landscape. As data center services support a broader range of industries, even infrequent failures now affect a greater number of stakeholders.

For operators, the increasing criticality of IT and data centers, a long-standing trend, is paired with another trend: the growing uncertainty around external risks. For example, extreme weather events, geopolitical tensions, and third-party provider and supply chain issues (including power) can be unpredictable and are largely outside of operator control. These challenges may obscure some of the industry's underlying progress — improvements in resiliency are often invisible in public reporting or media coverage.

This may explain why Uptime's annual survey results have shown little change in the reported severity of outages in the past year. One in 10 say their most recent impactful outage was serious or severe — up just one percentage point from 2024 (9%). This may be viewed as success, but one that still requires further improvement; continued investments in redundancy, system monitoring and maintenance, and staff training remain essential.

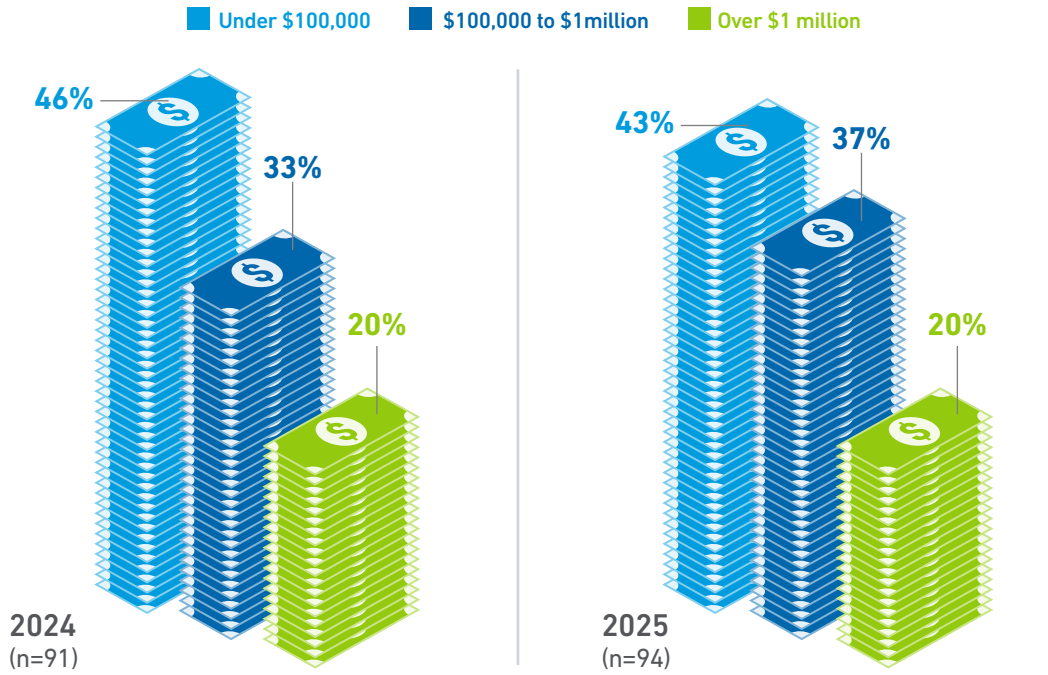
Outage costs rise steadily

The costs of outages are high, and there are signs they may be increasing. For the second consecutive year, one in five respondents who experienced a significant outage report that it cost more than $1 million (see **Figure 11**). Notably, there is a four-percentage point increase in those reporting outage costs between $100,000 and $1 million, suggesting that even moderate outages are becoming more expensive.

Figure 11

Outage costs rise steadily

Please estimate the total cost of this downtime incident (from outage to full recovery), including direct, opportunity and reputation costs using the following options.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Several persistent factors continue to drive these high costs: inflation, rising labor expenses, hardware replacement, penalties for service level agreement (SLA) breaches, and reputational damage. In some regions, regulatory fines and compliance failures — such as those under the EU's Digital Operational Resilience Act (DORA) — add further pressure (see *DORA update: what the EU act means for data centers*).

A key reason costs remain high is the growing criticality of digital services. As digital services become further embedded in everyday life, tolerance for downtime decreases. This growing reliance may lead to a wider "blast radius" when problems strike, with more people and services affected, stronger SLAs, tighter uptime guarantees, and harsher penalties when outages do occur.

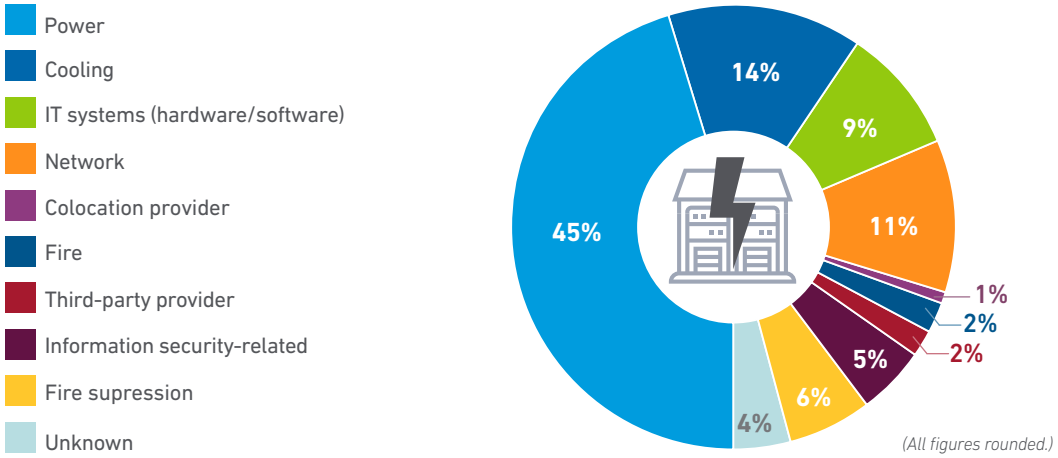
Assessing the causes of outages

Understanding the root causes behind outages is crucial for minimizing downtime risk and directing investment effectively. Identifying these causes can be difficult because failures often involve multiple interdependent systems that can obscure the initial trigger.

Figure 12

Power issues remain the top cause of impactful outages

What was the primary cause of your data center's most recent impactful incident or outage? (n=96)



(All figures rounded.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Power remains the top cause of impactful outages, accounting for 45% of incidents in 2025. These outages are most often associated with UPS issues. For example, separate research from the Uptime Institute Data Center Resiliency Survey 2025 suggests that the top causes of IT service outages stemming from a power-related issue are UPS failures (42%) followed by transfer switch (e.g., utility) failure (36%) and generator failure (28%) (see *Annual outage analysis 2025*).

At the same time, the share of outages caused by power issues has declined significantly from 2024 (54%, n=97), reversing an upward trend seen since 2020. This shift may suggest that recent investments in power redundancy, along with advances in distributed and software-based resiliency architectures, are starting to pay off. The overall gain may be short-lived: there are growing threats to the power chain from aging power grid infrastructure and the destabilizing effects of intermittent renewable energy sources.

Other causes of outages are also becoming more prominent. For example, the proportion of operators citing fire suppression systems as the cause of their most recent impactful outage rose six percentage points from 2024. This increase may reflect a growing concern about data center fires, and a combination of wider deployment of fire suppression systems and a higher rate of accidental discharges.

Outages due to information security-related issues also increased by four percentage points from the previous year (in some other Uptime surveys, security-related incidents in operational technology have grown faster). This category includes, but is not limited to, cybersecurity incidents, such as breaches or cyberattacks. As IT and OT systems become increasingly integrated, the potential for security tools themselves to introduce operational risk also increases. Misconfigurations, complex access controls, and flawed update or patching processes following (and during) a cybersecurity incident can all lead to service disruptions, even in the absence of malicious activity (see *Weak security processes can increase impact of failures*).

These shifts suggest that while traditional outage causes, such as power, remain an ongoing concern, overall operational risks are becoming increasingly diverse. The growing role of software, automation and integrated systems means operators now need to guard against a broader array of threats — including those emerging from the technologies designed to enhance resiliency.

Building more resilient systems

Although reliance on IT is increasing and external risks are growing, operators are mostly relying on redundant architectures to protect against equipment failures. When asked to describe the most common level of redundant power equipment for their primary data center, the highest number of survey respondents (42%, n=679) selected 2N, closely followed by N+1 (41%).

Physical redundancy alone, however, can be insufficient protection against outages. Operators are increasingly adopting distributed and software-based resiliency strategies to ensure service availability — mostly in addition to, not instead of, hardware redundancy. These approaches include using multi-site availability zones, distributed microservices and applications, and platform-level failover mechanisms. Although these approaches offer scalability and flexibility, they also introduce management and configuration challenges.

Importantly, the 2025 data suggests that a growing share of outages are preventable. For example, 87% of respondents who experienced an impactful outage in the past three years (n=98) say it could have been avoided with better management/processes or configuration — a seven-percentage point increase from 2024.

These findings underline the need to pair infrastructure investment with a continued re-examination of existing systems and processes. Efforts to improve regular testing, staff training and change management procedures will remain essential to adapting to an expanding risk landscape. Given that human error plays a role in a high proportion of outages, training and testing is — among all the options for increasing resiliency — likely to be both the cheapest and most effective.

Cloud and provisioning

For the past two decades, enterprises have typically been using a mix of cloud, colocation and their own data centers to host mission critical IT. Given the variety of workloads, and the different ways of measuring work, it is difficult to determine exactly how this usage is distributed. Over the past 10 years, Uptime has been asking managers to supply a rough estimate of where their workloads are distributed (note that small and newly formed enterprises without data centers or colocation sites are unlikely to participate in the survey).

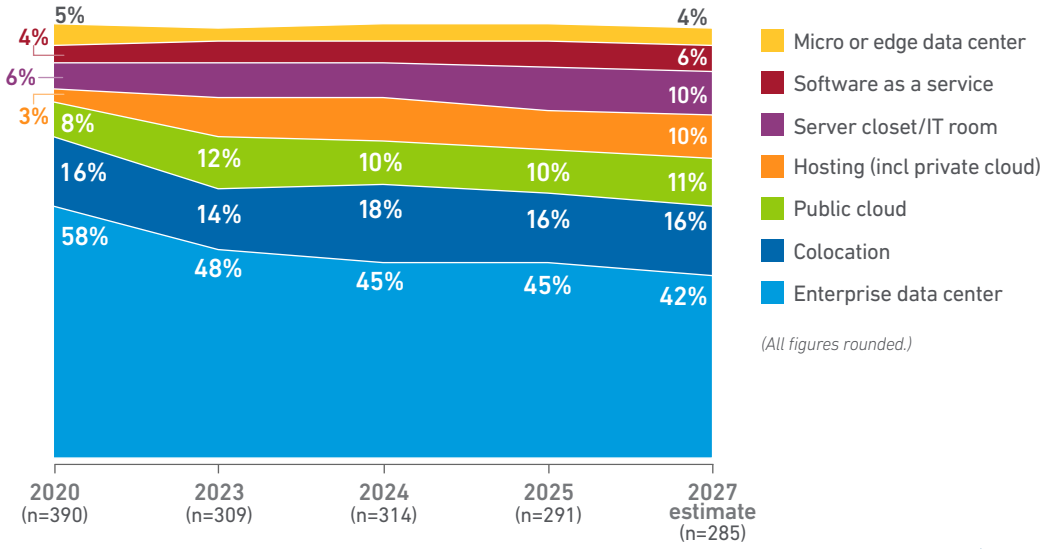
The future remains hybrid

The proportion of IT workloads hosted remotely in off-premises facilities in 2025 remains unchanged from last year (55%), with a modest increase to 58% anticipated for 2027 (see **Figure 13**). Despite the share of IT situated in owned and operated data centers declining over the past five years, they remain the most popular venue for IT workloads.

Figure 13

On-premises facilities remain central to hybrid IT strategies

Approximately what percentage of your organization's total IT would you describe as running in the following IT environments today?



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Hybrid IT is the driving force behind the growth of off-premises facilities: organizations choose the best venue for each workload based on individual requirements, with financial drivers or constraints usually one of the many factors considered. With this approach, some workloads have already moved, or are likely to move, from on-premises to cloud and colocation facilities. However, few large organizations are comfortable moving all workloads out of their own data centers, with data sovereignty and regulatory concerns being top of mind. In addition, enterprise data centers become uneconomic if they are operating too far below capacity.

The long-term trend in IT has, for the past two decades, been a strong and growing use of public cloud that is unlikely to change in the foreseeable future. Over the past year, however, there have been considerable shifts in geopolitical alliances that could affect organizations' choices of data center venues. US-based hyperscale cloud providers are under scrutiny from European customers and government agencies (as well as other nations) regarding the safe and compliant storage of data (see *Tariff tensions undermine trust in hyperscalers*). This uncertainty might help explain why the share of workloads in the public cloud has remained stable at 10% over the past year.

Without reliance on a single venue, a hybrid approach enables organizations to pause their plans and change direction if needed. In an uncertain world, it is likely that some organizations are waiting before continuing to move workloads outside their own facilities and into the cloud. Similarly, they are likely holding off on investing time and money in moving their workloads out of the cloud until there is more clarity (see *Cloud repatriation is overstated*).

The value of AI remains unclear

Although the appeal of AI is strong, the value of AI is still uncertain for most organizations. Despite the significant hype and media attention regarding generative AI, few projects have so far demonstrated a clear return on investment. Organizations may be experimenting with AI today to stay ahead in this potentially game-changing technology, but Uptime's data (from multiple sources) suggests that most are holding off on major investments until AI's value becomes more evident. For now, maintaining the status quo, supplemented by pilots and proof-of-concept projects, is a safer bet amid uncertainty around the reliability of other options.

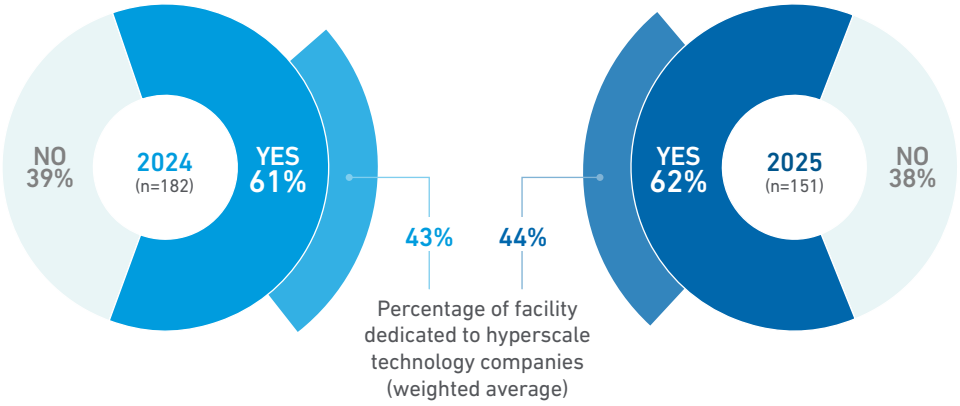
Although colocation represents only 16% of venue share from the enterprise perspective, demand from hyperscale tenants is driving the continued expansion of large colocation campuses. Demand for these cloud services is expected to continue to grow, regardless of geopolitical tensions. How much they use cloud services, however — and for what workloads — may change if US hyperscalers fail to reassure non-US customers that their data is secure.

Uptime's data shows the extent to which colocation providers are vital partners for hyperscalers. Almost two-thirds of colocation providers say they support hyperscale customers (a similar level to last year). The share of data center space they have allocated to hyperscalers has a weighted average of 44%, which is also a steady percentage in a growing market (see **Figure 14**).

Figure 14

Most colocation providers support hyperscale customers

Does your colocation facility host hyperscale technology companies, such as AWS, Google, Microsoft, IBM, Meta, Tencent, Alibaba or Baidu (n=182)? If so, how much of your facility is dedicated to hosting these companies?



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

AI is a major — and new — driver of hyperscaler demand for colocation space. Hyperscalers not only require this space to deliver infrastructure as a service to cloud customers, but also to train their own models to be used subsequently by consumers. However, AI is not the only driver: much of the growth is due to business as usual: more customers, using more services, in more regions.

There have been some well-publicized cases of hyperscalers canceling data center construction projects or rescinding lease agreements. Uptime's analysis suggests that these changes in strategy do not indicate a pullback in their investments or a slowdown in growth. Instead, they reflect minor adjustments in the rapid build out of capacity and highlight a degree of underlying uncertainty. Hyperscaler plans, like other organizations, may evolve in response to changing market conditions or circumstances, and some may occasionally pause data center projects while they wait for uncertainties to subside.

Staffing

The data center industry has faced shortages of staff and skills for more than a decade. This year, Uptime's annual survey indicates that the labor shortfall has not worsened. This small "victory" may be partly due to improvements in training and recruitment strategies.

Talent shortage is neither worse nor better

Hiring and retention difficulties remain widespread and justify a sustained effort to attract talent from a broad range of sources and groups of candidates. There have been some shifts over previous years, with skills gaps affecting different job roles. There is currently a greater need for operations management candidates, and talent shortfalls in electrical and mechanical trades continue to be significant.

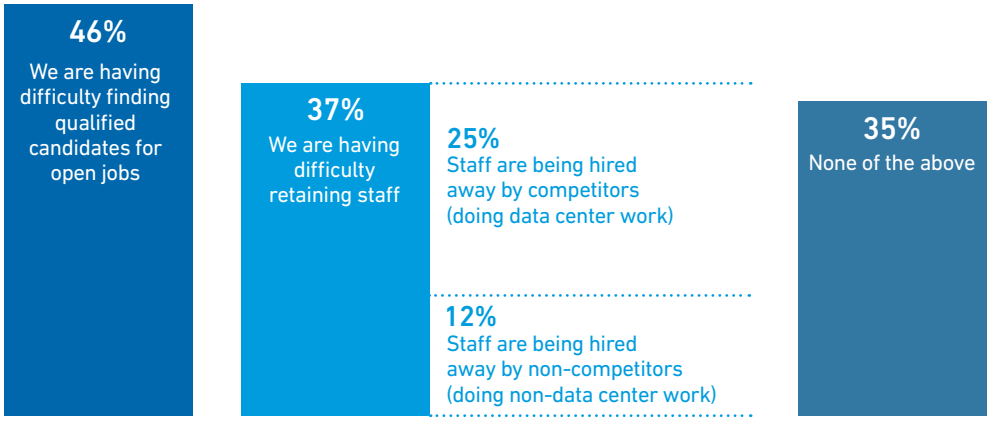
Operators say hiring and retention remain as challenging as they have been over the past two years. Nearly half (46%) of operators report difficulties finding qualified candidates for vacant roles, and more than one in three (37%) struggle to retain their staff (see **Figure 15**).

The shortage stems partly from a growing demand for skilled workers and partly from the loss of existing staff. These workers, however, are generally not leaving the sector. Of the operators grappling with staff retention, most report workers being hired away by competitors (doing data center work) rather than leaving the industry entirely. This suggests most departing staff are content with data center employment, but are looking for greater benefits, such as higher salaries or more flexible schedules.

Figure 15

Most operators struggle with employee hiring, retention, or both

Please select any of the following statements that apply. (n=650)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Poaching of workers from other data centers is a widely used and popular approach, but is only a short-term strategy to meet an organization's most urgent staffing requirements. The industry's labor pool is already constrained, and hiring from competitors only fuels higher worker turnover, driving up labor costs and diverting resources from programs that could result in a stronger long-term talent pipeline.

Uptime has tracked gender demographics in data center design, build and operations teams since 2018. The balance has not moved meaningfully, and women remain poorly represented. Only 5% of operators report that women account for half of their data center staff. Three in four operators (75%) say their workforce comprises 10% women or less, while a majority (58%) report employing just one woman (or fewer) per 20 workers.

The sector's persistent staffing challenges justify renewed effort to hire more women, but only around one in three (38%) organizations report having a formal initiative to boost the recruitment of women candidates. Other under-represented groups could also provide fresh talent pools.

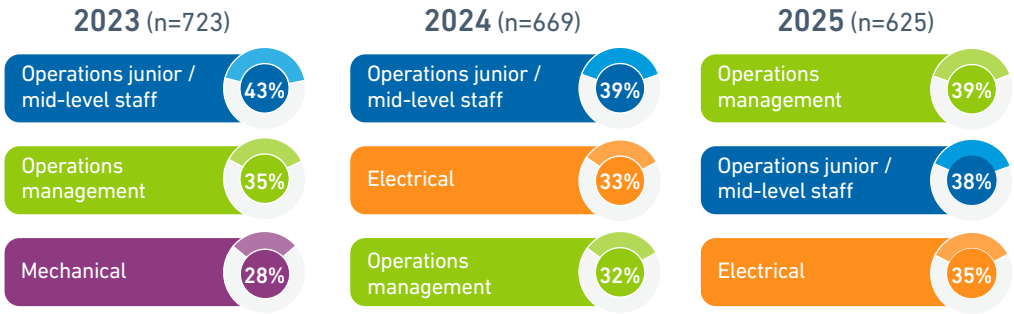
Management sees widening skills gap

Uptime survey data tracks which job categories show the largest skills gaps. There have been shortfalls in virtually every job category since 2023, but these gaps are not evenly distributed. Until this year, junior and mid-level operations roles topped the list of talent shortages, with smaller — but significant — shortages in operations management, electrical and mechanical roles. In 2025, operations management is ranked as the job category with the most serious skills gap, cited by 39% of operators (see **Figure 16**).

Figure 16

More operators see skills gap in operations management

In which of the following areas is your organization experiencing skills gaps? Choose all that apply.



(Only the top three response categories for each year are included.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

This is not to suggest junior and mid-level operations staff, and electrical and mechanical roles have become less of a concern: the share of operators identifying skills gaps in each of these roles has not decreased since 2024 — all have remained above 30%.

The talent shortage in operations management may be partly due to a growing number of new facilities, as well as a loss of experienced senior staff through retirement. It also suggests, however, structural problems with data center careers. Junior operators accumulate experience over time, and some will attain operations manager positions. A growing management shortage could suggest a failure to pass on knowledge to junior employees as more senior staff retire. Industry stakeholders have warned, repeatedly, of a "silver tsunami," where older, experienced operators and managers retire faster than can be replaced by new workers.

Crucially, shortfalls in skilled management can have a knock-on effect. Unless there are enough experienced managers to train new incoming workers, organizations will never be able to acquire and develop the staff and skills they need in junior-level operations roles.

The current surge in the construction of densified data centers for generative AI and other applications is likely worsening the shortage of electrical and mechanical skills. These skilled workers are needed in greater numbers to build power distribution and cooling infrastructure for increasingly large and numerous facilities.

Finding qualified skilled trades workers, such as electricians and mechanical technicians, can be especially challenging for data center operators. While operations teams often grow talent internally, skilled trades professionals are typically hired with experience from other industries, such as utilities or manufacturing and industrial facilities. Meeting the industry's demand for qualified electricians (and similar roles) may require more collaboration with underutilized avenues, such as local trade schools.

Appendix

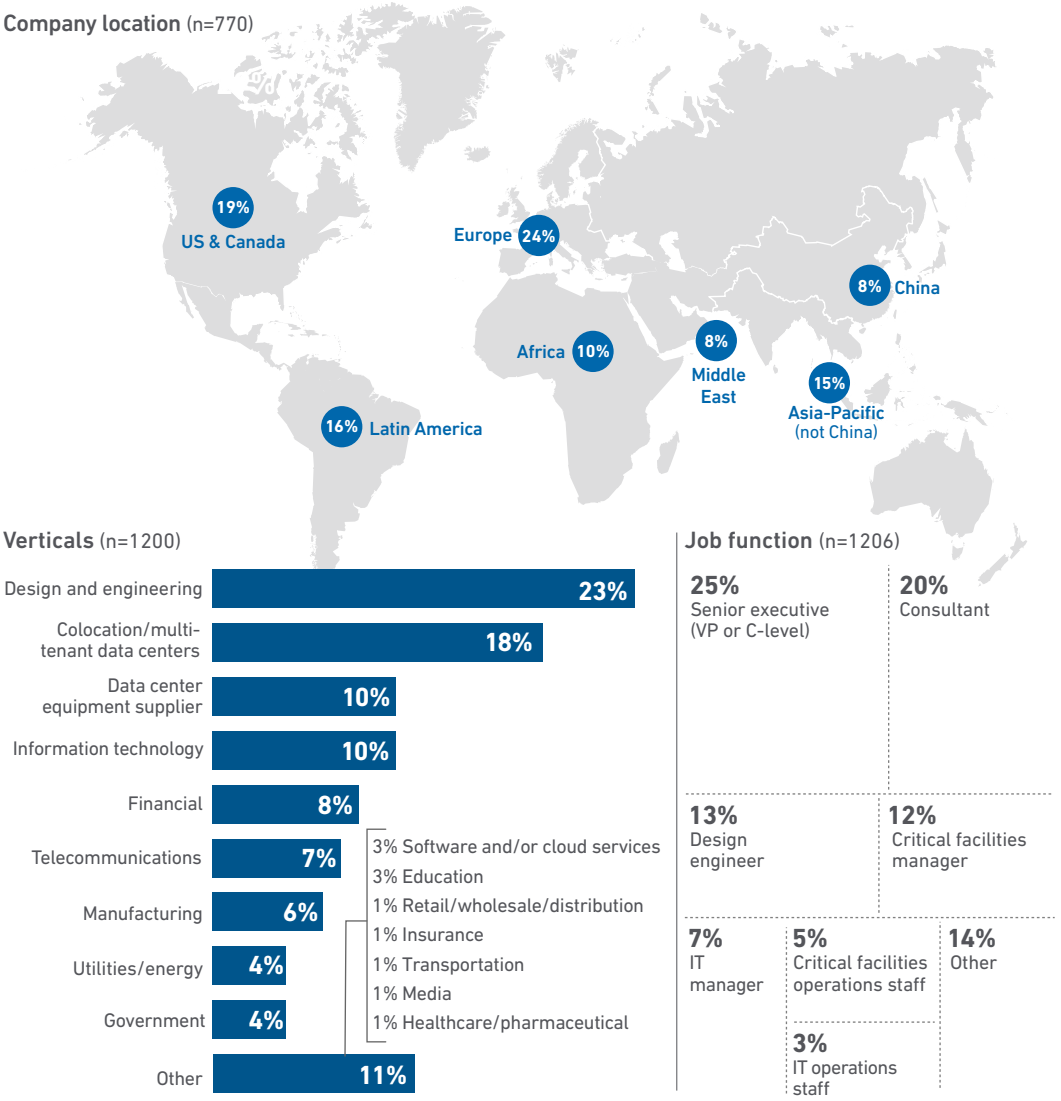
Survey methodology and demographics

Uptime Institute's Global Data Center Survey, now in its 15th year, is conducted annually online and by email. The 2025 survey was conducted in the first half of the year.

This report focuses on responses from the owners and operators of data centers, including those responsible for managing infrastructure at the world's largest IT organizations. Job titles include senior executive, IT manager, IT operations staff, critical facilities manager, critical facilities operations staff, design engineer and consultant.

Figure 17

Respondents by location, industry vertical and job function



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025

Appendix (*continued*)

The participants represent a wide range of industry verticals in multiple countries. Nearly half (43%) are located in North America and Europe. Approximately one in five respondents work for professional IT/data center service providers — that is, staff with operational or executive responsibilities for a third-party data center, such as those offering colocation, wholesale, software or cloud computing services.

A total of 835 end users registered for the survey and answered at least one question. The number (n) of respondents varies between individual questions because respondents are not required to answer every question.

If you have questions, comments or seek further insights, please contact:
research@uptimeinstitute.com

This report is based on the entire survey sample. If you are interested in additional data (e.g., by region, sector, data center size, etc.) please contact: Simon Carruthers at scarruthers@uptimeinstitute.com

About the authors

Douglas Donnellan

Douglas Donnellan is a Research Analyst at Uptime Institute covering sustainability in data centers. His background includes environmental research and communications, with a strong focus on education.

ddonnellan@uptimeinstitute.com

Andy Lawrence

Andy Lawrence is Uptime Institute's Executive Director of Research. He has spent three decades analyzing developments in IT, emerging technologies, data centers and infrastructure. He also advises companies on their technical and business strategy.

alawrence@uptimeinstitute.com

Daniel Bizo

Daniel Bizo is Uptime Institute's Research Director. He has been covering the business and technology of enterprise IT and infrastructure in various roles, including more than a decade as an industry analyst and advisor.

dbizo@uptimeinstitute.com

Dr. Owen Rogers

Owen Rogers is Uptime Institute's Research Director of Cloud Computing. Dr. Rogers has been analyzing the economics of cloud for over a decade as a chartered engineer, product manager and industry analyst. He covers all areas of cloud, including AI, FinOps, sustainability, hybrid infrastructure and quantum computing.

orogers@uptimeinstitute.com

Peter Judge

Peter Judge is a Senior Research Analyst at Uptime Institute. His expertise includes sustainability, energy efficiency, power and cooling in data centers. He has been a technology journalist for 30 years and has specialized in data centers for the past 10 years.

pjudge@uptimeinstitute.com

Jacqueline Davis

Jacqueline Davis is a Research Analyst at Uptime Institute covering global trends and technologies that underpin critical digital infrastructure. Her background includes environmental monitoring and data interpretation in the environmental compliance and health and safety fields.

jdavis@uptimeinstitute.com

Max Smolaks

Max Smolaks is a Research Analyst at Uptime Institute. His expertise spans digital infrastructure management software, power and cooling equipment, and regulations and standards. He has 10 years' experience as a technology journalist, reporting on innovation in IT and data center infrastructure.

msmolaks@uptimeinstitute.com

Rose Weinschenk

Rose Weinschenk is a Research Associate at Uptime Institute covering staffing and education in data centers. Her background includes psychology research, with a focus on ethics.

RWeinschenk@uptimeinstitute.com

All general queries

Uptime Institute
405 Lexington Avenue
9th Floor
New York, NY 10174, USA
+1 212 505 3030

info@uptimeinstitute.com

About Uptime Institute

Uptime Institute is the Global Digital Infrastructure Authority. Its Tier Standard is the IT industry's most trusted and adopted global standard for the proper design, construction, and operation of data centers — the backbone of the digital economy. For over 25 years, the company has served as the standard for data center reliability, sustainability, and efficiency, providing customers assurance that their digital infrastructure can perform at a level that is consistent with their business needs across a wide array of operating conditions. With its data center Tier Standard & Certifications, Management & Operations reviews, broad range of related risk and performance assessments, and accredited educational curriculum completed by over 10,000 data center professionals, Uptime Institute has helped thousands of companies, in over 100 countries to optimize critical IT assets while managing costs, resources, and efficiency.

Uptime Institute is headquartered in New York, NY, with offices in Seattle, London, Sao Paulo, Dubai, Singapore, and Taipei.

For more information, please visit **www.uptimeinstitute.com**