Uptime Institute | RESEARCH

ADVISORY
2018

PLANNING & STRATEGY

Uptime Institute Data Center Supply-Side Survey 2018

Designers, consultants, and suppliers prepare for a smart, automated, and complex future

ADVISORY I 2018

Uptime Institute Data Center Supply-Side Survey 2018

Uptime Institute research suggests more free cooling and smart power are being installed in data centers today—with AI and micro sites coming tomorrow.

Data center design engineers, consultants, and product vendors say that their customers are adopting increasingly efficient and sophisticated power and cooling approaches and are ever-more interested in technologies such as Lithium-ion batteries. Against this backdrop, they also say that many of their customers are struggling with capacity forecasting and hybrid IT approaches. In the years ahead, many of the respondents expect that AI will become widely adopted, as well as micro data centers for use at the "edge" of the network.

These are some of the key findings from the first annual Uptime Institute Data Center Supply-Side Survey, which complements the large annual Uptime Institute global survey of owners, operators, and IT practitioners, now in its eighth year.

The supply-side survey offers additional perspectives into the adoption of technologies and where investments are being made. In this report, we also discuss some of the practices and technologies and why they are being adopted (or not). For more forward-looking analysis, refer to our report Disruptive Technologies in the Data Center: 10 Technologies Driving a Wave of Change (https://insidetrack.uptimeinstitute.com/member/resource/show/24192).

Key findings

The survey asked 271 data center design engineers, consultants, and product vendors about their customers' buying habits, plans, and challenges. Among the findings:

- **Forecasting capacity and hybrid IT management are top challenges**. The use of mixed data center environments, such as privately-owned facilities, public clouds, colocation, and XaaS (software, platforms, and infrastructure as a service), appears to be complicating forecasts for future capacity requirements, among other issues.

Uptime Institute Data Center Supply-Side Survey

- **Large operators are warming to Lithium-ion batteries**. Most operators are either taking a wait-and-see approach or are in limited deployment projects, with a small portion of (mostly their largest) customers deploying Li-ion at scale.
- **Many, but not most, adopt smart power distribution.** For almost a third of our respondents, most of their customers have adopted so-called “smart” PDUs, a technology that continues to be developed.
- **Automatic transfer switches are still generally favored.** Static transfer switches tend to be added over time, and their deployment lags behind ATS units.
- **DCIM deployment is becoming more common.** The operators and suppliers in our two surveys agree that more data centers of all types are implementing DCIM software.
- **Free-cooling economization gains traction.** The use of indirect outside air, along with additional evaporative cooling, is becoming more accepted, especially in temperate climates and for mission-critical data centers.
- **N+1 power and cooling redundancy is becoming more common.**
- **AI will become ubiquitous in data centers.**
- **Edge computing will necessitate micro data centers.** Respondents predict that most data center operators will deploy small edge facilities within five years.
- **Staffing challenges: more robots, more women?** Unlike end users, the industry’s supply-side view is generally that staffing issues will threaten future growth. Some believe robots will play a large role; many believe a more diversified talent pool will too.

Uptime Institute Data Center Supply-Side Survey

Survey demographics

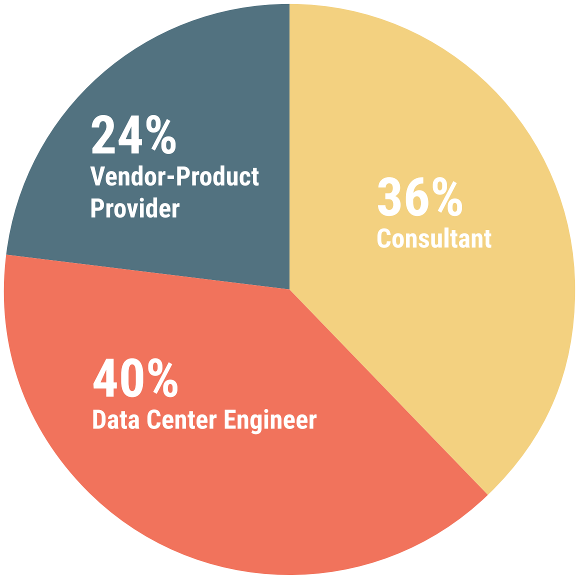
The survey, conducted via email between February and May 2018, includes responses from 271 data center service and equipment suppliers–people responsible for designing and building data centers, as well as suppliers of data center equipment. The majority are data center design engineers (40% of respondents), and data center consultants (36%). Fewer than one quarter (24%) are data center equipment product vendors.

Forecasting capacity is a top challenge

According to nearly half of our survey respondents, the single largest challenge facing their customers is forecasting data center capacity requirements. Being able to accurately predict capacity growth has long been problematic for operators, particularly when looking ahead more than one or two years—but the emergence of cloud and hybrid architectures has made this challenge more difficult.

Job Function

Source: Uptime Institute Research, 2018, n = 271



Note: Respondents were not required to answer all questions, which means the number of respondents for individual questions ('n') varies. We did not ask respondents for additional information, such as their geographic location.

The ability to forecast capacity more accurately (and easily) is the top reason operators deploy DCIM software. However, even advanced DCIM-based analysis may not adequately take into consideration factors such as increased computing throughput from server refreshes or from step changes such as a retrofit project. We are beginning to see new artificial intelligence (AI)-driven approaches from DCIM suppliers and others to model future capacity requirements for these types of scenarios.

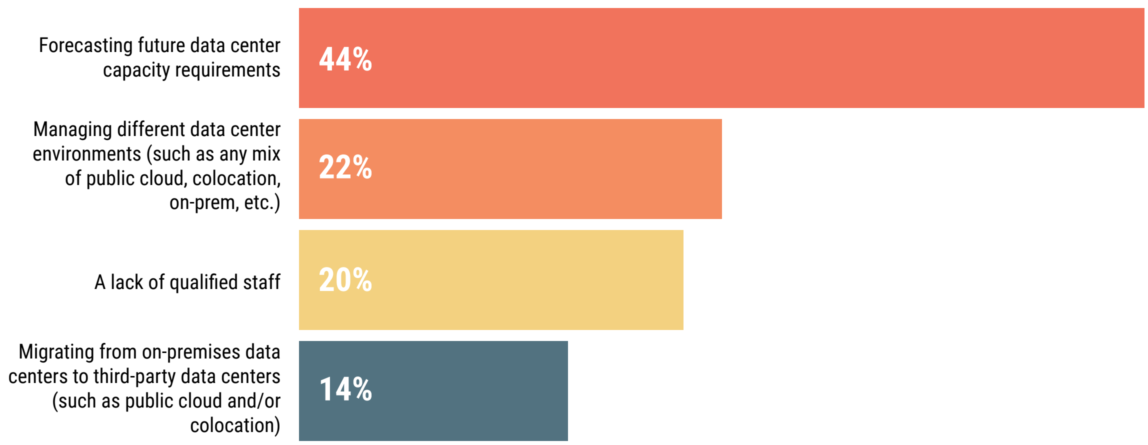
Capacity forecasting has also been complicated by the ever-expanding capacity options available to organizations (SaaS, PaaS, IaaS, serverless, managed hosting, colocation, and so on). CIOs may make major commitments to move in one direction or another (and sometimes back again), upsetting planning.

Uptime Institute Data Center Supply-Side Survey

Managing different data center environments ("hybrid IT capacity"), such as any mix of privately owned "on-premises," public cloud, colocation, and XaaS capacity, was cited as the next largest challenge for data center customers (22% of respondents), followed closely by a lack of qualified staff (20%), and migrating from on-premises data centers to third-party data centers (14%). Unlike the other answer options, migrating from an on-premises facility to a third-party one is a one-time event, which might explain why it's lower on the list.

Capacity forecasting remains the number 1 challenge facing data center operators, according to designers, vendors, and consultants. Operating hybrid environments and a lack of qualified staff are also concerns.

Source: Uptime Institute Research, 2018, n = 264



Which of the following do you believe is the single largest challenge for the majority of your customers? (Choose one)

Hybrid IT capacity has become the norm, with more than half of enterprises utilizing two or more different data center capacity options (based on our previous surveys and research). Outsourcing is being driven by the need for faster time to deployment for new applications and by the general move from capital to operational expenditure, among other factors. Demand for colocation capacity, in particular, is coming from organizations seeking highly connected capacity, including interconnects to public cloud services and partners.

Organizations tend to favor multi-supplier or multi-cloud strategies to best balance performance and cost and to meet their specific workload needs. However, there is a significant hidden cost with hybrid strategies: increased management and integration complexity.

Many organizations that are taking a hybrid IT capacity approach also often struggle to find IT staff with adequate skill sets; hybrid IT management skills are different to traditional on-premises-only IT and facilities roles. For example, managing different service provider contracts and service-level agreements could, in some cases, require dedicated personnel, particularly in regulated industries. Hybrid environments may also introduce more software-defined technologies for workload orchestration and automation, for example, that require additional scripting and software management skills.

Hybrid IT: Uptime Institute Recommendations

Business-led goals are critical for a successful hybrid IT capacity strategy—lowest-cost or “cloud-first” is not necessarily the best starting point. To meet business objectives, evaluation methods must be developed to determine the best execution venue for each IT workload. Evaluation methods should assess key factors in addition to cost, such as application performance, latency requirements, and the overall impact on business risk, including security, availability, and redundancy.

Increasingly DCIM software is being deployed by all types of organizations to enable visibility across data centers to allow for real-time monitoring of critical systems and for data center site- or portfolio-level capacity planning. DCIM data can also be used to create standardized metrics and key performance indicators (KPIs) to help support best-execution venue decisions. New metrics and KPIs can be created by combining (in an automated way) data from DCIM and IT systems management software, as well as from financial tools, to help with venue decisions.

We recommend creating standardized internal process blueprints and KPIs, where possible, to assess each data center location and service.

It is important to plan for an investment in resources and staffing to manage hybrid IT environments, including implementing initiatives for retaining and retraining staff. In particular, training on software management tools, the creation of custom dashboards, and new reporting functions may be required to efficiently and cost-effectively execute on hybrid IT approaches.

Technology adoption

In our supply-side survey, we asked respondents about the adoption of various data center technologies by their customers and also their future views of the data center industry.

Lithium-ion battery adoption has begun

Until recently, the market for lithium-ion (Li-ion) batteries in data centers has been largely dormant, but the density, cost, and safety profiles of Li-ion battery power systems have all improved. The adoption of large Li-ion batteries in electric vehicles has driven down manufacturing and supply chain costs to the point that their use in UPS systems—either centralized or distributed—can deliver significant savings in large data centers. As costs continue to fall, we expect Li-ion will gradually replace traditional batteries in UPS systems for data centers of all sizes.

The main advantages of Li-ion power batteries are clear, when compared with traditional VLRA batteries:

- Significantly higher energy density, which means significantly smaller (by a factor of two to three) and lighter batteries. These require less cooling and enable savings on racks, in addition to easing system installation.

- Roughly twice the life span and much lower maintenance requirements.

- Faster recharging. Li-ion batteries do not experience nearly the same reduction in effective energy as the load increases, which means fewer batteries may be required in some applications.

The table below details the specifics of some of these favorable (typical) performance attributes for Li-ion versus incumbent VRLA batteries in data center UPS applications.

High-level comparison of battery performance characteristics

Performance attribute	Lithium-ion	Lead-acid (VRLA)
Energy density (kWh/kg)	High, 70-260	Low, 15-50
Life span (years)	10-15	4-6
Charge/discharge cycles (no)*	>1,000	200-400
Recharge time (hr)	0.5-1.0	6-12

*The actual number is highly dependent on the specific battery design and the depth of the discharges. Source: Schneider Electric, White Paper 231, 2016

Higher ambient temperatures reduce the lifespans of both Li-ion and VRLA batteries. Each increase of 10 degrees Celsius (18 degrees Fahrenheit) will halve the lifespan of a Li-ion battery. Li-ion batteries are often designed for higher nominal temperatures and so the operational lifespan is typically better. For example, the lifespan of some Li-ion power batteries used in data centers today will typically degrade by 70-80% when operating at a supply ambient temperature of 40°C or (104°F)—still a relatively good profile.

Li-ion batteries have higher purchase and shipping costs and greater safety concerns. Li-ion power batteries can cost roughly 1.5-3 times more to purchase than VRLAs with comparable power levels and runtime—a significant delta, although far less than just a few years ago when Li-ion batteries cost about 10 times more.

Transportation regulations can increase the shipping costs. High-density Li-ion batteries, such as those used in data centers, are a Class 9 hazardous material and must be shipped by freight air and ground cargo only, adding cost and lead times in some situations.

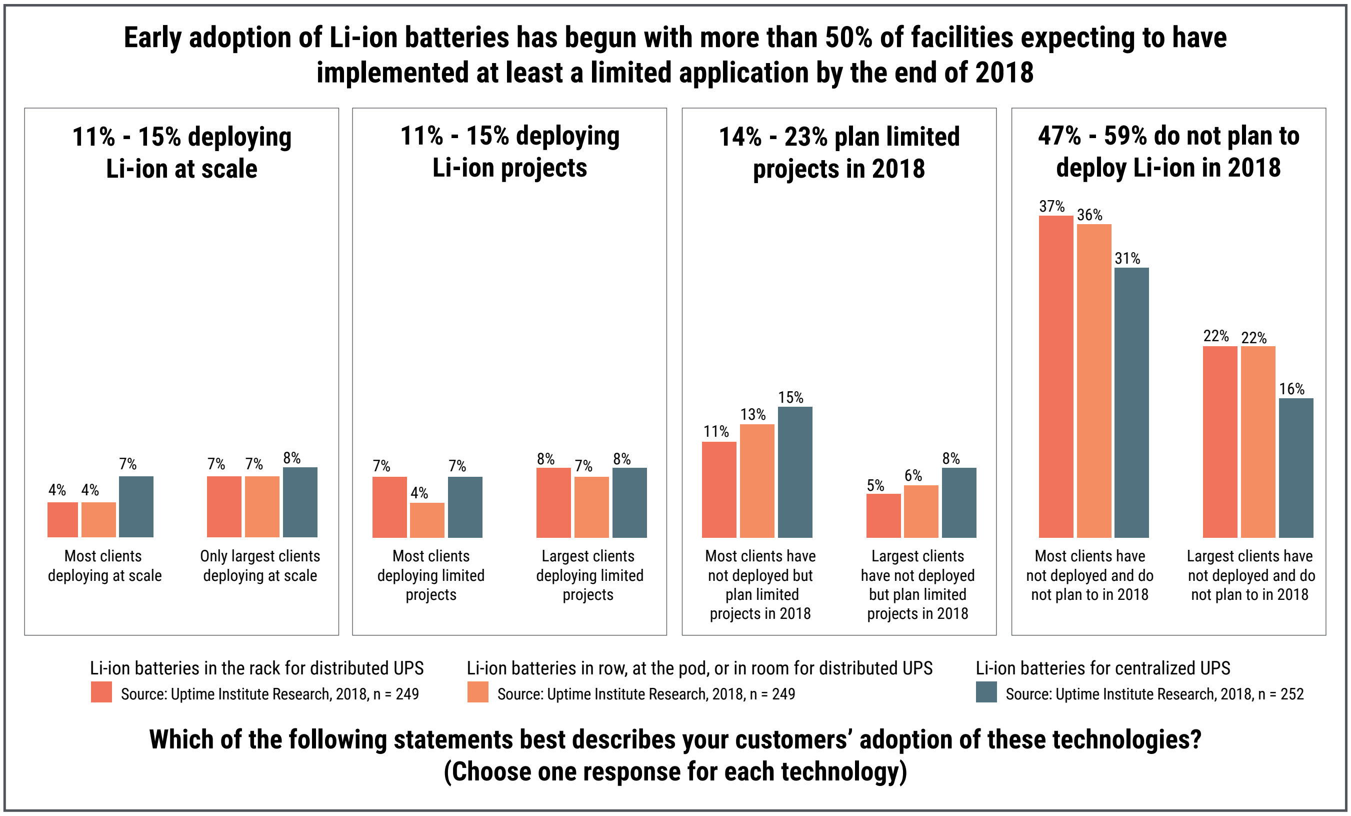
Li-ion batteries include several layers of safety measures built into the systems, including over-current and over-voltage devices in the battery cells themselves to manage combustibility. The first line of defense is a battery management system, which typically includes monitoring, protection, recharge control, and battery cell equalization, as well as management features such as multiple warning levels. (The software can also be integrated into other systems to enable new efficiency opportunities.)

Finally, disposing of Li-ion batteries is a factor. Most of the Li-ion battery can now be fully recycled but at a much higher cost than VRLAs. This situation will likely change over time as high-density Li-ion batteries are adopted in more electric vehicles.

The bottom line is that, on balance and over a 10-year span, Li-ion can be less expensive than VRLA as a large data center UPS.

All major data center equipment suppliers are pushing Li-ion technology, although the industry has yet to standardize on a set of Li-ion configuration specifications. That will just be a matter of time.

Meanwhile, according to our survey respondents, most of their customers, including their largest ones, are either taking a wait-and-see approach or are in limited deployment projects (most likely proof of concepts.) Only a small portion of customers are deploying Li-ion at scale, and mostly these are their largest customers.



Uptime Institute Data Center Supply-Side Survey

As operators become confident with the technology and overcome preconceived safety concerns, it's likely that more will begin thinking about using Li-ion, including in the white space. We expect this may in turn lead to more distributed Open Compute Project-type electrical designs in more data centers, promising greater cost savings and differentiation, particularly for data center service providers.

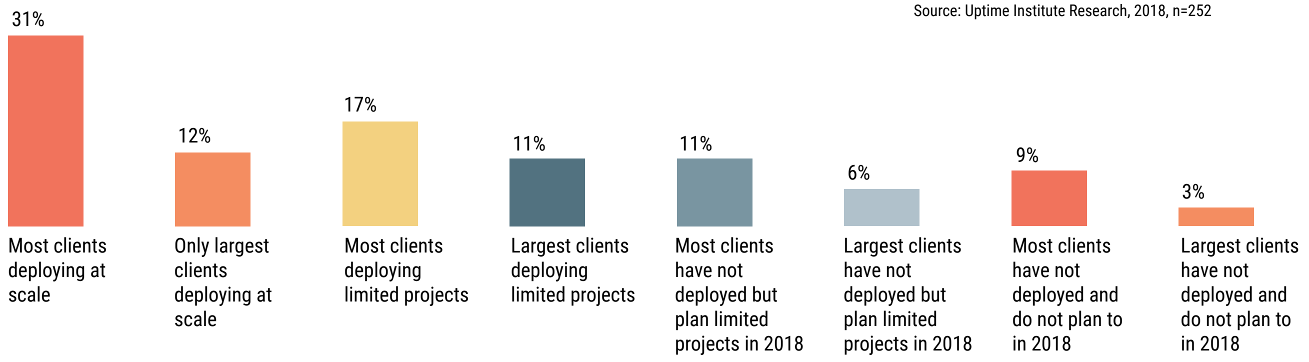
Energy-storage systems using banks of Li-ion batteries may even eventually replace engine generators—but the economics will only make sense if deployed with a demand-response or micro-grid-type system, at least in the short term.

Many, but not most, adopt smart power distribution
According to nearly half of our survey respondents, most of their data center customers are using smart PDUs, either at scale (31%) or in limited deployments (16%). Smart PDU is a catch-all term referring to power distribution units that have an embedded computing chip (controller) and a user interface that shows alarms and alerts/status, power usage, and so on. Some smart PDUs just meter inlet and outlet power, others monitor it (with additional features), and some enable switching. Smart PDUs can be centrally managed, power cycled, and report data to DCIM and other software.

Relatively few large clients have deployed smart PDUs, which is not surprising. Many internet giants, for example, favor standard/commodity PDUs rather than the more expensive intelligent devices or else have adopted alternative Open Compute Project (OCP) or similar rack designs.

Many data centers will have installed at least a limited number of smart PDU deployments by the end of 2018

Source: Uptime Institute Research, 2018, n=252



Which of the following statements best describes your customers' adoption of smart PDUs?

Uptime Institute Data Center Supply-Side Survey

Fierce competition among PDU suppliers is driving innovation, which appears to be helping drive adoption among enterprise customers. More than 90% of our respondents say that most of their clients that have not already deployed smart PDUs will do so in 2018. Over time, we expect some of the more advanced smart PDU features to become cheaper.

How smart are smart PDUs?

Smart PDUs can have various capabilities and configurations and may include some of the following features:

- The ability to connect PDUs from an A and a B feed, so that their controllers share power. If there is an outage in one feed the affected controller can still communicate (via alerts) so the operator can determine, for example, whether the fault lies with the PDU or the rack-power supply.
- Gigabit Ethernet connectivity to allow for future network upgrades without the need to replace PDUs.
- Secondary networking ports enabling, for example, a colo customer to allow the provider to access its PDU for power monitoring on a separate network (without giving the colo access to its main local area network).
- Either:
  - daisy-chaining multiple devices (via USB ports of Ethernet) with one IP address to reduce IP connectivity costs ($300-$500 per IP connection that is not required) or
  - multiple devices share one IP address with a designated master, so if the master loses power, the network card remains operational and maintains network connectivity and management, and with the first link unit providing power redundancy.
- Data exchange with building management system (BMS) platforms (including controls).
- Alternating phase outlets that distribute phases across each outlet of the device to balance the load and reduce cabling.

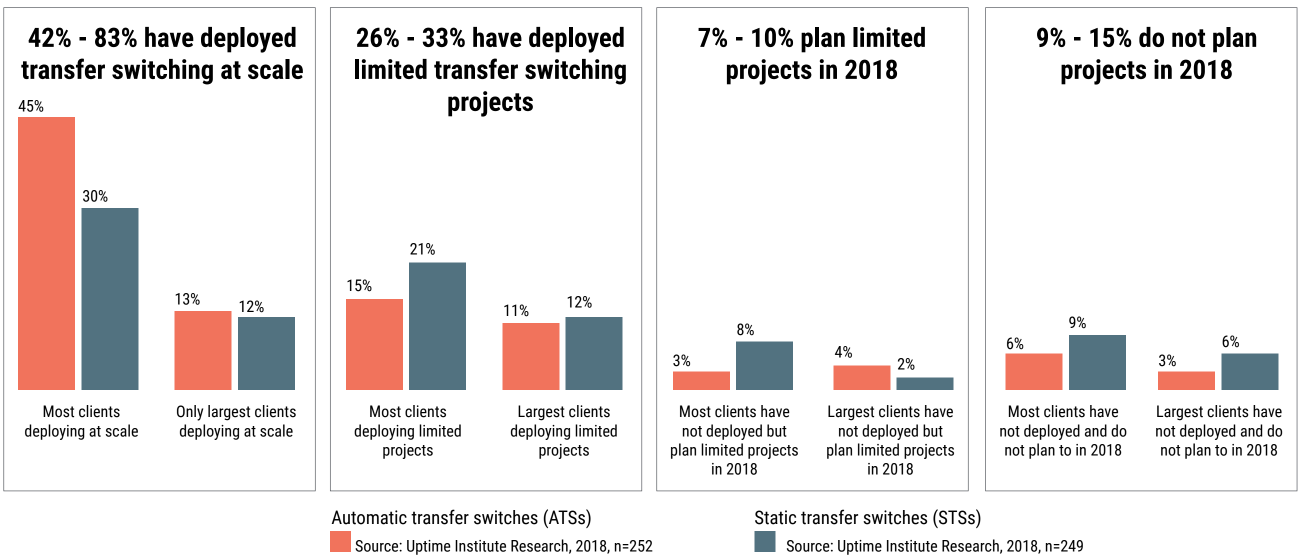
Uptime Institute Data Center Supply-Side Survey

Power switching: automatic versus static

We also asked about the adoption of power switching technology: automatic transfer switches (ATS) and static transfer switches (STS). Each has trade-offs in terms of cost, product lifecycle, and response times. Uptime Institute often sees STS deployments in N+1 data center power designs and in colocation data centers (colos tend to deploy STS on PDUs as part of their SLA promises and to enable safer live-load maintenance work).

In our survey, more than half of the respondents said that their customers are deploying ATSs, either at scale (45%) or in limited projects (15%).

ATS and STS projects are mainstream. Vendors, consultants, and designers see customers deploying transfer switching at scale.



Which of the following statements best describes your customers' adoption of these technologies? (Choose one response for each technology)

The survey data correlates to Uptime Institute's field experiences, whereby ATS are typically installed as part of a data center's original design for engine generator systems and facility systems that support cooling or other equipment. STSs are often used in direct support of IT equipment and, therefore, tend to be added over time—so their deployment can lag behind ATSs. STSs are more likely to be deployed downstream of the UPS.

STSs are slightly more prone to failures (due to their more complex design) and are significantly more expensive than ATSs. (Generally speaking, STS costs can be lowered if purchased as part of a smart switching PDU.) But the cost premium can be outweighed by the faster load transfer capabilities of STSs, which is key when switching IT devices to an alternative power source (in the event of a power loss.)

Beyond this requirement, ATSs are typically the technology of choice because of their high reliability, lower cost, and their ability to switch large loads (relative to STSs.)

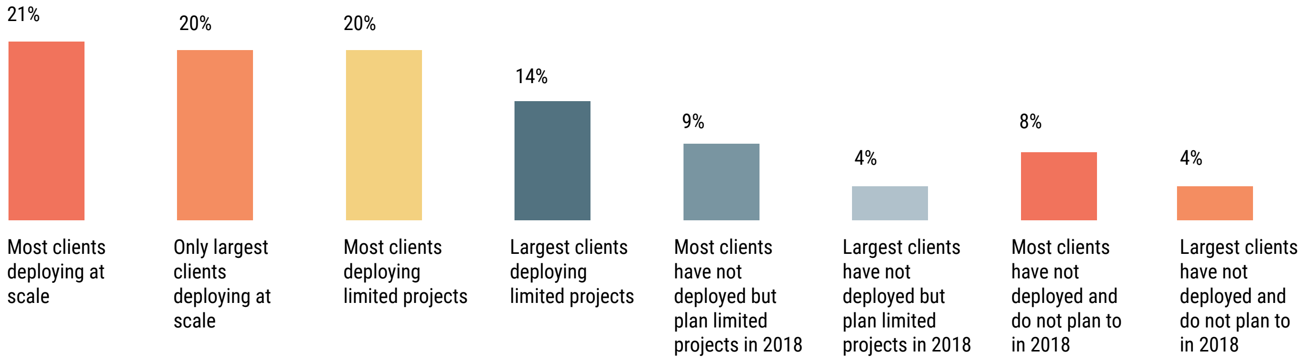
DCIM software is becoming more common

Data center infrastructure management (DCIM) software appears to have evolved from being a contentious and under-deployed technology just a few years ago to a mainstream one today.

In our survey, when we asked suppliers whether most of their customers had deployed the software at scale or in limited deployments, the answer was, for the most part: both.

DCIM adoption remains slow, but steady, with some holdouts remaining.

Source: Uptime Institute Research, 2018, n=248



Which of the following statements best describes your customers' adoption of DCIM software?

The results mirror findings from our recent Eighth Annual Uptime Institute Data Center Survey of 867 data center operators and IT practitioners (globally in 2018), where almost two thirds of the respondents said they had deployed DCIM (54% had purchased commercial DCIM software, with an additional 11% having installed homegrown DCIM.)

In this survey, we asked our supply-side respondents to distinguish between at-scale (that is, large scale) and limited DCIM deployments. There was almost an equal number of respondents who said most of their customers were deploying at scale (21%) and those who said most were deploying limited projects (20%). One fifth of respondents said only their largest clients were deploying at scale, with 14% of their largest clients deploying limited projects.

The high number of limited projects reflects what Uptime Institute sees in the field (and also what we recommend to many operators)—proof-of-concepts deployments ahead of wider deployments are common, as are phased DCIM implementations over time. Operators with multiple data centers typically deploy one or two DCIM components, such as monitoring or asset management, at one data center at a time and incrementally add new features.

Uptime Institute is seeing DCIM being deployed in all types of data centers, ranging from hyperscale to micro facilities. The most active vertical has been colocation data centers, where many providers are making DCIM features available to their customers. We anticipate greater adoption of DCIM by colocation providers, in particular, driven by:

- Operators' need to better manage and forecast capacity and to drive down risks.

- Demand from end-user customers for greater transparency of facility operations, the delivery of services (such as remote hands and other work), and their consumption of data center resources (such as conditioned power)

Most colo providers today are deploying just DCIM monitoring features to provide their customers with online portals for visibility about an individual customer's power usage, environmentals, such as temperature and humidity (typically at the cage level, but some also at the rack), and connectivity. A smaller number of colos have also deployed DCIM for their own operations, and some are making limited data available to their customers to give them better transparency into the provider's supporting infrastructure and the delivery of services (such as remote hands and other work).

Some colos are offering their customers more advanced features, or plan to, including what-if scenario planning and customer capacity forecasting. Artificial intelligence is on the roadmap for some.

Most colos that are offering basic DCIM monitoring data are not charging their customers for it (or they charge a nominal amount); however, we believe they will charge a premium for more advanced forecasting and analysis features in the future.

Free cooling economization gains traction

As the industry's broader energy-efficiency expectations have increased over the recent past, more data centers are being designed with some level of non-mechanical cooling, free cooling modes for conventional chillers, or extremely efficient mechanical cooling. This move was largely triggered by ASHRAE's decision to widen its recommended operating temperature and humidity envelope for data centers, which is taken as an industry-standard best practice by many operators.

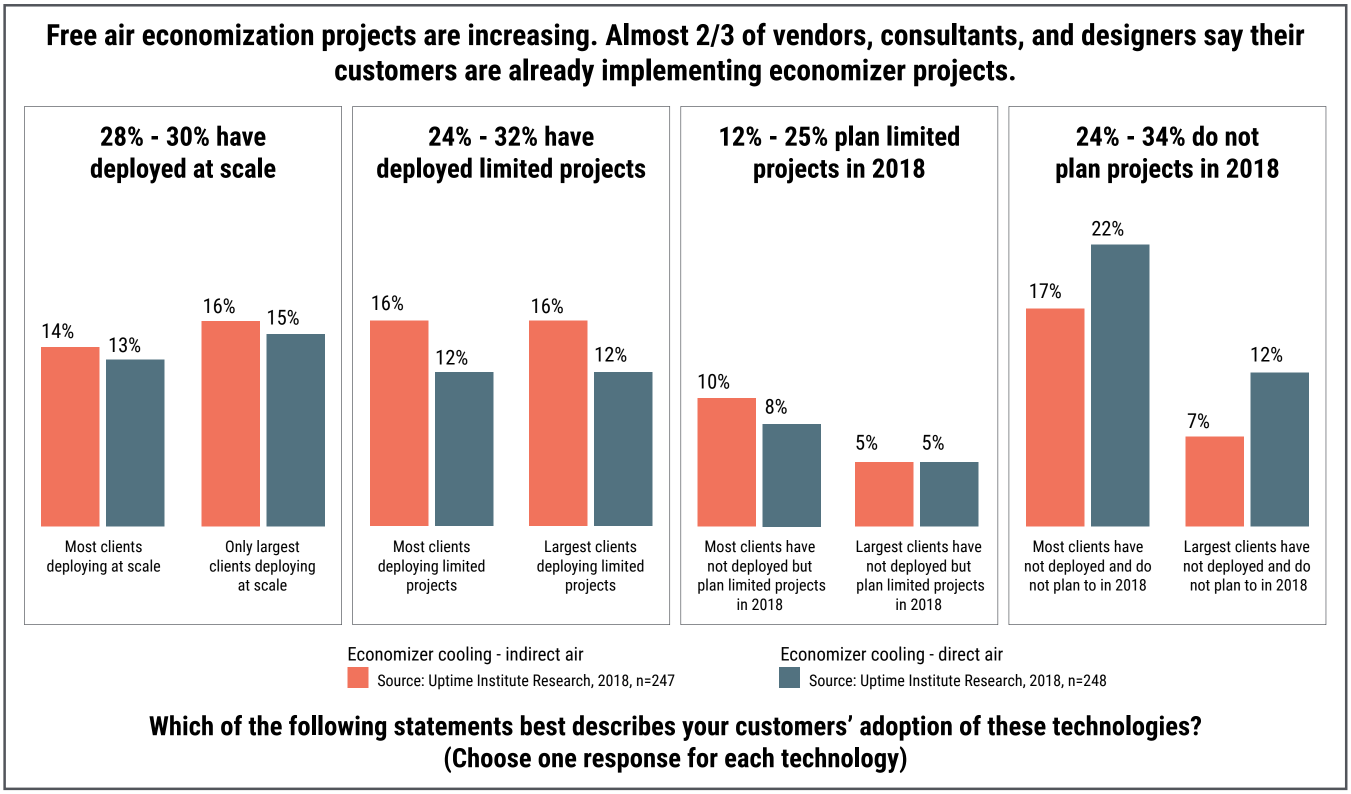
There is no *de facto* standard cooling system or technology for the data center of today and tomorrow, and there almost certainly won't be. However, more data centers are using outside-air cooling (known as free cooling or free-air cooling) and evaporative techniques, in combination with traditional mechanical cooling, which has dramatically improved the energy efficiency of data centers, enabling very good PUE ratios. New data centers are also being designed to optimize airflow and cooling, including optimization strategies that use much less energy than in the past. Retrofits involving cooling are not common, but some owners are also overhauling or replacing their CRAC and CRAH systems.

There is, of course, a difference between mechanical cooling equipment with economizer modes and the use of outside air or combination of water and air to supplement mechanical cooling (known as air economizer cooling). It's also worth noting that the term free cooling is problematic because, while less electricity is required compared with conventional mechanical cooling, free cooling is not cost free; power is still required for fans to move the air and for other infrastructure.

We asked our survey respondents about their customers' adoption of air economizer cooling (the use of outside air or combination of water and air to supplement mechanical cooling) using the following approaches:

- **Indirect air.** Outside air passes through a heat exchanger that separates air inside the data center from the cooled outside air to prevent particulates from entering the white space.

- **Direct air.** Outside air passes through an evaporative cooler and then directed via filters to the data center cold aisle. When the temperature outside is too cold, the system mixes the outside air with exhaust air to achieve the correct inlet temperature for the facility.



As the data suggests, the use of indirect outside air, along with additional evaporative cooling, is becoming an increasingly popular technology, especially in temperate climates and in mission-critical data centers where there is some caution around using outside air directly. The use of direct-air economizer cooling continues to lag due to concerns about humidity and contaminants.

The benefits of these technologies depend on how much cooling can be achieved without mechanical equipment, on the risk appetite of the operator, and on the service-level agreements (SLA) that data centers offer (many operators agree to keep in the middle or low end of the ASHRAE recommended range so equipment reliability is not compromised).

But even relatively risk-averse data centers can save significantly on mechanical cooling equipment (those less risk-averse can eliminate it altogether), in addition to reducing their overall energy consumption. For example, according to some supplier estimates, indirect-air cooling systems cost 15% less than an equivalent chilled-water system, but the operational cost is 90% less—adding up to a 75% reduction over the total life. Depending on the configuration, lower cooling energy requirements can also mean that diesel generator size can be reduced by 60% and transformer size by 70%—figures that are impressive.

However, the data also makes it clear that free cooling remains an under-deployed technology. Conventional mechanical cooling continues to dominate the industry and, given the costs associated with retrofitting existing cooling systems, will likely continue to dominate in the near term, even taking into account the recent growth in direct liquid cooling approaches (for high performance computing, artificial intelligence, and other applications).

N+1 power and cooling redundancy is becoming more common

One of the most talked about trends in data centers (and one that is associated with the use of smart PDUs and static transfer switches,) is the apparent increase in the use of N+1 data center designs, as opposed to 2N.

According to Uptime Institute's operator survey, N+1 designs are becoming more popular, with 51% of operator respondents having N+1 cooling equipment and 41% having N+1 power equipment configurations, representing the highest portion of respondents in both categories. The supplier survey shows that suppliers think about half of all primary data centers now have N+1 on power distribution, with most of the rest opting for higher levels of redundancy. Unsurprisingly, for cooling, the figure for N+1 rises to 70%.

A 2N design includes 100% replication of components/capacity, and so it would require a major failure of multiple components to cause downtime. In an N+1 design, a certain risk is taken on the available capacity, because extra capacity/component (or sometime an extra two components) will cover for a single failure, or perhaps a limited amount more, but there may not be enough spare capacity to cover for a complete failure of the primary equipment. The use of N+1 is usually justifiable because the risk of multiple failures is so low.

Uptime Institute Data Center Supply-Side Survey

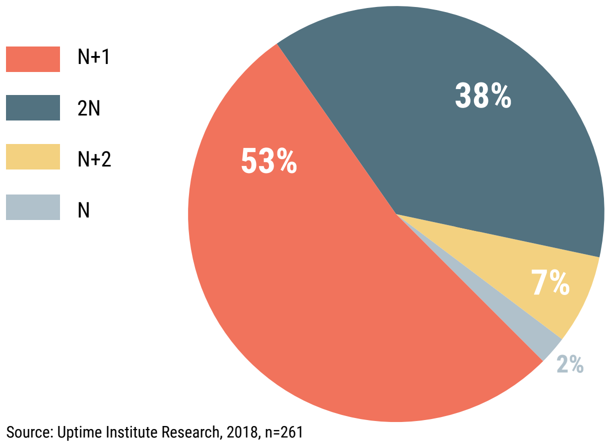
N+1 designs are becoming more popular for four reasons:

- N+1 designs are cheaper and more energy efficient because the physical infrastructure is lighter and uses less energy.
- New products and designs from major suppliers, sometimes involving more use of static transfer switches and possibly some deployment of Li-Ion batteries, are being used to manage capacity and to share reserves of power as a single, virtual pool. With ever-more advanced management software embedded into the equipment or used for monitoring separately, these products are becoming ever more reliable.
- Many colocation companies are under pressure from their biggest clients—cloud service providers—to build an N+1 infrastructure for cost reasons.
- N+1 designs configured as part of a distributed resiliency plan may help organizations meet operational requirements as well as or better than traditional 2N infrastructure.

It is important to understand that both 2N designs and N+1 designs can be receive Tier III Certification from Uptime Institute. The key is that all the power equipment must be concurrently maintainable.

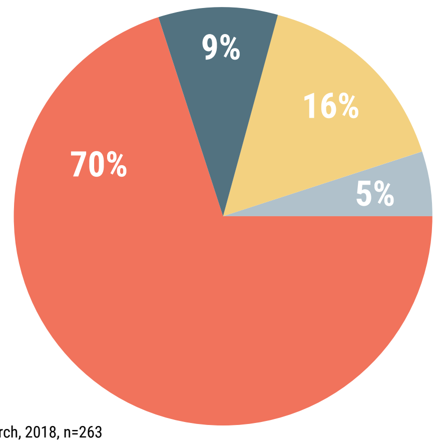
N+1 power and cooling redundancy configurations are most common

For your customers' primary data centers, which of following best describes the most common level of redundant power equipment?



Source: Uptime Institute Research, 2018, n=261

For your customers' primary data centers, which of following best describes the most common level of redundant cooling equipment?



Source: Uptime Institute Research, 2018, n=263

The growth of N+1 is an important trend with implications for vendors and operators alike. One finding from our operator survey is that, of those with a 2N architecture, 22% had experienced an outage in the past year, rising to 35% in the past three years (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. The message is that the cheaper solution (N+1) is currently less resilient—but that distributed resiliency plans based on N+1 designs could prove equally as resilient as a 2N facility at lower cost.

Looking ahead

Design engineers, consultants, and suppliers—the respondents in our survey—can have a solid perspective of future trends, driven by the interests of their customers. They can also influence their customers by way of the types of technologies they invest in for research and development. While their customers' choices and investments will ultimately determine the future of the data center industry, we asked our supply-side respondents for their view on several hot trends.

AI will become ubiquitous in data centers

Artificial intelligence is not new in data centers, but its application is broadening. For several years, cooling control systems have been available that are based on machine-learning AI software that determines and continually learns relationships between variables such as rack temperature, cooling unit settings, cooling capacity, cooling redundancy, power use, and risk of failure. Oftentimes the software is used to control cooling equipment, including variable frequency drives (VFDs), by turning units on and off, adjusting VFDs up or down, and adjusting temperature setpoints. More recently, the software can also predict what would happen if an operator took a certain action, such as shutting off a cooling unit or increasing the set-point temperature.

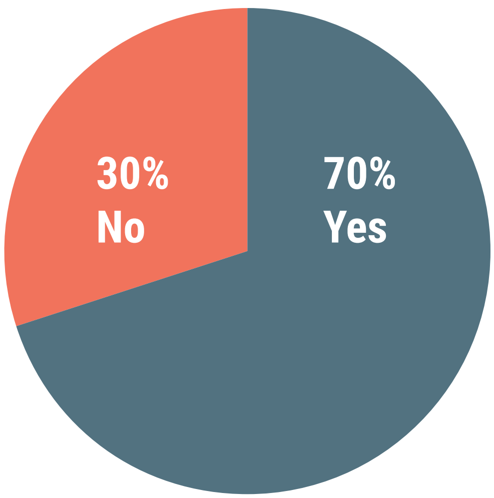
Similar AI approaches are now being applied to various data center facilities equipment. One notably development has been AI-driven predictive maintenance for UPSs, for example. Suppliers ranging from entrenched equipment vendors to specialist AI firms are developing AI software and services to support and in some cases create new insights. A common goal is to increase data center efficiencies, including for operations, equipment configurations, and even new facility designs, among other use cases. The majority (70%) of respondents in our survey believe AI will be widely used in data centers in the future.

Uptime Institute Data Center Supply-Side Survey

The advent of AI-driven data center management as a service (DMaaS), in particular, has made the technology readily accessible. Cloud-delivered DMaaS, a new market development, aggregates and analyzes large sets of anonymized DCIM data, which is enhanced with machine learning to spot anomalies and patterns, optimize operations, and predict and forecast.

Some DMaaS vendors are using general-purpose AI engines (such as IBM's Watson IoT) that include cognitive-learning capabilities such as identifying data patterns that are not part of a pre-set data model to produce analysis that is data-center specific. The outputs are reports and actions, such as alarming and recommendations, tailored for an individual customer's data center.

Do you agree that AI will be widely used in data centers?



Source: Uptime Institute Research, 2018, n=265

As in other areas of operational technology AI, DMaaS is initially being used to improve existing functions by widening alarm lead times, spotting anomalies, predicting risks with greater accuracy, and so on. Over time, additional capabilities and services are highly likely.

Micro data centers are coming (within 5 years)

The Internet of Things, or IoT, (and the advent of robust cloud computing) has led to new approaches to edge computing and the deployment of new types of device and software. These include IoT gateways, which often are modified edge routers with industrial-type interfaces that translate proprietary protocols to IP and can be a link between sensor data and wide-area networks.

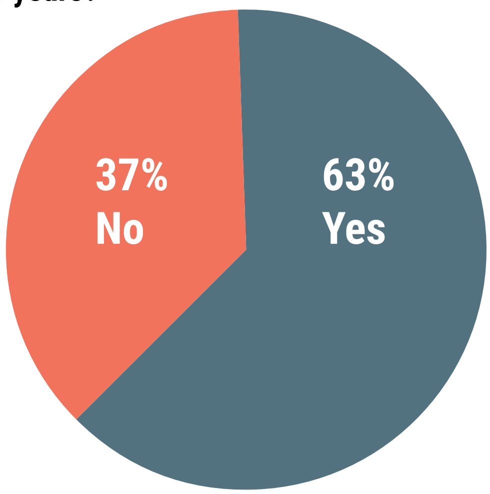
IoT gateways are part of the layer of first-line data ingestion and processing, where raw, streamed data is aggregated, translated, and normalized. Not all IoT applications will require a gateway, but most IoT data will need to be manipulated in at least one software platform and/or will require higher-level functions, such as data blending and orchestration (for example, to concurrently analyze multiple data feeds)—which will require either thick IoT gateways, with significant computing and storage capacities or dedicated data center capacity. Even if thick

IoT gateways are deployed, many believe dedicated data center capacity will still be required for (relatively low-cost) redundancy.

The need for data center capacity at the edge (close to where data is generated and/or consumed) is driving new and improved types of micro data centers. Nearly two thirds (63%) of respondents in our survey believe that within five years most of their customers will own small edge data centers (less than 100 kW.)

If our respondents' prediction (and our own bullish forecasts) are correct, the proliferation of micro edge data centers will not be at the expense of large data center growth. The second line of edge computing (sometimes called the core layer) is likely to happen at lower-cost centralized or public cloud data centers, where key edge data will be sent to be integrated, further analyzed, and probably managed, and visualized as part of a broad enterprise application or service (and then archived).

Do you believe that most of your customers will own very small (less than 100 kW) edge data centers within 5 years?



Source: Uptime Institute Research, 2018, n=266

There is likely to be many different owners and operators of micro data centers at the edge. Public clouds, colo providers, and telcos will likely have the largest appetite to own and operate data centers across both edge and core layers, along with an elite group of big enterprises that have mission-critical requirements. Micro edge data centers are likely to act as scalable extensions to their large-scale core facilities. But as the IoT sector evolves and matures, we expect that the lines of edge data center ownership will become increasingly blurred.

Staffing challenges: more robots, more women?

In our recent Eighth Annual Uptime Institute Data Center Survey of 867 data center operators and IT practitioners (globally in 2018), it appeared that most operators are not paying enough attention to the sector's skills shortage.

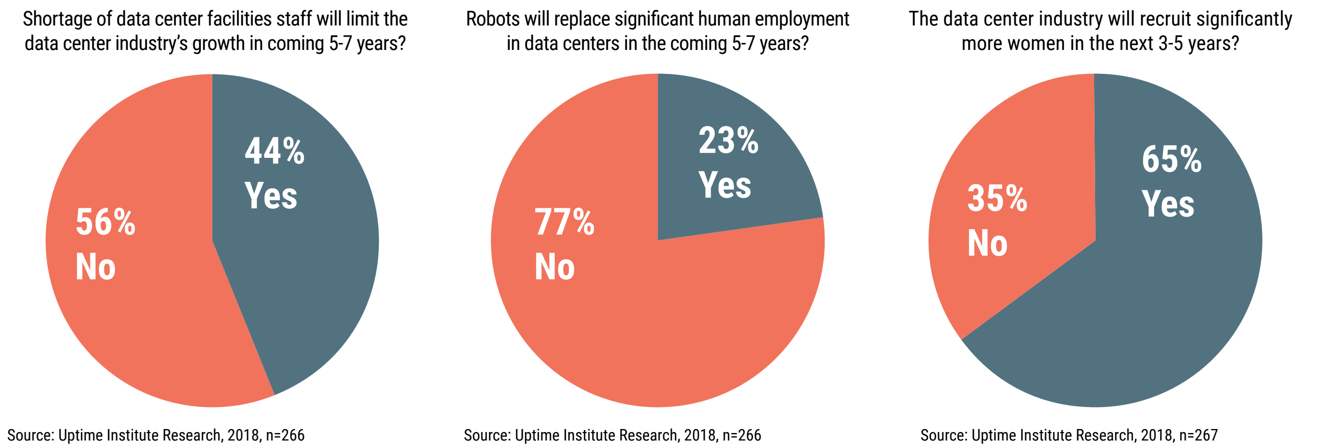
This is an aging industry; more than half of end users had more than 20 years' of work experience. Only 5% were new to the industry, having fewer than five years' of experience. Most operators told us they are struggling with staffing issues. Just 35% of these end users reported that they did not have any of the hiring or staffing issues we identified. Roughly the same number had recent staff cuts, were expecting staffing cuts, or, conversely, were having difficulty finding qualified candidates for open jobs.

The industry's workforce is also overwhelmingly male: 56% of respondents said women make up less than 6% of their data center design, build, or operations staff. However, 70% said a lack of women in the data center profession is not a threat to their business or the industry at large, but 38% reported struggling to find qualified candidates for open jobs.

This raises a central question: How will this fast-growing industry adequately meet staffing requirements in the future? We asked our supply-side respondents for their perspective on some related areas:

- Will a shortage of data center facilities staff limit the sector's growth in the coming five to seven years? Just a small majority (56%) believe it will.
- Will robots replace significant human employment in data centers in the coming five to seven years? Most (77%) said no.
- Will the data center industry recruit significantly more women in the next three to five years? Most (65%) said it will.

In the short term, vendors, designers, and consultants expect their customers to experience a staffing pinch, but robots and more women in the industry may help to address the problem.



Each of the three questions were asked separately; they were not grouped under the guise of potential staffing problems or solutions. Yet taken together they may be reasonably interpreted as a snapshot of the sector's near-term future state of employment: staffing issues will intensify and will be met, at least in part, by a combination of robotic technology and a concerted effort to attract more women.

The practitioners in our end-user survey mostly did not believe a lack of diversity in their ranks was an issue to be concerned about; 70% said a lack of women in the data center profession is not a threat to their business or the industry at large. It appears that data center design engineers, consultants, and suppliers are taking a more pragmatic view—that the future success of the data center business will depend on building a diverse workforce. It appears that the supply-side of the industry take an opposing view to their customers at least according to our data, and believe that their customers will need to hire more women.

In Summary

There were some divergences between some of the results uncovered across Uptime Institute's 2018 supply-side and operator surveys, notably the potential threat of data center skills shortages to the sector's growth.

For the most part, however, the thinking with regard to most other major trends across the two groups was aligned, including:

• Hybrid IT capacity approaches are creating additional complexity.

• Data center energy efficiency continues to be a focus.

• To meet edge computing demands, new data center capacity is expected to be built/installed.

• DCIM is becoming mainstream.

• There is an apparent increase in the use of N+1 data center designs, as opposed to 2N.

In our ongoing research and in future surveys, we will track the degree to which these trends play out over time.

Uptime Institute Global Data Center Survey

Authors

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, Ms. Ascierto 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.

Andy Lawrence is Uptime Institute's director of Research. Mr. Lawrence has built his career focusing on innovative new solutions, emerging technologies, and opportunities found at the intersection of IT and infrastructure.

He is the author of reports covering data center management, power/energy management, 5G mobile networks, disruptive data center technologies, and data center and application resiliency. He has represented 451 Research at the Green Grid and contributed to European Union projects on cloud services and renewable energy.

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