Uptime Institute® | RESEARCH

ADVISORY
2018

PLANNING & STRATEGY

Uptime Institute Global Data Center Survey

Operators struggle with constraints, change, and complexity

ADVISORY I 2018

Uptime Institute Global Data Center Survey

The data center industry has made significant efficiency gains in recent years, yet the rate and severity of outages are disturbingly high. The Uptime Institute's annual survey of almost 900 data center operators and IT practitioners worldwide reveals some of the key forces at play.

- Efficiency–and outages–increase
- Edge computing is coming
- Issues include rack density, climate change, and staffing

The Eighth Annual Uptime Institute Data Center Survey provides an overview of the major trends shaping IT infrastructure delivery and strategy. The survey was conducted via email between February and May 2018. It includes responses from 867 data center operators and IT practitioners globally from enterprise and service provider facilities.

For the better part of a decade, operators have agonized over how to reduce energy waste. By and large, they have succeeded–the average PUE, the industry's most common infrastructure efficiency metric, was a record low in this year's survey.

Most respondents believe that their hybrid data center approach–a mix of off-premises and privately owned on-premises capacity–has made their IT operations more resilient. If this is the case, it is not supported by the evidence: The number of respondents that experienced an IT downtime incident or severe service degradation in the past year (31%) increased over last year's survey (about 25%). And in the past three years, almost half of our 2018 survey respondents had an outage. This is a higher-than-expected number.

Almost 80% said their most recent outage was preventable, with on-premise power failures, network failures, and software or IT systems errors being the most common primary causes. For most, it took one to four hours to fully recover, with one third of respondents reporting a recovery time of five hours or longer. (For more details on outages data, including cause and effect, please see our report: 'Uptime data shows outages are common, costly, and preventable.')

ADVISORY I 2018

Uptime Institute Global Data Center Survey

Why are so many data centers vulnerable to outages and why are they so severe? Most operators have or are in the process of deploying purpose-built management tools, and most say they are operating their facilities within the boundaries of their expected lifespan. In this overview report, we delve into some of the key challenges and complexities facing data center managers globally.

KEY FINDINGS

- **Efficient but more complex:** PUE across the industry has improved, signaling an improvement in data center efficiency. Today, operators are grappling with new challenges, including the business case and cost of hybrid IT approaches.
- **Edge computing is coming:** Operators are expecting to deploy significant new edge computing capacity, which will add a layer of operational and management complexity.
- **Rack density issues are growing:** The highest rack densities reported at enterprise and service provider data centers suggests that many operators face cooling challenges.
- **DCIM is now mainstream:** A small majority of data centers now have some type of DCIM, and typically their implementation has been successful (contrary to widespread industry reports).
- **Many data centers are unprepared for, or their managers are not expecting to be affected, by climate change:** Despite being vulnerable to increased temperatures, water shortages, and extreme and sometimes disastrous weather events, most data center operators have determined that they either won't be impacted or are ignoring the problem.
- **Data center skill shortages will intensify:** In this aging and overwhelming male sector, most operators are struggling with staffing issues. Yet most do not believe a lack of diversity in their ranks is an issue to be concerned about.

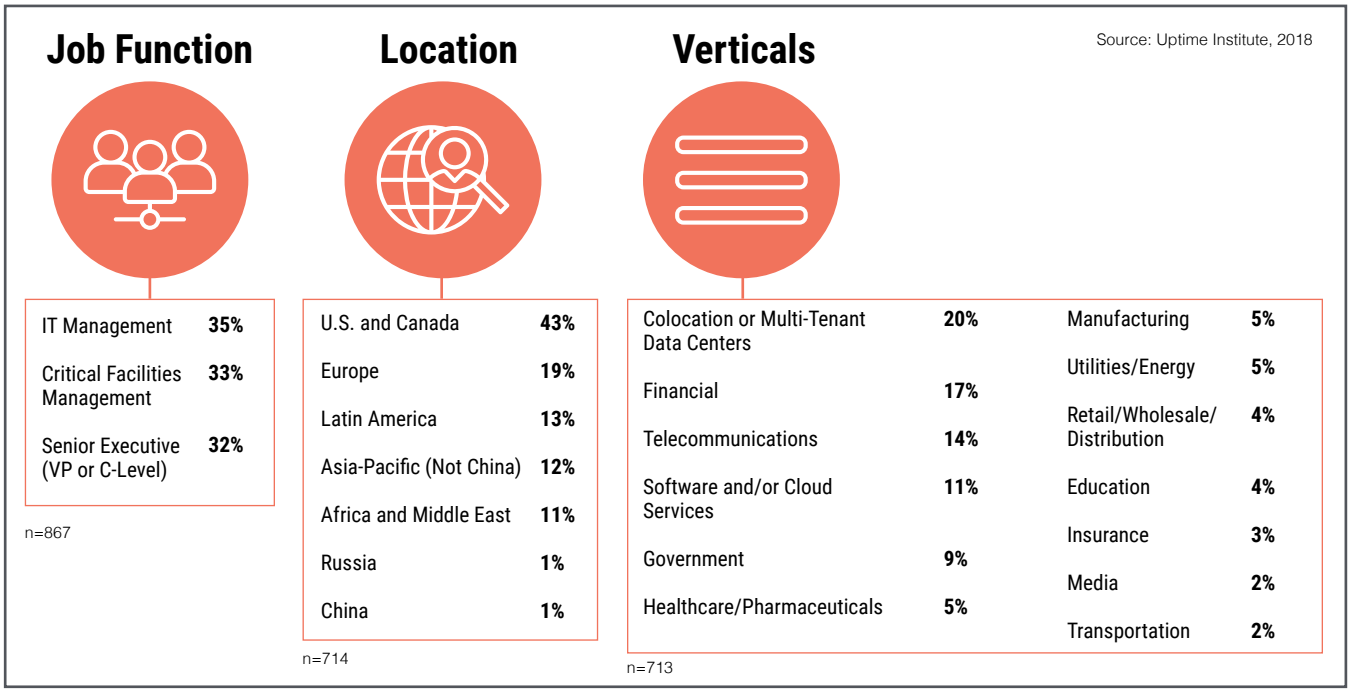
Uptime Institute Global Data Center Survey

This Uptime Institute Research report is an overview of the results from Uptime Institute's 2018 Global Data Center Survey. The survey includes additional insights not covered in this overview, including the cause and effect of data center outages, power and cooling redundancy configurations, the impact of legacy IT assets, power generation and energy storage adoption rates, and data center technology trends, including among multi-tenant data center customers.

Clients of Uptime Institute and members of the Uptime Institute Network can obtain additional drill-down reports and data cuts by contacting Rhonda Ascierto, rascierto@uptimeinstitute.com. If you are not a member or client but are interested please contact Matt Stansberry, mstansberry@uptimeinstitute.com.

Survey demographics

The survey respondents are end-users–people responsible for managing infrastructure at the world's largest IT organizations. The participants represent a wide range of industries, with about a 70-30 split between enterprise IT managers and service providers (those with operational or executive responsibilities for a third-party data center, such as colocation, wholesale, software, or cloud computing services).



Uptime Institute Global Data Center Survey

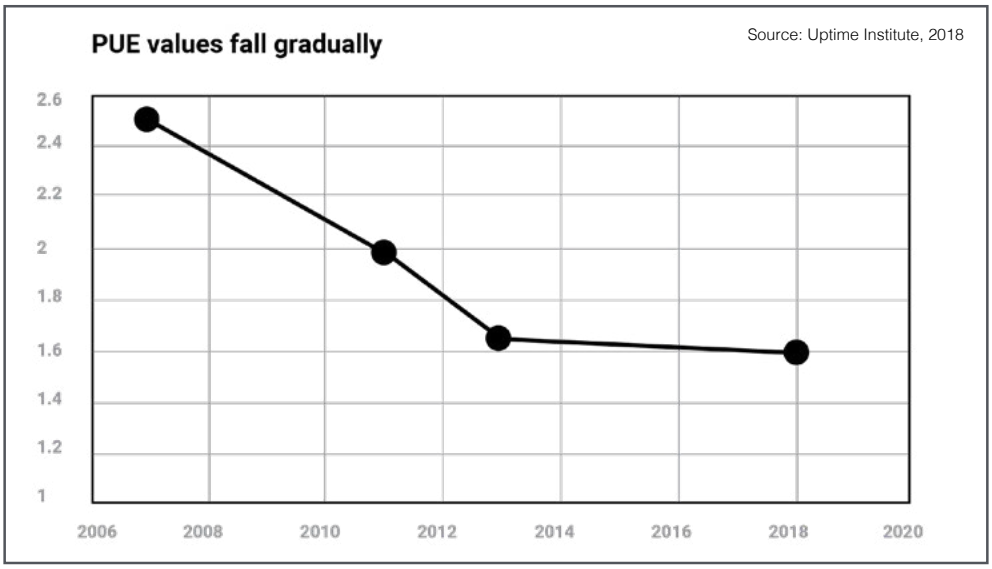
The roles of the participants range from IT and facilities management to executive, with senior-level participants at the VP level and above. Multiple geographic regions are represented, providing a global perspective. Most respondents (62%) are from the large, mature data center markets of the U.S., Canada, and Europe, with the balance spread across a mix of emerging data center markets and smaller regions.

Note: respondents were not required to answer all questions, which means the number of respondents for individual questions ('n') varies.

Efficient but more complex

The Green Grid's power usage effectiveness (PUE) metric has become the industry-preferred metric for infrastructure energy efficiency for data centers. PUE does not take into account the efficiency of IT or networking equipment (or software applications), and it was designed only to benchmark efficiency gains over time at an individual site rather than to compare one facility against another. While the data center industry has long acknowledged these limitations of the metric, PUE remains the *de facto* efficiency standard and one that the Uptime Institute has been tracking for over the past decade.

In 2007, average PUE was 2.5 in a survey of Uptime Institute's Network Members (a global user group of large data center owners and operators). This improved to 1.98 in our inaugural industry survey in 2011, and to 1.65 in our 2013 survey. Improvements since then have been incremental; in 2018, the average PUE was 1.58.



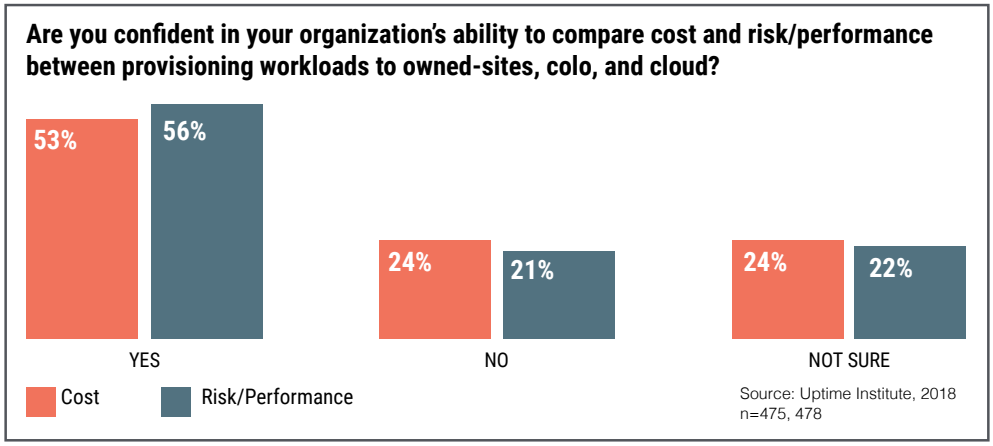
Uptime Institute Global Data Center Survey

One takeaway is that the biggest infrastructure efficiency gains happened five years ago. Further improvements will require significant investment and effort, with increasingly diminishing returns. Organizations will continue to increase efficiency in a bid to lower operating costs or to maximize available power (or both), including with artificial intelligence-driven data center management as a service (DMaaS), software-defined power, and other approaches. But Uptime Institute advises that organizations should concentrate on total energy consumption related to their IT load, rather than solely on mechanical/electrical losses.

For most data center operators, especially at enterprises, the key challenges of today are different from five years ago, driven by increased change and complexity. Specifically, managers must effectively manage the proliferation of hybrid IT architectures, defined as a mix of on-premises data center capacity and off-premises resources such as colocation, cloud, and hosting. Hybrid IT is now the norm, creating technology, organizational, and management complexity.

The ongoing march toward digital transformation across society and business will continue to shape new composite data center approaches. A goal for organizations is placing workloads in the 'best execution venue' according to cost, availability, compliance, and other factors–yet clearly there is much work ahead.

Many survey respondents struggle to assess the business case and effectiveness of their hybrid approaches. Only about half are confident in their organizations' ability to compare costs and risk/performance across their on-premises, colocation, and cloud facilities.



Uptime Institute Global Data Center Survey

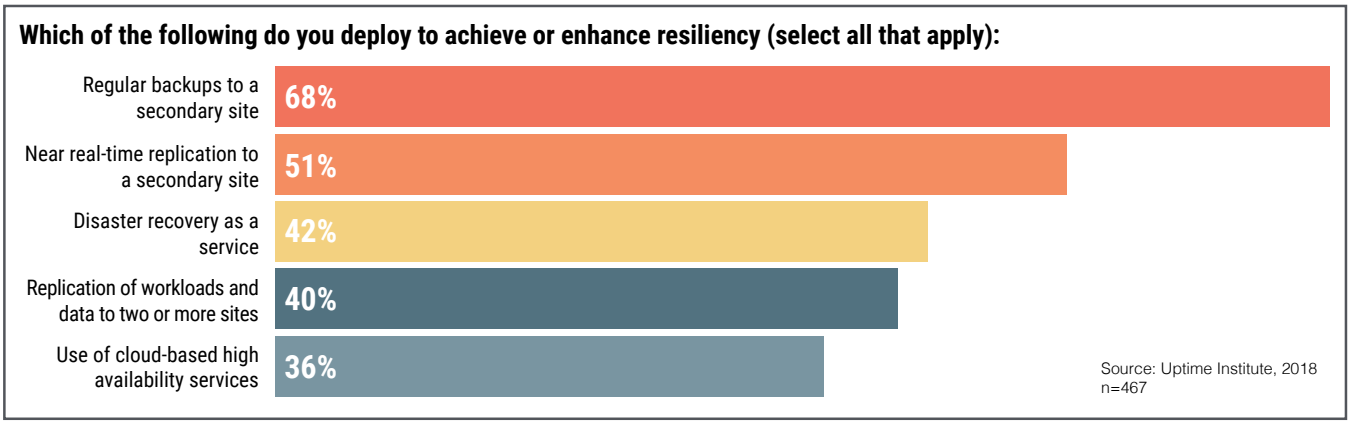
The need for end-to-end visibility and management across different infrastructures is one of the most common and greatest challenges enterprises say they face, and, we believe, a driver for data center infrastructure management (DCIM) software deployments (see “DCIM is now mainstream” below). This represents an opportunity for suppliers and colocation companies, which can offer services to help with this challenge.

Is having workloads spread across on-site, colocation, and cloud deployments making overall IT more resilient? More than half of respondents (61%) said that it has, despite about one third of them having suffered an outage in the past year. About 30% said their hybrid approach has made them less resilient and 9% did not know (n=475).

It is clear is that managing multiple types of data center environments creates increased complexity. We hear this often directly from operators, and previous 451 Research surveys support this conclusion.

One obvious area for improvement is assigning ownership of the issue: only about half of respondents (49%) have a single department head or executive who is charged with resiliency across their various on-premises, cloud, and colocation assets. (Note that Uptime Institute now offers an advisory service, the Hybrid Resiliency Assessment, to help organizations identify possible vulnerabilities).

While the results may be patchy at this stage, it is clear that hybrid approaches are being leveraged to achieve or enhance IT resiliency. While most are taking traditional approaches, such as regular backups and/or replications to other sites, a notable portion are using cloud-based, high-availability services (40%) and disaster recovery as a service (35%). We expect this trend will accelerate.



Uptime Institute Global Data Center Survey

By analyzing just those survey respondents with 2N architectures for cooling and power, we are also able to see if N+1 architectures, sometimes viewed as riskier, are any less reliable.

The results were clear: Of those with a 2N architecture, 22% had experienced an outage in the past year, rising to 35% in the past three years (this is for all failures, not just on-site facility failures). But those with an N+1 architecture did not fare so well: 33% said they had an outage in the past year, rising to 51% in the past three years–higher than the overall average.

This response is clear and given that 2N redundancy costs more and is designed to reduce failures, perhaps somewhat expected: 2N architectures, according to the data, are more effective at preventing outages than N+1 solutions. The data, however, may change in the years ahead, as N+1 systems, aided by management software, become more effective.

Edge computing is coming

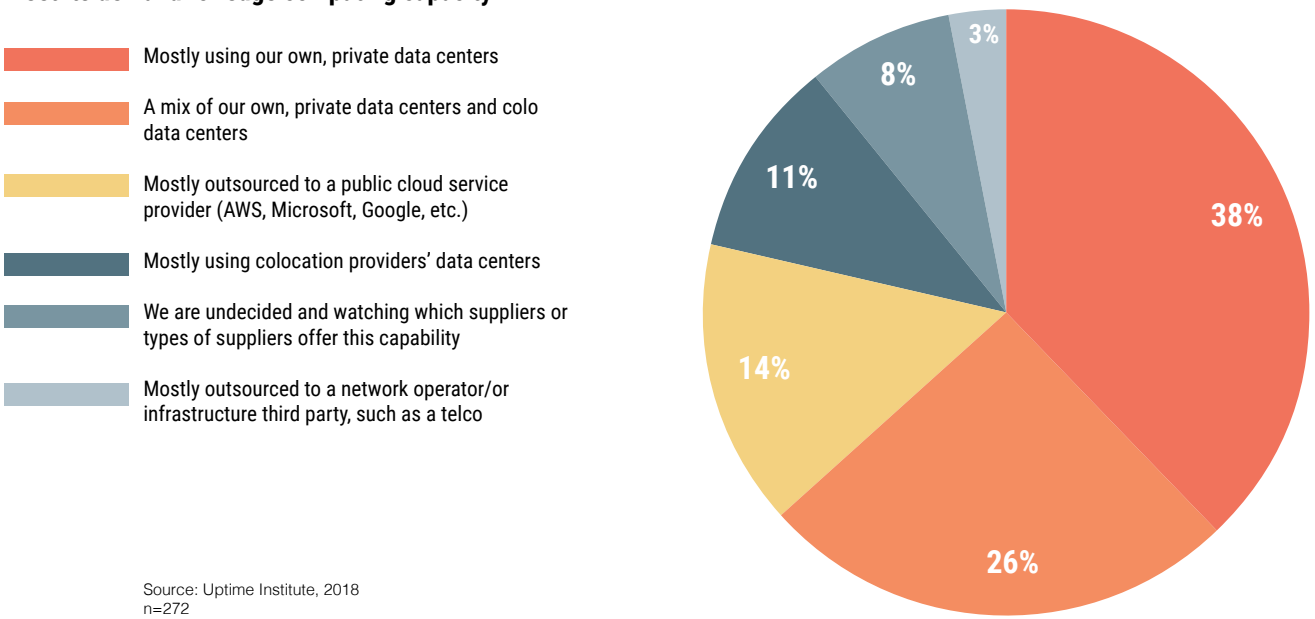
An anticipated build out of new edge computing capacity will add a new layer of operations and management complexity in the years ahead. More than 40% of respondents expect their organization will require edge computing capabilities, defined as requirements that will necessitate processing data closer to the source of its generation/use.

Just 27% of respondents said they do not anticipate additional edge computing requirements. Tellingly, 30% said they are not yet sure, which we believe is a reflection of the immaturity of next-generation edge applications such as for the Internet of Things (IoT). While IoT promises to be a high-growth opportunity for edge data centers, there is a wide range of vertical use cases, and most deployments are still in the early stages. IoT technology standards, including around security, and supplier ecosystems are evolving. Given the large volume of emerging IoT applications, we would not be surprised if the majority of organizations in the future will require new edge data center capacity.

Most survey respondents who anticipate the need for new edge capacity say they plan to use their own private data centers (37%) or a mix of colocation and their own data centers (26%). A smaller portion plan to mostly outsource–with slightly more favoring a public cloud service provider to handle their edge requirements versus using colocation data centers. This supports the view of Uptime Institute Research that data center capacity in enterprises will not so much shrink but fragment from large and often inefficient data centers into distributed “fleets” of smaller, hardened data center “nodes.”

Uptime Institute Global Data Center Survey

[For those who do anticipate that their organization will require edge computing capabilities] How will your organization meet its demand for edge computing capacity?



Source: Uptime Institute, 2018
n=272

As more edge applications are introduced and mature, it is likely that the 'edge' will expand and move out beyond the proximity of their own private facilities, driving demand for data center capacity in multi-tenant facilities and in strategically placed, small distributed, micro-modular data centers.

Micro-modular data centers are IT cabinets that are encapsulated in their own protective shell with built-in cooling, connectivity, physical security, shock absorption, and, when required, uninterruptible power supply. They have a smaller footprint than 'macro' room environments and 'plug-and-play' installation, making them a fit-for-purpose form factor for distributed IT at edge locations.

There is lively debate in the industry as to whether many or most micro-modular data centers will be shared infrastructure, either as a managed hosted, public cloud, or colocation provider service offering (aaS), or as privately owned capacity. It is likely that all scenarios will play out in the coming years.

Uptime Institute Global Data Center Survey

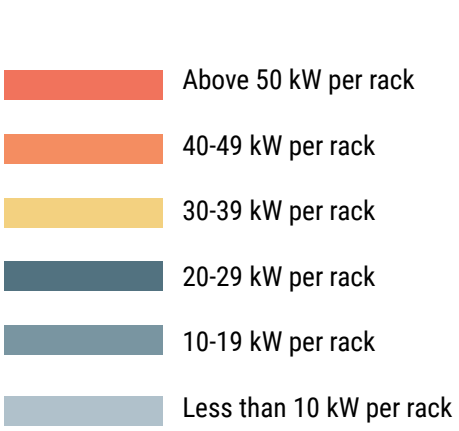
Rack density extremes climb higher

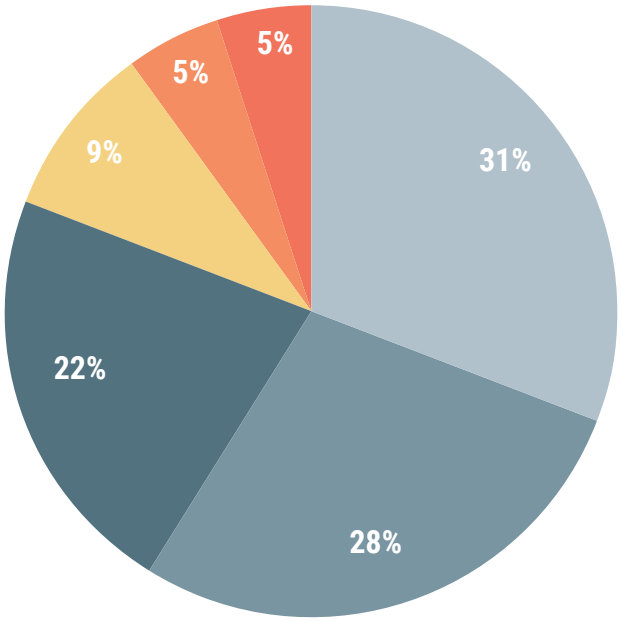
Average rack densities in data centers remain fairly low. In our 2017 survey, 67% of respondents reported an average of below 6 kilowatt (kw) per rack. Just 9% had average densities of 10 kW per rack or higher.

The high level of consolidation and the movement of workloads to public cloud has rendered the metric of average rack density less meaningful than it used to be. This year, we asked for respondents' highest server density deployed. Half said that their highest was between 10 kW and 29 kW. This range, showing an upward drift, is in line with our findings of six years ago: in 2012, our survey of a sample of operators in the Uptime Institute Network (n=59) found that 26 kW per rack was the single highest-density rack among them. These were operators of high-availability data centers.

However, in 2018, about one fifth of respondents said their highest-density rack was 30 kW or higher, which suggests that density extremes in data centers are escalating. Uptime Insitute's experience is that data center cooling systems are often poorly optimized; unless an IT environment was designed specifically for these high densities, operators are likely to experience issues.

What is the HIGHEST server density deployed in your site?

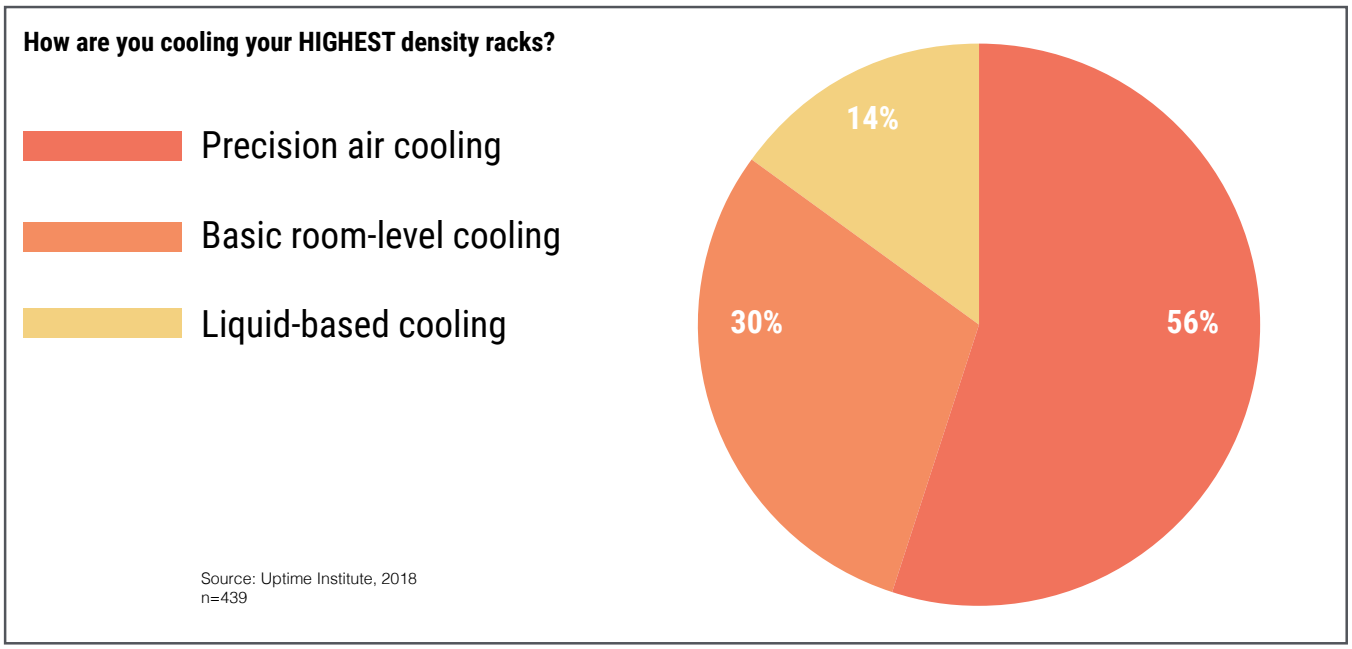




Source: Uptime Institute, 2018
n=431

Uptime Institute Global Data Center Survey

The rise in rack density is far from universal. Roughly one third of respondents said that their highest-density rack was less than 10 kW per rack. Clearly, density extremes are mixed, although our findings show that +10 kW racks are a reality for most data centers. Most respondents said they are cooling their highest-density racks using precision air cooling (59%), followed by basic room-level cooling and liquid cooling approaches.



In data centers such as Alibaba, Google, Facebook, and Microsoft that host artficial intelligence (AI) densities are clearly rising. Server hardware that is optimized for AI workloads, including GPUs, FPGAs, and ASICs, have much higher power/cooling requirements than standard x86 servers (up to 66% higher thermal design power [TDP or the maximum amount of a heat that a component is expected to generate under load] for GPUs vs CPUs, for example.)

While very high-density IT environments are likely to be confined to operators of AI applications and high-performance computing (including gaming and IoT applications with high IO), some colocation data center providers serving these types of customers will need to adapt. Prefabricated modular data center components, equipped with precision or liquid cooling, are increasingly being viewed as a retrofit tactic to enable mixed-density colocation (and other) environments.

Uptime Institute Global Data Center Survey

We believe micro-modular data centers will increasingly be designed for high densities. Operating as remotely managed, lights-outs facilities and often being sited in space-constrained areas, they are well suited to higher densities (and operating temperatures). Very high-density micro-modular data centers are available today, including designs for +40 kW racks and some that are liquid cooled.

DCIM is now mainstream

Data center infrastructure management (DCIM) software is often described as one of the most powerful and productive technologies that can be applied in the modern data center. DCIM promises accurate information about data center assets, resource use, and operational status. DCIM software collects, normalizes, and reports data about power utilization, availability, redundancy, and quality, as well as environmental conditions such as temperature, humidity, and airflow. Some DCIM also tracks connectivity, from the network switch to the patch panel and for individual ports, among other metrics.

When combined with IT management data, DCIM can deliver insight from the lowest levels of the facility infrastructure to the higher echelons of the IT stacks for end-to-end IT service management taking into account the availability and resiliency of physical data center resources (power, cooling, space, and networking).

The best-run data centers have implemented DCIM, but deployment can be difficult and the benefits are not always easily understood or measured. For these and other reasons, it has historically been a controversial and under-deployed technology.

However, it appears that DCIM has (finally) reached the point of being a mainstream data center technology. More than half of our survey respondents (54%) said they had purchased some sort of commercial DCIM software, with an addition 11% having deployed homegrown DCIM. In a testament to the maturity of DCIM technology, 75% of these users said their deployment was successful, and nearly half (47%) of these are currently supplementing their implementation with more DCIM tools.

These deployment figures raise certain questions about DCIM suppliers' costs and pricing and what the user community qualifies as DCIM. A 54% deployment level across the industry should point to higher aggregated DCIM revenues than our forecasted $1.6 billion by 2020 (global)–a fraction of the $6.9 billion total addressable market that we

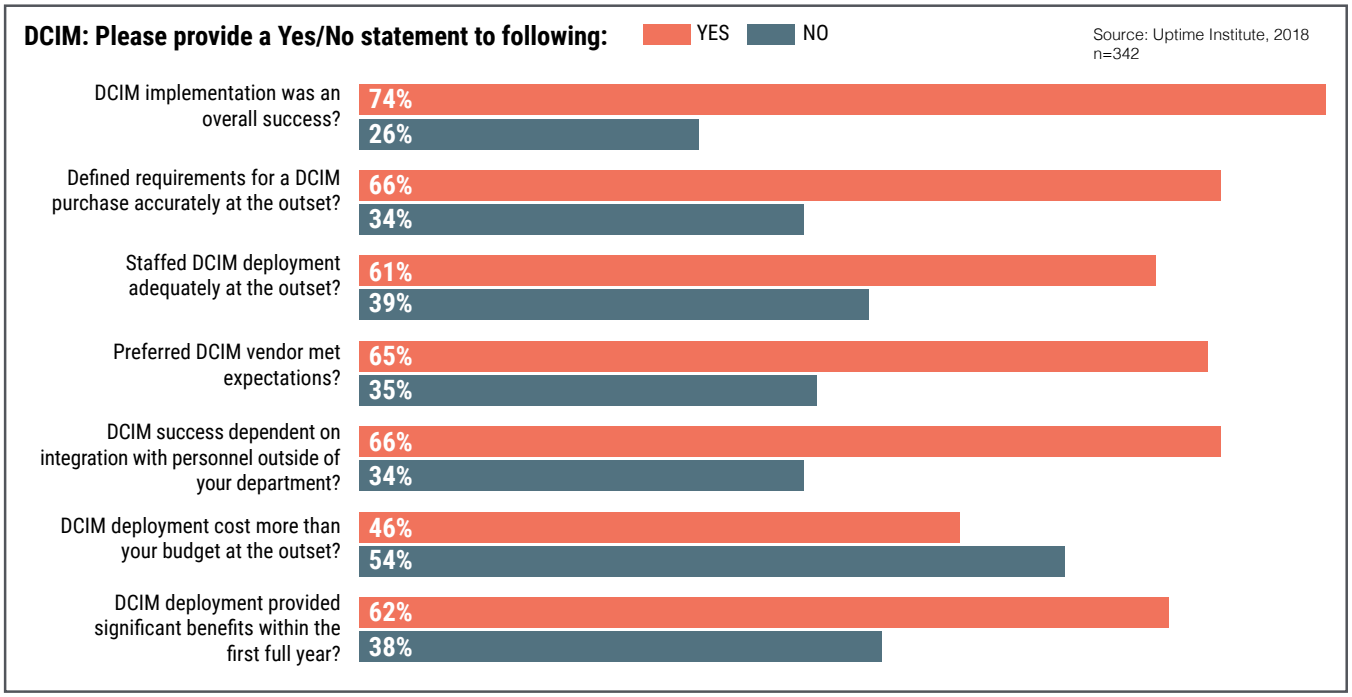
Uptime Institute Global Data Center Survey

estimate for the sector–suggesting either low prices/heavy discounting or a highly variable interpretation of what qualifies as DCIM software.

Of the 25% of survey respondent who had an unsuccessful deployment, one-third said they had no current plans to pursue the technology. The remaining two-thirds are working on secondary DCIM implementation options.

Among these respondents, the most common motivation for deploying DCIM was capacity planning (76%) and power monitoring (74%). Other reasons ranged from giving executives and/or customers (of multi-tenant data centers) visibility or reports (52% of respondents) to compliance (35%).

Digging into their DCIM deployments, most respondents adequately scoped their requirements for a successful implementation, including staffing, and said that the software delivered on their expectations, including benefits and cost.



Uptime Institute Global Data Center Survey

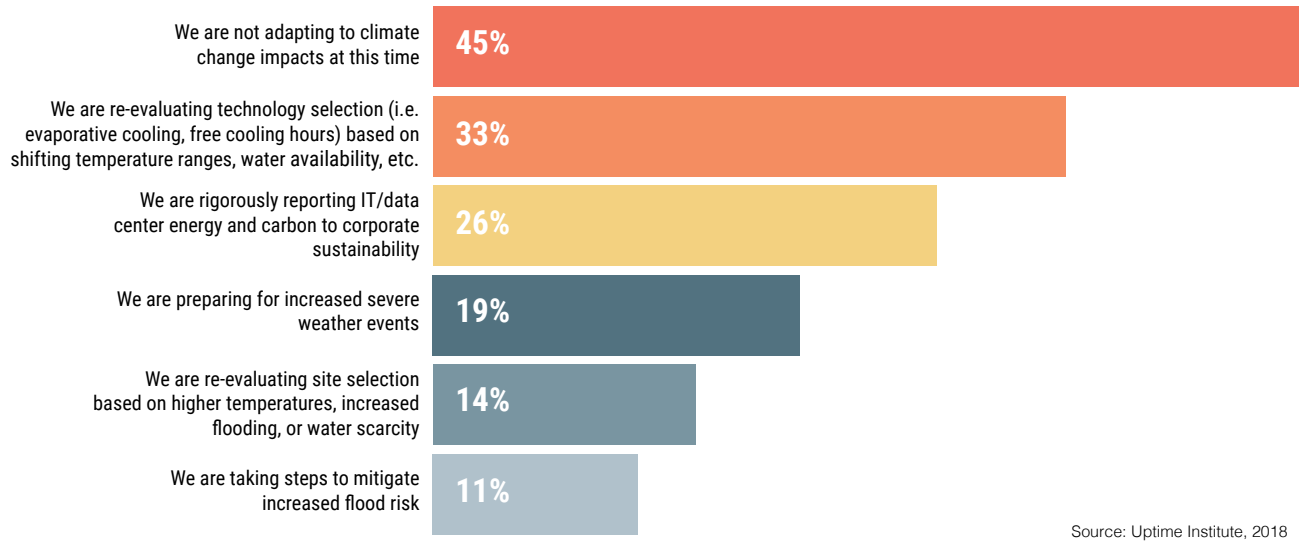
Many data centers are unprepared for climate change

In Uptime Institute's analysis, organizations around the world need to plan for increased extreme and sometimes disastrous weather. A warmer climate will drive a surge in the severity and frequency of destructive events, including (but not limited to) heat waves, flooding rains, droughts, and hurricanes. Data centers are vulnerable in many ways. Rising sea levels and heavier precipitation from storms are making current floodplain maps irrelevant, with 500-year floods becoming regular occurrences in certain geographies, such as New York City.

Yet about half (46%) of the respondents surveyed said that their organizations were not addressing potential climate-change disruption to their data centers. It appears that they had determined that they either won't be impacted or are ignoring the problem.

Of those that were preparing, the most common approach was re-evaluating their technology selection (such as evaporative cooling and free-cooling hours) based on shifting conditions, such as temperature ranges and water availability.

Is your organization revising policies and/or planning to revise data center policies to adapt to climate change? Select all that apply:



Source: Uptime Institute, 2018
n=709

Uptime Institute Global Data Center Survey

As region-wide disasters become more common, planning an organization's IT resiliency becomes more complicated. Uptime Institute advises that disaster and emergency planning can no longer be isolated to equipment testing and emergency drilling procedures but instead must be conducted in the context of an organization's broader emergency and business-continuity plans.

Climate change will also drive new considerations for siting and technology choices. Maps for estimating economizer or free-cooling hours may no longer be accurate, which could in turn change the availability and usability of some technology as well as the ROI on equipment and require a review of design decisions. Precipitation changes will result in drought and water shortages in some areas, making evaporative cooling options untenable, for example. Even small changes in temperature for a certain number of days a year can make some technologies more expensive or unviable. In addition, even if the data center itself is well prepared, staff and suppliers (of fuel) could be affected by transport and infrastructure issues.

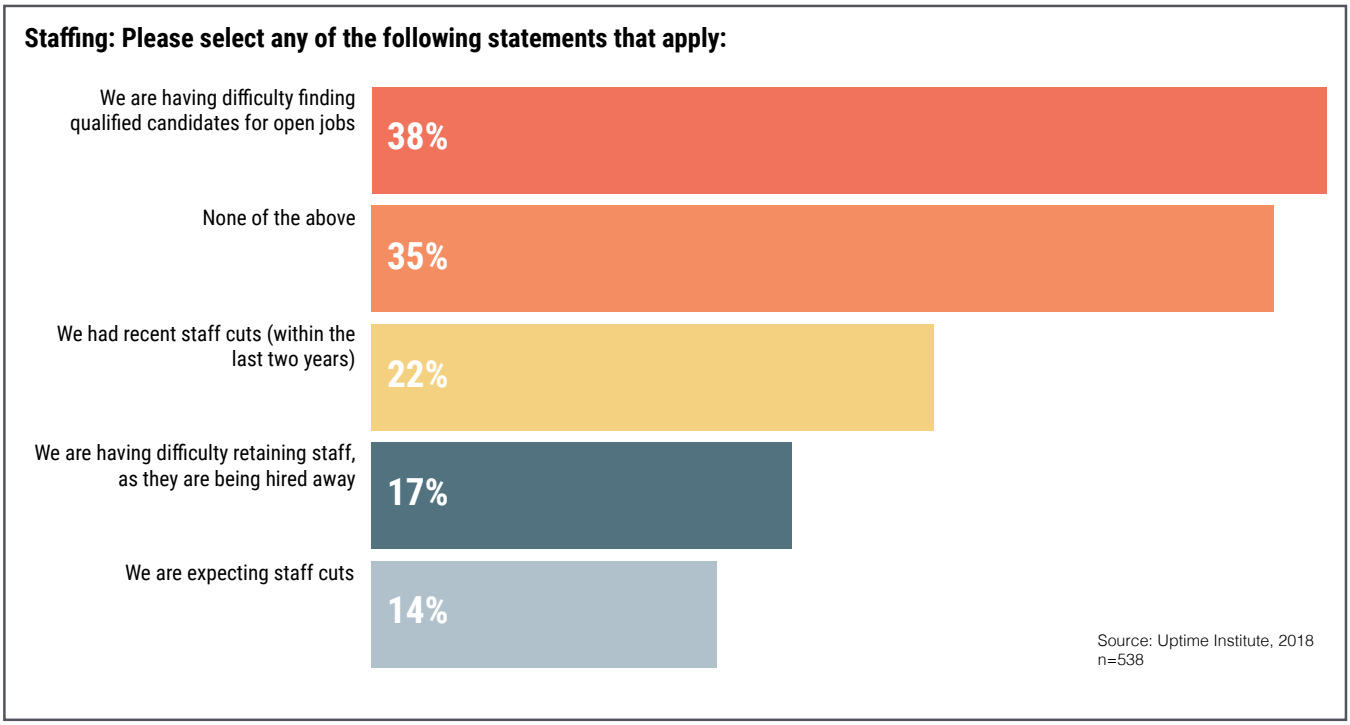
As climate impacts become more severe, organizations in every region should expect increased scrutiny of data center carbon emissions. Even if certain governments may be relaxing rules, the long-term trend is clear: new and expanded regulations are highly likely across the globe. It appears, however, that most of the data center industry today will be unprepared: just one-quarter of respondents said that they are rigorously reporting IT/data center energy and carbon to corporate sustainability.

Data center skill shortages will intensify

It is becoming harder to recruit and train staff for the skills required to operate and support increasingly hybrid IT environments. There is a growing need for new skills, including overseeing and managing SLAs for off-premises workloads. At the same time, software-defined and automation data center technologies are increasingly being deployed and may require less data center facilities staff–however, those staff typically require skills, such as software, that are not readily found in traditional data center staffing sources.

Just 35% of survey respondents reported that they did not have any of the hiring or staffing issues we identified. Roughly the same number had recent staff cuts or were expecting staffing cuts and/or were having difficulty finding qualified candidates for open jobs.

Uptime Institute Global Data Center Survey



When we parsed the data by respondents' industry vertical, those in the colocation or multi-tenant data center and software and/or services sectors reported a lower level of staffing cuts but had more difficulty finding qualified candidates for open jobs than the remaining total respondent pool.

Across all respondents (for all industry verticals), 17% said they are having difficulty retaining staff as they are being hired away. Staff retention difficulties and a shortage of qualified candidates is partly due to the intense hiring by hyperscale cloud and internet operators, as well as large colocation providers, which we have heard anecdotally is driving up salaries across the board.

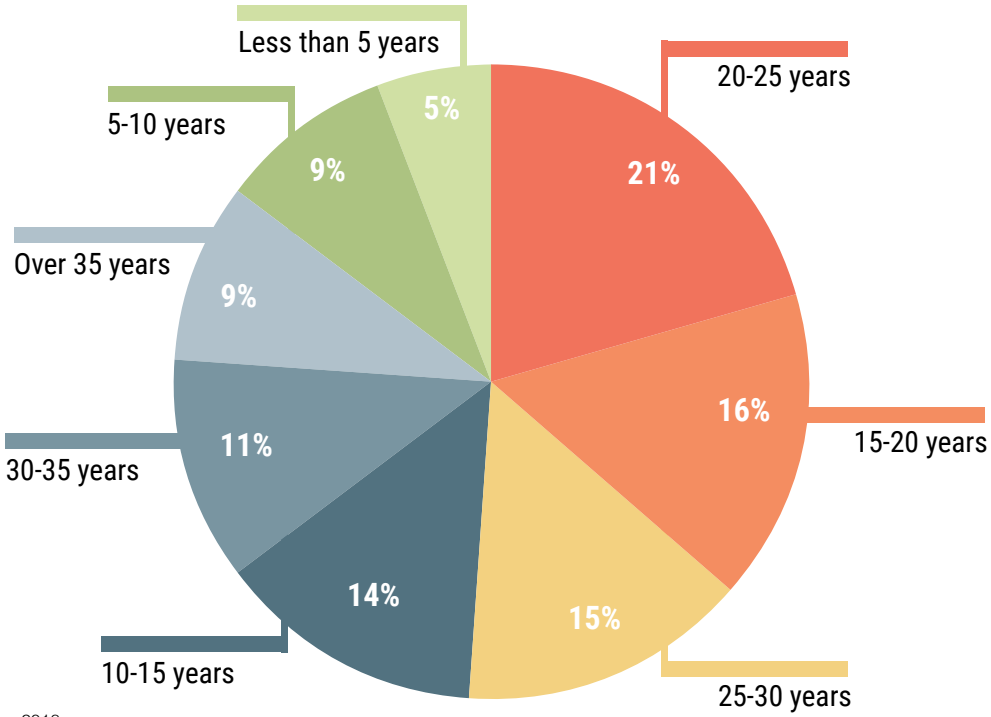
The leading area of expertise that is particularly critical and yet difficult to hire for is operations and management, according to more than 50% of respondents. Security, networking, electrical engineering, cloud (provisioning, load balancing, and resiliency), and mechanical engineering were also cited by at least one-third of respondents.

Uptime Institute Global Data Center Survey

Operators that are seeing success in their staffing strategies tend to focus on staff training, including cross-training existing personnel for both IT and facilities skills–essentially merging the two roles into a single generalist position. This has proven particularly attractive in organizations with a stated cloud-first strategy, which can leave internal staff feeling uncertain about their long-term future.

The broader issue is a staffing shortage facing the industry. Even the Internet giants, with their attractive salary packages, struggle to fill open positions. The problem is set to intensify; this is a fast-growing industry with an aging workforce–more than half (56%) of respondents in our survey had more than 20 years work experience. Only 5% were new to the industry, with fewer than 5 years experience. While our survey sample focused on managers (with some engineers), who tend to have longer industry tenures, this is still a smaller portion than we had anticipated.

How many years experience do you have in the technology industry?



Source: Uptime Institute, 2018
n=713

Uptime Institute Global Data Center Survey

Based on the Uptime Institute's joint efforts with large operators and other industry initiatives, there will be an increased focus on hiring outside the 'traditional' staffing routes with concerted efforts to attract 'non-traditional' candidates, including women and people with diverse backgrounds.

The sectors' workforce is overwhelming male: 56% of respondents said women make up less than 6% of their data center design, build, or operations staff. Yet 70% of the respondents in our survey said a lack of women in the data center profession is not a threat to their business or the industry at large.

56% say women make up less than 6% of their data center design, build or operations staff

73% of data center practitioners say it's not a problem

38% are struggling to find qualified candidates for open jobs

As study after study shows, a lack of diversity typically represents not just a lack of pipeline for hiring but also a threat of technical stagnation, negative publicity, and, ultimately, a loss of market share. There is growing consensus among data center industry leaders, and elsewhere, that the future success of the data center business will depend on building a diverse workforce.

Uptime Institute Global Data Center Survey

Author

Rhonda Ascierto, Uptime Institute

Rhonda Ascierto is VP of Research at the Uptime Institute. She has spent nearly two decades at the crossroads of IT and business as an analyst, speaker, adviser, and editor covering the technology and competitive forces that shape the global IT industry. Ms. Ascierto focuses on innovation and disruptive technologies in data centers and critical infrastructure, including those that enable the efficient use of all resources.

Before joining the Uptime Institute, Rhonda was research director of Datacenters at 451 Research. She joined 451 Research from Ovum where she was a senior analyst in energy and sustainability IT. Previously, she had a successful career as an IT reporter and editor, mostly in Silicon Valley, at a number of well-known industry publications.

This Executive Summary report is an overview of the results from Uptime Institute's 2018 survey of data center operators worldwide. The study includes additional insights not covered in this overview, including the cause and effect of data center outages, power and cooling redundancy configurations, the impact of legacy IT assets, power generation and energy storage adoption rates, and data center technology trends, including among multi-tenant data center customers.

Clients of the Uptime Institute and members of the Uptime Institute Network can obtain additional drill-down reports and data cuts by contacting Rhonda Ascierto, rascierto@uptimeinstitute.com. If you are not a member or client but are interested please contact Matt Stansberry, mstansberry@uptimeinstitute.com.

Uptime Institute®

Inside Track®

Uptime Institute Inside Track provides its subscribers with actionable resources and access to an elite peer community to ensure that IT functions as a competitive advantage rather than cost center. Uptime Institute is the IT industry's most trusted and adopted global standard for the proper design, build, and operation of data centers–the backbone of the digital economy. For over 20 years, Uptime Institute has been providing customers with the assurance that their digital infrastructure can perform at a level that is consistent with their business needs, across a wide array of operating conditions. Today, thousands of companies rely on Uptime Institute to enable their digital-centric business success. Uptime Institute, a division of The 451 Group, has office locations throughout the world, including those in North and South America, Europe, U.A.E., Russia, and Asia.

For more information on Inside Track please contact Matt Stansberry
mstansberry@uptimeinstitute.com

Uptime Institute is an unbiased advisory organization focused on improving the performance, efficiency, and reliability of business critical infrastructure through innovation, collaboration, and independent certifications. Uptime Institute serves all stakeholders responsible for IT service availability through industry leading standards, education, peer-to-peer networking, consulting, and award programs delivered to enterprise organizations and third-party operators, manufacturers, and providers. Uptime Institute is recognized globally for the creation and administration of the Tier Standards & Certifications for Data Center Design, Construction, and Operations, along with its Management & Operations (M&O) Stamp of Approval, FORCSS® methodology, and Efficient IT Stamp of Approval.

Uptime Institute – The Global Data Center Authority®, a division of The 451 Group, has office locations in the U.S., Mexico, Costa Rica, Brazil, U.K., Spain, U.A.E., Russia, Taiwan, Singapore, and Malaysia.
Visit www.uptimeinstitute.com for more information.

All general queries:
Uptime Institute
5470 Shilshole Avenue NW, Suite 500
Seattle, WA 98107
USA
+1 206 783 0510

info@uptimeinstitute.com