uptime INTELLIGENCE

Uptime Institute Global Data Center Survey 2026

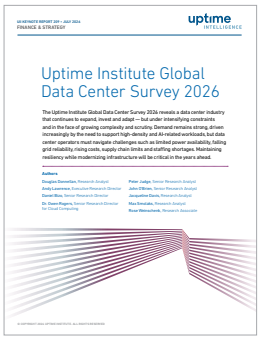
The Uptime Institute Global Data Center Survey 2026 reveals a data center industry that continues to expand, invest and adapt — but under intensifying constraints and in the face of growing complexity and scrutiny. Demand remains strong, driven increasingly by the need to support high-density and AI-related workloads, but data center operators must navigate challenges such as limited power availability, falling grid reliability, rising costs, supply chain limits and staffing shortages. Maintaining resiliency while modernizing infrastructure will be critical in the years ahead.

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
John O'Brien, Senior Research Analyst
Jacqueline Davis, Research Analyst
Max Smolaks, Research Analyst
Rose Weinschenk, Research Associate

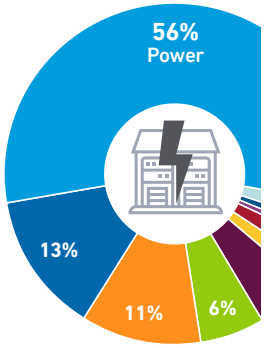
Key points

- High costs remain the greatest concern for digital infrastructure management teams in 2026, but concerns around capacity forecasting, power availability and supply chain disruptions are growing.
- Average PUE levels continue to improve in 2026, but the gains remain gradual. Newer, more efficient data centers are being built, but legacy infrastructure and cooling constraints are slowing progress.
- Average modal rack densities continue to rise slowly, and more operators are now reporting peak rack densities of 30 kW or above.
- The share of operators who say their organization collects sustainability metrics returned to its gradual upward trend in 2026, and more than half now say they track water use.
- Expectations for AI's operational benefits declined slightly in 2026. Trust remains highest for lower-risk applications, such as sensor data analytics and predictive maintenance.
- Fewer operators report experiencing an impactful outage in the past three years, continuing a long-standing trend. However, one in ten outages is still classified as serious or severe, and the costs of outages continue to rise.
- For the first time, third-party data center facilities and services account for a larger share of IT workloads than enterprise-owned data centers (by facility, not total capacity). However, corporate data centers remain central to hybrid IT strategies.
- Staffing and skills shortages may be intensifying. More than half of respondents now report difficulty finding qualified candidates for open jobs in 2026, and staff turnover remains a persistent challenge.



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 16th 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 to May 2026 and collected responses from more than 800 data center owners and operators, and over 800 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**.

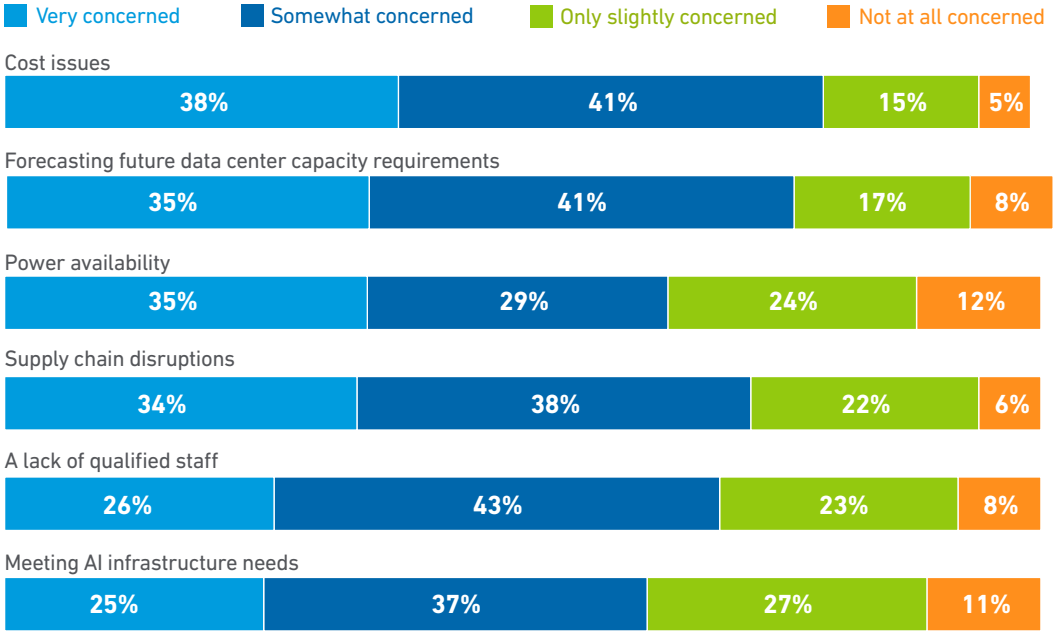
For the fourth consecutive year, Uptime's survey asked data center operators to identify their management team's top concerns related to digital infrastructure. Cost issues remain the highest-ranked concern, with 79% of respondents saying management is at least somewhat concerned (see **Figure 1**). While the share saying management is "very concerned" about costs is below the 44% reported in 2024, the high prices of power, staff and equipment — particularly for AI-related infrastructure and IT — persist.

The lower priority given to costs compared with previous years does not necessarily mean those concerns have eased. Rather, costs are being weighed alongside other issues that affect operators' ability to expand, modernize and maintain resiliency.

Figure 1

Cost issues lead, but management concerns are widening

Looking at the next 12 months, how concerned is your digital infrastructure management regarding each of the following issues? (n=537)



(Only the top six response categories are shown. All figures rounded.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

Capacity forecasting is a growing concern. The share of respondents saying management is at least somewhat concerned about forecasting future data center capacity requirements has risen steadily, from 68% in 2024 to 76% in 2026. This likely reflects uncertainty about the scale, timing and location of future demand, especially as operators assess how AI and higher-density workloads may affect overall demand for data center capacity, as well as the demand for power and cooling.

Concerns about power availability remain stable compared with 2025, while the share of respondents who say that management is very concerned about supply chain disruption rose eight percentage points. While the findings around power availability may run counter to the known industry challenges, they may be explained by the fact that most respondents are already operating facilities and most have sufficient power to meet near-term plans.

None of these concerns stands independently from the others. For example, a shortage of qualified staff can delay projects and make operations harder to manage. AI infrastructure needs can also complicate decisions about future capacity, power, cooling, cost and technical skills. The 2026 results therefore show an industry that is still growing, but with more limits on how quickly and easily operators can add capacity. These limitations, as more studies are now beginning to find, can act as a brake on the speed of the data center build-out.

Industry benchmarks

The 2026 Uptime Institute Global Data Center Survey draws upon 16 years of data from a large sample of operators — a long timeline that illuminates high-level trends across key operational and design metrics in the industry. Respondents shared data regarding their facility energy performance, rack power density, and related aspects of capacity management. This industry-wide data shows a sustained, measured pace toward more efficient facilities and denser racks, even against a backdrop of accelerating capacity growth.

The trends, however, are beginning to change. The build-out of dedicated AI facilities, with their distinct power and density needs, is beginning to exert a greater influence on industry averages. While such facilities are still few in number compared with existing conventional data centers, a growing number of facilities are being built expressly for generative AI. Currently, however, their influence on industry benchmarks remains tempered.

PUE edges downward

Introduced by The Green Grid in 2007, power usage effectiveness (PUE) is a metric that is widely used as a proxy for data center energy efficiency — especially to monitor the energy performance of a given facility over time. The data upon which PUE relies is almost universally available to operators, and its calculation is straightforward: total facility power divided by power consumption of IT equipment. This ease of calculation has made PUE the most successful metric of facility energy performance.

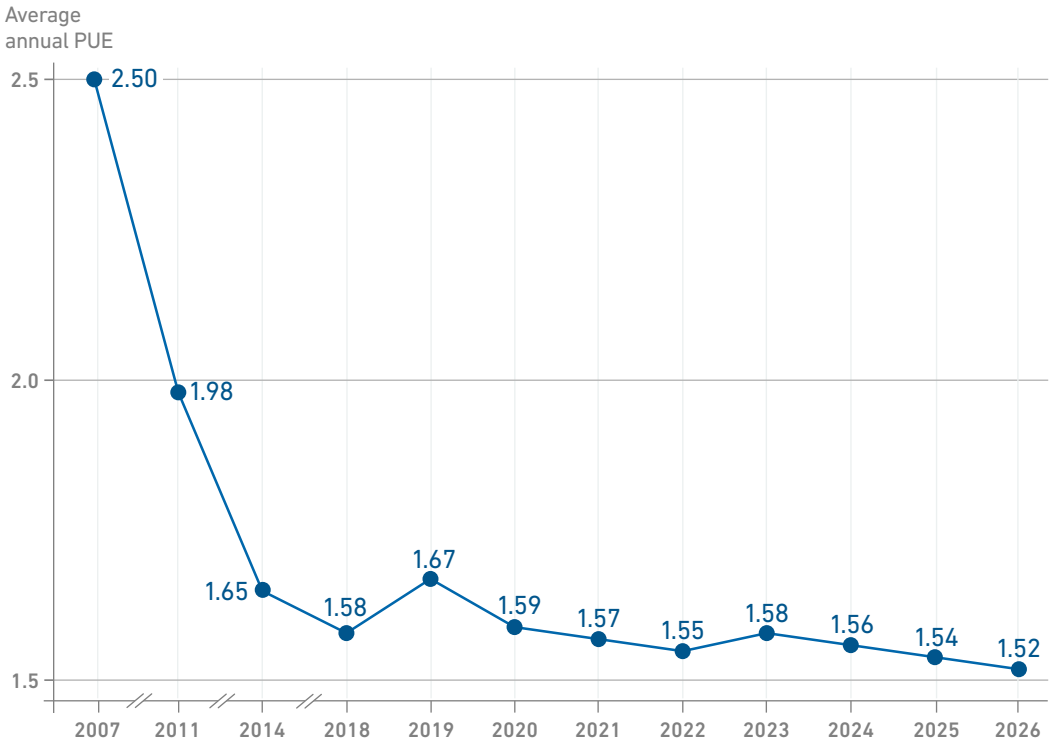
Its simplicity and ubiquity are tempered by its limitations. Key data center design characteristics such as business objectives, resiliency objectives and climatic conditions all exert an influence on PUE and vary markedly between facilities. Important trade-offs between facility power needs and water consumption, temperature setpoints and IT energy efficiency are outside the scope of PUE as well. For this reason, PUE is best suited to evaluate a single site over time or to observe directional motion in large sample sizes, rather than comparing between only a few differing facilities.

Uptime Intelligence has been collecting average annual PUE figures from a diverse cross section of data center operators since 2007 to document trends in facility efficiency. The 2026 industry average PUE is 1.52 (see **Figure 2**), a figure that falls right in line with a seven-year trend of relative stasis. The stability of this calculated mean value obscures the myriad variations in the data centers it represents — but also suggests that in our sample, their opposing influences are roughly in balance.

Figure 2

Average PUE still inching toward 1.5

What is the average annual PUE for your data center? (n=644)



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2007-2026

PUE's precipitous drop from 2007 to 2014 denotes a period when operators first discovered the extent of their inefficiency and found ample opportunities for cost-effective and comparatively easy facility efficiency improvements. Chief among these are relaxed air supply temperatures, aisle containment systems, blanking panels, optimized cooling system controls, and more efficient electric motors and distribution systems. In legacy data centers, upgrades more substantial than these can be costly, or might not be possible without disrupting application availability.

As the most cost-effective efficiency measures became widely adopted, average PUE improvement became far more gradual. After this point, further progress owes largely to the construction of new facilities that tend to be more efficient than their predecessors. The pace of this construction has been sufficient to lower the average PUE, but only incrementally, as older data centers have remained in operation to meet capacity demand. Even today, legacy facilities are not decommissioned at the pace of new construction — more than a quarter (27%) of survey respondents work in data centers over 15 years old.

Newer and larger data centers outperform the industry average, often by a significant margin. Globally, the average annual PUE of respondents' largest data centers (which may be different from the one they work in) was 1.45. This effect is partially because very large data centers in our sample are also newer, and recent data center builds routinely produce PUE of 1.3 or better where design optimization and climate take advantage of fresh air cooling or evaporative cooling to reduce energy consumption. Importantly, almost one in four respondents (23%) work in a data center averaging below 1.3 PUE.

The ongoing rapid build-out of new data centers with optimized designs will continue to exert downward pressure on PUE, but Uptime Intelligence expects the overall pace will remain slow. Persistent capacity demand, substantial equipment price increases and long lead times incentivize keeping legacy data centers operational. Furthermore, energy-efficient cooling techniques which minimize or eliminate reliance on mechanical cooling are not yet practical in all climates.

These factors will temper industry-wide PUE improvement in the next few years as data center development accelerates in subtropical and desert regions. Adoption of direct liquid cooling (DLC) operating at elevated temperatures will help over the longer term. Legislation in some countries (such as Germany) may force down PUEs over time, but the effect will likely not be strong or dramatic.

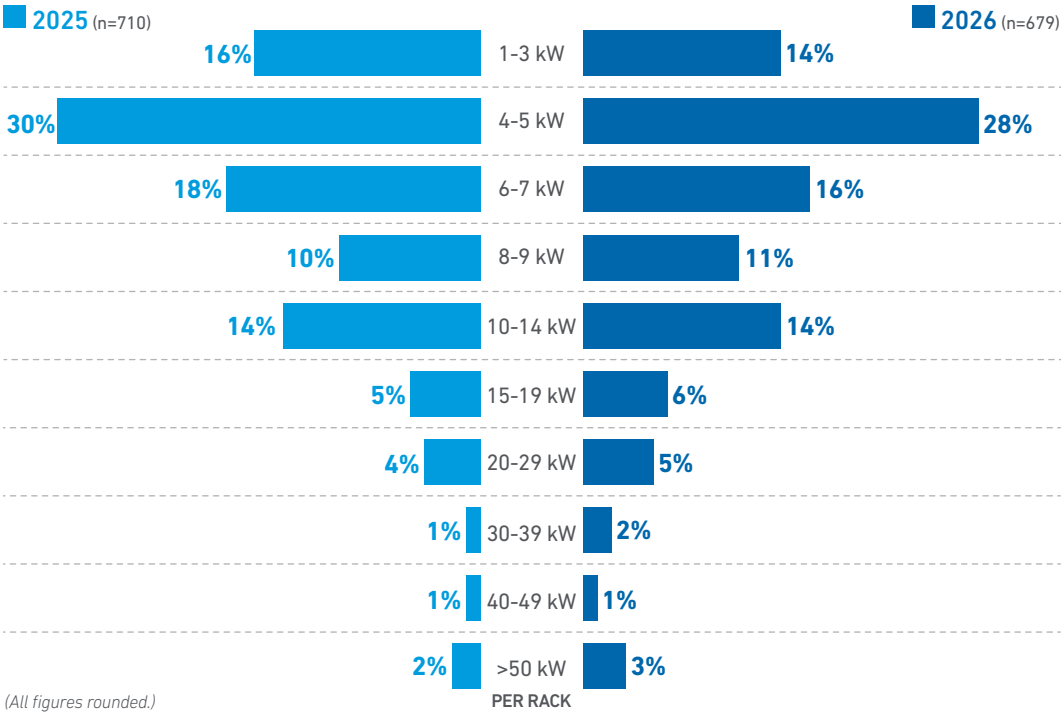
High-density racks exert force

The long-term trend of densification in data centers (defined by IT rack power draw) continues unabated, according to our 2026 survey. For the first time, the average of modal (most typical) rack densities surpassed 11 kW in our sample, as the gradual shift toward higher-powered hardware was compounded by a small number of new high-density facilities with typical racks above 30 kW. These factors produced a visible increase from last year's figure of 9 kW. Without these high-density facilities, the average modal density sits at 7.8 kW, up from 7.5 kW in 2025. **Figure 3** shows respondents' reported average densities.

Figure 3

Average density rises, with and without super-dense leaders

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



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

We expect this trend of rising rack densities to continue as IT organizations gradually upgrade their server fleets with new systems and seek operational efficiencies. Recent and next-generation servers boost both workload capacity and energy performance, but maximizing these benefits means a corresponding rise in system power. As legacy systems are retired, typical rack densities will edge higher too. Some operators are now pursuing more aggressive technology refresh timelines of less than 4 years (see section: **Sustainability and metrics**).

Where are densities headed? It is not yet clear whether a new dominant standard will emerge to displace the 4-5 kW average rack design density. Data centers have centered on this range for more than a decade, and it remains the most common rack density in deployment. Future trends will largely depend on the operational discipline of driving IT utilization higher. The incentives for this are becoming stronger: IT hardware costs are high, and the pressure to make the most of the facility power capacity (or the contracted power in colocation sites) is escalating.

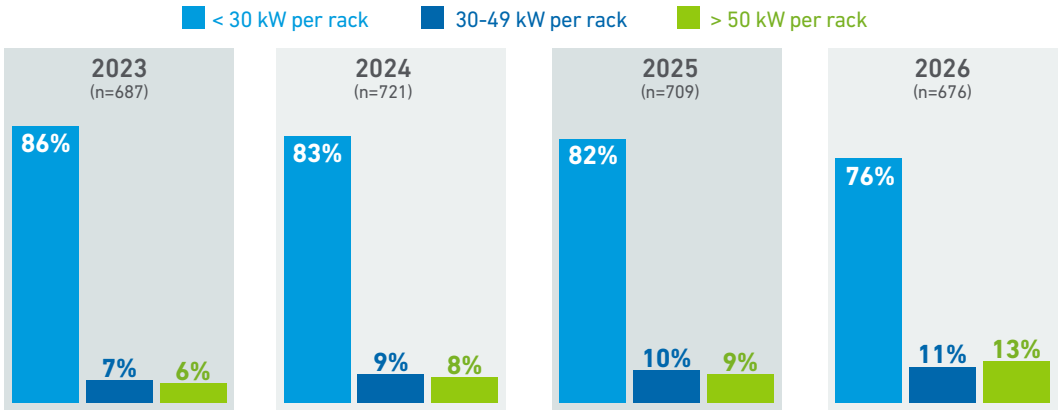
Most new racks populated with current and next-generation servers for balanced switch port counts (e.g., 16, 24 or 32 servers) will converge on 10 kW from both higher and lower bands. The survey data bears out this theory for now: racks between 8 kW and 15 kW represent the modal density band in a quarter of facilities in our sample.

When it comes to respondents' peak densities, 2026 brings a clear uptick in more powerful racks after a fairly uneventful reading in 2025 (see **Figure 4**). The majority of facilities still do not have any racks at 30 kW or above in deployment. Statistically, 30 kW represents the lower bound of high density. However, more respondents (24%) now say they have some of these racks, up from 19% in 2025. This increase was seen mostly in the 50-plus kW ultra-high-density range (about one in seven respondents in 2026), including facilities with some racks above 100 kW. Clearly, as AI and GPU servers are more widely deployed, densities will rise — and the highest will rise in line with Nvidia's product roadmaps.

Figure 4

More operators host high-density racks

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



(All figures rounded.)

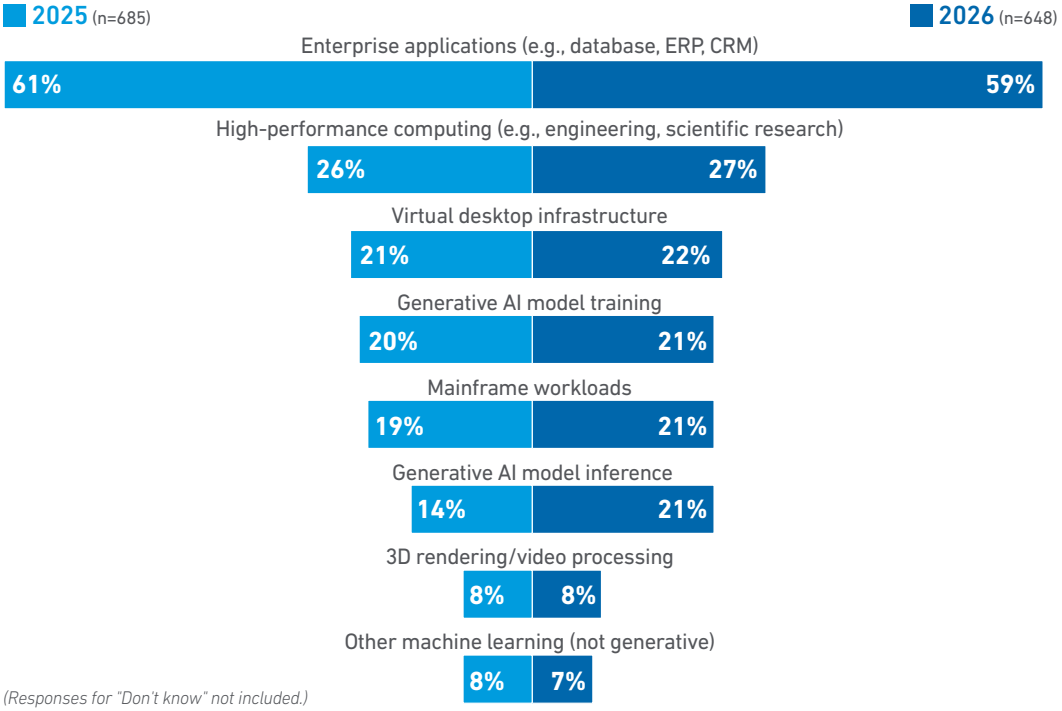
UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2023-2026

The workloads that drive the installation of the highest-density racks are diverse, but enterprise software and applications dominate. **Figure 5** shows that applications such as databases, enterprise resource planning, mainframe applications, and virtual desktop infrastructure continue to drive densification under 50 kW. Above this mark, high-performance computing (HPC) and AI training take the lead. These observations align with past findings. Notably, AI inference has risen to match AI training for the first time in this survey, with more than a fifth (21%) of respondents now reporting each as a workload driver for more powerful racks.

Figure 5

Enterprise workloads remain key driver for rack densification

Which of the following types of applications are driving the highest density deployments in your data center? Choose all that apply.



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

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

Sustainability and metrics

As the AI infrastructure frenzy has taken hold in recent years, the importance of collecting sustainability and efficiency data appeared to have subsided. But the survey data confirms what many operators have been saying: sustainability and efficiency are critically important to the long-term effectiveness of most data centers.

After faltering in 2025, data collection for all sustainability metrics resumed a slow upward trend (see **Figure 6**), partly driven by continued requirements for this data by governments and other parties.

The wavering of data collection in 2025 may have been partly due to a perceived loosening of regulatory requirements in some regions. At the time, Uptime commented that disengaging from sustainability issues would be shortsighted because operators would sooner or later face increased demand for data. In both the US and Europe, for example, the relaxing of some reporting requirements is being counterbalanced by other, more stringent rules. As a result, the long-term trend toward increased reporting remains.

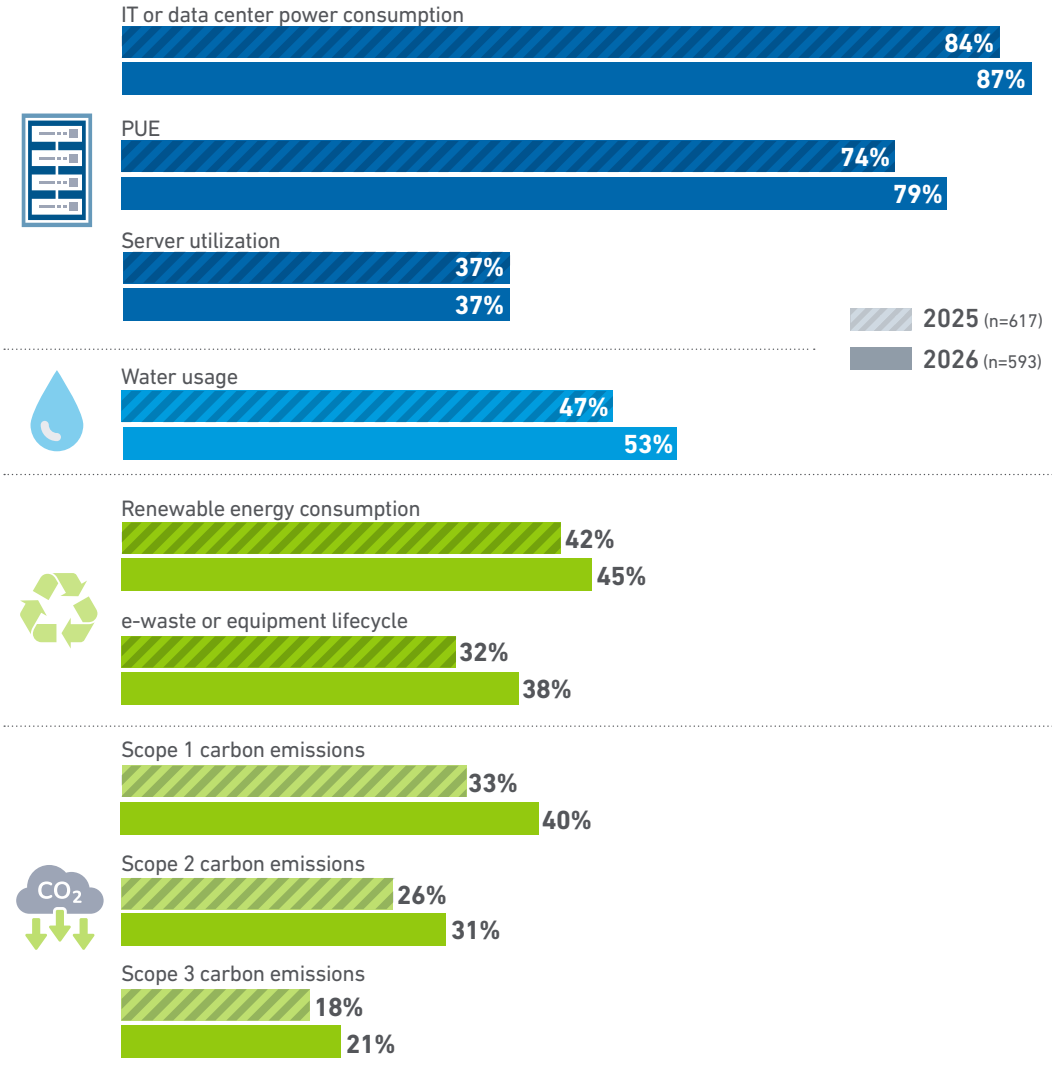
In 2026, the most significant increases in data collection were in water usage, e-waste and carbon emissions reporting, particularly Scope 1 (direct emissions). Server utilization is the only metric that stagnated, remaining on par with 2025. The increase in water consumption data collection is likely due to growing public pressure and new legislation, while stalled server utilization may stem from the complexities and volatility from deploying new GPU and CPU processors and AI workloads.

Energy consumption and PUE remain the most tracked efficiency and sustainability metrics, at 87% and 79%, respectively. They are often required by regulators or customers, are easy to collect, and can help reduce operational costs.

Figure 6

Tracking sustainability metrics resumes an upward trend

Which of the following IT or data center metrics does your organization collect for corporate sustainability purposes whether or not you report it externally? Choose all that apply.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

Besides PUE, energy usage and water consumption, other sustainability metrics remain poorly recorded. These include renewable energy use and carbon emissions, which are essential for measuring an organization's contribution to global warming via its operational and embedded carbon footprint.

While the base level of reporting these metrics is low, more operators are now collecting and reporting this data. This may be due to customer pressure — stakeholders and customers increasingly require the data for their own environmental assessments.

There may also be a new driver at play: some governments, newly aware of the importance of digital infrastructure, have classified it as critical national infrastructure (CNI). This may, in turn, lead to stronger requirements to align data centers with government net-zero plans.

The increased level of reporting comes with a drawback: few operators are on track to meet their publicly stated carbon emissions targets (usually net zero by 2030 to 2040), and as more build new data centers, many are going backwards. This has led, as Uptime long predicted, to a widespread rewriting of carbon emissions goals, with much less ambitious targets (see *Net-zero timelines are becoming more realistic*).

There are significant and long-standing regional variations in sustainability reporting, with Europe and Asia more focused on these issues — especially regarding the use of renewable energy and carbon emissions. This is partly due to emerging legislative requirements.

Site-level environmental reporting

Although it is not always comfortable for the operator, Uptime Institute has often advised that environmental reporting is most useful when done per site rather than aggregated across an entire portfolio. The increased granularity allows operators to target local inefficiencies and address issues such as the role of renewable energy on local grids.

Most of the required data is already available but is not collected and reported at a site level. The metrics most likely to be reported per site are PUE, power consumption, renewable energy, water usage and greenhouse gas (GHG) scope emissions. European operators are more likely than their North American counterparts to collect data per site, driven largely by the EU's Energy Efficiency Directive reporting requirements.

Uptime asked which per-site metrics are passed on to third parties, whether to governments, customers, or others. Power consumption (75%), PUE (75%) and Scope 1 GHG (50%) are the leading metrics most likely to be shared.

Scope 2 and Scope 3 GHGs are recorded by only 31% and 21% of all survey respondents, respectively. However, when they are recorded, they rank much higher on the list of metrics reported to third parties (at 42% and 29%). This indicates that where these metrics are gathered, they are collected for a clear reason — because an external party requires them.

Variation by sector shows few surprises. Colocation providers report recording PUE (94%), renewable energy (60%) and water usage (71%) at higher rates than enterprise or cloud operators, primarily because their customers expect it. Correspondingly, colocation providers report server utilization (22%) far less often than enterprise or cloud operators because they would have to request this data from clients and have a valid reason (e.g., legislative requirements) for doing so.

Into the racks

Uptime has long argued that the drive to improve efficiency should expand its scope beyond the mechanical and electrical systems to the IT estate, where energy and hardware use can be significantly optimized.

Server utilization is an obvious first step. Underutilized servers waste energy while idling, whereas fully utilized servers perform more work per kilowatt-hour. Servers are instrumented so that their utilization can be readily measured, and many tools (such as power management) can help improve both utilization and energy efficiency. For cloud service providers, higher server utilization helps generate revenue, but operators need to balance this against availability and responsiveness.

Although server utilization is monitored and managed by IT teams, it is not widely regarded as a key sustainability metric. Excluding colocation operators, which do not have access to this data, around 42% of operators collect it. This is lower than for all other metrics except e-waste and the more complex, emerging measures, such as GHG Protocol Scope 2 and 3.

Regional variations in server utilization reveal a surprise: it is more widely tracked in North America. It is the only metric collected at a significantly higher rate in North America (44%) than in Europe (35.5%). Meanwhile, Latin America (35.8%) is on par with Europe.

Server utilization is reported to third parties less often (21%) than it is recorded, reflecting the fact that regulators and customers are not yet requesting this metric.

Server refresh: an increasingly watched metric

IT sustainability specialists have long paid attention to server refresh rates (how long organizations keep servers before replacing them) — because IT equipment contains valuable materials and contributes to embodied carbon and e-waste. Keeping servers in service for longer can reduce e-waste and spread the embodied carbon of the equipment over a longer useful life. However, there is a trade-off: older servers may consume more energy for a given level of work, which can offset some of the sustainability benefits of delaying replacement.

Refresh rates are watched for other reasons too. During periods of rapid innovation, they can shorten as newer and more efficient equipment is brought in; and accountants — and insurers — can be caught out by sudden swings in the cost of replacement equipment. When refresh rates are shorter, older equipment is likely to lose value more quickly.

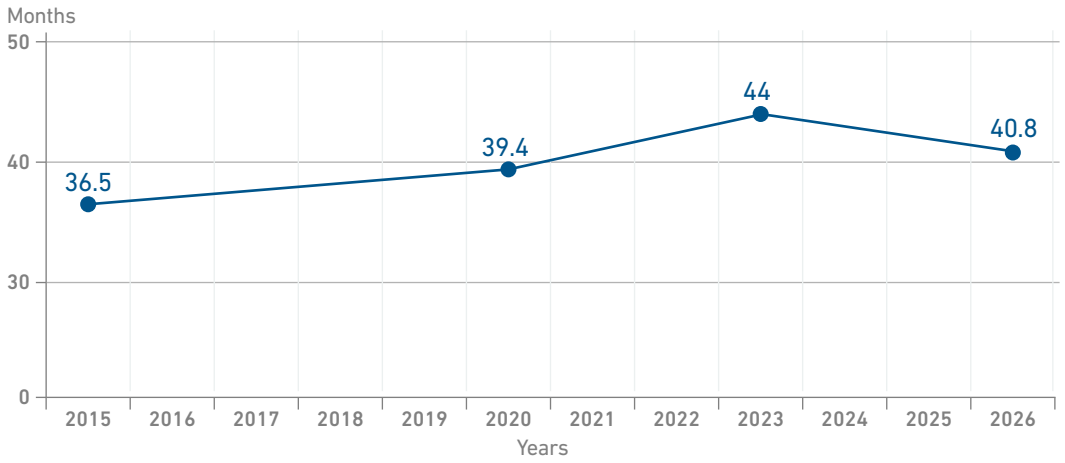
Between 2015 and 2023, operators reported increasing server life cycles, from around 36 to 44 months. Since 2023, this trend has reversed: operators say they are retiring servers sooner, and the reported server lifetime has decreased to around 41 months, slightly above the 2020 level (see **Figure 7**).

This partly relates to the current rapid build-out of data centers and a shift toward higher-density applications, including HPC and AI, where newer and more powerful servers are more attractive because they deliver higher performance. However, there is a more specific explanation too: in the period leading up to 2023, Intel, still the largest incumbent supplier of CPUs to the commercial server market, was unable to introduce new processors with enough performance gains to encourage widespread replacement. With the introduction of new processors, Intel (along with competitor AMD) is now correcting this.

Figure 7

Server refresh cycles shorten modestly after 2023 peak

How often do you typically refresh servers in your data center?



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2015-2026

Operators trust their data more

Operational data quality has long been a source of concern for data center operators. In private discussions, many admit they do not trust their operational data because of unreliable legacy systems, data silos and increasing system incompatibility.

This is the first time Uptime's annual survey has asked operators about the quality of operational data collected from sensors, software and equipment. The results are positive, revealing a high degree of trust in the accuracy, timeliness and completeness of data (see **Figure 8**). On average, more than 90% of respondents say they are very or mostly confident in this data across the four characteristics, and only one in ten is either not very confident or not confident at all.

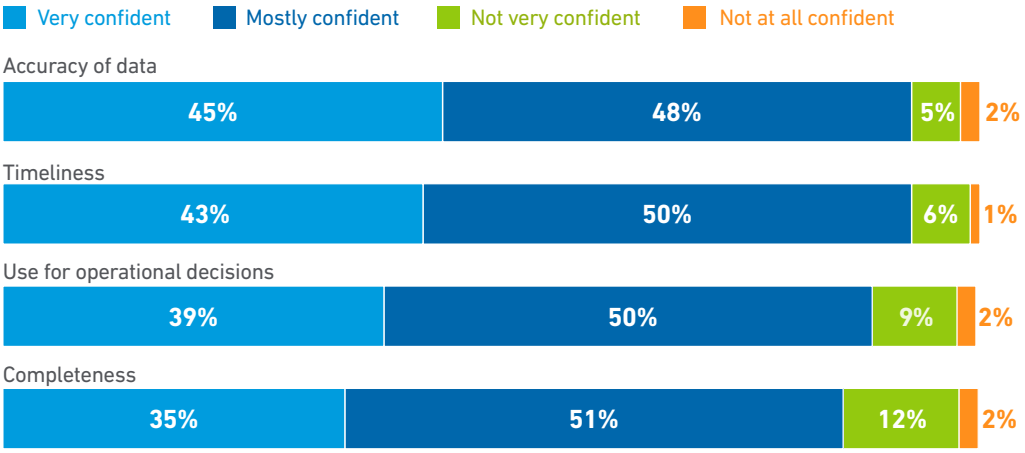
While this is a strong message in favor of more telemetry, the growing use of generative AI and automation may test — or even undermine — this confidence (see section: **Expectations for AI continue to shift**). Operational AI applications depend on high-quality, often real-time, data for functions such as for live monitoring and predictive analysis.

Live telemetry data, for example, is now sought across multiple systems and distributed environments to identify spare capacity and to inform and adapt operations for AI infrastructures. Any problems with data quality could have severe consequences. This high level of confidence is valuable only when it is supported by effective data governance, validation and maintenance processes.

Figure 8

Operators are mostly confident in their operational data

How confident are you in the quality of your data center's operational data collected from sensors, software, and/or equipment? (n=610)



(All figures rounded.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

Expectations for AI continue to shift

In 2026, many data center operators moved from theoretical discussions of generative AI in data center management to early real-world experimentation. Vendors, too, are now turning broad claims about automation, predictive maintenance and optimization into practical tools for operators.

The first wave of AI-based products and services for data center operations was driven primarily by niche players, specialist private companies and startups. Some of these pre-date large language models (LLMs), using more deterministic neural network-based tools.

The next phase is now being shaped by the industry's market leaders. Some of the largest DCIM developers are adding LLM-based functionality to their software suites; major cooling equipment suppliers are building machine learning models into their management tools; and startups that initially focused on individual facility systems are expanding their products to cover more of the data center.

Despite this technical progress, operators' expectations of the benefits of AI in data center operations declined in 2026 (see **Figure 9**). This caution should not be seen as an outright rejection of AI in the data center. Rather, it is likely a reaction to repeated reports of errors made by AI-based systems in enterprise environments.

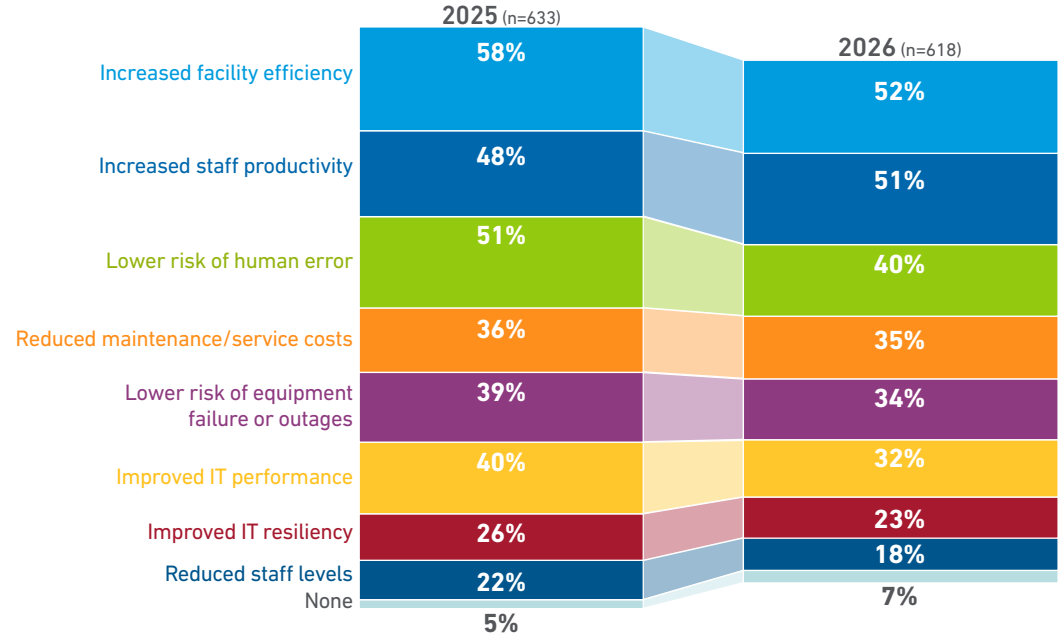
Few organizations have experience with domain-specific AI for data centers, and those that do are not in a rush to share the results of their projects out of fear of giving away a competitive advantage. This lack of public case studies is slowing further adoption, and the situation is likely to persist for some time — especially for the most innovative, and therefore most unorthodox applications. Data center applications of AI are likely to be mission-critical, and the models require rigorous testing.

The three primary drivers for deploying AI-based software in data center operations remain the same as in the previous two years. Increasing facility efficiency remains the leading benefit (52%). Increased staff productivity has risen slightly to 51%, while reducing the risk of human error has seen a notable drop by 11 percentage points, to 40%.

Figure 9

The perceived benefits of using AI in operations

Which of the following — if any — do you consider to be benefits for using AI in your data center operations? Choose all that apply.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

The results suggest a more realistic assessment of the capabilities of AI-based tools. Most survey respondents do not believe these tools will help reduce the risk of facilities or IT outages — and indeed, that is not what most are designed to do. Much of the emerging AI-based functionality aims to improve data center efficiency and staff productivity, and just over half of operators see these as realistic goals.

Overall, trust in AI has declined across most data center use cases, with the proportion of respondents who would not permit any form of AI in their facility increasing by three percentage points since 2025. However, trust continues to vary significantly based on the application, with some use cases remaining highly favored by respondents (see **Figure 10**). Note that this survey did not specify which type of AI might be deployed.

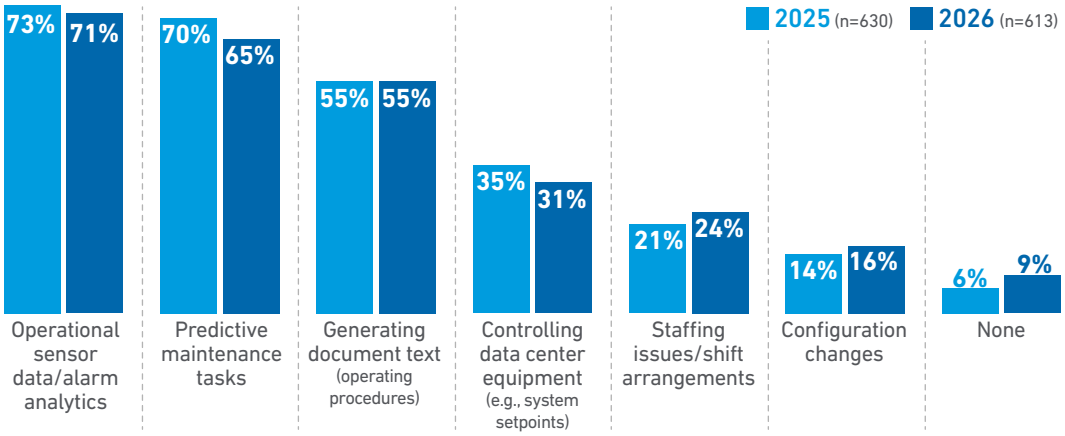
In 2026, 71% of respondents say they would trust an adequately trained AI system to perform automated analytics on sensor data or prioritize alarms. The second most trusted application is predictive maintenance (65%), followed by generation of procedural documentation (55%). Using AI models to set parameters such as cooling equipment setpoints within pre-defined ranges would be permitted by 31% of respondents.

Trust in the least favored application — allowing AI-based systems to make automated configuration changes to mechanical and electrical equipment — has increased slightly to 16%.

Figure 10

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.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

These results are likely to change significantly as operators gain firsthand experience with specialist AI tools rather than basing their perceptions on reports about consumer chatbots. The recent increase in AI-based services from established vendors is likely to accelerate this process.

Resiliency and outages

Uptime's annual survey results for 2026 continue to show gradual improvement in data center outage frequency, the result of years of investment in redundancy, monitoring, operational processes and incident response. It also reflects the changing makeup of the data center population: larger, newer and optimally run data centers make up a larger portion of the sample each year.

However, this does not suggest there is a lower risk of outages in the future. Many of the factors that can drive up the risk of outages — such as reduced or unstable power availability, local grid reliability, supply chain constraints and extreme weather — have intensified. Equally, the scale and potential impact of any outage or disruption has become ever greater as data centers become more central to all aspects of economic and consumer activity.

This changing context means that operators are making steady yet incremental gains in resiliency, but they are doing so against an expanding risk environment and with the threat of consequences extending beyond individual facilities.

Tracking outage data is inherently difficult (see *Annual outage analysis 2026*), and the increasingly distributed nature of the infrastructure, the applications, and the impact of outages adds complexity and uncertainty. Many outages are partial, ill-defined and widely distributed; some involve multiple parties.

Despite these data collection challenges, findings from the 2026 annual survey remain broadly consistent with previous years: fewer respondents say they suffered at least one significant outage in the past three years, the proportion of outages that are serious or severe is largely stable, the leading causes of impactful incidents are unchanged, and the cost and complexity of outages continue to rise.

However, the past does not predict the future: the growing number of external threats and the increasing pace of technical change within data centers may slow or halt this recent progress, at least temporarily.

Outages decline, but challenges persist

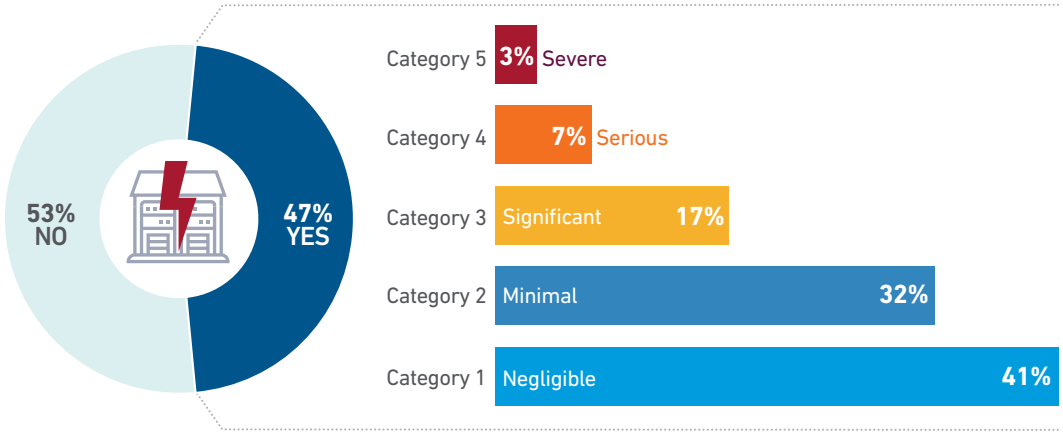
For the sixth year in a row, Uptime's annual survey results show improvement in reported outage frequency. The share of respondents who experienced an outage in the past three years fell by three percentage points from 2025, continuing the steady reduction seen since 2020 (see **Figure 11**). However, this pace of improvement has slowed since 2023, and nearly half of respondents still experienced at least one outage over the three-year period (see **Figure 12**).

Among respondents who experienced an outage, the share reporting a serious or severe incident (those that caused the most significant damage) also remained unchanged at 10%. However, as organizations become more dependent on digital infrastructure, these outages may carry greater operational, financial and reputational impact than in the past. In other words, perceptions of what constitutes a "serious" or "severe" outage may change over time, and an event considered serious in 2015 may appear relatively immaterial by today's standards.

Figure 11

Fewer than half report an outage 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=715)



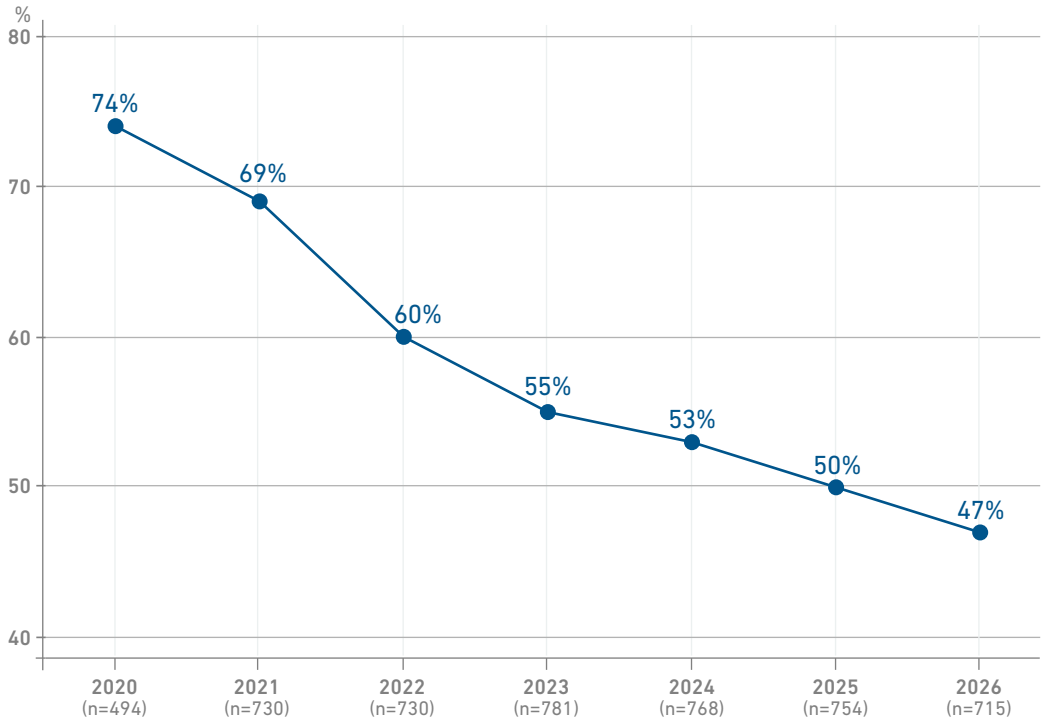
(Responses for "Don't know" not included.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

Figure 12

Outage frequency continues its gradual decline

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



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2020-2026

The findings in Uptime's annual survey are largely supported by separate Uptime research studies, which have found that resiliency risk is increasingly shaped by factors outside the direct control of individual operators (see *Annual outage analysis 2026*). While these factors — such as grid reliability, extreme weather and third-party service dependencies — may not directly translate into more outages in a given survey year, they can make outages harder to prevent, more difficult to contain and more costly to recover from.

Operators may therefore be making incremental progress in reducing outage frequency, while the consequences of the most damaging incidents may be intensifying. As digital services become more interconnected, business-critical and dependent on external infrastructure, a stable severity profile can still mask rising financial and operational impacts.

Outage costs continue to rise

Although the proportion of outages in each severity category has been stable, Uptime's annual survey indicates that the costs of outages continued to rise for the third consecutive year. This reinforces the distinction between outage severity and outage consequence described in the previous section.

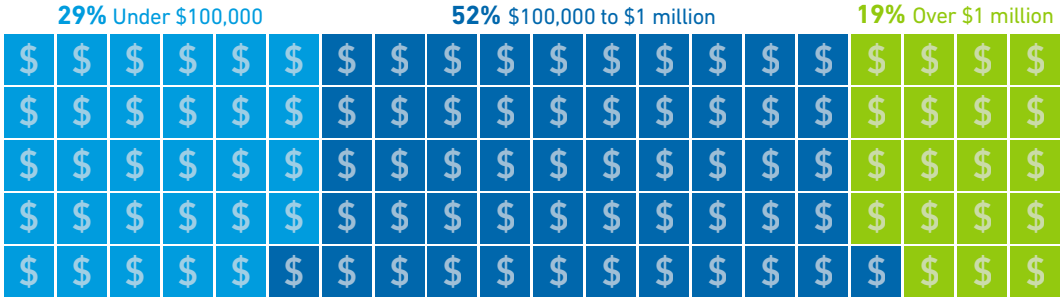
In 2026, 29% of operators who experienced an outage say their most damaging incident cost less than $100,000, down from 43% in 2025 (see **Figure 13**). At the same time, the share who report costs between $100,000 and $1 million increased from 37% to 52%.

The share of respondents who report outage costs of more than $1 million did not change meaningfully in 2026, moving down only one percentage point from 2025. Instead, the main shift in the data is away from relatively low-cost outages and toward the middle-cost range. Overall, 71% of respondents report that their most damaging outage cost at least $100,000, compared with 57% in 2025.

Figure 13

More outages cost over $100,000

Estimate the total cost of this downtime incident (from outage to full recovery), including direct, opportunity and reputation costs using the following options. (n=77)



(Responses shown are only from those who reported experiencing either a significant, serious or severe outage in the past three years. All figures rounded.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

Many overlapping factors explain the overall increase in outage costs. Direct repair and recovery costs are rising as labor, equipment and replacement components become more expensive. At the same time, the business impact of downtime is increasing as more organizations depend on digital infrastructure. Service level agreements (SLA) penalties, customer compensation, delayed transactions, lost productivity and reputational harm can all increase the cost of an outage, even when its operational severity appears unchanged.

Some outages cost many millions of dollars, largely because of the business losses experienced by those who own or operate the affected applications. In extreme cases, outages can lead to a material loss of revenue, compensation payments, large insurance payouts, and even a drop in company valuation.

These growing costs place greater importance on financial impact analysis in resiliency planning. Operators who analyze how outage costs accumulate across recovery work, SLA exposure, customer disruption, delayed operations and reputational damage are better positioned to make realistic assessments of the financial fallout from failures.

The causes of outages

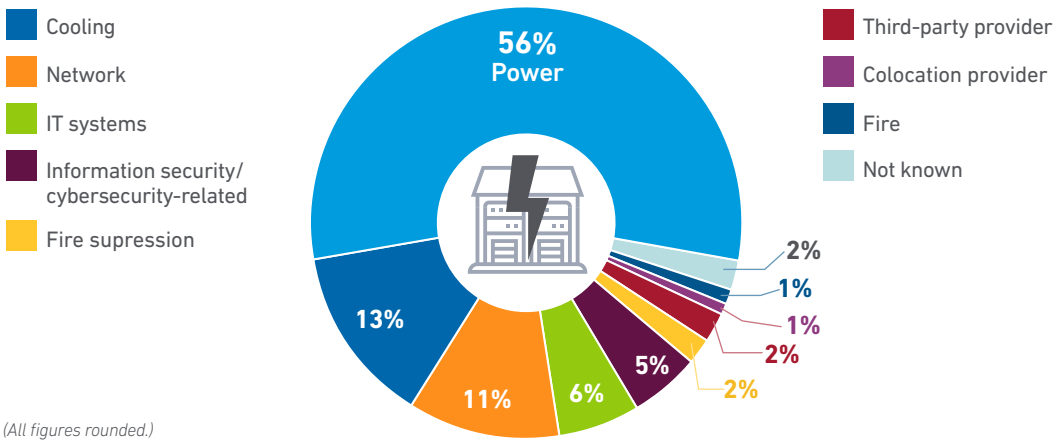
Outage cause data should be treated with caution: outages often result from multiple interrelated factors, and the root cause of an incident can be difficult to isolate. For example, an external power disturbance may expose a UPS weakness, a cooling issue may be exacerbated by monitoring or maintenance gaps, or a third-party failure may disrupt services even if the facility itself continues to operate as designed.

Uptime's annual survey continues to show that power is the leading identified cause of respondents' most recent impactful data center outage — a finding that has been consistent since 2020. In 2026, 56% of respondents who experienced an outage in the previous three years say power was the primary cause of their most damaging incident, up from 45% in 2025 (see **Figure 14**).

Figure 14

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=85)



(All figures rounded.)

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

This increase is notable, but it should be interpreted carefully. The 2025 result appears unusually low compared with the longer-running pattern since 2023, in which power has accounted for more than half of respondents' most recent impactful outages. As a result, the 2026 finding may represent a return to the historical baseline rather than a sudden deterioration in power resiliency.

Additional findings from the Uptime Institute Resiliency Survey 2026, which explores root causes behind IT service outages (including data center outages), reinforce the persistence of challenges around power infrastructure. Among power-related IT service outages, UPS failures remain the most commonly cited driver, at 39% (n=129). This is consistent with previous years, but the 2026 results also show growing shares of outages linked to transfer switch (34%) and generator (28%) failures (see *Annual outage analysis 2026*).

Power failures can occur at any point in the electrical chain — from backup power and transfer systems to distribution and the IT equipment being supported. Although grid failures can reveal on-site problems, they are not counted as a cause of outages; Uptime does not consider the grid to be the primary source of power-related failures.

As data centers become more power-intensive and operators face greater exposure to utility capacity, interconnection timelines, grid congestion, fuel logistics and equipment supply chains, the causes and types of power failure may change — especially if more operators use on-site power. Some of these issues sit partly outside the traditional boundaries of the data center, but they increasingly affect data center resiliency directly.

The persistent dominance of power as an outage risk should not lead operators to discount non-power threats. While the data shows little year-over-year change, risks from cooling failures may become more complex as facilities support higher-density workloads and new cooling architectures. IT, networking and cybersecurity risks remain important as operations become more software-defined and dependent on interconnected systems — and as GPUs strain capacity limits and introduce workload and electrical volatility.

Building more resilient systems

In 2026, operators continue to rely mostly on redundant architectures to reduce the risk of equipment failure and service disruption. The annual survey results show that 44% of respondents use 2N as their most common level of redundant power equipment (up two percentage points from 2025), while another 40% report N+1. High equipment costs and space constraints may encourage greater use of N+1, especially in newer, high-density facilities.

Good design and architecture do not, of course, prevent all outages. The value of redundancy depends on whether systems are configured, maintained, tested and operated effectively. In 2026, 89% of respondents who experienced an impactful downtime incident say it could have been avoided with better management, processes or configuration, up from 87% in 2025 and 80% in 2024.

This finding reinforces a conclusion that remains highly relevant from previous analyses: physical redundancy must be paired with operational discipline. It should also be viewed alongside software-based resiliency strategies, such as distributed applications, multisite availability zones, automated failovers and platform-level recovery mechanisms. These approaches can improve availability, but they also highlight the importance of strong processes, attentiveness to all causes of failures, robust configuration and change management, continuous staff training and regular testing.

Workload placement

Every year, Uptime asks its enterprise respondents to roughly estimate what percentage of their IT workloads are run in-house versus in third-party venues. The data helps to show, at a high level, where these important workloads are carried out, and the relative use of different venue types. It does not, however, show the overall work handled by data centers, as many services, such as software and consumer applications, are likely not captured. Nor does it measure the scale of this work (such as by megawatts).

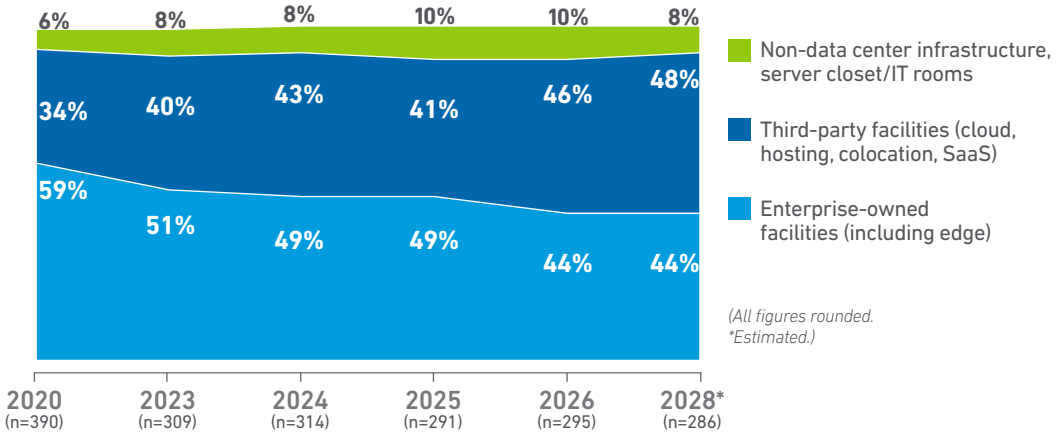
Most corporate IT is off premises

Third-party data center facilities and services account for 46% of IT workloads in 2026, compared with 44% in enterprise-owned corporate data centers. This is the first time in Uptime's annual survey history that third-party data center venues have held the larger share (see **Figure 15**).

Figure 15

Workload placement by data center ownership model

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 2020-2026

In this analysis, third-party facilities include colocation, the public cloud, software as a service (SaaS) and hosting services. Edge and micro data centers are treated as enterprise-operated, as the survey does not specify their ownership or operating model. If they were instead classified as third-party-operated, the share of workloads hosted in third-party infrastructure would be even higher.

IT rooms and server cabinets are not considered data centers because they lack the key features of dedicated facilities. Even so, it is surprising that their use has not declined more sharply given the growing use of online workloads.

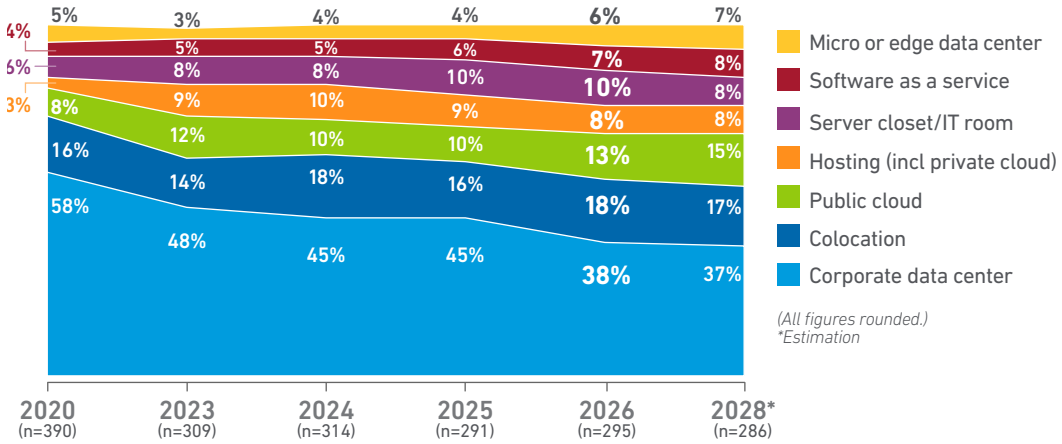
This shift between infrastructure ownership models does not represent an existential threat to corporate data centers. Rather, it reflects the continued normalization of hybrid IT, in which organizations place workloads according to technical, operational, regulatory and financial requirements.

Between 2020 and 2026, the share of workloads hosted in colocation and the public cloud has risen from 24% to 31%. Hosting and SaaS have also expanded over that period, although their combined share has levelled off at around 15% in recent years (see **Figure 16**). Over the same period, the share of workloads hosted in corporate data centers has fallen from 58% to 38% (excluding IT rooms and server cabinets).

Figure 16

Workload placement by venue

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



PTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2020-2026

Hybrid IT strategies explain this fragmented pattern of workload distribution. Organizations are increasingly choosing the most suitable venue for each workload, rather than defaulting to a single operating model (see *Cloud scalability and resiliency from first principles*). Improved hybrid IT management software tools have done much to enable this flexibility. At the same time, container-based architectures have improved workload portability.

As a result, some workloads have already moved, or are likely to move, from private facilities to cloud and colocation environments. These venues offer faster access to capacity, more flexible pricing and a way to manage persistent cost pressures.

Even so, few large organizations are comfortable moving all workloads out of their own data centers. Data sovereignty, fears of vendor lock-in, and regulatory requirements remain major constraints.

Since 2023, data security and regulation have consistently ranked as the two biggest barriers to cloud adoption in the survey. Over the past two years, concern has intensified: the share of respondents citing data security as an obstacle has risen from 60% in 2024 to 68% today, while the share citing regulation has increased from 44% to 55%.

Geopolitical realignment is also shaping venue decisions. US-based hyperscale cloud providers face increasing scrutiny from some customers and regulators in Europe and other regions over the secure and compliant storage of data (see *What cloud sovereignty really means*). These concerns have not halted cloud adoption, but they have encouraged enterprises to assess workload placement more carefully.

At the same time, strong investment in AI is increasing demand for data center capacity, GPU infrastructure and specialist operational expertise. Colocation, hosting, SaaS and cloud services can offer faster access to these capabilities, adding further momentum to third-party adoption (see *Where to deploy AI training: a guide to the economics*). Over time, AI is expected to shift more workloads to cloud services, whose providers have the scale, budget and operational commitment to support higher densities.

The short-term outlook points to a further modest decline in the share of workloads hosted in corporate data centers. Respondents forecast that 48% of workloads will be hosted in third-party facilities and services by 2028, up two percentage points from today, while the share hosted in non-data center infrastructure is projected to fall from 10% to 8%. Historically, enterprise operators have often overestimated the percentage of workloads they expect to move away from their corporate data centers.

Over the longer term, enterprises are likely to continue using a mix of facilities and services. AI demand, the need for rapid scalability and continuing cost pressures all support the case for cloud and colocation.

At the same time, geopolitical uncertainty, security concerns and regulatory requirements mean that most enterprises are unlikely to outsource everything. A hybrid approach gives organizations the flexibility to pause, adapt or change direction as conditions evolve.

Staffing

For more than a decade, the data center industry has consistently faced shortages of skilled labor, and many operators have also struggled to retain staff. This is not surprising: the industry has been expanding so rapidly that recruiters have inevitably struggled to keep pace, especially given the specialist nature of many of these roles.

In recent years, however, the severity of the talent shortage has somewhat stabilized, neither worsening nor improving. This year's data suggests that the industry's recruitment challenges are beginning to intensify again.

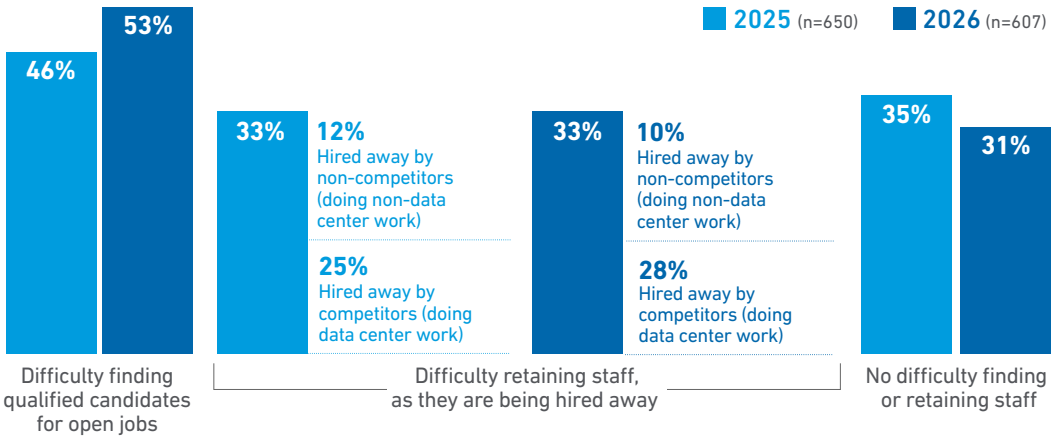
Recruitment challenges rise again

As operators contend with surging demand for data center capacity, they face an unprecedented need for skilled labor. Large, specialized AI facilities compound this challenge because they require more staff for major new projects and evolving skill sets to support higher thermal densities. This year, more than half (53%) of operators report difficulties finding qualified candidates for vacant roles — a statistically significant increase from 46% in 2025 (see **Figure 17**).

Figure 17

Staffing difficulties persist in retention, intensify in hiring

Which of the following staffing conditions apply to your organization? Choose all that apply.



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2025-2026

The ongoing talent shortage arises from both growing demand for skilled workers and the retirement of experienced senior operations staff. At the same time, the industry's pipeline of new talent remains insufficient, leaving data center operators and their expansion plans under increasing pressure.

The parallel challenge of employee turnover continues on a broadly flat trend this year. More than a quarter (28%) of respondents say they have had staff hired away by competitors, while one in ten report employees leaving the data center business altogether. Overall, one in three (33%) cite one or both of these retention issues, highlighting the potentially significant costs faced by a large minority of operators. For years, many operators have responded by increasing salary spending, both to retain existing employees and to recruit trained candidates from competitors.

Ambitious capacity expansion plans require operators to recruit and train staff from diverse sources, including universities, trade schools and industries with overlapping skill sets. Recruitment efforts may also benefit from the active pursuit of underrepresented demographic groups. While operators have ample motivation to hire from all viable sources, formal data center mentoring programs have shown little growth in recent years.

Recruitment challenges can be very region-specific and, in some cases, cultural. Uptime's tracking of the gender imbalance in data center staffing suggests that women are an underutilized labor pool, and they are likely only one of many. Data center design, build and operations teams worldwide remain heavily male-dominated. Nearly three-quarters of respondents (74%) still say that women make up around 10% of their data center staff, while more than half (54%) estimate the ratio at 20:1 — equivalent to just 5% women.

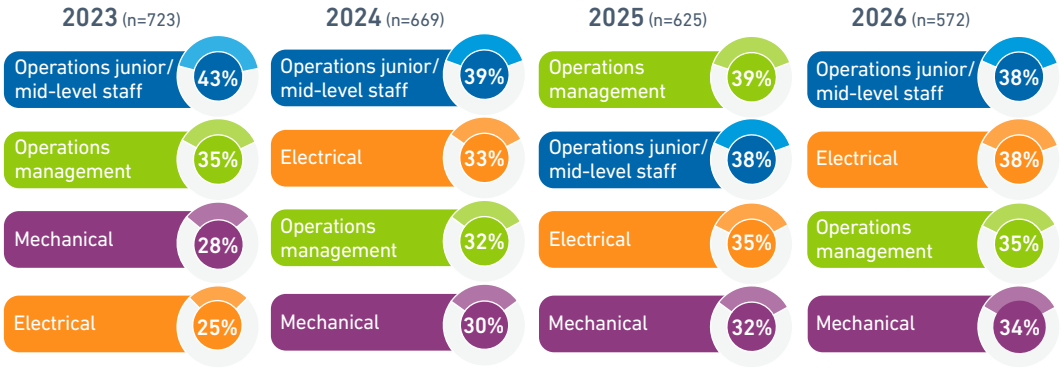
Staff skills remain strained

Each year, Uptime gathers data on skills gaps across the major data center job categories. Shortfalls persist across all roles, although their severity varies. In 2026, operators report the greatest skills gaps in electrical (38%), junior level operations (38%), operations management (35%) and mechanical (34%) roles (see **Figure 18**).

Figure 18

Electrical needs intensify within broad skills gaps

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



*(*Only the top four response categories are shown.)*

UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2023-2026

Operators' recruitment struggles are pervasive, and no single job category is clearly in crisis. However, the areas of greatest need have shifted noticeably in recent years. Since 2023, reported electrical skills gaps have increased by 13 percentage points, bringing them level with junior/mid-level operations in the top slot (38%). The ongoing construction of very large data centers amounts to gigawatts of new electrical infrastructure being built, commissioned and operated. This, in turn, brings a corresponding large demand for electricians at a time when many regions are already facing shortages of qualified electrical workers.

To improve access to electrical and other skilled trades, data center hiring managers may want to expand past successful partnerships with local trade schools and community colleges. Companies that fund education and training programs can strengthen their ties with local communities and build a more consistent talent pipeline.

Appendix

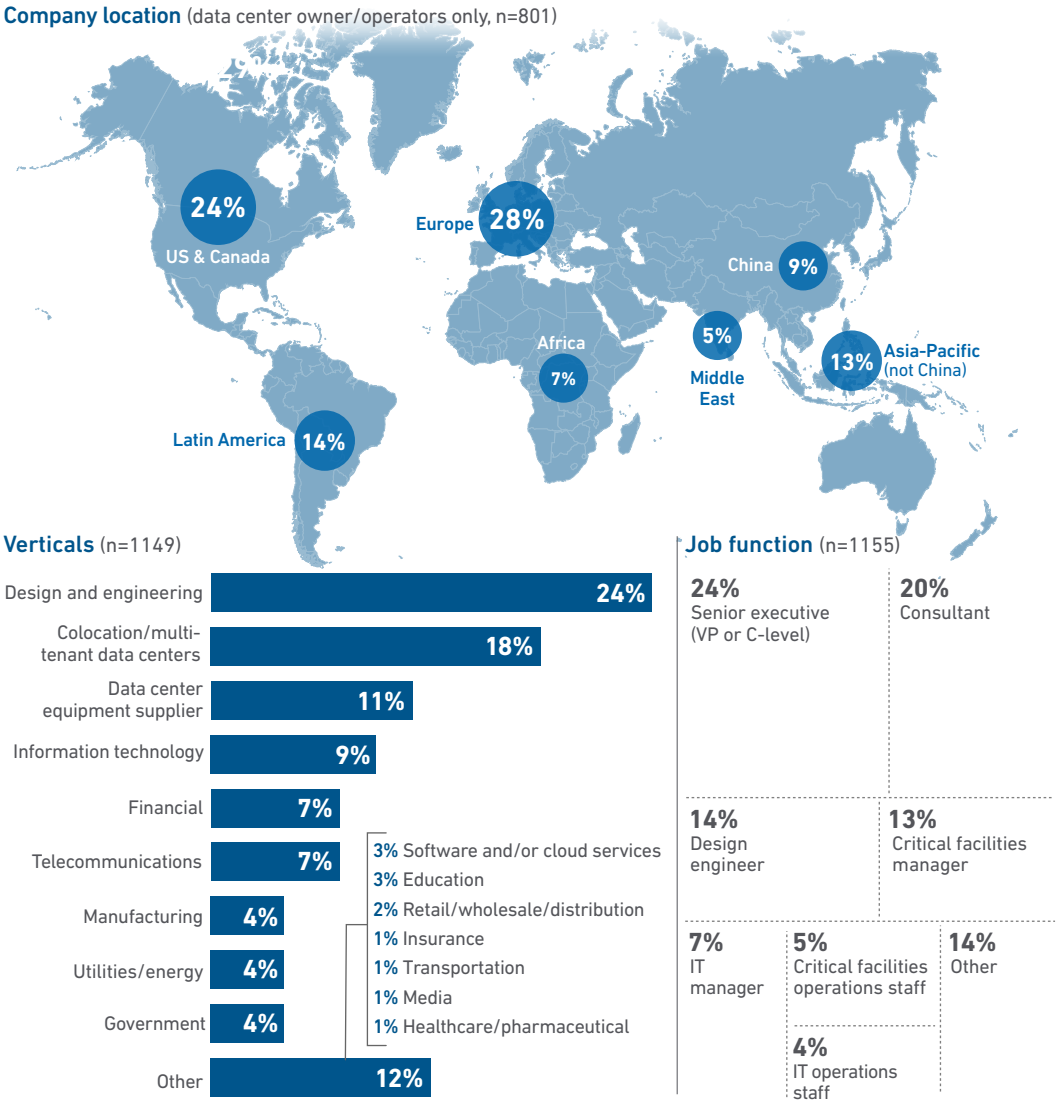
Survey methodology and demographics

Uptime Institute's Global Data Center Survey, now in its 16th year, is conducted annually online and by email. The 2026 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 19

Respondents by location, industry vertical and job function



UPTIME INSTITUTE GLOBAL DATA CENTER SURVEY 2026

Appendix (*continued*)

The participants represent a wide range of industry verticals in multiple countries. More than half (52%) are located in North America and Europe. Nearly one in four 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 801 end users registered for the survey and answered at least one question. The number of respondents (n) 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

John O'Brien

John is Uptime Institute's Senior Research Analyst for Cloud and Software Automation. As a technology industry analyst for over two decades, John has been analyzing the impact of cloud migration, modernization and optimization for the past decade. John covers hybrid and multi-cloud infrastructure, sustainability, and emerging AIOps, DataOps and FinOps practices.

jobrien@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. For over 30 years, the company has established industry-leading benchmarks for data center performance, resilience, sustainability, and efficiency. These benchmarks assure customers that their digital infrastructure will perform consistently with their business needs under any operating conditions. With more than 4,300 awards issued globally and over 1,100 active projects in 120+ countries, Uptime has helped tens of thousands of companies optimize critical IT assets while managing costs, resources, and efficiency. The organization's Tier Standard is the IT industry's most trusted and adopted global standard for data center design, construction, and operations. Core offerings include Tier Certifications, Management & Operations reviews, AI Infrastructure Advisory, Data Center Risk Assessments, Energy & Sustainability Assessments, Cybersecurity Assessments & Ratings, and IT Advisory Services.

Uptime Education programs have been completed by more than 100,000 data center professionals, establishing Uptime as the Learning & Development partner of choice worldwide. Uptime Institute is headquartered in New York, NY, with offices in London, Sao Paulo, Dubai, Riyadh, Singapore, and Madrid, and full-time Uptime professionals based in over thirty-four countries around the world.

For more information, visit www.uptimeinstitute.com